

# **DEMAND DRIVEN OPERATIONS MANAGEMENT IN MOTION PICTURE EXHIBITION**

by

Katherine Goff Inglis

Master of Spatial Analysis, Ryerson University, Toronto, 2008

Bachelor of Environmental Studies, York University, Toronto, 2002

A dissertation presented to Ryerson University in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Mechanical & Industrial Engineering

Toronto, Ontario, Canada, 2017

© Katherine Goff Inglis 2017

## AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners. I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my dissertation may be made electronically available to the public.

**Title: Demand Driven Operations Management in Motion Picture Exhibition**

**Degree: Doctor of Philosophy in the Program of Mechanical & Industrial Engineering  
at Ryerson University, 2017**

**By Katherine Goff Inglis**

**Abstract.** New opportunities for operational efficiency in movie exhibition exist as a result of recent developments in the industry, such as the mass-scale conversion to digital cinema, the explosion of customer data sources, and the availability of new channels for watching movies. This dissertation begins with an industry overview and discussion of these trends. A review of existing research on forecasting and scheduling problems in movie exhibition is utilized to identify eight factors for decision support systems for operations management in motion picture exhibition. A prototype decision support system (DSS) is constructed using customer loyalty and point of sale data from Cineplex Entertainment. The DSS that is built considers six out of the eight decision support factors that are identified through four modules. The first module projects audience composition for new release movies using loyalty data and identifies a set of the most important movie attributes for each age group. The second module leverages output from the first module with additional data to forecast ticket sales for individual theatre locations. In constructing the second model three different methods (gradient boosted regression, random forest and traditional multiple regression) are tested and the best performing method is utilized in the DSS. The third module applies the attendance forecasts to make labour recommendations. The fourth module uses output from the second module and extends the micro forecast with a concession food demand forecast which is applied to labour recommendations. The

DSS is tested empirically in two different ways; firstly forecasted box office sales are compared to actual box office sales to demonstrate that the forecasts being produced are reasonably accurate. Secondly, the labour recommendations from the DSS are compared to the recommendations from the existing DSS at Cineplex and against the theatre manager's schedules. The DSS performs better than the current schedule and the theatre manager schedule. The labour recommendations with and without the fourth module are compared to demonstrate the incremental value of the concession food module. The study concludes by highlighting further opportunities to extend the system as well as context for practical applications in the motion picture industry.

## **Acknowledgements**

This study was made possible with data from Cineplex Entertainment.

## Table of Contents

Abstract.....	iii
Table of Figures.....	viii
List of Nomenclature & Acronyms.....	x
Chapter 1. Introduction: Theatrical Industry Trends and Research Objectives .....	1
1.1. Theatrical Industry Overview.....	3
1.2. Consumer Behaviour in a Digital Era of Movie-going .....	5
1.3. Decision Support Factors .....	10
1.4. Research Objectives .....	13
Chapter 2. Literature Review: A review of the quantitative models relevant to Demand Driven Operations in Motion Picture Exhibition .....	18
2.1. Introduction.....	18
2.2. The Micro-Forecasting Problem.....	21
2.3. Movie Scheduling .....	28
2.4 Scheduling Model Extensions.....	30
2.5. Genetic Algorithms and Movie Scheduling.....	32
2.6. Summary.....	33
Chapter 3. Methodology: Four Modules that make up the labour recommendations decision support system for movie exhibitors .....	35
3.1. Audience Composition Models.....	36

3.2. Box Office Micro Forecasting Model .....	41
3.3. Labour Recommendations.....	45
3.4. Concession Micro Forecasting Extension.....	47
<b>3.5. Summary.....</b>	<b>49</b>
Chapter 4. Results: What drives different types of people to the movies and measuring performance of the DSS .....	51
4.1. Audience Composition.....	52
Results by Demographic Model .....	53
Model Results by Independent Variables .....	57
4.2. Box Office Micro Forecasting .....	60
New Release Forecast.....	62
Second-Week & Hold-Over Forecasts.....	69
4.3. Labour Recommendations.....	71
4.4. Concession Micro Forecasting.....	76
4.5. Illustrative Example of Data Outputs for a Single Location and Day .....	77
4.6. Summary.....	80
Chapter 5. Conclusion: Discussion for use of DSS in practice, summary of contributions and future research opportunities .....	83
References.....	92

## Table of Figures

Figure 1.1 - Domestic Theatrical Demand (derived using data from MPAA Theatrical Market Statistics 2012) .....	4
Figure 1.2 - 3D Share of Domestic Box Office Revenue (derived using data from MPAA Theatrical Market Statistics 2012) .....	4
Figure 1.3 - Prototype decision support system flow-chart and data outputs for business applications .....	15
Figure 3.1 - Data preparation summary for audience composition dataset .....	36
Figure 3.2 - Data preparation summary for Box Office Micro forecasting dataset .....	42
Figure 3.3 - Data preparation overview for labour recommendation analytical dataset .....	46
Figure 3.4 - Data preparation overview for concession forecast & labour recommendations extension .....	48
Figure 4.1 – Adjusted R Square by Customer Segment Model .....	53
Figure 4.2 - Standardized coefficients significant at the 90% confidence level for independent variables across 11 customer segment models .....	59
Figure 4.3 - New release forecast accuracy comparison between three methods at the location/week level of detail .....	62
Figure 4.4 - New release forecast accuracy comparison between three methods at the location/day level of detail .....	63
Figure 4.5 - New release forecast accuracy comparison between three methods at the location/day/film level of detail .....	64



Figure 4.6 - New release forecast accuracy comparison between three methods by weekday at the location/day/film level of detail.....	65
Figure 4.7 - Independent variables significant at the 90% confidence level in the new release micro forecasting model for a selection of locations .....	70
Figure 4.8 - Distribution of new release accuracy for all locations .....	71
Figure 4.9 - Forest model accuracy by weekday for week 2 films .....	72
Figure 4.10 - Forest model accuracy by weekday for hold-over films .....	73
Figure 4.11 - Forest model accuracy by release week for hold-over films.....	73
Figure 4.12 - Sum of demand forecasts for new release, second week and hold-over films depicted as sum_score_forest plotted against the actual seats sold.....	74
Figure 4.13 - Sum of demand forecasts for new release, second week and hold-over films depicted as sum_score_forest plotted against the actual seats sold for three select locations .....	75
Figure 4.14 - Topline labor recommendation delta (actual vs. predicted by the model) based on median (top) and average (bottom) .....	76
Figure 4.15 - Labor recommendation effectiveness - model versus manager and system ...	77
Figure 4.16 - Summary of model's recommendations by Job Function (labelled as Team (group)) .....	78
Figure 4.17 - Food station labour recommendations results without concession extension .....	80
Figure 4.18 - Food station labour recommendations results with concession extension .....	80
Figure 5.1 - Flow of models in DSS mapped to data output and business applications .....	86

## List of Nomenclature & Acronyms

- i. **Moviegoer** - a person who attends movies in the cinema
- ii. **Cinema** - the location where movies are shown, used interchangeably with 'multiplex' and 'theatre' in this paper.
- iii. **Multiplex** - a cinema which has multiple auditoriums
- iv. **3D** - refers to a movie being shown in three dimension, movie-goers wear special glasses to see the effects
- v. **IMAX** - a name brand of multiplex auditorium with enhanced audio, visual and production techniques
- vi. **DSS** - acronym for decision support system
- vii. **Theatrical release** - refers to the first date that a movie is available in cinemas
- viii. **Holdover** - refers to movies playing in cinema after their second week of release
- ix. **Play constraints** - refers to constraints on consecutive movie play imposed by distributors
- x. **Concession** - refers to food offerings in cinemas, typically fountain drinks, popcorn and candies
- xi. **Pre-release** - refers to the period of time before a film's release
- xii. **Premium Offering** - refers to innovative and enhanced experiences in cinemas that moviegoers pay more including IMAX and 3D
- xiii. **Loyalty program** - a program that customers can opt into where they are rewarded for behaviour usually with a specific company or group of companies
- xiv. **MPAA** - acronym for Motion Picture Association of America
- xv. **Cineplex** - the largest movie theatre chain in Canada, used interchangeably with Cineplex Entertainment
- xvi. **Regal** - One of the largest movie theatre chains in America
- xvii. **Netflix** - A company that customers subscribe to in order to stream unlimited movies to any personal device or home entertainment system

- xviii. DCIP** - The largest digital cinema integration company.
- xix. EWH** - acronym refers to the study by Eliashberg et al 2008
- xx. Macro vs. Micro** - in this study macro refers to the entire country, cinema chain or continent, micro refers to an individual cinema location or show time
- xxi. Crossover** - refers to the consumers who frequent multiple different cinemas to see movies

## **Chapter 1. Introduction: Theatrical Industry Trends and Research Objectives**

Motion picture exhibition is a popular research topic with a number of different perspectives publicized in mainstream media and academic journals. The research focuses on demand forecasting and operations management problems in theatrical exhibition, a rich area for practical applications of Operations Research techniques. Several researchers have noted the similarity between scheduling movies on multiple screens to scheduling jobs on parallel machines and moreover, the similarity to scheduling human resources to specific job functions. However, the movie-screen problem has unique constraints, such as play period restrictions, where the movie has to play for a minimum number of weeks, and it has to play consecutively. More notably, movie scheduling problems are different from machine scheduling because the value of the job (movie) is not fixed (Dawande et al. 2010). Movie scheduling algorithms require demand forecasts to produce results and the quality of the results is dependent on the quality of the forecast. Demand forecasting for movies, a problem in the marketing science domain, has garnered a significant amount of academic interest largely due to the presence of publicly available box office data. The movie-screen scheduling problem has been the focus of many studies with regard to both film distribution and exhibition. Since the industry has been widely studied, some review papers exist which provide overviews of the various topics and research directions relating to motion picture (i.e., Moul 2005, Weinberg 2006, Eliashberg et al. 2006, McKenzie 2012, Goff Inglis & Zolfaghari 2017). This work utilizes customer behaviour and point of sale data to develop a system for demand forecasting and that can be applied in the areas of

cinema labour, marketing and advertising scheduling. This work also extends the current body of literature on movie exhibition with a five year theatrical exhibition industry overview, a review of recent trends in the digital era of multiplex management and a new integrated approach to demand driven labour recommendations for large cinema chains.

The digital era has caused significant shifts in the movie-going industry, both in consumer behaviour and theatrical distribution and exhibition. On one hand, the availability of digital movies on demand and through a variety of channels has resulted in fierce competition for movie exhibitors. On the other hand, the shift from 35mm in the theatrical industry has the potential to impact movie scheduling significantly. The most recent review published relating to motion picture describes models in motion picture distribution, exhibition and home video, and differs from the other review work because it focuses specifically on models that are feasible for practical application in the movie industry (Eliashberg et al. 2008 – referred to as EWH from here on). EWH conclude this work by describing emerging trends and opportunities for future research in the motion picture industry. This provides a most interesting preamble for an industry overview, given that much has changed over a short period of time in motion picture. EWH highlighted five emerging trends:

1. Increased focus on bottom line performance
2. Availability of multiple distribution outlets (digital)
3. Decline in US theatrical movie attendance
4. Competition from other forms of media
5. Increased worldwide focus (globalization of supply and demand)

The first has to do with a fundamental shift in pre-production decision making due to increases in investor driven financing. As a result of this shift, there is more demand for

quantitative methods early in the life cycle of a movie to identify success factors, profitability drivers and identify potential box office hits or flops. The second trend they describe (multiple distribution outlets) has developed significantly in the motion picture industry over the past few years. EWH suggested quantitative models to recommend omni-channel distribution strategies based on content and address new cannibalization challenges. Over the past decade, this trend has changed the industry, largely due to explosive consumer adoption of tablet devices, digital video on demand services (i.e., Netflix) and moviegoer loyalty programs, like Regal's Crown Club or Cineplex's SCENE card. This trend is discussed further in the next section.

### **1.1. Theatrical Industry Overview**

Global box office receipts totalled a record high 34.7 billion in 2012 after five years of consistent growth (MPAA 2013). A significant part of this growth can be attributed to international box office revenue growth. While Hollywood maintains its dominant position, new trends are unfolding. In 2012 China became the world's largest international theatrical market, having surpassed Japan (MPAA 2013). Within North America, Hollywood's dominance is being challenged with an explosion of ethnic (i.e., Bollywood) and independent product<sup>1</sup>. New technologies and web platforms, make it easier for independent producers to gain crowd-sourced funding for their projects. This is resulting in more movies for exhibitors to choose from, and thus more competition among studios.

---

<sup>1</sup> The number of films released in 2012 grew 11% (total 677) however the number films released by MPAA members was down by 9%, while the number of independent films was up by 17% (MPAA 2013)

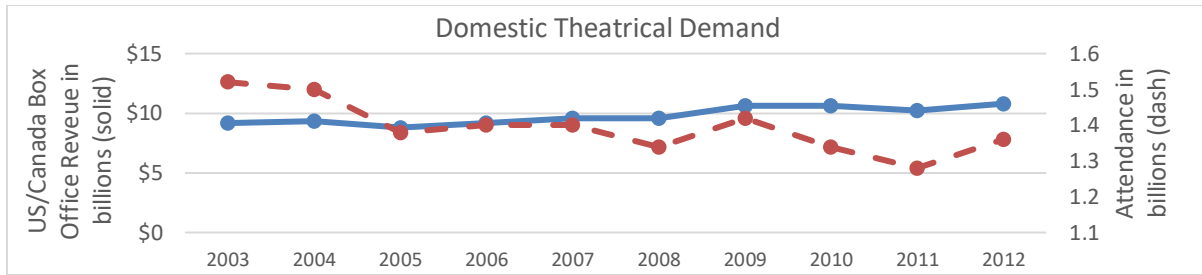


Figure 1.1 Domestic Theatrical Demand (derived using data from MPAA

Theatrical Market Statistics 2012)

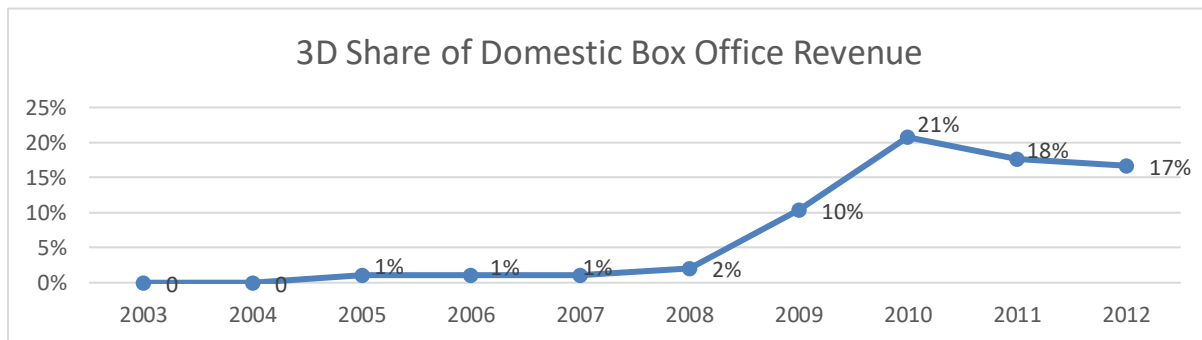


Figure 1.2 3D Share of Domestic Box Office Revenue (derived using data from

MPAA Theatrical Market Statistics 2012)

In 2012, domestic box office revenue hit an all-time high of \$10.8 billion with a 6% increase in attendance (1.36 billion) compared to 2011. As depicted in Figure 1.1, box office revenues have been trending upward over the last decade, but attendance levels are more volatile with changes in film product and consumer demand. While the adult ticket price remained stable, downward attendance trends are not reflected in revenue trends because of increase in 3D product, commanding a premium on the ticket price and accounting for almost 20% of box office revenue over the past 3 years. The attendance

spike in 2009 was prompted by a surge in 3D revenues (depicted in Figure 1.2) and the smash box office hit, *Avatar*, in 3D.

## **1.2. Consumer Behaviour in a Digital Era of Movie-going**

Theatrical movie-going has been a favourite pastime of North American and international leisure seekers for decades. It's known in industry and academia that not all moviegoers are the same, and that there are homogeneous sub-segments that exist with differing behaviors and attitudes. One of the most notable ways to segment moviegoers is by frequency (total visits to the movie theatre in a month or year). High frequency moviegoers (people who go 1x/month or more) drove 57% of US/Canada box office revenue in 2012 (MPAA 2013). Segmentation by age, gender and ethnicity are also a common practice in both industry and academic studies.

Several recent trends in consumer technology are changing the way people watch movies. Netflix and digital video on demand are arguably the largest disruptions to theatrical movie-going, giving consumers instant access to entire libraries of movies on demand and across devices. Parallel to this, home theatre systems continue to improve and tablet devices make cloud based movie libraries accessible from virtually anywhere. Moreover, piracy continues to threaten the industry as a whole - cloud based technologies and pocket gadgets make it possible for movie pirates to disseminate digital bootlegs to millions with the click of a button. Some studies have been done around piracy in movies (i.e., Ho & Weinberg 2011, Danaher & Smith (2014), McKenzie & Walls 2016), but there is still a need for further research to better understand and combat it.

Among several actions, the motion picture industry is responding to the changing landscape by shortening release windows and offering new theatrical experiences.



Approximately two thirds of the world's screens are digital, with one half of them being 3D capable (MPAA 2013). Over the past decade, the motion picture industry has undergone a mass scale conversion from analog (physical movie reels) to digital (virtual prints). For distributors, there are significant cost savings to produce and transfer digital prints versus large film reels. Exhibitors, alike benefit in a digital era, with savings resulting from operational efficiencies such as in production booth operations. In an analog world, moving large film reels between auditoriums was a cumbersome process and film reels had to be rewound manually and required close supervision. In a digital world, one audio visual technician can supervise projection on multiple screens. Despite these savings, the conversion to digital is costly, involving the replacement all 35mm theatre projectors. DCIP, the largest digital cinema integrator, facilitated the transition to digital for the majority of exhibitors in US and Canada, using a Virtual Print Fee (VPF) mechanism to help exhibitors finance the cost of equipment. From a scheduling perspective, the digital era presents new opportunities, and thus new management decisions for exhibitors, because digital prints can easily and dynamically be moved between screens.

New theatrical experiences available in a digital environment command premium prices and are centered around the latest technology offerings like IMAX and 3D, as well as upgraded audio and video. There is also an emerging stream of alternative content, such as opera, theatre, concerts, sports, being streamed digitally for moviegoers to enjoy on the big screen. These digital experiences are coupled with service offerings where movie goers can do things like reserve their seat in advance, receive in-seat food service and choose from gourmet food and beverage options. Customers are able to make purchases online or through mobile applications and use digital tickets presented on their phones instead of

paper. As EWH suggested several years ago, the study of different content channels and their consumers has grown to become even more relevant today, but there is still a scarcity of academic studies in this area.

For quantitative scheduling models to impact the theatrical industry they must be used in practice. This is especially relevant, because a dominant perspective in the motion picture industry recognized by many scholars (i.e., Wierenga et al. 1999, Eliashberg et al. 2001) is one that favours intuition-based decision making. This perspective is largely justified by the artistic nature of films, but also by the dynamic, complex, relationship-driven environment that distributors and exhibitors operate in, which makes model implementation challenging. In recent years, the non-analytical perspective has been countered by a growing collective of academics and industry champions alike, who are successfully demonstrating that, despite dynamic complexities, the industry's decision making environment can be modeled quantitatively.

Complexities around contract structures, new film forecasting and play constraints have been explored in the current literature on scheduling. Spatial interdependencies between multiplexes have been explored in the literature on location research, dealing with topics such as location-allocation strategies and cannibalization (i.e., Davis 2005, Davis 2006, Chisholm & Norman 2012, Hermann 2012). Quantitative models promise exhibitors and distributors the potential to maximize the value of their assets using technology and analytics to allocate resources more effectively.

Eliashberg et al. (2009) demonstrate an “evolutionary approach to decision support systems” (DSS) for practical implementation of quantitative models in the theatrical industry. The demonstration of the DSS is documented based on a collection of academic

case studies with the largest exhibitor in The Netherlands, Pathé. Eliashberg et al. (2009b) observed a 60/40 split between model recommendation and manager intuition. The implementation of the DSS at Pathé resulted in a significant impact (totalling almost \$1M on an annual basis) to the bottom line. This example is promising for taking an analytical perspective to decision support in theatrical exhibition. Three trends are identified below that are leading to a growing number of analytics champions in the motion picture industry.

(1) **Big Data, Big Technology.** There has been an explosive growth in the term “big data” which refers to the mass amounts of data that are being collected every day in the digital world from non-traditional sources like social media, sensors and scanners. Characterised by the 3V’s (volume, velocity and variety), big data’s requirements for more powerful computing power have been answered by emerging and leading technology vendors, and the usage of predictive analytics has grown to become mainstream in business. Accordingly, there is an increased interest, and buzz, around quantitative models in the motion picture industry. Perhaps the most famous example of this is the Netflix competition where hundreds of data analyst teams competed to develop an algorithm that would result in an incremental increase in the Netflix recommendation engine accuracy (PR Newswire 2008). Concurrently with the explosion of big data and big technology, the motion picture industry’s shift to digital projection opens up a plethora of new management questions for distributors and exhibitors alike. With digital projection and distribution, it becomes far more feasible to move film prints between screens and to distribute digital copies or stream digital feeds. This feasibility could

theoretically enable a five screen multiplex to operate like a ten screen multiplex by showing different prints on the same screen at different times of day, or different days of week. This suggests that exhibitor-distributor contracts could evolve, a new avenue for future research, and an opportunity that both exhibitors and distributors could benefit from. Distributors can push out more film content (i.e., product from their historic libraries), or alternate content (i.e., streaming of sports events, concerts, arts) to exhibitors, enabling them to serve a broader range of consumer preferences. Furthermore, with 3D technology and premium auditorium experiences there is the question of how to best schedule films across these different offerings (i.e., should a movie be scheduled in 2D, 3D or both?). In this digital era, there are more management decisions to make than ever and thus, more opportunities for DSS.

(2) **Competitive landscape.** The landscape is shifting and exhibitors are in more competition with in-home entertainment and the availability of multiple distribution outlets (a trend described by EWH). As described previously, digital on demand services and the availability of multiple outlets are giving movie lovers more choice around where, when and how they consume film product. With shortening release windows making for shorter shelf life of films, exhibitors are turning to mathematical modeling techniques to help make better decisions and maximize the profitability of their assets.

(3) **Omni-Channel Marketing.** Omni-channel marketing has drastically changed the way films and other entertainment experiences are communicated to consumers. New ways of connecting and engaging with movie-fans, using social media and

direct messaging to their emails and cell phones, have proven to impact film and television performance so much that media ratings giant, Nielsen, has started to incorporate social analytics as a standard audience metric (Vanderbilt 2013).

Furthermore, in this omni-channel world, analytics and quantitative models play an important role in developing and executing integrated marketing communication strategies, providing a 360 degree view of the consumer and helping marketers ensure that film products are promoted to the right people, at the right time, using the right channel.

In the next section, the key decision support factors, which movie exhibitors must consider when operating in a digital era, are summarized.

### **1.3. Decision Support Factors**

Based on the existing literature and using a market driven approach based on an analysis of common attributes of the world's most successful cinema networks, this paper identifies eight decision support factors that impact movie exhibitor operations management. These factors differentiate movie scheduling problems from commonly studied scheduling problems in the operations research domain. These factors can in theory, be modeled quantitatively for the purpose of creating decision support systems to improve movie scheduling. The eight factors are presented and defined below.

(1) **Location-specific attributes.** Not all multiplex movie theaters are the same; variances exist in terms of the brick-and-mortar space, the composition of the population in the primary, secondary, and tertiary trade areas, and the size and

distance of those trade areas. These variances between individual locations are often correlated with variances in consumer demand.

(2) **Auditorium attributes.** Within an individual multiplex, variances often exist between auditoriums in the same building, which impact demand and scheduling. Seating capacity refers to the number of seats in an auditorium, which typically varies between auditoriums in a single complex. Seating type and the possibility of reserved seating (i.e., booking a specific seat as one would for a sporting event) can vary between auditoriums within a single location. Technology can also vary between screens; some auditoriums offer bigger and better screens, louder and clearer sound systems, and the possibility for 3D projection.

(3) **Showtime schedule.** Scheduling of showtimes varies between multiplexes but can also vary within a multiplex depending on the day of the week. With digital projection capabilities, changing the content (movie) playing after each showtime within a single auditorium is theoretically possible.

(4) **Film product competition (cross-elasticity).** Cross-elasticity is a measurement of how much one product's demand shifts as a result of another product's demand. In movie exhibition, this could measure how the demand for one movie changes based on another movie's allocation of space (Raut et al. 2009).

(5) **Network considerations.** Most multiplex cinema locations are concentrated in major urban markets. Hermann (2012) demonstrated empirically that crossovers exist between locations. Consumers in urban markets may choose a location based on the

movie they want to see or the showtime convenience. The availability of premium offerings, such as those described in the auditorium attributes factor, can also be a driver of crossover visits between theater locations. For large networks of multiplexes, such as Cineplex Entertainment in Canada with its 163 theaters and 1,666 screens (Cineplex 2016), scheduling decisions must account for cannibalization between locations to ensure cumulative profit maximization.

**(6) Price elasticity (premium and discount offerings).** Premium theatrical experiences command premium prices and include upgraded offerings, such as IMAX and 3D, advanced seat reservations, and in-seat food service. Premium offerings and discount offerings (e.g., cheap Tuesdays) both cannibalize demand for standard offerings and have varied profit margins; therefore, both should be incorporated into demand forecasts and scheduling models to accurately account for overall profits.

**(7) Multiplex staffing.** Labor scheduling requirements in a multiplex are similar to movie scheduling requirements, because they are driven by box office demand; however, they are unique in that they have the additional complexity aspect of operational factors, such as janitorial, guest, and food services, which become increasingly complex with premium offerings (e.g., in-seat service and upgraded food offerings).

**(8) Marketing personalization.** Digital on-demand services and the advancement of mobile and in-home technologies are giving movie lovers more choices on where, when, and how they consume film and entertainment content. The shift to digital projection equipment has increased the availability of more varied entertainment content in

cinema. New digital channels of communication, such as social media and short-message service (SMS), have resulted in an omni-channel world—a world in which companies aim to provide their customers with integrated messaging across multiple communications channels. Companies in the theatre industry must consider this when promoting movies to consumers. In the digital landscape, identifying granular trends in customer behavior to define customer communication strategies, which provide the right message to the right customer at the right time, is necessary. Personalized communication strategies aim to maximize profits through more efficient scheduling of marketing communications.

In Chapter 5 - Conclusion, the eight factors are revisited as a framework to evaluate the existing literature and identify new research opportunities. The next section details the objectives of this research and describes the organization of the balance of the paper.

#### **1.4. Research Objectives**

This research utilizes a market driven perspective to develop a demand driven labour recommendations model. To provide real world context, data from a major motion picture exhibitor will be used for the case study. The exhibitor is Cineplex Entertainment, Canada's largest movie chain and an industry leader in theatrical exhibition with one of the most modernized circuits in the world. The operations management decisions faced by Cineplex are nontrivial, with a national network of multiplexes having the number of variables and constraints can easily be in the millions for a short term planning horizon. In the era of big data and analytics, a customer driven perspective to business strategy is essential. Customer loyalty, research and social media programs run by most major theatrical exhibitors are rich sources of consumer behaviour data that can be leveraged to



inform business operations. The research opportunity to explore applications of operations research techniques with movie exhibitor customer behaviour data is large, however the proliferation for academic studies is limited due to the proprietary nature of the data. In the private sector, complex analytics solutions are required to effectively manage a movie exhibitor chain, especially in the areas of omni-channel marketing strategy, advertising sales and multiple labor staffing. Model implementations such as that at Pathé in the Netherlands (Eliashberg et al. 2009b), show that a balance between managerial input and quantitative methods provide the best results, and fuel the growing perspective that favours the use of quantitative models to maximize revenue in theatrical exhibition.

This research utilizes customer loyalty data to derive approaches for audience and film segmentation that are applied to the micro forecasting and labour scheduling problems. The overarching objective is to construct a decision support system which utilizes both management input and data to derive micro demand forecasts and audience composition projections that can be utilized to improve the scheduling of cinema labour staffing, marketing communications and advertising. Using customer loyalty data, concession purchase behaviour is modeled and applied to refine the labour recommendations model to account for variation in concession behaviour as it relates different types of movies. This is a fruitful area for exhibitors, because concession sales offer much higher profit margins versus box office. No existing studies address the exhibitor's labour scheduling problem or micro concession demand forecasting. Theatre staffing is a complex problem involving staffing several customer touch points (i.e., box office, guest services, core concessions, ancillary concessions, bar/lounge, games and VIP

food service). The decision support system is tested empirically, leveraging micro attendance forecasts to make labour recommendations and the results are compared to the benchmark system schedule and the actual schedule used by managers. Through the application of the decision support system the potential upside of utilizing a concession forecasting extension in conjunction with the micro box office forecasting model is demonstrated.

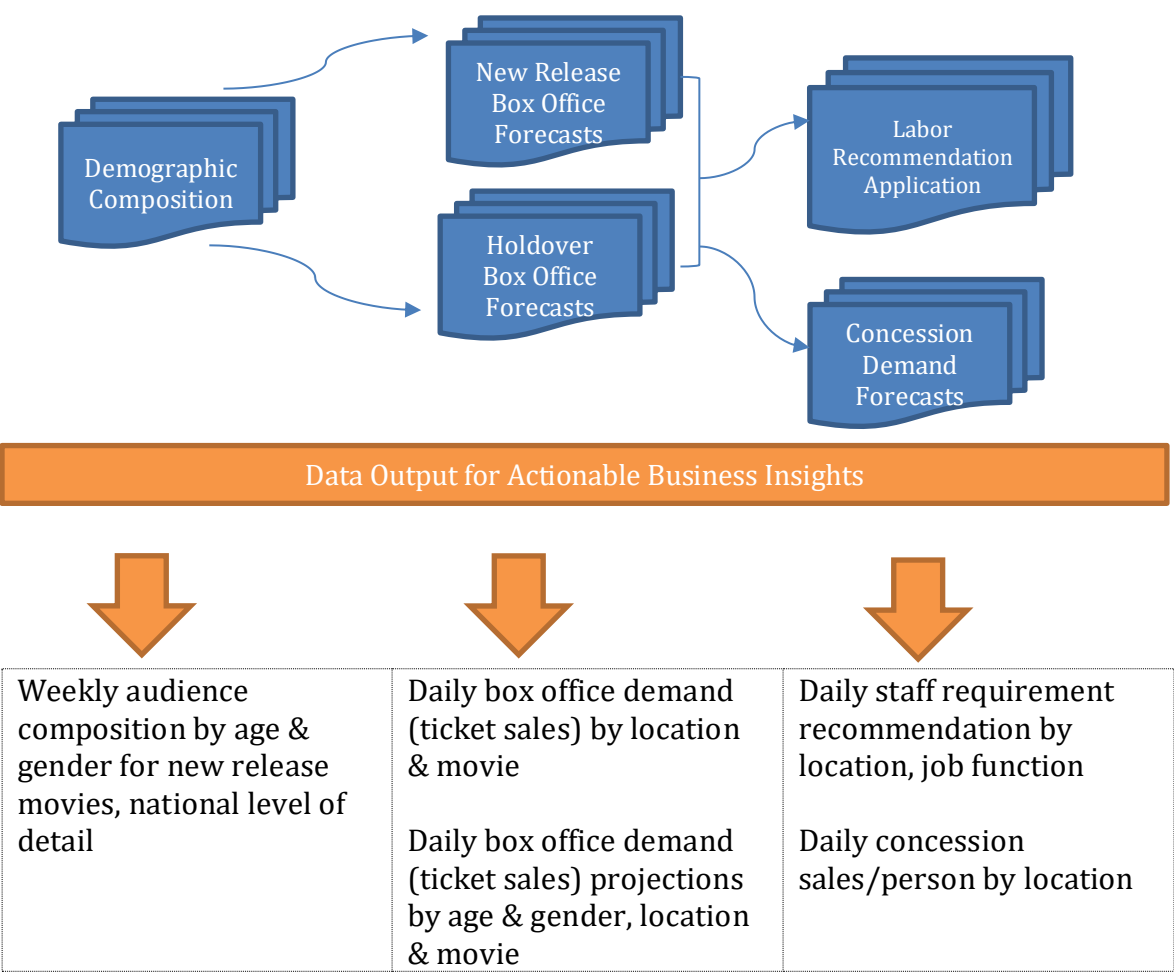


Figure 1.3 - Prototype decision support system flow-chart and data outputs for business applications

Figure 1.3 illustrates the flow of models that make up the decision support system and outlines the data output from each step in the process. In summary, the key objectives of this research are:

1. Build a prototype decision support system for demand driven recommendations that can be applied to scheduling of cinema labour, advertising and marketing communications.
2. Utilize consumer behaviour data to develop micro box office demand forecast and audience projection models.
3. Develop a multiplex theatre concession food demand micro forecasting model which accounts for moviegoer, film and theatre specific factors.
4. Conduct empirical analysis of prototype decision support system forecasting accuracy and labour recommendations effectiveness.

The balance of this paper is organized into four chapters. Chapter 2 is a literature review broken out into five sections. The first section provides an overview of quantitative models in the academic literature that pertain to the motion picture industry. The second section provides an overview of the academic work that has been done on box office forecasting models and the difference between macro and micro forecasting problems. The third section details the basic model for movie to screen scheduling. The fourth section details scheduling model extensions and the fifth discusses alternate approaches (genetic algorithms) for solving the movie to screen scheduling problem. Chapter 3 details the methodology for the demand driven labor labour recommendations system that was built starting with the details of the data sets that were used. The system has four components and Chapter 3 has a section for each. The audience composition model is detailed first - it utilizes loyalty program data to predict the key demographics for new release movies. Next

the box office micro forecasting model is detailed. It uses output from the audience composition model, film attributes and past box office sales to forecast demand for individual dates and cinema locations. Then the labor recommendations model is detailed. The output is the number of staff recommended by work station type. Lastly in Chapter 3, the micro concession food demand model extension is detailed. The results are presented in Chapter 4 broken out in the same sections as Chapter 3. Finally Chapter 5 is the concluding section where limitations of this research are highlighted along with opportunities for future research organised into the eight decision support factors defined in this chapter.

## **Chapter 2. Literature Review: A review of the quantitative models relevant to demand driven operations management in Motion Picture Exhibition**

### **2.1. Introduction**

To provide broad context in the domain of OR problems, the key areas in which movie scheduling problems differ from classic OR challenges are identified. As previous studies (Swami et al. 1999) demonstrate, traditional OR approaches, with some modifications, can be utilized to solve movie scheduling problems. Two common OR problems exemplify how the movie scheduling problem differs:

(1) Retail labor scheduling: The labor scheduling problem for movie exhibitors is much like the classic retail labor scheduling problem when working under the assumption that consumer demand for a given movie is known. The various work stations within a theater (e.g., box office, concession, food) are parallel to the work stations within a store (e.g., produce and cashier at a grocery store). Moreover, the standard constraints used in retail labor scheduling apply to movies, such as minimum and maximum number of hours an employee can work, employee availability, and employee vacation time. The key difference between the multiplex staffing problem and the retail labor scheduling problem is in the demand forecast. Retail stores, such as grocery stores, are able to utilize seasonal trends to compute demand forecasts; however, this is not feasible for multiplexes in which the unique attributes of movies (e.g., actors, genres, and critical acclaim) are key drivers of consumer demand.

(2) Job scheduling on parallel machines: Several movie-industry researchers have noted the similarity between scheduling movies on multiple screens and scheduling jobs on parallel machines. However, the movie screen problem has unique constraints, such as play-period restrictions, in which the movie must play for a minimum number of weeks, and the requirement that it must play consecutively. More notably, movie scheduling problems differ from machine scheduling problems because the value of the job (movie) is not fixed (Dawande et al. 2010).

In theatrical exhibition, demand forecasting has been widely studied on a macro level, looking at box office estimation models and relevant factors (e.g., Smith and Smith 1986, Litman 1983, Sawhney and Eliashberg 1996, Simonoff and Sparrow 2000), and a micro (location) level (e.g., Chisholm and Norman 2002, Somlo 2005, Somlo et al. 2011, Swami et al. 1999). Some studies approach the problem from the distributor's perspective, while others approach it from the exhibitor point of view. Exhibitors have a unique opportunity because they can easily move pictures between auditoriums using analytical models to plan and adjust to changes in the market (Sawhney and Eliashberg 1996). We define the types of forecasting and scheduling problems that movie exhibitors face as the Exhibitor's Scheduling Problem (ESP).

ESP has many facets and challenges, making it a fruitful area for research. Generally, the goal of the ESP is to maximize financial gain (e.g., film profit, concession sales attendance, box office revenues) through a better allocation of film products; however, this could also be achieved through a better allocation of labor and marketing communications. The ESP problem has been addressed in the literature on both macro and micro levels, with

empirical studies (e.g., Eliashberg et al. 2001) demonstrating that quantitative models yield an incremental lift in attendance. The macro level addresses the problem of selecting which films to play at a location and for how many weeks (Dawande et al. 2010, Somlo 2005, Somlo et al. 2011, Swami et al. 1999); the micro problem involves identifying the time segments and auditoriums in which to play films in a multiplex (Eliashberg et al. 2009a). The movie screen scheduling problem has also been modeled as a multiple-period shelf-space optimization problem factoring for product perishability, pricing contracts, and cross-elasticity (Raut et al. 2009). To solve micro-scheduling problems (in which optimization occurs within a given cinema location), micro forecasting of demand at individual theatre locations is required. For organizational purposes, the models that were reviewed in this study are grouped into three subtopics, which are summarized in Table 2.1.

Table 2.1 - Quantitative models reviewed grouped by three subtopics: macro forecasting, micro forecasting, and movie scheduling.

Movie scheduling subtopic and authors	Macro forecasting	Micro forecasting	Movie scheduling
Ainslie et al. (2005)	X		
Asur and Huberman (2010)	X		
Chisholm and Norman (2002)		X	
Dawande et al. (2010)			X
Einav (2007)	X		
Elberse and Eliashberg (2003)	X		
Eliashberg et al. (2001)			X
Eliashberg et al. (2009a)			X
Eliashberg et al. (2009b)			X
Gaytán-Iniestra et al. (2006)			X
Litman (1983)	X		
Litman and Ahn (1998)	X		
Radas and Shugan (1998)	X		
Raut et al. (2009)			X
Ravid (1999)	X		
Sawhney and Eliashberg (1996)	X		
Sharda and Delen (2006)	X		
Simonoff and Sparrow (2000)	X		
Smith and Smith (1986)	X		
Somlo (2005)		X	X
Somlo et al. (2011)		X	X
Swami et al. (1999)		X	X

## 2.2. The Micro-Forecasting Problem

Forecasting is the basis for profit maximization, the objective function of the movie scheduling problem. Box office revenue is interrelated with the scheduling of movies to screens (Elberse and Eliashberg 2003); consequently, the most practical forecasting and scheduling models combine these processes. For new product releases, such as movies, forecasting can be viewed as a two-part process: postrelease forecasting (with the availability of some data points following film release) and prerelease forecasting. Prerelease forecasting involves using classification methods, nonanalytical methods, or



some hybrid of the two to match or group newly released movies with historical movies. Both have been addressed in the literature and are discussed in this section.

Prerelease forecasting is particularly challenging and can be a time-consuming task for movie industry managers to do manually (Eliashberg et al. 2009b). In practice, it is common to match new release movies with past movies using a combination of manual heuristics and film industry expertise. In the literature, the challenge of prerelease forecasting has been tackled using the data-driven approach with movie classification methods that compare the attributes of newly released movies to the attributes of movies released in the past. In motion picture decision support systems, new-release movie segmentation can be either a manual or an automated process, depending on the requirements of a given practical situation. Either way, it is typically done by segmenting historical movies into groups, based on their decay rate and opening strength, and then determining which new releases match best with which segments. In the manual process, managers assign new releases to categories using intuition-based heuristics; in the automated process, new releases are assigned to the category that is most likely based on defined factors. Jededi et al. (1998) define a four-class scheme that other studies, such as Raut et al. (2009) also use: Type 1 movies open with high demand and decay rapidly; Type 2 movies open with high demand and decay slowly; Type 3 movies open with low demand and decay rapidly; and Type 4 movies open with low demand and decay slowly. In their micro-forecasting process, Eliashberg et al. (2009b) utilized a 16-segment classification scheme, with four classes of opening strength and four classes of decay rate.

Macro-forecasting models involve forecasting aggregate commercial success, and many researchers have studied these models; examples include Neelamegham and

Chintagunta (1999), Sawhney and Eliashberg (1996), Litman and Ahn (1998), and Zufryden (1996). The most common movie-forecasting models use simple regression analysis with various independent variables to predict box office revenue success (Litman and Ahn 1998, Ravid 1999). The factors that contribute significantly to box office success, along with other movie features, are used as the basis of movie matching or classification for prerelease forecasting. Lilien and Rangaswamy (2005) match new releases to historic releases using movie characteristics, such as genre, Motion Picture Association of America (MPAA) ratings, lead actors, warnings, and sequels in a prerelease box office revenue-forecasting model. The market share model, which Ainslie et al. (2005) developed, found that other movies with the same MPAA rating or genre have an adverse effect on market share.

The exploration of the presence of violence and objectionable content in popular media is the subject of the book *Hollywood vs. America: Popular Culture and the War on Traditional Values* written by film critic Michael Medved (1992). De Vany & Walls (2012) explore the same topic empirically in their work titled “Are There too Many R-Rated Movies?”. Their results suggest that movie studios produce a large number of R-rated movies, despite G and PG movies being more profitable and affirm findings from the study by Ravid (1999), which concludes that G and PG ratings are positively correlated with film success. De Vany & Walls (2012) and Medved (1992) suggest that studio rationale for making so many R-rated movies is driven by an artistic endeavor to earn insider respect and praise in the movie industry.

A number of studies have demonstrated that positive critics’ reviews are related to movie performance (e.g. Eliashberg & Shugan (1997), Basuroy et al. (2003)). A positive

correlation with film performance and production budget has also been demonstrated empirically (e.g. Ravid (1999), Litman (1983)). Seasonality and market share are key factors that have been modeled in many studies for pre-release box office forecasting. Observed seasonality was the focus of some earlier studies, such as Radas & Shugan (1998). Aggregate box office performance varies significantly by season, but shows stable seasonal trends year over year.

Pre-release buzz in the entertainment industry is an interesting field of study with an emerging stream of research using social media data to predict box office performance. Westland (2012) discusses various studies that focus on various niches in social media adoption for cinema releases and investigates the effectiveness of social networking on film success using event studies techniques to analyse Google trends. Key findings show that social media activities<sup>2</sup> targeted around theatrical release increase revenue by 64% and search activity by 48%. This corroborates similar findings from a study done by Asur & Huberman (2010), where they utilized Twitter data to forecast box office revenue, achieving improved accuracy over market-based predictors. The Hollywood Stock Exchange (HSX) is another, public, example of using crowd sourced data to predict film performance – models based on its auction data achieve very high forecasting accuracy (Westland 2012).

The basic model for forecasting gross box office revenue (Sawhney and Eliashberg 1996) draws on queuing theory to conceptualize the moviegoer's adoption process. It provides a foundation that many studies have utilized in their work and have extended to

---

<sup>2</sup> Such as (1) official film websites (studio sponsored) (2) viral marketing using movie trailers and (3) rapid distribution of content using email and blogs (Westland 2012 P2)

account for additional factors. Part of the analysis that follows is adapted from their work and from the review of it by Eliashberg et al. (2008). Using the first two to three weeks of box office data, the basic model calibrates fairly accurate forecasts, that is, accuracy rates of  $\pm 10$  percent in most cases. However, it is not successful in predicting the prerelease forecast. The model is the combination of two behavioral processes modeled independently over time, accounting for consumer behavior and movie-specific factors. Specifically, time-to-decide behavior (i.e., if moviegoers want to see the movie) and time-to-act behavior (i.e., if moviegoers decide to see the movie at the theater) are modeled as independent stochastic subprocesses with stationary parameters. This is based on the premise that after a decision has been made to see a particular movie, the next decision is where to see it. The time-to-decide behavioral subprocess is influenced by movie attributes, such as genre, star power, and marketing budgets. The time-to-act parameter is the key differentiator between the basic model and simple regression models, which consider movie attributes, and factors in distribution and characteristics of consumers, such as moviegoing behavior, free time, and distance to the theater. The authors propose to extend the model by making the time-to-act parameter proportional to the number of screens, using either a share of the total-screens metric or an absolute number of total screens to control for the relationship between distribution intensity and time to act.

Somlo et al. (2001) developed a relevant extension to the basic model by adding location attributes to the time-to-decide parameter. These authors use a multiple-step regression model to estimate model parameters and empirically test it. They use (1) nonlinear regression models to estimate parameters of the extended model, which includes theater characteristics, and run the model using a historic data set of location-specific

movie performance; (2) multiple regression models to link historic data with movie and theater attributes, resulting in a function to estimate model parameters for a given set of theater and movie attributes; and (3) parameter estimates to forecast box office revenue over time for each potential location. They identify the following key challenges with forecasting location-specific box office revenue over time:

- (1) Understanding which movie attributes and theater attributes affect box office revenue, what the relationships among them are, and how the relationships change over time; and
- (2) Understanding the individual's moviegoing decision-making process and how to incorporate this process into the model.

The challenges are pertinent to the ESP, and could be addressed using models based on consumer-insight data derived from consumer research surveys, loyalty programs, and social media. In addition, an opportunity exists for further model segmentation to factor in premium offerings, such as 3D, IMAX, and licensed dinner and movie offerings, which allow moviegoers to enjoy comfortable, reserved seats in an adults-only environment with upscale food and beverage offerings. Furthermore, the mass volume of social media data could be referenced geographically to further extend the time-to-decide parameter. The extensions developed by Somlo et al. (2001) and Eliashberg et al. (2001) refine the basic forecasting model, making it more realistic for practical use in solving exhibitor operations problems. The bottom-up approach proposed by Somlo et al. (2011) better forecasts macro box office revenue based on distribution and adoption patterns.

Theatrical forecasting models must factor in both prerelease forecasting and postrelease forecasting for use in practice. The prerelease forecast is critical because new-release movies typically account for the majority of box office revenue each week.

Conversely, the postrelease forecast is critical for leveraging the best possible data. Several researchers have demonstrated that model accuracy improves significantly by using actual data points versus comparable titles; see Jedidi et al. (1998), Lehmann and Weinberg (2000), and Sawhney and Eliashberg (1996).

Swami et al. (1999) propose a prerelease forecasting model using data from matching movies. However, in practical implementation, manager estimates are used for prerelease forecasting (Eliashberg et al. 2001). Managers cannot feasibly provide estimates for all locations; see Eliashberg et al. (2009b), which focuses on the micro-scheduling problem. Consequently, the authors developed a 16-class segmentation system, which they used, in conjunction with a theater-specific regression model, to forecast demand by film, location, and hour. Sharda and Delen (2006) define nine classes of movies based on gross box office revenue and use neural networks to predict class membership for prerelease films. Additionally, they compare neural networks against other types of classification models, including logistic regression, discriminant analysis, and classification decision trees. Results show that the neural network approach gives greater classification accuracy. In Redondo and Holbrook (2010), the authors show that heterogeneous moviegoer groups with varying demands for specific movie features exist. The existence of heterogeneous groups of moviegoers is of particular relevance to the micro-forecasting problem in which groups of moviegoers vary between theater locations. Furthermore, variations in the types of moviegoers at a theater location could have a significant implication for concession food-category sales. Modeling the segment of moviegoers and their behavior at both the box office and concession stand at the micro level is critical for cinema operations management, and provides a rich pool of data for future research, because exhibitors have much higher

profit margins on concession food than on movie tickets. Some studies (e.g., Swami et al. 1999) incorporate concession forecasts into box office sales; however, they assume a uniform purchase incidence across locations and films.

### **2.3. Movie Scheduling**

Basic scheduling model: Macro scheduling is the method used to allocate films to screens. It differs from the micro-scheduling problem in that it does not consider showtime allocation. In the ESP, the objective is to maximize revenue for a network of multiplex cinemas; conversely, the distributor's scheduling objective is to maximize revenue for specific films. The SilverScreener model (Swami et al. 1999) addresses the macro-scheduling problem using an integer programming model to schedule movies in a multiplex and maximize profitability. It is aimed at helping exhibitors determine which movies to play at a multiplex and the number of weeks during which they should play those movies. The SilverScreener model handles standard constraints required for movie-to-screen scheduling. These constraints include consecutive play, which ensures that a movie is played only in consecutive weeks, and capacity, which ensures that the number of movies scheduled does not exceed the number of available screens. The SilverScreener model assumes:

- (1) Titles for the upcoming season are known in advance;
- (2) Weekly revenues can be estimated in advance;
- (3) Replacement decisions are made on a weekly basis;
- (4) All screens in the multiplex are of equal capacity; and
- (5) No lag time exists between placing an order for a new movie and its arrival.

Swami et al. (1999) relax the second assumption by incorporating forecasts into the adaptive approach and utilizing a revenue-prediction matching model to support decisions in film planning. The objective function is to maximize profit over the season; however, the model can also be used for weekly decisions on whether to hold over existing movies for another week using the adaptive approach and rolling weekly time windows. The empirical analysis of the SilverScreener model leverages 1989 box office revenue from a six-auditorium multiplex in New York City; the results showed that the model could help exhibitors increase profit by choosing fewer right movies (i.e., movies that will generate the most demand at a given location) and running them longer. A key limitation of the SilverScreener is screen capacity (the model assumes that all screens are the same size); however, the authors recommend some modifications to the model to handle this. They point out that, as a result of these modifications, different sets of constraints are required to ensure exhibitor business rules are met, and that the resulting model would be quite complex.

Swami et al. (1999) propose another research direction that is highly relevant to the ESP—competitive issues in which demand varies when theaters in close proximity to each other play competing titles, versus playing the same titles. Krider and Weinberg (1998) and, more recently, Chisholm and Norman (2002) also explore competitive issues; however, as Swami et al. (1999) point out, incorporating these models into an exhibitor's weekly decision-making model is challenging; however, it might be possible if the exhibitor has some historical data that could be leveraged to estimate competitive effects on demand.



## 2.4 Scheduling Model Extensions

Eliashberg et al. (2001) implemented a special instance of SilverScreener at Pathé. The model ran each week to determine which movies should be played. It was a special implementation because only one week of the model's recommendations were used and the results were updated each week with new data. In this implementation, Eliashberg et al. (2001) relaxed the assumption that equal screen capacity uses a simple heuristic to recommend film-to-screen allocation based on forecasted attendance (i.e., the model recommends putting the movies with the highest demand into the largest auditorium, and scaling down). It accounts for competitive considerations (e.g., the effect of movie-choice location on nearby locations) by using input from Pathé management (Eliashberg et al. 2001).

Gaytán-Iniestra et al. (2006) developed an adapted version of the SilverScreener model, which accounts for the capacity of auditoriums and market-specific requirements, such as a minimum audience requirement and no obligatory period. Four constraints ensure the following conditions: (1) minimum audience size is met and forecasted demand does not exceed screen capacity; (2) only one movie is scheduled per screen; (3) movies are played in consecutive weeks; and (4) the decision variable (e.g., whether the movie plays) is binary.

Somlo et al. (2011) developed a macro-scheduling integer programming model utilizing empirical theater-level forecasts to solve the distributor's location-selection problem (DSLPP). The DSLPP is a complex problem, because it includes hundreds of locations and movies. Accordingly, Somlo et al. (2001) used a heuristic to solve this problem. The researchers provide a variation of the integer programming model developed by Swami et

al. (1999) by factoring in business rules around competitive play zones in which only one location of a set of locations can play a given film. The objective function is to maximize box office profits for a movie by making an optimal choice of theaters using the following constraints, which must be satisfied for all movies:

- (1) Total number of theaters is less than or equal to the maximum input parameter;
- (2) Total number of theaters in each region is greater than equal to the minimum input parameter;
- (3) Competitive theaters do not book the same film;
- (4) All variables are equal to 0 or 1.

They evaluate the heuristics by computing upper bounds using a linear programming relaxation; however, they prove that this may not be a feasible solution, because the relaxation results in an integrality constraint. The authors propose myopic and greedy heuristics to solve the problem; the former relaxes all constraints and iterates until a feasible solution is found, and the latter searches for the most profitable theaters with least amount of competition. The sensitivity test they performed showed that minimum play length (i.e., between two, three, or four weeks) had a big impact on the optimal theaters, but did not show a big shift in the number of locations. When the play length was shorter, the model recommended small neighborhood theaters in mid-income areas instead of multiplexes in higher-income areas. When play length increased, the total number of theaters decreased from 740 to 670, and the model favored large multiplexes over small ones. Their model suggests a potential lift of 12 percent when benchmarked against the distributors' current method, which utilizes a mathematical model developed by Somlo (2005) to determine minimum play length.

Raut et al. (2009) extend the basic model by introducing more realistic characteristics, such as product demand perishability, pricing contract, and film product competition (cross-elasticity). This is achieved by adding more constraints. The researchers model cross-elasticity using an  $N$  matrix (where  $N$  is the total number of movies) and compute the forecasted incremental profit and loss due to cross-product elasticity. The function is incorporated through a movie-type input, and movie demand and point in time are other inputs to the profit model. The results show a potential 11 percent lift compared to the SilverScreener model in Swami et al. (1999) (Raut et al. 2009).

Dawande et al. (2010) also modeled the macro-scheduling problem and estimate the implications of the consecutive-play constraint from the exhibitor perspective, looking at profitability between consecutive- and nonconsecutive-play policies. They show that the consecutive-play problem is solvable in polynomial time and solve it using a minimum-cost flow problem (each path represents a screen and  $N$  paths represent a schedule). The obligatory period is met by construction, and the best collection of flows through the network is solved as an integer problem. The results of the comparative analysis suggest that the nonconsecutive-play policy could result in incremental revenue of 44 percent, suggesting the need for new contract structures between exhibitors and distributors and further academic work to explore this issue.

## **2.5. Genetic Algorithms and Movie Scheduling**

Genetic algorithm (GA) methods are increasingly being used to solve complex optimization problems by mimicking biological evolution in a process that mirrors natural selection. Recent studies demonstrate that GA-based heuristics are able to solve complex exhibitor scheduling problems with significant computational efficiency and without

sacrificing model performance (measured in terms of proximity to the optimal solution). The work of Dawande et al. (2010) and Raut et al. (2009) provides a compelling foundation for further research in the niche area of genetic algorithms for movie scheduling. Dawande et al. (2010) develop a GA heuristic to solve the nonconsecutive scheduling problem, which performs favorably when benchmarked against optimal results. In Raut et al. (2009), the GA is compared to three greedy heuristic methods: (1) a simple revenue-based heuristic, (2) a product-based heuristic, which factors in film contracts and rents, and (3) a hybrid heuristic, which is both product and revenue based according to Eliashberg et al. (2001). The end result is that GA performance is within 2 percent of optimality with a faster solve rate (i.e., up to 10 times faster), versus the integer programming model (Raut et al. 2009).

## **2.6. Summary**

There is a wide body of academic literature that covers topics related to the motion picture industry, so this literature review focuses only on the subset that involves the development of quantitative models. Of these, the majority focus on developing models for macro (all of North America) box office projections which are relevant to this work because they prove that demand can be modeled (most movies follow an exponential demand curve). Also, they inform which independent variables are potentially relevant to predicting demand in the context of this study (forecasting by individual cinema locations). The literature on box office forecasting models also emphasizes the importance of using separate quantitative techniques for demand forecasting for films before and after theatrical release when sales data becomes available.

In the scheduling domain, there is a concentration of studies in the niche area of movie to screen scheduling. The pioneering study in this niche describes the basic movie to screen scheduling model which can be used as a planning tool to set the total weeks of consecutive play for individual films in a single cinema. Several studies extend on this work by developing and testing methods to integrate different data sources and account for additional factors such as handling for variances in auditorium capacity and planning over large networks of cinemas to maximize revenue for a given film. More recently there are a handful of studies which demonstrate the potential of genetic algorithms for movie to screen scheduling - this approach is less computationally expensive than linear programming models but returns comparable results. The importance of computational efficiencies is emphasized when considering the micro scheduling problem which optimizes the allocation of movies to screens *and showtimes*, growing the number of variables exponentially.

At the beginning of this chapter the Exhibitor's Scheduling Problem (ESP) that the decision support system aims to address was described. The next chapter describes in detail the methodology and datasets used to build the prototype decision support system.

### **Chapter 3. Methodology: Four Modules that make up the labour recommendations decision support system for movie exhibitors**

This chapter details the data sources and the quantitative methods utilized. The next chapter provides the results for these models. The primary data for this study is sourced from Cineplex Entertainment and represents aggregate ticket sales and loyalty card usage from May 1 2011 to May 1 2014. Attributes of the relevant movie, theatre, consumers, dates and schedule were also obtained from Cineplex and have been summarized at the location, film and date level of detail. A series of Transact-SQL statements were developed and executed in Microsoft SQL Server to reduce the largest datasets. Following this, the more compacted data tables were connected to Alteryx Data Blending and Advanced Analytics software for further aggregation, transformation and integration. Alteryx software can execute custom R scripts in addition to having over 200 of the most commonly used R functions built into the user interface. The Alteryx procedure has been built out to run the models described in this section over a 52 week sample period, producing a total of 53,345 forecasts every run (each output record a unique location - film - day). Using Tableau Data Visualization software, dashboards have been developed to aggregate and disaggregate forecast results for analysis of overall module performance and accuracy of the individual models.

The details of the methodology are broken out into sections with each one corresponding to a different set of models. Each section details the preparation of the analytical dataset used for modeling, followed by the details of the methodology and specific variables used. The first section details audience composition models which predict the demographic audience composition for new release movies. The second section

details box office micro forecasting models and the differentiation between forecasts for movies that are already releases versus new release movies with no prior sales data. The third section details the labor recommendations rules-based model, including the concession micro forecasting model extension and how it is applied.

### 3.1. Audience Composition Models

Figure 3.1 shows the flow for creating the analysis dataset for audience composition models by combining a sample of data from three sources: 1) a film attribute dataset from a movie industry data solutions provider, Baseline Intelligence, 2) a production budget dataset sourced from the-numbers.com (The Numbers 2016) and 3) a transactional customer loyalty dataset (anonymized) from Cineplex Entertainment.

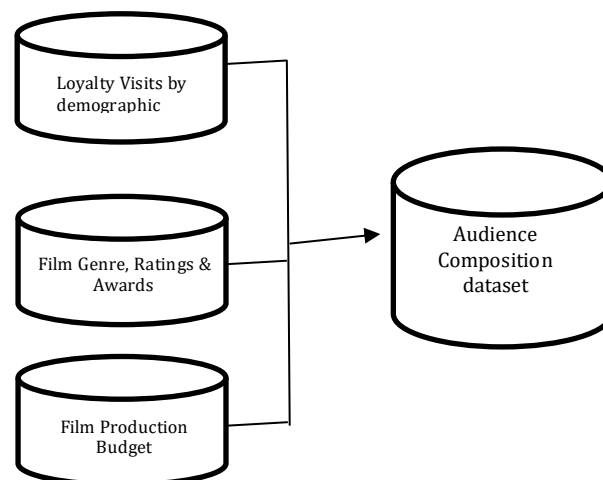


Figure 3.1 - Data preparation summary for audience composition dataset

The dataset includes 350 films released in Canada between April 29, 2011 and April 25, 2014. The release date utilized represents the wide release of the film in North America.

The loyalty data is limited to the film's first week of release because the intent of this study is to demonstrate a method that can be used when no sales data is available for a film. After the first week of release, sales data exists and can be used to inform the appeal of a film to segments of moviegoers. Each film in the dataset has at least 5,000 loyalty cardholders attending it during the first week. This rule ensures that small and limited release films are not included in the dataset and that there is a sufficiently large sample size for each age and demographic cohort group, even if the cohort group represents as little as 1% of the audience. Loyalty cardholders represent approximately 30% of the box office ticket sales for the major exhibitor who provided the data; the author believes that directional insights based on their behavior can be used as a proxy for all moviegoers, but also highlights the potential for some bias in the data due to the universe of loyalty program card holders having an over or under representation of certain segments of movie-goers. It is expected that the loyalty program over represents movie goers who have a high frequency of theatrical visits and under represents occasional movie goers. This bias is accepted for the purpose of the DSS built in this paper, since the DSS is being tested empirically, but the author recommends future work to investigate the bias further.

The titles in the dataset provide a representative sample of movies across seasonality, ratings, release strength and genre. The dependent variables used in model testing and development are percentages between 0 and 1 representing the proportion of the audience that each demographic group makes up. The gender and age of the loyalty cardholder is used to determine categorize behavior with the exception of one; the group titled "Prop\_Parents" represents the behavior of children and the adults that take them to the movies. Table 3.1 details all dependent variables used:



Table 3.1 - Itemized list of Dependent Variables for Audience Composition model

<b>Prop_Parents</b>	Represents the percent of loyalty member cardholders attending the film with children.
<b>Prop_M14_18</b>	Represents the percent of loyalty member cardholders attending the film that are males between 14 and 18 years old. Coincides with the age of secondary or high school students.
<b>Prop_F14_18</b>	Represents the percent of loyalty member cardholders attending the film that are females between 14 and 18 years old. Coincides with the age of secondary or high school students.
<b>Prop_M19_34</b>	Represents the percent of loyalty member cardholders attending the film that are males between 19 and 34 years old.
<b>Prop_F19_34</b>	Represents the percent of loyalty member cardholders attending the film that are females between 19 and 34 years old.
<b>Prop_M35_49</b>	Represents the percent of loyalty member cardholders attending the film that are males between 35 and 49 years old.
<b>Prop_F35_49</b>	Represents the percent of loyalty member cardholders attending the film that are females between 35 and 49 years old.
<b>Prop_M50_64</b>	Represents the percent of loyalty member cardholders attending the film that are males between 50 and 64 years old.
<b>Prop_F50_64</b>	Represents the percent of loyalty member cardholders attending the film that are females between 50 and 64 years old.
<b>Prop_M65P</b>	Represents the percent of loyalty member cardholders attending the film that are males aged 65 and up. Coincides with retirement age.
<b>Prop_F65P</b>	Represents the percent of loyalty member cardholders attending the film that are females aged 65 and up. Coincides with retirement age.

Independent variables are selected based on access to data and the academic literature and include: Genre (one movie can have more than one genre), MPAA rating, Film Runtime, Production Budget, Seasonality, Big Star Count and Award Nominations. Seasonal and

Ratings variables have been adapted from what was found in the literature to the Canadian market to align with the study dataset.

Table 3.2 details all independent variables used.

Table 3.2 - Itemized list of independent variables in analytic dataset for audience composition model

<b>Genre_Action</b>	Binary variable equal to 1 if film genre is tagged with action
<b>Genre_Adaptation</b>	Binary variable equal to 1 if film genre is tagged with adaptation, refers to some type of literature converted to a screenplay
<b>Genre_Adventure</b>	Binary variable equal to 1 if film genre is tagged with adventure
<b>Genre_Comedy</b>	Binary variable equal to 1 if film genre is tagged with comedy
<b>Genre_Drama</b>	Binary variable equal to 1 if film genre is tagged with drama
<b>Genre_Family</b>	Binary variable equal to 1 if film genre is tagged with family
<b>Genre_Horror</b>	Binary variable equal to 1 if film genre is tagged with horror
<b>Genre_Period</b>	Binary variable equal to 1 if film genre is tagged with period, refers to films set in a period of the past
<b>Genre_RomCom</b>	Binary variable equal to 1 if film genre is tagged with romantic comedy
<b>Genre_Romance</b>	Binary variable equal to 1 if film genre is tagged with romance
<b>Genre_Sci_Fi</b>	Binary variable equal to 1 if film genre is tagged with science fiction
<b>Genre_Sequel</b>	Binary variable equal to 1 if film genre is tagged with sequel, meaning that one or more films were previously released in a series of films
<b>Genre_Thriller</b>	Binary variable equal to 1 if film genre is tagged with thriller
<b>Production_Budget</b>	Production budget in dollars
<b>Film_Runtime</b>	Length of film in minutes
<b>Summer_May_Aug</b>	Binary variable equal to 1 if film theatrical release date is in May, June, July or August

<b>March_Spring_Break</b>	Binary variable equal to 1 if film theatrical release date is in March
<b>Valentines</b>	Binary variable equal to 1 if film theatrical release date is between February 8 and 14
<b>October_Halloween</b>	Binary variable equal to 1 if film theatrical release date is in October
<b>Winter_Holidays</b>	Binary variable equal to 1 if film theatrical release date is between Dec 24 and Jan 2
<b>Rating_14A</b>	Binary variable equal to 1 if the film is rated as 14A in the study geography, meaning that it's suitable for audiences 14 years or age and older
<b>Rating_18A</b>	Binary variable equal to 1 if the film is rated as 18A in the study geography, meaning that it's suitable for audiences 18 years or age and older
<b>Rating_PG</b>	Binary variable equal to 1 if the film is rated as PG in the study geography, meaning that it's suitable for all audiences but parental guidance is advised
<b>Award_Level_0</b>	Binary variable equal to 1 if the film does not have any award nominations
<b>Award_Level_1</b>	Binary variable equal to 1 if the film has 1-5 award nominations
<b>Award_Level_2</b>	Binary variable equal to 1 if the film has 6-24 award nominations
<b>Big_Star_Count</b>	Count of actors in the film who are in the top 100 films ranked by gross revenue for the study period

A total of 11 models are run to generate a demographic appeal score for each demographic segment. The dependent variable (ie: “Parent Appeal” for the movie “Hunger Games”) is in the range of 0-1, computed using loyalty program data and standardized to represent a % of total loyalty sales by demographic segment. The higher the score, the higher the appeal to that demographic group. Independent variables were selected for testing based on the work by Redondo & Holbrook (2010) which shows that different types of movies appeal to specific demographic groups. Relevant data points that will be tested

include film attributes such as genres, awards and film format (ie: IMAX). Stepwise regression will be used to refine input variables and generate demographic appeal scores. The resulting scores will be used as an input to subsequent models.

A data analysis workflow is built using the Alteryx Analytics platform and R programming language. A multiple regression approach is utilized to predict the appeal of movie features for each demographic group using the independent variables in Table 3.2, resulting in a total of 11 models. A multiple regression approach was chosen to present detailed results, as it offers a simplified output for interpretation of the significance and contribution of the predictor variables versus more advanced methods like neural networks. The initial regression equation is the same for each model and detailed in (1). In this equation Y represents the percent of the audience for a given demographic and gender segment.

$$Y = \alpha + \beta_1 \text{Genre\_Adaptation} + \beta_2 \text{Genre\_Adventure} + \beta_3 \text{Genre\_Comedy} + \beta_4 \text{Genre\_Drama} + \beta_5 \text{Genre\_Family} + \beta_6 \text{Genre\_Horror} + \beta_7 \text{Genre\_Period} + \beta_7 \text{Genre\_Romance} + \beta_9 \text{Genre\_RomCom} + \beta_{10} \text{Genre\_Sci\_Fi} + \beta_{11} \text{Genre\_Sequel} + \beta_{12} \text{Genre\_Thriller} + \beta_{13} \text{Production\_Budget} + \beta_{14} \text{Film\_Runtime} + \beta_{15} \text{Summer\_May\_Aug} + \beta_{16} \text{March\_Spring\_Break} + \beta_{17} \text{Valentines} + \beta_{18} \text{October\_Halloween} + \beta_{19} \text{Winter\_Holidays} + \beta_{20} \text{Rating\_14A} + \beta_{21} \text{Rating\_18A} + \beta_{22} \text{Rating\_PG} + \beta_{23} \text{Award\_Level\_0} + \beta_{24} \text{Award\_Level\_1} + \beta_{25} \text{Award\_Level\_2} + \beta_{26} \text{Big\_Star\_Count}$$

(1)

Detailed results presented in the next chapter show the final regression equations resulting from excluding predictors that are not significant at the 90% confidence level, which provides a unique regression equation for each customer segment.

### 3.2. Box Office Micro Forecasting Model

Figure 3.2 - Data preparation summary for Box Office Micro forecasting dataset illustrates the integration of six datasets to create an analytic dataset which is further split

out into 3 subsets based on film release week. The *New Release Dataset* contains data for films in their first week of release. The *Second Week Dataset* contains data for films in their second week of release. The *Holdover Dataset* contains data for films in the third plus week of release.

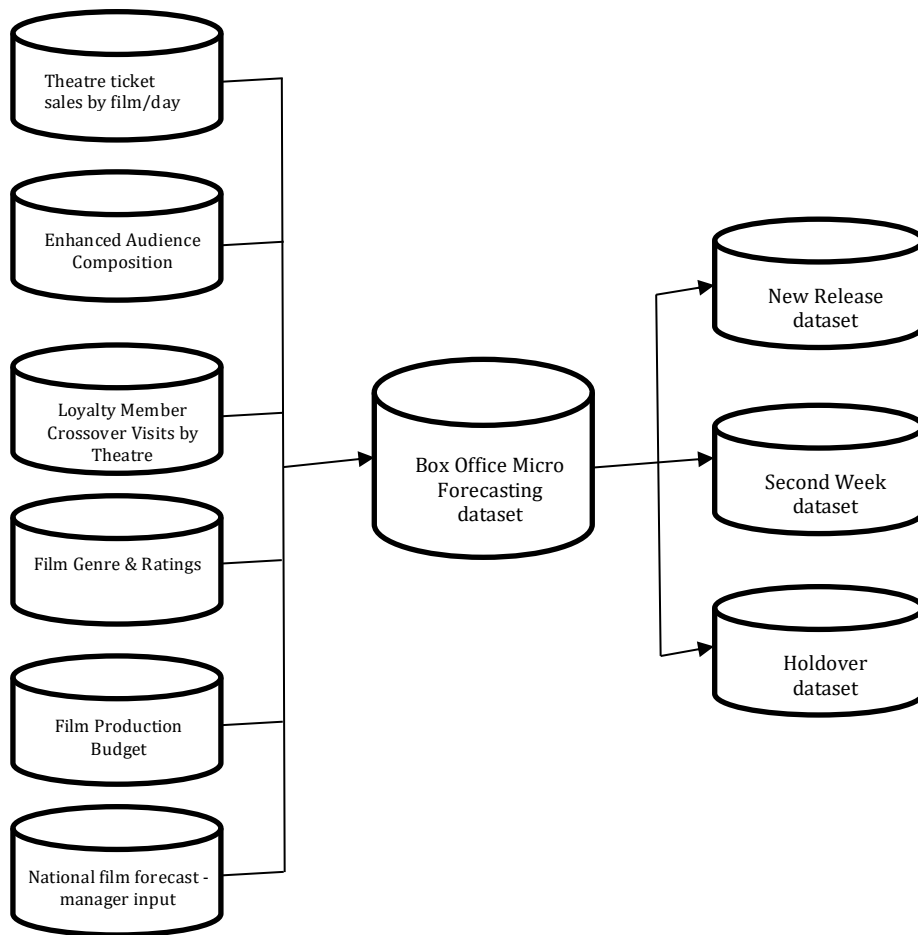


Figure 3.2 - Data preparation summary for Box Office Micro forecasting dataset

The itemized list that follows details all fields in the Box Office Micro Forecasting Analytic Dataset which are all present in the subsets for each release week grouped by their source dataset.

1. **Location Unique Identifier**
2. **Film Unique Identifier**
3. **Date**
4. **Weekday** - binary variables to indicate day of week (Friday is excluded).
5. **Tickets Sold** - dependent variable - total attendance across all shows for the location/film/date
6. **Seat Capacity** - total seats available across all shows for the location/film/date
7. **Number of shows** - total screenings of a film
8. **Concurrent auditoriums** - number of concurrent auditoriums playing the film
9. **Auditorium attributes**
10. **Tickets Sold Last Week** - Computed at the weekly level for each location/film based on data 1-7 days prior to the Friday starting date of the week.
11. **Tickets Sold The Week Before Last** - Computed at the weekly level for each location/film based on data 8-14 days prior to the Friday starting date of the week.
12. **Advanced Tickets Sold** - as of last Monday 4 days before
13. **Average Seasonal Attendance** - Seasonal function computed based on the average attendance for all films at the location for the past 2 years.
14. **First Week Gross Manager Estimate** - National level data. Manager estimate of national gross revenue for a film's first week 5 days prior to first release Friday. Estimates are made by a central team of film booking managers and based on industry experience and ad hoc analysis of past performance of comparable films.
15. **First Week Drop Manager Estimate** - National level data. Manager estimate of the film's weekly revenue percent change between week 2 and week 1, 5 days prior to first release Friday.
16. **Crossover locations playing film** - A count of the top 5 crossover locations that are also playing the film. Each location is assigned a top 5 crossover locations based on aggregate loyalty member behaviour across groups of theatres in the same network.
17. **Genre Competition** - Number of films with the same genre playing at the location.
18. **First Week Demographics** - National level data. Refer to Table 3.1 - Itemized list of Dependent Variables for Audience Composition model.
19. **First Week Predicted Demographics** - National level data. Predicted values for each dependent variable in Table 3.1 (this is the output from the audience composition model - refer to page 36).
20. **Film Genre, Ratings, Production Budget** - National level data. Refer to Table 3.1 - Itemized list of Dependent Variables for Audience Composition model.

Three predictive models are used to produce micro location attendance forecasts for 156 theatres belonging to a single exhibitor. Forecast granularity within a location is by film and business day. Forecasts are produced for a week at a time with Friday

utilized as the start of the week to align with new film release in North America. The models use the most recent data available as of Monday which includes box office sales data to the end of Sunday five days prior to the Friday start of the week. The rationale for selecting Monday is to allow time for theatre labor or marketing scheduling during the business week.

It is necessary to have three different models to compute the micro film forecasts to ensure that the model is always utilizing the most relevant data possible. Previous academic studies, most notably Sawhney & Eliashberg (1996), demonstrate that past box office performance is the best predictor of future performance, so this data is utilized as soon as it becomes available. For example, for movies in their second week the model leverages first weekend box office performance data to predict attendance for the balance of the week. The R function `randomForest` is utilized in all three forecasts, with an experiment in the new release forecast to benchmark the performance of the Forest model to multiple regression and gradient boosted regression. The Forest model is a machine learning methodology that predicts a dependent variable using an ensemble of decision tree models that are constructed using bootstrapping where each tree model is created using a random sample of the input data. For specific details on the R 'randomForest' package see the documentation by Liaw (2015). Gradient boosted regression is another machine learning technique that uses ensembles of decision trees to minimize a specific loss function. More details on each of the three models is provided below. For specific details on the R package 'gbm' see the documentation by Ridgeway (2017). More details on the method for each forecast are presented on the next page.

- i. **New Release Forecast** – The new release forecast utilizes predictions from the audience composition model, along with film attributes and distribution intensity to produce micro forecasts for films with no prior sales data available. Three different predictive models are tested: Forest, Boosted and Multiple Regression. The Forest model (R RandomForest method) produces the best results and is utilized to make labor requirements predictions.
- ii. **Second-Week Forecast** – For films going into their second week, the only available sales data is for opening weekend. Data from Friday, Saturday and Sunday is utilized with distribution intensity in a random forest model to forecast week two attendance. A random forest model is utilized to produce hold-over forecasts.
- iii. **Hold-Over Forecast** – Following the example set by Sawhney & Eliashberg (1996), past ticket sales and distribution intensity (ie: # of shows) are expected to have the strongest predictive power for attendance for films going into their 3<sup>rd</sup> + week of release. This model produces forecasts by location, day and film using historic sales data from the two prior weeks. A random forest model is utilized to produce hold-over forecasts.

Validation output and details on the specific variables used in each section is presented in the Results chapter.

### **3.3. Labour Recommendations**

The goal of the micro forecasting model is to minimize labour waste through demand driven staffing. Success of the micro forecasting model is dependent on forecasting accuracy; the ideal solution is one that has no discrepancy between the forecasted & actual optimization solutions. From a case study perspective, the model will be considered practical for production deployment if it performs better than the actual



staffing decisions that were made by theatre managers, and the recommended staffing decision made by the current software system. Theatre manager staffing decisions are made based on human judgement while the current software uses a retail forecast based on recent and seasonal sales trends. The human judgement of the theatre manager's typically considers what films are getting released in addition to sales trends while the system's forecast is unaware of the context of film performance.

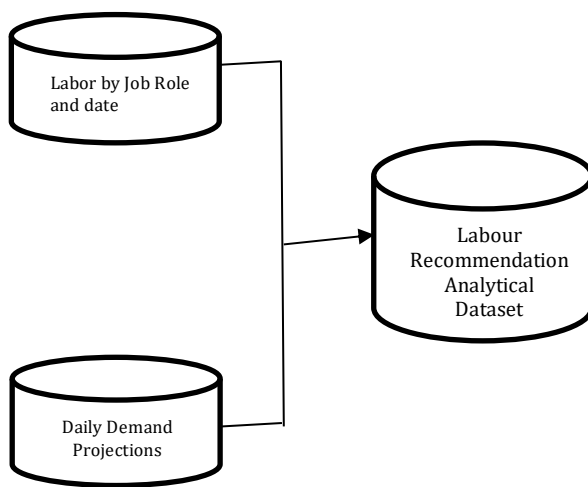


Figure 3.3 - data preparation overview for Labour recommendation analytical dataset

The itemized list that follows details the fields present in the Labour Recommendation Analytical dataset while Figure 3.3 shows the data preparation flow.

- 1. Location unique identifier**
- 2. Date**
- 3. Role** - the labor role (for example: Box Office, Concession, Booth, Front of House, Back of House, Floor, Guest Services)
- 4. Actual Admissions** - at the daily level of detail
- 5. System Forecasted Admissions** - at the daily level of detail
- 6. Manager Forecasted Admissions** - at the daily level of detail
- 7. Micro Attendance Model Forecasted Admissions** - the output of the model described in the previous methodology section
- 8. System Recommended Labor count** - at the daily & role level of detail

- 9. Manager Scheduled Labor count** - at the daily & role level of detail
- 10. System Scheduling Rule** - Represents the attendance required per labor head count by job role. Computed as  $[\text{System Forecasted Admissions}] / [\text{System Recommended Labour count}]$

Using the labour recommendation analytical dataset, an *Adjusted Schedule Rule* is developed to obscure and protect proprietary data from Cineplex by using the average of the *System Scheduling Rule* field for each location. Using this field additional fields are computed:

- A. Actual Required Labor** - Computed and rounded to the nearest whole number:  $[\text{Actual Admissions}] / [\text{Adjusted Schedule Rule}]$
- B. System Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{System Forecasted Admissions}] / [\text{Adjusted Schedule Rule}]$
- C. Manager Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{Manager Forecasted Admissions}] / [\text{Adjusted Schedule Rule}]$
- D. Model Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{Micro Attendance Model Forecasted Admissions}] / [\text{Adjusted Schedule Rule}]$
- E. System Delta Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{System Projected Labor}] - [\text{Actual Required Labor}]$
- F. Manager Delta Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{Manager Projected Labor}] - [\text{Actual Required Labor}]$
- G. Model Delta Projected Labor** - Computed and rounded to the nearest whole number:  $[\text{Model Projected Labor}] - [\text{Actual Required Labor}]$

Finally to measure effectiveness of the labor schedule the percent of cases where labor is predicted exactly and within one class is compared using fields E, F & G from the itemized list above.

### **3.4. Concession Micro Forecasting Extension**

The concession forecasting module is an extension of the box office forecasting & labour recommendations module discussed above that outputs an alternate value for '**Model Projected Labour**' for up to three potential food staffing areas within a location: 1)

café/yogurt/pretzels, 2) hot foods and 3) concession. Figure 3.4 depicts the data preparation workflow for the concession micro forecasting model extension, a description of the variables in each dataset follows.

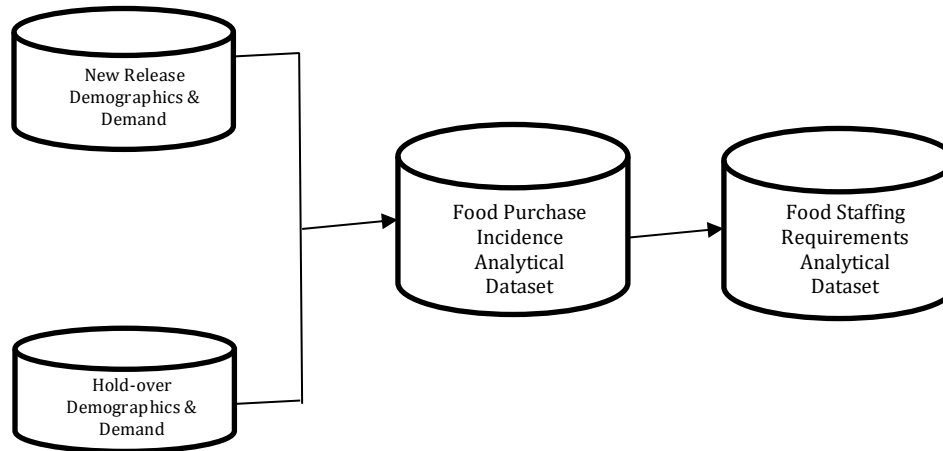


Figure 3.4 - data preparation overview for concession forecast & labour requirements extension

New Release Demographics & Demand Dataset Included fields:

1. Location
2. Film
3. Week Start Date
4. Predicted attendance (location)
5. Predicted audience composition (national)
6. Predicted new release demand by demographics (location): predicted audience composition national \* predicted attendance location

Hold-over Demographics & Demand Included fields:

1. Location
2. Date
3. Weekday
4. Last Week's demographics (at the location - weekly)
5. Predicted Demand - Holdovers
6. Predicted holdover demand by demographics (location): Predicted Demand - Holdovers \* Last Week's demographics
7. Predicted Demand - All films

8. Café/Yogurt/Pretzels - \$/visit
9. Hot Foods - \$/visit
10. Concessions - \$/visit

Food Purchase Incidence Analytical Dataset:

1. Location
2. Date
3. Weekday
4. Predicted demand for demographics (location): [Predicted new release demand by demographics] + [Predicted holdover demand by demographics]
5. Predicted Demand - All films
6. Café/Yogurt/Pretzels - \$/visit [Dependent]
7. Hot Foods - \$/visit [Dependent]
8. Concessions - \$/visit [Dependent]

Food Staffing Requirements Analytical Dataset:

1. Location
2. Date
3. Weekday
4. Actual Required Staff [Dependent]
5. Predicted Café/Yogurt/Pretzels - \$/visit
6. Predicted Hot Foods - \$/visit
7. Predicted Concessions - \$/visit
8. Predicted Demand - All films
9. Predicted demand for demographics (location): [Predicted new release demand by demographics] + [Predicted holdover demand by demographics]

For each location a separate forest model is created for each food area to forecast daily purchase incidence represented as the Spend/Visit. The daily purchase incidence forecasts are utilized in conjunction with additional predictor variables to predict the daily number of staff required by food station.

### **3.5. Summary**

Our prototype decision support system includes four modules that are built using proprietary and third party data sourced from Cineplex Entertainment. Audience Composition, the first module, consists of 11 multiple regression models that predict the

national (macro) audience composition for new release films by age cohort and gender. The dependent variable in these models is the percent of a film's audience for a specific age cohort and the predictor variables are film attributes. The second module is Box Office Forecasting where ticket sales for a specific date and location are forecasted using one of three model types (New Release, Second Week, Holdovers) depending on the film's point in its lifecycle. Three different methods are tested for New Release micro forecasts and the best performing method (Random Forest) is utilized for the Second Week and Holdover models. The Box Office forecasting models are built for individual cinema locations, a total of five models per location. In the third module, labour scheduling details (proprietary data from Cineplex) is anonymized by generating cinema location specific volume based scheduling rules by workstation type. These rules are applied to ticket sales forecasts to determine staffing recommendations by workstation type and date. The fourth module is a Concession Micro Forecasting extension which utilizes the output from the first two modules (Audience Composition and Box Office Forecasting) in addition to actual sales and demographic data to project the dollar spend per loyalty member visit by food type (Café/Yogurt/Pretzels, Hot Foods, Core Concessions). Food spend predictions are then used as independent variables in a Random Forest model where the dependent variable is the required staff. Similarly to the Box Office Forecasting module, concession models are built for individual locations with a separate model for each food type (Café/Yogurt/Pretzels, Hot Foods, Core Concessions). Lastly, the labour recommendations utilizing the concession extension are compared to the labour recommendations without it to determine and quantify additional gains resulting from the extension module. In the next Chapter the results for each module are presented.

## **Chapter 4. Results: What drives different types of people to the movies and measuring performance of the DSS**

This chapter presents the results for each of the four modules described in Chapter 3 - Audience Composition, Box Office Micro Forecasting, Labour Recommendations and Concession Micro Forecasting, and is organized into subsections accordingly. The first section on Audience Composition begins with a high level comparison of the Adjusted R Square values between the 11 models (one for each age cohort and gender) that make up the module. This is followed by the detailed results and analysis of coefficients statistically significant at the 90% confidence level for each model. The first subsection then concludes with a comparison of the significant coefficients between the 11 models highlighting similarities and differences between the movie tastes of demographic segments.

The second subsection on the Box Office Micro Forecasting model is organized in two parts: 'New Release Forecast' and 'Second-Week & Holdover Forecast'. For new release forecasts, each location has three models built to identify the best performing method between: multiple regression, boosted regression and random forest. The random forest model is identified as the best performing and is utilized for the Second-Week & Holdover Forecasts. In total 600 models (5 models x 120 locations) were built to arrive at the final forecasting module which consists of 360 models (3 models x 120 locations). Since it is not feasible to report on the detailed statistical output for each of the 600 models built, a measure of forecast accuracy is used to report topline results and compare methods. For new releases forecast accuracy is presented at different levels of aggregation ranging from most detailed - by cinema location, date and film - to least detailed - by cinema location

and week. The second week and hold-over forecast results reported in this section are reported at the most granular level of detail, by film, date and cinema location.

The third section on Labour Recommendations begins by comparing forecasted to actual attendance in aggregate for all cinema locations. This is followed by the results for sensitivity testing on the methodology for generating the labour scheduling rules (median, average). The section concludes with a comparison of labour recommendations made by the DSS system, the existing system and the cinema manager. The fourth section, Concession Micro Forecasting, compares the labour recommendations generated by the DSS system with and without the concession forecast extension. Chapter 4 concludes with a summary of results.

#### **4.1. Audience Composition**

Across the 11 models adjusted R-square values range from 0.39 to 0.84, results are visualized in *Figure 4.1 – Adjusted R Square by Customer Segment Model*. The top performing model is *Parents* and the bottom performing model is *Females 14-18*. Regression equations for each of the 11 models are presented in this section with an analysis of the results offering further insight into the demographic cohort of moviegoers. The results provided in this section are extracted from *Figure 4.2- Standardized coefficients significant at the 90% confidence level for independent variables across 11 customer segment models*.

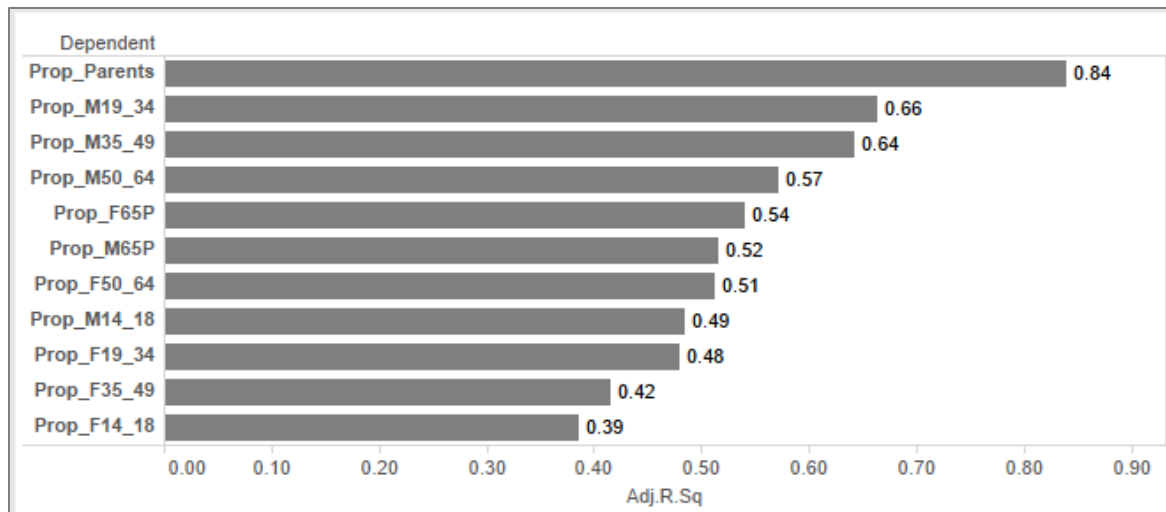


Figure 4.1 – Adjusted R Square by Customer Segment Model

## Results by Demographic Model

### Parents/Guardians

$$\text{Prop\_Parents} = \alpha + \beta_1 \text{Rating\_14A} + \beta_2 \text{Rating\_18A} + \beta_3 \text{Genre\_Comedy} + \beta_4 \text{Genre\_Family} + \beta_5 \text{Genre\_Sequel} + \beta_6 \text{Genre\_RomCom}$$

(4.1)

The R Square for the parent segment is 0.84, indicating that the predictors used in this model explain most of the variance in the proportion of the audience that is made up of parents. Film ratings are more explanatory than genre variables for parents, the ratings 18A and 14A have negative coefficients, indicating that parents are not likely to take kids to films with these ratings. The genres Family, Sequel & Comedy yield positive coefficients, and indicate that parents are more likely to see film tagged with any of these genres versus films that do not have these genres. The regression equation is presented in (4.1).



#### 14-18 Age Cohort

$$\text{Prop\_F14\_18} = \alpha + \beta_1\text{Genre\_Action} + \beta_2\text{Genre\_Family} + \beta_3\text{Genre\_Period} + \beta_4\text{Genre\_Romance} + \beta_5\text{Genre\_Horror} + \beta_6\text{Genre\_RomCom} + \beta_7\text{Award\_Level\_0} + \beta_8\text{Summer\_May\_Aug} + \beta_9\text{March\_Spring\_Break} + \beta_{10}\text{Production.Budget}$$

$$\text{Prop\_M14\_18} = \alpha + \beta_1\text{Rating\_18A} + \beta_2\text{Genre\_Action} + \beta_3\text{Genre\_Adaptation} + \beta_4\text{Genre\_Drama} + \beta_5\text{Genre\_Family} + \beta_6\text{Genre\_Horror} + \beta_7\text{Award\_Level\_0} + \beta_8\text{Production.Budget}$$

(4.2)

The R Square for the female segment in the 14-18 age cohort group is 0.34 while the R Square for the male segment is 0.49, indicating that the predictors tested in this study explain much more of the variance for males in this age cohort versus females. For females, positive coefficients that indicate appeal are found for Horror, Romance, RomCom, March Spring Break and Award Level 0. For males, positive coefficients that indicate appeal are found for Horror, Action, Rating 18A, Award Level 0 and production budget. The regression equations are presented in (4.2).

#### 19-34 Age Cohort

$$\text{Prop\_F19\_34} = \alpha + \beta_1\text{Rating\_14A} + \beta_2\text{Rating\_18A} + \beta_3\text{Genre\_Action} + \beta_4\text{Genre\_Family} + \beta_5\text{Genre\_Romance} + \beta_6\text{Genre\_Horror} + \beta_7\text{Genre\_RomCom} + \beta_8\text{Award\_Level\_2}$$

$$\text{Prop\_M19\_34} = \alpha + \beta_1\text{Film\_Runtime} + \beta_2\text{Rating\_14A} + \beta_3\text{Rating\_18A} + \beta_4\text{Genre\_Action} + \beta_5\text{Genre\_Adaptation} + \beta_6\text{Genre\_Comedy} + \beta_7\text{Genre\_Drama} + \beta_8\text{Genre\_Family} + \beta_9\text{Genre\_Romance} + \beta_{10}\text{Genre\_Sci\_Fi} + \beta_{11}\text{Genre\_Horror} + \beta_{12}\text{Summer\_May\_Aug} + \beta_{13}\text{Valentines} + \beta_{14}\text{Production.Budget}$$

(4.3)

The R Square for the female segment in the 19-34 age cohort group is 0.48 while the R Square for the male segment is 0.66, indicating that, similarly to the 14-18 age cohort, the predictors tested in this study explain much more of the variance for males in this age

cohort versus females. For females, positive coefficients are found for Rating 18A, Rating 14A, Romance, RomCom, Horror and Award Level 2. For males, positive coefficients are found for Rating 18A, Rating 14A, Action, Production Budget, Horror and Sci Fi. For males, positive coefficients are found for Rating 18A, Rating 14A, Action, Production Budget, Horror and Sci Fi. The regression equations are presented in (4.3).

#### 35-49 Age Cohort

$$\text{Prop\_M35\_49} = \alpha + \beta_1\text{Rating\_18A} + \beta_2\text{Genre\_Action} + \beta_3\text{Genre\_Comedy} + \beta_4\text{Genre\_Family} + \beta_5\text{Genre\_Romance} + \beta_6\text{Genre\_Sequel} + \beta_7\text{Genre\_Thriller} + \beta_8\text{Genre\_Sci\_Fi} + \beta_9\text{Genre\_RomCom} + \beta_{10}\text{Award\_Level\_0} + \beta_{11}\text{Big\_Star\_Count} + \beta_{12}\text{Production.Budget}$$

$$\text{Prop\_F35\_49} = \alpha + \beta_1\text{Film\_Runtime} + \beta_2\text{Rating\_PG} + \beta_3\text{Rating\_14A} + \beta_4\text{Genre\_Action} + \beta_5\text{Genre\_Drama} + \beta_6\text{Genre\_Family} + \beta_7\text{Genre\_Sequel} + \beta_8\text{Genre\_Sci\_Fi} + \beta_9\text{Genre\_Horror} + \beta_{10}\text{Award\_Level\_0} + \beta_{11}\text{Summer\_May\_Aug} + \beta_{12}\text{Valentines} + \beta_{13}\text{Production.Budget}$$

(4.4)

The R Square for the female segment in the 35-49 age cohort group is 0.42 while the R Square for the male segment is 0.64, indicating that, similarly to the younger age cohorts, the predictors tested in this study explain more of the variance for males in this age cohort versus females. For females, positive coefficients are found for Rating 14A, Film Runtime, Rating PG, Award Level 0, Summer, Drama and Valentines. For males, positive coefficients are found for Action, Production Budget, Rating 18A, Sci Fi, Thriller, Award Level and Big Star Count. The regression equations are presented in (4.4).

#### 50-64 Age Cohort

$$\text{Prop\_F50\_64} = \alpha + \beta_1\text{Rating\_14A} + \beta_2\text{Genre\_Action} + \beta_3\text{Genre\_Drama} + \beta_4\text{Genre\_Family} + \beta_5\text{Genre\_Period} + \beta_6\text{Genre\_Sequel} + \beta_7\text{Genre\_Sci\_Fi} + \beta_8\text{Genre\_Horror} + \beta_9\text{Big\_Star\_Count} + \beta_{10}\text{March\_Spring\_Break}$$

$$\text{Prop\_M50\_64} = \alpha + \beta_1\text{Rating\_PG} + \beta_2\text{Rating\_14A} + \beta_3\text{Rating\_18A} + \beta_4\text{Genre\_Action} + \beta_5\text{Genre\_Comedy} + \beta_6\text{Genre\_Drama} + \beta_7\text{Genre\_Family} + \beta_8\text{Genre\_Period} + \beta_9\text{Genre\_Romance} + \beta_{10}\text{Genre\_Sequel} + \beta_{11}\text{Genre\_Thriller} + \beta_{12}\text{Genre\_Horror} + \beta_{13}\text{Big\_Star\_Count} + \beta_{14}\text{March\_Spring\_Break}$$

(4.5)

The R Square for the female segment in the 50-64 age cohort group is 0.51 while the R Square for the male segment is 0.57. Similarly to the younger age cohorts, the predictors tested in this study explain more of the variance for males in this age cohort versus females, but the male-female difference in variance explained is much smaller than younger age cohorts. For females, positive coefficients are found for Drama, Rating 14A, Period and Big Star Count. For males, positive coefficients are found for Rating 14A, Rating 18A, Rating PG, Drama, Action, Period, Thriller and Big Star Count. The regression equations are presented in (4.5).

#### 65+ Age Cohort

$$\text{Prop\_F65P} = \alpha + \beta_1\text{Film\_Runtime} + \beta_2\text{Rating\_14A} + \beta_3\text{Genre\_Action} + \beta_4\text{Genre\_Adventure} + \beta_5\text{Genre\_Drama} + \beta_6\text{Genre\_Period} + \beta_7\text{Genre\_Romance} + \beta_8\text{Genre\_Sequel} + \beta_9\text{Genre\_Sci\_Fi} + \beta_{10}\text{Genre\_Horror} + \beta_{11}\text{March\_Spring\_Break} + \beta_{12}\text{Production.Budget}$$

$$\text{Prop\_M65P} = \alpha + \beta_1\text{Film\_Runtime} + \beta_2\text{Rating\_14A} + \beta_3\text{Rating\_18A} + \beta_4\text{Genre\_Comedy} + \beta_5\text{Genre\_Drama} + \beta_6\text{Genre\_Family} + \beta_7\text{Genre\_Period} + \beta_8\text{Genre\_Romance} + \beta_9\text{Genre\_Sequel} + \beta_{10}\text{Genre\_Thriller} + \beta_{11}\text{Genre\_Sci\_Fi} + \beta_{12}\text{Genre\_Horror} + \beta_{13}\text{March\_Spring\_Break}$$

(4.6)

The R Square for the female segment in the 65+ age cohort group is 0.54 while the R Square for the male segment is 0.52, indicating that the predictors tested in this study explain only slightly more of the variance for females in this age cohort versus males. For

females, positive coefficients are found for Drama, Film Runtime, Rating 14A and Period. For males, positive coefficients are found for Drama, Rating 14A, Rating 18A, Thriller, Period and Film Runtime. The regression equations are presented in (4.6).

### **Model Results by Independent Variables**

*Figure 4.2- Standardized coefficients significant at the 90% confidence level for independent variables across 11 customer segment models* provides a comparative view of the independent predictor variables used by model. Blank cells in Figure 4.2 indicate that the independent variable was not significant at the 90% confidence level for the respective model. The results are summarized in this section by drawing similarities and differences between the demographic cohorts.

**Production Budget** has positive coefficients for males aged 14-49, indicating that males are more likely to see films with a bigger budget.

**Film Runtime** has positive coefficients for females 35 to 49 and both genders 65 plus. Males 19-34 have a negative coefficient on film runtime, indicating a preference for shorter films.

**Big Star Count** was only significant for three cohorts: both genders aged 50 to 64 and males 35 to 49.

The categorical **award nomination variables** indicate that films with no nominations appeal to teens and ages 35-49 while the opposite is true for females aged 19-34.

**Action**, where significant, shows positive coefficients for males and negative coefficients for females. With the exception of Females 65 plus, Romance shows the

opposite of Action with the models for female segments having positive coefficients and males having negative, this is also comparable to the results from Redondo & Holbrook (2012). The genre **RomCom** appeals to younger females 14-34.

**Drama**, where significant, yields negative coefficients for younger demographics and positive coefficients for mature demographics. This is similar to the results from Redondo & Holbrook (2010) where the fourth dimension indicated that a more mature audience was positively correlated with drama.

Across all but one model, **Family** was significant and offered strong explanatory power compared to most other independent variables. As expected, the parent cohort is likely to see Family movies, and all other age cohorts are unlikely to see movies tagged with Family. This finding is comparable with the results from Redondo & Holbrook (2010) where the Family factor had the greatest model contribution and most distinct segment.

**Horror** has positive coefficients for younger cohorts and negative coefficients for older cohorts. **Period** appeals to a mature crowd with positive coefficients for demographics aged 50 plus. **Sci Fi** appeals to males 19-49 and Thriller appeals to males 35 and up.

**Rating 18A** is significant and positively correlated to all male age cohorts as well as females 19 to 34. **Rating 14A** is significant and has positive coefficients for all female age cohorts except 14 to 18 year olds.

The temporal variables **March Spring Break** and **Summer May Aug** are significant and have positive coefficients for high school aged females 14 to 18. Summer May Aug also has a positive coefficient for females aged 35 to 49. **Valentine's Day** has a positive coefficient for females aged 35 to 49.

Independents	Dependent										
	Prop_Parents	Prop_F14_18	Prop_M14_18	Prop_F19_34	Prop_M19_34	Prop_F35_49	Prop_M35_49	Prop_F50_64	Prop_M50_64	Prop_F65P	Prop_M65P
Production.Budget		-0.116	0.122		0.208	-0.198	0.170			-0.159	
Film_Runtime					-0.072	0.213				0.202	0.075
Big_Star_Count							0.062	0.072	0.113		
Award_Level_0		0.106	0.138			0.123	0.079				
Award_Level_2				0.110							
Genre_Action		-0.189	0.177	-0.137	0.248	-0.205	0.285	-0.087	0.134	-0.096	
Genre_Adaptation			-0.158		-0.086						
Genre_Adventure										-0.114	
Genre_Comedy	0.079				-0.129		-0.232		-0.180		-0.097
Genre_Drama			-0.259		-0.152	0.083		0.302	0.154	0.248	0.233
Genre_Family	0.477	-0.155	-0.343	-0.277	-0.186	-0.340	-0.279	-0.236	-0.244		-0.181
Genre_Horror		0.337	0.281	0.135	0.143	-0.120		-0.273	-0.221	-0.326	-0.300
Genre_Period		-0.123						0.103	0.120	0.099	0.120
Genre_Romance		0.212		0.272	-0.078		-0.178		-0.148	-0.076	-0.136
Genre_RomCom	-0.048	0.119		0.141			-0.064				
Genre_Sci_Fi					0.118	-0.145	0.126	-0.121		-0.140	-0.085
Genre_Sequel	0.076					-0.115	-0.084	-0.210	-0.195	-0.179	-0.212
Genre_Thriller							0.096		0.116		0.131
Rating_PG						0.211			0.225		
Rating_14A	-0.535			0.346	0.251	0.343		0.193	0.472	0.198	0.219
Rating_18A	-0.642		0.147	0.475	0.598		0.134		0.397		0.166
March_Spring_Break		0.118						-0.089	-0.065	-0.093	-0.086
Summer_May_Aug		0.119			-0.068	0.102					
Valentines					-0.057	0.077					

Figure 4.2- Standardized coefficients significant at the 90% confidence level for independent variables across 11 customer segment models

## **4.2. Box Office Micro Forecasting**

Table 4.1 provides a comparative view of the independent variables used in each part of the box office forecasting module. Independent variables are selected based on the film's point in its release cycle. For films in their first week of release with no actual sales data available the independent variables are a collection of film attributes (including genre, ratings, predicted demographic appeal), film scheduling at nearby locations, capacity at the location and advanced ticket sales. The models for films in their second or third week of release have similar sets of independent variables (days since release, actual ticket sales and capacity at the location) with the key difference being that the holdover model (for films in their third week of release or later) has more actual sales data. The results presented in this section are the results of the models applied to the cross validation dataset which spans from Jan 1 2014 to April 30 2014.

Forecast Accuracy is computed by comparing forecasted box office sales to actual box office sales. Different aggregations of granularity are compared to show the trade off between precision and accuracy. For each aggregation, if the forecast is within 35% of the actual sales then the forecast is considered accurate. The topline forecast accuracy that is reported in this section is the percent of total cases that are accurate.

Table 4.1 - variables used in final model broken out by forecast model

		Used in Week 1 Micro Forecast	Used in Week 2 Micro Forecast	Used in Week 3P Micro Forecast
1				
2	<i>Rating_PG</i>	Yes		
3	<i>Rating_14A</i>	Yes		
4	<i>Rating_18A</i>	Yes		
5	<i>Binary Genres (12 in total)</i>	Yes		
6	<i>Big_Star_Count</i>	Yes		
7	<i>Film_Runtime</i>	Yes		
8	<i>Production_Budget</i>	Yes		
9	<i>Predicted Demographic Appeal (11 in total)</i>	Yes		
10	<i>Days_Since_Release_Friday</i>	Yes	Yes	Yes
11	<i>Holiday_Name</i>	Yes	Yes	Yes
12	<i>Show Day of Week (6 binary variables Saturday-Thursday)</i>	Yes	Yes	Yes
13	<i>Sum_Sum_Seats_Sold</i>	*dependent variable	*dependent variable	*dependent variable
14	<i>Sum_Total_Sessions</i>	Yes	Yes	Yes
15	<i>P2year_AvgSeatsPerSession</i>	Yes		
16	<i>Estimated_First_Week_Gross_Revenue</i>	Yes		
17	<i>Has_Genre_Comp_at_Location</i>	Yes		
18	<i>Playing_at_all_nearby_locations</i>	Yes		
19	<i>Not_playing_nearby</i>	Yes		
20	<i>Sum_Sum_Seats_Available</i>	Yes	Yes	Yes
21	<i>Max_Is_3D</i>	Yes	Yes	Yes
22	<i>Max_Is_IMAX</i>	Yes	Yes	Yes
23	<i>Max_Is_AVX</i>	Yes	Yes	Yes
24	<i>Max_Is_VIP</i>	Yes	Yes	Yes
25	<i>Max_Is_French_Language</i>	Yes	Yes	Yes
26	<i>Week1_Total_Sessions</i>		Yes	Yes
27	<i>Week1_Seats_Sold</i>		Yes	Yes
28	<i>Week1_Seats_Available</i>		Yes	Yes
29	<i>Week2_Total_Sessions</i>			Yes
30	<i>Week2_Seats_Sold</i>			Yes
31	<i>Week2_Seats_Available</i>			Yes



## New Release Forecast

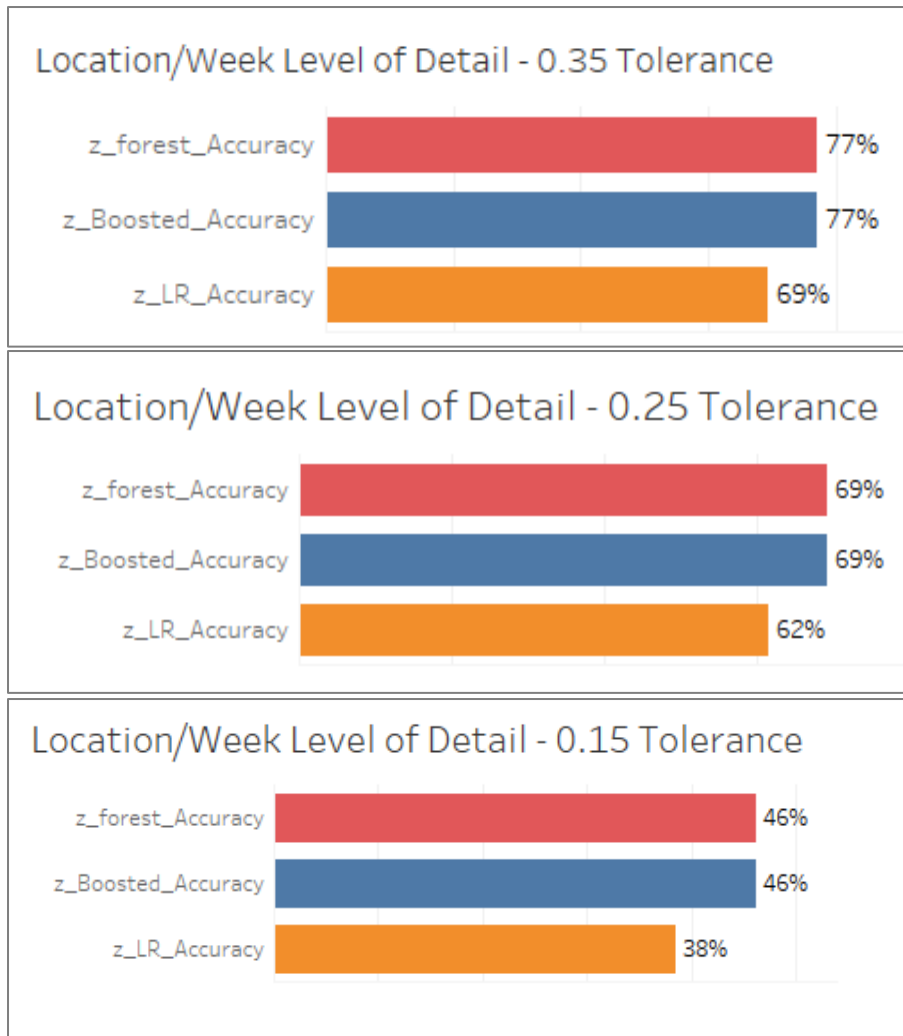


Figure 4.3 - New release forecast accuracy comparison between three methods at the location/week level of detail (red = forest model, blue = boosted model, orange = multiple regression)

Figure 4.3 shows Forecast Accuracy for the three different methods with the least precision at the location and week level of detail, meaning that each case represents the total sum of forecasted box office admits for all week 1 titles in a single week at a single location. At this level of detail the forest model and boosted regression model result in the

same accuracy with 77% of location weekly forecasts being within 35% of actual sales. The multiple regression model performs slightly worse with an accuracy of 69%. Results are presented at three different tolerance levels (35%, 25% and 15%) for sensitivity testing and to demonstrate that the forest model is best performing at different threshold levels.

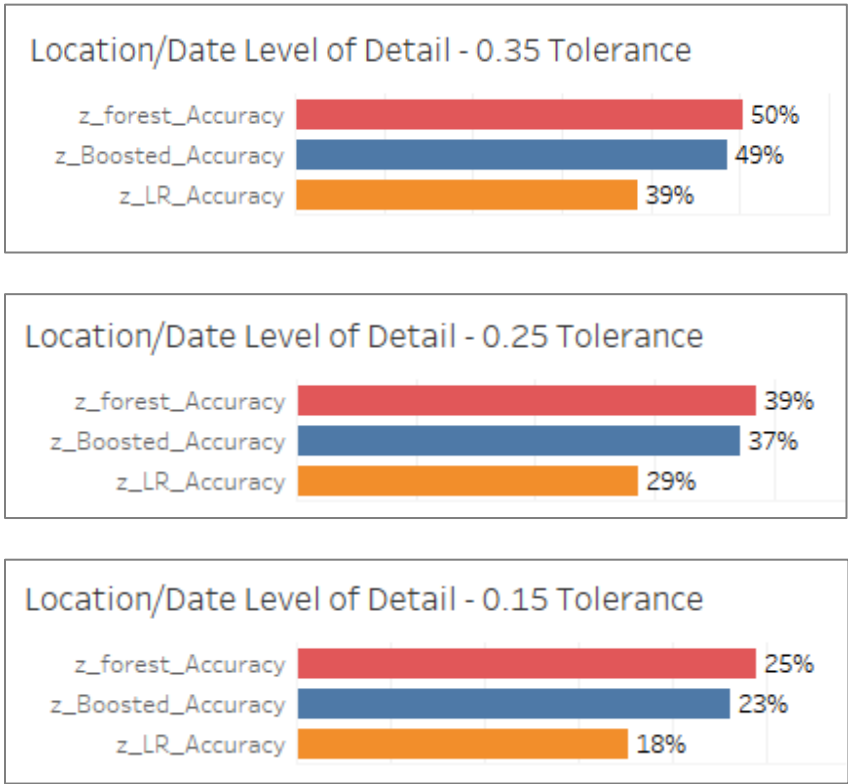


Figure 4.4 - New release forecast accuracy comparison between three methods at the location/day level of detail (red = forest model, blue = boosted model, orange = multiple regression)

Figure 4.4 shows Forecast Accuracy for the three different methods with more precision at the location and date level of detail, meaning that each case represents the total sum of forecasted box office admits for all week 1 titles in a single day at a single location. At this level of detail the forest model performs slightly better than the boosted regression model with 50% of location daily forecasts being within 35% of actual sales,

compared to 49%. The multiple regression model performs slightly worse with an accuracy of 39%. Results are presented at three different tolerance levels (35%, 25% and 15%) for sensitivity testing and to demonstrate that the forest model is best performing at different threshold levels.

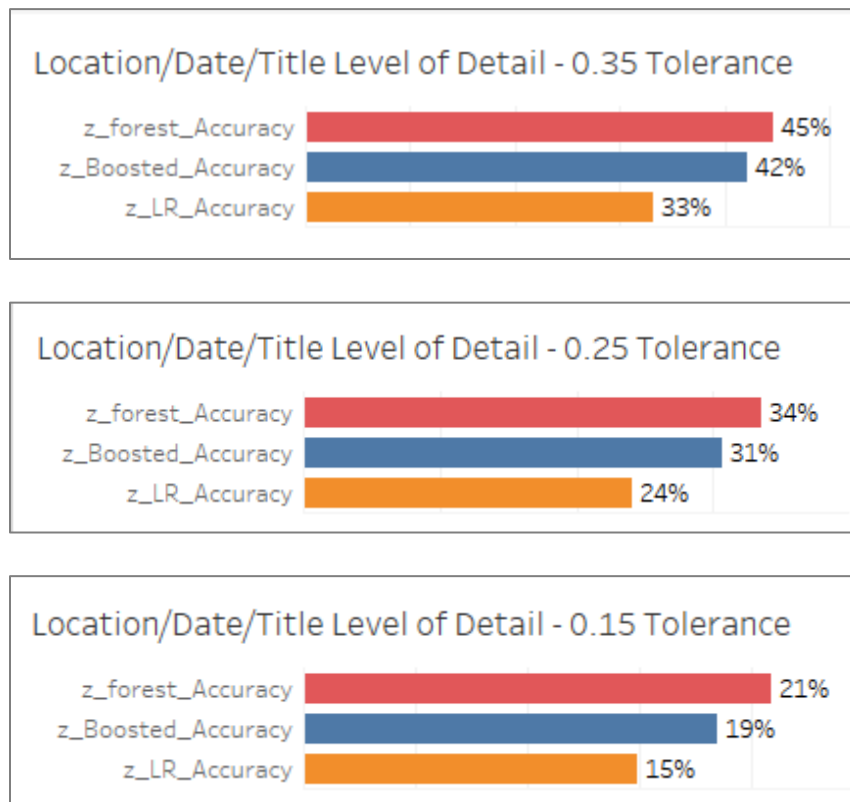


Figure 4.5 - New release forecast accuracy comparison between three methods at the location/day/film level of detail (red = forest model, blue = boosted model, orange = multiple regression)

Figure 4.5 shows Forecast Accuracy for the three different methods with most precision at the location, date and film level of detail, meaning that each case represents the forecasted box office admits for a single week 1 title on a single day at a single location. At this level of detail the forest model performs better than the boosted regression model with

45% of location daily forecasts being within 35% of actual sales, compared to 42%. The multiple regression model performs slightly worse with an accuracy of 33%. Results are presented at three different tolerance levels (35%, 25% and 15%) for sensitivity testing and to demonstrate that the forest model is best performing at different threshold levels.

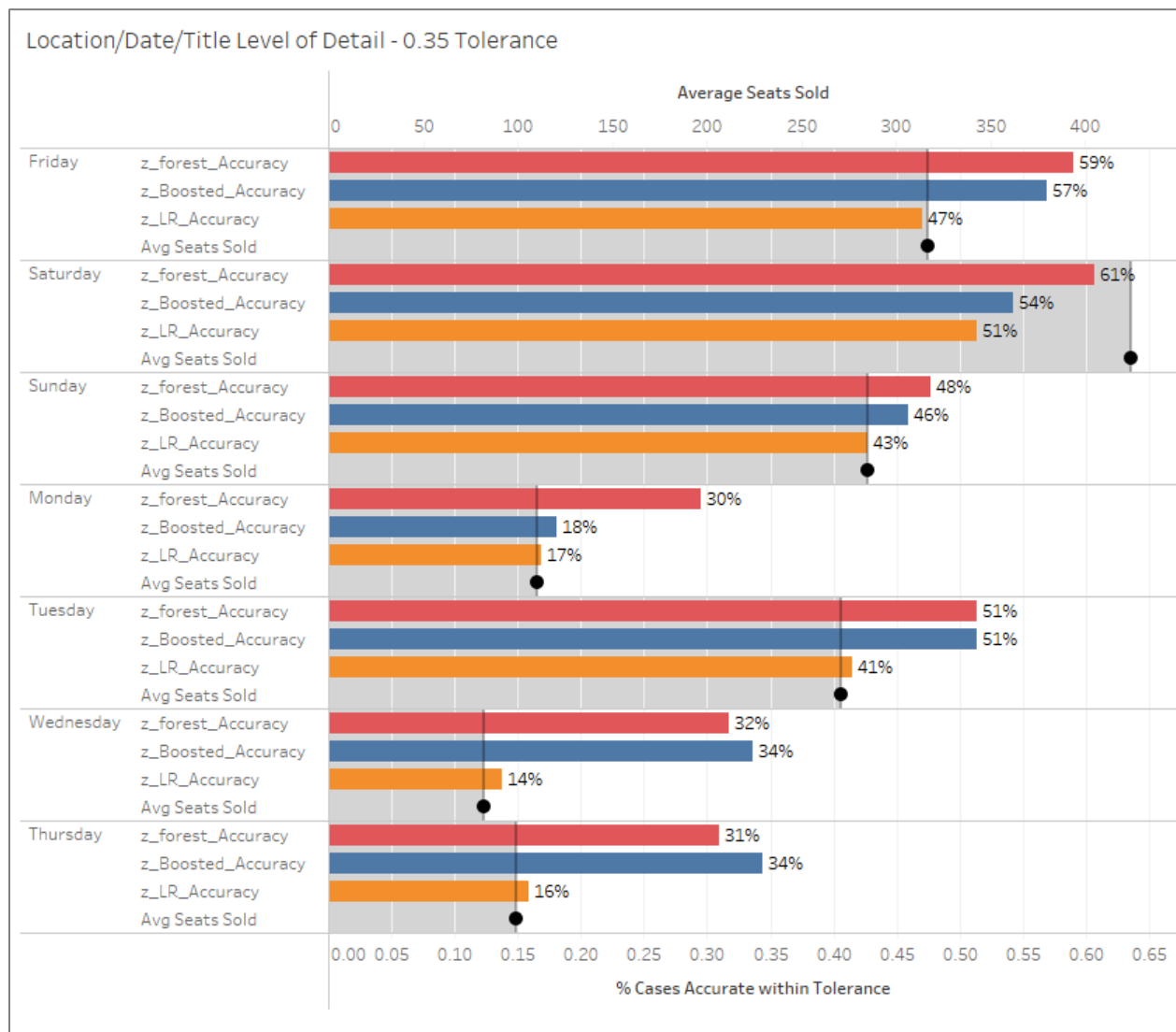


Figure 4.6 - New release forecast accuracy comparison between three methods by weekday at the location/day/film level of detail (red = forest model, blue = boosted model, orange = multiple regression)

Figure 4.6 shows Forecast Accuracy for the three different methods at the location, date and film level of detail broken out by weekday plotted against the historic average seats sold. As expected the average seats sold is highest on weekends: Saturdays followed by Fridays and then Sundays having the most sales. It is observed that all the models perform better on the busier weekdays compared to the weekdays with less sales (Monday, Wednesday, and Thursday). The forest model has the best results for Friday (59%), Saturday (61%) and Sunday (48%) with accuracy on these days approximately twice as good as the accuracy on slower weekdays - Monday (30%), Wednesday (32%), Thursday (31%). The boosted model provided the best results on Wednesdays (34%) and Thursdays (34%).

Figure 4.7 depicts the independent variables significant at the 90% confidence level in the new release micro forecasting model for a selection of locations. Variables which are significant differ between locations. For the locations depicted, most of the coefficients have a common direction of relationship (either positive or negative) with demand for new release movies, however the values of coefficients vary between locations. In the locations depicted there is one independent variable - Has Genre Comp at Location - which displays two different relationships with new release demand, for the British Columbia - Not Vancouver location a positive coefficient is observed while a negative coefficient is observed for the Greater Toronto Area Suburb location. The positive coefficient in the more rural location could be explained by a spillover where patrons may encounter a sold out show and elect to see a different movie of the same genre.

	Location ID String / Adj.R.Sq				
	British Columbia - Not Vancouver Area	Town in South-western Ontario	Greater Toronto Area Suburb	Greater Toronto Area Suburb 2	Montreal Suburb
Independents	0.734346492	0.690034720	0.732745168	0.707407357	0.674306740
Avg_Prop_F14_18				0.36	0.25
Avg_Prop_F19_34		0.44			
Avg_Prop_F35_49		-0.50			-0.65
Avg_Prop_F50_64			-0.45		
Avg_Prop_F65P	0.82		0.30		
Avg_Prop_M14_18		0.49	0.23	0.31	
Avg_Prop_M19_34		-1.36	-1.40	-0.76	-2.06
Avg_Prop_M35_49				-0.70	
Avg_Prop_M50_64			0.52		
Avg_Prop_Parents	0.11	0.11	0.10	0.10	0.12
Film_Runtime					0.13
Genre_Action	0.40			0.38	0.42
Genre_Adaptation			-0.09		-0.12
Genre_Adventure		-0.14	-0.09	-0.12	-0.10
Genre_Comedy				-0.24	-0.23
Genre_Drama					-0.24
Genre_Family	-0.86				-0.81
Genre_Fantasy			0.06		
Genre_Horror		-0.20		-0.20	
Genre_Period	-0.14			0.13	0.10
Genre_Romance		-0.24		-0.19	-0.22
Genre_RomCom		-0.11		-0.09	
Genre_Sci_Fi			0.11	0.15	0.19
Genre_Thriller			-0.09		
Has_Genre_Comp_at_Location	0.12			-0.11	
Monday		0.05			
Not_playing_nearby				0.00	0.00
Playing_at_all_nearby_locations		-0.08	-0.05	-0.29	
Production_Budget	0.68		0.21	0.37	0.35
Rating_14A					0.81
Rating_18A			0.73		1.33
Saturday	0.24	0.24	0.20	0.22	0.30
Sum_Total_Sessions	0.01		0.01		0.01
Sunday	0.13	0.09	0.08	0.13	0.20
Thursday	0.05		0.11	0.04	
Tuesday	0.46	0.34	0.21	0.26	0.33
Wednesday		0.05			

Figure 4.7 - independent variables significant at the 90% confidence level in the new release micro forecasting model for a selection of locations

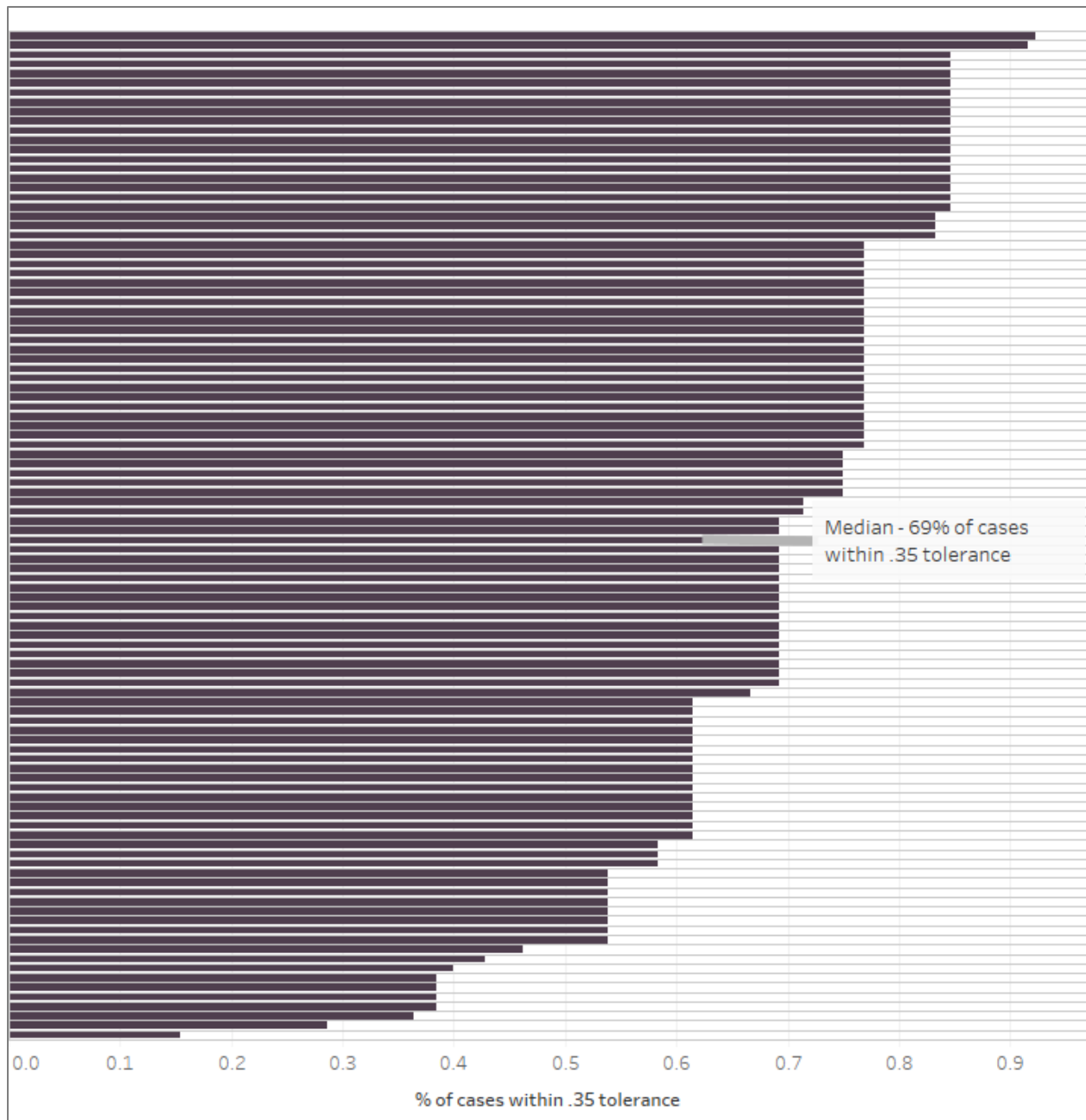


Figure 4.8 - distribution of new release accuracy for all locations

Figure 4.8 shows the distribution of new release forecasting accuracy for all locations, with the median being 69% at the 35% tolerance level.

## Second-Week & Hold-Over Forecasts

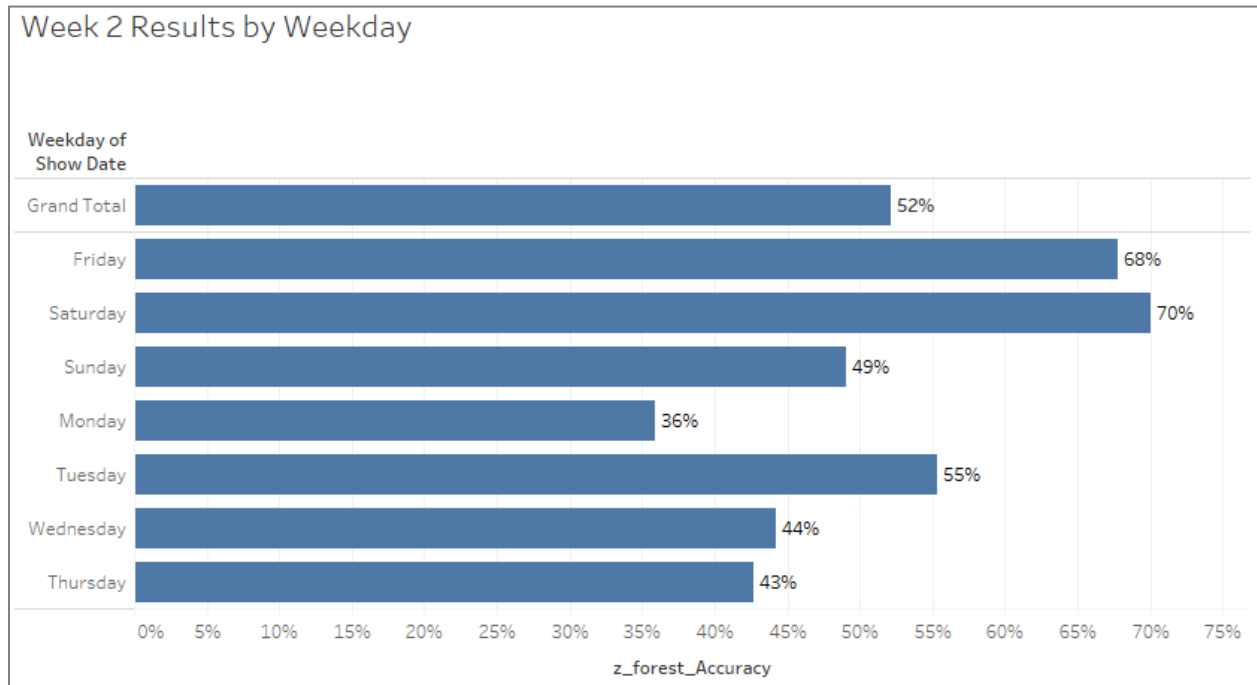


Figure 4.9 - Forest model accuracy by weekday for week 2 films

The second week and hold-over forecast results reported in this section utilize the most granular level of detail, by title, date and location. As expected, with more actual sales data, the forecasts for second week (52%) and hold-over (49%) are more accurate than week 1 (45%). Figure 4.9 visualizes second week results by weekday, the results follow the same distribution as the week 1 results, with Saturdays having the highest accuracy (70% vs. 61% for week 1), followed by Fridays (68% vs. 59% for week 1), followed by Tuesdays (55% vs. 51% for week 1).



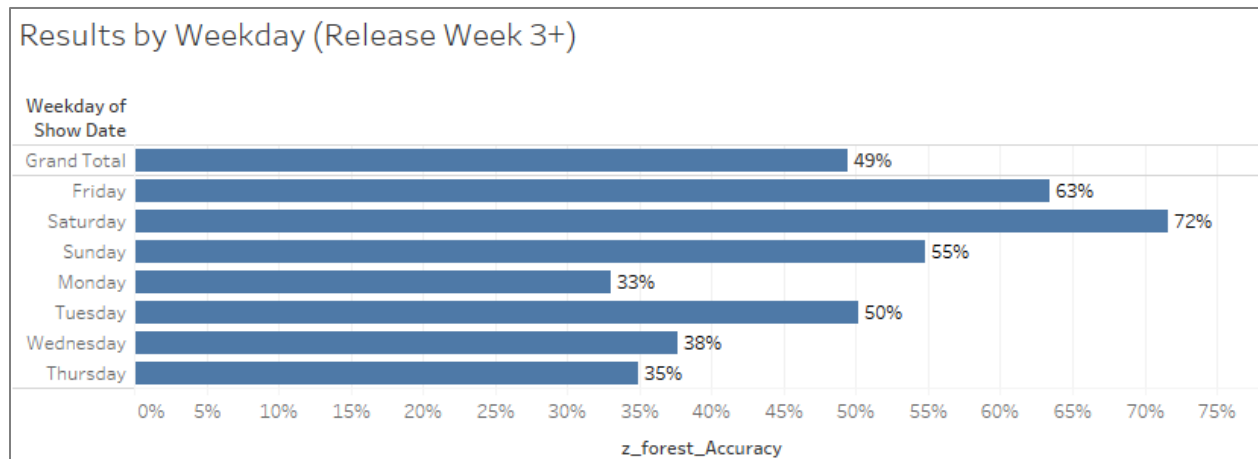


Figure 4.10 - Forest model accuracy by weekday for hold-over films

Figure 4.10 depicts the hold-over results by weekday, compared to the second week model, accuracy is lower on all days except Saturdays (72% vs. 70% for second week) and Sundays (55% vs. 49% for second week).

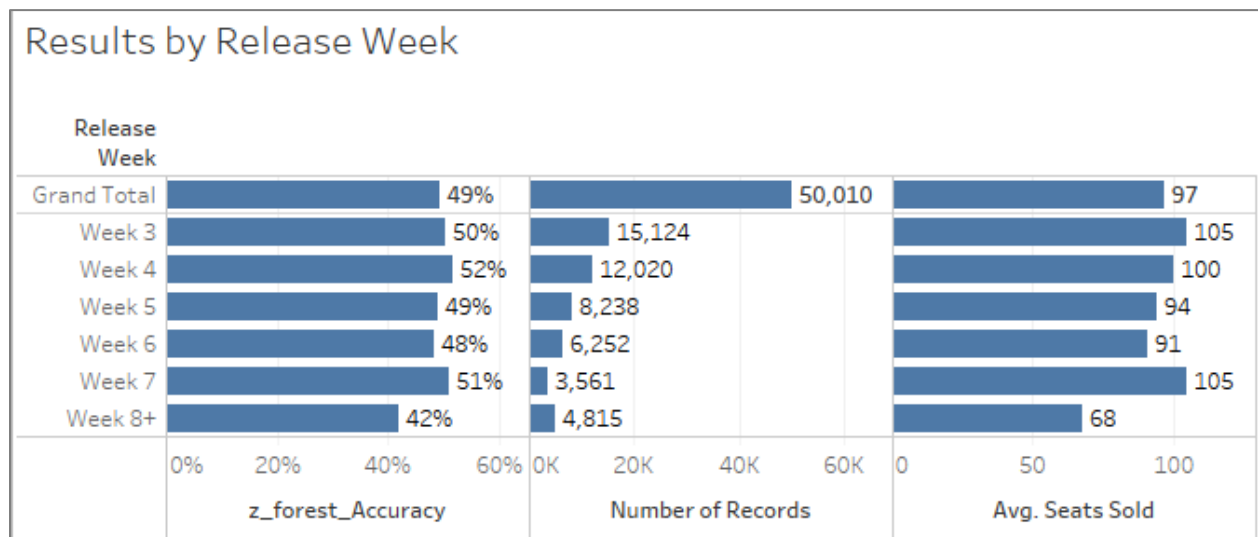


Figure 4.11 - Forest model accuracy by release week for hold-over films

Looking at Figure 4.11, lower accuracy for hold-overs could be related to the release week, as the release week increases there are fewer show times with a decreasing average seats

sold. Week 3 (50% accuracy) and week 4 (52% accurate) have significantly higher accuracy rates than week 8+ (42%).

### 4.3. Labour Recommendations

Labour recommendations are made by location, date and job function based on the sum of demand forecasts for new release, second week and hold-over films and the location's *Adjusted Scheduling Rule*. Figure 4.12 provides an aggregate view of the sum of demand forecasts across the 3 model types for all locations compared to actual ticket sales, and Figure 4.13 provides a micro view for 5 example locations.

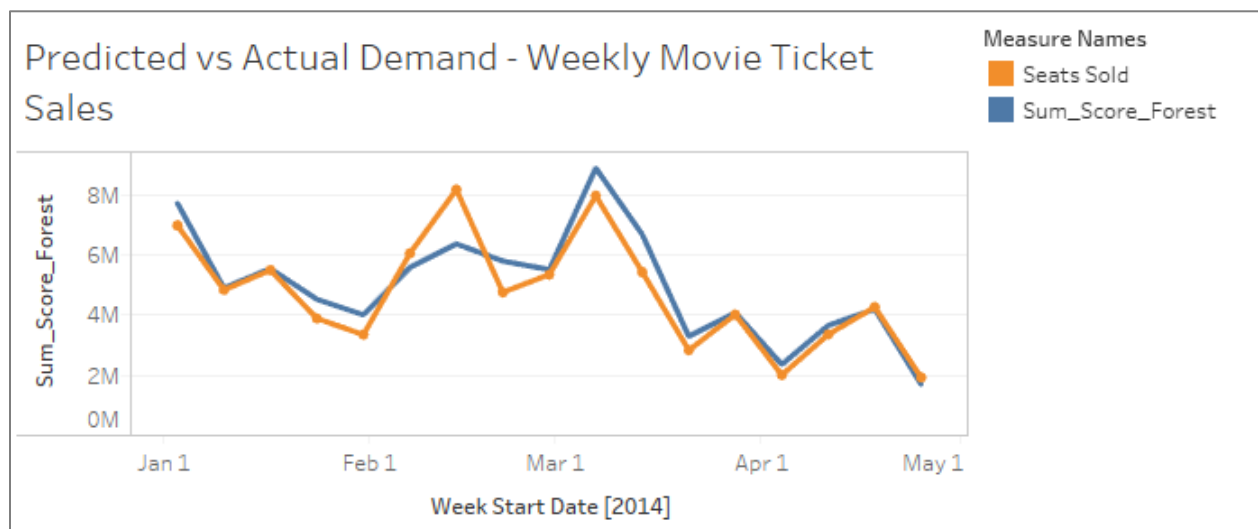


Figure 4.12 - sum of demand forecasts for new release, second week and hold-over films depicted as sum\_score\_forest plotted against the actual seats sold

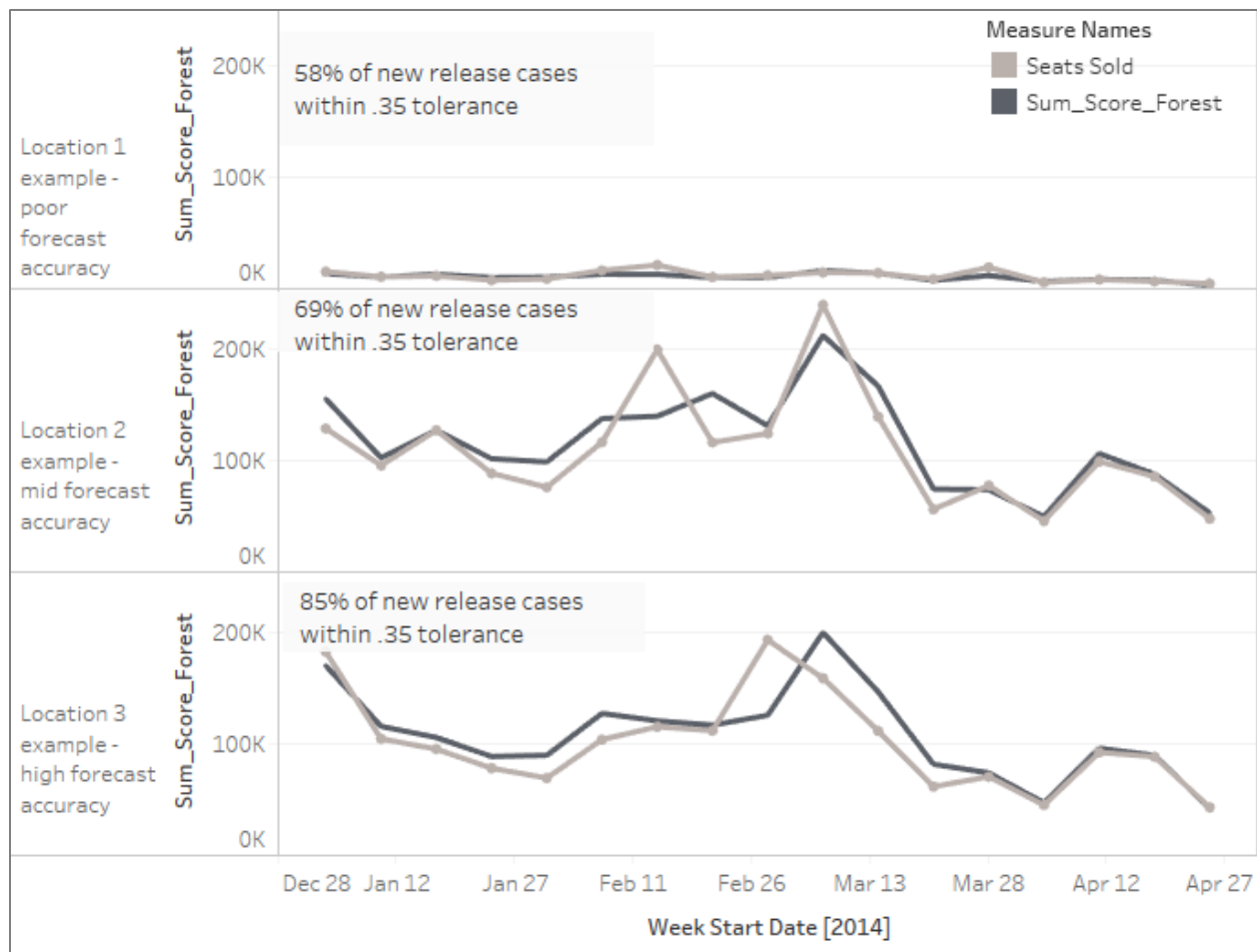


Figure 4.13 - sum of demand forecasts for new release, second week and hold-over films depicted as sum\_score\_forest plotted against the actual seats sold for three select locations

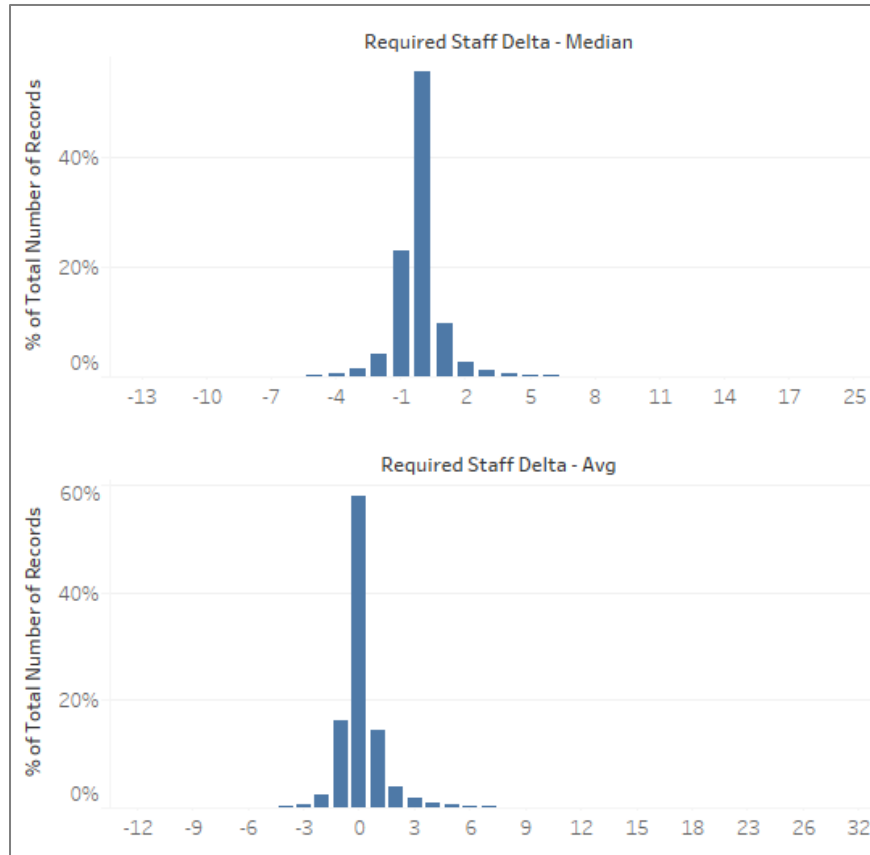


Figure 4.14 - topline labor recommendations delta (actual vs. predicted by the model) based on median (top) and average (bottom)

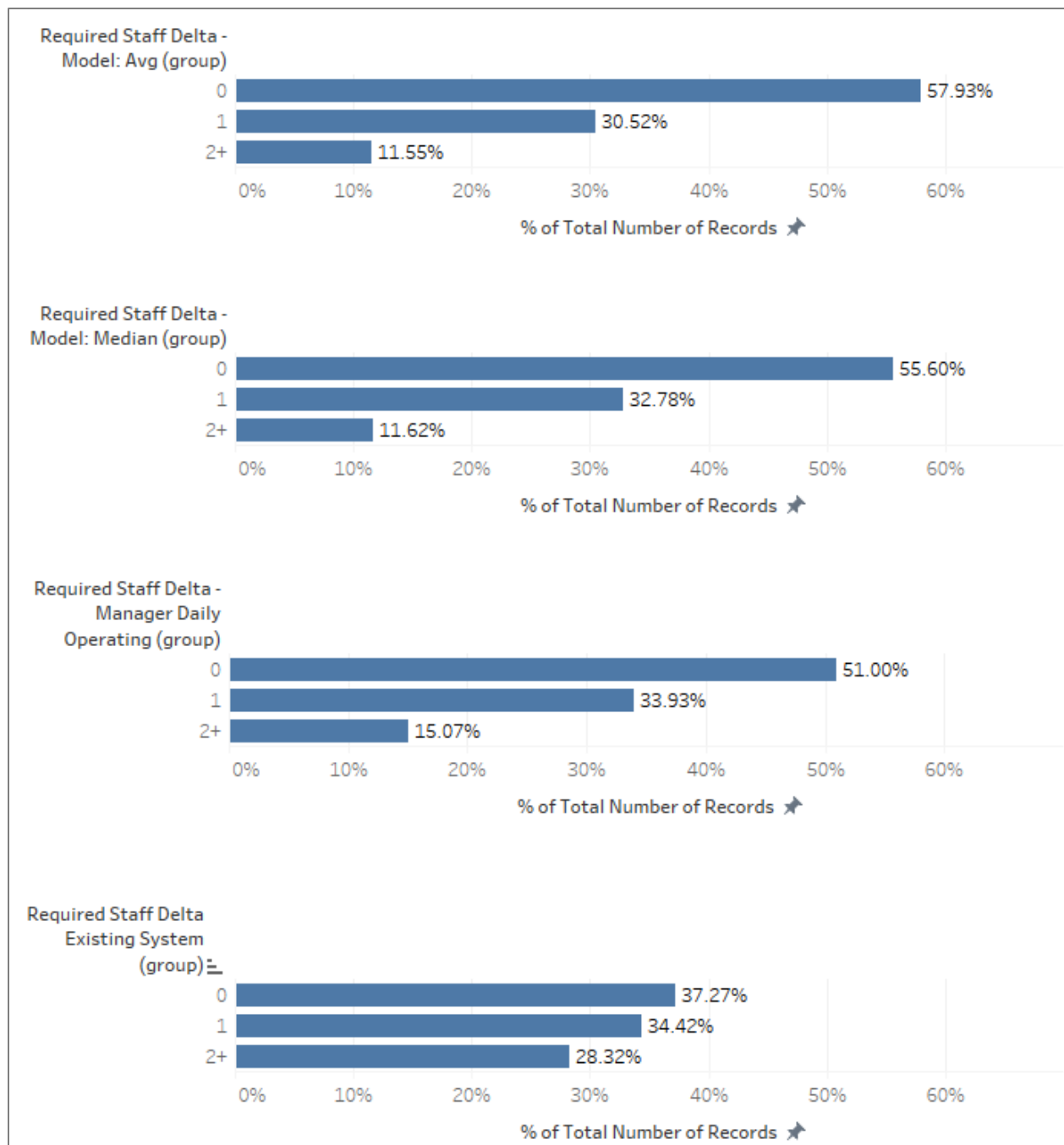


Figure 4.15 - labor recommendation effectiveness - model versus manager and system

The methodology for the *Adjusted Scheduling Rule* is detailed in section 4.3. Sensitivity testing was done to select the most appropriate measure of central tendency.

Figure 4.14 depicts the topline labor recommendations deltas (actual vs. predicted by the model) based on median (top chart) and average (bottom chart). Both methods result in a high proportion of cases having the correct labor recommendation. Figure 4.15 shows a more summarized view of the results charting with the proportion of labor recommendations with 0 person, 1 person and 2+ people delta. The top chart of Figure 4.15 shows the results for average (57.9% with delta of 0, 30.5% with delta of 1), second from the top shows median (55.6% with delta of 0, 32.8% with delta of 1), third from the top shows the results based on the manager's forecast and average method (51.0% with delta of 0, 33.5% with delta of 1), and bottom shows the results based on the system's forecast with average method (37.3% with delta of 0, 34.4% with delta of 1).

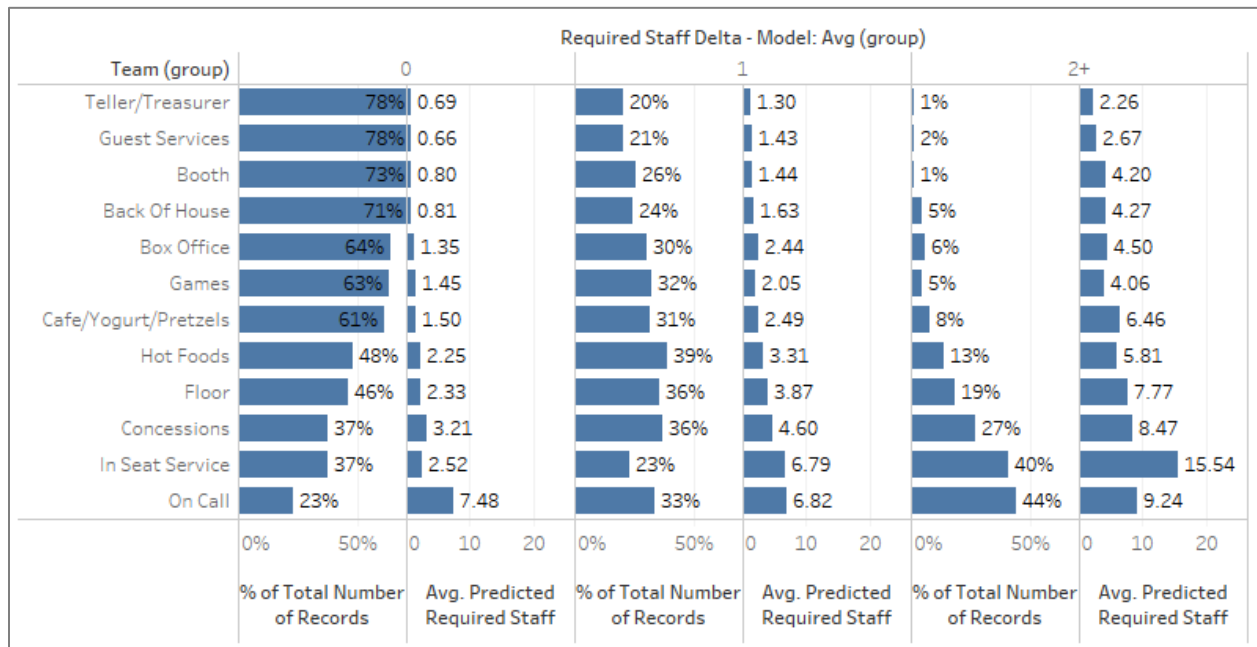


Figure 4.16 - Summary of model's recommendations by Job Function (labelled as Team (group))

Figure 4.16 shows how the model's staffing recommendations break out by job function, compared to the average predicted required staff. The model's recommendations have a higher success rate for predicting the required staff with a 0 person delta for the job functions that require fewer staff; Teller/Treasurer (78%) is the top ranked function based on the % of recommendations with a 0 person delta, followed by Guest Services (78%), Booth (73%), Back Of House (71%), Box Office (64%), Games (63%), Café/Yogurt/Pretzels (61%), Hot Foods (48%), Floor (46%), Concessions (37%), In Seat Service (37%) and On Call (23%). Note that job functions which require more staff on average are more likely to have a larger delta when looking at the number of staff required because the 0 person delta measure is not standardized to account for the number of staff recommended.

#### **4.4. Concession Micro Forecasting**

The concession forecasting extension outputs daily staffing recommendations by location for the three food areas (as applicable): Café/yoghurt/pretzels, hot foods and concessions. Of these three areas, the concessions area is most profitable serving the highest volume of food with fountain drinks and popcorn making up a large fraction of sales and having the most favourable profit margins. Figure 4.17 and Figure 4.18 depict the food station staffing results with and without the concession extension. The concession extension performs better than the original model in all three food areas: Café/Yogurt/Pretzels improved to 87% correct recommendations (from 37%), Hot Foods to 78% (from 29%) and Concessions to 65% (from 34%).

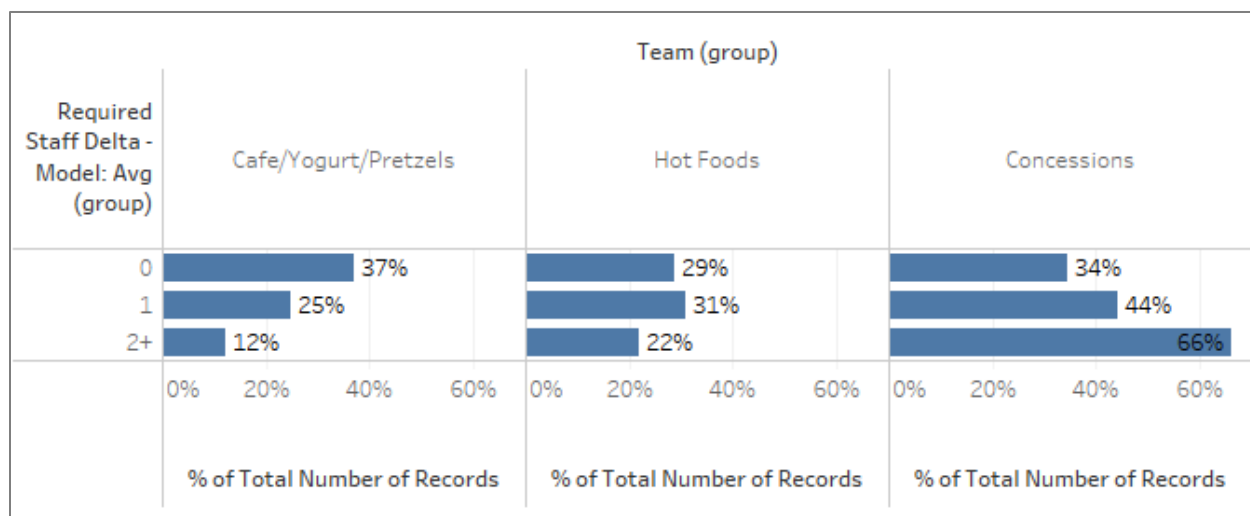


Figure 4.17 - Food station labour recommendations results without concession extension

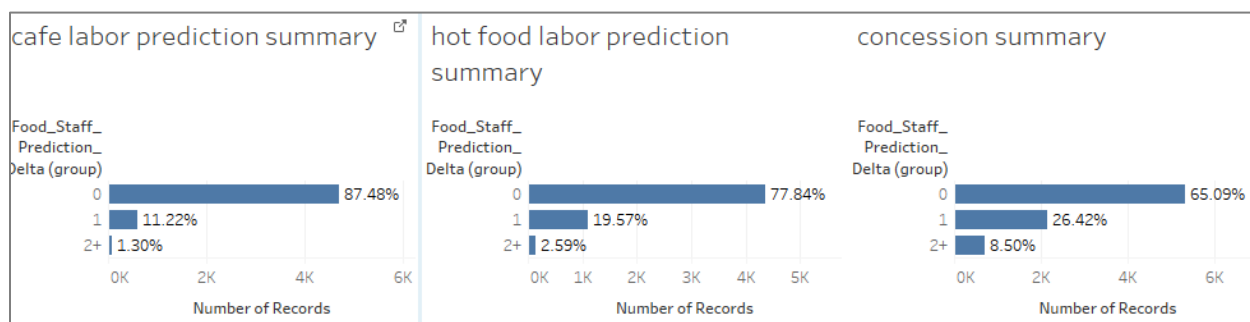


Figure 4.18 - Food station labour recommendations results with concession extension.

#### 4.5. Illustrative Example of Data Outputs for a Single Location and Day

In order to illustrate the end to end nature of the prototype system, this subsection provides a numerical example showing the data output from each module of the system for a single cinema location and business day. Table 4.2 illustrates the data that is output from the box office forecasting models which is the predicted number of tickets sold by title and date. Also included in this table for the purposes of this illustrative example are the actual tickets sold and the delta percent between predicted and actual tickets sold.



Table 4.2 data output from the box office forecasting model

<b>Show_Date</b>	<b>Title_Name</b>	<b>Days Since Release Date</b>	<b>Actual Tickets Sold*</b>	<b>Predicted Tickets Sold</b>	<b>Predicted vs. Actual Tickets Delta %*</b>
3/7/2014	300: Rise of An Empire	0	1412	1251	-11%
3/7/2014	Mr. Peabody And Sherman	0	927	796	-14%
3/7/2014	Non-Stop	7	758	747	-2%
3/7/2014	Son Of God	7	185	172	-7%
3/7/2014	Pompeii	14	179	292	63%
3/7/2014	RoboCop	23	104	134	29%
3/7/2014	The Lego Movie	28	662	445	-33%
3/7/2014	Monuments Men	28	151	139	-8%

\*not part of decision support system, included for illustrative purposes

Table 4.3 illustrates the micro demographic projections that are output from the decision support system. For each location, the number of tickets sold is predicted for each demographic and age cohort and show date.

Table 4.3 micro demographic projections

<b>Field</b>	<b>Value</b>
<b>Show Date</b>	3/1/2014
<b>Predicted Tickets Sold</b>	3977
<b>predicted_location_F14_18</b>	85
<b>predicted_location_M14_18</b>	113
<b>predicted_location_F19_34</b>	916
<b>predicted_location_M19_34</b>	908
<b>predicted_Location_F35_49</b>	490
<b>predicted_location_M35_49</b>	412
<b>predicted_Location_F50_64</b>	256
<b>predicted_Location_F65P</b>	98
<b>predicted_Location_M65P</b>	73
<b>predicted_Location_Parents</b>	627

Table 4.4 depicts the predictions and recommendations that are output from the concession forecasting extension. For each location and date projections are made to recommend the staffing requirements and forecast concession spend, hot food spend and yogurt/café spend per loyalty visit. In this table actual spends have been included for illustrative purposes.

Table 4.4 recommendations from concession forecasting extension

<b>Location_ID</b>	77777
<b>Conformed_Show_Date</b>	3/1/2014
<b>Forecast Concession Spend per Loyalty Visit</b>	\$ 5.79
<b>Actual Concession Spend per Loyalty Visit</b>	\$ 5.52
<b>Forecast Concession Staff Required</b>	14.6
<b>Actual Concession Staff Required</b>	15
<b>Forecast Hot Food Spend per Loyalty Visit</b>	\$ 0.76
<b>Actual Hot Food Spend per Loyalty Visit</b>	\$ 0.66
<b>Forecast Hot Food Staff Required</b>	8.9
<b>Actual Hot Food Staff Required</b>	9
<b>Forecast Yogurt Cafe Spend per Loyalty Visit</b>	\$ 0.45
<b>Actual Yogurt Cafe Spend per Loyalty Visit</b>	\$ 0.50
<b>Forecast Yogurt Cafe Staff Required</b>	14.8
<b>Actual Yogurt Cafe Staff Required</b>	15

Table 4.5 illustrates an example of the volume based scheduling rules for a single location.

Table 4.5 Volume Based Scheduling Rules for a single location

<b>Team (group)</b>	<b>Demand Per Headcount</b>
<b>Back Of House</b>	666
<b>Booth</b>	436
<b>Box Office</b>	129
<b>Cafe/Yogurt/Pretzels</b>	73
<b>Concessions</b>	54
<b>Floor</b>	49
<b>Guest Services</b>	741
<b>Hot Foods</b>	132

#### **4.6. Summary**

Empirical testing was done by applying the DSS micro attendance forecasts to labour recommendations and the system's recommendations were closer to optimal compared to the actual schedule used by managers. The micro concession extension demonstrates the potential upside of utilizing concession forecasting in conjunction with the micro box office forecasting model for more optimal labour scheduling. In the audience composition model, the independent variables used are good predictors for most age cohorts with all models having an explanatory power over 39%, which is acceptable in the social sciences for gaining directional insights into the variables that impact the behavior of humans. These results are also directionally similar to the overall explanatory power of the model built by Redondo & Holbrook (2010) with 45% of variance explained. This study extends the work of Redondo & Holbrook (2010) in a new and interesting way, exploring the definition of the characteristics of heterogeneous moviegoer groups using customer loyalty data. Through this approach, conventional wisdom is reaffirmed and new insight is gained into the

relationship between the key film attributes (genres, movie ratings, critic reviews, production budgets, film length) and sub-segments of the movie-going population.

There are several practical use cases for predictive audience insights to inform the decision making processes of industry stakeholders (such as producers, distributors exhibitors, customers) (see Somlo et al. 2011 for a full explanation of the functions of the different industry stakeholders). Many stakeholders obtain this information from a media industry ratings and measurement leader, Nielsen, which offers pre-release customer research in major American markets to compile a number of metrics for new release films, such as interest for key demographic and ethnic segments. Movie industry stakeholders utilize this information to inform key operational, advertising and marketing processes for new release films. The key audience for new release films can be utilized to make scheduling decisions like whether the film should play at the theatre, how many shows should be scheduled at that location or what size auditorium should the film play at that location. Media executives can use the key audience by movie to place relevant advertising with films based on the advertiser's target customer profile. Moreover, the predicted composition of the audience can be used in conjunction with demand forecast data to provide reach and frequency predictions by film for granular demographic segments. The target marketing applications for the film itself are vast. For example, demographic customer segments are used to inform decisions around:

- affiliate advertising with complementary products targeted at key demographics
- targeted advertising buys on social platforms like Facebook
- targeted mass media advertising, based on key demographics
- personalized digital communications via the loyalty program communication channels (e.g. email, SMS, website, mobile app)

Overall, the results are promising and indicate that the DSS system described in this paper is suitable for use in practice for demand driven labour recommendations and optimization of marketing and advertising in cinemas. In the next section, Chapter 5 - Conclusion, the 8 decision support factors are revisited with a summary of this study's research contributions and opportunities for future research to improve the prototype DSS.

## **Chapter 5. Conclusion: Discussion for use of DSS in practice, summary of contributions and future research opportunities**

For quantitative models to impact the theatrical industry, they must be used in practice. Events, such as the unprecedented success of the movie Avatar in 2009, contribute to a common view that the movie industry is different from other industries because of its unpredictable scenarios, which are challenging to model quantitatively and require human judgement and expertise to manage. This perspective is typically justified by the artistic nature of films, but also by the dynamic, complex, relationship-driven environment in which distributors and exhibitors operate. In recent years, the non-analytical perspective has been countered by a growing collective of academics and industry champions, who are successfully demonstrating that, despite dynamic complexities, the industry's decision-making environment can be modeled quantitatively. One example of the power of data-driven decision support systems in the movie industry is described in Chung et al. (2011), which uses Blockbuster as a case study to demonstrate the potential of integer programming models for revenue optimization in the retail movie industry. Moreover, model implementations, such as that at Pathé in the Netherlands (Eliashberg et al. 2009b), show that a balance between managerial input and quantitative methods provides the best results (over \$1 million annually to the bottom line in this case), and fuel the growing perspective that favors the use of quantitative models to maximize revenue in theatrical exhibition. New challenges, such as increased variety and types of film products, scaled price points, and varied technological and auditorium experiences, make exhibitor decision making more complex in a digital era, thus resulting in more opportunities for quantitative modeling in management practices.

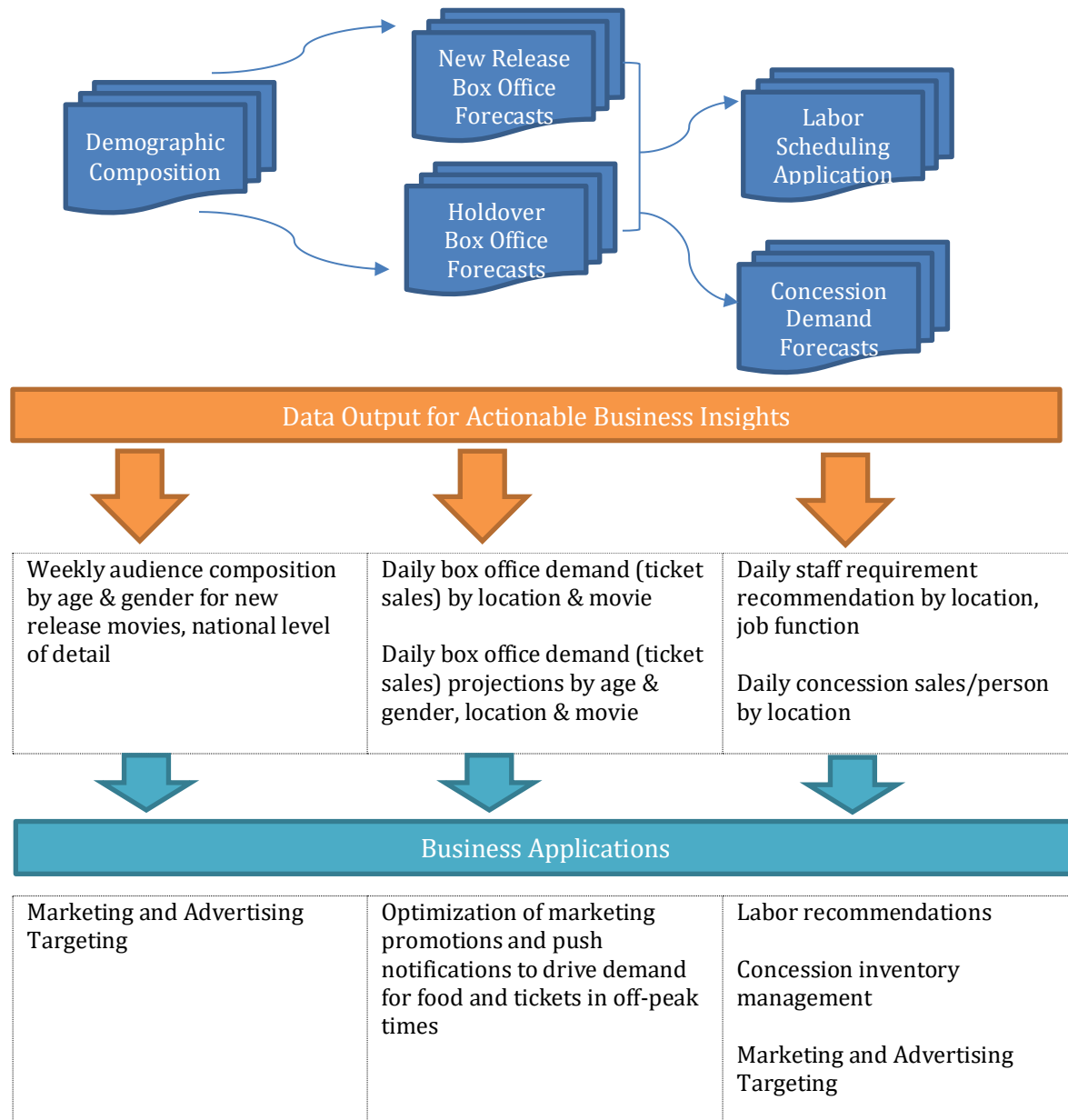


Figure 5.1 - An integrated system for DSS in movie exhibition, models are mapped to data output and business applications

Figure 5.1 depicts the main contribution of this dissertation, an integrated system for decision support in movie exhibition. Through the prototyping of this integrated decision support system the potential for improving movie exhibitor profitability through demand

driven labour recommendations is demonstrated. The integrated system developed through this research is unique to the existing body of academic work because it uses both customer loyalty and point of sale data to generate recommendations for demand-driven labor recommendations and demographic audience projections that can be used to inform marketing and advertising at a micro (location) scale. Another contribution is that the potential use cases and incremental value of micro demographic forecasts are demonstrated through empirical testing of the integrated DSS system using actual movie exhibitor data. In summary the key contributions are:

1. Development of an integrated decision support system for movie exhibitor operations
2. Demonstration of potential use cases and incremental value of micro demographic forecasts through empirical testing of the integrated DSS system using actual movie exhibitor data
3. The integrated DSS developed in this study considers more of the eight critical decision support factors than any of the systems developed in existing studies. The contributions for each decision support factor are discussed further in the section that follows.

In Table 5.1 the quantitative models described in Chapter 2 of this paper are compared to the eight decision support factors for movie scheduling. An additional contribution of the system developed in this dissertation is that it considers more of the critical decision support factors than any of the systems developed in existing studies. The remainder of this paper is organized by these decision support factors, for each factor new perspectives



for future research are detailed and where applicable the research contributions of this dissertation are summarized.

Table 5.1 - The Quantitative Models Described in This Paper are compared to the Eight Decision Support Factors for Movie Scheduling and the prototype system

Authors and Factor	1. Theater-specific factors	2. Auditorium attributes	3. Showtime schedule	4. Film product Competition	5. Network considerations	6. Price elasticity	7. Multiplex staffing	8. Marketing personalization
Eliashberg et al. (2001)	X	X						
Gaytán-Iniestra et al. (2006)		X						
Eliashberg et al. (2009a)	X		X					
Eliashberg et al. (2009b)	X	X	X					
Raut et al. (2009)		X	X	X				
Dawande et al. (2010)		X	X					
Somlo et al. (2010)	X				X			
De Roos & McKenzie (2014).						X		
<b>This Study</b>	<b>X</b>	<b>X</b>		<b>X</b>	<b>X</b>		<b>X</b>	<b>X</b>

**(1) Theatre-Specific Attributes.** Customer loyalty, research, and social media programs run by most major theatrical exhibitors provide rich insights that can be leveraged to inform business operations. This dissertation work utilizes demographic data sourced from a loyalty program as an input for demand for specific movies at individual theatre locations. Some of the demographic data used in this study was at a macro (national) level of detail. As future research an experiment to quantify the benefit of using micro level demographic projections is recommended. This study also identifies segments of moviegoers based on movie attribute preferences. Future research could analyse how the composition of the trade population of each cinema location is unique and identify clusters of theatres with similar audiences.

(2) **Auditorium Attributes.** At the time of writing this paper, no studies that consider the premium capabilities found in some auditoriums in scheduling models were found. This study accounts for premium capabilities in the micro forecasting module. There is an opportunity to extend the existing basic scheduling model to provide a data-driven approach to (a) determine which movies to offer as a two-dimensional film versus a three-dimensional film, (b) which movies to offer in auditoriums with premium seating, and (c) forecast how these decisions will impact traditional business.

(3) **Showtime Schedule.** The advent of digital projection offers many new challenges related to the scheduling of showtimes. Distributors can push out more film content (i.e., products from their historic libraries) or alternate content (e.g., stream sports events, concerts, arts) to exhibitors, enabling them to serve a broader range of consumer preferences. There is a need to create demand-forecasting models for content, which does not follow the standard consecutive play constraints. This model only includes film presentations which make up the majority of ticket sales, further research could extend the demand forecasts to other types of film content that can be an input to an extension of the basic scheduling model for new types of content. Another interesting area of research is simulating how exhibitor-distributor contracts could evolve to maximize profit for all stakeholders by using a data-driven approach and relaxing constraints, such as consecutive play.

(4) **Film Product Competition.** As the amount of content grows, decision support systems will become more critical to theatrical operations. To maximize cinema profit through scheduling, understanding the impact that demand for one movie has on the

demand for another is necessary. Most of the existing scheduling models do not account for cross-elasticity at a micro level. In this study, cross-elasticity is considered at a micro level by factoring for the number of other films playing at the location in the same genre. This could be extended to consider the overall demand projection for specific movies (for example: so that different weight would be given to major blockbusters versus small budget limited appeal movies). Modeling the cross-elasticity of movies is a rich area of research opportunity with great potential to incorporate these metrics to showtime scheduling and marketing optimization models.

**(5) Network Considerations (Cannibalization).** Spatial interdependencies between multiplexes have been explored in the literature on location research, addressing topics such as location-allocation strategies and cannibalization (Chisholm and Norman 2002). However, there is an opportunity to extend the current scheduling models in the literature to incorporate geographic interdependencies between nearby locations (i.e., those that are part of the same theater chain). For leading exhibitors with large networks of multiplexes, such as Cineplex, Regal, Cinemark and AMC, network considerations are an important aspect of operations management. Network considerations may include: nearby theaters playing (or not playing) the same movie, nearby theaters playing movies of the same genre, nearby theaters having special events, etc. Urban centers exhibit significant crossover between nearby locations. Therefore, accounting for the proximity of locations playing the same film, or playing that film in an enhanced auditorium, such as IMAX, may provide more accurate forecasts. A contribution of this study is that the DSS developed a preliminary method to factor in network considerations in the micro forecasts by constructing a variable which indicates the number of nearby locations the film is playing at. This

variable was found to be of statistical significant for many locations in the micro forecasting model.

(6) **Price Elasticity** (Premium and Discount Offerings). Existing models assume uniform pricing across all films. Therefore, an opportunity exists to build models that assess price sensitivity relating to individual cinema locations, individual film products, and individual customers. These models can help movie exhibitors determine which movies to offer on premium screens, where to build premium technology, and how much to charge for admission.

(7) **Multiplex Staffing**. This study demonstrates the application and potential for applying operations research methods to improve movie exhibitor profitability through demand driven scheduling. Using customer loyalty data, concession purchase behavior was modeled and the output was combined with the micro attendance forecasts to recommend staffing by work station type at a micro (location) level. Further testing, by implementing this DSS in practice in select pilot locations and comparing the results to a control group, is recommended.

(8) **Marketing Advertising Personalization and Optimization**. In the era of big data and analytics, a customer-driven perspective to business strategy is more important than ever. Location-based mobile messaging opens up the potential for exhibitors to deliver messages and offers to moviegoers while they are visiting a theater. Movie previews and big-screen commercials offer an immersive experience for brands to communicate with consumers, and movie-loyalty data make it theoretically possible to optimize media rate cards based on granular audience segmentation, targeting, and profiling. In the private

sector, leading exhibitors managing omnichannel marketing and distribution strategies will demand complex analytics, as consumers gravitate toward movie brands that engage them with personalized content (i.e., Netflix). Omnichannel marketing has drastically changed the way films and other entertainment experiences are communicated to consumers. New ways of connecting and engaging with movie fans using social media and direct messaging to their emails and cell phones have proven to impact film and television performance so much that the media-ratings giant Nielsen has started to incorporate social analytics as a standard audience metric (Vanderbilt 2013). There is an opportunity to incorporate this data into the DSS developed in this study.

This study demonstrates the application and potential for customer loyalty data in projecting demand and key customer segments for new products. The opportunities to explore the use of models based on customer data are vast; however, the proliferation of academic studies is limited due to the proprietary nature of the data. Operational decisions around film and screen booking require the key customer segment predictions be compared with location-specific customer attributes. To inform affiliate advertising with complementary products, key customer segment predictions must be compared with product-specific customer attributes. For large organizations, there may be hundreds or thousands of locations or products, which is too much data for managers to sift through manually. Therefore, programmatic methods enable the most efficient use of customer segment predictions. In an omnichannel world, analytics and quantitative models play an important role in developing and executing integrated marketing communication strategies, providing a 360-degree view of the consumer and helping marketers ensure that

film products are promoted to the right people, at the right time, and through the right channels.

## References

- Ainslie, A., X. Dreze & F. Zufryden. (2005). Modeling Movie Life Cycle and Market Share. *Marketing Science* 24(3) 508–517
- Asur S, & B. Huberman. (2010). Predicting the future with social media. In: Proceedings of the IEEE/WIC/ACM international conference on web intelligence, IEEE Computer Society Press, Washington, DC, pp 492–499
- Basuroy, S., Chatterjee, S. & A. Ravid. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power and Budgets. *The Journal of Marketing*. 67(4), 103-117.
- Chisholm, D. & G. Norman. (2002). Spatial Competition and Demand: An Application to Motion Pictures. *Discussion Papers Series, Department of Economics, Tufts University with number 0216*
- Chung, C., Dawande, M., Rajamani, D., & C. Sriskandarajah. (2011). *A short range scheduling model for Blockbuster's order-processing operation*. *Interfaces* 41(5):466–484.
- Cineplex Entertainment Press Releases accessed at <http://cineplexgalaxy.disclosureplus.com/About-Us/Default.aspx> dated June 22 2011 & Feb 08 2012 accessed on July 29 2013.
- Davis, P. (2006). Spatial competition in retail markets: movie theaters. *RAND Journal of Economics*. 37(4) 964-982.
- Davis, P. (2006). Measuring the business of stealing, cannibalization and market expansion effects of entry in the U.S. Motion Picture Exhibition Market. *Journal of Industrial Economics*. 54(3) 293-321.

- Danaher, B. & M. Smith (2014). Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales. *International Journal of Industrial Organization*, 33(3), 1-8
- Dawande, M., Drobouchevitch I., Rajapakshe, T. & C.Sriskandarajah. (2010). Analysis of Revenue Maximization under Two Movie-Screening Policies. *Production and Operations Management*, 19 (1), 111-124
- De Roos, N. & J. McKenzie (2014). Cheap Tuesdays and the demand for cinema. *International Journal of Industrial Organization*, 33(3), 93-109.
- Einav, L. (2007). Seasonality in the U.S. Motion Picture Industry. *RAND Journal of Economics* 38(1) 127–145.
- Elberse, A. & J. Eliashberg. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science* 22(3) 329–354.
- Eliashberg, J., S. Swami, C. Weinberg & B. Wierenga. (2001). Implementation and Evaluation of SilverScreeners: A Marketing Management Support System for Movie Exhibitors. *Interfaces: Special Issue on Marketing Engineering* 31(3) 108–127.
- Eliashberg, J., Elberse, A. & M. Leenders. (2006). The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions. *Marketing Science* 25(6) 638–661.
- Eliashberg, J., Weinberg, C. & S. Hui (2008) Decision Models for the Movie Industry in Wierenga, B. (ed), *Handbook of Marketing Decision Models*. DOI: 10.1007/978-0-387-78213-3\_13, Springer Science Business Media, LLC 2008



- Eliashberg, J., Hegie, Q., Ho, J., Huisman, D., Miller, S., Swami, S., Weinberg, C & B. Wierenga. (2009a). Demand-driven scheduling of movies in a multiplex. *International Journal of Research in Marketing*, 26(2) 75–88.
- Eliashberg, J., Swami, S., Weinberg, C. & B. Wierenga. (2009b). Evolutionary approach to the development of decision support systems in the movie industry. *Decision Support Systems*.47(1) 1-12.
- Eliashberg, J. & S. Shugan. (1997). Film Critics: Influencers or Predictors? *The Journal of Marketing*, 61(2), 68-78.
- Gaytán-Iniestra J., Arroyo-López P., & N. Viktorovna-Gorina. (2006). Optimal allocation of movies to screens in movie theaters. Presentation, Proc. Indust. Engrg. Res. Conf. (IERC), Institute of Industrial and Systems Engineers, Norcross, GA.
- Goff Inglis, K. & S. Zolfaghari. (2017). A Review of Scheduling Problems and Research Opportunities in Motion Picture Exhibition. *Interfaces*, 47(2), 175-187.
- Hermann, B. (2012). Cineplex SCENE member crossover and trade area analysis: A comparison of the Toronto and Vancouver CMA. Unpublished master's thesis, Ryerson University, Toronto, Ontario.
- Ho, J. (2005). Marketing models of entertainment products. *Unpublished doctoral thesis, University of British Columbia*.
- Ho, J. & C. Weinberg. (2011). Segmenting consumers of pirated movies. *Journal of Consumer Marketing*, 28(4) 252–260
- Jedidi, K., Robert K., & C. Weinberg. (1998). Clustering at the Movies. *Marketing Letters* 9, no. 4. 393-405.

- Krider, R. & C. Weinberg. (1998). Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game. *Journal of Marketing Research* 35(1) 1–15.
- Lehmann, D. & C. Weinberg. (2000). Sales Through Sequential Distribution Channels: An Application to Movies and Videos. *Journal of Marketing* 64(3) 18–33.
- Liaw, A. (2005). *Breiman and Cutler's Random Forests for Classification and Regression*. accessed at <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> dated October 7 2015 accessed on May 2 2017.
- Lilien, G. & A. Rangaswamy. (2005). *Marketing Engineering*. Trafford Publishing.
- Litman, B. (1983). Predicting Success of Theatrical Movies: An Empirical Study. *Journal of Popular Culture*, 16 (2) 159-175.
- Litman, B. & H. Ahn. (1998). Predicting Financial Success of Motion Pictures: The Early '90s Experience. Litman B.R., Ed. *The Motion Picture Mega-Industry*. Needham Heights, MA: Allyn & Bacon.
- McKenzie, J. (2012). The Economics of Movies: A Literature Survey. *Journal of Economic Surveys*, 26(1), 42-70.
- McKenzie, J. & D. Walls. (2016). File Sharing and Film Revenues: Estimates of Sales Displacement at the Box Office. *Journal of Economic Analysis & Policy*, 16(1), 25-57.
- Medved, M. (1992). *Hollywood vs. America: Popular Culture and the War on Traditional Values*. HarperCollins. ISBN 0-06-016882-X.
- Motion Picture Association of America. (2013). *Theatrical Market Statistics*. Industry Publication

- Moul, C. (2005). *A Concise Handbook of Movie Industry Economics*. Cambridge University Press, Cambridge, UK.
- Neelamegham, R. & P. Chintagunta. (1999). A Bayesian model to forecast new product performance in domestic and international markets, *Marketing Science*. 18 (2) 115–136.
- PR Newswire (2008, Dec 10). Netflix awards \$50,000 progress prize in year two of multi-year, multi-national netflix prize competition. Retrieved from <http://ezproxy.lib.ryerson.ca/login?url=http://search.proquest.com/docview/448530543?accountid=13631>
- Radas, S. & S. Shugan. (1998). Seasonal Marketing and Timing Introductions. *Journal of Marketing Research*, 35 (3) 296-315.
- Raut, S., Swami, S. & M. Moholkar. (2009). Heuristic and meta-heuristic approaches for multi-period shelf-space optimization: the case of motion picture retailing. *Journal of the Operational Research Society*. 60(10) 1335 – 1348.
- Ravid, S. (1999). Information, Blockbusters, and Stars: A Study of the Film Industry. *Journal of Business* 72(4) 463–492.
- Redondo, I. & M. Holbrook. (2010). Modeling the appeal of movie features to demographic segments of theatrical demand. *Journal of Cultural Economics*, 34(4), 299-315
- Ridgeway, G. (2017). *Generalized Boosted Regression Models*. accessed at <https://cran.r-project.org/web/packages/gbm/gbm.pdf> dated March 21 2017 accessed on May 2 2017.
- Sawhney, M. & J. Eliashberg. (1996). A Parsimonious Model for Forecasting Gross Box Office Revenues of Motion Pictures. *Marketing Science*, 15(2), 113-131.

- Sharda, R. & D. Delen. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications*, 30(2), 243–254.
- Simonoff, J. & I. Sparrow. (2000). Predicting Movie Grosses: Winners and Losers, Blockbusters and Sleepers. *Chance Magazine*, 13(3), 15-24.
- Smith, S. & V. Smith. (1986). Successful Movies: A Preliminary Empirical Analysis. *Applied Economics*, 18(5), 501-507.
- Somlo, B. (2005). Distribution planning for the motion picture industry: Models, applications and analysis. *Unpublished doctoral thesis, University of California, Los Angeles*.
- Somlo, B., Rajaram, K. & R. Ahmadi. (2011), Distribution Planning to Optimize Profits in the Motion Picture Industry. *Production and Operations Management*, 20(4), 618-636.
- Swami, S., Eliashberg, J. & C. Weinberg. (1999). SilverScreeners: A Modeling Approach to Movie Screens Management. *Marketing Science*, 18(3), 352-372.
- Swami, S. (2006). Research Perspective at the Interface of Marketing and Operations: Applications to the Motion Picture Industry. *Marketing Science*, 25(6), 670-673.
- The Numbers (2016). Movie Budgets accessed from the-numbers.com (<http://www.the-numbers.com/movie/budgets/all>)
- Vanderbilt, T. (2013). *The New Rules of The Hyper-Social, Data-Driven, Actor-Friendly, Super-Seductive Platform Age of Television*. *Wired: Underwire* (accessed at <http://www.wired.com/underwire/2013/03/nielsen-family-is-dead/> on May 30 2013)
- Wallace, W. & M. Holbrook. (1993). The Role of Actors and Actresses in the Success of Films: How Much is a Movie Star Worth? *Journal of Cultural Economics*, 17(1), 1-27.

- Wierenga, B., G. Bruggen & R. Staelin. (1999). The Success of Marketing Management Support Systems. *Marketing Science* 18(3), 196–207.
- Weinberg, C. (2006). Research and the Motion Picture Industry. *Marketing Science*, 25(6), 667-671.
- Westland, J. (2012). The adoption of social networking technologies in cinema releases. *Information Technology Management*. 13(3), 167-181.
- Zufryden, F. (1996). Linking Advertising to Box Office Performance of New Film Releases: A Marketing Planning Model. *Journal of Advertising Research* July–August 29–41.