STARS Mission Request Data Analysis
Project Report – ISYE 6748

Jesse Keith – 903650252 – jkeith37@gatech.edu

Abstract—In this project, Mission Request data for a helicopter ambulance service (STARS) is analyzed to understand and model periods were demand-for-service exceeds capacity. The goal of this project was to predict the need and timing for additional infrastructure. The modeling approaches in this project focus primarily on time-series approaches (including the Holt-Winters Exponential Smoothing & Prophet algorithms).

1 THE ORGANIZATION

For my practicum, I partnered with an organization that is headquartered in my home city of Calgary, Alberta, Canada. STARS (Shock Trauma Air Rescue Service) is a helicopter air rescue and ambulance service that operates across the western Canadian provinces of British Columbia, Alberta, Saskatchewan, and Manitoba. STARS' mission statement is to provide *critical care, anywhere* [1], providing rapid critical care to a dispersed population across a large geographic area. STARS is one of the most recognizable and well-known not-for-profit organizations in Western Canada due to the highly visible role it plays within the healthcare system.

2 PROBLEM STATEMENT

Throughout the year, STARS must decline some mission requests for safety or technical reasons, such as weather that is not suitable for helicopter flight, or inadequate fuel availability along the route. However, in recent years, **STARS has seen an increase in the number of mission declines because the helicopter covering the call area is already busy on another mission.** This shows that at certain times STARS' existing infrastructure (helicopters, crews, etc.) is over-capacity and that investing in additional infrastructure/assets would lead to higher mission acceptance rates and to increased lives saved. However, investing in backup infrastructure is costly, and will likely be utilized only a portion of the time. Prior to making this investment, STARS requires a thorough quantitative evaluation of these resource allocation challenges. **The goal of this project will be to use the STARS Emergency Link Centre database to arrive at a data-driven understanding of the need and timing for additional infrastructure**. [2]

3 DATA SET

3.1 Non-Disclosure Agreement

In order to participate in this project, I was required to sign a non-disclosure agreement covering STARS' data and data warehouse. However, in this report I have included as much detail as possible on data exploration and modeling, redacting only specific data observations that may have made it possible to reconstruct portions of the underlying dataset.

3.2 Data Terminology

STARS' Emergency Link Centre data is stored according to several key terms or codes associated with STARS' operations. The key terms required for understanding this project are outlined below [3]:

<u>Request:</u>

> A request regarding the availability of the STARS helicopter to fly a mission (scene, interfacility) or SAR (search and rescue).
>
> Will either turn into a mission or be declined (STARS) or cancelled (requestor)

<u>Mission:</u>

> STARS helicopter lifts off (goes skids up) with the intent of a Scene or IFT or SAR event OR the AMCs go to a call in an Ambulance or Fixed Wing

without involvement of the STARS helicopter. Zero, one, or more patients can be associated with a mission.

Request Declined:

When a Request is decline by STARS because of a technical inability to accept the mission (example: weather that is not acceptable for helicopter flight).

Request Declined - Other Mission:

This sub-category of Request Declined is the focus of this project. This occurs when a Request is made for the STARS helicopter, but it is already occupied with an "Other Mission" (OM) leading to the request being declined.

3.3 Data Collection

STARS' mission and request data from 2010 onwards is stored in a data warehouse (DW). The first step in constructing a dataset that could be used for modeling was to explore and understand the DW. The data required for this analysis was stored across several database tables that host data views for Missions, Requests and Request-Declines.

Research, and collaboration with STARS' Database Administrator was required to understand the DW and construct the queries required to get a dataset. SQL queries were made to the DW using Python and the SQLAlchemy library.

Several consultations were needed with STARS' leadership and subject-matter-experts (SMEs) to vet and agree on an appropriate dataset. During these consultations, two important factors were uncovered that were incorporated into the dataset construction:

1. Emergency Link Centre SMEs noted that there was change in data coding practices in 2018. This led to a change in the trend of "other mission" classifications within the data from 2018-onward. These SMEs indicated that the data pre-2018 was of lower confidence/quality. For this reason, a higher level of focus has been placed on the data from 2018-onward.

2. STARS SMEs had differing views on the correct dataset to use for modeling. When Mission Requests are declined they are often coded with more than one reason for the decline (example: weather & fuel availability, etc), so it cannot be known which reason was the most decisive. As well, some Requests are later canceled by the requestor. Therefore, some SMEs felt that the number of Mission Declines that were coded as Other Mission was too high, and that not all these missions would have been executed even if the helicopter had been available.

   To account for this, a complex query was created that linked the Request data tables with the Mission data tables to find Other Mission Declines that were made during the time that the STARS helicopter was in the air. However, many SMEs felt that this number was too low because a helicopter may be occupied with a Mission for significant periods before and after it is in the air (preparation, post-mission checks and maintenance, etc). In the end, STARS leadership made the decision to proceed with the larger dataset of bulk OM declines. This dataset is the focus of this report. However, some data exploration and modeling were done with other data views.

Overall, the data collection and dataset preparation portion of this project took longer than initially planned, which led to compression in other portions of the project schedule. Understanding the structure of the data warehouse, building queries to gather the data, and then getting agreement and buy-in from STARS SMEs was a complex process that involved several meetings and collaboration sessions that were sometimes challenging to schedule. Ultimately, teamwork and discussion led to STARS SMEs and leadership making a decision on the best dataset to move forward with.

4 METHODOLOGY

The following section outlines the methodology that was used to approach this problem, and approximately follows the methods proposed in my midterm progress report [2].

4.1 Data exploration & Visualization

The initial stage in approaching this problem was to explore the data in the data warehouse by creating various datasets, aggregating summary statistics, and creating visualizations of the data, and by discussing and vetting these with the subject-matter-experts at STARS to ensure the correct dataset had been developed and that the project problem was clearly understood.

Constructing visualizations of the dataset allowed me to understand the problem, the dataset, and develop an approach for modeling. Some of the key visualizations that outline the problem/opportunity, or highlight important features in the data are outlined below:

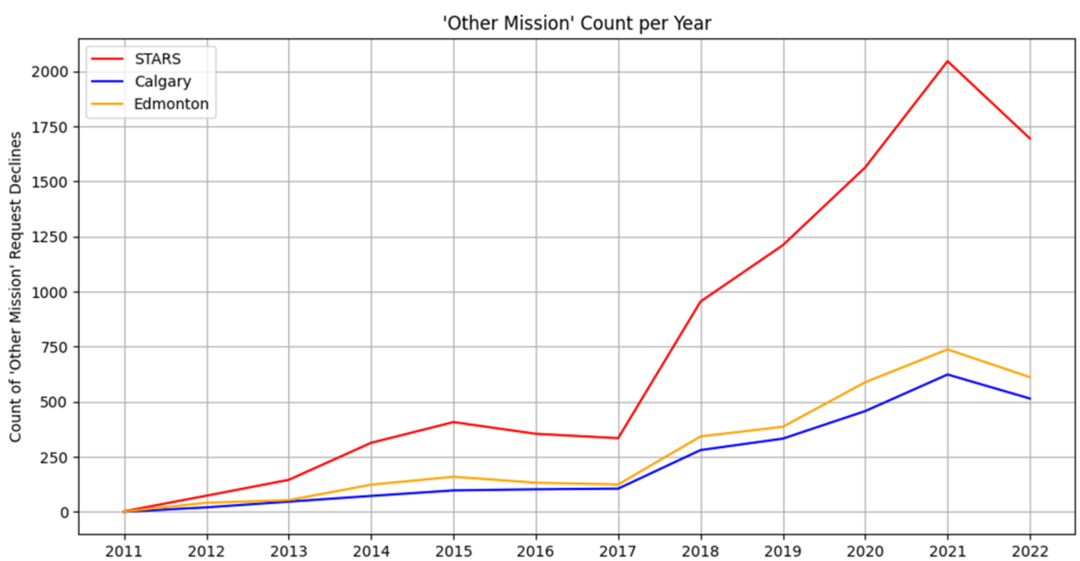

Figure 1 – **Total OM decline count per year (2011 – 2022)**

Figure 1, provides an illustration of the problem statement outlined in Section 2: Requests for service that are being declined due to OM have been increasing in recent years.

Another illustration of the problem statement is shown in Figure 2, which shows this challenge using the bulk data for Missions and Requests. We can see that mission requests have been increasing but the number of missions being flown from each base are generally flat, or increasing at a slower rate.

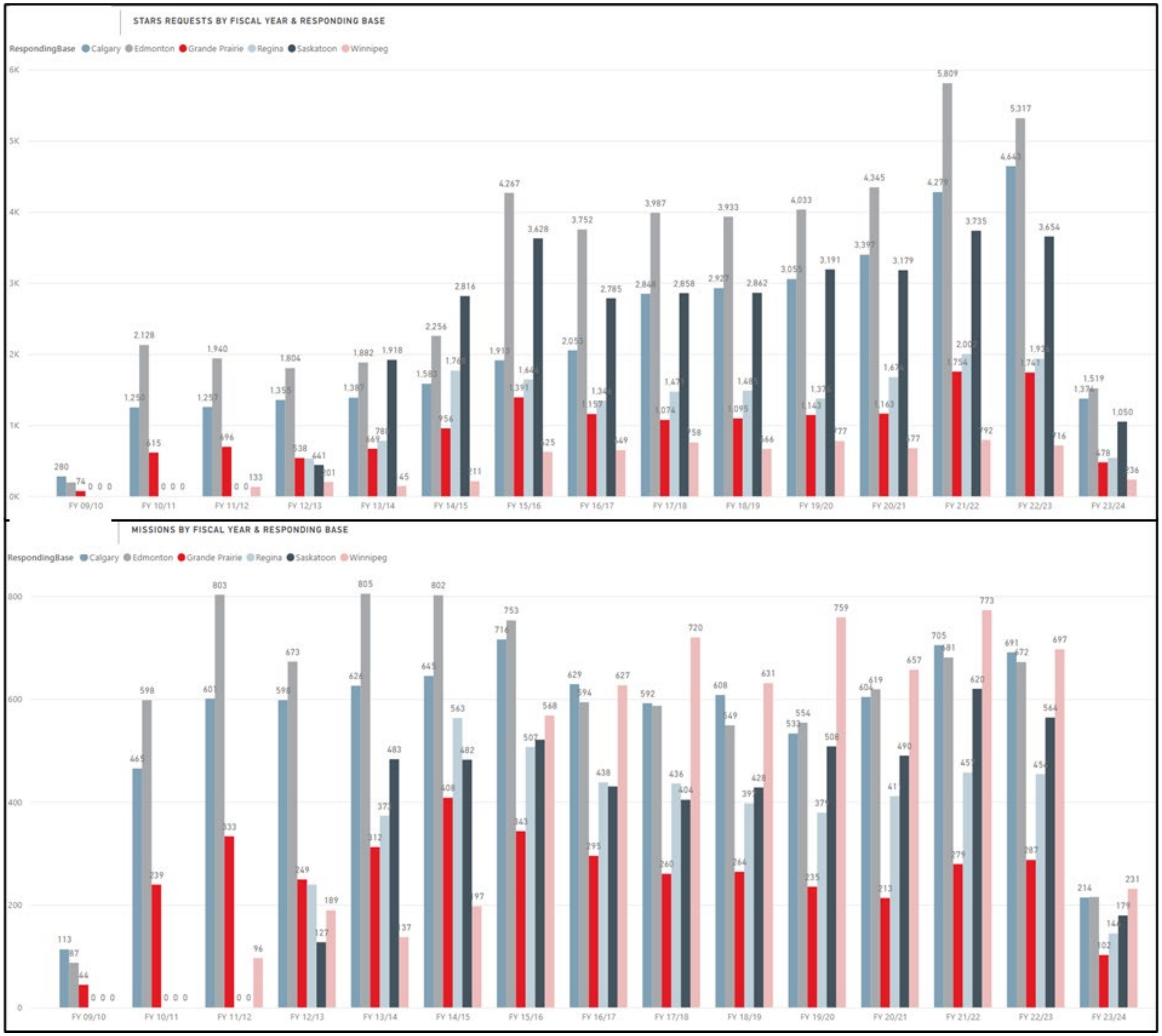

Figure 2 – **Missions Flow vs Requests (by Base) [4]**

Focusing specifically on OM Declines, I created several different time-scale aggregations to understand the frequency, trend and seasonality within the data. Figures 3 & 4, show the daily frequency of OM for the Calgary and Edmonton bases, respectively.

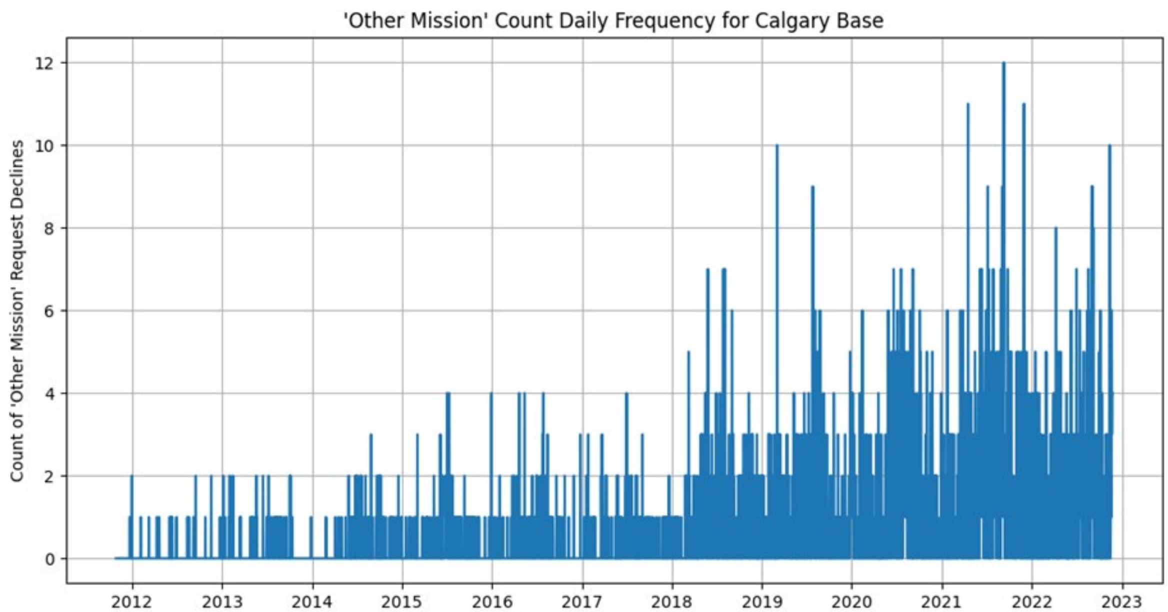

Figure 3 – **OM Count Daily Frequency (Calgary Base)**

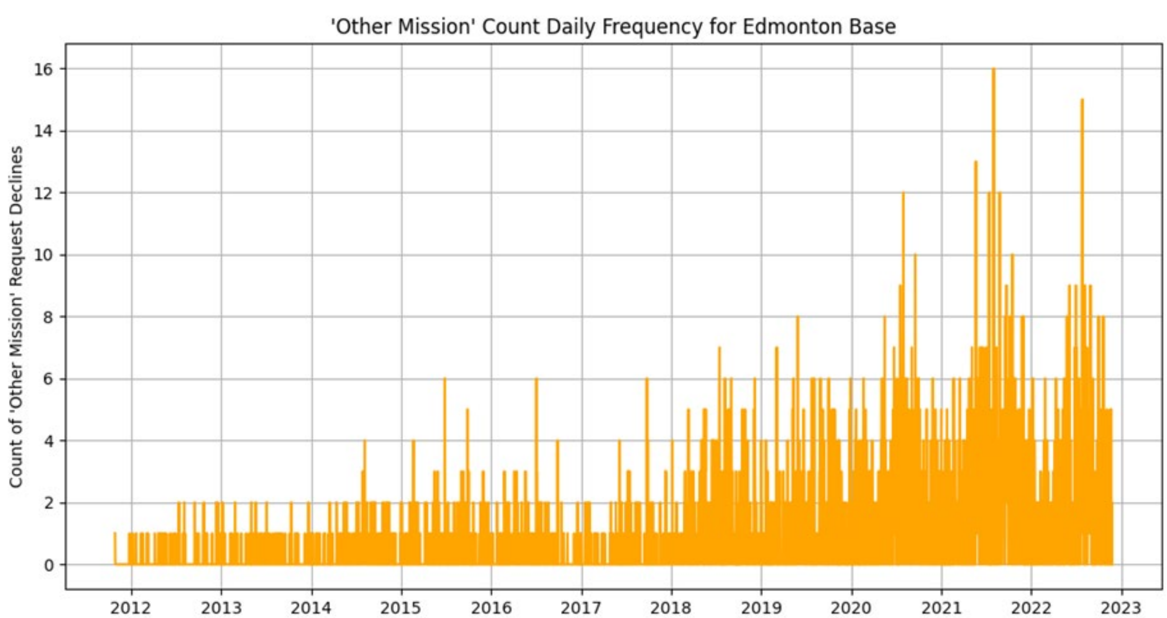

Figure 4 – **OM Count Daily Frequency (Edmonton Base)**

Figure 3 & 4, illustrate a couple of key features of the dataset. Firstly, they provide another illustration of the problem statement, namely that OM missions are on an increasing trend. Secondly, they show that there is a clear seasonality component to the dataset. Both points – trend & seasonality – indicate that this dataset might be best approached using time-series methods.

Looking at the dataset from a weekly perspective, Figures 5 & 6, also indicate a clear seasonal component. As well, these figures highlight the variability within the data; i.e., there are some weeks that much more busy than other weeks, which might be considered as possible outliers.

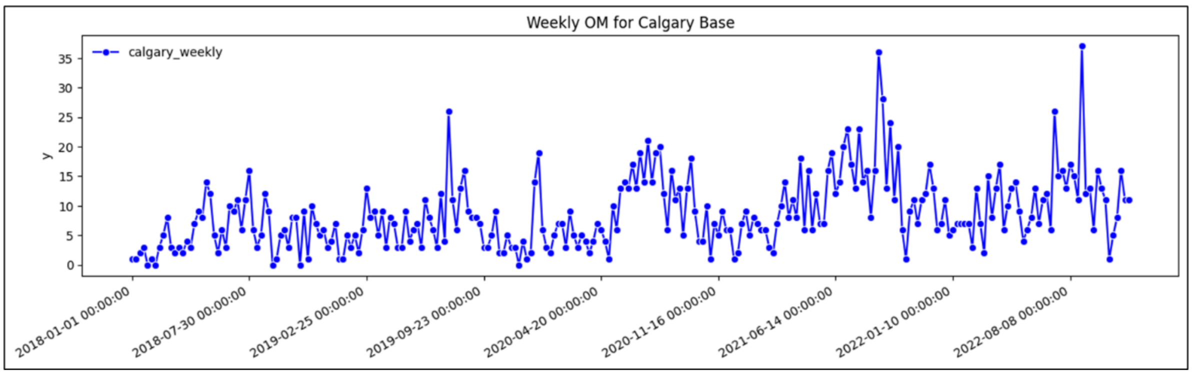

Figure 5 – **Weekly OM Mission Declines (Calgary Base)**

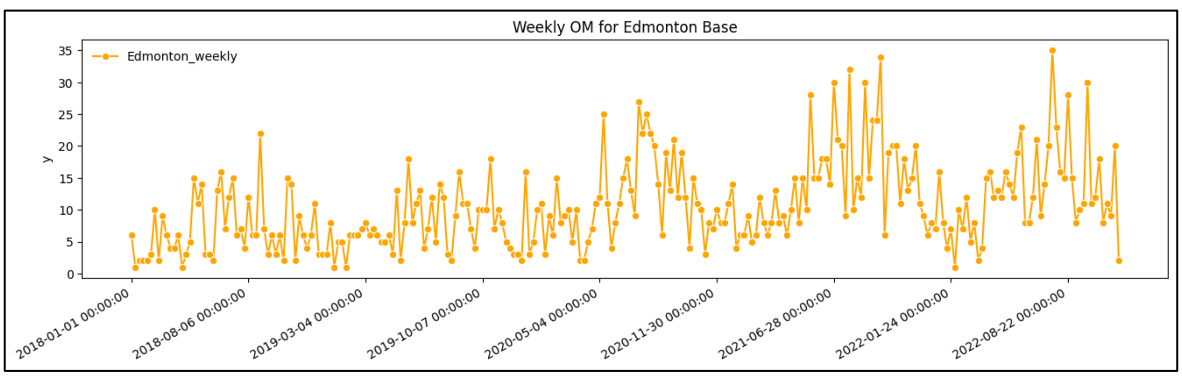

Figure 6 – **Weekly OM Mission Declines (Edmonton Base)**

Given the clear indication of trend and seasonality within the data, a time-series decomposition was performed on the weekly data, using the Statsmodels Time-Series-Analysis library to get an illustration of these underlying components, which is shown in Figure 7. Again, we can see representations of the trend and seasonality.

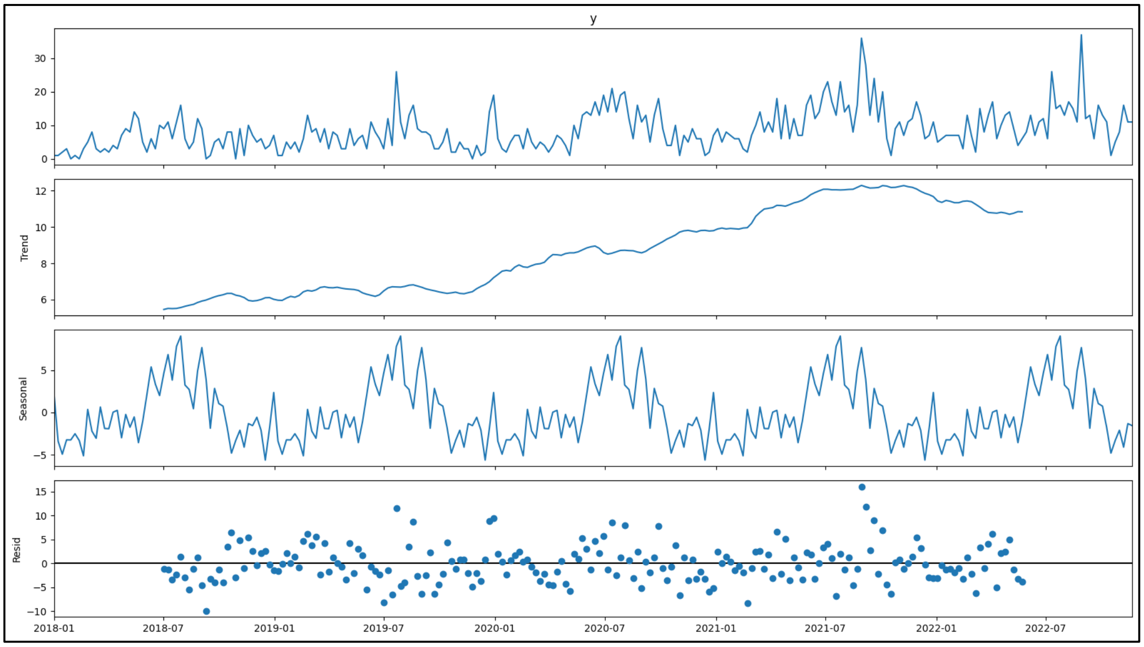

Figure 7 – **Time-series Decomposition for Weekly OM Declines (Calgary Base)**

Based on the clear indications of seasonality within the data, I further investigated the composition of this seasonality on different timescales. Indications from SMEs, and general understanding of the problem, indicated a couple key periods to investigate for high demand for the helicopter.

Looking at the data in terms of days-of-the-week and isolating for days when OM declines are highest (>=5 in Figure 8 & >=8 in Figure 9). We can see that a higher-than-expected number of high OM days occur on the weekends (note: an even split for the 7 days of the week would results in 14.3% of events occurring on each day of the week).

Additionally, looking at the annual seasonality within the data, we can see that the key months are during the summer (June – September), as shown in Figures 10 & 11.

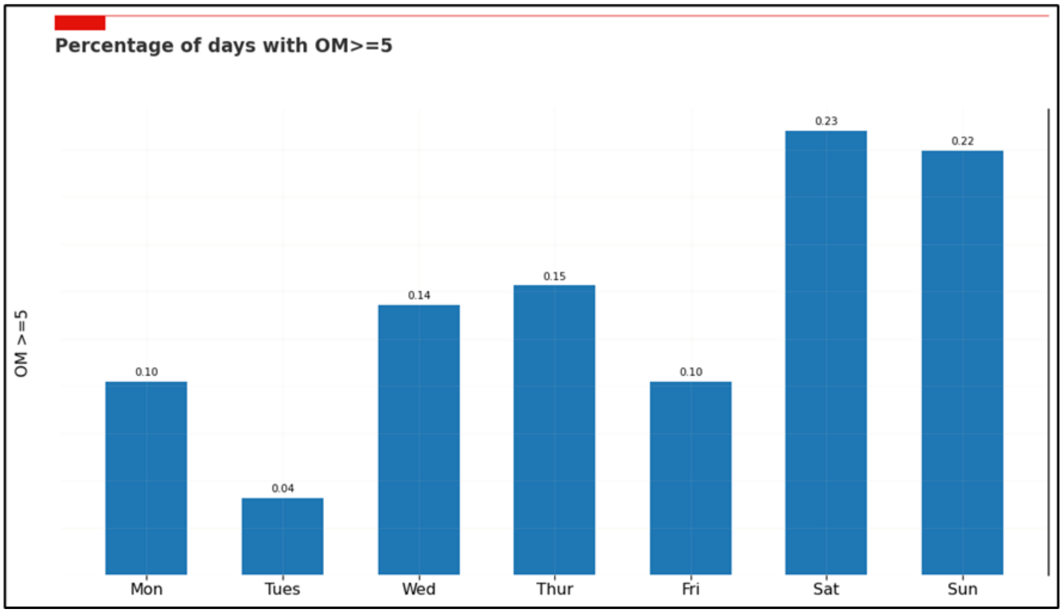

Figure 8 – **Percentage split for high OM occurrences (>=5 OMs) (days-of-the-week)**

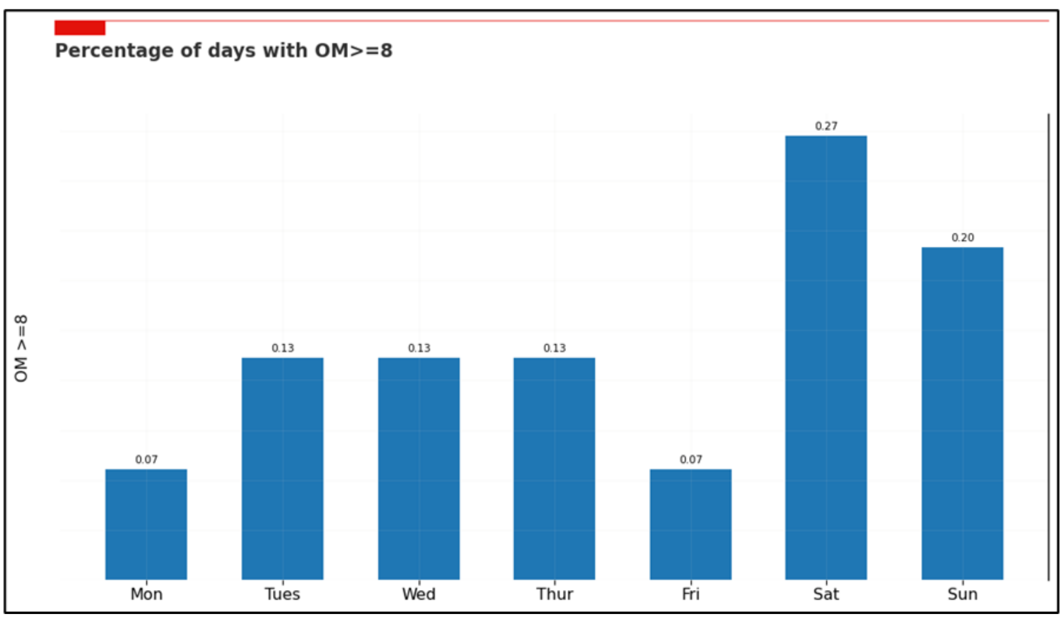

Figure 9 – **Percentage split for high OM occurrences (>= 8 OMs) (days-of-the-week)**

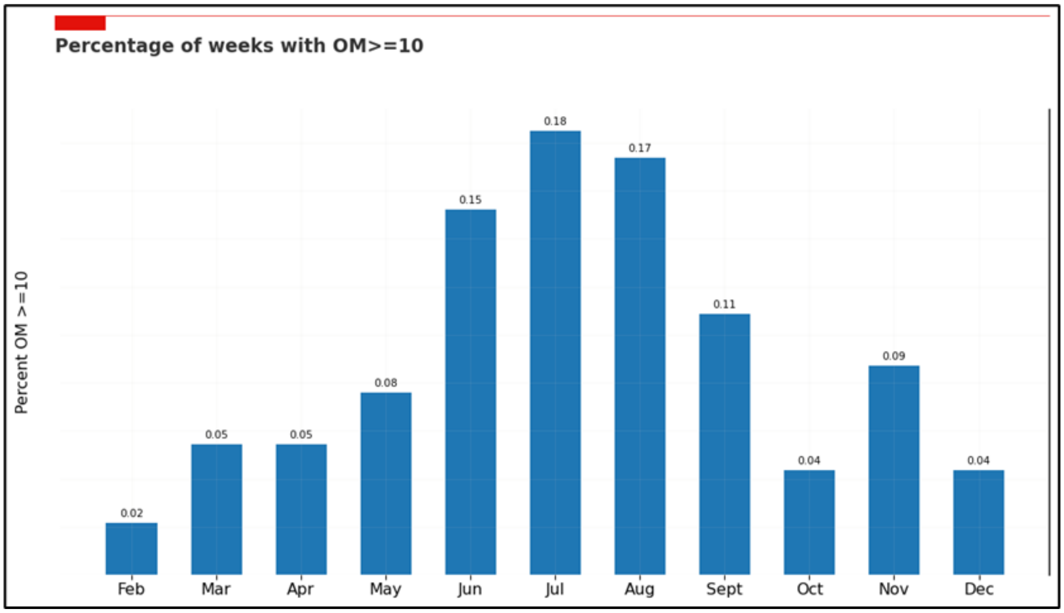

Figure 10 – **Percentage split for high OM occurrences (>= 10 OMs per week) (by month)**

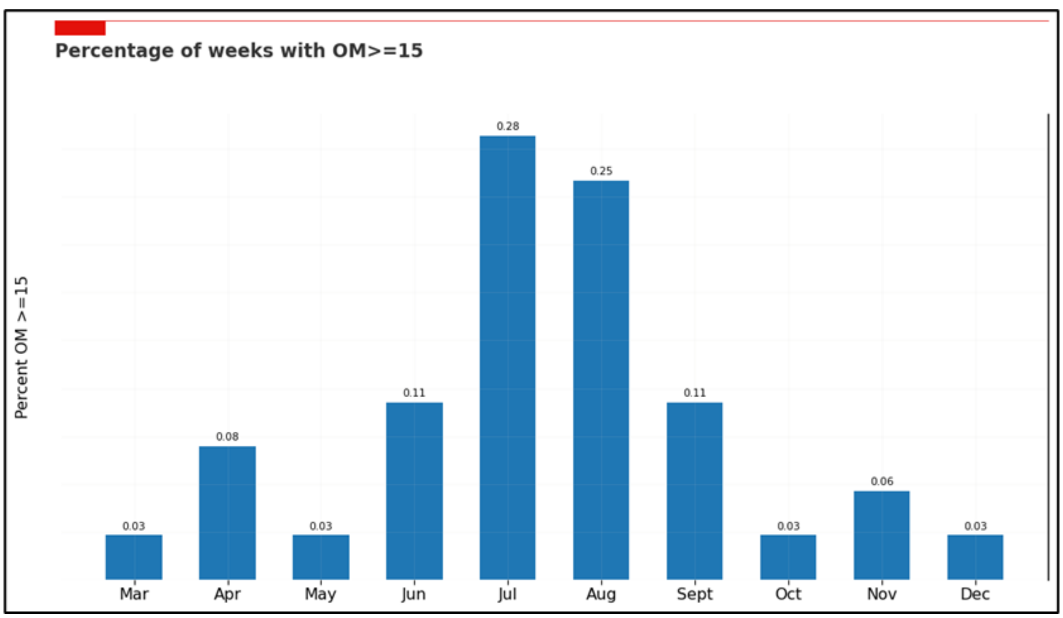

Figure 11 – **Percentage split for high OM occurrences (>= 15 OMs per week) (by month)**

4.2 Data Cleaning & Manipulation

Data Manipulation

Because the data for this problem is stored as individual events (Missions, Requests, Request-Declines), the first step in preparing the data for modeling was determining how to structure the dataset. Given the clear presence of time-series components, as highlighted, in Section 4.1, it was determined that the initial modeling approach would be to aggregate the data for time-series approaches. For this reason, the data was aggregated into daily, weekly and monthly time-series. However, the focus of the modeling effort was on the daily and weekly aggregations, as STARS SMEs indicated that monthly level estimates would not be granular enough for operational use.

Data Cleaning

Following the aggregation of the data into temporal datasets, further cleaning and manipulation of the data was required prior to arriving at datasets that could be used for modeling. The key steps are outlined below:

- **Missing data** – Because the data was aggregated by time increments (days, weeks) it was necessary to populate missing time increments (i.e., time increments when no events took place) with zeros.
- **Separation by Base** – Because the STARS helicopters operate out of different bases with different demands and asset constraints, it was necessary to model each based separately. Based on discussions with STARS SMEs, the decision was made to focus modeling on the Calgary and Edmonton bases, which are largest and have the most OM declines. For brevity, the modeling results shown in the sections below are for the Calgary base, but models were created for both key bases.
- **Changes in DW coding** – During collaboration sessions with STARS SMEs, it was uncovered that there was a change in data coding practices in 2017/2018, which results in a shift in the trend of missions being coded as OM. This can be seen clearly in Figure 1. These SMEs had lower confidence in the data prior to 2018, and for this reason most of the modeling was focused on data >= 2018.

4.3 Modeling

Due to the presence of trend, as well as weekly and monthly seasonality in the data, the primary focus of the modeling stage of the project was on time-series models. Machine learning regression models were also evaluated, but these approaches were unsuccessful and are briefly included for completeness.

Model Selection

Time-series models were selected based on their applicability to the dataset. The models planned for initial evaluation were Holt-Winters Exponential Smoothing (HW), Prophet (produced by Meta), and ARIMA.

However, it quickly became clear that ARIMA is not a good fit for short-term data (daily or weekly) that contains longer term seasonality. In this case there is a clear annual component to the data, which in ARIMA would requires high auto-regressive periods (~52 for weekly data, or ~365 for daily data). ARIMA models with these kinds of high periods are poor in terms of run-time and results. Therefore, the models that received the most focus were HW and Prophet.

Ultimately, the main models selected for utilization were HW and Prophet using the weekly timescale. In general, the daily timescale contained such a high degree of variability that it was difficult to achieve good modeling results. The reasons for this are discussed in Section 6.

Following the exploration of time-series approaches, some machine learning regression approaches using regression trees and regression forests were tested on daily data, but these approaches were unsuccessful. Again, the reasons for this are discussed in Section 6.

Model Training

Once the initial modeling approach was decided, models were trained on the cleaned datasets. Because the focus was using time-series approaches, the models were trained according to time-series best practices.

Hyper-parameters for the models were optimized using a grid-search approach combined with walk-forward 5-fold cross-validation, meaning that models were trained and evaluated across 5-segments of the training data that maintained the correct temporal structure of the data (i.e., models were not allowed to "cheat forward" by looking at future time periods, and models were evaluated across different periods of the data, leading to a more robust evaluation). The following hyper-parameter ranges were evaluated for each model:

Holt-Winters Exponential Smoothing

smoothing level (alpha)	0.01	0.1	0.25	0.5	0.75	-
smoothing seasonal	0.01	0.015	0.1	0.25	0.5	-
smoothing trend	0.01	0.015	0.05	0.1	0.25	0.5
box cox	0.25	0.5	0.75	0.85	-	-
seasonal periods	51	52	53	-	-	-

Prophet

change-point prior scale	**0.001**	**0.01**	**0.05**	**0.08**	**0.1**	**0.25**
seasonality prior scale	**0.01**	**0.1**	**1**	**5**	**10**	**12**

4.4 Model Evaluation

Models were evaluated using the Mean Absolute Error (MAE) metric. Because the goal from STARS' perspective is to estimate the number of missions that will be declined during a given period, it is best to use an error in terms of absolute OMs. Additionally, because the range of the data is relatively small (there are very few datapoints with >10 OM observations at a daily level), and because there are a significant number of zero observations (meaning percentage-based metrics are a poor fit), it was determined that MAE was the best overall metric for evaluating model fit in this context.

Following the hyper-tuning and cross-validation process, the optimized models were evaluated on a test set of the final 20% of the dataset, which corresponds approximately to the latest calendar year (2022).

5 RESULTS

The results for the best overall models, which were selected for utilization are outlined below.

5.1 Models selected for utilization

Holt-Winters Exponential Smoothing

HW is a well-known time-series algorithm that combines exponential smoothing with components for trend and seasonality. Note that only additive trend and seasonality components are compatible with this dataset due to the presence of zeros (as opposed to multiplicative). The HW version used for this project was from the Statsmodels Time-Series-Analysis library for Python.

The parameter set that achieved the best fit for the HW algorithm during the grid-search and cross-validation processes, outlined above, were:

smoothing level (alpha)	0.01
smoothing seasonal	0.15
smoothing trend	0.015
box cox	0.75
seasonal periods	53

The HW algorithm was then evaluated on the test set using these hyper-parameters. The performance on the test set is illustrated below, in Figure 13. The MAE for this model when evaluated against the test set was 5.12, which is reasonable when evaluated at a weekly level.

Prophet

Prophet is an additive time-series regression model created by Facebook/Meta that combines a "growth" term or trend with multiple seasonality effects (annual, monthly, weekly), making it a good fit for this dataset [5,6]. The prophet library contains functionality for extracting the different components for each model, as shown below, in Figure 12, which is useful when making operational decisions.

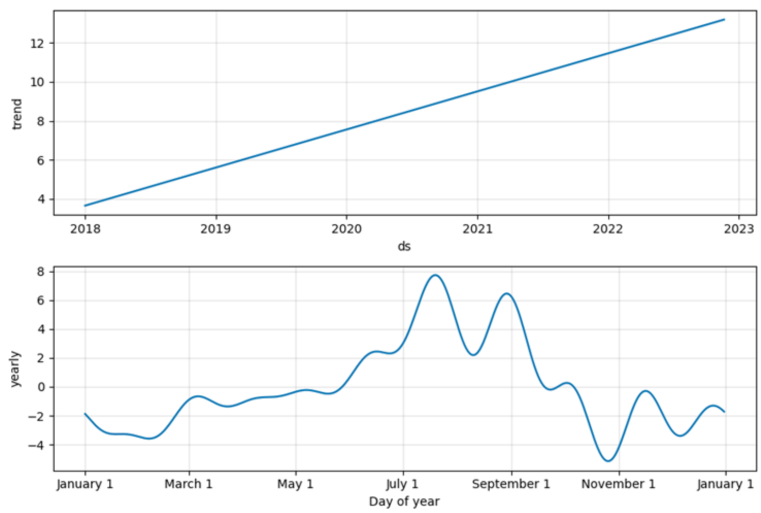

Figure 12 – **Trend and seasonality from best-fit prophet model**

The hyper-parameters for the prophet model were evaluated using the grid-search and cross-validation approaches outlined above, and resulted in the following set of best-fit parameters:

change-point prior scale	0.01
seasonality prior scale	10

The prophet algorithm was then evaluated on the test set using these hyper-parameters. The performance on the test set is illustrated below, in Figure 14. The MAE for this model when evaluated against the test set is 5.35, which is slightly higher than the best HW model, but reasonable at a weekly level.

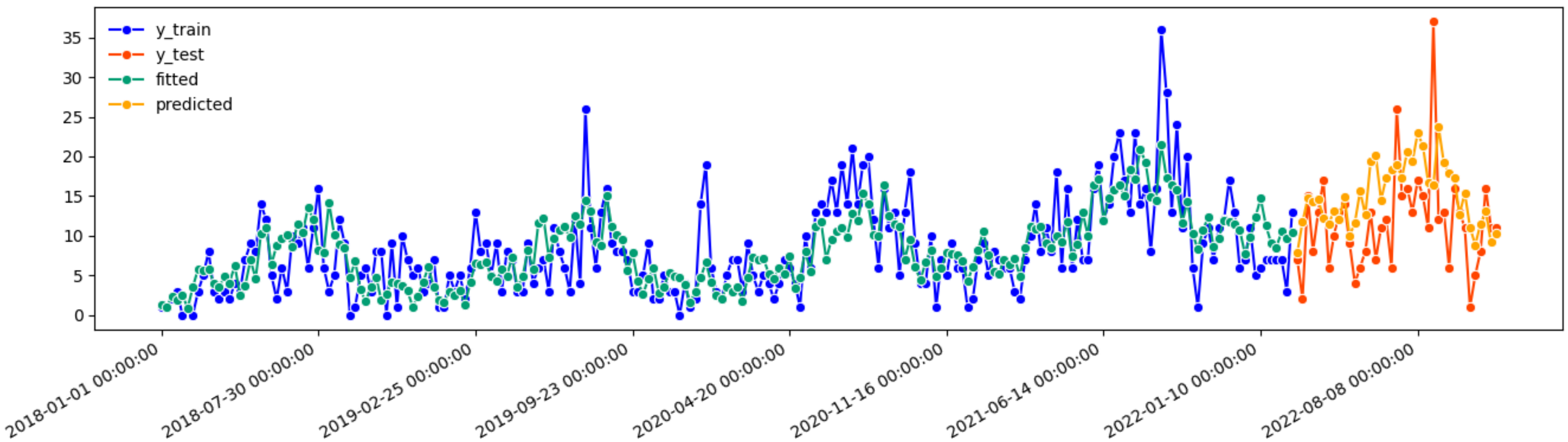

Figure 13 – **Best fit model for Holt-Winters Exponential Smoothing**

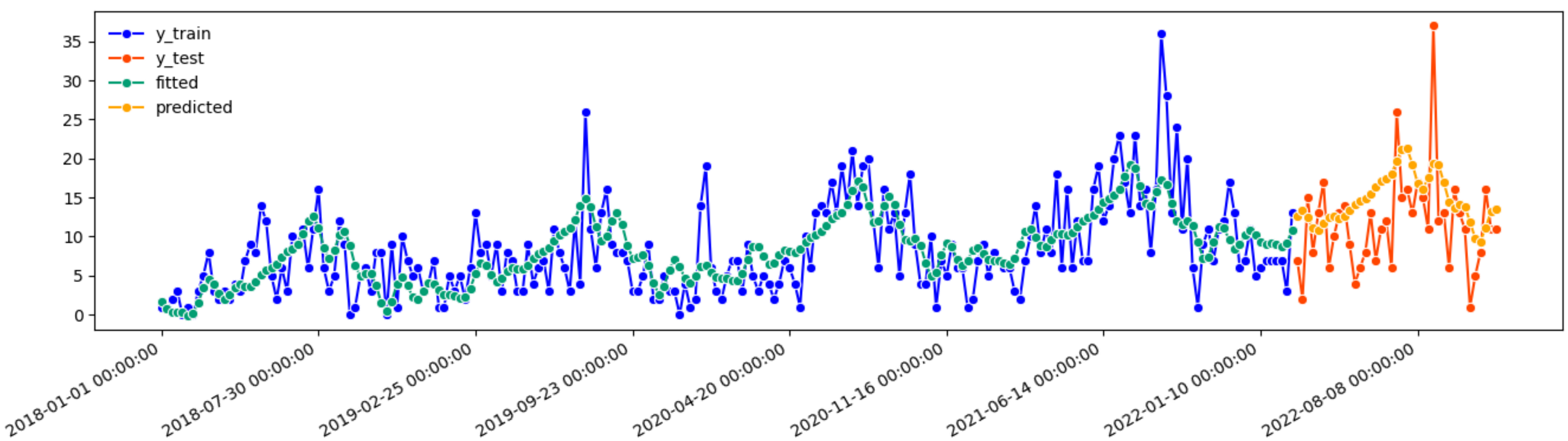

Figure 14 – **Best fit model for the Prophet algorithm**

5.2 Unsuccessful models

Several algorithms were evaluated in an attempt to create a useable model on the dataset for daily frequency data. These included the top algorithms for the weekly data (HW, prophet) but also some machine learning regression models (regression trees, regression forest). None of these models achieved results that were robust enough for utilization. I believe this is driven by high variability within the daily frequency data that is driven by exogenous factors that are not captured in the dataset, as explained in further detail in Section 6.

An example of a regression tree model, shown in Figure 15, that was trained using the available time-series parameters (month, day-of-week, holiday status) results in a model that provides seasonal averages without capturing the day-to-day variability of the dataset. This averaging or smoothing effect was seen in all the algorithms (prophet, regression trees, regression forests). This is because these temporal variables (season, day, holiday) do not capture the full variability of the data, which is further discussed in Section 6.

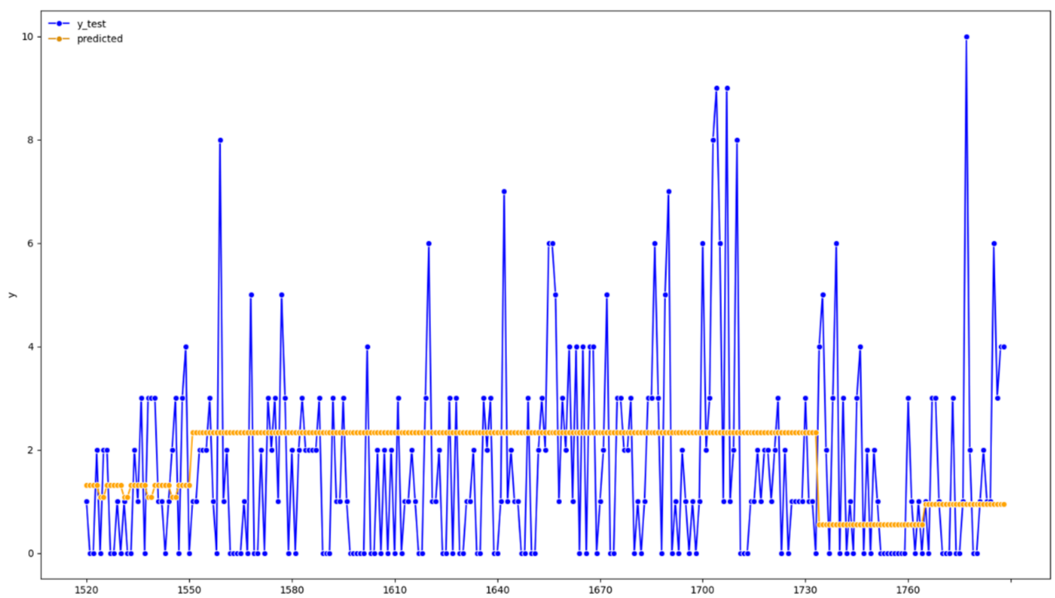

Figure 15 – **Regression Tree Results vs the Test Set (daily)**

6 ANALYSIS AND OUTCOMES

6.1 Analysis

This turned out to be a challenging data science problem, largely due to the high variability in the event occurrence/frequency within the dataset. This variability could be smoothed at slightly longer timescales – which is why the focus for modeling was at the weekly timescale – but could not be matched with high confidence at the daily timescale.

I believe the reason that this variability could not be matched by the algorithms is because the drivers of this variability are exogenous to the current dataset. While some of the key components of the data, such as weekly and annual trends could be quantified, I believed that there are other important factors that are not currently captured.

I believe the most significant factor that is not captured within the existing dataset is weather. Weather events are very significant drivers of human behavior and traffic events in the Alberta region, particularly snow/ice storms, which are a common occurrence. I believe that weather events are a key driver behind daily-variability and the presence of "outlier" days in the dataset. One future avenue of investigation might be incorporate weather into the model. However, in practice this is difficult to achieve because weather, itself, is highly unpredictable. Showing the impact of weather on past helicopter requests would be possible using historical weather data but using these models for planning would be unfeasible, except in the very short-term, because weather events cannot be predicted accurately beyond short timescales (daily or even hourly). For this reason, weather was not a key focus to this initial round of investigation.

Another potentially important exogenous factor is covid effects. These are difficult to account for in the modeling, because there is only a small amount of post-covid data (covid restrictions in the Alberta-region began to be removed in the spring/summer of 2022). Additionally, covid policy in this region was highly variable during the covid period, with some periods of severe-restriction and other periods where the restrictions were milder.

Overall, weekly models, and explorations at the daily level, were enough to create models that are functional for medium term planning at STARS (annual and seasonal planning cycles) and were able to quantify key trends/components in the data that will assist with this planning, notably that key high demand points occur on weekends and during the summer months. This understanding alone will allow STARS to capture a large percentage of OM declines.

Although, these trends appear relatively simple in retrospect, they had not previously been quantified or understood, beyond the anecdotal level, at STARS.

6.2 Outcomes

A main outcome of this project was the development of time-series models that can be used to estimate periods were OM declines are expected to be high. STARS SMEs can use these models, and the components they have uncovered to plan for periods when additional resource are most likely to be utilized. One of the next steps for STARS SMEs is to determine the required threshold for expected OM declines required to justify the scheduling of additional assets.

Additionally, these models can be used to add rigor to discussions with STARS funders (STARS is a not-for-profit that is partly funded by government) to explain and justify the need for additional assets.

The first operational outcome from this project is development of an experimental evaluation or "hypothesis-test" for the upcoming Labour Day period. The time-series models and dataset explorations have highlighted this as a period where OM declines are expected to be high. STARS is planning a project to track the number of missions that an additional asset (helicopter) would capture during this period.

Because this period is approaching too quickly for operational planning, and because there are differing views on the correct OM count among STARS SMEs (as outlined in Section 3.3) this test will provide a low-cost way to validate the modeling and underlying dataset. Based on the results, STARS will have the confidence to schedule additional assets for predicted high periods in the future.

Date	OM Count	holidays	weekday	month
2021-09-03	[redacted]	0	4	9
2021-09-04	[redacted]	0	5	9
2021-09-05	[redacted]	0	6	9
2021-09-08	[redacted]	0	2	9
2021-09-09	[redacted]	0	3	9

Date	OM Count	holidays	weekday	month
2022-09-01	[redacted]	0	3	9
2022-09-04	[redacted]	0	6	9
2022-09-07	[redacted]	0	2	9

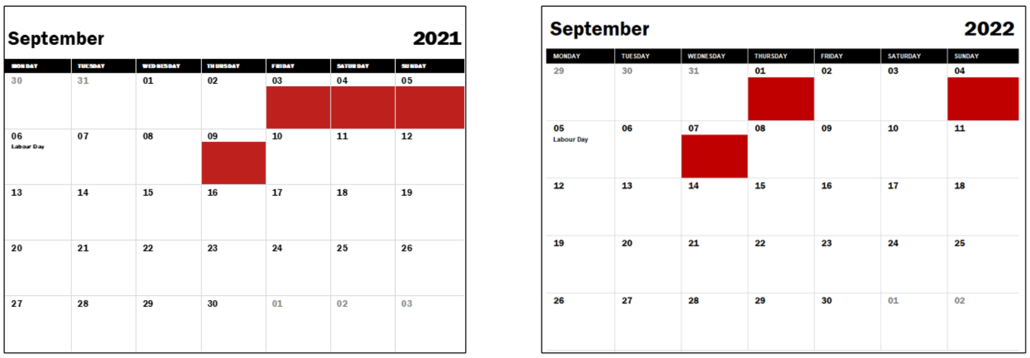

Figure 16 – **High OM Days during Labour Day Week (2021 & 2022) (Data Observations Redacted)**

6.3 Project Summary

Overall, this turned out to be a challenging but rewarding project. Although the dataset was a relatively simple time-series, it was challenging to model due to a high level of inherent variability, multiple seasonality components, and due to the number of different avenues of approach. My work on this project strengthened my understanding of time-series approaches, and the strengths and weaknesses of these algorithms.

However, the most challenging aspect of this project may have come during the data collection and consultation phases. I had not anticipated how challenging and time consuming it would be to understand the data warehouse and arrive at a dataset for modeling as someone who is not a subject-matter-expert within STARS. Additionally, I had not considered that there would be disagreement on the correct dataset among STARS SMEs, or the possibility that there could be multiple ideas for the correct dataset. This put the schedule of this project at risk and required collaboration and the making of compromises to maintain progress. In the end, I gained a lot of learnings that will be useful in my career as a data scientist.

The most personally rewarding aspect of this project is the impact this work might have in the future. One of the reasons that I chose to work with STARS is due to the interesting and high-impact nature of their work. If, in the future, STARS can schedule additional resources and fly additional missions because of this investigation then it will have had an impact on peoples' lives.

6.4 Next Steps & Ideas for Further Investigation

Next Steps

- Perform an evaluation test during the Labour Day 2023 period
- Utilize models and trends for future scenario planning
    - Expected value thresholds
    - Funding requirements

Ideas for Further Investigation

- Incorporation of exogenous variables:
    - Weather
    - Traffic
    - Demographic Trends
- Modeling of bulk Requests, as opposed to OM declines
    - Apply mission acceptance rates

7 REFERENCES

1. STARS Website. Website. Accessed July 5, 2023. Home - STARS
2. Midterm Progress Report. Jesse Keith. Georgia Institute of Technology OMSA Practicum Project.
3. STARS Business Terms Glossary. STARSgate Home Intranet. Accessed June 9, 2023.
4. STARS Operations Mission/Request Power BI dashboard. STARS Intranet. Accessed June 25, 2023.
5. Prophet Algorithm. Amazon Web Service Documentation. Website. Accessed July 2, 2023. Prophet Algorithm - Amazon Forecast
6. Time Series Analysis with Facebook Prophet: How it works and How to use it. Mitchell Krieger. Towards Data Science. Website. Accessed July 2, 2023. Time Series Analysis with Facebook Prophet: How it works and How to use it | by Mitchell Krieger | Towards Data Science