A picture containing screenshot, snowboarding, sky, skiing

Description automatically generated

McKenzie Gowing

Student #: 500730860

CIND820

Professor: Ceni Babaglu

July 17th, 2023

[Git Hub Repository](https://github.com/KZiEG)

Trend Analysis of Capacity Rates for Toronto Homeless Shelters

Table of Contents

[Introduction 3](#_Toc140524781)

[Data Sources 4](#_Toc140524782)

[Literature review: 5](#_Toc140524783)

[Methodology 10](#_Toc140524784)

[Overall Trend 11](#_Toc140524785)

[Proportion of Shelters at or Over Capacity 20](#_Toc140524786)

[Refugee Programs 26](#_Toc140524787)

[Limitations and Next Steps 35](#_Toc140524788)

[Conclusion 36](#_Toc140524789)

[Literature Assessment 36](#_Toc140524790)

[Q & A 36](#_Toc140524791)

[References 37](#_Toc140524792)

## Introduction

Having adequate shelter is a foundational component to a person’s health, however even in areas like Toronto too many people experience homelessness on a consistent basis. Homelessness can take on many forms but it is most often understood as a situation where a individual or group of individuals are living without an stable, safe, and permanent housing accommodation (Uppal, 2022). The homelessness situation in Canada has evolved to the point where homelessness affects more of diverse group of individuals. Now more women, families and youth are experiencing homelessness, whereas historically it was a phenomenon that mostly affected single, older males (Gaetz, 2010). The objective of this research paper is to describe the trend and patterns of homeless shelter programs capacity rates and proportion of shelters hitting maximum capacity. Homelessness is a complex issue and the city of Toronto has taken multiple steps to support shelter programs however understanding the trend in capacity rates and the relationship between shelter use and independent environmental and socio-economic variables will be necessary in making data driven decisions to effectively assist Toronto’s shelter programs.

Due to the different features, intervals and level of detail of publicly available datasets, this project will make use of time-series and cross-sectional statistical methods. Where appropriate, insights will be presented and analyzed at the overall level for the city of Toronto and at other lower levels of granularity. Answering the following research questions will provide direction for this project:

* What does the trend look like for capacity rates for Toronto shelters?
  + Are there seasonal patterns?
  + What has been the movement of the trend (upward or downward)?
  + What is the relationship between daily capacity rates and prior capacity rates?
  + How often do shelters go over capacity?
* What are the characteristics of the time series for Refugee shelter programs?
  + Is there seasonality?
  + How often do they go over capacity?
  + Is there a difference in capacity rates between refugee and other shelter programs?
* Based on socio-economic factors of the city what areas have the highest capacity rates?
  + What neighborhoods have the highest capacity rates?
  + Is there a difference in capacity rates between neighbourhood improvement areas and those that are not improvement areas?
* Is there a difference in capacity rates between neighborhoods based on socio-economic characteristics (median income, education, age, home ownership)?

## Data Sources

This project will use three main data sources. The city of Toronto’s *Daily Shelter & Overnight Service Usage* dataset will be the central focus and used to analyze the dependent variables, shelter capacity rates and total intakes. This dataset reports information at the shelter program level at daily intervals for shelters located within the city of Toronto going back to 2021. This data source contains additional information related to the location of the shelter programs, the name of each program, the program model, sector, and capacity type ([City of Toronto](https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/shelter-census/)). The overall quality of the data is very good. The integer variables related to bed & room occupancy rates (the last 12 columns) have a high rate of missing values because shelter capacity is based on available rooms or beds and these are exclusive categories but both types are captured in the dataset. A calculated field will need to be created to consolidate the capacity. The daily shelter & overnight usage data source is updated daily. At the time of the analysis the information is value up until July 13 2023.

The following table provides a data dictionary for the fields used for this analysis. Neighbourhood improvement areas we identified using the Neighbourhood Improvement Area found on the city of [Toronto’s open data portal.](https://open.toronto.ca/dataset/neighbourhood-improvement-areas/)

|  |  |  |
| --- | --- | --- |
| Variable\_Name | Variable\_Type | Description |
| OCCUPANCY\_DATE | Date | Date of shelter program operation |
| LOCATION\_ADDRESS | character | Address of the Shelter |
| LOCATION\_POSTAL\_CODE | character | Shelter Postal Code |
| LOCATION\_CITY | character | Shelter City |
| PROGRAM\_ID | integer | Shelter Program ID. |
| PROGRAM\_NAME | character | Program Name |
| SERVICE\_USER\_COUNT | integer | Total Shelter Users |
| OVER\_OCCUPIED | character | Indicates if a program hit capacity. 1 = Hit Capacity 0 = Did Not Hit Capacity |
| Improv\_Status\_Area | Character | Identifies if a program is operating in a neighborhood improvement area or not. |
| OCCUPANCY\_RATE | Character | Shelter Occupancy Rate. |

## Literature review:

Jenkinson et al. (2021), “Nowhere to go: exploring the social and economic influences on discharging people experiencing homelessness to appropriate destinations in Toronto, Canada”

Jenkinson et al. (2021) address the issue of hospitals in Toronto discharging patients experiencing homelessness to either the streets or shelters. Semi-structured interviews were conducted on hospital and shelter workers. The researchers write about detailed gaps in the hospital system when it comes to discharging homeless people. Often times hospitals ignore shelter admission policy and simply leave patients with the shelter system because the hospital lacks resources however Toronto’s shelters regularly operate at over capacity and shelter workers do not have the resources to care for people experiencing homelessness and other physical or mental problems. They note that an aging population that is more reliant on the health care system and the lack of affordable housing strains financial and physical resources in the health care system that could be used to help unhoused hospital patients. This article is useful to this research paper as it provides Toronto specific context to some of the systematic challenges shelters face while they are operating at over capacity and factors behind homelessness.

Gaetz, S. (2010) The struggle to end homelessness in Canada: how we created the crisis, and how we can end it.

Gaetz, S (2010) explains the evolution of homelessness and the structural conditions that create homelessness in Canada. The article provides details about how the main cause to homelessness is a lack of affordable housing. People have always experienced homelessness in Canada but it became a social problem after the 1980’s because the government cut spending on social and hosing programs and a economic shift to economy that took on characteristics of trade liberalization and deindustrialization. This accelerated financial inequalities within the Canadian population. Even though nationally there has been an overall increase in wealth during the past decades, these financial gains have mostly been felt by the upper class while middle income earnings have remained stagnate and lower income earnings have actually declined. The author also notes how single parent women, visible minorities and new Canadians are also disproportionately affected by homelessness. This article is relevant for this research paper as it identifies economic factors behind homelessness and groups that are effected by homelessness. This paper will be particularly useful when working with the census data.

Clemens et al. (2022), Impact of Ontario’s Harmonized Heat Warning and Information System on emergency department visits for heat-related illness in Ontario, Canada: a population-based time series analysis

Clemens et al. conducted a time-series analysis to examine visits to hospital for heat-related illness. The population of interest was Ontario residents living in urban centers. The researchers used a Autoregressive integrated moving average (ARIMA) model to examine if there was a change in admissions to hospitals with the implementation of Ontario’s Heat Warning Information System. Homeless individuals were identified as a sub-population in the study which provided relevant information for this specific research project. They found that individuals with a recent history of homelessness had higher rates of visits for heat-related illness and admission rates increasing overtime was positively associated with temperature and peaked in June and July. This paper is relevant to my research topic as it provides knowledge around the methodology of time-series analysis in relation to temperature and people experiencing homelessness. It also provides context around the relationship between weather patterns and how unhoused individuals are affected. A important take away from the paper is that the researchers limited the study period to May-September to exclude the impact of colder temperatures, I anticipate I will need to take a similar approach and extend it when looking at the relationship between capacity rates and colder temperatures.

Zhang et al. (2019), Cold Weather Conditions and Risk of Hypothermia Among People Experiencing Homelessness: Implications for Prevention Strategies.

Zhang et al. (2019) conducted a time-series analysis to analyze the relationship between weather conditions and the risk of hypothermia among people experiencing homelessness in Toronto. A part of their methodology I found interesting is that they used a time-stratified case-crossover design where if a hypothermic event occurred it was considered a case and up-to-four cases falling on the same weekday of that month were chosen as controls. The datasets the researchers used were a hospital admission data and weather data from Toronto Pearson Airport weather station. The research was limited to events occurring during the colder months of the year (November 15th to March 31st). A conditional logistic regression model was used to examine the relationship between minimum temperature and precipitation and hospital admissions for hypothermia for homeless individuals. The researchers found that most of the hypothermic events occurred when the temperature was warmer that -15 degrees. This study is relevant to my research as it provides information around time-series analysis and regression modeling when understand cold weather patterns and its effects on the Toronto homeless population. This research provides insights that will be useful for me when understanding the seasonality of winter weather patterns.

Jadidzadeh, A., & Kneebone, R. D. (2015). Shelter from the Storm: Weather-Induced Patterns in the Use of Emergency Shelters.

Jadidzadeh, A., & Kneebone, R. D. (2015) used a data source that captured daily overnight stays in Calgary homeless shelters and daily Calgary weather reports to understand weather-induced shelter use from January 2008 to May 2014. The researchers used a multiple linear regression model to understand the relationship between temperature and precipitation (independent variables) and total overnight shelter stays. They found that there is a highly significant relationship between precipitation and colder temperatures on overnight shelter use. When the minimum overnight temperature is between -0C and -10 C and if there is precipitation, it causes a increase in overnight shelter use. This is a very useful and relevant paper to my research as the researchers went into great detail in their methodology. One thing I found requires further investigation because it is applicability to my research project is the use of dummy variables in the regression models for temperature.

City of Toronto’s Social Policy Analysis and Research (March 2014). TSNS 2020

Neighbourhood Equity Index. https://www.toronto.ca/legdocs/mmis/2014/cd/bgrd/backgroundfile-67350.pdf

As part of the *Toronto Strong Neighbourhood Strategy* the city of Toronto have recently updated it’s list of neighbourhoods identified as *Neighbourhood Improvement Areas.* Neighbourhood Improvement Areas are identified based on a equity index which assess inequity between Toronto neighbourhoods. Ares of focus for identifying inequity include lower income, marginalization in terms of residential instability & ethnic concentration, and unemployment. Toronto current has identified 31 neighbourhoods as improvement areas. Since residential instability is a major component to identifying a neighbourhood improvement area, this information will be incorporated into the analysis with the aim of understanding shelter capacity rates.

Uppal,S (2022, March 14th). A portrait of Canadians who have been homeless.

Uppal examined several socio-economic and demographic characteristics of Canadians who are currently head of households but have previously experienced homelessness. The researcher used the 2018 Canadian housing survey to identify the population of interest. They found that First Nations people living outside of a reserve were more likely to experience homelessness than non-indigenous. I am not going to dive into all the findings but overall they found individuals who faced past financial difficulties, had lower education, were sexual minorities, female visible minorities, or females between the ages of 15-39 were more likely to have experienced homelessness. The researcher presented their findings as a cross-tab and reported if results were significantly different between groups. This paper is relevant when it comes to understanding socio-economic factors behind homelessness. This article will be used to assess appropriate variables to pull from the census profile dataset.

Freeman, J. (July 14th, 2023). 'I feel like I'm not welcome': Refugees describe 'crisis' on

Toronto's streets.

This article details the relatively recent events of Toronto’s shelters inability to house refugees. Toronto is increasingly seeing more refugees arrive in the city and end up sleeping on the streets. Shelters located in Toronto have experienced a 440% Increase in refugees since 2021. The article notes how the city is short almost $70 million dollars in founding to house refugees. This article is useful for the analysis as it provides more context and experiences for those most at risk of homelessness. Part of my analysis will look to analyze the trend in capacity rates amongst refugee programs specifically.

## Methodology

Each section of the analysis will analyze capacity rates and the proportion of shelters operating at or over capacity. The analysis first use visuals to display the characteristics of the trend. Mann-Kenndall Trend Test and Sen’s slope will be statistical test used to understand the trend. Then AFT and KPSS will be used to make conclusions about a time-series stationarity. The forecast models used will be Naïve, Random Walk with Drift, Simple Exponential Smoothing and ARIMA models. The forecast will be trained on a dataset of all the records except the last 30 and the test set will be made of all the last 30 days of the time series. Forecast will be compared using MAPE, Mean Absolute Percentage Error.

## A graph of a number of times Description automatically generatedA graph with numbers and lines Description automatically generatedOverall Trend

Figure 3: Capacity Rate - ACF plot 1

Figure 2: Capacity Rate: QQplot

Figure 1: Average Daily Shelter Capacity Rates

| Metric | Daily Capacity Rate |
| --- | --- |
| Mean | 95.65 |
|  |  |
| Median | 96.37 |
| Standard Deviation | 2.43 |
| Minimum | 89.5 |
| Maximum | 99.04 |
| Range | 89.5 - 99.04 |

Table 1: Average Daily Shelter Capacity Summary Statistics

Referring to figure 1: Average Daily Shelter Capacity Rates, there appears to be a increasing trend from the lowest levels in daily average capacity rates occurring on July 10 2021 to the highest daily capacity rates on September 24th 2022. This insight here is that there was a gradual increase in shelter capacity rates however after the peak it looks it is staying around that 97.5% capacity range. This consistency after the peak in daily capacity rates suggests that Toronto shelters are becoming increasingly over burden and staying over burden.

The QQplot highlights the distributional characteristic of the average daily shelter capacity rate trend. The plot shows that the distribution is left skewed. This means that statistical methods being used will need to be non-parametric. Furthermore when analyzing the ACF plot it shows all lagged values are significant however they are slowly tapering off, with the first lag value being the most significant. This suggest that there is not seasonality and that current values are highly correlated with past value, with the highest correlation being with lag one. This means that each observation is highly correlated with the previous observation one day prior

Based on the appearance of the time series and the results from the ACF plot, the trend in the Average Daily Capacity Rates for Toronto shelters does not appear to show characteristics of white noise, were there is a stationary mean or variance. Instead we see characteristics of a random walk where the current value is highly related to the value 1 day prior. These are necessary characteristics to understand as the forecsting models will need to take this into account.

We can confirm the trend exhibits non-stationary by running a KPSS test for stationary and Augmented Dickey Fueller Test (ADF). The ADF test is a statistical test that test for non-stationarity in a time series while he KPSS test is similar except it test for stationarity. Both methods are appropriate for non-parametric distributions ([Analytics Vidhya](https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/#:~:text=ADF%20test%20is%20conducted%20with,series%20has%20no%20unit%20root.)).. The null and the alternate hypothesis test is listed below for both test. Since the p-value of the ADF test is larger than alpha level of 0.05 we fail to reject the null hypothesis concluding that there is significant evidence to suggest that the trend is non-stationary. The KPSS test also supports this conclusion, as the results of the test is highly significant which allows for the rejection of the null hypothesis and the conclusion that there is enough evidence to suggest that the time series is non-stationary.

ADF 1: Daily Average Capacity Rates

Null Hypothesis: The time series has a unit root (is non-stationary)

Alternative Hypothesis: The time series does not have a unit root (is stationary)

KPSS 1: Daily Average Capacity Rates

Null Hypothesis: The time series does not have a unit root (stationary)

Alternative Hypothesis: The timer series does have a unit root (non-stationary)

A graph of a sound wave

Description automatically generated

Differenced plot: Lag 1

Once the data is differenced at lag once the trend now shows stationarity. Interestingly based on the plot of the differenced data it looks like there is more variance to the left of the visual compared to the right again suggesting that after the peak there is more stationarity.

We can assess the trend using the Mann-Kendall trend test and Sen’s slope. Both are non-parametric test which would be appropriate based on the distribution of the trend. Mann-Kendall Trend Test for a monotonic trend in the time series data while Sen’s slope estimate the magnitude of the trend. Refer to the null and alternative hypothesis for both tests below (Pohlert, 2023). Based on the results of the Mann-Kendall trend test there is a significant small positive trend/correlation (tau) of around 0.6777, which allows for us to reject the null hypothesis and conclude that there is a significant evidence to suggest that there is a increasing trend in the time series data. The Sen’s slope further supports this statement as the test found significant evidence to support there is a trend and based on the sen’s slope estimate a positive slope of 0.0074.

Null Hypothesis: No monotonic trend

Alternative Hypothesis: There is a trend, decreasing or increasing, in the time series.



Null Hypothesis: There is no slope or no significant trend in the time series.

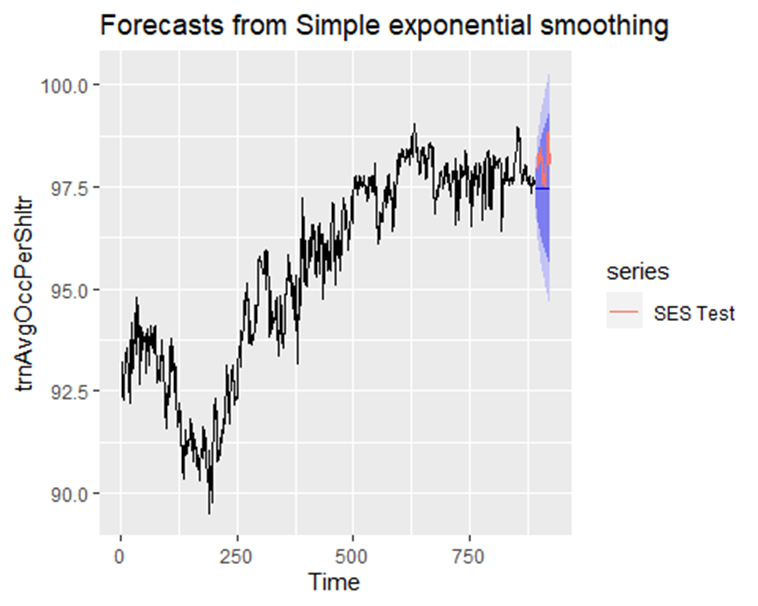
Alternative Hypothesis: There is a slope, indicating a significant trend.

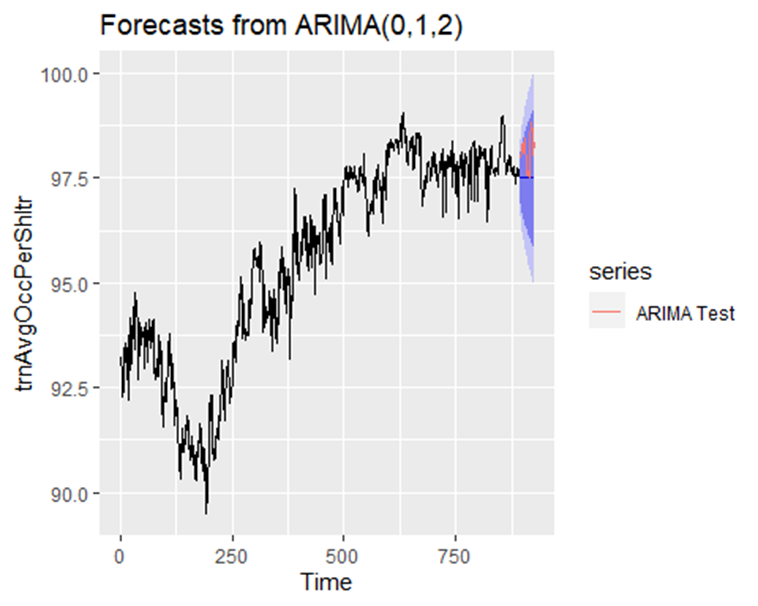
**Forecasting**

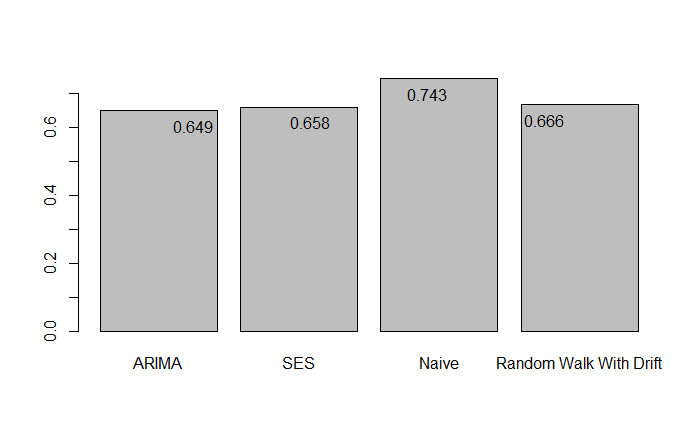
A graph of a graph showing a graph

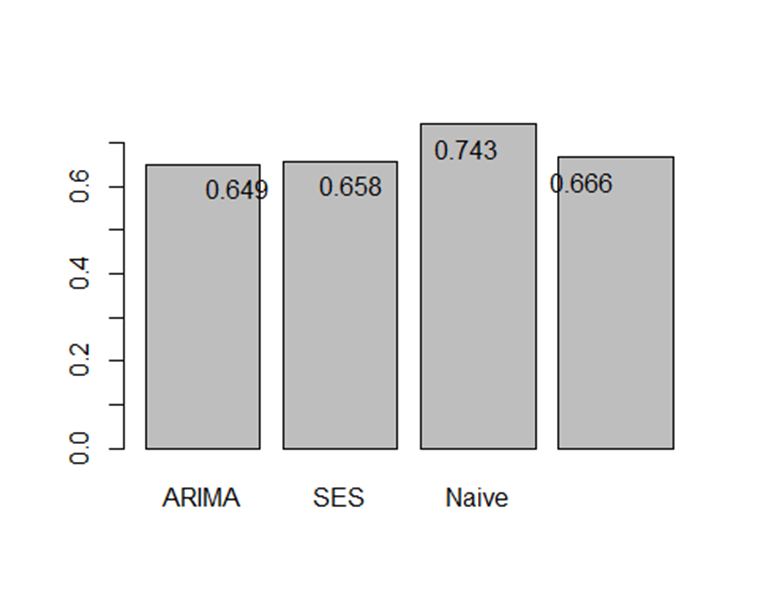
Description automatically generatedA graph showing a graph of a method

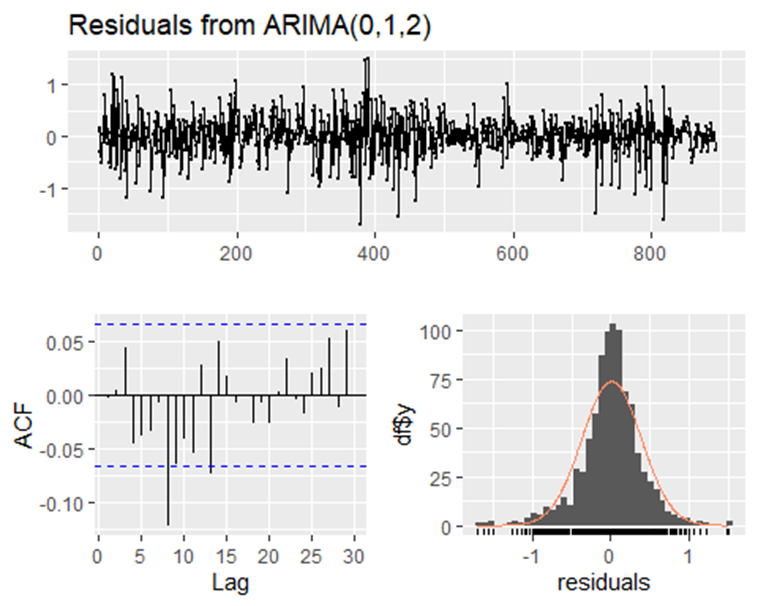
Description automatically generated The following are the results of the 4 forecasts that were run. The training dataset was all records minus the last thirty days and the test set was based on the results from the last 30 days. For the purpose of the analysis the same four forecasting methods will be applied for every forecasting section. These methods were selected based on their common usage and also the fact that the time series chows characteristics of non-stationarity. The naïve method was chosen because it is a very basic and common method for forecasing where the most recent observation is used to forecast the predicted values. The random walk with drift method is similar to the naïve method but it extends to the previous value, meaning that the forecast allows for a increasing or decreasing trend based on the average change from the previous to current observation. The Simple Expenential Smoothing (SES) forecasting methods takes into consideration the average of whole trend when predicting values and applies more weight to the most recent values. The ARIMA forecast models are more dynamic models that allow for differencing of a non-stationary time series and combines both autoregression of previous values and a moving average (Hyndman, R.J., & Athanasopoulos, G. ,2018).







Model Performance 1



When comparing the MAPE values we can conclude that the ARIMA model has the best fit and most accurate. The autoregressive order of 0 indicates current value of the time series is not dependent on its own previous values when the time series is differenced at lag 1 and a moving average of 2. The moving average of order 2 means that the current value of the time series depends on the errors of the two previous days. The MAPE stands for Mean absolute percentage error and is calculated against the test set. The forecasted error are on the same scale as the time series (%), so the ARIMA model with a MAPE of 0.649 indicated a Mean absolute percentage error of 0.649 % on average for daily average shelter capacity rates. All forecast were relatively close however. When looking at the plots for the residuals of the ARIMA model we see that difference between the observation and its predictied value are normally distributed and do not show autocorrelation, this are essential assumption for model assessment.

**Trend Segmentation**

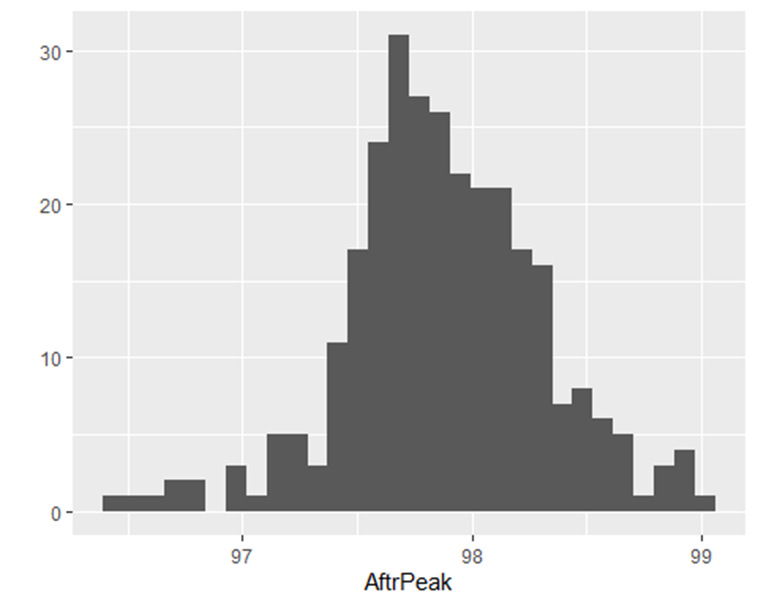
A graph with numbers and lines

Description automatically generated

As mentioned previously, after the peak it looks like the trend takes on different characteristics. It starts to look like it takes on more qualities of a white noise model. We can confirm this by running another KPSS test and ADF test on the time series after the peak. Interestingly when we just focus our analysis on the period after the peak the results of the ADF indicate that there is significant evidence to reject the null hypothesis, suggesting that the trend is indeed stationary after the peak. The results of the KPSS support this since the results of the test indicate we fail to reject the null hypothesis, concluding that time series after the peak is indeed stationary. The histogram further supports this as it indicated that after the peak the distribution appears to be normal and taking on more characteristics of white noise.







Now if we run the Mann-Kenndall test on the time series after the peak, it shows that there is not significant evidence to reject the null hypothesis, suggesting that there is no monotonic trend. This makes forecasting more straightforward as the predicted values would be based on the mean and the more consistent variance of the time series after the peak and more towards the end of the time series.



## A graph with a line drawn on it Description automatically generatedA graph of a number of sheltering Description automatically generatedProportion of Shelters at or Over Capacity

To get a better understanding of how shelters are dealing with the increase in capacity rates looking the proportionof shelters operating or over capacity is crucial. Similarly to the Average Daily Capacity Rate, there appears to be a increasing growth trend in the proportion of shelters hitting capacity. Interestingly here we see a wider range between the dates of minimum and maximum levels. Actually the level were the proportion of shelters operating at or over capacity was the highest occurred relatively recently. The same forecast methods will be applied so it will be interesting to see how the forecast predict into the future without having the peak in the training dataset.

The QQplot suggest that the time series is rightly skewed. This means that statistical methods being used will need to be non-parametric. Again the ACF plot show all lagged values are significant and slowly tapper off while lag 1 is the most significant lag. This indicates that there is no seasonality in the time series when measuring of daily proportion of shelters at or over capacity.



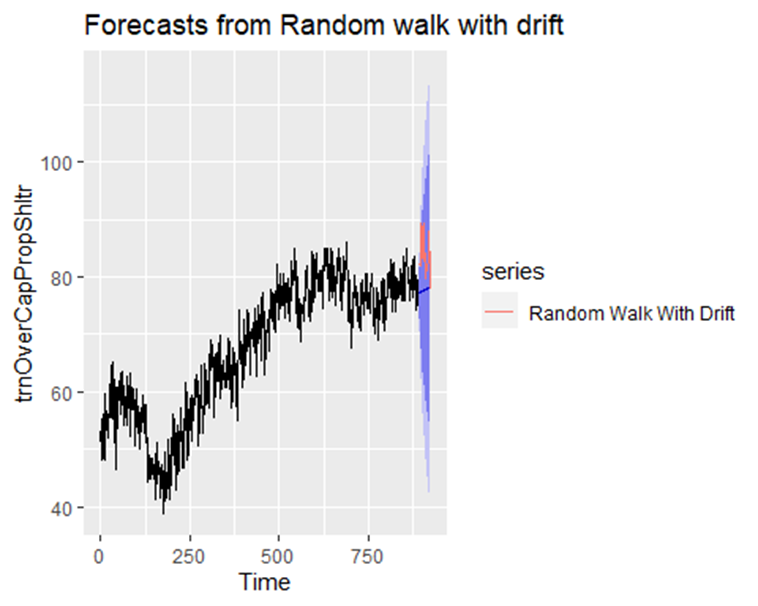


Again, like the daily average capacity rate, based on ADF and KPSS test and can investigate if the trend is stationary. The results from the ADF test indicate that there is not significant evidence to suggest that the trend is stationary. We fail to reject the null hypothesis and conclude the trend is non-stationary. The results from the KPSS test also support these findings.

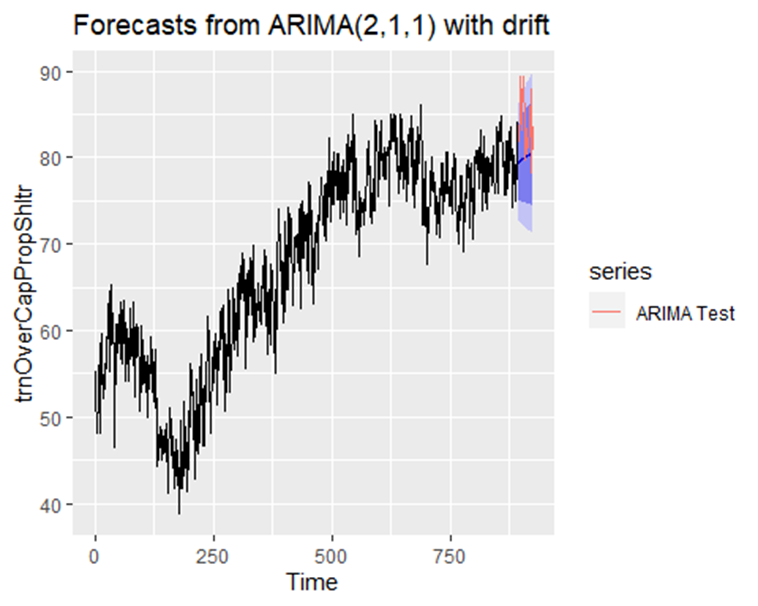
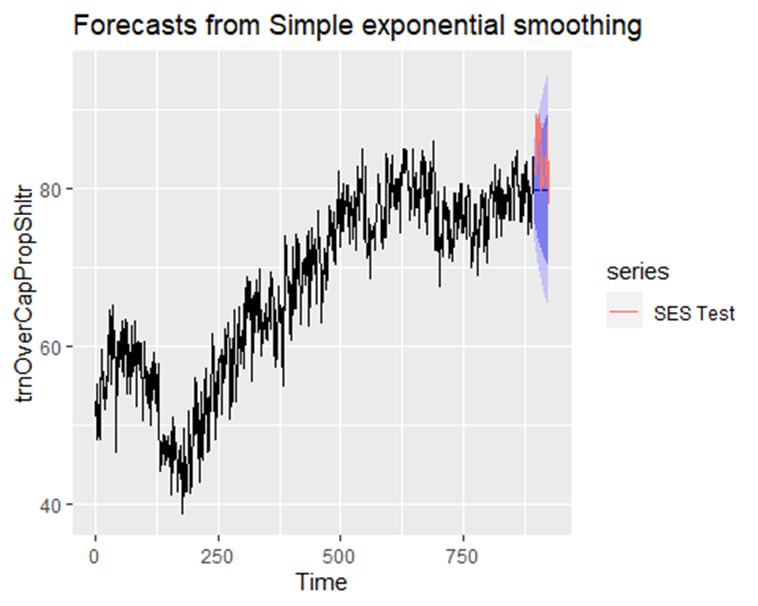
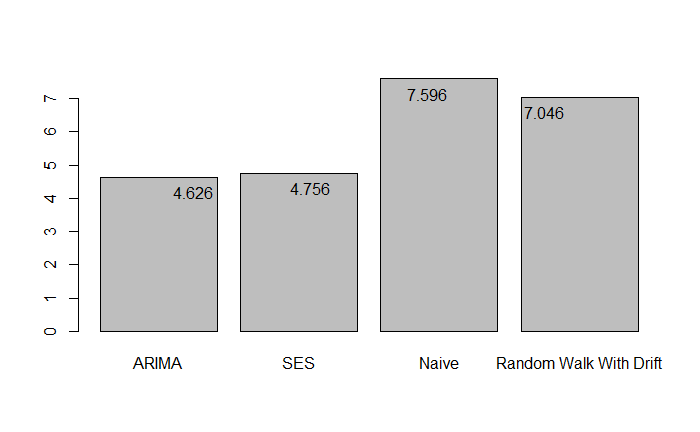


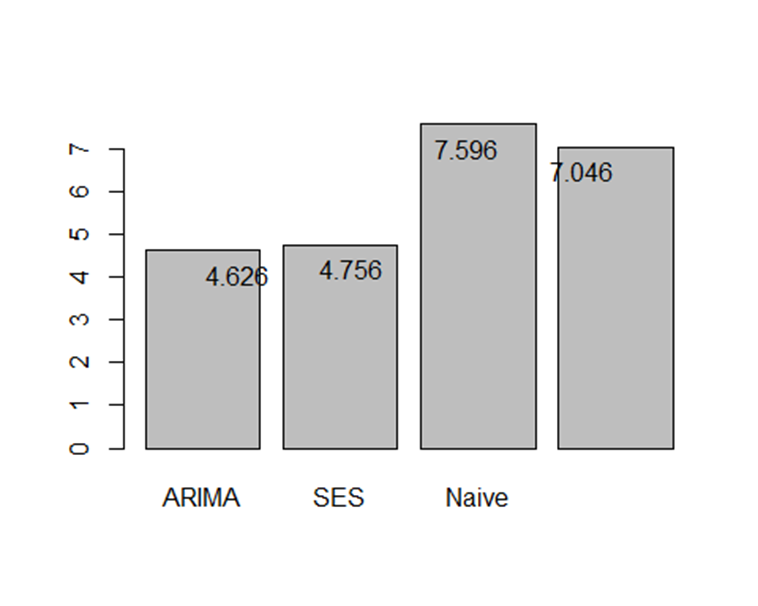


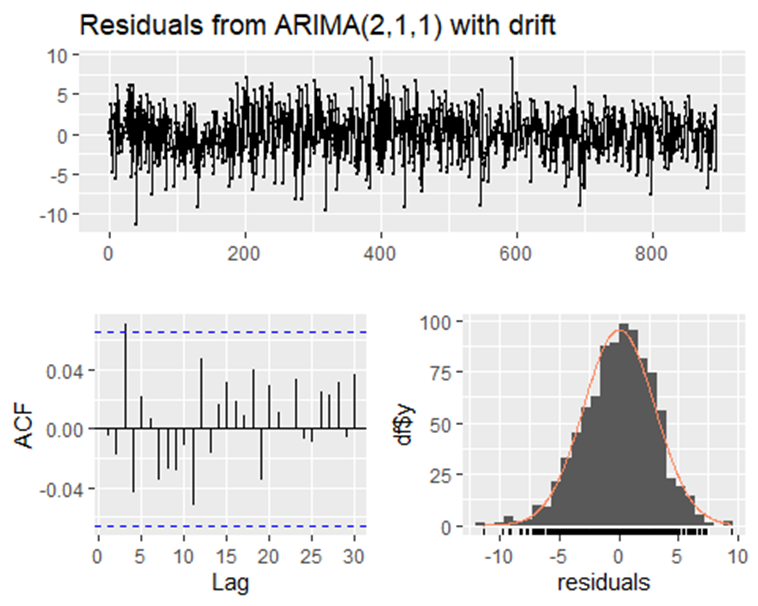
The Mann-Kendall trend test returns results indicating that there is significant evidence to reject the null hypothesis. There is a positive monotonic trend in the time series. The Sen’s Slope further supports this. Based on the test we see there is significant evidence to suggest there is a slope with a slope estimate of an increasing slope of 0.0368.

A graph showing a graph of a graph

Description automatically generated**Forecasting**

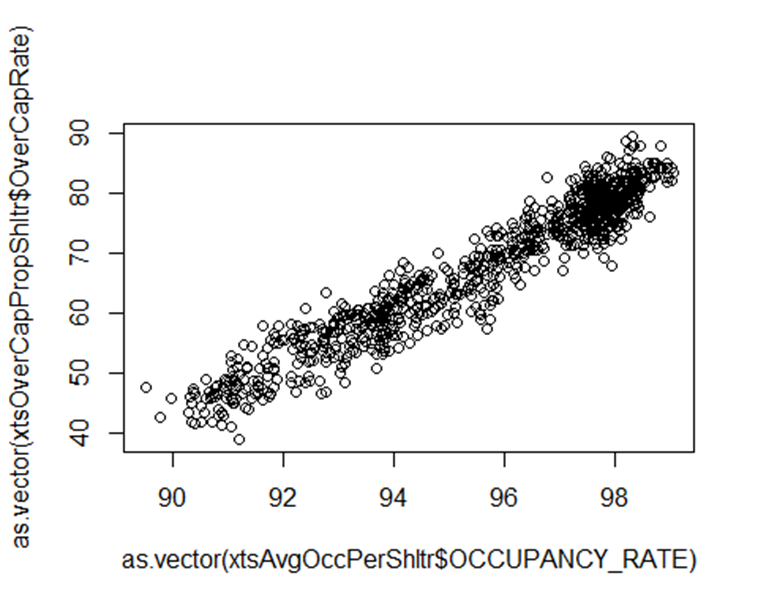






The ARIMA and SES model both have very close MAPE values however the MAPE is slightly lower for the ARIMA with the model differenced at lag one, with a autoregression of 1 and a moving average of 2. The plots pf the residuals also confirm that this is a good forecast as the residuals show normal distribution and constant mean and variance of the residuals.

**Correlation of Average Daily Capacity Rates and Proportion of Shelters Operating at or Over Capacity**

****



When running a Pearson correlation on daily average capacity rates and daily proportion of shelters at or over capacity we see a strong positive relationship, when occupancy rate goes up the proportion of shelters hitting or going over capacity increases. What this suggest is the more the capacity rates rise the more shelters are operating at or over capacity. If the mission is to provide shelter space for individuals experiencing homelessness having a high capacity rate is okay however if more shelters are reaching their limits it suggests that demand is increasing more than availability.

## Refugee Programs

Average Daily Capacity Rates

A graph of different numbers

Description automatically generated

| Metric | Refugee Programs Average Daily Capacity Rate | Other Programs Average Daily Capacity Rate |
| --- | --- | --- |
| Mean | 96.8 | 95.58 |
| Median | 99.22 | 96.26 |
| Standard Deviation | 4.61 | 2.39 |
| Minimum | 75.92 | 89.51 |
| Maximum | 100 | 99.08 |
| Range | 75.92 - 100 | 89.51 - 99.08 |

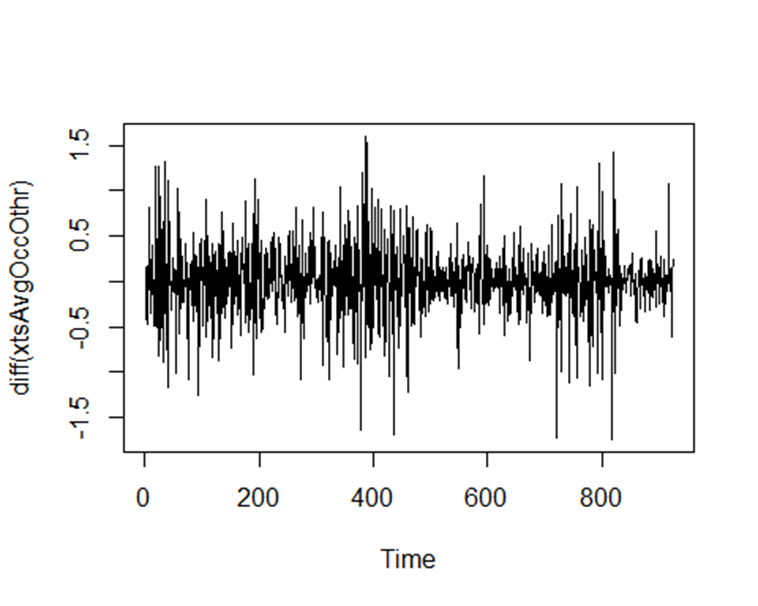


When comparing capacity rates between programs specific to refugees and other programs we see that after the peak for refugee programs, the capacity rates have consistently stayed high with large dips. Another important insight is the increase in trend appears to not be as drastic for non-refugee programs. Referring to the measures of central tendency it appears that refugee programs on average have higher capacity rates. The results of the Wilcoxon rank sum test show that there is significant evidence to reject the null hypothesis, concluding that there is a difference in medians between refugee and other programs.

A graph of black dots

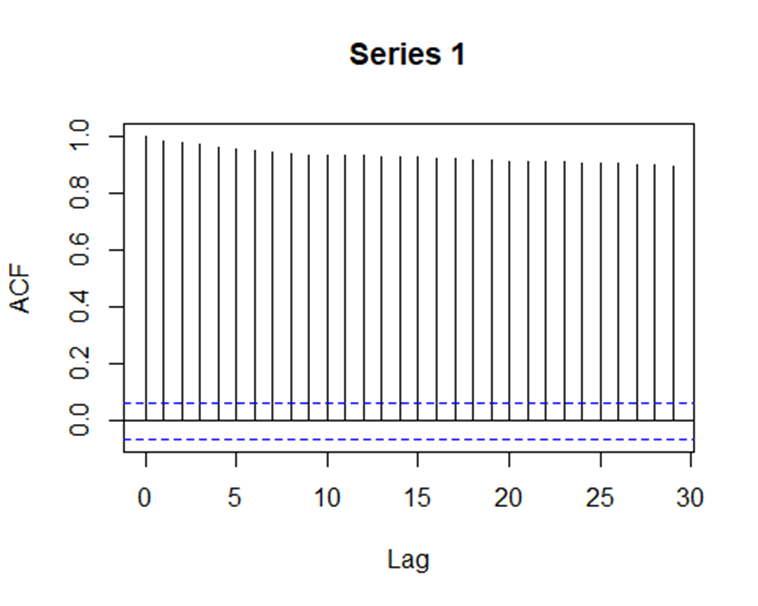
Description automatically generatedBased on the Pearson correlation of the daily average capacity rates between refugee and other programs it suggest there is a moderately strong relationship between the two. Indicating that they are both positively correlated, so when refugee programs see a increase in capacity rates so do other programs.

A graph of sound waves

Description automatically generated

A graph with lines and numbers

Description automatically generatedWhen comparing the plots of the differenced values for both programs we see that refugee programs have more variance towards the start of the time series however towards the end of the time series there are less fluctuations.



Both ACF plots indicate that there is significant autocorrelation for both refugee and other program types, however there appears to be higher significantly correlated lags for the refugee programs.



Interestingly the test for stationarity shows conflicting results. The ADF results are statistically significant, meaning we reject the null hypothesis indicating that there is enough evidence to conclude that the time series is stationary. However the KPSS results are also statistically significant which also means we reject the null hypothesis, which suggest that there is enough evidence to suggest that the time series is non-stationary. Referring back to the differenced plot, since both the test provide conflicting results the time series will need to be differenced when it comes to forecasting.

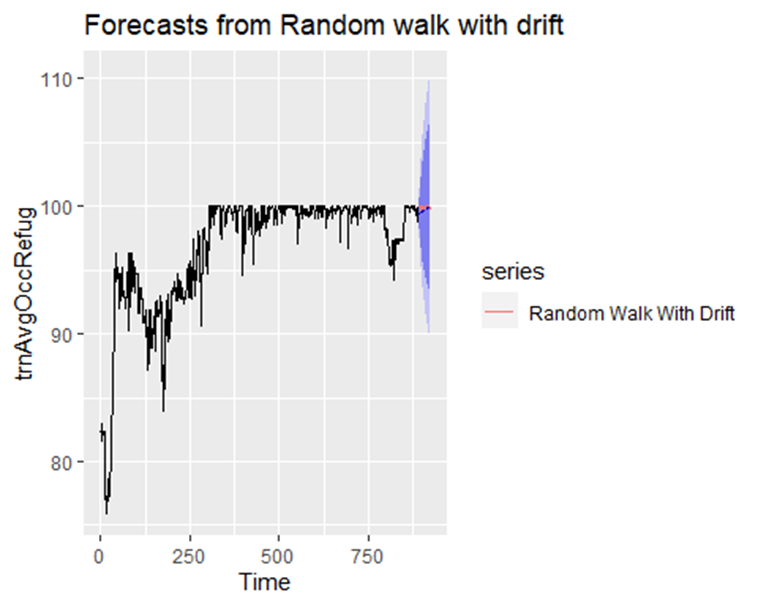


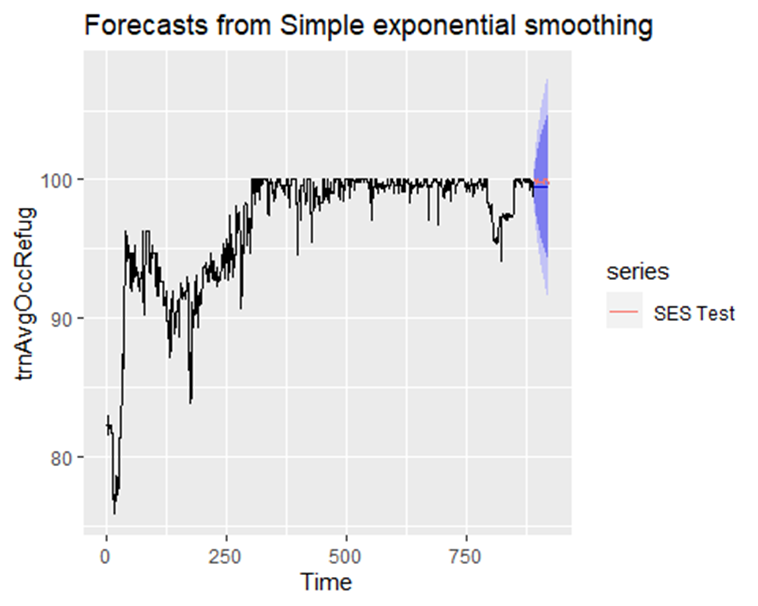
Investigating the trend further the Mann-Kendall test returns significant results. We can say that there is enough evidence to reject the null hypothesis and conclude that there is a trend in the average daily capacity rates for refugee programs, and the tau value suggest a positive trend. The results of the sen’s slope further confirm this and indicate that the trend has a slope of 0.0075.

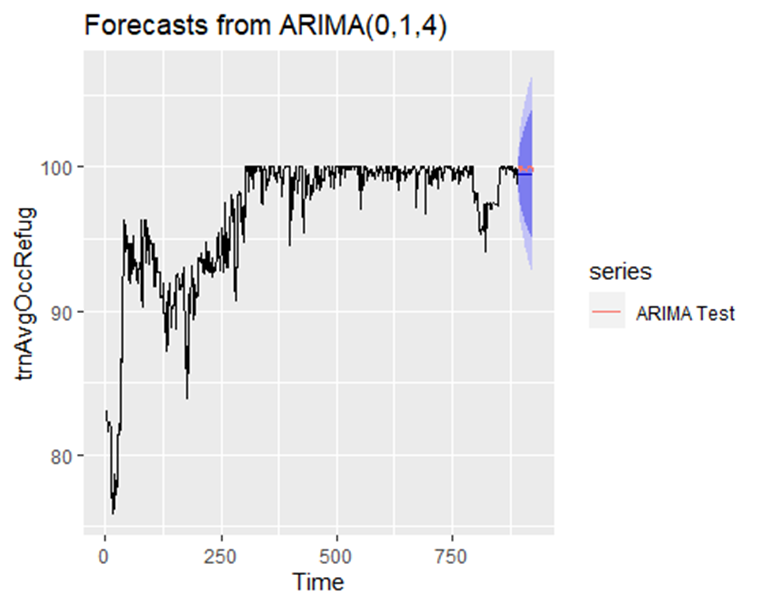
**Forecasting**

A graph showing a graph of a method

Description automatically generatedThe following methods are forecast for the daily average capacity rates just for refugee shelter programs.

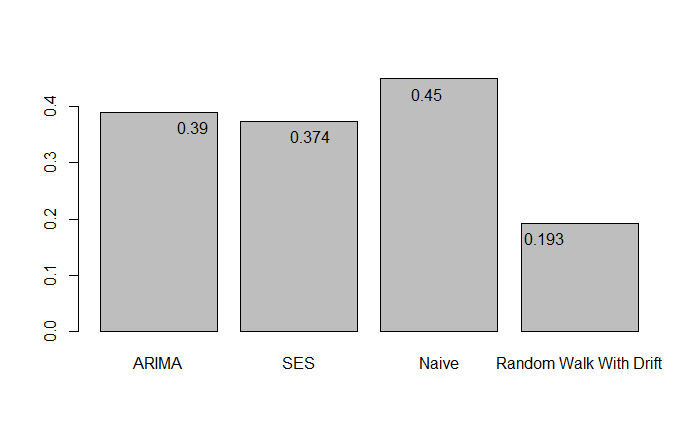


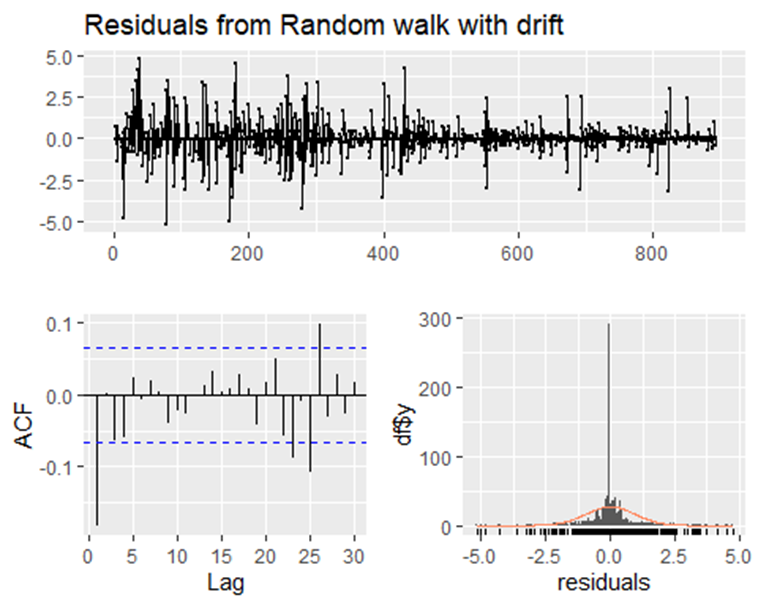




A graph of numbers and a number of objects

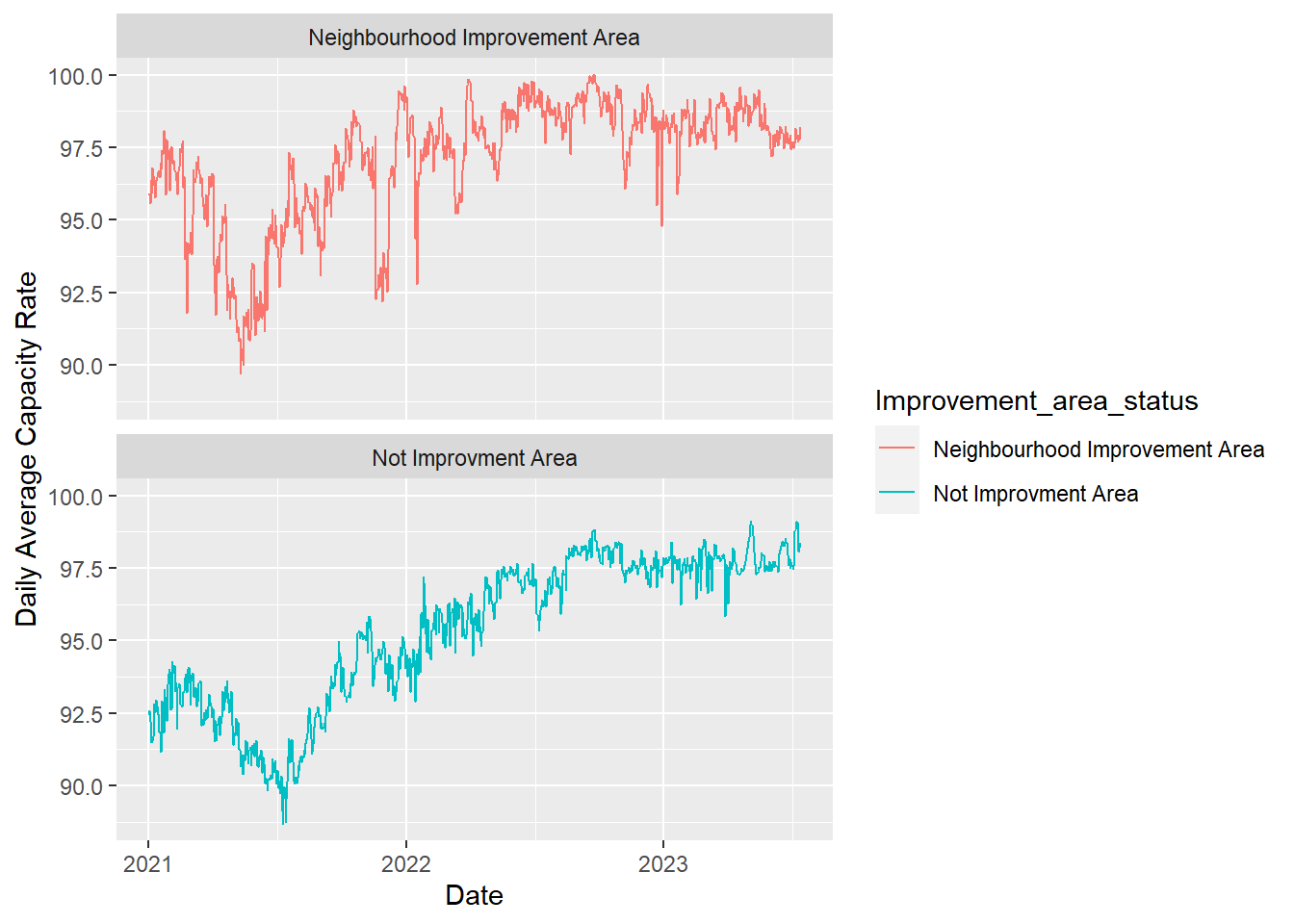
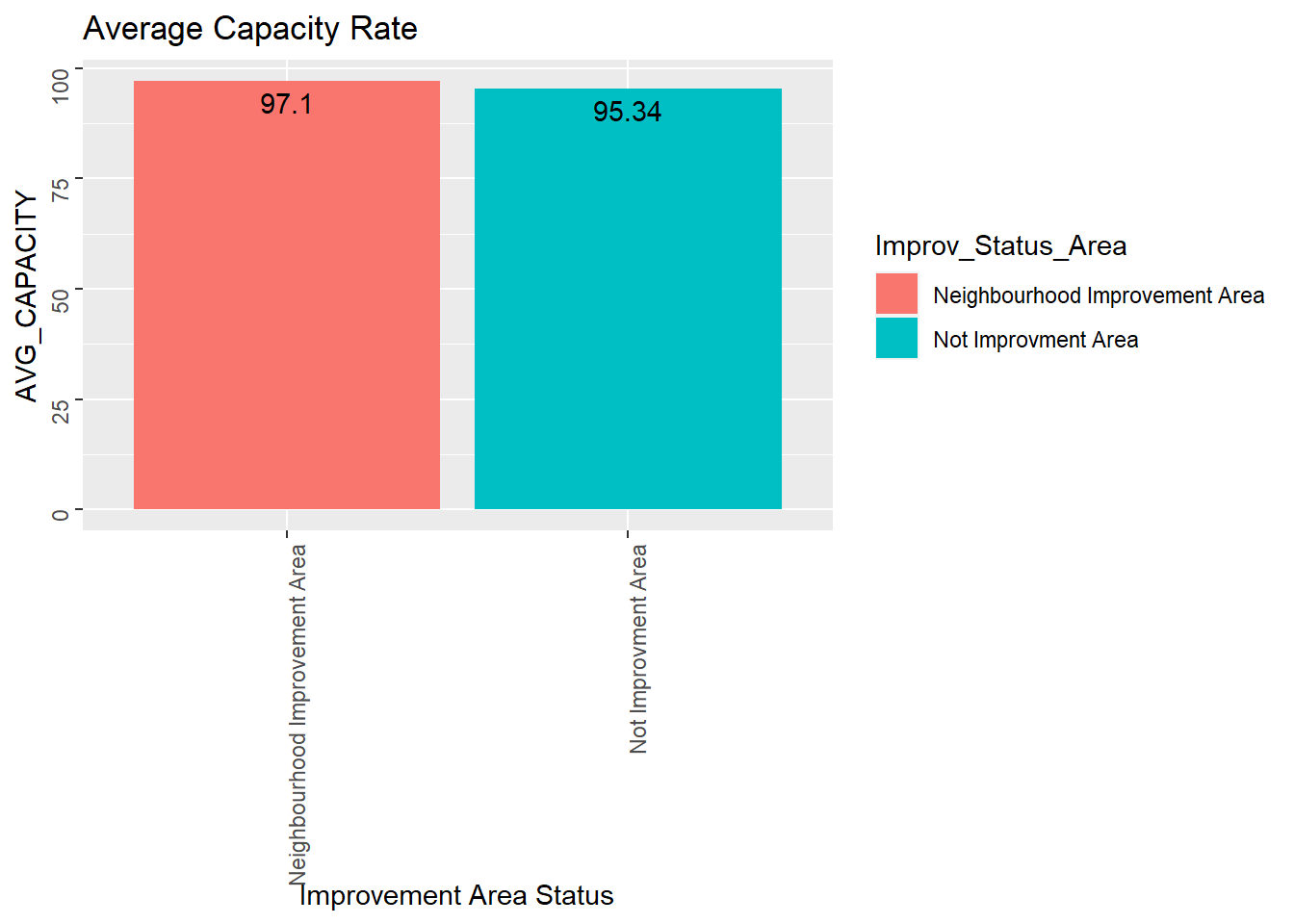
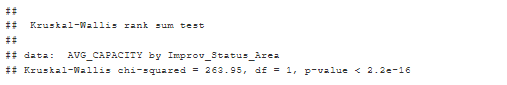
Description automatically generated



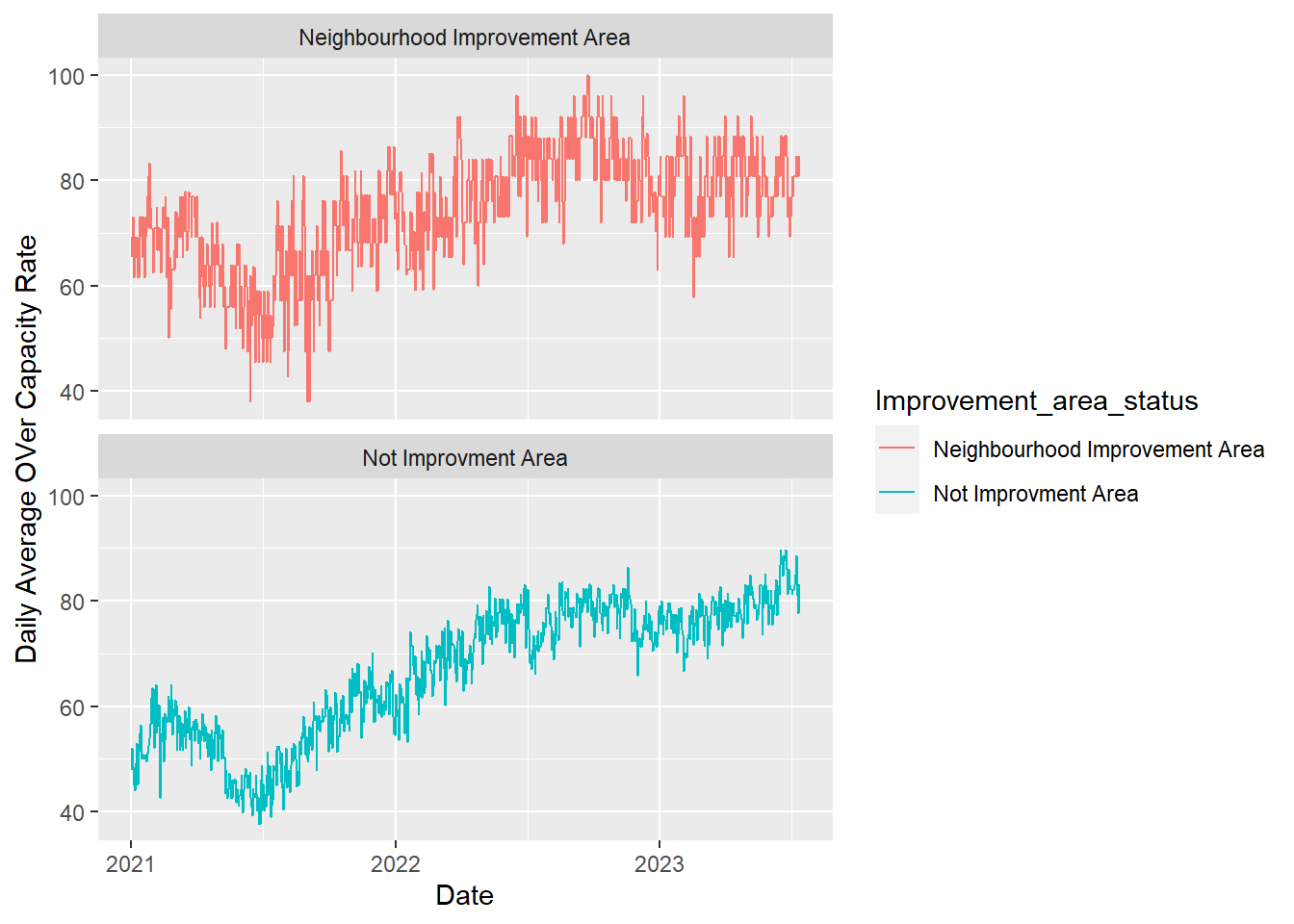


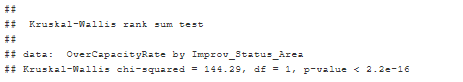
The Random Walk with drift produced the model with the best fit. The MAPE values are the lowest and the residual plots display normal distribution, no autocorrelation and white noise. I suspect the random walk with drift has the lowest MAPE values because it appears to take into consider for the changes at the end of the time-series.

**Neighbourhood Improvement Areas**

****

**A graph showing a number of different colored squares

Description automatically generated**

****

An investigation into the trend of capacity rates and over or at capacity rates for neighbourhood improvement areas allows for us to compare the difference in rates based on regional socio-economic variables. Based on the trend in average capacity rates and the overall average we can assume that neighbourhood improvement areas have higher shelter capacity rates. This is supported by the Kruskal-Wallis test, which shows statistically significant results indicating that there is a significant difference in averages between Neighbourhood Improvement Areas and other Toronto neighbourhoods. When comparing the trends it also looks like there are deeper dips and more variability in capacity rates for Neighbourhood Improvement Areas.

When comparing the trend in shelters that are over or at capacity, we see more variability with refugee programs and a higher increase. Neighbourhood improvement areas on average have a higher rate of operating at over or at capacity and this is further supported by the Kruskal-Wallis test. We reject the null hypothesis and conclude that there is enough evidence to suggest that there is a difference between neighourhood improvement areas and other neighbourhoods when it comes to the proportion of shelters operating at or over capacity.

## Limitations and Next Steps

The main contribution of the analysis is it provides up to date insights related to burden shelters are facing in Toronto. Understanding the proportion of shelters operating at or over capacity is a unique quality of this project and is not something that has been talked about in any of the literature.

The analysis focused on mainly very high level continuous measure that that speak to the burden most shelters are experiencing. Next steps would be to analyze discrete measures which add more context, for example the trend in the size of the programs or total intakes or total operating programs. Additionally more research into the correlation and causation would also be necessary, like further examining the peaks and dips against other datasets to see what else is overing during that time. For example understanding the synchronicity between the shelter measures and housing affordability, or CPI index or inflation or unemployment or lockdown periods.

## Conclusion

Homelessness is not something that is easy to solve however it appears that Toronto is struggling to address the crisis more than ever. Indivudals experiencing homelessness are some of the most vulnerable and without emergency shelters most will have to resort to the streets as seen more and more recently. After the analysis it is evident that supply does not meet demand and Toronto needs more support and funding to deal with the crisis.

## Literature Assessment

Unlike many other of the literature I read, my time series did not experience seasonality. Also falling in line with what others have mentioned I noticed that neighbourhood that are unequitable in sense to socio-economic characteristics are seeing a larger % of shelters operating at or over capacity, suggesting that these areas are again not receiving the appropriate funding in reference to shelter programs.

## Q & A

Q: What does the trend look like for capacity rates for Toronto shelters?

A: The results of the analysis show that there has been a increasing trend for both average capacity rates and the proportion of shelters operating at or over capacity. The trend does not show seasonality but the trend appears to be more stabile at the end of the time series. Based on the autocorrelation of the trend there is a significantly strong correlation between observed values and their prior values 1 day prior.

Q: What are the characteristics of the capacity rates between refugee and other shelter programs?

A: Refugee programs appear to have higher rates in capacity and proportion of shelters operating at or over capacity than other shelter programs. This trend seems to increasing and staying high. There does not appear to be seasonality with refugee program capacity rates and proportion of programs operating over or at capacity. However there is a significant difference between the program types and on average refugee shelters have a higher capacity rate and proportion of shelters operating at or over capacity.

Q: Based on socio-economic factors of the city what areas have the highest capacity rates?

A: Based on the results of the analysis Neighbourhood Improvement Areas have higher capacity rates and a larger proportion of shelters operating at or over capacity.

## References

Analytics Vidhya (2021), Statistical Tests to Check Stationarity in Time Series.

https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/#:~:text=ADF%20test%20is%20conducted%20with,series%20has%20no%20unit%20root.

Clemens, K. K., Ouédraogo, A. M., Le, B., Voogt, J., MacDonald, M., Stranberg, R., Yan, J. W.,

Krayenhoff, E. S., Gilliland, J., Forchuk, C., Van Uum, R., & Shariff, S. Z. (2022*). Impact of*

*Ontario’s Harmonized Heat Warning and Information System on emergency department visits*

*for heat-related illness in Ontario, Canada: a population-based time series analysis*. Canadian

Journal of Public Health, 113(5), 686-697. <https://doi.org/10.17269/s41997-022-00665-1>

Gaetz, S. (2010). *The struggle to end homelessness in Canada: how we created the crisis, and how we*

*can end it*. The Open Health Services and Policy Journal, 3, 21–26

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts:

Melbourne, Australia. OTexts.com/fpp2. Accessed on 07-17-2023.

Jadidzadeh, A., & Kneebone, R. D. (2015). *Shelter from the Storm: Weather-Induced Patterns in the*

*Use of Emergency Shelters.*The School of Public Policy Publications

*(SPPP), 8*https://doi.org/10.11575/sppp.v8i0.42500

Jenkinson, J. I., R., Strike, C., Hwang, S. W., & Di, R. E. (2021). Nowhere to go: exploring the social

and economic influences on discharging people experiencing homelessness to appropriate

destinations in Toronto, Canada.*Canadian Journal of Public Health, 112*(6), 992-1001.

<https://doi.org/10.17269/s41997-021-00561-0>

Pohlert, T (2023, March 26). A Non-Parametric Trend Test and Change-Point Detection.

https://rdrr.io/cran/trend/f/inst/doc/trend.pdf

Uppal,S (2022, March 14th). A portrait of Canadians who have been homeless. Statistics Canada.

https://www150.statcan.gc.ca/n1/pub/75-006-x/2022001/article/00002-eng.htm

Zhang, P., Wiens, K., Wang, R., Luong, L., Ansara, D., Gower, S., Bassil, K., & Hwang, S. W.

(2019). Cold Weather Conditions and Risk of Hypothermia Among People Experiencing

Homelessness: Implications for Prevention Strategies.*International Journal of Environmental*

*Research and Public Health, 16*(18)https://doi.org/10.3390/ijerph16183259

Stephanie Glen. "ADF — Augmented Dickey Fuller Test" From StatisticsHowTo.com: Elementary

Statistics for the rest of us! https://www.statisticshowto.com/adf-augmented-dickey-fuller-test/