Case Study: Austin-Bergstrom International Airport

Joaquin Maroto, Ryan Daher, Qiji Xiang and Karl Juhl

IE University

Professor: Borja González Del Regueral

Table of Contents

Abstract

The Austin-Bergstrom International Airport has been collecting monthly data on their number of visitors for the past 6 years. This data can prove to be of great importance to companies like the airport, which the data shows has been receiving more and more visitors over this short period. With more visitors the airport may need to expand, which is the question this paper will explore. Using R Studio, ETS and ARIMA models, the solution was found; the visitors of the airport would continue to increase steadily, and the expansion will certainly be required in order to continue operating at comfortable levels. This paper explores which method was found to be the best, and how it was found to be the best.

Introduction

The Austin-Bergstrom International Airport is now facing a difficult situation. There has been a rise of visitors for the past five year and the number of visitors is approaching the airport's capacity. The airport's current capacity is 15 million visitors per year. Last year (2018), the airport experienced 15.8 millions visitors and a lot of the visitors complained about over-capacity in terms of long waiting time, over-crowded and so on. The dataset contains data up until October 1st, 2019. From the dataset, a record high 14.4 millions visitors was observed for the first ten months of this year. This paper aims to model the growth of the number of visitors and predict how many visitors are expected to visit next year which helps the local government decide if an expansion of the airport is required to facilitate the growing demand. The modeling techniques used in this paper is primarily ETS and Arima. In particular multiplicative decomposition and seasonal Arima. Different models are constructed and compared. The best model is selected and the number of visitors regarding the rest of the year and next year is predicted using the best model.

Theory

Holt-Winters' multiplicative method[1]

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}$$

[1] (Iranica, 6)

Holt-Winter' multiplicative method is an extension to Holt's method to capture seasonality. The method consists of the forecast equation and three three smoothing equations - one for the level l_t, one for trend b_t, and one for the seasonality component s_t. As the image above shows, these smoothing equations also come with smoothing parameters α, β*, and γ. m denotes the frequency of the time series.

Seasonal Arima[2]

Arima is obtained by combining differencing with autoregression and a moving average model. Seasonal Arima obtained by adding an additional seasonal terms in ARIMA. From an ARIMA (p,d,q)(P,D,Q)m:

$$(1 - {}_1B - ... - {}_pB^p)(1 - {}_1B - ... - {}_PB^{Pm})(1 - B)^d(1 - B)^{Dm}y_t =$$

$$c + (1 + {}_1B + ... + {}_qB^q)(1 + {}_1B + ... + {}_QB^{Qm})_t$$

As the image above shows, The additional seasonal terms are simply multiplied by the non-seasonal terms.

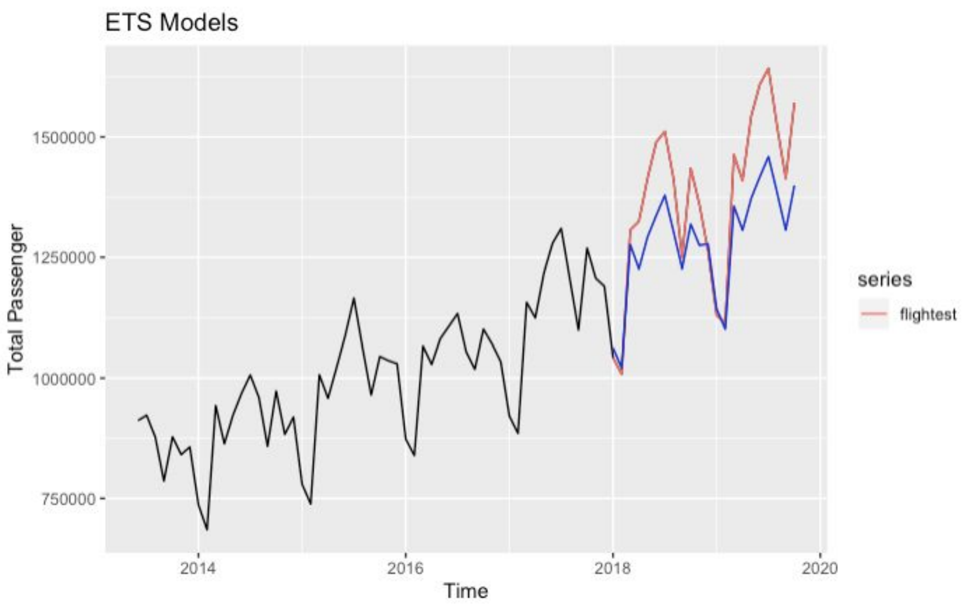
Implementation

We needed to install four different types of packages. To begin with, we installed "fpp2", which includes the basics and examples of forecasting. Moreover, we also used "forecast" to be able to apply arima models. Then, we installed "ggplot2" to create data visualization (e.g. autoplot(), ggsubseries()...). And finally, we also used the "seasonal" package to apply seasonality to the arima model.

[2] (Springer, 22)

We used the dataset Airport_Monthly_Operational_Report dataset (a csv file) obtained from *Austin Texas Government*[3] webpage which contains a total of 49 variables, from which we only used the number of passengers per month throughout a period of six years: starting in 2013 and ending in 2019.

We conducted some basic analysis from where we observed that there is a trend, cyclicity and seasonality. We looked at ACF and observed a clear increase in variance throughout time. Therefore, we decided to use a multiplicative decomposition. We used holt's winters and holt's winters damped for the train data and obtained that the best model is MMM. This is the best fit model with ETS:



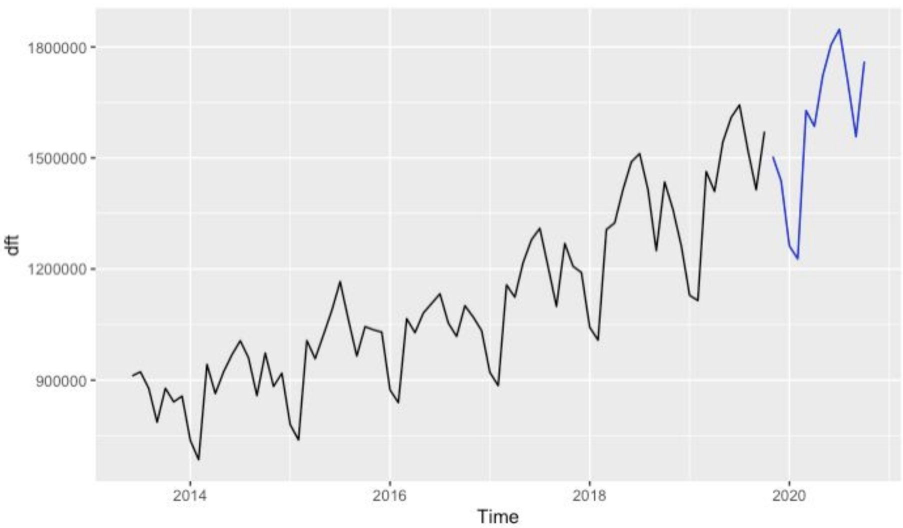
Nevertheless, we compared them to benchmark methods and model AAA and saw that the RMSE was slightly higher. Therefore, we looked at the residuals and decided to use seasonal ARIMA for the best fit model. We saw that there was a trend and difference in the variance throughout time and therefore, we used BoxCox to transform the data set and better fit the model. Consequently, we looked at ndiffs() for both the trend and seasonality and

[3] (Ciry of Austin, 1)

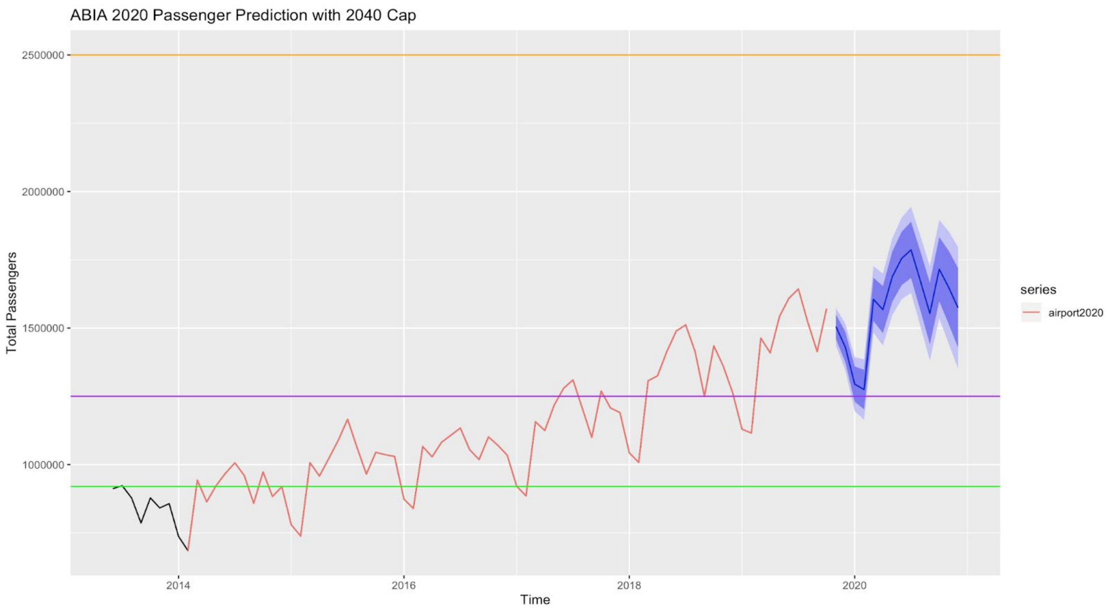
differentiated them once each. After having a stationary model, we looked at different candidates for the best fit model (including lambda as model had transformed with BoxCox):

- Trend (1,0,0) and Seasonal (0,0,1)
- Trend (0,0,1) and Seasonal (0,0,1)
- Trend (1,0,1) and Seasonal (1,0,1)
- Trend (1,0,1) and Seasonal (0,0,1)
- Trend (2,0,0) and Seasonal (0,0,1)
- Trend (0,0,2) and Seasonal (0,0,1)

The lowest AICc were Trend (1,1,0) and Seasonal (0,1,1) and Trend (1,1,0) and Seasonal (1,1,0) including differentiation. We plotted them, check residuals and normality and conducted an accuracy test for both of them. The model with lowest RMSE was the second one. Later, the auto.arima test showed that the best model is Trend (0,1,1) and Seasonal (0,0,1). This is the best fit model for the dataset:



Measuring total passengers per month, the plot shows the historical data with our forecast fitted. The horizontal green line represents the original monthly capacity of the airport before the first expansion (in 2013), and the purple line represents the current monthly capacity (2019). Overall, our findings provide evidence that the airport clearly made the correct decision in initiating the 2040 master plan, which, when finished, will hold the orange line as the monthly capacity. Nevertheless, it is important to understand that throughout time, reliability towards forecasting decreases. This is the prediction with the three means of the number of passengers per month (2013, 2020 and 2040):



Conclusion

The goal of this project was to explore the data of the Austin-Bergstrom Airport and see if modelling it using the methods learned in the past semester would show any meaningful results. This goal was surpassed by the insights extracted from the data. The capacity limits given by the airport as well as the forecasts produced from the models in this project gave a window into the real-life application of analysing a time series and forecasting on it. Before finishing this project our group was unaware of the fact that the airport already had a 2040 master plan of expanding the airport. Our models gave us similar information through our forecasting methods which the airport's own researcher's most likely found. Realizing the 2040 plan existed confirmed our conclusions and gave our team confidence in our new-found forecasting & time series analysis skills.

References

I. City of Austin. (2019). *Airport Monthly Operational Report:* https://data.austintexas.gov/City-Government/Airport-Monthly-Operational-Report/67kt-ykxh

II. City of Austin. (2019). *Airport Master Plan:* http://www.austintexas.gov/page/airport-master-plan

III. Facebook. (2019). *Quick Start Prophet:* https://facebook.github.io/prophet/docs/quick_start.html#r-api

IV. Scientia Iranica (2016). Robust Holt-Winter based control chart:

http://scientiairanica.sharif.edu/article_3902_fe55e2d9d1b83aad0df15e460a838a60.pdf

V. Springer Link (2015). Seasonal ARIMA model with limited input data

https://link.springer.com/article/10.1007/s12544-015-0170-8