JULY 31, 2018



# AIRCRAFT TAXI TIME PREDICTION

BY: SEYED REZA MIRHOSSEIN CEU ID: 165066

Capstone project for MSc in Business Analytics Public Summary Report

<u>Supervisors:</u> Prof. Gabor Bekes (CEU) Mr.Roland Guraly (Slot Consulting)

#### **1. Introduction**

Flight delays are a common hassle for many travelers and aviation stakeholders. It directly affects everyone's daily life and truly results in a considerable amount of financial and environmental loss. AviolQ as the latest development at Slot Consulting, developed a mobile information sharing system which connects existing airport systems with mobile devices (smartphones, tablets and other wearable devices) carried by ground staff. The system provides the right information at the right place and time to the right people, making it easier for planners to make optimal use of resources.

This helps reducing primary and reactionary delays, improving the reliability of estimates, and allowing airlines to make smarter decisions about turnaround buffers, that will be shorter than before. AviolQ also offers a simple way to simulate turnaround activities, making it possible to run operational exercises using the same mobile tools, which simplifies and improves the quality of training, as well as providing the basis for designing longer-term improvements in ground handling processes.

The essential element of the solution is to define default values for each turnaround task. So far, aircraft taxi time is considered as one of the crucial phase of the turnaround which can greatly affect the aircraft ground time estimation, particularly the slot allocation and apron management at the airports.

### 2. Study Background

Aircraft turnaround process requires a great number of resources. For each process there is the equipment and the staff that operates it. To manage the resources a number of factors should be taken into consideration. First, there is the schedule that represents the timing and the aircraft types that will be arriving and departing at particular date in their own time. It also provides information on which airline to operate the particular flight. Each aircraft type and airline has its own requirements, so the handling organizations should be prepared to perform their activities accordingly.

Usually the staff and equipment are not present during the whole turnaround at one flight they have their role at a certain point of the turnaround and then they are migrating to the next turnaround. To be able to conduct efficient management of the resources it is important to know how long it would take the particular unit to service a particular flight or rather aircraft at particular stand and estimate how long it would take that particular unit to relocate to the next stand, where the next task should be performed. AvioIQ is designed to ensure proper turnaround management through smooth information flow between different stakeholders for more optimal decisions in critical turnaround phases.

The uncertainty in prediction of the ground delays results in tangible demand of an IT management control system. AviolQ provides variety of functionalities through four different solution packages namely; AirlineIQ, ApronIQ, DepartureIQ and AirportIoT in order to serve all stakeholders individually or in an integrated module.

The fundamental phases of aircraft turnaround are ground handling sequential activities and taxiing to the gate or runway. The ambiguity of the duration of aircraft taxi time beside the involvement of

various known and unknown factors have resulted in importance of better understanding of its behaviors and identify a series of methods in order to predict the duration of taxi time for both departure and arrival flights. Accordingly, a prediction methodology based on data analytics techniques are introduced as a potential future approach to enhance airport scheduling and reduce flight delays. Associated data source is extracted from US Department of Transportation (Bureau of Transportation Statistics) for JFK airport in 2017. The analysis is focused on all departure and arrival flights in JFK airport for four months in 2017 (Jan, Apr, Jul, Oct). Moreover, the average daily data are merged to the original database which helps us to investigate the weather impacts as well.



# 3. Methodology

Aircraft taxi time is the time that an aircraft spends taxiing between its parking stand and the runway or vice versa. The main intention is to find the best method to predict the duration of this variable considering all other variables. Having a good prediction model leads to achieve better scheduling by knowing the taxi time of the future flights according to other variables such as destination, origin, day of week, time of flights, airlines, weather and so on. In this regard, the initial phase is to explore the existing data source and try to understand the underpinned data characteristics to show their relationships to the target variable. At the same time ,the best method to handle missing data and extreme values should be defined. Although, the extreme values for flight delays and taxi time seemed extraordinary, the essence of the study leads to keep most of them and remove a small portion very high and extreme small values of aircraft delays and taxi time (Concluded to cause bias in our prediction). Moreover, remove missing values for all variables and replace them with zero for binary ones. In order to reach the most realistic prediction, the taxi time values below 5 minutes were removed.



Exploratory data analysis takeaways were to convert the target variable and identify the relevance of predictors to the taxi time. Based on the initial visualization and correlation matrixes we can guess about the possible trend in data and important predictors (i.e. positive correlation of Dew points and temperature with taxi time duration as well as flight duration). After cleaning, visualization and explore data characteristics, the proposed methodology is to try different prediction models such as regression and machine learning algorithms and evaluate their performance on out of sample dataset. The prediction process provides information about the importance of predictors as well as prediction power, however the interpretability of machine learning models are significantly lower comparing to linear regression and regression trees.

According to the Initial exploratory analysis, the impacts of the weather data in to aircraft taxi time is residual expect tough weather events such as thunderstorm or snowy days. On the other hand the obvious trend is tangible for taxi time and different weekdays or monthly days and more importantly the contribution of flight time during the day has influence in taxi time. Moreover, it is expected to see found hidden trends in different flight routes corresponding to various airlines. within analysis phase the deployed prediction models and their performance power are defined in below table and would help us to apply the best model in out of sample (entire population) after careful validation.

No	Model Name	Model Explanation	Performance measure (R squared)
1	Linear Regression	Simple linear regression with all predictors	0.7581
2	LASSO	Simple linear regression with shrinkage of unimportant variables	0.7579
3	Regression Tree	A decision tree with recursive partitioning of the most important predictors	0.6864
4	Random Forest	constructing a multitude of decision trees at training time and outputting the mean prediction (regression) of the individual trees.	0.7731
5	Gradient Boosting Machine	builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.	0.7747

# 4. Conclusion

This document has presented a series of conventional data mining steps to reach a final best performing model in out of sample observations. According to the machine learning performance measure, we can claim a pretty good prediction power which also can be applied in external datasets, while it is expected that the external validity for other airports requires additional analysis in order to validate the existing models. Moreover, the benefit of this validation process might be to recognize the impacts of other hidden variables or so called confounders in predictions. the apron movement tracking data can reveal an interesting trends and improve the prediction power of the model. These movements is not limited to the aircraft but also ground handling vehicles and equipment movement.

However, the existing analysis provides a valuable information about the most important variables in taxi time prediction which are highlighted in details of last two models as well as tree based model. like any data analysis project, the causal effects should be considered in a more detailed study with extensive variables. Obviously, the prediction would be more and more accurate by retrieving more variables from airport operational database (i.e. aircraft parking slots, hourly weather data, flight queuing and airport operational limitations).

The later use cases of this methodology would be applied in a much more complex machine learning models which will present relatively higher prediction power.