

# Prediction of Flight Delays and Cancellation

Keerthana K, Viswa Harini T N, Deivarani S

**Abstract**— In this hasty world, there is no time to get ahead of our work and no one has tranquility to wait. For example, flight delay plays a major role in this aspect. The scope of our project is to deliver to the people on which is the most consistent airline and the one which provides a more consistent travel experience with respect to delays. The time delay maybe caused due to disruptions like technical breakdowns, late check-ins and airspace congestion. Therefore, the consideration of timeliness and robustness has become an important topic in analysis of variance. In analysis of variance, the parameters are taken into consideration are arrival, departure, and the delay times. On the consideration of these parameters, ANOVA (Analysis of variance) is performed, which shows the most reliable and consistent airlines.

**Index Terms**— Prediction, forecasting, flight delays, alysis of variance, measure of central tendency, measure of dispersion

## I. INTRODUCTION

The worldwide transportation systems consider commercial airlines as their backbone, bringing significant utility by enhancing efficient and easier long distance travel. Delay is one of the major problems that affect any transportation system. Flight delay is considered to be the crucial problem faced by air travelers as it causes major impact on their time. Delay is the difference between the scheduled time and real times of departure or arrival of the flight. Modern passengers are most probably affected because of flight delays as they don't recognize which Airline Company could reach their destination as fast as possible. Thus flight delay is considered as a model problem and is also widely understood among the air travellers. From the maintainability point of view, delay may also cause environmental effect and damage to resources and increasing the fuel consumption and gas emission.

In the context of air transportation system, flight delay plays a vital role. Flight delay not only impact airlines and passengers but also the airport. This is because, the time allocation for the arrival of the flight changes from flight to flight based on their departure time but, if any one of the flight delays, then the total system gets affected and thus cause an inconvenience to passengers. The prediction of flight delays can improve the technical and operational decisions of airport authorities and airline managers to inform the passengers in advance such that, passengers could reschedule their travel plans and relatively reduce the delay impact.

In this study, data of 14 top airlines have been collected, analyzed, establishing flight delay data, understanding the various types of problems and observing highlights about flight delay prediction problem.

**Manuscript received September 22, 2019**

**Keerthana K**, Student, Dept. of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

**Viswa Harini T N**, Student, Dept. of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

**Deivarani S**, Assistant Professor, Dept. of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

## II. REASON FOR FLIGHT DELAY AND CANCELLATIONS

Flight delay occur when a flight takes off and or lands later than its scheduled time. A cancellation occurs when the airline does not manipulate the flight at all for a certain reason. The main factors for flight delay and cancellations are bad weather, exceptional circumstance, mechanical or technical reasons, congestion in air traffic, engineering issues, waiting for connecting passengers and baggage, security clearance, delay in boarding passengers, waiting for crew and issues pertaining to signals from the aircraft traffic control authorities and etc... Normally, aircraft preparation takes a lot of time to get the flight prepared and ready to fly. The operation team is in-charge and lots of communication will be going behind the screen. But then, sometimes such communications or tasks goes misaligned and delays the flight. All these reasons happen frequently leading to an increase in the situations of inconvenience and unwarranted waiting time of passengers.

## III. EFFECTS ON AIR TRAVELLERS

Flight delay can cause more inconvenience to air travellers as they are the modern passengers; they depend more on air transportation system. Passenger behaviour and health concern is also important for flights to be operated as per schedule. Thus flight delay is considered as a major problem among passengers, as they cannot waste their time for the welfare of that particular airlines. The environmental impact of aviation also occurs due to aircraft engines which emit heat, noise, emit gases and contribute to climatic change and global dimming. The engines of airplane work by burning fossil fuel and produces carbon dioxide and water vapour thus results in affecting not only air travellers but every stakeholder.

## IV. DATASET AND EXAMINATION OF DATA

The dataset of flight delays and cancellation is collected from The United States Department of Transportation's (DOT) Bureau of Transportation Statistics which tracks the on-time performance of domestic flights contains the information on flight numbers, number of on-time, delayed, cancelled, and diverted flights is published in DOT's monthly Air Travel Consumer Report of 2015.

In this dataset there are 14 Airlines around the world, each airline contains corresponding International Air Transport Association (IATA). IATA code is known as location identifier and designating many airports all over the world.

Table 1: Airline names with IATA code

Sl. No.	IATA Code	Airline Name
1	UA	United Air Lines Inc.
2	AA	American Airlines Inc.
3	US	US Airways Inc.
4	F9	Frontier Airlines Inc.

5	B6	JetBlue Airways
6	OO	SkyWest Airlines Inc.
7	AS	Alaska Airlines Inc.
8	NK	Spirit Air Lines
9	WN	Southwest Airlines Co.
10	DL	Delta Air Lines Inc.
11	EV	Atlantic Southeast Airlines
12	HA	Hawaiian Airlines Inc.
13	MQ	American Eagle Airlines Inc.
14	VX	Virgin America

The attributes used in all the 14 airlines are follows:

- Year - The dataset of the year of 2015 flight delays and cancellation
- Airline – An organization that provides air transport services to one or more destinations
- Flight number – It refers to the numeric identity assigned to a flight for operation on a specific route
- Tail number – It refers to the identification number delineated on an aircraft commonly called as tail where it can be represented as an aircraft registration number
- Origin airport – It refers to the taking-off area or starting point of an aircraft
- Destination airport – It refers to the landing area or final point of an aircraft
- Departure delay – A departure delay is when a flight takes off later than its scheduled time and it is calculated by the difference of scheduled departure time and actual departure time
- Scheduled time - The scheduled time of an airline flight is calculated at the beginning of the occurrence that the flight should arrive at a given scheduled time
- Elapsed time – The elapsed time is the amount of time that the flight travels from the beginning of an event to its end or the difference between the beginning time and the ending time.
- Air time – The amount of time that the flight travels in air
- Distance – The full length of flight travels from the start to end point
- Arrival delay – An arrival delay is when a flight lands later than its scheduled time of arrival and it is calculated by the difference of scheduled arrival time and actual arrival time
- Diverted flight – A diverted flight is the one which is routed from the scheduled origin point to a new destination point other than the scheduled destination point or arrival destination point.

Upon considering these attributes of flight delays and cancellation dataset, efforts are taken to predict the best airliner, which is as well more consistent in its operation among the identified 14 airlines; without much delays or causes of disruption which shall help passengers to save their time instead of waiting for a long time. This study attempts to predict flight delays using analysis of variance (ANOVA) considering the attributes listed in the previous section.

V. USE OF ANOVA

Analysis of Variance (ANOVA) is a hypothesis-testing technique used to test the equality of two or more population means by analyzing the variances of samples that are taken. It is a collection of parametric statistical technique used to compare the dataset and their association estimation procedures such as variation among groups and also used to analyze the differences over the group means. ANOVA is an extension of t-test and z-test when we have samples from more than two populations; F-distribution is used to test the significance of difference between means. ANOVA assumes the population involved follows a normal distribution, ANOVA falls into the category of hypothesis tests known as parametric tests. It helps to find out whether survey results are significant or not. The null hypothesis for the test assumes the flight delays and cancellation are least (H0) and the alternative hypothesis assumes the flight delays and cancellation are most (H1). Using F-distribution table, the observed value and the critical value are compared and thus reject or accept the null hypothesis. Hence, the most consistent and reliable airline could be predicted.

VI. OBSERVATIONS AND CONCLUSION

Since flight delays cause economic significance to all the passengers and airlines, diagnosing them through our prediction may improve the marketing decisions. It may also increase costs to passengers and operational costs to airlines. In this context, for flight delay prediction

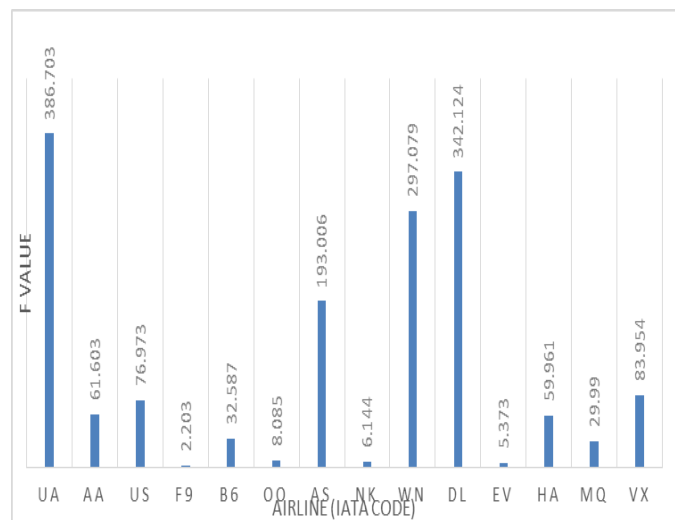


Figure1: An analysis of Airline with F-value

Analysis of variance test is performed for 14 airlines and checked the significance of each airline. As we compared the parameters of departure delay and arrival delay, thus obtained the F values for the all airlines and the F critical value is constant of 3.8417 for all the airlines.

From the analysis, it is concluded that Frontier Airline (F9) is more significant as compared to all other airlines. As F calculated value is lesser than the F critical value (2.203<3.841) and so the null hypothesis accepted. The remaining airlines are not significant as all F calculated values are greater than the F critical value, and so the null hypothesis is rejected.

Therefore, it is concluded that the Frontier Airline is more consistent and reliable taking in to account all the parameters

identified and included in the dataset of flight delays and cancellation. This analysis may be useful to air travellers while choosing the best airline with less time delay and low cost. Further, shall pave way for providing a feasible environment for the air transportation system.

#### REFERENCES

- [1] Hooda R P, 2013, Statistics for Business and Economics, 5e, Vikas Publishing House Pvt. Ltd., Noida, India.
- [2] Levine D M, Berensen M L, Krehbiel T C, Viswanathan P K, 2011, Business Statistics: A First Course, 5e, Pearson Education, New Delhi, India.
- [3] Richard I Levin, David S Rubin, 2011, Statistics for Management, 7e, Pearson Education, New Delhi, India.
- [4] <http://www.cs229.stanford.edu/proj2016/report/MenonMvva-PredictingFlightDelays-report.pdf>
- [5] <https://www.kaggle.com/usdot/flight-delays>
- [6] [https://www.researchgate.net/publication/315382748\\_A\\_Review\\_on\\_Flight\\_Delay\\_Prediction](https://www.researchgate.net/publication/315382748_A_Review_on_Flight_Delay_Prediction)
- [7] <https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/anova/>
- [8] <https://www.statisticssolutions.com/anova-analysis-of-variance/>

#### ACKNOWLEDGMENTS

The authors whole heartedly acknowledge and thank all the authors and publishers whose content has been referred and cited in this research paper.