Flights Dataset Report

By Alhassan Atama Isiaka

This report contains an analysis of domestic flights in US in 1989. The analysis in this report deals with a dataset of about 1,300,000 records and 21 columns.

We exploited the dataset to look at factors that cause Arrival delays based on Flights. Here we compared several kinds of variables: univariate, bivariate and multivariate data. We basically looked at the relationship between it an departure delay with the distance to see how it affects this delay among other observations

Summary of Findings

The dataset has records of the year 1987 that has a lot of missing information however, I had been able to establish some factors that causes delay of flights. The data has numeric, categorical, data which are spread across several data types: int, string, Boolean, etc.
The data has some noise which we had to clean-up before analyzing. We removed NULLS, duplicates and blank values from the dataset to reduce the noise in the data

Key Insights for Presentation

1. From findings, distance is a factor that affects the delay of Flights. A larger distance affects delay of flights. It is logical to think that those travelling to a very far location are less compared to those trying to assess a less proximal area.
2. Weather and other delay factors such as Late Aircraft delay, carrier delay, etc. would have been another interesting trend to look at in the data but there had been Null values in the column so we had to remove it
3. It is correct to say that travelling in the evenings have more delay than in the mornings
4. The delay increases when end of the year is approached. This is logical because most people go on holidays within that period