

Flight Data Exploration

Dataset:

The dataset reports flights in the United States, including carriers, arrival and departure delays, and reasons for delays. I have combined data from four different years: 2003,2004,2005 and 2006.

Gathering Data:

After downloading the respective csv files, I have read them into different pandas dataframes and finally combined them into a single dataframe using pandas 'concat' function.

Not much cleaning was required for the dataset. I have just made a couple of corrections in data types.

Exploration of Data:

Univariate Exploration:

1. I have chosen a histogram plot to depict the frequency of trips on a particular day of the week.
2. I have chosen a bar plot to show the most sought after flight carrier in the US.

Bivariate Exploration:

1. I have used a scatter plot to show the relation between Arrival delay and Distance to be covered. I have used the parameters of alpha and jitter for better understanding of the plot.
2. I have chosen a clustered bar chart to show the change in the factors for cancellations over multiple years.