REDICT FLIGHT DELAYS

Name: Issa Mohamed Issa Alasire
Email: asiry999@hotmail.com
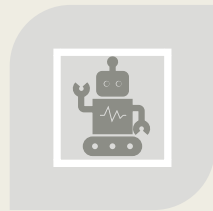Mobile Number: 0509418641

T5 Bootcamp Data Science Project
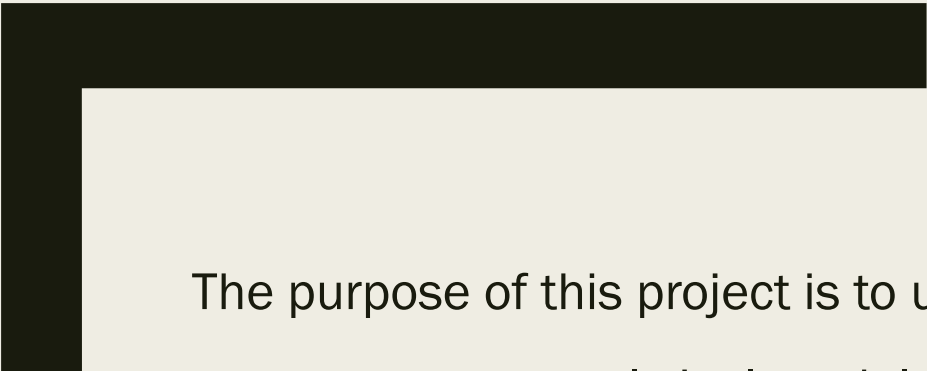
DESIGN
INTRODUCTION

DATA

PROCESS
DATA

VISUALIZATIONS

ALGORITHMS

TOOLS

Table of Content

The purpose of this project is to use publicly available data and data mining techniques to construct an analytical model to predict flight delays based on flight attributes such as origin, destination, date/time, distance, etc. Additional models will be created to determine the most likely cause of a flight delay and to predict the approximate length of the delay.

Design Introduction

Data for this project is available on Kaggle website. The database contains flight information, including scheduled and actual arrival times, in CSV-formatted data files. Each row in the data set represents an individual flight. Data files are available by month dating back to 1987. Each monthly file contains approximately 123520 observations.

The following attributes are included in each observation:

Year - Month - DayofMonth - DayOfWeek - DepTime - CRSDepTime - ArrTime - CRSArrTime

UniqueCarrier - FlightNum - TailNum - ActualElapsedTime - CRSElapsedTime -AirTime

ArrDelay - DepDelay - Origin - Dest - Distance - TaxiIn – TaxiOut - Cancelled - CancellationCode

Diverted - CarrierDelay - WeatherDelay - NASDelay - SecurityDelay - LateAircraftDelay

DATA

- Explore the features
- Drop any unnecessary features
- Apply the TfidfVectorizer
- Train and test a naive bayes model to classify the data
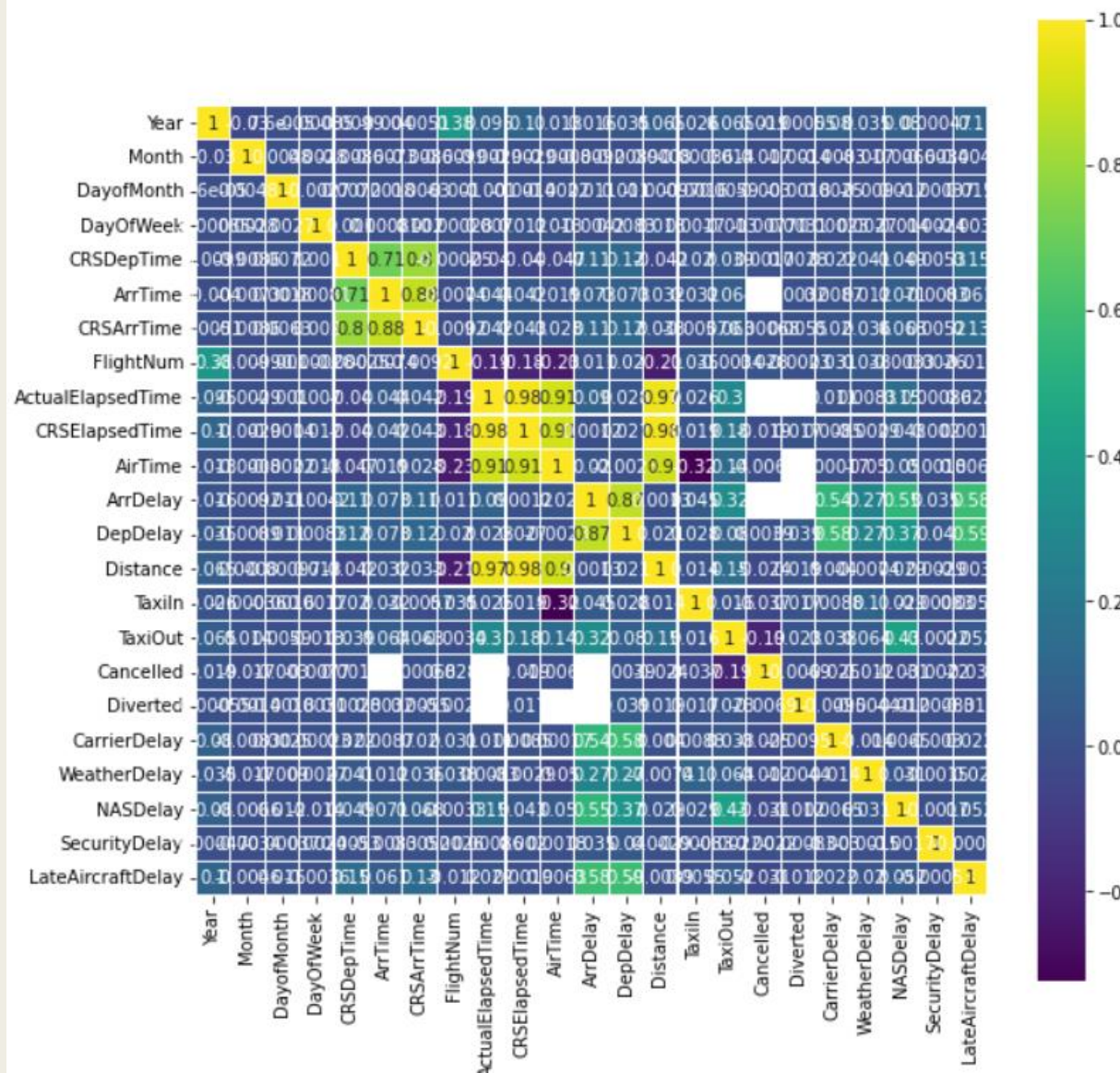- The following figure illustrates the ratio of AirTime to DepDelay:
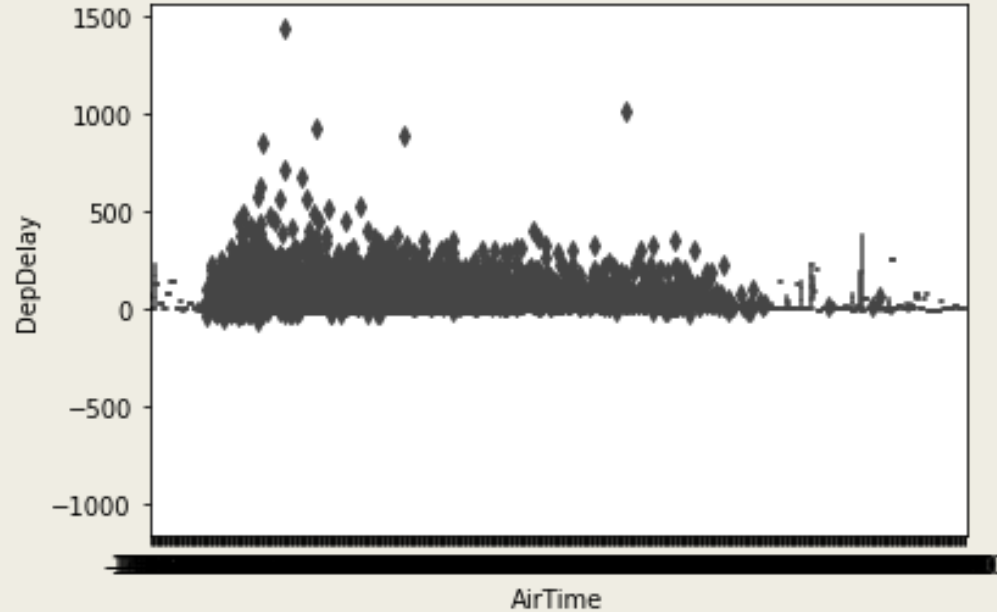
Process Data
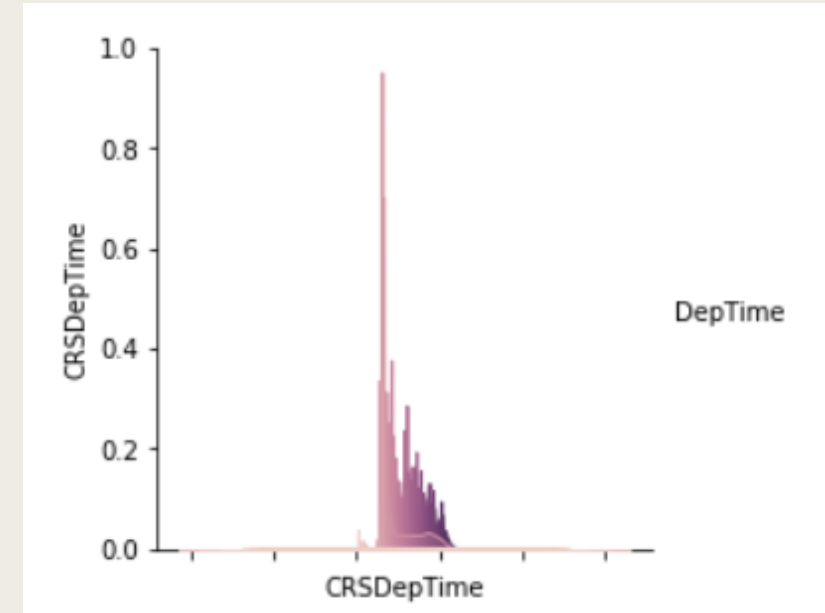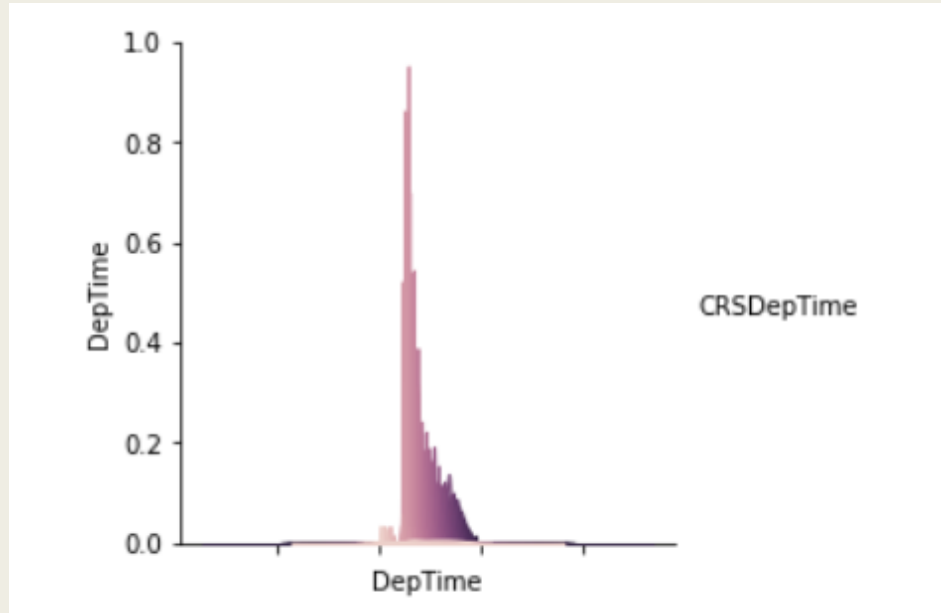
Visualizations

- Matplotlib
- Seaborn

Algorithms

Tools

Pandas for data manipulation

Scikit-learn for modeling

nltk for natural language processing

Matplotlib for plotting

streamlit for interactive visualizations

THANK YOU