Scientific study on the discipline of flight schedules

SOUTHWEST AIRLINES

JANUARY 2020

NAWAF JAMIL NAZER

In this project we will study the discipline in the flights schedule, and we will generate graphical report to represent it to the management

we will be using python as a programming language and data science libraries such as import pandas, numpy, matplotlib & seaborn

In this study we will depend on data provided from: kaggle.com

https://www.kaggle.com/akulbahl/covid19-airline-flight-delays-and-cancellations

WE WILL BE CONSIDERING ONLY JANUARY DATA FOR SOUTHWEST AIRLINES

This source provide important data for this study such as:

Flight Departure and Arrival Airport

Scheduled Departure and Arrival Time
Actual Departure and Arrival Time
Departure Delay
Scheduled Total Flight Time
Reason for the delay by Time (Carrier, Weather, Security or National Aviation)

Project Plan:

analyzing the requirement
gathering the data
understand and reading the data
cleaning the data
generate the reports

Used tools :

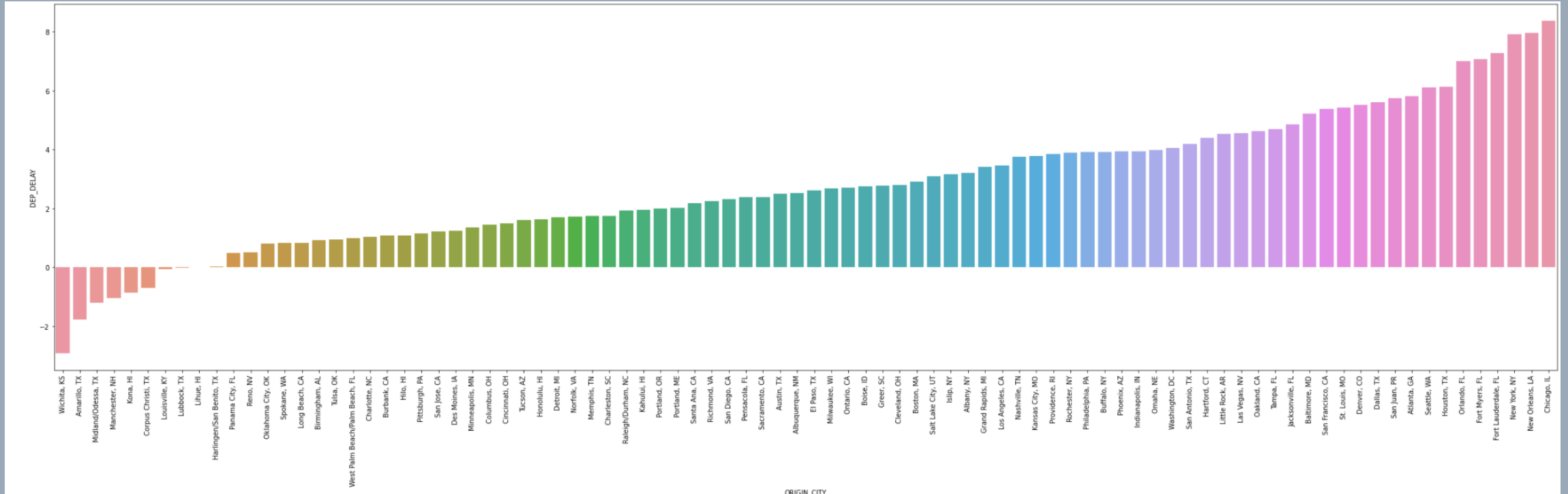Python as a programming language

pandas and numpy as a data frame librares

seaborn as a Python data visualization library based on
matplotlib

Matplotlib as a comprehensive library for creating static,
animated, and interactive visualizations in Python.
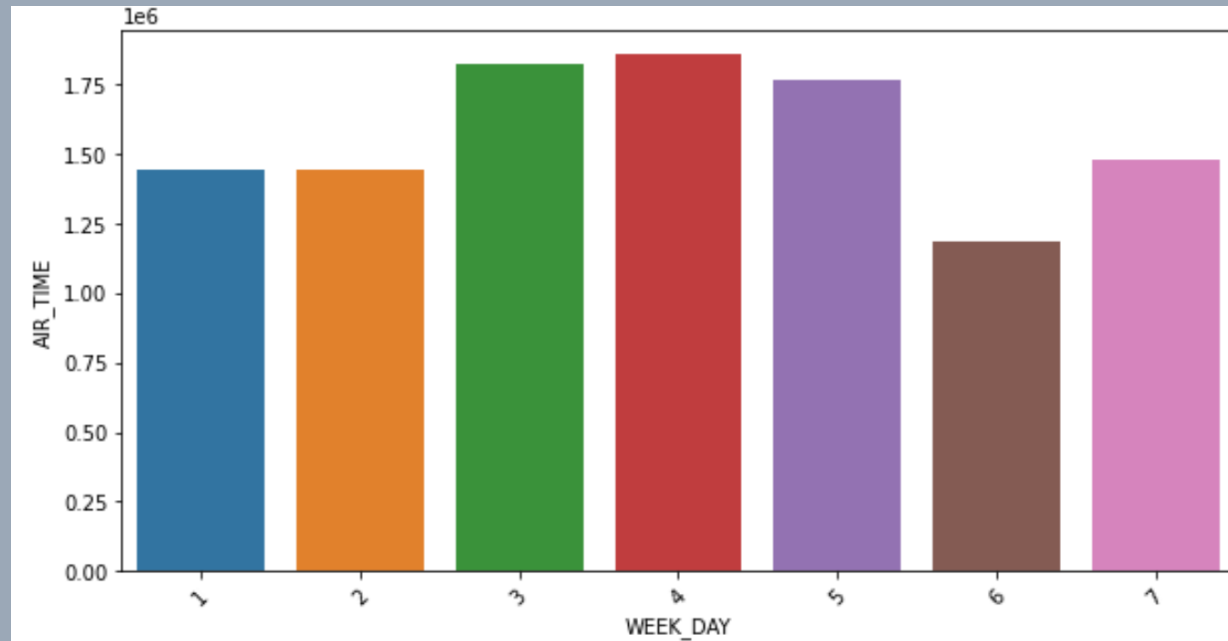
scikit-learn : tools for predictive data analysis
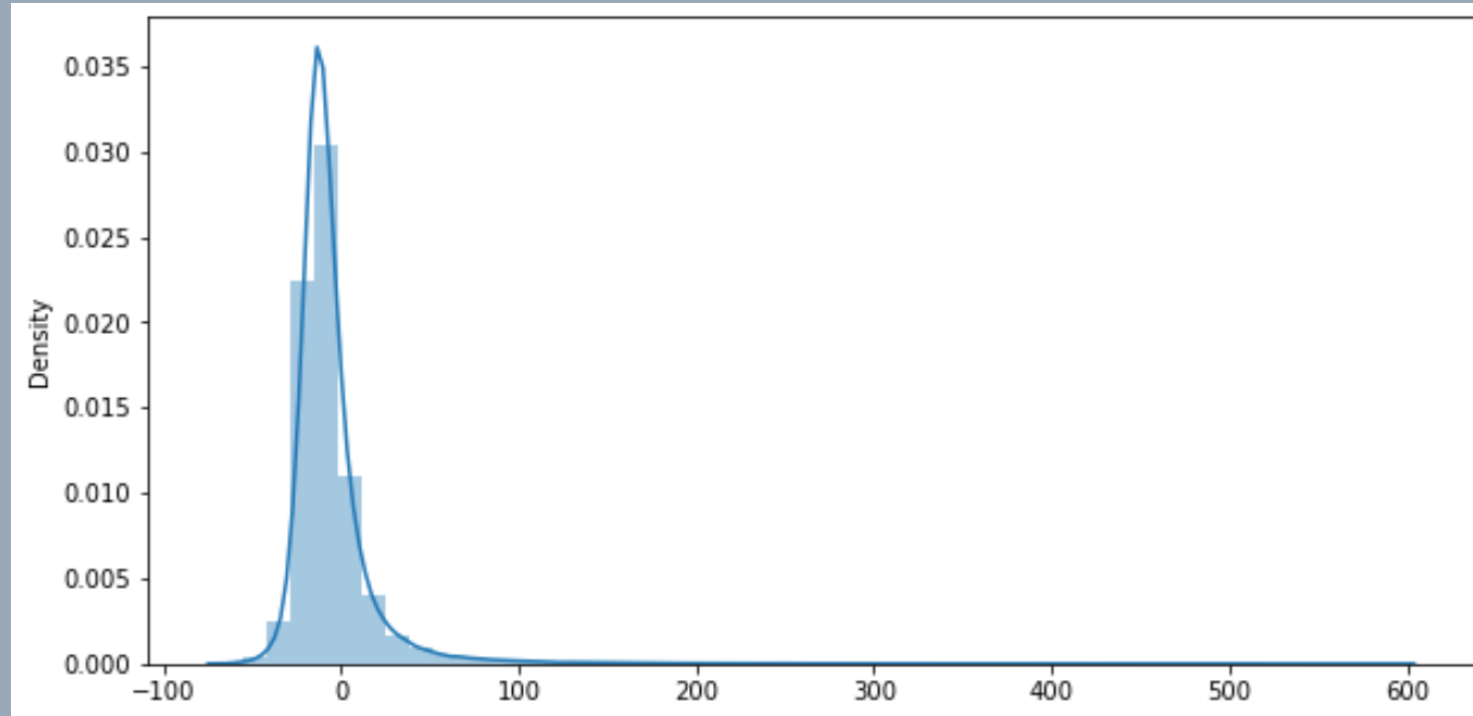
Jupyter as a text editor

Report Sample

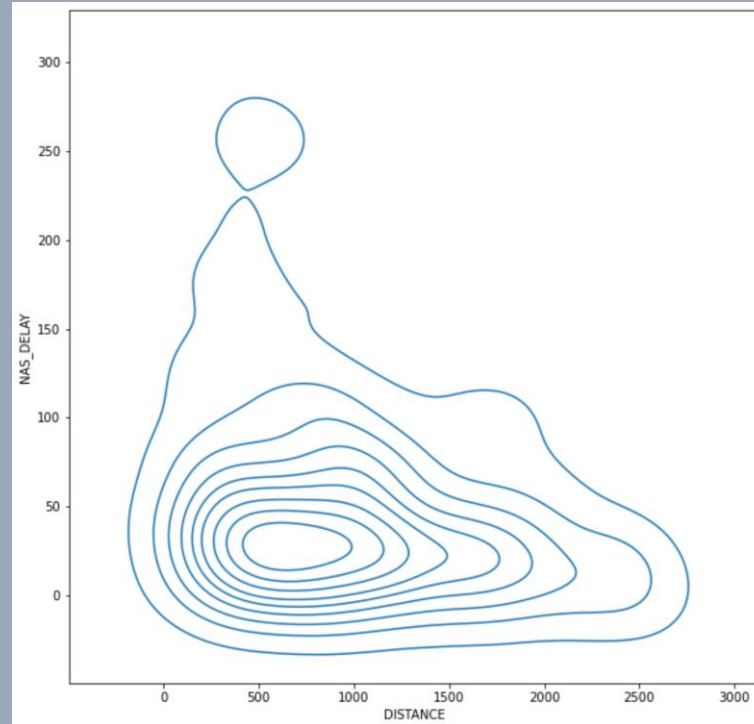THE DELAY DEPARTURE TIME FOR EACH CITY, THE MINUS NUMBER MEANS EARLY DEPARTURE

Report Sample

THE NUMBER OF FLIGHTS TIME IN EACH DAY OF THE WEEK (AIR TRAFFIC)
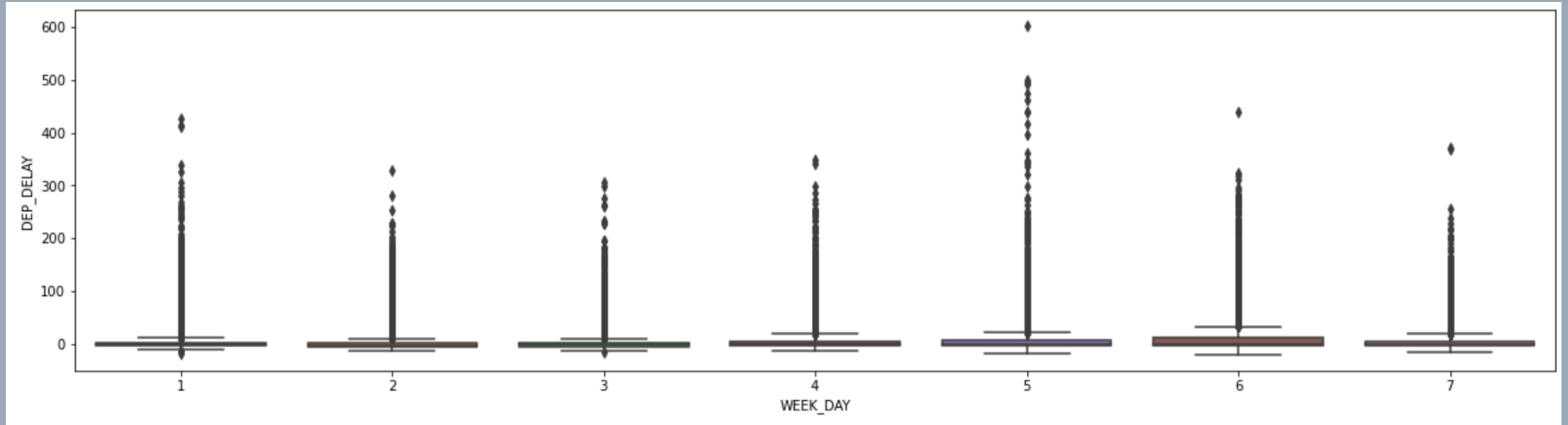
Report Sample

DIFFERENCE BETWEEN THE SCHEDULED ACCESS TIME AND THE ACTUAL TIME
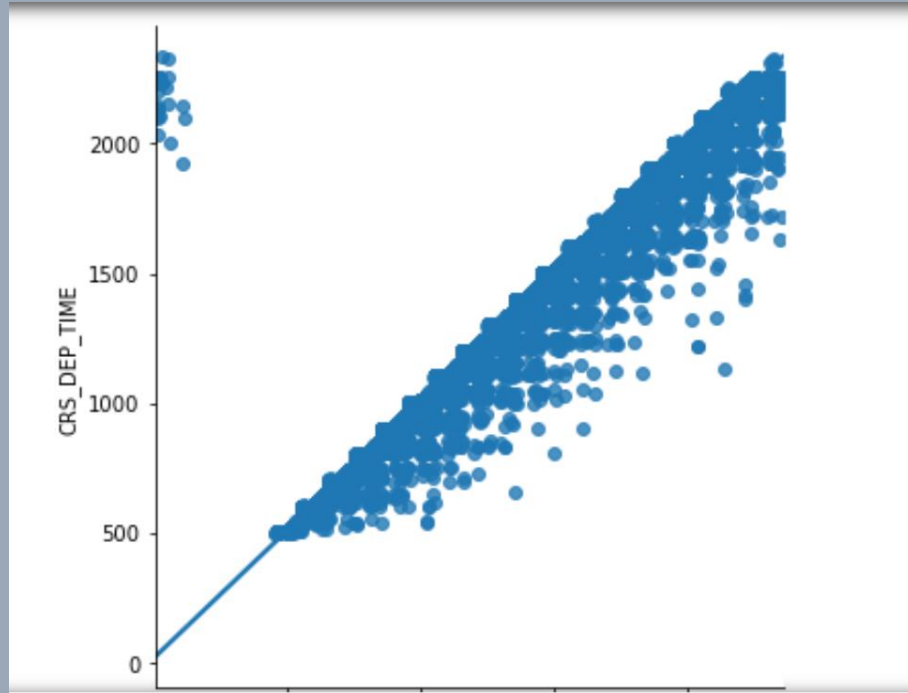
Report Sample

THE RELATIONSHIP BETWEEN FLIGHT DELAYS DUE TO CIVIL
AVIATION AND THE DISTANCE BETWEEN AIRPORTS

Report Sample

THE DEPARTIAL DELAY FOR EACH DAY OF THE WEEK

Report Sample


```python
# import model and fit
from sklearn.metrics import r2_score
from sklearn.linear_model import LinearRegression

# grab the DEP_TIME and CRS_DEP_TIME
#define the independent column
X1 = df.DEP_TIME.values.reshape(-1, 1)
#then the dependent column?
y1 = df.CRS_DEP_TIME.values.reshape(-1, 1)

#1 eate an object to Linear Regression model formula
linreg_model = LinearRegression()

#2 tting the model == Train Model
linreg_model.fit(X1, y1)
```

```
LinearRegression()
```

```python
round(linreg_model.score(X1, y1)*100)
```

```
99
```

```python
import numpy as np
x2 = np.array(100)
x2 = x2.reshape(-1,1)

linreg_model.predict(x2)
```

```
array([[117.81314602]])
```

MODELING WITH SKLEARN TO KNOW THE DISCIPLINE IN TAKE-OFF TIMES