 Project Report Titles

# 1 INTRODUCTION

1.1 Overview

A brief description about your project

1.2 Purpose

The use of this project. What can be achieved using this.

# 2 LITERATURE SURVEY

2.1 Existing problem

Existing approaches or methods to solve this problem

2.2 Proposed solution

What is the method or solution suggested by you?

# 3 THEORETICAL ANALYSIS

3.1 Block diagram

Diagrammatic overview of the project.

3.2 Hardware / Software designing

Hardware and software requirements of the project

1. **EXPERIMENTAL INVESTIGATIONS**

Analysis of the investigation made while working on the solution.

1. **FLOWCHART**

Diagram showing the control flow of the solution

1. **RESULT**

Final findings (Output) of the project along with screenshots.

**7 APPLICATIONS**

The areas where this solution can be applied

**8 CONCLUSION**

Conclusion summarizing the entire work and findings.

# 9 FUTURE SCOPE

Enhancements that can be made in the future.

 Project Report Titles

# 10 BIBILOGRAPHY

References of previous works or websites visited/books referred for analysis about the project, solutions previous findings, etc.

# APPENDIX

A. Source Code

Attach the code for the solution built.

**Flight Delays Prediction Using Machine Learning**

**1 Introduction**

* 1. **Overview**

Over the last twenty years, air travel has been increasingly preferred among travelers mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. An increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and the air. These delays are responsible for large economic and environmental losses The main objective of the model is to predict flight delays accurately to optimize flight operations and minimize delays.

* 1. **Purpose The use of this project.**

Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vectors like departure date, departure delay, the distance between the two airports, scheduled arrival time, etc. We then use a decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when the difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare the decision tree classifier with logistic regression and a simple neural network for various figures of merit.

**2 LITERATURE SURVEY**

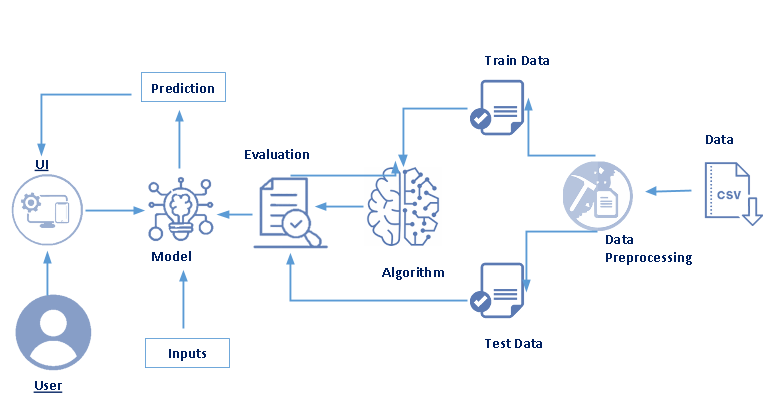
**2.1 Existing problem**

* You’ll be able to understand the problem to classify if it is a regression or a classification kind of problem.
* You will be able to know how to pre-process/clean the data using different data pre-processing techniques.
* You will be able to analyze or get insights into data through visualization.
* Applying different algorithms according to the dataset and based on visualization.
* You will be able to know how to build a web application using the Flask framework.

**2.2 Proposed solution**

* **Data Collection.**
  + Collect the dataset
* **Data Pre-processing.**
  + Import the Libraries.
  + Importing the dataset.
  + Checking for Null Values.
  + Data Visualization.
  + Taking care of Missing Data.
  + Label encoding.
  + One Hot Encoding.
  + Feature Scaling.
  + Splitting Data into Train and Test.
* **Model Building**
  + Training and testing the model
  + Evaluation of Model ( Decision Tree Classification)
* **Application Building**
  + Create an HTML file
  + Build a Python Code

**3 THEORETICAL ANALYSIS**

**3.1 Block diagram Diagrammatic overview of the project**

**3.2 Software designing**

we will be using

**Jupyter** notebook

**Spyder**

because it is a free and open-source distribution of the Python for data science and machine learning related applications

* Flask (Web applications)

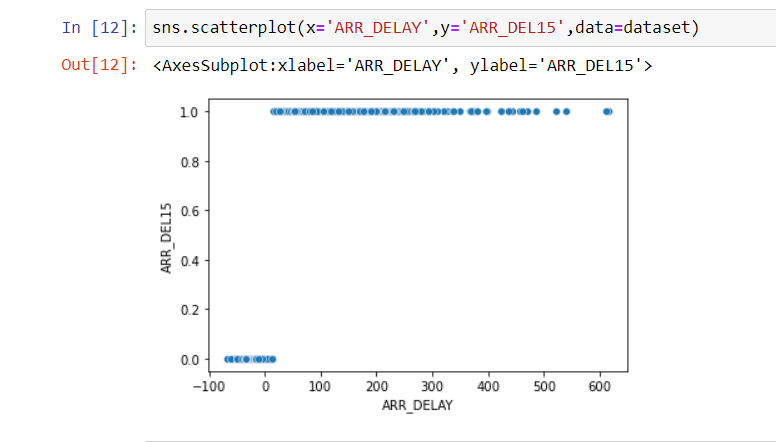
**Hardware:**

* **Processor:** Processor Intel CORE i5 and above Internet.
* **System architecture :** Windows- 64-bit x86, 32-bit x86; MacOS- 64-bit x86; Linux- 64-bit x86, 64-bit aarch64 (AWS Graviton2 / arm64), 64-bit Power8/Power9, s390x (Linux on IBM Z & Linux ONE).
* **RAM**:4 GB or above.

**4. Analysis or the investigation made while working on the solution.**

Scatterplot

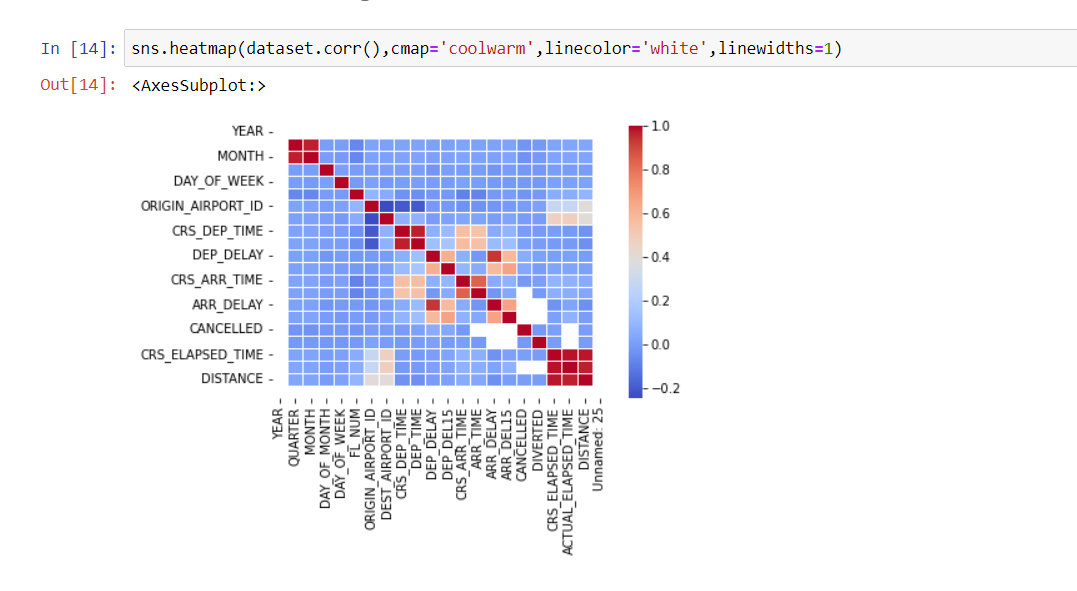
A scatter plot (also called a scatterplot, scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data.



From this scatterplot, comparing the two columns we can see many flights were delayed from their arrival time.

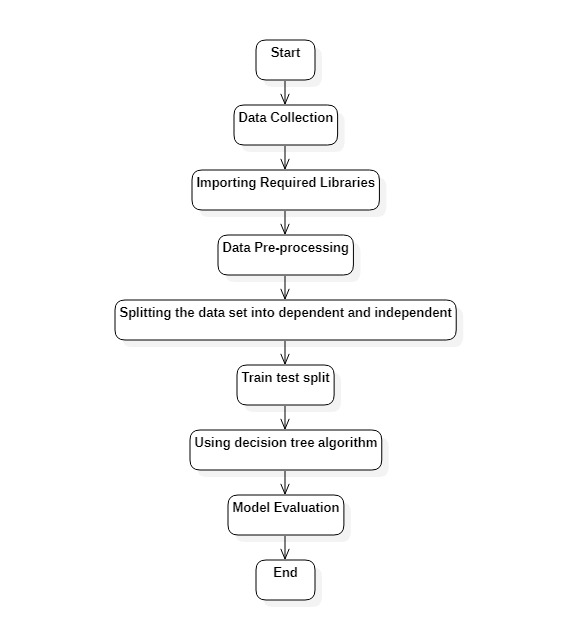
Heatmap

**Heatmap** is defined as a graphical representation of data using colors to visualize the value of the matrix. In this, to represent more common values or higher activities brighter colors and reddish colors are used, and to represent less common or activity values, darker colors are preferred.



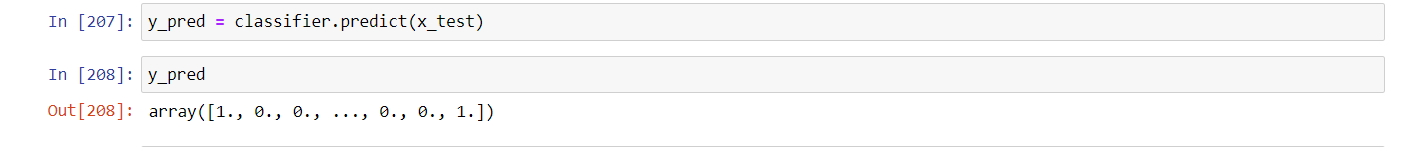
If you observe the heatmap, the lighter the color the correlation between the two variables will be high. And correlation plays a very important role in extracting the correct features for building our model.

**5 Flow Chart**



**6. Result**

**Result final findings (output) of the project along with screenshots.**

****

**Decision Tree Model Accuracy**



**Confusion Matrix**

****

**7 Application**

Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. An accurate estimation of flight delay is critical for airlines because the results can be applied to increase customer satisfaction and the incomes of airline agencies.

**8 CONCLUSION**

* By measuring the performance of the models using real data, we have seen interesting results on the predictability of the delays.
* This is the main page of Prediction of Flight Delay .where you may know about the inputs.
* The prediction page user gives the input for predicting the output where they can give input as Flight Number, Month, Day of Month, Week, Origin, Destination, Schedule Departure Time, Schedule Arrival Time, Actual Departure Time then click to submit the output.
* On the prediction page, the user will get the output based on the inputs they were given on the prediction page.

**9 Future Scope**

There have been many kinds of research on modeling and predicting flight delays, where most of them have been trying to predict the delay through extracting important characteristics and most related features. However, most of the proposed methods are not accurate enough because of the massive volume of data, dependencies, and an extreme number of parameters. This paper proposes a model for predicting flight delays based on Deep Learning (DL).DL is one of the newest methods employed in solving problems with a high level of complexity and a massive amount of data. Moreover, DL is capable of automatically extracting the important features from data. Furthermore, because most flight delay data are noisy, a technique based on stack denoising autoencoder is designed and added to the proposed model. Also, the Levenberg-Marquart algorithm is applied to find weight and bias proper values, and finally, the output has been optimized to produce highly accurate results. To study the effect of stack denoising autoencoder and LM algorithm on the model structure, two other structures are also designed. The first structure is based on autoencoder and LM algorithm (SAE-LM), and the second structure is based on denoising autoencoder only (SDA). To investigate the three models, we apply the proposed model to the U.S flight dataset which is the imbalanced dataset. To create a balanced dataset, the undersampling method is used. We measured the precision, accuracy, sensitivity, recall, and F-measure of the three models in two cases. The accuracy of the proposed prediction model was analyzed and compared to the previous prediction method. results of three models on both imbalanced and balanced datasets show that precision, accuracy, sensitivity, recall, and F-measure of the SDA-LM model with the imbalanced and balanced dataset is an improvement in SAE-LM and SDA models. The results also show that the accuracy of the proposed model in forecasting flight delay on imbalanced and balanced datasets respectively has greater than the previous model called RNN.

**10 Biblography**

**References**

E.R. Mueller and G.B. Chatterji. “Analysis of aircraft arrival anddeparture delay characteristics”. In: Proceedings of the AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference, Los Angeles, CA. 2002.

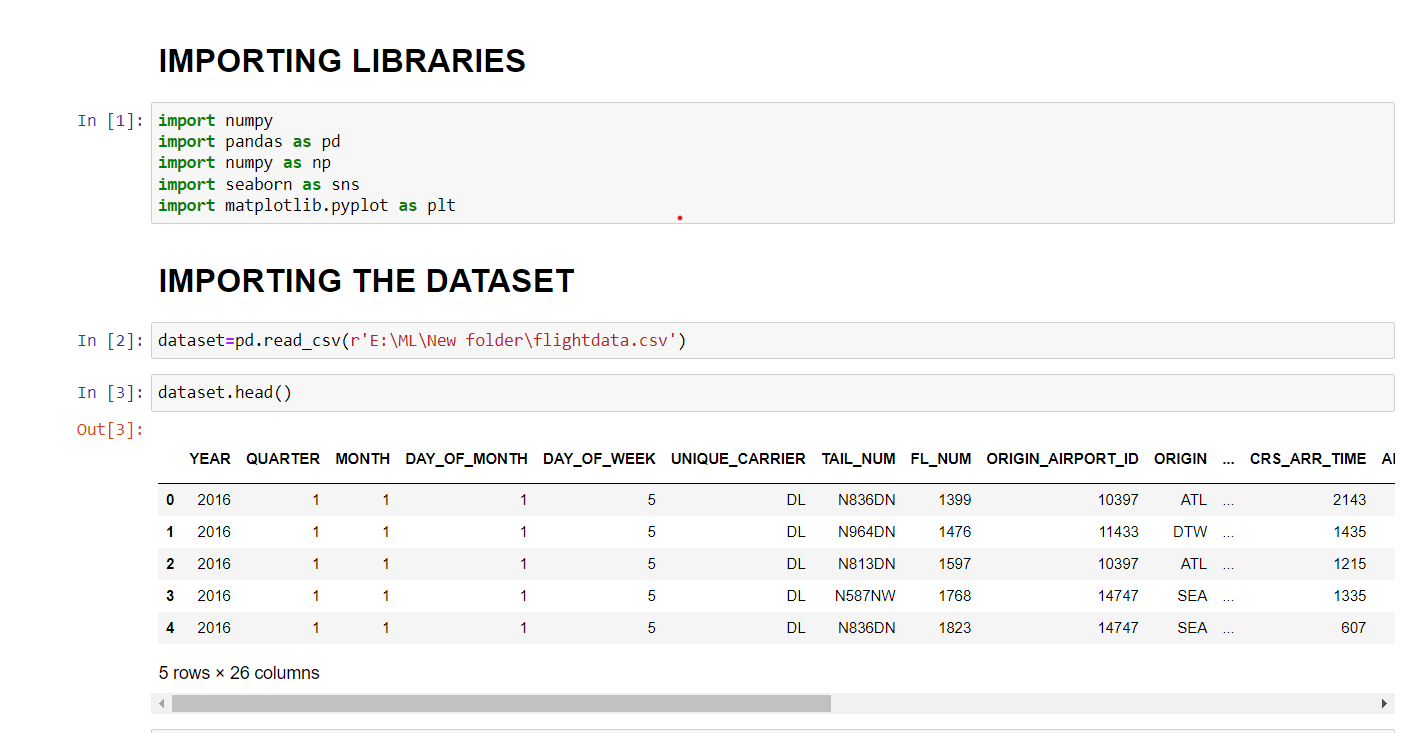
SS Allan et al. “Analysis of delay causality at Newark InternationalAirport”. In: 4th USA/Europe Air Traffic Management R&D Seminar. 2001.

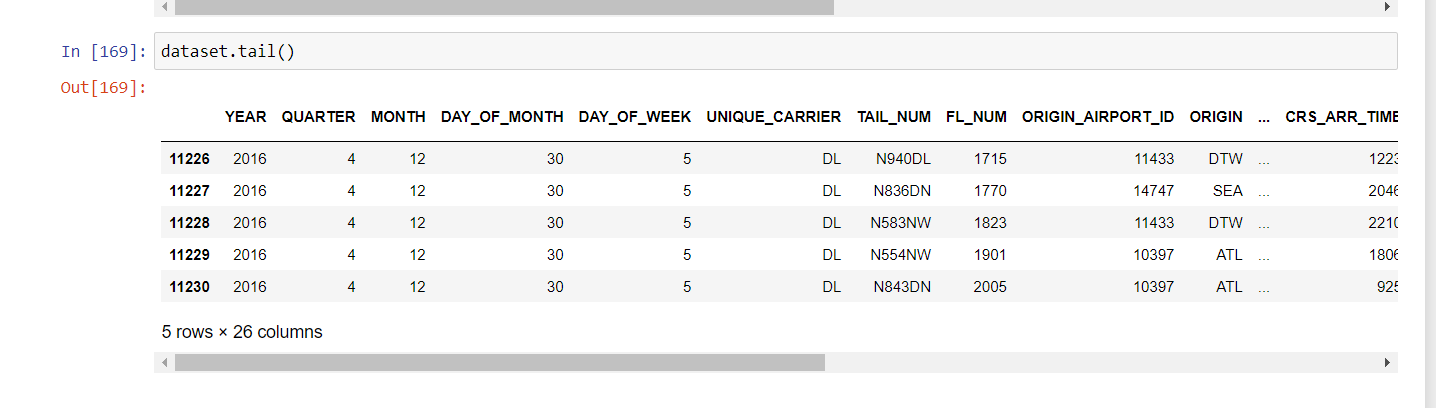
Lu Zonglei, Wang Jiandong, and Zheng Guansheng. “A new method toalarm large scale of flights delay based on machine learning”. In: Knowledge Acquisition and Modeling, 2008.

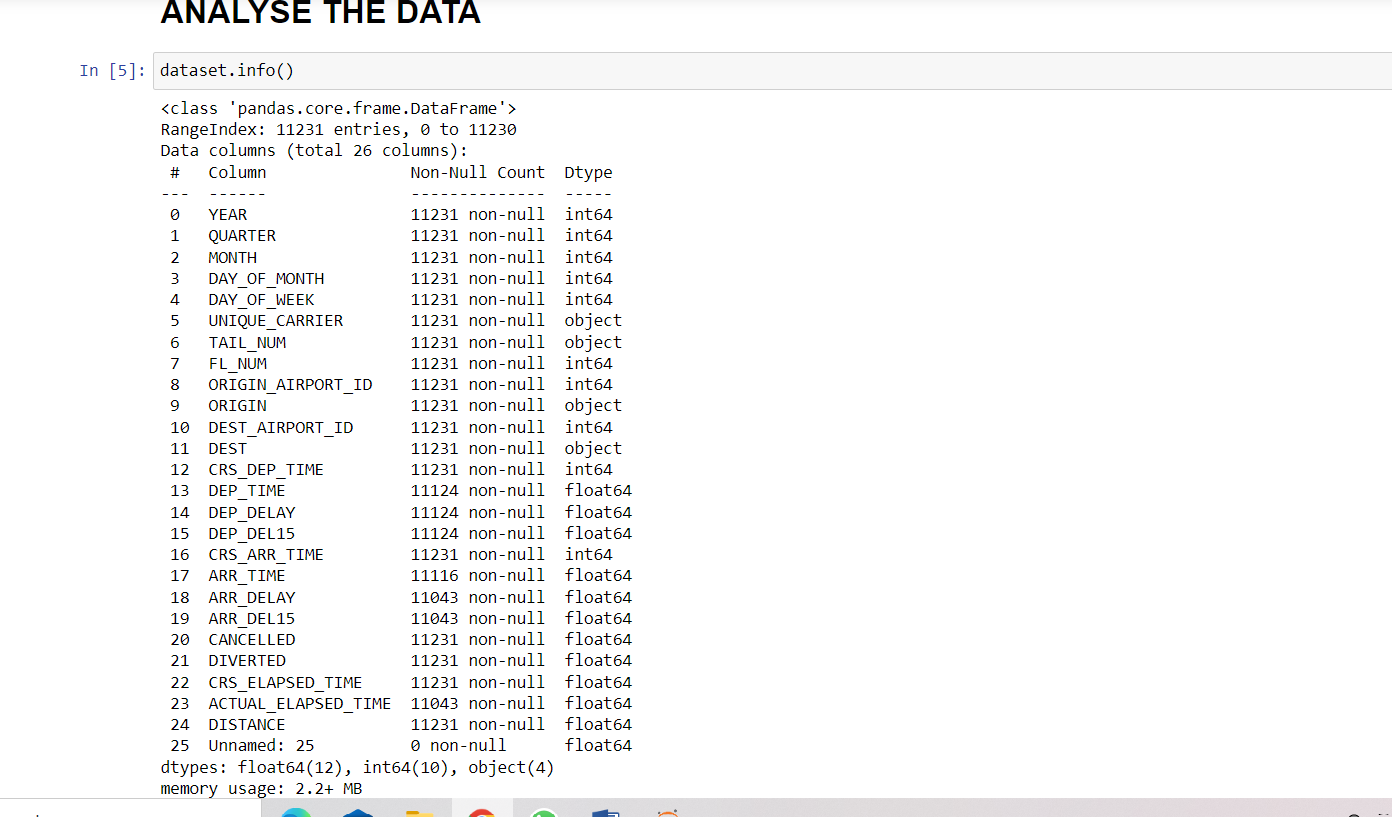
Y. Tu, M.O. Ball, and W.S. Jank. “Estimating flight departure delaydistributions – a statistical approach with the long-term trend and short-term pattern”. In: Journal of the American Statistical Association 103.481 (2008), pp. 112–125.

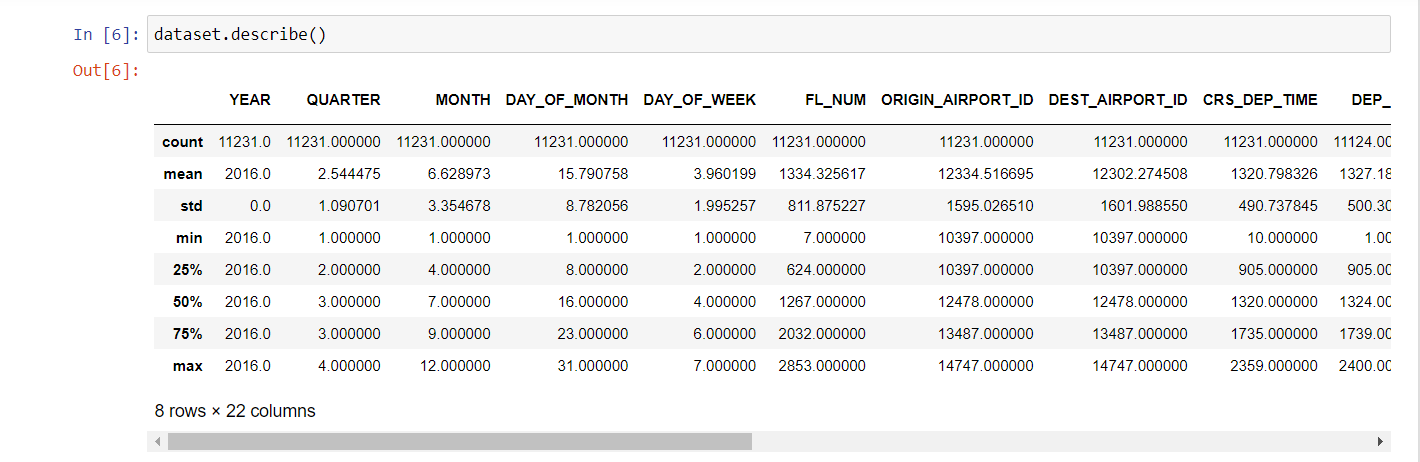
B.W. Silverman. Density estimation for statistics and data analysis. Vol. 26. Chapman & Hall/CRC, 1986.

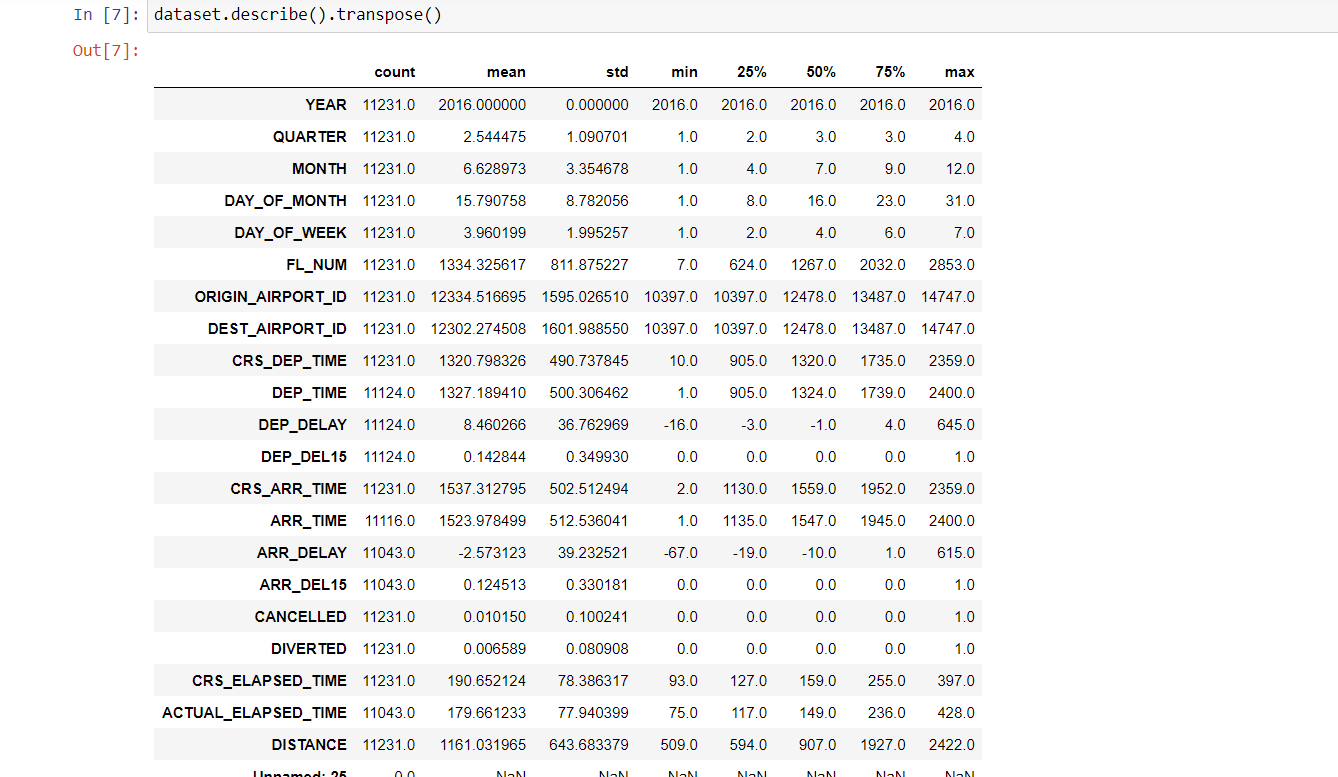
**Appendix**

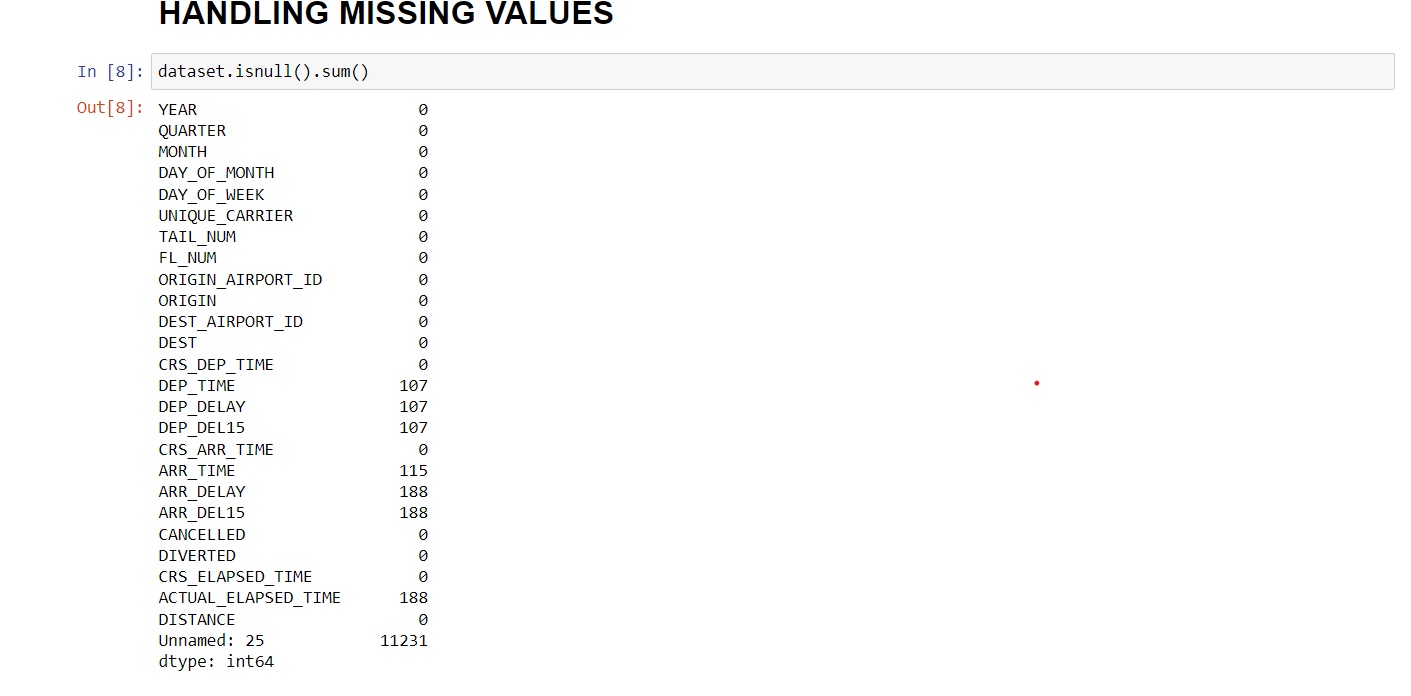
****

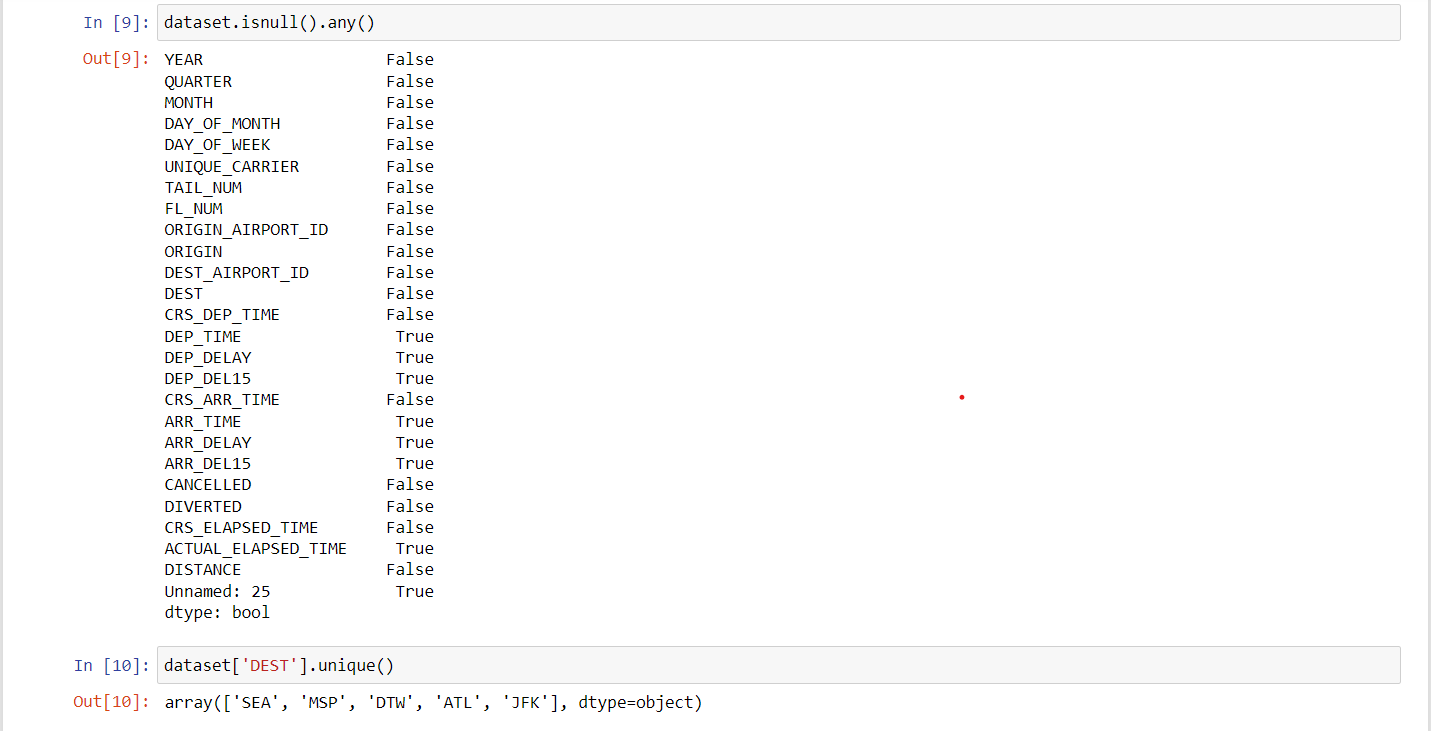
****

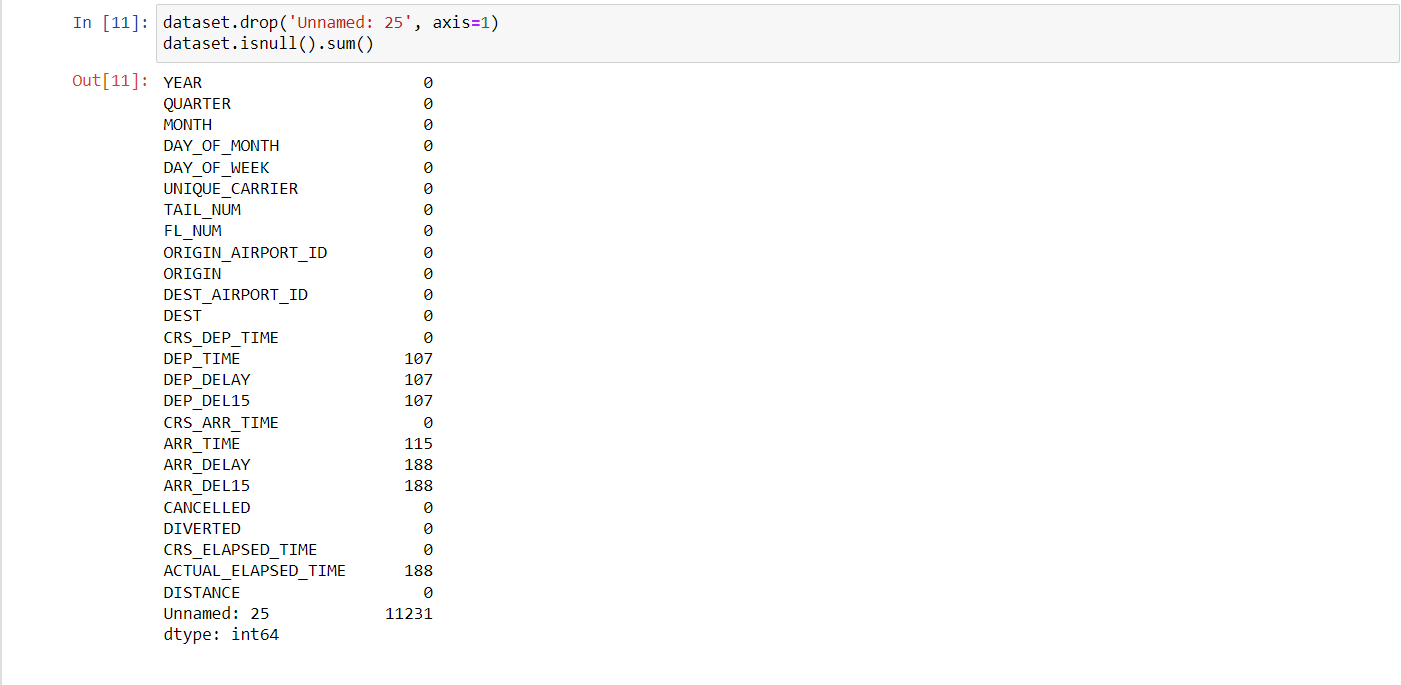
****

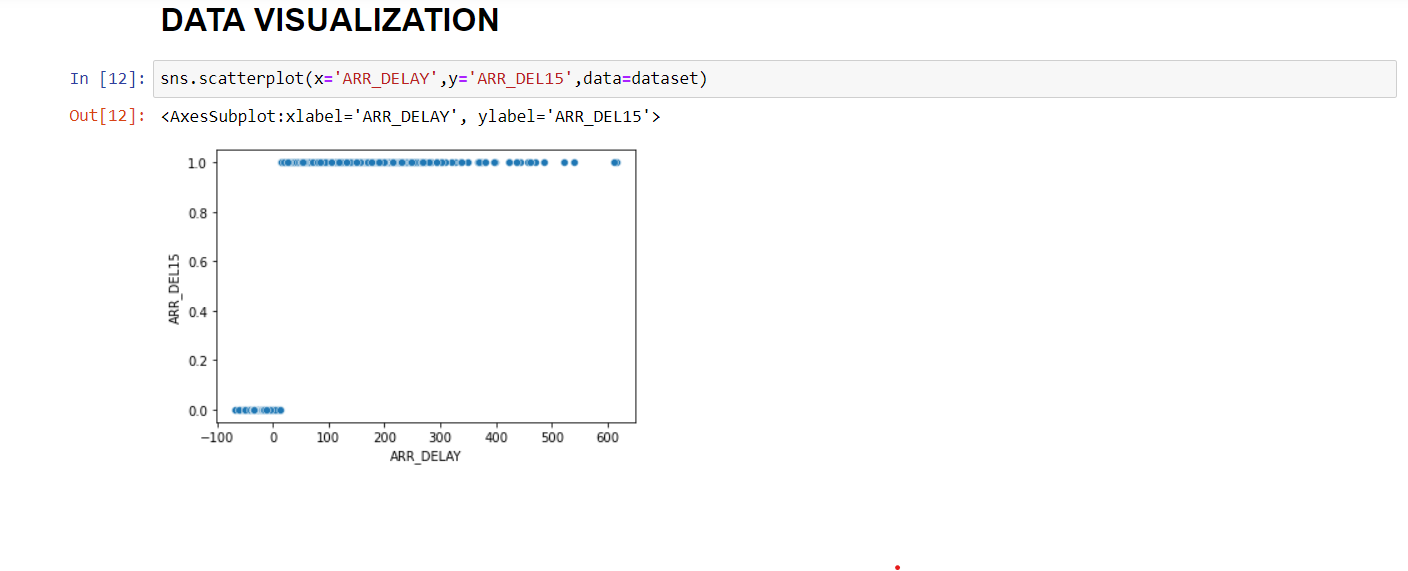
****

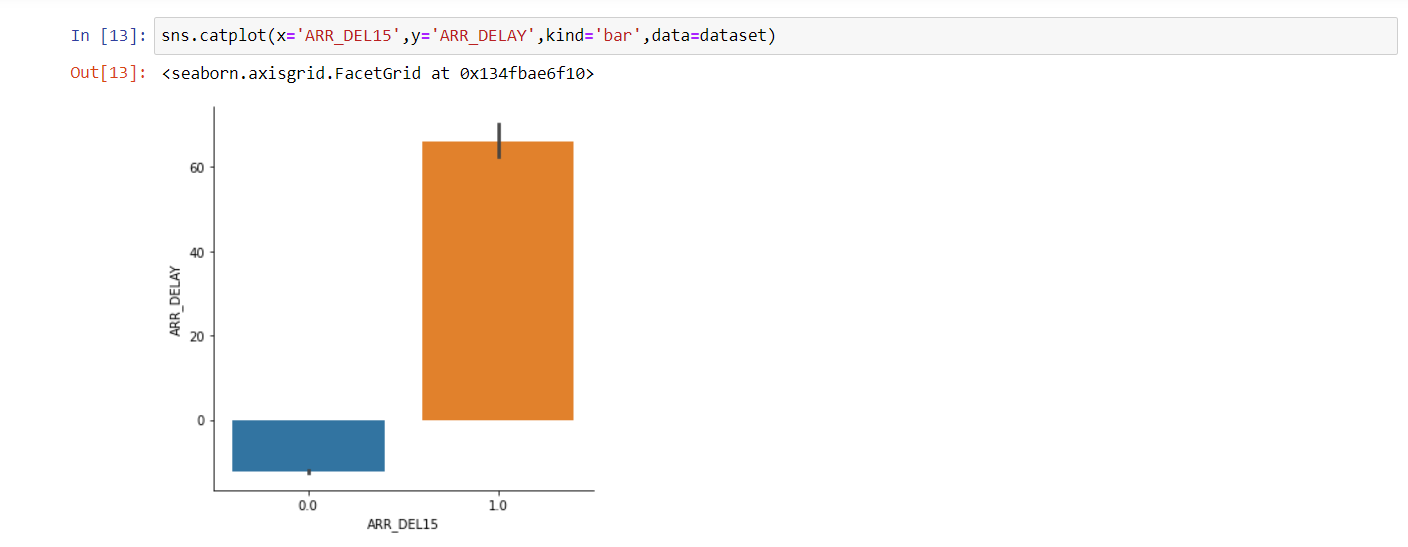
****

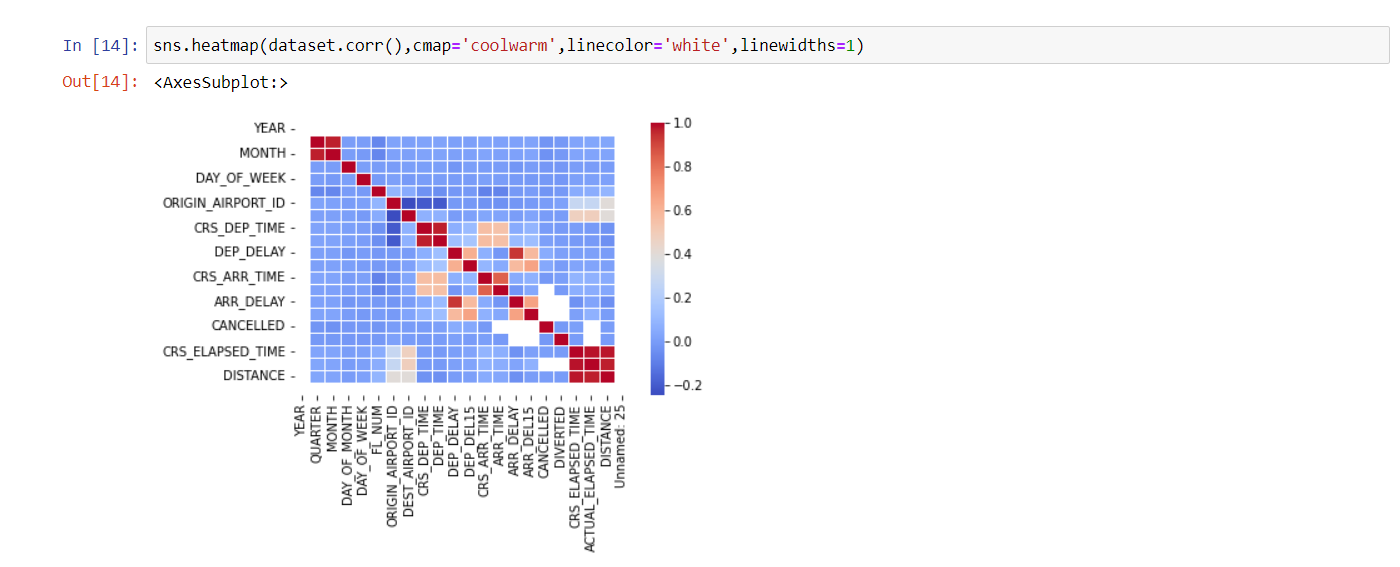
****

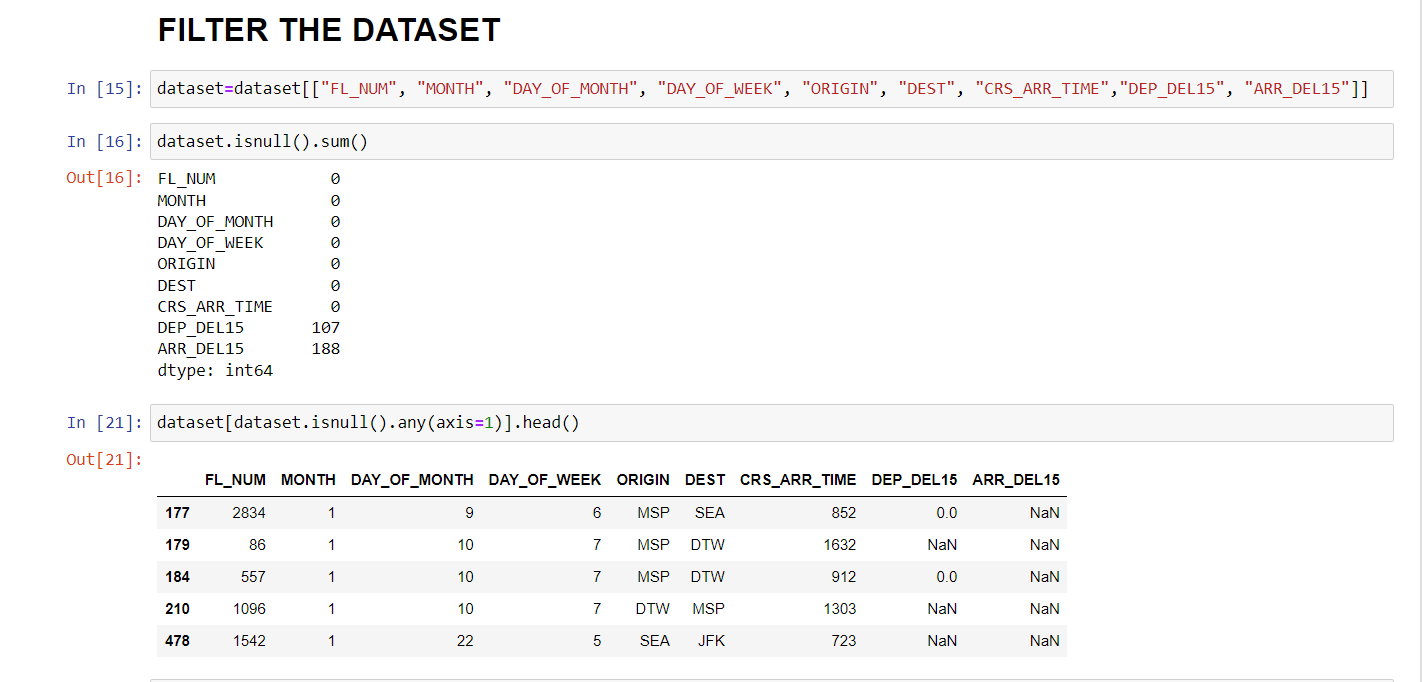
****

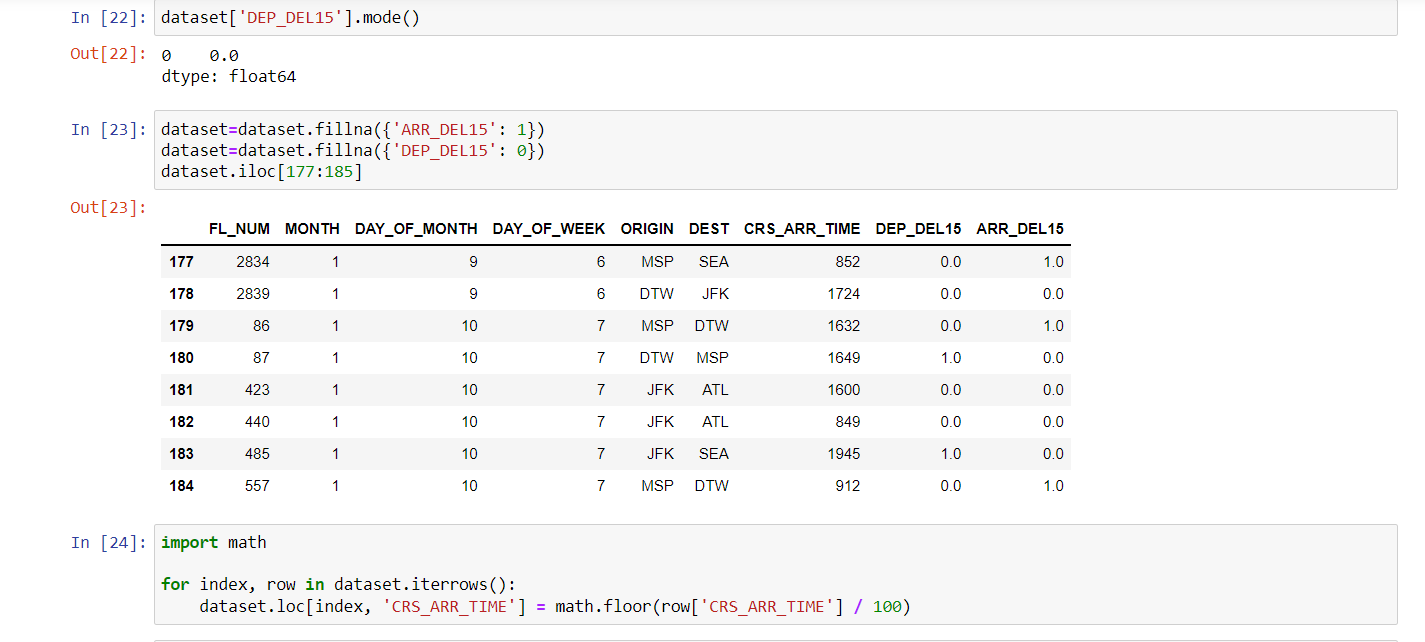
****

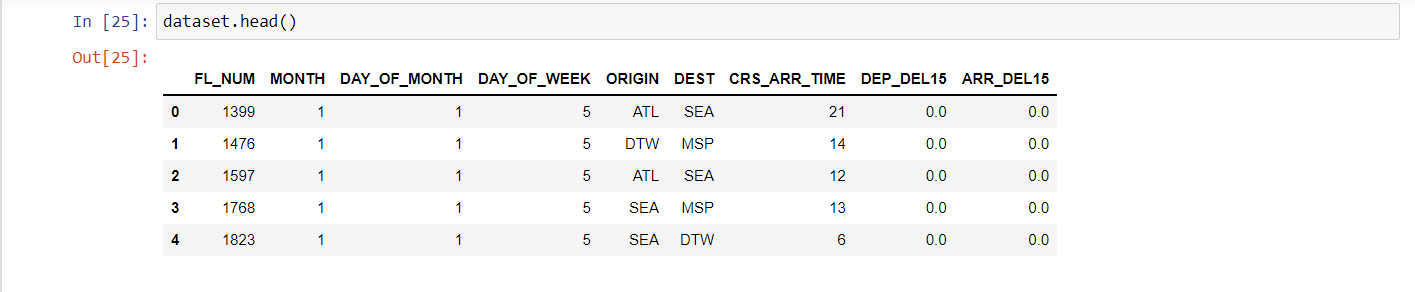
****

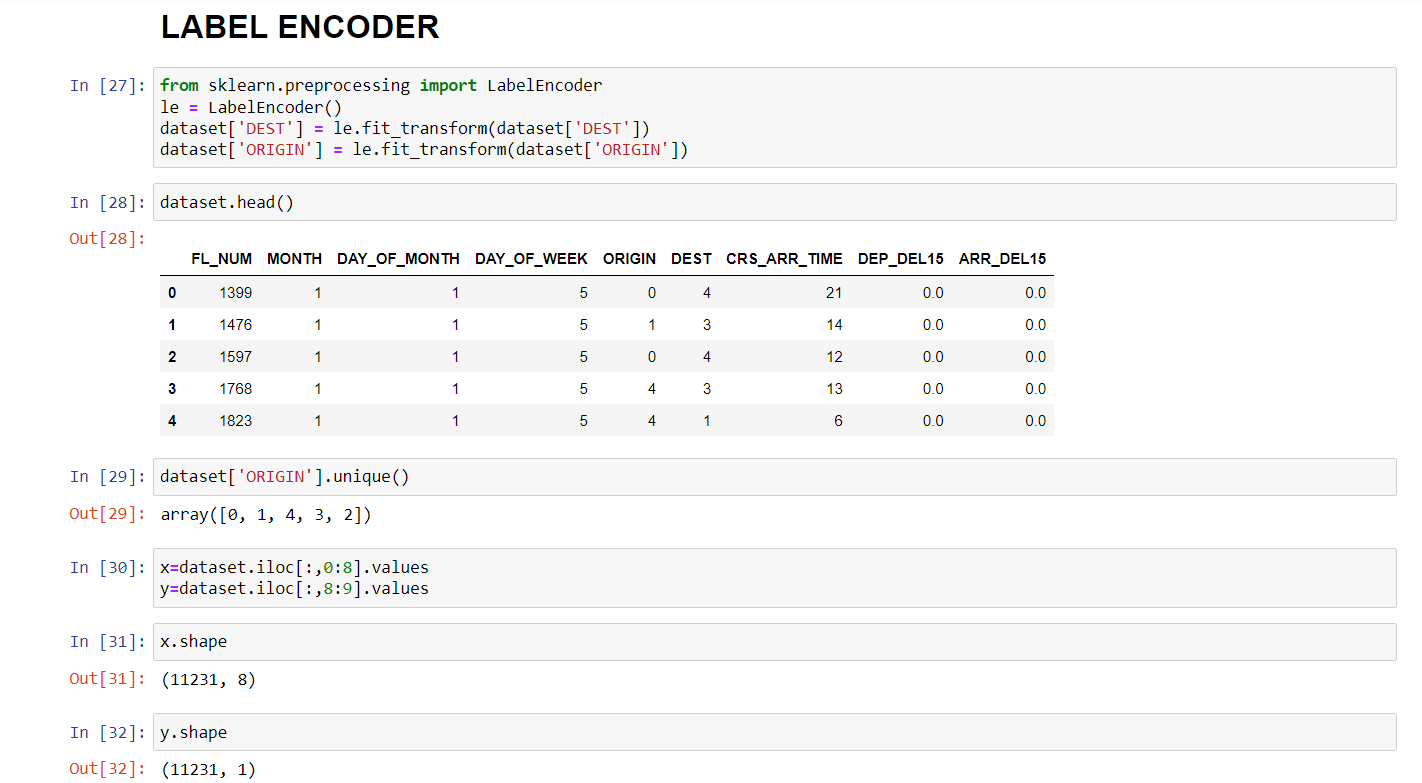
****

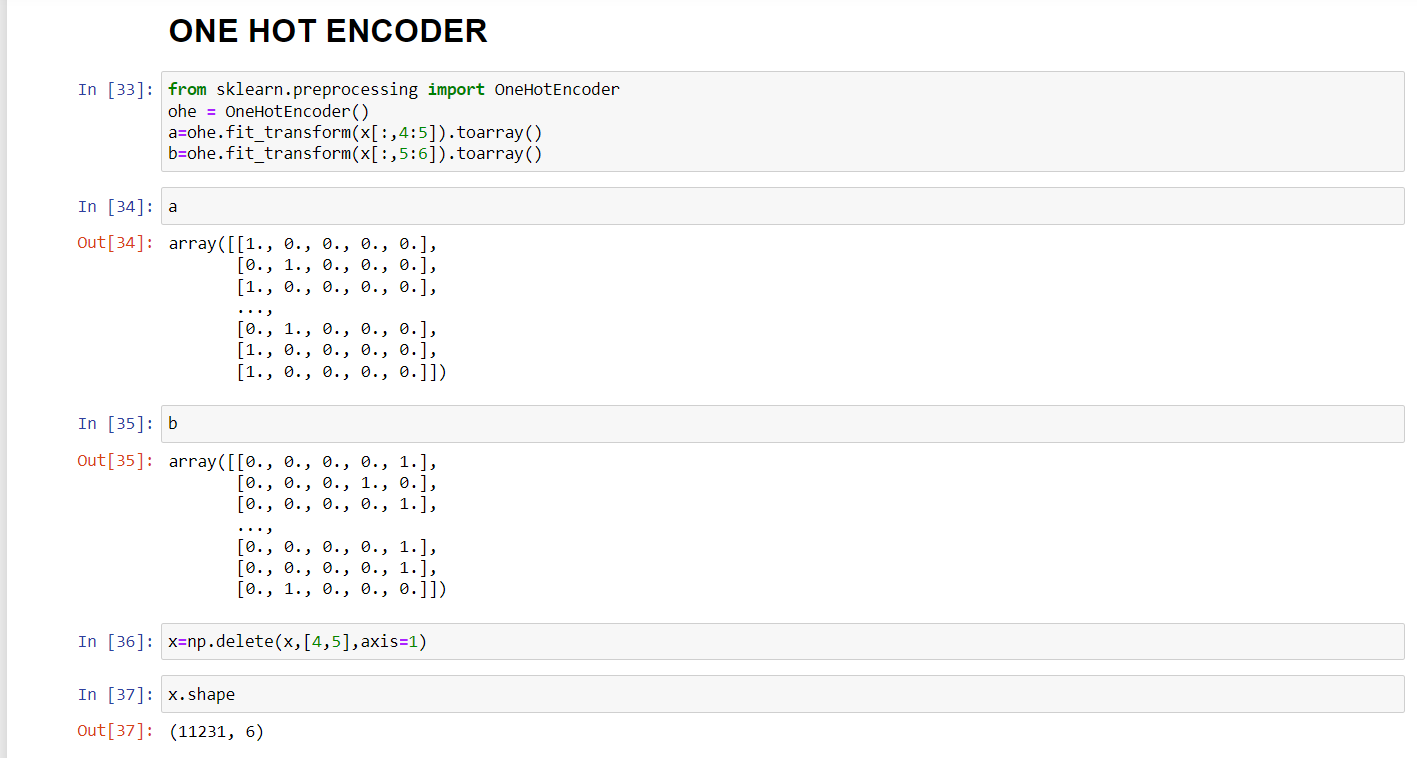
****

****

****

****

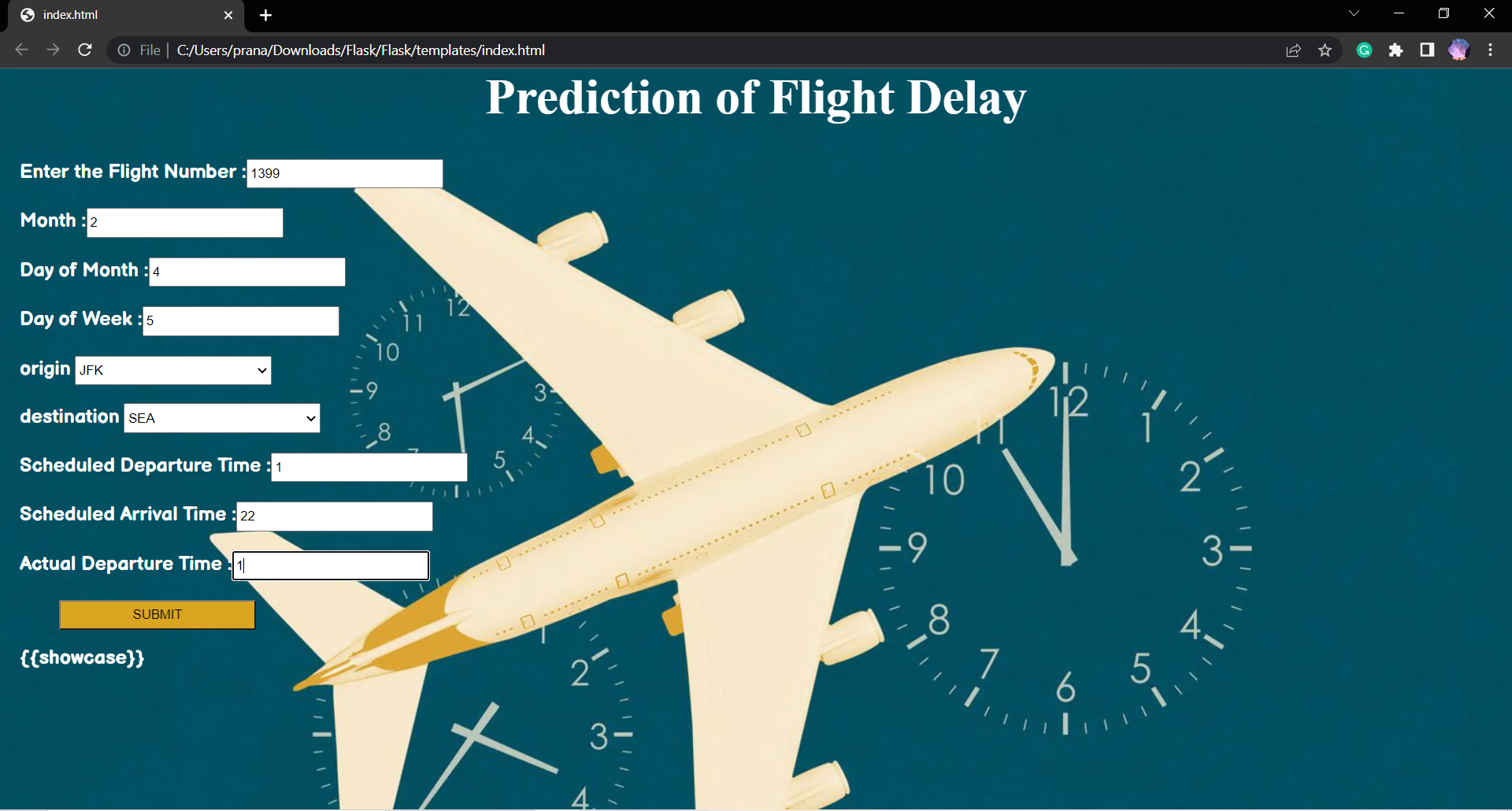
****

****

****

****

****

****

****