

Applied Data Science Capstone

Peer-graded Assignment

► Introduction

The Applied Data Science Capstone project was carried out as a requirement for the Coursera-based IBM Professional Certification in Data Science. The research involved using real-world datasets and doing data extraction, cleaning, exploration, and analysis in order to uncover important insights.

► Data Collection

The practise of obtaining information or data from various sources for analysis and decision-making is known as data collection.

- •Datasets are collected from Rest SpaceX API and webscrapping Wikipedia
- •The information obtained by the API are rocket, launches, payloadinformation.
- •The Space XREST API URL is api.spacexdata.com/v4/
- •The information obtained by the webscrapping of Wikipedia are launches, landing, payloadinformation.
- •URL
[ishttps://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922](https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)

► Data Wrangling

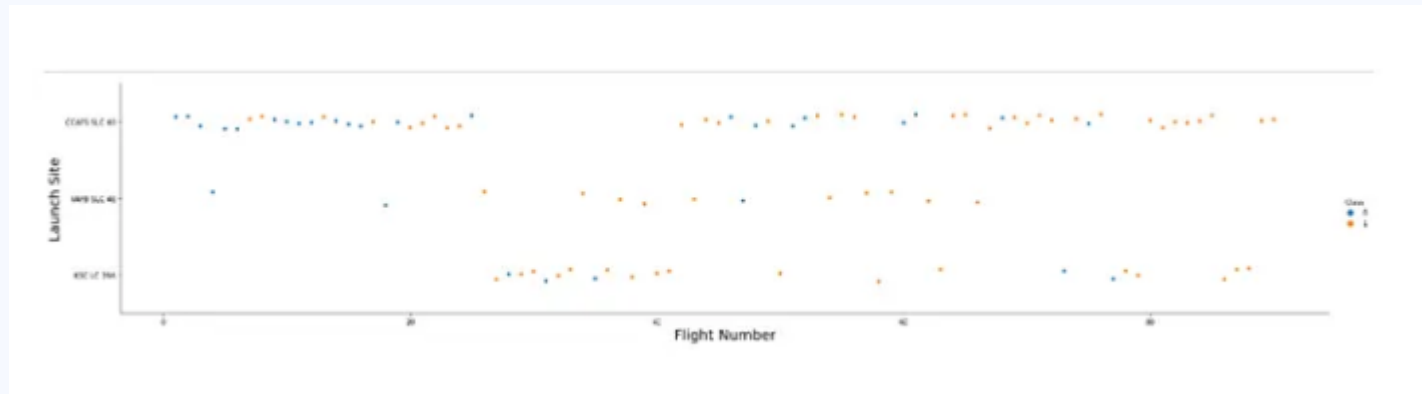
After collecting the data we check the missing data ,and data types and do on of the following to clean the data :

- *Replace the missing data with one-Using mean or so.*
- *Change data type of the data*
- *Represent categorical data using integer or float dummy numbers -one hot encoding.*

► Exploratory Data Analysis

After Data cleaning the we can proceed to Analyzing the data using visualization to get some insights of the launches .

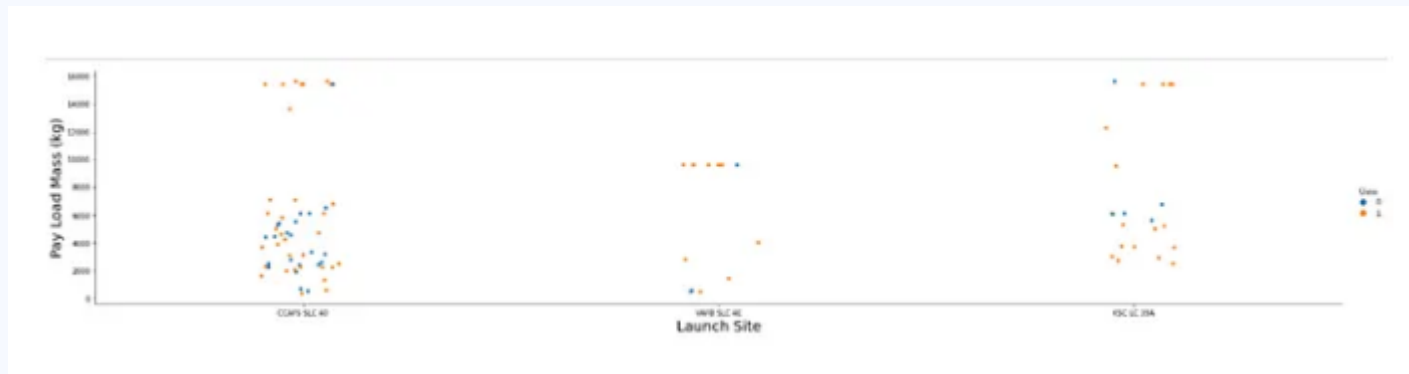
► Here are some of the screenshots



► Exploratory Data Analysis

From the Visualization we can conclude that:

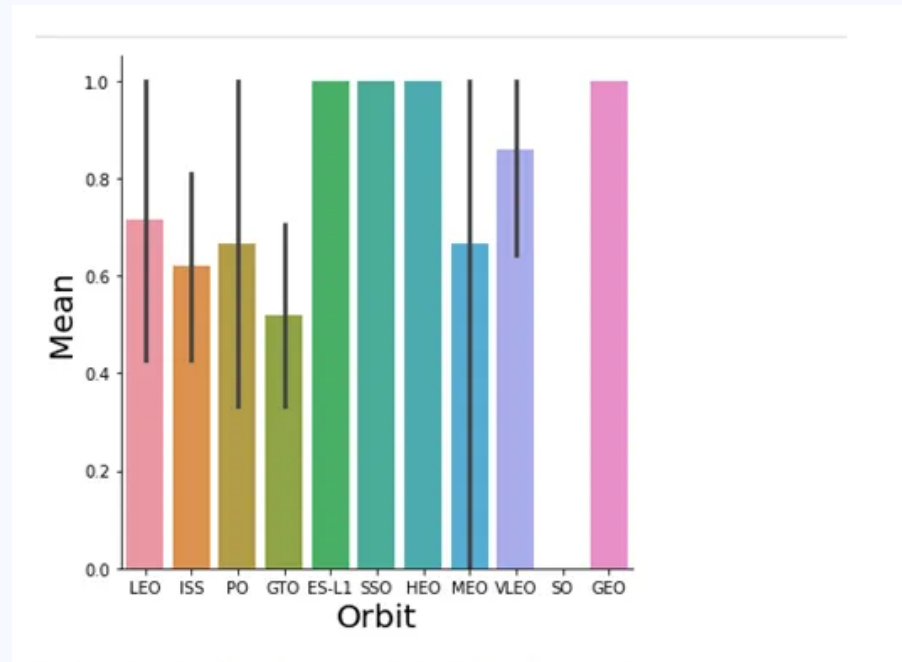
- 1, Earlier flights launch were from CCAFS-SLC-40 site , Followed by KSC-LC-39A
- 2, Most Launches are Launched from CCAFS-SLC-40
- 3, Fewer Launches from VAFB SLC 4E site



► Exploratory Data Analysis

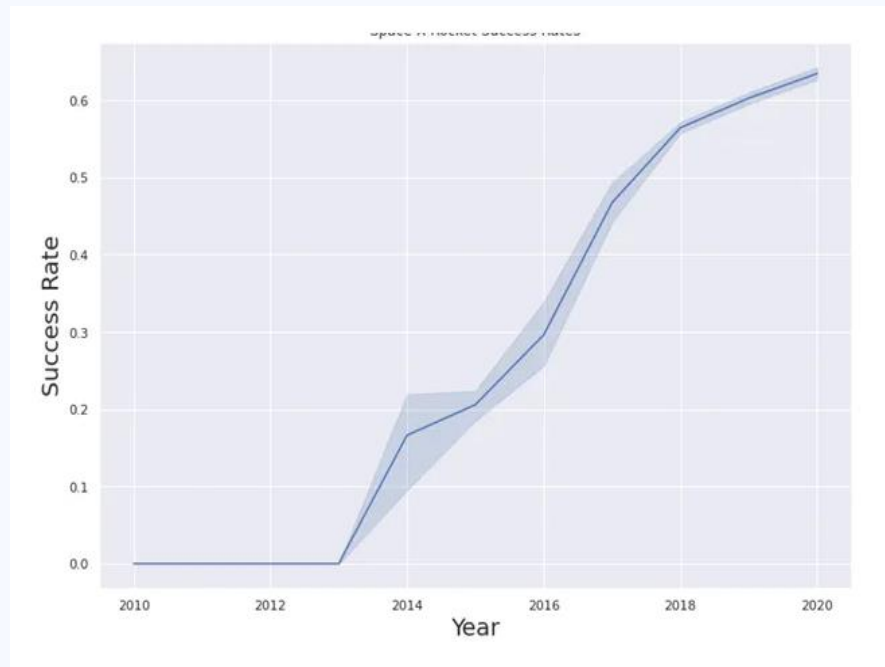
From the Visualization we can concluded that:

- VAFB SLC 4E has Low Payload launches
- CCAFS SLC 40 has more Higher Payload Launches and Low Payload Launches .



► Exploratory Data Analysis

Other analysis include Exploratory analysis of data from db2 database using sql statement to get insights

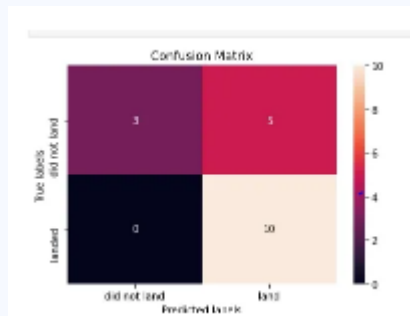


► Exploratory Data Analysis

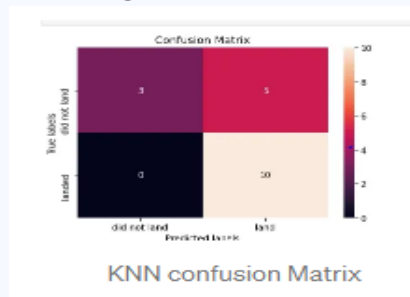
We analyze the data and see that their success rate which shows increase in landing success probability

.

► Predictive Analysis

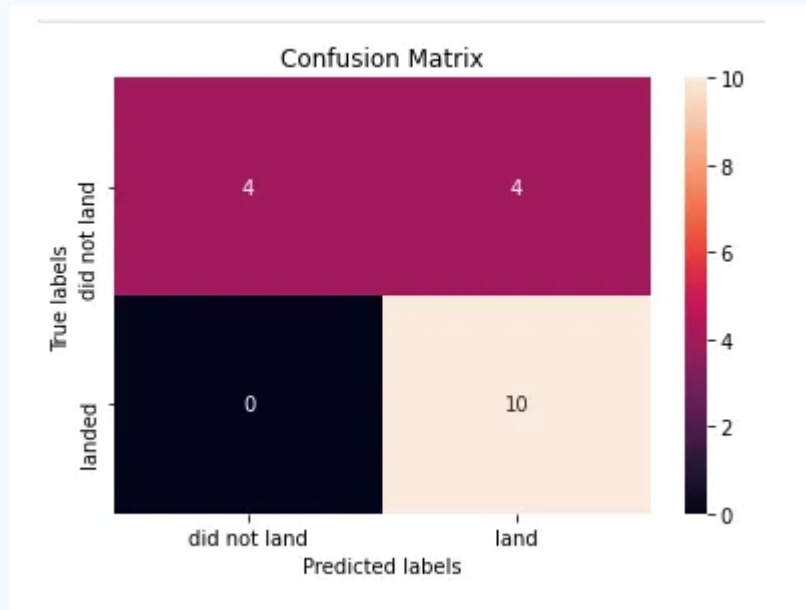


Using the data I trained the Machine learning models such as: KNeighborsClassifier

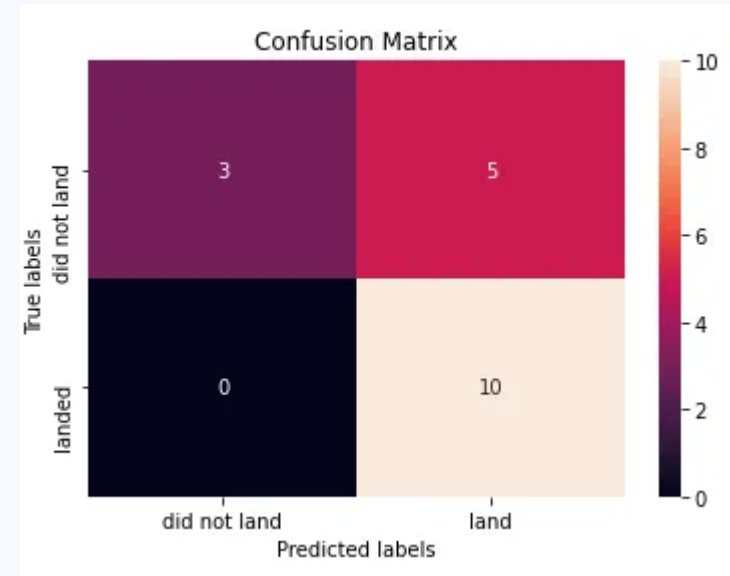


► Exploratory Data Analysis

Decision Tree classifier

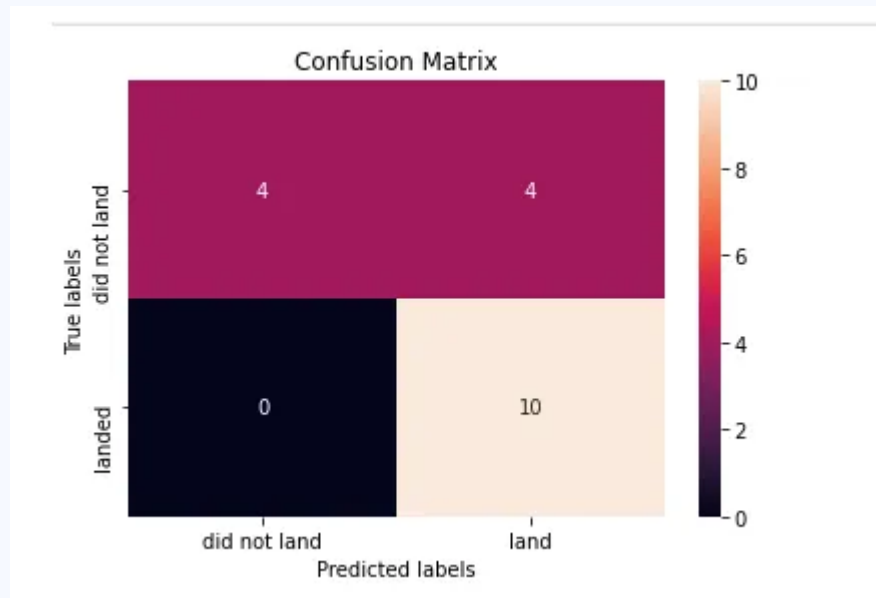


Logistic regression



► Exploratory Data Analysis

Support Vector Machine



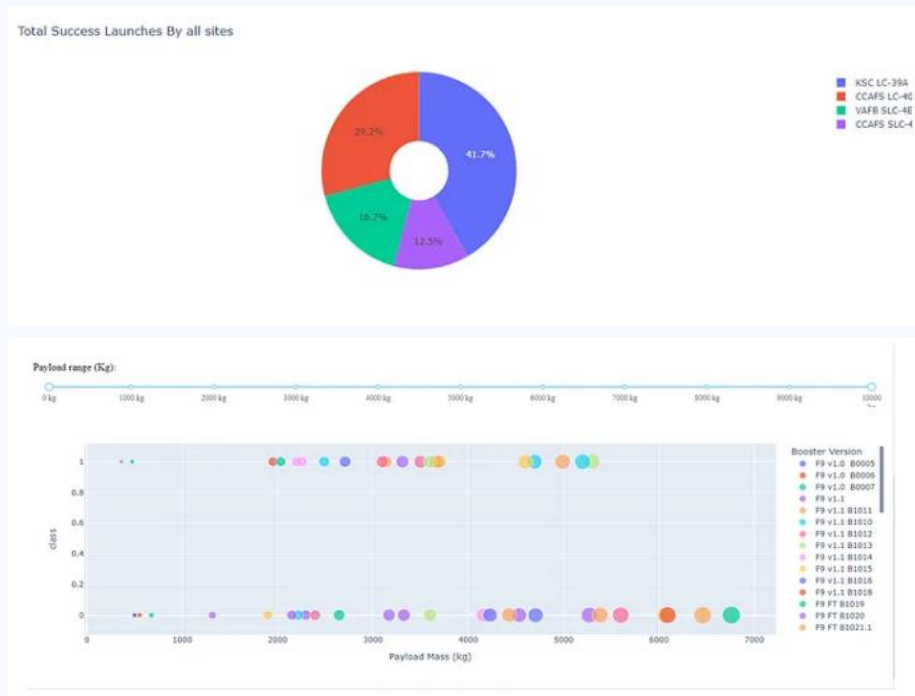
Interactive Map with Folium

Since spaceX launches come from different launch sites I displayed the information of failed and successful launches as a cluster on the map. Through zooming in and out you can observe the clusters of successful launches and failed launches.



Interactive Dashboard with Plotly Dash

Plotly Dash is Python library that makes it easier to create a dashboard for us as Data Scientist . With a simple interactive dashboard one can change the inputs to see representation of values in graphs .



Results

Most launches were from KSC PAD 39A since most of them were to VLEO, GEO or ISS which makes it a good site to launch from .

Falcon heavy launches mostly to full payload to maximize use of the falcon payload capacity .

Probability of booster landing increases over time by use of the data collected from failing .

SpaceX first successful booster Landing happened on 06/05/2016.

Conclusion

Using Existing Data and Analyzing the data ,SpaceX and other rocket companies can be able to see the best way to reduce the cost of launches,and evolve before there tradition costly launches lead to their absoluteness and losing their client .

THANK YOU