<SpaceX Falcon 9>

<Aditya Lokhande>

<05-05-2023>

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Problem Statement:

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

Solution Suggested:

- We can try to get some insights from this data so that we will have some prediction about the launch. Here we will analyze the different components involved in a rocket launch like launch site, engine model etc. and try to predict the result

- With the prediction in hand in advance, we can save lot of money. If any other company wants to bid against the SpaceX this insights will be very helpful because they can have a very good estimation of the price to bid for a project

- Those insights can give you provides the probability of the success of the launch which are very helpful and can save a lot of time and money.

INTRODUCTION

Purpose:

- To predict the result of launch (successful or failure) and if it is successful then try to closely estimate the cost of overall launch.

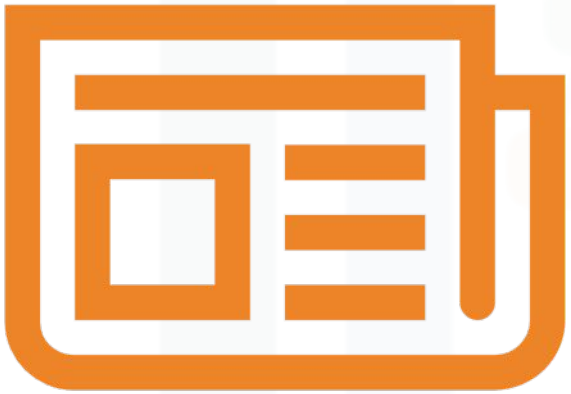
Objective :

- Get the insights by using the previous launch data to predict the launch result and the cost in the launch

Methods:

- Methods to get insights includes:
- Exploratory data Analysis
- Interactive visual Analytics
- Predictive Analysis

METHODOLOGY



- Data collection :
  - collect the data using APIs and web scraping
- EDA using SQL and visualization
- Interactive data visualization using folium and plotly dash
- Predictive analysis using classification models
  - Grid Search CV

Data Collection:

Data collection is a systematic process of gathering observations or measurements. Whether you are performing research for business, governmental or academic purposes, data collection allows you to gain first-hand knowledge and original insights into your research problem.

Important columns are: ['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']

Data Collection using APIs

Steps to follow for data Collection using APIs:

- Request (Spacex APIs)
- JSON file + Lists(Launch Site, Booster Version, Payload Data)
- Json_normalize to DataFrame data from JSON
- Dictionary relevant data
- Cast dictionary to a DataFrame
- Filter data to only include Falcon 9 launches
- Replace missing Payload Mass values with mean

Data Collection Web scraping

Steps to follow for Data Collection using Web Scarping:

- Request Wikipedia html
- BeautifulSoup html5lib Parser
- Find launch info html table
- Cast dictionary to DataFrame
- Iterate through table cells to extract data to dictionary
- Create dictionary

Data Wrangling:

Data Wrangling is the process of gathering, collecting, and transforming Raw data into another format for better understanding, decision-making, accessing, and analysis in less time. Data Wrangling is also known as Data Munging.

There are four broad steps in the munging process:

- Discovery
- Transformation
- Validation
- Publishing

Exploratory Data Analysis

EDA is a phenomenon under data analysis used for gaining a better understanding of data aspects like:
– main features of data
– variables and relationships that hold between them
– identifying which variables are important for our problem

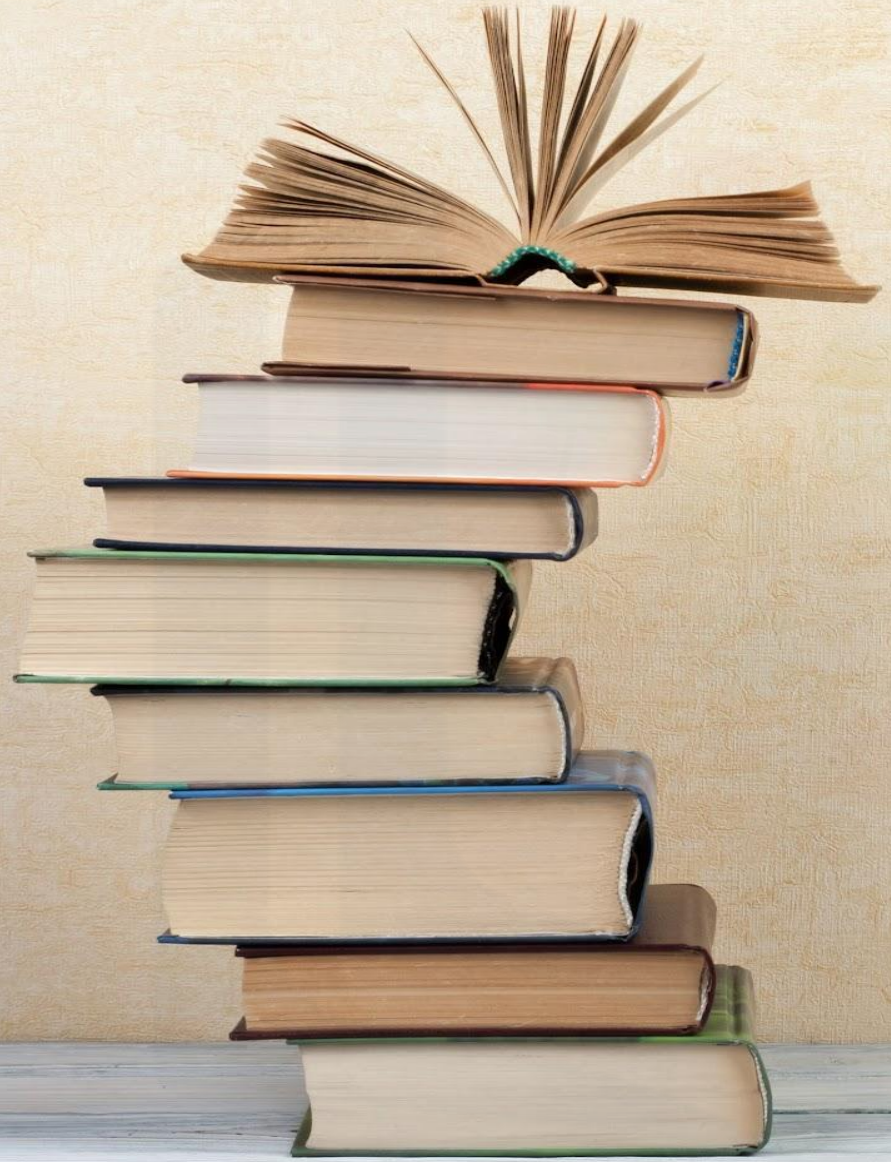
- Descriptive Statistics, which is a way of giving a brief overview of the dataset we are dealing with, including some measures and features of the sample
- Grouping data [Basic grouping with *group by*]
- ANOVA, Analysis Of Variance, which is a computational method to divide variations in an observations set into different components.
- Correlation and correlation methods

Predictive Analysis:

Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques and machine learning. Companies employ predictive analytics to find patterns in this data to identify risks and opportunities. Predictive analytics is often associated with big data and data science.

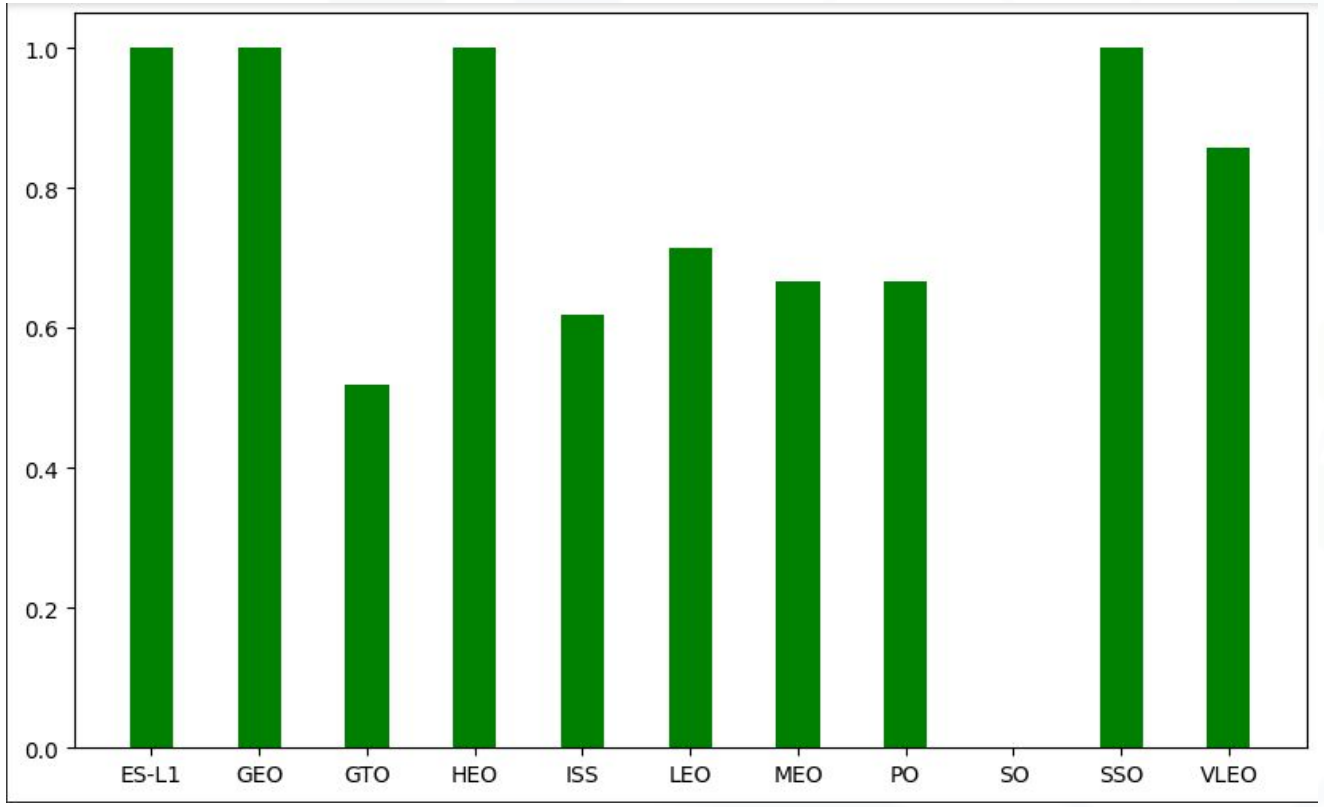
Methods used in the Solution :

- Support Vector Machine
- Decision tree classifier
- Logistic regression
- K-Nearest neighbour



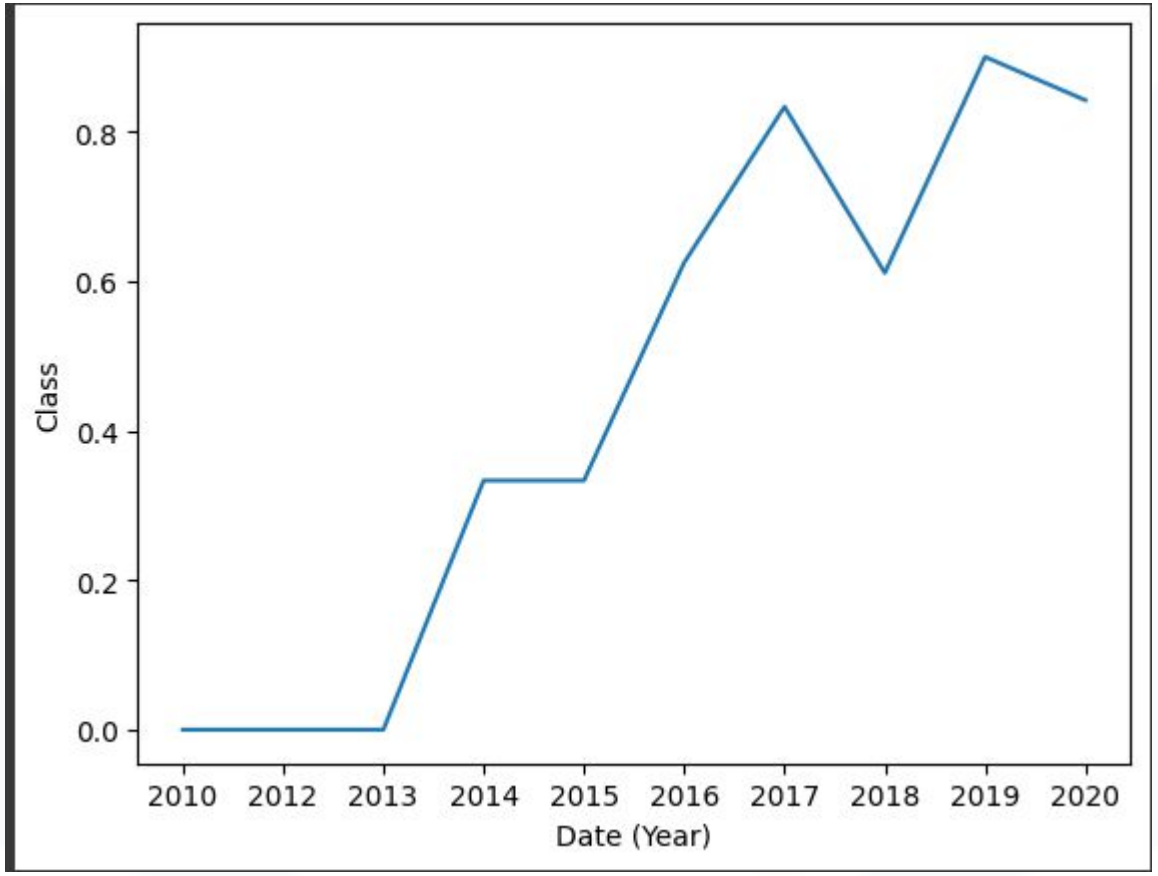
Results

EDA with Visualization:

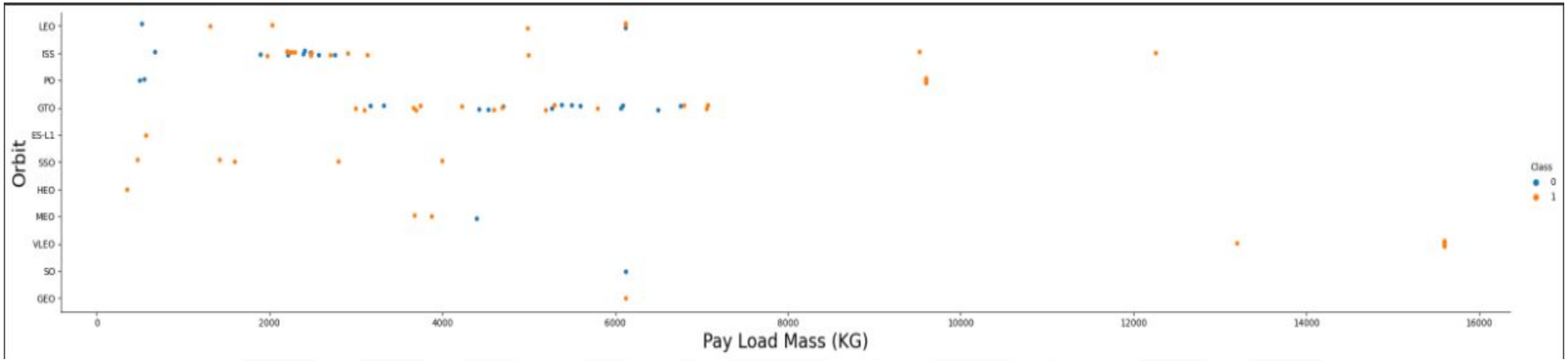


launch site ES-L1, GEO, HEO, SSO has 1.0 class mean where as SO has 0.0 class mean

Exploring and Preparing the Data:



This graph shows that the success rate after year 2013 is kept rising



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

EDA with SQL

the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[13]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Done.

[14]:

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

List the total number of successful and failure mission outcomes

Done.

[12]:

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

Average payload mass carried by booster version F9 v1.1

Done.

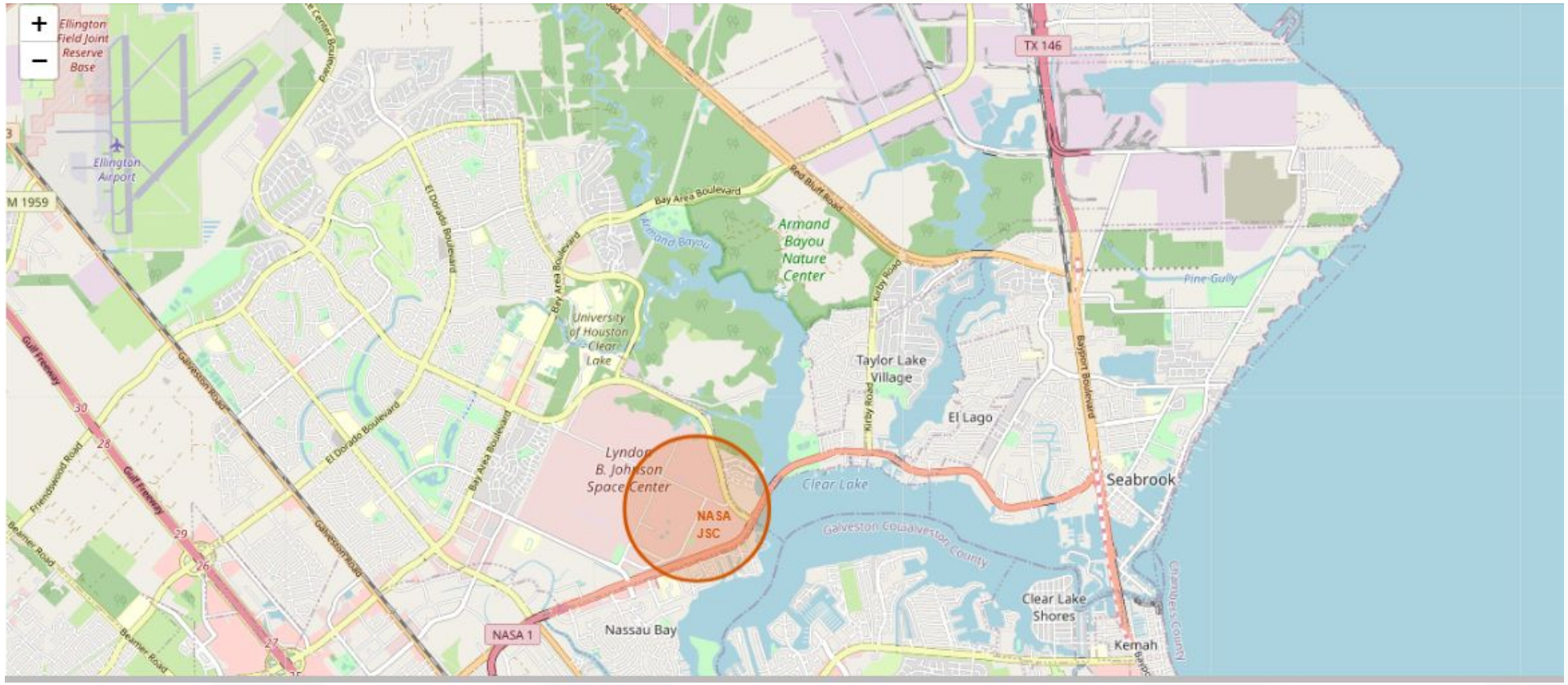
[15]:

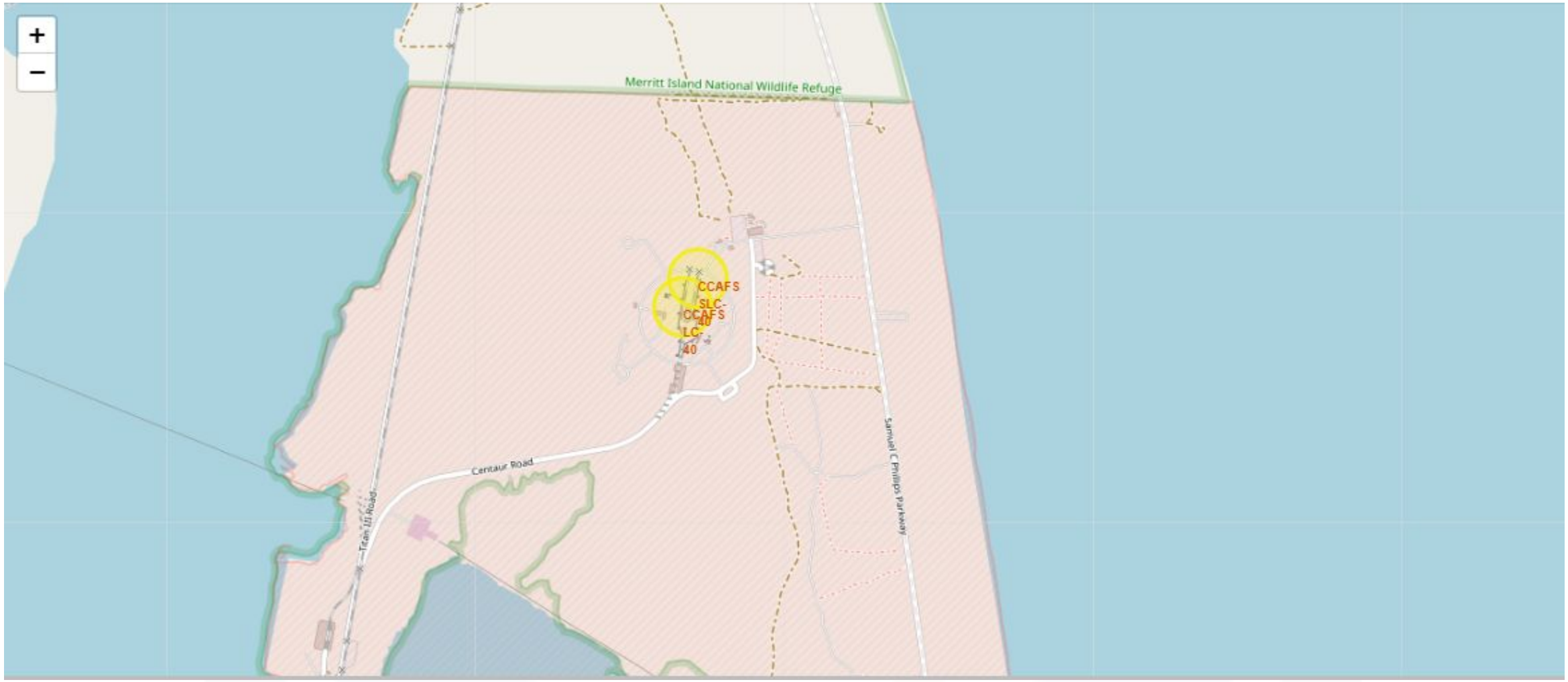
Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

List the names of the booster_versions which have carried the maximum payload mass.

This shows that F9 B5 B1048.4 can carry the maximum payload mass with success.

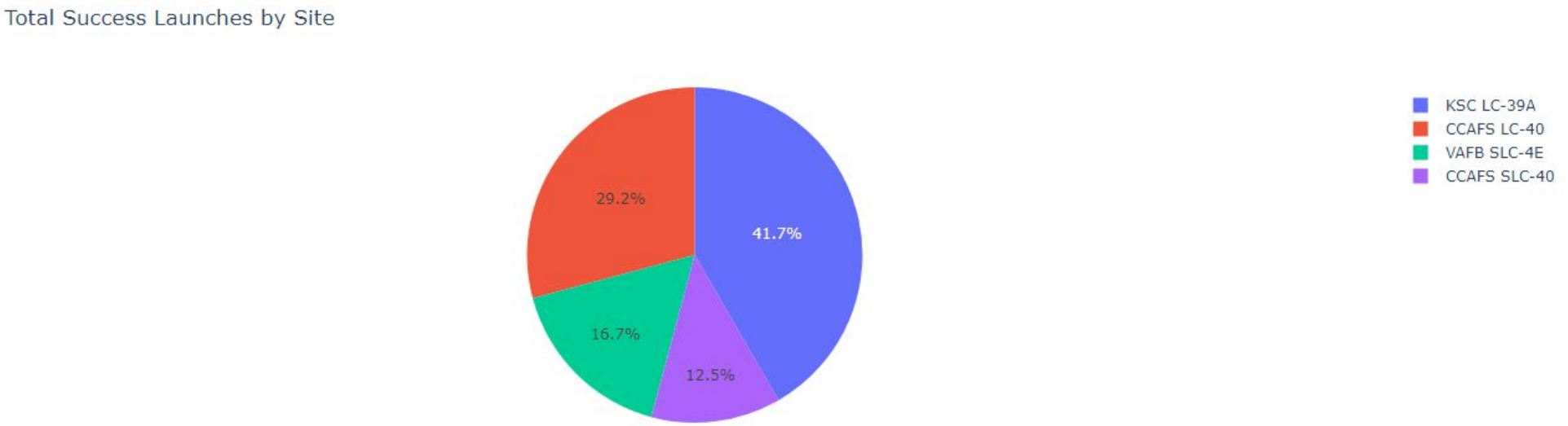
Folium map:


Ellington Field Joint Reserve Base
Ellington Airport
M 1959
Gulf Freeway
Galveston Road
El Dorado Boulevard
Bay Area Boulevard
Bay Area Boulevard
Red Bluff Road
TX 146
Armand Bayou
Armand Bayou Nature Center
Pine Gully
University of Houston Clear Lake
Kirby Road
Taylor Lake Village
El Lago
Baypoint Boulevard
Seabrook
Lyndon B. Johnson Space Center
NASA JSC
Clear Lake
Galveston County
Clear Lake Shores
Kemah
Chambers County
NASA 1
Nassau Bay
Friendswood Road
Beamer Road
30
28
29
26
27

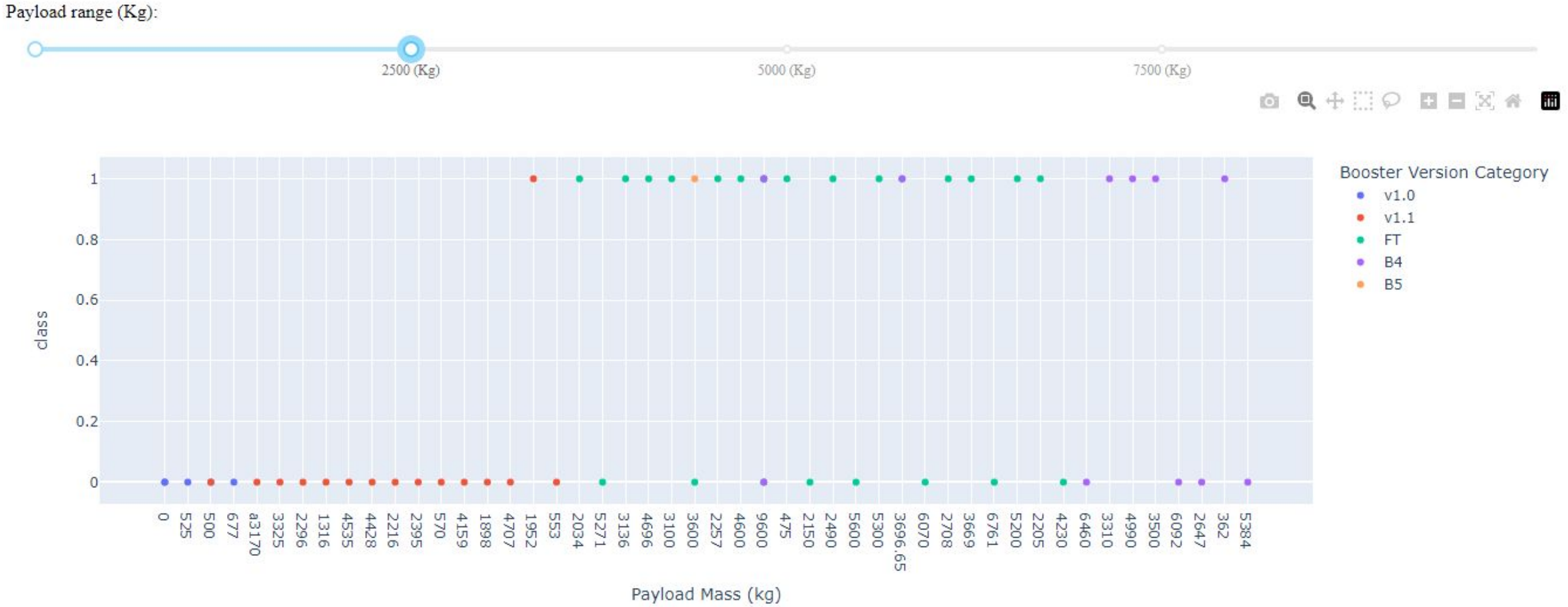


It is a polyline between launch site to the selected coastline point

Interactive dashboard with plotly dash:



This is the distribution of successful landings across all launch sites. CCAFS LC-40 is the old name of CCAFS SLC-40 so CCAFS and KSC have the same amount of successful landings, but a majority of the successful landings were performed before the name change.



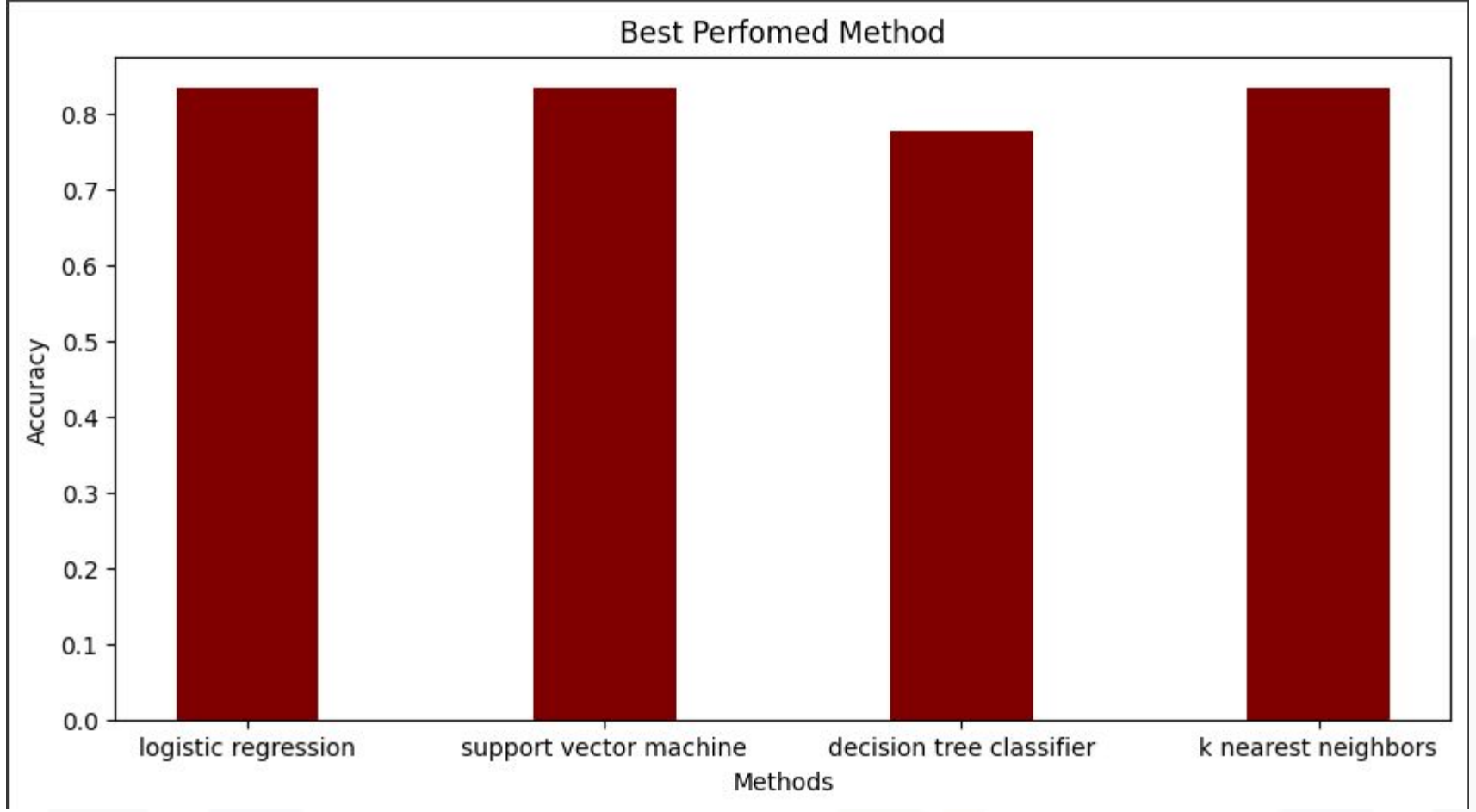
This shows the logistic regression graph between pyload mass and class
This also includes the data about different version of booster.

Predictive Data Analysis

Applied various Data Analysis methods to the data which are:

- Support machine vector
- tree_cv
- logreg_cv
- K-Nearest Neighbour

Those different methods gives different results with varying accuracy. Comparing the accuracy is one of the methods to compare all the different methods

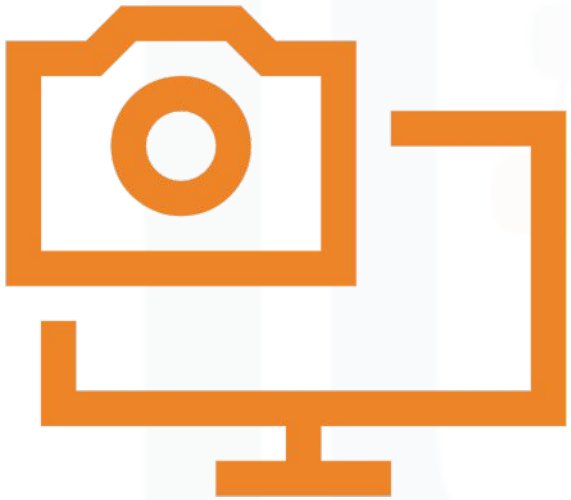


The image shows the accuracy of the different Data Analysis methods. **Logistic regression**, [support vector machine](), [k nearest neighbour]() show very similar accuracy and we can select any one of them.

CONCLUSION

- Our task: to develop a machine learning model for Space Y who wants to bid against SpaceX
- The goal of model is to predict when Stage 1 will successfully land to save ~$100million USD
- Used data from a public SpaceX API and web scraping SpaceX Wikipedia page
- Created Data Labels and stored data into a DB2SQL database
- Created a dashboard for visualization
- We created a machine learning model with an accuracy of 83%

APPENDIX



- Github repo link: https://github.com/adii492/DS_Capstone_final_project.git