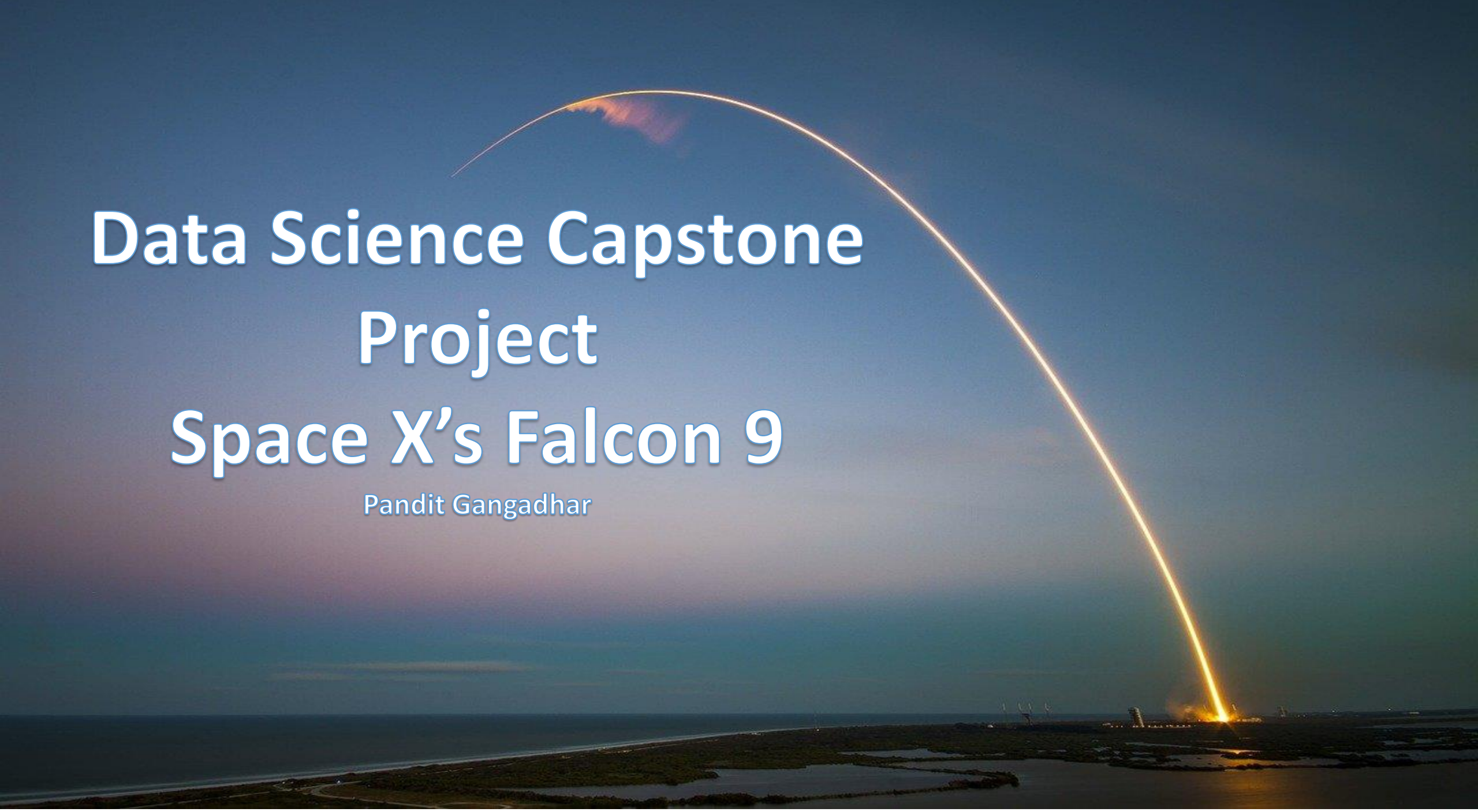


Data Science Capstone Project

Space X's Falcon 9

Pandit Gangadhar

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context

Space X 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 Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch.

- Problems you want to find answers

This Project is focused on Predicting if the falcon 9 first stage will land successfully.

METHODOLOGY

- Data collection
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from Space X REST API (https://api.spacexdata.com/v4/launches/past)
- Perform data wrangling
  - Using the Python package BeautifulSoup for web scraping and using html table to find relevant data of falcon 9
  - Data frame was created by parsing the launch HTML tables
  - Convert the raw data into clean dataset to make its meaningful
  - Data wrangling by using some technique by dealing with null value of dataset
  - Creating the new column of class and gave the outcome of 0 and 1
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification model

Data Collection

- Describe how data sets were collected.
  - Data were collected in two ways

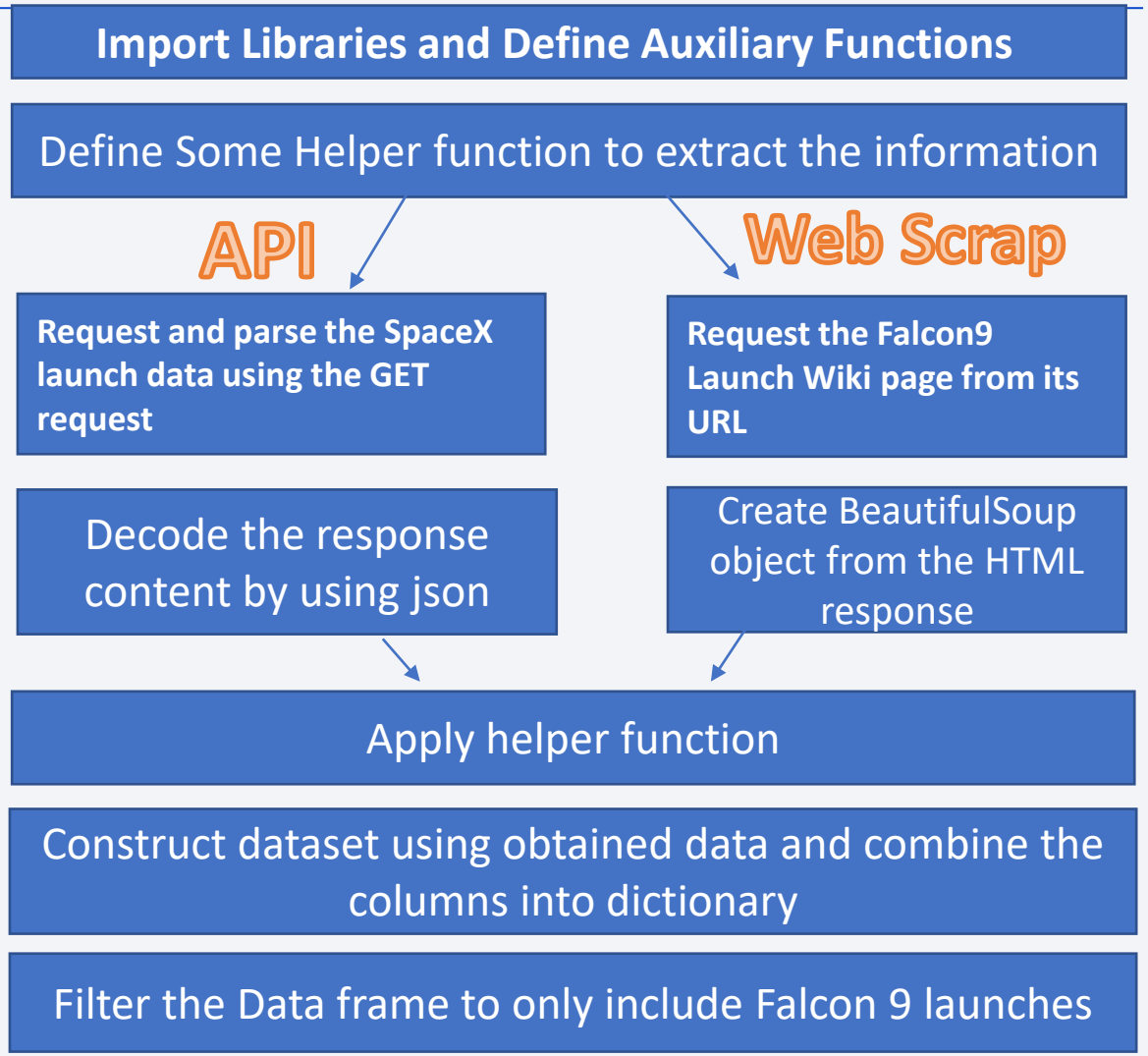
1. By using Space X REST API

   (https://api.spacexdata.com/v4/launches/past)

1. By using Python package BeautifulSoup and

   scraping from Wikipedia

(https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)

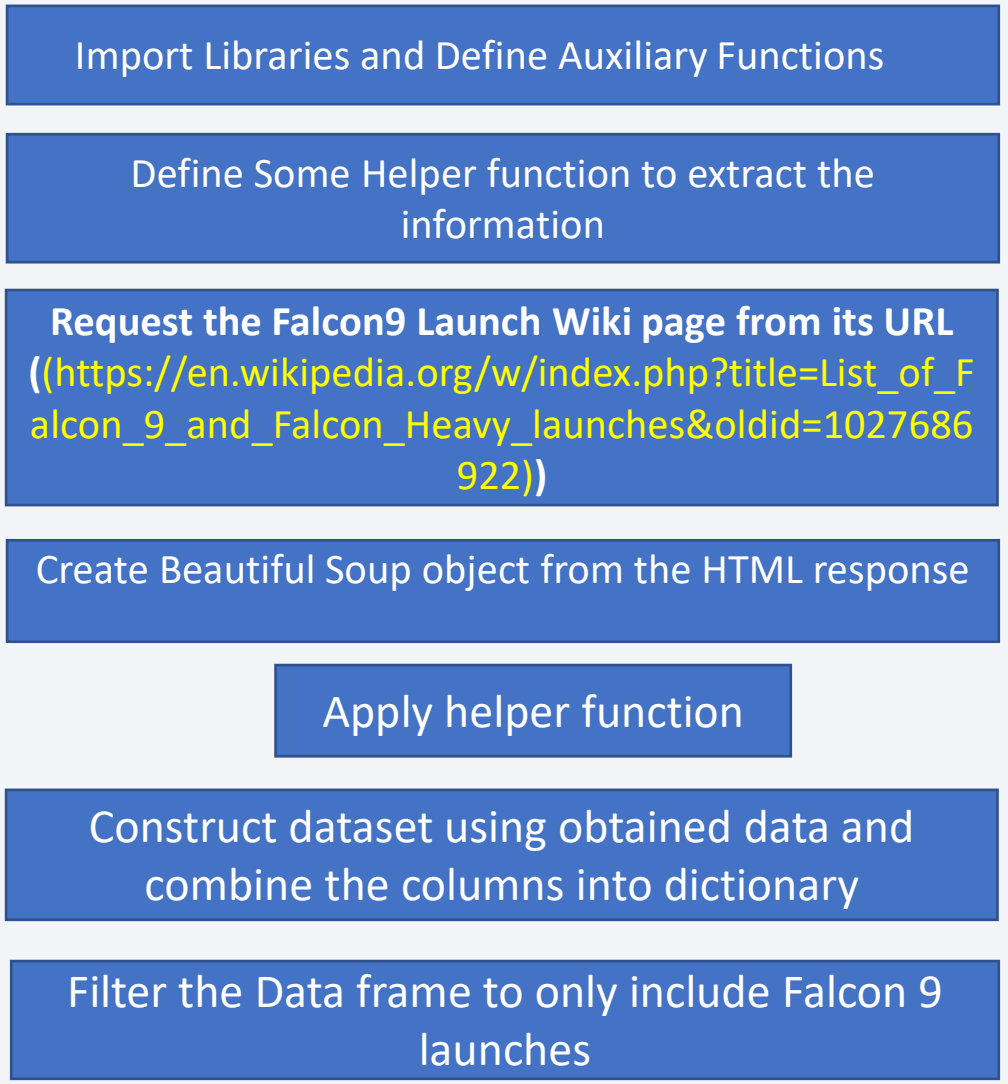


Data Collection - Scraping

Web scraping process using flowcharts

GitHub URL of the completed web scraping notebook

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Web%20scraping%20of%20data.ipynb)

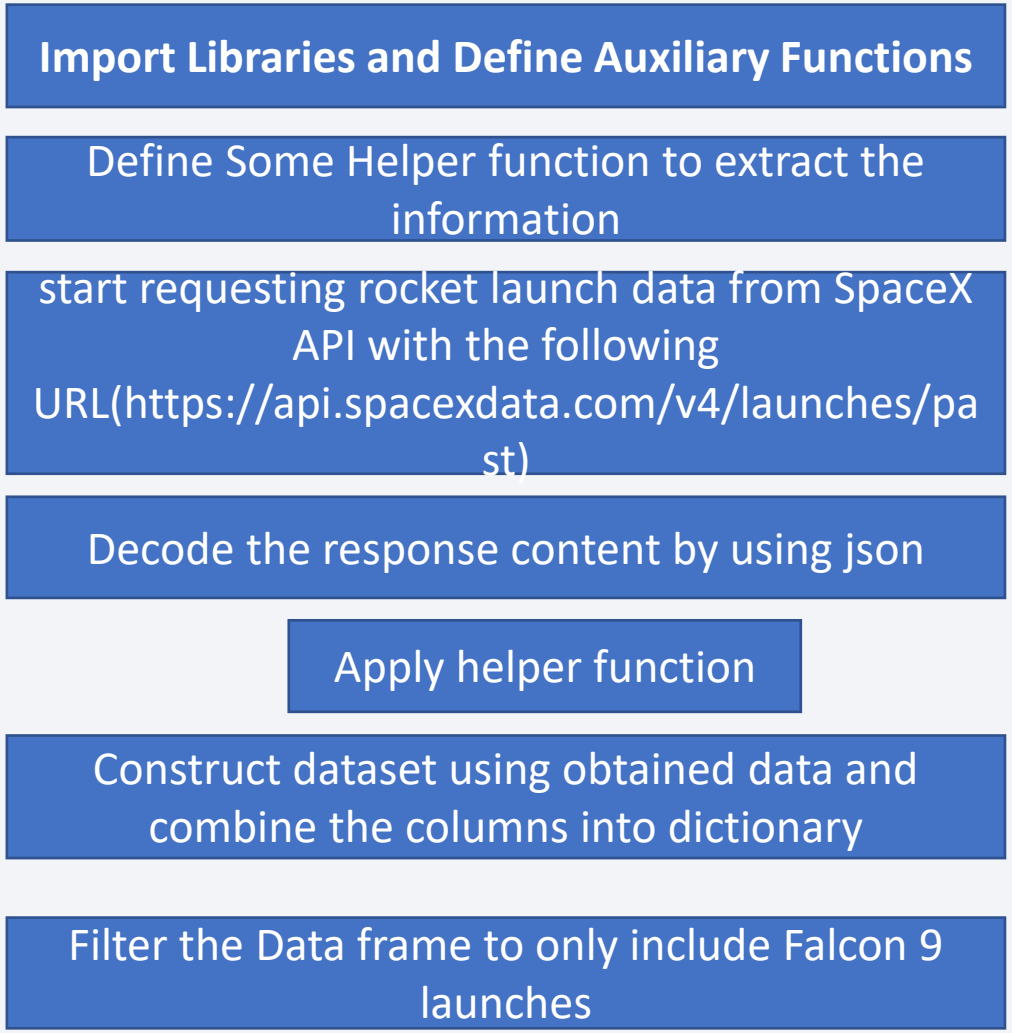


Data Collection – SpaceX API

Data collection with SpaceX REST calls using flowchart

GitHub URL of the completed SpaceX API calls notebook

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Data%20Collection%20From%20Api%20(1).ipynb)



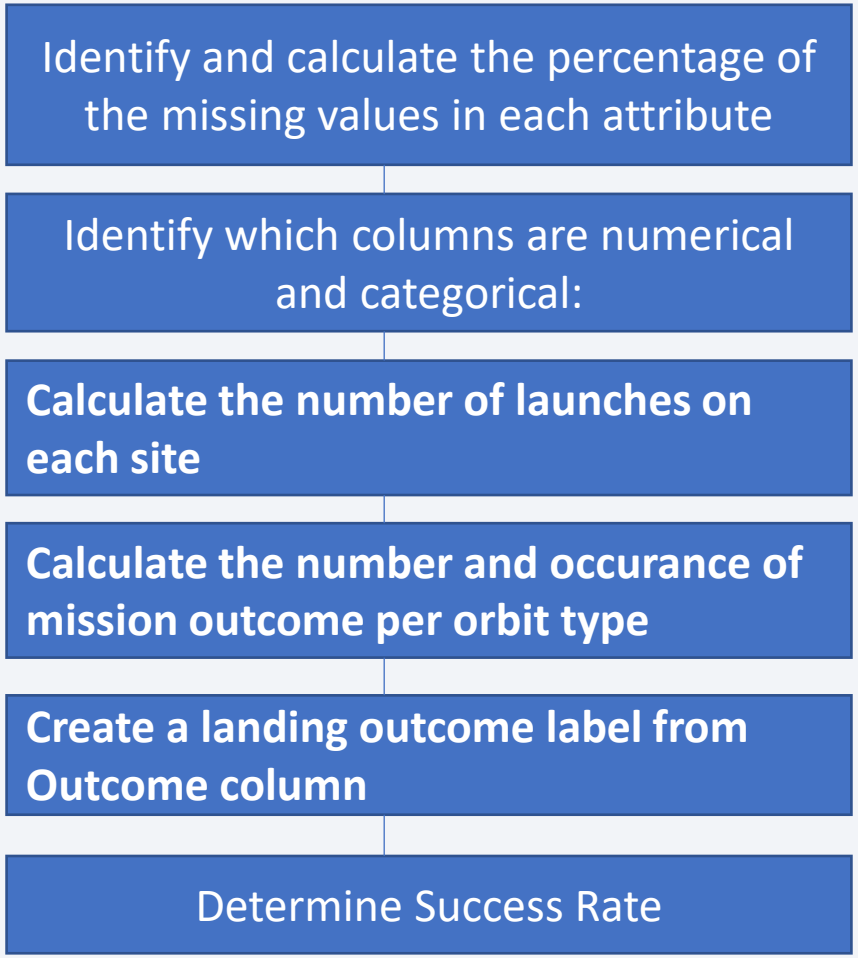
Data Wrangling

Data Wrangling using Flowcharts

Github url of completed

Space X API calls Notebook

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Exploratory%20Data%20Analysis.ipynb)



EDA with Data Visualization

To understand the data very clearly and find the relevant feature that is more effective on target data, charts plot is more useful to explain data.

The following charts plot were used:

- Scatter plot between Payload Mass and Flight Number
- Scatter plot between Flight Number and Launch Site
- Scatter plot between Launch site and Payload
- Bar chart for success rate of each orbit type
- Scatter plot between Flight Number and Orbit type
- Scatter plot between Payload and Orbit type
- Line plot for launch yearly on average success rate

Github url Space X API calls Notebook:

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Visualization%20with%20EDA%20spaceX.ipynb)

EDA with SQL

Following SQL queries were performed:

- *Display the names of the unique launch sites in the space mission*
- *Display 5 records where launch sites begin with the string 'CCA'*
- *Display the total payload mass carried by boosters launched by NASA (CRS)*
- *Display average payload mass carried by booster version F9 v1.1*
- *List the date when the first successful landing outcome in ground pad was acheived*
- *List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
- *List the total number of successful and failure mission outcomes*
- *List the names of the booster_versions which have carried the maximum payload mass. Use a subquery*
- *List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015*
- *Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order*

- Add the GitHub URL of your completed EDA with SQL notebook

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/sql%20EDA%20.ipynb)

Build an Interactive Map with Folium

Here lists of map objects that is created by folium map:

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

By plot this map object it was found that some locations such as railway, highway and city were in close proximity to launch site

GitHub URL of completed interactive map with Folium map:

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Interactive%20Visual%20Analytics%20with%20Folium.ipynb)

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

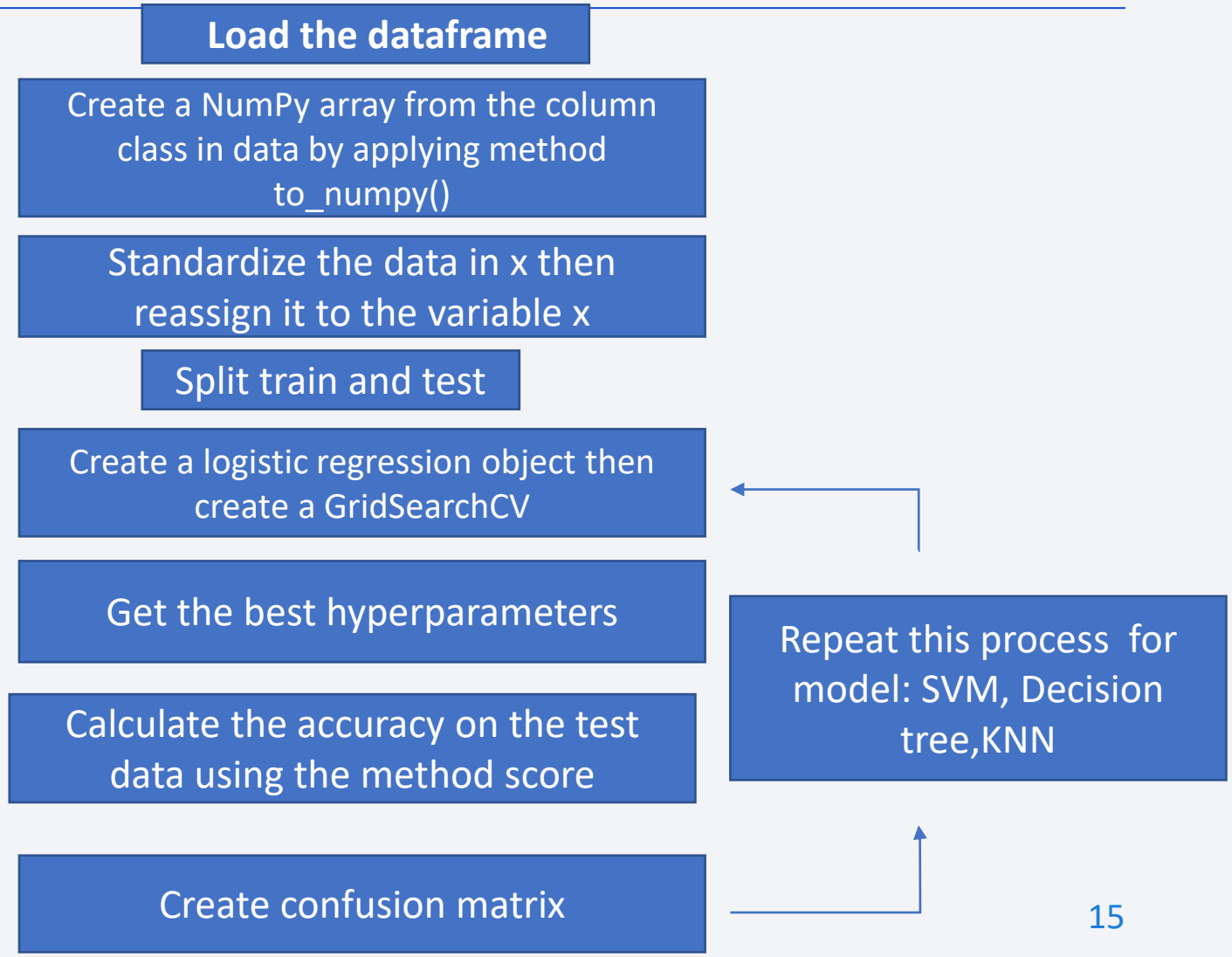
Predictive Analysis (Classification)

Process of Predictive analysis

- Load the data set and create numpy from column class in data by applying method to_numpy().
- Before splitting the data. It is import to standardize the the data to avoid the biased model then split it.
- Create the model of logistic regression then create GridSearchCV and choose best hyperparameters to calculate the accuracy score of the test data
- Create confusion matrix
- After that repeat the for model svm, decision tree and knn.

Github URL:

(https://github.com/pandit0305/SpaceX_Falcon9_analysis/blob/8628d6ac191e0621a7a232eee9b5ae5b97980972/Machine%20Learning%20Prediction.ipynb)



Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Insights Drawn
From Data

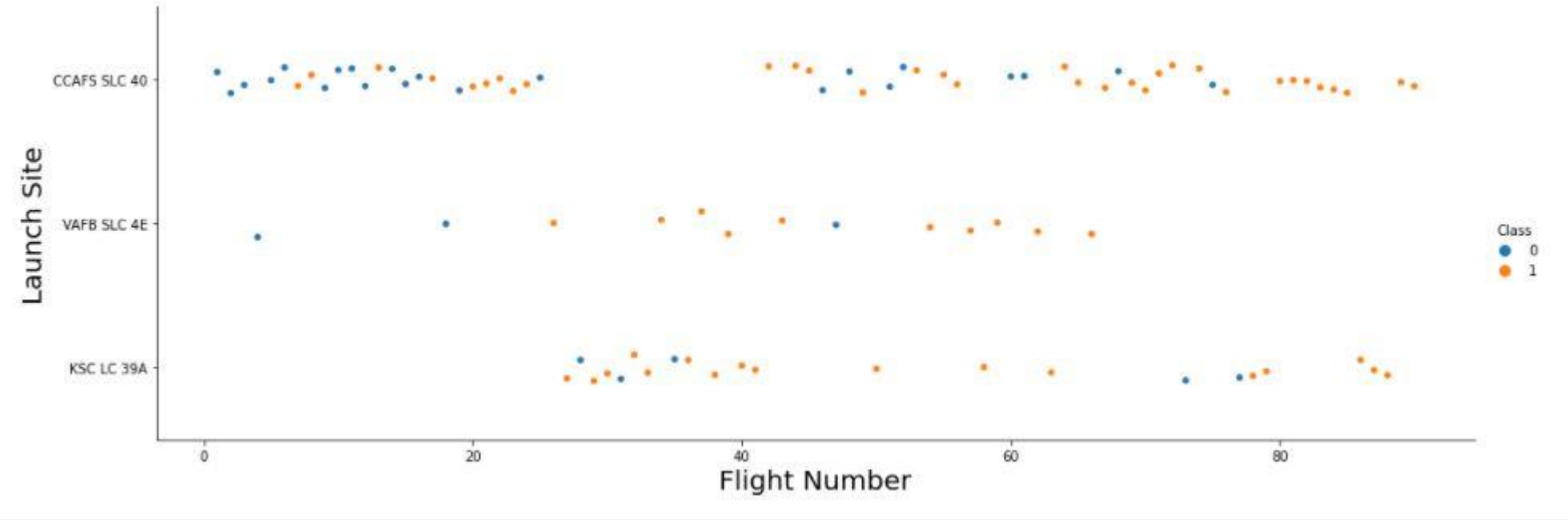
EDA with Visualization

- Scatter plot between Payload Mass and Flight Number
- Scatter plot between Flight Number and Launch Site
- Scatter plot between Launch site and Payload
- Bar chart for success rate of each orbit type
- Scatter plot between Flight Number and Orbit type
- Scatter plot between Payload and Orbit type
- Line plot for launch yearly on average success rate

Flight Number vs. Launch Site

Scatter plot of Flight Number vs. Launch Site

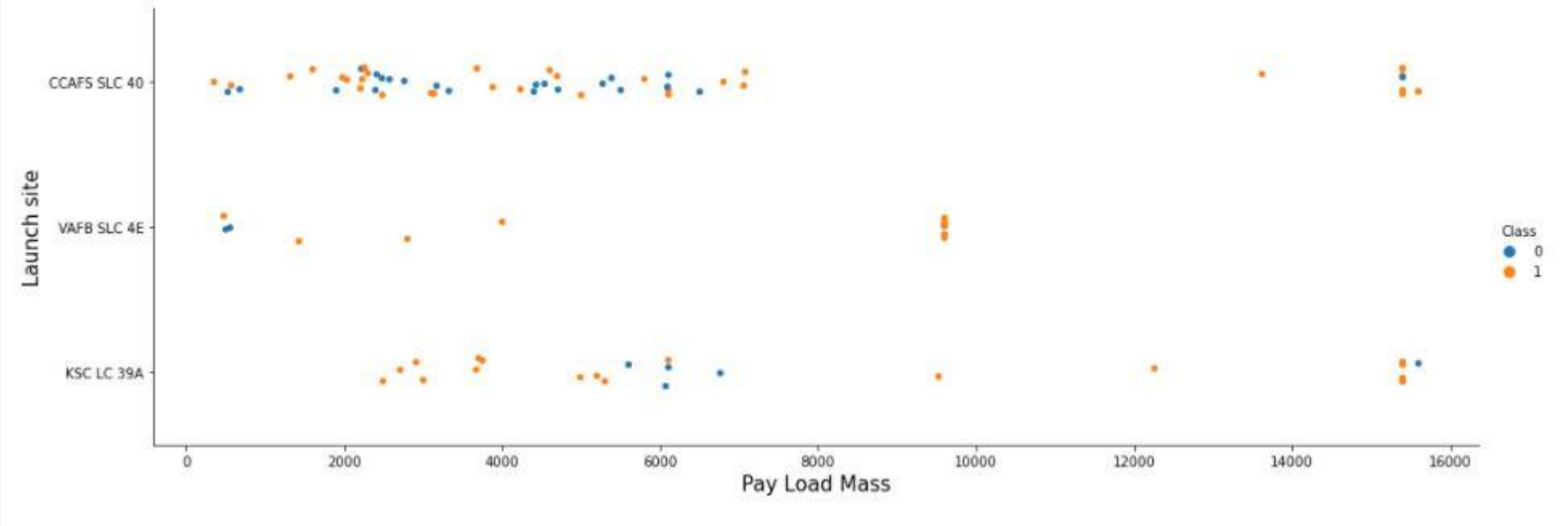
- The plot shows on launch site CCAFS SLC 40, there are many flights launched successfully and rapidly increases with increase flight number.



Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

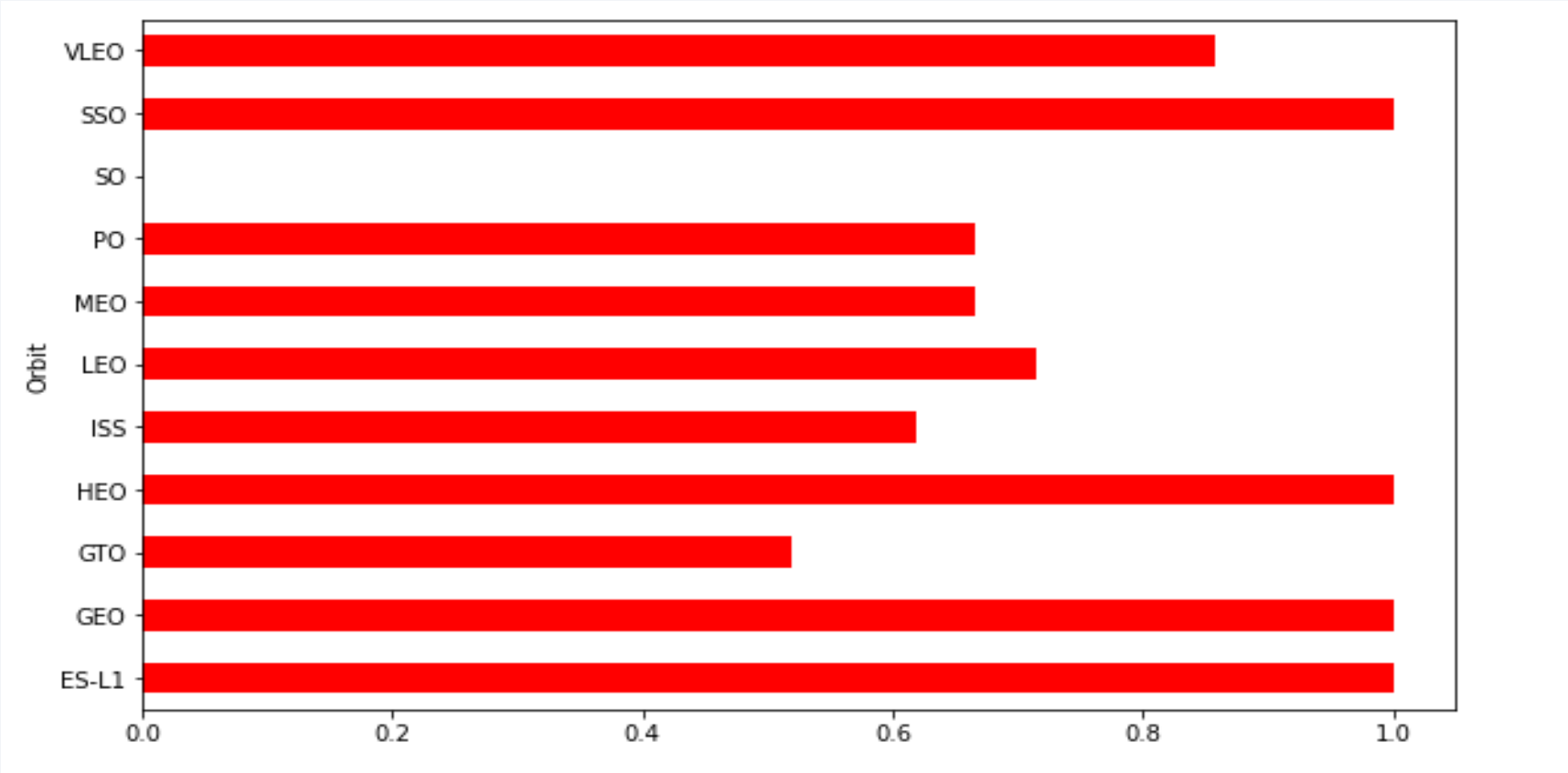
The plot shows that launch site VAFB SLC 4E there is no any rockets are launched having heavy load mass that is greater that 10000 kg



Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type

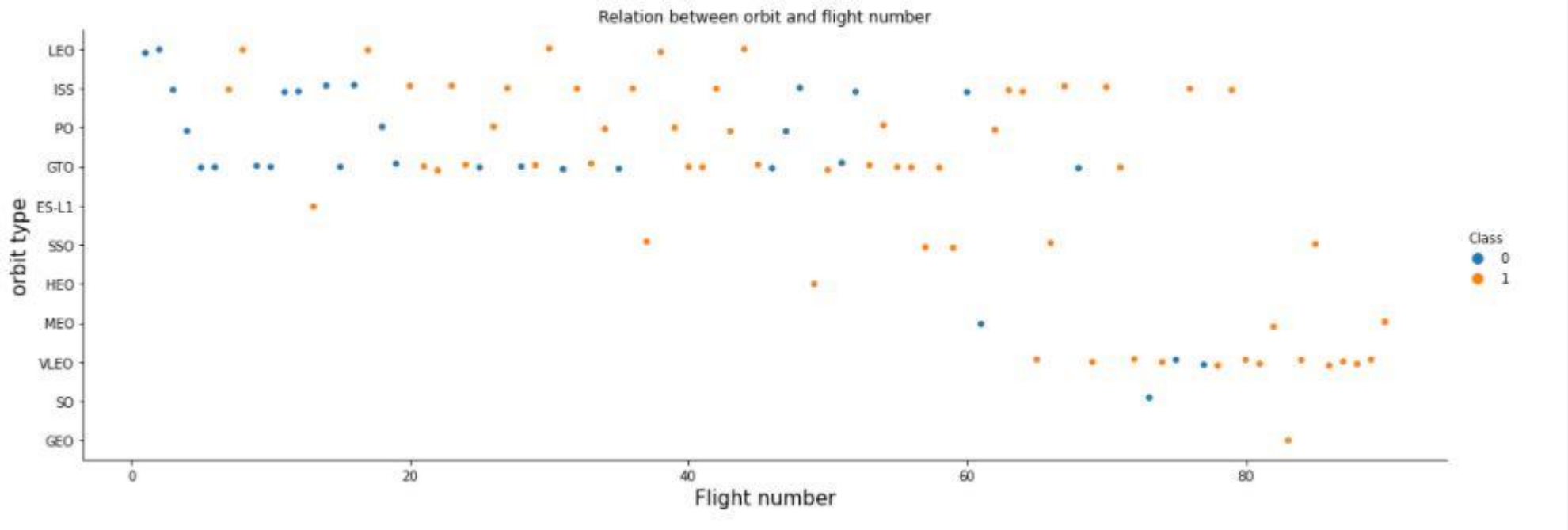
Bar Chart shows that orbit SSO, HEO, GEO, and ES-L1 have high success rate and VLEO orbit has the second highest success rate.



Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit type

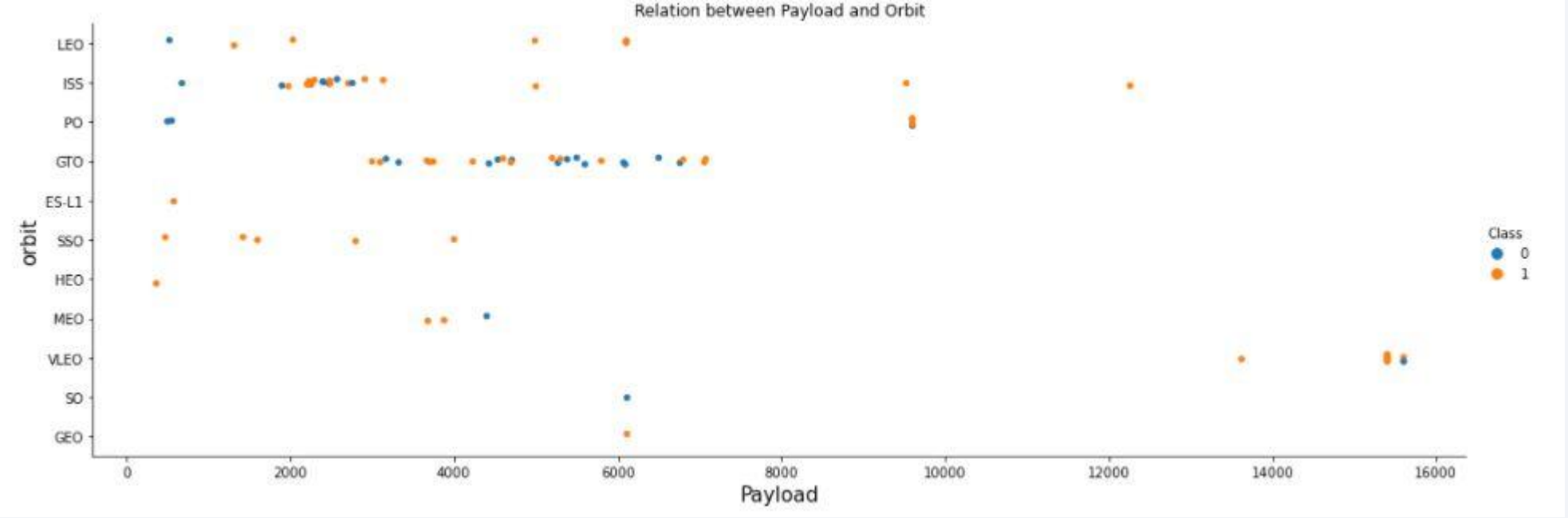
The plot shows that the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

Scatter point of payload vs. orbit type

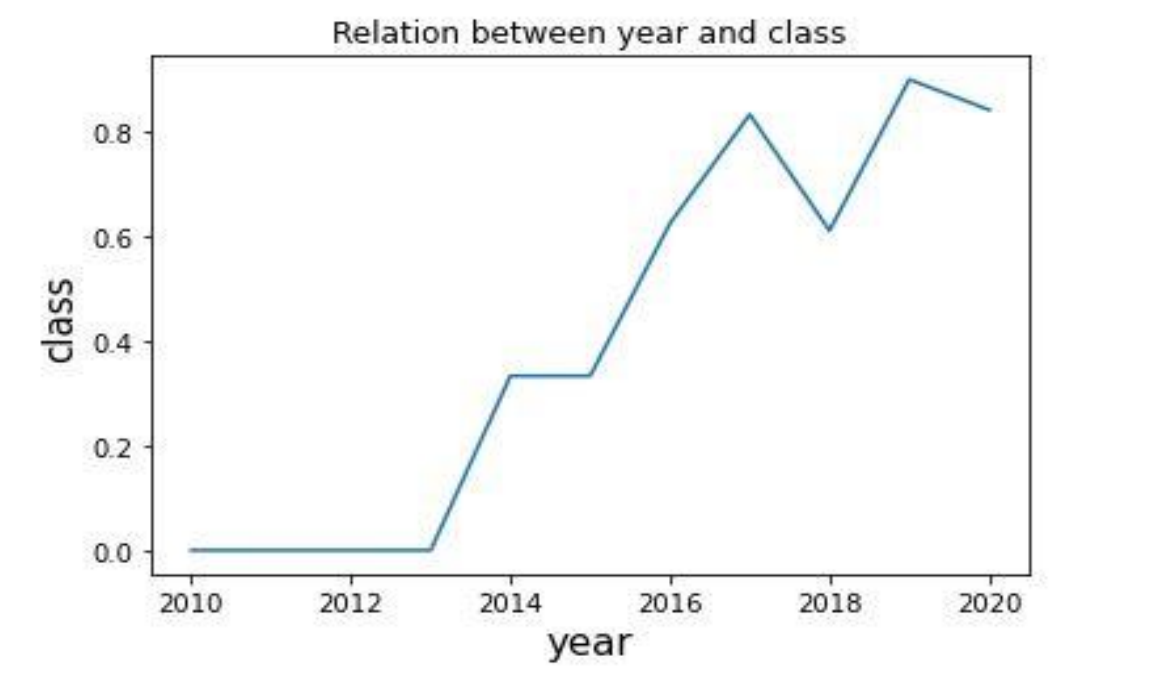
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

Line chart of yearly average success rate

The line plot shows that success rate since 2013 kept increasing till 2020



EDA with SQL

- *Display the names of the unique launch sites in the space mission*
- *Display 5 records where launch sites begin with the string 'CCA‘*
- *Display the total payload mass carried by boosters launched by NASA (CRS)*
- *Display average payload mass carried by booster version F9 v1.1*
- *List the date when the first successful landing outcome in ground pad was acheived*
- *List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
- *List the total number of successful and failure mission outcomes*
- *List the names of the booster_versions which have carried the maximum payload mass. Use a subquery*
- *List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015*
- *Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order*

All Launch Site Names

The names of the unique launch sites

Out[14]:

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Records where launch sites begin with `CCA`

Out[22]:

launch_site
CCAFS LC-40
CCAFS SLC-40

Total Payload Mass

Total payload carried by boosters from NASA

1
45596

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

1
2534

First Successful Ground Landing Date

Dates of the first successful landing outcome on ground pad

DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Total number of successful mission outcomes

100

and total number failure mission outcomes

1

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	total
Success (drone ship)	5
Success (ground pad)	3

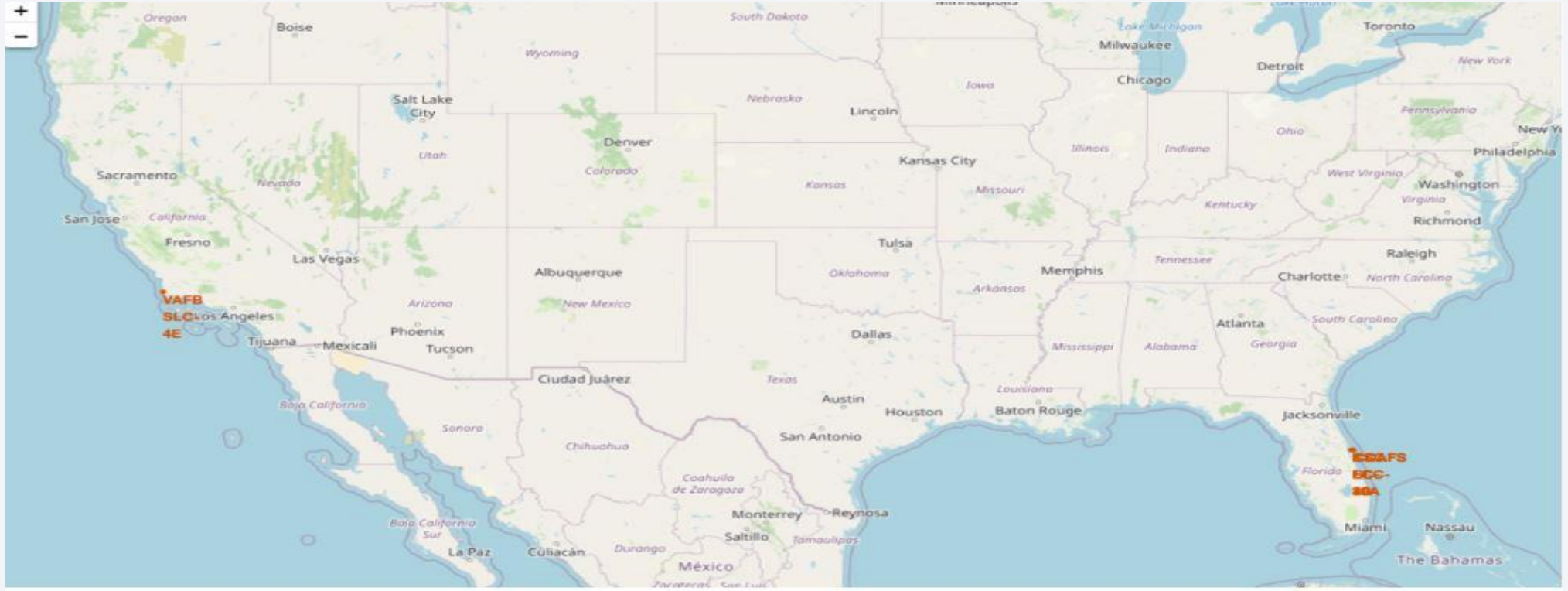
Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Folium Map: Mark all launch sites on a map

Add each site's location on a map using site's latitude and longitude coordinates

In this map, one launch site is in west location and rest launch sites location are in east location and all of them is in the proximity of the coastline



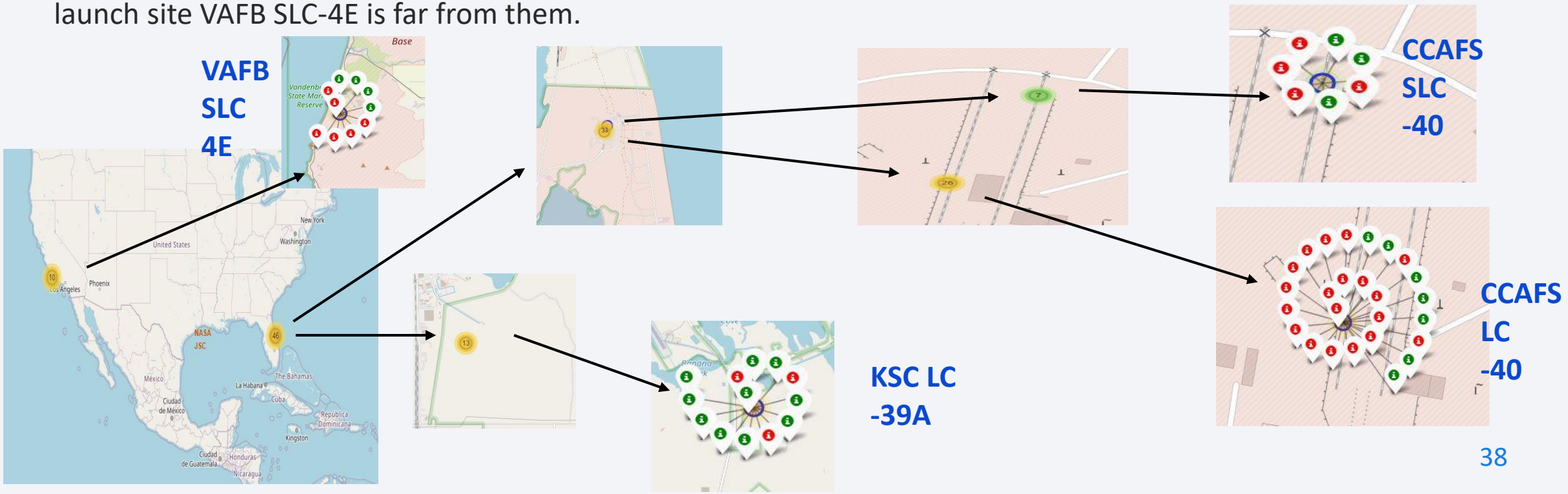
Folium Map: Color- labeled launch outcome on the map

Launch site CCAFS SLC-40 having lowest number of launches but launch success rate is highest.

Launch site CCAFS LC -40 having large number of launches but success rate of launches is low.

Launch site KSC LC -39A having large number of launches and success rate also.

Map shows the launch site CCAFS SLC-40, CCAFS LS-40, and KSC LC-39A are at close distance to each other only launch site VAFB SLC-4E is far from them.



Folium Map: Launch site to its proximities such as railway, highway, coastline with distance calculated and displayed

Launch site VAFB SLC-4E is at close to railway location about 0.19 km after looking other location such as highway, city and coastline, it can be observed

that :

Launch sites are in close proximity to railways.

Launch sites are in close proximity to highway.

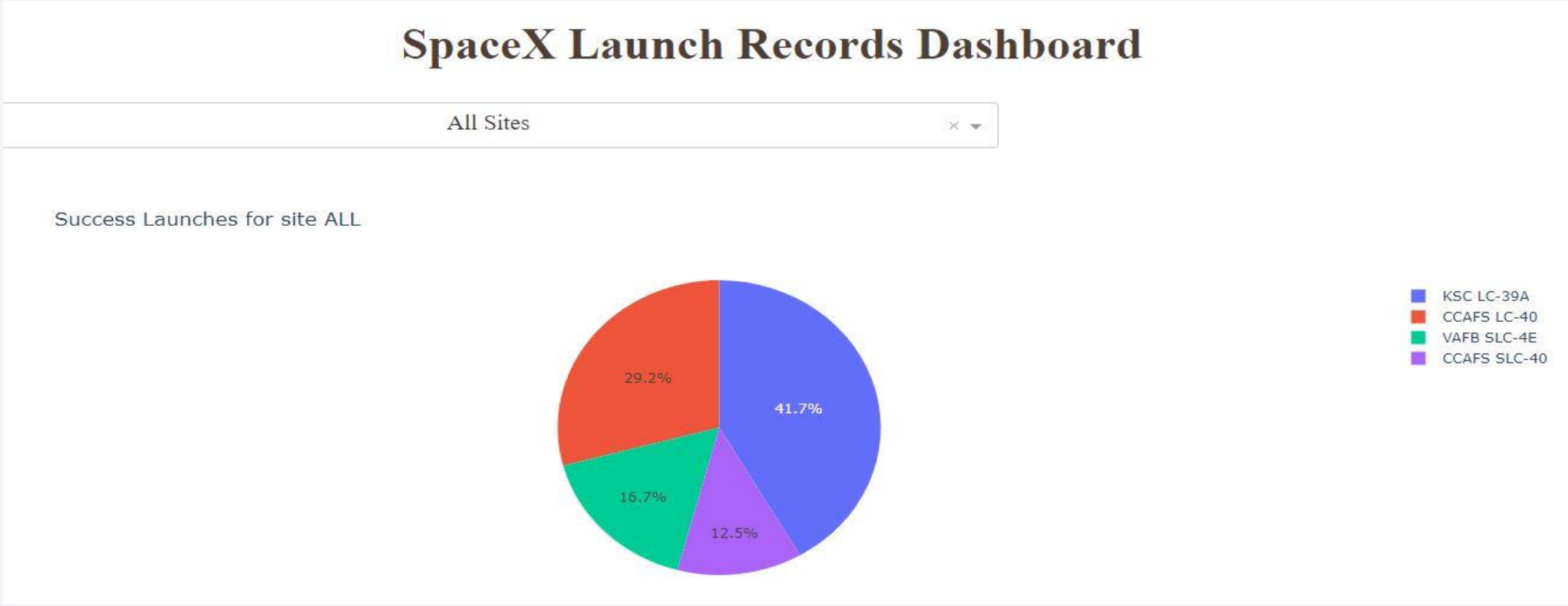
Launch sites are in close proximity to coastline.

Launch sites keep certain distance away from cities.

Dashboard

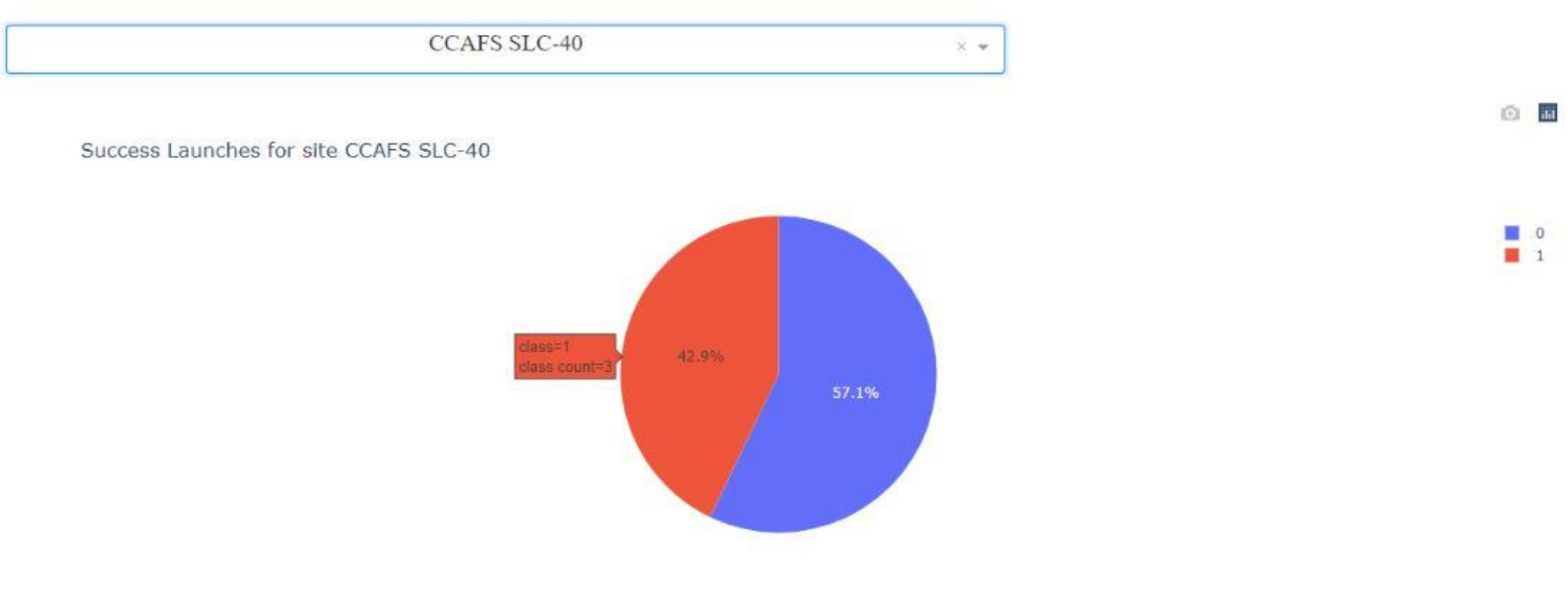
Launch Success Count For All Sites: Pie-chart

Here, four sites are shown in pie chart such as KSC LC-39A, CCAFS LC-40, VAFB SLC-4E, and CCAFS SLC-40. In this pie-chart, CCAFS SLC-40 has low percentage with low class count.



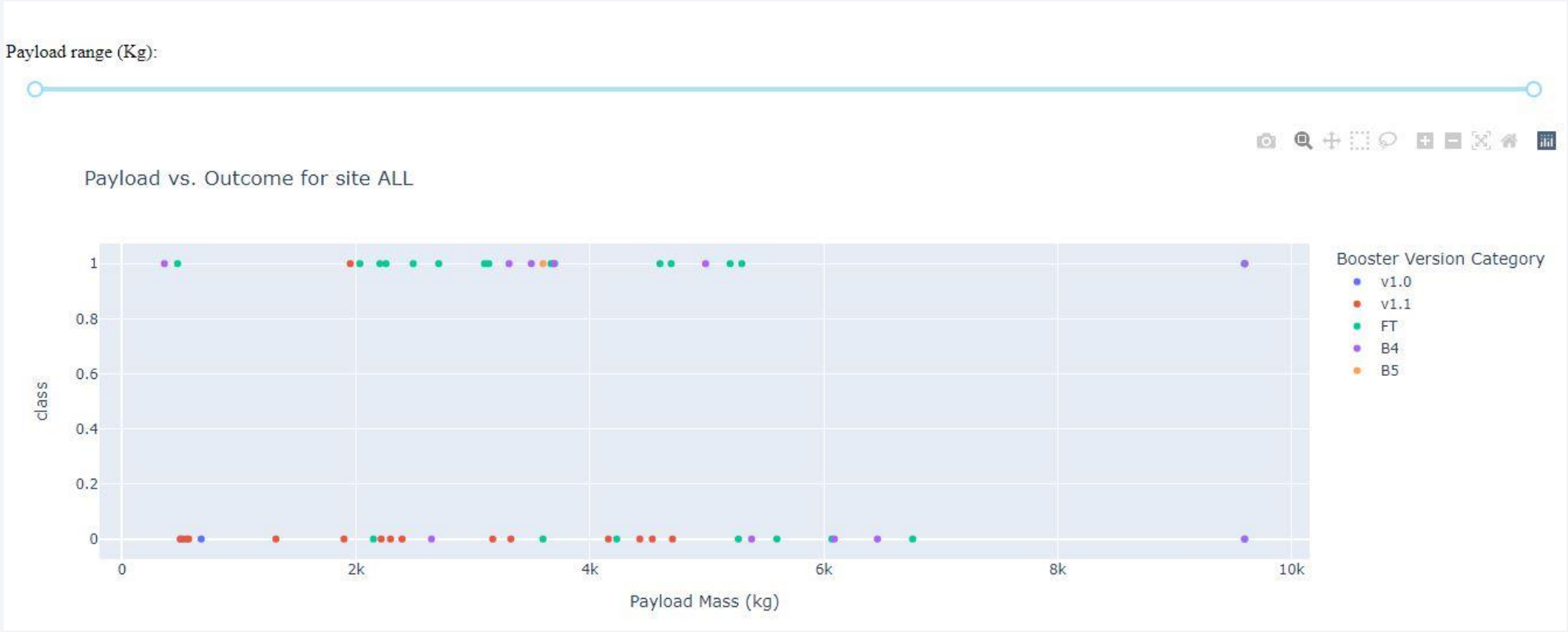
Pie-chart for the launch site with highest launch success ratio

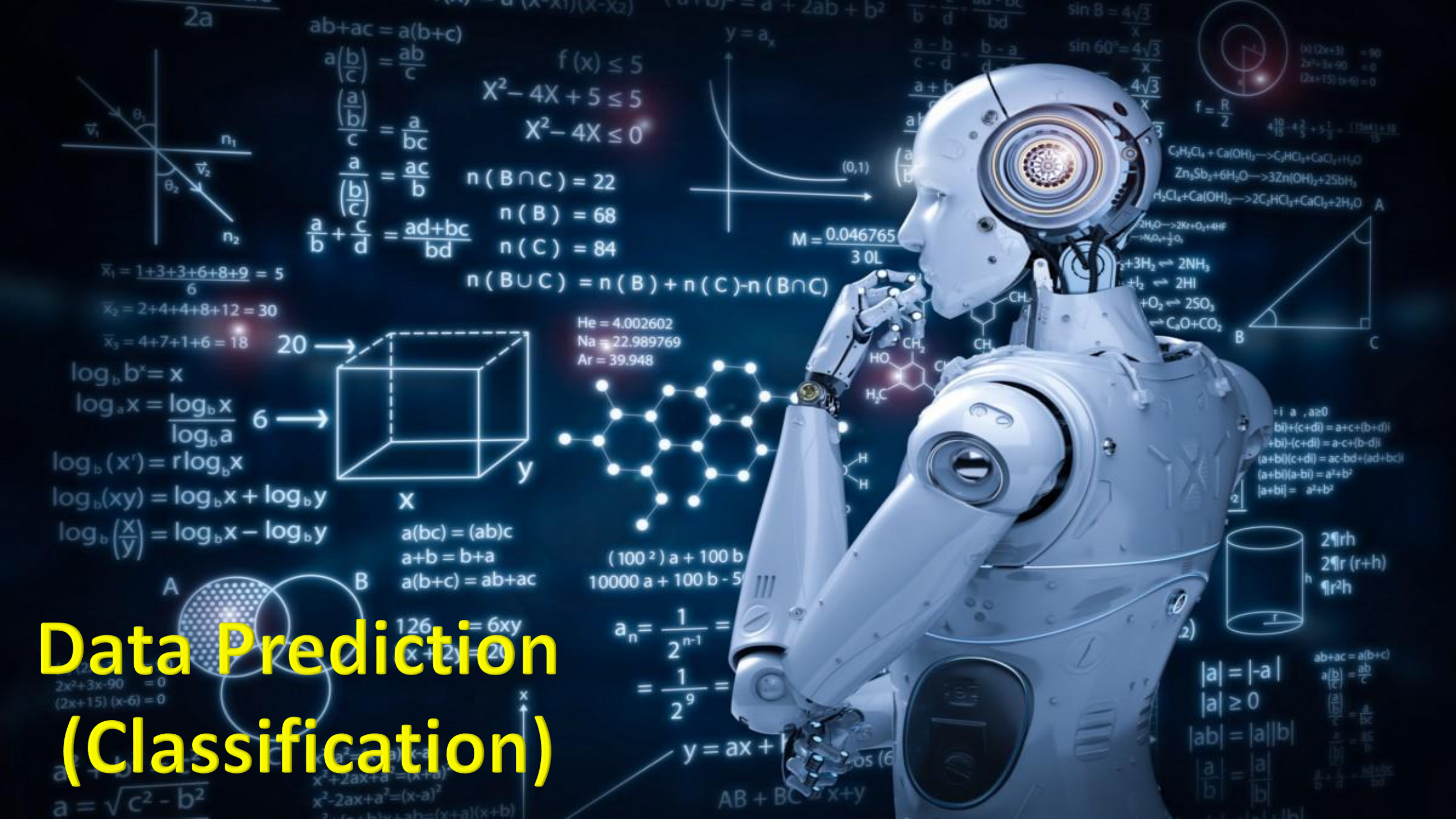
Launch site CCAFS SLC-40 has highest success ratio.



Scatter plot between Payload and Launch outcome for the all sites

Some Booster version FT having payload mass(kg) greater than 5k has highest success rate and Booster version B5 having payload mass(kg) greater than 3k has high outcome but success rate is low.




Data Prediction
(Classification)

Classification Accuracy

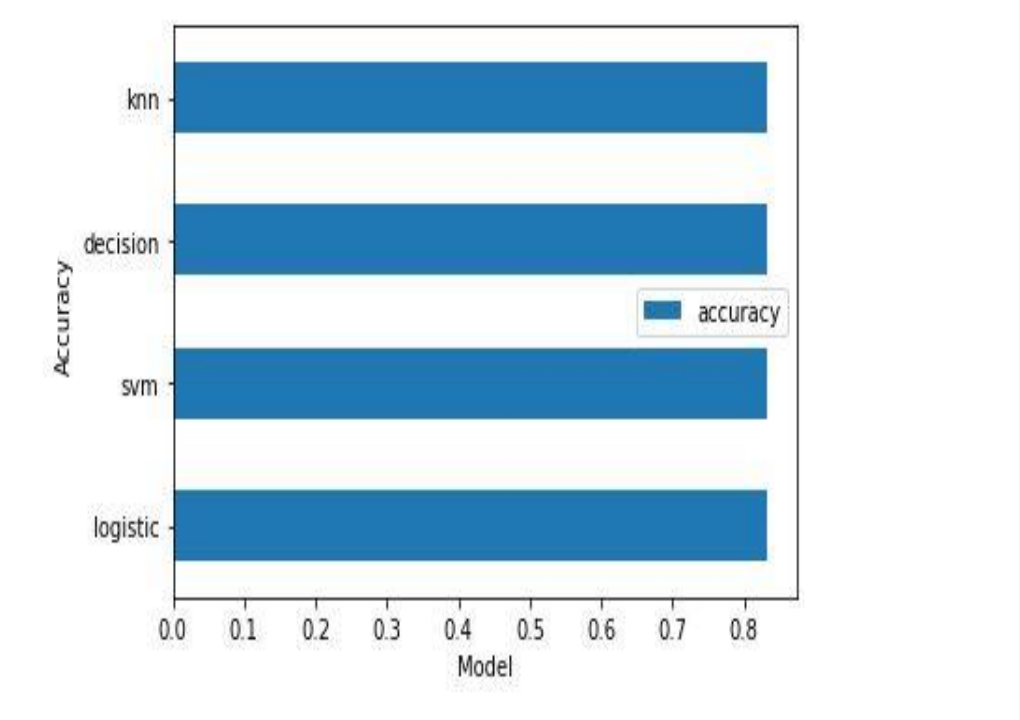
Visualize the built model accuracy for all built classification models, in a bar chart

There are four models are used for analysis:

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K nearest Neighbors

All have same accuracy score like 0.8333333334

These models have good performance for predicting the success rate.



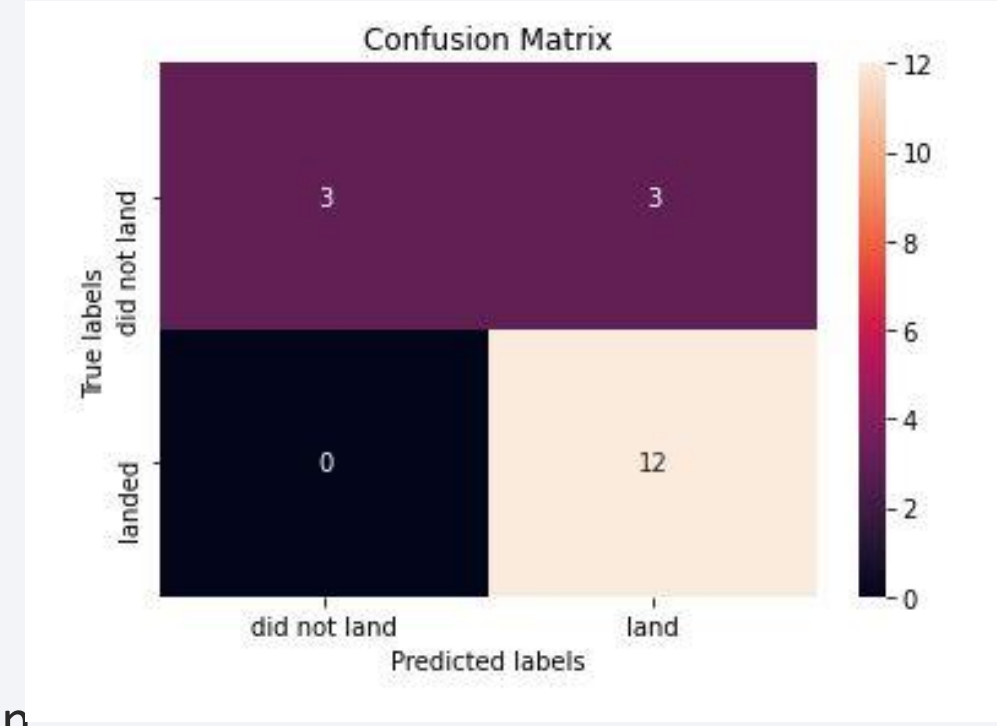
Confusion Matrix

Confusion MATRIX of model :

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K Nearest Neighbors

In prediction analysis all model shows that

the accuracy score are same i.e. all are good in performance for prediction of test data and confusion matrix of these model are same.



Conclusions

- Launch site CCAFS SLC 40, there are many flights launched successfully and rapidly increases with increase flight number.
- Orbit SSO, HEO, GEO, and ES-L1 have high success rate and VLEO orbit has the second highest success rate.
- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- If the mass of payload decreases, the performance of success rate will increase but success rate decreases if very high mass of payload is used.
- All launch sites are not so far to each other except the launch site VAFB SLC-4E. This is far and situated in west coast and CCAFS SLC-40, CCAFS LC-40, AND KSC LC-39A are in east coast.
- Launch sites are in close proximity to railways.
- Launch sites are in close proximity to highway.
- Launch sites are in close proximity to coastline.
- Launch sites keep certain distance away from cities
- In prediction analysis all model (logistic regression, svm, decision tree and k nearest neighbor) shows that the accuracy score are same i.e. all are good in performance for prediction of test data and confusion matrix of these model are same.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!