IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Subhrajit Chaudhuri
23-Nov-2022

Outline

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

Executive Summary

- Summary of methodologies
  - The necessary data was collected by web scraping from Wikipedia page, and SpaceX API.
  - Date wrangling, exploratory data analytics(EDA),
  - data visualization ,and interactive visualization was done.
  - Machine learning models were used for prediction.
- Summary of all results
  - EDA and interactive visualization were used to identify most important features relating to the success of launches.
  - Machine Learning models predicted the success of launches based on features.

Introduction

- **Project background and context**
- Falcon 9 rockets are set of low cost rockets provided by SpaceX. They offer launches for 62 million dollars while cost from other providers is more than 165 million dollars each. This reduction in price comes from the fact that SpaceX reuses first stage of the rocket by re-landing the first stage. By determining if the stage will land, we can determine the cost of a launch. This information is interesting for another company if it wants to compete with SpaceX for a rocket launch. Repeating this process will make the price down even further.
- **Problems you want to find answers**
- What are the main characteristics of a successful or failed landing ?
- What are the effects of each relationship of the rocket variables on the success or failure of a landing ?
- What are the conditions which will allow SpaceX to achieve the best landing success rate ?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from SpaceX REST API and by web-scrapping from Wikipedia page.
- Perform data wrangling
  - Dropping unnecessary columns.
  - Dealing with missing values.
  - One hot encoding was done for categorical data.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

- Perform predictive analysis using classification models
    - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

- Describe how data sets were collected.

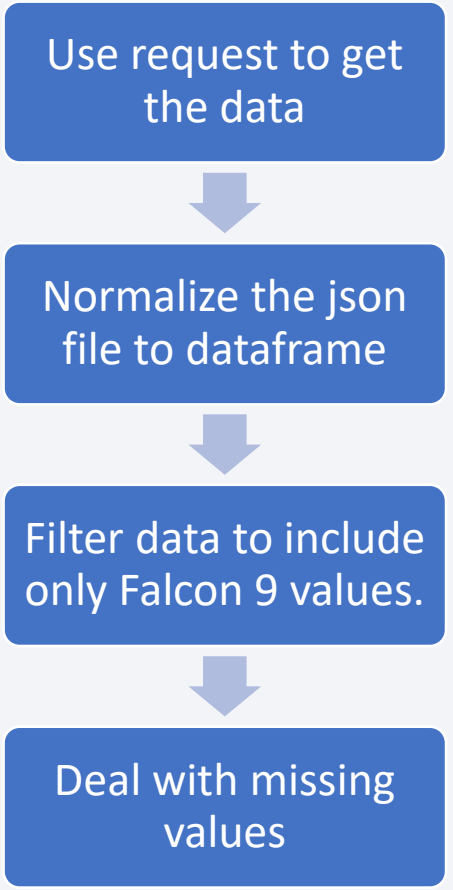
Data was collected by two methods. First set of data was connected from SpaceX API(https://api.spacexdata.com/v4/) via get requests. Second set of data was collected from Wikipedia page of Falcon 9 and Falcon Heavy launches (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches). Both sets of data were used for the analysis.

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts is given.

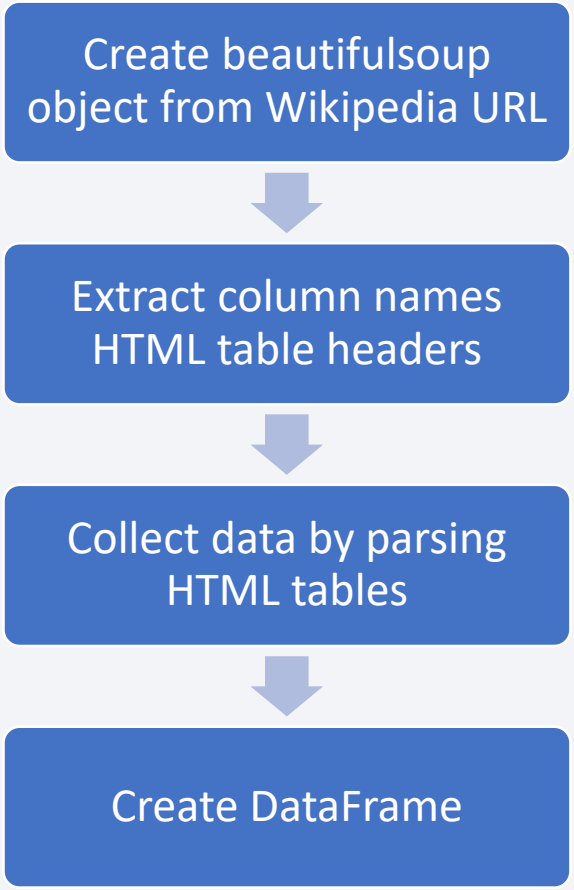
- Source code: Github Code Link



Data Collection - Scraping

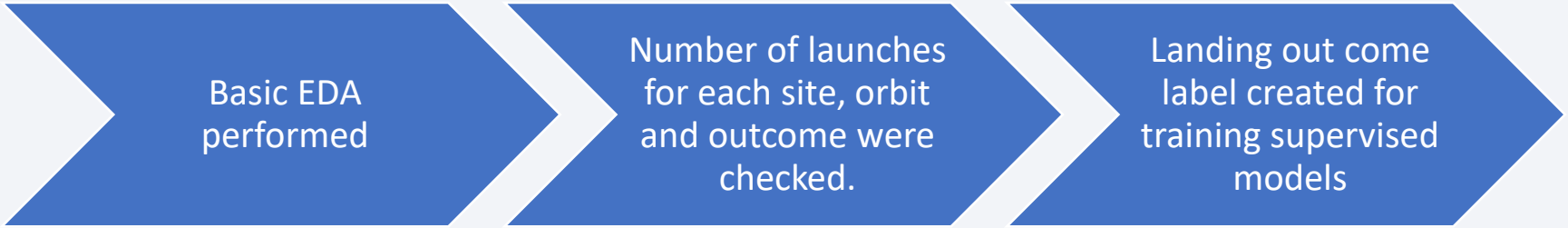
- Flow chart for data collection using web scraping process is given

- Source Code: Github Code Link



Data Wrangling

- In data wrangling data is analyzed , cleaned, sorted for further analysis and model creation. In this case first basic EDA was performed on the data. Number of launches based on site, orbit, outcomes were studied. Finally, landing outcome label was created for training supervised model.



- Source Code: Github Code Link

EDA with Data Visualization

- The following charts were plotted
    - Flight Number and Launch Site(scatter), Payload and Launch Site(scatter), success rate of each orbit type(bar), Flight Number and Orbit type(scatter), Payload and Orbit type(scatter), launch success yearly trend(line).
- The scatter plots were helpful to visualize relations between two features, bar plots were helpful to compare between discreet variables, and line plot was use to compare trend over time.

- Source code: Github Code Link

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 achieved.
  - 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 records which will display the month names, failure landing-outcomes in drone ship ,booster versions, launch-site for the months in year 2015
  - Rank the count of successful landing-outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- Source Code: Github Code Link

Build an Interactive Map with Folium

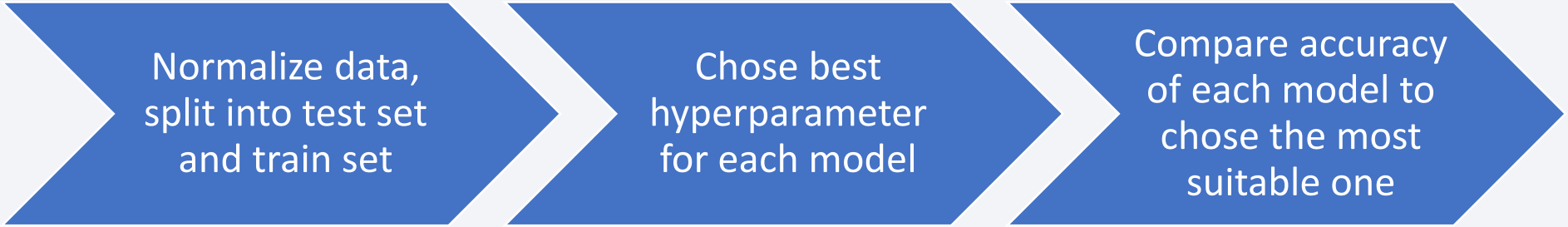
- Circles and markers added for NASA Johnson Space Centre and the launch sites. The markers contain text label denoting the name of the sites.
- Marker clusters were added for each launch sites denoting successful and unsuccessful landing outcomes.
- Lines were added to mark distance from launch sites to key locations.
- The objects were added analyze importance of locations and close proximities of the
- Source Code: Github Code Link

Build a Dashboard with Plotly Dash

- First part of the dashboard shows total success and failure for the launch site chosen in drop down or all sites in a pie chart.
- In the second part a range slider was used to select pay load mass range and successful and unsuccessful landings were shown in a scatter plot along with their version.
- The dashboard enables users to interactively analyze and visualize successful and unsuccessful landings based on location, payload mass, booster version.
- Source Code: Github Code Link

Predictive Analysis (Classification)

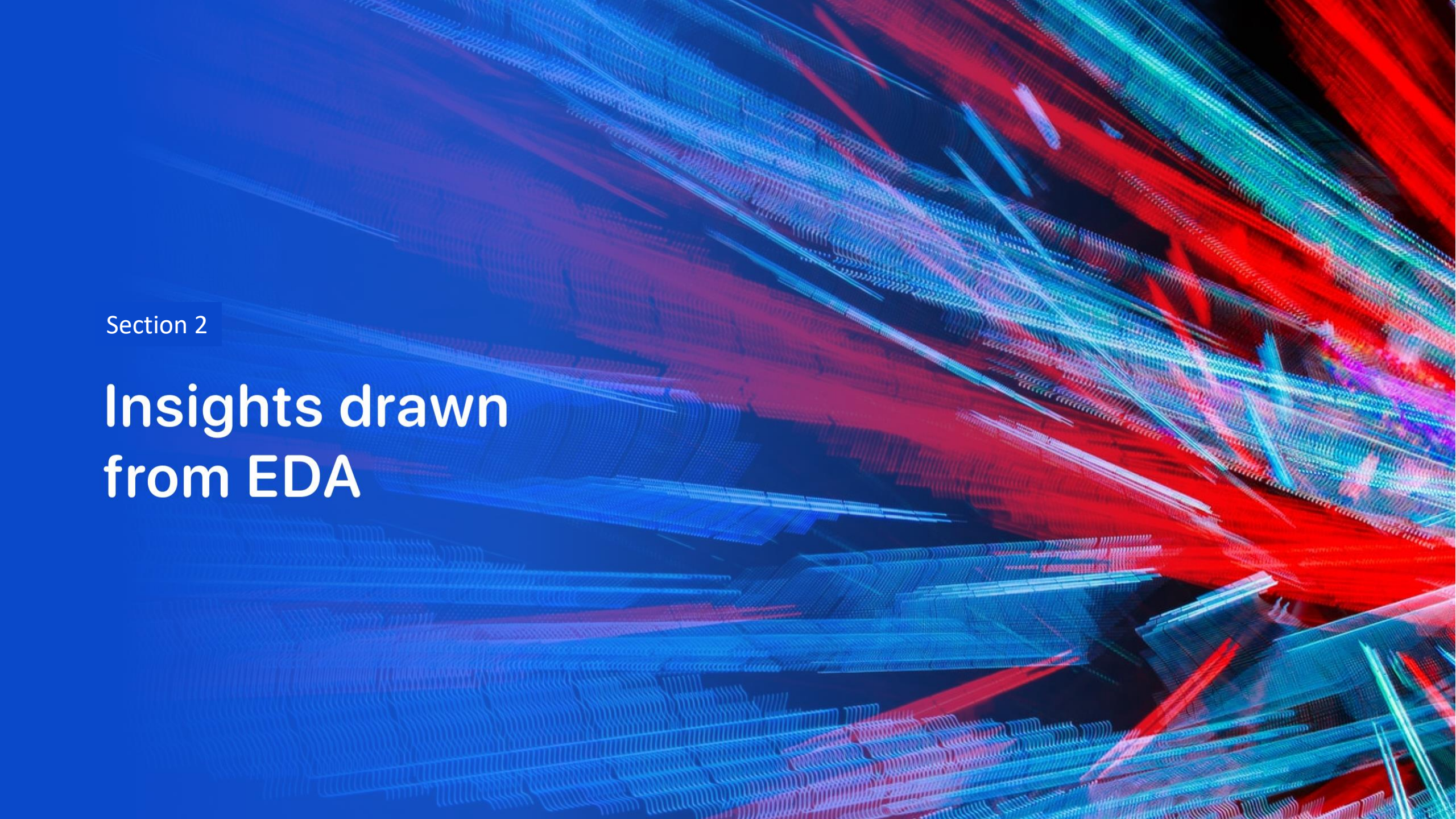
- The data were normalized and split into test train sets. Logistic regression, support vector machine, decision tree and k nearest neighbors classification models were used. Best hyperparameters were chosen for each model by using GridSearchCV. Accuracy of each model were compared to chose the best model.



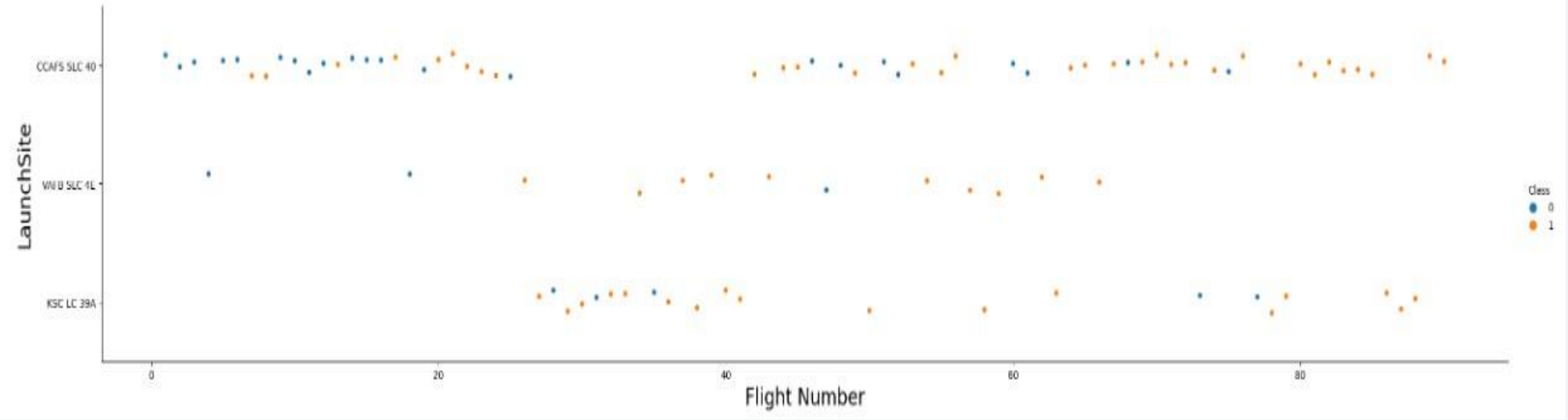
- Source Code: [Github Code Link]

Results

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

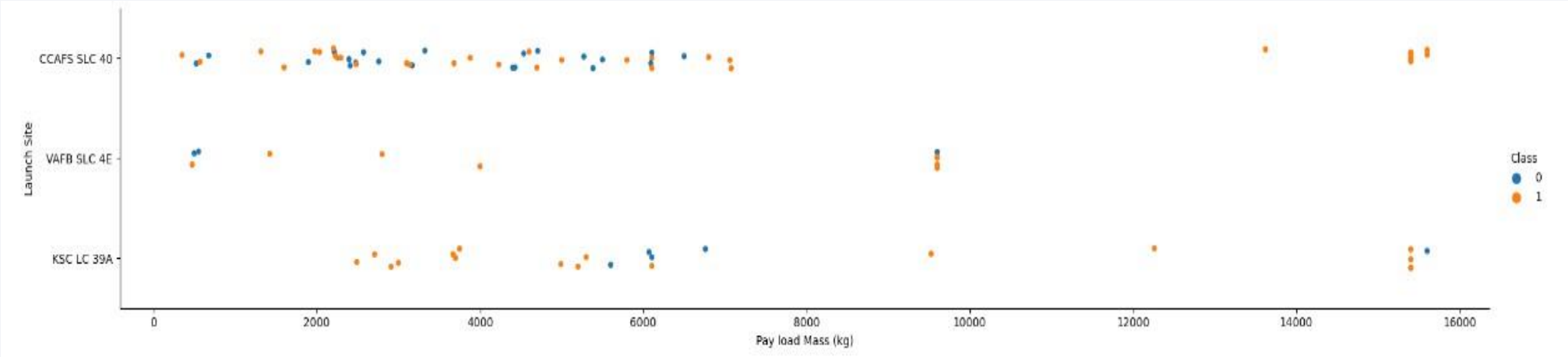

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



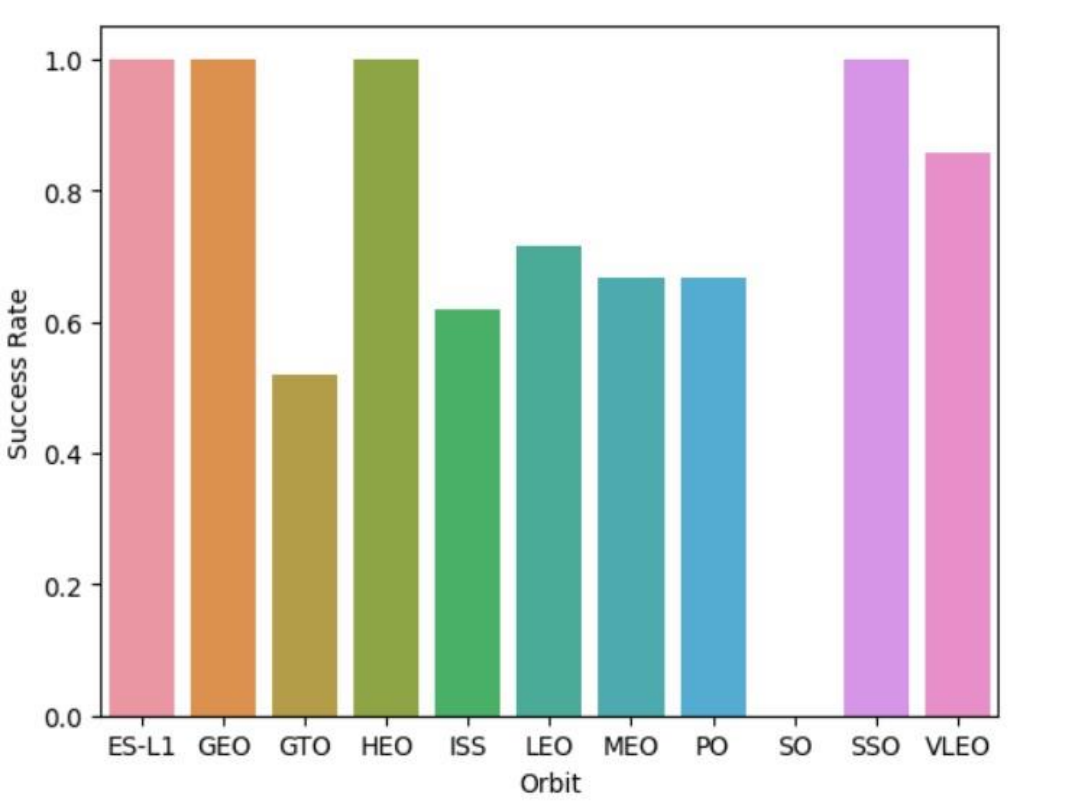
- It can be seen were launched from CCAFS SLC-40 site specially in the earlier years. Most of the earlier flights ended in failure.

Payload vs. Launch Site



- Most of the payload weigh less than 8000kg but higher success rate can be observed for payload weight more then 8000kg

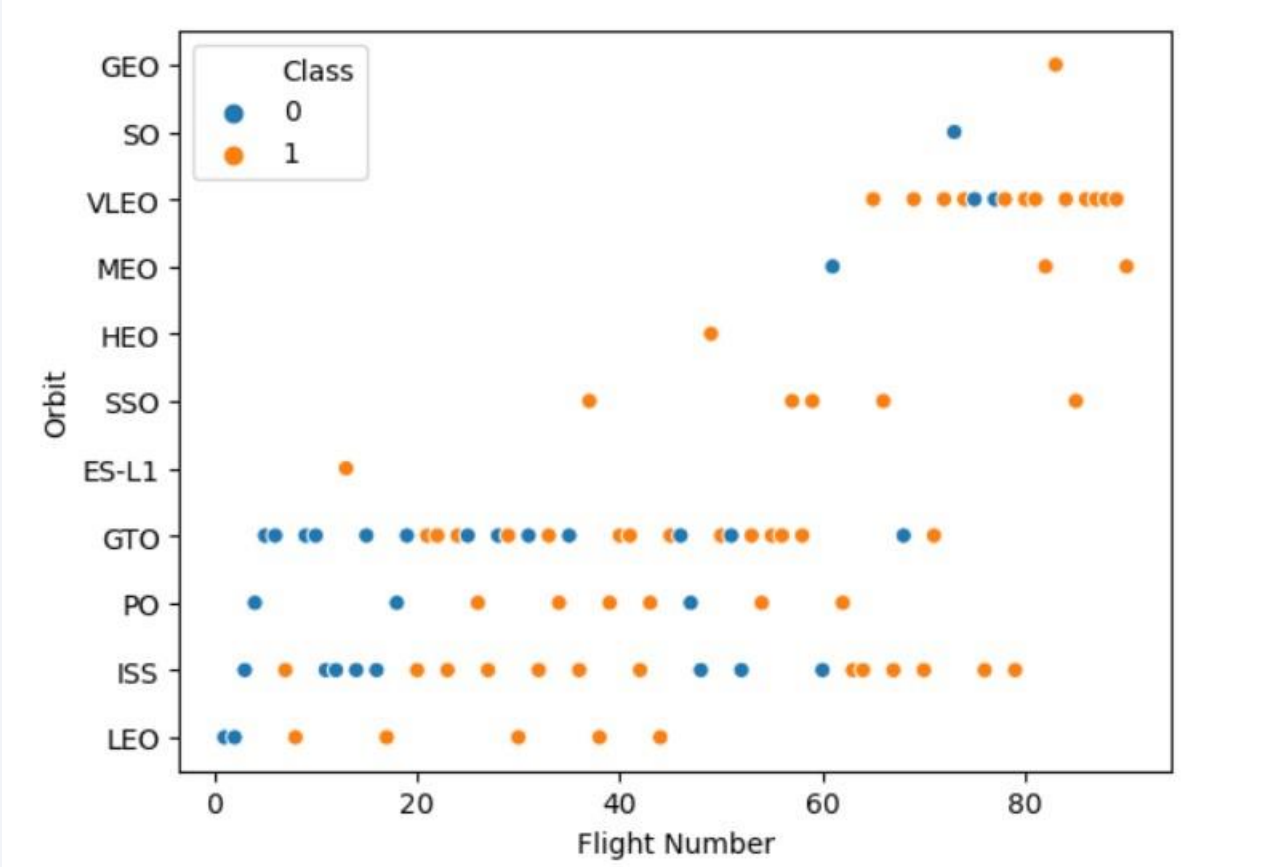
Success Rate vs. Orbit Type



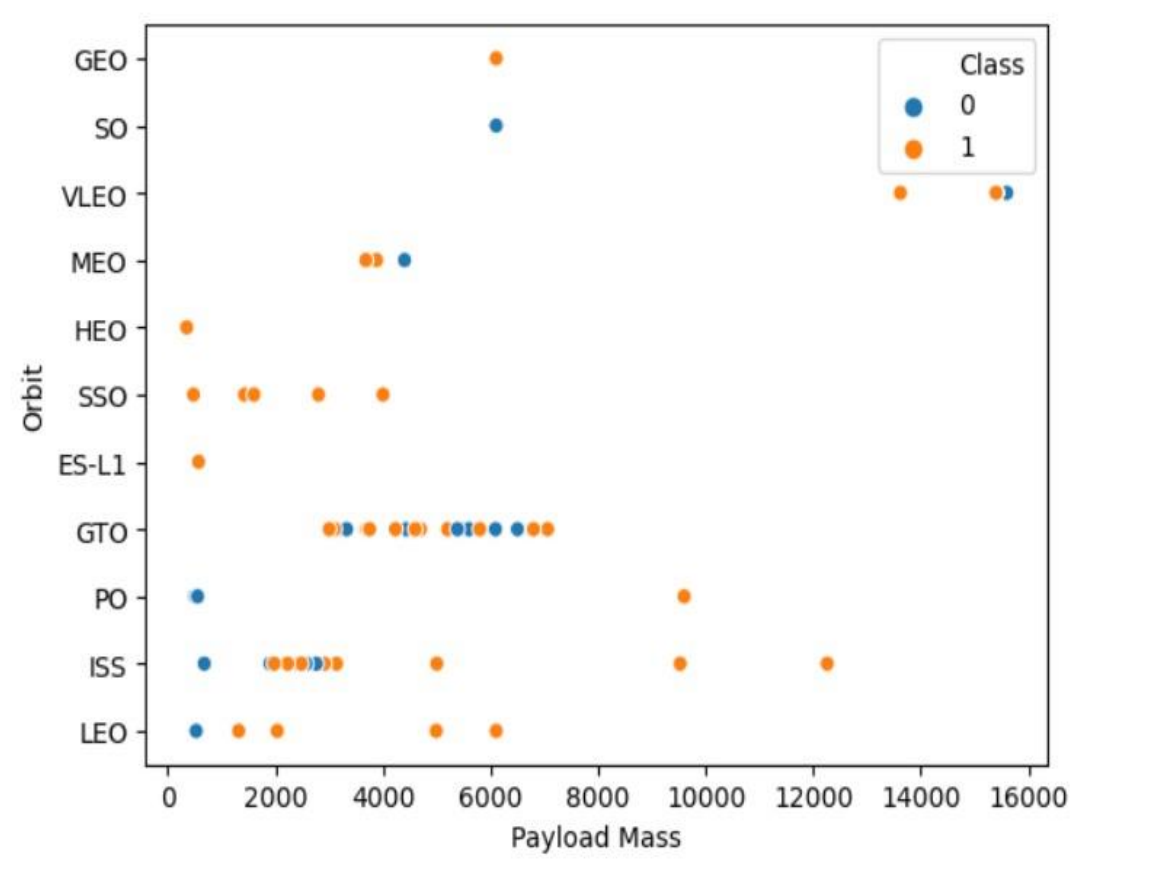
- Orbits ES-L1,GEO,HEO,SSO have success rate of 1 meaning all launches have been successful; while orbit SO has no successful launches.

Flight Number vs. Orbit Type

- Initially most launches were limited to GTO,PO,ISS,LEO orbits. Later on flights were aimed for VLEO ,SO ,GEO ,MEO ,SSO orbits.
- GTO and ISS are most used.

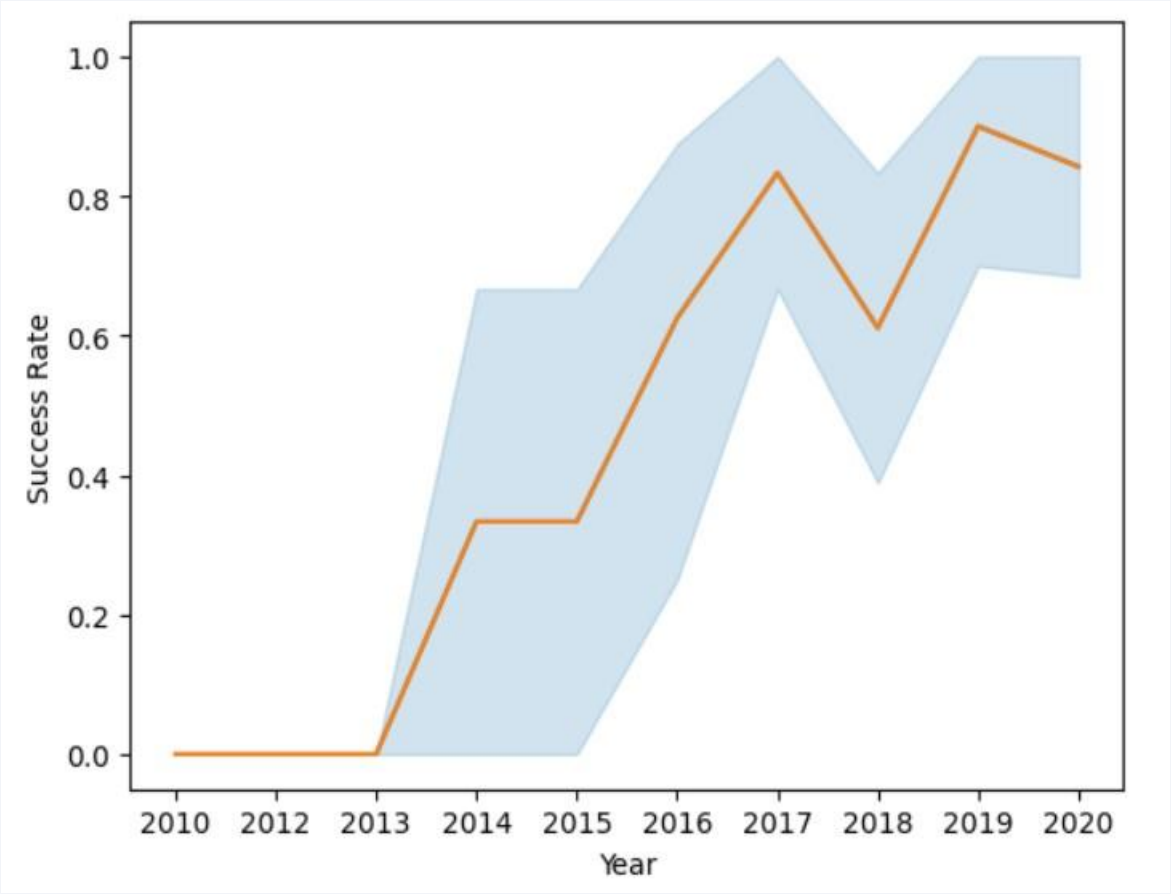


Payload vs. Orbit Type



- Higher payload mass launches were used in PO, ISS, VLEO orbits.
- ISS,GTO have most range in payload use.

Launch Success Yearly Trend



- Sharp increase in success rate is observed from 2013.
- Thereafter success has increase with only dip in 2018.

All Launch Site Names

- There are 4 unique launch sites as per data. The names of launch sites obtained from data with SQL query is give below.

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

Launch Site Names Begin with 'CCA'

- 5 records with launch site name beginning with CCA are given

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters is 45596kg, result as obtain with SQL query.

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2928.4kg. Result obtained with SQL query given below

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- First successful landing outcome on ground pad is 01-05-2017. min function was used to obtain oldest date for successful launch.

min(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- The list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are given below.

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

Total Number of Successful and Failure Mission Outcomes

- The total successful and failure mission outcomes are 100 and 1 respectively.

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

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass are given below as obtained from data set using SQL query.

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

- There were two failed landing_outcomes in drone ship in 2015. Their booster versions, and launch site names given below.

MONTH	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

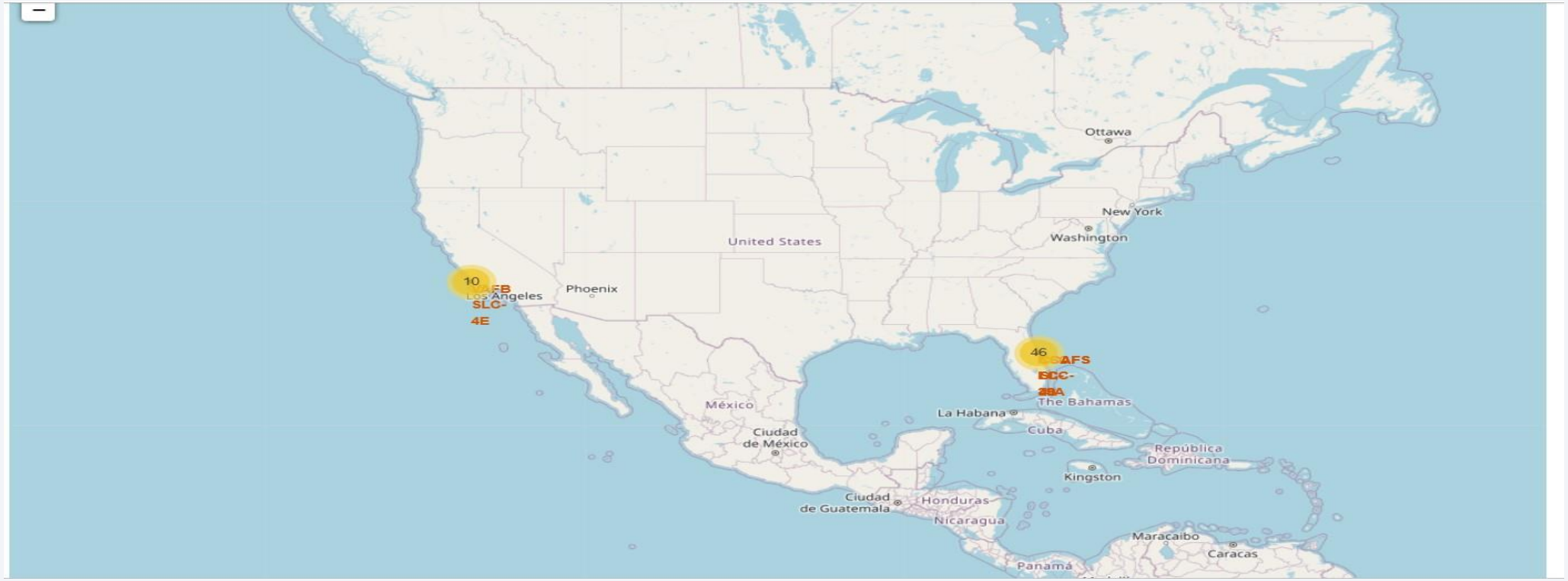
- The count of successful landing outcomes between the date 2010-06-04 and 2017-03-20 ranked in descending order

Landing _Outcome	COUNT("LANDING _OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

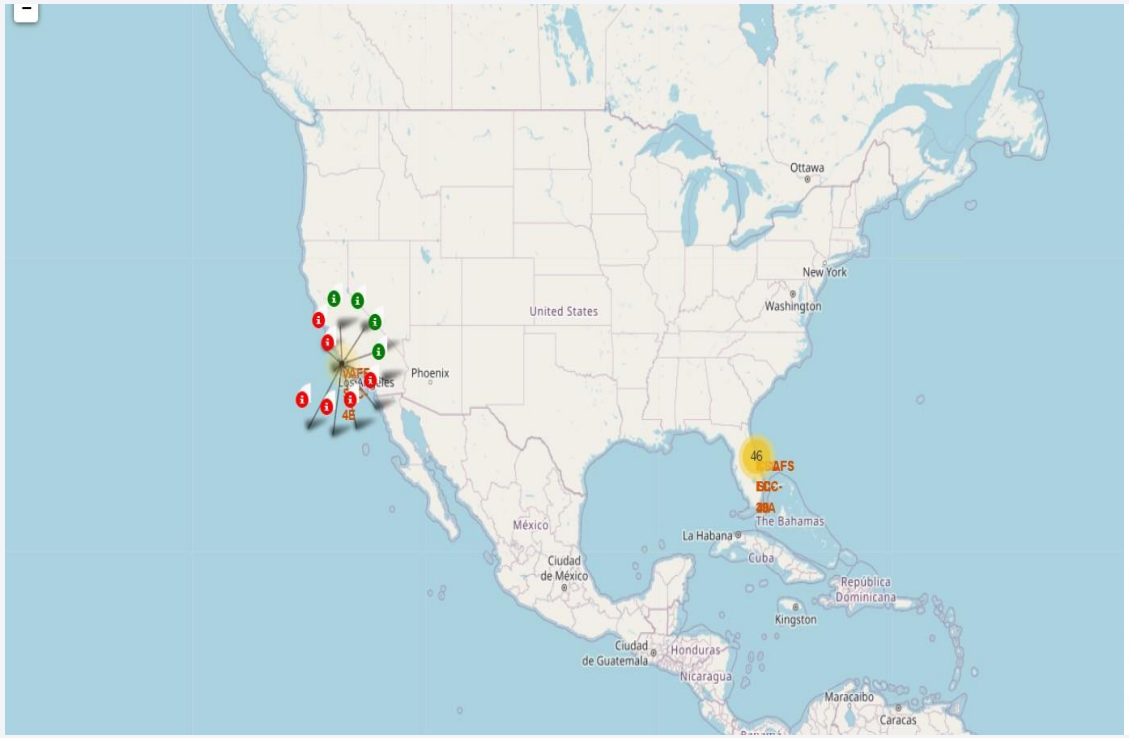
Launch Sites Proximities Analysis

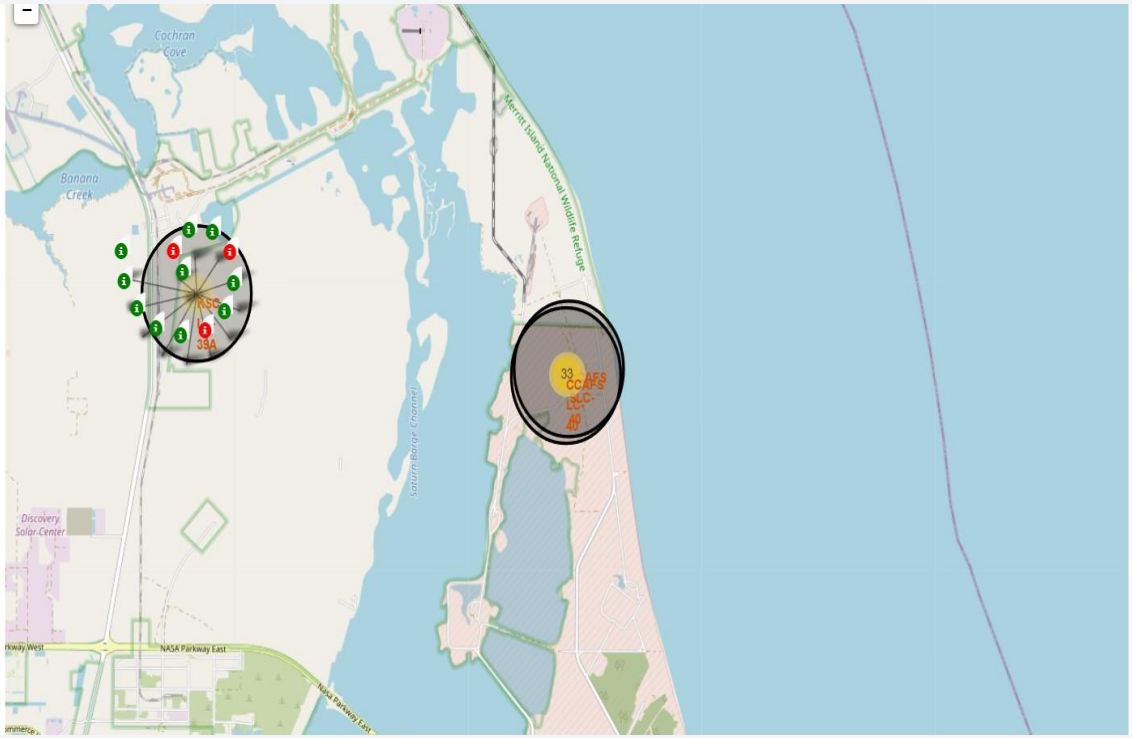
Launch Site Locations



- Launch Site locations are shown in the map. It is clearly seen that all the launch sites are near to the oceans.

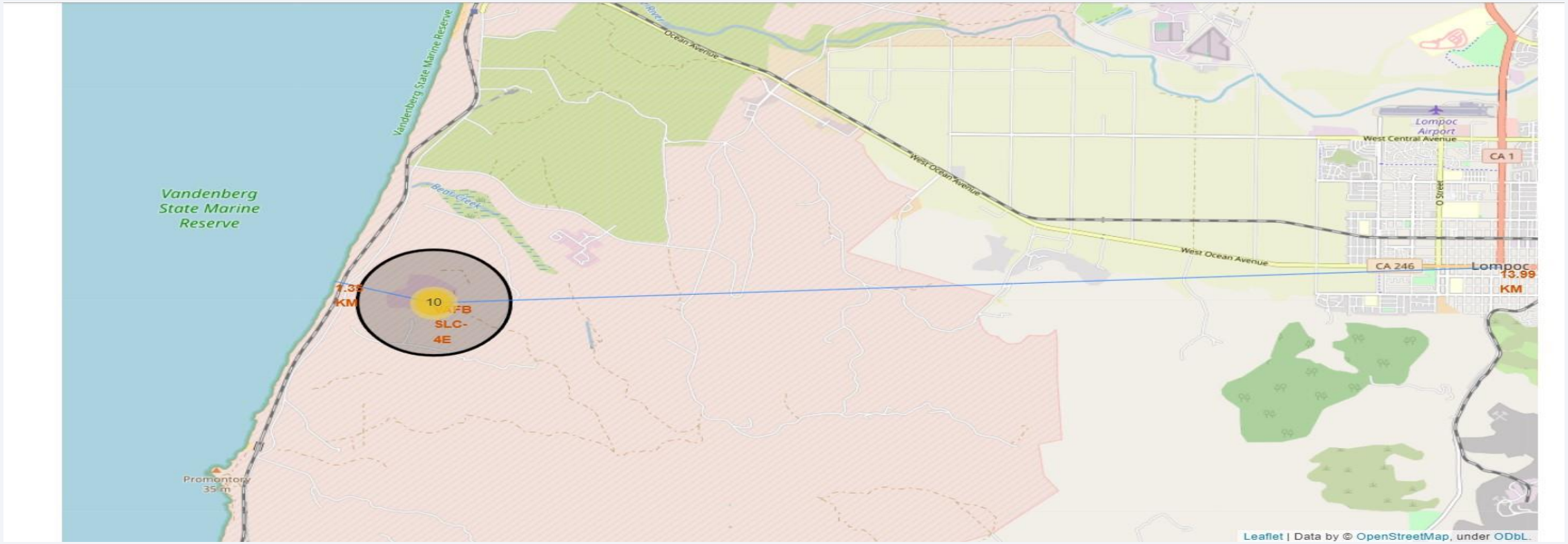
Launch Site Color Labeled Marker



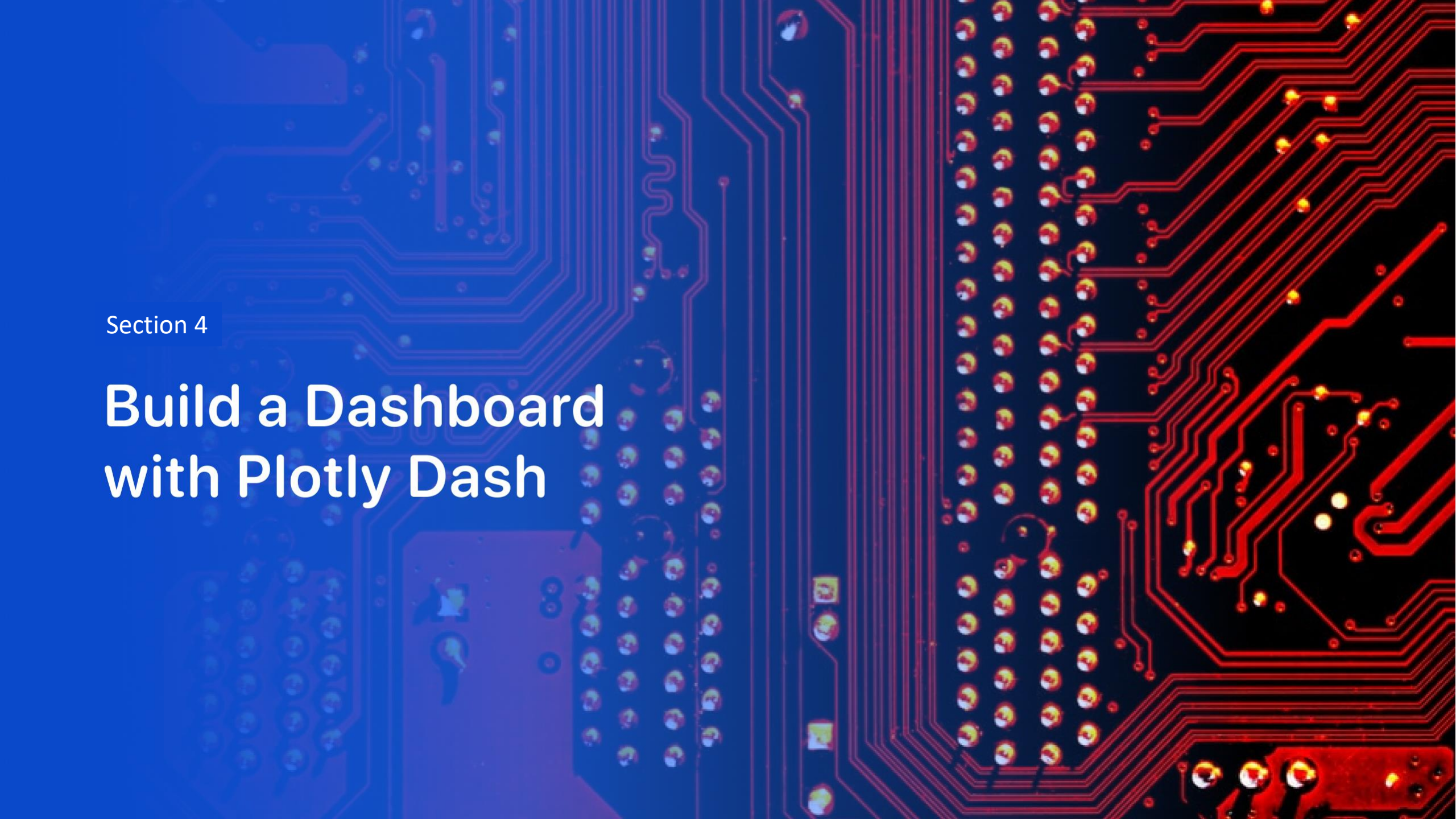


- Launch Site are shown with color labeled markers. Green color indicating successful launches , red indicating failures.

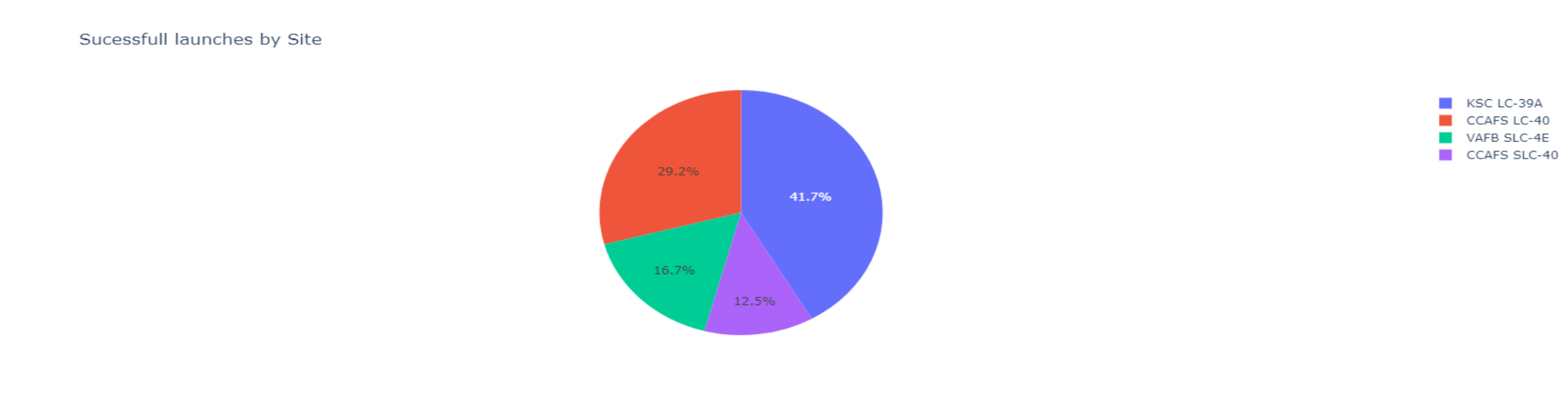
Launch Site Proximity in Map



- VAFB SLC-4E launch site distance from coastline, and from nearest city is shown. Both of which are near to the launch site

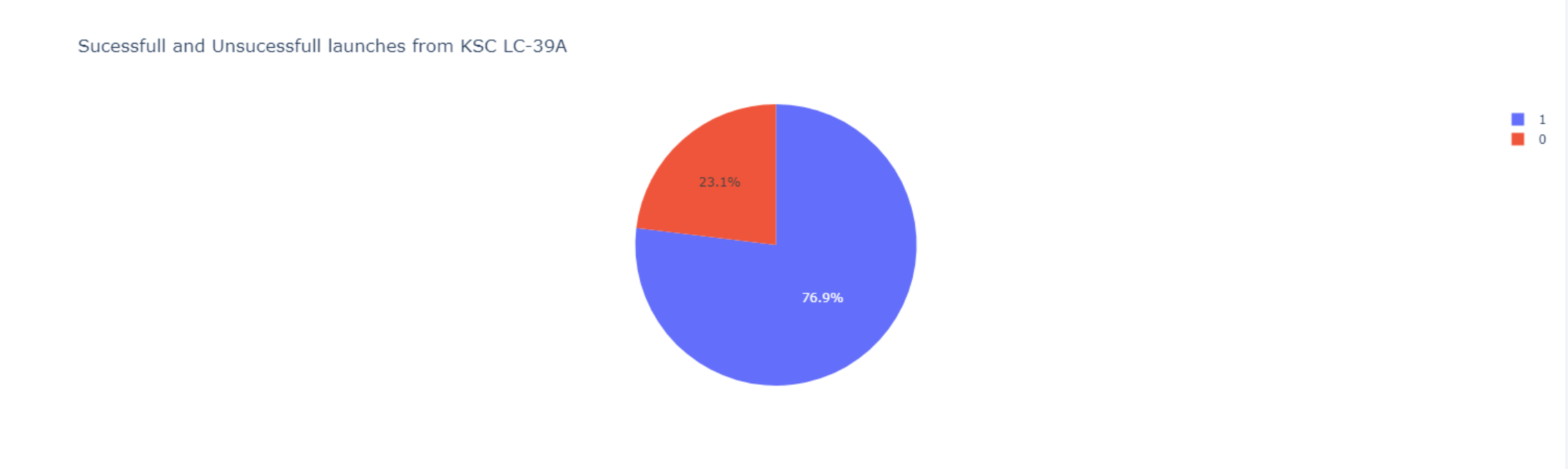

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



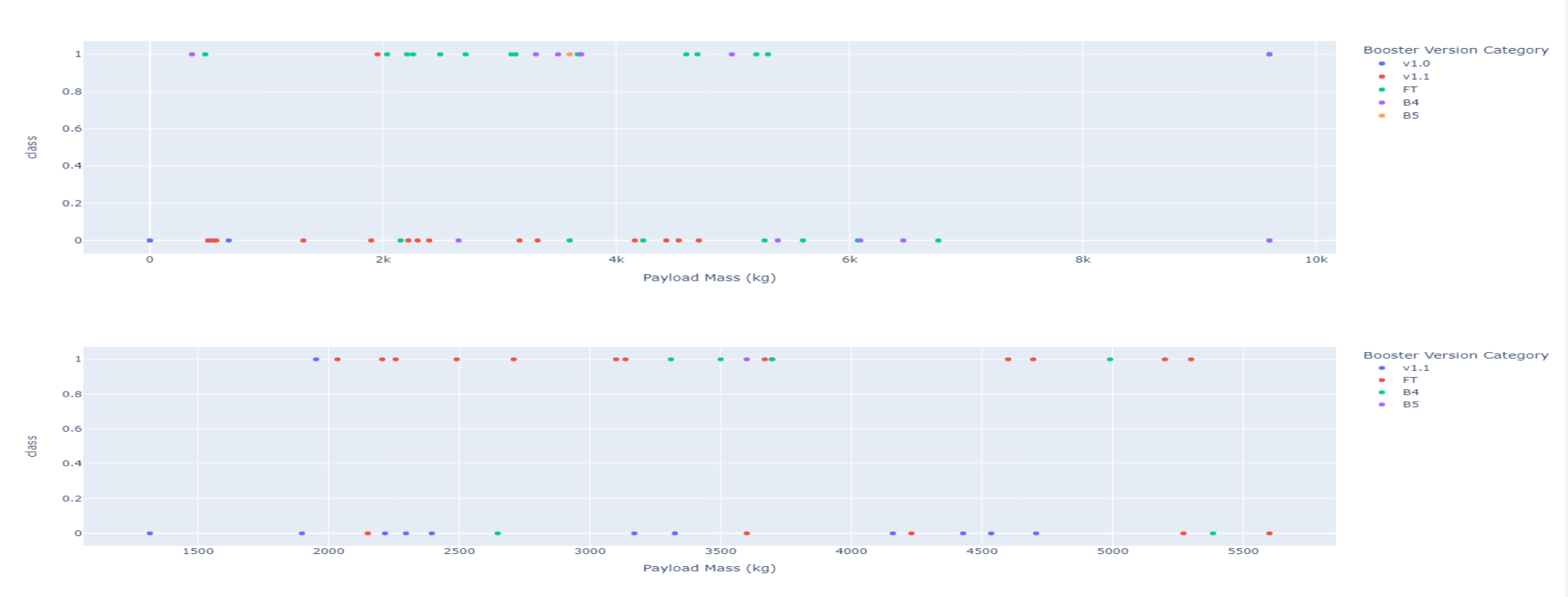
- It is clearly seen that most successful launches is done from KSC LC-39A site

Successful and Unsuccessful Launches from KSC LC-39A



- It is seen that success percentage of launches from KSC LC-39A is 76.9%

Payload Mass vs Outcome for different scales

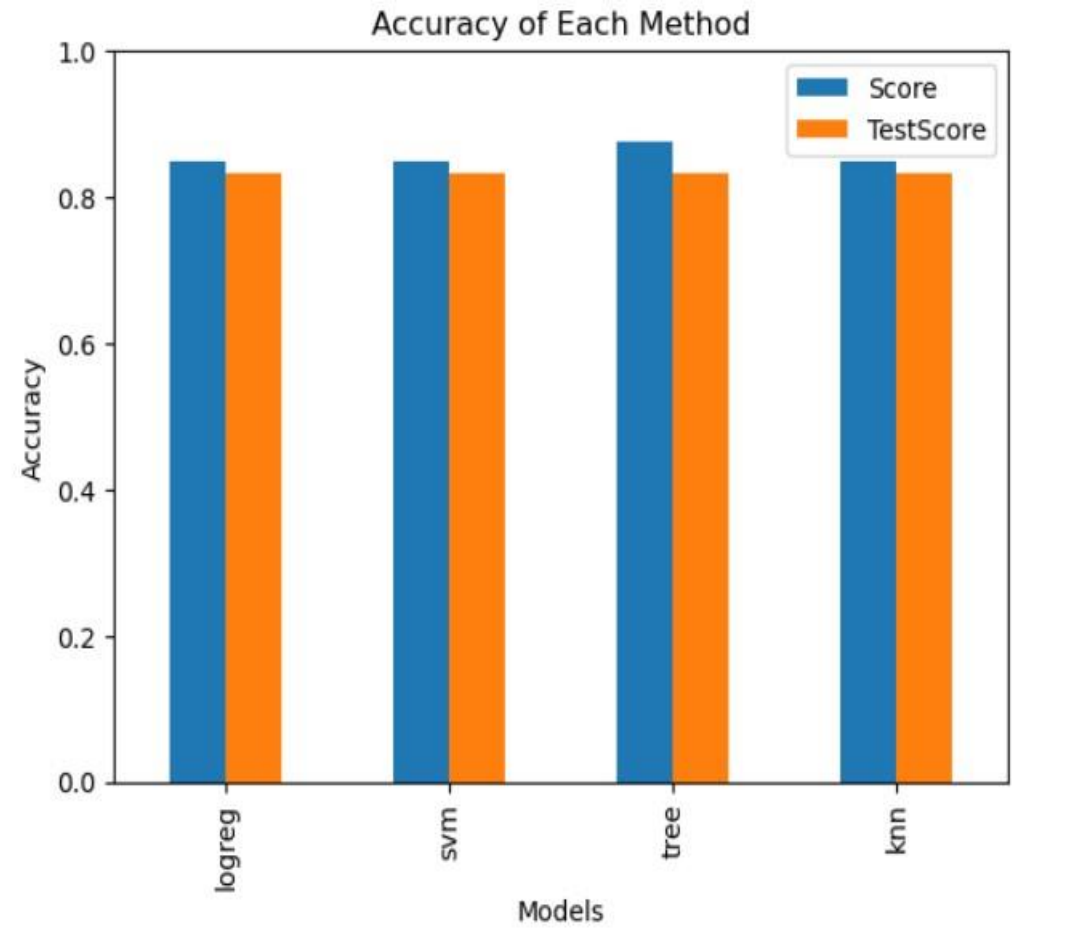


- The first scatter plot shows outcome for the whole payload mass range , while the second plot show the same for 0 to 5500 kg range

Section 5

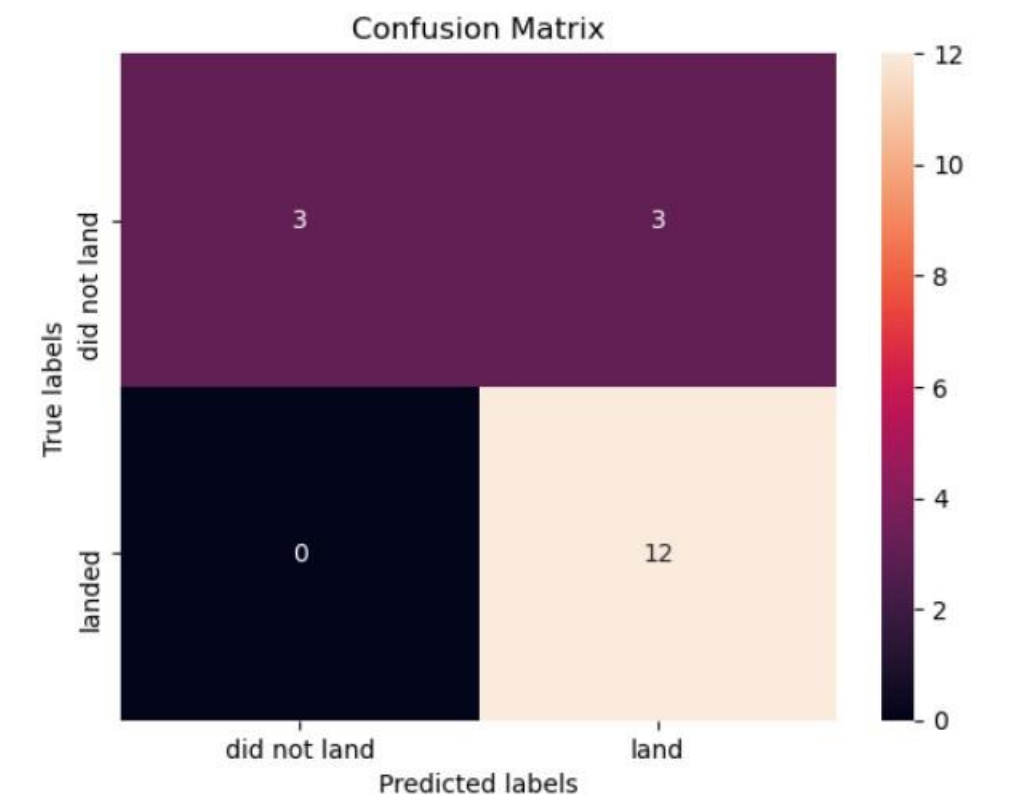
Predictive Analysis (Classification)

Classification Accuracy



- While the test score of all the models are same, the training score is highest for Decision tree model.
- Training set accuracy of decision tree model is 87.5%.

Confusion Matrix



- The confusion matrix shows that it predicted 3 times to land when the flight actually did not land. In all other cases it predicted well

Conclusions

- Success of landing is influence by factors like payload mass, orbit type, launch site, booster version. Several other factors also may be at work.
- There has been steep rise in success rate since 2013, may be due to technology upgradation and experiences gained from previous launches.
- It is observed that KSC LC-39A has the highest success rate among the sites.
- Higher payload mass has better landing outcomes
- ES-L1,GEO,HEO,SSO orbits have the highest success rates

Appendix

- Link for the github repository : Link
- Kindly view the ipynb files in nbviewer.org to view them completely.

Thank you!