Winning Space Race with Data Science

Yaksh Shah

7nth May 2022

Outline

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

Executive Summary

- Summary of methodologies

The Launch Data Was Collected using Web Scrapping and fetching data from SPACEX API. Than the Data Was Wrangled to take care of NULL values in the data. After that EDA was done using Dash, Folium Maps and Seaborn Libraries to fetch insights from the data. Using the fetched insights ML algorithms were used to predict the outcome of Rocket Launch on the basis of different Parameters that affected the outcome of the launch.

Summary of all results

From EDA we obtained the Launch Site wise Success rate and obtained the dependency of Flight number and Success Rate for each Launch site. Got to know the relationship of payload vs Launch Site and also analyze the relationship between Payload and Flight Number vs orbit type . ML algorithm showed better prediction of Success of the Launch.

Introduction

- Project background and context

To determine the Price of each launch by making the Prediction of the Successful landing of the First Stage of the Falcon 9 Rocket . so, that it can reduce the cost of launch and can be reused for the launch of the another rocket

- Problems you want to find answers

To determine whether SPACEX will reuse the First stage of rocket or not.

Section 1

Methodology

Methodology

Executive Summary

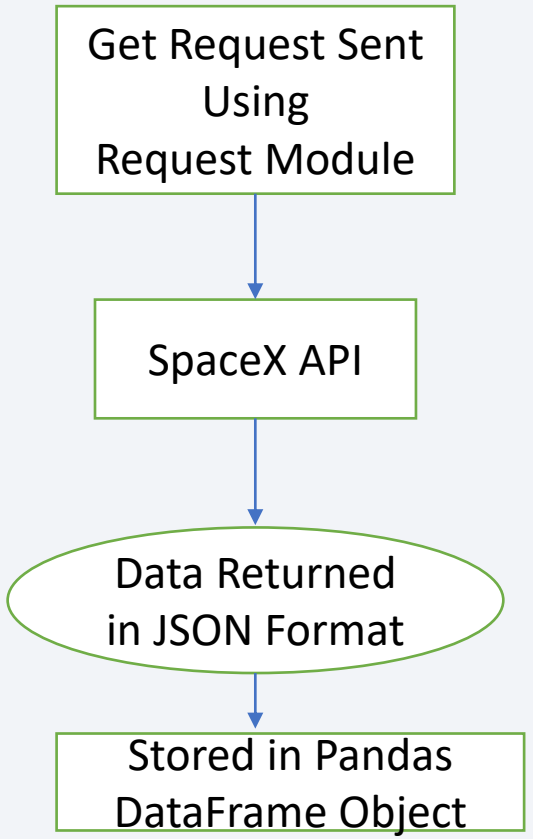
- Data collection methodology:
  - Data was collected using SPACEX API and By Scraping the Falcon 9 launch data Provided on Wikipedia and the data was filtered to get only Launch data of Falcon 9
- Perform data wrangling
  - Data was used to get the number of NULL or missing values in each column and was replaced by relevant value.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data was collected using SPACEX API and By Scraping Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches
- For Web scrapping Beautiful Soup and Request Libraries were used.
- From the API JSON data was fetched using Request module and stored into pandas DataFrame object.

Data Collection – SpaceX API

- Here is the Representation of the Data Collection Process From SPACEX API

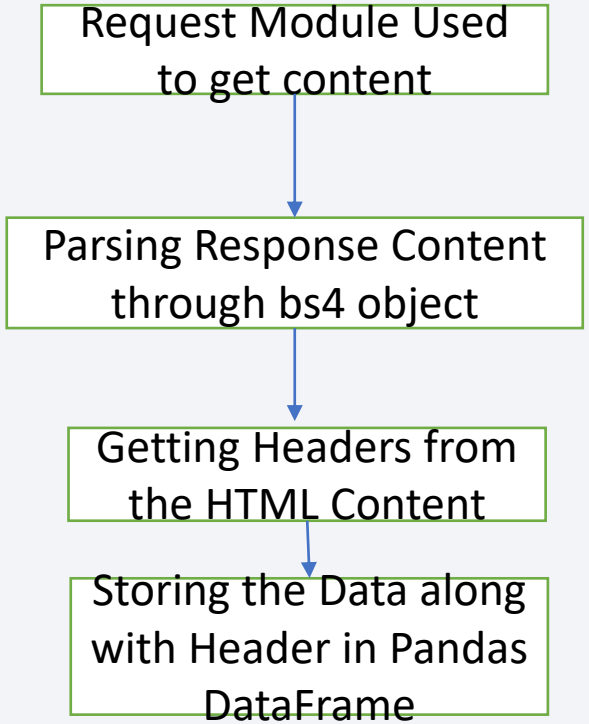


- GitHub URL

IBM-Data-Science-Capstone-Project/Data Gathering and Wrangling.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Data Collection - Scraping

- Flow Chart of Data Collection using Web Scraping of Table Data From Falcon 9 Launch Data on Wikipedia



- GitHub URL

IBM-Data-Science-Capstone-Project/Data Gathering and Wrangling using Web Scraping.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Data Wrangling

- The Data Obtained from the Wikipedia had some unrequired data so we ignored it and just obtained the Launch Data of Falcon 9 Rocket.
- We had to first obtain the column names and than the corresponding data.
- Than Search for the missing or NULL values from the data was carried out and than the necessary operations were performed to remove them from the data.
- GitHub URL:
- IBM-Data-Science-Capstone-Project/Data Gathering and Wrangling using Web Scraping.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)
- IBM-Data-Science-Capstone-Project/Data Gathering and Wrangling.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

EDA with Data Visualization

❖The Scatter Plots Were used to draw insights for following relationships

- Flight Number vs Payload Mass, Payload Mass vs Launch Site, Flight Number vs Launch Site, Flight Number vs Orbit type, Payload Mass vs Orbit Type.
- Scatter Plots were used as we can get the relationship between two parameters and also success rate very easily compared to other charts.

❖The Bar Chart Were used to draw insights for following relationships:

- Success Rate vs Orbit Type
- Bar Chart was used because its very easy to get Orbit type which had the highest success rate of Landings compared to other orbit

❖Line Chart for Year vs Success Rate was used to visualize Year wise Increase in Success rate of Landings

URL: IBM-Data-Science-Capstone-Project/EDA With Data Visualization.ipynb at main · shahyaksh/IBM-Data-Science-Capstone- Project (github.com)

EDA with SQL

- There are 4 Launch Sites.
- The total Payload mass carried by the boosters launched by NASA CRS is 45596 Kg
- Average Payload Mass carried by booster Falcon 9 v1.1 is 2928.400 Kg
- On 22-12-2015 the first successful landing in ground Pad was achieved
- 4 boosters have success in drone ship and have payload mass greater than 4000 but less than 6000
- There are 100 successful Mission outcomes and 1 failed Mission outcome
- There are 12 Boosters which have carried maximum payload mass
- There are 5 Failed(drone ship) Landing Outcomes and 3 Success (ground pad) Landing Outcomes between 2010-06-04 and 2017-03-20

URL: IBM-Data-Science-Capstone-Project/jupyter-labs-eda-sql-coursera.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Build an Interactive Map with Folium

- All the Launch site were marked with Red Circular Marker and a popup displaying name of the Launch Site.
- After that All the Successful Outcomes of the Site were marked with green marker and all Failed once with red Marker.
- And than the Distance Between Close Proximate sites Like City, Coastline, Highway and Railway Line with respect to Launch Site were calculated by plotting line between two sites.
- This objects were added so, that we could get answers on question like how close the Launch sites are to City, Coastline, Highway and Railway Line ?, how close they are form the equator?, do Launch Sites have proximate distance from cities? Etc.

GitHub URL:

IBM-Data-Science-Capstone-Project/Interactive Visual Analytics Using Folium Maps.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

(As by default GitHub takes Notebook as Untrusted the Output Maps will not be seen in the Notebook)

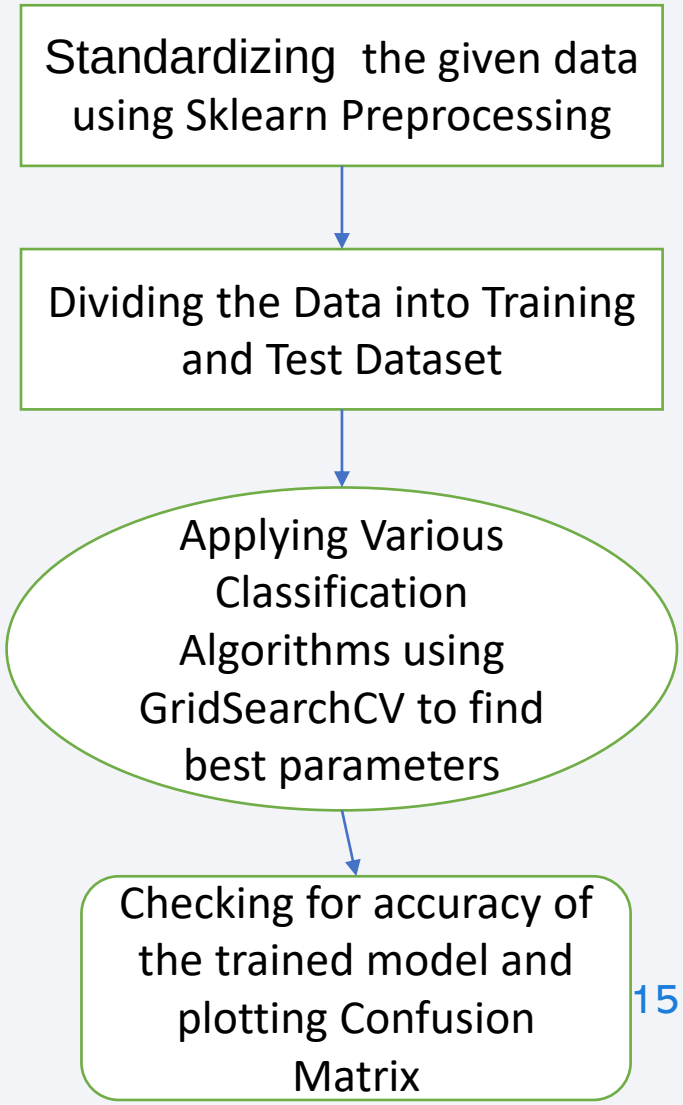
Build a Dashboard with Plotly Dash

- Pie Chart was added to get Success and Failure Percentage .The Launch Site is to be selected from the Dropdown Menu and the Pie Chart will display the Output of Success rate accordingly
- Launch Site from Dropdown and Payload Mass Range is taken as input for Scatter Plot which will show the average success rate using the boosters that are capable to lift payload which lies in given Payload range.
- Pie Chart is easy to understand when you require to compare the outcomes in this case Success Rate vs Failure Rate
- Scatter Plot is used to Provide visualization that is very easy to interpret.
- GitHub URL:
- IBM-Data-Science-Capstone-Project/spacex_dash_app.py at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Predictive Analysis (Classification)

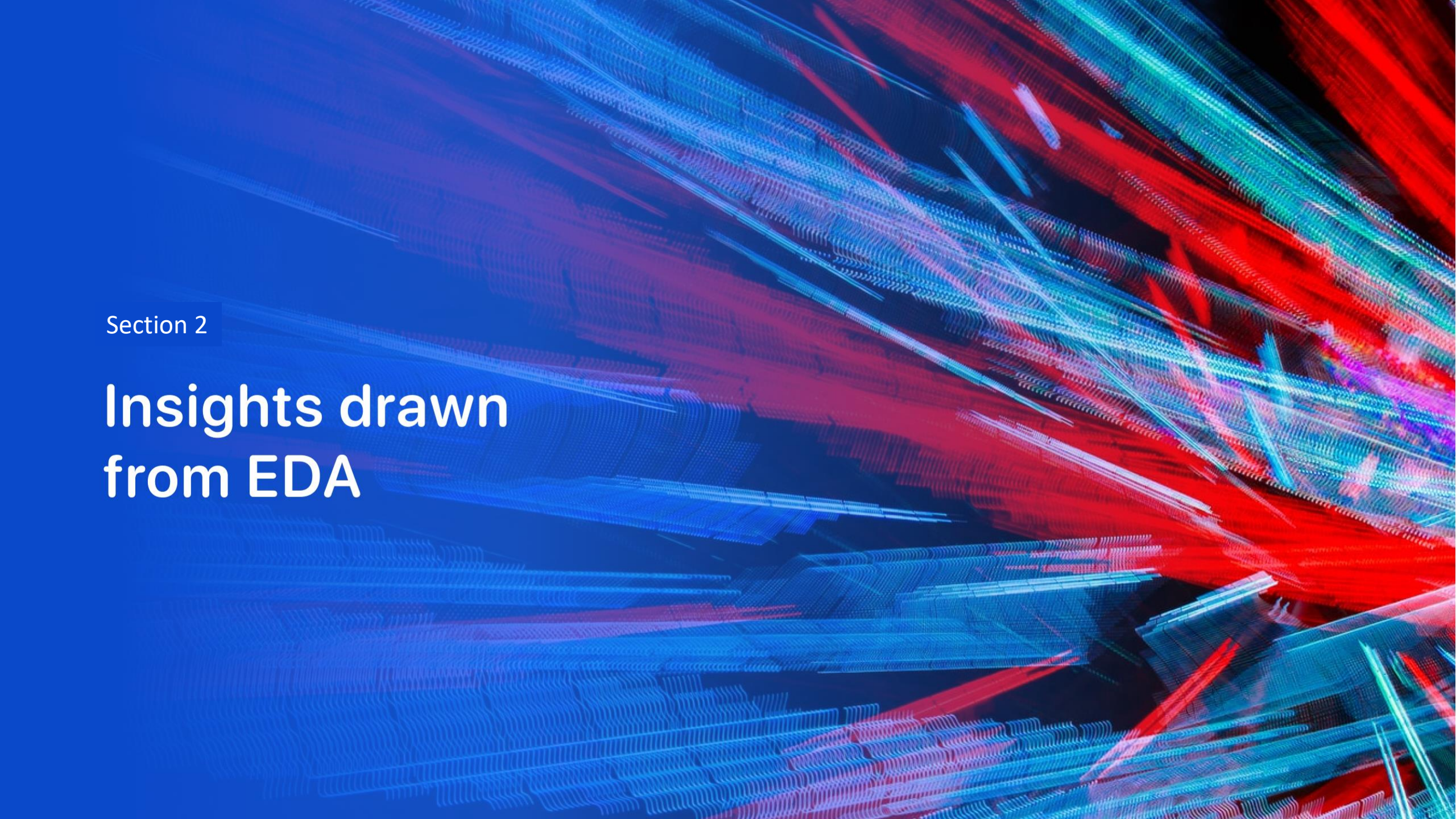
- Logistic Regression, SVM, Trees and KNN were used to find out the best Algorithm based on the accuracy of the output after Hyperparameter tuning Using GridSearchCV.
- The Highest accuracy was achieved from Logistic Regression, SVM and Tree which was 83.3333%
- GitHub URL:

IBM-Data-Science-Capstone-Project/Predictive Analysis Using ML.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

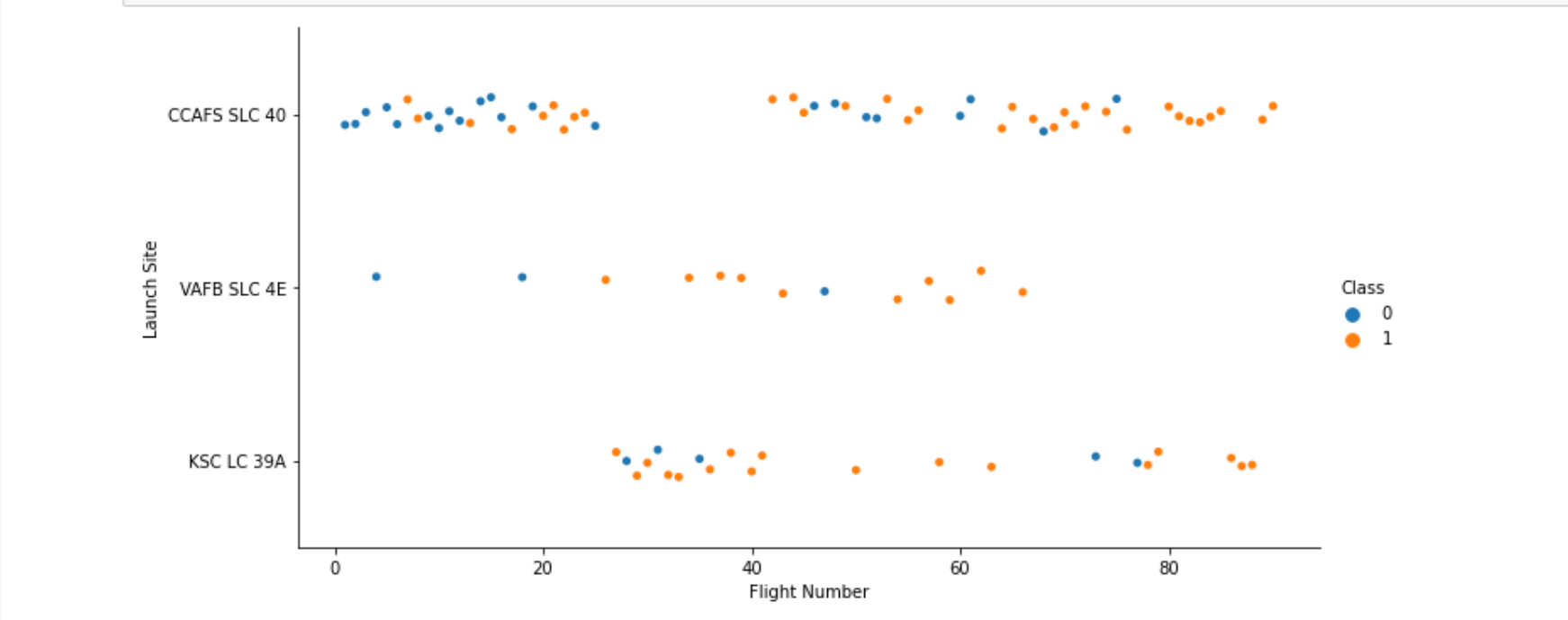


Results

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

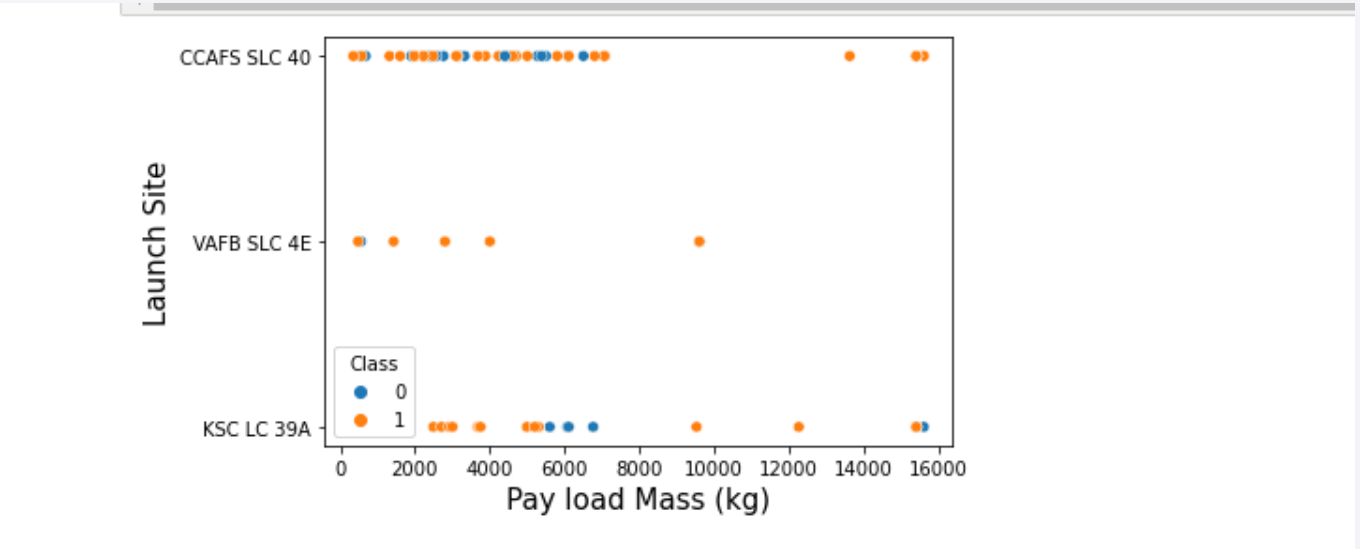

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



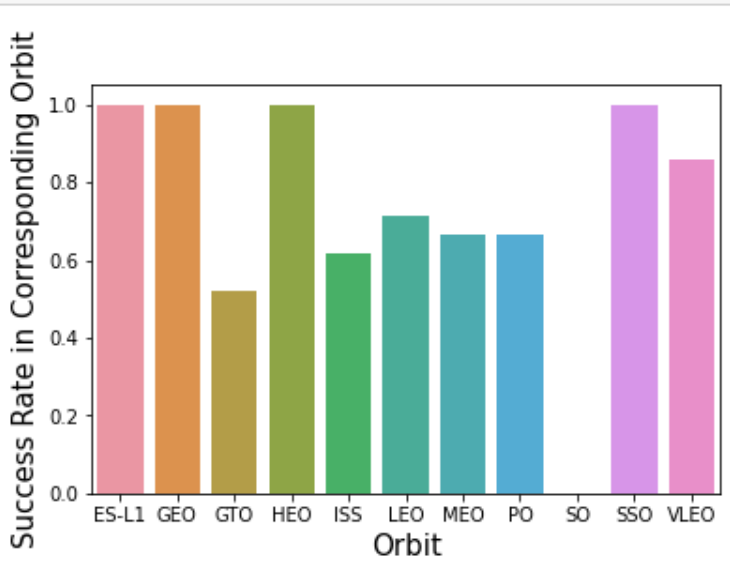
- With the Increase in flight number the No. of Successful outcomes also Increased
- Site CCAFS SLC 40 has highest number of Launches.
- Site KSC LC 39A has highest percentage of Successful Flight Outcomes.

Payload vs. Launch Site



- For Site CCAFS SLC 40 2 rockets carried the highest Payload Mass that is 16000 Kg and also landed Successfully
- For Site VAFB SLC 4E there are no rockets that have High Payload mass i.e greater than 10000 Kg
- Most of the Payload carried by the Rockets is between 0 Kg to 8000 Kg

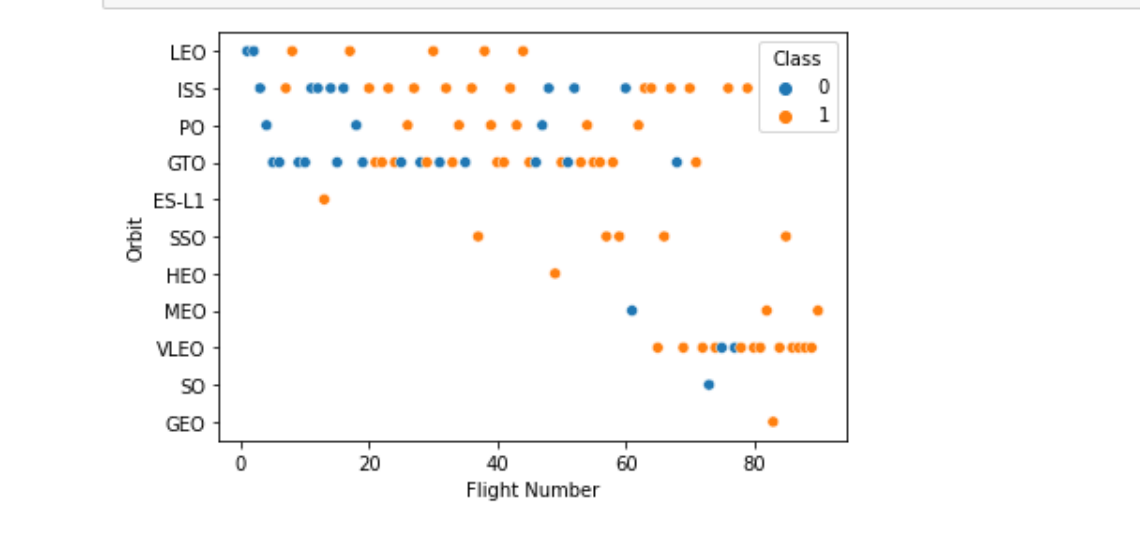
Success Rate vs. Orbit Type



Analyze the ploted bar chart try to find which orbits have high sucess rate.

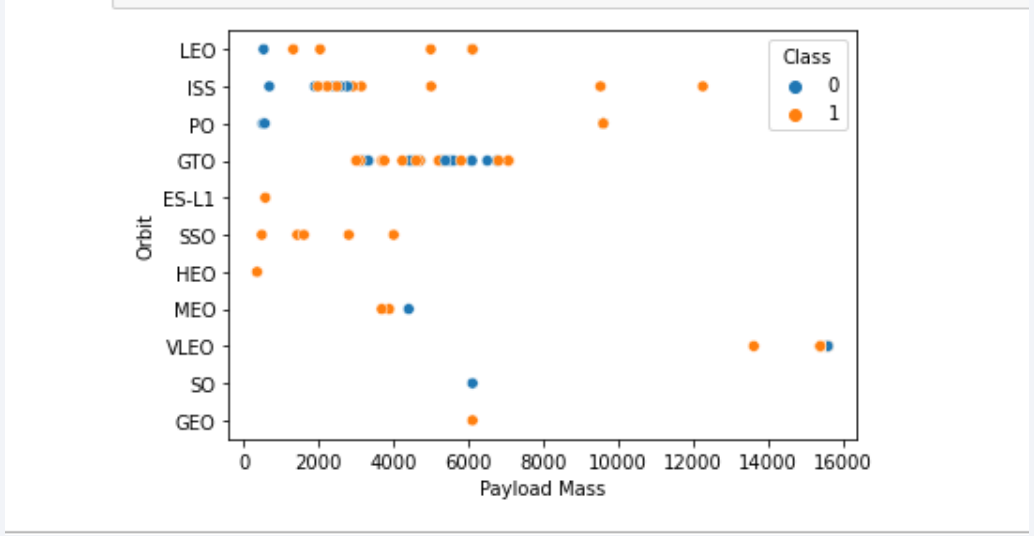
- The Rocket Launch to orbits ES-L1, GEO, HEO and SSO have the highest average success rate i.e. 1
- Least success rate was when the Rocket was Launched to GTO Orbit
- There was no successful Rocket Launch to SO orbit type

Flight Number vs. Orbit Type



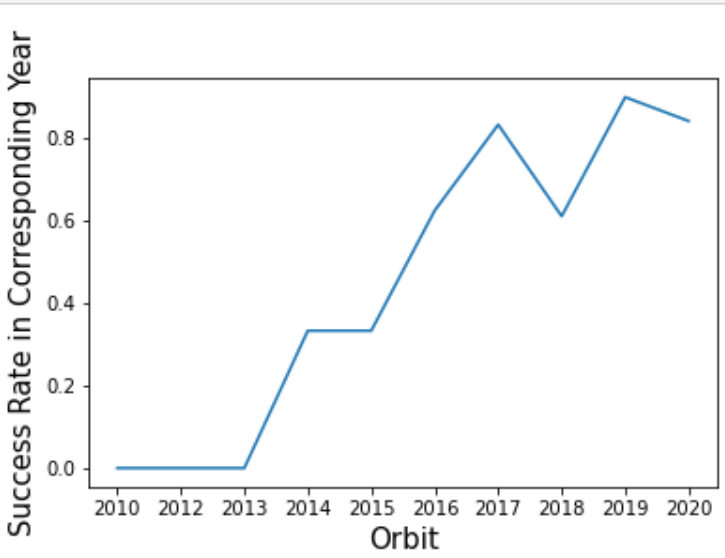
- In LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- There are 3 Orbit types in which there is only one launch and was successful
- Rocket Launch to orbits ES-L1, GEO, HEO and SSO showed maximum successful launches with increase in flight number
- With increase in flight number the number of successful launches increased significantly

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- For GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there.
- There are 5 Launches in which payload mass was greater then 10000 Kg and the outcome was successful.
- Highest Payload carried by rocket was 16000 Kg which was Launched to VLEO Orbit type and the landing outcome was a success

Launch Success Yearly Trend



- The Success rate increased significantly with Years passing by from 2013 to 2020.

All Launch Site Names

- There are 4 Unique site names in space mission

1) CCAFS LC-40
2) VAFB SLC-4E
3) KSC LC-39A
4) CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- There are 2 Launch Site name that begin with ‘CCA’

1) CCAFS LC-40
2) CCAFS SLC-40

Total Payload Mass

- Total payload mass carried by boosters launched by NASA (CRS) is 45,596 Kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.400 Kg

First Successful Ground Landing Date

- On 22-12-2015 the first successful landing outcome on ground pad was achieved

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are

1) F9 FT B1022
2) F9 FT B1026
3) F9 FT B1021 2
4) F9 FT B1031 2

Total Number of Successful and Failure Mission Outcomes

- total number of successful and failure mission outcomes are

1) Success:98
2) Failure (In Flight):1
3) Success(Payload Status Unclear):1

Boosters Carried Maximum Payload

- There are total 12 boosters which have carried the maximum payload mass

Out[18]:

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

Out[15]:

Flight_Date	LandingOutcome	Booster_Version	Launch_Site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- There are total 2 failed landing outcomes in drone ship, their booster versions, and launch site names in year 2015.
- Both the Rockets have CCAFS LC-40 as Launch site

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

- 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 is,

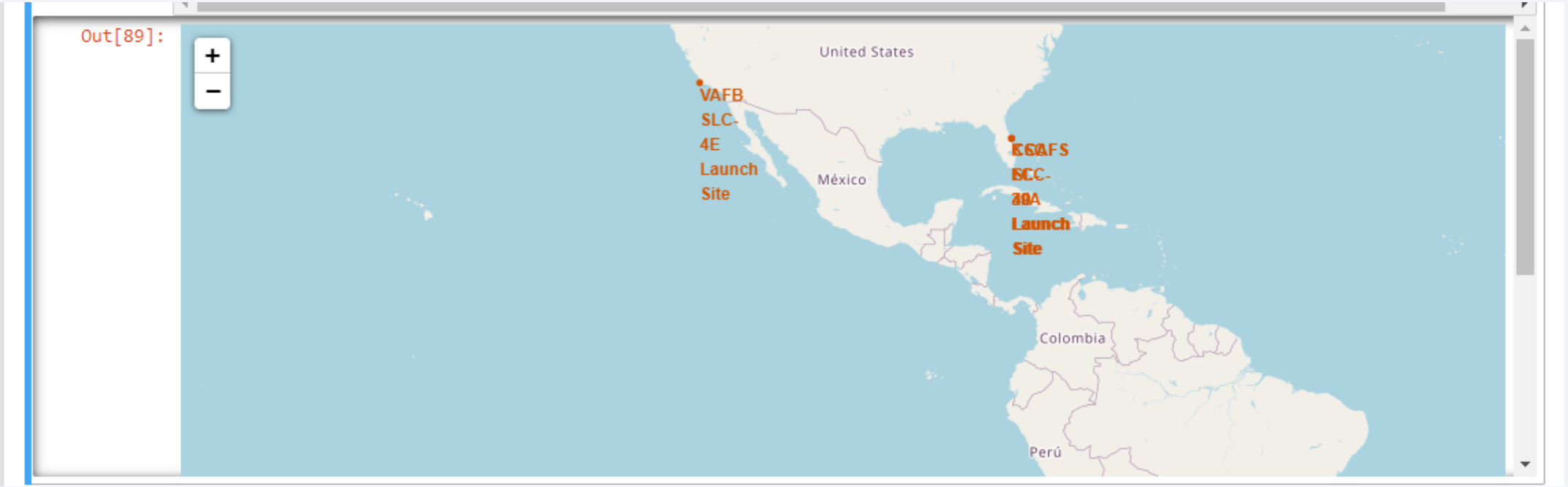
Out[13]:

LandingOutcome	Total
Failure (drone ship)	5
Success (ground pad)	3

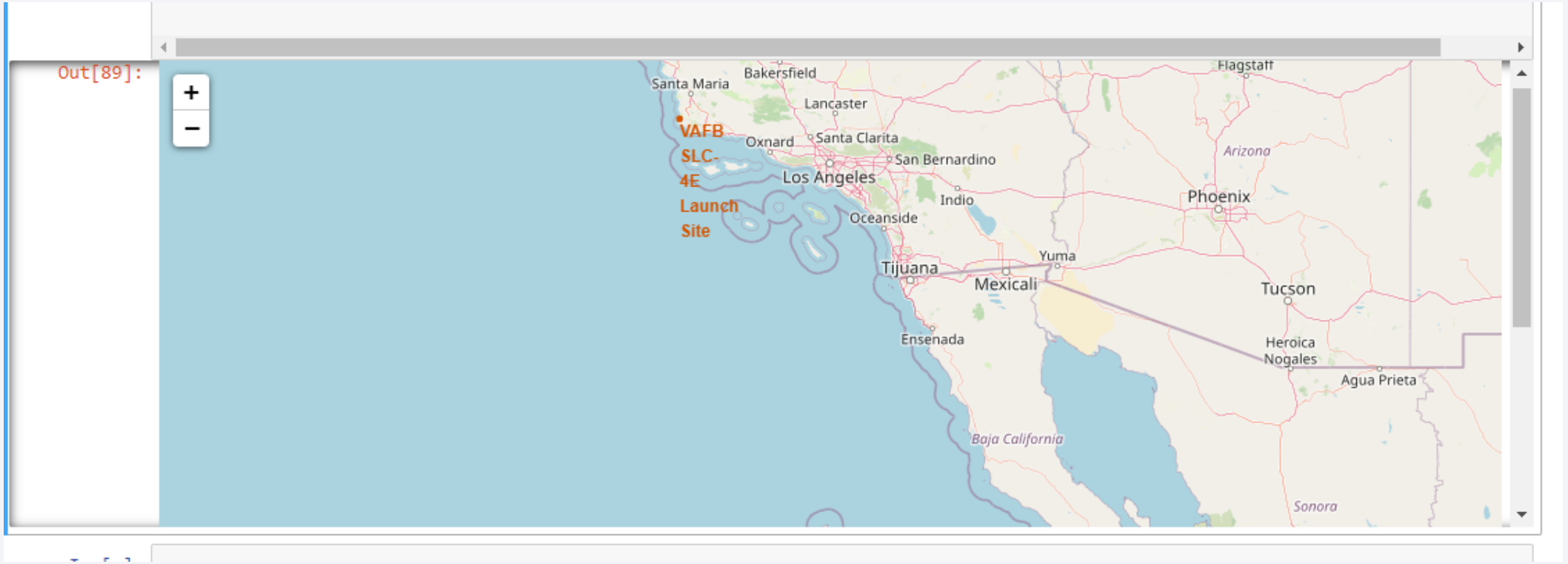
Section 3

Launch Sites Proximities Analysis

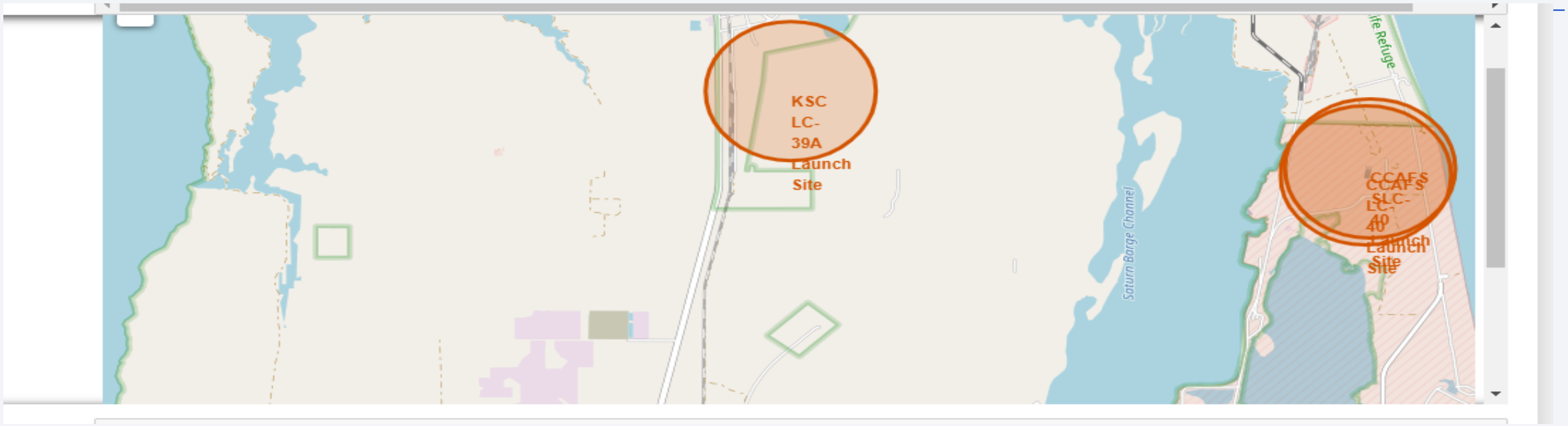
Map with Marked Launch Sites



Map with Marked Launch Sites

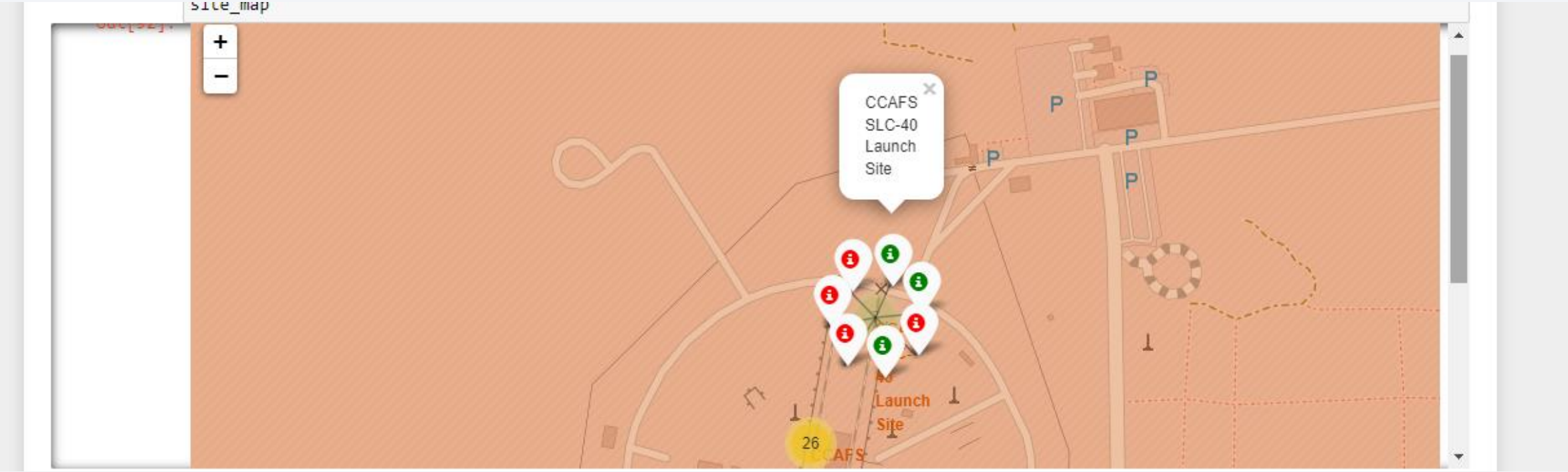


Map with Marked Launch Sites

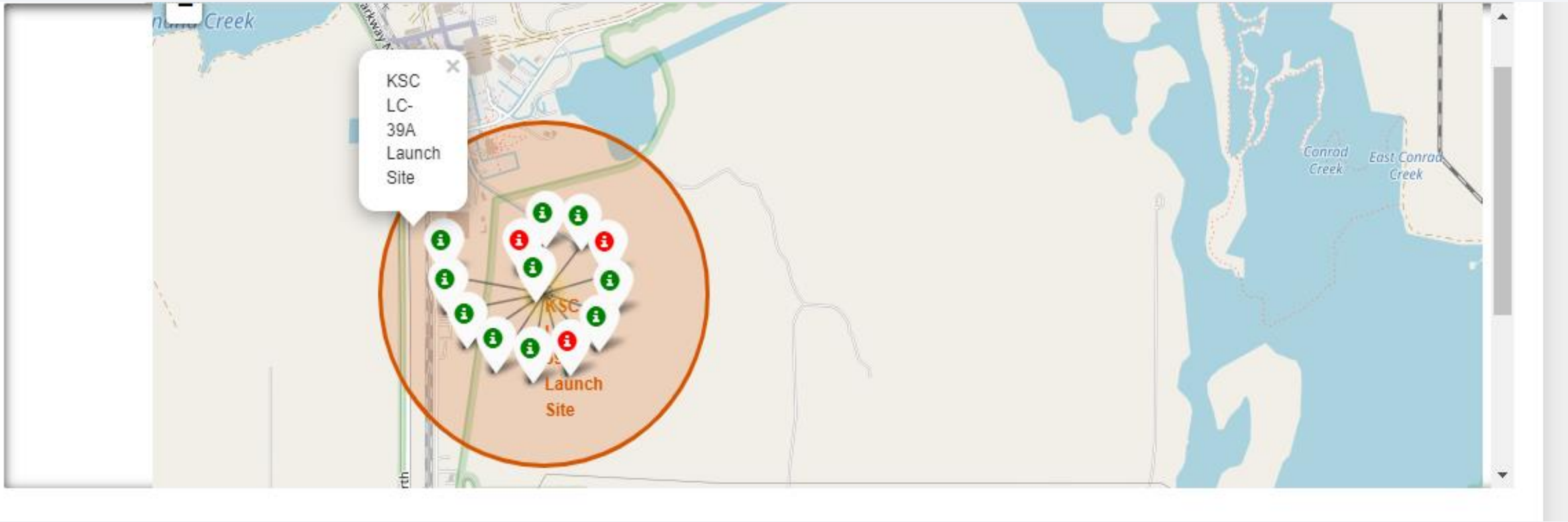


- From the map we can see know that Launch Sites are far away from equator
- And all Launch Sites are very close to the cost.

Map with Labeled Launch Outcome

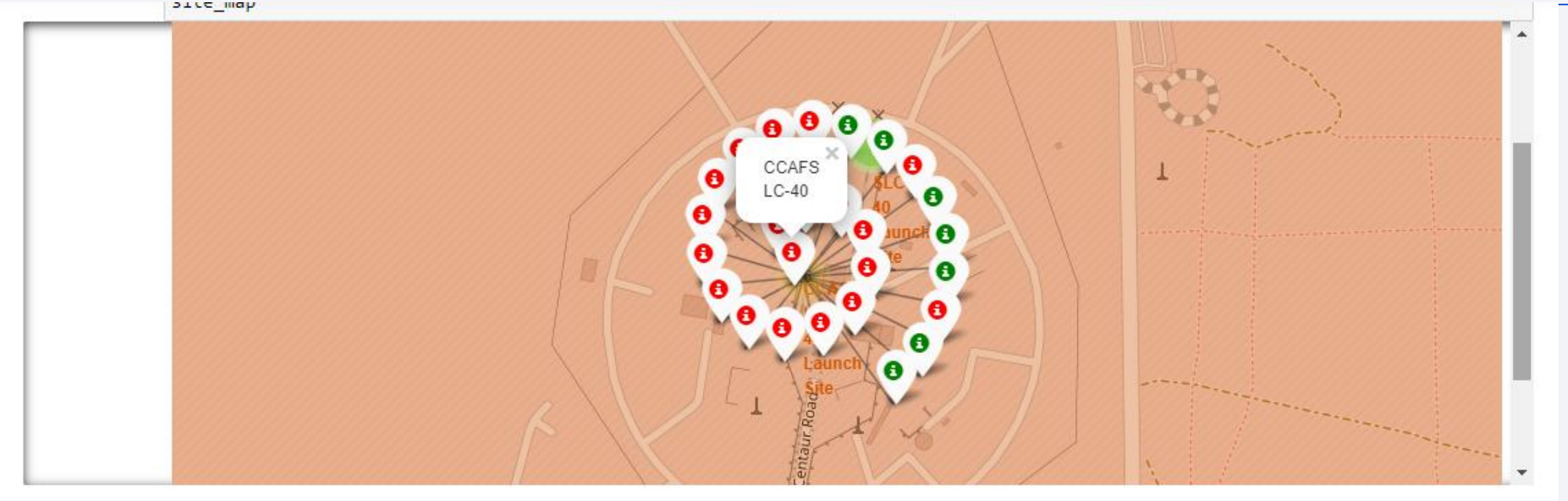


Map with Labeled Launch Outcome



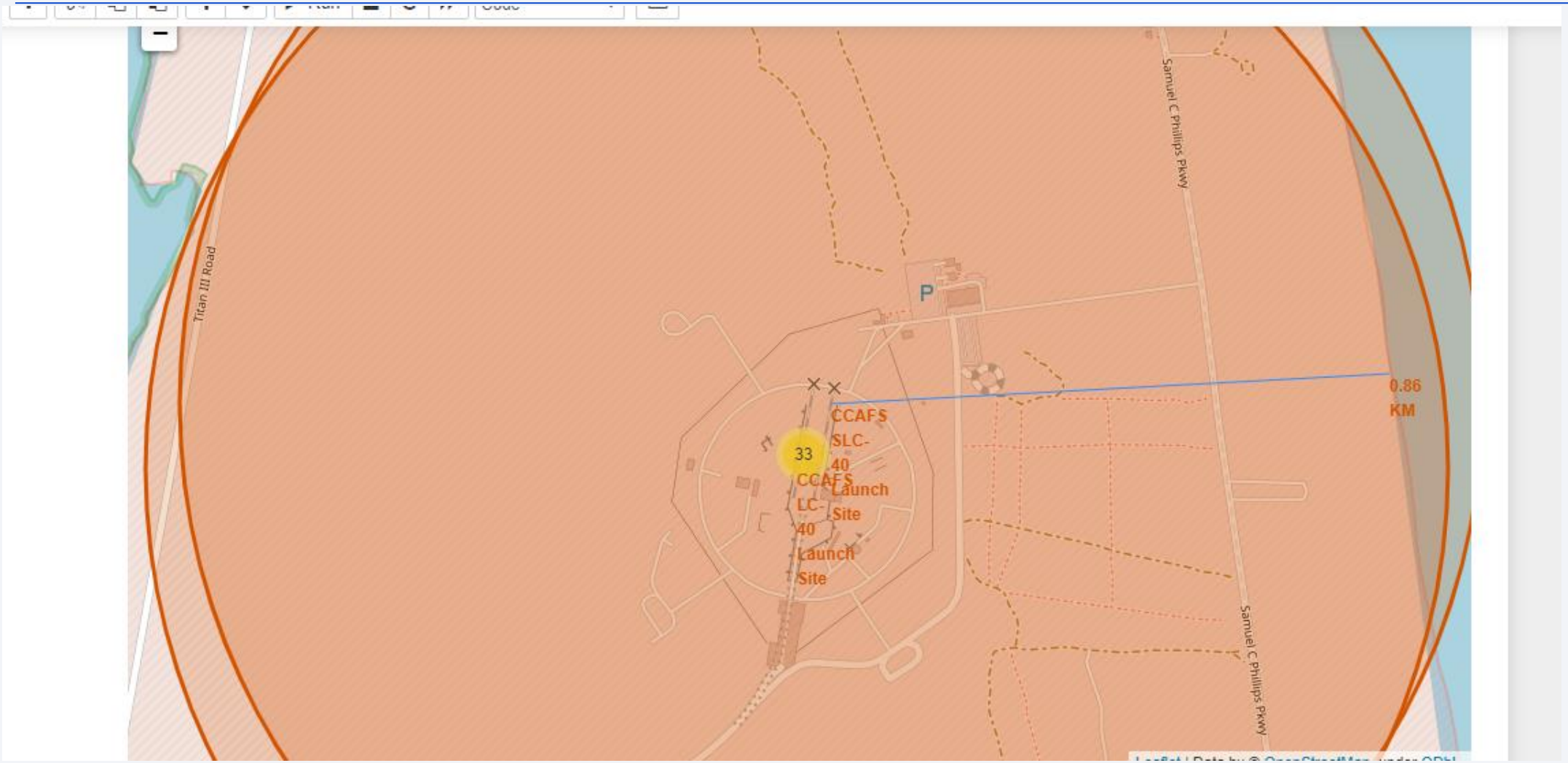
Map with Labeled Launch Outcome

Map with Labeled Launch Outcome

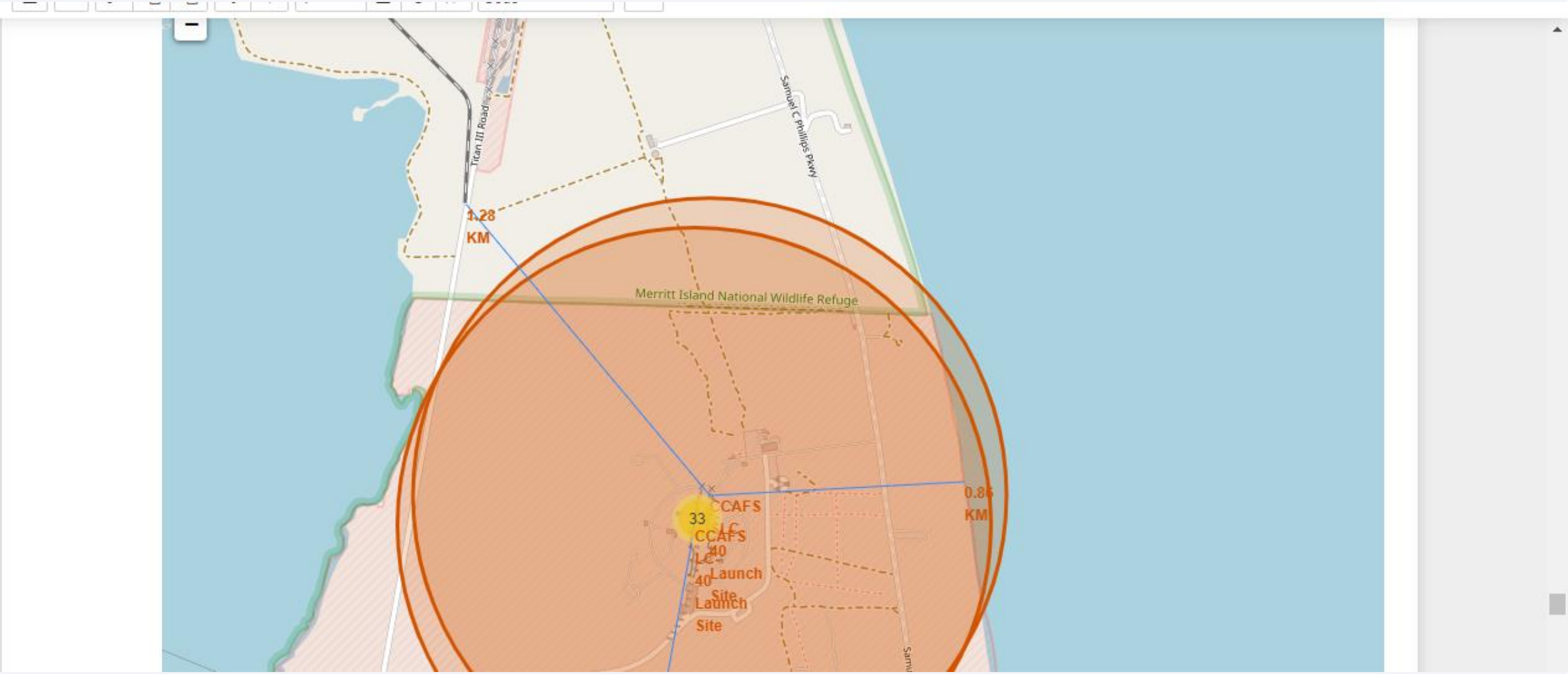


- Here Red Labels are for Failed outcomes and Green Labels are for Successful outcomes
- We can easily come to know the success rate of the launches for each site
- We can see here that the CCAFS LC-40 Lunch site have more Failed outcomes compared to successful outcomes

Map for Distance of Launch Site From Sea Coast

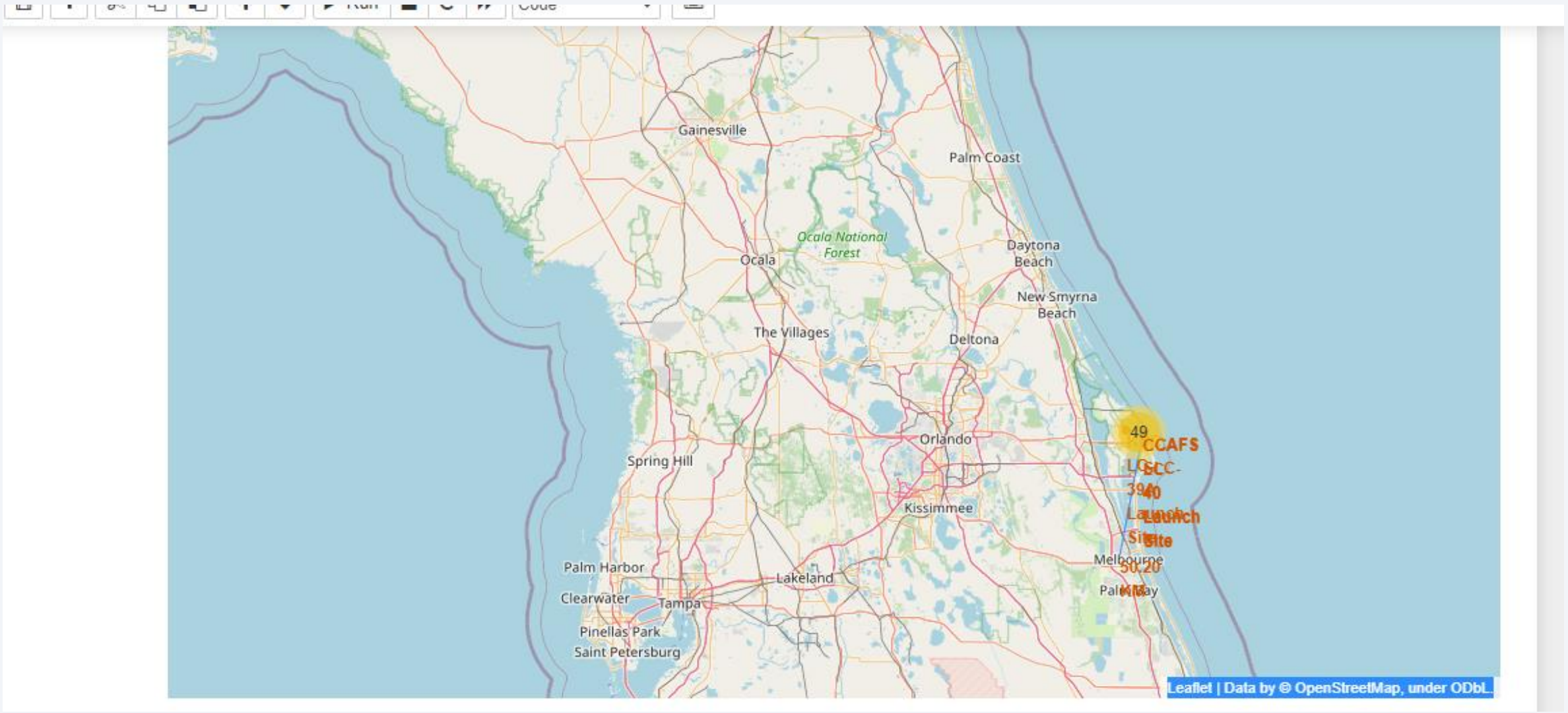


Map for Distance of Launch Site From Rail Road



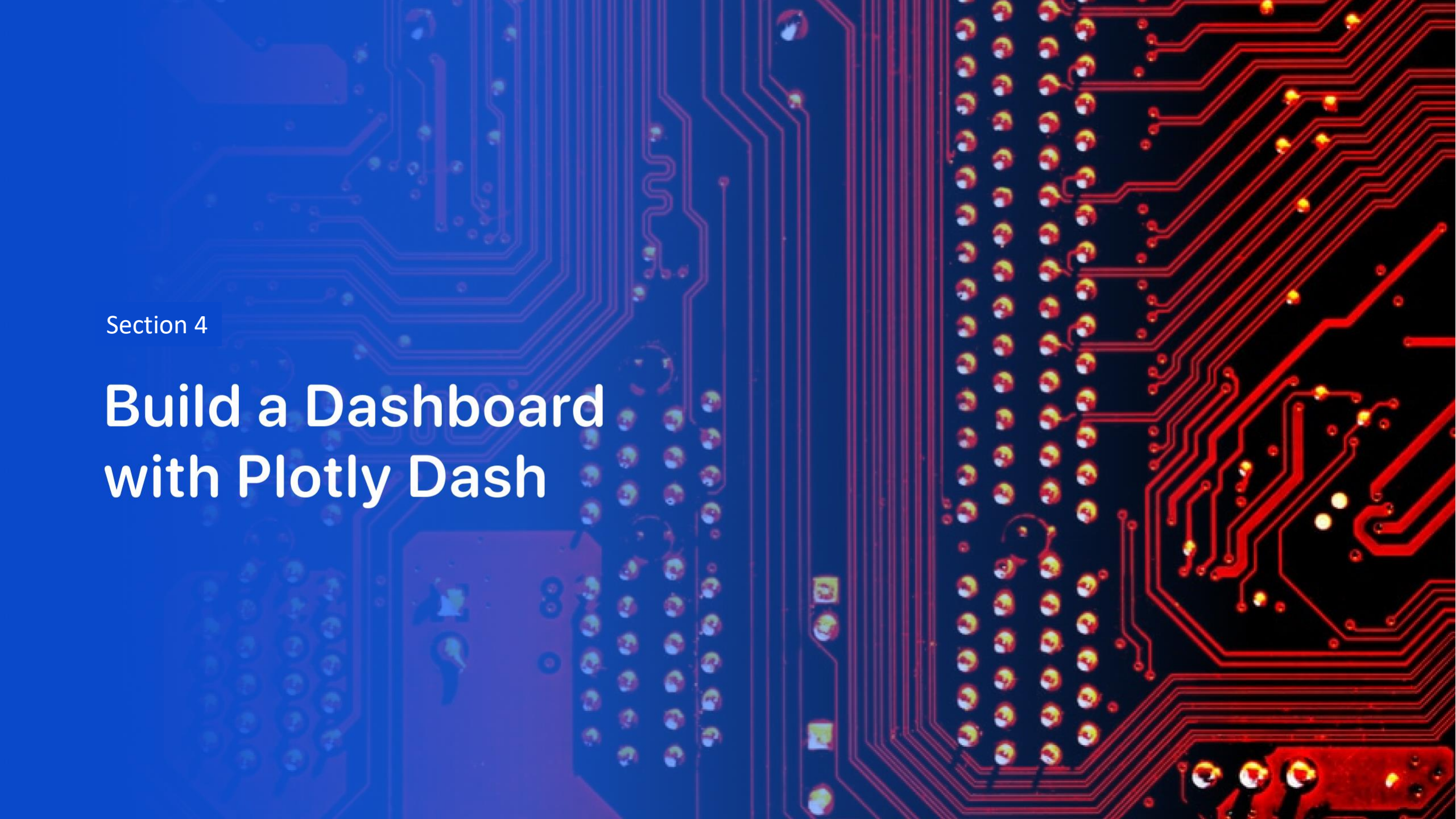
Map for Distance of Launch Site From Highway

Map for Distance of Launch Site From City

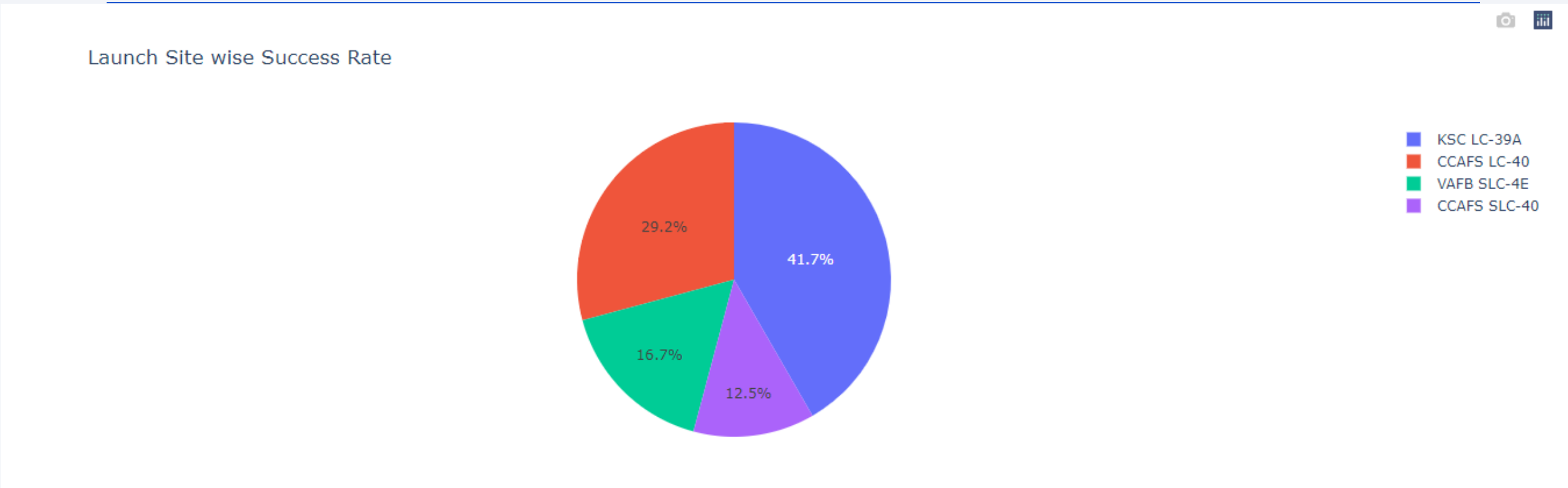


Map for Distance of Launch Site

- The Distance of Launch Site CCASF SLC-40 from NASA Railroad is 1.28 Km
- The Distance of Launch Site CCASF SLC-40 from Coastline is 0,.86 Km
- The Distance of Launch Site CCASF SLC-40 from Samuel C Phillips Highway is 0.63 Km
- The Distance of Launch Site CCASF SLC-40 from Melbourne is 50.20 Km
- Hence, we can say that the Launch Sites keep Significant distance from Cities

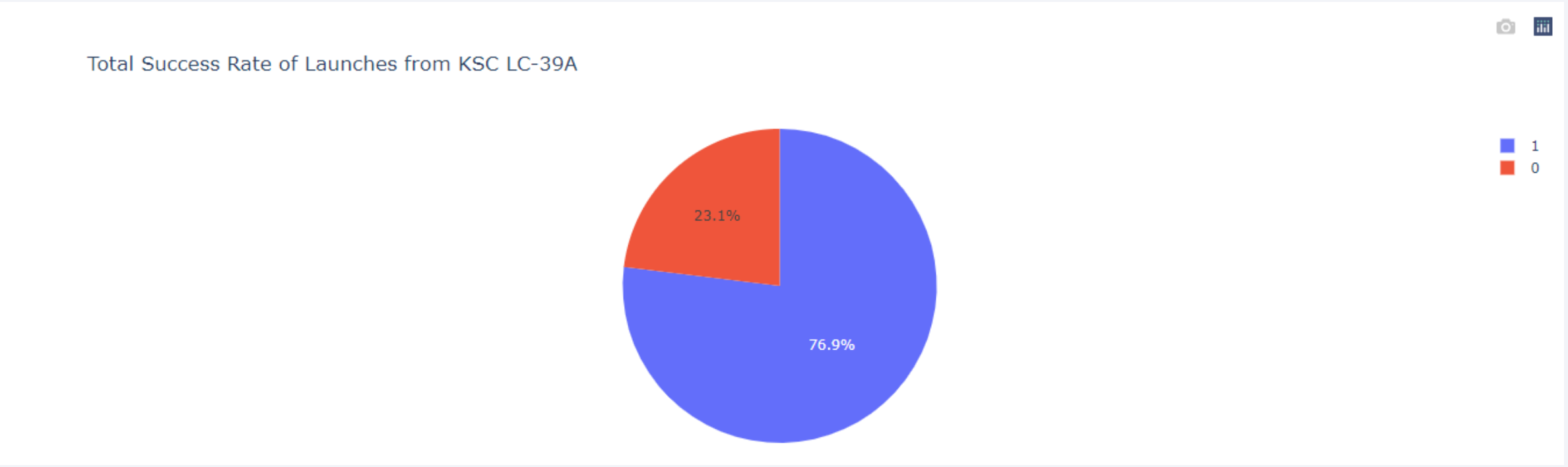

Section 4
Build a Dashboard with Plotly Dash

Pie Chart For Launch Outcome Success Count For All Launch Sites



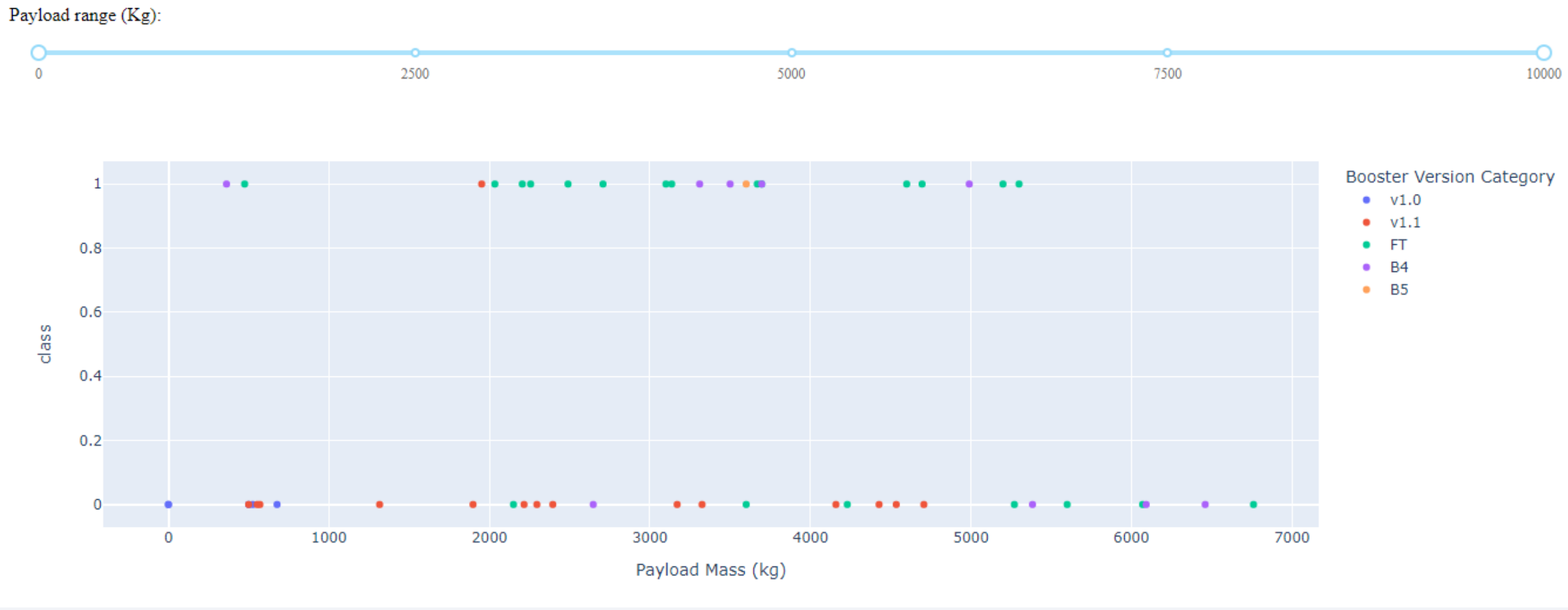
- KSC LC-39A has the highest Launch Site Success Rate i.e. 41.7%
- CCAFS SLC-40 has the lowest Launch Site Success Rate i.e. 12.5%

Pie Chart For Site With The Highest Launch Success Ratio

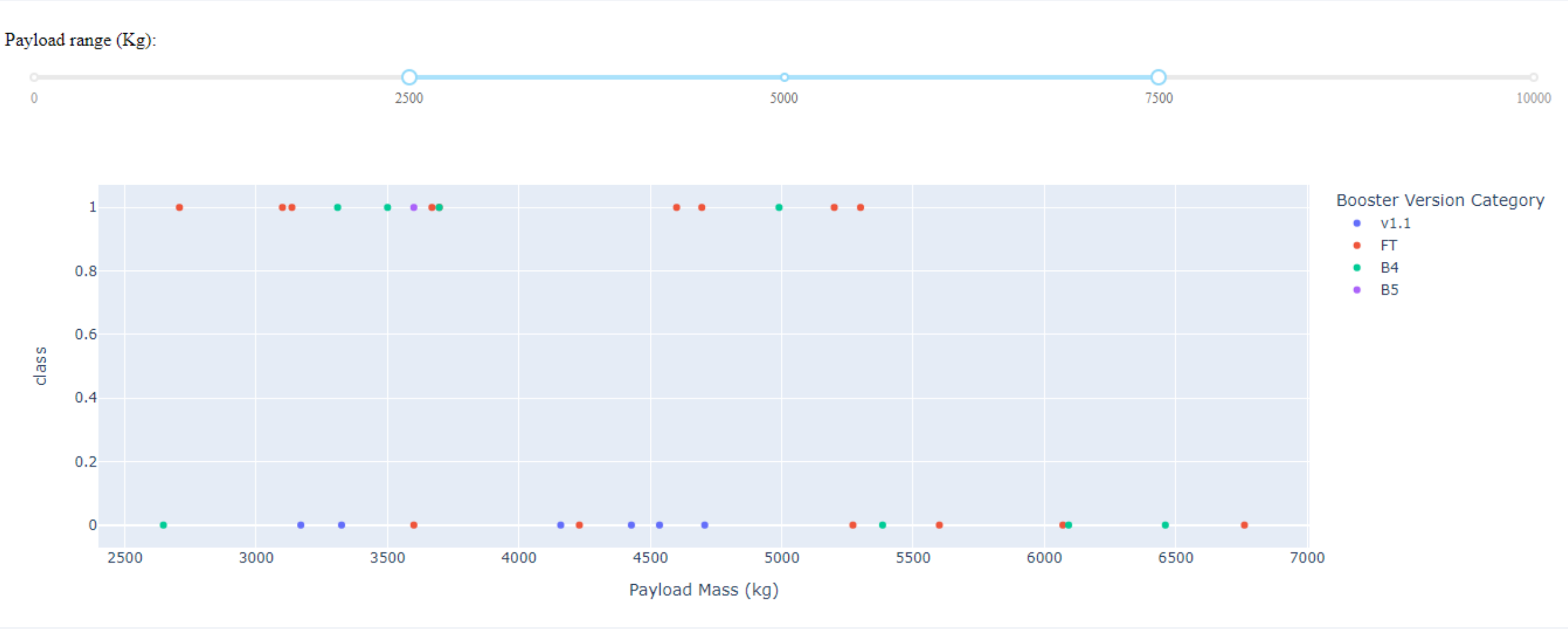


- KSC LC-39A have highest Launch Success Rate i.e. 76.9% compared to other Launch Sites

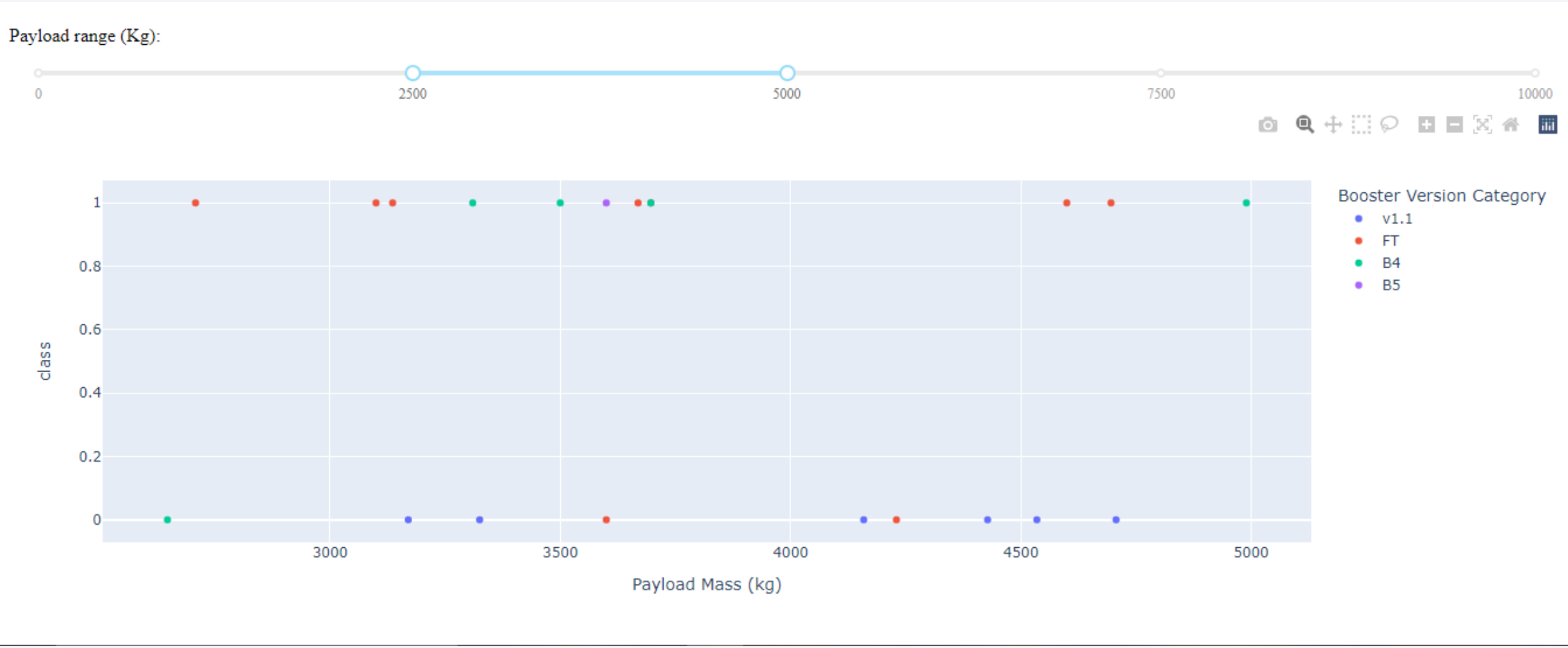
Scatter Plot Launch Outcome VS Variable Payload Range



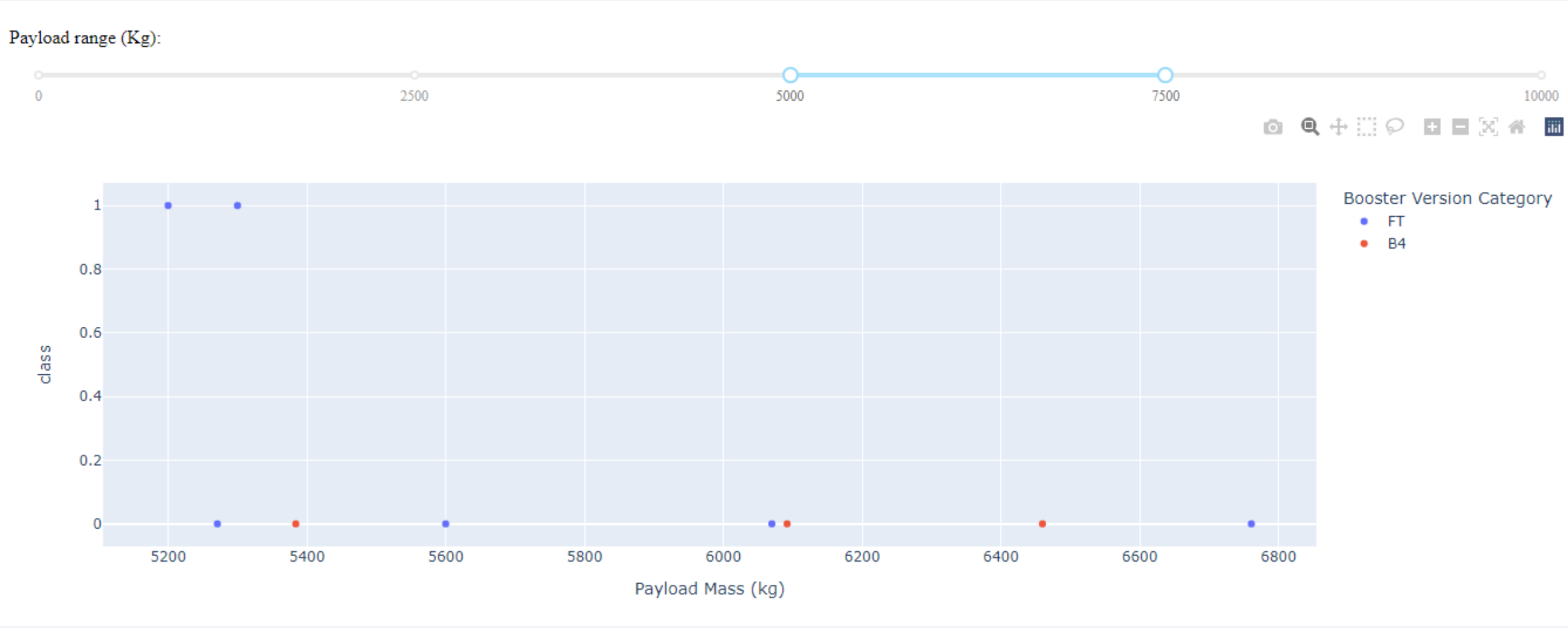
Scatter Plot Launch Outcome VS Variable Payload Range



Scatter Plot Launch Outcome VS Variable Payload Range



Scatter Plot Launch Outcome VS Variable Payload Range



Scatter Plot Launch Outcome VS Variable Payload Range

Scatter Plot Launch Outcome VS Variable Payload Range

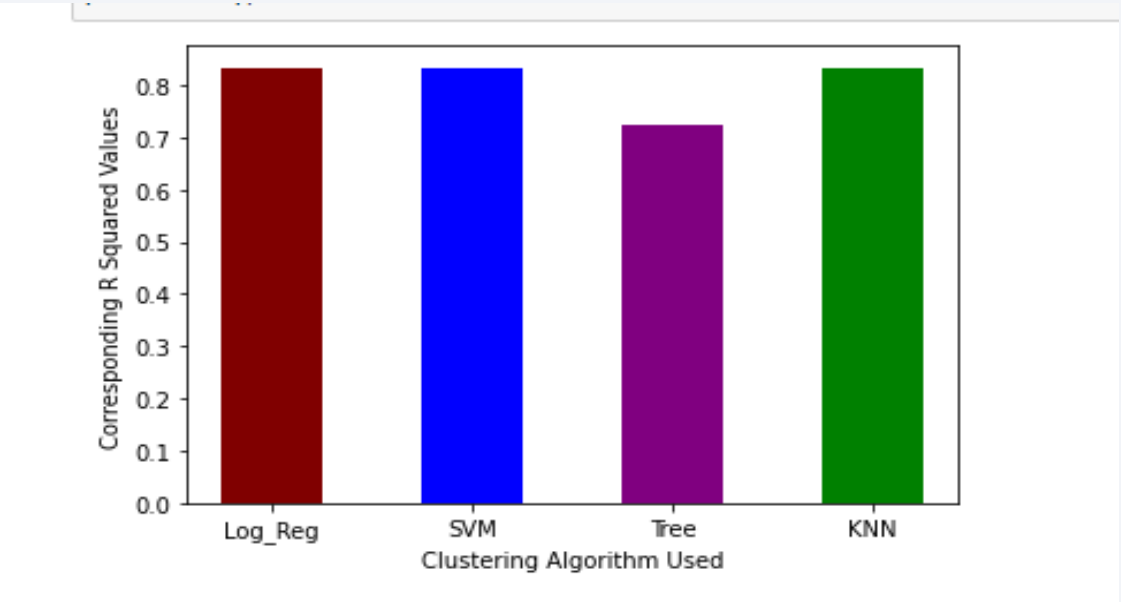
- Booster version FT has highest launch success rate compared to other booster version
- There is only one version of Booster i.e. B4 which has a successful outcome carrying highest payload 9600 Kg
- With increase in payload the launch success rate decrease significantly
- The Range with highest launch success rate is 2500 Kg to 5000 Kg
- The Range with lowest launch success rate is 0 Kg to 7500 Kg

Section 5

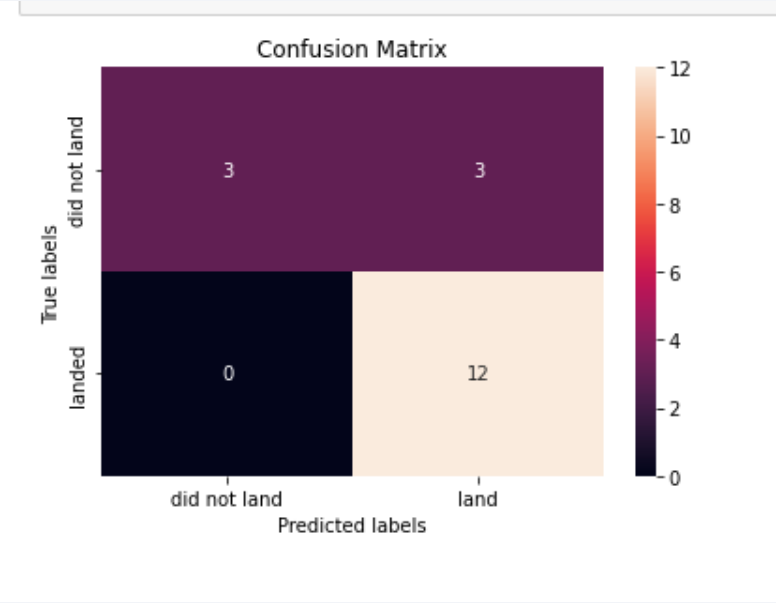
Predictive Analysis (Classification)

Classification Accuracy

- Logistic Regression, SVM, KNN have highest accuracy i.e. 83.3333%



Confusion Matrix



- From the Confusion Matrix we come to know that the model falsely predicted outcome **successful landing** for 3 flights from total 6 flights that **did not land** successfully.
- And it predicted 12 out of 12 correct outcome for successful landing outcome

Conclusions

- Having KSC LC 39A as launch site
- Launch destination to one of the ES-L1, GEO, HEO or SSO Orbits
- Using Booster version FT for the Rocket
- Payload carried by the Rocket must be in range 2500 Kg to 7500 Kg
- If take all this point into consideration than it can increase the chances of Successful landing to very great extent

Appendix

Datasets and extra Notebooks made in order to complete project

Dataset 1: IBM-Data-Science-Capstone-Project/dataset_part_2.csv at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Dataset 2: IBM-Data-Science-Capstone-Project/dataset_part_3.csv at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Data Wrangling Notebook: IBM-Data-Science-Capstone-Project/Space Data Wrangling.ipynb at main · shahyaksh/IBM-Data-Science-Capstone-Project (github.com)

Thank you!