IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thant Thiri Maung
2024. 02. 16

Outline

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

Executive Summary

- **Summary of methodologies**

Data Collection by Web Scraping

Data Wrangling

Data Analysis and visualization with sql and visualization tools

Prediction models

Introduction

- **The objective is to evaluate the viability of the new company Space Y to compete with Space X**

- **Problems**
- The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets
- Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
- 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 model

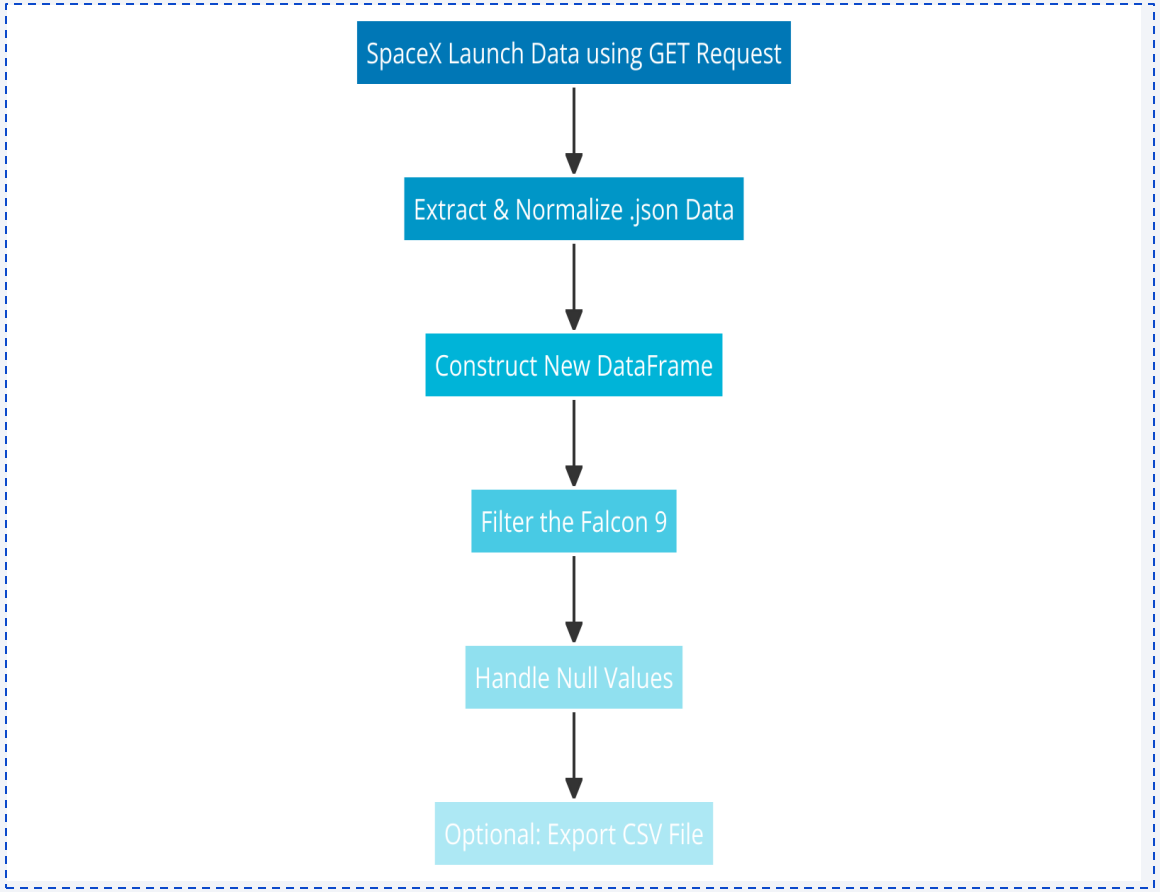
Data Collection

- Data is collected from two sources.
  1. https://api.spacexdata.com/v4/rockets/
  2. https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- SpaceX launch data using the GET request
- Falcon 9 and Falcon Heavy Launches Records from Wikipedia with web scraping

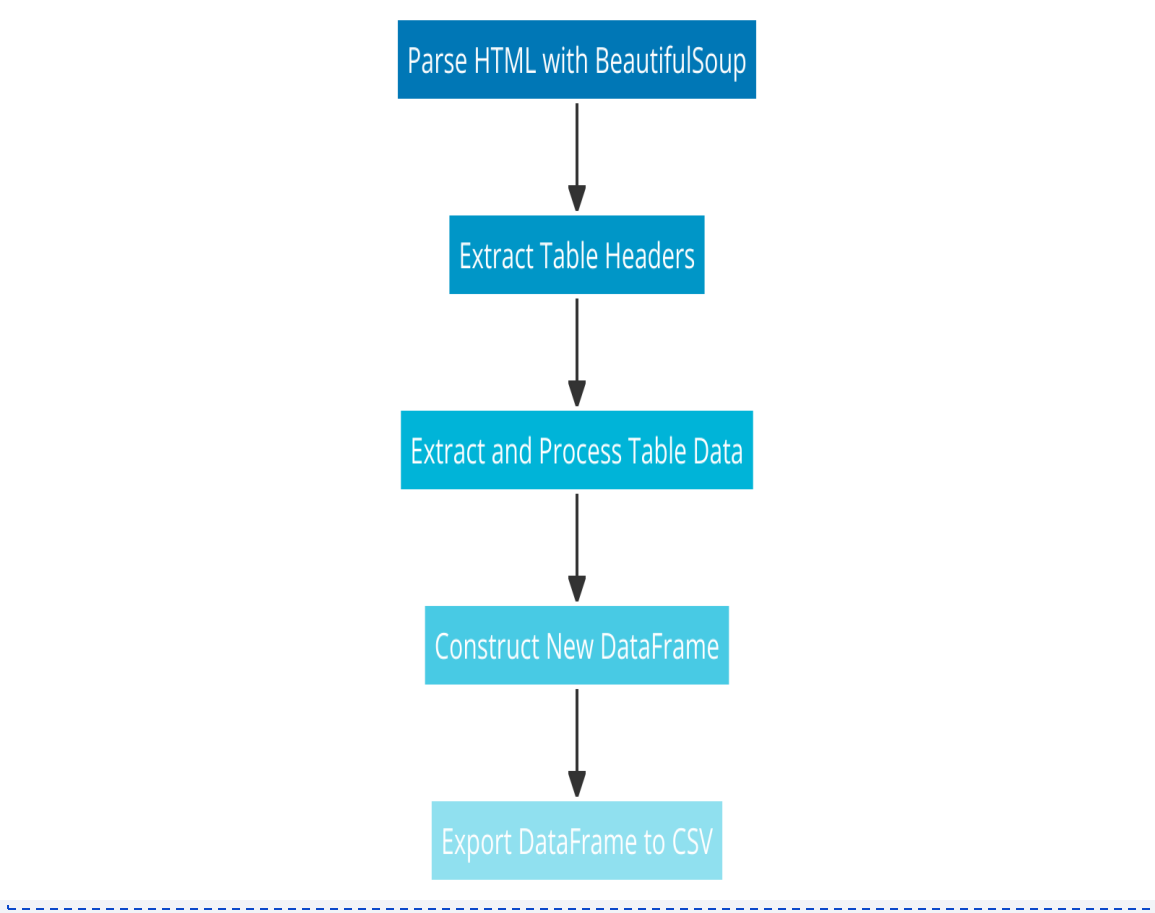
Data Collection – SpaceX API

- SpaceX launch data using the GET request.
- Extract Normalize the .json Data
- Construct New DataFrame by retrieving details data
- Filter the Falcon 9
- Handle with Null Values
- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science



Data Collection - Scraping

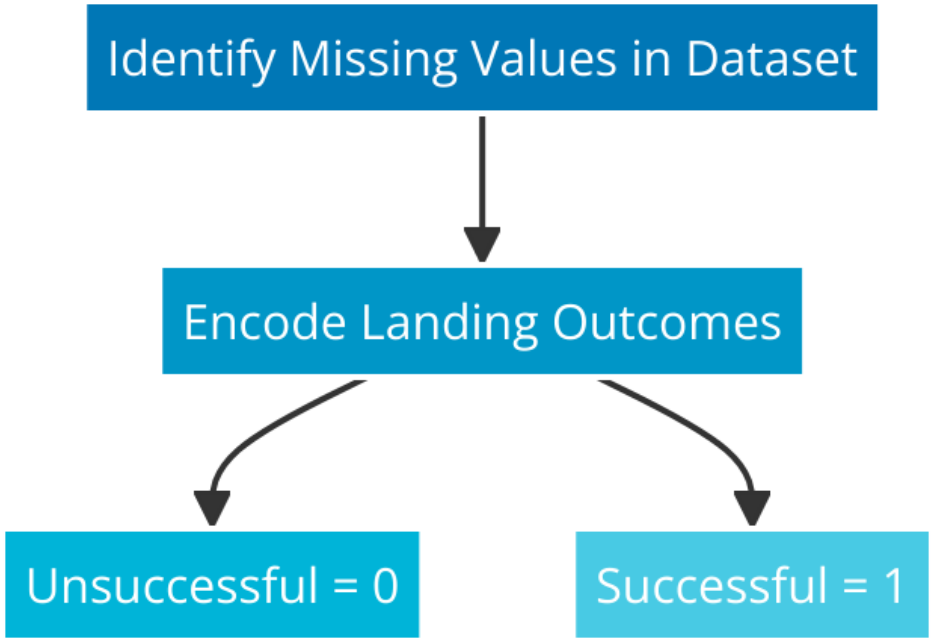
- Request Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Parse HTML with BeautifulSoup and process Web Scraping to get the dataset
- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science



Data Wrangling

- First identify if there any missing values in each attribute
- In the "Mission Outcome" we have different category (True Ocean, False Ocean, True RTLS, False TRLS, True ASDS, False ASDS, None ASDS, None None)
- Encoded these outcomes as unsuccessful = 0 and successful = 1
- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

Data Wrangling



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Scatter plot between fligh number and launch site
- Scatter plot between Payload and launch site
- Bar chart for success rate of each orbit
- Scatter plot between FlightNumber and Orbit type
- Scatter plot between payload and Orbit type
- Line chart of launch success yearly trend
- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

EDA with SQL

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

- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

Build an Interactive Map with Folium

- **Markers, circles, lines and marker clusters were used in Folium Maps**
- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates
- Marker clusters indicates groups of events in each coordinate
- Lines are used to indicate distances between two coordinates.
- Source Code - https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

Build a Dashboard with Plotly Dash

- The Dashboard for
    - Percentage of launches by site
    - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.
- Source code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

Predictive Analysis (Classification)

- Predict whether first stage of Falcon 9 will land successfully
- Test with 4 models

1. Logistic Regression
2. Support Vector Machine
3. Decision Tree Classifier
4. K-nearest Neighbors

- Source Code – https://github.com/thantthirimaung/Winning-Space-Race-with-Data-Science

Results

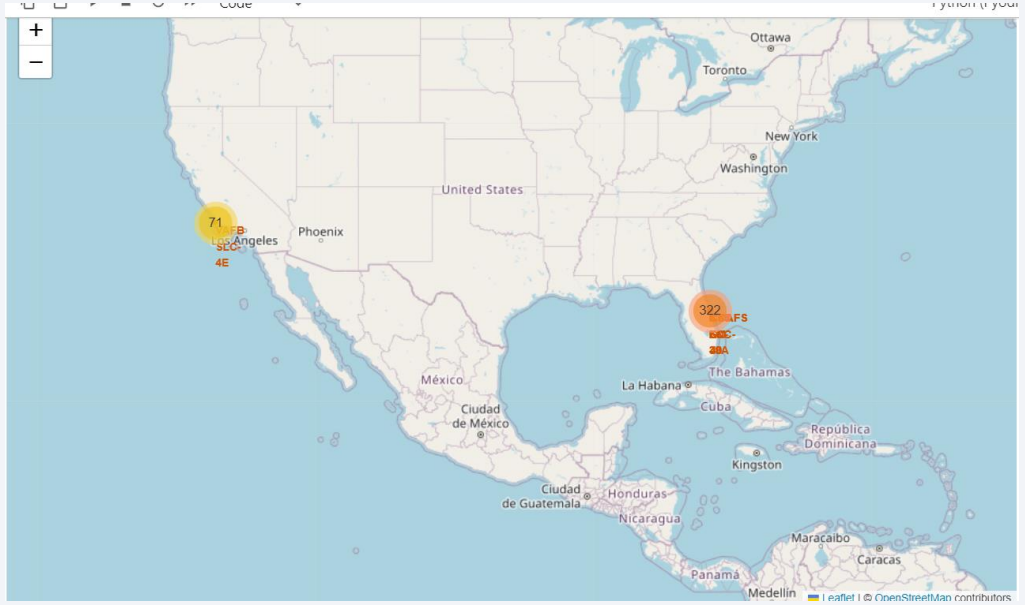
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

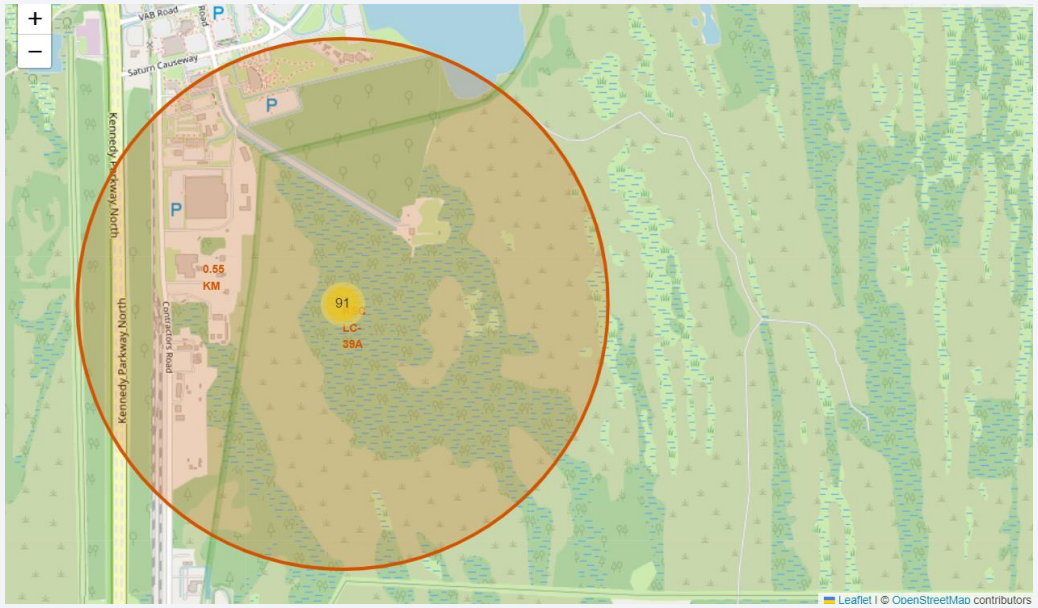
Results

- Exploratory data analysis results
- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA
- The average payload of F9 v1.1 booster is 2,928 kg
- The first success landing outcome happened in 2015 fiver year after the first launch
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
- Almost 100% of mission outcomes were successful
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
- The number of landing outcomes became as better as years passed.

Results

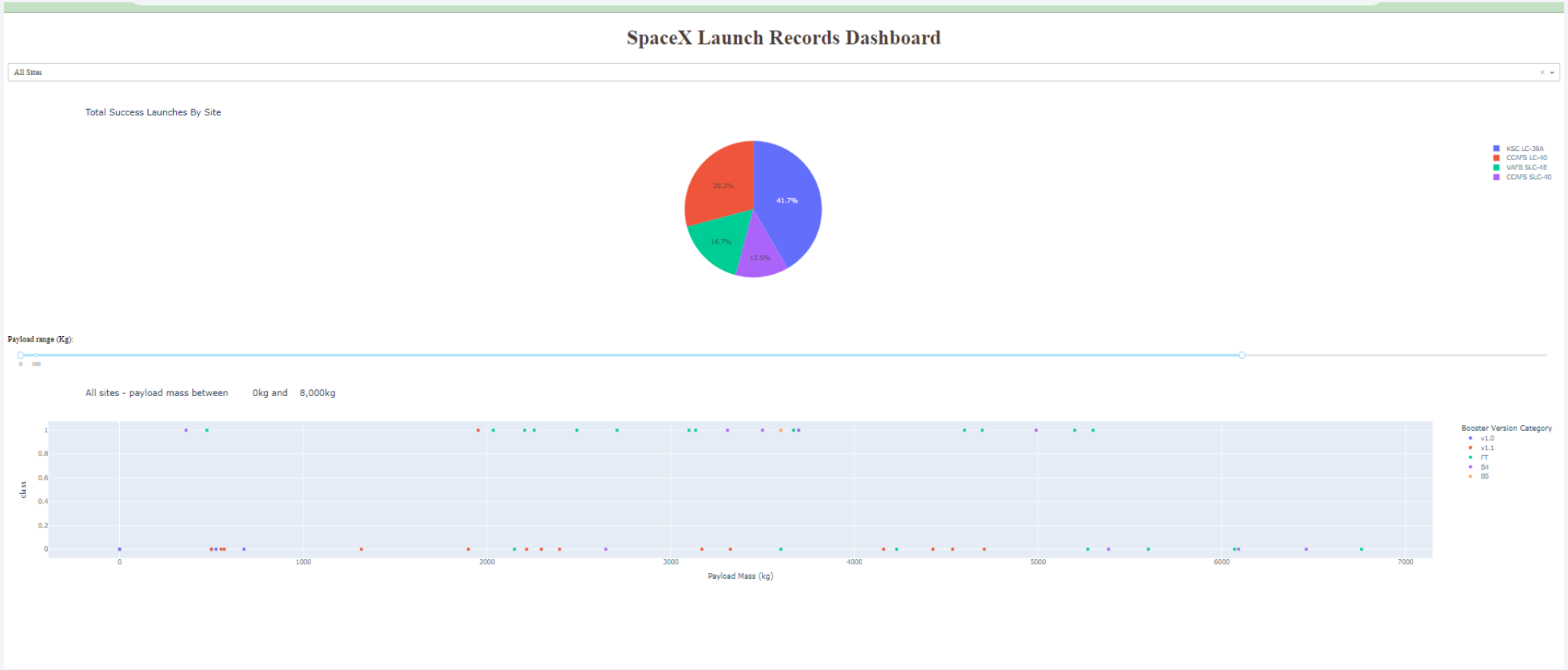
Interactive demo has shown that launch are built near sea.





Results

Space X Launch Records Dashboard



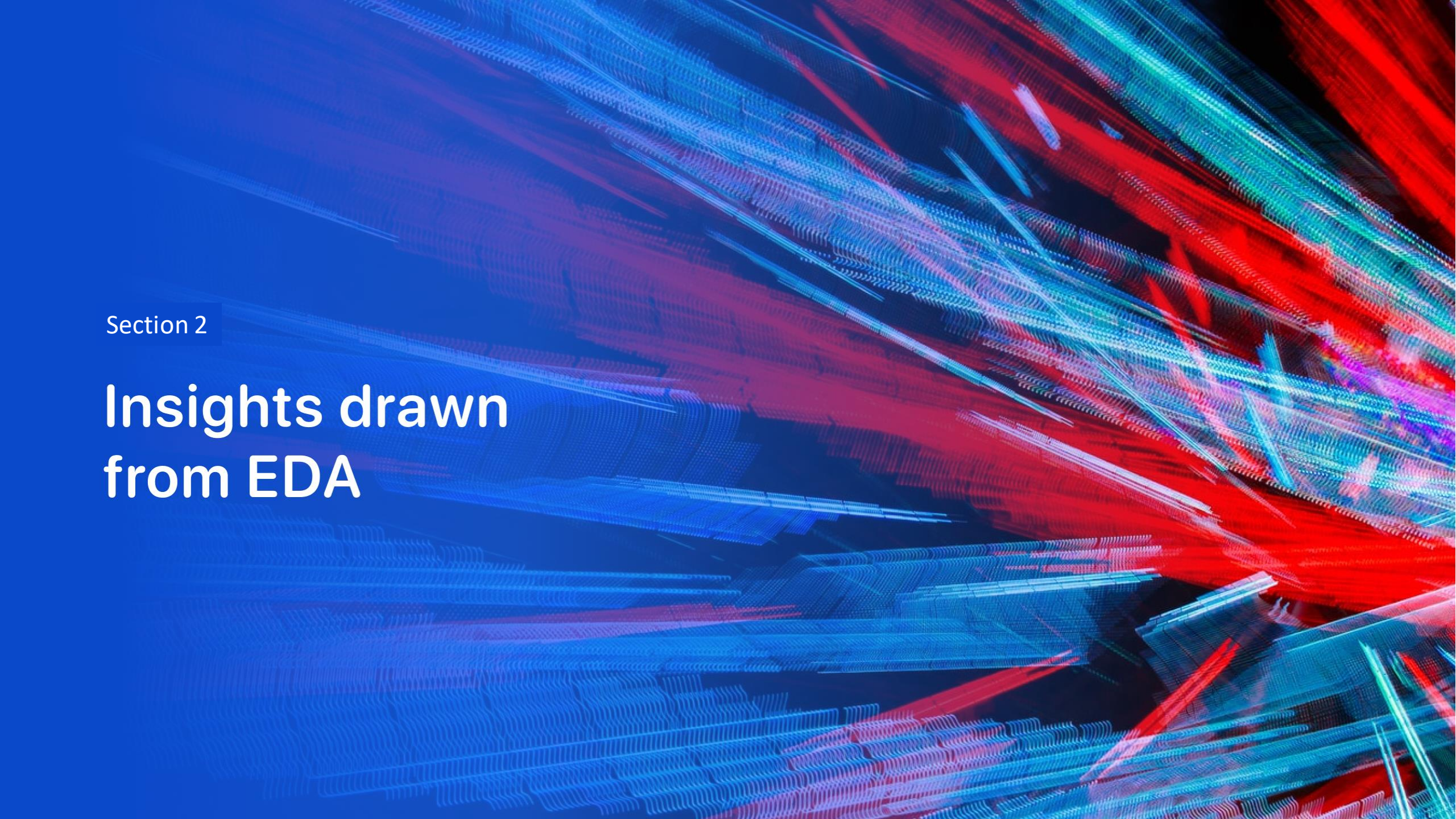
Results

The prediction with 4 different models show that Decision Tree Classifier has the higher accurate value than other models.

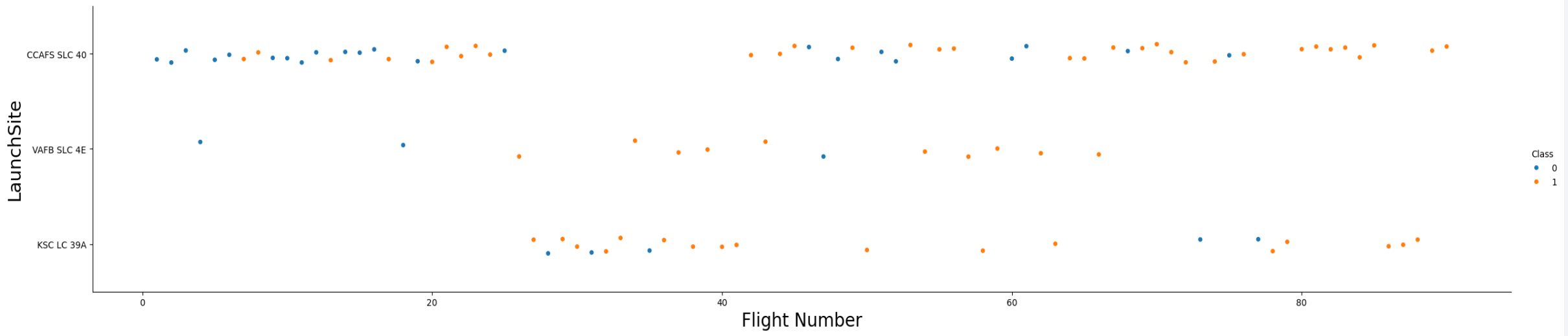
Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.89107	0.83333
KNN	0.84821	0.83333

Section 2

Insights drawn from EDA

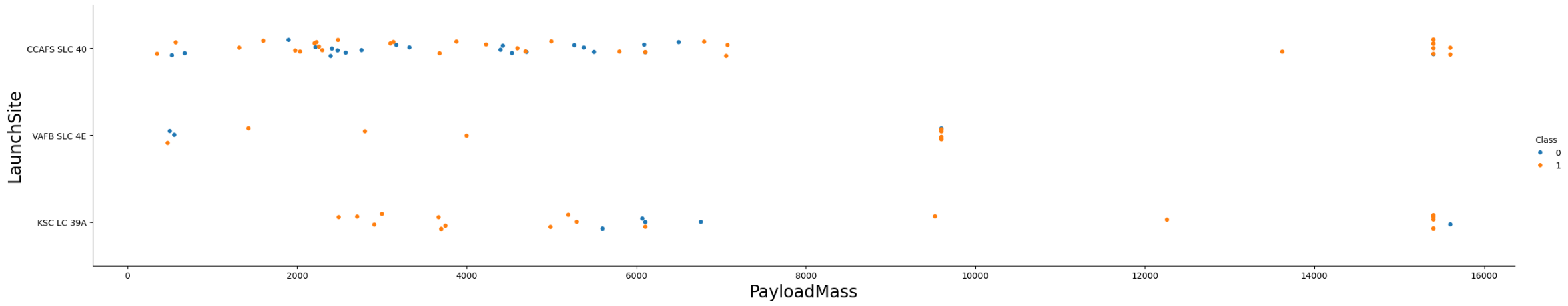


Flight Number vs. Launch Site



- Most of the flights are launch at the CCAFS SLC 40 site.
- The later flights numbers seem to launch successfully.

Payload vs. Launch Site

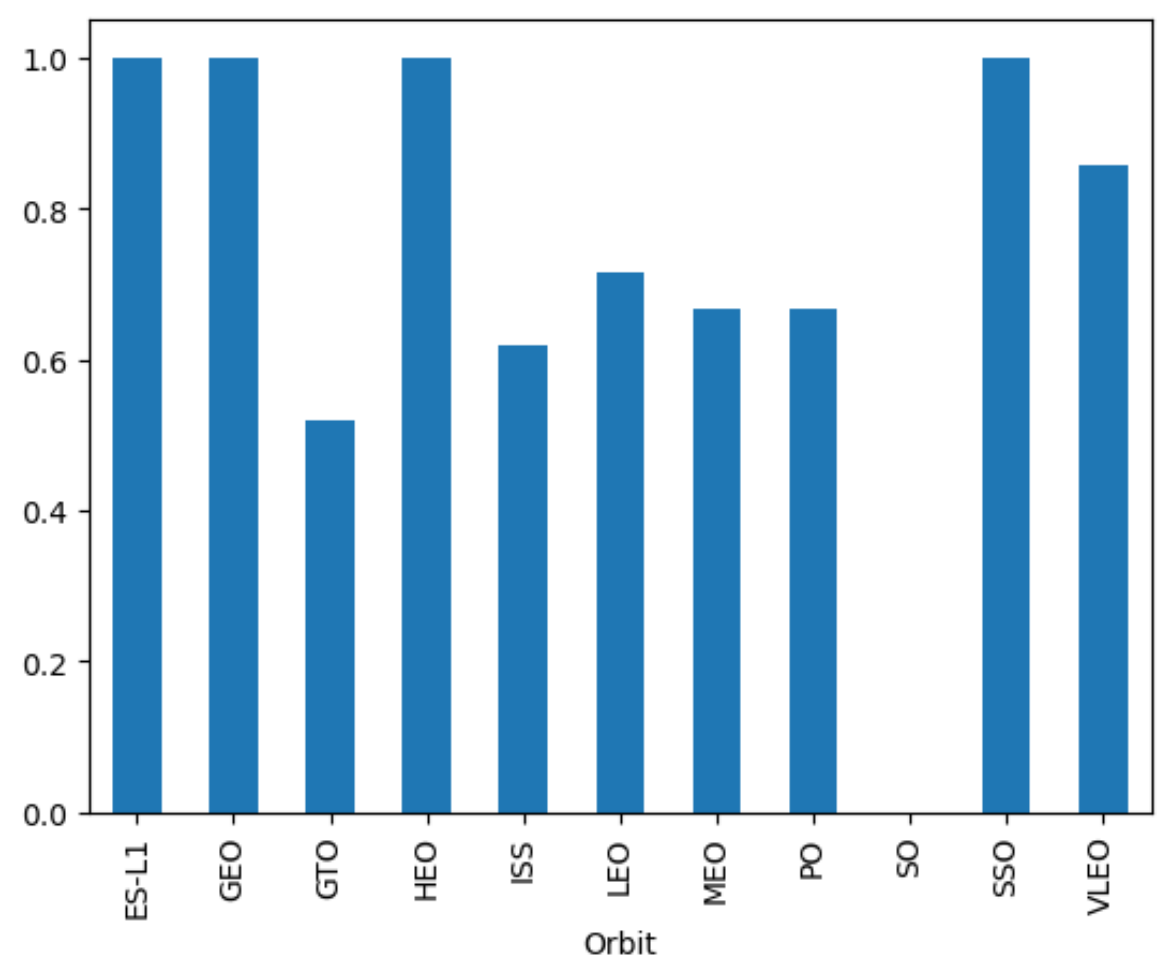


- Payload over 8000 has highly successful rate.
- Payload under 8000 are mostly launch at CCAFS SLC 40 and KSC LC 39A.

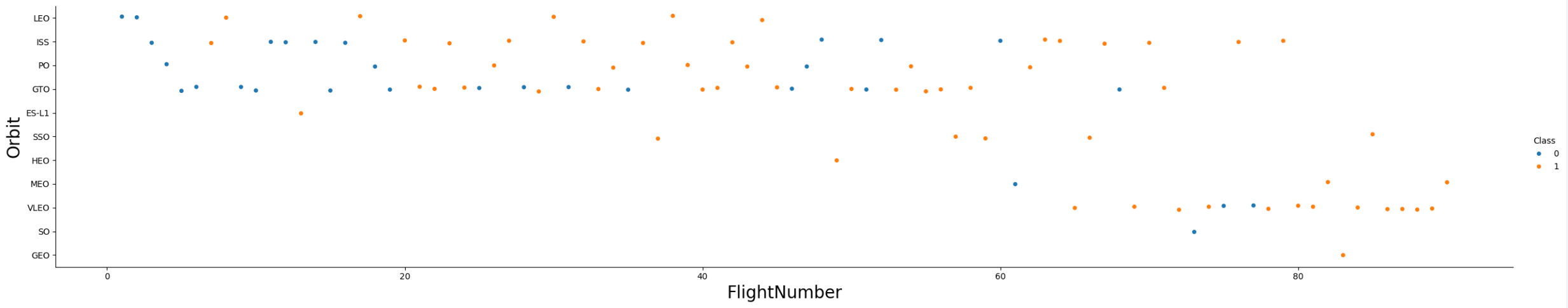
Success Rate vs. Orbit Type

- Orbit ES-L1, GEO, HEO, SSO has high success Rate.

One significant thing is
: SO orbit has no success rate shown.

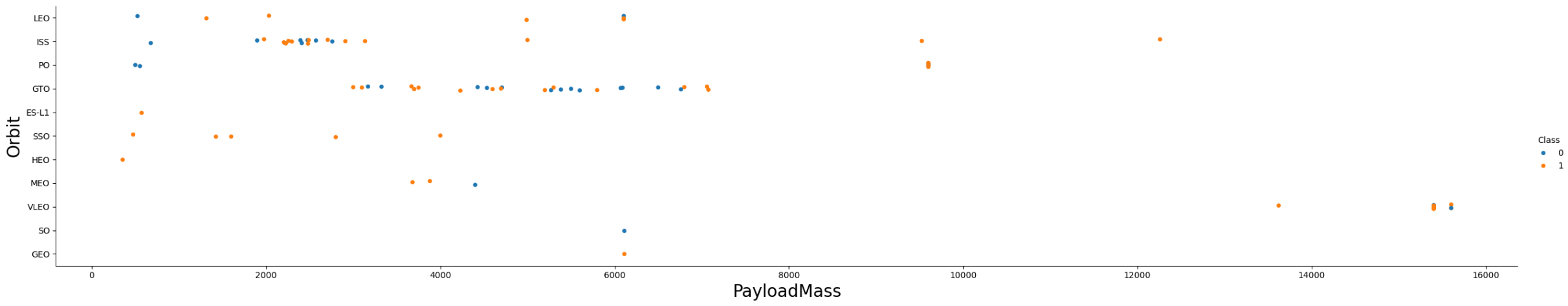


Flight Number vs. Orbit Type



- Flight number above 20 has greater chance to successfully launched.

Payload vs. Orbit Type



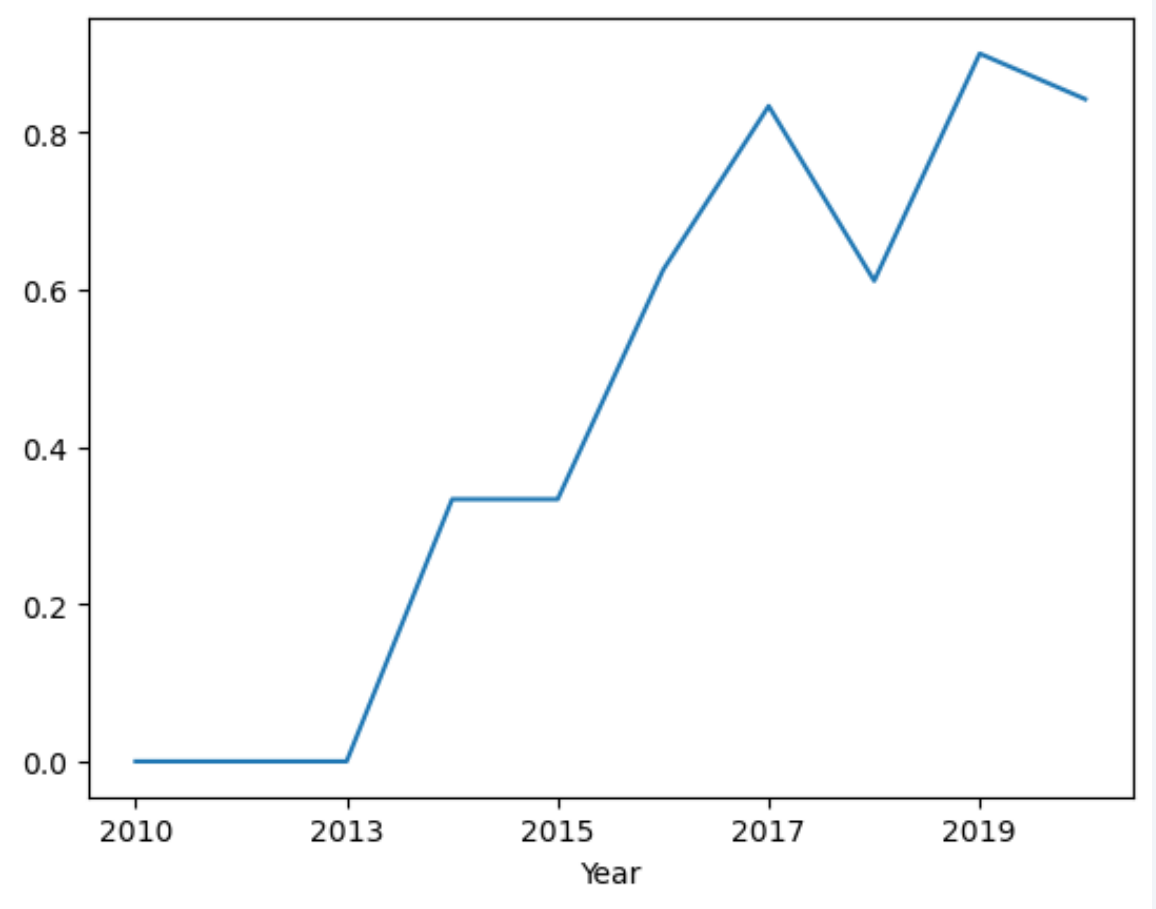
- Most of the orbit are around the payload mass 2000 – 8000.

Launch Success Yearly Trend

- The Success rate is increasing year by year.

2013-2015, we can see the success rate freeze around 0.4 and after that sharply increase.

•2017-2019, there can be seen a little bit of decrease in success rate but then increase again.



All Launch Site Names

- 4 Launch Site
- Select

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Select all attributes from DataSet,like CCA.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the sum of 'payload_mass__kg_' where PAYLOAD is like '%CRS%'and put that sum value to TOTAL_PAYLOAD

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Calculate the average of 'payload_mass__kg_' from the DataSet where the Booster_Version is 'F9 v1.1' and name that average as 'AVG_PAYLOAD'

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Use min function to select first success date from the data set and condition is 'LANDING_OUTCOME' must be 'Success(ground pad)'

FIRST_SUCCESS_GP
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Select Distinct Booster_Version where the payload_mass_kg_ is between 4000 and 6000 and the 'Landing_Outcome' is success drone ship.

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

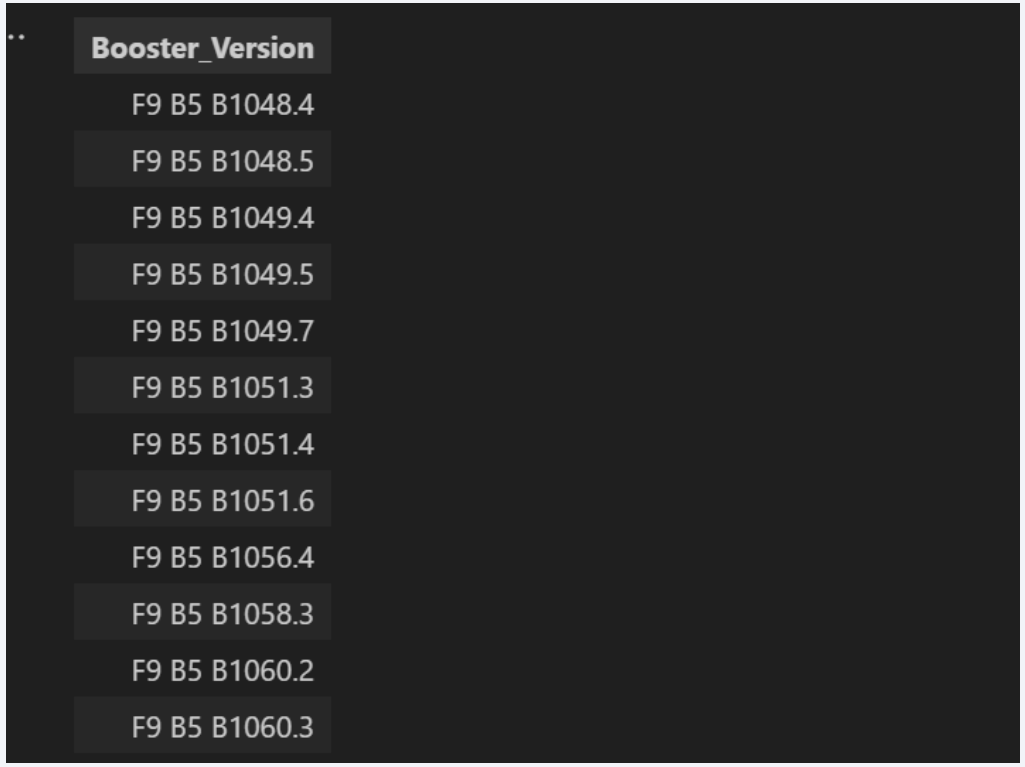
Total Number of Successful and Failure Mission Outcomes

- Select the 'Mission_Outcome' and count the numbers of each outcome group by 'Mission_Outcome'

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

Boosters Carried Maximum Payload

- Select distinct 'Booster_Version' from the dataset, where 'Payload_mass__kg' is maximum



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

2015 Launch Records

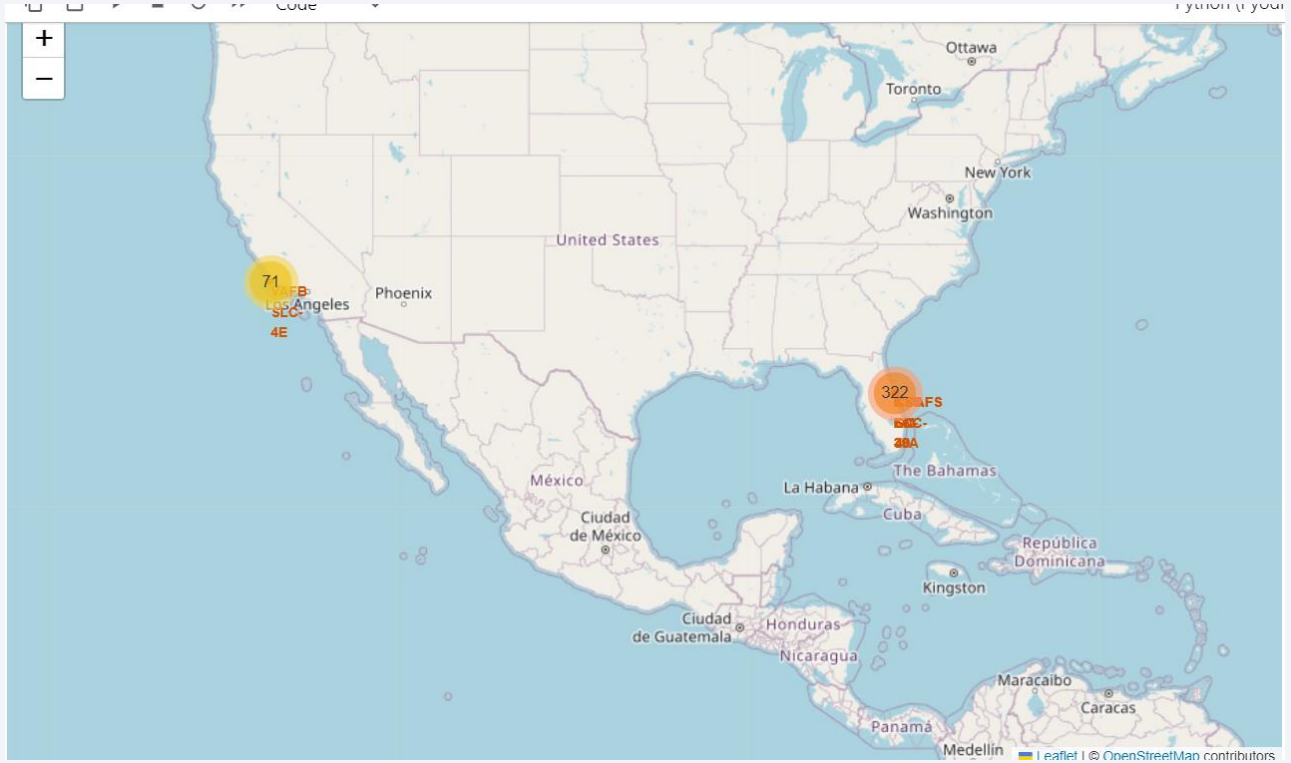
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	QTY
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

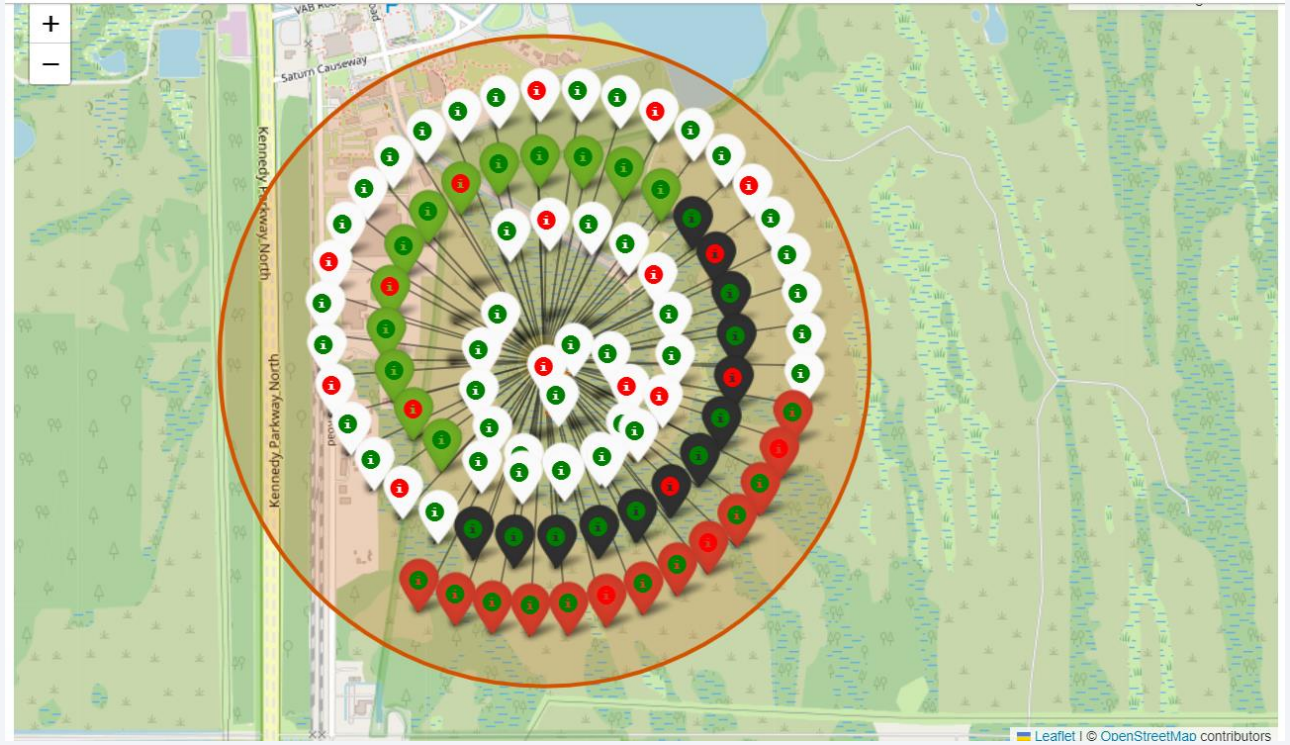
Launch Site Locations

Launch sites are near sea, probably by safety, but not too far from roads and railroads

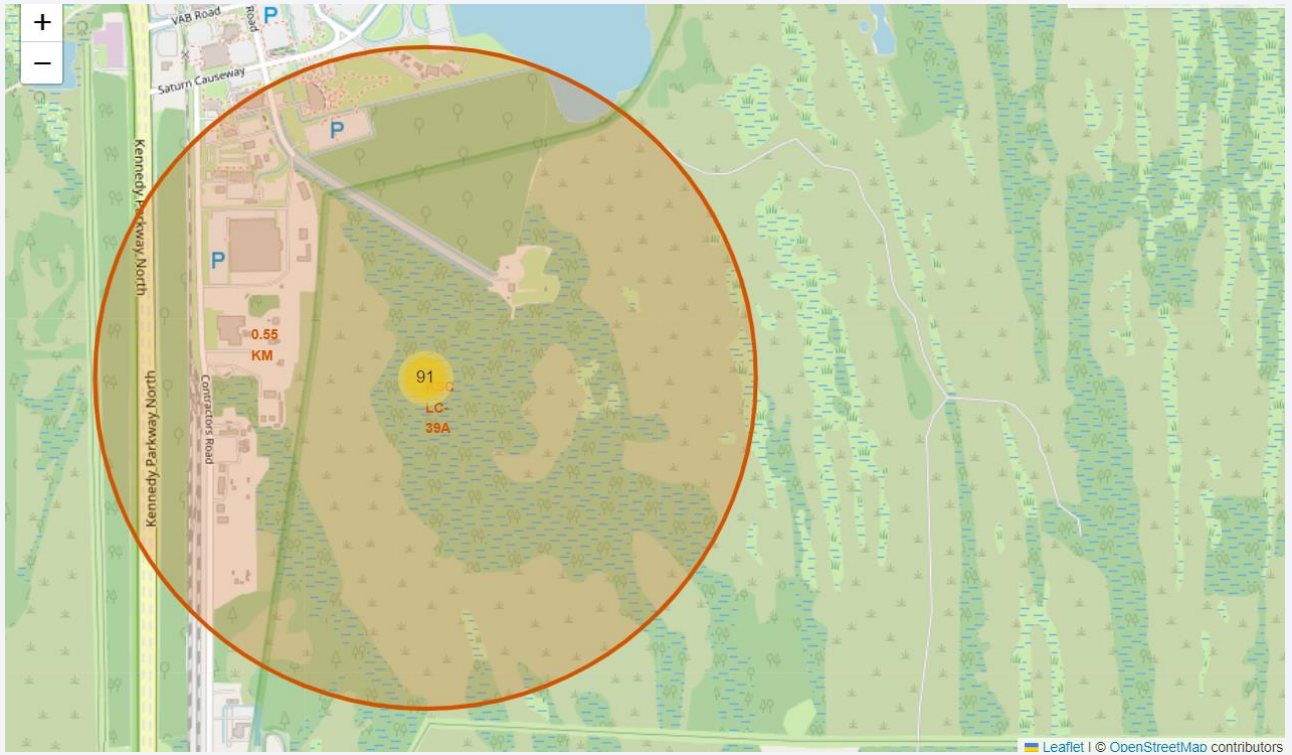


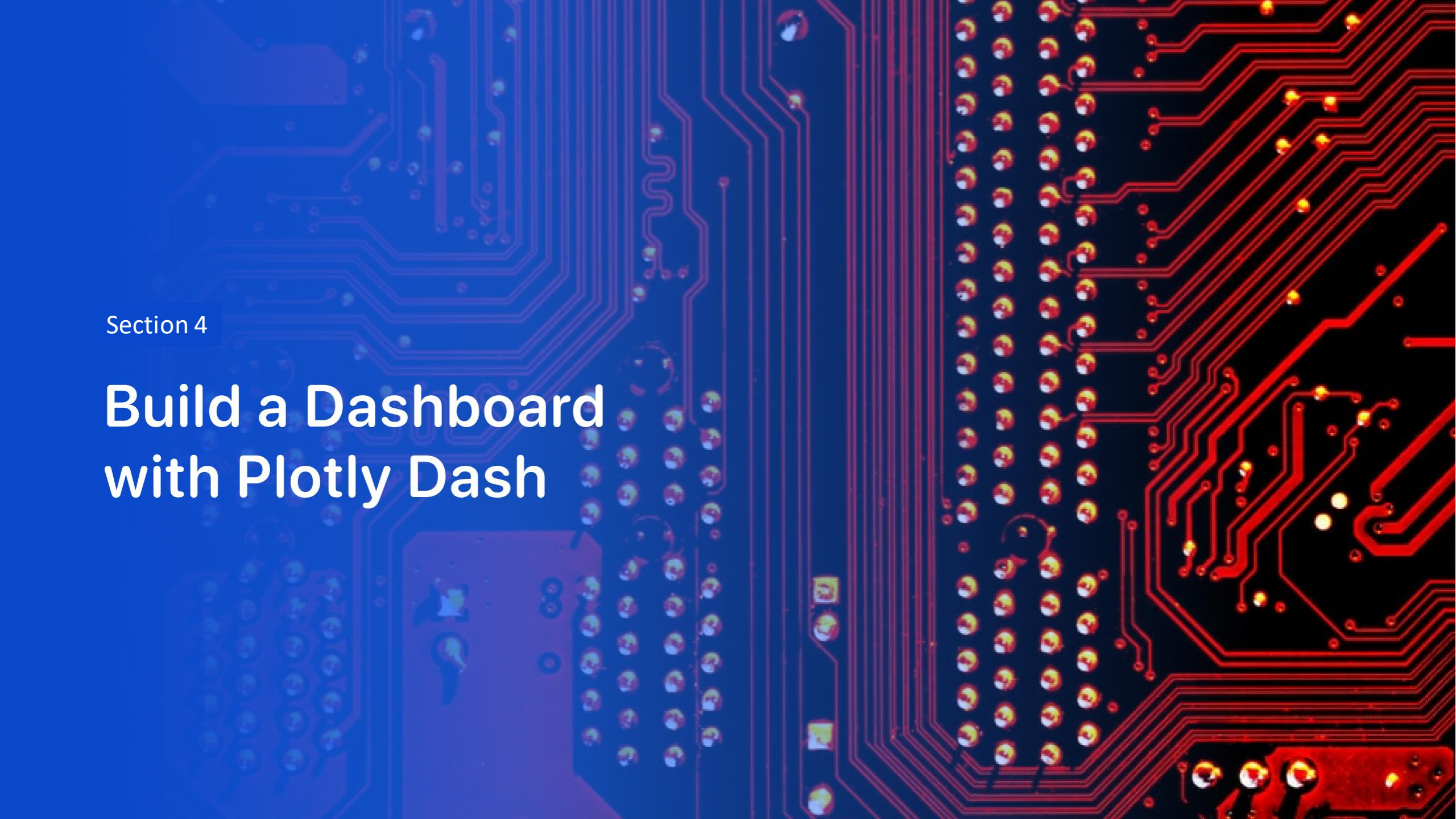
<Folium Map Screenshot 2>

- Launch Successfully are colored by green and unsuccessful launch are colored by red.



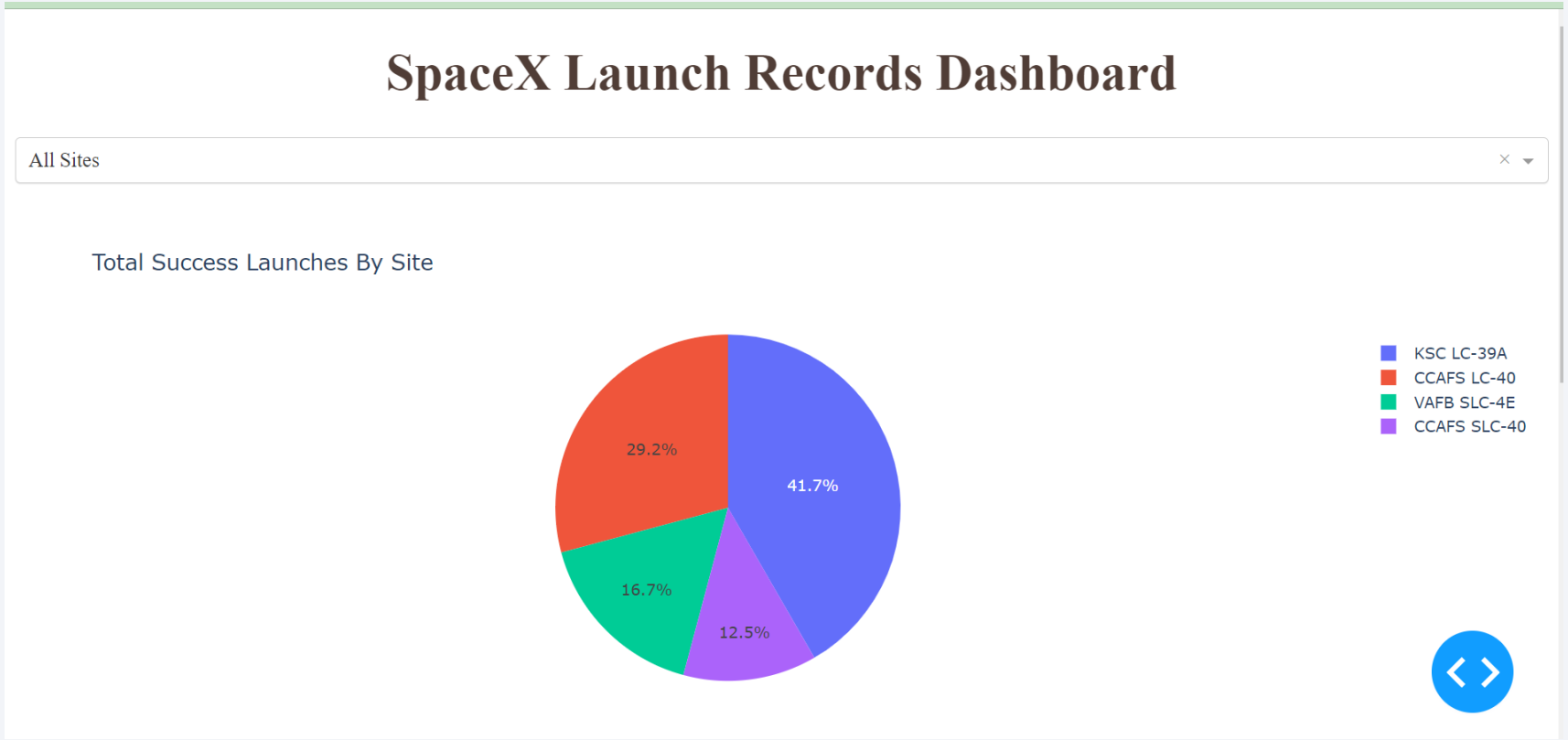
Safety Environment of Launch Sites




Section 4
Build a Dashboard with Plotly Dash

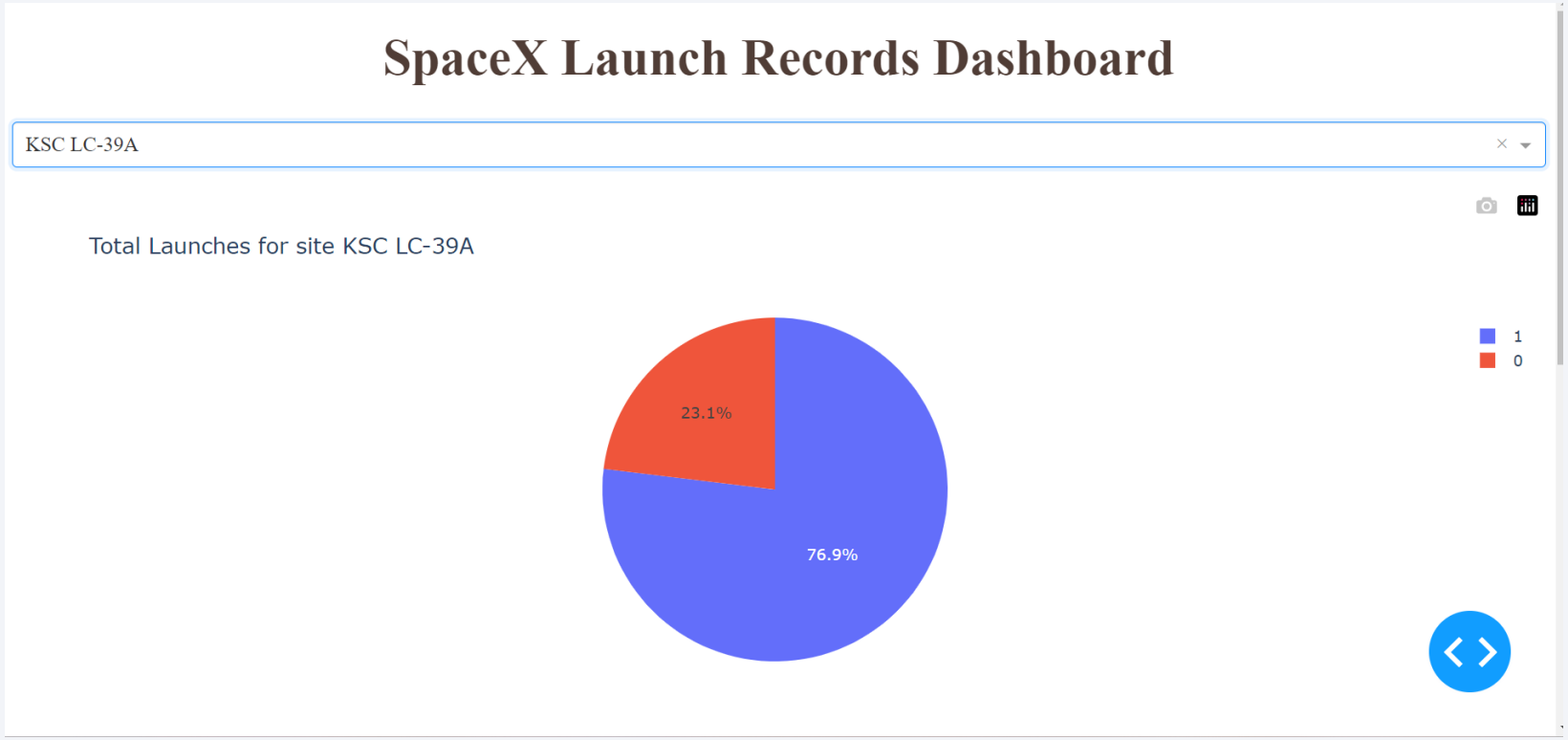
Total Success Launches By Site

The site KSC LC-39A has the most success launches among other sites

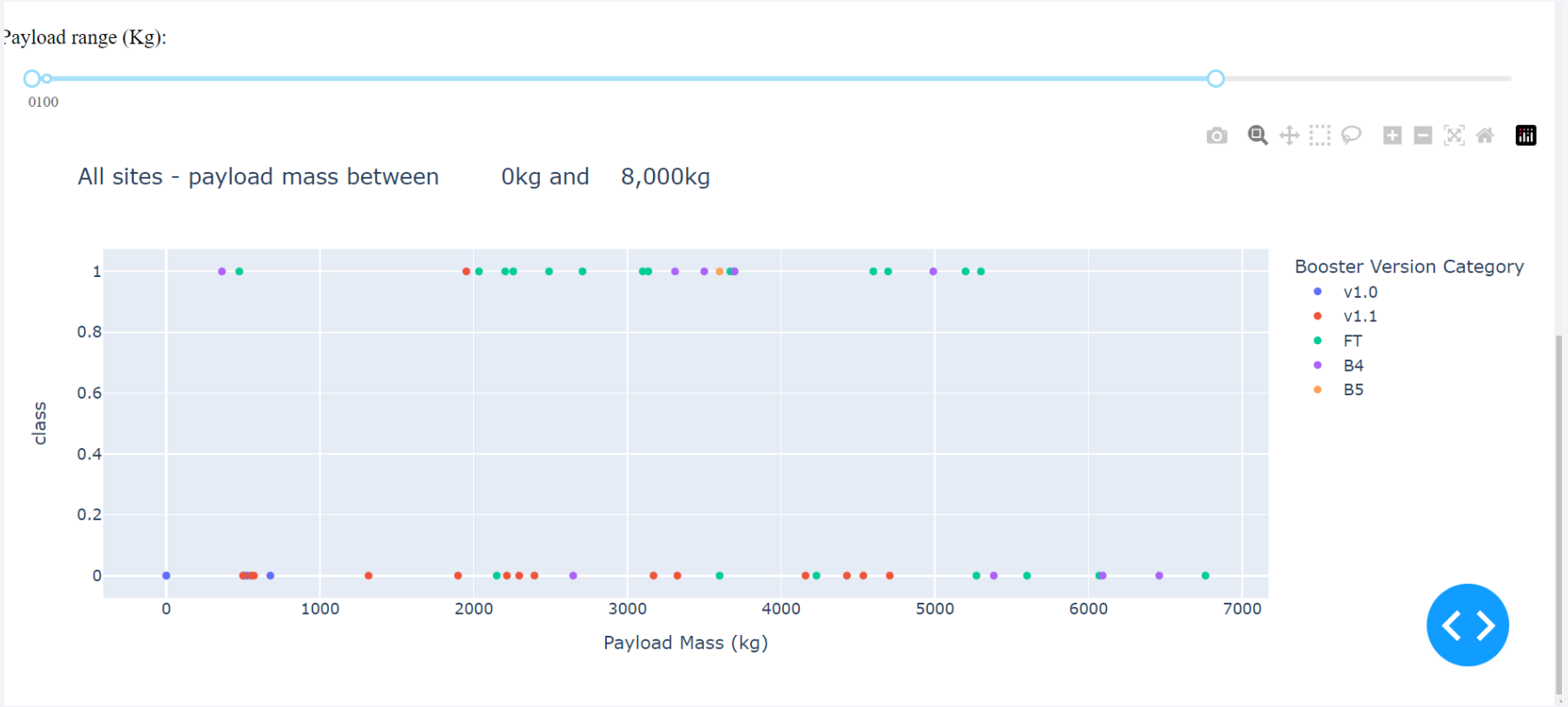


Success Launches Ratio of KSC LC-39A

The site KSC LC-39A has the success launches rate of 76.9& and unsuccess rate of 23.1%



Payload vs. Launch Outcome scatter plot for all sites

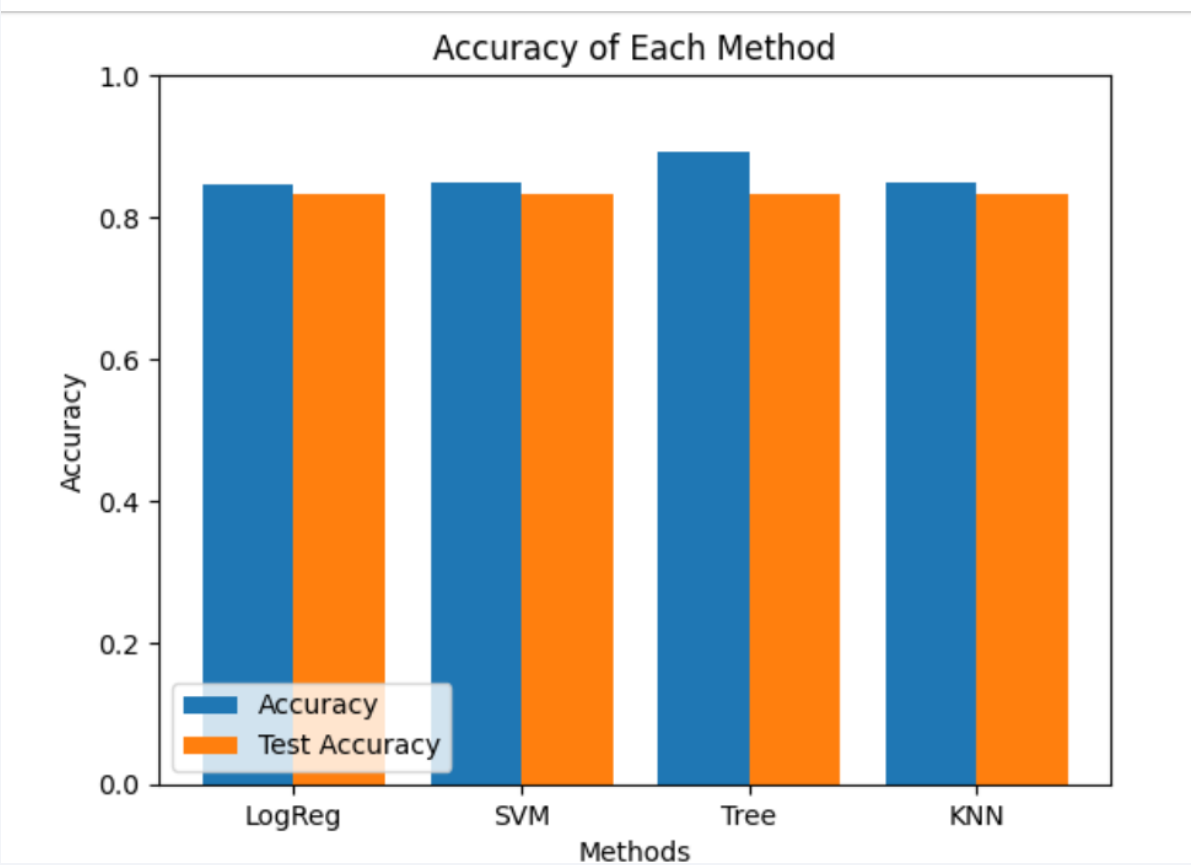


Section 5

Predictive Analysis (Classification)

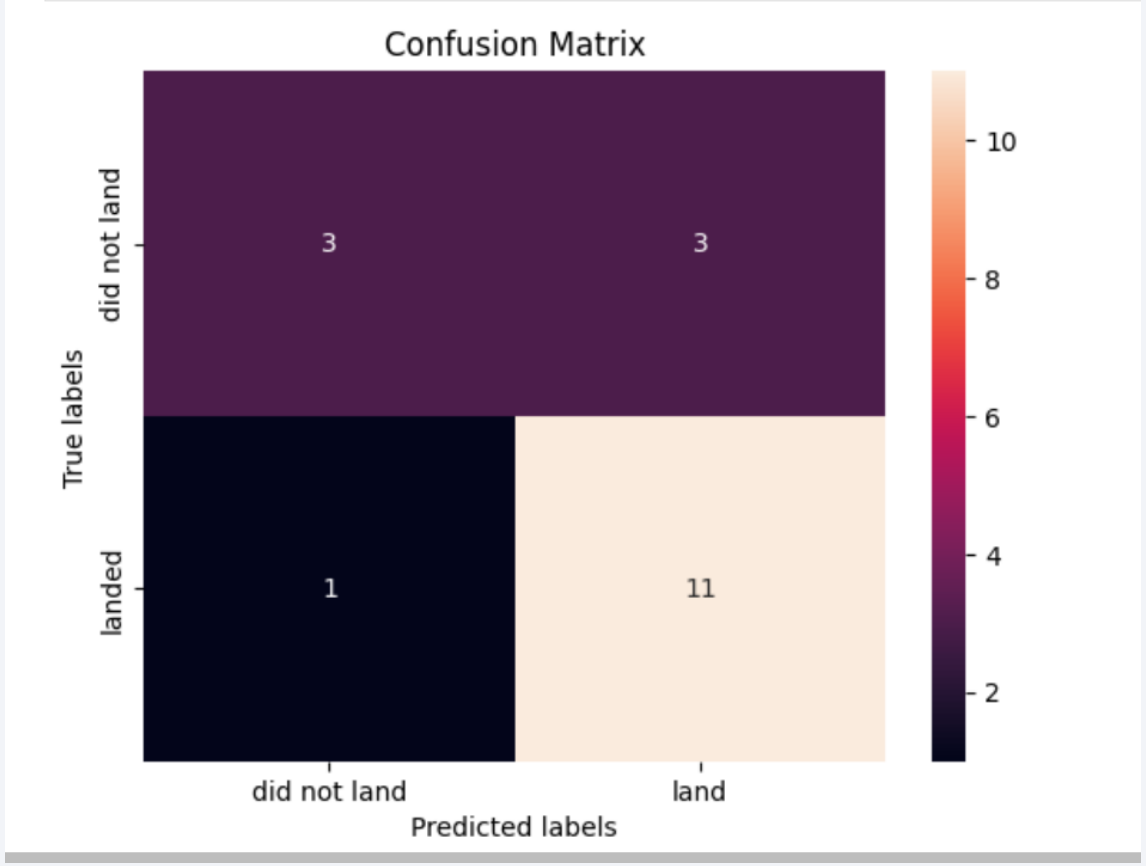
Classification Accuracy

- This Bar Chart show that among 4 models, Decision Tree Classification has the higher accuracy score.



Confusion Matrix

- This is Confusion Matrix of Decision Tree Classification with our data.



Conclusions

- After analysing, this dataset, we have known some important facts
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

Thank you!