IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmed Mohamed Mostafa
2023/02/22

Outline

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

Executive Summary

- Summary of methodologies

Here is the methodologies we use in our project

❑ Data collection about the Falcon 9 first-stage landings from SpaceX site using web scraping API

Source of data https://api.spacexdata.com/v4/rockets

❑ Perform some data wrangling ,Exploratory Analysis data and Visualization using python and interactively with Plotly Dash.

❑ Machine learning using SVM, Classification Trees, and Logistic Regression.

- Summary of all results

Find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression. Then find the method that performs best using test data.

Introduction

- Project background and context

❑The aim is to assess the feasibility of Space Y, the new company, to rival Space X.

- Problems you want to find answers

❑Predicting the successful landings of the first stage of rockets is the optimal approach to estimate the total cost of launches.

❑ Determine the ideal location for conducting launches.

Section 1

Methodology

Methodology

Executive Summary

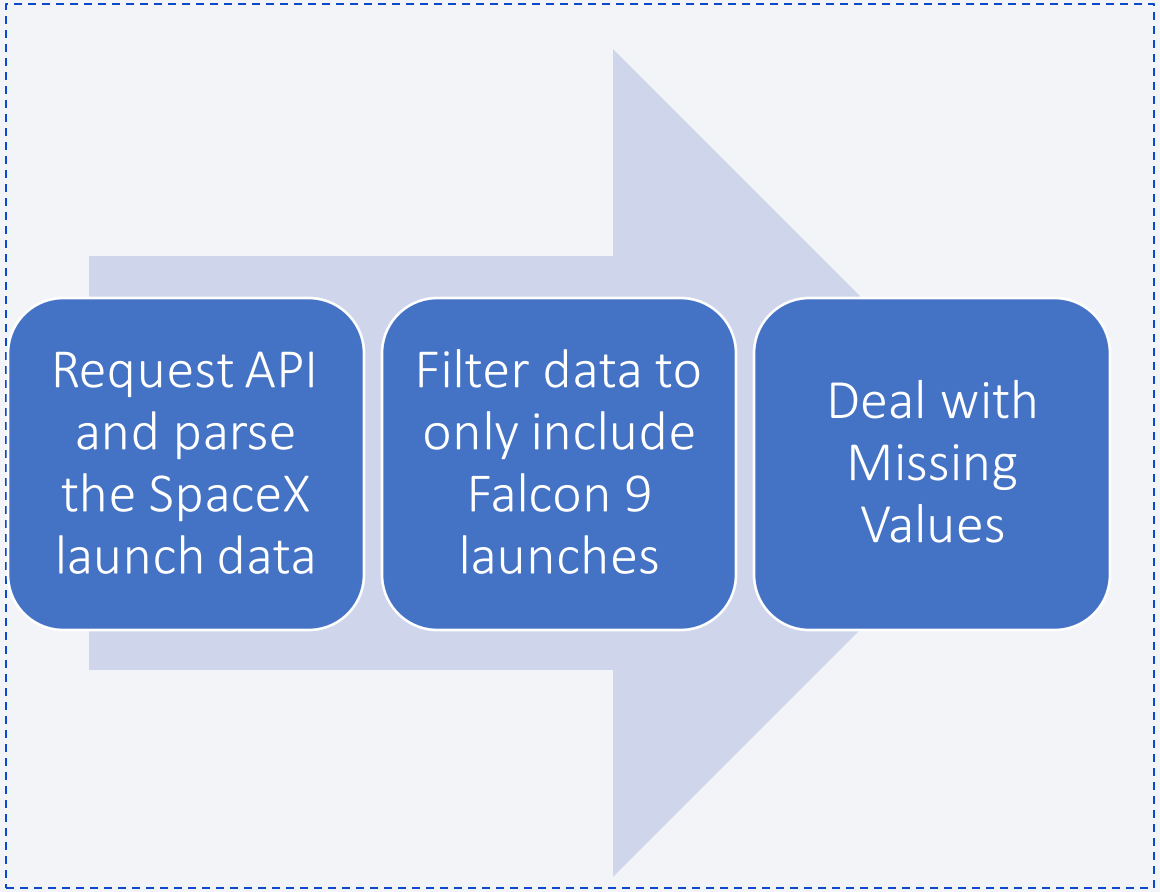
- Data collection methodology:
  - Using web scraping API python from 2 sources
  - https://api.spacexdata.com/v4/rockets/
  - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
- Perform data wrangling
  - We'll conduct Exploratory Data Analysis (EDA) to identify data patterns and establish the training label for supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

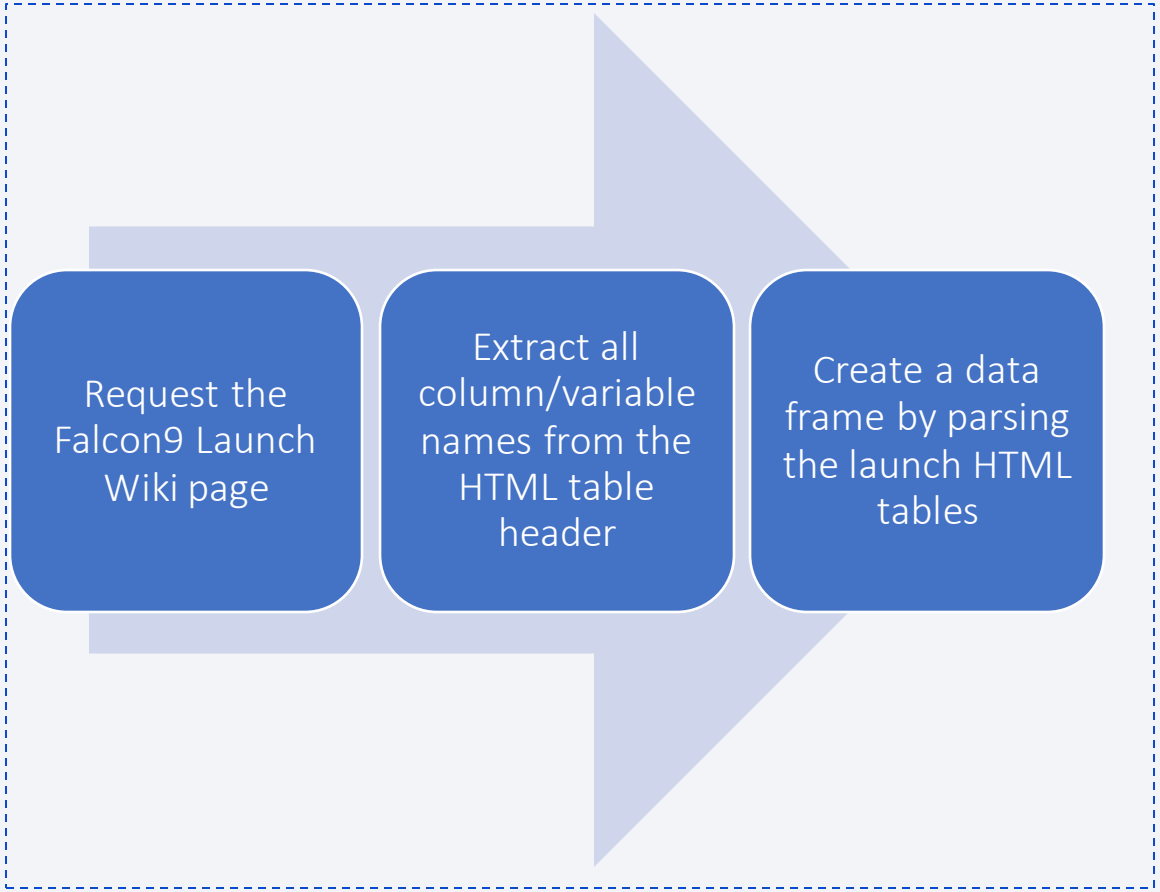
Data Collection – SpaceX API

- SpaceX provides a public API that allows access to data;
- The flowchart indicates that the API was utilized, and the resulting data was stored.
- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20(2).ipynb



Data Collection - Scraping

- Information about SpaceX launches can be acquired from Wikipedia;
- The flowchart illustrates that the data is obtained from Wikipedia and then stored.
- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



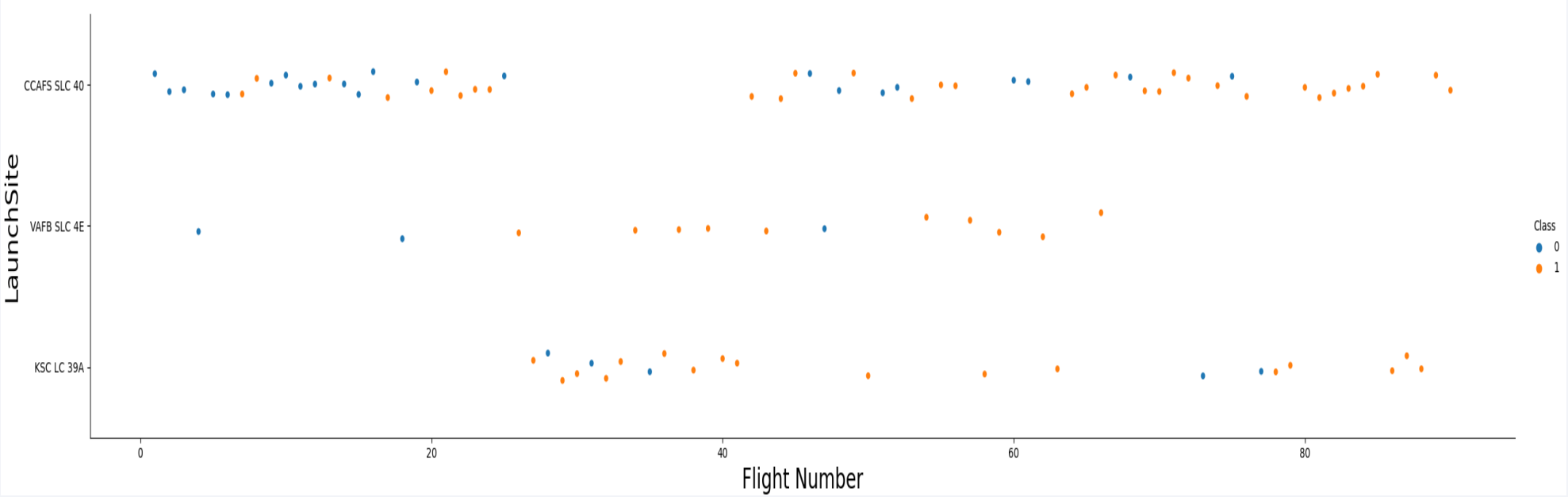
Data Wrangling

- The dataset was first subjected to Exploratory Data Analysis (EDA).
- Next, the number of launches per site, the frequency of each orbit, and the frequency of mission outcomes per orbit type were computed.
- Subsequently, the landing outcome label was derived from the Outcome column.

- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Use the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber,set the y to Launch Site and set the parameter hue to 'class'



- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

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
- 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps

❑Mark all launch sites on a map

❑Mark clusters to show the success/failed launches for each site on the map

❑Lines to Calculate the distances between a launch site to its proximities

- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The following graphs and plots were used :
- Percentage of launches by site
- Payload range
- Adding those plots To answer the following questions:

❑ Which site has the largest successful launches?

❑ Which site has the highest launch success rate?

❑ Which payload range(s) has the highest launch success rate?

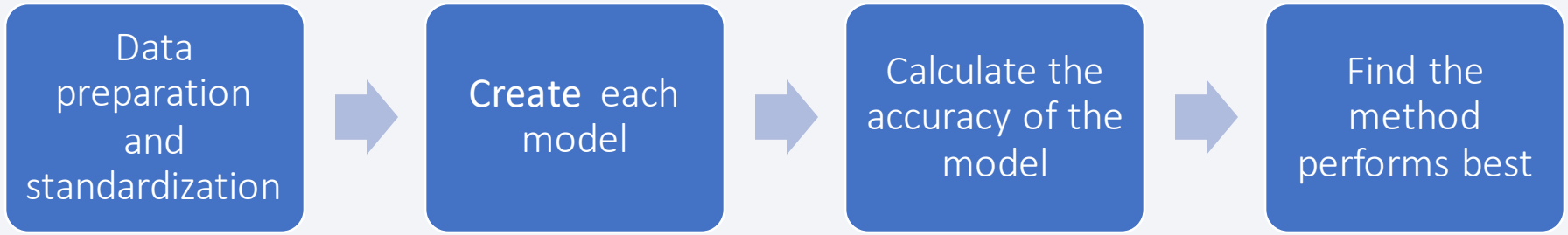
❑ Which payload range(s) has the lowest launch success rate?

❑ Which F9 Booster version (`v1.0`, `v1.1`, FT, B4, B5, etc.) has the highest

- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- By Find best Hyperparameter for SVM, Classification Trees ,Logistic Regression and k nearest neighbors



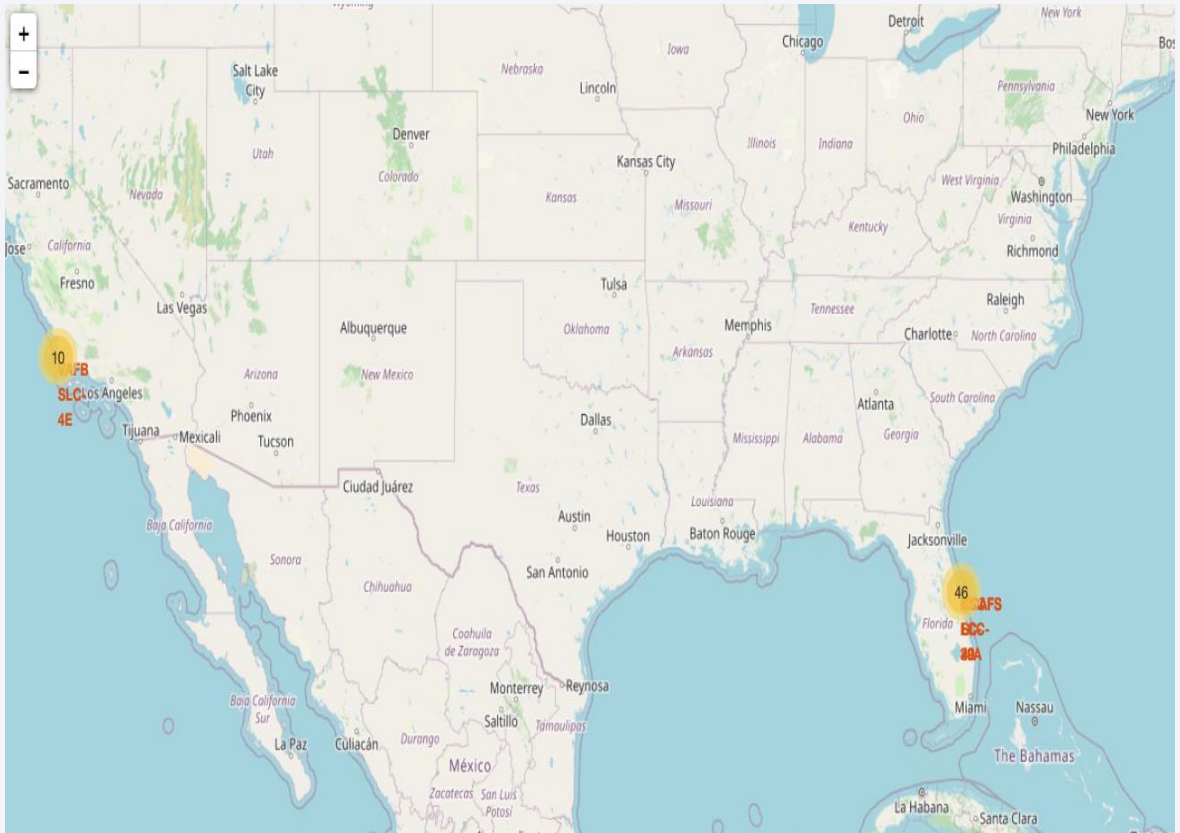
- https://github.com/Ahmed-Elshreef/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

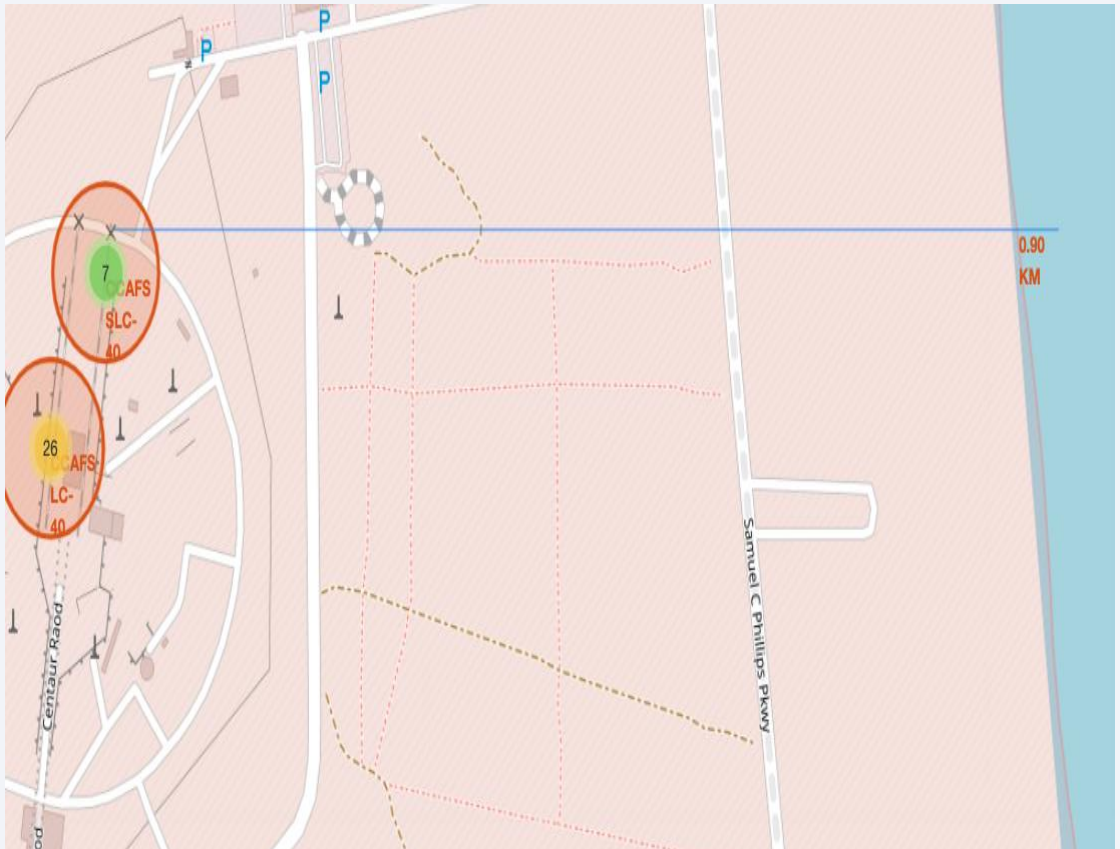
Results

- Exploratory data analysis results
- We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- We observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavypayload mass(greater than 10000.
- in 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.
- the average launch success trend we observe that the sucess rate since 2013 kept increasing till 2020

Results

- Interactive analytics demo in screenshots
- Most launches happens at east cost launch sites





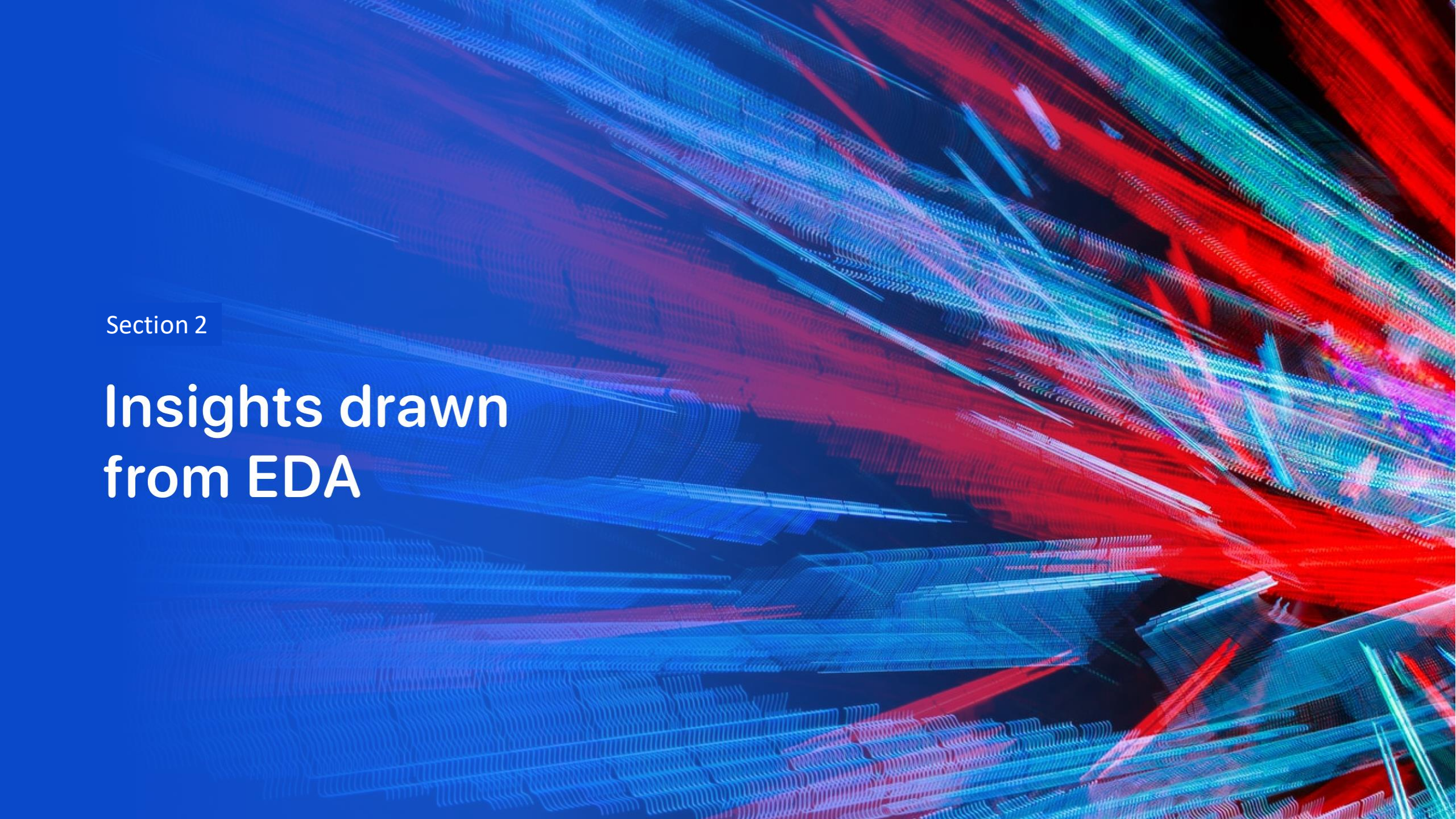
Results

- Predictive analysis results

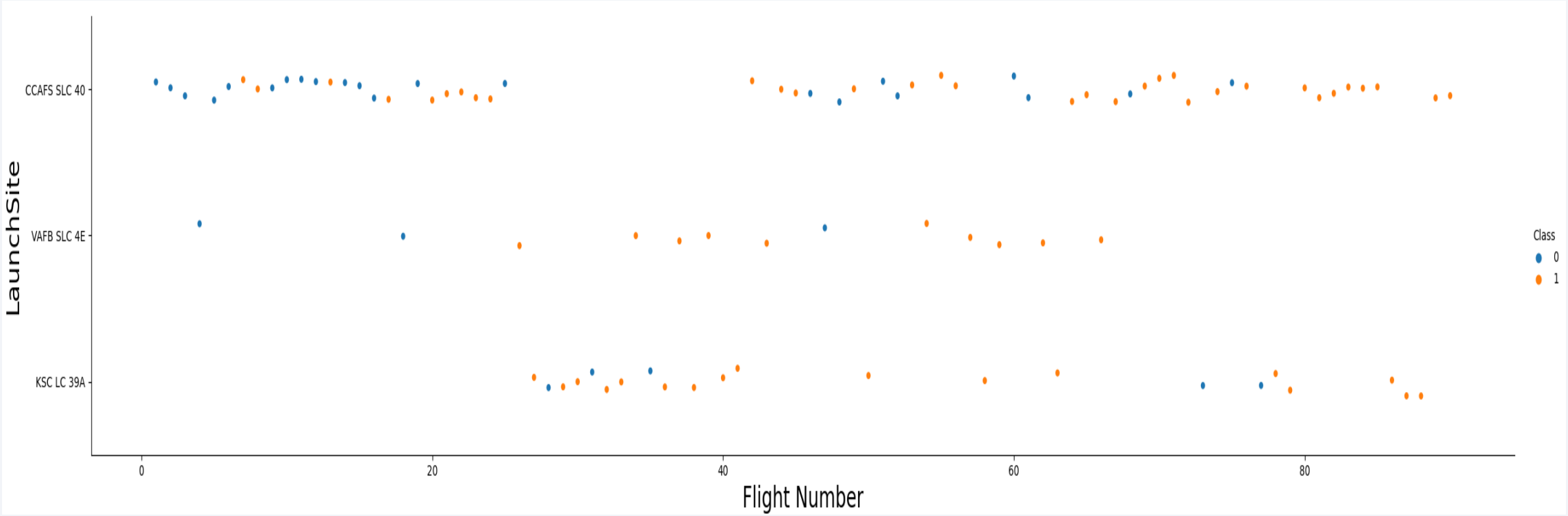
Model	Accuracy	Test Accuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.90536	0.77778
KNN	0.84821	0.83333

Section 2

Insights drawn from EDA

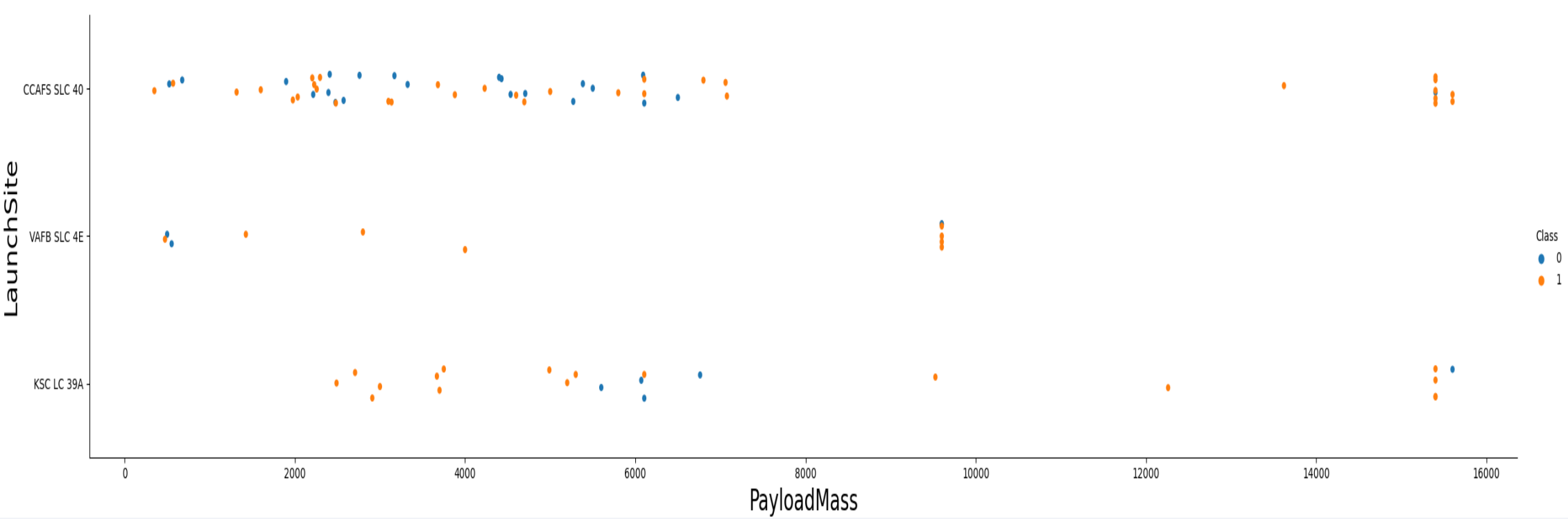


Flight Number vs. Launch Site



the best launch site nowadays is CCAF5 SLC 40 , In second place VAFB SLC 4E and third place KSC LC 39A;

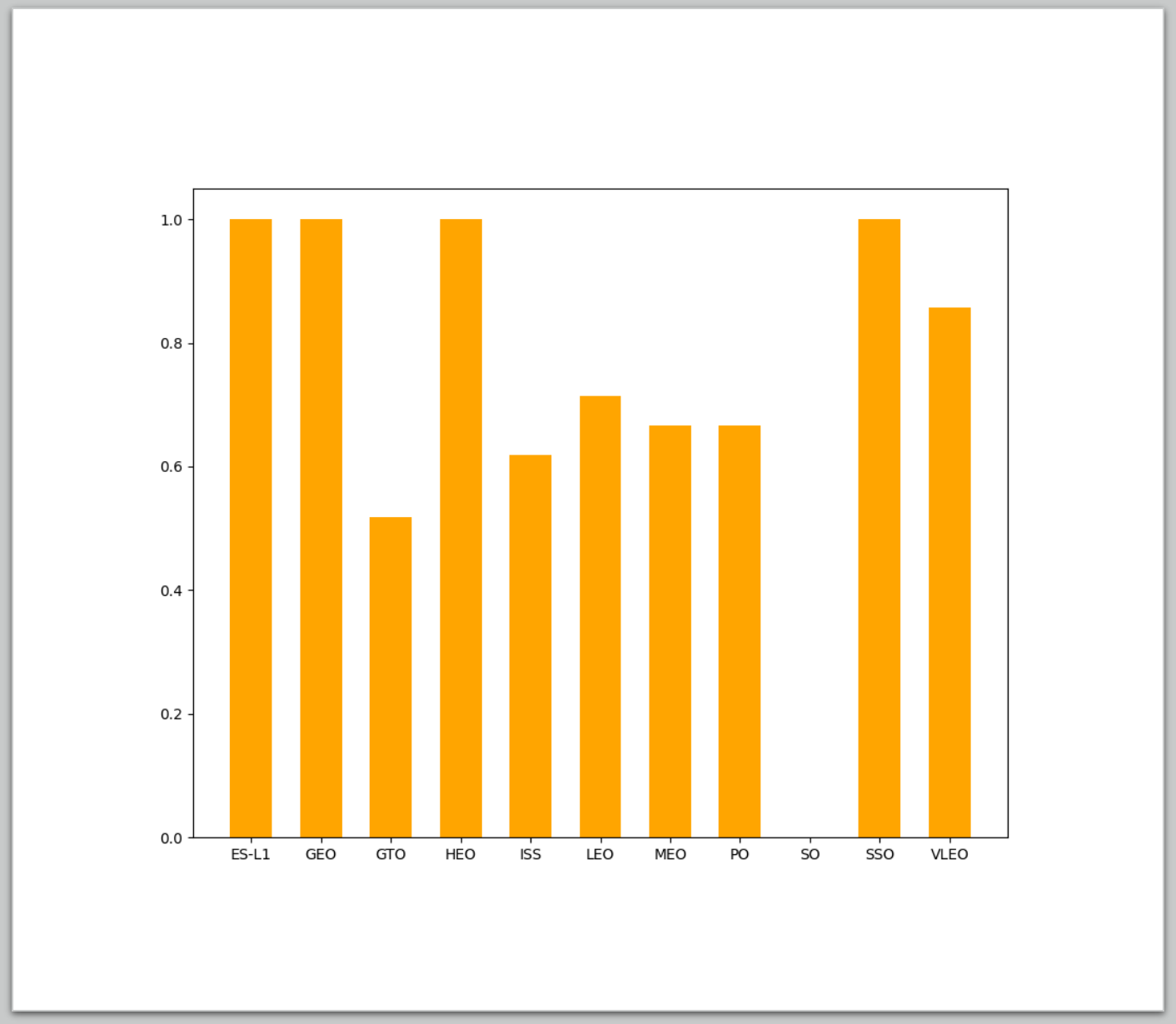
Payload vs. Launch Site



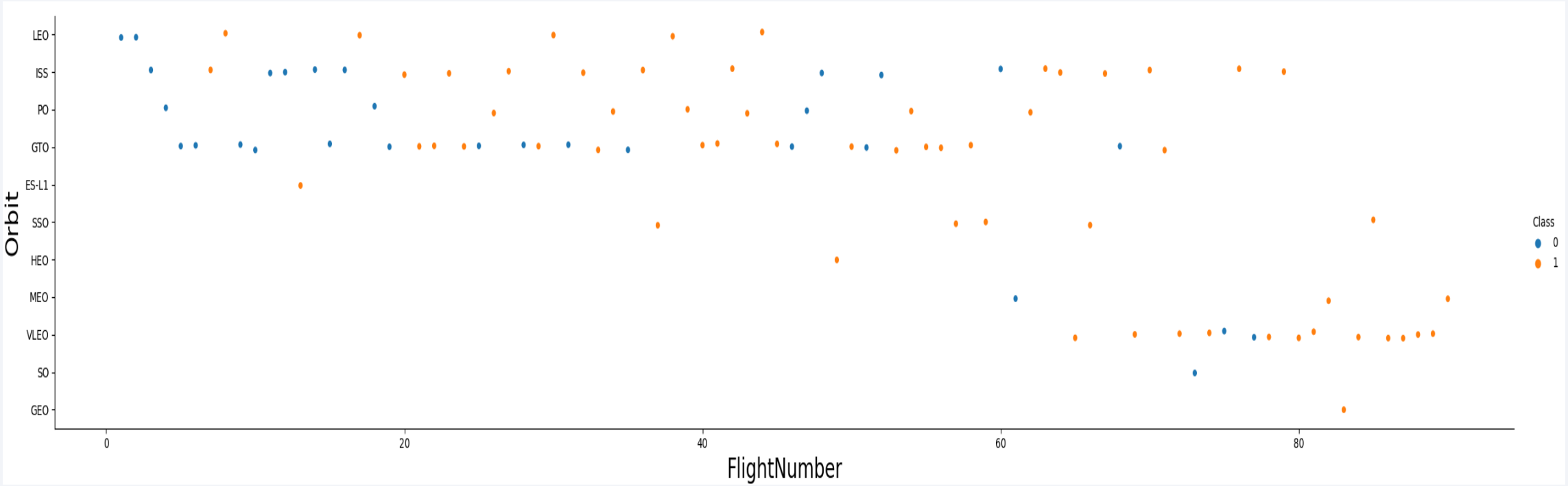
- We will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
- ES-L1;
- GEO;
- HEO;
- SSO.

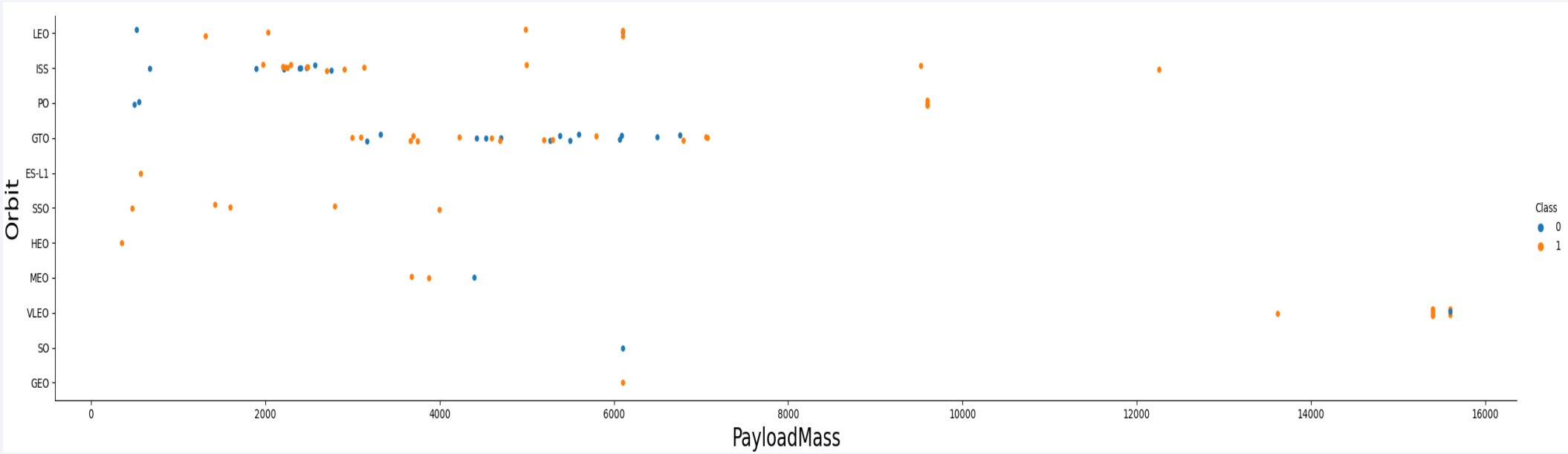


Flight Number vs. Orbit Type



- We should see that in 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.

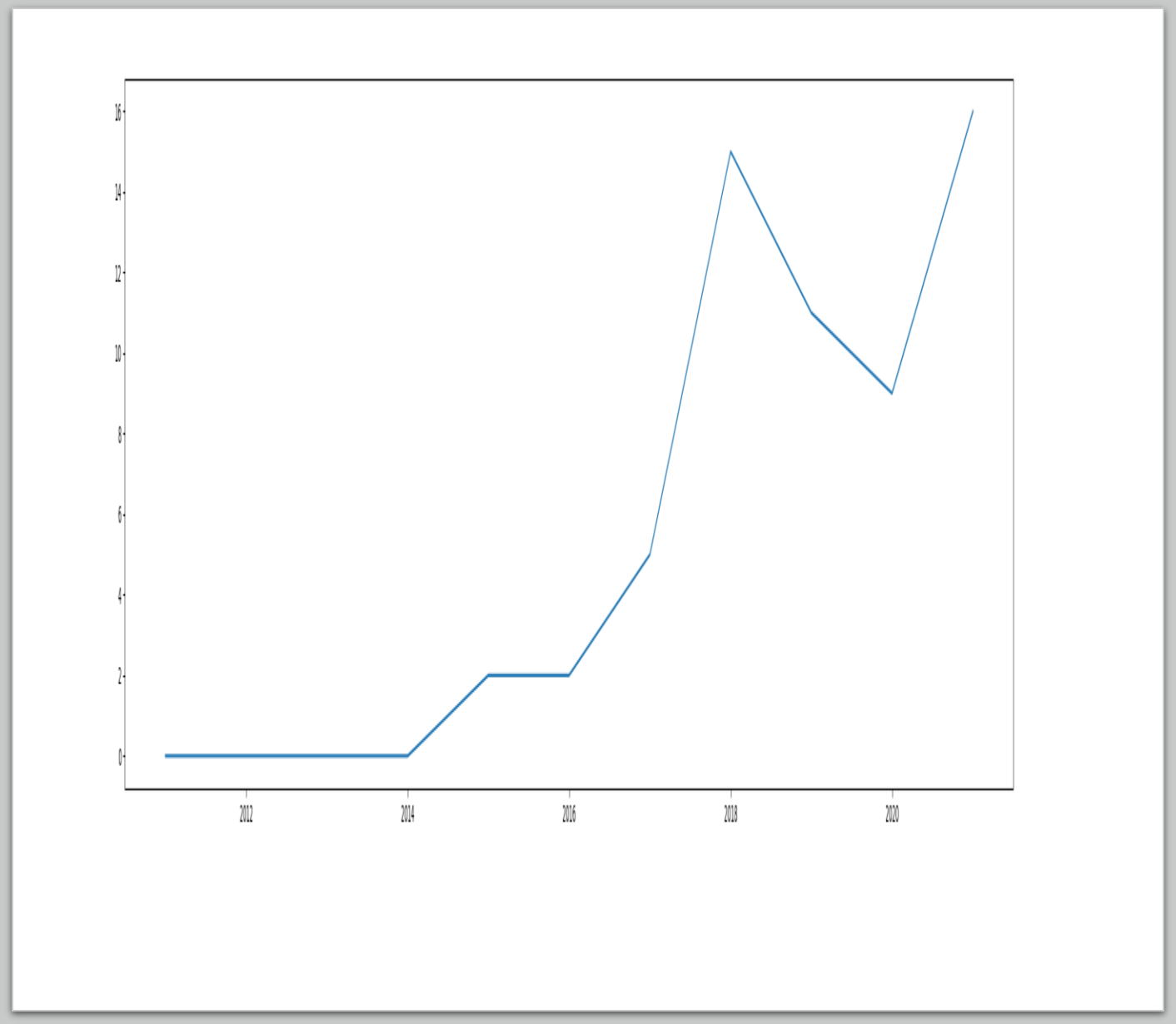
Payload vs. Orbit Type



- 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

- you can observe that the sucess rate since 2013 kept increasing till 2020



All Launch Site Names

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

- selecting unique occurrences of “launch_site” values

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

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

- There are only one site in the simple begin with "CCA"

Total Payload Mass

- the total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- the average payload mass carried by booster version F9 v1.1

AVG_PAYLOAD
2928.4

- The condition is BOOSTER_VERSION

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Min Date
2015-12-22

By filtering data by successful landing outcome on ground pad

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

- select BOOSTER_VERSION according to the filters above,

Total Number of Successful and Failure Mission Outcomes

- the total number of successful and failure mission outcomes

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

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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

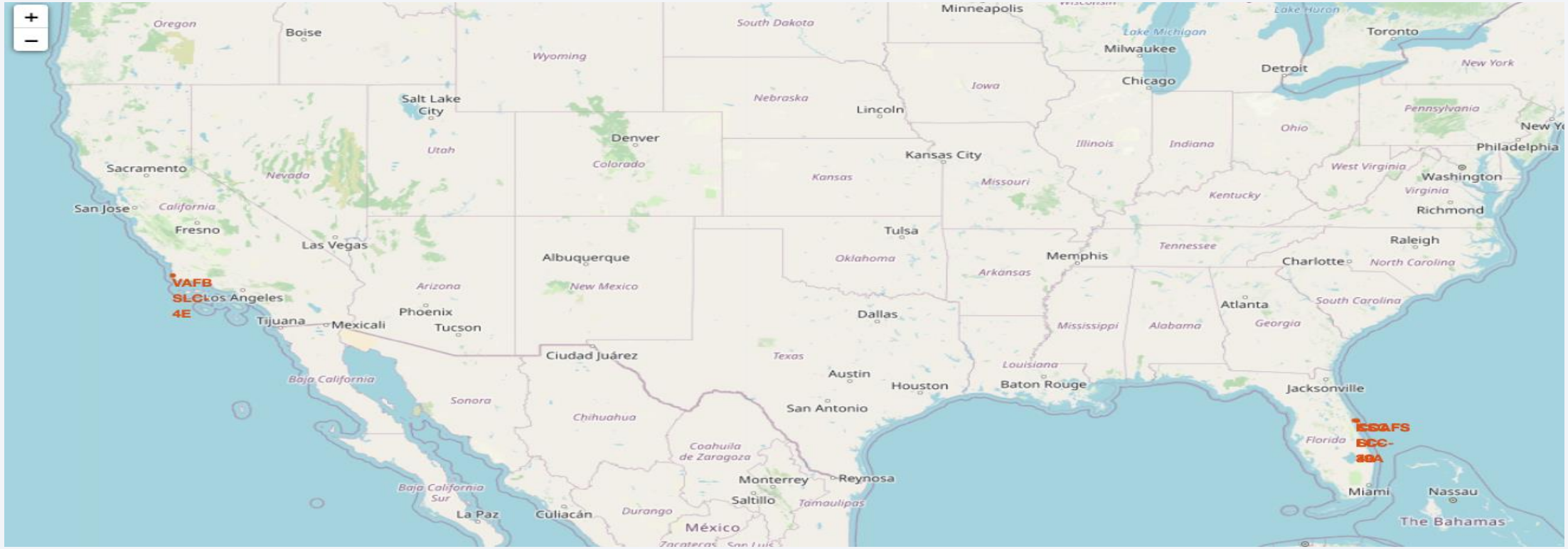
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

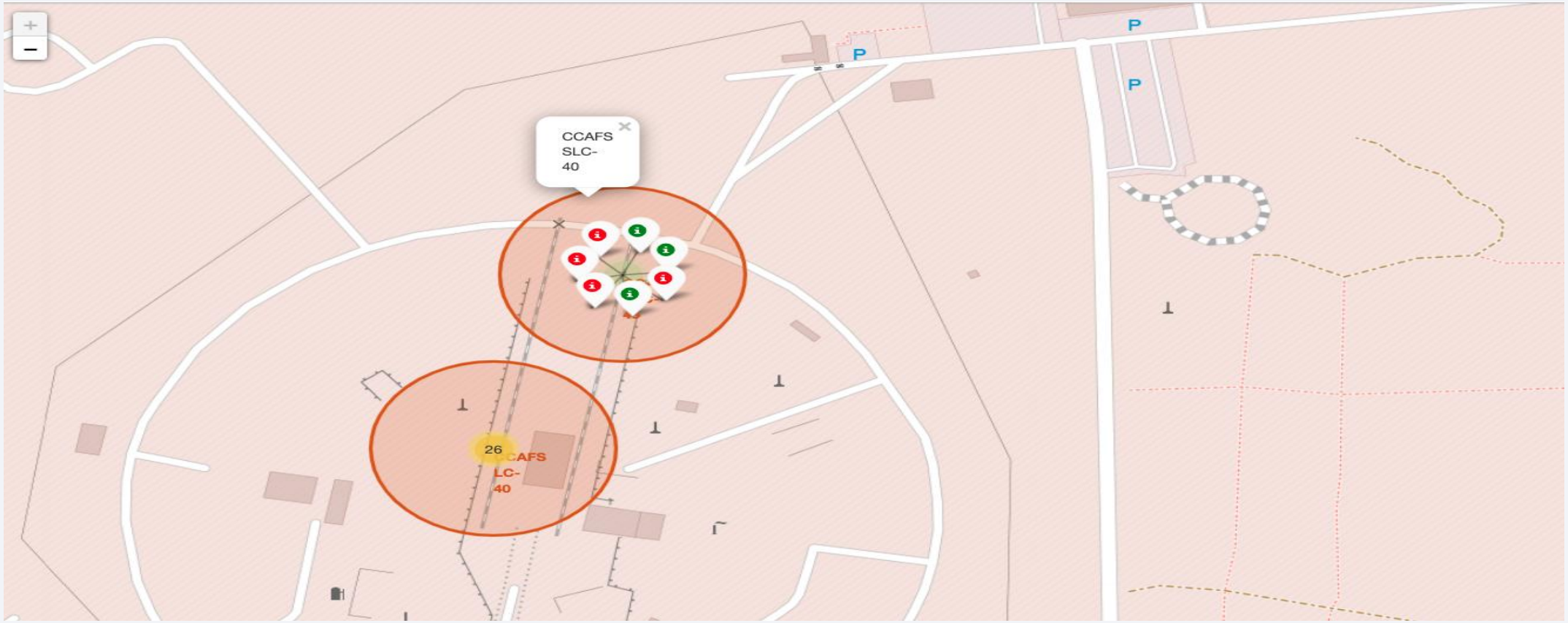

Section 3
Launch Sites
Proximities Analysis

All launch sites



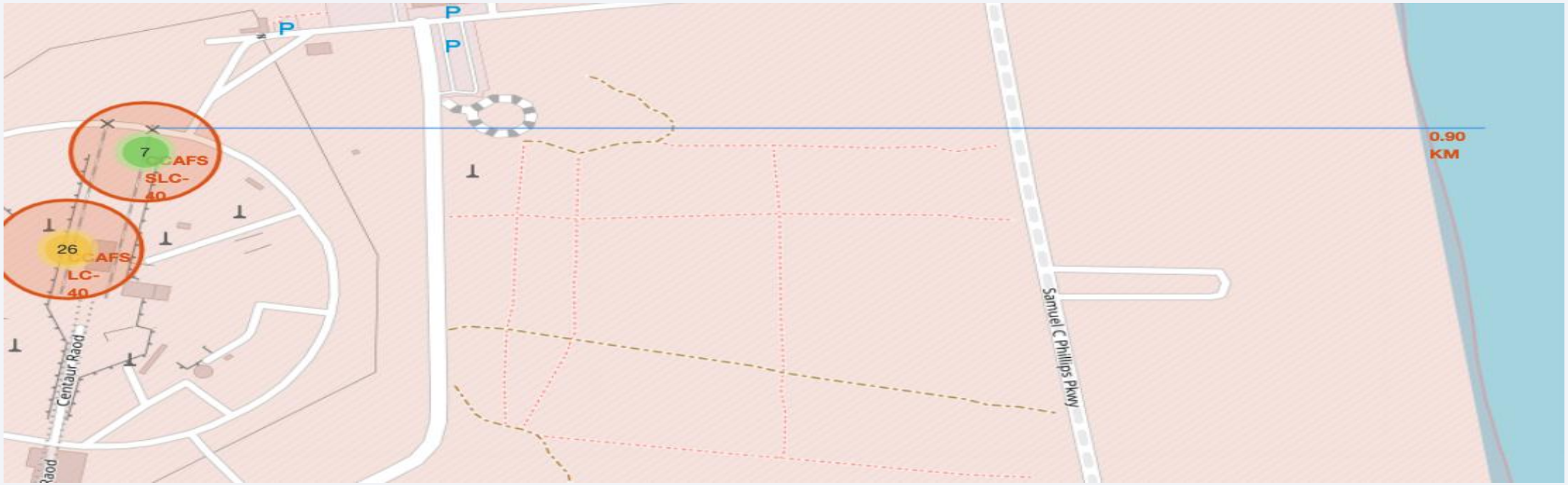
All launch sites in very close proximity to the coast and near to the Equator line

successful and failed launch site

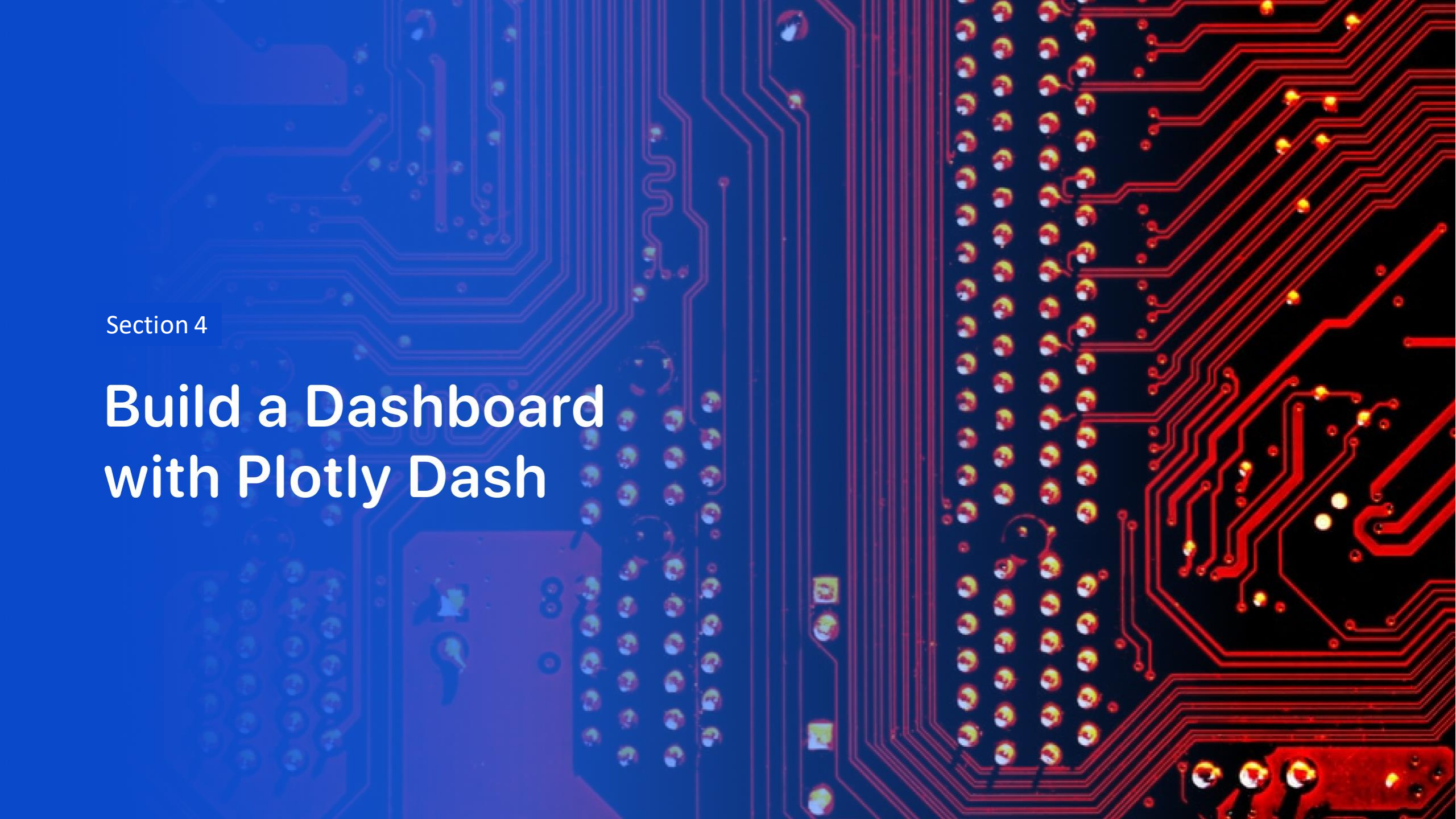


- If a launch was successful we use a green marker and if a launch was failed, we use a red marker

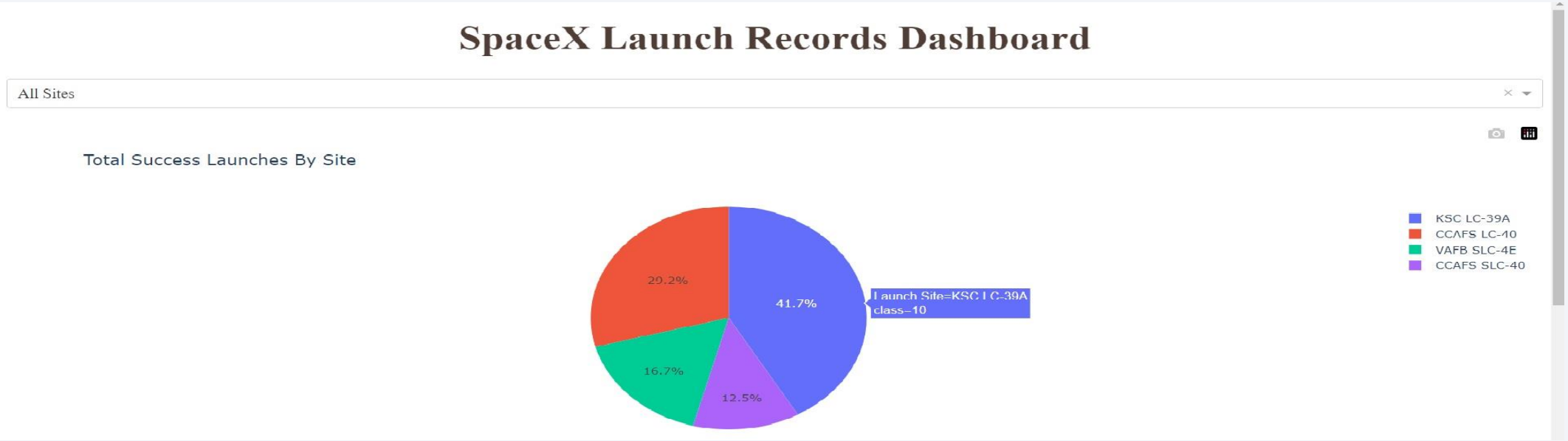
The distance between logistics and the launch site



- The distance between a launch site to its closest city, railway, highway

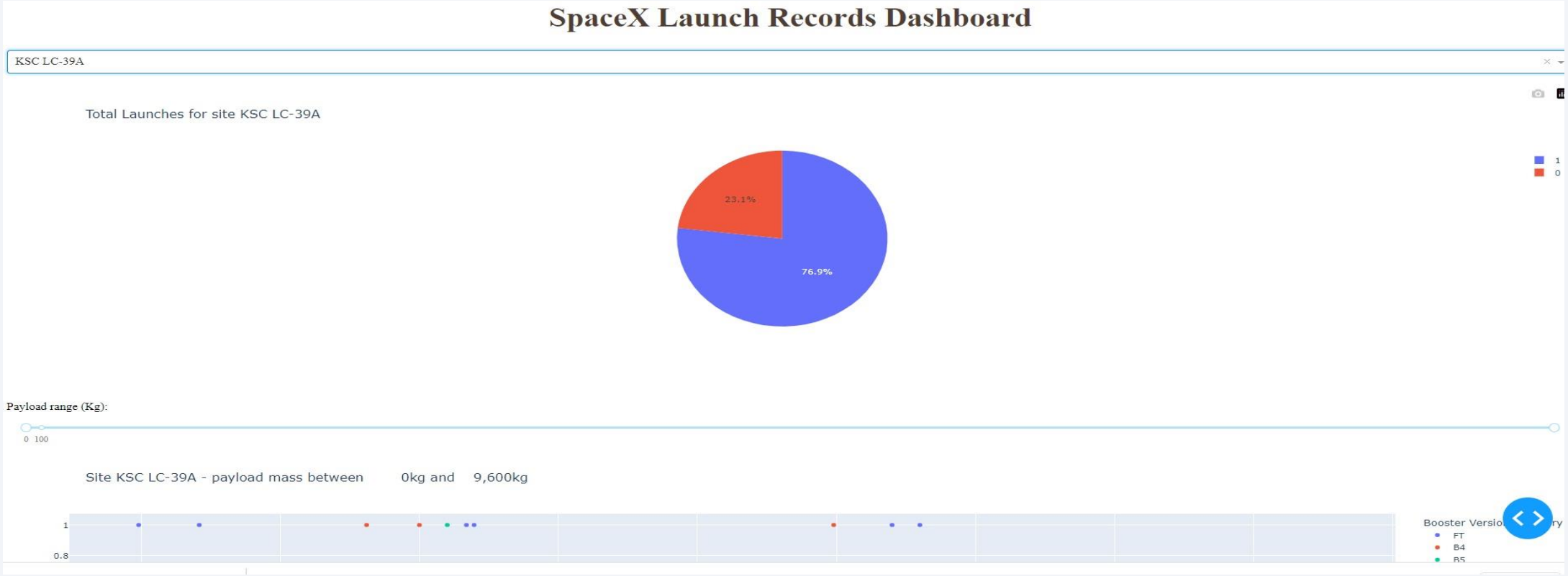

Section 4
Build a Dashboard with Plotly Dash

launch success count for all sites



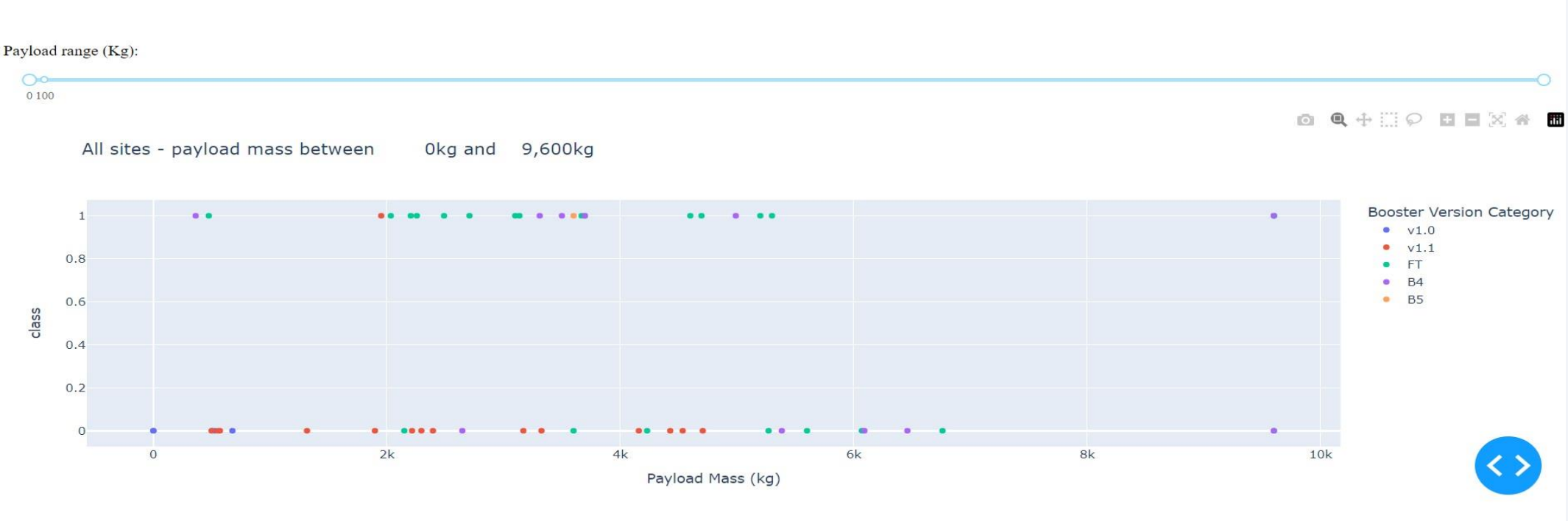
- The place from where launches are done seems to be a very important factor of success of missions

launch site with highest launch success ratio



- 76.9% of launches are successful in KSC LC-93A site.

Launch Outcome scatter plot for all sites



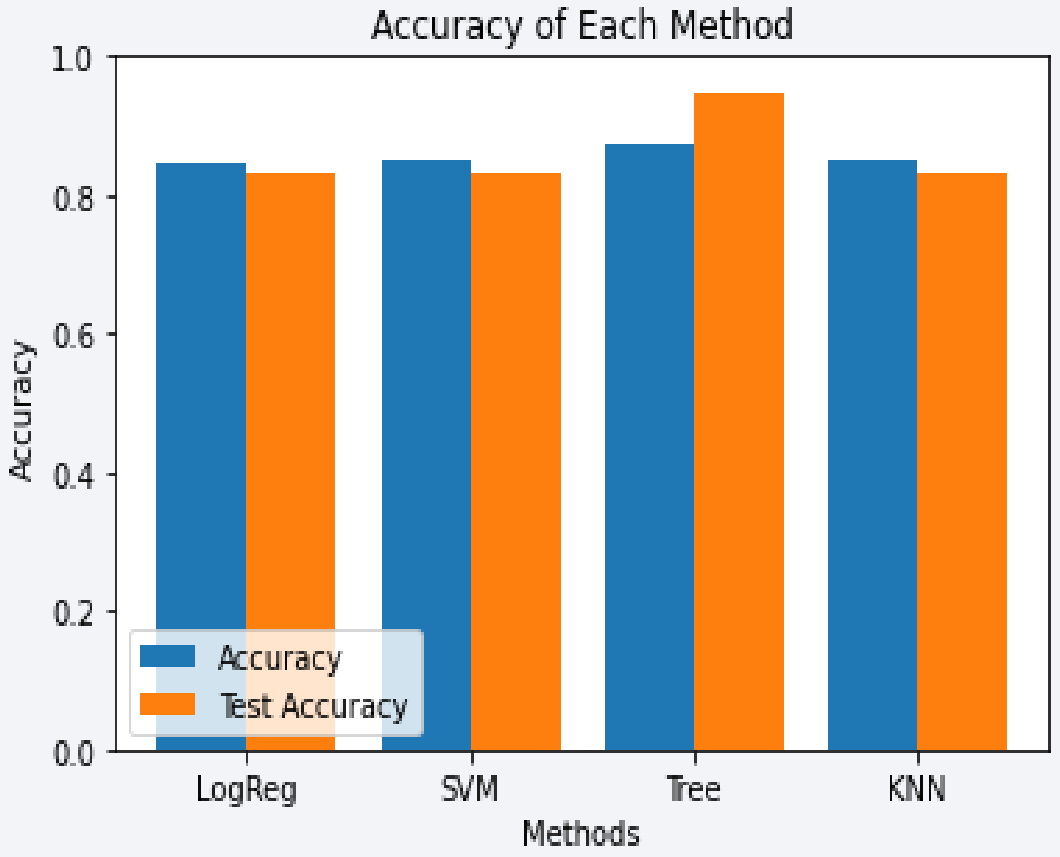
- •Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

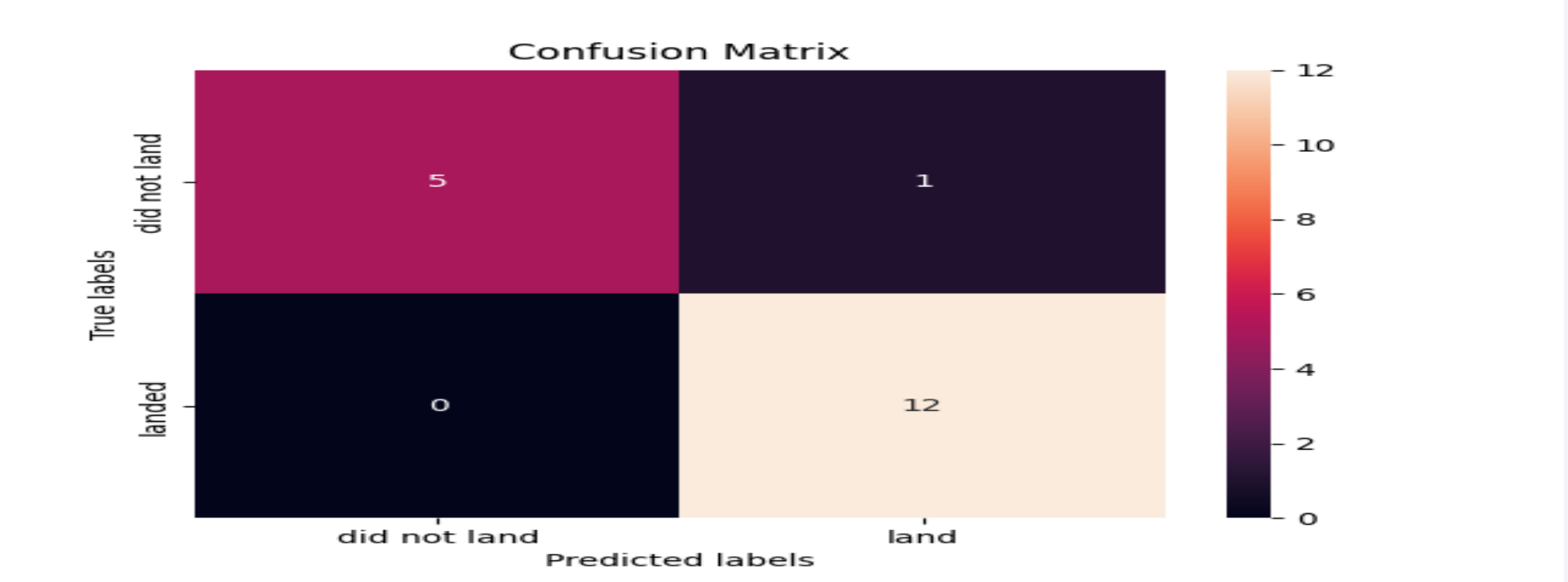
Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%



Confusion Matrix of Decision Tree Classifier



- We showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- Throughout the analysis of various data sources, conclusions were iteratively refined.
- It was determined that the optimal launch site is KSC LC-39A, and that launches involving payloads exceeding
- 7,000kg are associated with a lower level of risk.
- successful landing outcomes appear to improve over time due to advancements in processes and rockets. ,
- it was found that a Decision Tree Classifier could be implemented to forecast successful landings and ultimately increase profits.

Appendix

- Python code ,
- SQL queries,
- charts,
- Notebook outputs,

Thank you!