IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Wenqi Ni>
<Aug 20th 2022>

Outline

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

Executive Summary

Summary of methodologies:

- Data Collection and Wrangling
- Exploratory Data Analysis using Visualization and SQL
- Interactive Visual Analytics using Folium and Plotly Dash
- Predictive Analysis using Classification Models

Summary of all results:

- The maximum payload mass is 15,600 kg.
- Launch success rate kept increasing from 2013 to 2020.
- F9 FT booster version and a payload range of 2,000 kg and 5,000 kg have the largest success rate.
- The highest prediction accuracy is 83% and 3 classification models have excellent performance, including Support Vector Machine, logistic regression, and k nearest neighbors.

Introduction

A growing number of companies are making space travel affordable for everyone, such as Virgin Galactic, Rocket Lab, and SpaceX. The most successful example is SpaceX and one reason that SpaceX can send manned missions to Space is its relatively inexpensive price for rocket launches. SpaceX can reuse the first stage, which can save much money.

“SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, other providers cost upwards of 165 million dollars each”.

Problems:

- to understand the scale of the Falcon 9;
- to explore the factors to determine whether the first stage will land successfully;
- to train a machine learning model to predict if SpaceX will reuse the first stage.

Section 1

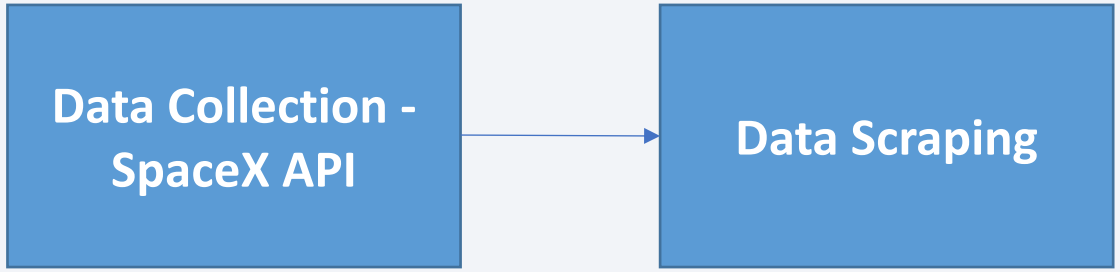
Methodology

Methodology

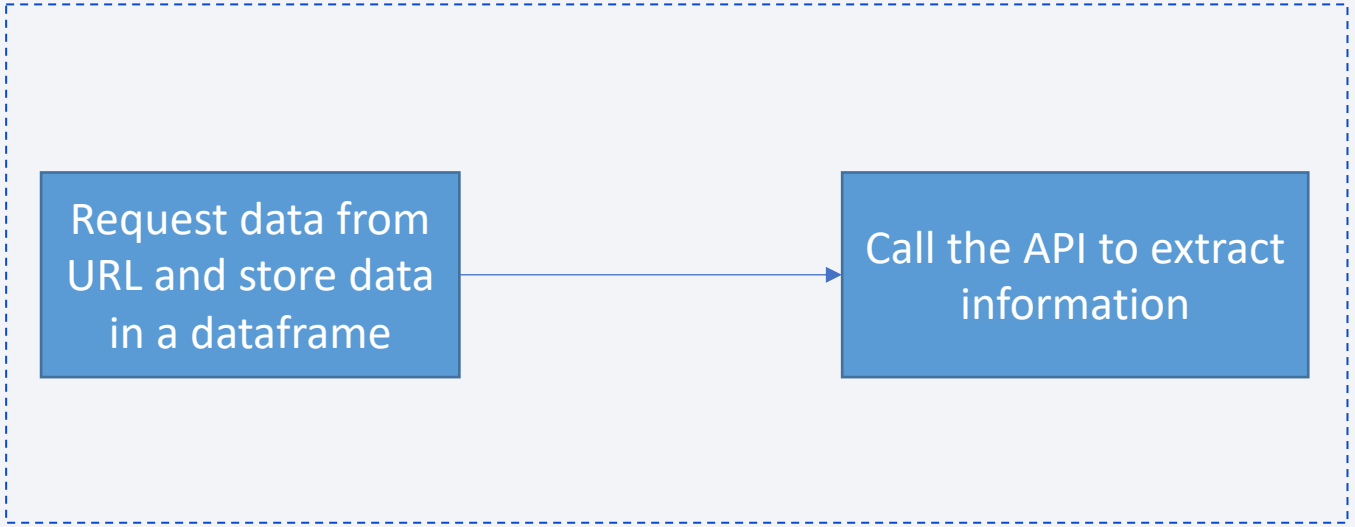
Executive Summary

- Data collection methodology:
  - SpaceX API; Scraping
- Perform data wrangling
  - Deal with missing values; Filter the data for further analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create GridSearchCV objects to build classification models
  - Calculate the accuracy of models on test data and Perform confusion matrix

Data Collection

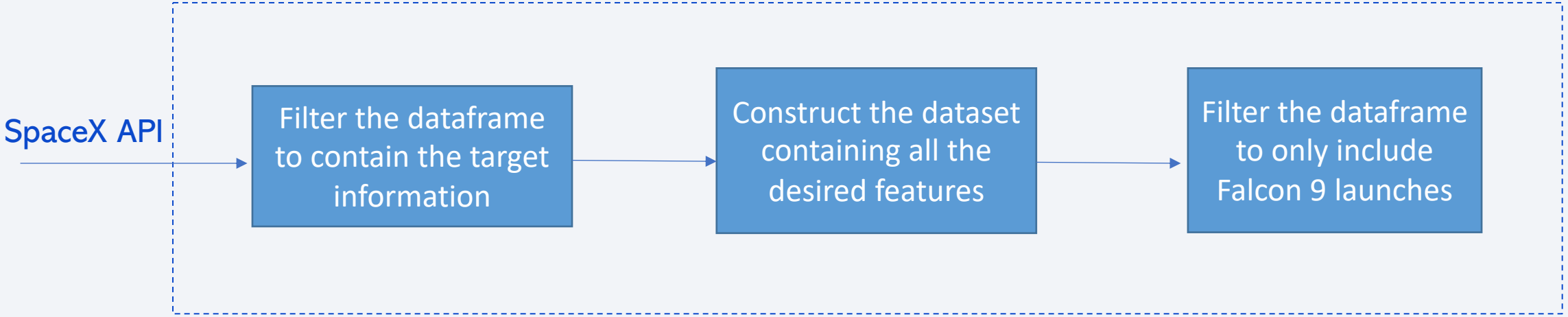


Data Collection – SpaceX API



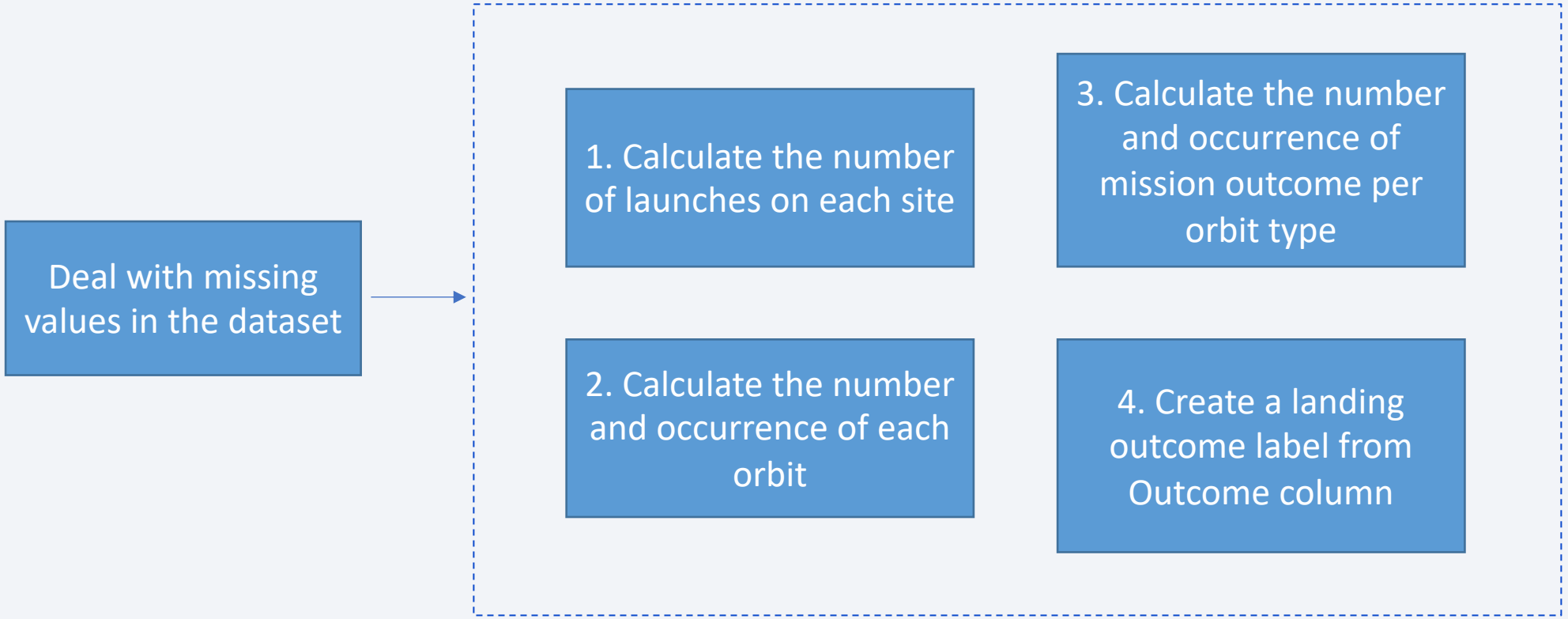
- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



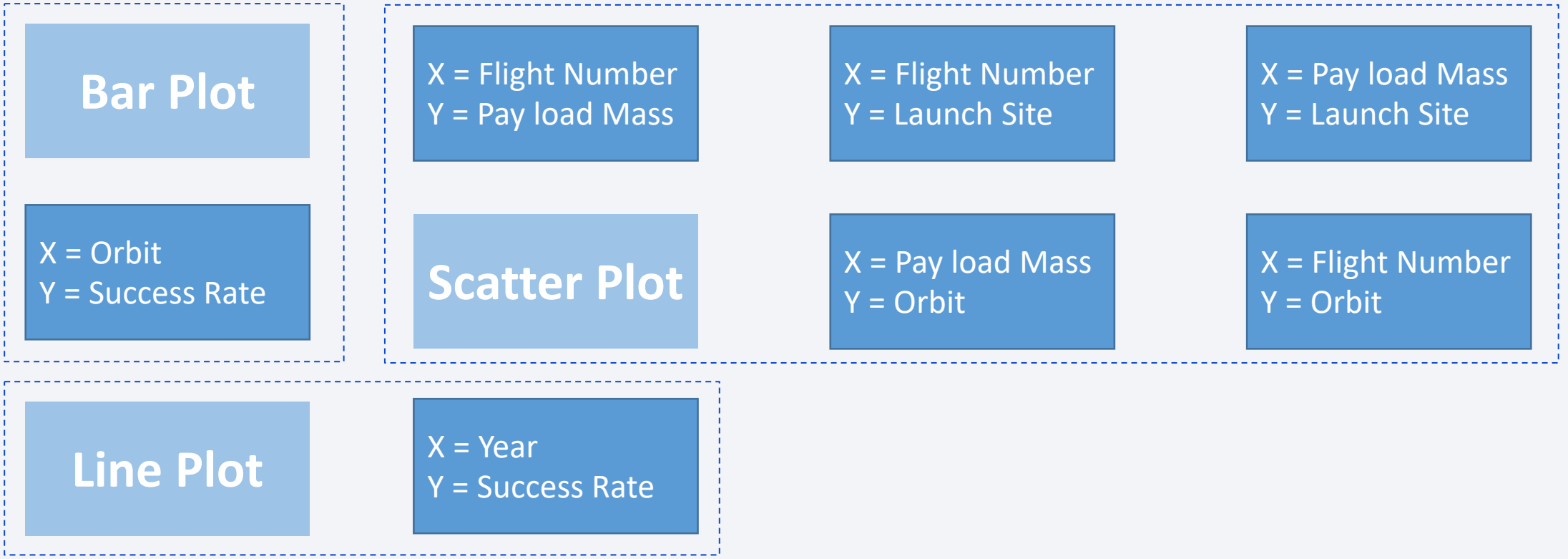
- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Data Wrangling



- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/Data-Wrangling.ipynb

EDA with Data Visualization



- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Display the name 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.
- 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.

- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

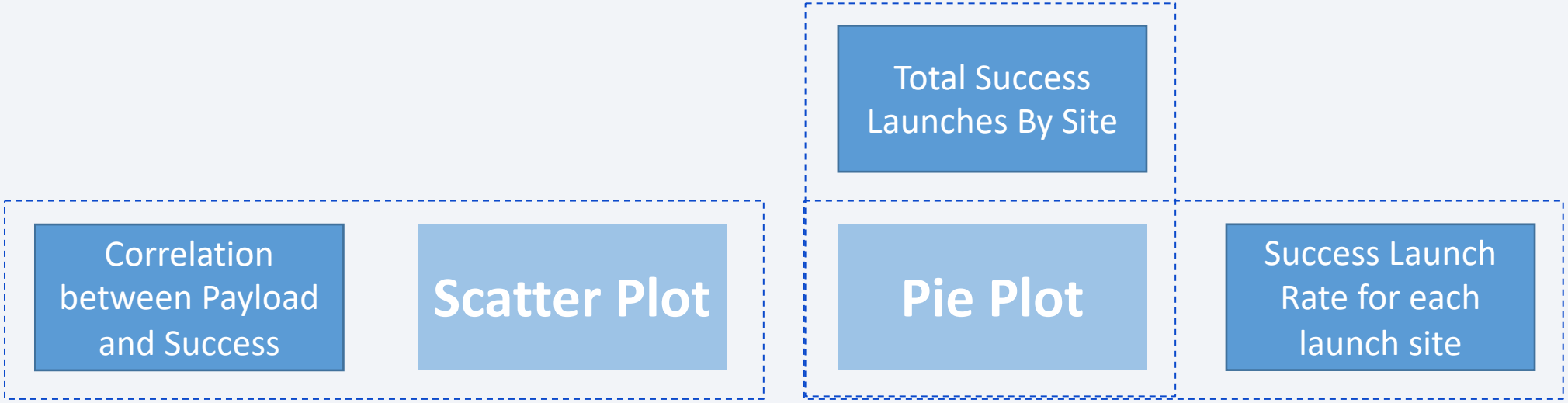
Build an Interactive Map with Folium

- Mark all launch sites on a folium map
- Mark the success/failed launches for each site on the map
- Plot the line and calculate the distances between a launch site to its proximities

[These map objects, such as markers, circles, lines, enable to find some geographical patterns about launch sites.]

- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/Interactive-Map.ipynb

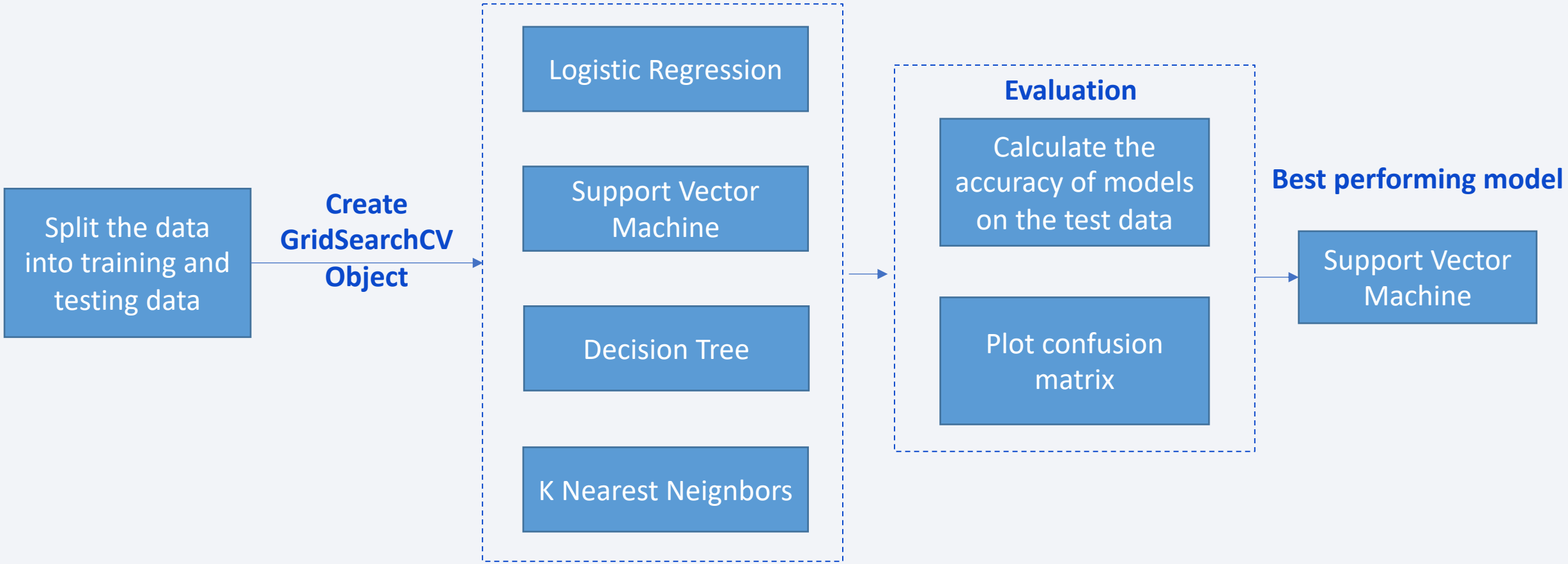
Build a Dashboard with Plotly Dash



[These plots can help to explore the key determinants of Success Launch Rates.]

- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/spacex_dash_app.py

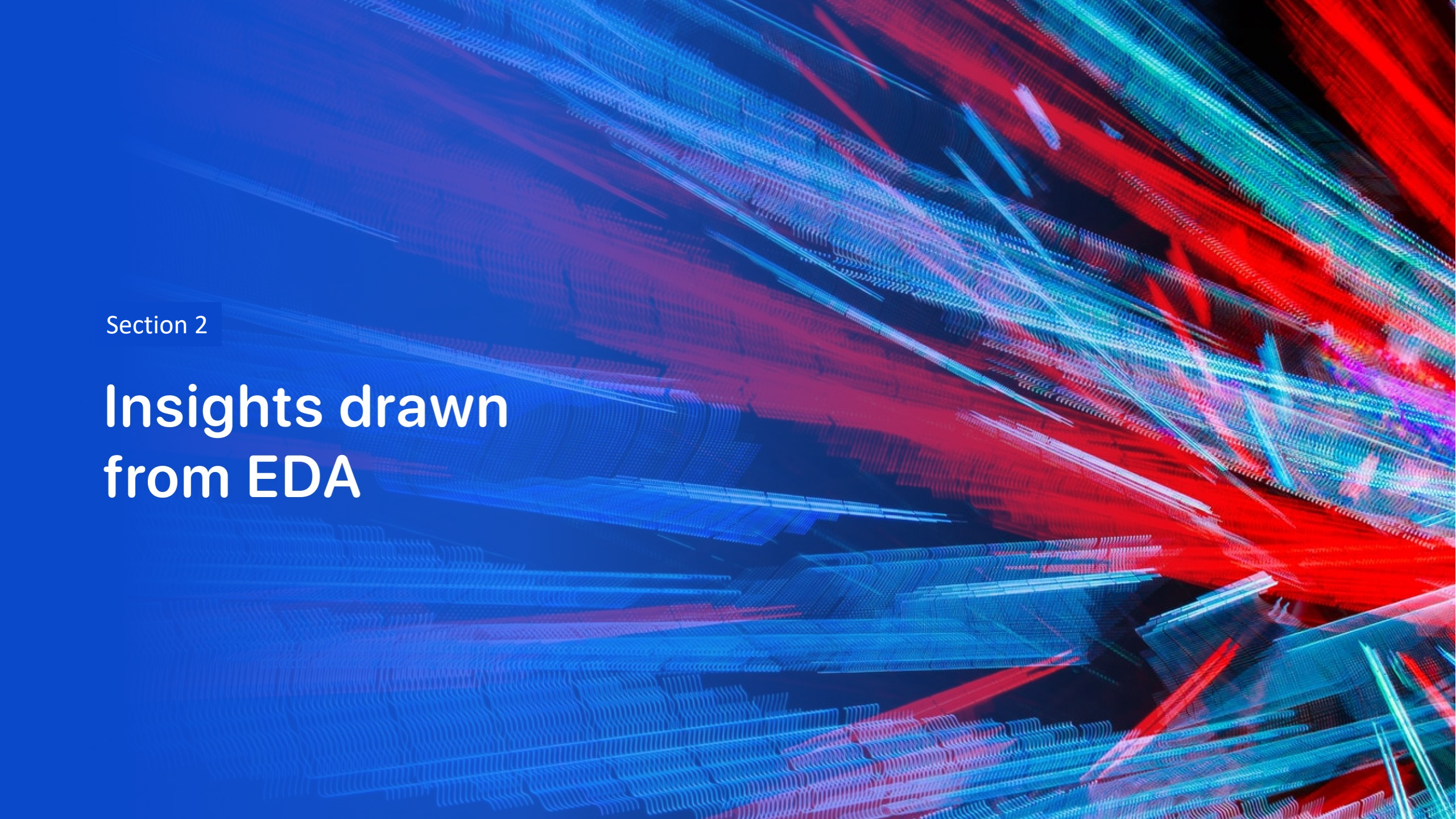
Predictive Analysis (Classification)



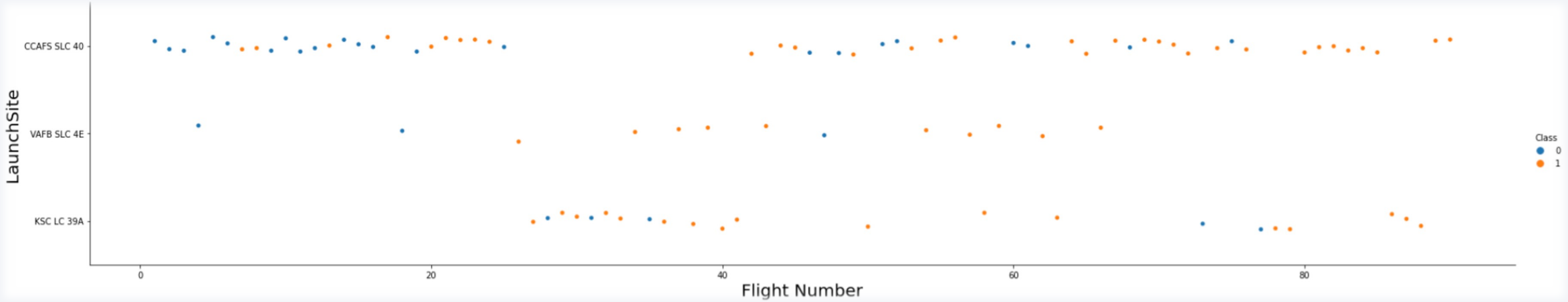
- GitHub URL: https://github.com/niwq15/ADSC-Assignment/blob/master/Predictive-Analysis-(Classification).ipynb

Results

- Exploratory data analysis results – Section 2
- Interactive analytics demo in screenshots – Section 3/4
- Predictive analysis results – Section 5

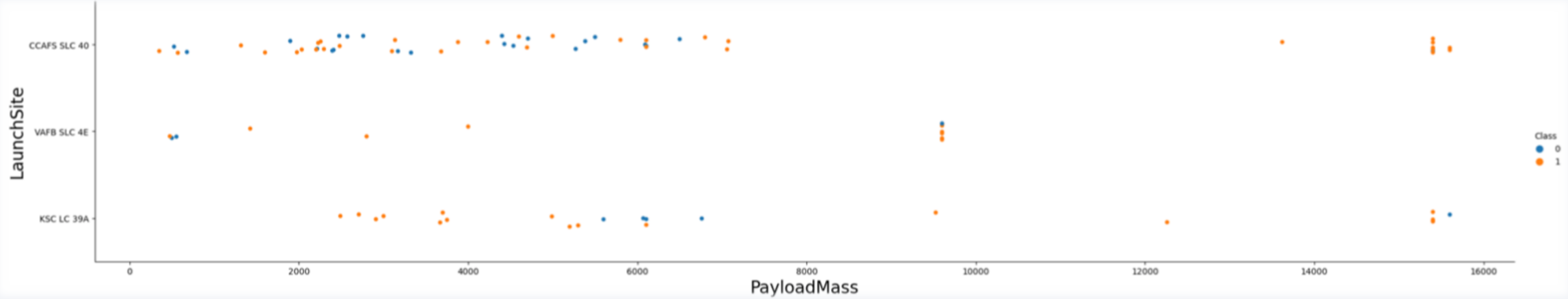

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



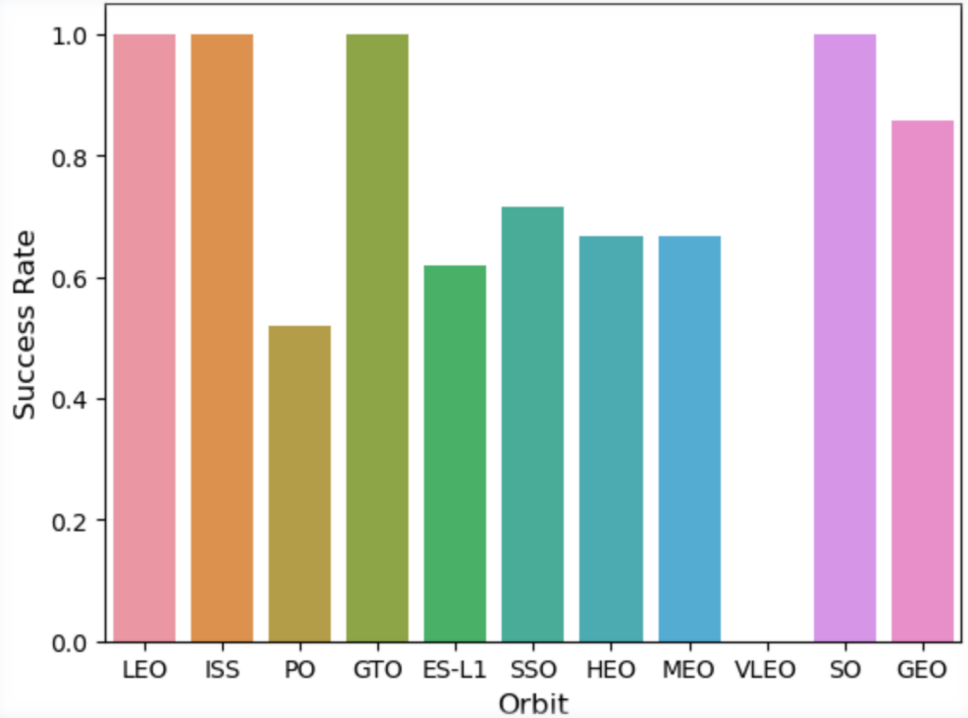
For the VAFB SLC 4E launchsite, all the rockets are launched with flight number no more than 70, while for the KSC LC-39A launchsite, there are no rockets launched with flight number less than 20.

Payload vs. Launch Site



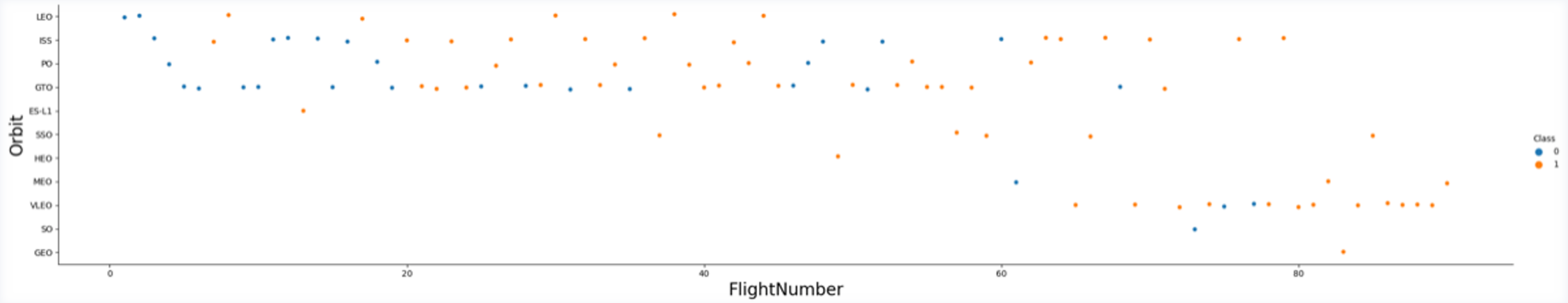
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB SLC 4E launch site there are no rockets launched for heavy payload mass (greater than 10000). For the CCAFS SL-40 and KSC LC-39A launch sites, most rockets are launched for heavy payload mass (less than 8000).

Success Rate vs. Orbit Type



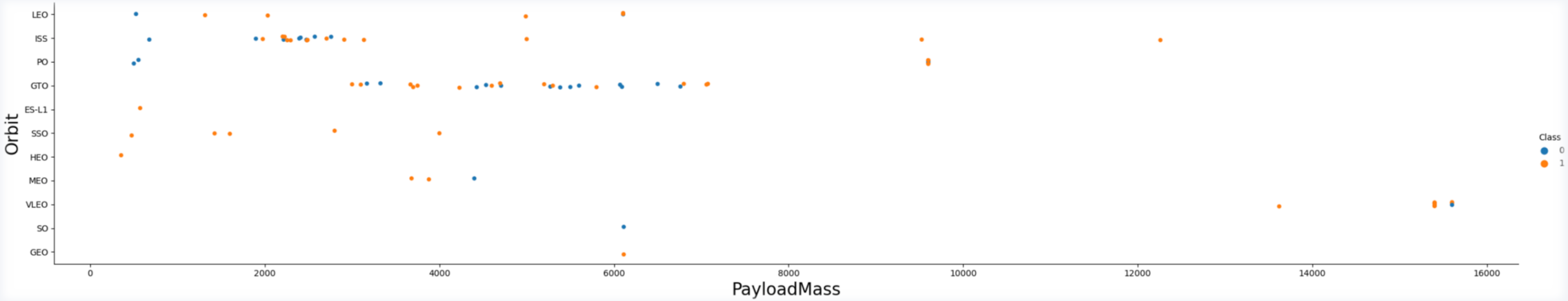
The LEO, ISS, GTO, and SO orbits have the highest success rate equal to 100%. The success rate for the VLEO orbit is 0, which is the lowest. Other orbits such as SSO and ES-L1 have success rates between 50% and 70%.

Flight Number vs. Orbit Type



You 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

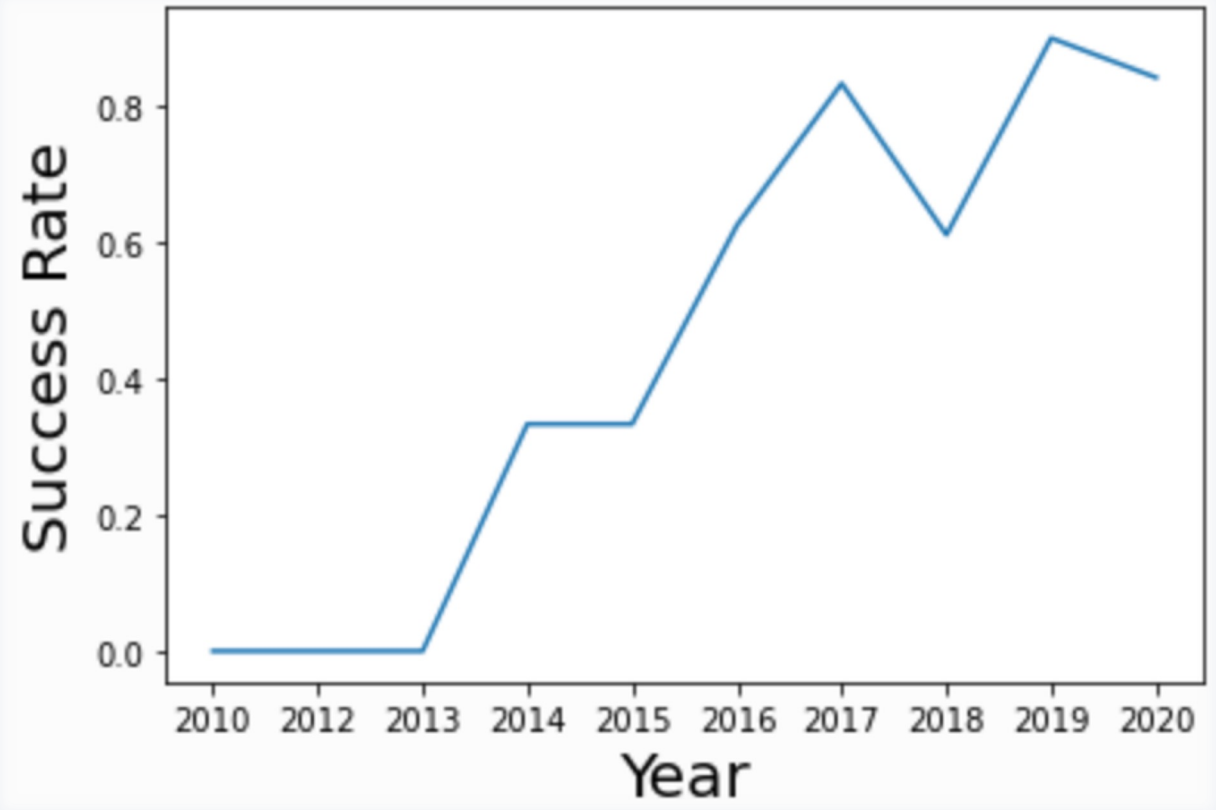


With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

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

Since 2013 the success rate kept increasing till 2020.



All Launch Site Names

There are four unique launch sites in the space mission.

	Launch_Site	count
0	CCAFS LC-40	26
1	CCAFS SLC-40	34
2	KSC LC-39A	25
3	VAFB SLC-4E	16

Launch Site Names Begin with 'CCA'

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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 five records where launch sites begin with the string ‘CCA’.
All the launch sites are CCAFS LC-40. All the mission outcomes are ‘Success’.

Total Payload Mass

After specifying the customer as ‘NASA (CRS)’, we calculate the sum of payload mass.

	Total_Payload_Mass
0	45596

Average Payload Mass by F9 v1.1

After specifying the booster version, ‘F9 v1.1’, we calculate the average payload mass.

	Avg_Payload_Mass
0	2928.4

First Successful Ground Landing Date

After specifying the successful landing outcome, which is on ground pad, we use ‘Limit 1’ to obtain the information for the first first successful landing outcome.

The date of the first successful landing outcome on ground pad:

	Date	Landing _Outcome
0	22-12-2015	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

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

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Use 'WHERE' to set two conditions:

1. Landing Outcome = Success (drone ship)
2. PAYLOAD_MASS__KG_ in (4000, 6000)

Total Number of Successful and Failure Mission Outcomes

Note there are in total 101 cases.

The following results can be obtained through setting

“Mission_Outcome like 'Success%’” or “Mission_Outcome like 'Failure%’”

	Success_Mission_Outcome
0	100

	Failure_Mission_Outcome
0	1

Boosters Carried Maximum Payload

The right table shows the names of the booster which have carried the maximum payload mass.

All the 12 cases have the maximum payload mass, which is 15,600 kg.

	Booster_Version	PAYLOAD_MASS__KG_
0	F9 B5 B1048.4	15600
1	F9 B5 B1049.4	15600
2	F9 B5 B1051.3	15600
3	F9 B5 B1056.4	15600
4	F9 B5 B1048.5	15600
5	F9 B5 B1051.4	15600
6	F9 B5 B1049.5	15600
7	F9 B5 B1060.2	15600
8	F9 B5 B1058.3	15600
9	F9 B5 B1051.6	15600
10	F9 B5 B1060.3	15600
11	F9 B5 B1049.7	15600

2015 Launch Records

The follow table presents the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

	Month	Landing _Outcome	Booster_Version	Launch_Site
0	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

For the period between 2010-06-04 and 2017-03-20, the rank of the count of landing outcomes, including Success, Success (drone ship), and Success (ground pad)), in descending order:

	Landing _Outcome	count
0	Success	20
1	Success (drone ship)	8
2	Success (ground pad)	6


Section 3
Launch Sites
Proximities Analysis

Mark all Launch Sites on a Map

4 launch sites:

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

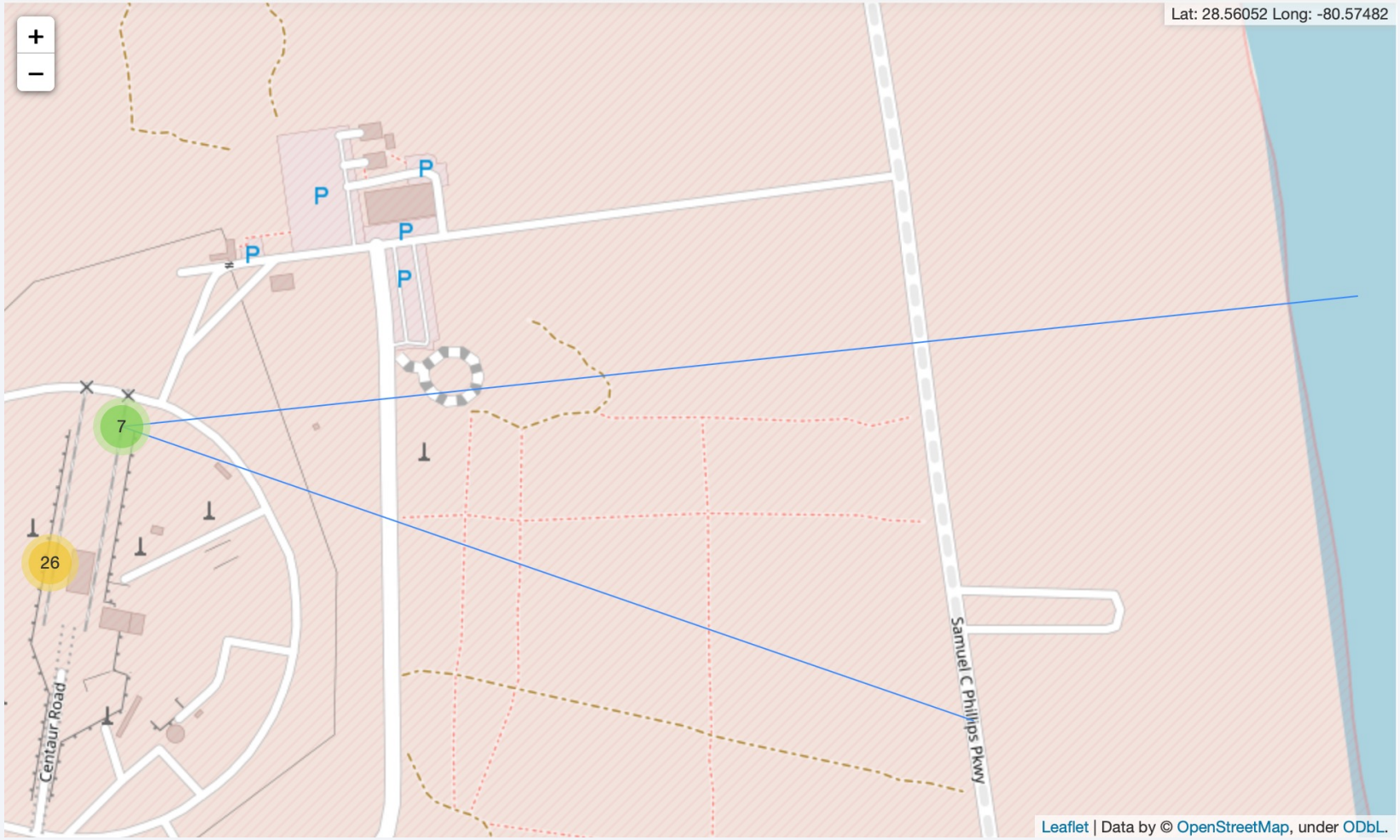
Mark the Success/Failed Launches for Each Site on the Map

Create a MarkerCluster object

Add the marker color based on the ‘class’ value, which represents the outcome of each launch.

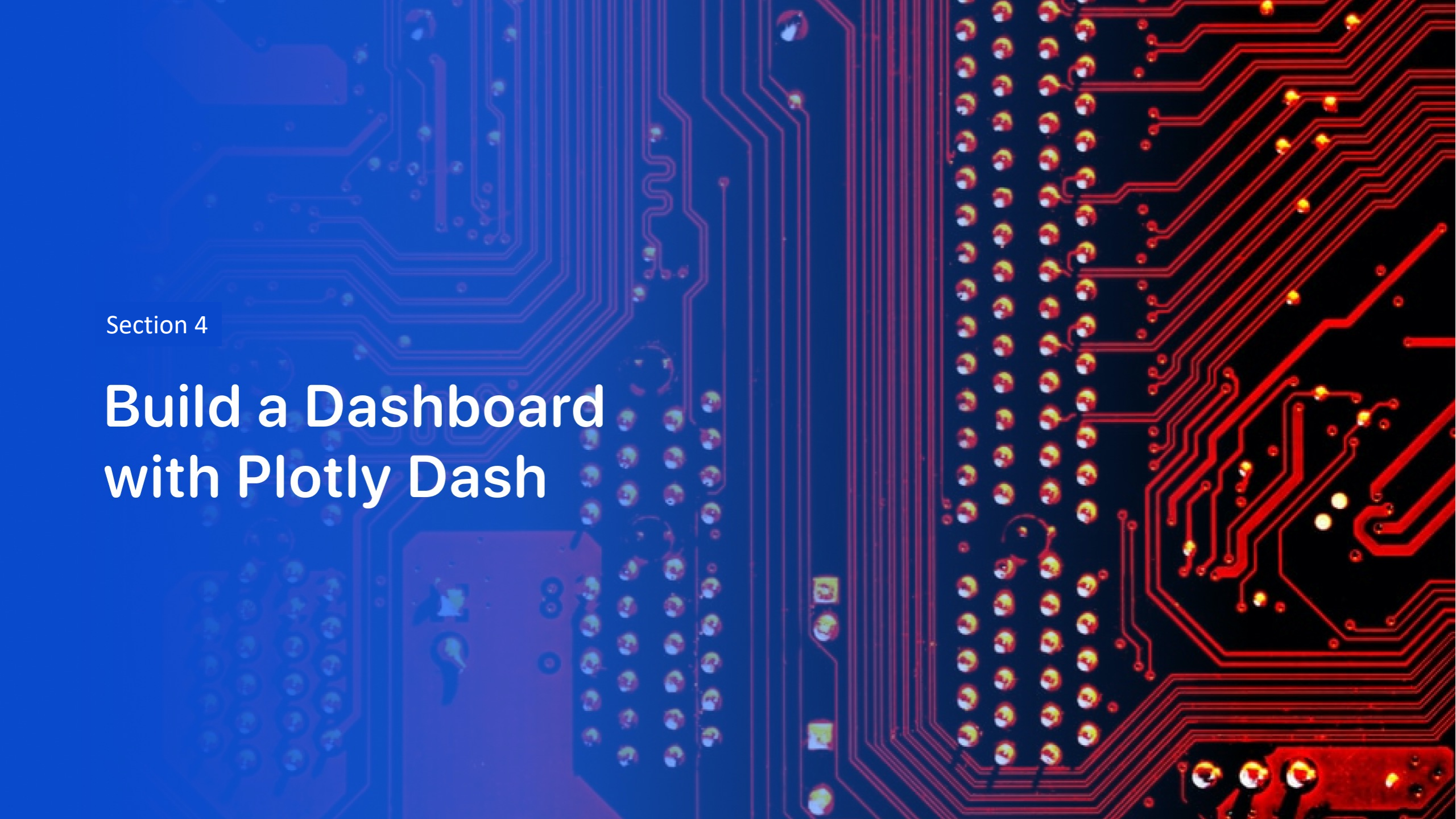
If the first stage did not land successfully, the ’class’ value is equal to 1; if the first stage landed successfully, the ‘class’ value is 0.

Distances between a launch site to its proximities

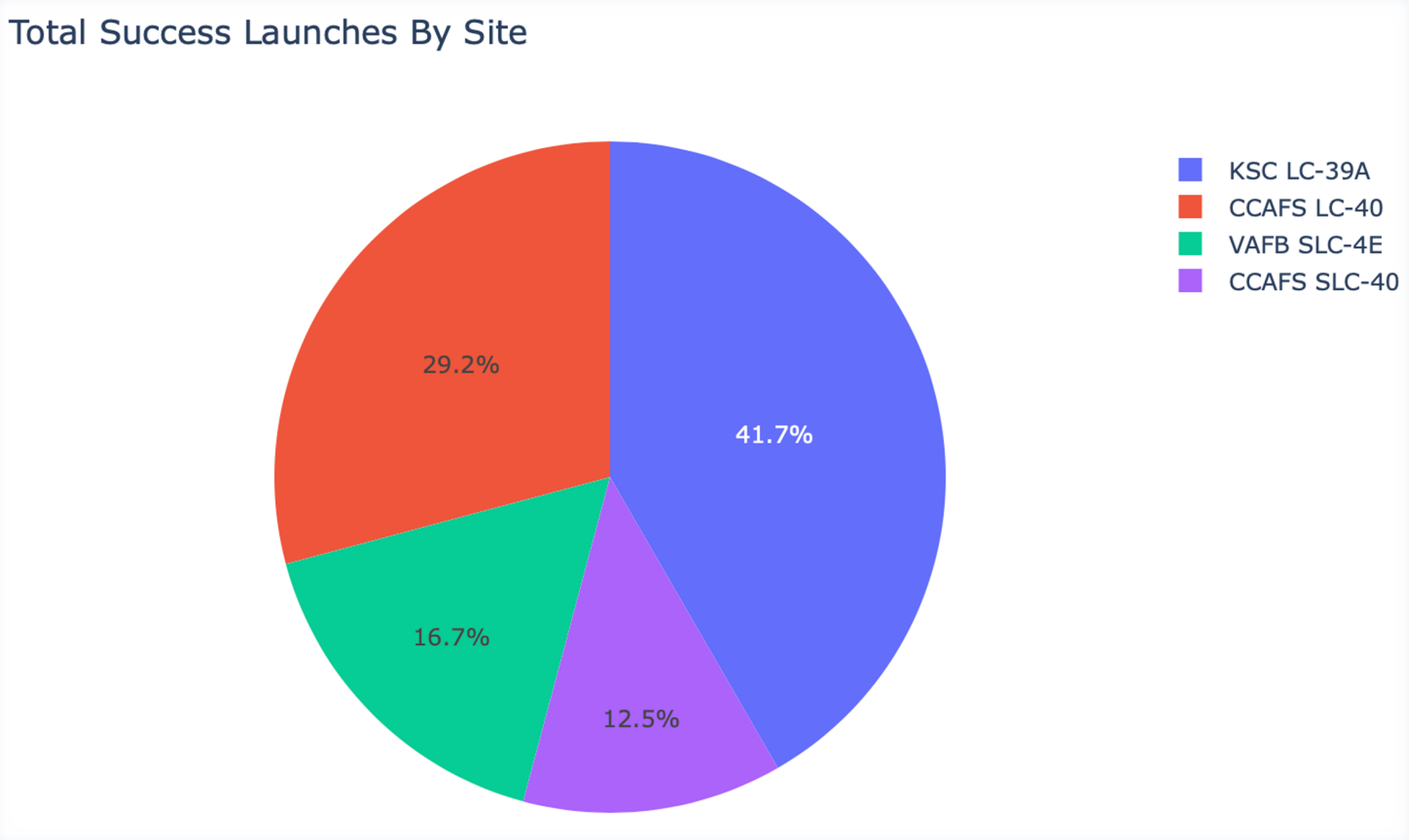


The figure shows the CCAFS SLC-40 launch site and its proximities, including both coastline and railway.

The CCAFS SLC-40 is closer to the railway, Samuel C Phillips Pkwy, compared to the coastline.

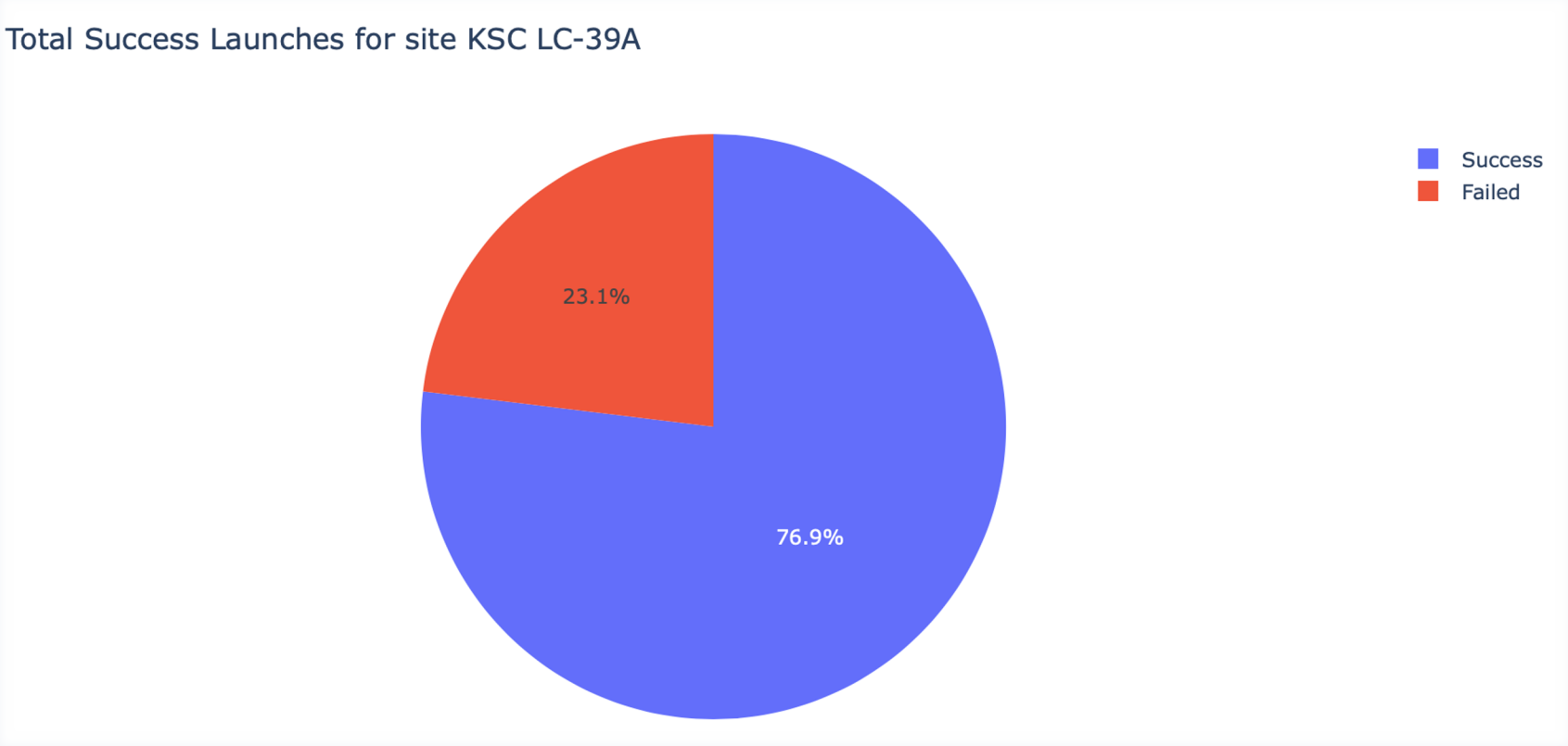

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches By Site



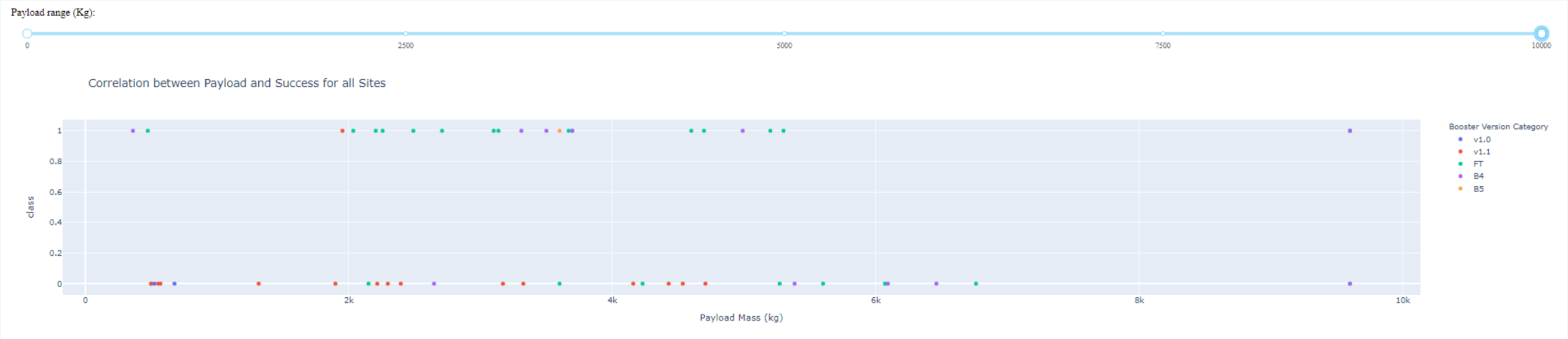
The CCAFS LC-4 launch site accounts for the largest amount of success launches, which is about 42%.

Total Success Launches for Site KSC LC-39A



The KSC LC-39A launch site has the highest launch success ratio, which is about 76.9%.

Correlation between Payload and Success for all Sites

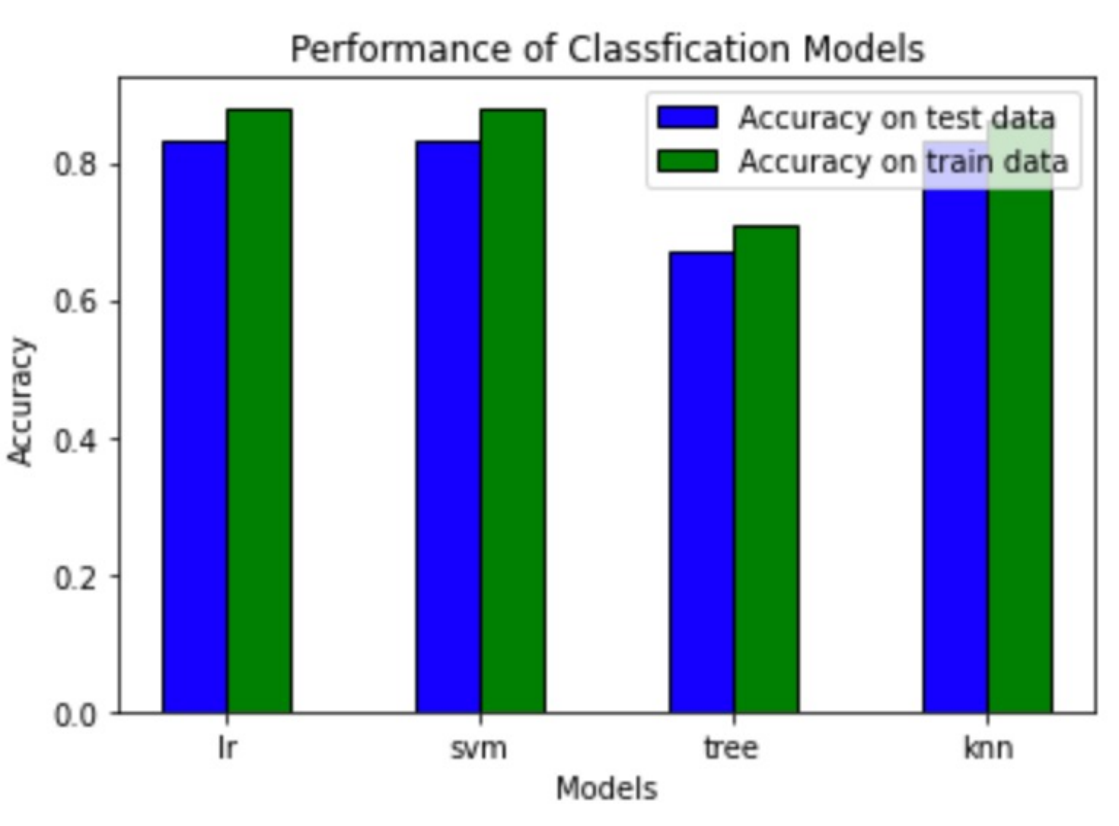


The figure presents that FT booster version and a payload range of 2,000 kg and 5,000 kg have the largest success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



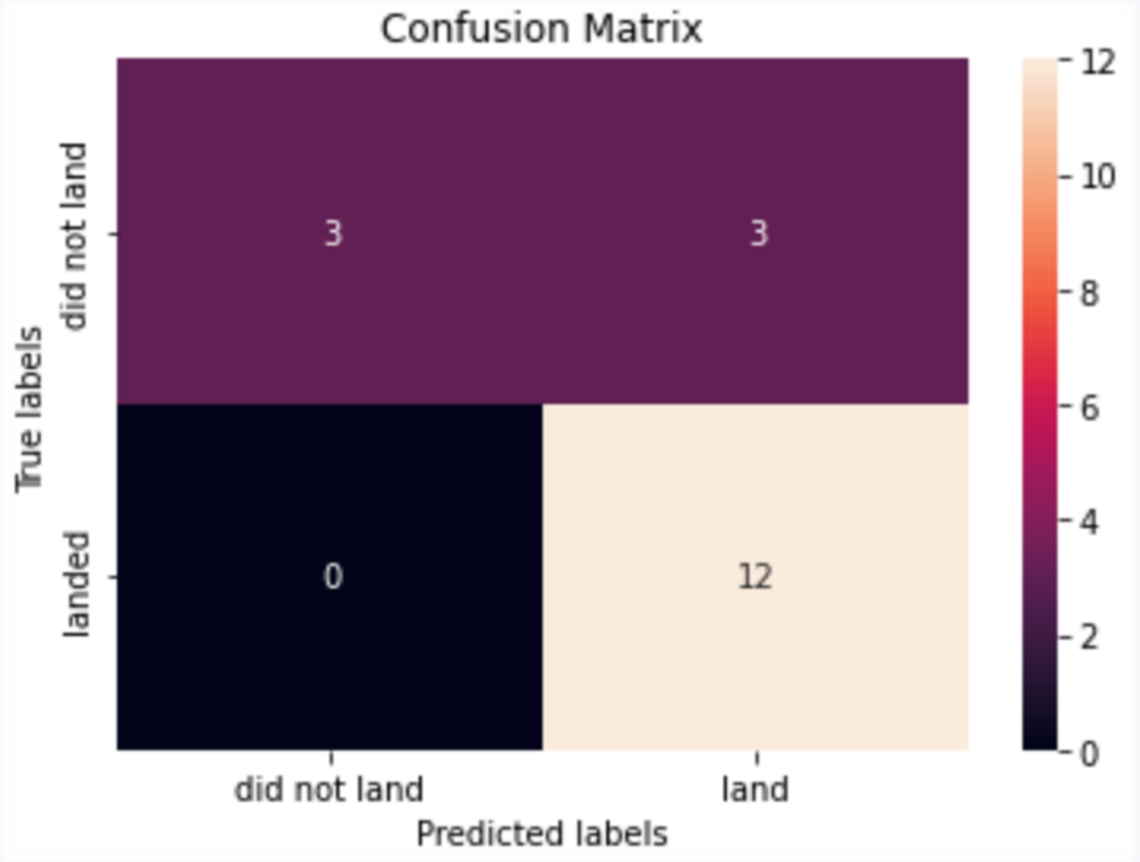
According to the accuracy of models on the test data, the highest accuracy on the test data is 83%, which is achieved by Support Vector Machine (svm), logistic regression (lr), and k nearest neighbors (knn).

In the dataset, the decision tree (tree) classification model has inferior performance on both the train and the test data.

Confusion Matrix

The right table is the confusion matrix for the Support Vector Machine model.

It can distinguish between the different classes and the major problem is false positives.



Conclusions

- During the period of 2010 and 2020, LEO, ISS, GTO, and SO orbits had the highest success rate equal to 100% and Launch success rate kept increasing from 2013 to 2020.
- The highest accuracy on the test data is 83%, which can be achieved by Support Vector Machine, logistic regression, and k nearest neighbors.
- The CCAFS LC-4 launch site accounts for the largest amount of success launches, which is about 42%.
- The KSC LC-39A launch site has the highest launch success ratio of 76.9%.
- F9 FT booster version and a payload range of 2,000 kg and 5,000 kg have the largest success rate.

Appendix

- GitHub URL: https://github.com/niwq15/ADSC-Assignment/tree/master

[The link above includes all the relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that have been created during this project]

Thank you!