Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

Summary of methodologies

- Data collection using the space X API and web scrapping using Beautiful soup library from python.
- Data wrangling using pandas to clean the data, drop non falcon9 booster versions and remove missing numbers.
- Analyze the data using pandas library and sql database
- Use MatPlotlib, seaborne, pandas to visualize and discover relationships between features
- Use folium and potly to design interactive visualization and dash board.

Summary of all results

The model performed well on the test data with an accuracy score 83% predicting the whether the space mission will be a success .

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.
- Therefore we can use models to predict whether a mission launch will be successful or not.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Date was collected using the get request to the Space X APL and web scrapping using beautiful soup
- Perform data wrangling
  - Data was modified by removing missing values and replacing them with mean and then remained with Falcon9 as the booster version only.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - One hot encoding was performed to make categorical data numerical and a class was assigned to the data frame.

Data Collection

Data set were collected using two data collection methods

- SpaceX API

- Web-Scraping

Data Collection – SpaceX API

1. Spacex URL
2. Response using get request function
3. Print response content
4. Static response object to get json documents

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

1. Data was scrapped from a snapshot of the List of Falcon9 and Falcon Heavy launches Wikipage updated on 9th June 2021 as a static url
2. Use requests. Get() method with the provided static url
3. Create a beautiful soup object from the html response

Print the title

4. Extract columns and variables from table headers

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/jupyter-labs-webscraping%20(1).ipynb

Data Wrangling

- Read csv fail into data frame
- Find the missing values
- Find the data types
- Calculate the number of launches from the launch sites
- Calculate the number of launch occurrences in each orbit
- Calculate the landing outcomes whether they are success or failures
- Calculate the average space mission launches

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

EDA with Data Visualization

The following charts were plotted

- How the Flight Number (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome using a cat plot and scatter plot. As the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; the more massive the payload, the less likely the first stage will return.
- Visually check if there are any relationship between success rate and orbit type using a bar chart.
- For each orbit, to see if there is any relationship between Flight Number and Orbit type using a scatter plot
- Payload vs. Orbit scatter point charts to reveal the relationship between Payload and Orbit type.
- A line chart to show a trend of success rate in different years.

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/edadataviz.ipynb

EDA with SQL

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

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Circles were used to add a highlighted circle area with a text label on a launch site
- Markers were used to locate a coordinate on the map
- Marker Clusters were used to store markers colors based on the class value.
- Marker colors were used to separate success and failed launches
- Mouse positions were added to get a coordinate for a mouse over a point on the map.
- Polyline to show the distance between the launch site and the proximity for example a railway

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- An app layout
- A title
- A dropdown to have all launch site
- A pie chart to show the success count rate per launch site and all sites
- A Slider to show the pay load mass in Kilograms ranges
- A scatter chart to show the correlation between payload and launch success

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/spacex_dash_app.py

Predictive Analysis (Classification)

- Standardized the data using Standard Scalar and transformed
- Split the data into training and testing
- The model were improved using Grid Search Cross validation
- Use the Hyper parameter for SVM, Classification Trees and Logistic Regression
- Use a confusion matrix to see the performance of the model predictions
- Calculate the accuracy using the score method on the test data on each model.
- Get the model that performs best using the test data

https://github.com/artpet19/SpaceX-Falcon9-Launch-and-Landing-Predictor/blob/3f2eea93f7dd5b9b04c4915e769de90be3edc680/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

Exploratory data analysis results

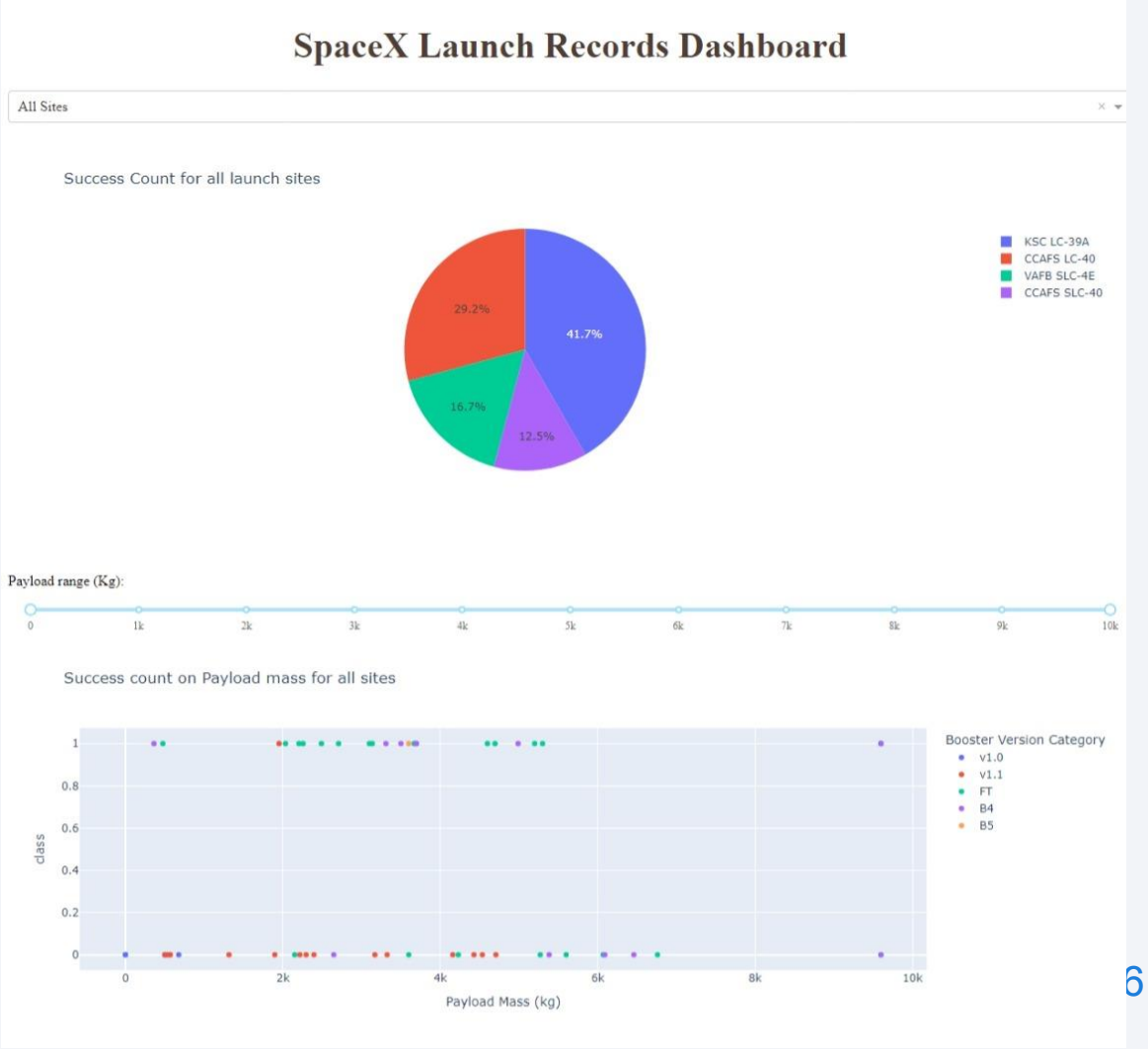
- There is a relationship among Flight Number and Payload variables and how they affect the launch outcome
- As the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; the more massive the payload, the less likely the first stage will return.

Interactive analytics demo in screenshots

- An interactive dashboard was created that represents relationships between success count rate, launch site and pay load mass.

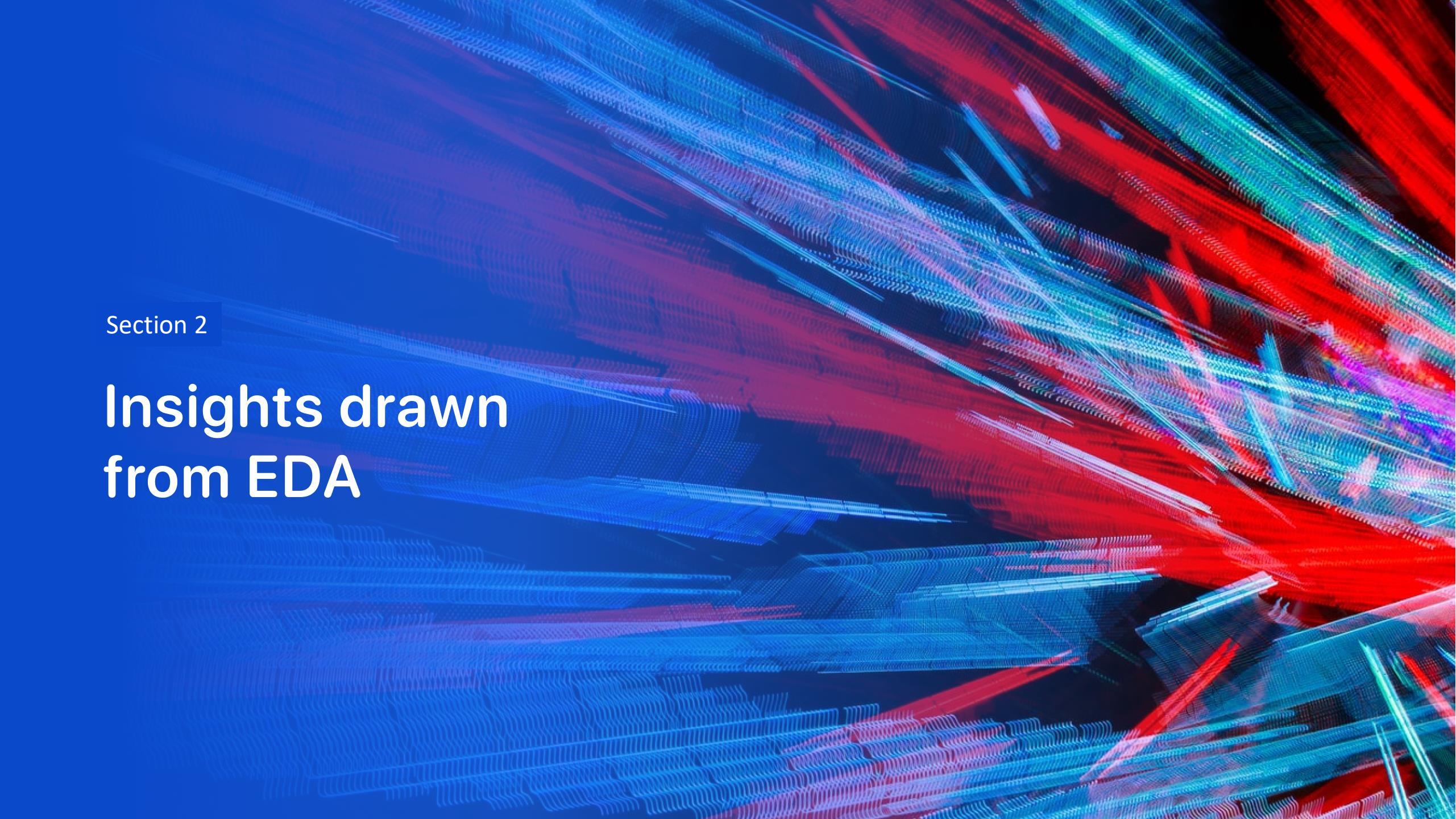
Predictive analysis results

- All the models performed well on the test data with an accuracy of 83%

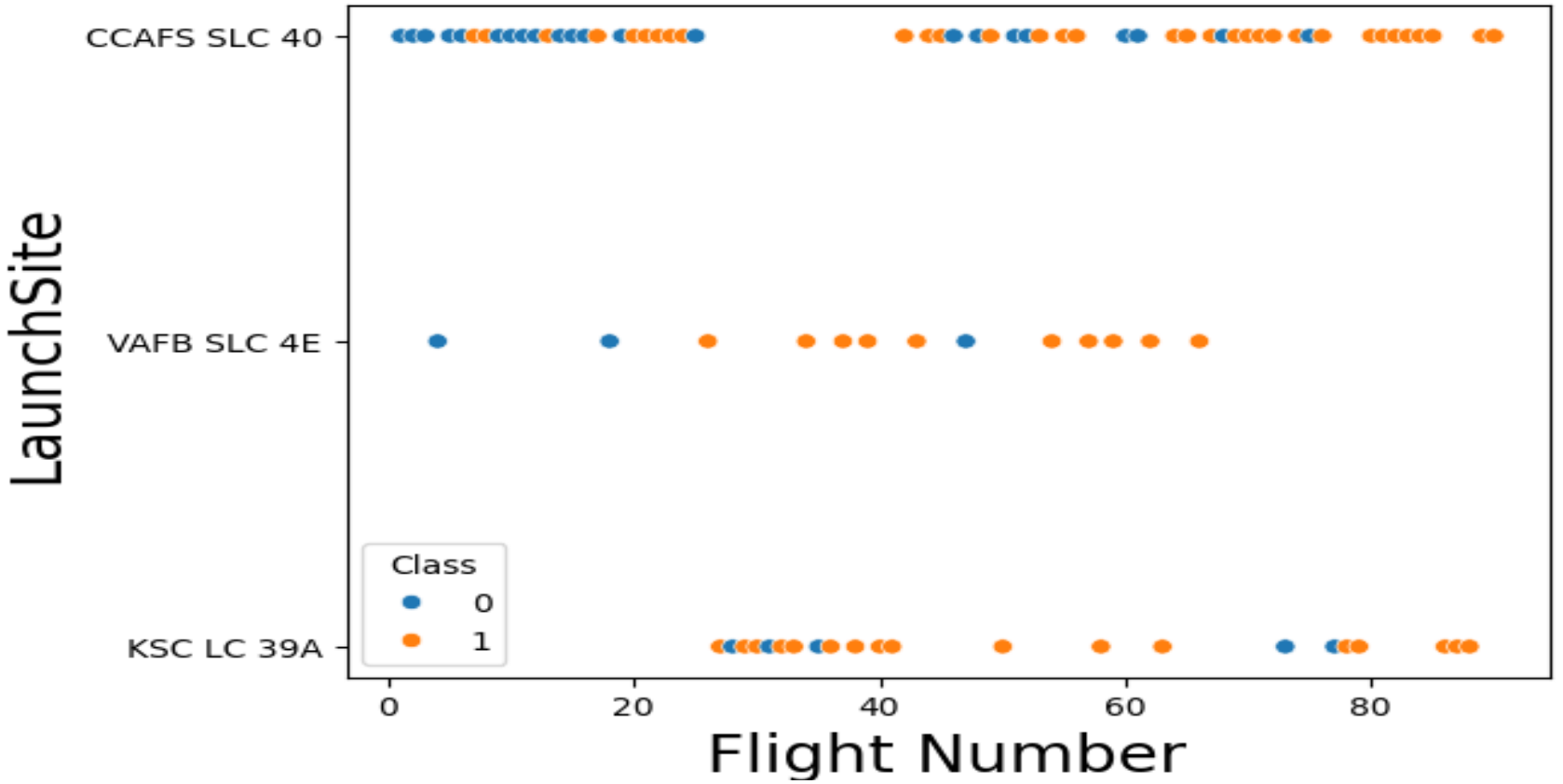


Section 2

Insights drawn from EDA



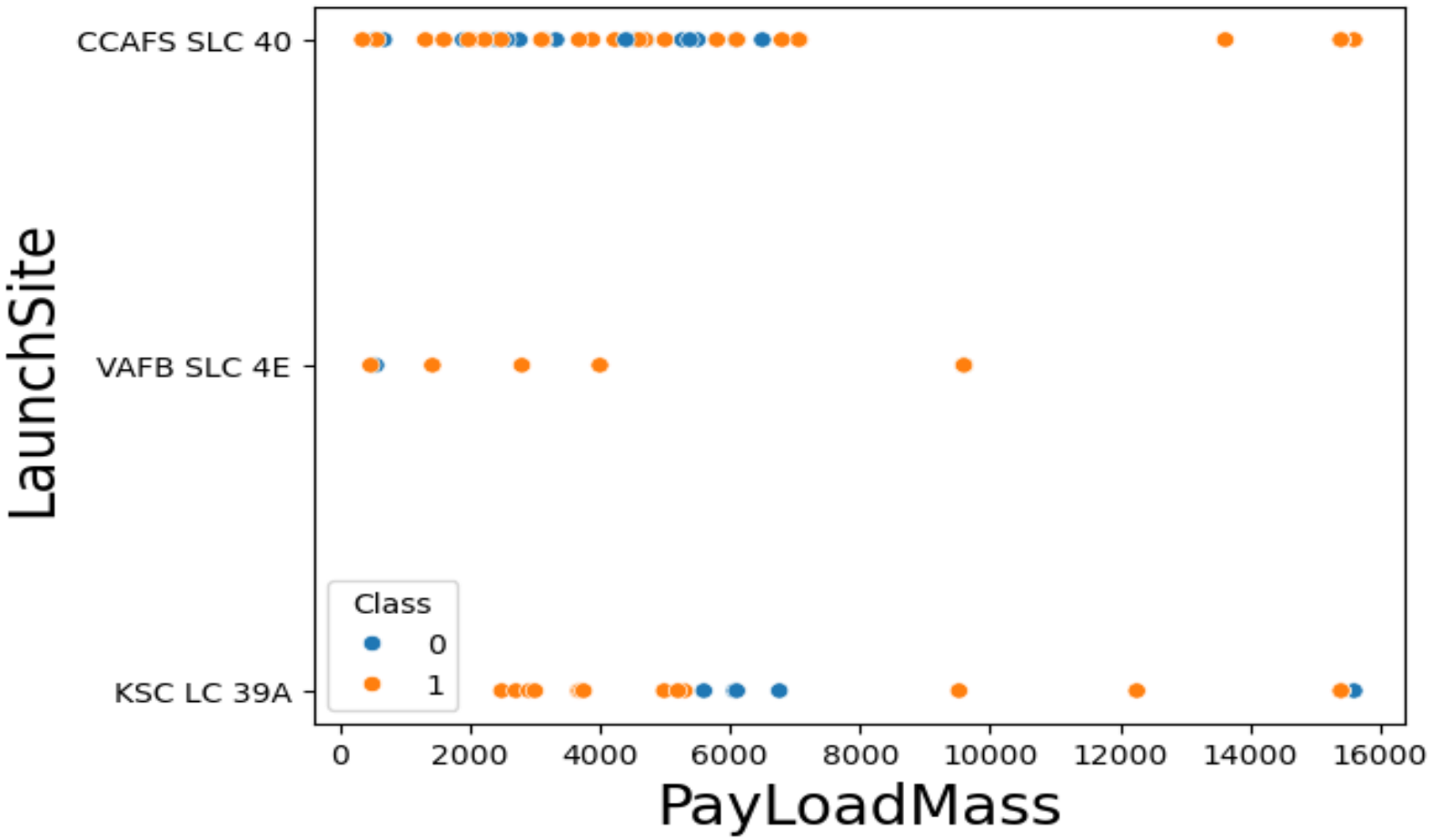
Flight Number vs. Launch Site



Explanation:

- There is no correlation between the success rate with the number of flight at CCAFS SLC 40 and KSC LC 39A launching sites.
- For Launch site VAFB SLC 4E, the success rate increased with the number of flights.

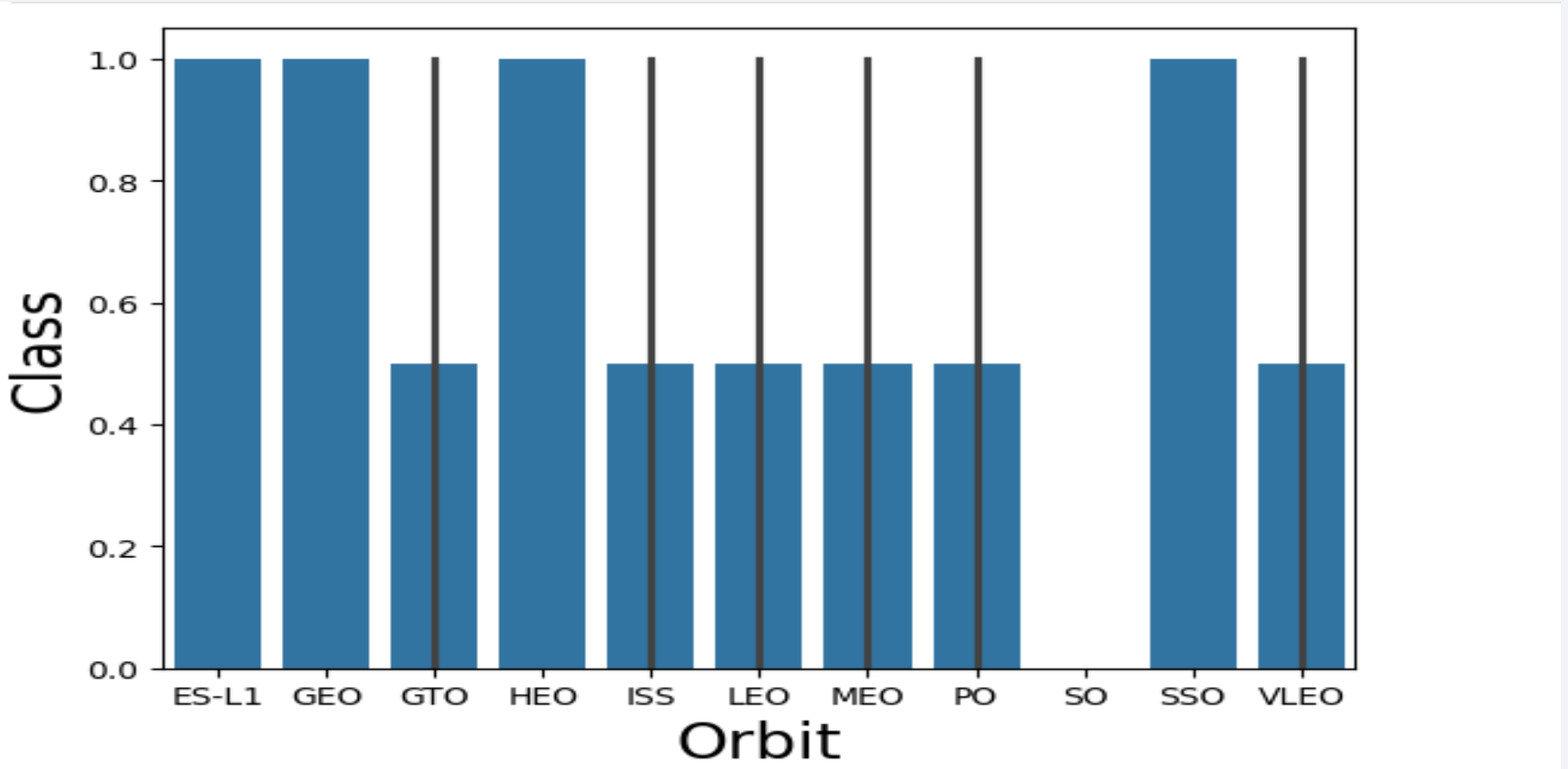
Payload vs. Launch Site



Explanation:

- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

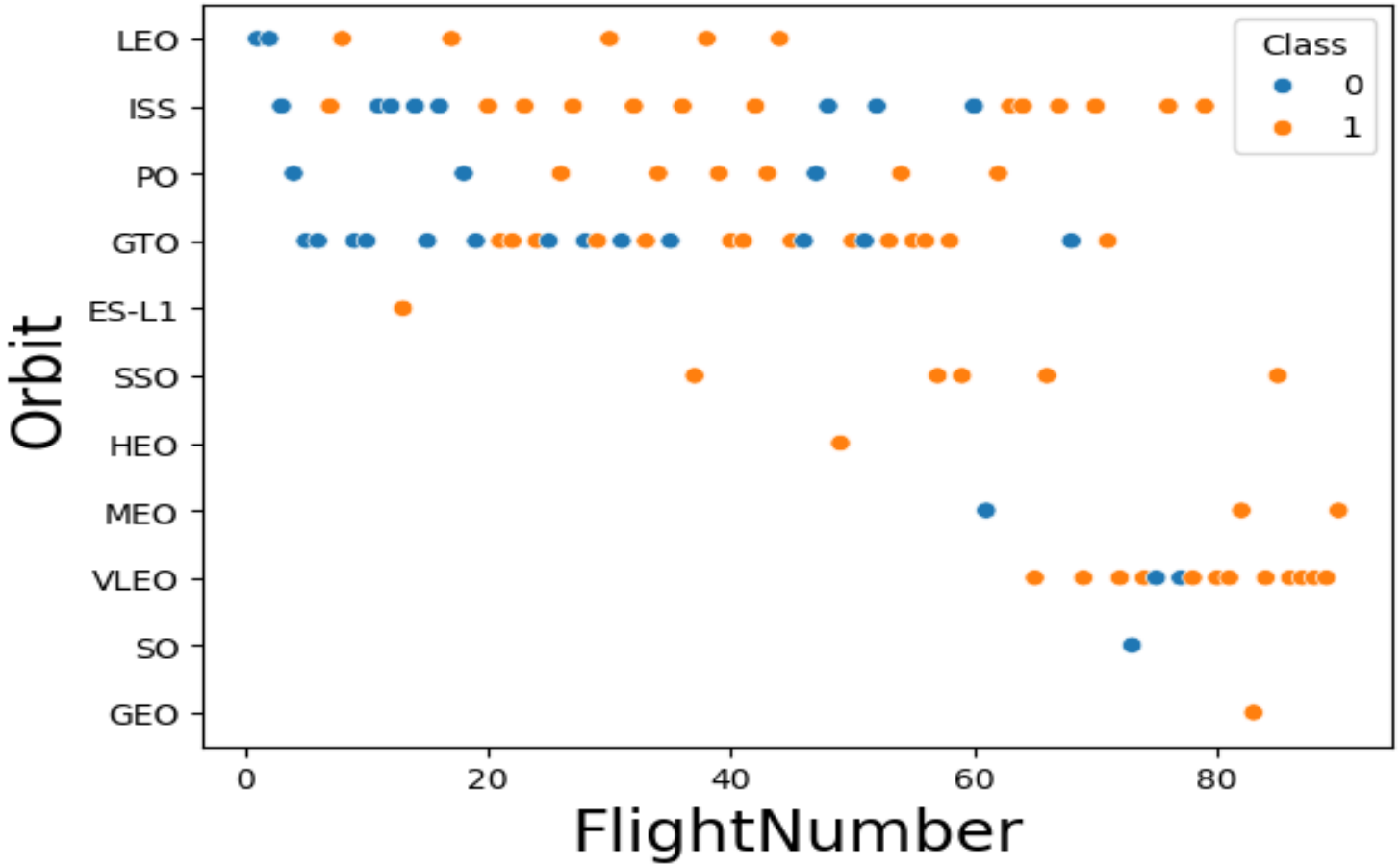
Success Rate vs. Orbit Type



Explanation:

- Falcons launched in the ES-L1, GEO, HEO, SSO Orbits have a high success rate as compared to those launched in the GTO, ISS, LEO, MEO, PO,VLEO which have a low success rate

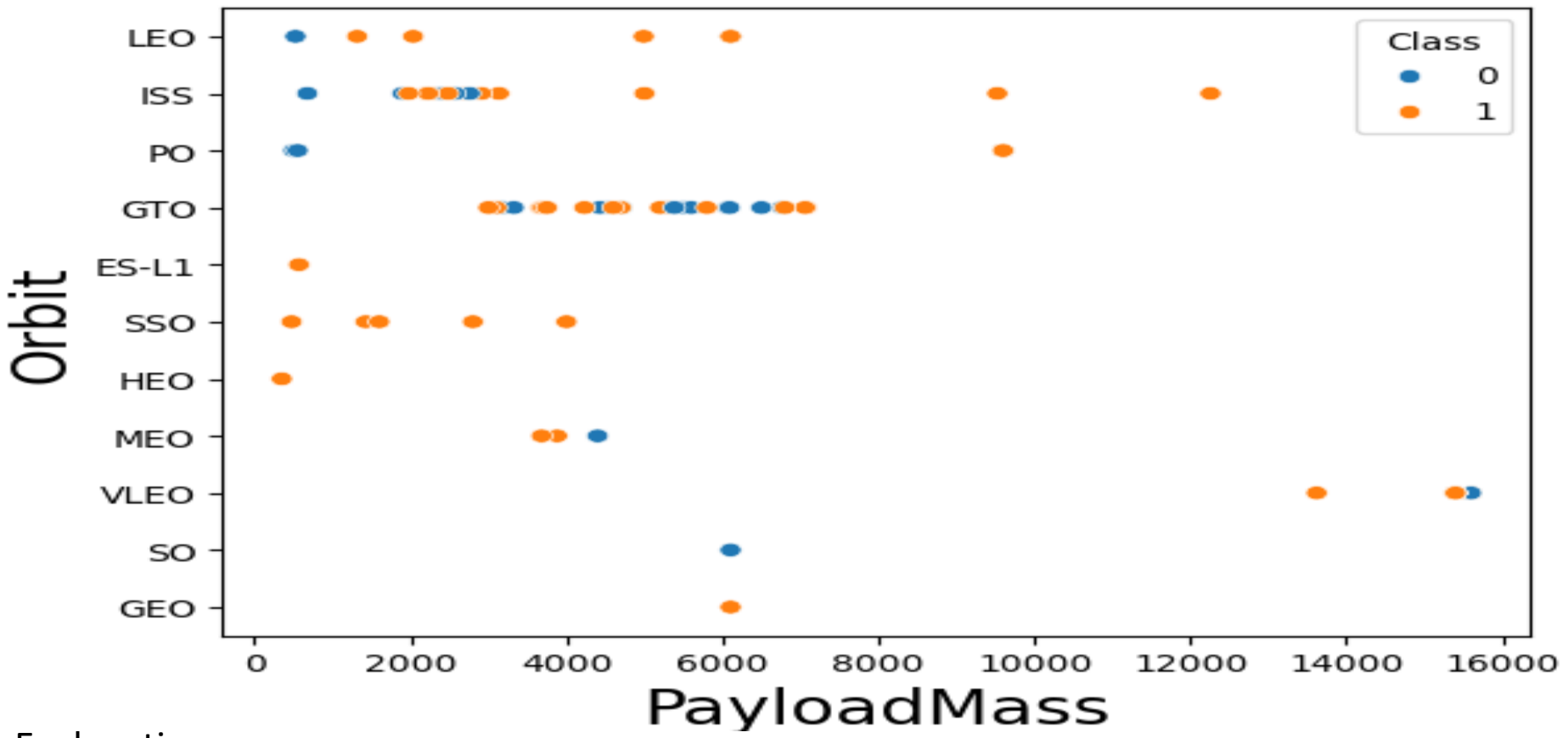
Flight Number vs. Orbit Type



Explanation:

- 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.

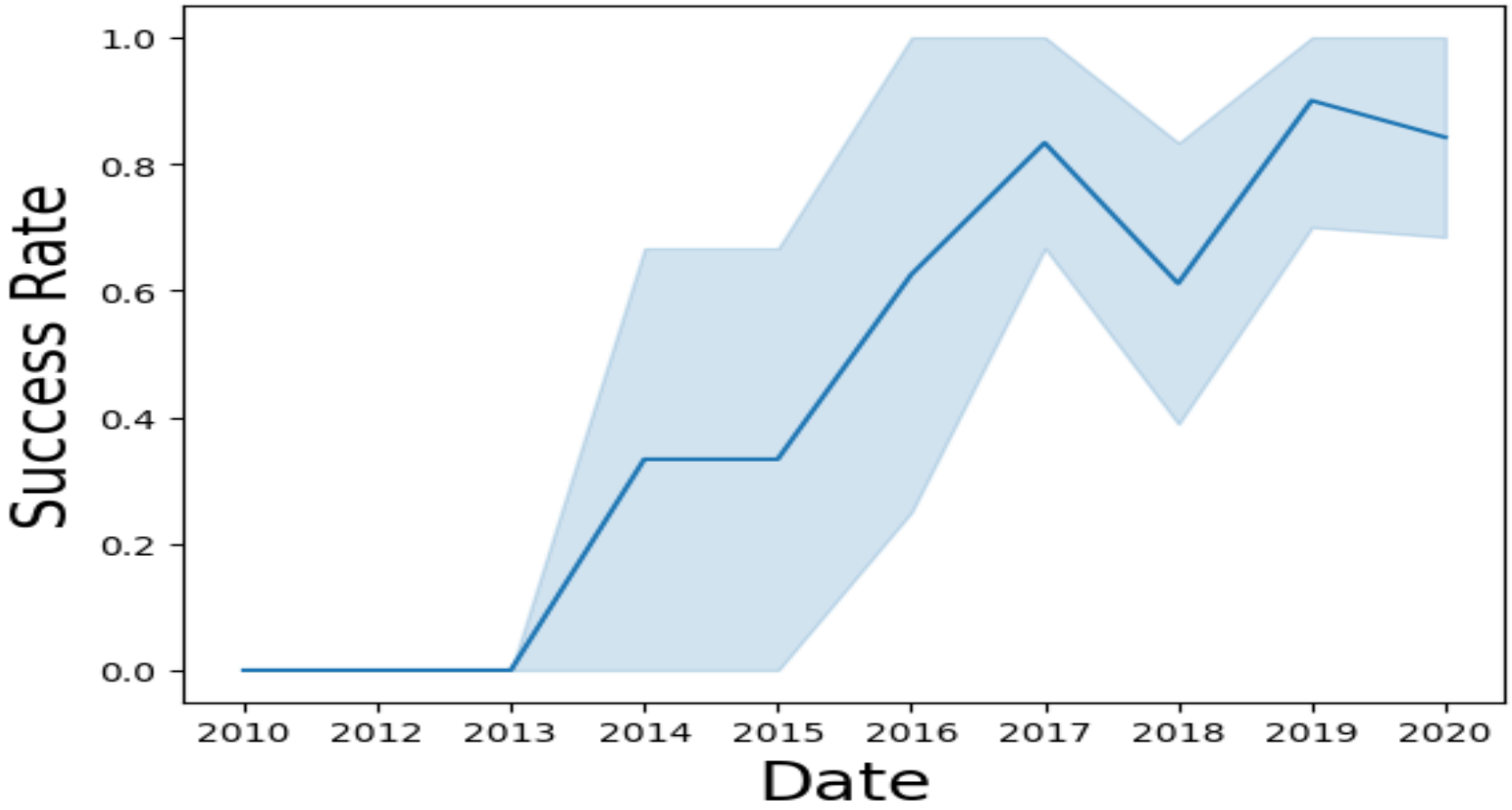
Payload vs. Orbit Type



Explanation:

- With heavy payloads the successful landing or positive landing rate are more for 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



Explanation:

- The general trend in success rate has been increasing since inception however the success rate since 2013 kept increasing till 2020 significantly.

All Launch Site Names

There are four launching sites that participated in the space mission namely;

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

Launch Site Names Begin with 'CCA'

Below is a table showing the first five records of Launching Sites starting with “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

The total payload carried by boosters from NASA is zero

SUM(PAYLOAD_MASS__KG_)
None

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2928.4 Kg

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

The dates of the first successful landing outcome on ground pad was the 22nd of December the year 2015.

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The following is a list of 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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes was 99

COUNT(Mission_Outcome)
99

Boosters Carried Maximum Payload

The following is a list of names of the booster versions which have carried the maximum payload mass

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

- The following list shows the failed landing outcomes in drone ship, their booster versions, and launch site names and their month for in the year 2015.

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
01	F9 v1.1 B1017	VAFB SLC-4E	Failure (drone ship)
03	F9 FT B1020	CCAFS LC-40	Failure (drone ship)
06	F9 FT B1024	CCAFS LC-40	Failure (drone ship)

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

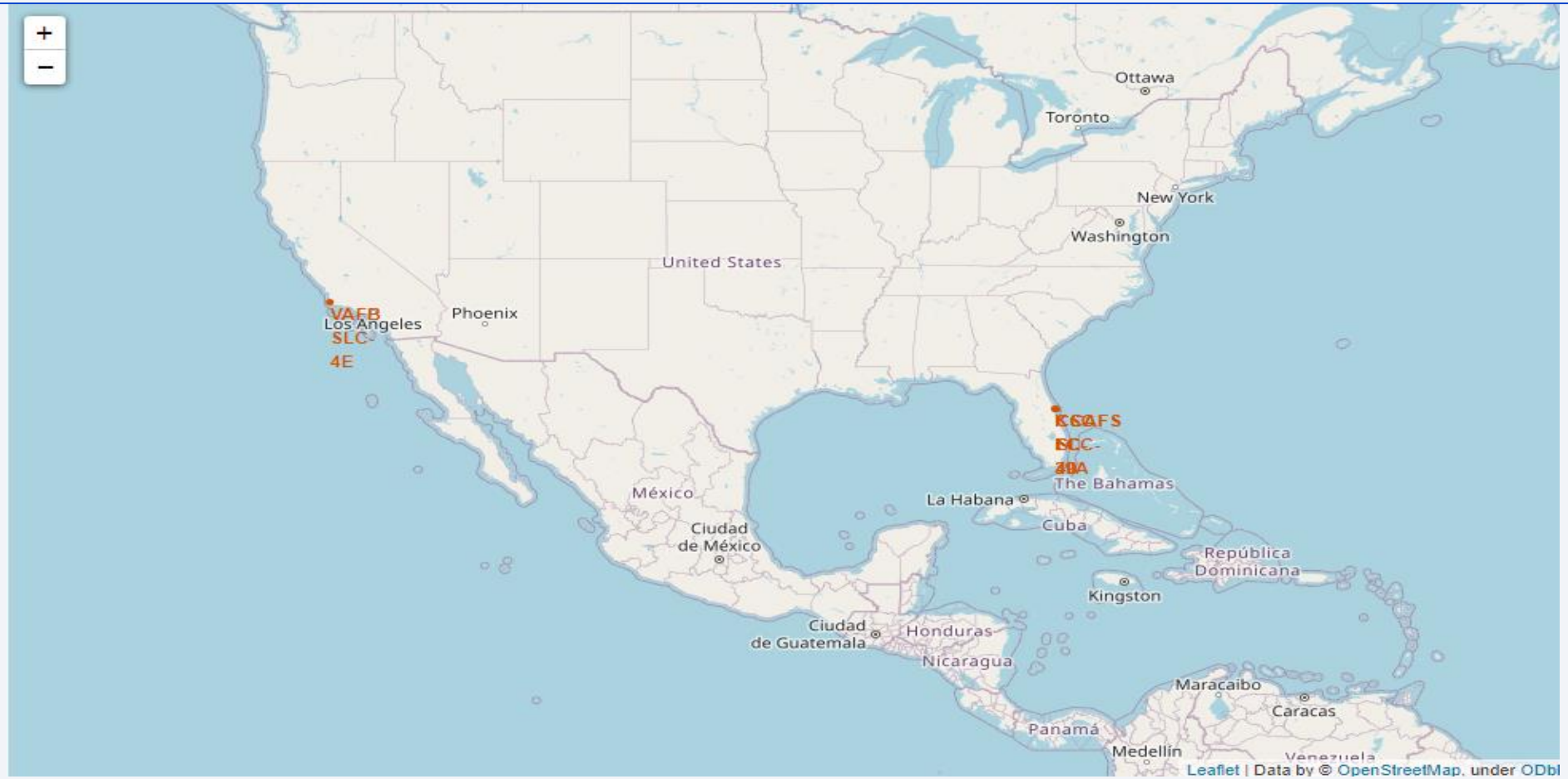
The table below ranks the count of landing outcomes for Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Date	Time (UTC)	Booster_ Version	Launch_S ite	Payload	PAYLOAD _MASS__ KG_	Orbit	Customer	Mission_ Outcome	Landing_ Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

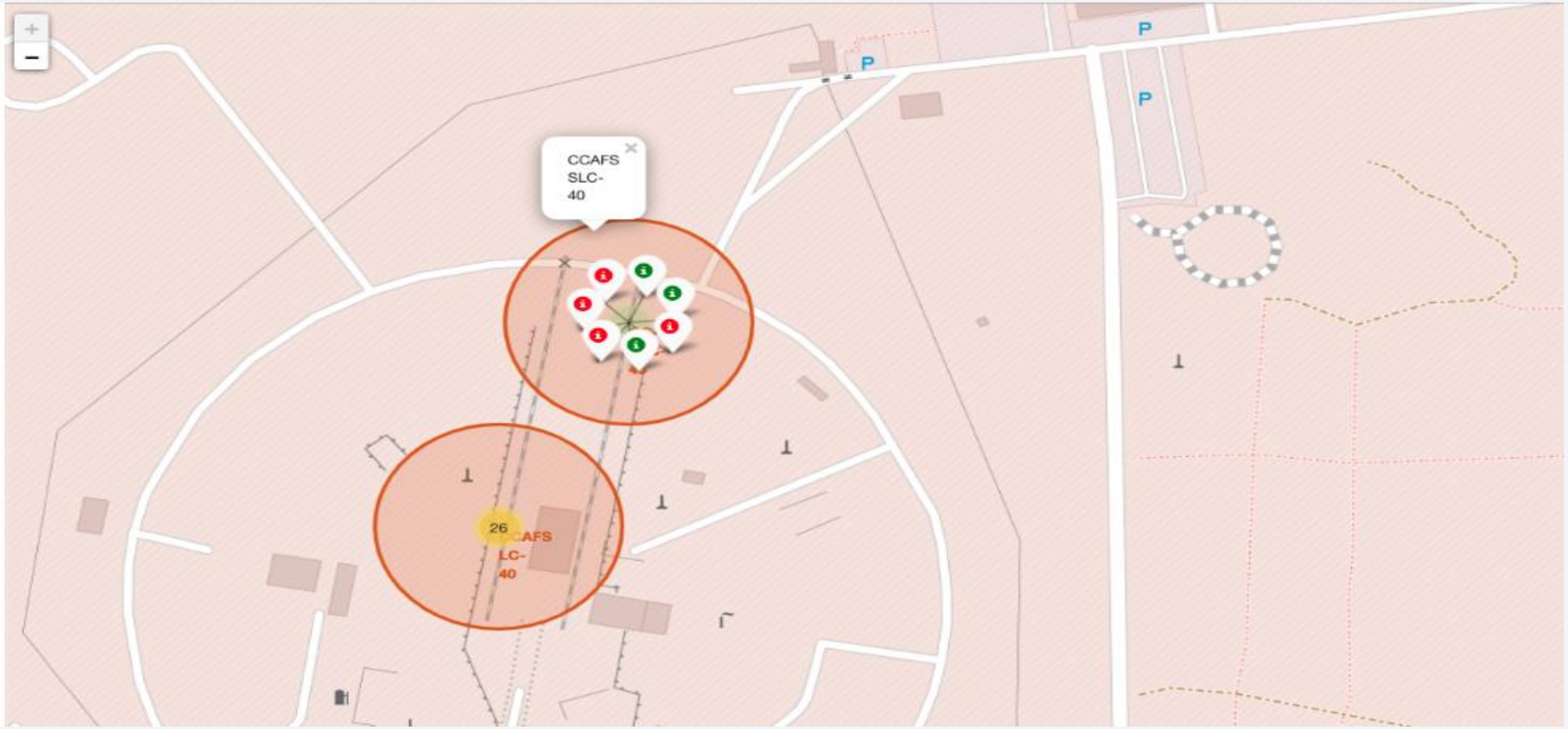
Launch Sites Proximities Analysis

Launch Site Geo-Positions on a Global Map



- All launch sites are in proximity to the Equator line?
- All launch sites are in very close proximity to the coast?

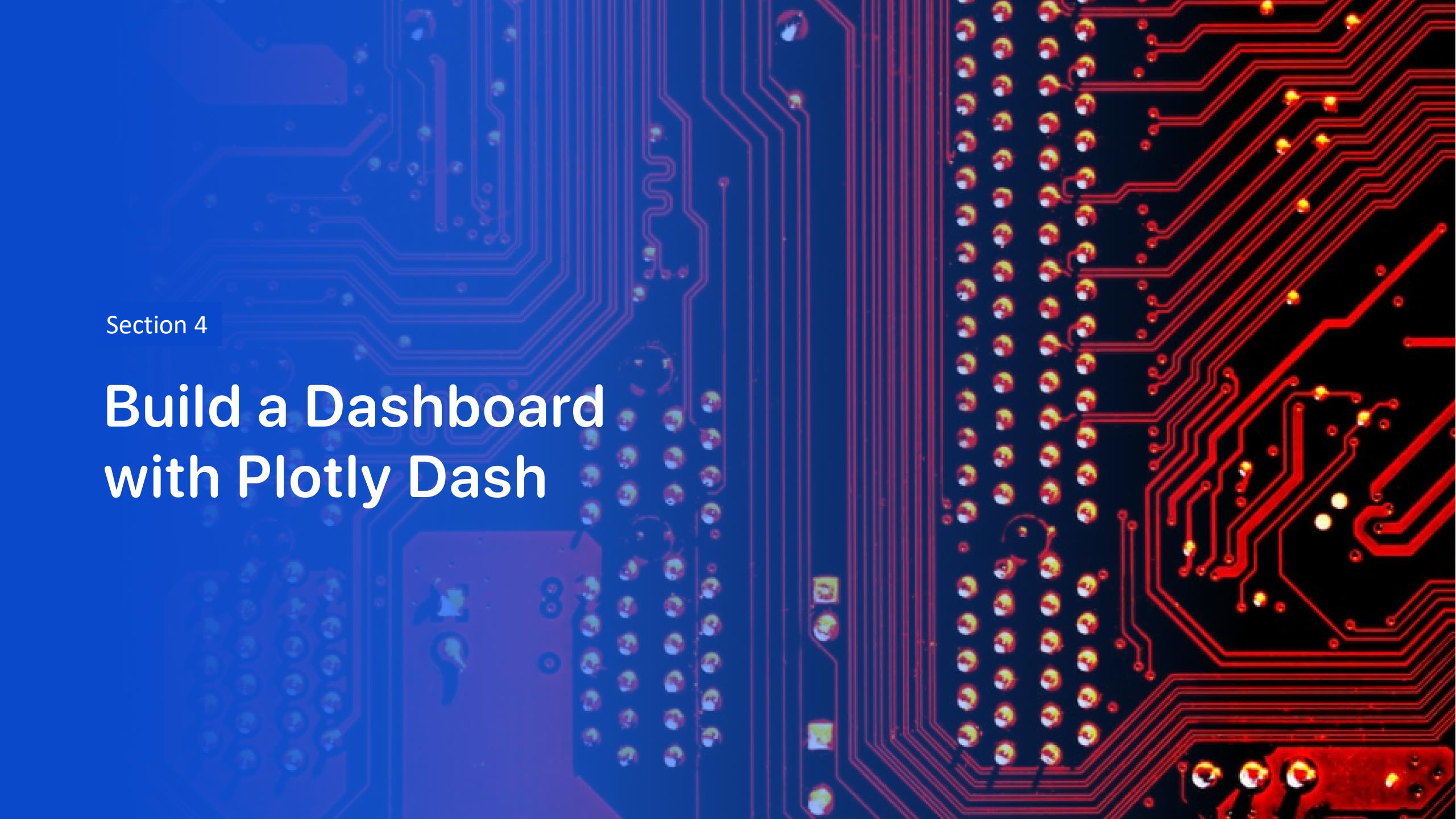
Map Showing Success Rate Per Launch Site



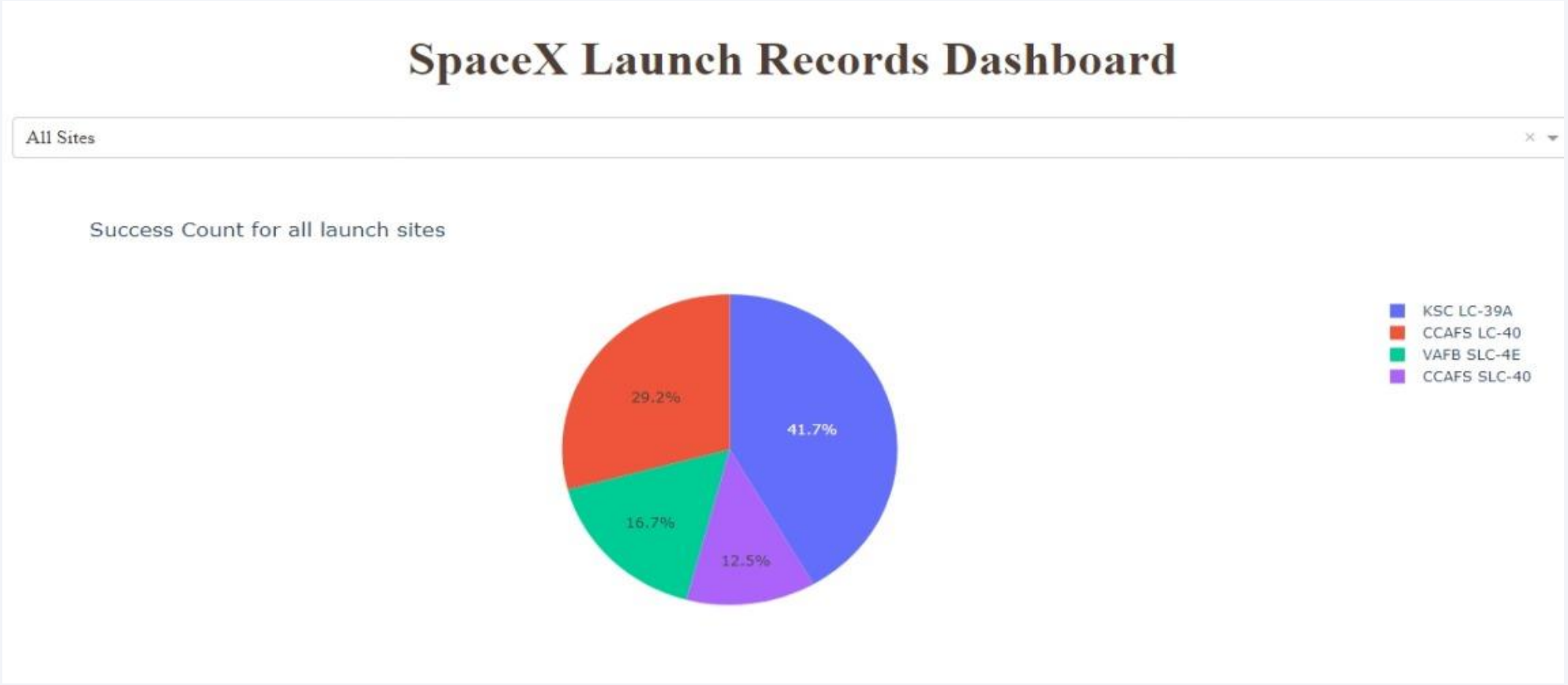
Launch site CCAF3 SLC 40, has a relatively moderate success rate with four successful launches to 3 failures

A MAP Showing Proximities Of Launch sites with Distance

From the map above we can see that the distance between launch site CCAFS SLC-40 with the coastline in 0.9KM

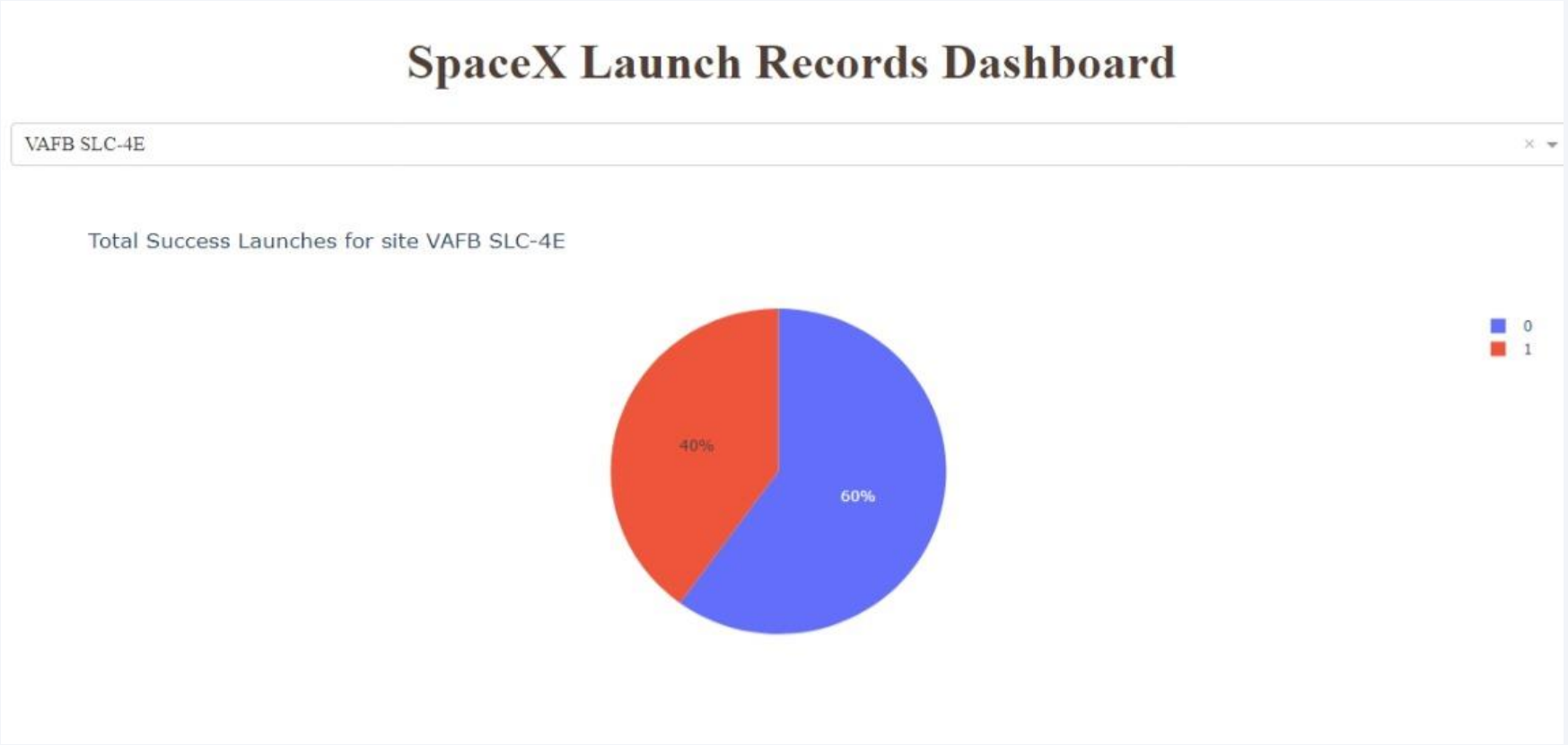

Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Pie Chart For All Sites-Dashboard



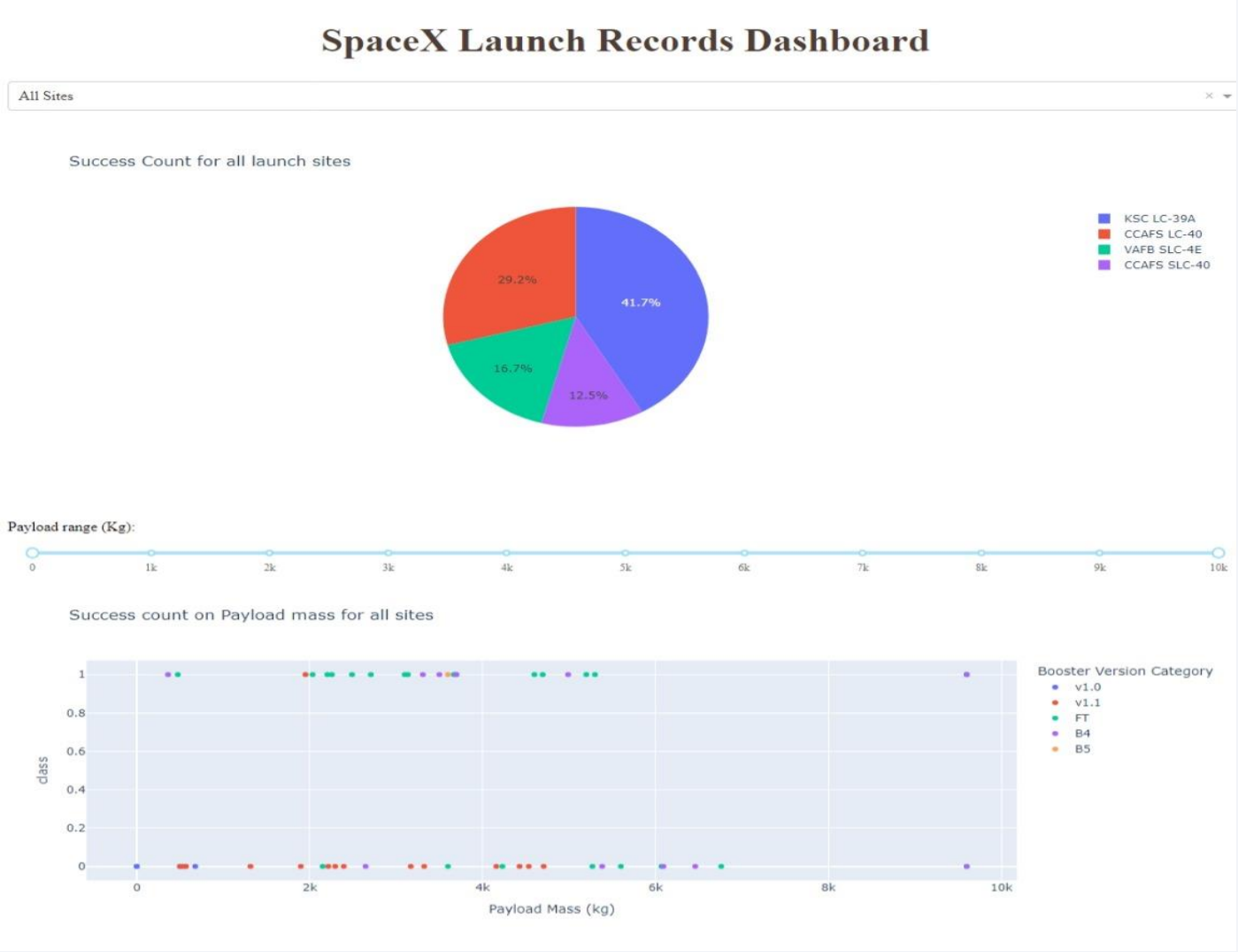
From the pie chart, the Success count was highest in KSC LC-394 with 41.7% and lowest in CCAFS SLC -40 with 12.5%

Highest Launch Success ratio Pie Chart For Launch Site



From the pie chart, VAFC SLC -4E has the highest launch success ratio with 40% success rate

Payload vs. Launch Outcome scatter plot for all sites

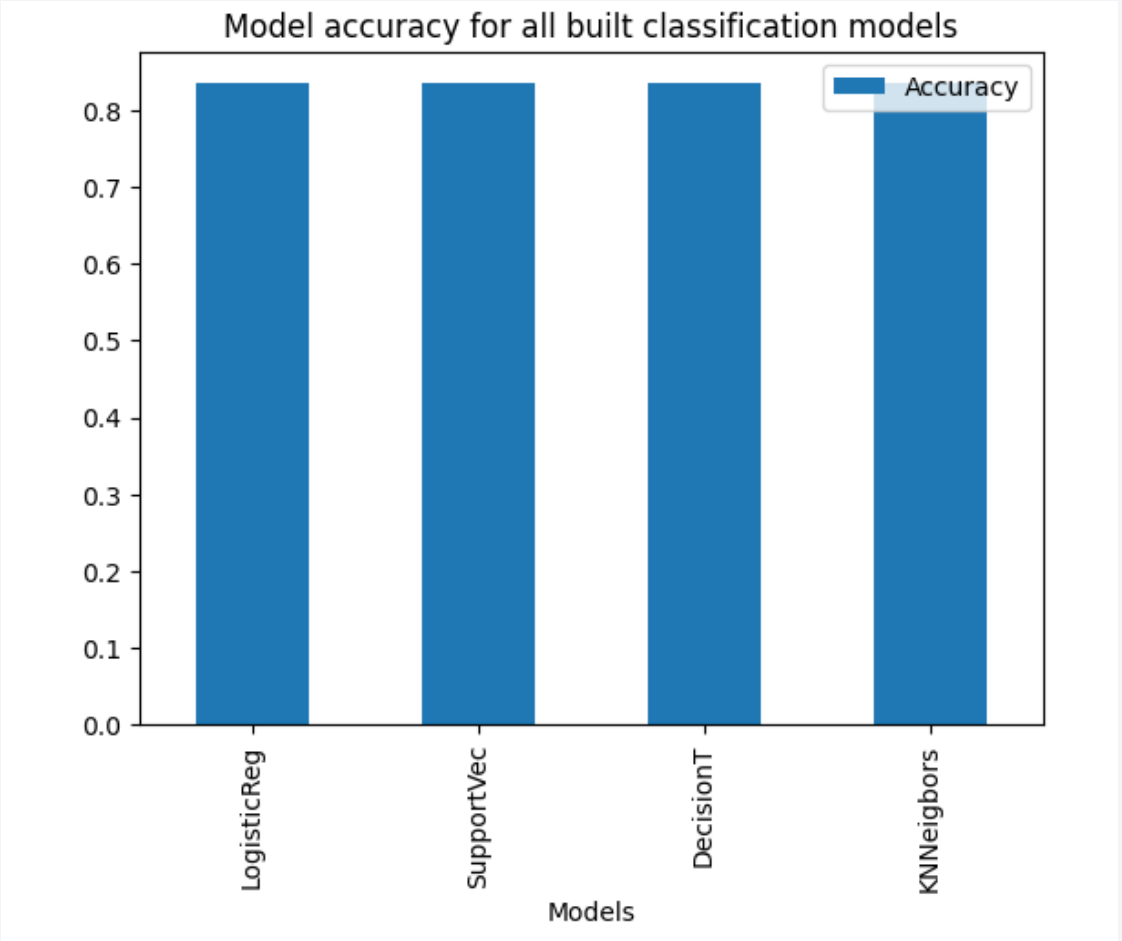


The scatter shows a variation in the points as the payload range slide is shifted from 0 to other points .

Section 5

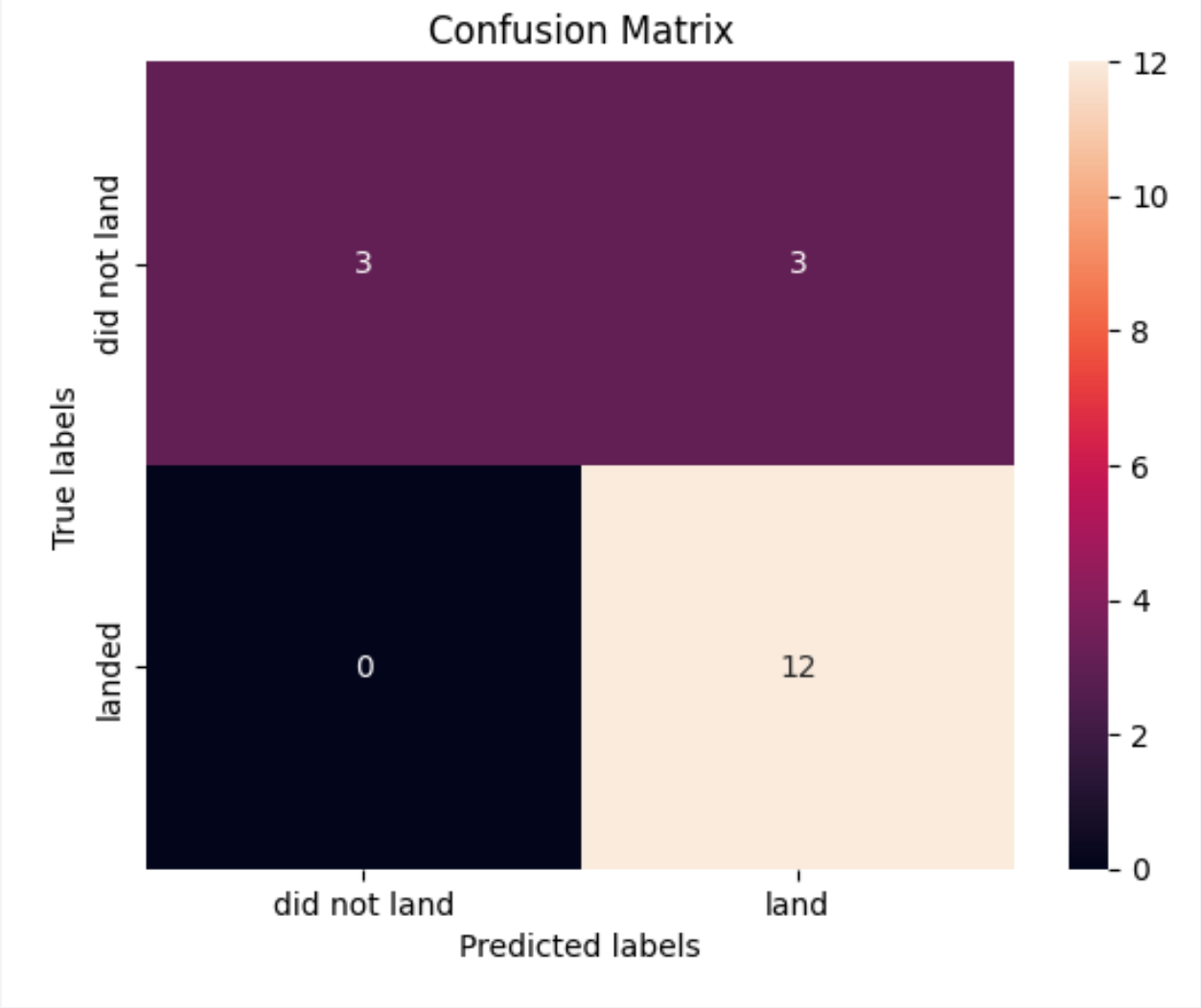
Predictive Analysis (Classification)

Classification Accuracy



All models had the same classification accuracy 0f 83%.

Confusion Matrix



Conclusions

- The more the pay load mass the lower the success rate for a space mission launch
- Falcons launched in the ES-L1, GEO, HEO, SSO Orbits have a high success rate
- All the classifications performed well on the test data with an 83% accuracy score
- The average payload mass carried by booster version F9 v1.1 is 2928.4 Kg
- In the LEO orbit the Success appears related to the number of flights

Appendix

- "https://api.spacexdata.com/v4/payloads/"+load).json()
- "https://api.spacexdata.com/v4/rockets/"+(x)).json()
- "https://api.spacexdata.com/v4/launchpads/"+(x)).json()
- "https://api.spacexdata.com/v4/launches/past"
- "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

Thank you!