IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rajni Gulati
January 11, 2022

Outline

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

Executive Summary - A

Summary of methodologies

Research Method

- Data collection via Web Scraping, API's
- Understand the data using SQL queries, Python.
- Visualizing the data with charts and graphs
- Using Machine learning algorithms to predict if the first stage will land

Data Sources

- SpaceX API
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Executive Summary - B

Data Collection Process

- API's and Web Scraping to extract the data from the web.
- The results were stored in .csv/excel files and converted to python data frames.
- Python and SQLite to clean, organize and analyze the data.
- Matplotlib, Seaborn and Folium to visualize key findings.
- Machine Learning Models for Predictive Analysis

Executive Summary - C

Summary of all results

- There are 4 launch sites - one in CA and the rest in FL
- Majority of rocket launches took place at Launch site CCAFS SLC 40
- Flight No. 60 and above had more successful launches
- No rockets with payload mass above 10,000 kg were launched at site VAFB-SLC
- Orbit Types ES-LI, GEO, HEO, SSO had the most success rates
- Orbit Types GTO, SO had the lowest success rates
- Majority of flights were for Orbits GTO, PO, ISS and LEO

Executive Summary - D

Summary of all results

- Most rockets with payload mass under 8000 were for Orbit type GTO
- Yearly success rate over 80% since 2019
- Total Payload Mass carried by NASA (CRS) = 45596
- Average Payload Mass for Booster F9v1.1 = 2928.4
- First successful landing outcome (ground pad) was in 12-22-2015
- 4 boosters successfully landed on drone ship with payload mass between 4000-6000.

Executive Summary - E

Summary of all results

- There were 100 successful mission outcomes and 1 failed outcome
- There are 12 boosters which carried maximum payload mass of 15,600
- There were 2 failed landing outcomes in drone ship in 2015
- Successful landing outcomes (drone ship) ranked first followed by (ground pad)
- 4 classification models used to determine success accuracy. All 4 showed the same accuracy score of 83%.

Introduction

Project background and context

- The purpose of this project is to determine if the Falcon 9 first stage will land successfully.
- SpaceX Falcon 9 rocket launches cost 62 million dollars; other providers cost 165 million dollars. The savings is a result of SpaceX reusing the first stage.
- First stage landing successfully translates into millions of dollars in saving.

Problems you want to find answers

- Determine if the first stage will land.

Section 1

Methodology

Methodology

- Data collection methodology
  - Extract the data from the web.
  - The results converted to Pandas data frames.
- Data wrangling
  - Converted booster landing outcome into 1/0; 1 = successful 0 = unsuccessful
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Prepared, standardized data to split into training and testing sets.
  - Found the best parameters for the models
  - Calculated accuracy of each model to determine which performs the best.

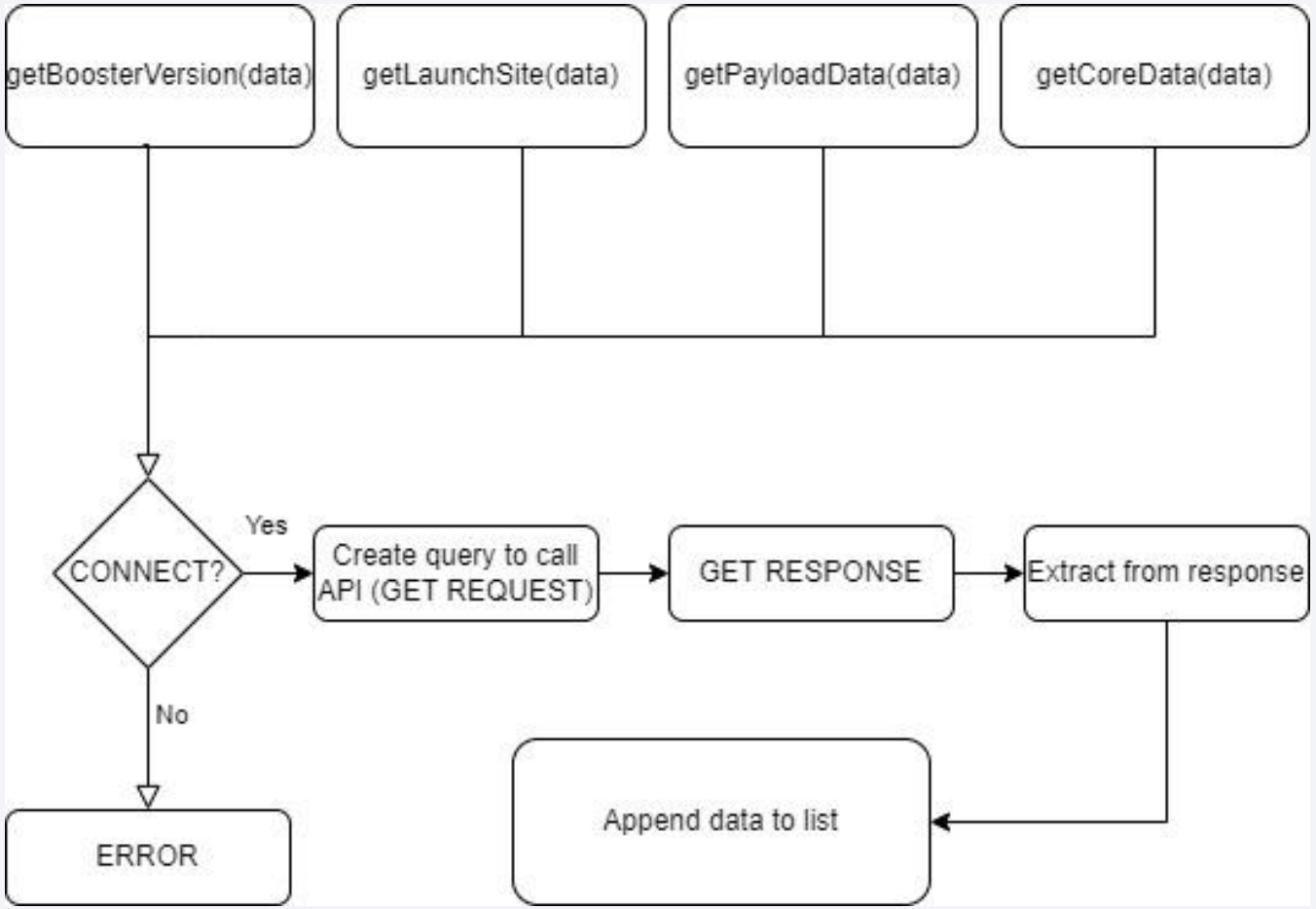
Data Collection

Data collection process in brief:

- Extracted data via SpaceX API
- Web Scraped Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`.
- The results were stored in .csv/excel files and converted to python data frames.
- Python and SQLite to clean, organize and analyze the data.

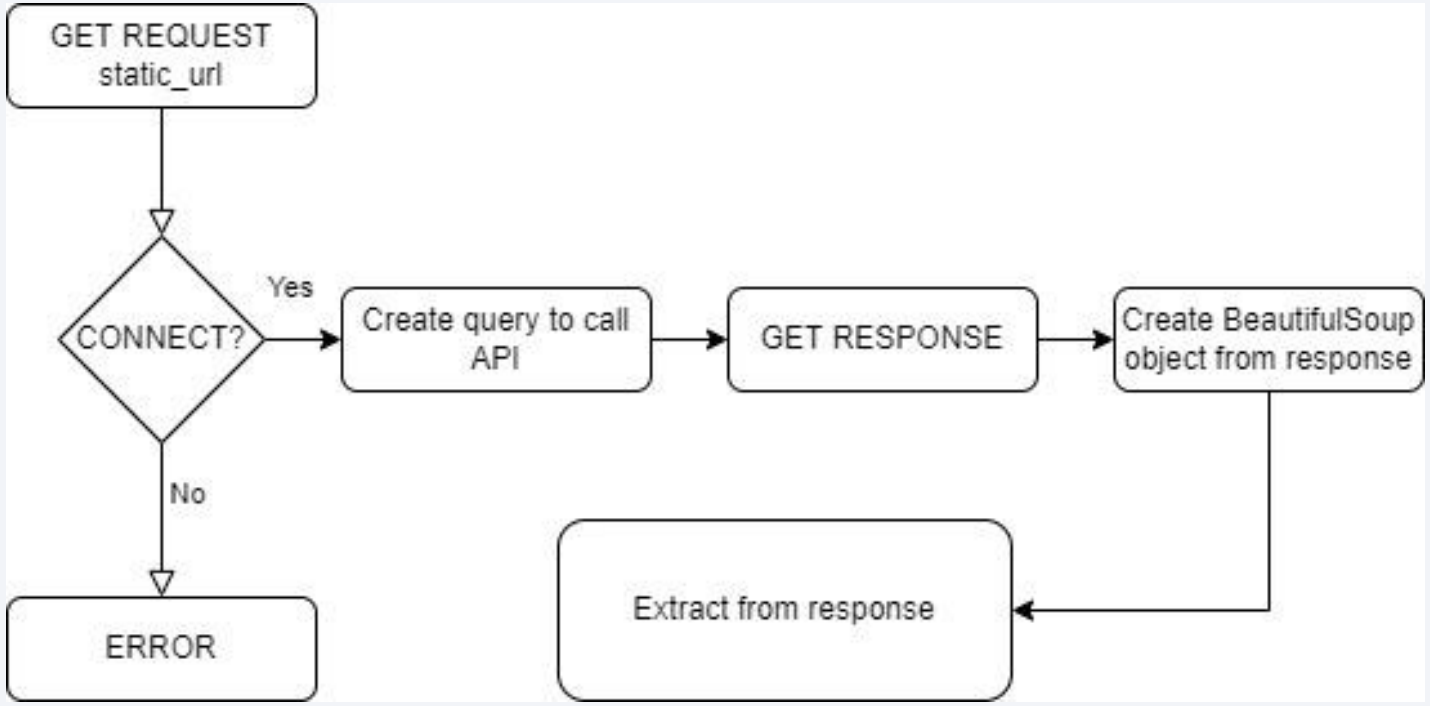
Data Collection – SpaceX API

- Used the SpaceX API to extract the data
- Requested and parsed data with the GET method
- Converted data into a data frame
- Filtered data frame to include only Falcon 9 launches
- Data Collection Notebook



Data Collection - Scraping

- Performed Web Scraping on Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`.
- Extracted column names from HTML table header
- Created data frame by parsing the HTML tables
- Web Scraping Notebook



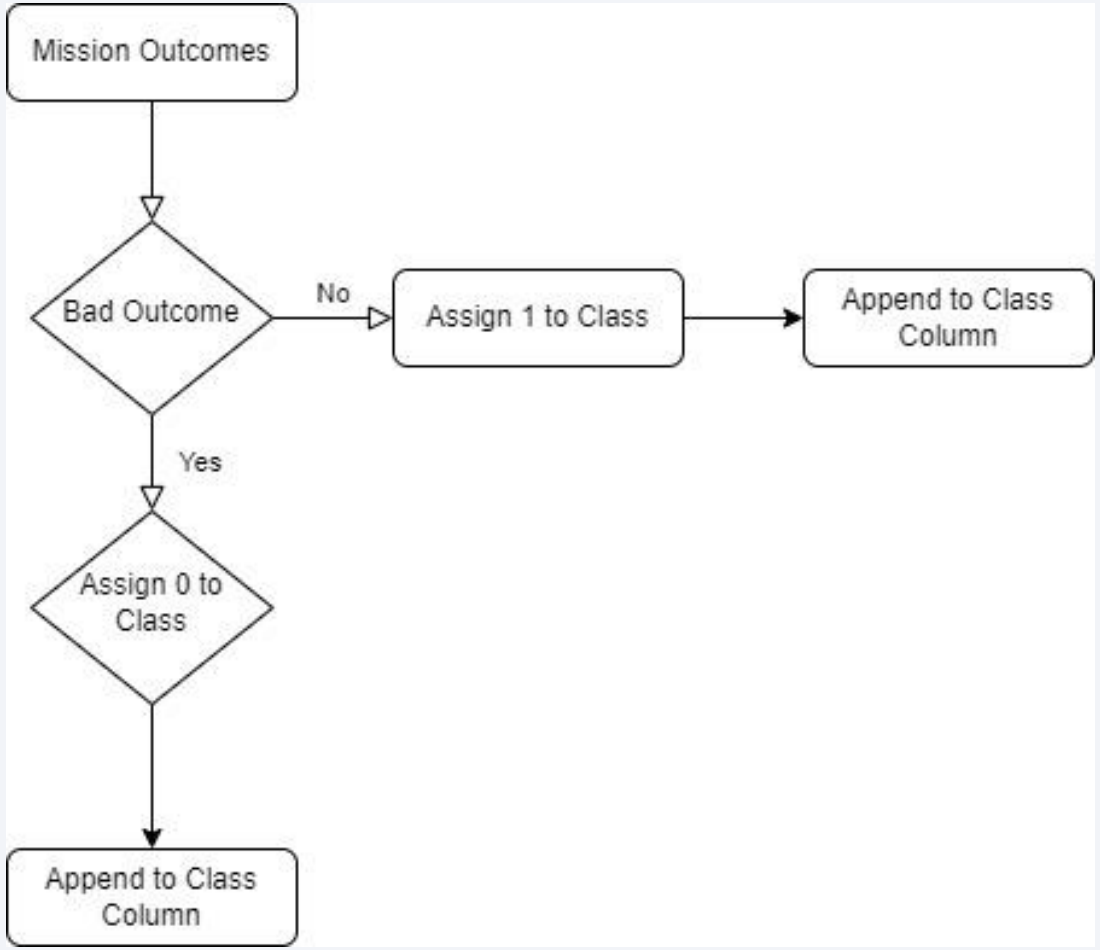
Data Wrangling

- Converted Mission Outcomes
- Assigned outcome as:

  1 = Successful

  0 = Unsuccessful
- Appended results into the Class column

- Data Wrangling Notebook



EDA with Data Visualization

Following charts were used to perform exploratory data analysis in order to determine the correlation between the variables.

- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- Success Rate vs Orbit Type
- Flight Number vs Orbit Type
- Payload Mass vs Orbit Type
- Launch Success Yearly Trend
- EDA with Data Visualization notebook

EDA with SQL - A

SQL Queries Performed

- Names of the unique launch sites
- 5 records where launch sites begin with `CCA`
- Total payload carried by boosters from NASA
- Average payload mass carried by booster version F9 v1.1
- Date of the first successful landing outcome on ground pad
- Names of boosters successfully landed on drone ship with payload mass greater than 4000 but less than 6000
- EDA with SQL Notebook

EDA with SQL - B

SQL Queries Performed – Con't

- Total number of successful and failure mission outcomes
- Names of the boosters which have carried the maximum payload mass
- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes (Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- EDA with SQL Notebook

Build an Interactive Map with Folium

The following objects were added to find the optimal location for launch site as there is a correlation between launch site and success rate.

- Marked all launch sites on map using folium circle and marker
- Marked success/failed launches for each site by assigning green and red colors to class
- Added the assigned colors to folium marker and then to folium cluster.
- Created line between launch site and railway to determine distance between the two points
- Interactive Map with Folium Notebook

Build a Dashboard with Plotly Dash

Summary of plots/graphs and interactions added to dashboard

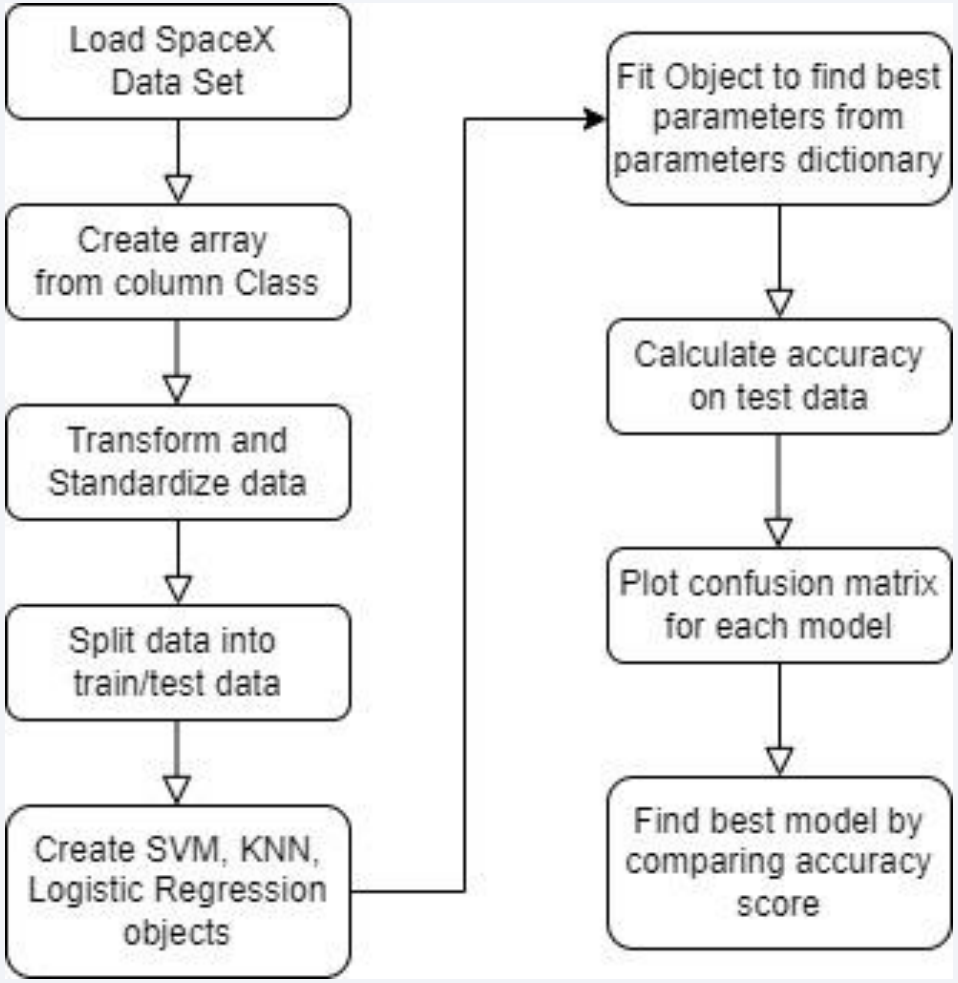
- Added dropdown input component – for launch site selection
- Added pie chart - to depict success/failure percentage for selected launch site
- Added range slider – to select payload mass
- Added scatter chart – to show correlation between payload and launch success
- Added call back functions – to render the pie chart and scatter chart
- Plotly Dash Lab Notebook

Predictive Analysis (Classification)

Summary

- Created a column for the class
- Standardized the data
- Split into training data and test data
- Found the best Hyperparameter for SVM, Classification Trees and Logistic Regression by using the parameters dictionary
- Found the best performing classification model by calculating the accuracy score with the method 'score' and comparing the results

- Predictive Analysis Lab Notebook

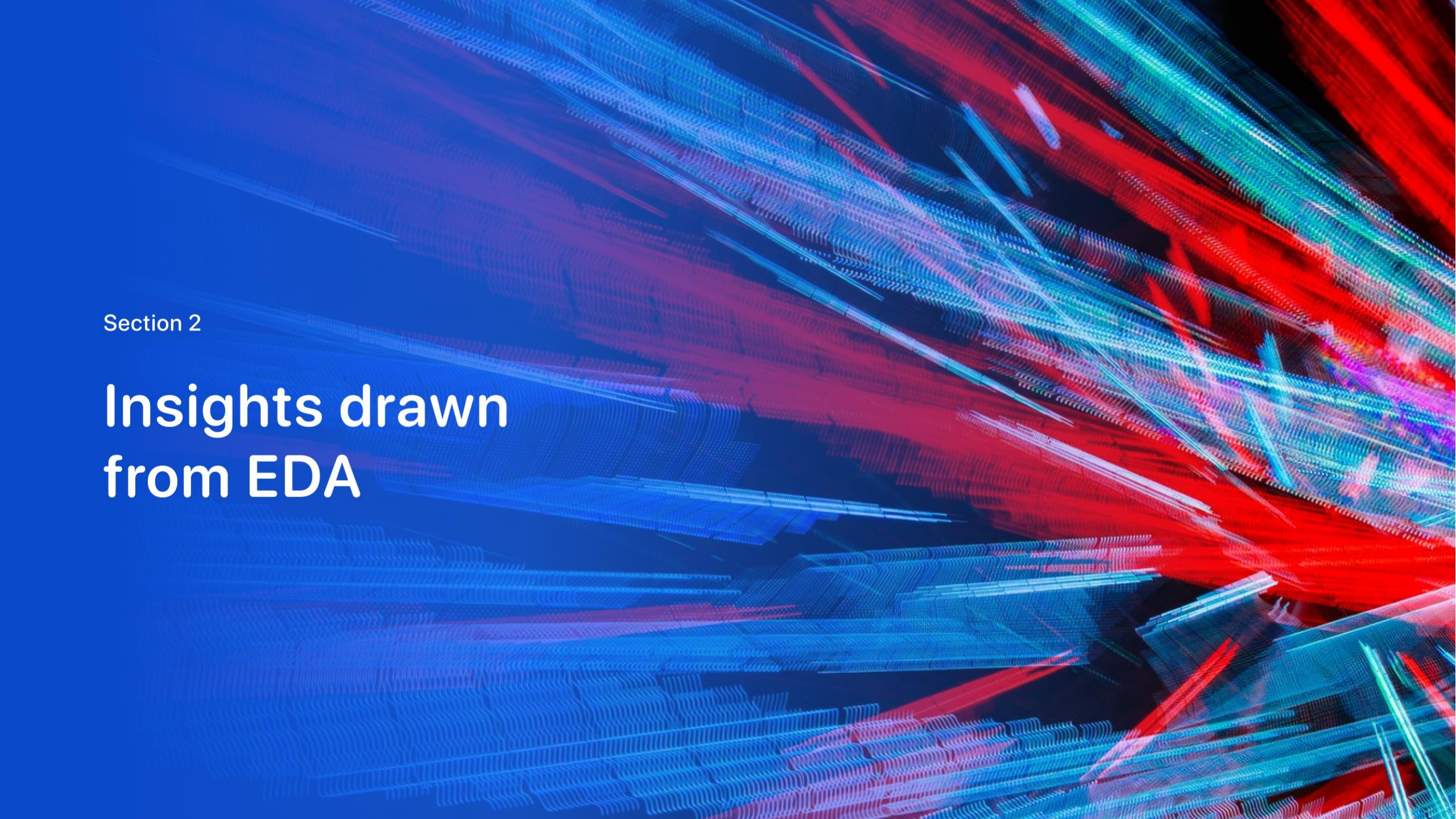


Results

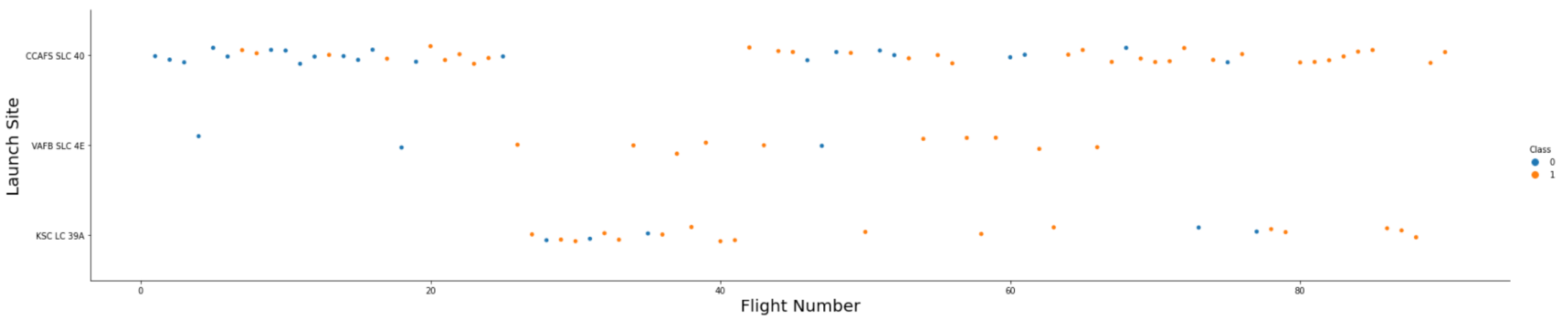
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

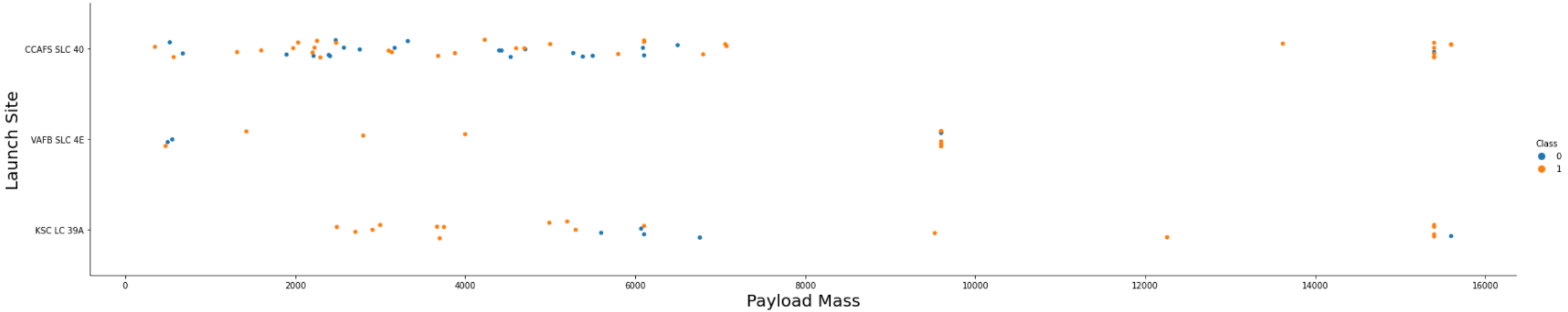


Flight Number vs. Launch Site



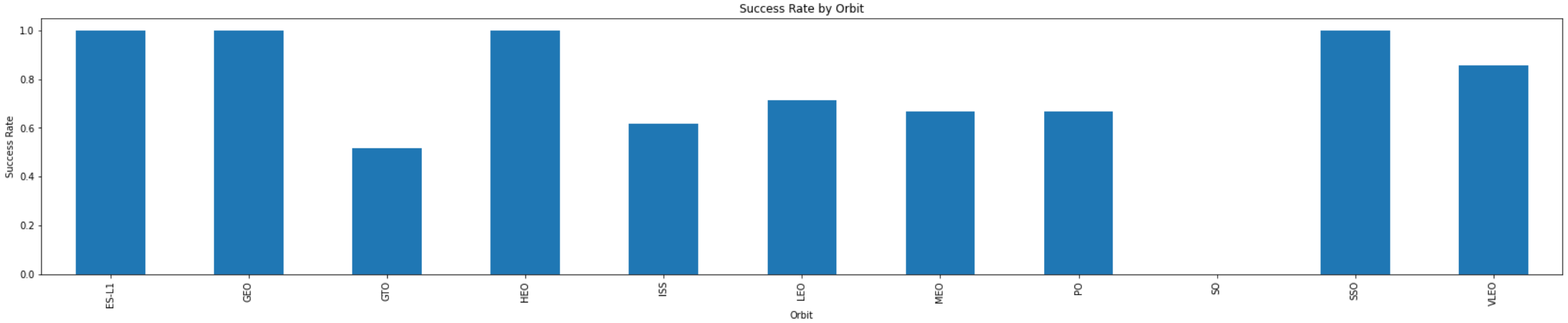
- Majority of rocket launches took place at Launch site CCAFS SLC 40
- Flight No. 60 and above had more successful launches

Payload vs. Launch Site



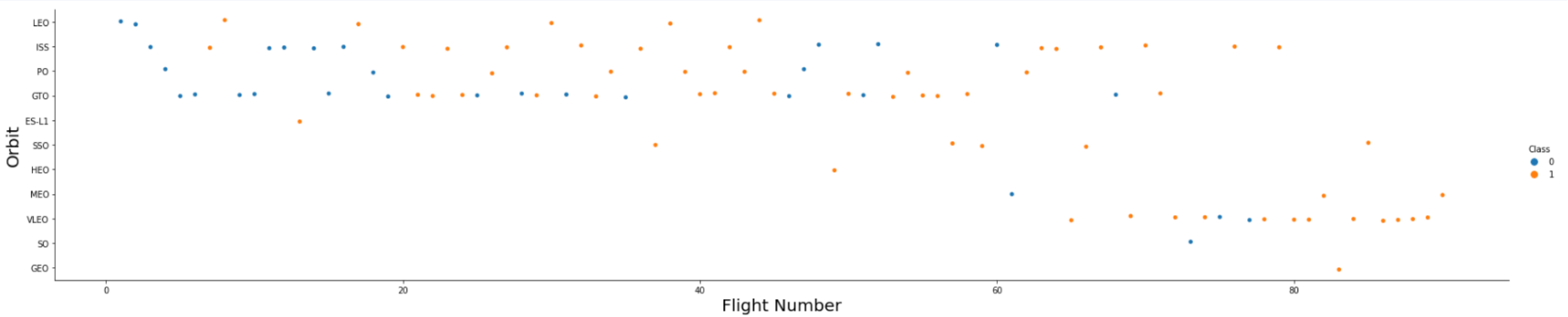
- No rockets with payload mass above 10,000 kg were launched at site VAFB-SLC

Success Rate vs. Orbit Type



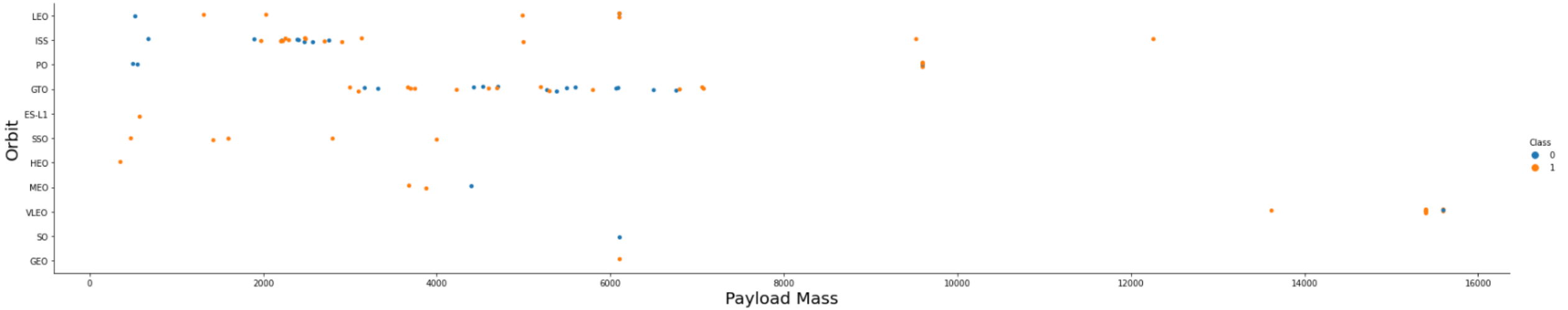
- Orbit Types ES-LI, GEO, HEO, SSO had the most success rates
- Orbit Types GTO, SO had the lowest success rates

Flight Number vs. Orbit Type



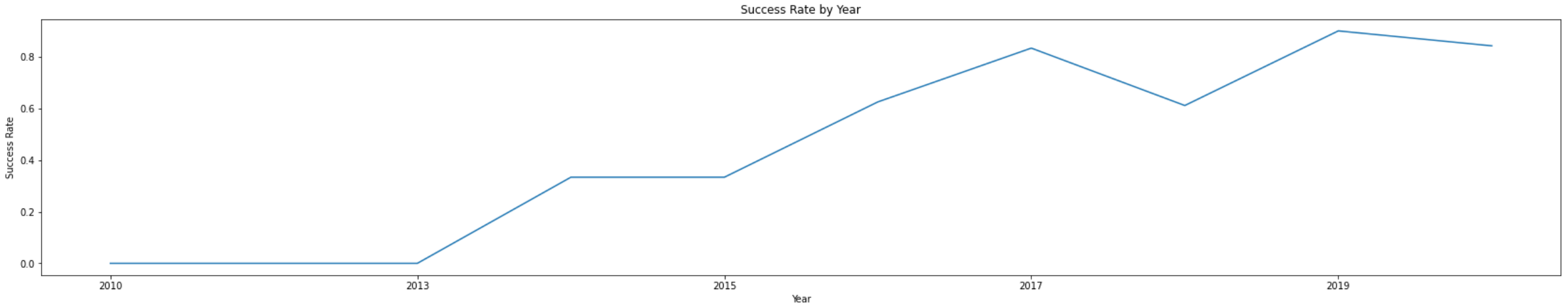
- **Majority of flights were for Orbits GTO, PO, ISS and LEO**

Payload vs. Orbit Type



- **Most rockets with payload mass under 8000 were for Orbit type GTO**

Launch Success Yearly Trend



- **Yearly success rate over 80% since 2019**

All Launch Site Names

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- **Above are the names of the 4 unique launch sites**

Launch Site Names 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

- **Above are the 5 records where launch sites begin with `CCA`**

Total Payload Mass

Total Payload Mass for NASA
107010

Total Payload Mass for NASA (CRS)
45596

- Total Payload Mass carried by NASA = 107010
- Total Payload Mass carried by NASA (CRS) = 45596

Average Payload Mass by F9 v1.1

Avg_Payload_Mass	Booster_Version
2928.4	F9 v1.1

- Average Payload Mass for Booster F9v1.1 = 2928.4

First Successful Ground Landing Date

Date	Landing_Outcome
22-12-2015	Success (ground pad)

- First successful landing outcome (ground pad) was in 12-22-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	Payload_Mass
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

- **Above are the 4 boosters successfully landed on drone ship with payload mass between 4000-6000.**

Total Number of Successful and Failure Mission Outcomes

Outcome	Total
Failure	1
Success	100

- There were 100 successful mission outcomes and 1 failed outcome

Boosters Carried Maximum Payload

Booster_Version	Max_Payload_Mass
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- There are 12 boosters which carry maximum payload mass of 15,600

2015 Launch Records

Month_Name	Year	Booster_Version	Launch_Site	Landing_Outcome
January	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- **There were 2 failed landing outcomes in drone ship in 2015**

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

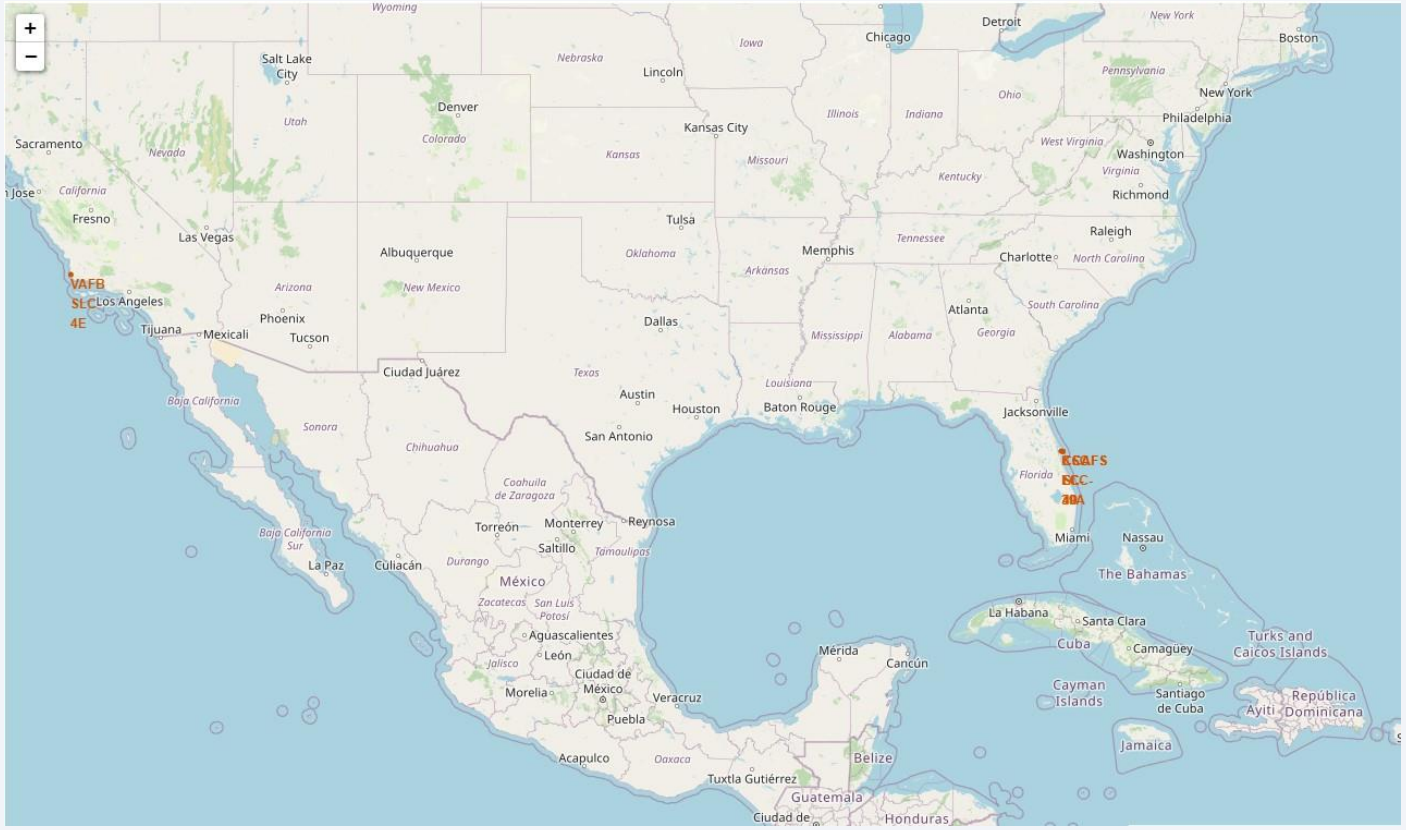
Date	Landing_Outcome	Count	Rank
22-12-2015	Success (ground pad)	3	2
08-04-2016	Success (drone ship)	5	1

- **Successful landing outcomes (drone ship) ranked first followed by (ground pad)**

Section 4

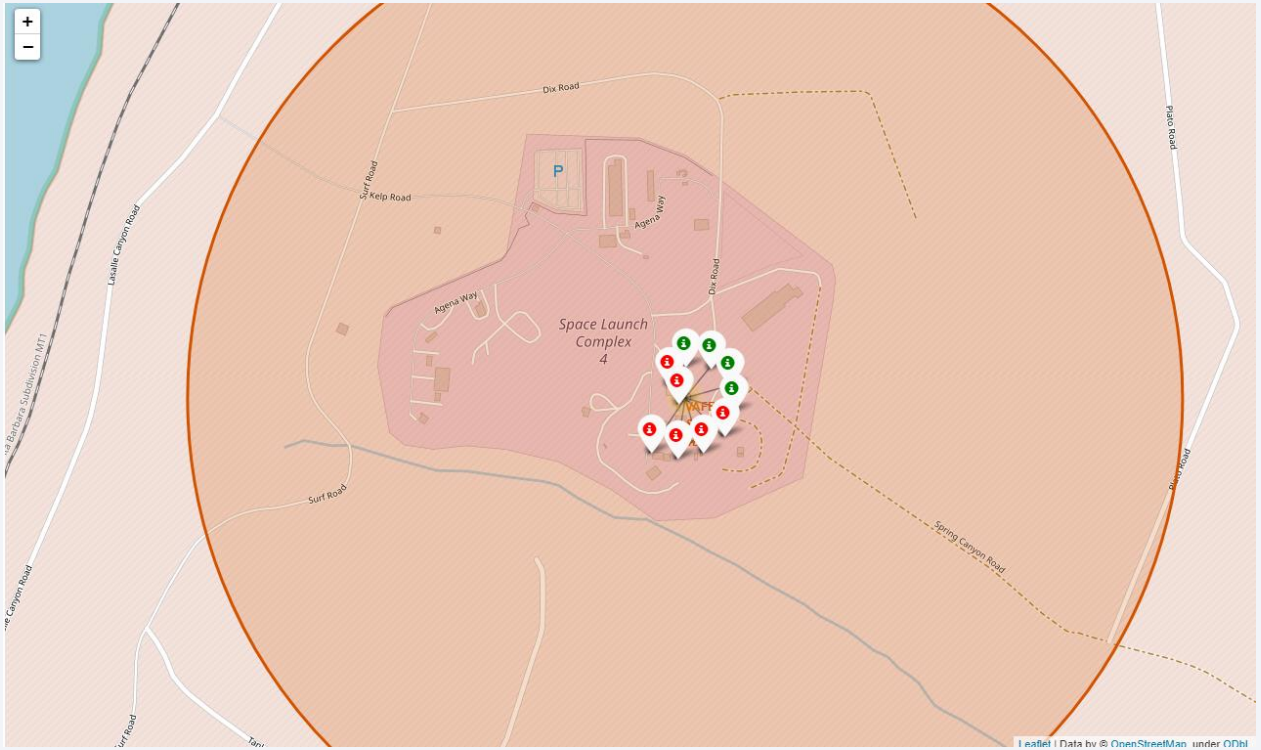
Launch Sites Proximities Analysis

Launch Sites



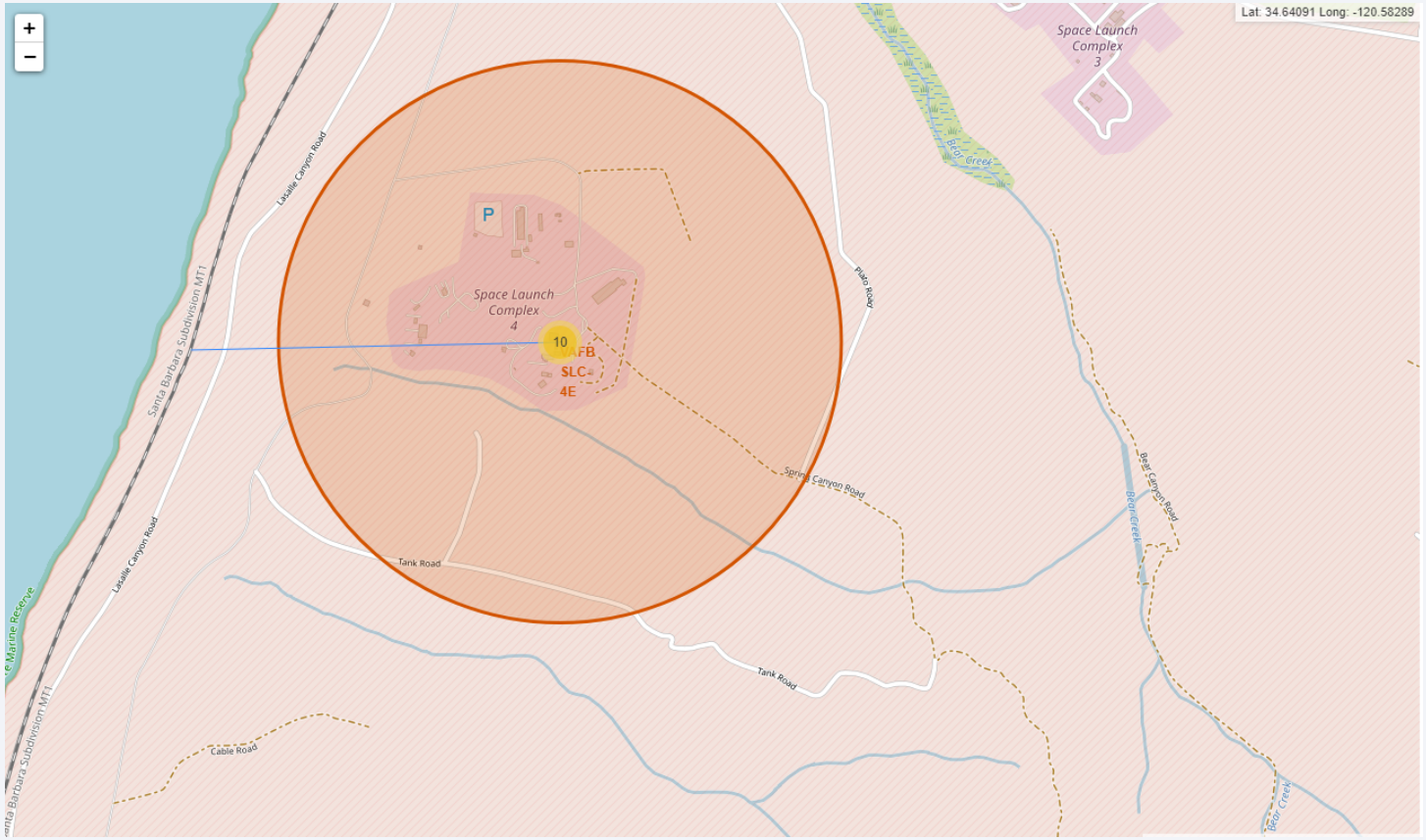
- There are 4 launch sites - one in CA and the rest in FL

Success/Failed Launches



- Successful and failed launches have been marked for each site on the map
- Green = Successful; Red = Failed outcome

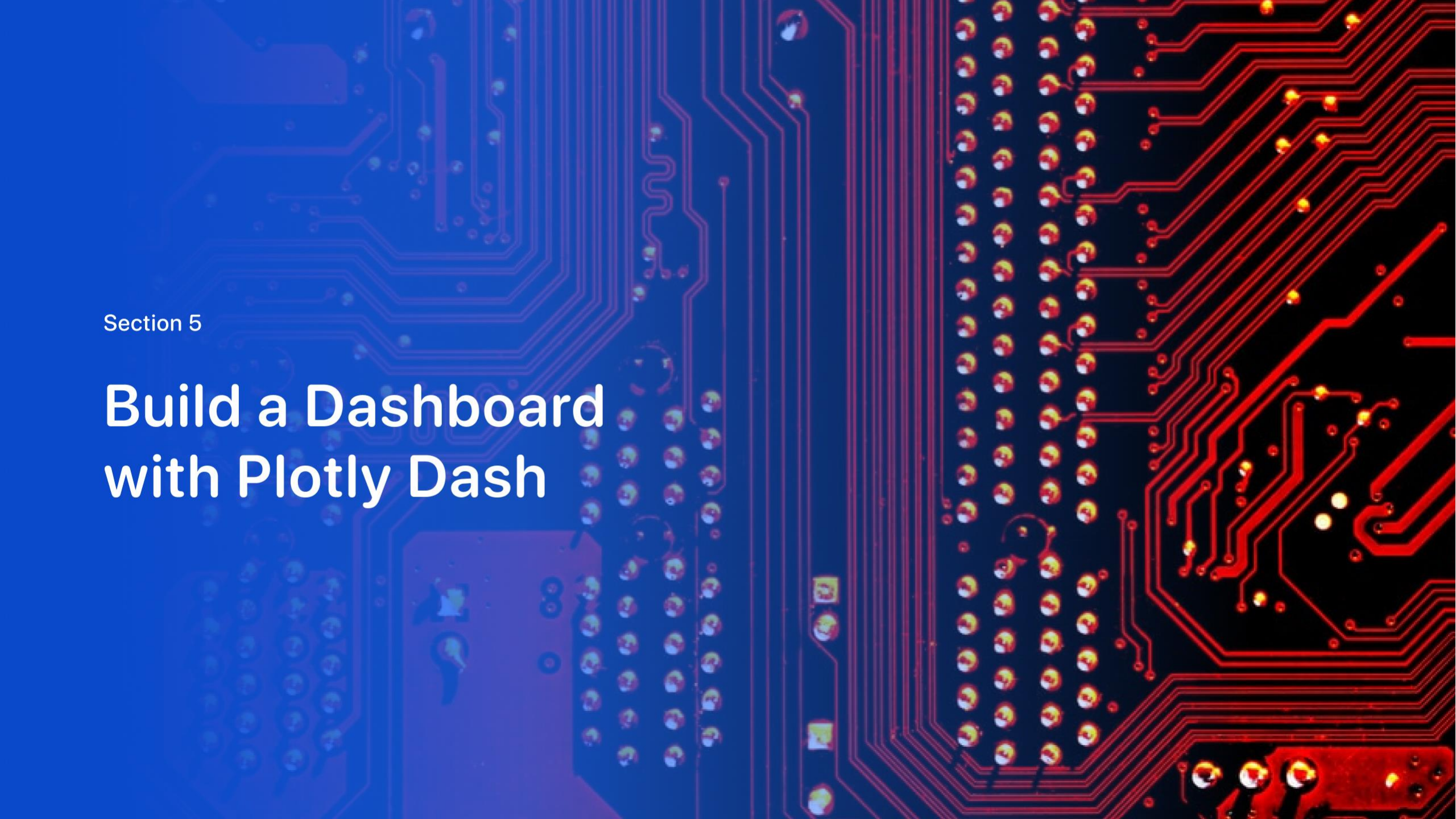
Launch Site and Proximities Distance



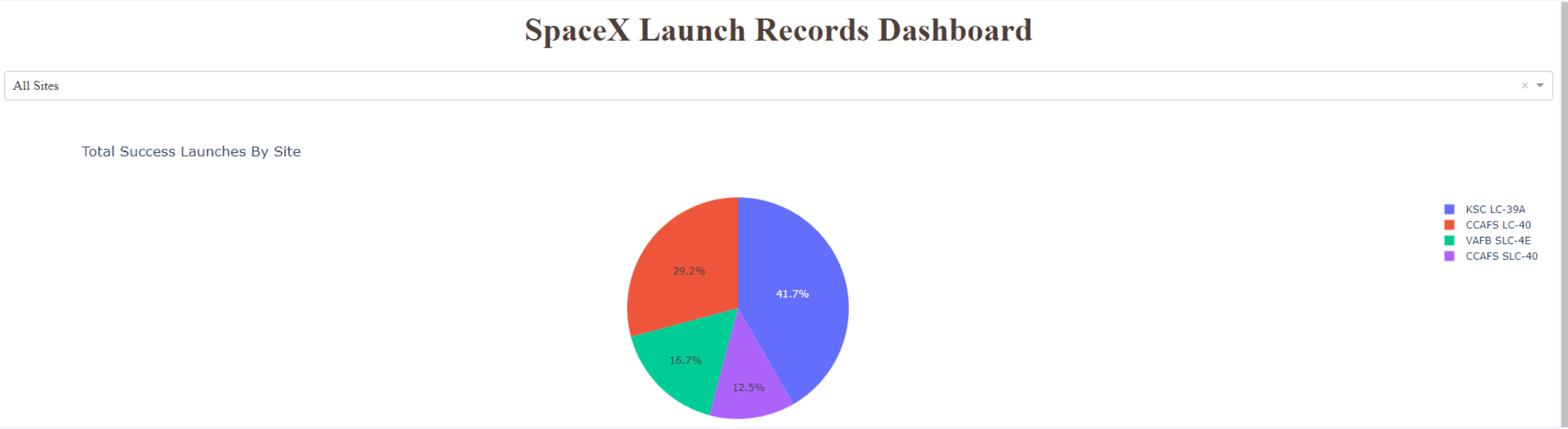
- Distance between launch site and nearest railway is marked by a blue line. Any proximities such as railway, highway, coastline can be calculated.

Section 5

Build a Dashboard with Plotly Dash

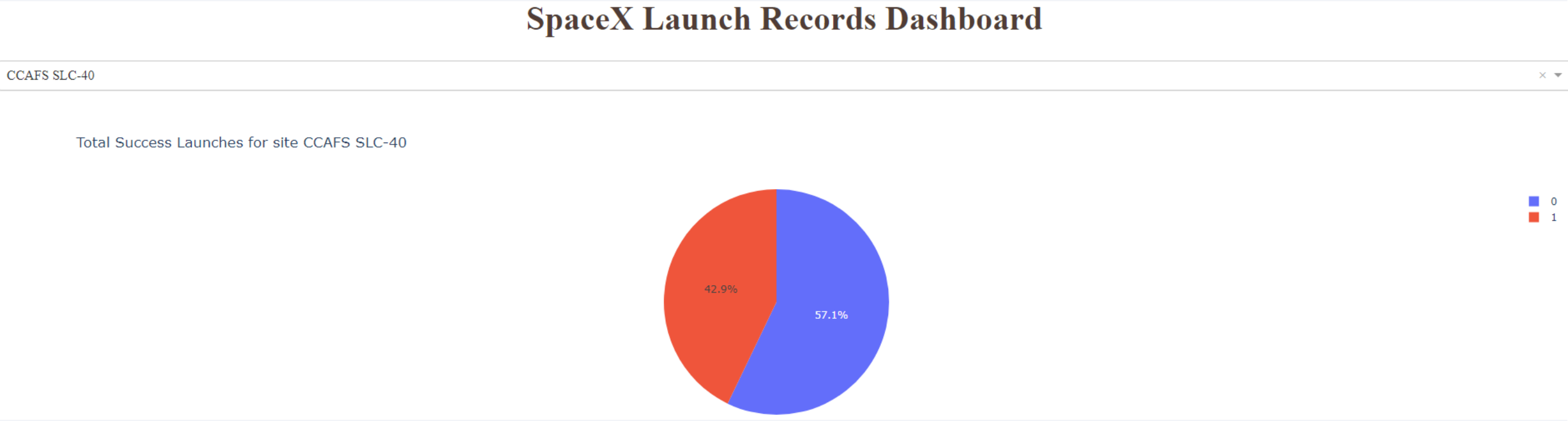


SpaceX Launch Dashboard – All Launch Sites



- Above is a pie chart displaying total success rate for each Launch Site.
- KSC LC-39A has had the most successful launches (41.7%)

SpaceX Launch Dashboard – Highest Success Ratio



- Above is a pie chart displaying launch site CCAFS SLC-40 as having the highest success ratio
- 57.1% successful; 42.9% Failed

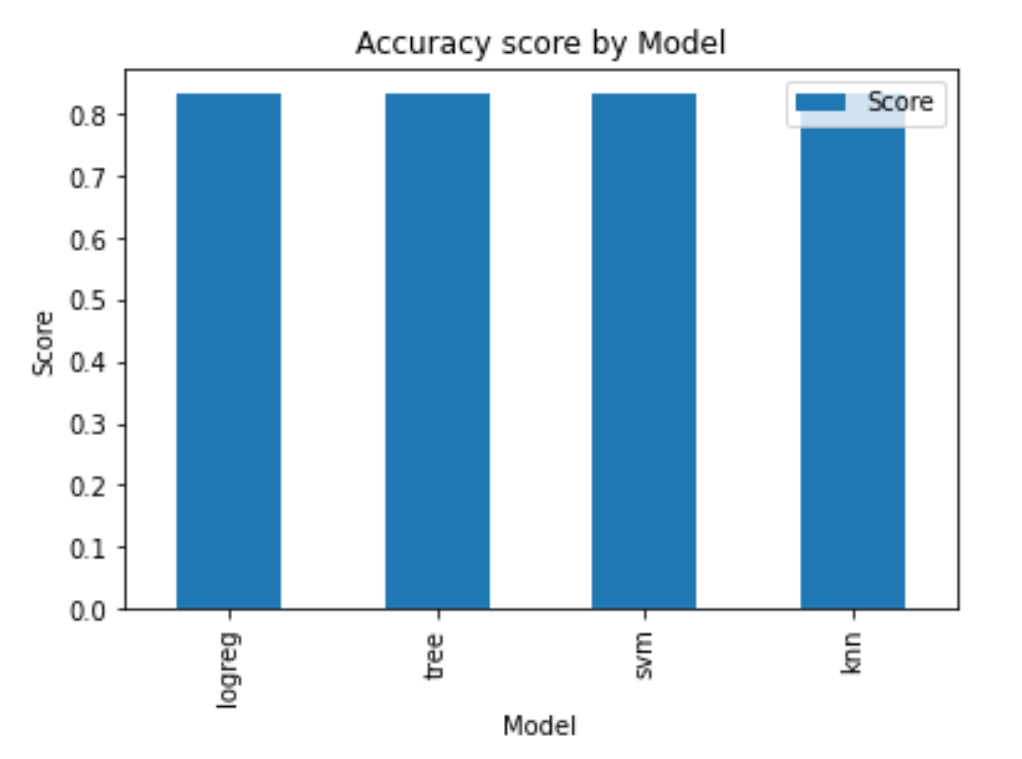
Payload Range vs Success Rate

- Above is a scatter plot is showing booster version category FT having the most success at payload mass of 2000 KG.

Section 6

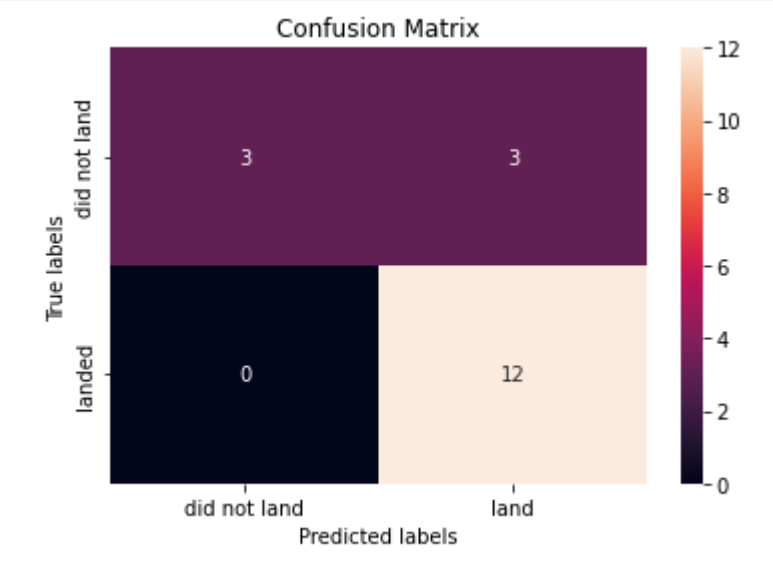
Predictive Analysis (Classification)

Classification Accuracy



- Above is a bar chart displaying the model accuracy of each classification model
- As you can see all display the same result. Each showed a score of 83%

Confusion Matrix



- Below is the confusion matrix of the all the performing models
- All showed the same result

Conclusions

- Launch site CCAFS SLC 40 had the highest success ratio.
- Orbit Types ES-LI, GEO, HEO, SSO had the most success rates.
- Most of successful landing outcomes were on (drone ship)
- All four classification models showed an accuracy score of 83%
- Therefore a first stage will likely land if the above results are noted

Appendix - A

- Below is the SQL query (in sqlite) to 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.

```
%sql select \
case substr(Date,4,2) \
when '01' then 'January' \
when '02' then 'Febuary' \
when '03' then 'March' \
when '04' then 'April' \
when '05' then 'May' \
when '06' then 'June' \
when '07' then 'July' \
when '08' then 'August' \
when '09' then 'September' \
when '10' then 'October' \
when '11' then 'November' \
when '12' then 'December' \
else ' ' \
end as 'Month_Name',\
substr(Date,7,4) as Year, Booster_Version, Launch_Site, Landing_Outcome \
from spacextbl where substr(Date,7,4)='2015' and Landing_Outcome = 'Failure (drone ship)';
```

Appendix - B

- **Below is the result of the SQL query (in sqlite)**

Month_Name	Year	Booster_Version	Launch_Site	Landing_Outcome
January	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Thank you!