IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Collect data from SpaceX API
- Clean and format collected JSON data into table

Data Collection - Scraping

- BeautifulSoup to do web scraping Falcon 9 historical launch records wiki page https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Extract required data from Html into our dataframe

Data Wrangling

- Analyze different cases where the booster did not land successfully.
- Introduce Class column, to label 1 being successfully landed and 0 otherwise.
- Github

EDA with Data Visualization

- In order to predict Falcon 9 first stage will land successfully or not, we need to carefully analyze and understand the relationships between each columns and features.
- Correlation map, scatter plot, bar charts, line graph are best to visualize the relationships.
- Github

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

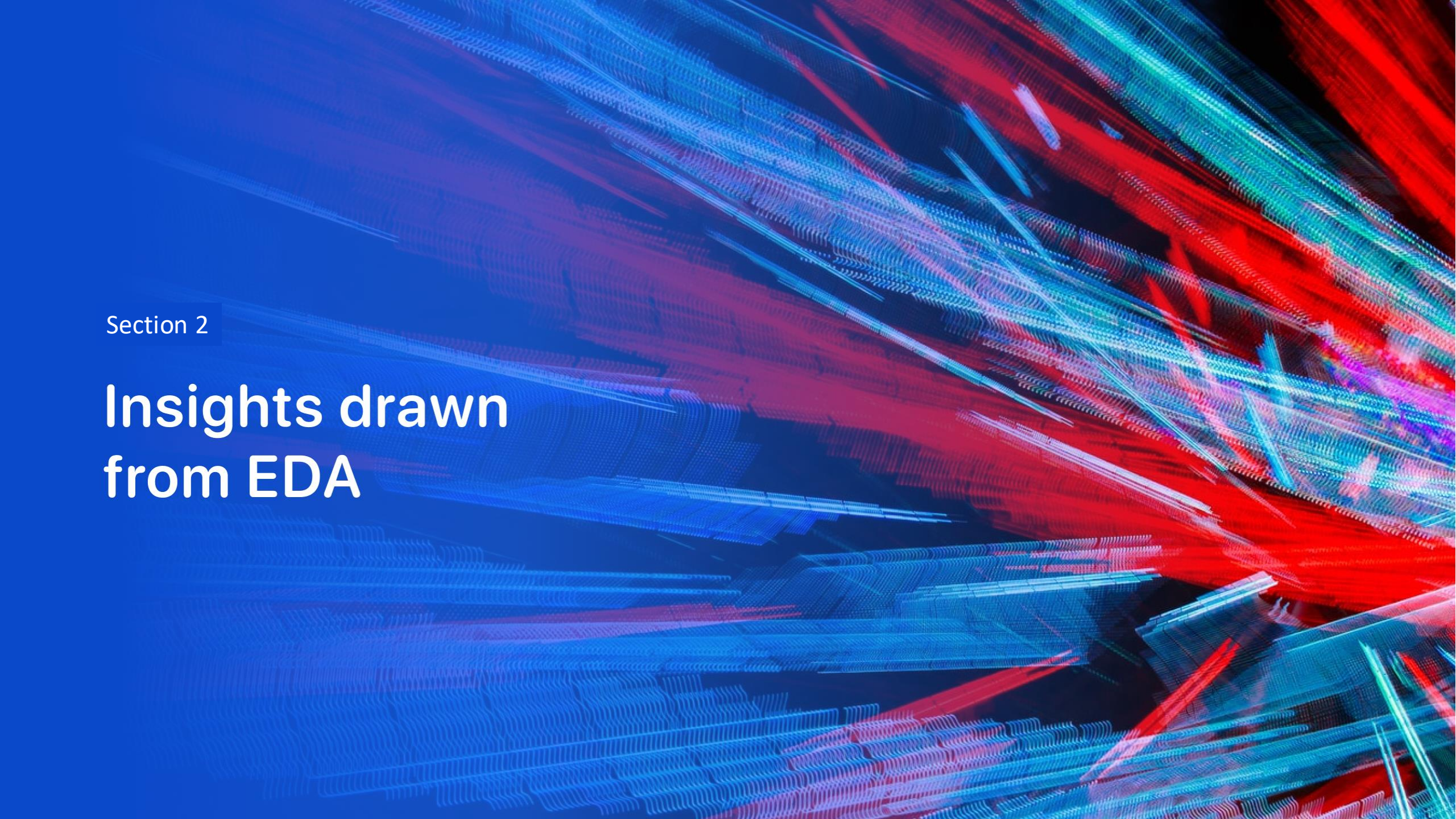
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

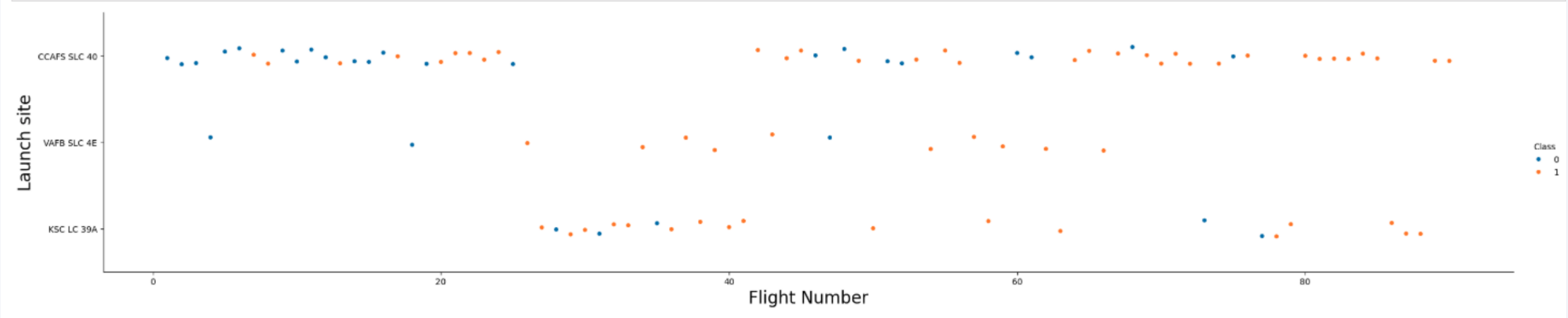
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

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

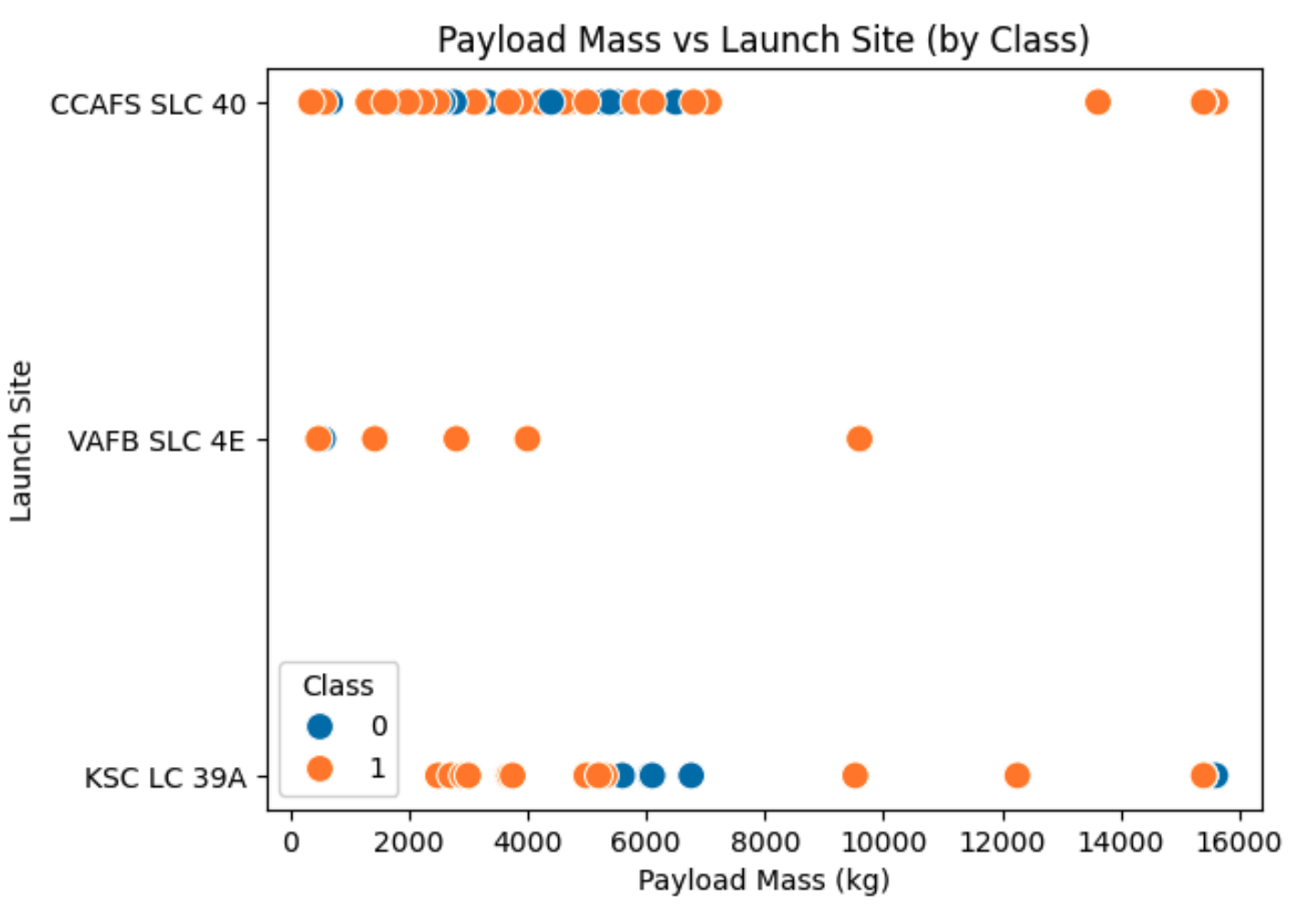

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



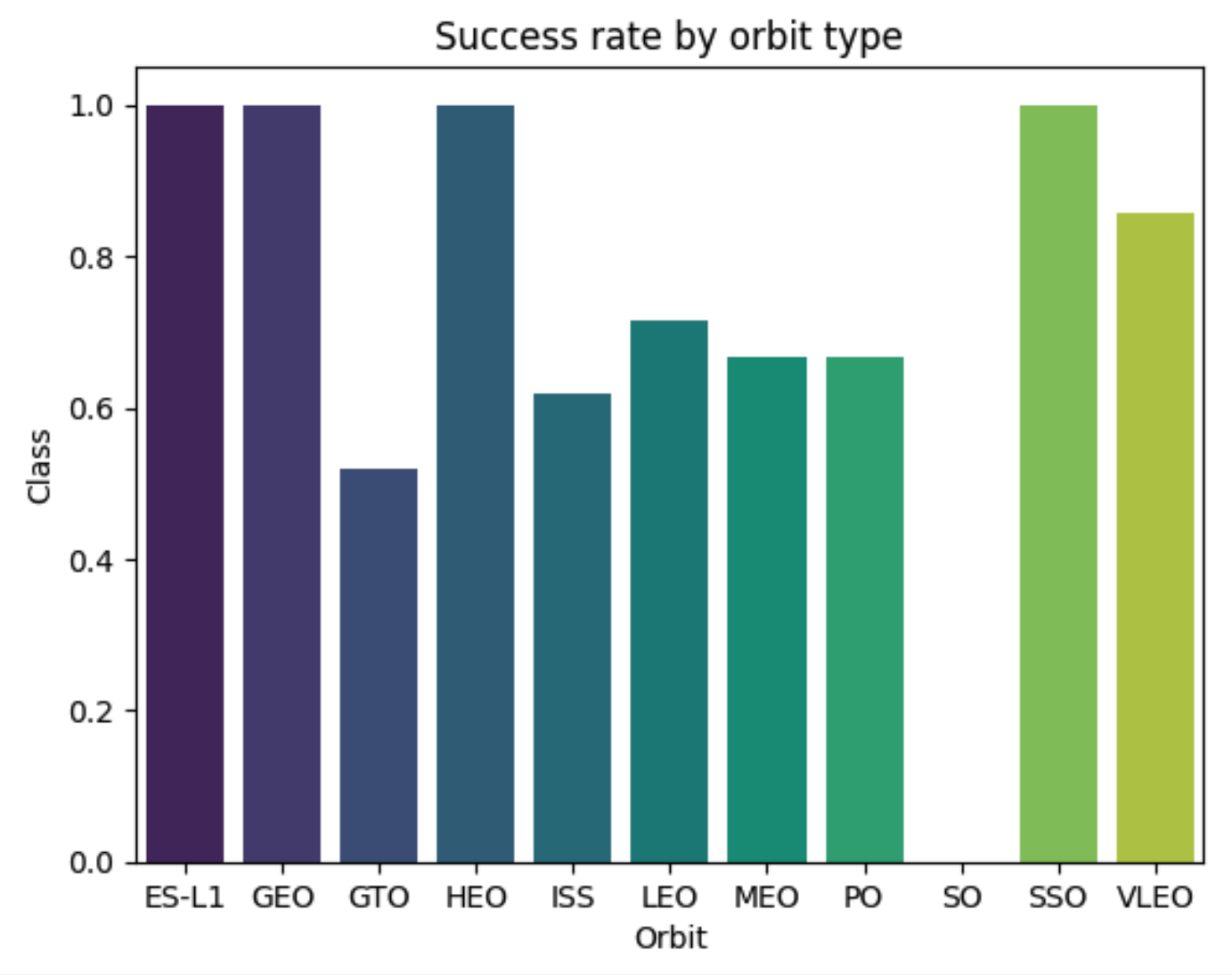
- Launch site CCAFS SLC 40 has the most flights launched compared to the other two sites. However, VAFB SLC 4E seems to be the site with the most success.

Payload vs. Launch Site



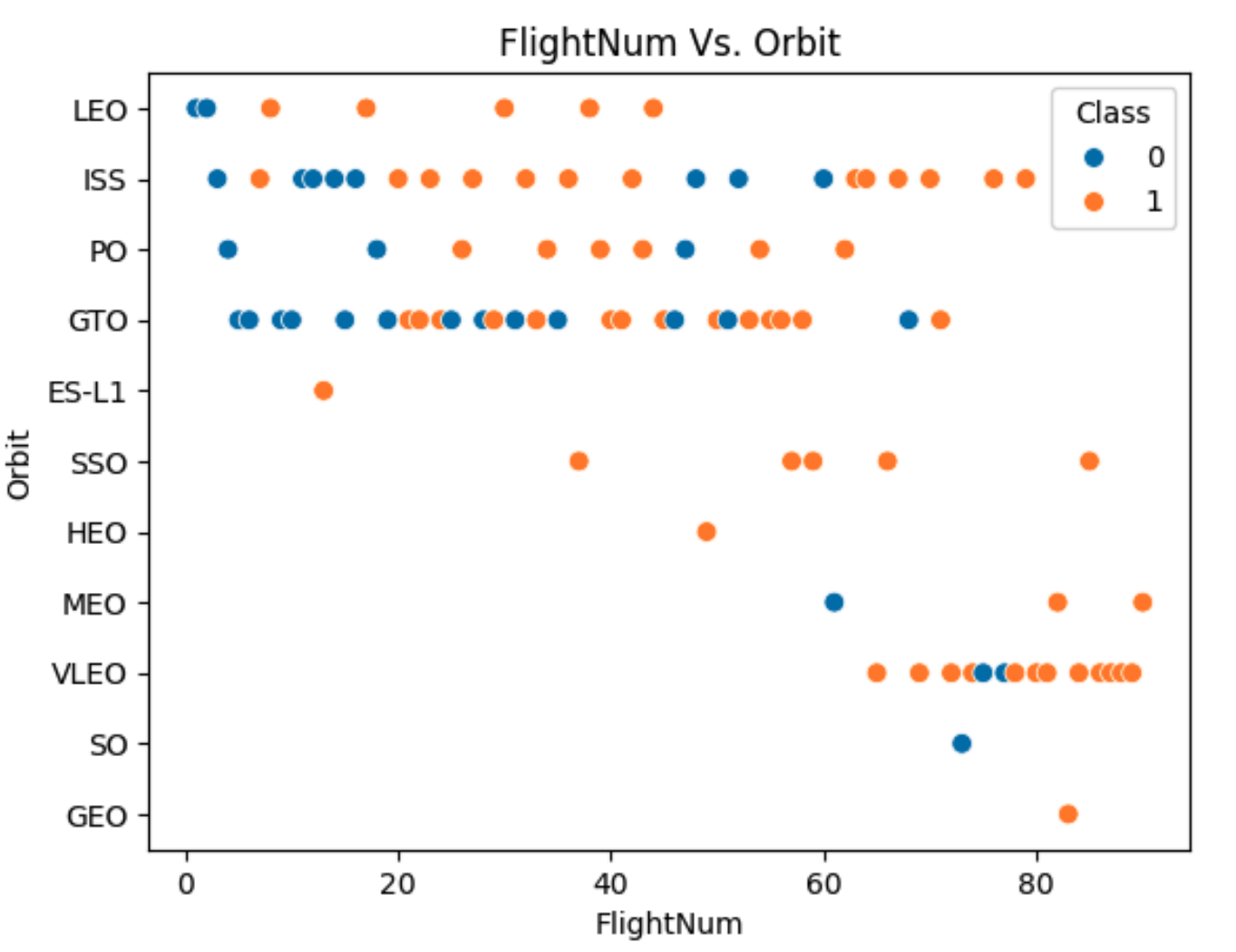
- VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

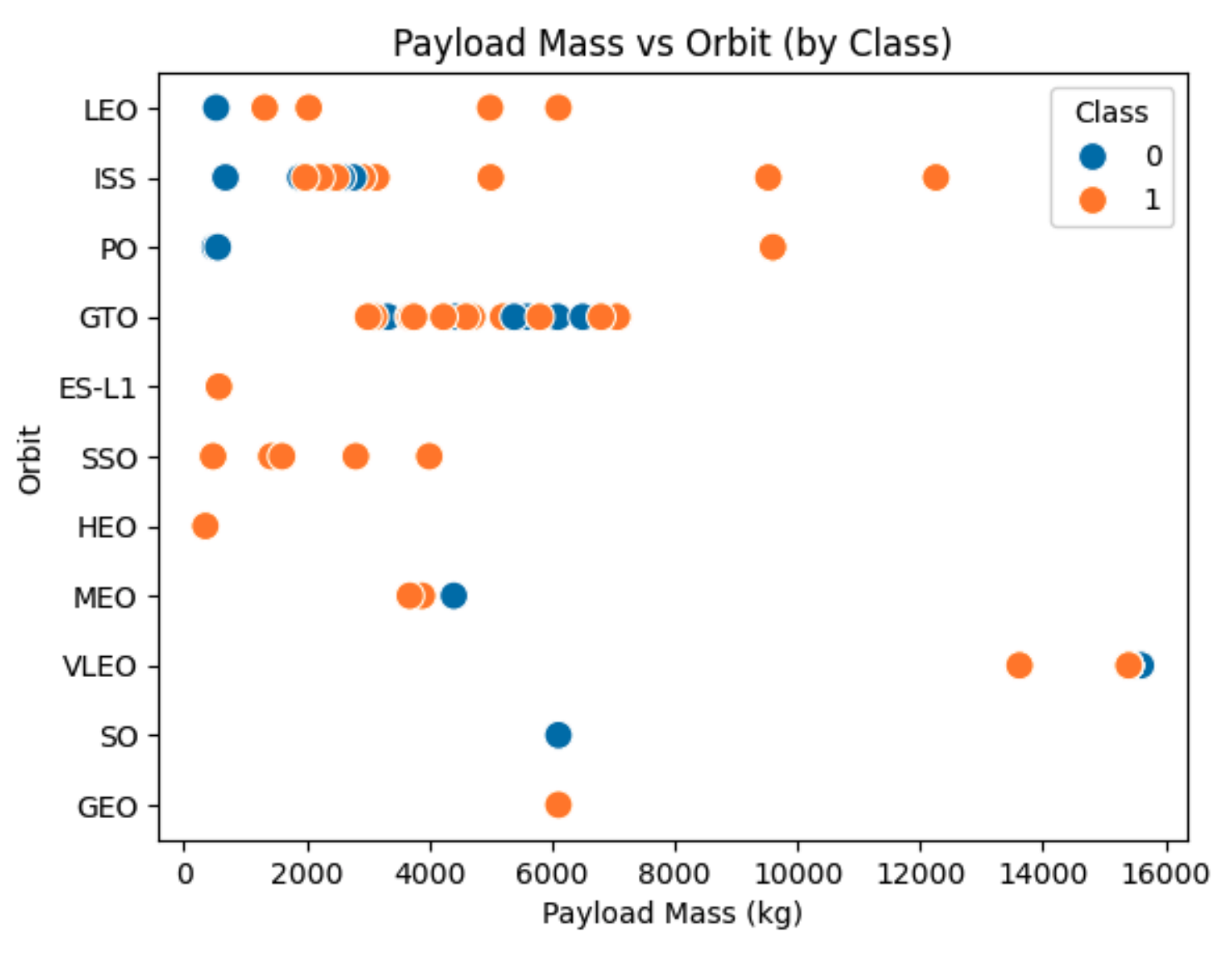


- We can clearly see orbit SO has zero success rate whereas, ES-L1, GEO, HEO, SSO has highest.

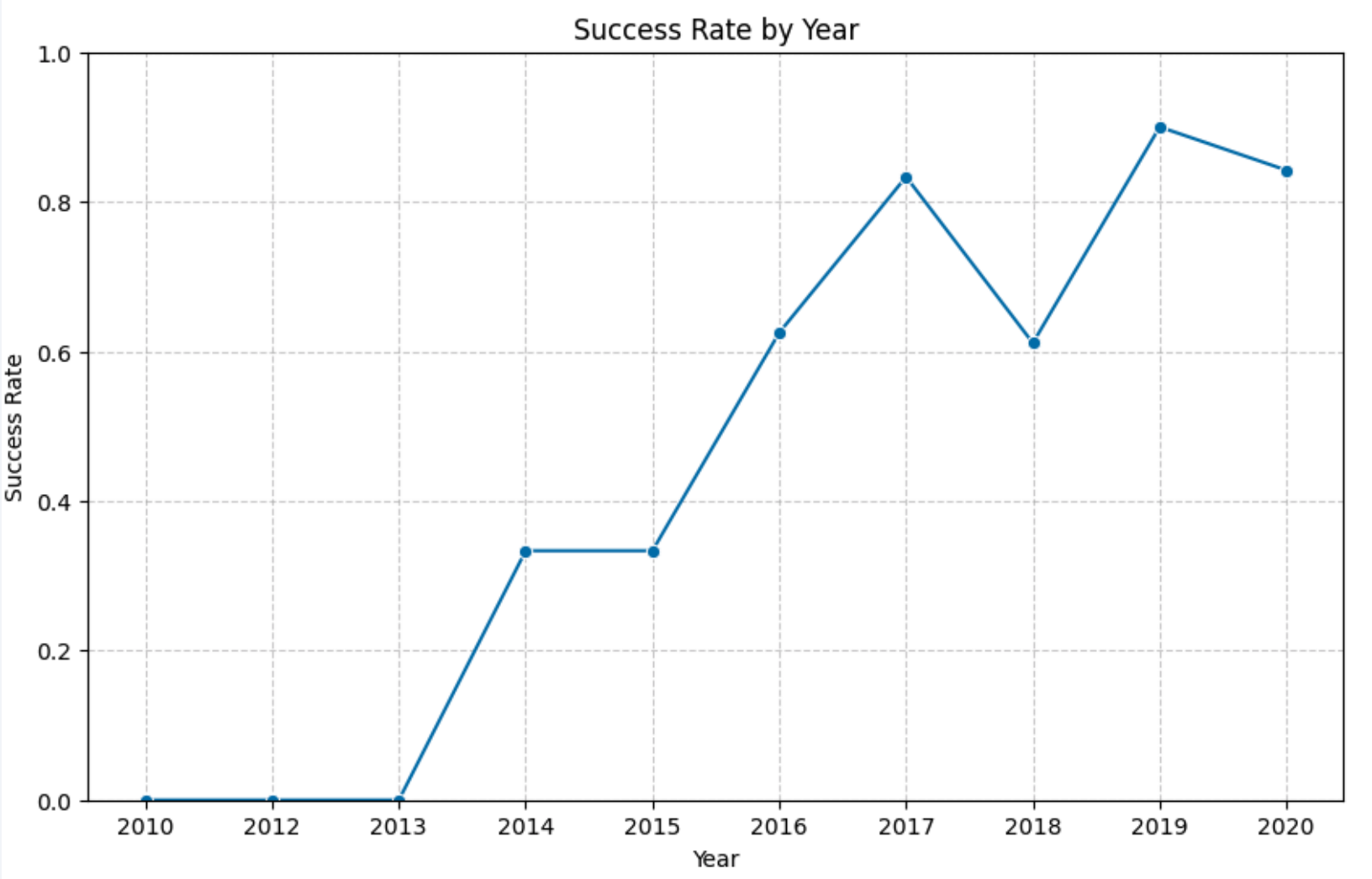
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Find the names of the unique launch sites

```
%sql select distinct(launch_site) from Spacextable
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%sql select * from Spacextable where launch_site like 'CCA%' limit 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql select SUM(Payload_mass__kg_) as 'Total payload mass' from Spacextable where Customer like 'NASA (CRS)%';
```

```
 * sqlite:///my_data1.db
Done.
```

Total payload mass
48213

Average Payload Mass by F9 v1.1

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

```
%sql select AVG(Payload_mass__kg_) as 'AVG payload mass' from Spacextable where Booster_Version like 'F9 v1.1%';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG payload mass
2534.6666666666665

First Successful Ground Landing Date

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

```
%sql select MIN(date) as 'First successful landing date' from Spacextable where Landing_Outcome like '%Success%';
```

```
 * sqlite:///my_data1.db
Done.
```

First successful landing date
2015-12-22

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

```
%sql select booster_version, payload_mass__kg_ from Spacextable where payload_mass__kg_ > 4000 and payload_mass__
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 v1.1	4535
F9 v1.1 B1011	4428
F9 v1.1 B1014	4159
F9 v1.1 B1016	4707
F9 FT B1020	5271
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1030	5600
F9 FT B1021.2	5300
F9 FT B1032.1	5300
F9 B4 B1040.1	4990
F9 FT B1031.2	5200
F9 B4 B1043.1	5000
F9 FT B1032.2	4230
F9 B4 B1040.2	5384
F9 B5 B1046.2	5800
F9 B5 B1047.2	5300
F9 B5B1054	4400

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT CASE WHEN Mission_Outcome LIKE 'Success%' THEN 'Success' ELSE 'Failure' END AS OutcomeType, COUNT(*)
```

```
 * sqlite:///my_data1.db
Done.
```

OutcomeType	Total
Failure	1
Success	100

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql select booster_version, payload_mass__kg_ from Spacextable where payload_mass__kg_ = (select max(payload_mass
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
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

2015 Launch Records

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

```
%sql SELECT substr(Date,6,2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SpacexTable WHERE Landi
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	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

```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) AS OutcomeCount FROM SpacexTable WHERE Date BETWEEN '2010-06-0
```

```
 * sqlite:///my_data1.db
Done.
```

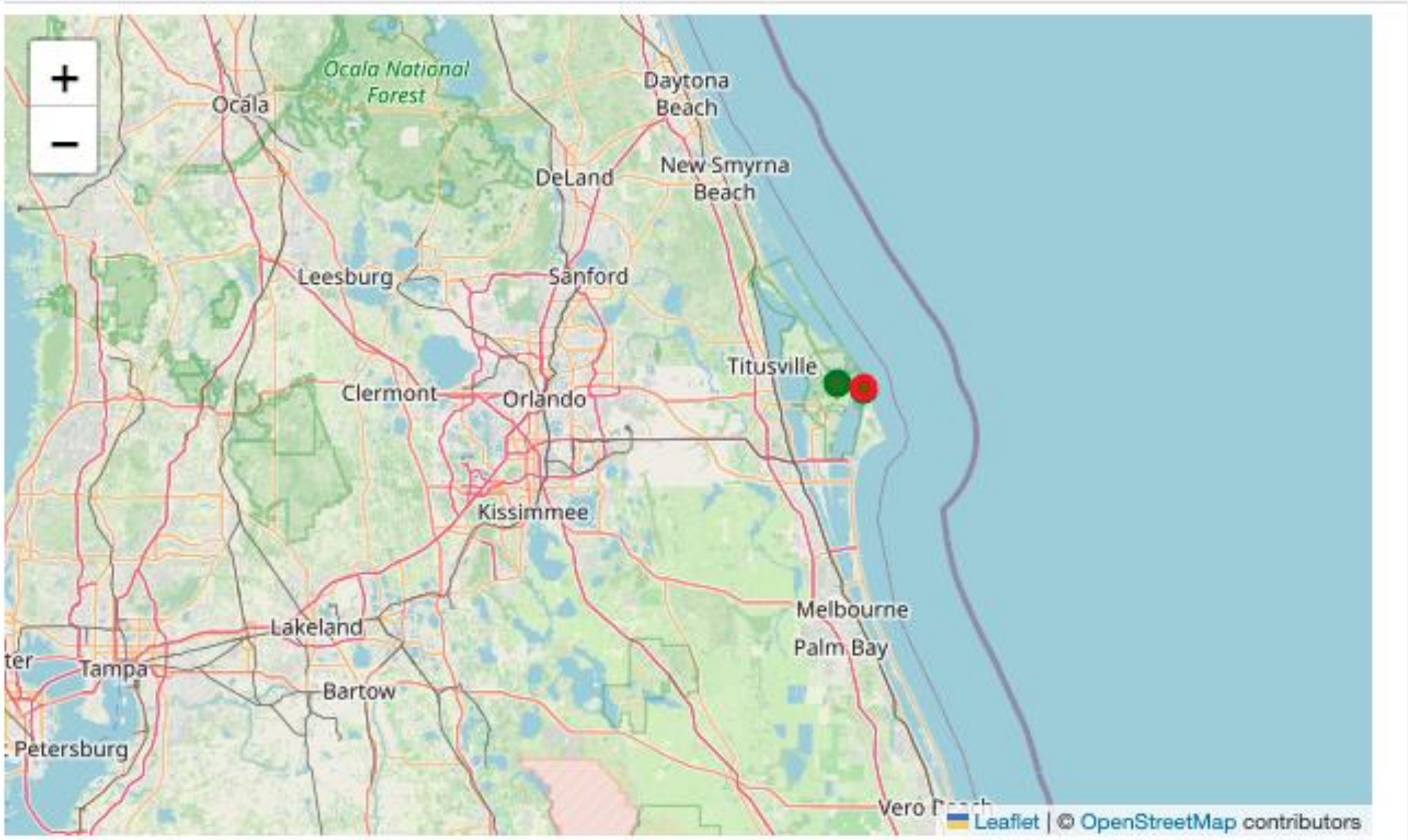
Landing_Outcome	OutcomeCount
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

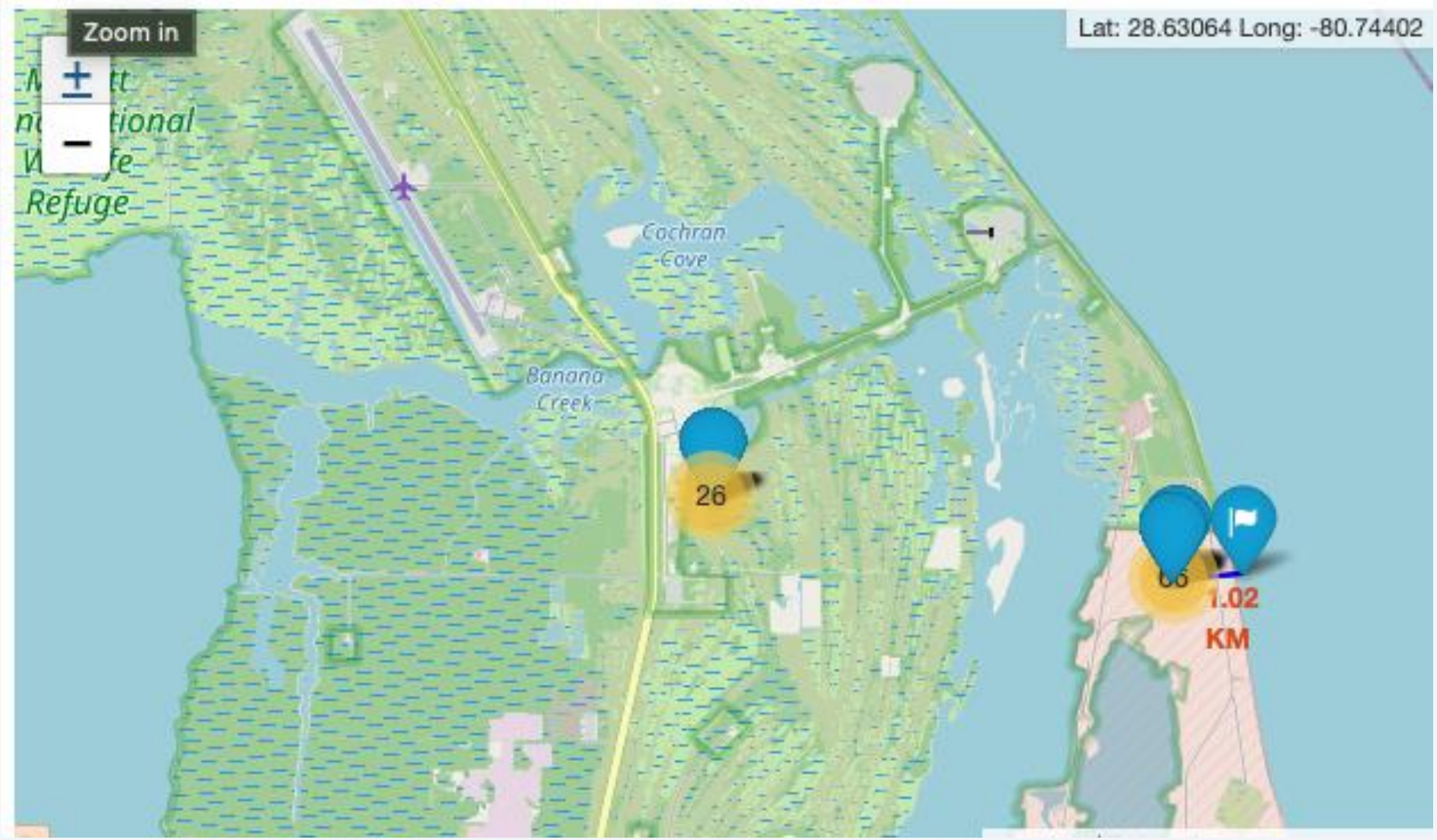
Launch Sites Proximities Analysis

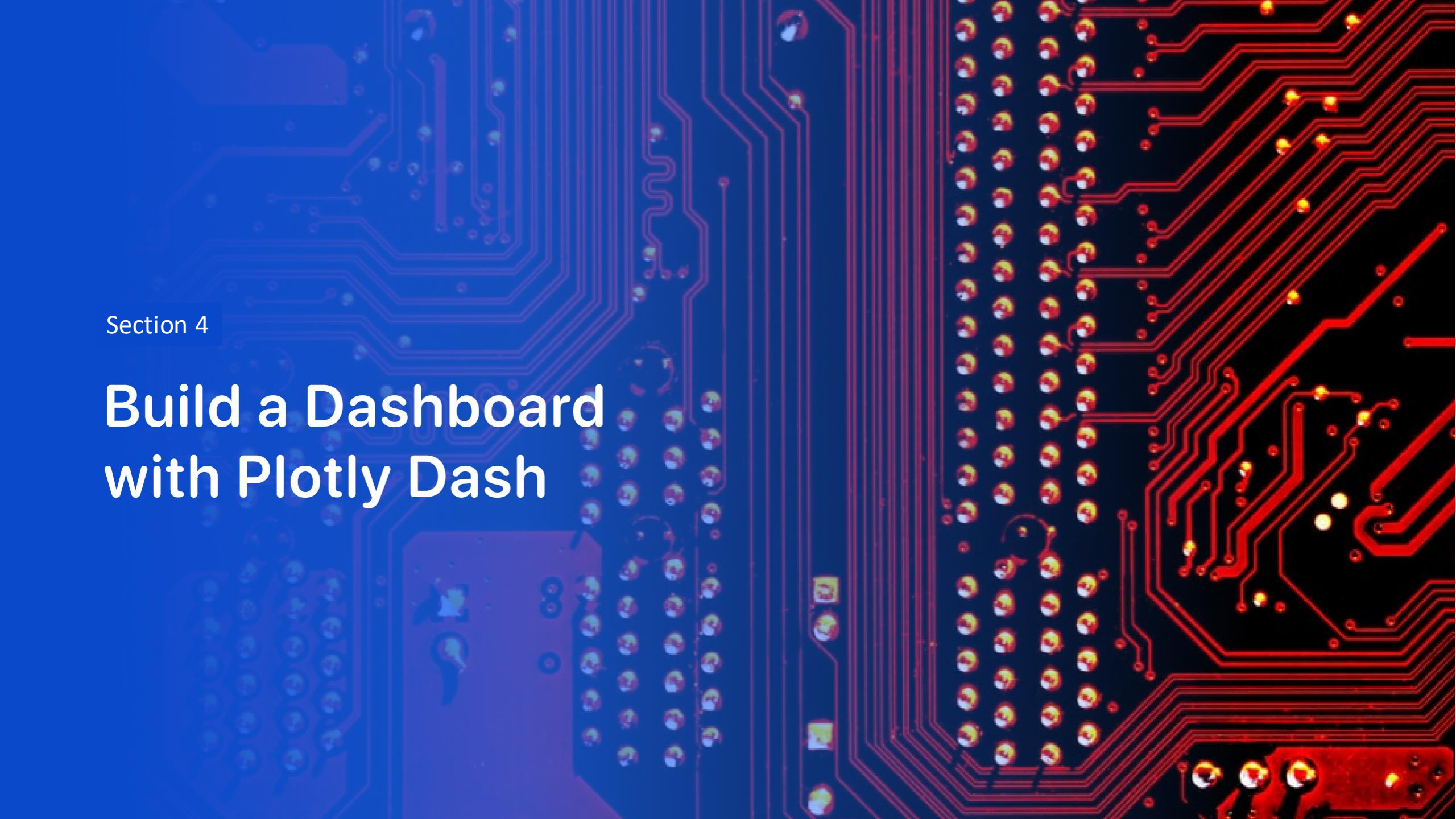
Map - all launch sites

Map – green for success launch, red for else.

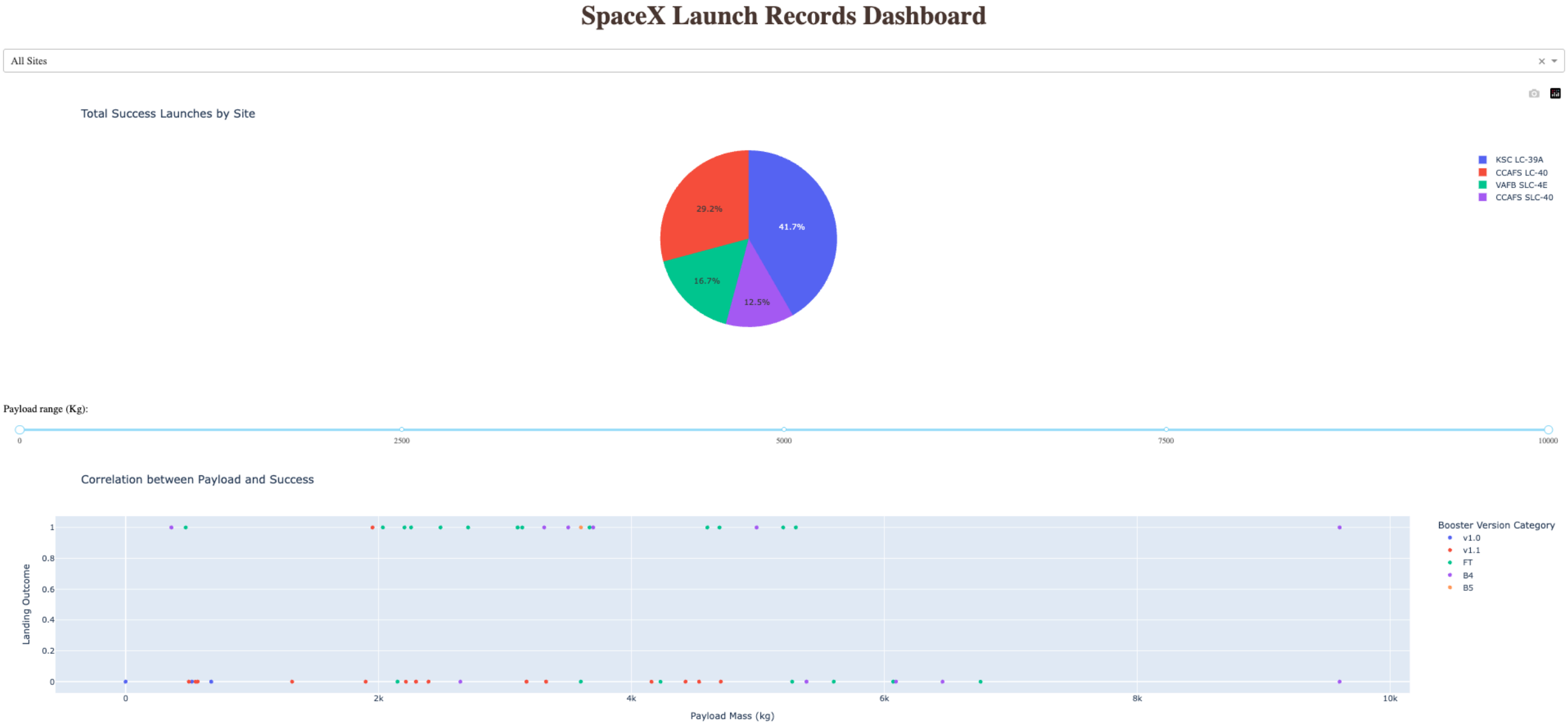


Map 3

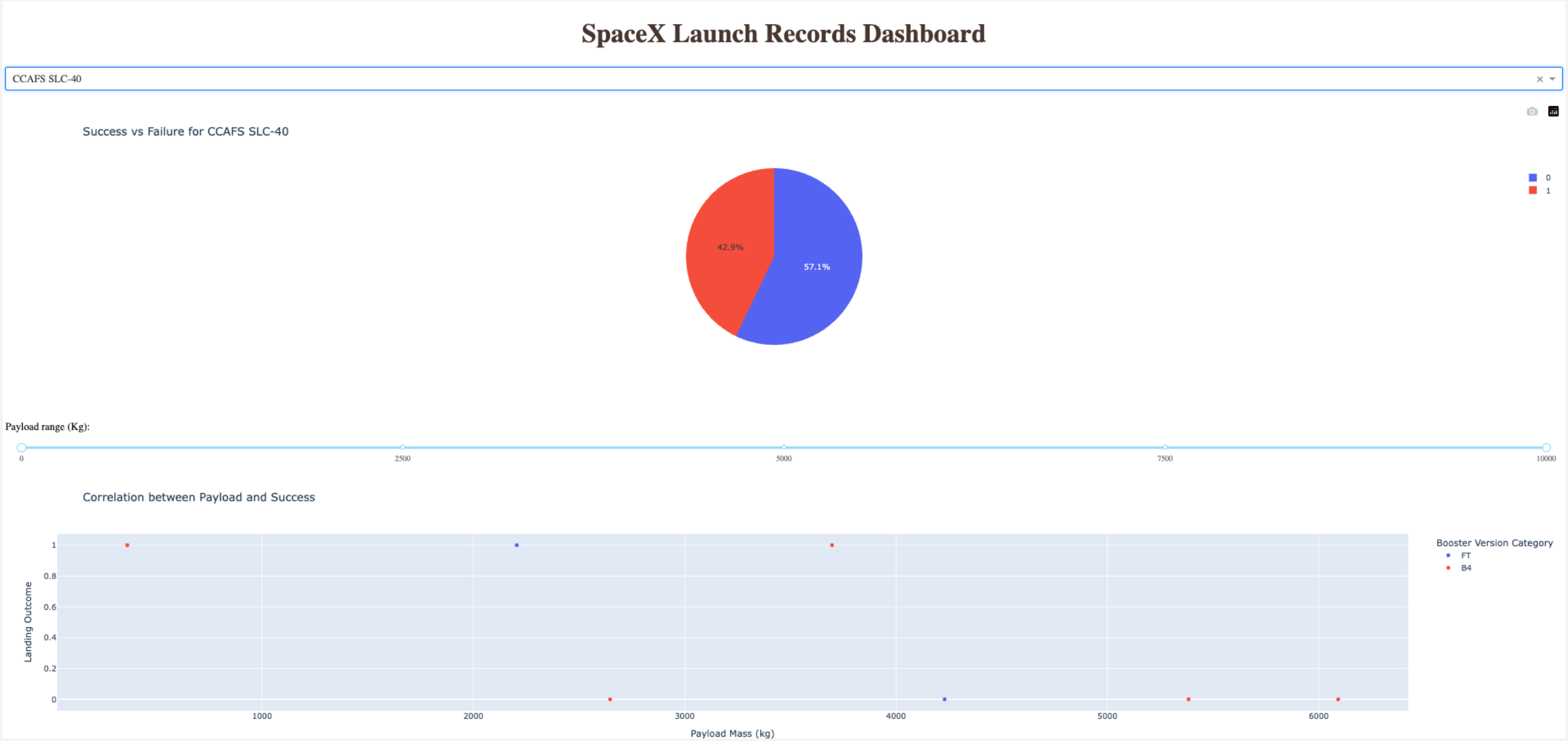

Zoom in
Lat: 28.63064 Long: -80.74402
Refuge
Cochran
Cove
Banana
Creek
26
1.02
KM


Section 4
Build a Dashboard with Plotly Dash

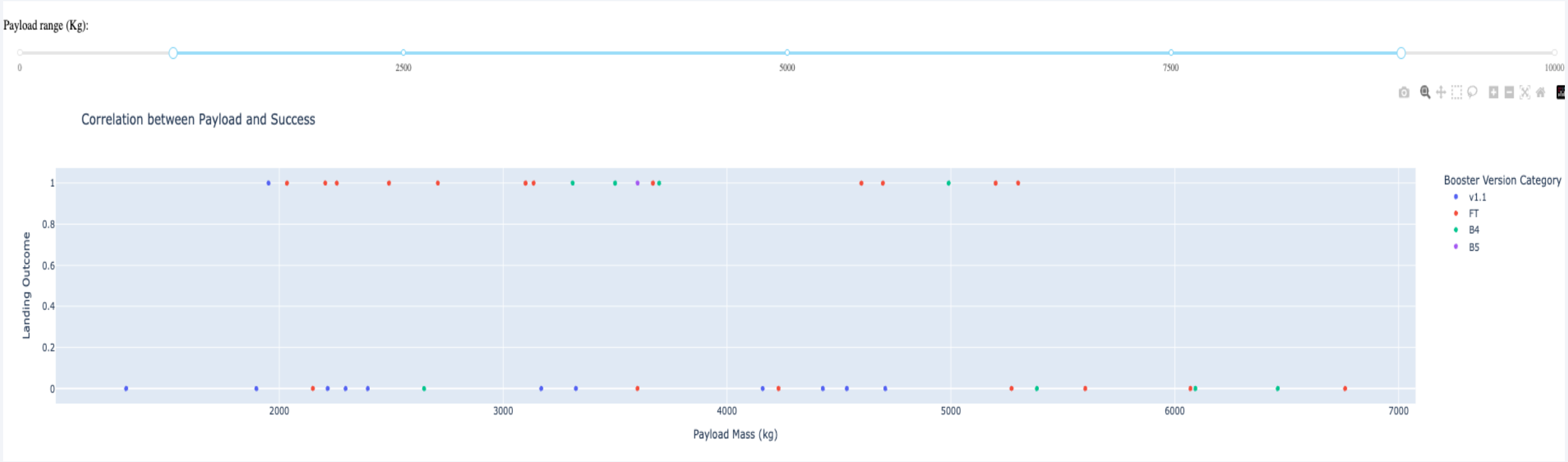
Dashboard - Total success launces by site



CCAFS SLC-40 site has highest success launch rate



Payload range vs. landing outcome

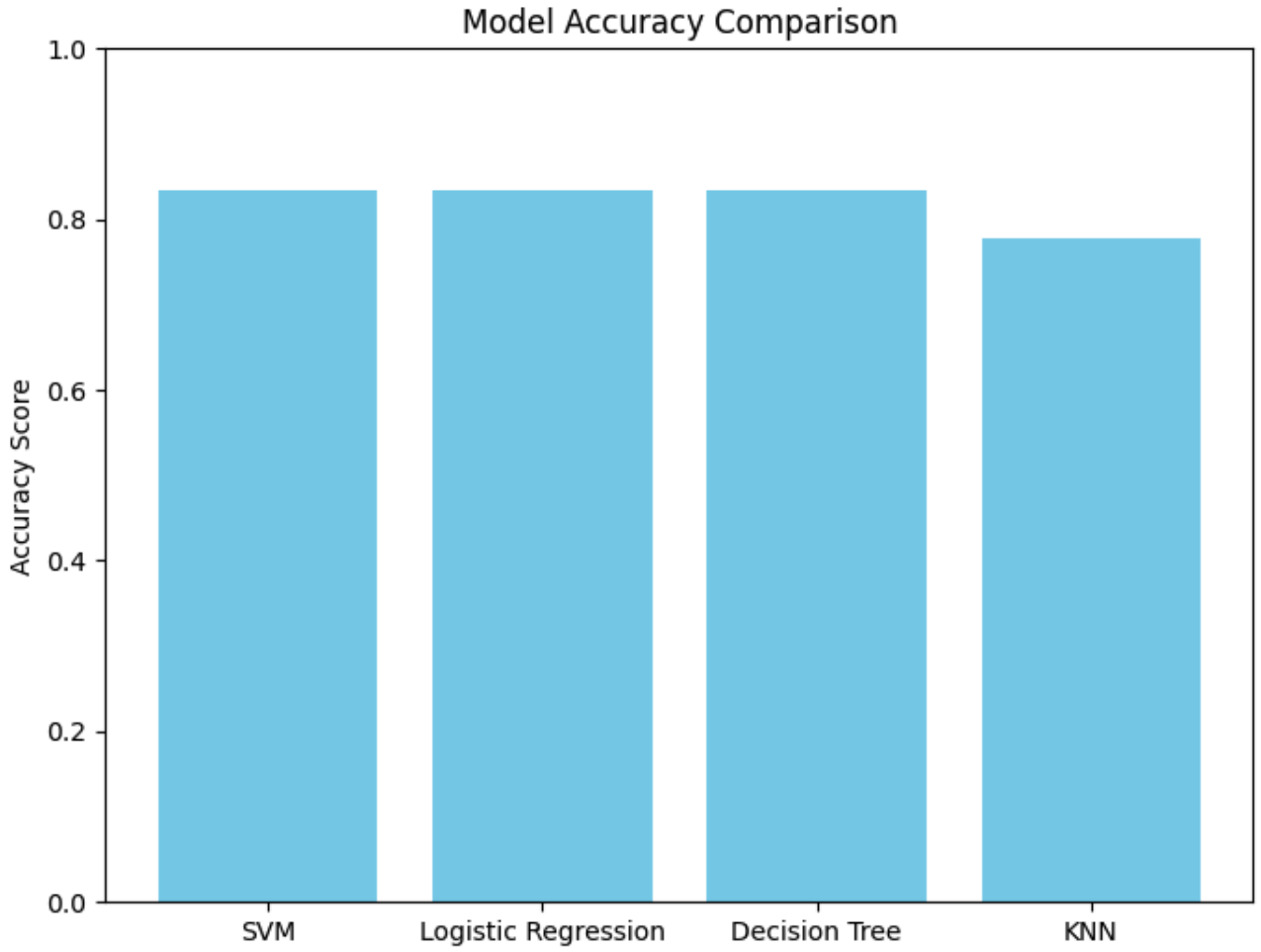


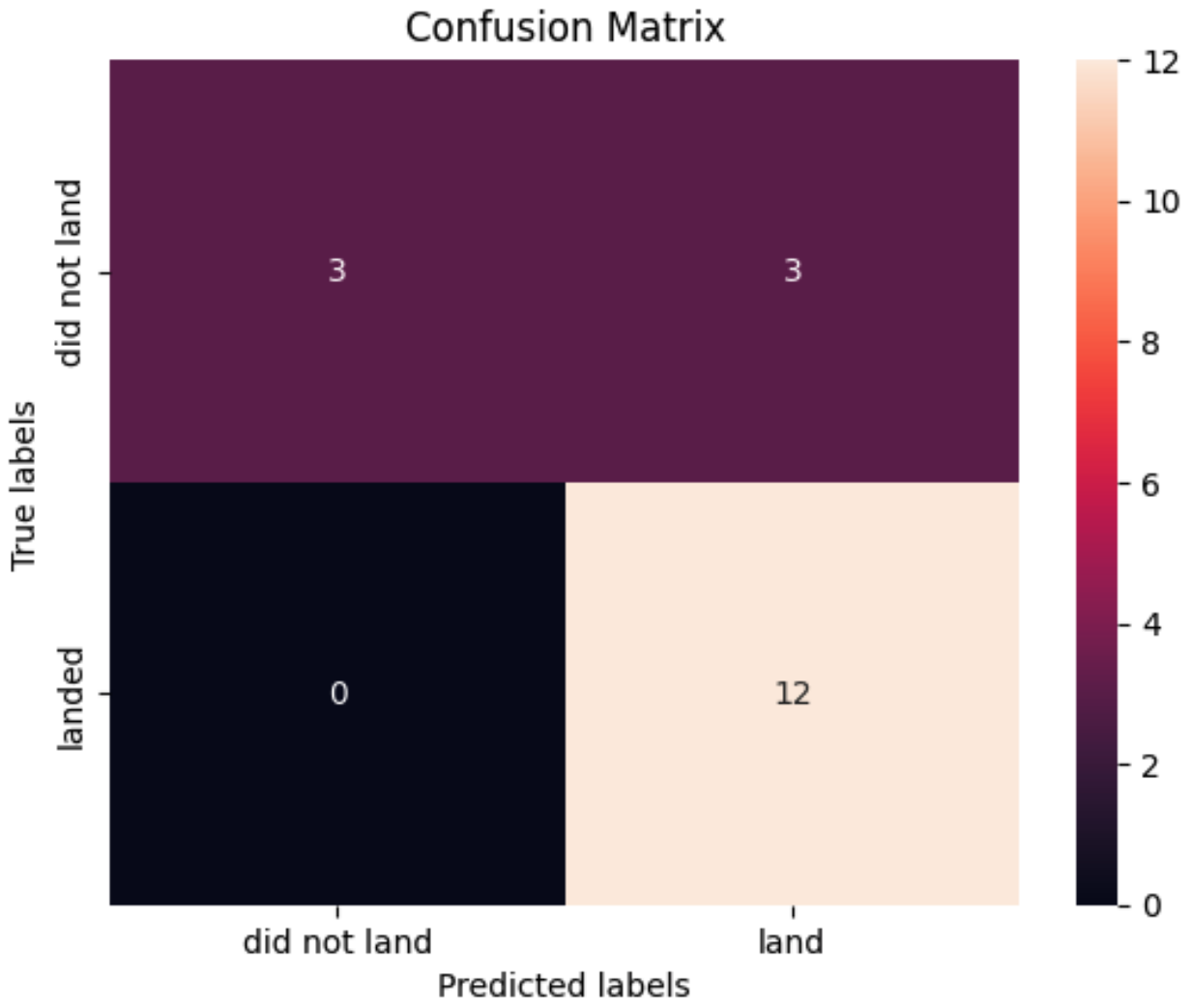
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- SVM, LR and DecisionTree have same accuracy score of 83%, whereas KNN is 78%
- Since all three best performed models have same accuracy, **LogisticRegression** would be best suited for our prediction
- For the reason as it's simple, interpretable (coefficients show how features affect probability), computationally efficient.





Confusion Matrix of LogisticRegression

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!