IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nelba Barreto
February 2024

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Interactive Visual Analytics
- Interactive Dashboard
- Predictive Analysis

Summary of all results

- Exploratory Data Analysis Results
- Interactive Analytics Demo in Screenshots
- Predictive Analysis Results

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, while other providers cost upward of 165 million dollars each. Much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

? Will the first stage land successfully?

? What variables influence the success or failure of a launch?

Section 1

Methodology

Methodology

- Data collection methodology:
  - Data was obtained from the SpaceX REST API and by scraping the Wikipedia's List of Falcon 9 and Falcon Heavy launches page
- Perform data wrangling
  - Data preparation involved encoding categorical variables into numerical representations, addressing missing data, and applying data transformations for enhanced 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
  - After splitting the data into training and test sets, various models were employed to predict the success (1) or failure (0) of the Falcon 9 first stage landing. Optimal hyperparameters were selected using GridSearchCV and evaluation metrics were then applied to assess the performance of each model

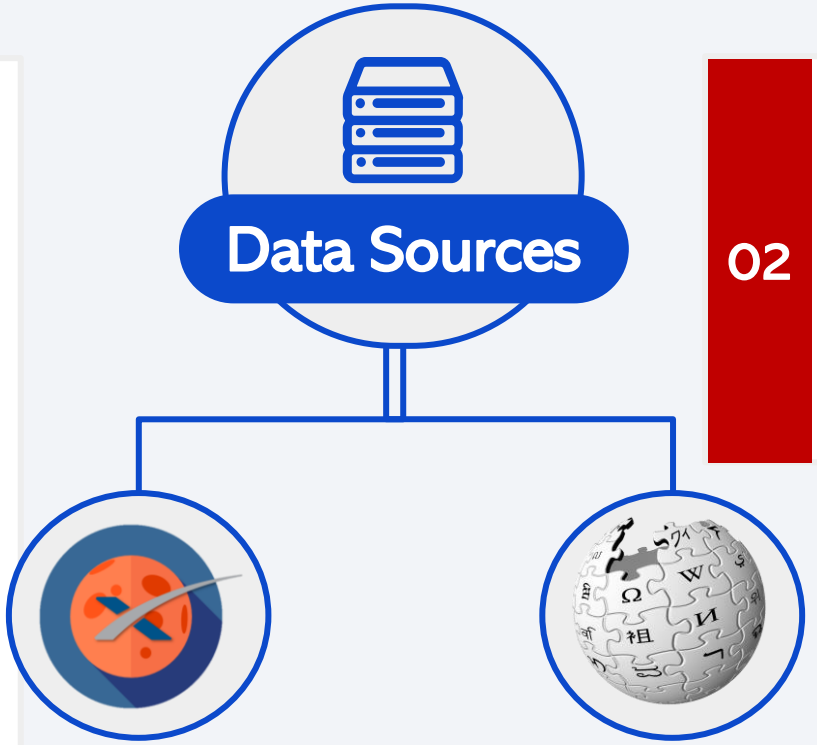
Data Collection

01

https://api.spacexdata.com/v4/launches/past

Additional information was extracted using nested API calls:

- **Rocket**: Booster name
- **Launchpad**: Launch site, longitude, latitude
- **Payload**: Mass, orbit
- **Cores**: Landing outcome, type, flight count, gridfins, reuse, legs, landing pad, block, reuse count, serial



SpaceX REST API

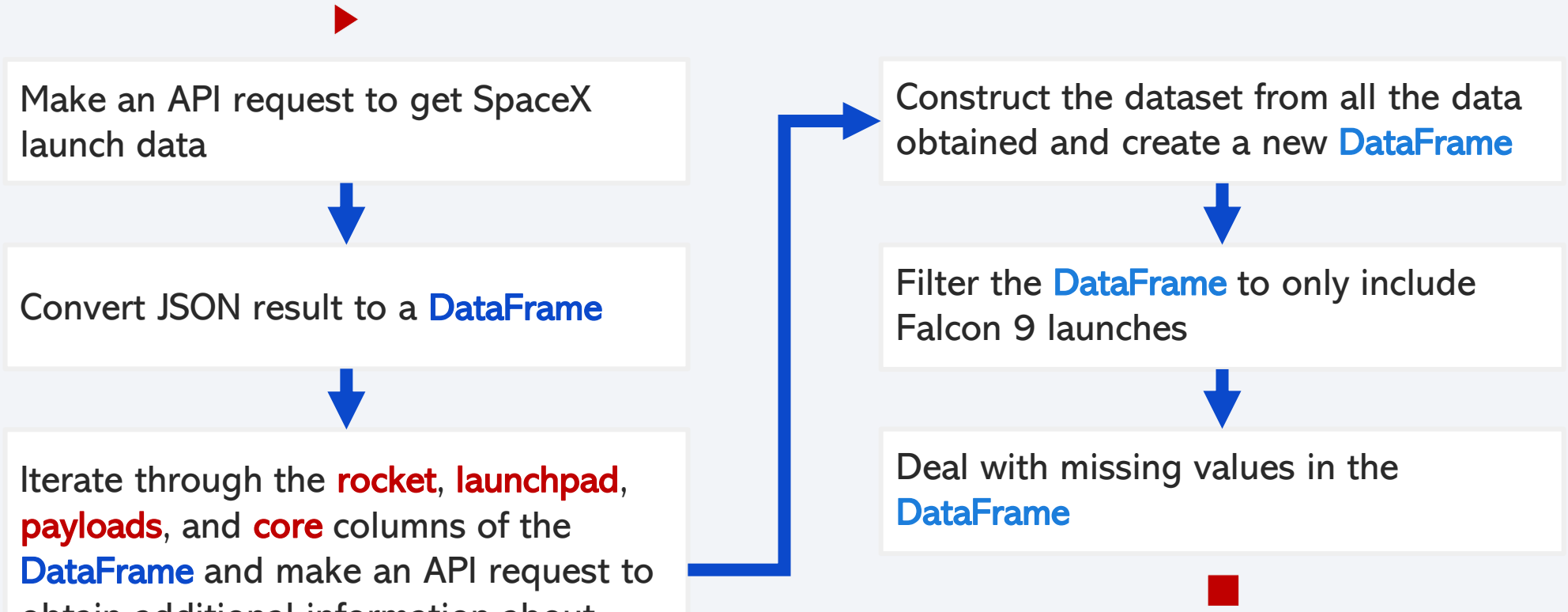
Wikipedia's List of Falcon 9 and Falcon Heavy launches page

02

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

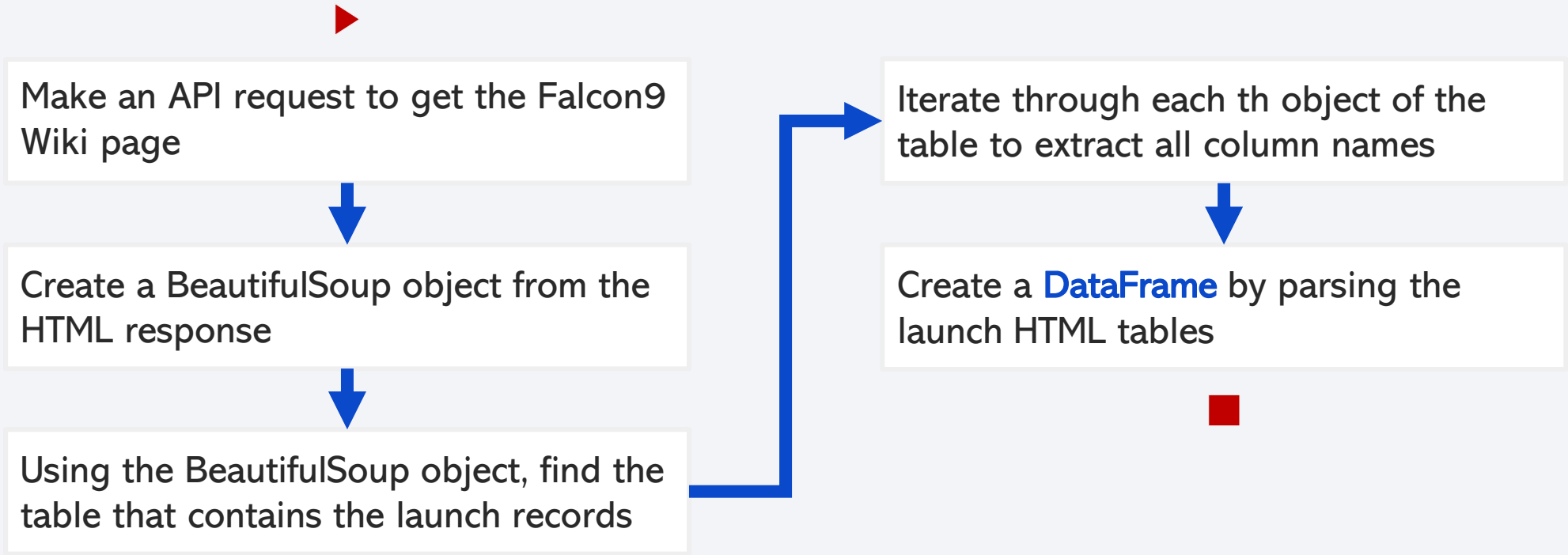
Web scraping was performed to extract an HTML table containing Falcon 9 launch records

Data Collection – SpaceX API



https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/2_jupyter-labs-webscraping.ipynb

Data Wrangling

Perform some Exploratory Data Analysis (EDA) on the dataset

Determine the **number of launches on each site**, the **number** and **occurrence of each orbit**, and the **number of landing outcomes**

Create a **Class** column using the **Outcome** column

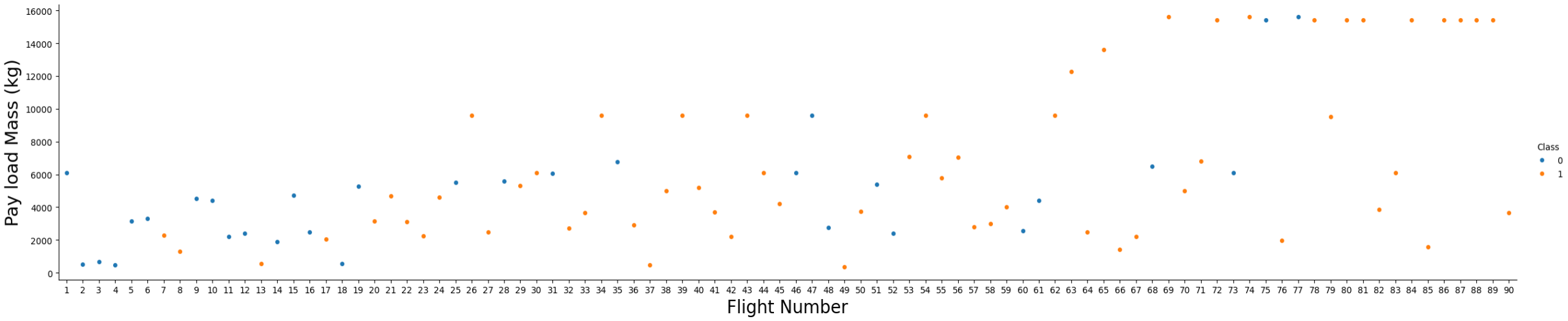
Landing Outcomes	Landing Class	Meaning
True ASDS; True RTLS; True Ocean	1	Success
None None; False ASDS; False Ocean; None ASDS; False RTLS	0	Failure

https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

01 **Scatter plots** were used to examine the relationship between two continuous variables

FlightNumber vs PayloadMass



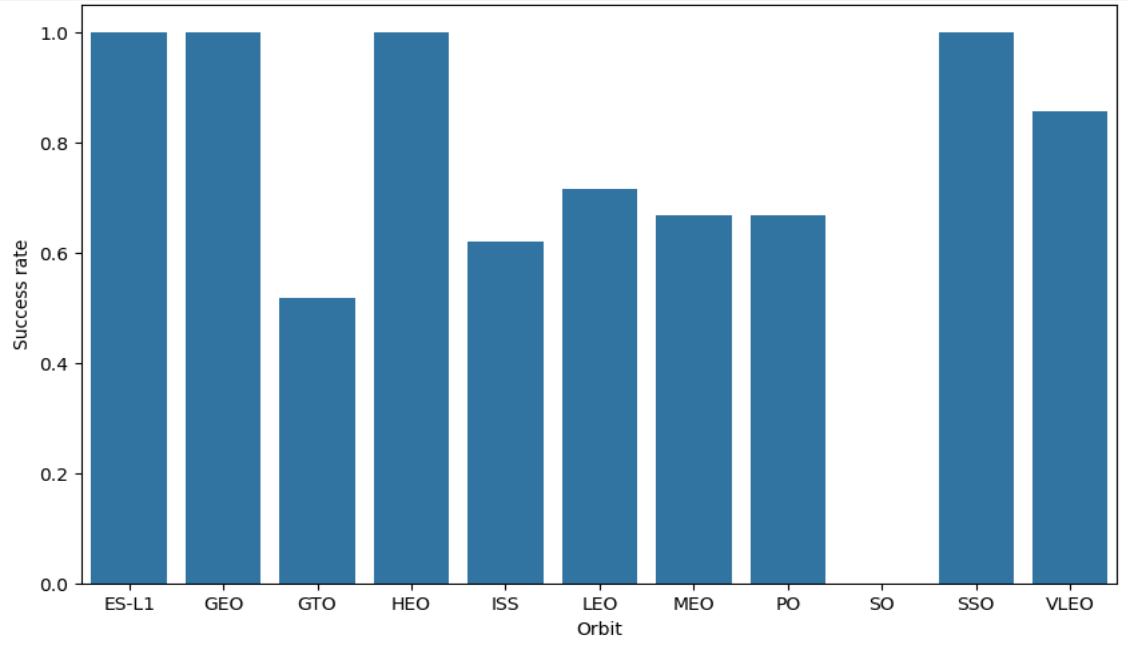
https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/5_jupyter-labs-eda-dataviz.ipynb

EDA with Data Visualization

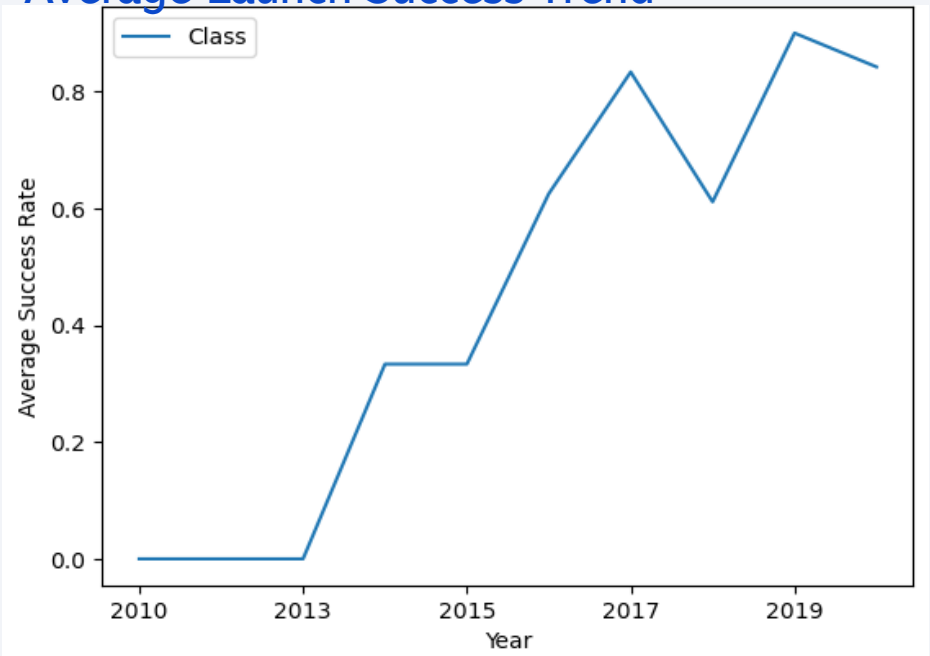
02 A **bar chart** was used to make a comparison between categories

03 A **line chart** was used to track the average success rate over time

Success Rate of Each Orbit Type



Average Launch Success Trend



https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/5_jupyter-labs-eda-dataviz.ipynb

EDA with SQL

The following information was obtained through SQL queries:

- The names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string ‘CCA’
- The total payload mass carried by boosters launched by Nasa (CRS)
- The average payload mass carried by booster version F9 v1.1
- The month names, failure landing_outcomes in drone ship, booster versions and launch_site for the months in year 2015
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- The total number of successful and failure mission outcomes
- The names of the booster_versions which have carried the maximum payload mass
- The date when the first successful landing outcome in group pad was achieved
- The ranking of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

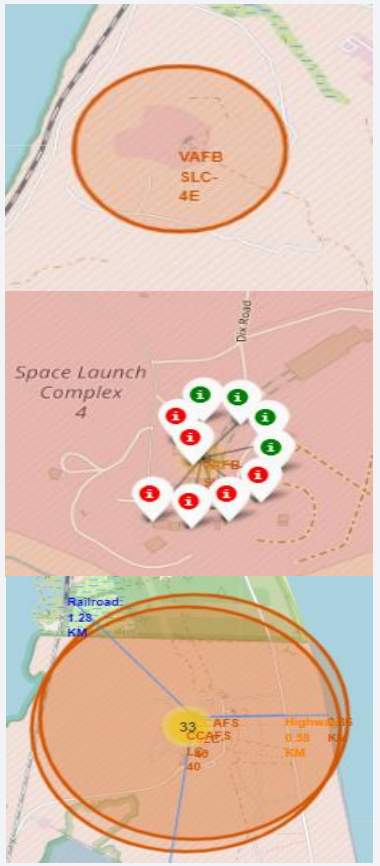
Folium was employed to uncover any discernible geographical patterns related to launch sites

Markers were used to add a marker on specific coordinates, representing each launch site and each launch

Markers clusters were used to group markers in each launching site

Circles were used to add a highlighted circle area on specific coordinates

Lines were used to show the distances between a launch site to its proximities



https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

The dashboard includes the following plots, graphs, and interactive features:

01 A **drop-down menu** for selecting launch sites

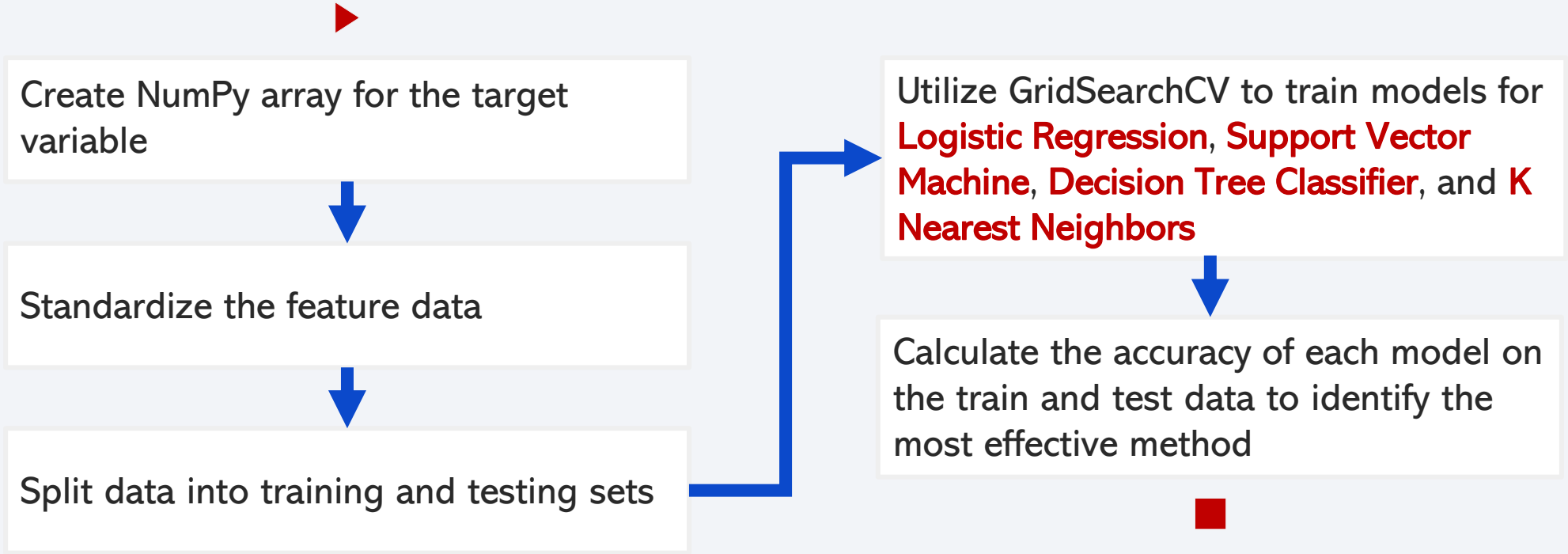
02 A **callback function that renders a pie chart**, visualizing the counts of launch successes based on the selected launch site from the dropdown

03 A **range slider** for selecting payload parameters

04 A **callback function that renders a scatter plot** depicting **Payload vs. Launch Outcome**, based on the selected payload parameters

https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/7_spacex_dash_app.py

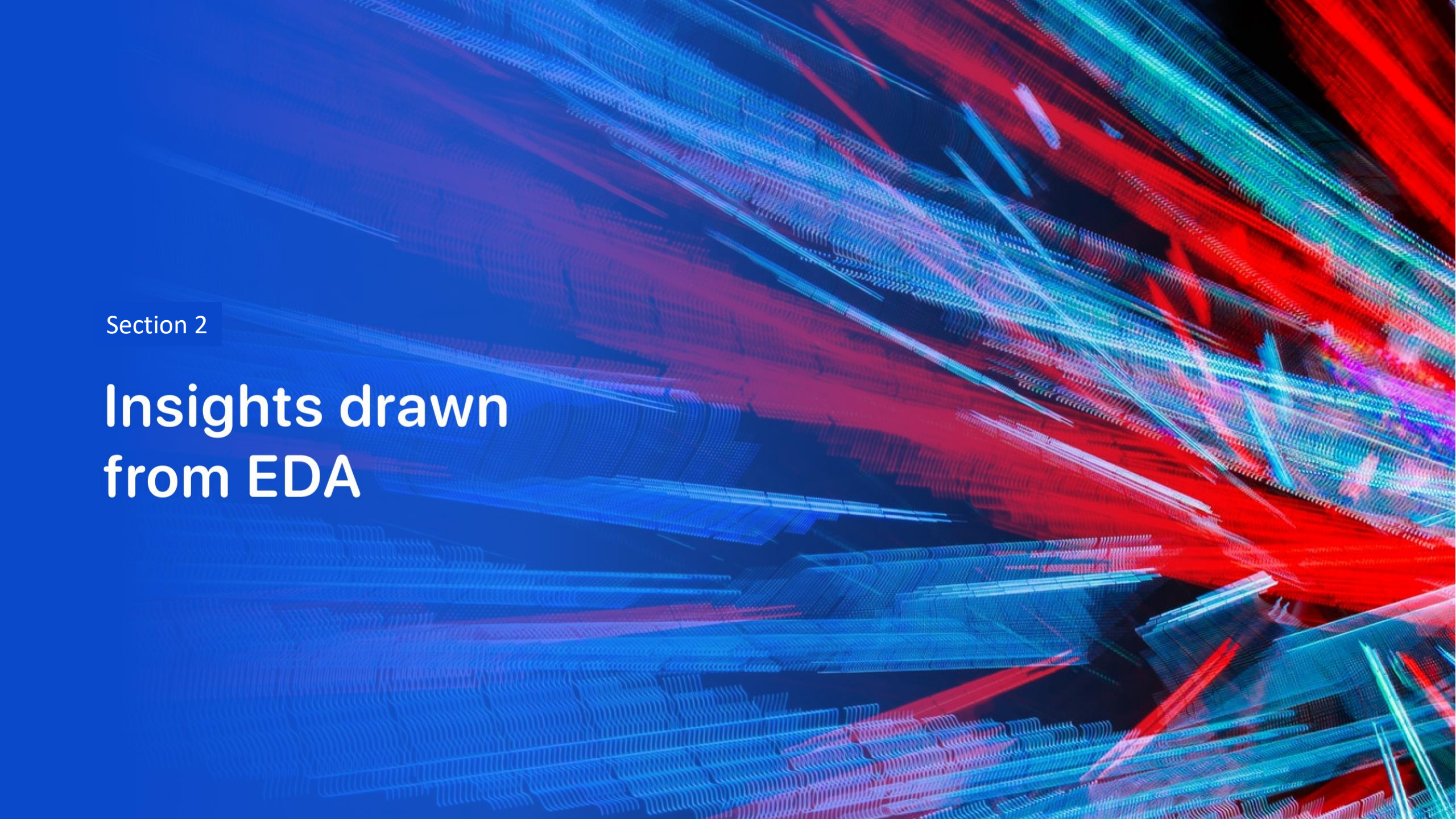
Predictive Analysis (Classification)



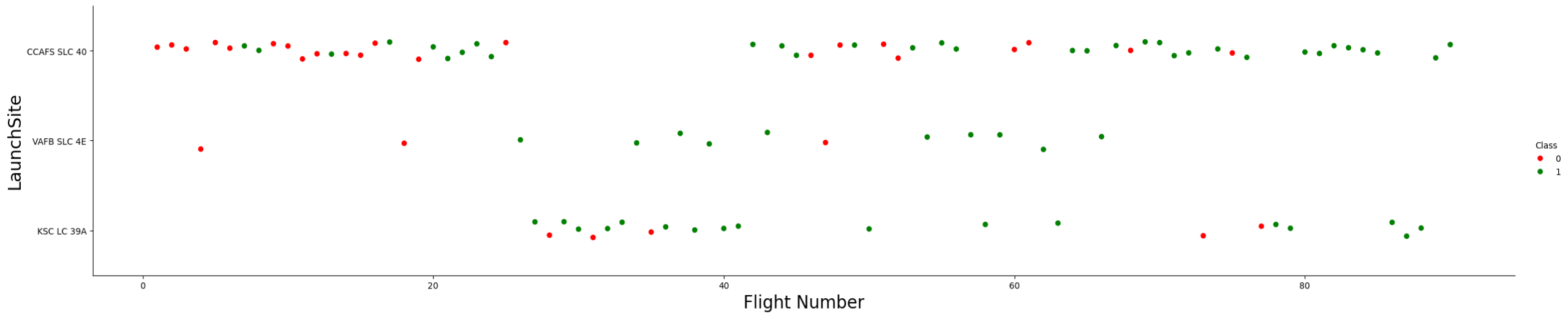
https://github.com/NelbaBarreto/applied-data-science-capstone/blob/main/8_SpaceX_Machine_Learning_Prediction_Part_5.ipynb

Results

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

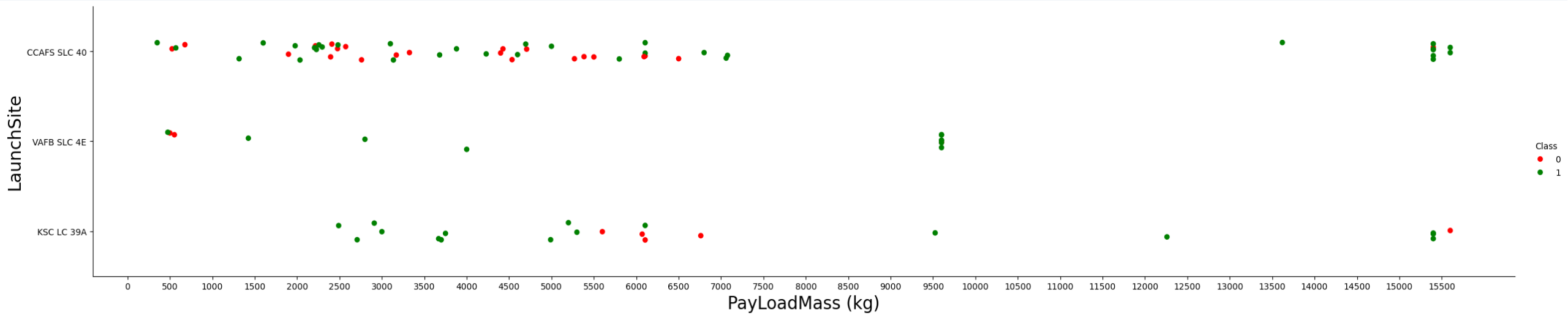

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



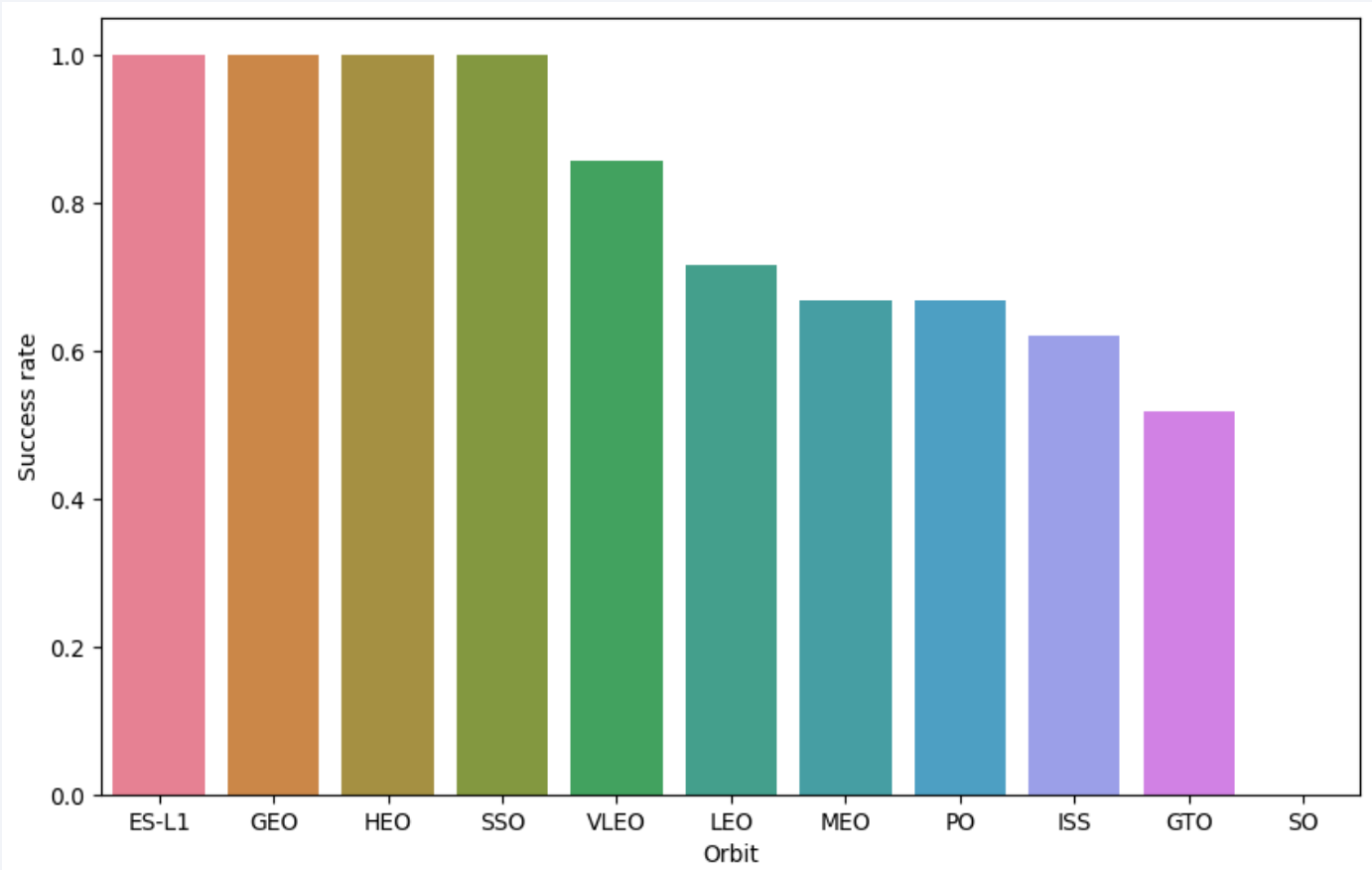
- The launch site with the **highest** number of launches is **CCAFS SLC 40**
- The launch site with the **least** number of launches is **VAFB SLC 4E**
- As **the flight number increases**, the **likelihood of success also increases**

Payload vs. Launch Site



- For the **VAFB-SLC** launch site there are no rockets launched for heavy payload mass (greater than 10000 kg)
- **KSC LC 39A** has a 100% success rate for payload mass under 5500 kg
- The **likelihood of success** is very high for heavier payload mass (>9000 kg)

Success Rate vs. Orbit Type



01 **ES-L1 (Lagrange Point 1):** A point between the sun and the Earth where gravitational forces cancel out, allowing small objects to remain in equilibrium.

02 **GEO (Geosynchronous Orbit):** A circular orbit 35,786 kilometers above Earth's equator, following the direction of Earth's rotation.

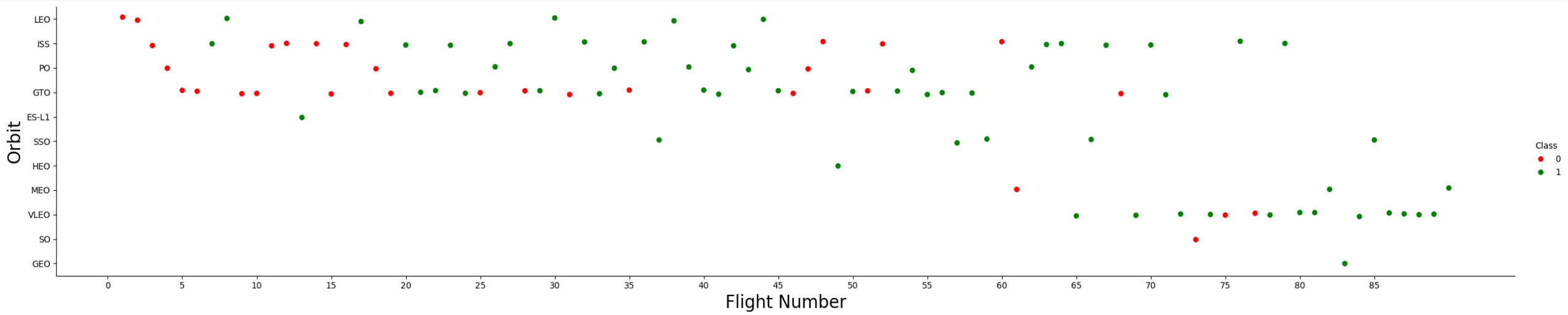
03 **HEO (Highly Elliptical Orbit):** An elliptic orbit with high eccentricity, usually around Earth.

04 **SSO (Sun-synchronous Orbit or SO):** A nearly polar orbit around a planet, ensuring the satellite passes over any given point at the same local mean solar time.

ES-L1, **GEO**, **HEO** and **SSO** have a 100% success rate

SO has a 0% success rate

Flight Number vs. Orbit Type



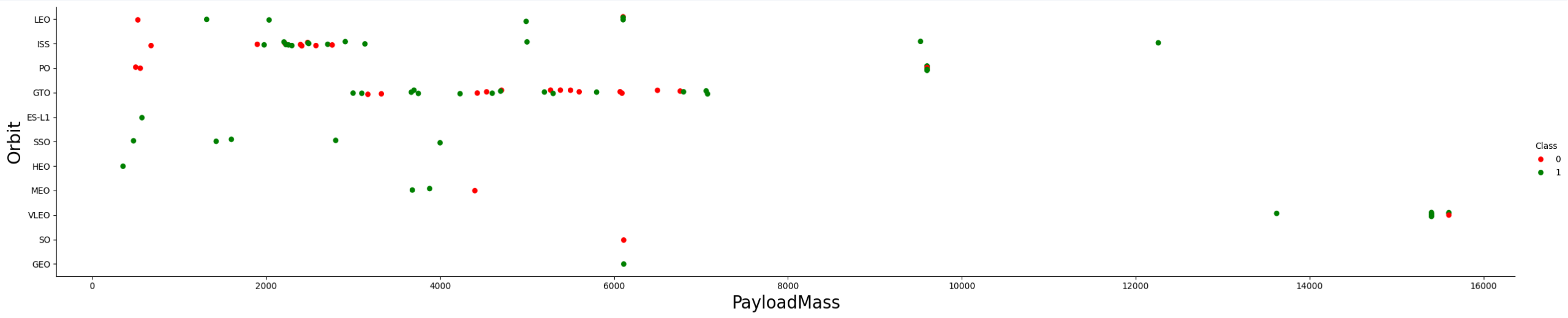
In the LEO orbit the Success appears related to the number of flights

On the other hand, there doesn't seem to be a clear increase or decrease in success rate with increasing flight number

01

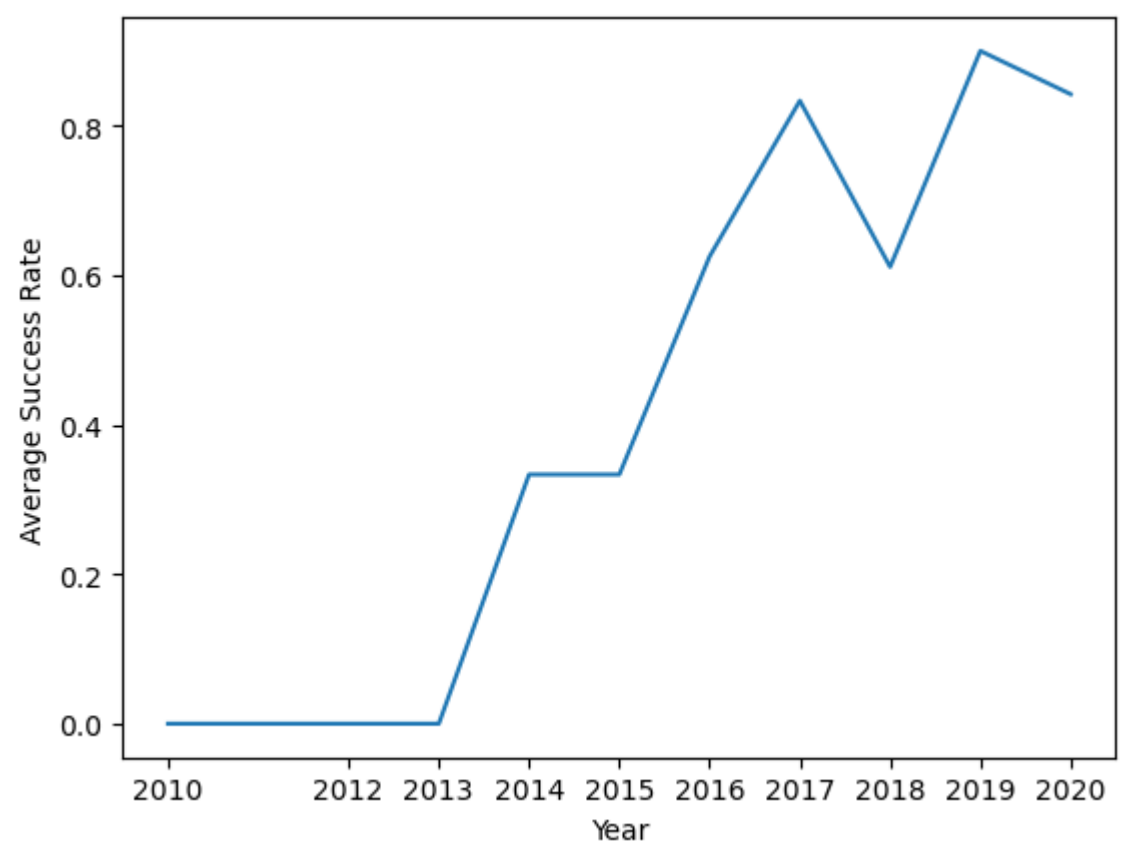
LEO (Low Earth Orbit): This orbit is at an altitude of 2,000 km or less, containing most manmade objects in outer space.

Payload vs. Orbit Type



- Heavy payloads tend to have a higher success rate for landing, especially in PO, LEO and ISS
- There doesn't seem to be a clear increase or decrease in success rate with increasing payload mass in GTO
- The other orbits have too few launches to draw any conclusions

Launch Success Yearly Trend



The average success rate of Falcon 9 launches has been increasing steadily over time

There was a noticeable rise in success rates between 2013 and 2015, possibly due to technological advancements or improvements in manufacturing procedures.

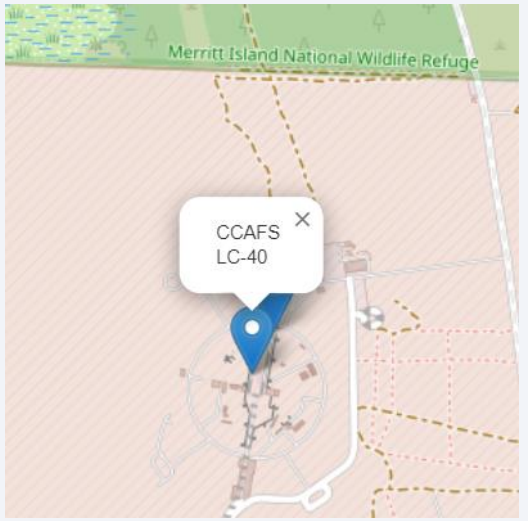
All Launch Site Names

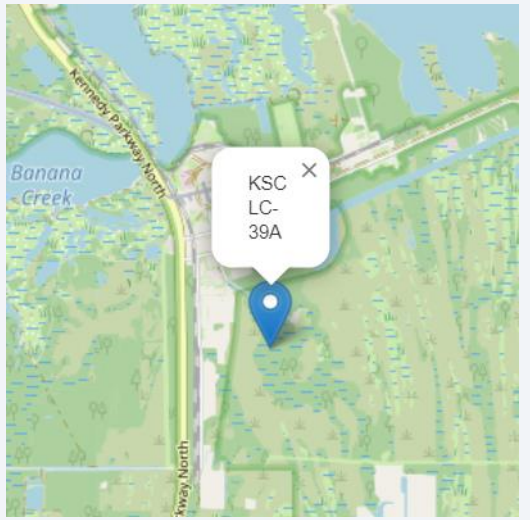
Unique launch sites in the space mission

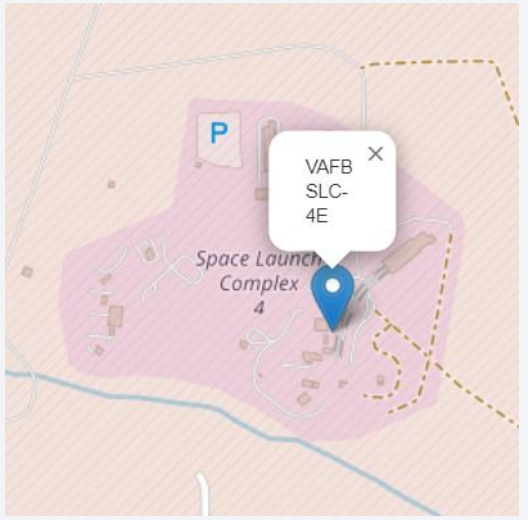
```
%sql SELECT DISTINCT launch_site FROM spacextable
```

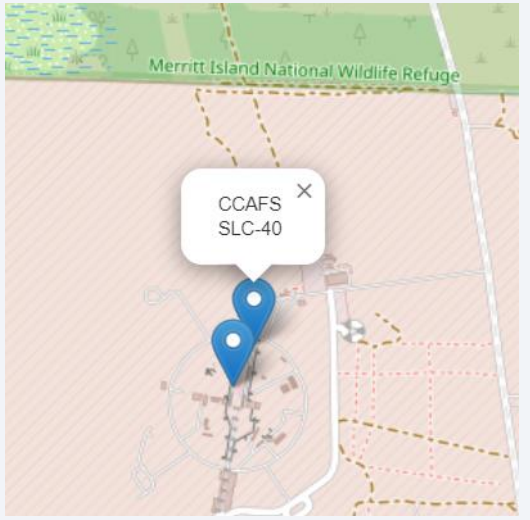
```
 * sqlite:///my_data1.db
Done.
```

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









Launch Site Names Begin with 'CCA'

Five records where launch sites begin with the string '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_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

Total payload mass carried by boosters launched by **NASA (CRS)**

```
%sql SELECT sum(payload_mass__kg_) total_payload_mass FROM spacextable WHERE customer = 'NASA (CRS)'
```

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

total_payload_mass
45596

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version **F9 v1.1**

```
%sql SELECT avg(payload_mass__kg_) average_payload_mass from spacextable WHERE booster_version = 'F9 v1.1'
```

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

average_payload_mass
2928.4

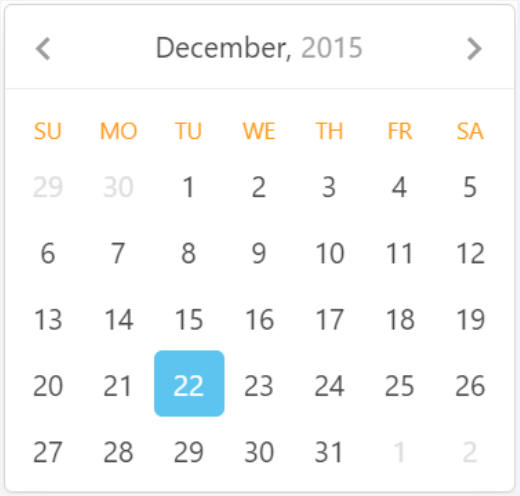
First Successful Ground Landing Date

Date of the first successful landing outcome in ground padC

```
%sql SELECT min(date) first_successful_landing FROM spacextable WHERE landing_outcome = 'Success (ground pad)'
```

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

first_successful_landing
2015-12-22



Successful Drone Ship Landing with Payload between 4000 and 6000

Names of the boosters which have **success in drone ship** and have payload mass greater than **4000** but less than **6000**

```
%sql SELECT booster_version FROM spacextable WHERE \
landing_outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 and payload_mass__kg_ < 6000
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Total number of **successful** and **failure** mission outcomes

```
%sql SELECT count(1) total, mission_outcome FROM \
(SELECT upper(trim(mission_outcome)) mission_outcome FROM spacextable) GROUP BY mission_outcome
```

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

total	mission_outcome
1	FAILURE (IN FLIGHT)
99	SUCCESS
1	SUCCESS (PAYLOAD STATUS UNCLEAR)

Boosters Carried Maximum Payload

Names of the booster_versions which have carried the maximum payload mass

```
%sql SELECT booster_version FROM spacextable WHERE \
payload_mass__kg_ = (SELECT max(payload_mass__kg_) FROM spacextable)
```

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

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

Month, **failure landing_outcomes in drone ship**, booster version, and launch site for the months in year **2015**

```
%sql SELECT substr(date, 6,2) month, landing_outcome, booster_version, launch_site FROM \
spacextable WHERE substr(Date,0,5) = '2015' AND landing_Outcome = 'Failure (drone ship)'
```

```
 * 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

Count of landing outcomes between the date **2010-06-04** and **2017-03-20**, in descending order

```
%sql select count(1) total, landing_outcome FROM \
spacextable WHERE date BETWEEN '2010-06-04' AND \
'2017-03-20' GROUP BY landing_outcome ORDER BY 1 DESC
```

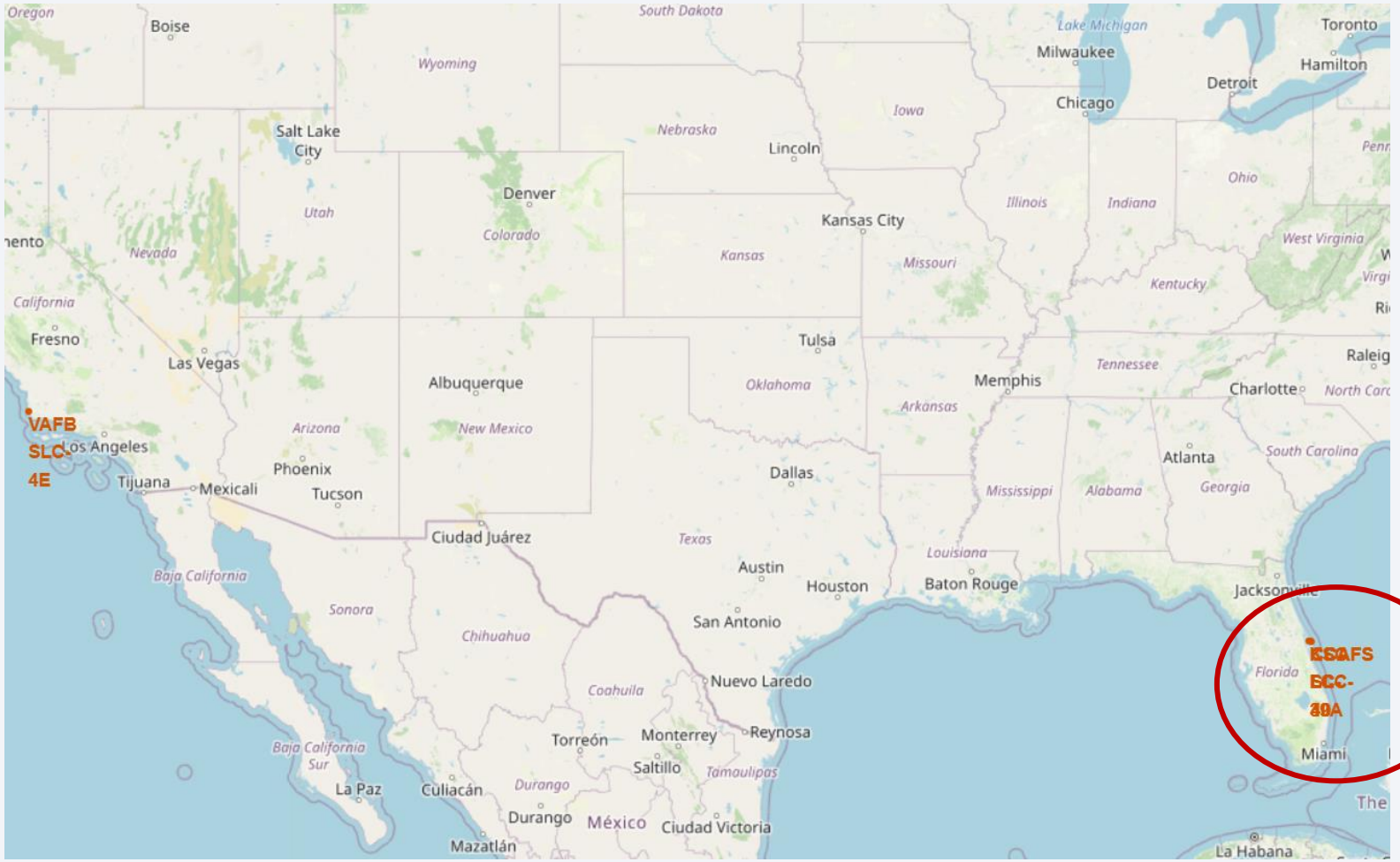
```
 * sqlite:///my_data1.db
Done.
```

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

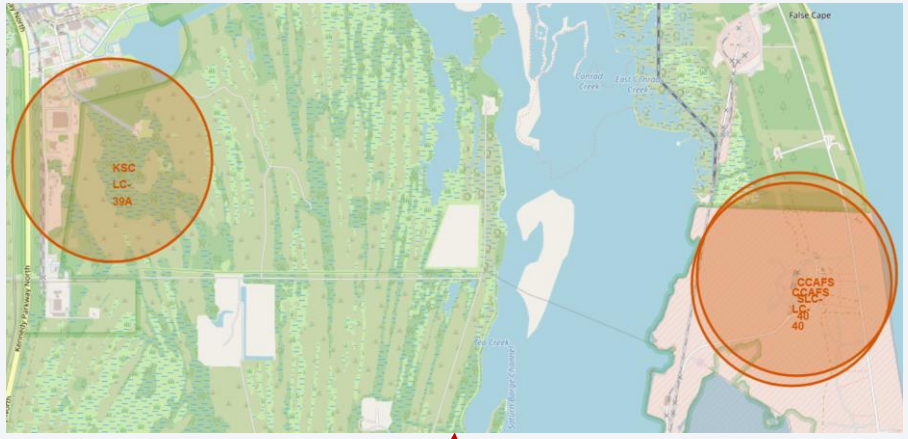
Section 3

Launch Sites Proximities Analysis

All Launch Sites



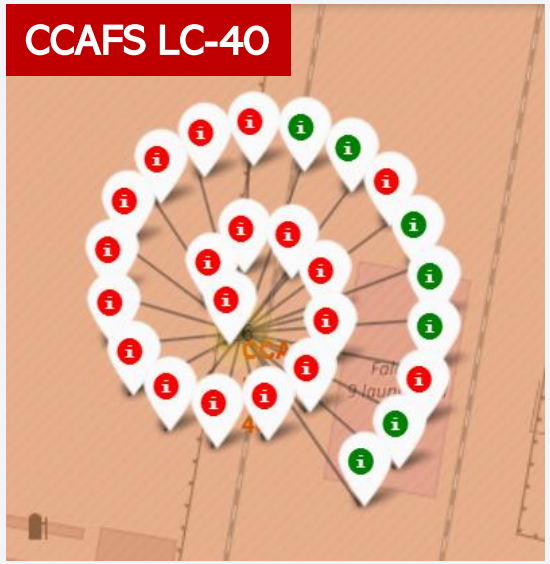
All launch sites are located near coastlines to provide flexibility for trajectory options and safety considerations

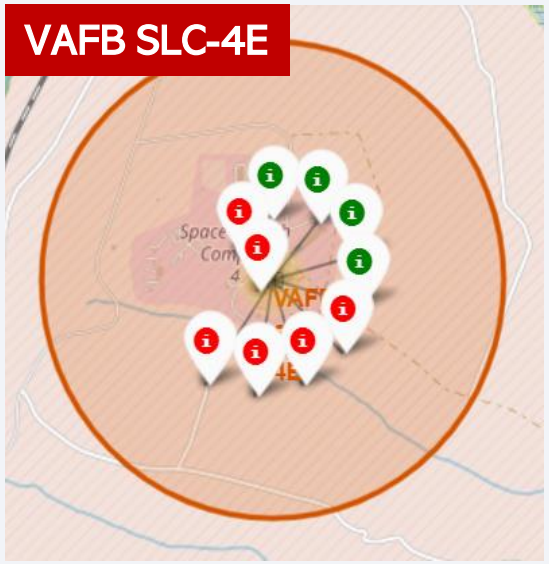


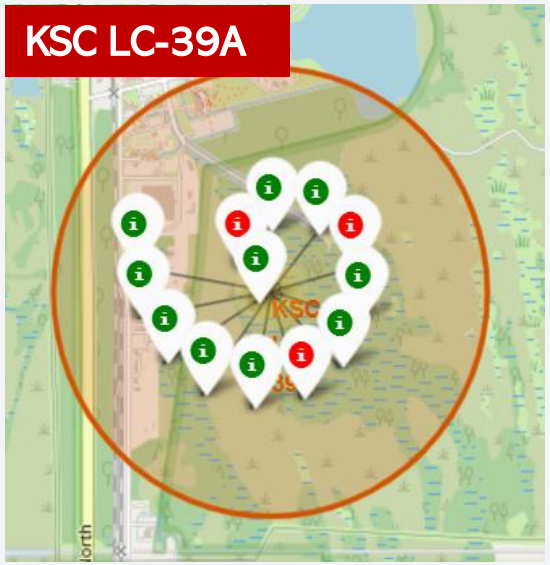
Launch Outcomes by Site

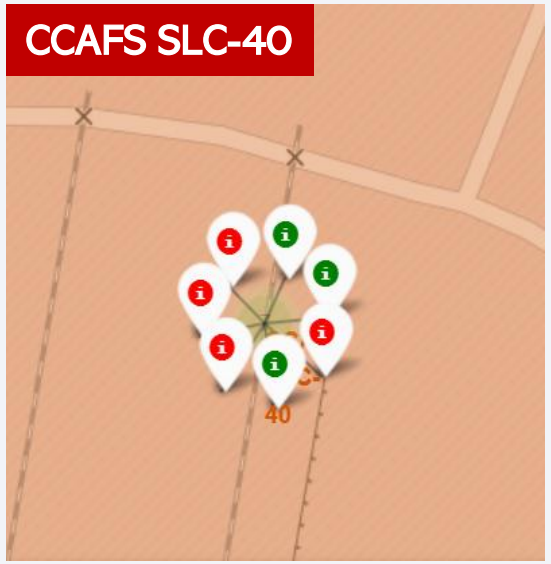
Successful launch

Failed launch





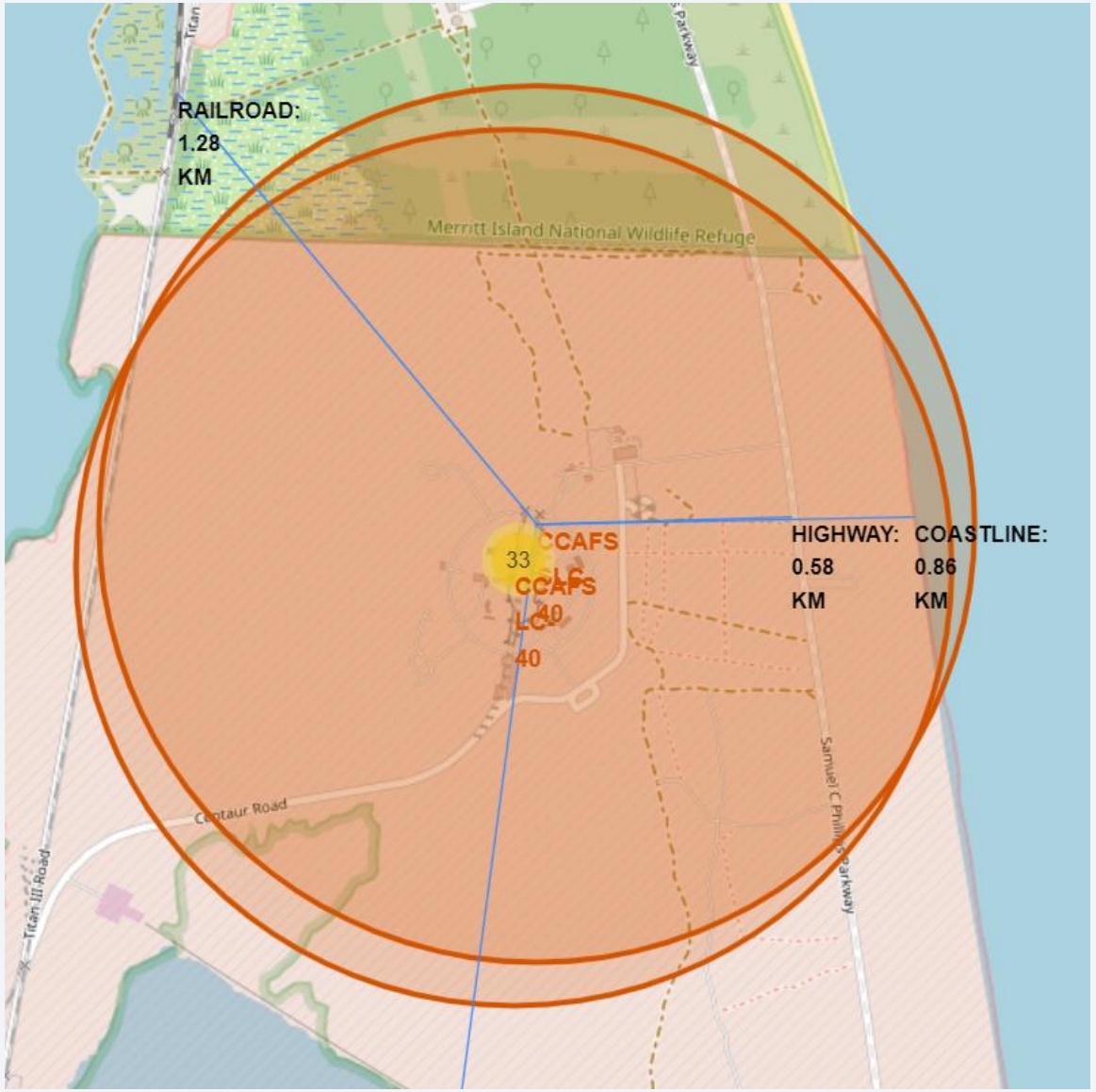




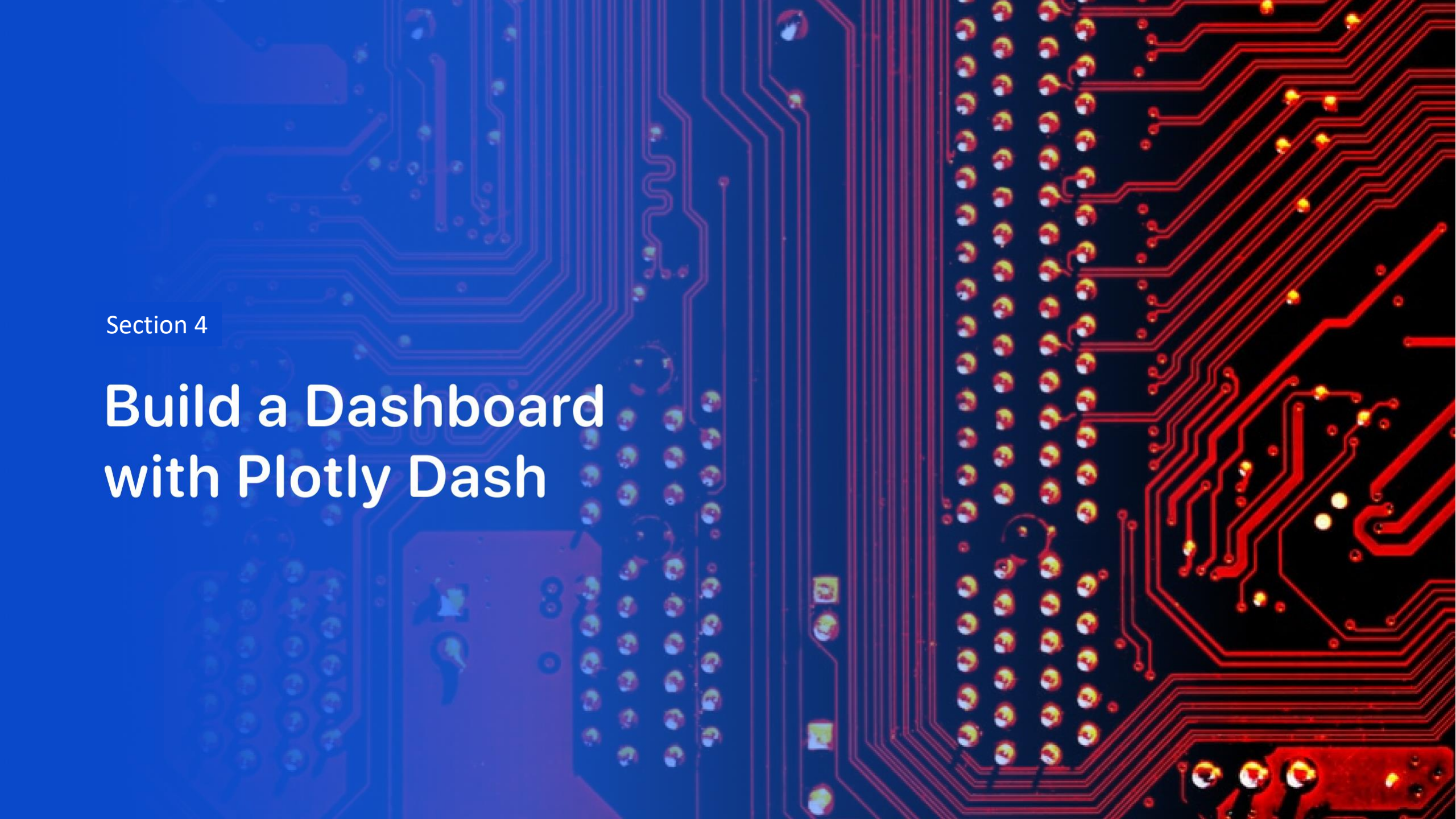
The launch site with the **highest** number of launches is CCAFS LC 40, however, it has a high percentage of failed launches

KSC LC-39A has a very high success rate

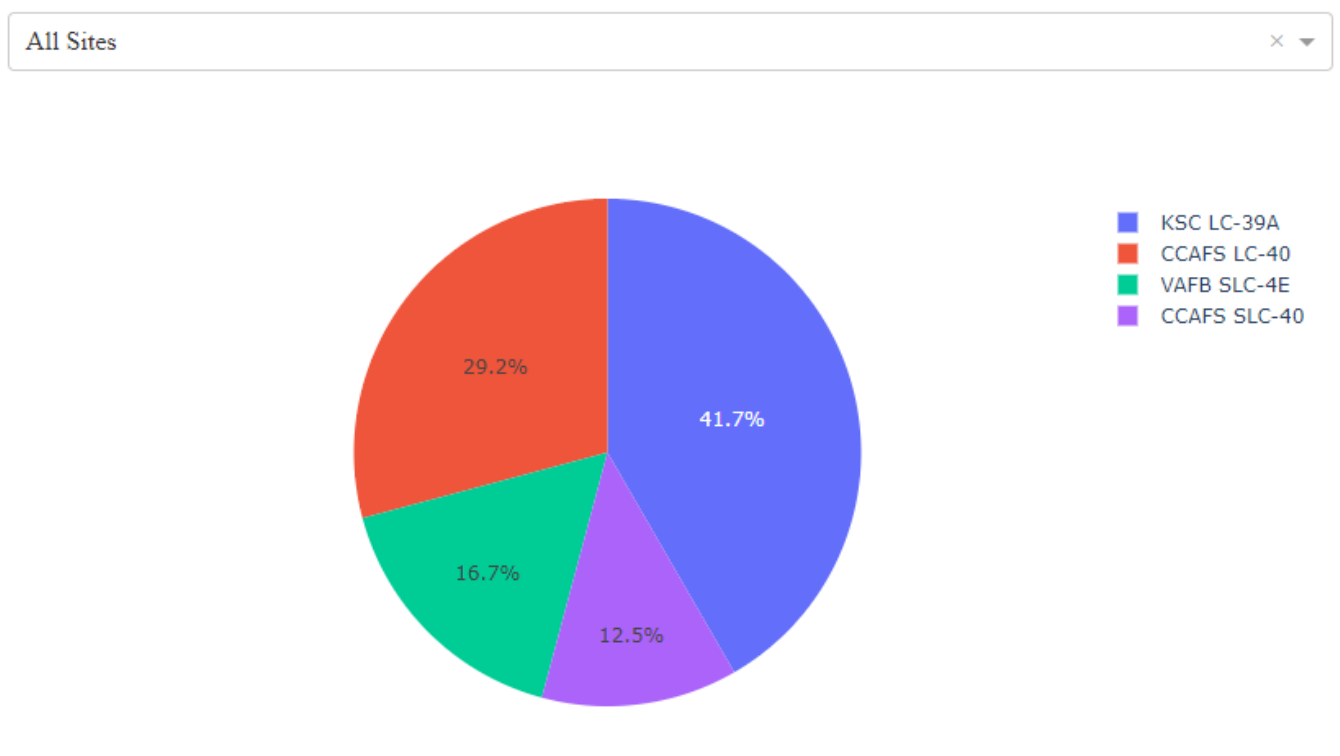
Distance Between CCAFS SLC 40 to its Proximities



CCAFS SLC 40 is situated in proximity to a coastline, a highway, and a railroad; however, it is not near inhabited areas


Section 4
Build a Dashboard with Plotly Dash

Launch Success Count – All Sites

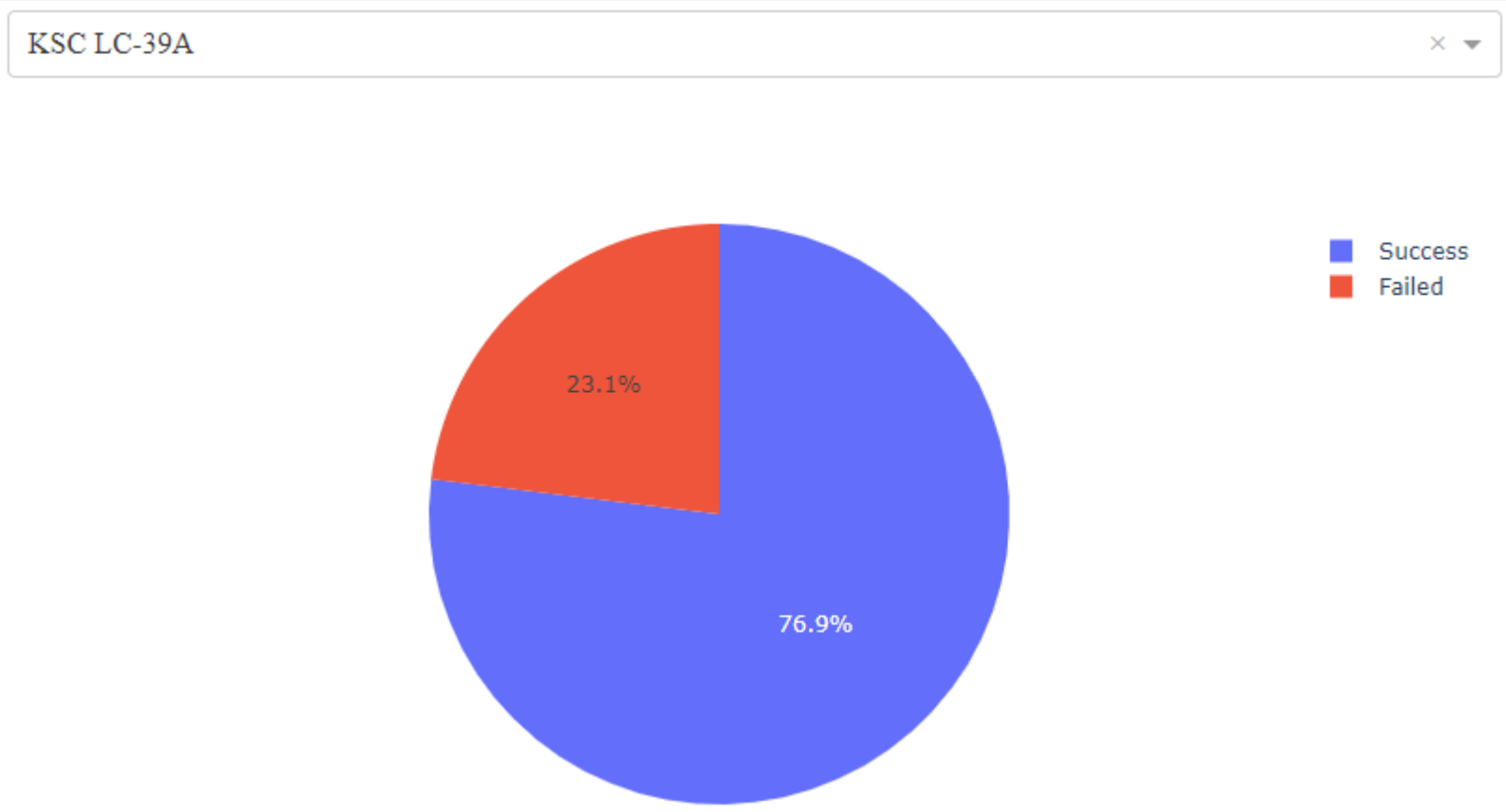


KSC LC 39A is the launch site with the highest number of successful launches

Although **CCAFS LC-40** has the highest number of launches among the launch sites, its success count is not particularly high

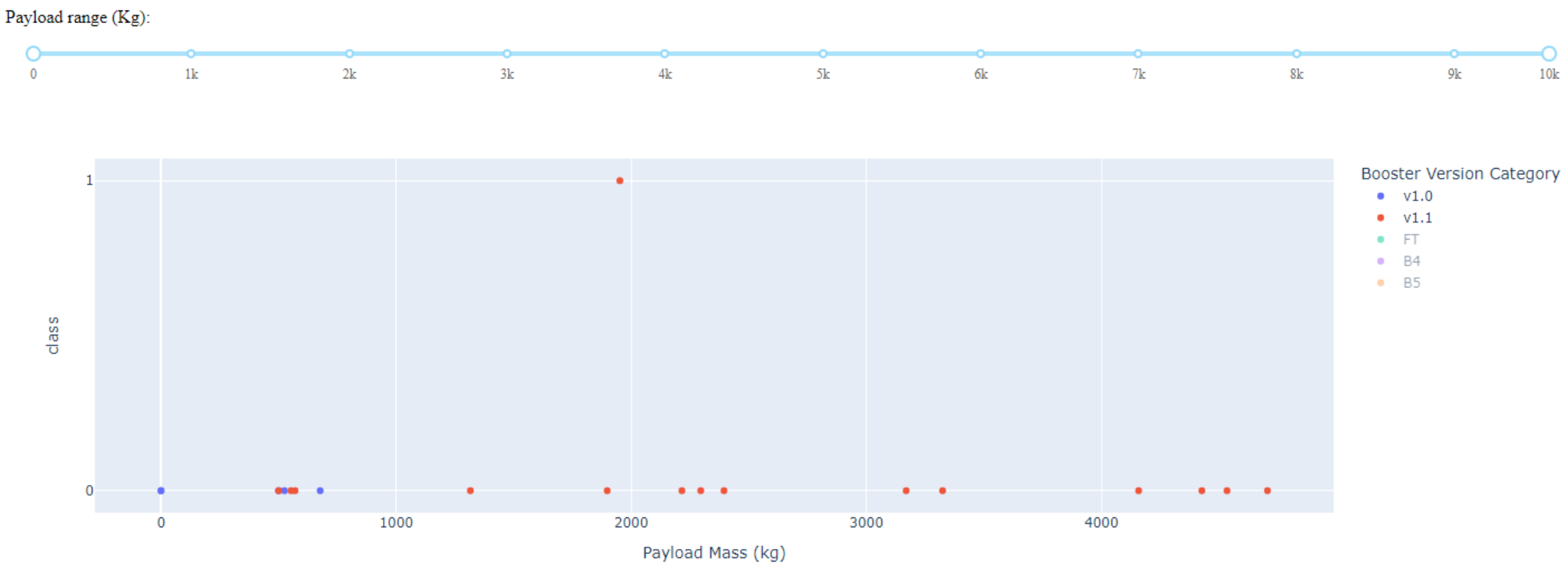
CCAFS SLC-40 has the lowest number of successful launches, but also the smallest number of launches.

Highest Launch Success Rate



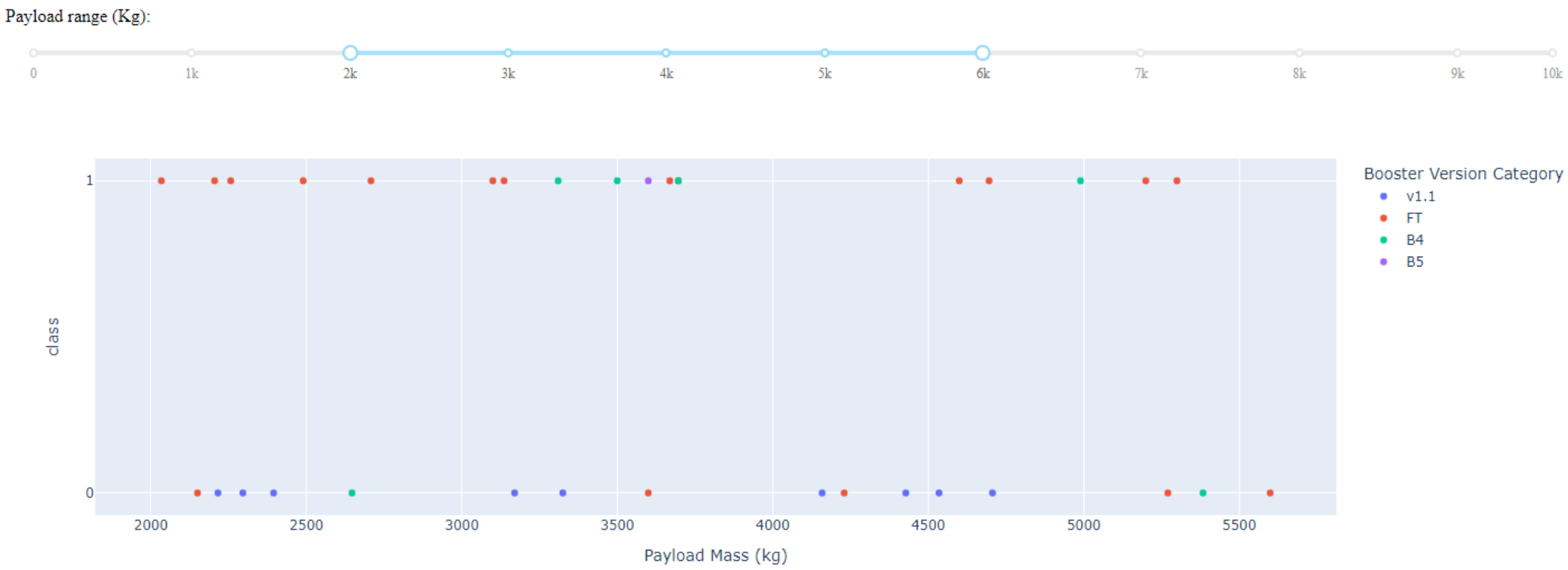
KSC LC 39A is the launch site with the highest success rate with **10 successful launches** and **3 failed** ones

Lowest Launch Success Rate



Booster **v1.0** has only **failed launches**, while booster **v1.1** has **only one successful launch**

Highest Launch Success Rate



Payloads ranging between **2000** and **5500** have the highest launch success rate

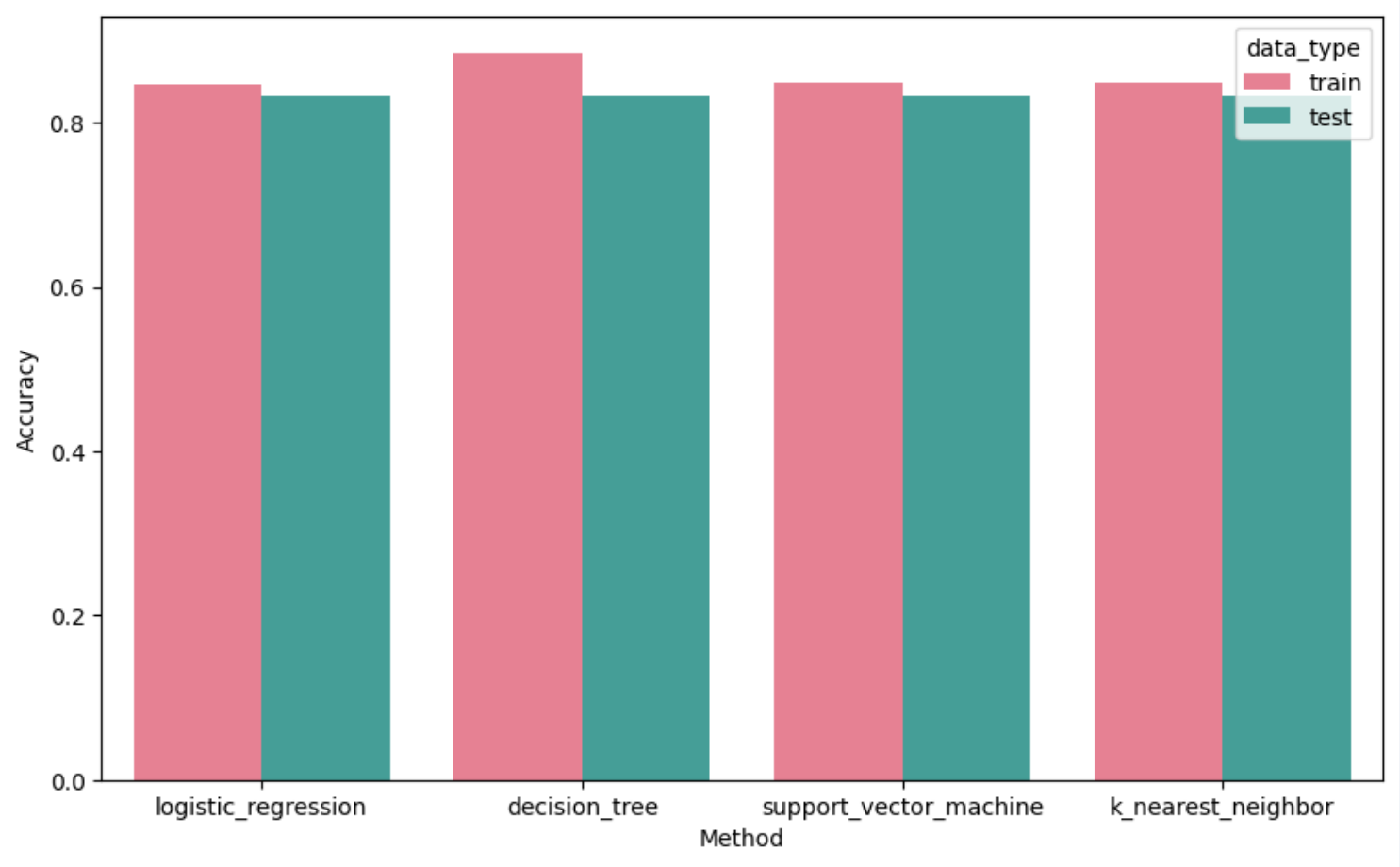
FT boosters have the highest success rate

Most unsuccessful launches within this range are associated with **v1.1** boosters, which only has one successful launch overall (not in this range)

Section 5

Predictive Analysis (Classification)

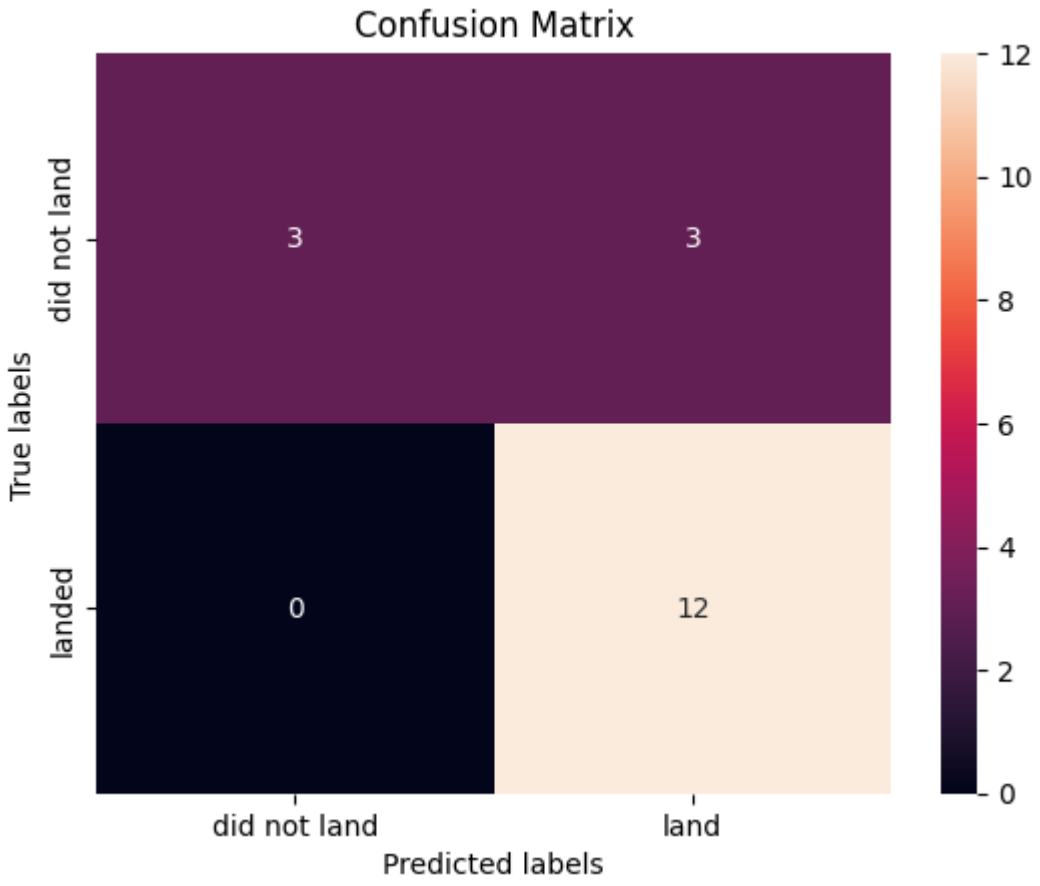
Classification Accuracy



method	data_type	accuracy
logistic_regression	train	0.846429
logistic_regression	test	0.833333
decision_tree	train	0.885714
decision_tree	test	0.833333
support_vector_machine	train	0.848214
support_vector_machine	test	0.833333
k_nearest_neighbor	train	0.848214
k_nearest_neighbor	test	0.833333

All methods exhibited nearly identical performance, with the exception that the **decision tree classifier** slightly outperformed the others on **training data**

Confusion Matrix



True Positives: 3
False Positives: 3
False Negatives: 0
True Negatives: 12

While most results were accurate, the issue appears to lie in false positives. It is possible that expanding the dataset could lead to improved accuracy

Conclusions

All launch sites for Falcon 9 are near a coastline

The best launch site in terms of the quantity of launches vs success rate is KSC LC-39A

Payload mass ranging between 2000 and 5500 have the highest launch success rate

FT booster version category has the highest launch success rate

Logistic regression, decision tree, support vector machine, and k-nearest neighbors methods exhibit nearly identical performance in predicting the successful first stage landing of the Falcon 9. The decision tree classifier, however, shows a slightly better performance on the training data.

The likelihood of launch success shows an upward trend both over the years and as the flight number increases

Thank you!