IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Houston Erwin
20-April-2025

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection and Wrangling using Python and SQL
  - Comprehensive Data Analysis and Visualization with Python
  - Predictive Analysis with Machine Learning Algorithms (Logistic Regression, Random Forest, KNN)
- **Summary of all results**
  - Descriptive and Exploratory Analytics
  - Interactive Dashboard
  - Machine Learning Modeling

Introduction

- The purpose of this analysis was to predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
- This report will show our analysis results to determine if the first stage will land and determine if an alternate company wants to bid against SpaceX.

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

Using SpaceX REST API to make a Get Request

Perform data wrangling

Preprocessing data to include a binary “Class” to determine success or failure and updating N/A values

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

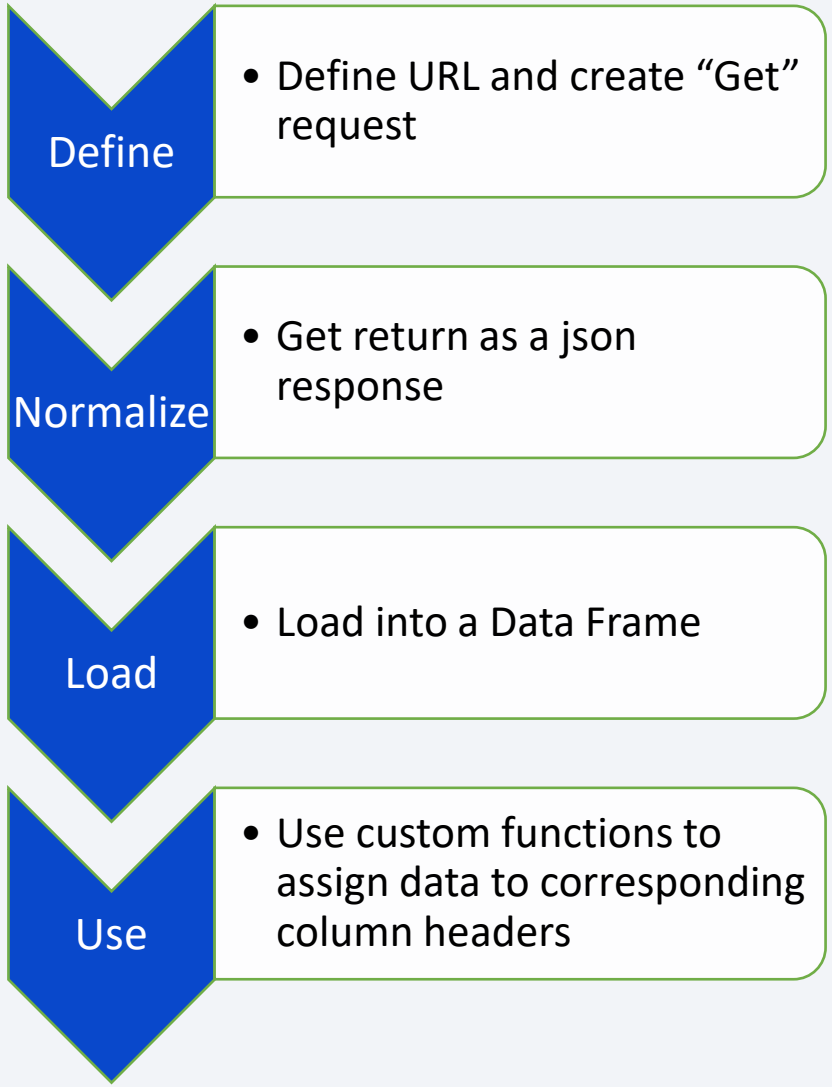
Perform predictive analysis using classification models

Tuning ML models with Grid Search and grading via Confusion Matrix

Data Collection – SpaceX REST API

- **Get** data from SpaceX URL
- **Confirm** response is 200
- **Normalize** data from the json response into a data frame,
- Use **custom functions** to get data and assign to columns,
- **Filter** for only Falcon 9 models

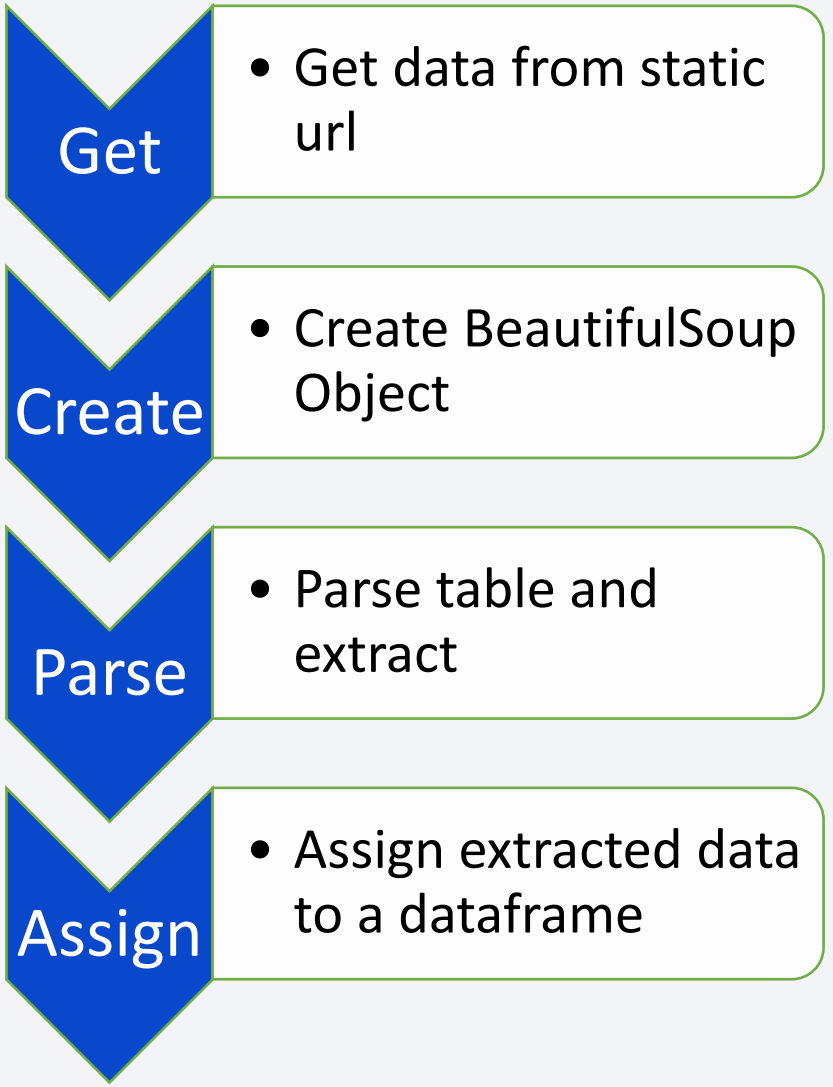
https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_API_Calls.ipynb



Data Collection – Web Scraping

- Using a get request and beautiful soup, an html request were parsed, and the Falcon 9 table rows were extracted and assigned to column headers using custom functions.

- https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_Webscraping.ipynb

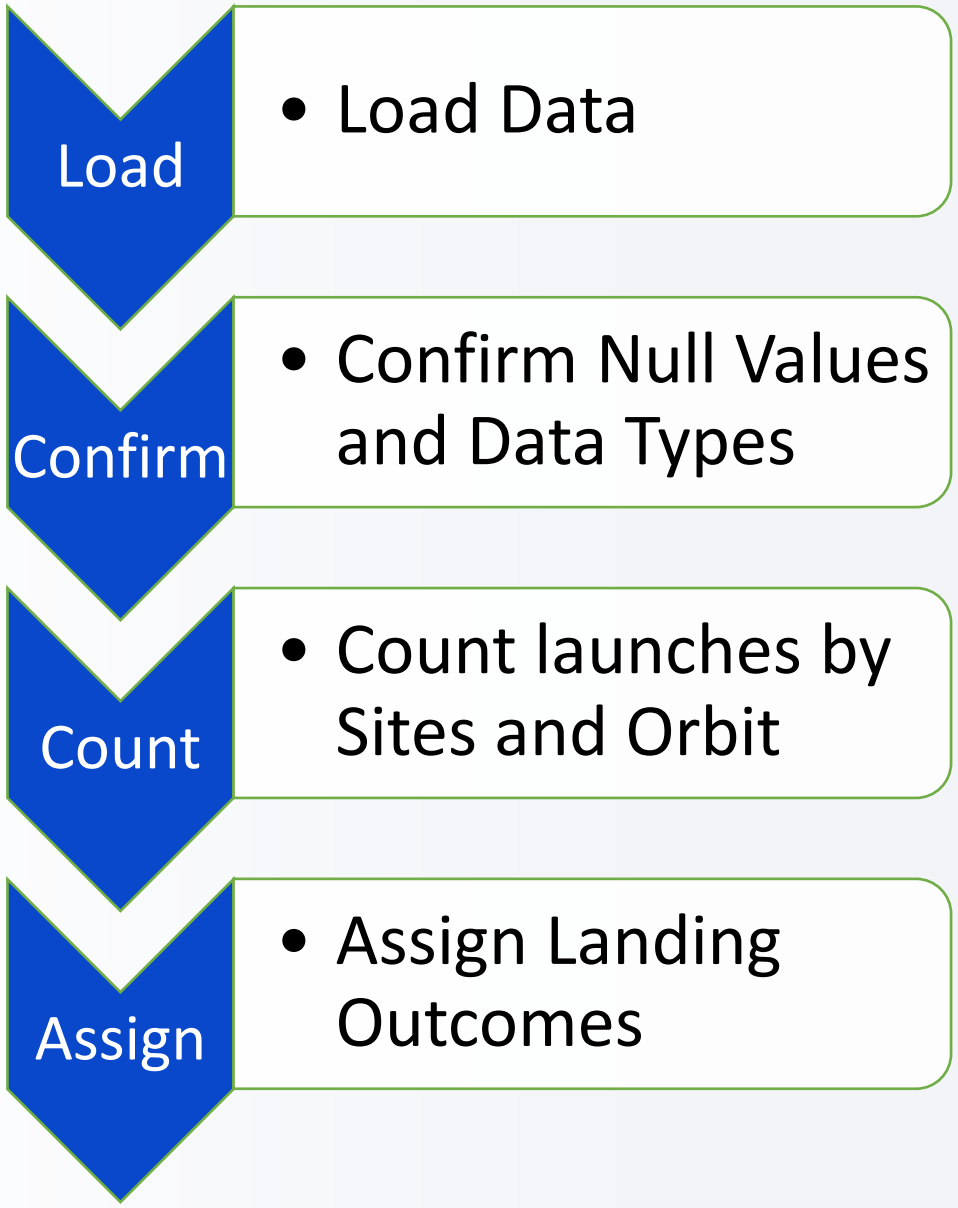


Data Wrangling

- Insights about our data including **count** of **null** values and data types were obtained first
- Questions were then answered about how often launches take place at different **site locations** and **orbits**
- Data was **modified** to include a new column to create a binary class of successful launches (0 for failure, 1 for success)

GitHub

- https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_Data_Wrangling.ipynb



EDA with Data Visualization

GitHub

https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_EDA_Data%20Visualization.ipynb

Line Plot (best for time-based variables)

- Successful Flights Over Time

Scatter Plot (efficient for simple correlation)

- Flight Number and Launch Site
- Launch Site and Payload Mass
- Flight Number and Orbit
- Payload Mass and Orbit

Bar Plot (best for categorical data)

- Orbit and Class

EDA with SQL

The following SQL commands were performed:

1. **Display the names of the unique launch sites in the space mission**

2. **Display 5 records where launch sites begin with the string 'CCA'**

3. **Display the total payload mass carried by boosters launched by NASA (CRS)**

4. **Display average payload mass carried by booster version F9 v1.1**

5. **List the date when the first successful landing outcome in ground pad was achieved.**

6. **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**

7. **List the total number of successful and failure mission outcomes**

8. **List all the booster versions that have carried the maximum payload mass. Use a subquery.**

9. **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.**

10. **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.**

GitHub

https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_EDA_SQL.ipynb

Build an Interactive Map with Folium

Method:

- Markers were placed at each launch site and color coded based on whether the launch was a success or failure. Lines were then generated between the launch sites and geographic points of interest such as the ocean and Cocoa Beach.

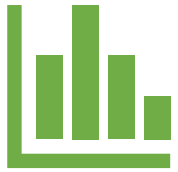
Rationale:

- Marking launch sites and parsing them by success rates gives us insight into whether successes or failures occur more frequently in a particular location or within proximity to a common geographic marker.

GitHub

https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_EDA_Visual_Folium.ipynb

Build a Dashboard with Plotly Dash



Pie Chart

Clearly see the distribution of successful launches by site

Scatter Chart

Clearly determine the correlation between Payload and Successful Launches at all sites with a slicer to filter the data by Payload Mass

https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/spacex-dash-app%20(1).py

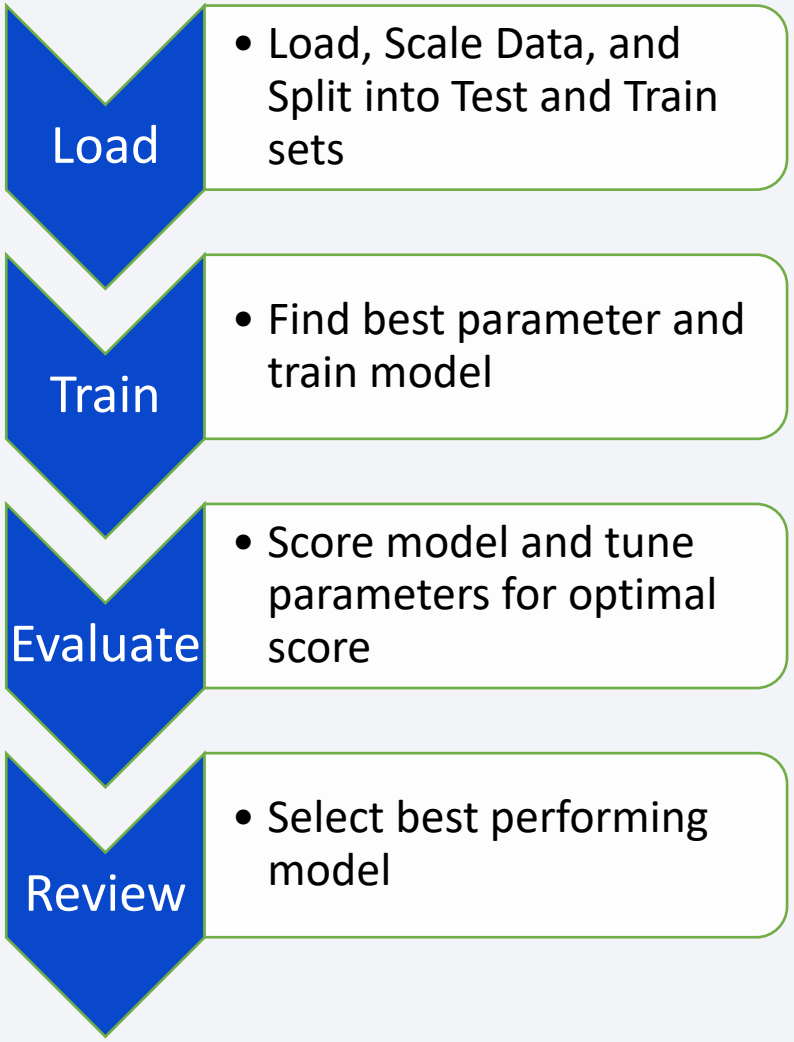
Predictive Analysis (Classification)

ML models were utilized and evaluated to best predict successful outcomes of the Falcon 9

- Model Types
  - Logistic Regression
  - Support Vector Machine
  - Random Forest
  - K Nearest Neighbor
- Model Evaluation
  - Accuracy Score
  - Confusion Matrix
- Model Tuning
  - GridSearch CV

GitHub

https://github.com/erwinw2/Public/blob/3e52eb5b0fa5509bc415eb5e13bd36e55d365904/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

Exploratory data analysis results

Interactive analytics demo in screenshots

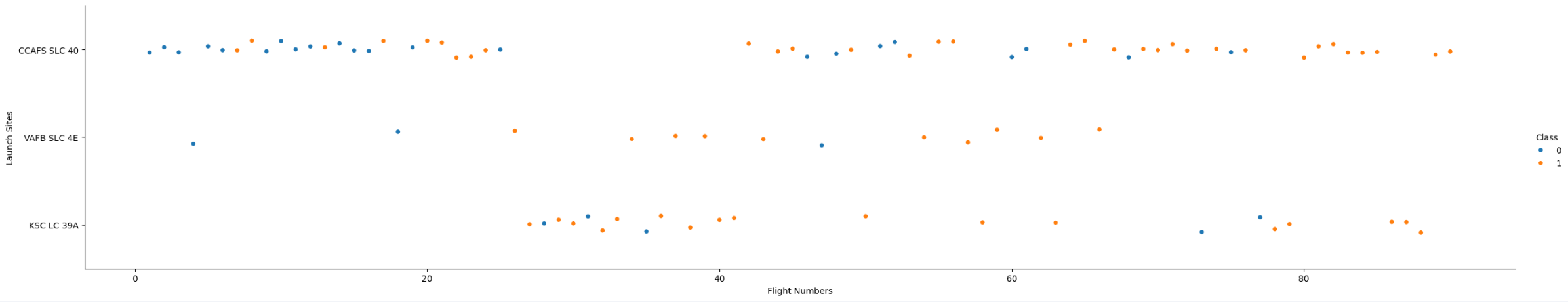
Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

1. CCAPS SLC 40
   - Main launch site in the beginning
   - Failures significantly decreased as more flights took place over time
2. VAFB SLC 4E
   - Fewest number of launches
   - More successes than failures but site not utilized consistently
3. KSC LC 39A
   - Launches at this site began mid-program
   - Fewer failures at the beginning most likely due to lessons learned from CCAPS SLC 40

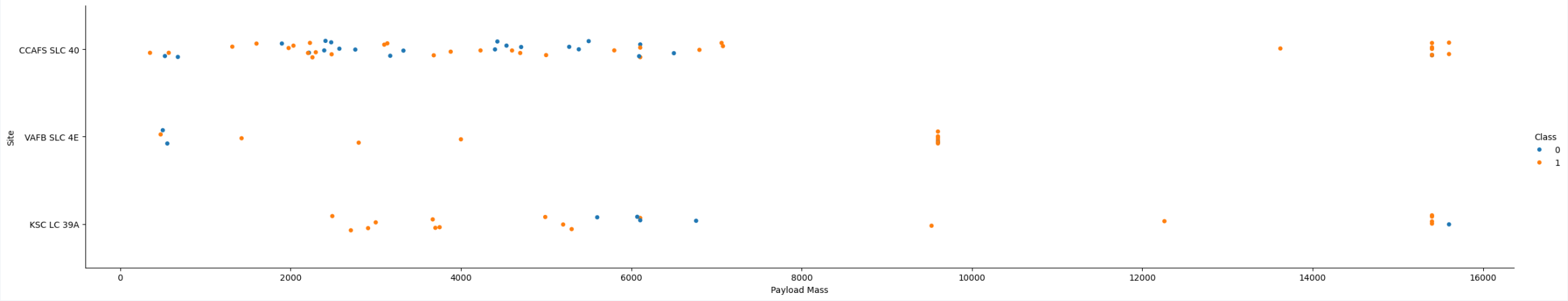


Payload vs. Launch Site

1. CCAPS SLC 40
   - Payloads primary stayed less than 8000kg
   - Higher payload masses did not correlate with higher failure rate

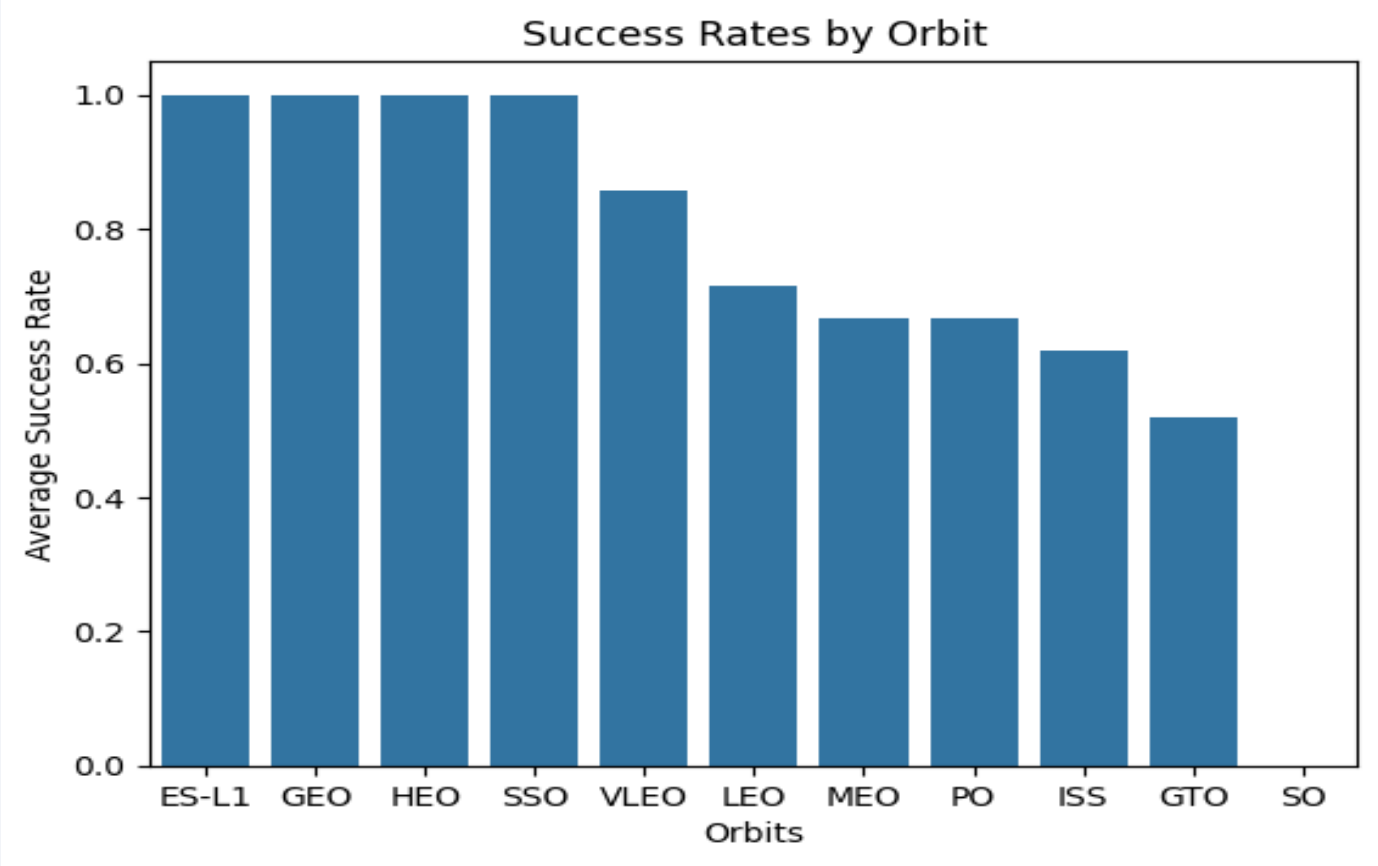
2. VAFB SLC 4E
   - Site had fewer launches, but did not appear to have failure at payloads around 10,000 kg

3. KSC LC 39A
   - Primary payload weight around 2000-6000 kg with no clear failure correlation



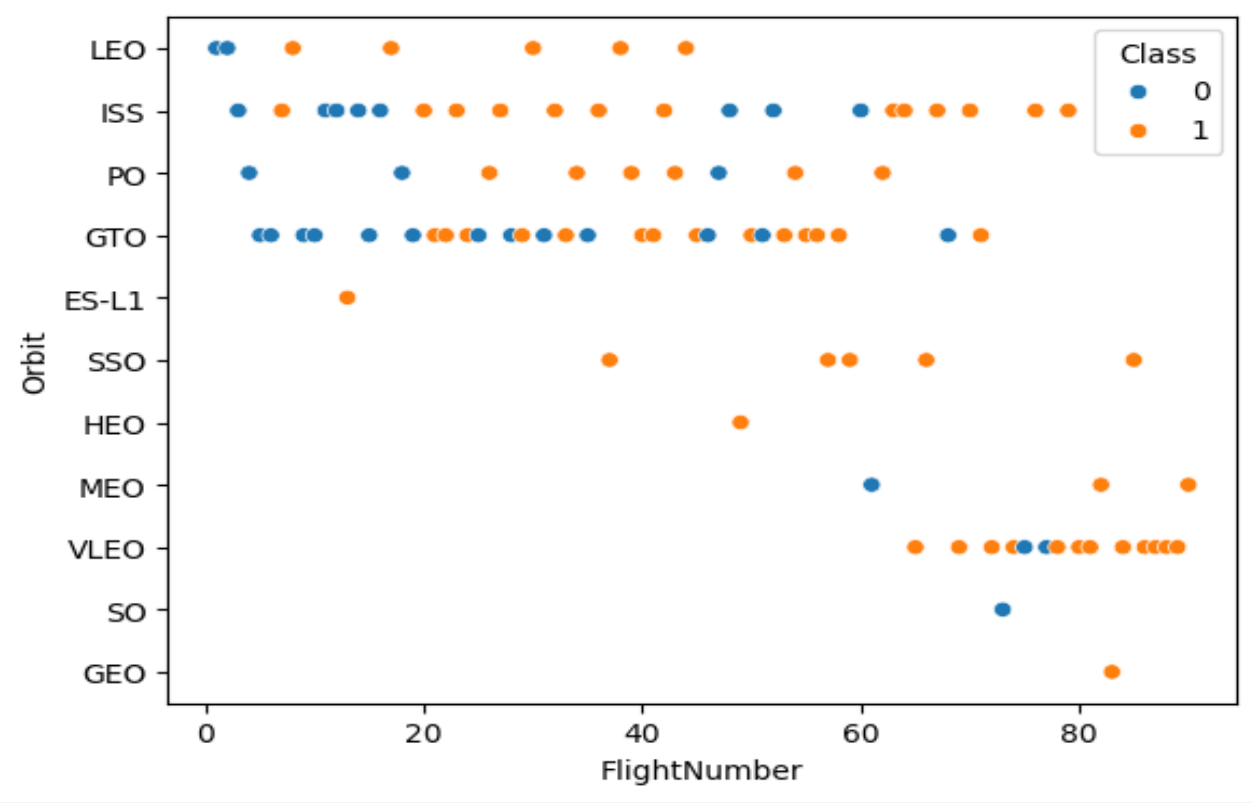
Success Rate vs. Orbit Type

- Insights
  - Highest average success rate was seen at ES-L1, GEO, HEO, and SSO
  - SO has no successes
  - Not enough information on number of launch attempts to draw inference.



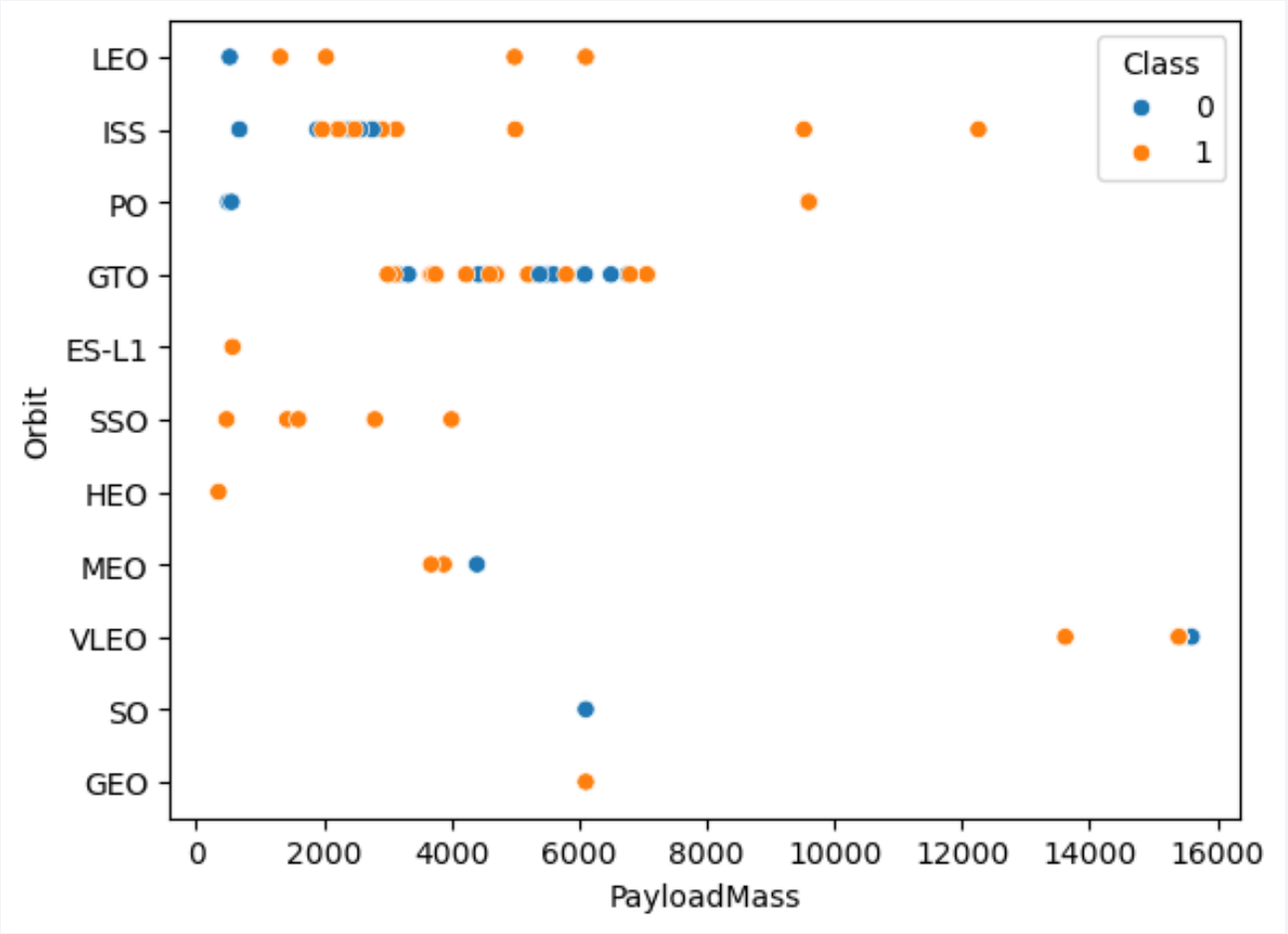
Flight Number vs. Orbit Type

- SSO, GEO, and HEO have a 100% success rate, but not enough launches to draw insight.
- **VLEO** appears to have the highest success rate with enough samples and may be preferred in later flights for this reason.
- **ISS** remains consistently successful as well possibly due to historical familiarity.



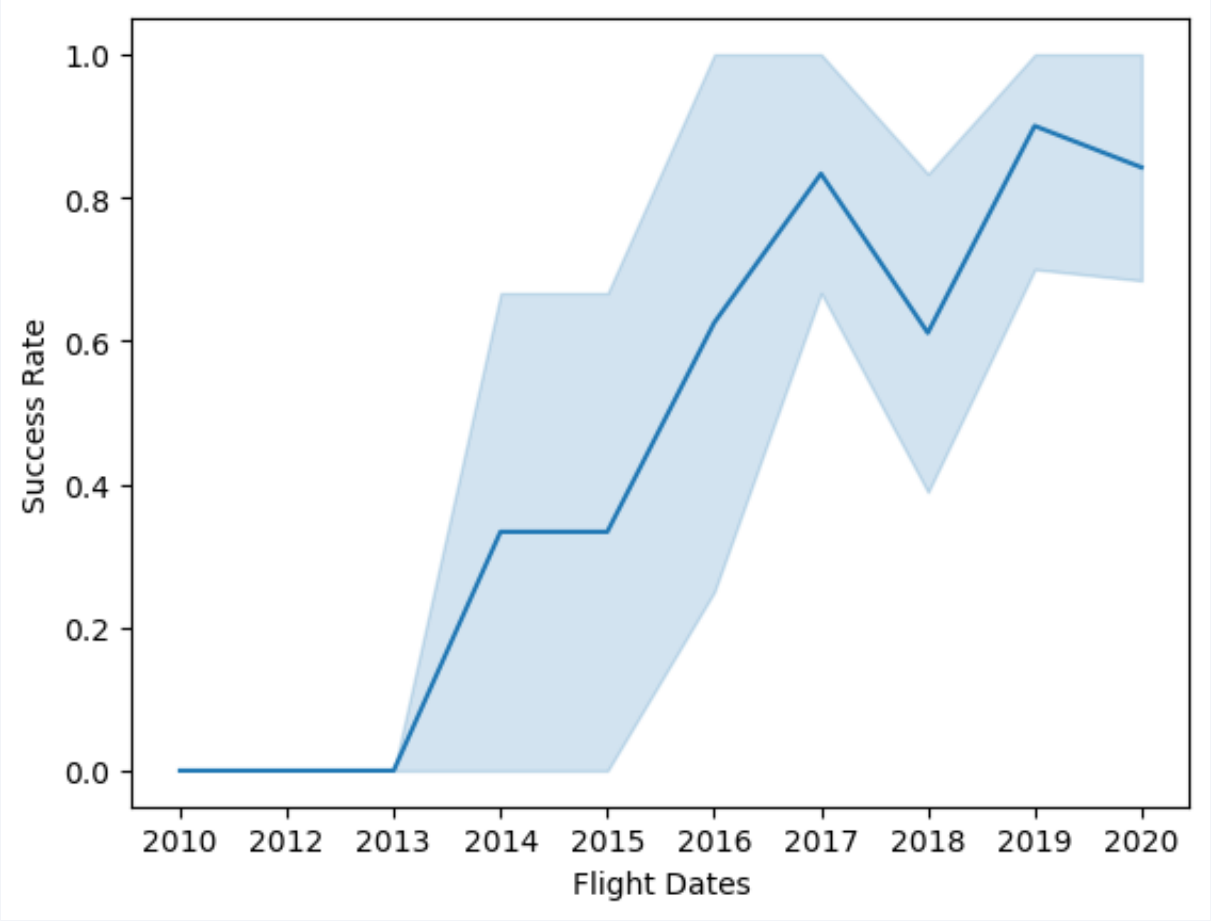
Payload vs. Orbit Type

- Insights
  - VLEO has the highest payloads
  - SSO has a 100% success rate but low launch cases
  - ISS has success in both low and high Payload ranges
  - GTO has a very concentrated payload range between 2000 and 8000.



Launch Success Yearly Trend

- Insights
  - Successful flights have been consistently increasing since 2013.
  - While there is a perceived dip in successful landings in 2018, all payloads were reported to successfully be deployed to space regardless.



All Launch Site Names

SQL query = %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE

Command will show distinct values

in the “Launch_Site” column in the table

“SPACEXTABLE”

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

Launch Site Names Begin with 'CCA'

SQL query = %sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE "CCA%" LIMIT(5)

Command will show 5 records

From the table “SPACEXTABLE”

Where the value in the “Launch_Site”

Column is begins 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

SQL query = %sql SELECT SUM("PAYLOAD_MASS__KG_") AS "Total Payload Mass" FROM SPACEXTABLE WHERE "Customer" LIKE "NASA%"

Command will show the sum of the

values in column “PAYLOAD_MASS__KG_”

with a column header as “Total Payload Mass”

when the value in the customer column begins

with “Nasa”

Total Payload Mass
99980

Average Payload Mass by F9 v1.1

SQL query = %sql SELECT AVG("PAYLOAD_MASS__KG_") AS "Average Payload Mass" FROM SPACEXTABLE WHERE "Booster_Version" LIKE "F9 v1.1%"

Command will show the average of the

values in column “PAYLOAD_MASS__KG_”

with a column header as “Average Payload Mass”

when the value in the booster version column begins

with “F9 v1.1”

Average Payload Mass
2534.6666666666665

First Successful Ground Landing Date

SQL query = %%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE “Mission_Outcome” = “Success”

Command will show the lowest date

value in the “Date” column of the rows

where the mission outcome was a success

MIN("Date")
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL query = %%sql SELECT DISTINCT("Booster_Version") FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000

Command will show the distinct values

in the booster version column when

the payload mass is between 4000

and 6000 kg.

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

SQL query = %%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") FROM SPACEXTABLE GROUP BY ("Mission_Outcome")

Command will show the count of the different mission outcomes grouped by mission outcome type.

Mission_Outcome	COUNT("Mission_Outcome")
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

SQL query = %%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
ORDER BY Booster_Version

Command via subquery will

show the booster versions

That have carried the maximum

payload of the table.

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

2015 Launch Records

SQL query = %%sql SELECT Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Failure (drone ship)' AND
BETWEEN '2015-01-01' AND '2015-12-31'

Command will show the failure drone ship

launch records that occurred in 2015.

Booster_Version	Launch_Site	Landing_Outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

SQL query = %%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) FROM SPACEXTABLE WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC

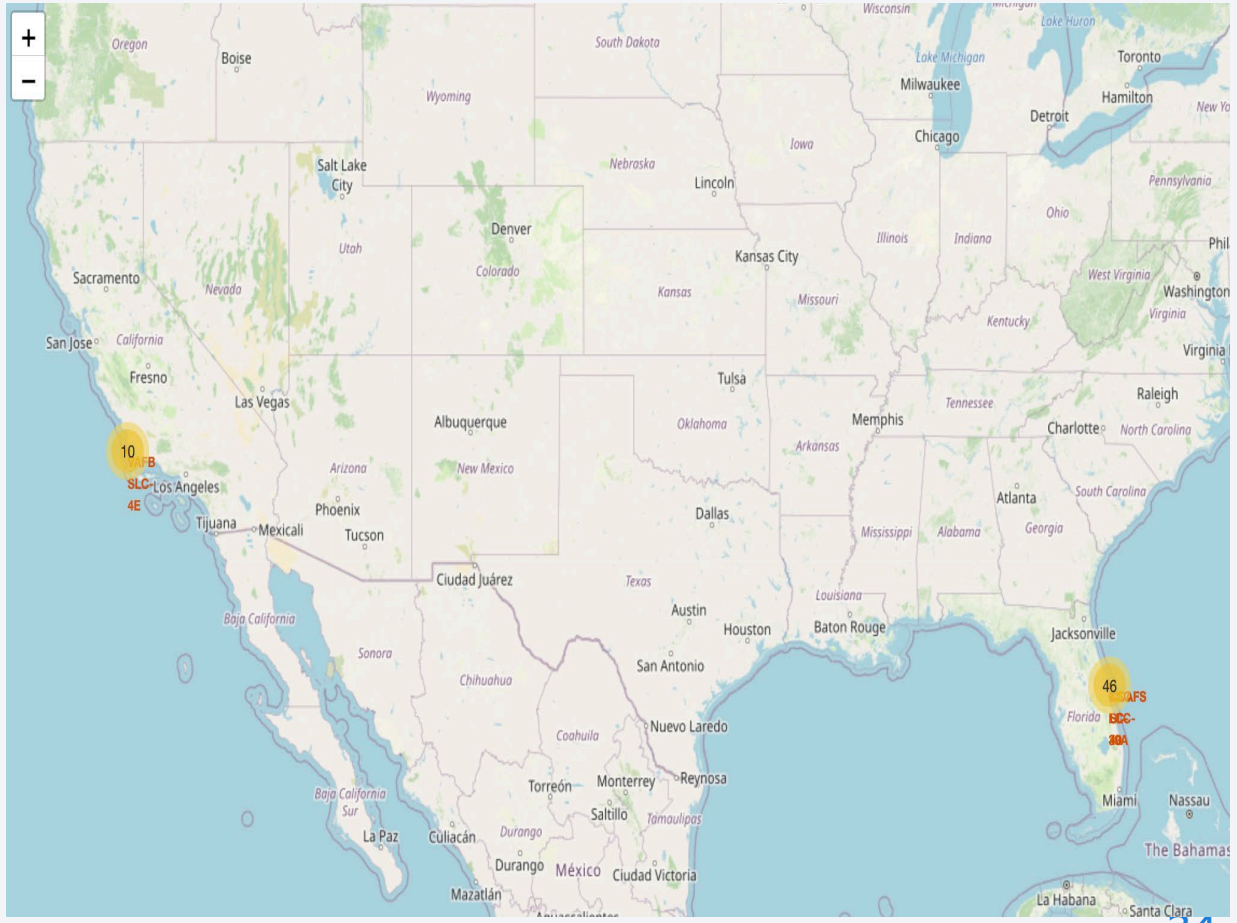
Command will show the landing outcomes between the corresponding dates grouped by landing outcome type and then shown in descending order

Landing_Outcome	COUNT(Landing_Outcome)
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

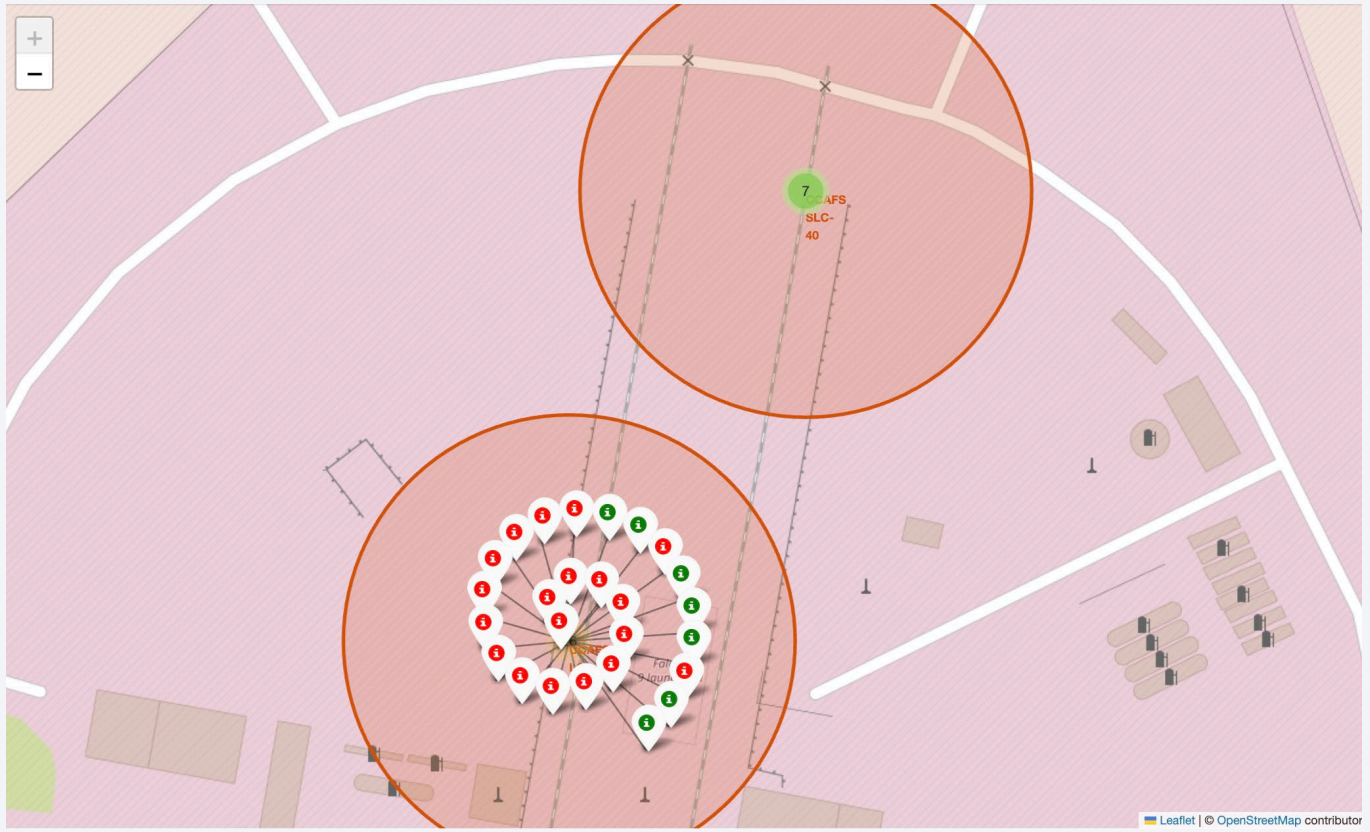
Folium Map: Country Level Launch Sites

- Most of SpaceX launches take place in Florida
- A few launches have occurred in California
- No launches have occurred at the NASA space station



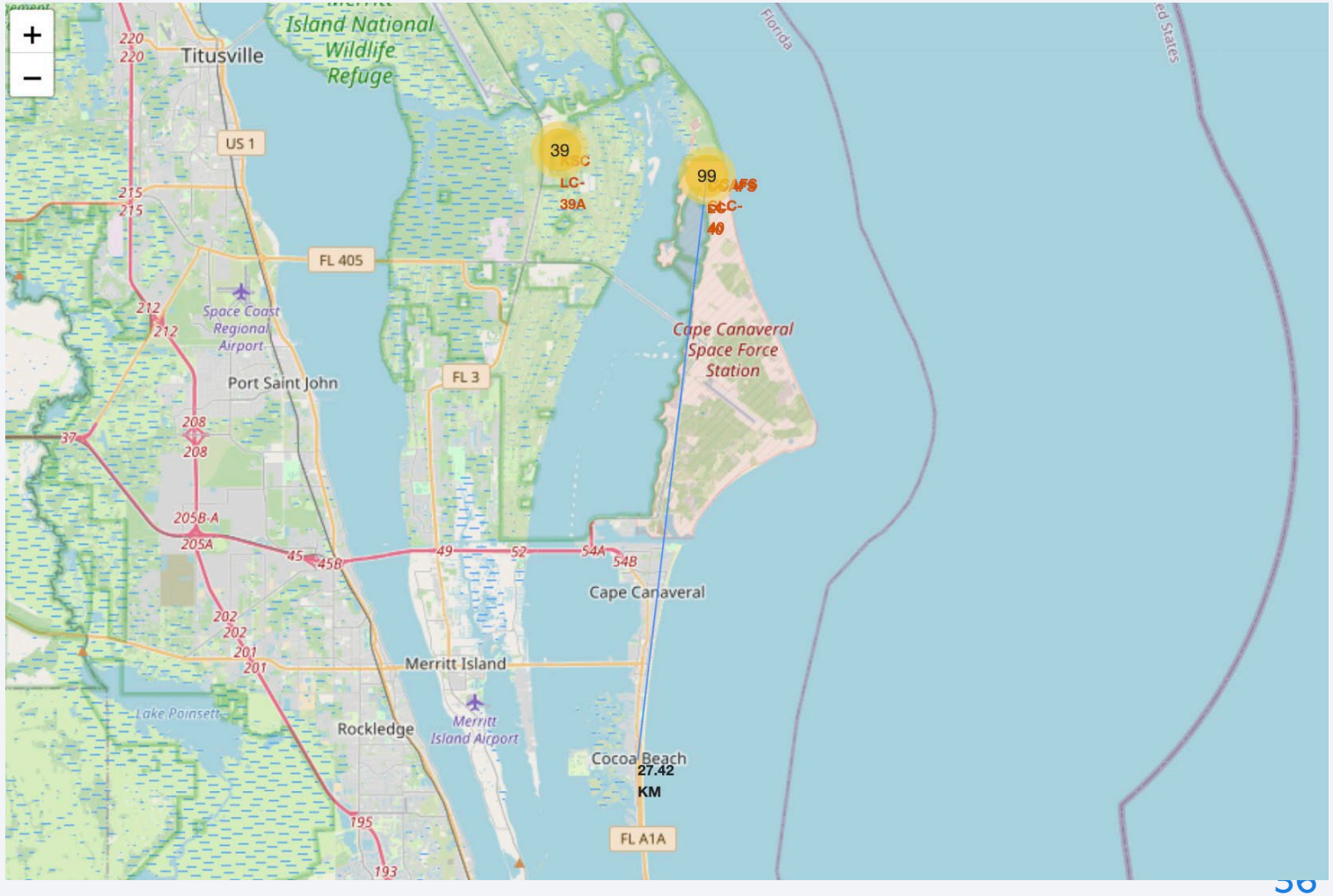
Folium Map: Color Coded Launch Outcomes

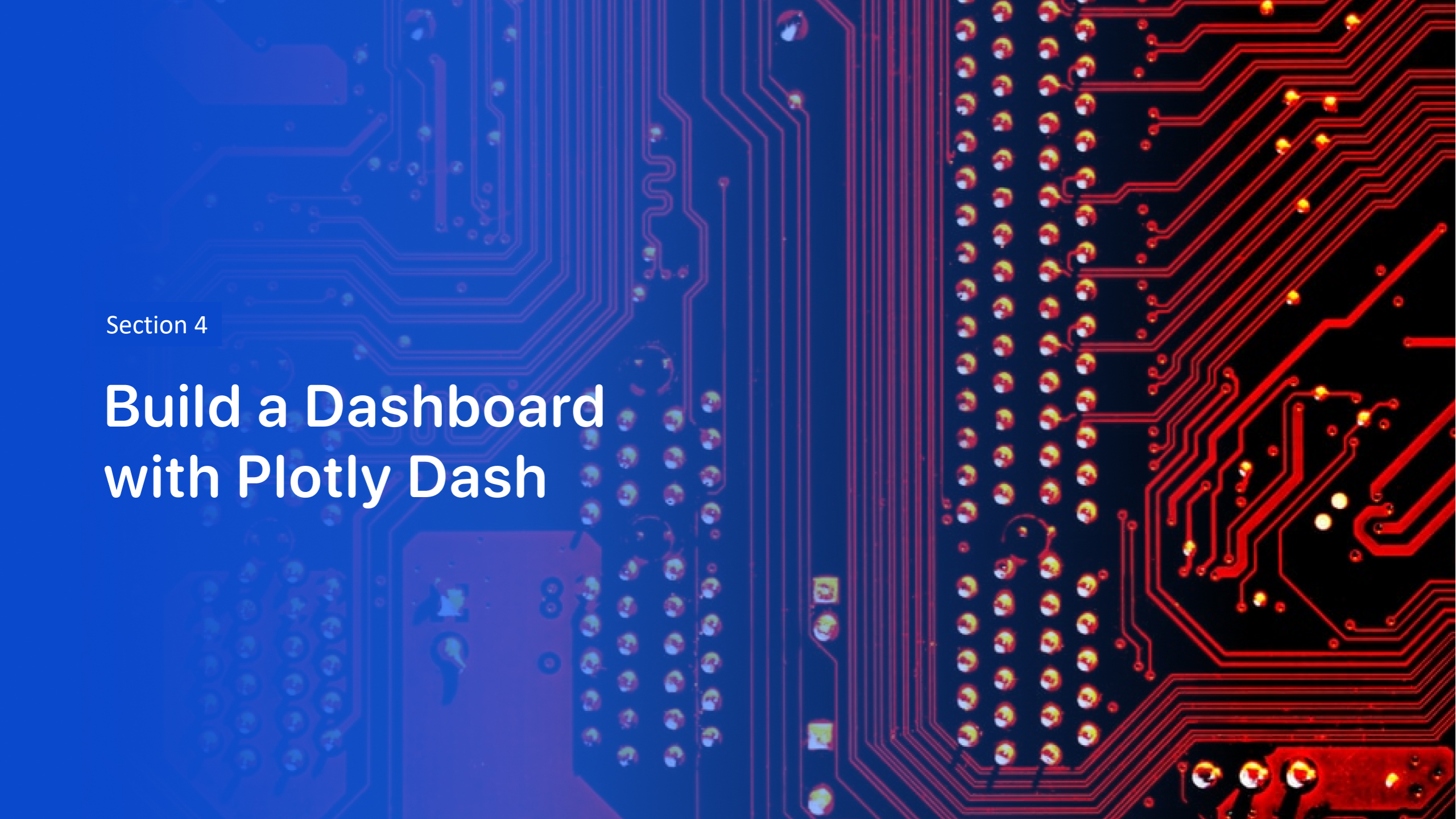
Launches are clustered by location and marked based on whether or not they were a success or a failure



Line Markers to Points of Interest

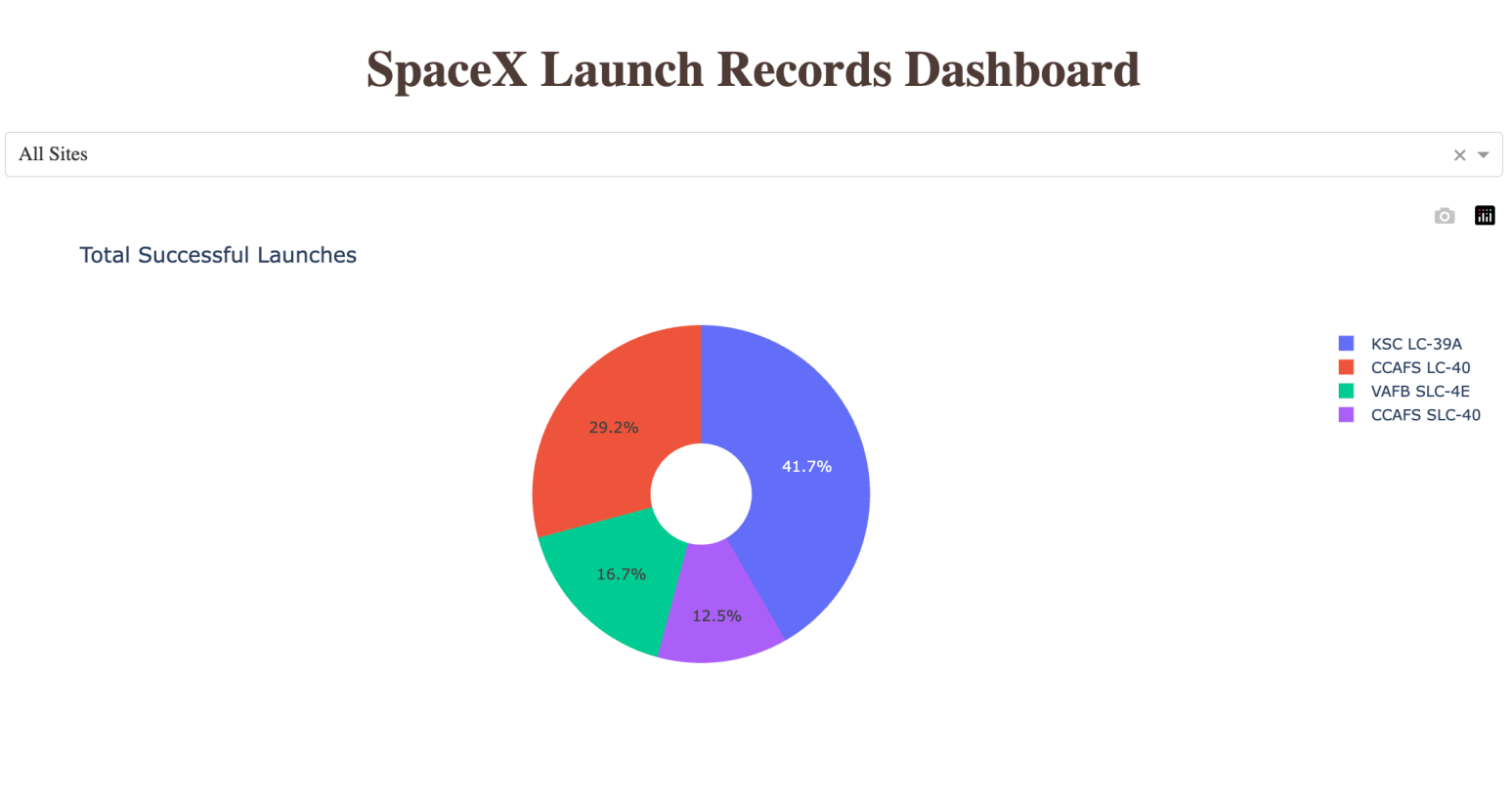
- Lines can be plotted between geographic locations of interest.
- Endpoints can be marked to display distance based on a functional calculation of longitude and latitude.
- Example is distance to Cocoa Beach




Section 4
Build a Dashboard with Plotly Dash

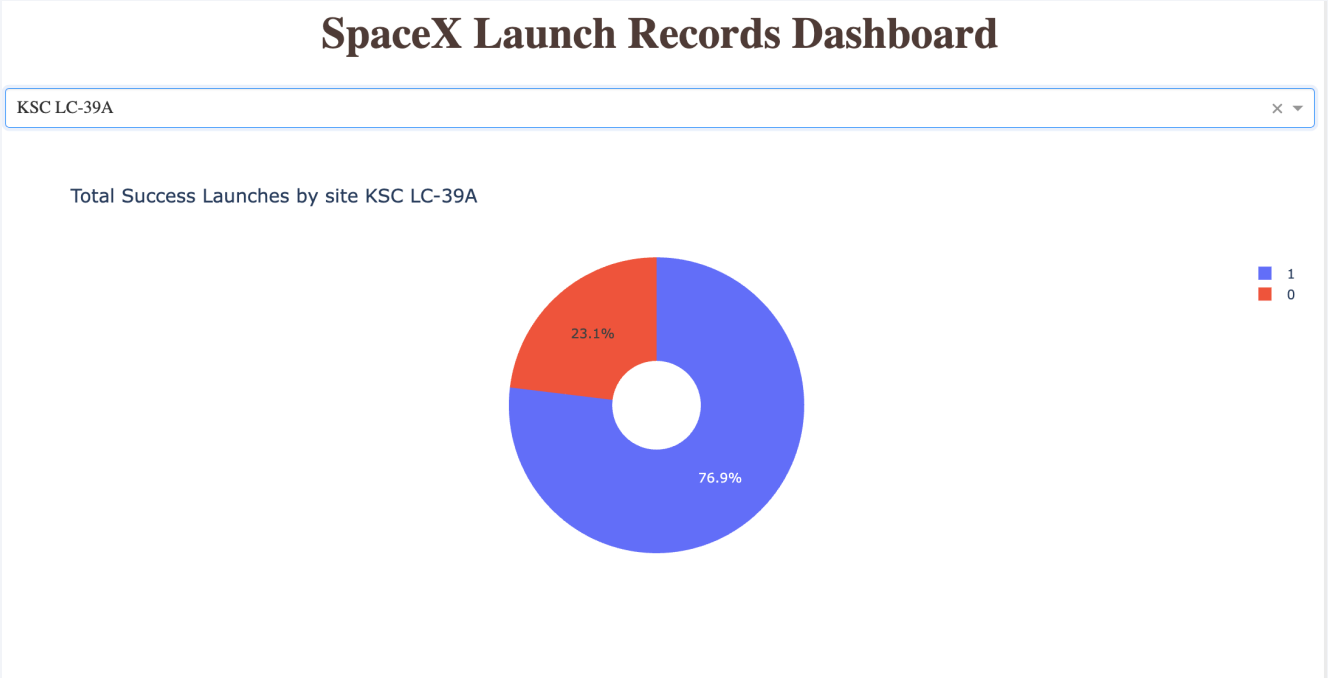
Pie Chart Dashboard (All sites)

- KSC LC-39A has the highest count of launches followed by CCAFS LC-40



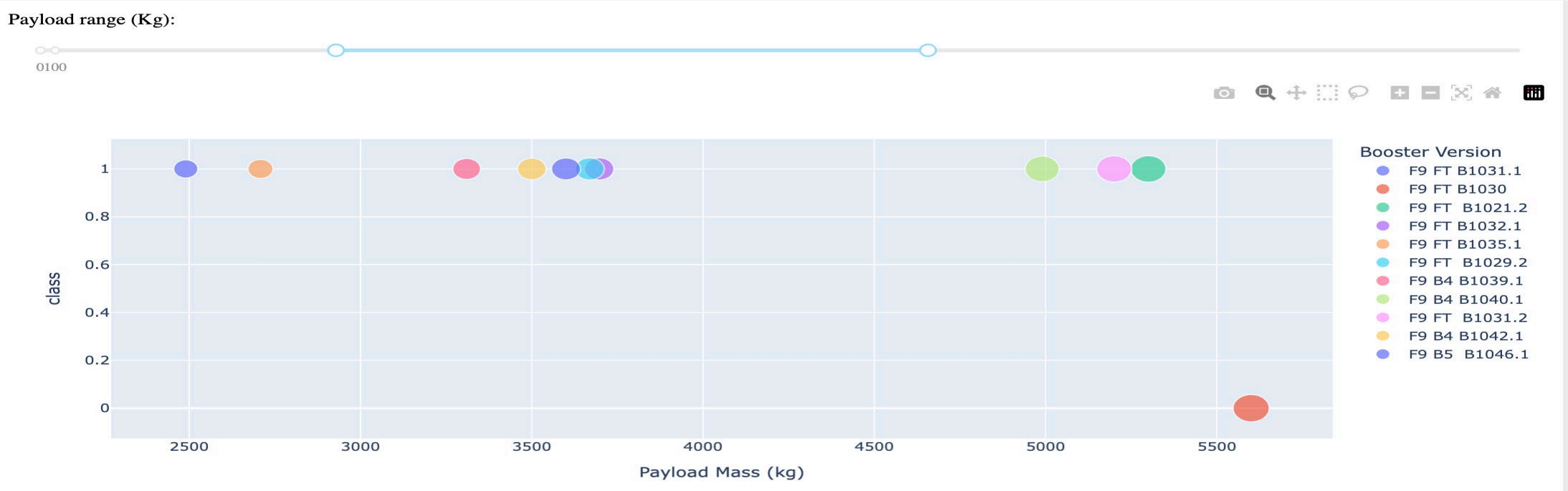
<Dashboard Screenshot 2>

- Of the launches at KSC LC-39A, about 77% were successful launches while 23% were not



Scatter Plot: Payload vs. Launch Outcome

Key Insights: As an example, higher payloads tend to be fewer in frequency at each site with more successful outcomes occurring at the lower payload weights

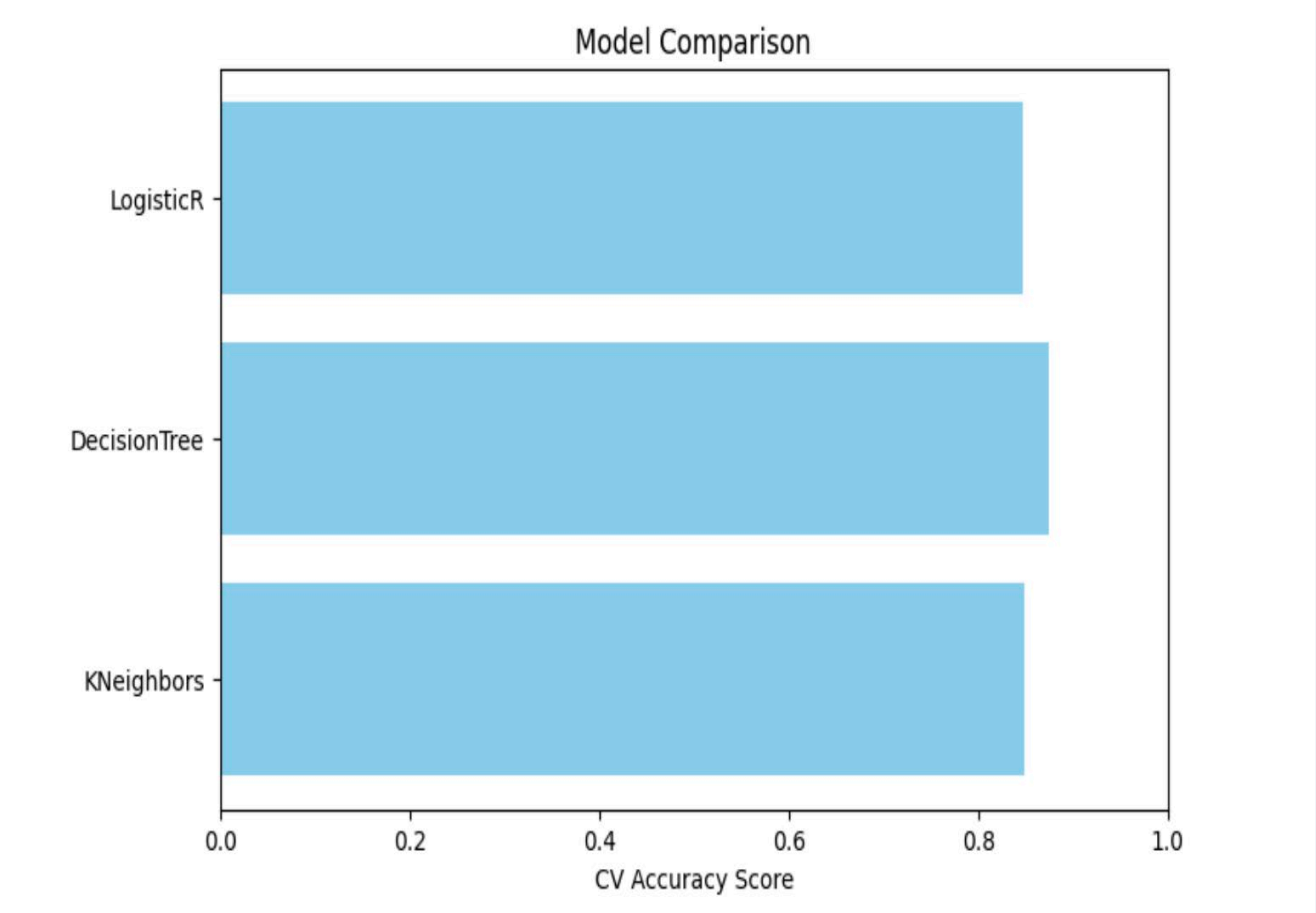


Section 5

Predictive Analysis (Classification)

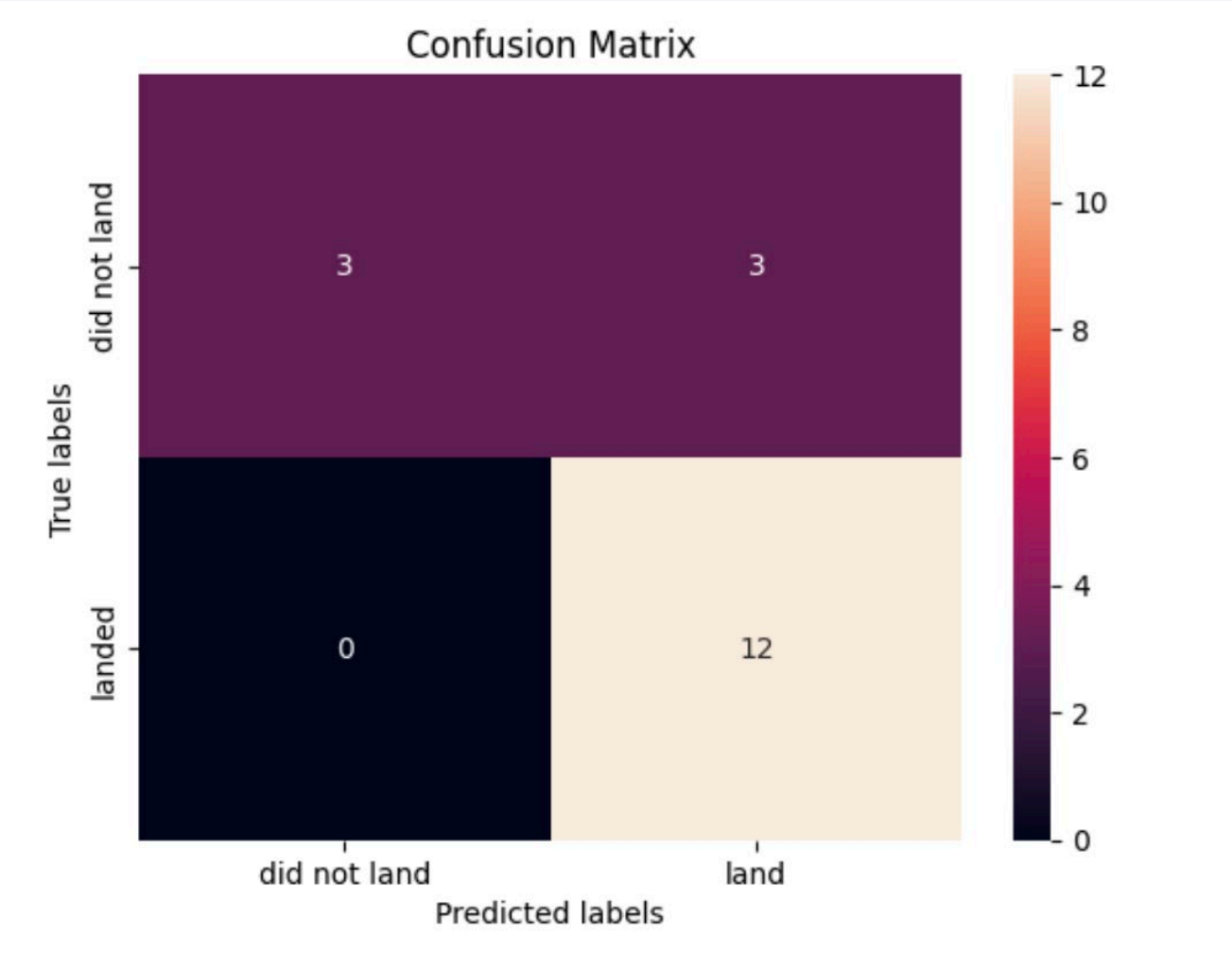
Classification Accuracy

- When determining the best model to use for this data, the decision tree outperformed other models to predict successful landing.



Confusion Matrix

- When assessing the Random Forest confusion matrix, the model was accurate in predicting true successful landings, while also not falsely predicting failures when the landing was a success



Conclusions

- As SpaceX continue to perform flights and landings, the chances of a successful landing consistently increased.
- While some orbit types had a 100% success rate, VLEO appears to have the best chance of success with more samples and may be preferred in later flights for this reason.
- 77% of KSC LC-39A launches were successful making this site the highest success rate.
- Heavy payloads seemed to correlated with fewer failures, however, more samples were present at the lower payload weight

Thank you!