IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James Jackson
<Date>

Outline

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

Executive Summary

- The methods that were used to analyze the data included:
  - Collecting data through the use of web-scraping and the SpaceX API
  - Data wrangling, data visualization, and visual analytics (Exploratory Data Analysis)
  - Machine learning predictions
- Summary of all results
  - The data that was needed was obtained using public sources
  - Exploratory data analysis was used to help show which features were the best indicators of a launch's success or failure
  - The model, that was used to find the best predictions of the features which lead to the highest success rate, was found using machine learning

Introduction

- The goal of this project is to see if SpaceY will be able to compete with SpaceX
- Questions to be answered:
  - What is price of a given launch or how can that be determined?
  - Additionally, will SpaceX reuse the first stage or not?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX data was obtained from two public sources:
    - The SpaceX API
      - https://api.spacexdata.com/v4/rockets/
    - Data from Falcon Launch Wikipedia was web-scraped
      - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Perform data wrangling
  - After analyzing and summarizing the features, a landing outcome label was created based on the outcome data

Methodology

Executive Summary

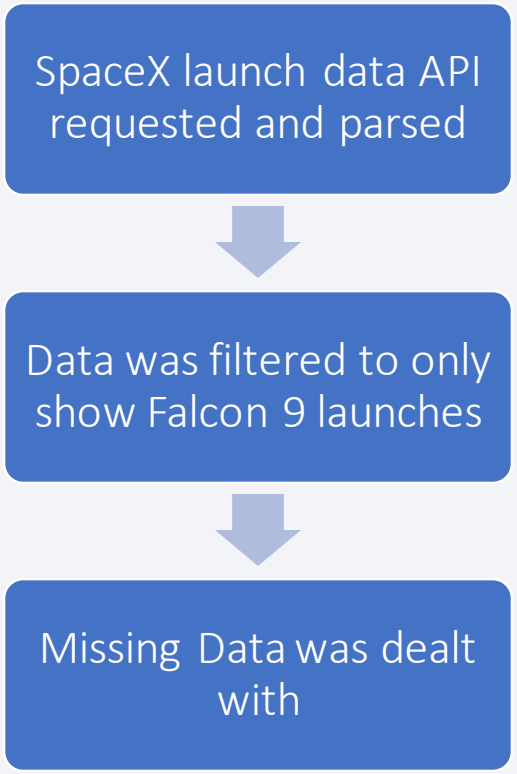
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Collected data was normalized, split into training and test sets, and then evaluated using four different classification models. The accuracy of each model was then evaluated to determine which worked best.

Data Collection

- SpaceX data was obtained from two public sources:
  - The SpaceX API
    - https://api.spacexdata.com/v4/rockets/
  - Data from Falcon Launch Wikipedia was web-scraped
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

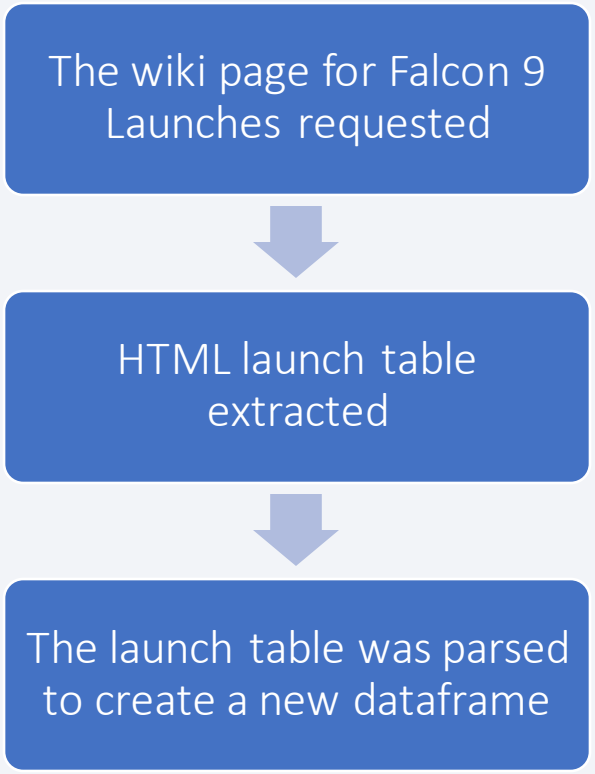
- The public API from SpaceX was one of the sources of data
- The provided flowchart shows basic steps taken with the data
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

- Additional data was obtained from the Falcon 9 launch Wikipedia

- The provided flowchart shows basic steps taken with the data
- Github Source Code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Exploratory data analysis was carried out on the data
- Data was summarized to find launches per site, number launches for each orbit, and how many of each launch outcome occurred for each orbit type.
- Github Source Code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Data%20Wrangling%20Lab.ipynb

EDA with Data Visualization

- Scatter plots were used to show the relationships between flight number and launch site, payload mass and launch site location, flight number and orbit type, and payload and orbit type
- Bar charts were used to how successful each orbit type was
- Line plots were used to show the general launch success trend per year
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Exploratory%20Data%20Analysis%20using%20Visualization.ipynb

EDA with SQL

- The following tasks were performed on the data using SQL queries:
  - Listed the names of the unique launch sites used in the space missions
  - Listed 5 records where the name of the launch site began with 'CCA'
  - Listed the total payload mass carried by NASA (CRS) boosters
  - Found the average payload mass carried by booster version F9 v1.1
  - Found the first date of a successful landing on a ground pad
  - **Continued on next slide**

EDA with SQL (Continued)

- The following tasks were performed on the data using SQL queries:
  - Found the names and corresponding payload mass of boosters that had success in drone ship and had a payload mass between 4000 and 6000 kg
  - Listed the total number of successful and failed missions
  - Listed the booster versions which carried the maximum payload mass
  - Listed the date, booster version, and launch site for failed landings in drone ships
  - Sorted the counts of landing outcomes between the dates of 2010-06-04 and 2017-03-20 in descending order
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/EDA%20with%20SQL%20lab.ipynb

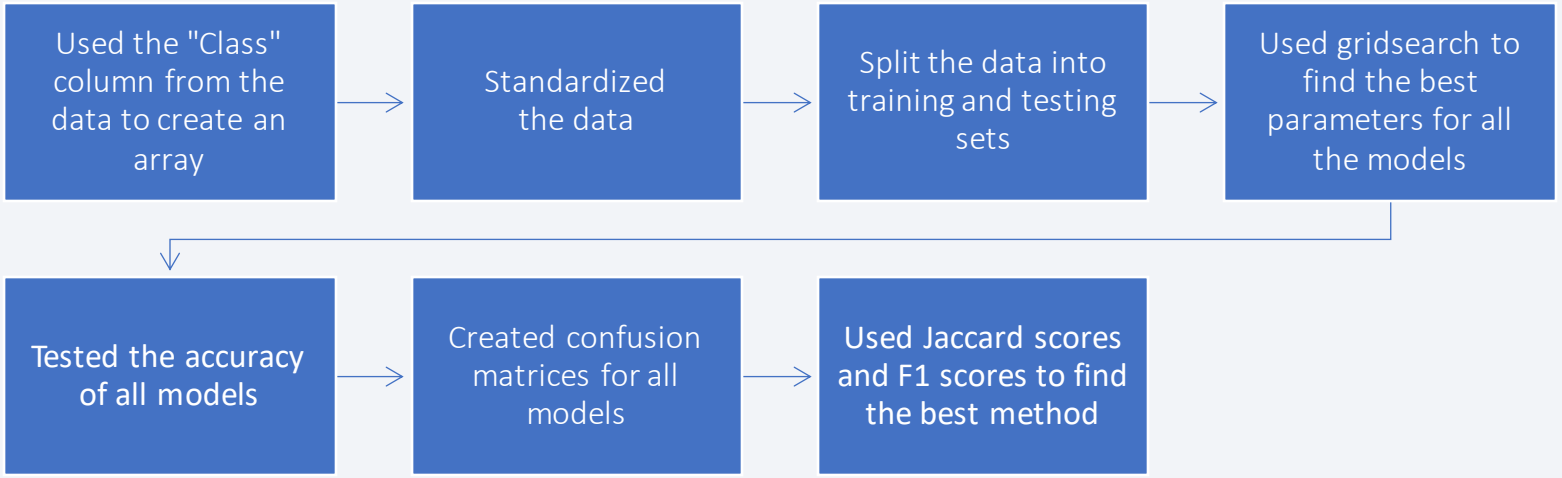
Build an Interactive Map with Folium

- Markers, circles, lines, and marker clusters were created and added to the folium map of the United States
- Markers were added in order to label the launch sites
- The purpose of the circles is to define the area that these launch sites occurred
- The lines show the distance from a launch site to the given coastline
- Marker clusters were used to show and differentiate between the successful and failed launches from a launch site
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- A dashboard was created which shows the following graphs:
  - A pie chart that can show either the percentage of launches each launch site has, or the percentage of successful and failed launches for a given launch site.
  - A scatter plot that lets you compare different launch factors with each other for a given payload range.
- These charts were added to show how different combinations of launch factors and sites affect launch success.
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)



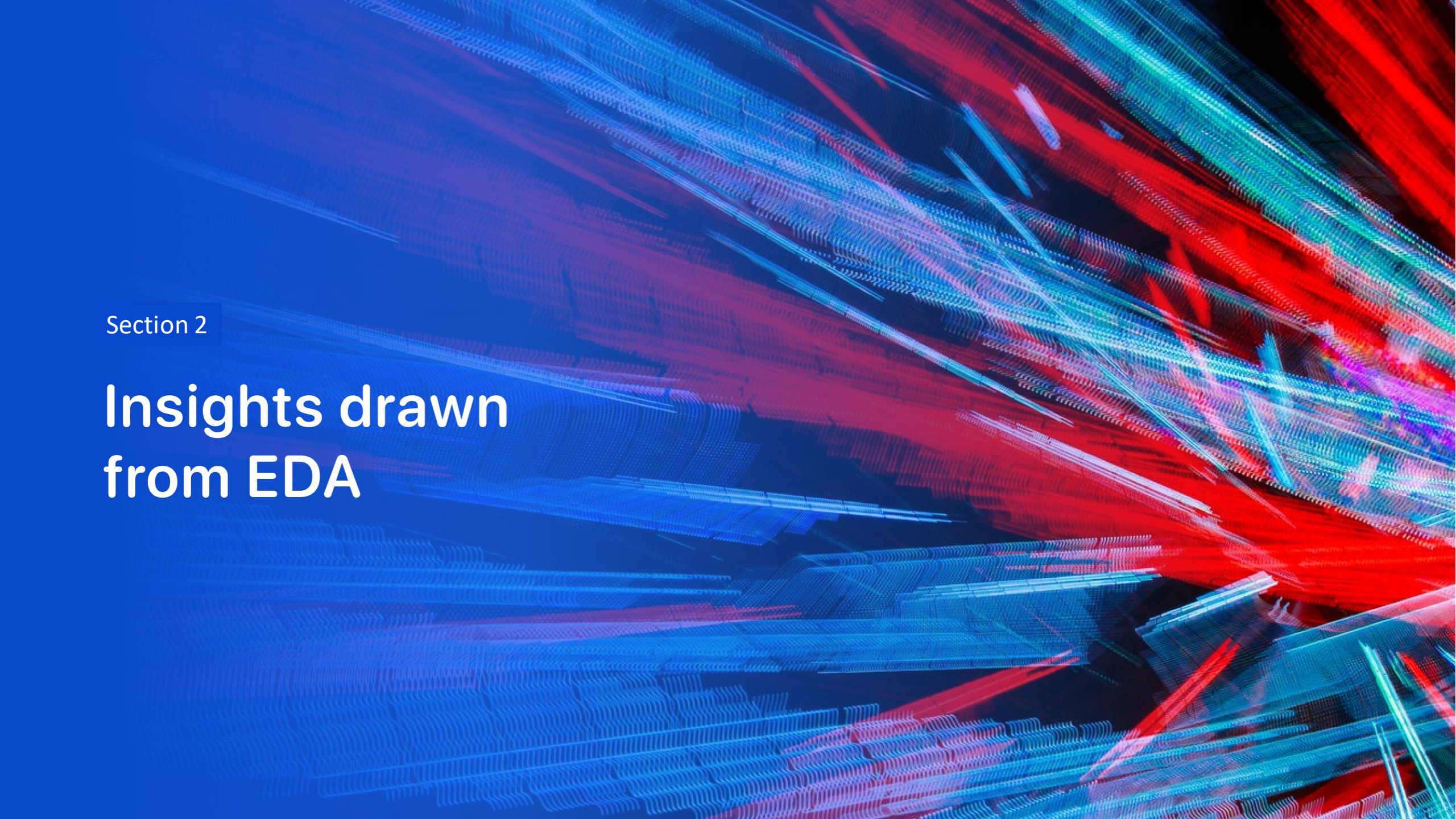
- Github source code:
  - https://github.com/J2Nature/Applied-Data-Science-Capstone/blob/main/Machine%20Learning%20Predictions.ipynb

Results

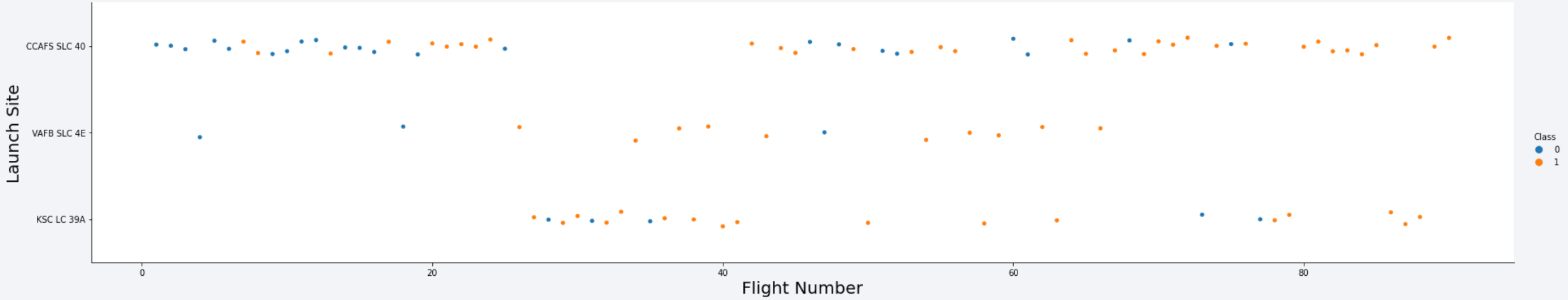
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

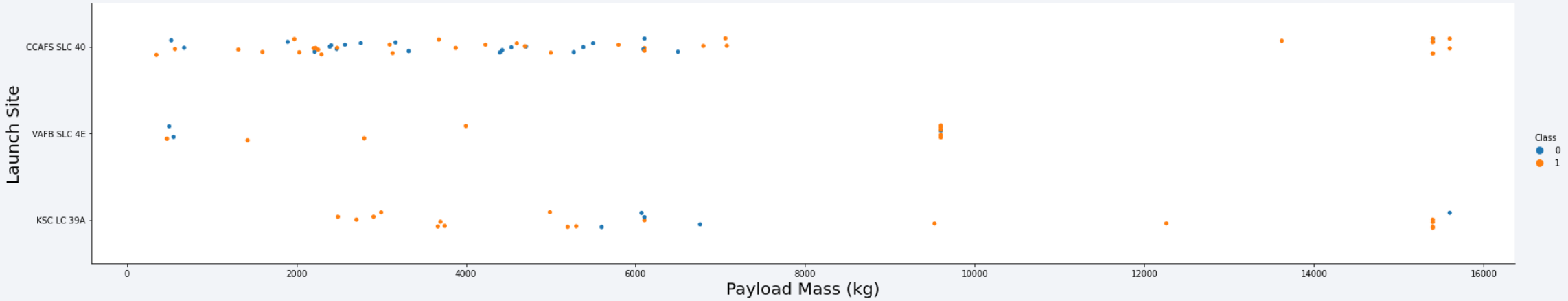


Flight Number vs. Launch Site



- The CCAFS SLC-40 site appears to be the first site that the rockets were launched from and also has the lowest success rate.

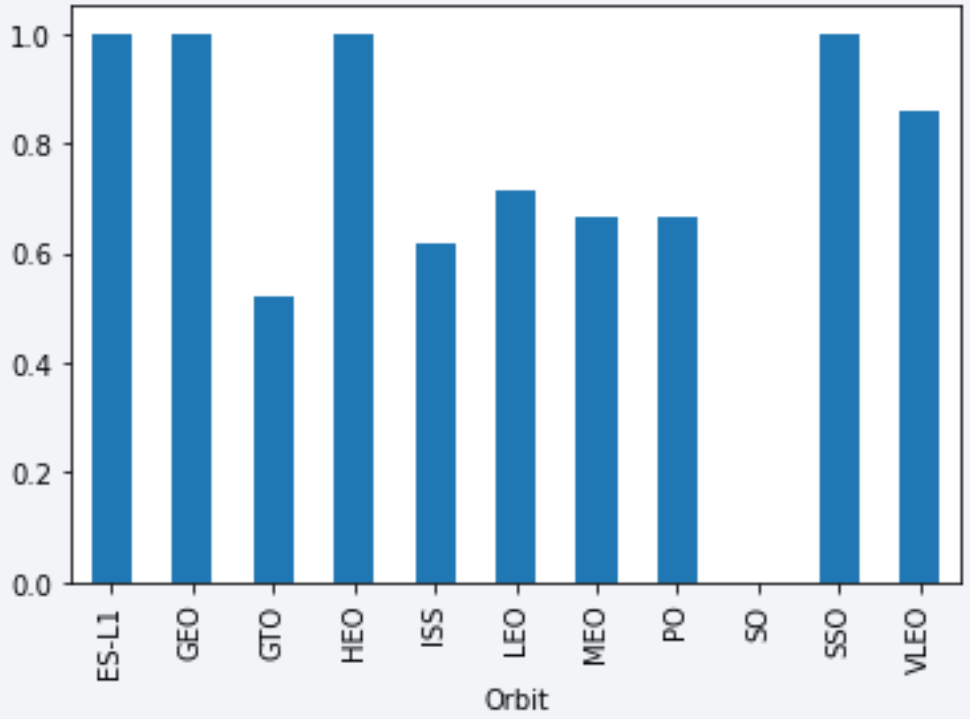
Payload vs. Launch Site



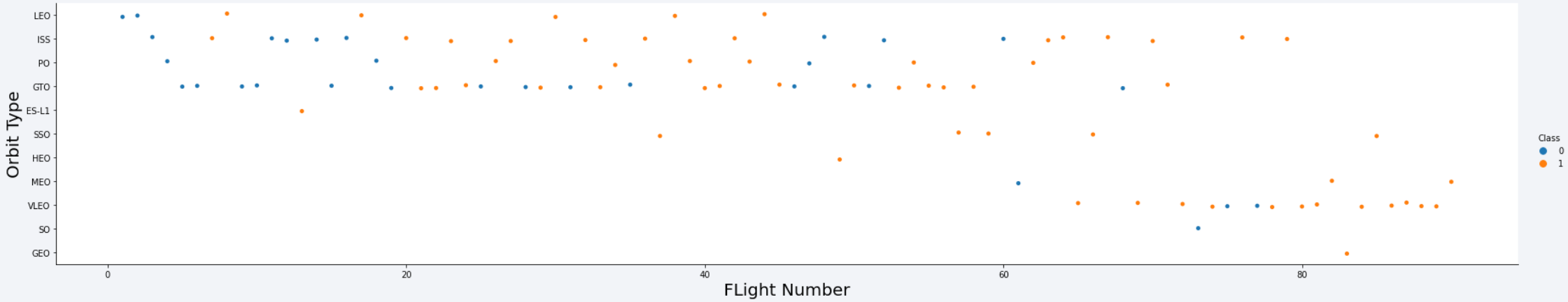
- The launch sites CCAFS SLC-40 and KSC LC-39A appear to be used more for heavier payloadds than VAFB SLC-4E, but that might just be because the VAFB site has fewer launches.

Success Rate vs. Orbit Type

- The orbit types with the highest success rates include ESL1, GEO, HEO, and SSO.
- The orbit type with the lowest success rate is GTO.

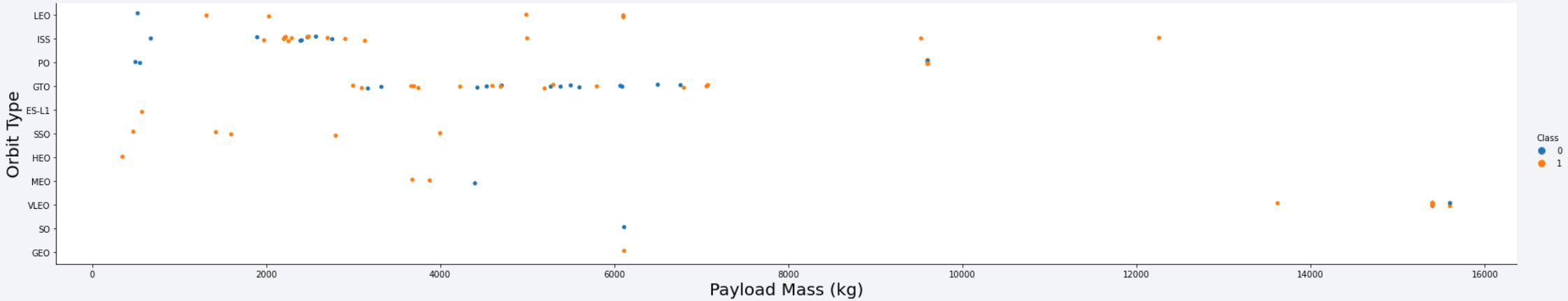


Flight Number vs. Orbit Type



- There is a positive correlation between flight number and first stage recovery for all orbit types.

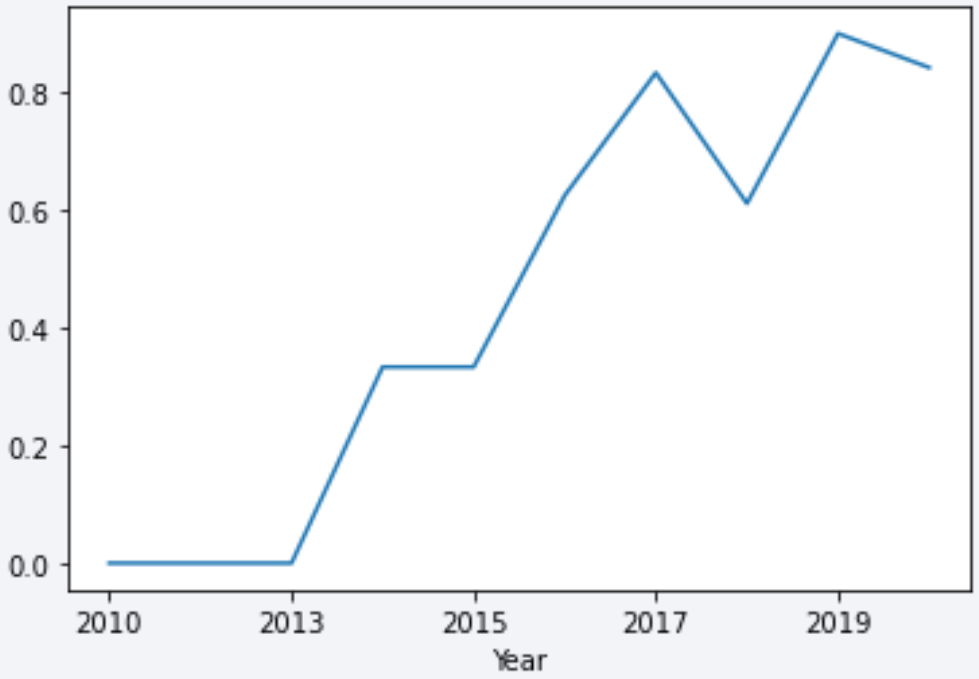
Payload vs. Orbit Type



- GTO orbits have a lower success rate with heavier payloads, while ISS orbits appear to have a higher success rate with heavier payloads.

Launch Success Yearly Trend

- This graph shows that the yearly trend in the launch success rate overall has been increasing since 2013, but has started to fluctuate since the year 2017.



All Launch Site Names

- The names of four unique launch sites were found in the SpaceX data:

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

- These names were found by using SQL to select the unique launch sites in the data

Launch Site Names Begin with 'CCA'

- 5 of the records where launch sites begin with `CCA` include:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- This was accomplished by using SQL to select the results where the name was like "CCA" and limiting the results to 5

Total Payload Mass

- Total payload carried by boosters from NASA:

total_payload_kg
45596

- This was accomplished by using SQL to add up all the payload masses for boosters where the customer was like NASA

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is:

booster_version	avg_payload_mass
F9 v1.1	2928

- This was accomplished by using SQL to select the average payload mass from the data that was grouped by booster version and using only the results for the F9 v1.1 booster

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad:
  - 2010-06-04
- This was found by using SQL to select the first date where the mission outcome was a success

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

- This was accomplished by using SQL to select the booster version and payload mass from the data where the landing outcome was a success and the payload mass was between 4000 and 6000 kg

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

mission_outcome	quantity
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- This was accomplished by using SQL to select the mission outcomes and counting the results from the data that was grouped by mission outcome

Boosters Carried Maximum Payload

- The names of the boosters which have carried the maximum payload mass are in the chart to the right.
- This was accomplished by using SQL to select the booster versions from the data where their payload mass equaled the maximum payload mass for the data

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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

- This was accomplished by selecting the date, booster version, and launch site from the data where the landing outcome was a failure and the year equaled 2015.

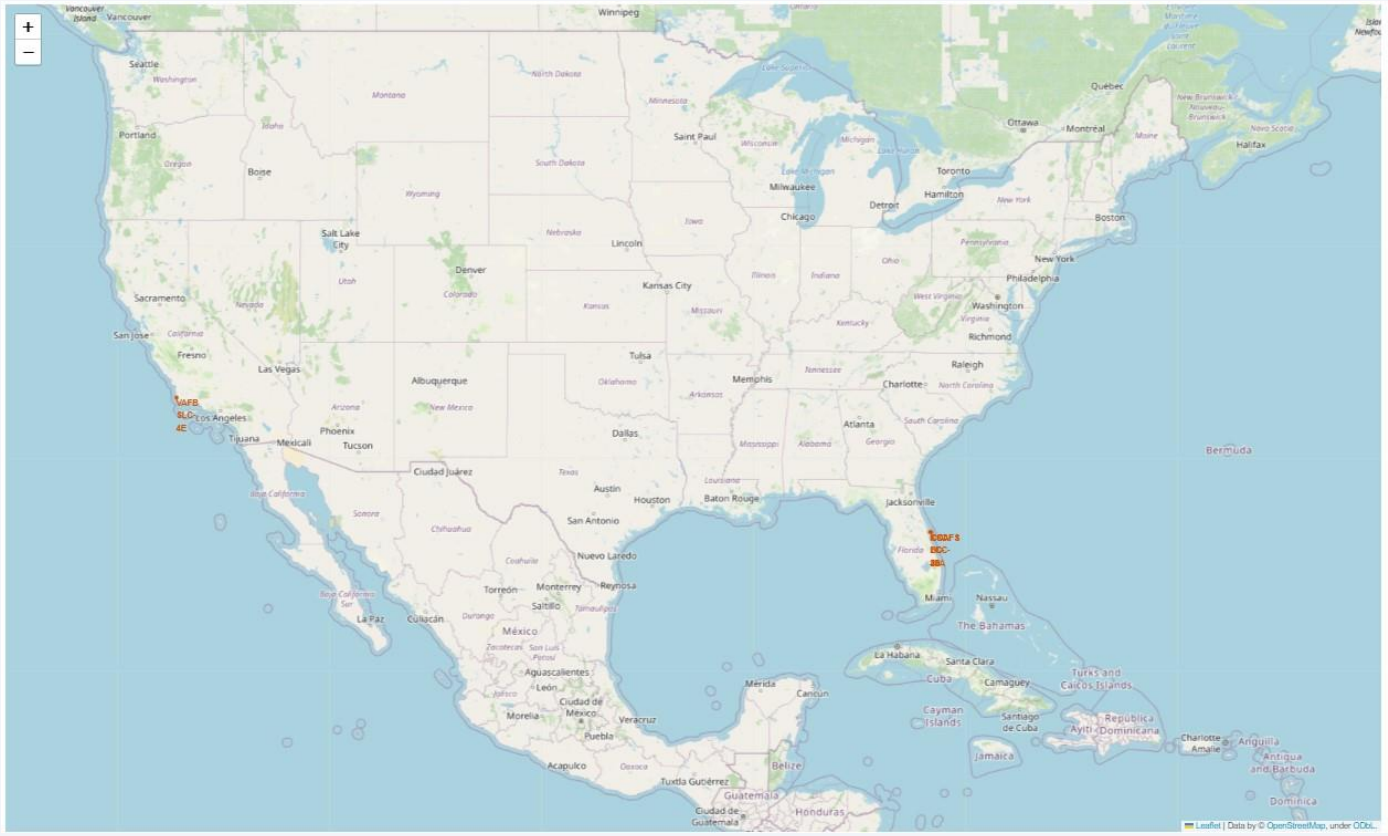
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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.
- This was accomplished by using SQL to select the landing outcome and counting the results from the data grouped landing outcome and ordered by quantity where the date was between the two before mentioned dates.

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

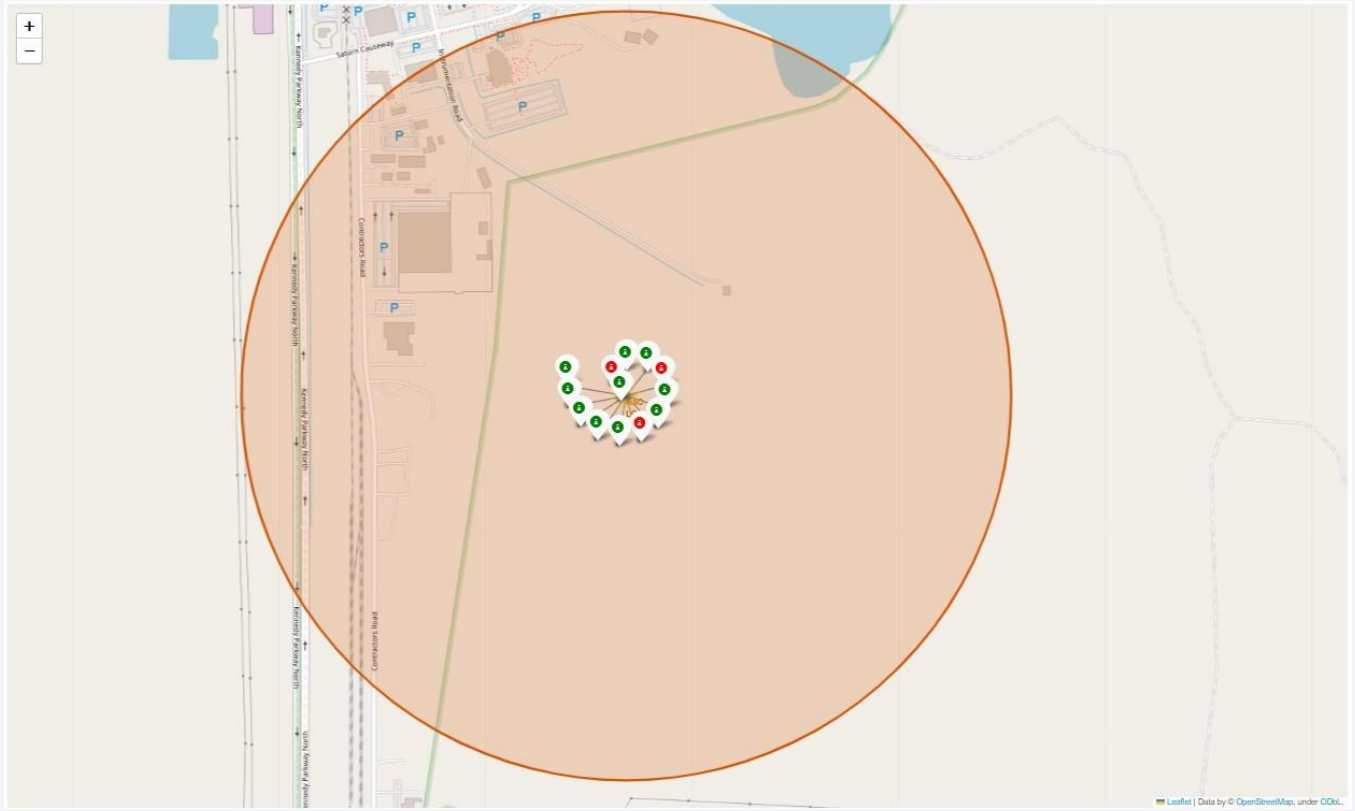

Section 3
Launch Sites
Proximities Analysis

Locations of all Launch Sites



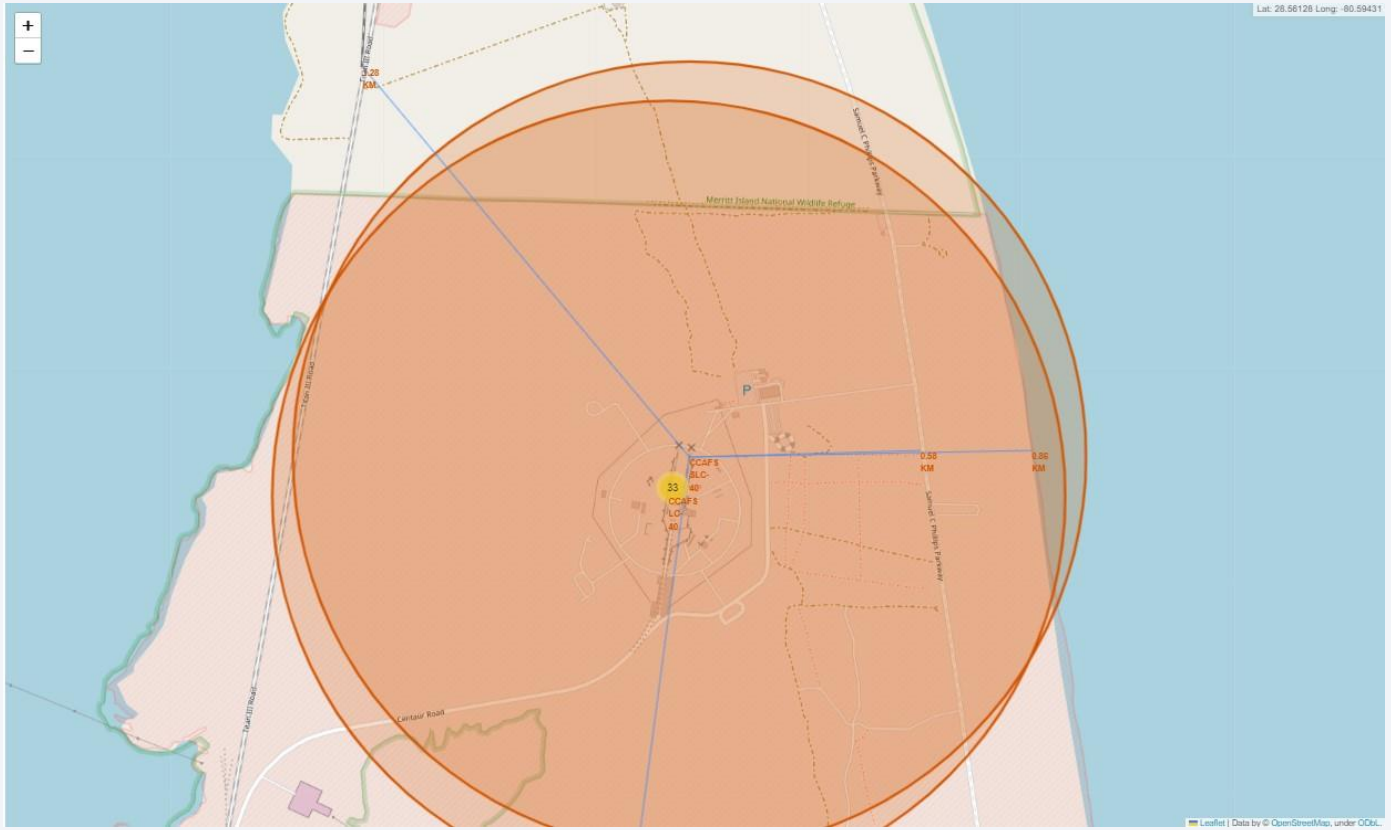
- Three of the launch sites used are on the east coast of the east coast of the USA and one is on the west coast.

Map of Launch Outcomes for Each Site

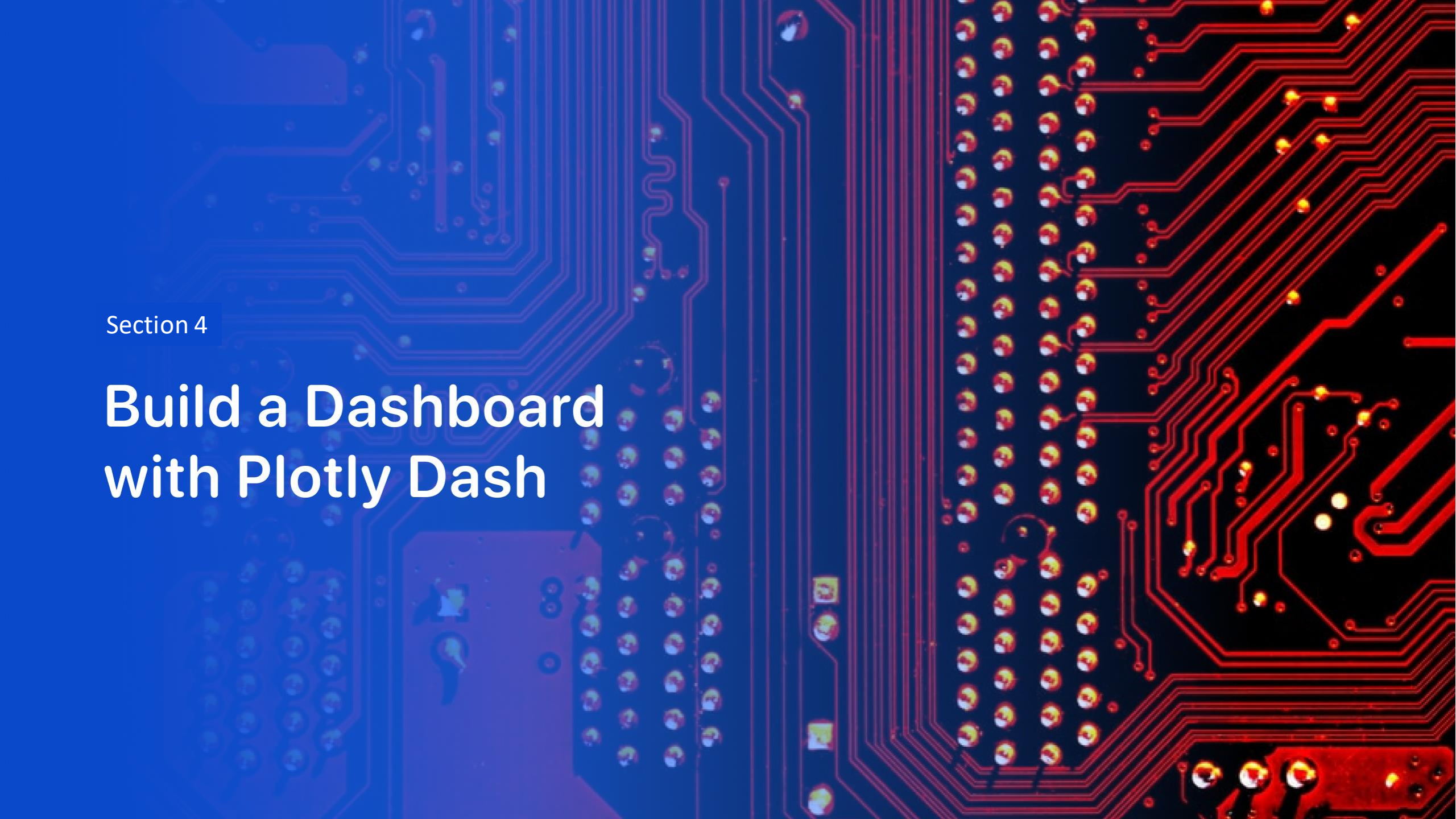


- The launch site shown here as an example is KSC LC-39A. The green markers are successful launches while the red markers are failed launches.

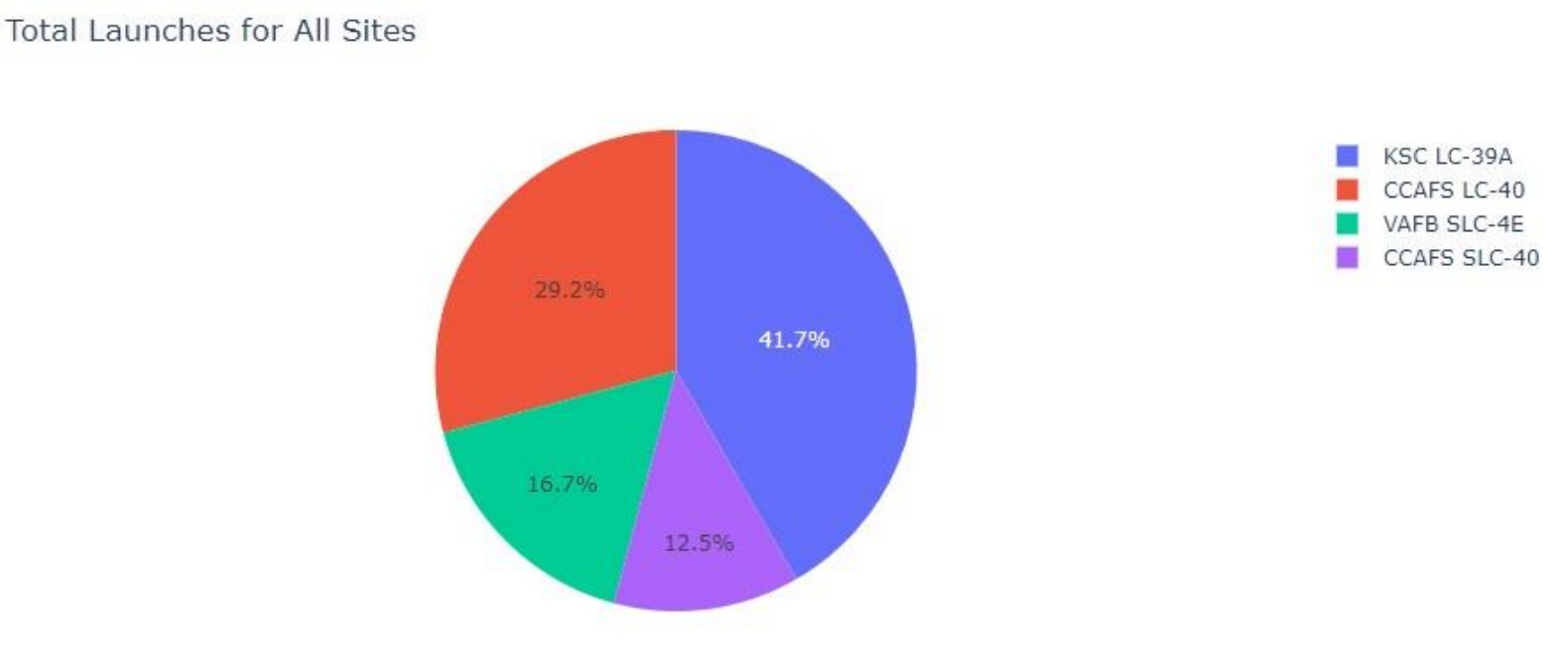
Proximity of Launch Sites to Transportation, etc



- Launch sites CCAFS LC-40 and CCAFS SLC-40 are close to the coast, railroad, and road, but are farther away from the airport.

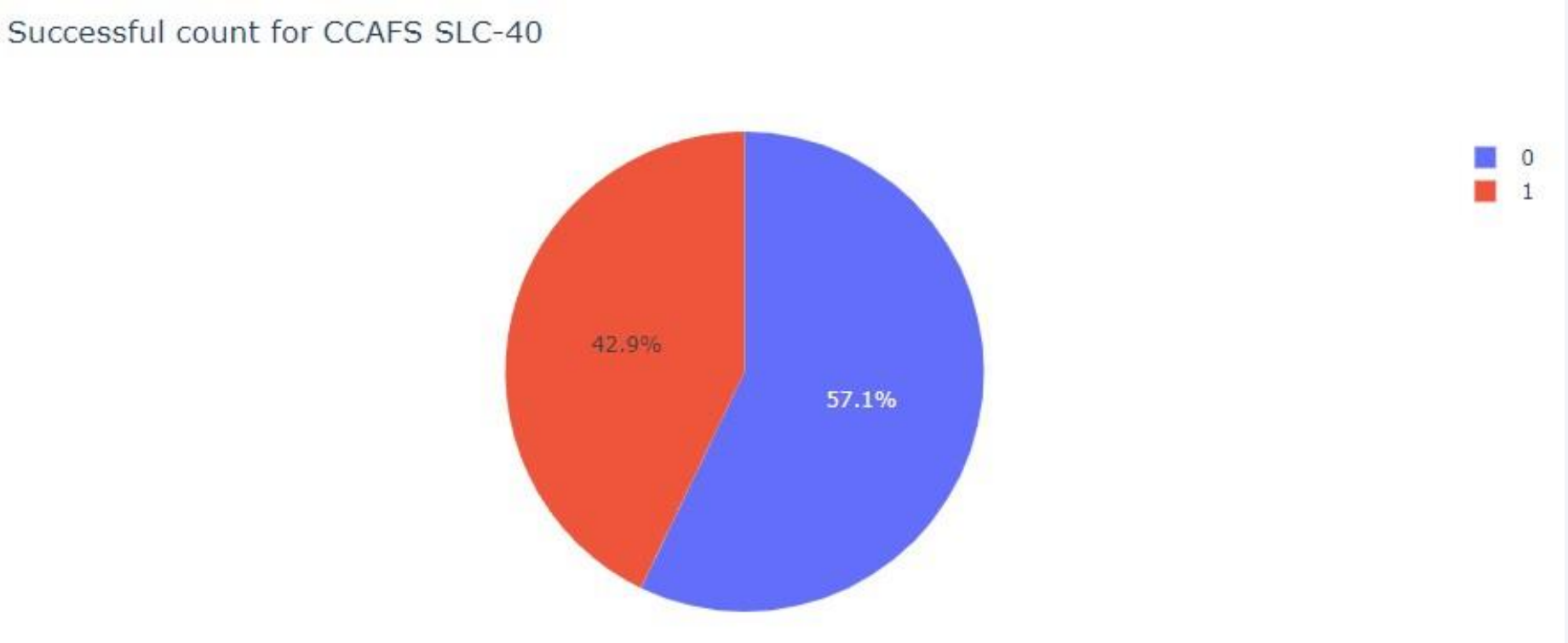

Section 4
Build a Dashboard with Plotly Dash

Total Launches for All Sites



- The pie chart shown above displays the percentage of all launches for each launch site.
- Most of the launches come from the KSC LC-39A site with 41.7% of the launches.
- The launch site with the least amount of launches is the CCAFS SLC-40 site with 12.5% of the launches.

Launch Site with Highest Success Ratio



- The pie chart shown above displays the percentages for successful and failed launches from the CCAFS SLC-40 site.
- The number of successful launches, shown in red, is 42.9%.
- The percentage of failed launches for this site, shown in blue, is 57.1%.

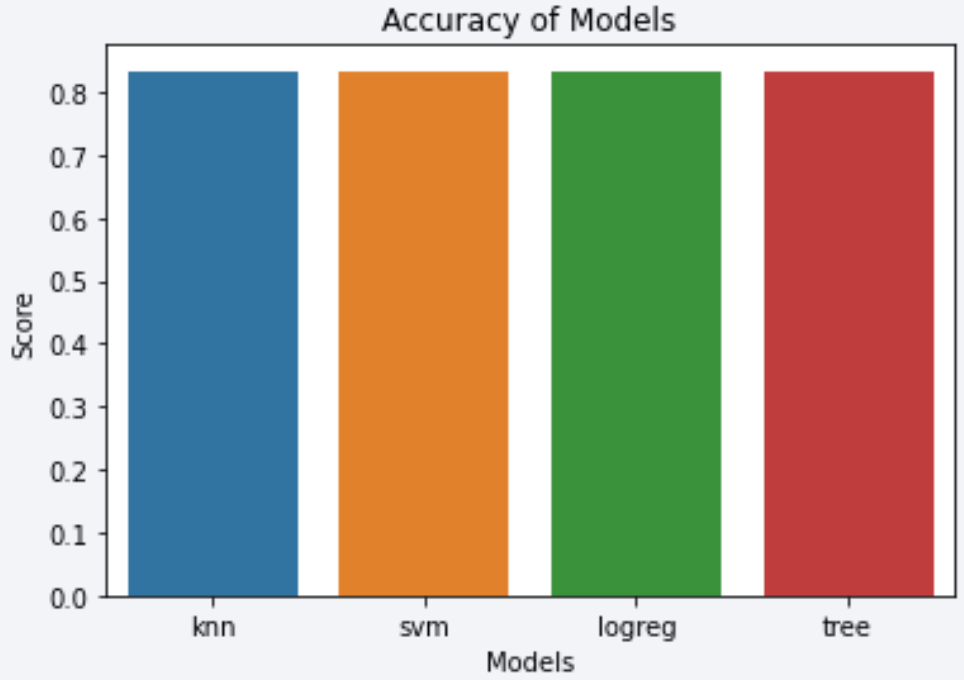
Payload Range with the Highest Success Rate

- The scatter plot shown above displays the success rate for booster versions carrying varying payload masses in kilograms ranging from 2500 kg to 5000 kg.
- This range of payloads has the best launch success rate than the other 2500 kg intervals.

Section 5

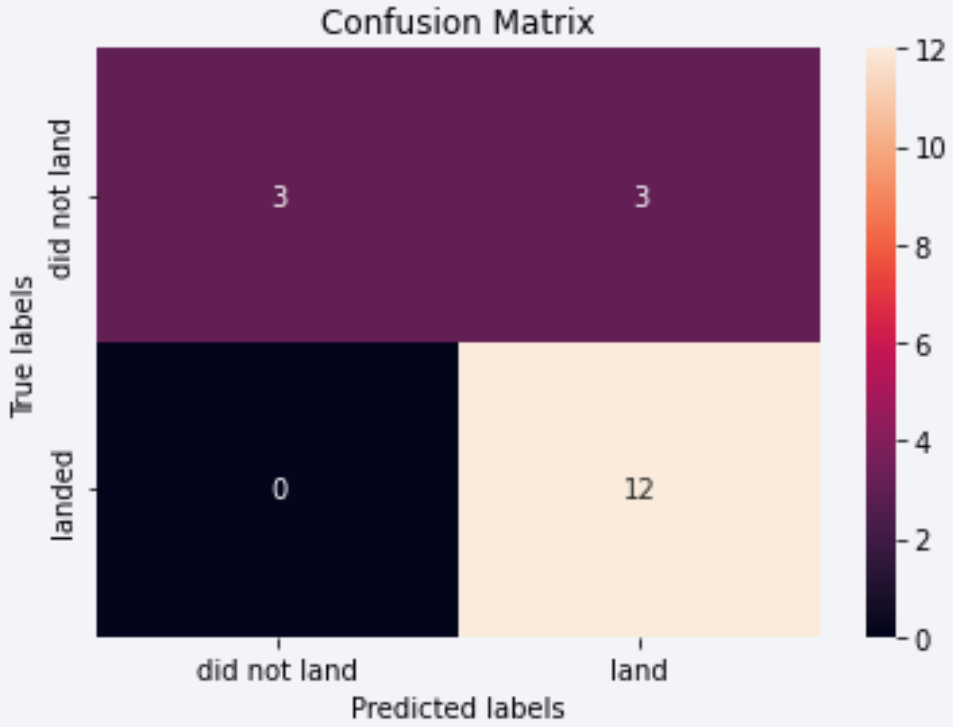
Predictive Analysis (Classification)

Classification Accuracy



- All models tested had an accuracy score of 0.8333...

Confusion Matrix



- This is the confusion matrix for the decision tree classifier. It shows that the accuracy of this model is pretty high because the true positive results are much higher than the false positives. However, the false negative and true negative results are equal.

Conclusions

- All models have and 83.333% accuracy, so any of the four models will work.
- All the launch sites are close to the coast and most are close to the equator.
- Lighter payloads tend to have a better success rate than heavier payloads.
- As expected, the launch success rate has increased over time.
- Orbits that haven't failed yet include GEO, HEO, SSO, and ES-L1.
- The launch site with the highest success rate is the KSC LC-39A site.

Thank you!