IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adam Cersosimo
06/08/2002

Outline

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

Executive Summary

- **Summary of methodologies:**

Collected data using web scraping

Data wrangling and data visualization

Utilized machine learning techniques to make predictions

- **Summary of all results:**

Used accessible data to create visualizations that identify what factors play a key role in a successful launch

Apply knowledge gathered to create machine learning predictions

Introduction

- **Objective:**

We will determine if a new company, Space Y, can compete with Space X

Need to see what drives success and profitability

- **Problems you want to find answers**
  - We are predicting the success rate of the first stage of the rocket launch, where part of the rocket lands successfully and can be reused
  - What locations are optimal for launches

Section 1

Methodology

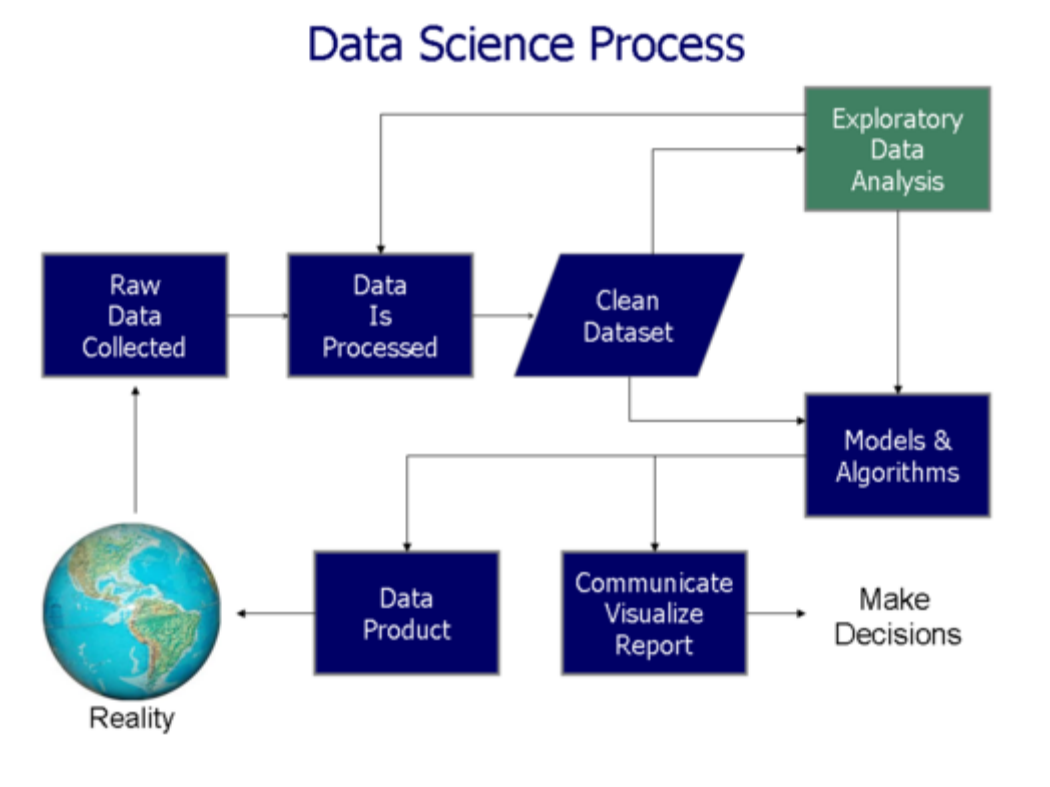
Methodology

Executive Summary

- Data collection methodology:
  - Space X API
  - Webscraping from Wikipedia
- Perform data wrangling
  - Created an outcome label based on the outcome and feature data
- 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

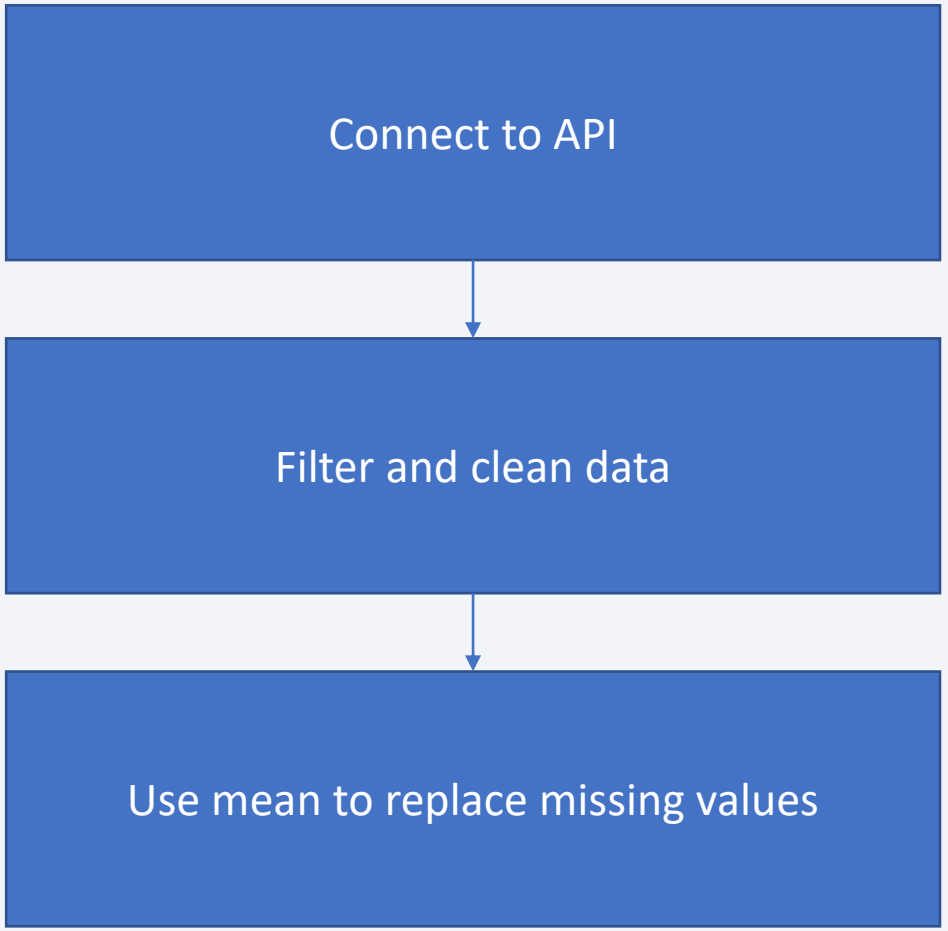
Data Collection

- Used Space X API (https://api.spacexdata.com/v4/rockets/) and Wikipedia



Data Collection – SpaceX API

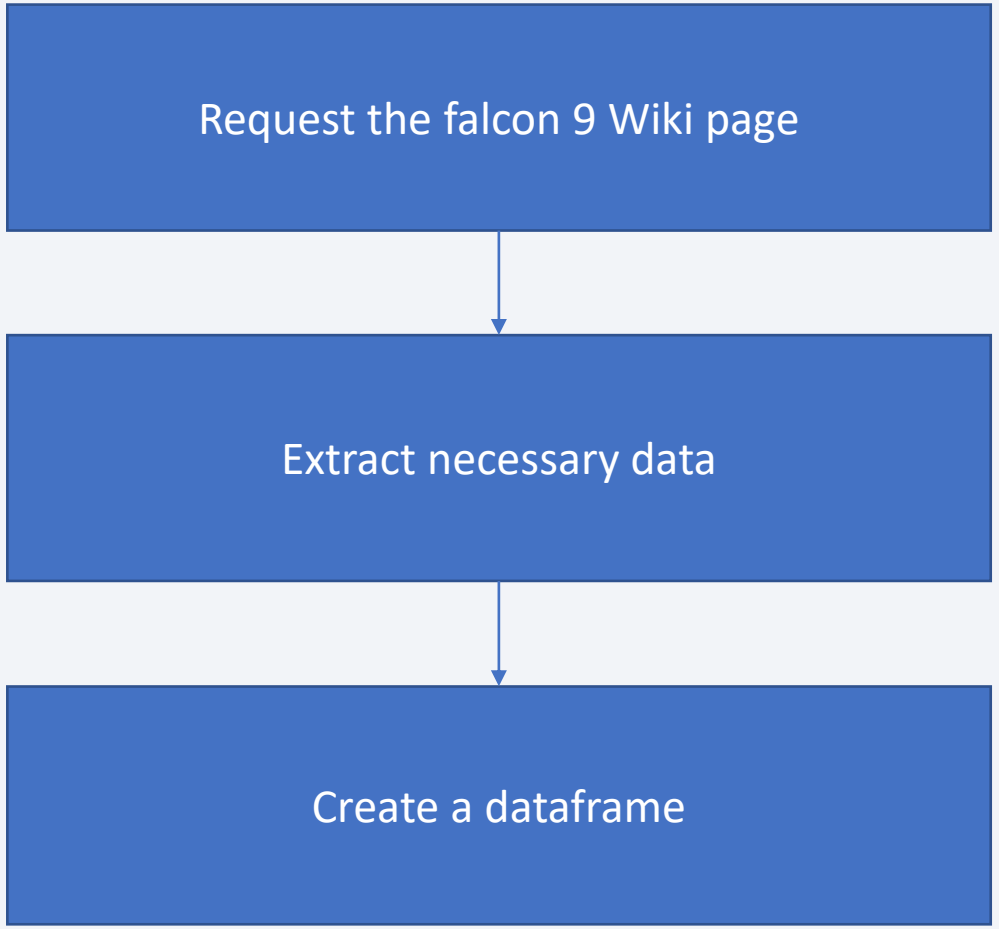
- The data is acquired from the API and Wikipedia
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

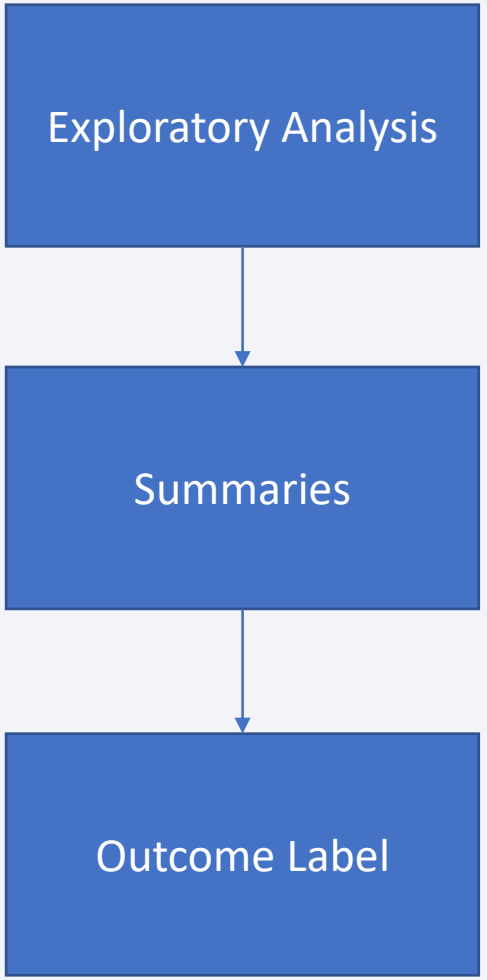
- Use Wikipedia to scrape the data necessary

- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/jupyter-labs-webscraping.ipynb



Data Wrangling

- Performed Exploratory Analysis
- Explore and determine summaries of the launches
- Created an outcome label based on the landing outcome
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb



EDA with Data Visualization

- Used Scatterplots to display different views of the data
- Some of the features visualized together are Payload Mass and Flight Number, Launch Site and Flight Number, Launch Site and Payload Mass, Orbit and Flight Number, Payload and Orbit
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA’
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in the ground pad was achieved.
- List the names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failed mission outcomes
- List the names of the booster versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing outcomes in drone ship, booster versions, and launch site for the months in the year 2015.
- Rank the count of successful landing outcomes between the dates 04-06-2010 and 20-03-2017 in descending order.

- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

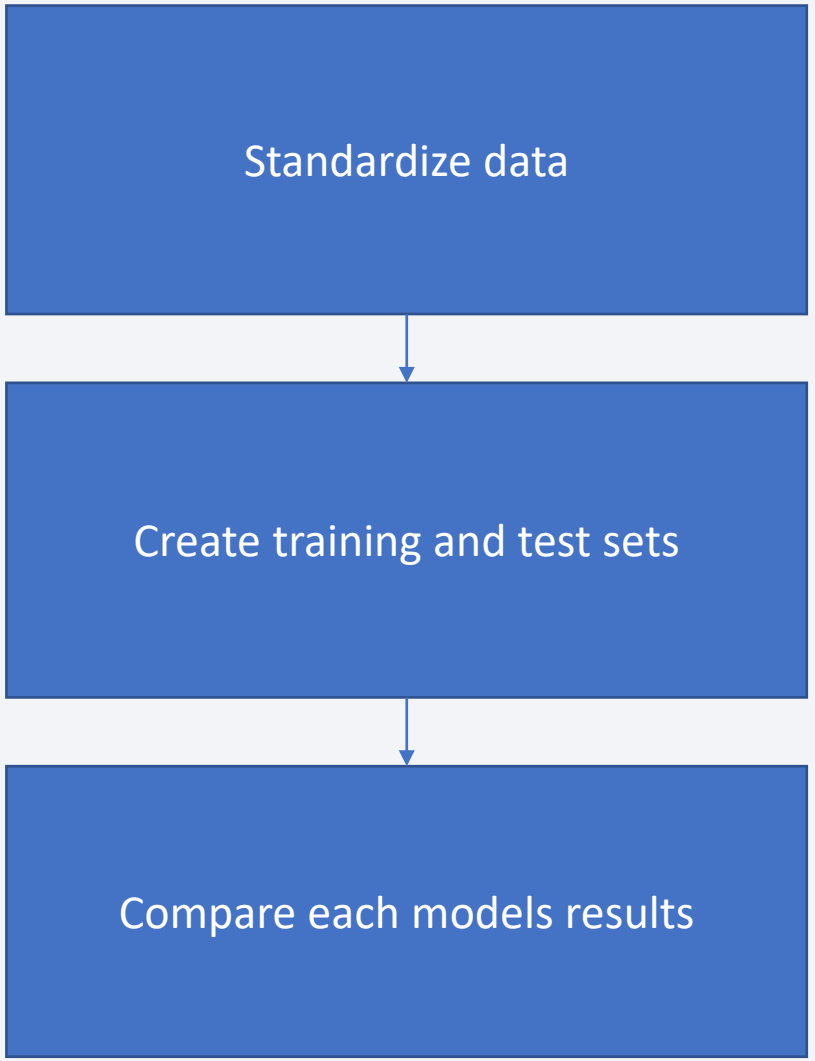
- Markers, circles, lines, and marker clusters were used with Folium Maps to:
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The data was visualized using the following graphs and plots:
- Percentage of launches by site
- Payload range
- This combination facilitated a rapid analysis of the relationship between payloads and launch sites. It helped in identifying the optimal launch site based on the payload requirements.
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Predictive Analysis (Classification)

- A comparison was conducted among four classification models, namely logistic regression, support vector machine, decision tree, and k-nearest neighbors.
- You need present your model development process using key phrases and flowchart
- https://github.com/adacersos/IBM-Data-Science-Capstone/blob/64e2f5f1c48255b48a9b76ddd8314ec2682d8b10/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

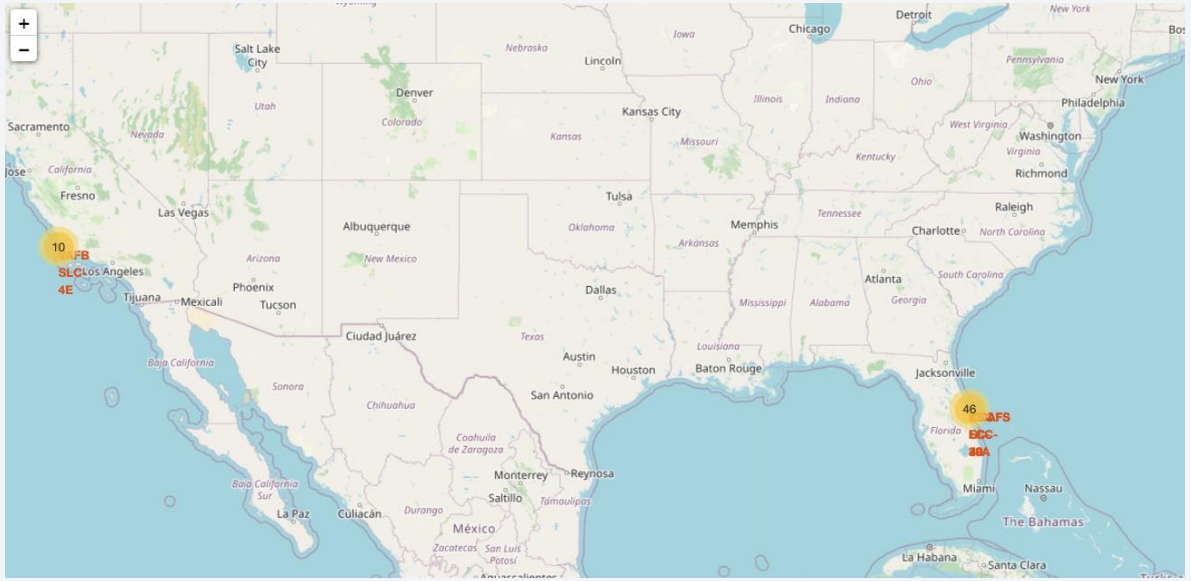


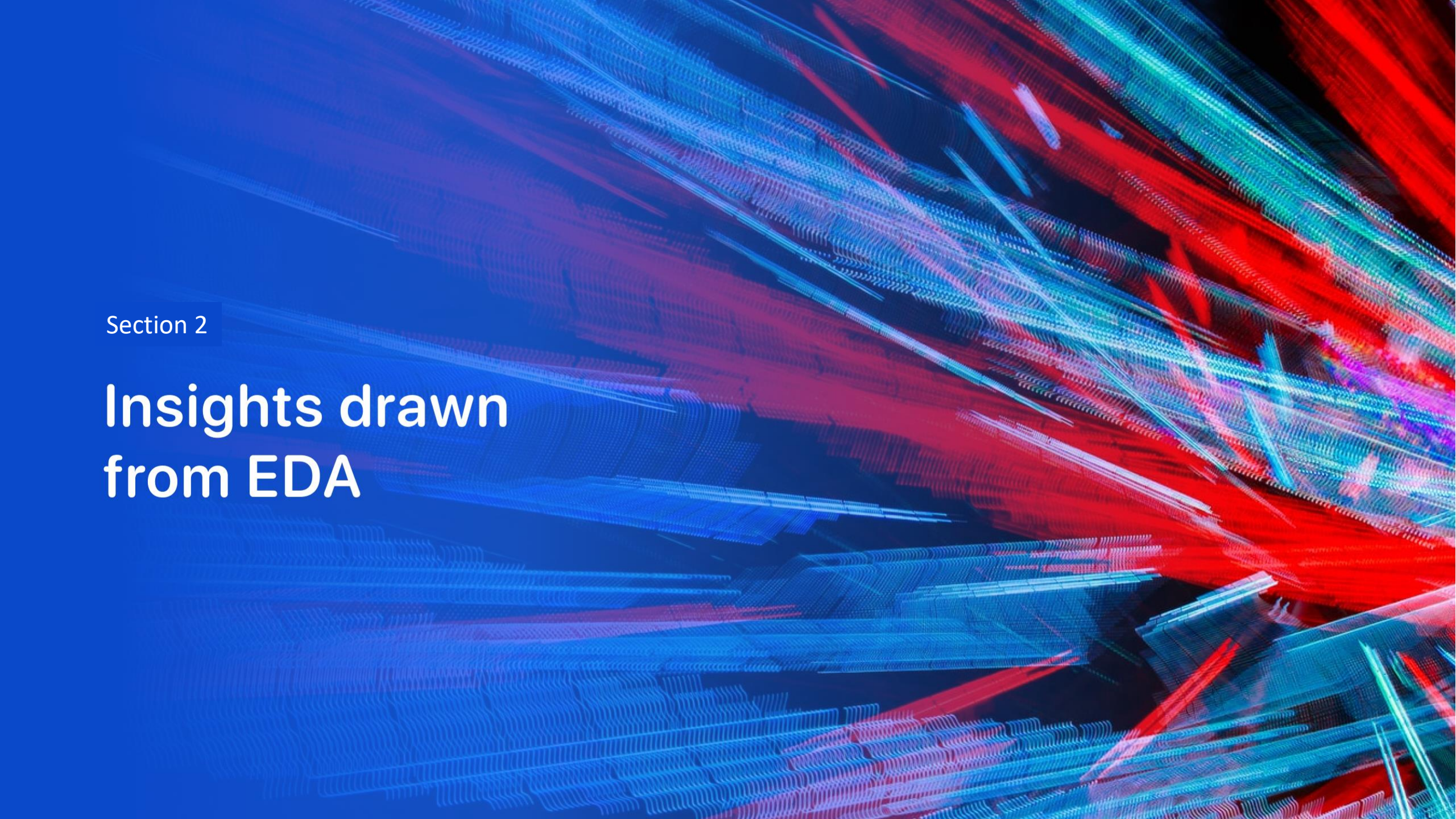
Results

Exploratory data analysis results:

Space X utilizes 4 different launch sites.
The initial launches were conducted by Space X itself and NASA.
The average payload of the F9 v1.1 booster is 2,928 kg.
Several Falcon 9 booster versions achieved successful landings on drone ships with payloads exceeding the average.
Two booster versions, F9 v1.1 B1012 and F9 v1.1 B1015, failed to land on drone ships in 2015.
The number of successful landing outcomes improved over the years.

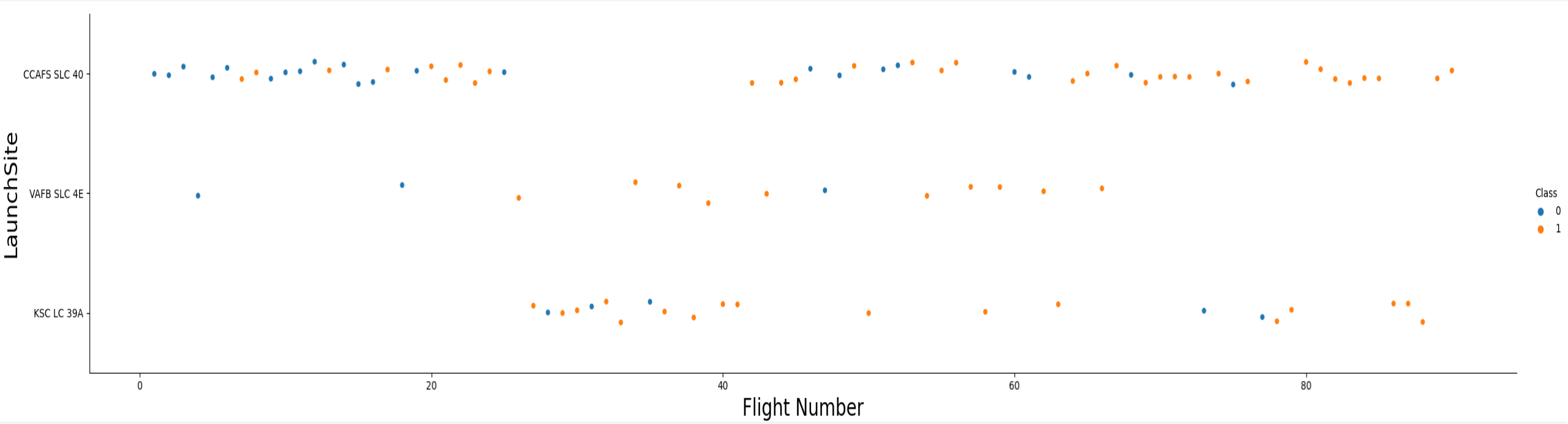
The predictive analysis highlights the decision tree classifier as the best option with a test accuracy of 83.3%




Section 2
Insights drawn from EDA

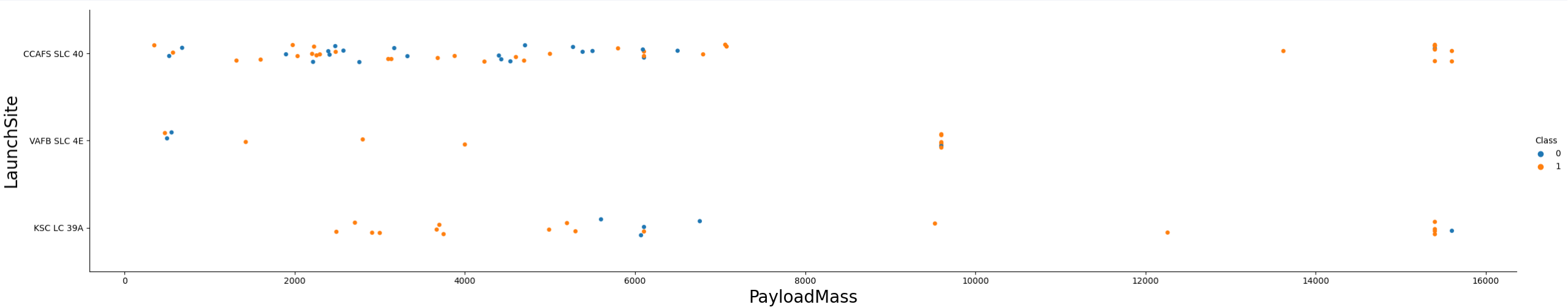
Flight Number vs. Launch Site

- By far the vast majority of the launches were from CCAF5 SLC 40, and recently it has been a very successful location
- The overall success rate improved over time.



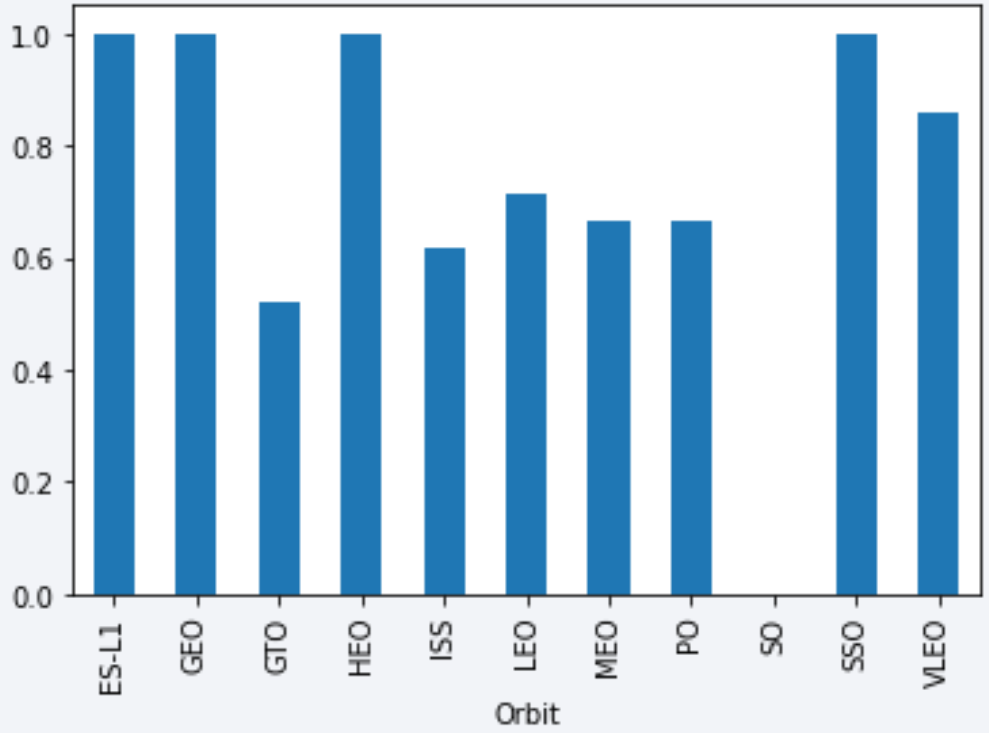
Payload vs. Launch Site

- Heavy payloads have been much more successful in comparison to lighter ones
- It seems as if VAFB SLC 4E is only capable of payloads under 10000 kg
- KSC LC 39A does not have any launches under 2000 kg



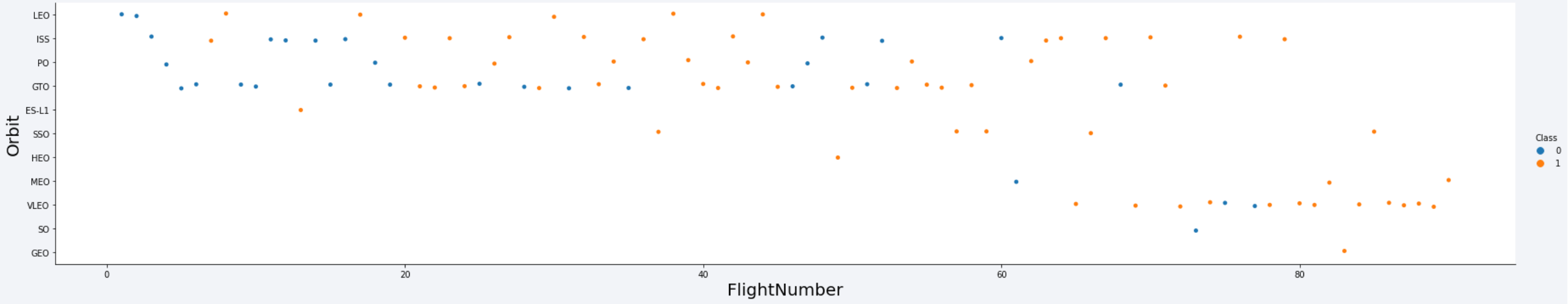
Success Rate vs. Orbit Type

- The orbits with the biggest success rates are ES-L1, GEO, HEO, and SSO



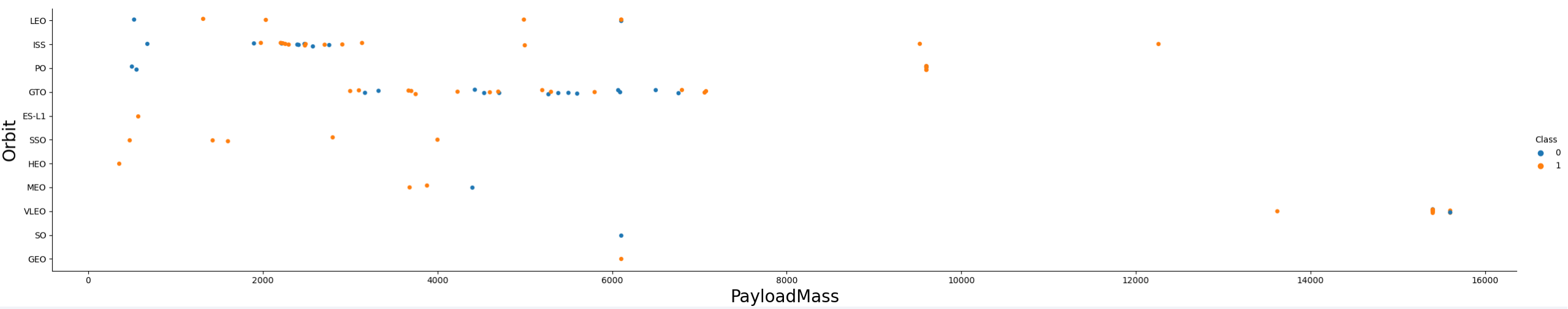
Flight Number vs. Orbit Type

- Again, we see clear overall improvement over time
- It seems that more recent launches favor orbits that are on the bottom half of the graph
- The orbits with the highest success rate also have a low number of launches



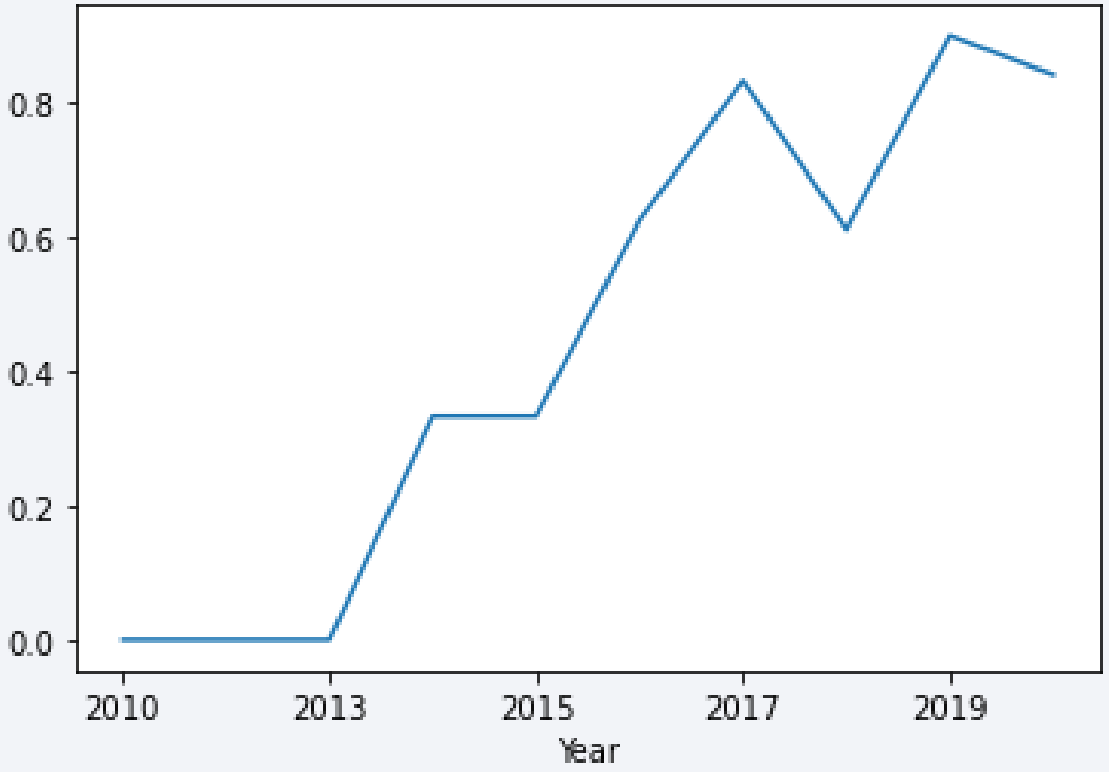
Payload vs. Orbit Type

- Payload mass seems to have a big impact on what orbit is taken because each orbit seems to have certain weights that it does often
- ISS seems to be the only exception to this



Launch Success Yearly Trend

- There is an upward trend in success since 2010
- There was a slight decrease from 2017 to 2018 but then the increase began again



All Launch Site Names

- Used an SQL query to determine these as the four launch site

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin 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	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

Total Payload Mass

- Total payload calculated by summing up all the payloads whose codes contain CRS

total_payload
111268

Average Payload Mass by F9 v1.1

- Total payload calculated by finding the average of the payloads whose codes contain CRS

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Used the minimum date to find the first occurrence

FIRST_SUCCESS_GP
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

- Selecting distinct booster versions

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

Total Number of Successful and Failure Mission Outcomes

- Utilize counting and grouping to perform this query

Mission_Outcome	QTY
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- These are the boosters that carried the maximum payload

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

2015 Launch Records

- Only two failed occurrences recorded

month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	01/10/2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14/04/2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

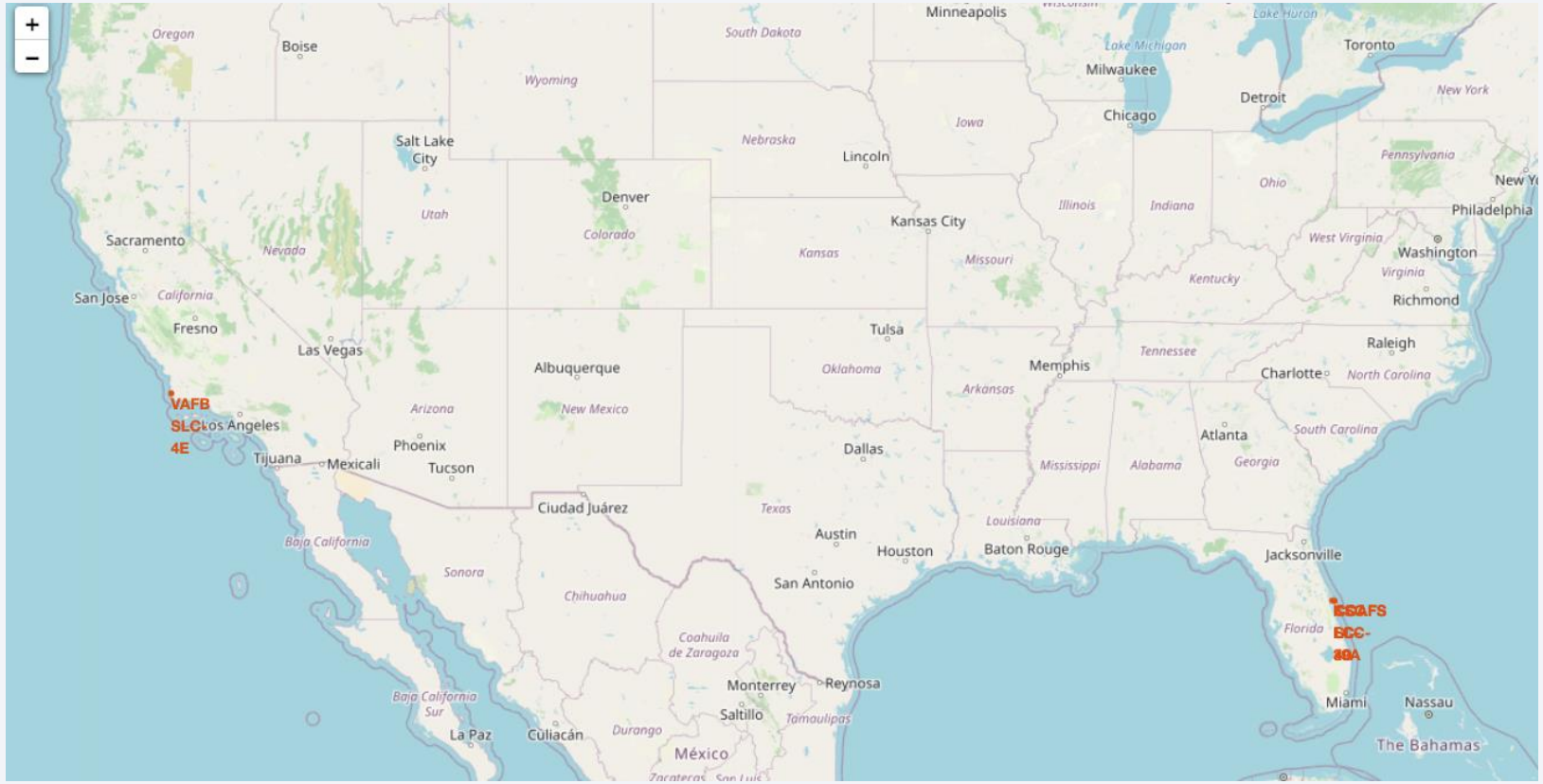
Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

Section 3

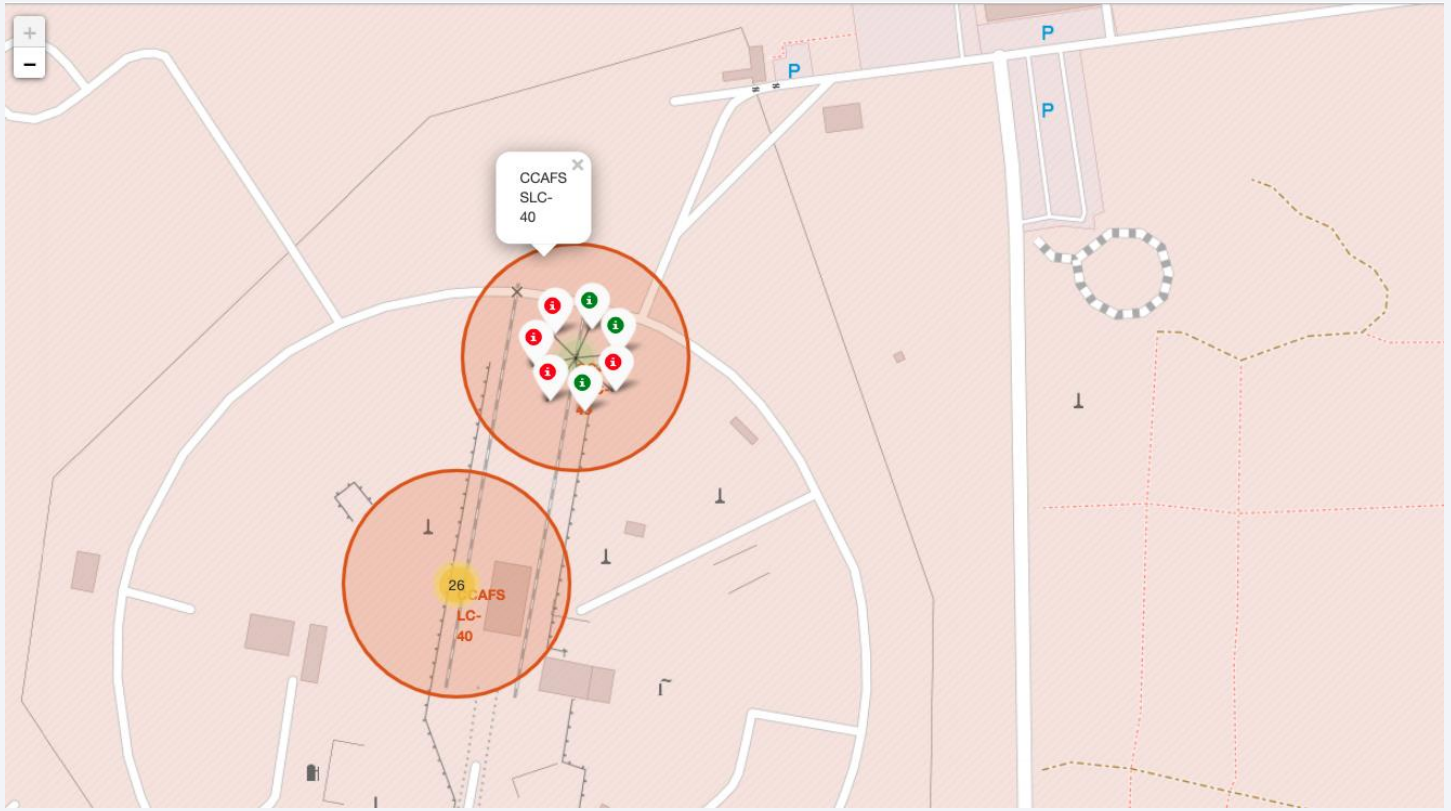
Launch Sites Proximities Analysis

Launch Locations



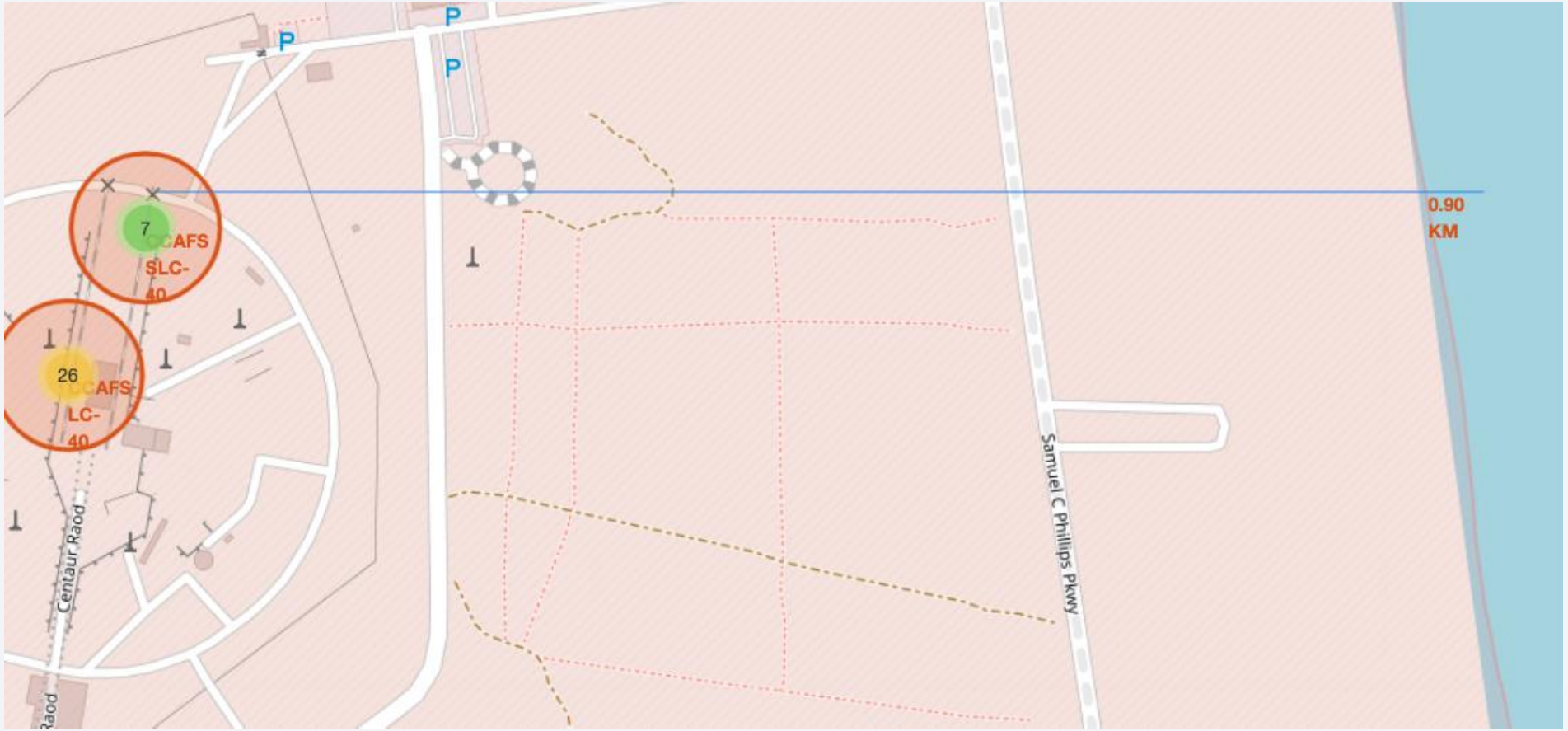
- Launches are on either side of the coast where there is less risk of unintentional injury to civilians

Success Rate

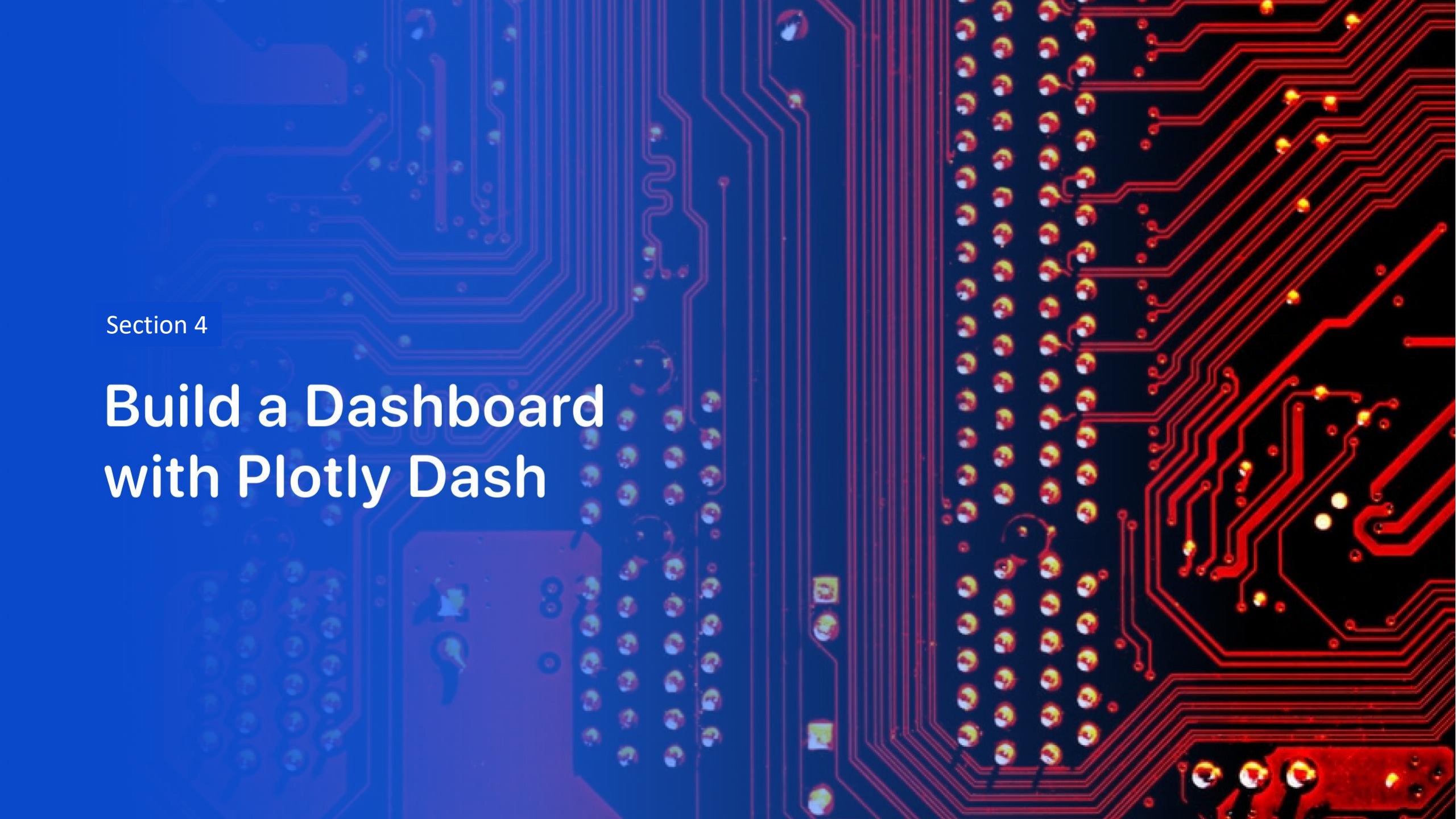


- Green indicates success

A Closer Look

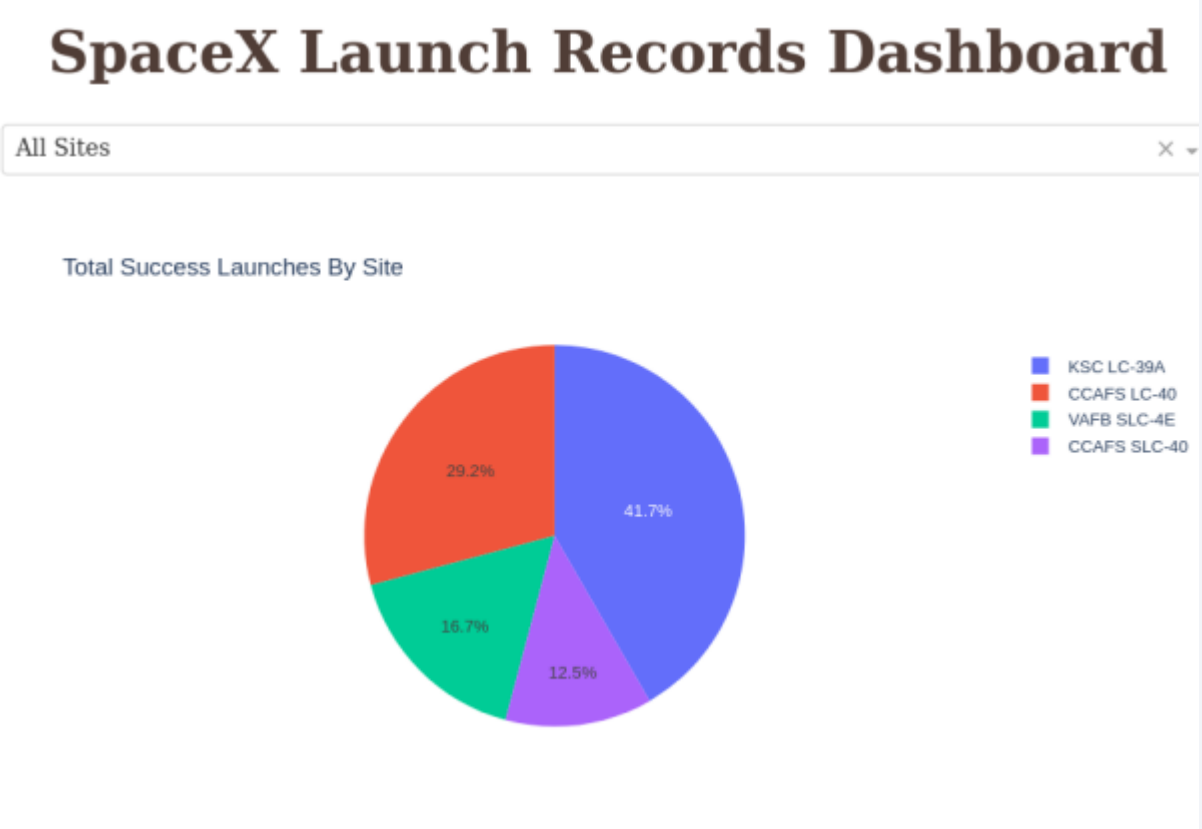


- Shows the distance to the water and other things located around it


Section 4
Build a Dashboard with Plotly Dash

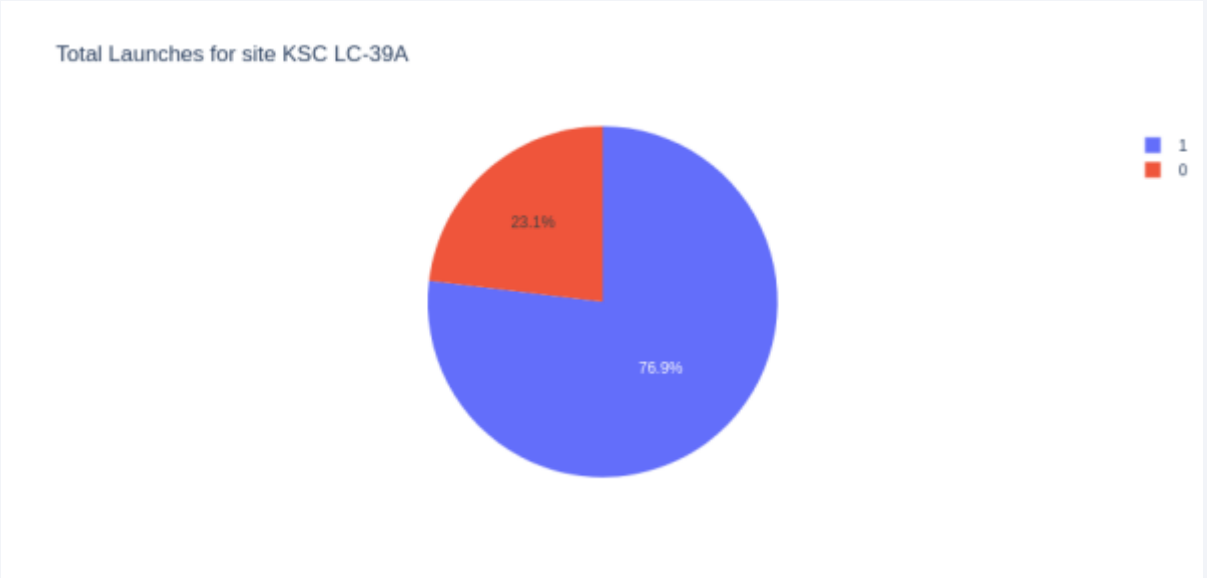
Successes by Location

- While there needs to be more context to draw definitive conclusions, it can be seen that the location has a major impact on the success of the launch



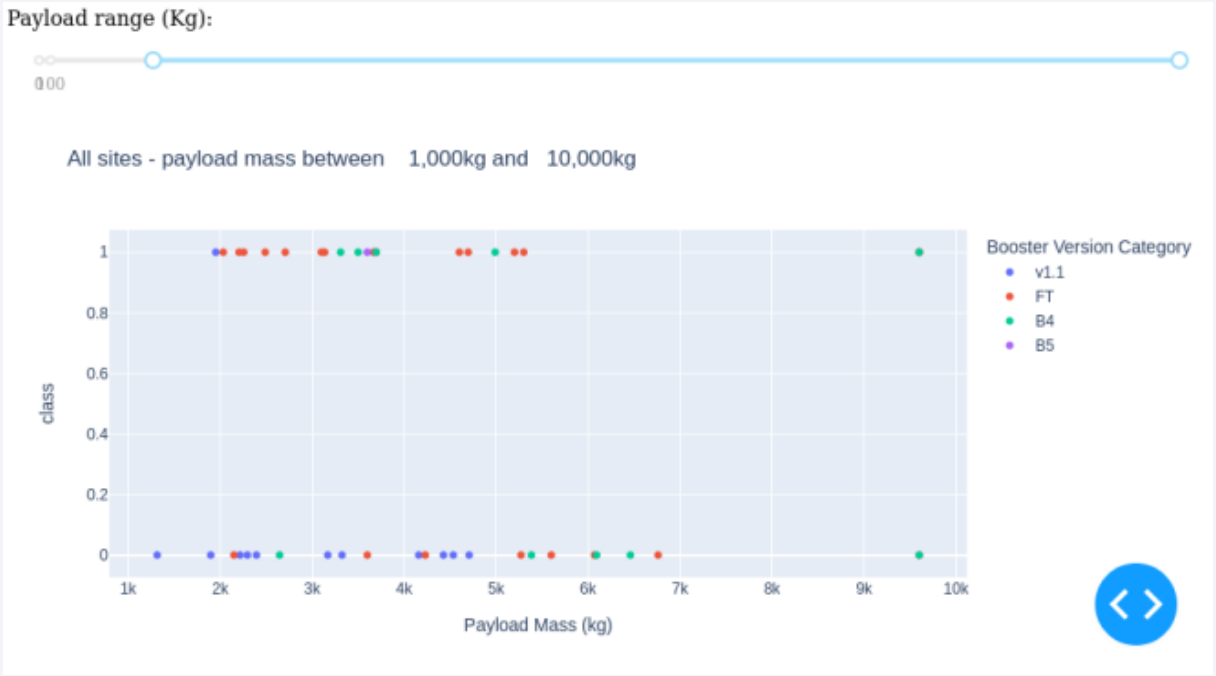
KSC LC-39A Success Ratio

- 76.9% success rate



Payload vs Launch Outcome

- Not a lot of data for 7000 and up

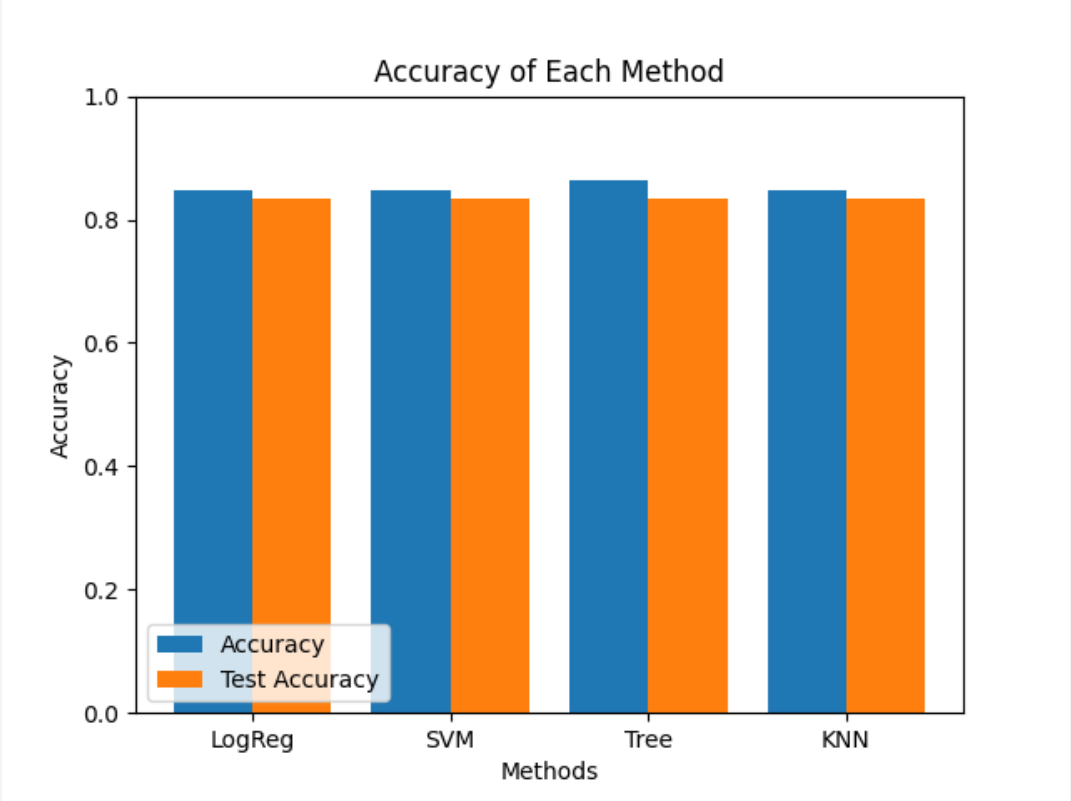


Section 5

Predictive Analysis (Classification)

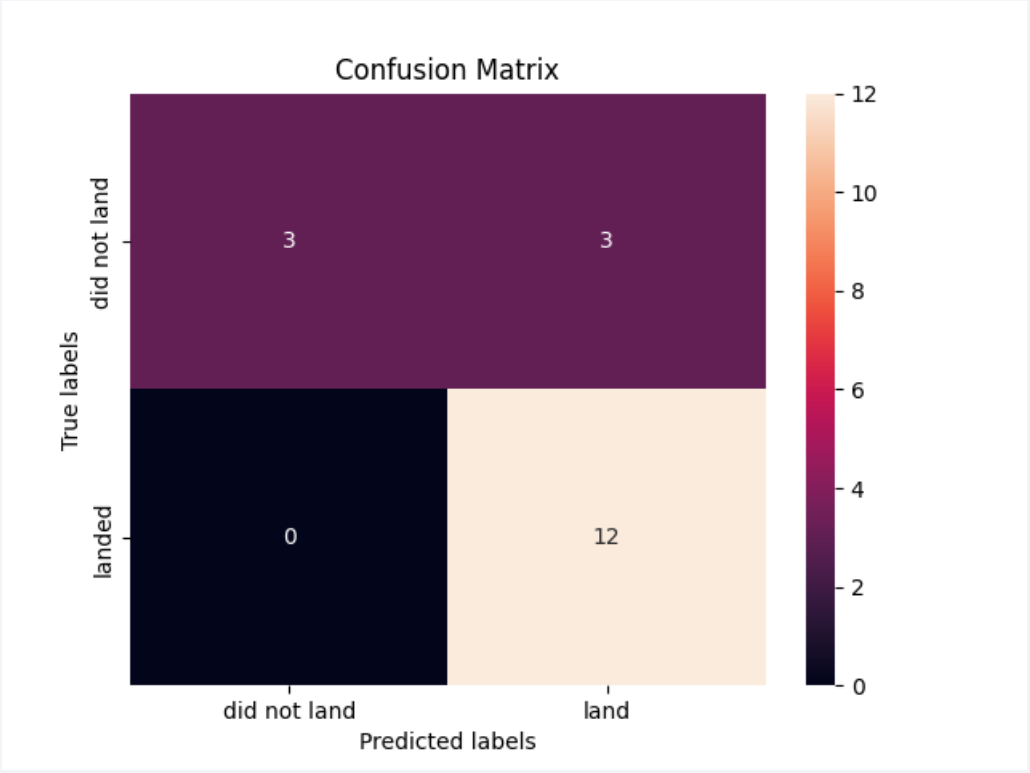
Classification Accuracy

- The tree method has the highest accuracy
- The test accuracies are very similar which means that they all likely perform similarly



Confusion Matrix

- The confusion matrix shows the accuracy of predicting positive and negative results
- This matrix predicts the positive (lands) very well but the negatives (did not land) rather poorly

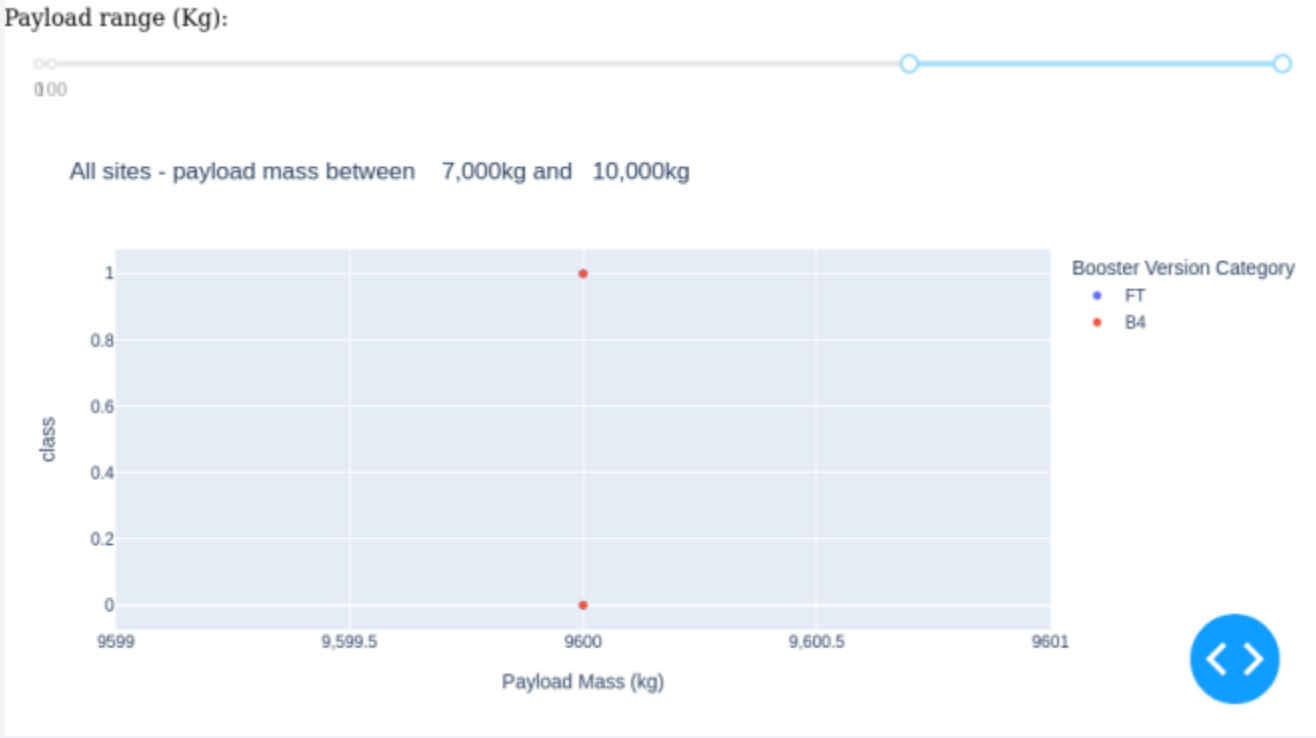


Conclusions

- The most obvious point is how launches are improving over time as technology is improved and experience is gained
- The use of the decision tree classifier will help increase profits because it allows us to make predictions on future launches based on the data of past launches. We will be able to continue doing so as more launches occur in the future
- Larger and heavier launches seem to be successful more often
- Every launch is different and it seems that aspects of the launch such as orbit, location, etc. are chosen based on what kind of launch is occurring. Would need more data to confirm this with more specificity

Appendix

- Here is one



Thank you!