IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

John Elliot Stowers Jr
11/28/2025

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion
6. Appendix

Executive Summary

<u>Methodologies included:</u>

- Collection
- Wrangling
- Exploring
- Analyzing
- Geomapping
- Visualization
- Model Construction

OBJECTIVE: OUR OBJECTIVE WAS TO DETERMINE THE OPTIMAL TRAITS IN A LAUNCH

PAYLOAD SIZE(IN KG)	LOCATION	ORBIT	RECOVERY SYSTEM

OUTCOMES: OUR RESULTS SHOWED THAT:

2K-4K KG	STIE KSC LC-39A	ES-L1, GEO, HEO, SSO	DRONE SHIP (ASDS)

Introduction

- Our project began with the objective of insuring the success of our company, “SpaceY,” by building upon the experiences of the existing company “SpaceX”.

- To do this, we sought to determine the optimal conditions for launching reusable rockets to maximize recoverability and succuss rates.

This presentation aims to explain our findings and provide an understanding of the principles used to determine at them

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected via Scrapping and SpaceX API
- Perform data wrangling
  - Data standardized according to industry standards (more on this latter)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built predictive models based using a Decision Tree Classification model.

Data Collection

The data used for our investigation was collected using 2 primary sources:

SpaceX API

Web scrapping Wikipedia

SpaceX API is a free API that provides excellent data freely available to the public.

Wikipedia, while not ideal as a source, certainly provides a free and convenient information repository.

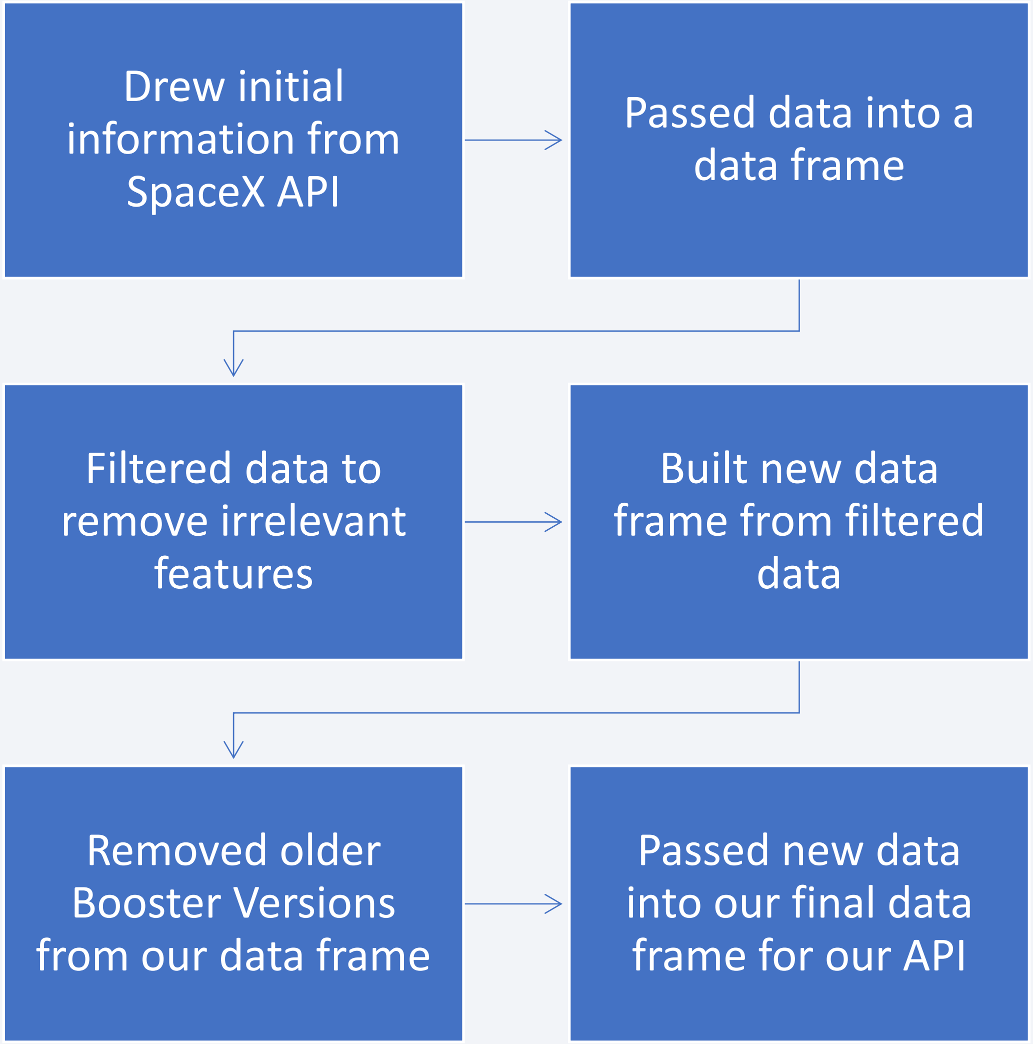
Data Collection – SpaceX API

Data from SpaceX API was gathered using a variety of libraries:

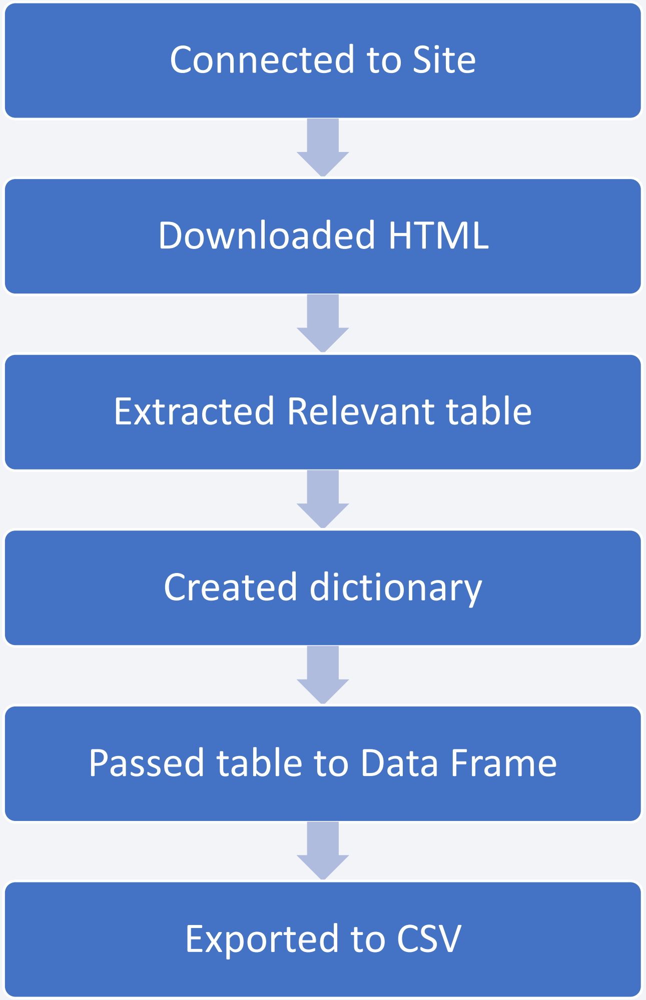
- Requests
- Pandas
- Numpy
- Datetime

Code used to preform these functions can be viewed at the following location:

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%201%20-%20jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping



Webscraping was done using BeautifulSoup to draw information from tables stored on Wikipedia.

Code used to preform these functions can be viewed at the following location:

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%202%20-%20jupyter-labs-webscraping.ipynb

Data Wrangling

- While wrangling the data it became obvious that some additional changes to our data were necessary, namely:
  - Null values needed resolving
  - Values for launches from each site were needed
  - Orbit types needed binning
  - Landing types needed classification
  - ‘Success’ and ‘Failure’ needed binary classification

Code used to preform these functions can be viewed at the following location

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%203%20-%20labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- In this section, we conducted basic exploration of the data visually. We used Matplotlib and Seaborne to help visualize our information. The following chart types were used:

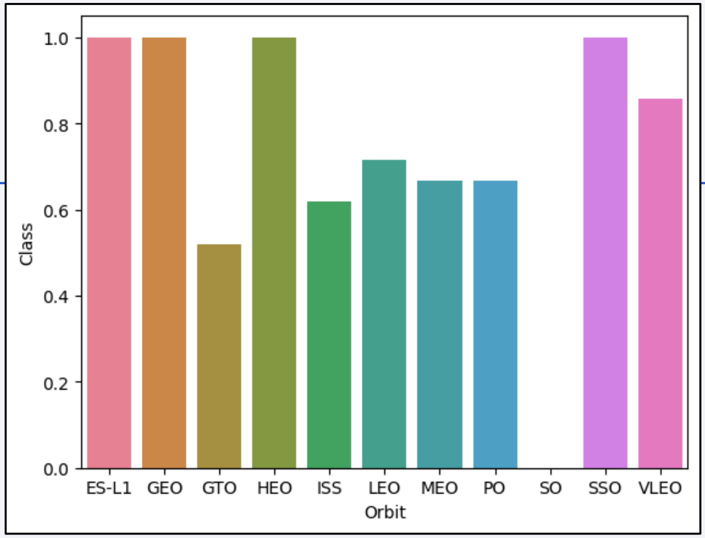
Scatter - Showed Success/Failure of launches based on related features

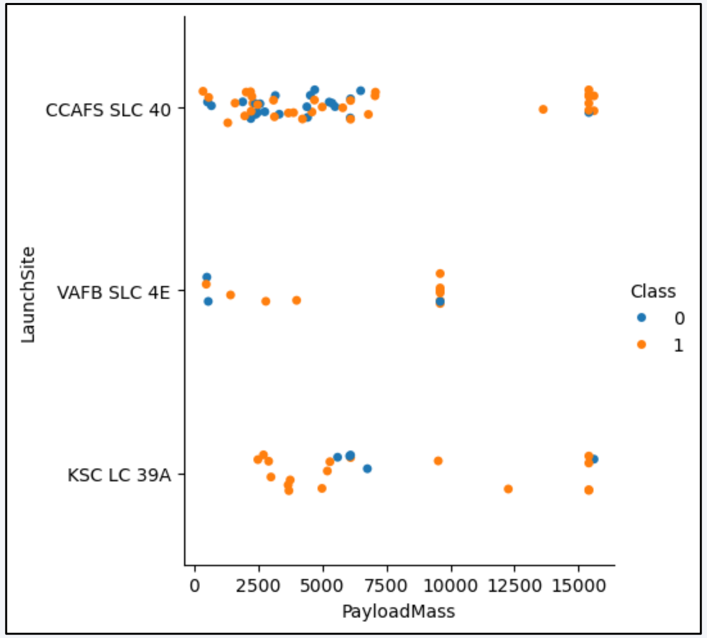
Bar – Showed Success/Failure of launches based on Orbit

Line – Showed Success/Failure of launches based on year

Code used to preform these functions can be viewed at the following location

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%205%20visualisation%20edadataviz.ipynb





EDA with SQL

- Some of the data exploration was conducted using sqlite3. The following is a non-exhaustive list of queries that were performed:

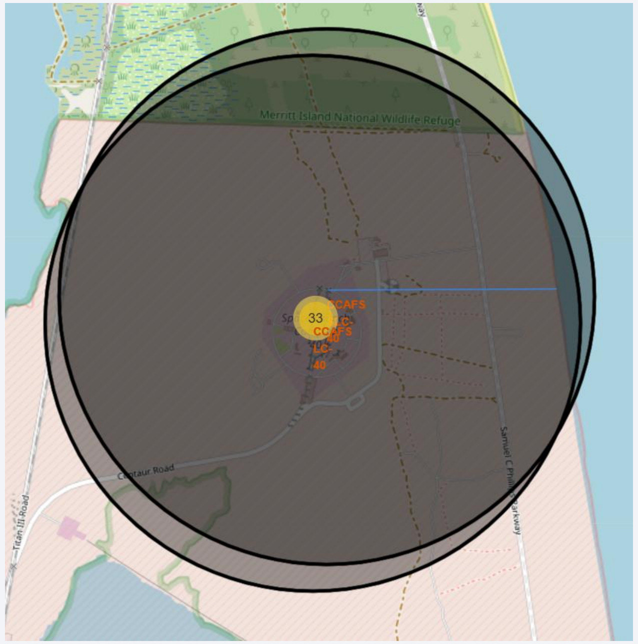
Query	Purpose
SELECT DISTINCT Launch_Site FROM SPACEXTABLE	Find all unique launch sites
SELECT SUM(PAYLOAD_MASS__KG_) as 'TOTAL_PAYLOAD_MASS_KG_' FROM SPACEXTABLE	Collect the total lifetime payload sent on this platform
SELECT AVG(PAYLOAD_MASS__KG_) as 'TOTAL_MODULE_MASS' FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%';	Collect the total lifetime payload sent on the Falcon9 V1.1 booster specifically
SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome == 'Success'	Find the earliest success of this system
SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)	Find the successful launches with the largest payloads
SELECT Landing_Outcome, COUNT(*) AS 'Total' FROM SPACEXTABLE GROUP BY Landing_Outcome ORDER BY 'Total' DESC;	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

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%205%20visualisation%20edadataviz.ipynb

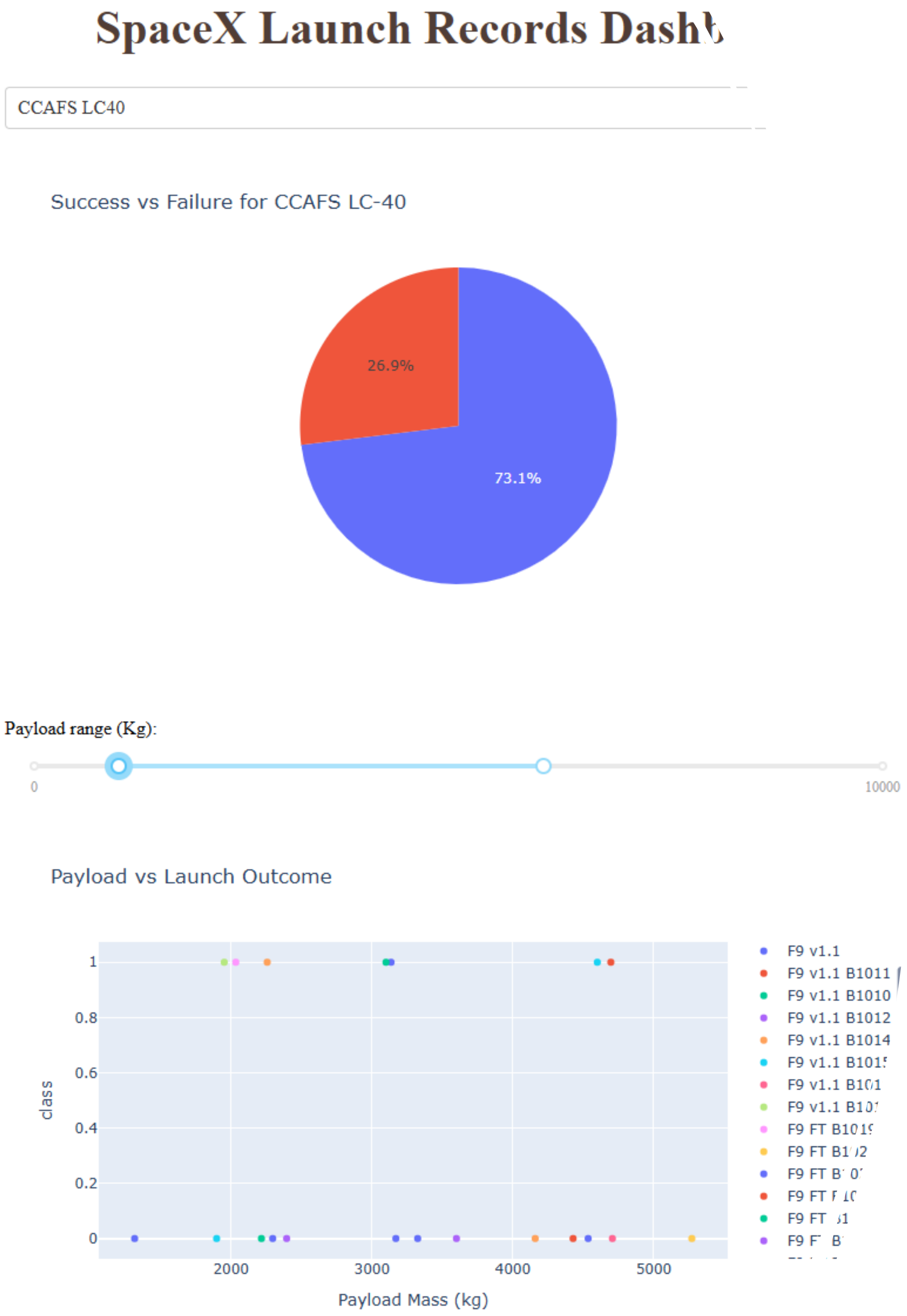
Build an Interactive Map with Folium

Folium was used to gather Geographical Data and apply markers to the map, such as:

- Lines – showing distances to relevant geographical features(i.e. Highways, coasts, railways, etc.),
- Pop-ups – showing information regarding the launch itself
- Circles – showing clusters of markers
- Markers – showing individual launch sites



https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/pt%206%20folium%20lab_jupyter_launch_site_location.ipynb



Build a Dashboard with Plotly Dash

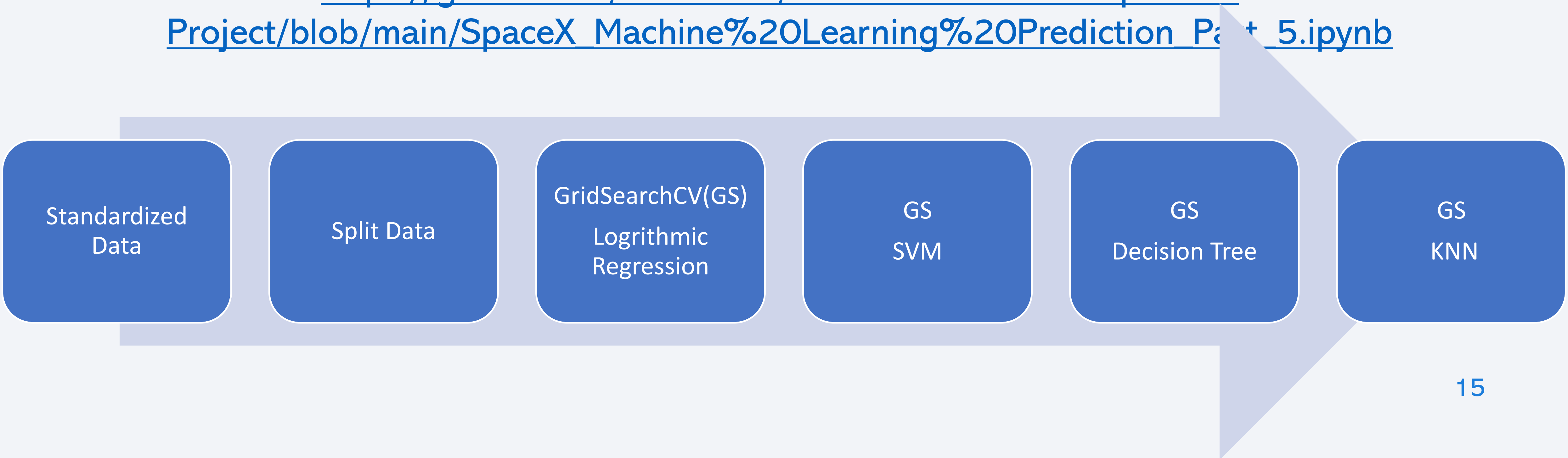
- An interactive Dashboard was built using Plotly. In this dashboard you can see scatter charts showing success and failure instances for any given booster, as well as a pie chart showing success and failure rates as percentages for any given booster
- These interactions help visually identify which boosters have the highest success rates.

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/spacex-dash-app-finished.py

Predictive Analysis (Classification)

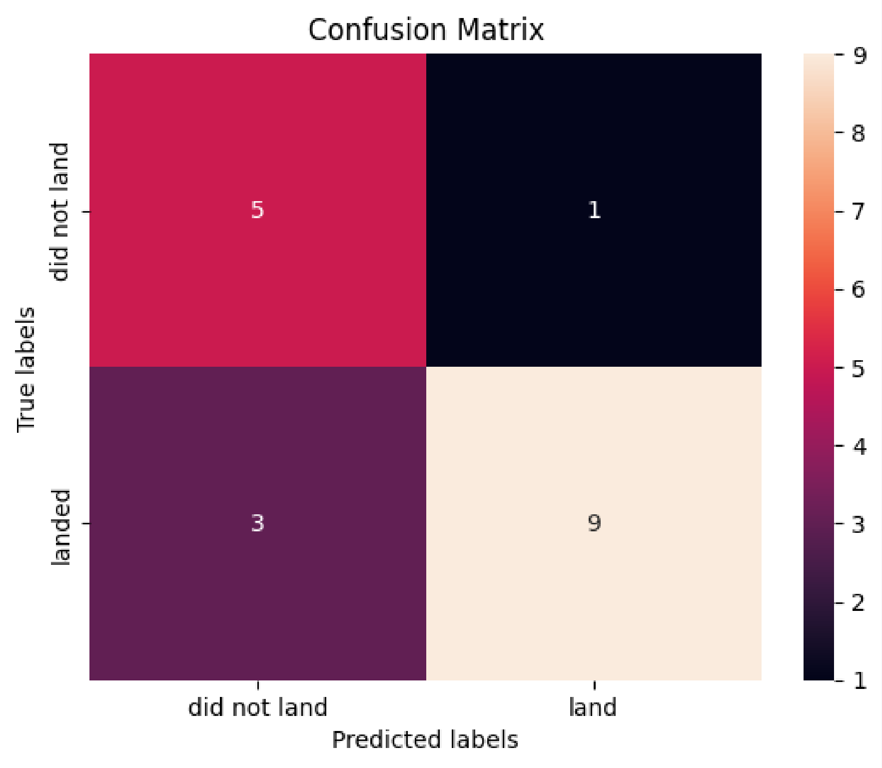
- Data was standardized, transformed, and fit to an X and Y. It was then split into testing and training sets at a ratio 1/5 and given a random state of 2. A GridSearchCV was then created and iteratively fit with different classifiers to determine the strongest candidate for modeling.

https://github.com/Bootscum/IBM-Data-Science-Capstone-Project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

- Exploratory data analysis results
  - 100% of ES-L1(Lagrange Point), GEO(Circular Geosynchronous), HEO(Highly Elliptical), and SSO(Sun-synchronous) orbits were successful.
  - The highest number of successful recoveries were on drone ships
- From our interactive analysis, we found:
  - That the ideal location was Launch Stie KSC LC-39A near Melbourne, Florida, with ~77% of launches being successful.
  - That the ideal payload is between 2000kg and 4000kg.
- Predictive analysis results
  - Using our model, we can predict with 89% accuracy the outcome of our launch.



Section 2

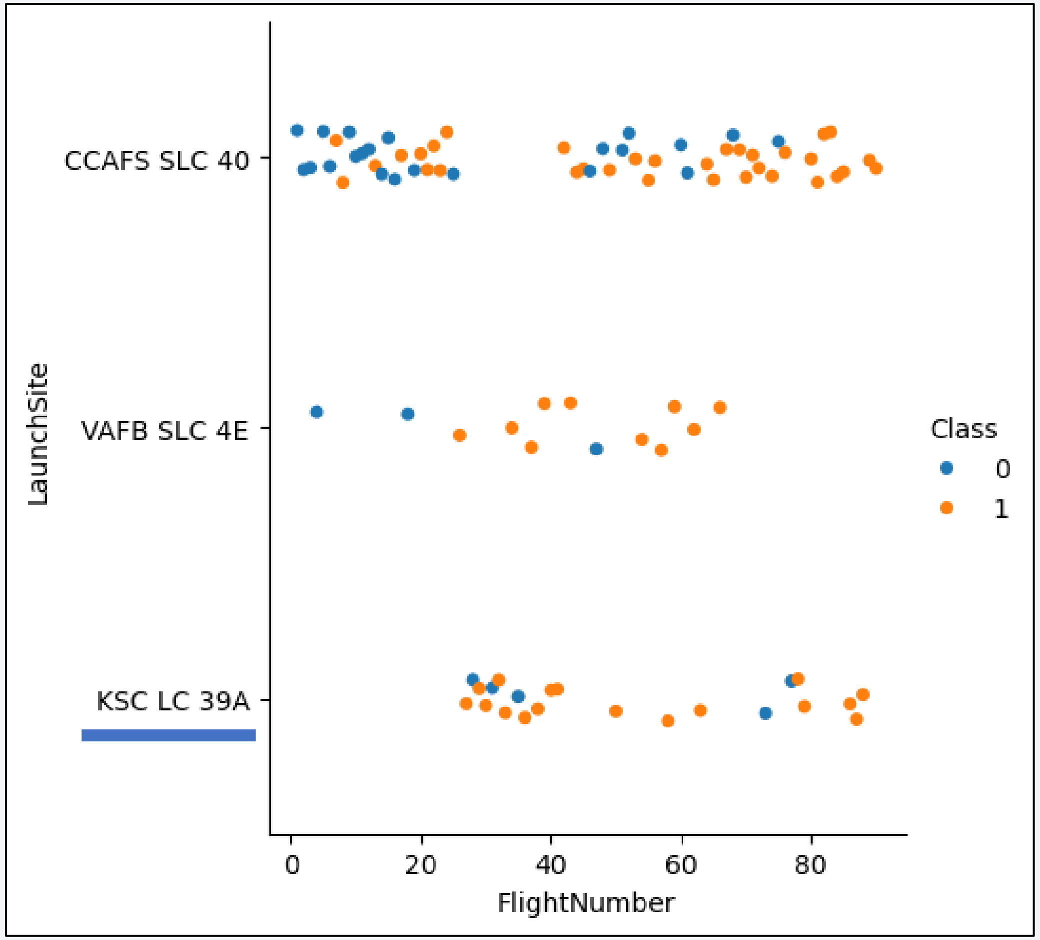
Insights drawn from EDA

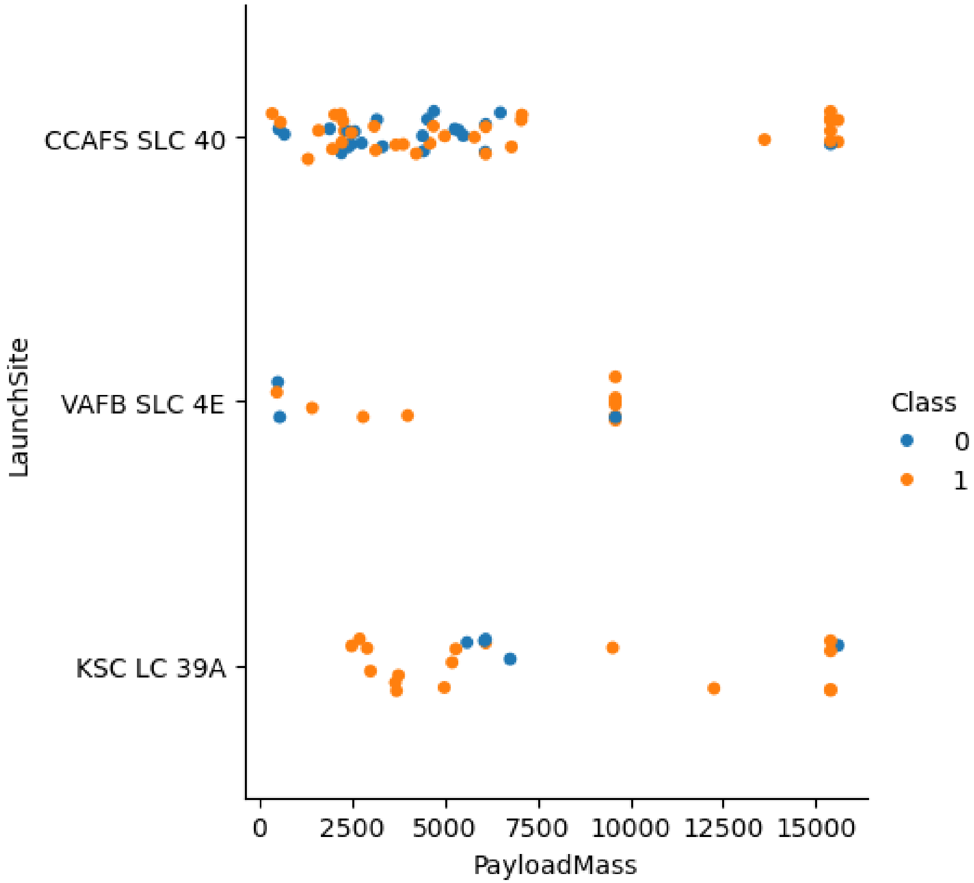
Flight Number vs. Launch Site

A Class of 0 indicates a failure, while a Class of 1 indicates a success.

We can see here that, generally speaking, as more launches are conducted from a given site, their success rate increases.

Additionally, it is apparent that KSC LC-39A has the highest ratio of success





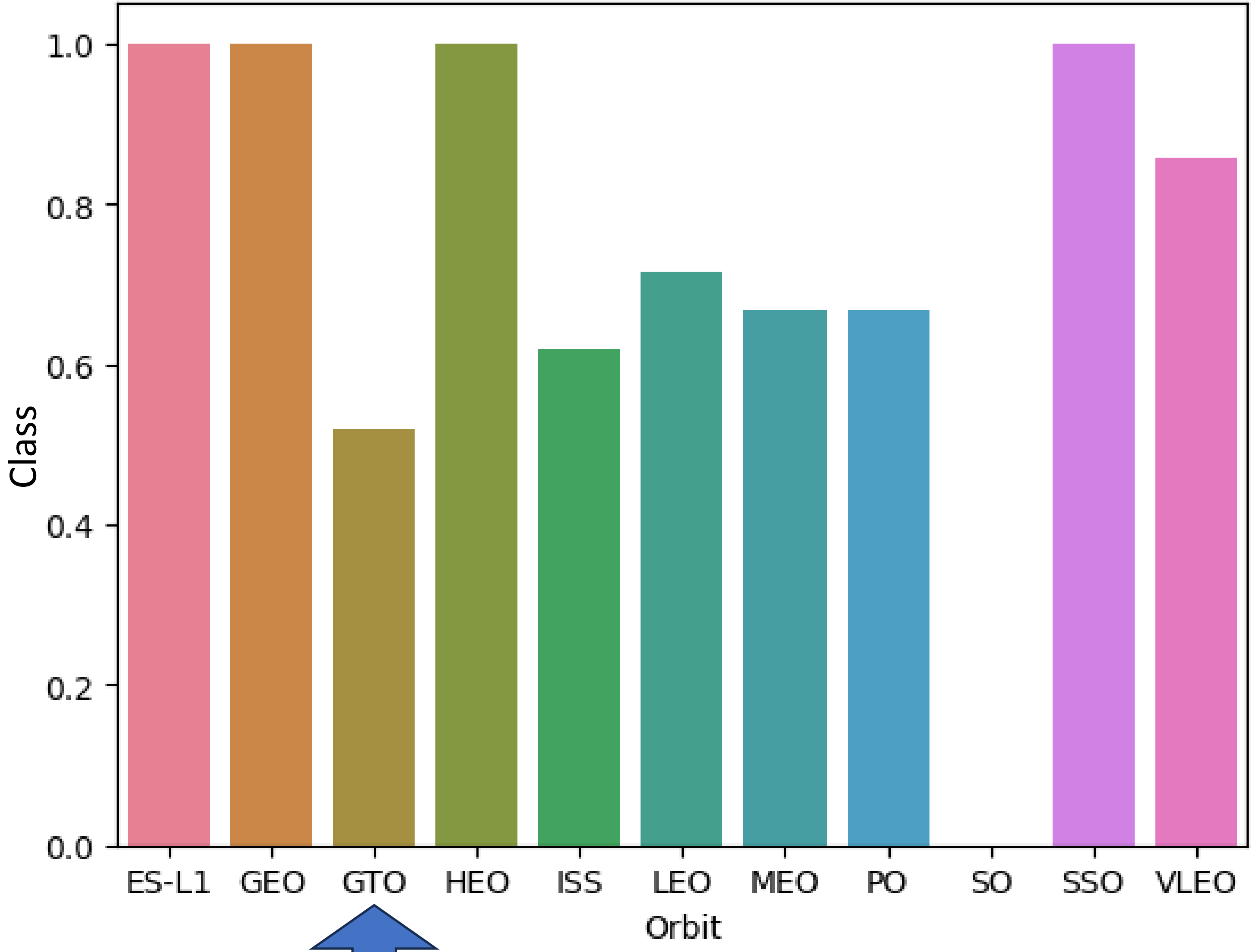
Payload vs. Launch Site

As before, a Class of 0 indicates a failure, while a Class of 1 indicates a success.

While most site can launch a variety of payload weights, KSC LC-39A has a 100% success rate with payloads 2000-4000 kg. While the same is true of VAFB SLC-4E, the sample size is much smaller, and therefore less indicative.

CCAFS SLC-40, by comparison, has a much more mixed rate.

Success Rate vs. Orbit Type



Note: GTO is a transitory orbit and therefore not a "True orbit"

Orbit	Count
ES-L1	1
GEO	1
GTO	27
HEO	1
ISS	21
LEO	7
MEO	3
PO	9
SO	1
SS)	5
VLEO	14

Here we can see that orbit and class have a strong correlation. We can see that ES-L1, GEO, HEO, and SSO have a 100% success rate.

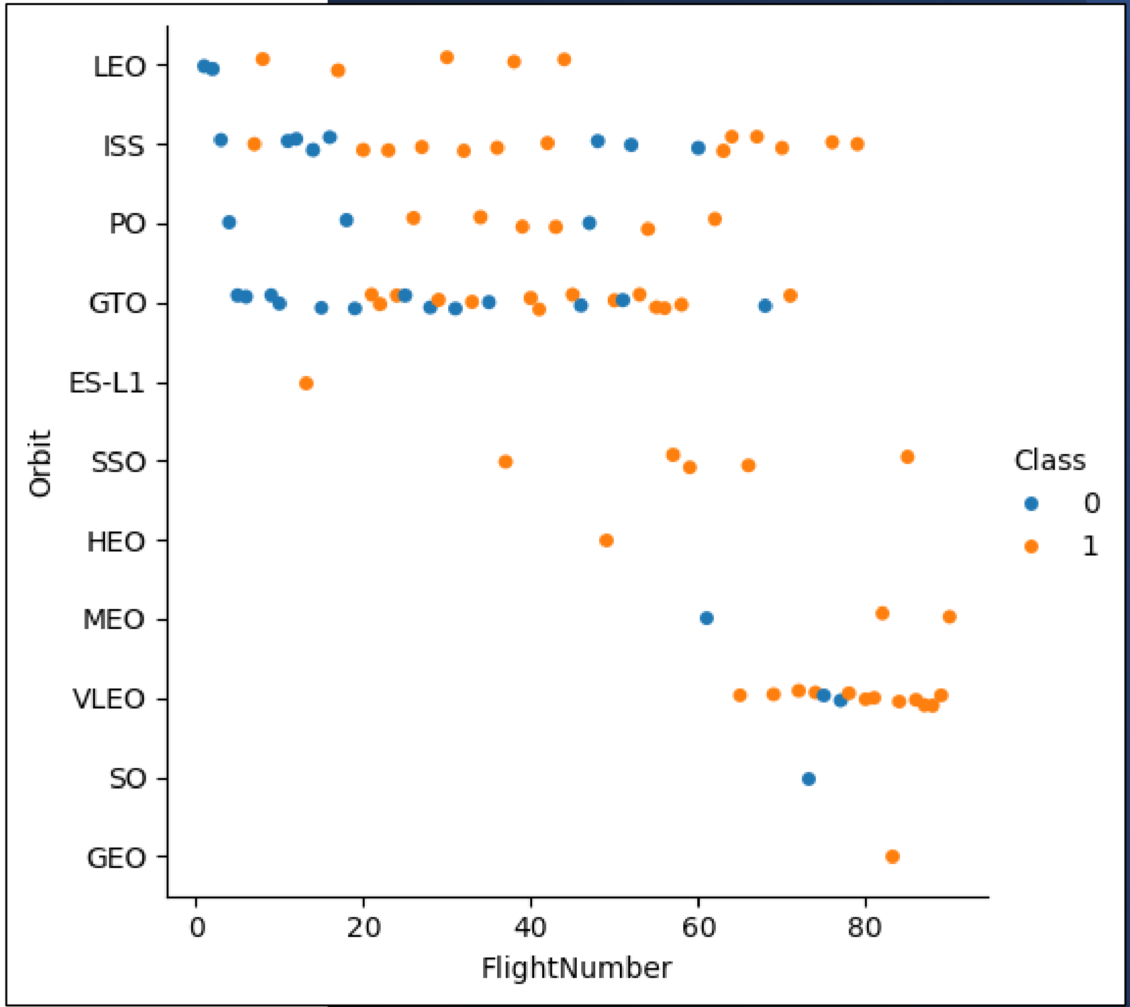
Again, a Class of 0 indicates a failure, while a Class of 1 indicates a success.

Flight Number vs. Orbit Type

The previous data could prove to be somewhat deceptive unless without some idea of how many launches have been directed towards each orbit. Here, that information is charted for your convenience. It also allows us to see how orbital launches have shifted over time.

Notice that over time Very Low Earth Orbits (VLEO) have become increasingly more common. It is possible this indicates a shift in demand, but that is beyond the scope if this project.

- You know the drill, a Class of 0 indicates a failure, while a Class of 1 indicates a success.

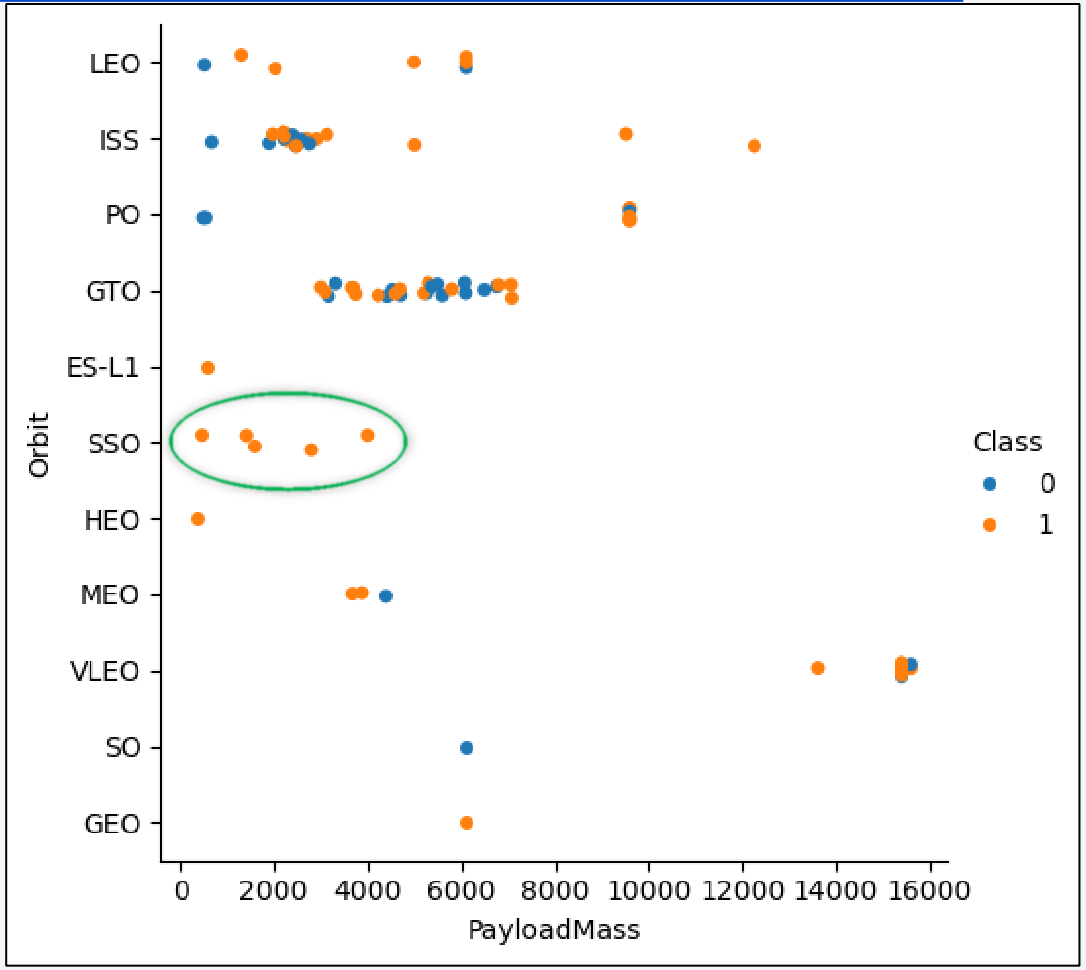


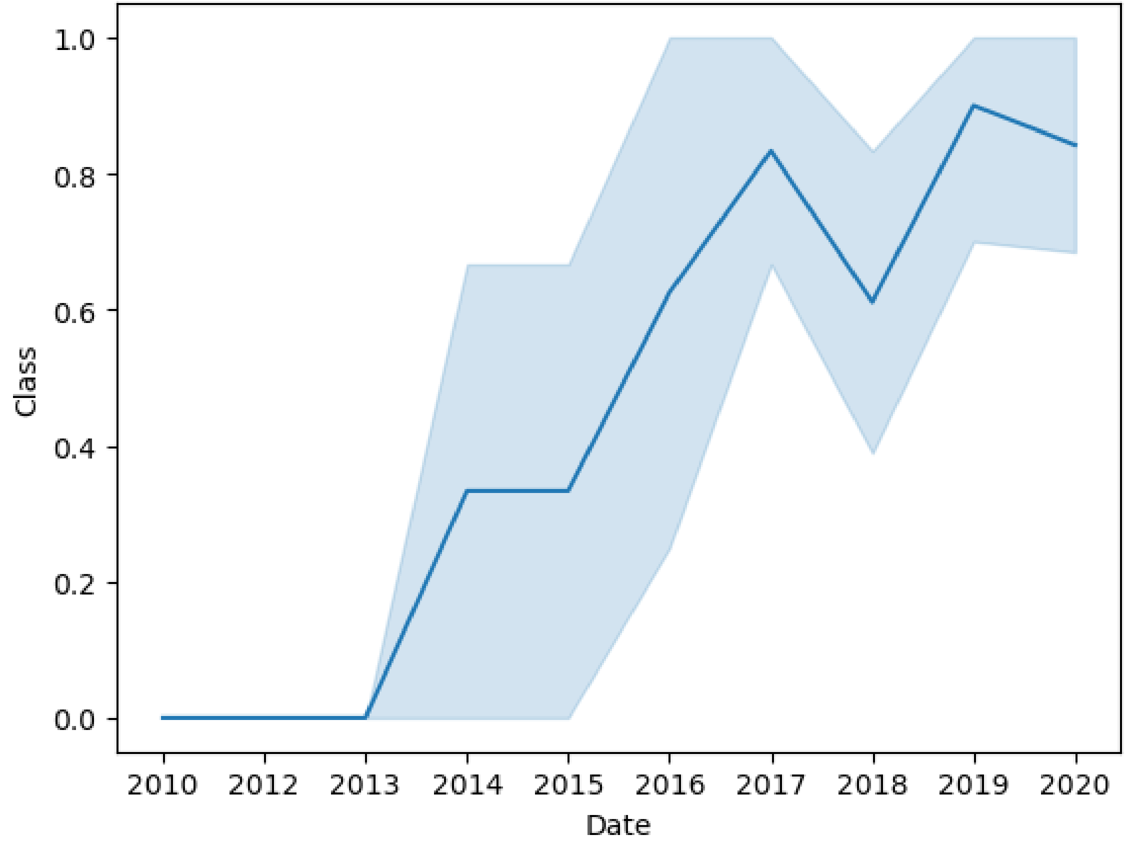
Payload vs. Orbit Type

- Here we can see the orbit types matched to the payload mass. We can see our ideal payload range exhibits no failures across the SSO orbit type.

- This is the last time I'll say it, but….

A Class of 0 indicates a failure, while a Class of 1 indicates a success.





Launch Success Yearly Trend

- With any emerging technology, it is key to consider whether there is a track record of improvements in the field to decide whether it’s possibilities have peaked or if there is room for growth.

- Here we can see that, except for a dip in 2018, the success of this technology has improved year over year.

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

All Launch Site Names

Here is a short list of all distinct Launch Sites

Launch Site Names Begin with 'CCA'

- These are the first 5 records where launch sites begin with `CCA`. This sample shows some of the types of data utilized.

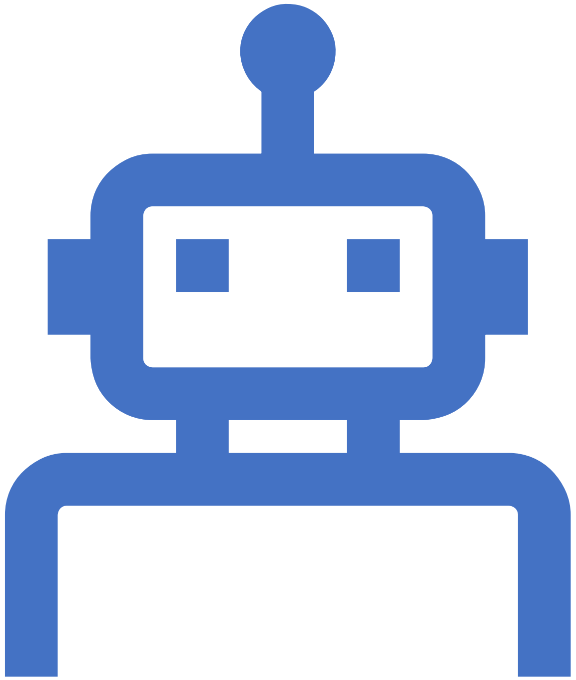
Date	Time (UTC)	Booster Version	Launch Site	Payload	PAYLOAD MASS(KG)	Orbit	Customer	Mission Outcome	Landing Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight 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

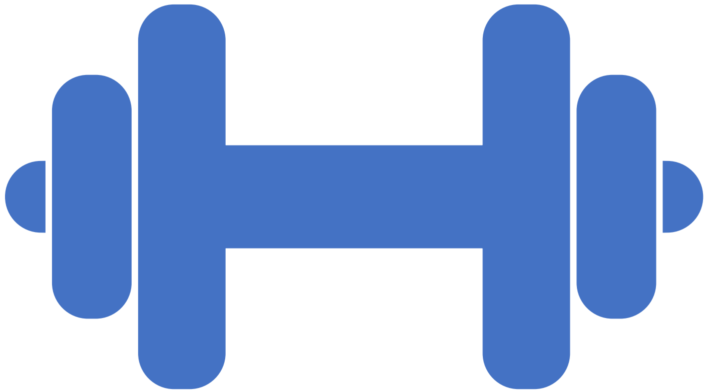
Total Payload Mass

- The Query I used was :

```
SELECT SUM(PAYLOAD_MASS__KG_) as 'TOTAL_PAYLOAD_MASS_KG_
FROM SPACEXTABLE;
```

- The result was a total of 619967 KG





Average Payload Mass by F9 v1.1

Our chosen booster was F9 v1.1 . A helpful piece of information to know would be what the average mass it can carry is. To that end, using SQLite I ran the following query:

SELECT AVG(PAYLOAD_MASS__KG_) as 'TOTAL_MODULE_MASS' FROM SPACEXTABLE

WHERE Booster_Version LIKE 'F9 v1.1%’;

The result was: 2534.66kg

This provided us with a helpful approximation of what weights to consider.

First Successful Ground Landing Date

- Wanting to know more about the history of the technology, our team examined the initial success date and found the first successful landing was 22 December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

There are 5 boosters that have payloads between 4000 and 6000

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

Below is the count of how many mission For the F9 Booster rocket were successes or failures.

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

Important to note here is that these are MISSION outcomes and not LANDING outcomes. These do not indicate safe retrieval.

Boosters Carried Maximum Payload

This is a list of the specific Booster models that have carried the maximum Payload

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

2015 seems like it was generally a poor year for SpaceX with 2/7(28.6%) failed launches

Month	Booster_Version	Launch_Site	Date
01	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
04	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

- Between 2010 and 2017, we can see that about 1-in-3 launches made no attempt at recovery. Furthermore, it appears that Parachutes have the highest failure rate as a percentage.

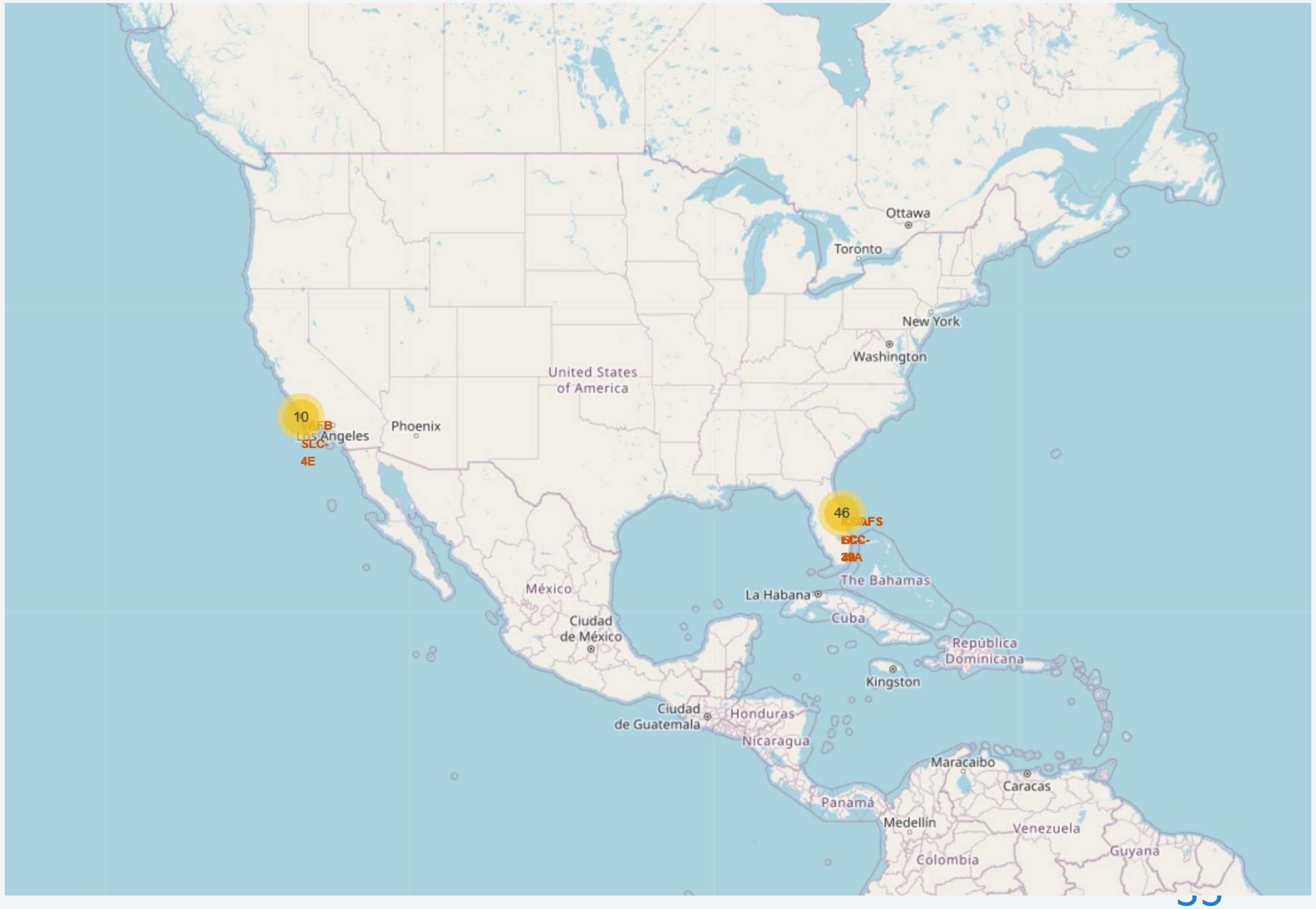
Landing_Outcome	Total
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

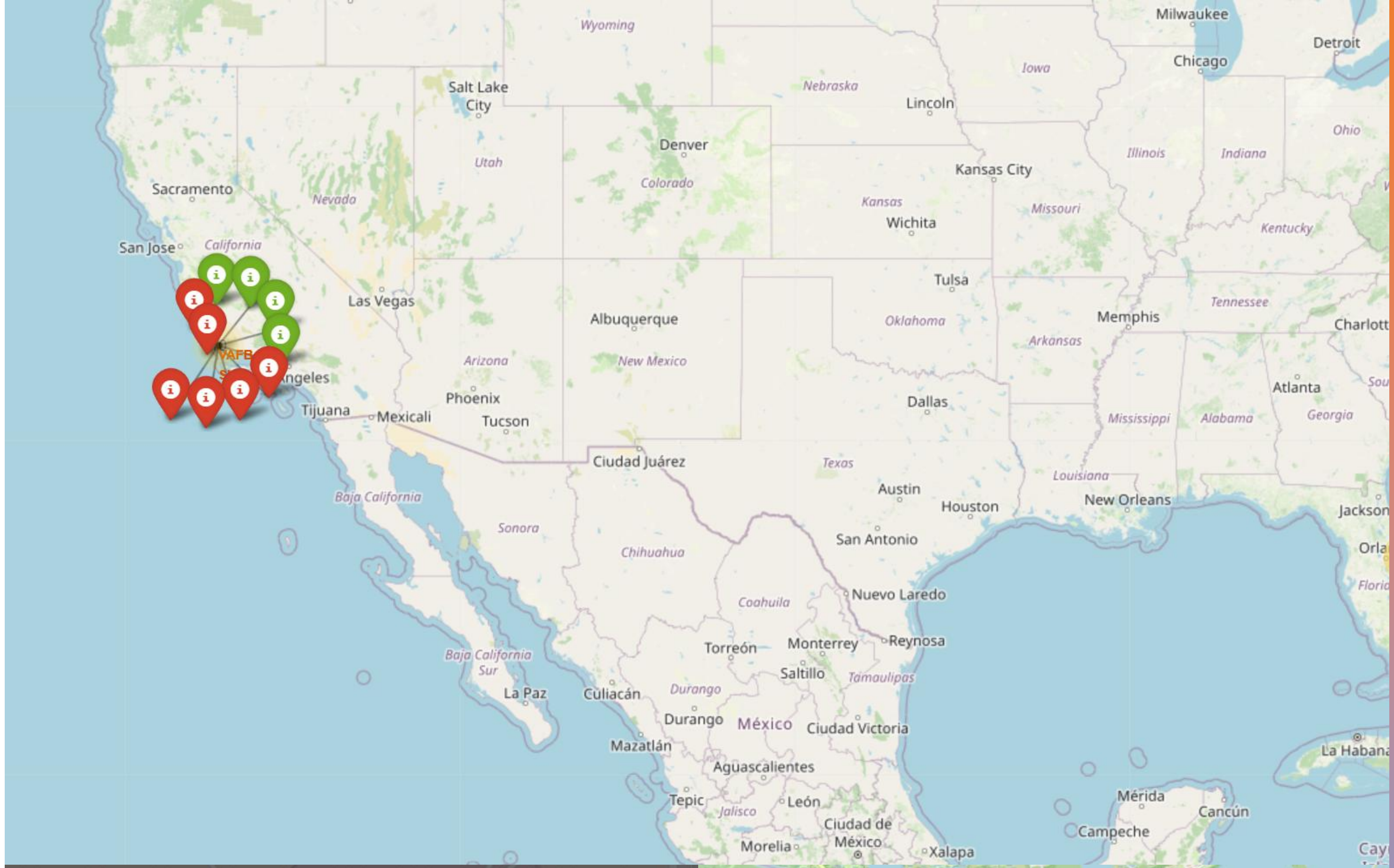
Available Launch Sites

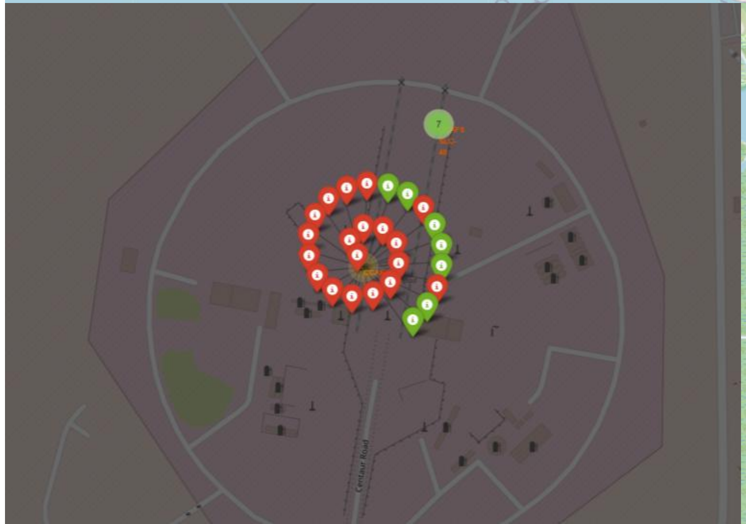
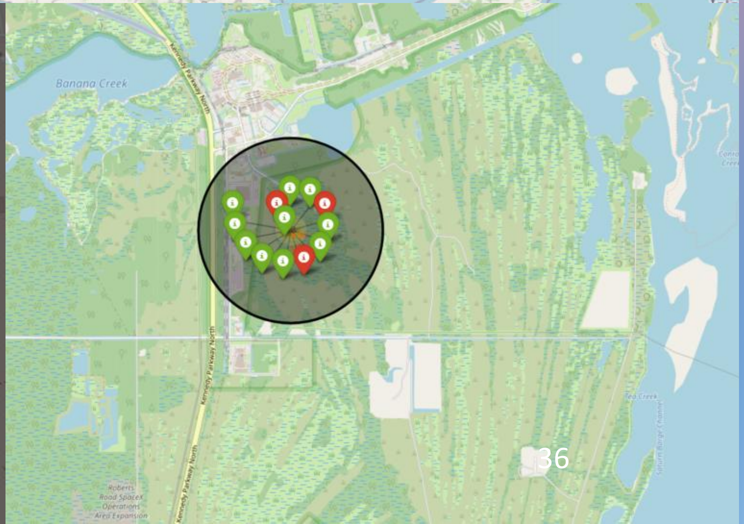
While other launch sites certainly exist, data is available for only a few. I have decided against a global map as most of it would be pointlessly blank.



Pop-up Folium Map

- Here we can see the outcomes from a variety of sites.
- Each of these pop-ups contains helpful information on each launch



Distance to Coast

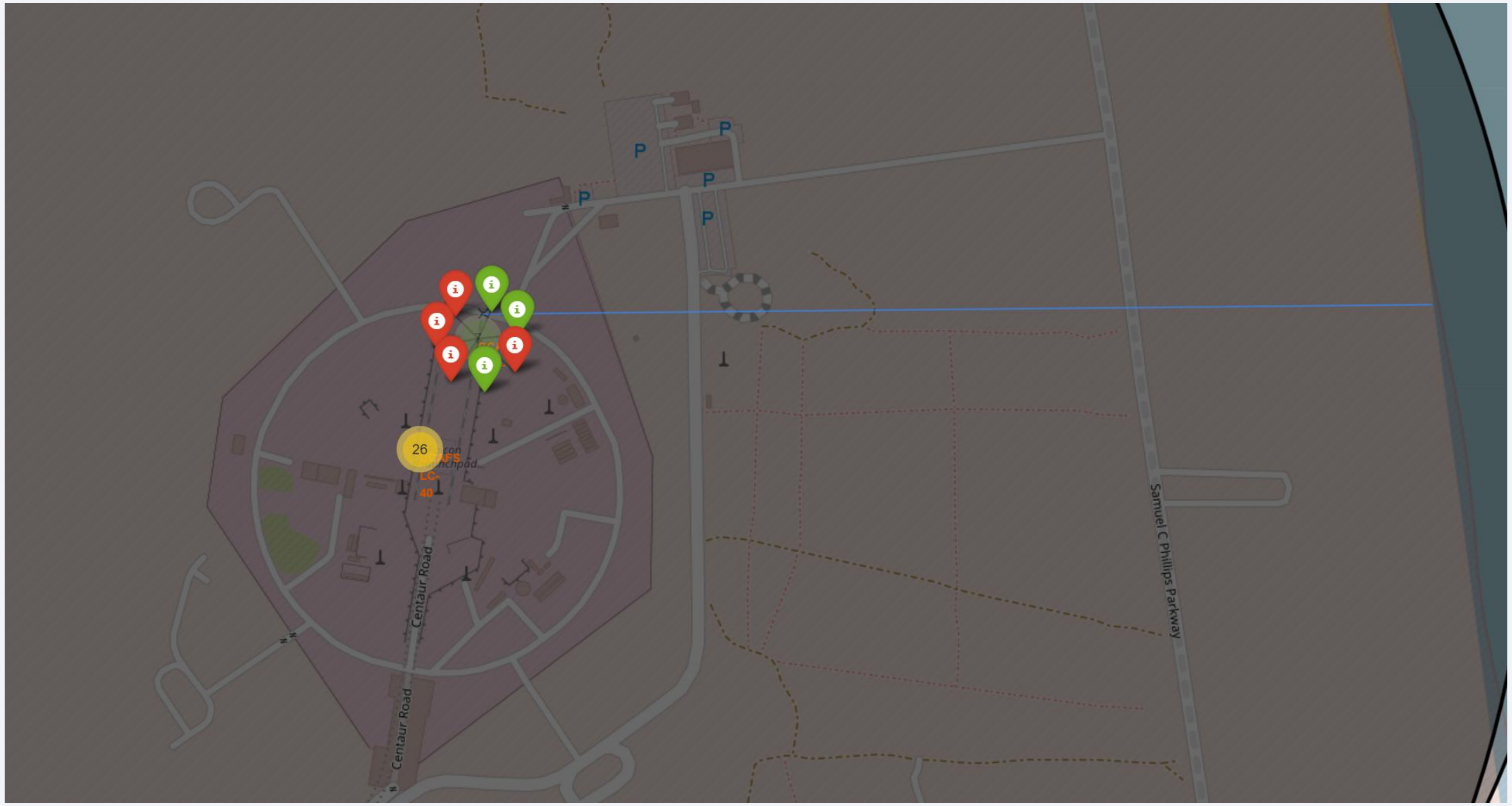
This is an example of how we tracked distances between key geographic features.

Each site has its distance

tracked to better understand

better how other features

might impact outcomes.

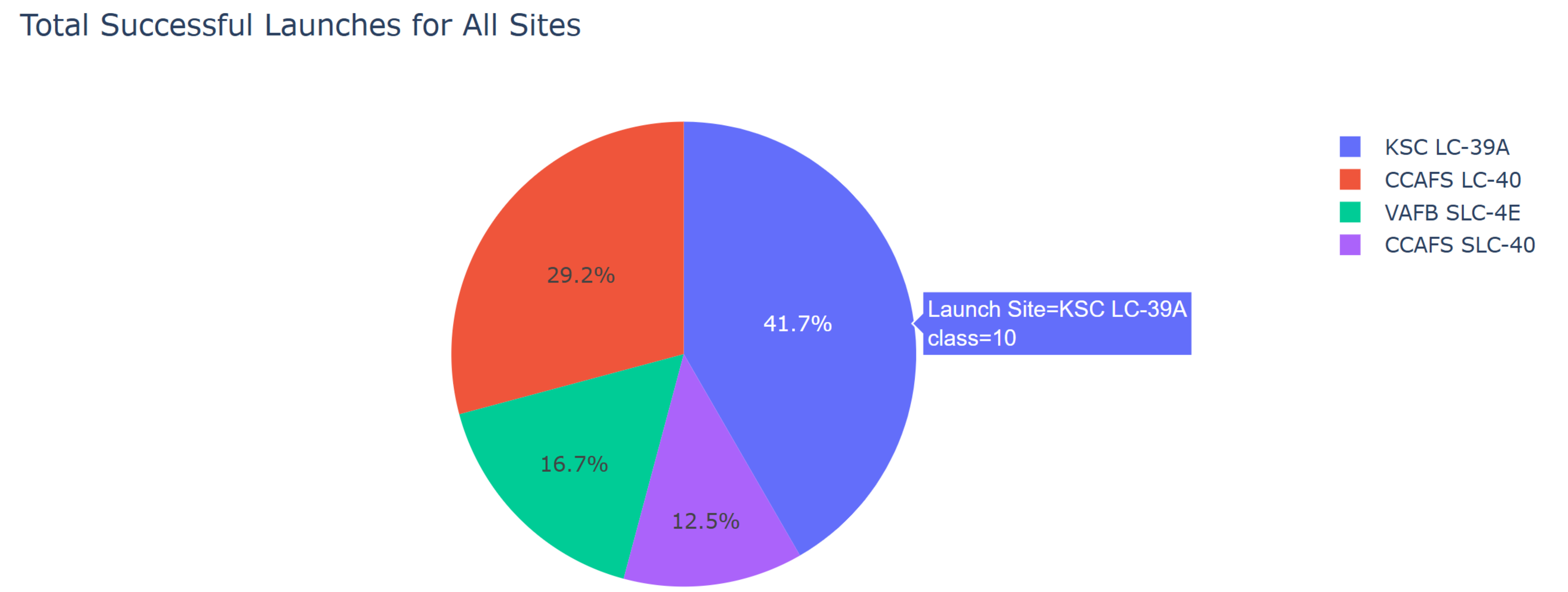


Section 4

Build a Dashboard with Plotly Dash

Total Successful Launches for All Sites

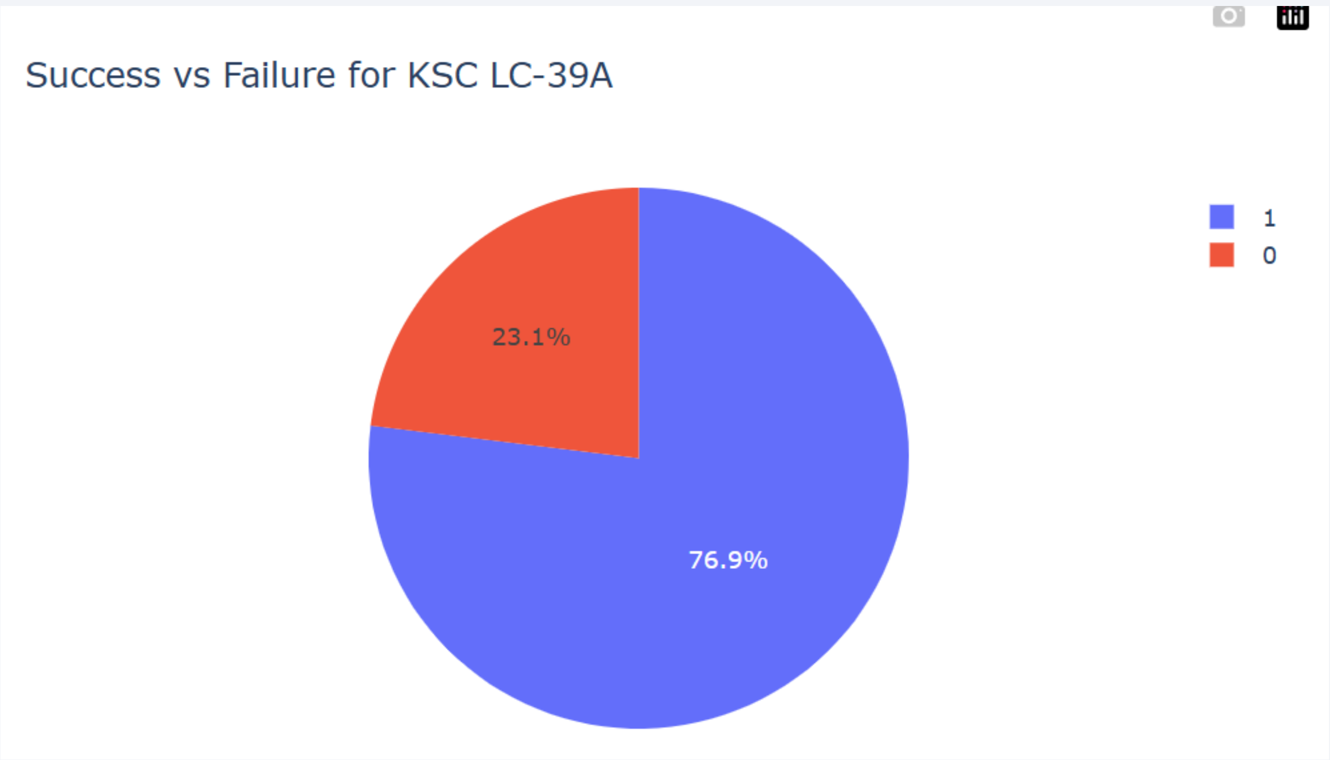
This chart shows where successful launches are coming from. From here, we can see that the largest share of successes come from KSC LC-39A



Success and Failure for KSC LC-39A

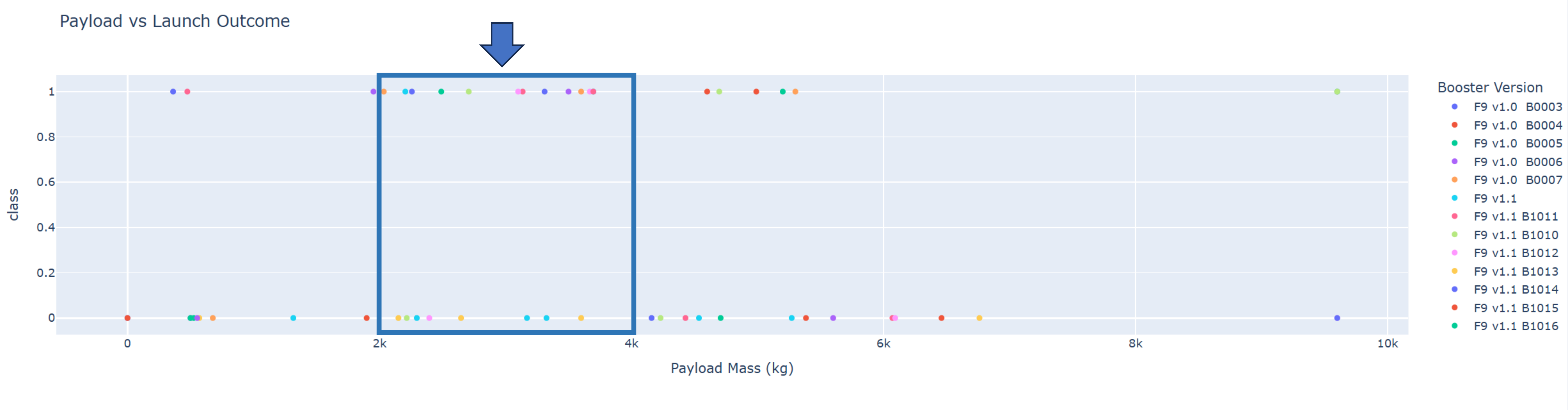
I know what're you're thinking: I said wouldn't say it anymore, but here I am, saying it. Hopefully this is REALLY the last time, but…. a Class of 0 indicates a failure, while a Class of 1 indicates a success.

We can see an almost 77% success rate at site KSC LC-39A



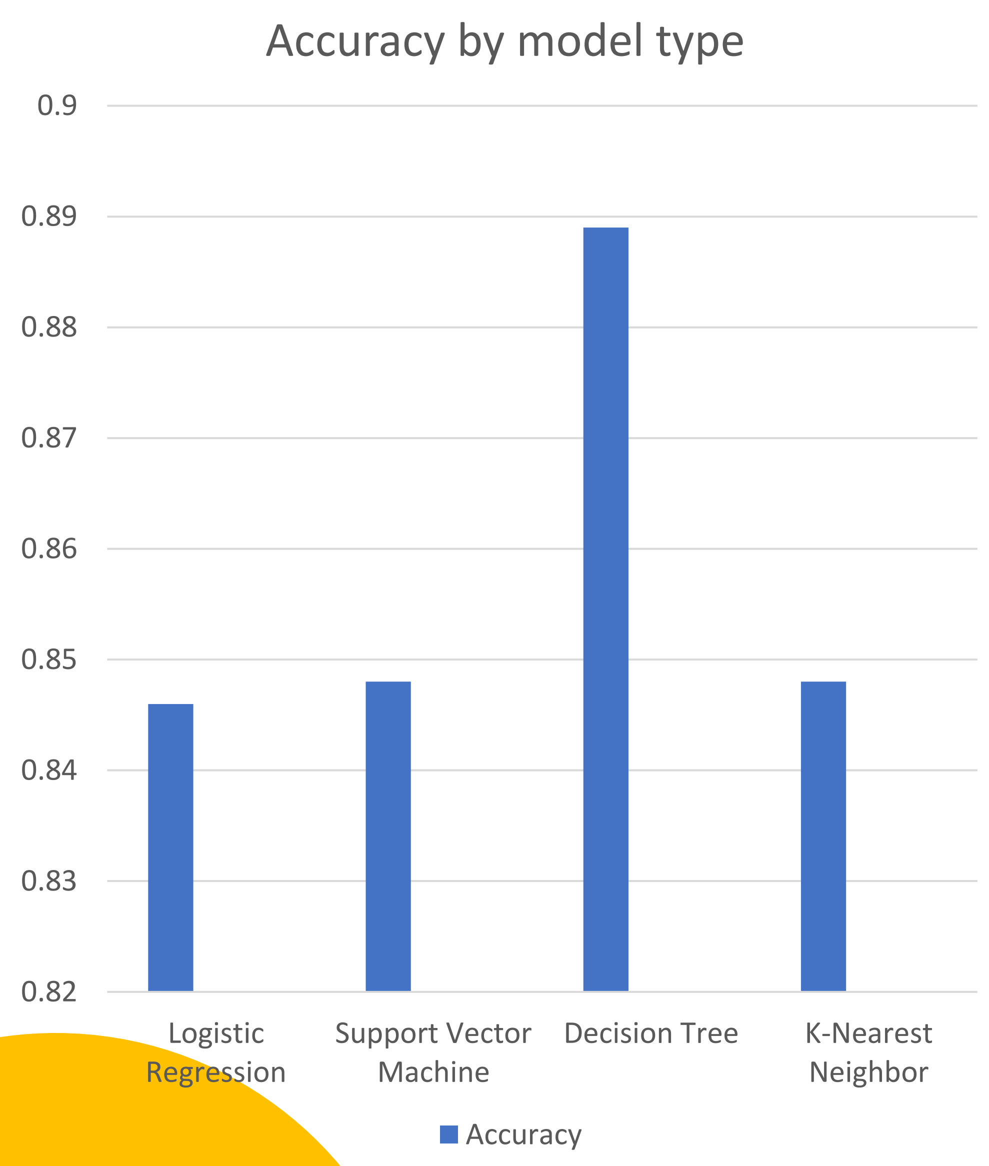
Payload vs. Launch Outcome

The first thing to do is to visualize the graph in terms of bands. The band with the most successes is 2000kg to 6000kg. Narrowing the band down even further takes us to 2000kg to 4000kg



Section 5

Predictive Analysis (Classification)

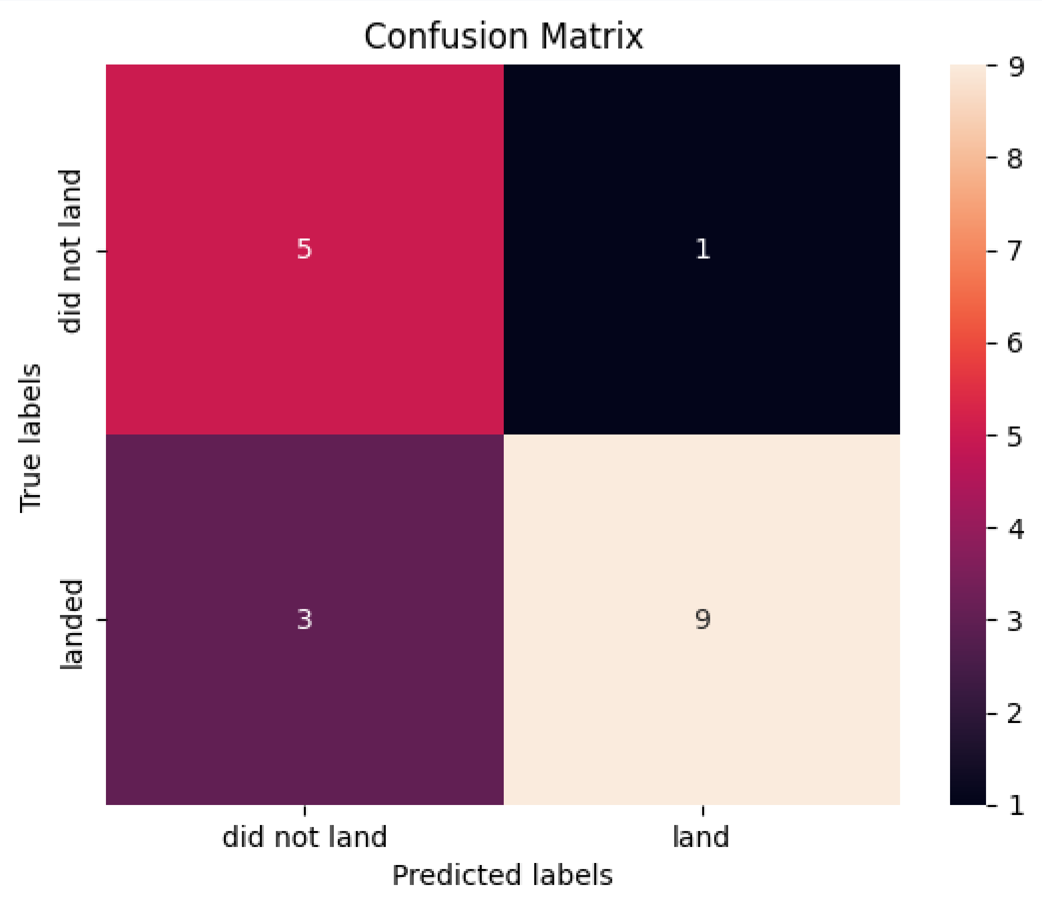


Classification Accuracy

This chart shows accuracy as a percentage.

Here we see that the Decision Tree has by far the highest accuracy

Confusion Matrix for Tree Decision



Based on our Confusion Matrix, we can see that there are the fewest False Positives. While there is room for improvement, any thing more accurate could be suspected of overfitting the data

Conclusions

To restate the conclusions:

1) From our Plotly Visualizations we can see that the optimal weight is between 2k and 4k kg.
2) Again using our Plotly Dashboard, we see that the site with the most success is KSC LC-39A
3) From our advance visualization with Seaborne, we can see that there are a variety of good choices for our orbit.
4) From our initial Data Wrangling, we see that the optimal recovery method is Drone Ship(ASDS)

2K-4K KG

STIE KSC LC-39A

ES-L1, GEO, HEO, SSO

DRONE SHIP (ASDS)

Thank you!