IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

David Villers
01/19/2024

Outline

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

Executive Summary

- Summary of Methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Visualization
  - EDA with SQL
  - Interactive Map using Folium
  - Interactive Dashboard using Plotly Dash
  - Predictive analysis using Machine Learning Techniques
- Summary of all Results
  - EDA Results
  - Interactive Results using Screenshots of Map/Dashboard
  - Predictive Analysis Results

Introduction

Project background and context

SpaceX is the most successful company of the commercial space age, making space travel affordable. The company advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; 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. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

Questions to be answered?

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

- Data collection methodology
  - SpaceX Rest API
  - Webscrape SpaceX Wikipedia page
- Perform data wrangling
  - Filter Data
  - Handle missing values
  - Utilize One Hot Encoding to prepare data for binary classification
- Perform EDA using visualization and SQL
- Perform interactive EDA with Folium and Dash
- Perform predictive analysis using classification models
  - Build and test various classification models using machine learning techniques

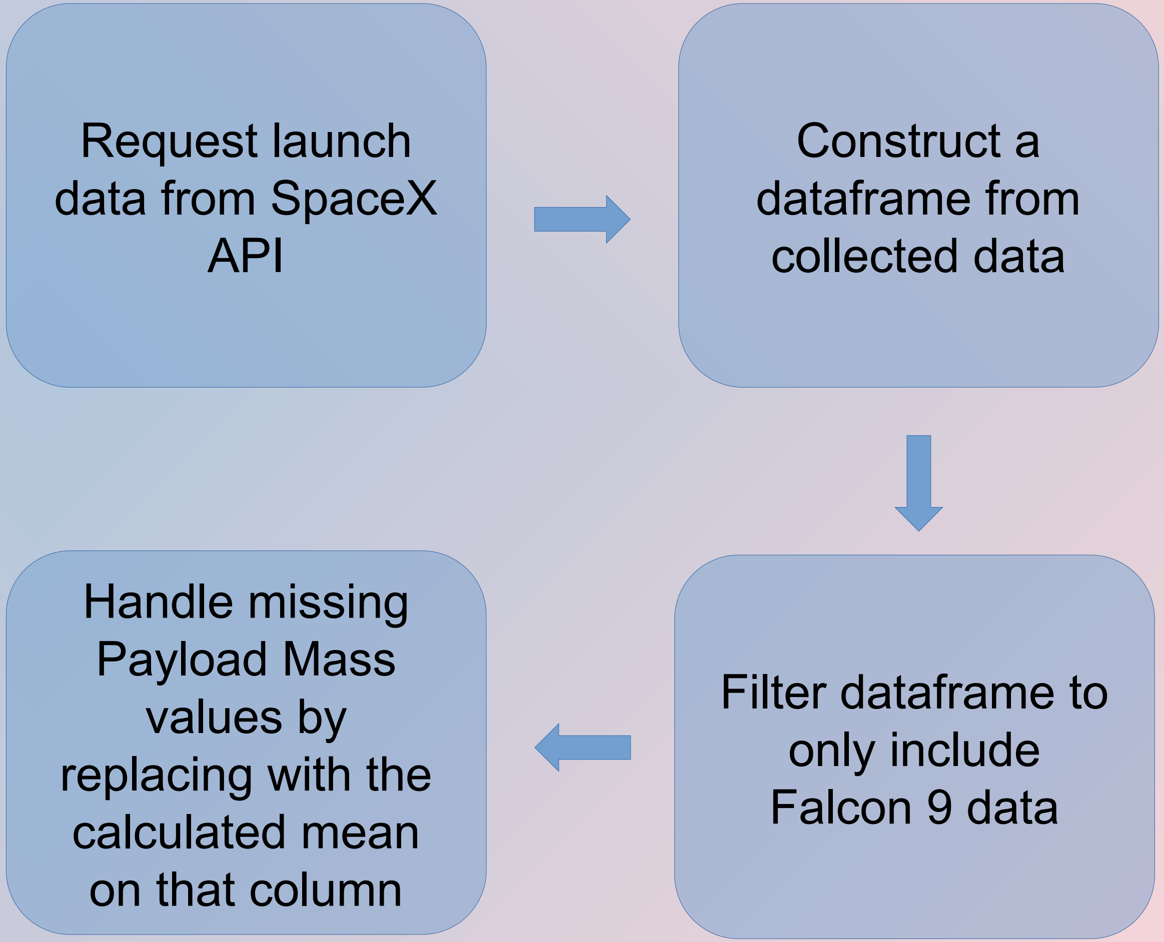
Data Collection

Data sets used we gathered from the following sources

- SpaceX API → https://api.spacexdata.com/v4/rockets/
- SpaceX Wikipedia → https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

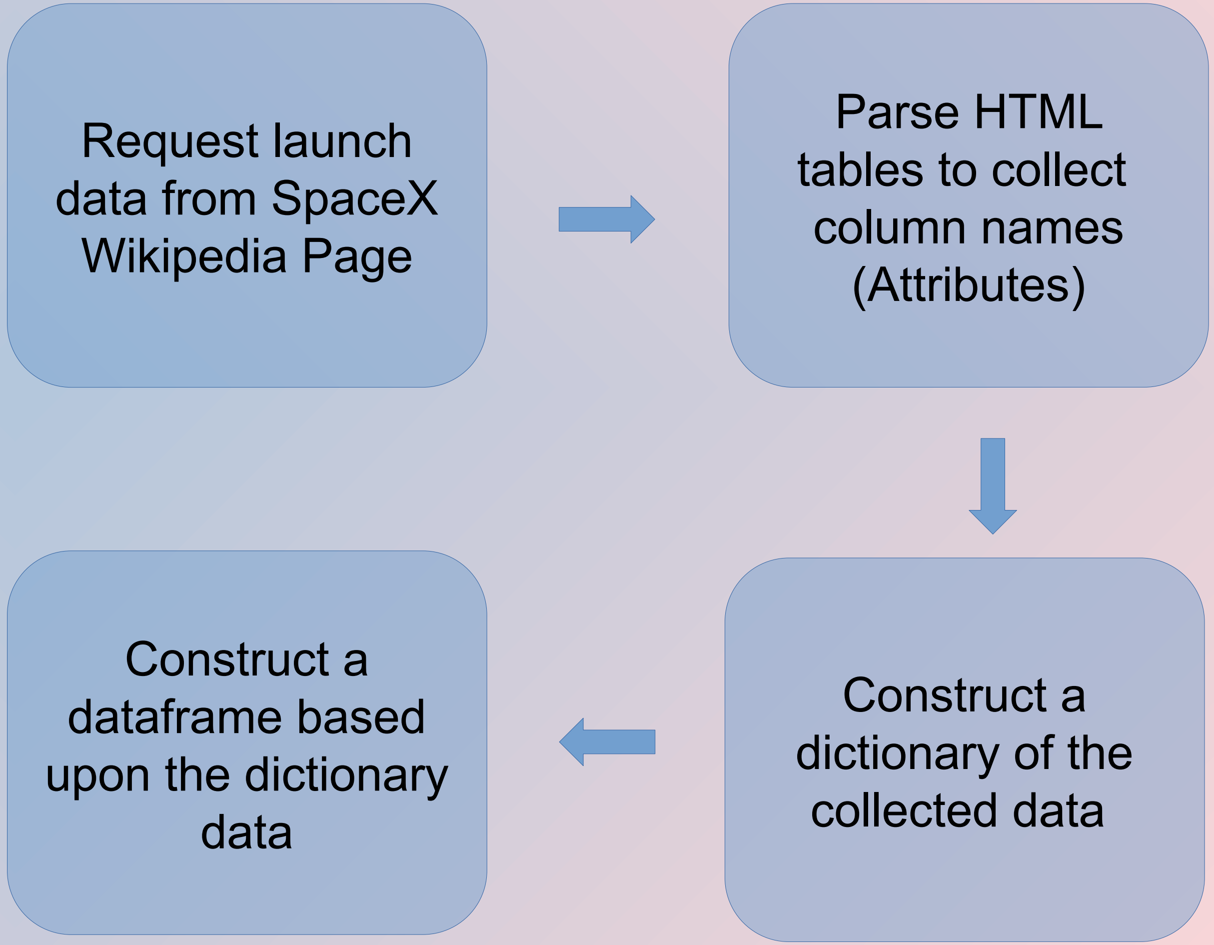
Data Collection – SpaceX API

- The SpaceX API is available to the public. We can use python to make a request to gather data and construct our database from it.
- Source code :
  Github: Data Collection API



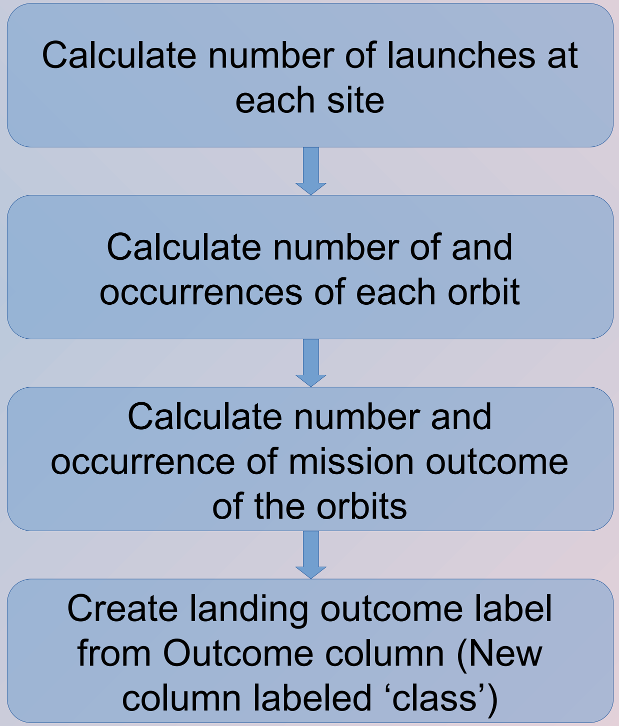
Data Collection – Web Scraping

- Using python we can perform webscraping on the SpaceX Wikipedia webpage to collect launch data.
- Source code:
  Github: Data Collection Web Scraping



Data Wrangling

Our data set indicates if the booster landed successfully. The result was listed as either True or False followed by a location such as “Ocean” or “Ground Pad.” We want to condense these results simply as to True or False and assign the results to a binary class (1 for True, 0 for False).



Github: Data Wrangling

EDA with Data Visualization

To help us better understand our data, we utilize various visualization tools and in this particular section we made several charts and graphs.

- Scatter Plots → Shows the relationship between variables
- Bar Plots → Shows a comparison among categories
- Line Plots → Shows trends in data over time

Gitgub: EDA with Visualization

EDA with SQL

Tasks to be performed using 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 succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- 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.
- 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: EDA with SQL

Build an Interactive Map with Folium

The following objects were displayed on the interactive map:

- Makers → Indicates locations of launch sites
- Circles → Highlights areas around specific launch sites
- Cluster Markers → Groups nearby launch sites together when zoomed out
- Lines → Indicates distance between launch site and nearest rail, city, ocean

Github: Folium

Build a Dashboard with Plotly Dash

User interaction features:

- Dropdown List → Contained option to select particular launch site or select ALL sites
- Min/Max Slider → Allows user to select a range (0-10000kg) for the payload mass

Plots displayed based upon user inputs:

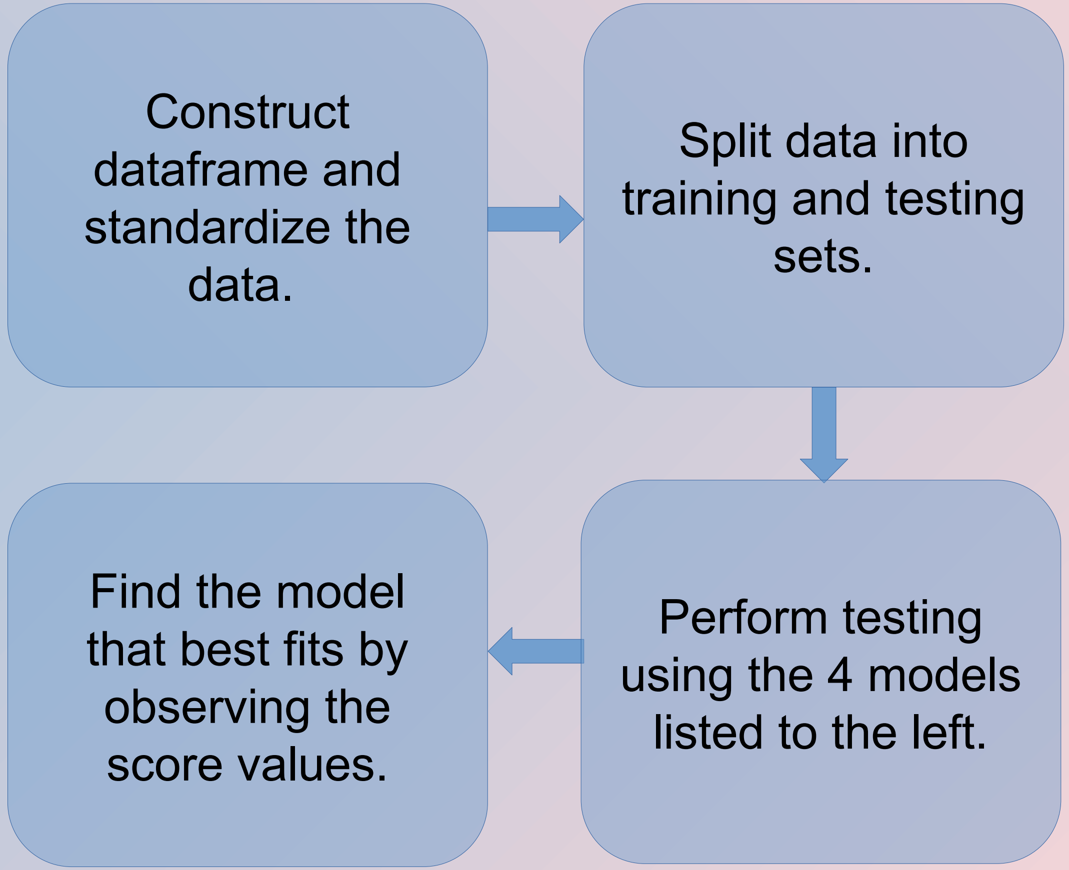
- Pie Chart → Shows success rate of launches
- Scatter Plot → Shows correlation between payload mass and launch success

Github: Dashboard

Predictive Analysis (Classification)

The following algorithms were used:

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K-Nearest Neighbors



Github: Predictive Analysis with Machine Learning

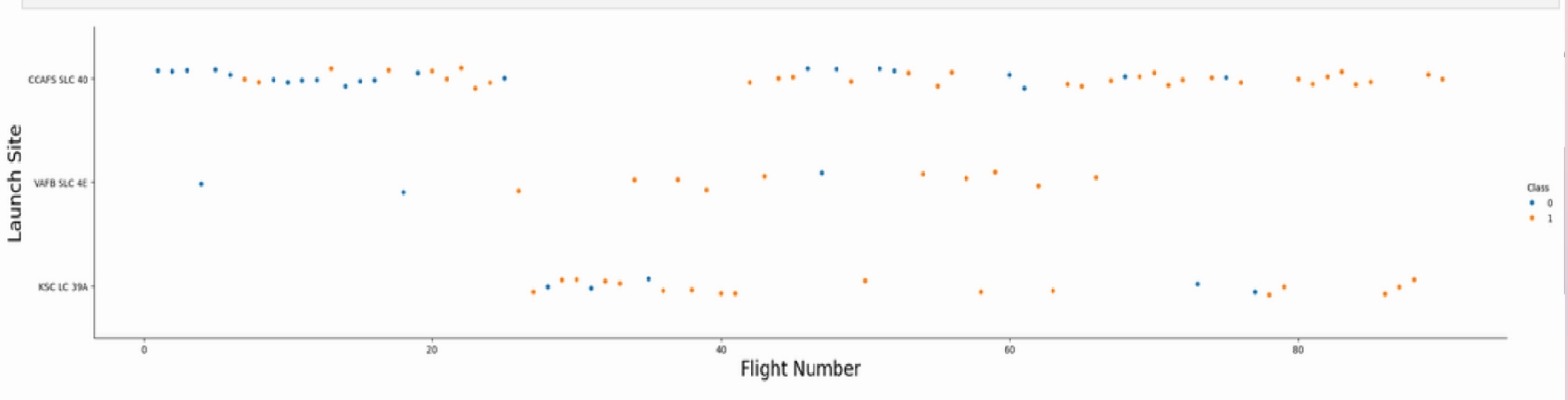
Results

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

Section 2

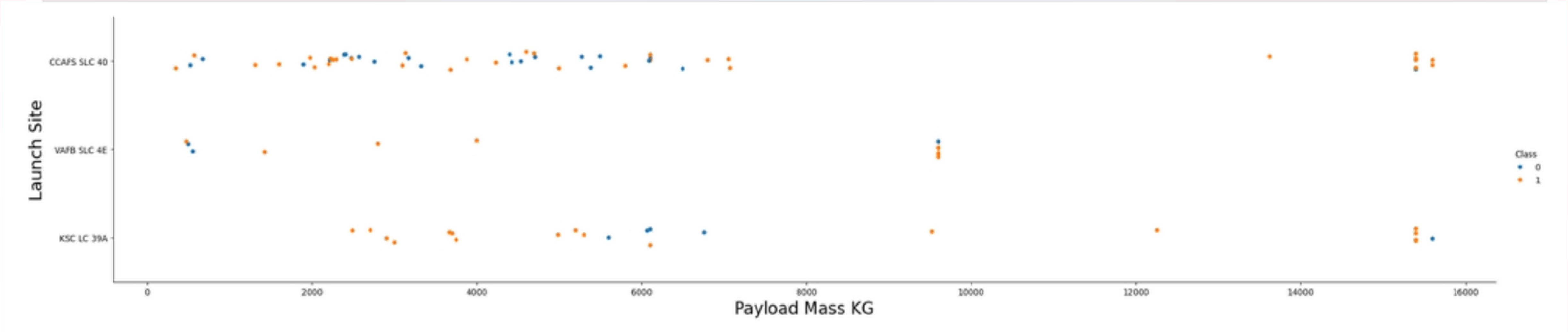
Insights drawn from EDA

Flight Number vs. Launch Site



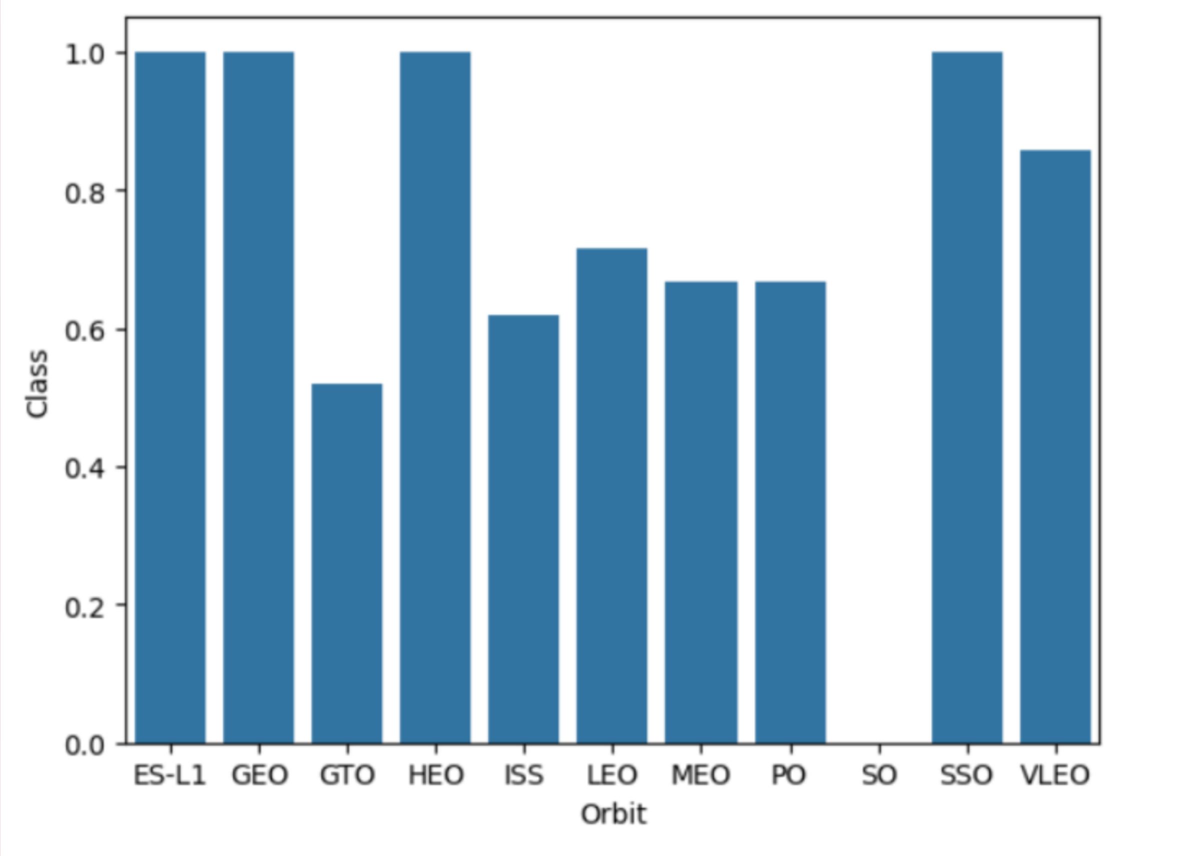
- Takeaways:
  - As time goes along, the success rate of the flights increases greatly
  - VAFB and KSC have better success rates compared to CCAFS
  - CCAFS accounts for nearly 50% of all launches, with success rate being high with later launches

Payload vs. Launch Site



- Takeaways:
  - The success rate increases as the mass increases among all 3 launch sites
  - KSC does show successful launches when the mass is under 5000 kg

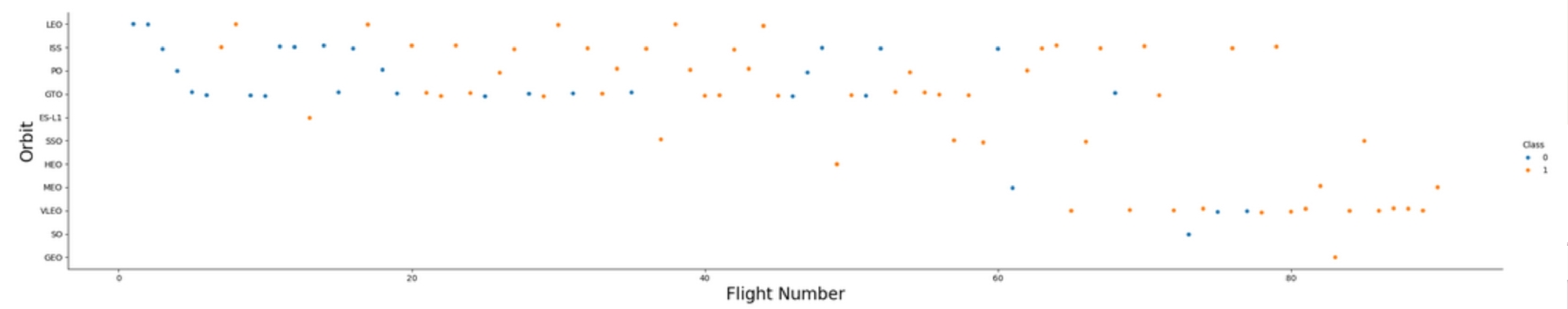
Success Rate vs. Orbit Type



Takeaways:

- Launches in orbits ES-L1, GEO, HEO, and SSO had a 100% success rate
- Launches into the SO orbit have yet to succeed
- Launches into the ISS, LEO, MEO, PO orbits yield roughly a 65% success rate

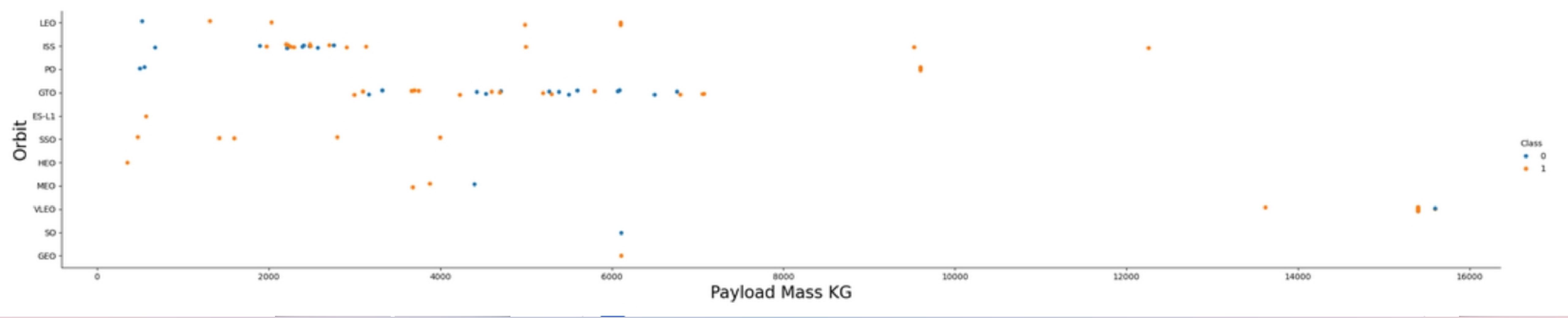
Flight Number vs. Orbit Type



Takeaways:

- In general, success rate has increased over time in each orbit
- The orbit VLEO has been used more frequently as of late and the success rate indicates this orbit will allow for a successful launch

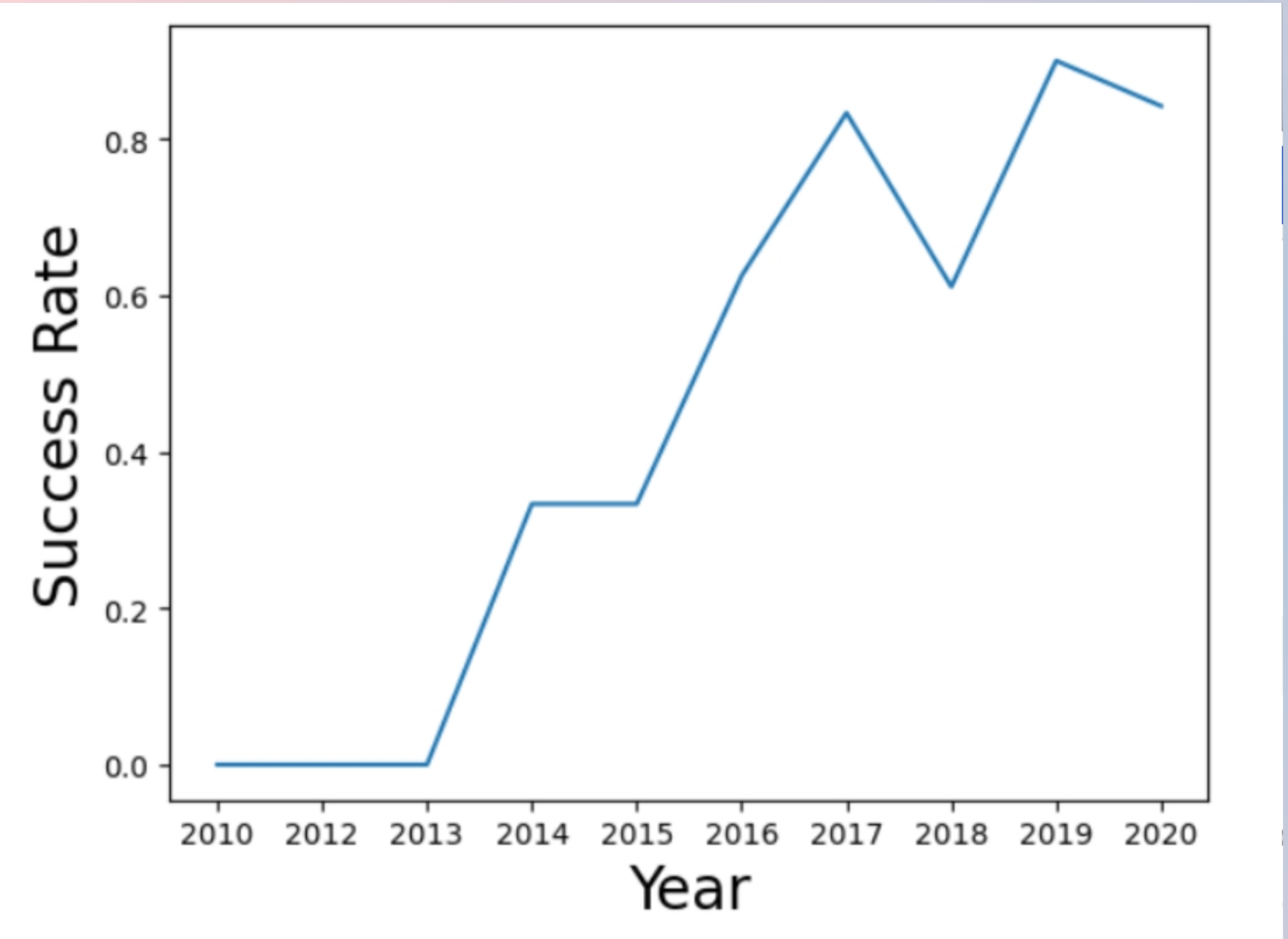
Payload vs. Orbit Type



Takeaways:

- For GTO we cannot gather much insight since there are a 50/50 mix of successful landings and unsuccessful landings
- PO, LEO, ISS do indicate the success rate increases as the payload mass increases

Launch Success Yearly Trend



- Takeaways:
  - Since the year 2013, the success rate has increased over time.

All Launch Site Names

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

Explanation:

- Performed SQL query to display the names of the Launch Sites

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (
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 (
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

Explanation:

- Performed SQL query to display the names of the Launch Sites that begin with ‘CCA’ and displaying the first 5 results

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
619967

Explanation:

- Performed SQL query to calculate the total mass of the payload launched by NASA

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

Explanation:

- Performed SQL query to calculate the average mass of the payload with booster version F9 v1.1

First Successful Ground Landing Date

MIN(Date)
2015-12-22

Explanation:

- Performed SQL query to view the fist successful ground landing date

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Explanation:

- Performed SQL query to view the booster versions that successfully landed on a drone ship while having a payload mass between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

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

Explanation:

- Performed SQL query to list the total number of successful and failure mission outcomes

Boosters Carried 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

Explanation:

- Performed SQL query to list different boosters that carried the maximum payload

2015 Launch Records

Month	Landing_Outcome	PAYLOAD_MASS__KG_	Launch_Site
01	Failure (drone ship)	2395	CCAFS LC-40
04	Failure (drone ship)	1898	CCAFS LC-40

Explanation:

- Performed SQL query to show the launch failures that occurred in 2015

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

Landing_Outcome	Counter
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

Explanation:

- Performed SQL query to show the landing outcomes between 6/4/2010 and 3/20/2017 in descending order

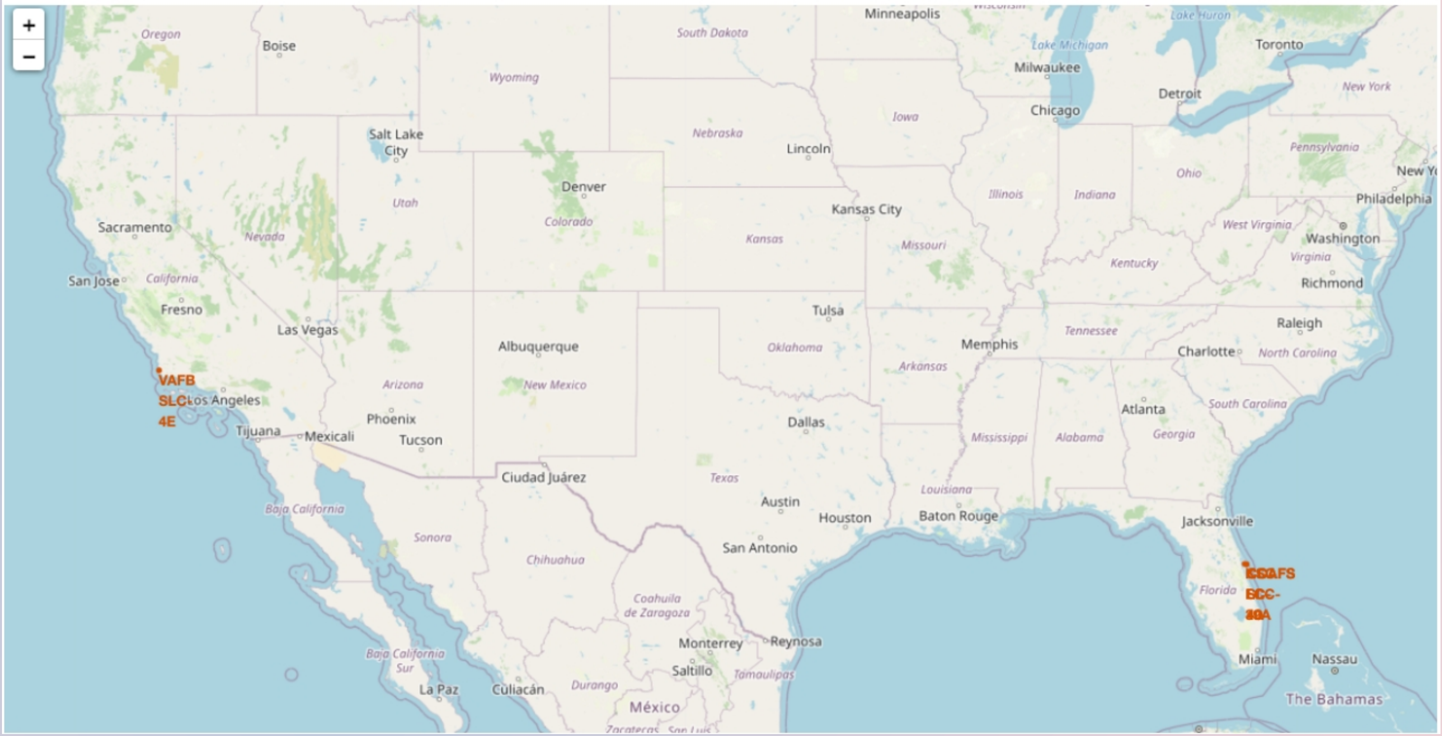
Section 3

Launch Sites Proximities Analysis

All Launch Sites

The launch sites have 2 noticeable characteristics about there locations

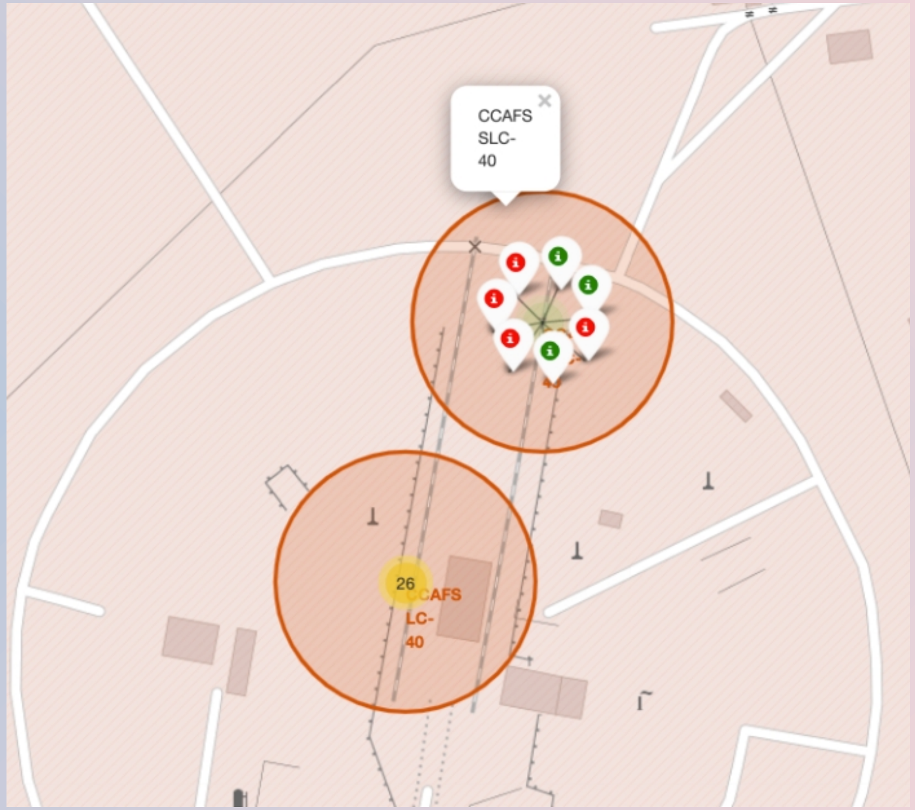
- The closer to the equator the rockets are when launched will help them stay in orbit since the land is rotating faster at the equator.
- The launch sites are located along the coasts which is likely due to safety concerns when launching and landing.



Launch outcomes by site

When we zoom in onto a launch site we can quickly discover how success that site has been with it's launches.

- **Red = Failed Launch**
- **Green = Successful Launch**



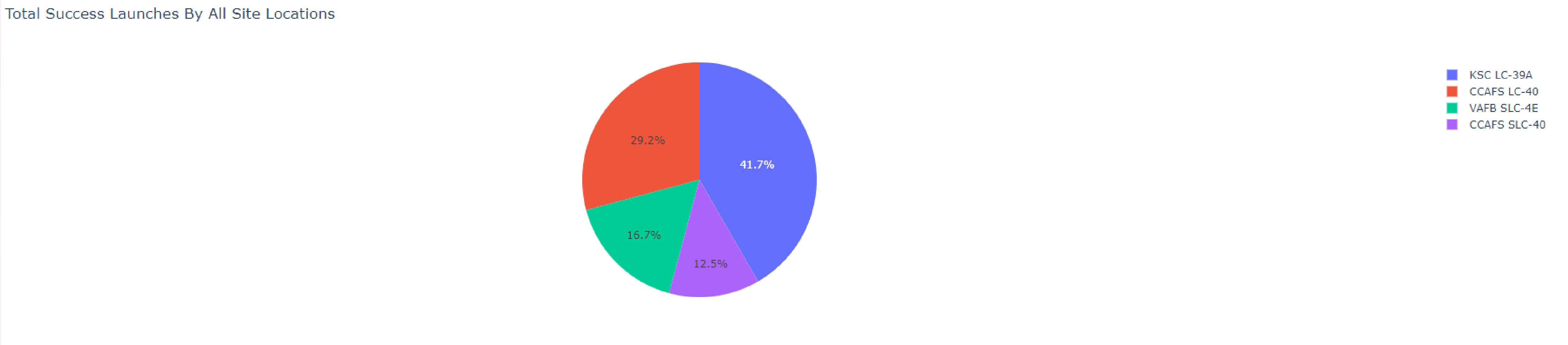
Proximity to coastline

Here we can show with a line the distance to the nearest coastline. We can assume the launch site is in close proximity to the coast to help mitigate safety concerns. Debris from the launch and landing process could potentially be harmful to civilians in the immediate area.

Section 4

Build a Dashboard with Plotly Dash

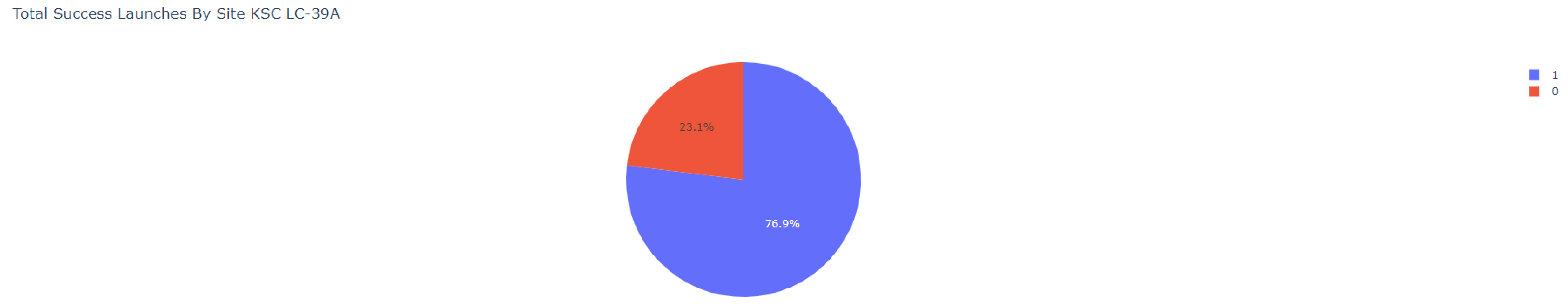
Successful Launches for All Sites



Takeaways:

- We can see that launches from the KSC LC-39A location have the best success rate. Nearly 42% of all the successful launches came from here.
- The opposite is true for the CCAFS SLC-40 and VAFB SCL-4E sites.

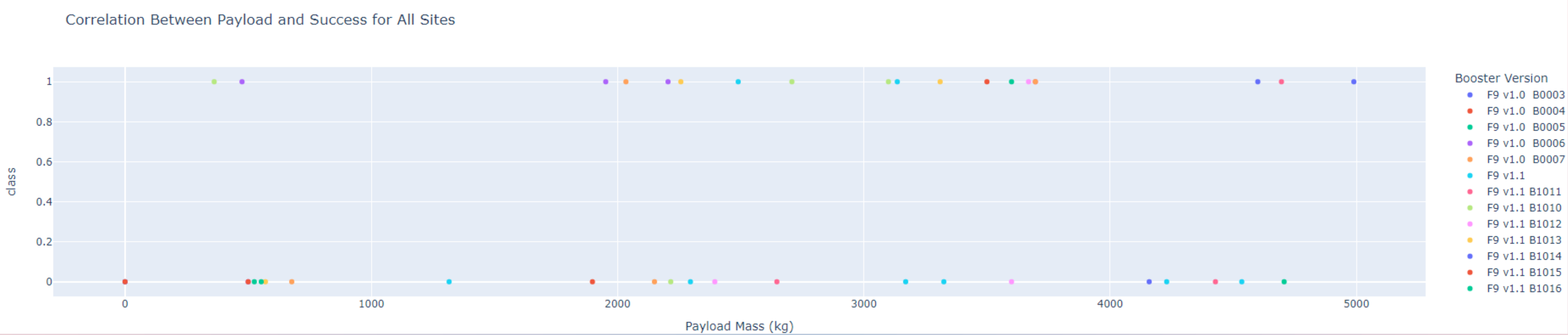
Most Successful Launch Site



Takeaways:

- We are able to see that 3 out of 4 launches have been successful at the KSC LC-39A site location. It has the highest success rate of any launch site in our study.

Payload Mass (0-5000 kg) vs Launch Outcomes



Takeaways:

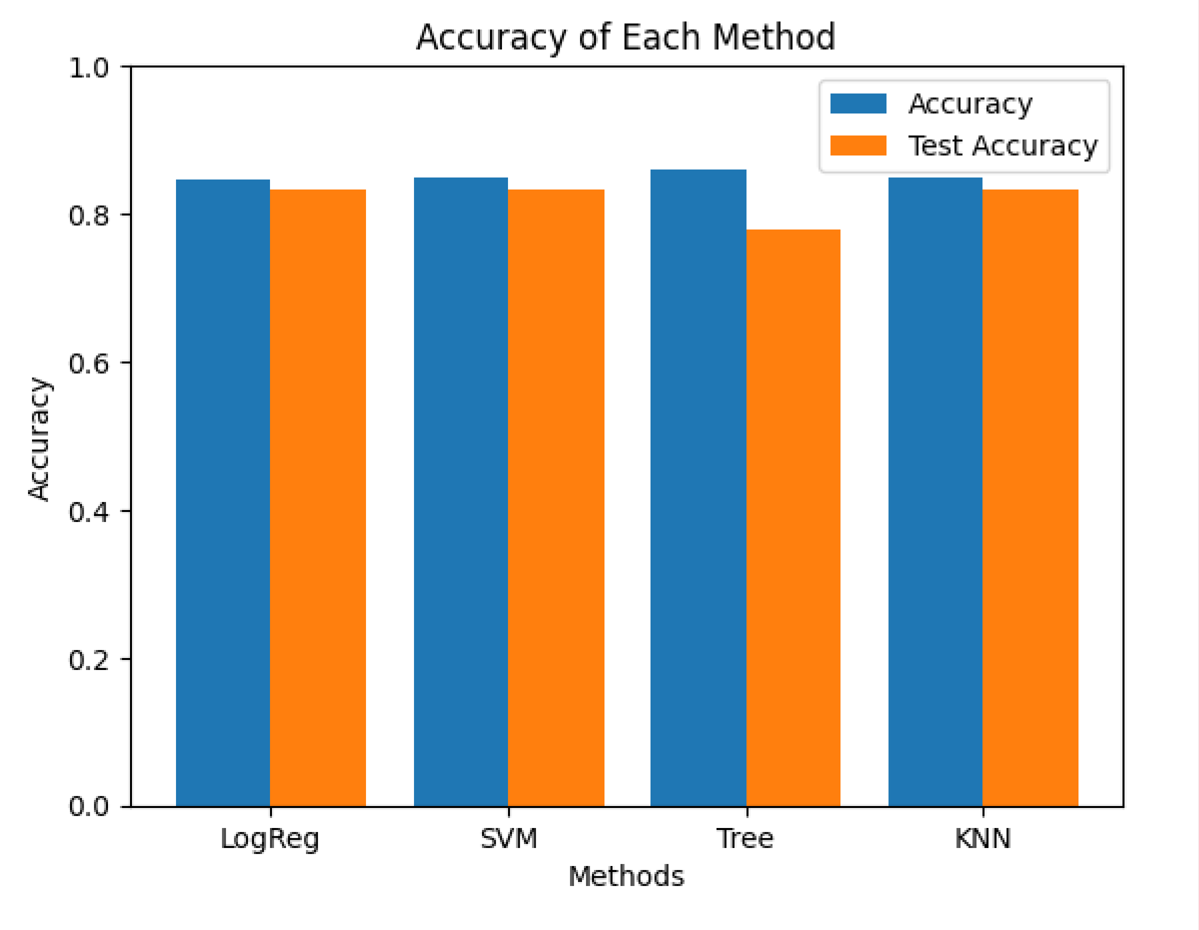
- Since a majority of the launches had a payload mass under 5000kg we can observe those outcomes easily with the given chart.
  - Under 2000kg → Very poor success rate
  - 2000-5000kg → Evenly split between success and failure

Section 5

Predictive Analysis (Classification)

Classification Accuracy

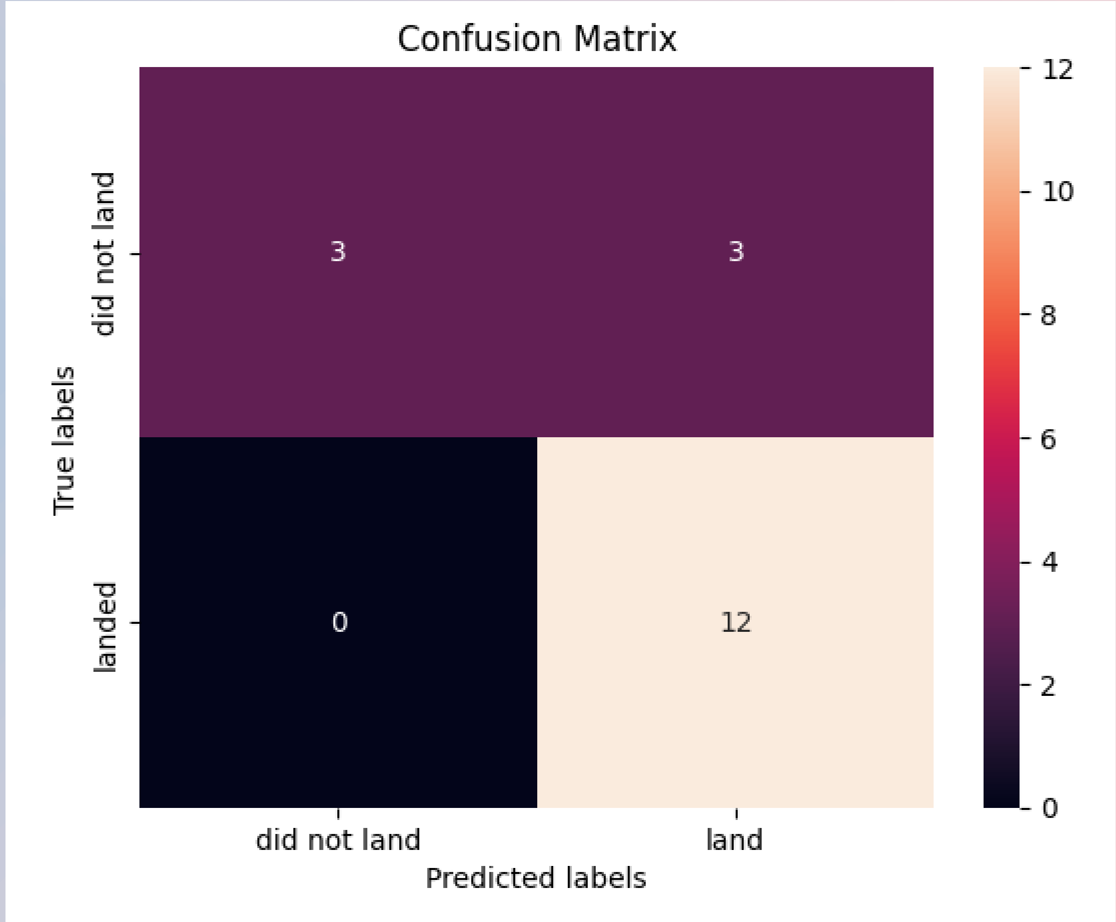
- According to our chart we can see the decision tree algorithm gave us the highest accuracy just edging out the other methods tested.



Confusion Matrix

According to our decision tree confusion matrix we can see our results:

- True Negative = 3
- False Positive = 3
- False Negative = 0
- True Positive = 12



Conclusions

- Our best algorithm to use was the decision tree
- Light payload masses (under 2000kg) showed they are likely to fail
- Heavier Payload masses (over 6000kg) showed better success
- Overall success rate has greatly improved over time
- KSC LC-39A is the most successful launch site (77%)
- CCAFS SLC-49 is the least successful launch site (57%)
- SSO, GEO, HEO, ES-L1 are good orbits to launch into (100% success rate)
- SO orbit has yet to yield a success launch

Thank you!