IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kyle House
May 17th, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection through API and web scraping
  - Data wrangling
  - Exploratory data analysis with SQL and data visualization
  - Interactive visual analysis with Folium
  - Machine learning predictions
- Summary of all results
  - Exploratory data analysis results
  - Screenshots of interactive analytics results
  - Predictive analysis results

Introduction

- Project background and context

  Space X offers delivery of a payload into outer space at a cost of 62 million US dollars. While the best their competition can offer for the same service is 165 million US dollars. This reduction in cost is primarily due to Space X's efforts into making large and expensive portions of the rocket assembly reusable. For a company to bring down their costs to compete with Space X, what is the cost effective path to target to streamline their own systems while trying to develop Space X's reusability.

- Problems you want to find answers
  - What factors have a strong correlation on whether the first stage of a Space X Falcon 9 lands successfully?
  - What are the ideal conditions to target to start to develop a launch system of our own?

Section 1

Methodology

Methodology

Executive Summary

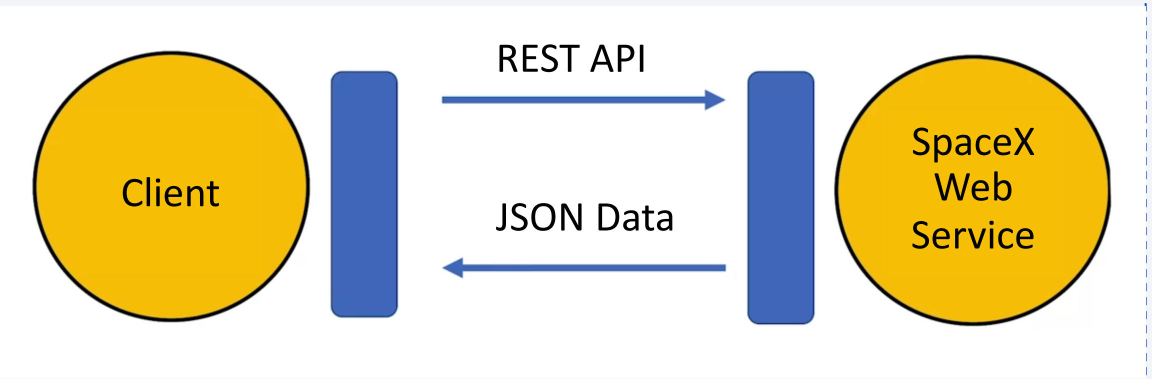
- Data collection methodology:
  - Data was gathered from the SpaceX website via a REST API with additional data from web scraping from Wikipedia.
- Perform data wrangling
  - Data was cleaned up with some missing values having the mean added to those entries and to change the first stage landing success from true/false to 1/0.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - After being wrangled, the data was normalized, divided into training and test sets, and evaluated and optimized with four different learning methodologies.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

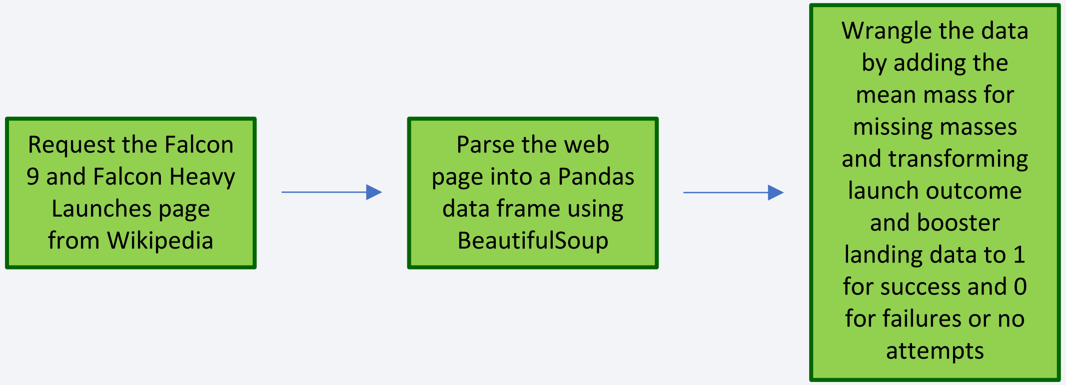
Data Collection – SpaceX API

- Space X had available a public API where the history of their entire program can be accessed
- The methodology in accessing this resource can be followed starting from the Client circle on the accompanying flowchart.
- Source code: Link



Data Collection - Scraping

- Data pertaining to the Falcon 9 launches can also be retrieved from Wikipedia.
- This data is accessed, parsed, and tabled in the accompanying flowchart.
- Source code: [Link]



Data Wrangling

- To get an idea of what data was available to us we used some Exploratory Data Analysis (EDA) to get column headings as well as to find any missing elements in each launch's entry.
- Once an understanding of what data was provided was grasped, missing data in priority columns were filled with the best data available, and alphabetical data that could be represented with binary or mathematical were updated to reflect that option.
- Source code: Link

EDA with Data Visualization

- To investigate the data, scatterplots and bar plots were used to visualize the relationships between priority elements.
  - Payload Mass vs. Flight Number
  - Launch Site vs. Flight Number
  - Launch Site vs. Payload Mass
  - Orbit vs. Flight Number
  - Orbit vs. Payload Mass
- Source code: Link

EDA with SQL

- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission
  - Top 5 launch sites whose name began with the string ‘CCA’
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome on ground pad was achieved
  - Names of the boosters which have successfully landed on a drone ship with a payload between 4000 and 6000 lbs.
  - Total number of successful and failed mission outcomes
  - Names of the Booster Versions which have carried the maximum payload mass
  - Booster Version, Launch Site, Launch Month, for Failed Landings in 2015
  - Count the successful landing outcome between June 4th, 2010, and March 20th, 2017
- Source Code: Link

Build an Interactive Map with Folium

- Markers, circles, lines and marker cluster were created and place at points of interest with the Folium Maps Python package
  - Circles were used show the area where the launch sites utilized by the Falcon 9 launches were located
  - Markers were used as an identification tag for the created circles so we could see which launch site was which circle
  - Lines were used to measure the distance between launch sites and nearby points of interest
  - Marker Clusters were used to show, with a click of the main site marker, the landing fate of each Falcon 9 first stage landing attempt
- Source code: Link

Build a Dashboard with Plotly Dash

- Utilizing the Plotly package we created the following
  - Pie charts
    - When no launch site is selected it shows the percentage and count of the successful launches from each site
    - When a specific launch site is selected it shows the percentage and count of the success or failure of the first stage landing attempt
  - Scatter plot
    - With an accompanying slider, the scatter plot shows the success of the launch between Payload Mass and Launch Success.
    - The points are also colored based upon their Booster Version
- Source Code: Link

Predictive Analysis (Classification)

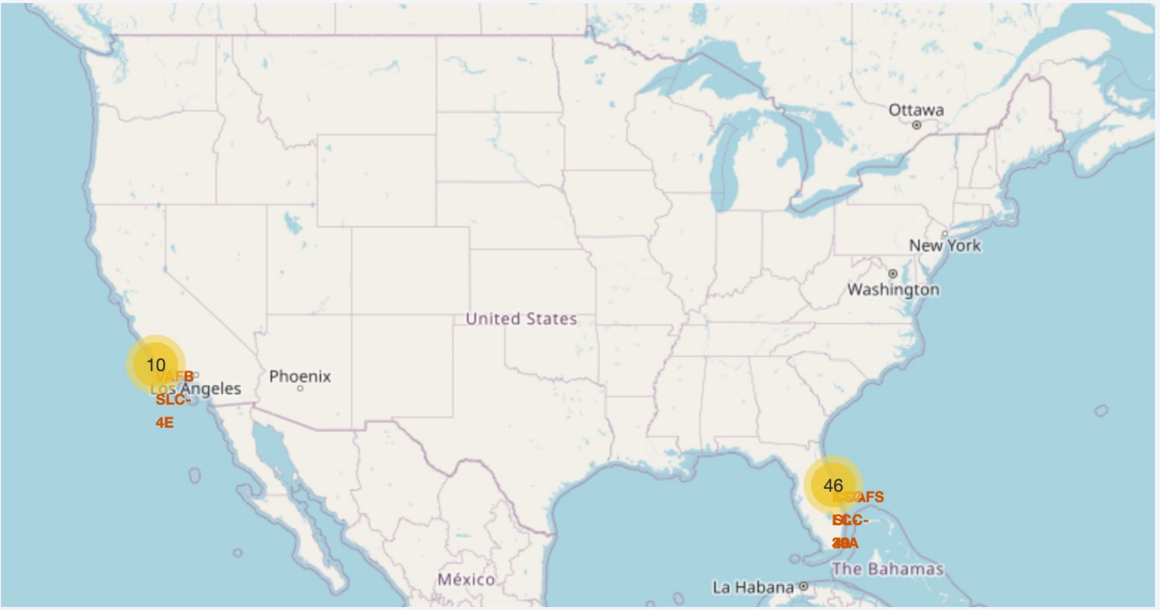
- Utilizing Numpy and Pandas, we transformed our data, and divided it into Training and Testing sets.
- Using GridSearchCV and hyperparameters, we optimized four different machine learning models:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K - Nearest Neighbor
- To judge if a change in parameter improved a model and which model was best optimized, we used accuracy results from the test set of data.
- Source Code: Link

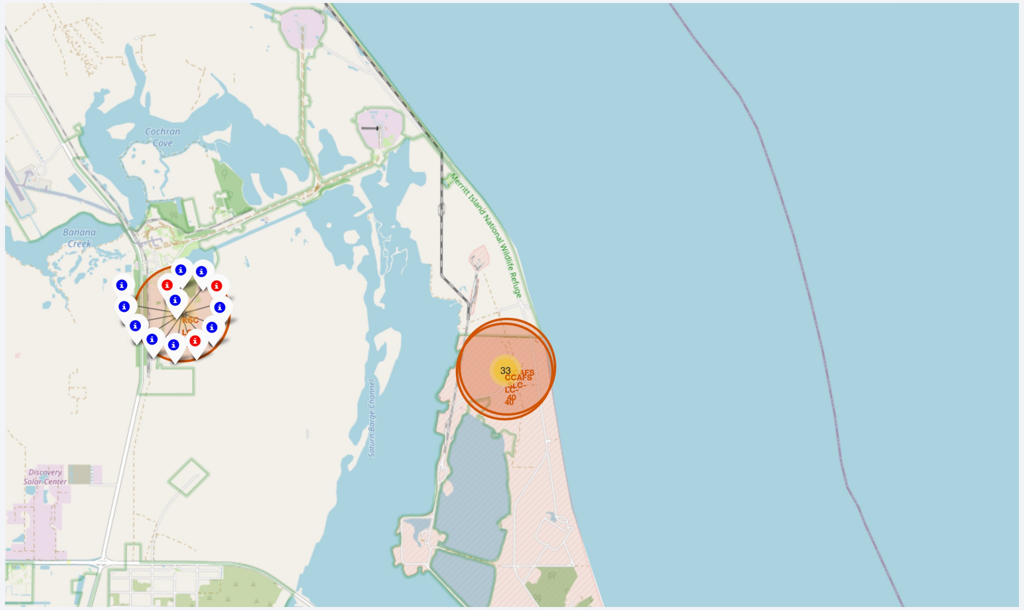
Results

- Exploratory data analysis results:
  - Space X has used four different locations to launch Falcon 9 and Falcon Heavy launch systems
  - The average payload of the Falcon 9 v1.1 is 6442 lbs.
  - The first successful, first stage landing occurred five years after the first launch.
- Predictive analysis results

Results

- Using Interactive analysis we were able to identify that there are two key features that make an ideal launch site
  - Existing launch infrastructure
  - Public safety down range from the launch site. Either over an ocean or sparsely populated desert environments.



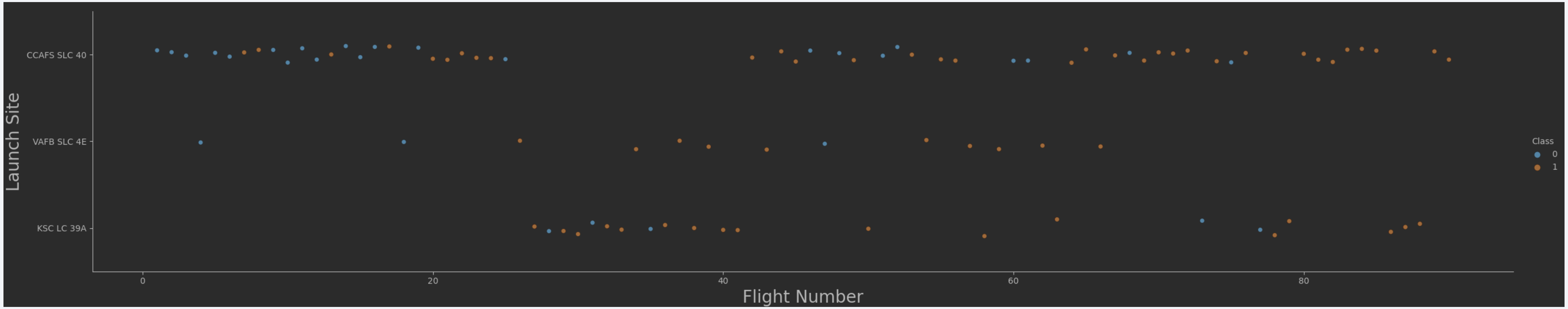


Section 2

Insights drawn from EDA

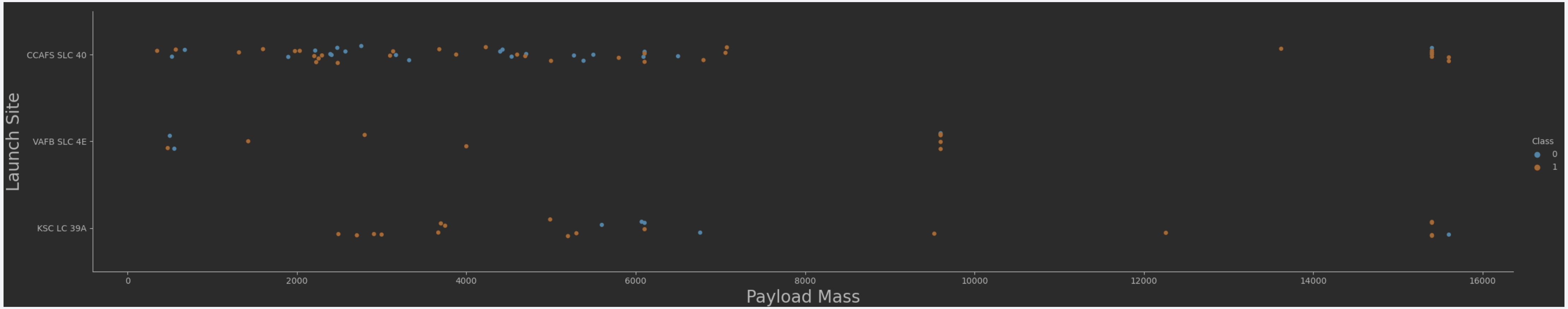
Flight Number vs. Launch Site

- With more experience at a launch site, the probability of a successful launch has increased.



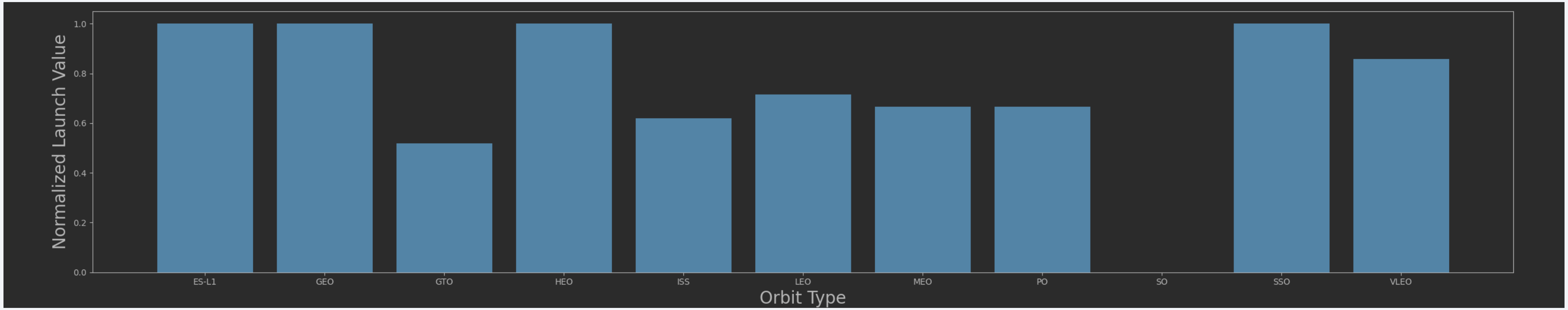
Payload vs. Launch Site

- The payload mass of each launch from each launch site, colored by successful/unsuccessful launches.



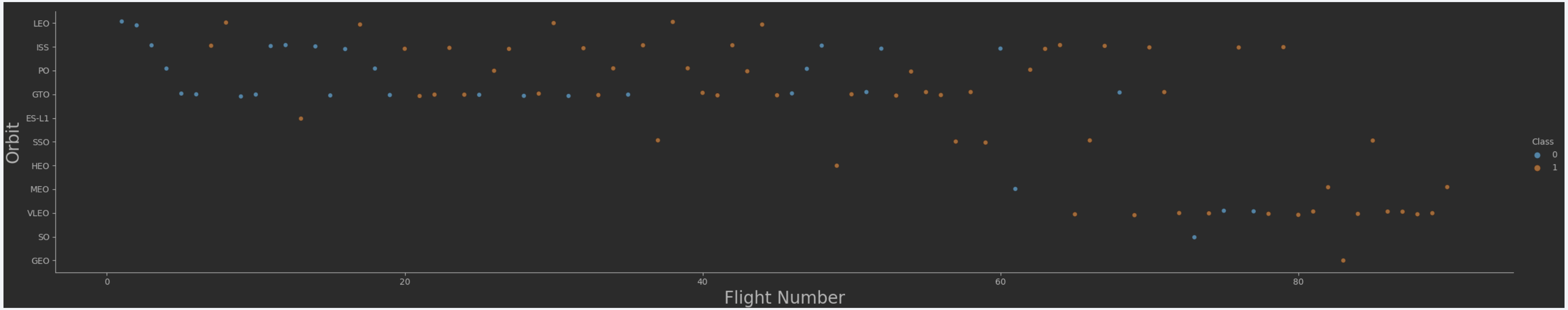
Success Rate vs. Orbit Type

- The success rate of launch by each target's orbit



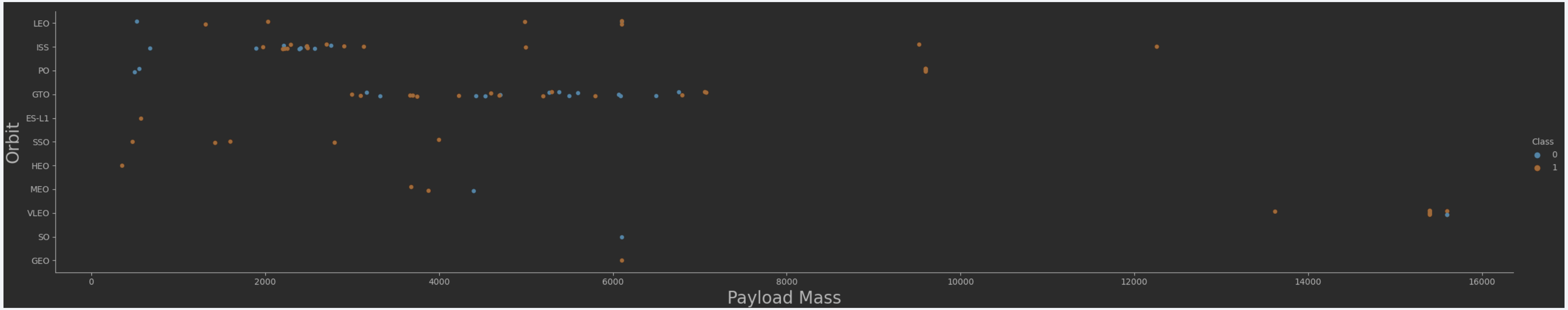
Flight Number vs. Orbit Type

- The orbit target by flight number, color coded based on successful launch



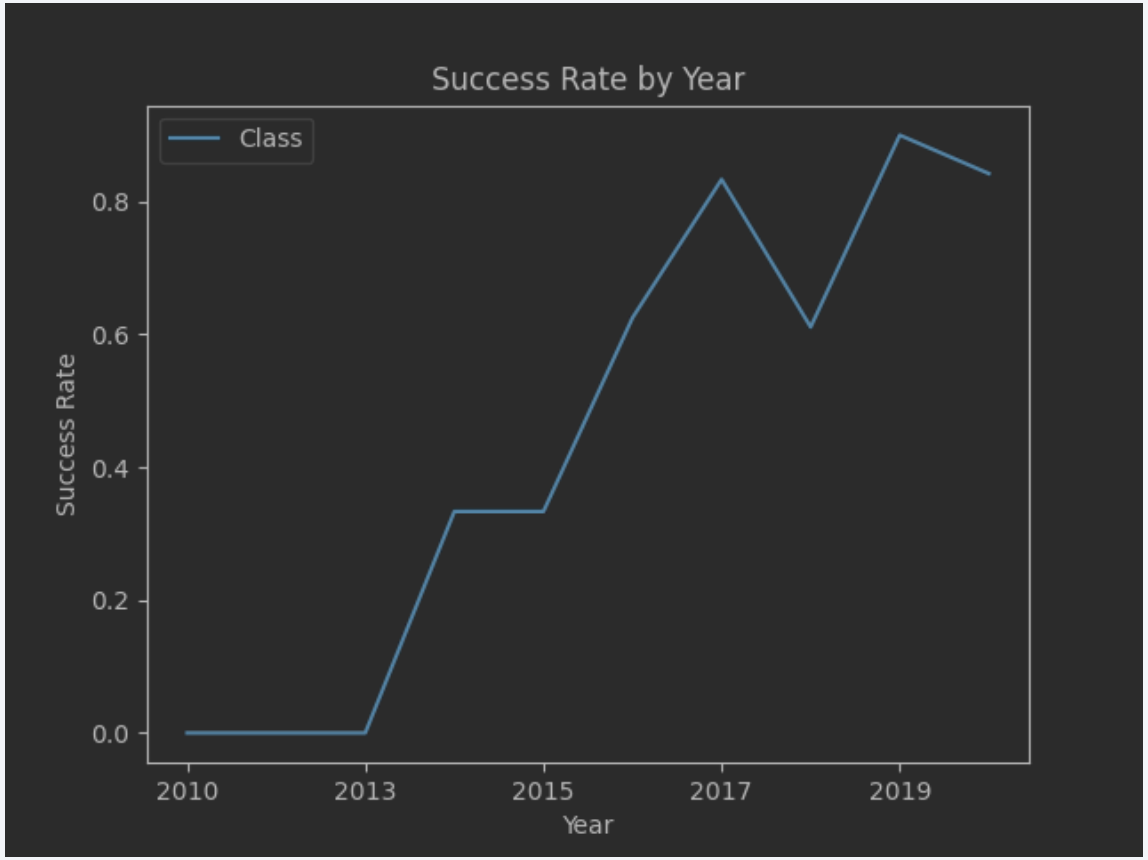
Payload vs. Orbit Type

- The orbit target by payload mass, color coded based on successful launch



Launch Success Yearly Trend

- Line plot of the launch success rate by year



All Launch Site Names

- Result of SQL query for a list of distinct launch sites

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

Launch Site Names Begin with 'CCA'

- First 5 records with a launch site starting with “CCA”

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Ou
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
08-12-2010	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
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

- Total mass from SQL query

sum(payload_mass__kg_)
45596

Average Payload Mass by F9 v1.1

- Average mass of Falcon 9 v1.1 payload from SQL query

avg(payload_mass__kg_)
2928.4

First Successful Ground Landing Date

- First successful first stage on a ground pad landing date from SQL query

min(date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 from SQL query

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes from SQL query

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

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

2015 Launch Records

- List of failed landing attempts on a drone ship, in the year 2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- Ranked count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

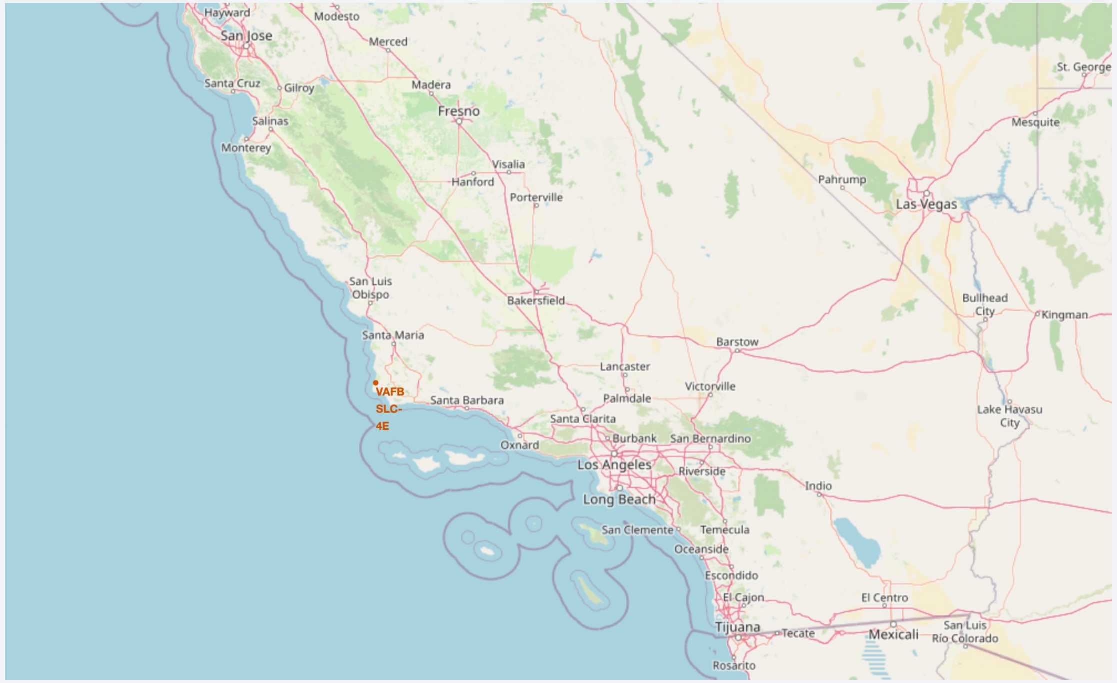
1	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

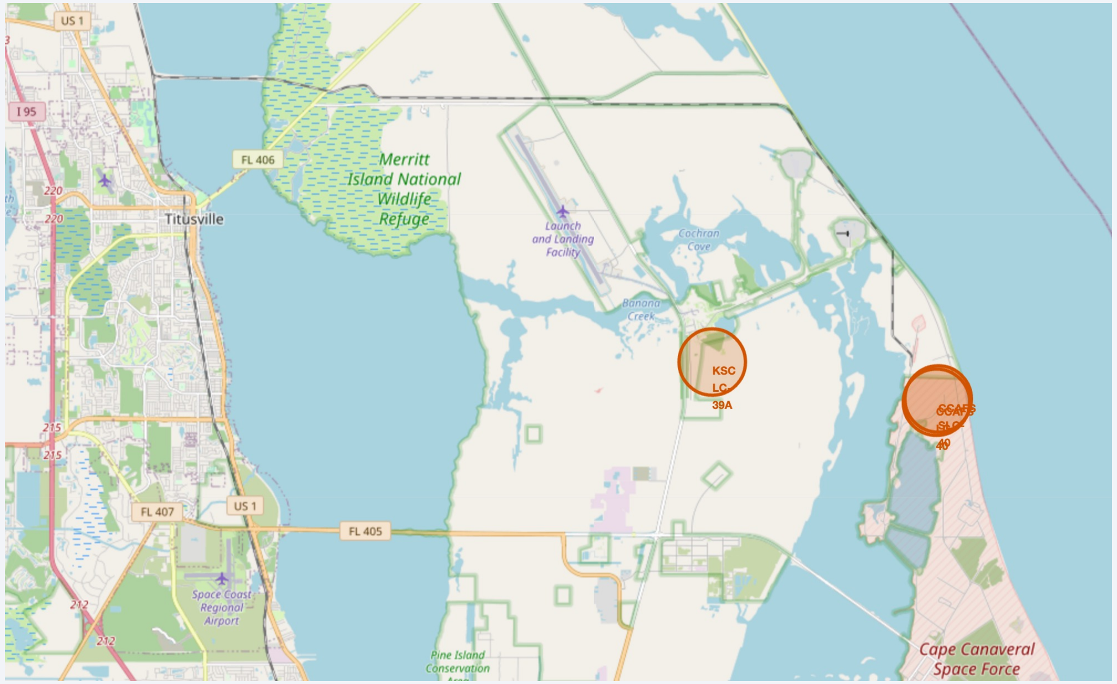
Section 3

Launch Sites Proximities Analysis

Launch Site Locations

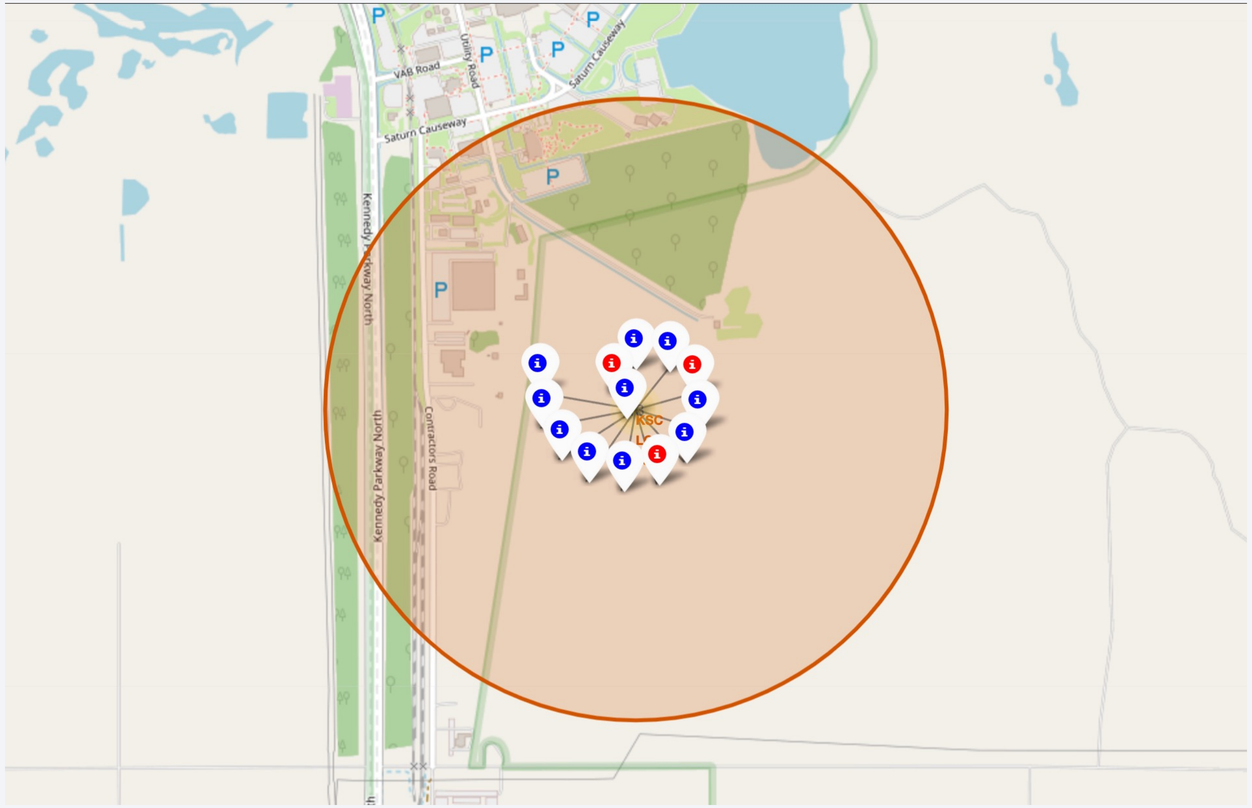
- The four called out locations in the maps below are the launch sites for the Falcon 9 launch system. It is a little unclear but the farthest east site has two separate launch pads that counts as two sites.





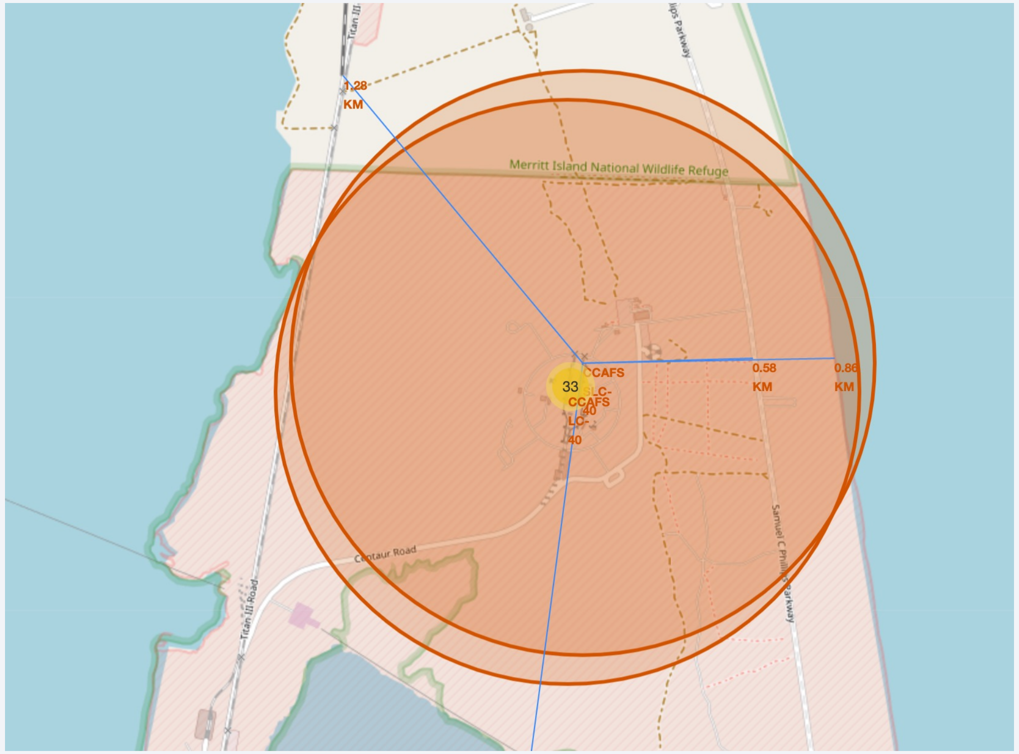
Success Rate of a Launch Site

- This is an example of the color coded labels at each launch site. Blue is for a successful launch, red is for a launch that did not complete its goal. This example is KSC LC-39A.



Distances from Points of Interest

- This show the chord distances from launch site CCAFS SLC-40 to various points of interest nearby. The two lines running east are the distance to the closest road (0.58 km) and shoreline (0.86 km) while the line to the northwest shows the closest railroad (1.28 km). The line to the south is going to the nearest international airport, located in Melbourne, FL 51.4 km, away.

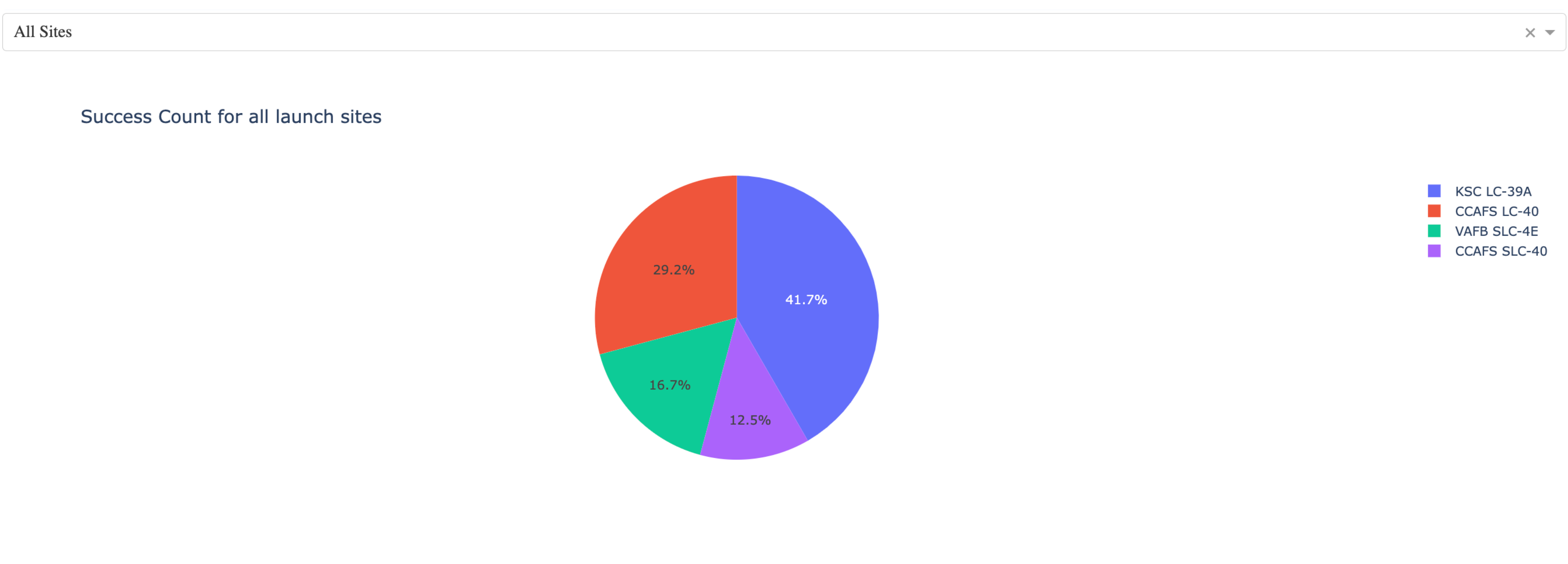


Section 4

Build a Dashboard with Plotly Dash

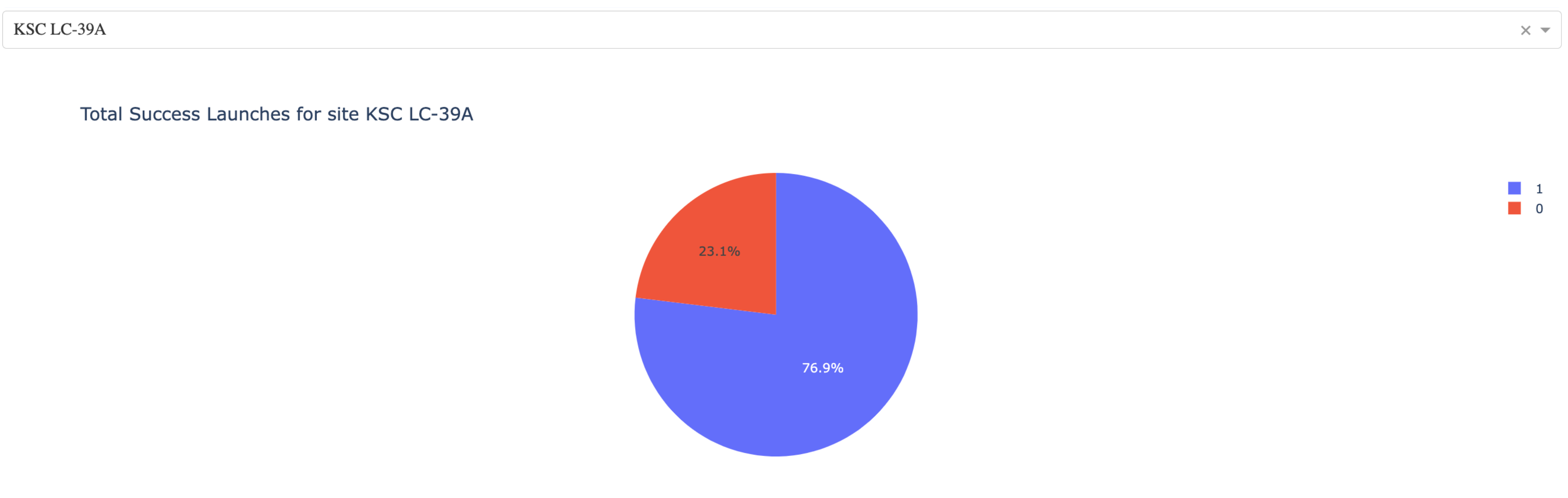
Launch Success broken down by Launch Site

- Pie chart showing the rate of successful launches broken down by launch site.



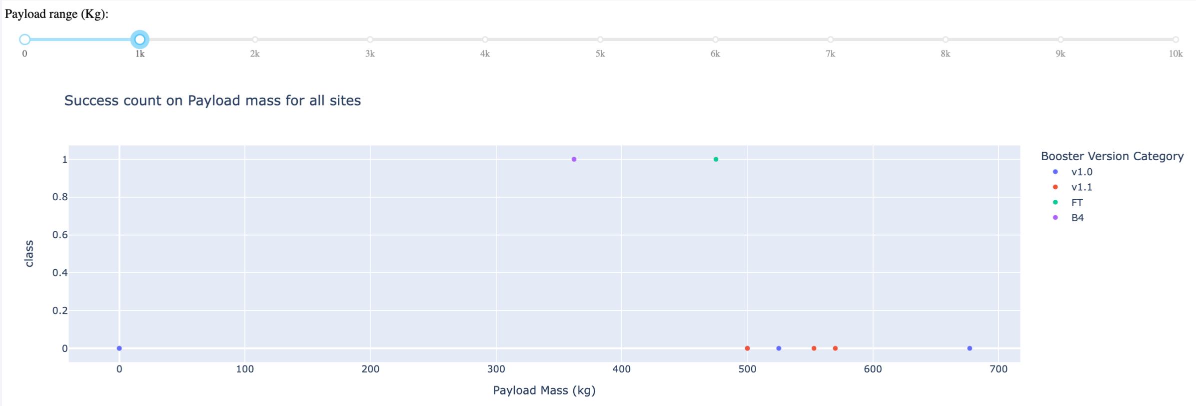
KSC LC-39A Success Rate

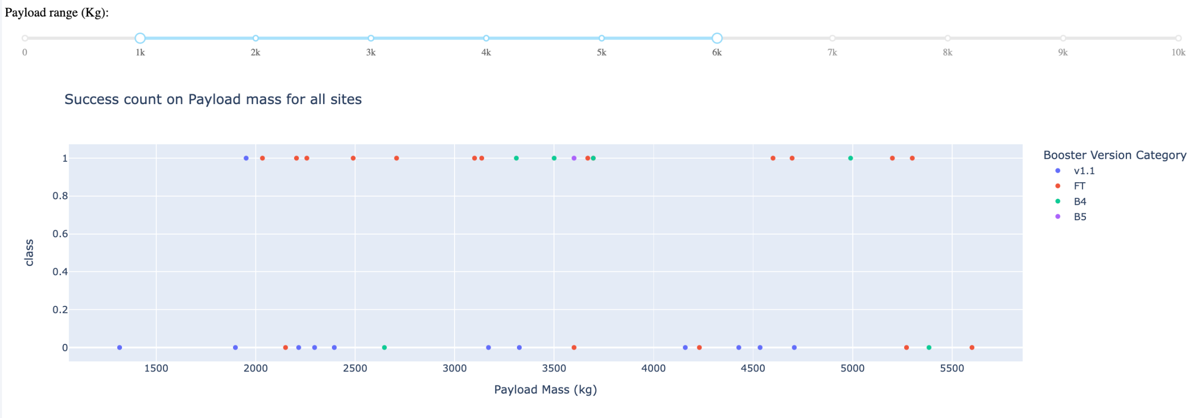
- With a 76.9% successful launch rate, launches from KSC LC-39A have the best track record.



Effect of Payload Mass on Launch Outcome

- Launches with very light (sub 1000 kg) and very heavy (over 6000 kg) payload masses have a greater rate of failure. This is due to these launches being early on in development for the lighter masses, and also larger strain tests on the launch system for the heavier masses.



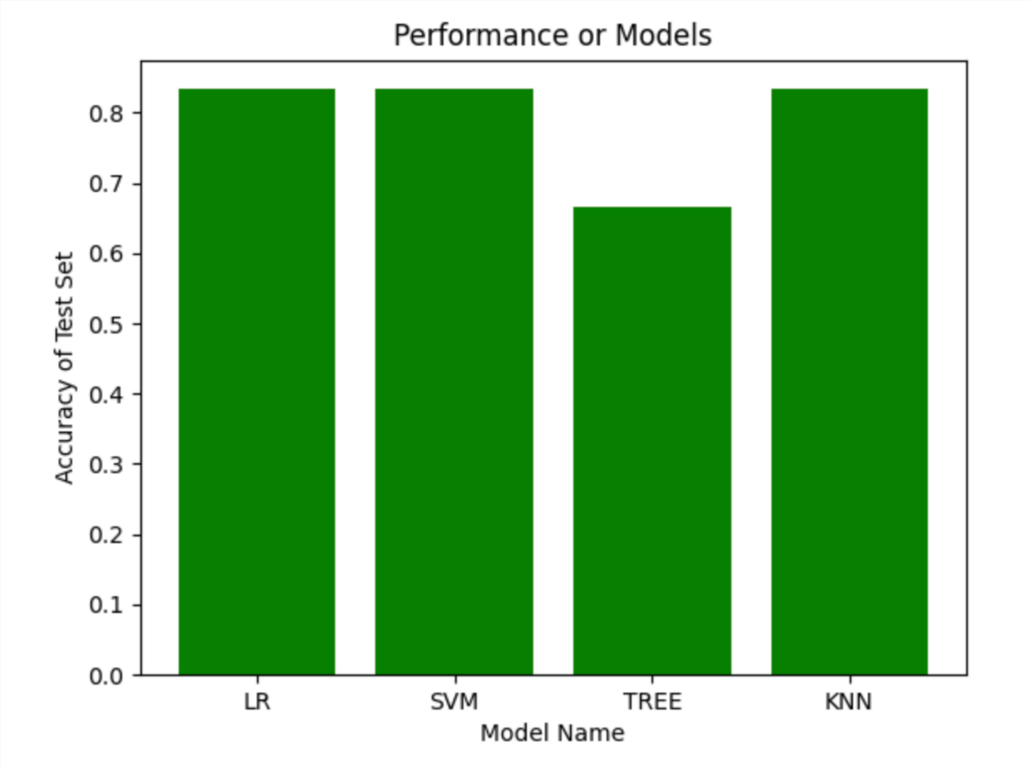


Section 5

Predictive Analysis (Classification)

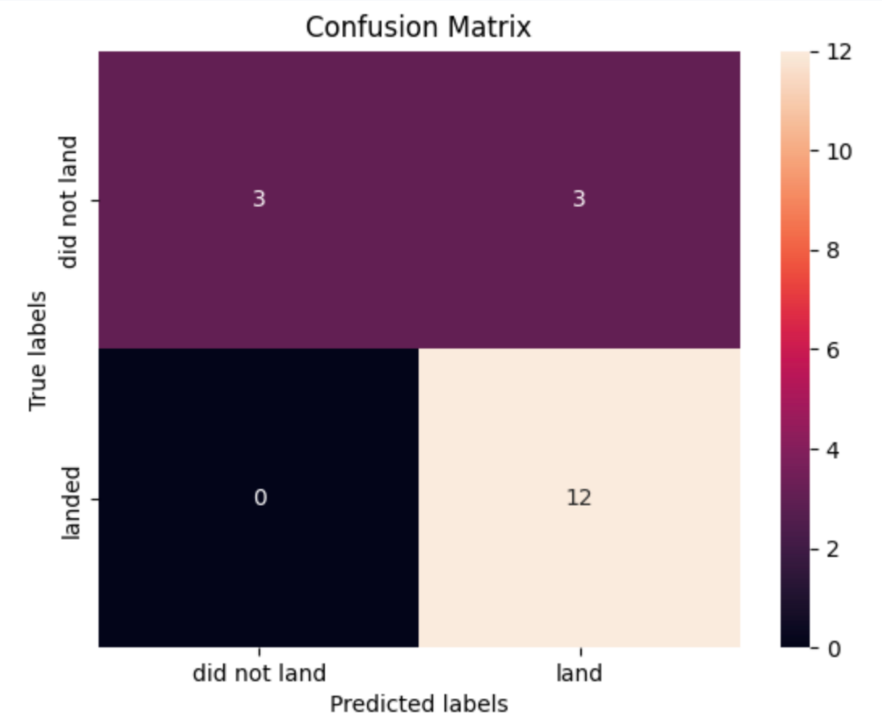
Classification Accuracy

- With optimization Linear Regression, Support Vector Machine, and K-Nearest Neighbor all produce an equally accurate model.
- With its inherent randomness the accuracy of a Decision Tree Classifier can produce a better accuracy model but its inconsistency in getting to that result make over fitment a larger risk compared to other methods in this application.



Confusion Matrix

- The confusion matrix for each of the best performing model. All three well performing models (Linear Regression, Support Vector Machine, and K-Nearest Neighbor) all produce the same confusion matrix.



Conclusions

- More experience at a launch site increases the likelihood of a successful launch.
- More experience launching a particular system increases the likelihood of a successful launch.
- With the exception of early teething problems, a payload mass between the 25^{th} and 75^{th} percentile of the maximum payload mass leads to increased launch success.

Thank you!