IBM Developer
SKILLS NETWORK

Rocketing to Success: Predicting Falcon 9 First Stage Landings for Cost-Efficient SpaceX Launches

Leigh Cohen
5/13/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

Executive Summary

I started by gathering data from the SpaceX API and Wikipedia using HTTP requests and web scraping with Beautiful Soup. After creating a Pandas data frame, I conducted exploratory data analysis (EDA) and used SQL queries to analyze SpaceX launch data, identifying trends in payload capacities and mission outcomes.

Using Pandas and Matplotlib, I explored launch success rates and orbital patterns, highlighting insights like varying success rates across different sites and orbits over time. Visualizing launch sites on a map with Folium revealed geographic clustering and proximity to key features like coastlines, influencing launch logistics.

I developed a Plotly Dash application for interactive analytics of SpaceX launches, featuring dynamic pie charts and scatter plots based on user-selected launch sites and payload ranges. The application's methodology included creating interactive components and implementing callback functions for real-time data visualization and analysis.

Lastly, I conducted exploratory data analysis, defined training labels, standardized data, and optimized SVM, Classification Trees, and Logistic Regression models using GridSearchCV. Most of the models performed equally, with specific hyperparameters, achieving high accuracy in predicting launch outcomes, as confirmed by the Confusion Matrix's metrics. However, there is definitely room for improvement and more analysis can be done to optimize the combination of features that would produce a successful Falcon 9 rocket launch.

Github URL: SpaceX Project Repository including all of the Notebooks and Python files

Intro

In this project, the goal is to determine the cost of the Falcon 9 rocket launch.

SpaceX can reuse the first stage and we know that Falcon 9 rocket launch cost 62 million dollars. Therefore, if we can find out whether the first stage will land, we can determine the cost of the rocket launch.

Problem: Find out whether the first stage will land.

Methodology

Methodology- Data Collection

Method 1: HTTP GET Request

1. Define functions that take in the dataset as an argument and send requests to a web server using the HTTP GET method to retrieve rocket launch data from the SpaceX API.

2. Use Pandas' json_normalize method to normalize JSON data into a flat table and then convert the JSON result into a Pandas data frame for further visualization and analysis.

3. Prepare the data obtained from the API by selecting relevant columns, filtering out certain rows based on specified conditions, transforming data formats, and restricting dates to a specified range. This processed data can then be used for further analysis or visualization.

Github URL: Data Collection- Sending HTTP GET requests to the API

FLOW CHART

Send HTTP GET requests to API → Parse data from API → Create Pandas Data frame

Methodology- Data Collection

Method 2: Web Scraping from Wiki Page

1. Perform web scraping using Panda's BeautifulSoup library to extract Falcon 9 historical launch records from an HTML table on a Wikipedia page displaying the List of Falcon 9 and Flacon Heavy Launches.

2. Parse the table and convert it into a Pandas data frame for further visualization and analysis.

[Github URL: Data Collection- Web Scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia]

FLOW CHART

Web scrape from Wiki page ➜ Parse data from Wiki page ➜ Create Pandas data frame

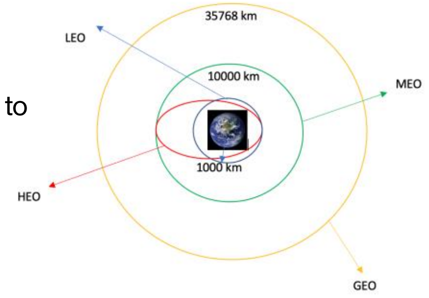
Methodology- Data Wrangling

Perform exploratory data analysis (EDA) using Pandas and NumPy to find patterns in the data and determine the labels for training supervised models.

<u>Exploratory data analysis:</u>

- Find the % of missing/NULL values in each attribute
- Identify numerical vs. categorical attributes
- Count the number of launches per site
- Count the number and occurrence of each orbit

Each launch aims to a dedicated orbit:



<u>Determine the training labels for training supervised models:</u>

- Create a set of 'bad' outcomes referring to launches in which the second stage didn’t land successfully
- Create a landing outcome label from the 'Outcome' column by applying a lambda function that creates a list where its element = 0 if the corresponding row is in the 'bad' outcomes set and element = 1 if corresponding row is not in the set. Then convert it to a list.
- Add this list as a new column ‘Class’ to the data frame
- Apply the mean function on that column ‘Class’ to determine the success rate

(since all values are 0 or 1, if the mean > 0.5 → it is more successful & vice versa)

<u>Training Labels:</u>

- True Ocean: mission outcome successfully landed in a specific region of the ocean
- False Ocean: mission outcome unsuccessfully landed in a specific region of the ocean
- True RTLS: mission outcome successfully landed to a ground pad
- False RTLS: mission outcome unsuccessfully landed to a ground pad
- True ASDS: mission outcome successfully landed to a drone ship
- False ASDS: mission outcome unsuccessfully landed to a drone ship

[Github URL: First Stage Landing Prediction via Data Wrangling]

<u>FLOW CHART</u>

Exploratory data analysis → Determine training labels

Methodology-
Exploratory Data Analysis using Visualization

Perform exploratory data analysis and feature engineering using Pandas, Matplotlib, and Seaborn for visualization.

- Create several catplots (scatterplots displaying the relationship between a categorical variable and a numerical variable), bar charts, line plots to check for correlations between features.
- After gaining preliminary insights into which features influence the success rate, select the features that will be used in success prediction later on. Then, use One Hot Encoding using the get_dummies() function to convert each categorical features into numerical dummy variables, in order to prepare the data frame for features engineering.

[Github URL: Exploratory Data Analysis with Visualization and Feature Engineering]()

FLOW CHART

Create plots → One Hot Encoding

Methodology-
Exploratory Data Analysis using SQL

Load data set into a SQL table using SQLAlchemy. Then, perform SQL queries to explore and analyze the data.

Run queries to….

- Display launch sites by name
- Display payload mass by boosters launched by NASA (CRS)
- Run aggregations to find the average payload mass of Booster Version F9 v1.1
- Display the first successful landing outcome in ground pad using min(DATE)
- Display Booster Version names which had success in drone ship landing and payload mass in between 4000-6000
- List the total # of successful and failure mission outcomes using CASE expressions and conditional aggregations, summing up the occurrences of successful and failed mission outcomes.
- List the names of Booster Versions that have carried the maximum payload mass using nested SELECT statements.
- List the records including month, Failure Landing Outcomes in Drone Ship, Booster Versions, and Launch Sites for each month in the year 2015 using CASE statements and extracting dates using substr().
- Rank the count of landing outcomes i.e. Failure by Drone Ship, between the date **2010-06-04 and 2017-03-20**

[Github URL: Exploratory Data Analysis using SQL]()

FLOW CHART

Convert data set into a Pandas df ➜ Convert Pandas df into SQL Table ➜ Run SQL queries

Methodology-
Launch Sites Location Analysis with Folium

Create an interactive Folium map using Folium library with NASA Johnson Space Center at Houston, Texas as the starting location and center point of the map. Mark all the launch sites on the map, color-coding successes vs. Failures, and clustering launch sites close in proximity.

The success rate of launches can vary based on factors like payload mass, orbit type, and launch site location. Identifying an optimal launch site involves considering various factors, including locations and proximities of launch sites. Our objective is to analyze existing launch site locations to uncover some of these influential factors.

1) Create a Folium map object with NASA Johnson Space Center at Houston, Texas as the starting location and center point of the map.
2) Create a marker and circle object including corresponding site name around NASA Johnson Space Center.
3) Create markers and circle objects including their launch site names around each launch site on the map.
4) Color-code the success and failed launches for each site on the map by marking successes as green and failures red based on the 'Class' value.
5) Use a MarkerCluster() object to cluster the launch sites within close proximities.
6) Calculate the distances between a launch site to its proximities, such as its nearest coastline point or nearest airport.
7) Create a Polyline object to display a line that represents the distance between a launch site and its nearest landmarks.

Github URL: Launch Sites Location Analysis with Folium

FLOW CHART

Create a Folium Map → March all launch sites on a map → Color-code success and failed launches per site →
Cluster launch sites within close proximities → Calculate distances and draw distance lines

Methodology- Launch Sites Location Analysis with Plotly Dashboard

Build a Plotly Dashboard application for users to perform interactive visual analytics on SpaceX launch data in real-time.

The dashboard features a pie chart that illustrates the total successful launches categorized by site. Users have the option to choose a particular launch location from the dropdown menu, enabling them to observe the percentage of success vs failure for each site separately. Alternatively, users can select 'All sites' to visualize the overall percentage of successes across all sites.

Additionally, the dashboard contains a scatter chart depicting payload versus launch outcome, where the launch outcomes are indicated by the class attribute (0 = failure, 1 = success), covering all sites across various payload ranges. A sliding payload range feature above the scatter plot enables users to select specific payload ranges of interest, facilitating zooming in and focusing on particular payload segments.

1) Create a Dash application
2) Create an app layout / design: structure, and arrangement of different components within the dashboard. Use HTML and CSS to structure the dashboard's appearance. Then, incorporate Plotly graphs (pie chart and scatter chart with interactive elements such as a drop down menu select launch sites in the pie chart and a payload range slider to zoom in on the scatter chart).
   - TASK 1: Add a Launch Site Drop-down Input Component
   - TASK 2: Add a callback function to render success-pie-chart based on selected site dropdown
   - TASK 3: Add a Range Slider to Select Payload
   - TASK 4: Add a callback function to render the success-payload-scatter-chart scatter plot

GIthub URL: SpaceX Dash App Python Script

FLOW CHART

Create Dash App ➜ Create app layout ➜ Add Launch Site Drop-Down Input Component ➜

Add Callback Function for Pie Chart ➜ Add Range Slider for Scatter Chart ➜ Add Callback Function for Scatter Chart

Methodology- First Stage Landing Prediction using Machine Learning; Finding the Best Classification Model

Perform exploratory data analysis and determine training labels using Pandas, Numpy, and Matplotlib. Create a column for the class. Standardize the data and split into training data and test data using Sklearn's preprocessing and train_test_split libraries. Use GridSearchCV to test various hyperparameters of classification algorithms on test data and find the best performing ones. Classification algorithms tested are Logistic Regression, Support Vector Machine (SVM), Decision Trees and K-Nearest Neighbors.

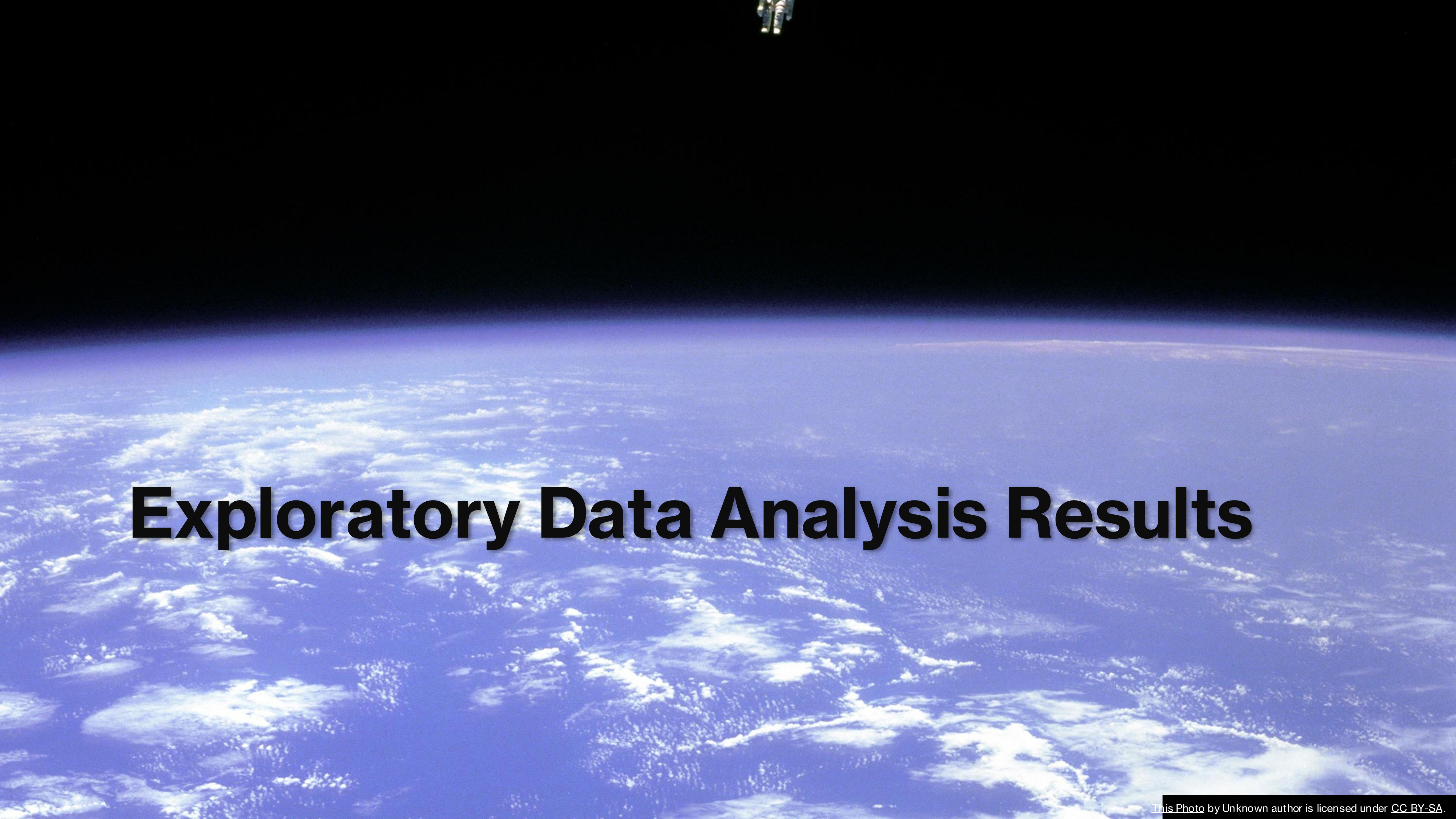
1) Create a Numpy array from the 'Class' attribute to use as test data.
2) Standardize the data using preprocessing.StandardScalar() method.
3) Fit the data using fit_transform(data) method.
4) Split the data into training and test sets using train_test_split function. The training data is divided into validation data and training data.
5) For each type of model:
   - Create a classification algorithm object
   - Create a GridSearchCV object with 10-fold cross validation that takes in the classification algorithm object and hyperparameters for hyperparameter turning.
   - Perform accuracy measures: best_params_ to display the best parameters, best_score_ to display the accuracy on the validation data, score() to display the accuracy on the test data, and a confusion matrix to create a visual summary of the predictions made by the model on the test data compared to the actual labels.
6) Create a dataframe that displays all of the models' accuracy measures and make comparisons to find the best performing model, including the best combination of classification algorithm and hyperparameters.

Github URL: First Stage Landing Prediction using Machine Learning; Finding the Best Classification Model

FLOW CHART

Create Numpy array as test data → Standardixe & fit data → Split data into training and test data → Create a classification algorithm object → Creat a GridSearchCV → Perform accuracy measures → Find the best model!

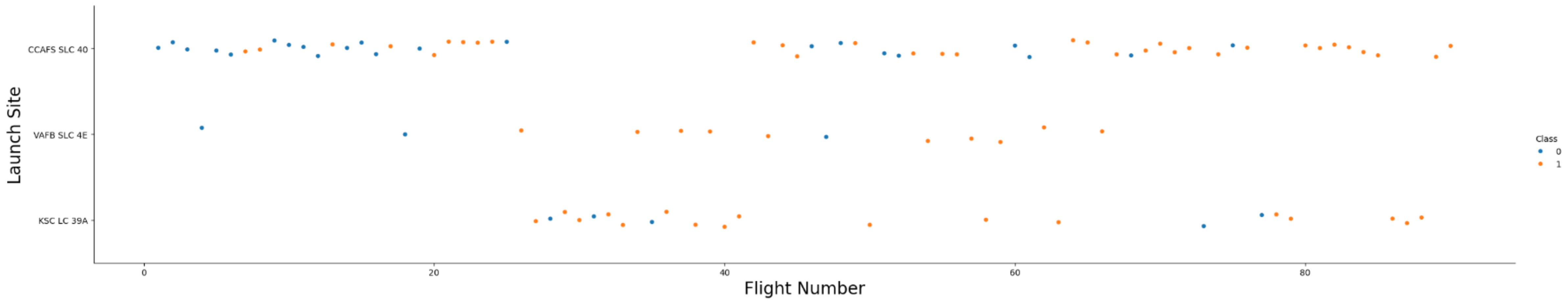
Results



Exploratory Data Analysis Results

Exploratory Data Analysis Results

Flight Number vs. Launch Site

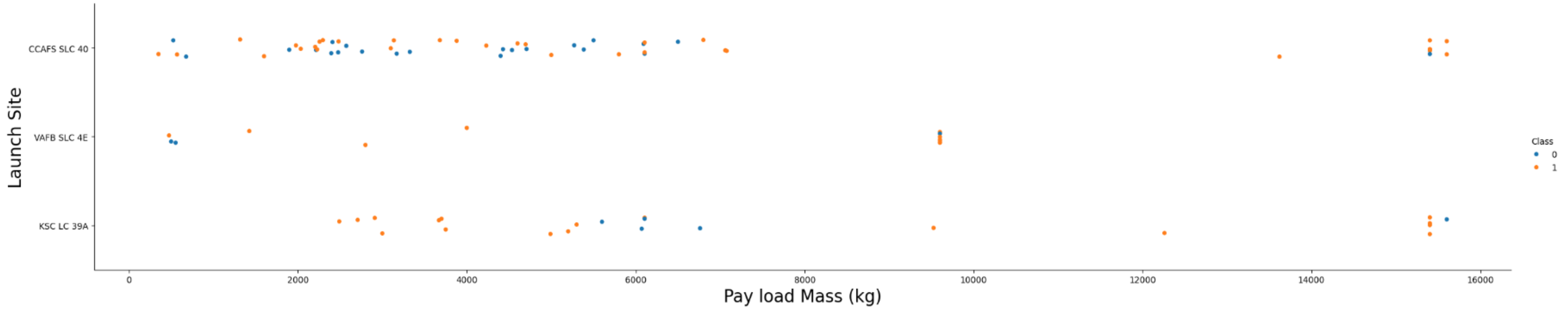


The blue labels represent failed launches and the orange labels represent successful launches.

It appears that CCAFS SLC 40 launch site had mostly unsuccessful launches during the first 20 flights, but the number of successful launches improved tremendously in the later flights (flights 40-80). The VAFB SLC 4E launch site had much fewer flights, yet also showed the same pattern of improvement in success as the flight numbers increased. The KSC LC 39A launch site started being used around the 25-30th flight, with a consistent majority of successful launches all throughout.

Exploratory Data Analysis Results

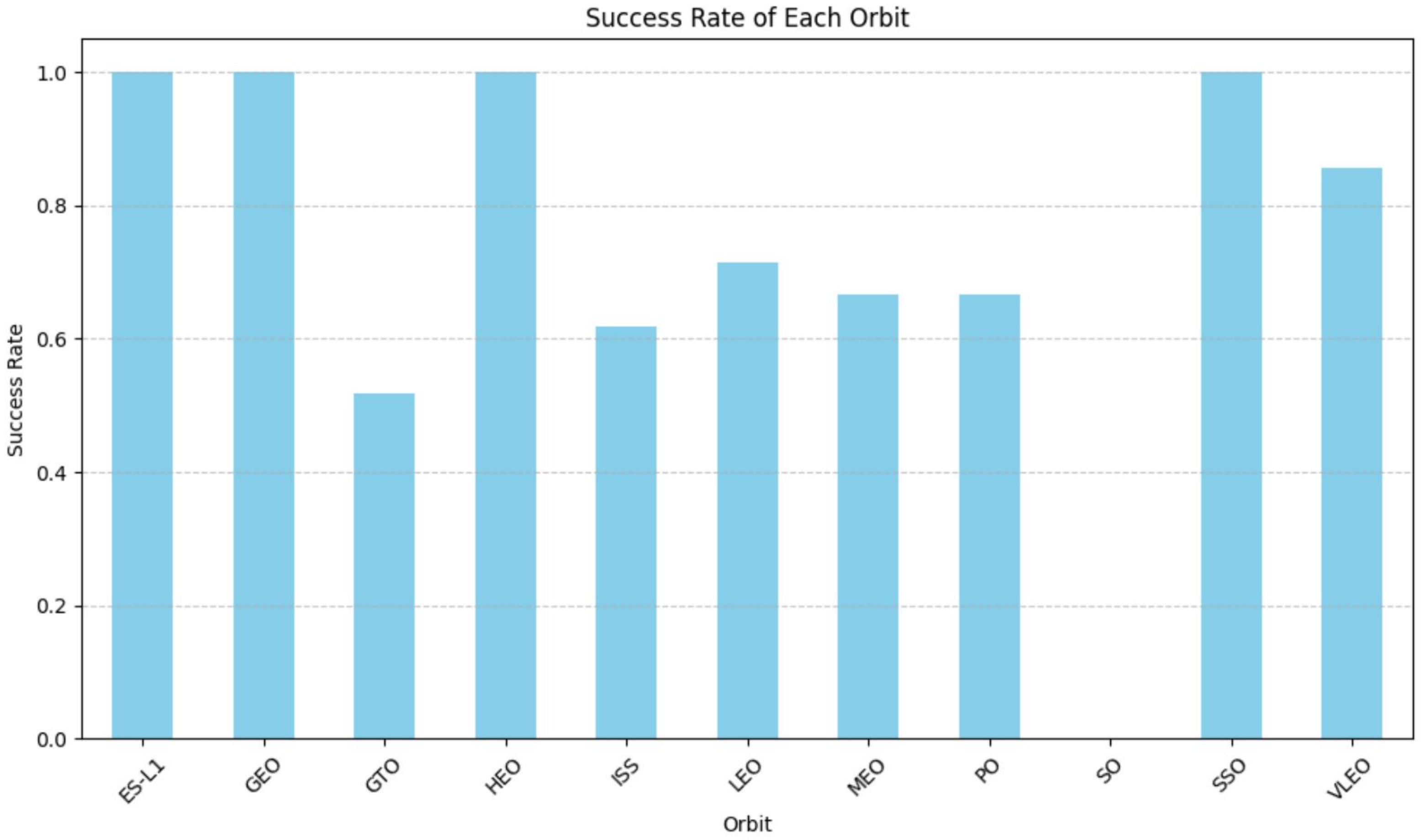
Payload Mass (kg) vs. Launch Site

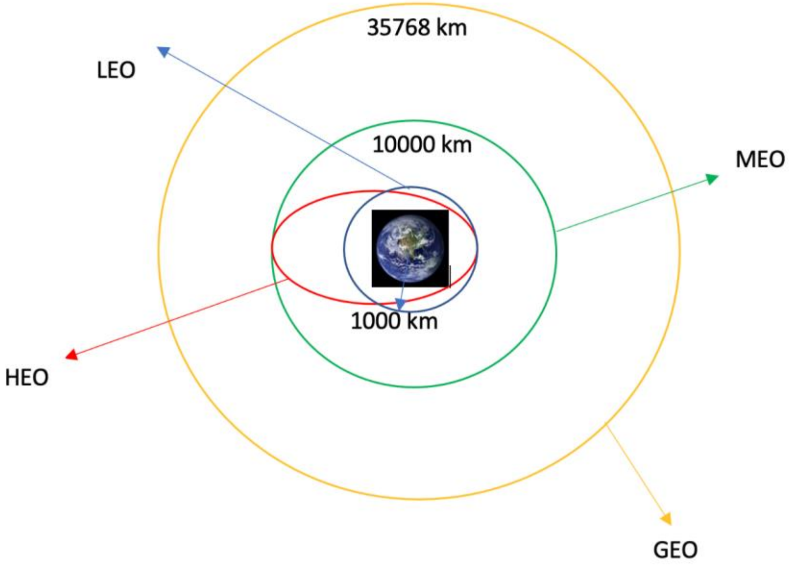


The blue labels represent failed launches and the orange labels represent successful launches.

It appears that for the VAFB-SLC 4E launch site, there are no rockets launched for heavy payload mass (greater than 10,000 kg). Although there does not seem to be much correlation between the payload mass and the success of the launches in each site, this chart does show that most of the launches' payload masses were within the range of 0-8,000 kg.

Exploratory Data Analysis Results

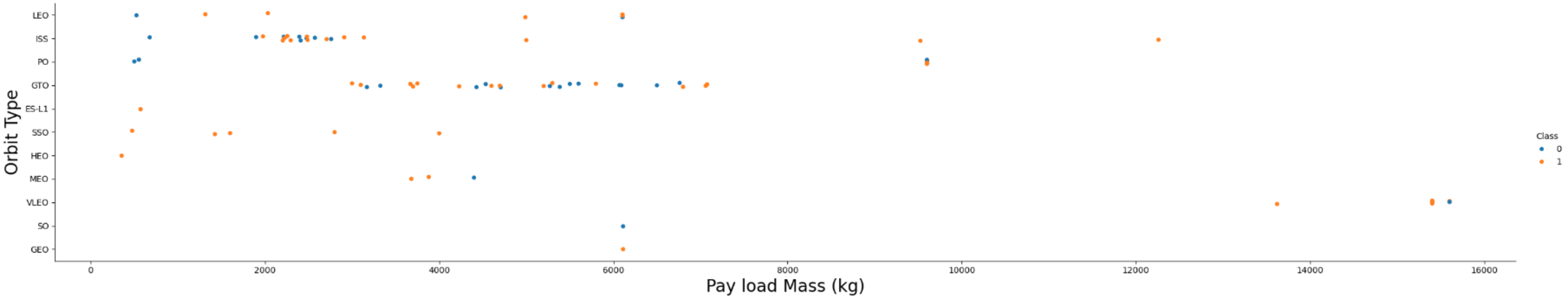




The ES-L1, GEO, HEO, and SSO orbits are shown to have the highest success rates, whereas SO has the lowest success rate at 0.

Exploratory Data Analysis Results

Payload Mass (kg) vs. Orbit Type



The blue labels represent failed launches and the orange labels represent successful launches.

It appears that launches tend to become more successful with heavier payload mass for Polar, LEO, and ISS orbit types. However, for GTO we cannot distinguish this well as both successful and failure outcomes are intertwined all through the range of payload masses. For SSO, it appears that there are few launches in the lower range of payload masses between 0 and 4,000 kg, of which are all successful.

Notably, GTO and ISS have the most launches overall, whereas GEO, SO, and ES-L1 seem to only have 1 launch. This explains why SO showed a success rate of 0 in the previous bar chart showing the success rate of each orbit displayed SO at a success rate of 0. Since there was only 1 launch in SO and it failed, the entire success rate of SO is 0.

Exploratory Data Analysis Results

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

Listed above are the distinct Launch Site names.

Exploratory Data Analysis Results

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

Listed above are 5 records with launch sites names that start with "CCA". Note that these records all belong to the launch site CCAFS LC-40, but have various booster versions, orbits, customers, and landing outcomes.

Exploratory Data Analysis Results

Listed here are the total payload masses carried by boosters from NASA customers. The payload masses values range from 0 to 12,530 kg.

PAYLOAD_MASS__KG_	Customer
0	NASA (COTS) NRO
525	NASA (COTS)
500	NASA (CRS)
677	NASA (CRS)
2296	NASA (CRS)
2216	NASA (CRS)
2395	NASA (CRS)
570	U.S. Air Force NASA NOAA
1898	NASA (CRS)
1952	NASA (CRS)
553	NASA (LSP) NOAA CNES
3136	NASA (CRS)
2257	NASA (CRS)
2490	NASA (CRS)
2708	NASA (CRS)
3310	NASA (CRS)
2205	NASA (CRS)
2647	NASA (CRS)
362	NASA (LSP)
6460	Iridium Communications GFZ , NASA
2697	NASA (CRS)
2500	NASA (CRS)
12055	NASA (CCD)
2495	NASA (CRS)
2268	NASA (CRS)
2617	NASA (CRS), Kacific 1
12050	NASA (CTS)
1977	NASA (CRS)
12530	NASA (CCDev)
12500	NASA (CCP)
1192	NASA / NOAA / ESA / EUMETSAT
2972	NASA (CRS)

Exploratory Data Analysis Results

Average Payload Mass carried by Booster Version F9 v1.1
2928.4

Listed above is the average payload mass of rocket launches carried by Booster Version F9 v1.1 with varying launch times and dates, launch sites, customers, and orbits.

Exploratory Data Analysis Results

First Successful Ground Landing Date:

December 22, 2015

SpaceX rocket landings occurred in regions of the ocean, on drone ships, and land. The first successful landing outcome on the ground (land) occurred on December 22nd, 2015.

Exploratory Data Analysis Results

Successful Drone Ship Landing with Payload Masses between 4,000-6,000 kg

Listed here are the Booster Versions of rocket launches that successfully landed on drone ships and had payload masses within the range of 4,000 and 6,000 kg.

Booster_Version	PAYLOAD_MASS__KG_
F9 v1.1	4535
F9 v1.1 B1011	4428
F9 v1.1 B1014	4159
F9 v1.1 B1016	4707
F9 FT B1020	5271
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1030	5600
F9 FT B1021.2	5300
F9 FT B1032.1	5300
F9 B4 B1040.1	4990
F9 FT B1031.2	5200
F9 B4 B1043.1	5000
F9 FT B1032.2	4230
F9 B4 B1040.2	5384
F9 B5 B1046.2	5800
F9 B5 B1047.2	5300
F9 B5 B1046.3	4000
F9 B5B1054	4400
F9 B5 B1048.3	4850
F9 B5 B1051.2	4200
F9 B5B1060.1	4311
F9 B5 B1058.2	5500
F9 B5B1062.1	4311

Exploratory Data Analysis Results

Total Succeses	Total Failures
100	1

The total number of successful SpaceX mission outcomes was 100 and the total number of failed mission outcomes was 1.

The **mission outcome** concerns the primary objectives of the launch (e.g., payload delivery), while the **landing outcome** pertains to the recovery and potential reuse of the rocket's first stage. Both outcomes are important for SpaceX's goals, with the mission outcome being crucial for the immediate success of the customer's needs, and the landing outcome being critical for SpaceX's long-term cost savings and sustainability through rocket reusability.

Exploratory Data Analysis Results

Listed here are the Booster Version names of the rocket launches that carried that maximum payload mass.

The **booster** is the first stage of the rocket, which provides the initial thrust to propel the rocket off the ground and into space. It plays a crucial role in determining the payload capacity because it must generate sufficient thrust to lift the total mass of the rocket and its payload.

The **maximum payload** is the heaviest cargo (satellites, scientific instruments, cargo for the ISS, etc.) that the rocket can carry to its intended destination. This includes considerations of weight, volume, and the specific requirements of the mission.

The payload mass is a huge factor in determining and improving performance, cost-efficiency of the launches, and optimizing missions.

Boosters Carried Max 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

Exploratory Data Analysis Results

Failed Landing Outcomes in
Drone Ships in 2015

Month	Failure_Landing_Outcomes	Booster_Versions	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Listed here are the months, launch sites, and booster versions of the rocket launches that had failed landing outcomes in drone ships in 2015.

It appears that both of these rockets had the same booster version, v1.1 and launched from the same launch site, CCAFS LC-40.

Exploratory Data Analysis Results

Landing Outcomes between
June 4th, 2010 and March 20th, 2017

Landing_Outcome	Outcome_Count
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

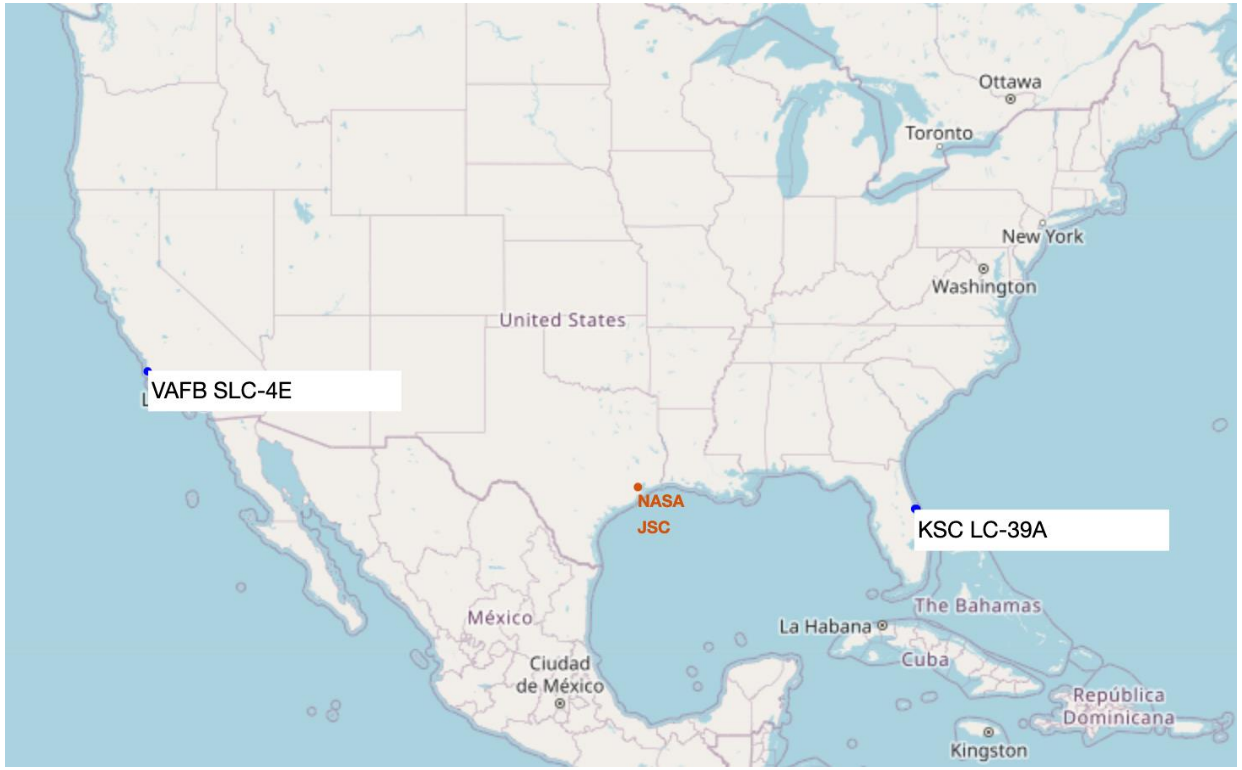
Listed here are the counts of landing outcomes bewteen June 4th, 2010 and March 29th, 2017 in descending order.

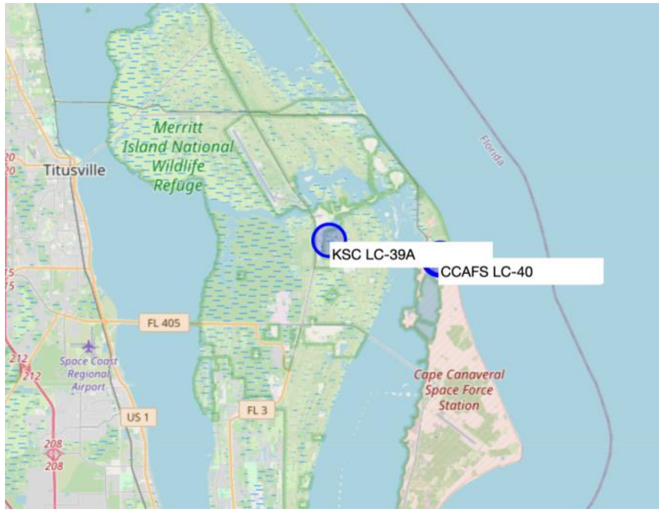
It appears the highest count of outcomes was for rocket launches with no attempt of landing outcomes and the lowest count was for precluded drone ships.

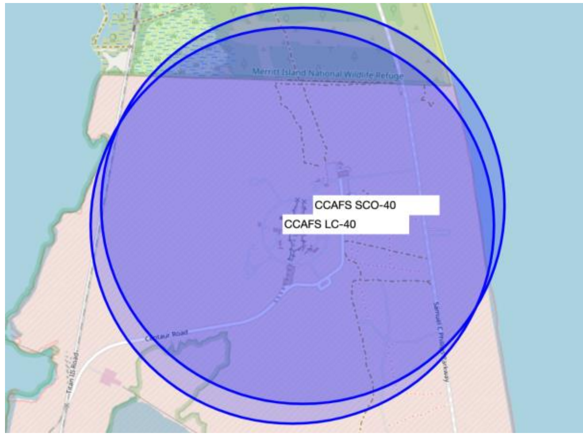
Launch Sites Proximities Analysis Results

Launch Sites Proximities Analysis

Folium Map with Launch Sites markers







Shown above is a map of the United States with the launch sites marked in a blue dot and the name inside a white label.

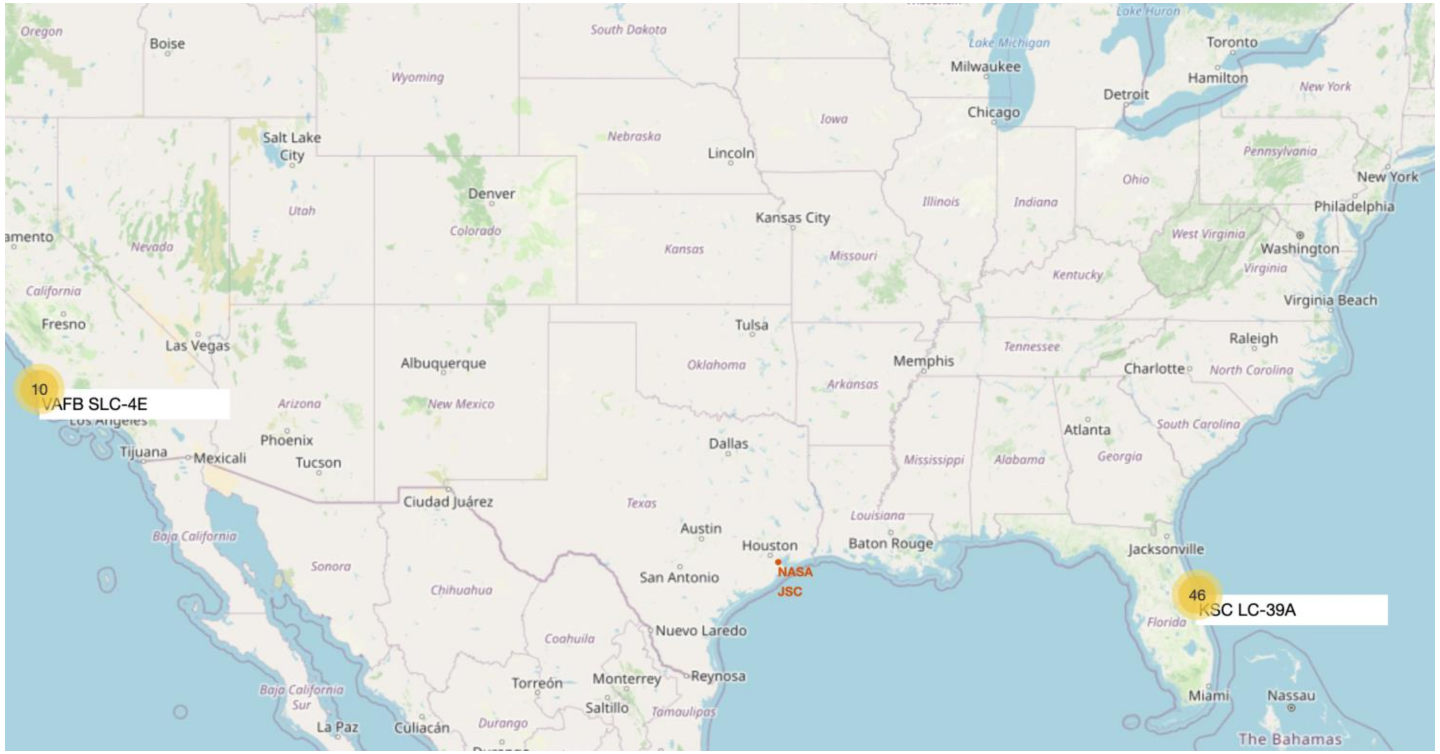
CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A are all grouped together in the East Coast of Florida, whereas VAFB SLC-4E is located on the opposite West coast in California. Therefore, you can see each of the East Coast launch sites when you zoom in, as shown in the two smaller images to the right of the Folium map.

All the launch sites are in close proximity to the equator line and they are in very close proximity to their respective coasts.

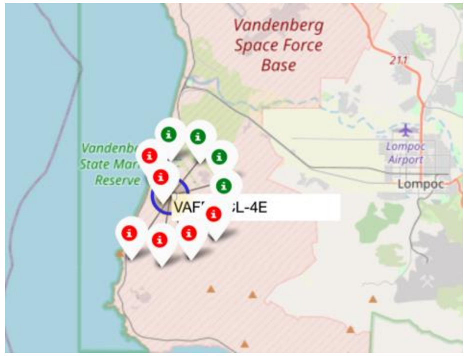
Launch Sites Proximities Analysis

Red = Failed
Green = Successful

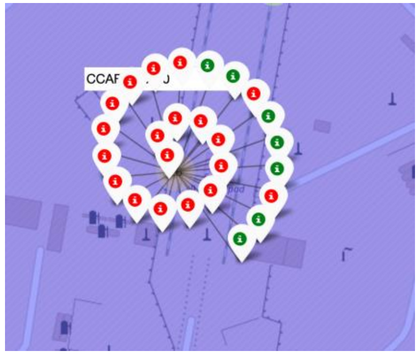
Launch Site Clusters



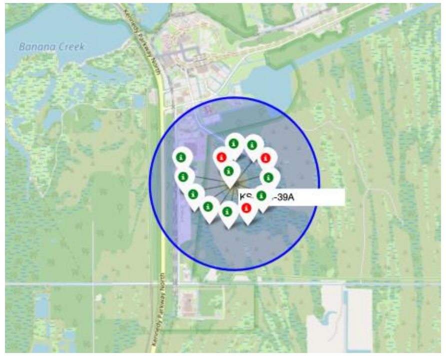
VAFB SLC-4E



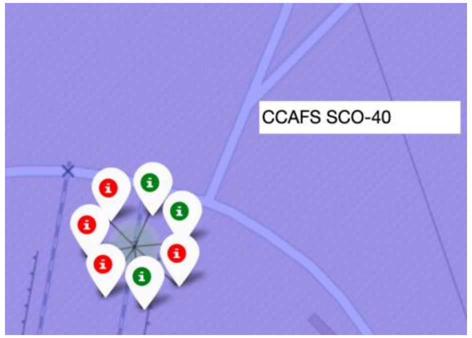
CCAFS LC-40



KSC LC-39A



CCAFS SCO-40



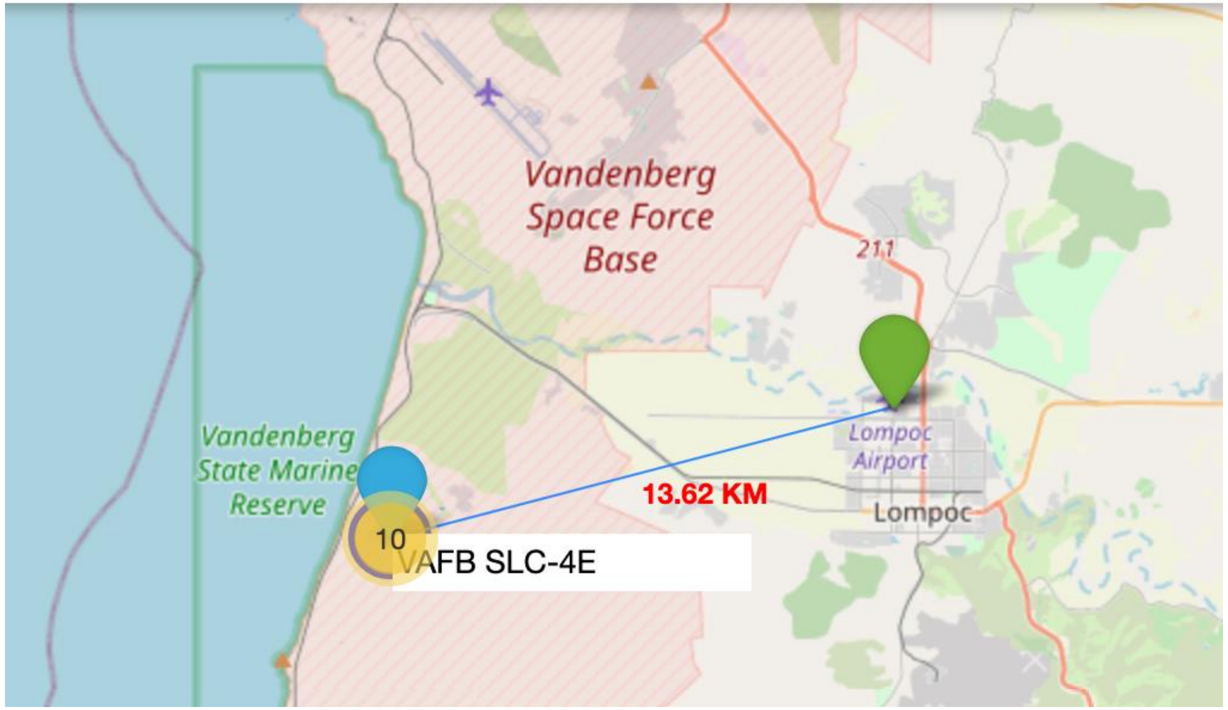
The images above display clusters of launches for each site, with color-coding to indicate successful – marked by green, and failed – marked by red, launches.

Recall that VAFB SLC –4E is the only launch site on the West Coast, whereas CCAFS LC-40, CCAFS SCO-40, and KSC LC-39A are all in close proximities on the East Coast, with CCAFS LC-40 and CCAFS SCO-40 overlapping.

The color-coded launches display the proportion of successful launches per site.

Launch Sites Proximities Analysis

Distance between VAFB SLC –4E Launch Site and Lompoc airport



Displayed above is a distance line between the VAFB SLC –4E launch site and its closest airport, Lompoc Airport. The distance between the two is displayed on the line: 13.62 km.

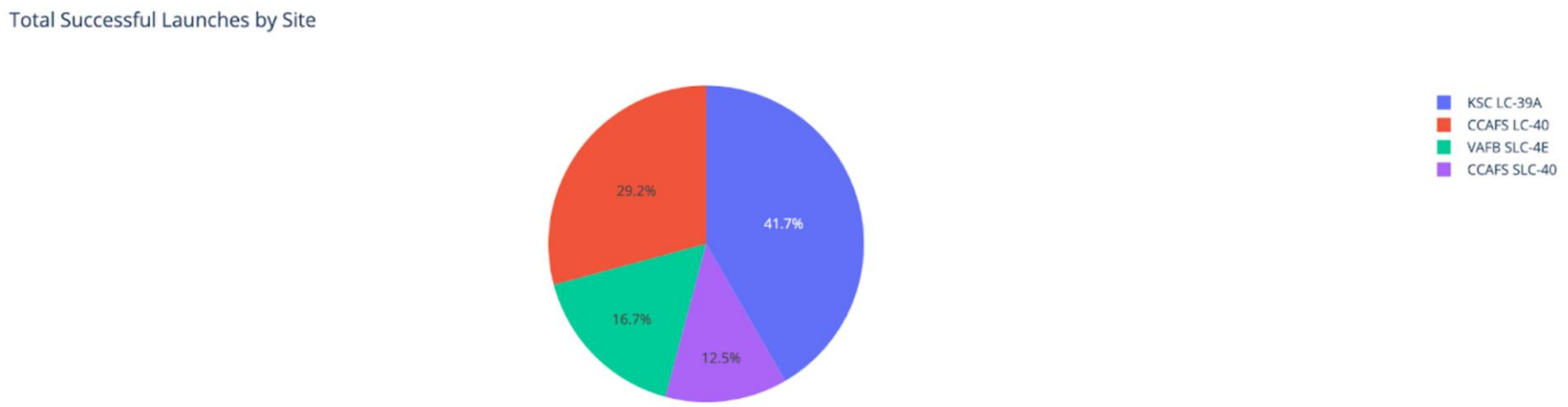
Distance between CCAFS SCO-40 Launch Site and closest coastline

Displayed above is a distance line between the CCAFS SCO –40 launch site and its closest coastline point (with the corresponding Latitude and Longitude of the coastline point displayed on the top right corner). The distance between the two is displayed on the line: 0.86 km.

Build a Dashboard with Plotly Dash Results

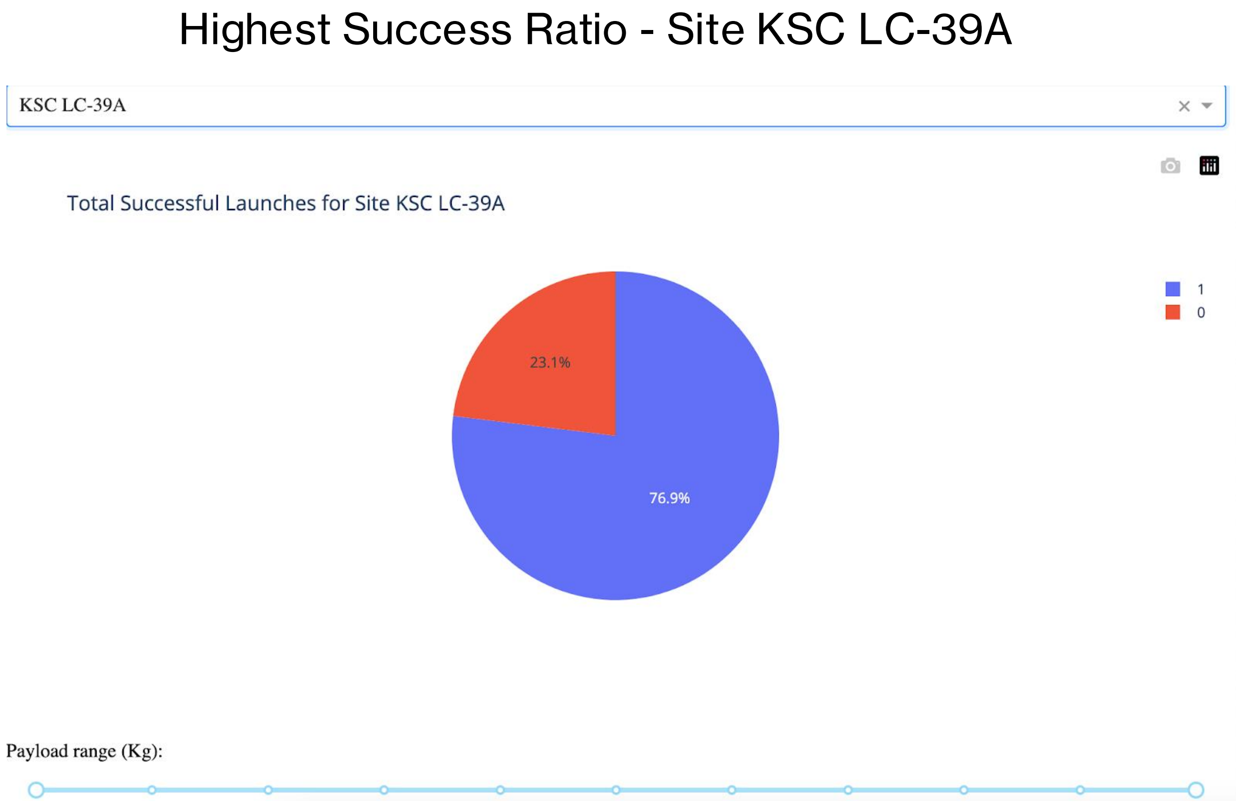
Build a Dashboard with Plotly Dash

Total Successful Launches by Site (for all sites)



This pie chart displays the distribution of successful launches across all launch sites. On the right hand side, there is a key that shows which colors represent which launch site. It appears that KSC LC –39A has the highest proportion of successes from all the sites, and CCAFS SCL –40 has the lowest proportion of successes.

Build a Dashboard with Plotly Dash

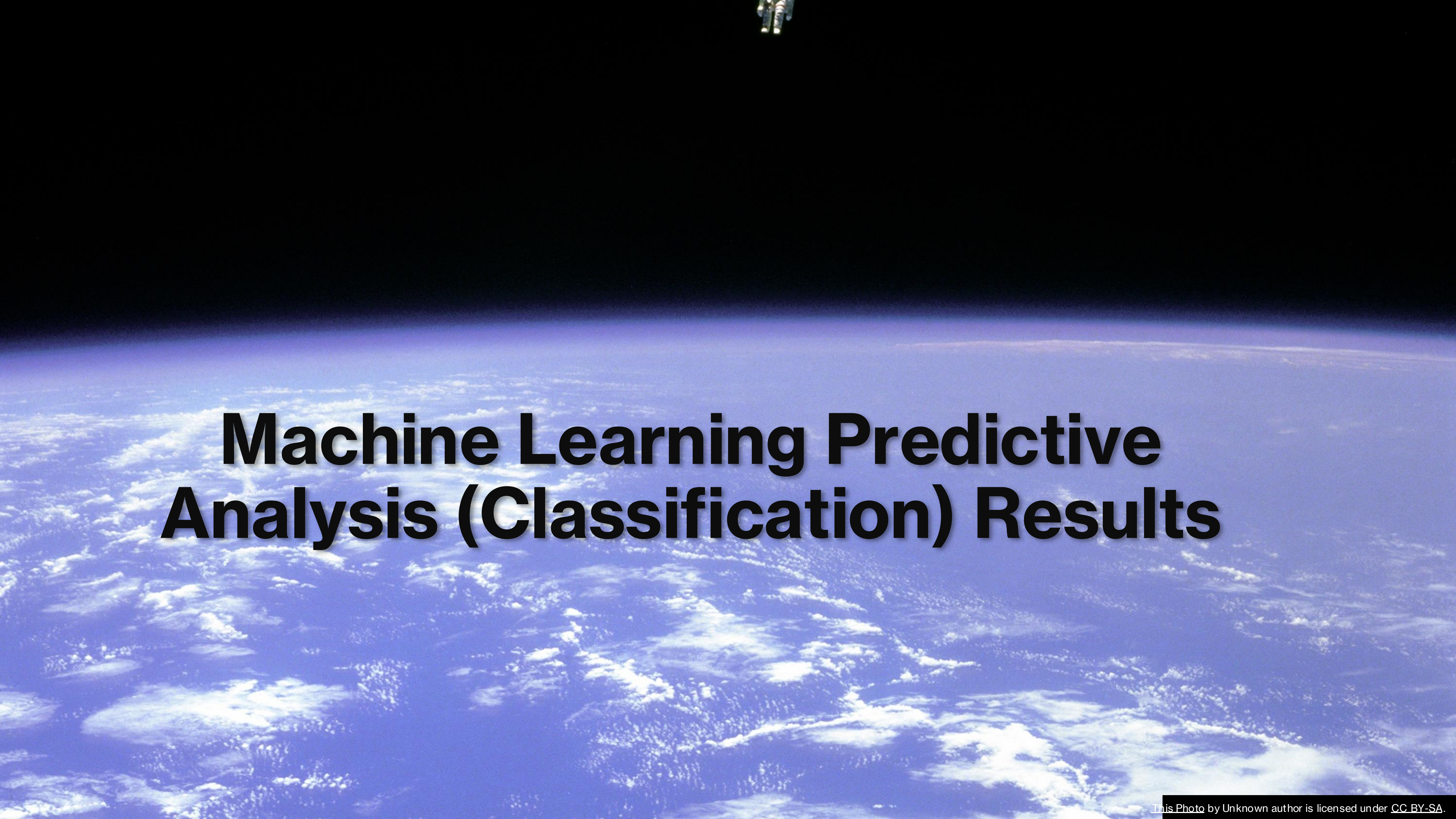


This pie chart displays the proportion of successful and failed launches for Site KSC LC-39A, which is the site that has the highest launch success ratio.

Build a Dashboard with Plotly Dash

Scatter Chart for Entire Payload Range

This scatter chart displays Payload vs. Launch Outcome for all sites across different payload ranges. It allows us to zoom in and focus on specific interesting ranges by using the range slider on the top. In the image above, it is set for the entire payload range from 0-10k. The payload range with the highest number of successful launches is 2,000-3,750 kg, in which there were 12 successful launches. The booster version with the most successful launches is FT.



Machine Learning Predictive Analysis (Classification) Results

Build a Dashboard with Plotly Dash

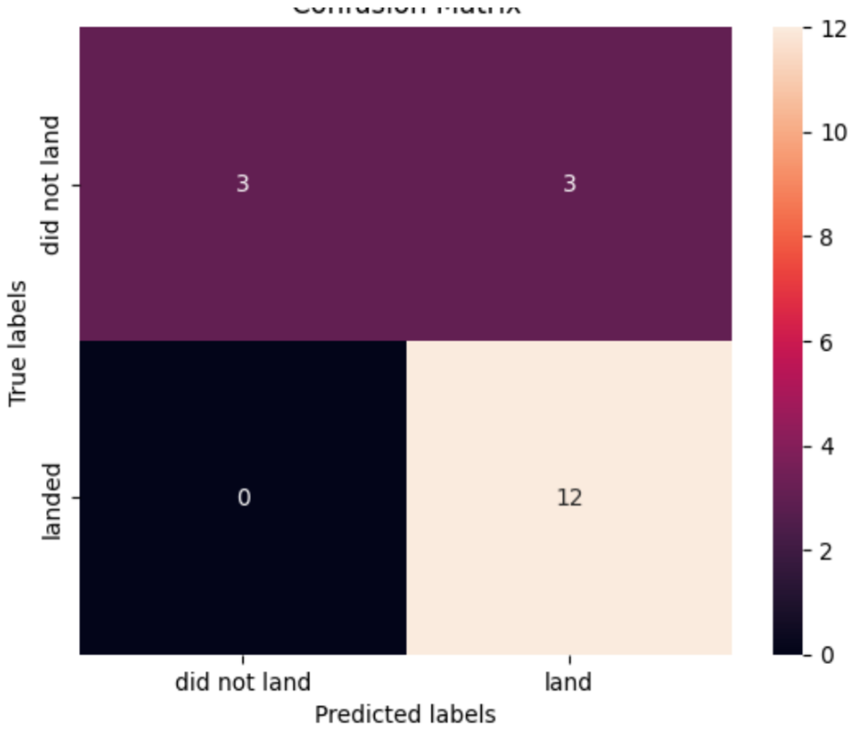
Classification Accuracy

	Model	Accuracy Score	Best Score	Best Parameters
0	Logistic Regression	0.833333	0.833333	{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
1	SVM	0.833333	0.833333	{'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
2	Decisiono Tree	0.666667	0.666667	{'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
3	KNN	0.833333	0.833333	{'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}

After training different models and hyperparameters on the data and performing accuracy measures, we find that the most of the models have the same accuracy scores. This is likely due to the fact that the dataset is small and has lesseer values. However, the data frame does show the best parameters for each model.

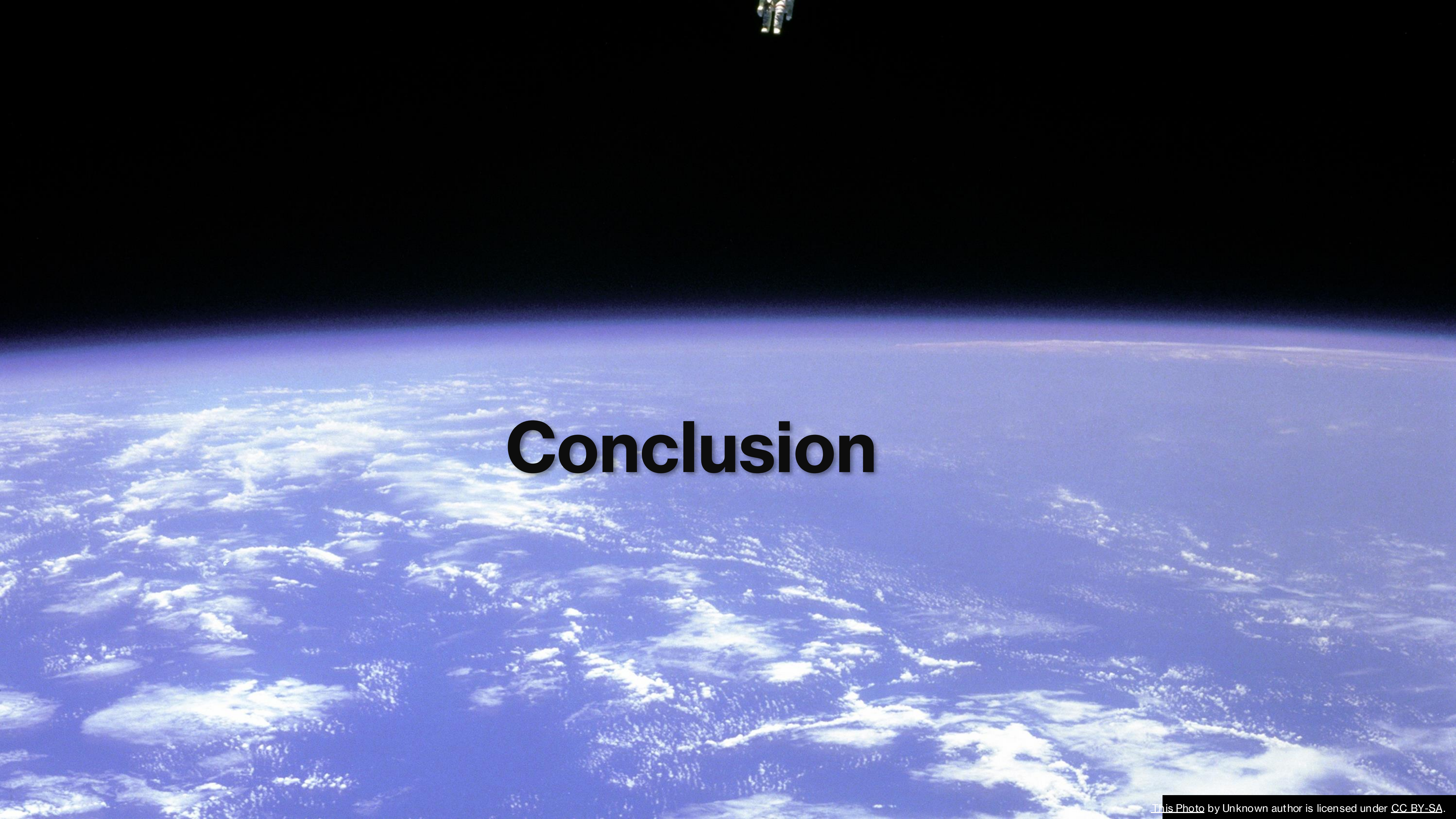
Build a Dashboard with Plotly Dash

Confusion Matrix



The Confusion Matrix shows in the top left square that the predicted labels of 3 launches not landing was indeed true and in the bottom right square that the predicted labels of 12 launches successfully landing was indeed true. However, in the top right square, the predicted labels of 3 launches landing was not true. The last square on the bottom left shows that the predicted labels of 0 launches not landing was true. This shows a strong accuracy measure of the model, with a slight mistake represented by the top right square in which the predicted labels don't match the true labels.

Conclusion



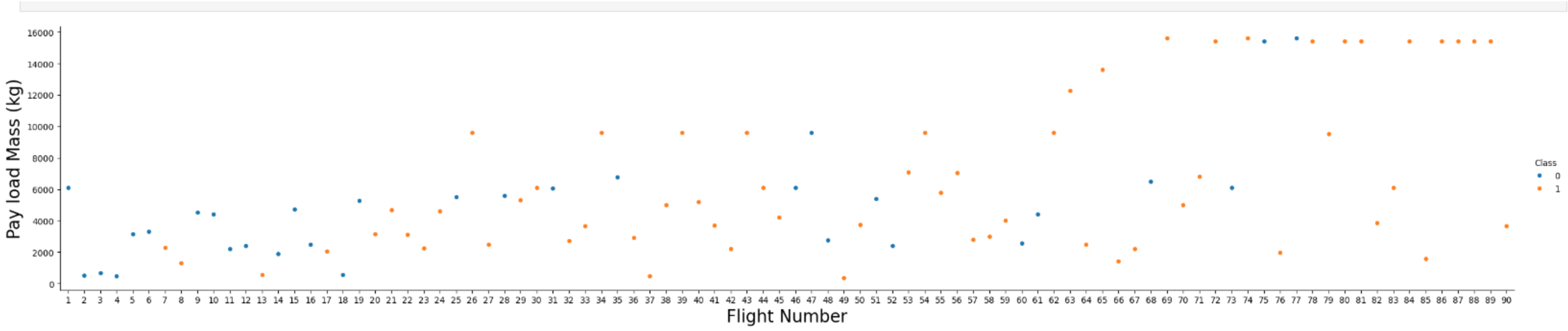
Conclusion

Our project aimed to determine the cost of a Falcon 9 rocket launch by assessing the likelihood of the first stage landing successfully. Given that SpaceX can reuse the first stage, knowing this would help us accurately calculate the launch cost, which is $62 million.

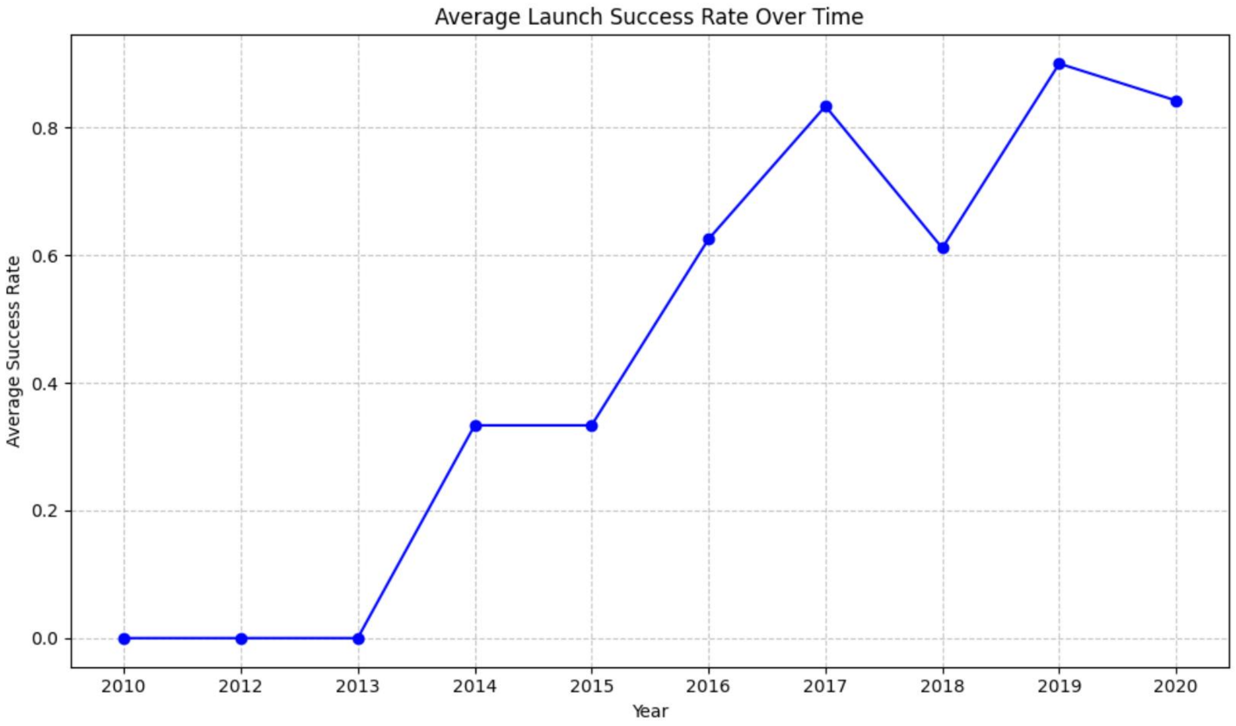
Our findings revealed that launch success rates improve with number of flights at many sites. Specifically, ES-L1, GEO, HEO, and SSO orbits achieved perfect success rates. We also identified that certain orbit types, when matched with specific payload ranges, significantly enhance launch success. Out of 101 missions, SpaceX achieved 100 successful outcomes, highlighting their commitment to mission objectives, such as payload delivery. The most successful launch site was KSC LC-39A, boasting a 76.9% success rate, while the FT booster version had the highest success rate among boosters.

With these insights, we can now make informed decisions to ensure the successful landing of the Falcon 9's first stage. By selecting the optimal combination of launch site, booster version, orbit type, and payload range, we can maximize our success rates. Future analyses will focus on refining these combinations to achieve even higher success rates, ultimately enhancing our ability to predict and plan successful launches.

Appendix



This graph shows that the proportion of failed launches (blue dots) to successful launches (orange dots) decreases as flight numbers increase and payload mass increases, suggesting that the success rate increases with number of flights and increasing payload masses.



This graph shows that the launch success rate has increased over time. Therefore, the analysis conducted and applied to refine the combinations of launch features is evidently productive and a good indicator of whether the first stage of Falcon 9 will launch successfully or not.