IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Layth Alsamarray
5/3/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Data Collection: Utilized SpaceX REST API and web scraping.
  - Data Wrangling: Categorized landing outcomes to create a 'Class' column for predictive modeling.
  - Exploratory Data Analysis (EDA): Conducted using visualization tools and SQL.
  - Interactive Visual Analytics: Employed Folium and Plotly Dash for dynamic data exploration.
  - Predictive Analysis: Built and tuned classification models (Logistic Regression, SVM, Decision Tree, K-Means) using GridSearchCV for optimal hyperparameters.
- Summary of all results
  - Exploratory Data Analysis:
    - Flight and Payload Analysis: CCAFS SLC 40 had the most flights; most payloads were under 7000 kg, with the highest at KSC LC 39A and CCAFS SLC 40.
    - Success Rates by Orbit: Highest success rates were in ES-L1, GEO, HEO, and SSO orbits, with GTO having the lowest.
    - Yearly Trends and Site Insights: Successful launches rose after 2013, with CCAFS SLC 40 being the most active; four distinct launch sites were identified.
  - SQL Query Findings:
    - Key Insights: The first successful ground landing occurred in 2015; successful drone ship landings had payloads between 4000-6000 kg; highest payloads reached MEO orbit.
  - Predictive Analysis:
    - Model Performance: Logistic Regression, SVM, and K-Means achieved 0.83 accuracy on test data; Decision Tree underperformed with 0.44.
    - Classification Challenges: Logistic Regression, SVM, and K-Means were strong in predicting positives but struggled with negatives, per confusion matrix results.

Introduction

- As the space industry burgeons, with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX revolutionizing space travel, a new era has dawned where space is increasingly accessible to all.
- SpaceX, stands out for its pioneering achievements, including satellite launches, manned missions, and the groundbreaking Starlink project. Central to SpaceX's success is its cost-effective approach, notably its ability to reuse rocket components, driving down launch costs significantly.
- However, the key question remains: Can we reliably predict the reusability of SpaceX's first stage? The answer holds immense implications for cost estimation and competitive strategy in the space industry.
- Our mission at Space Y is twofold: to ascertain the price of each launch and to predict the likelihood of SpaceX reusing the first stage.
- Rather than relying solely on traditional rocket science, we will leverage machine learning techniques and publicly available data to train predictive models. Through meticulous analysis and dashboard creation, we aim to empower Space Y with actionable insights to navigate the competitive landscape of space commerce.

Section 1

Methodology

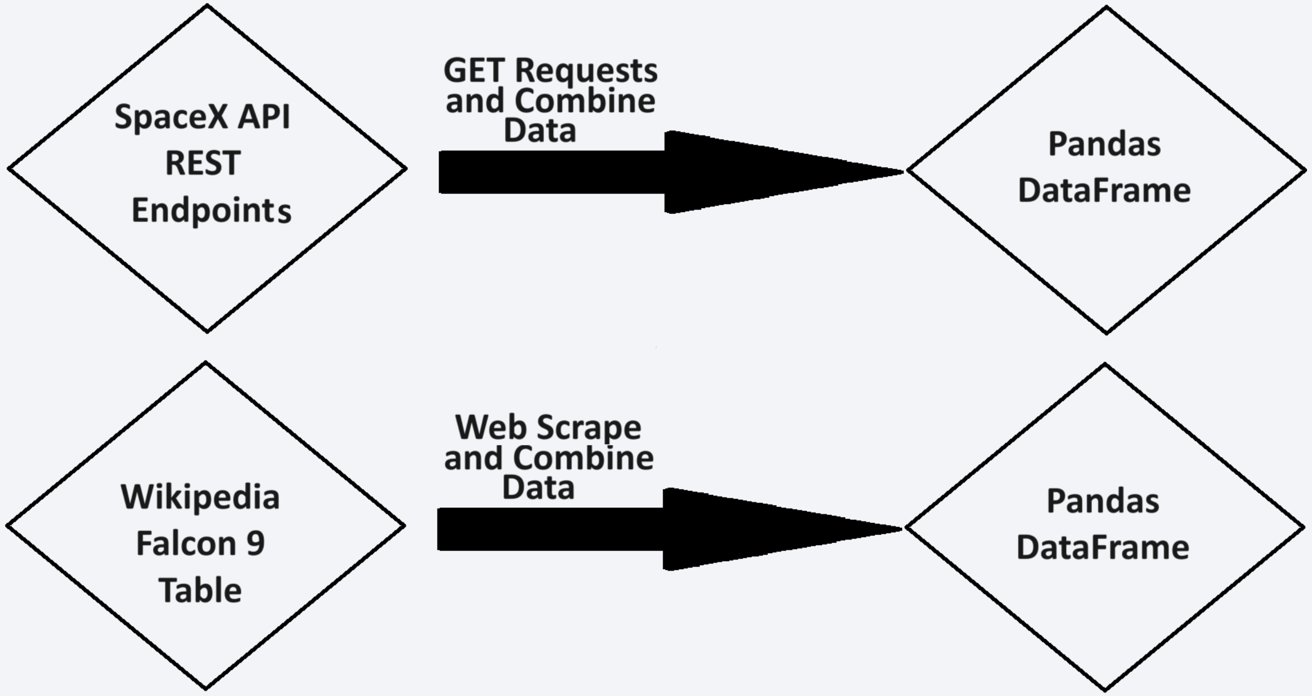
Methodology

Executive Summary

- Data collection methodology:
  - Data collection involved SpaceX REST API Request calls and web scraping.
- Perform data wrangling
  - We categorized SpaceX landing outcomes by checking against predefined undesired outcomes and creating a 'Class' column for predictive modeling.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Tuned models like Logistic Regression, SVM, Decision Tree, and K-Means used GridSearchCV for optimal hyperparameters.

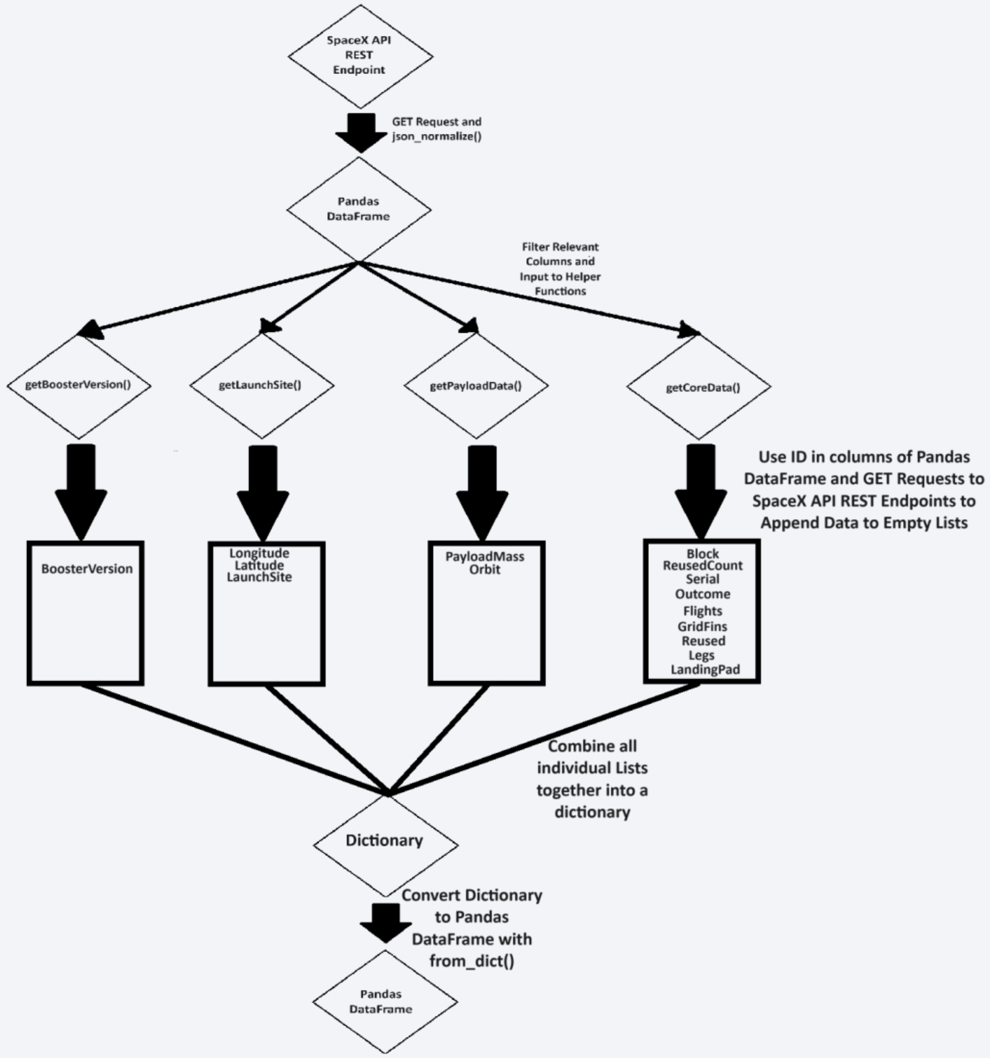
Data Collection

- One data set was collected first from extracted information from GET Requests to public SpaceX REST ENDpoints at api.spacexdata.com/v4.
- Another data set collected through web scraping extracted information with the Beautiful Soup Library from a Wikipedia article detailing Falcon 9 launches.



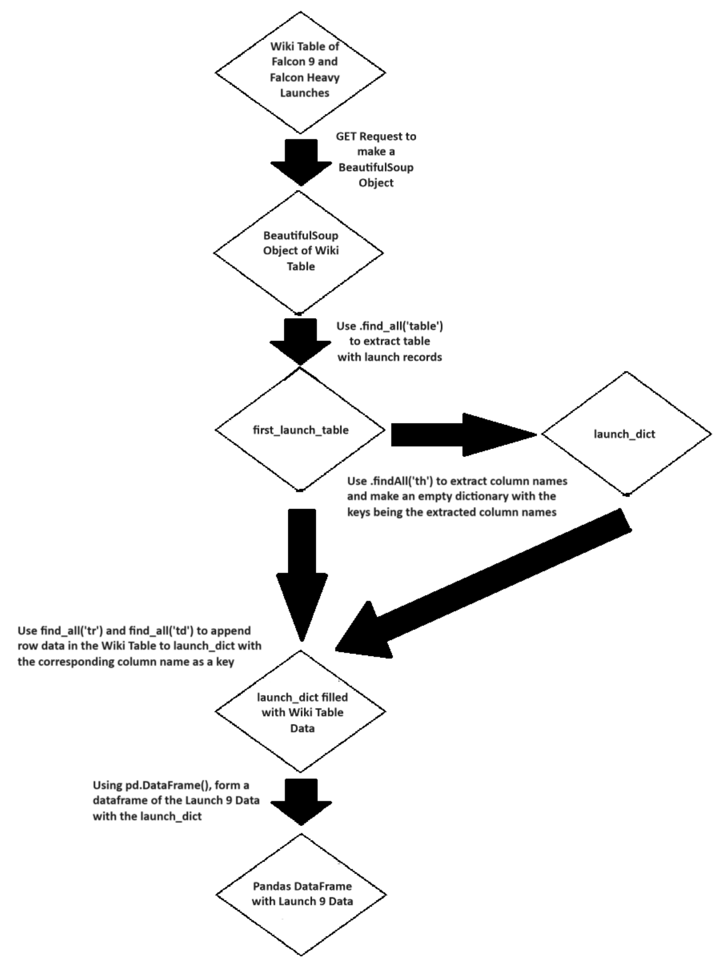
Data Collection – SpaceX API

- **Objective:** Predict if the Falcon 9 first stage will land successfully, impacting launch costs significantly.
- **Steps:**
  1. **Request Data from SpaceX API:**
     1. Extract comprehensive launch data including rocket, payloads, launchpad, and cores information.
     2. Parse JSON responses to gather detailed information on each aspect of the launch.
  2. **Data Cleaning and Preparation:**
     1. Filter and clean the dataset to remove irrelevant or incomplete data.
     2. Convert IDs to meaningful names using additional API calls.
     3. Handle missing values, ensuring a complete dataset for analysis.
  3. **Create and Format DataFrame:**
     1. Combine extracted information into a structured DataFrame.
     2. Filter to include only Falcon 9 launches, ensuring relevance.
     3. Handle missing values by replacing them with the mean where appropriate.
- **Outcome:** A clean, comprehensive dataset ready for predictive analysis on Falcon 9 first stage landings.
- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- **Objective:** Extract Falcon 9 launch records from the Wikipedia page "List of Falcon 9 and Falcon Heavy launches" and convert the data into a Pandas DataFrame.
  - **Process Overview:**
    - **Request the Wiki Page:** Use the HTTP GET method to retrieve the HTML content from the provided static URL.
    - **Create BeautifulSoup Object:** Convert the HTML content into a BeautifulSoup object to facilitate data extraction.
    - **Extract Column Names:**
      - Locate the HTML table containing the launch records.
      - Extract column names from the table header elements.
      - Extracted column names include: “Flight No.”, “Date and time (UTC)”, “Version, Booster”, “Launch site”, “Payload”, “Payload mass”, “Orbit”, “Customer”, “Launch outcome”
  - **Parsing Falcon 9 Launch Records from Wikipedia**
    - **Initialize Dictionary:** Create an empty dictionary with column names from the extracted table headers.
    - **Populate Dictionary:** Extract and append data from each table row to the corresponding dictionary keys.
    - **Data Cleaning:** Handle annotations, missing values, and inconsistent formatting.
    - **Create DataFrame:** Convert the populated dictionary into a Pandas DataFrame.
  - **Outcome:** A DataFrame with comprehensive Falcon 9 launch data, ready for analysis.
- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- To categorize the landing outcomes in the SpaceX dataset, we first compiled a list of undesirable outcomes ('False ASDS', 'False Ocean', 'False RTLS', and 'None None').
- Employing a for loop to iterate through the 'Outcome' column, we checked each value against this list. If a value matched any of the undesired outcomes, we appended '0' to a new list called 'landing_class'; otherwise, we appended '1'.
- Leveraging the 'landing_class' list, we created a new column named 'Class' in the SpaceX dataframe, populating it with the corresponding values from 'landing_class'.
- This data wrangling step lays a crucial foundation for subsequent predictive modeling aimed at determining the quality of SpaceX landing outcomes.
- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- **Flight Number vs. Launch Site Data Visualization**
  - Scatter plot showcasing the relationship between flight number and launch site.
- **Payload vs. Launch Site Data Visualization**
  - Scatter plot illustrating payload mass variations across launch sites.
- **Orbit Success Rates Data Visualization**
  - Bar chart depicting success rates for each orbit type.
- **Flight Number vs. Orbit Type Data Visualization**
  - Scatter plot revealing success outcomes based on flight number and orbit type.
- **Payload vs. Orbit Type Data Visualization**
  - Scatter plot indicating success rates concerning payload mass and orbit type.
- **Yearly Success Trend Data Visulaization**
  - Line chart demonstrating the trend of success rates over the years.
- Selected features for success prediction:
  - FlightNumber, PayloadMass, Orbit, LaunchSite, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial
- Encoded categorical columns using dummy variables.
- Converted numeric columns to float64 type.
- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/edadataviz.ipynb

EDA with SQL

- Loaded the dataset into a Db2 database.
- Executed SQL queries to inspect and manipulate the dataset.
- Conducted various SQL queries to exploratory data analysis, including:
    - Displaying unique launch sites.
    - Filtering records based on specific criteria (e.g., launch site, payload mass).
    - Calculating aggregate statistics (e.g., total payload mass, average payload mass).
    - Identifying key milestones (e.g., first successful landing on ground pad).
    - Extracting insights on mission outcomes and booster versions.
- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Add markers and circles to all launch sites on the map using coordinates from our SpaceX Dataframe. This will provide a clearer visual representation of where each launch site is located.
- Create a new column in the SpaceX dataframe to designate marker colors based on launch success or failure. These markers will be added to a MarkerCluster object to display each launch site on the map.
- To calculate distances from a launch site to nearby locations, integrate the MousePosition plugin on the map to retrieve coordinates by hovering.
    - Additionally, develop a function to compute the distance between any two coordinates.
- Determine closest coastline by marking the nearest coastline point and compute its distance from the launch site.
- Calculate distances to other points of interest through computing and displaying distances to key points like cities, railways, and highways.

- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Objective: Build an interactive Plotly Dash application for real-time visual analytics on SpaceX launch data.
- Components:
  - Launch Site Drop-down: Select from four launch sites to analyze launch success.
  - Success Pie Chart: Displays success counts based on selected launch site.
  - Payload Range Slider: Allows selection of payload range (0-10000 Kg) for visual analysis.
  - Payload Scatter Plot: Shows correlation between payload mass and launch outcome, color-labeled by booster version.
- Insights:
  - Largest Successful Launches: Determine which site has the highest count of successful launches.
  - Highest Launch Success Rate: Identify the site with the highest percentage of successful launches.
  - Highest Launch Success Rate by Payload Range: Analyze which payload ranges correlate with the highest and lowest success rates.
  - Booster Version Success: Explore which F9 booster version achieved the highest launch success rate.
- GitHub Links:
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Creating%20Dashboards%20with%20SpaceX%20Data.txt
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Total%20Success%20and%20Failed%20Launches%20at%20VAFB%20SLC-4E%20Pie%20Chart.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Total%20Success%20and%20Failed%20Launches%20at%20KSC%20LC-39A%20Pie%20Chart.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Total%20Success%20and%20Failed%20Launches%20at%20CCAFS%20SLC-40%20Pie%20Chart.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Total%20Success%20and%20Failed%20Launches%20at%20CCAFS%20LC-40%20Pie%20Chart.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Correlation%20between%20Payload%20and%20Success%20for%20VAFB%20SLC-4E%20Scatter%20Plot.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Correlation%20between%20Payload%20and%20Success%20for%20KSC%20LC-39A%20Scatter%20Plot.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Correlation%20between%20Payload%20and%20Success%20for%20CCAFS%20SLC-40%20Scatter%20Plot.png
  - https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/Correlation%20between%20Payload%20and%20Success%20for%20CCAFS%20LC-40%20Scatter%20Plot.png

Predictive Analysis (Classification)

- Task 1: Extracting Class Labels
  - You need to create a NumPy array from the Class column in the data dataframe, which contains the labels indicating whether the first stage landed successfully.
- Task 2: Standardizing the Data
  - Standardize the features in the X dataframe using StandardScaler from sklearn.preprocessing.
- Task 3: Splitting the Data into Training and Test Sets
  - Split the standardized data into training and test sets using an 80-20 split.
- Task 4: Using Multiple Classifiers with GridSearchCV
  - Create different classifiers and use GridSearchCV to find the best hyperparameters.
  - Calculate accuracy on the test data and plot the confusion matrices.
- Task 5: Model Evaluation and Selection
  - After fitting all models and evaluating them on the test set, compare the performance to determine which model performs best.
  - By running this code, you will have completed all tasks required for your machine learning pipeline. You can then determine which model performed best based on the test accuracies and confusion matrices.

- GitHub Link: https://github.com/LAA-tech/SpaceX-Rocket-Predictive-Model/blob/main/SpaceX_Machine%20Learning%20Prediction.ipynb

Results

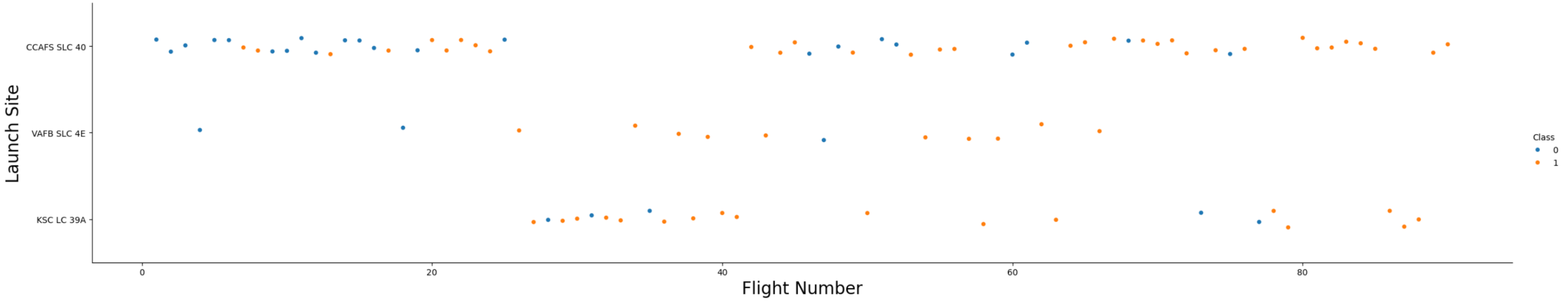
- Exploratory data analysis results
  - Flight and Payload Analysis: CCAFS SLC 40 had the highest number of flights, and most payloads were under 7000 kg, with the highest payloads at KSC LC 39A and CCAFS SLC 40.
  - Success Rates by Orbit: ES-L1, GEO, HEO, and SSO orbits had the highest success rates, while GTO had the lowest.
  - Yearly Trends and Site Insights: Successful launches increased significantly after 2013, with CCAFS SLC 40 showing the highest activity. Four distinct launch sites were identified.
  - SQL Query Findings: Key insights include the first successful ground landing in 2015, successful drone ship landings with payloads between 4000-6000 kg, and the highest payloads reaching MEO orbit.
- Predictive analysis results
  - Model Performance: Logistic Regression, SVM, and K-Means models achieved similar accuracies of 0.83 on test data, while the Decision Tree model underperformed with an accuracy of 0.44.
  - Classification Challenges: Logistic Regression, SVM, and K-Means models demonstrated strong performance in predicting positive outcomes but struggled with negative predictions, as shown by the confusion matrix results.

Section 2

Insights drawn from EDA

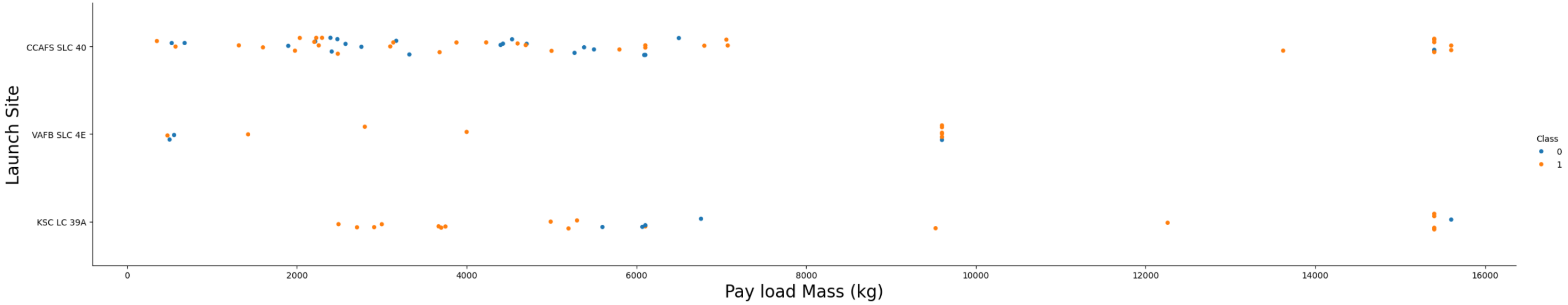
Flight Number vs. Launch Site

- This scatter plot visualizes the relationship between the Flight Number and Launch Site for SpaceX missions.
- Each point represents a launch, with the x-axis indicating the Flight Number and the y-axis representing the Launch Site.
- The color of each point corresponds to the outcome of the launch (success or failure), providing insight into the performance of each launch site over time.
- Analyzing the scatter plot reveals that CCAFS SLC 40 had the highest number of flights. Additionally, it's evident that as the number of flights increases, so does the number of successful flights.



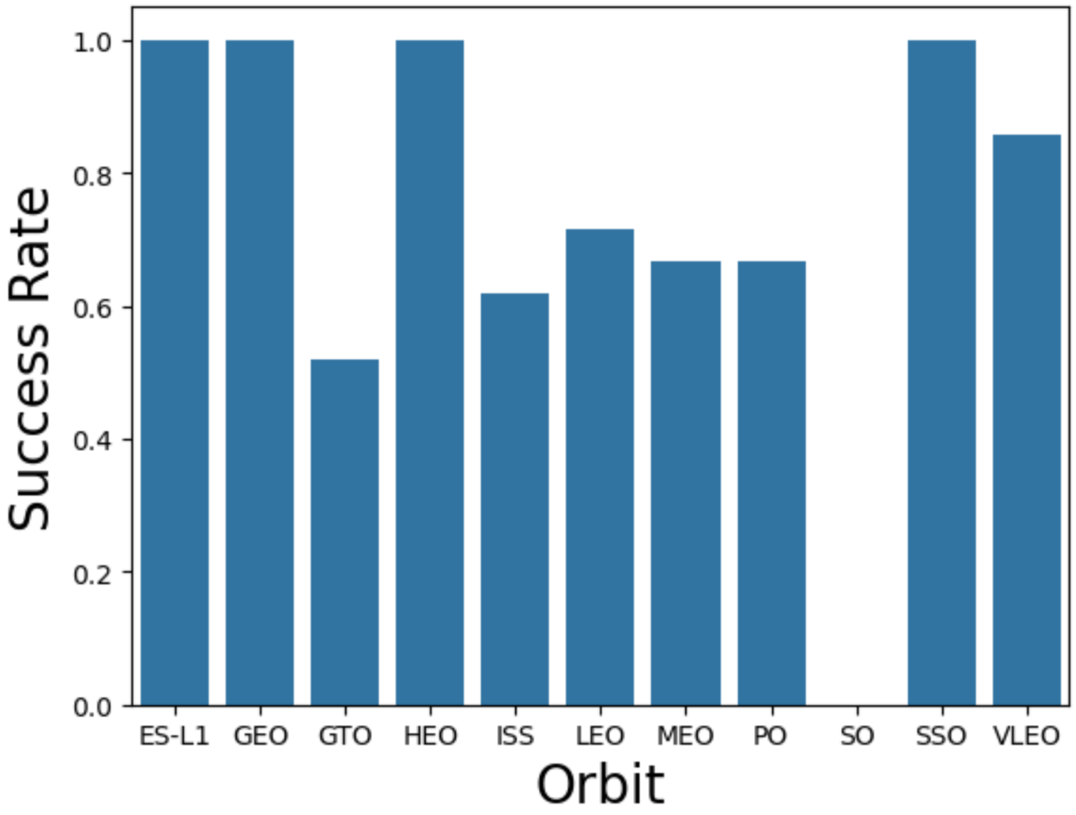
Payload vs. Launch Site

- This scatter plot illustrates the relationship between the payload mass and the launch site for SpaceX missions.
- By examining the data points, we can observe how the payload mass varies across different launch sites.
- This visualization reveals that the majority of SpaceX flights had a payload mass of less than ~7000 kg across all launch sites. Additionally, it demonstrates that the highest payload were reserved for the KSC LC 39A and CCAFS SLC 40 launch sites.



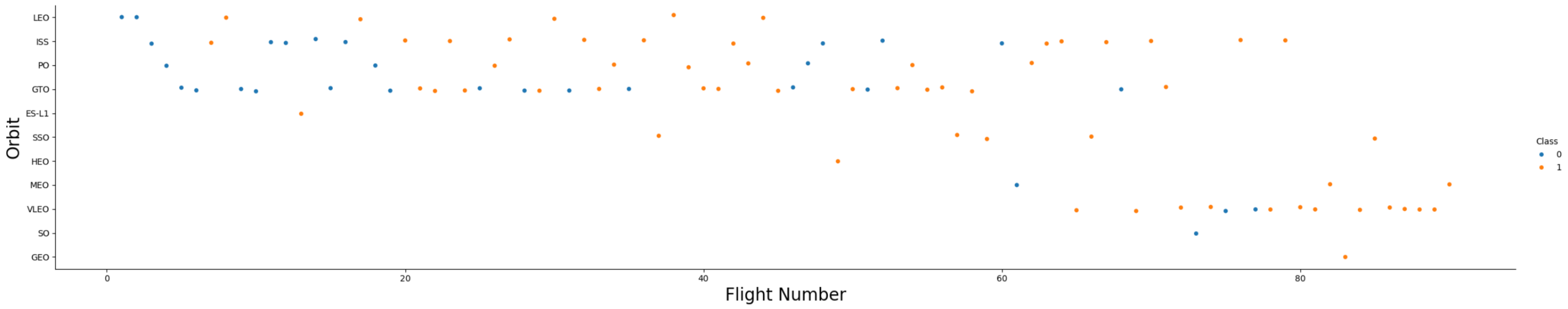
Success Rate vs. Orbit Type

- This bar plot illustrates the success rates of SpaceX missions across different orbit types.
- It provides a clear comparison, indicating which orbits have higher success rates.
- Orbits with highest success rate are ES-L1, GEO, HEO, and SSO.
- The Orbit with the lowest success rate is GTO.



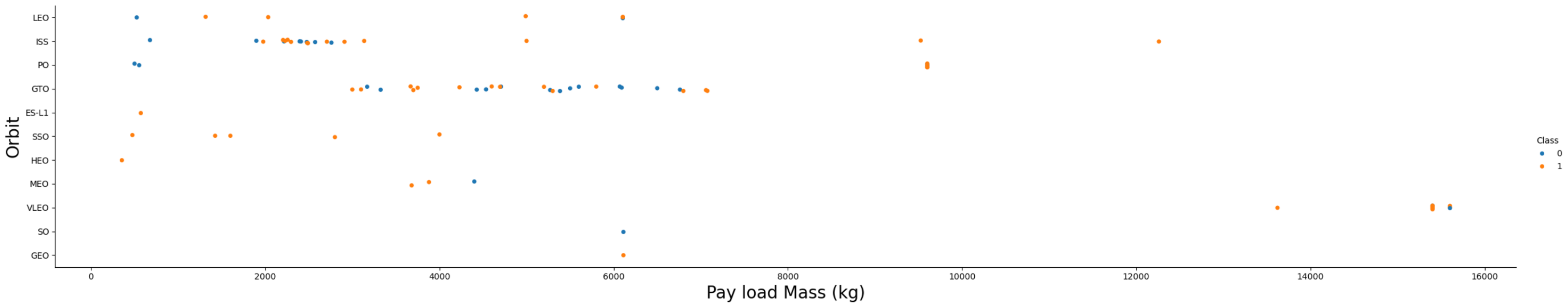
Flight Number vs. Orbit Type

- This scatter plot visualizes the relationship between the flight number and orbit type for SpaceX missions.
- Most flights occurred in the orbits LEO, ISS, PO, and GTO with increasing successful outcomes as number of flights increased.



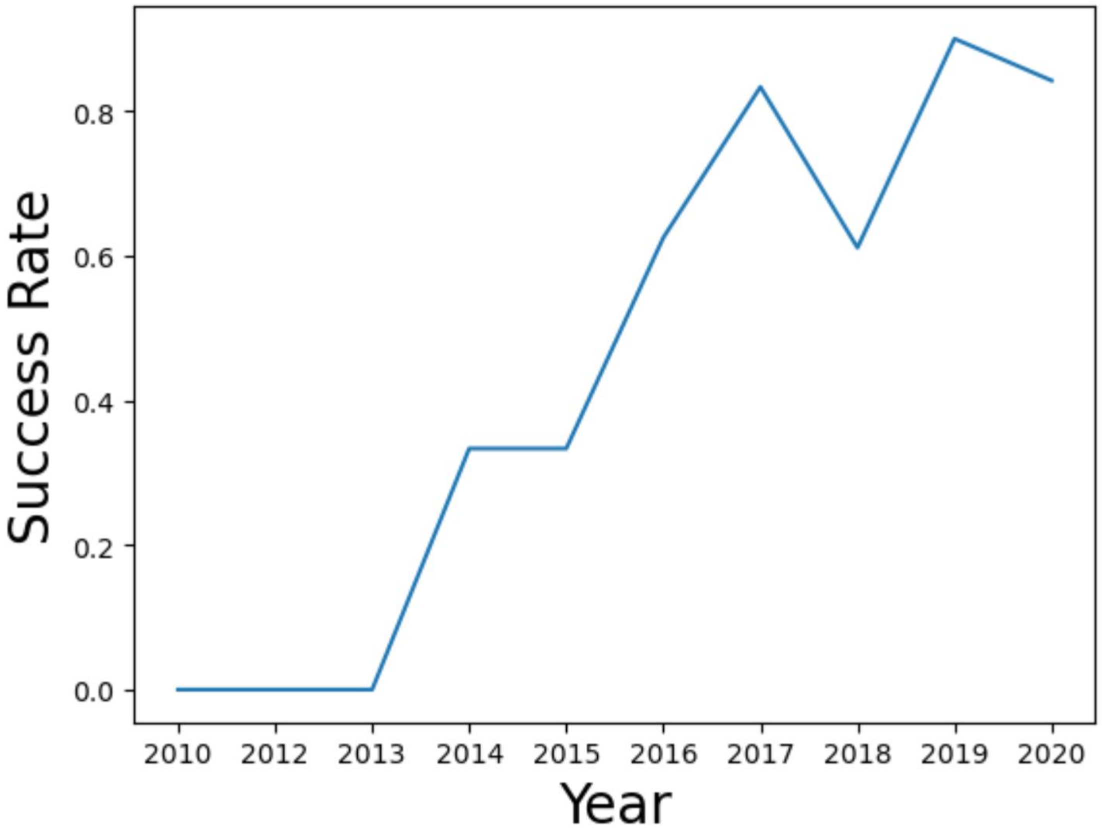
Payload vs. Orbit Type

- This scatter plot visualizes the relationship between the pay load mass and orbit type for SpaceX missions.
- The highest payload mass around 15500 kg was used for the VLEO orbit.
- The lowest payload mass around 500 kg was used for the HEO orbit.
- The highest number of flights occurred in the GTO orbit.



Launch Success Yearly Trend

- This line chart visualizes the relationship between the years 2010 to 2020 and success rate for SpaceX launches.
- Successful launches didn’t start to occur until 2013 but then plateaued from 2014 to 2015.
- Successful launches increased from 2015 to 2017 but then dipped until 2018.
- After a steady rise in successful launches starting in 2018, they slightly dipped again from to 2019 to 2020.



All Launch Site Names

- The initial SQL query aimed to list all the launch site names by utilizing 'SELECT DISTINCT’.
- In earlier data visualizations, only three launch sites were depicted. However, this particular SQL query reveals four.
- The prevalence of flights at the CCAFS SLC-40 launch site might be attributed to the fact that it encompasses two launch sites: CCAFS LC-40 and CCAFS SLC-40, as indicated in this SQL query.

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

Launch Site Names Begin with 'CCA'

- The second SQL query aimed to list five rows with launch site names that begin with ‘CCA’.
- This SQL query was achieved with a WHERE clause that queries rows with Launch Sites names LIKE ‘CCA%’ and LIMIT 5 rows.
- The boosters in this query reach the orbit LEO using a max payload of 677 Kg with all landing outcomes being failure or no attempt.
- This lines up with the previous flight number vs orbit type visualization where the landing outcomes for orbit LEO did not become successful until after the payload was over 1000 Kg.

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 third SQL query aimed to show the total payload mass used for boosters by the customer 'NASA(CRS)'.
- This SQL query was achieved with a SUM aggregate function on the Payload Mass column with a WHERE clause that queries rows with the Customer column being equal to 'NASA(CRS)'.
- From previous visualizations, the total payload mass is very likely the sum of payload masses for different missions from 500 Kg to 7000 Kg.

Total_Payload_Mass_By_NASA(CRS)
45596

Average Payload Mass by F9 v1.1

- The fourth SQL query aimed to show the average payload mass used for missions with the F9 v1.1 booster.
- This SQL query was achieved with a AVG aggregate function on the Payload Mass column with a WHERE clause that queries rows with the Booster_Version column being equal to ‘F9 v1.1’.
- From the previous visualization Payload vs Orbit Type, one can conclude that most of the missions with the F9 v1.1 booster reached the orbits of ISS, GTO, and SSO given the typical payload below.

AVG_Payload_Mass_By_F9v1.1
2928.4

First Successful Ground Landing Date

- The fifth SQL query aimed to show the date of the first successful landing on a Ground Pad.
- This SQL query was achieved with the MIN function on the Date column with a WHERE clause that queries rows with the Landing_Outcome column being equal to ‘Success (ground pad)’.
- The previous Yearly Average Success Rate Line Chart shows a plateau in success ending in 2015 probably due to the successful landing on the ground pad occurring that year.

Date of First Success Landing in Ground Pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The sixth SQL query aimed to a list of the names of boosters with successful landings on a drone ship and with a Payload between 4000 Kg and 6000 Kg.
- This SQL query was achieved with a WHERE clause that queries rows with the Landing_Outcome column being equal to ‘Success (drone ship)’ and the ‘PAYLOAD_MASS__KG’ column being between 4000 and 6000.
- From the previous visualization of the Payload vs. Launch Site scatter plot, we can conclude that the majority of successful landings took place at the CCAFS SLC 40 Launch Site. This is particularly evident for payloads weighing between 4000 kg and 6000 kg, where most successful landings occurred.

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

Total Number of Successful and Failure Mission Outcomes

- The seventh SQL query aimed to show the count of the total number of successful missions and failed missions.
- This SQL query was achieved by selecting the COUNT function for all rows WHERE the ‘Mission_Outcome’ column was LIKE ‘%Success%’.

Total_Number_Of_Successful_Missions
100

Total_Number_Of_Failed_Missions
1

Boosters Carried Maximum Payload

- The eighth SQL query aimed to show a list of all the boosters which carried the maximum payload.
- This SQL query uses a nested subquery in the WHERE clause to select from the 'Booster_Version' column. The selection is based on the 'PAYLOAD_MASS__KG_' column being equal to the maximum value in the 'PAYLOAD_MASS__KG_' column, determined using the MAX function.
- From the previous Payload vs. Launch Site and the Payload vs. Orbit Type scatter plots, the launch sites these booster were launched from were either the KSC LC 39A or the CCAFS SLC 40 launch sites and the orbit they reached was the MEO orbit.

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

- The ninth SQL query aimed to 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.
- This SQL query uses the substring function to extract the month from the Date column in the SELECT clause and the year 2015 under the WHERE clause.
- These two failed landings probably contributed to the plateau success rate in the previous Yearly Average Success Rate Line Chart in 2015.

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- The final SQL query performed ranks 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.
- This SQL query SELECTS the ‘Landing_Outcome’ column and the Count function of all records WHERE the Date column is BETWEEN the two dates specified above.
- This SQL query also uses a GROUP BY clause on the ‘Landing_Outcome’ column and an ORDER BY clause on the total count in DESC order.

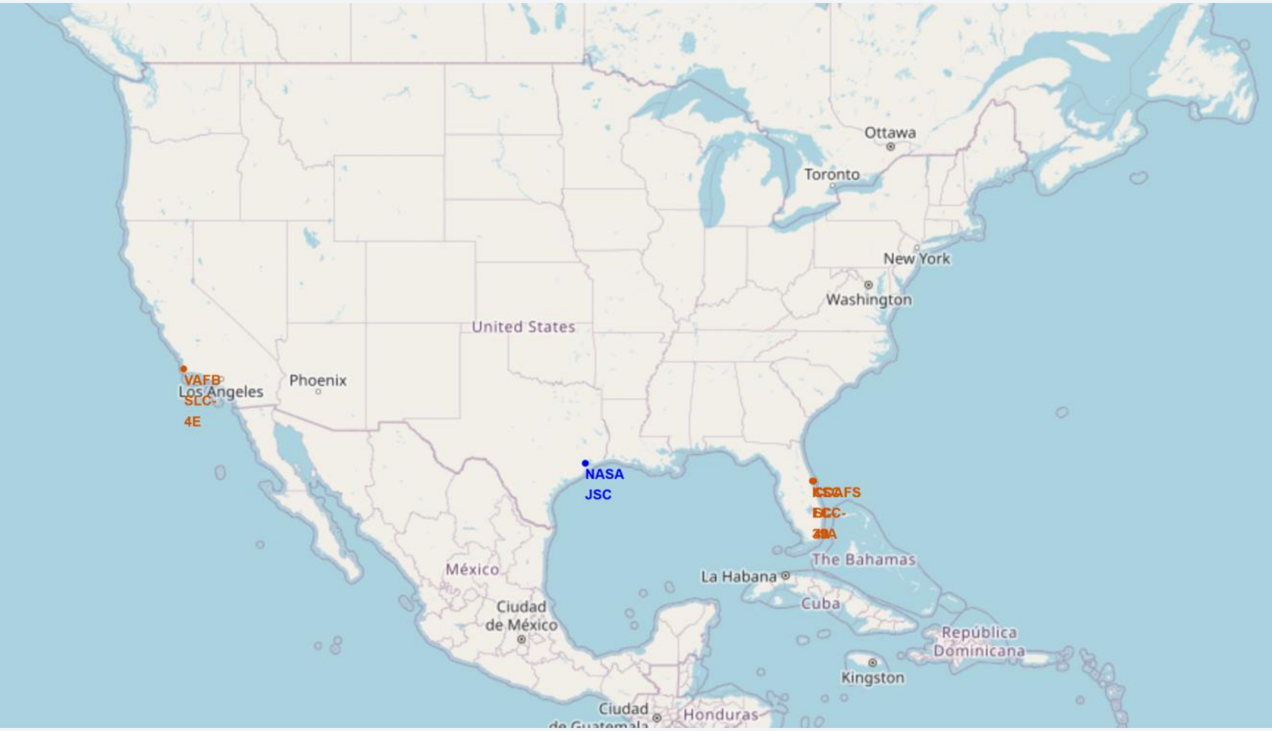
Landing_Outcome	Count_Of_Landing_Outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

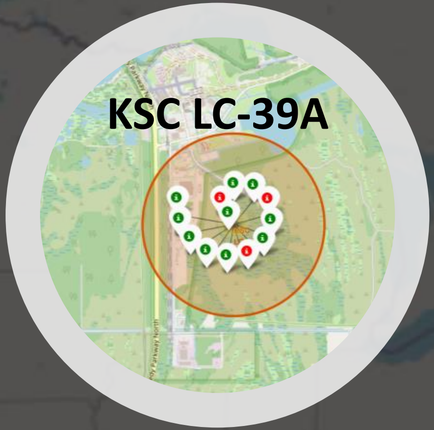
Map of All SpaceX Launch Sites

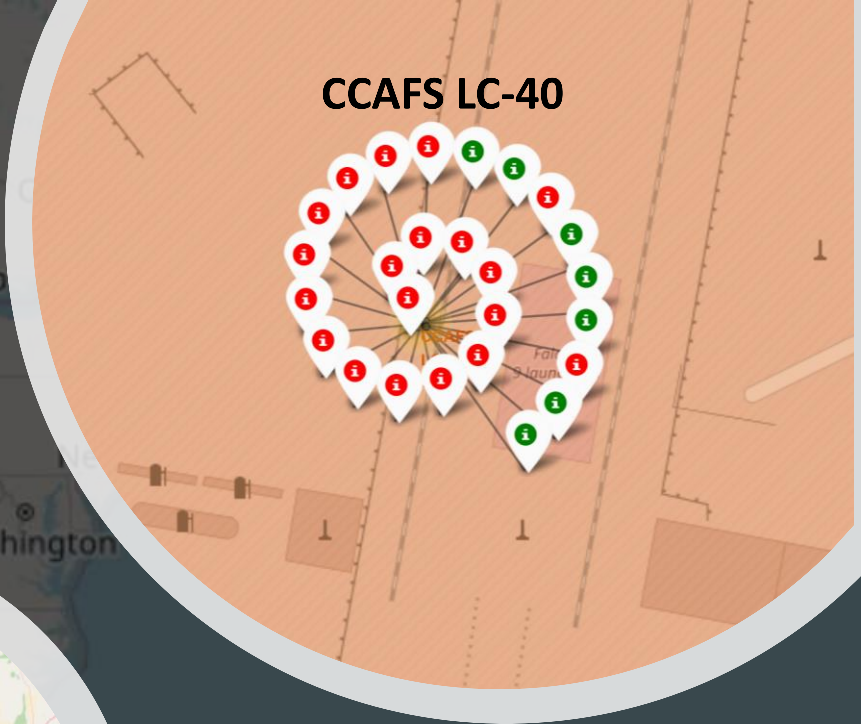
- **Map Insights and Findings:**
  - Launch sites are not all close to the Equator, indicating diverse geographic locations.
  - All launch sites are in close proximity to coastlines, facilitating safer launches and recoveries.
  - Launch sites are strategically chosen for their geographic advantages in terms of safety and logistics.

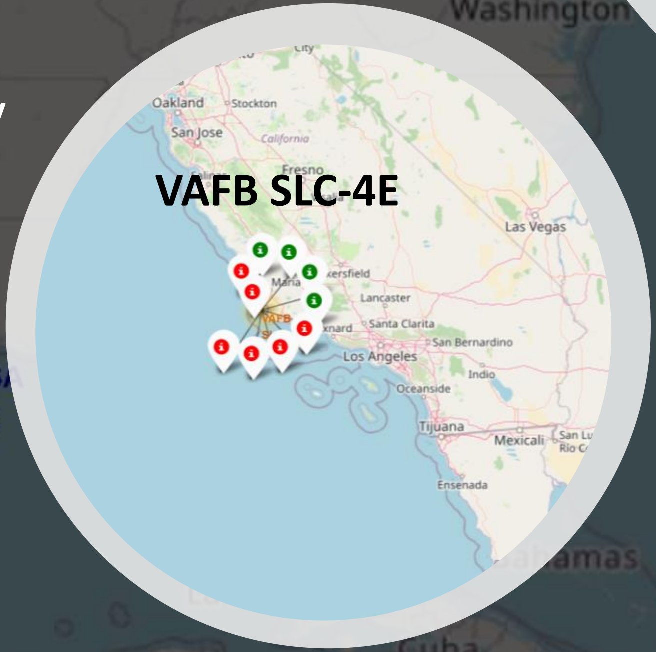


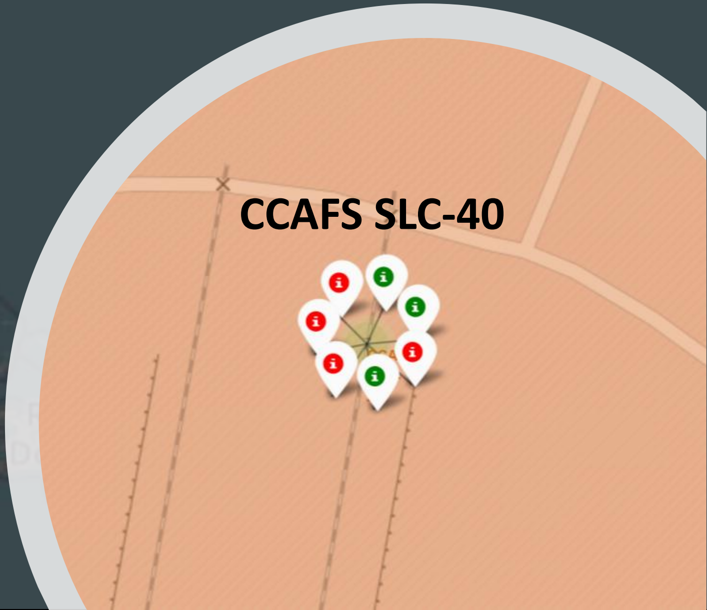
Maps of Successes and Failures for Each Launch Site

- **Map Insights and Findings:**
  - **Success Rate Distribution:**
    - Green markers indicate successful launches.
    - Red markers indicate failed launches.
    - Patterns in marker colors help identify sites with higher success rates.
  - **Visual Clarity:**
    - Marker clusters simplify the visualization of dense data points.
  - **Launch Site Performance:**
    - Sites with more green markers have higher success rates.
    - Sites with more red markers indicate areas with improvement needs.



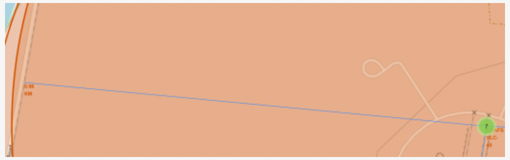






Map of the Proximity of CCAFS SLC-40 Launch Site to Railway, Highway, and Coastline

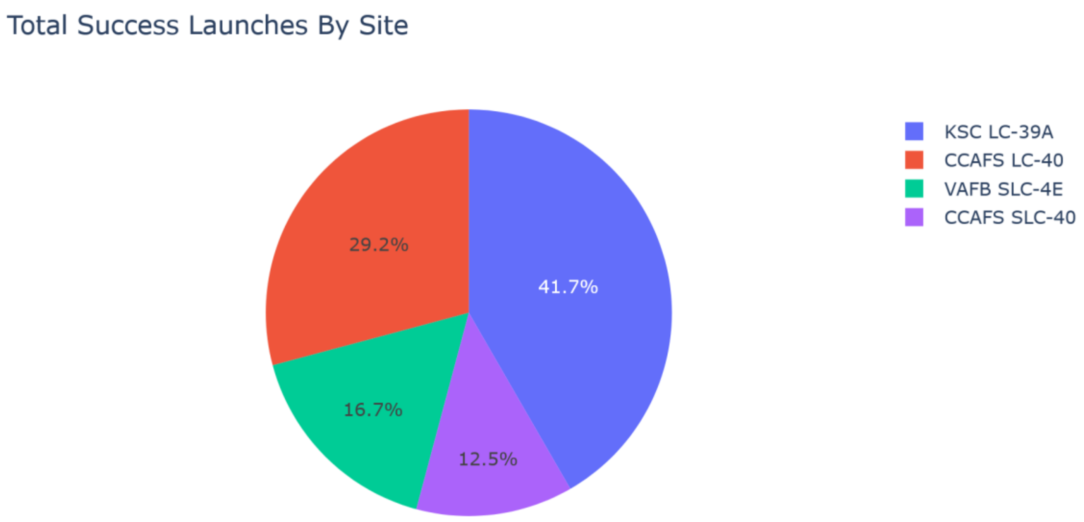
- **Explanatory Notes:**
  - **Map Features:**
    - Orange markers show distances to points of interest.
    - PolyLines indicate the paths measured.
  - **Proximity Insights:**
    - Launch sites are generally close to coastlines and highways.
    - They maintain a safe distance from cities.
    - Proximity to railways varies but is generally within a few kilometers.



Section 4

Build a Dashboard with Plotly Dash

Total Success Launches By Site

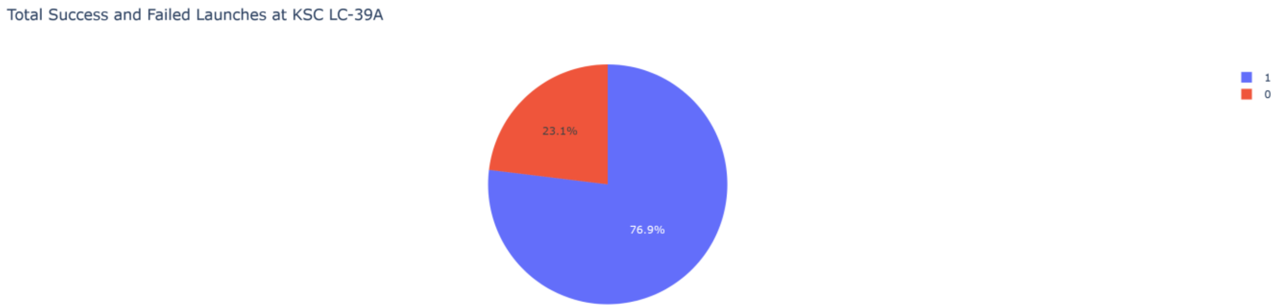


- The pie chart clearly shows that the KSC LC-39A launch site has the highest number of successful launches.
  - This conclusion is supported by the previous slide, which illustrated that the majority of launches at KSC LC-39A were successful.
- On the other hand, the CCAFS SLC-40 launch site has the lowest number of successful launches.
  - This is likely because CCAFS SLC-40 has the fewest total launches among all the sites.

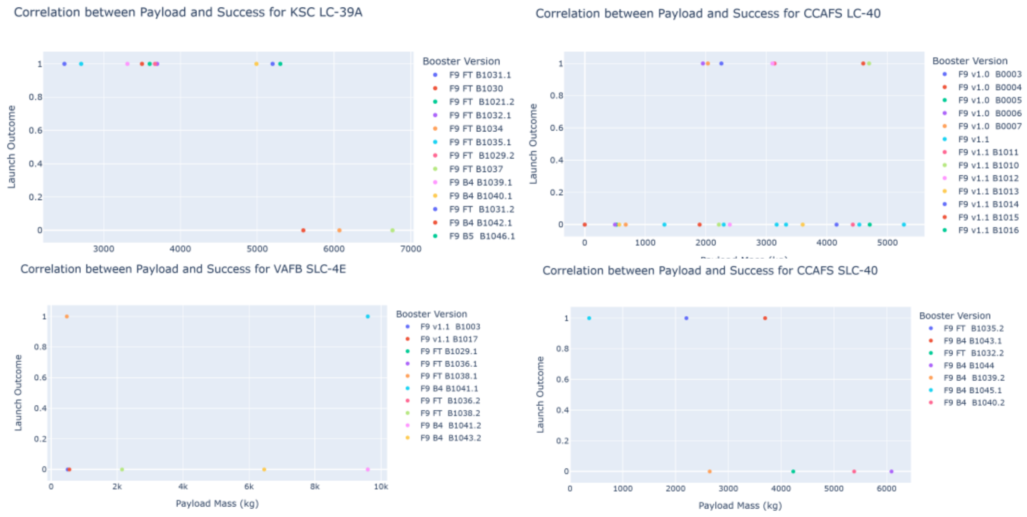
Total Success and Failed Launches at KSC SL-39A

As shown on the previous slide, the KSC SL-39A launch site has the highest success launch ratio among all sites.

This could be because the KSC SL-39A launch site is located closer to the center of the United States compared to the other launch sites.



Correlation between Payload and Success for Different Launch Sites



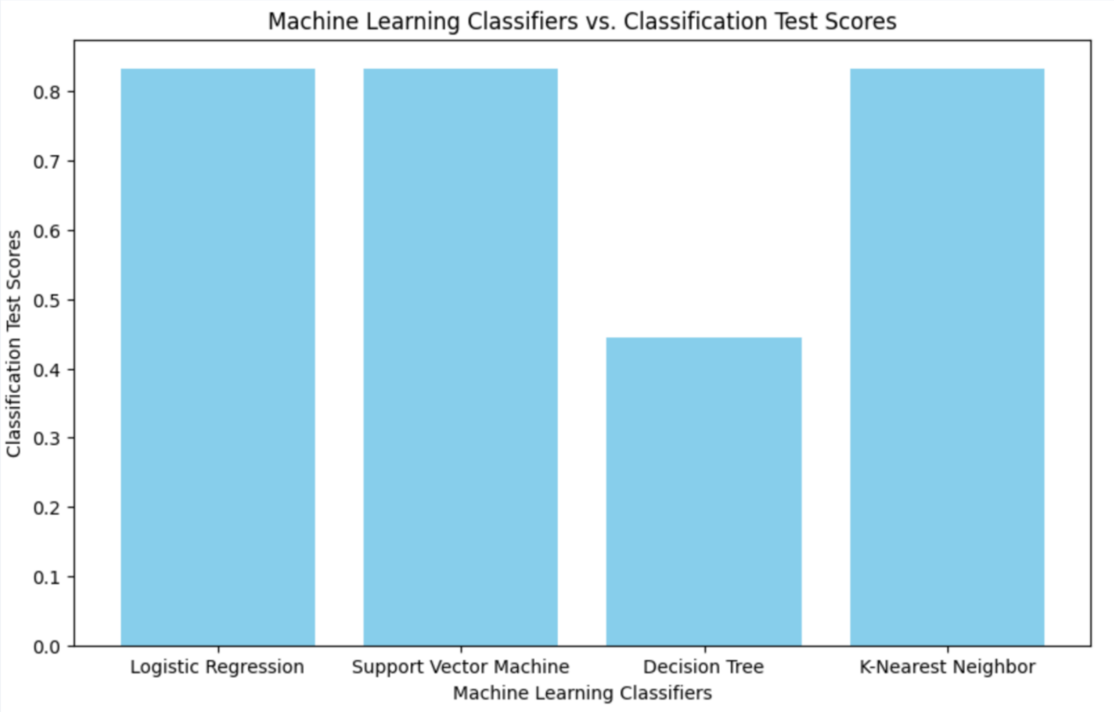
- The scatter plots reveal that the FT booster version with payloads under 5500 kg at the KSC LC-39A launch site has the highest success rate.
- Conversely, the v1.0 and v1.1 booster versions at the CCAFS LC-40 launch site exhibit the highest failure rates.
- At the VAFB SLC-4E launch site, launches tend to be successful only if the payload is either very high or very low, within the range of 0 to 10,000 kg.

Section 5

Predictive Analysis (Classification)

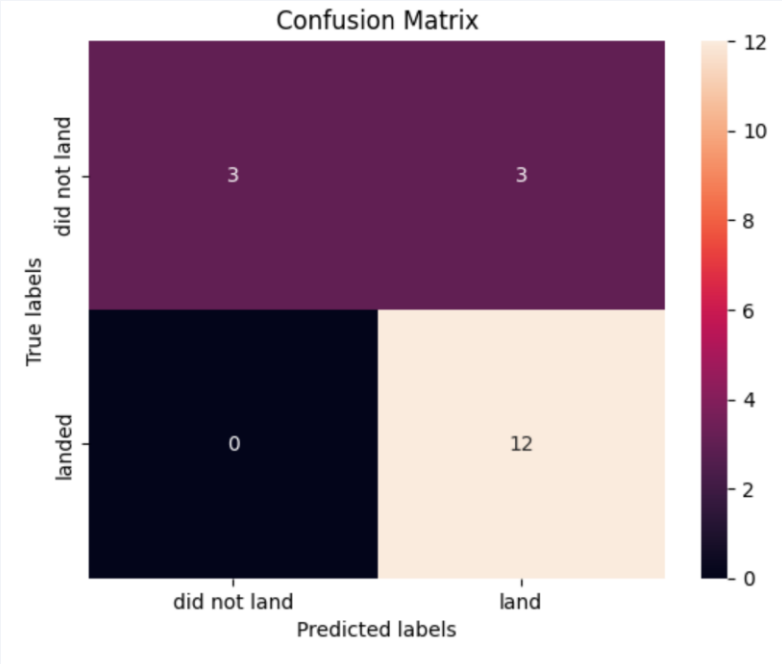
Classification Accuracy

- Logistic Regression with GridSearchCV
  - Using GridSearchCV with a CV of 10, the best hyperparameters were {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}.
  - Accuracy of training data was 0.85 while accuracy of test data as shown in the bar chart was 0.83.
- Support Vector Machine with GridSearchCV
  - Using GridSearchCV with a CV of 10, the best hyperparameters were {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}.
  - Accuracy of training data was 0.85 while accuracy of test data as shown in the bar chart was 0.83.
- Decision Tree Classifier with GridSearchCV
  - Using GridSearchCV with a CV of 10, the best hyperparameters were {'criterion': 'entropy', 'max_depth': 16, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'best'}.
  - Accuracy of training data was 0.86 while accuracy of test data as shown in the bar chart was 0.44.
- K-Means Clustering Classifier with GridSearchCV
  - Using GridSearchCV with a CV of 10, the best hyperparameters were {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}.
  - Accuracy of training data was 0.85 while accuracy of test data as shown in the bar chart was 0.83.



Confusion Matrix

- The confusion matrix below for the Logistic Regression, Support Vector Machine, and K-Means Clustering classifiers shows similar results.
- All these classifiers perform exceptionally well at predicting positive values.
- However, they struggle with predicting negative values, evidenced by the presence of 3 true negatives and 3 false positives.



Conclusions

- The most likely scenario for a successful and reusable SpaceX rocket launch is using an FT booster with a payload under 5500 kg at the KSC LC-39A Launch Site.
- A SpaceX rocket launch at the CCAFS LC-40 site with a v1.0 or v1.1 booster, regardless of payload, is most likely to result in a failure where the first stage will not be reused.
- The Logistic Regression, Support Vector Machine, and K-Means Clustering models all performed best on the test data, each achieving an accuracy of 0.83.

Thank you!