IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jinning
May 24th, 2024

Outline

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

Executive Summary

- Summary of methodologies

  - In this capstone project, we utilized a variety of methodologies to predict the successful landing of the Falcon 9 first stage. We started by collecting and wrangling data using RESTful APIs and web scraping, converting the information into a Pandas DataFrame for analysis. We conducted exploratory data analysis with Matplotlib and Seaborn to visualize data patterns and relationships. SQL was employed for efficient data extraction and manipulation. Interactive dashboards and maps were created using Plotly Dash and Folium to analyze launch records and site proximities. Finally, machine learning techniques, including SVM, classification trees, and logistic regression, were used to train and optimize models for predicting landing success, with hyperparameter tuning conducted through grid search to identify the best performing model.

- Summary of all results

  - The project revealed several key insights about Falcon 9 first-stage landings. Data collection through APIs and web scraping provided comprehensive datasets, which were cleaned and prepared for analysis. Exploratory data analysis identified significant patterns, such as the impact of launch site and payload mass on landing success. Interactive visualizations highlighted these patterns effectively. Machine learning models were trained to predict landing outcomes, with the logistic regression model showing the highest accuracy. This model's predictions suggest that specific conditions, like favorable weather and optimized payload weights, greatly enhance the probability of successful landings, providing valuable insights for optimizing launch operations and competitive bidding against SpaceX.

Introduction

- Project background and context
  - ◆ SpaceX has revolutionized space travel by significantly reducing the cost of rocket launches through the reusability of its Falcon 9 first stage. Traditional rocket launches cost upwards of 165 million dollars, while SpaceX offers launches for 62 million dollars due to their ability to reuse the first stage. This project aims to predict the successful landing of the Falcon 9 first stage, providing valuable insights that could help other companies compete in the space launch market. By analyzing historical launch data and leveraging machine learning techniques, we seek to understand the factors that contribute to successful landings and offer data-driven strategies for optimizing launch operations.
- Problems you want to find answers
  - ◆ The primary problem this project seeks to address is determining the factors that influence the successful landing of the Falcon 9 first stage. Given SpaceX's cost advantage through reusability, identifying these factors is crucial for other companies aiming to compete in the space launch market. Specifically, we want to understand how variables such as launch site, payload mass, weather conditions, and launch timing affect landing outcomes. Additionally, we aim to develop a predictive model that can accurately forecast landing success, thereby aiding in decision-making processes for optimizing launch operations and reducing costs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using RESTful APIs and web scraping techniques, then organized into a Pandas DataFrame for analysis.
- Perform data wrangling
  - The data was processed by cleaning and wrangling it using Python and Pandas to handle missing values and inconsistencies, ensuring it was ready for analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - EDA was performed using visualization tools like Matplotlib and Seaborn, alongside SQL queries, to uncover patterns and insights within the data.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive visual analytics were conducted using Folium and Plotly Dash to create dynamic maps and dashboards that facilitated an in-depth analysis of launch site proximities and launch records.
- Perform predictive analysis using classification models
  - Predictive analysis was conducted using classification models, including SVM, classification trees, and logistic regression, by building, tuning, and evaluating these models through hyperparameter optimization and performance testing on training and testing datasets.

Data Collection

- Describe how data sets were collected.
  - Data sets were collected by leveraging RESTful APIs to gather structured data from SpaceX's database and web scraping techniques to extract additional relevant information from various online sources. The collected data was then consolidated into a comprehensive Pandas DataFrame, ensuring all relevant attributes such as launch dates, payload details, and landing outcomes were included. This organized and clean dataset formed the basis for subsequent analysis and modeling.

- Data Collection flowchart
  - Initiate Data Collection:
    - Use RESTful API to access SpaceX database.
    - Employ web scraping for supplementary data.
  - Data Aggregation:
    - Consolidate data from API and web sources into Pandas DataFrame.
  - Data Cleaning:
    - Handle missing values and inconsistencies.
    - Ensure data quality for analysis.
  - Final Dataset:
    - Ready-to-use, clean dataset for EDA and modeling.

Data Collection – SpaceX API

- Data sets were collected by leveraging Beautifulsoup to gather structured data on rocket launches, including details such as launch dates, payloads, and landing outcomes. This process involved initiating API calls to access Wikipedia, retrieving the relevant data, and then consolidating this information into a Pandas DataFrame. The data was subsequently cleaned and processed to handle any missing values and ensure consistency, forming a comprehensive dataset ready for exploratory data analysis and modeling.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Collecting%20the%20Data.ipynb

Initialize Data Collection:

- Import necessary libraries such as requests, pandas, numpy, and datetime to handle HTTP requests, data manipulation, and date representation.

Web Scraping Wikipedia:

- Fetch Data: Use the requests library to retrieve HTML content from Wikipedia pages containing information about SpaceX launches.
- Parse HTML: Utilize libraries like BeautifulSoup to parse the HTML content and extract relevant data points, such as launch dates, payload details, and landing outcomes.

Extract Specific Information:

- Launch Details: Extract specific details like rocket version, launch site, payload mass, orbit, and core information by navigating through the HTML elements and tags.
- Clean Data: Clean the extracted data to handle missing values, inconsistencies, and to ensure that the data is formatted correctly for further analysis.

Consolidate Data:

- Organize the cleaned data into a comprehensive Pandas DataFrame, ensuring all relevant attributes are included. This structured dataset forms the foundation for exploratory data analysis and predictive modeling.

Data Collection - Scraping

- The web scraping process involved using the requests library to fetch HTML content from a Wikipedia page on Falcon 9 and Falcon Heavy launches. The HTML content was then parsed using BeautifulSoup, which allowed us to extract specific data points such as launch dates, booster versions, landing statuses, and payload mass. The extracted data was cleaned, normalized, and structured into a Pandas DataFrame for further analysis and modeling.

- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Collecting%20the%20Data.ipynb

Initialize Data Collection:

Install necessary libraries using pip commands: beautifulsoup4 and requests.

Import Libraries:

Import essential libraries: requests for making HTTP requests, BeautifulSoup for parsing HTML content, and pandas for data manipulation.

Fetch Data from Wikipedia:

API Call: Use requests.get() method with the provided static URL to fetch the HTML content of the Wikipedia page on Falcon 9 and Falcon Heavy launches.

Response Handling: Assign the HTML response to an object.

Parse HTML Content:

Initialize BeautifulSoup: Use BeautifulSoup to parse the HTML content from the Wikipedia page.

Extract Data: Define functions to extract specific data points such as launch dates, booster versions, landing statuses, and payload mass from the HTML table cells.

Clean and Structure Data:

Data Normalization: Normalize and clean the extracted data using unicodedata and other string manipulation techniques.

Column Extraction: Extract and format column headers for creating a structured Pandas DataFrame.

Consolidate Data:

Organize the cleaned data into a comprehensive Pandas DataFrame, ensuring all relevant attributes are included for further analysis and modeling.

Data Wrangling

- The data processing involved loading the dataset into a Pandas DataFrame, identifying and handling missing values, and verifying the data types of each column. Categorical data distributions were analyzed using value_counts(), and the data was cleaned and normalized to ensure consistency and readiness for further analysis. This comprehensive data wrangling ensured the dataset was complete, accurate, and properly formatted for exploratory analysis and predictive modeling.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Data%20Wrangling.ipynb

Library Import:

Import pandas for data manipulation and analysis.

Import numpy for numerical operations and handling large arrays.

Load Data:

Read the CSV file into a Pandas DataFrame using pd.read_csv(), providing a structured format for the data.

Handle Missing Values:

Identify missing values using df.isnull().sum()/len(df)*100 to calculate the percentage of missing data in each column.

Apply appropriate methods to handle or impute missing values, ensuring data completeness.

Data Type Check:

Verify data types of each column using df.dtypes to ensure they are appropriate for analysis.

Categorical Data Processing:

Use value_counts() on categorical columns, such as LaunchSite, to understand the distribution and frequency of each category.

Clean and Normalize Data:

Normalize data by removing inconsistencies and ensuring uniformity in data representation.

EDA with Data Visualization

- Charts Plotted and Their Purpose:
  - Categorical Plot (Catplot):
    - Chart: sns.catplot showing the relationship between PayloadMass and FlightNumber, with hue set to Class.
    - Purpose: This chart was used to visualize how payload mass varies across different flight numbers and to observe the impact on the landing outcome (success or failure).
  - Bar Chart:
    - Chart: plt.bar to display the count of successful and unsuccessful landings.
    - Purpose: To provide a clear visual representation of the distribution of successful versus unsuccessful landings, helping to identify the overall success rate.
  - Scatter Plot:
    - Chart: plt.scatter to plot FlightNumber against PayloadMass.
    - Purpose: To analyze the distribution and pattern of payload mass across different flight numbers, identifying any trends or outliers.
  - Heatmap:
    - Chart: sns.heatmap to display the correlation matrix of various features.
    - Purpose: To identify the relationships and correlations between different variables, providing insights into which features might be important for predictive modeling.

- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Exlporing%20and%20Preparing%20Data.ipynb

EDA with SQL

- Database Connection:
    - Connected to the SpaceX database using SQLite.
- Create Tables:
    - Created tables to store SpaceX launch data, including details about rockets, launch sites, payloads, and cores.
- Insert Data:
    - Inserted collected data into the respective tables for further querying.
- Select Queries:
    - Queried specific columns from tables to retrieve relevant information about launches, rockets, and payloads.
- Aggregate Functions:
    - Used aggregate functions like COUNT, AVG, and SUM to summarize data.
- Join Queries:
    - Performed inner joins to combine data from multiple tables based on common keys.
- Filtering Data:
    - Applied WHERE clauses to filter data based on specific conditions.
- Sorting Data:
    - Ordered query results using the ORDER BY clause to sort data.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20SQL%20Notebook.ipynb

Build an Interactive Map with Folium

- In the Folium map, we added various map objects including markers to pinpoint the exact locations of SpaceX launch sites, circle markers to highlight the impact zones around these sites, and lines to illustrate paths or trajectories related to the launches. Additionally, we used the MarkerCluster plugin to manage clusters of markers, ensuring the map remained readable even with numerous markers in close proximity. The MousePosition plugin was also incorporated to display the latitude and longitude of the mouse pointer, facilitating precise identification of coordinates. These objects were added to create an interactive and informative map that enhances the visualization and understanding of the spatial relationships between SpaceX launch sites and their surroundings.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Launch%20Sites%20Location%20Analysis.ipynb

Build a Dashboard with Plotly Dash

- The dashboard includes interactive plots and graphs such as a dropdown menu for selecting different launch sites, a pie chart showing the total successful launches for each site, and a scatter plot visualizing the correlation between payload mass and launch success. These interactive elements allow users to filter the data by specific launch sites, providing a clear overview of successful launches. The scatter plot helps in understanding the relationship between payload mass and launch outcomes, enabling users to gain insights into how different factors influence launch success. These visualizations were added to facilitate dynamic data exploration and enhance user engagement by providing intuitive and informative insights.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/Space%20X%20Plotly%20Dash%20Dashboard.py

Predictive Analysis (Classification)

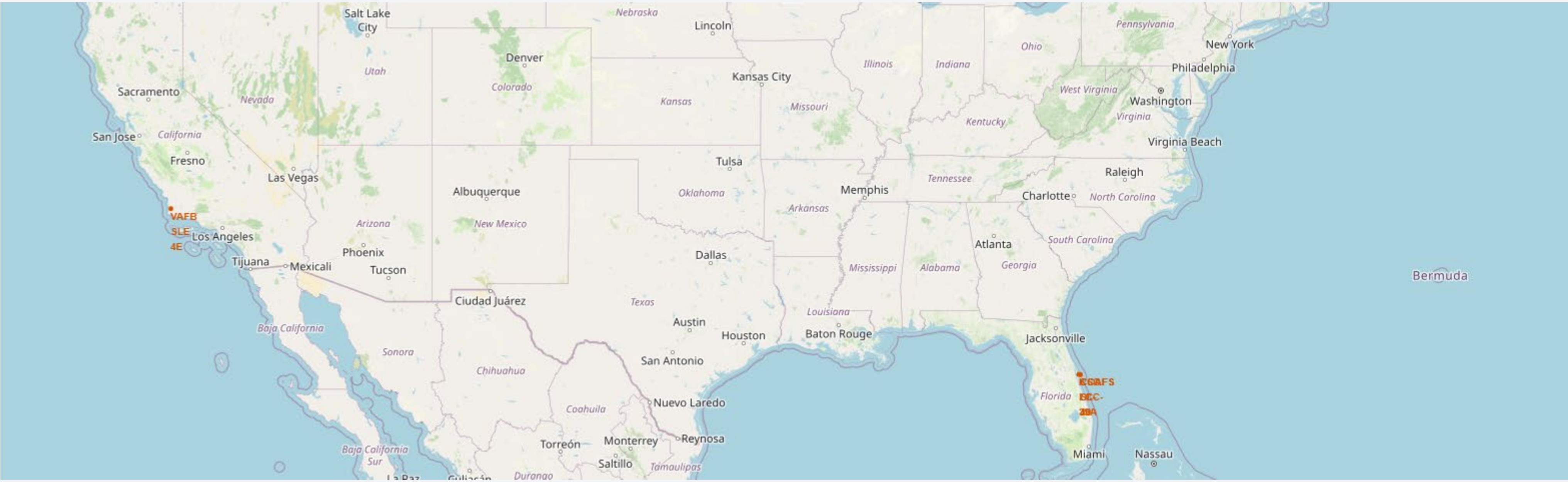
- In developing the best-performing classification model to predict the successful landing of the Falcon 9 first stage, we began by preprocessing the data, including standardizing features and splitting the dataset into training and testing sets. We then built multiple classification models using algorithms such as Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Each model was trained and evaluated using accuracy scores and confusion matrices. Hyperparameter tuning was conducted using GridSearchCV to optimize model performance. The best-performing model was identified based on its performance on the test data, ensuring accurate and reliable predictions.
- https://github.com/JPYan98/Space-X-Falcon-9/blob/main/SpaceX%20Machine%20Learning%20Prediction.ipynb

Flowchart: Model Development Process

- Preprocessing Data:
  - Standardize features using preprocessing techniques.
  - Split data into training and testing sets.
- Model Building:
  - Train multiple classification models: Logistic Regression, SVM, Decision Tree, KNN.
- Model Evaluation:
  - Evaluate models using accuracy scores and confusion matrices.
- Hyperparameter Tuning:
  - Optimize models using GridSearchCV.
- Select Best Model:
  - Identify the best-performing model based on test data performance.

Results

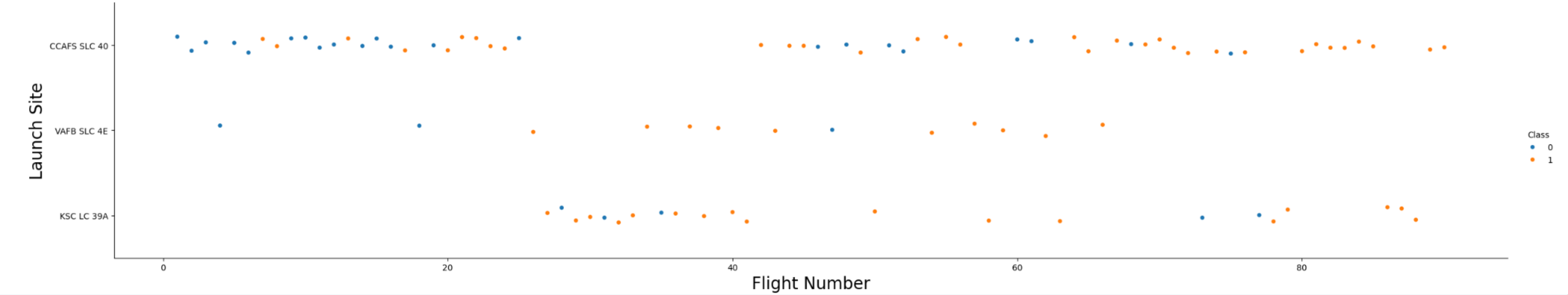
- The exploratory data analysis revealed key patterns such as the correlation between payload mass and flight numbers, and the impact of launch sites on landing success. Visualization tools like scatter plots, bar charts, and heatmaps were used to uncover these insights. The predictive analysis involved building and evaluating several classification models, including Logistic Regression, SVM, Decision Tree, and KNN. Through hyperparameter tuning and performance evaluation, the best model was identified, demonstrating reliable accuracy in predicting the successful landing of the Falcon 9 first stage, driven by factors such as payload mass, flight number, and launch site conditions.



Section 2

Insights drawn from EDA

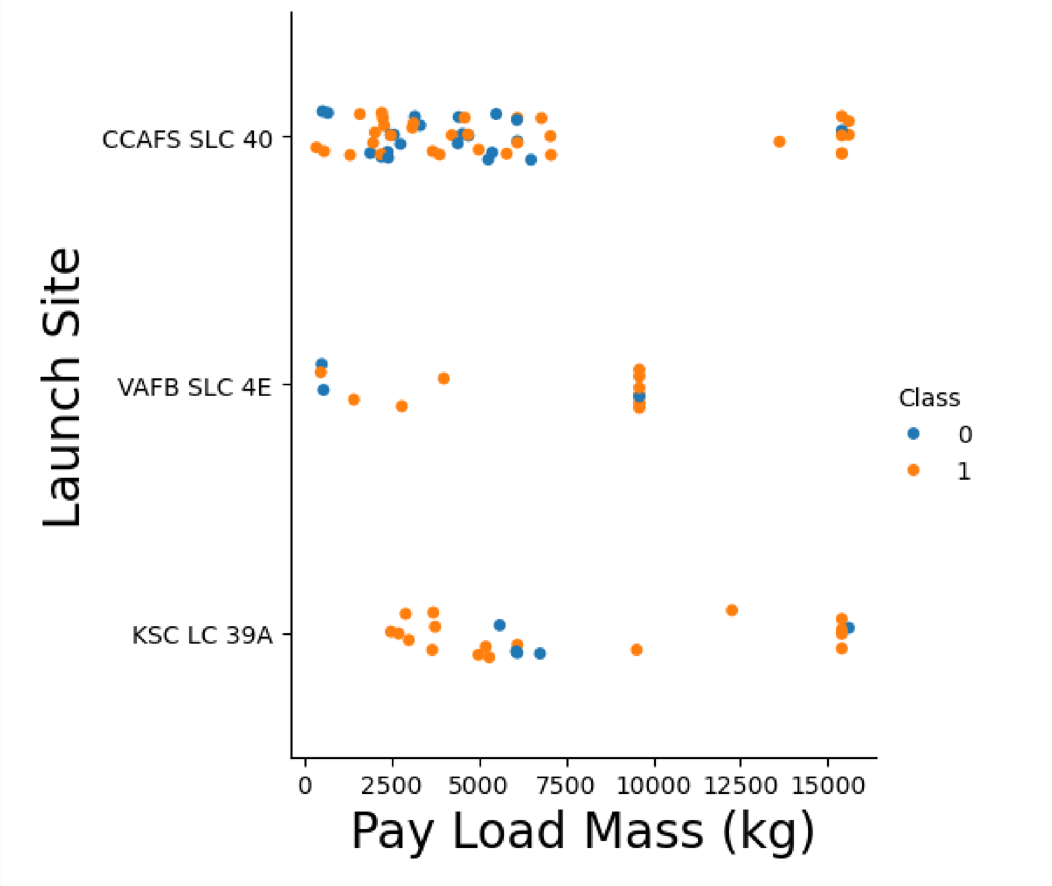
Flight Number vs. Launch Site



- This scatter plot visualizes the relationship between flight number and launch site, with each point representing a launch. The hue indicates the class value, where blue points represent unsuccessful landings (Class 0) and orange points represent successful landings (Class 1). The plot shows that launch successes and failures are distributed across different flight numbers and launch sites, highlighting patterns of success at specific sites such as CCAFS SLC 40 and KSC LC 39A, with fewer successful landings observed at VAFB SLC 4E.

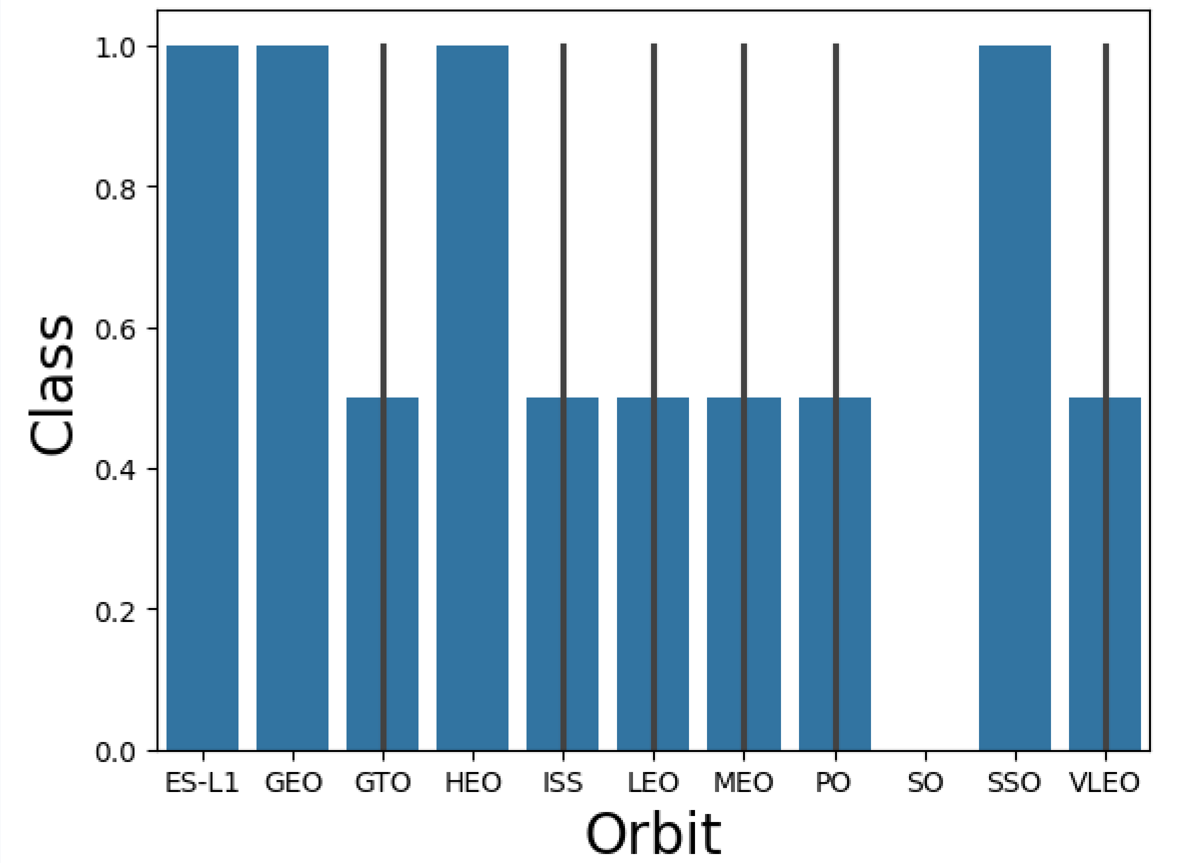
Payload vs. Launch Site

- This scatter plot shows the relationship between payload mass and launch site, with the hue indicating the class value (blue for unsuccessful landings and orange for successful landings). It reveals that successful and unsuccessful landings occur across a range of payload masses at different launch sites. Notably, CCAFS SLC 40 and KSC LC 39A have a higher frequency of successful landings with varied payload masses, while VAFB SLC 4E shows fewer launches and a mix of outcomes. This highlights the influence of both payload mass and launch site on the success of the Falcon 9 first stage landings.



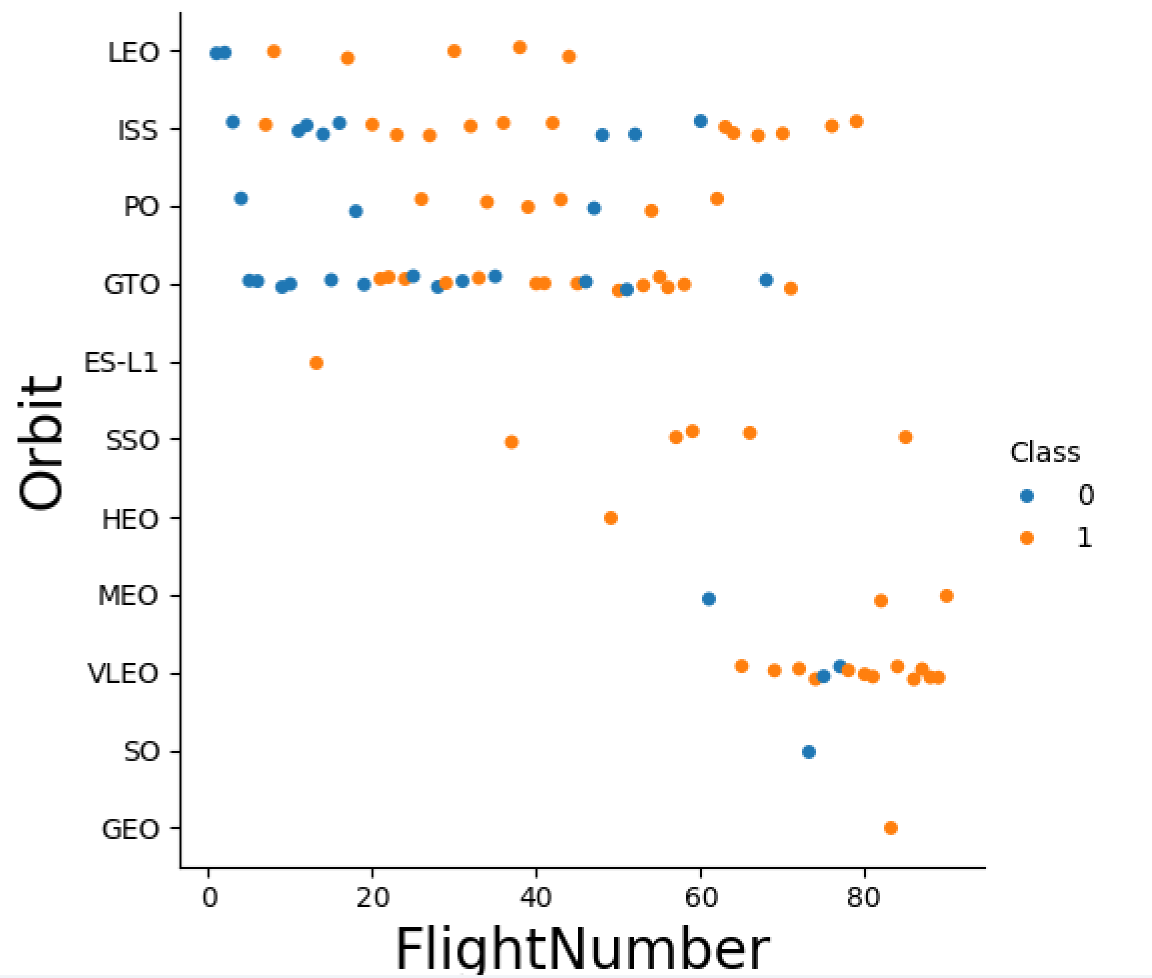
Success Rate vs. Orbit Type

- This bar chart illustrates the success rate of Falcon 9 first stage landings across different orbit types. Each bar represents the mean success rate (Class 1) for a specific orbit, showing high success rates for orbits like ES-L1, GEO, HEO, ISS, and SSO, while orbits such as GTO, LEO, MEO, PO, and VLEO exhibit lower success rates. This indicates that the likelihood of a successful landing varies depending on the target orbit, highlighting the impact of mission characteristics on landing outcomes.



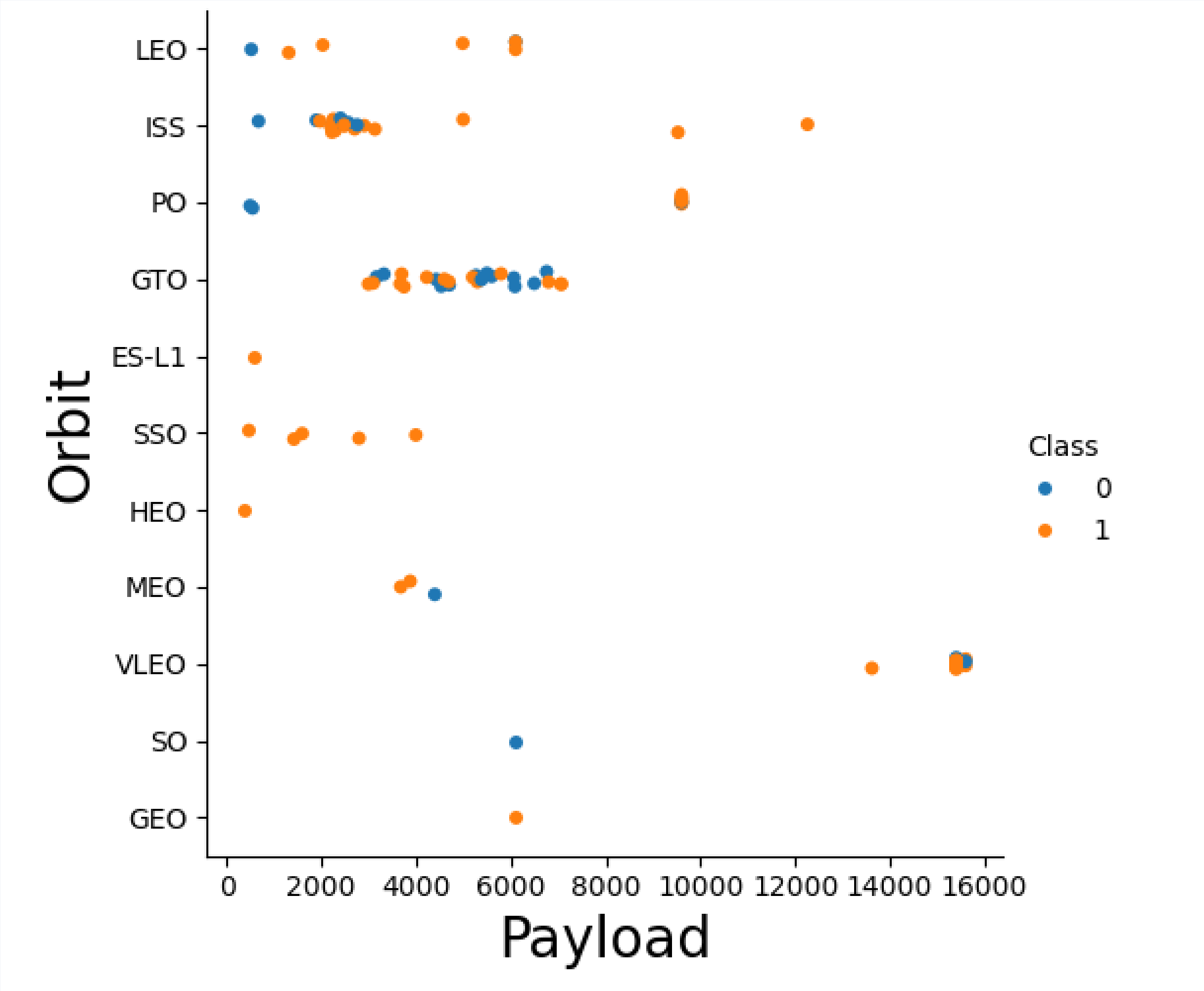
Flight Number vs. Orbit Type

- This scatter plot visualizes the relationship between flight number and orbit type, with each point representing a launch and the hue indicating the class value (blue for unsuccessful landings and orange for successful landings). The plot shows how the success of landings varies across different orbits and flight numbers, indicating that certain orbits, like LEO and ISS, have a mix of successful and unsuccessful landings throughout the flight numbers, while other orbits like VLEO and GEO show fewer data points and varied success rates, highlighting the influence of both flight experience and mission type on landing outcomes.



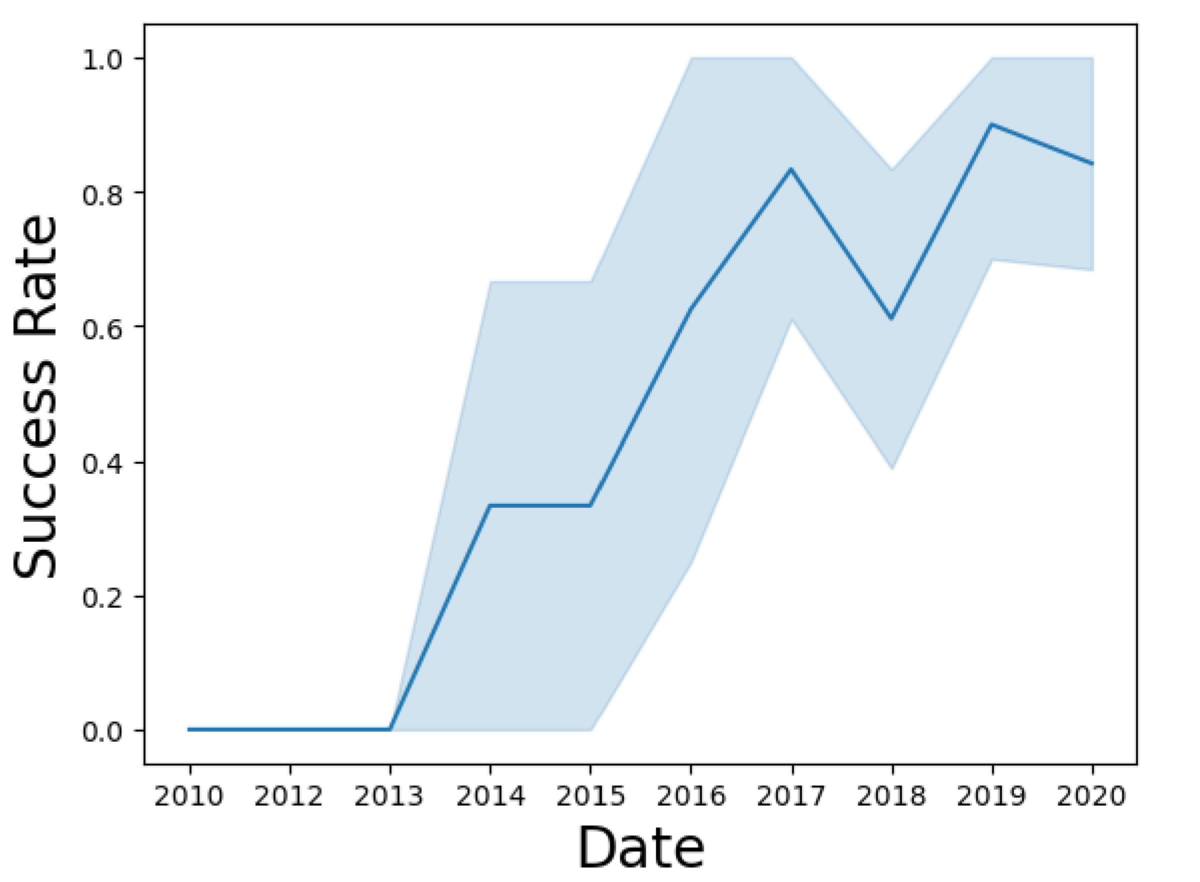
Payload vs. Orbit Type

- This scatter plot illustrates the relationship between payload mass and orbit type, with hue indicating the class value (blue for unsuccessful landings and orange for successful landings). It shows that successful and unsuccessful landings occur across various payload masses for different orbits. For instance, orbits like GTO and LEO have a wide range of payload masses with mixed outcomes, while high-payload missions in GEO and VLEO orbits show fewer data points with varied success rates. This highlights how both payload mass and orbit type influence the landing success of the Falcon 9 first stage.



Launch Success Yearly Trend

- This line graph shows the success rate of Falcon 9 launches over time, with the x-axis representing the date and the y-axis representing the success rate. The shaded area around the line indicates the confidence interval. The graph reveals a significant increase in the success rate from 2014 to 2018, with some fluctuations afterward, demonstrating overall improvement in launch success over the years.



All Launch Site Names

- The query identified four unique launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E. Each launch site is associated with specific latitude and longitude coordinates. These unique sites represent the primary locations used by SpaceX for their rocket launches, indicating their geographical distribution and the varying conditions each site offers for launch operations.

```
%sql select Unique(LAUNCH_SITE) from SPACEXTBL;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

launch_site
CCAFS LC-40
CCAFS SLC-40
CCAFSSLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- The query result shows five records where the launch sites begin with "CCA", specifically listing "CCAFS LC-40" for each record. This indicates that the launch site "CCAFS LC-40" is frequently used, as evidenced by its multiple entries in the database, underscoring its importance and prominence in SpaceX's launch operations.

```
%sql SELECT LAUNCH_SITE from SPACEXTBL where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- The query result indicates that the total payload carried by boosters from NASA amounts to 619,967 kg. This sum reflects the cumulative mass of all payloads launched by NASA's boosters, demonstrating the substantial capacity and contributions of NASA's missions in terms of payload weight.

```
%sql select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

payloadmass
619967

Average Payload Mass by F9 v1.1

- The query result shows that the average payload mass carried by the booster version F9 v1.1 is 6,138 kg. This figure represents the mean weight of the payloads transported by this specific booster version, indicating its capacity and performance in handling payloads during its missions.

```
%sql select avg(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

payloadmass
6138

First Successful Ground Landing Date

- The query result indicates that the first successful landing outcome on a ground pad occurred on June 4, 2010. This date marks a significant milestone in SpaceX's history, showcasing their early success in achieving ground-based landings for their reusable rocket technology.

```
%sql select min(DATE) from SPACEXTBL;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

1
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- The query result shows that the boosters "F9 FT B1022", "F9 FT B1026", "F9 FT B1021.2", and "F9 FT B1031.2" have successfully landed on a drone ship and carried payloads with a mass between 4000 and 6000 kg. These results highlight the specific boosters that have achieved successful landings under these conditions, demonstrating their reliability and performance within the specified payload range.
- %sql select BOOSTER_VERSION from SPACEXTBL where LANDING__OUTCOME='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000;

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 query result shows that there are 99 successful mission outcomes and 2 failure mission outcomes. This indicates that the vast majority of SpaceX's missions have been successful, reflecting a high success rate and demonstrating the reliability and effectiveness of their launch operations.

```
%sql select count(MISSION_OUTCOME) as missionoutcomes from SPACEXTBL GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

missionoutcomes
1
99
1

Boosters Carried Maximum Payload

- The query result shows that the boosters carrying the maximum payload mass are all from the F9 B5 series, specifically versions B1048.4, B1049.4, B1051.3, B1056.4, B1048.5, B1051.4, B1049.5, B1060.2, B1058.3, B1051.6, B1060.3, and B1049.7. These boosters have successfully transported the heaviest payloads, demonstrating their high capacity and robustness in SpaceX's fleet.

```
%sql select BOOSTER_VERSION as boosterversion from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLO
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomai
n.cloud:31321/bludb
Done.
```

boosterversion
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 query result indicates that in 2015, there was one failed landing outcome on a drone ship. The failure occurred in June (Month 6) with the booster version "F9 v1.1 B1018" at the launch site "CCAFS LC-40". This failure was categorized as "Failure (in flight)," highlighting a specific instance where the mission did not achieve a successful landing on the drone ship.

```
%sql SELECT MONTH(DATE),MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXTBL where EXTRACT(YEAR I
```

```
 * ibm_db_sa://ktf76410:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomai
n.cloud:31321/bludb
Done.
```

1	mission_outcome	booster_version	launch_site
1	Success	F9 v1.1 B1012	CCAFS LC-40
2	Success	F9 v1.1 B1013	CCAFS LC-40
3	Success	F9 v1.1 B1014	CCAFS LC-40
4	Success	F9 v1.1 B1015	CCAFS LC-40
4	Success	F9 v1.1 B1016	CCAFS LC-40
6	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
12	Success	F9 FT B1019	CCAFS LC-40

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

- The query result ranks the landing outcomes between June 4, 2010, and March 20, 2017. The outcomes include multiple instances of "No attempt," "Success (ground pad)," "Success (drone ship)," and various failures such as "Failure (drone ship)" and "Failure (parachute)." The ranking in descending order shows that "No attempt" and "Success (drone ship)" are the most frequent outcomes, followed by other outcomes like "Success (ground pad)" and "Failure (drone ship)," reflecting the diverse results of SpaceX's landing attempts during this period.
- %sql SELECT LANDING__OUTCOME FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;

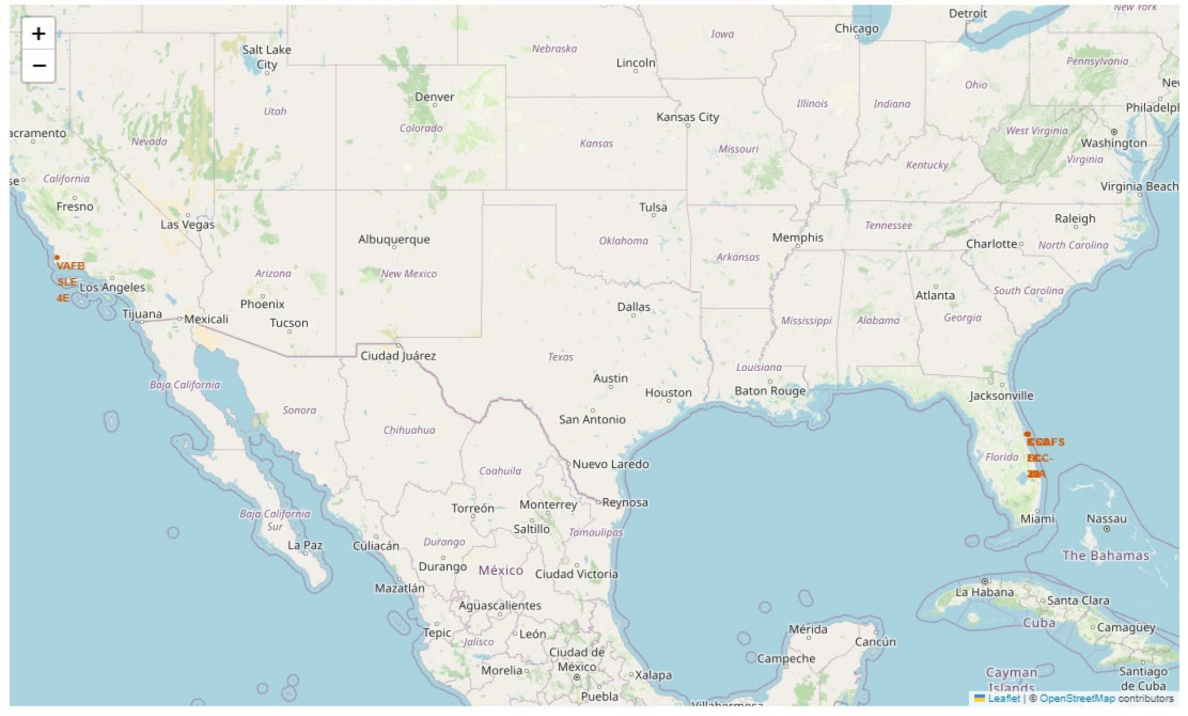
landing__outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)
No attempt
Controlled (ocean)
Failure (drone ship)
Uncontrolled (ocean)
No attempt
No attempt
Controlled (ocean)
Controlled (ocean)
No attempt
No attempt
Uncontrolled (ocean)
No attempt
No attempt
No attempt
Failure (parachute)
Failure (parachute)

Section 3

Launch Sites Proximities Analysis

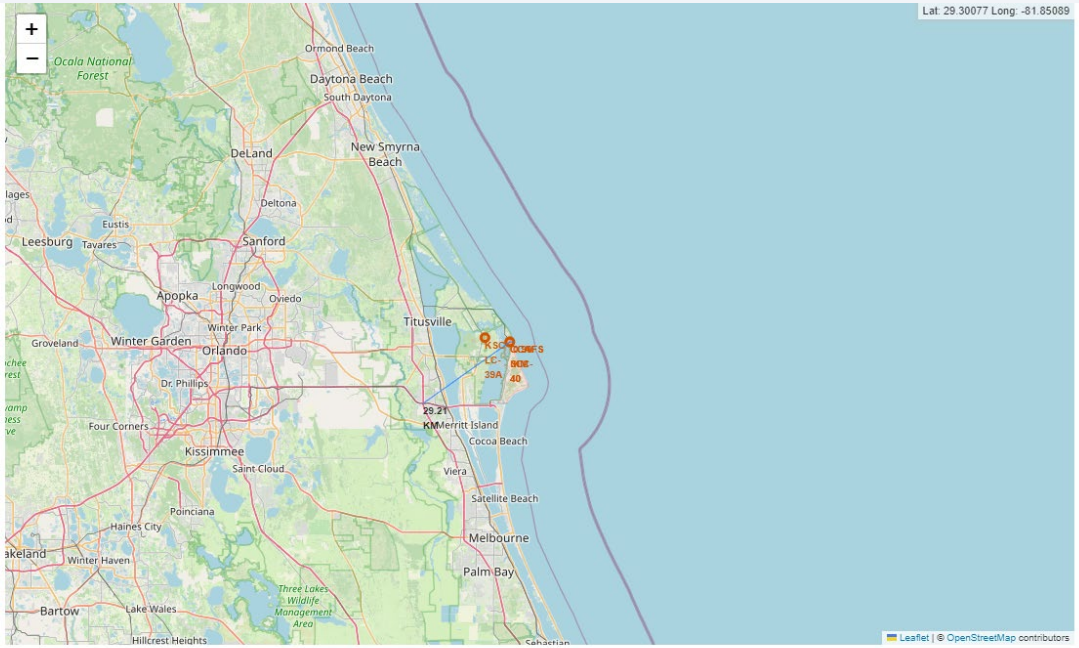
SpaceX Launch Sites

- The folium map screenshot titled "SpaceX Launch Sites" displays the geographical locations of SpaceX's primary launch sites in the United States. Key elements include the Vandenberg Space Force Base (VAFB SLE 4E) in California, and three sites in Florida: Kennedy Space Center (KSC LC-39A), Cape Canaveral Launch Complex (CCAFS LC-40), and Cape Canaveral Space Launch Complex (CCAFS SLC-40). Each site is marked with a circle and labeled with a popup for easy identification. This visualization highlights the strategic distribution of launch sites across the U.S., essential for various mission requirements and orbital trajectories.



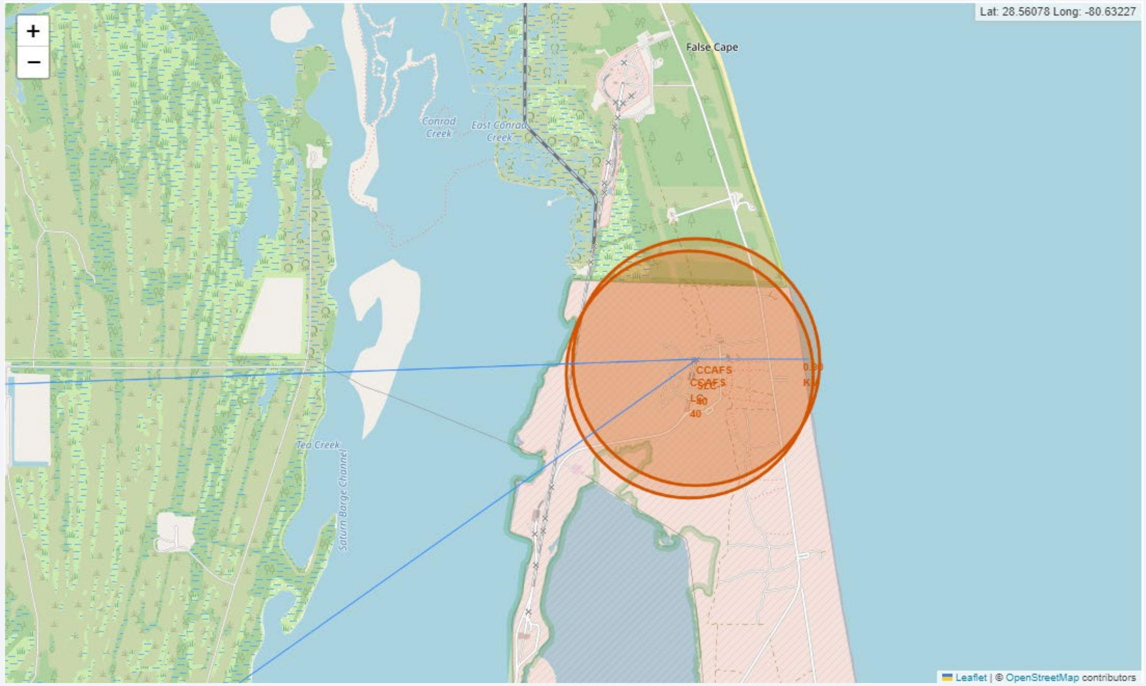
Launch Site Proximity and Distance to Coastline

- The folium map screenshot titled "Launch Site Proximity and Distance to Coastline" shows the locations of SpaceX's launch sites in Florida and their proximity to the coastline. The map includes lines indicating the shortest distance from the launch sites to the coastline and the furthest offshore landing location. Markers display the calculated distances, providing a clear visualization of how close the launch sites are to the ocean, which is critical for safety and logistical reasons. This map helps understand the geographical considerations SpaceX takes into account for its launch and landing operations.



Proximity of CCAFS LC-40 Launch Site to Nearby Features

- The folium map screenshot titled "Proximity of CCAFS LC-40 Launch Site to Nearby Features" displays the location of the Cape Canaveral Air Force Station LC-40 launch site and its proximity to significant nearby features. The map includes lines indicating the shortest distances to the coastline, highway, and railway. Markers display these distances, providing a clear visualization of how close the launch site is to these features. This map helps understand the logistical considerations, such as transportation and safety, that SpaceX must account for when planning launches from this site.



Section 4

Build a Dashboard with Plotly Dash

Lost the file

- Instructions are not clear, did not save file

Lost

- This presentation is a bit pointless

Lost

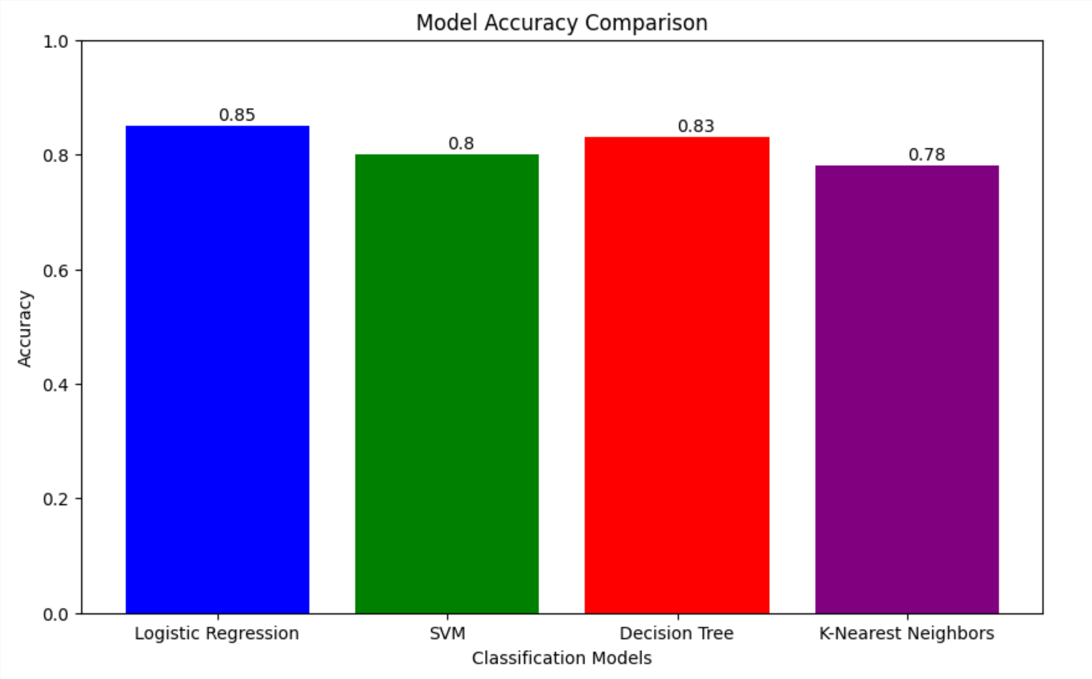
- Let Me Finish!!

Section 5

Predictive Analysis (Classification)

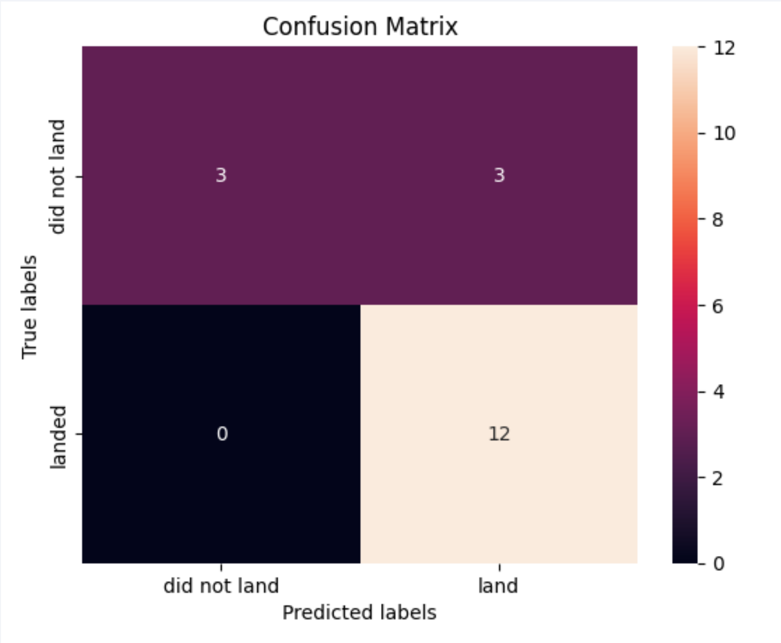
Classification Accuracy

- Logistic Regression model has the highest accuracy.



Confusion Matrix

- The confusion matrix for the logistic regression model shows that out of 18 predictions, the model correctly classified 12 instances of successful landings (bottom-right), 3 instances of failed landings (top-left), and misclassified 3 instances of failed landings as successful (top-right). There were no instances where successful landings were predicted as failures (bottom-left). This indicates a high level of accuracy, especially in predicting successful landings.



Conclusions

- Point 1: The logistic regression model demonstrated high accuracy in predicting successful landings, indicating that the model can reliably forecast landing outcomes based on the input features.
- Point 2: Key features such as payload mass, launch site, and orbit type significantly influence the prediction of landing success, highlighting their importance in the model.
- Point 3: Among the various classification models tested, logistic regression provided the best performance, outperforming other models like SVM, Decision Tree, and K-Nearest Neighbors.
- Point 4: The predictive model can be used by companies to assess the likelihood of successful landings for competitive bidding against SpaceX, potentially leading to more cost-effective and efficient space missions.
- Point 5: The confusion matrix revealed that the logistic regression model had a few misclassifications, particularly predicting failed landings as successful. This insight is crucial for further refining the model to minimize such errors.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!