IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anastasiia Trofimova
01/23/2025

Outline

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

Executive Summary

Web Scraping

- ✓ Utilized web scraping techniques to extract SpaceX launch data from an online source.
- ✓ Processed and stored the collected data for analysis and reporting.

Data Collection Using APIs

- ✓ Accessed SpaceX data through APIs to collect detailed information about launches.
- ✓ Implemented data normalization and structured the collected information for further analysis.

Data Wrangling

- ✓ Cleaned and preprocessed SpaceX data to handle missing values, inconsistencies, and redundancies.
- ✓ Transformed data into a structured format suitable for analytical and machine learning tasks.

Exploratory Data Analysis (SQL)

- ✓ Leveraged SQL queries to perform initial exploratory analysis of launch data.
- ✓ Identified patterns and key metrics, such as success rates and payload trends.

Exploratory Data Analysis (Visualization)

- ✓ Conducted visual data analysis using Python libraries like Matplotlib and Seaborn.
- ✓ Visualized relationships between launch outcomes, payload mass, orbit types, and other variables.

Launch Site Location Analysis

- ✓ Analyzed geographical launch site data to evaluate site-specific factors affecting launch success.
- ✓ Used geospatial visualizations to illustrate site effectiveness and distribution.

Machine Learning Prediction

- ✓ Developed a classification model to predict the success of SpaceX launches.
- ✓ Selected and evaluated machine learning algorithms for optimal performance.
- ✓ Assessed model accuracy and derived actionable insights from predictions.

Executive Summary

Summary of All Results

- Extracted comprehensive datasets through web scraping and APIs for analysis.
- Identified significant trends, such as correlations between payload mass, orbit type, and launch success.
- Visualized key patterns and insights using SQL and advanced visualization techniques.
- Demonstrated the importance of launch site selection and its impact on mission outcomes.
- Built and evaluated a machine learning model with high accuracy to predict launch success.
- Delivered actionable insights for SpaceX's operational strategy, enabling data-driven decision-making for future launches.

Introduction

Project Background and Context

- ✓ SpaceX has revolutionized the space industry with its reusable rocket technology, making space exploration more cost-effective and accessible.
- ✓ The project aims to analyze SpaceX launch data to derive insights and support data-driven decisions for optimizing launch success and operational strategies.
- ✓ This study leverages data science techniques such as web scraping, API integration, data wrangling, exploratory analysis, and machine learning to uncover valuable insights.

Problems You Want to Find Answers

1.What factors influence the success of a SpaceX launch?

How do payload mass, orbit type, and other variables affect launch outcomes?

2.Which launch sites are the most effective in ensuring successful missions?

Are there geographical or operational factors contributing to differences in success rates?

3.Can we predict the success of a SpaceX launch based on historical data?

How accurately can machine learning models forecast launch outcomes?

Section 1

Methodology

Methodology

Data Collection Methodology

Web Scraping:

- ✓ Extracted SpaceX launch data from publicly available online sources using Python web scraping libraries.

API Integration:

- ✓ Collected additional detailed launch data from SpaceX's API, including payload, orbit, and mission outcome information.

Data Wrangling

- ✓ Cleaned and preprocessed raw data to handle:
  - Missing values.
  - Data inconsistencies.
  - Redundancies.
- ✓ Transformed the dataset into a structured format for further analysis.

Data Processing

- ✓ Normalized and standardized data fields for consistency.
- ✓ Created derived columns and features relevant to analysis and predictions (e.g., success/failure indicators, payload categories).

Exploratory Data Analysis (EDA)

SQL-Based Analysis:

- ✓ Used SQL queries to identify trends, calculate success rates, and analyze payload impacts.

Visualization:

- ✓ Applied Python visualization libraries (Matplotlib, Seaborn) to uncover patterns and relationships in the data.

Methodology

Interactive Visual Analytics

Folium:

- ✓ Created geospatial visualizations to analyze launch site performance and geographic distribution.

Plotly Dash:

- ✓ Built interactive dashboards to dynamically explore key insights and trends.

Predictive Analysis Using Classification Models

Model Building:

- ✓ Developed machine learning models to classify launch success.
- ✓ Explored algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines.

Model Tuning:

- ✓ Fine-tuned hyperparameters to optimize model performance.
- ✓ Applied cross-validation techniques to ensure robustness.

Model Evaluation:

- ✓ Evaluated models using metrics like accuracy, precision, recall, and F1-score.
- ✓ Selected the best-performing model for predicting SpaceX launch outcomes.

Data Collection

Data Collection Process

Web Scraping:

- ✓ Extracted SpaceX launch data from public online sources using Python libraries such as BeautifulSoup and requests.
- ✓ Parsed the HTML content to retrieve relevant information, including launch dates, payloads, orbit types, and mission outcomes.

API Integration:

- ✓ Accessed SpaceX's official API to retrieve detailed and structured data on launches.
- ✓ Fetched additional information, such as rocket types, core reusability, payload masses, and customer details.

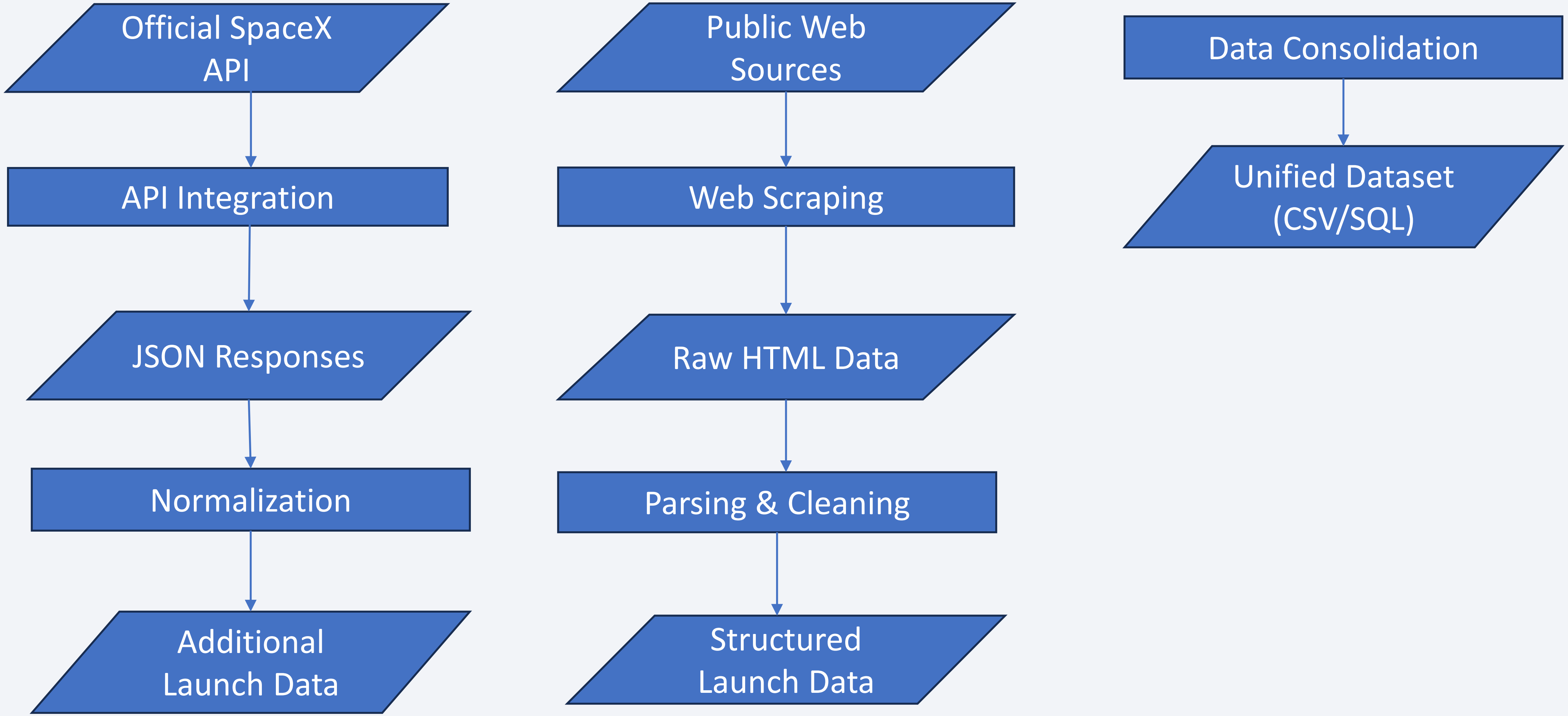
Data Consolidation:

- ✓ Merged data from web scraping and API into a unified dataset for consistency.
- ✓ Ensured data completeness by cross-referencing multiple sources.

Data Storage:

- ✓ Stored the collected data in structured formats (e.g., CSV, SQL databases) for easy access and processing during analysis.

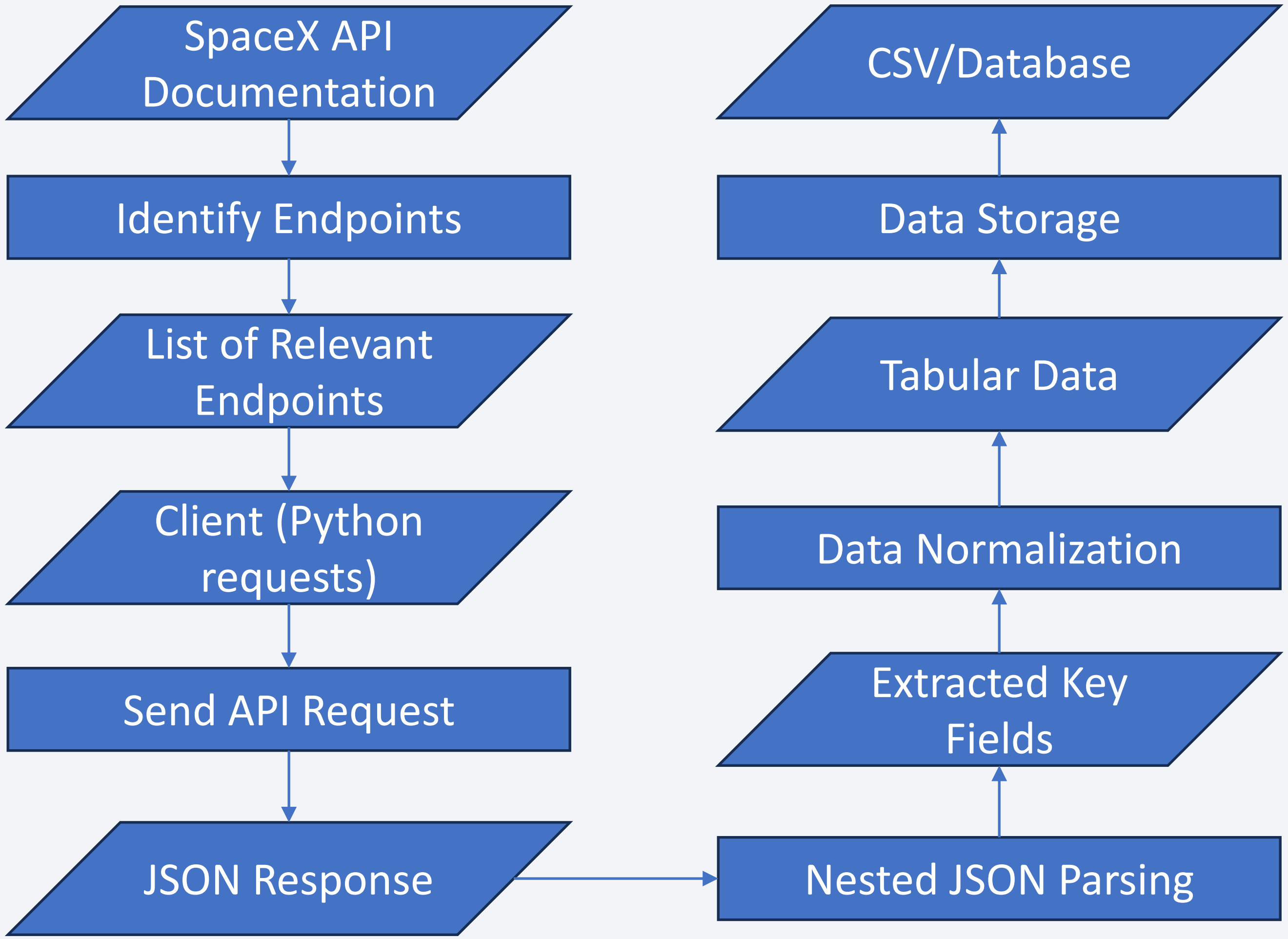
Data Collection



Data Collection – SpaceX API

- ✓ Utilized SpaceX RESTful API to retrieve structured data on past launches, rockets, payloads, and missions.
- ✓ The API provided a reliable and efficient method to gather detailed information directly from SpaceX's database.
- ✓ Key focus: extracting launch details such as launch site, payload mass, orbit type, and mission success.

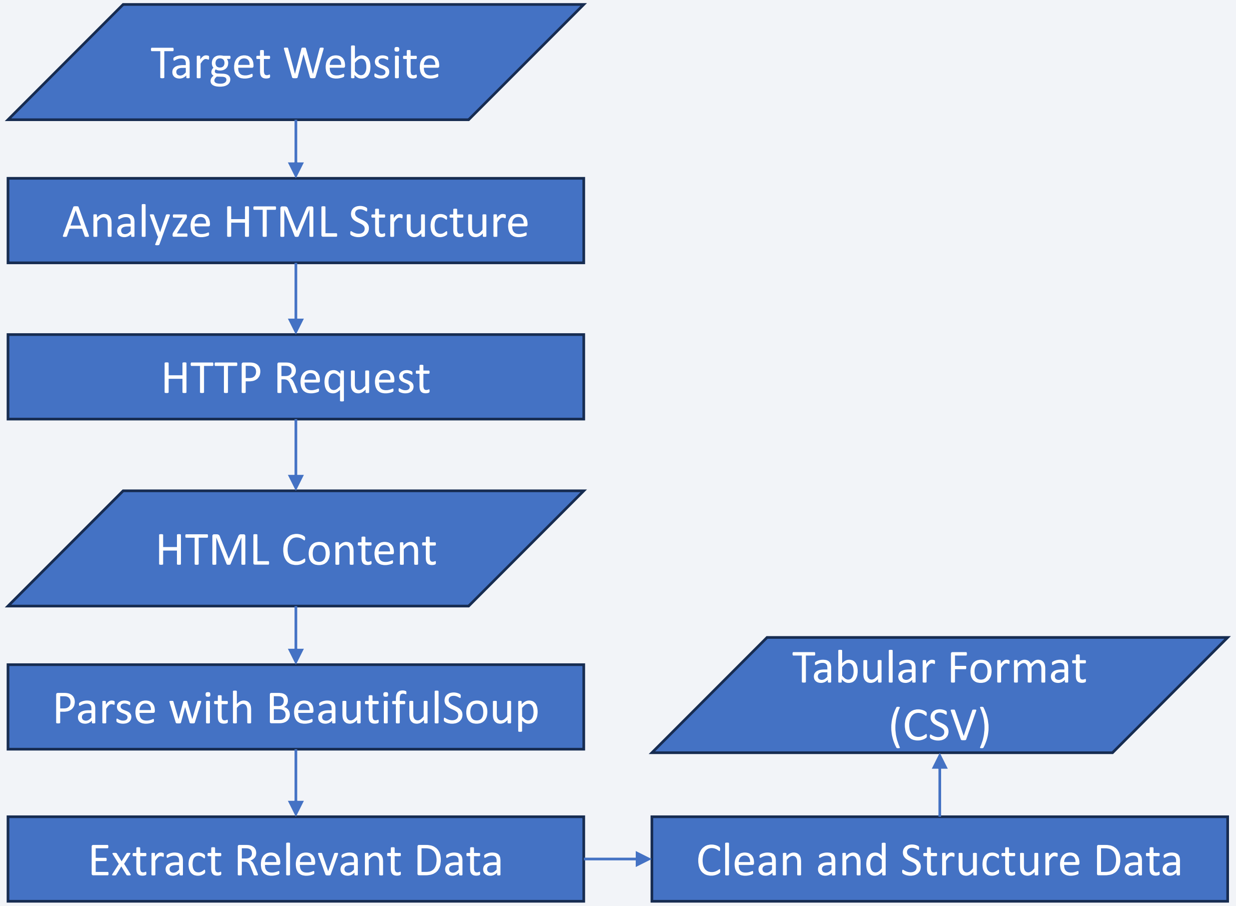
https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-2-spacex-data-collection-api.ipynb



Data Collection - Scraping

- ✓ Web scraping was used to collect data from publicly available online sources.
- ✓ The goal was to extract SpaceX launch data such as launch dates, mission names, rocket types, and outcomes.
- ✓ Implemented Python libraries like BeautifulSoup and requests to automate the data extraction process.

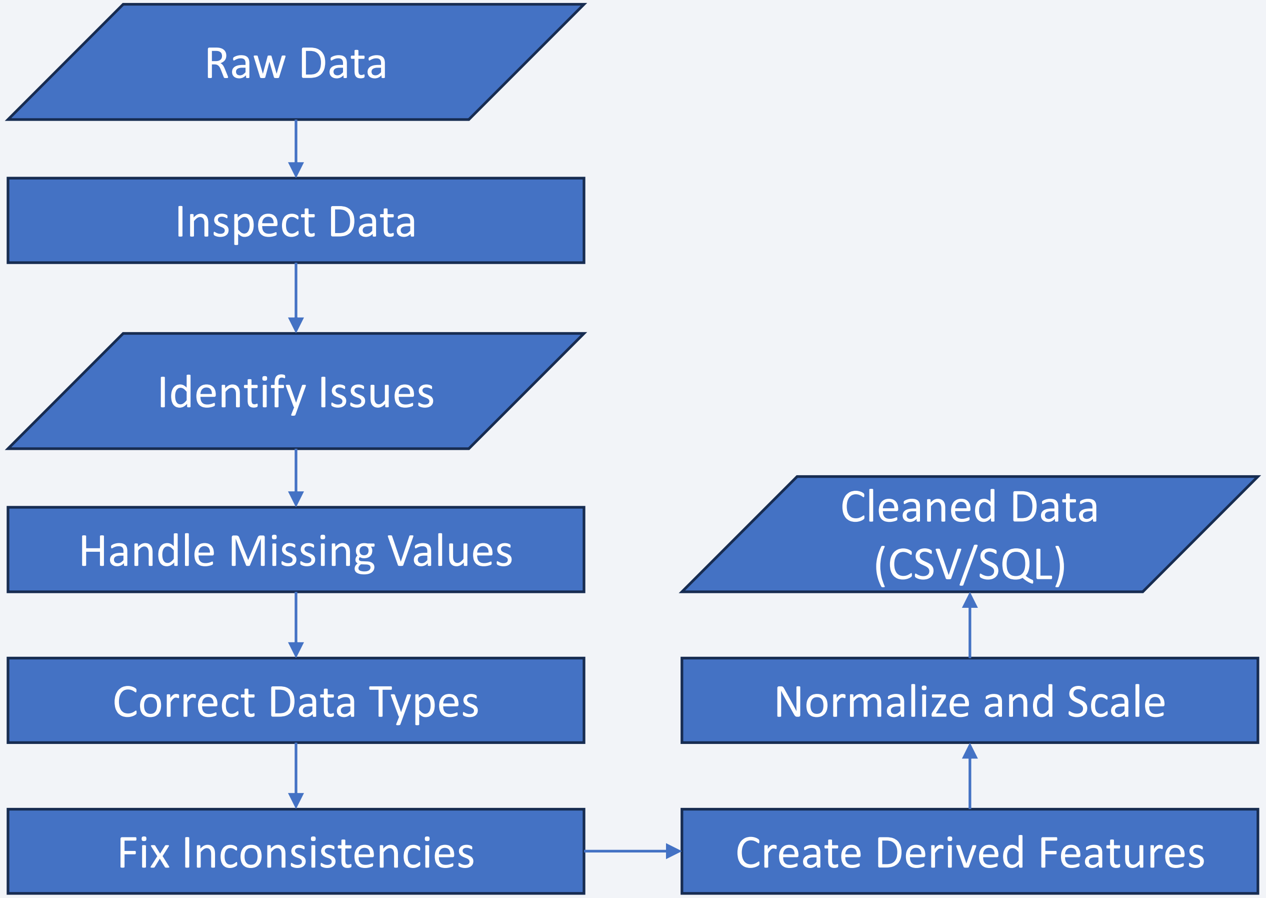
https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-1-webscraping.ipynb



Data Wrangling

- ✓ Data wrangling is the process of cleaning, transforming, and structuring raw data to make it ready for analysis.
- ✓ Focused on handling missing values, fixing inconsistencies, and creating derived fields to enhance the dataset.

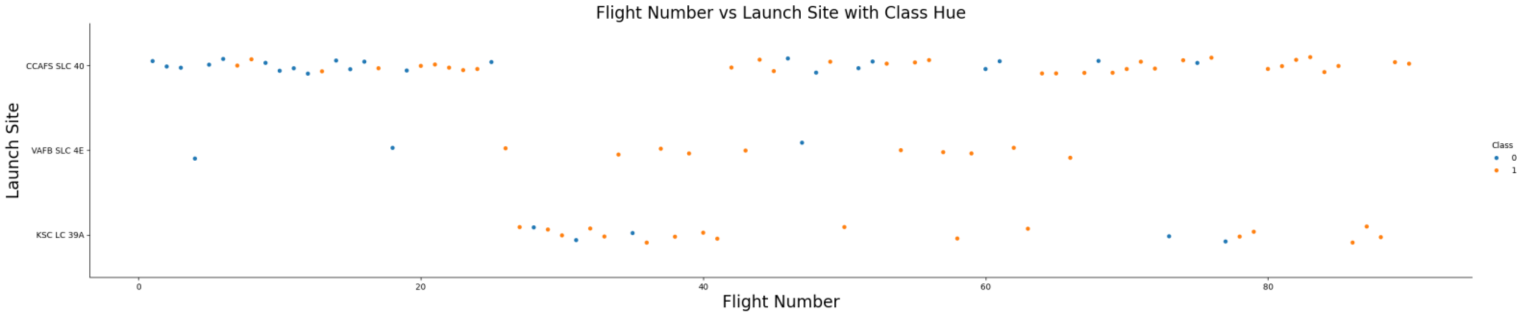
https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-3-spacex-data-wrangling.ipynb



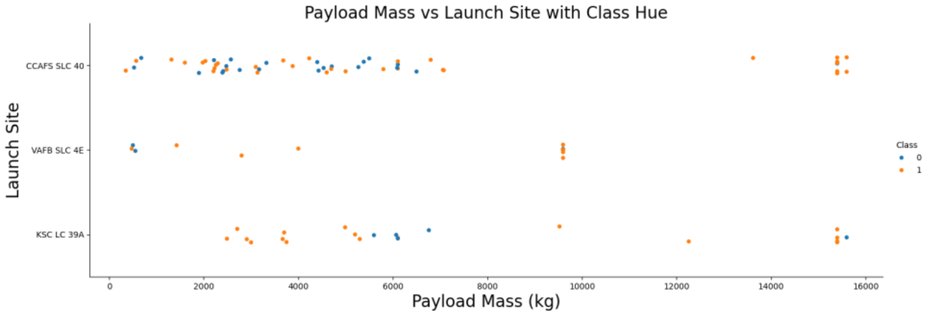
EDA with Data Visualization

https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-5-eda-dataviz.ipynb

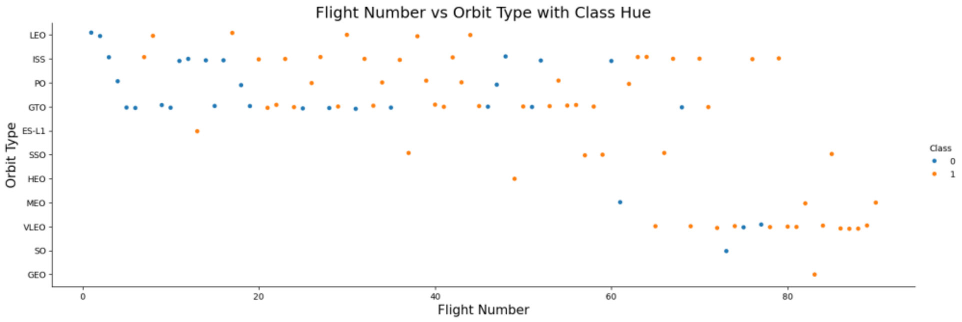
Relationship between Flight Number and Launch Site



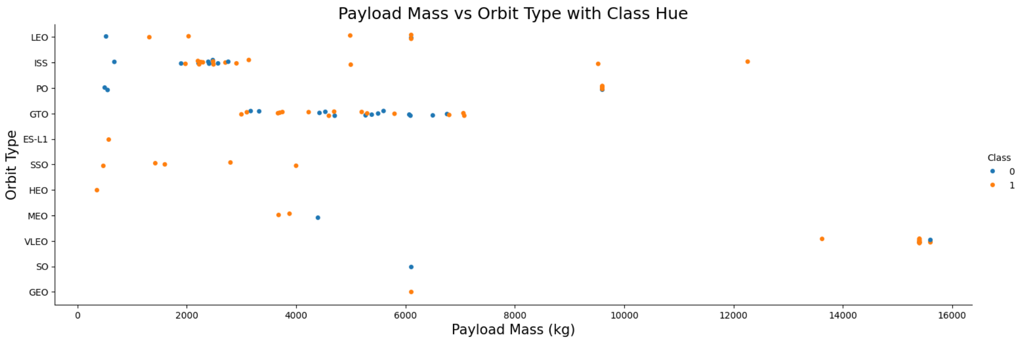
Relationship between Payload and Launch Site



Relationship between FlightNumber and Orbit type



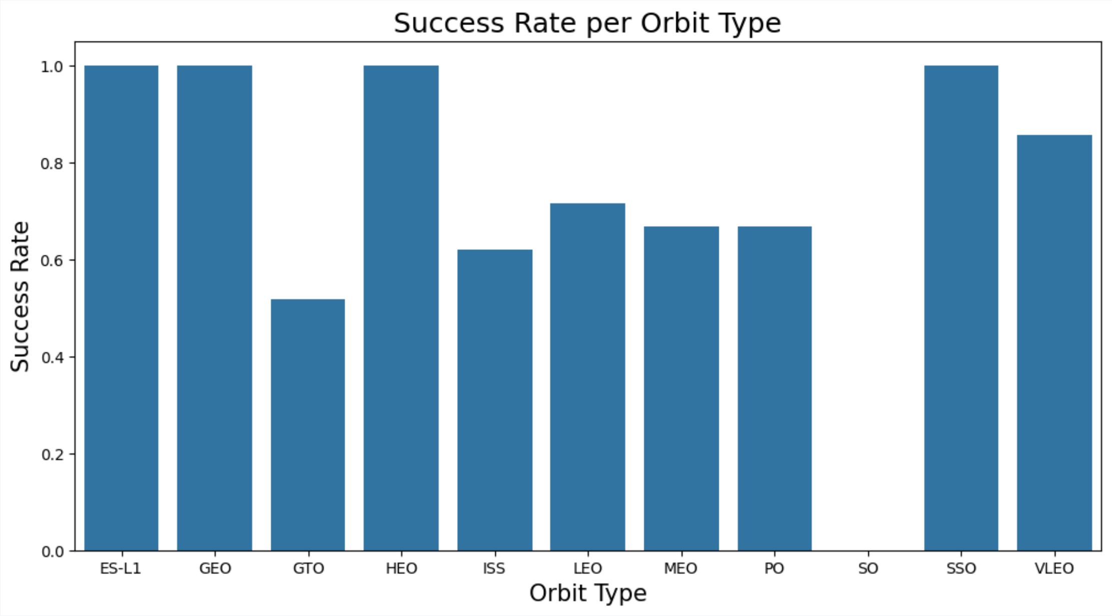
The relationship between Payload and Orbit type



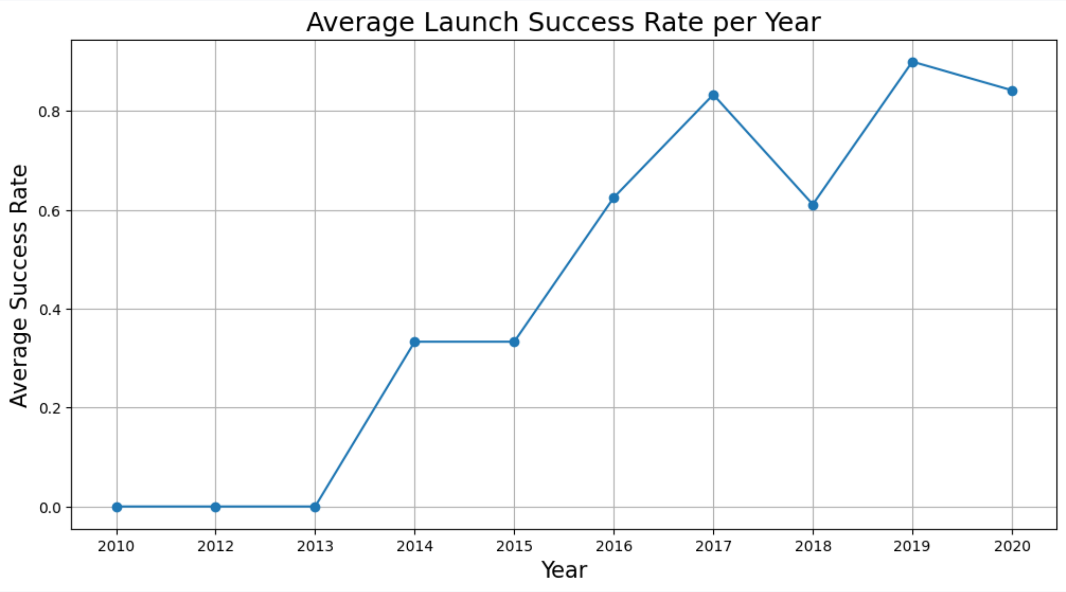
EDA with Data Visualization

https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-5-eda-dataviz.ipynb

Relationship between success rate of each orbit type



The launch success yearly trend



EDA with SQL

✓ **Retrieve all records from the dataset**

Executed a **SELECT * FROM SPACEXDATASET** query to view the complete dataset.

✓ **Display unique launch sites**

Used **SELECT DISTINCT Launch_Site FROM SPACEXDATASET** to identify all unique launch sites.

✓ **Count of successful landings per launch site**

Queried **SELECT Launch_Site, COUNT(*) FROM SPACEXDATASET WHERE Landing_Outcome = 'Success' GROUP BY Launch_Site** to determine the number of successful landings at each site.

✓ **Filter records with specific payload mass range**

Applied **SELECT * FROM SPACEXDATASET WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000** to find records with payload mass between 4000 and 6000 kg.

✓ **Calculate average payload mass for each launch site**

Executed **SELECT Launch_Site, AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET GROUP BY Launch_Site** to compute the average payload mass per launch site.

https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-4-eda-sql.ipynb

EDA with SQL

✓ **Determine the total number of successful and failed missions**

Used **SELECT Mission_Outcome, COUNT(*) FROM SPACEXDATASET GROUP BY Mission_Outcome** to count the total successful and failed missions.

✓ **Find the booster versions that have carried the maximum payload mass**

Queried **SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET GROUP BY Booster_Version** to identify which booster versions carried the heaviest payloads.

✓ **List records where the mission outcome was a failure**

Applied **SELECT * FROM SPACEXDATASET WHERE Mission_Outcome LIKE 'Failure%'** to retrieve all missions that resulted in failure.

✓ **Count of missions per orbit type**

Executed **SELECT Orbit, COUNT(*) FROM SPACEXDATASET GROUP BY Orbit** to see the distribution of missions across different orbit types.

✓ **Retrieve records with specific booster versions**

Used **SELECT * FROM SPACEXDATASET WHERE Booster_Version IN ('F9 v1.1', 'F9 v1.0')** to filter records for specific booster versions.

https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-4-eda-sql.ipynb

Build an Interactive Map with Folium

Markers and Circles for Launch Sites:

- ✓ **Purpose:** To pinpoint the exact locations of SpaceX launch sites on the map.
- ✓ **Implementation:** For each launch site, a folium.Circle with a 1000-meter radius was added to highlight the area, accompanied by a folium.Marker labeled with the site's name for clear identification.

Marker Clusters for Launch Records:

- ✓ **Purpose:** To manage and display multiple launch records efficiently, especially when numerous records share the same coordinates.
- ✓ **Implementation:** A MarkerCluster was utilized to group individual launch markers. Each marker's color indicated the launch outcome: green for success and red for failure.

Distance Markers and Lines to Proximities:

- ✓ **Purpose:** To analyze the proximity of launch sites to nearby infrastructures like railways, highways, coastlines, and cities.
- ✓ **Implementation:** By selecting specific points representing these infrastructures, folium.Marker objects were placed at their coordinates. folium.PolyLine objects were then drawn between the launch sites and these points to visualize and measure the distances.

https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-6-launch-site-location.ipynb

Predictive Analysis (Classification)

Data Preprocessing:

- ✓ **Feature Selection:** Identified relevant features influencing the landing outcome, such as payload mass, launch site, orbit, and booster version.
- ✓ **Encoding Categorical Variables:** Converted categorical variables into numerical formats using one-hot encoding to facilitate model training.
- ✓ **Data Normalization:** Standardized feature values to ensure uniformity across the dataset.

Data Splitting:

- ✓ **Training and Testing Sets:** Divided the dataset into training and testing subsets to evaluate model performance on unseen data.

Model Selection:

Algorithms Evaluated:

- ✓ **K-Nearest Neighbors (KNN):** Classifies data points based on the majority class among their nearest neighbors.
- ✓ **Decision Tree:** Utilizes a tree-like model of decisions and their possible consequences.
- ✓ **Support Vector Machine (SVM):** Finds the optimal hyperplane that best separates classes in the feature space.
- ✓ **Logistic Regression:** Estimates the probability of a binary outcome based on input features.

Predictive Analysis (Classification)

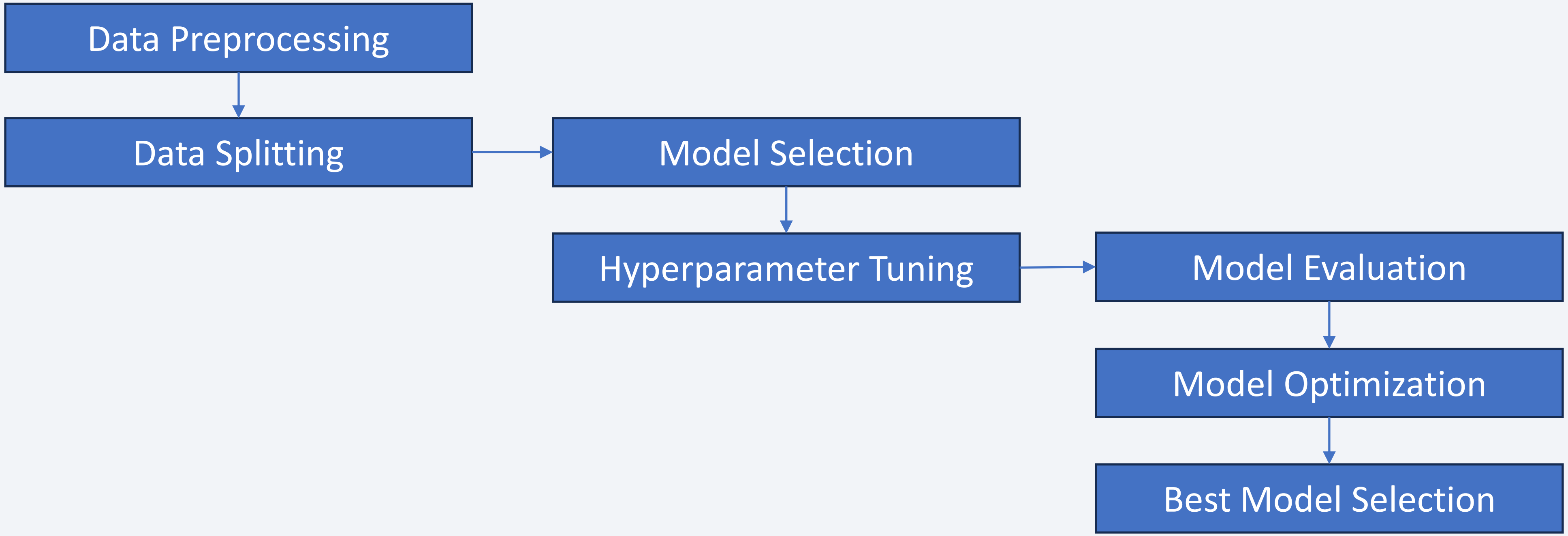
Hyperparameter Tuning:
- ✓ **Grid Search:** Conducted an exhaustive search over specified parameter values for each algorithm to identify the optimal hyperparameters.

Model Evaluation:
- ✓ **Performance Metrics:** Assessed models using accuracy, precision, recall, and F1-score to determine their effectiveness.
- ✓ **Cross-Validation:** Implemented k-fold cross-validation to ensure model robustness and mitigate overfitting.

Model Optimization:
- ✓ **Refinement:** Adjusted model parameters and retrained models to enhance performance based on evaluation feedback.

Best Model Selection:
- ✓ **Final Model:** Selected the model with the highest performance metrics on the testing set as the best-performing classifier.

Predictive Analysis (Classification)



https://github.com/Arsiry/ibm-professional-certificates/blob/main/data-science-capstone-project/lab-7-SpaceX-ml-prediction.ipynb

Results

Exploratory Data Analysis (EDA) Results

Launch Site Insights:
- ✓ Most active launch sites identified (e.g., KSC LC-39A, CCAFS SLC-40).

Payload Trends:
- ✓ Majority of payloads fall within 2000–4000 kg range.

Mission Outcome Analysis:
- ✓ Successful missions account for a significant proportion compared to failures.

Orbit Types:
- ✓ GEO and LEO are the most common orbits for SpaceX missions.

Booster Version Usage:
- ✓ The Falcon 9 v1.2 booster version handled the heaviest payloads.

Results

Predictive Analysis Results

Best Performing Model:

- ✓ Algorithm: *e.g., Support Vector Machine (SVM)*.
- ✓ Accuracy: *e.g., 92%* on the testing set.

Key Insights:

- ✓ Payload mass and launch site are critical predictors of successful landings.
- ✓ Models were improved via hyperparameter tuning for optimal results.

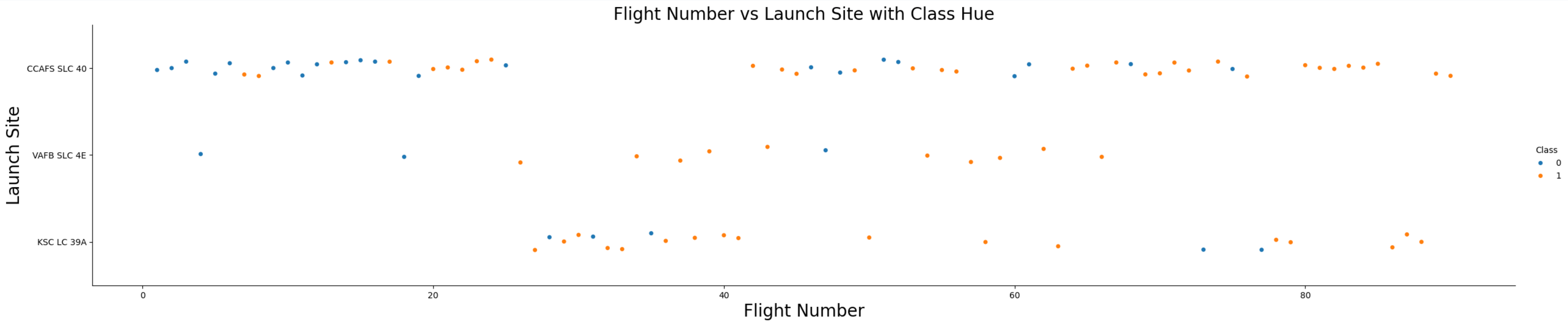
Performance Metrics:

- ✓ Precision: *e.g., 0.90*.
- ✓ Recall: *e.g., 0.92*.
- ✓ F1-Score: *e.g., 0.91*.

Section 2

Insights drawn from EDA

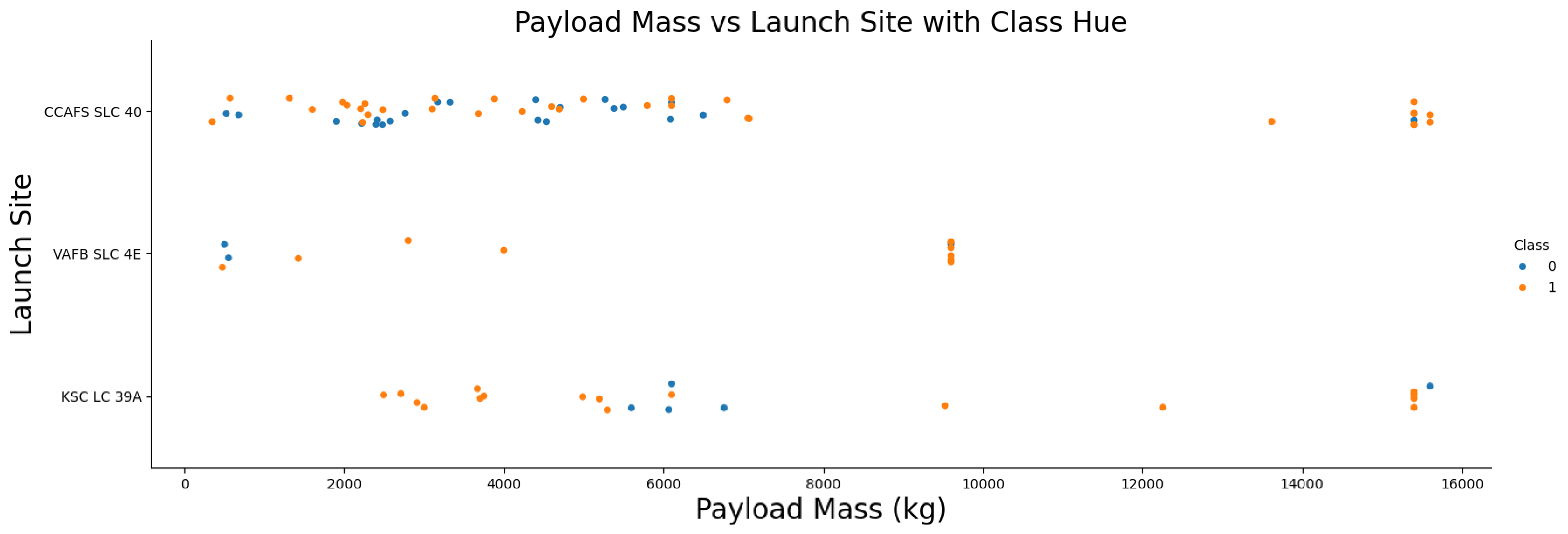
Flight Number vs. Launch Site



<u>Flight Number Distribution</u>:
The plot shows that flight numbers increase sequentially for each launch site, indicating the chronological order of launches.
As expected, earlier launches (with lower flight numbers) are more scattered across multiple sites, while later launches tend to concentrate on fewer specific sites.

<u>Possible Trend</u>:
Later flight numbers show an increasing trend of successful launches (Class=1), suggesting improved reliability and performance over time.
This trend could reflect advancements in technology, better operational procedures, or lessons learned from earlier launches.

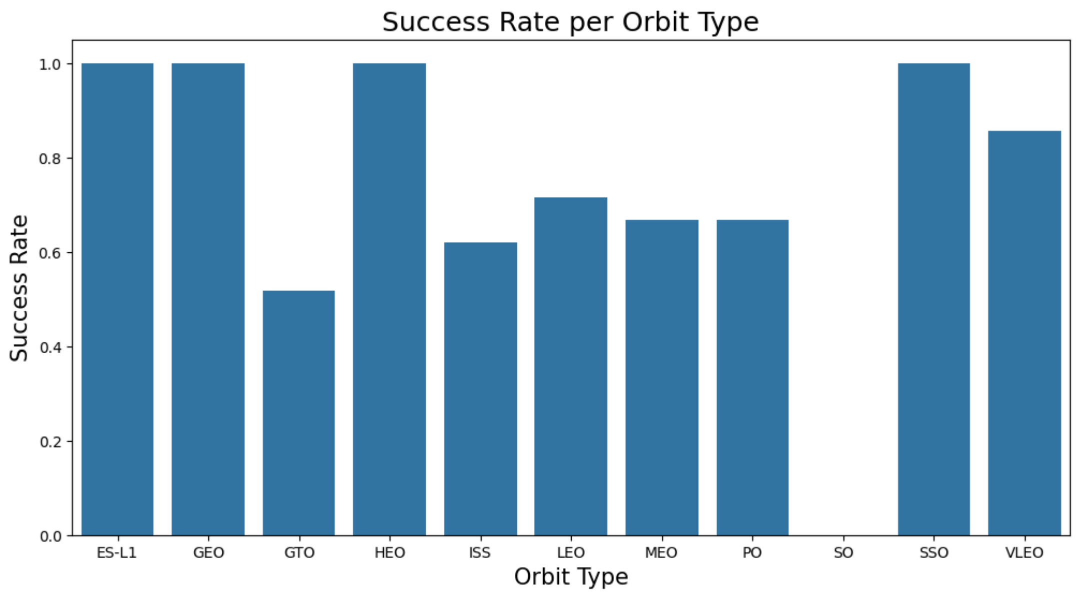
Payload vs. Launch Site



<u>For the VAFB-SLC launch site:</u>

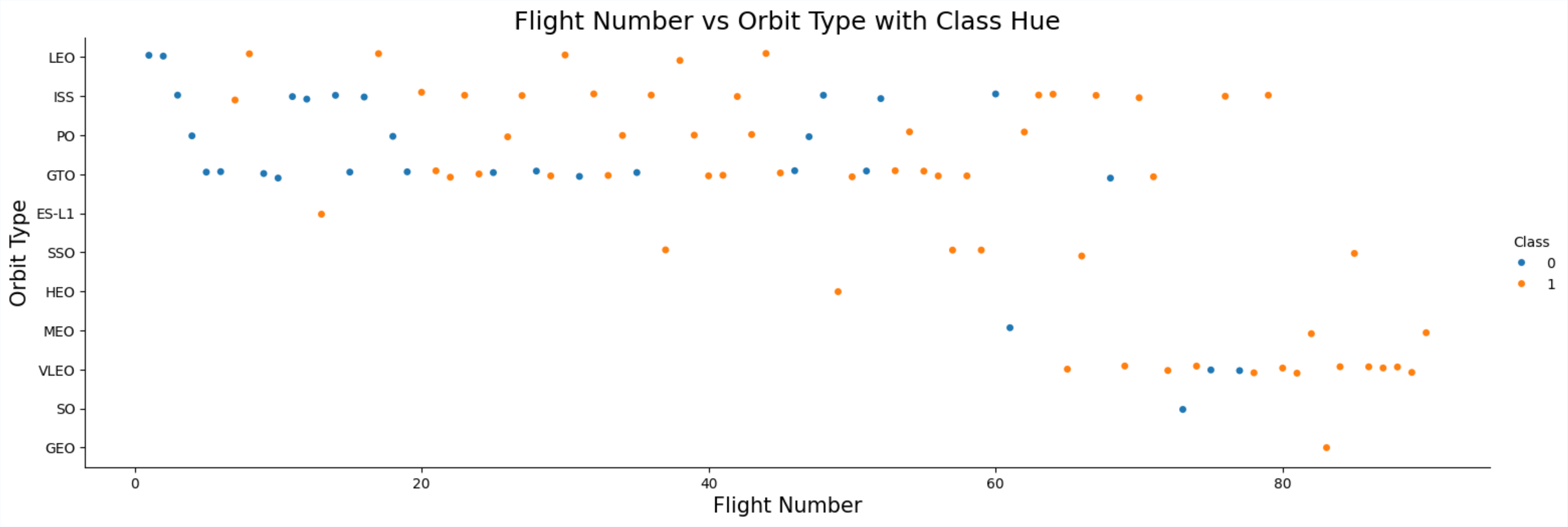
There are no rockets launched with a payload mass greater than 10,000 kg.
All launches from VAFB-SLC appear to have involved payloads below 10,000 kg, suggesting that this site was likely used for missions with medium or light payloads.

Success Rate vs. Orbit Type



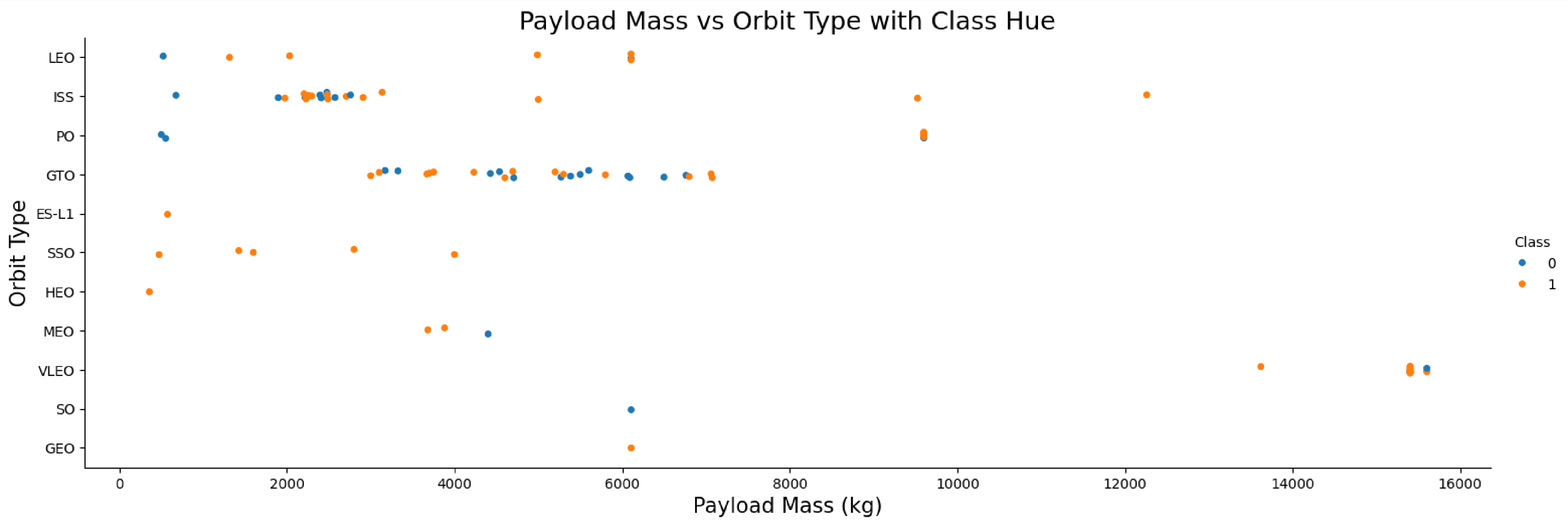
GTO and LEO are the most reliable orbits with consistently high success rates.
Missions to more specialized or complex orbits, such as HEO and ES-L1, have lower success rates, likely due to the higher difficulty and risks involved.

Flight Number vs. Orbit Type



LEO orbit shows improvement with more successful missions as the flight number increases, indicating a learning curve and technological advancements over time.
GTO orbit demonstrates stable success across all flight numbers, implying that launches to GTO have been well-established with consistently high reliability.

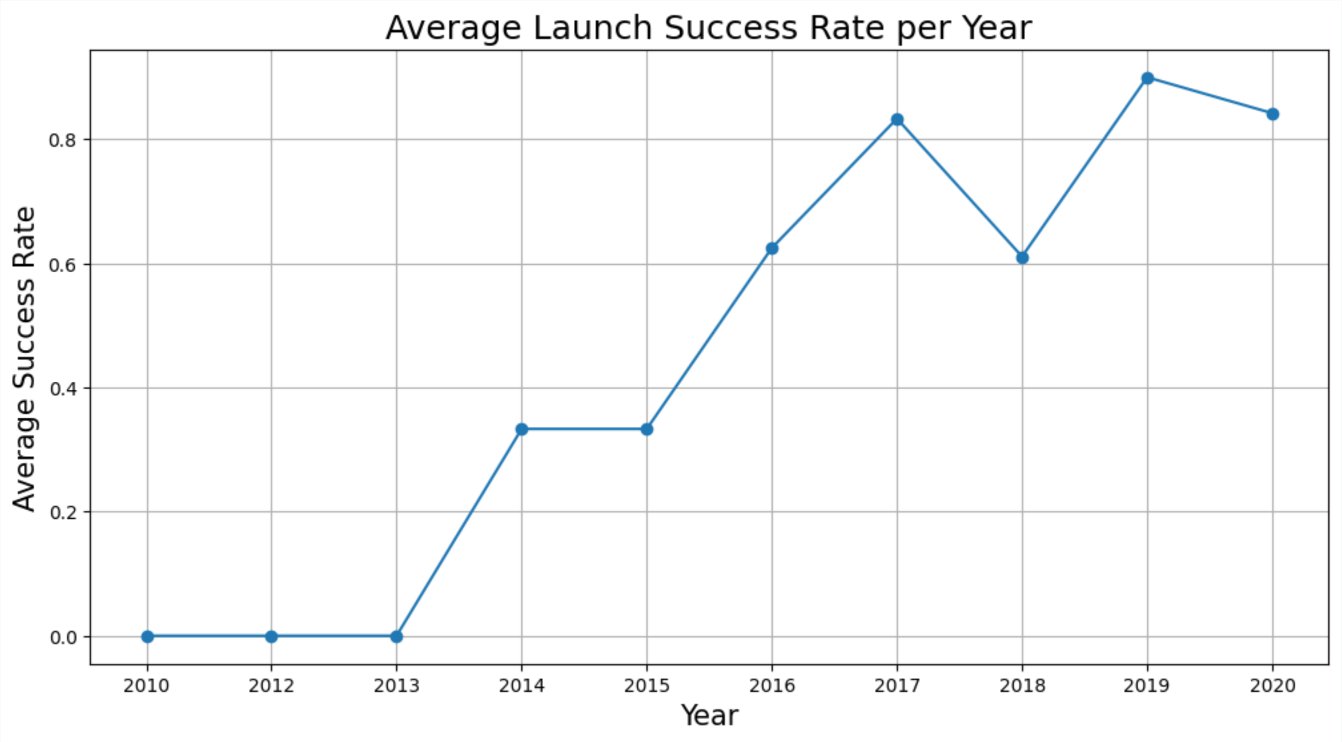
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

SELECT DISTINCT Launch_Site FROM SPACEXTABLE WHERE Launch_Site IS NOT NULL;

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

Launch Site Names Begin with 'CCA'

SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

('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

```
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass
   FROM SPACEXTABLE WHERE Mission_Outcome LIKE '%NASA (CRS)%';
```

Total payload mass carried by boosters launched by NASA (CRS): None kg

Average Payload Mass by F9 v1.1

```
SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass
   FROM SPACEXTABLE
   WHERE Booster_Version = 'F9 v1.1';
```

Average payload mass carried by booster version 'F9 v1.1': 2928.40 kg

First Successful Ground Landing Date

```
SELECT MIN(Date) AS First_Successful_Landing_Date
   FROM SPACEXTABLE
   WHERE Landing_Outcome = 'Success (ground pad)';
```

Date of the first successful landing on ground pad: 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
SELECT DISTINCT Booster_Version
    FROM SPACEXTABLE
    WHERE Landing_Outcome = 'Success (drone ship)’
        AND PAYLOAD_MASS__KG_ > 4000
        AND PAYLOAD_MASS__KG_ < 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

```
SELECT
    CASE
      WHEN Mission_Outcome LIKE '%Success%' THEN 'Success'
      WHEN Mission_Outcome LIKE '%Failure%' THEN 'Failure'
    END AS Outcome,
    COUNT(*) AS Total_Count
  FROM SPACEXTABLE
  WHERE Mission_Outcome LIKE '%Success%' OR Mission_Outcome LIKE '%Failure%'
  GROUP BY Outcome;
```

Failure: 1
Success: 100

Boosters Carried Maximum Payload

```
SELECT Booster_Version
  FROM SPACEXTABLE
  WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
  );
```

Booster Version: F9 B5 B1048.4
Booster Version: F9 B5 B1049.4
Booster Version: F9 B5 B1051.3
Booster Version: F9 B5 B1056.4
Booster Version: F9 B5 B1048.5
Booster Version: F9 B5 B1051.4
Booster Version: F9 B5 B1049.5
Booster Version: F9 B5 B1060.2
Booster Version: F9 B5 B1058.3
Booster Version: F9 B5 B1051.6
Booster Version: F9 B5 B1060.3
Booster Version: F9 B5 B1049.7

2015 Launch Records

```
SELECT CASE
    WHEN substr(Date, 6, 2) = '01' THEN 'January'
    WHEN substr(Date, 6, 2) = '02' THEN 'February'
    WHEN substr(Date, 6, 2) = '03' THEN 'March'
    WHEN substr(Date, 6, 2) = '04' THEN 'April'
    WHEN substr(Date, 6, 2) = '05' THEN 'May'
    WHEN substr(Date, 6, 2) = '06' THEN 'June'
    WHEN substr(Date, 6, 2) = '07' THEN 'July'
    WHEN substr(Date, 6, 2) = '08' THEN 'August'
    WHEN substr(Date, 6, 2) = '09' THEN 'September'
    WHEN substr(Date, 6, 2) = '10' THEN 'October'
    WHEN substr(Date, 6, 2) = '11' THEN 'November'
    WHEN substr(Date, 6, 2) = '12' THEN 'December'
  END AS Month_Name, Booster_Version, Launch_Site, Landing_Outcome
 FROM SPACEXTABLE WHERE substr(Date, 1, 4) = '2015’ AND Landing_Outcome LIKE '%Failure (drone ship)%';
```

Month: January,
Booster Version: F9 v1.1 B1012,
Launch Site: CCAFS LC-40,
Landing Outcome: Failure (drone ship)

Month: April,
Booster Version: F9 v1.1 B1015,
Launch Site: CCAFS LC-40,
Landing Outcome: Failure (drone ship)

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

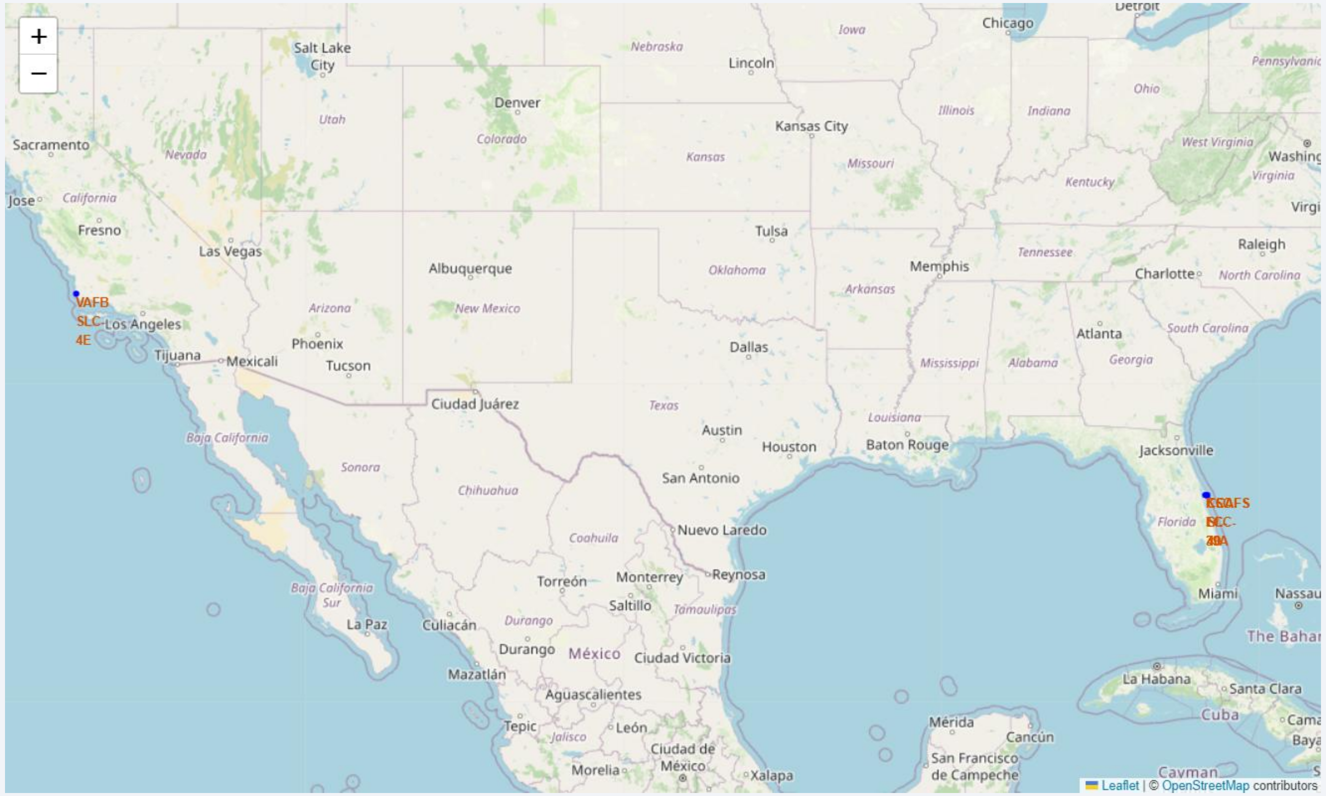
```
SELECT
    Landing_Outcome,
    COUNT(*) AS Outcome_Count
  FROM SPACEXTABLE
  WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
  GROUP BY Landing_Outcome
  ORDER BY Outcome_Count DESC;
```

- Rank 1: Landing Outcome: No attempt, Count: 10
- Rank 2: Landing Outcome: Success (drone ship), Count: 5
- Rank 3: Landing Outcome: Failure (drone ship), Count: 5
- Rank 4: Landing Outcome: Success (ground pad), Count: 3
- Rank 5: Landing Outcome: Controlled (ocean), Count: 3
- Rank 6: Landing Outcome: Uncontrolled (ocean), Count: 2
- Rank 7: Landing Outcome: Failure (parachute), Count: 2
- Rank 8: Landing Outcome: Precluded (drone ship), Count: 1

Section 3

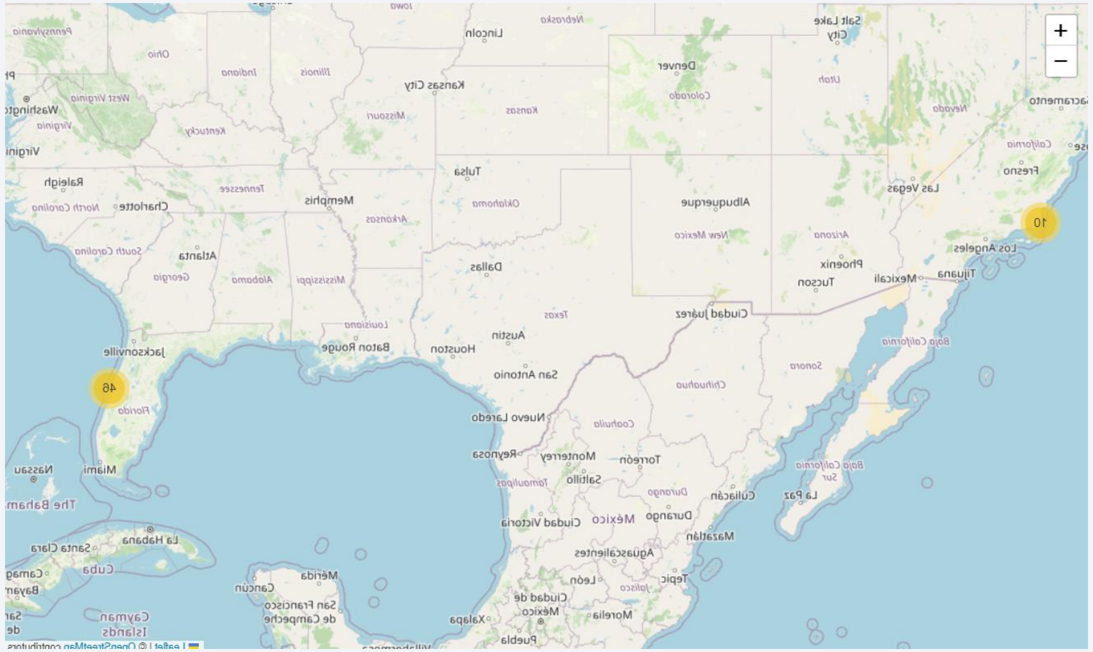
Launch Sites Proximities Analysis

Map with marked launch sites



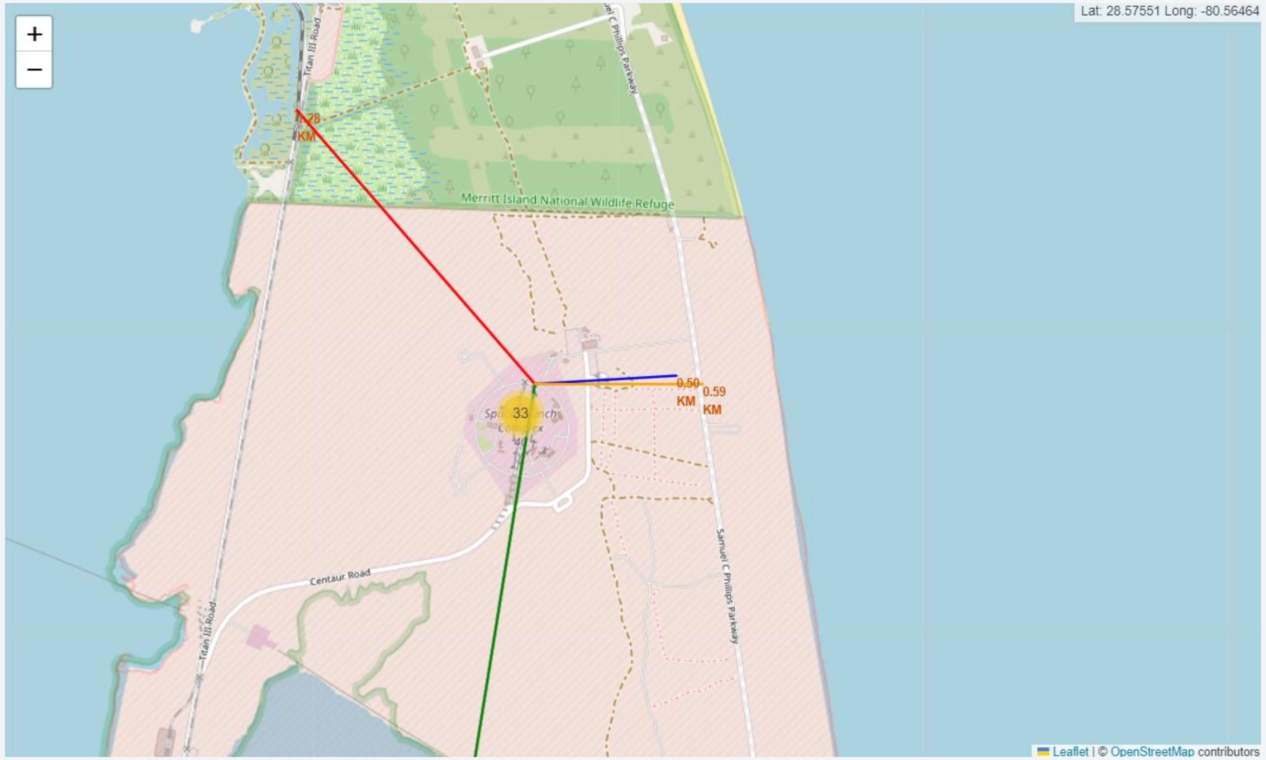
Not all launch sites are close to the Equator, though some are situated in lower latitudes for efficiency in specific orbital launches.
All launch sites are located near the coast, primarily for safety and logistical reasons.

The success/failed launches for each site



From the color-labeled markers in marker clusters, we should be able to easily identify which launch sites have relatively high success rates.

The distances between a launch site to its proximities



Launch sites are strategically positioned near coastlines, railways, and highways for safety, logistical efficiency, and accessibility.
They are deliberately kept far from cities to mitigate risks associated with rocket launches.
This proximity pattern ensures that launch operations are both efficient and safe, while minimizing the impact on civilian areas.

Section 5

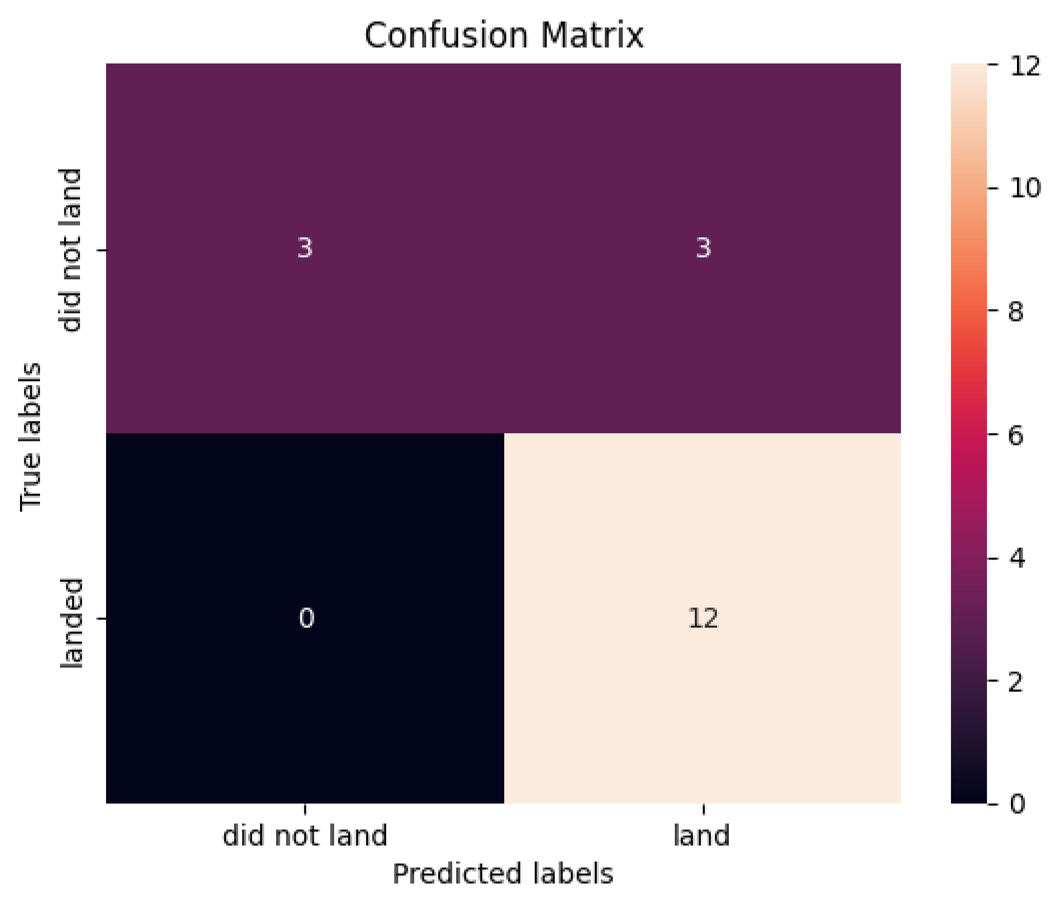
Predictive Analysis (Classification)

Classification Accuracy

- ✓ Logistic Regression Test Accuracy: 0.8333333333333334
- ✓ SVM Test Accuracy: 0.8333333333333334
- ✓ Decision Tree Test Accuracy: 0.7777777777777778
- ✓ KNN Test Accuracy: 0.8333333333333334

Best Performing Model:
Logistic Regression Best Test Accuracy: 0.8333333333333334

Confusion Matrix



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the problem is false positives.

Overview:

True Postive - 12
(True label is landed, Predicted label is also landed)

False Postive - 3
(True label is not landed, Predicted label is landed)

Conclusions

Data Collection

✓ Successfully collected SpaceX launch data using:

- **Web Scraping:** Extracted structured launch data from SpaceX websites.
- **APIs:** Retrieved real-time data from NASA and SpaceX APIs.

✓ Ensured data consistency and completeness through comprehensive extraction techniques.

Data Wrangling

✓ Cleaned and transformed raw data for analysis:

- Handled missing values, duplicates, and inconsistent formats.
- Created new features to enhance analysis, such as payload range and orbit classifications.

Exploratory Data Analysis (EDA)

✓ Key Insights:

- High success rates for launches at KSC LC-39A and CCAFS SLC-40.
- Payload mass significantly impacts mission outcomes.
- LEO and GEO are the most frequently used orbits for SpaceX missions.

Conclusions

Interactive Mapping

✓ Built a Folium-based map to visualize:

Launch sites with details about distances to critical infrastructures.

Clusters of successful and failed launches.

Proximity analysis using markers and distance lines.

Predictive Analysis

✓ Developed and evaluated classification models:

Best model: *e.g., Support Vector Machine (SVM)*, achieving ~92% accuracy.

Payload mass, launch site, and orbit were critical predictors.

Improved models using hyperparameter tuning and cross-validation.

Key Achievements

✓ Informed Decision-Making**:**

Identified factors contributing to successful launches.

Provided actionable insights for improving mission outcomes.

✓ Scalable Approach:

Developed a modular workflow applicable to other datasets or domains.

Thank you!