# Predicting First Stage Recovery SpaceY

# Competitive Edge

**M Huzaifa Khan**

*(An IBM Guided Project)*

# Context

# Executive Summary

SpaceY aims to compete with SpaceX by predicting the success of Falcon 9 first-stage recovery to optimize the launch cost. Develop a machine learning model to predict whether the stage of Falcon 9 will land successfully. Accurate predictions will enable SpaceY to offer competitive pricing and strategically position itself in the commercial space market.

# Introduction

The commercial space industry is rapidly advancing with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX leading the way. SpaceX’s Falcon 9 rocket is notable for its cost efficiency, largely due to its reusable first stage, which significantly reduces launch costs. Space Y aims to compete with SpaceX by predicting the success of Falcon 9 first stage recovery.

Utilize data science techniques to forecast first stage recovery and assess the implications for launch cost optimization.

# Methodology

- **Data Collection and Preparation:**

  Gather historical data on Falcon 9 launches from SpaceX, including features like payload, mission parameters, and recovery outcomes. Clean and preprocess the data to handle missing values and ensure consistency.

- **Exploratory Data Analysis (EDA):**

  Conduct initial analysis to understand data patterns and relationships. Use visualizations to identify key factors affecting first-stage recovery.

- **Predictive Modeling:**

  Select and train machine learning models (e.g., logistic regression, random forest) to predict the likelihood of successful first-stage recovery. Evaluate model performance using metrics such as accuracy, precision, and recall.

- **Visualization and Dashboards:**

  Develop interactive visualizations and dashboards to present insights and predictions. Use tools like Plotly and Folium for dynamic data representation.

- **Results Interpretation:**

  Analyze model outputs and visualizations to conclude first-stage recovery and its impact on launch costs.

# Exploratory Data Analysis with SQL

- **Launch Sites:**

  Identified unique launch site names, with a focus on sites starting with "CCA".

- **Payload Mass:**

  Calculated the total payload mass carried by NASA and the average payload mass carried by booster F9 v1.1.

- **Landing Outcomes:**

  Determined the date of the first successful landing outcome in a ground pad and listed boosters with successful drone ship landings with payload masses between 4000 and 6000 kg.

- **Mission Outcomes:**

  Calculated the total number of successful and failed mission outcomes.

- **Maximum Payload:**

  Identified boosters that have carried the maximum payload mass.

- **2015 Launch Records:**

  Calculated the total payload mass carried by NASA and the average payload mass carried by booster F9 v1.1.

- **Ranked Success Count:**

  Ranked the count of successful landings between 2010-06-04 and 2017-03-20 in descending order.

- **Insights:**

  The analysis provides a comprehensive understanding of SpaceX's launch sites, payload capacities, and landing outcomes.

The findings can be used to inform decisions on future mission planning, payload optimization, and landing site selection.

The findings can be used to inform decisions on future mission planning, payload optimization, and landing site selection.

- **Conclusion:**

  This analysis demonstrates the power of data analysis in extracting valuable insights from complex datasets. The findings provide a deeper understanding of SpaceX's mission capabilities and can be used to drive future success.

## Data Analysis SQL Results

- **Launch Site:**

  Identified four unique launch sites for SpaceX missions.

- **Launch Site Records:**

  Displayed records for launch sites starting with 'CCA' and found specific details for several missions.

- **Payload Mass:**

  Calculated total payload mass for NASA (CRS) missions and found no records; similarly, average payload mass for booster version F9 v1.1 was also not found.

- **Successful Landing:**

  Determined the first successful landing on a ground pad occurred on December 22, 2015.

- **Booster Success:**

  No boosters were identified with successful drone ship landings and payloads between 4000 and 6000 kg.

- **Landing Outcomes:**

  Analyzed landing outcomes and found 'Success' was the most common, while 'Failure (drone ship)' and 'Controlled (ocean)' were also significant.

- **Booster Versions:**

  Listed multiple booster versions associated with maximum payload masses.

- **Outcome Ranking:**

  Ranked landing outcomes between 2010-06-04 and 2017-03-20, with 'No attempt' as the most frequent outcome.

## Exploratory Data Analysis (EDA) and Feature Engineering using Pandas and Matplotlib

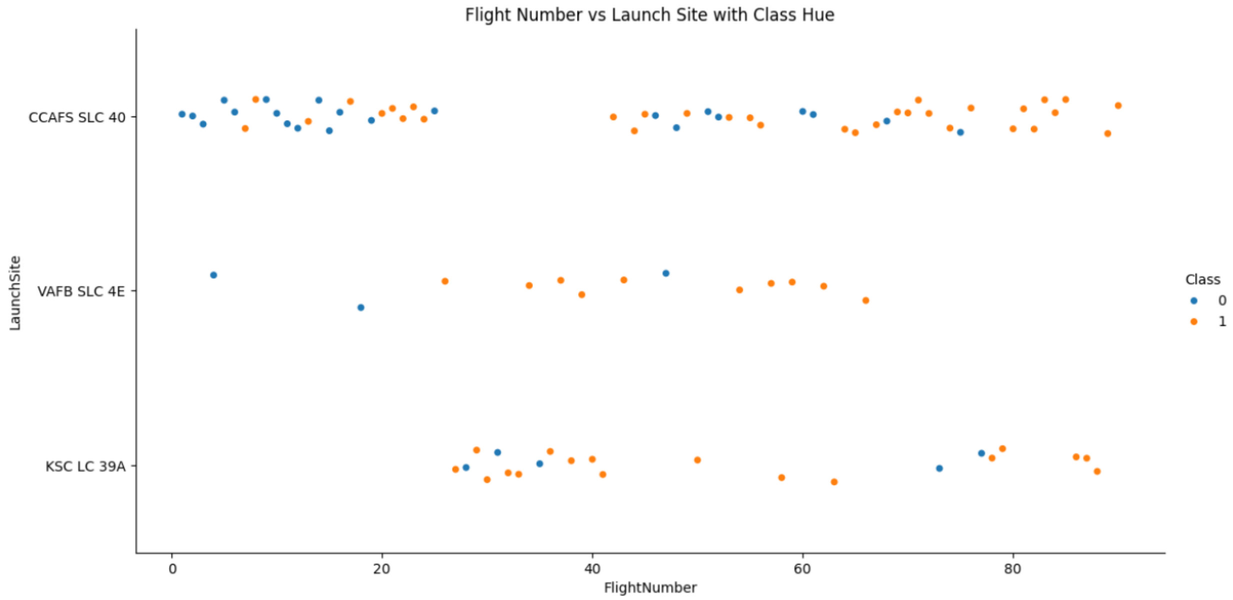
- **Preparing Data:**

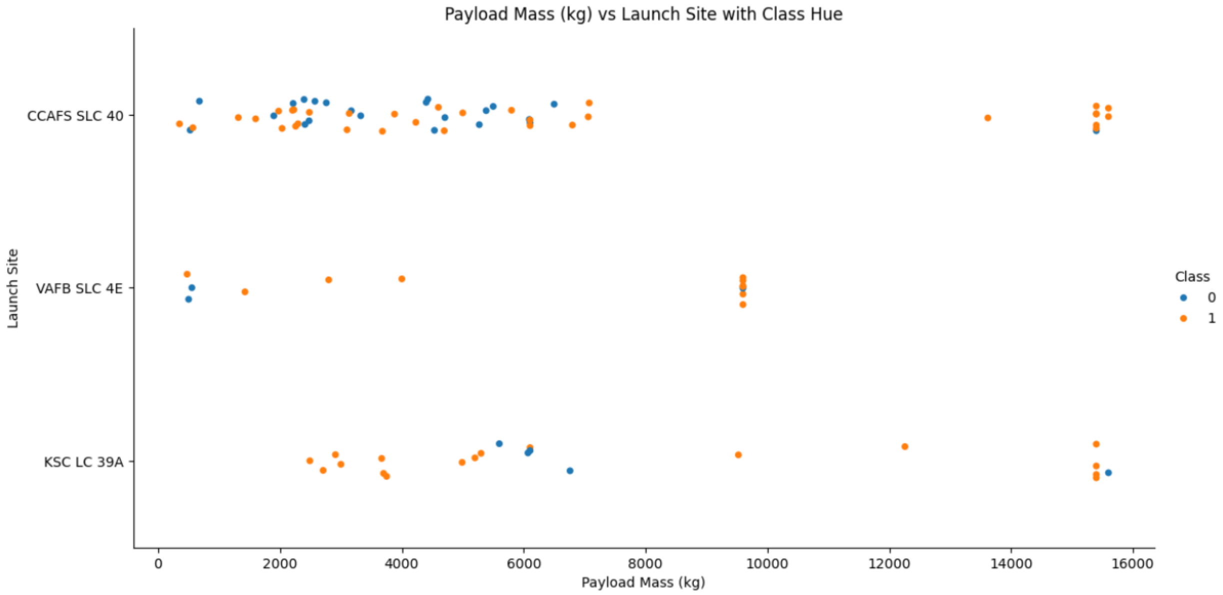
  Converted the ‘features_one_hot’ Dataframe to ensure that all the features are numerical by casting the entire DataFrame to the ‘float64’ data type. This step is essential as it guarantees that all data within the DataFrame is in a format suitable for numerical analysis and modeling. This conversion is crucial for maintaining consistency and avoiding potential issues during further data manipulation or application of analytical methods.

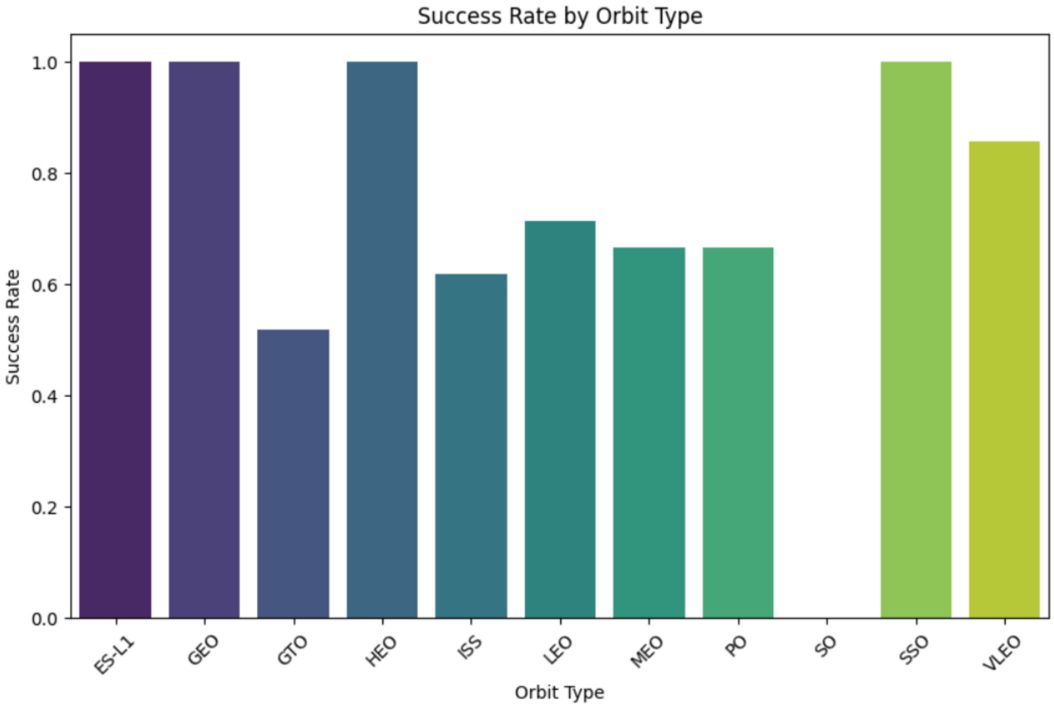
- **Feature Engineering:**

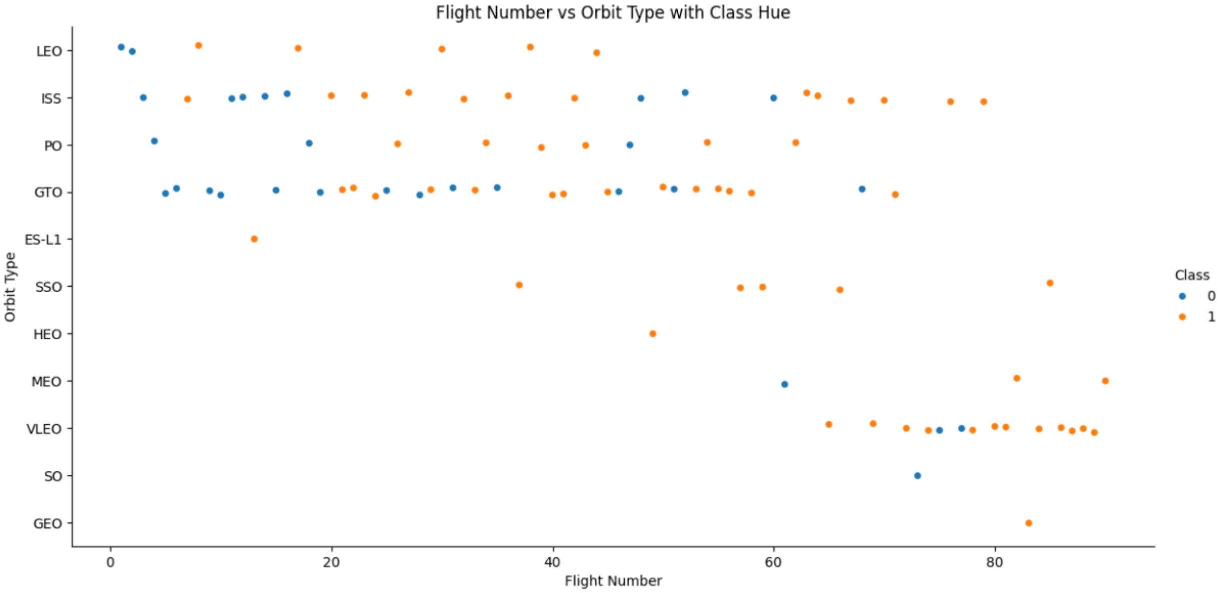
  Verified that the ‘features_one_hot’ Dataframe now exclusively contains columns of type ‘float64’. This uniformity in data type ensures compatibility with a wide range of analytical techniques and machine learning algorithms. The conversion to ‘float64’ provides a solid foundation for accurate data processing, model training, and subsequent analysis, making the dataset well-prepared for any advanced data science tasks.

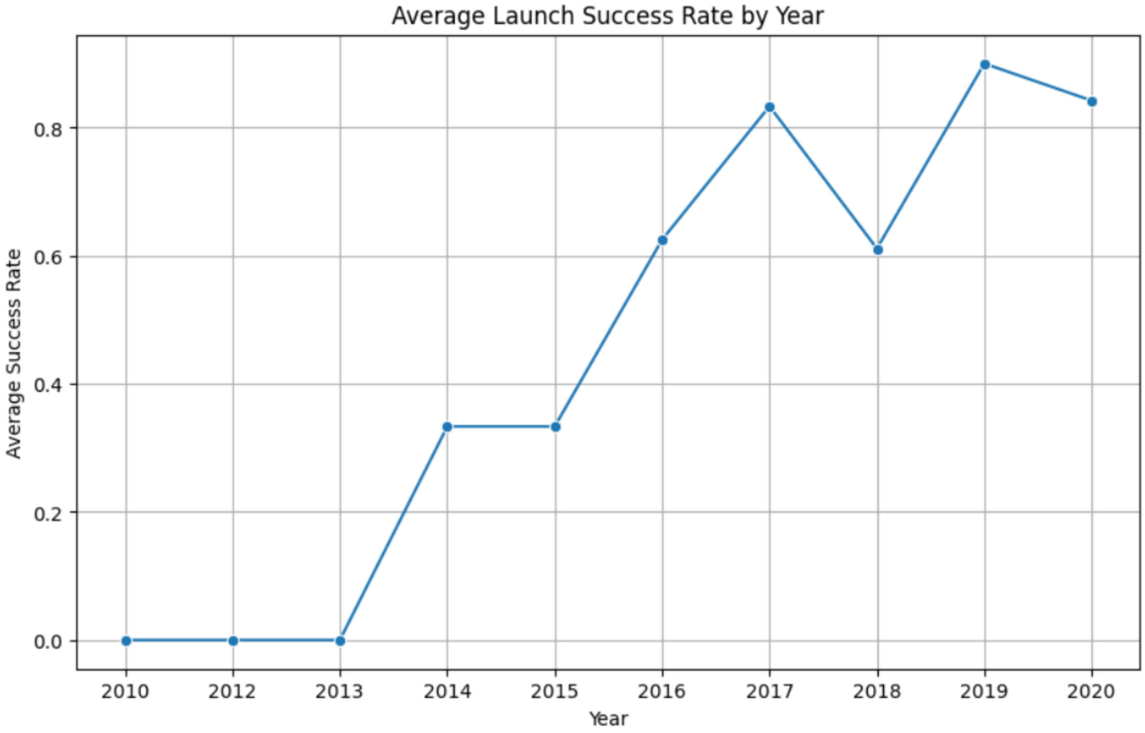
# Exploratory Data Analysis Visualizations

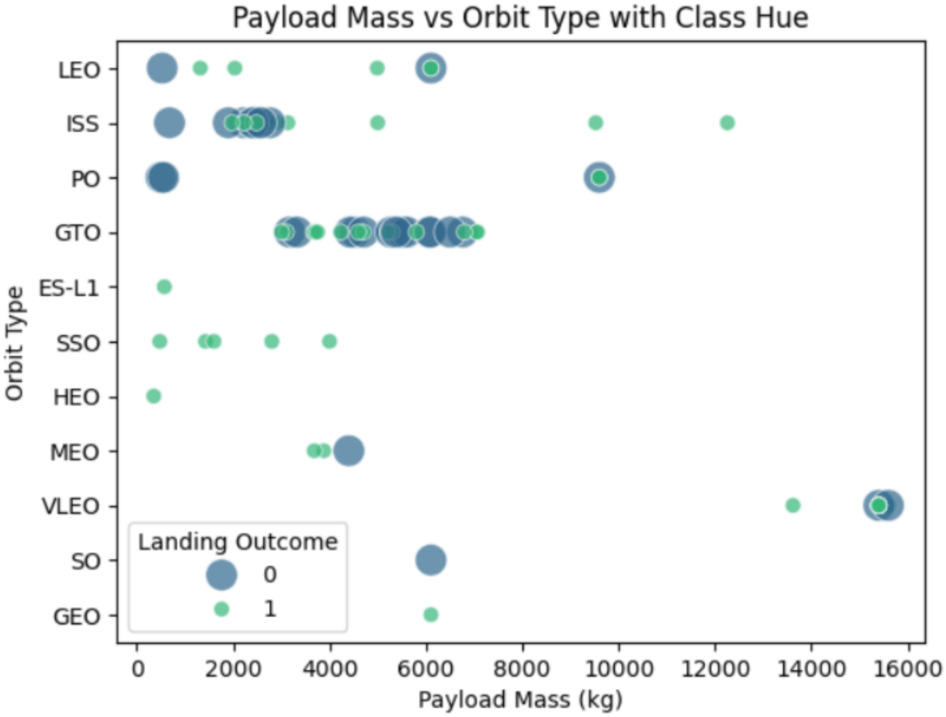





Success Rate by Orbit Type
Success Rate
1.0
0.8
0.6
0.4
0.2
0.0
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit Type


Flight Number vs Orbit Type with Class Hue
Orbit Type
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
0
20
40
60
80
Flight Number
Class
● 0
● 1





# Interactive Map with Folium

- **Data Import and Setup:**

  Import required libraries (‘folium’, ‘pandas’). Read the dataset ‘spacex_launch_geo.csv’ to extract launch site coordinates.

- **Mapping Launch Sites:**

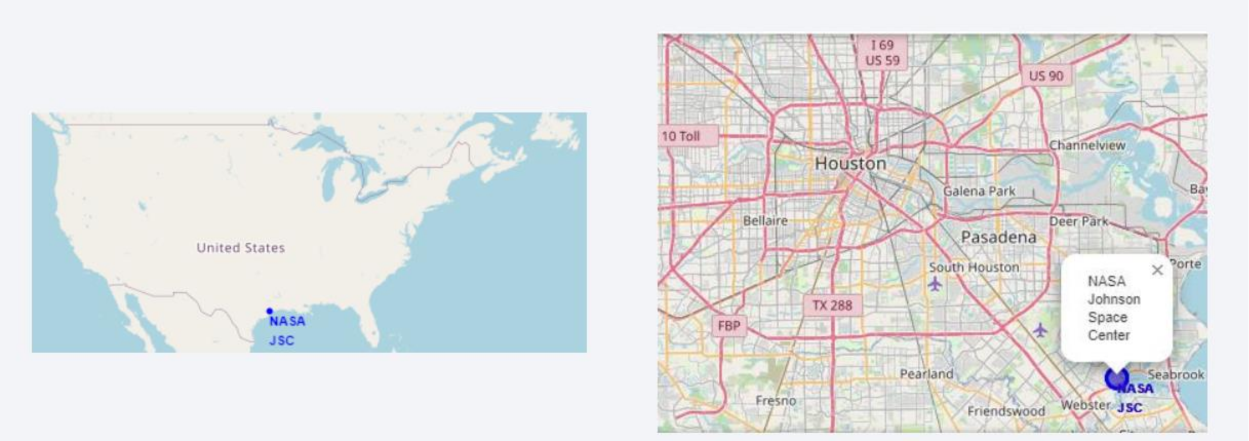
  Use ‘folium’ to create a map with markers for each launch site.

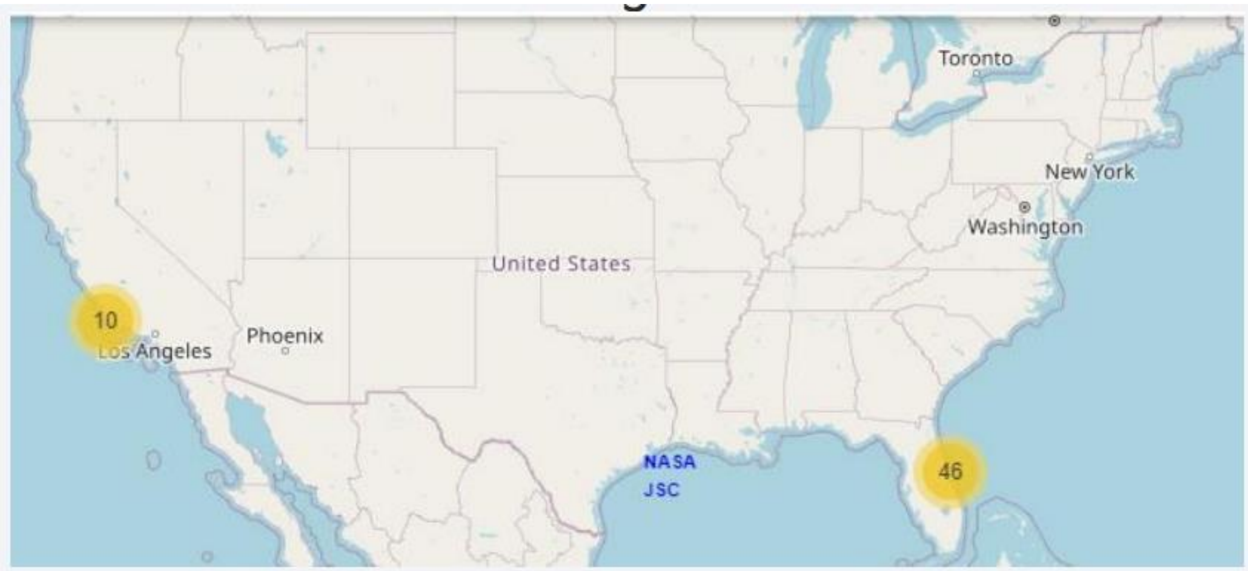
- **Marking Launch Success/Failure:**

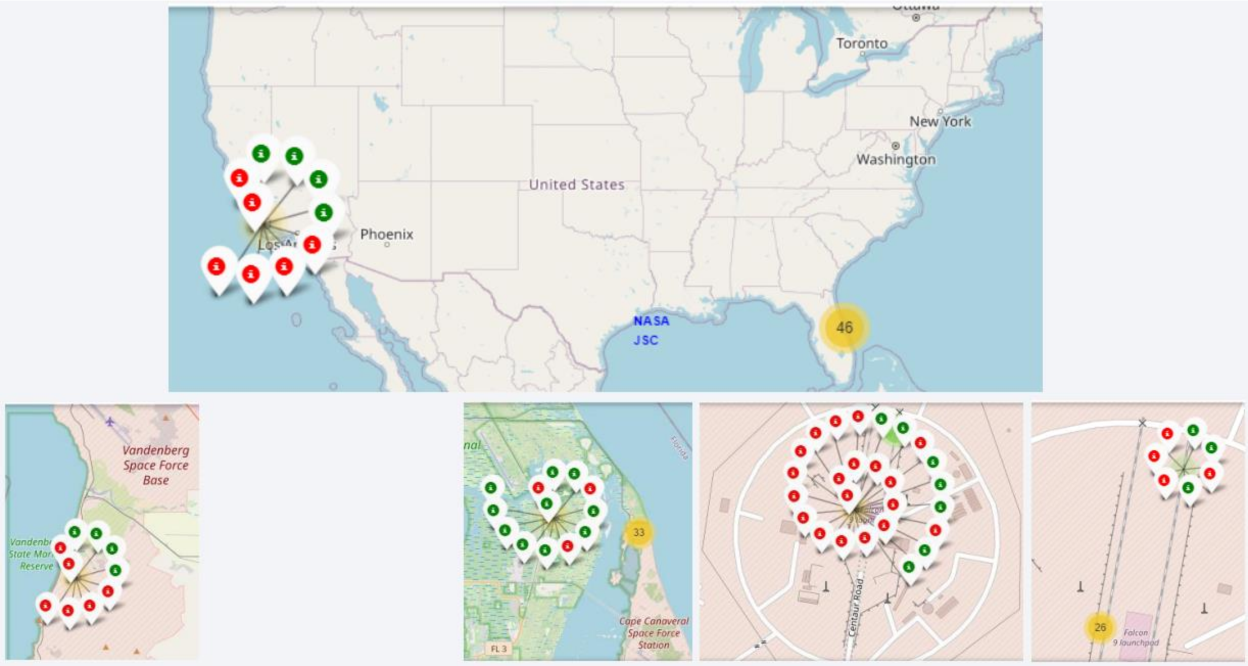
  Overlay success and failure data on the map to visualize performance by site.

- **Distance Calculation:**

  Compute distances from launch sites to nearby railways, highways, coastlines, and cities using the Haversine formula. Analyze the proximity of launch sites to these features to identify any geographical patterns.






Toronto
New York
Washington
United States
Phoenix
NASA
JSC
46
Vandenberg Space Force Base
Vandenberg State Marine Reserve
33
Cape Canaveral Space Force Station
FL 3
Centaur Road
26
Falcon 9 launchpad

# Predictive Analysis Methodology

- **Building the Models:**

Four classification models were built: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (k-NN). Each model was trained on the training data (X_train, Y_train) using cross-validation (cv) to tune hyperparameters.

- **Evaluating the Models:**

  Each model was evaluated on the test data (X_test, Y_test) using the score() method to calculate accuracy. Accuracies were stored in a dictionary for comparison.

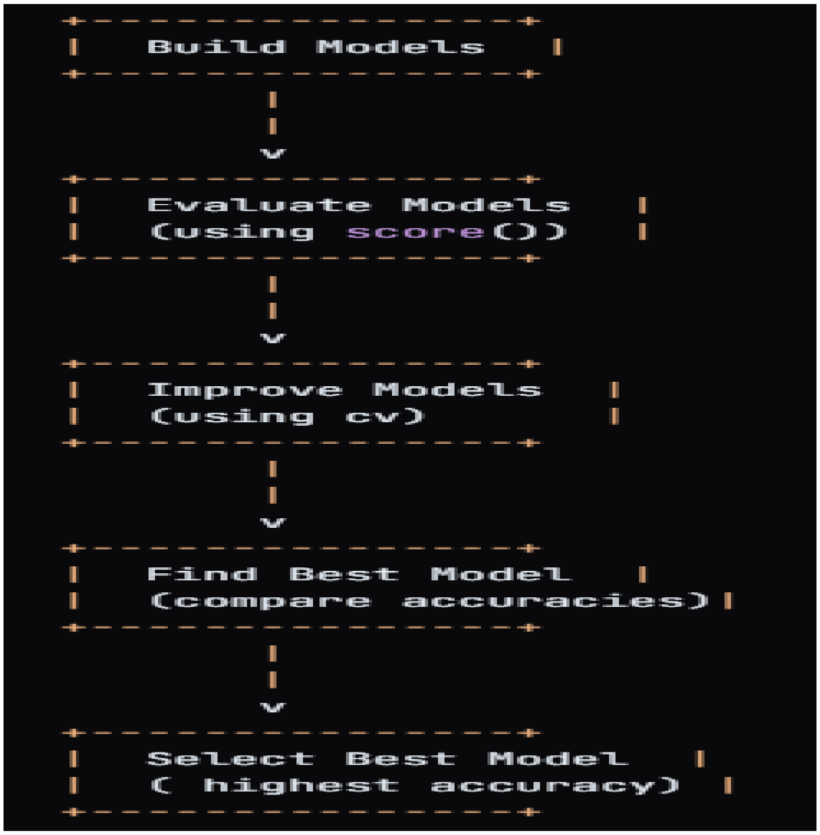
- **Improving the Models:**

  Hyperparameters were tuned using cross-validation (cv) to improve model performance. No feature engineering or data preprocessing was mentioned.

- **Finding the Best Model:**

  The best model was found by comparing the accuracies of each model. The model with the highest accuracy was selected as the best-performing model.

- **Flow chart:**



```
+------------------+
|   Build Models    |
+------------------+
        |
        |
        v
+------------------+
|   Evaluate Models    |
|   (using score())    |
+------------------+
        |
        |
        v
+------------------+
|   Improve Models   |
|   (using cv)       |
+------------------+
        |
        |
        v
+------------------+
|   Find Best Model    |
|   (compare accuracies)|
+------------------+
        |
        |
        v
+------------------+
|   Select Best Model   |
|   ( highest accuracy)  |
+------------------+
```

## Predictive Analysis

- **Model Evaluation:**

  Tested four different models: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (k-NN). Calculated the test accuracy for each model using the ‘.score()’ method on the test dataset.

- **Accuracy Comparison:**

  Printed the test accuracy for each model. Compiled the accuracy results into a dictionary and identified the model with the highest accuracy.

- **Results:**

  Logistic Regression, SVM, and k-NN models achieved an accuracy of 0.8333. The Decision Tree model had a slightly lower accuracy of 0.7778. Determined that Logistic Regression was the best-performing model based on the highest accuracy.

- **Findings:**

  All models except Decision Tree had the same accuracy score, but Logistic Regression was selected as the best-performing model due to its comparable performance and possibly other factors like simplicity or interpretability.

# Conclusion

Determined the impact of various factors (like orbit type and launch site) on launch success rates using data-driven insights. Compare different machine learning models (Logistic Regression, SVM, Decision Tree, k-NN) to identify the most accurate model for predicting launch outcomes.

Mapped launch sites and analyzed their proximity to key features like railways, highways, and coastlines to understand geographical influences on launch site selection. Suggested optimal launch site locations and improvements in model predictions based on the analysis, which can help in strategic planning and operational efficiency.

Provided a deeper understanding of factors influencing launch success and site selection, enabling better decision-making and planning for future missions.