IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alper Karayaz

Outline

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

Executive Summary

- **Summary of Methodologies:** The project combined exploratory data analysis, interactive analytics, and predictive modeling using classification algorithms. Data was cleaned, transformed, scaled, and split into training and test sets. Multiple classifiers were built, tuned with GridSearchCV, and evaluated to determine their predictive accuracy.
- **Summary of Results:** EDA revealed patterns in payload, launch site, and booster versions affecting success. Interactive visualizations highlighted key trends and allowed dynamic exploration. Predictive models all achieved 83% accuracy on test data, with Decision Tree slightly higher on training data (≈88.93%), and confusion matrices were similar across models, indicating consistent performance.

Introduction

- **Project Background:**
  - The goal of this project is to predict whether the Falcon 9 first stage will land successfully.
  - Landing success is critical because it determines whether the rocket can be reused, which significantly reduces launch costs.
  - Historical SpaceX launch data is used to analyze patterns and train machine learning models for landing predictions which can inform cost and planning decisions for launches.

- **Problems to Answer:**
  - Can historical launch data predict whether a Falcon 9 first stage will land successfully?
  - Which factors are most predictive of landing success?
  - How can machine learning, dashboards, and interactive visualizations help the team make informed decisions regarding launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data for this project was collected from SpaceX launch records using REST API calls to retrieve structured JSON data and from Wikipedia via web scraping to extract historical Falcon 9 and Falcon Heavy launch information.
- Perform data wrangling
  - The dataset was loaded, cleaned, and transformed, handling missing values, encoding categorical variables, scaling numeric features, engineering new columns, and splitting into training and test sets for modeling
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Classification models were built on the training data, tuned using hyperparameter search with cross-validation, and evaluated on the test set using accuracy and confusion matrices to identify the most effective model for predicting outcomes.

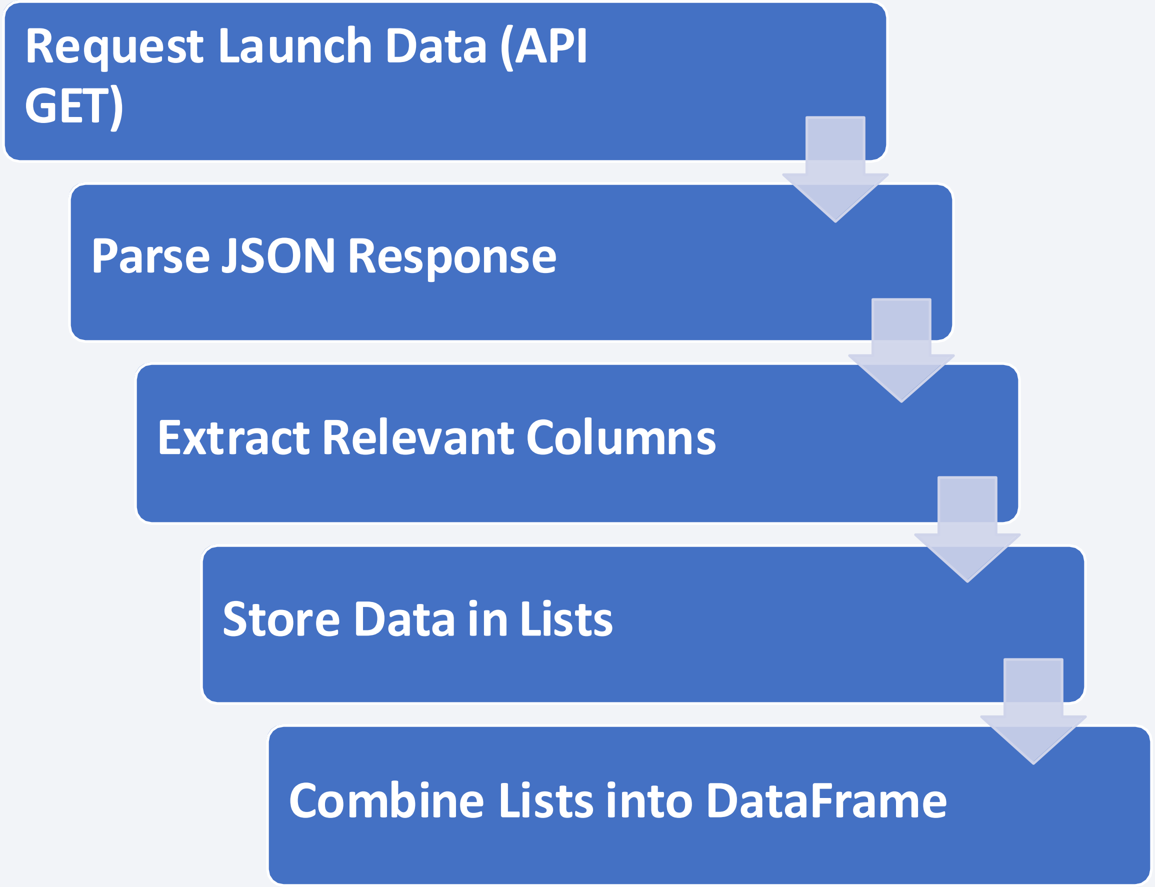
Data Collection

- Collected historical SpaceX launch data from the official API.
- Extracted detailed information for each launch using helper functions:
    - **Rocket:** Booster version
    - **Payload:** Mass, Orbit
    - **Launchpad:** Name, Latitude, Longitude
    - **Core:** Landing outcome, reuse info, type, legs, grid fins, block, serial
- Filtered dataset to include only Falcon 9 launches and removed multiple cores/payloads.
- Converted dates to datetime format and handled missing values (filled PayloadMass with mean).
- Created a clean, structured Data Frame ready for EDA and modeling.

Data Collection – SpaceX API

- Collected launch data using HTTP GET requests to the SpaceX API
- Extracted data from JSON responses
- Focused on rocket, payloads, launchpad, and cores columns
- Transformed API data into lists and then into a Pandas dataframe
- Handled missing values and filtered for Falcon 9 launches
- Exported cleaned dataset to CSV
- GitHub URL:

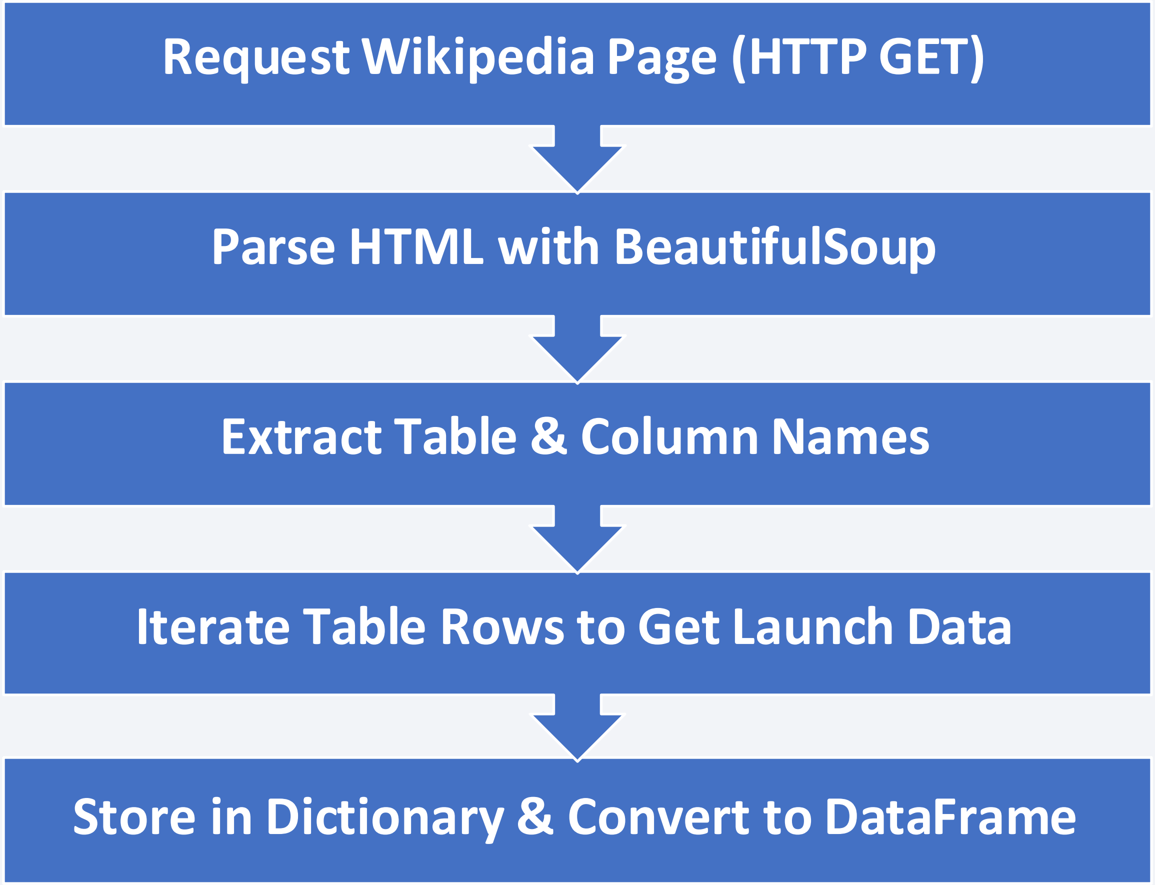
https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/data_collection_api.ipynb



Data Collection - Scraping

- Retrieved Wikipedia Falcon 9 launch page using HTTP GET request
- Parsed HTML tables with BeautifulSoup
- Extracted flight number, date, booster version, launch site, payload, orbit, outcome
- Stored extracted data in Python dictionary, then converted to Pandas dataframe
- Cleaned data (removed references, normalized payload mass, split date/time)
- Exported cleaned dataset to CSV
- GitHub URL:

https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/data_collection_webscraping.ipynb



Data Wrangling

- **Loaded and cleaned the dataset**
    - Checked column types, summary statistics, and missing values.
    - Removed duplicates and irrelevant records to ensure accuracy.
- **Transformed and encoded data for analysis**
    - Converted categorical variables to numeric formats.
    - Scaled numerical features for model compatibility.
- **Engineered features and prepared for modeling**
    - Created new columns like split date/time and landing outcomes.
    - Split data into training and testing sets for supervised learning.
- **GitHub URL:**

https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/data_wrangling.ipynb

Data Wrangling

- **Load Dataset**
  - Read CSV or API data
- **Clean Data**
  - Handle missing values
  - Remove duplicates
- **Transform & Encode**
  - Convert categorical to numeric
  - Scale numeric features
- **Feature Engineering**
  - Split date/time
  - Create landing outcome labels
- **Data Ready for EDA and Modeling**

EDA with Data Visualization

- **EDA Charts Purpose:** The following charts were created to help uncover patterns in the data and guide feature selection for the machine learning model.
- **Flight Number vs Payload Mass (scatter):** To examine how launch experience and payload weight affect first-stage landing success.
- **Flight Number vs Launch Site (scatter):** To identify site-specific trends in landing outcomes.
- **Payload Mass vs Launch Site (scatter):** To explore differences in payload handling across launch sites.
- **Success Rate by Orbit (bar chart):** To determine which orbit types have higher or lower landing success rates.
- **Flight Number vs Orbit & Payload vs Orbit (scatter):** To analyze how experience and payload impact success across different orbits.
- **Average Launch Success Rate by Year (line chart):** To visualize trends in landing success over time.

- **GitHub URL:** https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/eda_visualizations.ipynb

EDA with SQL

- **Unique Launch Sites:** Retrieved distinct launch sites.
- **Filtered Launches:** Displayed launches starting with “CCA”.
- **Payload Analysis:** Calculated total and average payloads for specific boosters/customers.
- **Landing Outcomes:** Found first successful ground pad landing, listed successful drone ship boosters, and counted mission outcomes.
- **Max Payload & Rankings:** Identified boosters with max payload and ranked landing outcomes.
- **2015 Failures:** Listed failed drone ship landings by month.
- **Landing outcome ranking:** Ranked landing outcomes by occurrence between 2010-06-04 and 2017-03-20 in descending order.
- **GitHub URL:** https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/eda_sql.ipynb

Build an Interactive Map with Folium

In the interactive Folium map, I created and added the following objects:

- **Markers** for each launch site and individual launches, color-coded by success (green) or failure (red) to visualize outcomes.
- **Circles** around launch sites to highlight their locations and provide popup labels for identification.
- **Marker Clusters** to group multiple launches at the same site for clarity and reduce clutter.
- **MousePosition Plugin** to allow easy reading of coordinates for any point of interest.
- **PolyLines** connecting launch sites to nearby coastlines, highways, railways, and cities, showing proximities and distances.
- **Distance Markers** at proximities, labeled with distances in kilometers to quantify closeness to critical infrastructure.
- These objects were added to explore geographical patterns, understand site accessibility, and evaluate safety and strategic advantages of existing launch site locations.
- **GitHub URL:** https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/interactive_folium.ipynb

Build a Dashboard with Plotly Dash

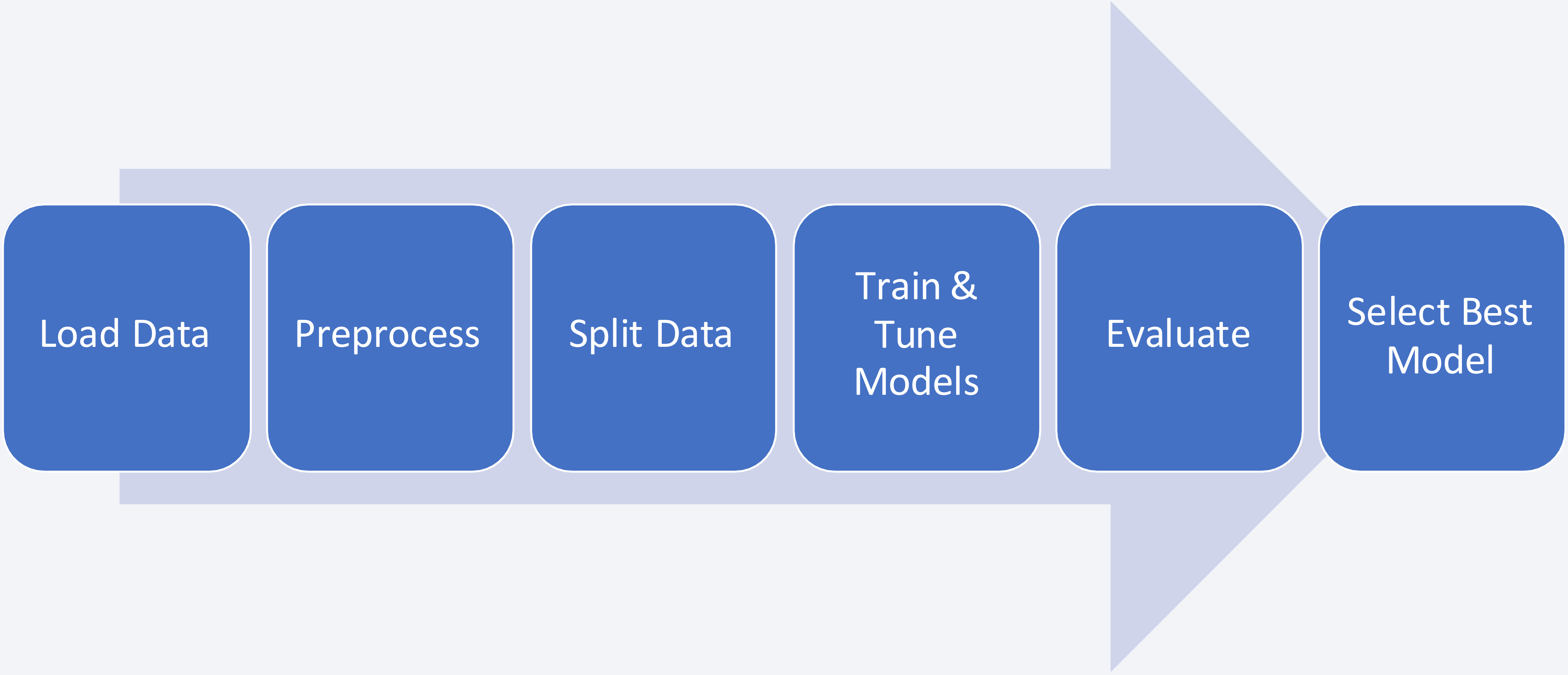
Interactive SpaceX Launch Dashboard

- **Plots & Interactions Added:**
- **Launch Success Pie Chart:** Shows success counts for all sites or a selected site.
  - Controlled by Launch Site Dropdown.
  - Highlights successful vs. unsuccessful launches (color-coded).
- **Payload vs. Launch Outcome Scatter Plot:** Displays payload mass against success/failure.
  - Points color-coded by **Booster Version.**
  - Controlled by **Launch Site Dropdown** + **Payload Range Slider**.
- **Purpose:** Enables visual comparison of launch success rates, exploration of payload effects, and booster performance analysis.

Predictive Analysis (Classification)

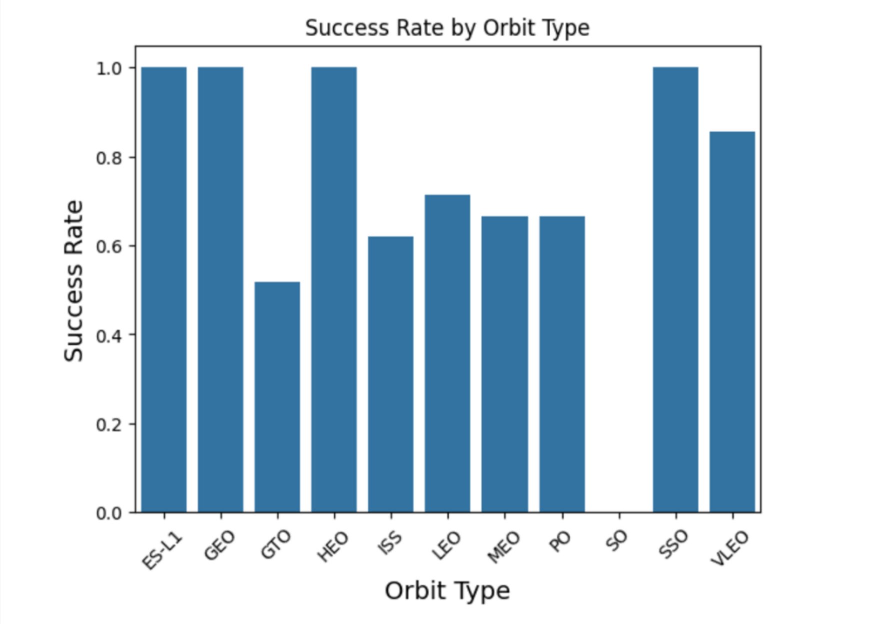
- **Data Preparation:** Loaded Falcon 9 launch datasets, standardized features, and split into training and test sets.
- **Model Training:** Trained Logistic Regression, SVM, Decision Tree, and KNN models, optimizing hyperparameters using GridSearchCV with 10-fold cross-validation.
- **Evaluation:** Assessed each model using test set accuracy and confusion matrices to identify prediction errors and performance differences.
- **Best Model Selection:** All models performed similarly, with Decision Tree showing slightly better validation performance, making it the most reliable for predicting first stage landings.
- **GitHub URL:** https://github.com/akarayaz/Falcon9-LandingPrediction-Project/blob/main/machine_learning_predicti

Predictive Analysis (Classification)



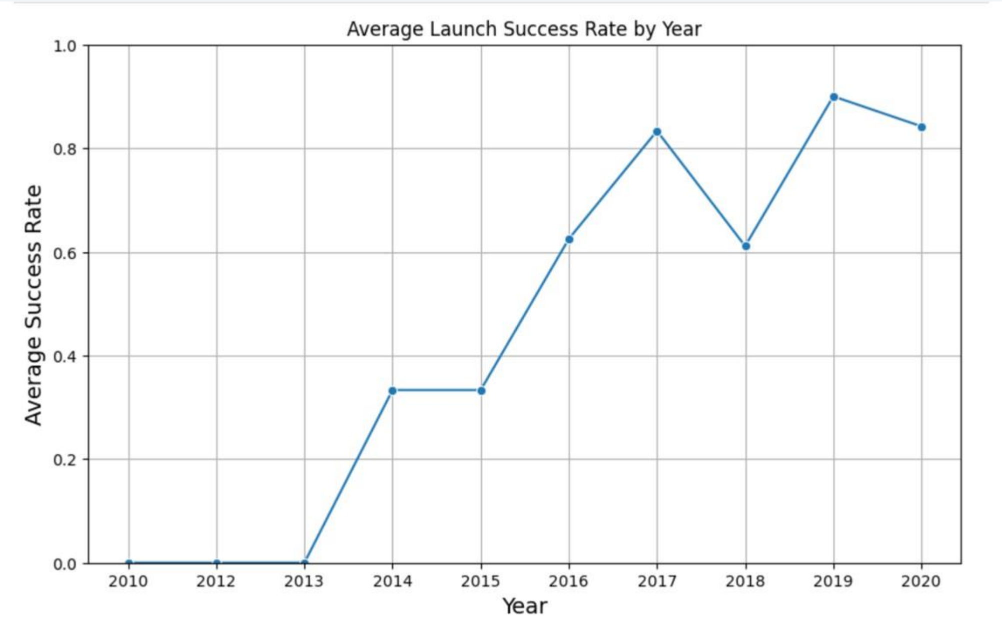
EDA Results

- **Launch Experience Improves Success:** Earlier flights had more failures, but success rates increased as flight numbers grew, showing that accumulated experience improves landing reliability.
- **Orbit Impacts Outcomes:** Some orbit types, like ES-L1, GEO, HEO, and SSO, have 100% success rates, while GTO is less consistent. This means the type of orbit naturally influences landing probability, allowing us to anticipate higher risks for challenging orbits.



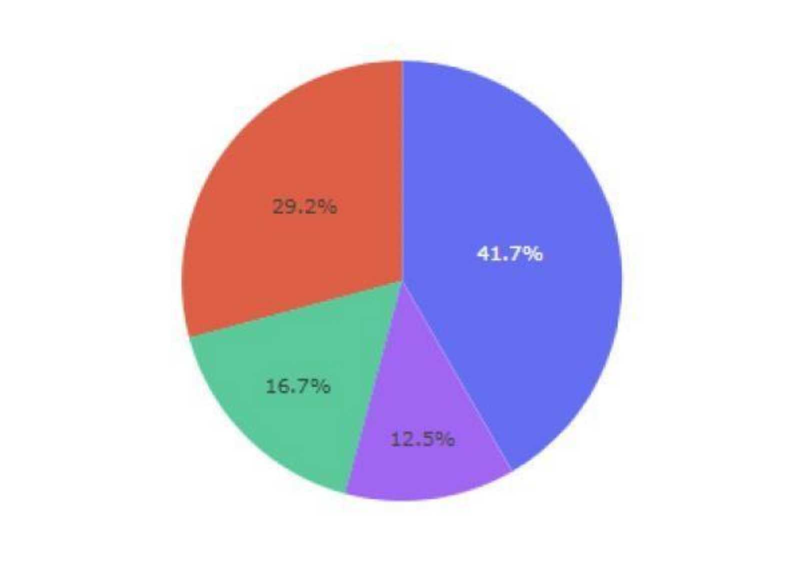
EDA Results

- **Payload and Site Matter:** Heavy payloads are more likely to succeed at certain launch sites, such as CCAFS SLC-40 and KSC LC-39A. This tells us that both payload mass and launch site choice are critical factors in predicting successful recoveries.
- **Temporal Improvements:** Landing success has steadily increased from 2013–2020, reflecting technological advances and process improvements, which means historical trends can guide expectations for future missions.



Interactive Analytics Results

- **Site Success Rates:** KSC LC-39A has the highest success rate (~77%), while CCAFS LC-40 has the lowest (~27%). This indicates that launch site is a strong predictor of first-stage landing success.
- **Payload Trends:** Launches under ~6000 kg and using FT or B5 boosters tend to land successfully more often, which helps predict outcomes based on payload and booster type.
- **Visualization Benefits:** Interactive dashboards highlight patterns in real time, allowing analysts to identify factors that influence landing success and improve prediction accuracy.



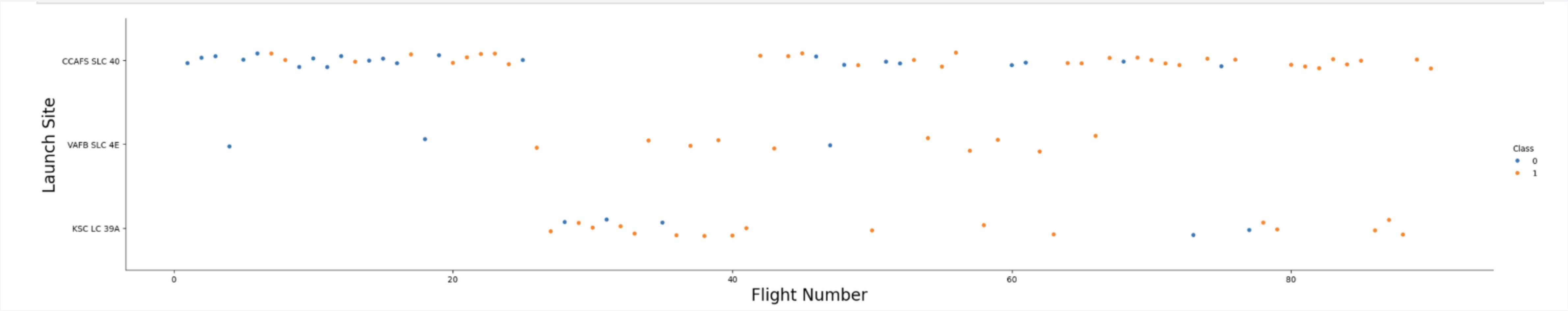
Classification Results

- **Reliable Predictions:** Logistic Regression, SVM, Decision Tree, and KNN classifiers all achieved 83.33% test accuracy, demonstrating that historical data can consistently predict first-stage landing outcomes.
- **Decision Tree Training Advantage:** The Decision Tree had slightly higher training accuracy (88.93%), showing it may capture subtle patterns in the data slightly better, which can inform simulations or scenario testing.
- **Error Analysis:** Confusion matrices were identical across models (12 TP, 3 FP, 3 TN, 0 FN), indicating that false positives are the primary type of error. This helps focus attention on overestimating failures when evaluating risk.
- **Practical Takeaway:** These predictive models serve as effective tools for assessing landing success probabilities, supporting data-driven evaluation and risk assessment.

Section 2

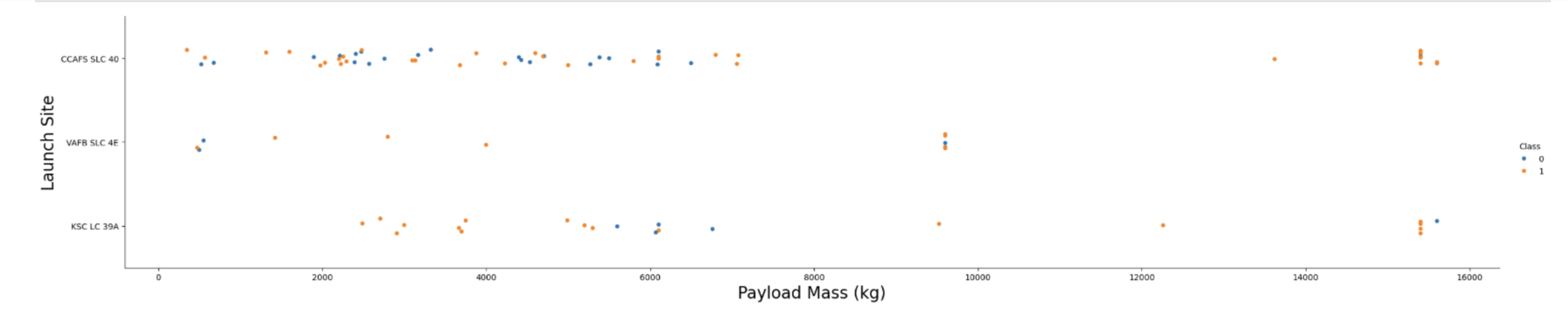
Insights drawn from EDA

Flight Number vs. Launch Site



- The plot shows that earlier flights at all launch sites had a higher frequency of landing failures, with success increasing as the number of flights grew.

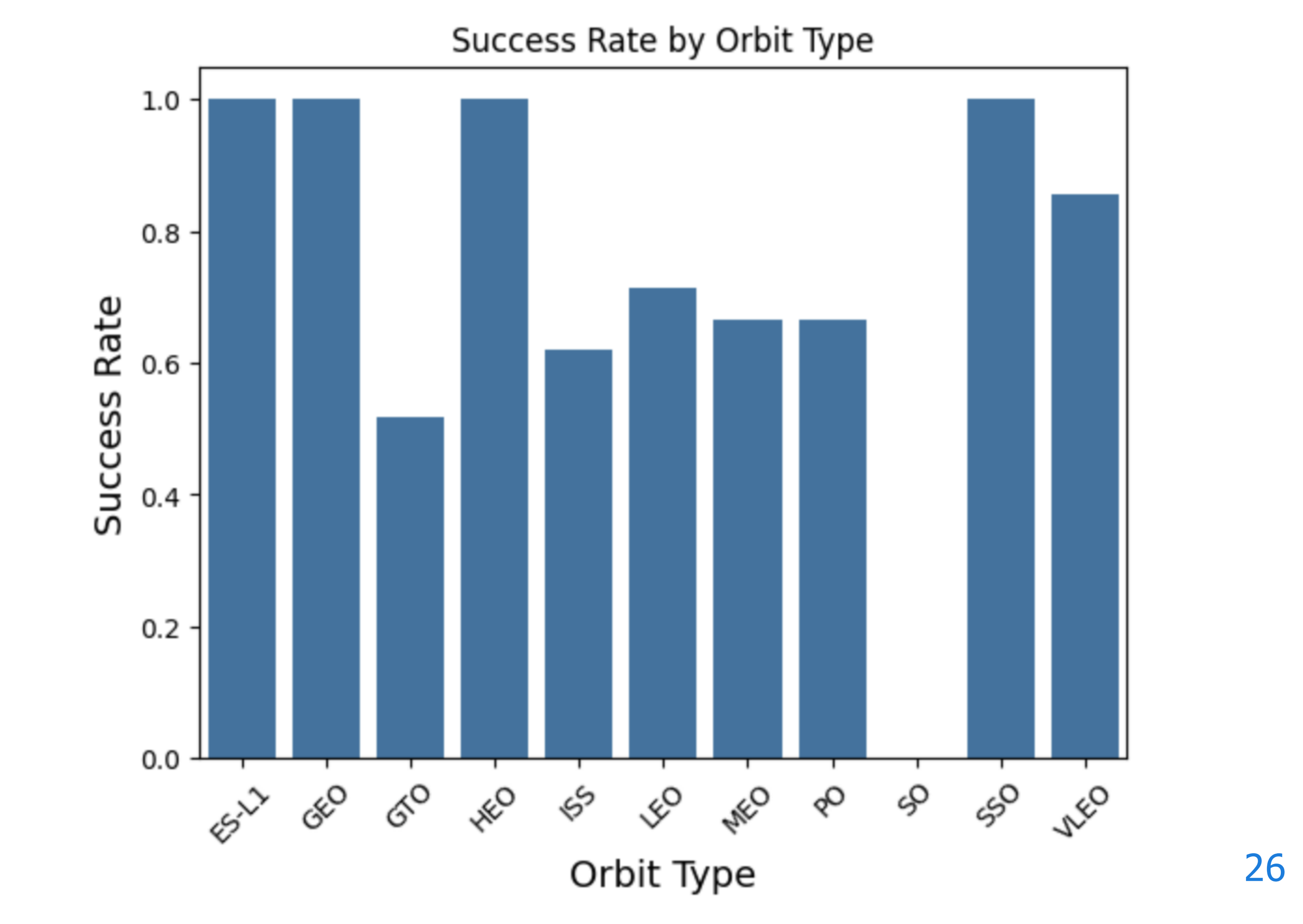
Payload vs. Launch Site



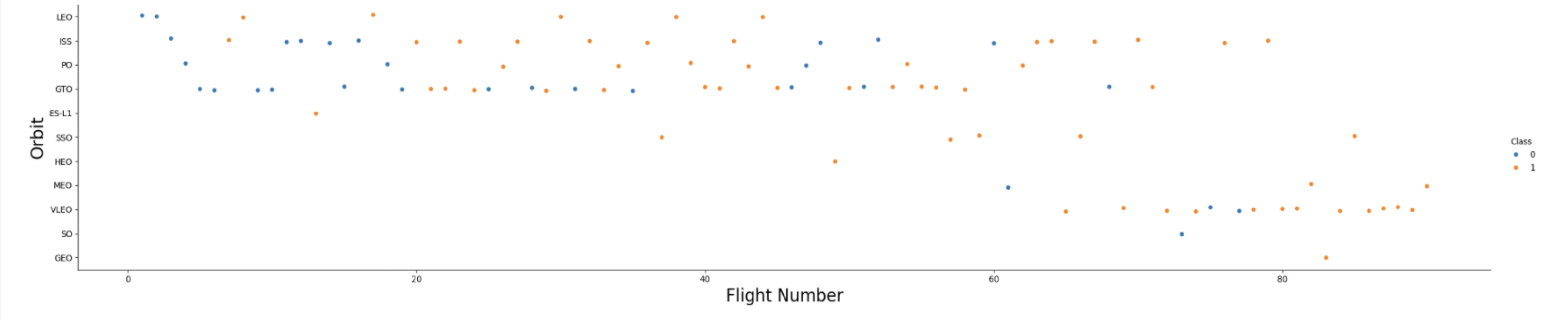
- The scatter plot shows that VAFB-SLC rarely launches very heavy payloads (over 10,000 kg), while CCAFS SLC-40 and KSC LC39-A have successfully handled heavy payloads, including launches around 15,500 kg.
- Landing success is influenced not just by payload mass but also by launch site - some sites, like VAFB-SLC, rarely attempt very heavy launches, while others, like CCAFS SLC-40 and KSC LC39-A, can successfully recover first stages even with heavy payloads.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits had a 100% landing success rate - every first stage launched to these orbits landed successfully.
- Orbit type is a strong indicator of first-stage landing success, with some orbits consistently achieving higher success rates than others.

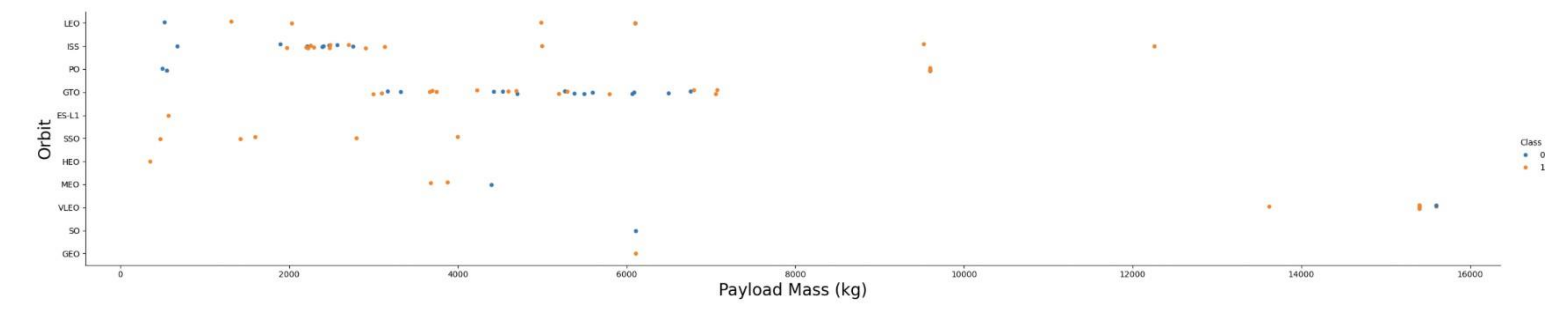


Flight Number vs. Orbit Type



- The plot shows that in LEO, landing success tends to improve with flight number, while in GTO, there is no clear relationship between flight experience and success.
- This tells us that landing experience improves success for some orbits, like LEO, but in more difficult orbits like GTO, other factors dominate outcomes.

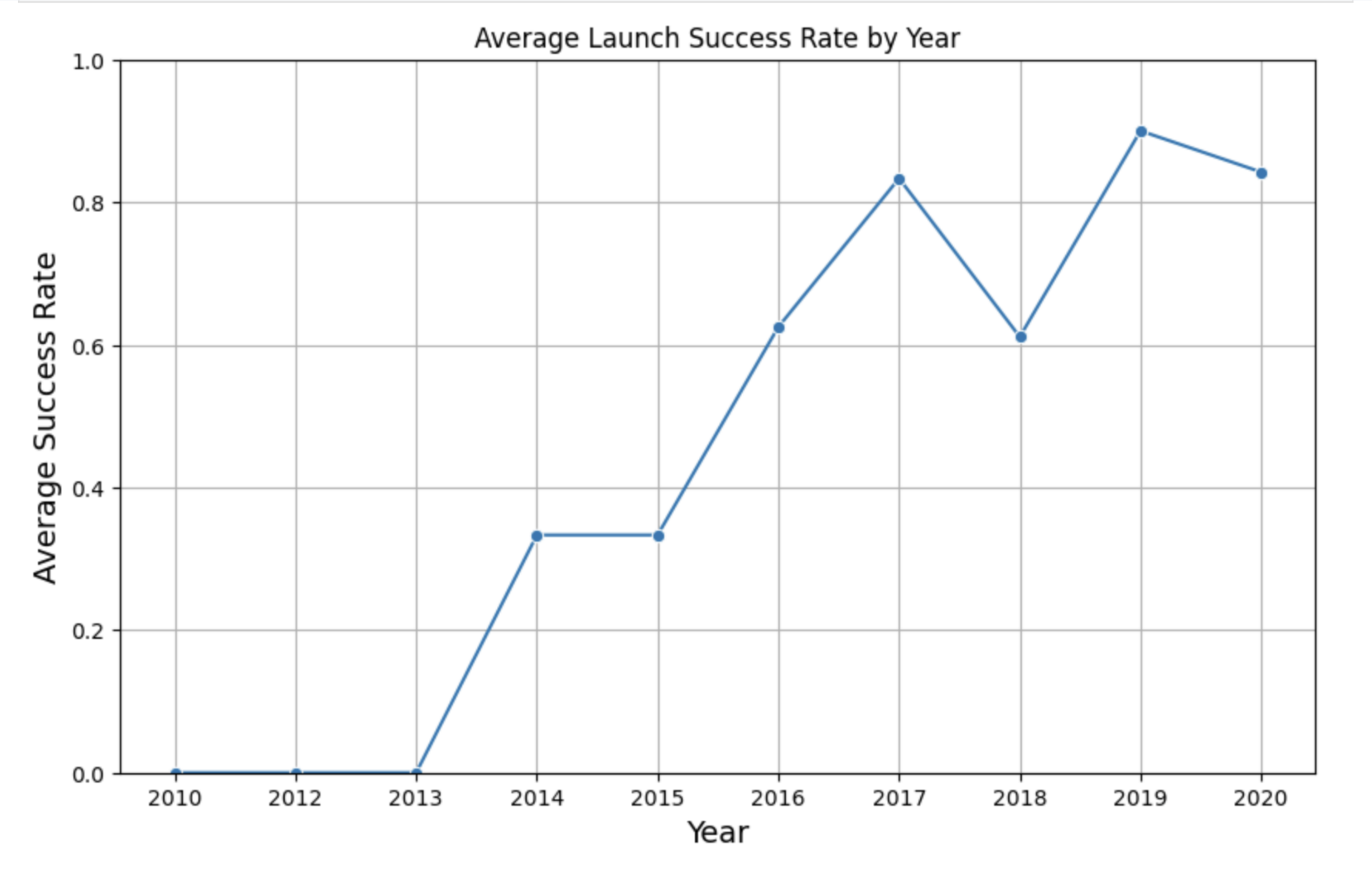
Payload vs. Orbit Type



- The plot shows that landing success varies by both payload mass and orbit, with some orbits handling heavy payloads successfully while others, like GTO, have inconsistent outcomes.

Launch Success Yearly Trend

- The chart shows that first-stage landing success has steadily increased from 2013 through 2020, reflecting general improvements in technology, processes, and accumulated launch experience over time.



All Launch Site Names

- **Unique Launch Sites:**
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

- **Explanation:**
  This query retrieves all distinct launch sites from the SpaceX dataset, showing the four locations where Falcon 9 rockets have been launched.

Launch Site Names Begin with 'CCA'

[11]:

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

- This query retrieves the first five records from the dataset where the launch site begins with ‘CCA’, showing Falcon 9 launches from CCAFS.

Total Payload Mass

- Total Payload Mass = 45,596 kg
- This query sums the payload mass for all Falcon 9 launches where NASA (CRS) was the customer, showing a total of 45,596 kg carried on these missions.

Average Payload Mass by F9 v1.1

- Average payload mass by booster version F9 v1.1 = 2,535 kg
- This query calculates the average payload mass (in kilograms) carried by Falcon 9 booster version F9 v1.1, which comes out to approximately 2,535 kg.

First Successful Ground Landing Date

- First successful landing date – 2015-12-22
- This query finds the earliest date when a Falcon 9 first stage successfully landed on a ground pad, which occurred on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Booster Version:
  F9 FT B1022
  F9 FT B1026
  F9 FT B1021.2
  F9 FT B1031.2

- **Explanation:**
  This query lists the unique booster versions that successfully landed on a drone ship while carrying a payload mass between 4,000 kg and 6,000 kg.

Total Number of Successful and Failure Mission Outcomes

- Mission Outcome:
  Failure (in flight) - 1
  Success - 98
  Success - 1
  Success (payload status unclear) - 1

- **Explanation:**
  This query counts the total number of missions for each outcome type, showing that most missions were successful, with a few failures or unclear payload outcomes.

Boosters Carried Maximum Payload

- Booster _Version:
  F9 B5 B1048.4
  F9 B5 B1049.4
  F9 B5 B1051.3
  F9 B5 B1056.4
  F9 B5 B1048.5
  F9 B5 B1051.4
  F9 B5 B1049.5
  F9 B5 B1060.2
  F9 B5 B1058.3
  F9 B5 B1051.6
  F9 B5 B1060.3
  F9 B5 B1049.7
- **Explanation:**
  This query identifies all booster versions that carried the maximum payload mass in the dataset using a subquery to find the highest payload value.

2015 Launch Records

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- **Explanation:**
  This query lists all failed drone ship landings in 2015, showing the month, booster version, and launch site, using substring functions to extract year and month from the date.

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

- This query ranks the landing outcomes by count between June 4, 2010, and March 20, 2017, showing which outcomes occurred most frequently in descending order.

Landing_Outcome	Outcome_Count	Rank
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	4
Controlled (ocean)	3	4
Uncontrolled (ocean)	2	6
Failure (parachute)	2	6
Precluded (drone ship)	1	8


Section 3
Launch Sites
Proximities Analysis

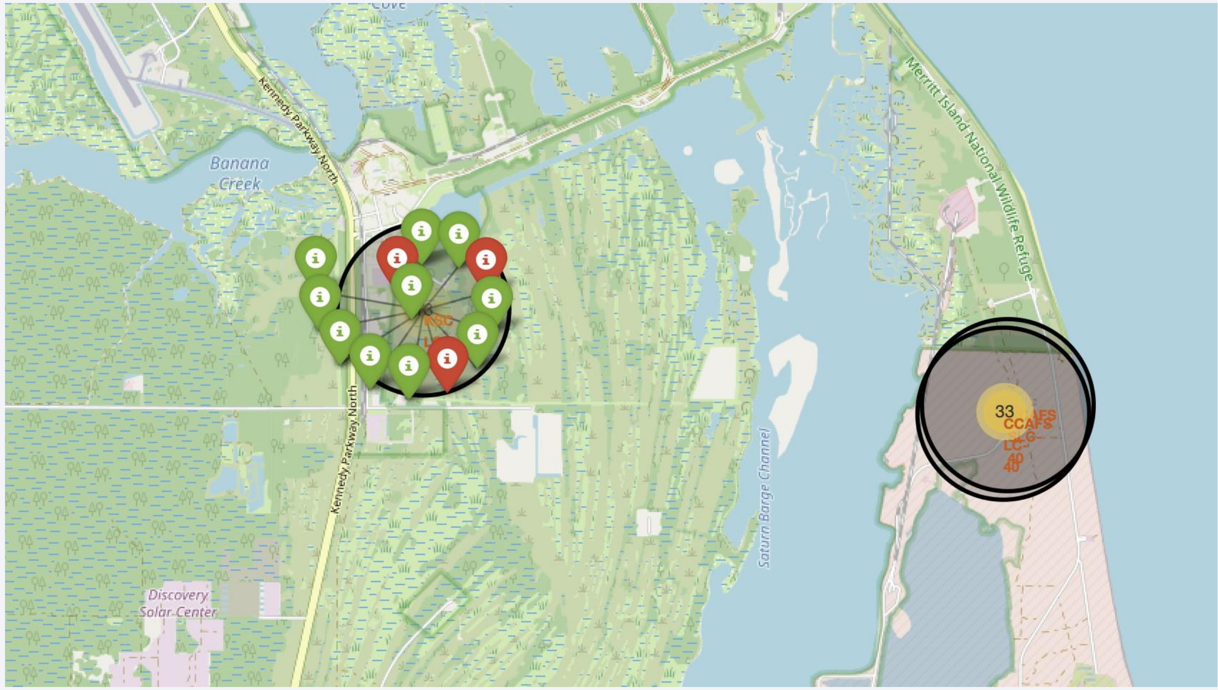
Global Launch Site Locations

Key Elements and Findings:

- **Launch Sites:** Four main sites are visible: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E.
- **Geographical Patterns:** Florida sites are clustered near the equator and coast for optimal launch trajectories and safe overwater landings. The California site (VAFB) is farther north, which makes sense as it's primarily used for polar orbits.

Launch Outcomes Color-labeled

- **Launch Outcomes by Site**
- **KSC LC-39A:** 13 total launches - **10 successes**, 3 failures
- **CCAFS SLC-40:** 7 total launches - 3 successes, 4 failures
- **CCAFS LC-40:** 26 total launches - 7 successes, 19 failures
- **VAFB SLC-4E:** 10 total launches - 4 successes, 6 failures



- **Key Takeaways:**
- KSC LC-39A shows the highest success rate (~77%).
- CCAFS LC-40 has the lowest success rate (~27%).

Launch Site Proximity Analysis

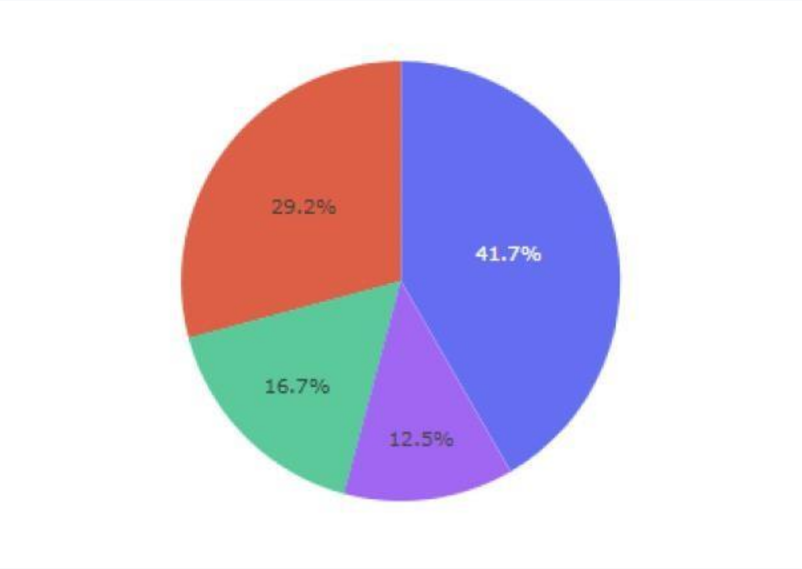
- **Key Findings:**
- Proximity to the coastline allows launches to head over the ocean, reducing risks to populated areas.
- Nearby railways and highways provide crucial logistics for transporting rockets, fuel, and equipment.
- Sites are are usually located farther from cities to reduce risks to people and property.

Section 4

Build a Dashboard with Plotly Dash

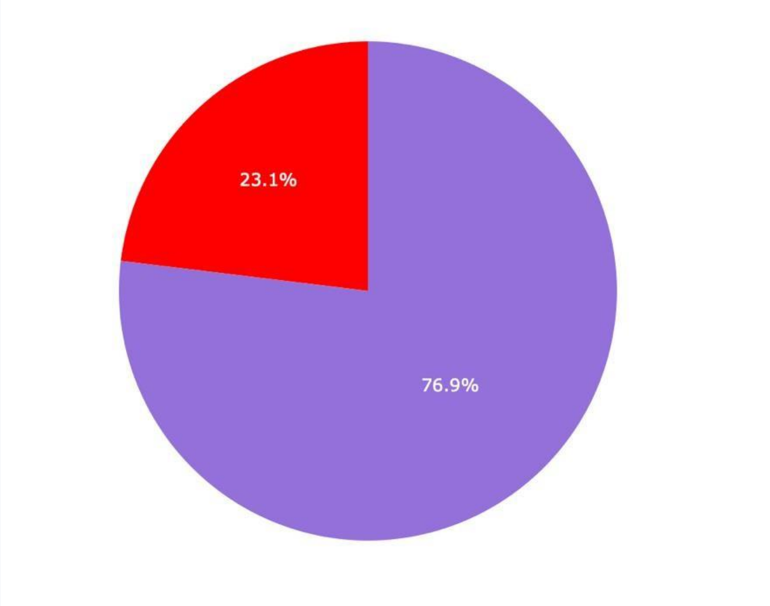
Launch Success Counts By Site

- This pie chart shows the distribution of successful launches across all SpaceX launch sites. Out of 24 total successful launches:
- **KSC LC-39A:** 41.7% of successes
- **CCAFS LC-40:** 29.2% of successes
- **VAFB SLC-4E:** 16.7% of successes
- **CCAFS SLC-40:** 12.5% of successes
- KSC LC-39A has the largest share of successful launches, while CCAFS SLC-40 contributes the least.



KSC LC-39A Launch Success Ratio

- The pie chart illustrates the launch success ratio for the KSC LC-39A, which has the highest success rate among SpaceX launch sites. Out of 13 total launches, **10 were successful** (approximately **77%**) and **3 were unsuccessful** (approximately **23%**).
- This visual clearly emphasizes the high reliability of KSC LC-39A launches compared to other sites.



Payload vs. Launch Outcome Scatter Plot

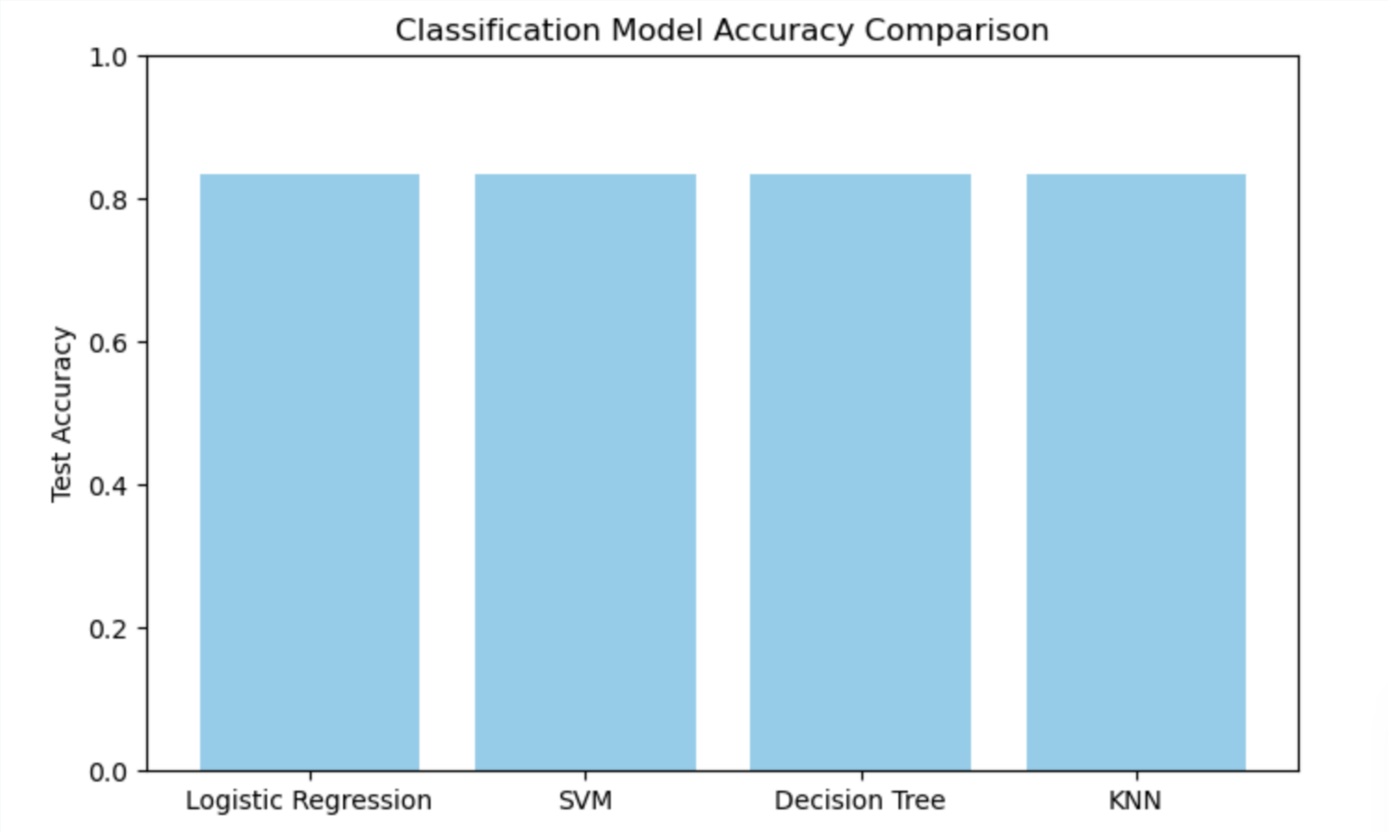
- Scatter plot shows how payload mass relates to launch success across all sites.
- Launches with payloads under ~6000 kg have a higher success rate.
- FT and B5 booster versions achieve the most consistent success.
- Range slider allows filtering by payload, revealing trends in success rates for different payload ranges.
- Point colors indicate booster type, helping compare performance visually across sites.

Section 5

Predictive Analysis (Classification)

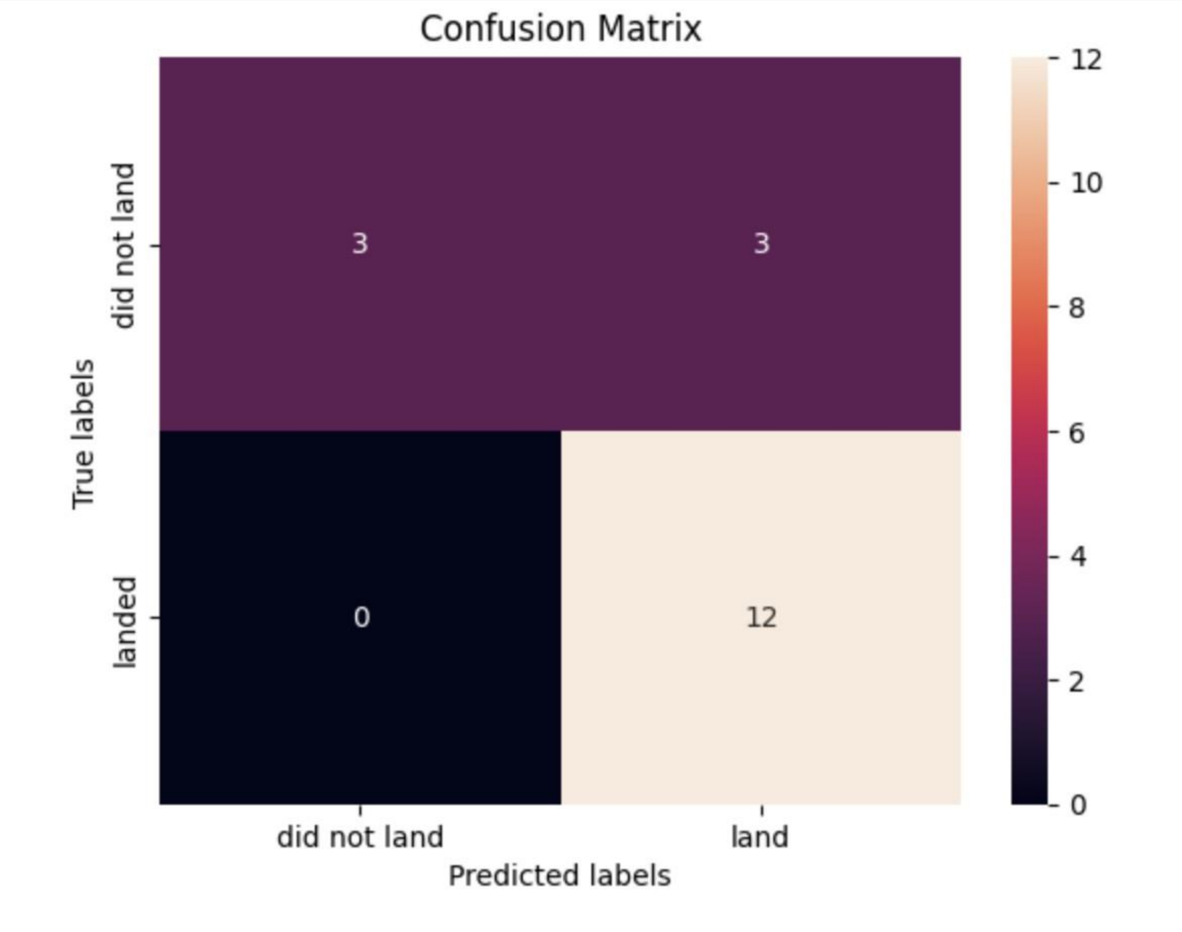
Classification Accuracy

- **Classification Model Accuracy:** A bar chart was created to compare the test accuracy of all built models (Logistic Regression, SVM, Decision Tree, KNN).
- **Observation:** All models achieved the same test accuracy of 0.8333, indicating that for this dataset and feature set, each model performed equally well.
- **Implication:** No single model outperformed the others, suggesting the dataset may be simple or balanced enough that multiple algorithms can achieve the same performance.



Confusion Matrix

- **Confusion Matrix:** Each classification model has the same confusion matrix: 12 true positives, 3 false positives, 3 true negatives, and 0 false negatives.
- **Explanation:** All models perform equally, correctly predicting most successful landings but making a few false positive errors.



Conclusions

- **Project Goal:** Predict SpaceX Falcon 9 first-stage landing success using historical launch data.
- **Data Preparation:** Loaded, cleaned, transformed, scaled, and split data into training and test sets.
- **Modeling:** Built and tuned Logistic Regression, SVM, Decision Tree, and KNN classifiers using GridSearchCV.
- **Model Accuracy:** While the Decision Tree achieved slightly higher accuracy on the training data (88.93%), all models including Logistic Regression, SVM, Decision Tree, and KNN, performed equally on the test data with 83.33% accuracy, and their confusion matrices were identical.
- **Insights:** Models can reliably predict landing outcomes, with most errors being false positives; Decision Tree may provide a slight advantage for real-world predictions.
- **Takeaway:** Classification models support cost analysis and risk assessment for reusable rocket launches.

Appendix

- API SpaceX data: https://api.spacexdata.com/v4/rockets/
- Wikipedia Falcon 9 launch dataset: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- GitHub Repo : https://github.com/akarayaz/Falcon9-LandingPrediction-Project

Thank you!