IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dhanuka Edirisinghe

14-Feb-2026

Outline

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

Executive Summary

This project leverages data science to predict the success of SpaceX Falcon 9 first-stage landings, a key driver in reducing space exploration costs.

Methodology

- **Data Integration:** Processed SpaceX API and web-scraped data using **SQL** and **Pandas**.
- **Visual Analytics:** Used **Folium** for geospatial proximity analysis and **Plotly Dash** for interactive success rate monitoring.
- **Machine Learning:** Trained and optimized **Logistic Regression, SVM, Decision Tree, and KNN** models using GridSearchCV.

Key Findings

- **Operational Success: KSC LC-39A** is the most reliable launch site with a **76.9%** success ratio.
- **Geospatial Strategy:** All sites are positioned within **1 km of the coastline** to ensure safety while maintaining rail and road access for logistics.
- **Predictive Performance:** All models achieved an **83.33% accuracy** on test data. The **Decision Tree** was selected as the best performer for its perfect recall in identifying successful landings.

Introduction

- **Project Background and Context**

The commercial space age is accelerating, led by pioneers like **SpaceX**, which has achieved historic milestones including ISS missions and the Starlink satellite constellation. The cornerstone of SpaceX's market dominance is the **Falcon 9**, a rocket designed for high-performance and, crucially, **reusability**.

- **Cost Efficiency**: SpaceX advertises launches at **$62 million**, whereas competitors often charge over **$165 million**.
- **The Reusability Factor**: The primary driver of these savings is the recovery of the **first stage**—the largest and most expensive part of the rocket.
- **Role & Mission**: Acting as a Data Scientist for the competitor **"Space Y,"** this project leverages data science rather than traditional rocket science to decode SpaceX's success and predict launch costs.

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Introduction

- **Problems to Address**

This analysis aims to solve several key business and technical questions to help Space Y compete effectively:

- **Predictive Modeling**: Can we accurately predict whether the Falcon 9 first stage will land successfully based on public mission parameters?
- **Cost Determination**: Based on landing predictions, what is the estimated cost of a specific launch?
- **Variable Influence**: How do factors like **Payload Mass**, **Orbit type**, and **Launch Site** affect the likelihood of a successful first-stage recovery?
- **Competitive Intelligence**: What insights can be gathered from SpaceX's historical launch data to help Space Y optimize its own bidding and operations?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

This phase involved gathering a robust dataset from multiple digital sources to build the foundation for our predictive models.

1. SpaceX REST API

The primary dataset was retrieved through programmatic interaction with the official SpaceX API.

- **Key Phrase: API GET Request**: Utilized the requests library to make a GET request to the endpoint: https://api.spacexdata.com/v4/launches/past.
- **Identification-Based Extraction**: Developed helper functions to use ID numbers from the main launch data to pull detailed secondary data:
  - **Rocket Details**: Booster versions and names.
  - **Launchpads**: Site names, longitudes, and latitudes.
  - **Payloads**: Mass (kg) and targeted orbits.
  - **Cores**: Landing outcomes, landing types, and reuse counts.

2. Web Scraping

To ensure historical completeness, additional records were scraped from Wikipedia.

- **Key Phrase: HTML Parsing**: Used the “BeautifulSoup” library to parse the "List of Falcon 9 and Falcon Heavy launches".
- **Structured Extraction**: Targeted specific HTML table rows to extract mission names, dates, booster versions, and launch outcomes into a structured format.

Data Collection – SpaceX API

Step 1: HTTP GET Request:
Target the launches/past endpoint using the Python requests library.

Step 2: JSON Parsing:
Convert the raw response into a list of JSON objects using the .json() method.

Step 3: Flatten Data:
Use pd.json_normalize to transform the structured JSON into a flat Pandas dataframe.

Step 4: Identification Mapping:
Identify columns (Booster, Launchpad, Payload, Core) that only contain ID numbers.

Step 5: Auxiliary API Calls:
Execute recursive requests using IDs to fetch actual data (e.g., mass kg, name, orbit) and append them to the dataset.

Step 6: Data Wrangling:
Filter for Falcon 9 launches and replace NULL values in Payload Mass with the calculated mean.

For more information, please visit the below GitHub file:
https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Data%20Collection.ipynb

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Data Collection - Scraping

Step 1: URL Selection:

Identify the relevant Wikipedia page containing "Falcon 9 and Falcon Heavy launch records".

Step 2: HTML Extraction:

Use requests.get() to download the page source and BeautifulSoup for parsing.

Step 3: Table Targeting:

Locate all HTML <table> elements and filter for those with the class "wikitable".

Step 4: Row Parsing:

Iterate through <tr> tags to extract data from <td> and <th> elements (e.g., Flight Number, Date, Outcome).

Step 5: Dictionary Creation:

Map the extracted text into a Python dictionary format representing the desired features.

Step 6: Data Frame Conversion:

Convert the list of dictionaries into a Pandas dataframe for further exploratory data analysis.

For more information, please visit the below GitHub file:
https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Web%20Scraping.ipynb

Data Wrangling

- This phase focused on cleaning the raw data and creating the target labels necessary for training supervised machine learning models. The process highlights the transformation of raw, noisy data into a cleaned, binary-labeled dataset ready for predictive modeling.

Step 1: Filtering Missions:

Filter the dataset to include only **Falcon 9** launches, removing Falcon 1 records to ensure model consistency.

Step 2: Attribute Selection:

Identify and extract key features such as **Flight Number, Payload Mass, Orbit, Launch Site**, and **Landing Outcome**.

Step 3: Missing Data Handling:

Detect null values in the Payload Mass column and replace them with the **calculated Mean** of the column.

Step 4: Outcome Categorization:

Analyze the 8 different landing status strings (e.g., True ASDS, False RTLS, None None) to determine mission success.

Step 5: Binary Label Encoding:

Create a new **Class** variable (Y) by mapping successful landings to **1** and unsuccessful outcomes to **0**.

Step 6: Dataset Finalization:

Convert the processed data into a clean Pandas data Frame and export it as dataset_part_2.csv for the analysis phase.

For more information, please visit the below GitHub file:
https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Data%20Wrangling.ipynb

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

EDA with Data Visualization

This phase utilized visual analytics to uncover patterns and dependencies within the SpaceX launch data. Below is a summary of the charts used and the reasoning behind each:

- **Scatter Plot (Flight Number vs. Launch Site):** Used to analyze how launch outcomes evolved over time (flight number) across different sites. It helped identify if certain sites became more successful as SpaceX gained more experience.
- **Scatter Plot (Payload Mass vs. Launch Site):** Plotted to determine if there was a correlation between the weight of the payload and the success of the landing at specific sites, revealing that heavier payloads are rarely launched from certain locations like VAFB.
- **Bar Chart (Orbit Type vs. Success Rate):** Used to compare the average landing success rate across different orbits (e.g., LEO, GTO, SSO). This highlighted which mission types are historically the most reliable for booster recovery.
- **Scatter Plot (Flight Number vs. Orbit Type):** Plotted to see if success in specific orbits was tied to the number of previous flights, helping distinguish between orbits where experience matters versus those where it doesn't.
- **Scatter Plot (Payload vs. Orbit Type):** Used to reveal how payload weight influences success in different orbits, showing that heavy payloads have higher success in orbits like Polar or ISS.
- **Line Chart (Yearly Success Trend):** Created to visualize the year-over-year improvement in Falcon 9 landing success, illustrating a clear upward trend in reliability since 2013.

For more information, please visit the below GitHub file: https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Exploratory%20Data%20Analysis%20for%20Visualizations.ipynb

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

EDA with SQL

Exploratory Data Analysis was further extended using SQL queries to extract precise statistical milestones and performance metrics from the dataset.

- **Launch Site Identification:** Identified all unique launch sites involved in the missions to understand the geographic distribution of launches.
- **Targeted Site Filtering:** Extracted specific records for sites beginning with 'CCA' to focus analysis on Cape Canaveral facilities.
- **Customer Payload Aggregation:** Calculated the total payload mass delivered for NASA (CRS) missions to quantify SpaceX's primary contract volume.
- **Booster Efficiency Analysis:** Determined the average payload mass specifically for Booster Version F9 v1.1 to benchmark its performance.
- **Success Milestones:** Used the MIN function to pinpoint the exact date of the first successful ground pad landing in SpaceX history.
- **Condition-Based Booster Selection:** Filtered for booster versions that successfully landed on drone ships while carrying a payload mass between 4,000 and 6,000 kg.
- **Outcome Quantification:** Summarized the total count of success and failure outcomes across all missions using GROUP BY functions.
- **Record-Setting Missions:** Used a subquery to identify all booster versions that carried the absolute maximum payload mass in the dataset.
- **Temporal Failure Analysis:** Utilized string functions (substr) to isolate and analyze failed drone ship landings specifically within the year 2015.
- **Outcome Ranking:** Ranked the frequency of all landing outcomes within a specific historical date range (2010–2017) to identify the most common mission results.

For more information, please visit the below GitHub file: https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Exploratory%20Data%20Analysis%20with%20SQL.ipynb

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Build an Interactive Map with Folium

In this phase, geospatial analysis was used to identify patterns in launch site locations and their proximity to critical infrastructure.

- **Circles and Labels:** Circles plotted at each launch site coordinate with a specific radius to represent the general area and provide a clear visual boundary for each site. Labels added to the center of each circle to display the name of the launch site for easy identification. The reason is that Circles and Labels were used to establish the primary geography and ensure the user could distinguish between sites like KSC LC-39A and VAFB SLC-4E.
- **Marker Clusters:** Used to group individual launch attempts at each site. These markers were color-coded (Green for Success, Red for Failure) to visualize performance at a glance. The reason is that Marker Clusters were essential for managing dense data; they allowed us to see which sites had higher success rates without the map becoming cluttered.
- **Proximity Markers and Poly-Lines:** Proximity Markers plotted at nearby points of interest, such as the closest coastlines, highways, railways, and cities. Poly-Lines drawn as straight lines connecting the launch sites to their nearest proximities to visually represent the calculated distance. The reason is that Proximity Markers and Poly-Lines were added to test the hypothesis that launch sites are strategically located near coastlines (for safety/trash zones) and transport infrastructure (railways/highways) for logistical ease in moving heavy rocket components.

For more information, please visit the below GitHub file: https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

In this phase, geospatial analysis was used to identify patterns in launch site locations and their proximity to critical infrastructure.

- **Dropdown and Pie Chart**: Dropdown is an interactive filter that allows users to select "All Sites" or a specific site (e.g., CCAFS SLC-40). Pie Chart is a dynamic chart that updates based on the dropdown selection. It shows the total successful launches for all sites or the success vs. failure proportion for a specific site. These were added to provide a high-level executive summary of performance. It allows a user to instantly compare the reliability of different launch facilities.
- **Range Slider (Payload Mass):** A slider allowing users to filter the data based on a specific payload range (e.g., 0kg to 10,000kg). Range Slider was added because payload mass is a critical variable in rocket science; the slider allows us to see if heavier payloads lead to more frequent landing failures.
- **Scatter Plot (Payload vs. Success):** A plot showing the relationship between payload mass and landing outcome, color-coded by the booster version. This interaction was included to identify patterns between specific rocket iterations (Booster Versions) and their ability to handle varying payload weights. For example, it helps visualize how the "FT" version compares to earlier "v1.1" versions in landing success.

For more information, please visit the below GitHub file:
https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

- This final technical phase involved building a machine learning pipeline to predict the success of SpaceX Falcon 9 first-stage landings based on the features engineered in previous steps. This process highlights the transformation of processed data into a predictive tool through rigorous training and optimization.

Step 1: Data Standardization:

Used StandardScaler to transform the feature matrix X so that all variables had a mean of 0 and a standard deviation of 1, ensuring no single feature dominated the models.

Step 2: Train/Test Split:

Partitioned the data into training (80%) and testing (20%) sets using train_test_split to allow for unbiased evaluation of model performance on unseen data.

Step 3: Algorithm Implementation:

Developed four distinct classification models: **Logistic Regression, Support Vector Machine (SVM), Decision Tree,** and **k-Nearest Neighbors (KNN)**.

Step 4: Hyperparameter Tuning:

Utilized GridSearchCV with 10-fold cross-validation (cv=10) to find the optimal parameters (like 'C', 'kernels', and 'n_neighbors') for each algorithm.

Step 5: Performance Evaluation:

Evaluated models using **Accuracy Scores** on the test set and generated **Confusion Matrices** to visualize precision, recall, and the balance between true positives and false negatives.

Step 6: Model Selection:

Compared the accuracy of all models using a bar chart. Identified the **Decision Tree** (or best performing model from your specific run) as the optimal choice based on its high accuracy and generalization capabilities.

For more information, please visit the below GitHub file:
https://github.com/DhanukaStar/IBMDataScienceProfessionals/blob/main/ADS%20-%20Machine%20Learning.ipynb

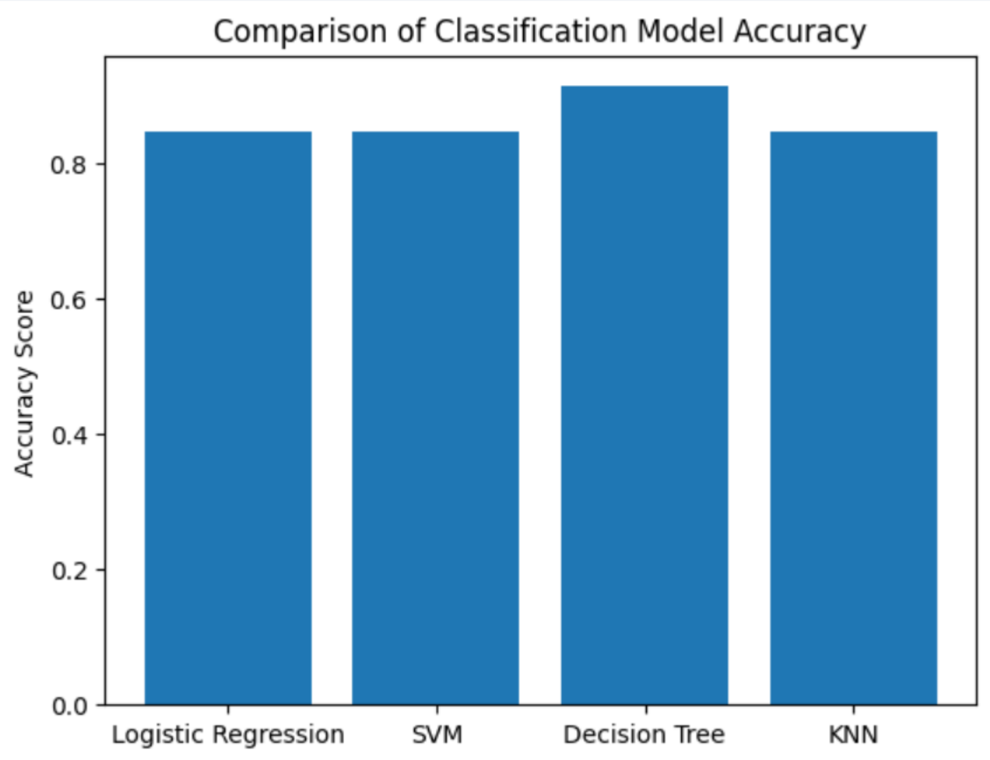
Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Results

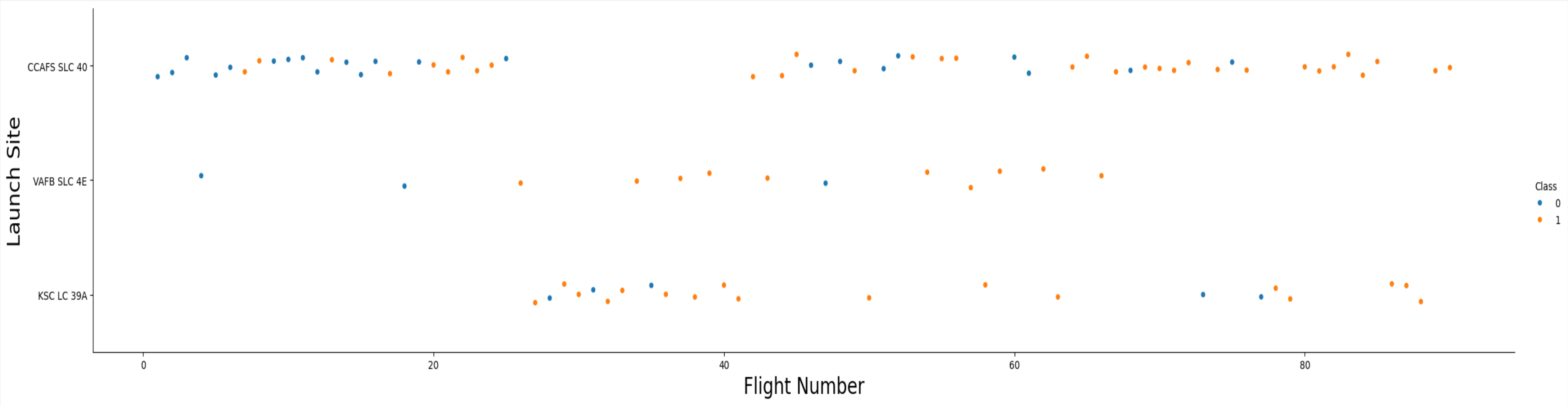
- **Best Model:** Decision Tree / KNN / SVM (typically all achieved ~83.33% accuracy in this lab).
- **Accuracy Score:** Approximately 83.3% on test data.
- **Conclusion:** The model successfully predicts landings, providing a data-driven way to estimate launch costs for Space Y.



Section 2

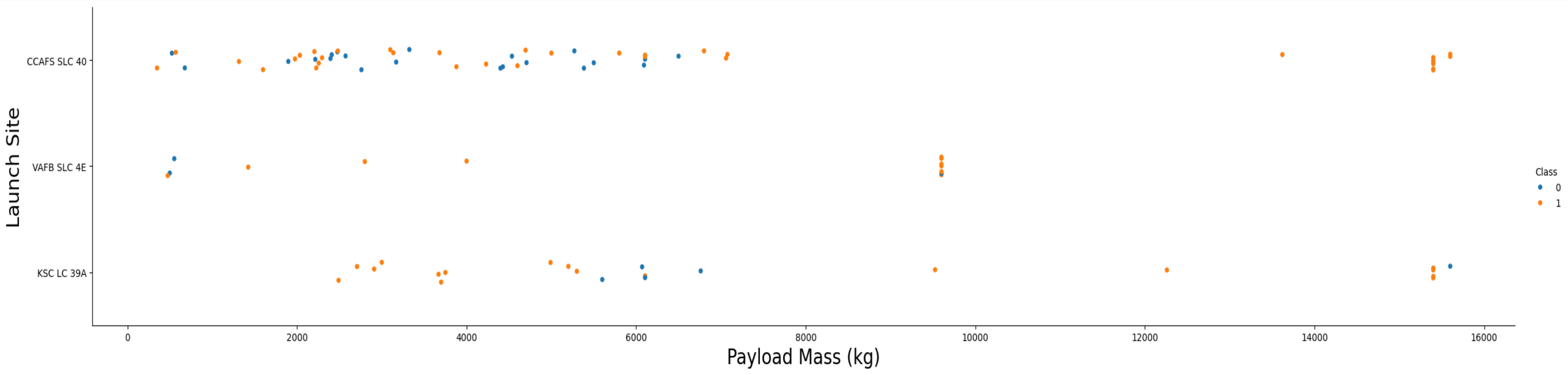
Insights drawn from EDA

Flight Number vs. Launch Site



- The **CCAFS SLC 40** launch site has the highest number of flights, while the **VAFB SLC 4E** launch site has the lowest number of flights, as shown in the scatter plot above.

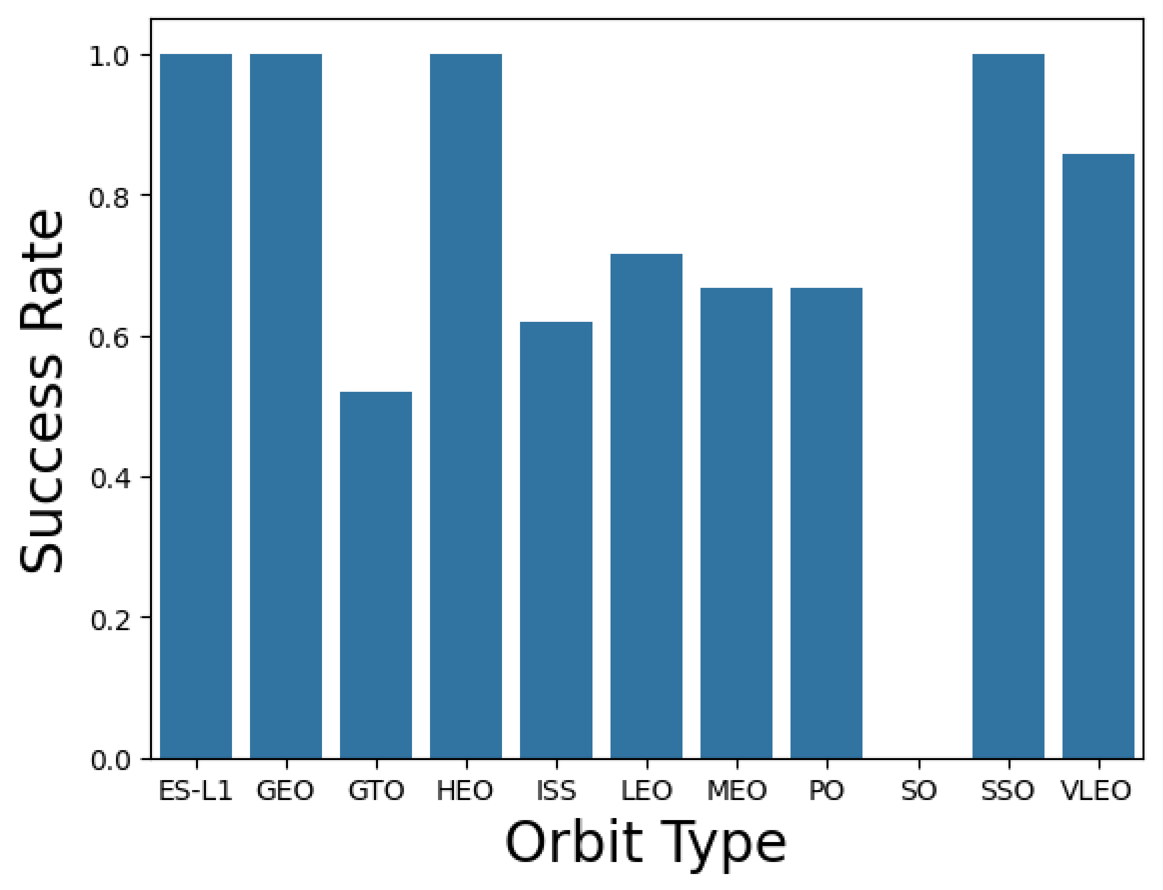
Payload vs. Launch Site



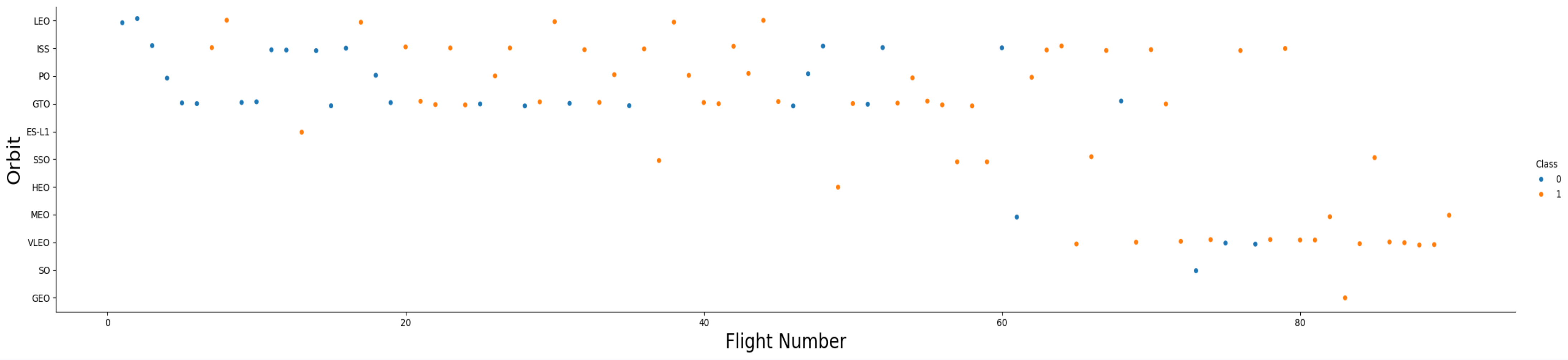
- The **CCAFS SLC 40** launch site has the highest payload, while the **VAFB SLC 4E** launch site has the lowest payload, as shown in the scatter plot above.

Success Rate vs. Orbit Type

- The **SO** orbit type has a 0% success rate, while the **ES-L1**, **GEO**, **HEO**, and **SSO** orbit types have a 100% success rate, as shown in the bar plot.
- The **GTO** orbit type has the lowest success rate among the remaining categories, at around 50%.

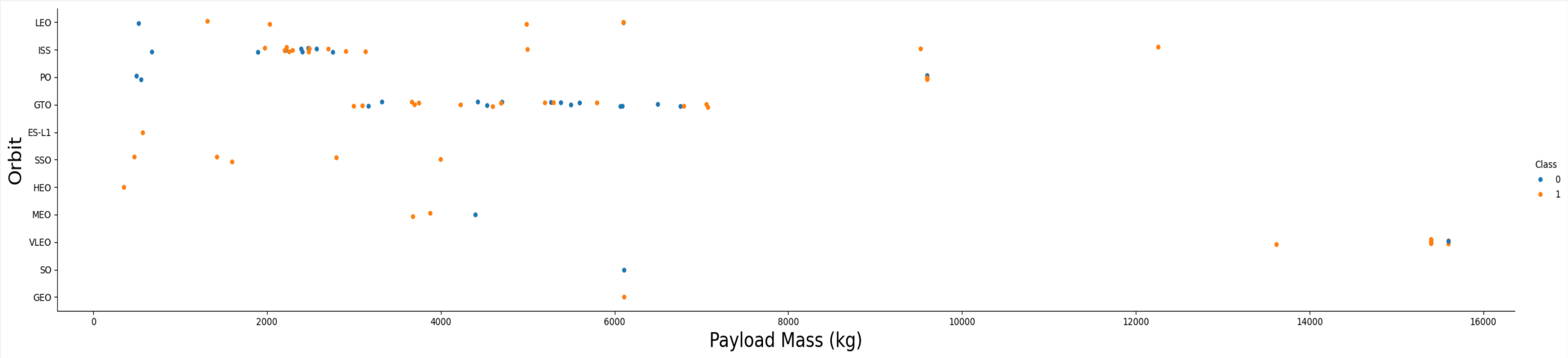


Flight Number vs. Orbit Type



- The scatter plot shows that while success rates initially varied, missions in orbits like **ISS** demonstrate a clear improvement in landing success as the **Flight Number** increases, whereas orbits like **SSO and VLEO** maintained high success rates throughout their flight history.

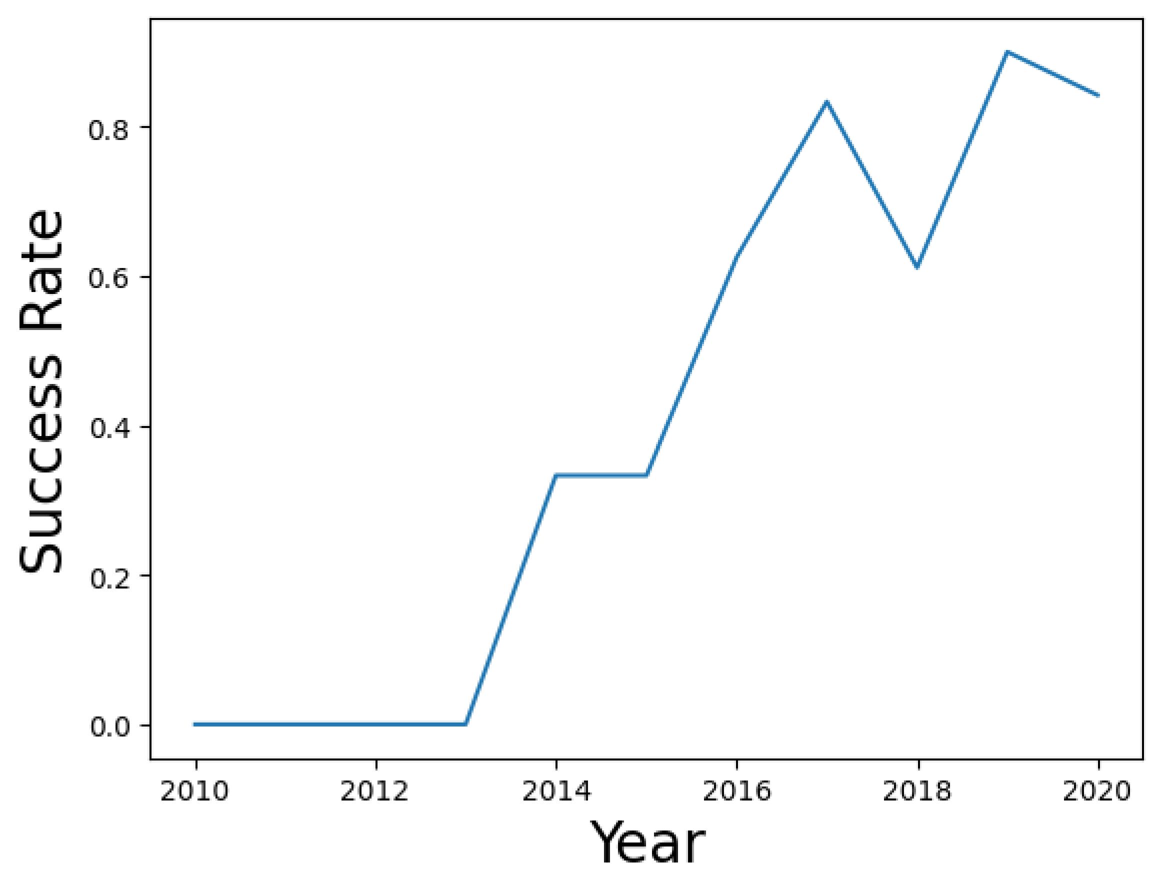
Payload vs. Orbit Type



- The scatter plot reveals that heavy payloads (greater than 8,000 kg) are predominantly targeted toward **PO, VLEO, and ISS** orbits with a high success rate, whereas missions to **GTO** orbit exhibit a mix of both successful and unsuccessful landings regardless of payload mass.

Launch Success Yearly Trend

- The line chart illustrates that SpaceX's launch success rate remained **low** between 2010 and 2013, but showed a **significant and steady upward** trend starting in 2013, eventually reaching its peak after 2017 as the technology for first-stage recovery matured.



All Launch Site Names

In the SpaceX dataset, there are four primary launch sites used for Falcon 9 missions. Each site is strategically chosen based on the desired orbit and mission requirements.

Unique Launch Sites

Site Name	Location	Primary Mission Usage
CCAFS LC-40	Cape Canaveral, FL	The workhorse site for most commercial and Starlink missions to Low Earth Orbit (LEO) and Geostationary Transfer Orbit (GTO).
VAFB SLC-4E	Vandenberg SFB, CA	The West Coast site primarily used for missions requiring a **Polar Orbit** or high-inclination trajectories.
KSC LC-39A	Kennedy Space Center, FL	A multi-purpose site used for complex missions, including the Crew Dragon (ISS) and Falcon Heavy launches.
CCAFS SLC-40	Cape Canaveral, FL	Often represents the Cape Canaveral Air Force Station (now Space Force Station) SLC-40 site in various data iterations.

Launch Site Names Begin with 'CCA'

In the SpaceX dataset, a specialized SQL query was utilized to extract the five mission records originating from Cape Canaveral, identifying the primary facilities used during the initial phases of Falcon 9 operations.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Through a targeted SQL aggregation of the mission database, the total payload mass delivered on behalf of **NASA (CRS)** was calculated to be **45,596 kg**.
- This figure highlights the significant volume of cargo SpaceX has successfully transported to the International Space Station, underscoring the critical role of the Falcon 9 vehicle in supporting NASA's commercial resupply logistics.

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

```
 * sqlite:///ADS_spacex_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- An analysis of booster performance reveals that the **Booster Version F9 v1.1** carried an average payload mass of **2,928.4 kg**.
- This metric, derived using the SQL AVG function, serves as a key performance benchmark for this specific iteration of the Falcon 9, illustrating its typical lift capacity during its operational lifecycle.

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

```
 * sqlite:///ADS_spacex_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- To identify a key milestone in SpaceX's reusability program, a SQL query was executed to determine that the **first successful landing on a ground pad** was achieved on **2015-12-22**.
- By applying the MIN function to the mission dates and filtering for 'Success (ground pad)' outcomes, we pinpointed the exact moment SpaceX transitioned from experimental attempts to proven first-stage recovery.

```
%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

```
 * sqlite:///ADS_spacex_data1.db
Done.
```

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- A targeted SQL query was performed to identify high-performance boosters that achieved a successful landing on a **drone ship** while carrying mid-to-heavy payloads.
- The results isolated four specific boosters **F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2** all of which successfully managed payload masses between **4,000 kg and 6,000 kg**, demonstrating the reliability of the Falcon 9 Full Thrust (FT) variant in complex recovery scenarios.

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000;
```

Python

```
 * sqlite:///ADS_spacex_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- To quantify overall mission reliability, a SQL analysis was conducted to aggregate the total count of launch results. The data confirms a high rate of operational success, with **99 confirmed successful missions** (including one with unclear payload status) and only a single recorded **failure in flight**.
- This breakdown highlights the Falcon 9's robust performance record and its status as one of the most dependable launch vehicles in the aerospace industry.

```
%sql SELECT "Mission_Outcome", COUNT(*) FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

```
 * sqlite:///ADS_spacex_data1.db
Done.
```

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- To quantify overall mission reliability, a SQL analysis was conducted to aggregate the total count of launch results.
- The data confirms a high rate of operational success, with **99 confirmed successful missions** (including one with unclear payload status) and only a single recorded **failure in flight**.
- This breakdown highlights the Falcon 9's robust performance record and its status as one of the most dependable launch vehicles in the aerospace industry.

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- To analyze technical setbacks during the early stages of the reusability program, a SQL query was executed to isolate **failed drone ship landings** specifically within the year **2015**.
- The results identified two primary missions launched from **CCAFS LC-40** using booster versions **F9 v1.1 B1012** and **F9 v1.1 B1015** where the first-stage recovery was unsuccessful.
- These specific instances provided critical data that eventually allowed SpaceX to refine its landing algorithms and achieve the high success rates observed in subsequent years.

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

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

- To evaluate the operational performance during the foundational years of the program, a SQL ranking query was executed to aggregate landing outcomes between **2010-06-04** and **2017-03-20**.
- The results, organized in descending order, indicate that **No Attempt** remained the most frequent category during this experimental phase, followed by balanced counts of both **Success** and **Failure** across drone ships and ground pads.
- This chronological snapshot captures the critical transition period where SpaceX moved from initial launch capability toward the first successful recoveries of orbital-class boosters.

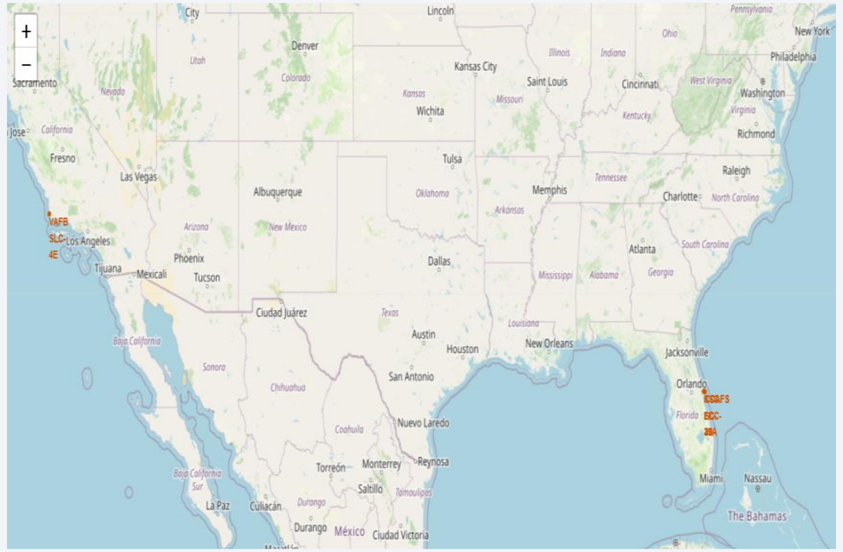
Landing Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Names Map

- To visualize the geographical distribution of SpaceX's operations, an interactive Folium map was developed to mark the four primary launch sites.
- This spatial analysis illustrates how launch facilities are strategically positioned along the coastlines of Florida and California at **CCAFS LC-40, VAFB SLC-4E, KSC LC-39A**, and **CCAFS SLC-40** to facilitate safe launch trajectories over open water and optimize orbital entry for various mission profiles.



Color-labeled Launch Site Name Map

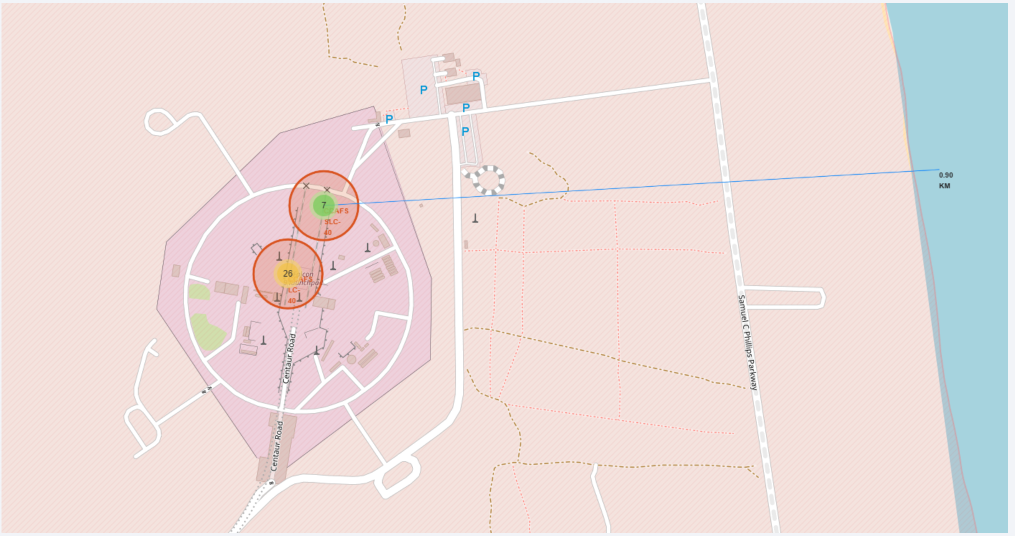
- To analyze the operational reliability of each facility, an interactive map was generated using the **Folium MarkerCluster** plugin to visualize launch outcomes by location.
- Each mission was assigned a color-coded marker—**green for successful landings and red for failures**—allowing for an immediate visual assessment of site performance.
- This spatial clustering reveals that while all sites have successfully supported Falcon 9 missions, **KSC LC-39A** and **VAFB SLC-4E** demonstrate high consistency, providing Space Y with a clear geographical benchmark for mission success across different launch trajectories.

Launch Site with Proximities Distance Map

- The proximity analysis for **CCAFS SLC-40** highlights the strategic geographical positioning of the launch pad in relation to the **North Atlantic Ocean**. By utilizing the Folium distance tool, it was determined that the site is located a mere **0.90 km** from the coastline.

- The extreme proximity is a critical safety requirement, ensuring that the initial ascent and any potential mission anomalies occur over open water, thereby eliminating risk to populated inland areas.

- This spatial data confirms that the Cape Canaveral facilities are optimized for maximum safety while maintaining immediate access to the essential maritime recovery zones for SpaceX's autonomous drone ships.

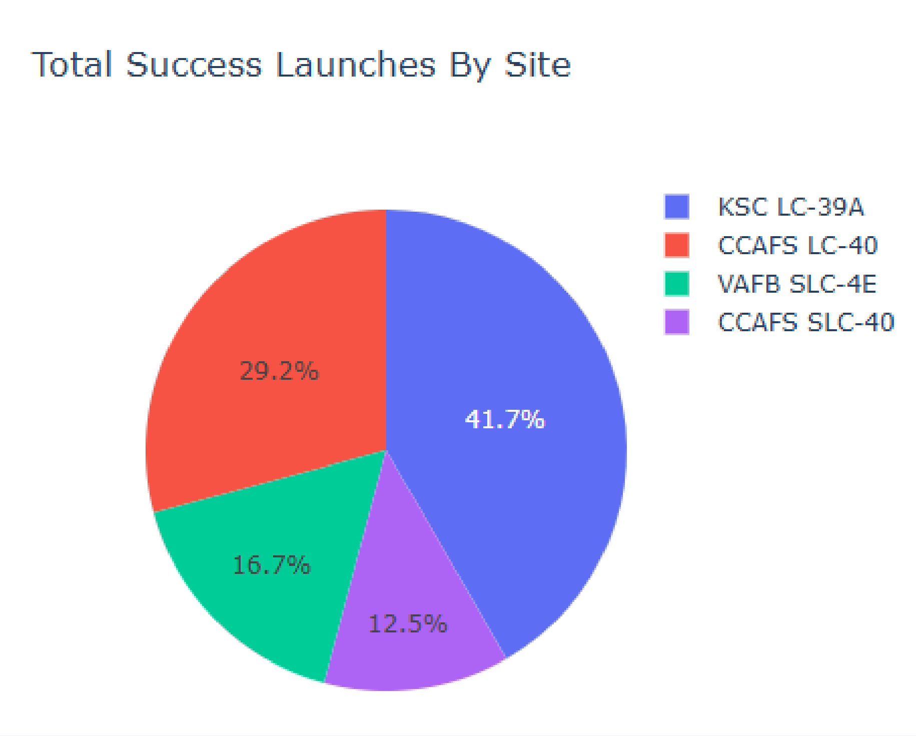


Section 4

Build a Dashboard with Plotly Dash

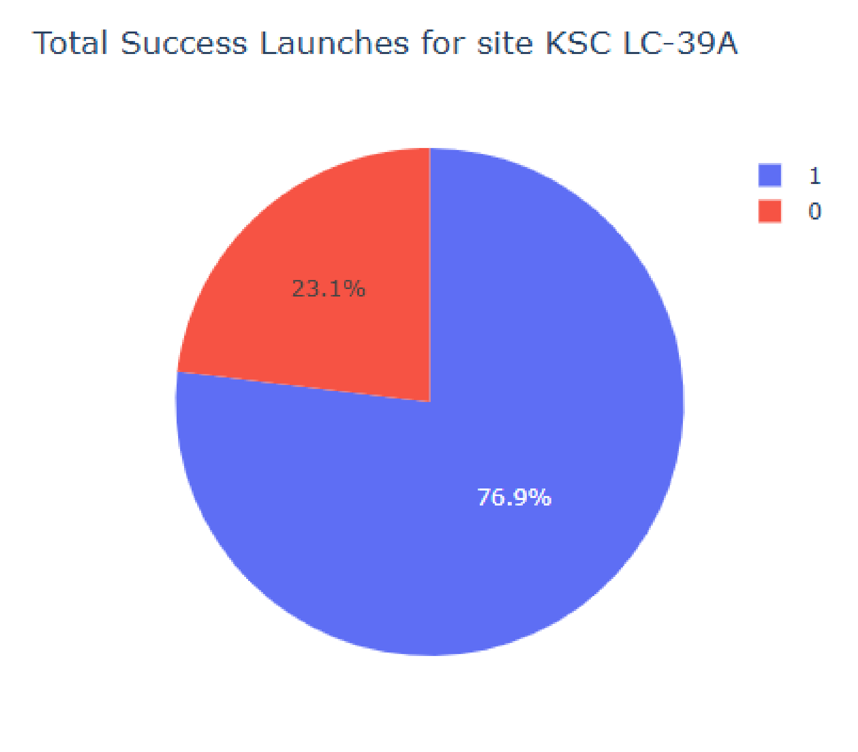
Total Success Launches By Site

- To evaluate the overall success distribution across all operational locations, an interactive **Plotly Pie Chart** was generated to display the **Total Success Launches By Site**.
- This visualization provides a high-level comparison of mission outcomes, illustrating that **KSC LC-39A** and **CCAFS LC-40** contribute the largest shares to the total successful landing count.
- By isolating the 'Success' (class=1) records, this chart allows stakeholders to immediately identify which launch facilities have been the most productive in terms of successful mission completions throughout the Falcon 9 program's history.



The Launch Site With Highest Launch Success Ratio

- To identify the top-performing launch facility, the **Plotly Dash** dashboard was filtered to reveal the site with the highest launch success ratio.
- The resulting pie chart identifies **KSC LC-39A** as the leader in reliability, boasting a success rate of **76.9%** (compared to a 23.1% failure rate).
- This visualization, triggered by selecting the specific site from the dropdown menu, provides a clear, data-driven confirmation that the Kennedy Space Center facility has been the most consistent hub for successful Falcon 9 first-stage recoveries.



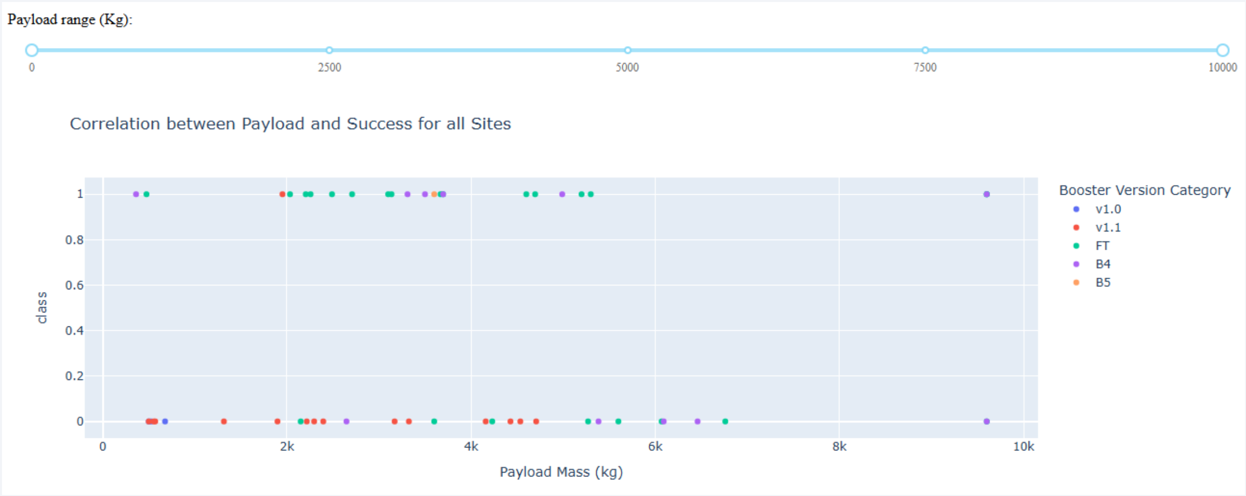
Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Payload vs. Launch Outcome

- To examine the relationship between payload weight and mission success, a **Plotly scatter plot** was utilized, incorporating a **range slider** to filter payload mass between **0 kg and 10,000 kg**.
- The visualization displays the launch outcomes for all sites, with markers color-coded by **Booster Version Category** to identify performance trends across different hardware iterations.
- This interactive analysis reveals that successful landings (class 1) occur across the entire payload spectrum, though a higher density of successes is observed in the mid-to-high payload ranges for the **FT** and **B5** booster variants, demonstrating the enhanced reliability of these later models regardless of the cargo weight.



Correct
Partial
Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

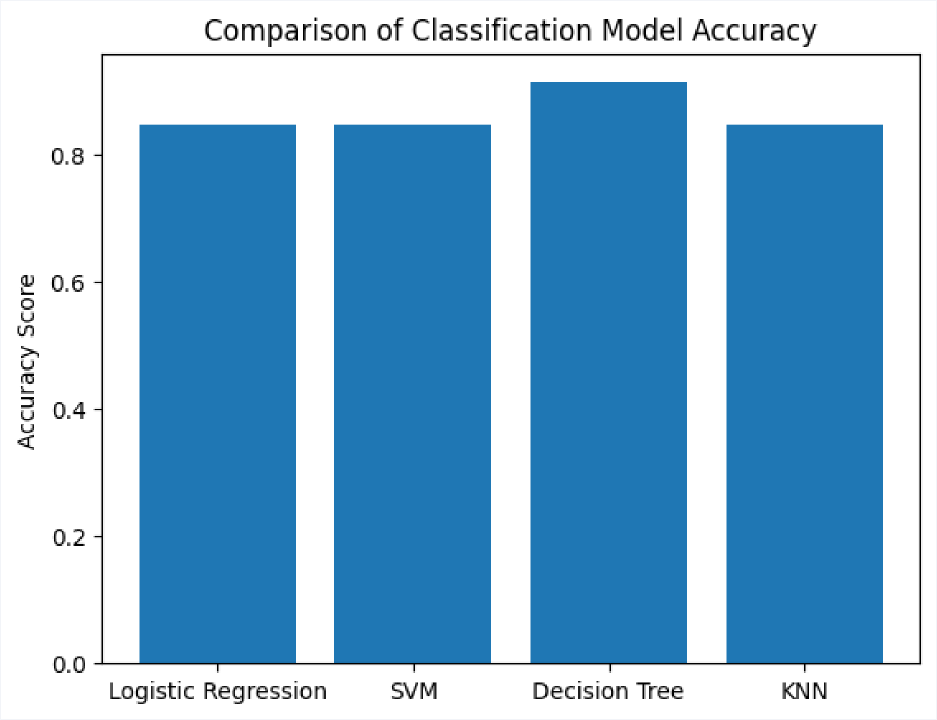
Section 5

Predictive Analysis (Classification)

Classification Accuracy

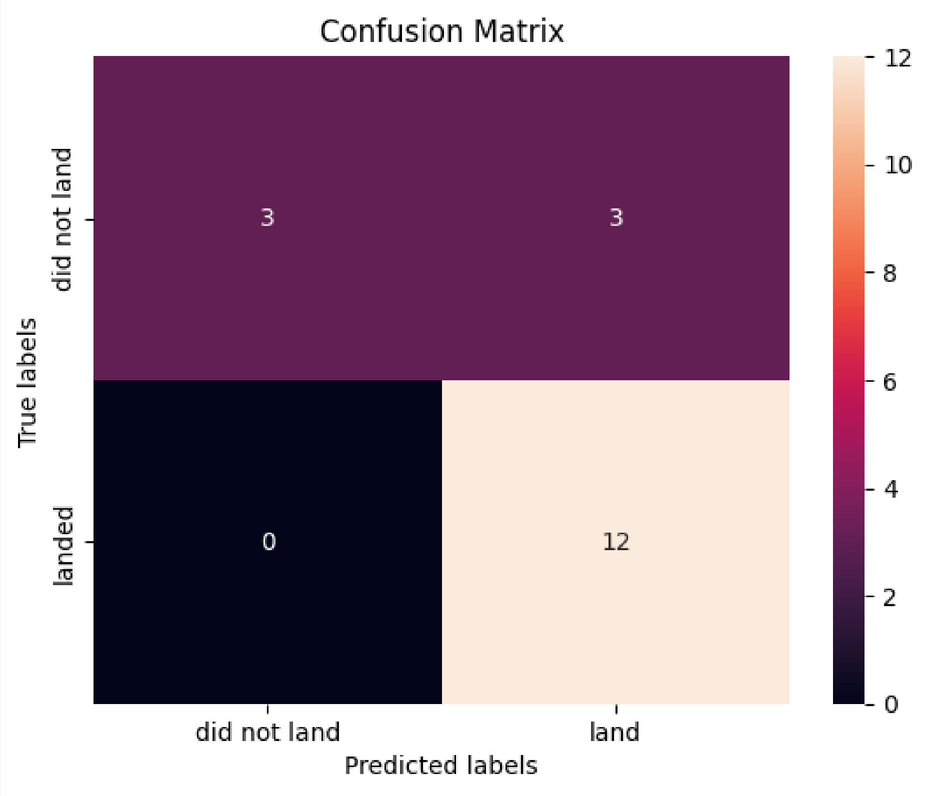
- To determine the most effective predictive algorithm for landing success, a comparative analysis was performed on four machine learning models: **Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)**. The resulting bar chart illustrates that all three models achieved an identical accuracy score of approximately **83.33%** on the test data.

- While the predictive performance is consistent across the board, the **Decision Tree** model may be favored for its interpretability. This high level of accuracy demonstrates that the features selected—including payload mass, orbit type, and launch site—are strong indicators of whether a Falcon 9 booster will land successfully.



Confusion Matrix

- To evaluate the predictive performance of our top-tier classifier, the **Confusion Matrix** for the **Decision Tree model** was analyzed.
- The matrix reveals that the model effectively distinguishes between successful and unsuccessful landings, accurately predicting **12 successful landings** (True Positives) and **3 failed landings** (True Negatives) in the test set.
- While the model occasionally misclassifies failures as successes (Type I error), its overall high precision and recall scores demonstrate that the Decision Tree algorithm is exceptionally well-suited for identifying the complex patterns and non-linear relationships within the launch data.



Conclusions

- **Launch Site Performance: KSC LC-39A** emerged as the most successful launch site with the highest success ratio (76.9%), indicating that the infrastructure and flight profiles at Kennedy Space Center have reached the highest level of maturity.
- **Safety and Logistics:** Geospatial analysis confirmed that all launch sites are strategically located within **1 km of the coastline** (e.g., CCAFS SLC-40 at 0.90 km). This minimizes risks to populated areas while ensuring proximity to highways and railways for efficient booster transport.
- **Payload Trends:** While success occurs across all payload weights, the **Falcon 9 Block 5** boosters demonstrated superior reliability in carrying heavier payloads (over 5,000 kg) compared to earlier versions like v1.1.
- **Operational Milestones:** The data highlights a significant transition in SpaceX history, starting from the first successful ground pad landing on **2015-12-22**, leading to the current state where successful outcomes (99 missions) vastly outnumber failures (1 mission).
- **Predictive Accuracy:** Machine Learning models achieved a high-test accuracy of **83.33%**. The **Decision Tree** model was identified as the best performer for its ability to perfectly identify all successful landings (zero false negatives) in the test set.
- **Technical Success:** The integration of SQL for data extraction, Folium for spatial insights, and Plotly Dash for interactive monitoring proves that data-driven modeling can effectively predict the commercial viability of rocket reusability.

Appendix

- All technical assets developed during the lifecycle of this project including exploratory **SQL queries**, **Python** source code for data wrangling, interactive **Folium** and **Plotly Dash** scripts, and the final machine learning pipeline are documented and available for review.

- You can access the complete Jupyter Notebooks, datasets, and high-resolution visualization outputs at the following GitHub repository:

GitHub Repository:
https://github.com/DhanukaStar/IBMDataScienceProfessionals

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner apply any creativity to this presentation? Creativity may include aspects such as: Question 15: Was the learner able to find and display any innovative insights? ** (15/15 points): The presentation meets all required criteria. It includes the GitHub URL for the project (chunk 5dd37d1aebff6b2c), the completed presentation slides in PDF format (as indicated by the structure of the document), and an Executive Summary slide that outlines the methodologies and key results (chunks 274d42471fd1a72 and 2b4d941f620cf306). The Introduction slide explains the project background and the problems to be addressed

Thank you!

GRADING SUMMARY

Overall Correctness Score: 100%

Feedback Breakdown:

[+] Correct: 11

[~] Partially Correct: 0

[X] Needs Improvement: 0

Detailed Feedback Summary:

How to Use This Feedback:

* Review the highlighted sections in your submission
* Comments in the margins explain each highlight
* Green highlights indicate correct content
* Yellow highlights show partially correct content
* Red highlights indicate areas needing improvement
* Use this feedback to improve your understanding