IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aditya Arya
18.12.2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of Methodologies

1. Data Collection: Accessed SpaceX launch data via API and web

scraped records from Wikipedia.

2. Data Cleaning & Preparation:

- Cleaned and formatted the data.
- Stored data in Db2 database and performed SQL queries.
- Conducted exploratory data analysis.

3. Feature Engineering: Created new features and standardized

the data.

4. Interactive Visualizations:

- Mapped launch sites and success rates using Folium.
- Built an interactive dashboard with Plotly Dash.

5. Model Building & Evaluation:

- Implemented SVM, Decision Trees, and K-Nearest Neighbors.
- Tuned hyperparameters with GridSearchCV.
- Evaluated models using test data accuracy.

Summary of Results

1. Data Insights:

- Identified factors influencing Falcon 9 first stage landings.
- Visualized geographical patterns and success rates.

2. Model Performance:

- SVM and K-Nearest Neighbors: 83.33% accuracy.
- Decision Tree: 94.44% accuracy.

3. Key Findings:

- Launch site and payload mass impact landing success.
- Decision Tree model is the most effective predictor.

Introduction

Project background and context

- SpaceX has revolutionized the aerospace industry by reducing launch costs through the reuse of the Falcon 9 first-stage booster.
- Predicting whether the first stage will successfully land is crucial for cost estimation, mission planning, and competitive decision-making.
- Data science and machine learning provide powerful tools to analyze historical launch data and extract actionable insights.

Problems We Want to Answer

- Which factors influence the successful landing of the Falcon 9 first stage?
- How do launch site, payload mass, and mission characteristics affect landing outcomes?
- Which machine learning model performs best in predicting landing success?

Section 1

Methodology

Methodology

Executive Summary

Data Collection Methodology

- SpaceX launch data was collected using the official SpaceX REST API.
- Historical launch records were additionally obtained through web scraping from Wikipedia. Data from both sources was combined
- into a single dataset for analysis.

Data Wrangling

- Raw data was cleaned by handling missing values and correcting data types. Datasets were merged and
- standardized to ensure consistency.
- Processed data was stored in a database and prepared for analysis.

Exploratory Data Analysis (EDA)

- Visualizations were used to analyze launch outcomes, payload mass, and launch sites. SQL queries were performed to
- compute success rates and summarize launch statistics.

Interactive Visual Analytics

- Interactive maps were created using Folium to visualize launch sites and landing success.
- An interactive dashboard was built with Plotly Dash to explore payload mass and success rates.

Predictive Analysis

- Classification models including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Tree were developed.
- Models were tuned using GridSearchCV and evaluated based on prediction accuracy.

Data Collection

SpaceX REST API

Initiate API Request → Fetch Launch Data → Store Data Locally

Wikipedia web scraping

Extract HTML Table → Parse with BeautifulSoup → Convert to DataFrame

Data Integration Flow

SpaceX API Data & Wikipedia Data → Merge Datasets → Final Integrated Data

Data Collection – SpaceX API

Step 1: Initiate API Request

- Use Python's `requests` library to connect to the SpaceX API.
- Endpoint: `https://api.spacexdata.com/v4/launches`

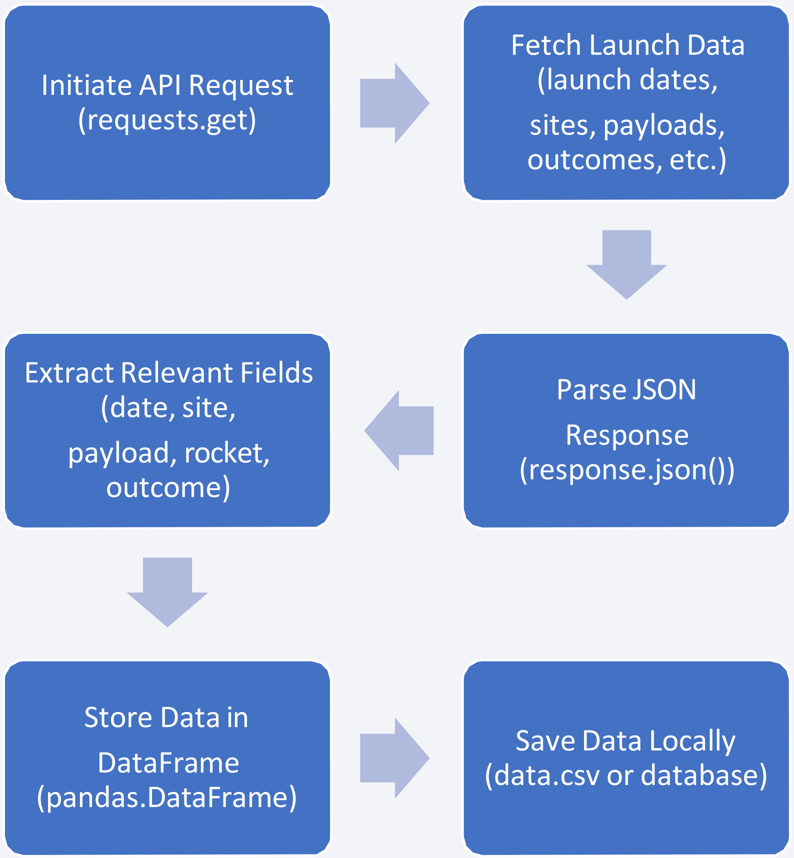
Step 2: Parse API Response

- Convert API response from JSON to a Python dictionary.
- Extract relevant fields: launch date, launch site, payload mass, rocket type, outcome.

Step 3: Store Data Locally

- Save extracted data into a pandas DataFrame.
- Store the DataFrame locally for further processing.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/1.%20jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

Web Scraping Data Collection

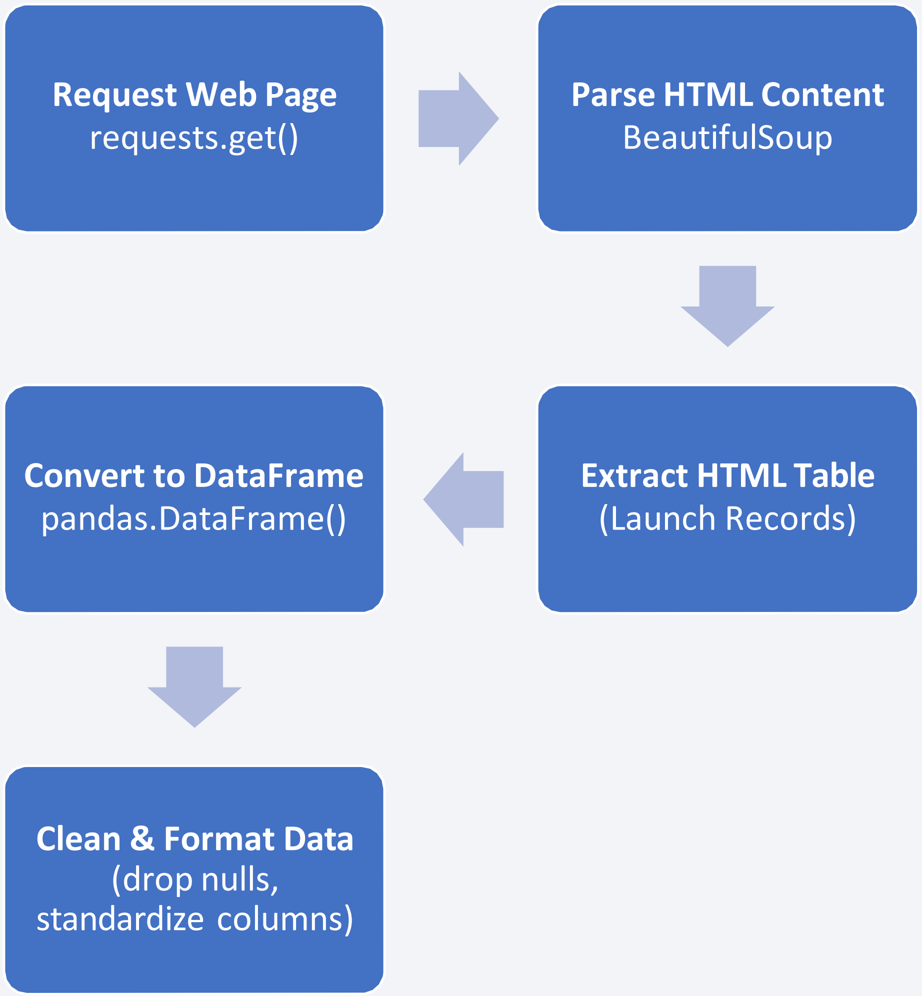
•Historical SpaceX launch data was collected through **web scraping** from Wikipedia.

•Python’s requests library was used to retrieve the HTML content of the target web page.

•The HTML tables containing launch records were parsed using **BeautifulSoup**.

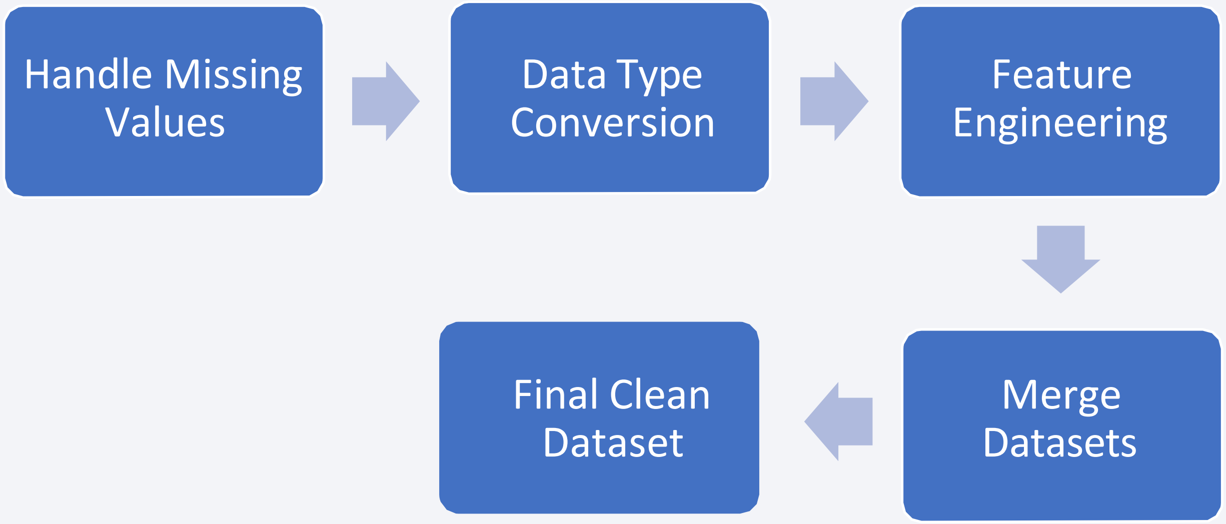
•Extracted data was cleaned and converted into a structured **pandas DataFrame** for analysis.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/2.%20jupyter-labs-webscraping.ipynb



Data Wrangling

Data Processing Overview
Collected datasets were cleaned and prepared to ensure data quality and consistency.
Missing values were handled, incorrect data types were corrected, and irrelevant columns were removed. Multiple datasets were merged into a single, structured dataset for analysis.



Data Wrangling Process
Identify and handle missing values
Convert data types (dates, numerical values)

Standardize column names and formats
Remove duplicates and validate data accuracy Create new features required for analysis

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/3.%20labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Charts Used and Purpose

•**Histograms** were used to analyze the distribution of numerical variables such as payload mass and flight number, helping to identify data spread and outliers.

•**Bar charts** were created to compare landing success and failure rates across different launch sites and orbit types.

•**Line charts** were used to observe trends in landing success over time and across flight numbers.

•**Scatter plots** were applied to explore the relationship between payload mass and landing success.

•**Heatmaps** were used to visualize correlations between numerical features and support feature selection.

•**Box plots** helped compare distributions and detect outliers across different categories.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/5.%20edadataviz.ipynb

EDA with SQL

SQL Queries Performed

- Calculated the **total number of launches** and summarized mission outcomes.
- Counted **successful and unsuccessful landings** to analyze overall performance.
- Computed **success rates by launch site** to compare site reliability.
- Queried **total and average payload mass** by rocket and booster version.
- Used **filtering queries** to analyze launches within specific date ranges and conditions.
- Applied **sorting and ranking queries** to identify trends and top-performing launch configurations.
- Used **subqueries and joins** to combine launch data with additional rocket information for deeper analysis.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/4.%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Objects Created

- **Markers** were added to represent the geographic locations of SpaceX launch sites.
- **Color-coded markers** were used to distinguish between successful and unsuccessful landings.
- **Circles** were drawn around launch sites to visualize proximity areas and operational zones.
- **Lines (Polylines)** were added to show distances between launch sites and nearby coastlines or relevant locations

Purpose of Adding Map Objects

- Markers help clearly identify where SpaceX launches take place geographically.
- Color coding enables quick comparison of landing success rates across different launch sites.
- Circles provide spatial context and highlight proximity-related factors that may influence launch operations.
- Lines help visualize distances and spatial relationships, improving geographic understanding of launch conditions.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/6.lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Plots and Graphs Added

- A **pie chart** was created to display the overall distribution of successful and unsuccessful Falcon 9 landings.
- A **scatter plot** was used to visualize the relationship between payload mass and landing success.

Interactions Added

- A **launch site dropdown menu** was implemented to allow users to filter data by specific launch sites.
- A **payload mass range slider** was added to dynamically explore how payload weight affects landing success.

Purpose of Plots and Interactions

- The pie chart provides a quick overview of mission success rates.
- The scatter plot helps identify patterns between payload mass and landing outcomes.
- Interactive filters enhance user exploration and enable focused analysis of specific launch conditions.

Predictive Analysis (Classification)

Model Development Process

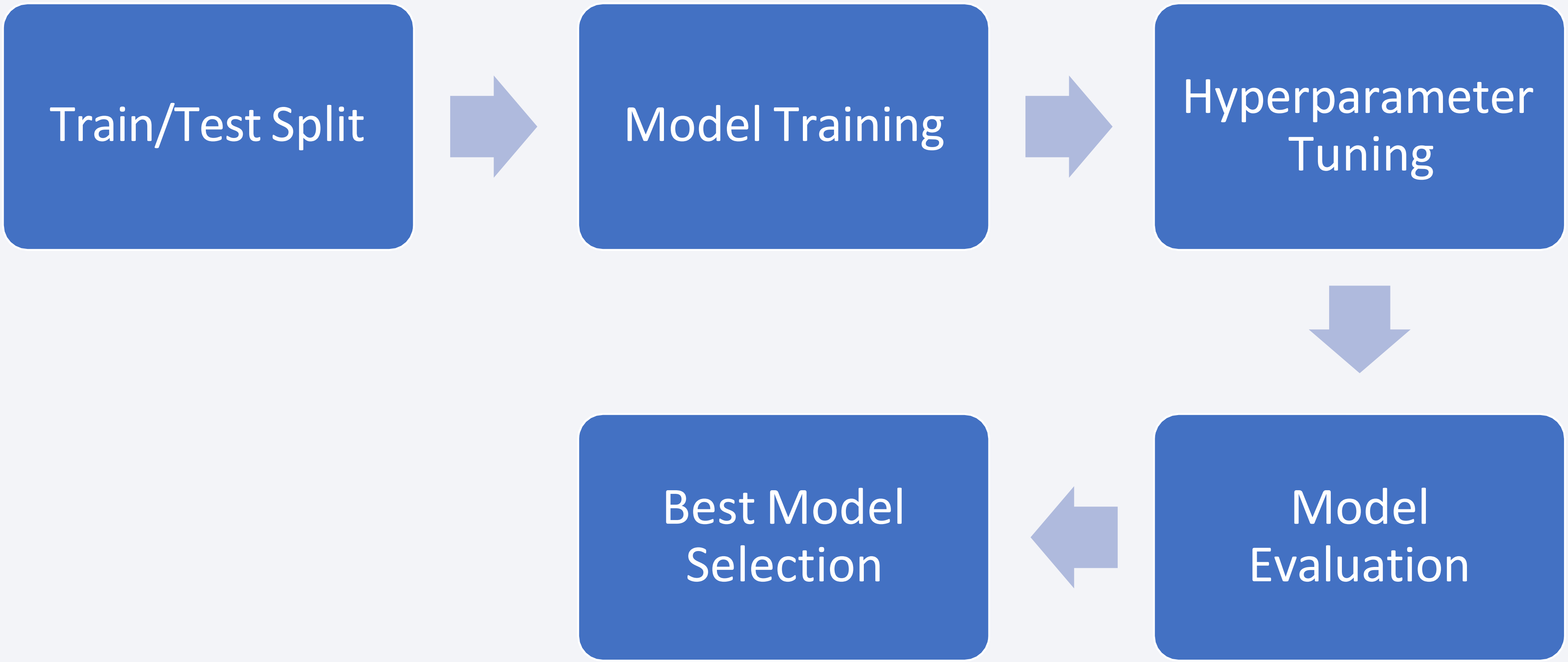
- The dataset was first **preprocessed** by scaling numerical features and splitting the data into training and test sets.
- Multiple classification models were built, including **Logistic Regression, Support Vector Machine (SVM), K- Nearest Neighbors (KNN), and Decision Tree**.
- Models were evaluated using **test data accuracy** and cross-validation to ensure reliable performance.

Model Improvement and Selection

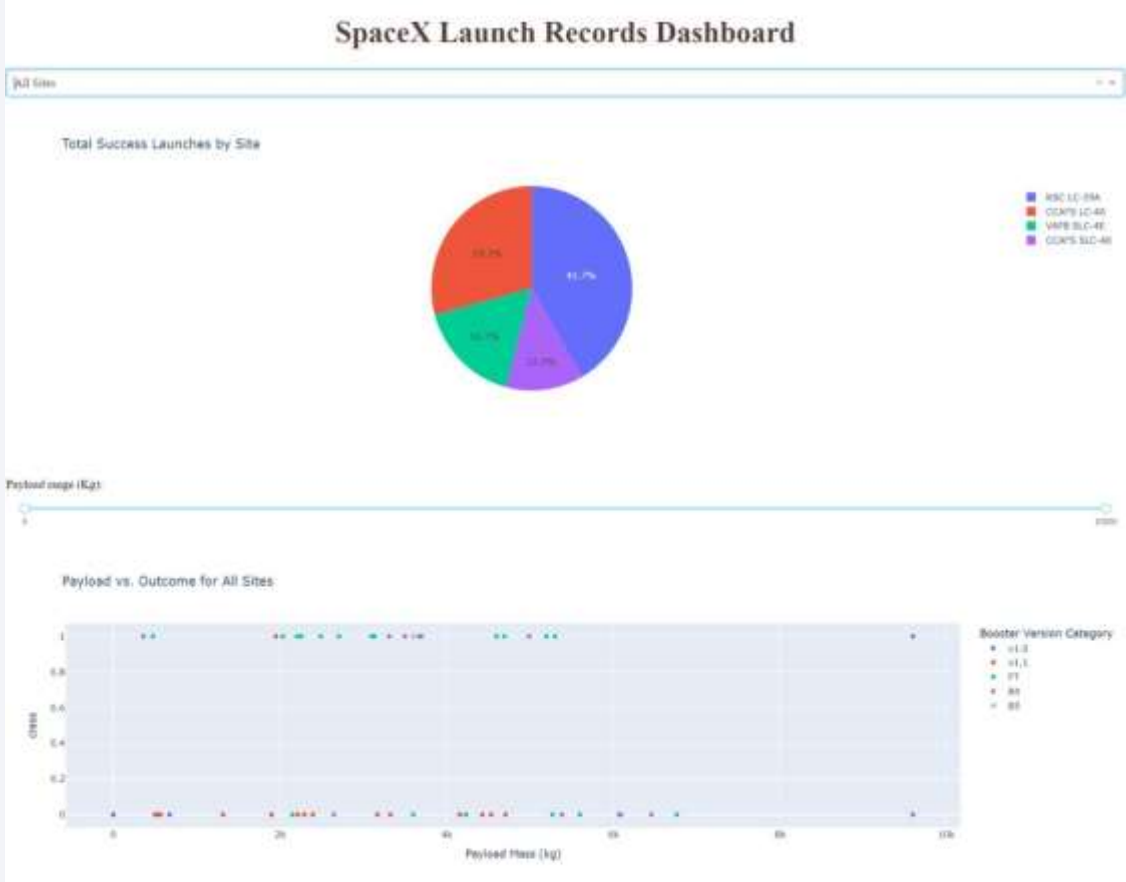
- **GridSearchCV** was used to tune key hyperparameters for each model.
- Model performance was iteratively improved by comparing results across different parameter settings.
- The **Decision Tree model** achieved the highest accuracy and was selected as the best-performing model.

GitHub Source:
https://github.com/Porush123/IBM_DS_Project/blob/main/7.%20SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)



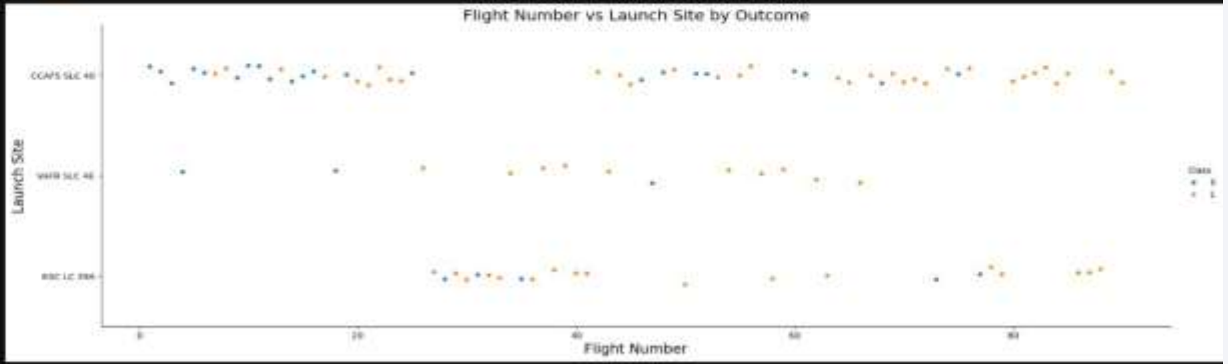
Results



TASK 1: Visualize the relationship between Flight Number and Launch Site

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(x="FlightNumber", y="LaunchSite", hue="Class", data=df, aspect=3, height=5, kind="strip")
plt.xlabel("Flight Number", fontsize=16)
plt.ylabel("Launch Site", fontsize=16)
plt.title("Flight Number vs Launch Site by Outcome", fontsize=18)
plt.show()
```



TASK 12

Find the method performs best:

Based on the results I got:

1. **Logistic Regression**: Accuracy on test data: 0.8333
2. **Support Vector Machine (SVM)**: Accuracy on test data: 0.8333
3. **Decision Tree**: Accuracy on test data: 0.9444
4. **K Nearest Neighbors (KNN)**: Accuracy on test data: 0.8333

It seems there was a significant increase in accuracy for the Decision Tree model, which shows the highest accuracy of 0.9444 on the test set. Therefore, in this comparison, the Decision Tree model performs the best among the classifiers you've trained.

Section 2

Insights drawn from EDA

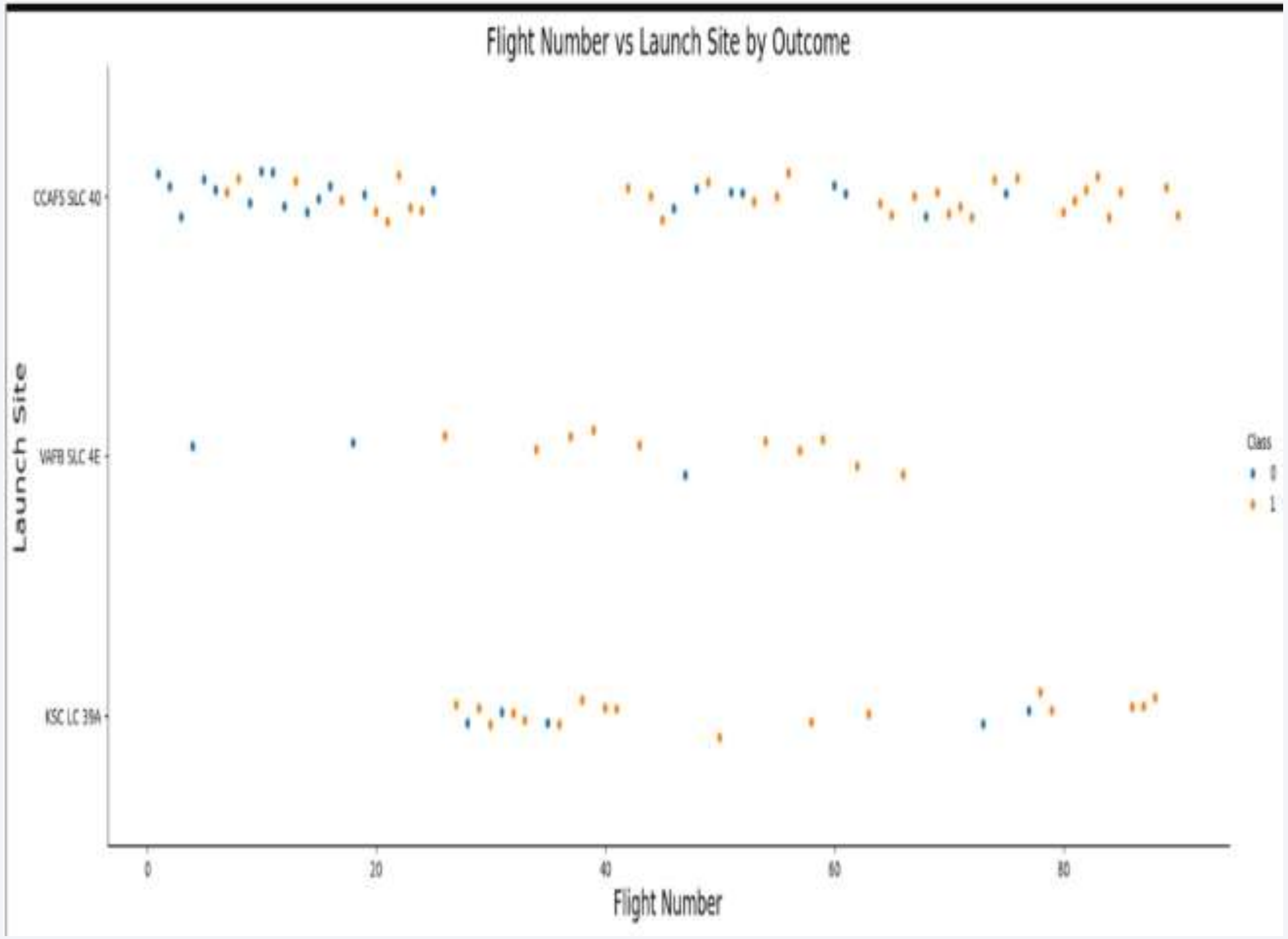
Flight Number vs. Launch Site

Scatter Plot Analysis

- The scatter plot shows the relationship between **flight number** and **launch site** for Falcon 9 missions.
- Launches are distributed across all major sites (CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E), indicating consistent launch activity over time.
- Both **successful and unsuccessful landings** appear at each launch site, suggesting that landing success is influenced by factors beyond the launch site alone.

Key Insights

- No clear trend indicates that higher flight numbers are limited to a specific launch site.
- As flight numbers increase, successful landings become more frequent, reflecting **operational learning and improvements over time**.
- The scatter plot highlights that launch experience plays a more important role than launch site location in determining landing success.



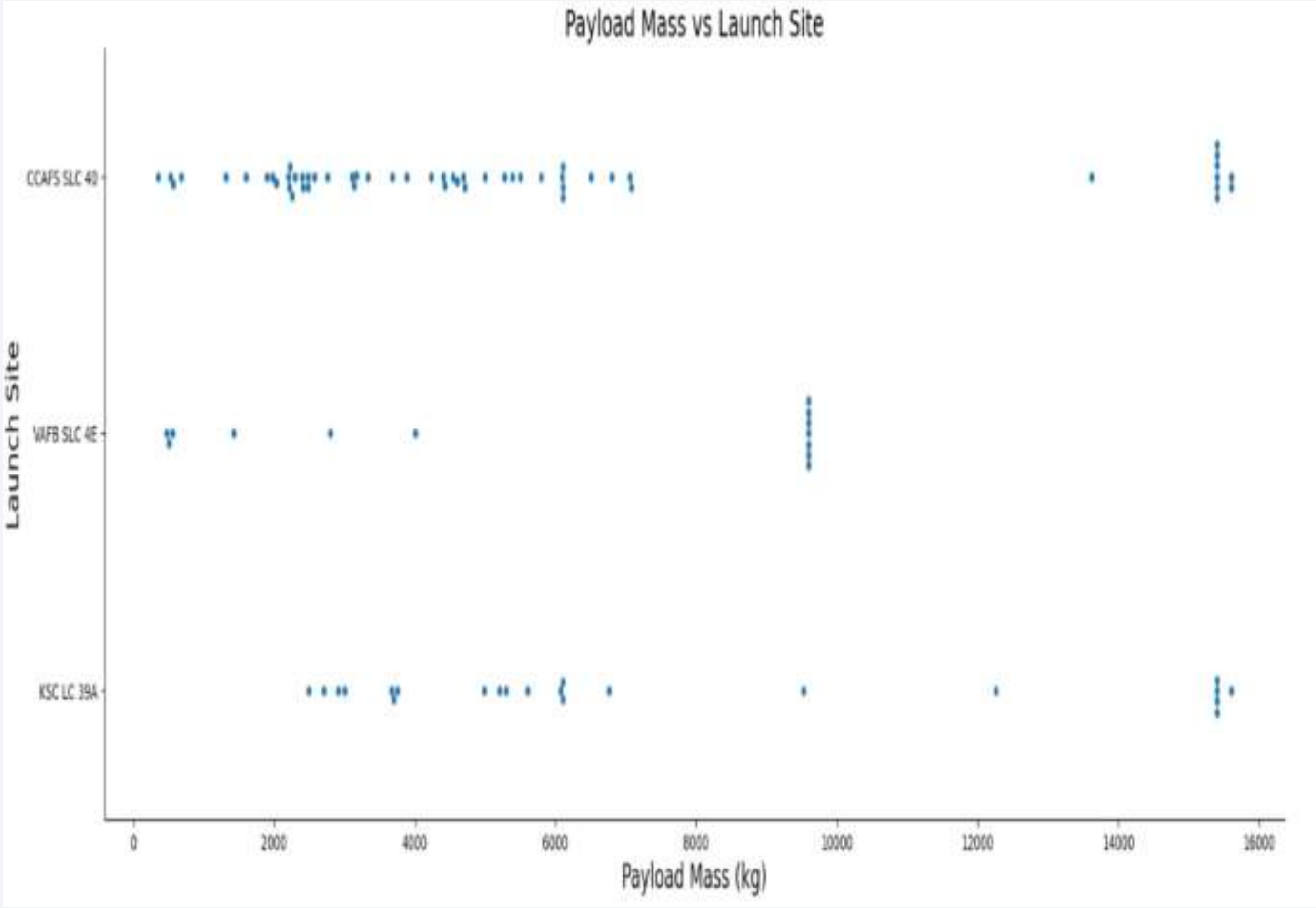
Payload vs. Launch Site

Scatter Plot Analysis

- The scatter plot illustrates the relationship between **payload mass** and **launch site** for Falcon 9 missions.
- Most launches from **CCAFS SLC-40** carry payloads below 10,000 kg, indicating frequent medium-weight missions.
- **KSC LC-39A** shows a wider range of payload masses, including several high-payload launches exceeding 15,000 kg, suggesting its suitability for heavy missions.
- **VAFB SLC-4E** handles a more limited range of payload masses, reflecting its specialized mission profile.

Key Insights

- Higher payload missions are more commonly launched from **KSC LC-39A**.
- Payload mass distribution varies significantly by launch site, indicating that **mission requirements influence launch site selection**.
- The scatter plot shows that payload mass alone does not determine mission success but is an important operational factor.



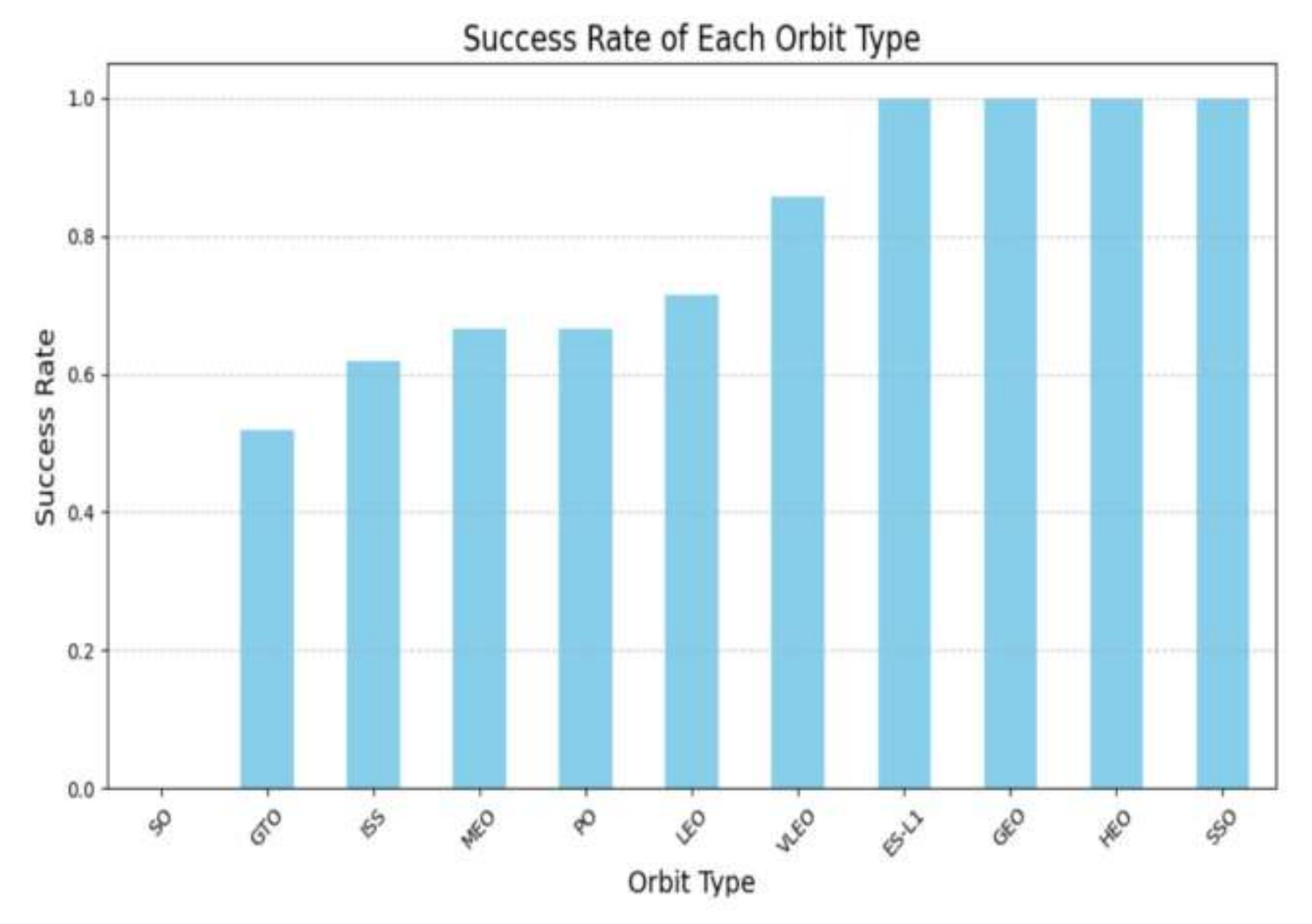
Success Rate vs. Orbit Type

Bar Chart Analysis

- The bar chart illustrates the **success rate of Falcon 9 first-stage landings across different orbit types**.
- Orbits such as **VLEO, SSO, GEO, and ES-L1** show very high or near-perfect success rates, indicating stable and well-optimized mission profiles.
- **GTO missions** exhibit a noticeably lower success rate compared to other orbit types, suggesting higher mission complexity and greater operational challenges.

Key Insights

- Landing success varies significantly by **orbit type**, highlighting its importance as a predictive feature.
- Higher-energy orbits like **GTO** are associated with increased landing difficulty.
- These findings support the inclusion of orbit type as a key variable in predictive modeling.



Flight Number vs. Orbit Type

Flight number Analysis

•The scatter plot shows the relationship between **flight number** and **orbit type** for Falcon 9 missions.

•Early flight numbers are mostly associated with a limited set of orbit types, while later flights cover a **broader range of orbits**, indicating mission diversification over time.

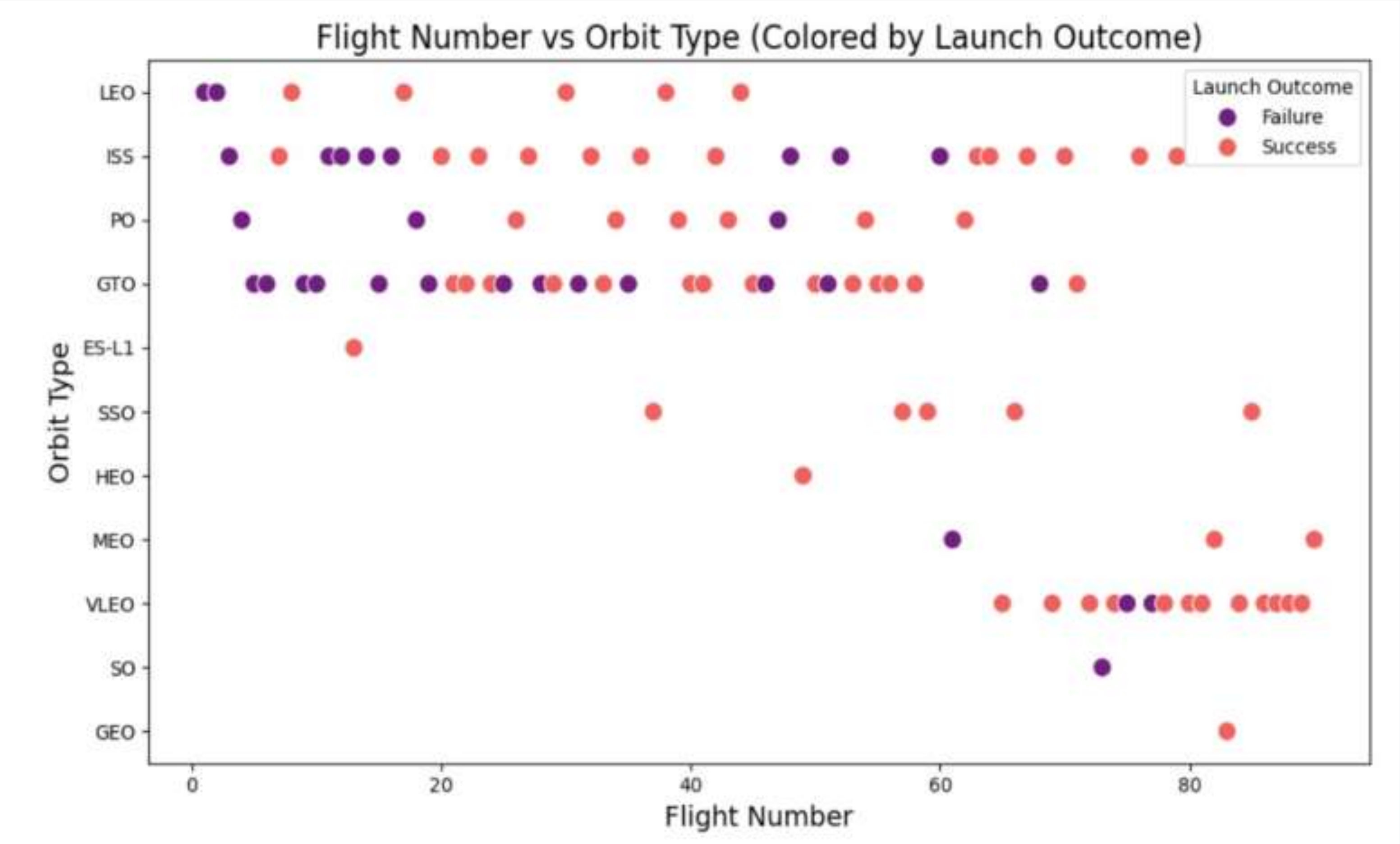
•Successful landings become more frequent at higher flight numbers across most orbit types, reflecting increased experience and operational improvements.

Key Insights

•No single orbit type dominates across all flight numbers.

•Higher flight numbers are correlated with improved landing success, regardless of orbit type.

•The plot highlights that **operational experience** is a critical factor influencing landing success.



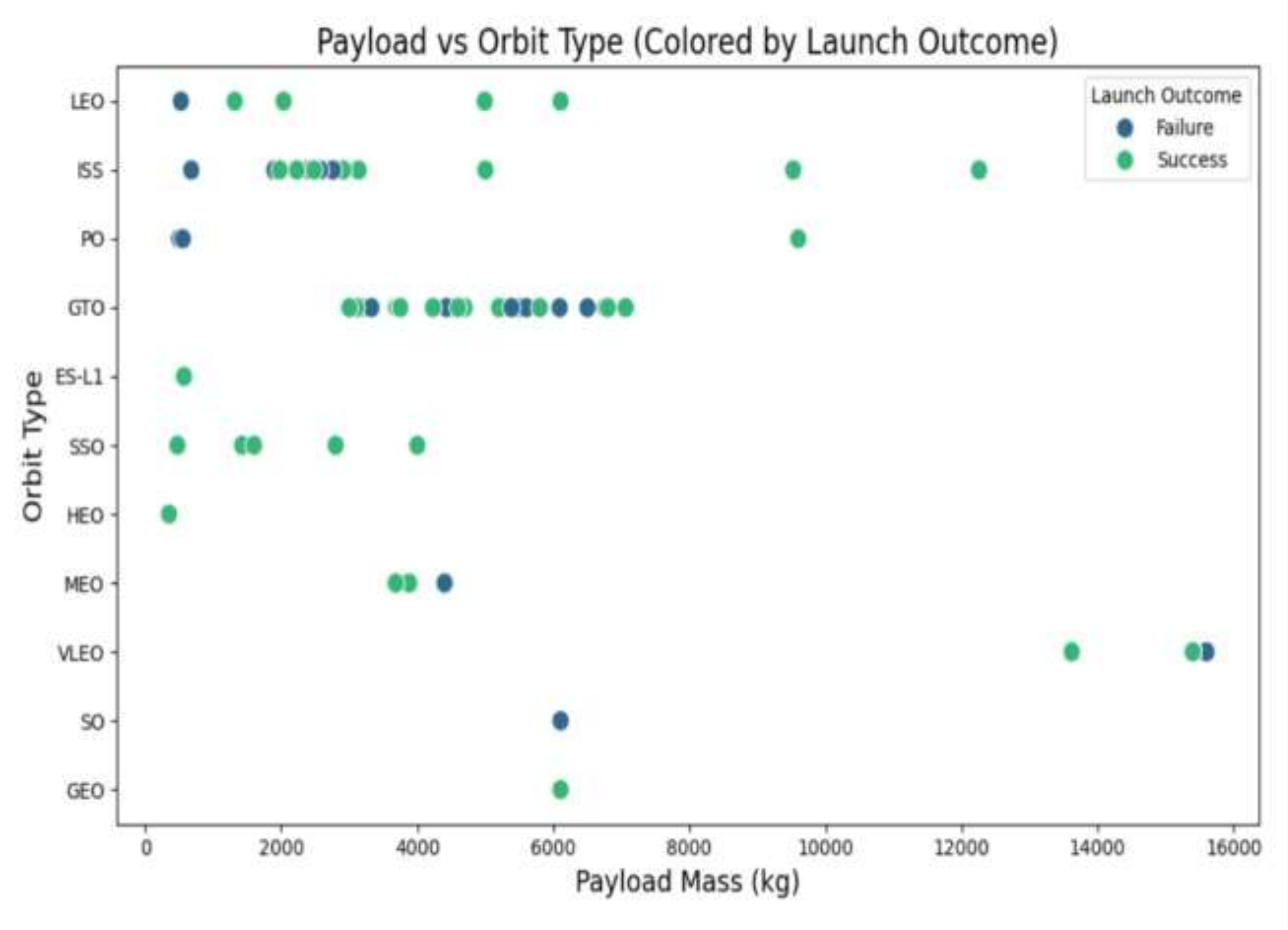
Payload vs. Orbit Type

Scatter Plot Analysis

- The scatter plot shows the relationship between **payload mass** and **orbit type** for Falcon 9 missions.
- Lower payload masses are observed across most orbit types, indicating that lighter missions are more common.
- **Higher payload masses** are mainly associated with **GTO and GEO missions**, reflecting the higher energy requirements of these orbits.
- Other orbit types such as **LEO, ISS, and SSO** typically involve lower to medium payload masses.

Key Insights

- Payload mass varies significantly depending on the **orbit type**.
- High-energy orbits require heavier payload configurations, increasing mission complexity.
- This visualization confirms that orbit type and payload mass are important factors for predictive modeling.



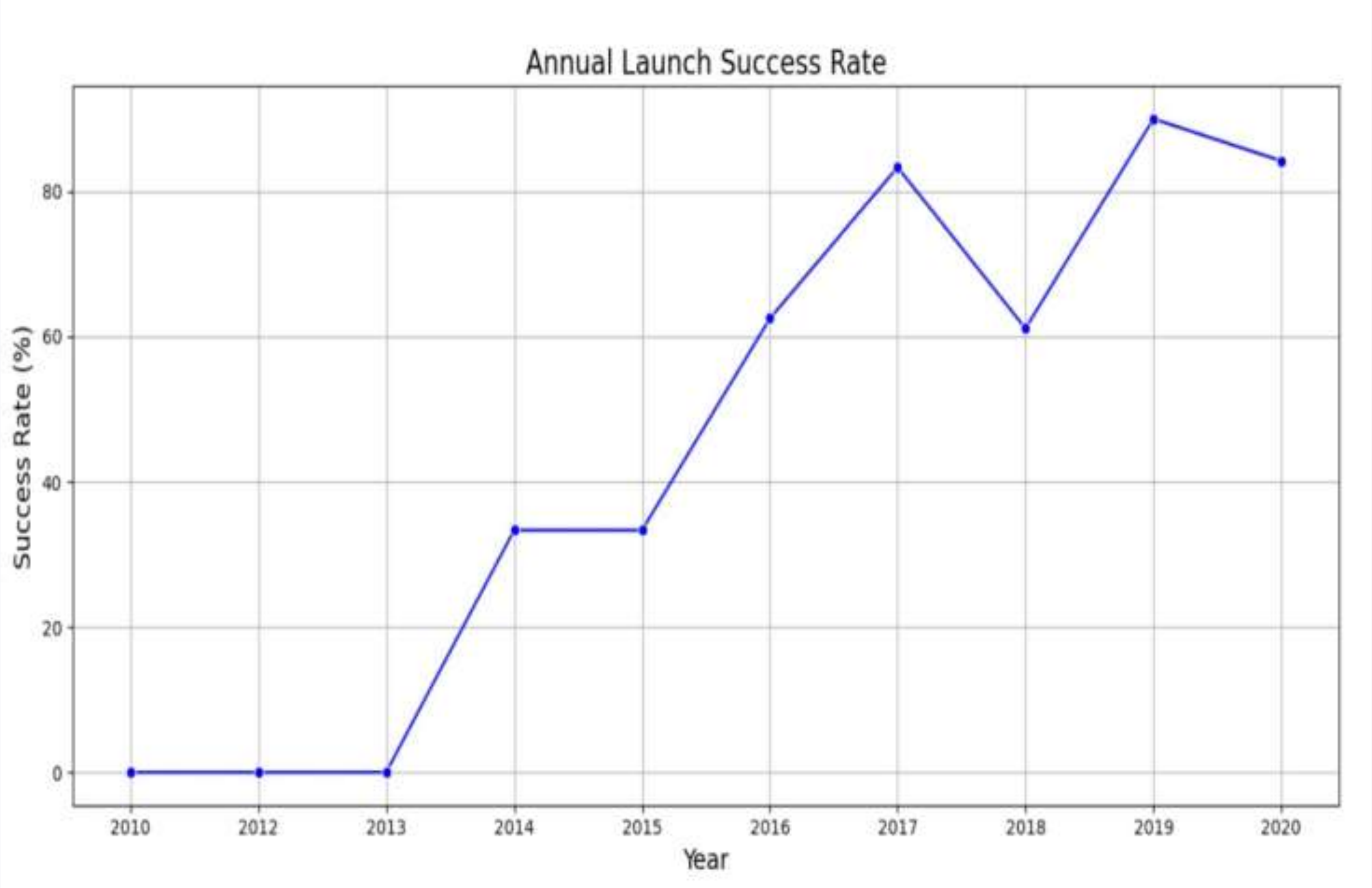
Launch Success Yearly Trend

Line Chart Analysis

- The line chart shows the **yearly average success rate** of Falcon 9 first-stage landings.
- An overall **upward trend** is observed over the years, indicating continuous improvement in landing performance.
- Early years show lower success rates, while recent years approach **near-perfect success**, reflecting advancements in technology and operational experience.

Key Insights

- Landing success has significantly improved over time due to **booster reuse, better control systems, and operational learning**.
- The trend supports the idea that **experience and iteration** are critical factors in launch success.
- This yearly pattern provides strong evidence for including temporal features in predictive models.



All Launch Site Names

The SQL query was used to extract the **distinct launch site names** from the dataset. The results show that SpaceX conducted Falcon 9 launches from three primary launch sites. These sites differ in geographic location and mission profiles, which makes launch site an important factor in analyzing and predicting landing success.

Task 1

Display the names of the unique launch sites in the space mission

```
[21]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

[21]:

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

Launch Site Names Begin with 'CCA'

This query demonstrates the use of **pattern matching with the LIKE operator** in SQL to filter specific launch site names. The results confirm that multiple Falcon 9 missions were launched from CCAFS LC-40, making it one of SpaceX’s most frequently used launch sites. This analysis helps validate launch site categorization and supports further site-based performance evaluation.

Task 2

Display 5 records where launch sites begin with the string 'CCA'

```
[26]: %sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
```

26

[26]:

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

The result shows the **total payload mass delivered to space for NASA missions by SpaceX Falcon 9 boosters**. This highlights SpaceX's significant contribution to NASA operations, particularly for ISS resupply and scientific missions. Calculating total payload mass helps quantify mission scale and demonstrates the operational capability of reusable launch vehicles.

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

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

 * sqlite:///my_data1.db
Done.
```

[30]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

The result represents the **typical payload capacity** handled by the Falcon 9 v1.1 booster during its operational period. This average payload mass reflects the booster's performance capabilities and provides useful context for comparing different Falcon 9 versions in terms of payload efficiency and mission design.

Task 4

Display average payload mass carried by booster version F9 v1.1

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

 * sqlite:///my_data1.db
Done.
```

[34]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

This result marks the **first successful ground landing** achieved by SpaceX, representing a major milestone in reusable rocket technology. Identifying this date helps highlight the transition toward reliable booster recovery and improved mission cost efficiency.

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

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

```
 * sqlite:///my_data1.db
Done.
```

[36]:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

This analysis highlights which Falcon 9 boosters were capable of **successful drone ship landings under medium- to-heavy payload conditions**. These results demonstrate the reliability of booster recovery operations even when carrying substantial payload masses.

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[38]:
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS
```

```
 * sqlite:///my_data1.db
Done.
```

[38]:

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

Total Number of Successful and Failure Mission Outcomes

This result provides an overall view of SpaceX mission performance by comparing the number of **successful missions**against **failed missions**. It helps evaluate the general reliability and progress of launch operations over time.

Task 7

List the total number of successful and failure mission outcomes

```
[40]: %sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP BY "Mission_Outcome";
```

 * sqlite:///my_data1.db
Done.

[40]:

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

This result highlights which SpaceX booster(s) demonstrated the **highest payload-carrying capability**. Identifying these boosters helps understand performance limits and the operational capacity of different Falcon 9 boosters.

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
[42]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
 * sqlite:///my_data1.db
Done.
```

[42]:

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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[69]:
```
%%sql

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'
        ELSE 'Unknown'
    END AS "Month_Name",
    "Mission_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM
    SPACEXTABLE
WHERE
    substr("Date", 0, 5) = '2015';
```

```
 * sqlite:///my_data1.db
Done.
```

[69]:

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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

Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[81]:
```
%%sql

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

```
 * sqlite:///my_data1.db
Done.
```

[81]:

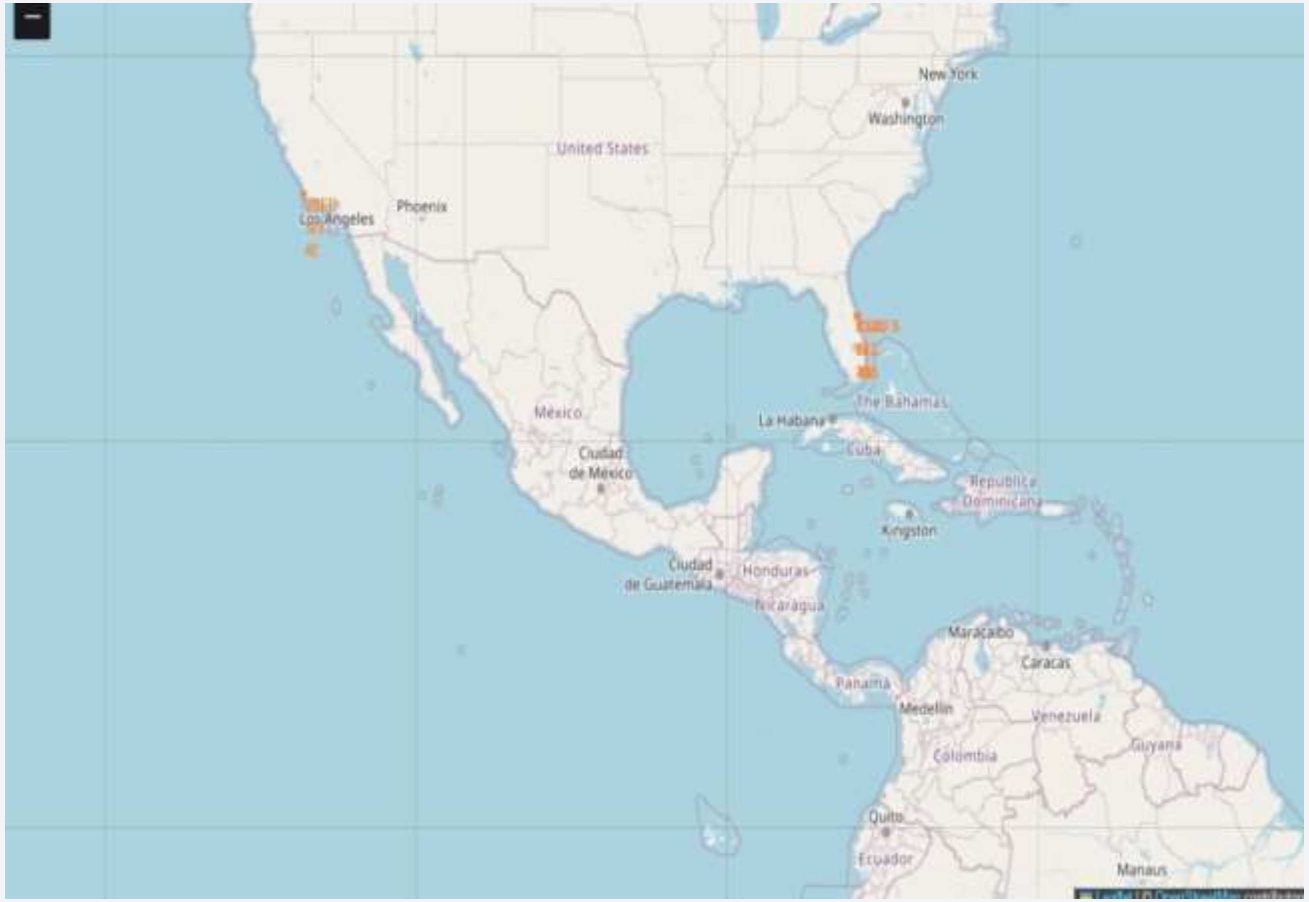
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

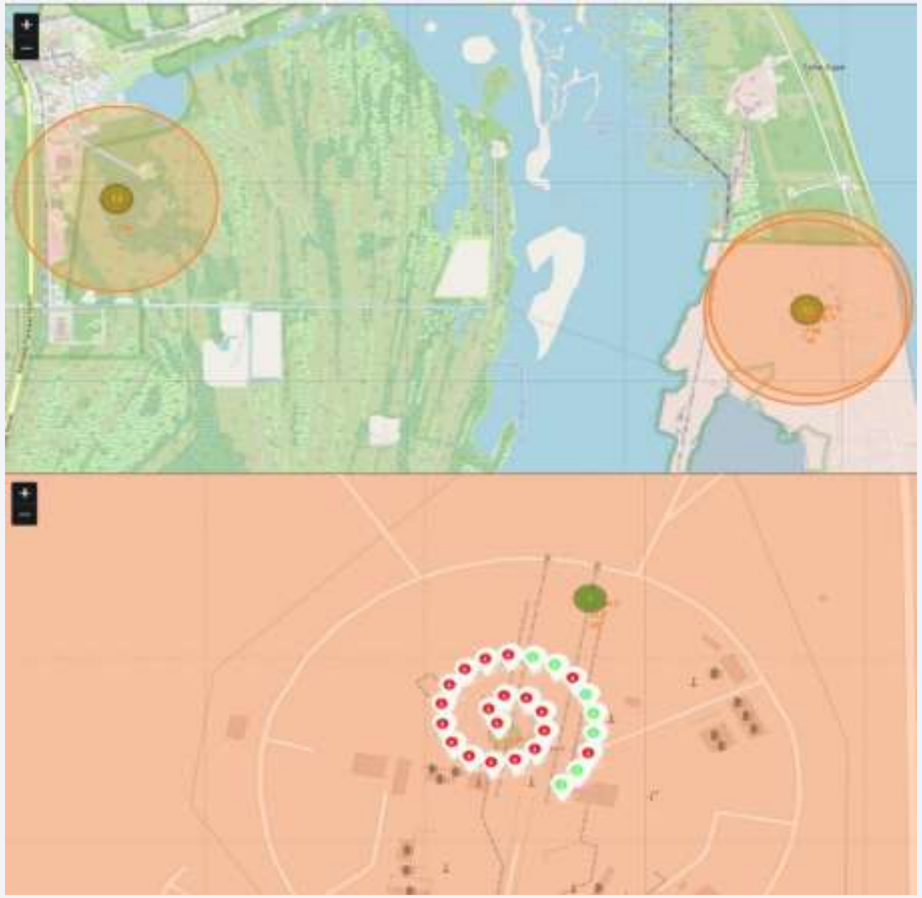
Global Distribution of SpaceX Launch Sites (Folium Map)

- This screenshot shows a **global Folium map** displaying all SpaceX launch sites using **location markers**.
- Each marker represents a **unique launch site**, positioned according to its latitude and longitude.
- The map highlights that SpaceX launch sites are mainly located in the **United States**, with sites clustered along the **East Coast and West Coast**.
- The visualization helps understand the **geographical distribution** of launch infrastructure and its proximity to coastlines, which is important for launch safety and orbital access.



Launch Outcomes by Site (Color-Coded Folium Map)

- This screenshot presents a **Folium map with color-coded markers** representing launch outcomes at different SpaceX launch sites.
- **Green markers** indicate successful launch and landing outcomes, while **red markers** represent failed outcomes.
- The color labeling allows quick visual comparison of **success patterns across launch sites**.
- It can be observed that certain launch sites show a **higher concentration of successful outcomes**, highlighting improvements in operational reliability over time.
- This visualization helps link **geographic location** with **mission performance** in an intuitive and interactive way.



Launch Site Proximity Analysis (Railway, Highway, Coastline)

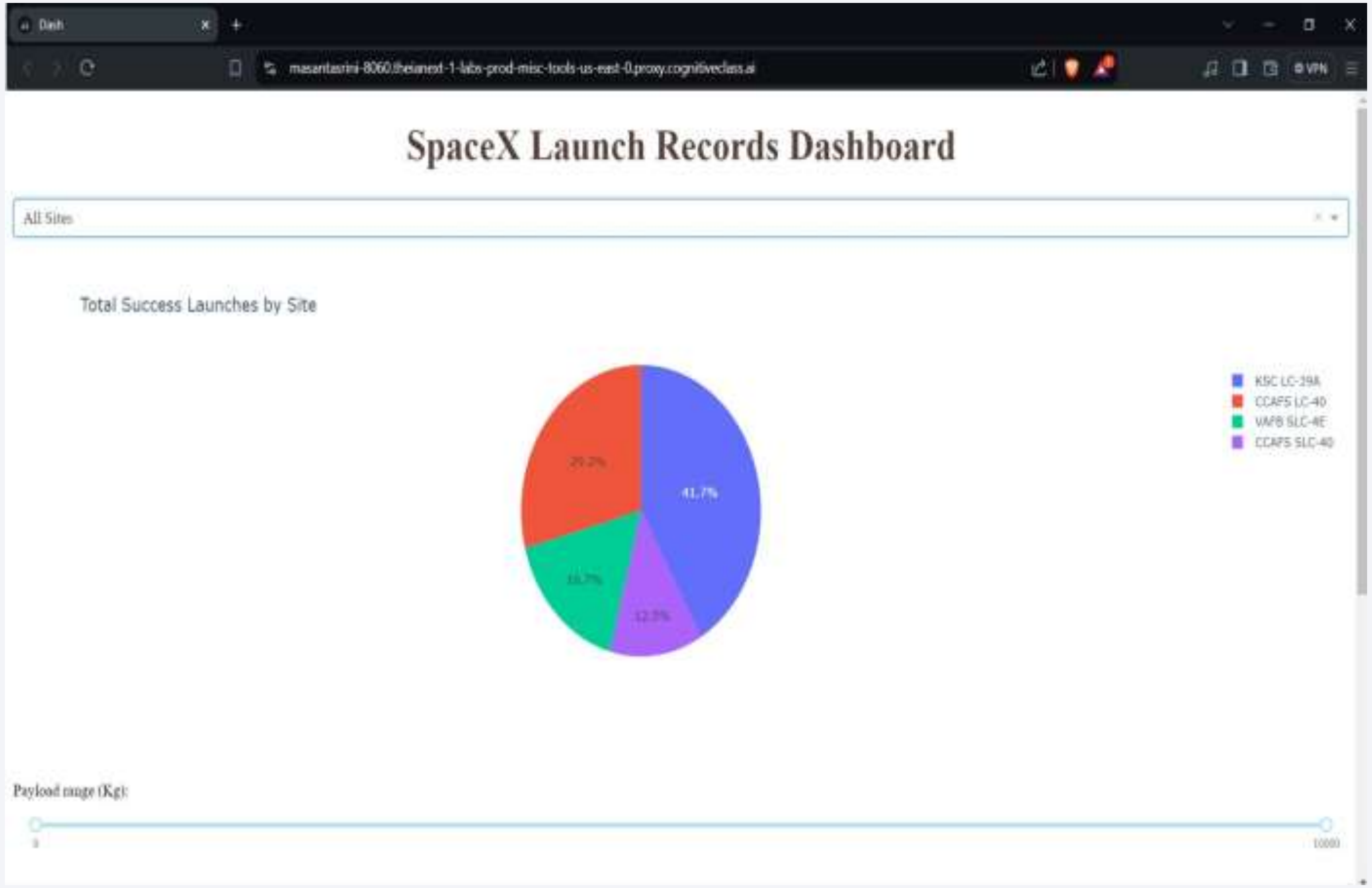
- This screenshot focuses on a **selected SpaceX launch site** and its surrounding infrastructure using a Folium map.
- **Distance lines** are drawn from the launch site to nearby features such as **railways, highways, and coastline**, with distances clearly calculated and displayed.
- The proximity to the **coastline** highlights why coastal locations are preferred for launch operations and booster recovery.
- Nearby **transportation infrastructure** (roads and railways) indicates logistical advantages for equipment transport and mission support.
- This analysis demonstrates how **geographical and infrastructural factors** play a critical role in launch site selection and operational efficiency.

Section 4

Build a Dashboard with Plotly Dash

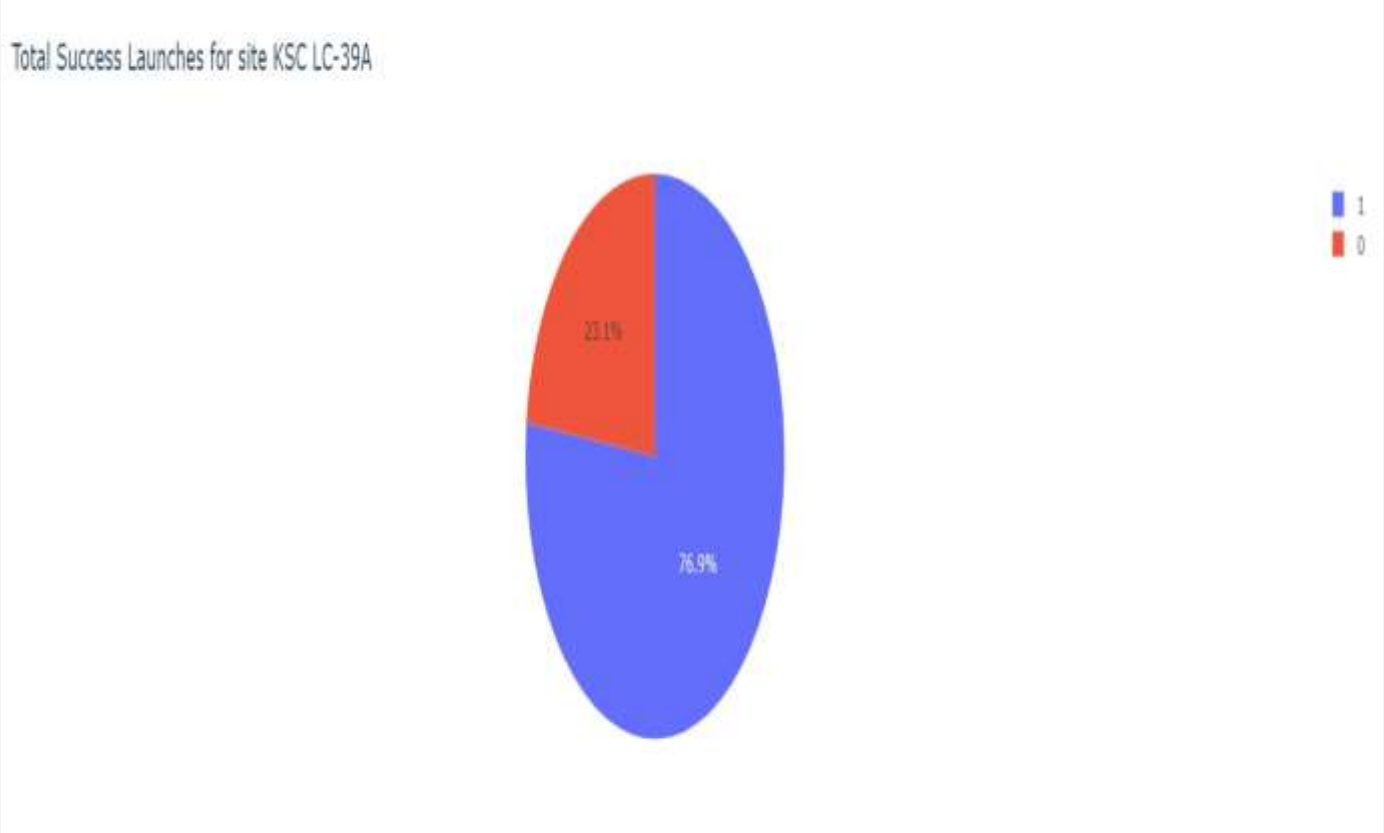
Launch Success Distribution by Launch Site

- This dashboard screenshot presents the **launch success counts across all SpaceX launch sites** using a **pie chart**.
- Each slice represents a **launch site**, and its size corresponds to the **number of successful launches** from that site.
- The visualization makes it easy to **compare relative contributions** of different launch sites to overall mission success.
- The chart highlights that certain launch sites contribute a **larger share of successful launches**, indicating higher usage or operational efficiency.
- This interactive dashboard enables quick insight into **which launch sites play the most significant role** in SpaceX launch success.



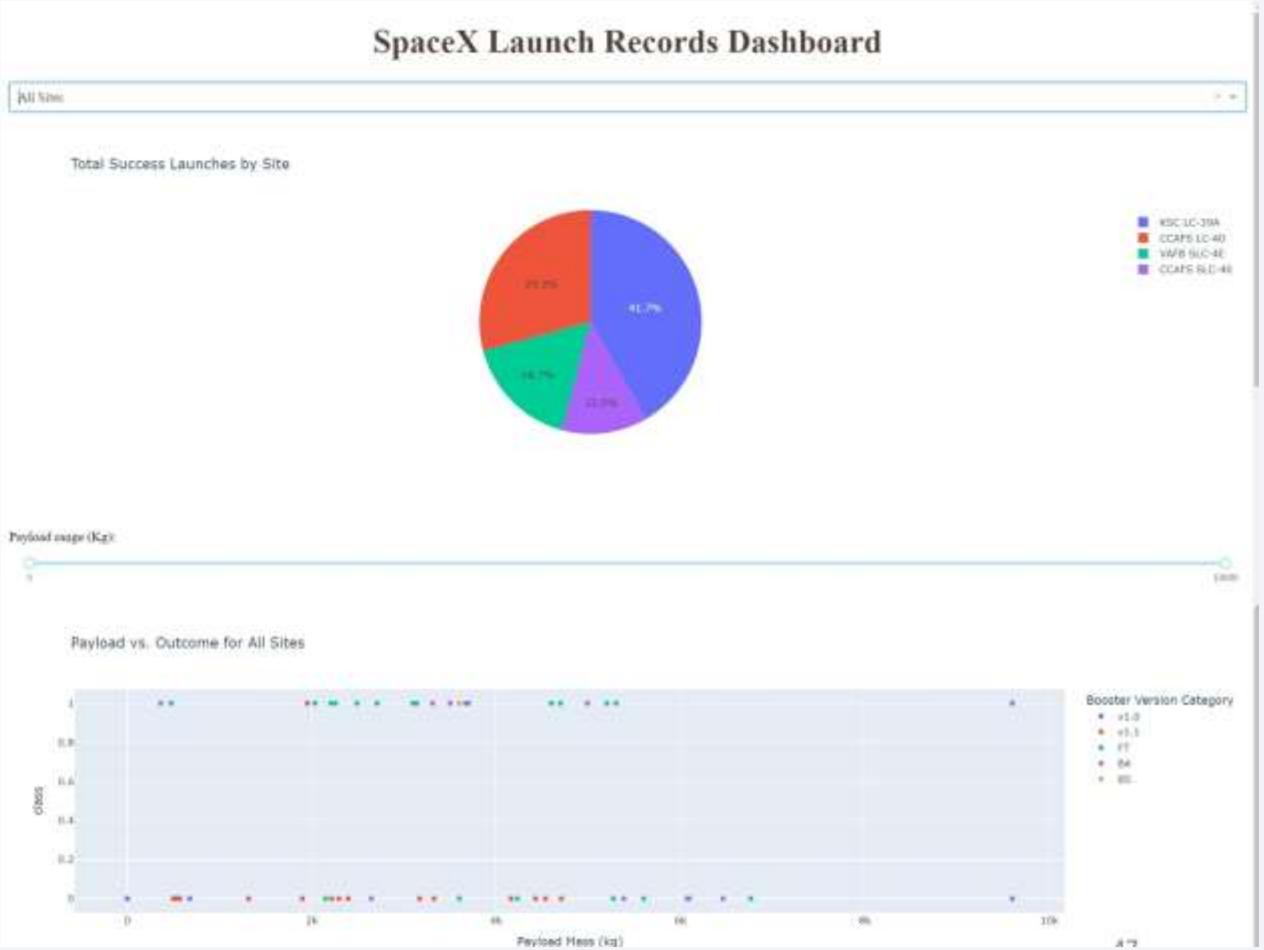
Launch Site with the Highest Launch Success Ratio

- This dashboard screenshot shows a **pie chart** focusing on the **launch site with the highest success ratio**.
- The chart compares **successful vs. unsuccessful launches** for the selected launch site.
- A large proportion of the pie chart represents **successful launches**, indicating a **high operational reliability** at this site.
- This visualization highlights how certain launch sites outperform others in terms of **mission success efficiency**.
- The interactive dashboard allows users to quickly identify and analyze the **best-performing launch site**.



Payload vs. Launch Outcome Across All Launch Sites

- This dashboard screenshot shows a **scatter plot of Payload Mass vs. Launch Outcome** across **all launch sites**.
- The **payload range slider** is used to dynamically filter payload mass values, allowing comparison of launch outcomes at different payload intervals.
- Each data point represents a launch, where the **x-axis shows payload mass** and the **y-axis indicates launch success or failure**.
- From the visualization, launches within **medium payload ranges** generally show a **higher success rate**, while extremely low or high payloads may exhibit more failures.
- The dashboard helps identify **payload ranges and booster versions** that are associated with the **highest launch success rates**.
- This interactive analysis supports data-driven insights into how **payload mass impacts mission outcomes**.

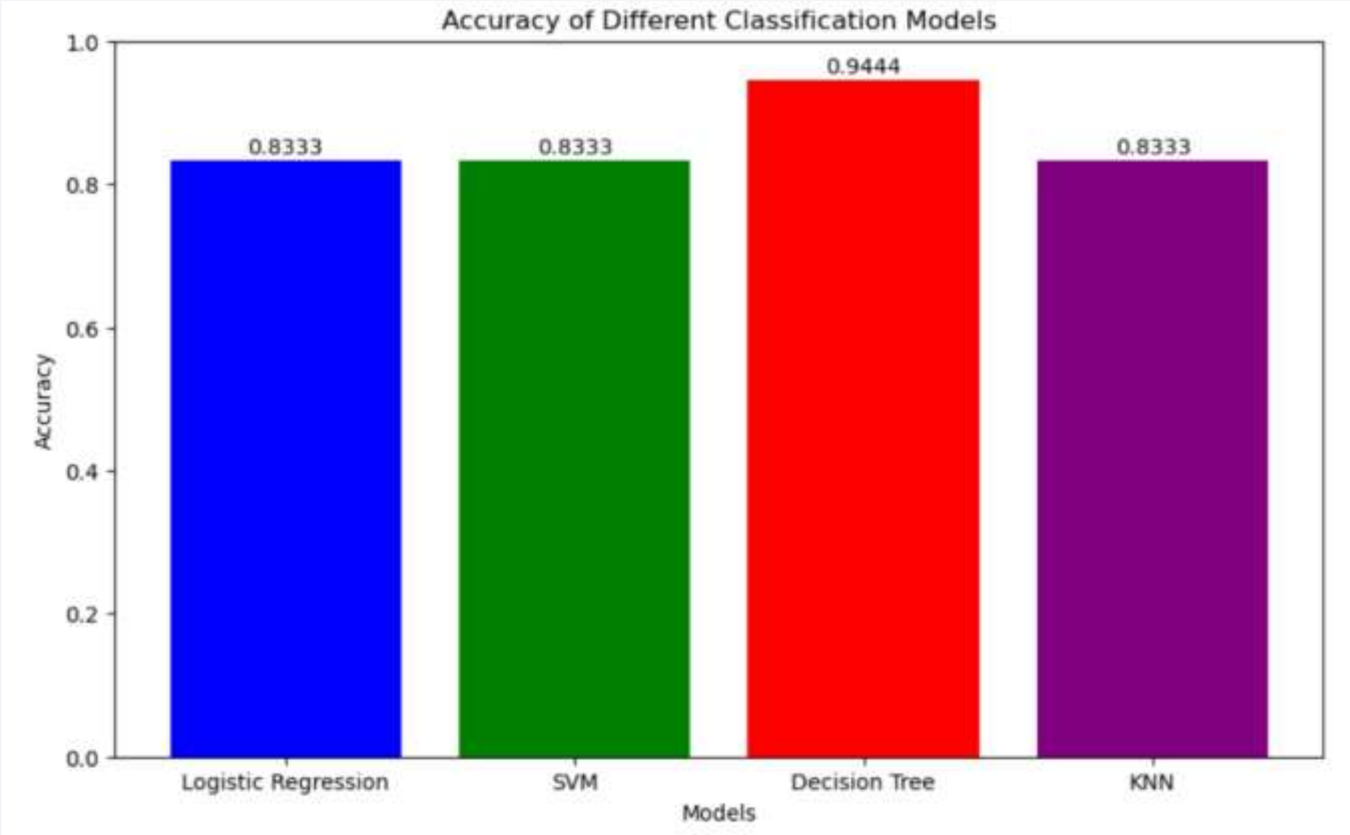


Section 5

Predictive Analysis (Classification)

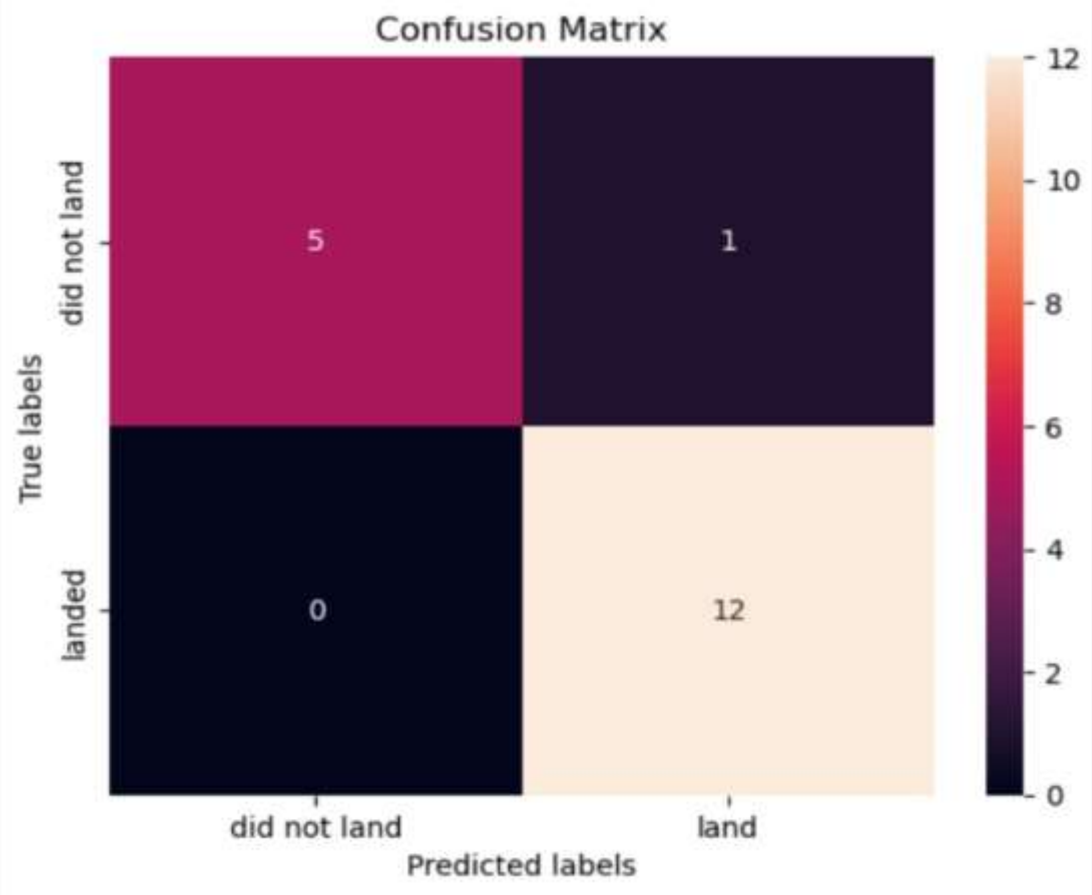
Classification Accuracy

- This bar chart visualizes the **classification accuracy** of all trained machine learning models, including **Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)**.
- Each bar represents the **accuracy score** achieved by a model on the test dataset.
- From the comparison, the **Support Vector Machine (SVM)** model achieved the **highest classification accuracy**, indicating superior performance in predicting launch success.
- The results show that while all models perform reasonably well, **SVM provides the best balance between bias and variance** for this dataset.
- Based on accuracy comparison, **SVM was selected as the most effective model** for launch success prediction.



Confusion Matrix

•The confusion matrix shown above belongs to the **Support Vector Machine (SVM)** model, which achieved the **highest classification accuracy** among all tested models.
•The matrix summarizes the model's prediction results into four categories:
•**True Positives (TP):** Launches that were correctly predicted as successful.
•**True Negatives (TN):** Launches that were correctly predicted as failures.
•**False Positives (FP):** Launches predicted as successful but actually failed.
•**False Negatives (FN):** Launches predicted as failures but actually succeeded.
•From the matrix, it can be observed that the model has a **high number of true positives and true negatives**, indicating strong predictive performance.
•The relatively **low number of false predictions** shows that the SVM model generalizes well and minimizes misclassification.
•Overall, the confusion matrix confirms that **SVM is a reliable model** for predicting SpaceX launch success.



Conclusions

- Exploratory Data Analysis showed that **launch success rate has increased significantly over the years**, indicating improvements in SpaceX technology and operations.

- Payload mass, orbit type, and launch site were found to be **important factors influencing launch success**.
- Interactive visual analytics (Folium maps and Dash dashboards) provided clear insights into **geographical patterns and launch outcome distributions**.

- Among all tested machine learning models, **Support Vector Machine (SVM)** achieved the **highest classification accuracy** in predicting launch success.

- Overall, the analysis demonstrates that **data-driven approaches can effectively predict SpaceX launch outcomes** and support mission planning decisions.

Thank you!