



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Noah Bonaguidi
10.30.2024



Table of Contents

• Executive Summary	3
• Introduction	4
• Methodology	5
• Results	17
• Insights from EDA	19
• Launch Site Proximity Analysis	36
• Plotly Dashboard	40
• Predictive Analysis	44
• Conclusion	47
• Appendix	48

Executive Summary

- We analyzed SpaceX launch data from the API and web scraping to identify factors influencing landing success. After data cleaning and feature engineering, we conducted exploratory data analysis and built machine learning models (Logistic Regression, Decision Tree, KNN, SVM) to predict landing success. We employed techniques like GridSearchCV and confusion matrices for model optimization and evaluation.
- Our findings reveal that **Number of Flights plays a significant role in determining landing success**, with more launches being associated with more successful re-entries. Over the past decade, the overall success rate of SpaceX flights has steadily improved. Specifically, launches targeting orbits such as ES-L1, GEO, HEO, and SSO **demonstrate the highest probability of successful landings**. Additionally, launches from CCAFS SLC-40 emerged as the most reliable in terms of landing success.

Introduction

- SpaceX has redefined the possibilities of space exploration, becoming the first private company to successfully land a rocket returning from low-Earth orbit—a feat that once seemed like science fiction. With each successful launch and landing, SpaceX not only brings us closer to the stars but also reshapes the future of space travel. In this project, we set out to uncover the secrets behind these remarkable achievements. What factors work in perfect harmony to ensure a rocket's safe return, and can this magic be reliably recreated? Through the lens of data, we embark on a journey to decode the mechanics of these celestial touchdowns.
- To guide our exploration, **we focused on answering three key questions:** *What are the critical factors that drive a successful launch and landing? How does the changing nature of payloads, orbit types, and flight conditions influence the outcome? And finally, can we leverage the power of machine learning to predict the success of future launches based on the lessons hidden within past missions?*

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
 - We sourced our data through two main channels: the SpaceX API, which provided real-time information on launches, and web scraping from Wikipedia to gather historical data for older missions. This allowed us to build a comprehensive dataset covering a wide range of launches.
- Perform data wrangling
 - We began by addressing missing or null values, either converting them where appropriate or removing them to maintain data integrity. Next, we engineered several new features, such as the number of launches per site, the distribution of orbit types, and categorized mission outcomes. To streamline our analysis, we added a binary column to distinguish between 'Good' and 'Bad' outcomes, enabling more straightforward evaluation of success factors.
- 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
 - Our predictive analysis started by organizing the cleaned data into a dataframe. We selected key features as our independent variable (X) and the mission outcome class as the dependent variable (Y). After splitting the data into training and testing sets, we built four classification models: Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Using GridSearchCV, we fine-tuned each model's parameters for optimal performance. To evaluate accuracy, we employed confusion matrices and scoring functions, with KNN emerging as the most accurate model.

Data Collection

To build a comprehensive dataset for our analysis, we employed two primary methods:

1. **SpaceX API:** The SpaceX API served as our primary source of data, providing detailed information on each launch. Key columns collected include:
 - **Booster Version, Payload Mass, Orbit, Launch Site, Outcome, Flights, Grid Fins, Reused, Legs, Landing Pad, Block, Reused Count, Serial, Longitude and Latitude**
2. **Web Scraping Wikipedia:** For historical missions, particularly those not covered in full by the API, we supplemented the dataset through web scraping. This process allowed us to retrieve and align missing data across all the same columns listed above, such as **Payload Mass, Orbit, and Launch Site**, ensuring completeness in our dataset for both current and historical launches.

After collecting the data, we combined it into a single dataset, ensuring consistency across both sources. This unified dataset was then cleaned, with any null values handled through imputation or removal, and prepared for analysis.

Data Collection – SpaceX API

To collect data from the SpaceX API, we made several RESTful API calls to retrieve detailed information on launches such as:

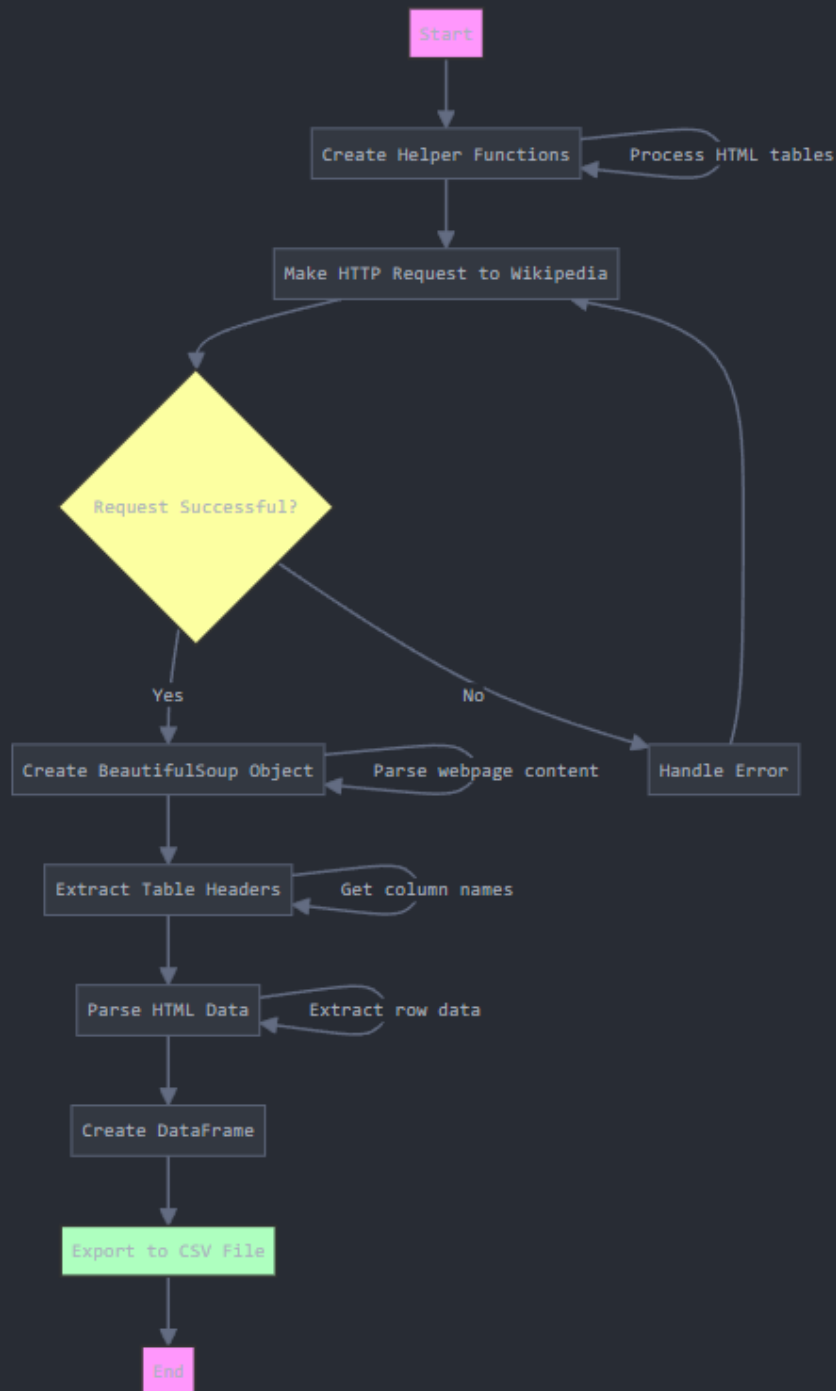
- Booster Version
- Payload Mass
- Orbit
- Launch Site
- Outcome
- And other key fields

See the flow chart to the right and [view our Github repository](#) for more information.

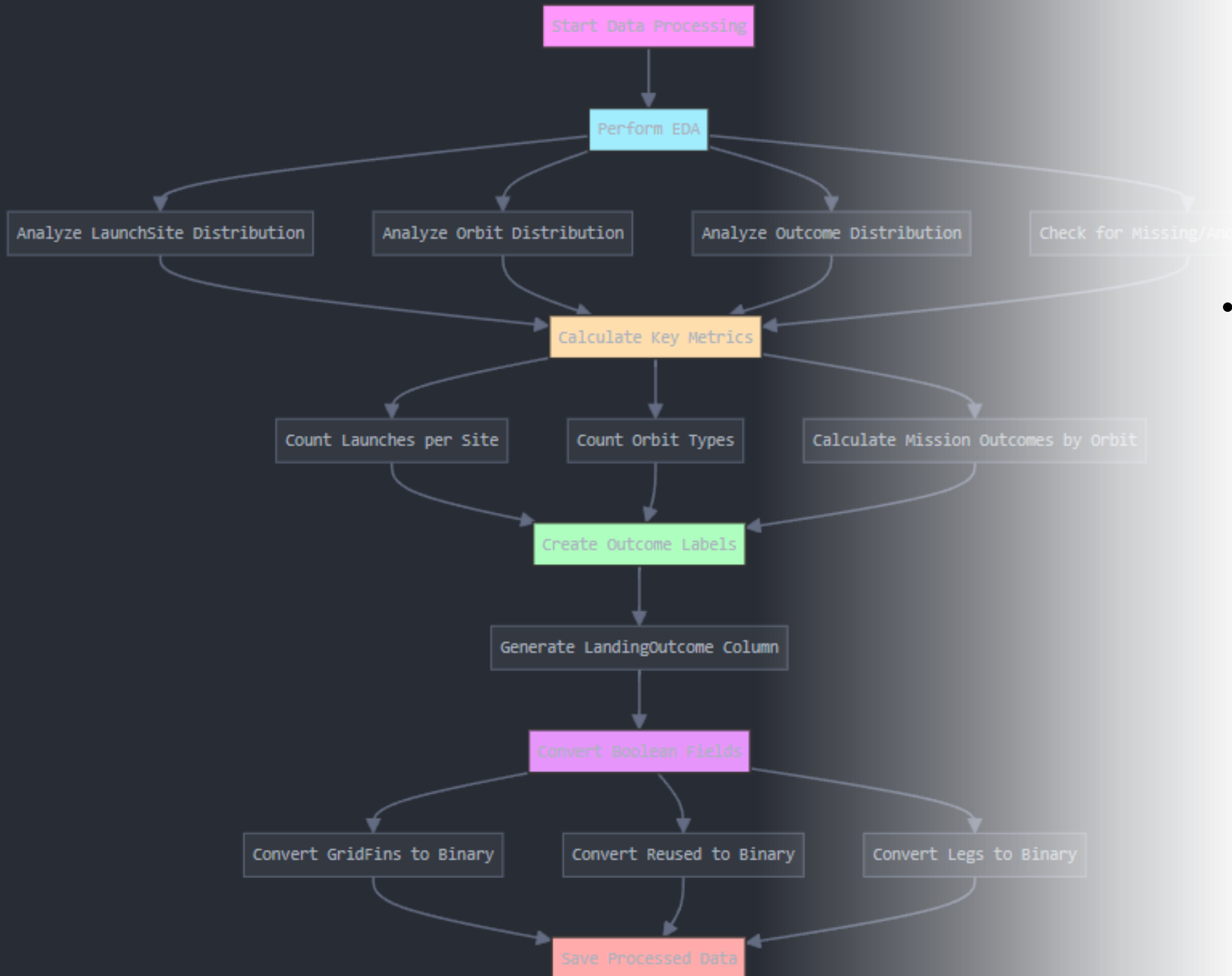


Data Collection – Web Scraping

- For the historical data that was not available via the SpaceX API, we implemented a web scraping pipeline to extract launch information from Wikipedia. We then cleaned parsed, cleaned, and formatted the data. The steps of our web scraping process can be seen in our flow chart to the left or [visit the Github link](#) for more detailed information about our web scraping methods.



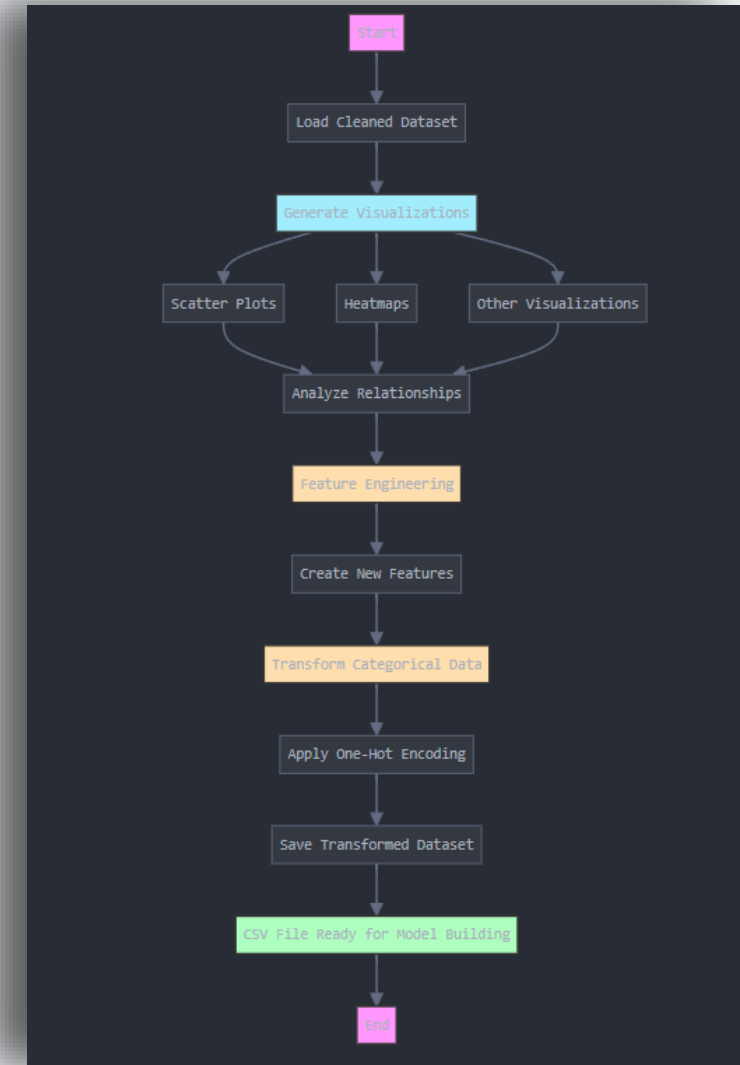
Data Wrangling



- For the data wrangling phase, we cleaned and analyzed necessary data to better understand the structure and distribution of the dataset and made necessary transformations while also creating new key features for model training in the next steps. The key steps in this process can be seen in the flow chart and [Github link](#) for more hands on information.

EDA with Data Visualization

In our exploratory data analysis phase, we compared various features to identify which had the strongest relationships with launch success. Using visualization tools such as **matplotlib** and **seaborn**, we gained valuable insights into the relationships between **Flight Number**, **Payload Mass**, **Orbit Type**, and **Launch Site**, which showed the most significant correlations with landings. Our visualizations helped uncover patterns and trends that informed the subsequent steps of feature engineering and model building. For more detailed analysis and visual outputs, please refer to our flowchart and [GitHub repository](#).



EDA with SQL

In addition to visualization tools, we utilized **SQL** to further manipulate and query our dataset, gaining deeper insights into key relationships. Through SQL queries, we explored various aspects such as the:

- **average payload mass carried by different rockets**
- **booster success rates on drone ships based on specific payload weights**
- **landing outcomes on ships versus ground pads over different time ranges**

This approach provided another layer of understanding, enabling us to analyze the data from new perspectives. For a comprehensive breakdown of the SQL queries and results, please [refer to our GitHub repository](#).

Build an Interactive Map with Folium

For this portion of our project, we used **Folium** to create an interactive map of the United States that highlights SpaceX launch sites. We added the following map objects to provide meaningful insights:

- **Markers for Launch Sites:** Each SpaceX launch site was marked on the map to visualize their geographic locations. This helps to contextualize the data and observe the distribution of launch sites across the US.
- **Success/Failure Markers:** We differentiated between successful and failed launches at each site by color-coding the markers. Green markers indicate successful launches, while red markers signify failures. This provides a quick visual reference to evaluate the success rates at various sites.
- **Distance Markers:** To analyze geographic influences, we calculated and marked the distances between launch sites and key geographic features, such as coastlines. This helped assess the potential impact of location and nearby water bodies on launch outcomes.

The interactive map can be explored further, and detailed code can be reviewed through our [GitHub repository link](#) provided for peer-review.

Build a Dashboard with Plotly Dash

Our dashboard features two primary interactive visualizations to provide insights into launch success patterns:

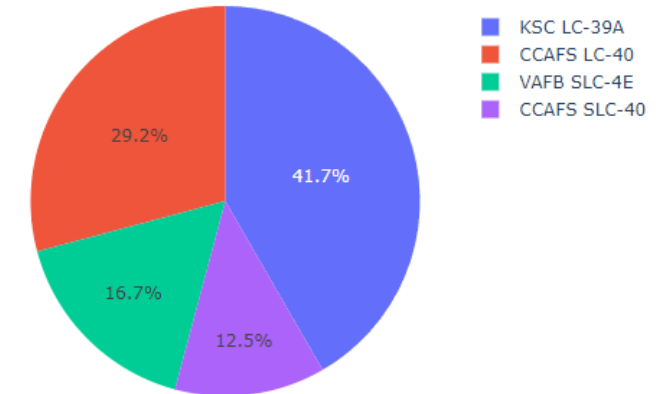
- **Launch Success Pie Chart:** This chart offers an at-a-glance view of the percentage of successful launches, defaulting to **All Sites**. Users can drill down to view **success versus failure rates at individual launch sites**. This visualization highlights site-specific performance, making it easy to compare success rates across locations.
- **Payload Mass Scatter Plot:** Equipped with a **Payload Mass slider**, this scatter plot visualizes launch outcomes by booster type and payload mass. Launch success (marked as 1) and failure (marked as 0) can be examined within any selected payload range, giving insight into how different payload weights and booster types affect launch outcomes.

These visualizations offer valuable, high-level insights in a dynamic and user-friendly format without over-cluttering the interface. For a hands-on experience with our dashboard, [please visit our GitHub repository](#).

All Sites



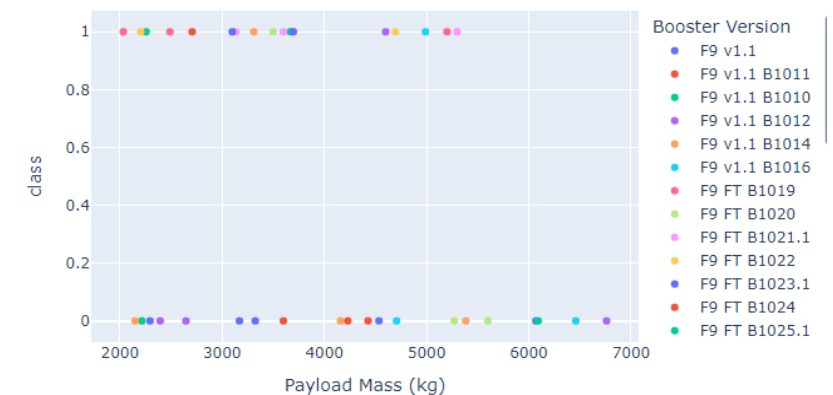
Total Success Launches By Site



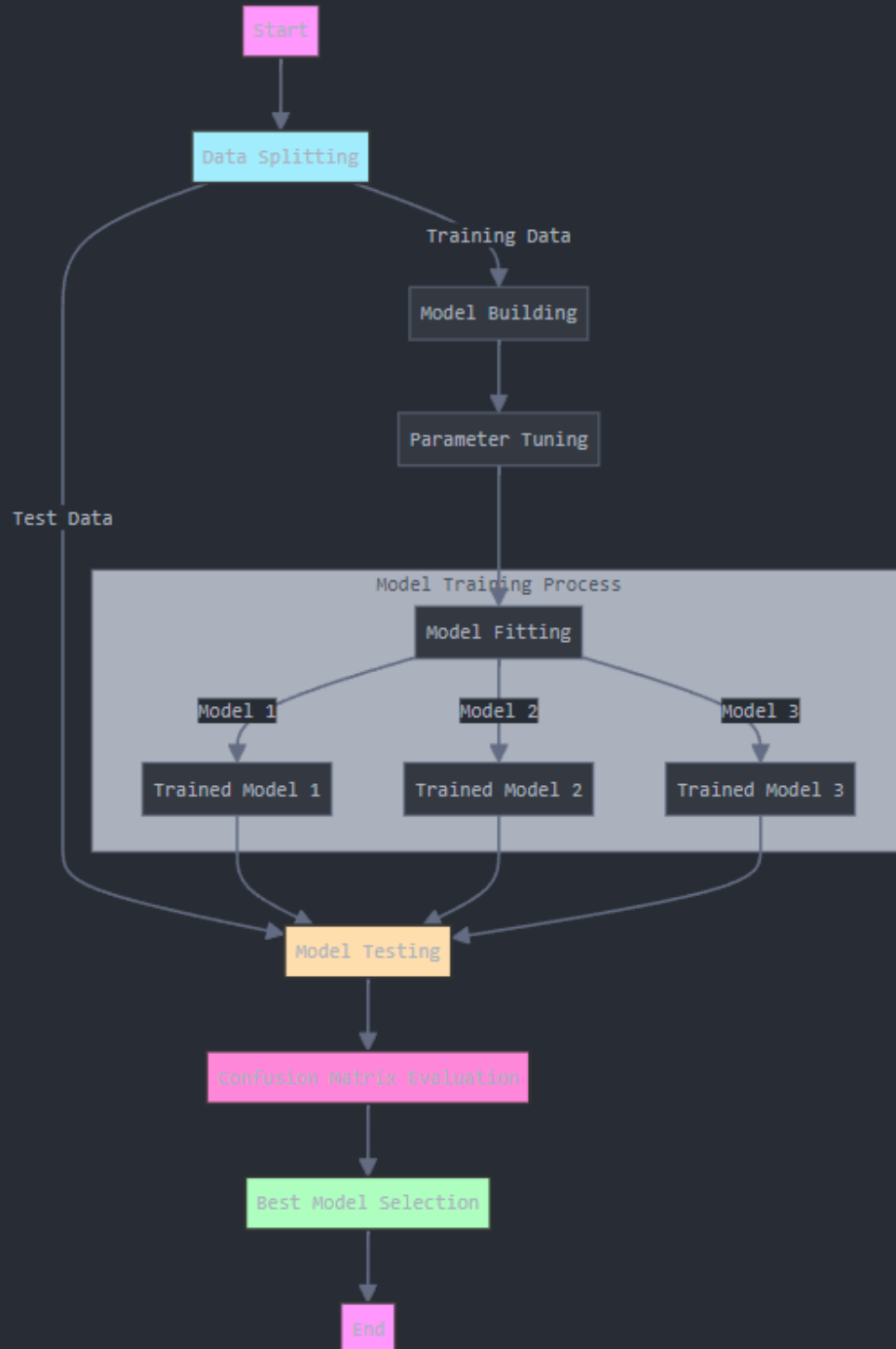
Payload range (Kg):



Correlation between Payload Mass and Success for all sites



Predictive Analysis (Classification)



We developed four classification models—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—to predict launch outcomes. Starting with data splitting into training and testing sets, we optimized each model using **GridSearchCV** to fine-tune parameters and **confusion matrices** to visualize results. Here's a summary of each model's process and performance:

- **Logistic Regression:** After setting ranges for parameters like regularization strength and solver type, GridSearchCV identified the optimal setup, achieving an **82% accuracy** on the training set.
- **Support Vector Machine (SVM):** By adjusting kernel type and regularization strength, we reached an **83% accuracy**, with clear classifications seen in the confusion matrix.
- **Decision Tree:** With depth, criterion, and split type optimized, the Decision Tree achieved a **86% accuracy**, showing promise in select cases but limited generalization.
- **K-Nearest Neighbors (KNN):** Fine-tuning parameters such as neighbor count and distance metric yielded the highest performance with **87% accuracy**, making KNN our best-performing model.

After evaluating all models, KNN showed the highest accuracy, though additional tuning could further improve its performance. For an in-depth analysis and code details, please visit our [GitHub repository](#).

Results

Exploratory Data Analysis (EDA) Key Findings

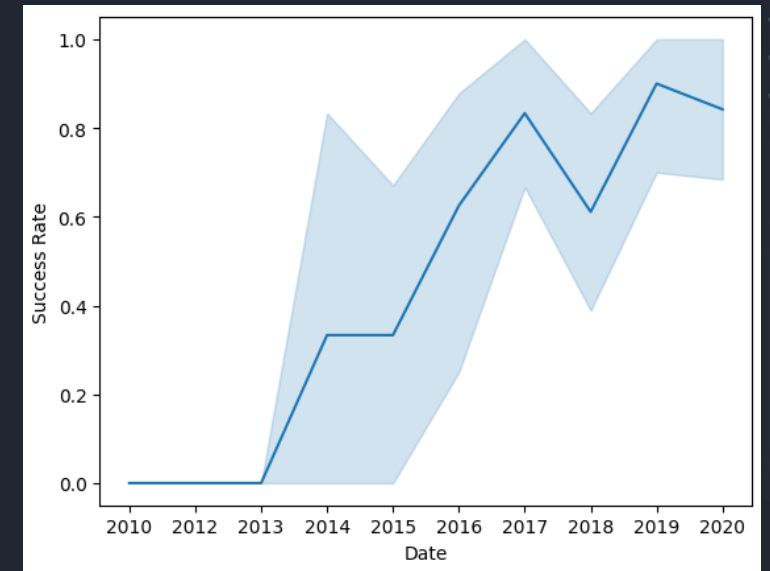
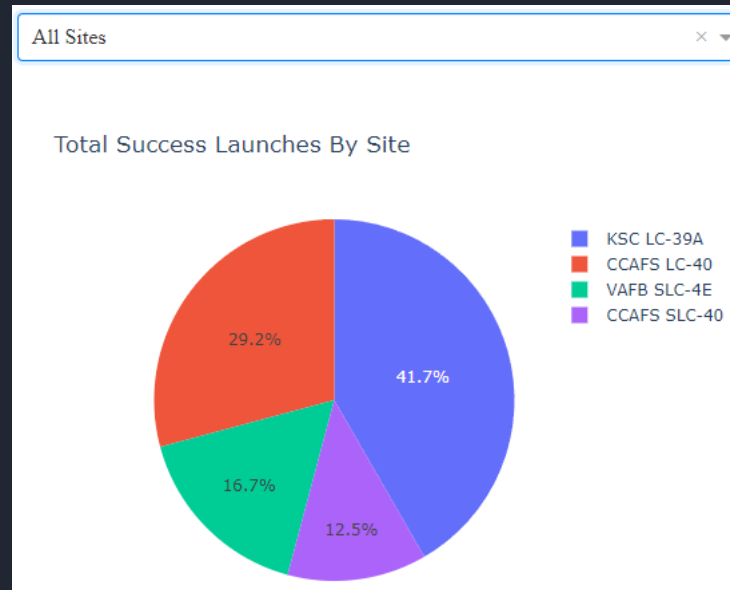
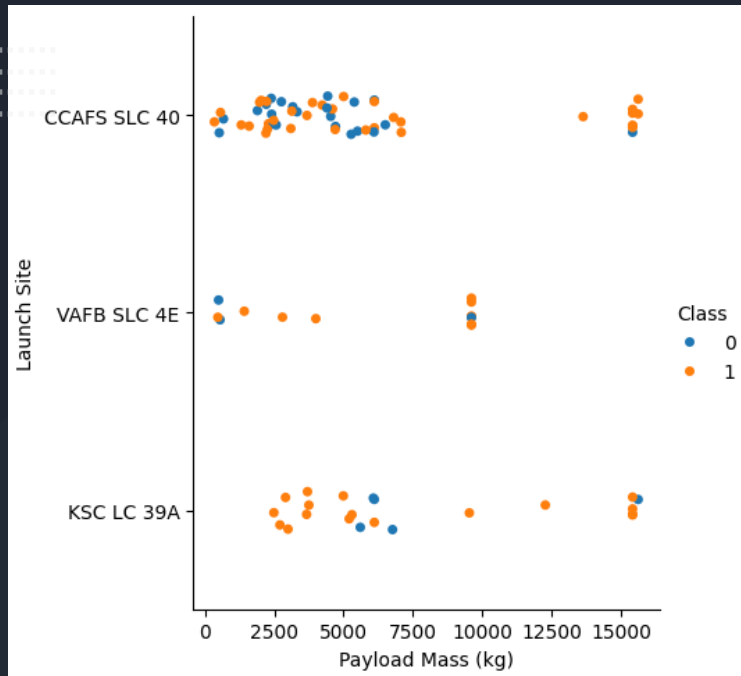
- **Launch Site Significance:** Certain launch sites consistently demonstrated higher success rates, with CCAFS SLC-40 achieving the highest number of successful launches.
- **Orbit Type and Payload Impact:** Heavier payloads in orbits like GEO, HEO, and SSO correlated strongly with successful re-entries and landings, highlighting specific orbit types as influential factors in landing success.
- **Flight Number:** Higher flight numbers corresponded to better success rates, suggesting operational improvements and increasing success rates over time.

Interactive Analytics (*Screenshots on next slide*)

- Dynamic exploration of launch success rates by site and orbit type.
- Payload success analysis by booster type and interactive filters to drill down on key patterns.

Predictive Analysis Highlights

- **K-Nearest Neighbors (KNN) Model:** Outperformed other models with an **87% accuracy** in predicting landing outcomes.
- **Model Insights:** High accuracy in distinguishing between successful and unsuccessful landings, providing reliable predictions across payload and launch site variables.
- For further interactivity, please view our [GitHub repository](#) with full code and demo.



Results (Cont.)

For our results, exploratory analysis revealed that certain factors like **launch site**, **orbit type**, and **payload mass** significantly impact landing success, with heavier payloads in orbits like GEO, HEO, and SSO achieving higher success rates. **CAAFS SLC-40** emerged as the most reliable launch site, while **flight number** positively correlated with mission outcomes, reflecting operational improvements over time. Our **K-Nearest Neighbors (KNN)** model achieved the highest accuracy at **87%** for predicting successful landings, outperforming other models in distinguishing success by payload and site variables. For a full interactive exploration, view our [GitHub repository](#)



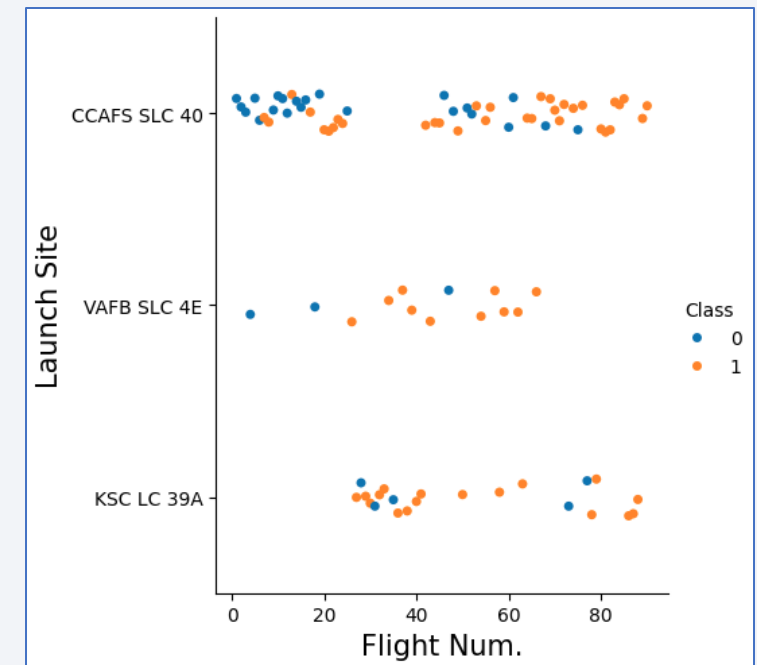
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

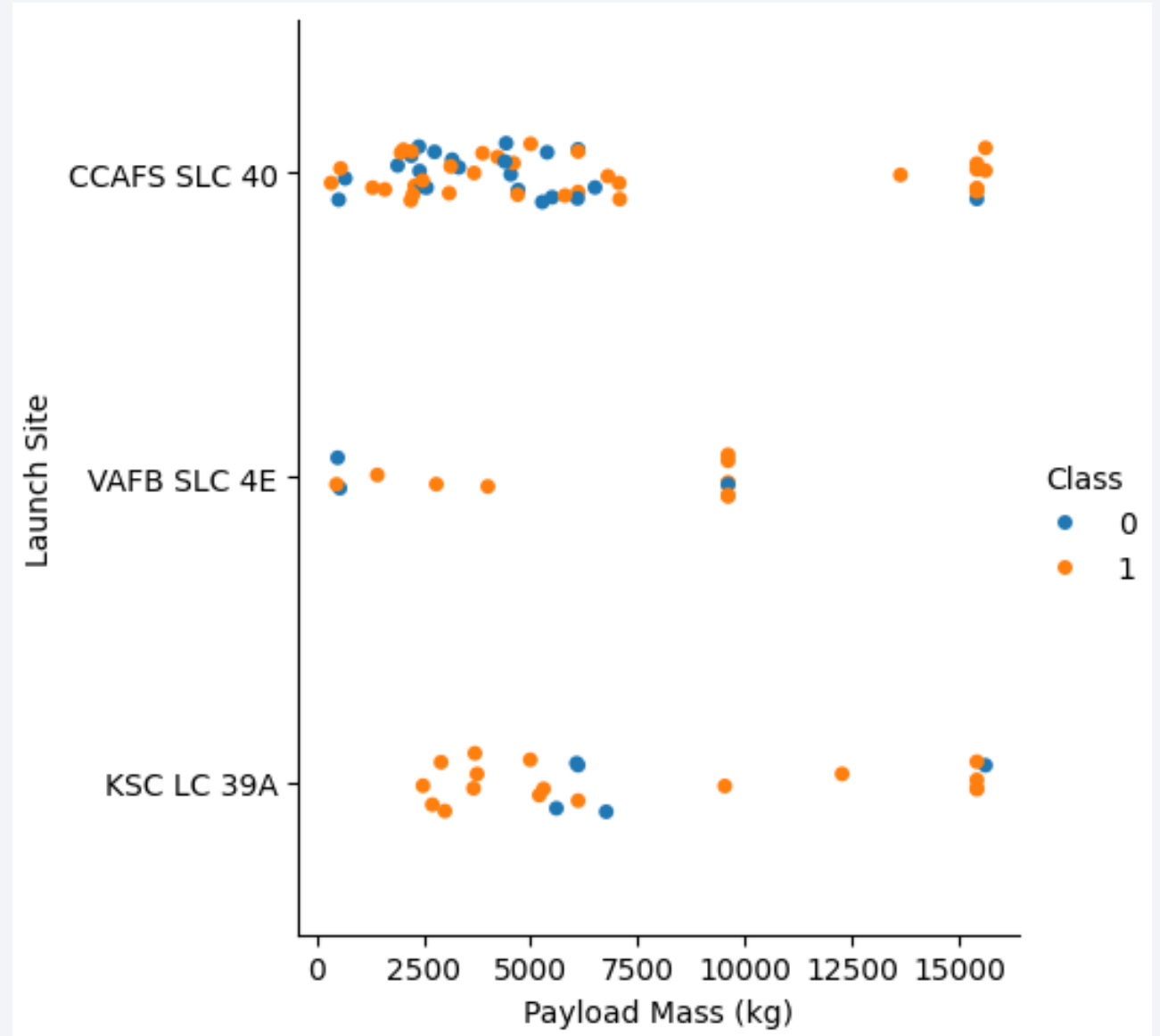
Our analysis shows a clear trend: as the frequency of flights increases, so does the likelihood of success. Initially, in the first 20-25 flights, there's a noticeable cluster of failed launches (blue dots), reflecting early operational challenges. However, beyond this point, launches become increasingly successful, with only occasional failures. This suggests a positive relationship between cumulative flight experience and mission success rates, likely due to iterative improvements in technology and processes with CCAFS SLC-40 having the most flights (and successes) out of our three sites.

FLIGHT	DESTINATION	TIME	GATE	STATUS
9587	NEW YORK	08:30	7	BOARDING
6254	TOKYO	08:55	32	BOARDING
1389	MOSCOW	09:20	15	ON TIME
6137	BERLIN	10:45	25	ON TIME
2648	COPENHAGEN	11:15	10	ON TIME
8582	BUENOS AIRES	12:20	18	ON TIME
3425	SINGAPORE	13:35	22	ON TIME
9836	TORONTO	14:10	14	ON TIME
8231	SYDNEY	15:30	24	ON TIME
6580	PARIS	16:15	19	ON TIME

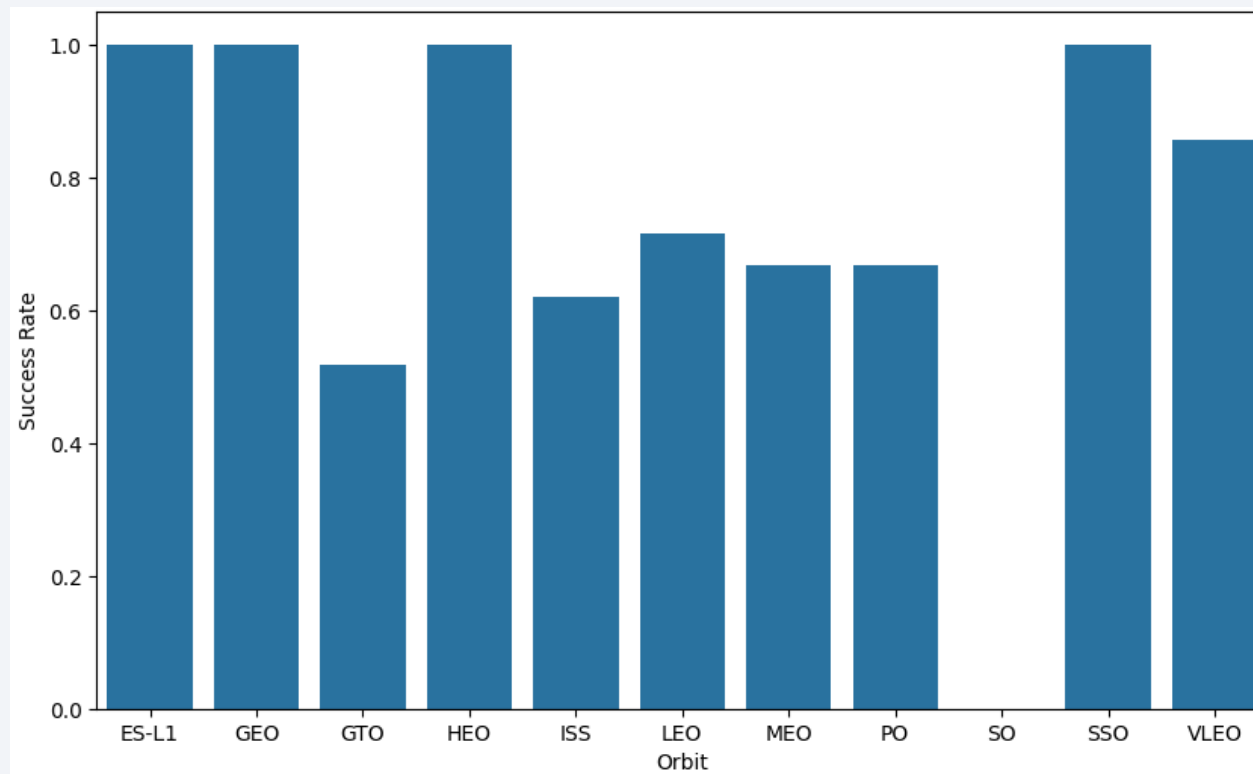


Payload vs. Launch Site

Our analysis indicates **no clear relationship between payload size and launch site success rates**. While there is a cluster of successful launches with payloads above 12,500 kg at one particular site, **the overall data remains too varied** to conclude that heavier payloads consistently contribute to higher success rates. The mixed outcomes across sites suggest that other factors may play a more significant role in launch success.



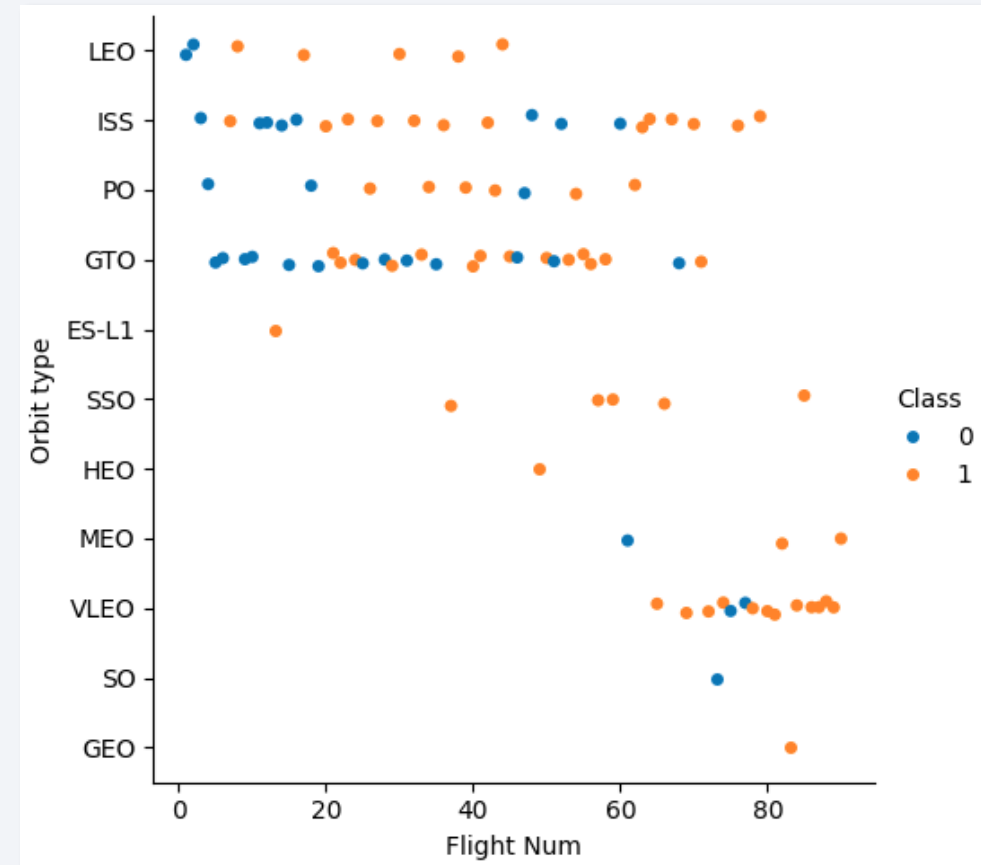
Success Rate vs. Orbit Type



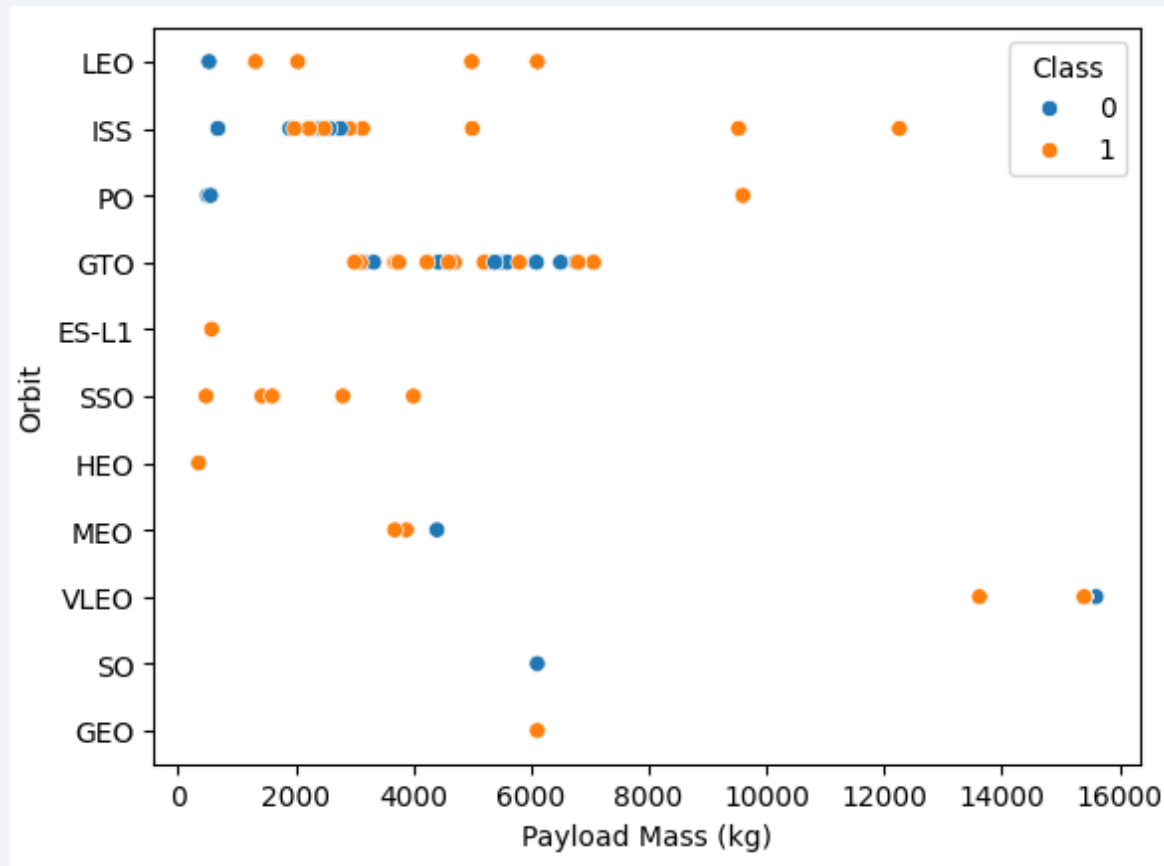
Our findings show that certain orbit types— **ES-L1, GEO, HEO, and SSO** —achieve perfect success rates, standing out among the 11 types analyzed. In contrast, the **SO orbit type** presents an anomaly, having only a single recorded launch, which resulted in failure. When viewed alongside **orbit type versus flight number**, this anomaly confirms that SO's lower success rate is due to limited launch frequency rather than an inherent risk factor in the orbit type itself.

Flight Number vs. Orbit Type

Similar to the trend observed with launch sites, our analysis of **flight number versus orbit type** shows that increased flight frequency generally correlates with higher success rates. While occasional outliers are present, these deviations may be attributed to external factors or adjustments in mission parameters. This pattern underscores the impact of cumulative experience and iterative improvements on the success rates across various orbit types.



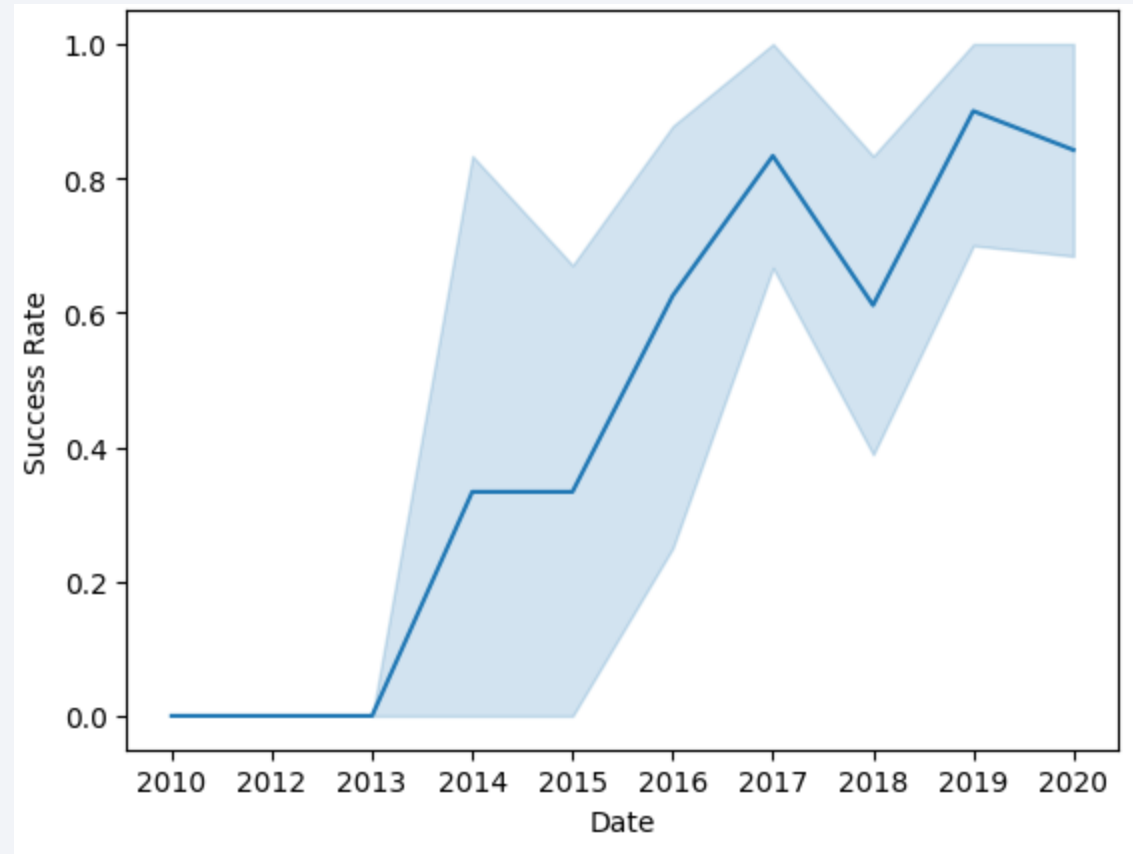
Payload vs. Orbit Type



Our analysis reveals no strong correlation between **payload mass and orbit type**. While certain orbit types show higher success rates at lower payload sizes, the overall data remains too varied to draw definitive conclusions. This suggests that factors other than payload mass are likely more influential in determining launch success across different orbit types.

Launch Success Yearly Trend

The yearly trend in launch success shows a consistent increase over time, with a minor dip observed in 2018. This trend aligns with expected advancements and refinements in launch technology, reflecting SpaceX's steady progress in achieving higher success rates.



All Launch Site Names

```
SELECT DISTINCT "Launch_Site" FROM "SPACEXTABLE";
```

This query selects distinct values from the **Launch_Site** column in the **SPACEXTABLE** database, effectively listing each unique launch site without repetition. The resulting launch sites are:

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

This analysis helps us understand where SpaceX's launches have historically taken place, aiding in our site-specific success analysis.

Launch Site Names Begin with 'CCA'

```
SELECT * FROM spacetable WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

This query selects all columns from **spacetable** where the **Launch_Site** column starts with "CCA," using the **LIKE** operator with a wildcard (%). By setting a **LIMIT** of 5, we get the first five matching records. This search helps us analyze site-specific patterns focused on Cape Canaveral (CCA) sites, which are key to SpaceX's historical launch data.

Done .									
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
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

To calculate the total payload mass carried for launches where NASA was the customer (specifically under NASA's Commercial Resupply Services), we used the following query:

```
SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM  
spacetable WHERE "Customer" LIKE '%NASA (CRS)%';
```

This query calculates the **SUM** of payload masses in the **PAYLOAD_MASS__KG_** column, filtered to include only records where the **Customer** column contains "NASA (CRS)." The total payload mass delivered under these conditions is **48,213 kg**. This measure allows us to assess SpaceX's payload contributions specifically for NASA's resupply missions.

Average Payload Mass by F9 v1.1

To find the average payload mass carried by SpaceX's **F9 v1.1** booster versions, we used the following query:

```
SELECT AVG("PAYLOAD_MASS__KG_") AS avg_booster_mass  
FROM SPACEXTABLE  
WHERE "Booster_Version" LIKE 'F9 v1.1%',
```

This query calculates the **average (AVG)** of payload masses in the **PAYLOAD_MASS__KG_** column, specifically filtering for records where the **Booster_Version** begins with "F9 v1.1." The resulting average payload mass for this booster version is approximately **2,534.67 kg**, giving insight into its typical payload capacity across recorded missions.

First Successful Ground Landing Date

To identify the date of SpaceX's first successful ground landing, we used the following query:

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

This query selects the **earliest date (MIN)** from the Date column, filtering specifically for rows where the **Landing_Outcome** was recorded as a successful landing on a **ground pad**. The result reveals that SpaceX achieved its first successful ground landing on **December 22, 2015**, marking a significant milestone in their reusability program.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

This query filters the **Booster_Version** column for rows where the **Landing_Outcome** was a successful drone ship landing and where **PAYLOAD_MASS_KG_** is within the specified range. The result shows the specific **booster versions** that managed these mid-weight payloads, providing insights into the payload capacity and landing consistency of different booster versions. Below are our results:

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

Total Number of Successful and Failure Mission Outcomes

```
SELECT
  CASE
    WHEN "Mission_Outcome" LIKE '%Success%' THEN 'Success'
    ELSE 'Failure'
  END AS outcome_type,
  COUNT(*) AS total_outcomes
FROM SPACEXTABLE
GROUP BY outcome_type;
```

This query groups **Mission_Outcome** data by filtering for results that contain "Success" to classify each mission as either a **success** or **failure**. Here, a mission success does not necessarily mean a successful landing but rather that the primary mission goals were met, whether it was launch, payload delivery, or orbit deployment. The result helps clarify mission consistency, irrespective of landing outcome.

Boosters Carried Maximum Payload

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

This query first determines the maximum payload mass from the **PAYLOAD_MASS__KG_** column, then retrieves the **Booster_Version** entries associated with this maximum payload. The result lists each **Booster_Version** that achieved this maximum capacity, providing insight into the highest-performing boosters in terms of payload capacity.

```
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

```
SELECT
```

```
    substr("Date", 6, 2) AS  
month,
```

```
    "Landing_Outcome",
```

```
    "Booster_Version",
```

```
    "Launch_Site"
```

```
FROM spacetable
```

```
WHERE
```

```
    "Landing_Outcome" LIKE '%Failure  
(drone ship)%' – Filters for failed  
landings on drone ships
```

```
    AND substr("Date", 1, 4) = '2015';
```

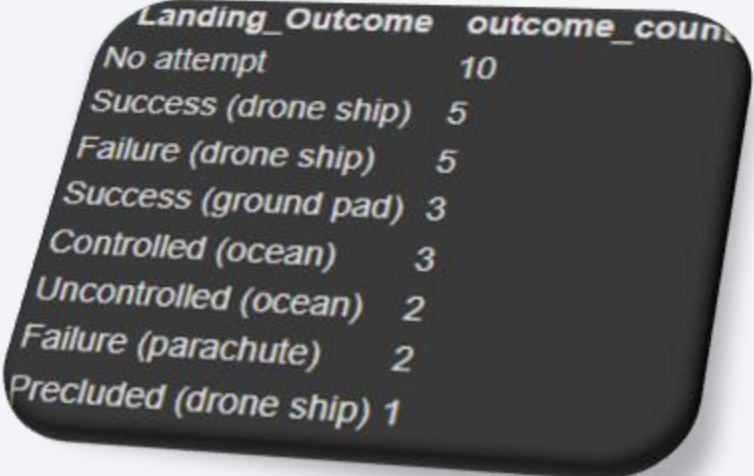
month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

This query filters for **failed landings on drone ships** in 2015, returning the **month**, **Landing_Outcome**, **Booster_Version**, and **Launch_Site** for each record. The output allows us to analyze unsuccessful landing attempts on drone ships, specifically by month and location, helping to identify trends or patterns in these instances.

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

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

This query provides a ranked list of landing outcomes during the specified period, allowing us to see the most common outcomes. The results show that there were **10 instances of 'No attempt,'** followed by **5 successes and failures** for drone ship landings, indicating a predominance of no attempts during this timeframe. This ranking highlights trends in landing outcomes and can be useful for further analysis on mission success rates.



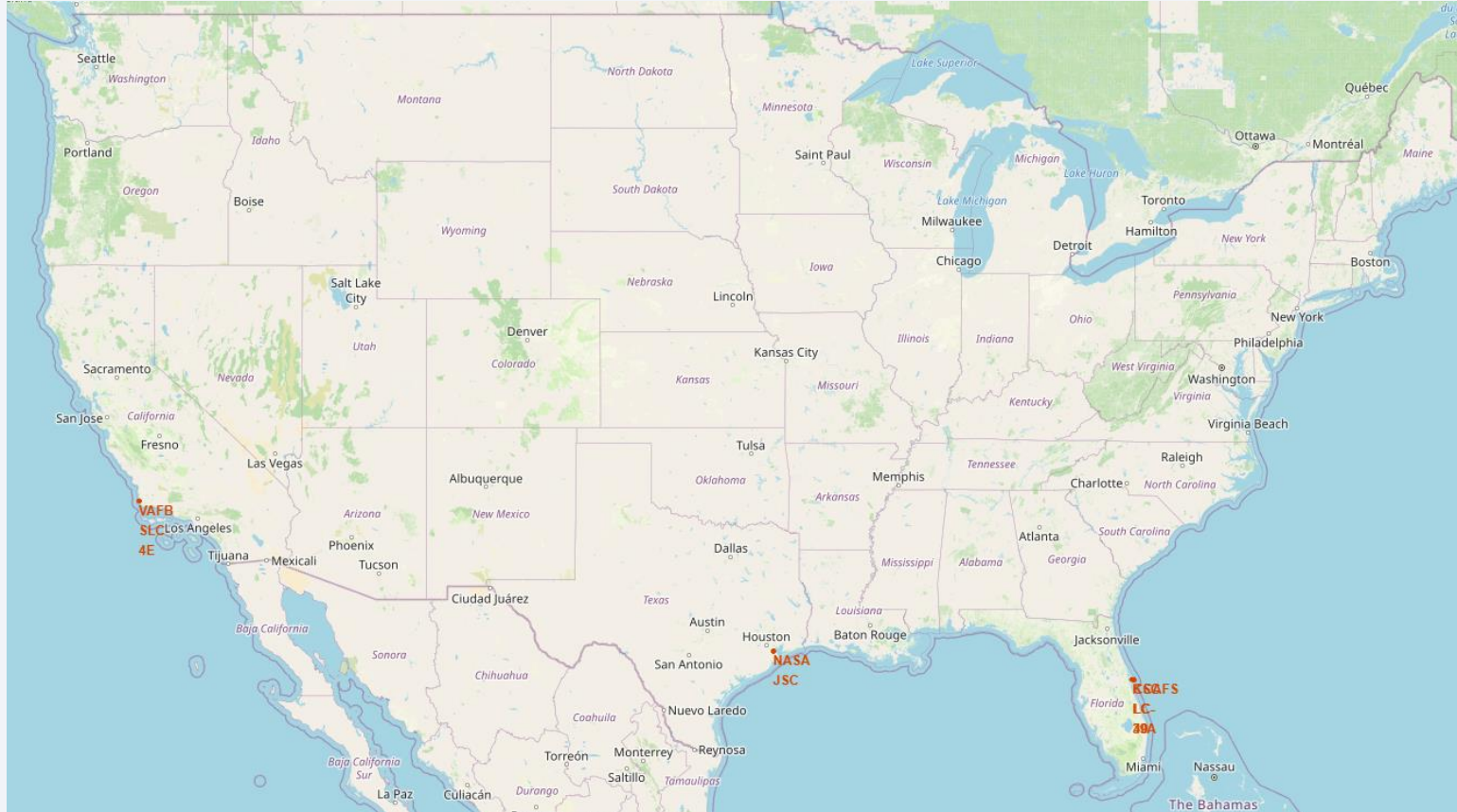
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

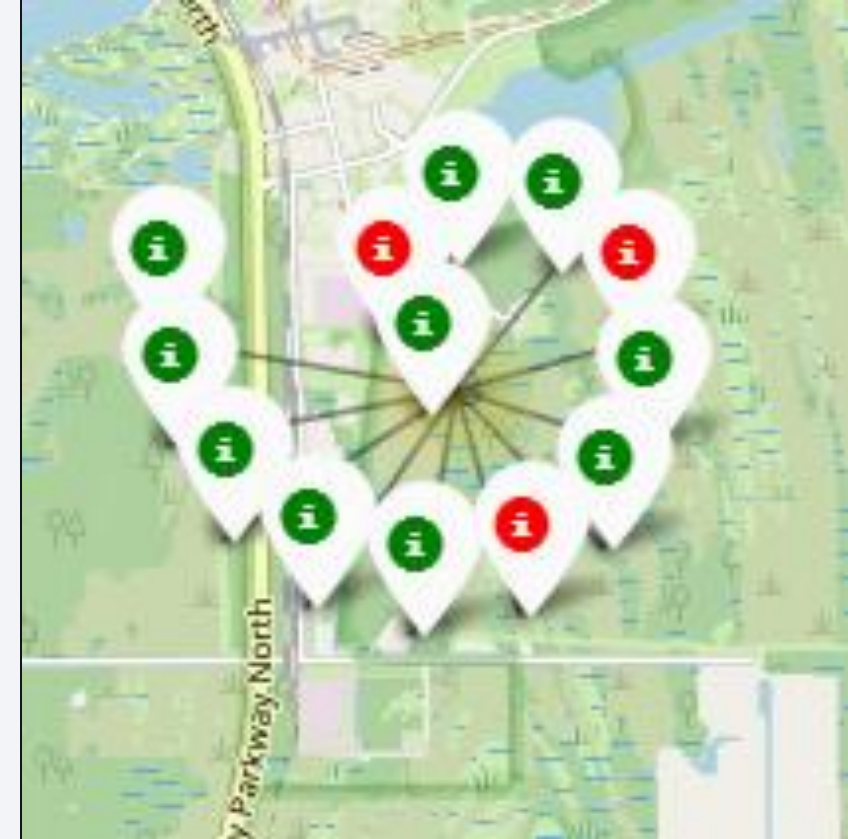
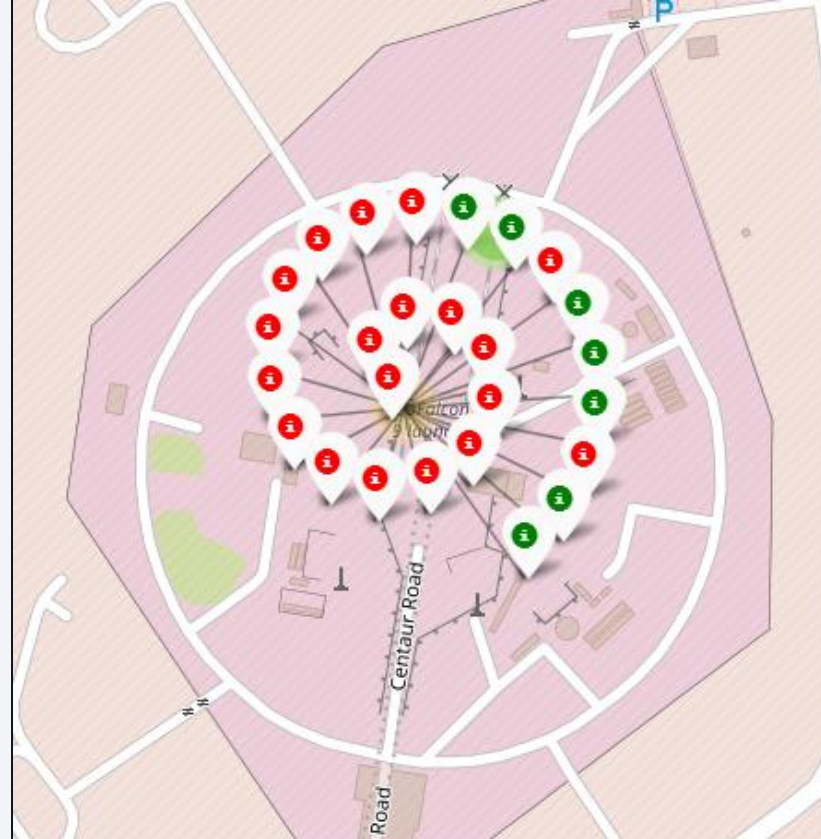
Launch Sites Proximities Analysis

Launch Site Locations - Folium



We have 4 main launch sites in the U.S.

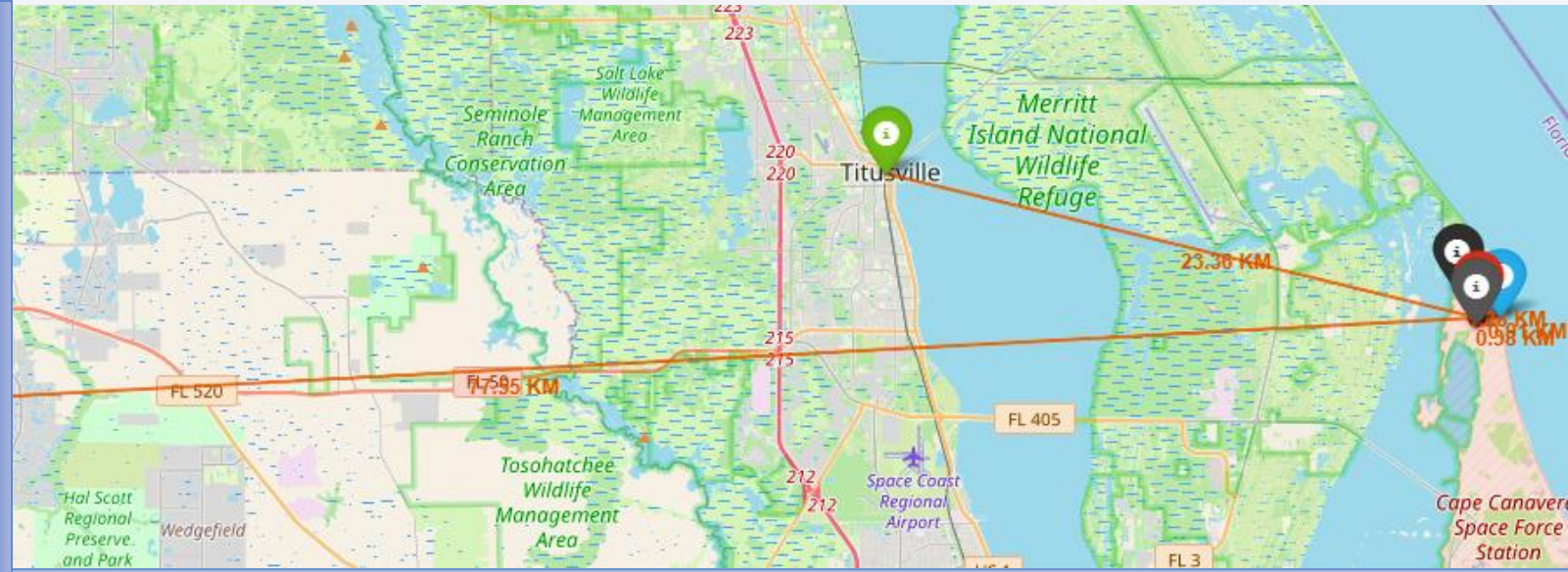
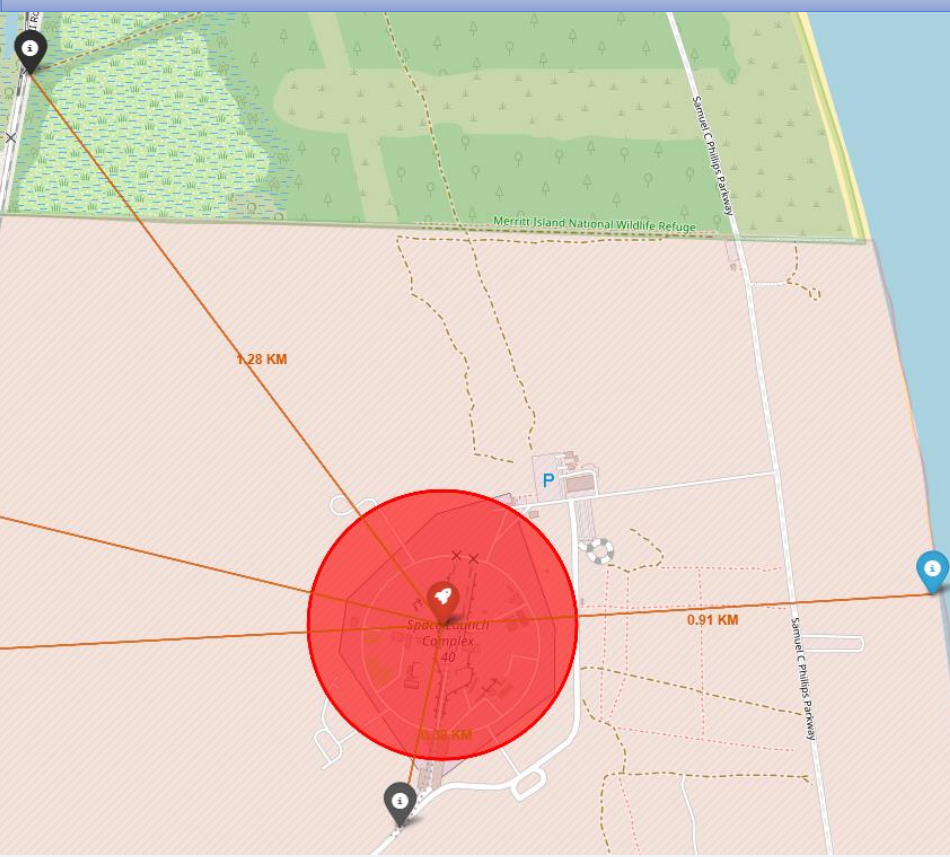
- Two located in Florida
 - CCAFS LC-40
 - KSC LC-39A
- One in Texas
 - NASA JSC
- One in California
 - VAFB SLC-4E



Success Rate per Site - Folium

To better understand launch outcomes, we visualized each launch site's success and failure data in a distinct spiral pattern. This layout provides a clear view of individual launch details, accessible by clicking any data point for more specific information. For an interactive experience, [visit our GitHub repository to explore the full map!](#)

Proximity Landmarks - Folium



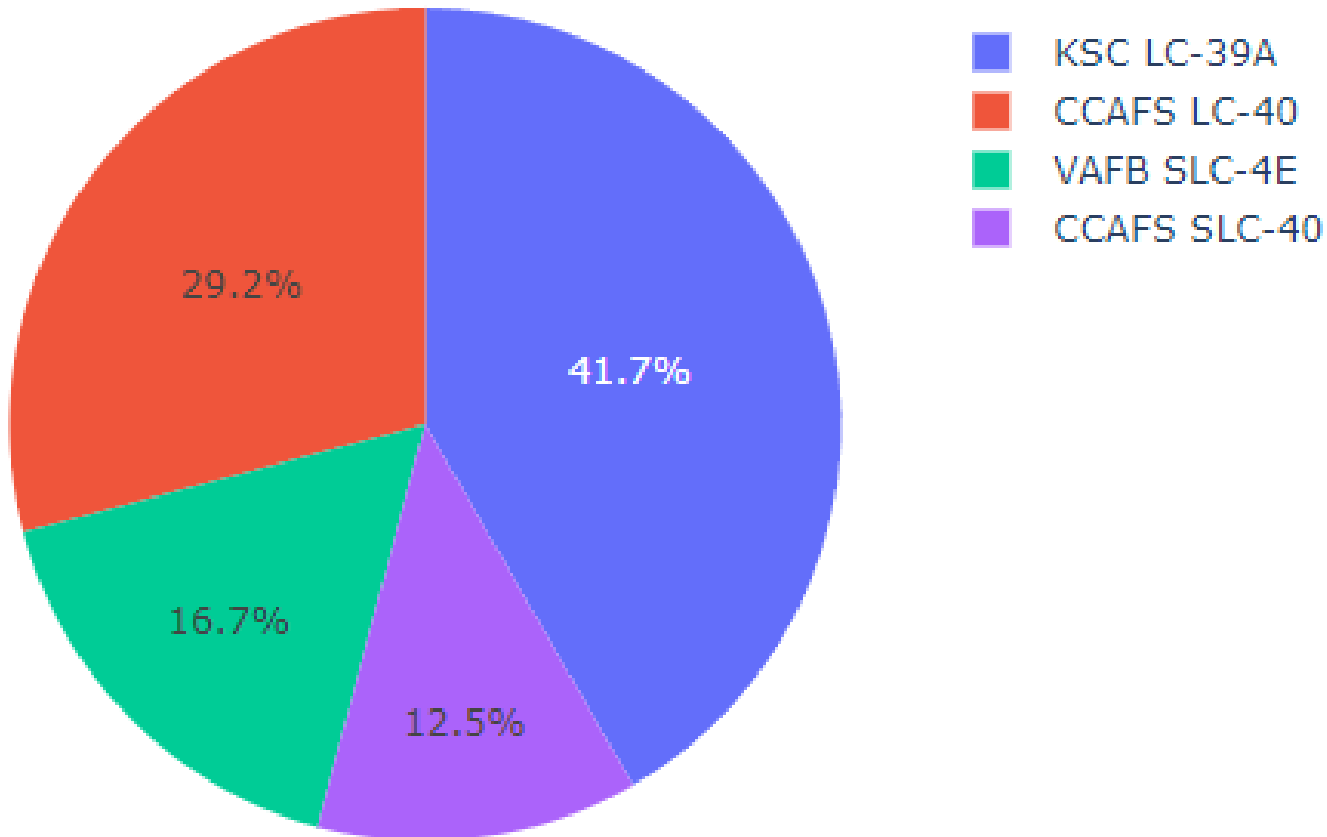
- In the selected screenshot, CCAFS LC-40 is positioned approximately 1 km from the coastline, ~75 km from Orlando, and ~20 km from smaller nearby cities like Titusville. This launch site, like others, maintains a safe distance from major urban areas while staying close to a large body of water. These proximities are likely intentional, balancing safety and logistical access to water routes for transport and potential recovery operations.



Section 4

Build a Dashboard with Plotly Dash

Total Success Launches By Site

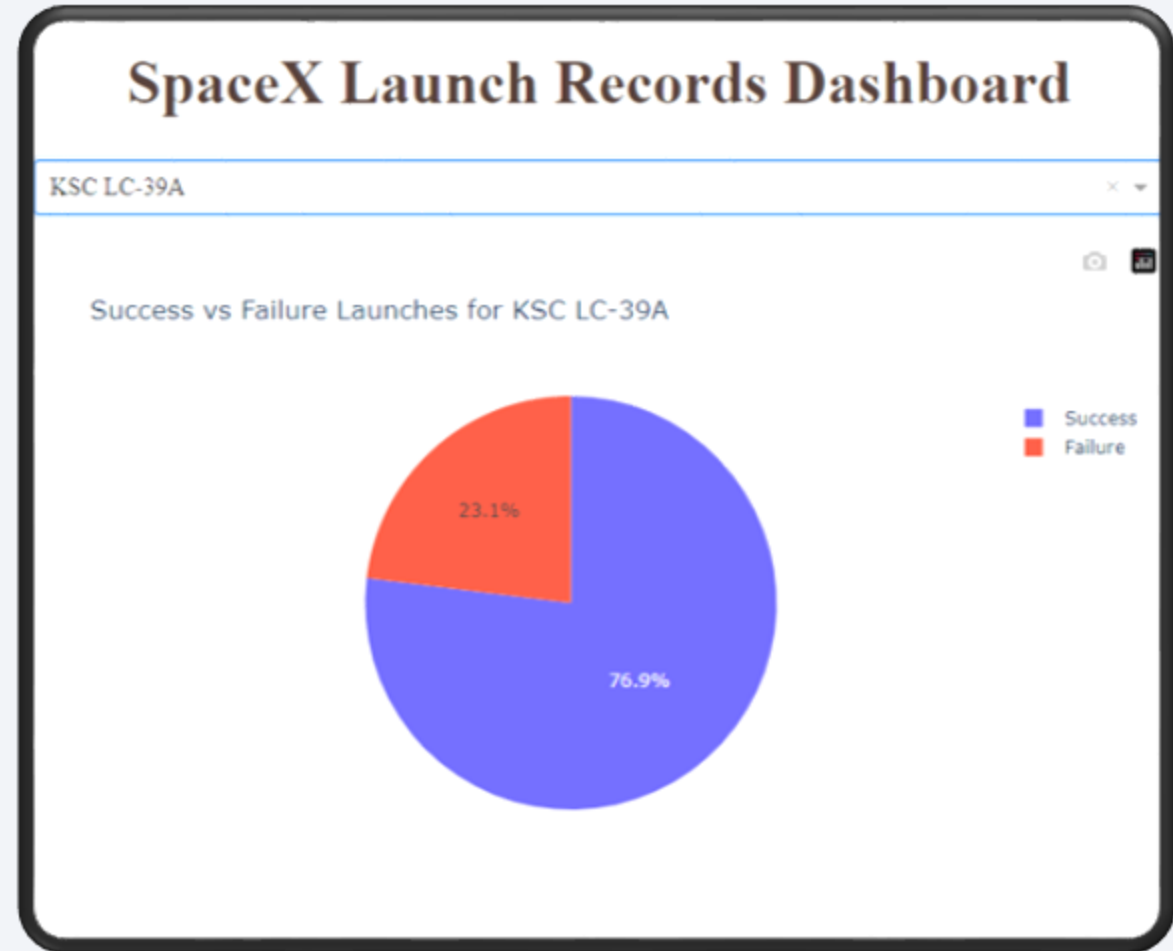


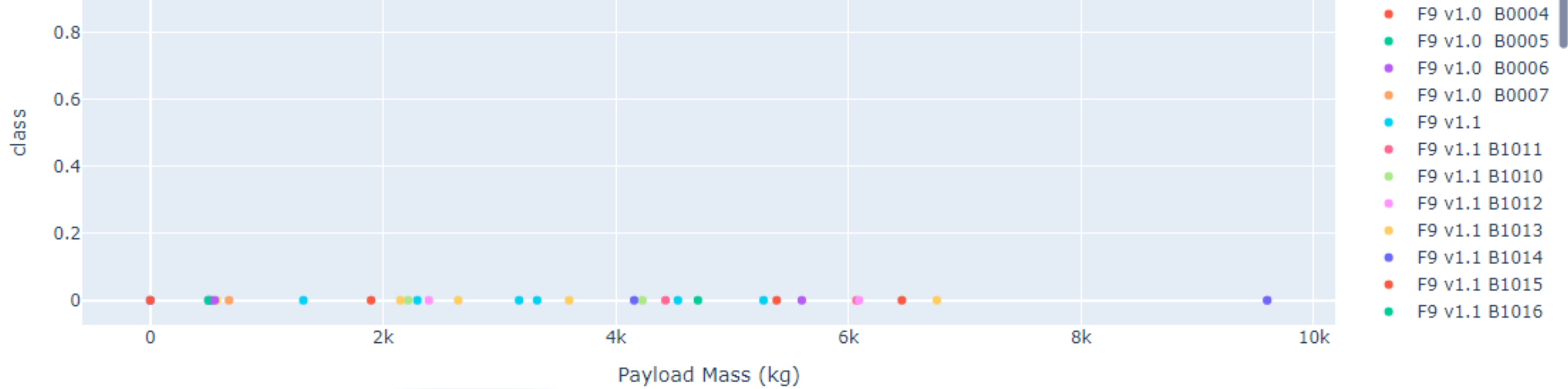
Total Success Pie Chart - Dashboard

Our dashboard's default pie chart shows the overall success rate by launch site, allowing us to drill down to view each site's specific success-to-failure ratio. See the next slide for further insights.

Highest Success - Dashboard

Drilling down further, we see that KSC LC-39A holds the highest success rate among our four launch sites, with a 76.9% success rate.





Payload Mass and Success - Dashboard

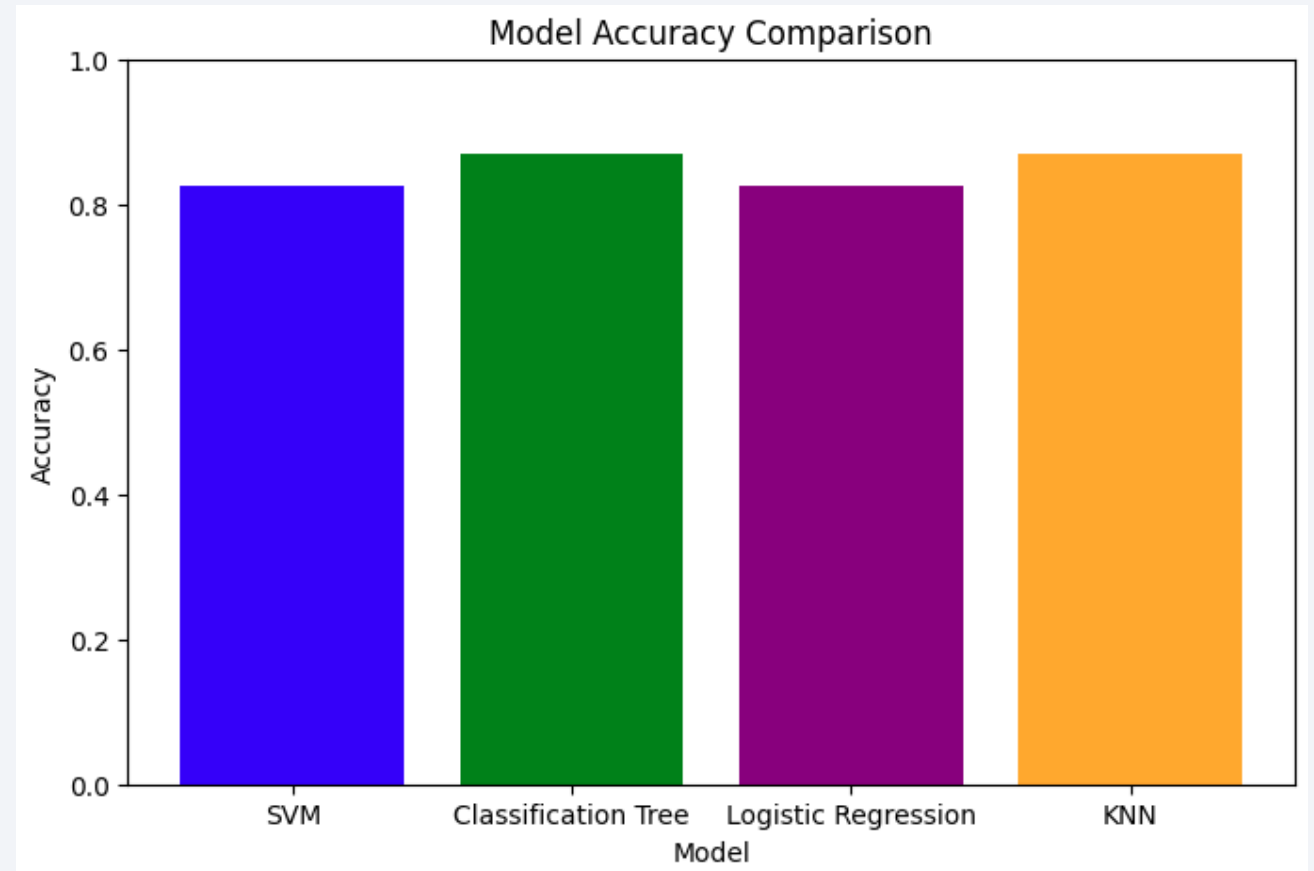
Analyzing Payload Mass versus Launch Outcomes by booster type, we observe no strong correlation up to 5300kg. However, failures become more frequent starting at 5600kg, with one notable success outlier at the maximum payload weight of 9600kg

Section 5

Predictive Analysis (Classification)

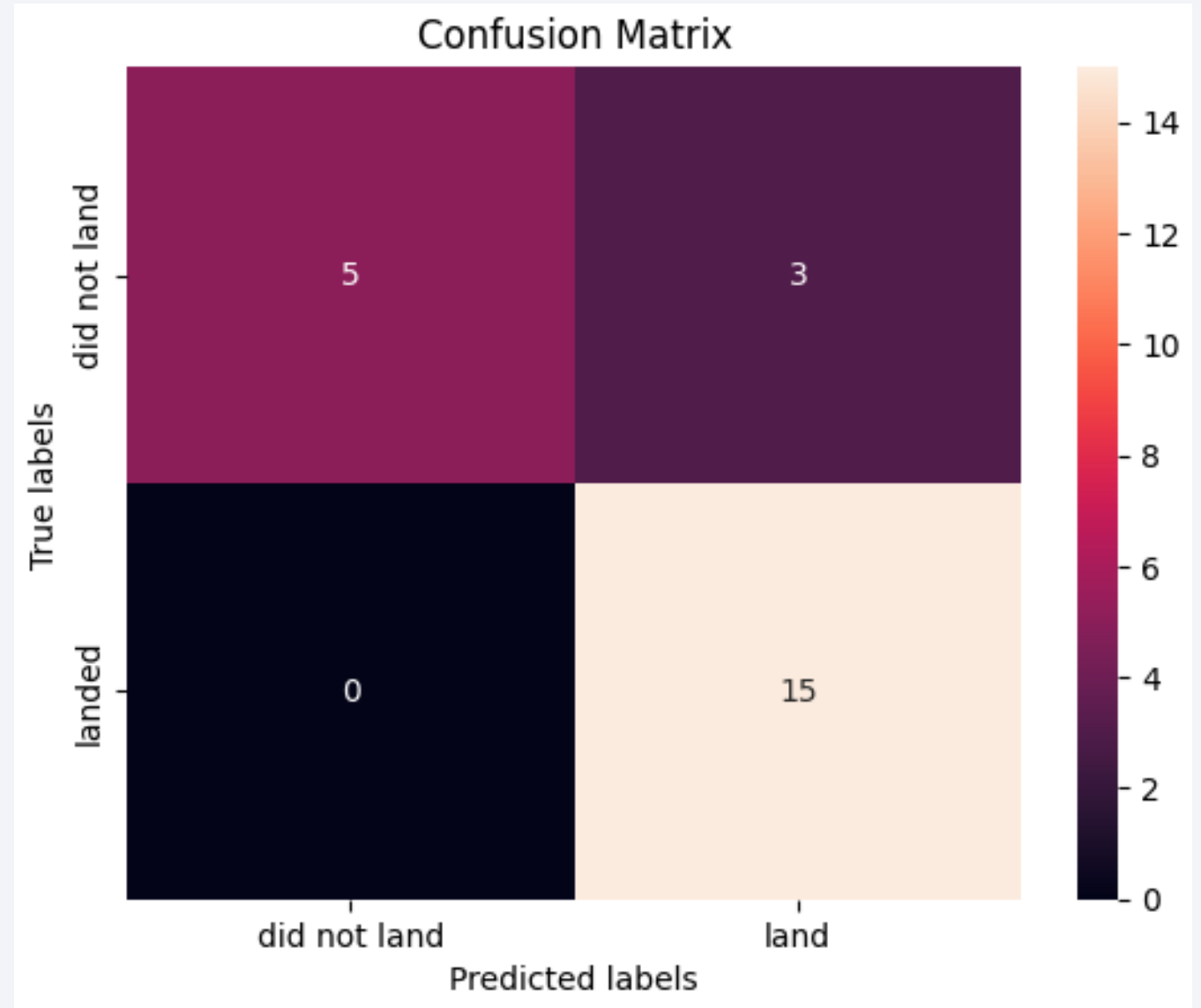
Classification Accuracy

As shown in our bar chart, the **Classification Tree (86.7%)** and **KNN (86.7%)** models demonstrate the highest accuracy among the models tested.



Confusion Matrix

This confusion matrix indicates that the model has a reasonably good accuracy in predicting whether a SpaceX landing was successful or not, with 15 true positives and only 3 false negatives.



Conclusions

- Analyzed SpaceX launch data to **identify patterns and improve predictive models** for launch success, particularly **focusing on factors such as payload mass, orbit type, and launch site**.
- **Identified significant trends**, such as an increase in launch success with a higher number of flight experiences at certain sites.
- Developed a **user-friendly dashboard in Plotly Dash** to enable **dynamic filtering** and exploration of launch data by site, payload, and success rates.
- Explored multiple models (SVM, Decision Tree, Logistic Regression, KNN) for predicting launch success, with **KNN models showing the highest accuracy at 86.7%**.
- The data suggests **launch site experience and type of orbit have tangible effects** on launch success rates.
- Refining models by incorporating additional external factors (e.g., weather data, technological advancements) could **further improve prediction accuracy**.

Appendix

[Github Repo \(click here\)](#)

Python

- # Pandas is a software library written for the Python programming language for data manipulation and analysis.
 - import pandas as pd
 - import numpy as np
 - import matplotlib.pyplot as plt
 - import seaborn as sns
 - from sklearn import preprocessing
 - from sklearn.model_selection import train_test_split
 - from sklearn.model_selection import GridSearchCV
 - from sklearn.linear_model import LogisticRegression
 - from sklearn.svm import SVC
 - from sklearn.tree import DecisionTreeClassifier
 - from sklearn.neighbors import KNeighborsClassifier

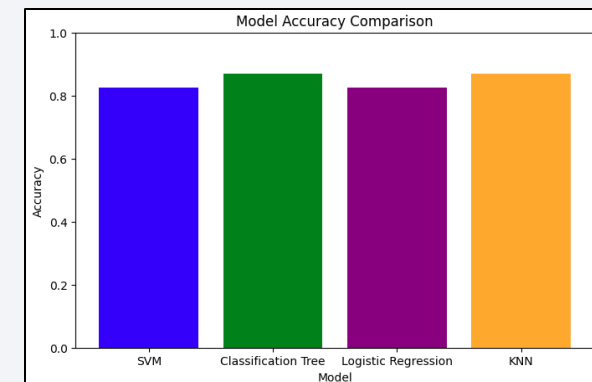
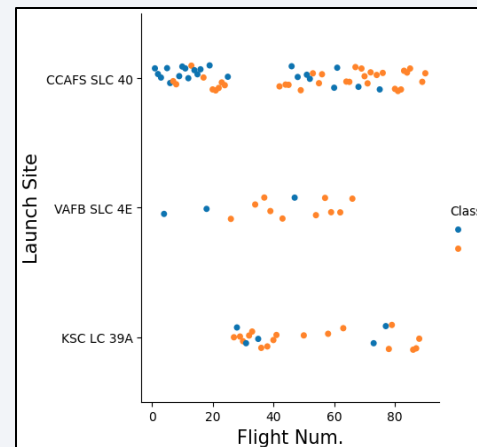
SQL

```
%%sql
select distinct * from
"SPACEXTABLE";
```

Data Source

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv

Key Charts



Thank you!

