IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eirene Michella Tjhan
15 October 2025

Outline

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

Executive Summary

This project analyzed SpaceX launch data to understand mission outcomes and predict future success.

Methodologies

Data was collected using the SpaceX REST API and web scraping from Wikipedia. After cleaning and merging the data, we performed exploratory data analysis (EDA) using SQL and visualizations to find key trends in payloads, launch sites, and orbit success rates. Interactive maps and charts were created using Folium and Plotly Dash to explore launch locations and mission results more clearly.

Predictive Analysis

For predictive analysis, several classification models were trained, including Logistic Regression, SVM, Decision Tree, and KNN. After tuning with GridSearchCV, the best model was chosen based on accuracy performance. Results showed that SpaceX's success rate improved greatly over time, with consistent successful landings after 2016. The final model accurately predicted mission success, showing that data-driven analysis can support better planning for future launches.

Introduction

Project Background and Context

SpaceX has carried out many orbital launches aimed at advancing reusable rocket technology and reducing space transportation costs. However, not all missions have been successful, and analyzing these outcomes helps identify the key factors influencing mission performance. This project focuses on examining launch data, visualizing patterns, and applying predictive models to better understand what drives successful landings.

Problems

- What technical or environmental factors (e.g., payload mass, orbit type, or launch site) most affect mission success?
- How have SpaceX's launch outcomes evolved from 2010 to recent years, and what trends can be observed?
- Can a classification model accurately predict the success of future SpaceX missions based on past data?
- What insights can be used to support future decision-making and reliability improvements in rocket launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was gathered using the SpaceX REST API and web scraping from Wikipedia to obtain complete launch and mission details.
- Perform data wrangling
  - Cleaned, merged, and standardized datasets by handling missing values and duplicates for consistent analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Trained and tuned Logistic Regression, SVM, Decision Tree, and KNN models, selecting the one with the highest accuracy.

Data Collection

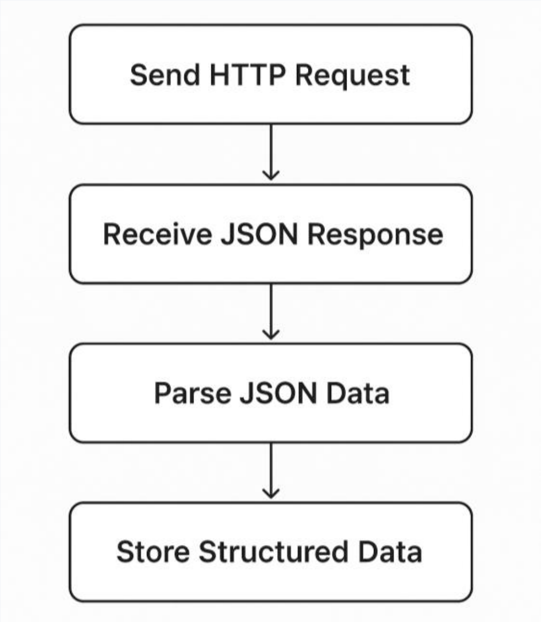
- **API**: Data were collected using the SpaceX REST API, which provides structured and real-time launch data directly from SpaceX's official database.

- **Web Scraping**: Additional data were extracted from SpaceX's official website using web scraping techniques to gather launch details not available through the API.

Data Collection – SpaceX API

SpaceX REST API Data Collection

- Accessed structured JSON data directly from SpaceX REST API endpoints.
- Retrieved information about launches, payloads, rockets, and launch sites.
- Used requests library in Python to send GET requests and parse responses.
- Converted JSON responses into a pandas DataFrame for analysis.
- Ensured data consistency and real-time accuracy through official API access.

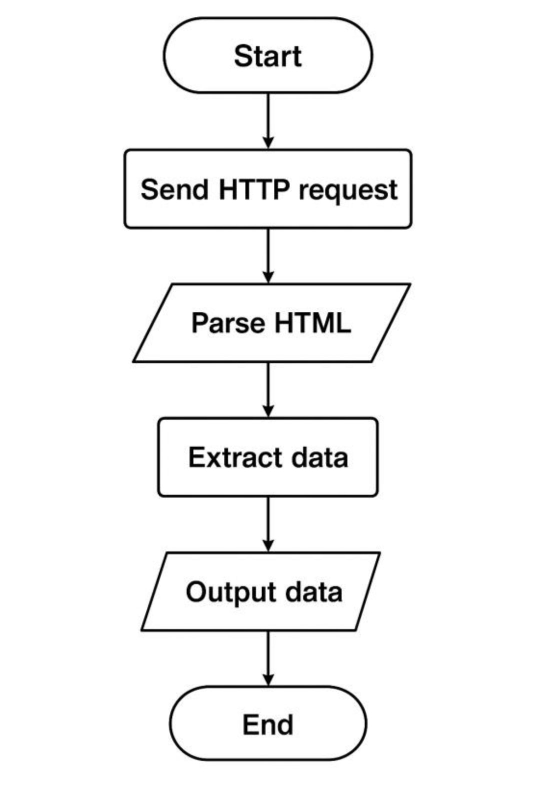


[GitHub Link]

Data Collection - Scraping

Web Scraping Data Collection

- Collected additional launch details from SpaceX's official website.
- Used BeautifulSoup and requests libraries for HTML parsing and extraction.
- Targeted tables containing historical launch data (dates, payloads, outcomes).
- Cleaned and structured scraped data into a DataFrame.
- Merged scraped data with API data to complete missing fields and enhance dataset richness.

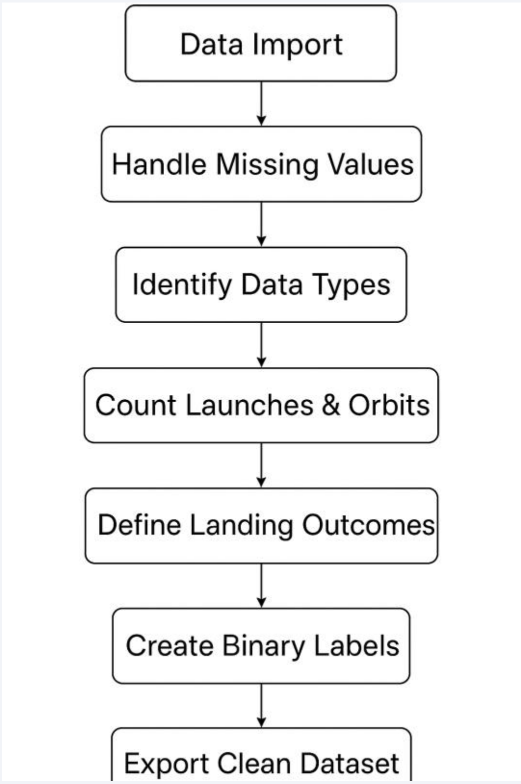


[GitHub Link]

Data Wrangling

Data Processing Summary:

- **Data Loading:** Imported SpaceX dataset (dataset_part_1.csv) using pandas.
- **Missing Values:** Checked null values; found 28.88% missing in the LandingPad column.
- **Data Types:** Identified categorical (e.g., Orbit, LaunchSite) and numerical (e.g., PayloadMass, Longitude) columns.
- **Feature Exploration:** Counted launches per site (value_counts()), orbit types, and mission outcomes to understand data distribution.
- **Data Cleaning:** Grouped similar outcomes into success/failure sets using logical conditions.
- **Label Creation:** Created a binary Class column (1 = successful landing, 0 = failed) based on the Outcome column.
- **Export:** Saved the cleaned dataset as dataset_part_2.csv for further analysis.



GitHub Link

EDA with Data Visualization

Data Visualization Summary:

- **Flight Number vs. Launch Site (Scatter Plot):** Used to identify launch frequency and distribution across different sites.
- **Payload vs. Launch Site (Scatter Plot):** Visualized how payload mass varied by launch site.
- **Orbit Type vs. Success Rate (Bar Chart):** Compared success rates across orbits; ES-L1, GEO, HEO, and SSO showed nearly 100% success, while GTO had the lowest.
- **Flight Number vs. Orbit Type (Scatter Plot):** Showed the relationship between launch sequence and targeted orbits.
- **Payload vs. Orbit Type (Scatter Plot):** Illustrated how payload capacities differ across orbit categories.
- **Success Rate Over Time (Line Graph):** Displayed SpaceX's performance trend, highlighting steady improvement after 2014 and consistent success post-2016.

GitHub Link

EDA with SQL

SQL Query Summary:

- Retrieved unique launch site names using SELECT DISTINCT.
- Filtered launches starting with “CCA” using LIKE 'CCA%'.
- Calculated total and average payload mass for NASA and specific boosters using SUM() and AVG().
- Identified first successful ground pad landing by sorting dates.
- Listed boosters with successful drone ship landings and payloads between 4000-6000 kg.
- Counted successful and failed mission outcomes using COUNT() and conditional grouping.
- Found boosters carrying the maximum payload using MAX().
- Filtered failed drone ship landings in 2015 using substr() and date filtering.
- Ranked landing outcomes between specific dates using GROUP BY and ORDER BY.

GitHub Link

Build an Interactive Map with Folium

1. Map Objects Created

- **Markers:**
  - Placed at each **launch site** (e.g., CCAFS SLC-40, KSC LC-39A, etc.)
  - Added **colored pins (green/red)** to indicate **launch success (1)** or **failure (0)**.
- **Circles:**
  - Highlighted each **launch site location** with orange circles to make them more visible on the map.
- **Marker Clusters:**
  - Grouped multiple launches at the same coordinates to avoid overlapping markers and improve readability.
- **Lines (Polylines):**
  - Connected each launch site to nearby **highway, railway, coastline, and cities** to visualize proximity.
- **Distance Labels (DivIcons):**
  - Displayed **calculated distances (in km)** between launch sites and nearby features directly on the map.
- **MousePosition Plugin:**
  - Enabled real-time display of **latitude and longitude** when hovering over the map.

Build an Interactive Map with Folium

2. Purpose of Adding These Objects

- **Markers and Circles:** To identify and visually differentiate each launch site.
- **Colored Icons:** To represent **launch outcomes**, making success/failure patterns easy to interpret.
- **MarkerCluster:** To simplify visualization of multiple launches from the same site.
- **Lines & Distance Labels:** To analyze **proximity relationships** between launch sites and nearby infrastructures.
- **MousePosition Plugin:** To interactively measure and locate new points of interest (e.g., coastline, railway).

GitHub Link

Build a Dashboard with Plotly Dash

In this Plotly Dash dashboard, I added multiple interactive plots to analyze SpaceX launch performance data. The dashboard includes:

1. **Launch Success Pie Chart** – Displays the proportion of successful and failed launches. When “All Sites” is selected, it shows the total success rate across all launch sites. When a specific site is chosen, it updates to show that site’s individual success rate.
2. **Payload vs. Launch Outcome Scatter Plot** – Shows the relationship between payload mass (kg) and launch outcomes (success or failure). The data points are color-labeled according to the **Booster Version Category**, allowing us to identify which booster versions perform best under different payload ranges.
3. **Interactive Components:**
   - A **dropdown menu** for selecting the launch site (all sites or specific ones).
   - A **payload range slider** that dynamically filters the scatter plot by payload mass range.

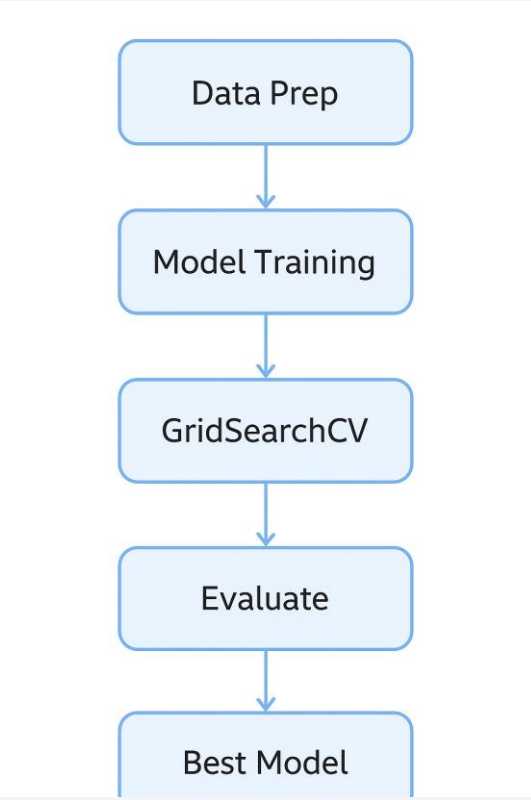
GitHub Link

Predictive Analysis (Classification)

Process:

GitHub Link

- **Data Preparation:**
  - Loaded datasets, extracted Class as target (Y)
  - Standardized features (X) with StandardScaler
  - Split data (80% train, 20% test)
- **Model Building & Tuning:**
  - Trained **Logistic Regression**, **SVM**, **Decision Tree**, and **KNN**
  - Used **GridSearchCV (cv=10)** to find best hyperparameters
- **Evaluation:**
  - Compared test accuracies using .score()
  - Visualized with bar chart (gray = all models, blue = best model)
- **Result:**
  - Selected the **best-performing classifier** based on highest accuracy



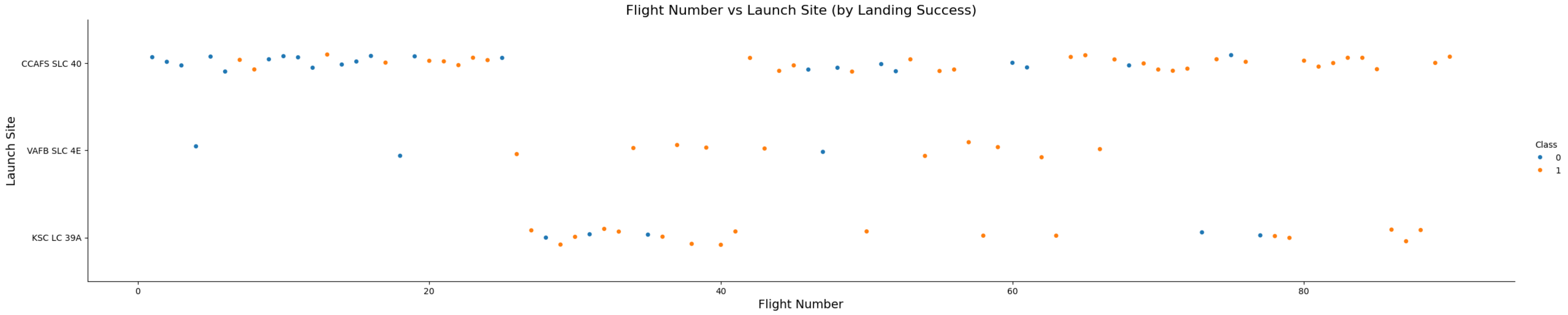
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

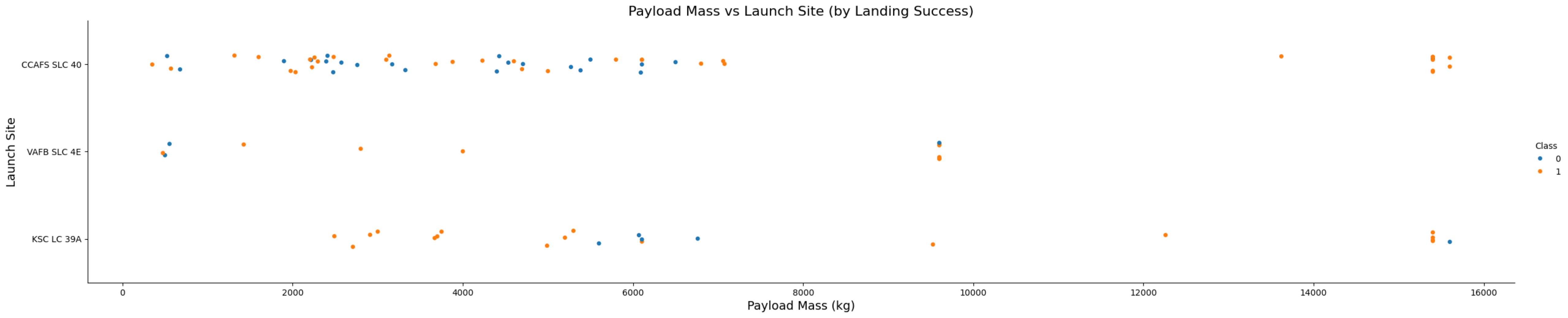
Insights drawn from EDA

Flight Number vs. Launch Site



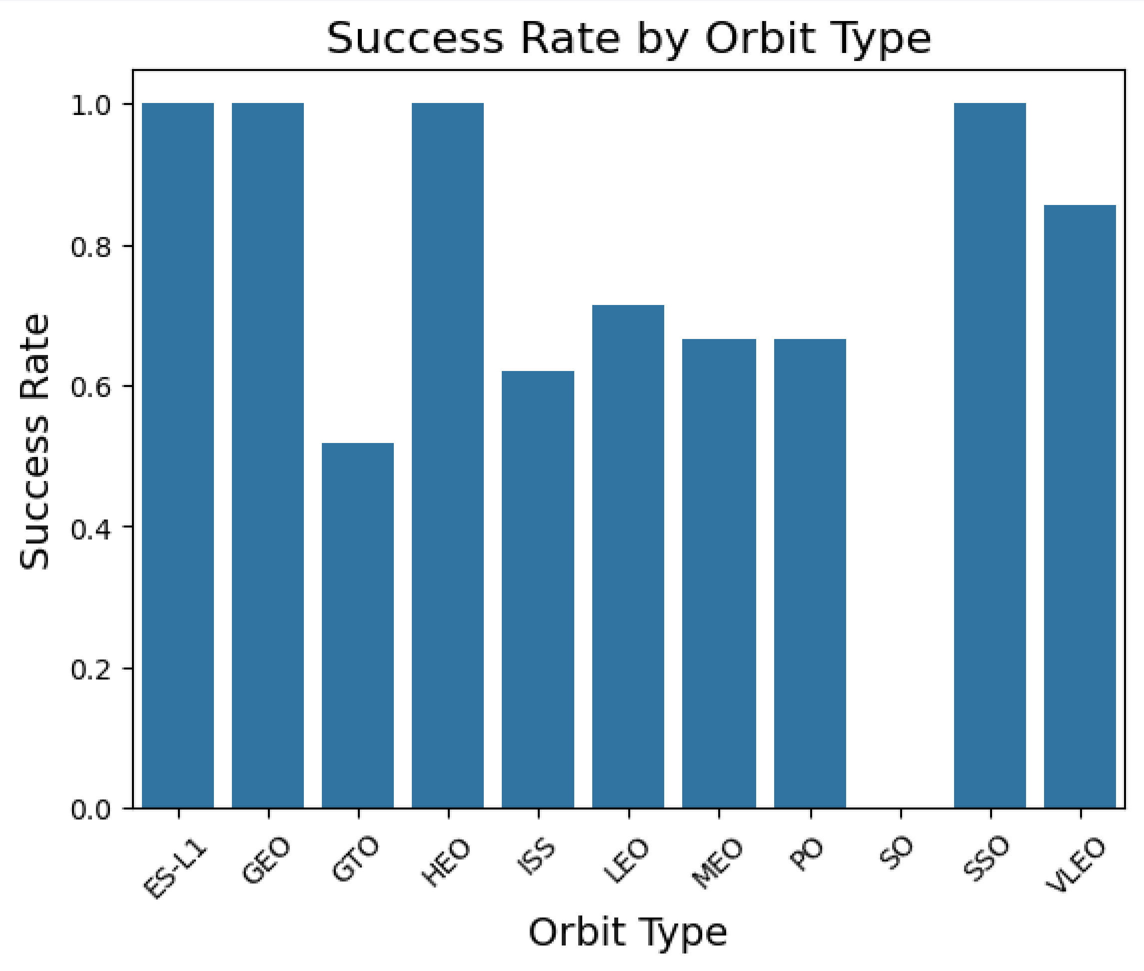
From the plot, we can see that in the early flights (lower Flight Numbers), there were more failed landings (blue points) at all launch sites. As the flight numbers increased, the number of successful landings (orange points) became higher, showing that SpaceX improved its landing technology over time. At the **CCAFS SLC 40** site, there were many launches and a mix of both successes and failures. The **KSC LC 39A** site shows mostly later flights, and most of them were successful. The **VAFB SLC 4E** site had fewer launches overall but showed consistent improvement too.

Payload vs. Launch Site



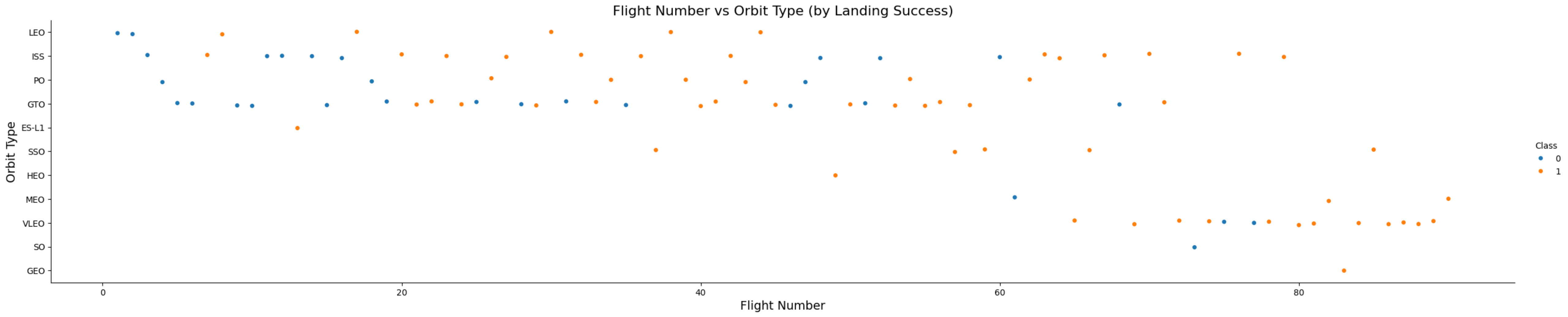
From the chart, we can see that the **VAFB SLC 4E** launch site did not handle rockets with heavy payloads (above 10,000 kg). Most high-payload launches happened at **CCAFS SLC 40** and **KSC LC 39A**, where both successful and failed landings occurred. This suggests that heavier payload missions were mainly launched from these two sites, while VAFB SLC 4E focused on lighter missions.

Success Rate vs. Orbit Type



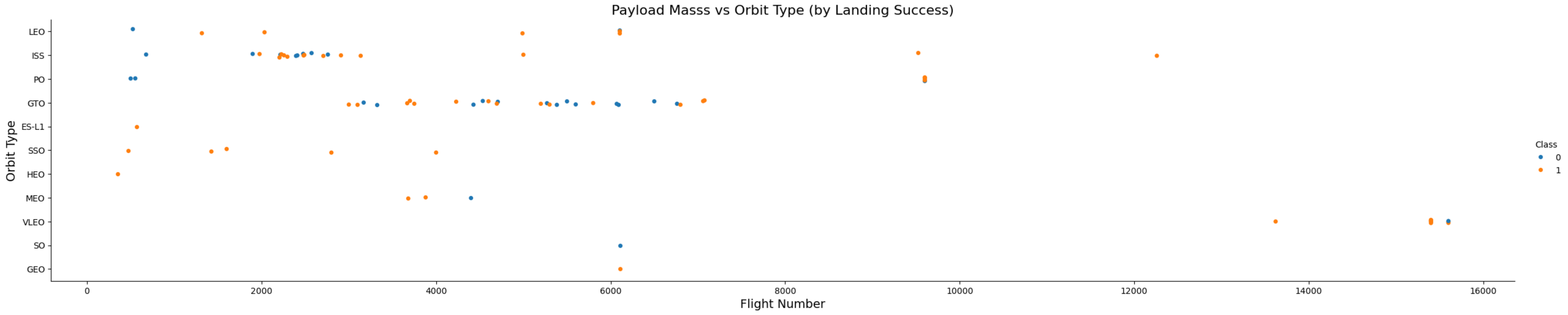
From the chart, the orbits **ES-L1**, **GEO**, **HEO**, and **SSO** show the highest success rates, all close to 100%. In contrast, **GTO** has the lowest success rate, meaning missions to that orbit faced more landing failures.

Flight Number vs. Orbit Type



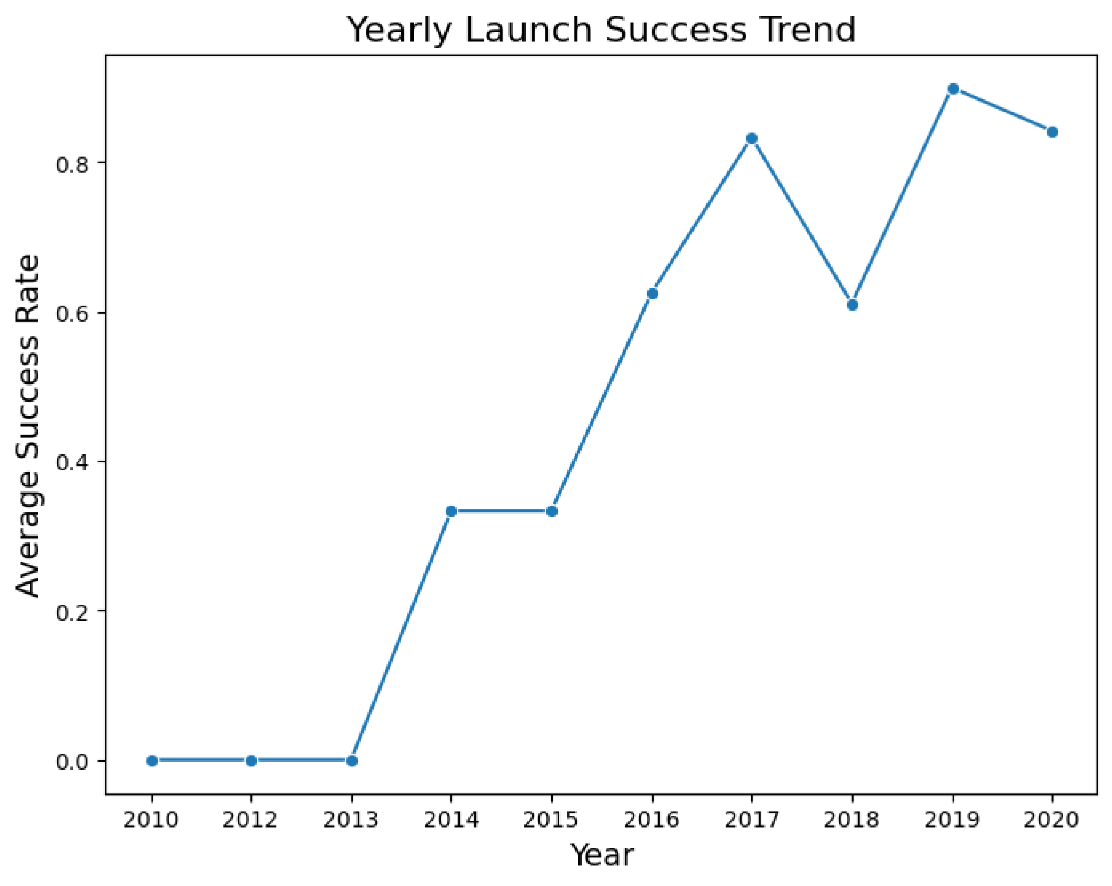
From the chart, we can see that in the **LEO orbit**, success increases with higher flight numbers, showing improvement over time. However, for the **GTO** orbit, there is no clear pattern, both successful and failed landings occur across different flight numbers.

Payload vs. Orbit Type



From the chart, we can see that heavier payloads tend to have more successful landings in **Polar, LEO, and ISS** orbits. In contrast, for the GTO orbit, both successful and failed landings occur, making it harder to see a clear pattern of success.

Launch Success Yearly Trend



The chart shows that SpaceX's success rate increased over time. From 2010 to 2013, most launches failed, but starting in 2014, the success rate rose steadily. After 2016, SpaceX achieved consistent success, showing major improvement in rocket reliability.

All Launch Site Names

Here we're trying to obtain all the **Launch Site** value that are available on the dataset. As we can see there are 4 (four) Launch Site name.

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

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

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

Launch Site Names Begin with 'CCA'

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

Python

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

The query uses `LIKE 'CCA%'` to filter launch sites that start with "CCA" and `LIMIT 5` to show only the first five records. The results all come from **CCAFS LC-40**, showing early SpaceX missions between 2010-2013, mostly successful launches with varying landing outcomes.

Total Payload Mass

```
%sql SELECT "Booster_Version", SUM("Payload_Mass__kg_") AS "Total_Payload_kg" FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)' GROUP BY "Booster_Versi
```

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

Booster_Version	Total_Payload_kg
F9 B4 B1039.2	2647
F9 B4 B1039.1	3310
F9 B4 B1045.2	2697
F9 B5 B1056.2	2268
F9 B5 B1058.4	2972
F9 B5 B1059.2	1977
F9 B5B1050	2500
F9 B5B1056.1	2495

This query not only filters launches for NASA (CRS) but also groups them by each booster version, calculating the total payload mass per booster.

Average Payload Mass by F9 v1.1

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

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

Average_Payload_Mass
2534.6666666666665

The query uses `AVG()` to calculate the **average payload mass** for all records where the booster version starts with '**F9 v1.1**'. The result, **2534.67 kg**, shows that on average, each F9 v1.1 booster carried about **2.5 tons** of payload during its missions.

First Successful Ground Landing Date

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

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

First_Successful_Landing_Date	Landing_Outcome
2015-12-22	Success (ground pad)

This query filters rows where the landing outcome was '**Success (ground pad)**', sorts them **by date** in ascending order, and returns the earliest record. It identifies the **first time SpaceX successfully landed a booster on a ground pad**.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

The query filters records where the landing outcome was '**Success (drone ship)**' and the payload mass was **between 4000 and 6000 kg**. It returns boosters **F9 FT B1022, F9 FT B1026, F9 FT B1021.2**, and **F9 FT B1031.2**, all of which successfully landed on a drone ship carrying medium-to-heavy payloads.

Total Number of Successful and Failure Mission Outcomes

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

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

Outcome	Total_Missions
Failure	1
Success	100

- This query **standardizes** mission outcomes containing “Success” and groups everything else as “Failure.” It then counts each category, giving a clear total number of successful and failed missions.
- The query result shows **1 failed mission** and **100 successful missions**, meaning SpaceX achieved a **99% success rate**. This indicates a very high mission reliability, with only one launch failure throughout the dataset.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE
    WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

The query identifies boosters with the **maximum payload mass** by comparing each record's payload to the dataset's highest value using a subquery with MAX().

The result shows that several **Falcon 9 Block 5 (F9 B5)** boosters, such as **B1048.4, B1049.4, B1051.3,** and others, each carried the **maximum payload of 15,600 kg**, highlighting the strong lifting capability and consistent performance of the Block 5 series.

2015 Launch Records

```
%sql SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE
    WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date, 1, 4) = '2015';
```

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

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

The query filters records where the **landing outcome was "Failure (drone ship)"** and the **launch year was 2015**, then extracts the month using substr(Date, 6, 2).

The result shows two failed drone ship landings in **January (B1012)** and **April (B1015)**, both launched from **CCAFS LC-40**, indicating early challenges in SpaceX's 2015 drone ship recovery attempts.

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

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

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

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

The query groups landing outcomes between **2010-06-04** and **2017-03-20**, counts each type using COUNT(*), and ranks them in descending order.
The results show **"No attempt" (10)** as the most frequent, followed by **5 successful** and **5 failed drone ship landings**, indicating that during this early period, many missions did not yet include landing attempts, while SpaceX was actively developing and testing recovery techniques.

Section 3

Launch Sites Proximities Analysis

Launch Sites' Location

Launch Sites' Location

1. Are all launch sites in proximity to the Equator line?

- No, **not all launch sites are close to the Equator**.
- The **Kennedy Space Center (KSC)** and **Cape Canaveral Space Force Station (CCAFS)** in Florida are located around **latitude 28° N**, which is relatively closer to the Equator (0°) compared to other sites.
- The **Vandenberg Space Force Base (VAFB)** in California, however, is located around **latitude 34° N**, which is much farther from the Equator.
- **Explanation:** Launch sites near the Equator are preferred for missions requiring equatorial orbits because Earth's rotational speed provides an additional velocity boost, reducing fuel consumption. However, sites farther from the Equator, like Vandenberg, are used for **polar or sun-synchronous orbits**, which require launches toward the north or south rather than eastward.

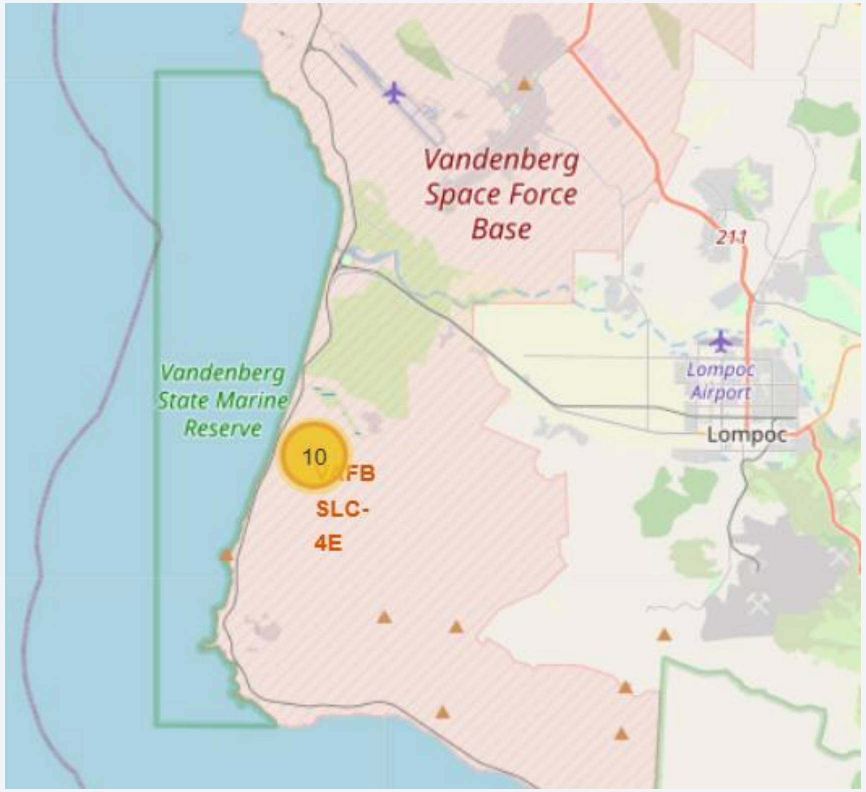
2. Are all launch sites in very close proximity to the coast?

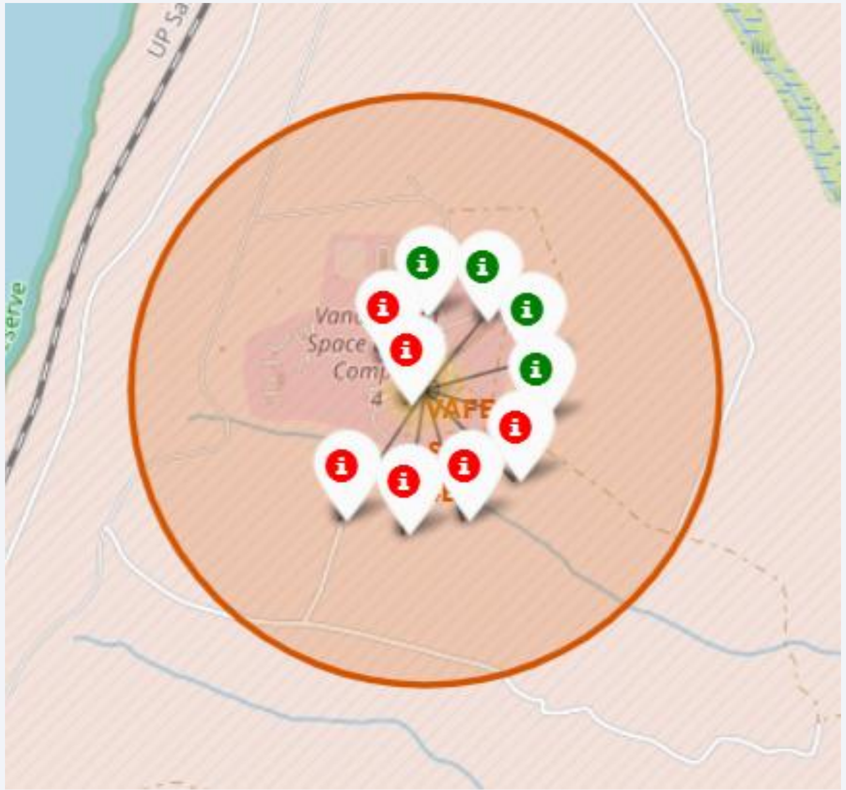
- Yes, **all launch sites are located near coastlines**.
- **Explanation:** Launch sites are deliberately placed near the ocean for **safety and trajectory reasons**:
- Rockets are launched over the sea to avoid populated areas in case of failure or debris fall.
- Coastal access allows rockets to take various orbital trajectories without crossing land.
- It's also easier to transport large rocket parts by ship.

Conclusion:

- While all SpaceX launch sites are **close to the ocean**, only the **Florida sites** are **relatively near the Equator**. This reflects their **different mission profiles**—Florida for equatorial launches, California for polar ones.

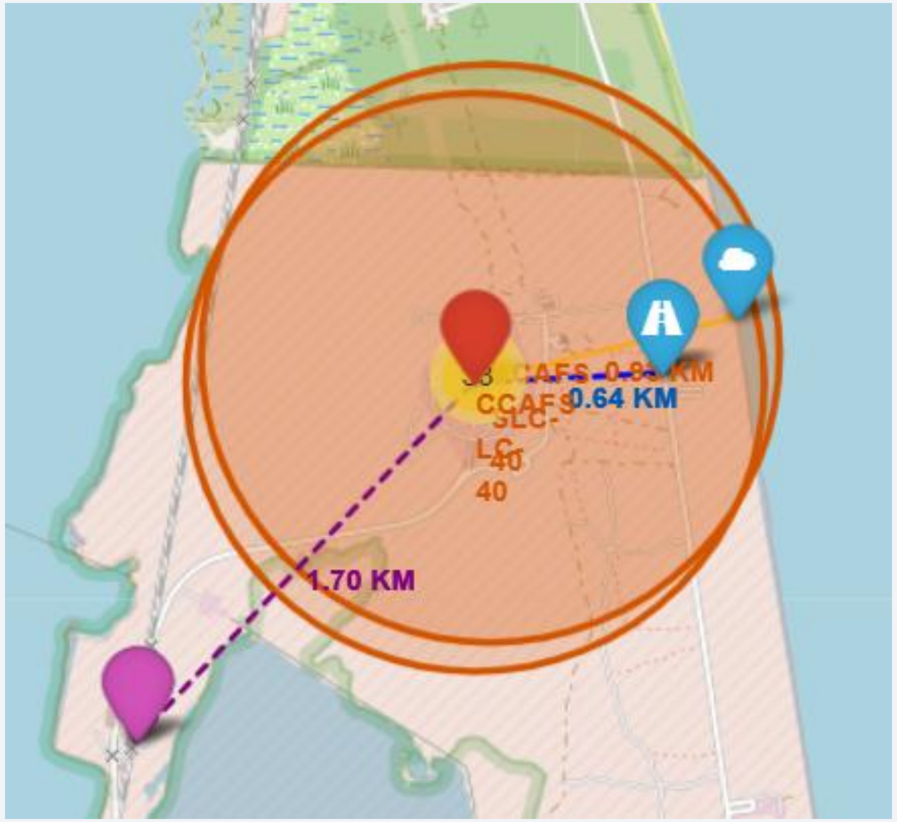
Success & Failed Launches





From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

Distances Between a Launch Site to its Proximities



1. Are launch sites in close proximity to railways?

- **Yes**, launch sites are relatively close to railways—about **1.70 km away**.
- **Explanation:**
- This short distance indicates that the sites are strategically placed near rail lines to **facilitate easy transportation** of heavy rocket components, fuel, and equipment from manufacturing facilities to the launch site.

Distances Between a Launch Site to its Proximities

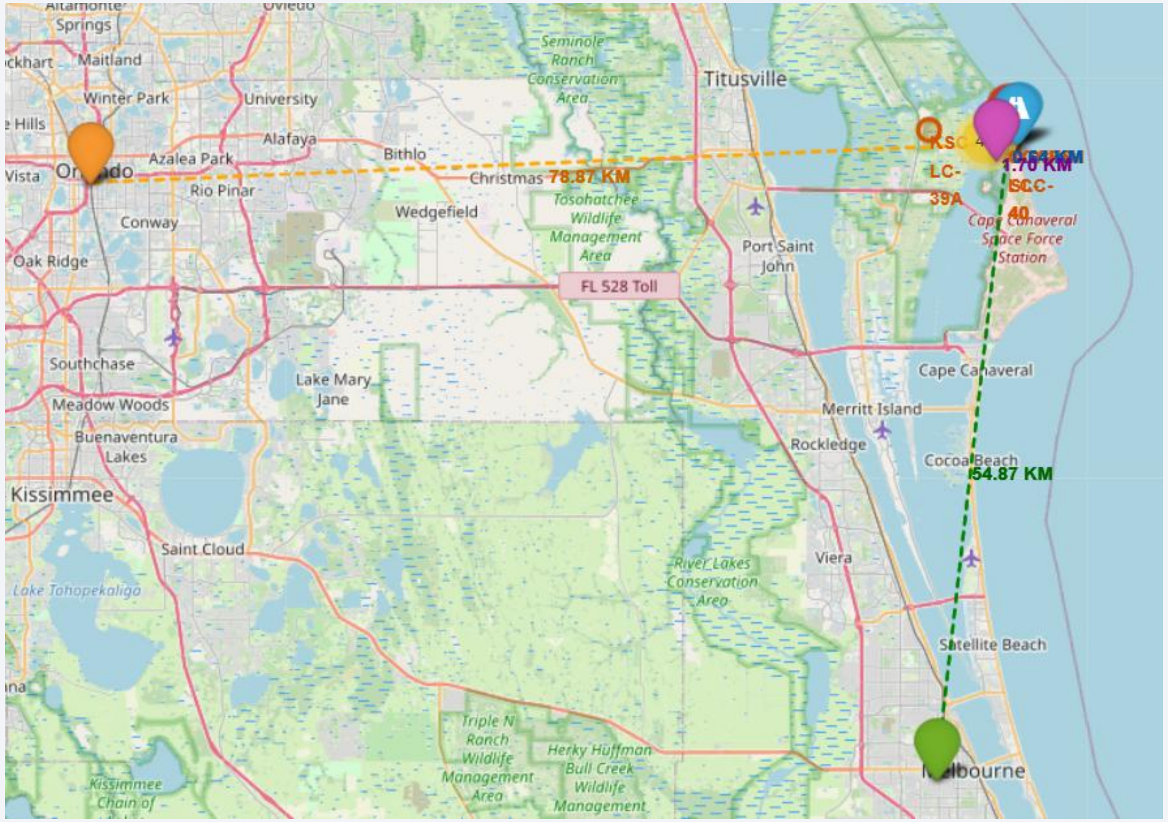
2. Are launch sites in close proximity to highways?

- **Yes**, the sites are **very close** to major highways—only **0.64 km away**.
- **Explanation:**
- Proximity to highways allows for **efficient ground transport and logistics**, especially for personnel and smaller cargo. It helps ensure smooth operations and accessibility during launch preparations.

3. Are launch sites in close proximity to the coastline?

- **Yes**, each site is **less than 1 km (0.93 km)** from the coastline.
- **Explanation:**
- Being close to the coast is intentional—launches near water reduce risk to populated areas if a failure occurs, and rockets can **safely fly eastward over the ocean**, taking advantage of Earth's rotation for **fuel efficiency**.

Distances Between a Launch Site to its Proximities



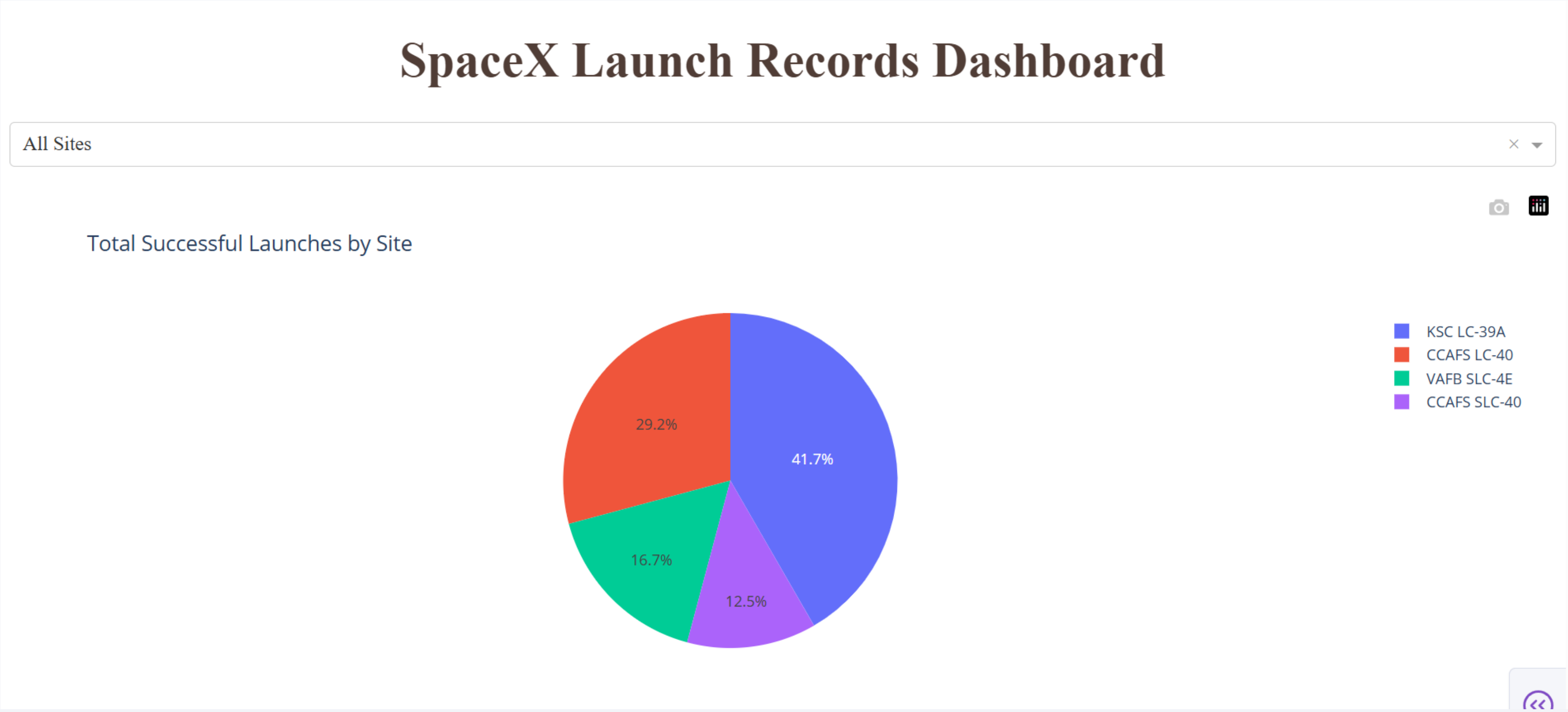
4. Do launch sites keep a certain distance away from cities?

- **Yes**, launch sites are **far from major cities**—about **54.87 km from Melbourne** and **78.87 km from Orlando**.
- **Explanation:**
- Maintaining a safe distance minimizes **risk to human populations** and infrastructure in the event of accidents or debris falls. It also reduces **noise pollution and vibration impacts** during liftoffs.

Section 4

Build a Dashboard with Plotly Dash

Total Successful Launches by Site



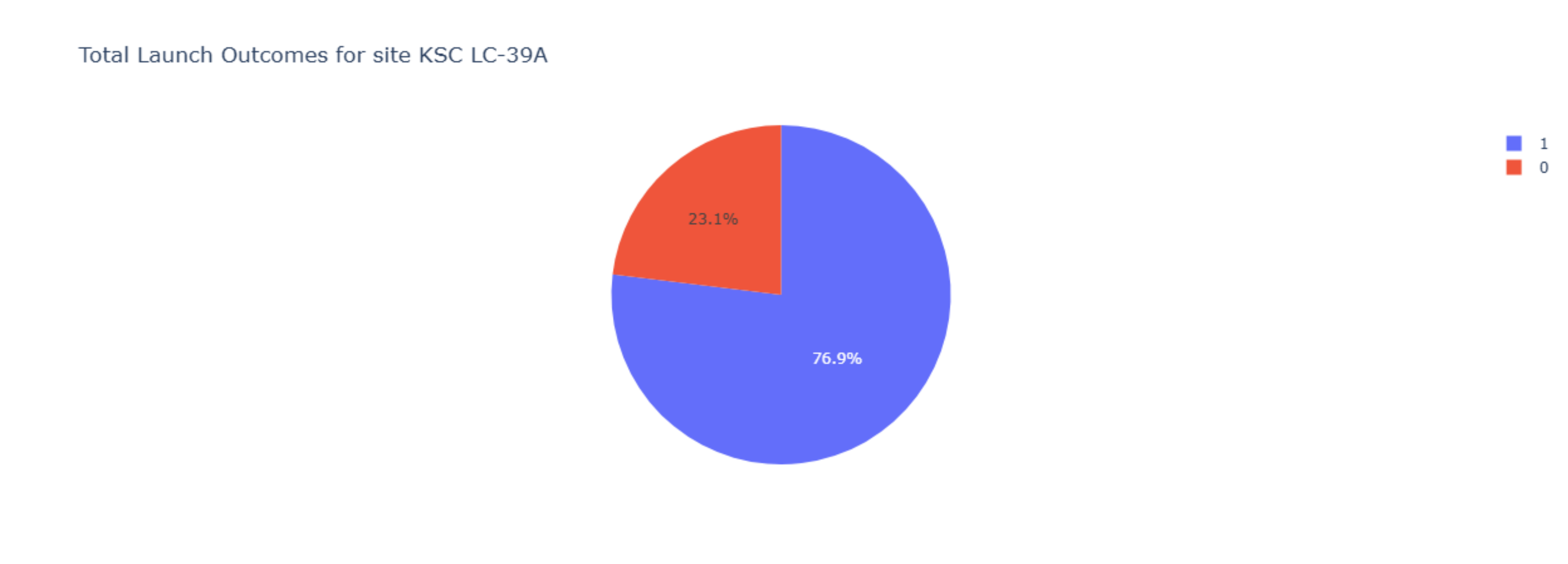
Which site has the largest number of successful launches?
KSC LC-39A has the largest proportion of successful launches (~41.7%).

Highest Launch Success Rate

KSC LC 39A

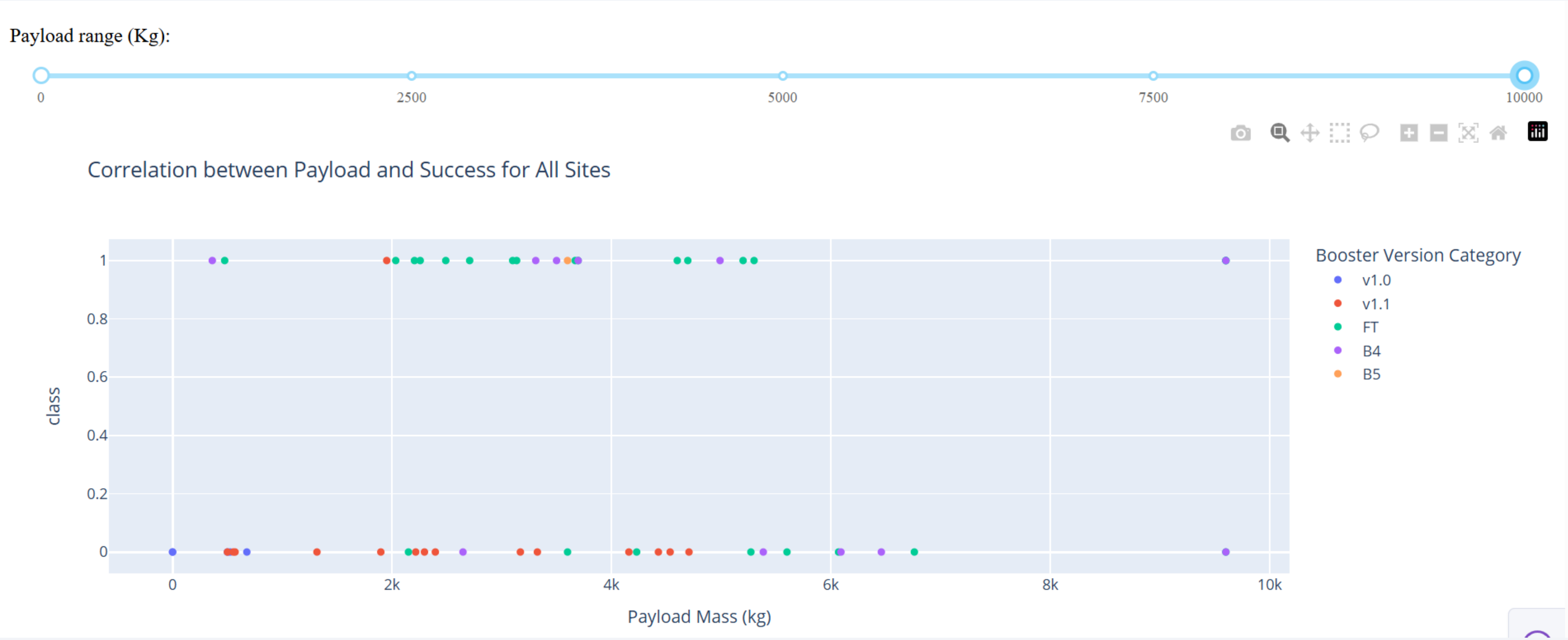
76.9 Success Rate

Highest Launch Success Rate



Which site has the highest launch success rate?
From further analysis (using the interactive dropdown to view each site's success vs failure), **KSC LC-39A** also shows the **highest success rate**, meaning it has the most consistent mission success among all SpaceX sites.

Payload vs. Launch Outcome for All Sites



Payload vs. Launch Outcome for All Sites

Highest Launch Success Rate

- The **payload range between 2000 kg and 6000 kg** shows the **highest launch success rate**.
- Most of the data points in this range are at **class = 1 (success)**.
- Most successful launches in this range are using **Booster Version FT** (green dots), which appears to perform very reliably.

Lowest Launch Success Rate

- The **payload range below 2000 kg** and around **6000-8000 kg** show the **lowest success rate**.
- In these ranges, there are several **class = 0 (failure)** points, especially for **older booster versions** such as **v1.0** and **v1.1** (blue and red dots).

Summary

- **Most successful payload range: 2000-6000 kg** (high success rate, mainly FT boosters).
- **Least successful payload ranges: 0-2000 kg** and **6000-8000 kg** (more failures, mostly older booster versions).

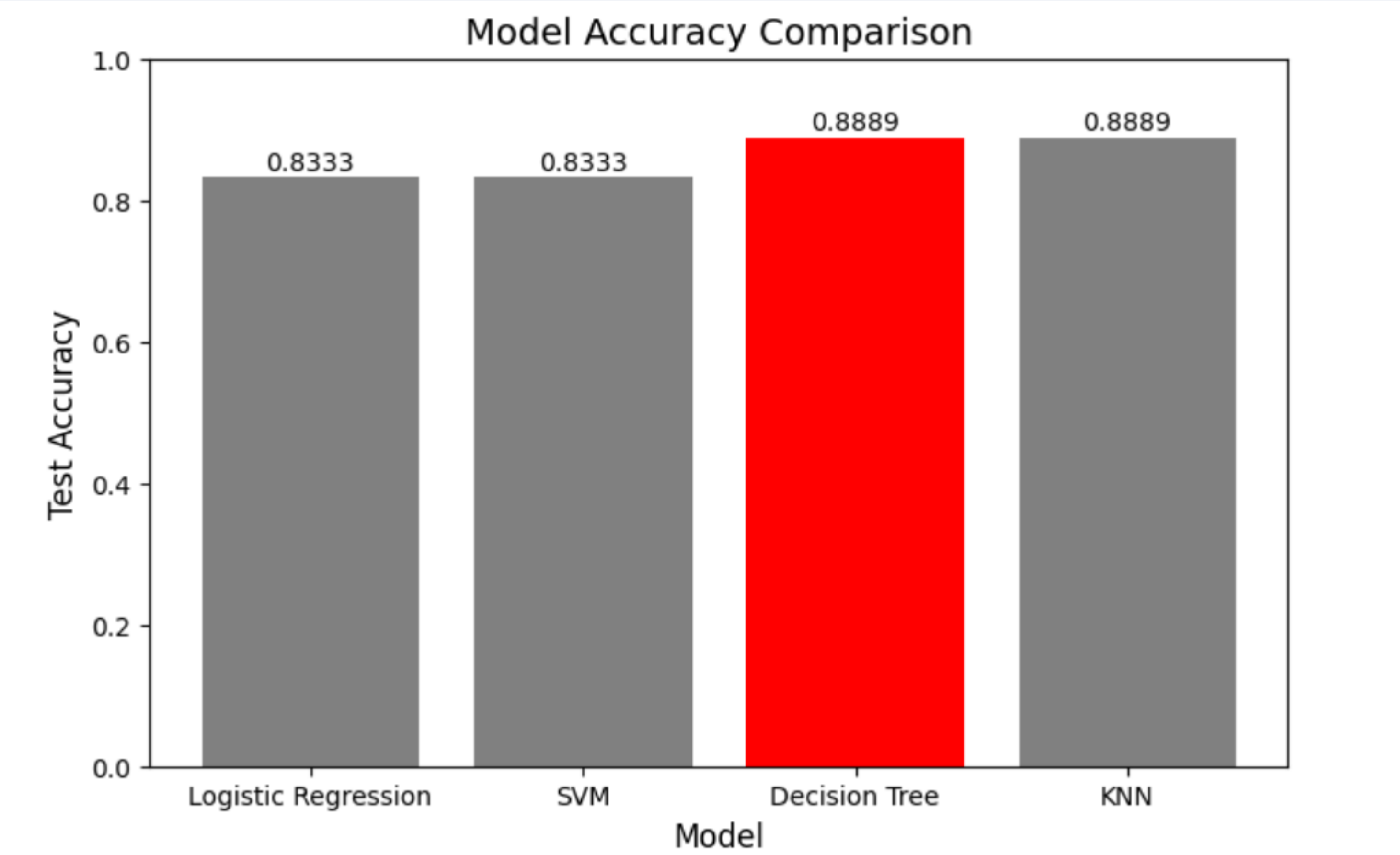
This suggests that **medium payloads (2-6 tons)** are the most stable and reliable range for SpaceX launches, especially when using **modern booster versions**.

Section 5

Predictive Analysis (Classification)

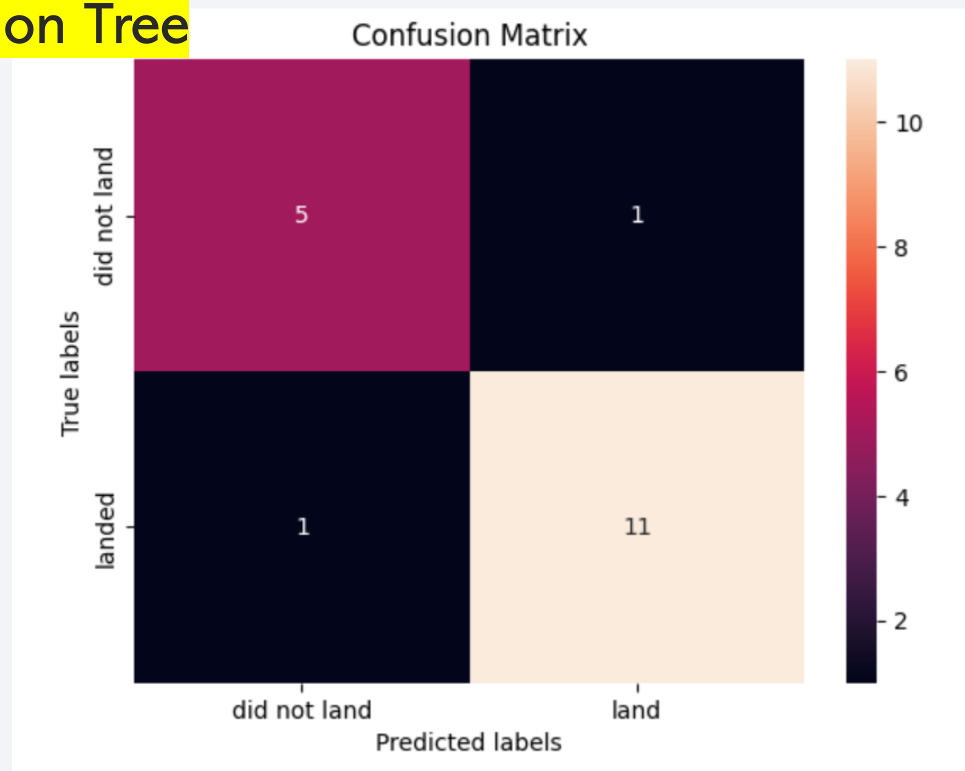
Classification Accuracy

Both Decision Tree and KNN achieved the same test accuracy of 0.8889.

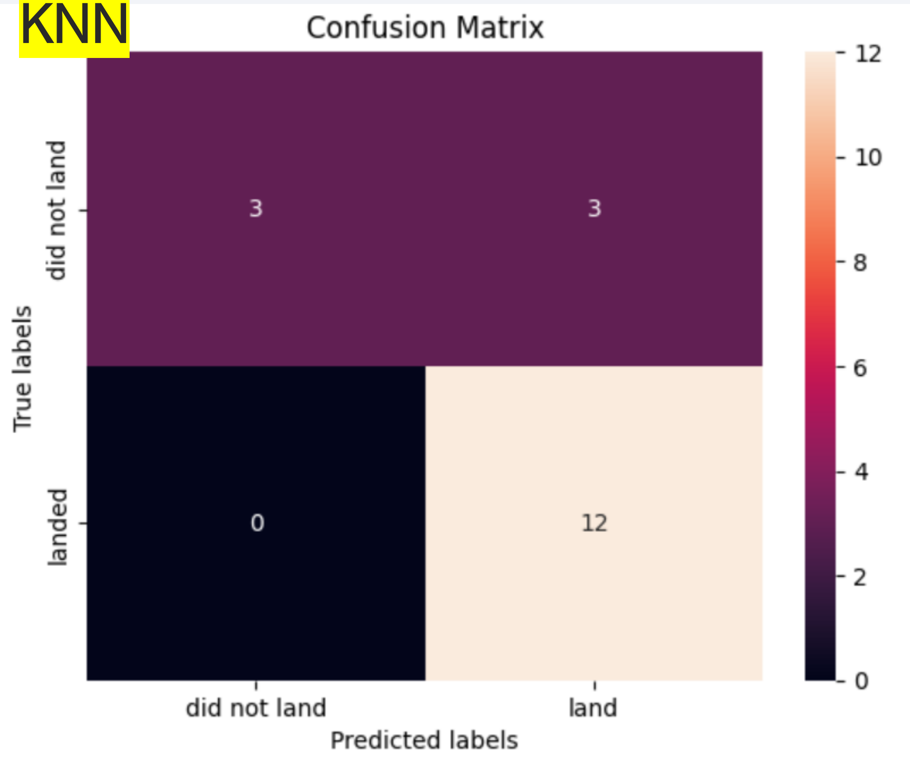


Confusion Matrix

Decision Tree



KNN



But based on the confusion matrix, the Decision Tree performs better because it correctly predicts more of both classes (landed and did not land) with fewer misclassifications.

Conclusions

After hyperparameter tuning and evaluation:

- **Logistic Regression** performed with test accuracy **0.8333** using parameters C=0.01, penalty=l2, solver=lbfgs.
- **SVM** achieved **0.8333** test accuracy with C=1.0, gamma=0.0316, kernel=sigmoid.
- **Decision Tree** reached **0.8889** test accuracy with criterion=entropy, max_depth=8, max_features=sqrt, min_samples_leaf=2, min_samples_split=2, splitter=random.
- **KNN** also achieved **0.8889** test accuracy with n_neighbors=10, algorithm=auto, p=1.

Although **Decision Tree** and **KNN** have the same test accuracy, the **confusion matrix** shows that the **Decision Tree** provides better class separation and fewer misclassifications.

Appendix

- Dataset:

  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- Full Source Code:

  GitHub - IBM Data Science Capstone Project

Thank you!