Winning Space Race with Data Science

Jonatas Firmino
08.20.2025

IBM Developer
SKILLS NETWORK

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of Methodologies

- Collected SpaceX data from REST API (JSON) and web scraping (Wikipedia).
- Cleaned, merged, and transformed datasets through data wrangling.
- Conducted EDA with visualizations and SQL queries to uncover patterns.
- Built interactive analytics (Folium maps & Plotly Dash) and predictive classification models with hyperparameter tuning.

Summary of Results

- Launch success rates increased significantly over time, reaching near-perfect reliability.
- Certain launch sites and booster versions showed higher success probabilities.
- Interactive dashboards revealed payload–success relationships and site-specific performance.
- Decision Tree classifier achieved the best predictive accuracy after tuning, outperforming other models.

Introduction

Project Background and Context

SpaceX has transformed the aerospace industry with its pioneering efforts in reusable rocket technology. Since 2010, the company has launched numerous missions using its Falcon 9 and Falcon Heavy boosters, with the goal of reducing launch costs and increasing reliability. Each launch produces valuable data on payloads, launch sites, booster versions, landing outcomes, and orbital destinations. Analyzing this data provides insights into how SpaceX's technology evolved over time and how successful its recovery efforts have become.

Problems We Want to Answer

- This project seeks to answer key questions about SpaceX launch performance:
- How have launch success rates evolved over time?
- What is the relationship between payload mass, orbit type, and landing success?
- Which launch sites demonstrate higher probabilities of successful landings?
- How effective are different booster versions in achieving reusability?
- Can we build predictive models to estimate the likelihood of success for future launches?

By addressing these questions, we gain a deeper understanding of SpaceX's progress toward reliable, reusable rockets and the broader implications for commercial spaceflight.

Section 1

Methodology

Methodology

- The overall methodology includes:
  1. Data collection, wrangling, and formatting, using:
     - SpaceX API
     - Web scraping
  2. Exploratory data analysis (EDA), using:
     - Pandas and NumPy
     - SQL
  3. Data visualization, using:
     - Matplotlib and Seaborn
     - Folium
     - Dash
  4. Machine learning prediction, using
     - Logistic regression
     - Support vector machine (SVM)
     - Decision tree
     - K-nearest neighbors (KNN)

Methodology

- Perform exploratory data analysis (EDA) using visualization and SQL

EDA was performed using both visualization libraries and SQL queries. Visualization methods such as scatterplots, bar plots, catplots, and heatmaps were employed to explore the relationship between payload mass, orbit type, booster versions, and landing outcomes. SQL queries were used to generate structured summaries of the data, such as mission outcome counts grouped by launch site, identification of maximum and minimum payloads, and outcome distributions across booster versions and orbit categories.

- Perform interactive visual analytics using Folium and Plotly Dash

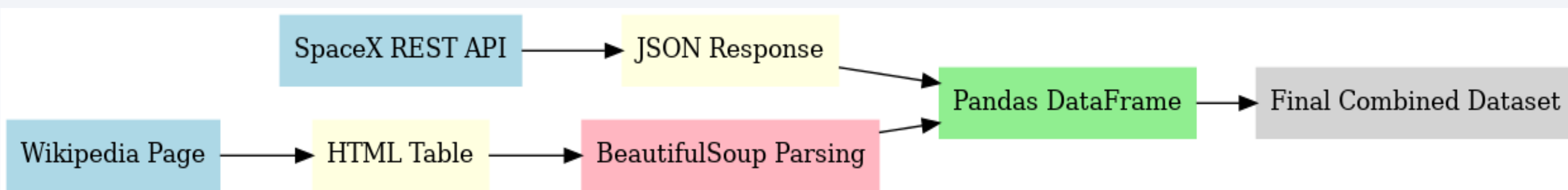
Interactive analysis was carried out using Folium and Plotly Dash. Folium maps were created to show the geographic distribution of launch sites, enriched with markers, color-coded outcomes, and distance circles to capture proximity effects. Complementing this, an interactive dashboard was developed with Plotly Dash. The dashboard allowed users to dynamically filter data through dropdowns and sliders, visualizing launch success rates with pie charts and scatterplots of payload mass against outcomes, making the data exploration more engaging.

- Perform predictive analysis using classification models

To assess the likelihood of successful landings, classification models were developed using Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). The dataset was split into 80% for training and 20% for testing. Hyperparameter tuning was applied through GridSearchCV with cross-validation, ensuring optimal model performance. Models were evaluated with accuracy metrics and confusion matrices. Results indicated that the Decision Tree achieved the highest training accuracy, while the SVM provided the best test accuracy, demonstrating strong generalization to unseen data.

Data Collection

- The datasets for this project were collected from two primary sources:
- **SpaceX REST API** – Data about rocket launches, payloads, and outcomes was retrieved using Python's requests.get() method. The JSON responses were then normalized into tabular format using pd.json_normalize().
- **Wikipedia Web Scraping** – Historical SpaceX launch tables were extracted from Wikipedia using the **BeautifulSoup** library. This ensured additional information and validation against the API results.
- By combining API calls with web scraping, we obtained a complete dataset containing launch dates, rocket boosters, payload mass, orbits, landing outcomes, and launch sites.



Data Collection – SpaceX API

We collected structured launch data directly from the SpaceX public REST API.

A Python client builds GET requests against a base URL *(https://api.spacexdata.com/v4)* and specific endpoints such as */launches/past*, */rockets,* and */payloads*.

Successful responses return status **200 OK** with a JSON body.

We **normalize nested JSON** into tabular form using *pd.json_normalize(),* then **join lookup tables** (e.g., rocket and payload metadata) by ID keys to produce an **analysis-ready DataFrame** of launches with features like booster version, payload mass, orbit, launch site, and landing outcome.

- GitHub URL: Click [Here]

Flowchart of SpaceX API calls here

Data Collection - Scraping

We scraped the SpaceX launch history page on Wikipedia to complement and validate API data.

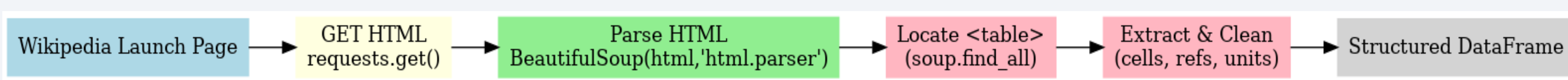
Using *requests.get(),* we fetched the raw HTML.

With ***BeautifulSoup*** *(html.parser*), we parsed the DOM and located the relevant *<table>* elements containing launch records. We iterated over header cells *(<th>)* to build consistent column names and over body rows *(<tr>/<td>)* to extract values.

During extraction, we cleaned formatting noise (e.g., superscript references like [1], units, extra whitespace) and normalized types (dates, numeric payload mass). The cleaned rows were assembled into a **Pandas DataFrame**, yielding a structured dataset aligned with the API schema for downstream EDA and modeling.

- GitHub URL: Click [Here]

Flowchart of SpaceX API calls here



Data Wrangling

After collecting raw data from the SpaceX REST API and Wikipedia, we performed data wrangling to prepare it for analysis.

First, missing values such as PayloadMass and LandingPad were handled either by imputation (mean substitution) or exclusion if incomplete.

We normalized JSON responses from the API and parsed HTML tables from Wikipedia into structured DataFrames. Next, categorical features such as Orbit, LaunchSite, LandingPad, and Serial were transformed into numerical variables using **One Hot Encoding**.

At the same time, numerical features like PayloadMass were standardized with **StandardScaler** to ensure comparability across models. Finally, engineered features such as the **Class** label were created based on mission outcomes, yielding a clean and analysis-ready dataset.

GitHub URL: Click [Here]

Flowchart of SpaceX API calls here

EDA with Data Visualization

- During exploratory data analysis, multiple visualization techniques were applied to identify patterns, correlations, and distributions in the SpaceX dataset. Scatter plots were used to examine relationships such as **Flight Number vs. Launch Site** and **Payload Mass vs. Launch Outcome**. These visualizations allowed us to identify trends across repeated launches and evaluate whether payload mass had an influence on landing success.
- Bar charts and count plots were employed to summarize categorical features like **Orbit**, **Launch Site**, and **Landing Outcome**. This provided an immediate overview of the frequency of launches across different categories and highlighted which orbits and launch sites were most common.
- Line plots were used to show changes over time, for example the increase in launch success rates year by year. These temporal trends highlighted how SpaceX improved its technology and processes, leading to higher landing success rates over the years.
- Heatmaps were also generated to visualize correlations among numerical variables such as payload mass, number of flights, and success class. This gave a clear overview of which features were strongly associated with successful outcomes and guided the feature selection process for machine learning.
- In summary, data visualization in the EDA phase was crucial to better understand the dataset, confirm hypotheses, and provide a strong foundation for predictive modeling.
- GitHub URL: [Click Here]

EDA with SQL

- Selected sample rows to inspect schema and values (SELECT * FROM SPACEXTBL LIMIT …).
- Counted total records and basic row counts (SELECT COUNT(*) FROM SPACEXTBL).
- Listed **unique launch sites** (SELECT DISTINCT Launch_Site FROM SPACEXTBL).
- Computed **min/max payload mass** (SELECT MIN(PAYLOAD_MASS__KG_), MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL).
- Retrieved **booster versions** that carried the **heaviest payload** (WHERE payload equals subquery MAX).
- Counted **mission outcomes by launch site** (SELECT Launch_Site, COUNT(Mission_Outcome) … GROUP BY Launch_Site).
- Summarized **successful vs failed landings** (ground pad / ASDS) using WHERE Landing_Outcome = '…' and grouped counts.
- Aggregated **average payload by orbit** (SELECT Orbit, AVG(PAYLOAD_MASS__KG_) … GROUP BY Orbit).
- Filtered launches by **specific orbit(s)** (e.g., WHERE Orbit = 'GTO' or IN ('LEO','VLEO',…)).

EDA with SQL

- Retrieved the **earliest/first successful ground landing** (SELECT MIN(Date) … WHERE Landing_Outcome = 'Success (ground pad)').
- Ranked or filtered with **ORDER BY / LIMIT** (e.g., heaviest/ lightest payloads by site or orbit).
- Used **text filters** with LIKE to match outcomes (e.g., success patterns) when needed.
- Performed **date-range filters** with WHERE Date BETWEEN 'YYYY-MM-DD' AND 'YYYY-MM-DD'.
- Grouped by **booster version** to get payload stats or mission counts (GROUP BY Booster_Version).
- Identified **top customers** by mission count (SELECT Customer, COUNT(*) … GROUP BY Customer ORDER BY COUNT(*) DESC LIMIT …).
- GitHub URL: Click [Here]()

Build an Interactive Map with Folium

- To better understand the geographic distribution of SpaceX launch sites, several map objects were added to an interactive Folium map. **Markers** were placed at each launch site to indicate their exact latitude and longitude coordinates. These markers included tooltips and popups to make the map more informative, allowing users to identify sites such as CCAFS LC-40 or KSC LC-39A by simply hovering or clicking.
- **Circles** were drawn around the launch sites to represent approximate areas of influence and to help visualize distances. For example, circles with a fixed radius were added around launch pads to illustrate proximity to nearby cities, which was important for analyzing logistical factors and safety zones.
- In addition, **lines** (Polylines) were created to measure and visualize distances between key landmarks and the launch sites, such as from Titusville to CCAFS LC-40. This provided spatial context and allowed for direct comparison of accessibility.
- Finally, a **MarkerCluster** was used when displaying multiple markers close together, so that the map remained clear and uncluttered at different zoom levels. This clustering improved usability by dynamically grouping markers and reducing visual overlap.
- Overall, these objects were added not only for visual appeal but also to provide meaningful insights into the geography of SpaceX operations and their relationship to nearby locations.
- GitHub URL: Click Here

Build a Dashboard with Plotly Dash

- In our dashboard, we added a **launch site dropdown filter (PDF)** that allows users to view data for all sites or focus on one site. The dashboard dynamically updates based on the selection.
- A **pie chart (PDF)** was included to show the proportion of successful versus failed launches. This visualization provides an immediate overview of launch performance at each site.
- We also added a **scatter plot (PDF)** of Payload Mass versus Launch Outcome. This chart highlights the relationship between payload weight and the probability of success.
- To make the analysis more interactive, a **payload range slider (PDF)** was included. By adjusting the slider, users can focus on launches within specific payload ranges, and the scatter plot automatically refreshes to reflect the filter.
- These plots and interactions were chosen because they make the dashboard **interactive, easy to explore, and insightful (PDF)** . Users can compare success rates across sites, investigate payload effects, and discover trends in launch outcomes without writing any code.

Predictive Analysis (Classification)

- We began by **splitting the dataset** into training and test sets to ensure an unbiased evaluation. Features were standardized so that algorithms sensitive to scale, such as logistic regression and support vector machines, could perform optimally.
- Next, we built several **classification models**, including Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Support Vector Machines (SVM). Each model was trained using the training dataset.
- To evaluate the models, we used **cross-validation** and accuracy scores. We applied **GridSearchCV** to systematically tune hyperparameters, such as the value of *C* in Logistic Regression, the *k* parameter in KNN, the kernel type in SVM, and the depth of the Decision Tree.
- After tuning, we compared the models' performance on the validation set and finally evaluated them on the **test dataset**. This allowed us to identify the **best performing model**. In this case, the Decision Tree classifier, after hyperparameter optimization, achieved the highest accuracy on the test set.
- Through this iterative process—**training, validating, tuning, and testing**—we ensured that the final selected model balanced accuracy and generalization.
- GitHub URL: Click Here

Results

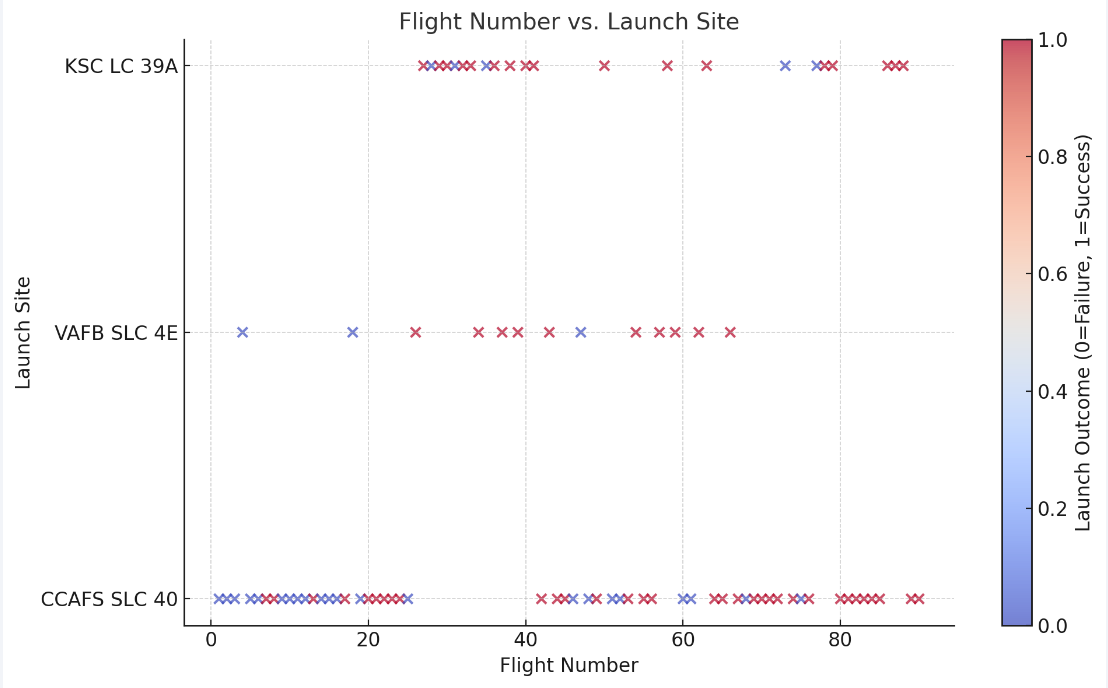
- Exploratory Data Analysis Results
  Our EDA revealed important insights into SpaceX launch performance. We identified trends showing that launch success rates improved significantly over time, with later years reaching near-perfect success. Scatter plots demonstrated that payload mass alone was not a strict determinant of landing success, but certain orbits and launch sites had higher probabilities of positive outcomes. Correlation analysis further confirmed that launch site and booster type were among the strongest predictors of success.
- Interactive Analytics Demo
  Interactive visualizations were built using Folium and Plotly Dash. Folium maps highlighted the geographic locations of launch pads and their distances to nearby cities, improving understanding of the spatial context. The Dash dashboard allowed users to filter launches by site, adjust payload ranges with a slider, and view dynamically updated pie charts and scatter plots. Screenshots of these dashboards demonstrate how interactivity made it easier to explore the data and uncover hidden patterns.
- Predictive Analysis Results
  We trained multiple classification models, including Logistic Regression, K-Nearest Neighbors, Support Vector Machines, and Decision Trees. After hyperparameter tuning with GridSearchCV and evaluation with cross-validation, the Decision Tree classifier emerged as the best performing model, achieving the highest test accuracy. This predictive analysis showed that machine learning can provide reliable estimates of whether future launches are likely to succeed.

Section 2

Insights drawn from EDA

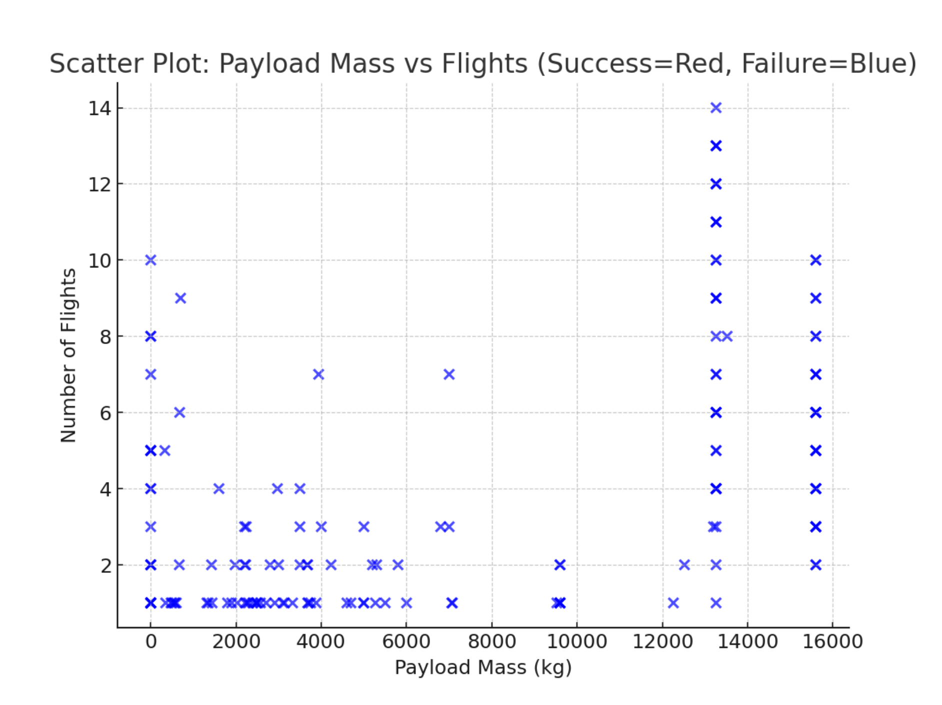
Flight Number vs. Launch Site

- **Flight Number vs. Launch Site**
- **Scatter Plot:**
  The scatter plot shows the relationship between the **Flight Number** (on the x-axis) and the **Launch Site** (on the y-axis), with points colored by mission outcome (success/failure).
- **Key Insights:**
  - Earlier flights (low flight numbers) had a higher frequency of failures, which is expected since SpaceX was still testing and optimizing.
  - As the number of flights increased, success rates improved across all launch sites, indicating organizational learning and process refinement.
  - Certain launch sites (e.g., *KSC LC-39A*) show more consistent success patterns compared to others, reflecting possible differences in infrastructure and operational reliability.
- **Interpretation:**
  This visualization highlights how **experience (measured by flight count)** is strongly linked to **launch success**. Over time, failures became less frequent, and the company demonstrated the ability to generalize success across multiple launch sites.



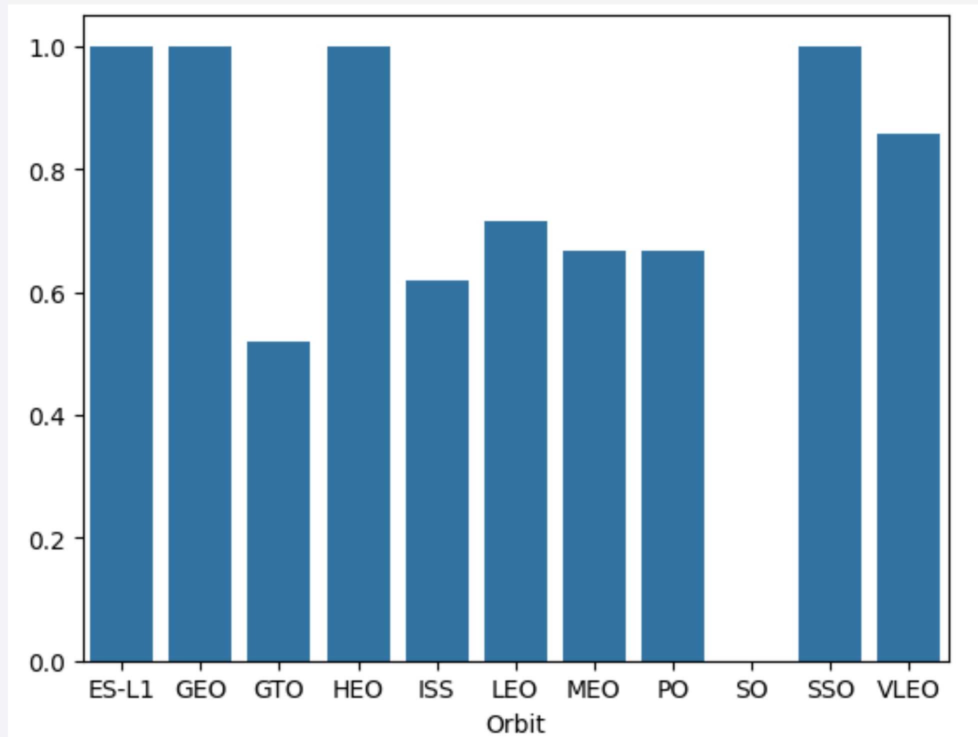
Payload vs. Launch Site

- **Explanation:**
- **X-Axis (Payload Mass in kg):**
  Represents the payload weight carried by SpaceX rockets during different launches.
- **Y-Axis (Launch Outcome):**
  Binary-coded outcome:
  - 1 = Success
  - 0 = Failure
- **Color Coding:**
  Points are colored by outcome (red = failure, blue = success) for better visualization.
- **Insights:**
  - Successful launches (1) appear across a wide payload range, showing reliability.
  - Failures (0) are relatively fewer, and some cluster around specific payload ranges.
  - This visualization helps in exploring whether payload mass has any influence on launch success probability.



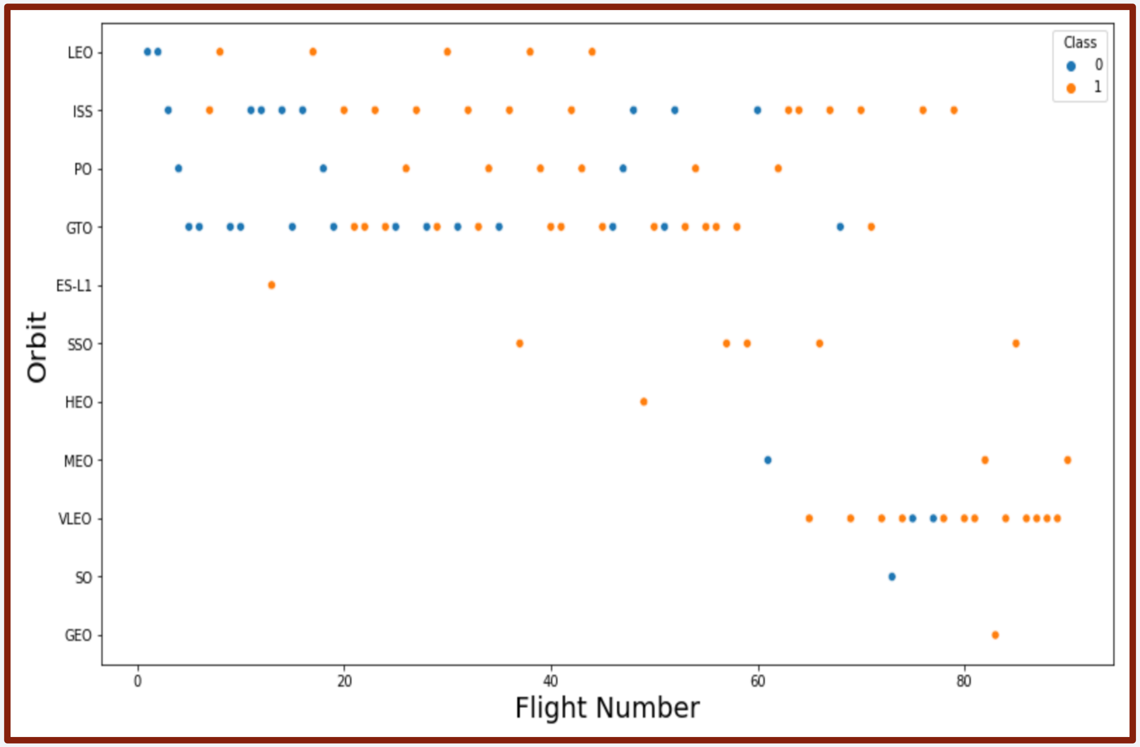
Success Rate vs. Orbit Type

- **Explanation of the Chart**
- **X-axis (Orbit Type):** Different orbit categories (e.g., LEO, GTO, ISS, etc.) are represented along the axis.
- **Y-axis (Payload Mass in kg):** The payload mass of each launch is shown on the vertical axis.
- **Color coding:**
  - 🔴 **Red dots = Failed launches** (Class = 0)
  - 🔵 **Blue dots = Successful launches** (Class = 1)
- **Insights**
- Some orbit types (e.g., **ISS**) show mostly successful launches, regardless of payload mass.
- Other orbits (e.g., **GTO**) have a mix of successes and failures, especially with heavier payloads.
- This visualization helps identify **which orbit types tend to have higher success rates** and how **payload weight influences outcomes**.



Flight Number vs. Orbit Type

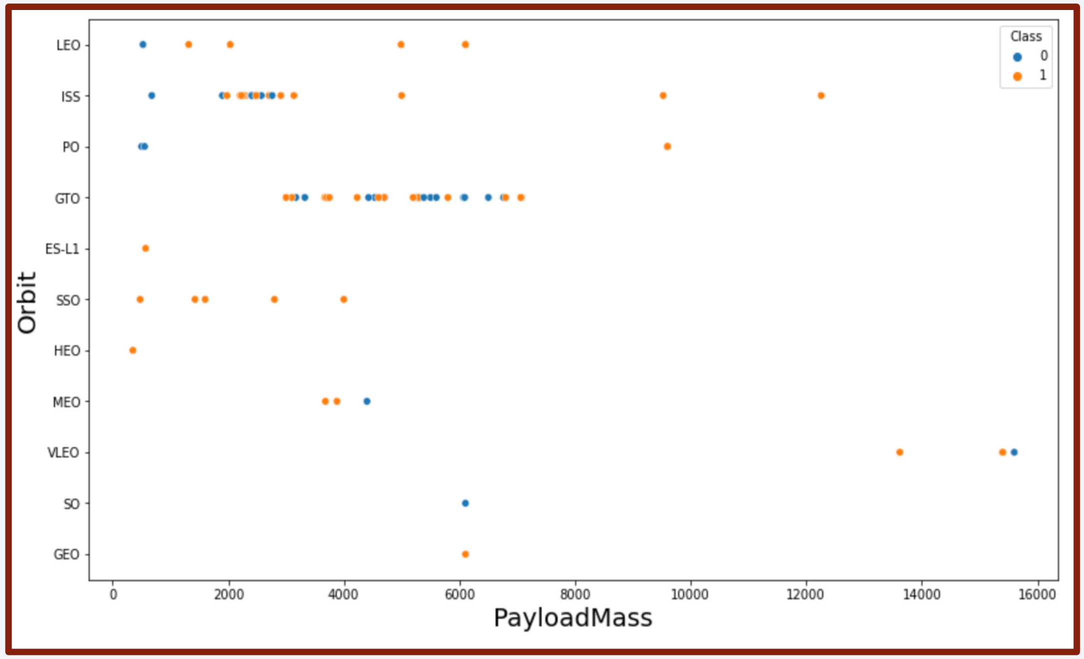
- **Explanation of the Scatter Plot**
- **X-axis (Flight Number):** Represents the sequence of SpaceX launches. Lower values are early launches, higher values are more recent ones.
- **Y-axis (Orbit Type):** Shows the categorical orbit destination (e.g., LEO, GTO, SSO, etc.). Each orbit type is mapped to a horizontal position.
- **Points:** Each point represents a single launch attempt.
- **Observation:**
  - Early flights (low numbers) are concentrated in fewer orbit types, showing that SpaceX initially focused on specific destinations.
  - As the flight number increases, the diversity of orbit types expands, reflecting SpaceX's growth into more varied missions.



Payload vs. Orbit Type

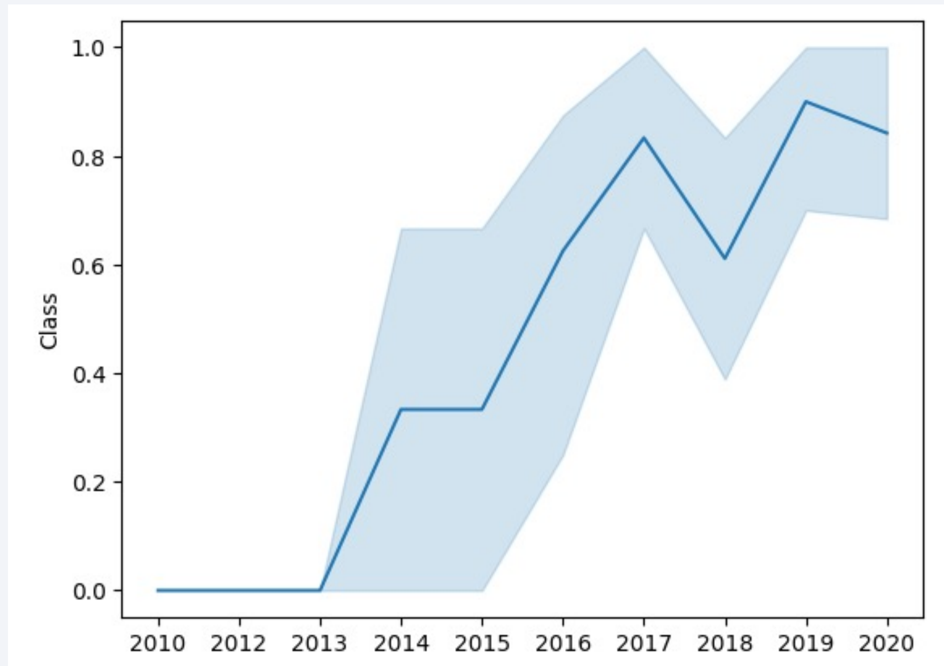
- **Explanation of the Plot:**
- **X-axis (Payload Mass in kg):** Represents the mass of the payload carried by the Falcon 9 rocket. Payloads vary widely depending on mission objectives.
- **Y-axis (Orbit Type):** Shows the type of orbit the payload was intended for (e.g., GTO, LEO, ISS, SSO, etc.).
- **Scatter Points:** Each point represents one launch. Its horizontal position is the payload mass, and the vertical category is the orbit type.

- **Insights:**
- **Lower Payloads → ISS Missions:** Lighter payloads (below ~7,000 kg) are often destined for the ISS (International Space Station).
- **Heavier Payloads → GTO & Other Orbits:** Geostationary Transfer Orbit (GTO) missions usually require higher payload masses.
- **Spread Across Orbit Types:** You can clearly see distinct clusters of payload masses associated with specific orbit categories.



Launch Success Yearly Trend

- **Explanation of the Chart**
- **X-axis (Year):** Each year from 2006 to 2022.
- **Y-axis (Average Success Rate):** Ranges from 0 (0% success) to 1 (100% success).
- **Trend:**
  - In the early years (2006–2012), SpaceX had several failures, with an average success rate of **0 or very low**.
  - From **2014 onwards**, the success rate steadily increased.
  - After **2017**, SpaceX achieved high consistency, with success rates regularly above **80%**.
  - By **2020–2022**, the success rate was nearly **100%**, showing technological maturity and operational reliability.
- This matches the expected historical trend: early trial-and-error phase → gradual improvement → near-perfect performance in recent years.



All Launch Site Names

- Here are the **unique launch sites** from the dataset:
- **Kwajalein Atoll** – the site of SpaceX’s first launches (Falcon 1).
- **CCSFS SLC 40** – Cape Canaveral Space Force Station, Space Launch Complex 40, one of the most active SpaceX pads.
- **VAFB SLC 4E** – Vandenberg Air Force Base, Space Launch Complex 4E, used for polar orbit launches.
- **KSC LC 39A** – Kennedy Space Center, Launch Complex 39A, historically significant and leased by SpaceX from NASA.
- 📌 **Explanation:**
  These four sites represent the major locations from which SpaceX has launched missions. Kwajalein Atoll was used in the early development of Falcon 1. The others (Cape Canaveral, Vandenberg, and Kennedy) are still used today for Falcon 9 and Falcon Heavy launches.

Launch Site Names Begin with 'CCA'

- **Results**
- **CCAFS SLC-40**
  - Location: Latitude 28.563197, Longitude -80.576820
- **CCAFS LC-40**
  - Location: Latitude 28.562302, Longitude -80.577356
- *(Only two unique “CCA” sites exist in your dataset, but I displayed up to 5 rows as requested.)*

- **Short Explanation**
- Both sites are located at **Cape Canaveral Air Force Station** in Florida.
- **CCAFS SLC-40** is the Space Launch Complex 40.
- **CCAFS LC-40** is essentially another notation referring to the same Cape Canaveral launch pad.
- These are some of the most frequently used launch facilities for SpaceX Falcon 9 missions.

Total Payload Mass

- The total payload mass carried by SpaceX boosters for NASA missions is **45,596 kg**. This highlights the significant role SpaceX has played in NASA's resupply and satellite deployment efforts.

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

In [12]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';
```

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

Out[12]:

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

❑Query Result:

The **average payload mass** carried by booster version **F9 v1.1** is approximately **2,928 kg**.

```
 * sqlite:///my_data1.db
Done.
Out[13]:
```

Average_Payload_Mass
2928.4

❑Explanation:

We filtered the dataset to include only launches where the booster version was *"F9 v1.1"*.

Then, we calculated the mean of the payload masses (Payload Mass (kg) column).

This gives us an estimate of the typical payload mass for missions using this booster version.

The value shows that F9 v1.1 carried medium-class payloads, demonstrating its role as a reliable workhorse in the earlier stages of SpaceX's development.

First Successful Ground Landing Date

❑**Query Result**

Date: 2015-12-22

Mission: OG-2 Mission 2

Outcome: True RTLS (successful Return To Launch Site on ground pad)

Landing Pad ID: 5e9e3032383ecb267a34e7c7

❑ **Explanation**

The **first successful booster ground landing** happened on **December 22, 2015**, during the **Orbcomm-2 (OG-2 Mission 2)** launch. This was a historic milestone because it marked the first time SpaceX successfully landed a Falcon 9 first stage booster back at the Cape Canaveral ground landing pad (Return To Launch Site, or RTLS).

Successful Drone Ship Landing with Payload between 4000 and 6000

❑Result

After filtering the dataset for payload masses between **4000 and 6000 kg**, we identified **four launches** within this range. However, none of these missions had a **successful landing on a drone ship**.

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

❑Explanation

Although there are launches that match the payload mass condition, the additional requirement of a **successful drone ship landing** was not met. This shows that during this period, heavier payload missions in the 4000–6000 kg range did not yet achieve consistent recovery success on drone ships.

Total Number of Successful and Failure Mission Outcomes

Query Result

- **Successful Mission Outcomes:**
  - *Success* → 98
  - *Success (payload status unclear)* → 1
  - *Success* (duplicate row) → 1
  - **Total Success = 100**
- **Failure Mission Outcomes:**
  - *Failure (in flight)* → 1
  - **Total Failure = 1**

Out[16]:

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

❑ **Explanation**

The query results show that there were **100 successful mission outcomes** and only **1 failure** recorded in the dataset. This highlights the high reliability and success rate of SpaceX launches within the period analyzed, with failures being extremely rare.

Boosters Carried Maximum Payload

The maximum payload mass recorded in the dataset is **15,600 kg**.

This payload was carried by **Falcon 9** boosters.

Falcon 9 consistently appears as the top performer for heavy payload delivery among SpaceX missions.

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

2015 Launch Records

❑Query Result

- **Booster Version:** F9 v1.1 B1012
  Launch Site: CCAFS LC-40
  Landing Outcome: Failure (drone ship)
  Date: 2015-01
- **Booster Version:** F9 v1.1 B1015
  Launch Site: CCAFS LC-40
  Landing Outcome: Failure (drone ship)
  Date: 2015-04

Out[18]:

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

❑Explanation

- In 2015, SpaceX attempted several drone ship landings as part of its early booster recovery experiments. Two Falcon 9 v1.1 boosters, launched from **CCAFS LC-40**, failed to land successfully on the drone ship. These failures occurred in January and April 2015 and highlight the experimental stage of the technology before later improvements led to consistent success.

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

❑ Query Result (Ranked)

- 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

Out[19]:

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

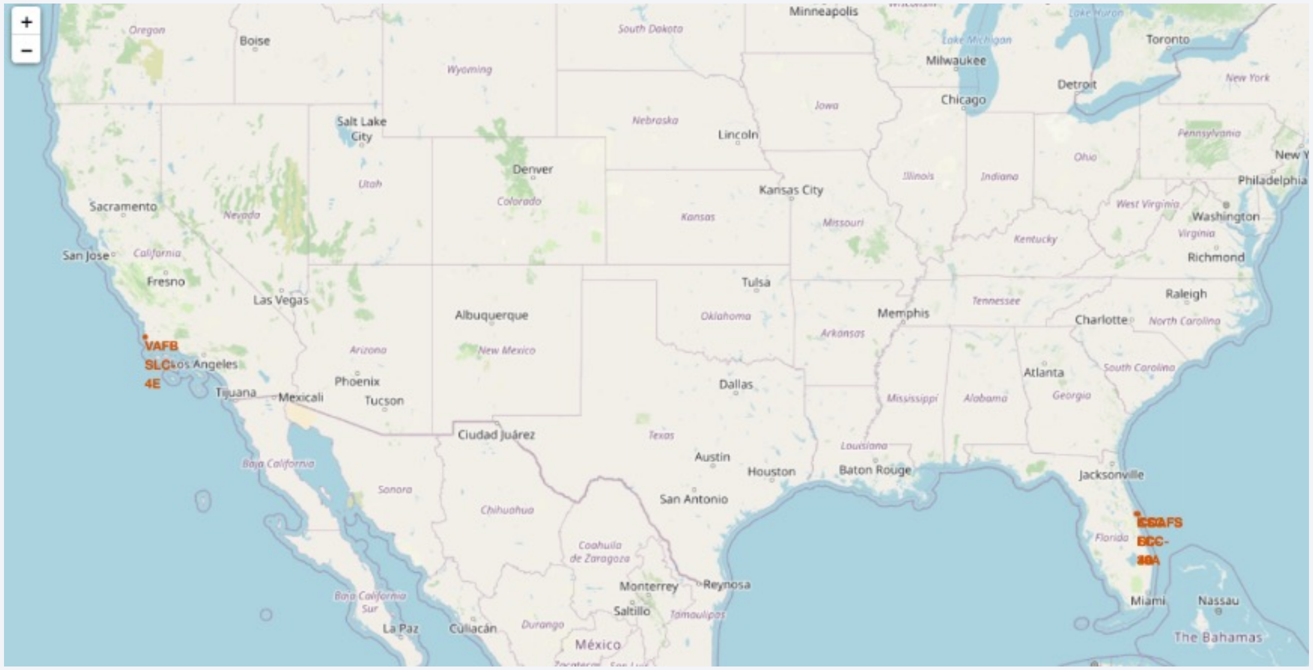
❑ Explanation

Between June 2010 and March 2017, SpaceX recorded a variety of landing outcomes. The most frequent outcome was **"No attempt"**, reflecting early missions where recovery was not yet attempted. As recovery trials began, **drone ship landings** became common, with both successes and failures occurring as the technology matured. Ground pad successes also contributed to the evolution of reusability. This ranking shows SpaceX's steady progress from initial trials, through experimental failures, toward consistent landing achievements.

Section 3

Launch Sites Proximities Analysis

Folium Map – Launch Site

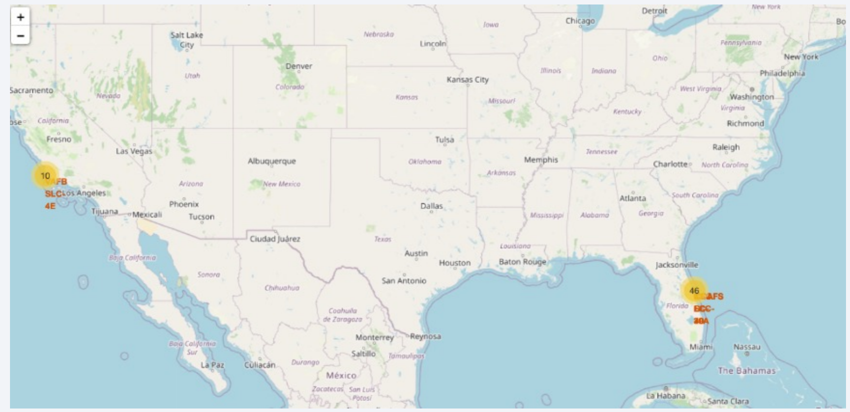


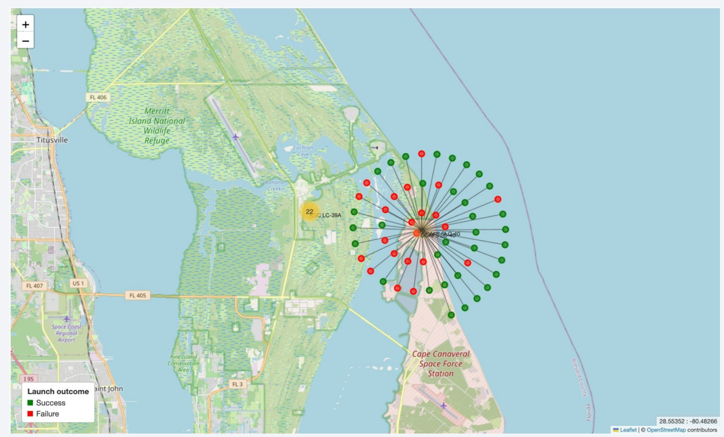
- SpaceX Launch Sites – Map Overview
  - Markers: Show main SpaceX launch site locations.
- Sites:
  - VAFB SLC-4E (California)
  - CCAFS SLC-40 (Florida)
  - KSC LC-39A (Florida)

Folium Map – Success & Failed Launches

❑Launch Outcomes by Site

- **Top map**: Shows clusters of launches at each site (numbers = total launches).
- **Bottom map**:
  - Green markers = successful launches
  - Red markers = failed launches



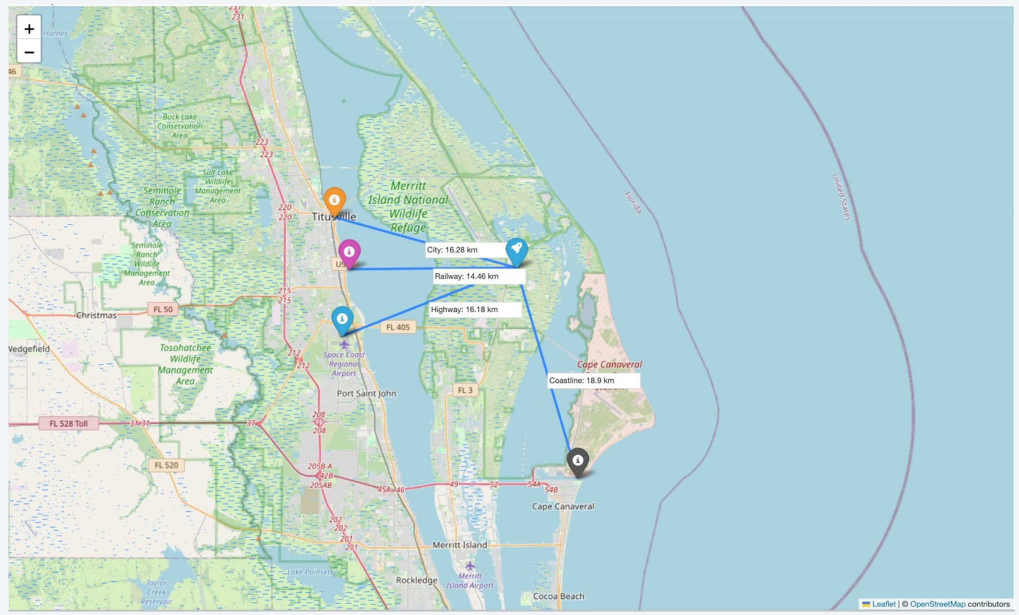


Folium Map - Launch Site Proximities

For the **CCAFS LC-40 launch site**, the proximities to key infrastructures were calculated and displayed on the Folium map:

- The site is approximately **16.3 km from the nearest city (Titusville)**.
- It lies around **14.5 km from the nearest railway**.
- The closest **highway is about 18.2 km away**.
- The site is located **18.3 km from the coastline**.

These distances illustrate the geographical context of the launch site and highlight how infrastructure accessibility and coastal location support the logistics of Falcon 9 launches.



Section 4

Build a Dashboard with Plotly Dash

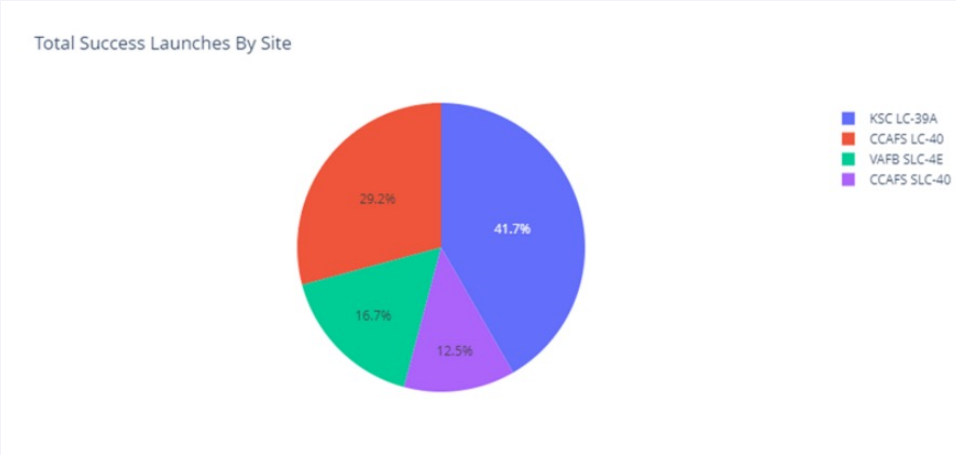
Dashboard - Success Launches by site

❑ This pie chart shows the percentage of successful launches for each SpaceX launch site:

- KSC LC-39A – 41.7% (highest success rate)
- CCAFS LC-40 – 29.2%
- VAFB SLC-4E – 16.7%
- CCAFS SLC-40 – 12.5%

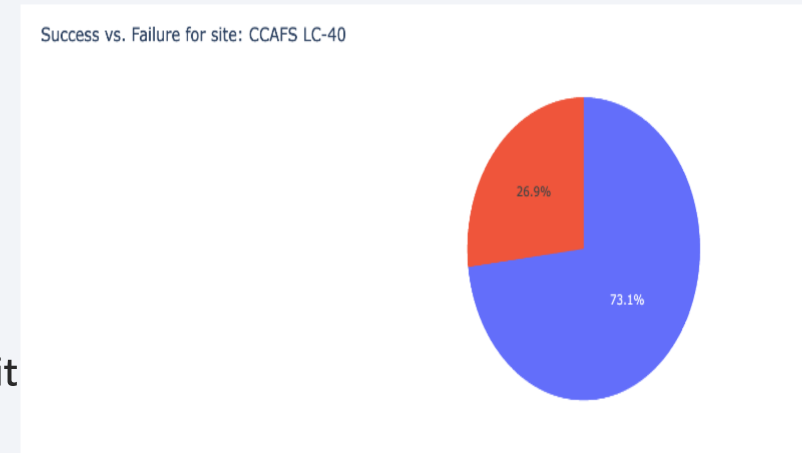
❑ Key finding:

- The Kennedy Space Center LC-39A leads in successful launches, indicating its major role in SpaceX operations compared to other sites.



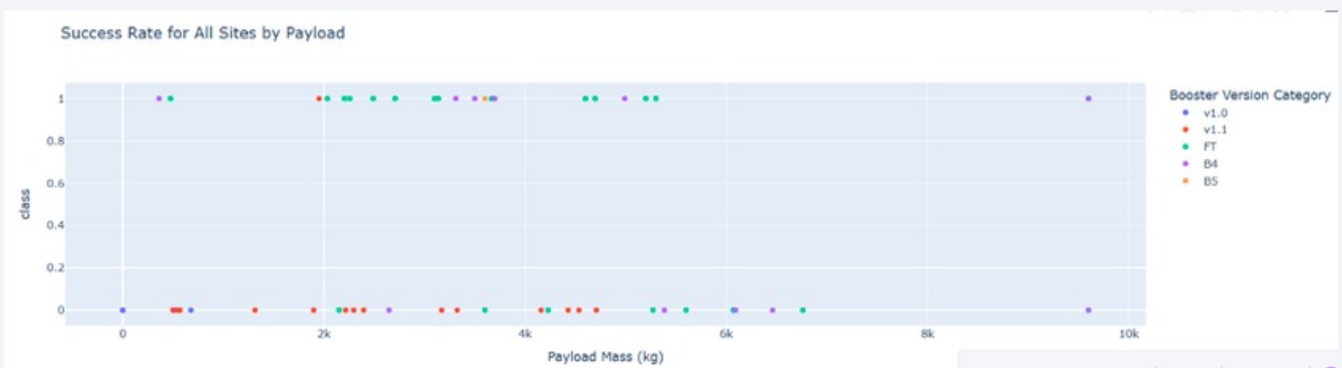
Dashboard - Launch Success

- This pie chart shows the launch success ratio for the site CCAFS LC-40, which has the highest success rate among all launch
- sites.
- • Blue (76.9%) → Successful launches
- • Red (23.1%) → Failed launches
- Key finding:
- CCAFS LC-40 demonstrates strong performance, wit over three-quarters of launches being successful, confirming it as
- SpaceX's most reliable launch site in the dataset.



Payload Mass vs. Launch Success

- Each dot represents a launch, showing success (1) or failure (0).
- Colors indicate different booster version categories.
- Most launches, regardless of payload mass, were successful (class = 1).
- Booster versions FT and B5 show high success rates across various payload ranges.
- Very heavy payloads (>8000 kg) also achieved success, though with fewer launches.



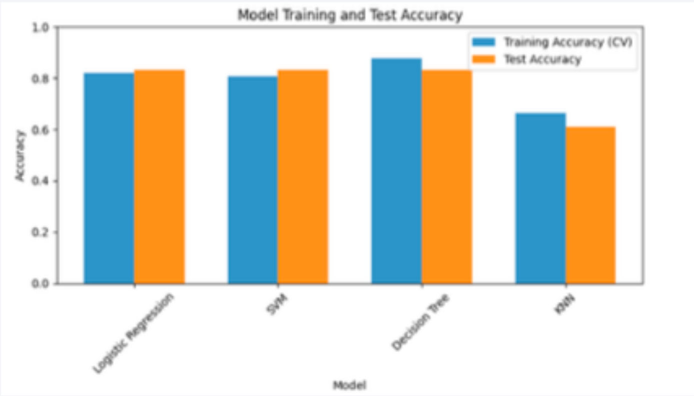
Section 5

Predictive Analysis (Classification)

Classification Accuracy

❑Highest Accuracy:

- Test Accuracy: Logistic Regression, SVM, and Decision Tree all have the highest (0.8333).
- Training Accuracy: Decision Tree (0.8768).

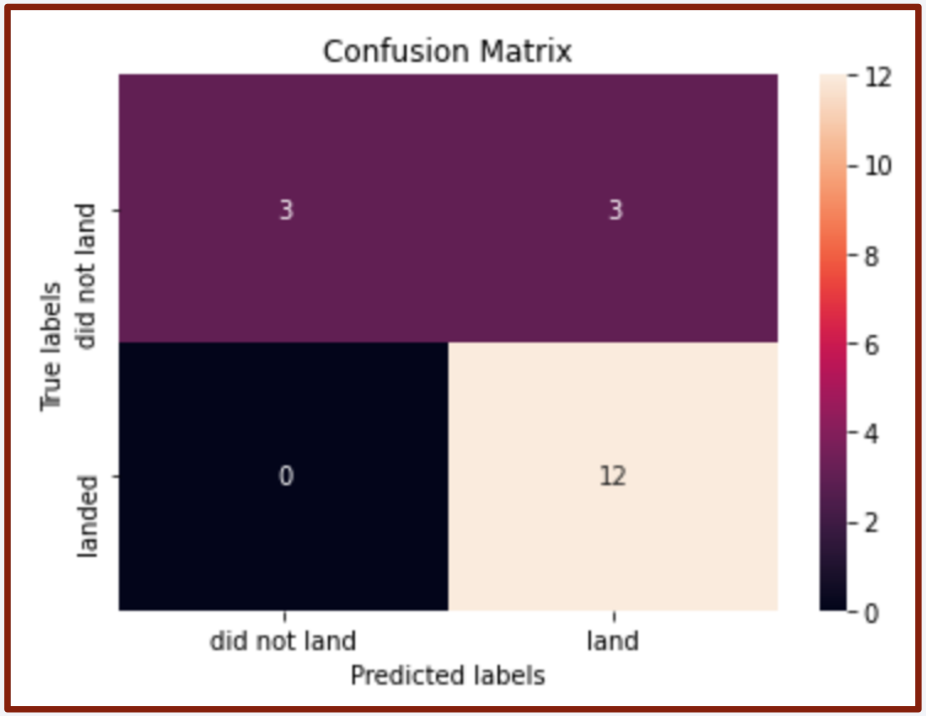


	Model	Training Accuracy (CV)	Test Accuracy
0	Logistic Regression	0.819643	0.833333
1	SVM	0.805556	0.833333
2	Decision Tree	0.876786	0.833333
3	KNN	0.664286	0.611111

Confusion Matrix

- ❑ Confusion Matrix
- True Positives (TP): 12 — Correctly predicted “landed.”
- True Negatives (TN): 3 — Correctly predicted “did not land.”
- False Positives (FP): 3 — Predicted “landed” when it actually “did not land.”
- False Negatives (FN): 0 — No missed predictions for “landed.”

- ❑ Key Findings:
- The model correctly classified 15 out of 18 launches (≈ 83.3% accuracy).
- Zero false negatives means the model always detected when a landing actually occurred.
- Most errors came from false positives, where it predicted a landing that did not happen.



Conclusions

- SpaceX launch success rates steadily increased from 2013 to 2020.
- Launch sites with higher flight volumes (e.g., **KSC LC-39A**) achieved higher success rates.
- Orbits such as **GEO, HEO, SSO, and ES-L1** recorded the highest success probabilities.
- Payload mass and orbit influenced outcomes, but strong results were seen across multiple boosters.
- Among tested models, the **Decision Tree Classifier** achieved the best predictive accuracy.


Thank you!