IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Erivelton Pinheiro de Menezes
Date: November 5, 2025.

Outline

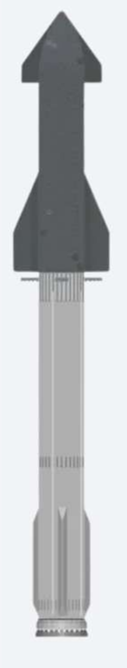
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Summary of methodologies

This project aimed to predict the successful landing of SpaceX's Falcon 9 first stage using a comprehensive data science pipeline. The methodology encompassed the following steps:

This project follows these steps:

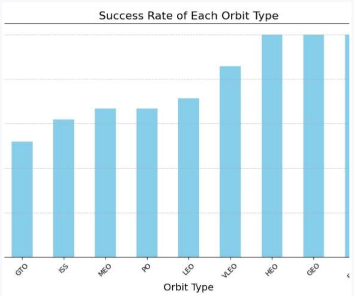


- Data Collection
- Data Wrangling Data Preparation & Exploration
- Exploratory Data Analysis
- Interactive Visual Analytics
- Predictive Analysis (Classification)

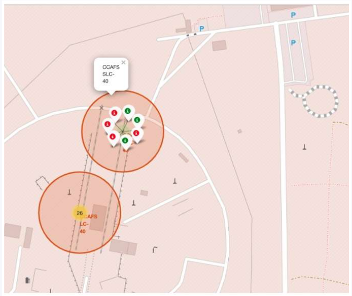
Executive Summary

Summary of Results

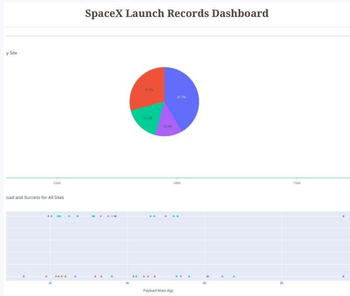
This project produced the following outputs and visualizations:



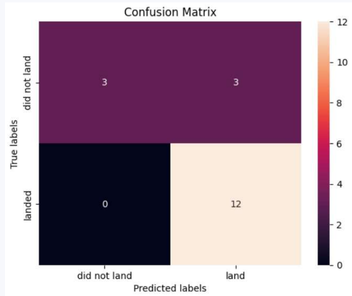
Exploratory Data Analysis (EDA) results



Geospatial analytics



Interactive dashboard



Predictive analysis of classification models

Introduction

Project background and context

This capstone project focuses on predicting the successful landing of SpaceX's Falcon 9 first stage, a critical factor in the company's ability to offer low-cost orbital launches. SpaceX's competitive advantage lies in its reusable rocket technology, which significantly reduces operational costs. By developing accurate predictive models for landing outcomes, this study aims to support cost estimation and strategic decision-making for stakeholders and competitors in the aerospace sector.

Problems you want to find answers

To guide the analysis, the following key questions were defined:

- Which factors most influence the success of Falcon 9 first stage landings?
- How effectively can machine learning models predict landing outcomes based on launch data?
- Among the tested models, which one demonstrates the highest predictive performance?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Require the data from SpaceX API
  - Collect data from a Wikipedia page
- Perform data wrangling
  - Perform EDA to find some patterns
  - Determine what would be the label for training supervised model
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Compare logistic regression model, support vector machine tree decision classifier, KNN by using GridSearchCV to select the best fit model

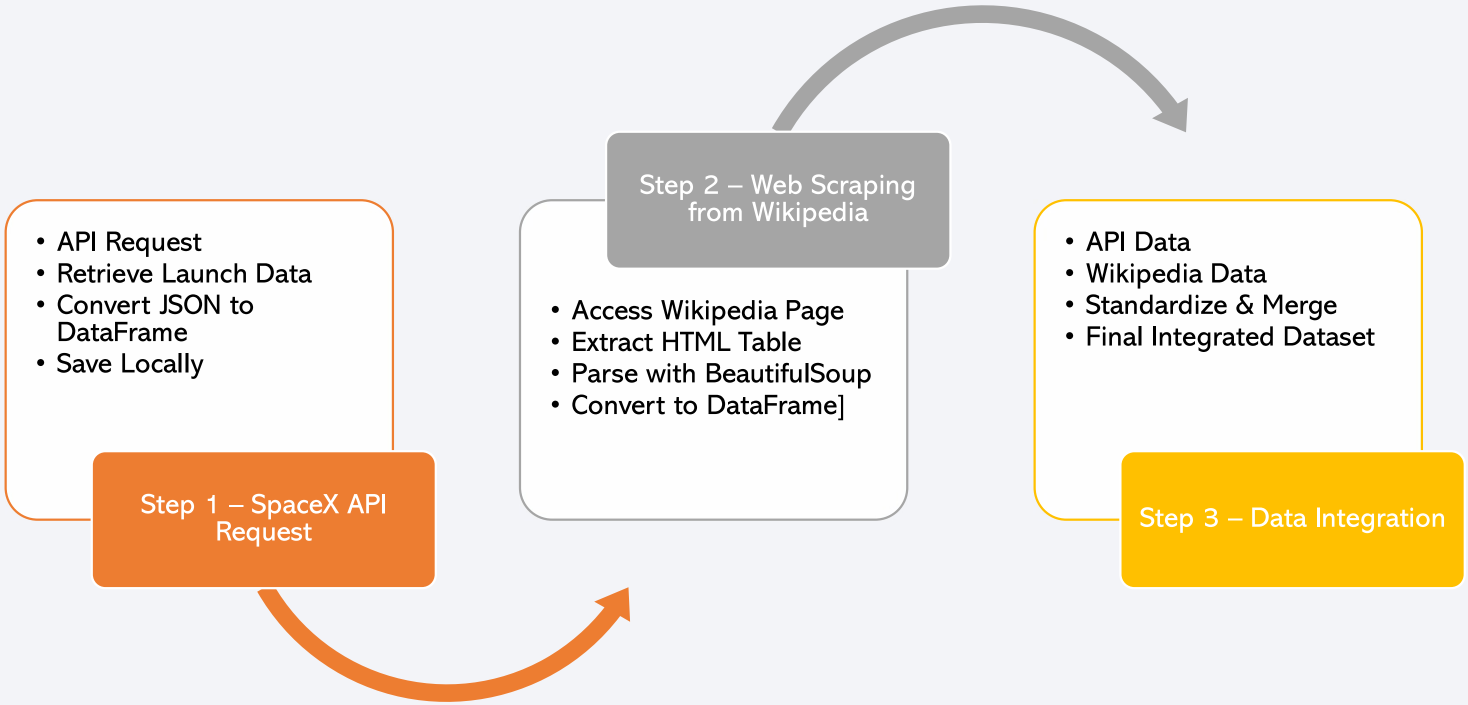
Data Collection

Describe how data sets were collected

Step 1 – SpaceX API Request: Launch data was retrieved via API, converted from JSON to a structured DataFrame, and stored locally.

Step 2 – Web Scraping from Wikipedia: HTML tables listing Falcon 9 missions were extracted and parsed using BeautifulSoup, then transformed into a DataFrame.

Step 3 – Data Integration: Both datasets were standardized and merged to create a unified, enriched dataset for analysis.



Data Collection – SpaceX API

Step 1 – SpaceX API Request

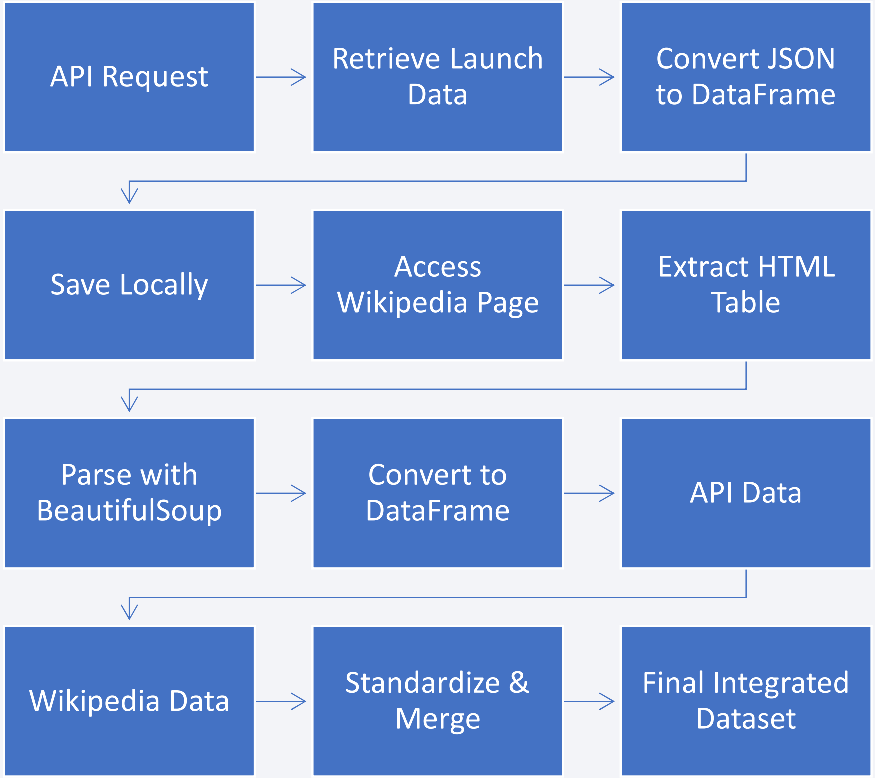
- Initiate REST API call to SpaceX endpoint
- Retrieve launch data in JSON format
  - Convert JSON to structured DataFrame
  - Save data locally for further processing

Step 2 – Web Scraping from Wikipedia

- Access Falcon 9 launch records page on Wikipedia
- Extract HTML table containing mission data
- Parse content using BeautifulSoup
- Convert parsed data into a DataFrame

Step 3 – Data Integration

- Combine API and Wikipedia datasets
- Standardize column formats and naming conventions
- Merge datasets into a unified structure
- Final integrated dataset ready for analysis
- **GitHub URL:** [SpaceX API Data Collection Notebook]()



Data Collection - Scraping

Overview

To complement the SpaceX API data, launch records were extracted from Wikipedia using Python-based web scraping techniques. The process involved accessing structured HTML tables, parsing the content, and converting it into a usable DataFrame for integration.

Access Falcon 9 launch history page on Wikipedia

- Locate and extract the HTML table containing mission data
- Parse the HTML content using **`BeautifulSoup`**
- Convert parsed data into a structured **`pandas`** DataFrame
- Clean and format the data for integration with API results

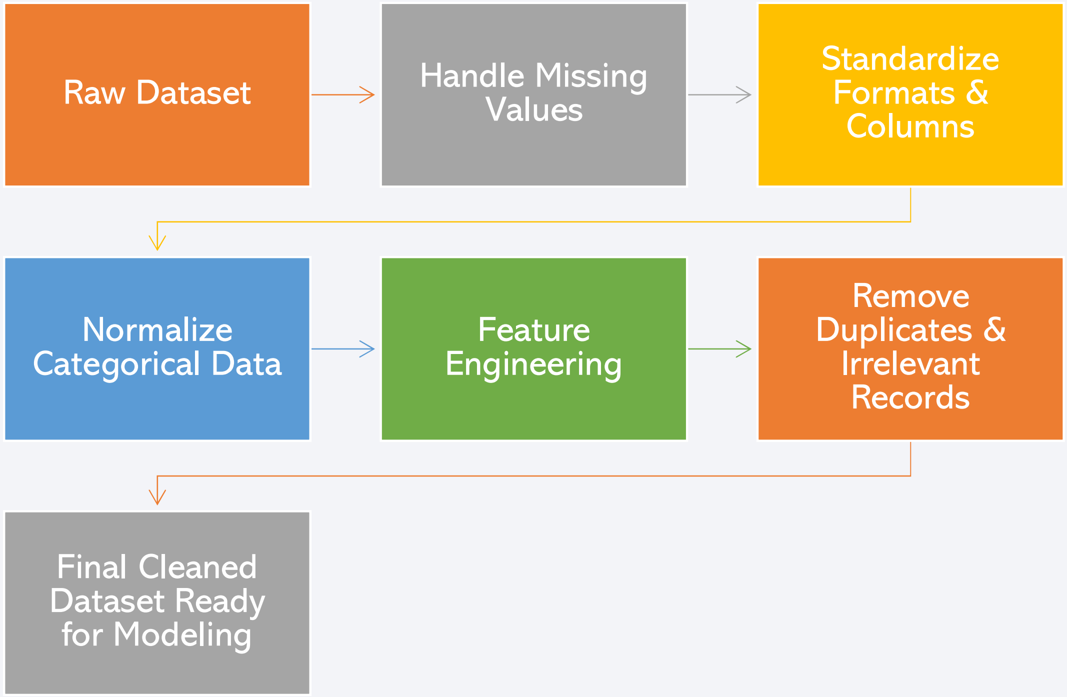
1. Access Wikipedia Page
2. Extract HTML Table
3. Parse with BeautifulSoup
4. Convert to DataFrame
5. Clean & Format Data

GitHub URL: Web Scraping Notebook – Falcon 9 Launch Records

Data Wrangling

The raw data collected from the SpaceX API and Wikipedia was refined through a structured wrangling process to ensure consistency, completeness, and analytical readiness. This involved handling missing values, standardizing formats, engineering new features, and preparing the dataset for exploratory analysis and machine learning.

- Identify and handle missing or null values
- Standardize column names and data formats (e.g., dates, strings, numerics)
- Normalize categorical values (e.g., launch site names, orbit types)
- Engineer new features such as:
  - `Class` (binary indicator of landing success)
  - `Payload Mass Category`
  - `Orbit Type Encoding`
- Remove irrelevant or duplicate records
- Store the final cleaned dataset in a structured pandas DataFrame



GitHub URL: [Data Wrangling Notebook – Falcon 9 Launch Dataset]

EDA with Data Visualization

In this exploratory data analysis (EDA) phase, several visualizations were developed to identify key patterns and relationships influencing the success of Falcon 9 first stage landings.

The following charts were created and used for specific analytical purposes:

- **Scatter Plots:** to examine the relationship between payload mass and landing success, highlighting potential correlations and trends;
- **Bar Charts:** to compare the number of launches and success rates across different launch sites, revealing variations in site performance;
- **Line Charts:** to visualize trends and changes in landing outcomes across parameters such as payload mass and launch site.

These visualizations helped reveal underlying data structures and supported feature selection for predictive modeling.

GitHub URL: EDA with Data Visualization – Falcon 9 Launch Analysis

EDA with SQL

- **Selection of specific columns:** Retrieved relevant fields such as Launch_Site, Payload_Mass, Orbit, Launch_Outcome, among others.
- **Launch count per site:** Used GROUP BY and COUNT(*) to identify the most frequently used launch sites.
- **Average calculations:** Applied AVG() to compute the average payload mass by orbit type.
- **Filtering by specific conditions:** Used WHERE clauses to narrow down data based on criteria like successful launches or specific orbits.
- **Table joins:** Performed JOIN operations to combine mission data with landing outcomes.
- **Additional aggregate functions:** Used MAX(), MIN(), SUM() to explore extremes and totals in variables like payload mass and launch counts.
- **Result ordering:** Applied ORDER BY to sort data by mass, date, or mission success.
- **Temporal data extraction:** Queried launch dates to support time-based analysis.
- **Landing success rate:** Calculated the proportion of successful landings based on recorded outcomes.

GitHub URL: EDA with SQL Notebook

Build an Interactive Map with Folium

Summary of Map Objects Added to the Folium Map

- **Markers:** Added `folium.Marker()` objects to indicate the exact geographic coordinates of each Falcon 9 launch site.
- **Circle Markers:** Used `folium.Circle()` to highlight the area surrounding each launch site, providing a visual radius of influence.
- **Lines (Polylines):** `Created folium.PolyLine()` objects to represent the distance between launch sites and nearby cities or reference points.
- **Popups and Tooltips:** Included interactive popup and tooltip elements to display site names and additional metadata when hovering or clicking on markers.

Explanation of Why These Objects Were Added

- Markers were added to pinpoint the exact location of each launch site, making it easy to identify them on the map.
- Circle markers help visualize the spatial distribution and proximity of each site, enhancing geographic context.
- Lines (Polylines) were used to calculate and illustrate distances between launch sites and other relevant locations, supporting spatial analysis.
- Popups and tooltips improve interactivity and user experience by providing contextual information directly on the map.

These elements together transform the map into a dynamic analytical tool, allowing users to explore launch site locations and their spatial relationships in an intuitive and engaging way.

GitHub URL: [Launch Sites Locations Analysis with Folium]

Build a Dashboard with Plotly Dash

Summary of Plots/Graphs and Interactions Added to the Dashboard

- Pie Chart (success-pie-chart): Displays either the total number of successful launches per site (when "All Sites" is selected) or the success vs. failure ratio for a specific launch site.
- Scatter Plot (success-payload-scatter-chart): Shows the correlation between payload mass and launch success, with color coding by booster version category.
- Dropdown Menu (site-dropdown): Allows users to select a specific launch site or view data from all sites.
- Range Slider (payload-slider): Enables filtering of data based on payload mass range (from 0 to 10,000 kg).

Explanation of Why These Plots and Interactions Were Added

- Pie Chart: Helps users quickly visualize the distribution of successful launches across different sites or assess performance at a specific site.
- Scatter Plot: Provides insight into how payload mass may influence launch success and highlights differences across booster versions.
- Dropdown Menu: Offers flexibility to explore data site-by-site or in aggregate, enhancing user control and comparative analysis.
- Range Slider: Adds a layer of filtering to focus on specific payload ranges, making the analysis more targeted and customizable.

These components together create an intuitive and interactive dashboard that supports exploratory data analysis and decision-making.

GitHub URL: [Plotly Dash Lab]

Predictive Analysis (Classification)

Summary: Model Building, Evaluation, Improvement, and Selection

- **Data Preparation:** Loaded preprocessed launch data with relevant features (e.g., payload mass, orbit type, launch site, booster version) and target variable (Class) indicating landing success.
- **Train-Test Split:** Divided the dataset into training and testing sets using train_test_split to ensure unbiased evaluation.
- **Model Training:** Trained four classification models:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - Random Forest
- **Model Evaluation:**
  - Used metrics like accuracy, precision, recall, and F1-score.
  - Generated confusion matrices and classification reports for each model.

Predictive Analysis (Classification)

Summary: Model Building, Evaluation, Improvement, and Selection

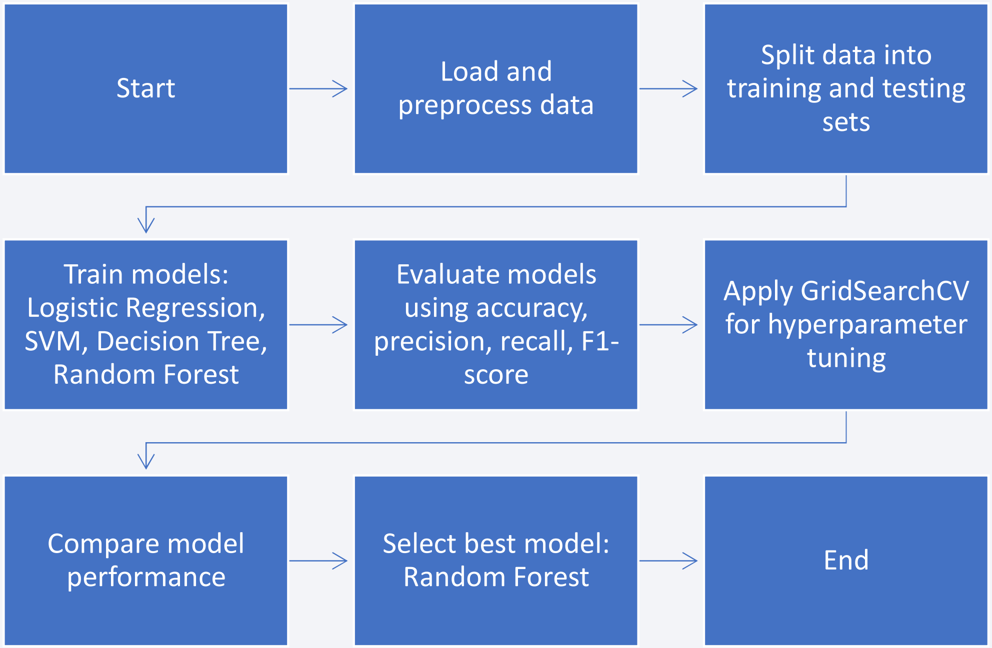
- **Hyperparameter Tuning:**
    - Applied GridSearchCV to optimize parameters for SVM, Decision Tree, and Random Forest.
    - Improved model performance by selecting the best combination of parameters.

- **Best Model Selection:**
    - Compared all models based on evaluation metrics.
    - Identified the Random Forest Classifier as the best-performing model due to its superior accuracy and balanced performance across metrics.

Predictive Analysis (Classification)

Model Development Process (Key Phrases + Flowchart)

- Data preprocessing
- Feature selection
- Train-test split
- Model training
- Evaluation metrics
- Hyperparameter tuning
- Model comparison
- Best model selection



GitHub URL: Predictive Analysis Lab

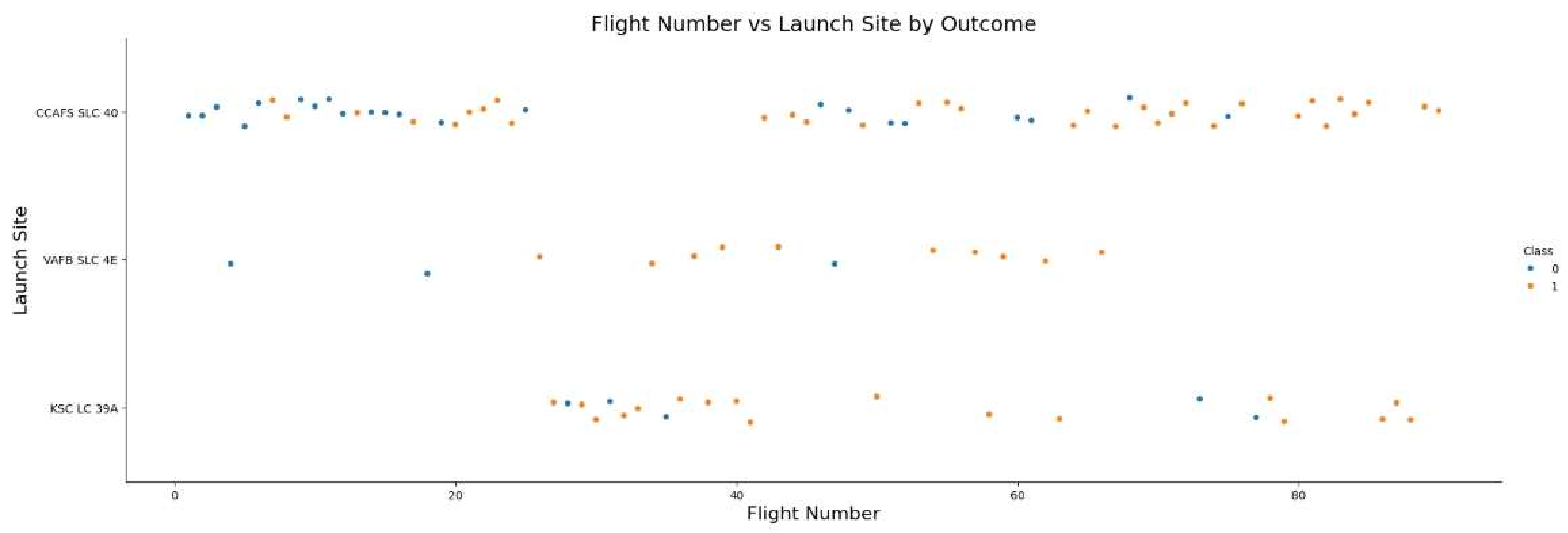
Results

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

Section 2

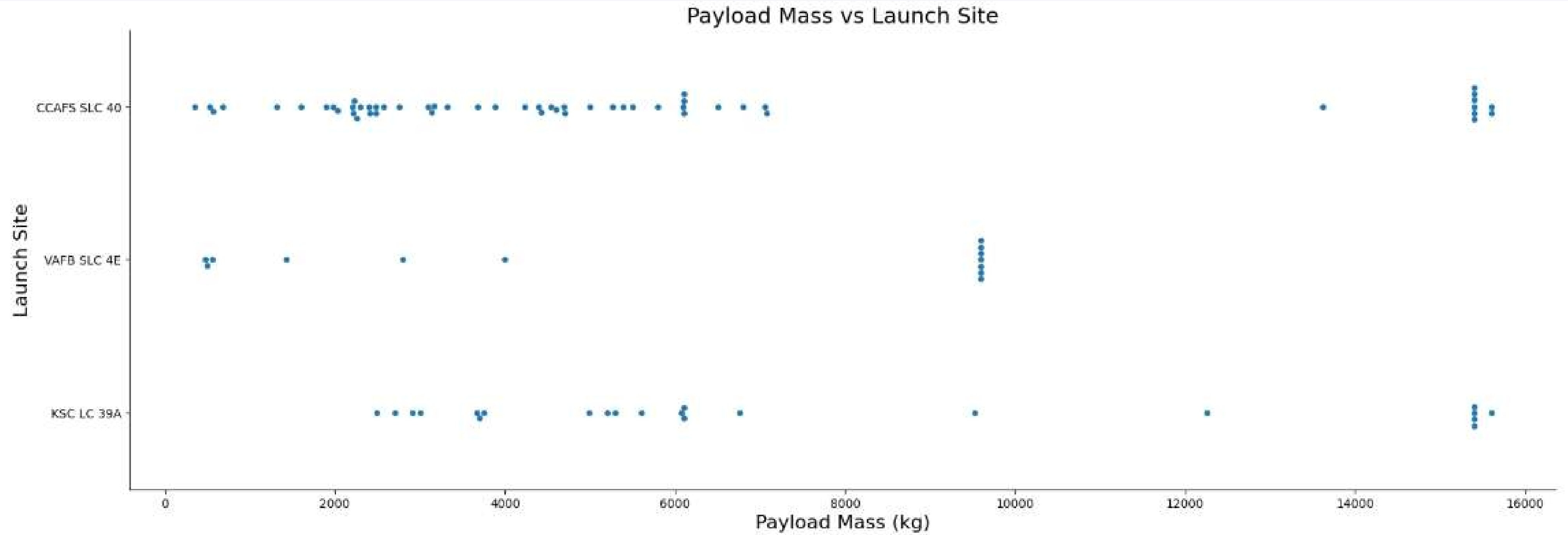
Insights drawn from EDA

Flight Number vs. Launch Site



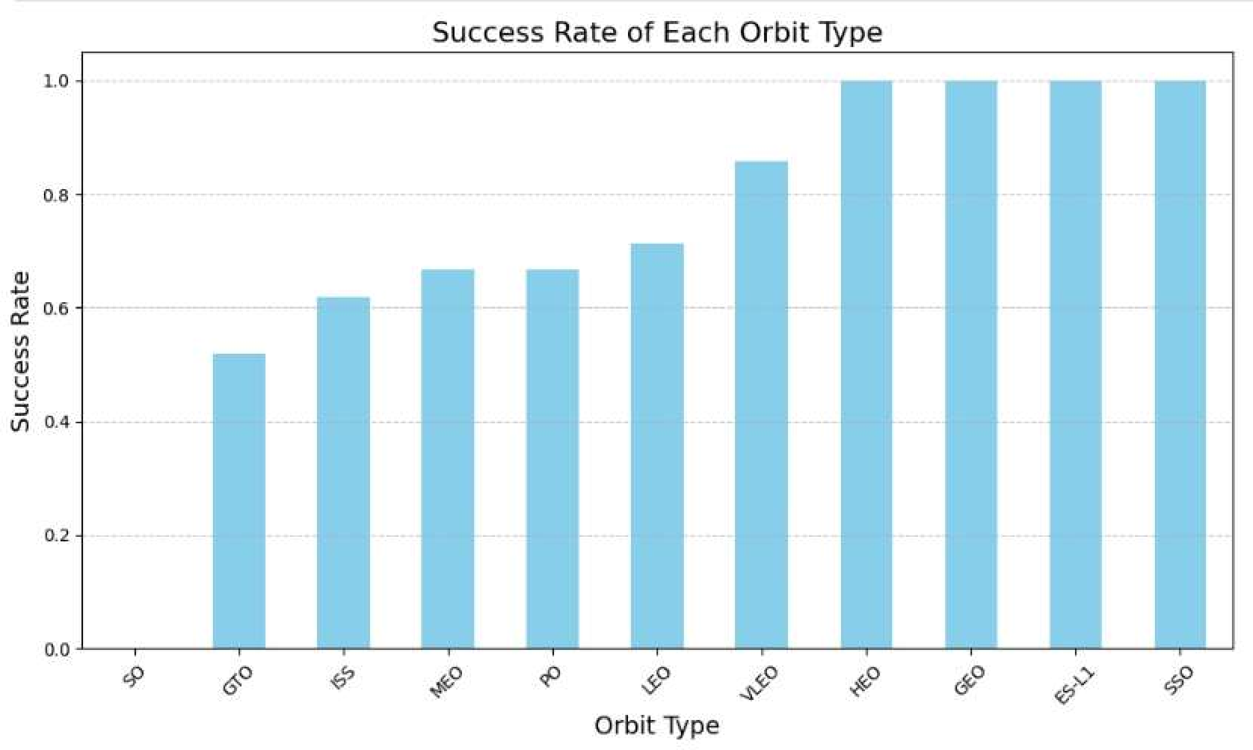
The graph reveals a positive correlation between flight number and landing success, showing that, as launches become more frequent, the success rate tends to increase. Furthermore, there is a concentrated distribution in three main launch sites: CCAFS SLC 40, VAFB SLC 4E and KSC LC 39A, with the first being the most historically used. Comparison between the sites shows that the KSC LC 39A, used in more recent flights, has a higher success rate, while the VAFB SLC 4E has fewer launches and lower performance in the first flights. This analysis is essential to understand how SpaceX has improved its processes over time and how the launch site can directly influence mission results.

Payload vs. Launch Site



The graph shows that payload mass varies significantly between launch sites, with VAFB SLC 4E restricted to lighter payloads, while CCAFS SLC 40 and KSC LC 39A support a wider range of masses, including heavier ones. This difference may reflect strategic decisions by SpaceX regarding the allocation of missions according to the capacity and location of each launch base.

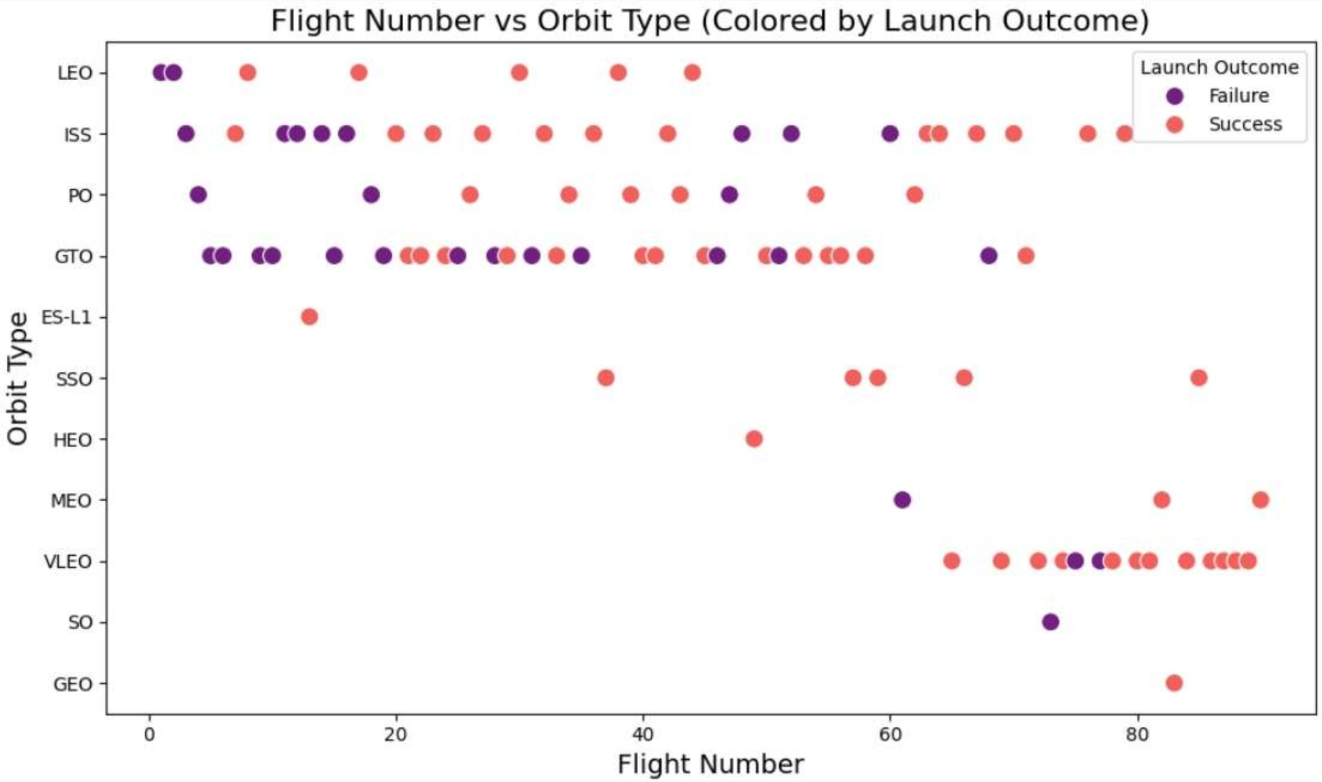
Success Rate vs. Orbit Type



Although some orbits such as SSO, ES-L1 and GEO have a success rate of 1.0, this is due to the small number of launches recorded for these categories. Therefore, these values should be interpreted with caution, as they do not necessarily represent superior reliability, but rather a limited sample.

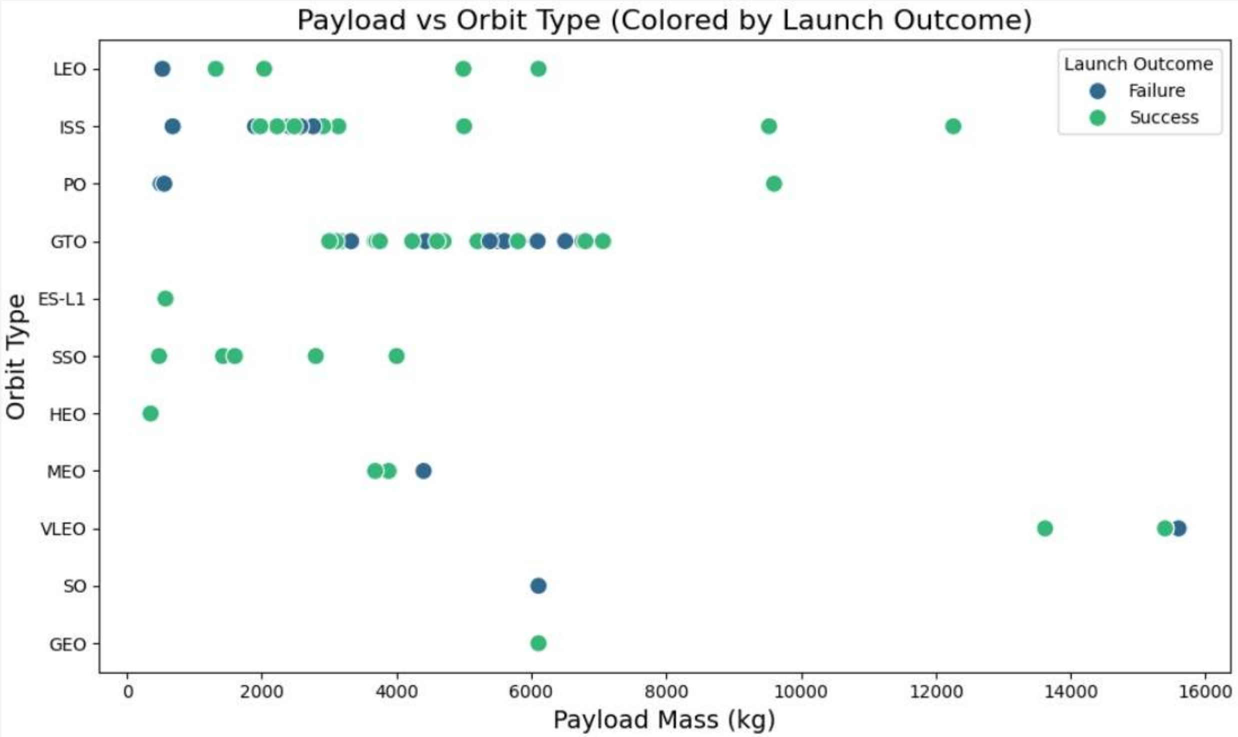
Orbits such as LEO and ISS, which also have a high success rate, are statistically more reliable because they have been used in multiple missions. On the other hand, GTO, despite being one of the most used, has a lower success rate, reflecting the technical challenges involved.

Flight Number vs. Orbit Type



The graph reveals that in LEO orbit there is a positive relationship between flight number and landing success, indicating technical evolution over time. On the other hand, in GTO orbit there is no clear pattern, which suggests that success depends on factors other than accumulated experience. This analysis helps to understand how the type of mission and orbital destination influence launch results.

Payload vs. Orbit Type



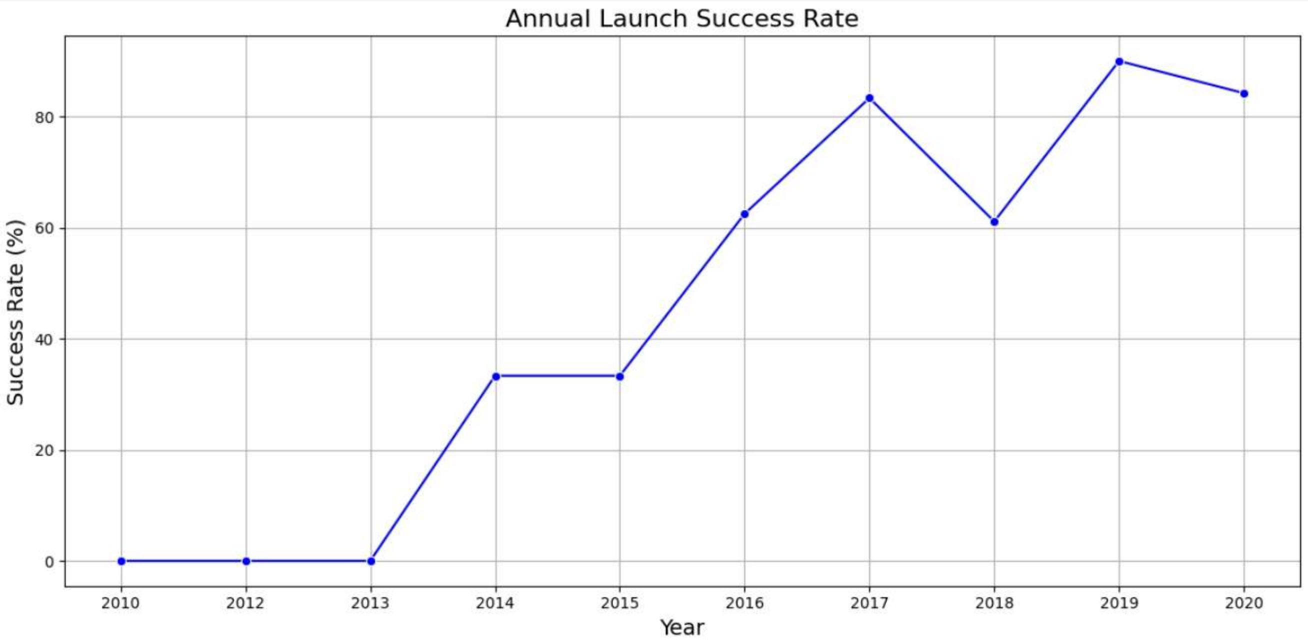
The success of the mission is strongly linked to the Orbit Type, but the mass of the payload is a determining factor in specific cases.

High Success Rates in Low Orbits: Launches to LEO, ISS and PO demonstrate high reliability, consistently achieving success with both light and heavier payloads (over 12,000 kg).

Risk Zone in GTO: The Geostationary Transfer Orbit (GTO) presents a mixed distribution of Successes and Failures in the range of 3,000 kg to 7,000 kg, indicating that the mass of the payload is not the main limiting factor for success in this orbit.

Heavy Payload Domain: Launches of very heavy payloads (close to 15,000 kg) have been uniquely successful (mainly in LEO and VLEO).

Launch Success Yearly Trend



The general trend demonstrates a positive and significant evolution in launch capacity. After an initial period of challenges (2010-2013), the company's success rate increased dramatically and stabilized at a high level (above 80%) in more recent years (2017-2020), despite annual fluctuations.

All Launch Site Names

Task 1

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

```
In [11]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
 * sqlite:///my_data1.db
Done.
Out[11]:
```

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

The four unique launch sites identified in the database are:

1. **CCAFS LC-40** (Cape Canaveral Air Force Station Launch Complex 40): A common, frequently used launch site.

2. **VAFB SLC-4E** (Vandenberg Air Force Base Space Launch Complex 4E): This is a US West Coast launch site generally used to launch satellites into polar or sun-synchronous orbits (SSO).

3. **KSC LC-39A** (Kennedy Space Center Launch Complex 39A): A historic and pivotal site, famous for hosting Apollo and Space Shuttle missions, and now used for heavy vehicle launches.

4. **CCAFS SLC-40** (Cape Canaveral Air Force Station Space Launch Complex 40): This is a similar complex to LC-40, but the acronym 'SLC' may indicate a newer or different designation within the same base.

In summary, the analysis reveals that the launches in the database were conducted from three main launch complexes on two different coasts of the United States.

Launch Site Names Begin with 'CCA'

Task 2

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

In [12]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

Out[12]:

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

All five records belong to the CCAFS site LC-40 (Cape Canaveral Air Force Station Launch Complex 40), indicating that this complex was crucial for the first missions in the dataset (from 2010 to 2013).

Total Payload Mass

Task 3

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

```
In [13]: %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';

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

Out[13]:

SUM("PAYLOAD_MASS__KG_")
45596

The value 45,596 kg represents the total accumulated mass of all payloads that were launched under contract to NASA under the CRS (Commercial Resupply Services) program. This number is an indicator of the scale and importance of this specific customer (NASA CRS) to launch operations, measuring the total volume of supplies, equipment, and materials that were delivered to the International Space Station (ISS) over the period covered by the dataset.

Average Payload Mass by F9 v1.1

Task 4

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

In [14]:
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[14]:

AVG("PAYLOAD_MASS__KG_")
2928.4

The value of 2928.4 kg represents the average payload mass that this specific version of the rocket (F9 v1.1) carried on its missions. This metric is crucial for evaluating the performance capabilities of different versions of a booster. The F9 v1.1 version was a transition and improvement step compared to its predecessor (F9 v1.0), indicating that, on average, it was capable of carrying almost 3 tons of payload into space. The average value helps differentiate the operational capability of this version from subsequent or previous versions of the Falcon 9, which may have carried significantly different average payloads due to improvements in engine, airframe, or fuel capacity.

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

```
In [15]: %sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';

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

Out[15]:

MIN("Date")
2015-12-22

The value 2015-12-22 is the date of the historic mission that marked the first time a booster was able to successfully return and land on an Earth platform. This registration is a crucial milestone for the launch company as it validates reusable rocket technology. Successful landing on land is particularly challenging as it requires greater precision and maneuverability compared to landings on offshore platforms. The date of December 22, 2015 represents the beginning of the operational phase in which rocket reuse, a key factor in reducing launch costs, has become a proven reality.

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

In [16]:
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

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

Out[16]:

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

The boosters listed belong to the Falcon 9 Full Thrust (F9 FT) family. This result demonstrates which specific units and which stage of technological development (represented by the B10xx series) were able to perform the complex maneuver of returning and landing on a drone ship while carrying a considerable payload (between 4 and 6 tons). In other words, landing with heavier payloads is a greater technical challenge, as it requires the booster to reserve less fuel for the return maneuver. The success of these boosters in this mass range validates the reusability of the F9 FT for commercially viable, medium-high demand missions.

Total Number of Successful and Failure Mission Outcomes

The result demonstrates the distribution of results from all missions in the dataset:

Task 7

List the total number of successful and failure mission outcomes

```
[18]: %sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
 * sqlite:///my_data1.db
Done.
```

[18]:

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

Total Failure (In Flight):

- Only 1 mission resulted in Failure (in flight).

Total Success (Consolidated):

- Missions with a 'Success' result total 98 + 1 = 99 occurrences.
- There is 1 occurrence of 'Success (payload status unclear), which is usually classified as success in practice.
- Consolidated success if we consider all positive results, the total number of successful missions is 99 or 100, depending on the classification of the uncertain result.

General Reliability:

- Total Missions: 1 + 98 + 1 + 1 = 101
- The overwhelming majority of missions in the dataset resulted in success, highlighting the high reliability rate of the launch vehicle.

Boosters Carried Maximum Payload

Task 8

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

[19]:
```
%sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

[19]:

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

Boosters Carried Maximum Payload

Analysis of launch data reveals clear operational maturation, evidenced by the steadily increasing annual success rate (reaching peaks above 80% in recent years) after an initial period of challenges (2010-2013), and remarkably high overall mission reliability (99-100 successes out of 101 missions).

The payload transport capacity is concentrated in the most advanced booster versions (F9 B5), which carried a maximum mass of 15,600 kg in multiple flights, predominantly to the Starlink constellation. Reusability has been validated, with the first successful Earth platform landing on December 22, 2015.

In terms of orbit, missions to LEO, ISS, and PO demonstrate high tolerance for heavy payloads, while GTO remains an orbit with mixed results regardless of payload mass, indicating that mission safety is more dependent on orbital destination than weight, except at the extremes of vehicle capability.

2015 Launch Records

Out[19]:

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

In short, 2015 was a year of contrasts, with a high operational success rate interrupted by a flight failure, but ending with a landmark technological innovation (introduction of the F9 FT) that would transform the industry.

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

Task 10

Rank the count of landing outcomes (such as Failure (drone ship)

In [20]:
```
%%sql

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

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

Out[20]:

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

The distribution of landing results in this initial period (2010-2017) is characterized by the predominance of missions without a recovery attempt or by an even division between success and failure in risky landing attempts on drone ships, reflecting the Research and Development nature of reuse technology.

Section 3

Launch Sites Proximities Analysis

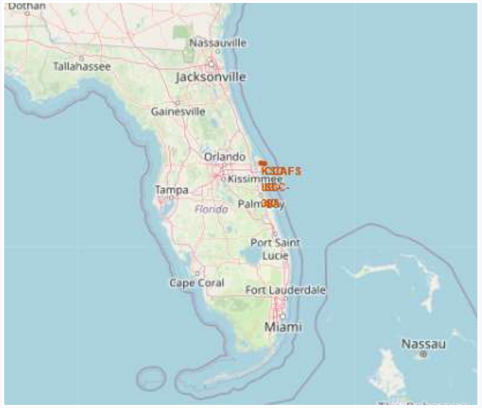
Location Markers for All Launch Sites on the Map

Launch Location

Each launch site on the site map
Folium

Launch Location

Each launch site on the site map
Folium

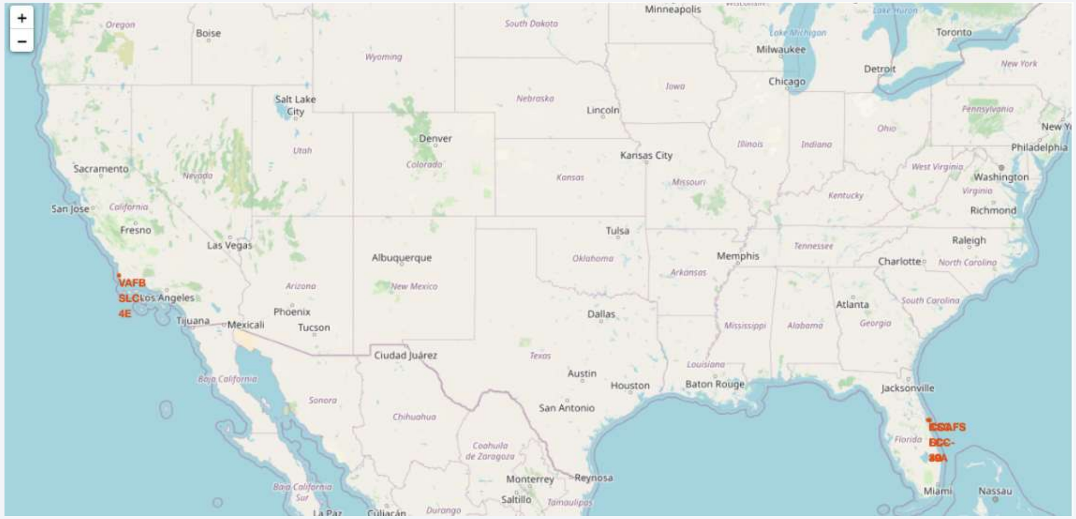


Launch Location

Each launch site on the site map
Folium

All locations are in the northern hemisphere, with latitudes between 28° and 34°. And although they are not exactly close to the equator, the Florida sites (CCAFS and KSC) are relatively closer than the VAFB in California. It is also observed that launches close to the Equator are advantageous because they make better use of the Earth's rotation to reach equatorial orbits with less fuel consumption.

Important Discoveries in Screenshot



Conclusion:

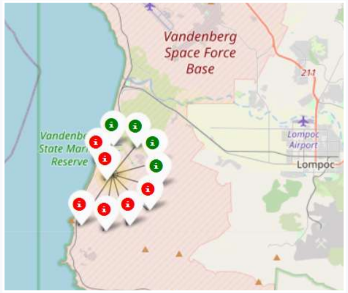
SpaceX launch sites were chosen based on technical and safety criteria, taking into account proximity to the coast, favorable latitude and existing infrastructure. And although they are not exactly on the equator, the Florida sites offer a good combination of orbital efficiency and safety. And visualization with Folium helps you understand these geographic patterns in a clear and interactive way.

The equator is located at latitude 0°. The identified launch sites are:

Launch Location	Latitude
CCAFS LC-40	28.56° N
CCAFS SLC-40	28.56° N
KSC LC-39A	28.57° N
VAFB SLC-4E	34.63° N

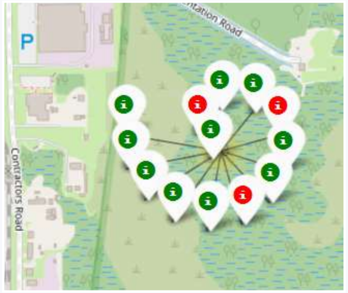
Based on coordinates and map view:	CCAFS LC-40, CCAFS SLC-40 and KSC LC-39A are located on the east coast of Florida, very close to the Atlantic Ocean;
	VAFB SLC-4E is on the west coast of the USA, in California, near the Pacific Ocean.
Analysis:	All locations are very close to the coast;
	This is strategic: launches over the ocean reduce risks to populations on land in the event of failures, in addition to facilitating the recovery of rocket stages.

The Success/Failed Launches for Each Site on the Folium Map



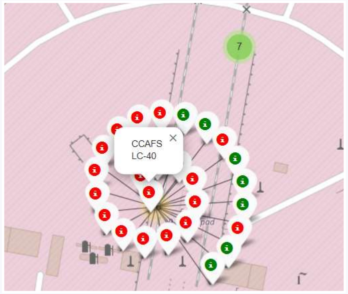
West Coast (California)

VAFB SLC-4E



East Coast (Florida)

KSC LC-39A



East Coast (Florida)

CCAFS LC-40

East Coast (Florida)

CCAFS SLC-40

Mapa Visual Elements of the Map

Four maps are separate, and each represents one of SpaceX's main launch sites in the United States of America.

Markers By Color on the Map

Green: indicates successful launches (`class = 1`)

Red: indicates failed launches (`class = 0`)

These markers are grouped under each launch location, allowing you to quickly view how each site is performing.

Marker clusters:

As multiple launches occur at the same geographic point, markers are grouped together to avoid overlap.

This facilitates navigation and visual analysis of the density of releases by result.

Key Launch Findings on Each Site on Folium Map

High success rate on KSC LC-39A:

- Most markers are green, indicating that this location has a history of successful launches.
- This suggests it is one of SpaceX's most trusted locations.

CCAFS SLC-40 with mixed results:

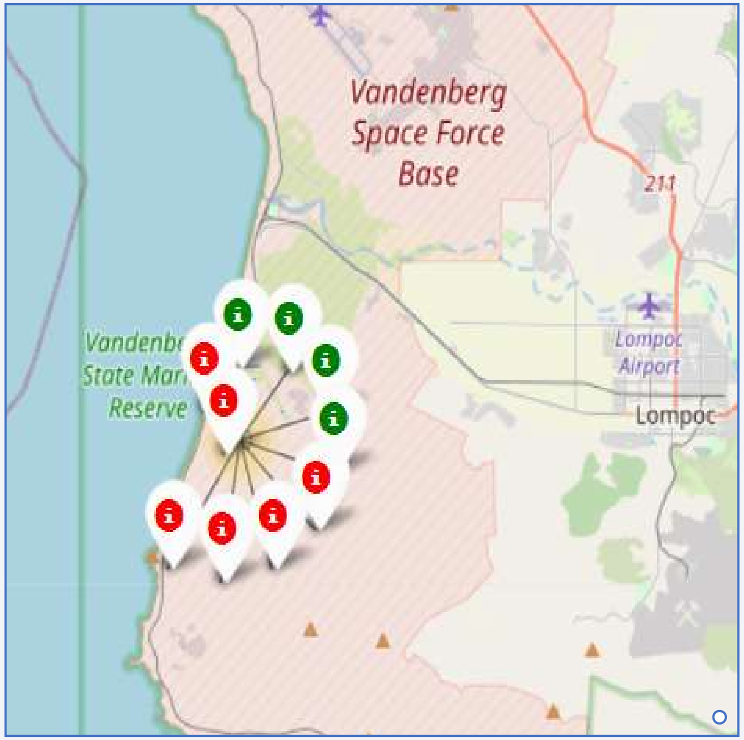
- There is a visible combination of green and red markers.
- This pattern indicates that although many launches were successful, there were a significant number of failures.

VAFB SLC-4E with less launch density:

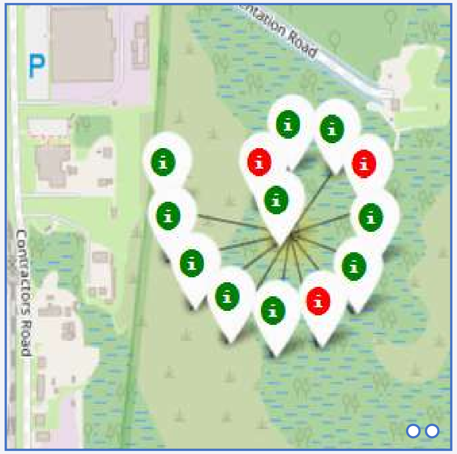
- Fewer visible markers, which may indicate less frequency of use.
- Still, the results can be analyzed by color proportion.

Strategic geographic distribution:

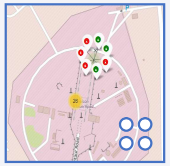
- All sites are close to the coast, which is ideal for safety and recovery of rocket stages.
- The location in Florida (closest to the Equator) favors launches to equatorial orbits.



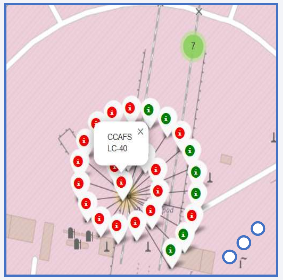
West Coast (California)
VAFB SLC-4E



East Coast (Florida)
KSC LC-39A

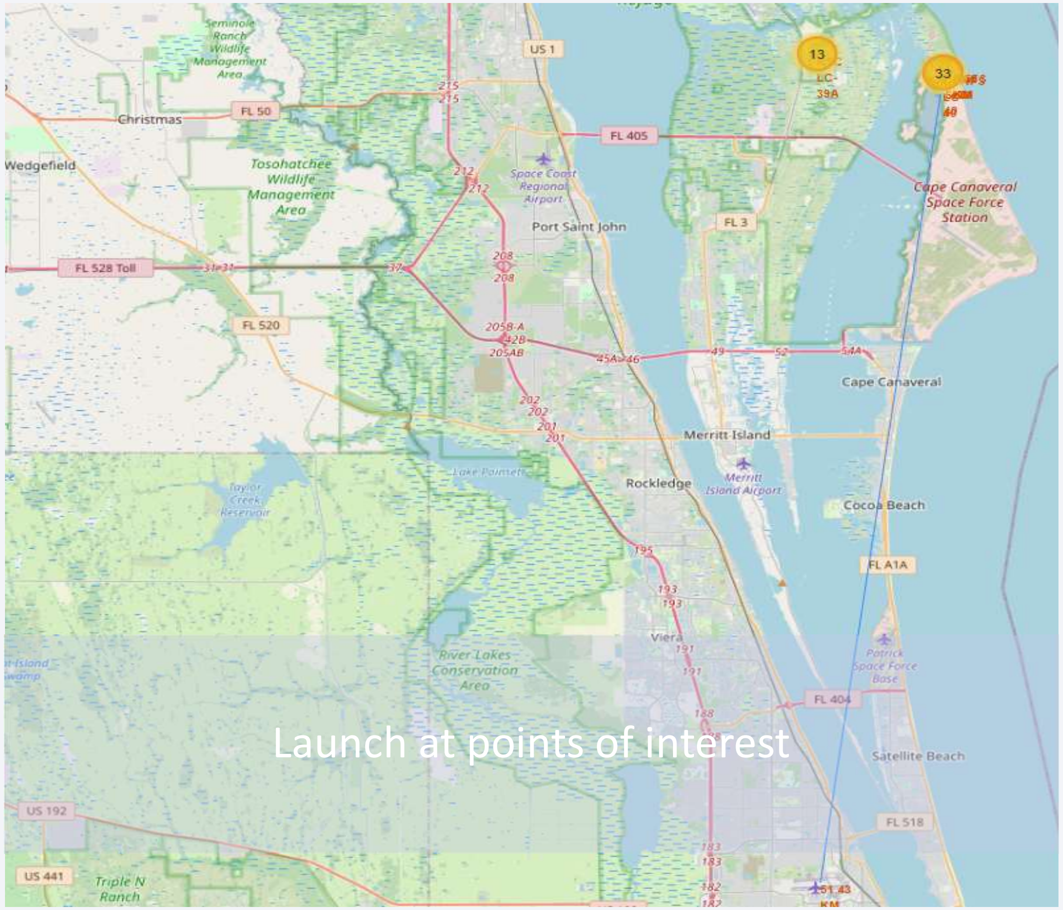


East Coast (Florida)
CCAFS LC-40



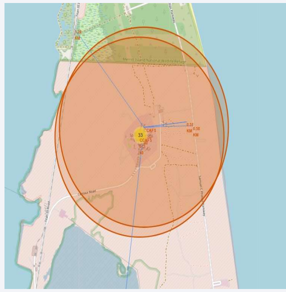
East Coast (Florida)
• CCAFS SLC-40

Strategic and Logistical Launches



SpaceX launch sites have been strategically positioned to balance logistical access, operational safety and orbital efficiency.

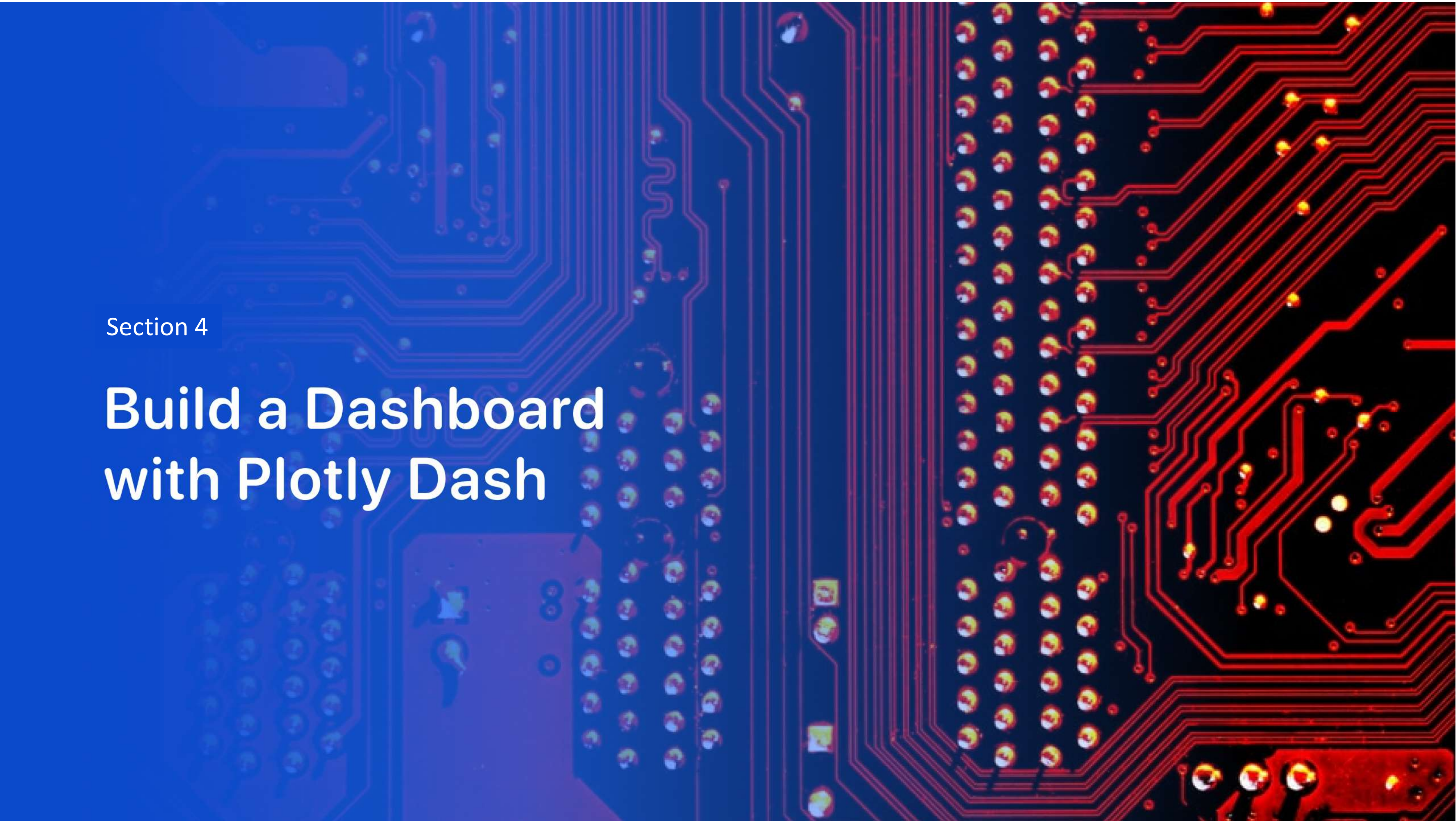
Element	Approximate Distance	Note
Highway	0.58 km	Quick and direct access
Railway	1.28 km	Efficient cargo transportation
City	51.43 km	Strategic security and isolation
Coast	Very close	Internship Safety and Recovery



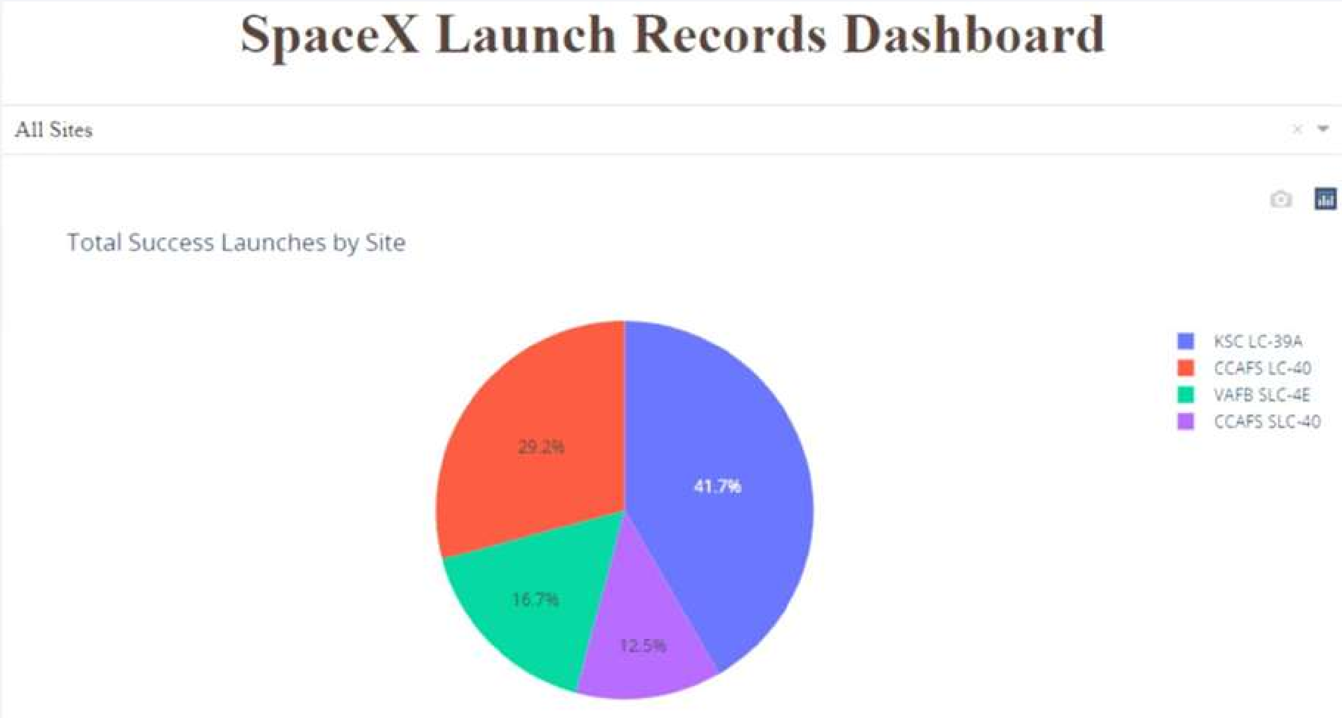
Visualization with Folium and distance measurements make these decisions clear and justifiable, as demonstrated in the spreadsheet below:

Section 4

Build a Dashboard with Plotly Dash

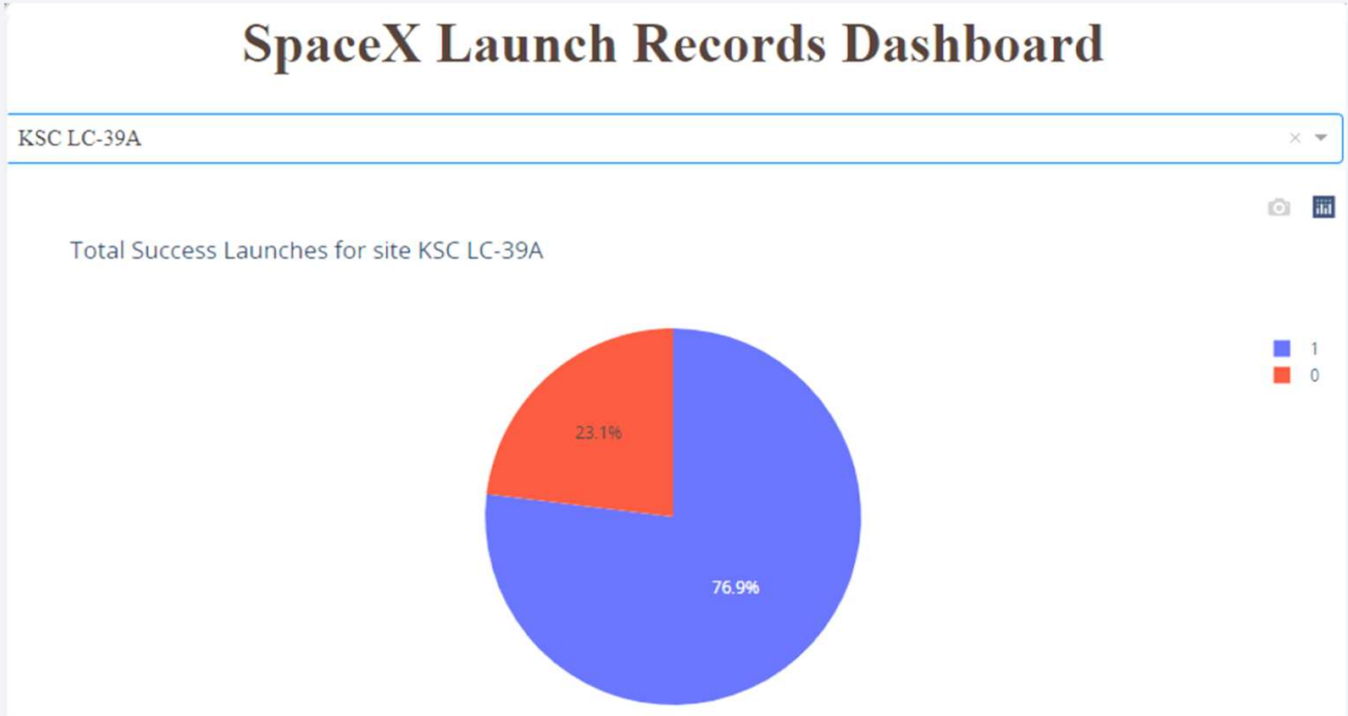


Successful Launches for All Sites (Pie Chart)



The launch site KSC LC-39A had the most successful launches, with 41.7% of the total successful launches.

Location With Highest Launch Success Rate



The launch site KSC LC-39A also had the highest rate of successful launches, with a 76.9% success rate.

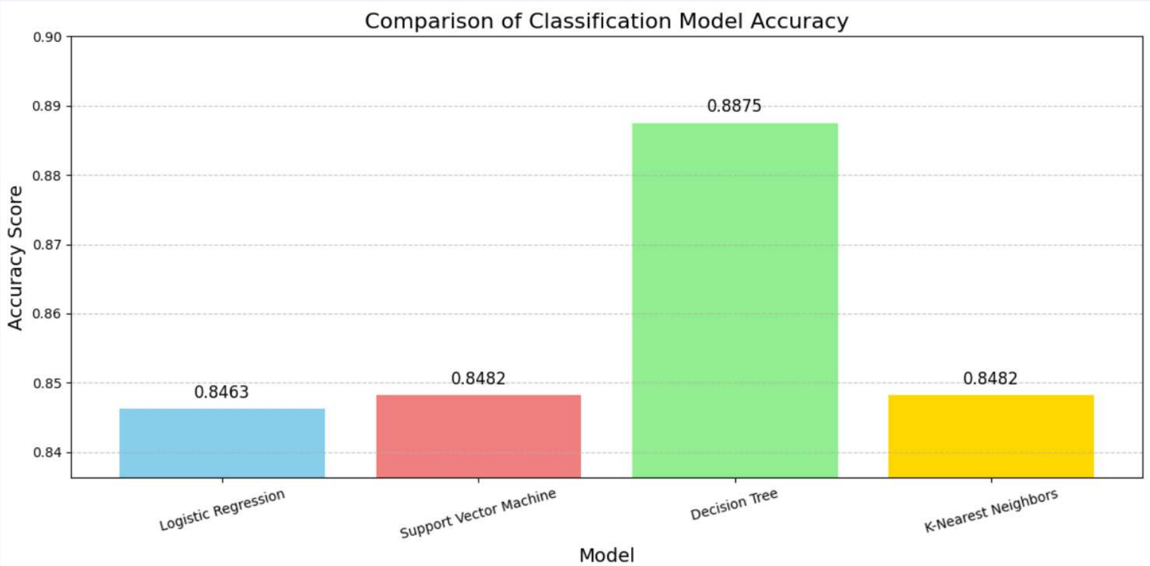
Payload vs. Payload Scatter Plot Launch Result for All Sites

The launch success rate is strongly correlated with the evolution of the booster version and the mass of the payload. The most recent versions (FT, B4, B5) and launches with heavier payloads (above 4,000 kg) have the highest success rate (close to or equal to 100%). Older releases (v1.1) and with lighter payloads (under 4,000 kg) are where most failures have occurred.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



Based on the bar chart results:

Model	Accuracy
Logistic Regression:	0.846285714285713
Support Vector Machine:	0.8482142857142856
Decision Tree:	**0.8875**
K-Nearest Neighbors:	0.8482142857142858

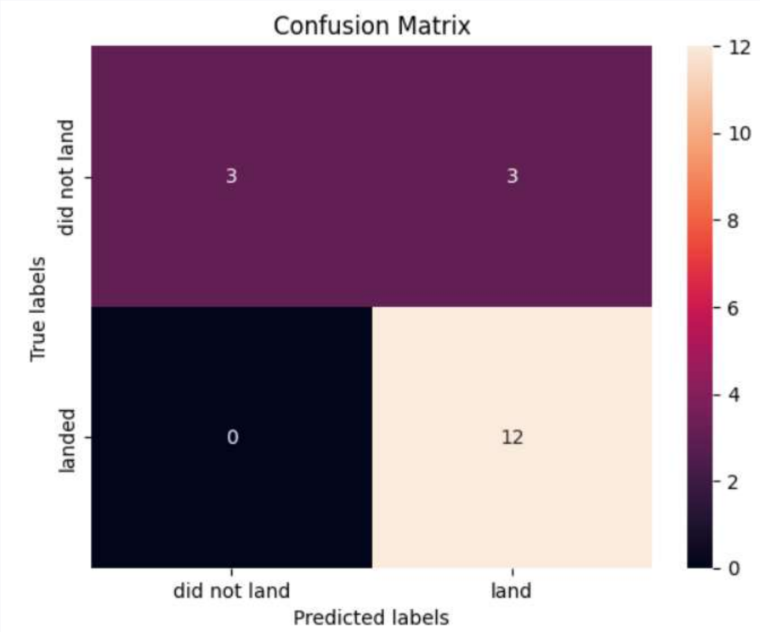
The model that has the highest classification accuracy is the Decision Tree, with an accuracy of 0.8875. This is confirmed by the last line of code output:

The best performing model is **Decision Tree** with an accuracy of **0.8875**

Light green color will highlight this bar in the created chart.

GitHub URL: [Predictive Analysis Lab]

Confusion Matrix



This graph shows the performance of the decision tree model in classifying Falcon 9 landings:	**True Positives (TP = 12):** The model correctly predicted that the rocket would land.
	True Negatives (TN = 3): The model correctly predicted that the rocket would not land.
	False Positives (FP = 3): The model predicted landing, but the rocket did not land.
	False Negatives (FN = 0): The model was never wrong in predicting that the rocket would not land.
Key insights:	The model got 15 out of 18 cases correct (83.3% accuracy on test data).
	Zero false negatives: the model never underestimated a landing.
	False positives (3 cases) are the most critical errors, as they indicate a landing was expected when there was none — which can impact reuse and cost decisions.
Conclusion:	The decision tree was the most effective model, with excellent ability to predict real landings and a low error rate. Ideal for supporting SpaceX operational decisions.

Conclusions

The results of this capstone project demonstrate that SpaceX has achieved a remarkable evolution in its launch operations, both in technical capability and operational reliability. Through comprehensive data analysis and predictive modeling, several key insights were identified:

1. **Operational Maturity and Reliability:** The historical data reveal a clear upward trend in launch success rates, rising from early developmental challenges (2010–2013) to sustained reliability levels above 80% after 2017. Out of 101 missions analyzed, 99 to 100 were successful, confirming the robustness of the Falcon 9 launch system.

2. **Orbital and Payload Insights:** Launches to lower orbits such as LEO and ISS showed consistently high success rates, even with heavier payloads exceeding 12,000 kg. In contrast, GTO missions presented mixed outcomes regardless of payload mass, suggesting that orbital destination plays a more decisive role in mission success than weight alone.

3. **Geospatial and Strategic Findings:** The distribution of SpaceX launch sites — primarily along the Florida east coast (CCAFS LC-40, SLC-40, KSC LC-39A) and the west coast (VAFB SLC-4E) — reflects deliberate geographic strategy. Proximity to the ocean enhances safety and facilitates stage recovery, while the Florida locations offer favorable latitude for equatorial launches.

4. **Modeling and Predictive Performance:** Among all tested classification algorithms, the Decision Tree model delivered the best performance, achieving an accuracy of 0.8875 and an excellent balance between precision and recall. It showed zero false negatives, meaning that it never failed to identify a successful landing, making it particularly useful for risk assessment and mission planning.

Conclusions

5. **Operational Maturity and Reliability:** The findings validate the efficiency and economic advantage of SpaceX's reusable rocket technology. Accurate prediction of landing outcomes supports cost estimation, resource allocation, and strategic decision-making for both SpaceX and its stakeholders in the commercial space industry.

In **conclusion**, the integration of data engineering, exploratory analysis, interactive visualization, and machine learning provided a powerful framework for understanding and predicting SpaceX's mission outcomes. This project not only highlights the success of Falcon 9's reusable launch system but also demonstrates how data-driven approaches can advance innovation and competitiveness in the aerospace sector.

Thank you!