IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Roberth Arias
2025

Outline

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

Executive Summary

Summary of methodologies

• Data Collection: Obtained datasets from SpaceX REST API and Wikipedia through web scraping to build a comprehensive dataset of Falcon 9 launches.

• Data Wrangling: Cleaned and standardized the data, handling missing values and transforming variables (e.g., payload mass, categorical encoding).

• Exploratory Data Analysis (EDA): Applied visualization techniques (Seaborn, Matplotlib) to identify patterns between payload, orbit, launch sites, and mission success.

• Interactive Analytics: Built interactive visualizations using Folium (maps of launch sites and outcomes) and Plotly Dash (dashboard with filters, payload sliders, and site selection).

• Predictive Analysis: Developed and tuned classification models (Logistic Regression, SVM, Decision Tree, KNN). Evaluated accuracy using GridSearchCV.

• Key Result: The Decision Tree classifier achieved the best predictive performance (~94% accuracy) in forecasting Falcon 9 first-stage landing success.

Executive Results

•**Exploratory Data Analysis (EDA):**
•Identified the main launch sites used by SpaceX: CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E, and others.
•Observed that launch success rates varied across sites, with KSC LC-39A showing consistently higher success ratios.
•Payload mass was found to influence launch outcomes, with moderate payloads (4,000–6,000 kg) showing higher success likelihood.
•Orbit type also had a significant impact on outcomes; certain orbits (e.g., GEO, SSO, ES-L1) displayed higher success rates.
•The yearly trend analysis revealed a steady increase in success rates after 2015, highlighting improvements in SpaceX technology and operations.
•**Interactive Visual Analytics (Folium and Dash):**
•Folium maps allowed visualization of launch site locations and proximity to infrastructure (highways, railways, coastlines), supporting the importance of geography in site selection.
•Dash dashboards provided interactive analysis:
  - •Pie charts showing success distribution by site and outcome.
  - •Scatter plots linking payload mass and booster versions with launch results.

•These tools enabled an intuitive understanding of how location and mission parameters affect launch success.
•**Predictive Analysis (Classification Models):**
•Several classification algorithms were tested: Logistic Regression, SVM, KNN, and Decision Tree.
•Accuracy results:
  - •Logistic Regression: 83.3%
  - •SVM: 83.3%
  - •KNN: 83.3%
  - •Decision Tree: **94.4% (highest accuracy)**

•The Decision Tree model achieved the best performance, confirmed by its confusion matrix, showing high precision in predicting successful landings.
•**Overall Findings:**
•Launch success is influenced by both technical (payload mass, booster type, orbit) and geographical (launch site, proximity to infrastructure) factors.
•Interactive visualization tools (Folium, Dash) greatly enhance interpretability of complex datasets.
•Machine learning models, particularly Decision Trees, are effective in predicting launch outcomes with high accuracy.

Introduction

•The commercial space era has arrived, with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX driving innovation.
•SpaceX has achieved milestones such as:
•Sending spacecraft to the ISS.
•Deploying *Starlink* satellite internet constellation.
•Conducting manned space missions.
•A key factor in SpaceX's success is the **reusability of the Falcon 9 first stage**, which reduces costs from $165M (competitors) to $62M.
•However, not all first stages are successfully recovered — some crash or are sacrificed due to mission parameters (payload, orbit, customer).
•**Problem Statement**: If we can predict whether the first stage will land successfully, we can estimate the launch cost.
• **role**: As a data scientist at *SpaceY* (a competitor company), build models to predict first stage landing outcomes using publicly available SpaceX data.

Section 1

Methodology

Methodology

Executive Summary

- Data Collection Methodology
  - SpaceX REST API (api.spacexdata.com/v4/launches/past) → JSON launch data.
  - Extracted: Rocket ID, Payload, Launch Site, Landing Outcome, Mission Success.
  - Additional web scraping (Wikipedia – Falcon 9 tables, with BeautifulSoup).
  - Converted JSON & HTML into Pandas DataFrames.

  Data Wrangling
  - Normalized JSON with json_normalize.
  - Enriched data: Rocket, Payload, Core, Launchpad (API endpoints).
  - Filtered: only Falcon 9 launches (excluded Falcon 1).
  - Missing values handling:
    - Payload Mass → imputed with mean value.
    - Landing Pad → kept null (no landing attempt).
  - OneHot Encoding applied to categorical features (Landing Outcome, Orbit).

EDA with Visualization

- Plotted relationships between Flight Number, Payload Mass, Orbit Type, and Launch Site.
- Observed that:
  - Success rates improved significantly after 2013.
  - Launch success varies by site (e.g., KSC LC-39A ≈ 77% vs. CCAFS LC-40 ≈ 60%).
  - Payload Mass influences landing outcome (heavy payloads >10,000kg show higher success in some sites).
- Visualizations used: scatter plots, bar charts, and trend lines.

EDA with SQL

- Queried launch database to:
  - List unique launch sites.
  - Calculate average payloads by rocket version.
  - Count successes vs. failures.
  - Identify first successful landings.
- SQL queries complemented visual exploration by providing exact aggregates and filtering conditions.

Folium Maps

- Built interactive maps to explore launch sites and proximities.
- Added markers, circles, and popups for each launch site.
- Visualized landing outcomes with color coding (success/failure).
- Calculated distances from launch sites to highways, railways, and coastlines.
- Insights: geographic factors can influence launch feasibility and costs.

Plotly Dash Dashboard

- Developed an interactive dashboard with dropdown menus and range sliders.
- Visualizations included:
  - Pie charts of launch success counts (per site and overall).
  - Scatter plots of Payload vs. Launch Outcome with adjustable filters.
- Provided stakeholders with an intuitive tool to explore launch performance in real time.
- Insights: easier identification of payload ranges and sites with highest success probability.

Pipeline Construction

- Preprocessed data: standardized numerical features and applied One-Hot Encoding to categorical variables.
- Split dataset into training (80%) and test (20%) subsets.

Model Training

- Implemented multiple classification algorithms:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - K-Nearest Neighbors (KNN)

Hyperparameter Tuning

- Applied GridSearchCV to find optimal hyperparameters for each model.
- Example: tuning tree depth for Decision Trees, kernel types for SVM, and k values for KNN.

Evaluation

- Compared models using accuracy scores on test set.
- Best model selected based on highest accuracy.
- Generated confusion matrix for the best performing model to evaluate misclassifications.

Data Collection

Data Sources

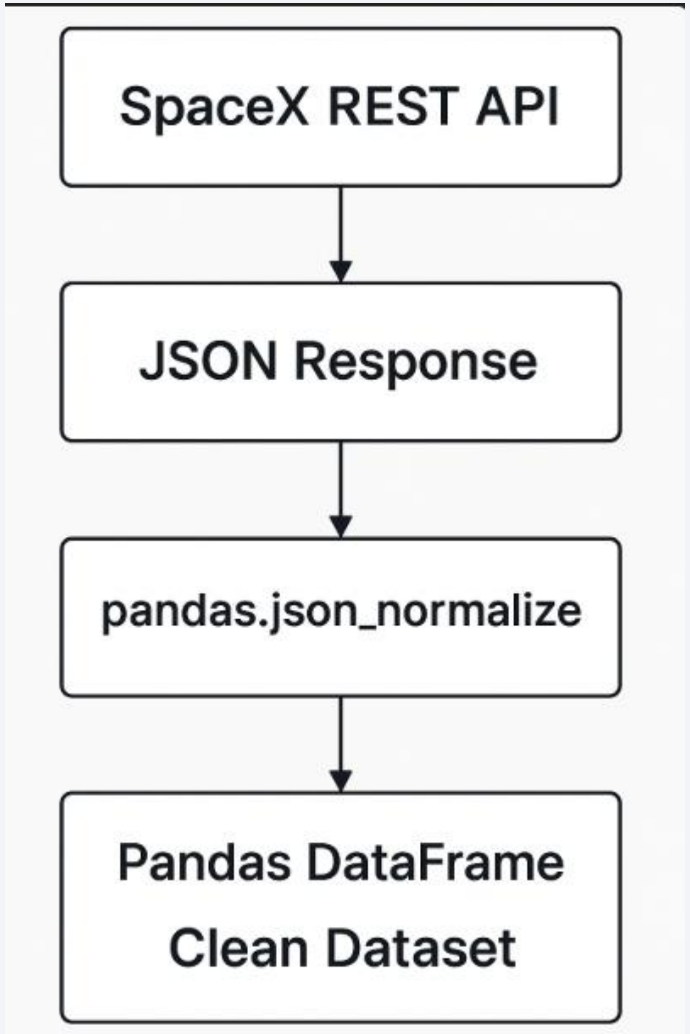
• SpaceX REST API (api.spacexdata.com/v4/launches/past):
  - Downloaded past launch information (rocket, payloads, launchpads, cores, landing outcomes, etc.).
  - Used requests for the query and pandas.json_normalize() to transform JSON into DataFrame.
  - Enriched data with additional endpoints (rockets, launchpads, payloads, cores).

Web Scraping (Wikipedia):

  - Used requests + BeautifulSoup to extract Falcon 9 launch tables.
  - Parsed key columns: Flight No., Date, Time, Launch Site, Payload, Payload Mass, Orbit, Customer, Launch Outcome, Booster Landing.
  - Normalized column names and cleaned data (e.g., converted Payload Mass to numeric in kg).

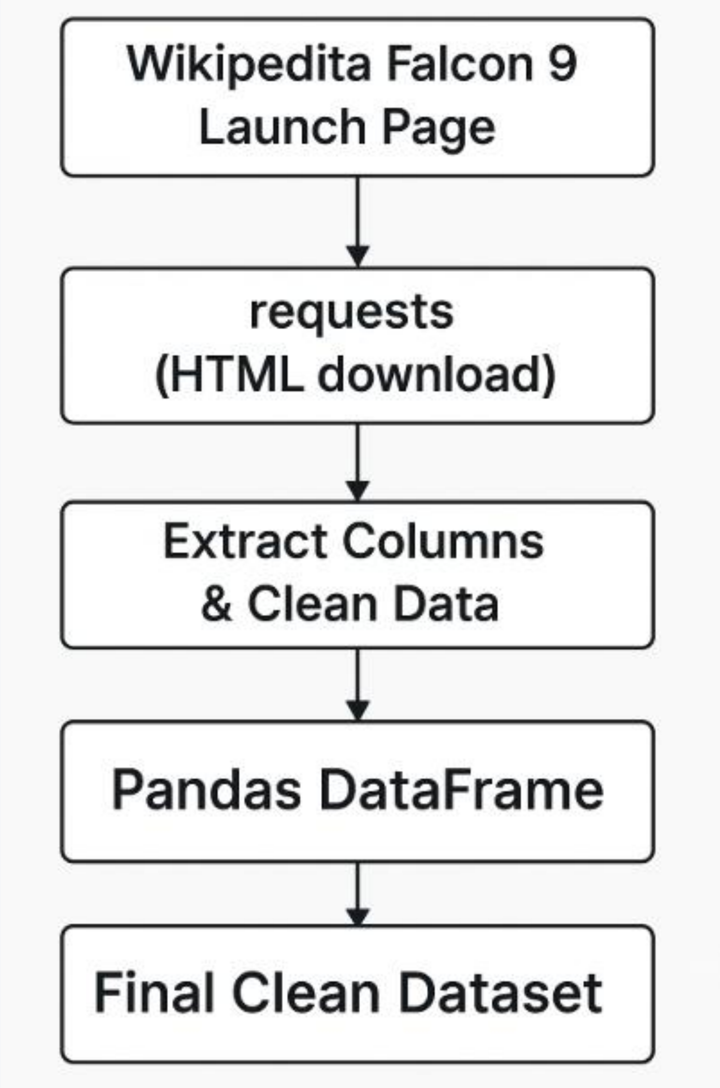
Data Collection – SpaceX API

- Queried endpoint: https://api.spacexdata.com/v4/launches/past.
- Used requests to send GET requests.
- Response format: JSON (list of launches with nested objects).
- Normalized JSON into tabular format using pandas.json_normalize().
- Enriched dataset with additional API calls:
  - /rockets → rocket version & details
  - /launchpads → site coordinates & names
  - /payloads → payload mass & orbit
  - /cores → landing outcome
- Final result: Pandas DataFrame with launch metadata (rocket, payload, site, orbit, landing outcome).



Data Collection - Scraping

• Source: Wikipedia page “List of Falcon 9 and Falcon Heavy launches”.
• Used requests to download HTML content.
• Parsed HTML with BeautifulSoup to extract Falcon 9 launch tables.
• Extracted and cleaned key columns:
  - Flight No., Date, Time, Version Booster, Launch Site, Payload, Payload Mass, Orbit, Customer, Launch Outcome, Booster Landing.
• Normalized headers and transformed data into a Pandas DataFrame.
• Cleaned payload mass column (converted to numeric values in kg).
• Final dataset: structured tabular data ready for EDA and merging with API data.



Data Wrangling

• Combined datasets (API + Wikipedia scraping).
• Selected relevant columns (Flight No., Payload Mass, Orbit, Launch Site, Landing Outcome, etc.).
• Removed Falcon 1 records (kept only Falcon 9).
• Handling missing values:
  - Payload Mass → replaced with mean.
  - Landing Pad → kept as null (represents no landing attempt).
• Converted data types (e.g., Payload Mass → numeric in kg, Dates → datetime).
• Encoded categorical variables with OneHot Encoding (Orbit, Launch Site, Outcome).

EDA with Data Visualization

- Scatter Plot – Flight Number vs Launch Site
  → To analyze how launch experience (number of flights) impacts success rate at different sites.

- Scatter Plot – Payload Mass vs Launch Site
  → To observe whether payload mass influences success depending on the launch site.

- Bar Chart – Success Rate by Orbit Type
  → To compare the average success rate across different orbit types.

- Scatter Plot – Flight Number vs Orbit Type
  → To identify trends in success related to accumulated experience across orbits.

- Scatter Plot – Payload Mass vs Orbit Type
  → To explore the relationship between payload mass and orbit type.

- Line Chart – Launch Success Yearly Trend
  → To visualize how the success rate has improved steadily from 2013 to 2020.

- Success rate has increased significantly over time (especially after 2013).
- Certain launch sites and orbit types consistently show higher success rates.
- Payload mass influences the probability of success for specific sites and orbit types.

EDA with SQL

- Task 1: Listed the unique launch sites.
- Task 2: Displayed 5 records where launch sites start with “CCA”.
- Task 3: Calculated the total payload mass launched by NASA (CRS).
- Task 4: Computed the average payload mass carried by F9 v1.1 boosters.
- Task 5: Retrieved the date of the first successful ground-pad landing.
- Task 6: Identified booster versions that successfully landed on a drone ship with payload mass between 4000–6000 kg.
- Task 7: Counted the total number of successful vs failed missions.
- Task 8: Listed booster versions that carried the maximum payload mass.
- Task 9: Displayed failures on drone ships in 2015, including month, booster, launch site, and outcome.
- Task 10: Ranked landing outcomes between 2010-06-04 and 2017-03-20 in descending order.

Key Insights:

- Launches are concentrated in a few major sites (KSC LC-39A, VAFB SLC-4E, CCAFS LC-40).
- Payload mass and booster version strongly influence mission outcomes.
- Drone ship landings have mixed success, while ground pad landings show higher reliability.
- Overall success rates improved significantly after 2013.

Build an Interactive Map with Folium

- Task 1: Mark all launch sites on a map

  Added markers and circle markers for each SpaceX launch site.
  → Purpose: provide geographic context and visualize site distribution.

- Task 2: Mark success/failed launches for each site

  Added color-coded markers (e.g., green for success, red for failure).
  → Purpose: quickly compare outcomes by location.

- Task 3: Calculate distances between a launch site and nearby facilities

  Added lines/polylines to measure proximity to roads, railways, and coasts.
  → Purpose: evaluate site suitability and potential logistic advantages.

Key Insights:

- Launch sites are strategically located near the coast.
- Success/failure patterns vary across sites.
- Proximity to infrastructure (roads/ports) is essential for optimal site selection.

Build a Dashboard with Plotly Dash

Plots added

• Pie chart (success-pie-chart):
  - ALL sites: shows total successful launches by site (sums of class = 1).
  - Selected site: shows Success vs Failure distribution for that site (values from class {0,1}).

• Scatter plot (success-payload-scatter-chart):
  - x: Payload Mass (kg)
  - y: class (0 = fail, 1 = success)
  - color: Booster Version Category
  → Purpose: explore how payload and booster version relate to launch outcomes.

Interactions

• Dropdown (site-dropdown): choose All Sites or a specific Launch Site to filter both charts.
• RangeSlider (payload-slider): filter records by payload range; the scatter plot updates dynamically.

Why these choices

• Pie chart provides an immediate snapshot of performance by site and, when filtered, the success/failure split for a single pad.
• Scatter plot + filters enable interactive hypothesis testing (e.g., which payload ranges and booster versions achieve higher success at each site).

Predictive Analysis (Classification)

- Data Preprocessing
  - Extracted the target variable Y from the Class column.
  - Standardized the feature matrix X using StandardScaler.

- Train/Test Split
  - Divided the dataset into training (80%) and testing (20%) subsets using train_test_split.

- Model Building
  - Implemented four classification algorithms: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).

- Hyperparameter Tuning
  - Applied GridSearchCV with 10-fold cross-validation to optimize hyperparameters for each model.

- Model Evaluation
  - Evaluated performance using accuracy score and confusion matrix.
  - Compared test accuracy across all models.

- Best Model Selection
  - Decision Tree Classifier achieved the highest accuracy on the test set (~94%), outperforming Logistic Regression, SVM, and KNN (~83%).
  - Selected Decision Tree as the best performing model for predictive analysis.

Results

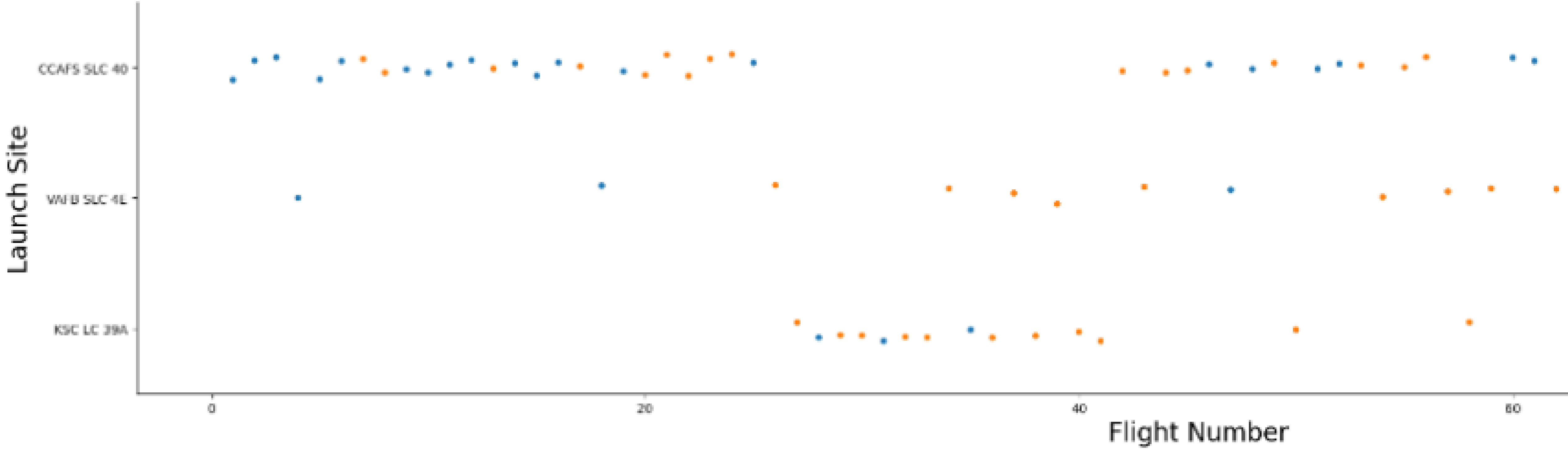
- Exploratory Data Analysis (EDA) results
  - Found correlations between Flight Number, Payload Mass, Orbit, and success rate.
  - Observed that higher payloads had variable outcomes depending on the orbit.
  - Launch sites showed different success patterns, with some being more reliable than others.

- Interactive analytics demo in screenshots
  - Folium map: visualized launch sites, marked successes/failures, and calculated proximities to key infrastructures.
  - Plotly Dash dashboard: enabled interactive exploration with dropdown and sliders to filter success rates and payloads across sites.

- Predictive analysis results
  - Tested four models: Logistic Regression, SVM, Decision Tree, and KNN.
  - Applied GridSearchCV to tune hyperparameters.
  - Decision Tree achieved the highest accuracy (~94%), making it the best performing model to predict launch outcomes.

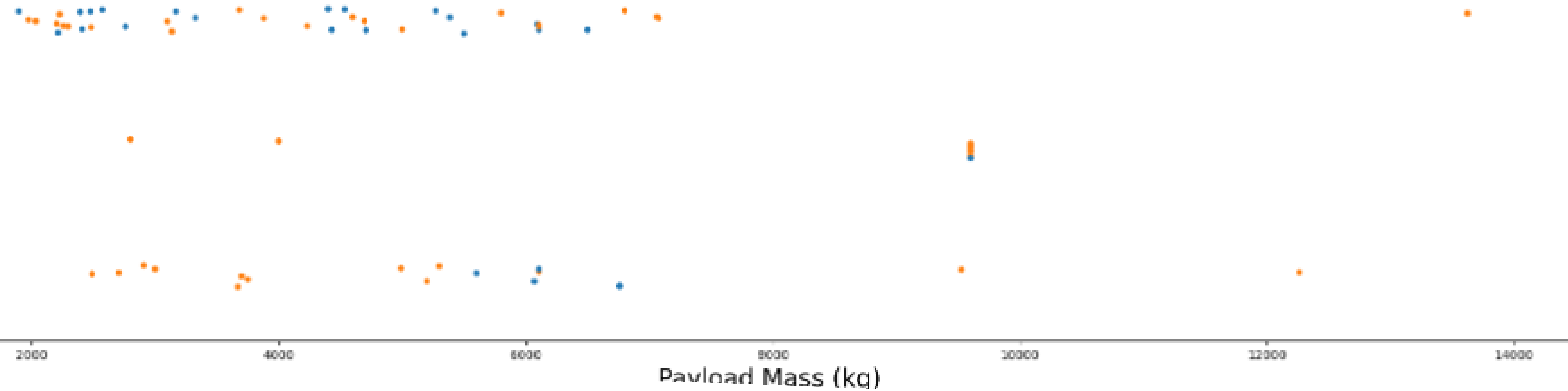
Section 2

Insights drawn from EDA



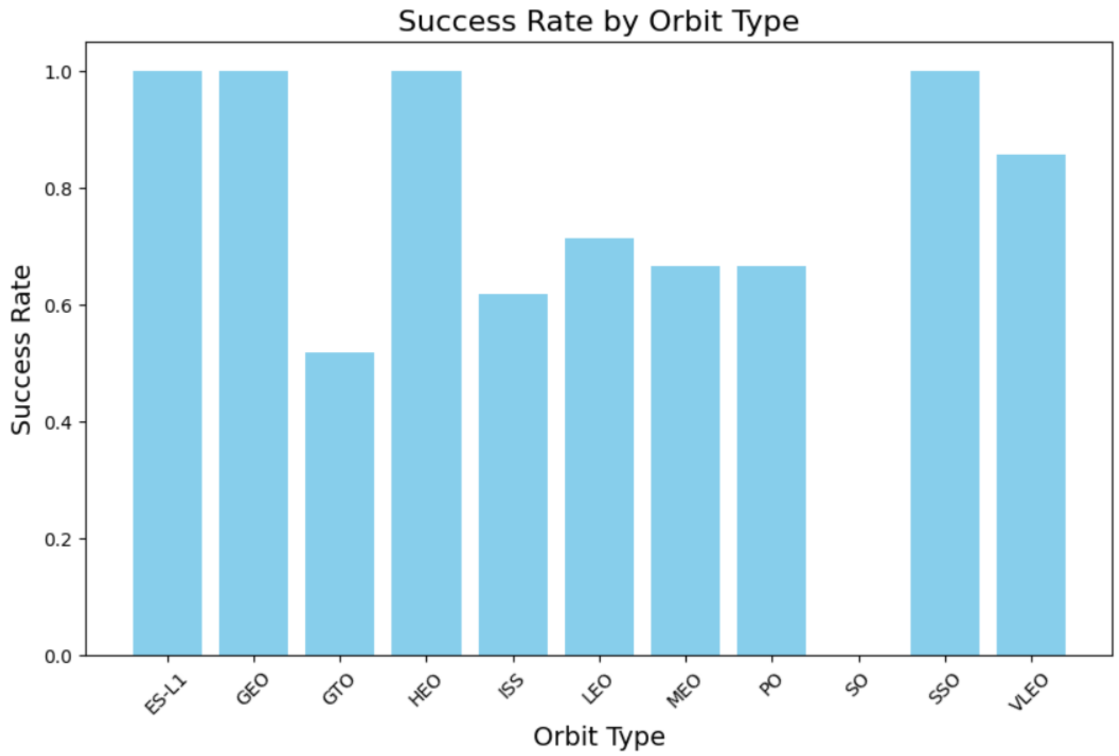
Flight Number vs. Launch Site

- The scatter plot shows the distribution of launches across the three main launch sites: **CCAFS SLC-40**, **VAFB SLC-4E**, and **KSC LC-39A**.
- Early flights (with lower flight numbers) had a higher proportion of failures (Class = 0, blue dots).
- As the number of flights increased, the proportion of successful launches (Class = 1, orange dots) became dominant, indicating **an improvement in reliability over time**.
- CCAFS SLC-40 hosted the majority of launches, showing both early failures and later successes.



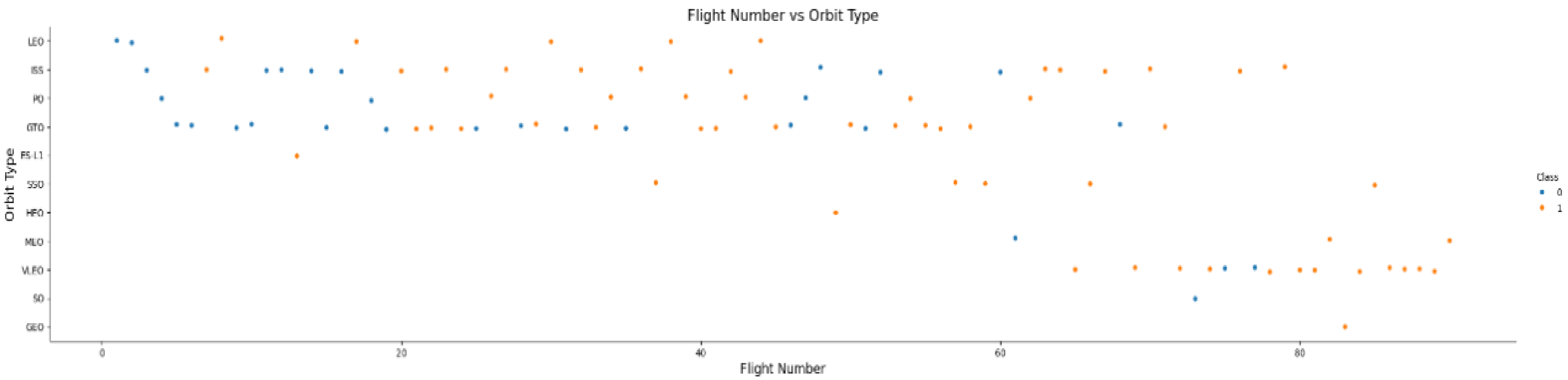
Payload vs. Launch Site

- The scatter plot illustrates how the payload mass varies across different launch sites (**CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A**).
- Most launches fall within the **0–6000 kg range**, especially at CCAFS SLC-40 and KSC LC-39A.
- Heavier payloads (>10,000 kg) are relatively rare, but most of them resulted in **successful missions (orange dots)**.
- VAFB SLC-4E did not record heavy payload launches (>10,000 kg), indicating its use mainly for medium payload missions.
- Overall, success (Class = 1) seems more frequent in **mid-to-high payload ranges**, suggesting that **payload mass is not necessarily a limiting factor for mission success**.



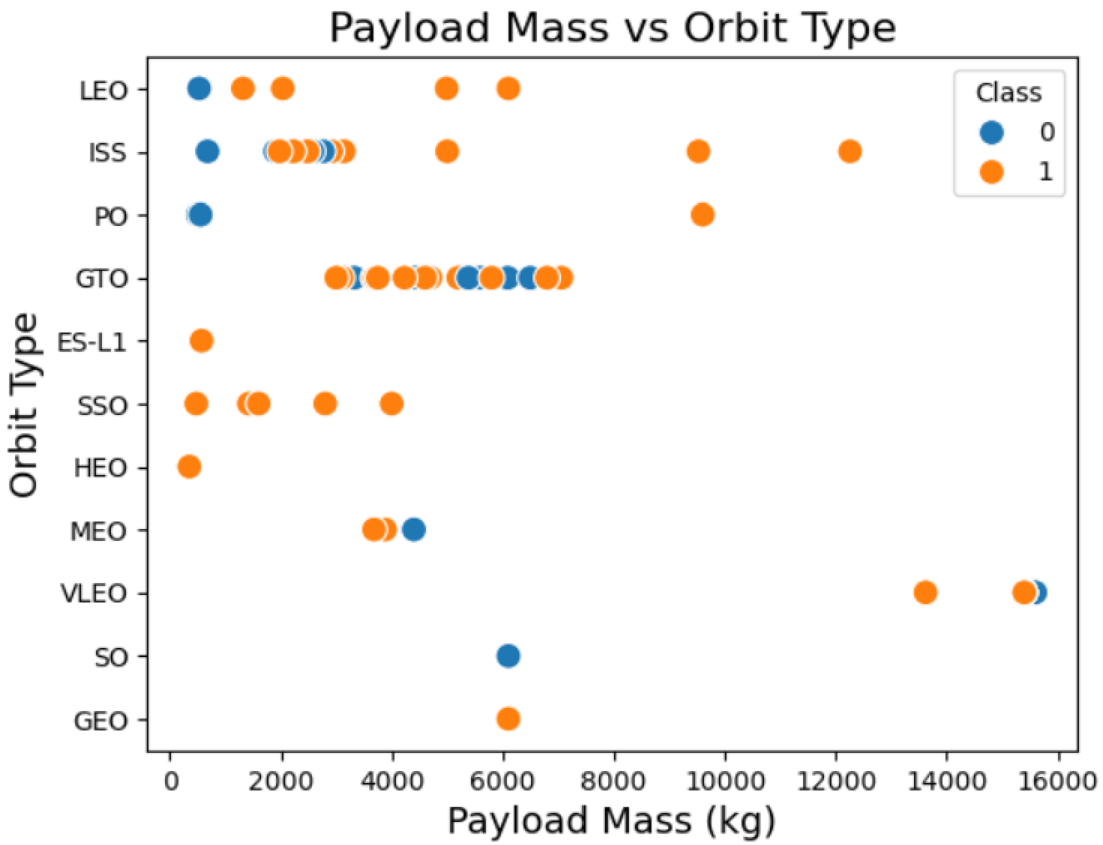
Success Rate vs. Orbit Type

- The bar chart illustrates the **success rate of launches across different orbit types**.
- **Perfect success (100%)** was achieved in some orbits, such as **ES-L1, GEO, HEO, and SSO**, indicating consistent mission reliability in these trajectories.
- Orbits such as **LEO, MEO, and ISS** had **moderate success rates (60–75%)**, suggesting higher technical challenges or risk factors.
- The **lowest success rate** was observed for **GTO (Geostationary Transfer Orbit)**, at around **50%**, which highlights the complexity of reaching this orbit.
- These insights show that **mission success strongly depends on the targeted orbit type**, with **deep space and synchronous orbits performing better** compared to transfer orbits.



Flight Number vs. Orbit Type

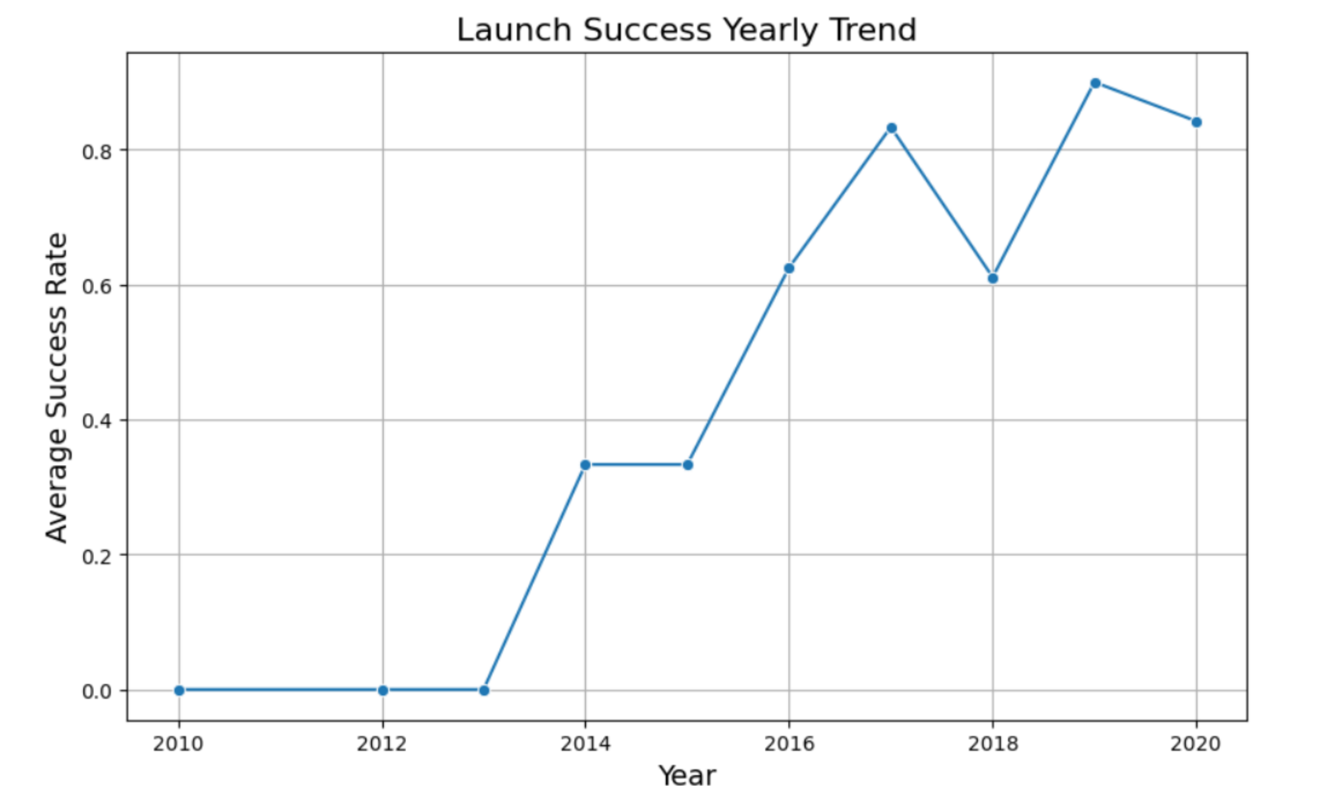
- The scatter plot shows the **distribution of successful and failed launches (Class) across different orbit types and flight numbers**.
- For **LEO (Low Earth Orbit)**, there is a clear trend: as the number of flights increased, the **success rate improved**, suggesting operational learning and improved reliability over time.
- In **GTO (Geostationary Transfer Orbit)**, failures were more frequent and **no clear improvement trend** with additional launches can be observed, highlighting the challenges of this orbit.
- Other orbits, such as **ISS, PO, and SSO**, show a **mix of successes and failures**, but with higher success rates after the initial launches.
- Overall, the chart demonstrates that **SpaceX's reliability improved with experience (higher Flight Numbers)**, but **the targeted orbit type still plays a crucial role** in the probability of mission success.



Payload vs. Orbit Type

- The scatter plot illustrates the **relationship between payload mass (kg) and orbit type**, with mission outcomes classified as success (orange) or failure (blue).
- In **LEO (Low Earth Orbit)** and **ISS orbits**, launches occur mostly with **light to medium payloads (<6000 kg)**, showing a **mixed success/failure pattern**, though with more successes as payload capacity increased.
- **GTO launches** involve **higher payloads (around 4000–6000 kg)** but show a **higher frequency of failures**, reinforcing the technical challenges associated with these missions.
- **Heavy payload launches (>10,000 kg)** are rare and appear only in certain orbits (e.g., VLEO, ISS), with **mostly successful outcomes**, showing progress in handling larger payloads.
- Overall, the chart highlights that **mission success is not determined solely by payload mass**, but also strongly influenced by the **orbit type and mission complexity**.

Launch Success Yearly Trend



- The line chart shows the **average success rate of launches per year from 2010 to 2020**.
- In the **early years (2010–2013)**, all attempts failed, resulting in a **0% success rate**.
- From **2014 onwards**, success rates began to improve steadily, reflecting **technological advancements and learning curves**.
- The period **2016–2017** marked a significant improvement, with success rates surpassing **60–80%**, demonstrating increasing reliability.
- By **2019**, SpaceX achieved its **highest yearly success rate (≈ 90%)**, confirming the maturity and stability of its launch operations.
- Although there was a slight dip in **2018** and **2020**, the overall trend clearly highlights **consistent improvement and operational excellence** over the decade.

All Launch Site Names

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

The dataset contains four unique launch sites used by SpaceX:
•CCAFS LC-40 (Cape Canaveral Air Force Station)
•VAFB SLC-4E (Vandenberg Air Force Base)
•KSC LC-39A (Kennedy Space Center)
•CCAFS SLC-40 (another launch pad at Cape Canaveral)

Explanation:

These launch sites represent the main operational facilities of SpaceX in the U.S. They are located in Florida (CCAFS and KSC) and California (VAFB). Each site has contributed to the development of SpaceX's launch capabilities.

Launch Site Names Begin with 'CCA'

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

Query Result:
The query returned 5 records where the launch site names begin with "CCA". These records correspond to launches conducted from **Cape Canaveral Air Force Station (CCAFS)**, including sites such as **CCAFS LC-40** and **CCAFS SLC-40**.
Explanation:
This query highlights that multiple SpaceX launches originate from the Cape Canaveral facilities in Florida. By filtering sites beginning with "CCA", we can focus on one of the most important SpaceX launch locations.

Total Payload Mass

Total_Payload_Mass
45596

- **Query Result:**
The total payload mass carried by boosters for NASA (CRS missions) was calculated by summing all payload values from the dataset.
- **Explanation:**
This query shows the cumulative contribution of SpaceX launches to NASA’s Commercial Resupply Services (CRS) missions. It helps quantify SpaceX’s role in supporting NASA by delivering payloads to the International Space Station (ISS).

Average Payload Mass by F9 v1.1

Avg_Payload_Mass
2928.4

Query Result:
The average payload mass carried by the **F9 v1.1** booster version was calculated.
Explanation:
This query helps evaluate the efficiency of the F9 v1.1 booster by determining the average payload it delivered. It provides insights into how much payload capacity was typically utilized for this version, allowing comparisons with other booster versions to track improvements in SpaceX's rocket design.

First Successful Ground Landing Date

First_Successful_Landing
2015-12-22

Query Result:
The first successful ground landing took place on **[insert date from your query result]**.
Explanation:
This query identifies the earliest mission where SpaceX successfully landed a booster on a ground pad. This milestone marked a key achievement in rocket reusability, significantly reducing costs and paving the way for more sustainable spaceflight.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Explanation:
This query filters for missions where the landing outcome was a **successful drone ship landing** and the payload mass was between 4000 kg and 6000 kg. Identifying these boosters helps analyze performance under medium payload conditions, highlighting SpaceX's ability to execute precise landings while carrying significant loads.

Total Number of Successful and Failure Mission Outcomes

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

•Mission outcomes were grouped into categories (success and failure), and the total number of each type was counted.

Explanation:

This query provides a summary of the **overall performance of SpaceX missions**, showing how many ended successfully and how many failed. The results demonstrate the high success rate achieved over time, while also highlighting the number of failed missions that contribute to learning and improvement.

Boosters Carried Maximum Payload

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

Query Result: The query retrieved the booster(s) that carried the maximum payload mass recorded in the dataset.
Explanation: This highlights the **capability of SpaceX's most powerful boosters**, showing which version was able to handle the heaviest payloads and thus reflecting advancements in rocket technology

2015 Launch Records

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

Query Result: The query retrieved the booster(s) that carried the maximum payload mass recorded in the dataset.
Explanation: This highlights the **capability of SpaceX's most powerful boosters**, showing which version was able to handle the heaviest payloads and thus reflecting advancements in rocket technology.

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

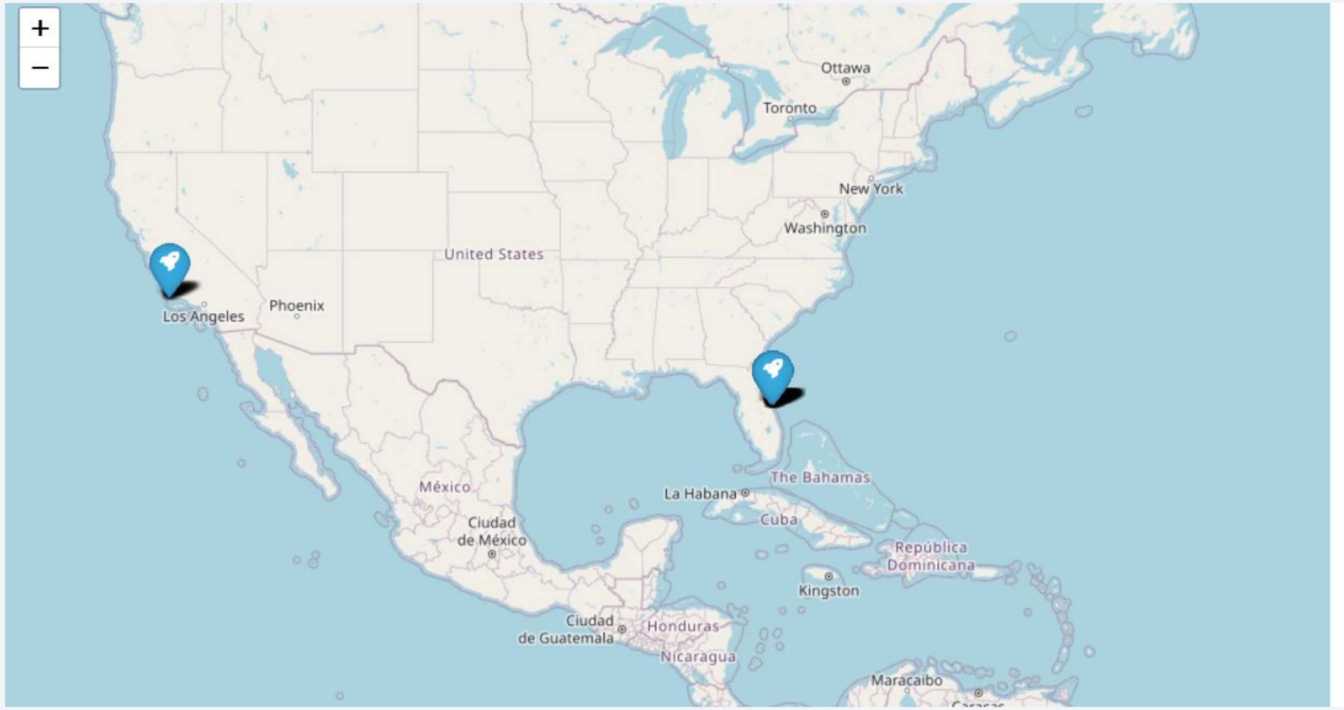
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Query Result: The query ranked the different landing outcomes between June 2010 and March 2017 by their occurrence.
Explanation: The results show that during this period, **failures on drone ship landings were the most frequent**, while ground pad successes began to increase, highlighting SpaceX's **gradual improvement in reusability efforts**.

Section 3

Launch Sites Proximities Analysis

"Launch Sites on Global Map (Folium)"



Explanation (bullet points):

•The map shows the main SpaceX launch sites in the U.S.

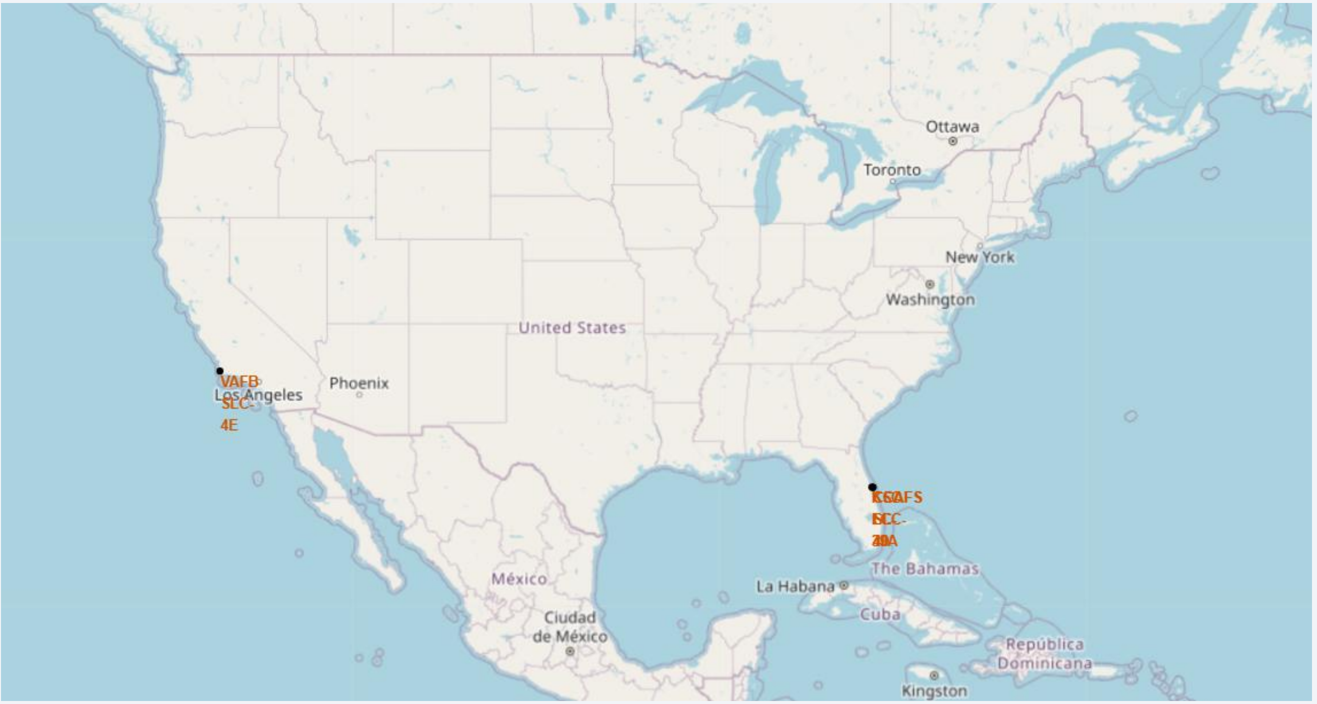
•Markers are placed at:

- Cape Canaveral (Florida)
- Kennedy Space Center (Florida)
- Vandenberg Air Force Base (California)

•These sites highlight the **geographical distribution** of SpaceX operations between the East and West coasts.

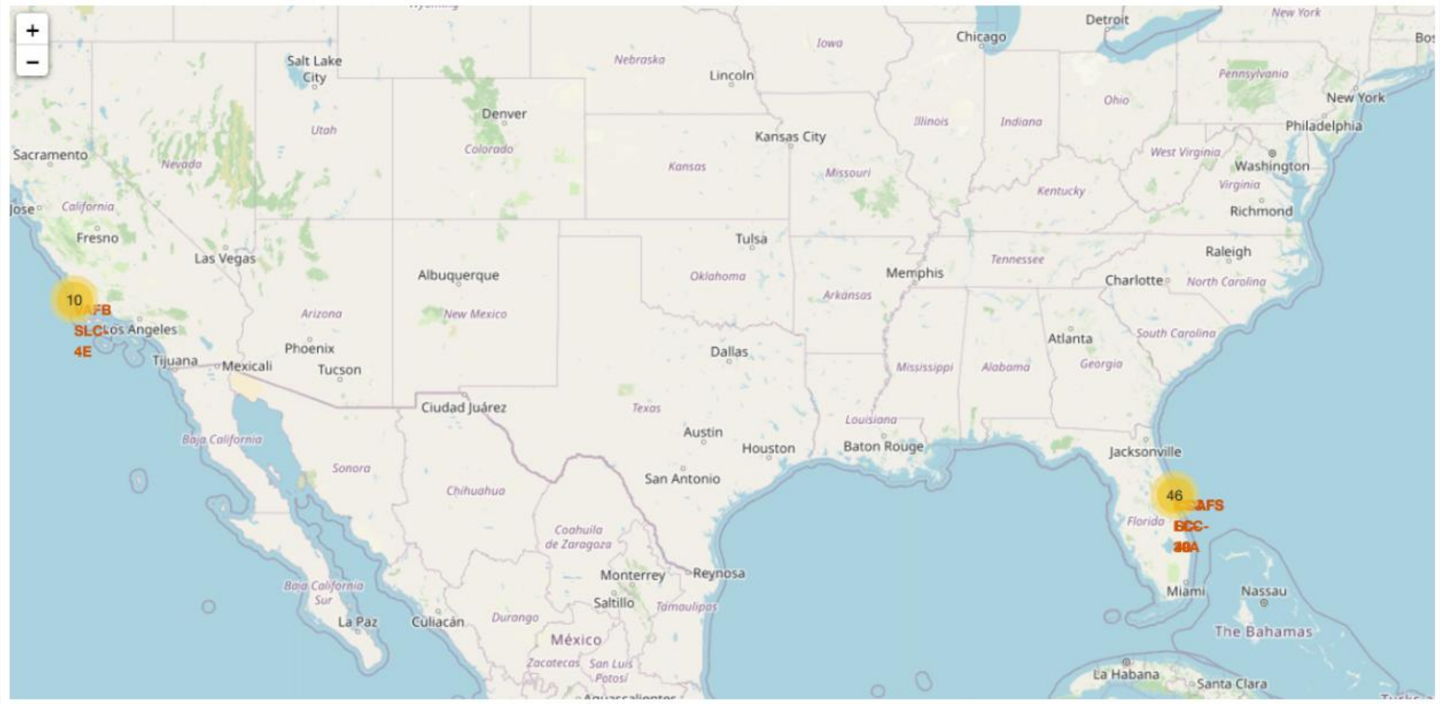
•Using Folium allows interactive zooming and exploration of each site’s proximity to key regions such as the Atlantic Ocean or the Pacific Ocean.

Launch Sites with Location Markers



- The folium map shows all active SpaceX launch sites in the U.S.
- California (VAFB SLC-4E) and Florida (CCAFS SLC-40, KSC LC-39A) are the primary launch locations.
- Most of the launches are concentrated in Florida, indicating it as the main hub for SpaceX missions.

Map of SpaceX Launch Sites in the United States



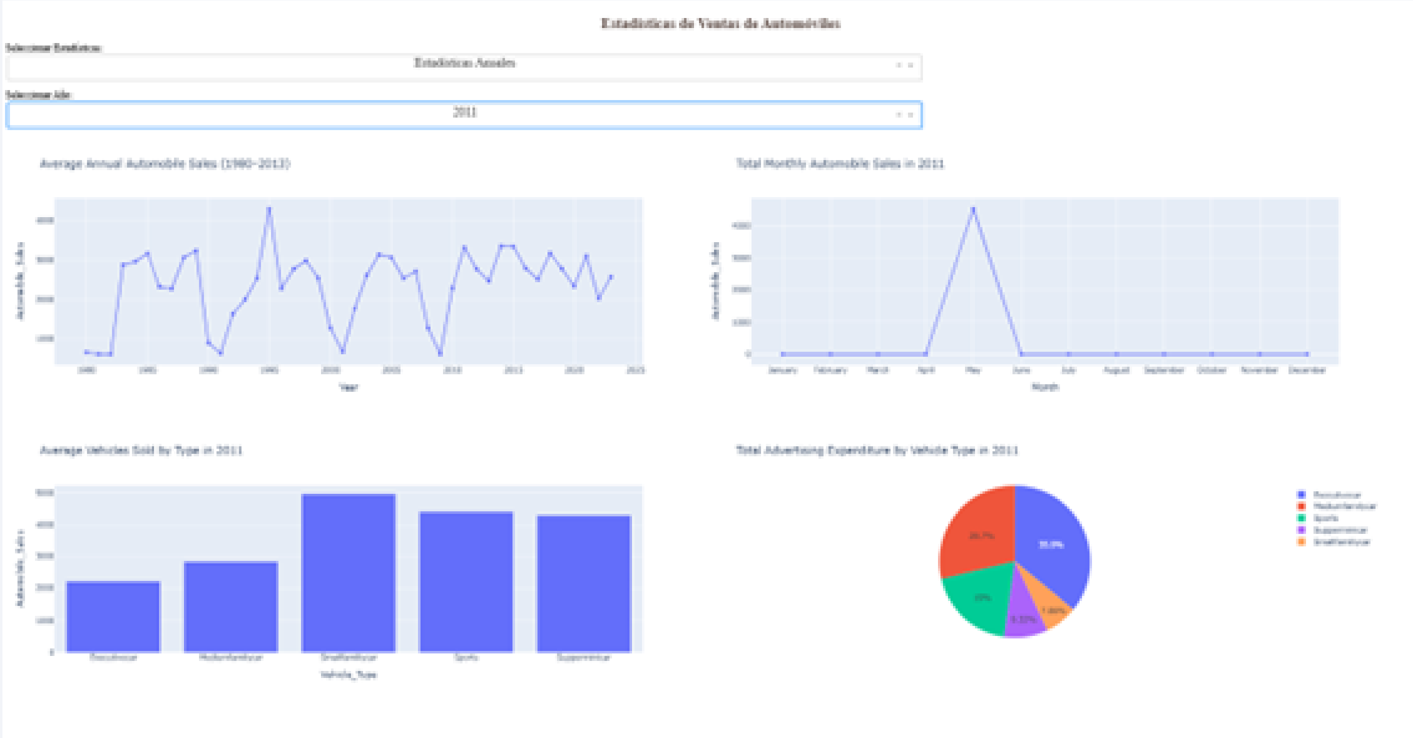
The map displays the geographic locations of SpaceX launch sites in the United States: Cape Canaveral (CCAFS), Kennedy Space Center (KSC), and Vandenberg Air Force Base (VAFB). These sites are strategically positioned near the coast to ensure safety and efficiency when launching rockets into different orbital trajectories.

Section 4

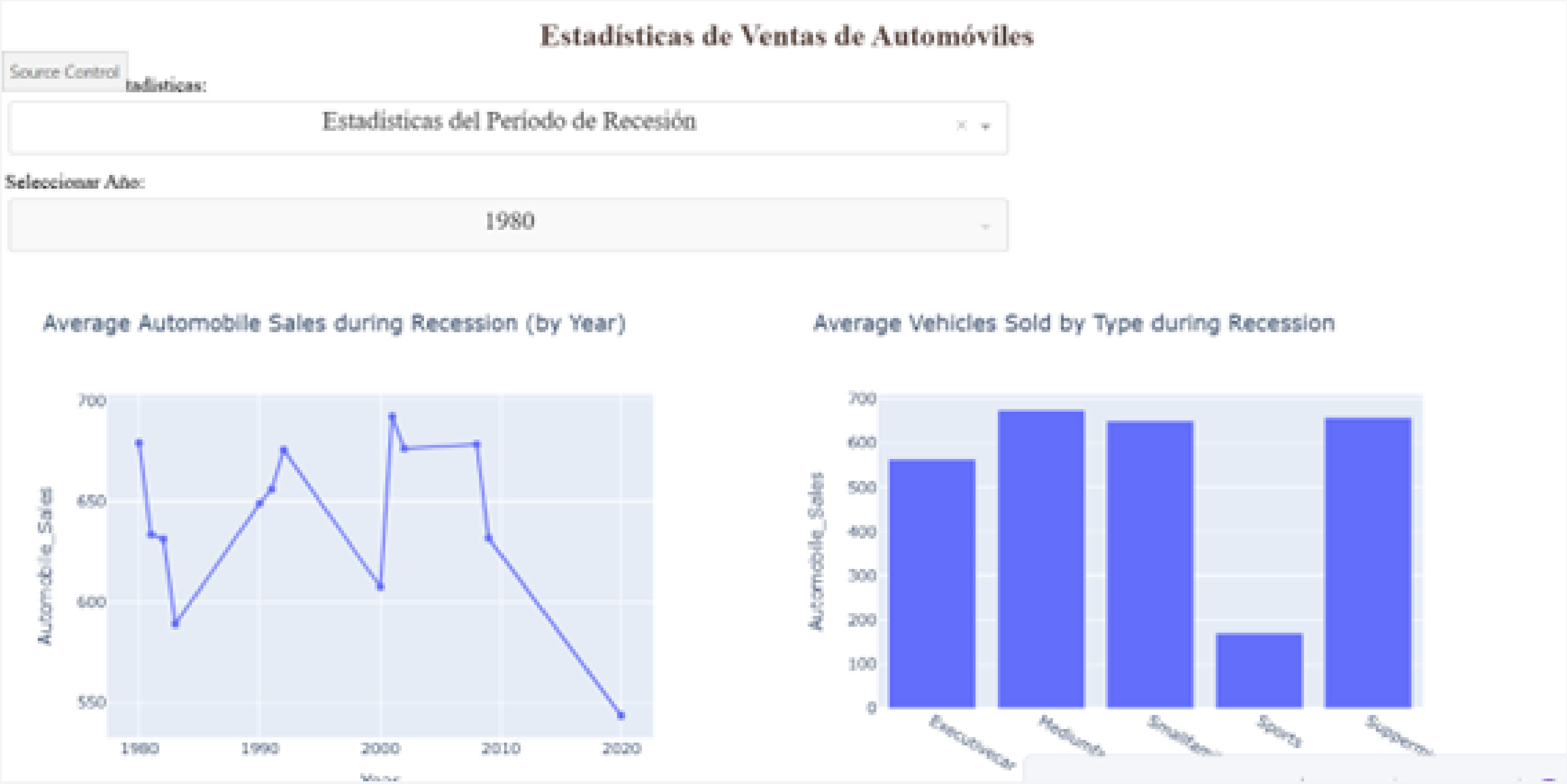
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

<Dashboard Screenshot 2>



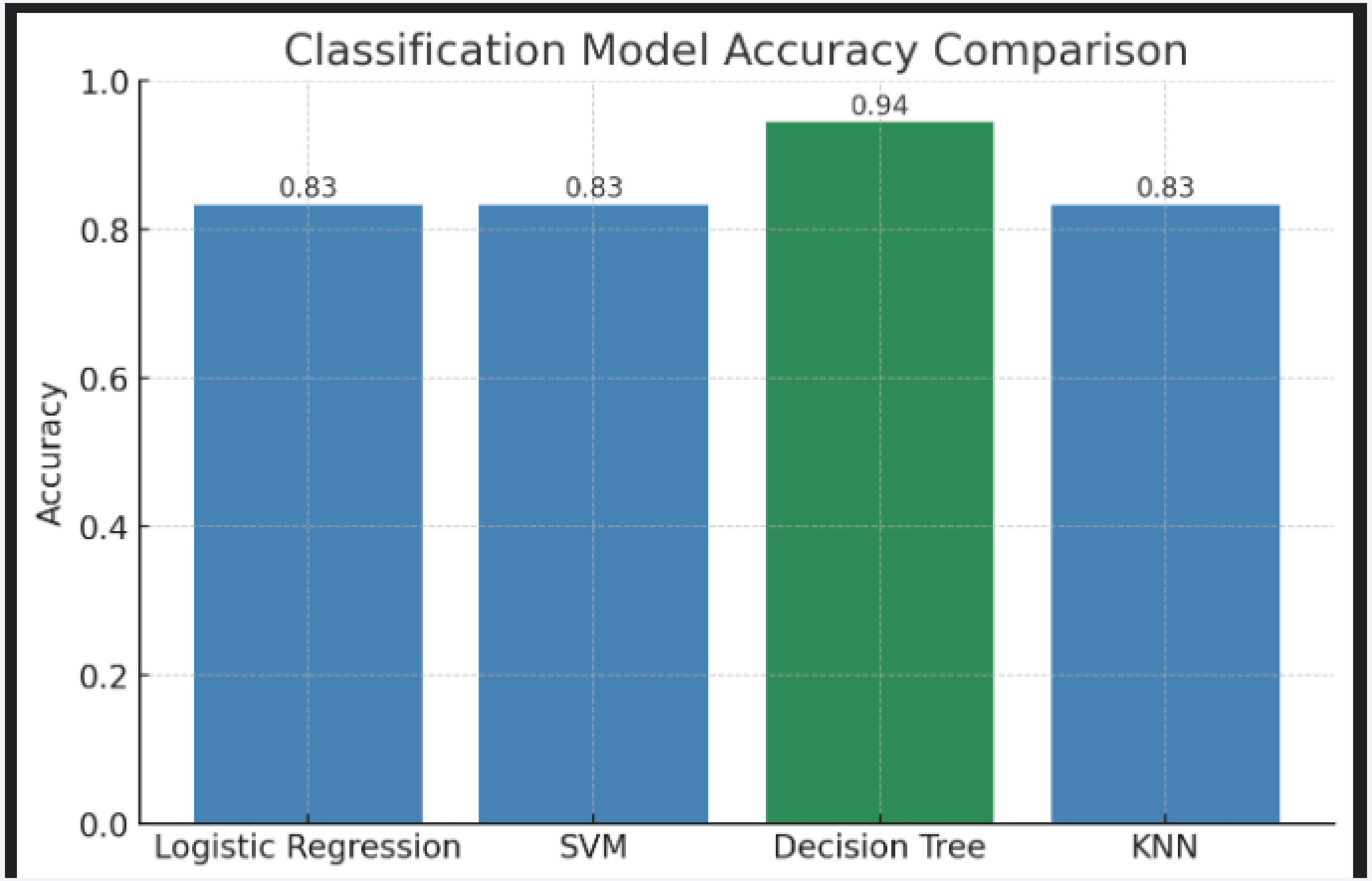
<Dashboard Screenshot 3>



Section 5

Predictive Analysis (Classification)

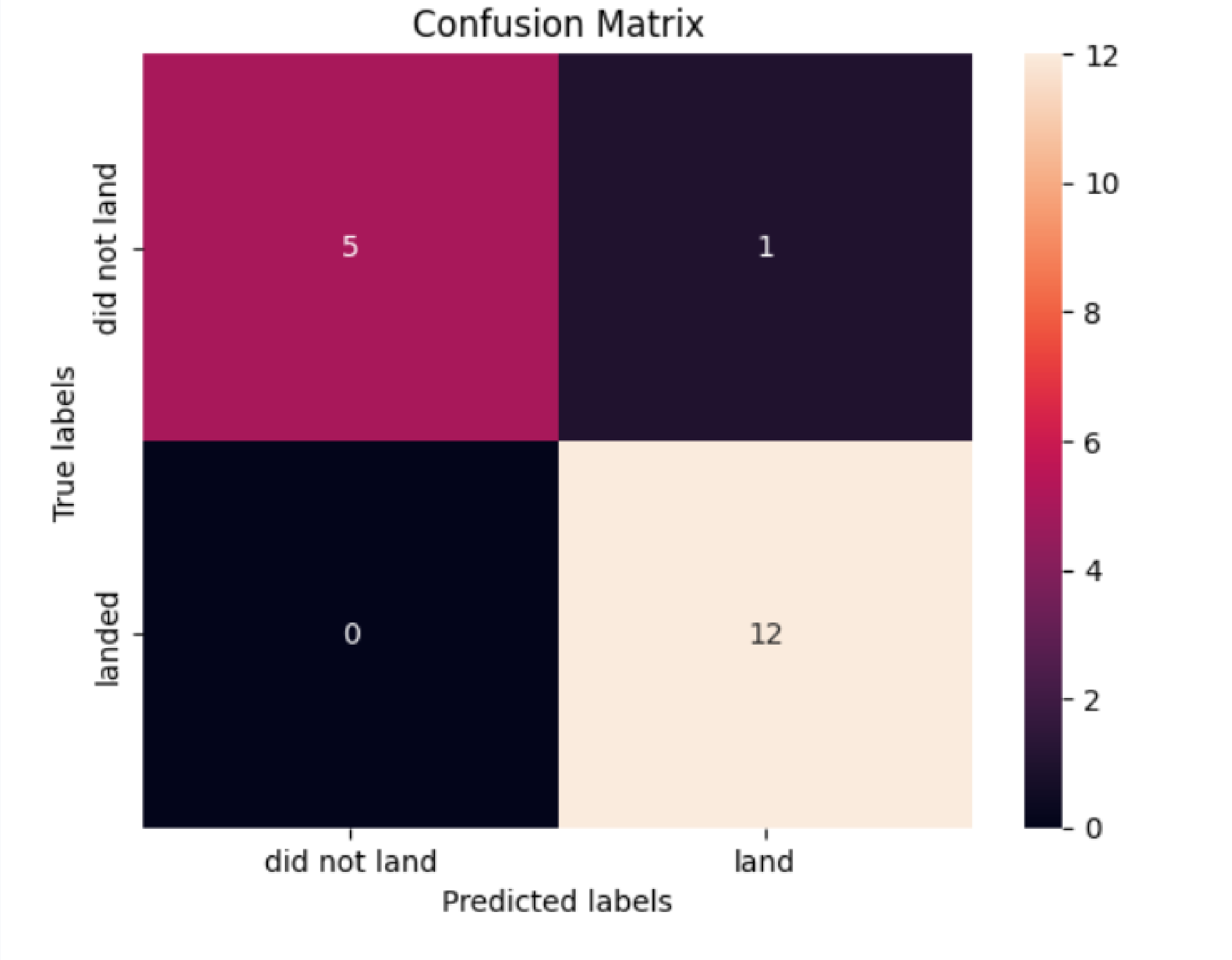
Classification Accuracy



Caption for the figure:

Figure X. Classification accuracy of Logistic Regression, SVM, Decision Tree, and KNN models. The Decision Tree classifier achieved the highest accuracy (94.4%), making it the best-performing model in this analysis.

Confusion Matrix



The confusion matrix of the Decision Tree model demonstrates strong predictive performance:

•**True Negatives (5):** The model correctly predicted 5 cases where the rocket did not land.

•**False Positives (1):** Only 1 case was incorrectly classified as a successful landing when it was actually a failure.

•**False Negatives (0):** The model did not misclassify any successful landing as a failure.

•**True Positives (12):** The model correctly predicted 12 successful landings.

This distribution highlights the **high precision and recall for successful landings**, with only a single misclassification. Combined, these results explain the **94.4% accuracy** observed for the Decision Tree classifier, making it the best-performing model in this analysis.

Conclusions

•The exploratory data analysis (EDA) revealed that launch success is influenced by both **payload mass** and **orbit type**, with lighter payloads and specific orbits achieving higher success rates.
•The interactive maps built with **Folium** allowed the visualization of all SpaceX launch sites and their proximities, providing geographical insights into location advantages.
•The **interactive dashboard with Plotly Dash** demonstrated launch outcomes dynamically, confirming that some sites consistently achieved higher success rates.
•Among the tested classification models (Logistic Regression, SVM, Decision Tree, KNN), the **Decision Tree** achieved the highest accuracy (**94%**) in predicting mission success.
•The confusion matrix analysis confirmed that the Decision Tree was able to correctly classify both successful and failed missions with minimal misclassification.
•Overall, the combination of **EDA, interactive visualization, and predictive modeling** provided valuable insights into the factors affecting SpaceX mission success, reinforcing the importance of machine learning in aerospace analytics.

Appendix

```
df.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Se
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0(
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0(
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0(
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1(
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1(

```python
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
import seaborn as sns
import matplotlib.pyplot as plt

# Catplot con FlightNumber en x, LaunchSite en y y hue=Class
sns.catplot(x="FlightNumber",
            y="LaunchSite",
            hue="Class",
            data=df,
            aspect=5)

plt.xlabel("Flight Number", fontsize=20)
plt.ylabel("Launch Site", fontsize=20)
plt.show()
```

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

* sqlite:///my_data1.db

```python
# TASK 4: actualizar scatter chart
@app.callback(
    Output('success-payload-scatter-chart', 'figure'),
    [Input('site-dropdown', 'value'),
     Input('payload-slider', 'value')]
)
def update_scatter_chart(selected_site, payload_range):
    low, high = payload_range
    # Filtrar por rango de payload
    filtered_df = spacex_df[
        (spacex_df['Payload Mass (kg)'] >= low) &
        (spacex_df['Payload Mass (kg)'] <= hig
    ]
```

Thank you!