IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

MUSKAAN RAHEJA
18 MAY, 2025

Outline

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

Executive Summary

This project investigates SpaceX launch data to identify factors affecting launch success rates.

Data was collected using the SpaceX REST API and web scraping from Wikipedia.

After cleaning and merging, we performed EDA using visualizations and SQL, then built interactive dashboards (Plotly Dash) and maps (Folium).

The final stage involved developing classification models to predict successful launches based on payload, orbit, booster type, and site.

Key outcomes:

- KSC LC-39A has the highest launch success rate
- Payloads in the 4000–6000 kg range have the most success
- F9 B5 boosters delivered the heaviest payloads

Introduction

SpaceX has revolutionized commercial spaceflight through innovation in reusability and launch efficiency.

This project aims to analyze historical launch data to uncover:

- Which factors impact launch success
- Which sites and boosters perform best
- What payload ranges are most reliable

These insights can guide future mission planning, resource allocation, and launch optimization strategies.

Section 1

Methodology

Methodology

Data Collection:

- Used **SpaceX API** for structured JSON launch data (rockets, payloads, launchpads)
- Performed **web scraping** for Falcon 9 mission tables from Wikipedia

Data Wrangling:

- Merged API and scraped datasets on key fields
- Cleaned missing/inconsistent values and standardized formats

EDA:

- Conducted visual analysis (scatter, bar, line charts) using Matplotlib & Seaborn
- Executed SQL queries to extract deeper insights from structured data

Interactive Visuals:

- Created Folium maps for site proximity and outcomes
- Built Dash dashboard for payload/site-specific launch filtering

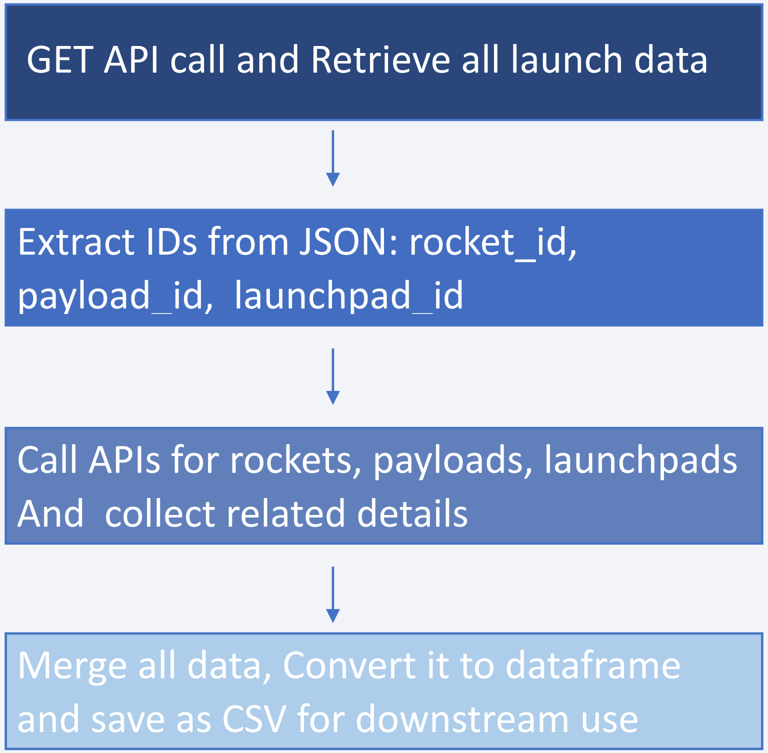
Predictive Analysis:

- Trained classification models (e.g., Logistic Regression, SVM) to predict mission success

Data Collection – SpaceX API

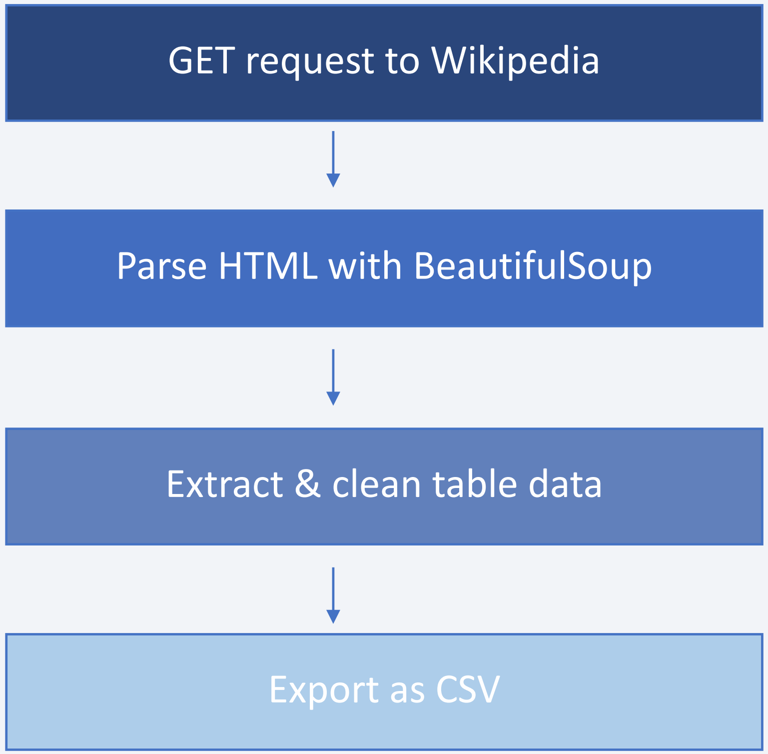
- Collected data via SpaceX REST API (https://api.spacexdata.com/v4/launches)
- Used Python, requests, and pandas for HTTP calls and parsing
- Retrieved JSON data and normalized nested structures (rockets, payloads, launchpads)
- Extracted key fields: flight_number, launch_site, payload_mass_kg, orbit, booster_version, landing_success
- Stored final dataset in a pandas DataFrame
- Exported to CSV for downstream analysis

🔗 **GitHub Notebook:**
[GitHub Notebook: SpaceX API Data Collection]



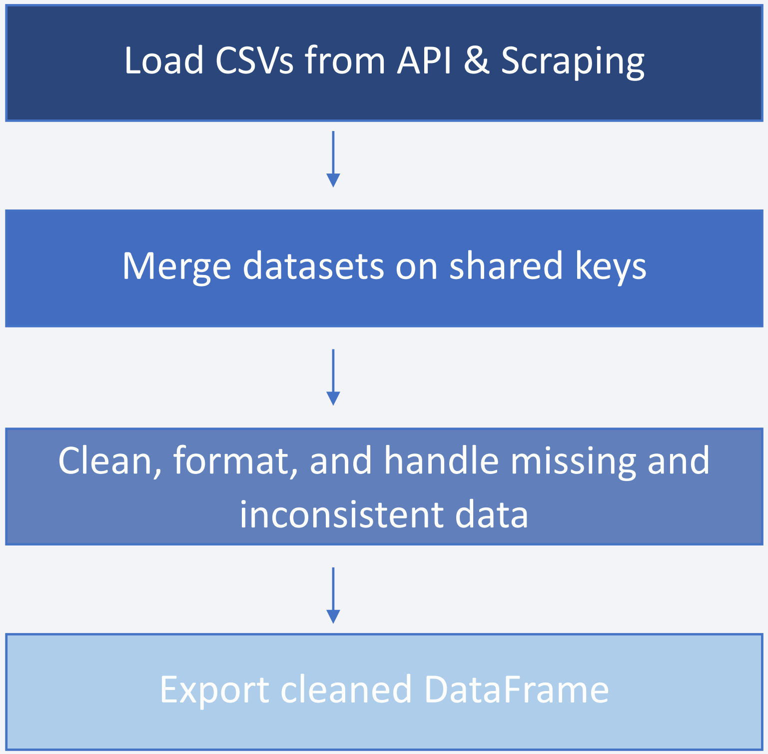
Data Collection - Scraping

- Target site: Wikipedia – Falcon 9 Launches
- Tools used: requests, BeautifulSoup, pandas
- Parsed HTML content and identified launch table
- Extracted key columns: Launch Date, Booster Version, Launch Site, Payload Mass, Landing Outcome
- Structured into a clean pandas DataFrame and exported to CSV
- 🔗 **GitHub Notebook Link:** Web Scraping Notebook – GitHub

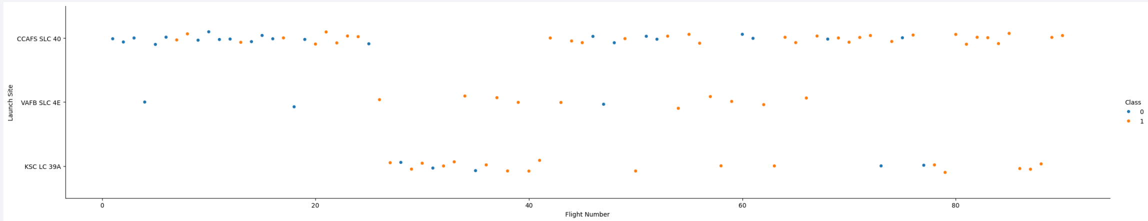


Data Wrangling

- Loaded datasets from:
  - spacex_api_data.csv (collected via SpaceX API)
  - spacex_web_scraped_data.csv (scraped from Wikipedia)
- Performed **data merging** based on common fields: Flight Number, Launch Site, and Booster Version
- **Data cleaning steps**:
  - Removed duplicates and irrelevant columns
  - Converted data types (e.g., Payload Mass (kg) to float, dates to datetime)
  - Identified and filled or dropped missing values
  - Standardized text values (e.g., unified landing outcomes)
- **Final output**: A clean and structured dataset saved as spacex_data_wrangled.csv for analysis and modeling
- 🔗 **GitHub Notebook (for Peer Review):** **View Data Wrangling Notebook on GitHub**

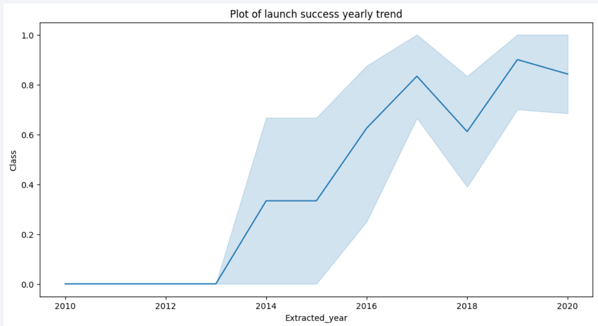


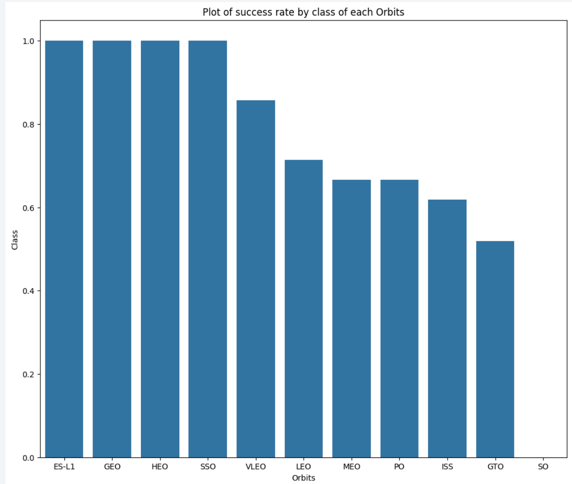
EDA with Data Visualization



📊 Charts Plotted and Purpose

1. **Flight Number vs. Launch Site (Scatter Plot)**: To observe the distribution of launches across sites over time
2. **Payload Mass (kg) vs. Launch Site (Scatter Plot)**: Visualizes how payload weight varies across different launch pads
3. **Success Rate vs. Orbit Type (Bar Chart):** Compares the launch success ratio across different orbit types
4. **Flight Number vs. Orbit Type (Scatter Plot):** Shows temporal patterns in orbit deployments
5. **Payload Mass vs. Orbit Type (Scatter Plot)**: Analyzes which orbit types handle heavier payloads
6. **Yearly Launch Success Trend (Line Chart)**: Highlights performance improvements over time.





🔗 GitHub Notebook Link

[EDA with Data Visualization – GitHub]

EDA with SQL

🔷 **Key SQL Insights:**

- Retrieved **unique launch sites** and filtered those starting with 'CCA'
- Calculated **total payload mass** for NASA missions
- Found **average payload** for booster F9 v1.1
- Identified **first successful ground landing date**
- Listed **boosters with 4000–6000 kg** payloads and successful drone ship landings
- Counted **success vs. failure** mission outcomes
- Found **maximum payload booster**
- Analyzed **failed drone ship landings in 2015**
- Ranked **landing outcomes (2010–2017)** by frequency

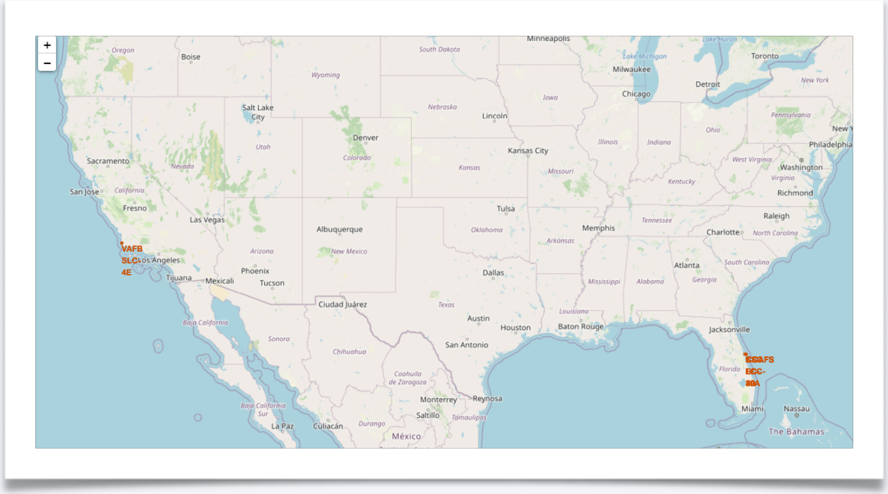
🔗 **GitHub Notebook:**

[EDA with SQL – GitHub]

Build an Interactive Map with Folium

🔷 Map Objects Used

- Markers: Plotted each launch site on the map
- Popups: Displayed launch site names
- Circle Markers: Showed impact zones around sites
- Lines: Connected launch sites to nearby features (e.g., coastlines)
- Color Coding: Indicated launch success vs. failure



🔗 GitHub Notebook

[Folium Map – GitHub]

Build a Dashboard with Plotly Dash

🔷 **Plots & Interactions Added**

- **Dropdown Menu:** To filter data by individual launch sites
- **Pie Chart:** Shows total successful launches (overall and per site)➝Helps identify most active and reliable sites
- **Payload Range Slider:** Allows dynamic filtering of data by payload mass (0–10,000 kg)➝Reveals success trends by payload range
- **Scatter Plot:** Displays Payload Mass vs. Launch Outcome, color-coded by Booster Version➝Highlights correlation between payload size and mission success

🔶 **Insights from Dashboard Largest number of successful launches: 🛰️ KSC LC-39A**

- Highest launch success rate: 🛰️ KSC LC-39A (76.9%)
- Highest success rate payload range: 📦 4000–6000 kg
- Lowest success rate payload range: 📦 0–2000 kg
- Top performing booster version: 🚀 B5 (Block 5)

🔗 **GitHub App Link**

📎 Plotly Dash App – GitHub

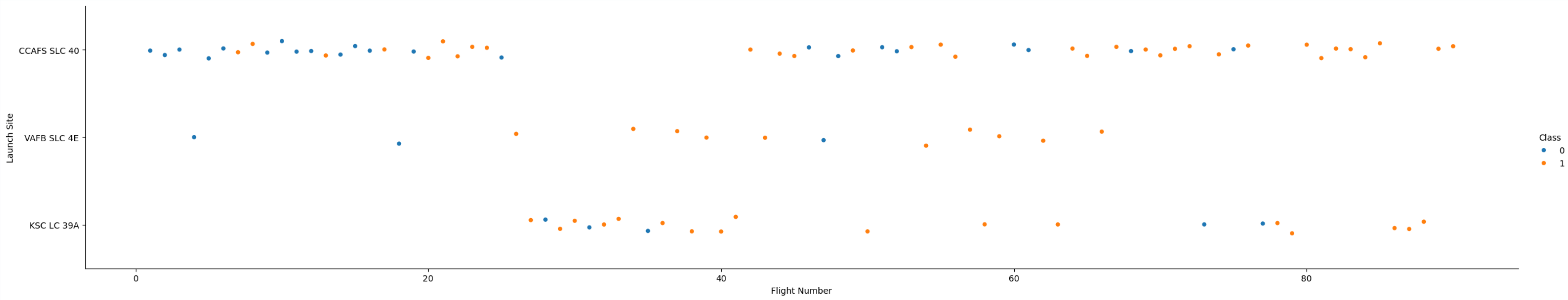
Results

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

Section 2

Insights drawn from EDA

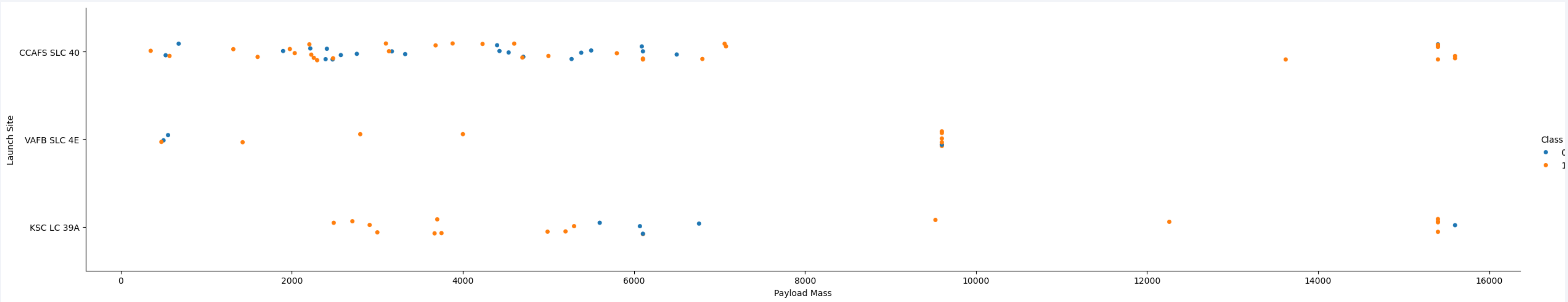
Flight Number vs. Launch Site



📊 Scatter Plot Insight: Flight Number vs. Launch Site

- **Success rate increases** with higher flight numbers, indicating improved reliability over time.
- **KSC LC 39A** shows a consistently high success rate in later missions.
- **CCAFS SLC 40** has both early and recent missions with mixed outcomes, reflecting varied performance across its timeline.

Payload vs. Launch Site



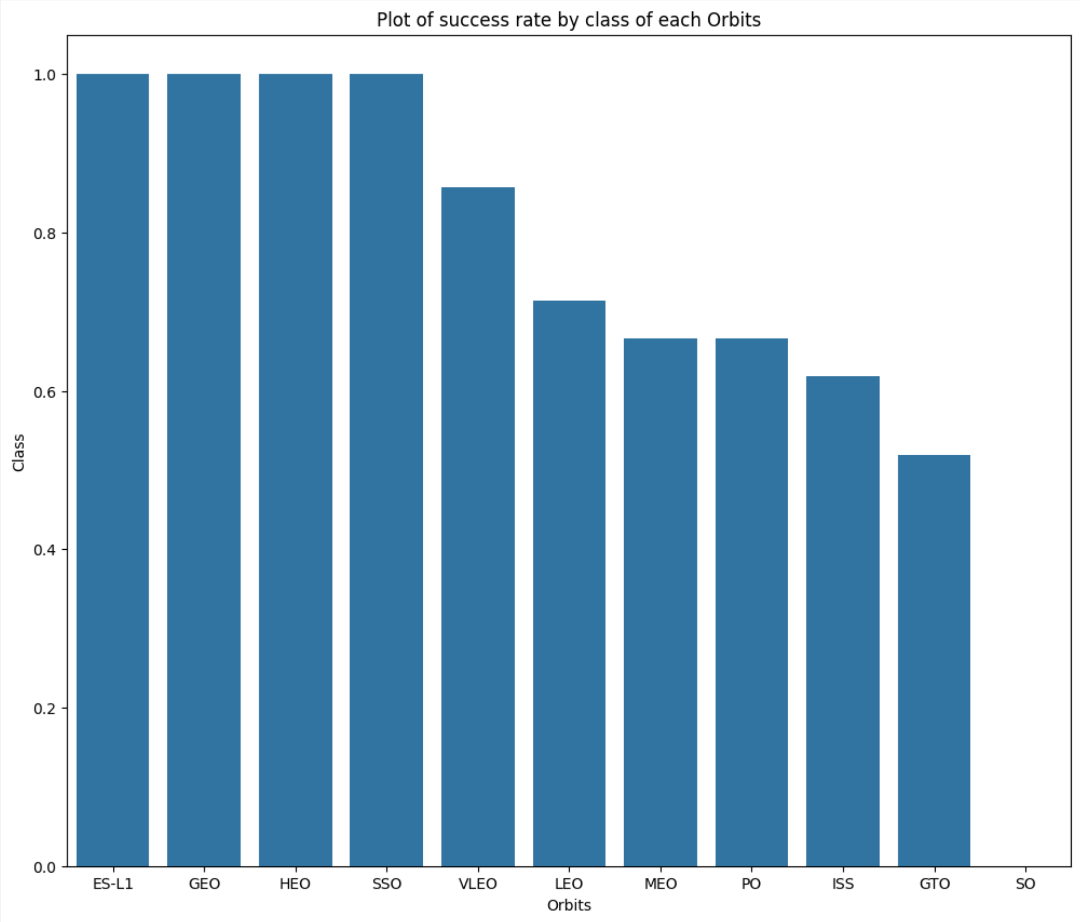
Scatter Plot Insight: Payload Mass vs. Launch Site:

- Most launches from CCAFS SLC 40 span a wide payload range and show a balanced mix of successes and failures.
- KSC LC 39A handled heavier payloads (above 6000 kg), with many successful missions.
- Success (orange) appears more frequent for payloads between 4000–16000 kg, especially from VAFB SLC 4Eand KSC LC 39A.

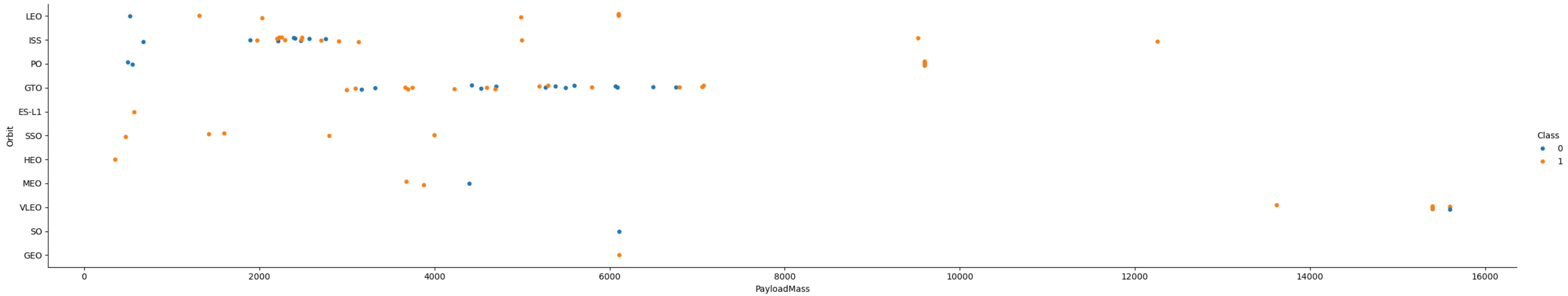
Success Rate vs. Orbit Type

Bar Chart Insight: Success Rate by Orbit Type

- Orbits like **ES-L1**, **GEO**, **HEO**, and **SSO** achieved a **100% success rate**, indicating high reliability.
- Common orbits such as **LEO** and **MEO** had moderate success rates.
- The **lowest success rate** was observed for **GTO** and **SO**, suggesting higher complexity or risk in these missions.



Payload Mass vs. Orbit Type



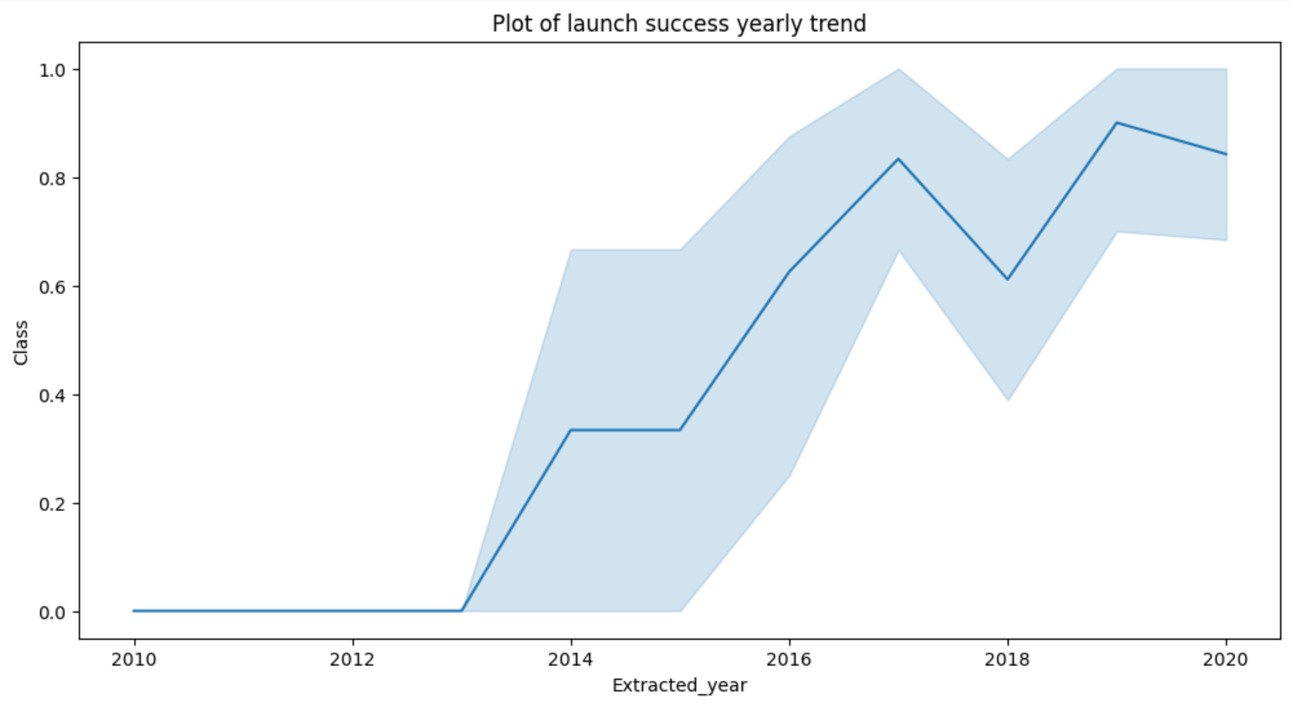
Scatter Plot Insight: Payload Mass vs. Orbit Type

- **GTO and LEO** orbits support the **heaviest payloads**, with varying success rates.
- Orbits like **SSO**, **ES-L1**, and **GEO** mostly handled lighter payloads with high success (more orange).
- **Failures (blue)** are more scattered in **mid-weight payloads** and lower orbits like **ISS** and **MEO**.

Launch Success Yearly Trend

Line Chart Insight: Yearly Launch Success Trend

- SpaceX's launch success rate **significantly improved after 2013**, showing technological and operational progress.
- By **2017**, success rates peaked near **100%**, reflecting greater reliability.
- Minor fluctuations occurred post-2017, but the overall trend remained **strongly upward**.



All Launch Site Names

The query retrieves distinct launch sites from the SpaceX mission data.
Result shows four unique launch sites used:

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

👉 This helps identify **where** SpaceX launches have historically occurred and supports site-based analysis in later visualizations.

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SpaceXTABLE WHERE Launch_site LIKE 'CCA%' LIMIT 5
```

This query filters and displays the **first 5 launch records** where the launch site name begins with 'CCA', which refers to **Cape Canaveral Air Force Station**. These sites, such as **LC-40** and **SLC-40**, are among the most frequently used by SpaceX for their early and commercial missions.

It helps narrow analysis by geographic location and identify trends or patterns specific to **Cape Canaveral-based launches**.

Total Payload Mass

This query calculates the **total payload mass** launched by SpaceX for **NASA's Commercial Resupply Services (CRS)**missions. It reveals that SpaceX has delivered **45,596 kg** of cargo to space under NASA CRS contracts, emphasizing SpaceX's significant role in supporting **ISS resupply operations**.

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

This query calculates the **average payload mass** delivered by the **F9 v1.1** booster version. With an average of **2,928.4 kg**, this highlights the booster's typical load capacity and performance in SpaceX's earlier missions before upgrades like Falcon 9 Full Thrust or Block 5 were introduced.

Avg_Payload_Mass
2928.4

First Successful Ground Landing Date

📅 **Date**: **2015-12-22**

✅ This marks the **first ever successful Falcon 9 landing on a ground pad**, a major milestone for rocket reusability.

First_GroundPad_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

🚀 **Boosters Identified**:

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

✅ These boosters successfully landed on drone ships carrying **payloads between 4000–6000 kg**.

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

Total Number of Successful and Failure Mission Outcomes

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

Indicates **high mission reliability**, with only **one in-flight failure**.

Boosters Carried Maximum Payload

🚀 Booster Versions (each carried **15,600 kg**):

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

✅ These **F9 Block 5 boosters** represent the **peak payload capability** in SpaceX's fleet.

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

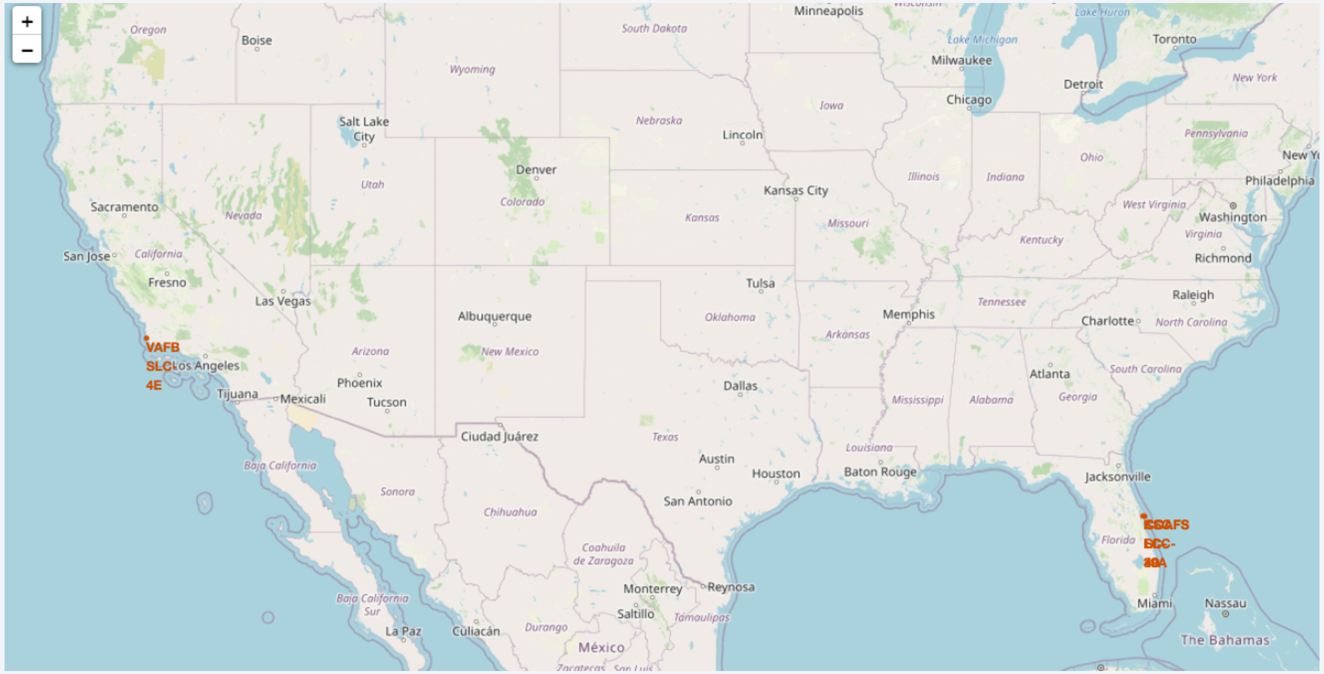
Section 3

Launch Sites Proximities Analysis

Launch Site Locations – Global Overview

The map displays **SpaceX launch sites** using **location markers** across the **United States**:

- **VAFB SLC-4E** (California)
- **CCAFS LC-40** and **SLC-40** (Florida)
- **KSC LC-39A** (Florida)
-



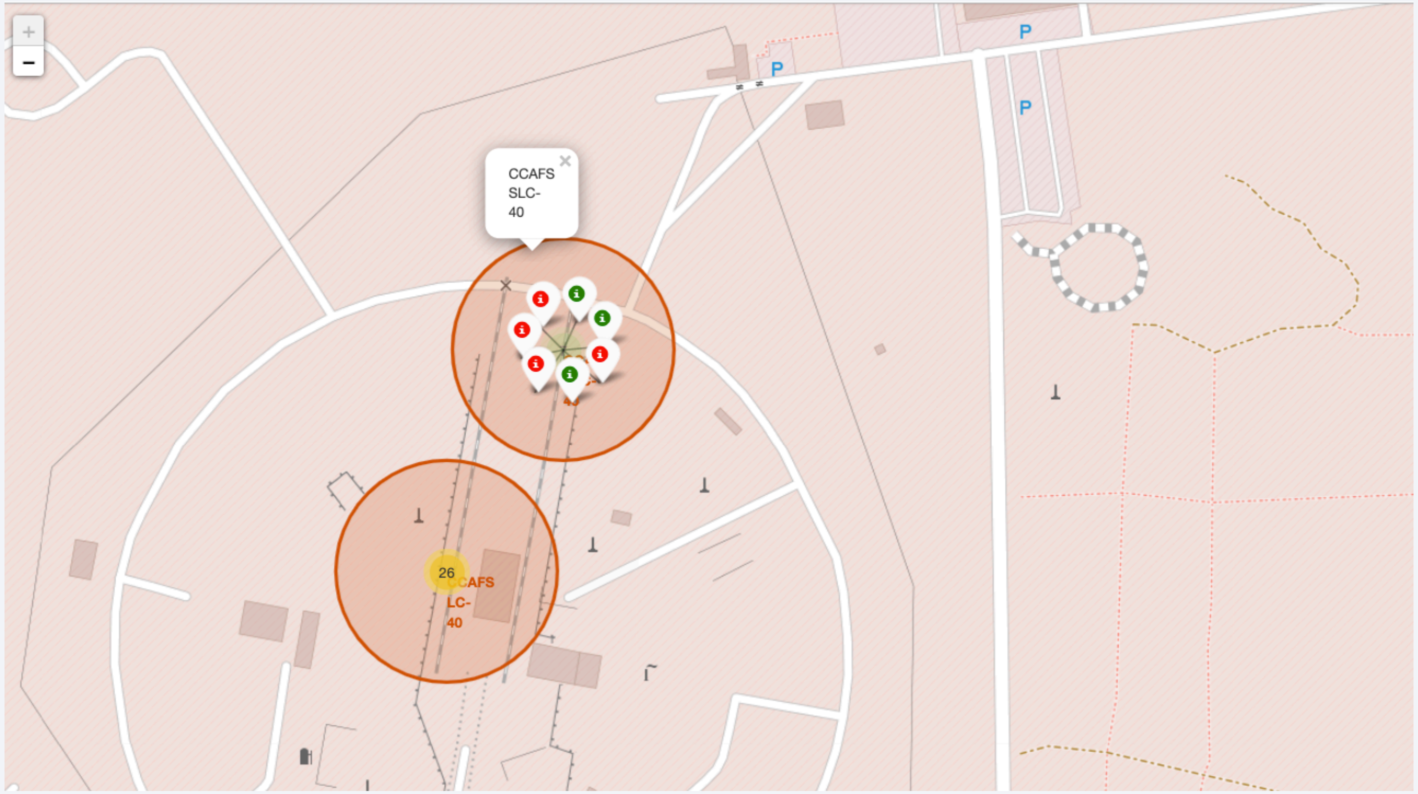
Launch Outcomes

This zoomed-in Folium map highlights the **CCAFS SLC-40 launch pad**.

Green markers indicate **successful launches**, while **red markers** show **failures**.

Each cluster is grouped with a popup label showing the site name.

Two **orange circles** indicate **impact or analysis zones** around the launch site.



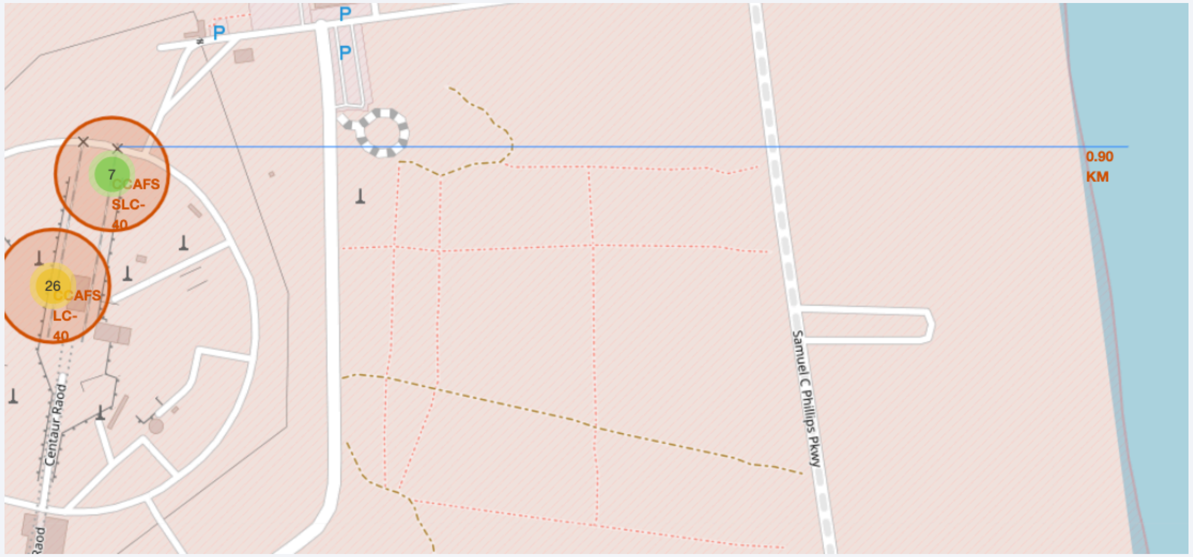
Launch Site Proximity to Coastline

The map shows the **CCAFS SLC-40 launch pad** and surrounding area.

A **blue line** measures the **distance (0.90 km)** from the launch site to the **coastline**, visualizing the site's proximity to water.

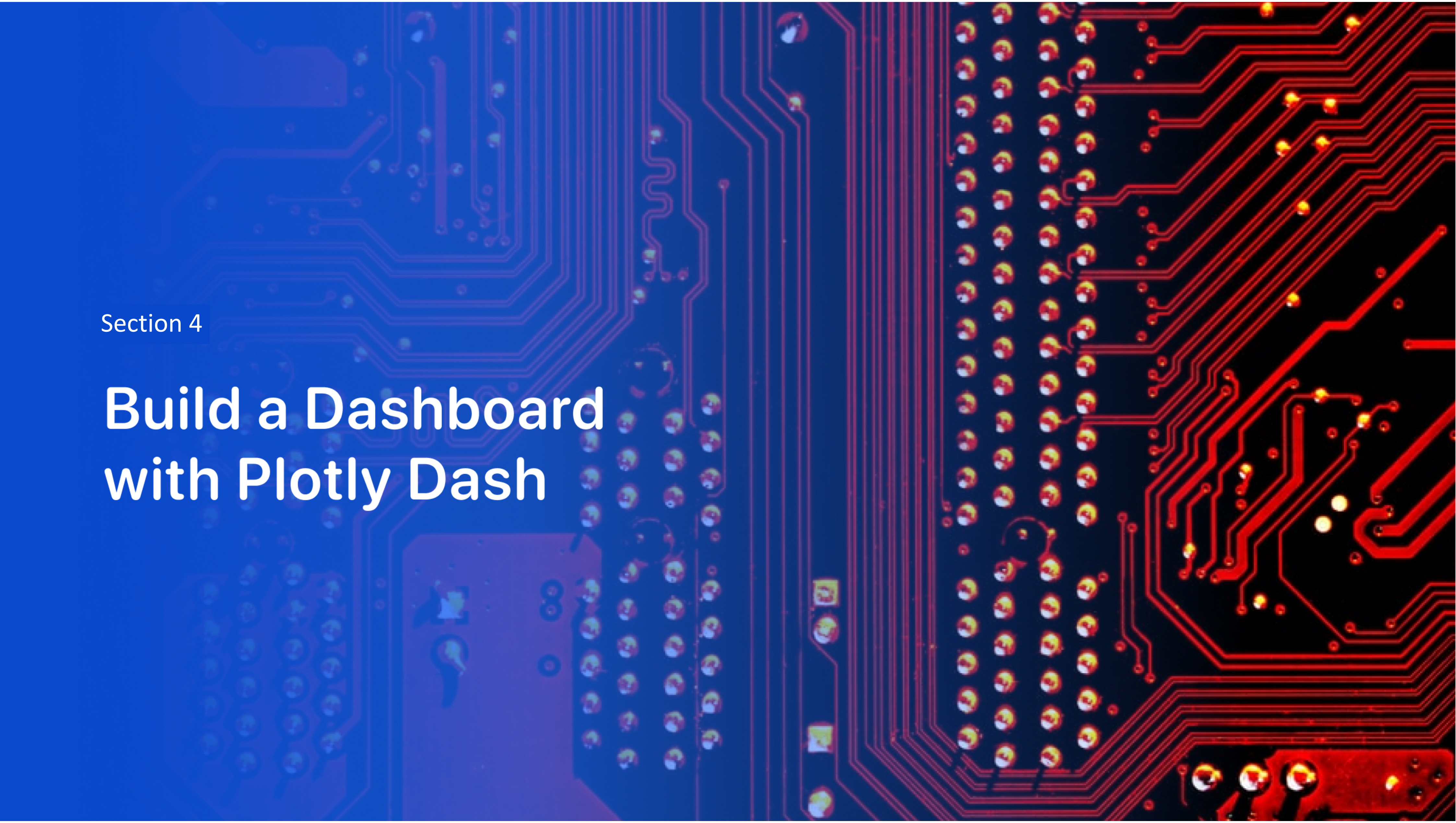
Nearby infrastructure such as **Samuel Phillips Parkway** and **parking areas** is also visible.

Circle markers and **clustered launch indicators** reflect activity at the launch pad.



Section 4

Build a Dashboard with Plotly Dash

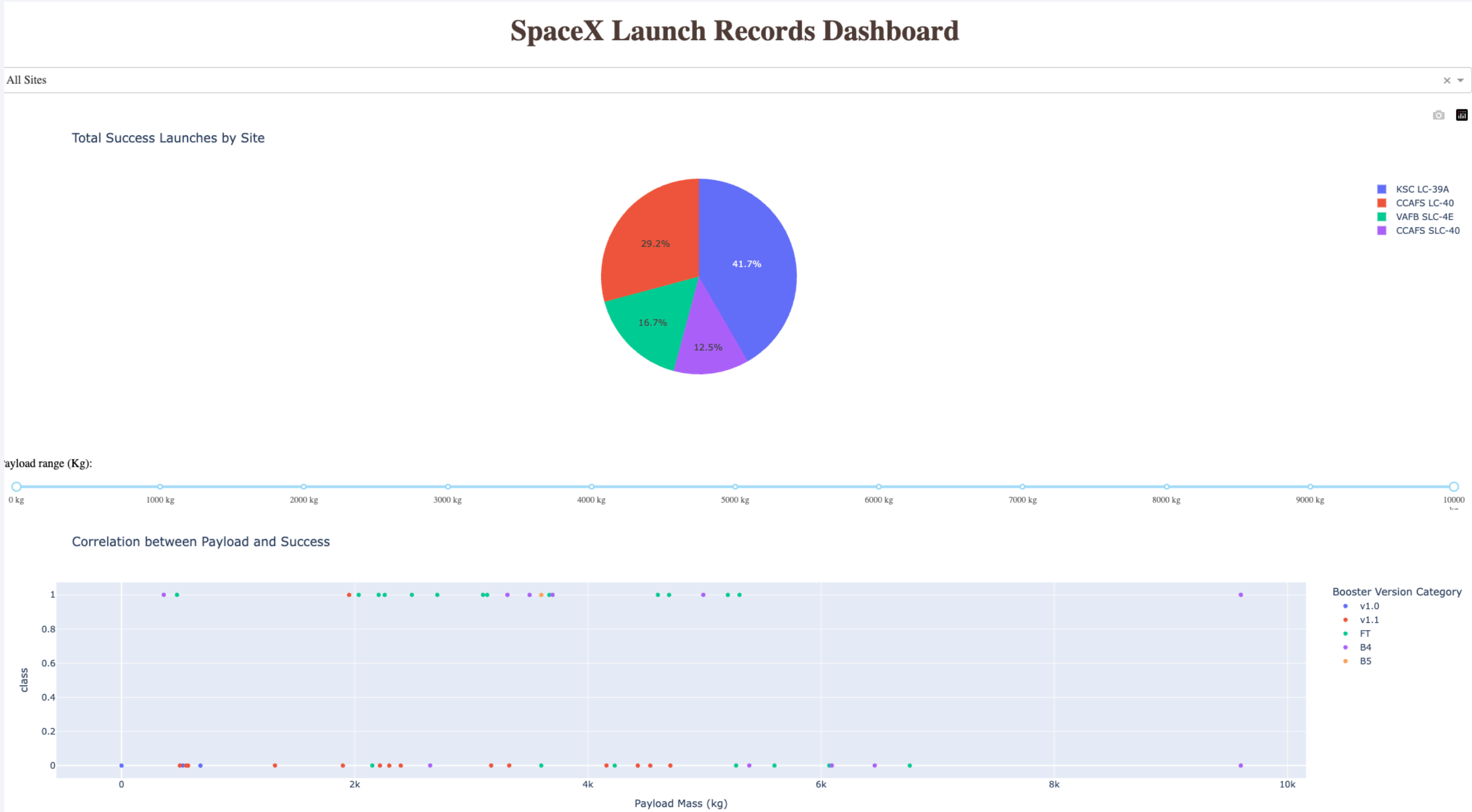


Launch Success Distribution Across All Sites

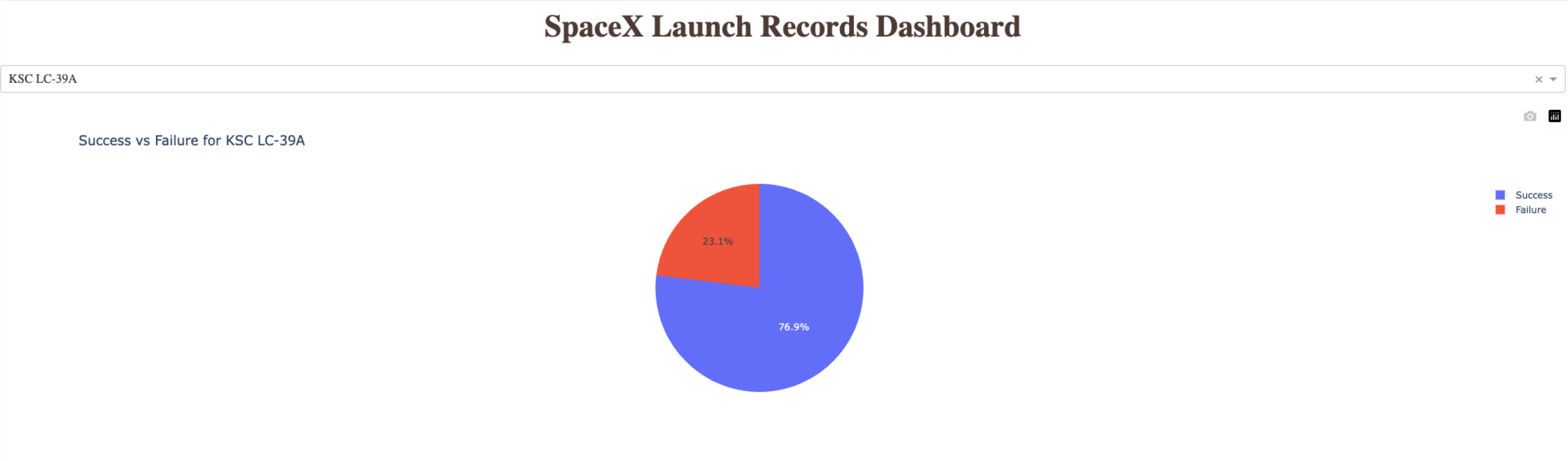
The **pie chart** displays the **percentage of successful launches** at each SpaceX launch site.

Sites included:

- **KSC LC-39A** – 41.7%
- **CCAFS LC-40** – 29.2%
- **VAFB SLC-4E** – 16.7%
- **CCAFS SLC-40** – 12.5%



KSC LC-39A – Site with Highest Launch Success Ratio



KSC LC-39A stands out as the **most successful launch site**, with a **76.9% success rate** across missions.

Conclusions

- SpaceX's launch performance shows clear improvement over time, with high success at specific sites like KSC LC-39A.
- Heavier payloads (4000–6000 kg) and Booster Version F9 B5 are linked with higher mission success.
- Interactive dashboards and maps provided intuitive ways to explore launch trends and outcomes.
- The final classification model supports mission planning by predicting launch success, contributing to safer and more reliable space operations.
- Future work could expand to include weather, launch window timing, and global launch site comparisons.

Thank you!