IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ngo Minh Phuong
22-Feb-2025

Outline

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

Executive Summary

- **Summary of methodologies**

Data Collection: SpaceX REST API (JSON → Pandas DataFrame) and BeautifulSoup web scraping for Falcon 9 data.

Data Wrangling: Filter irrelevant data, handle nulls (e.g., PayloadMass mean), convert data types, one-hot encoding.

EDA: Interactive visualizations (scatter plots, bar/line charts) with Folium and Plotly Dash to analyze flight number, payload, sites, orbits.

Interactive Analytics: Folium maps (site locations, outcomes), Plotly Dash dashboards (pie/scatter charts, dropdowns/sliders).

Predictive Analysis: Preprocess data, train Logistic Regression, SVM, Decision Tree, KNN models with GridSearchCV, evaluate with accuracy/confusion matrices.

Executive Summary

Summary of all results

- EDA Findings: Flight number, payload mass, launch site, orbit type impact landing success; orbits like ES-L1/GEO (100%) outperform GTO/LEO (60-70%); success rates rose to 0.6-0.8% since 2013.
- Launch Sites: CCAFS LC-40 (46.4% launches, 73.1% success); KSC LC-39A/VAFB SLC-4E less reliable; heavy payloads (>10,000 kg) tied to higher success.
- Predictive Results: All models (Logistic, SVM, Decision Tree, KNN) hit 83.33% accuracy; confusion matrix: 12 true positives, 3 false positives, 0 false negatives, 3 true negatives.
- Quantitative Data: NASA payload total: 45,596 kg; F9 v1.1 avg payload: 2,928.4 kg; first ground landing: 2015-12-22; 100 successful, 1 failed mission.

Introduction

Project background and context

The project's background and context are rooted in the burgeoning commercial space age, where companies are striving to make space travel more accessible and affordable. Several entities, including Virgin Galactic, Rocket Lab, and Blue Origin, are contributing to this evolving landscape. Among these, SpaceX stands out as particularly successful.

SpaceX's accomplishments include:

- Sending spacecraft to the International Space Station.
- Establishing Starlink, a satellite internet constellation.
- Launching manned missions into space.

A key factor in SpaceX's success is its relatively inexpensive rocket launches. SpaceX advertises Falcon 9 rocket launches at $62 million, significantly less than other providers that cost upwards of $165 million. This cost-effectiveness is largely attributed to SpaceX's ability to reuse the first stage of its rockets. Determining whether the first stage can be successfully landed is crucial in estimating the cost of a launch.

Introduction

Project background and context

The project simulates a scenario where the user takes on the role of a data scientist at Space Y, a new rocket companyseeking to compete with SpaceX. The primary task is to determine the price of each launch by:

- **Gathering information about SpaceX.**
- **Creating dashboards for the team.**
- **Predicting whether SpaceX will reuse the first stage of its Falcon 9 rockets.**

Instead of relying on rocket science principles, the project employs a machine learning model trained on public information to predict the reusability of the Falcon 9's first stage. This approach enables Space Y to make informed decisions and compete effectively in the commercial space market.

Section 1

Methodology

Methodology

Executive Summary

1. Data Collection Methodology

API Usage (SpaceX REST API)

- Source: api.spacexdata.com/v4/, e.g., /launches/past
- Data: Launches, rockets, payloads, launch/landing specs, outcomes
- Format: JSON (list of objects) → Pandas DataFrame via json_normalize
- Specific Endpoints: Target booster, launchpad, payload, core data using IDs

Web Scraping

- Tool: Python BeautifulSoup
- Task: Extract Falcon 9 launch data from HTML tables
- Output: Parsed data → Pandas DataFrame

Executive Summary

2. Data Wrangling

Filtering:

- Remove irrelevant data (e.g., Falcon 1 launches)

Handling Null Values:

- Replace PayloadMass nulls with mean
- Retain NULLs in LandingPad if no pad used

Methodology

Executive Summary

3. Exploratory Data Analysis (EDA

Key Techniques:

- Combine attributes for deeper insights (e.g., 100% landing success at CCAFS LC-40 for payloads >10,000 kg)
- Use interactive visualizations and dashboards over static graphs

SpaceX Insights:

- Launch Sites: CCAFS LC-40 (60%), KSC LC-39A/VAFB SLC 4E (~77%) success rates
- Temporal Trends: Improved success since 2013, usable as a feature via launch numberOptimal Site Selection: Analyze geos and proximities on maps

Executive Summary

4. Interactive Visual Analytics with Folium and Plotly Dash

Folium

- Purpose: Analyze launch site geography and proximities
- Functionality: Marks launch sites and nearby features on interactive maps
- Use Case: Explore patterns to select optimal launch sites

Plotly Dash

- Purpose: Build interactive dashboard applications
- Functionality: Includes inputs like dropdowns and sliders for chart interaction
- Use Case: Easily uncover insights from datasets

Methodology

Executive Summary

5. Predictive Analysis for Falcon 9 First Stage Landing

- ❑Data Preprocessing: Standardize the data
- ❑Data Splitting: Use train_test_split for training and testing sets
- ❑Model Training: Train models with Grid Search to optimize hyperparameters
- ❑Model Selection: Choose the model with best accuracy using training data
- ❑Classification Models:
  - Logistic Regression
  - Support Vector Machines
  - Decision Tree Classifier
  - K-Nearest Neighbors
- ❑Evaluation: Use confusion matrix to assess model performance

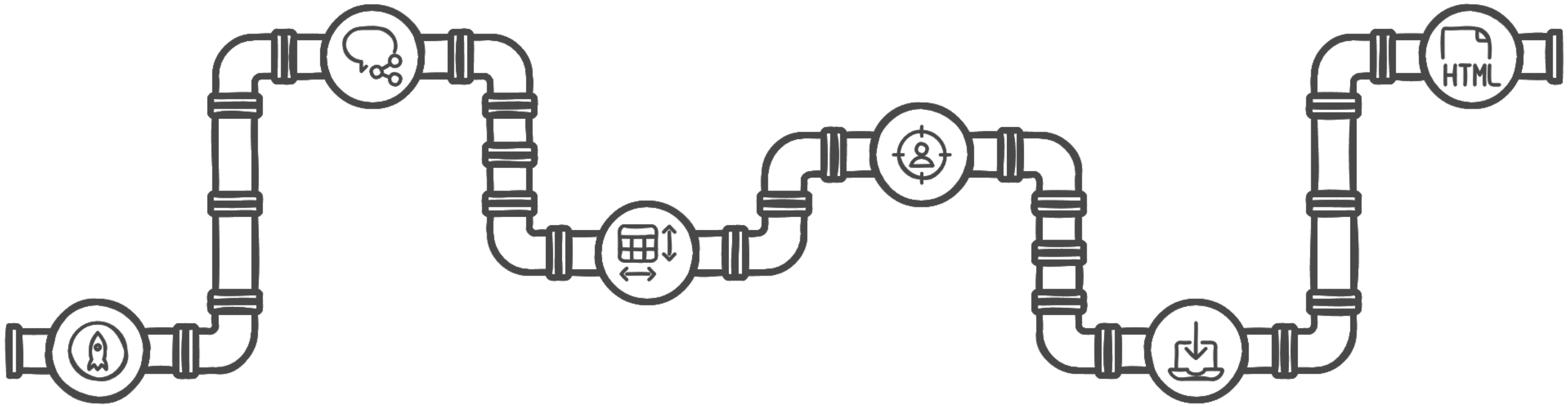
Data Collection

API and Web Scraping for SpaceX Data Analysis

This document outlines the methods for accessing and analyzing data from the SpaceX REST API and web scraping Falcon 9 launch records. It provides a step-by-step guide on how to retrieve JSON data, convert it into a Pandas DataFrame, and extract relevant information using Python libraries. The goal is to facilitate data analysis for launches, rockets, payloads, and other related information.

Data Collection

SpaceX Data Retrieval and Analysis



01 Access SpaceX API

Use the base URL to access SpaceX data

02 Send GET Request

Retrieve past launch data using specific endpoint

03 Convert JSON to DataFrame

Use Pandas to convert JSON data for analysis

04 Gather Specific Data

Target additional endpoints for detailed data

05 Install BeautifulSoup

Install the package for web scraping

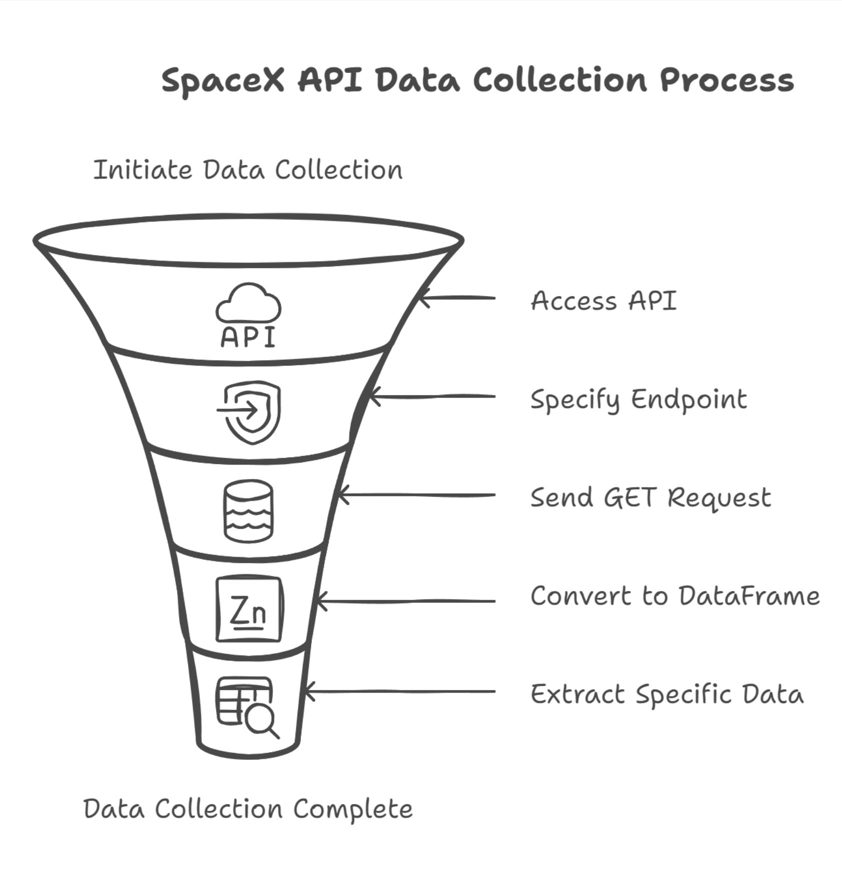
06 Parse HTML Tables

Extract data from HTML tables using BeautifulSoup

Data Collection - SpaceX API

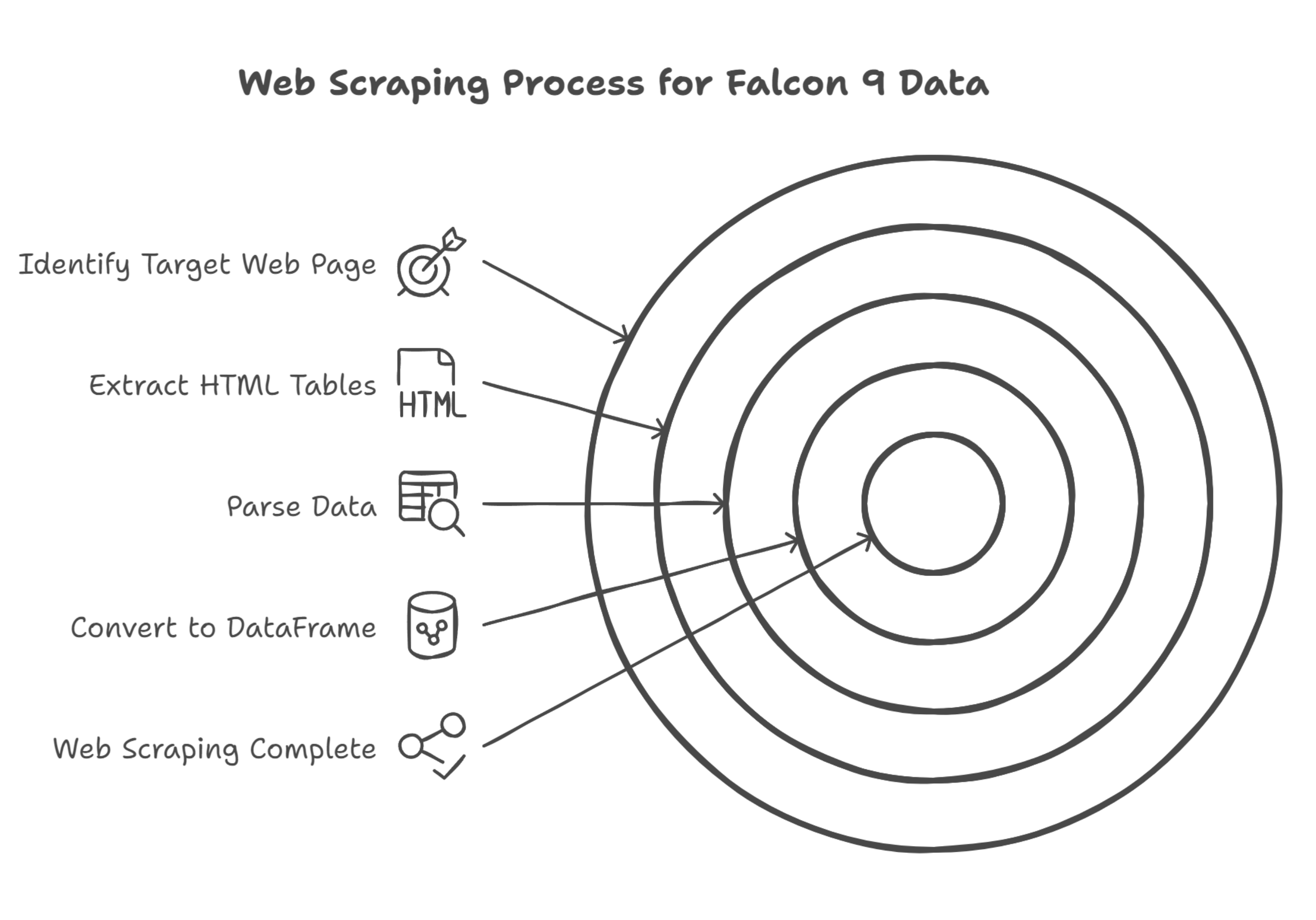
1. Start: Initiate data collection.
2. Access SpaceX REST API: Use base URL api.spacexdata.com/v4/.
3. Specify Endpoint: Target specific endpoint (e.g., /launches/past).
4. Send GET Request: Retrieve data in JSON format.
5. Convert to DataFrame: Use json_normalize() to create a Pandas DataFrame.
6. Extract Specific Data: Use ID numbers to target endpoints for boosters, launchpads, payloads, and cores.
7. End: Data collection complete.

- GitHub URL: https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/jupyter-labs-spacex-data-collection-api(solution).ipynb



Data Collection - Scraping

1. **Start**: Initiate web scraping
2. **Identify Target Web Page**: Locate the web page containing the desired HTML tables with Falcon 9 launch records
3. **Extract HTML Tables**: Use the **BeautifulSoup package** to extract the relevant HTML tables from the web page

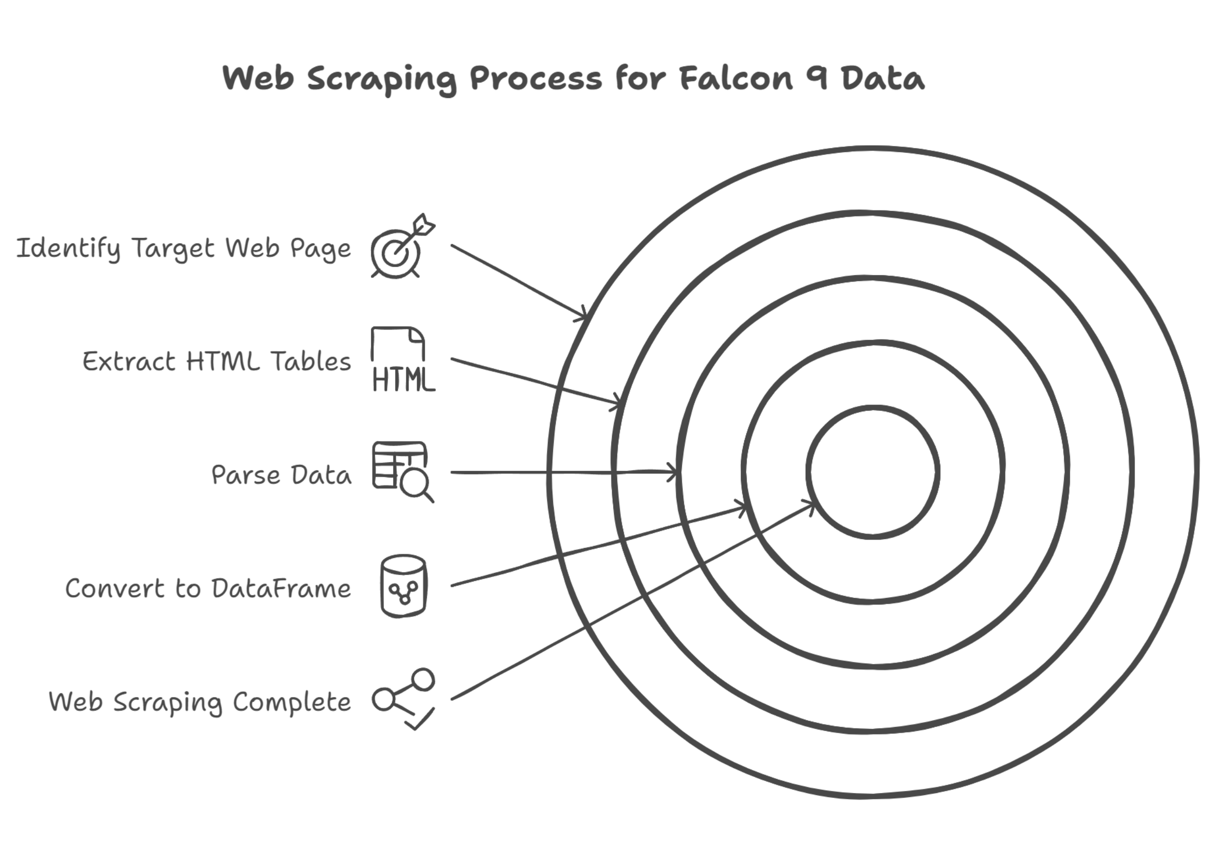


Data Collection - Scraping

4. **Parse Data**: Parse the extracted HTML data to identify and isolate the specific data elements within the tables
5. **Convert to DataFrame**: Convert the parsed data into a Pandas DataFrame to facilitate further analysis and manipulation
6. **End**: Web scraping complete

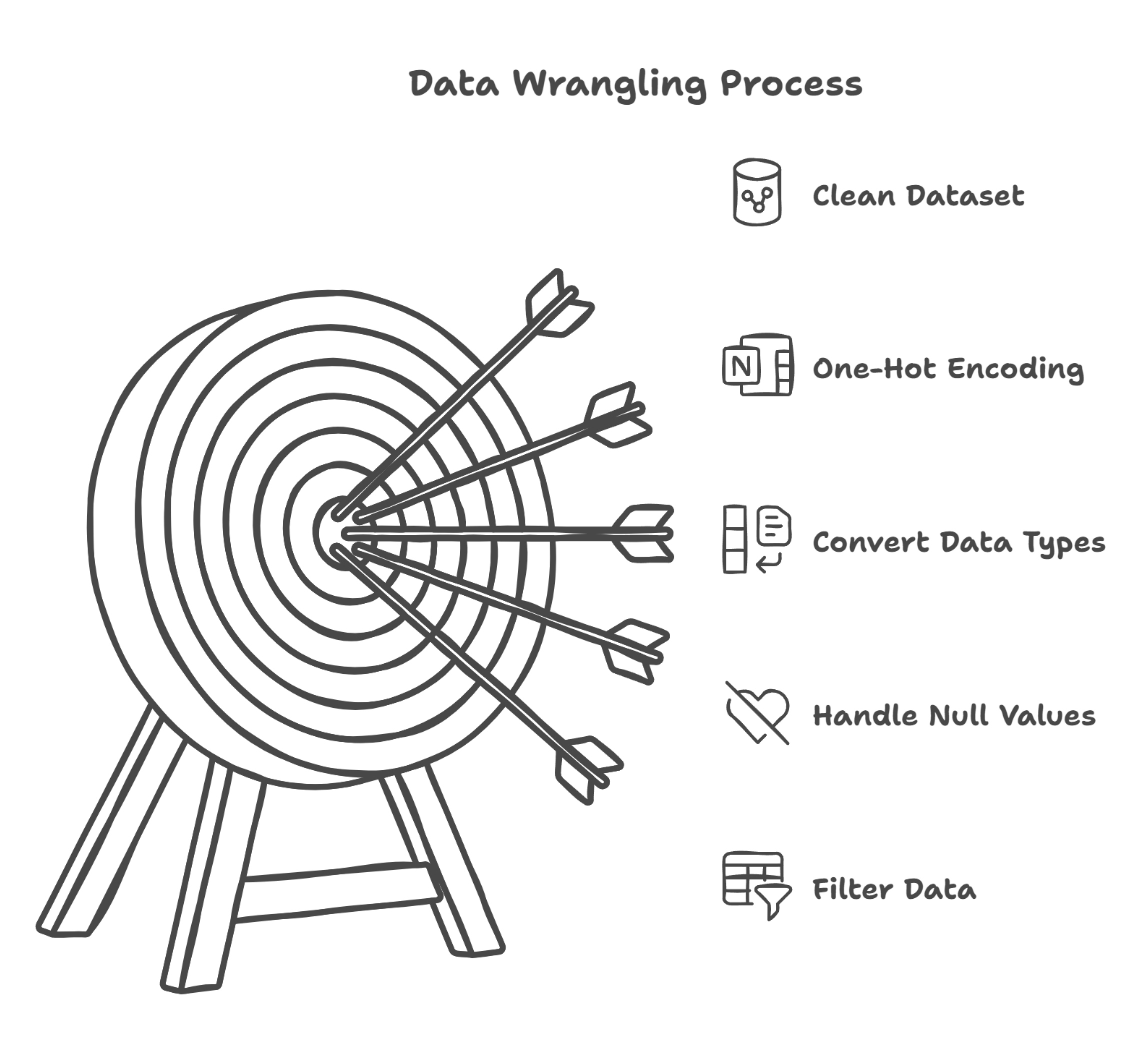
GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/jupyter-labs-webscraping%20(solution).ipynb



Data Wrangling

1. **Start: Begin data wrangling.**
2. **Filter Data: Remove irrelevant entries (e.g., Falcon 1 launches)**
3. **Handle Null Values: Identify and address null values in columns such as PayloadMass by replacing them with the mean.**
4. **Convert Data Types: Transform data types as needed, such as converting landing outcomes to binary classes.**

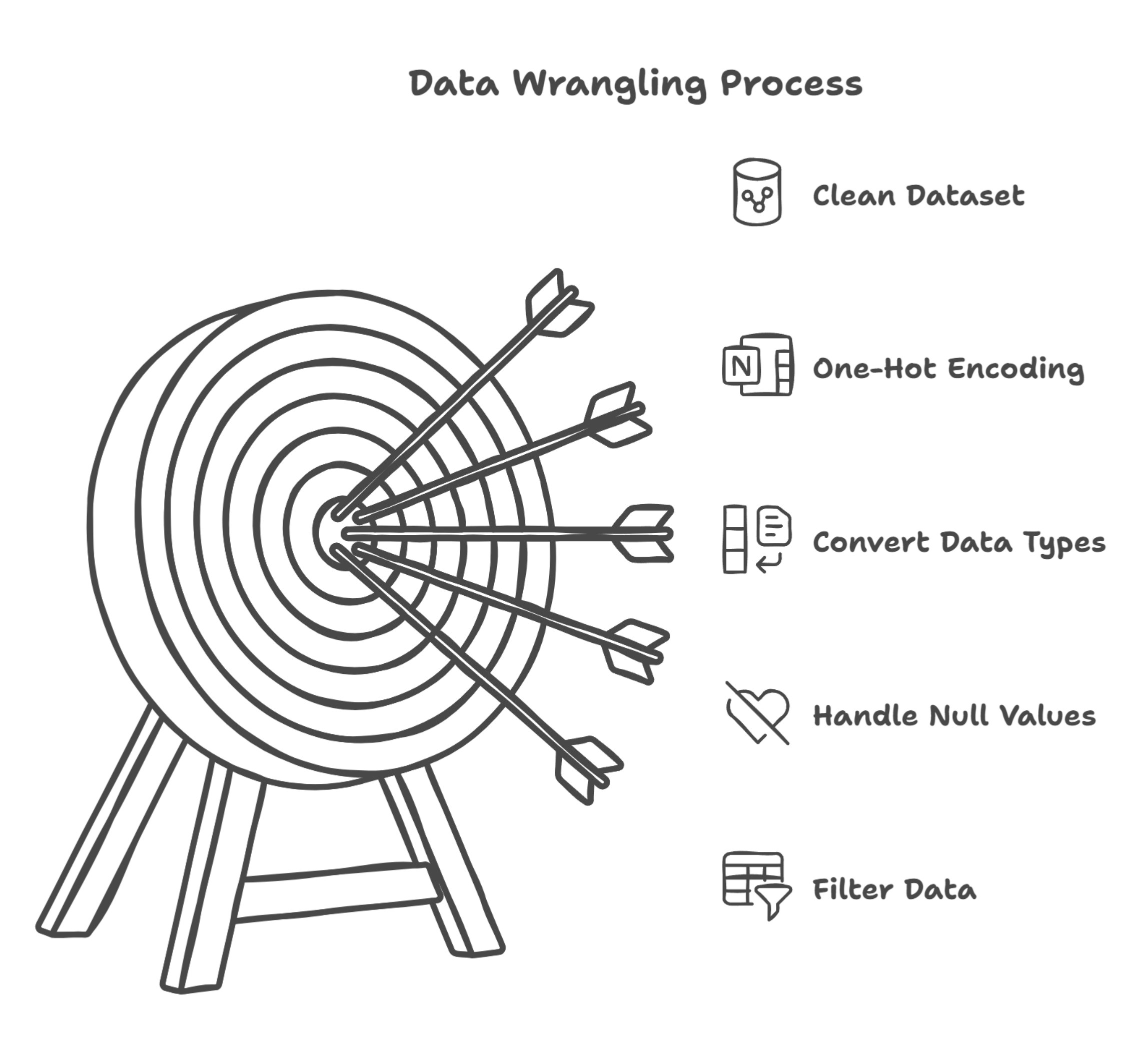


Data Wrangling

5. **Perform One-Hot Encoding: Convert categorical variables into numerical format.**
6. **End: Data wrangling complete, resulting in a clean and usable dataset**

GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/labs-jupyter-spacex-Data%20wrangling(solution).ipynb



EDA with Data Visualization

1. **Scatter Plot: Flight Number vs. Payload Mass**
   - Purpose: Visualize relationship between flight number, payload mass, and launch outcome (success/failure) to identify correlations with mission success.
   - Why Scatter Plot?: Shows relationship between two continuous variables (flight number, payload mass) and their impact on a categorical outcome.
2. **Scatter Plot: Flight Number vs. Launch Site**
   - Purpose: Explore relationship between flight number, launch site, and launch outcome to identify site-specific success trends over time.
   - Why Scatter Plot?: Visualizes relationship between flight number (continuous) and launch site (categorical) affecting launch outcome.

EDA with Data Visualization

3. **Scatter Plot: Payload Mass vs. Launch Site**
    - Purpose: Investigate relationship between payload mass, launch site, and launch outcome to assess payload capacity limits and success rates by site.
    - Why Scatter Plot?: Displays correlation between payload mass (continuous) and launch site (categorical) impacting launch outcome.
4. **Bar Chart: Success Rate of Each Orbit Type**
    - Purpose: Compare success rates across orbit types to identify high/low success orbits.
    - Why Bar Chart?: Effectively compares categorical variable (orbit type) against continuous variable (success rate) for clear visual differences.

EDA with Data Visualization

5. **Scatter Plot: Flight Number vs. Orbit Type**
   - Purpose: Analyze relationship between flight number, orbit type, and launch outcome to see if success improves with more flights in specific orbits.
   - Why Scatter Plot?: Visualizes relationship between flight number (continuous) and orbit type (categorical) affecting launch outcome.
6. **Scatter Plot: Payload Mass vs. Orbit Type**
   - Purpose: Examine relationship between payload mass, orbit type, and launch outcome to assess payload limits and success impacts in specific orbits.
   - Why Scatter Plot?: Shows correlation between payload mass (continuous) and orbit type (categorical) influencing launch outcome.

EDA with Data Visualization

7. **Line Chart: Launch Success Yearly Trend**
    - **Purpose**: Observe yearly trends in launch success to identify periods of improvement or decline.
    - **Why Line Chart?**: Displays trends over time, effectively showing changes in success rate year by year to spot patterns.

GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/edadataviz(solution).ipynb

EDA with SQL

Data Retrieval

- SELECT column1, column2 FROM table_name WHERE condition; – Retrieves specific columns from a table based on a condition.
- SELECT * FROM table_name; – Fetches all columns and rows from a table.

Data Filtering

- WHERE condition; – Filters data using operators like =, >, <, >=, <=, <> , LIKE, IN, BETWEEN.

Data Aggregation

- SUM(), AVG(), MIN(), MAX(), COUNT() – Calculates sums, averages, minimums, maximums, and counts of column data.
- GROUP BY column_name; – Groups rows by a column for aggregate calculations.
- HAVING condition; – Filters grouped results based on a condition.

EDA with SQL

Data Sorting

- ORDER BY column_name ASC/DESC; – Sorts data in ascending or descending order by a column.

Data Manipulation

- INSERT INTO table_name (column1, column2) VALUES (value1, value2); – Adds new data to a table.
- UPDATE table_name SET column1 = value1 WHERE condition; – Modifies existing data in a table.
- DELETE FROM table_name WHERE condition; – Removes data from a table

GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(solution).ipynb

Build an Interactive Map with Folium

Map Objects Created and Added

1. **Circles**: Highlight launch site locations, centered on coordinates (latitude/longitude) for clear geographical distribution.
2. **Markers:**
   - **Launch Site Markers**: Label each launch site with its name for clear visual identification.
   - **Launch Outcome Markers**: Use green (success) and red (failure) colors, clustered to avoid clutter in high-launch areas.
3. **Lines (PolyLines):** Connect launch sites to nearby points (coastline, city, railway, highway) to show proximity.
4. **Mouse Position:** Tool added to display coordinates on hover, aiding interactive location and distance identification.

Build an Interactive Map with Folium

Reasons for Inclusion

- **Visualizing Launch Site Locations**: Circles and site markers provide an overview of launch site distribution.
- **Representing Launch Outcomes:** Color-coded outcome markers and clustering ensure clarity in success/failure visualization.
- **Analyzing Proximity to Points of Interest**: Lines reveal launch sites' closeness to key features for strategic analysis.
- **Interactive Exploration:** Mouse Position enables real-time coordinate checking and distance calculations for user engagement.

GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/lab_jupyter_launch_site_location(solution).ipynb

Build a Dashboard with Plotly Dash

Plots/Graphs and Interactions:

- **Pie Chart:** A pie chart is included in the dashboard application. The source mentions that too much information in a pie chart can make it difficult to decipher the key message.
- **Scatter Point Chart:** A scatter point chart is included in the dashboard application.
- **Dropdown List:** A dropdown list is included as an input component to interact with the charts.
- **Range Slider:** A range slider is included as an input component to interact with the charts

Build a Dashboard with Plotly Dash

Reasons for Adding Plots and Interactions:

- **Interactive Visual Analytics:** To enable users to explore and manipulate data in an interactive and real-time way. Common interactions include zoom-in, zoom-out, pan, filter, search, and link.
- **Faster and More Effective Pattern Recognition:** Interactive visual analytics allows users to find visual patterns more quickly and effectively compared to static graphs.
- **Stakeholder Communication:** To build a dashboard for stakeholders, providing them with an appealing and informative way to understand the data

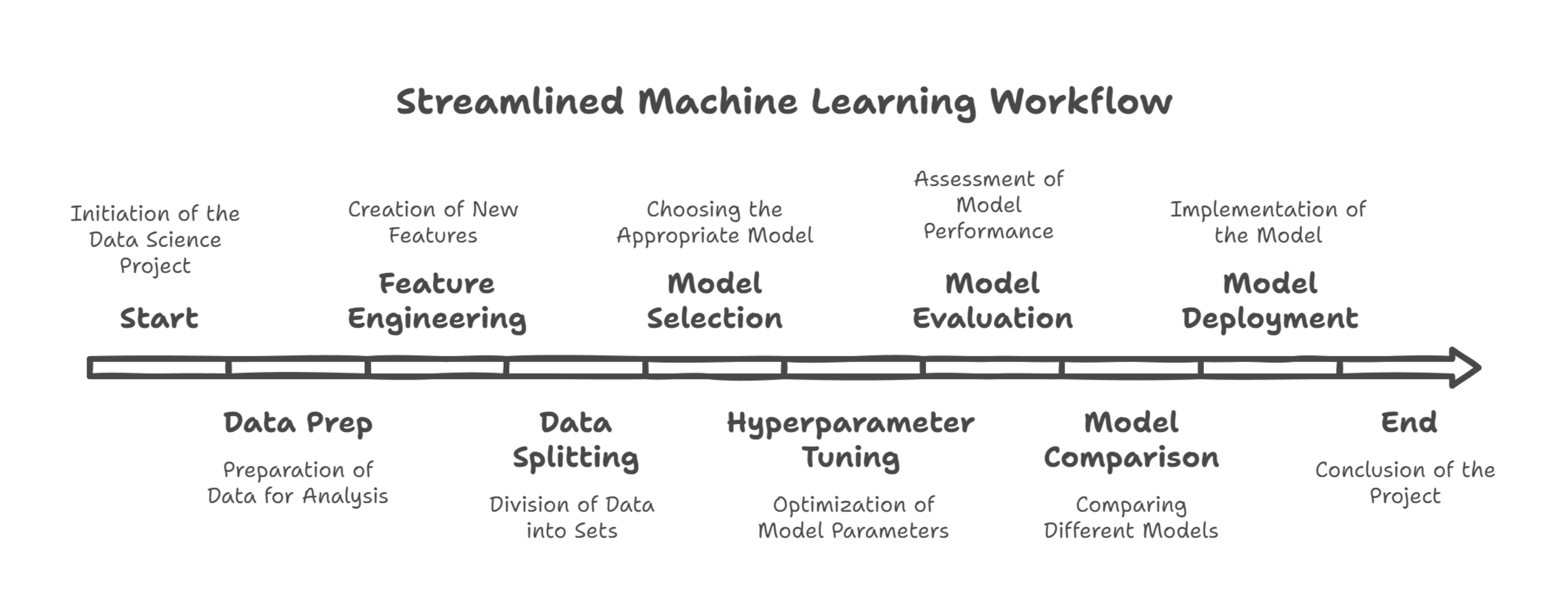
GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/spacex_dash_app%20(solution).py

Predictive Analysis (Classification)

1. **Data Loading & Preparation:** Import libraries (pandas, NumPy, etc.), load dataset, standardize features
2. **Feature Engineering:** Create/transform features for better performance
3. **Data Splitting:** Split data into training/testing sets (20% test, random_state=2)
4. **Model Selection:** Choose classification models (Logistic Regression, SVM, Decision Tree, KNN)
5. **Hyperparameter Tuning**: Optimize with GridSearchCV
6. **Model Evaluation:** Use accuracy, confusion matrices
7. **Model Comparison:** Compare accuracies to pick the best model
8. **Model Deployment:** Apply the best model for real-world predictions

Predictive Analysis (Classification)



GitHub URL

https://github.com/ngophuong1988/Applied-Data-Science-Capstone-2025/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5(solution).ipynb

Results

Exploratory data analysis results

- The flight number, payload mass, launch site, and orbit type are important factors that affect the success rate of SpaceX Falcon 9 first stage landings.
- Some orbits have higher success rates than others.
- The success rate has generally increased over the years.

Results

Predictive analysis results

Results:

Logistic Regression: Accuracy on test data: 0.8333

Support Vector Machine: Accuracy on test data: 0.8333

Decision Tree: Accuracy on test data: 0.8333

K-Nearest Neighbors: Accuracy on test data: 0.8333

Best Performing Model:

All four models achieve the same accuracy of 0.8333 on the test data. This indicates that, based on the provided dataset and evaluation metric, all models exhibit similar predictive capabilities for this specific task.

Results

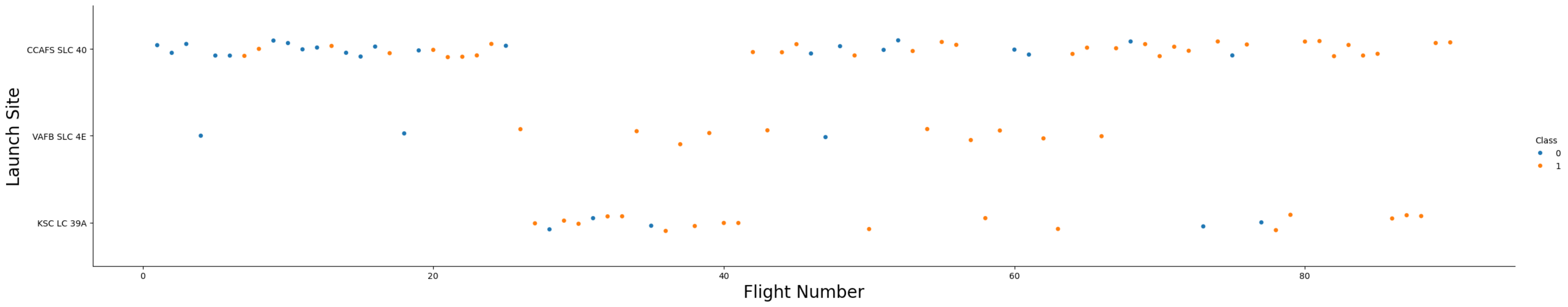
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

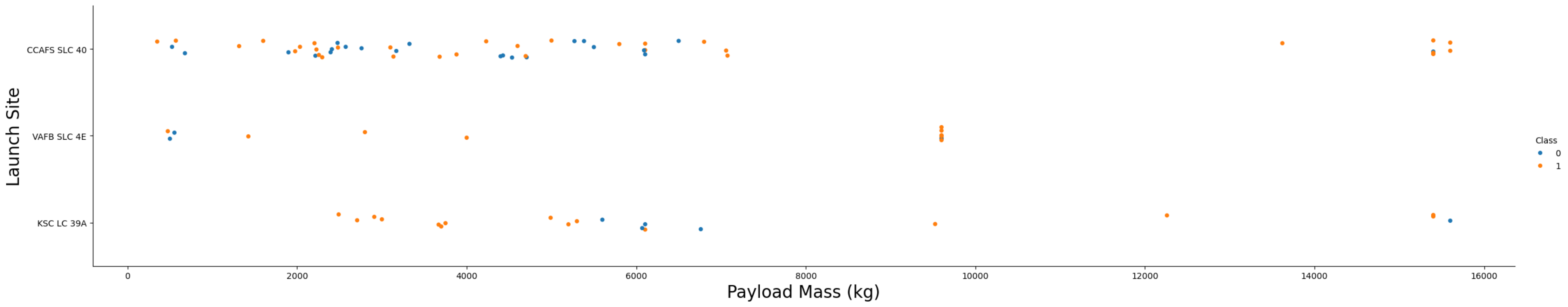
Scatter plot of Flight Number vs. Launch Site



- The scatter plot confirms that flight number is a strong indicator of launch success, with success rates improving over time across all launch sites.
- CCAFS SLC-40 and KSC LC-39A demonstrate higher and more consistent success, especially in later flights, aligning with their higher reported success rates (~60–77%).
- VAFB SLC-4E shows lower reliability in early flights and fewer launches overall, with some improvement later, but it remains less successful than the other sites, possibly due to mission type or payload constraints (e.g., no heavy payloads).
- The color coding (blue for failure, orange for success) effectively highlights the transition from early failures to later successes, supporting the EDA observation of increasing success rates since 2013.

Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

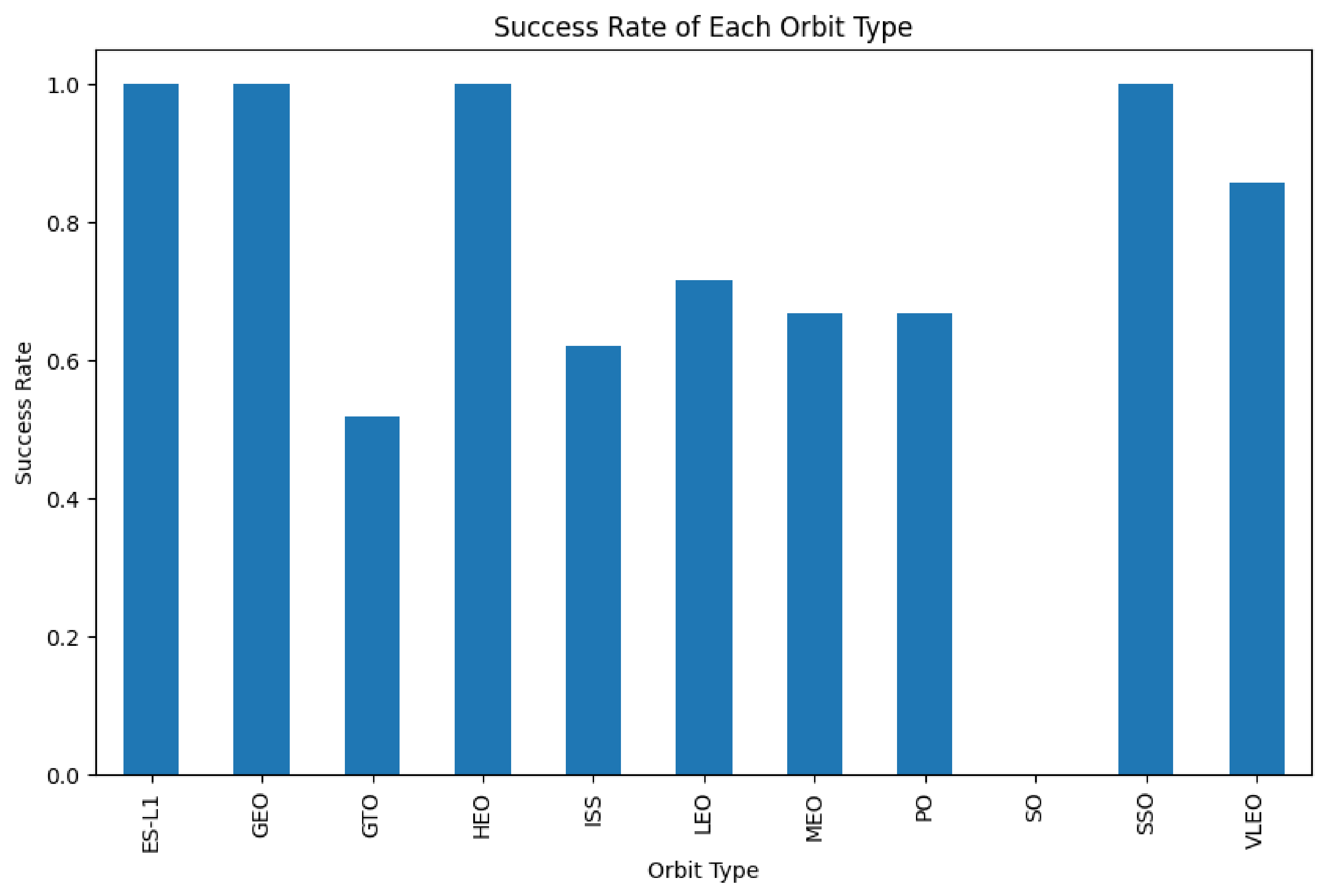


- Payload Success Correlation: Heavier payloads (>10,000 kg) strongly link to success, especially at CCAFS SLC-40.
- Launch Site Reliability: CCAFS SLC-40 and KSC LC-39A show high success across payloads; VAFB SLC-4E is less reliable with no heavy payloads.
- Early Failures with Light Payloads: Failures common at lower payloads (0–4,000 kg), indicating riskier early missions; heavier payloads signal mature success.
- Site-Specific Payload Constraints: VAFB SLC-4E focuses on smaller payloads (polar missions); CCAFS SLC-40 and KSC LC-39A handle broader, heavier payloads successfully.

Success Rate vs. Orbit Type

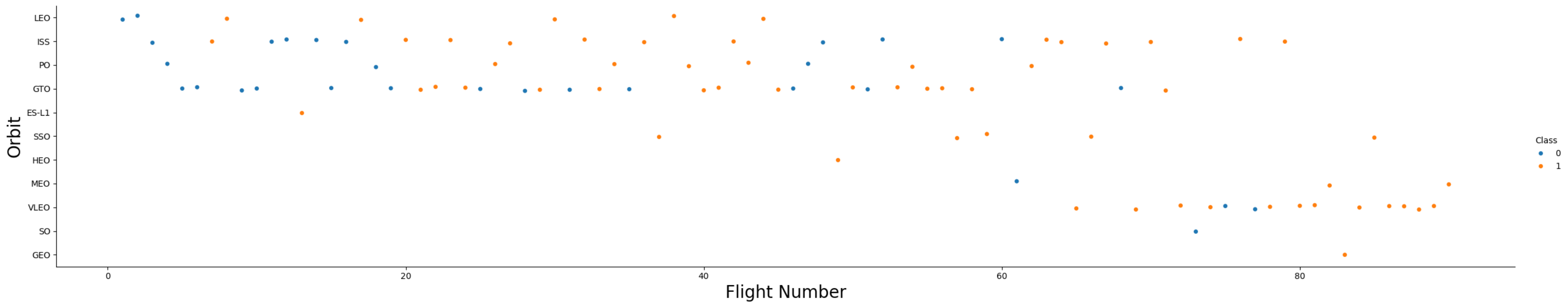
Bar chart for the success rate of each orbit type

- Highest Success Orbits: ES-L1, GEO, HEO, SSO (100% success), most reliable for Falcon 9 landings.
- Moderate Success Orbits: GTO, ISS, LEO, MEO, PO, SO (60–70% success), higher failure risk due to complexity, payload, or dynamics.
- VLEO's Strong Performance: 80–90% success, better than moderate orbits but below top performers.



Flight Number vs. Orbit Type

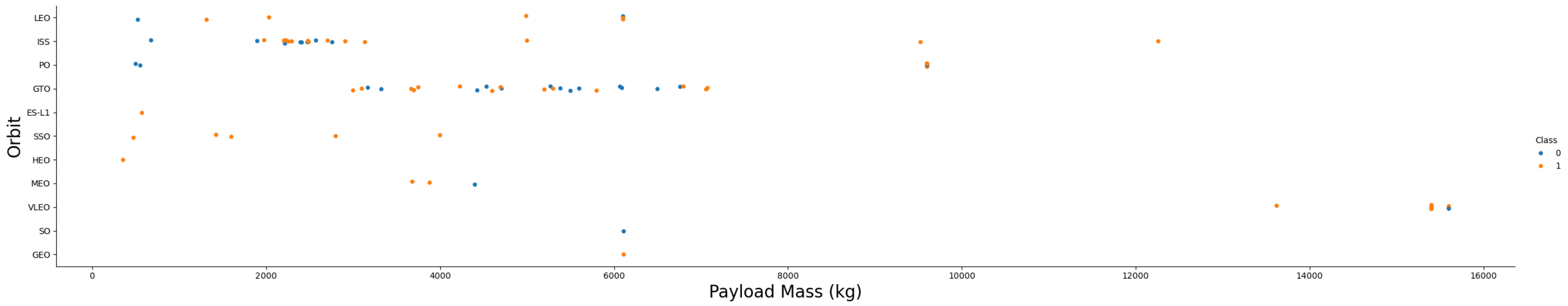
Scatter point of Flight number vs. Orbit type



- High Success Orbits: ES-L1, GEO, HEO, SSO (100% success), most reliable with no/minimal failures.
- Moderate Success Orbits: GTO, ISS, LEO, MEO, PO, SO (60–70% success), early failures improve over time due to complexity/payload challenges.
- VLEO's Strong Performance: 80–90% success, early failures decrease with higher flights, but below top orbits.
- Temporal Improvement: Success rates rise with flight numbers, showing SpaceX's advancements since 2013.
- Density and Focus: GTO, ISS, LEO most common but moderately successful; ES-L1, GEO, HEO, SSO less used but highly reliable

Payload vs. Orbit Type

Scatter point of Payload vs. Orbit type

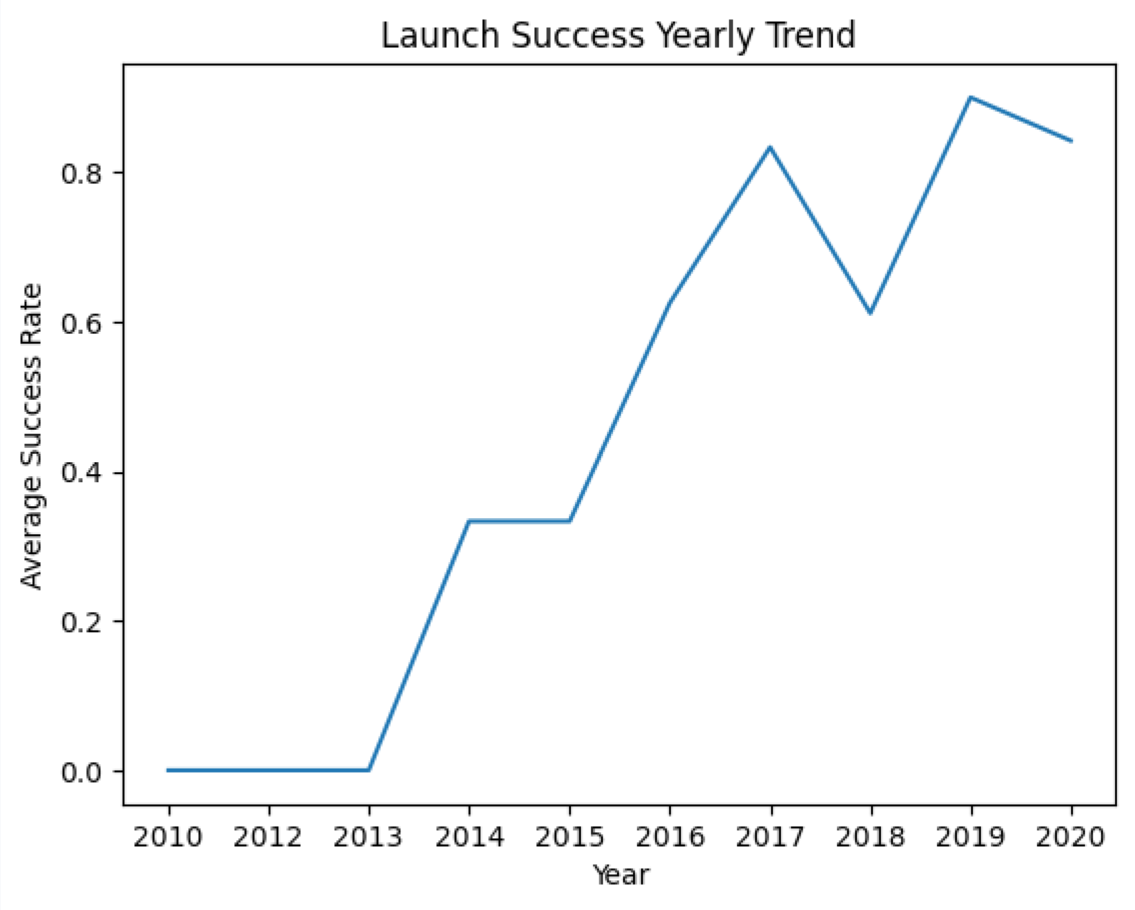


- Most Common Orbits: LEO, ISS, and GTO have the highest density of points, indicating frequent use across a wide range of payload masses (0–16,000 kg).
- Payload Mass Distribution: Most payloads are below 8,000 kg, with fewer missions exceeding 12,000 kg, primarily in LEO.
- Orbit-Specific Trends:
  - LEO and ISS handle the heaviest payloads (up to 16,000 kg), likely for large satellites and space station components.
  - GTO, SSO, HEO, MEO, and Lagrange points typically have smaller payloads (below 4,000 kg), reflecting specialized or lighter missions.

Launch Success Yearly Trend

Line chart of yearly average success rate

- Time Period: The chart tracks the average launch success rate from 2010 to 2020.
- Trend:
    - The success rate was near 0% from 2010 to 2013.
    - It jumped to around 0.4% in 2014 and remained stable until 2015.
    - From 2016 onward, the success rate increased sharply, peaking at around 0.8% in 2017.
    - There were fluctuations, with a dip to about 0.6% in 2018, followed by a rise back to 0.8% in 2019–2020.
- Overall Improvement: The data shows a significant improvement in launch success rates over the decade, stabilizing at a high rate (0.6–0.8%) in the later years.



All Launch Site Names

The query returns four unique launch sites used by SpaceX:

- CCAFS LC-40: Cape Canaveral Air Force Station Launch Complex 40
- VAFB SLC-4E: Vandenberg Air Force Base Space Launch Complex 4E
- KSC LC-39A: Kennedy Space Center Launch Complex 39A
- CCAFS SLC-40: Cape Canaveral Space Force Station Space Launch Complex 40

These are the distinct locations from which SpaceX has launched missions, as recorded in the SPACEXTBL table. The query successfully identifies and presents these unique launch sites.

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

The table shows SpaceX Falcon 9 launches from 2010-2013 at CCAFS LC-40, targeting LEO (ISS) orbits. Early missions had 0 kg payloads for testing, while 2013 missions carried 500-677 kg for NASA. All missions succeeded, but landings failed in 2010-2012 due to parachutes; no landing attempts were made in 2013. Customers included NASA and NRO.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

  45596 Kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)

2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

MIN(Date)

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Successful_Mission_Outcomes	Failed_Mission_Outcomes
100	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

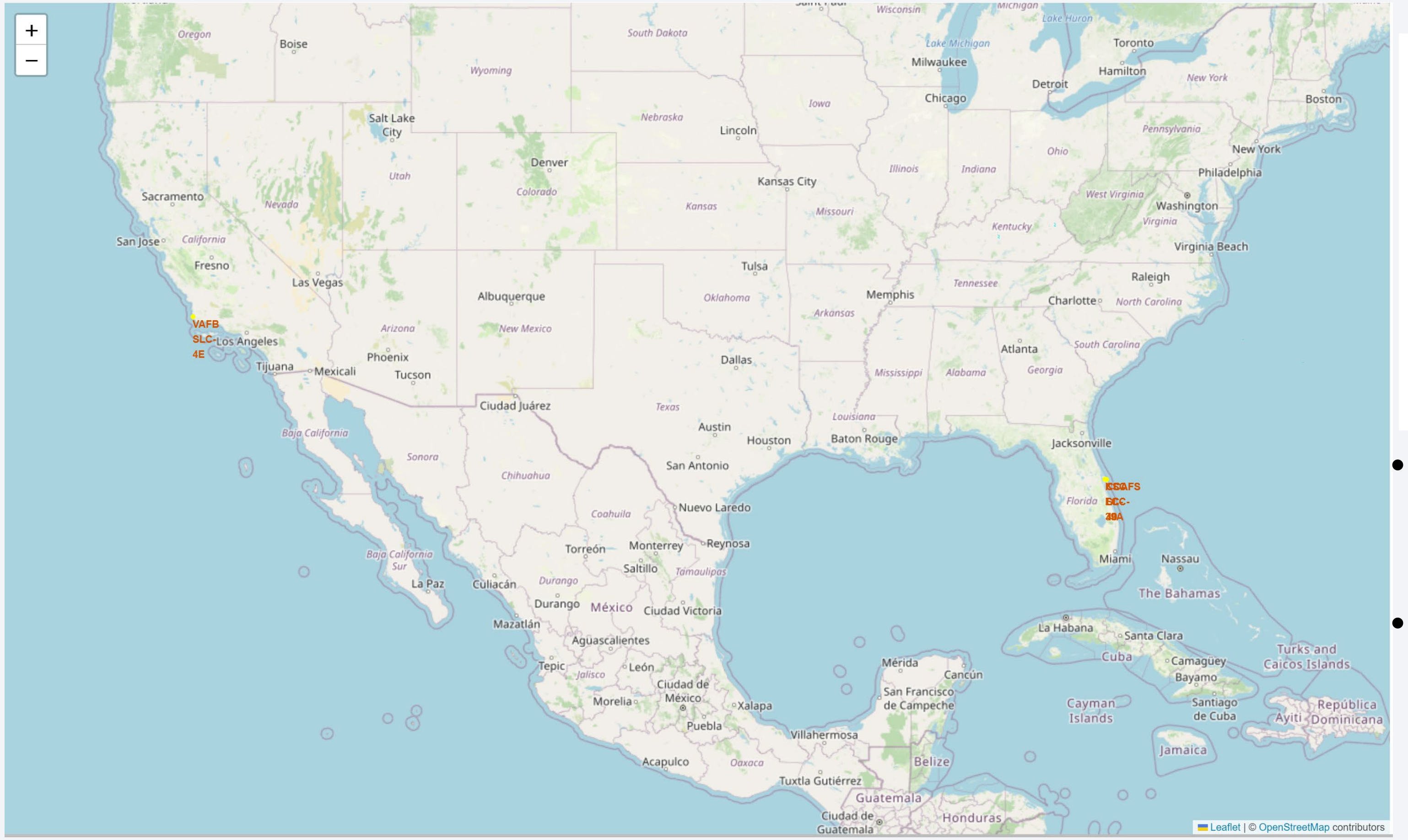
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	LandingOutcomeCount
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

Section 3

Launch Sites Proximities Analysis

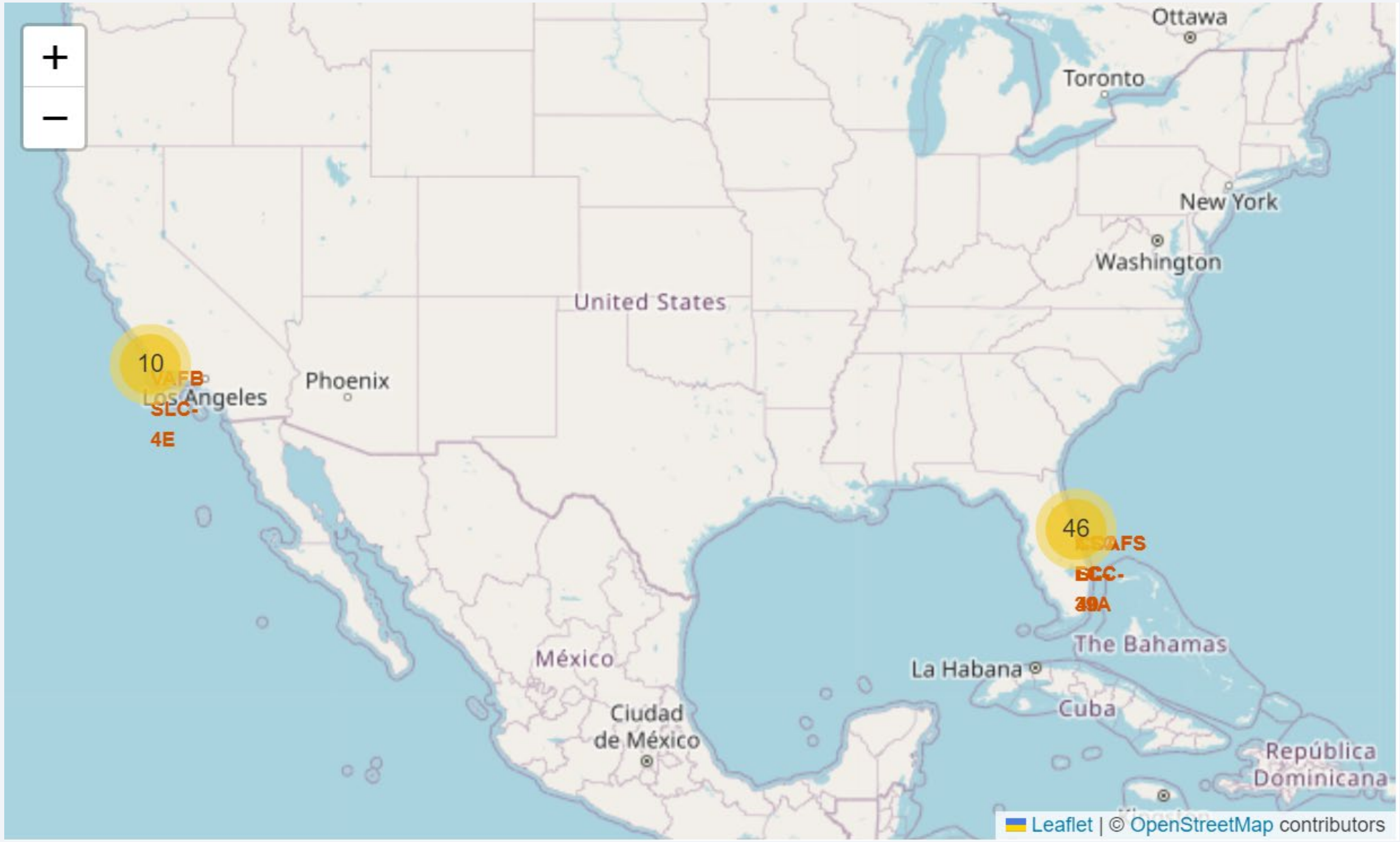
Folium Map of all launch sites' location markers



	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

- Launch sites are generally located closer to the Equator for fuel efficiency.
- All launch sites are located near coastlines for safety and logistical reasons.

Folium Map of the success/failed launches for each site

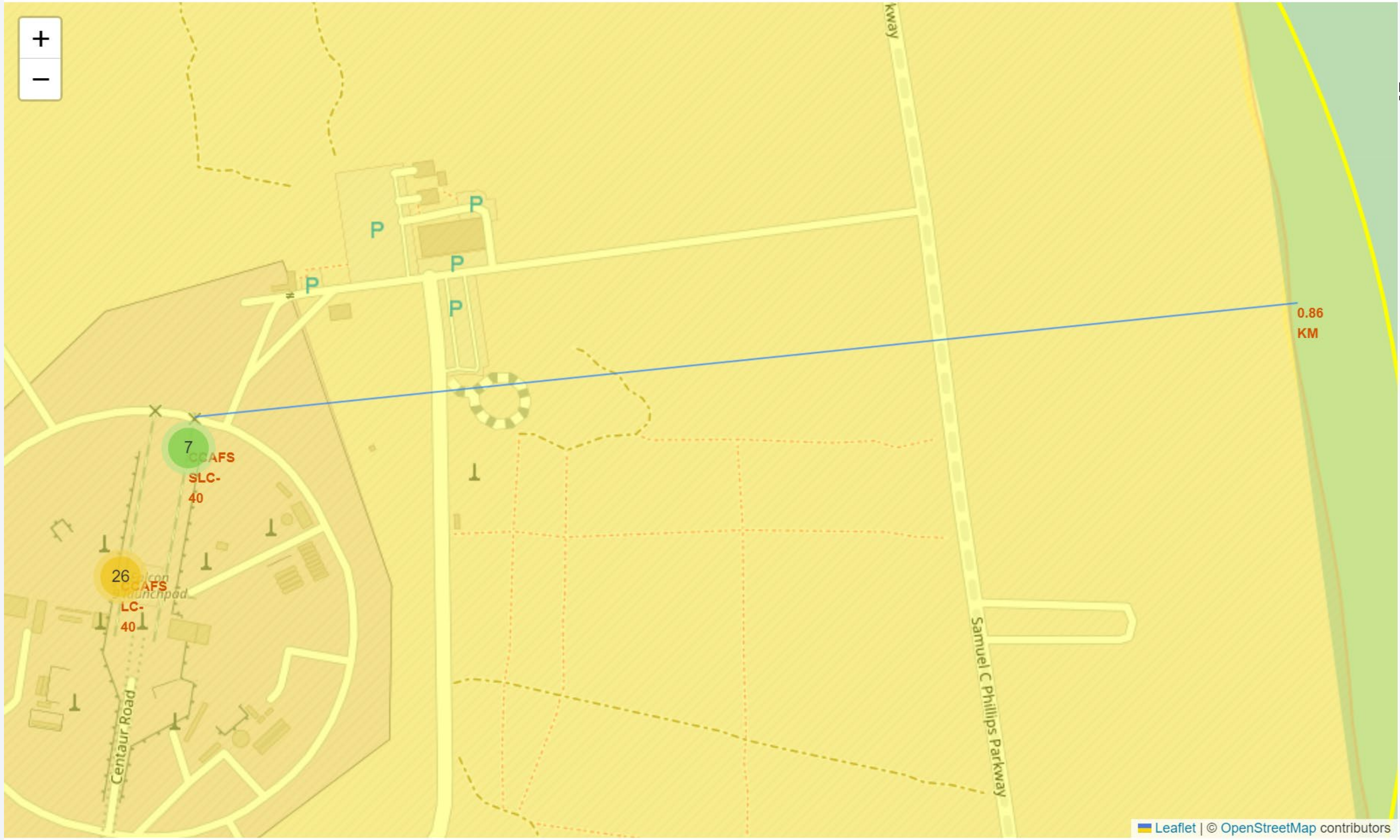


Folium Map of the success/failed launches for each site

CCAFS SLC-40: This launch site has a high concentration of green markers, indicating a relatively high success rate for launches from this location.

CCAFS SLC-40 and KSC LC-39A appear to have the highest success rates based on the visual analysis of the map

Distance between coastline point and launch site



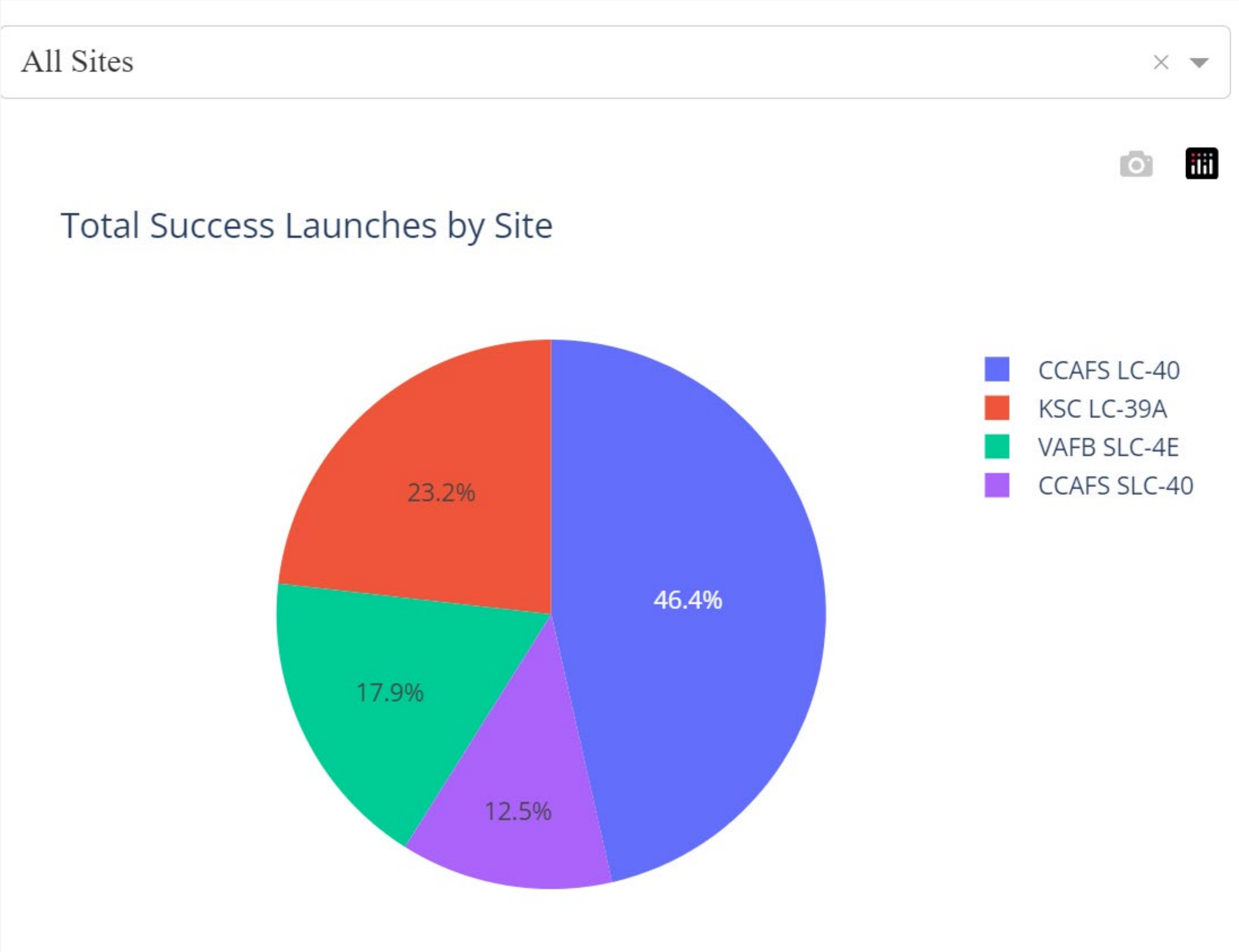
- **All launch sites are located very close to the coastline.**
- **Launch sites generally maintain a certain distance away from major cities**

Section 4

Build a Dashboard with Plotly Dash

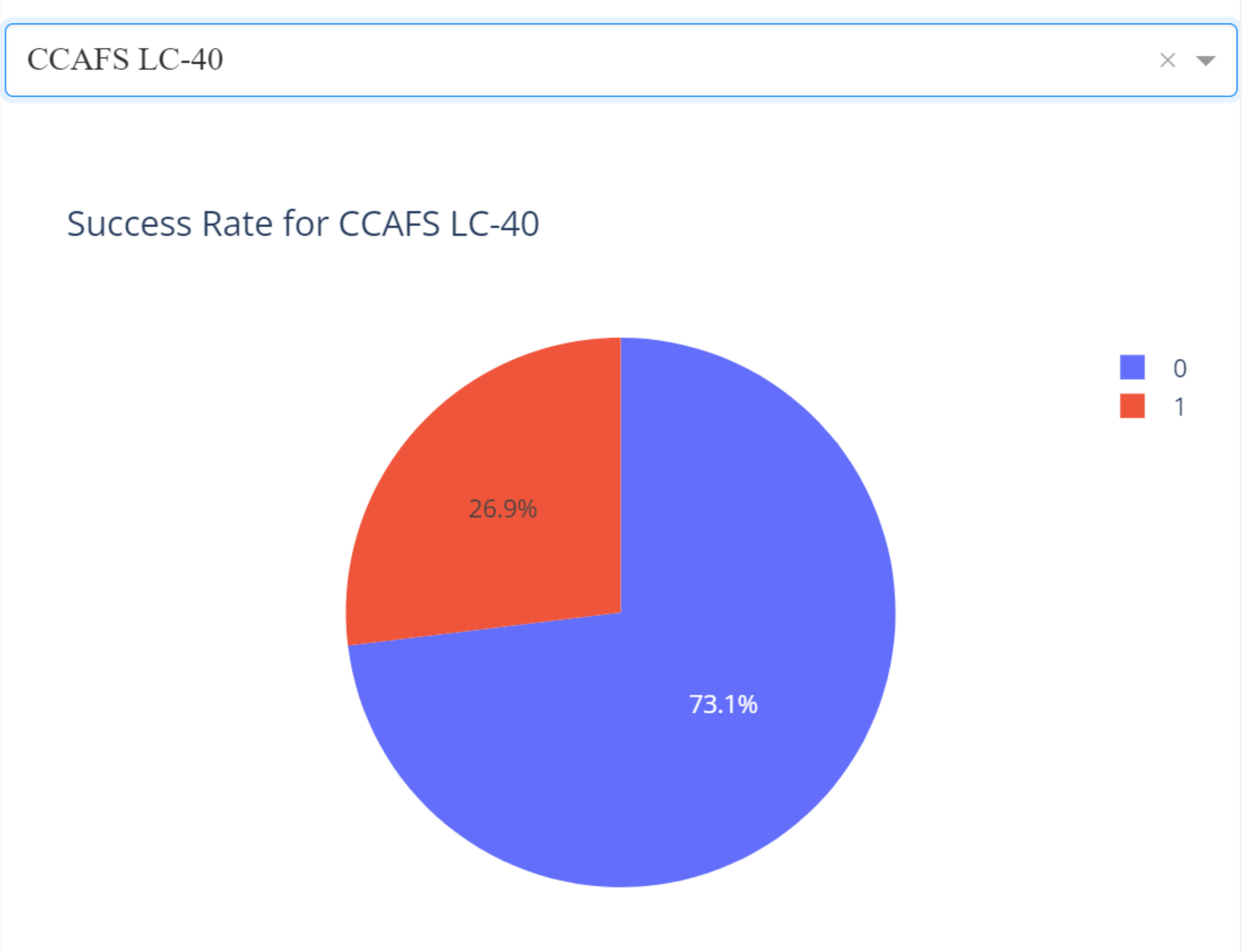
Total Success Launches by Site

- Pie Chart: The chart shows the distribution of total successful SpaceX launches by launch site:
    - CCAFS LC-40: 46.4% (largest share, blue)
    - KSC LC-39A: 23.2% (red)
    - VAFB SLC-4E: 17.9% (green)
    - CCAFS SLC-40: 12.5% (purple)
- Findings:
    - CCAFS LC-40 is the most frequently used site for successful launches, accounting for nearly half of all successes.
    - Other sites (KSC LC-39A, VAFB SLC-4E, CCAFS SLC-40) contribute smaller but significant portions, showing a diversified launch strategy.
    - The data highlights SpaceX's reliance on Cape Canaveral (CCAFS) sites, with CCAFS LC-40 being the dominant location



Highest launch success ratio-CCAFS LC-40

- Success Rate: 73.1% successful landings (class=1), 26.9% failed landings (class=0).
- Reliability: High success rate indicates CCAFS LC-40 is a reliable launch site for SpaceX Falcon 9 missions.



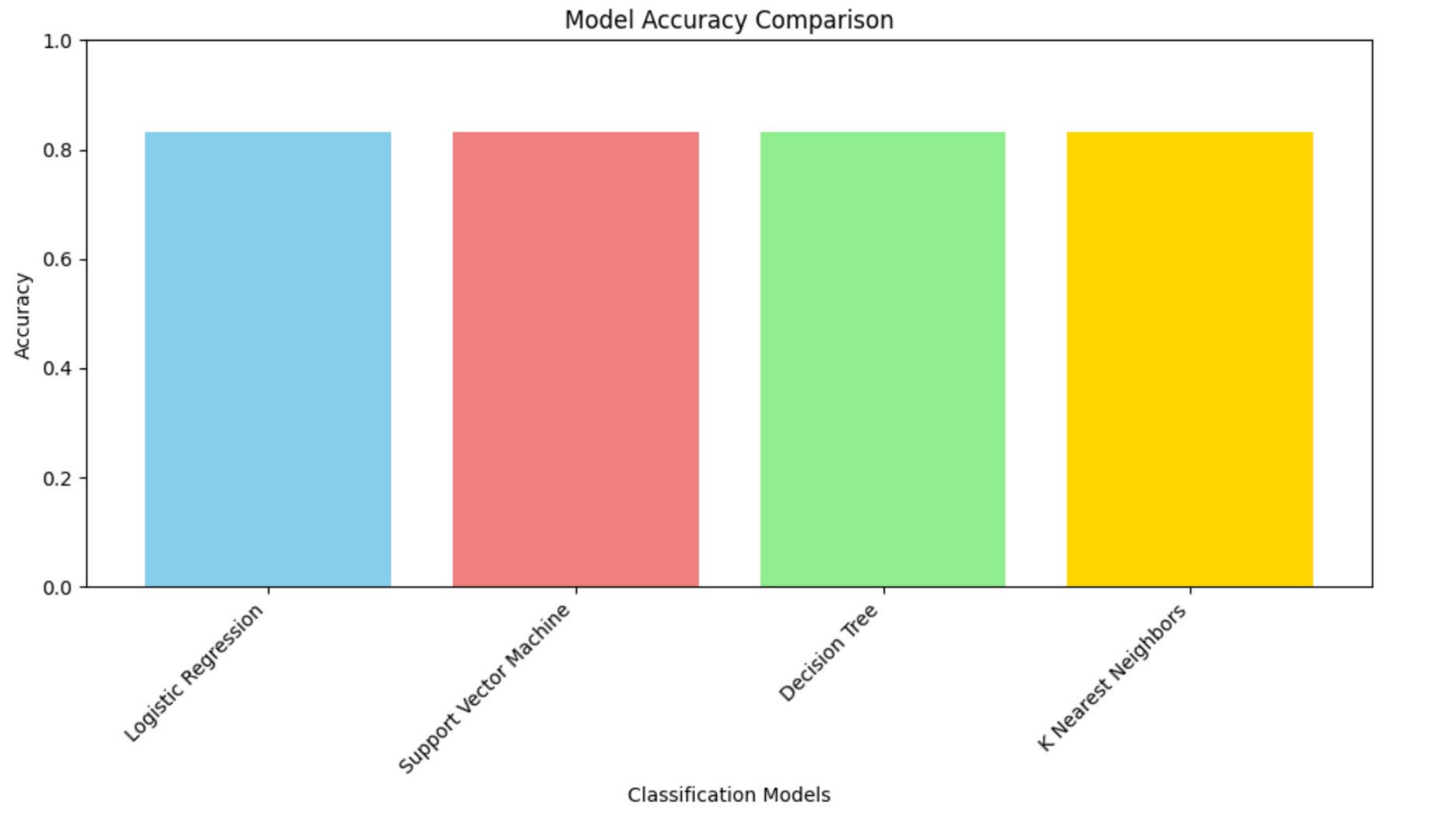
Payload vs. Launch Outcome scatter plot for all sites

- Payload Success Correlation: Higher payloads (up to 10,000 kg) generally show higher success rates (class=1), with most points above 0.5 on the y-axis.
- Booster Version Impact: Success varies by booster version (v1.0, v1.1, B4, B5); B4 and B5 boosters (green, orange) show more successes, while v1.0 and v1.1 (blue, red) have more failures at lower payloads.
- Payload Range: Most launches succeed with payloads between 2,000-10,000 kg; failures (y=0) are common at lower payloads (<2,000 kg).
- Overall Reliability: High success rates across all sites for moderate to heavy payloads, supporting SpaceX’s improving landing technology.

Section 5

Predictive Analysis (Classification)

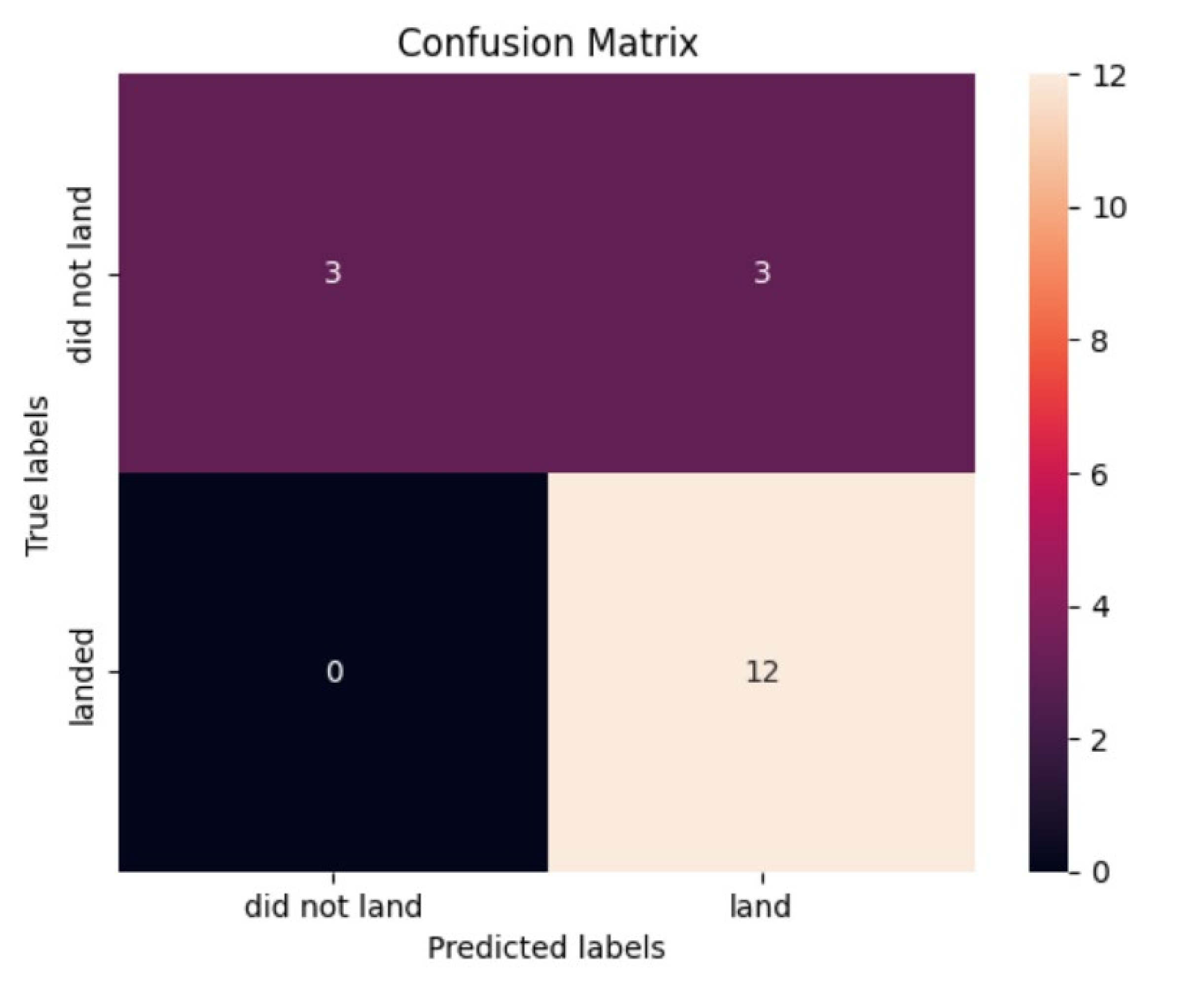
Classification Accuracy



All four models achieve the same accuracy of 83.33% on the test data

Confusion Matrix

- True Postive - 12 (True label is landed, Predicted label is also landed)
- False Postive - 3 (True label is not landed, Predicted label is landed)



Conclusions

- **Key Findings from EDA**: The slide can recap the most significant relationships between flight number, payload mass, launch site, orbit type, and their impact on the success rate of SpaceX Falcon 9 first stage landings. It can highlight which orbits have higher success rates and the general trend of increasing success rates over the years.
- **Insights from Interactive Visual Analytics**: This section summarizes insights gained from Folium maps and Plotly Dash dashboards. For example, the slide could mention the concentration of successful launches at CCAFS SLC-40 and the importance of proximity to coastlines for launch sites.
- **Performance of Classification Models**: All four classification models (Logistic Regression, SVM, Decision Tree, and K-Nearest Neighbors) achieved an accuracy of 83.33% on the test data.
- **Overall Summary**: The conclusion slide should reiterate the initial problem and summarize the main findings, stating the outcome of the analysis. It can also mention potential future steps or areas for further research.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!