IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nathan Osborn

03/01/2026

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

Summary of methodologies

Data was collected via SpaceX API and web scraping, then analyzed using SQL queries, visualizations, Folium maps, and a Plotly Dash dashboard. Four ML models (Logistic Regression, SVM, Decision Tree, KNN) were tuned using GridSearchCV with 10-fold cross-validation.

Analysis of 101 launches showed KSC LC-39A had the highest success rate, with landing success improving over time and Booster V5 performing best. The Decision Tree model achieved the highest accuracy at 88.9%.

Introduction

- **Project background and context**

- **SpaceX** offers **Falcon 9 launches at $62M** compared to **competitors at $165M+,** with savings driven by **first-stage booster reuse**. **Predicting landing success** allows estimation of **true launch costs**, valuable **for competitors bidding against SpaceX**.

- **Problems we want answers to**

- What factors influence **landing success**? Which **launch sites** perform best? How has **success rate changed over time**? Can we **reliably predict** whether a booster will **land successfully**?

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- SpaceX REST API for launch
- Web scraping (BeautifulSoup) from Wikipedia

Performed data wrangling

- Filtered to Falcon 9 launches only
- Handled missing values, created binary target variables

Performed exploratory data analysis (EDA) using visualization and SQL

Performed interactive visual analytics using Folium and Plotly Dash

Performed predictive analysis using classification models

Data Collection

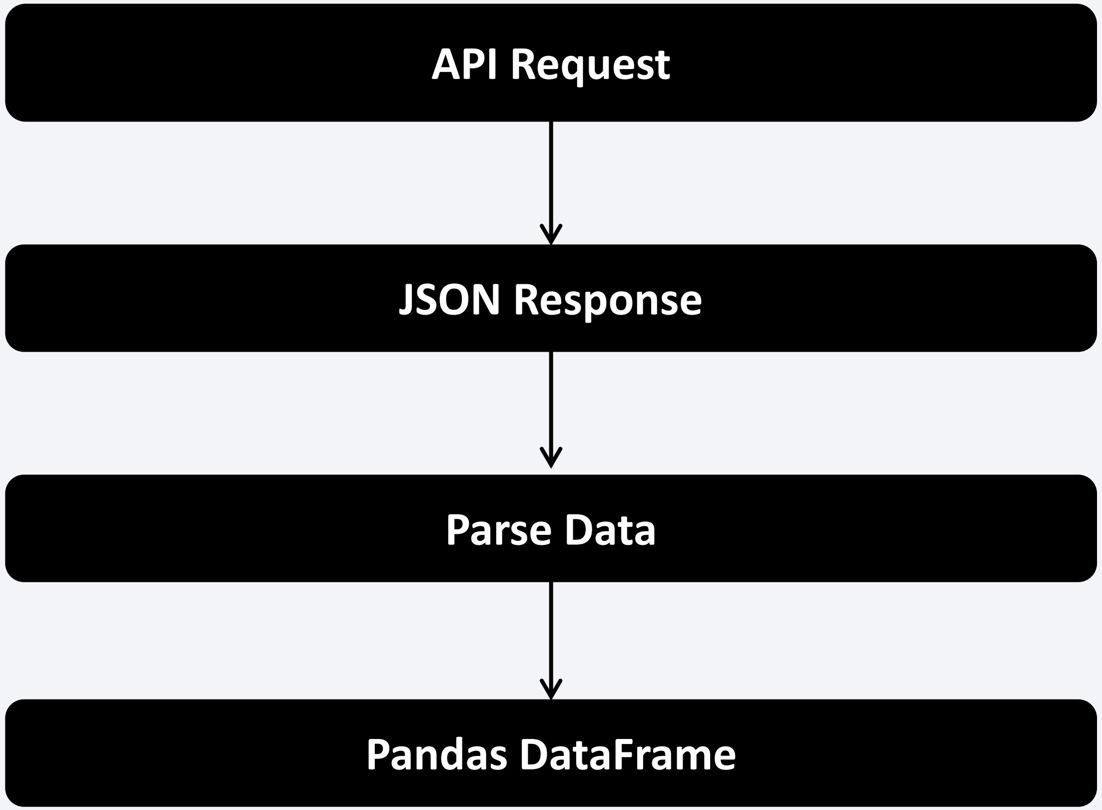
Launch data collected from **SpaceX REST API** (*flight number, date, booster version, payload mass, orbit, launch site, landing outcome*) and supplemented with historical records scraped from **Wikipedia** using BeautifulSoup

Dataset covers **101 Falcon 9 launches (2010-2020)** across four sites: *CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E*

Data wrangling: filtered for **Falcon 9** only, imputed missing **payload mass** with mean, created **binary target variable** (*1 = successful landing, 0 = failure*)

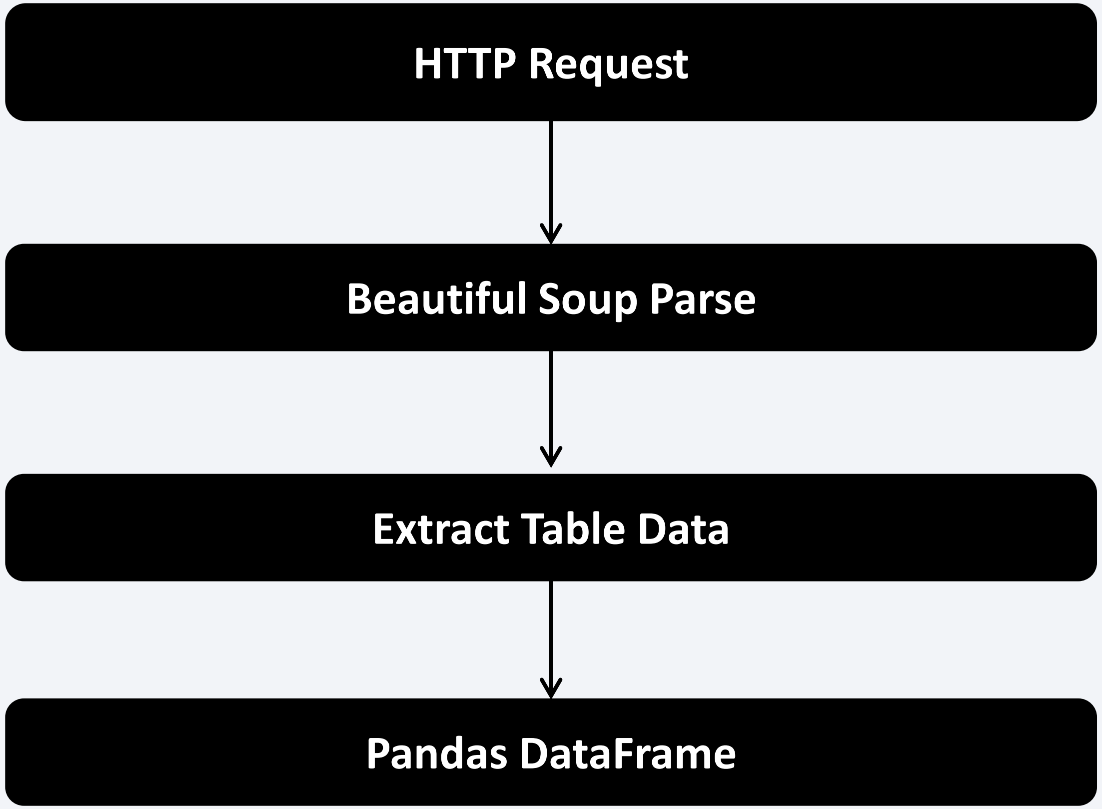
Data Collection – SpaceX API

- Sent GET requests to the **SpaceX REST API** endpoint to retrieve **Falcon 9** launch data. Response returned JSON containing ***flight number, launch date, rocket ID, payload details, launch site, and landing outcome***. Parsed JSON and converted to a Pandas DataFrame for analysis.
- https://github.com/NateOsborn/spacex-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



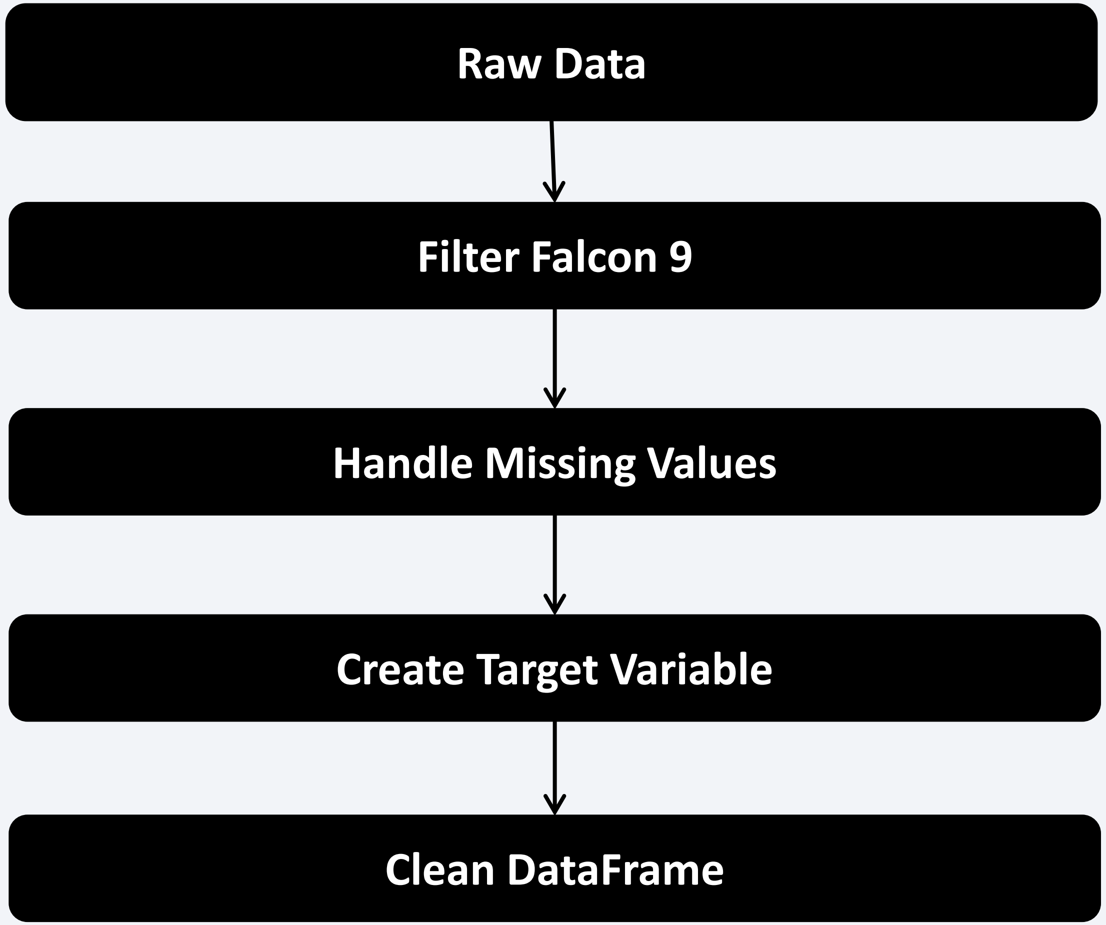
Data Collection - Scraping

- Used **Python Requests** library to fetch **Falcon 9** launch page from **Wikipedia**. Parsed **HTML** content using **BeautifulSoup** to locate and extract launch data from tables. Extracted fields ***including date, booster version, launch site, payload, and outcome.*** Converted parsed data into **Pandas DataFrame** and **merged** with **API data**.
- https://github.com/NateOsborn/spacex-capstone/blob/main/jupyter-labs-webscraping.ipynb



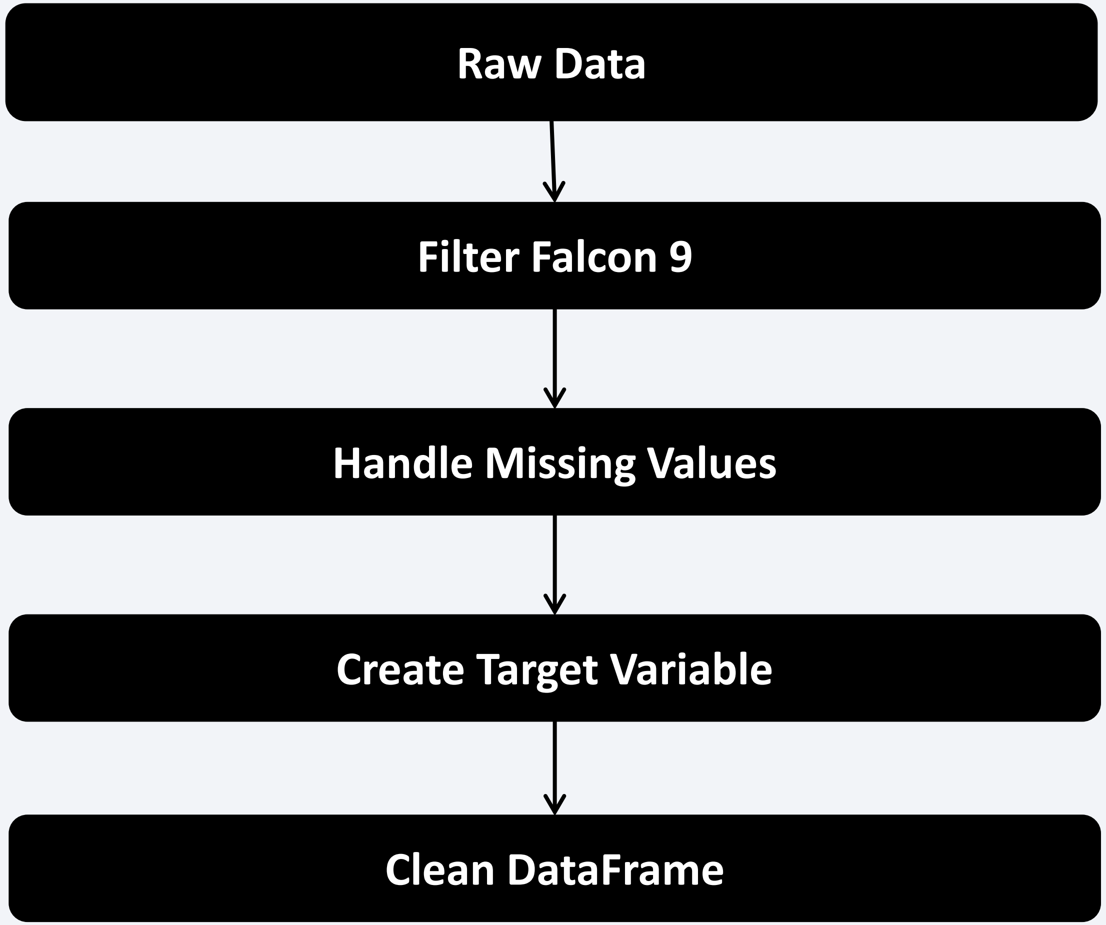
Data Wrangling

- **Filtered dataset** to include only **Falcon 9** launches. Identified and **handled missing values** by replacing null payload masses with the **column mean**. **Converted landing outcome** categories into a **binary target variable:** ***1 for successful landing, 0 for failure***. Removed irrelevant columns and ensured **consistent data** types across all features.
- https://github.com/NateOsborn/spacex-capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Explored **relationships between features** and landing success using **Matplotlib** and **Seaborn**. Created **scatter plots** to examine **payload mass vs. outcome** and **flight number vs. launch site**. Built **line charts** showing **success rate** improving over time. **Analyzed** success rates by ***orbit type, launch site, and booster version*** using **bar charts**.
- https://github.com/NateOsborn/spacex-capstone/blob/main/edadataviz.ipynb



EDA with SQL

- Retrieved **unique launch sites** from the dataset
- Calculated **total payload mass** carried for NASA (CRS) missions: 45,596 kg
- Found **average payload mass** for F9 v1.1 booster: 2,337 kg
- Identified **first successful ground landing** date
- Listed **boosters with successful drone ship landings** for payloads between 4,000-6,000 kg
- Counted **total successful vs. failed missions**: 100 successes, 1 failure
- Found **booster versions carrying maximum payload** (F9 B5 variants)
- Ranked **landing outcomes** by frequency between 2010-2017
- https://github.com/NateOsborn/spacex-capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Added **markers** at each launch site location to show geographic position and display site names on click
- Created **colored markers** to distinguish successful landings (green) from failures (red)
- Drew **circles** around launch sites to visualize proximity to coastline - sites located near water for safety during failed landings
- Added **lines** measuring distance to nearest railway, highway, and city to show transportation and logistics access

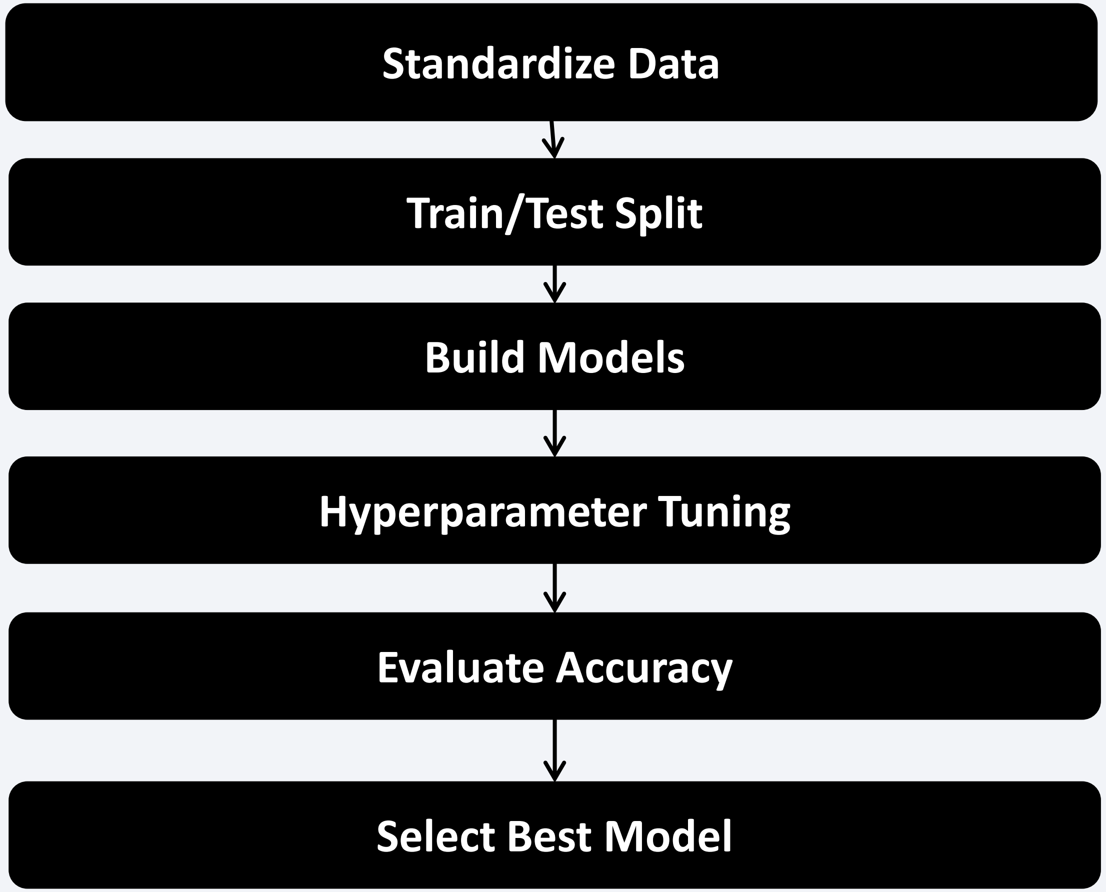
- https://github.com/NateOsborn/spacex-capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Dashboard**
- **Dropdown** filter for launch site, **range slider** for payload mass. **Pie chart** shows success rate by site. **Scatter plot** shows payload vs. outcome colored by booster version.
- **Why?**
- Allows users to ***compare sites, explore payload impact, and identify which booster versions perform best***.
- https://github.com/NateOsborn/spacex-capstone/tree/main/PlotyDashSpaceX

Predictive Analysis (Classification)

- **Standardized features** and **split data 80/20**. Built four models: ***Logistic Regression, SVM, Decision Tree, and KNN***. Used **GridSearchCV** with 10-fold **cross-validation** to **tune hyperparameters**. Evaluated using **accuracy scores** and **confusion matrices**. **Decision Tree** performed best at **88.9% accuracy**.
- https://github.com/NateOsborn/spacex-capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

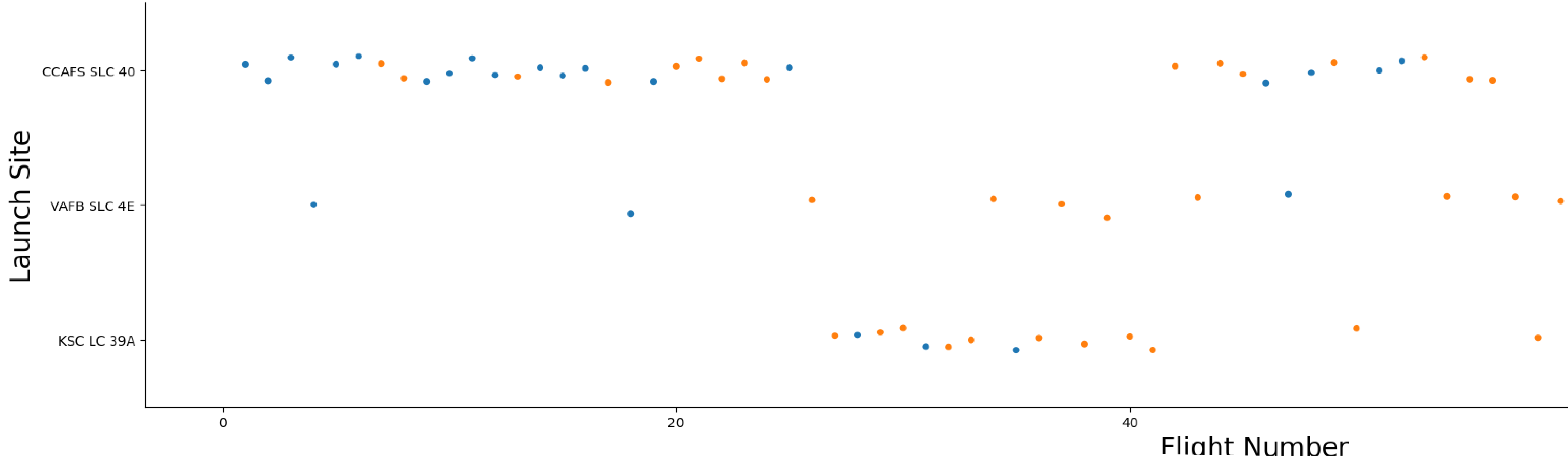


Results

- **Key Findings:**
- Success rate increases with flight number (learning curve effect)
- Highest success orbits: ***ES-L1, GEO, HEO, SSO*** (100%)
- KSC LC-39A launch site has the highest success rate
- Booster reuse and grid fins correlate with successful landings

Section 2

Insights drawn from EDA

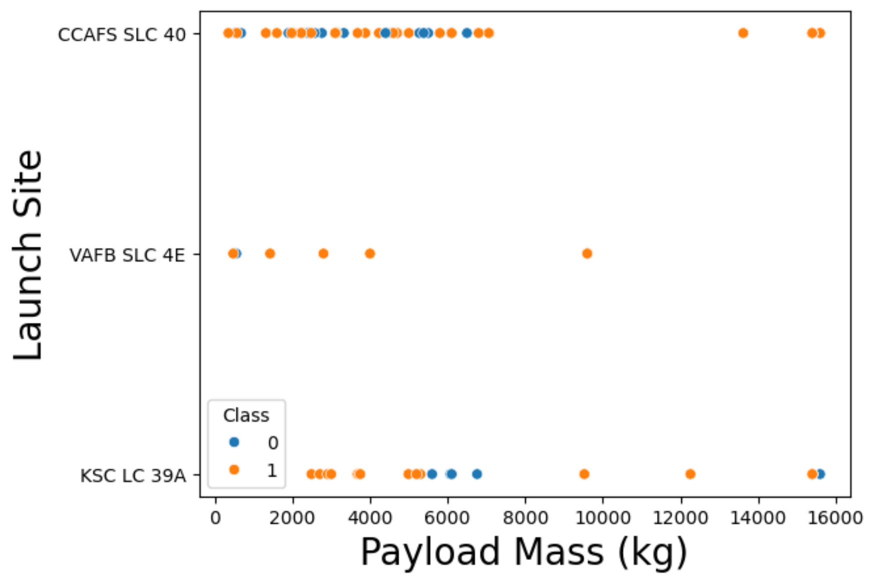


Flight Number vs. Launch Site

- **What it shows:** Landing outcomes (orange = success, blue = failure) across three launch sites over time
- **Key Insights:**
  - **KSC LC-39A:** Highest success rate, while **CCAFS SLC-40** had the most launches
  - **VAFB SLC-4E:** Fewer launches, but consistent success after flight 40
  - **Conclusion:** SpaceX significantly improved landing reliability over time at all launch sites

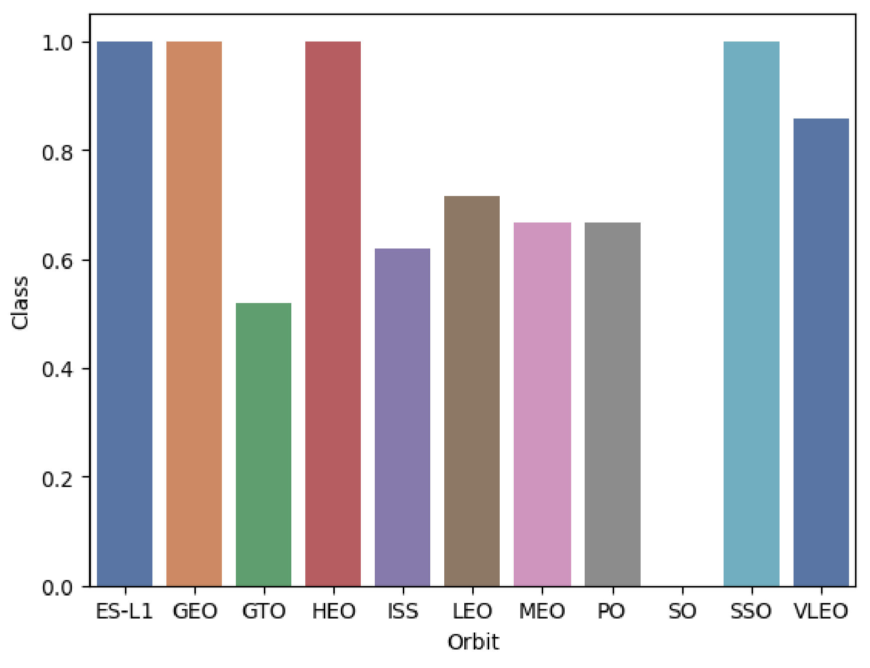
Payload vs. Launch Site

- **CCAFS SLC-40**: Handles full payload range (0–15,000 kg); mixed outcomes across all payload weights
- **VAFB SLC-4E**: Primarily lighter payloads (<10,000 kg); landings successful except for one
- **KSC LC-39A**: Handles wide range including heaviest payloads; mixed success/failure outcomes
- Heavy payloads (10,000+ kg): Only launched from CCAFS SLC-40 and KSC LC-39A with successful outcomes

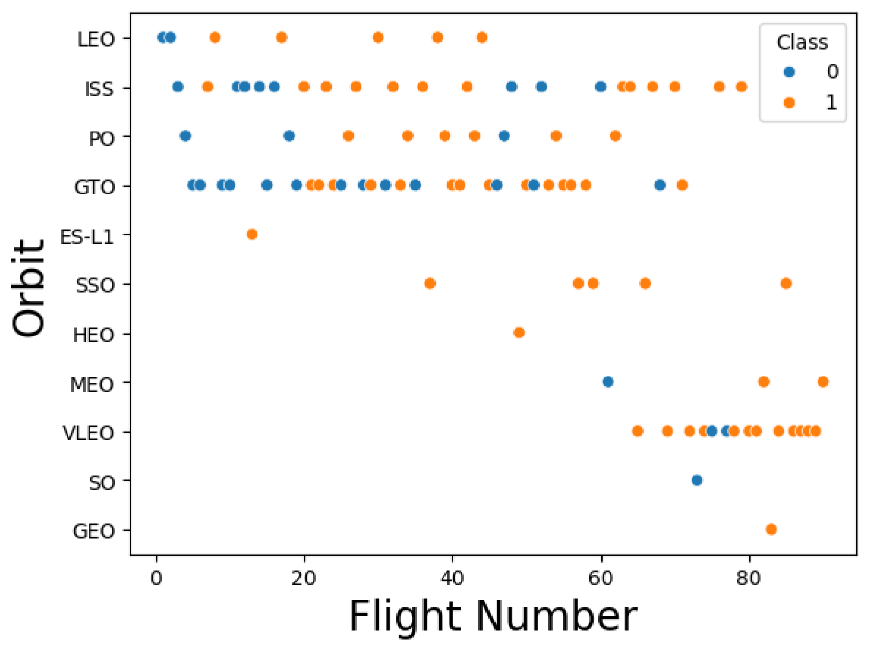


Success Rate vs. Orbit Type

- **What it shows:** Landing success rate (Class = 1) for each orbit type.
- **Key Insights:**
- 100% Success Rate: ***ES-L1, GEO, HEO, SSO orbits***
- High Success (>80%): ***VLEO (~85%)***
- Moderate Success (60-75%): ***LEO (~72%), MEO (~67%), PO (~67%), ISS (~62%)***
- Lowest Success: ***GTO (~52%)***



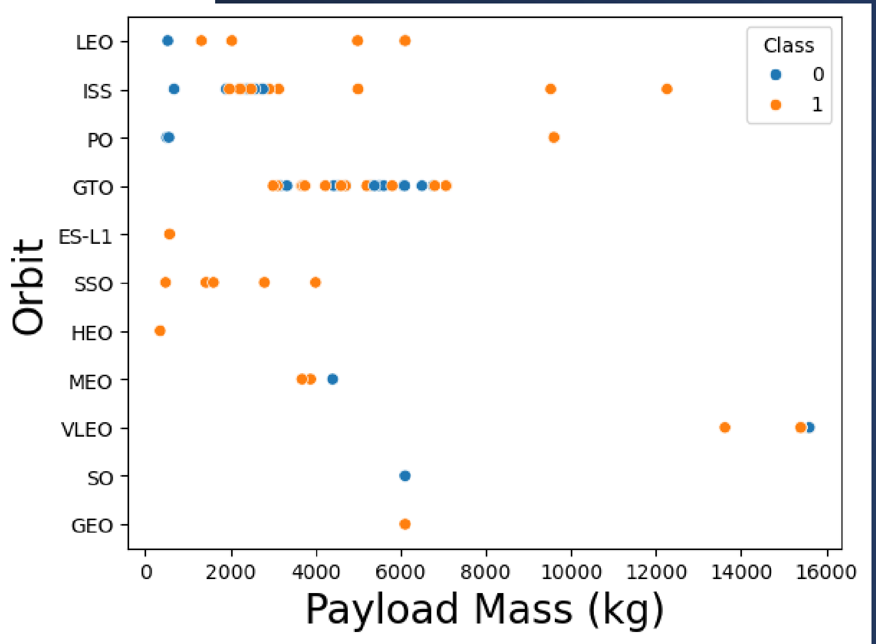
Flight Number vs. Orbit Type



- **What it shows:** Landing outcomes (orange = success, blue = failure) by orbit type over time (flight number).
- **Key Insights:**
- **GTO orbit:** Most frequent mission type; early flights show failures, improving success rate over time
- **ISS orbit:** Consistent missions throughout; mixed outcomes with clear improvement after flight ~60
- **LEO orbit:** Early missions (flights 1-10) had failures; later flights show success
- **VLEO, SSO, GEO, HEO, ES-L1:** Introduced in later flights (40+); predominantly successful outcomes

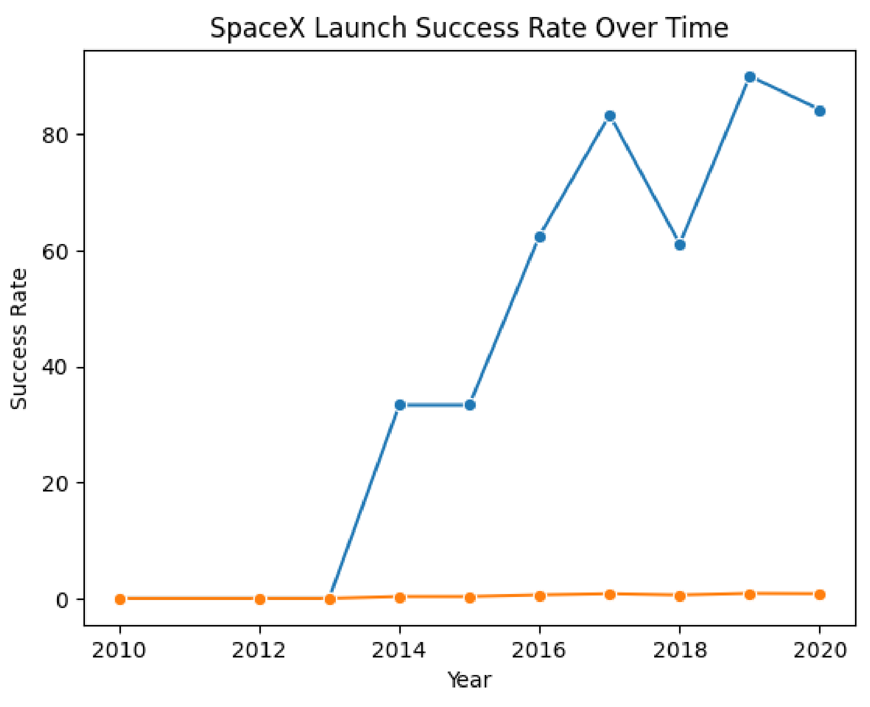
Payload vs. Orbit Type

- **What it shows:** Landing outcomes (orange = success, blue = failure) by payload mass for each orbit type.
- **Key Insights:**
- **GTO orbit:** Most failures concentrated in the 2,000–6,000 kg payload range
- **ISS orbit:** Mixed outcomes, mostly failures in lighter payloads
- **LEO orbit:** Light payloads <2,000 kg show failures; heavier LEO payloads ~10,000 kg successful
- **VLEO orbit:** Heaviest payloads ~14,000–15,500 kg; all successful landings
- **SSO, HEO, GEO, ES-L1:** Lighter payloads <6,500 kg were all successful



Launch Success Yearly Trend

- **What it shows:** The launch success rate (blue line) and the launch years (orange line)
- **Key Findings:**
- **Yearly Trend:** From 2013 the success rate has shown steady improvement.



All Launch Site Names

- Four Launch Sites:
- **CCAFS LC-40**
- **VAFB SLC-4E**
- **KSC LC-39A**
- **CCAFS SLC-40**

- **SQL query to find all launch site names:**
  - SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE

- **SQL query for finding five records for launch sites starting with ‘CCA’**:
  - SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5
- **Record 1:** 2010-06-04 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | LEO | SpaceX | Success | Failure (parachute)
- **Record 2:** 2010-12-08 | F9 v1.0 B0004 | CCAFS LC-40 | Dragon demo flight C1 | LEO (ISS) | NASA (COTS) NRO | Success | Failure (parachute)
- **Record 3:** 2012-05-22 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | LEO (ISS) | NASA (COTS) | Success | No attempt
- **Record 4:** 2012-10-08 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | LEO (ISS) | NASA (CRS) | Success | No attempt
- **Record 5:** 2013-03-01 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | LEO (ISS) | NASA (CRS) | Success | No attempt

Launch Site Names Begin with 'CCA'

- **SQL Query:**
  - SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA (CRS)’
- **Result:**
  - **Total Payload Mass:** 45,596 kg
- **Explanation:**
- NASA (CRS) refers to NASA's Commercial Resupply Services program
- This program contracts SpaceX to deliver cargo to the International Space Station (ISS)
- This demonstrates SpaceX's critical role as a logistics partner for NASA's space station operations

Total Payload Mass

- **SQL Query:**
  - SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" LIKE 'F9 v1.1%'
- **Result:** Average Payload Mass: 2,337.8 kg
- **Explanation:**
- The average payload of 2,337.8 kg (~2.3 metric tons) per mission reflects the typical cargo capacity utilized during F9 v1.1 missions
- This version was used primarily for International Space Station (ISS) resupply missions and satellite deployments

Average Payload Mass by F9 v1.1

- **SQL query:** For the date of the first successful landing outcome on ground pad
    - SELECT MIN("Date") FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE 'Success’
- **Answer:**
    - 2010-06-04

First Successful Ground Landing Date

Successful Drone Ship Landing with Payload between 4000 and 6000

- **SQL Query:**
  - SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000
- **Results:**
  - **Booster Version:**
  - F9 FT
  - B1022F9 FT
  - B1026F9 FT
  - B1021.2F9 FT
  - B1031.2
- **Explanation:** Four Falcon 9 Full Thrust (F9 FT) boosters met these criteria. The ".2" suffix on B1021 and B1031 indicates these were reused boosters on their second flight, demonstrating SpaceX's successful rocket reusability program. All four missions carried medium-weight payloads (4,000–6,000 kg) and landed on autonomous drone ships at sea.

Total Number of Successful and Failure Mission Outcomes

- **SQL Query:**
  - SELECT SUM(CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS Success_Count, SUM(CASE WHEN "Mission_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 END) AS Failure_Count FROM SPACEXTABLE;
- **Result:**
  - Success Count 100
  - Failure Count 1
- **Explanation:** SpaceX achieved an exceptional 99% mission success rate, with 100 successful missions and only 1 failure (the CRS-7 mission in June 2015). However, these successful mission outcomes did not always result in a successful landing.

Boosters Carried Maximum Payload

- **SQL Query:**
  - SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
- **Result:** Multiple F9 B5 boosters carried the maximum payload of 15,600 kg, including:F9 B5 B1048.4, B1049.4, B1051.3, B1056.4, B1048.5, B1051.4, B1049.5, B1060.2, B1058.3, B1051.6
- **Explanation:** The Falcon 9 Block 5 (B5) variant consistently handles the heaviest payloads, demonstrating SpaceX's engineering improvements across booster generations.

2015 Launch Records

- **SQL Query:**
  - SELECT "Booster_Version", "Launch_Site", "Date"FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)'   AND "Date" LIKE '2015%';
- **Result:**
  - 2015-01-10F9 v1.1 B1012CCAFS LC-40
  - 2015-04-14F9 v1.1 B1015CCAFS LC-40
- **Explanation:** Both 2015 drone ship landing failures occurred at **Cape Canaveral (CCAFS LC-40)** using **F9 v1.1** boosters. These early attempts were critical learning experiences before SpaceX achieved consistent landing success.

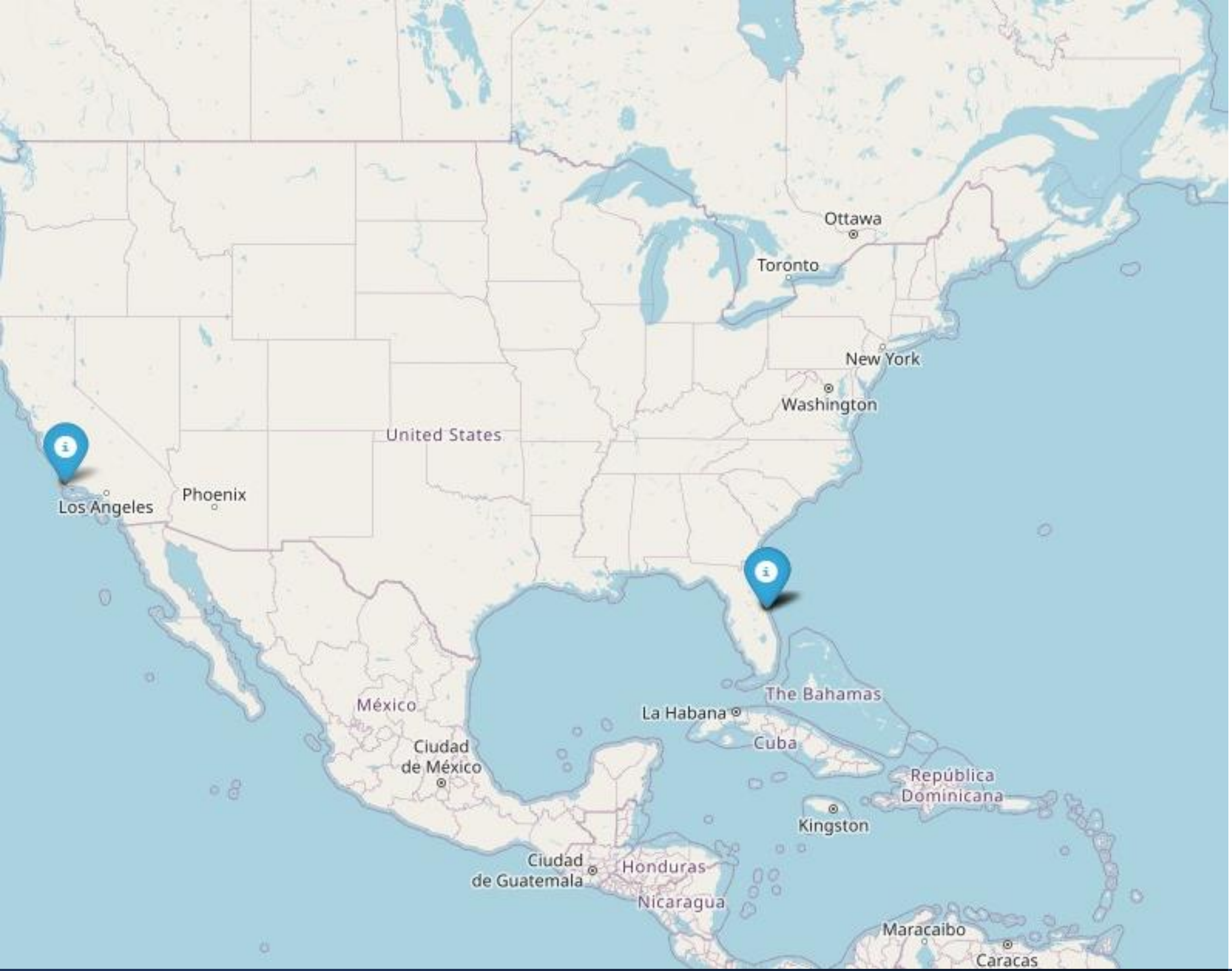
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- **SQL Query:**
  - SELECT RANK() OVER (ORDER BY COUNT(*) DESC) AS Rank, "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;
- **Explanation:** Early SpaceX missions (2010-2017) show the evolution from no landing attempts to achieving 50% drone ship success rate. Ground pad landings had higher success (3 successes vs. limited attempts).

Rank	Landing Outcome	Count
1	No attempt	10
2	Success (drone ship)	5
2	Failure (drone ship)	5
4	Success (ground pad)	3
4	Controlled (ocean)	3
6	Uncontrolled (ocean)	2
6	Failure (parachute)	2
8	Precluded (drone ship)	1

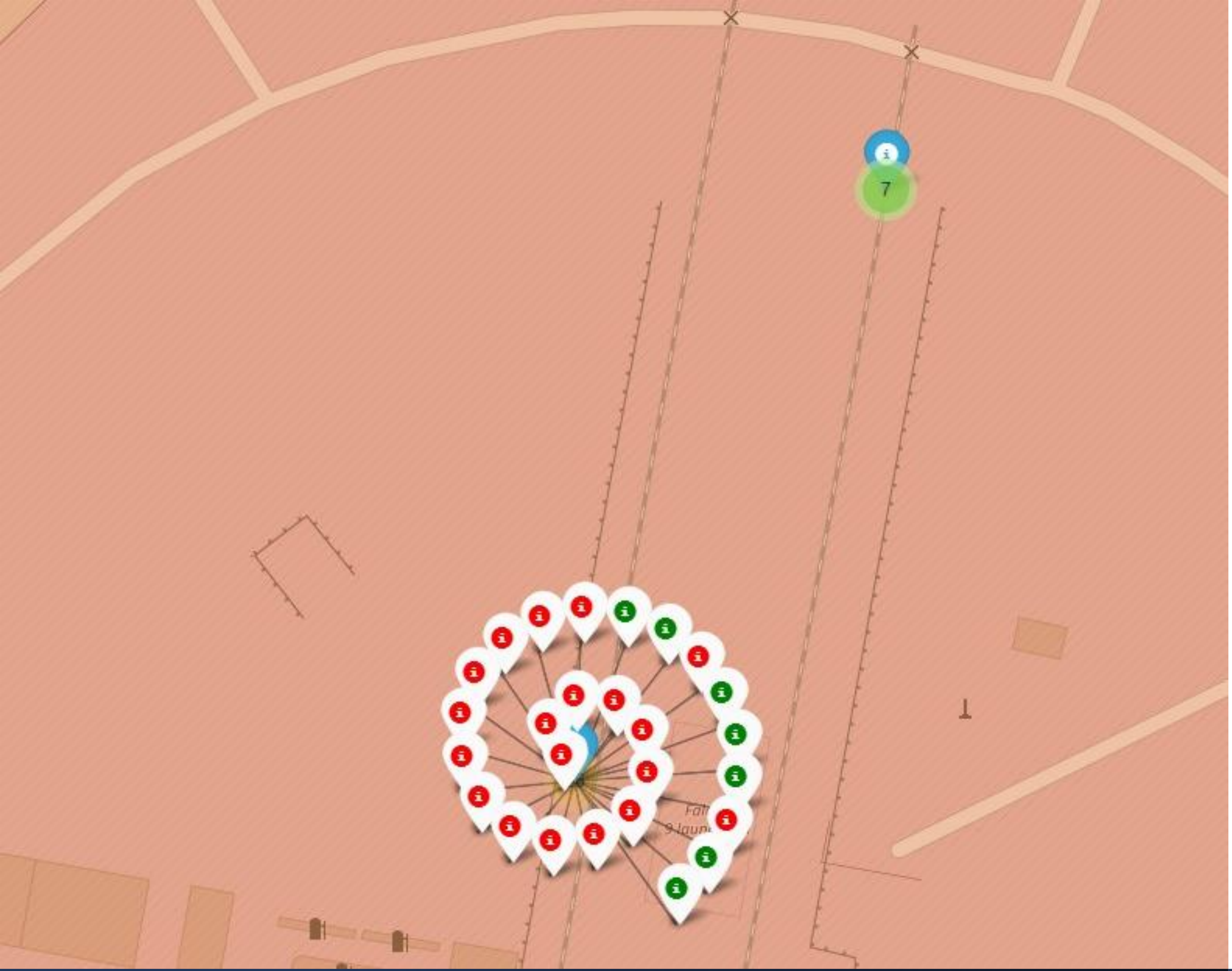
Section 3

Launch Sites Proximities Analysis



Launch Site Location on Folium Map

- **Two Main Launch Locations:**
  - **Florida (East Coast):** Cape Canaveral/Kennedy Space Center - handles majority of launches
  - **California (West Coast):** Vandenberg Space Force Base (VAFB SLC-4E) - used for polar orbit missions
- **Strategic Positioning:** Both sites are coastal, allowing rockets to fly over water for safety
- **Florida Advantage:** Closer to equator = more efficient for geostationary orbit launches



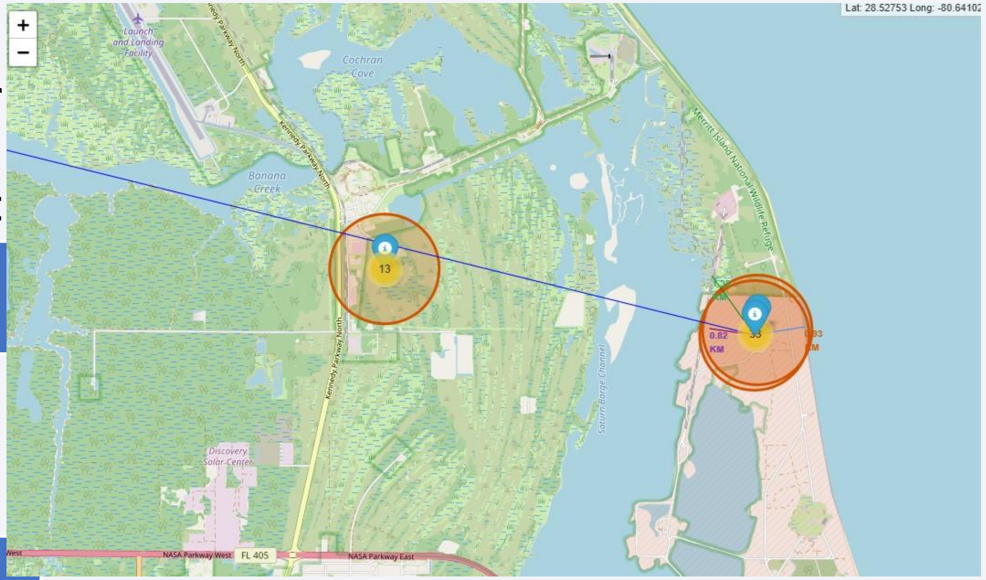
Color-labeled Launch Outcomes on Map

- **Color Coding:**
- Green markers = Successful landings
- Red markers = Failed landings
- **Pattern:** Mix of successes and failures at both sites, with success rate improving over time
- **Cluster at CCAFS:** Higher volume of launches, more varied outcomes.

Infrastructure Proximities on Folium Maps

Launch Sites Near Key Infrastructure:

- **Distance to coastline:** ~0.82 km (launches occur over water for safety)
- Proximity to NASA Parkway and FL-405 highway for logistics
- Launch and Landing Facility visible to the northwest



Strategic Factors:

- Coastal location allows debris to fall in ocean
- Highway access enables transport of rocket stages
- Nearby facilities support operations and recovery

Safety Buffer:

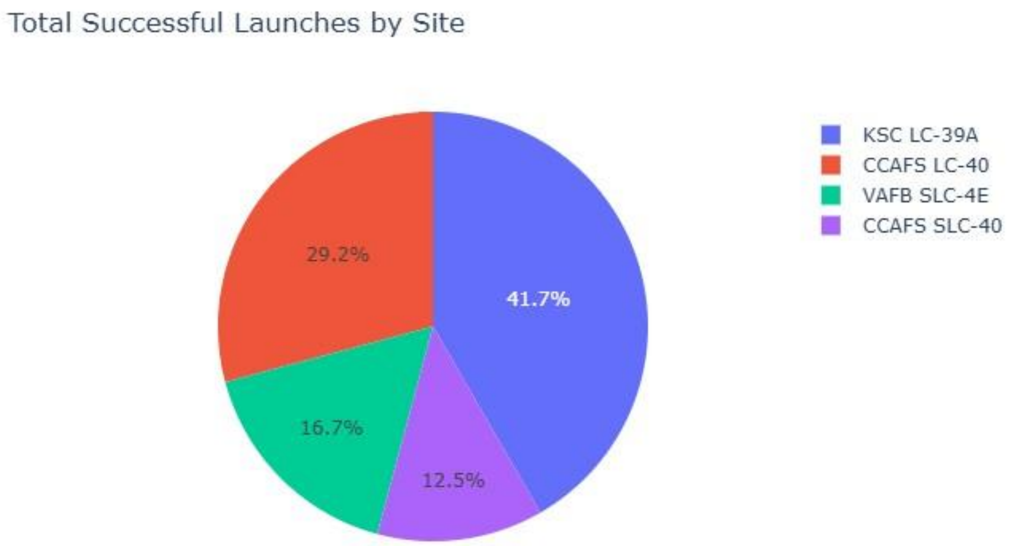
- Launch pads positioned away from populated areas

Section 4

Build a Dashboard with Plotly Dash

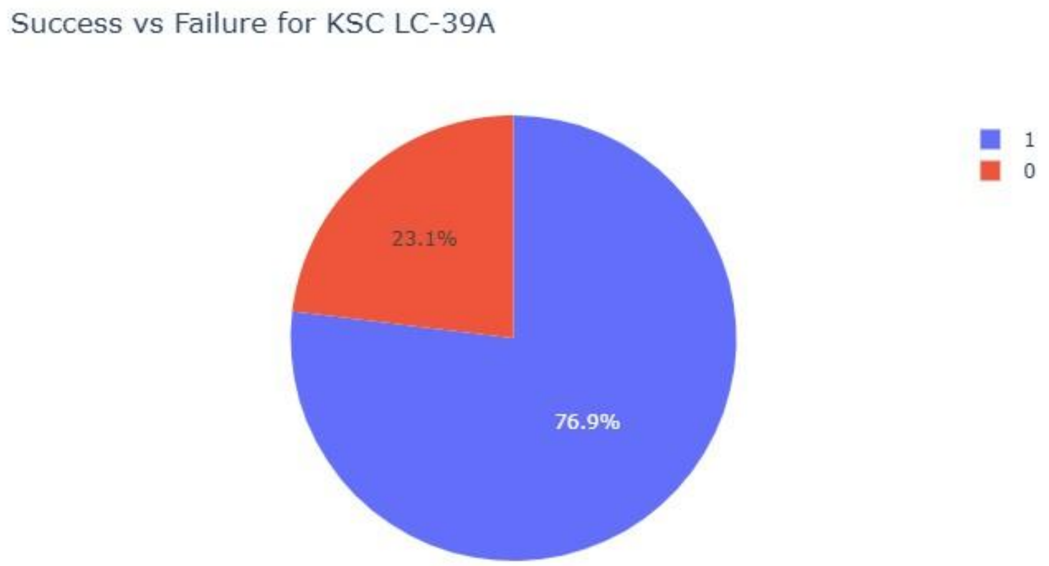
Plotly Dash Interactive Pie Chart (All sites)

- **Key Findings from Dashboard Screenshot:**
- **KSC LC-39A:** 41.7% of successful launches (largest share)
- **CCAFS SLC-40:** 29.2% of successful launches
- **VAFB SLC-4E:** 16.7% of successful launches
- **CCAFS LC-40:** 12.5% of successful launches
- **Explanation:** Kennedy Space Center's LC-39A (historic Apollo launch pad) has become SpaceX's most successful site, handling the plurality of missions. The Florida sites combined account for ~83% of all successful launches.



Interactive Pie Chart for Highest Launch Site Success

- **Success Rate:** 76.9%
- **Failure Rate:** 23.1%
- **Explanation: KSC LC-39A** has the highest launch success ratio among all SpaceX sites. This historic pad (used for Apollo and Space Shuttle missions) benefits from mature infrastructure, experienced personnel, and **optimal geographic positioning** for most mission profiles.

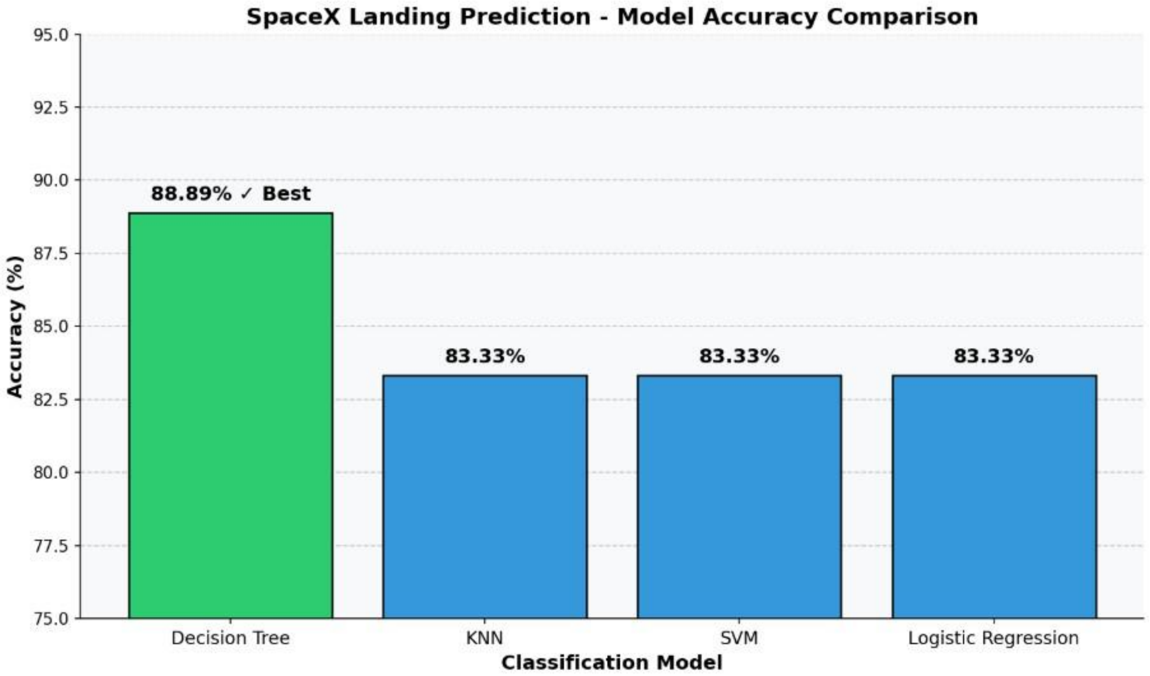


Payload vs. Launch Outcome (All sites)

- **Payload Range:** 0 - 10,000 kg shown
- **Success Pattern (Class=1):**
  - Higher success rates in 2,000-5,000 kg range
  - Very high success for heavy payloads (9,000+ kg) with B5 boosters
- **Booster Evolution:**
  - v1.0/v1.1 (blue/orange): Mixed results, more failures
  - FT (green): Improved success rate
  - B4/B5 (purple/orange): Highest success rates
- **Key Insight:** Newer booster versions (B4, B5) show consistently successful outcomes regardless of payload mass

Section 5

Predictive Analysis (Classification)

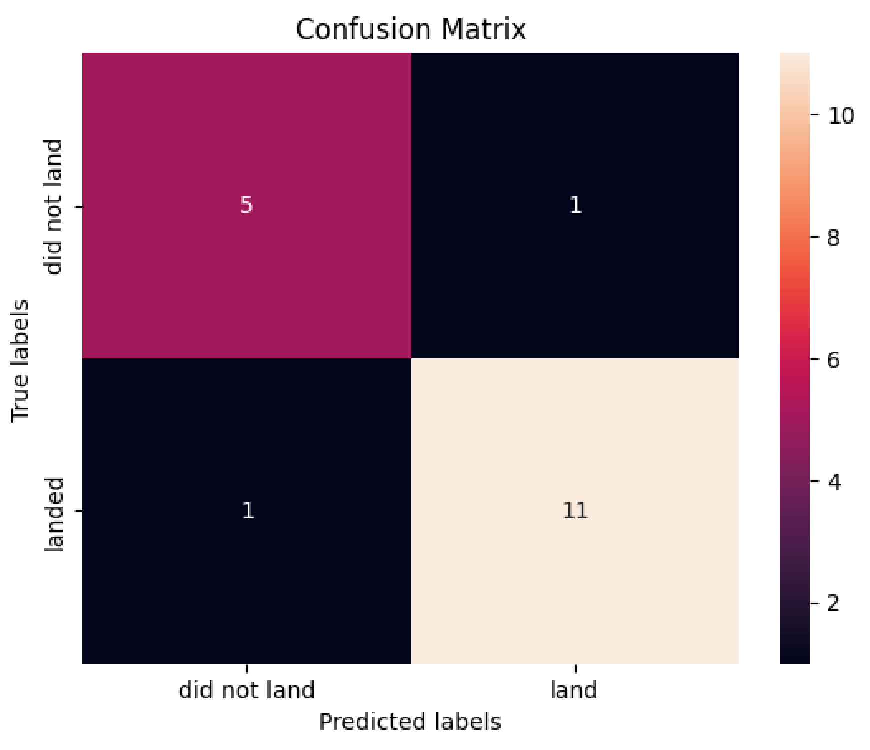


- The **Decision Tree Classifier** achieved the highest accuracy at **88.89%**, outperforming other models by ~5.5 percentage points.

Classification Accuracy

Confusion Matrix

- **Metrics:**
    - **Accuracy:** 88.89% (16/18 correct)
    - **Precision:** 91.7% (11/12)
    - **Recall:** 91.7% (11/12)
    - **True Negatives:** 5 - Correctly predicted non-landings
    - **True Positives:** 11 - Correctly predicted landings
- **Explanation:** The model excels at predicting successful landings with only **2 misclassifications** out of 18 test samples.



Conclusions

- **Data Collection & Processing:**
  - Successfully gathered SpaceX launch data via API, web scraping, and SQL databases
  - Dataset contains 90 Falcon 9 launches with 17 key features
- **Exploratory Analysis:**
  - KSC LC-39A has the highest success rate (76.9%)
  - Success rate improved significantly with newer booster versions (B4, B5)
  - Payload mass between 2,000-5,000 kg shows optimal success rates
- **Overall mission success rate:** 99% (100 successes, 1 failure)
- **Geographic Insights:**
  - Coastal launch sites enable safe over-water trajectories
  - Florida sites handle 83% of successful launches
  - Proximity to infrastructure supports efficient operations
- **Predictive Modeling:**
  - Decision Tree achieved best accuracy: 88.89%
  - Model correctly predicts landing outcomes with 91.7% precision
  - Key predictors: Payload mass, orbit type, launch site, booster version
- **Business Impact:**
  - Predicting landing success helps estimate launch costs ($62M reusable vs. $165M expendable)
  - Enables competitive bidding against SpaceX for rocket launch contracts

Appendix

- **Notebooks:**
- jupyter-labs-spacex-data-collection-api.ipynb - API data collection
- jupyter-labs-webscraping.ipynb - Web scraping from Wikipedia
- labs-jupyter-spacex-Data_wrangling.ipynb - Data cleaning & preparation
- jupyter-labs-eda-sql-coursera_sqllite.ipynb - SQL analysis
- edadataviz.ipynb - EDA & visualization
- lab_jupyter_launch_site_location.ipynb - Folium mapping
- SpaceX_Machine_Learning_Prediction_Part_5.ipynb - ML models

Thank you!