IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Debarjun Chakraborty
9th April, 2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Objective: Predict Falcon 9 first-stage landing success to optimize launch cost estimation.

- Significance: SpaceX’s competitive edge stems from reusability; accurate predictions aid resource allocation.

- Key Question: Can historical launch data determine if the first stage will land successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
  - API (JSON) + Web Scraping (Wikipedia).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Handling missing values, labeling outcomes (1/0).

Data Collection

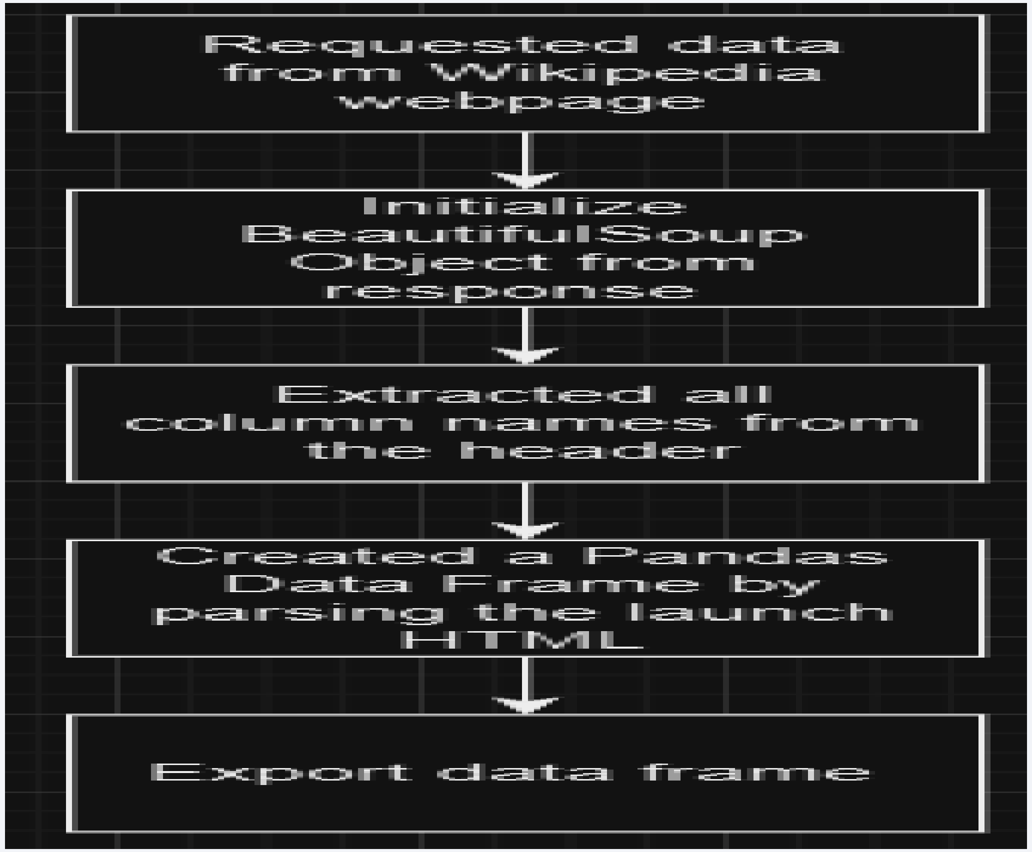
- Sources: SpaceX API (rocket specs, launch sites) + Wikipedia tables (historical launches).
- Tools: Python's requests, BeautifulSoup, Pandas for parsing and storage.

Data Collection – SpaceX API

- Collected rocket, launchpad, payload, and core data via SpaceX REST API (JSON format).
- Filtered Falcon 9 launches and cleaned payload mass data.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

Data Collection - Scraping

- Scraped Falcon 9 launch history from Wikipedia using BeautifulSoup.
- Parsed HTML tables into structured DataFrames for analysis.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git



Data Wrangling

- Labeled landing outcomes: 1 (successful—ASDS, RTLS, Ocean) or 0 (failed—None, False attempts).
- Analyzed launch sites, orbits, and mission outcomes.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

EDA with Data Visualization

- Key plots: Flight Number vs. Payload Mass, Launch Site, Orbit Type.
- Insights: Higher flight numbers correlate with success; certain orbits (ES-L1, GEO) have 100% success rates.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

EDA with SQL

Queries revealed:

- 4 unique launch sites (e.g., CCAFS LC-40).
- NASA's total payload: 45,596 kg.
- First successful ground landing: December 22, 2015.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

Build an Interactive Map with Folium

- Mapped launch sites with success/failure markers.
- Proximity analysis: Sites near coastlines/highways showed varied success ratios.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

Build a Dashboard with Plotly Dash

- Pie charts: KSC LC-39A had the highest success rate (76.7%).

- Scatter plots: FT booster versions succeeded most with payloads under 10,000 kg.
- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

Predictive Analysis (Classification)

- Tested 4 models; all achieved 83.33% accuracy (100% for successful landings).
- Best performers: Logistic Regression, KNN (faster training than SVM).

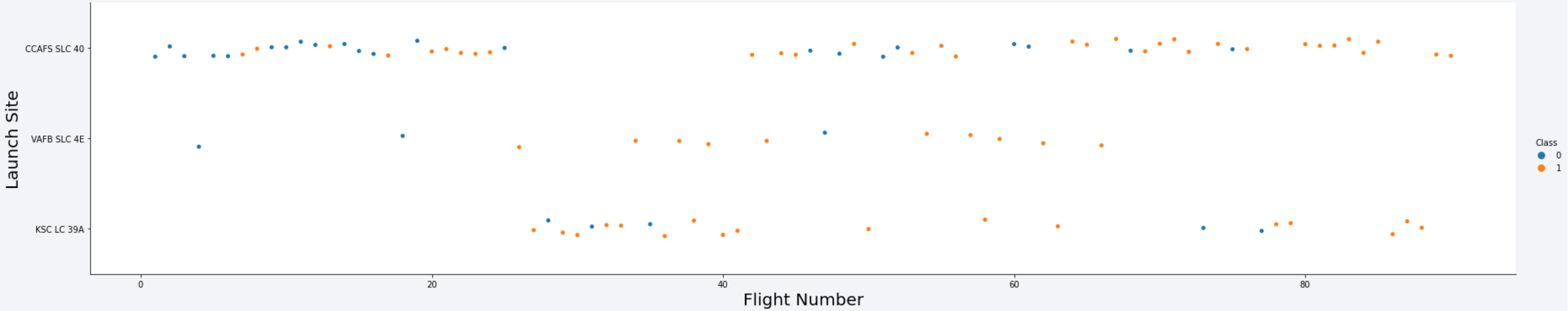
Results

- Flight Number vs. Launch Site: Success increased over time, especially at CCAFS SLC 40.
- Payload vs. Orbit: Heavy payloads (>10,000 kg) had higher success in LEO/ISS orbits.
- Yearly Trend: Success rates rose from 2013–2017, dipped in 2018, and recovered by 2020.

Section 2

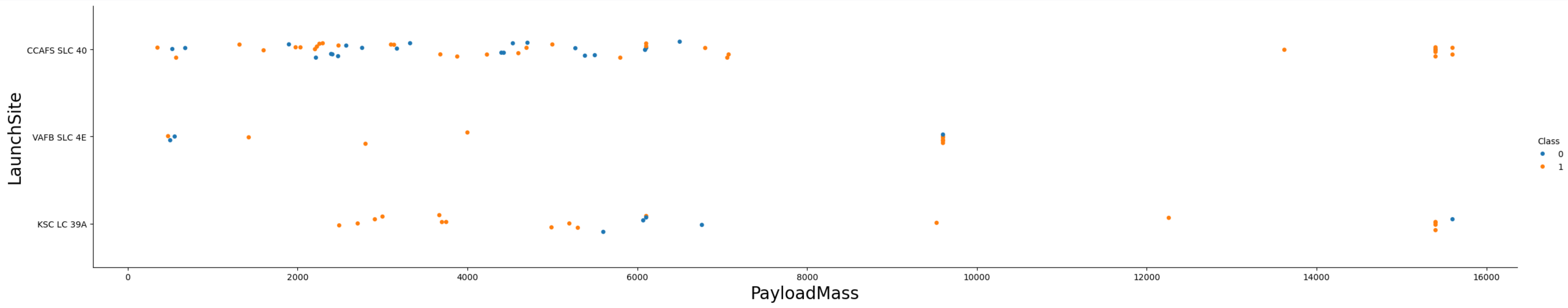
Insights drawn from EDA

Flight Number vs. Launch Site



- Higher flight numbers (more recent launches) show increased success rates across all launch sites.
- CCAFS SLC 40 has the most failures but also the most launches, with failures concentrated in earlier flights.
- Success trends suggest improved reliability over time.

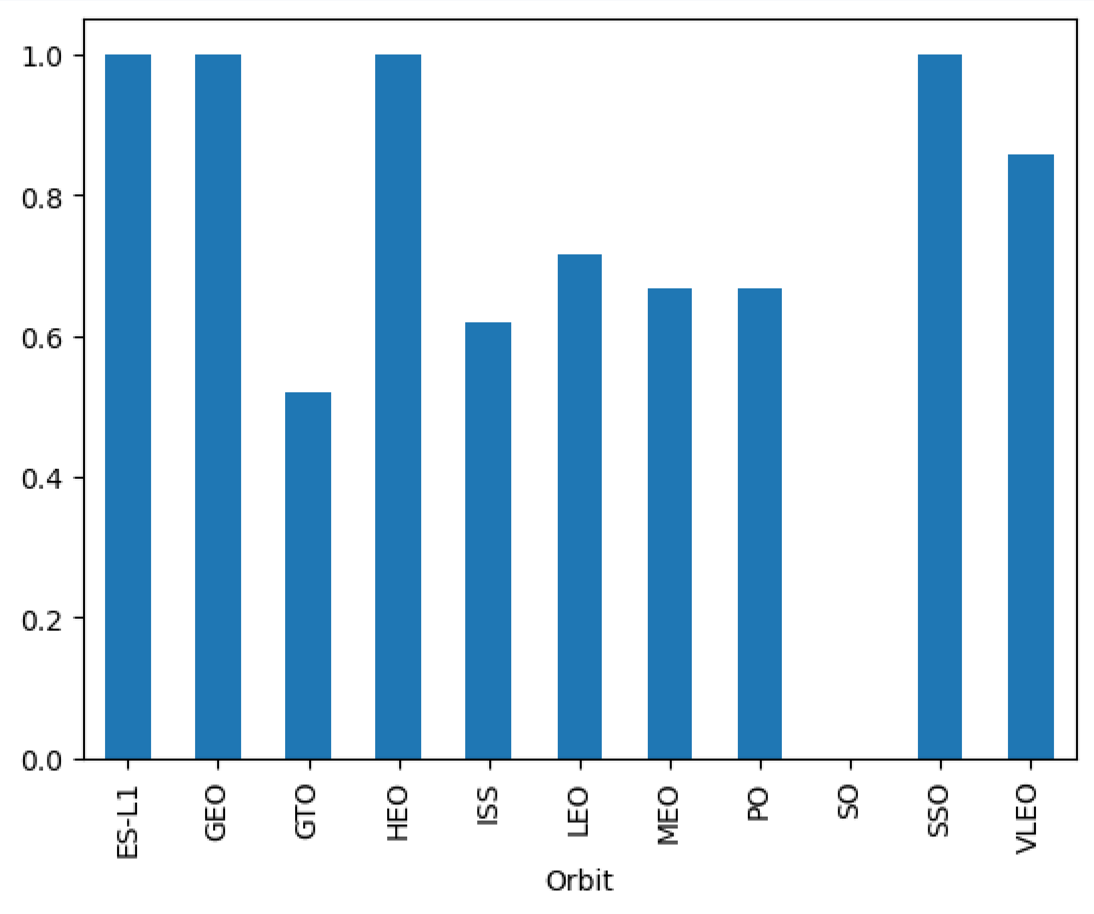
Payload vs. Launch Site



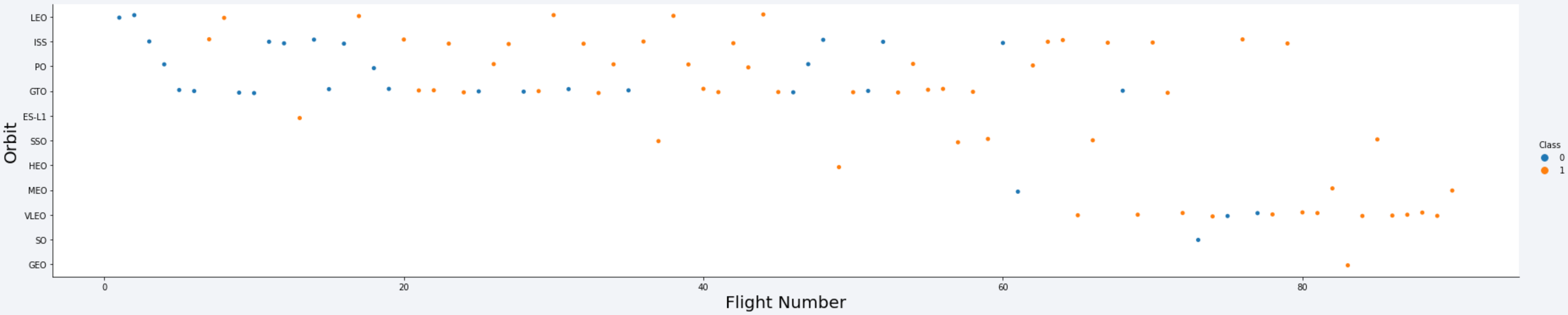
- VAFB SLC 4E has no heavy payload launches (>10,000 kg).
- Failures at other sites mostly occur with payloads below 10,000 kg.
- Payload mass may influence landing success but isn’t the sole factor.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits have a 100% success rate.
- VLEO shows >80% success; GTO, ISS, and LEO have moderate rates.
- SO orbit has no successful landings (only 1 attempt).

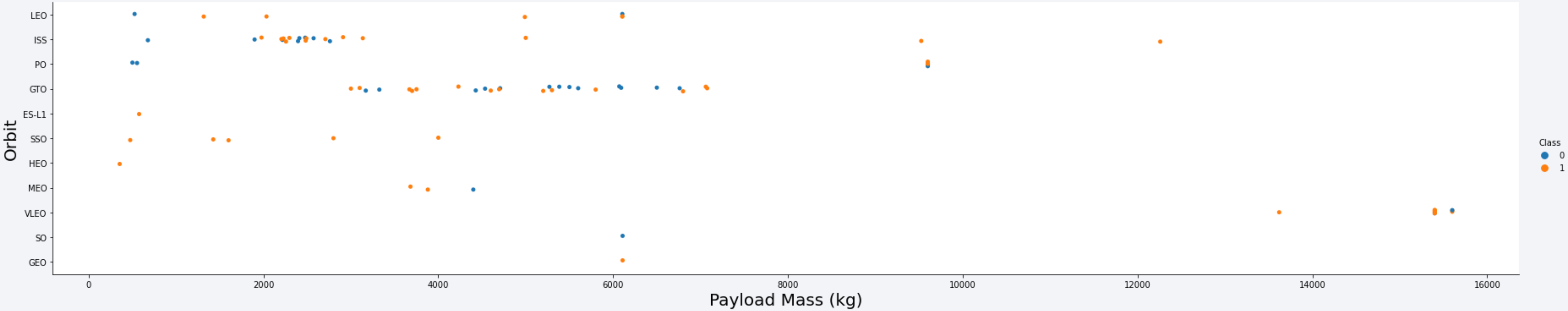


Flight Number vs. Orbit Type



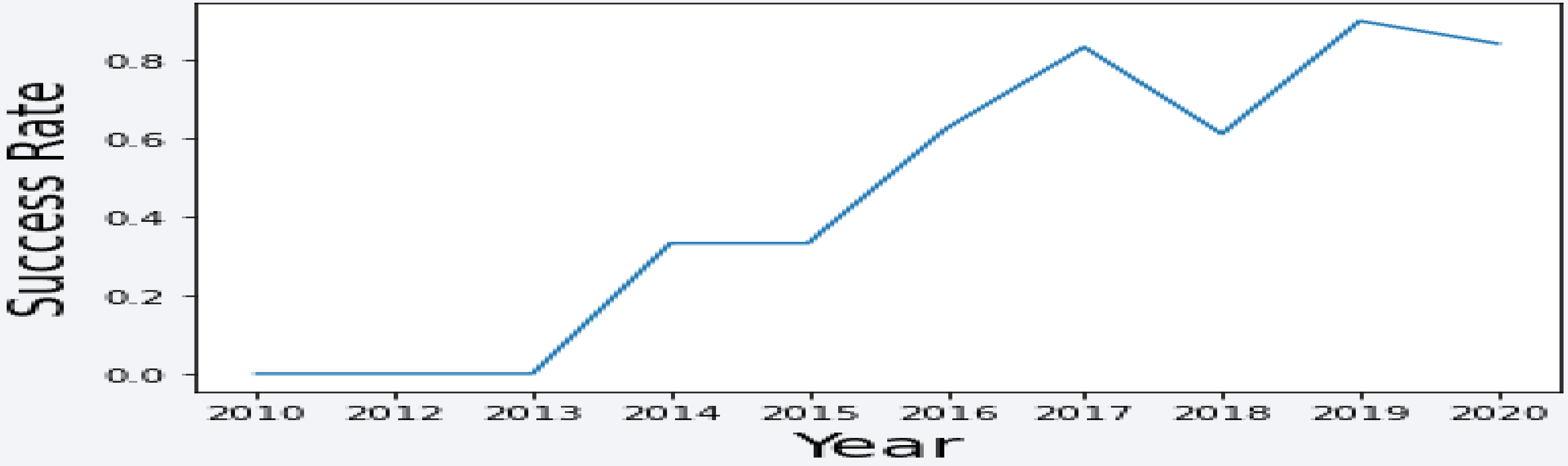
- Success in LEO and PO orbits improves with higher flight numbers (experience).
- GTO and ISS orbits show no clear correlation between flight number and success.
- SO orbit's low success is due to limited data (only 1 flight).

Payload vs. Orbit Type



- Heavy payloads (>10,000 kg) in PO, LEO, and ISS orbits have high success rates.
- GTO orbits show mixed outcomes regardless of payload mass.
- LEO and SO orbits have narrow payload ranges, limiting conclusions.

Launch Success Yearly Trend



- Success rates rose steadily from 2013 to 2017, dipped in 2018, and recovered by 2019.
- 2020 saw a minor decline but maintained high overall success.
- Reflects iterative improvements in SpaceX's landing technology.

All Launch Site Names

- Find the names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

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

average_payload_mass
2534

First Successful Ground Landing Date

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

FIRST_SUCCESS_GP
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

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

Boosters Carried Maximum Payload

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

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

2015 Launch Records

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

Month	Date	Booster_Version	Launch_Site	Landing_Outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

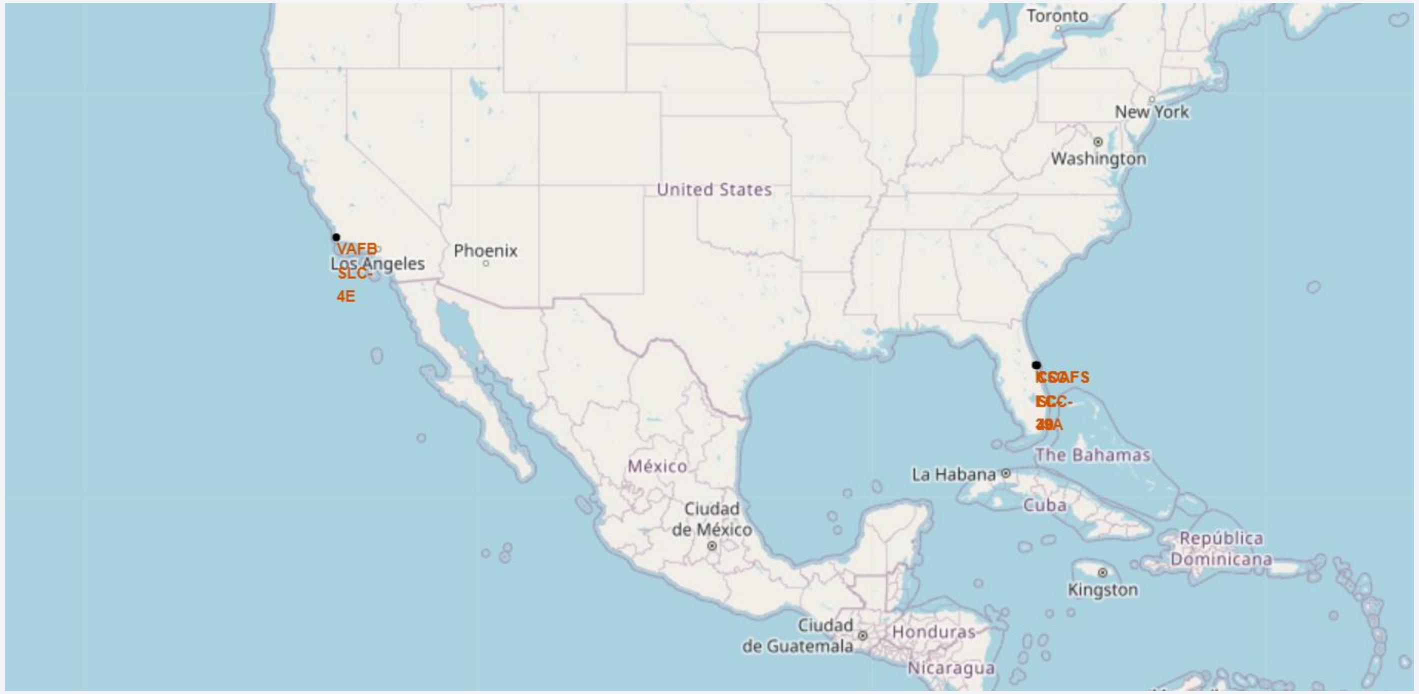
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

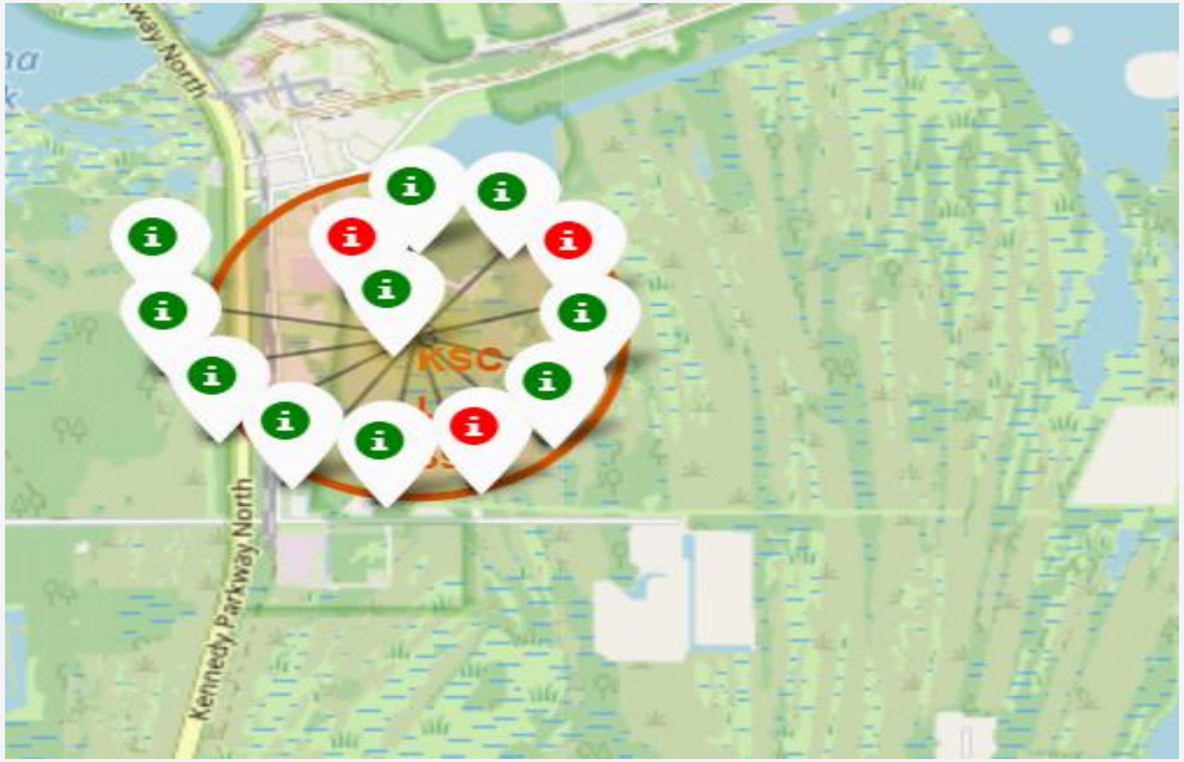
Section 3

Launch Sites Proximities Analysis

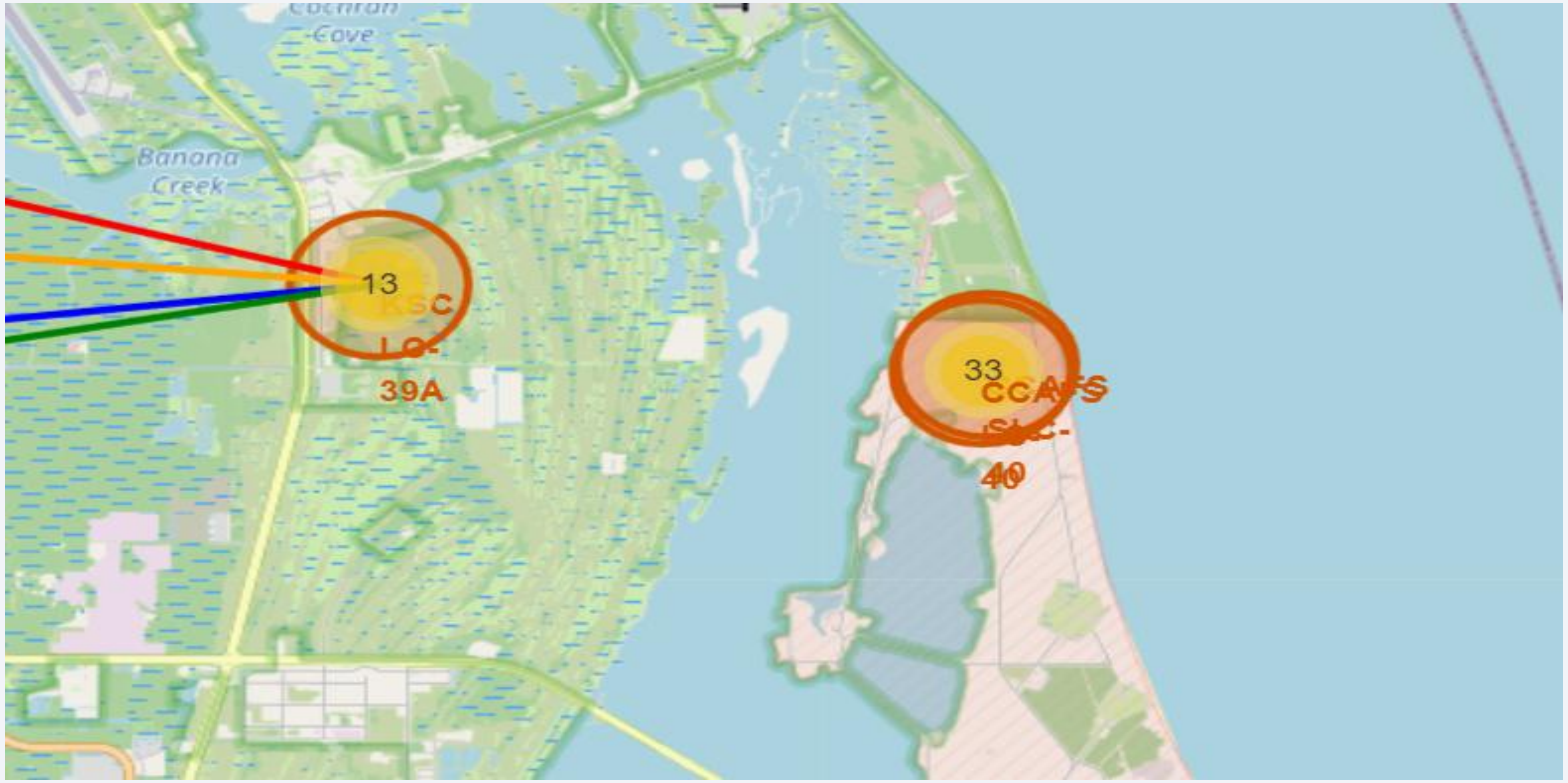
Folium Map Screenshot Launch Sites



Folium Map Cluster Marker



Folium Map Highway Railway



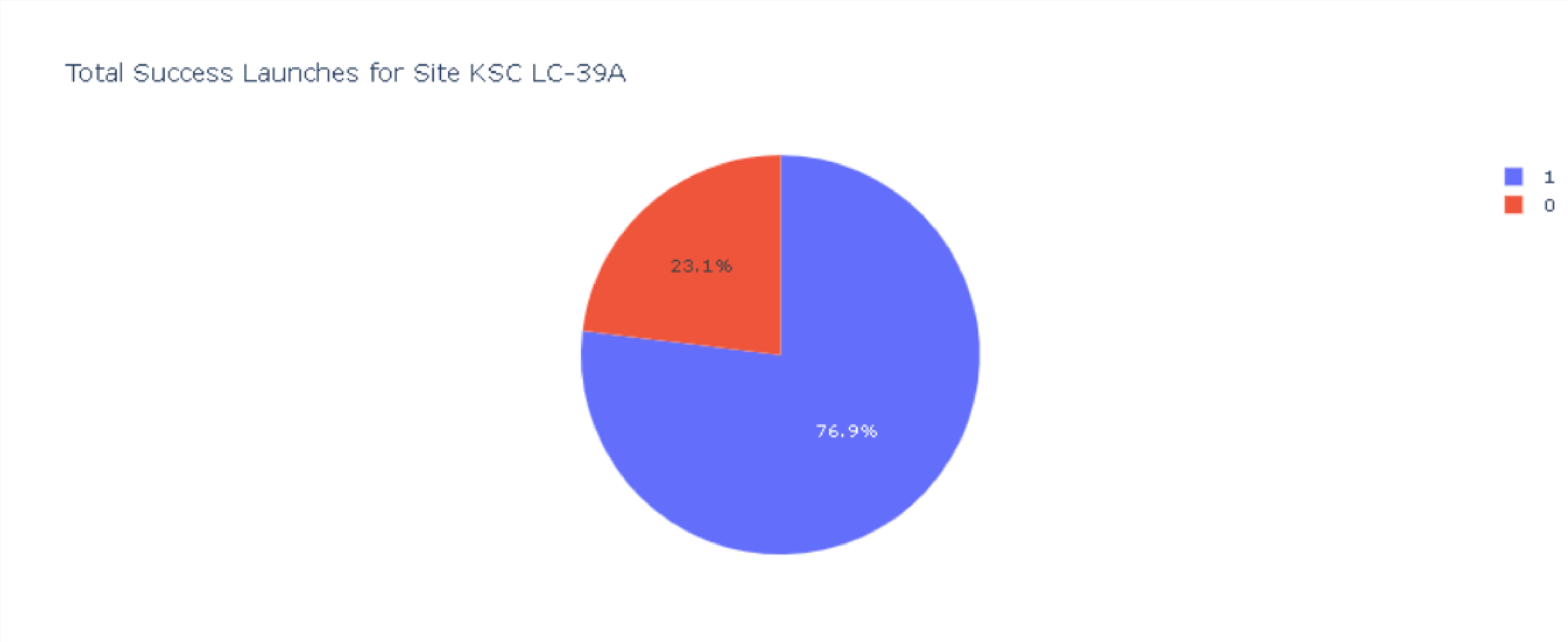
Section 4

Build a Dashboard with Plotly Dash

Dashboard Success Count

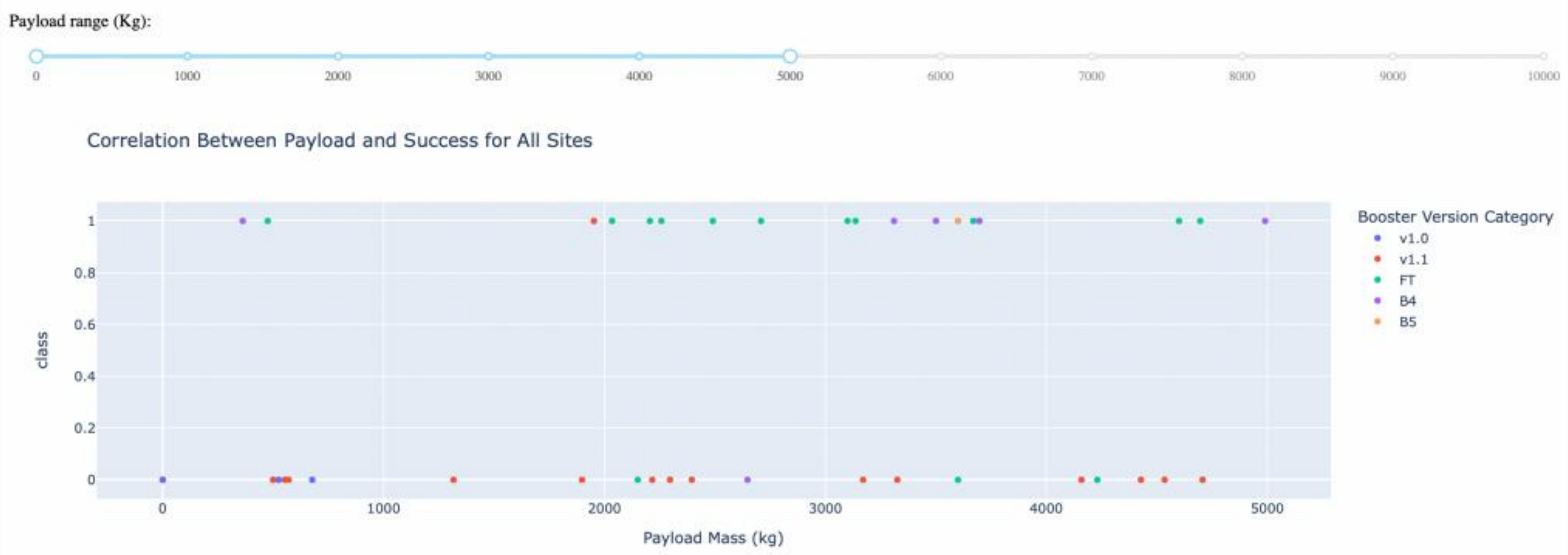
- Explain the important elements and findings on the screenshot

Dashboard Success Ratio



- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>



- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

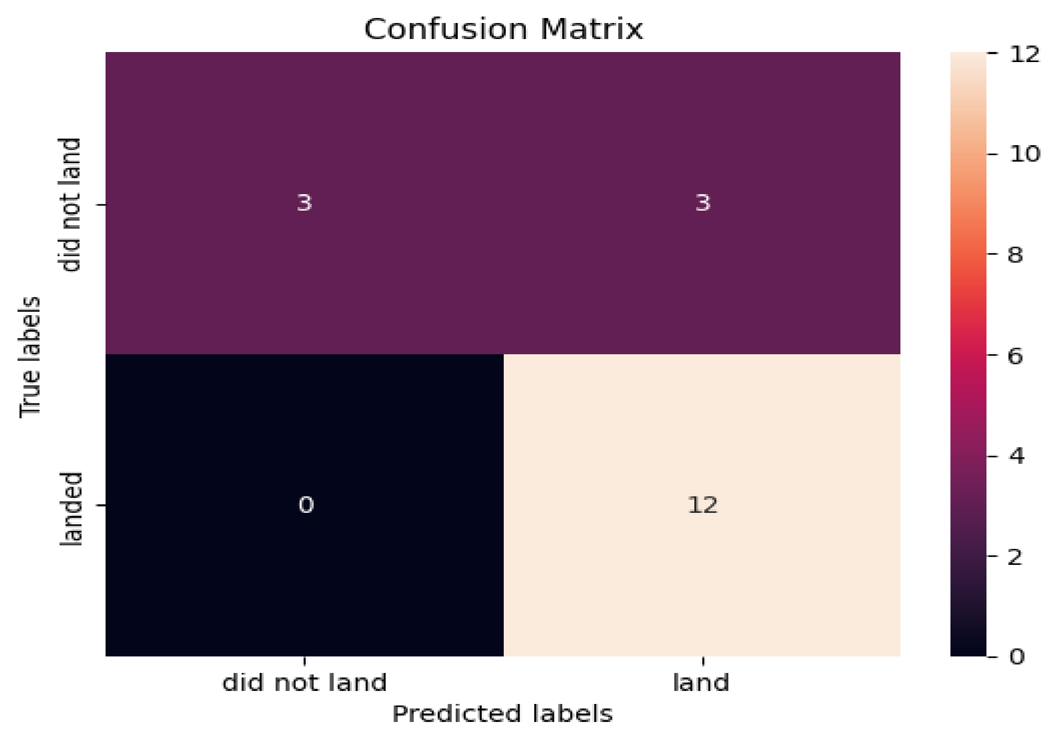
Predictive Analysis (Classification)

Classification Accuracy

- All models: 83.33% testing accuracy. Decision Tree had the highest training accuracy (88.93%)

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

Confusion Matrix



- Decision Tree missed 2 successful landings; others had 100% true positives for successes.

Conclusions

- Predictions are viable (83.33% accuracy), aiding cost estimates.
- Recommendations: Use Logistic Regression/KNN for speed; explore weather/seasonal data next.

Appendix

- https://github.com/DebarjunChakraborty/IBM-Data-Science-Specialization.git

Thank you!