IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Veda Janodia
23/03/2025

Outline

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

Executive Summary

- The project aims to analyze SpaceX launch data to determine factors influencing launch success.
- Data was collected from SpaceX API and web scraping, then processed through data wrangling and visualization.
- Predictive modeling was applied to classify successful vs. failed landings using logistic regression and other machine learning models.
- The project also involved interactive dashboards and geospatial mapping of launch sites.
- Key findings indicate a correlation between orbit type, payload mass, and landing success rates.

Introduction

- **Background**: SpaceX Falcon 9 launches aim for cost-efficient reusability by successfully landing boosters.
- **Problem Statement**: Identify factors influencing booster landing success.
- **Questions to Address**:
  - Which launch sites have the highest success rates?
  - How does payload mass impact landing success?
  - Which booster versions perform best?

Section 1

Methodology

Methodology

Executive Summary

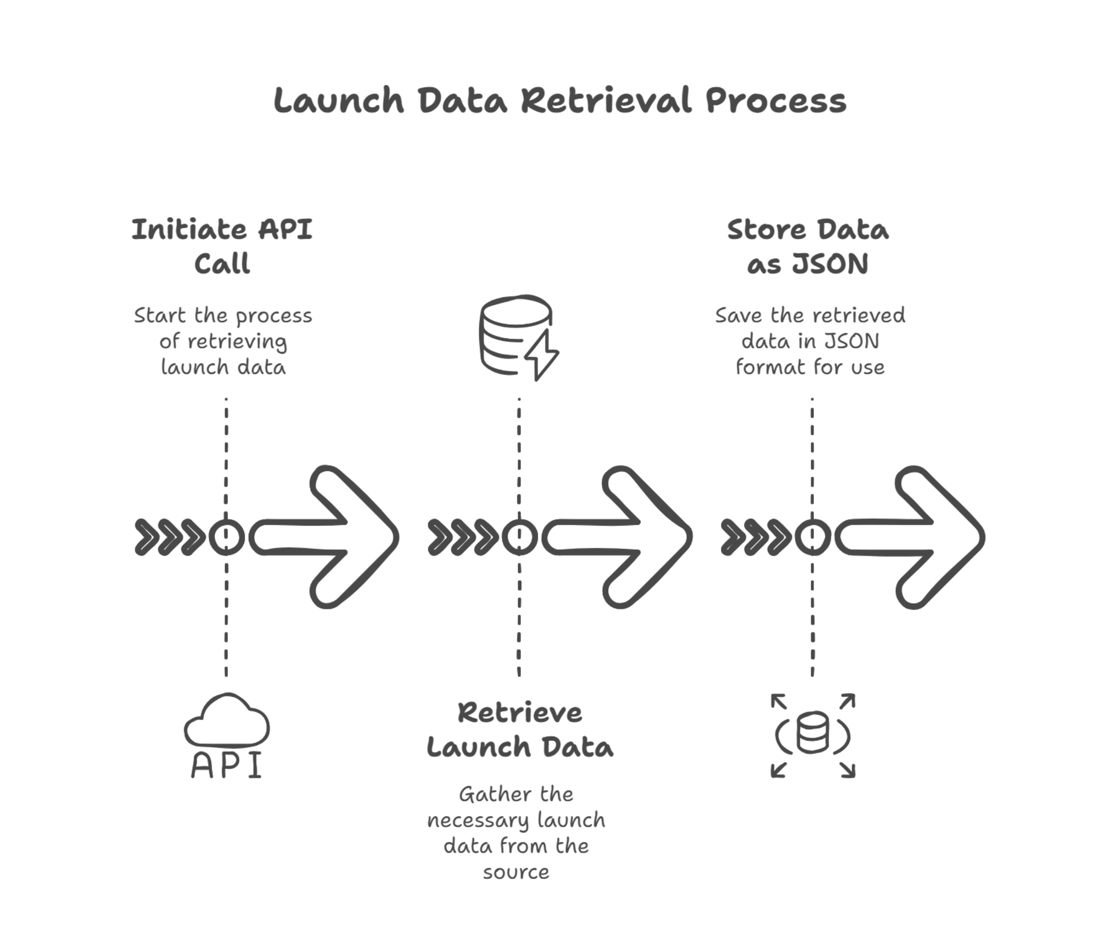
- Data Collection:
    - SpaceX API and Wikipedia scraping.
- Data Wrangling:
    - Cleaned and enriched data, created landing outcome labels.
- EDA:
    - Visualized data using scatterplots, barplots, and SQL queries.
- Interactive Analytics:
    - Built maps with Folium and dashboards with Plotly Dash.
- Predictive Analysis:
    - Tested classification models (Logistic Regression, SVM, Decision Tree, KNN).

Data Collection

- Sources:
  - SpaceX API: https://api.spacexdata.com/v4/rockets/
- Wikipedia:
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Techniques:
  - API requests and web scraping.

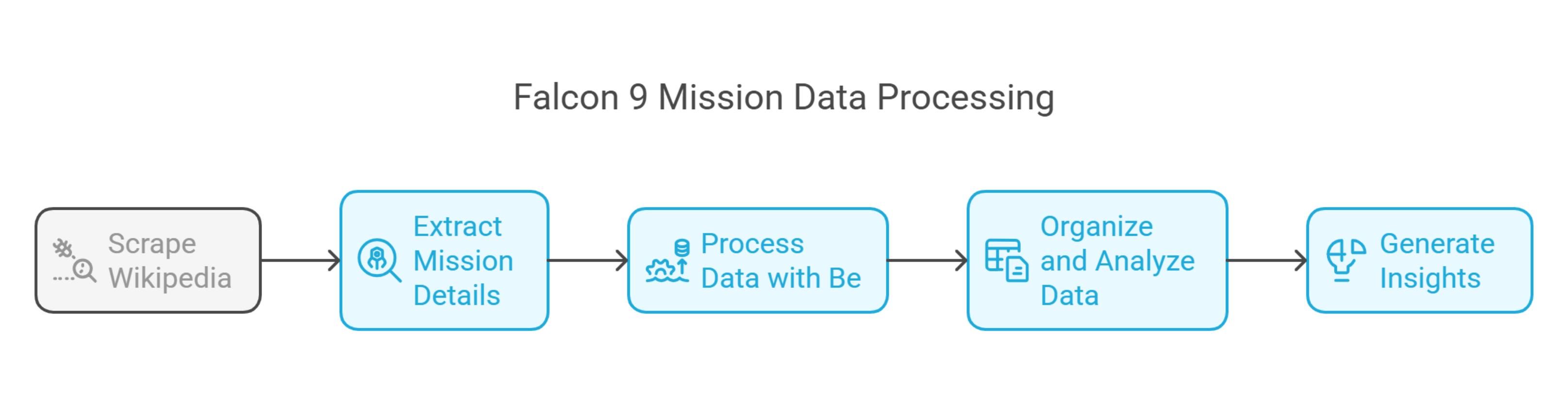
Data Collection – SpaceX API

- REST API calls made to retrieve launch data.
- Data stored as JSON and converted to DataFrame.
- GitHub link to API notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Scraped Wikipedia for Falcon 9 mission details.
- Used BeautifulSoup and Requests.
- GitHub link to web scraping notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Cleaned and processed data using Pandas.
- Handled missing values and inconsistent formats.
- GitHub link to data wrangling notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- **Visualizations**:
  - Success rate by launch site.
  - Impact of payload mass on success.
  - Flight trends over time.

GitHub link to EDA with data visualization notebook:
https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Extracted insights from structured launch data.
- Used CRUD Operation and conditions to extract SpaceX booster versions, classes to show if rockets were a success or failure etc.
- GitHub link to SQL notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- **Markers:**
  - NASA Johnson Space Center
  - SpaceX Launch Sites
- **Circles:**
  - NASA Johnson Space Center
  - SpaceX Launch Sites
- **Lines:**
  - Represent distances to key points of interest, including:
    - Coastline
    - City
    - Railway
    - Highway
    - **Purpose:**
- Identify and highlight important locations.
- Visualize spatial relationships between sites.
- Assist in analyzing distances and geographic context.
- Github notebook with Folium: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Pie Chart:**
- **Displays:** Success rates of SpaceX launches.
- **Interactions:** Dynamically updates based on the selected launch site.
- **Purpose:** Helps visualize the proportion of successful launches, making it easier to assess overall performance.
- **Scatter Plot:**
- **Displays:** Relationship between payload mass and launch outcomes.
- **Interactions:** Updates based on the selected launch site and payload range.
- **Purpose:** Analyzes how payload mass affects launch success and allows comparison between different booster versions.
- Git hub link to plotly dash notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/dash_interactivity.py

Predictive Analysis (Classification)

- **Hyperparameter Tuning:** Used **GridSearchCV** to optimize model parameters.
- **Evaluated Models:** Tested multiple algorithms, including:
    - Logistic Regression
    - Support Vector Machine (SVM)
    - Decision Tree
    - K-Nearest Neighbors (KNN)
    - **Model Evaluation:**
- **Performance Comparison:** Assessed models based on accuracy scores and confusion matrices.
- **Best Model Selection:** Chose the model with the highest test accuracy for final deployment.
- Git hub link to Machine learning prediction notebook: https://github.com/Vjanodia/Coursera-IBM_DS_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- **Optimal Launch Site:**
  - KSC LC-39A has the highest success rate (76.9%).
- **Payload Insights:**
  - Payloads under 6,000 kg with FT boosters are most successful.
  - Limited data for payloads over 7,000 kg.
- **Orbit Success Rates:**
  - Highest success rates for ES-L1, GEO, HEO, and SSO orbits.

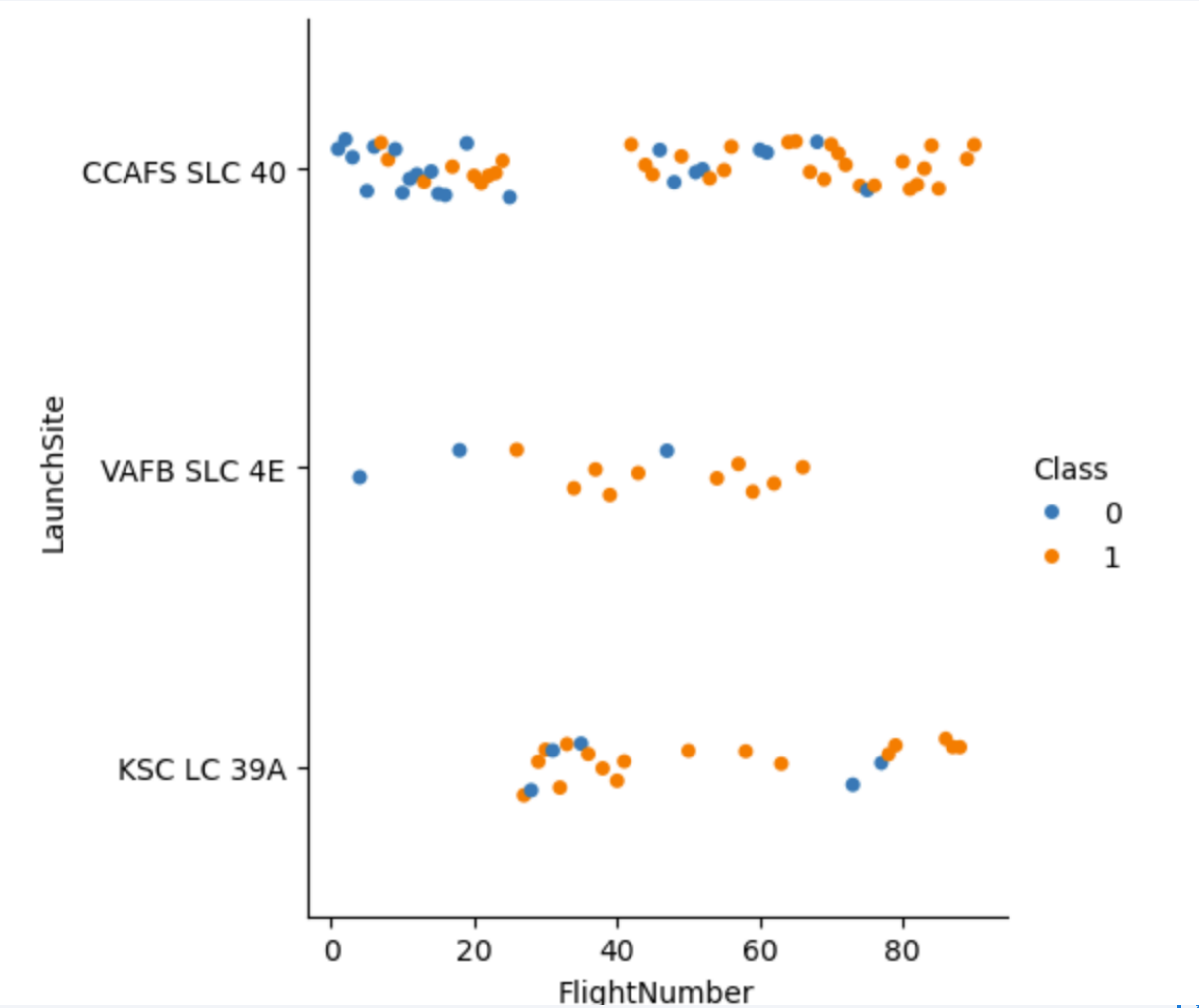
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

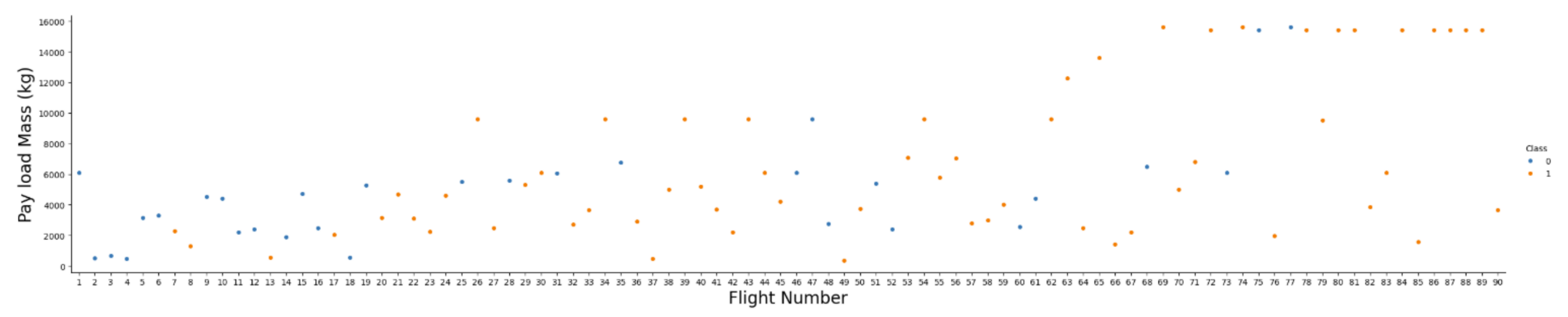
- Scatter plot of Flight Number vs. Launch Site

- Each type of launch site shows a mix of class 1 and 0 for different flight numbers



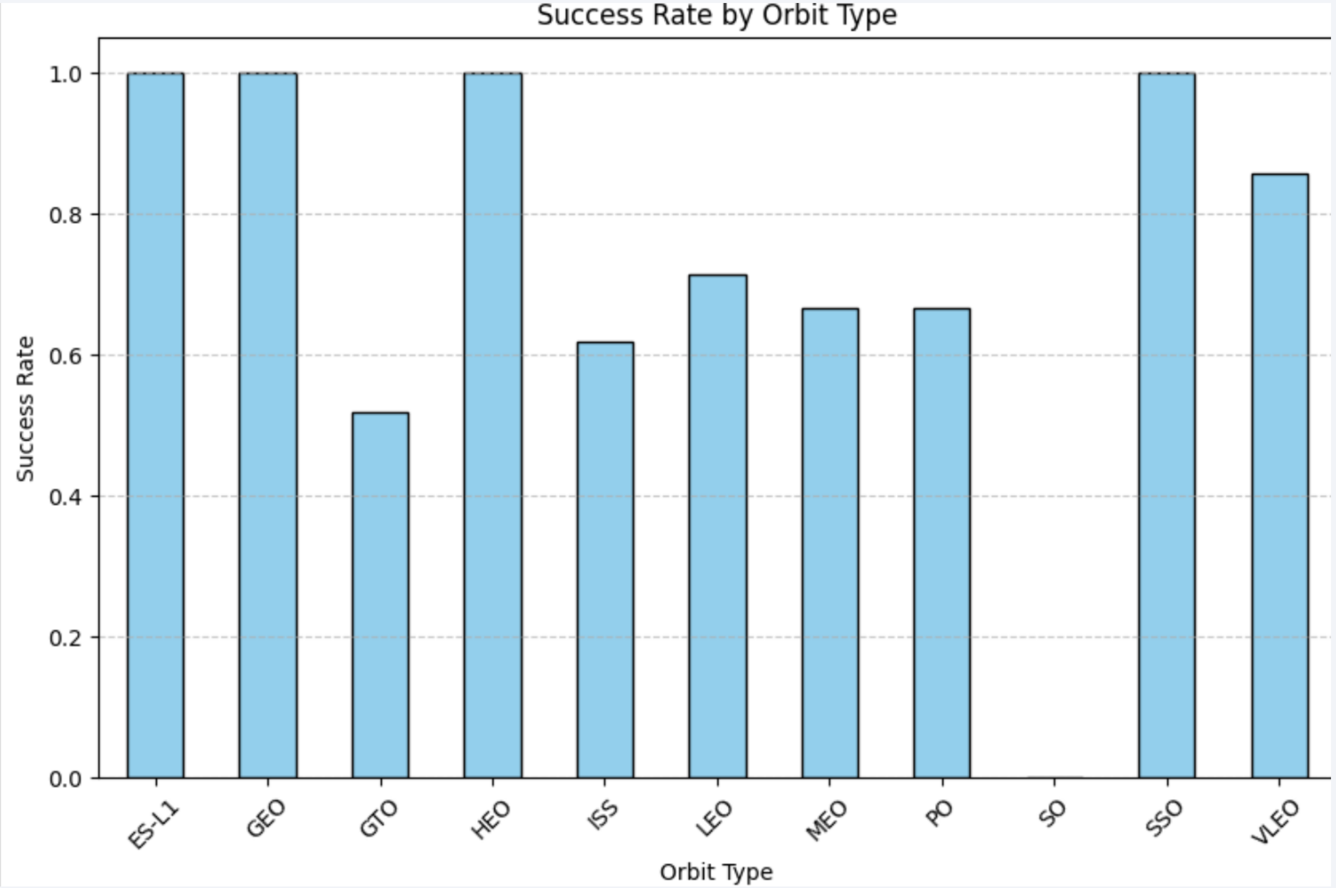
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site\
- Figure shows more of class 1 having a higher pay load mass



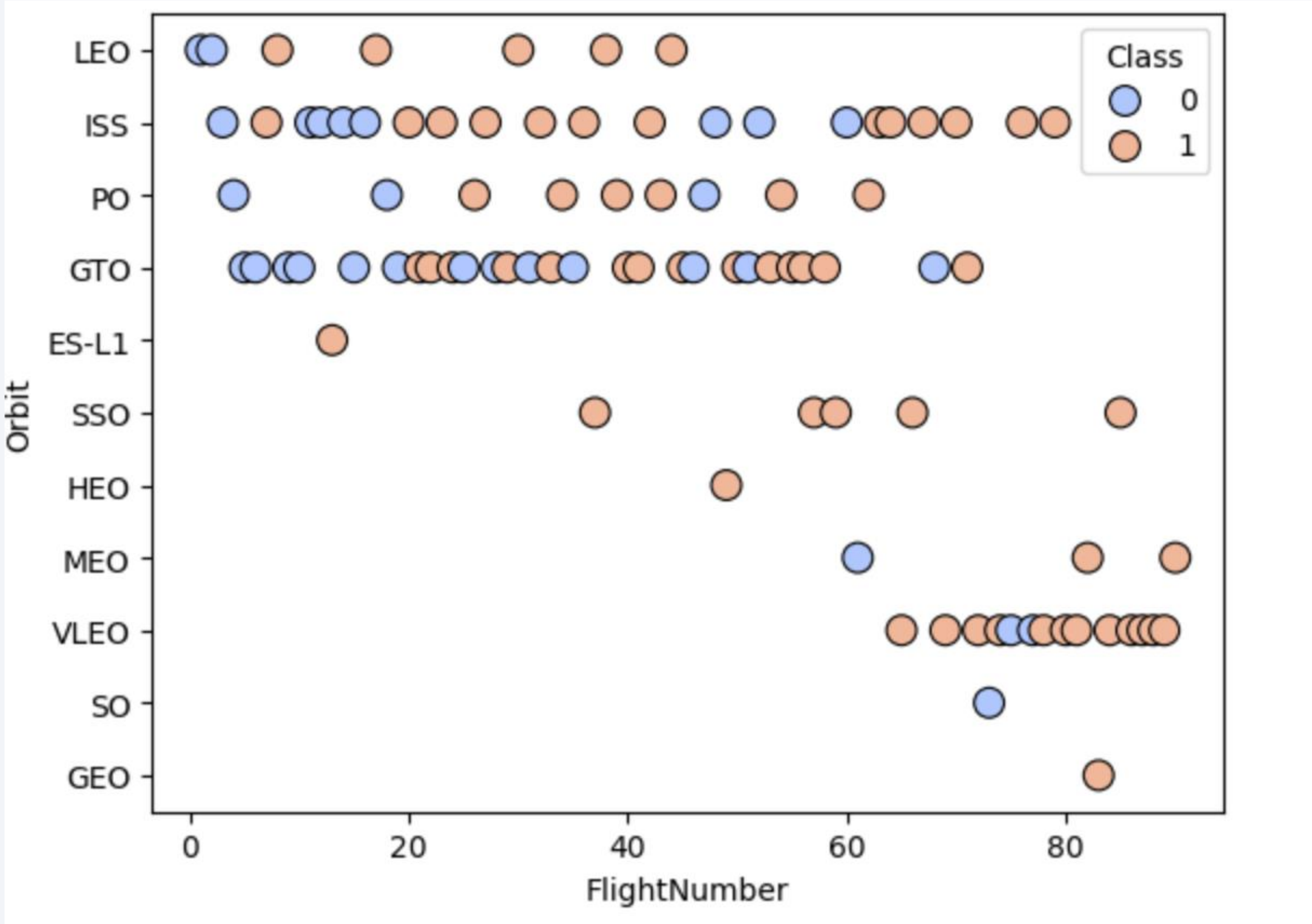
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type



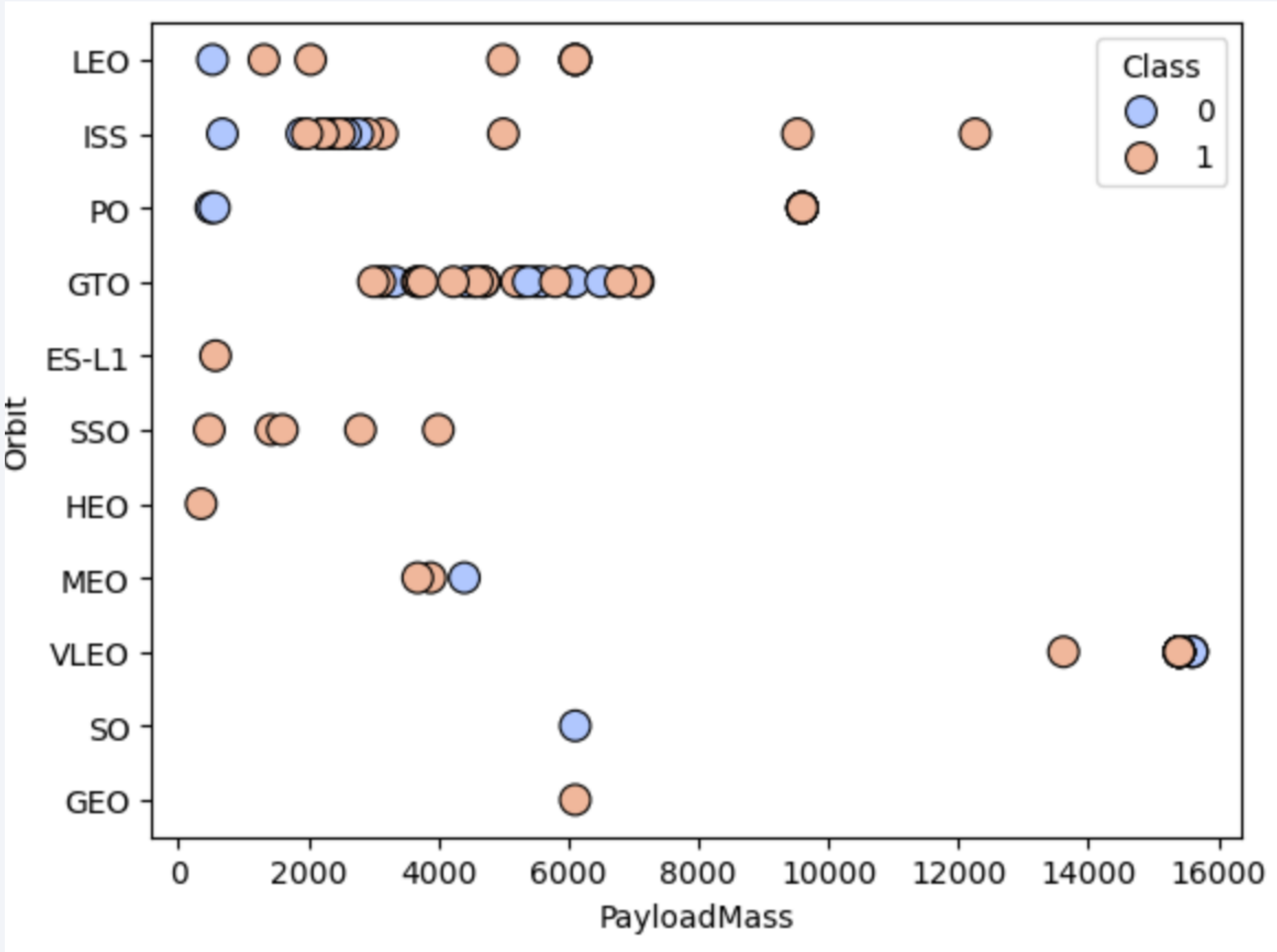
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type



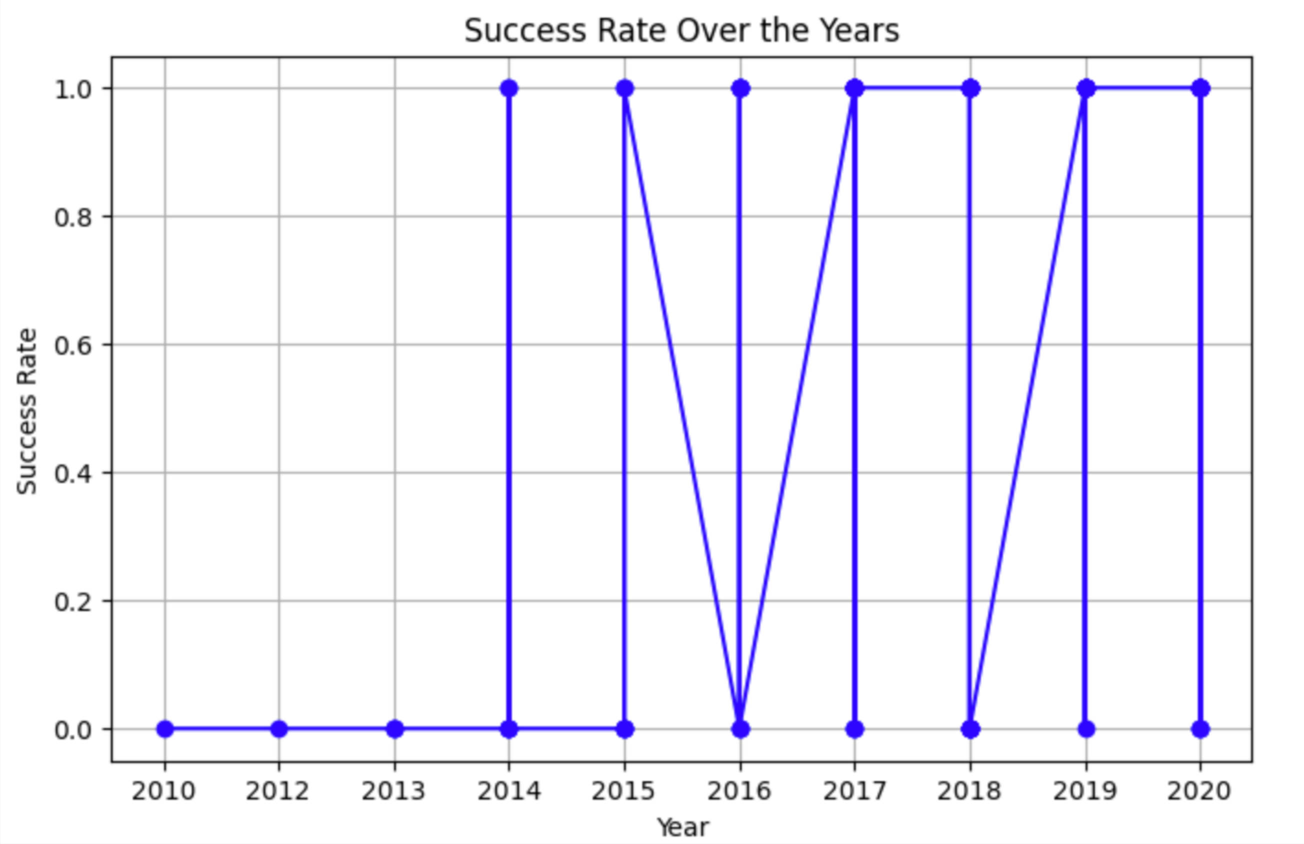
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Line chart of yearly average success rate



All Launch Site Names

- We found 4 launch sites

CCAFS LC-40

VAFB SLC-4

EKSC LC-39

ACCAFS SLC-40

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

Total Payload Mass

- Total Payload Mass: Sum: 619,967 kg
- This query calculates the total payload mass for all SpaceX launches in the table.

Average Payload Mass by F9 v1.1

- Average Payload Mass for Booster Version 'F9 v1.0’: 340.4 kg

First Successful Ground Landing Date

- Earliest Launch Record:
- Date: 2010-06-04
- Time (UTC): 18:45:00
- Booster Version: F9 v1.0 B0003
- Launch Site: CCAFS LC-40
- Payload: Dragon Spacecraft Qualification Unit
- Payload Mass: 0 kg
- Orbit: LEO
- Mission Outcome: Success
- Landing Outcome: Failure (parachute)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Payloads with Successful Drone Ship Landings (4000 kg < Mass < 6000 kg):
- JCSAT-14
- JCSAT-16
- SES-10
- SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- Mission Outcomes:
  - Success: 98 occurrences
  - Failure (in flight): 1 occurrence

Boosters Carried Maximum Payload

- Booster Versions with Maximum Payload Mass:
- 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

- January:
  - Booster Version: F9 v1.1 B1012
  - Launch Site: CCAFS LC-40

April:

Booster Version: F9 v1.1 B1015

Launch Site: CCAFS LC-4

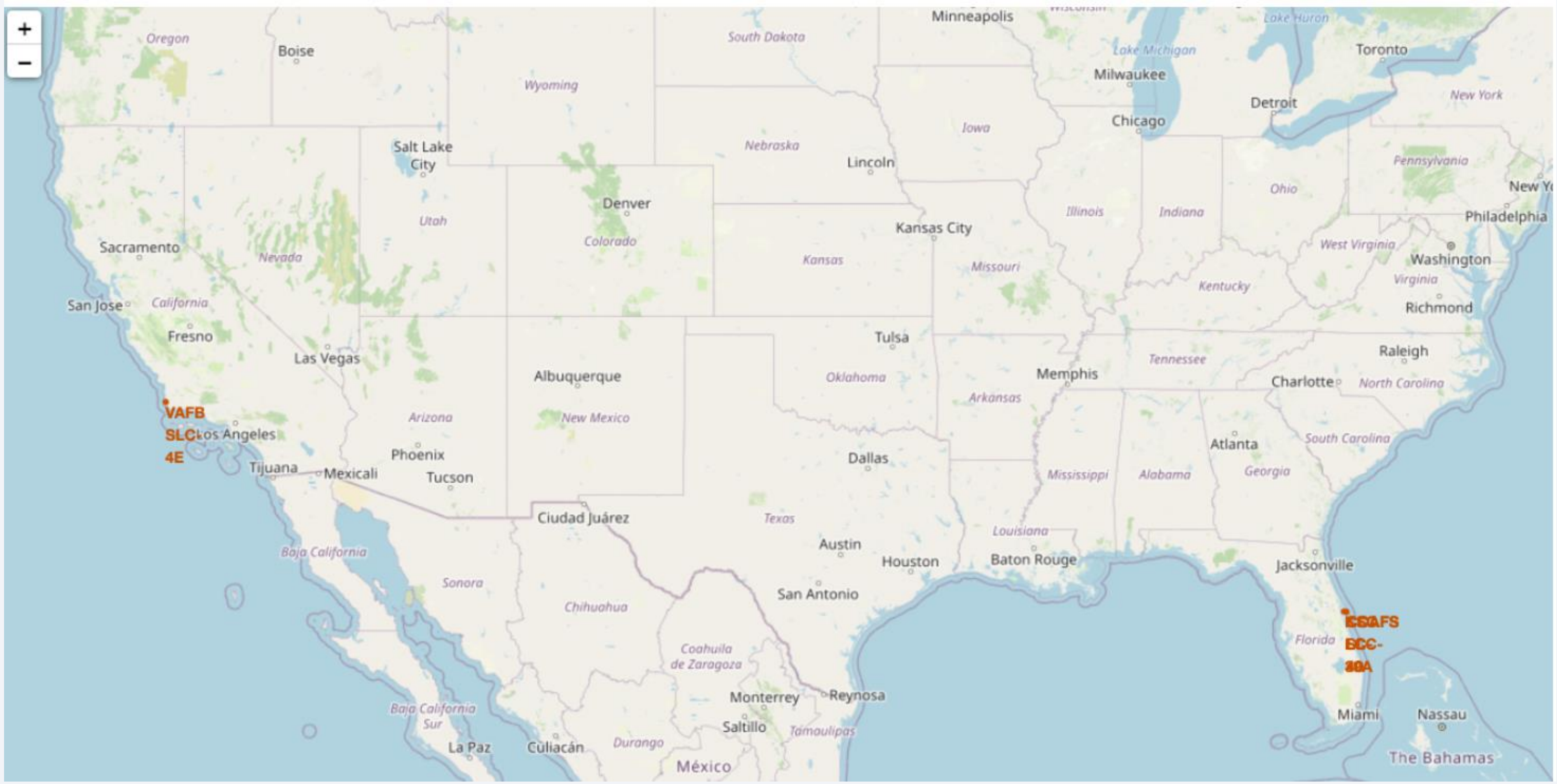
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

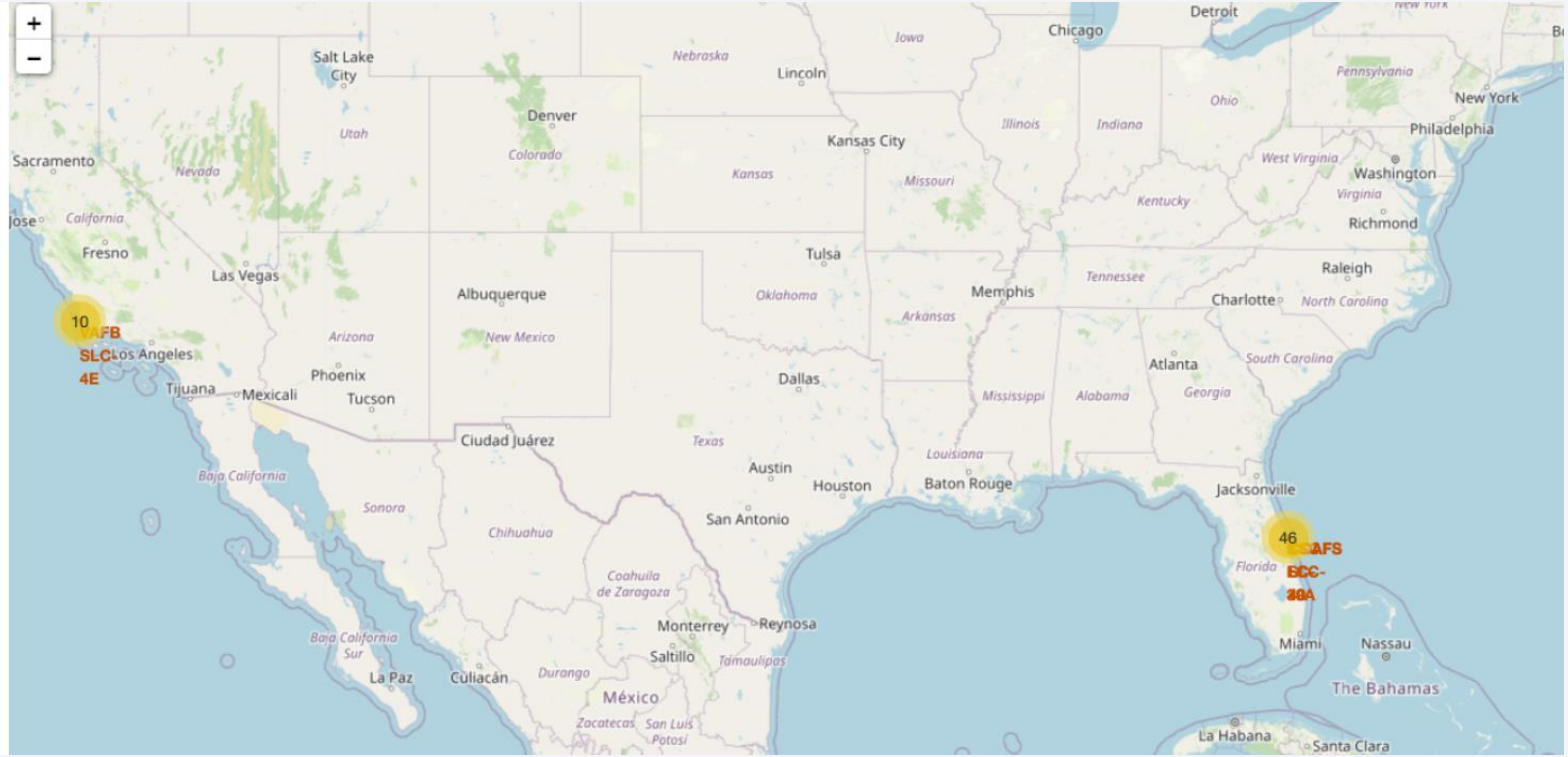
Section 3

Launch Sites Proximities Analysis

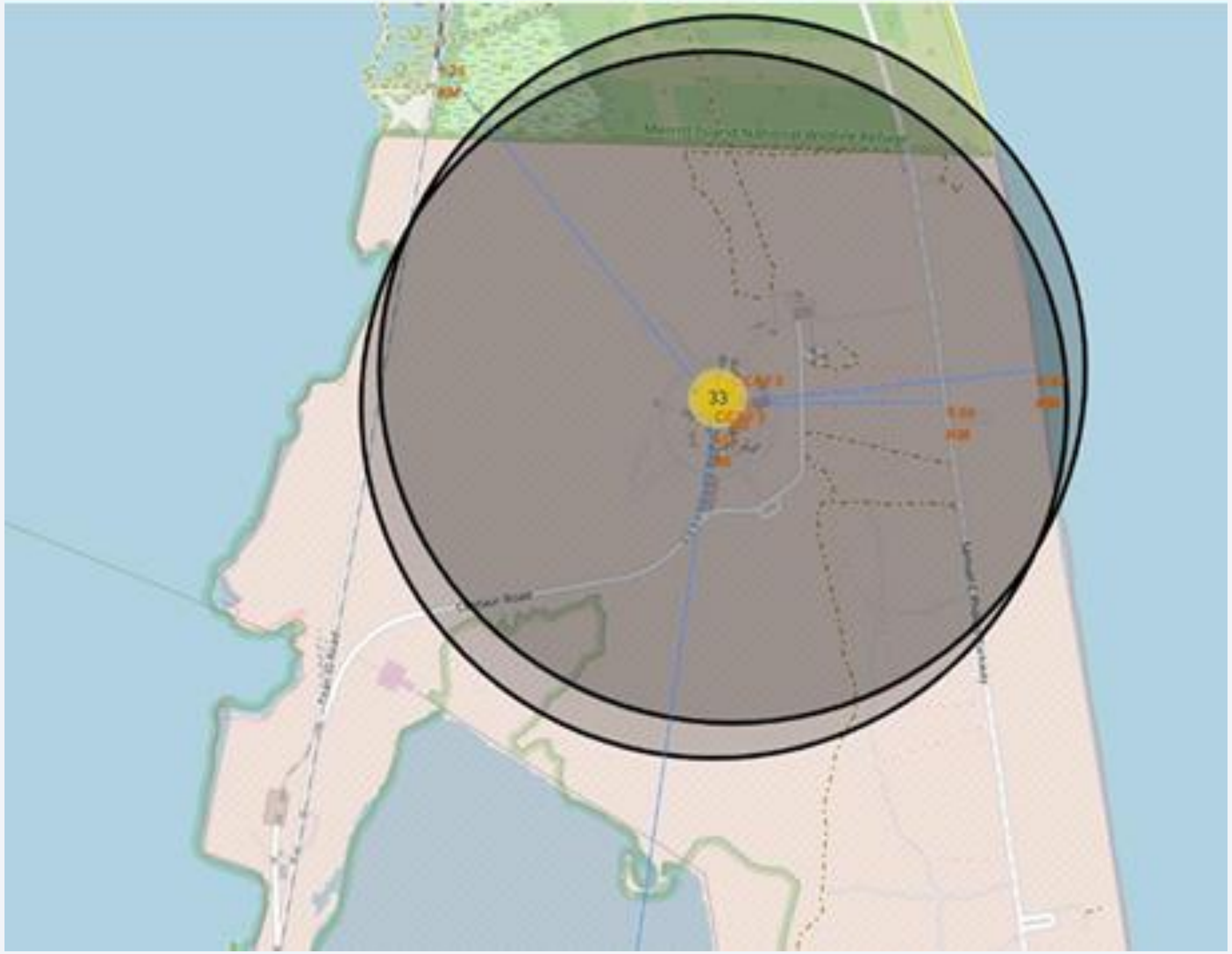
Folium Map Showing all launch sites marked



Folium Map highlighting the number and success rate of launches at each location



Folium Map showing proximity of launch site with other areas

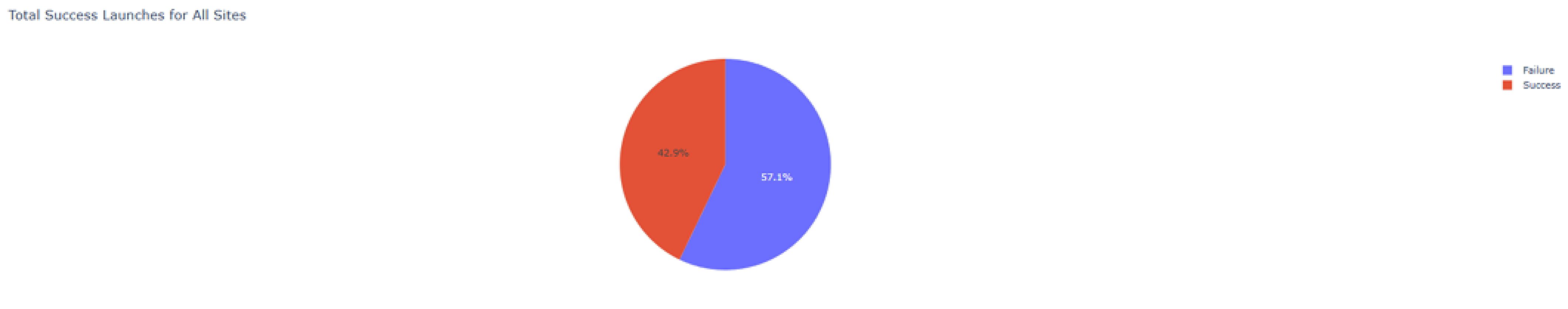


Section 4

Build a Dashboard with Plotly Dash

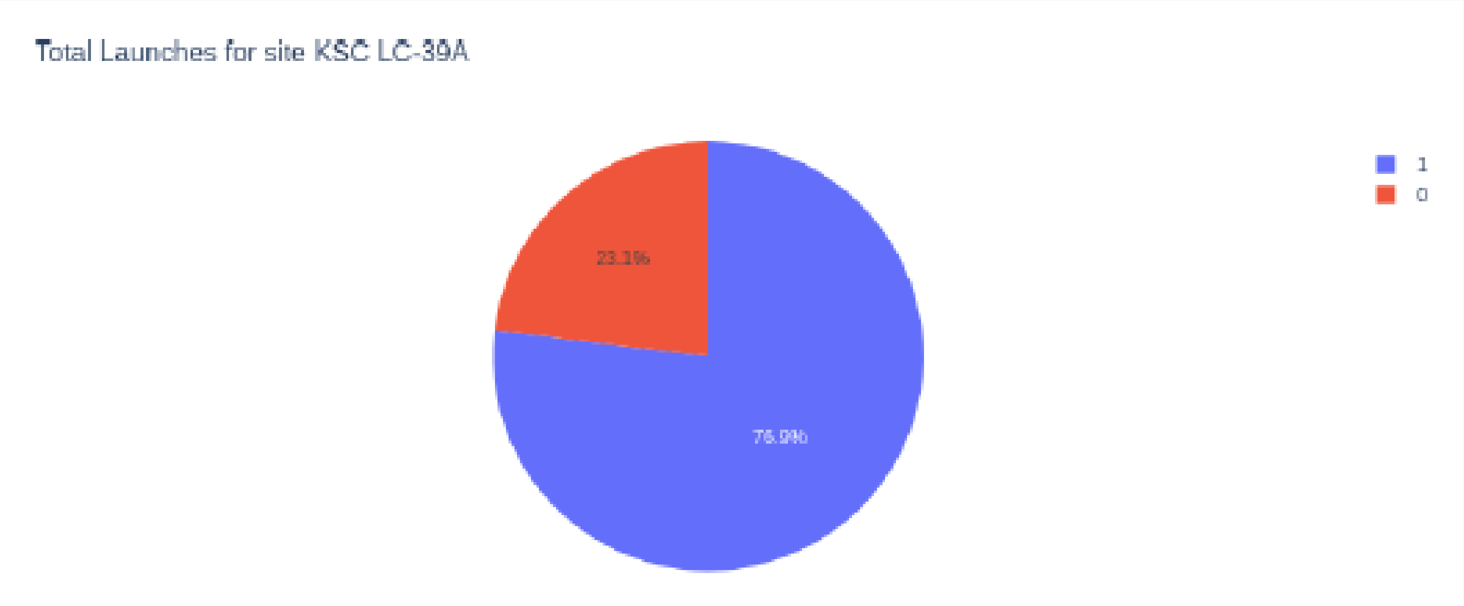
Total Success Launches for all sites

- Pie chart shows total success rate for all sites with 57.1% as success and rest failures.

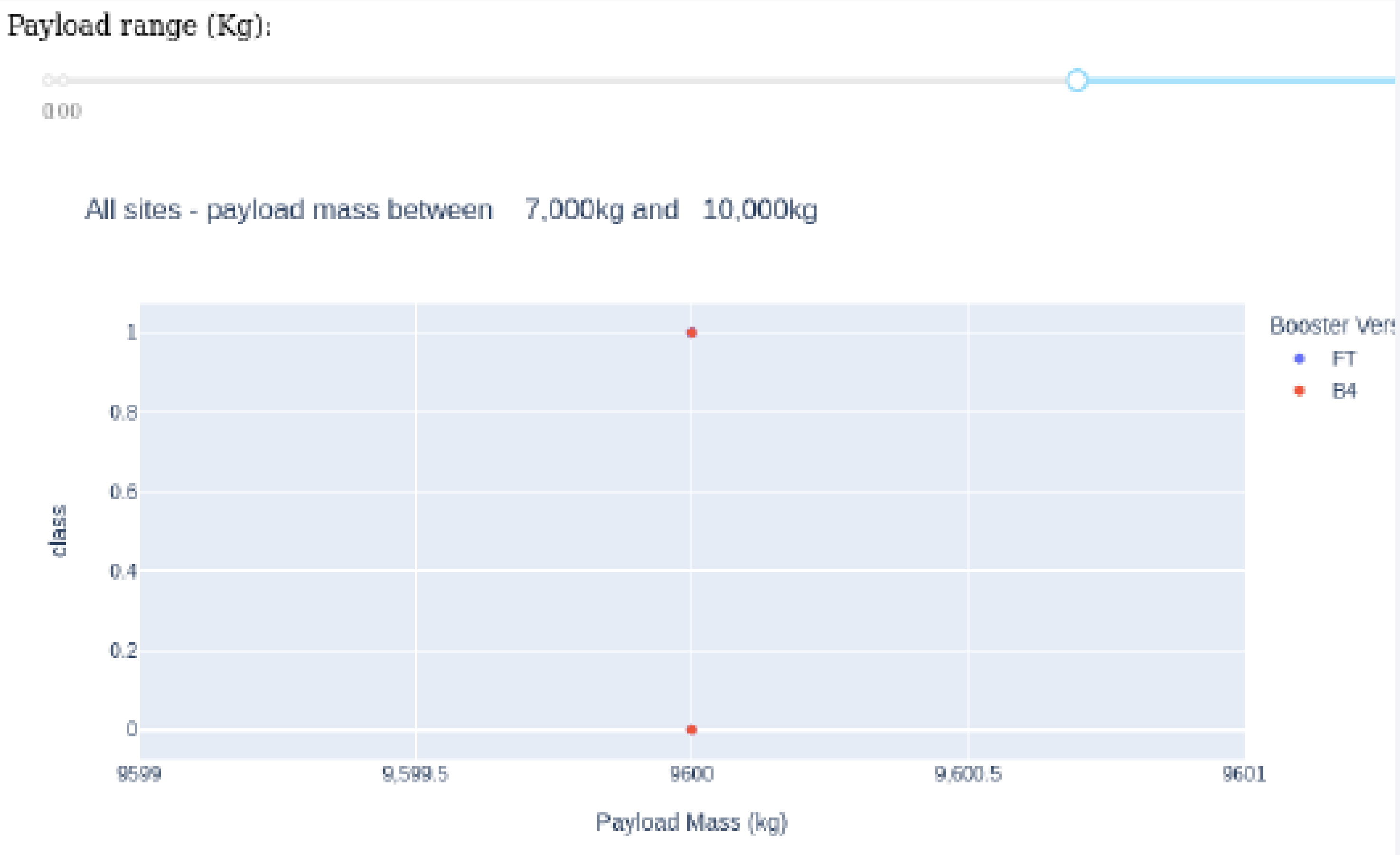


Launch Success Percentage for KSC LC-39A

- Pie shows total success launches for Site KSC LC-39A



Payload vs. Outcome for All Sites

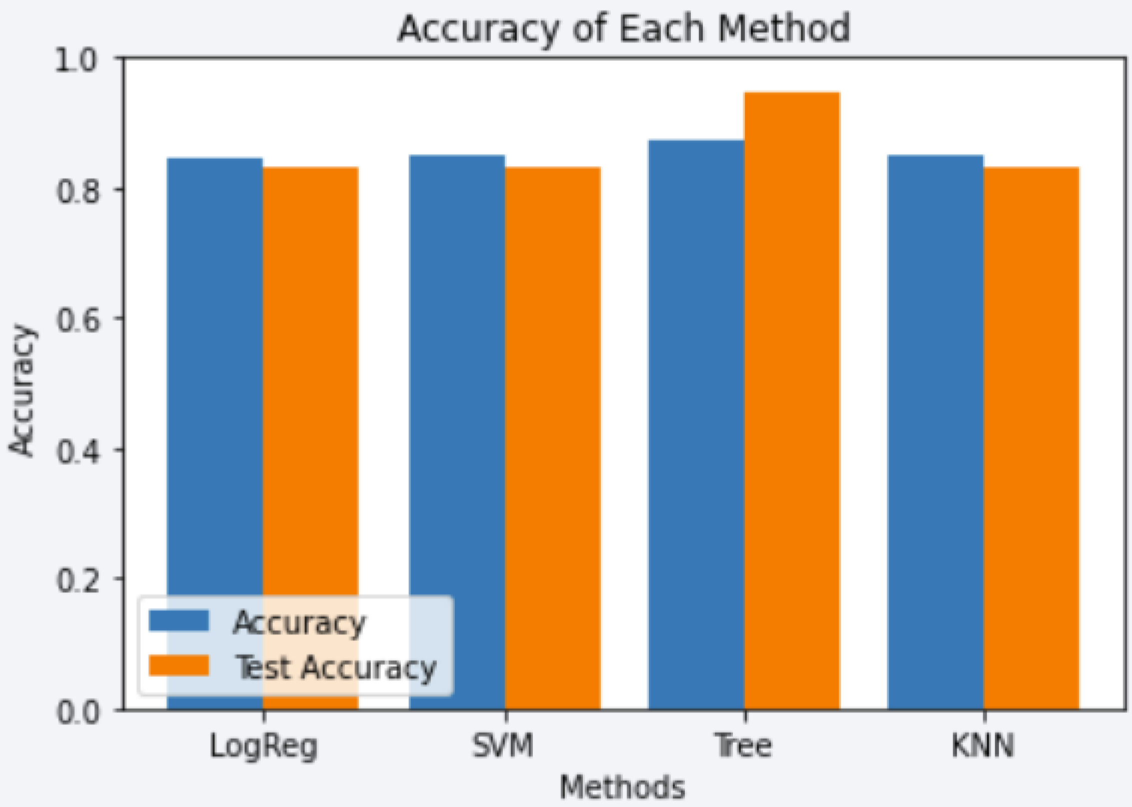


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest classification accuracy was Decision tree classifier
- Accuracy: 87%



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- Through extensive analysis of multiple data sources, we refined our conclusions at each step, uncovering valuable insights into SpaceX launches.
- **Best Launch Site:** KSC LC-39A emerged as the most reliable launch site, with the highest success rate (76.9%).
- **Payload & Risk:** Launches above **7,000kg** tend to be less risky, with a higher likelihood of mission success.
- **Trends in Mission Success:** While most missions have been successful, **landing outcomes have improved over time**, reflecting advancements in rocket technology and operational processes.
- **Predictive Modeling:** Among the evaluated models, **Logistic Regression, SVM, and KNN** achieved the highest accuracy (**83%**), with Decision Trees slightly lower at **78%**.
- **SVM Performance:** The **SVM model** demonstrated strong predictive capabilities with **12 true positives, 3 false positives, and no false negatives**.
- **Decision Tree for Landings:** Despite slightly lower accuracy, **Decision Tree Classifier** can effectively predict successful landings, offering insights that could help improve profitability.
- **Data Visualization & Interactive Dashboards:** Created **visualizations** to explore launch success rates, payload influences, and trends across different sites. **Interactive dashboards** enabled deeper analysis, allowing dynamic exploration of mission outcomes and payload factors

Appendix

- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN SkillsNetwork/datasets/API_call_spacex_api.json
- https://api.spacexdata.com/v4/rockets/

Thank you!