IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eduardo Dietrich Zimmermann
21/09/2025

Outline

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

Executive Summary

•**Objective:** Analyze SpaceX launch data to identify factors influencing mission success and build predictive models.
•**Approach:**
•Data collected via SpaceX REST API and web scraping.
•Data wrangling for consistency and missing value handling.
•Exploratory Data Analysis (EDA) using visualizations and SQL.
•Interactive analytics with Folium maps and Plotly Dash dashboard.
•Predictive analysis with classification models.
•**Key Findings:**
•Success rate of launches improved significantly after 2015 (>80% success in recent years).
•Payload mass and launch site strongly correlated with mission outcomes.
•The *Logistic Regression* achieved the best accuracy (**83,33%**).
•**Impact:** Provides insights for forecasting launch outcomes and guiding decision-making for future missions.

Introduction

•**Background:** SpaceX is a leader in commercial spaceflight, aiming to reduce costs and enable Mars colonization.
•**Problem Statement:** Which factors influence the success of Falcon 9 launches?
•**Guiding Questions:**
•Does payload mass affect mission success?
•Do different launch sites have higher/lower success rates?
•How have success rates evolved over time?
•Can we predict future launch success using historical data?

Section 1

Methodology

Methodology

Executive Summary

- **Data Collection:**
- Retrieved launch records from SpaceX REST API.
- Performed web scraping for supplementary mission details.
- GitHub Repo: https://github.com/eduardozimmermann1/PythonCapstone/tree/main
- **Data Wrangling:**
- Standardized column names, handled missing values, normalized data types.
- Created derived variables (e.g., success indicator).
- **EDA:**
- Visual exploration (scatter plots, bar charts, line charts).
- SQL queries for structured insights.
- **Interactive Analytics:**
- Folium maps to visualize launch sites and outcomes.
- Plotly Dash for interactive filtering and payload analysis.
- **Predictive Analysis:**
- Built classification models: Logistic Regression, Decision Tree, Random Forest, KNN, SVM.
- Compared performance to select the best predictor.

Data Collection

- Queried the SpaceX public REST API (/v4/launches/past)
- Retrieved structured launch data in JSON format (via API response)
- Converted JSON response directly into Pandas DataFrame
- Extracted relevant fields: flight number, date, payload mass, customer, booster version, outcome
- Normalized nested fields for consistency
- Ensured reproducibility by storing processed dataset in the project repository

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- GitHub URL of the completed SpaceX API calls notebook, as an external reference and peer-review purpose: <https://github.com/eduardozimmermann1/PythonCapstone/blob/main/jupyter-labs-spacex-data-collection-api_modified%20(1).ipynb>

we can apply the rest of the functions here:

In [24]:
```
# Call getLaunchSite
getLaunchSite(data)
```

In [25]:
```
# Call getPayloadData
getPayloadData(data)
```

In [26]:
```
# Call getCoreData
getCoreData(data)
```

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/Webscraping>

```
Flight No.                      object
Date and time (UTC)     datetime64[ns]
Version Booster                 object
Launch site                     object
Payload                         object
Payload mass                   float64
Orbit                           object
Customer                        object
Launch outcome                  object
Booster landing                 object
Date                            object
Time                            object
dtype: object
Flight No.             0
Date and time (UTC)    0
Version Booster        0
Launch site            0
Payload                0
Payload mass           0
Orbit                  0
Customer               0
Launch outcome         0
Booster landing        0
Date                   0
Time                   0
dtype: int64
Booster landing com \n: 0
```

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- GitHub URL of completed data wrangling related notebooks, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/Wranglig>

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- GitHub URL of completed EDA with data visualization notebook, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/EDAVisualization>

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- GitHub URL of completed EDA with SQL lab, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/EDAwithSQLPeerGraded>

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- GitHub URL of completed Interactive Map with Folium, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/VisualAnalyticswithFoliumLab>

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- GitHub URL of completed Dashboard with Plotly Dash, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/DashboardwithPlotly>

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- GitHub URL of completed predictive analysis lab, as an external reference and peer-review purpose: < https://github.com/eduardozimmermann1/PythonCapstone/tree/MachineLearningPredictionLab >

Results

Exploratory data analysis results:
•**All Launch Sites:** Extracted unique site names.
•**CCA Sites:** Filtered records beginning with CCA
•**Total Payload by NASA Missions:** Quantified NASA contributions.
•**Average Payload for F9 v1.1:** Derived mean payload for specific booster.
•**First Successful Ground Landing:** Identified milestone event date.
•**Drone Ship Landings with Payload 4000–6000 kg:** Extracted qualifying missions.
•**Total Successful vs. Failed Outcomes:** Showed improvement over time.
•**Ranking Landing Outcomes (2010–2017):** Revealed dominant patterns.
•**Flight Number vs. Launch Site:** Success rate increases with higher flight numbers (experience effect).
•**Payload vs. Launch Site:** Larger payloads associated with variable success rates across sites.
•**Success Rate vs. Orbit Type:** Certain orbits (e.g., GTO, LEO) show distinct success ratios.
•**Yearly Trend:** Success rate improves steadily, especially after 2015.

Interactive analytics results:
•**Global Map Markers:** All launch sites geolocated.
•**Color-Labeled Outcomes:** Successes in green, failures in red.
•**Distance to Coastline Example:** Measured proximity of launch pads to coastlines, highways, railways.
•**Launch Success by Site (Pie Chart):** KSC LC-39A achieved ~77% success.
•**Payload vs. Outcome Scatter (Interactive Slider):** Certain payload ranges show higher probability of success.
•**Filters Enabled:** Interactive selection by site, payload range, and booster version.

Predictive analysis results:
•**Classification Accuracy (Bar Chart):** Logistic Regression outperformed others with highest accuracy (0,8333**%**).
•**Confusion Matrix of Best Model:** Demonstrates strong precision and recall on positive (success) outcomes.
•**Insight:** Model can reasonably predict whether a future mission will succeed based on historical features..

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

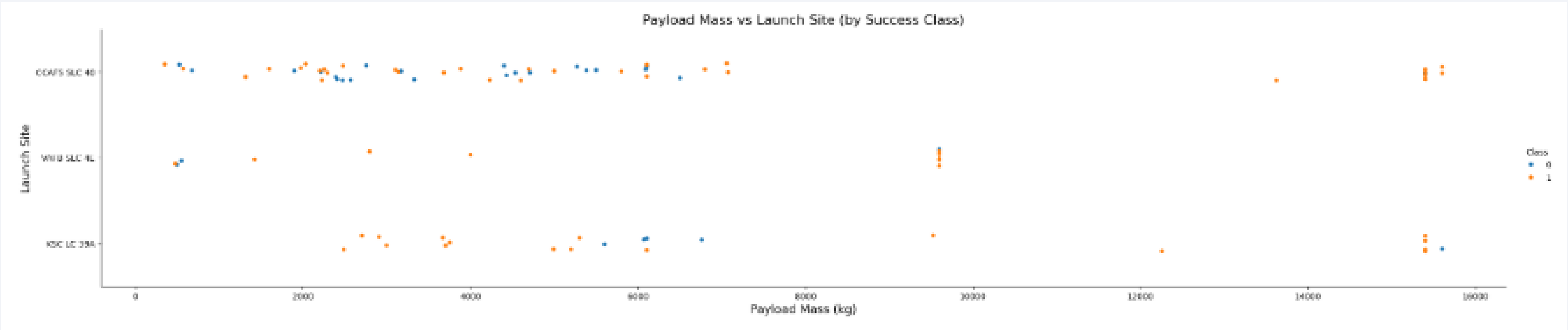
- Show a scatter plot of Flight Number vs. Launch Site

- CCAFS SLC40 had more flights than others.

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

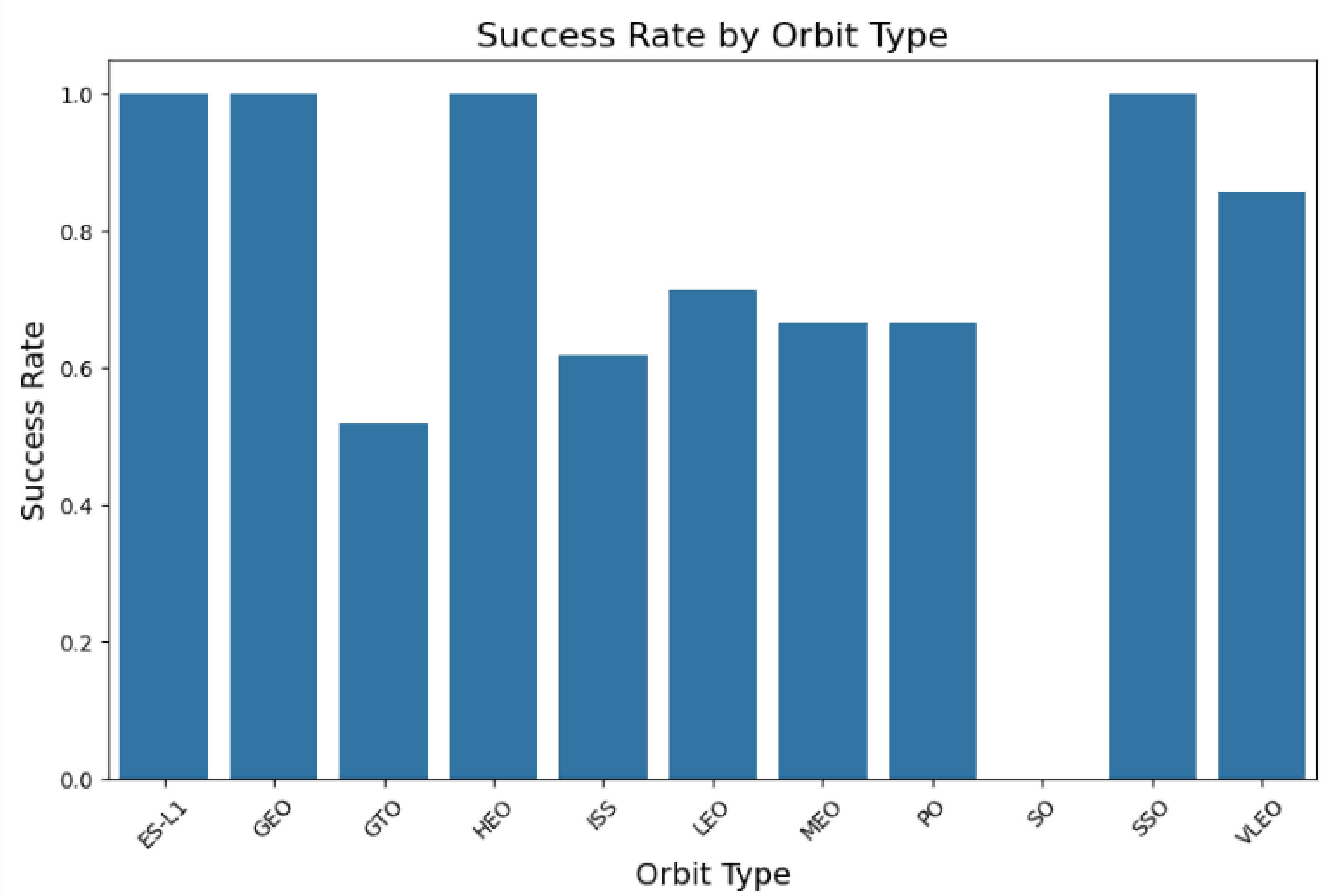
- Payload Mass less than 7500 Kg concentrates most launch sites.



Success Rate vs. Orbit Type

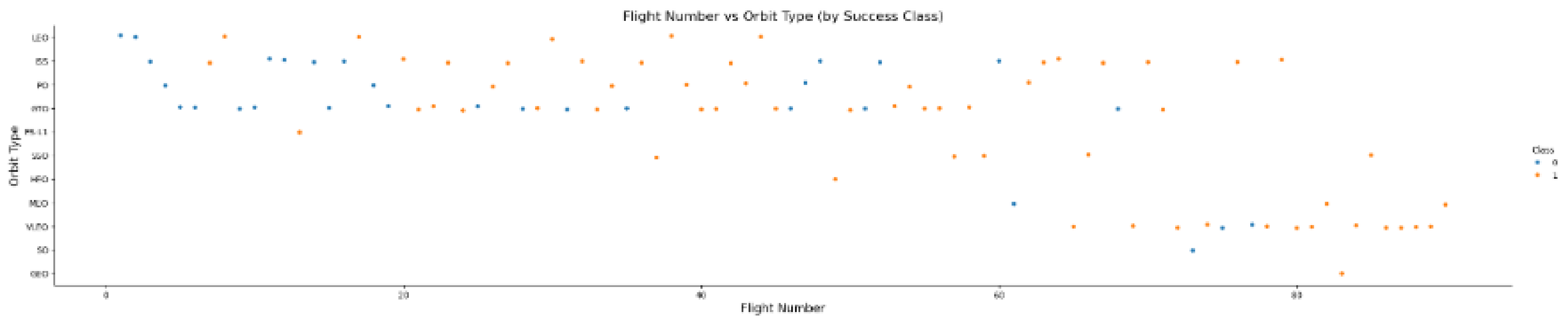
- Show a bar chart for the success rate of each orbit type

- ES-L1, GEO, HEO, and SSO, success rate of 100%.



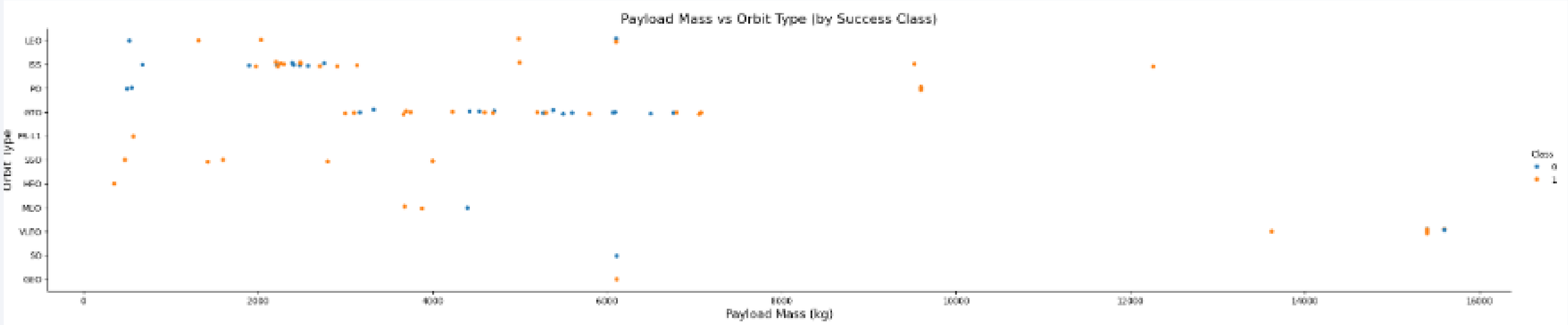
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Some orbit types started later than others.



Payload vs. Orbit Type

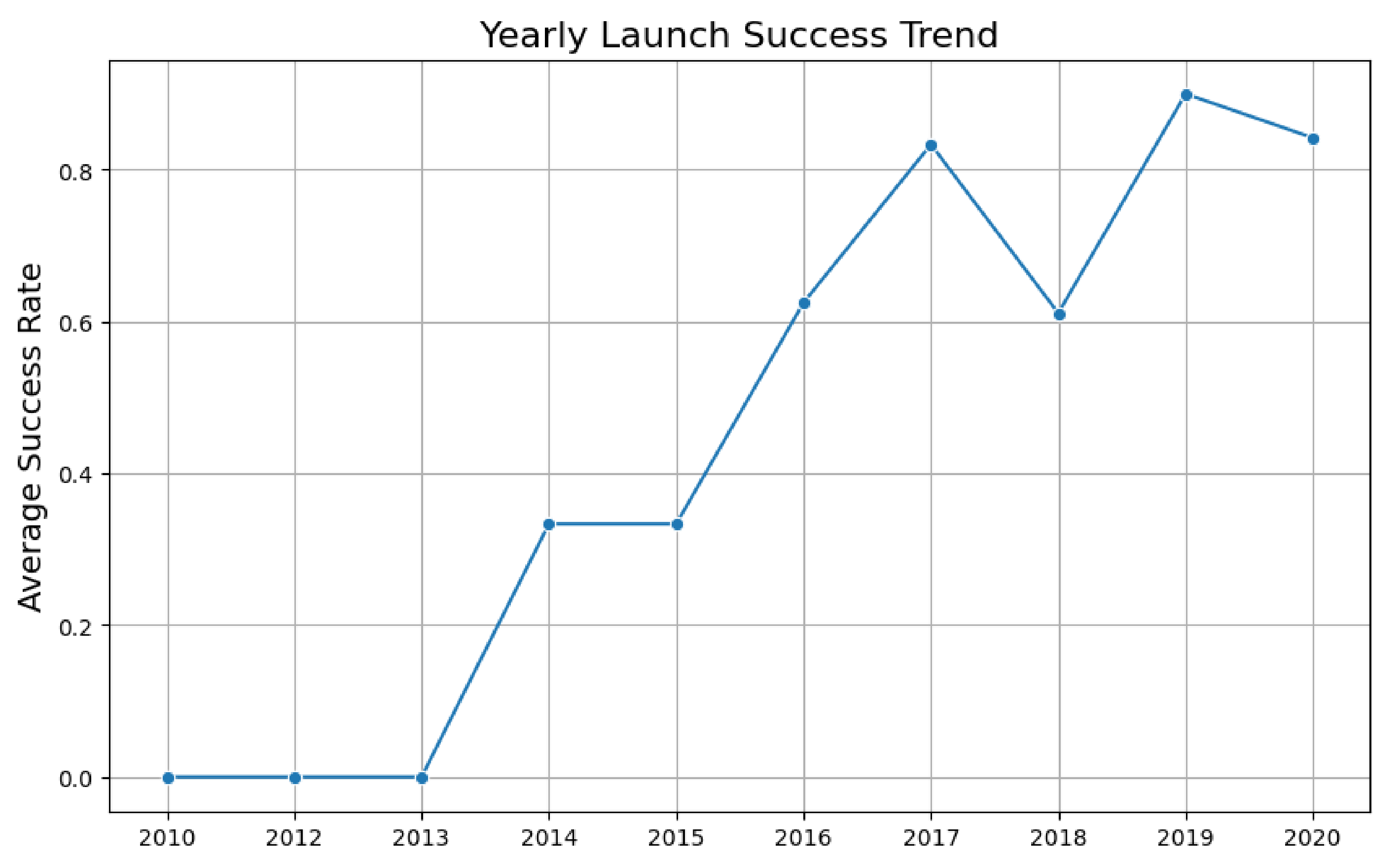
- Show a scatter point of payload vs. orbit type
- Paylod Mass 4000 to 6000 concentrates most orbit types.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- 2013 to 2014: increased;
- 2014 to 2015: neutral;
- 2015 to 2017: increased;
- 2017 to 2018: decreased;
- 2018 to 2019: increased;
- 2019 to 2020: decreased;
- Overall: increased average success rate.



All Launch Site Names

- Find the names of the unique launch sites
- 4 unique launch site names found

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with \`CCA\`
- Launch records found with site names beginning with CCA presented failure or no attempt.

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

- Calculate the total payload carried by boosters from NASA

- Total Payload Mass
    - 45596 Kg

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

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

- Average Paylod Mass by F9 v1.1:
  - 2928,4 Kg

Avg_Payload_Mass
2928.4

First Successful Ground Landing Date

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

- First successful ground landing occurred in 22 december 2015.

First_Ground_Pad_Success_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

- 4 Boosters with 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 success 98% rate.

Mission_Outcome	Total_Count
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

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (
SELECT MAX("PAYLOAD_MASS__KG_")
FROM SPACEXTABLE
);
```

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

- Failed landing identified in the database.

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

- Usual outcome is 50% chance of success landing.

Landing_Outcome	Outcome_Count
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 Maps – Global Map Markers

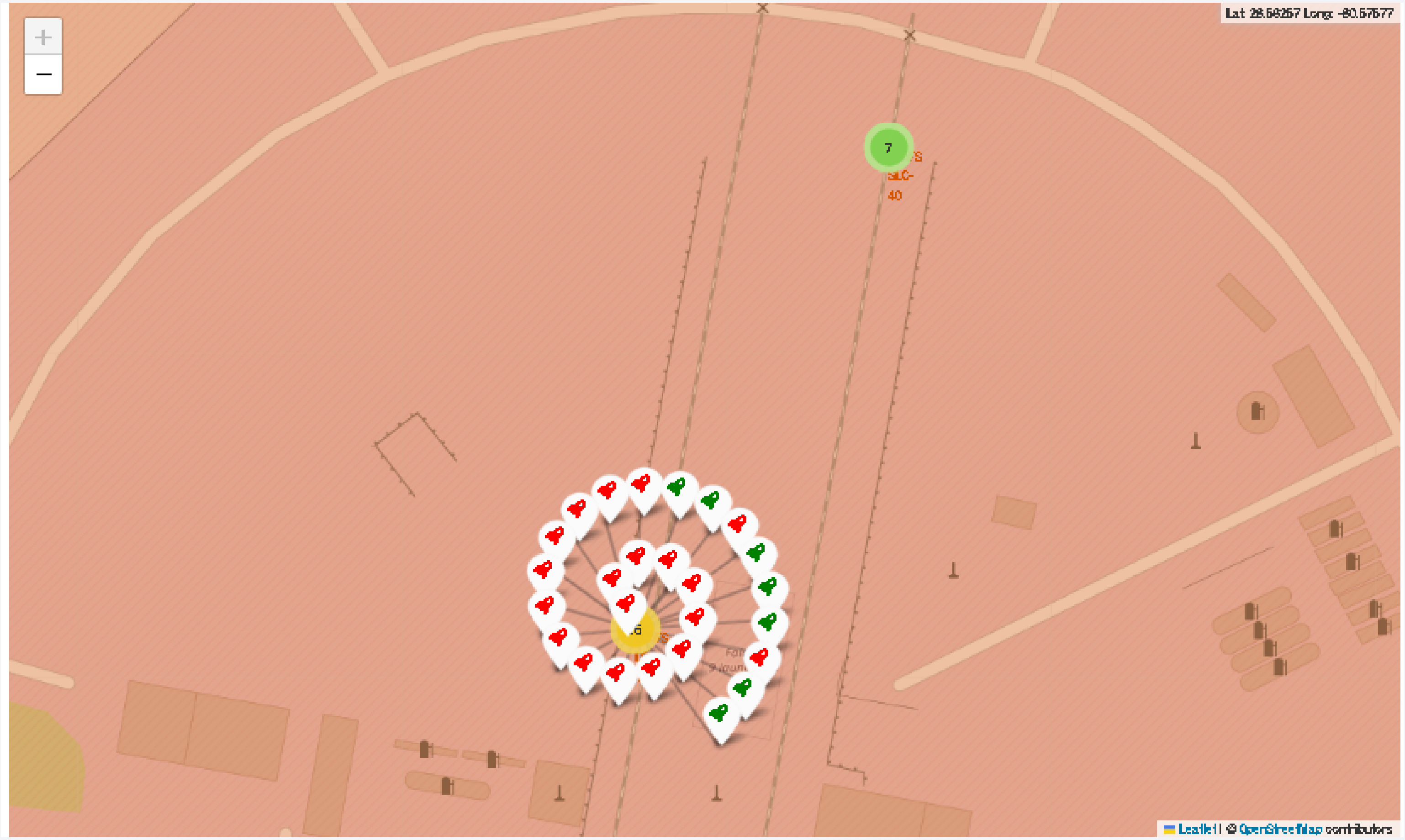
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

- Lauch sites are always next to coastlines, east or west.

Folium Maps – Color-labeled launch outcomes

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

- Green marker for failed launches.



Folium Maps – Distance to coastline

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Distance from launch site to coastline: 0,90 km.

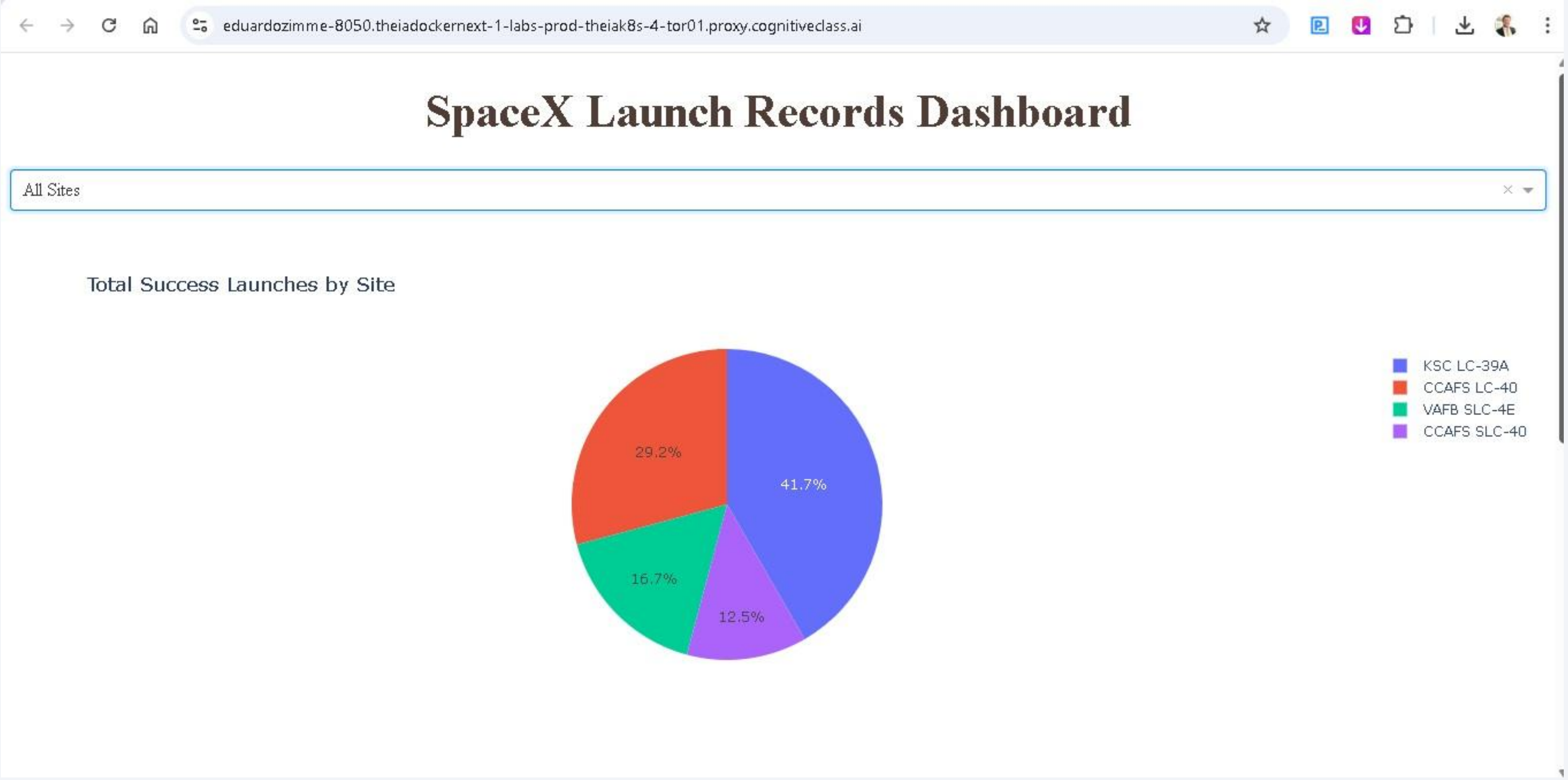
Section 4

Build a Dashboard with Plotly Dash

SpaceX Launch Records – All sites

- Show the screenshot of launch success count for all sites, in a piechart

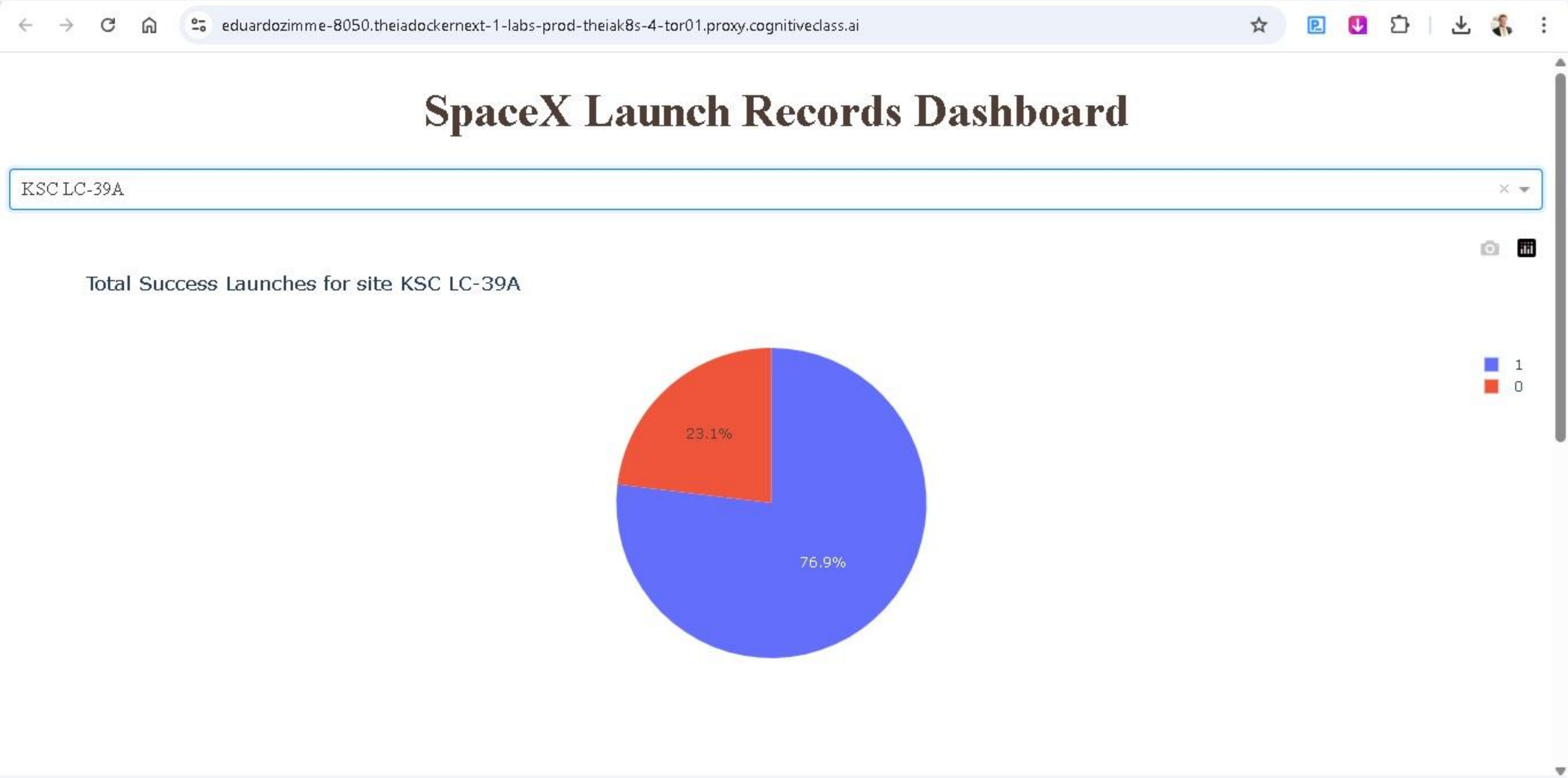
- KSC LC-39A: 41,7% of total success launches.



KSC LC-39A – success launch rate 76,9%

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Launch Success by Site (Pie Chart): KSC LC-39A achieved ~77% success.



CAFS LC-40 – Payload range 2000 to 8000 (Kg)

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

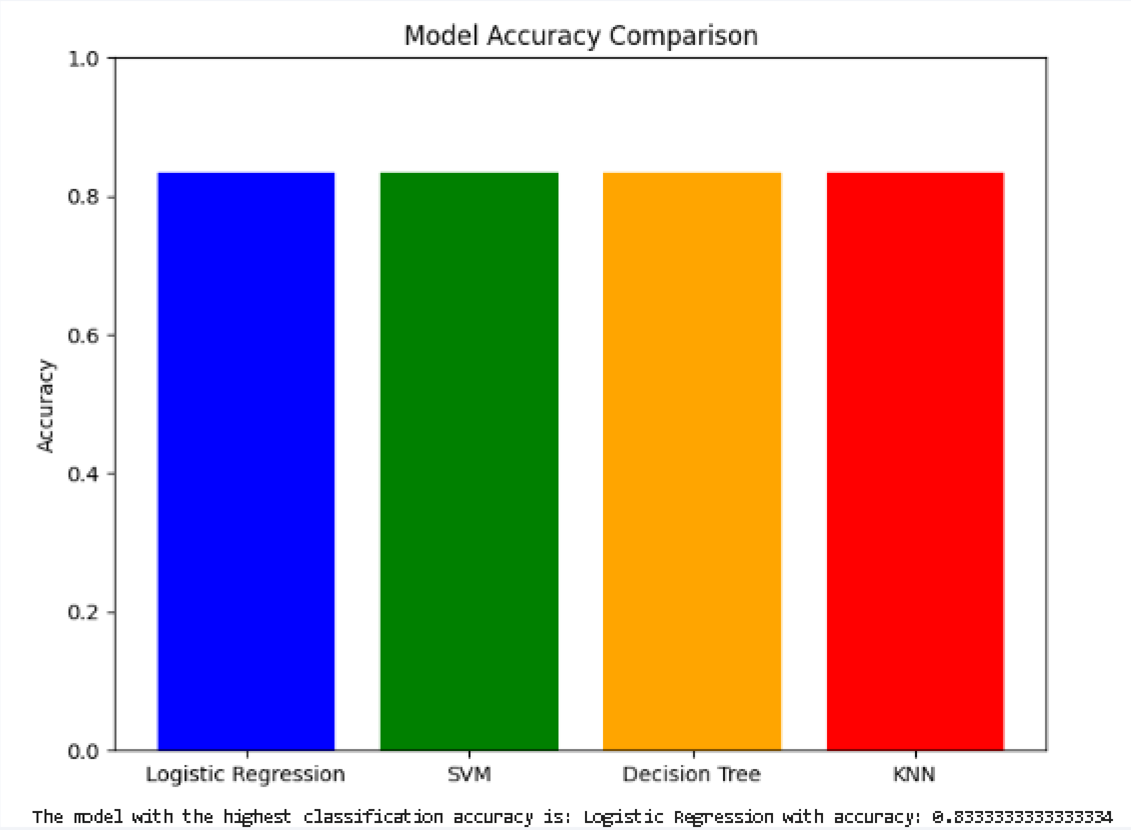
- Payload vs. Outcome Scatter (Interactive Slider): Certain payload ranges show higher probability of success.
- Filters Enabled: Interactive selection by site, payload range, and booster version.

Section 5

Predictive Analysis (Classification)

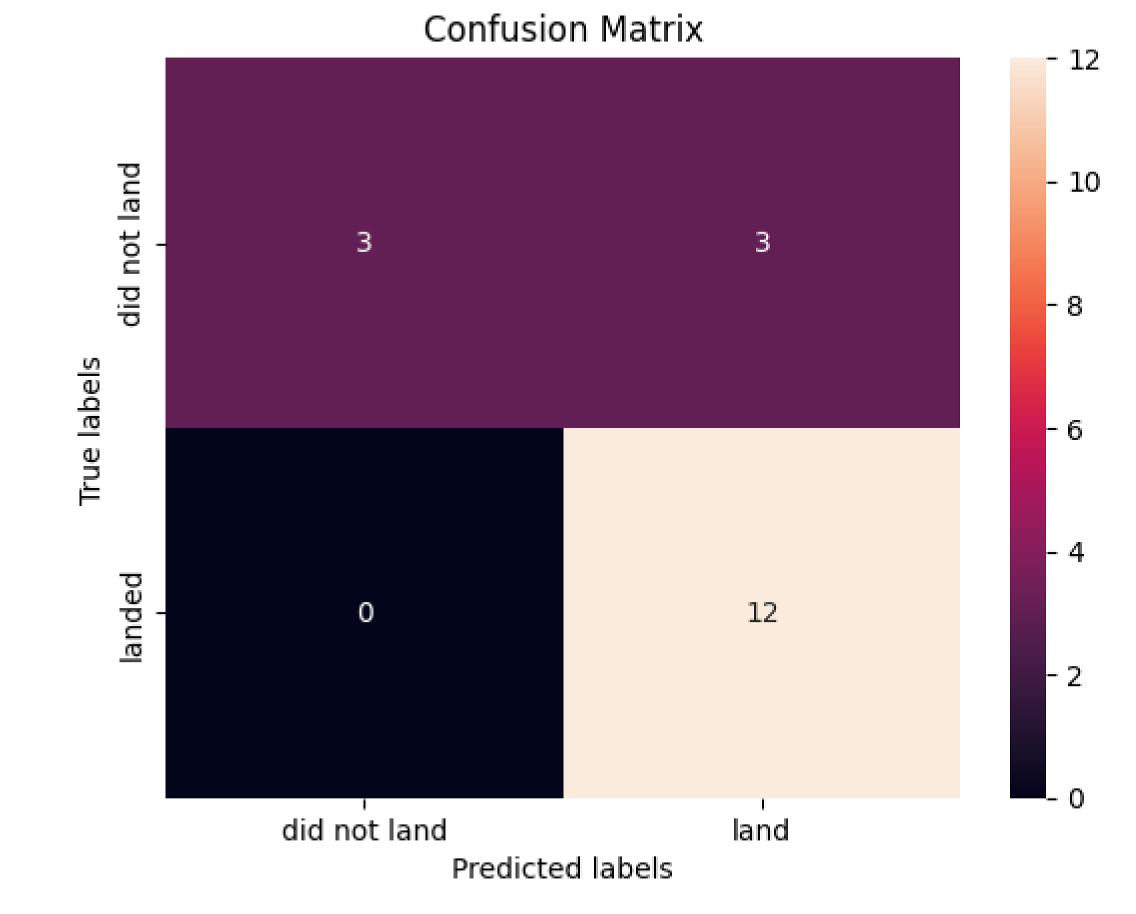
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Models tested: Logistic Regression, Decision Tree, SVM, KNNLogistic Regression achieved the highest accuracy: 83.3%Other models (SVM, Decision Tree, KNN) performed similarly (~83%), but not betterLogistic Regression was selected as the final model for interpretation
- Find which model has the highest classification accuracy
  - Logistic Regression: 83,33%



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
- Strengths:Perfect recall for successful landings (no false negatives)High precision on positive outcomes (landed)
- Weaknesses:Some false positives: predicted “land” when booster did not land (3 cases)
- Interpretation: Model is more reliable at confirming successful landings than predicting failures.

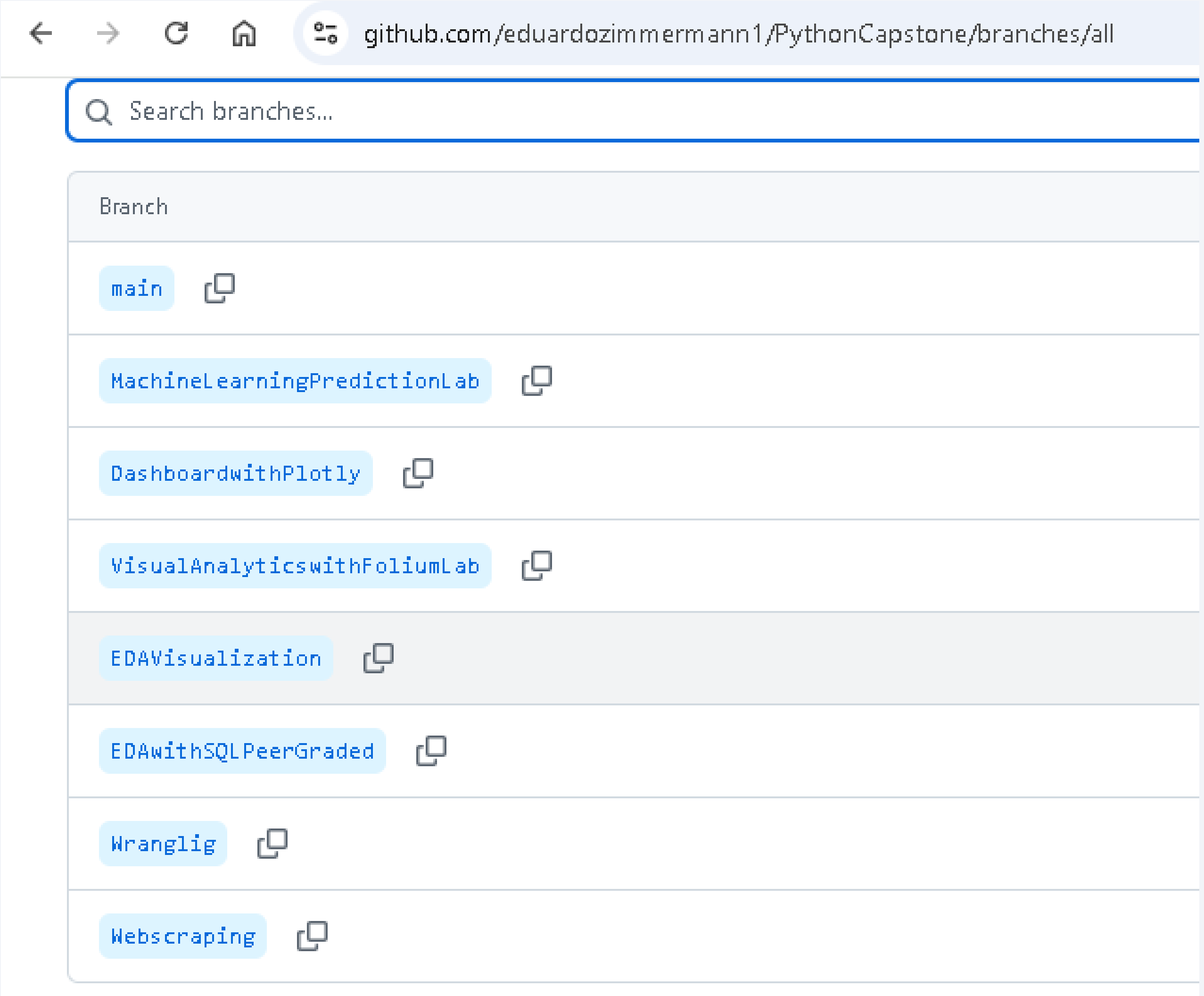


Conclusions

•**Findings Recap:**
•Mission success strongly influenced by payload mass and launch site.
•Success rate steadily improved over time.
•Interactive tools (Folium, Dash) enhance understanding of spatial and contextual factors.
•Predictive modeling achieved reliable performance (Logistic Regression best).
•**Implications:**
•Provides a foundation for decision-making on mission planning and risk assessment.
•Framework can be extended with more granular engineering and environmental data.
•**Next Steps:**
•Expand dataset with new launches.
•Experiment with deep learning for predictive modeling.
•Deploy dashboards for real-time monitoring.

Appendix

- GitHub Repo and view of all branches.



Thank you!