IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Patrick Rutledge
June 11, 2025

Outline

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

Executive Summary

- Summary of methodologies

  The question, problem and source of data is determined by SpaceY. The project methodology:

  1. **Collect and understand** type of data from SpaceX API and from Webscraping Wikipedia
  2. **Data Wrangling** - convert and transform data for visualization and analysis
  3. **Analysis** – select, train and optimize models: Logistic Regression, SVM, Decision Tree, KNN
  4. **Evaluate** – determine most accurate models
  5. **Report** – share findings of models in narrative and visualizations

- Summary of all results

  Reusing rocket components reduces the cost enough to affect the price of launches.

  Reuse is a new value proposition and the strategy is viable.

Introduction

Project background and context

In the new space age, SpaceX leads with successful, low-cost launches. Competitors can "fast-follow" by enhancing SpaceX's offerings and leveraging its market pioneering. This project aims to determine launch cost predictability and competitive strategies based on rocket component reusability, crucial for building a competing business case.

Problems you want to find answers

Specific Questions:

1. What is the viability of reusing rocket components?
2. What factors contribute to successful reuse of rocket components?
3. Can the launch cost be estimated?
4. Is the price of launch competitive when components are reused?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology: 2 sources, JSON & HTML
- Perform data wrangling: Panda & Libraries
- Perform exploratory data analysis (EDA) using visualization and SQL: Charts returning filtered and aggregates from DB
- Perform interactive visual analytics using Folium and Plotly Dash: Interactive & geographic visualizations
- Perform predictive analysis using classification models: Common models selected, optimized, trained and evaluated

Data Collection

- JSON files using get request from SpaceX REST API
- HTML File scraped from Wikipedia with BeautifulSoup

Extract the launch records

Data Collection – SpaceX API

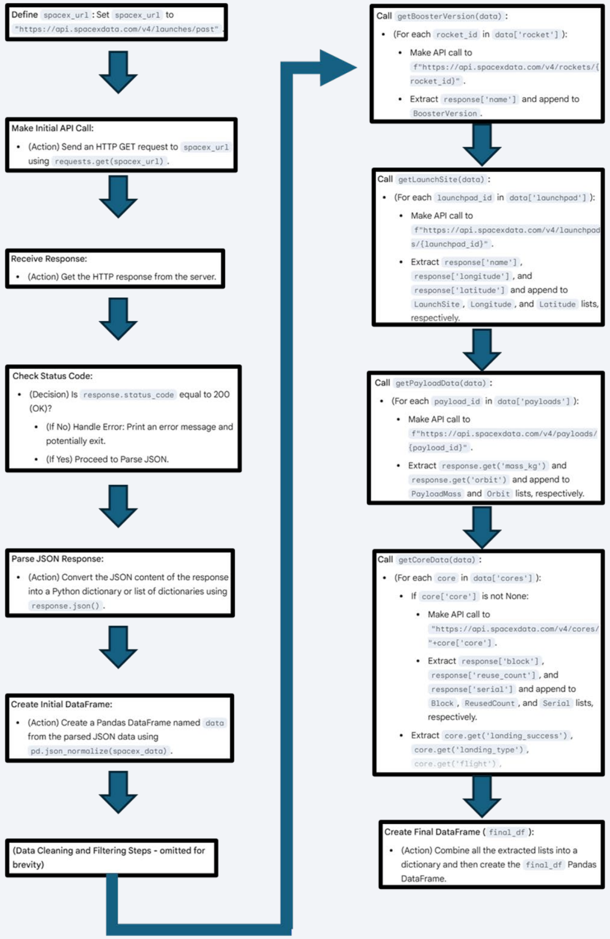
- Flowchart with phrases and calls

https://github.com/PatrickRutledge/IBM/blob/main/Spacex%20API%20Flowchart%20pdf.pdf

- GitHub URL

Jupyter Notebook

https://github.com/PatrickRutledge/IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

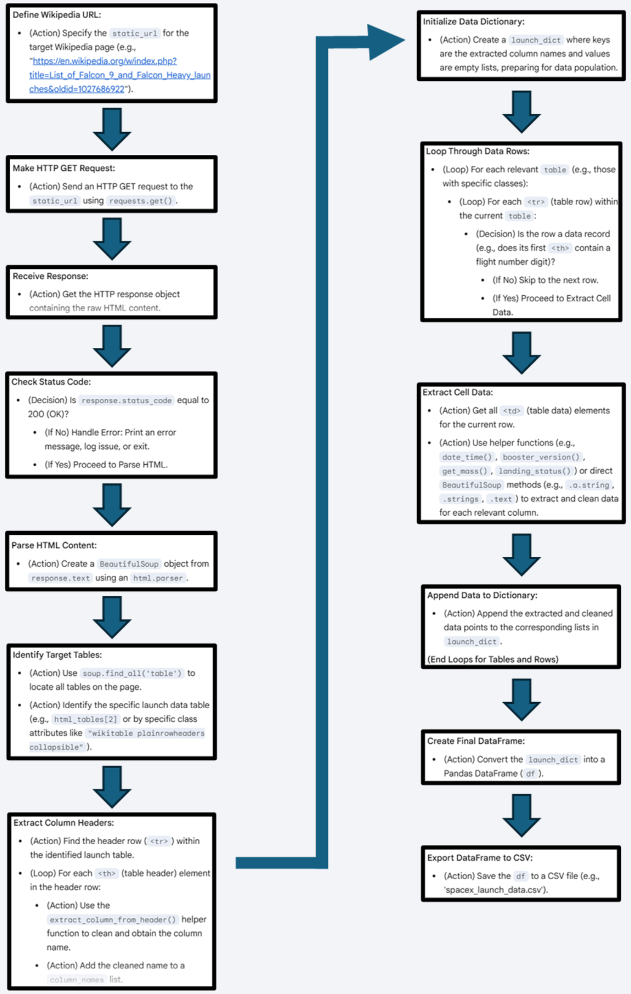
- Flowchart with phrases and calls

https://github.com/PatrickRutledge/IBM/blob/main/Spacex_webscraping_flowchart.pdf

GitHub URL

Jupyter Notebook

https://github.com/PatrickRutledge/IBM/blob/main/Falcon%209%20Webscraping%20Code.ipynb



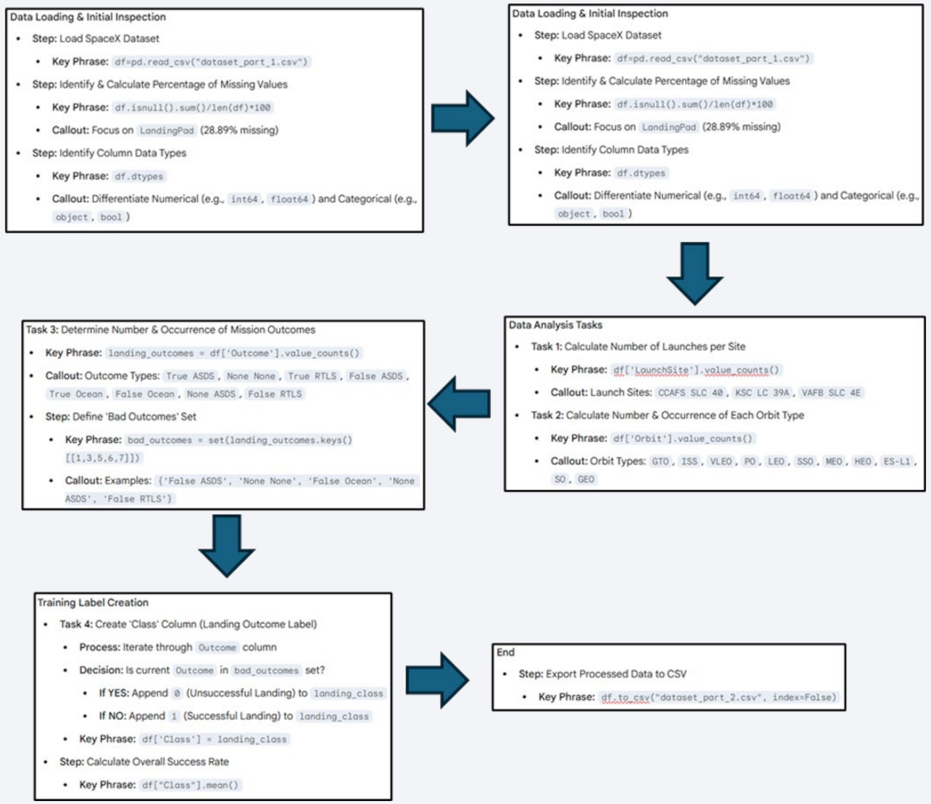
Data Wrangling

- Data inspected, cleaned, transformed, filtered, new features added and loaded into dataframes.

[DataWranglingFlowCart]

- Github URL for Jupyter Notebook

https://github.com/PatrickRutledge/IBM/blob/main/SpaceX_data%20wrangling_flowchart.pdf



EDA with Data Visualization

- Summarize what charts were plotted & why you used those charts

Five (5) Scatter Plots

1. Visualize launch outcome influenced to flight number & Payload
2. Observe relationship of launch outcome to flight number and launch site
3. Examine relationship of launch outcome to launch site and payload mass
4. Visualize relationship between flight number and orbit type considering launch outcome
5. Reveal relationship between payload mass, orbit type and landing success

One (1) Line Chart

1. Visualize the average launch success over the years

One(1) Bar Chart

1. Check the success rate for each distinct orbit type

- GitHub URL https://github.com/PatrickRutledge/IBM/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

- Retrieve all distinct launch facilities or sites
- Filter Launch Sites Starting with CCA
- Calculate Total Payload Mass for missions by customer, Nasa (CRS)
- Caclulate Average Payload Mass for Specific Booster
- Find Date of First Successful Ground Pad Landing
- List Boosters with Specific Payload Mass landing on a Drone Ship
- Identify Boosters with Maximum Payload Mass
- Filter Failed Drone Ship Landing in 2015
- Rank Landing Outcomes by Count and Date Range

GitHub URL: IBM/Assignment SQL Notebook for peer assignment_sqllite.ipynb at main · PatrickRutledge/IBM

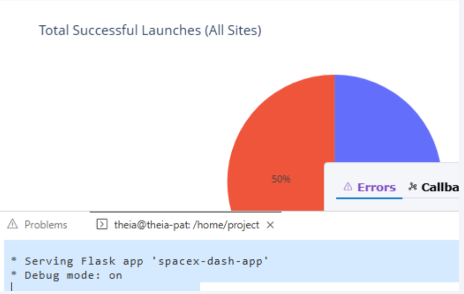
Build an Interactive Map with Folium

- folium.Map with zoom level is foundation.
  - Folium.Circle, Launch sites at exact coordinates to visually pinpoint locations
  - Folium.map.Marker with DivIcon, display the sites name
  - Folium.plugins.MarkerCluster, prevent overcrowding on the map
  - Folium.Marker,Folium.Icon, color coded for failed and successful launches
  - Folium.plugins.MousePosition, display latitude and longitude of mouse position on map
  - Folium.Polyline, show distance to closest coastline, city, railway, highway and beerjoint

GitHub URL [IBM/lab-jupyter-launch-site-location-v2.ipynb at main · PatrickRutledge/IBM]

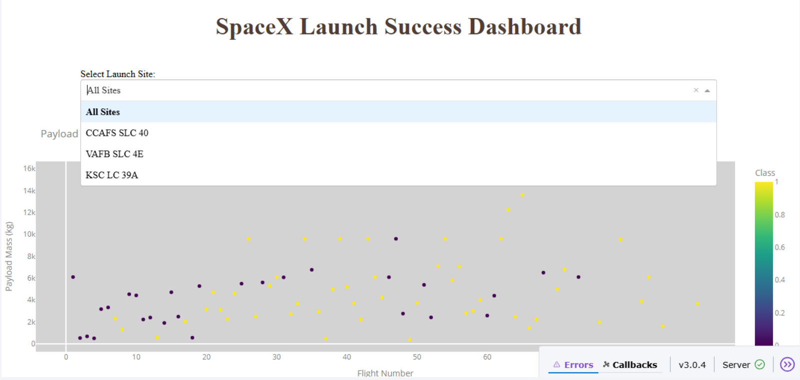
Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard



Pie chart: Total successful launches

Scatter Plot: Payload Mass vs. Flight Number by Launch Outcome with filtering by location



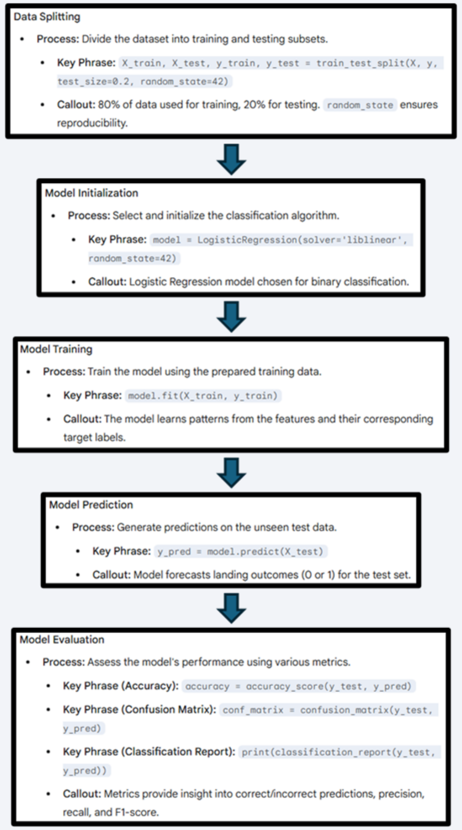
Purpose: Visualize relationship success across different launch sites.

GitHub URL: https://github.com/PatrickRutledge/IBM/blob/main/spacex-dash-app.py

********APP instructions are in Readme************

Predictive Analysis (Classification)

A combined dataset was built from previous labs and then logistic regression was performed. After training the model's performance was evaluated using accuracy score, confusion matrix and classification repot.



Flow Chart: https://github.com/PatrickRutledge/IBM/blob/main/Predictive%20Analysis_Classification%20flow%20chart.pdf

GitHub URL: https://github.com/PatrickRutledge/IBM/blob/main/Predictive%20Analysis_Classification%20flow%20chart.pdf

Results

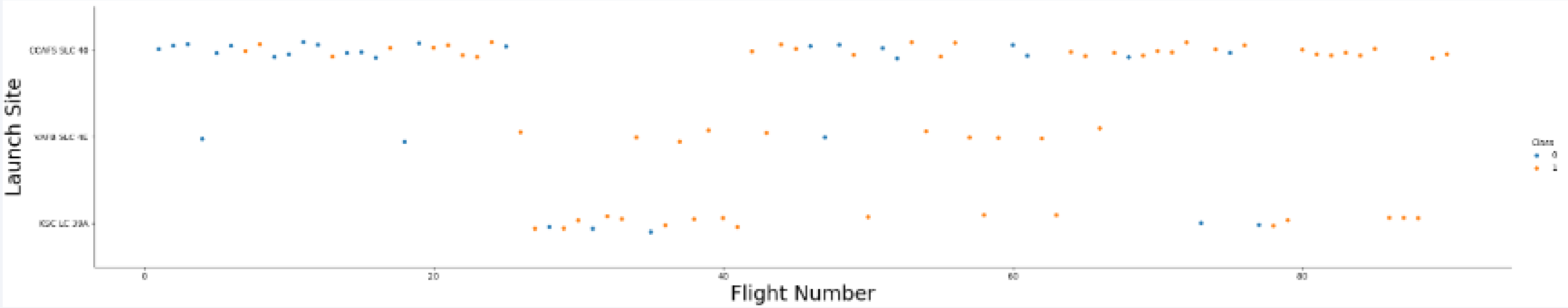
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

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

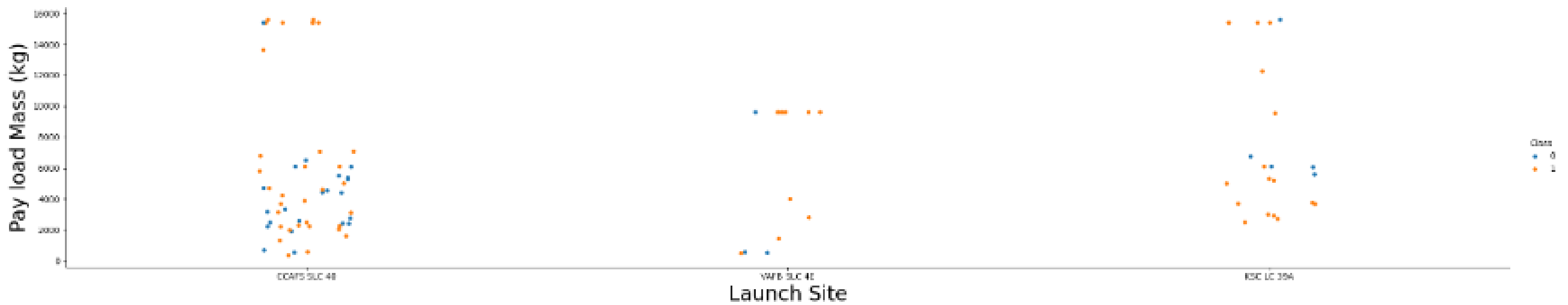


Explanations

Points correspond to specific launch sites and the flight number, the color of the point represents outcome. Success/Failure patterns by location can be recognized and the progression of flight numbers show the trends of launches because of sequential flight numbers. This plot allows for quick id of correlations and patterns.

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



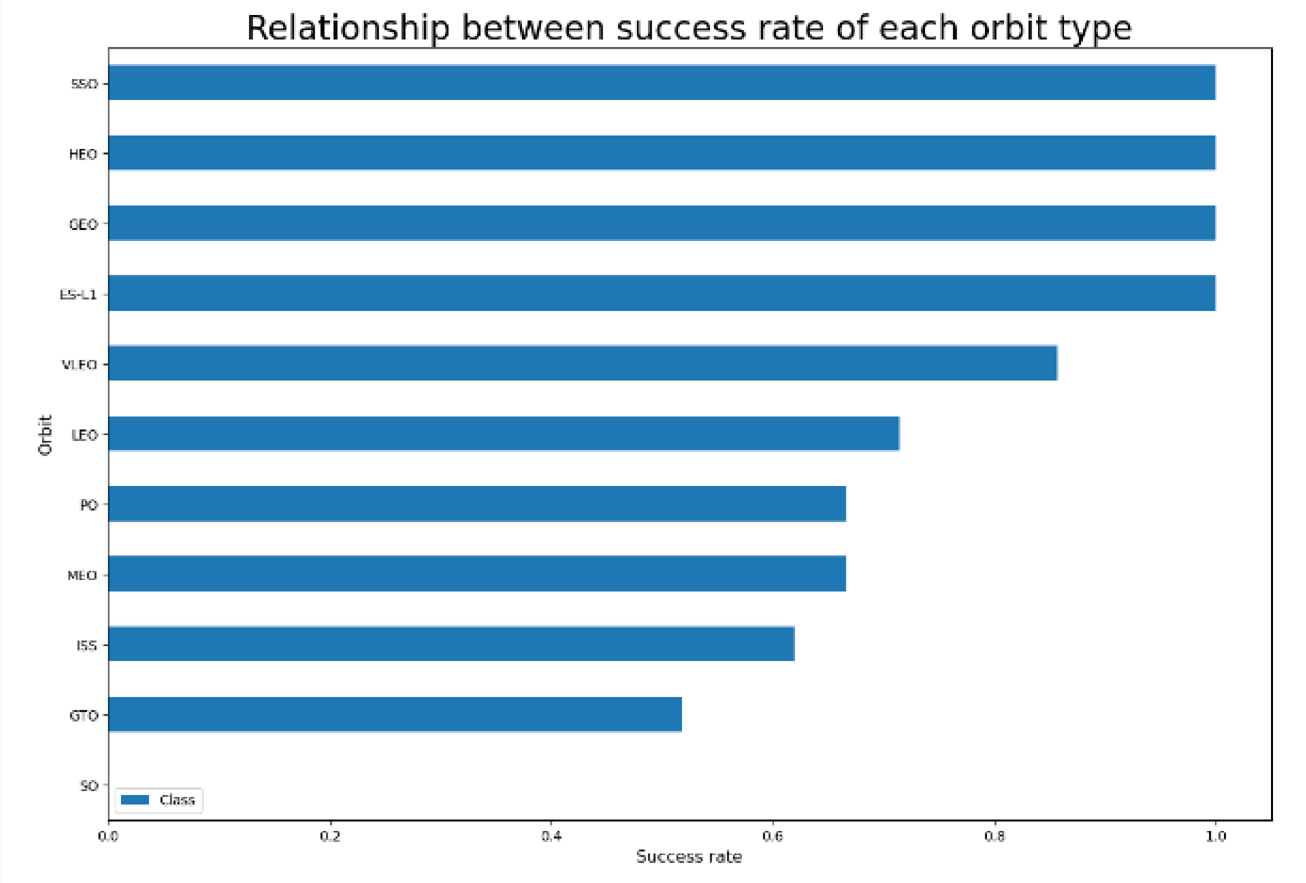
Explanations

Heaviest payloads launched on CCAFS SLC 40 were all successful, VAFB SLC 4E no heavy launches, KSC LC 39A high success of heavy launches, VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000

Success Rate vs. Orbit Type

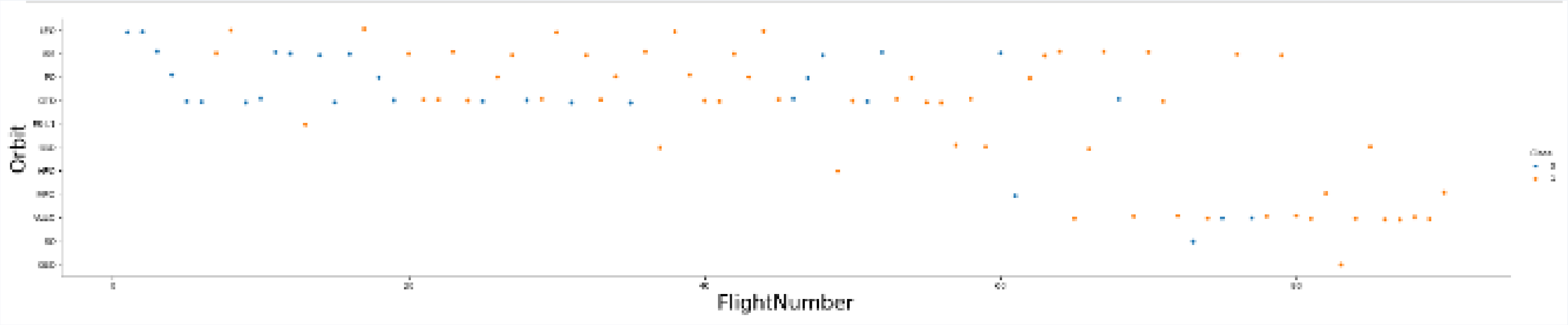
Explanations:

Orbit type is a strong predictor of landing success. Show gap where biggest savings for reuse could be achieved.



Flight Number vs. Orbit Type

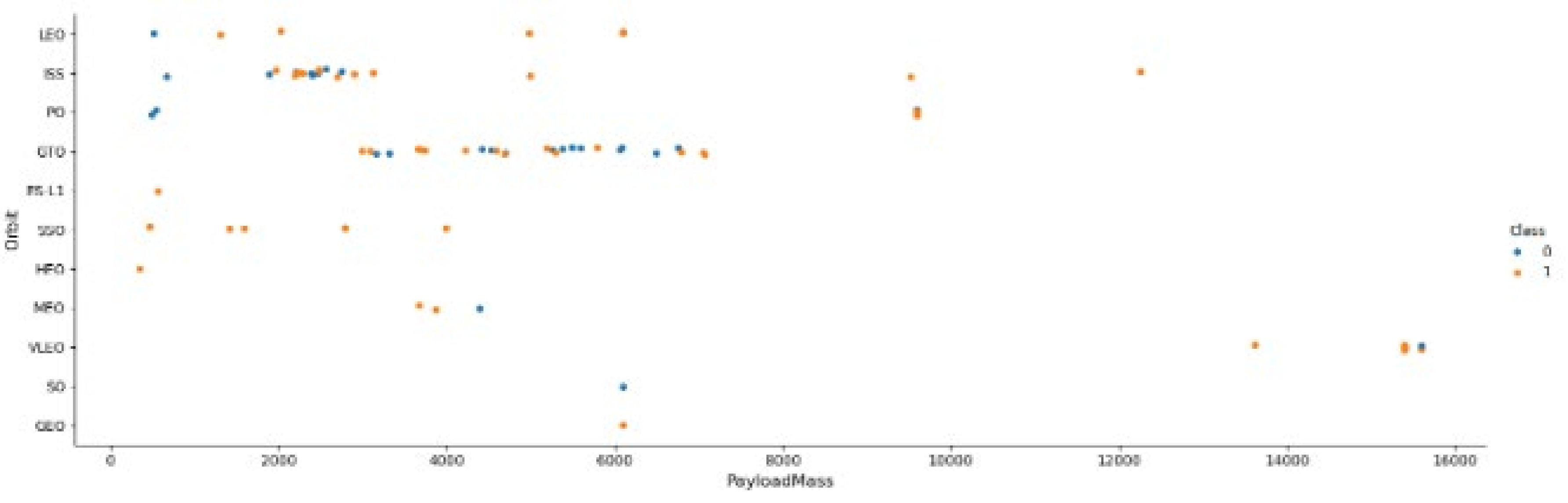
- Show a scatter point of Flight number vs. Orbit type



Explanations: LEO orbit success related to the number of flights, but no relationship between flight number & GTO Orbit

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

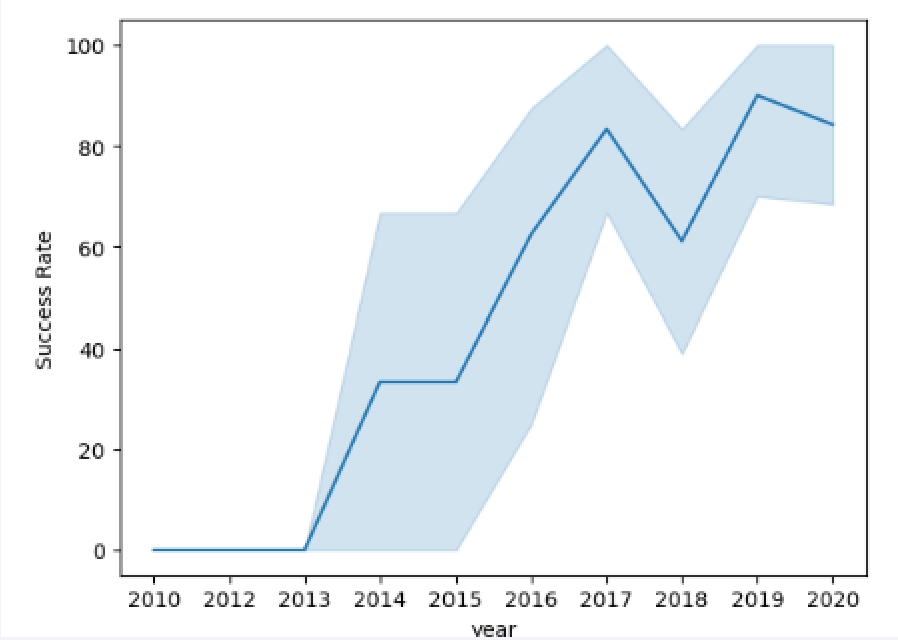


Explanations:

Polar, LEO and ISS have more success with heavy payloads. Cannot distinguish this very well because both the positive landing rate and negative landing rate are present.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



Explanations:

Three distinct increases: 2013, 2015, 2018

Two distinct decreases: 2017, 2019

All Launch Site Names

- Find the names of the unique launch sites

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

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

 * sqlite:///my_data1.db
```

- Present your query result with a short explanation here

We used SQLite to create a .db database file. We queried the database, specifically a table: SAPCEXTABLE to return distinct locations from the launch site column.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Date	Time (UTC)	Booster_Versio	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0[illegible]03	CCAFS LC-40	Dragon [illegible]pacecraft Q[illegible]lification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 [illegible]004	CCAFS LC-40	Dragon de[illegible]o flight C1, two [illegible]beSats, [illegible]arrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B[illegible]005	CCAFS LC-40	Dragon d[illegible]o flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0[illegible]6	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B000[illegible]	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
* sqlite:///my_data1.db
```

The query was written using a pattern matching filter with a wildcard character '%'

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

SUM(PAYLOAD_MASS__KG_)
45596

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

```
* sqlite:///my_data1.db
```

Query for the spacextable using Sum function for payload mass column filtering by customer Nasa

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

AVG(PAYLOAD_MASS__KG_)
2928.4

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

```
* sqlite:///my_data1.db
```

Average function on the payload mass column in spacextable using "where" to filter the booster version column by 'F9 V1.1'

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

MIN(Date)
2015-12-22

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

Queried the my_data1.db using select and the MIN() COMMAND to return year from Date column in spacextable and WHERE function on Landing Outcome column by Success(ground pad)

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
- Present your query result with a short explanation here

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS_
```

```
* sqlite:///my_data1.db
```

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

```
* sqlite:///my_data1.db
```

Queried my_data1.db from spacextable to select distinct booster versions showing successful drone ship landings and payload mass between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

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

```
%sql SELECT "Mission_Outcome", COUNT(*) AS Total FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

```
* sqlite:///my_data1.db
```

Listing the total number of successful and failure mission outcomes. Query is counting and totaling then grouping results for all outcomes.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

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

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

```
 * sqlite:///my_data1.db
```

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

```
 * sqlite:///my_data1.db
```

Query for spacextable to return distinct booster versions where the payload mass was the max

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

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

%sql SELECT substr(Date, 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr(Date, 0, 5) = '2015';

NOTE** **use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.**

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

- Present your query result with a short explanation here

Landing_Outcome	OutcomeCount
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

*use between to return results of certain Date range from spacextable; count function to return outcome counts, group and order to filter and sort results

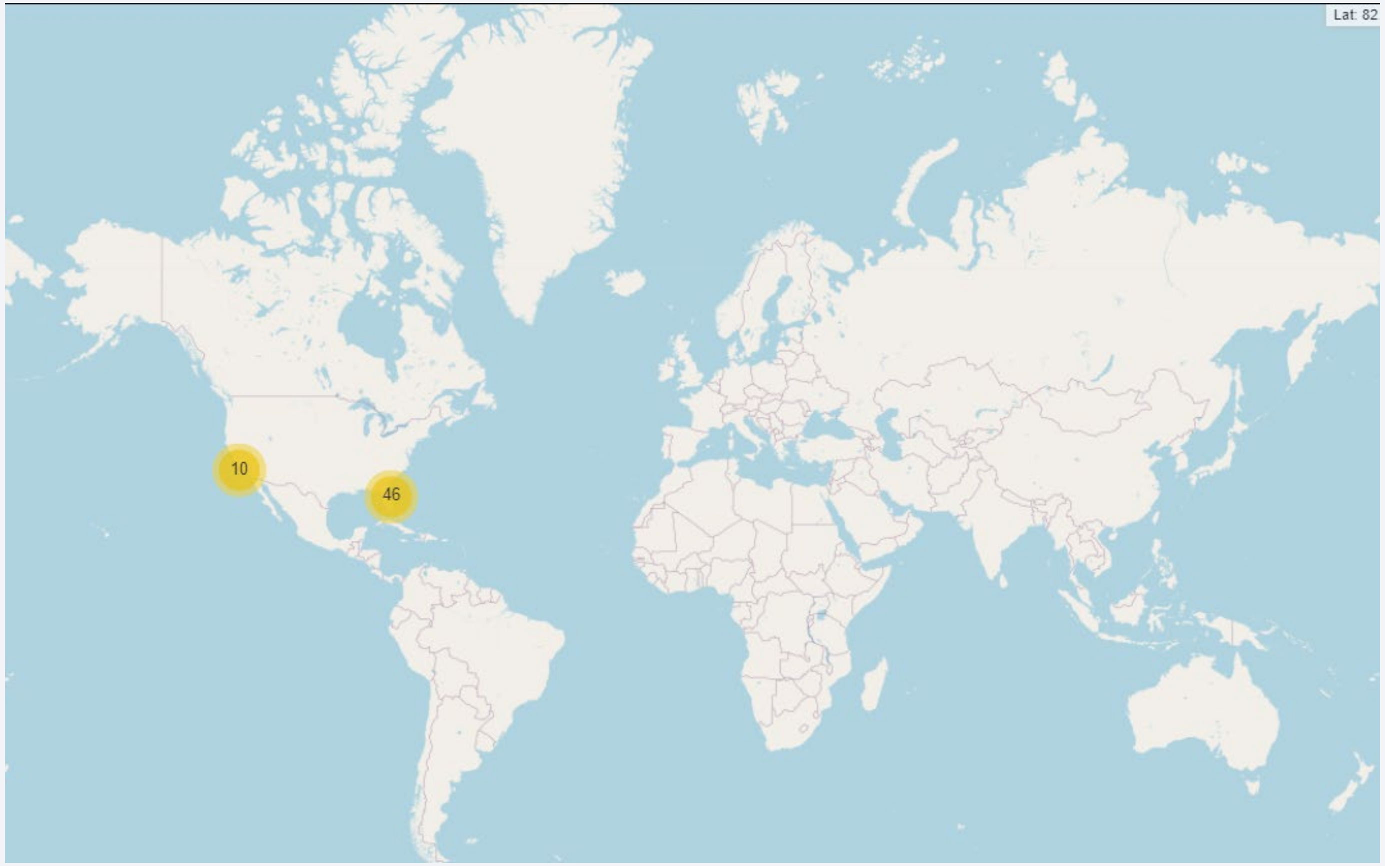
Query:

```
%sql SELECT "Landing_Outcome", COUNT(*) AS OutcomeCount FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY OutcomeCount DESC;
```

Section 3

Launch Sites Proximities Analysis

Launch Sites Global Map



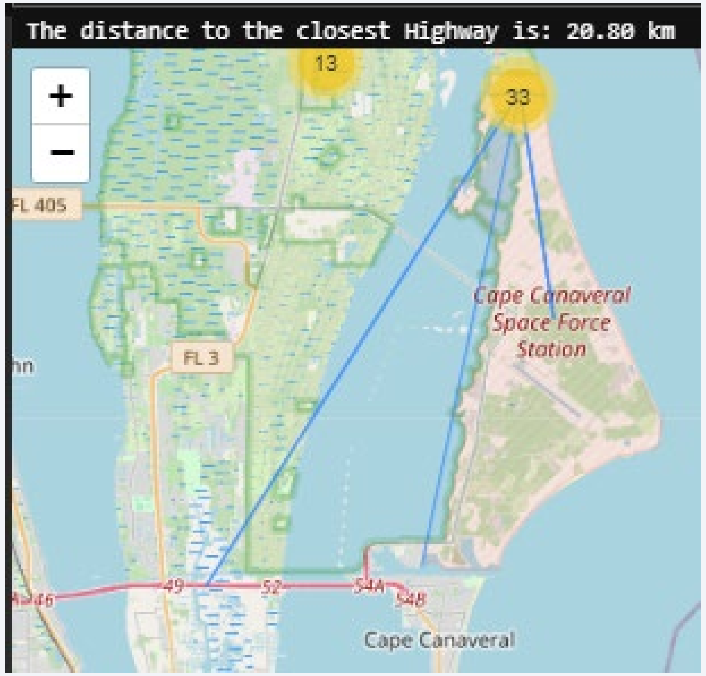
Launch Sites are near the coast and almost as close to the equator in the US as possible.

Launch sites appear to be strategically located to maximize efficiency and safety.

Color Labeled Launch Outcomes

- Zoom level is low enough to see one launch site labeled in green, the yellow dot is several launches grouped together. Hard to zoom in all the way without loosing site of launches.
- Circles represent the areas where rockets are launched and the color indicates successful and unsuccessful launches at those specific coordinates.
- Demonstrates the actual historical performance of the launch site.

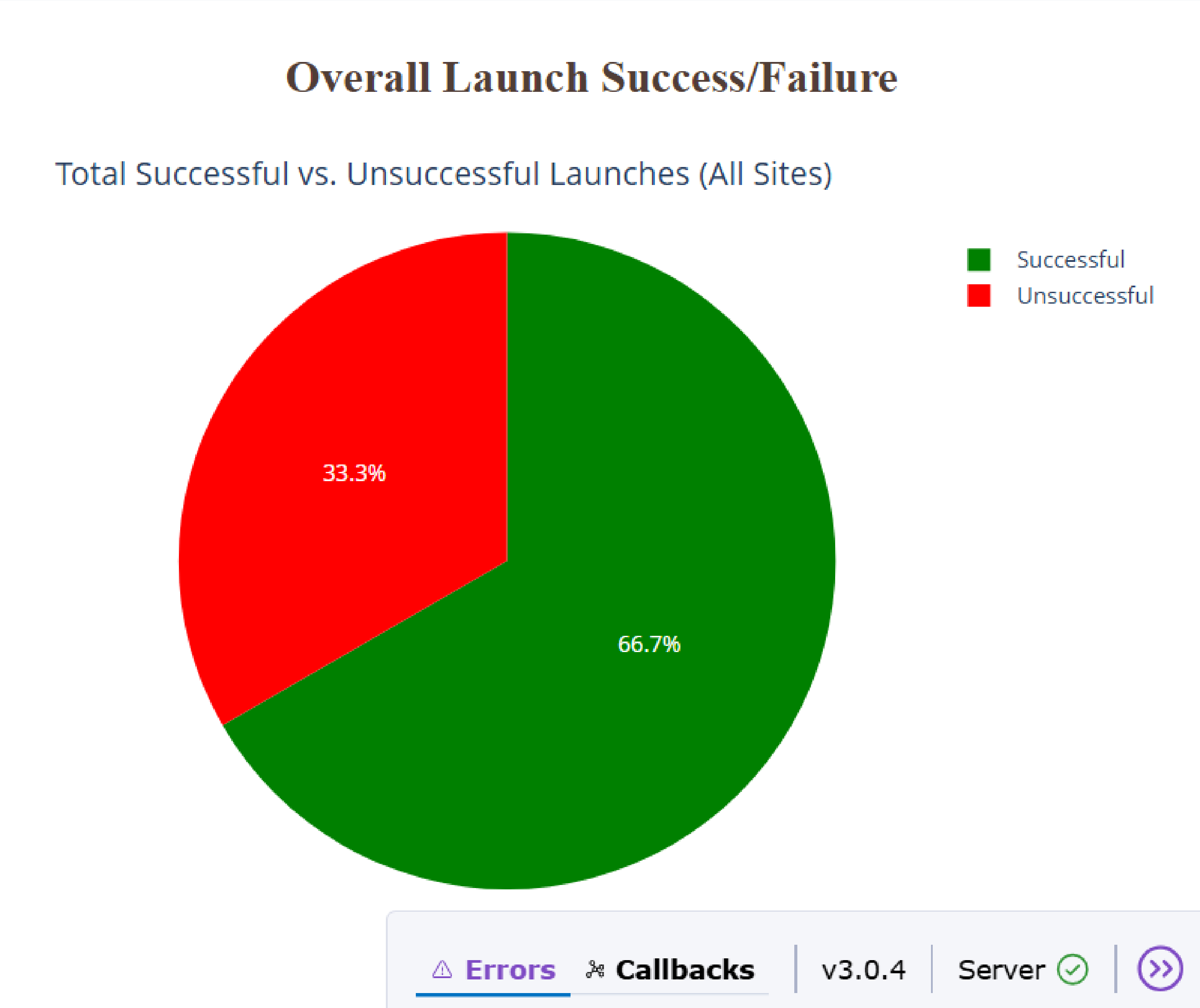
How Far is the Closest?



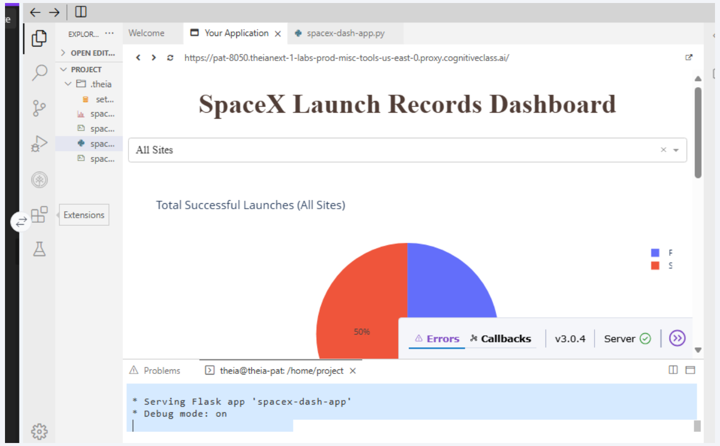
Section 4

Build a Dashboard with Plotly Dash

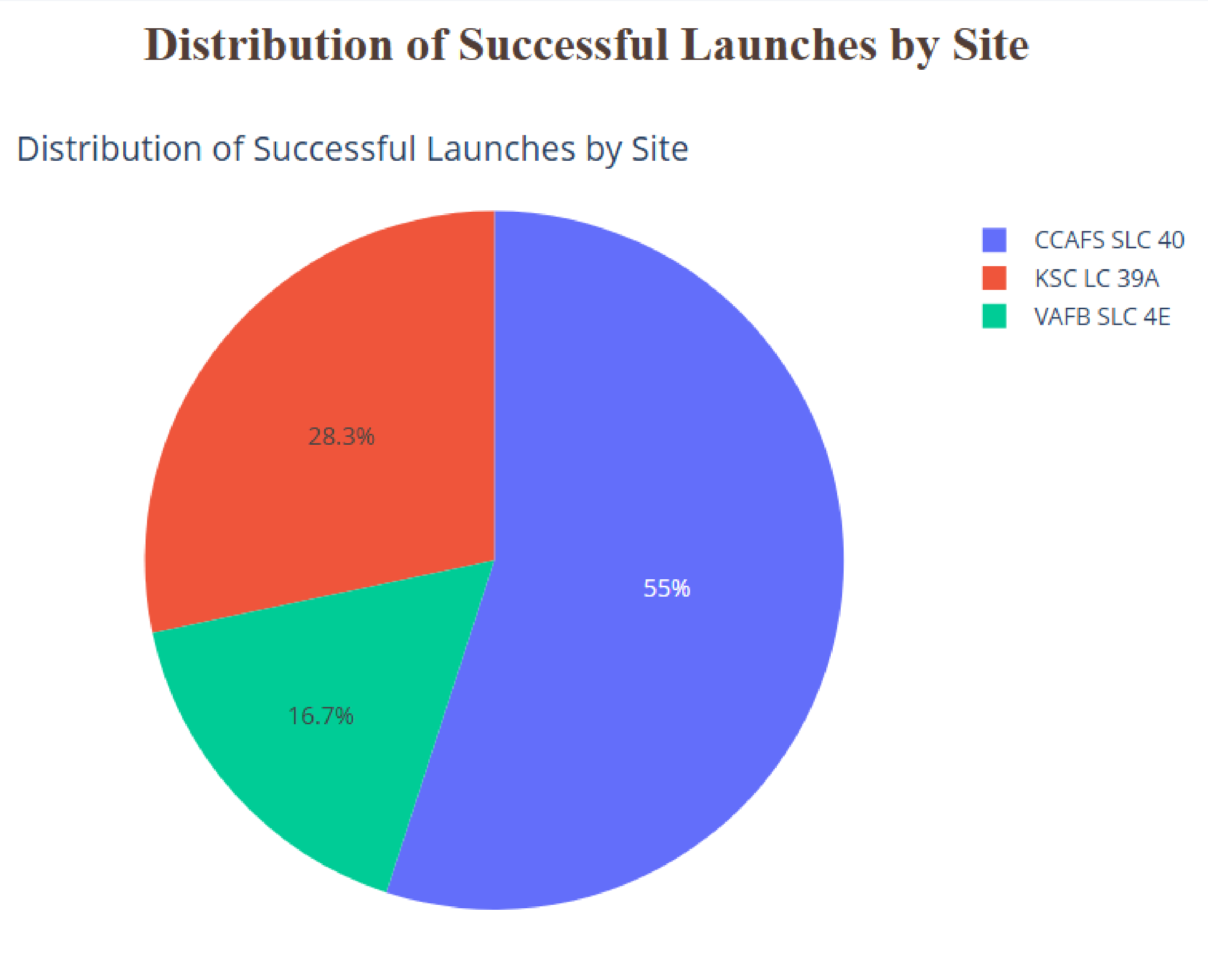
Overall Launch Success/Failure



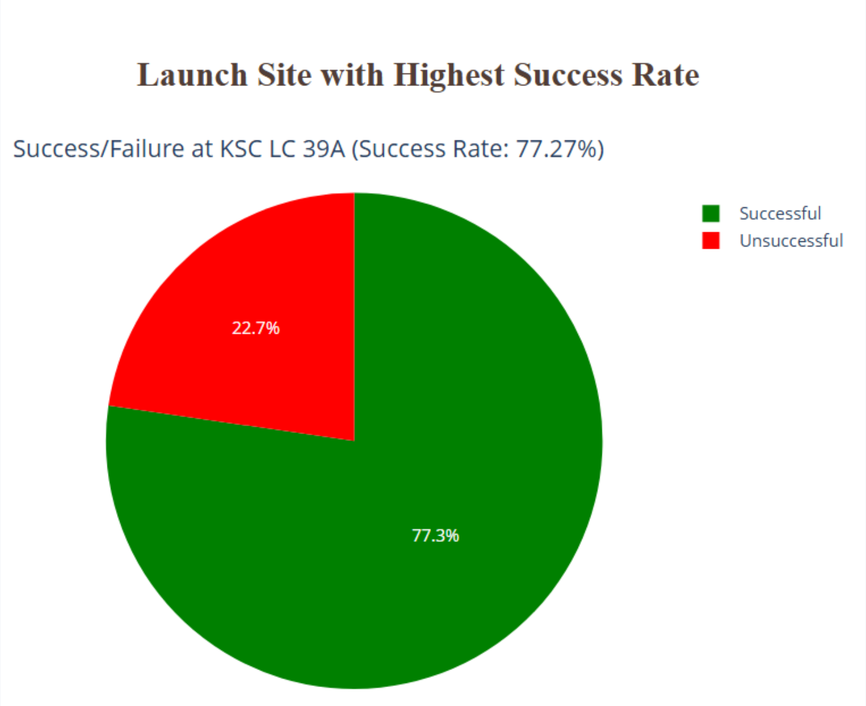
Dominant Success, More Green Less Red



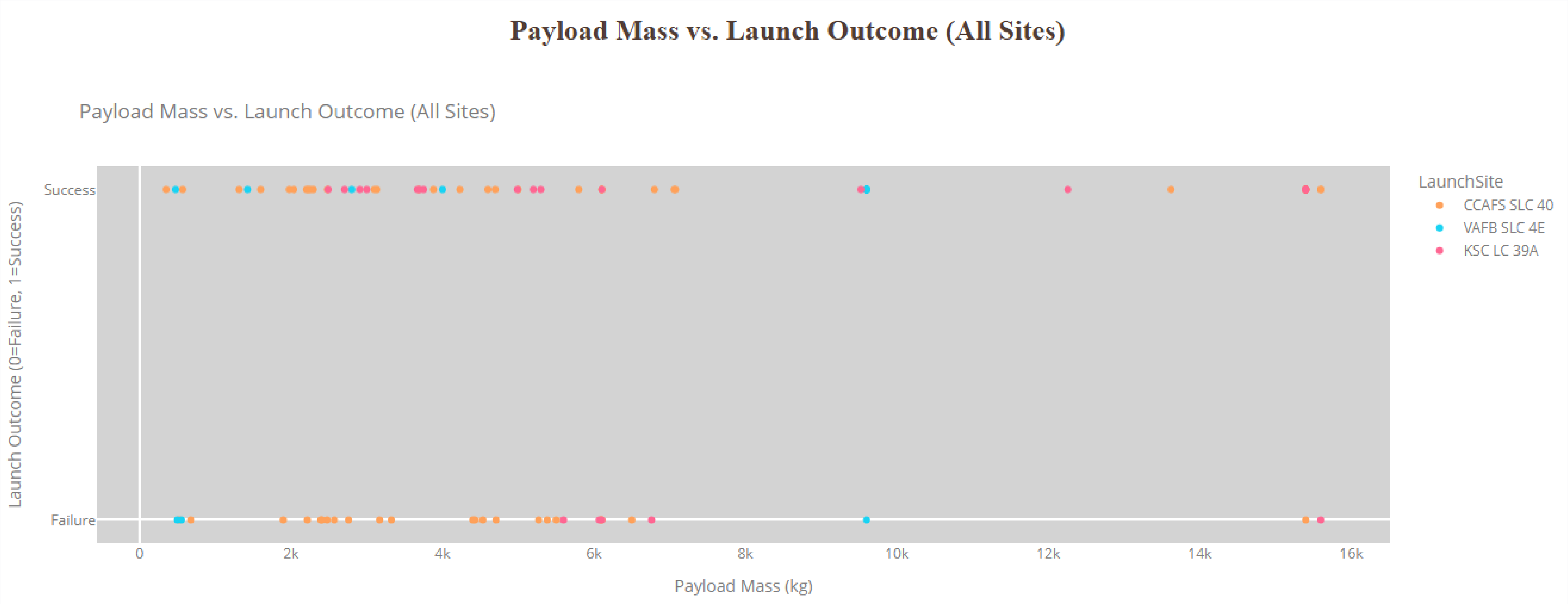
Successful Launches by Site & Highest Success Rate



Quickly and effectively communicates which the Success Rate for the Launch Site with the Highest Success Rate



Payload Mass vs. Launch Outcome Scatter Plot

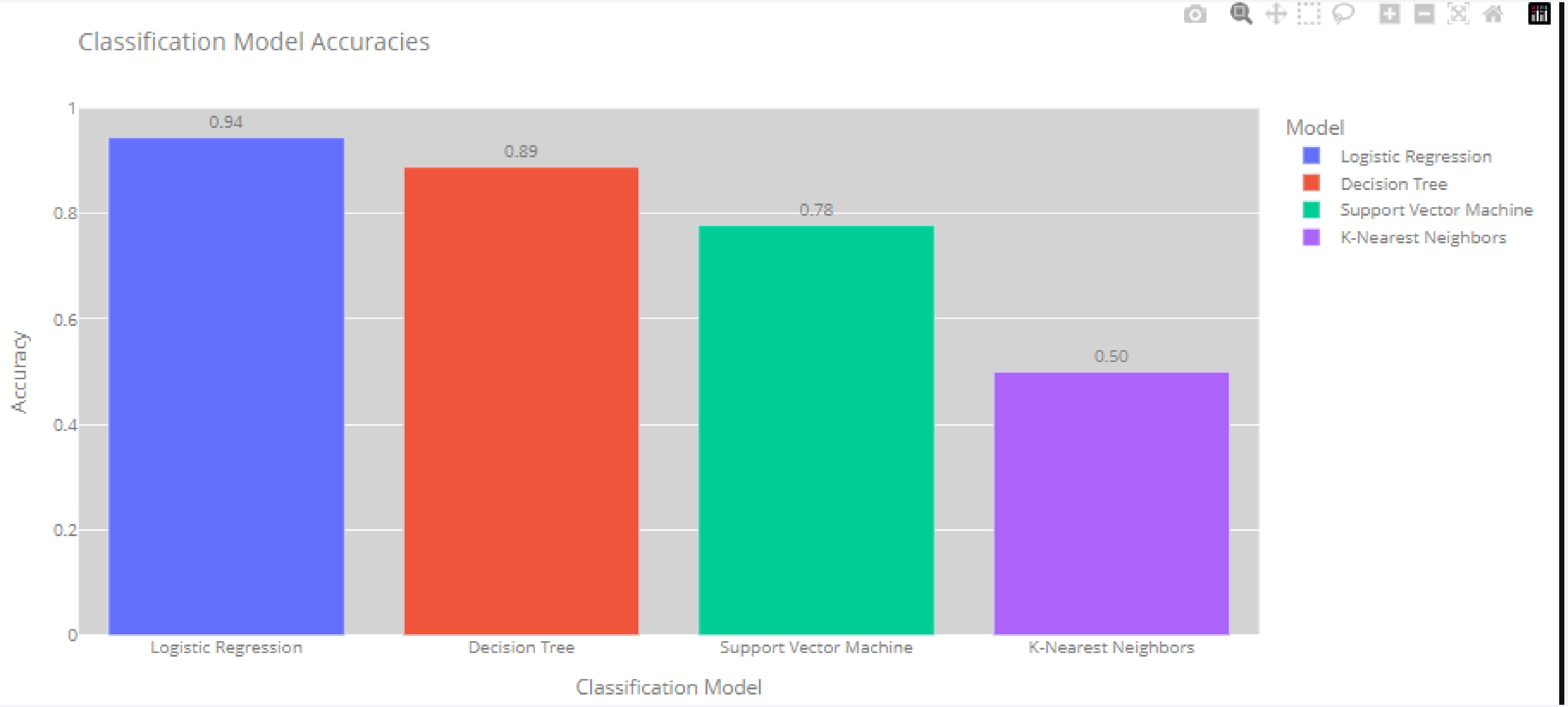


- Visualization shows the distribution of successful and unsuccessful launches across payload range while highlighting the launch site. The general trend of high success rates to certain payload ranges and specific launch sites.

Section 5

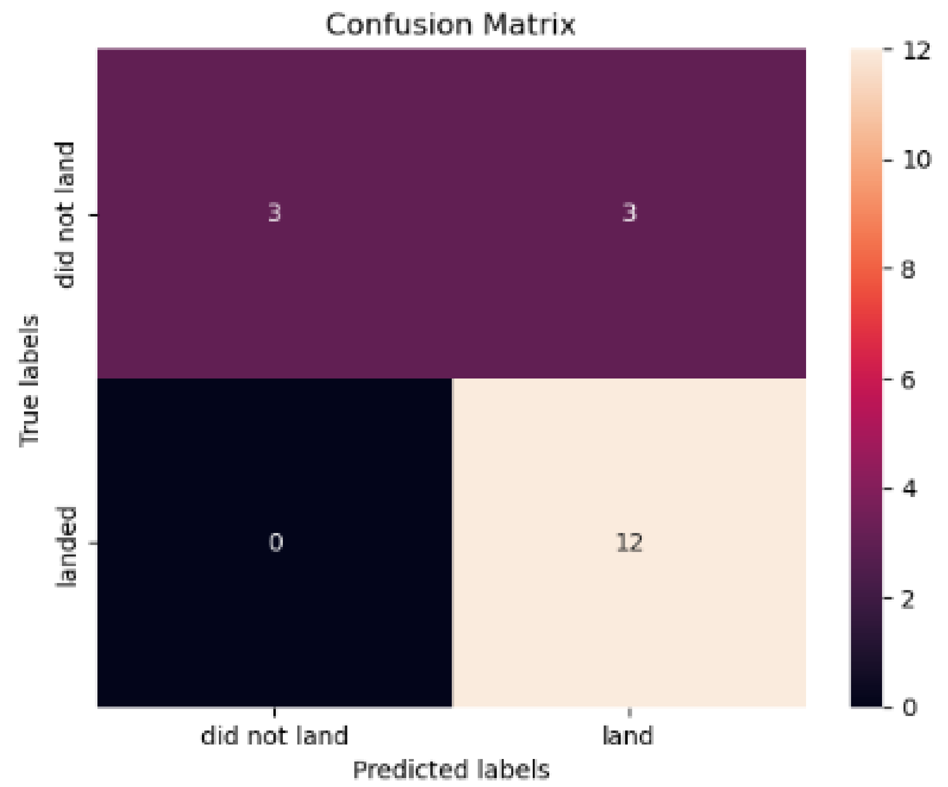
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

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



Top Left shows True Negatives, the model predicted 3 instances where the rocket did not land.

Top Right shows False Positives, the model incorrectly predicted 3 instances where the rocket would land but it actually crashed.

Bottom Right shows True Positives, the model correctly predicted 12 instances where the rocket would land and it did.

Bottom Left Cell shows False Negatives, the model incorrectly predicted 0 instances of the rocket not landing but the rocket actually landed.

Conclusions

- Reusing rocket components is viable.
- Orbit, Launch site, payload contribute to successful outcomes.
- Launch Cost can be reliably estimated.
- Price of launch is more competitive when components are reused.

Appendix

Space Age is here. Opportunity for Competitors.

Reusing Rocket Components reduces Cost

Thank you!