IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohsin Ali Shah
12 December, 2025

Outline

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

Executive Summary

- Summary of methodologies
- Collected SpaceX launch data using API calls and web scraping
- Cleaned, wrangled, and merged datasets for consistency
- Performed Exploratory Data Analysis (EDA) using visualizations and SQL
- Built interactive visual analytics using Folium maps and Plotly Dash
- Developed and evaluated multiple classification models to predict launch success
- Summary of all results
- Identified major factors influencing SpaceX launch outcomes
- Interactive maps revealed launch site performance and geographical patterns
- Dashboard showed strong correlations between payload, orbit, and success rate
- Best classification model achieved strong predictive accuracy using optimized hyperparameters

Introduction

- Project background and context
  - ➢ SpaceX operates frequent Falcon 9 launches with publicly accessible data
  - ➢ Understanding launch success factors improves mission planning and risk analysis

- Problems you want to find answers
  - ➢ What features most influence launch success?
  - ➢ How can geospatial and historical trends reveal opportunities for optimization?
  - ➢ Can machine learning models accurately predict launch outcomes?

Section 1

Methodology

Methodology

Executive Summary

- Collected data via SpaceX REST API and web scraping
- Cleaned and transformed data through structured wrangling steps
- Conducted EDA using charts, statistics, and SQL queries
- Built interactive visual analytics with Folium and Plotly Dash
- Built, tuned, and evaluated classification models
- Selected the best model based on accuracy, recall, and F1-score

Data Collection

- Describe how data sets were collected.

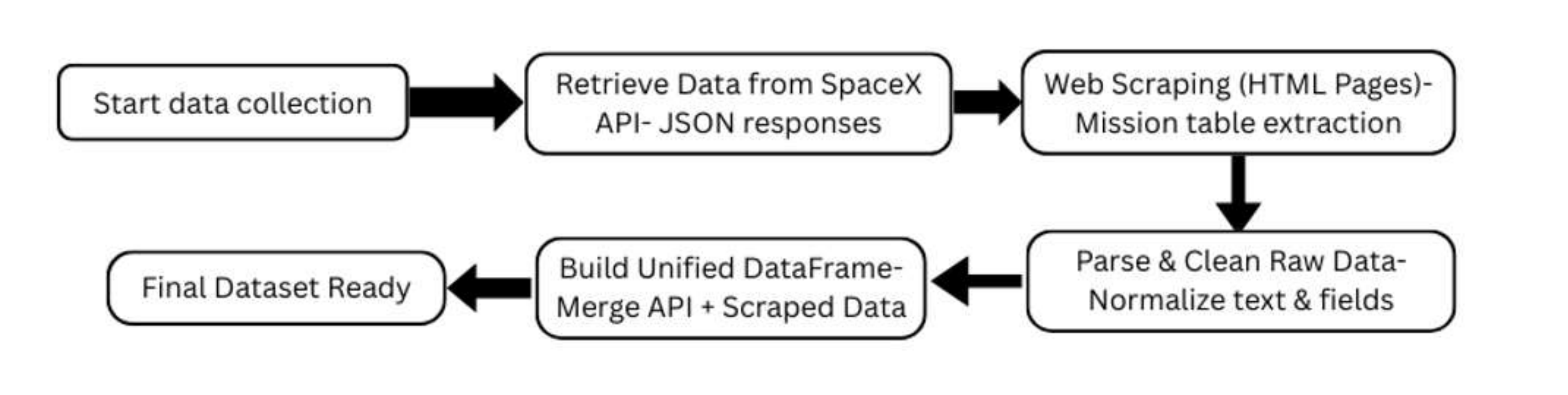
➢ Retrieved SpaceX launch records from official endpoints

➢ Parsed and structured datasets from web scraping

➢ Stored raw data for validation and replication

➢ Built a complete data pipeline from API → scrape → storage

- You need to present your data collection process use key phrases and flowcharts



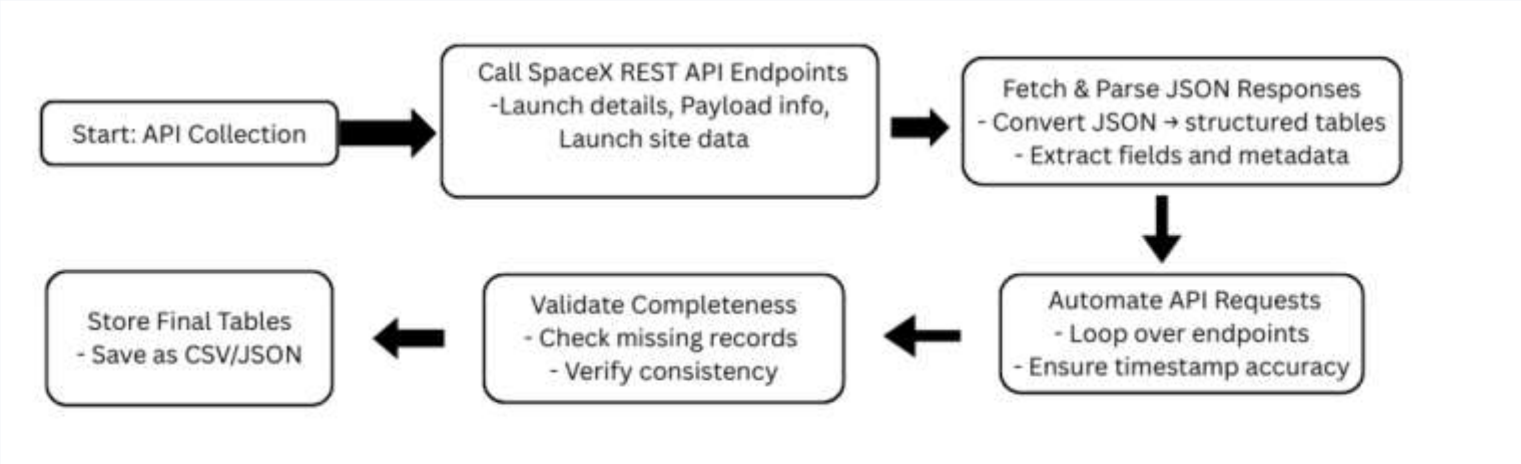
Data Collection – SpaceX API

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

- Used REST API endpoints to fetch launch details, payload data, and launch site info
- Automated requests and parsed JSON responses into structured tables
- Ensured data completeness and timestamp accuracy

- GitHub URL of the completed SpaceX API calls notebook

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb)

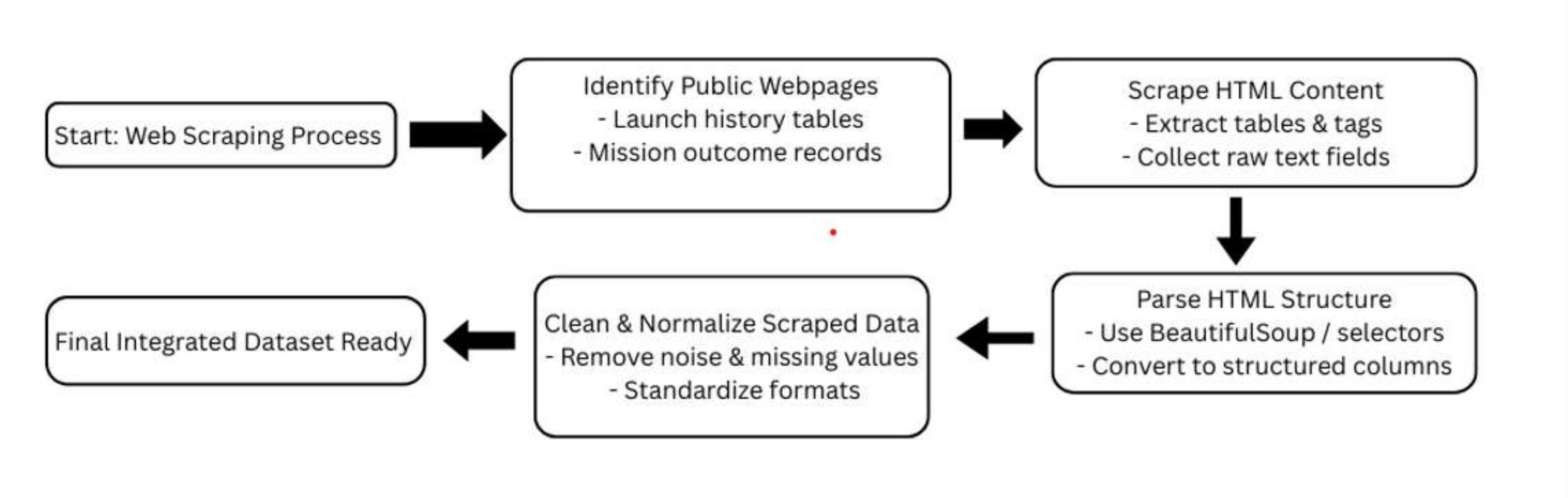


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- Scraped launch tables and historical data from public webpages
- Used HTML parsing to extract structured fields
- Cleaned scraped data and merged with API dataset

- GitHub URL of the completed web scraping notebook

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-webscraping.ipynb)



Data Wrangling

- Describe how data were processed
- Handled missing values, type conversions, and inconsistent formats
- Normalized categorical values and engineered new features (e.g., success label, launch site codes)
- Merged datasets using unique keys
- Validated data integrity before analysis

- GitHub URL of your completed data wrangling related notebooks

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb)

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Bar charts: launch success counts across sites
  - Scatter plots: payload vs. success relationships
  - Line charts: trends across launch years
  - Heatmaps: correlation between numerical features
- Purpose
  - Identify patterns, correlations, and performance trends
- GitHub URL of your completed EDA with data visualization notebook

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/edadataviz.ipynb)

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Filtered launch records by year and success
  - Aggregated success rates by launch site
  - Extracted payload and orbit statistics
  - Performed joins between related tables
- GitHub URL of your completed EDA with SQL notebook

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb)

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Launch site markers
  - Circle overlays showing launch radius
  - Pop-ups showing launch site names and success metrics
  - Lines indicating trajectory / proximity
- Explain why you added those objects
  - Highlight launch site distribution
  - Show spatial relationships affecting performance
- GitHub URL of your completed interactive map with Folium map

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb)

Build a Dashboard with Plotly Dash

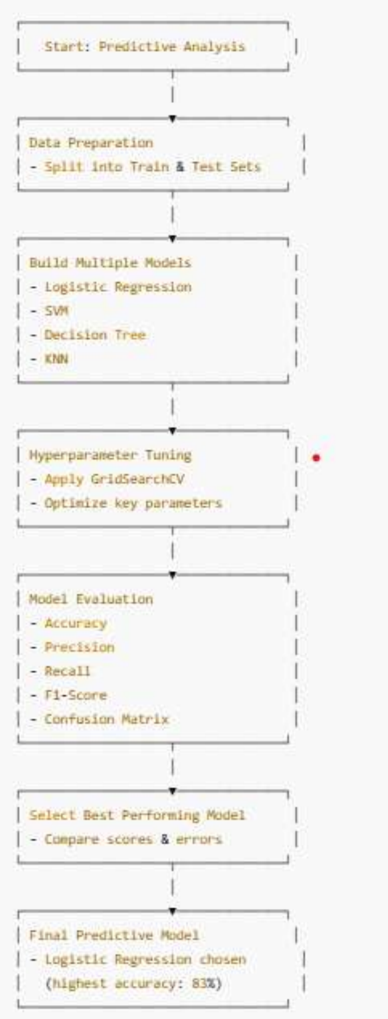
- Summarize what plots/graphs and interactions you have added to a dashboard
  - Interactive dropdowns for orbit and launch site
  - Success rate bar charts
  - Payload vs. success scatter plots
  - Dynamic filtering for real-time analysis
- Explain why you added those plots and interactions
  - Enable users to explore correlations and trends
  - Provide a decision-support interface through interactive visual analytics
- GitHub URL of your completed Plotly Dash lab

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex-dash-app.py)

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
    - Split dataset into training and testing sets
    - Built multiple models (Logistic Regression, SVM, Decision Tree, KNN)
    - Tuned hyperparameters using GridSearchCV
    - Evaluated models using accuracy, precision, recall, and F1-score
    - Selected best model based on metrics and confusion matrix
- You need present your model development process using key phrases and flowchart
- GitHub URL of your completed predictive analysis lab

(https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb)



Results (EDA)

- Exploratory data analysis results

- Launch success rates increased significantly over time, especially after 2015.
- KSC LC-39A showed the highest number of successful Falcon 9 landings.
- Payload mass showed weak correlation with mission outcome compared to booster version and flight experience.
- LEO, SSO, and Polar orbits demonstrated consistently higher success rates.

Results (Interactive Analytics)

- Interactive analytics demo in screenshots
- **Interactive Analytics Findings (Folium & Plotly Dash)**
- Folium maps revealed site proximity patterns: launch sites are close to coastlines and away from populated areas.
- Color-coded outcome markers highlighted which sites achieved the highest success ratios.
- Dashboard interactions confirmed relationships between payload ranges, orbit type, and landing success.
- Payload slider analysis showed heavier payloads (>5000 kg) have noticeably lower success probability.

Results (Predictive analysis)

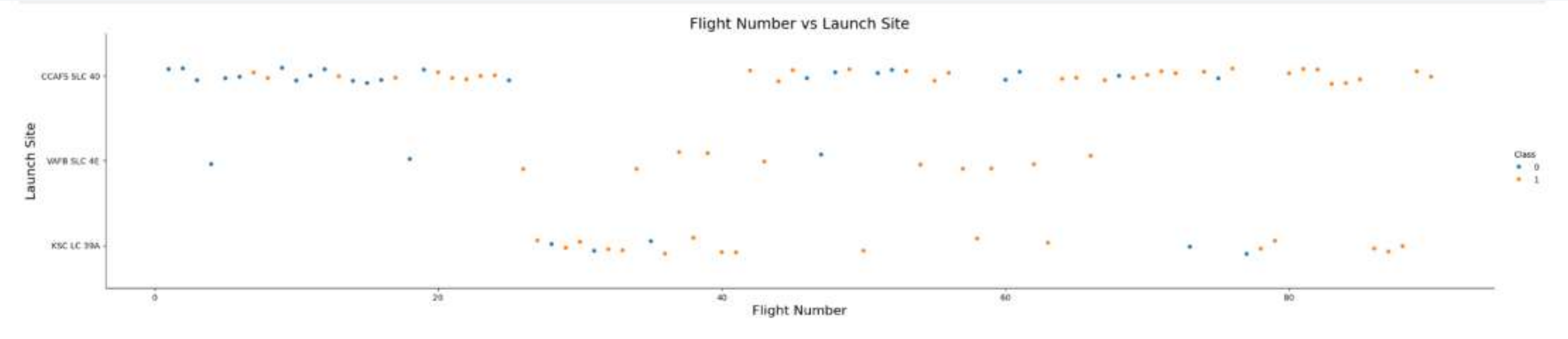
- **Predictive analysis results**
- Logistic Regression achieved the **highest accuracy of 83%** after hyperparameter tuning.
- Confusion matrix showed strong true-positive performance and minimal false positives.
- ML models effectively predicted landing success using features such as flight number, orbit, payload, and booster version.
- Overall, predictive modeling validated key drivers of Falcon 9 first-stage landing success.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

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

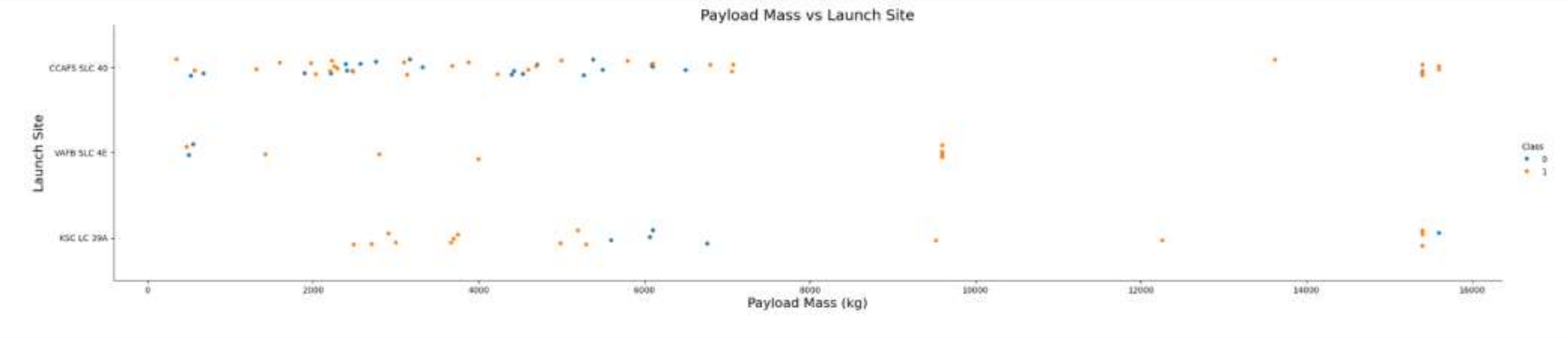


- Show the screenshot of the scatter plot with explanations

We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass also appears to be a factor; even with more massive payloads, the first stage often returns successfully.

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



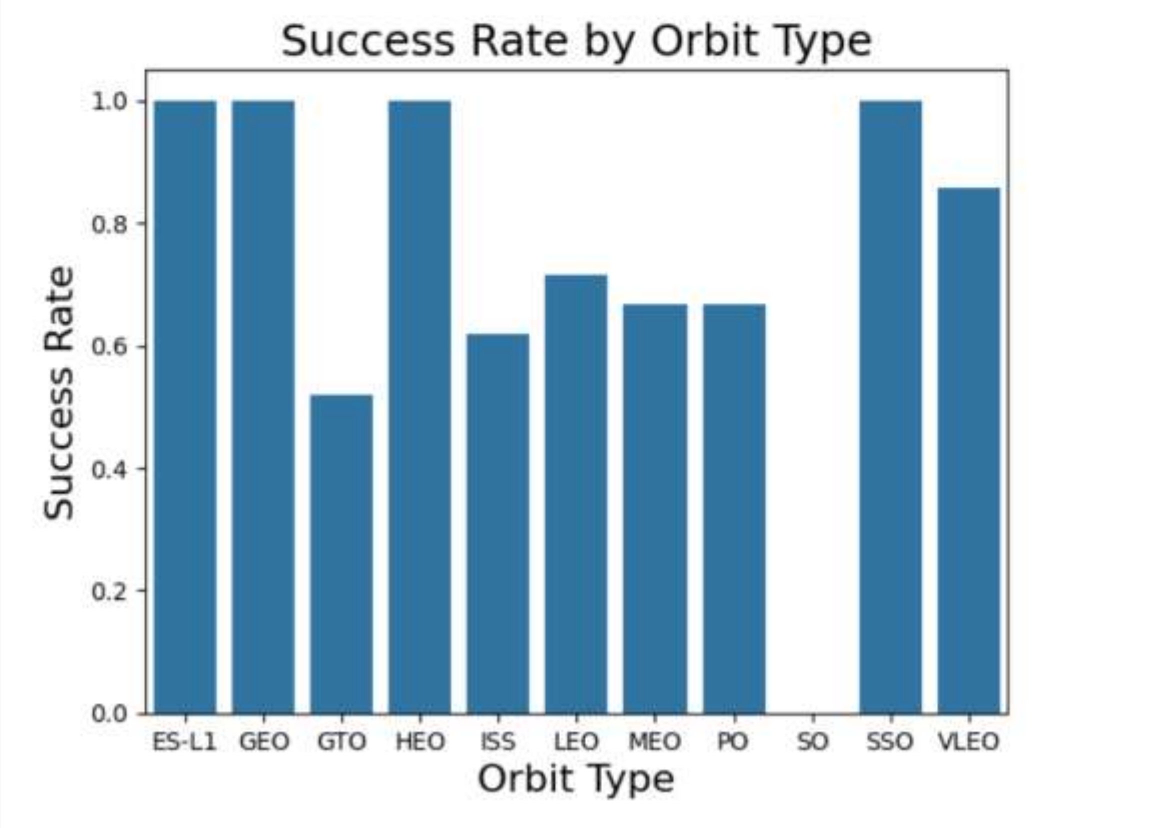
- Show the screenshot of the scatter plot with explanations
    - Payload mass does not strongly correlate with mission outcome.
    - All sites support a wide range of payload capabilities.
    - KSC LC-39A handles the heaviest payloads, confirming its use for big missions.
    - Success rates appear to be influenced more by other factors (booster version, mission complexity) than payload alone

Success Rate vs. Orbit Type

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

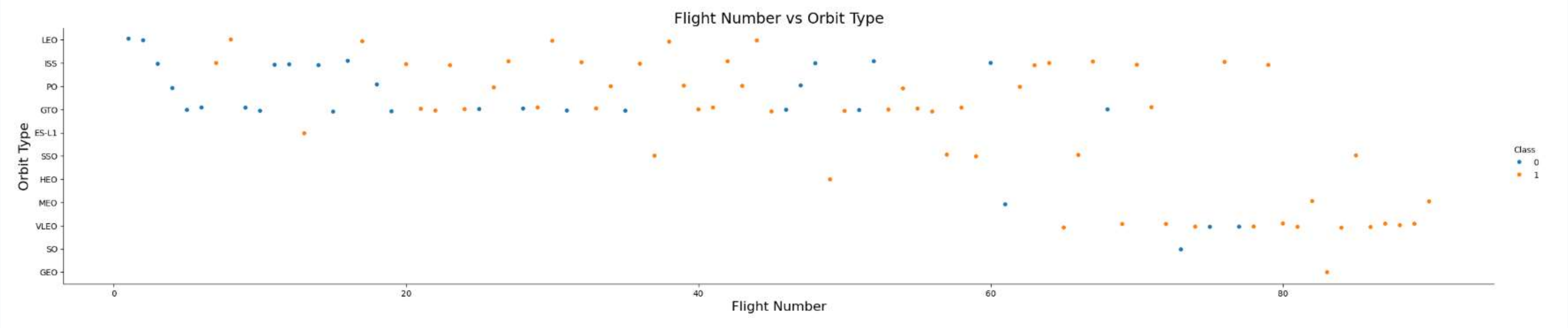
- Show the screenshot of the scatter plot with explanations

✓ From the bar chart we can see that Orbit ES-L1, GEO, HEO and SSO have the highest success rate compare to other orbits



Flight Number vs. Orbit Type

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

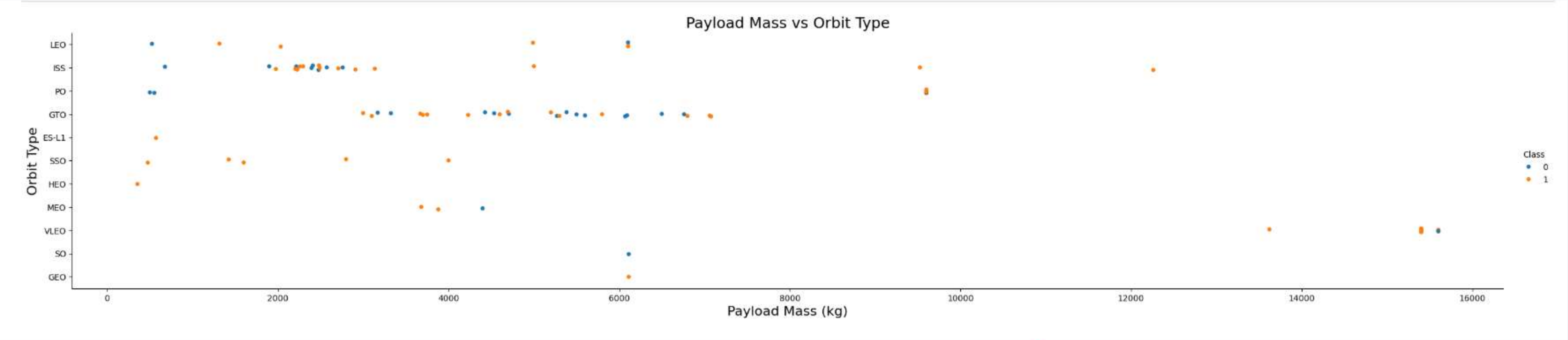


- Show the screenshot of the scatter plot with explanations

We can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

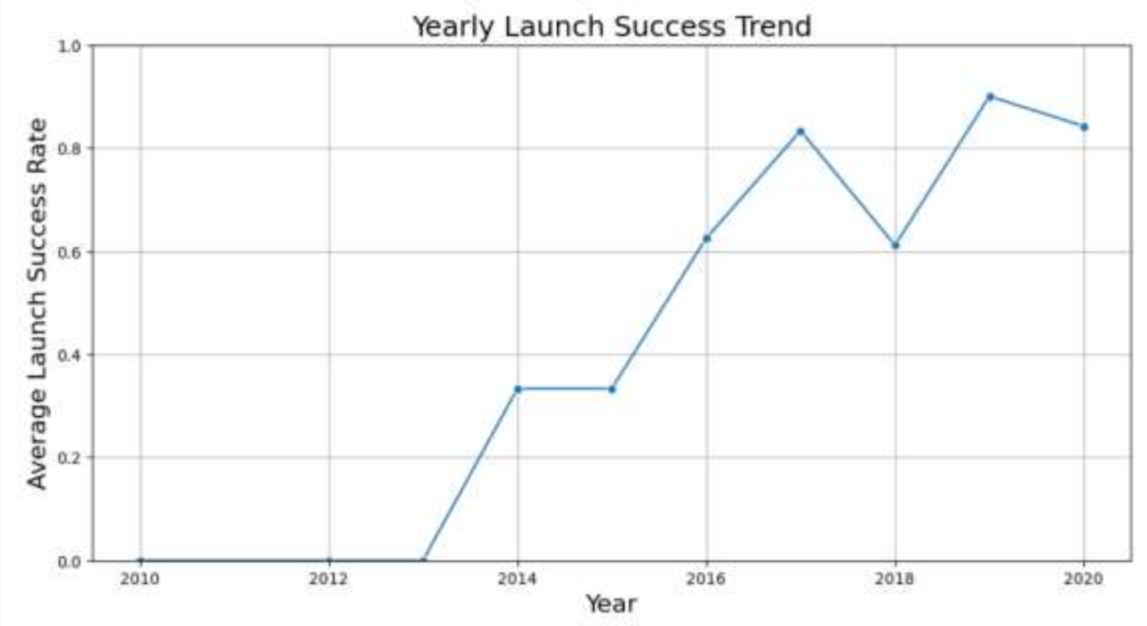


- Show the screenshot of the scatter plot with explanations
  - With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
  - However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations
  - We can observe that the success rate since 2013 kept increasing till 2020
  - There is a dip in 2018 but overall success rate kept increasing during this period



All Launch Site Names

- Find the names of the unique launch sites

Display the names of the unique launch sites in the space mission

In [17]:
```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

Out[17]:

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

- Present your query result with a short explanation here

By applying the above query using sql we successfully got the unique launch sites from the SpaceX Data.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

Display 5 records where launch sites begin with the string 'CCA'

In [20]:
```
%%sql SELECT *
    FROM SPACEXTABLE
   WHERE Launch_Site LIKE 'CCA%'
   LIMIT 5;
```

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

Out[20]:

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

- Present your query result with a short explanation here

By running the above query we were able to get the 5 records that begin with 'CCA'

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Display the total payload mass carried by boosters launched by NASA (CRS)

In [21]:

```
%%sql SELECT SUM(Payload_Mass__kg_) AS Total_Payload
FROM SPACEXTABLE
WHERE Customer = 'NASA (CRS)';
```

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

Out[21]:

Total_Payload
45596

- Present your query result with a short explanation here

By running the above query using SQL we can see that total payload carried by boosters from NASA is 45596 kg.

Average Payload Mass by F9 v1.1

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

Display average payload mass carried by booster version F9 v1.1

In [22]:

```
%%sql SELECT AVG(Payload_Mass__kg_) AS Avg_Payload
      FROM SPACEXTABLE
      WHERE Booster_Version = 'F9 v1.1';
```

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

Out[22]:

Avg_Payload
2928.4

- Present your query result with a short explanation here

By running the above query using SQL we can see that total Average payload mass carried by boosters version F9 v1.1 is 2928.4 kg.

First Successful Ground Landing Date

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

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

In [23]:
```
%%sql SELECT MIN(Date) AS First_Success_Ground_Pad
      FROM SPACEXTABLE
      WHERE Landing_Outcome = 'Success (ground pad)';
```

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

Out[23]:

First_Success_Ground_Pad
2015-12-22

- Present your query result with a short explanation here

By running the SQL query we got the date of first successful landing outcome on ground pad that is 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

```
In [24]: %%sql SELECT Booster_Version
             FROM SPACEXTABLE
             WHERE Landing_Outcome = 'Success (drone ship)'
               AND Payload_Mass__kg_ > 4000
               AND Payload_Mass__kg_ < 6000;

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

Out[24]:

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

- Present your query result with a short explanation here
- From above results we can see that Booster Version F9 FT B1022, F9 FT B1026, F9 FT B1021.2 and F9 FT B1031.2 have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

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

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

- Present your query result with a short explanation here

By running the above SQL query we can see failure is 1 where as success is 99 and 1 success has unclear payload status

Boosters Carried Maximum Payload

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

```
%%sql SELECT Booster_Version
      FROM SPACEXTABLE
      WHERE Payload_Mass__kg_ = (
          SELECT MAX(Payload_Mass__kg_)
          FROM SPACEXTABLE
      );
```

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

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

- Present your query result with a short explanation here

We ran the above query on SQL and on the right side we can see the list of the booster which have carried the maximum payload mass.

2015 Launch Records

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

```
%%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, 1, 4) = '2015';
```

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

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

- Present your query result with a short explanation here

We ran an SQL query and we can see 2 landing failures in drone ship in year 2015 with their site names, month and Booster Version

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

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

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

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

- Present your query result with a short explanation here

From running the above SQL query we found 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

Section 3

Launch Sites Proximities Analysis

Folium: all launch sites' location markers on a Global Map

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

In the above map created by Folium we can see the marker location of all launch sites on a global map. They appear stacked but when zoomed in we can see the exact location of the launch sites

Folium: Color-labeled Launch outcomes on the Map

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

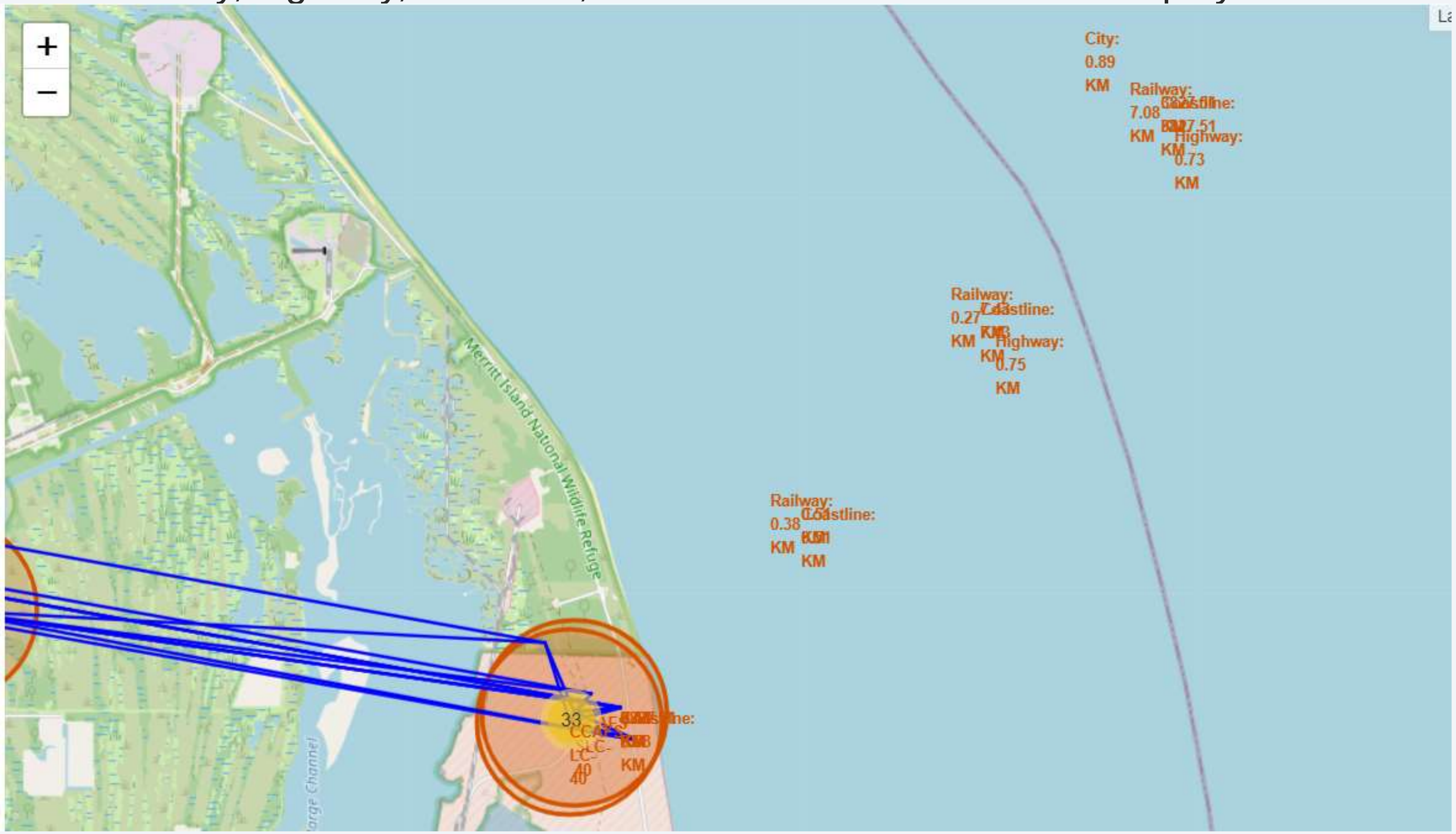
Folium: Color-labeled Launch outcomes on the Map (Continued)

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

- From the color-labeled markers in marker clusters, can easily identify which launch sites have relatively high success rates.
- The map on the previous slide gave zoomed out overview, whereas in this slide we can see that colored label sites on map give us clear view.

Folium: Launch sites’ & their Distance form Proximities

- 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



In the map we can see that distance from each launch site is displayed from its proximities such as railway, coastline and highway. By these useful findings we can actually tell that

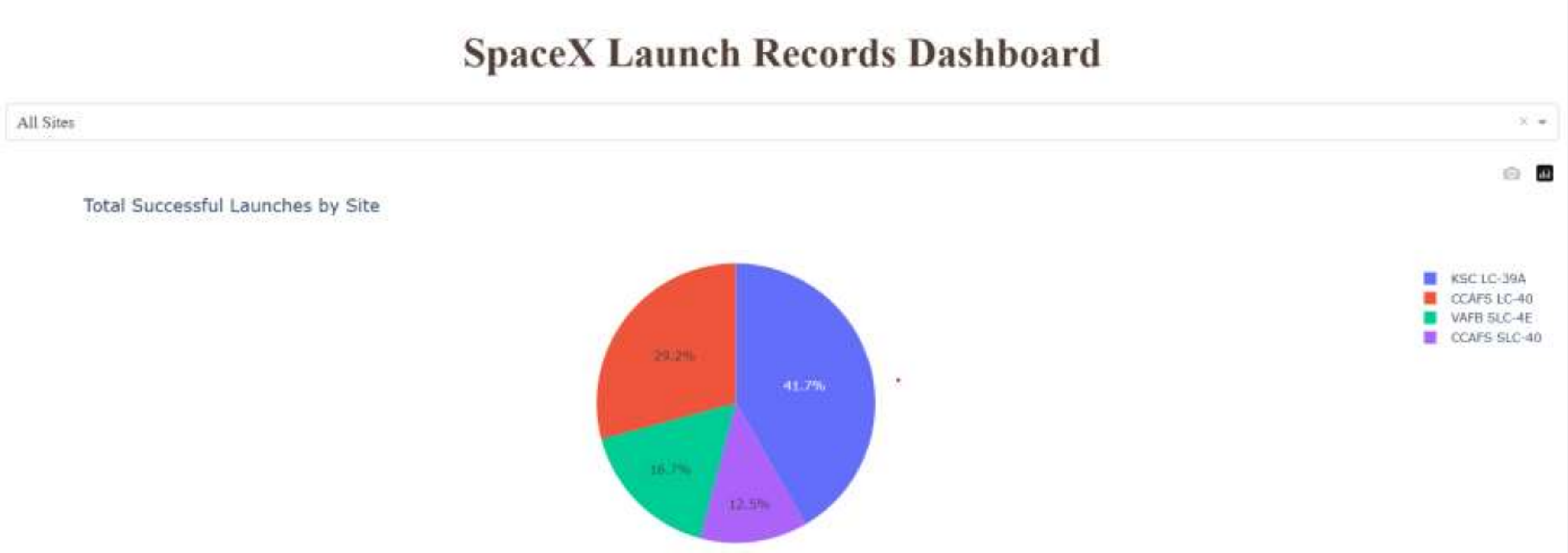
- Coastlines are closer to the sites
- They maintain certain distance from the city
- Railway are not far from launch sites
- Highways are at a certain distance from the sites.

Section 4

Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard pie-chart (all sites)

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

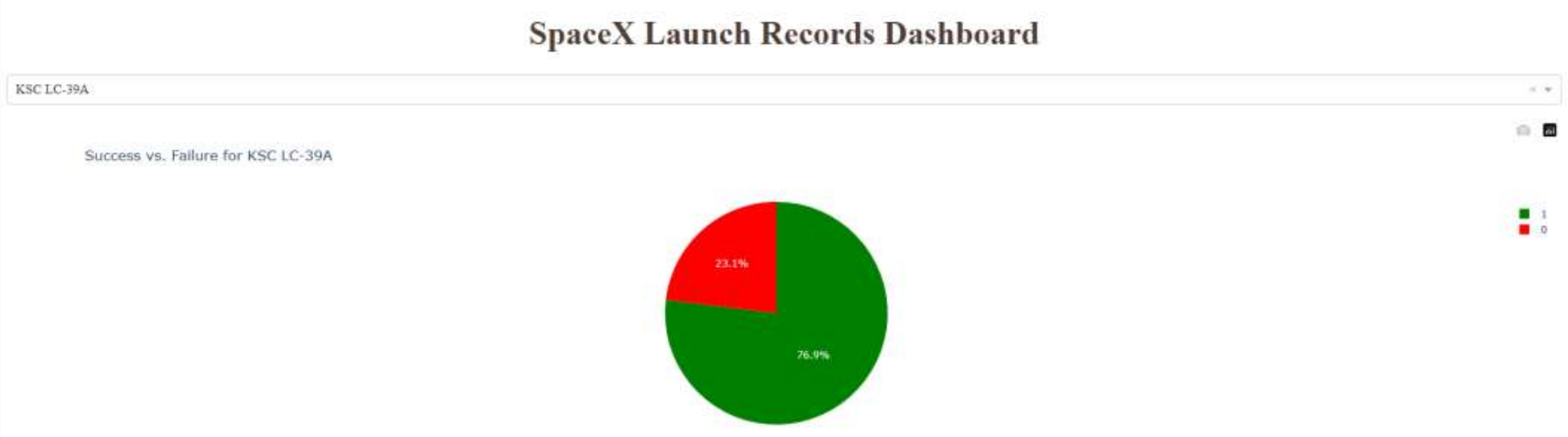


- Explain the important elements and findings on the screenshot

The pie chart above shows the count success for all sites, it has labels named at the right corner of the screenshot, showing what color represents which launch site.

Dashboard: piechart for the launch site with highest launch success ratio

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



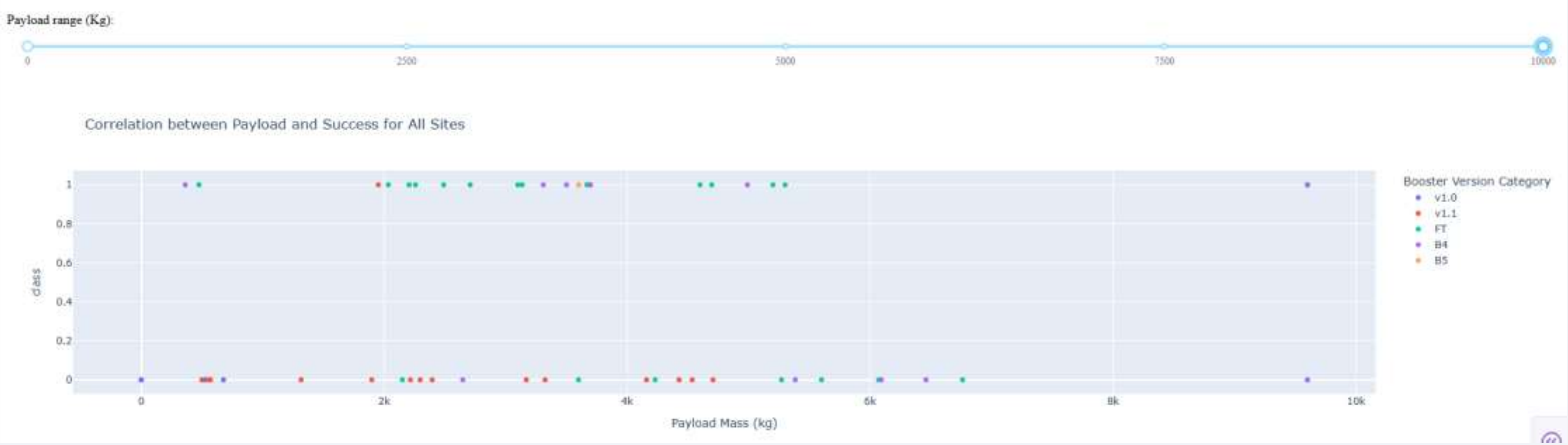
- Explain the important elements and findings on the screenshot

From the above pie chart we can see that overall KSC LC-39A has highest success count and when in this pie chart we further explored we can see that total 76.9% success rate and 23.1% is failure rate that makes it the site with highest success ratio.

Dashboard Payload vs. Launch Outcome Scatter Plot

- Payload vs. Launch Outcome scatter plot for all sites, with different payload

Payload Range (0-10k) kg

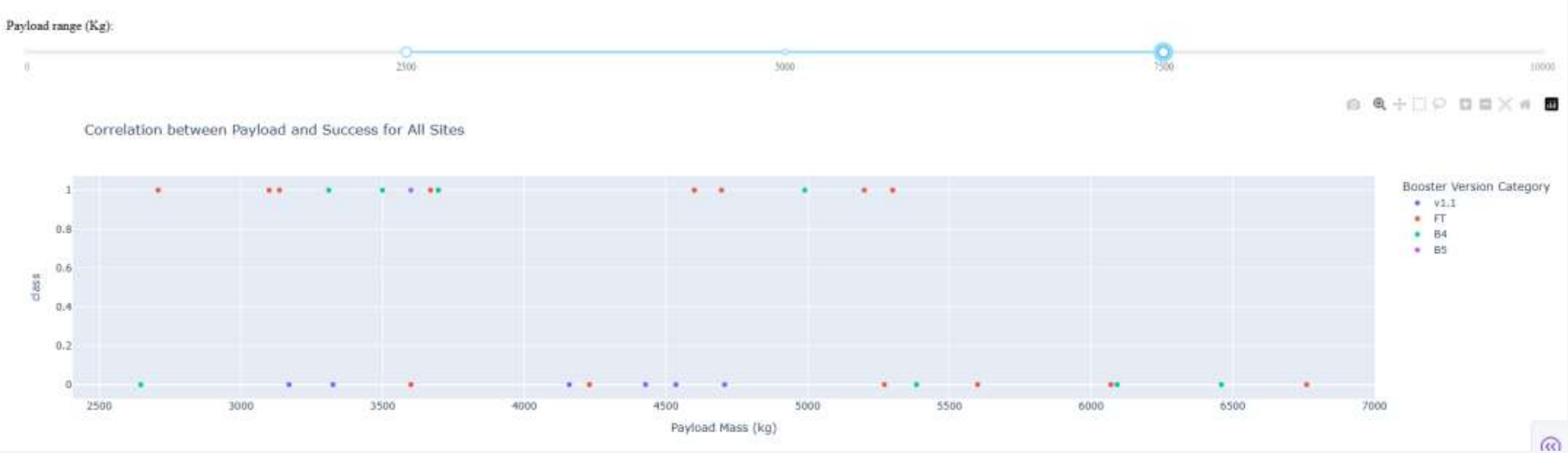


This scatter plot shows correlation between payload and success, on the right side in the indicator that shows which color shows which Booster version. Class 1 means its success and class 0 indicates failure

Dashboard Payload vs. Launch Outcome Scatter Plot

- Payload vs. Launch Outcome scatter plot for all sites, with different payload

Payload Range (2500-7500) kg

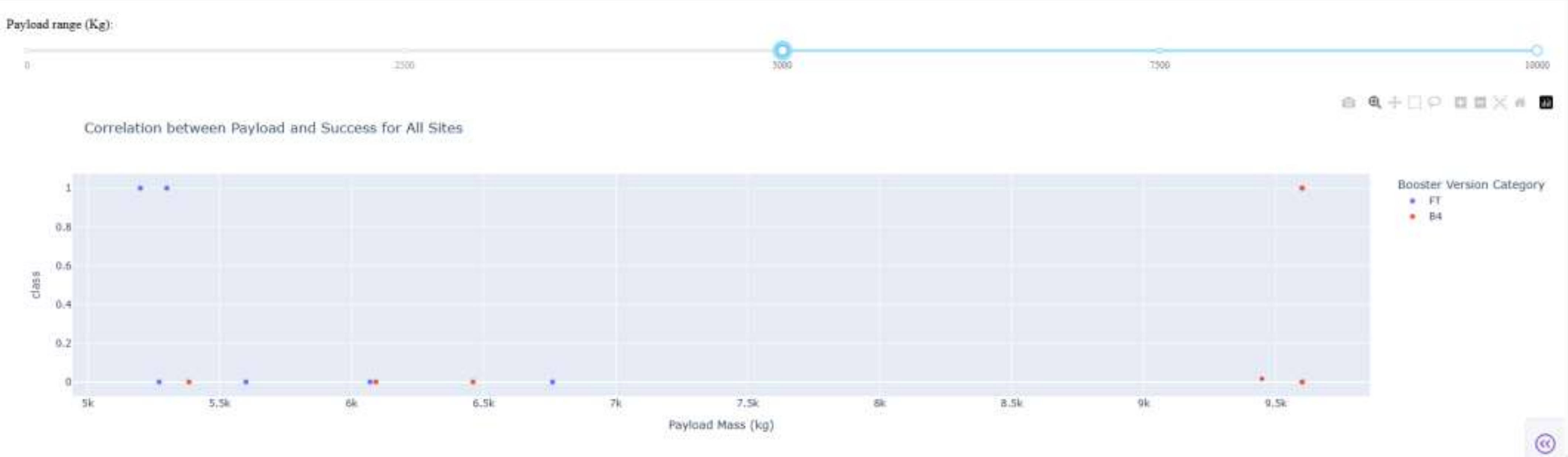


In this scatter plot we can see that between range 2.5k to 7.5k there is high success rate and the FT has the most success ratio among all other boosters.

Dashboard Payload vs. Launch Outcome Scatter Plot

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

Payload Range (5000-10k) kg



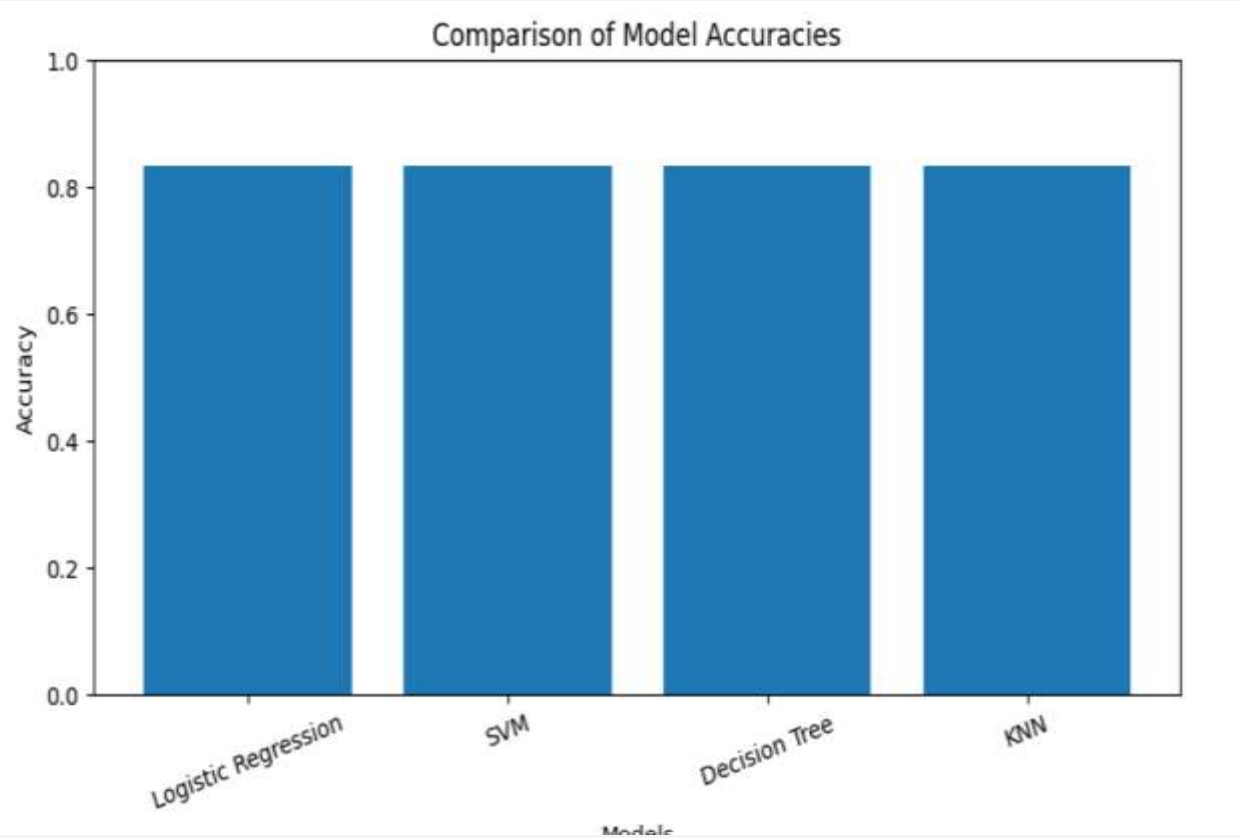
- In this scatter plot we can see that there is very low success rate in the range from 5000kg to 10k kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy
- From this we can see that Logistic Regression provides the best accuracy that is 83%.



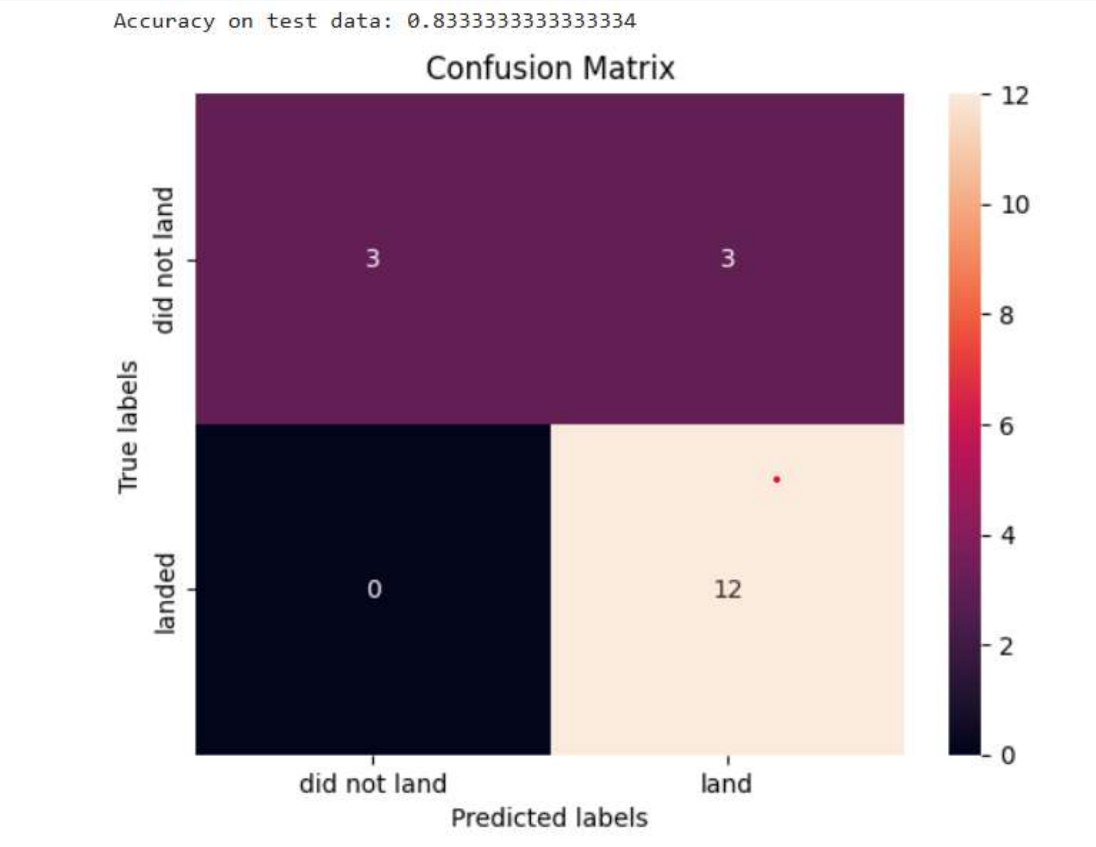
Confusion Matrix

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

So from this logistic regression model we can see that It has following results:

True positives are 12
False positive are 3
And overall accuracy is 83%

This is the best performing model compared to other models that show a high False positive rate.



Conclusions

- This project successfully analyzed SpaceX Falcon 9 launch and landing performance using data collected from API calls and web scraping.
- EDA revealed key factors influencing mission success, including booster experience, orbit type, and launch site performance.
- Interactive visual analytics using Folium and Plotly Dash provided clear insights into geographic patterns, payload effects, and overall success trends.
- Multiple machine learning models were built and evaluated; Logistic Regression achieved the best performance with an accuracy of 83%.
- Overall, the findings demonstrate that data-driven methods can effectively predict Falcon 9 landing outcomes and support improved mission planning and decision-making.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

1. https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/edadataviz.ipynb
2. https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb
3. https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb
4. https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
5. https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
6. https://github.com/Mohsin-ali-shah/SpaceX-Falcon-9-first-stage-Landing-Prediction

Thank you!