SPACEX FALCON 9 **LANDING PREDICITON**

Rolando John R. Aca-ac
April 27, 2024

Table of Contents

Outline

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

Executive Summary

Methodologies

- Utilized a combination of API requests from SpaceX and web scraping from Wikipedia to gather comprehensive launch data.
- Conducted extensive data cleaning, handling missing values, and restructuring the dataset for analysis.
- Employed visualization techniques to uncover insights into launch success trends, payload outcomes, and site characteristics.
- Utilized Folium and Plotly Dash to create interactive visualizations, providing dynamic exploration of launch site locations, success rates, and payload outcomes.
- Trained classification models to predict Falcon 9 landing outcomes, evaluating model performance and selecting the best-performing algorithm.

Executive Summary

Results

- Identified SpaceX's competitive advantage in cost reduction through Falcon 9 first stage reuse.
- Found that launch sites are strategically located near coastlines and away from cities for safety, with close proximity to railways and highways facilitating transportation.
- Folium and Plotly Dash visualizations provided insights into launch site locations, success rates, and payload outcomes, aiding in decision-making processes.
- The Decision Tree Classifier emerged as the best-performing model with 88.90% accuracy, offering valuable insights for stakeholders in launch planning and operations..

Introduction

Results

- Identified SpaceX's competitive advantage in cost reduction through Falcon 9 first stage reuse.
- Found that launch sites are strategically located near coastlines and away from cities for safety, with close proximity to railways and highways facilitating transportation.
- Folium and Plotly Dash visualizations provided insights into launch site locations, success rates, and payload outcomes, aiding in decision-making processes.
- The Decision Tree Classifier emerged as the best-performing model with 88.90% accuracy, offering valuable insights for stakeholders in launch planning and operations..

SECTION 1

METHODOLOGY

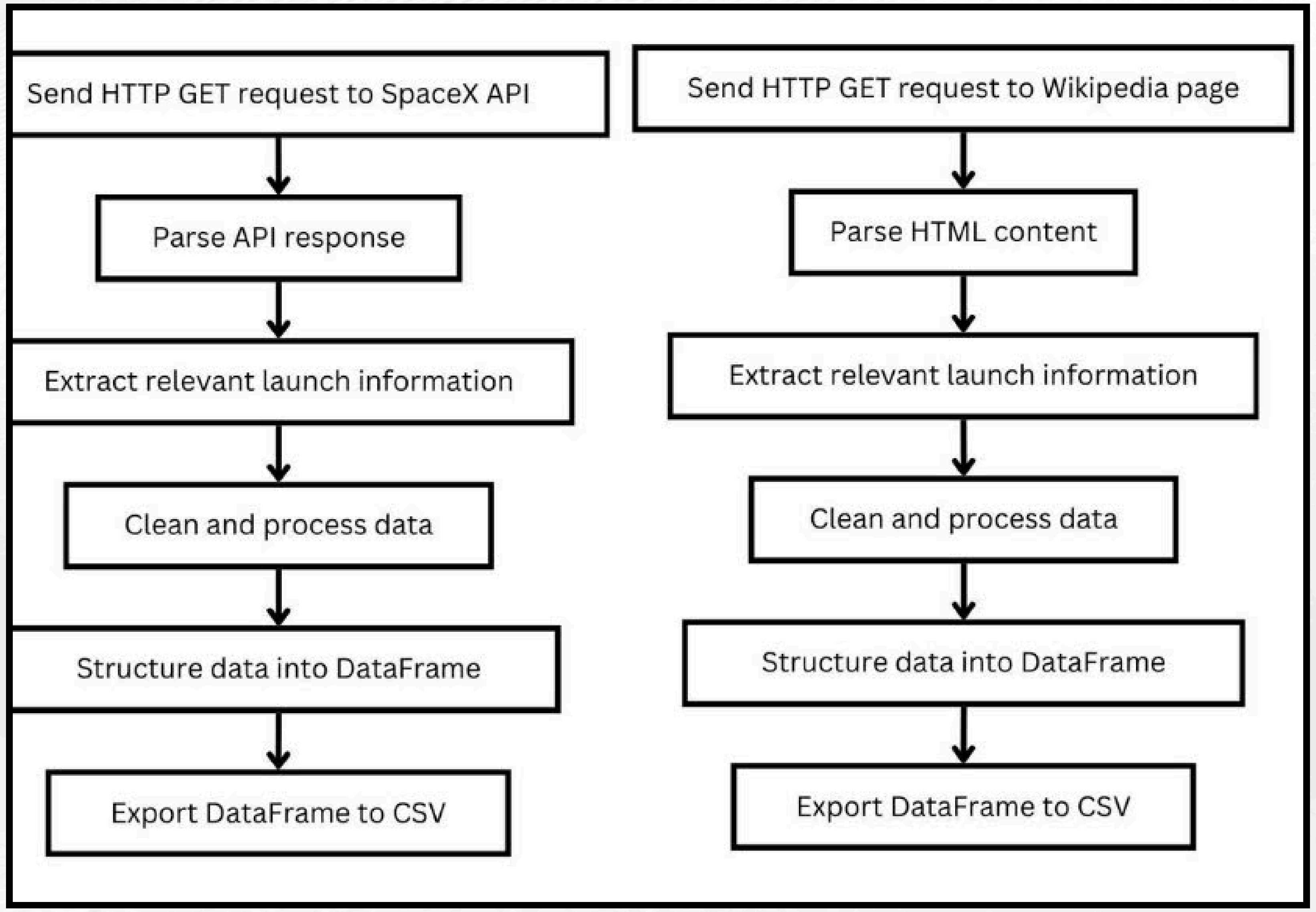
Methodology

Executive Summary

- Data collection methodology.
  - Data collection is performed in two ways: Requesting data from the SpaceX API and web scraping launch data from a Wikipedia page.
- Perform data wrangling.
  - Missing values are handled, data types are checked and exploratory data analysis techniques are applied to prepare the dataset for analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.
- Perform predictive analysis using classification models.
  - Classification model development includes selecting the right algorithm, refining its parameters, training it, and evaluating its accuracy.

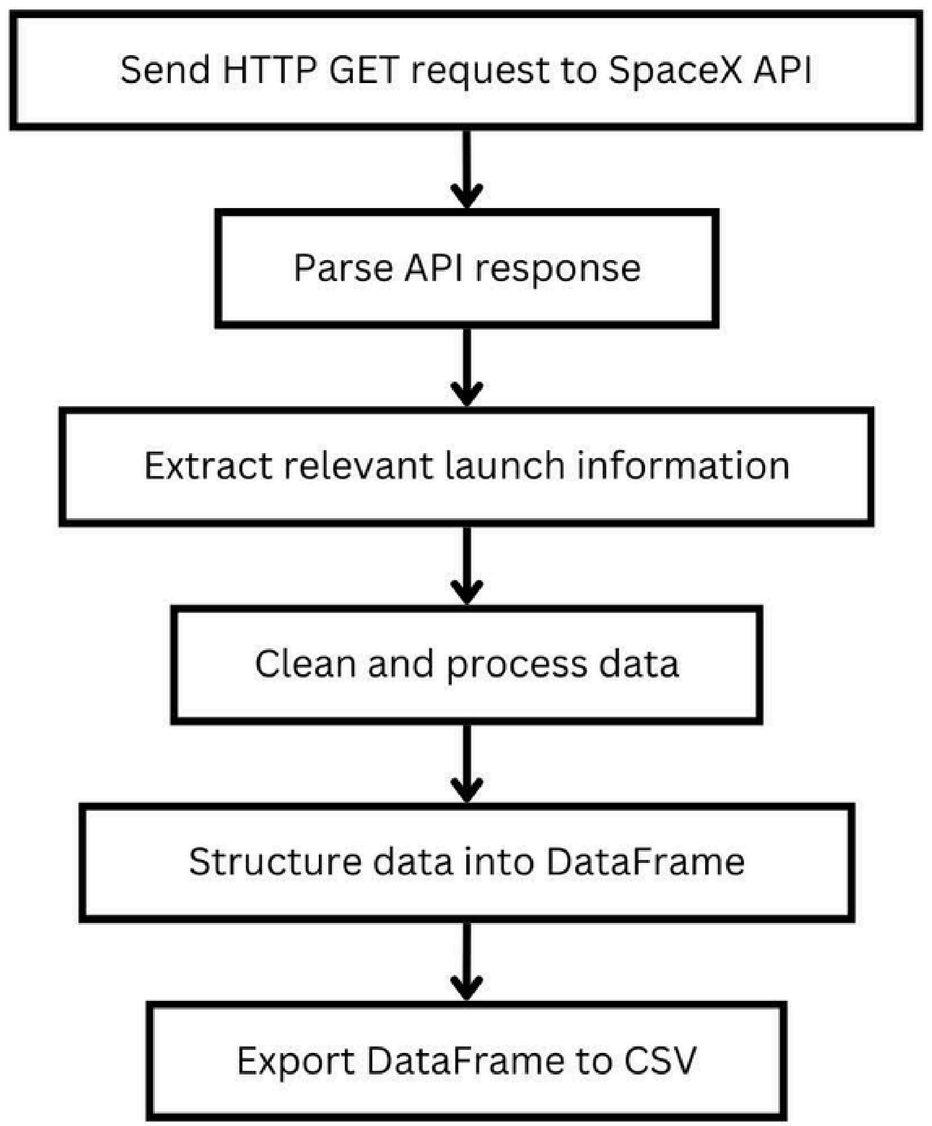
Data Collection

- Data collection involves two main methods: API and web scraping.
- API requests fetch detailed launch information directly from SpaceX's database.
- Web scraping extracts launch data from a Wikipedia page listing Falcon 9 and Falcon Heavy launches.
- For API data, responses are parsed, cleaned, and structured into a DataFrame.
- Web scraping involves making HTTP requests to the Wikipedia page, parsing HTML content, and extracting relevant launch details.
- Extracted data is organized, filtered, and exported for analysis.



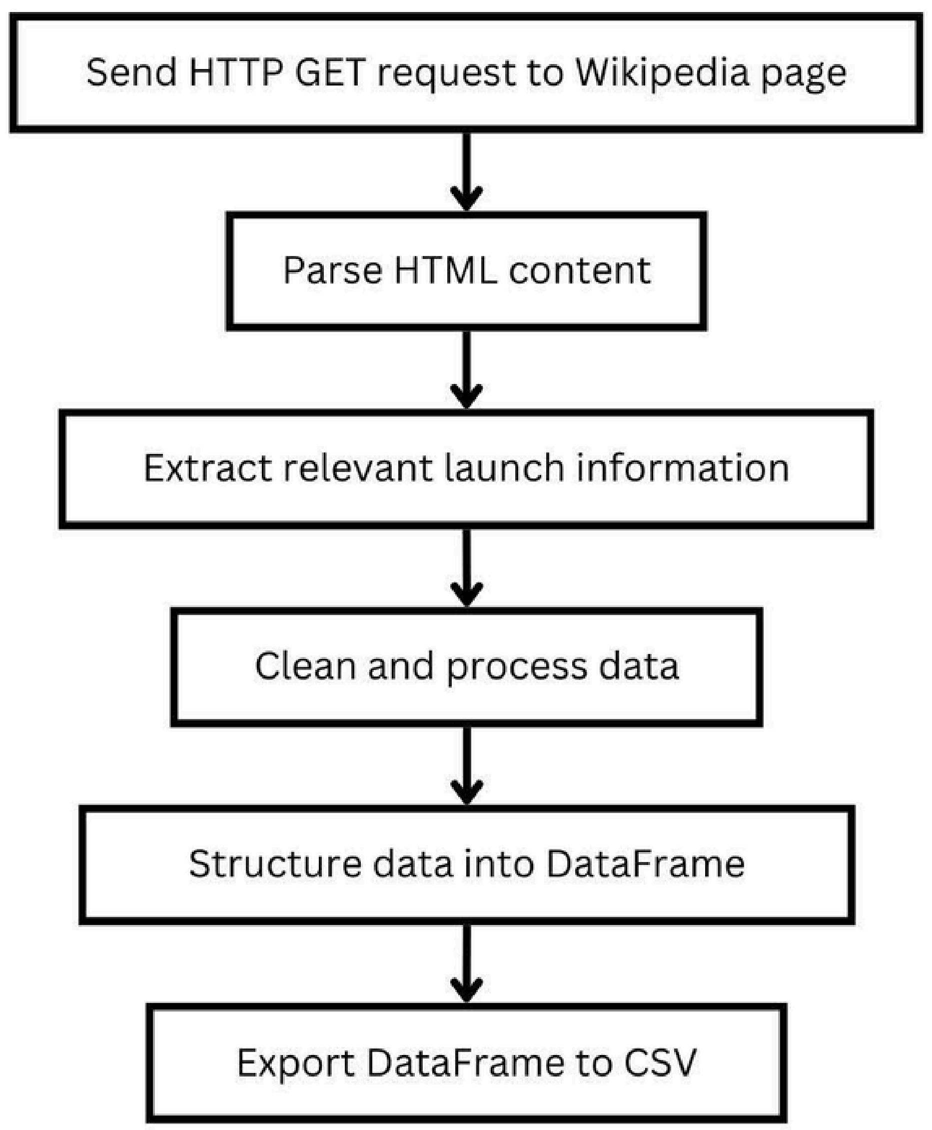
Data Collection – SpaceX API

- Data Collection using SpaceX API is given in flow chart for an overview.
- Complete Notebook link given below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/1_Data_Collection_API.ipynb



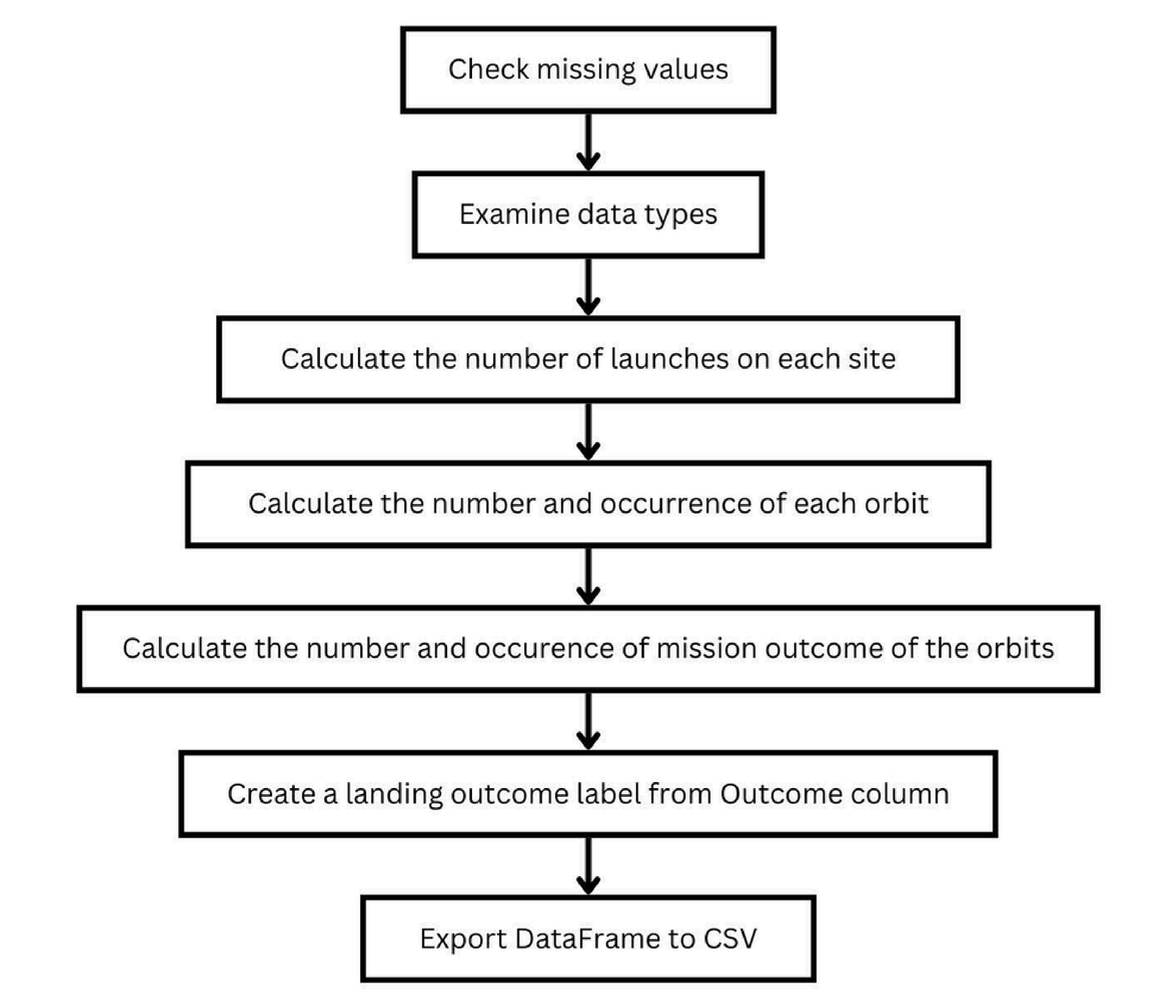
Data Collection – Scraping

- Data Collection through Web Scraping is given in flow chart for an overview.
- Complete Notebook link given below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/2_Data_Collection_Web_Scraped.ipynb



Data Wrangling

- Data Wrangling process is given in flow chart for an overview.
- Complete Notebook link given below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/3_Data_Wrangling.ipynb



EDA with Data Visualization

- Exploratory Data Analysis is performed by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly trend.
- Charts used:
  - Scatter plots, Bar chart, Line plot
- Complete Notebook link given below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/5_Exploratory_Data_Analysis_with_Visualization.ipynb

EDA with SQL

- Selected and displayed the first 5 rows of the "SPACEXTABLE".
- Retrieved distinct launch sites from the "SPACEXTABLE".
- Filtered and displayed the first 5 rows where the launch site starts with 'CCA'.
- Calculated the total payload mass for NASA (CRS) customers.
- Calculated the average payload mass for booster version 'F9 v1.1'.
- Determined the date of the first successful landing in a ground pad.
- Selected booster versions for successful drone ship landings with payload mass between 4000 and 6000 kg.
- Counted the total missions by mission outcome categories.
- Found the booster version with the highest payload mass.
- Extracted month names, landing outcomes, booster versions, and launch sites for missions in 2015 with failed drone ship landings.
- Counted the number of missions by landing outcomes within a specified date range.
- Complete Notebook link given below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/4_Exploratory_Data_Analysis_Using_SQL.ipynb

Interactive Map with Folium

- Markers represent SpaceX launch sites and individual launches, indicating locations and outcomes.
- Circles delineate launch site areas, enhancing visibility and differentiation.
- Marker clusters group nearby launches for improved map readability.
- Lines connect launch sites to coastal areas, railways, highways, and cities, displaying distances.
- Custom icons and popups provide additional information on launch sites and outcomes.
- Complete Notebook link given below.
    - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/6_Interactive_Visual_Analytics_with_Folium.ipynb

Web Dashboard with Plotly Dash

- Pie Chart
  - Shows total successful launches by site when "All Sites" is selected in the dropdown.
  - Shows successful and failed launches for a specific site when a site is selected in the dropdown.
- Scatter Plot
  - Displays the correlation between payload mass and launch success.
  - Shows data for all sites when "All Sites" is selected in the dropdown.
  - Displays data for a specific site when a site is selected in the dropdown.
- Complete Plotly Dash code link below.
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/7_Interactive_Dashboard_with_Dash/falcon9_launch_dash_app.py
- Plotly Dash web app link below.
  - https://spacex-falcon-9-dash-app.onrender.com/

Results

EDA with Visualization

- Scatter plot hows increasing success rate with flight numbers.
- Scatter plot Indicates weak correlation between payload, launch site, and outcome.
- Bar chart highlights orbits ES-L1, GEO, HEO, and SSO with highest success rates.
- Scatter plot Shows positive correlation in LEO, no correlation in GTO, and 100% success rate in SSO with fewer flights.
- Scatter plot indicates heavier payloads correlate with higher success rates in several orbit types.
- Line chart shows overall increase in landing success rate over years with dips in 2018 and 2020.

Results

EDA with SQL

- Identified launch sites are CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40.
- Total payload mass carried by NASA boosters is 45596 units.
- Average payload mass of booster version F9 v1.1 is 2928.4 units.
- Date of first successful ground landing is December 22, 2015.
- There are 4 boosters successfully landing on a drone ship with payload between 4000 and 6000 units.
- Total number of successful outcomes 99 and failed outcomes is 1.
- There are 12 boosters that carried the maximum payload mass.
- There are 2 records of failed landing outcomes for launches in 2015.
- Listed ranked landing outcomes between 2010-06-04 and 2017-03-20.

Results

Interactive Analytics with Folium

- All launch sites are situated near the Equator, predominantly southwards of the US map.
- Launch sites are located near railways and highways, perhaps facilitating personnel and equipment transportation.
- Launch sites are located near coastlines and distant from cities, likely for safety reasons in case of launch failures.

Results

Interactive Analytics with Plotly Dash

- KSC LC-39A boasts the highest success ratio among all sites, with a 76.9% success rate and 23.1% failure rate.
- Payloads between 2000 and 4000 units exhibit the highest success rate.
- Payloads below 5000 units generally have higher success rates compared to those between 5000 and 10000 units.
- The booster version FT demonstrates the highest success rate among all booster versions analyzed.

Results

Predictive Analysis

- The Decision Tree Classifier achieved the highest classification accuracy of 88.90%.
- The confusion matrix of the best performing model (Decision Tree Classifier) showed minimal errors, with only 1 false-negative and 1 false-positive prediction.

Results

Predictive Analysis

- The Decision Tree Classifier achieved the highest classification accuracy of 88.90%.
- The confusion matrix of the best performing model (Decision Tree Classifier) showed minimal errors, with only 1 false-negative and 1 false-positive prediction.

SECTION 2

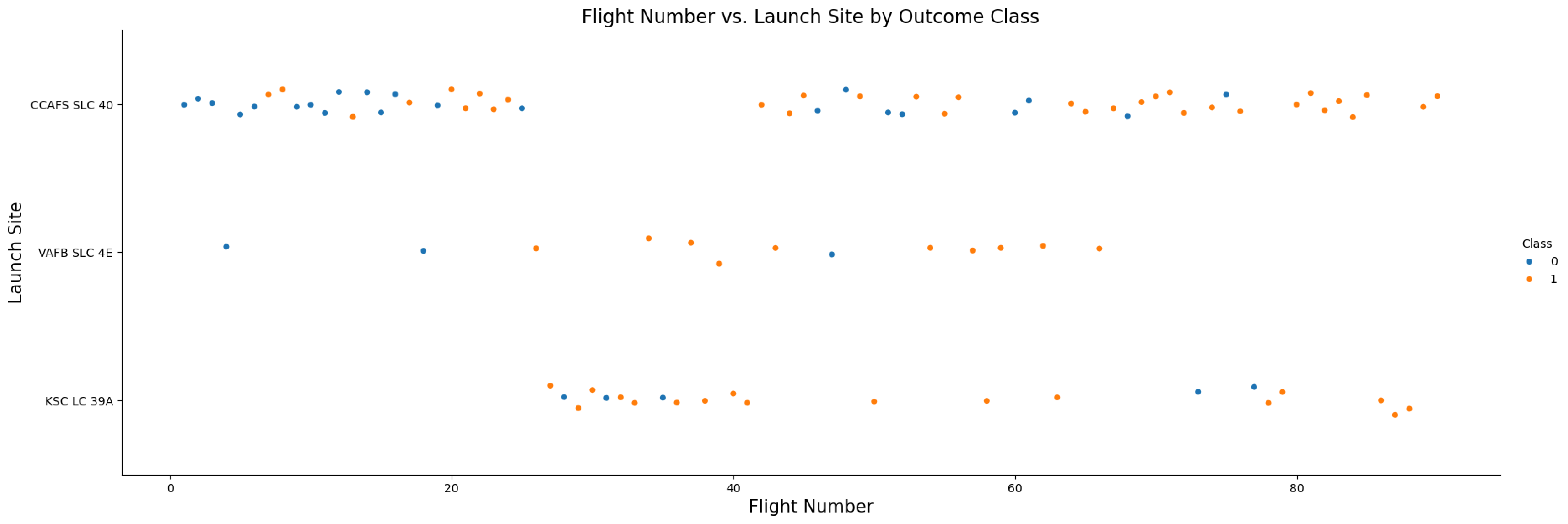
INSIGHTS DRAWN FROM EDA

NOTE for the Following Scatter Plots

- Class 0 (blue dots) represents unsuccessful launches.
- Class 1 (orange dots) represents successful launches.

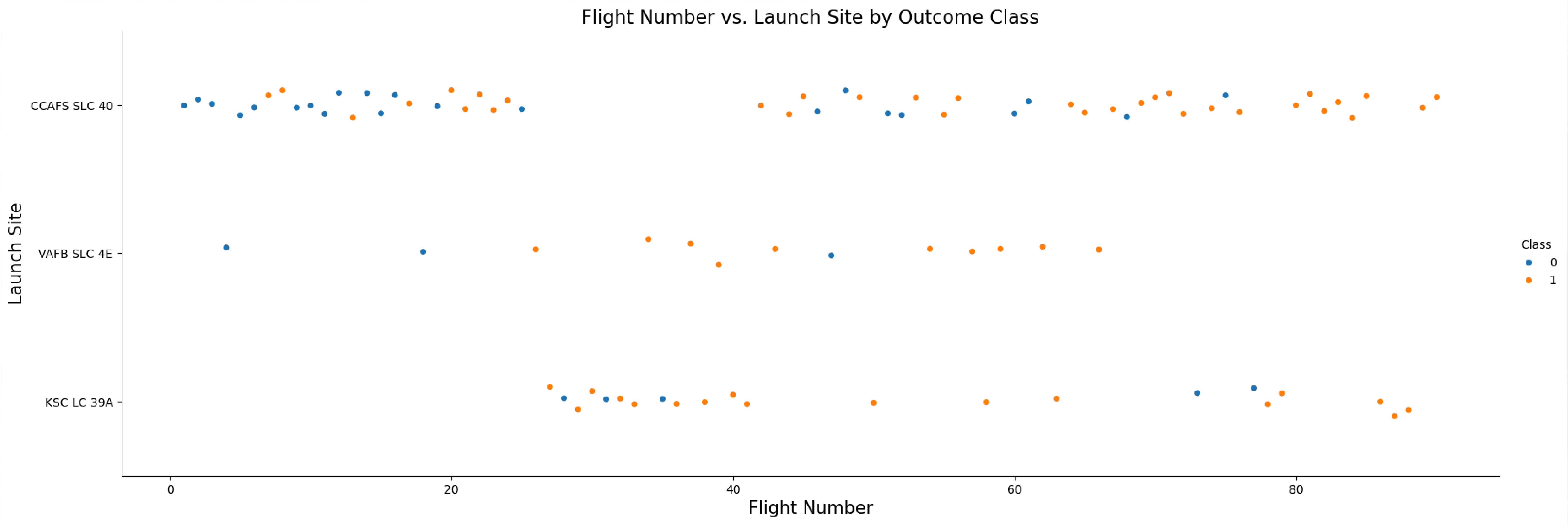
Flight Number vs. Launch Site

- This figure shows that the success rate increased as the number of flights increased.



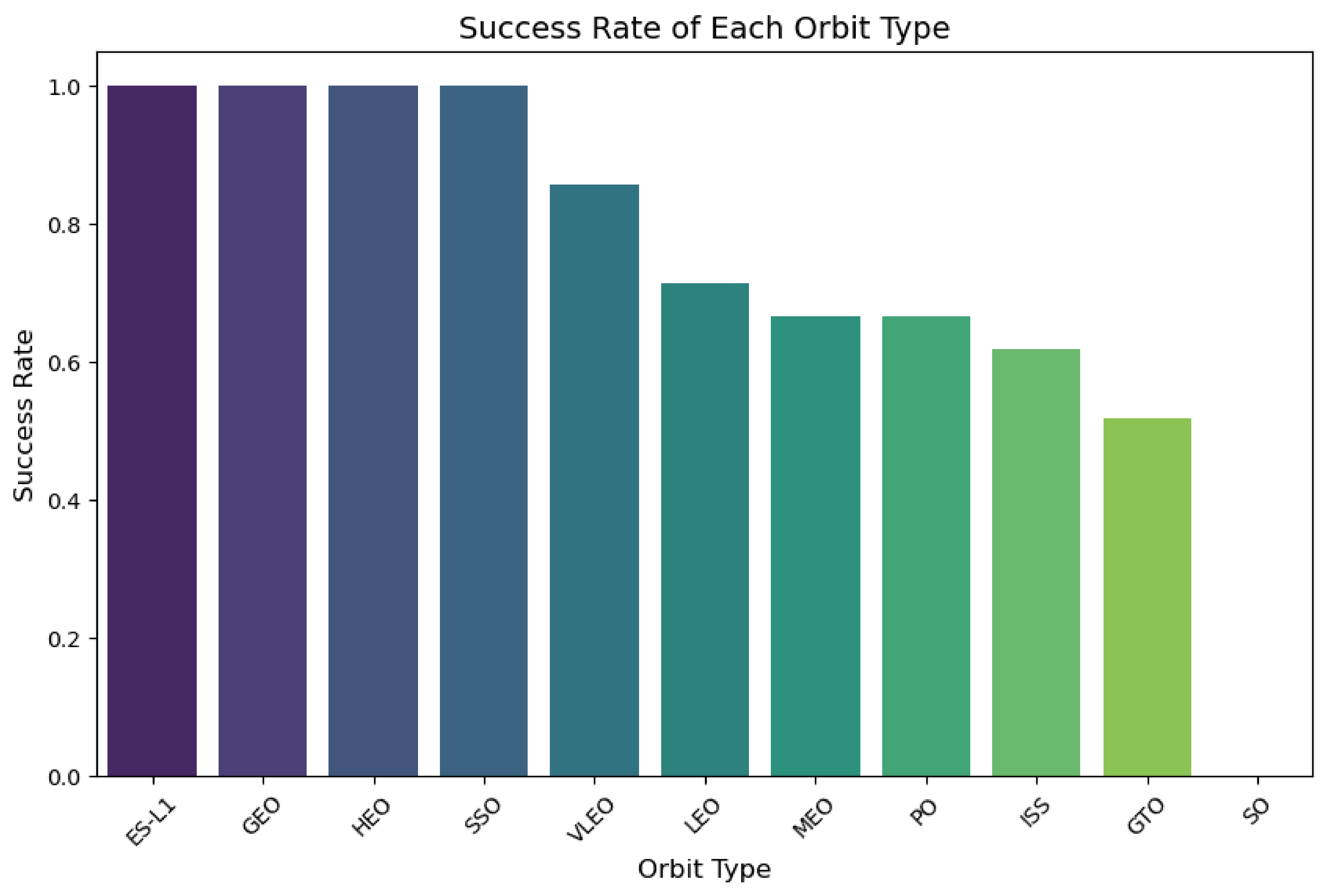
Payload vs. Launch Site

- There seems to be a weak correlation between payload, launch site and their outcome. Therefore, these are not strong predictors of whether a launch will be successful.



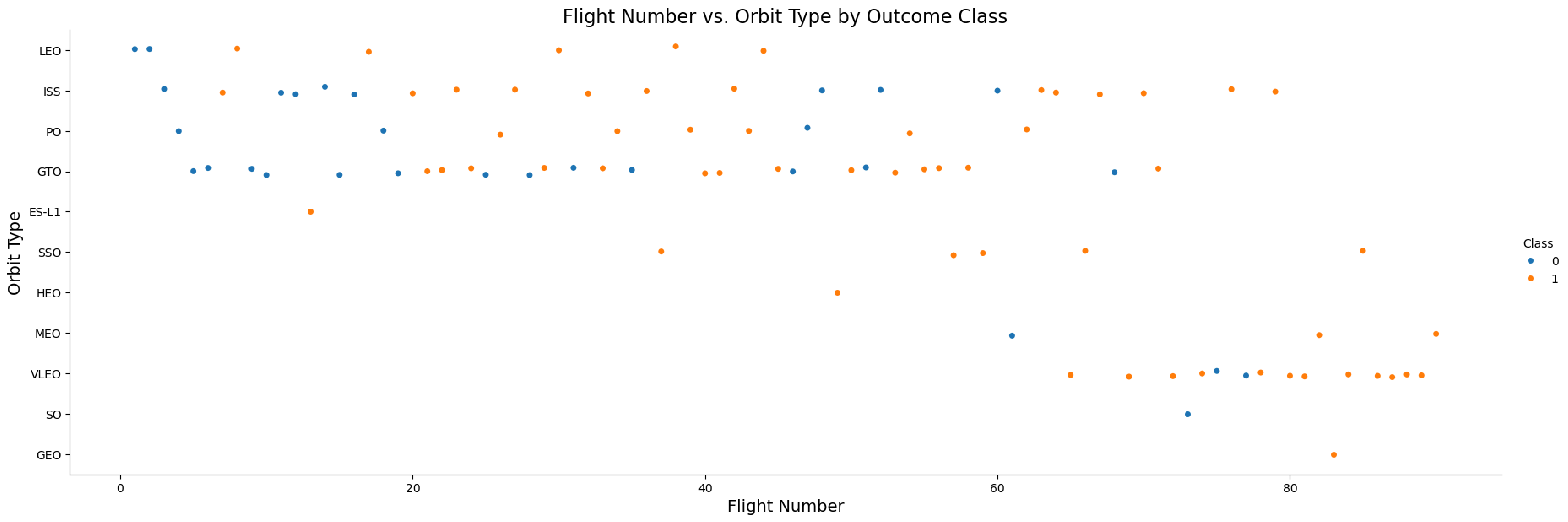
Payload vs. Launch Site

- Orbits ES-L1, GEO, HEO, and SSO have the highest success rates having 100% success rates.



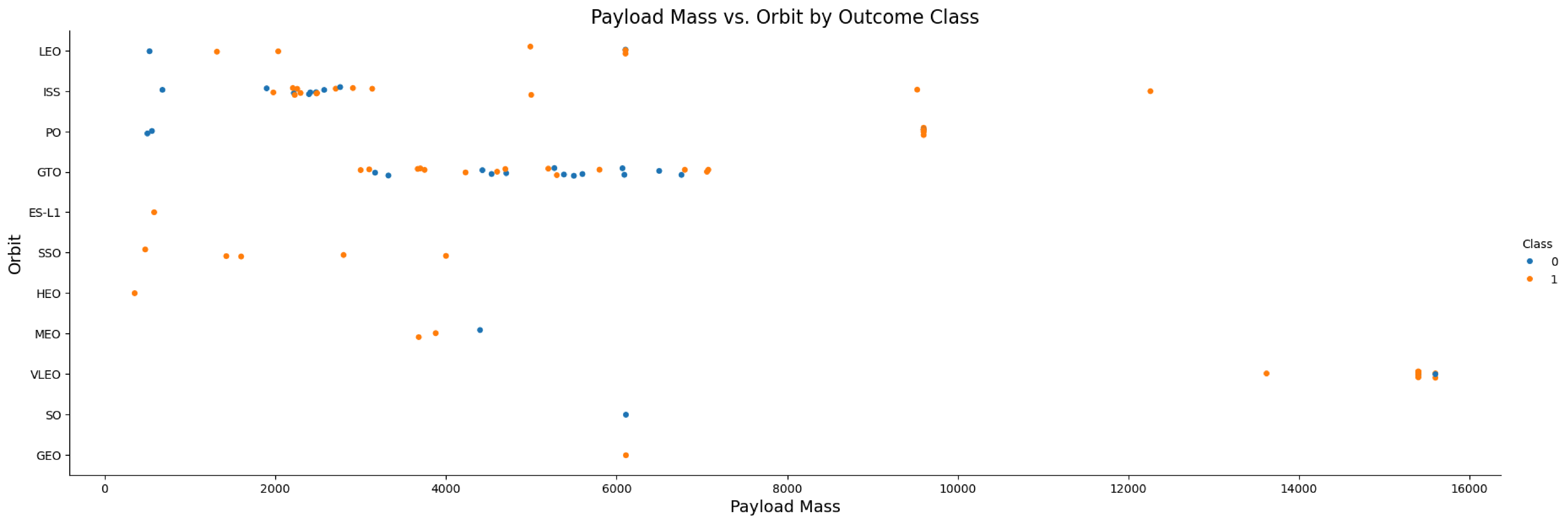
Flight Number vs. Orbit Type

- Success in LEO shows a positive correlation with flight numbers.
- There appears to be no correlation between flight numbers in the GTO.
- SSO shows a 100% success rate having fewer flights.
- Success rates are higher for flight numbers above 40.



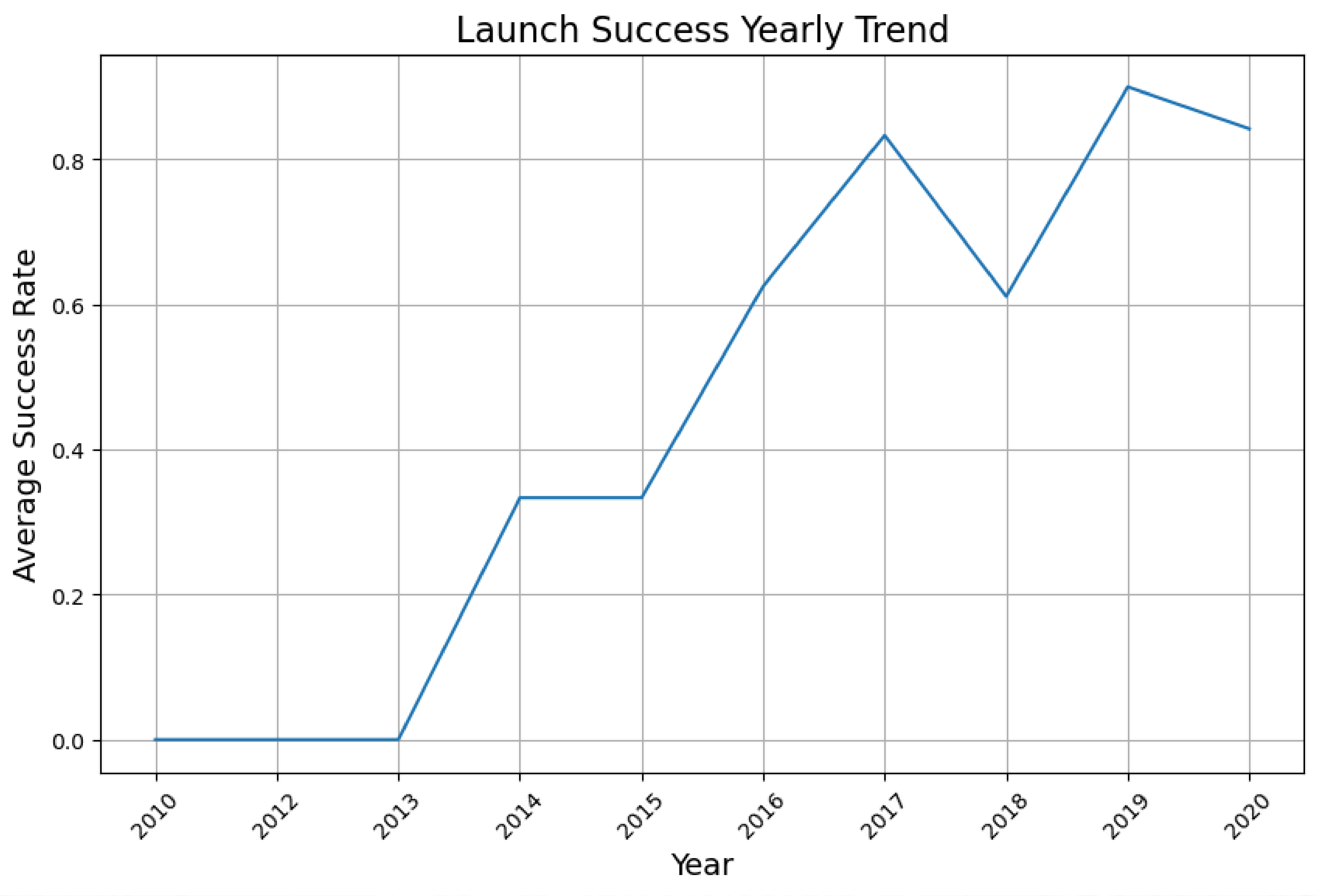
Payload vs. Orbit Type

- Heavier payloads correlates to higher success rates in PO, SSO, LEO, and ISS.
- There appears no correlation between orbit type and payload mass in GTO, as successful and failed launches are equally distributed.



Launch Success Yearly Trend

- The general trend of the chart shows an increase in landing success rate as the years pass.
- There is however a dip in 2018 as well as in 2020.



All Launch Site Names

- The DISTINCT clause was used to return only the unique launch sites from the ‘Launch_Site’ column in the SPACEXTABLE'.
- The Launch Sites are CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40.

```
SELECT DISTINCT launch_site
FROM SPACEXTABLE;
```

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

Launch Site Names Begin with 'CCA'

- LIMIT and LIKE clauses were used to display only the first 5 results of launch sites with names beginning with ‘CCA’

```
SELECT * FROM SPACEXTABLE
WHERE launch_site LIKE 'CCA%'
LIMIT 5;
```

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

- SUM() function and WHERE clause were used to calculate the total payload mass carried by boosters from NASA.
- The total payload mass carried by boosters from NASA is 45596

```
SELECT SUM(payload_mass__kg_) AS
Total_Payload_Mass, customer
FROM SPACEXTABLE
WHERE customer = 'NASA (CRS)';
```

Total_Payload_Mass	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- AVG() function and WHERE clause were used to calculate the average payload mass carried by booster version F9 v1.1.
- The average payload mass carried by booster version F9 v1.1 is 2928.4.

```
SELECT AVG(payload_mass__kg_) AS
Average_Payload_Mass, booster_version
FROM SPACEXTABLE
WHERE booster_version = 'F9 v1.1';
```

Average_Payload_Mass	Booster_Version
2928.4	F9 v1.1

Date of First Successful Ground Pad Landing

- MIN() function and WHERE clause were used to find the dates of the first successful landing outcome on ground pad.
- The date of the first successful outcome on ground pad is December 22, 2015.

```
SELECT MIN(date) AS
Date_of_First_Successful_Landing_in_Ground_Pad
FROM SPACEXTABLE
WHERE landing_outcome = 'Success (ground pad)';
```

Date_of_First_Successful_Landing_in_Ground_Pad
2015-12-22

Boosters Landed on Drone Ship (4k to 6k Payload)

- WHERE and AND clauses were used to list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.
- The boosters that have successfully landed on drone ship having payload mass greater than 4000 and less than 6000 are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2.

```
SELECT booster_version
FROM SPACEXTABLE
WHERE landing_outcome = 'Success (drone ship)'
AND payload_mass__kg_ > 4000
AND payload_mass__kg_ < 6000;
```

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

Total of Successful and Failure Mission Outcomes

- COUNT() function and IN and GROUP BY clauses were used to calculate the total number of successful and failure mission outcomes.
- The total number of successful missions are 99.
- The total number of failure missions is 1.

```
SELECT mission_outcome, COUNT(*) AS Total
FROM SPACEXTABLE
WHERE mission_outcome IN
('Success','Failure (in flight)','Success
(payload status unclear)','Success')
GROUP BY mission_outcome;
```

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

Boosters that Carried Maximum Payload

- WHERE and MAX() function in a subquery were used to list the names of the booster which have carried the maximum payload mass.

```
SELECT 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

- WHERE and substr(date,0,5) function to extract the month were used to list the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- CASE expression with WHEN clause and substr(date,6,2) function to extract the months were used to defined the month names.

```
SELECT
CASE
    WHEN substr(date,6,2) = '01' THEN 'January'
    WHEN substr(date,6,2) = '02' THEN 'February'
    ...
    ...
END AS Month_Name, landing_outcome,
booster_version, launch_site
FROM SPACEXTABLE
WHERE substr(date,0,5) = '2015'
    AND landing_outcome = 'Failure (drone ship)';
```

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Land Outcomes Between 06-04-10 and 03-20-17

- COUNT() function, WHERE, AND, GROUP BY, ORDER BY, and DESC clauses were used to list the landing outcomes between the dates 06-04-2010 and 03-20-2017, in descending order.

```
SELECT landing_outcome, COUNT(*) AS Count
FROM SPACEXTABLE
WHERE date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing_outcome
ORDER BY Count DESC;
```

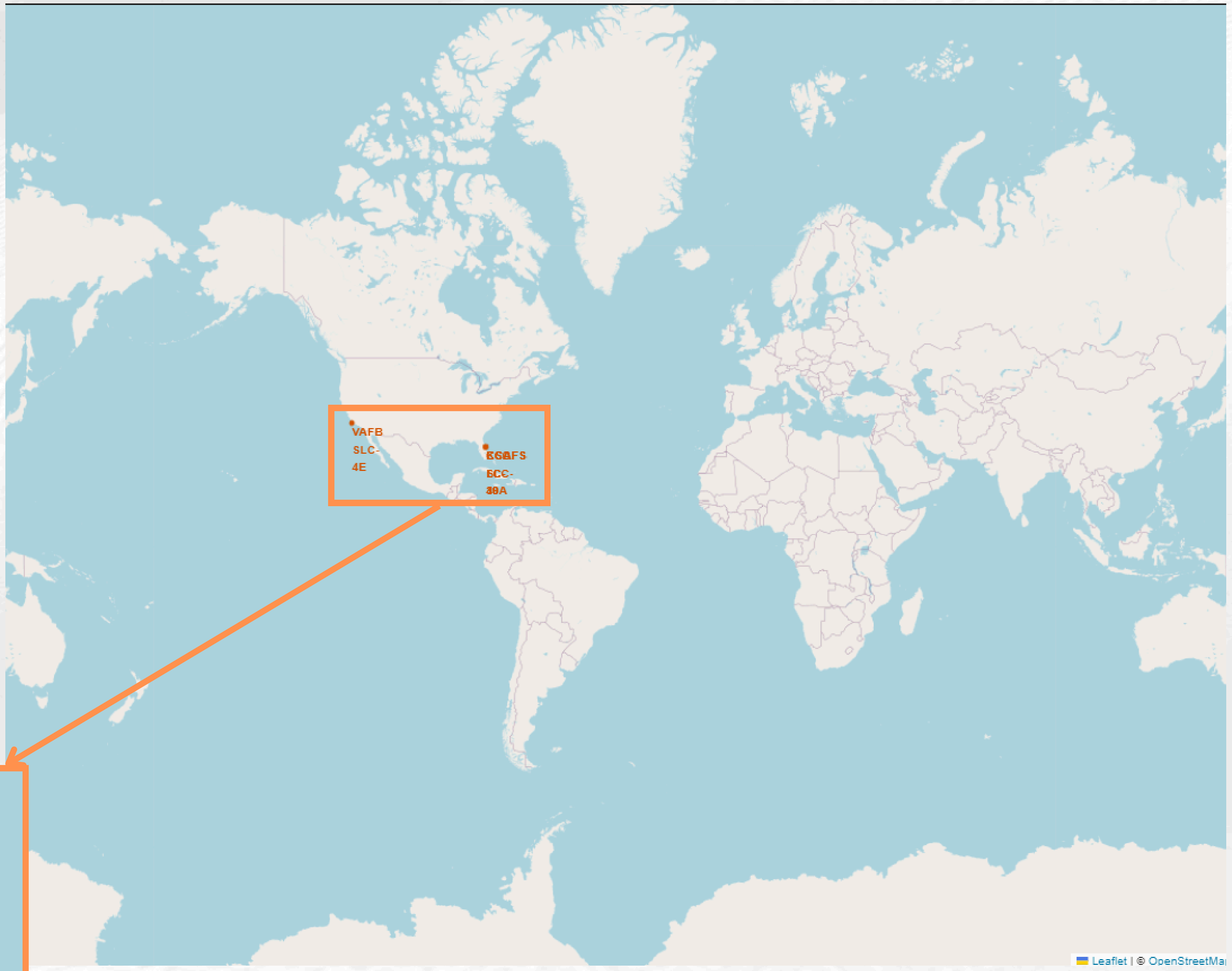
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

SECTION 3

LAUNCH SITES PROXIMITIES ANALYSIS

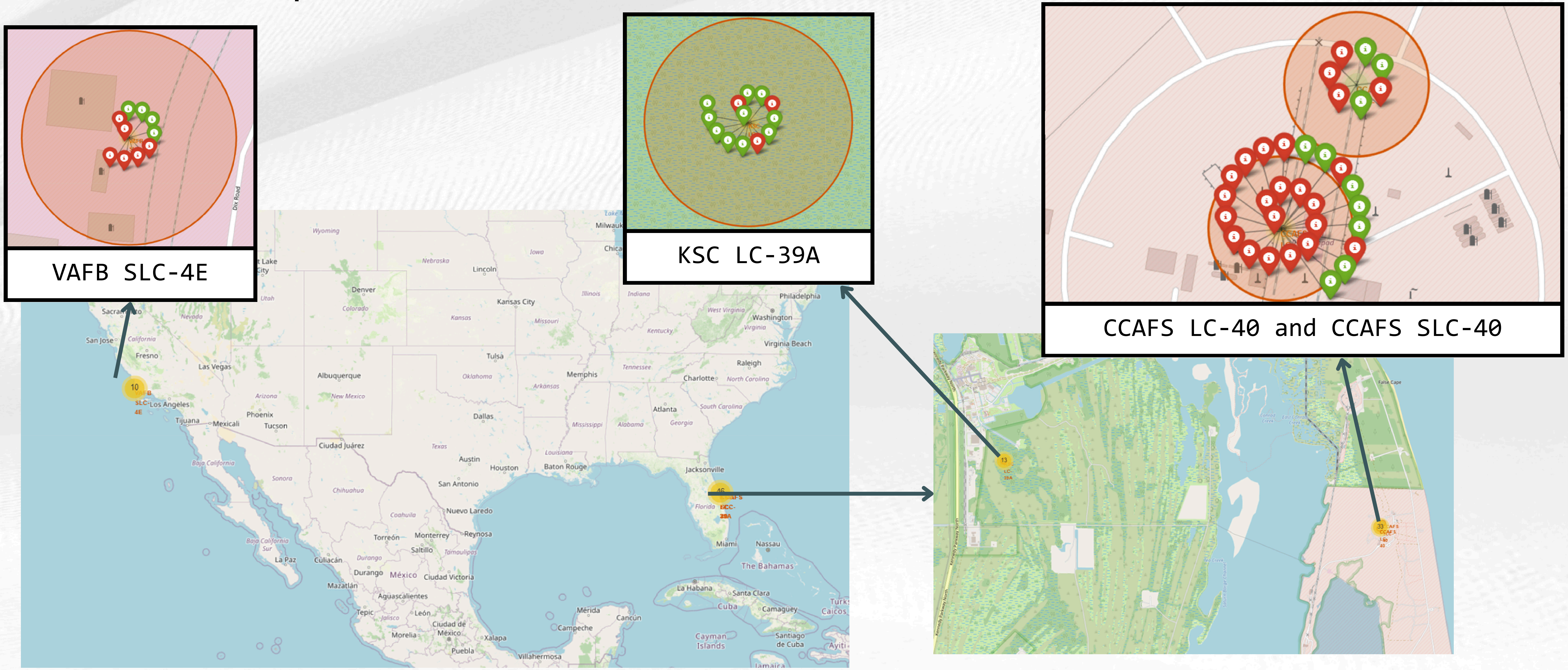
SpaceX Launch Sites Locations

- All launch sites are in proximity to the Equator, located southwards of the US map.
- All the launch sites are in very close proximity to the coast.



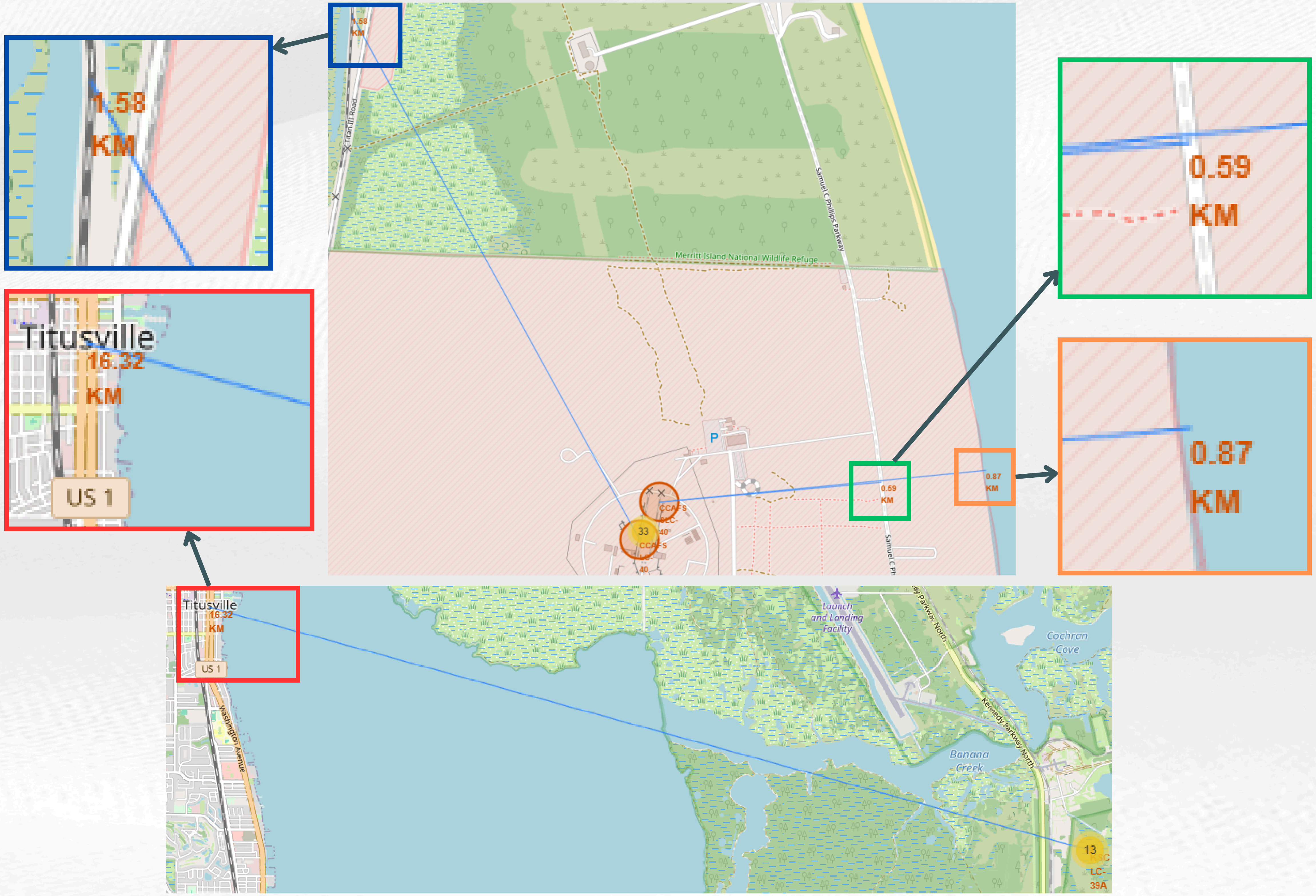
Falcon 9 Success/Fail Launches by Site

- The successful launches are represented by a green marker while the red marker represents failed rocket launches.



Launch Sites Proximities

- Launch sites are in close proximity to railways and highways perhaps due to transportation of personnel and equipment.
- They are close in proximity to coastlines and far away from cities perhaps due to safety concerns in failure of launches.

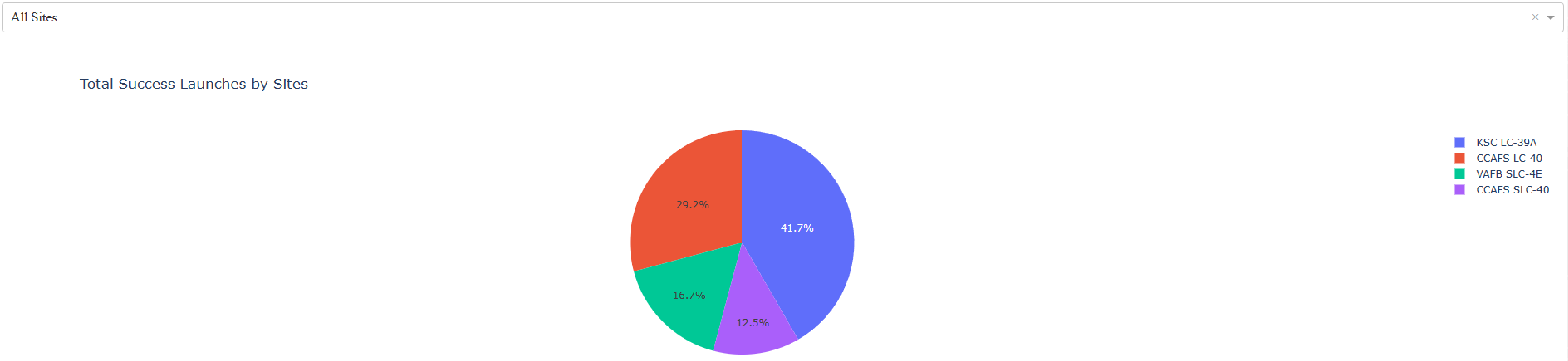


SECTION 4

DASHBOARD WITH PLOTLY DASH

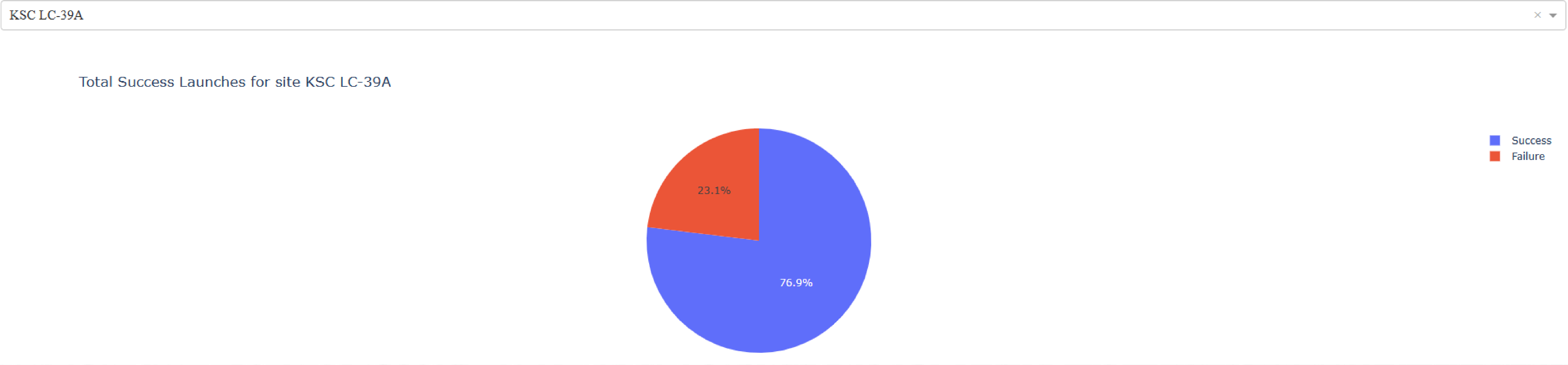
Pie Chart of Total Success Launches by Sites

- Launch site KSC LC-39A has the highest launch success rate at 41.7% followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7% and lastly launch site CCAFS SLC-40 at 12.5%.



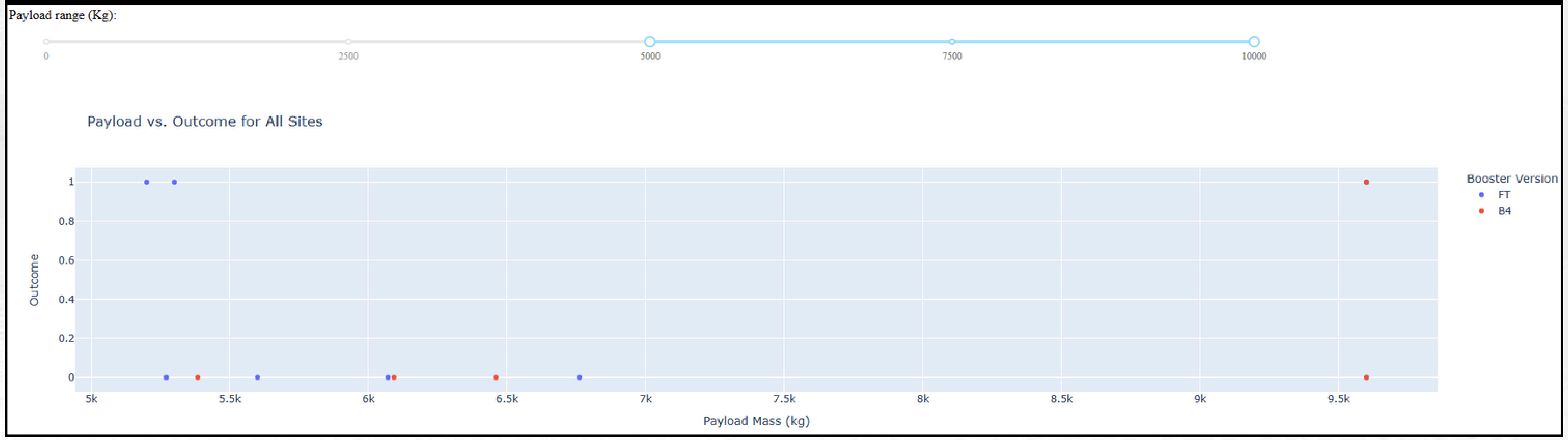
Pie Chart of Total Success for Site KSC LC-39A

- Launch site KSC LC-39A has the highest success ratio of 76.9% success against 23.1% failed launches.



Payload vs Launch Outcomes for All Sites

- Payload mass between 0 and 5000 have larger success rate compared to between 5000 and 10000.
- Payload mass between 2000 and 4000 have the highest success rate.
- Booster version FT has the largest success rate.

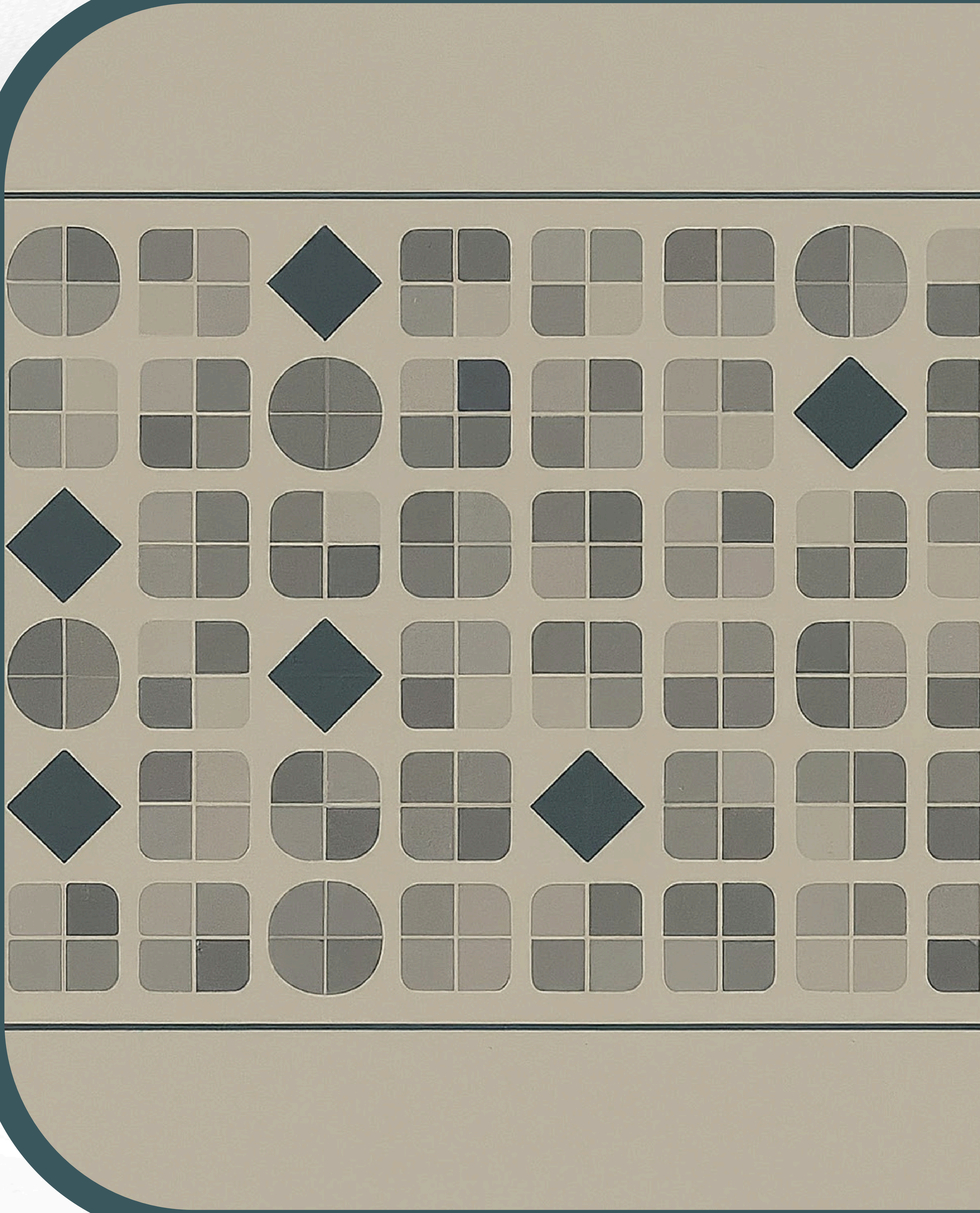


SpaceX Falcon 9 Plotly Dash Web App

- Explore the Dashboard through the link below.
- Plotly Dash App web app link.
  - https://spacex-falcon-9-dash-app.onrender.com/

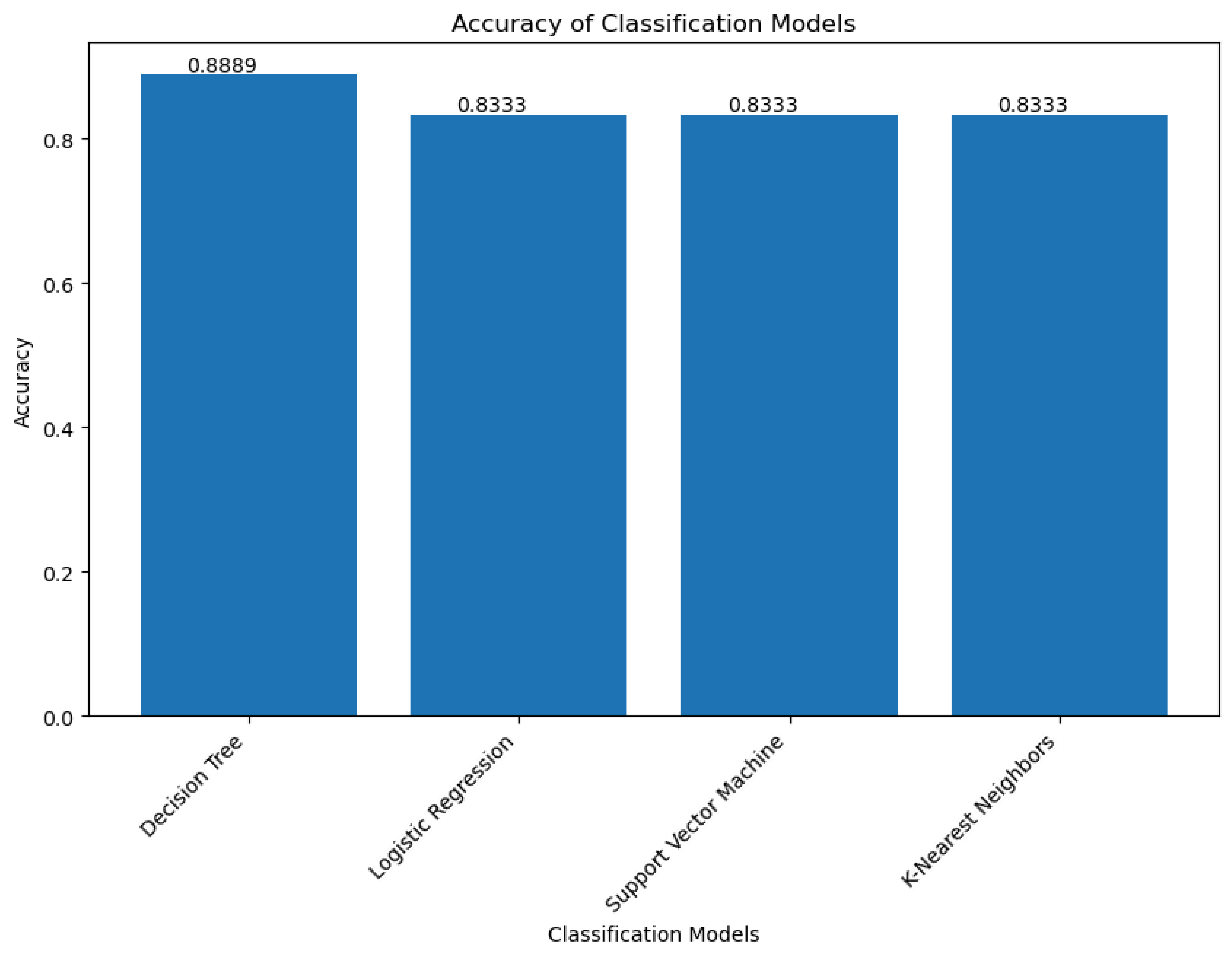
SECTION 5

PREDICTIVE ANALYSIS (CLASSIFICATION)



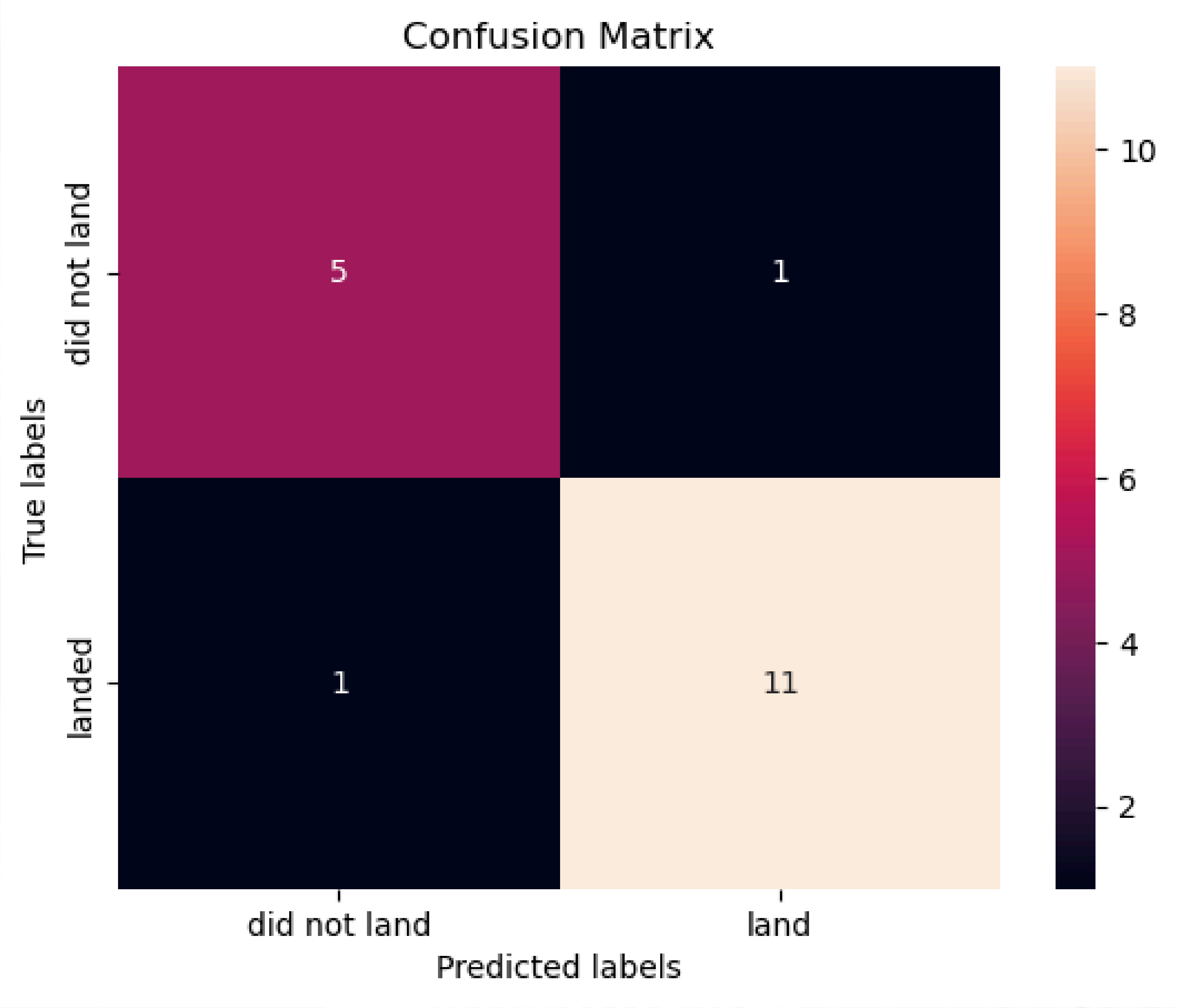
Classification Accuracy

- The model that has the highest classification accuracy is the Decision Tree Classifier having 88.90% accuracy.



Confusion Matrix

- This is the confusion matrix of the best performing model which is the Decision Tree Classifier.
- The model failed to predict only 1 false-negative and 1 false-positive.



Conclusions

- The project aimed to predict the successful landing of Falcon 9 first stages and identify factors influencing successful landings.
- Data collection involved API requests from SpaceX and web scraping from Wikipedia, followed by extensive data wrangling and exploratory data analysis (EDA).
- EDA revealed insights such as increasing success rates with flight numbers and varying success rates across different orbit types.
- Interactive analytics using Folium and Plotly Dash provided visualizations of launch site locations, success rates, and payload outcomes.

Conclusions

- Findings showed that launch sites are strategically located near coastlines and away from cities for safety considerations, with close proximity to railways and highways facilitating transportation of personnel and equipment.
- Predictive analysis involved training classification models to predict landing outcomes, with the Decision Tree Classifier emerging as the best-performing model with 88.90% accuracy.
- The project offers valuable insights for stakeholders in space exploration, aiding in decision-making processes related to launch planning and operations.

Conclusions

Answers to the problem questions

- How can we reliably predict whether the first stage of a Falcon 9 rocket will successfully land after launch?
    - We can reliably predict the successful landing of a Falcon 9 first stage using machine learning classification models trained on historical data. The Decision Tree Classifier, identified as the best-performing model in our project with 88.90% accuracy, can effectively predict landing outcomes.

Conclusions

Answers to the problem questions

- What factors contribute most significantly to the successful landing of the Falcon 9 first stage?
    - Findings from our exploratory data analysis reveal several significant factors influencing the successful landing of the Falcon 9 first stage. These include flight number, orbit type, payload mass, launch site characteristics, booster versions, and environmental conditions.

Conclusions

Answers to the problem questions

- Are there any observable patterns or indicators that could help us determine the likelihood of a successful landing?
  - Our analysis uncovered observable patterns indicating higher success rates with increasing flight numbers and specific orbit types like ES-L1, GEO, HEO, and SSO. Payload mass also correlates with success rates in various orbit types. These findings provide valuable indicators for determining the likelihood of a successful landing.

Appendix

- GitHub Repository of the Project
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction
- Jupyter Notebook Links from GitHub Repository
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/1_Data_Collection_API.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/2_Data_Collection_Web_Scraped.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/3_Data_Wrangling.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/4_Exploratory_Data_Analysis_Using_SQL.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/5_Exploratory_Data_Analysis_with_Visualization.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/6_Interactive_Visual_Analytics_with_Folium.ipynb
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/blob/main/8_Falcon9_First_Stage_Landing_Prediction.ipynb

Appendix

- Link of all the data Created/Used in the Project from GitHub Repository
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Landing-Prediction/tree/main/data
- Separate GitHub Repository Link of the Plotly Dash App for Deployment
  - https://github.com/rjacaac211/SpaceX-Falcon-9-Dash-App
- Plotly Dash App Web App Link
  - https://spacex-falcon-9-dash-app.onrender.com/
- SpaceX Rockets API
  - https://api.spacexdata.com/v4/rockets/
- Wikipedia Page (List of Falcon 9 and Falcon Heavy launches)
  - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

THANK YOU !

+639207005579

rj.acaac211@gmail.com

RJ R. Aca-ac