IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Bilal Gadajev
25/06/2025

Outline

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

Executive Summary

Methodologies

- **Data Preparation and Filtering:**
- Organized and sanitized launch data in the style of SpaceX
- Employed Pandas for analysis, filtering, and aggregation tasks.
- **SQL Analytics:**
- Inquired about metrics such as success rates, payload distributions, and the effectiveness of boosters.
- Metrics derived include total payload categorized by the type of mission and the earliest successful landings.
- **Geospatial Mapping (Folium):**
- Launch sites and the infrastructure around them, such as coastlines, highways, and cities, were visualized.
- Employed the Haversine formula to calculate the distances from launch pads to logistical points.
- **Interactive Dashboard (Plotly Dash):**
- Developed a pie chart that uses dropdown controls to display the performance of the launch site.
- Introduced a scatter plot and range slider for examining payload versus. relationship of success

Executive Summary

- Key Results & Insights

Metrics	Results
Top Launch Site	CCAFS LC-40 had the highest number of successful launches
Success vs Failure Breakdown	Highest reliability sites have >90% success
Payload Impact on Success	Mid-range payloads (4,000–6,000 kg) had higher success rates
Booster Performance	Booster version F9 FT showed highest reliability across most payloads
Spatial Proximity Analysis	Mapped proximity to coastline (3.5 km), nearest city (83 km), and highway (3 km)

Introduction

Project Overview

Background and Context:

- Space Y performs regular orbital launches at various locations, utilizing different boosters and payload capacities.
- A centralized tool for analysing performance trends across missions is lacking for business stakeholders.

Essential Inquiries:

- Which launch sites experience the greatest rates of success?
- In what ways does the mass of the payload influence the results of a mission?
- Which versions of boosters are considered to be the most dependable?
- Do specific infrastructure elements, such as closeness to coastlines, have a relationship with launch outcomes?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Launch data was collected from a structured CSV file stimulating SpaceX mission records
- Perform data wrangling
  - Cleaned and filtered using libraries such as Pandas; key fields were extracted transformed and grouped that allowed for data analysis and visualization.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Model training: trained classification models according by grouping training and testing data
  - Model Evaluation: Used GridSearchCV to tune the parameters and find the best model according to metrics such as accuracy, recall and precision.

Data Collection

[DataScienceEcosystem/jupyter-labs-spacex-data-collection-api.ipynb at main · Abahacker123/DataScienceEcosystem]

- **Data Collection:**

1. Request to the SpaceX API: make a get request to the SpaceX API.
2. Then, decoded the response content as a json file using .json(), and turn it into a Pandas data frame using .json_normalize()
3. From the rocket we got **booster name.**
4. From the payload we got **mass of the payload** and **the orbit** that it is going to.
5. From the launchpad got the **launch site** being used, **the longitude**, and **the latitude**.
6. Applied "getBoosterVersion" function method to get the booster version.
7. Created a Pandas data frame from the dictionary launch_dict.
8. Finally filtered the dataframe to only include **Falcon 9 launches**

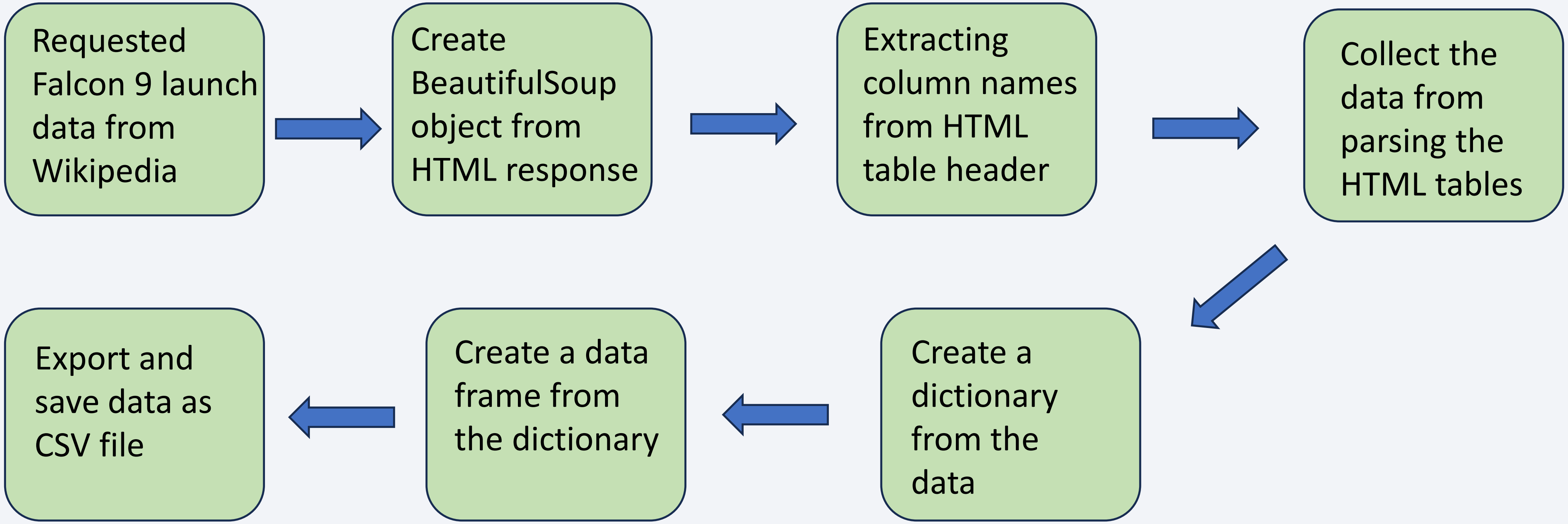
SpaceX API: get request → Convert json to Pandas data frame → Key data extraction e.g launch sites → Booster Version → Filter data: only Falcon 9

Data Collection – SpaceX API

- DataScienceEcosystem/jupyter-labs-spacex-data-collection-api.ipynb at main · Abahacker123/DataScienceEcosystem

1. REST API
2. API Endpoint
3. HTTP GET Request
4. JSON Response
5. Data Parsing
6. Json_normalize()
7. Data Filtering/Preprossing
8. Data Enrichment: getBoosterVersion
9. Data Storing

Data Collection - Scraping



[DataScienceEcosystem/jupyter-labs-webscraping.ipynb at main · Abahacker123/DataScienceEcosystem]()

Data Wrangling

Aim:

- Use EDA to establish patterns in the data → Data Analysis
- Identify labels of target for model training

Method:

- Calculate the number of launches at each site
- Calculate the number of orbits and their occurrence
- Calculate the number and occurrence of mission outcome per orbit
- Creating a landing outcome label for target column
- Export data as CSV file

DataScienceEcosystem/labs-jupyter-spacex-Data wrangling.ipynb at main · Abahacker123/DataScienceEcosystem

EDA with Data Visualization

DataScienceEcosystem/edadataviz.ipynb at main · Abahacker123/DataScienceEcosystem

Plotted Charts:

- Flight Number vs Payload Mass
- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- Orbit Type vs Success Rate
- Flight Number vs Orbit Type
- Payload Mass vs Orbit Type
- Yearly Success Rates

Reason of use: To see if there is a relationship between these parameters, can be useful to apply when building ML models for predictive analysis.

EDA with SQL

[DataScienceEcosystem/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · Abahacker123/DataScienceEcosystem]

- Performed SQL Queries:
- Display each unique launch site
- Show 5 instances where launch site names being with “CCA”
- Calculate the total payload mass carried by boosters launched by “NASA(CRS)”
- Calculate the average payload mass carried by the v1.1 Falcon 9 booster
- Find the date of the first successful landing
- Display a list of the successful landing boosters with payloads in the range: 4000-6000kg
- List the successful and unsuccessful missions
- List all booster versions with maximum payload mass
- List important information such as name, booster, launch site for mission that failed landing in 2015
- Show the landing outcome distribution between June 4th 2010 – March 20th 2017

Build an Interactive Map with Folium

- In order to establish geographical, understand of surrounding environment to the launch site, the following observations were made:
- Every launch site in the country
- Successful and unsuccessful launches
- Distance between launch sites and landmarks such as railways, high ways and cities.

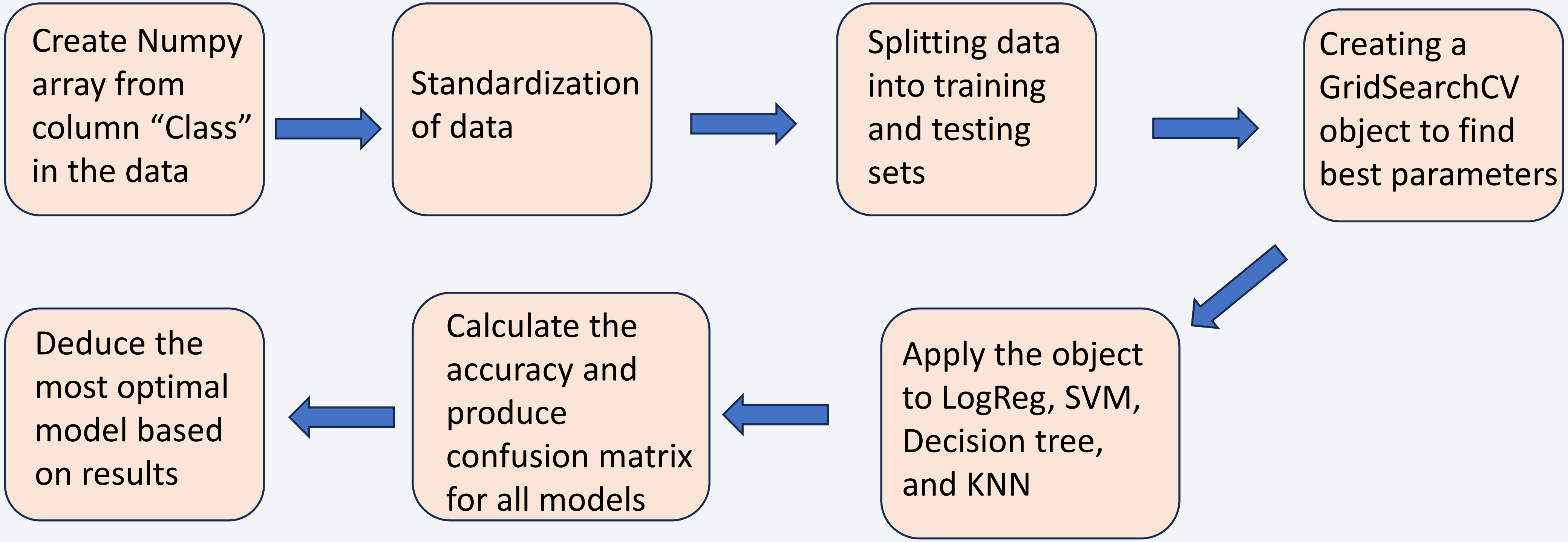
[DataScienceEcosystem/lab_jupyter_launch_site_location.ipynb at main · Abahacker123/DataScienceEcosystem]

Build a Dashboard with Plotly Dash

- **Dropdown list (Launch sites)**
  - You can select all launch sites but also specific site
- **Pie chart with Successful Launches**
  - Allows to see the successful launches as a % of all launches
- **Slides of Payload Mass Range**
  - You can choose which payload mass you want to study
- **Scatter Chart: Payload Mass vs. Success Rate by Booster Version**
  - You can see the patterns between booster type, payload mass and launch success

Abahacker123/DataScienceEcosystem: IBM Data Science Course 2 assignment Module 6

Predictive Analysis (Classification)



[DataScienceEcosystem/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · Abahacker123/DataScienceEcosystem]

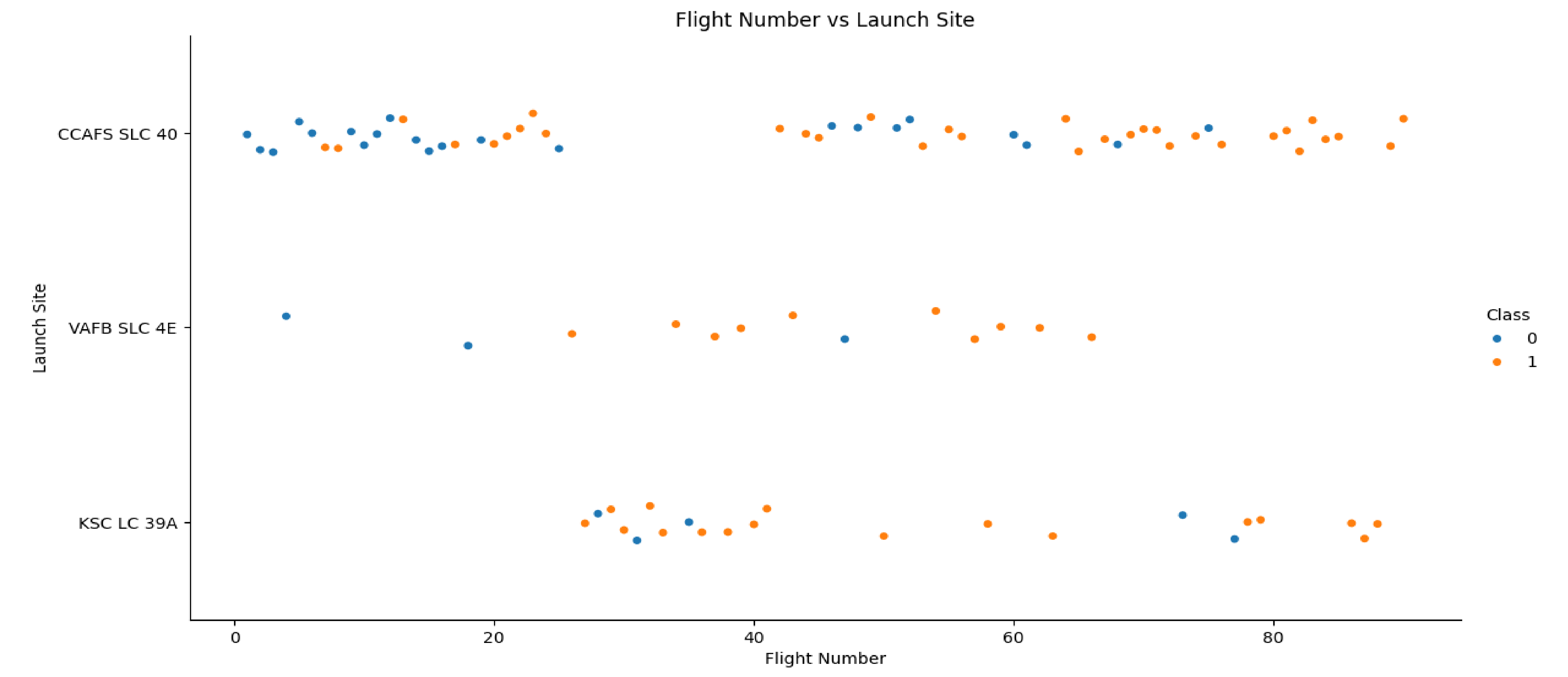
Results

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

Section 2

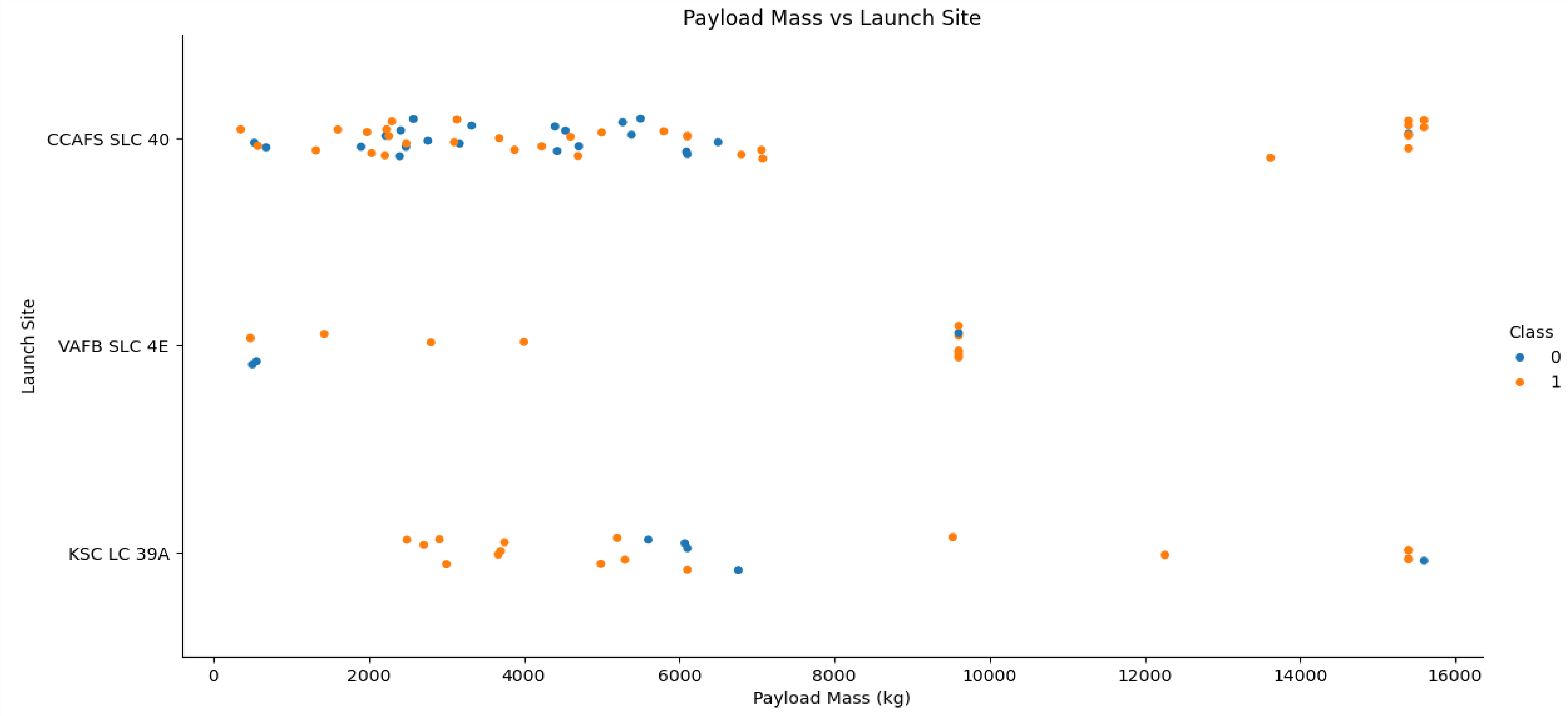
Insights drawn from EDA

Flight Number vs. Launch Site



- All sites has more failures of launches in the early stages, with improving strategy and launch tech, improvements of successful launches is visibly across all sites.

Payload vs. Launch Site



- Across all launch sites, a pattern exists with increasing payload, and increasing successful launches. It appears that the lighter the payload, the greater the risk of failure upon launch.

Success Rate vs. Orbit Type

- **Orbits with 100% success rate:**

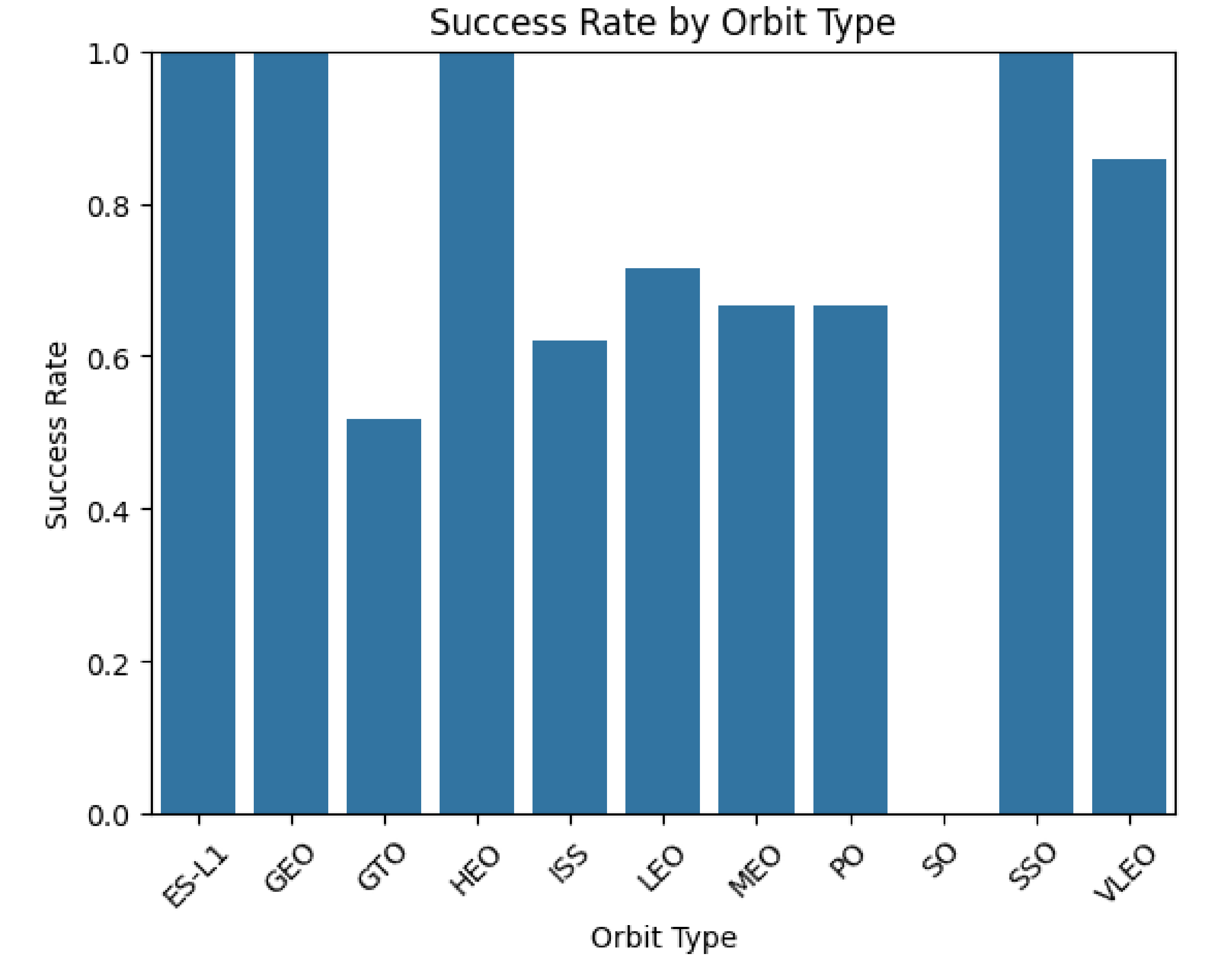
- ES-L1, GEO, HEO, SSO

- **Orbits with 0% success rate:**

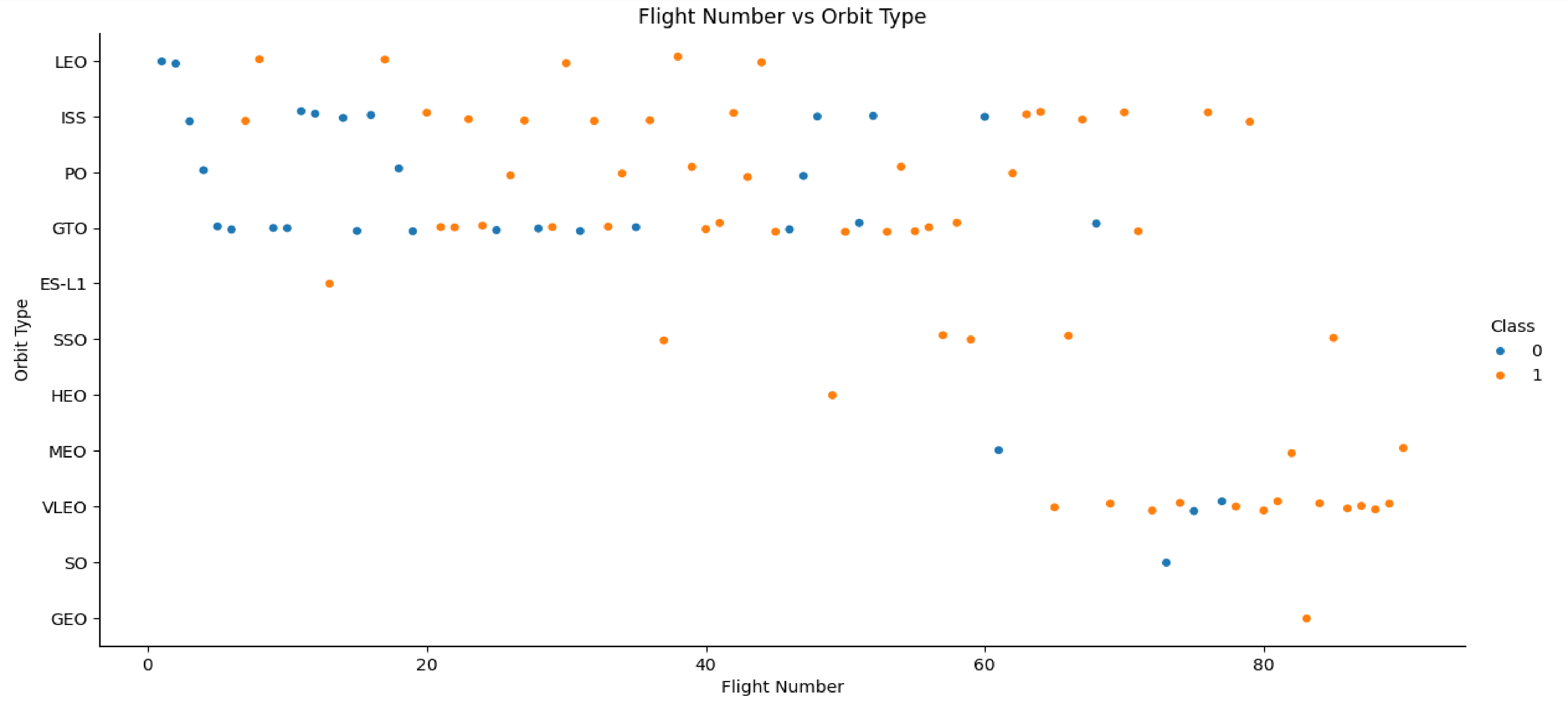
- SO

- **Orbits between 50-90% success rate:**

- GTO, ISS, LEO, MEO, PO, VLEO

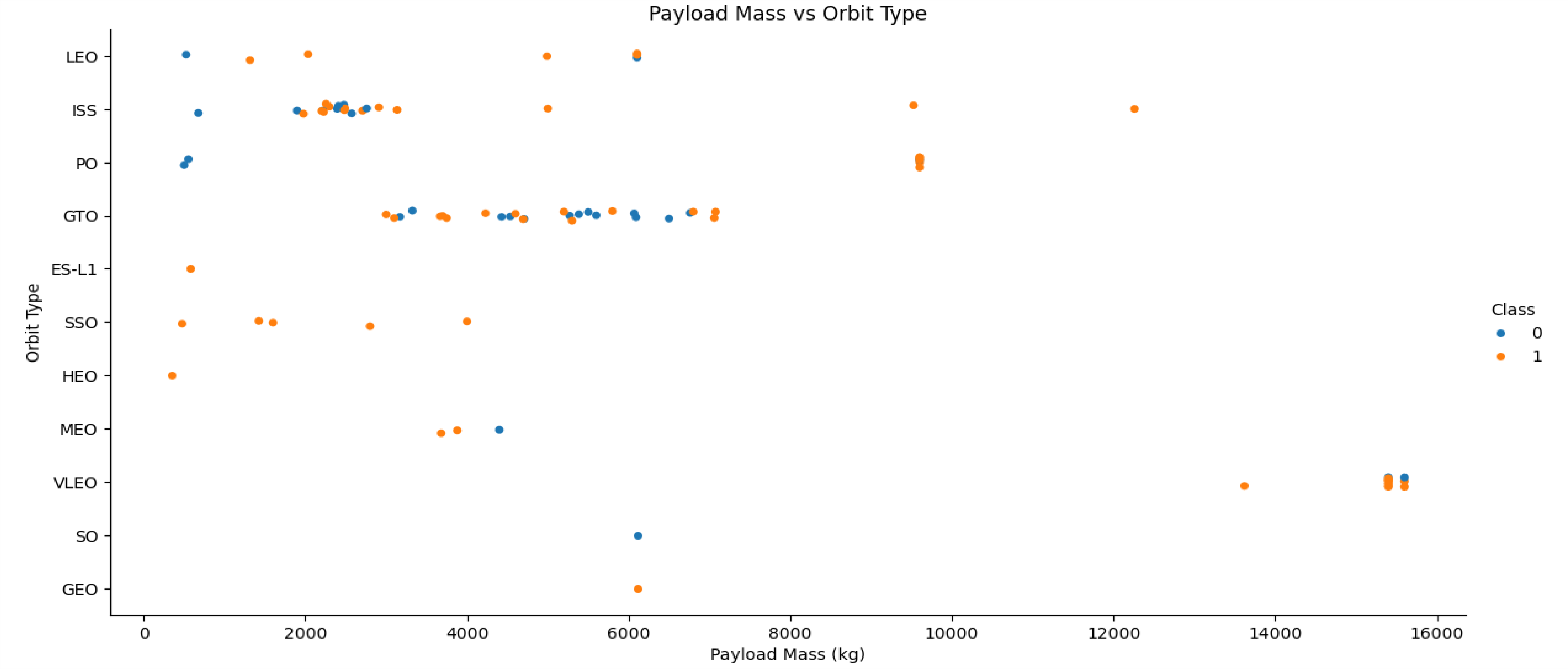


Flight Number vs. Orbit Type



- Success rate increases generally with higher number of launches: showing improving experience and increased tech development

Payload vs. Orbit Type

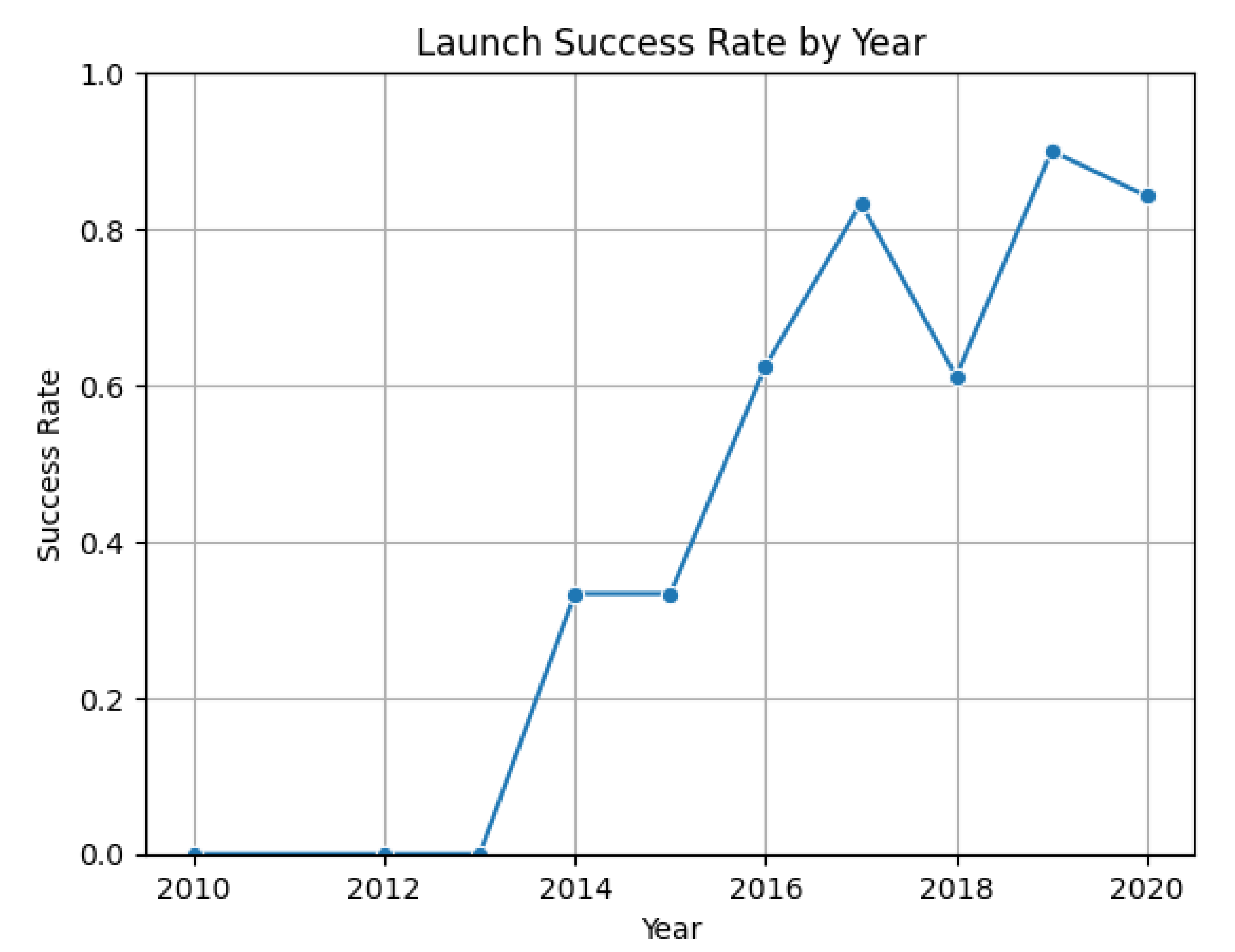


- The orbits LEO, ISS and PO are better with heavier payloads
- The GTO orbit shows inconsistent results with varying payload, thus inconclusive.

Launch Success Yearly Trend

- In 2013, the success rate increases until 2017, where there is a minor off peak, however it turn back in 2019.

- The success rate continues to increase until 2020.



All Launch Site Names

- There are 4 Launch Sites:

I. CCAFS LC-40

II. VAFB SLC-4E

III. KSC LC-39A

IV. CCAFS SLC-40

```
SELECT DISTINCT Launch_Site from SPACEXTABLE;
```

Launch Site Names Begin 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

Task 3

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

[15]:

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

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

[15]:

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

Task 4

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

[16]:
```
%%sql
SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

[16]:

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

[17]:
```
%%sql
SELECT MIN("Date") AS First_Successful_Landing_Date
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

[17]:

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[18]:
```
%%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
Done.
```

[18]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

[19]:

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

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

[19]:

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

Boosters Carried Maximum Payload

Task 8 ¶

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

[20]:
```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE "Payload_Mass__kg_" = (
    SELECT MAX("Payload_Mass__kg_")
    FROM SPACEXTABLE
);
```

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

[20]:

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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[21]:
```
%%sql
SELECT
  CASE substr("Date", 6, 2)
    WHEN '01' THEN 'January'
    WHEN '02' THEN 'February'
    WHEN '03' THEN 'March'
    WHEN '04' THEN 'April'
    WHEN '05' THEN 'May'
    WHEN '06' THEN 'June'
    WHEN '07' THEN 'July'
    WHEN '08' THEN 'August'
    WHEN '09' THEN 'September'
    WHEN '10' THEN 'October'
    WHEN '11' THEN 'November'
    WHEN '12' THEN 'December'
  END AS Month_Name,
  "Landing_Outcome",
  "Booster_Version",
  "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date", 1, 4) = '2015'
  AND "Landing_Outcome" = 'Failure (drone ship)';
```

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

[21]:

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

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

Task 10

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.

[22]:
```
%%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.
```

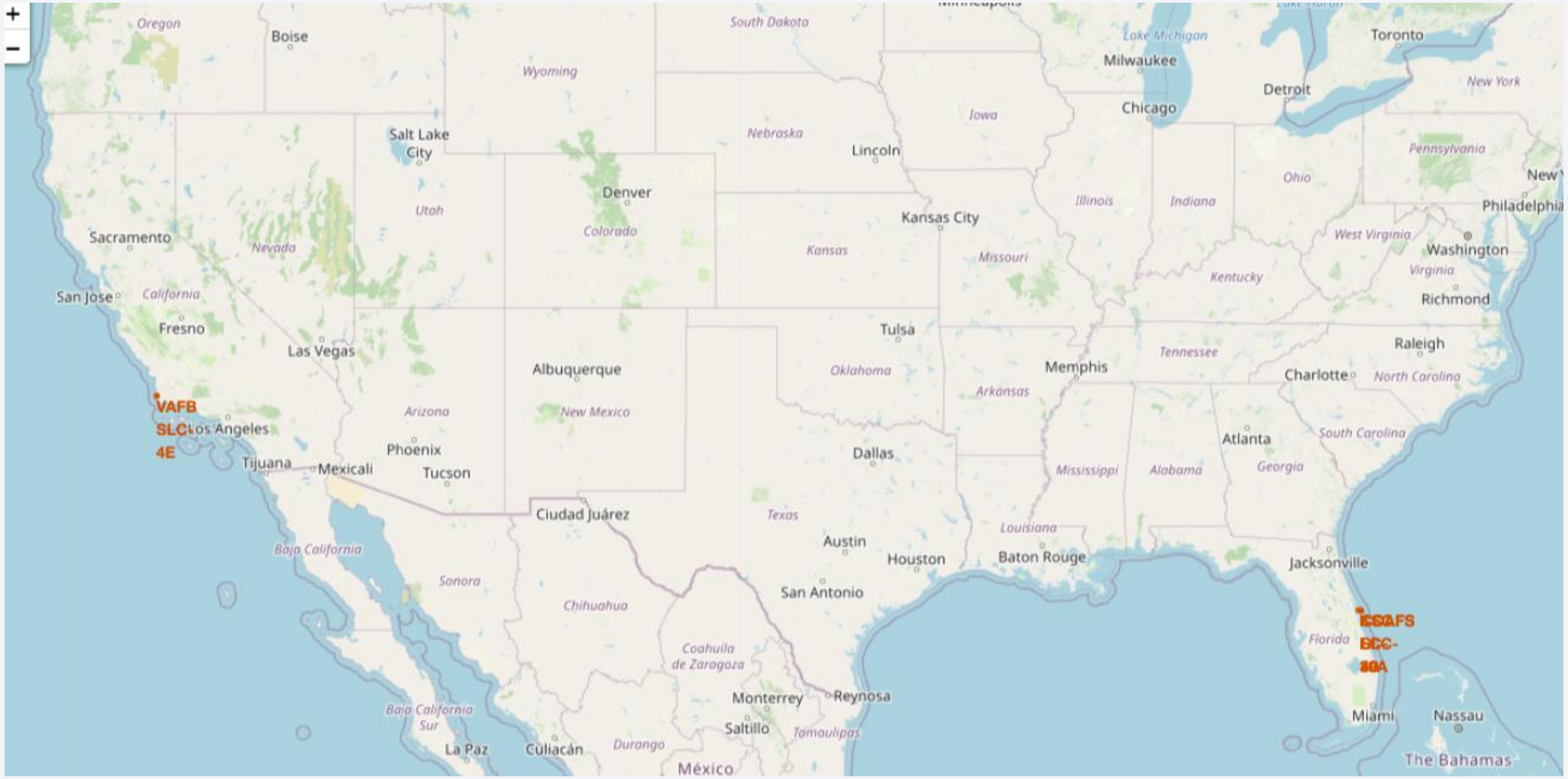
[22]:

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

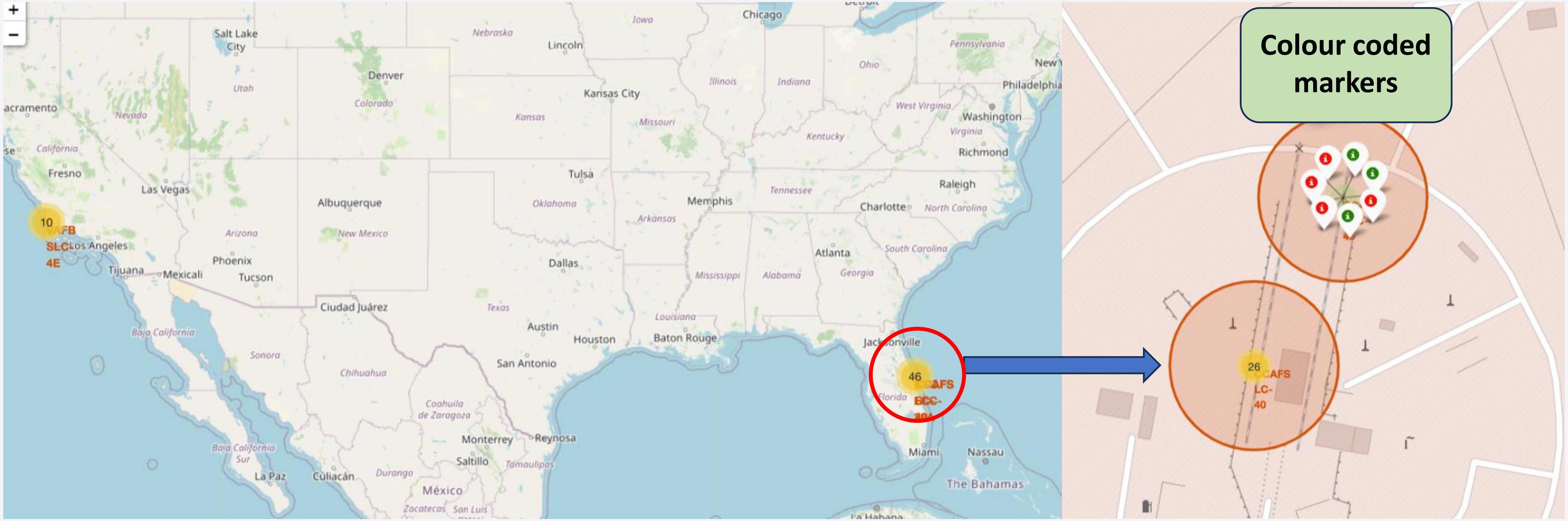
Launch Sites Proximities Analysis

Launch Sites Locations



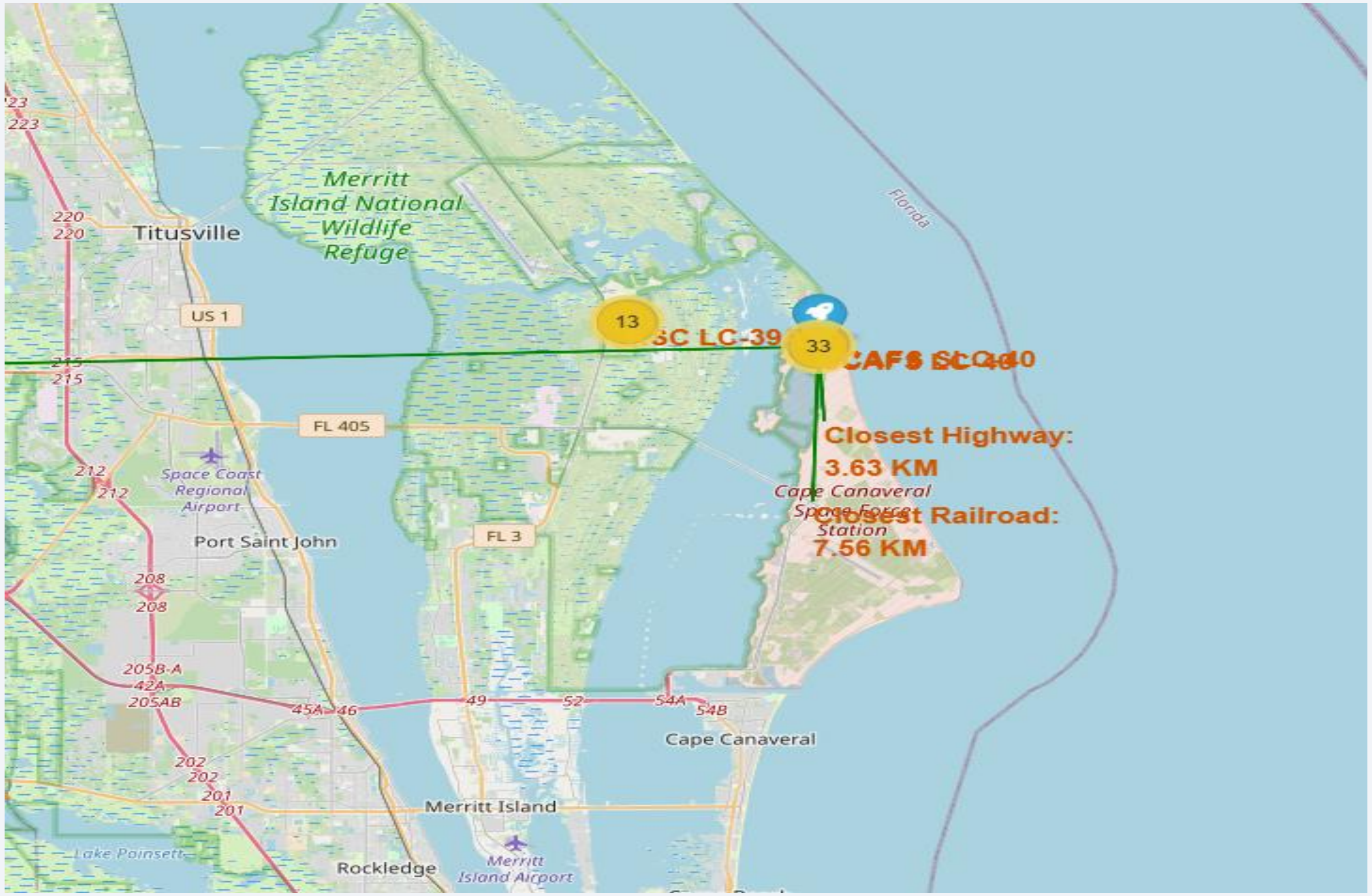
- All launch sites located and titled on the map of the US. Coastal sites are chosen to protect the public in case of launch failure.

Launch Site Outcomes



- Color coded markers showing the launch outcomes of this specific site.

<Folium Map Screenshot 3>



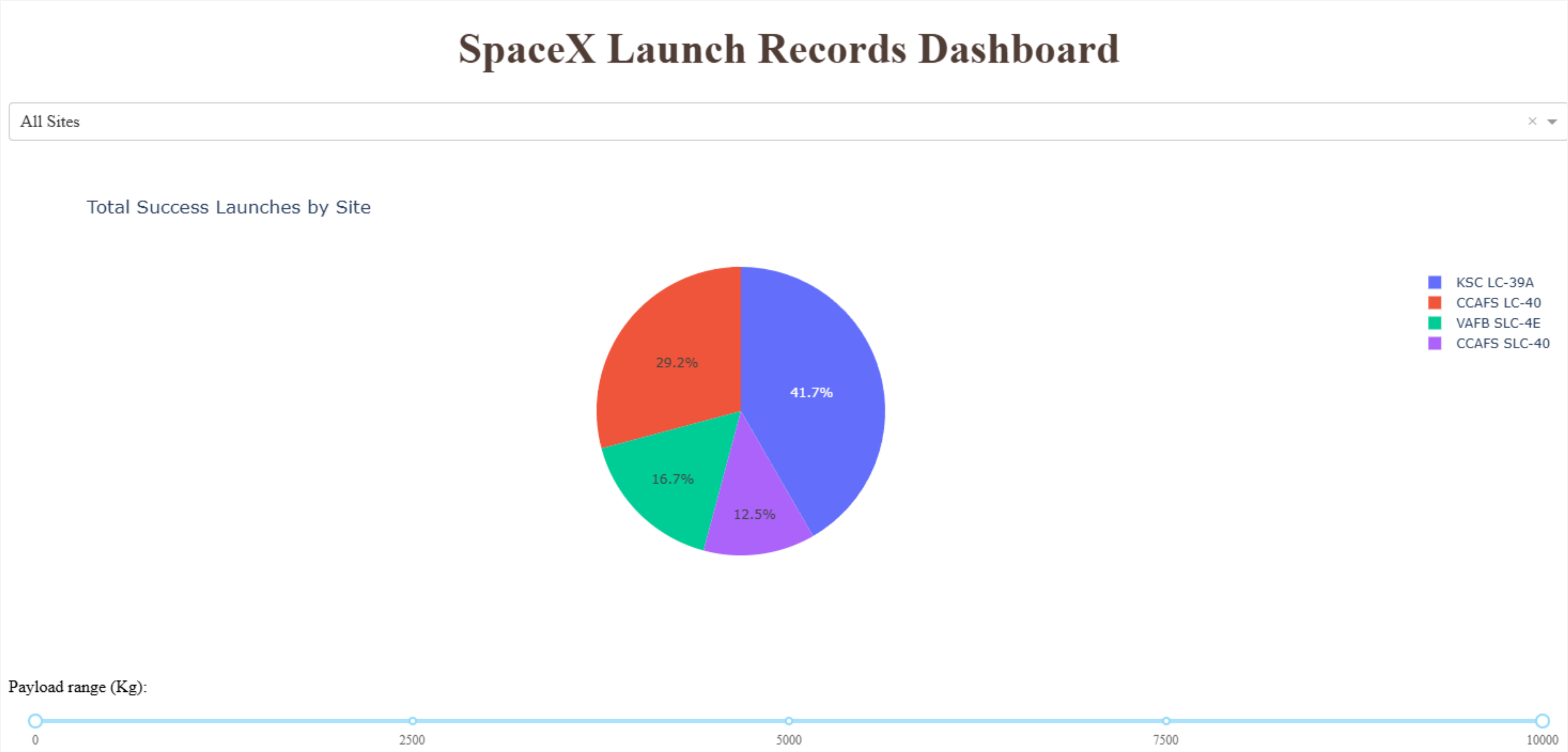
Close proximity shown to:

- Railway
- Highway
- City

Section 4

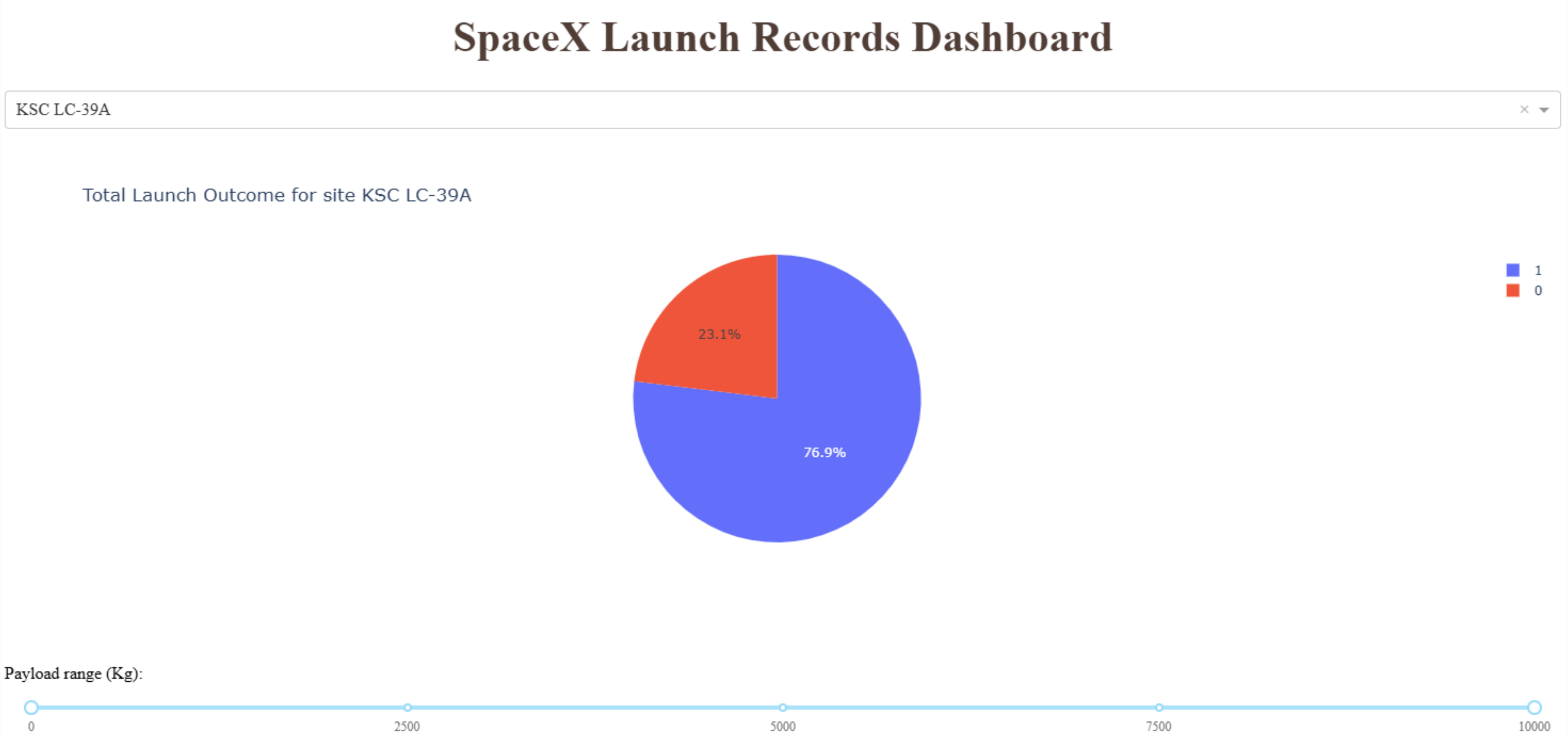
Build a Dashboard with Plotly Dash

SpaceX Launch Success Rates across all Sites



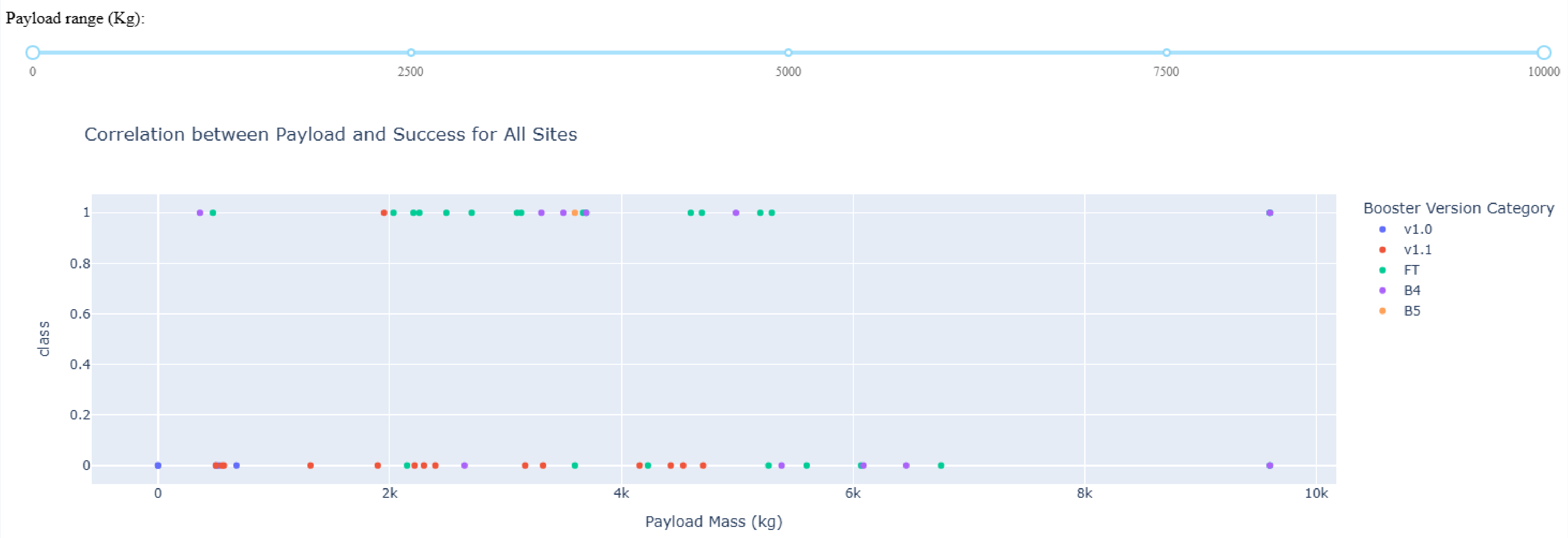
- Shows that the success is highest for KSC LC-39A, followed by CCAFS LC-40

Highest Launch Success Rate (KSC LC-39A)



- KSC LC-39A has the highest success : failure ration (76.9% success rate)

Payload Mass and Launch Success by Booster Version

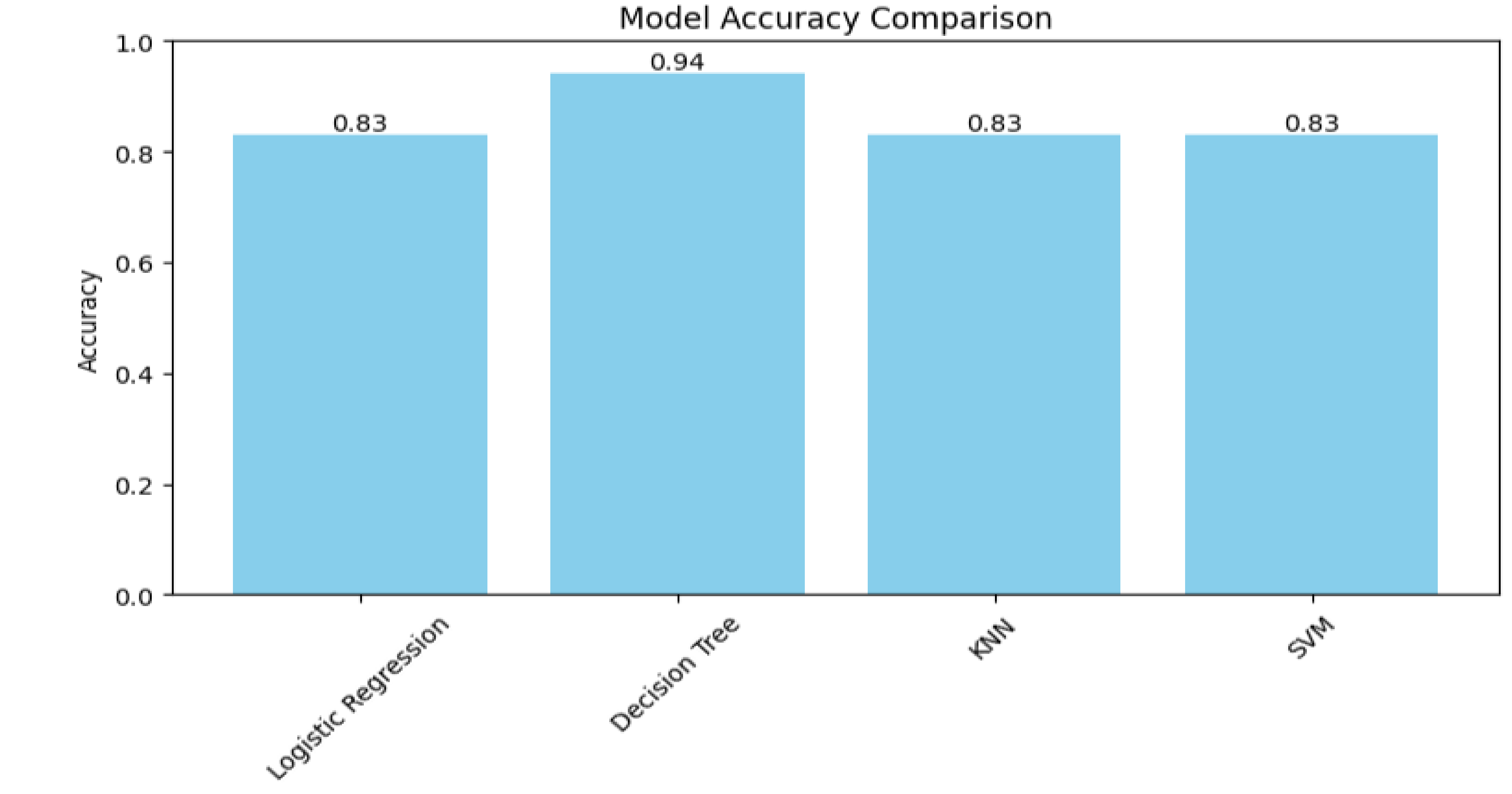


- All sites shown with payload slider and scatter plot by Booster Version

Section 5

Predictive Analysis (Classification)

Classification Accuracy

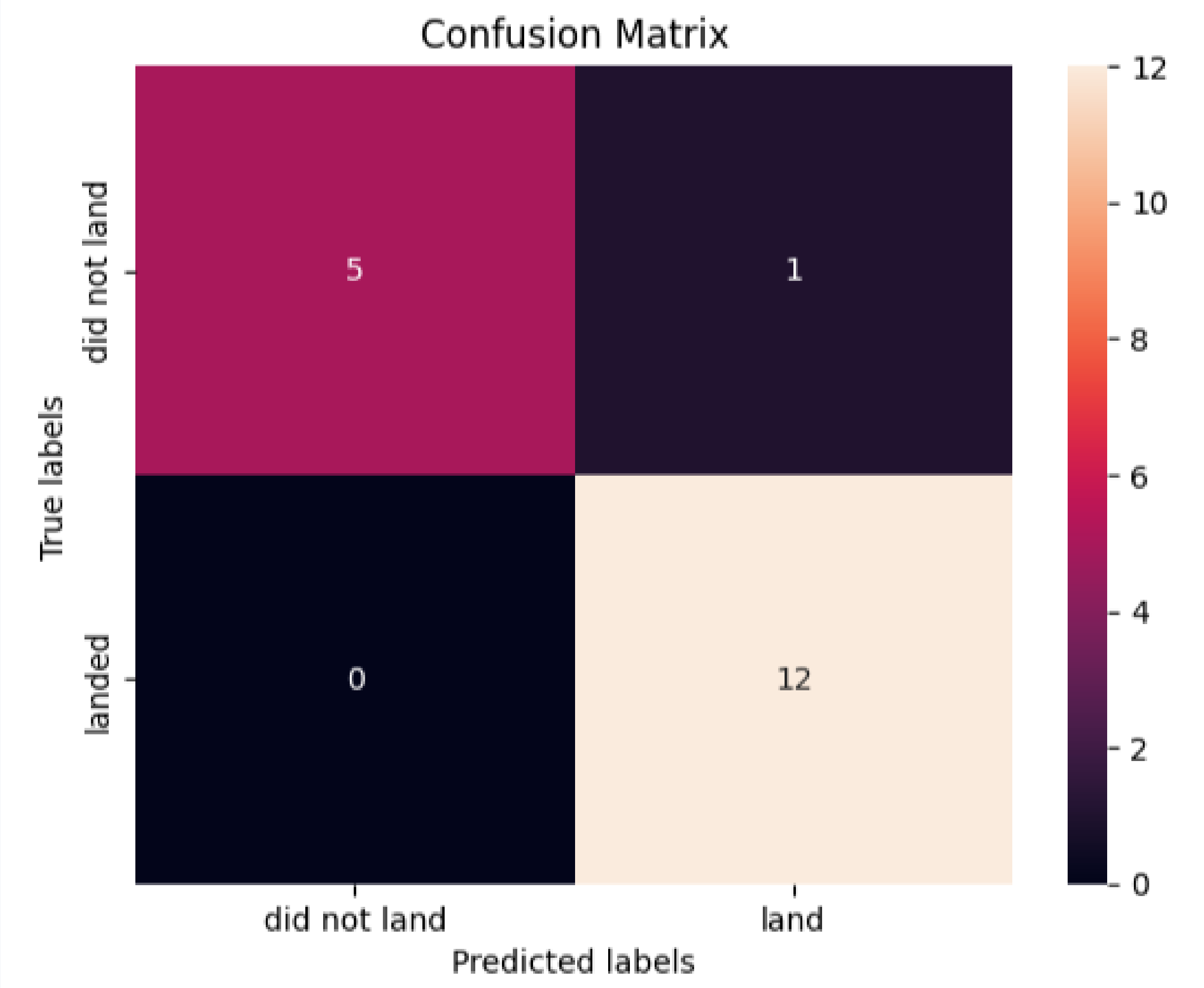



```
Best model: Decision Tree with accuracy 0.94
```

Confusion Matrix: Decision Tree

Confusion Matrix Outputs:

- 12 True positive
- 3 True negative
- 3 False positive
- 0 False Negative



Conclusions

- The highest performance for this dataset was shown by the Decision Tree algorithm.
- Lighter payloads in launches usually have a greater chance of success than heavier payloads.
- Almost all launch facilities are situated close to the equator and are found near the coast.
- An increasing trend in launch success rates has been observed over time.
- Of all the locations, KSC LC-39A had the highest percentage of successful missions.
- Missions aimed at orbits like ES-L1, GEO, HEO, and SSO accomplished an impeccable success rate of 100%.

Thank you!