IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

MOHAMED ZIYATI
30 August 2024

Outline

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

Executive Summary

- Summary of methodologies :
  - Data collection, Data Wrangling, Exploratory Data Analysis, Visualization, Dashboard, Machine learning prediction using Models .
- Summary of all results :
  - Decision Tree is the best Machine Learning model in our case, because he has the maximum accuracy.

Introduction

- Our goal is to predict whether the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website, costing $62 million, while other providers charge upwards of $165 million per launch. Much of SpaceX's cost savings come from their ability to reuse the first stage.
- Therefore, if we can predict whether the first stage will land successfully, we can estimate the cost of a launch. This information could be valuable for alternate companies looking to compete with SpaceX for rocket launches. To achieve our goal, we will use various data science tools for analysis and visualization.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - We used SPACEX API and WebScraping
- Perform data wrangling
  - Create, Calculate the number and occurrence …
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- We collected Data using a GET Request to the SpaceX API
- We have Filtering the DataFrame to only include `Falcon 9` launches

```
response = requests.get("https://api.spacexdata.com/v4
```

Data Collection – SpaceX API

- We collected Data using a GET Request to the SpaceX API

- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

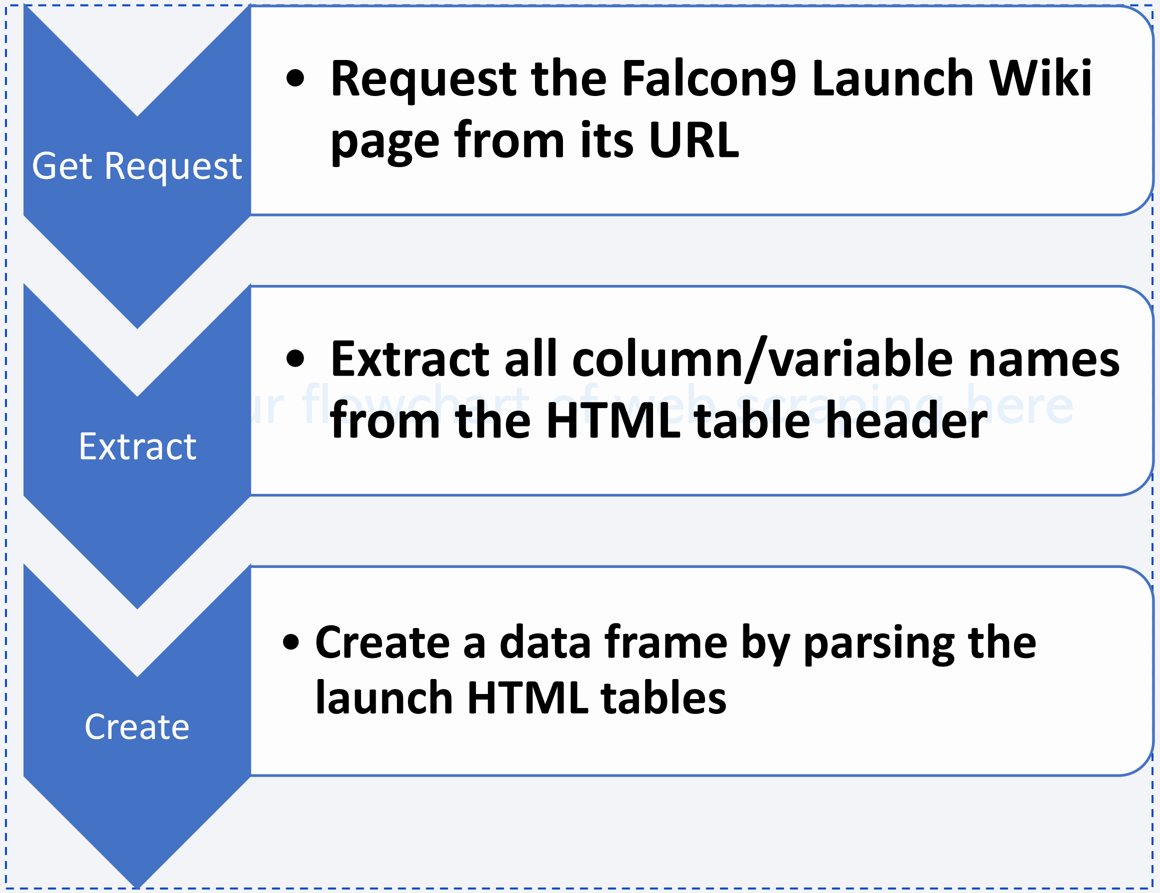
Get Request
- **Request and parse the SpaceX launch data using the GET request**

Filter
- **Filter the DataFrame to only include Falcon 9 launches**

Data Wrangling
- **Some Data Wrangling**

Data Collection - Scraping

- We used Web Scraping to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Data Wrangling were processed as follows:
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurence of mission outcome of the orbits
  - Create a landing outcome label from Outcome column
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- We visualized the relationship between Flight Number and Launch Site with a **scatter point chart** with the class 0 and 1 in the hue .
- We visualized the relationship between Payload Mass and Launch Site with a **scatter point chart**.
- We visualized the relationship between success rate of each orbit type with a **bar chart**.
- We visualized the relationship between FlightNumber and Orbit type with a **scatter point chart**.
- We visualized the relationship between Payload Mass and Orbit type with a **scatter point chart**.
- We visualized the launch success yearly trend with a **line chart**.

EDA with Data Visualization

- Why we used ? ,
  - Scatter plot : **To show the Relationship or correlation between two Numerical variables.**
  - Bart chart : **To compare different categories of data.**
  - Line chart : **To show Trends over time or Continuous data.**
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/edadataviz.ipynb

EDA with SQL

- Summary of SQL Queries :

1. SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE : **Displaying the Names of the Unique Launch Sites in the Space Mission**
2. SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5 : **Display 5 records where launch sites begin with the string 'CCA'**
3. SELECT SUM("PAYLOAD_MASS__KG_") AS "Total_Payload_Mass" FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA (CRS)' : **Display the total payload mass carried by boosters launched by NASA (CRS)**
4. SELECT AVG("PAYLOAD_MASS__KG_") AS "AVERAGE_payload" FROM SPACEXTABLE WHERE "Booster_Version" LIKE 'F9 v1.1%' : **Display average payload mass carried by booster version F9 v1.1**

EDA with SQL

5. SELECT MIN("Date") AS "First_Successful_Landing_Date“ FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)’ : **List the date when the first succesful landing outcome in ground pad was acheived.**
6. SELECT "Booster_Version“ FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)’ AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000 : **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
7. SELECT "Mission_Outcome", COUNT(*) AS "Total_Count“ FROM SPACEXTABLE GROUP BY "Mission_Outcome“ : **List the total number of successful and failure mission outcomes**
8. SELECT "Booster_Version“ FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE) : **List the names of the booster_versions which have carried the maximum payload mass.**

EDA with SQL

9. SELECT CASE WHEN substr("Date", 6, 2) = '01' THEN 'January' WHEN substr("Date", 6, 2) = '02' THEN 'February' .... WHEN substr("Date", 6, 2) = '11' THEN 'November' WHEN substr("Date", 6, 2) = '12' THEN 'December' END AS "Month_Name", "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr("Date", 1, 4) = '2015' : **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.**

10. 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 : **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.**

- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

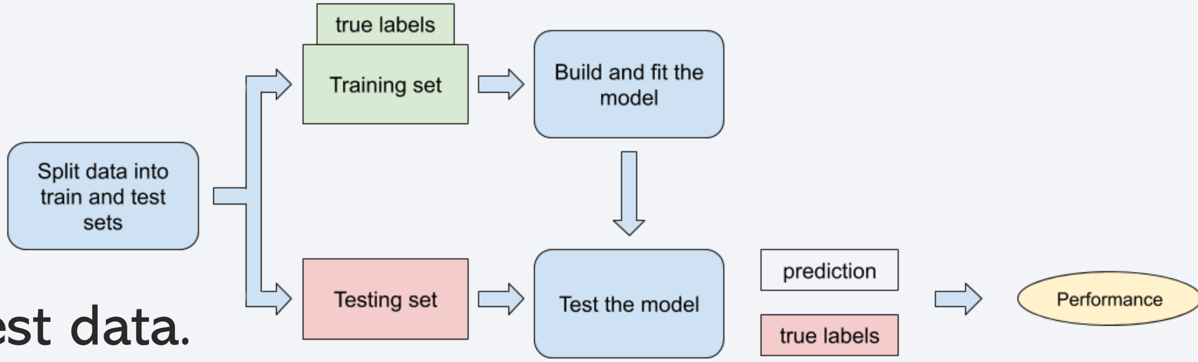
Build an Interactive Map with Folium

- We have added a markers, icons, circles to highlight our Area .
  - We have created a blue circle at NASA Johnson Space Center's with a icon showing its name
  - We have created a markers and an icon for Launch Site CCAFS LC-40
  - We have created a markers and an icon for Launch Site CCAFS SLC-40
  - We have created a markers and an icon for Launch Site KSC LC-39A
  - We have created a markers and an icon for Launch Site VAFB SLC-4E
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/lab_jupyter_launch_site_location%2C%20notC.ipynb

Build a Dashboard with Plotly Dash

- We have added to our dashboard a Pie chart to show Total Success Launches by Sites
- We have added a slider to select Payload range
- We have added a Scatter chart to show the correlation between payload and launch success
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/spacex_dash_app.py

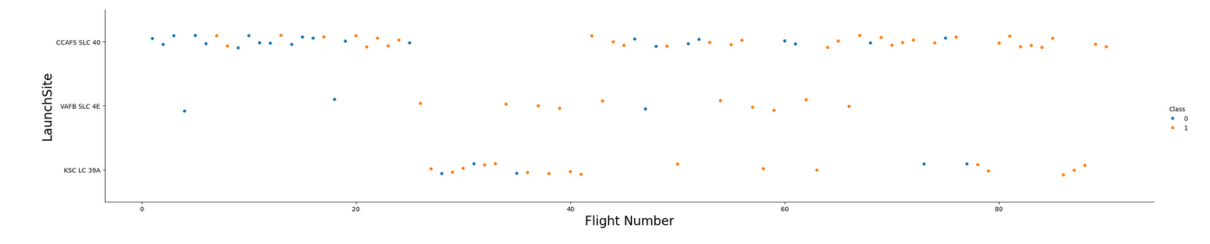
Predictive Analysis (Classification)



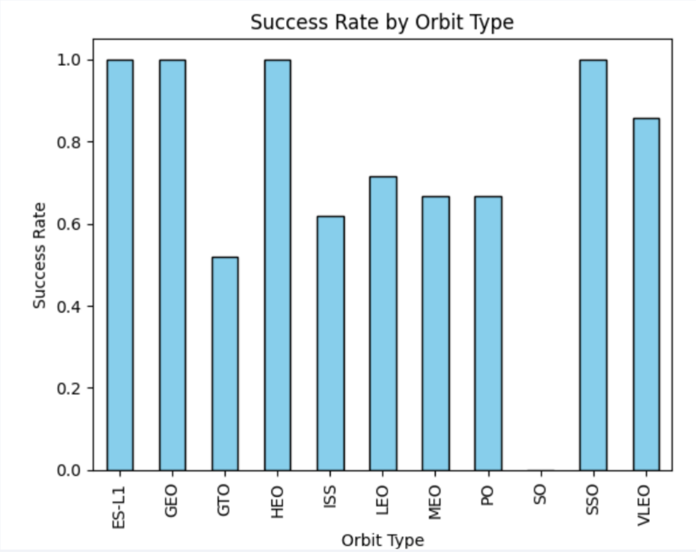
- We assign the Class column to the variable Y.
- We split our data X and Y into training and test data.
- Create a Logistic regression, SVM, Decision tree classifier objects, then create a GridSearchCV objects and fit it, to find the best parameters of the Models .
- Calculate the accuracy of each Methods
- You need present your model development process using key phrases and flowchart
- https://github.com/MOHAMED-ZIYATI/Data-Science-Capstone-NoteBooks/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots



- Predictive analysis : Logistic Regression, SVM, Decision Tree



Best parameters :

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

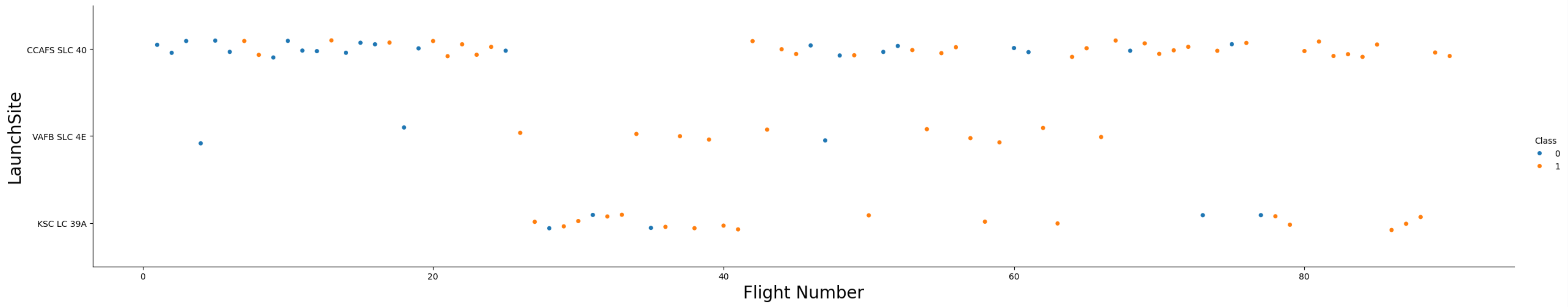
```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

```
tuned hyperparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 16, 'max_features': 'sqrt', 'min_samples_leaf
': 4, 'min_samples_split': 5, 'splitter': 'random'}
accuracy : 0.8767857142857143
```

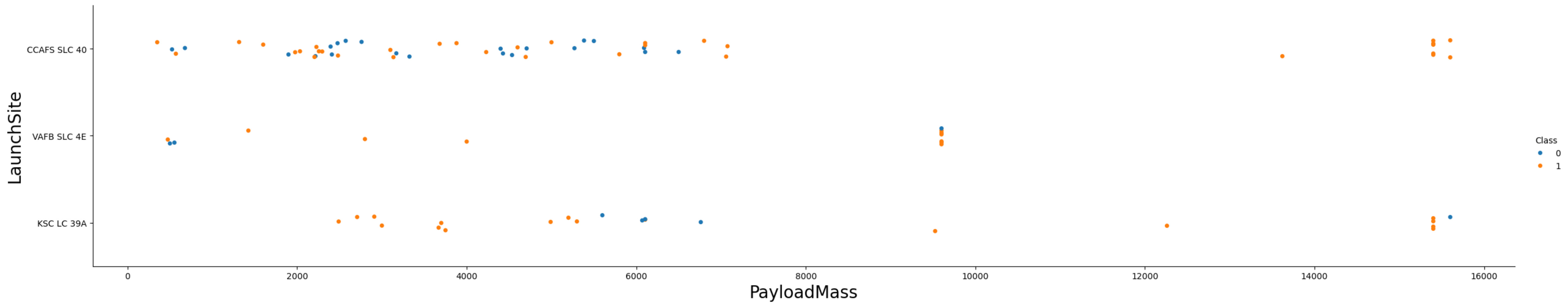
Section 2

Insights drawn from EDA

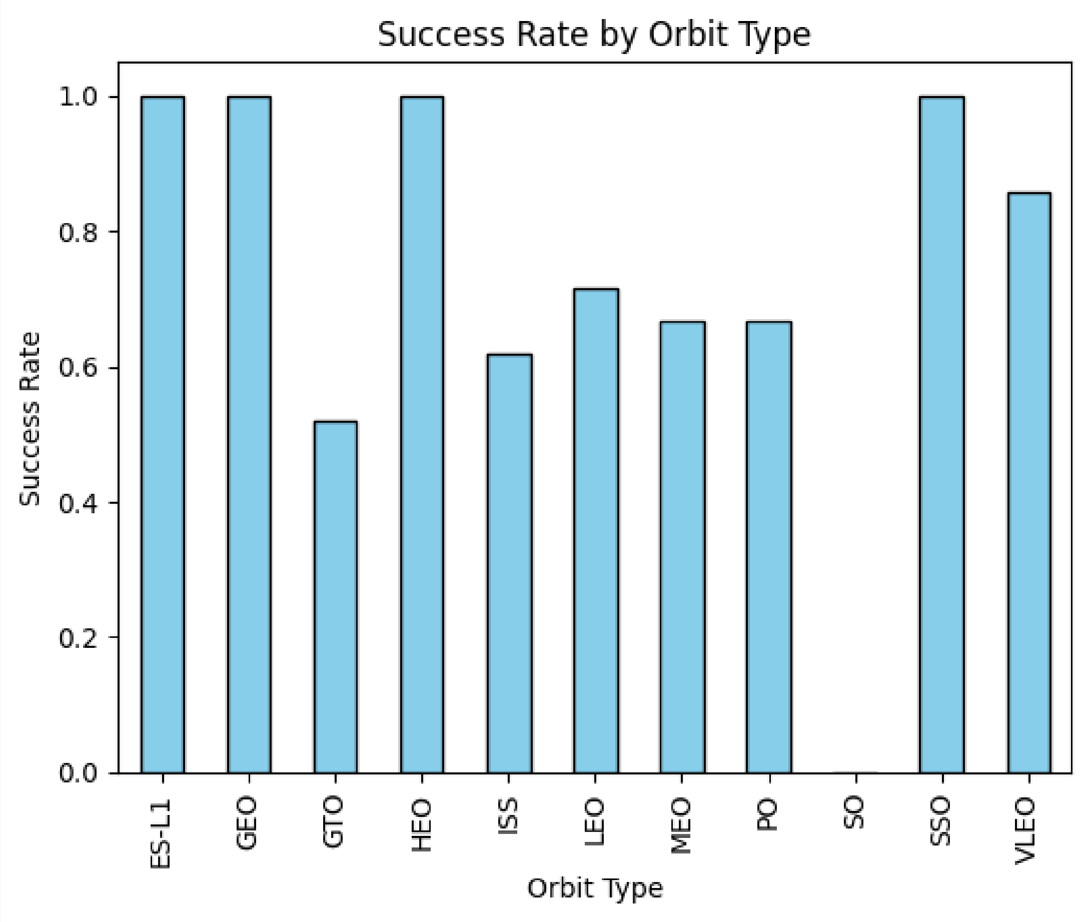
Flight Number vs. Launch Site



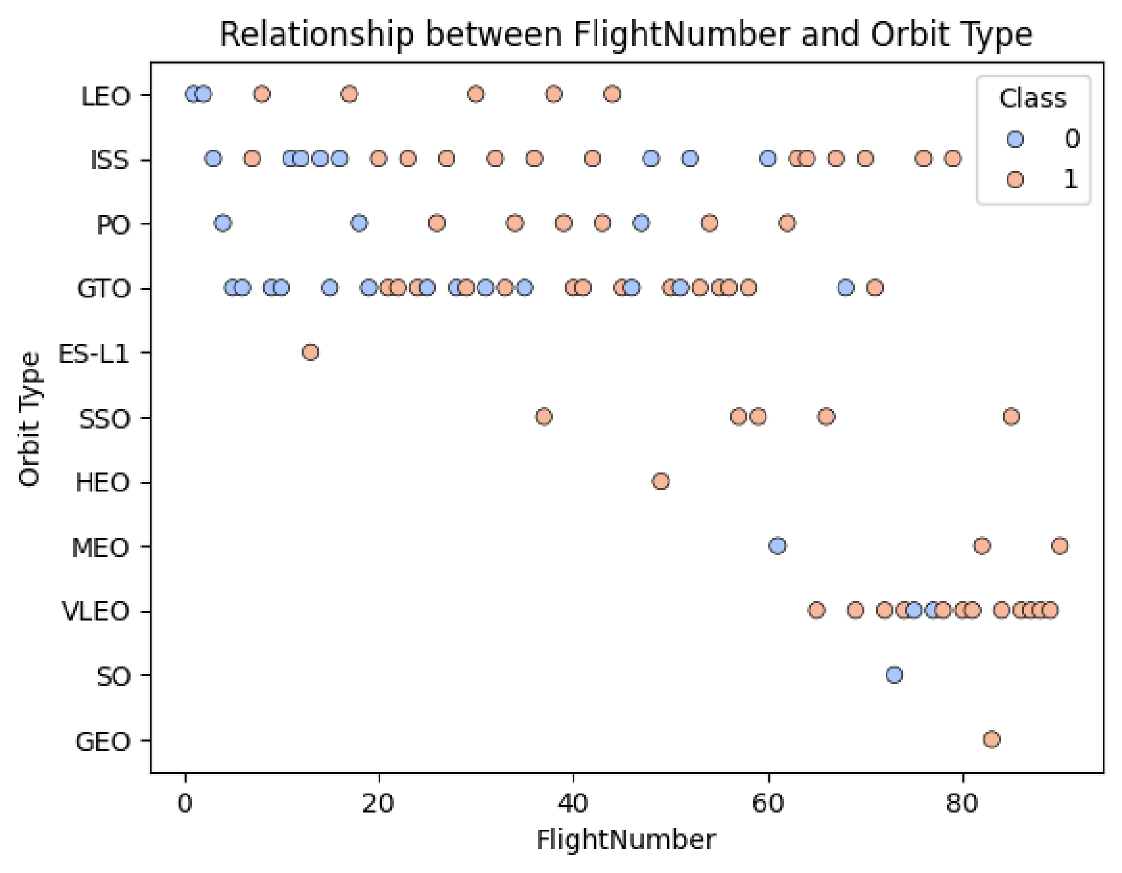
Payload vs. Launch Site



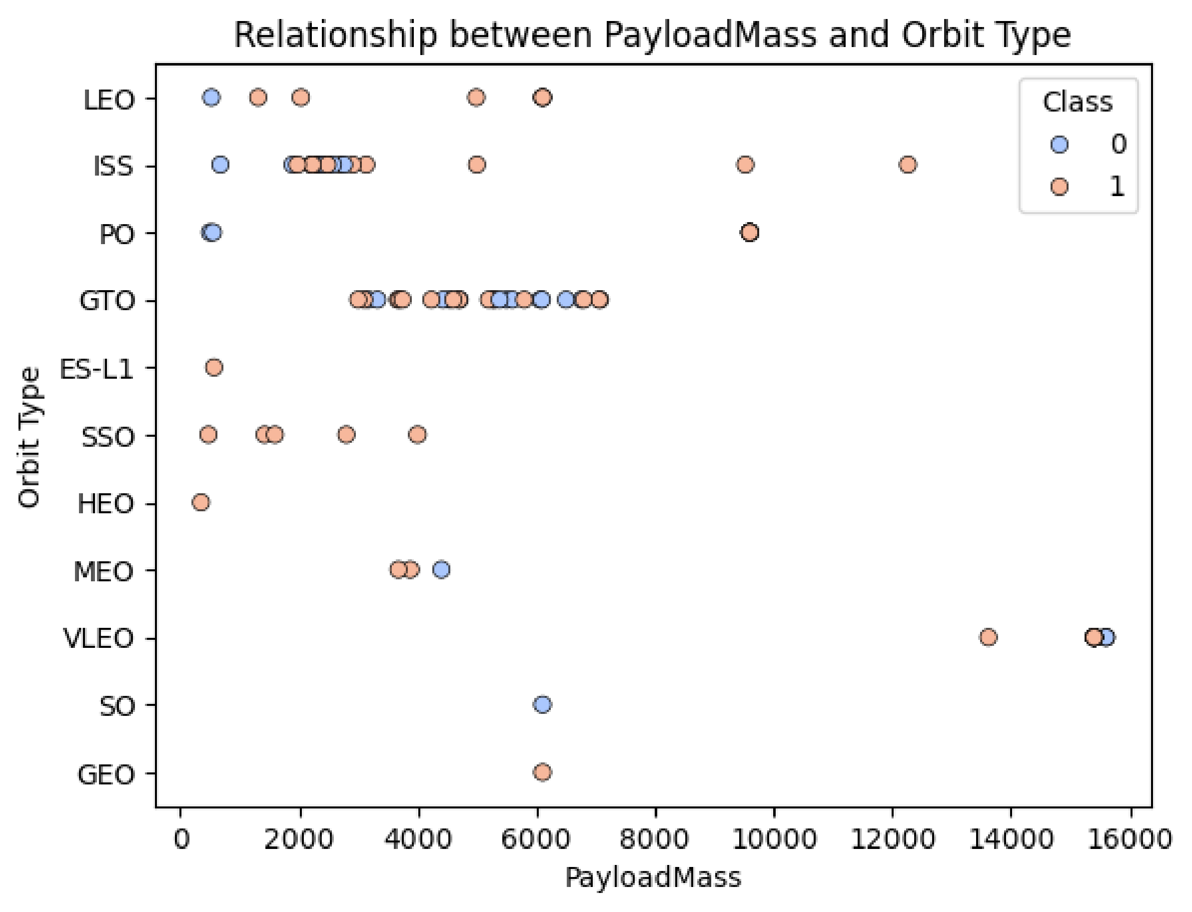
Success Rate vs. Orbit Type



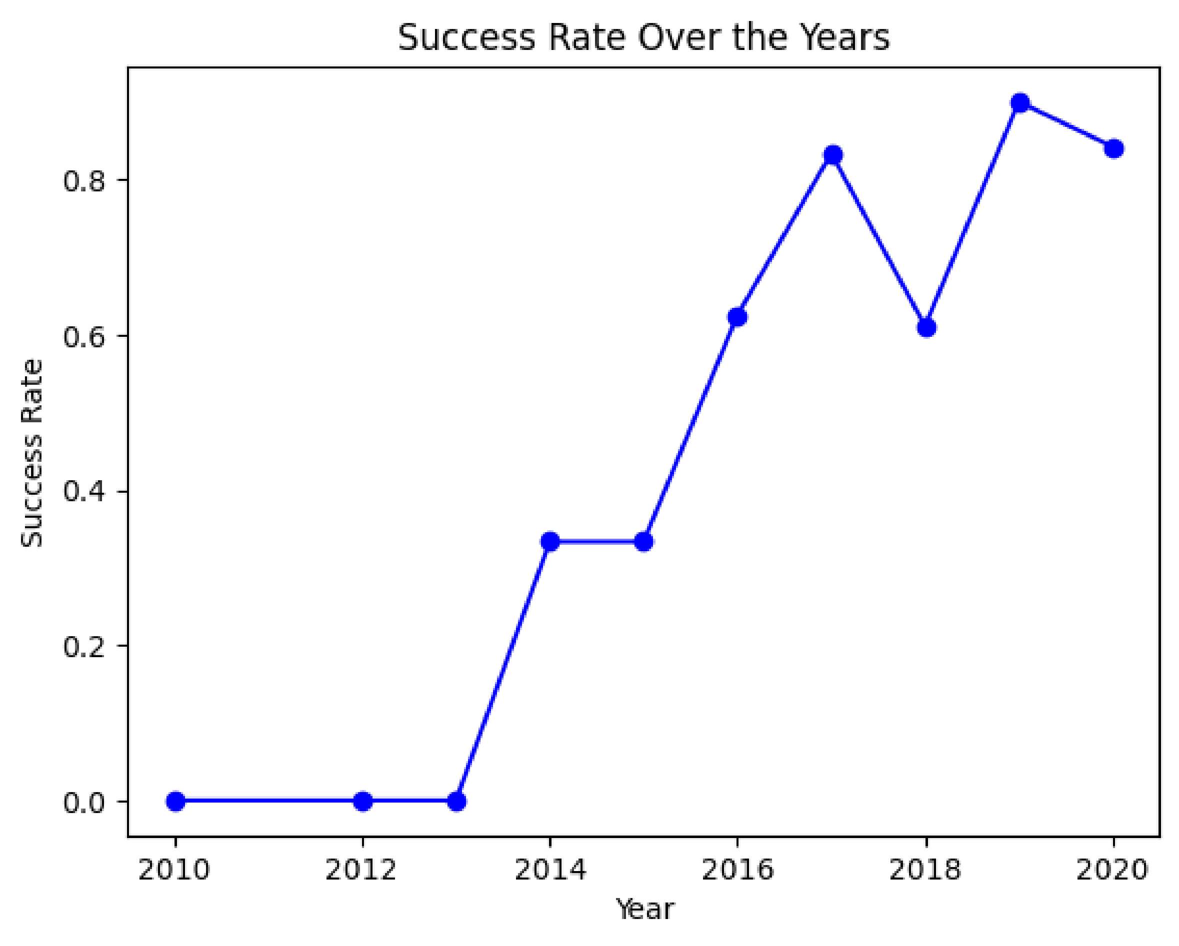
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

AVERAGE_payload
2534.6666666666665

First Successful Ground Landing Date

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

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

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

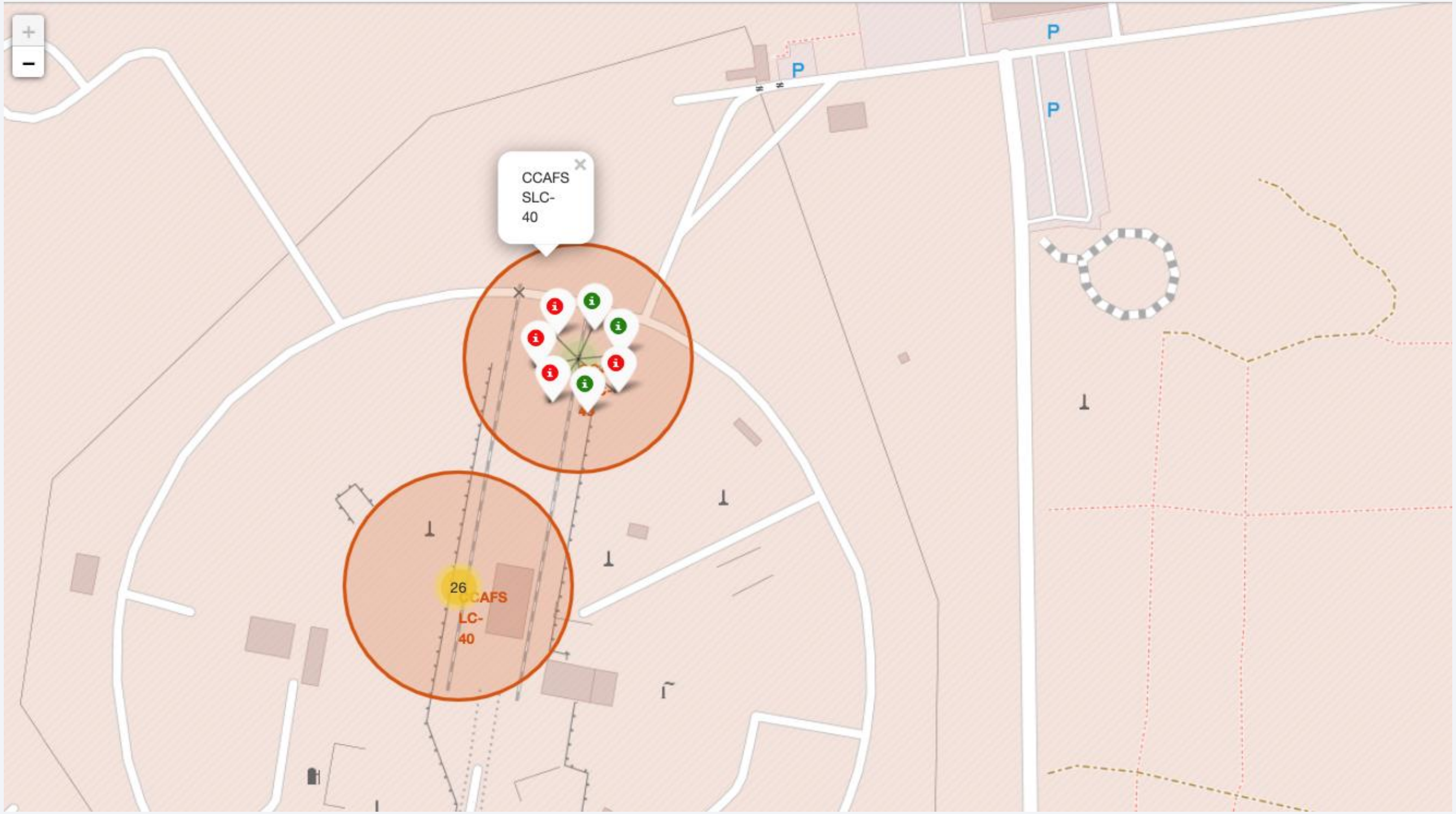
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

All launch Sites

Map with Markers

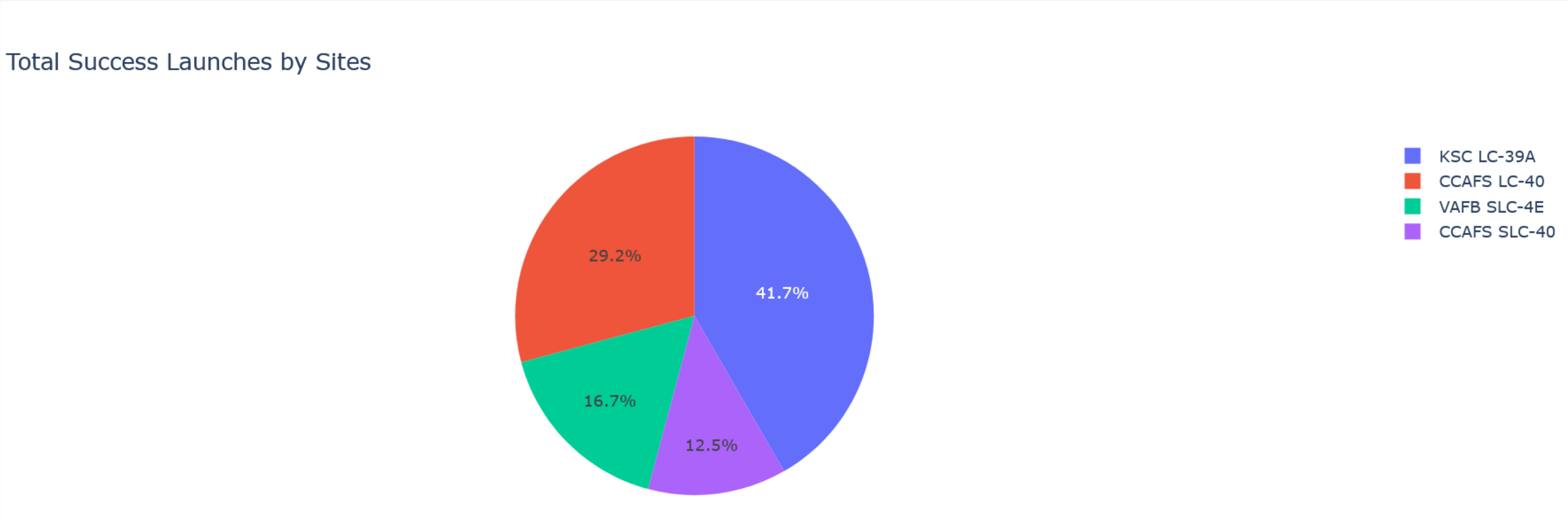


Map with distance line

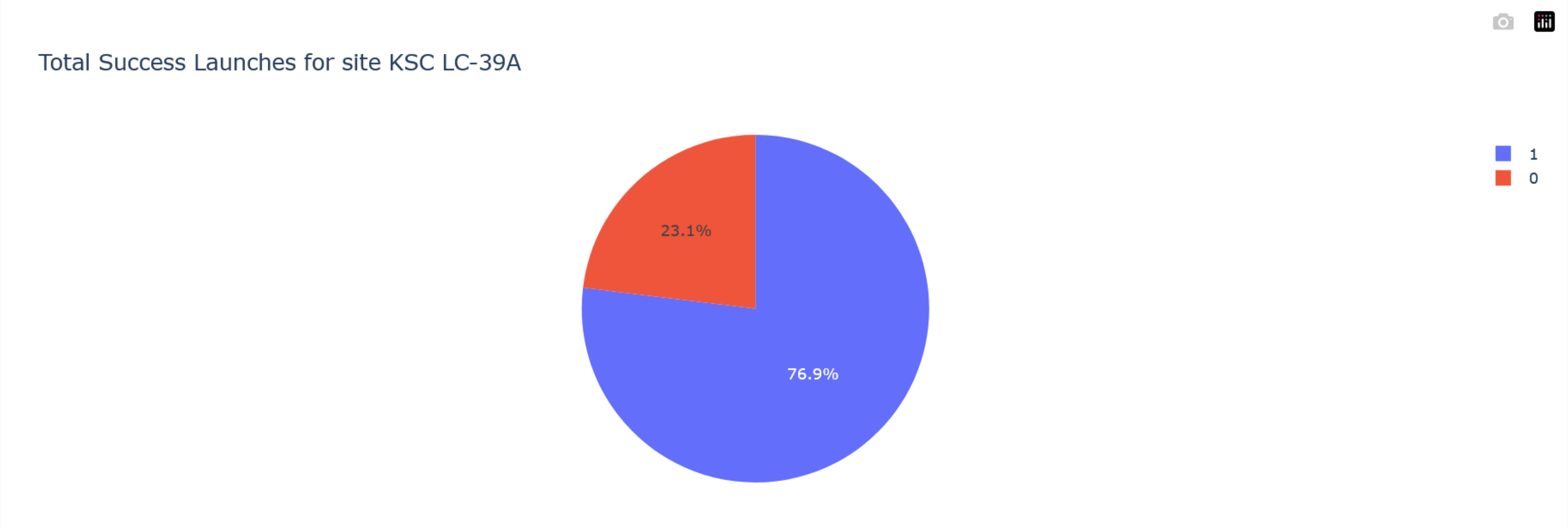
Section 4

Build a Dashboard with Plotly Dash

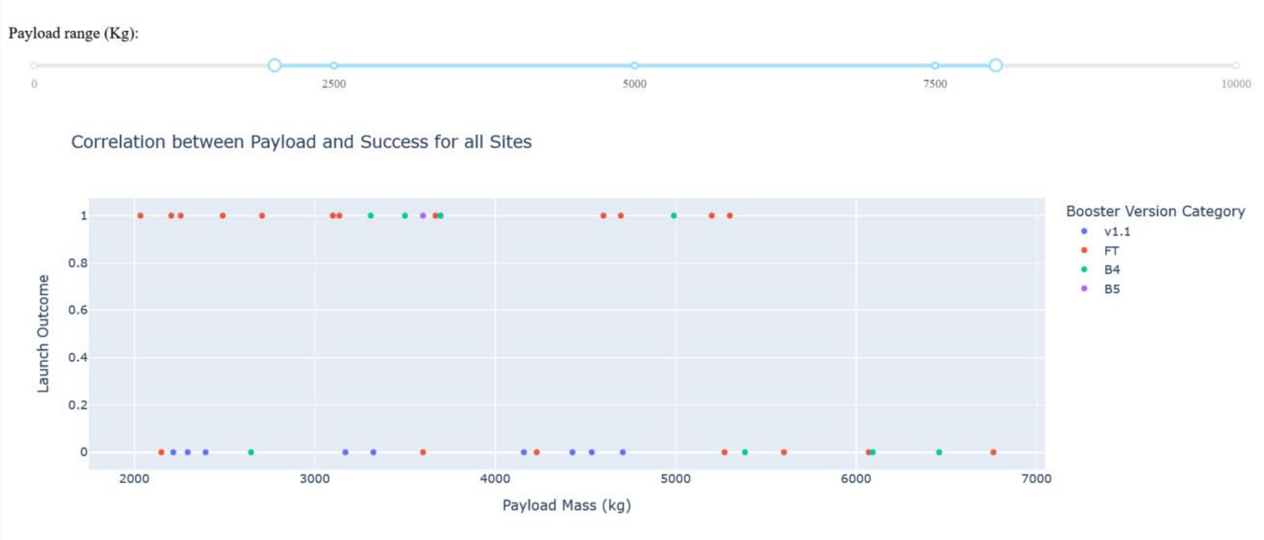
Total Success Launches by Sites



Total Success Launches for Site KSC LC-39A



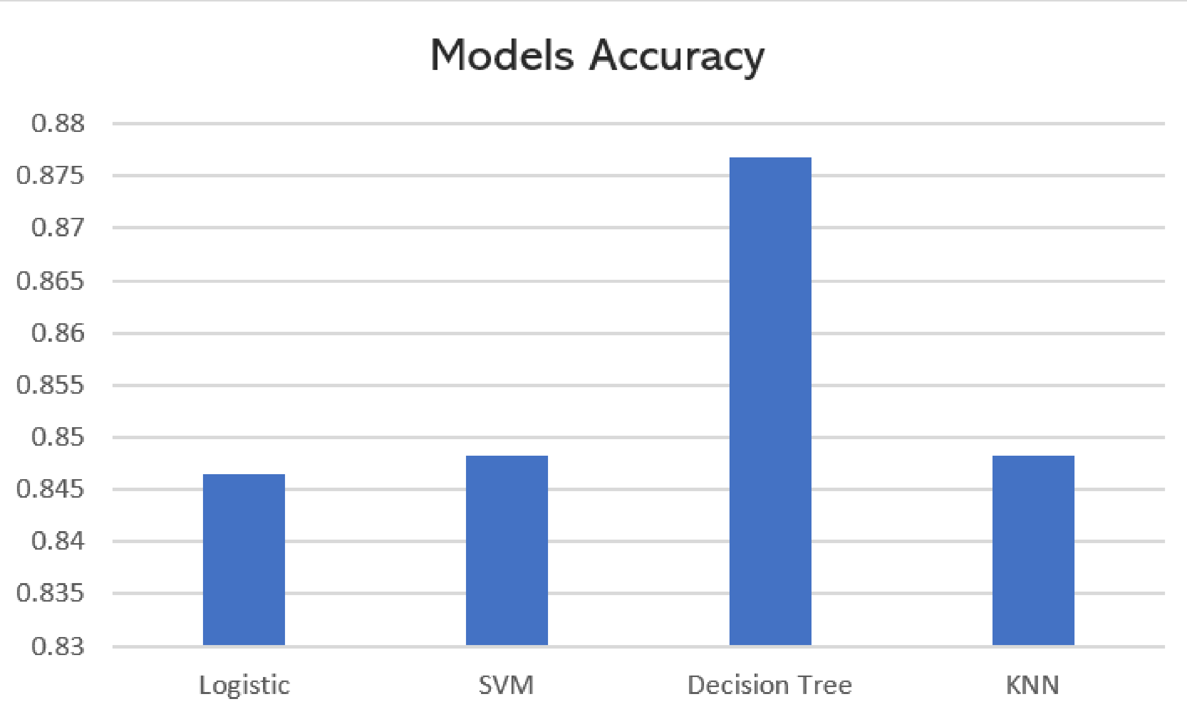
Scatter plot for different payload selected in the Range slider



Section 5

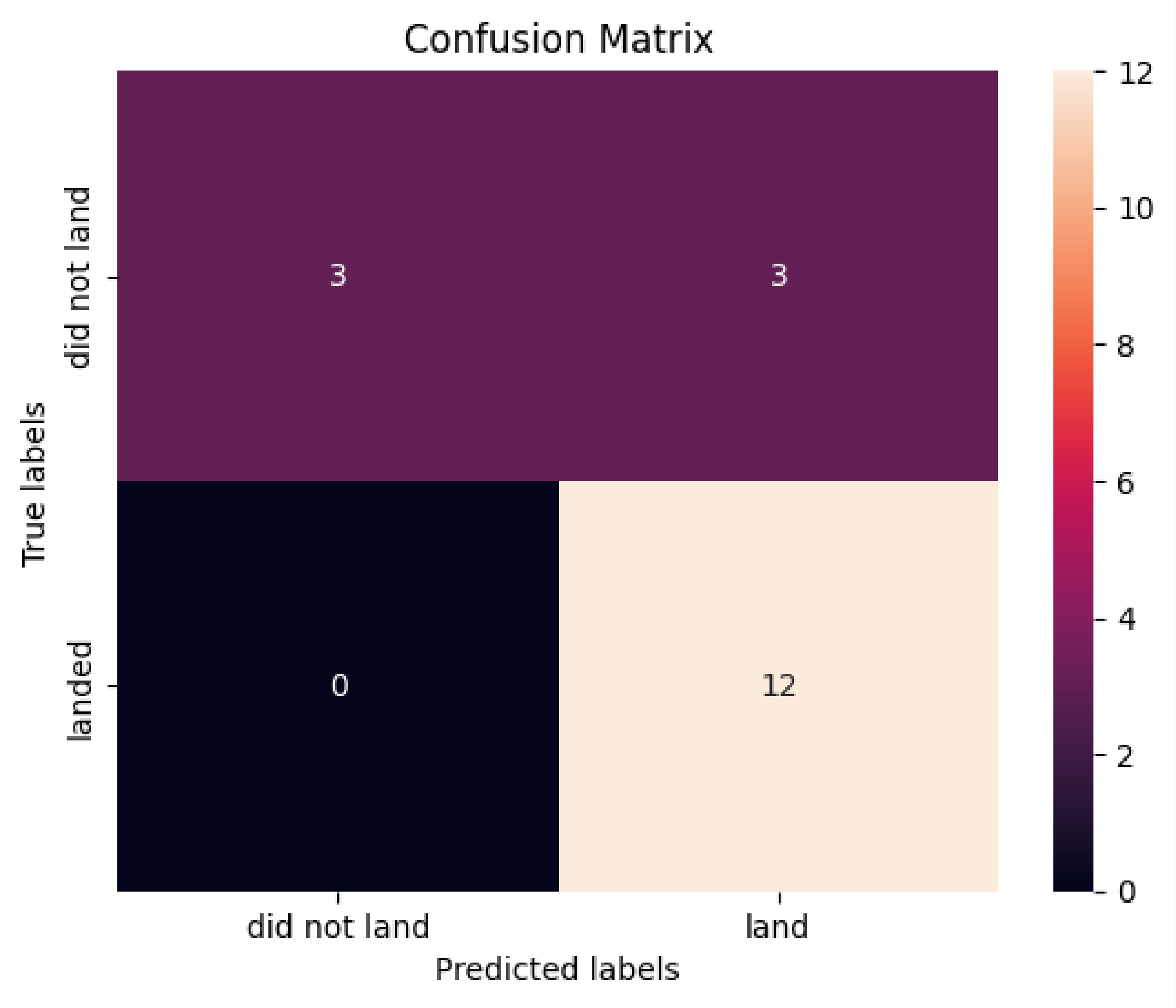
Predictive Analysis (Classification)

Classification Accuracy



- Decision Tree HAS THE BEST ACCURACY

Confusion Matrix



This confusion matrix indicates that the model has a good accuracy rate, especially in predicting successful landings, as evidenced by the zero false negatives.

Conclusions

- The orbit Type SSO, ES-L1, GEO, HEO, has the maximum success rate
- Success rate increased since 2015
- F9 B5 Boosters Carried Maximum Payload
- KSC LC-39A is the Best Launches Site to Success
- Decision Tree is the best Model in our case

Thank you!