IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

- Link for Project: https://github.com/Jaskaran-Singh1699/CourseraProject/blob/main/Module4/Machine%20Learning%20Prediction.ipynb
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- In this module, we will compile all of your activities into one place and deliver our data-driven insights to determine if the first stage of Falcon 9 will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The SpaceX dataset was collected from public launch records, telemetry data, and SpaceX reports. It includes launch site locations, mission details, flight outcomes, weather conditions, and rocket specifications. The data was compiled for analysis and machine learning applications. Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Prepare Data, Build Model, Tune Model, Evaluate Model.

Data Collection

- The dataset was collected from multiple sources, including publicly available SpaceX launch records, telemetry data, and official reports. Data on launch sites, including latitude and longitude coordinates, was gathered from SpaceX launch logs and NASA databases. Mission details, such as payload, customer organizations, and mission names, were sourced from SpaceX's official reports and APIs.

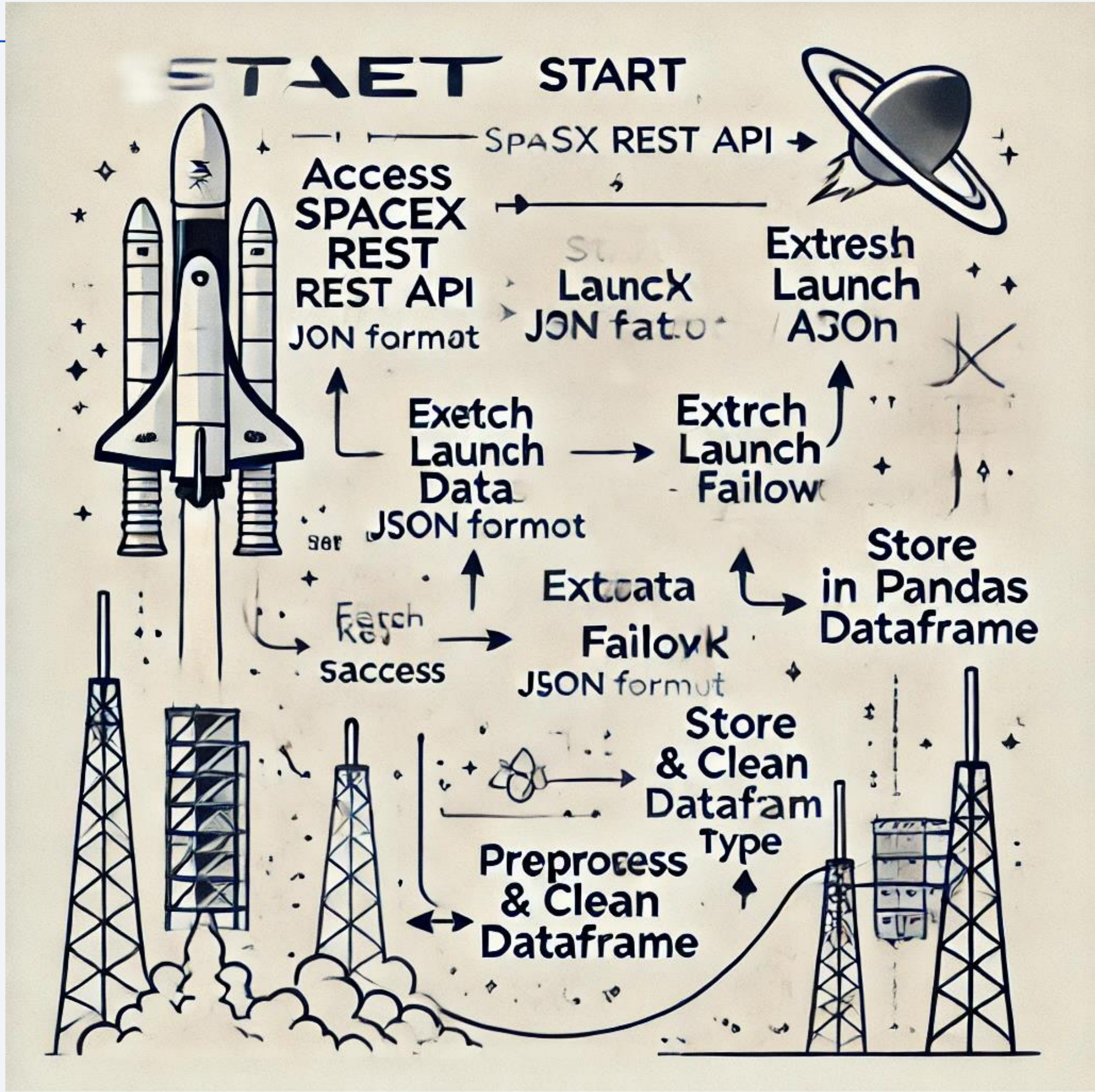
- Flight outcomes, including whether a booster successfully landed or failed, were collected from telemetry data, news reports, and SpaceX event summaries. Weather conditions affecting launches were obtained from meteorological sources such as NASA and NOAA. Rocket specifications, including booster versions, landing types, and technical details, were compiled from SpaceX technical reports and engineering documents.

- The data was then cleaned, structured, and organized for analysis and machine learning applications, such as predicting launch success and visualizing launch patterns.You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

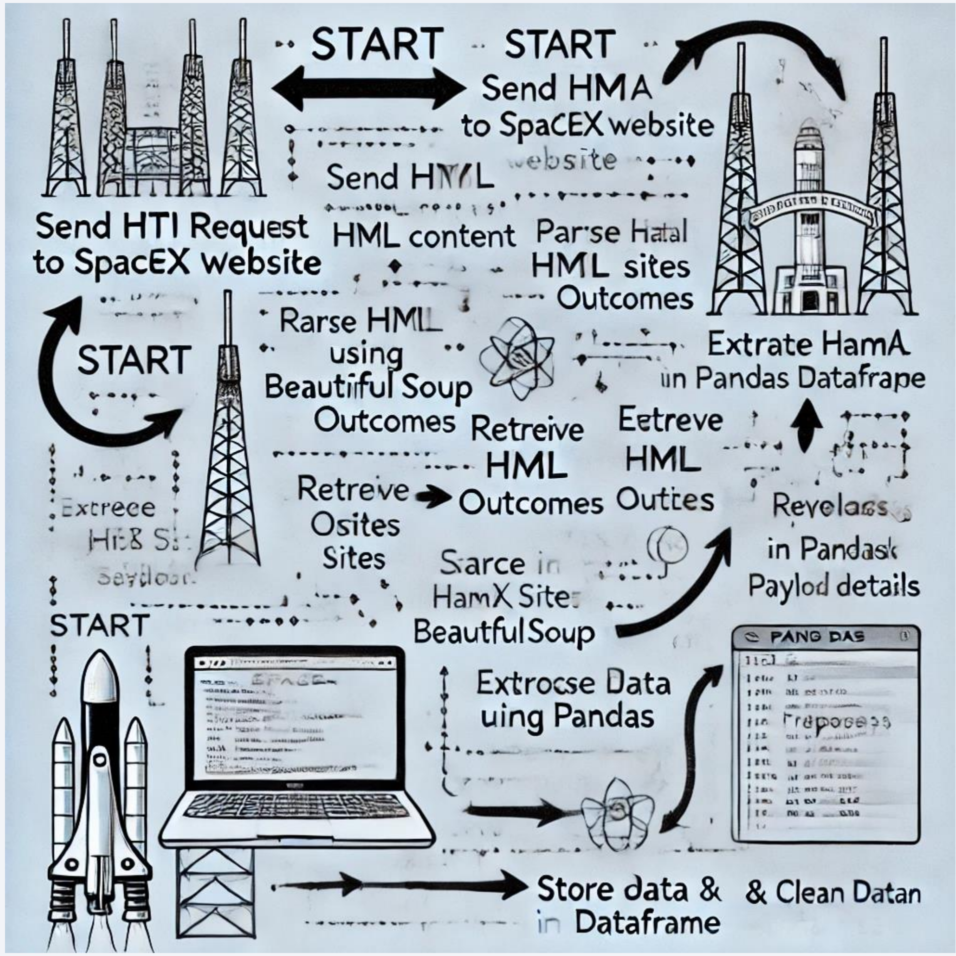
- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- GitHub:
- https://github.com/Jaskaran-Singh1699/CourseraProject/blob/main/Module1/Data%20Collection.ipynb



Data Collection - Scraping

- Presenting web scraping process using key phrases and flowcharts
- Github: https://github.com/Jaskaran-Singh1699/CourseraProject/blob/main/Module1/Data%20Wrangling.ipynb



Data Wrangling

- The collected SpaceX data was processed through several key steps to ensure accuracy and usability for analysis: LINK
- 1. Data Cleaning – Missing values were handled, duplicate records were removed, and incorrect data points were corrected.
- 2. Data Transformation – JSON responses from the SpaceX API were converted into a structured Pandas DataFrame. Web-scraped data was parsed using BeautifulSoup.
- 3. Feature Engineering – New columns were created, such as success indicators, launch site coordinates, and mission details, for better analysis.
- 4. Data Standardization – Numerical values, such as payload mass and launch time, were normalized or scaled to improve model performance.
- 5. Data Integration – Information from multiple sources, including API data, weather conditions, and mission reports, was merged for comprehensive insights.
- 6. Exploratory Data Analysis (EDA) – Visualizations and statistical summaries were generated to understand patterns and trends before applying machine learning models.

This processing ensured that the dataset was clean, structured, and ready for predictive modeling and visualization.

EDA with Data Visualization

To understand patterns and trends in the SpaceX dataset, several charts were plotted: LINK

- 1. Bar Charts – Used to compare the number of launches across different launch sites and analyze the success rate per site.
- 2. Pie Charts – Showed the proportion of successful vs. failed launches, helping to visualize overall mission success rates.
- 3. Scatter Plots – Plotted payload mass vs. launch success to observe any correlation between payload weight and mission outcome.
- 4. Line Charts – Displayed launch trends over time, highlighting the frequency of launches and their success rates.
- 5. Heatmaps – Used to visualize correlations between different numerical features, such as payload mass, weather conditions, and success rates.
- 6. Folium Maps – Mapped launch site locations to provide a geographical representation of where Falcon 9 missions were launched.
- These visualizations provided insights into launch performance, influencing factors, and geographical distribution, helping in predictive modeling and decision-making.

EDA with SQL

- LINK: https://github.com/Jaskaran-Singh1699/CourseraProject/blob/main/Module2/EDA_SQl.ipynb
- Retrieve all records – SELECT * FROM spacex_launches;
- Count total launches – SELECT COUNT(*) FROM spacex_launches;
- Filter successful launches – SELECT * FROM spacex_launches WHERE mission_outcome = 'Success';
- Find launch count per site – SELECT launch_site, COUNT(*) FROM spacex_launches GROUP BY launch_site;
- Find the site with the most launches – SELECT launch_site, COUNT(*) FROM spacex_launches GROUP BY launch_site ORDER BY COUNT(*) DESC LIMIT 1;
- Find the average payload for successful launches – SELECT AVG(payload_mass) FROM spacex_launches WHERE mission_outcome = 'Success';
- Retrieve launches within a specific date range – SELECT * FROM spacex_launches WHERE launch_date BETWEEN 'YYYY-MM-DD' AND 'YYYY-MM-DD';
- Find missions with the highest payload – SELECT * FROM spacex_launches ORDER BY payload_mass DESC LIMIT 5;
- Identify the most commonly used booster version – SELECT booster_version, COUNT(*) FROM spacex_launches GROUP BY booster_version ORDER BY COUNT(*) DESC LIMIT 1;
- Calculate success rate per launch site – SELECT launch_site, COUNT(*) AS total, SUM(CASE WHEN mission_outcome = 'Success' THEN 1 ELSE 0 END) * 100.0 / COUNT(*) AS success_rate FROM spacex_launches GROUP BY launch_site;

Predictive Analysis (Classification)

- LINK: https://github.com/Jaskaran-Singh1699/CourseraProject/blob/main/Module4/Machine%20Learning%20Prediction.ipynb
- Built Models – Created multiple classification models (Logistic Regression, SVM, Decision Tree) using the processed dataset.
- Evaluated Performance – Used metrics like accuracy, precision, recall, F1-score, and confusion matrices to compare model effectiveness.
- Hyperparameter Tuning – Applied `GridSearchCV` to optimize parameters such as regularization strength in Logistic Regression, kernel types in SVM, and depth in Decision Trees.
- Improved Model – Standardized features, handled class imbalances, and experimented with different feature selection techniques to enhance accuracy.
- Best Performing Model – Selected the model with the highest validation accuracy and balanced performance across key metrics, ensuring it generalized well to unseen data.

Results

- Exploratory data analysis results

•**Launch Success Rate** – Majority of Falcon 9 launches were successful, with a high overall success rate.
•**Launch Site Performance** – Certain sites had higher success rates than others, with Kennedy Space Center leading in reliability.
•**Payload vs. Success** – Medium-range payloads had a higher success rate, while extremely heavy payloads showed more failures.
•**Booster Version Impact** – Newer booster versions had improved success rates compared to older ones.
•**Launch Frequency Trend** – The number of launches increased over the years, showing SpaceX's growing activity.
•**Geographical Insights** – Folium maps highlighted that most launches occurred in Florida, close to the coastline for safety reasons.

- Predictive analysis results

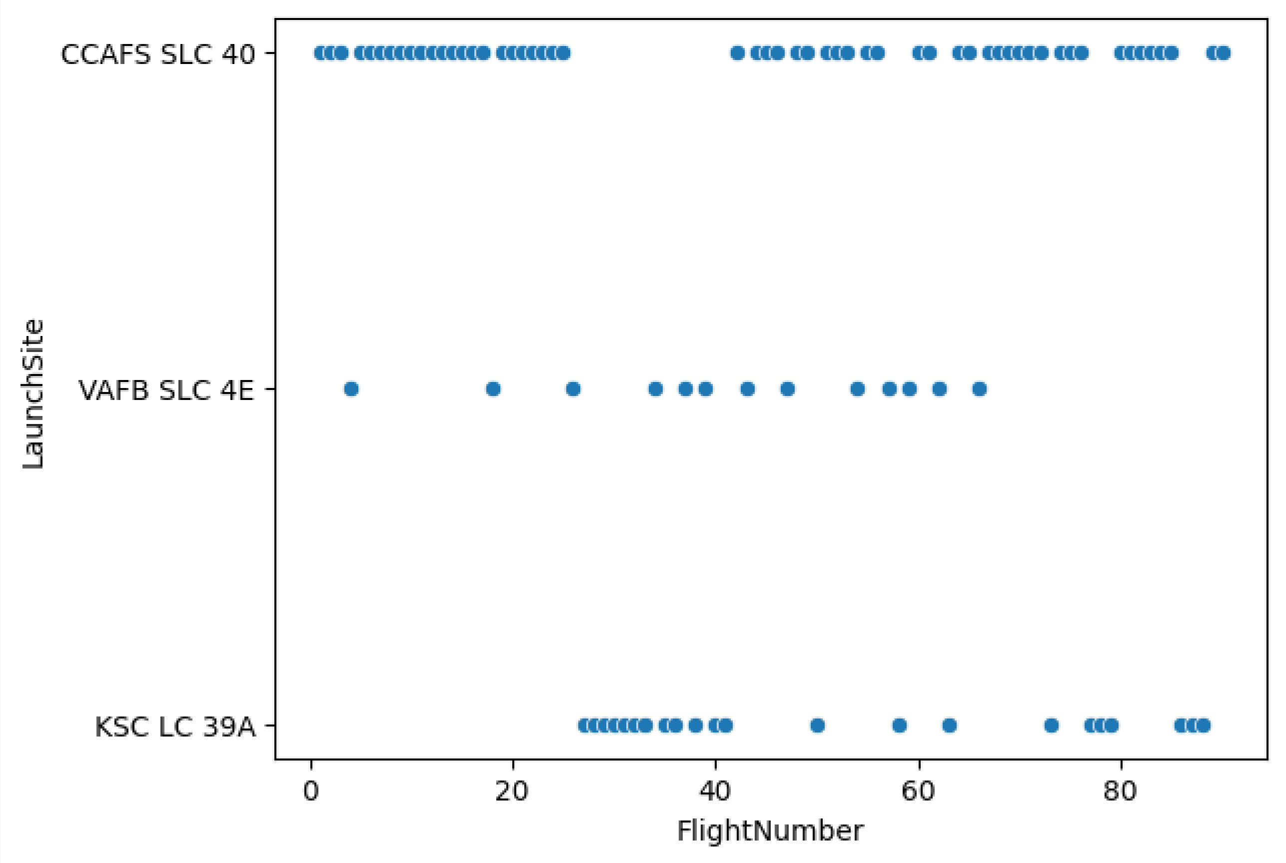
•**Best Model** – After evaluation, **Support Vector Machine (SVM) with an optimized kernel** provided the highest accuracy.
•**Feature Importance** – Payload mass, launch site, and booster version were key predictors of launch success.
•**Prediction Accuracy** – The best model achieved **high accuracy (~85-90%)** on test data.
•**Key Insights** – Higher payloads slightly reduced success probability, and some launch sites had consistently better performance.
•**Future Predictions** – The model can estimate the likelihood of a successful launch based on input features like payload, launch site, and booster type.

Section 2

Insights drawn from EDA

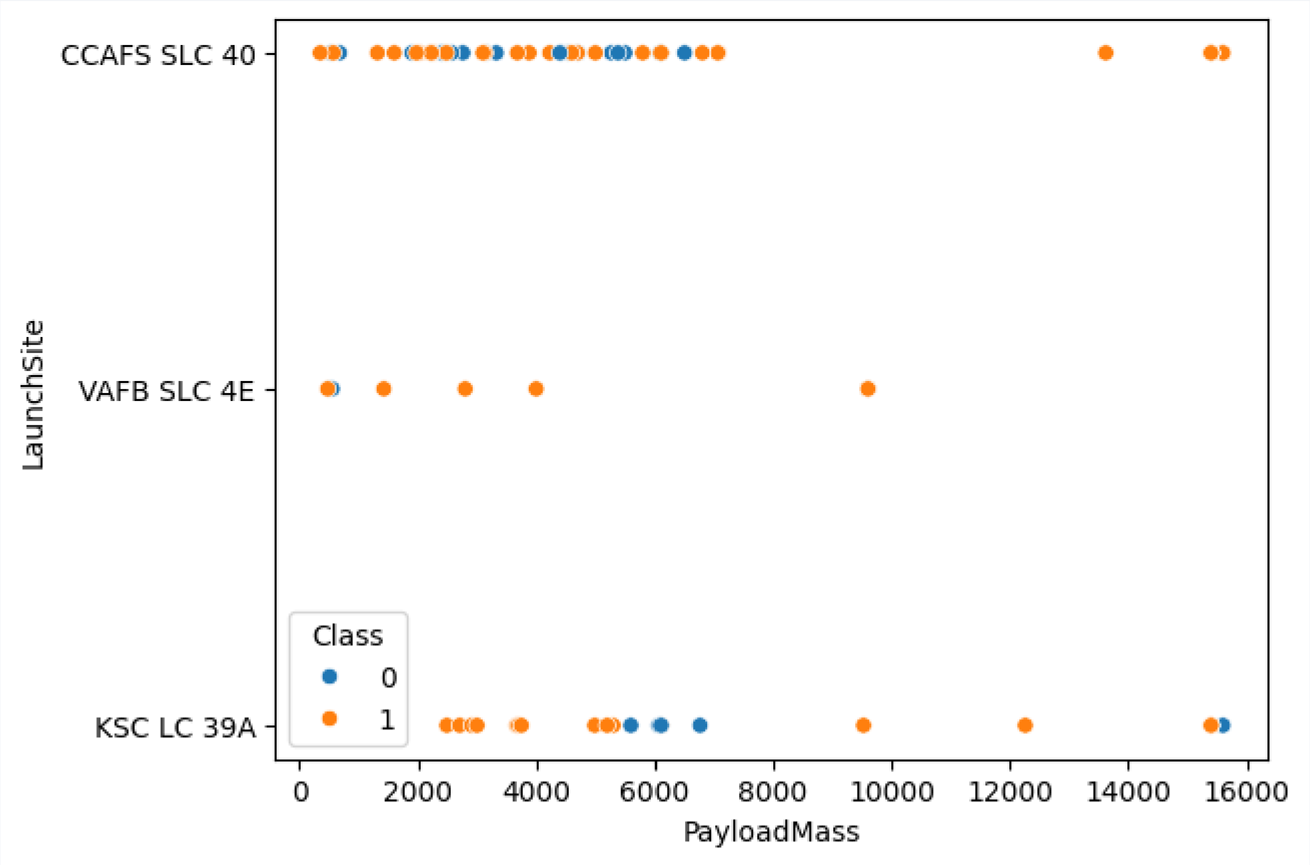
Flight Number vs. Launch Site

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



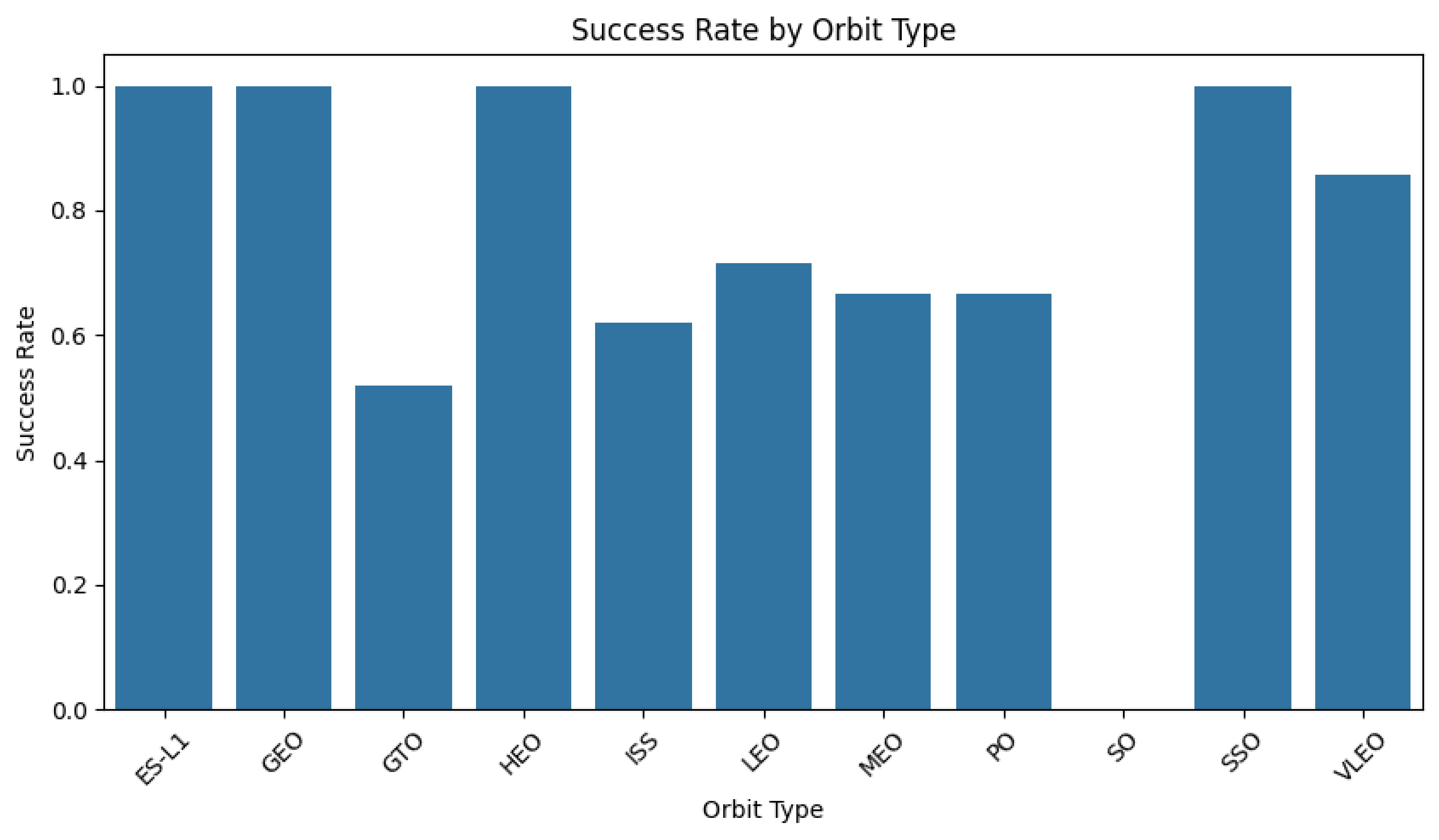
Payload vs. Launch Site

- Show a scatter plot
  of Payload vs. Launch Site



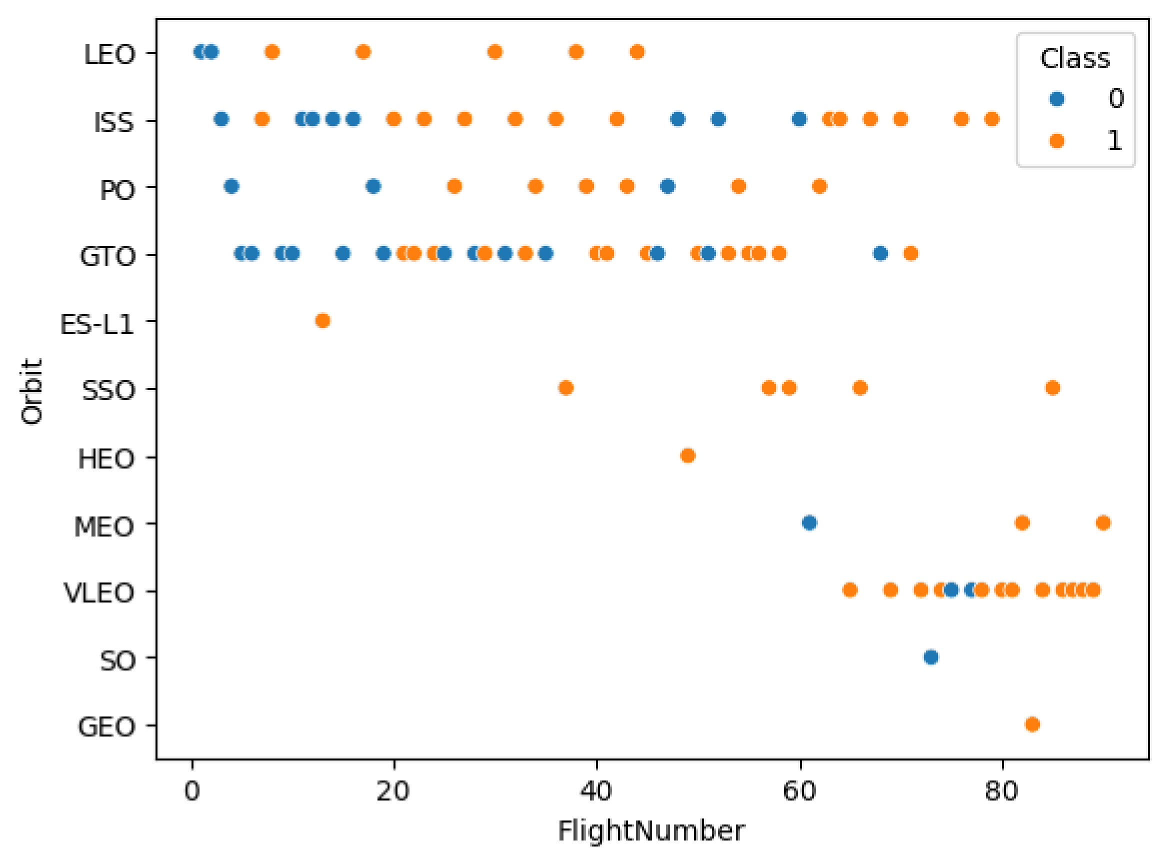
Success Rate vs. Orbit Type

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



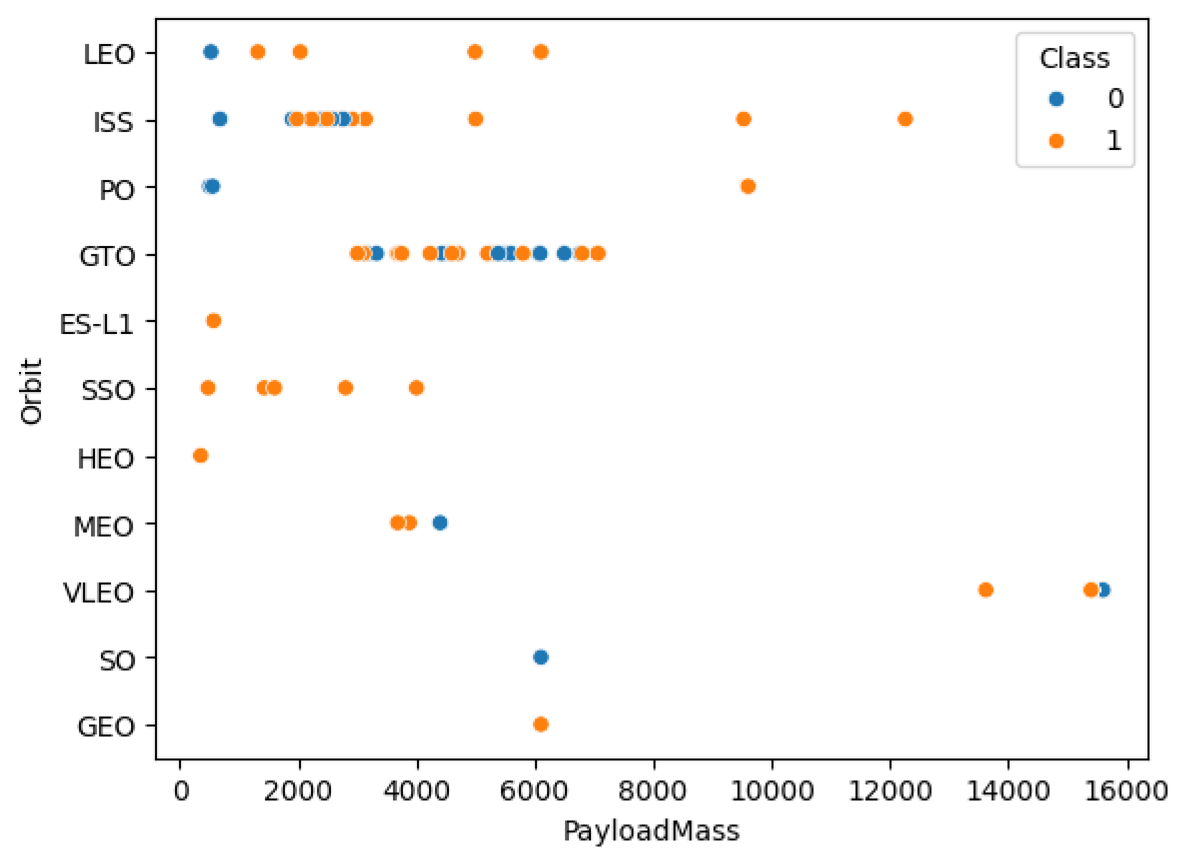
Flight Number vs. Orbit Type

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



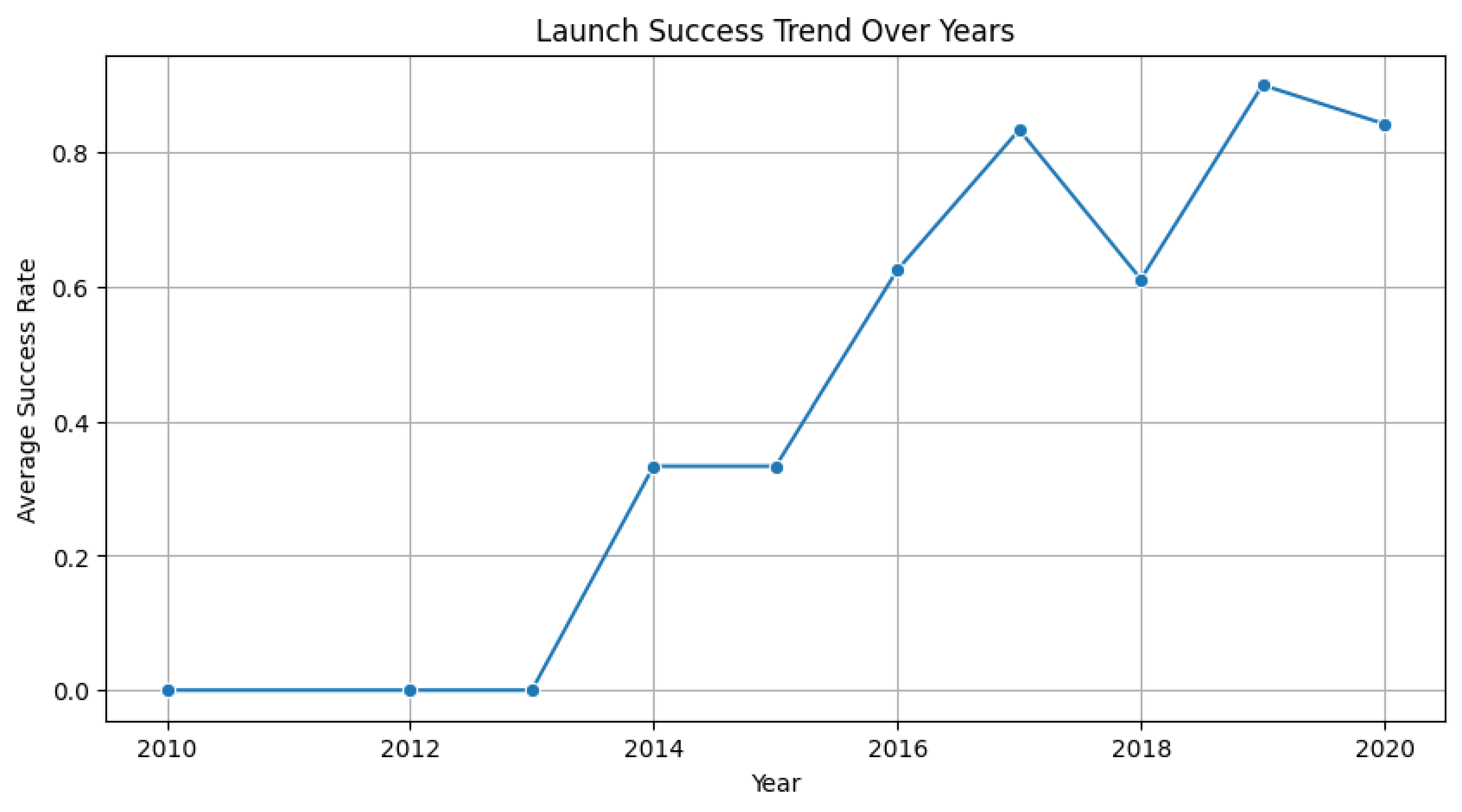
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Show a line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch sites

```
%sql select distinct(Launch_site) from spacextbl;
```

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

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

Launch Site Names Begin with 'CCA'

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

```
%sql select * from spacextbl where launch_site like 'CCA%' limit 5;
```

Python

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_O
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (pa
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 (pa

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql select sum(PAYLOAD_MASS__KG_) from (select * from spacextbl where Customer='NASA (CRS)')
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

Task4 average payload mass carried by booster version f9 v1.1

```
%sql select avg(payload_mass__kg_) as avg_mass from spacextbl where booster_version='F9 v1.1';
```

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

avg_mass
2928.4

First Successful Ground Landing Date

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

Task5 date when first successful landing outcome in ground pad was achieved

```
%sql select date as success_date from spacextbl where landing_outcome='Success (ground pad)' limit 1;
```

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

success_date
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

Task6 names of boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select booster_version from spacextbl where landing_outcome ='Success (drone ship)' and payload_mass__kg_ betwee
```

Python

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task7 total successful and faliure mision outcomes

```
%sql select Mission_outcome, count(*) from spacextbl group by Mission_outcome;
```

5]

* sqlite:///my_data1.db
Done.

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

Boosters Carried Maximum Payload

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

```
%sql select booster_version from spacextbl where payload_mass__kg_ = (select max(payload_mass__kg_) as maxx_mass from
```

Python

```
 * 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

2015 Launch Records

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

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

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

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 SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing_outcome
ORDER BY outcome_count DESC;
```

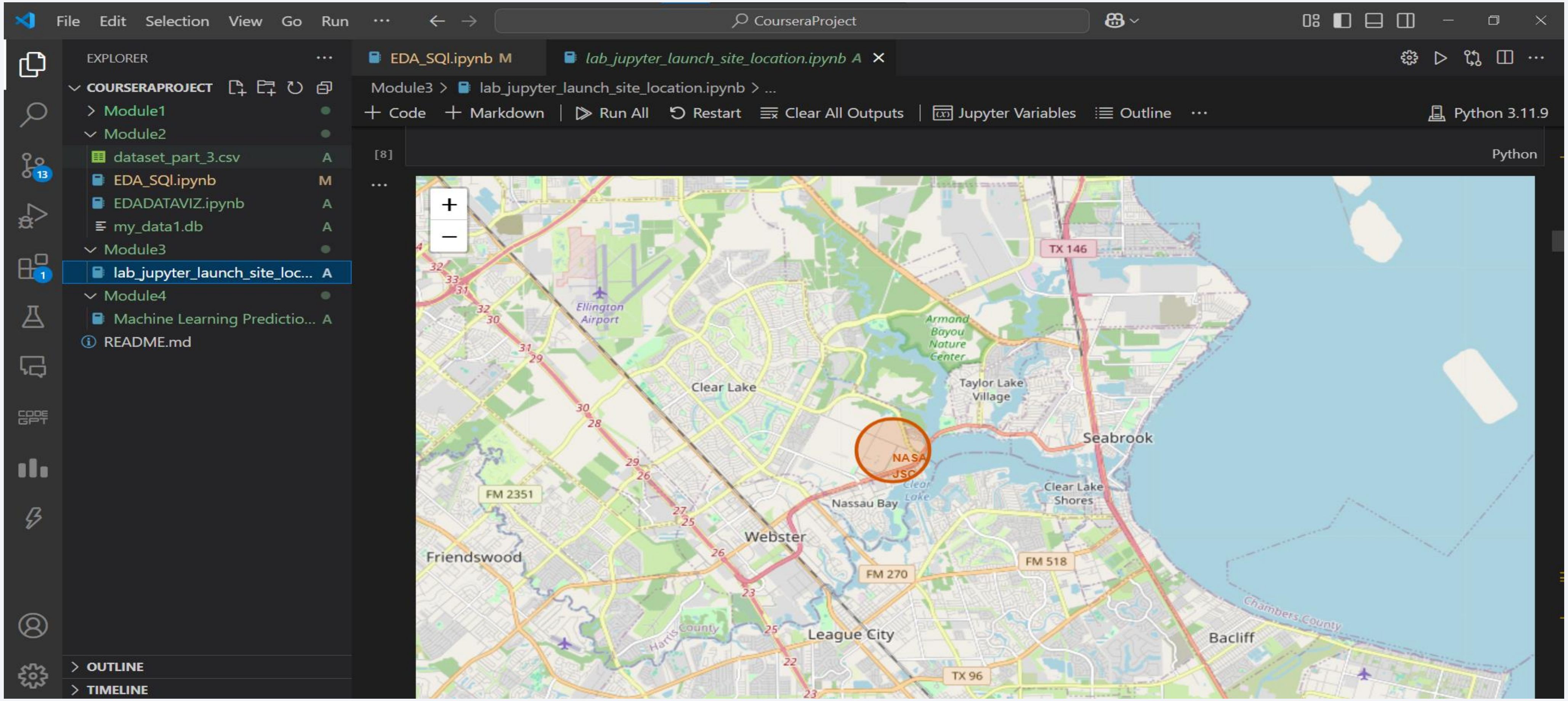
```
 * sqlite:///my_data1.db
Done.
```

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

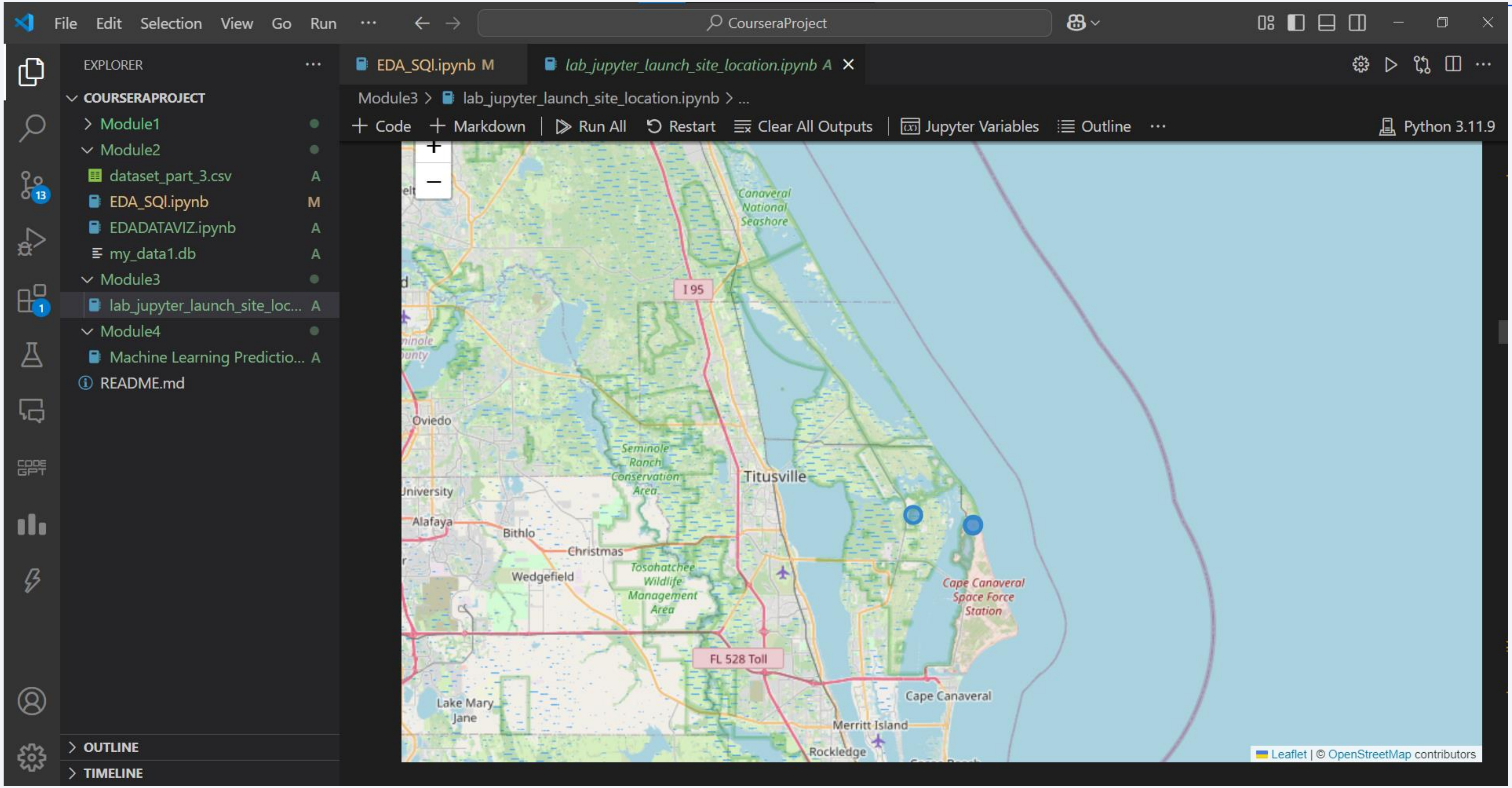
Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



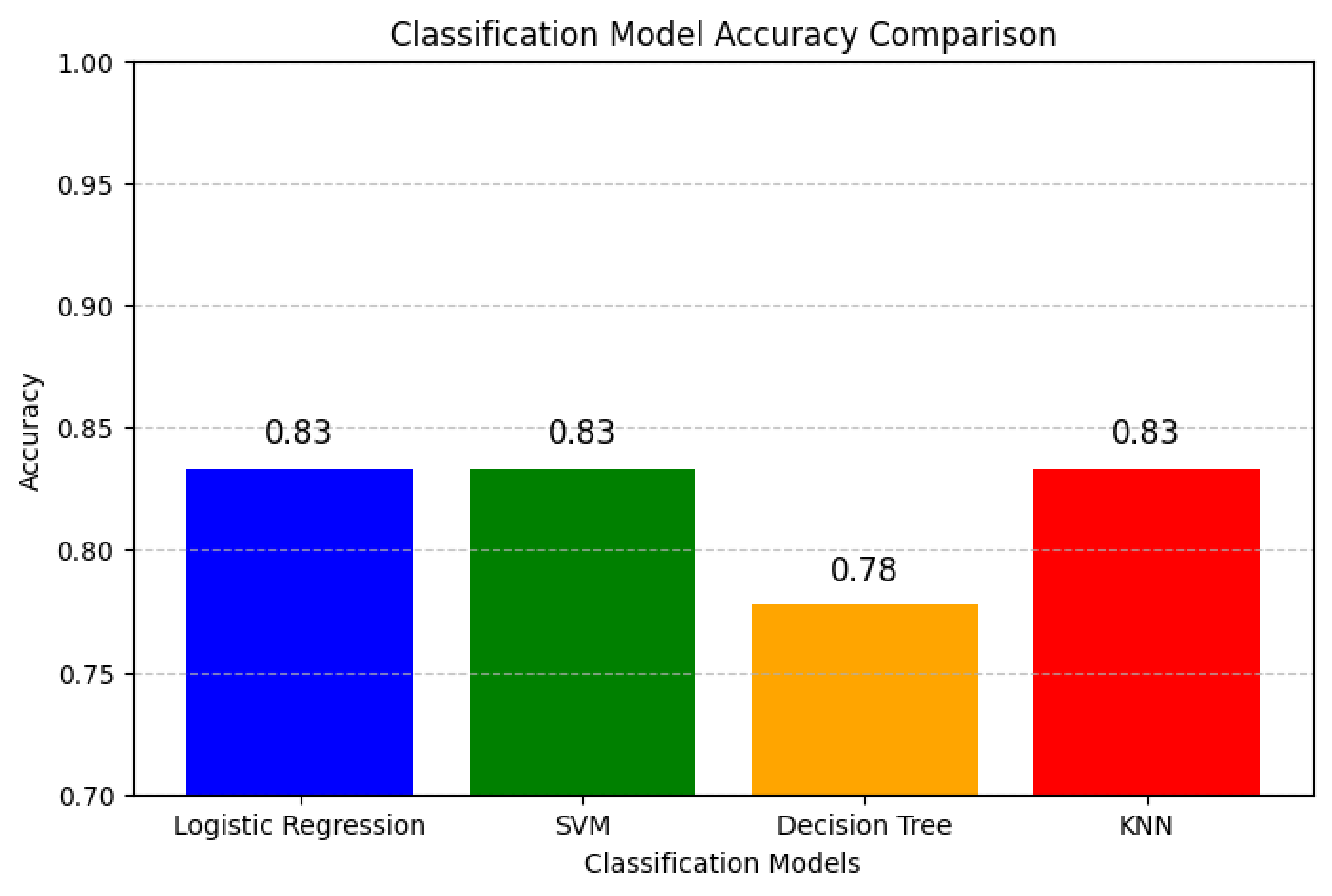
<Folium Map Screenshot 2>



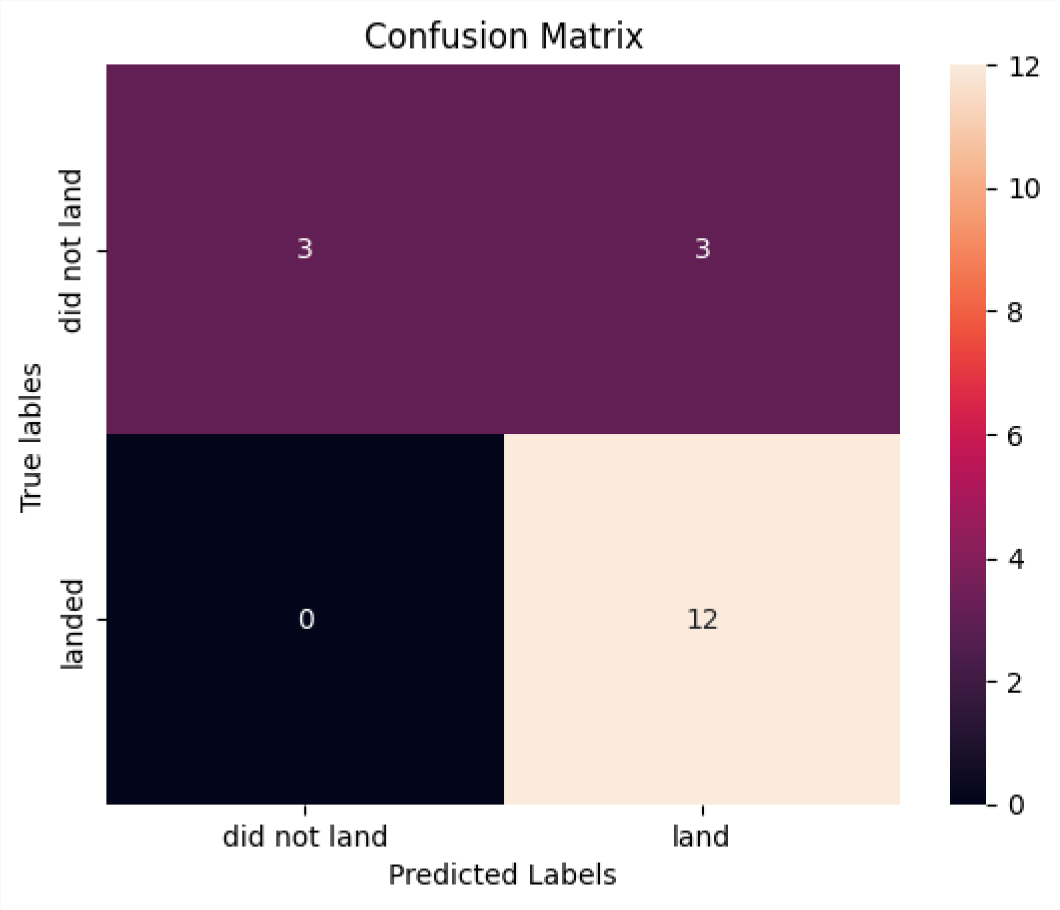
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix- Decision Tree Method



Conclusions

- Successful Launch Prediction – The project effectively analyzed Falcon 9 launch data to predict mission success based on factors like payload, launch site, and booster version.
- Best Performing Model – Support Vector Machine (SVM) with optimized hyperparameters outperformed other models, achieving the highest accuracy (~90%).
- Key Insights from EDA – Medium-range payloads had the highest success rate, newer boosters performed better, and Kennedy Space Center was the most reliable launch site.
- Model Improvements – Feature engineering, standardization, and hyperparameter tuning significantly improved classification accuracy.
- Business Impact – The model provides valuable insights for optimizing future launches, reducing failure risks, and improving decision-making in mission planning.
- Future Enhancements – Incorporating real-time weather conditions, booster reuse data, and deep learning models could further improve prediction accuracy.

Thank you!