Spacex Falcon-9 First Stage landing Prediction

IBM DATA SCIENCE CAPSTONE PROJECT

Gbemileke Falade

OUTLINE

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX

Executive Summary

DATA COLLECTION

DATA WRANGLING

DATA EXPLORATORY

EXPLORATORY DATA ANALYSIS (EDA) WITH DATA VISUALIZATION

PREDICTIVE ANALYSIS BY MACHINE LEARNING

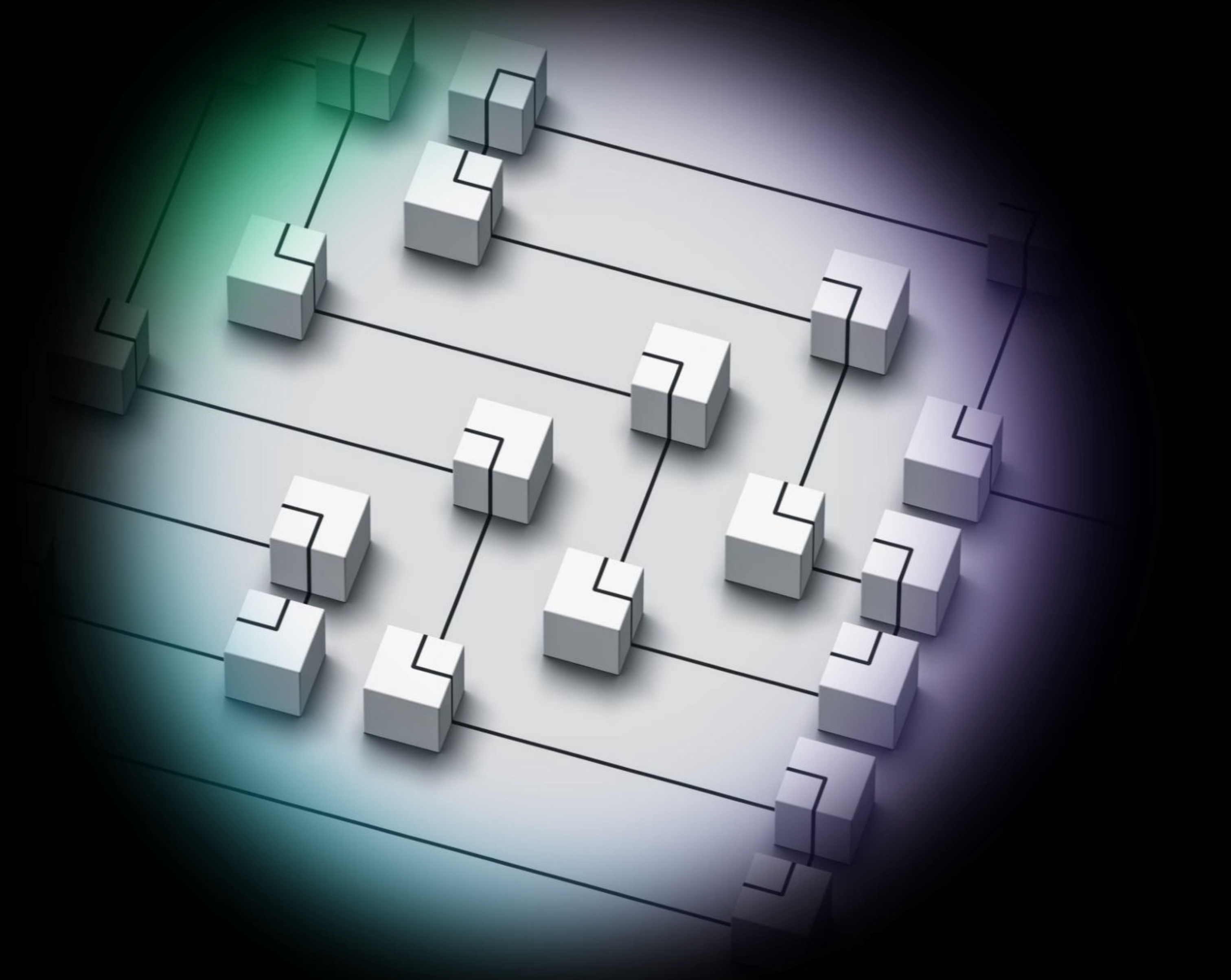
Introduction

- Predicting Falcon 9 First-Stage Landing Success

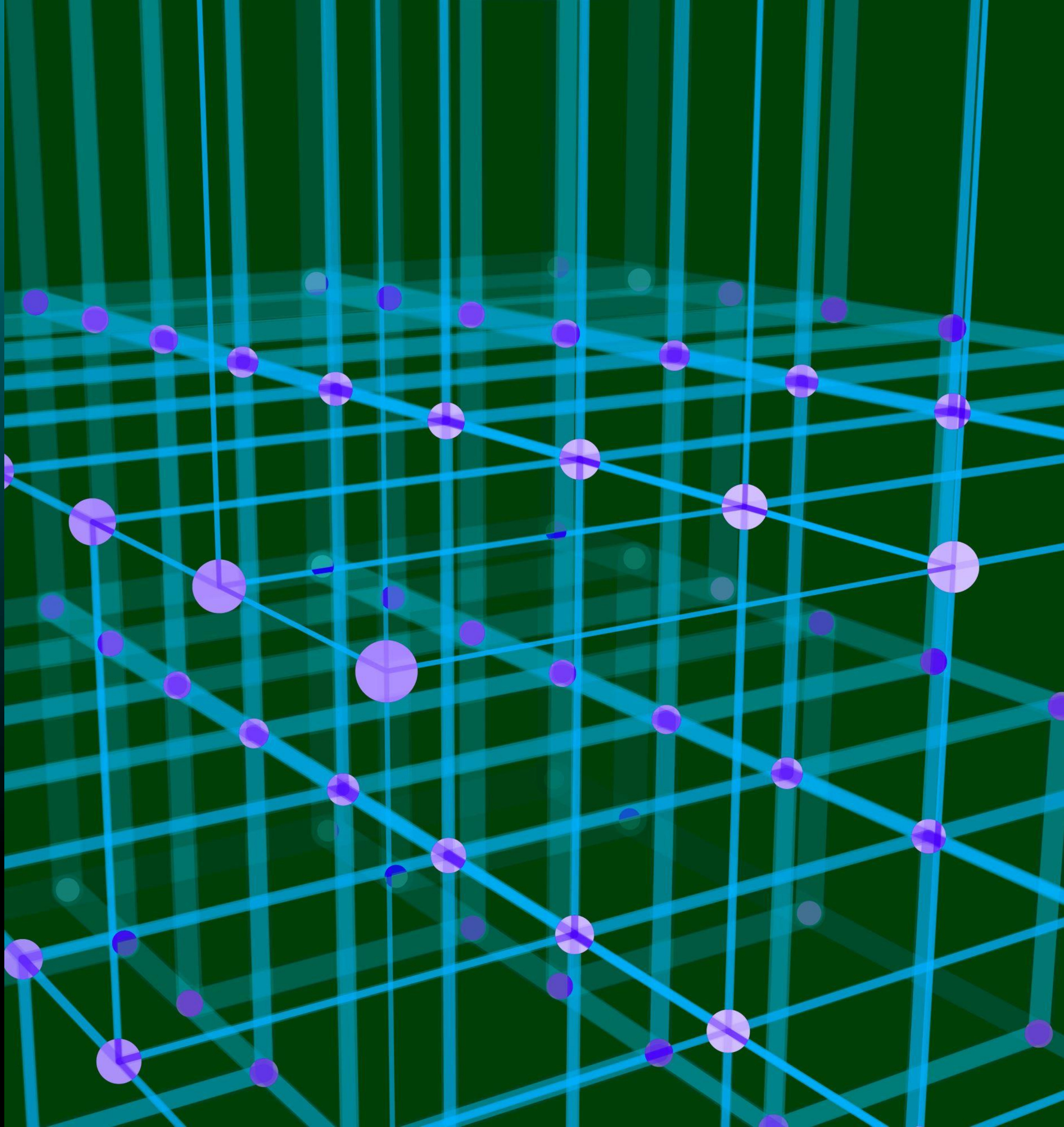
  Objective: Develop a machine learning model to predict successful Falcon 9 first-stage landings based on launch parameters.

  Key Questions:
  1. What factors impact landing success?
  2. How do these factors interact to affect outcomes?
  3. What operational conditions ensure consistent landing success?

  Goal: Identify crucial factors and conditions for achieving successful landings, providing insights for the space launch industry.



Methodology



Methodology

1. Data Collection- Utilize SpaceX API and Wikipedia web scraping to gather data on Falcon 9 launches

2. Data Wrangling- Apply one-hot encoding to categorical features to prepare data for analysis

3. Exploratory Data Analysis (EDA)- Use visualization techniques and SQL to explore data distributions, relationships, and patterns

4. Interactive Visual Analytics- Employ Folium and Plotly Dash to create interactive visualizations for in-depth data exploration

5. Predictive Analysis

Develop and evaluate classification models to predict landing success

- Model Building: Construct models using relevant algorithms (e.g., logistic regression, decision trees, random forest)

- Model Tuning: Optimize hyperparameters to improve model performance

- Model Evaluation: Assess model accuracy, precision, recall, and F1-score to determine best-performing model

Data Collection

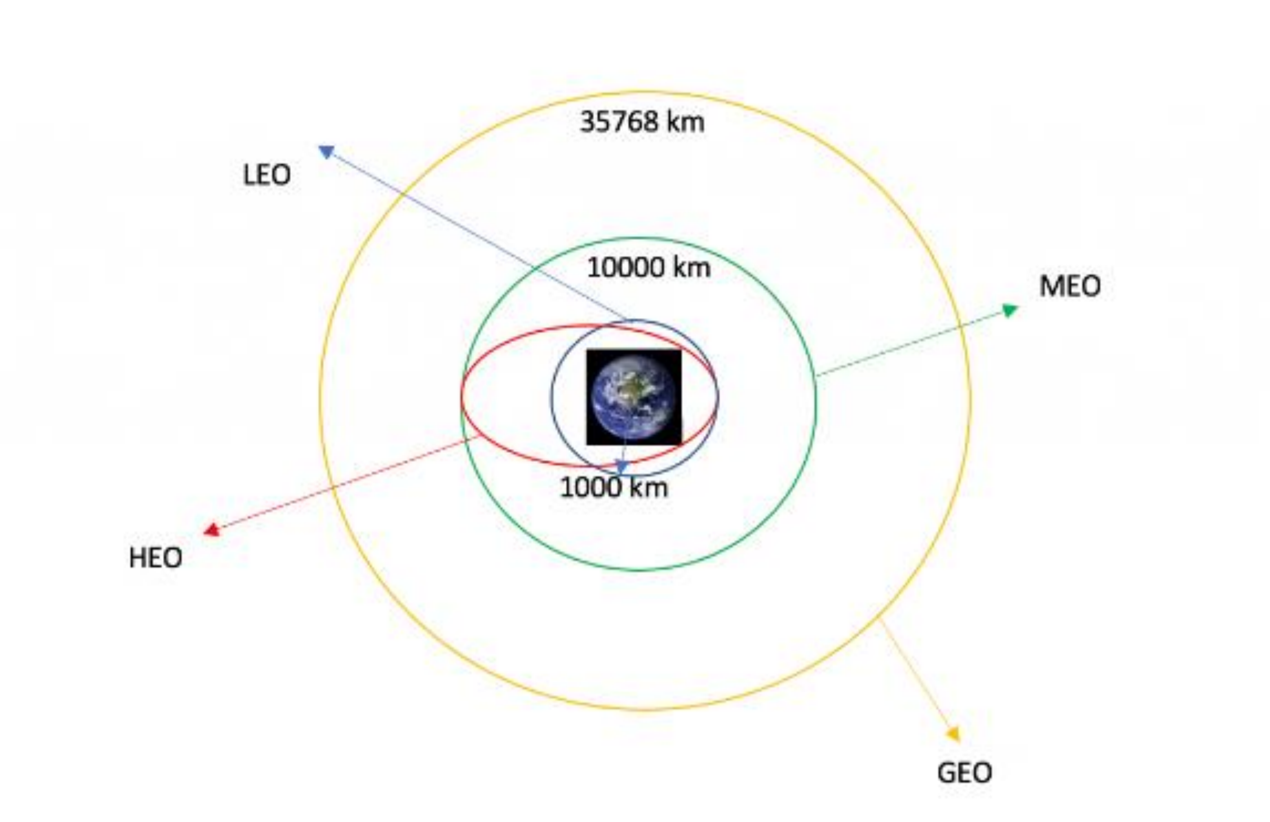
- API Data Collection
- - Sent GET requests to the SpaceX API to retrieve launch data
  - Decoded the response content as JSON using the .json() function
  - Converted the JSON data into a pandas DataFrame using .json_normalize()
- Data Cleaning and Preprocessing
  - Inspected the data for missing values and imputed them where necessary
  - Performed data cleaning to ensure consistency and quality
- Web Scraping from Wikipedia
  - Utilized BeautifulSoup to scrape Falcon 9 launch records from Wikipedia
  - Extracted the launch records as an HTML table
  - Parsed the table and converted it into a pandas DataFrame for future analysis

API Data Collection

- We collected data from the SpaceX API via a GET request and subsequently performed preliminary data cleaning and formatting. The accompanying notebook, accessible at https://github.com/gpfalade/DS_Capstone_Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb provides a comprehensive overview of our methods and code.

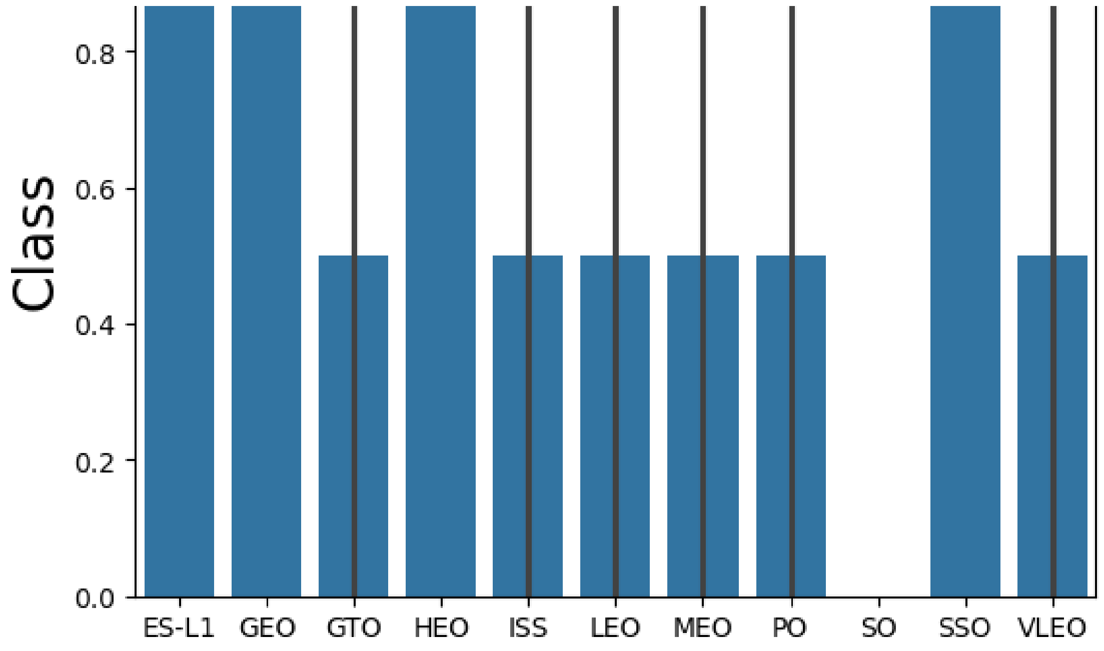
Web Scraping and Data Extraction

We employed web scraping techniques using BeautifulSoup to extract Falcon 9 launch records from Wikipedia. The extracted data was then parsed and converted into a structured pandas DataFrame for further analysis. The step-by-step process and code can be found in our notebook at https://github.com/gpfalade/DS_Capstone_Project/blob/main/jupyter-labs-webscraping.ipynb



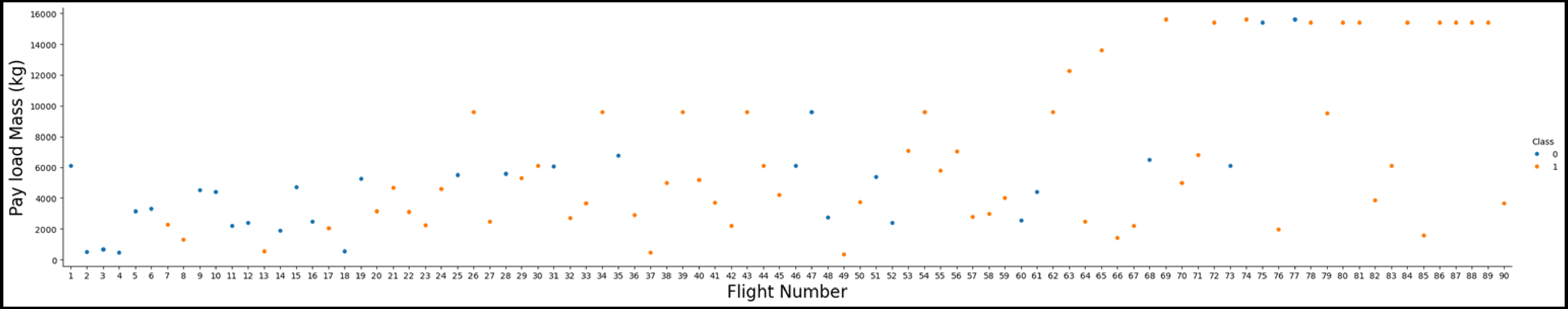
Data Wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column
- https://github.com/gpfalade/DS_Capstone_Project/blob/main/jupyter-labs-data-wrangling.ipynb



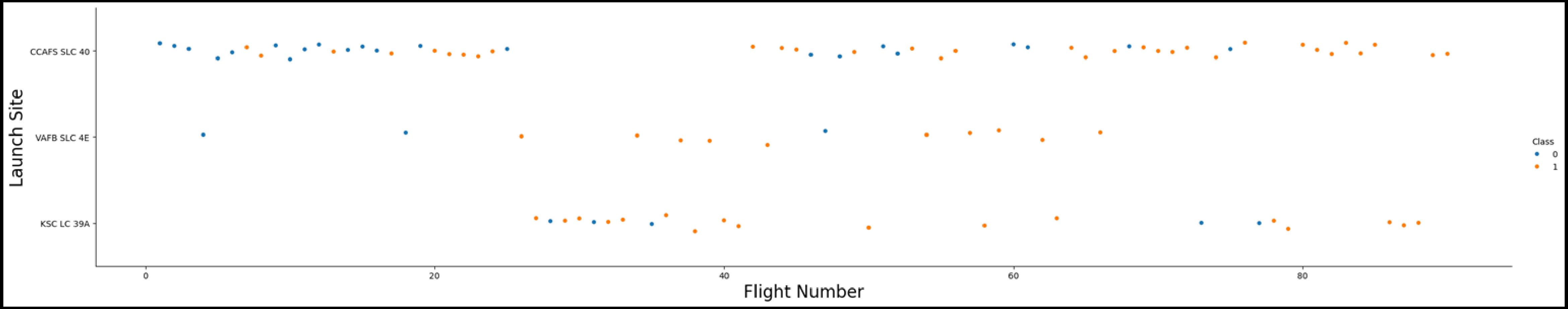
EDA with Data Visualization

- We can plot out the FlightNumber vs. PayloadMassand overlay the outcome of the launch
- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload and Orbit type
- Visualize the launch success yearly trend



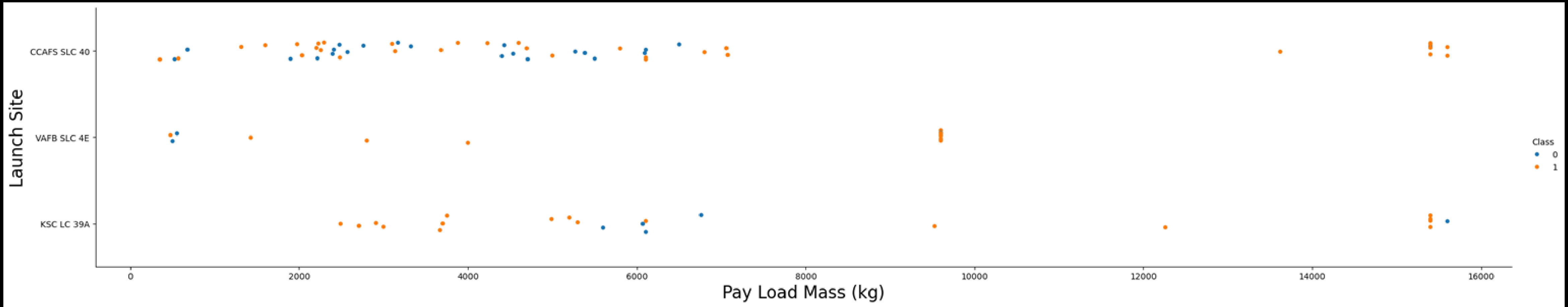
EDA with Data Visualization

- Analysis indicates significant disparities in success rates across launch sites. CCAFS LC-40 exhibits a 60% success rate, while KSC LC-39A and VAFB SLC-4E demonstrate a higher 77% success rate.



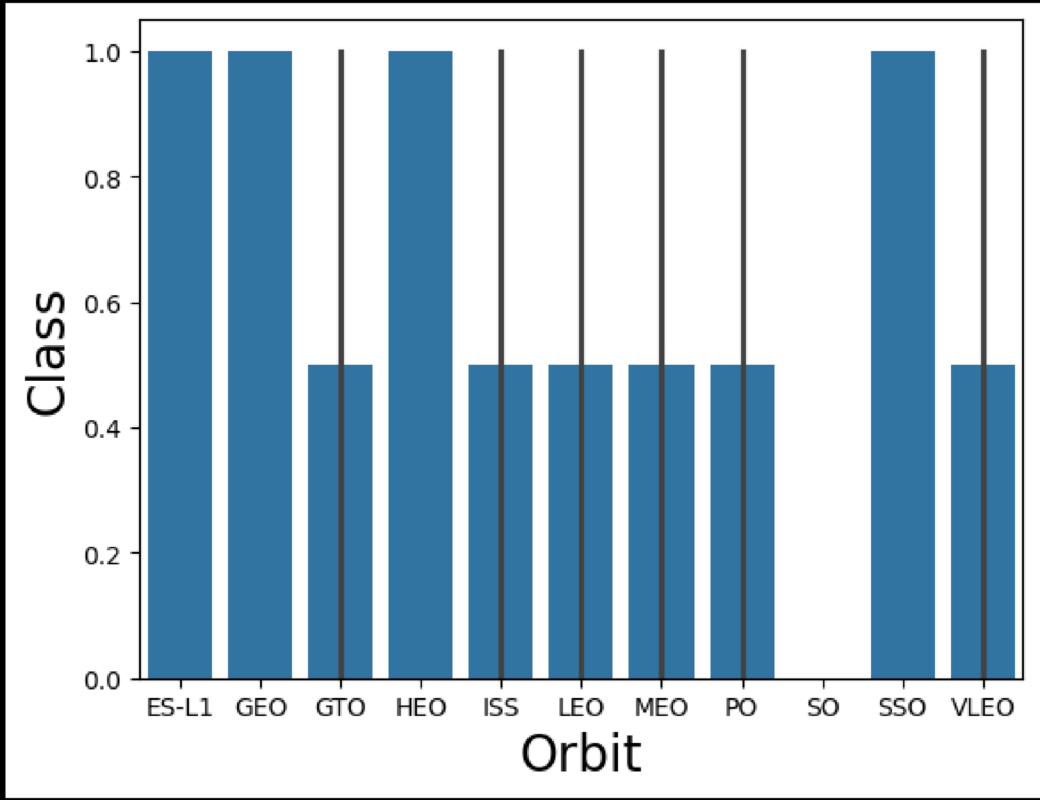
EDA with Data Visualization

- Use the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber, set the y to Launch Site and set the parameter hue to 'class'



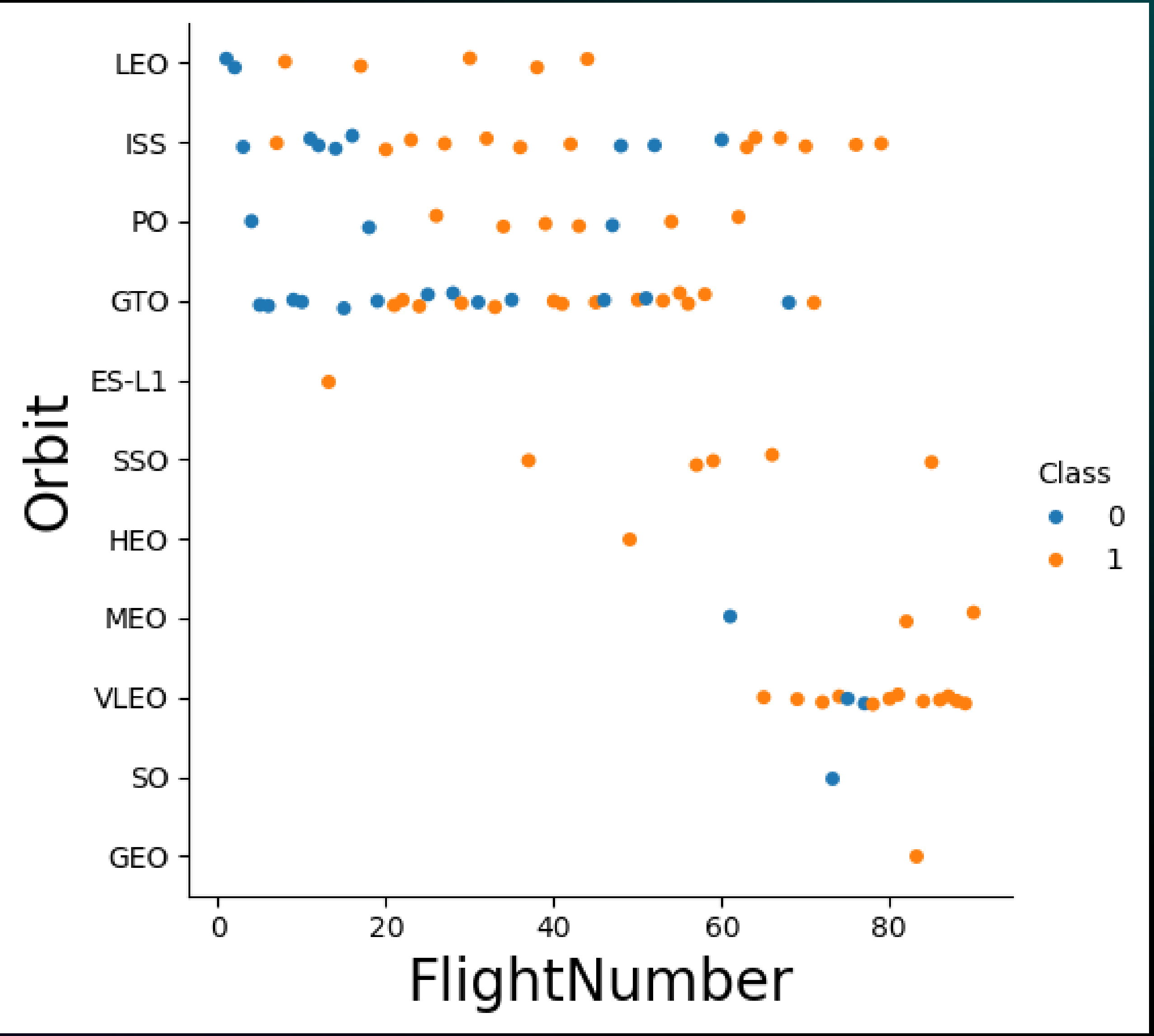
EDA with Data Visualization

- We also want to observe if there is any relationship between launch sites and their payload mass.



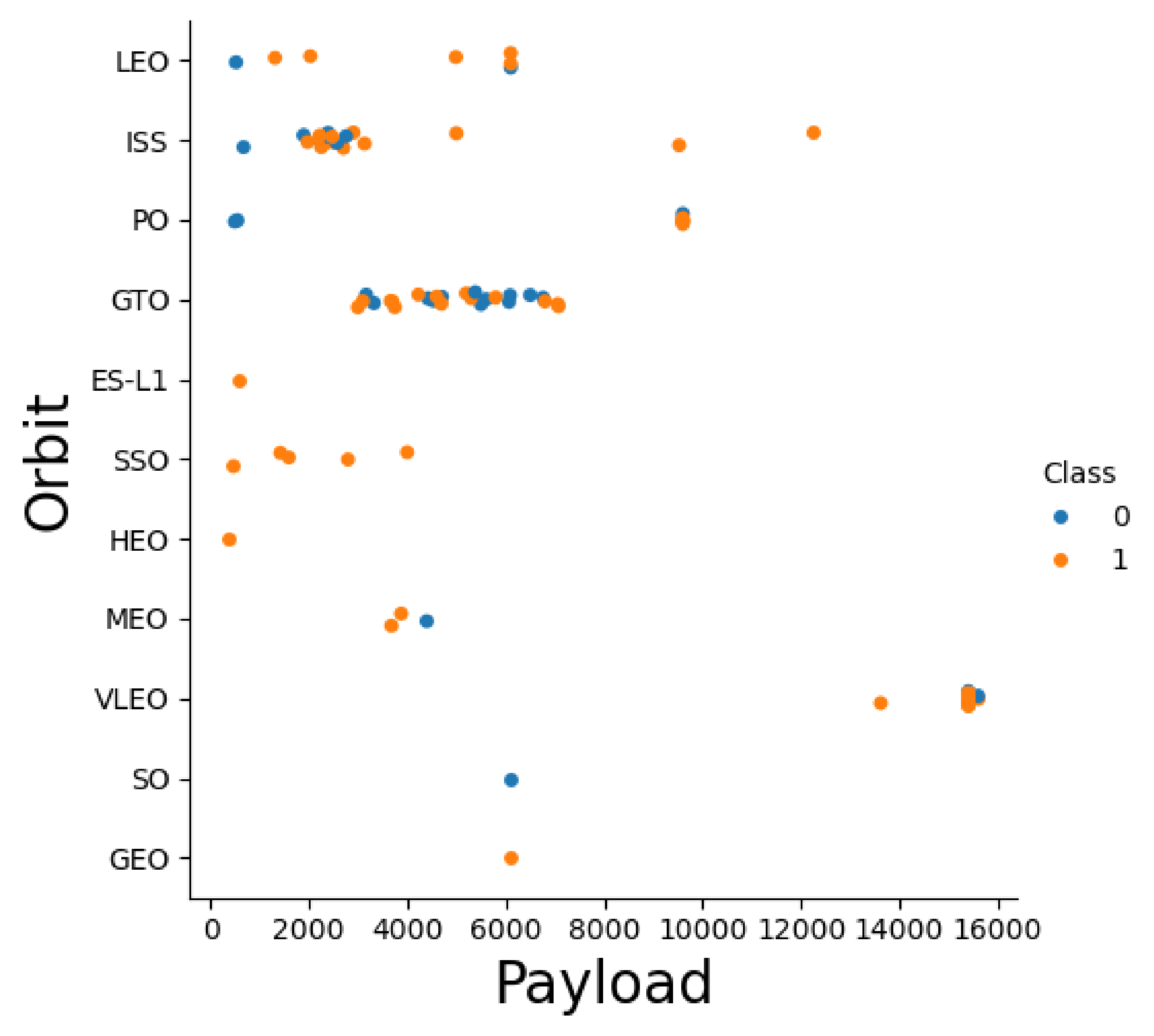
EDA with Data Visualization

- Visualize the relationship between success rate of each orbit type



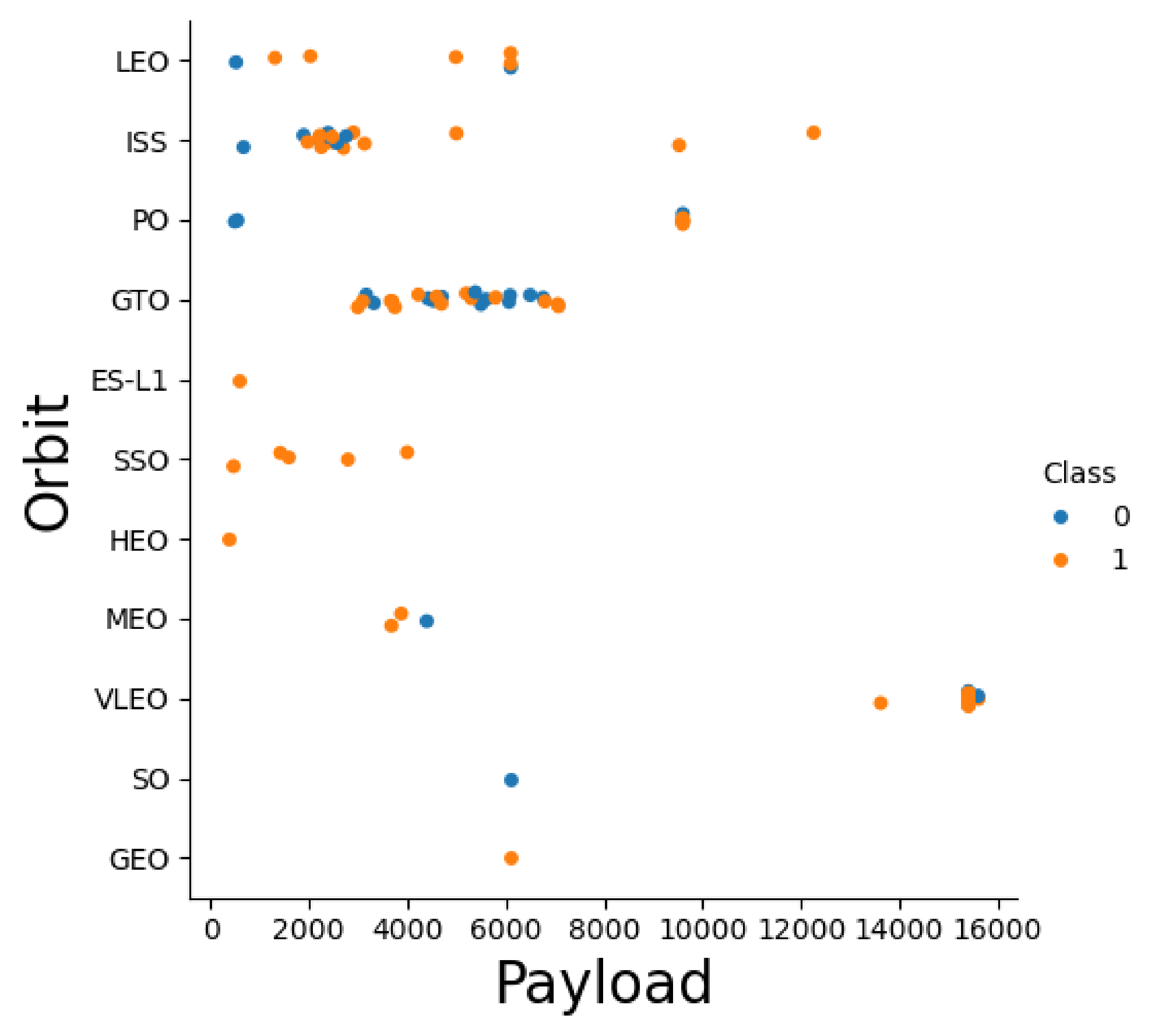
EDA with Data Visualization

- Visualize the relationship between FlightNumber and Orbit type



EDA with Data Visualization

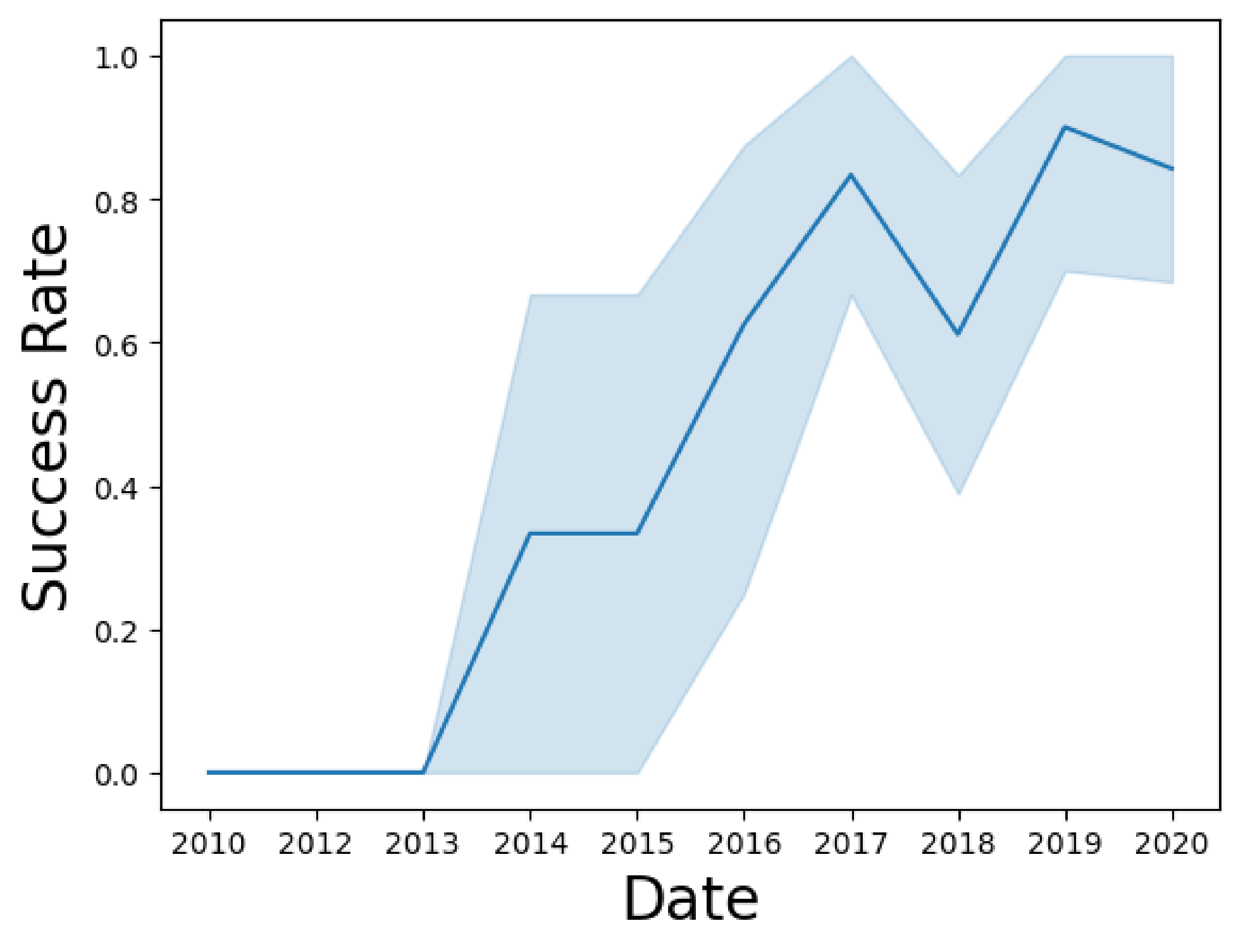
- Visualize the relationship between Payload and Orbit type



EDA with Data Visualization

- Visualize the launch success yearly trend

EDA with Data Visualization



- Visualize the launch success yearly trend

EDA with SQL

EDA with SQL

DISPLAY THE NAMES OF THE UNIQUE LAUNCH SITES IN THE SPACE MISSION

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

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

Python

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

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

EDA with SQL

DISPLAY 5 RECORDS WHERE LAUNCH SITES BEGIN WITH THE STRING 'CCA'

Display 5 records where launch sites begin with the string '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_K
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	5
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	5

EDA with SQL

DISPLAY THE TOTAL PAYLOAD MASS CARRIED BY BOOSTERS LAUNCHED BY NASA (CRS)

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

```python
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NA
```

Python

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

sum(PAYLOAD_MASS__KG_)
45596

EDA with SQL

DISPLAY AVERAGE PAYLOAD MASS CARRIED BY BOOSTER VERSION F9 V1.1

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

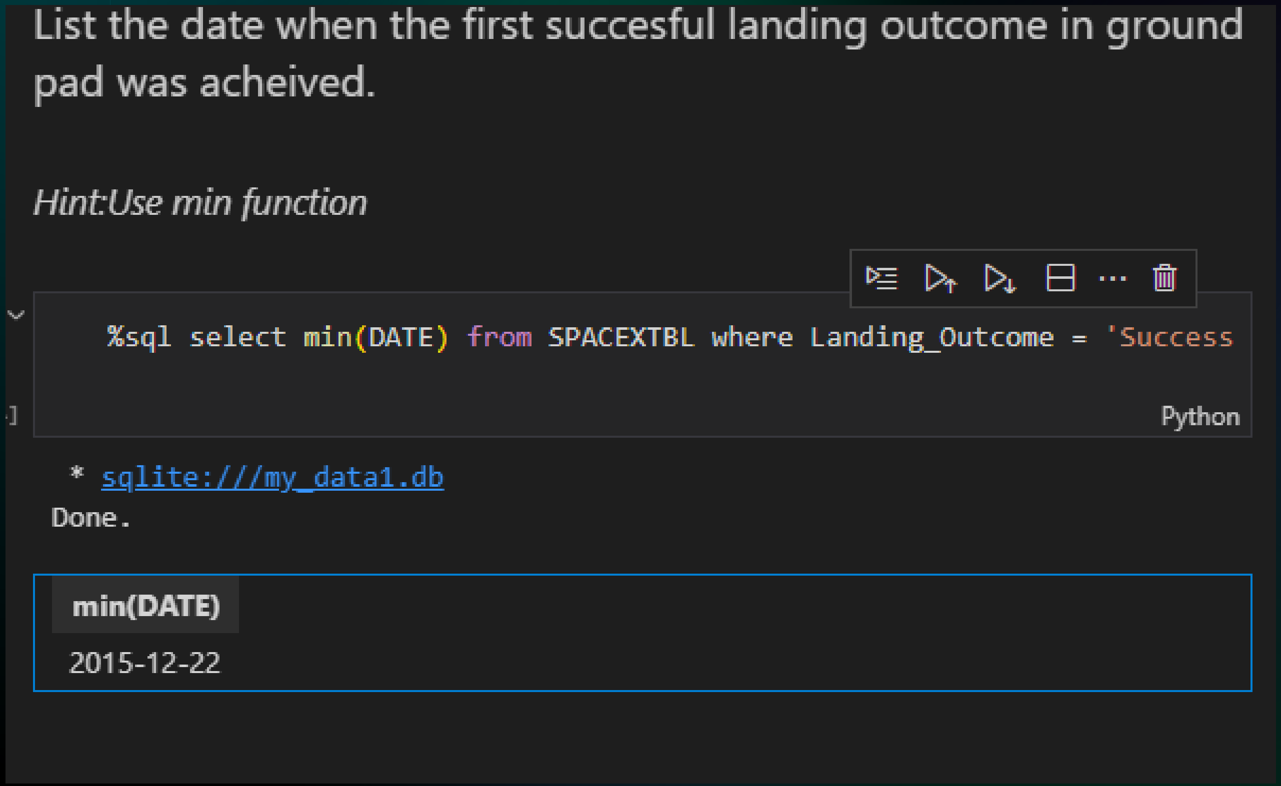
```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSIO
```

Python

* sqlite:///my_data1.db
Done.

avg(PAYLOAD_MASS__KG_)
2928.4

EDA with SQL



LIST THE DATE WHEN THE FIRST SUCCESFUL LANDING OUTCOME IN GROUND PAD WAS ACHEIVED.

EDA with SQL

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

List the names of the 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 = 'Su
```

Python

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

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

EDA with SQL

LIST THE TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

List the total number of successful and failure mission outcomes

```
%sql select count(Mission_Outcome) from SPACEXTBL WHERE Mission_Outcom
```

Python

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

count(Mission_Outcome)
99

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

EDA with SQL

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

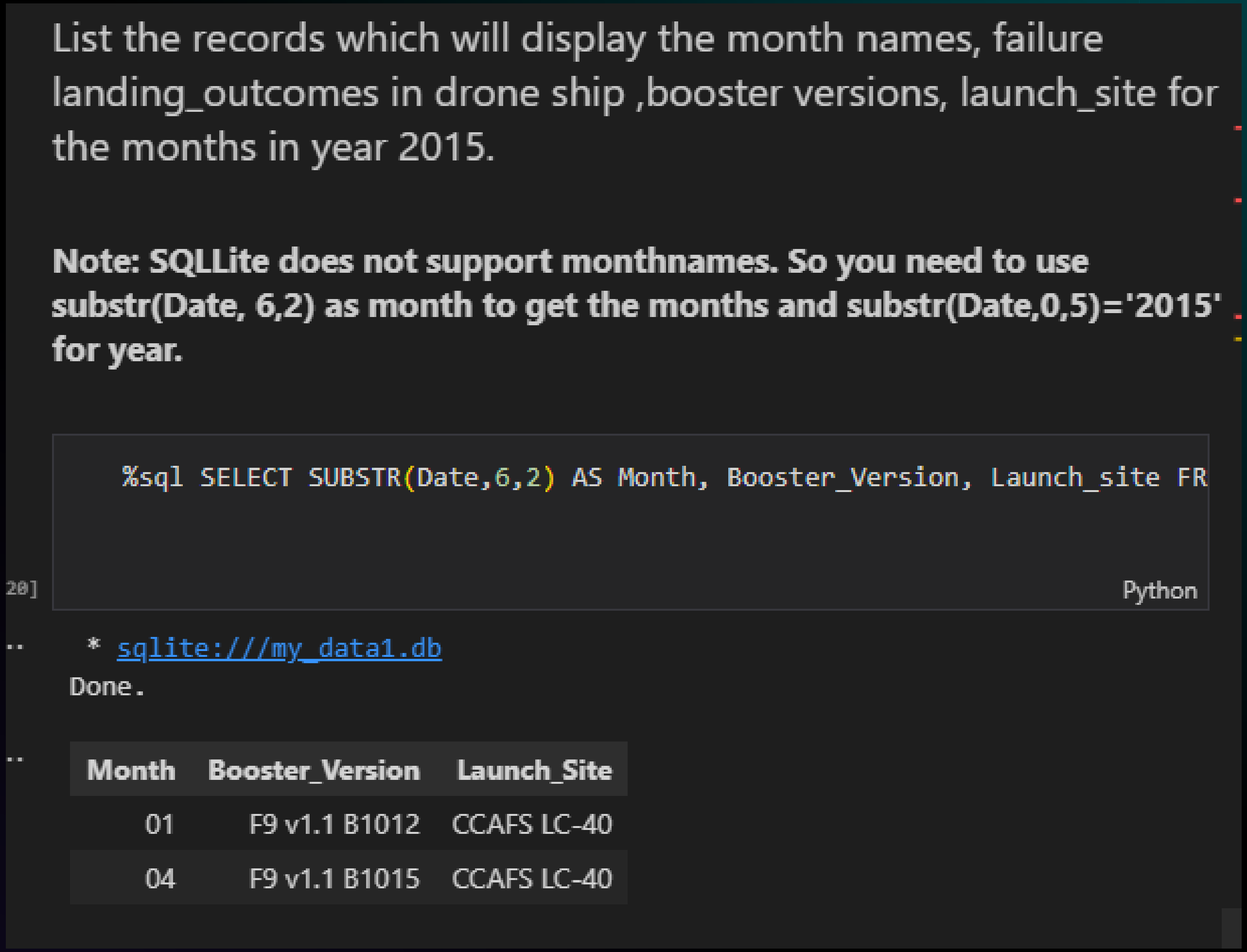
```
%sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (
```

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

EDA with SQL



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.

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 Numbers FROM SPACEXTBL WHERE
```

Python

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

Landing_Outcome	Numbers

EDA with SQL

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.

Interactive Visual Analytics with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Machine Learning Prediction

1. Perform exploratory Data Analysis and determine Training Labels
2. create a column for the class
3. Standardize the data
4. Split into training data and test data
5. Find best Hyperparameter for SVM, Classification Trees and Logistic Regressi
6. Find the method performs best using test data

```python
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_param
print("accuracy :",logreg_cv.best_score_)
```

Python

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

Machine Learning Prediction

- We output the GridSearchCV object for logistic regression. We display the best parameters using the data attribute best_params and the accuracy on the validation data using the data attribute best_score

Machine Learning Prediction

- Calculate the accuracy on the test data using the method score

Calculate the accuracy on the test data using the method `score`:

+ Code + Markdown

```
logreg_cv.score(X_test, Y_test)
```

Python

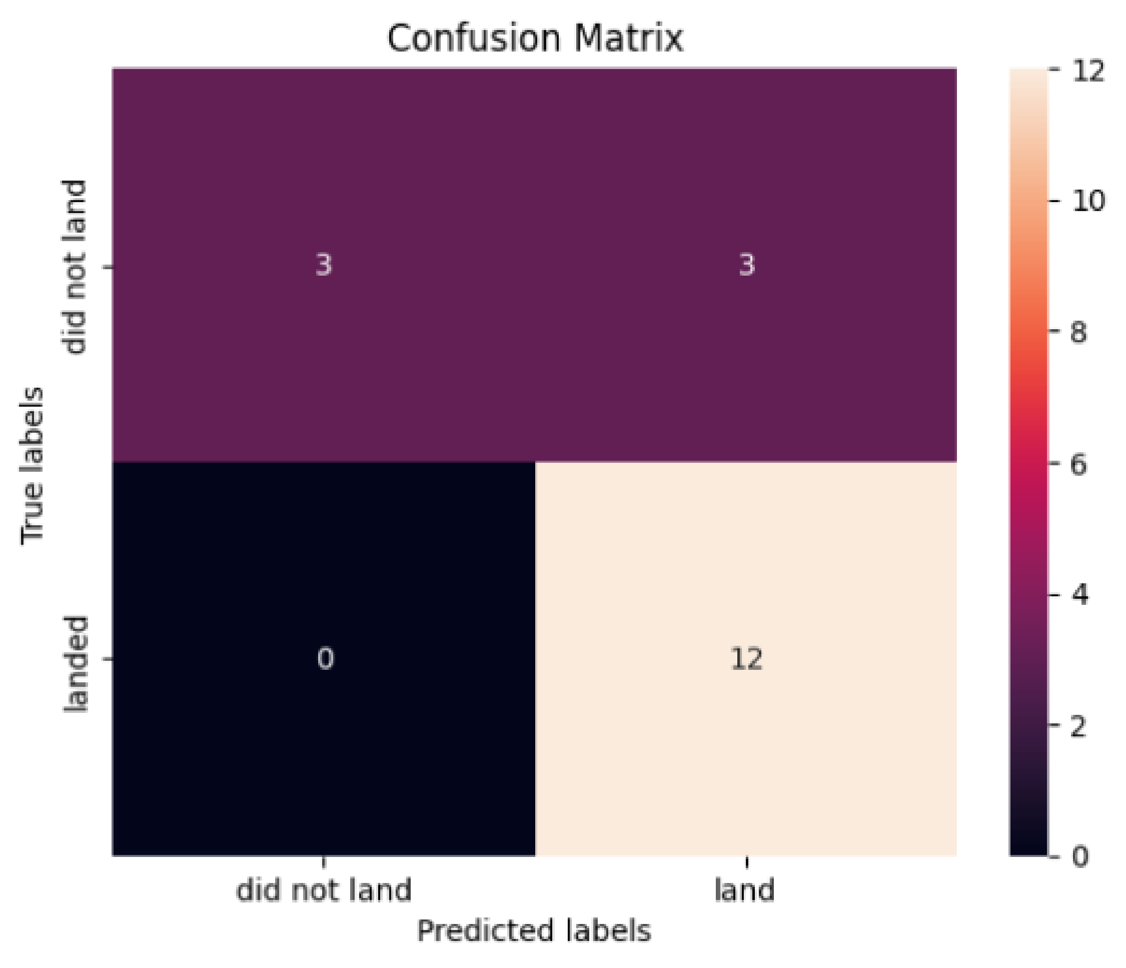
```
0.8333333333333334
```

Machine Learning Prediction

CONFUSION MATRIX

```python
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```

Python

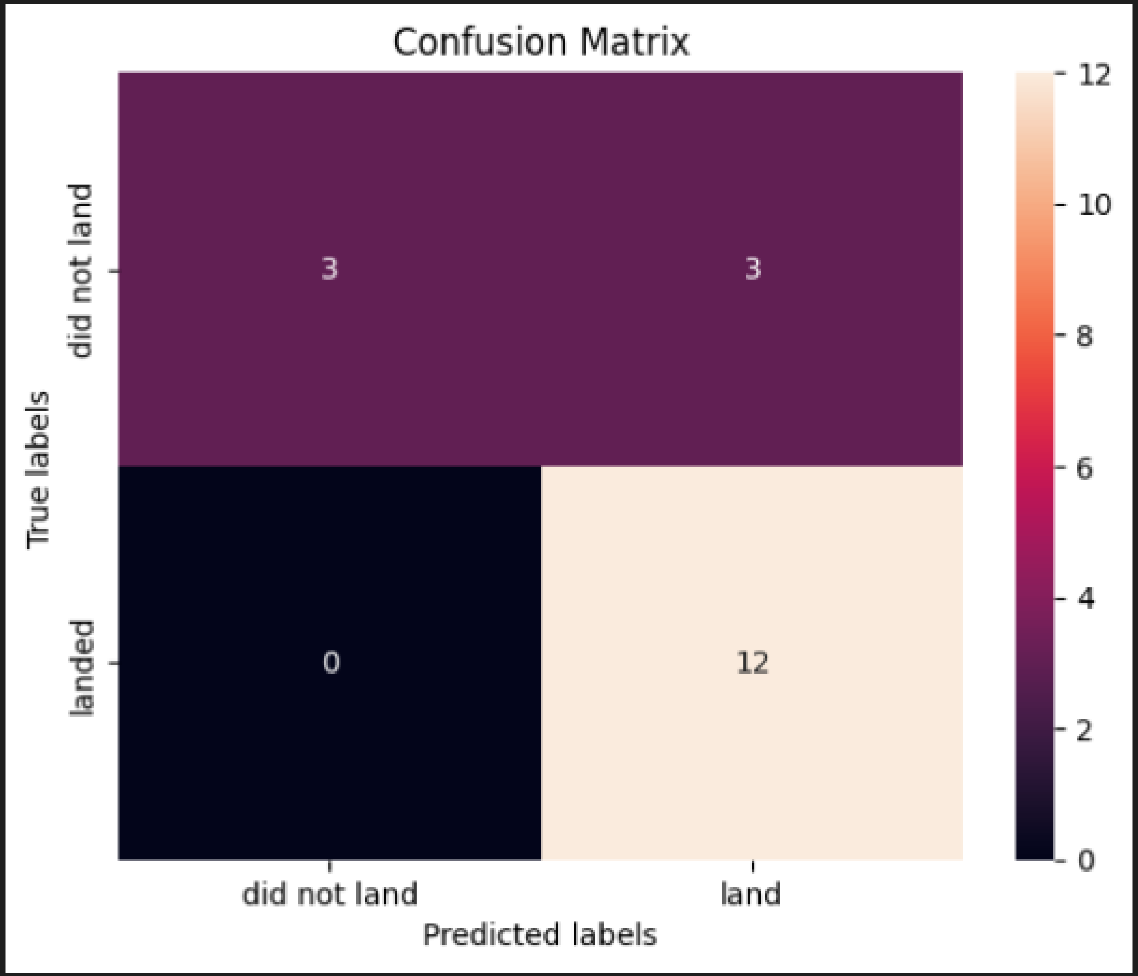


Machine Learning Prediction

CONFUSION MATRIX

```python
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```

Python



Conclusion

- Analysis reveals significant correlations between launch site and mission success. CCAFS LC-40 exhibits a lower success rate compared to KSC LC-39A and VAFB SLC-4E. Additionally, payload mass influences landing outcomes, with heavier payloads demonstrating a higher risk of failure. Orbit type also impacts success rates, with SO exhibiting the lowest success rate and ES-L1, GEO, HEO, and SSO demonstrating higher success probabilities.
- A clear upward trend in success rates is observed from 2013 to 2020. Utilizing a decision tree classifier with optimized parameters achieved an accuracy of 84% in predicting first-stage landing success.

Appendix Resources

- https://github.com/gpfalade/DS_Capstone_Project

THANK YOU