Applied Data Science Capstone

<Alpesh Chovatiya>

<17th Oct 2022>

https://github.com/alpdiv24/Data-Science-Capstone-Project

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- **Summary of methodologies**
  - Data collection methodology:
  - Perform data wrangling
  - Perform exploratory data analysis (EDA) using visualization and SQL
  - Perform interactive visual analytics using Folium and Plotly Dash
  - Perform predictive analysis using classification models
- **Summary of all results**
  - EDA results
  - Predictive Analysis results

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
Other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
We will predict if the Falcon 9 first stage will land successfully.

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data was collected by Rest API and Web Scrapping.

Rest API : Make a get request to SpaceX Rest API.

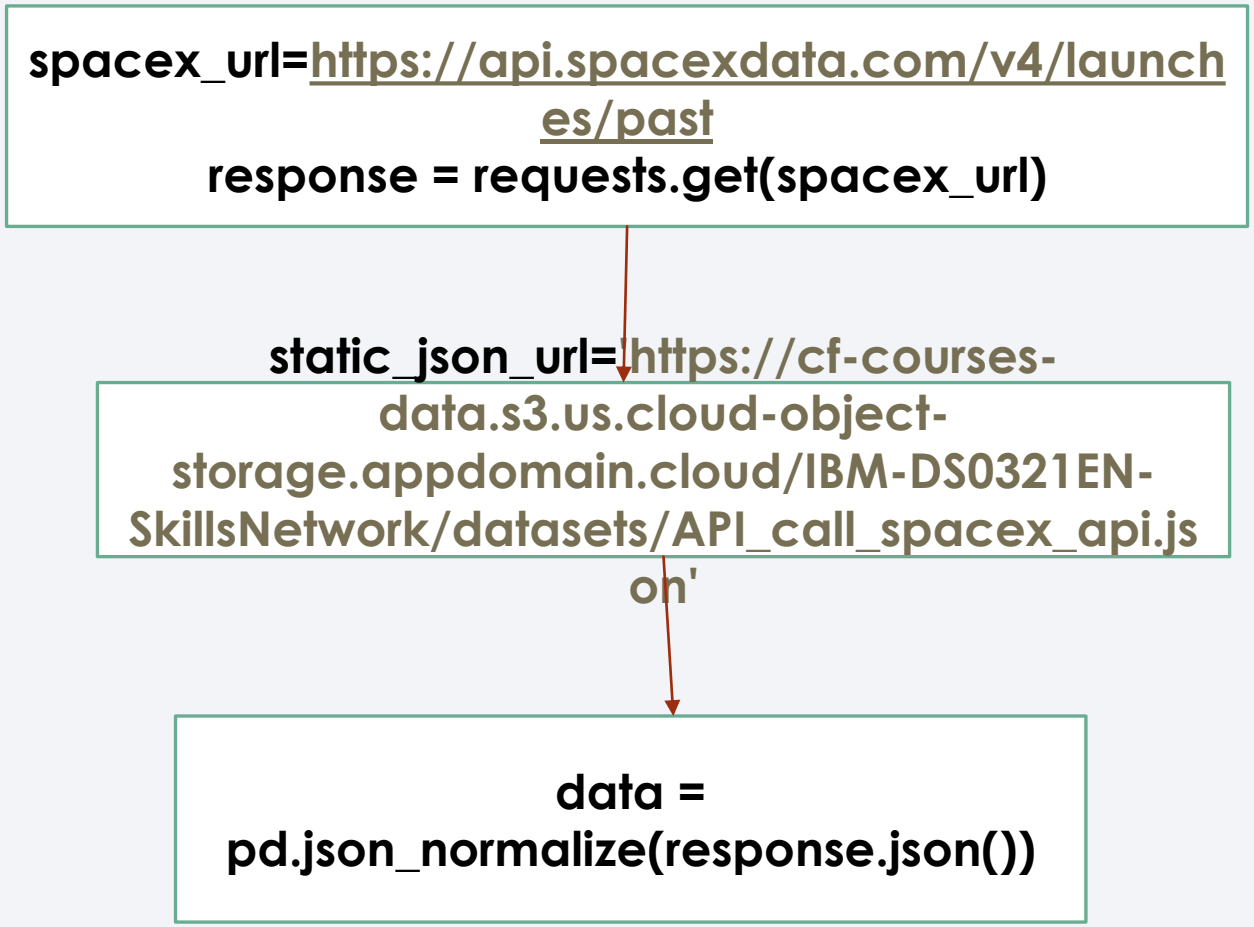
API returns the data in the form of JSON.

Transform data to a Dataframe using normalize method

Web Scrapping : Web Scrapping to collect Falcon 9 historical launch record from Wikipedia page.

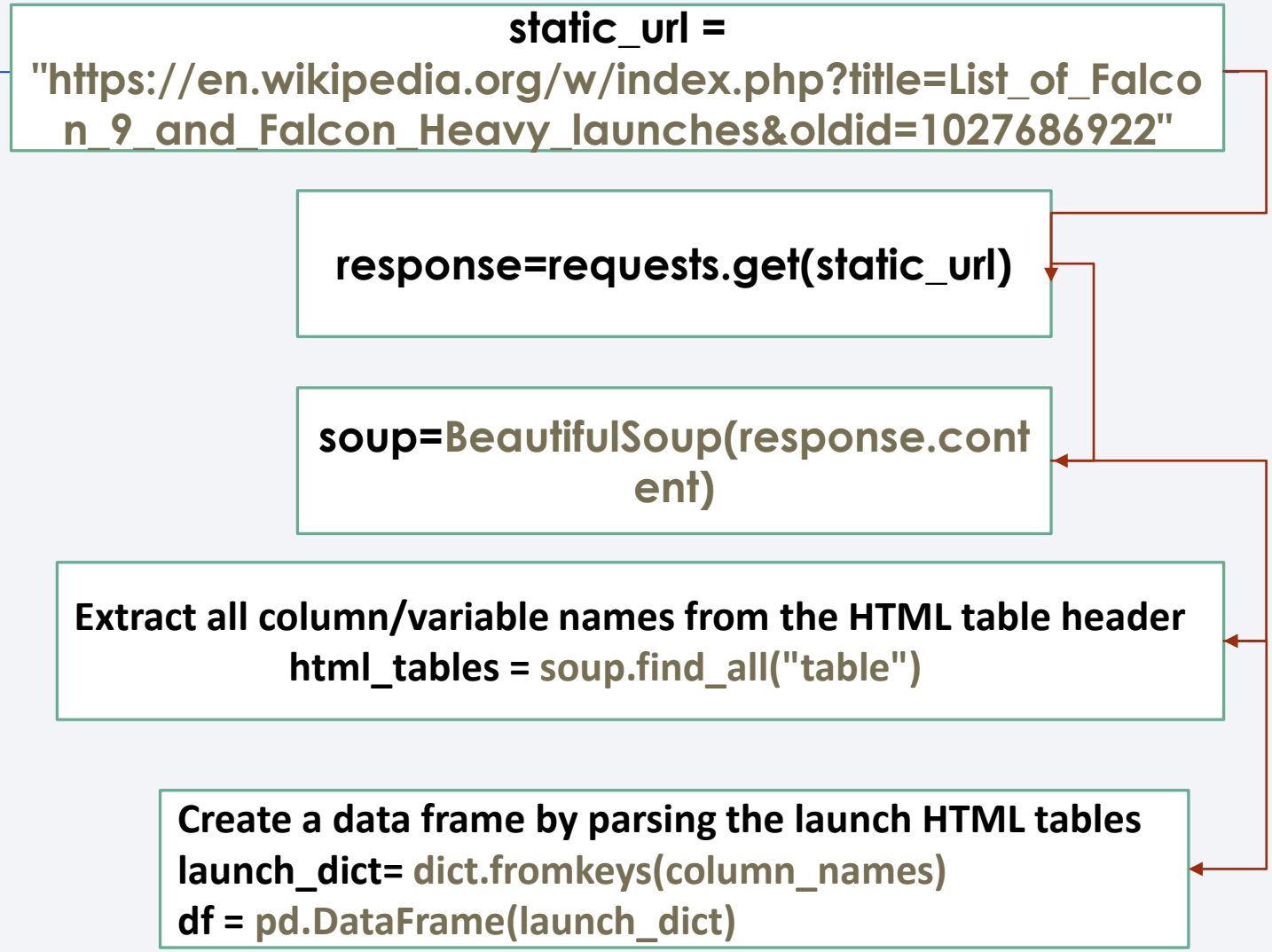
Data Collection – SpaceX API

- Make a get request to SpaceX Rest API.
- API returns the data in the form of JSON.
- Transform data to a Dataframe using the normalize method
- **GitHub URL =** https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/data-collection-api.ipynb



Data Collection – Web Scraping

- Web Scrapping to collect Falcon 9 historical launch record from Wikipedia page.
- **GitHub URL =** https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/webscraping.ipynb



Data Wrangling

- **In the dataset** there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad. True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.
- we will mainly convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.
- **GitHub URL =** https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/Data%20wrangling.ipynb

Data Wrangling

Perform Exploratory Data Analysis

Calculate the number of launches on each site

Calculate the number and occurrence of each orbit

In [5]:
```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

Out[5]:
```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

Each launch aims to an dedicated orbit, and here are some common orbit types:

Calculate the number and occurrence of mission outcome per orbit type

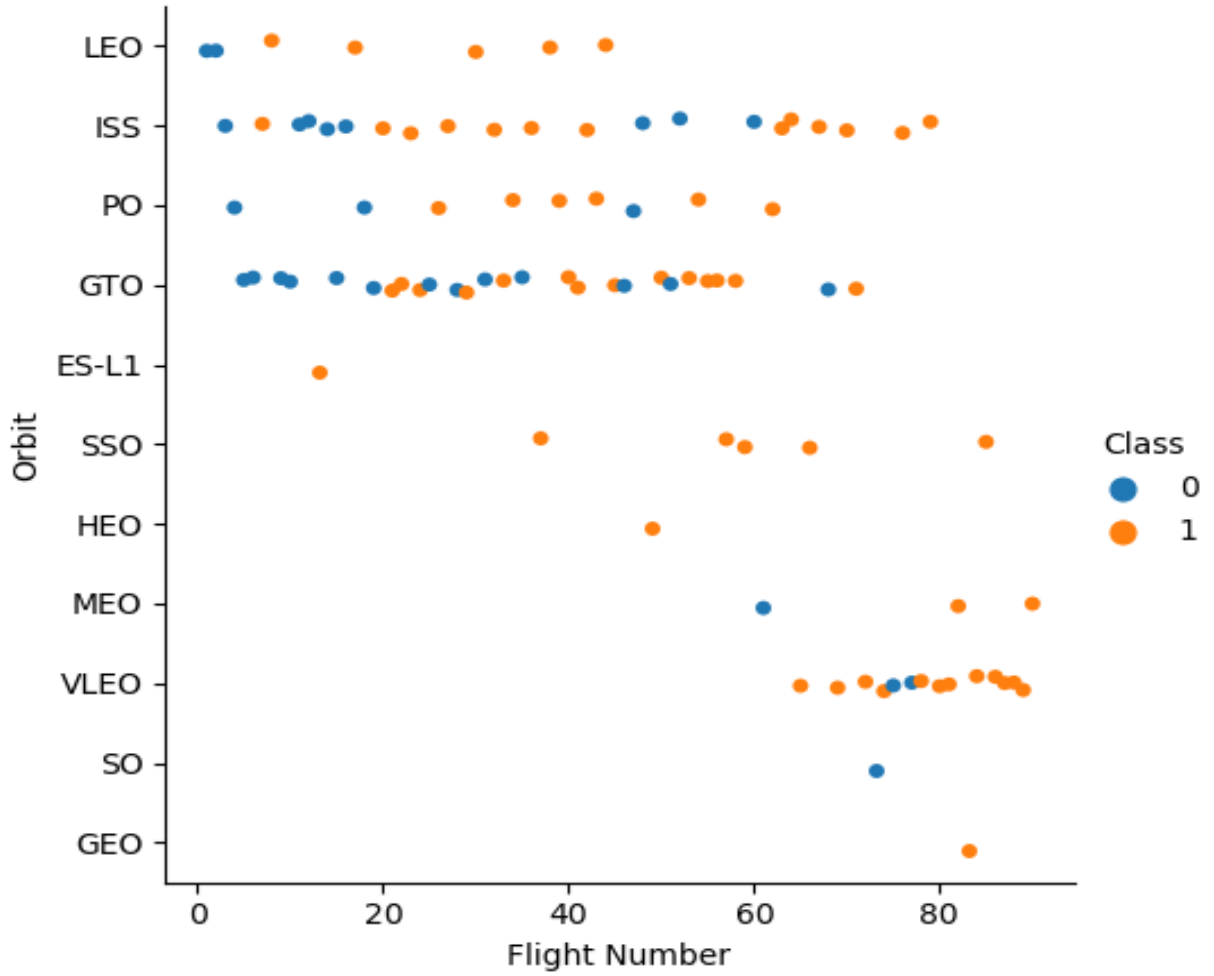
Create a landing outcome label from Outcome column

EDA with Data Visualization

- **Scatter Plot** : Visualize the relationship between two numeric variables.
  - FlightNumber vs. PayloadMass
  - FlightNumber vs LaunchSite
  - Payload vs Launch Site
  - FlightNumber vs Orbit type
  - Payload vs Orbit type
- **Bar Chart:** to visually check if there are any relationship between success rate and orbit type.
- **Line Chart:** To Visualize the launch success yearly trend

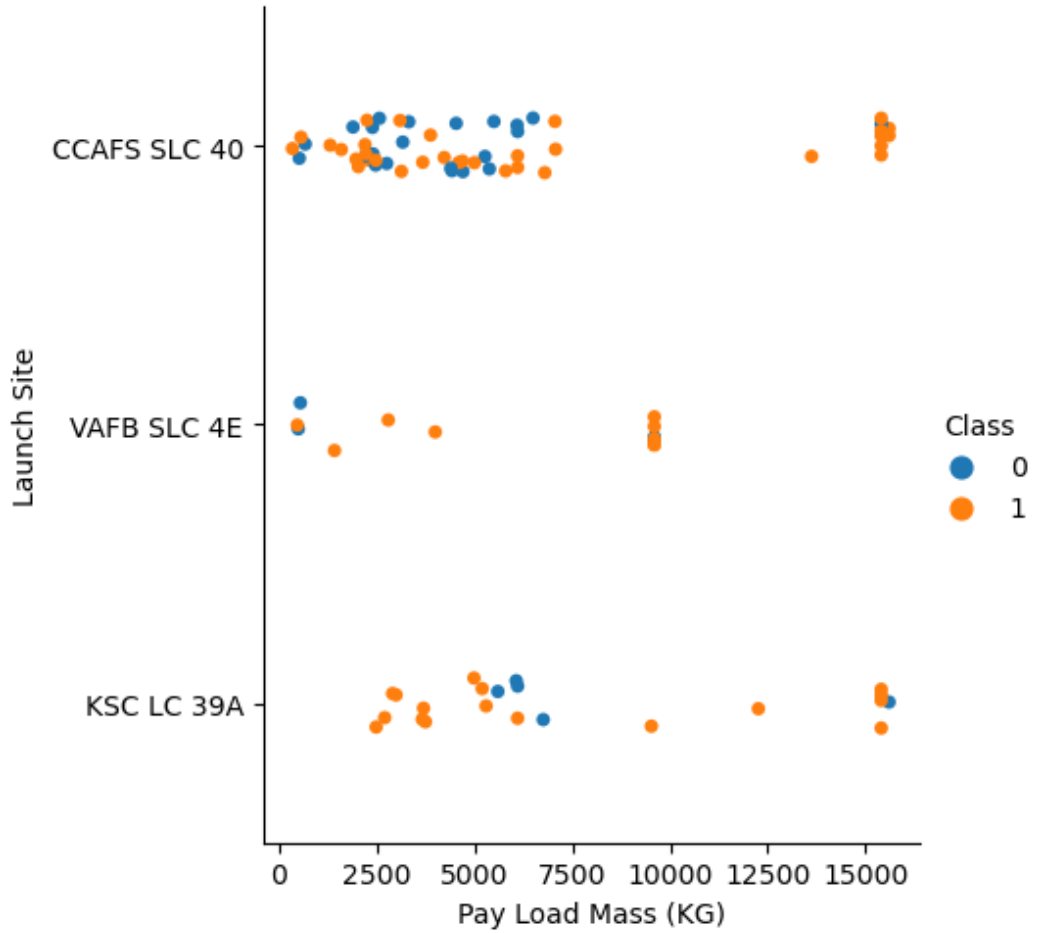
GitHub URL = https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/EDA-dataviz.ipynb

```
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
sns.catplot(x='FlightNumber',y='Orbit', data=df, hue='Class')
plt.xlabel("Flight Number")
plt.ylabel("Orbit")
plt.show()
```



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

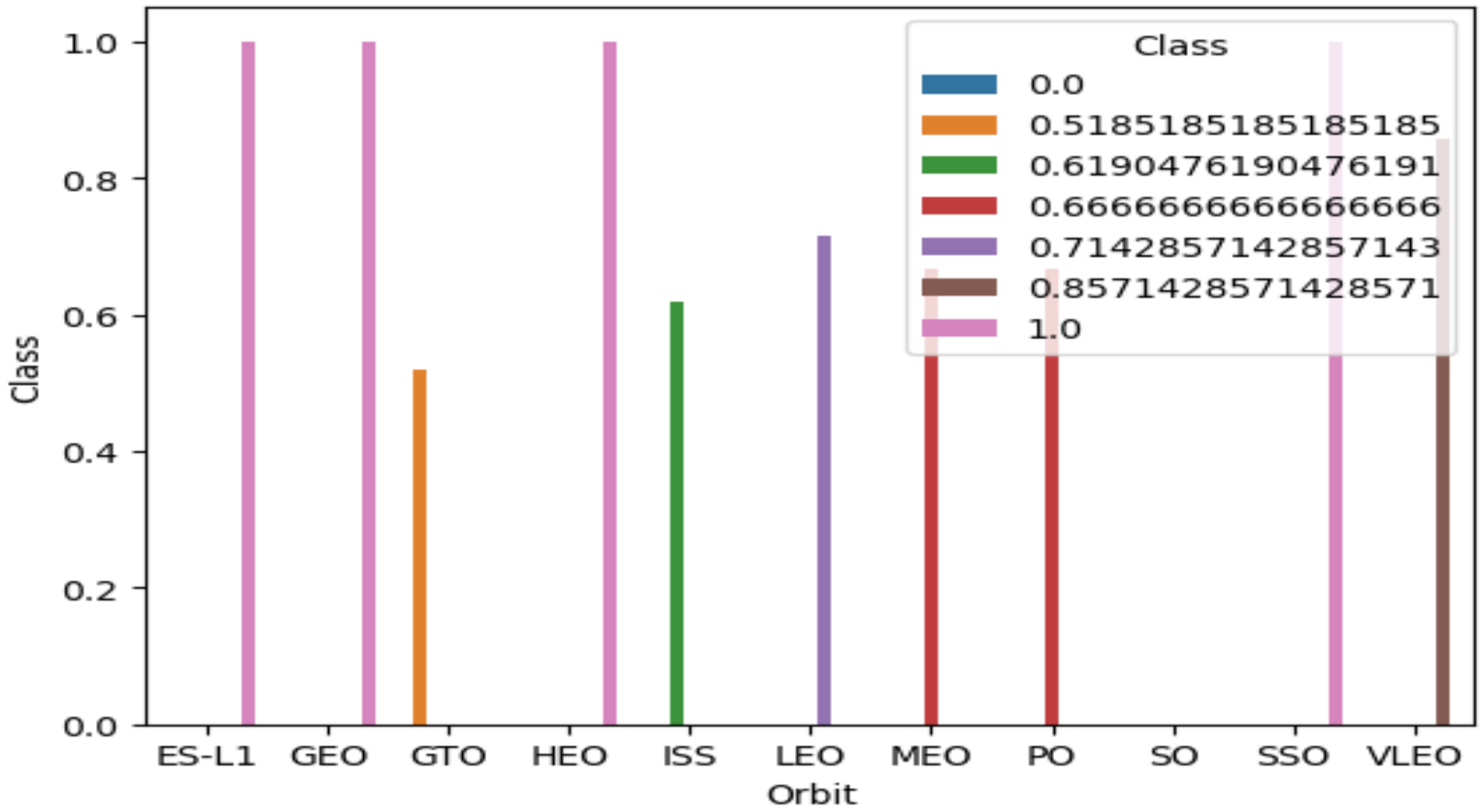
Payload vs. Launch Site



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

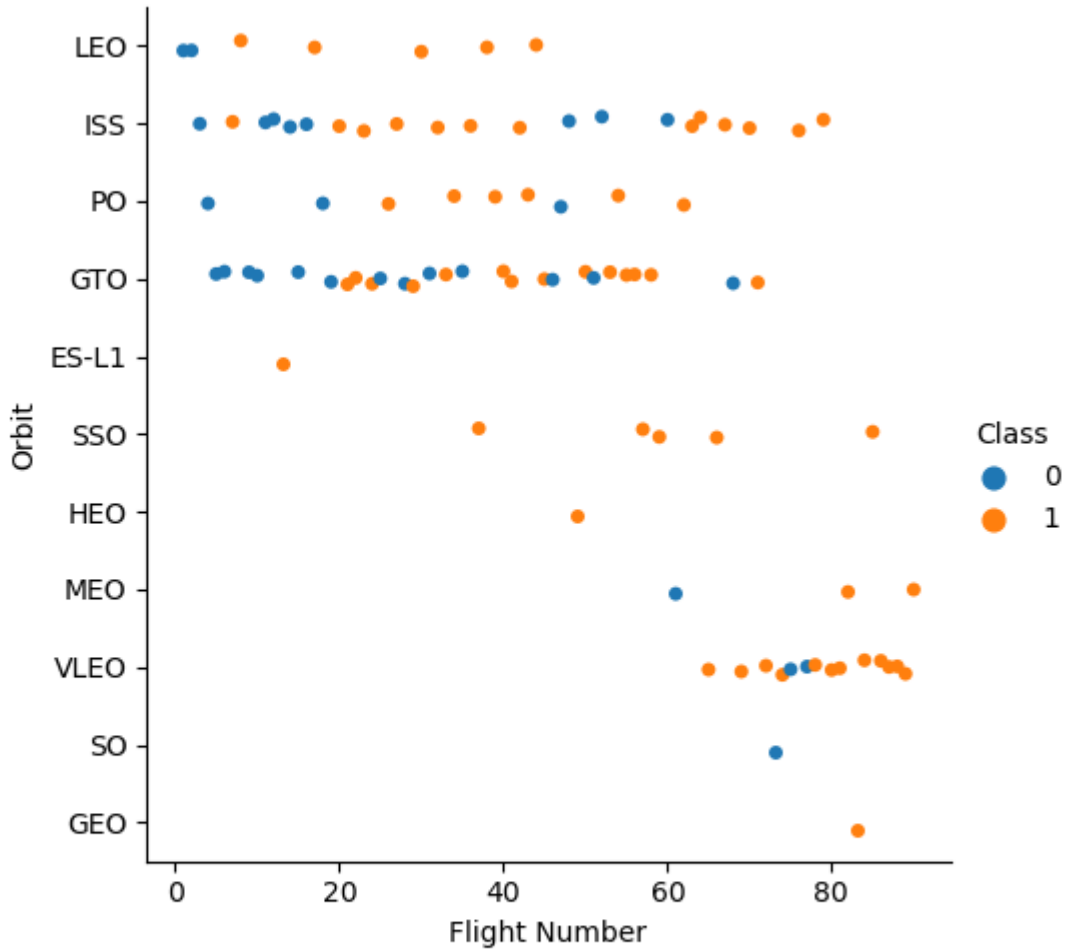
Success Rate vs. Orbit Type

Out[6]: <AxesSubplot:xlabel='Orbit', ylabel='Class'>



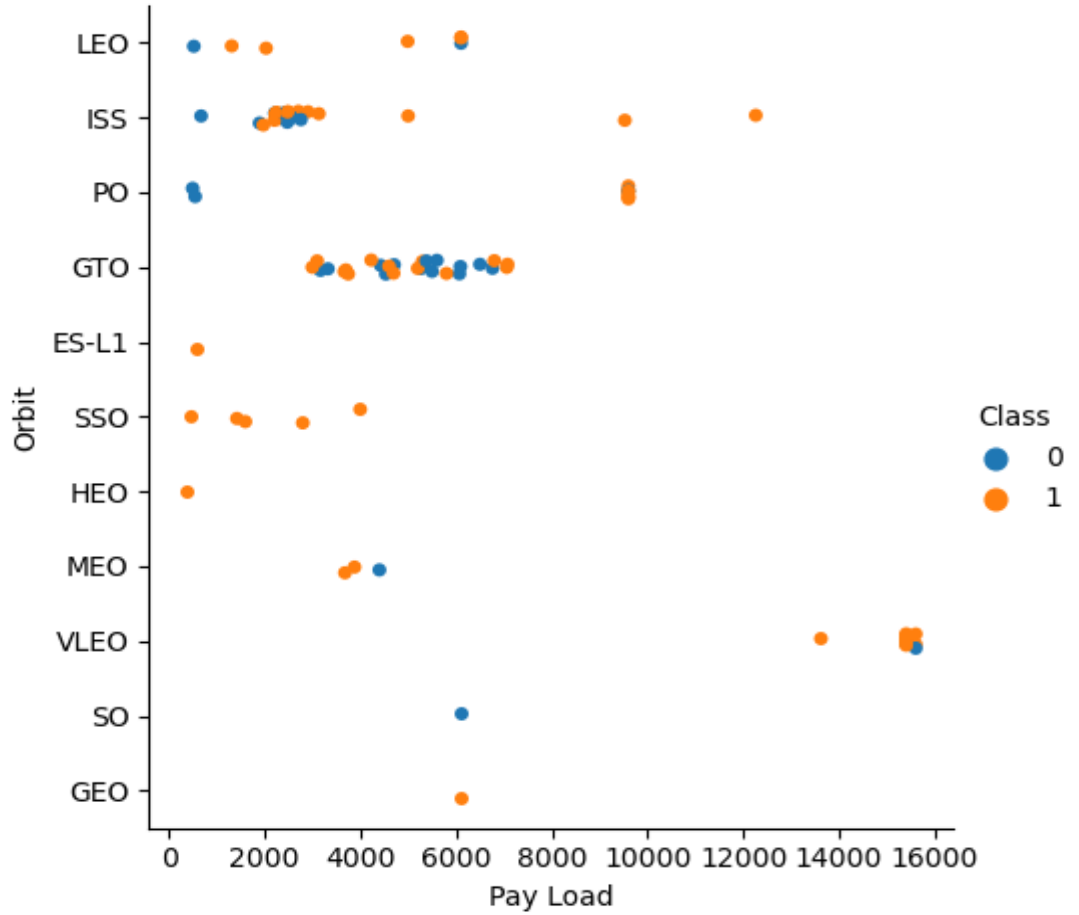
From the visualization, we can conclude that ES-L1, GEO,HEO & SSO have high success rate.

Flight Number vs. Orbit Type



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

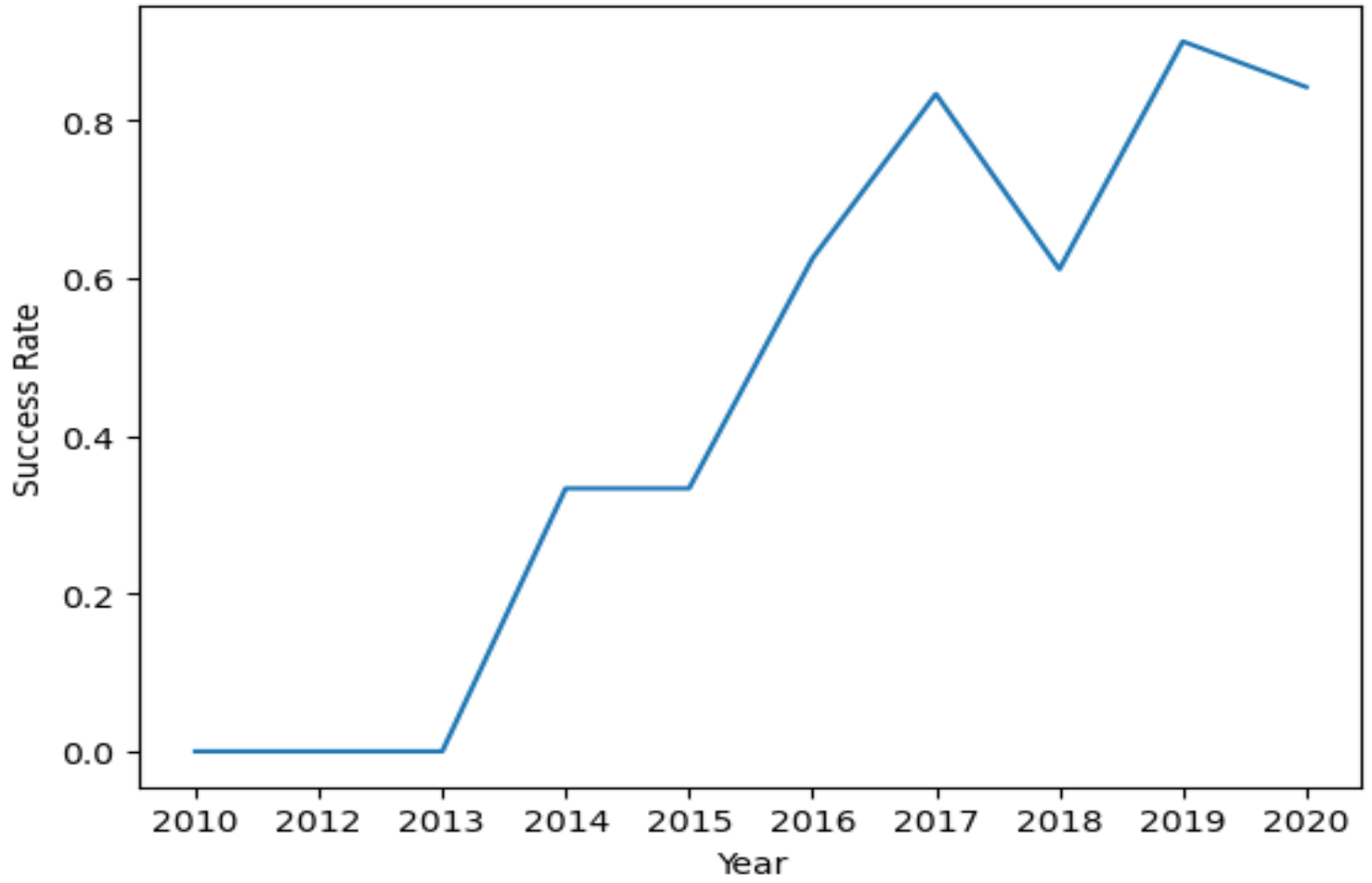
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



EDA with SQL

- Display the names of the unique launch sites in the space mission.
- Display 5 records where launch sites begin with the string 'CCA'.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1.
- List the date when the first succesful landing outcome in ground pad was acheived.
- 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 total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery.
- 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- GitHub URL = https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/da-sql-coursera_sqllite.ipynb

All Launch Site Names

- Find the names of the unique launch sites

In [7]:
```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

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

Out[7]:

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`

In [8]:
```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE "CCA%" LIMIT 2;
```

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

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)

Total Payload Mass

- Calculate the total payload carried by boosters from NASA (CRS)

```
In [9]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = "NASA (CRS)" ;
```

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

Out[9]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

Task 4

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

In [10]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version = "F9 v1.1" ;
```

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

Out[10]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

In [11]:
```
%sql SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (ground pad)";
```

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

Out[11]:

MIN(Date)
01-05-2017

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

In [12]:
```
%%sql SELECT "Booster_Version"  FROM SPACEXTBL
WHERE "Landing _Outcome" = "Success (drone ship)"
AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) ;
```

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

Out[12]:

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

In [14]:
```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

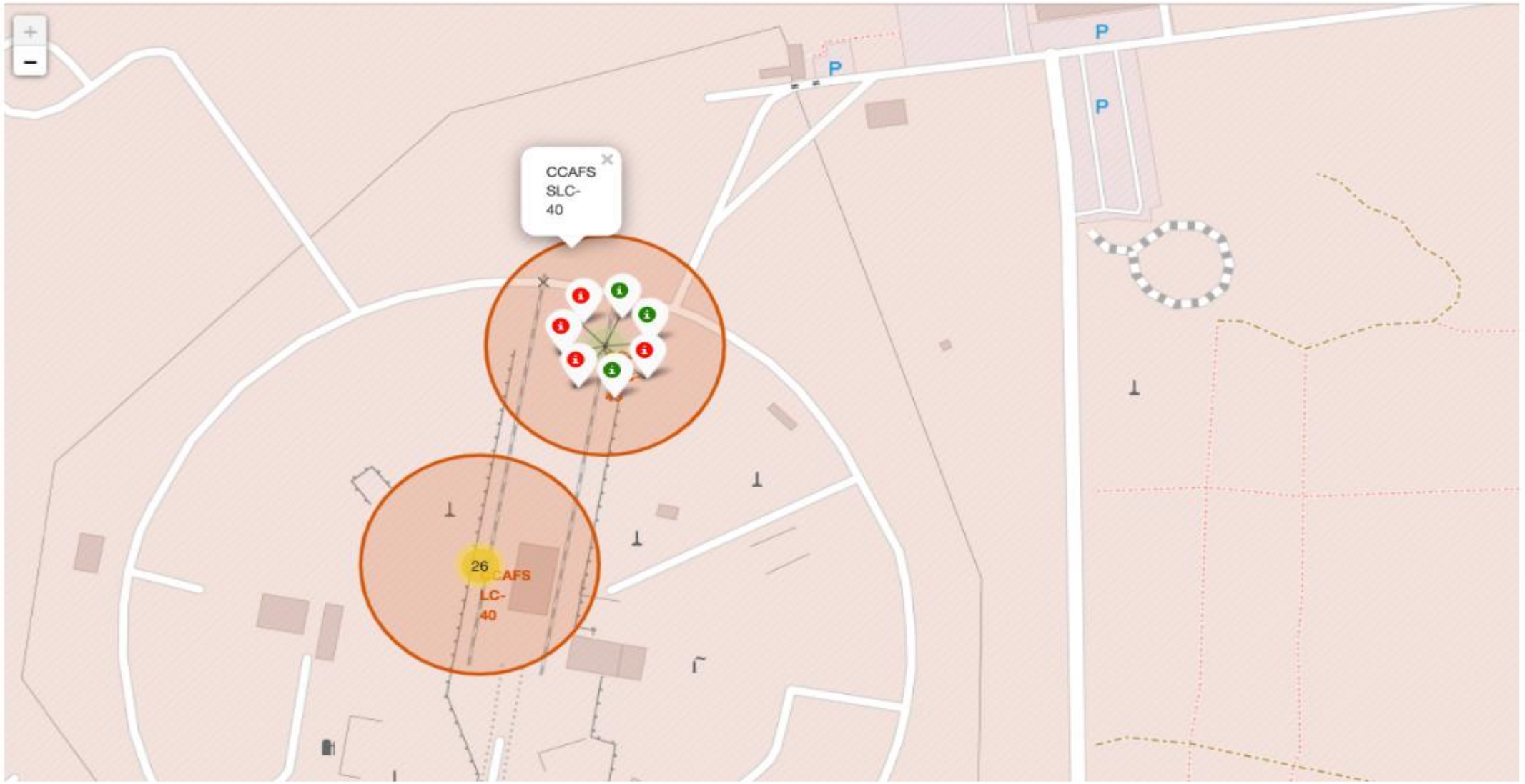
Out[14]:

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

Build an Interactive Map with Folium

- Folium library enables interactive map visualizations in Python and I used Map() function to create a Map.
- I created circles and markers objects and added to folium map.
- I used circle() function to circle the coordinates and marker() function to mark the location of the coordinates.
- **GitHub URL =** https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/Site_location_Folium.ipynb

Folium Map with Markers



From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

Folium Map With Distant Line

Build a Dashboard with Plotly Dash

- **Plotly Dash application** for users to perform interactive visual analytics on SpaceX launch data in real-time.
- This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.
- This allows easy hover, click and select actions on graphs.

GitHub URL = **https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/spacex_dash_app.py**

SpaceX Launch Records Dashboard

All Sites

Launch Success Rate For All Sites

SpaceX Launch Records Dashboard

VAFB SLC-4E

Launch Success Rate For VAFB SLC-4E

Predictive Analysis (Classification)

- We will create different machine learning models to predict if the first stage will land given the preceding data.
- **Models built:**
  - Logistic Regression
  - K Nearest Neighbors (KNN)
  - Support Vector Machines (SVM)
  - Decision Tree

GitHub URL = https://github.com/alpdiv24/Data-Science-Capstone-Project/blob/main/Machine%20Learning%20Prediction.ipynb

Predictive Analysis (Classification)

1. Collect & Load Data
2. Standardize Data
3. Split data into training data and test data
4. Choose Machine Learning Algorithm
5. Training the model
6. Evaluating the model
7. Predictions

Results

- Predictive analysis results

Find the method performs best:

Models with High Accuracy Score:

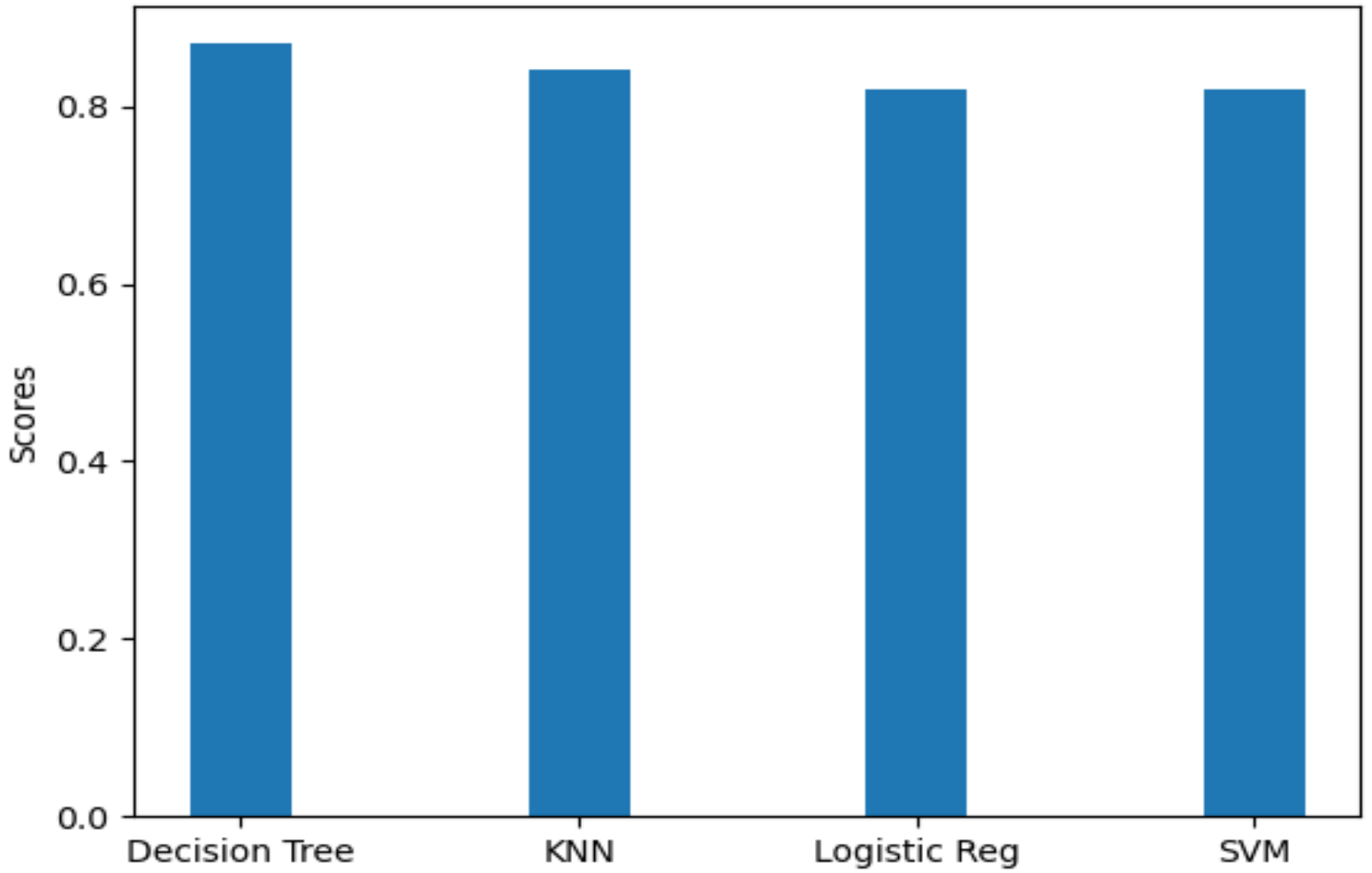
Decision Tree: 0.87

KNN: 0.84

Logistic Regression: 0.82 ¶

SVM: 0.82

Classification Accuracy

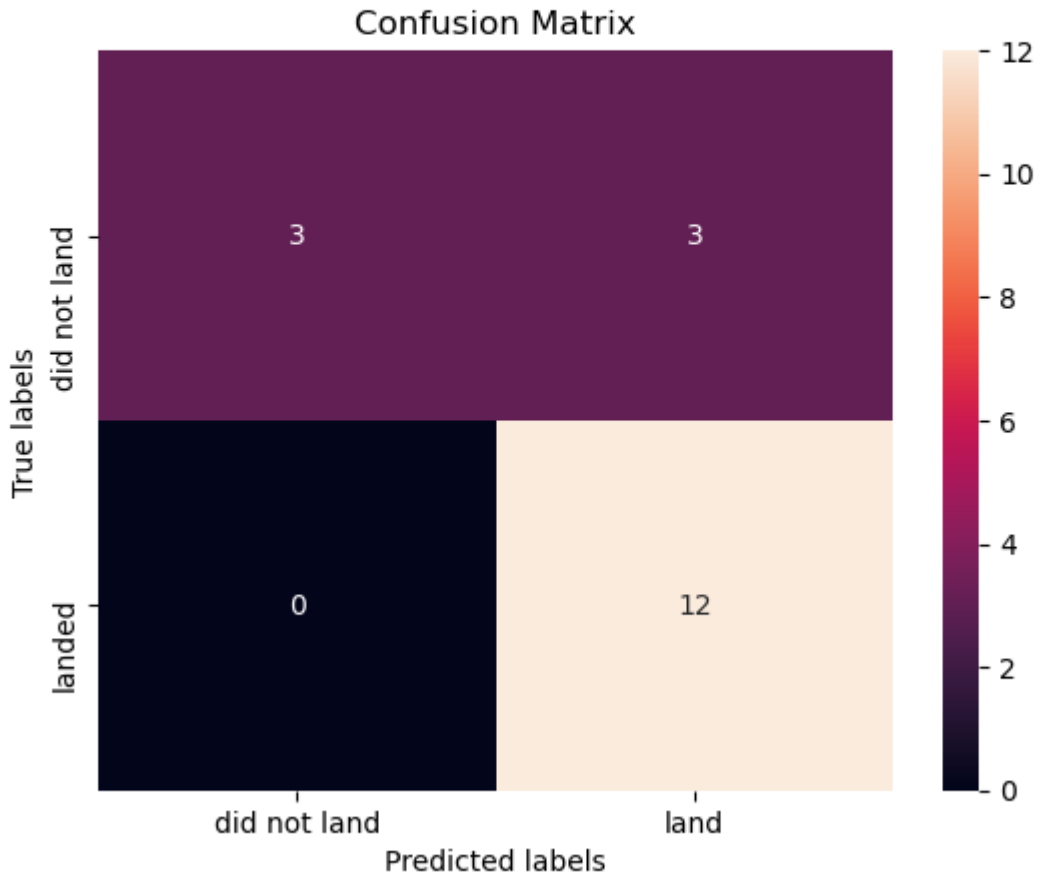


From the given bar chart, we can say that Decision Tree has a highest accuracy score in comparison to other models.

Confusion Matrix

We can plot the confusion matrix

In [27]:
```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- Decision Tree Algorithm has a highest Accuracy rate and it is a best suitable machine learning model for given dataset.
- KSC LC - 39A has a highest successful launches from All sites.
- Launch success rate has been increasing with every years.
- ES-L1, GEO, HEO & SSO have high success rate than other orbit types.
- With heavy playloads the successful landing or positive landing rate are more with LEO and ISS.

THANK YOU