Per aspera ad astra

"Winning Space Race
with Data Science"

Karol Bohdan Krawczyk
02.III.2025

OUTLINE

- Excecutive summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Summary of Methodologies:

Collecting data from a website using API

Collecting data using Web Scrapping

Data processing – Data wrangling

Data visualization inference

Data comparison

Creating machine learning models

Summary of all results:

Predictive analysis result

Finding correlation

INTRODUCTION

Project background and context

SpaceX is a controversial company worldwide, regardless of which side we are on, I think that each of us agrees on how big a breakthrough it has made in both space technology and the progress of civilization.

In such ventures as the conquest of another galaxy, there could not be no losses, errors or random situations - each of these events gives us data that, properly processed, interpreted and assimilated, gives us the fullest possible understanding of what is happening and why.

Problems I want to find answers to

In my work, I decided to find answers to the following questions:

1. In which orbits were the most successful SpaceX missions
2. Is there a relationship between the launch platform and successful missions
3. What is the trend of SpaceX rocket launch successes
4. What is the accuracy of the "Logistic Regression" model in assessing rocket landings

Section 1

Methodology

METHODOLOGY

Executive Summary

1.Data collection

Data was collected using SpaceX API and Web Scrapping from Wikipedia

2.Data Wrangling

Processed data to create orbital location

3.Visualization

Visualization was performed using standard libraries

4.Predicitve analysis

Predictive analysis was carried out in the form of creating and testing a landing assessment model.

DATA COLLECTION

The goal of the task was to retrieve SpaceX rocket launch data using a GET request to the API and transform the result into a Pandas DataFrame that will allow for further analysis.

- The URL (static_json_url) points to a static JSON file containing SpaceX rocket launch data.
- It is stored in the cloud and publicly available.
- Using the requests library to send a GET request to the provided URL.
- The result is assigned to the response variable.
- The data returned by the API is in JSON format.
- We retrieve it as a JSON object and assign it to result.

- Once we have the JSON data, we convert it to a Pandas DataFrame using pd.json_normalize().

- json_normalize() converts nested JSON structures to a tabular format.

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
[12]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.js
```

We should see that the request was successfull with the 200 status response code

```
[13]: response=requests.get(static_json_url)
```

```
[14]: response.status_code
```

```
[14]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
[15]: result = requests.get(static_json_url)
print(result.content)
```

...

Using the dataframe `data` print the first 5 rows

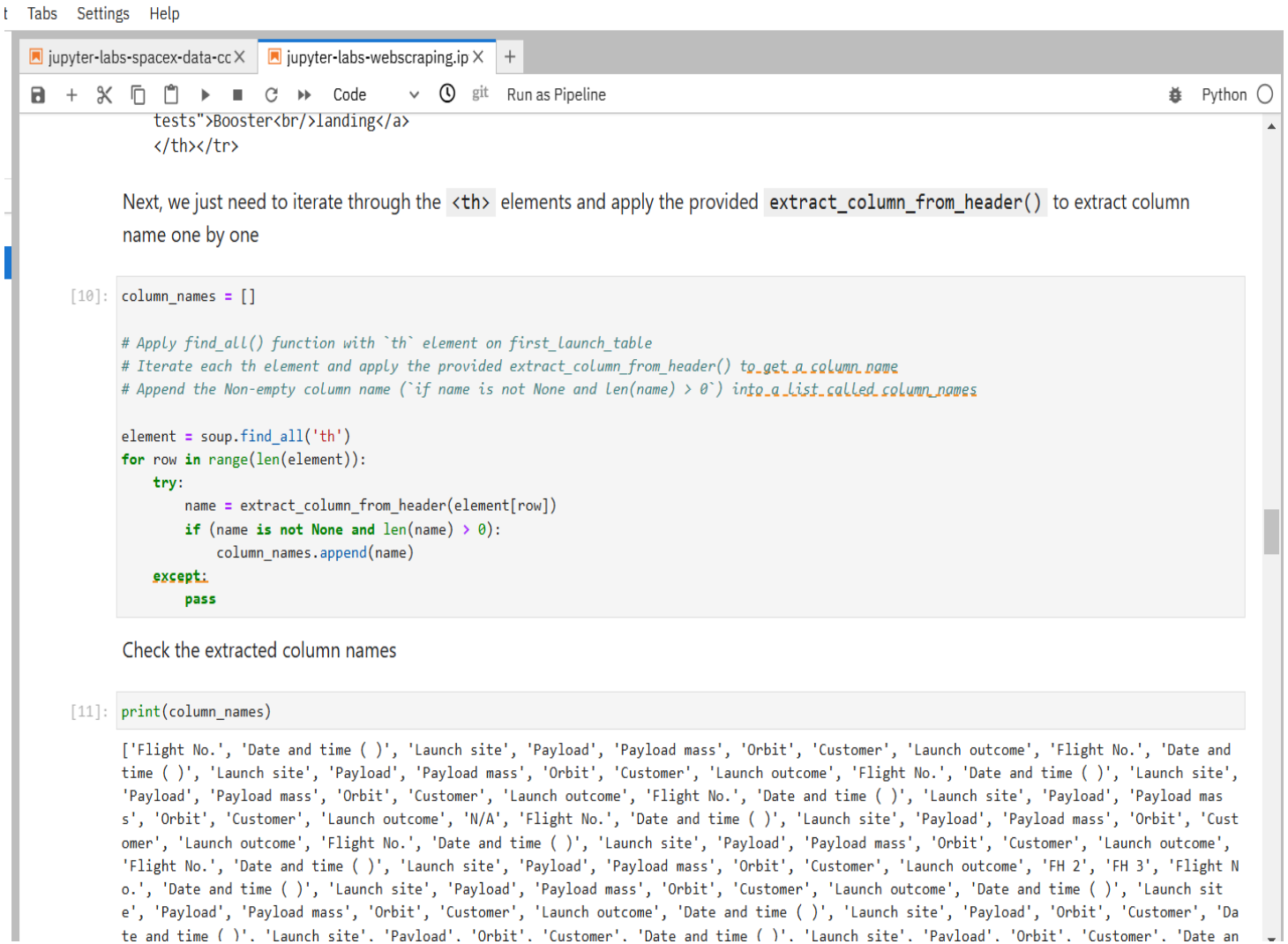
```
[20]: static_json_df = result.json()
data = pd.json_normalize(static_json_df)

data.head(5)
```

DATA COLLECTION - SCRAPPING

- Using web scraping to retrieve Falcon 9 launch information using BeautifulSoup

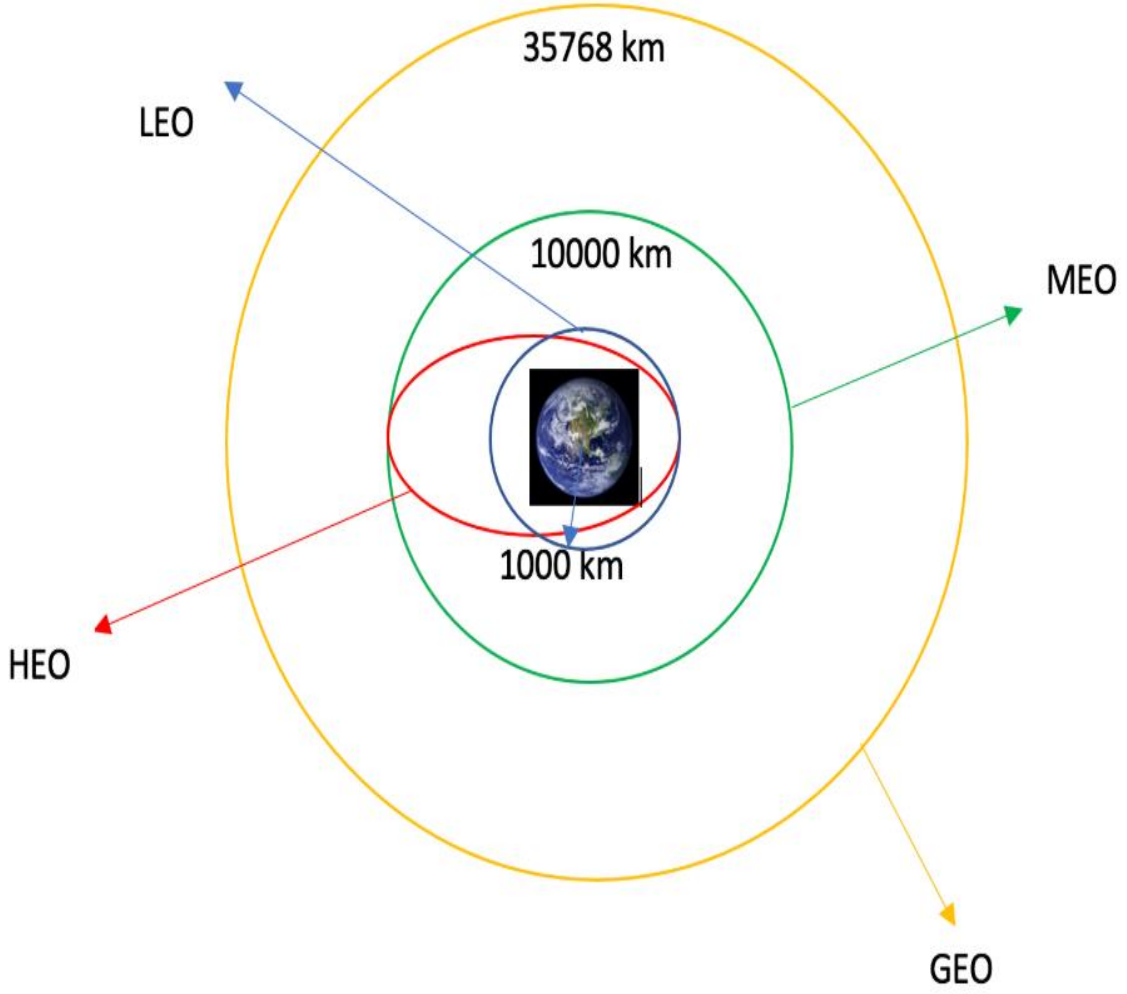
- We extract only the headers containing <th>, iterate through the table to discard those with empty records, and then convert it to Pandas format



DATA WRANGLING

- Data wrangling allows us to more easily analyze orbits from Earth and the distances covered by individual SpaceX missions.

- 1: Calculate the number of launches on each site
- 2: Calculate the number and occurrence of each orbit
- 3: Calculate the number and occurrence of mission outcome per orbit type
- 4: Create a landing outcome label from Outcome column



EDA WITH DATA VISUALIZATION

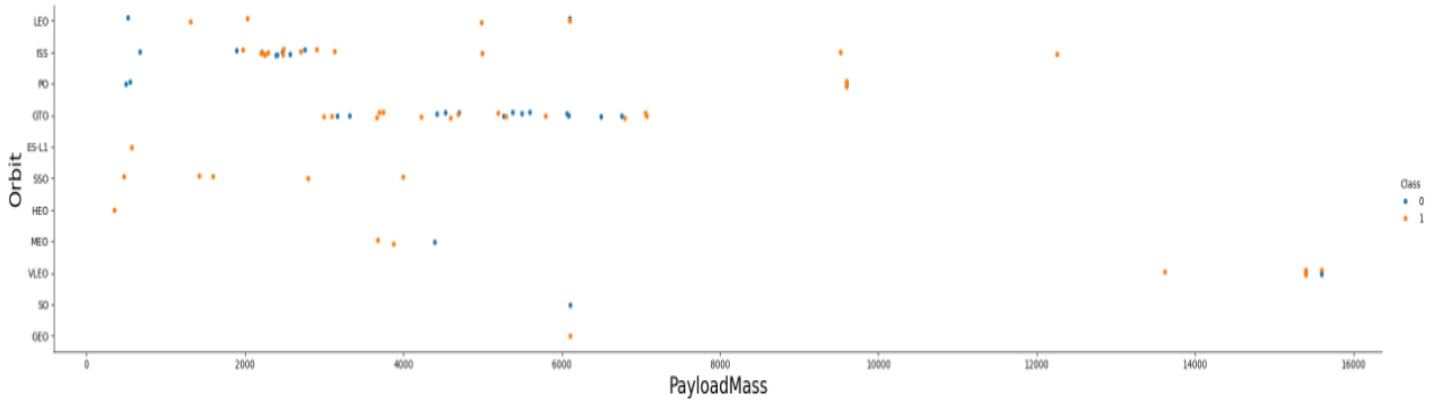
- The analysis aimed to investigate how the Payload Mass (Payload Mass) affects the orbit type (Orbit) and the mission success (Class), which is marked with colors (0 - failed, 1 - successful).
- The graph shows that for different orbits (LEO, GTO, ISS, MEO, etc.) there are specific ranges of payload mass.
- It can be seen that in some orbits (e.g. ISS, LEO there are more successful missions (more poir marked as 1).
- The analysis shows that the payload mass affects the choice of orbit, but is not directly related to the mission success.

You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

TASK 5: Visualize the relationship between Payload Mass and Orbit type

Similarly, we can plot the Payload Mass vs. Orbit scatter point charts to reveal the relationship between Payload Mass and Orbit type

[9]:
```
# Plot a scatter point chart with x axis to be Payload Mass and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

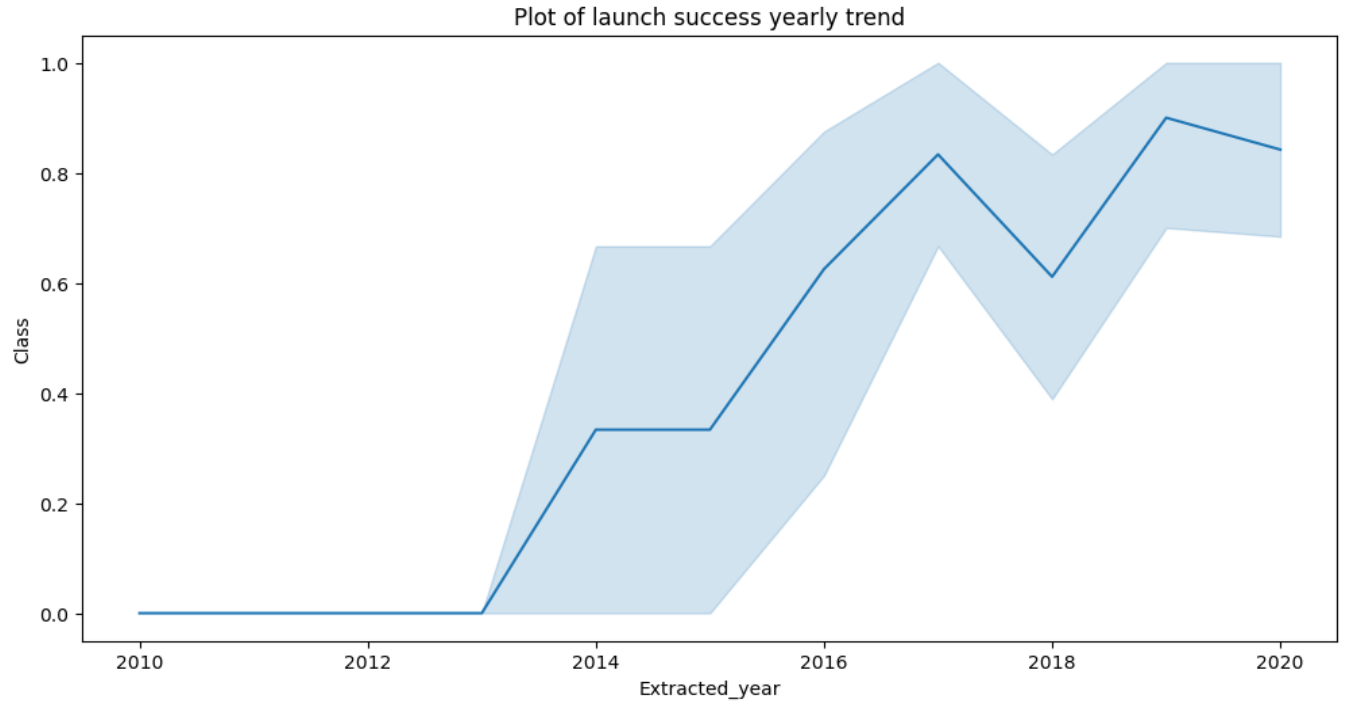


EDA WITH DATA VISUALIZATION

- Since 2013, there has been a significant improvement in the success rate of rocket launches, suggesting technological advances and better procedures.
- Since 2016, successes have become the norm, and the trend shows high reliability of launch systems.
- In 2020, the success rate was close to 100%, suggesting stability of technology and experience in rocket operations.

```
df_copy['Extracted_year'] = pd.DatetimeIndex(df['Date']).year

# plot line chart
fig, ax=plt.subplots(figsize=(12,6))
sns.lineplot(data=df_copy, x='Extracted_year', y='Class')
plt.title('Plot of launch success yearly trend');
plt.show()
```

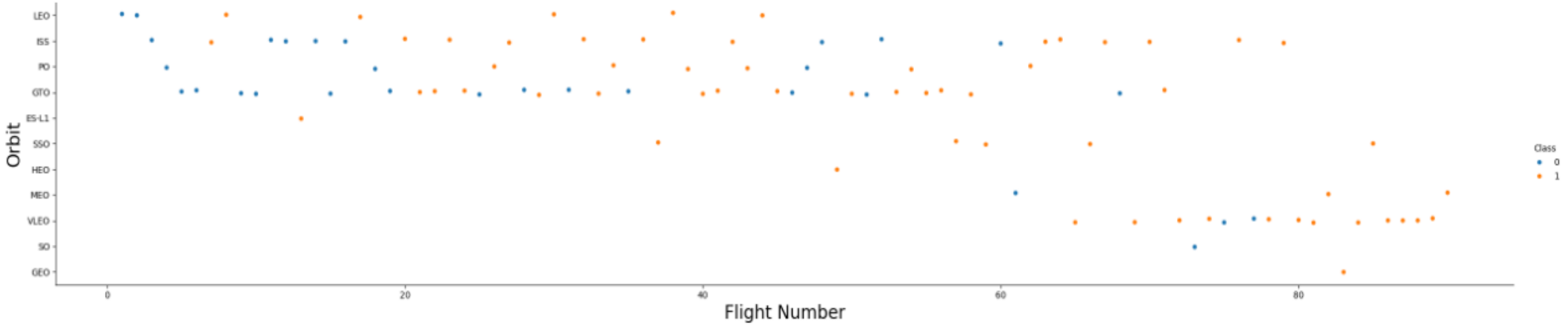


you can observe that the sucess rate since 2013 kept increasing till 2020

THERE IS A CLEAR CORRELATION BETWEEN MORE FLIGHTS AT THE LAUNCH SITE AND A GREATER CHANCE OF MISSION SUCCESS.

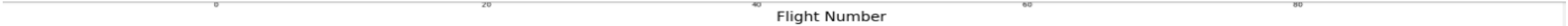
For each orbit, we want to see if there is any relationship between FlightNumber and Orbit type.

[8]:
```
# 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(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship

THE MOST SUCCESSFUL MISSIONS ARE SEEN IN THE RANGE OF 2000 TO 6000 PAYLOAD MASS

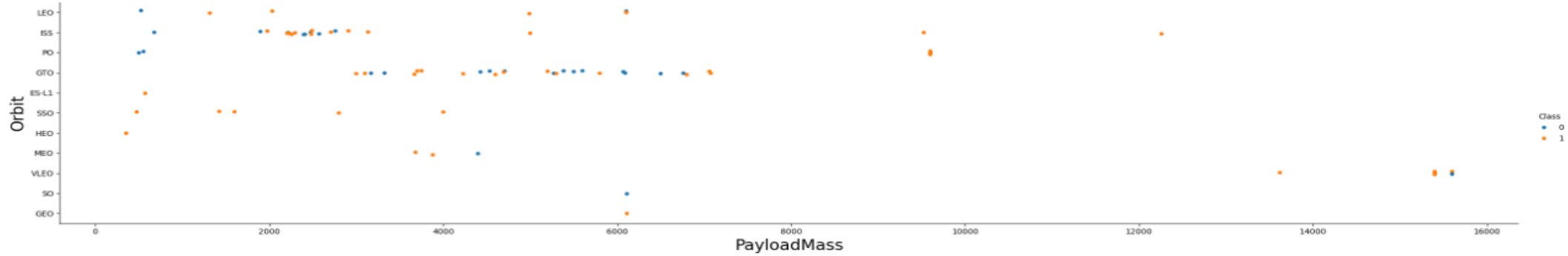


You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

TASK 5: Visualize the relationship between Payload Mass and Orbit type

Similarly, we can plot the Payload Mass vs. Orbit scatter point charts to reveal the relationship between Payload Mass and Orbit type

[9]:
```
# Plot a scatter point chart with x axis to be Payload Mass and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



Section 2

Predictive Analysis & Results

PREDICTIVE ANALYSIS

- The model achieved an accuracy of 83.3% (0.833).
- This means that the model correctly classifies rocket landings 83.3% of the time.
- 3 cases were correctly classified as "did not land" (True Negative).
- 3 cases were incorrectly classified as "landed" even though the rocket did not actually land (False Positive).
- 12 cases were correctly classified as "landed" (True Positive).
- 0 cases were incorrectly classified as "did not land" even though the rocket actually landed (False Negative).

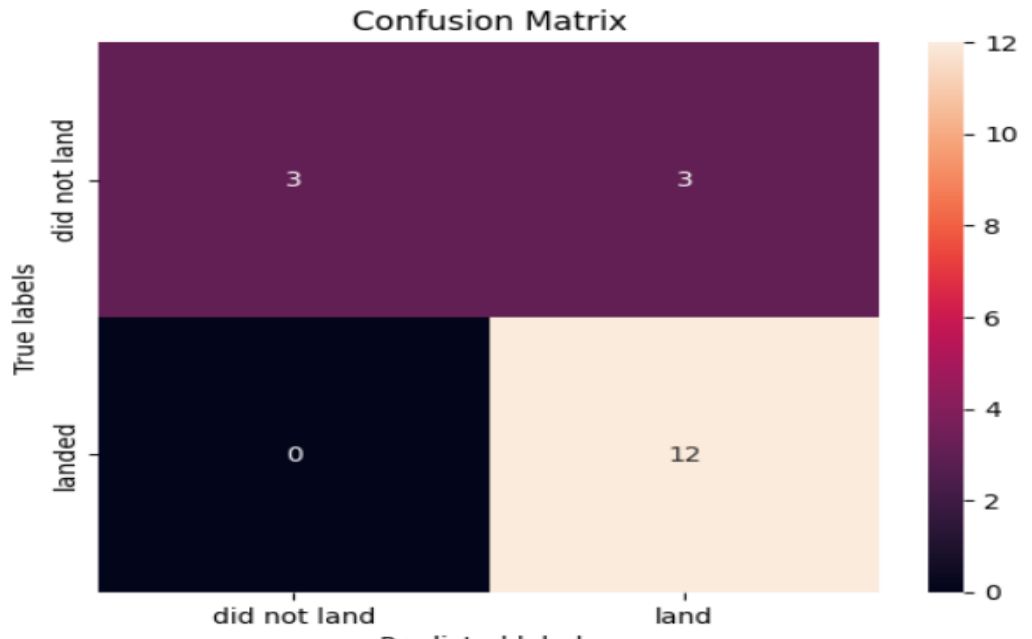
Calculate the accuracy on the test data using the method `score`:

```
[15]: print('Accuracy = {:.3f}'.format(logreg_cv.score(X_test, Y_test)))
```

```
Accuracy = 0.833
```

Lets look at the confusion matrix:

```
[16]: yhat=logreg_cv.predict(X_test)
      plot_confusion_matrix(Y_test,yhat)
```



RESULT

The methodology was based on the collection, processing and analysis of historical rocket launch data. Visualization and data mining techniques were used to identify key trends and relationships. The results of this analysis can help to better understand the factors that affect the success of space missions and optimize future launches.

Answers to my questions

1. The data analysis shows that LEO (Low Earth Orbit) had the most successful launches

2. The data suggests that the KSC LC-39A platform had the highest success rate

3. The trend shows a steady increase in the number of successful launches since 2013.

4. The "Logistic Regression" model can be accurate in the range of 80-90% with the right input data.