# WINNING SPACE RACE WITH DATA SCIENCE

Razan Albeladi

# Introduction

## Overview:

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.

## Goal:

- determine if the first stage will land
- determine the cost of a launch

# the link to repository

GitHub

# Data Collection

Data collection- web scraping → GitHub

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Data Collection API -> GitHub

- Request to the SpaceX API
- Clean the requested data

# Data Wrangling

GitHub

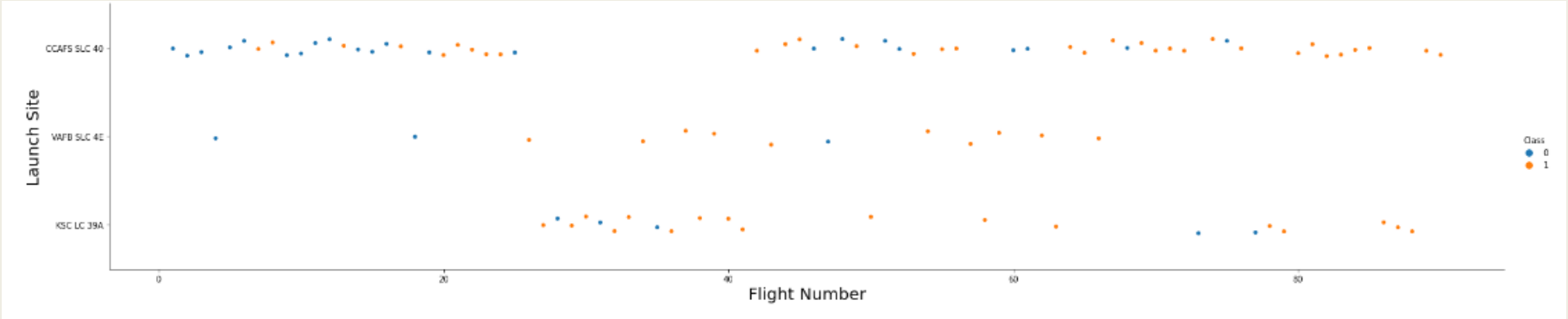
In this lab I performed exploratory Data Analysis and determine Training Labels

# Exploratory Data Analysis

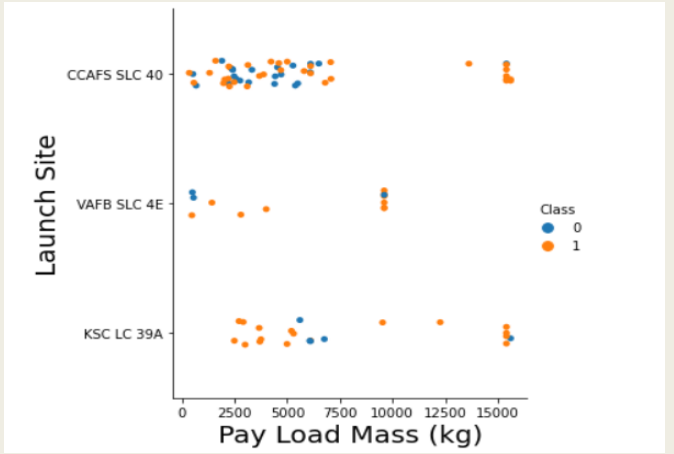
- Exploratory Data Analysis using Pandas and Matplotlib
- Exploratory Data Analysis using SQL

# Exploratory Data Analysis using Pandas and Matplotlib

- Relationship between flight number and launch site.

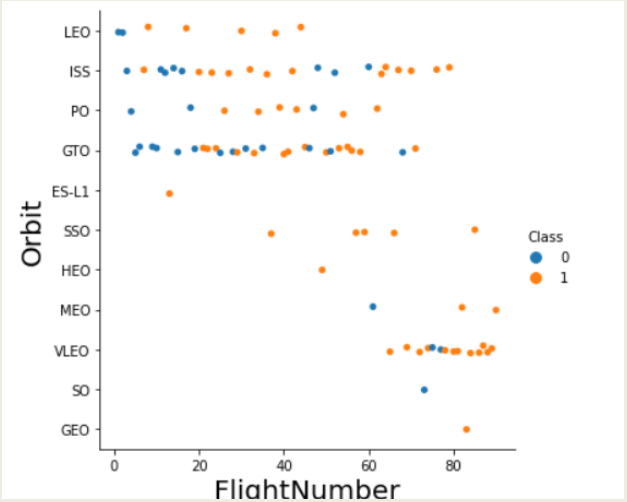



- Relationship between payload and launch site.

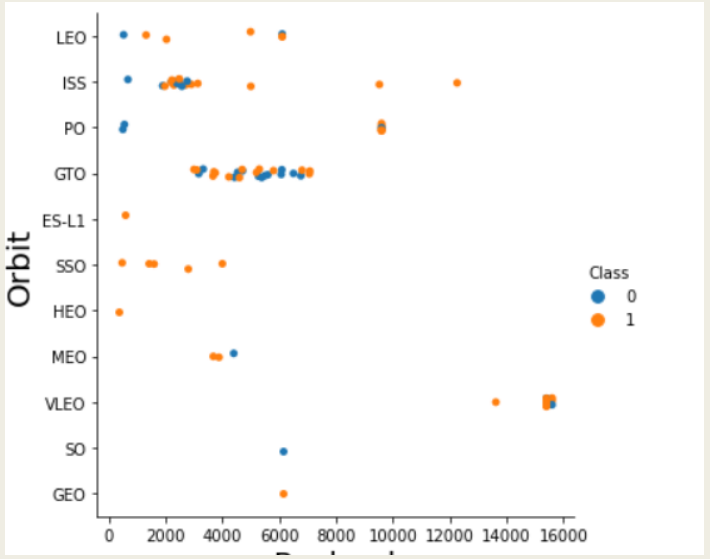


# Exploratory Data Analysis using Pandas and Matplotlib

- Relationship between orbit and flight number.

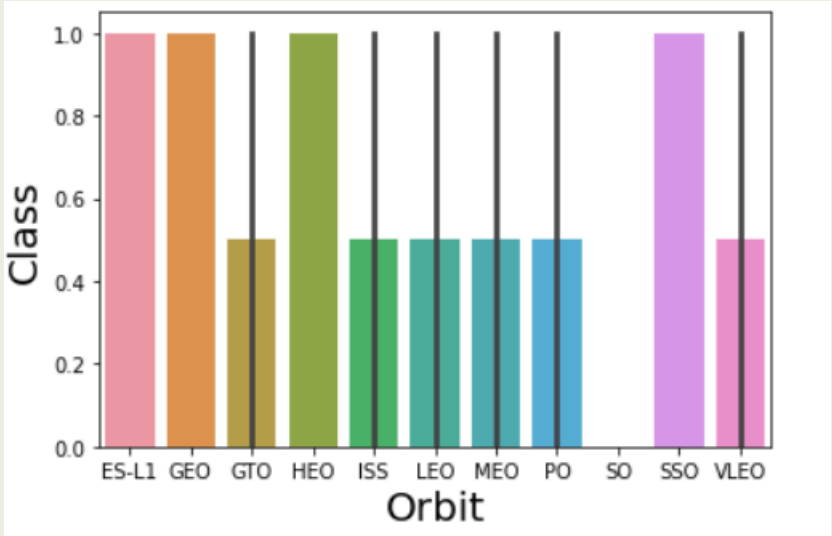



- Relationship between orbit and payload.

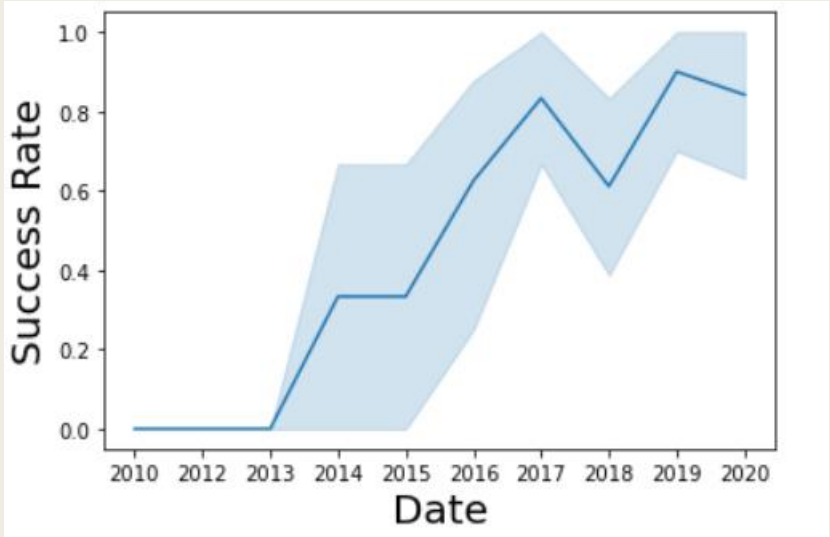


# Exploratory Data Analysis using Pandas and Matplotlib

- Relationship between success rate of each orbit type.




- launch success yearly trend.



# Exploratory Data Analysis using SQL

- The code → GitHub
- the names of the unique launch sites in the space mission.

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

- 5 records where launch sites begin with the string 'CCA'.

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 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 (parachute) |
| 2012-05-22 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 0:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

# Exploratory Data Analysis using SQL

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

| SUM(PAYLOAD_MASS__KG_) |
| --- |
| 45596 |

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

| AVG(PAYLOAD_MASS__KG_) |
| --- |
| 2928.4 |

# Exploratory Data Analysis using SQL

- the date when the first succesful landing outcome in ground pad was acheived.

| min(DATE) |
| --- |
| 2015-12-22 |

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

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

# Exploratory Data Analysis using SQL

- the total number of successful and failure mission outcomes

```
[21]: COUNT(*)
      101
```

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

```
[22]: 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
```

# Exploratory Data Analysis using SQL

- the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015

[45]:

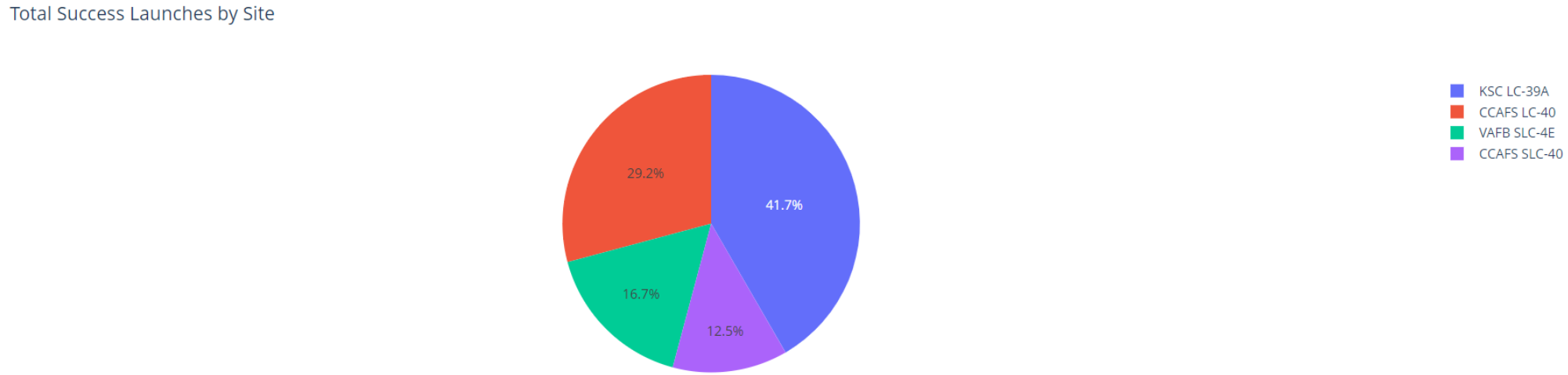
| Month | Landing_Outcome | Booster_Version |
|---|---|---|
| 01 | Failure (drone ship) | F9 v1.1 B1012 |
| 04 | Failure (drone ship) | F9 v1.1 B1015 |

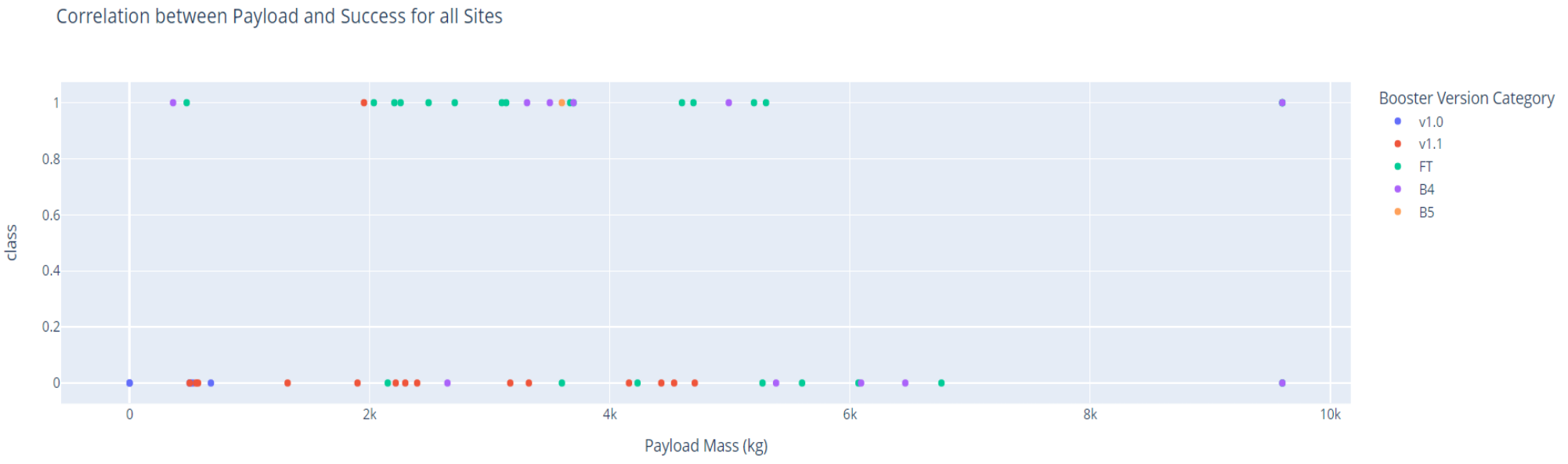
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[47]:

| Date | COUNT |
|---|---|
| 2017-02-19 | 1 |
| 2017-01-14 | 1 |
| 2016-08-14 | 1 |
| 2016-07-18 | 1 |
| 2016-05-27 | 1 |
| 2016-05-06 | 1 |
| 2016-04-08 | 1 |
| 2015-12-22 | 1 |

# Plotly Dash dashboard results

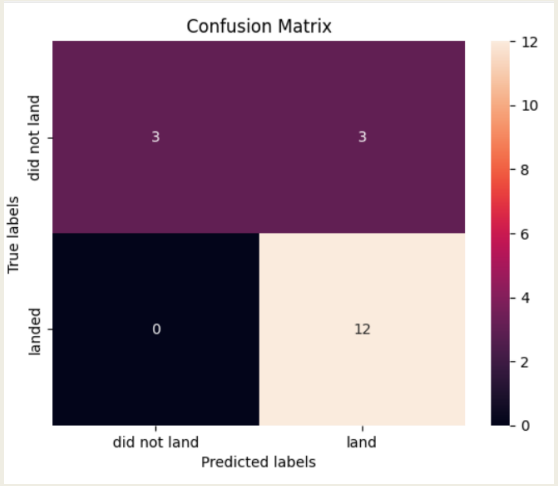


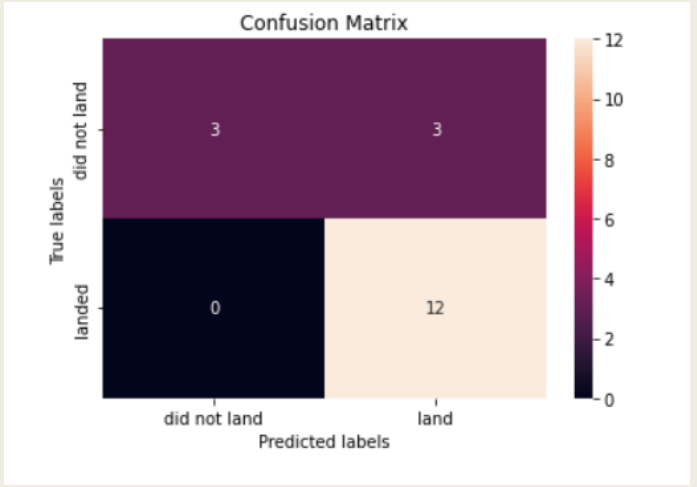



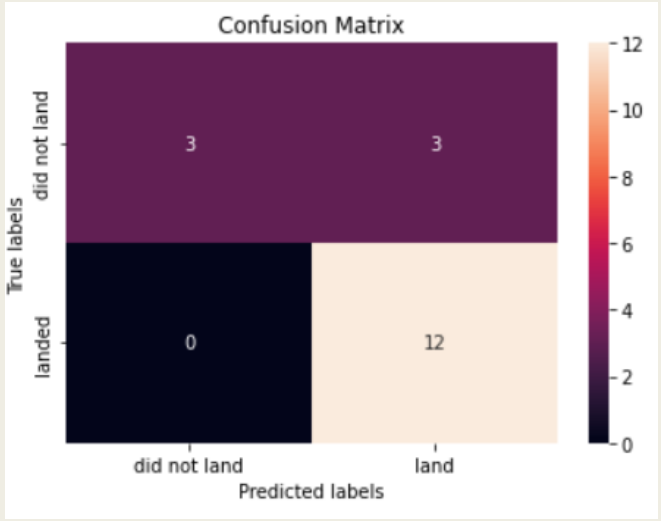
# Interactive Map with Folium

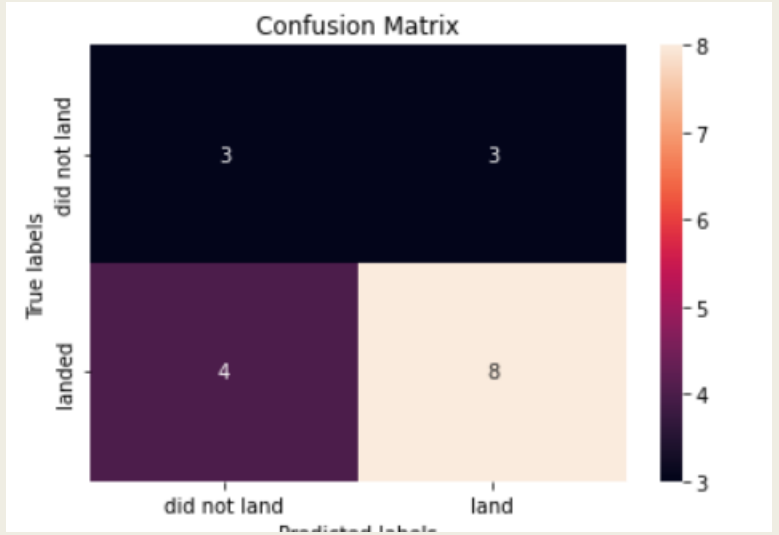
GitHub

# Machine learning Predictive Analysis



# conclusion

- KSC LC-39A the most successful launch – see slide 15 –
- In terms of accuracy prediction, it is recommended to use SVM,KNN,
- HEO, GEO,SSO AND ES-L1 →the top success rate – see slide 9 -