# SpaceX-Capstone-Project

- Michael E. Baker
- September 28, 2022

# Outline

- Executive Summary
- Introduction
- Methodology
- Conclusion

# Executive Summary

## Purpose of Report

- The space Flight industry is in a close race for dominance to deliver payloads. Cost and safe landings are the primary considerations when biding for payload delivery. The purpose of this report is to:

- Determine which factors contribute most reducing cost and guaranteeing a safe landing.

- Suggest ways to decrease cost by insuring a safe landing.

Methods

- Data Collection-SpaceX API

- Exploring Data Analysis (EDA) using visualization and SQL

- Interactive visual analysis using Folium and Ploly Dash

- Predictive Analysis using Classification Medels

Findings and Conclusions

1. Payload size plays a big part in determining a successful launch.

2. Launch location is a factor in determining a successful landing

# Introduction

- Project background and context
- The goal of this project is to create a machine learning pipeline to predict if the first stage of a rocket launch will land successfully. If the prediction is correct, other companies can more successfully bid against SpaceX's Falcon 9 rocket launches. SpaceX advertises a cost of 62 million due to its reuse of its first stage.

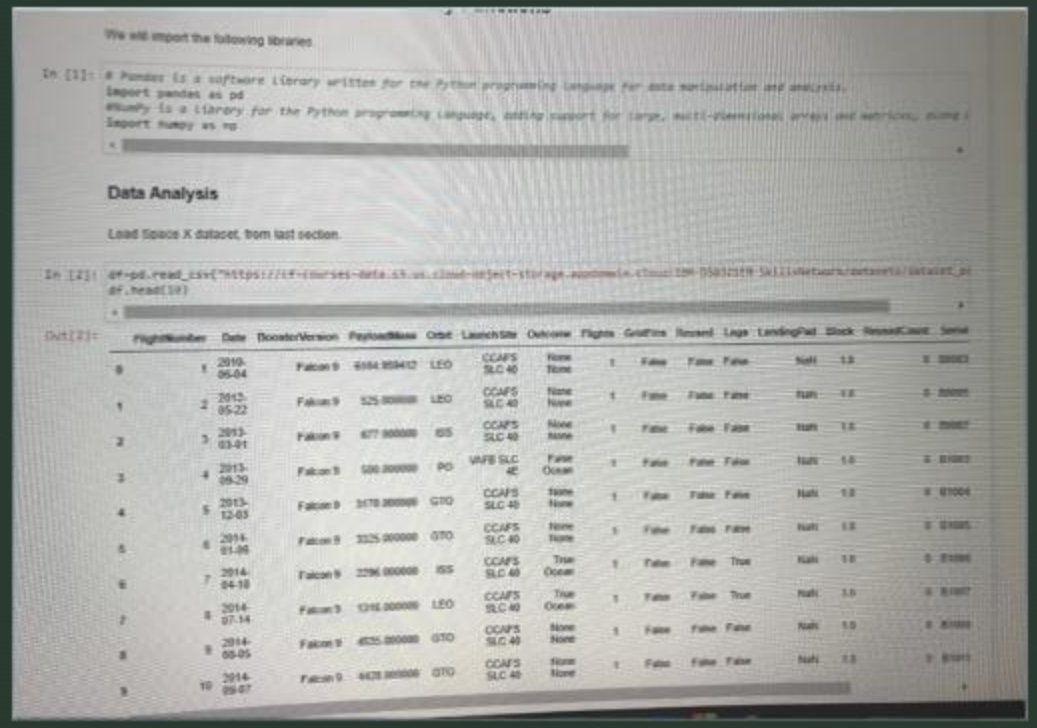
- Problems you want to find answers
- What factors need to be in-place to guarantee a successful launch and landing
- How much do payload mass and location factor into a successful launch and safe landing

Section 1

# Methodology

# Data Collection- SpaceX API and Data Wrangling methodology

- The libraries and calls used to extract data from the SpaceX API
- https://github.com/MikeDataCourse/Data-Science-Final.git



We will import the following libraries

```
In [1]: # Pandas is a software library written for the Python programming language for data manipulation and analysis.
        import pandas as pd
        #Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along
        import numpy as np
```

**Data Analysis**

Load Space X dataset, from last section.

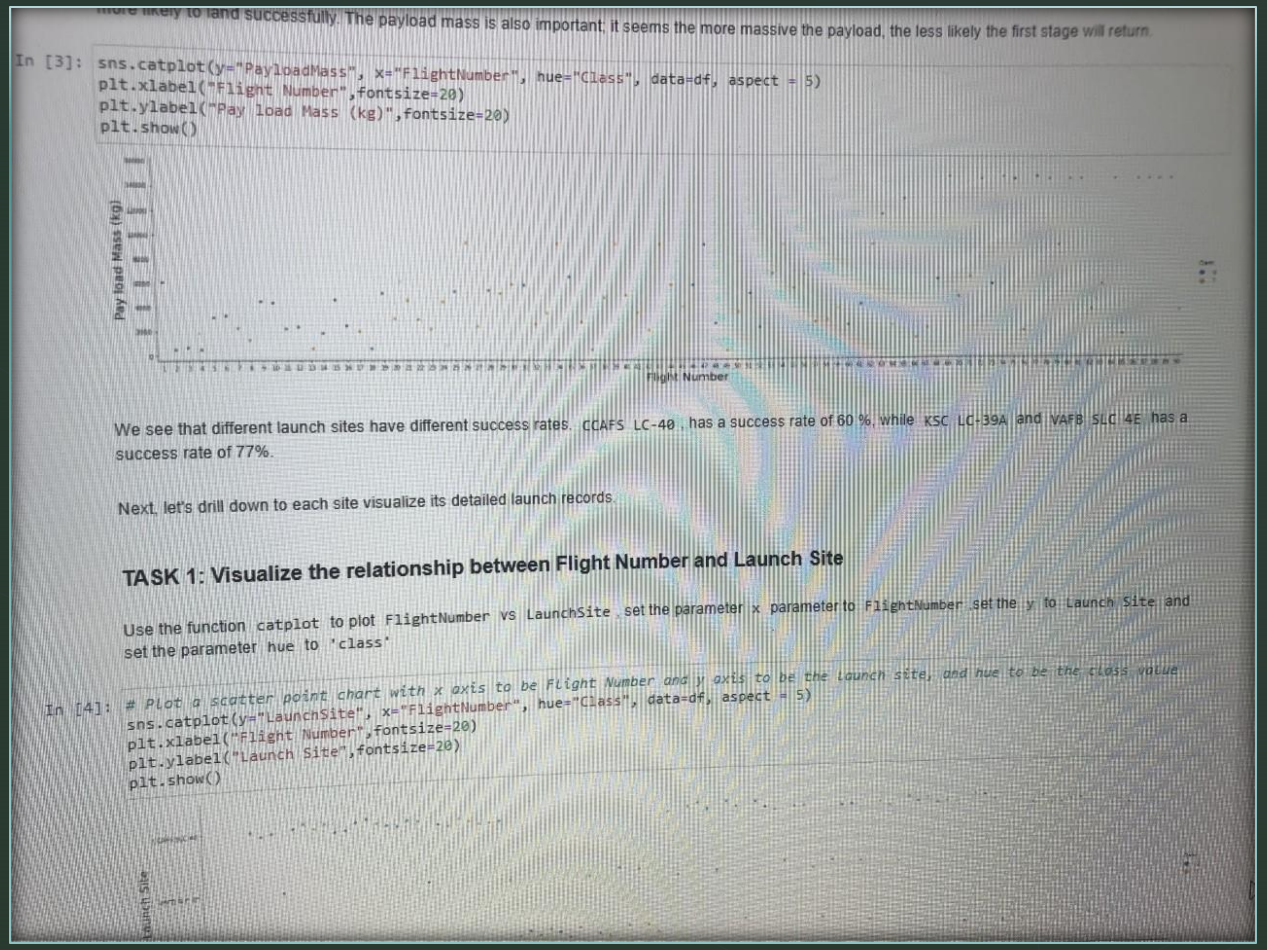
```
In [2]: df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_p
        df.head(10)
```

Out[2]:

| | FlightNumber | Date | BoosterVersion | PayloadMass | Orbit | LaunchSite | Outcome | Flights | GridFins | Reused | Legs | LandingPad | Block | ReusedCount | Serial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2010-06-04 | Falcon 9 | 6104.959412 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0003 |
| 1 | 2 | 2012-05-22 | Falcon 9 | 525.000000 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0005 |
| 2 | 3 | 2013-03-01 | Falcon 9 | 677.000000 | ISS | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0007 |
| 3 | 4 | 2013-09-29 | Falcon 9 | 500.000000 | PO | VAFB SLC 4E | False Ocean | 1 | False | False | False | NaN | 1.0 | 0 | B1003 |
| 4 | 5 | 2013-12-03 | Falcon 9 | 3170.000000 | GTO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B1004 |
| 5 | 6 | 2014-01-06 | Falcon 9 | 3325.000000 | GTO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B1005 |
| 6 | 7 | 2014-04-18 | Falcon 9 | 2296.000000 | ISS | CCAFS SLC 40 | True Ocean | 1 | False | False | True | NaN | 1.0 | 0 | B1006 |
| 7 | 8 | 2014-07-14 | Falcon 9 | 1316.000000 | LEO | CCAFS SLC 40 | True Ocean | 1 | False | False | True | NaN | 1.0 | 0 | B1007 |
| 8 | 9 | 2014-08-05 | Falcon 9 | 4535.000000 | GTO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B1008 |
| 9 | 10 | 2014-09-07 | Falcon 9 | 4428.000000 | GTO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B1011 |

# EDA and interactive data Visualization methodology

- Charts Plotted

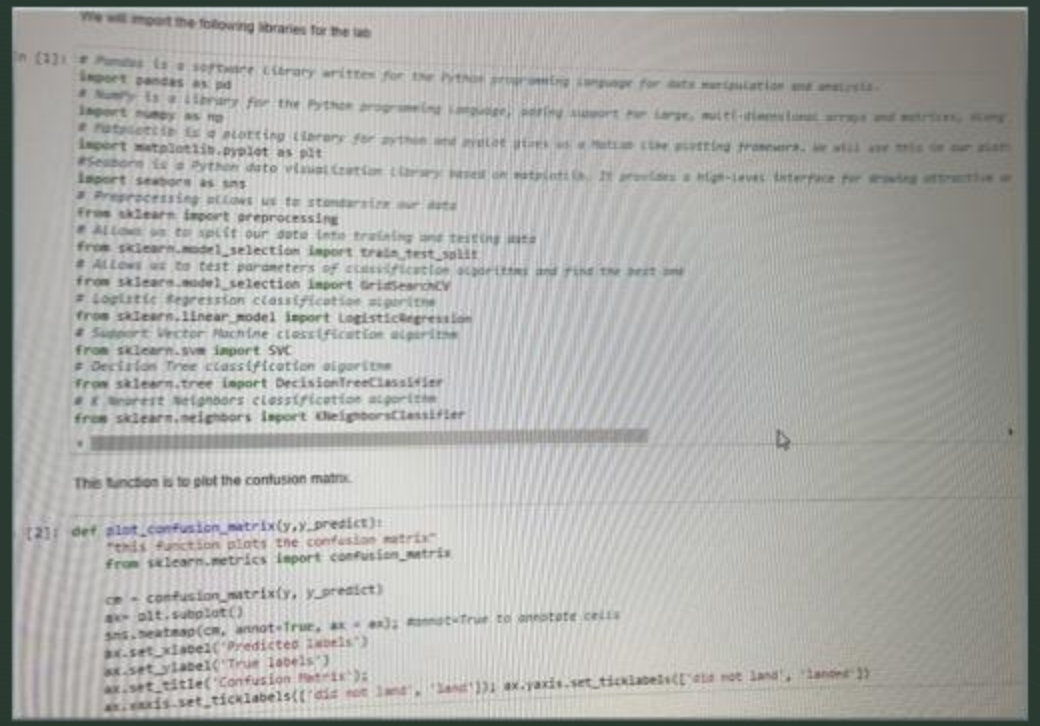



- Launch number & Pay load Mass
- Flight number & Launch Site. These factors are two of the most critical in determining the success of a launch
- https://github.com/MikeDataCourse/Data-Science-Final.git

# Predictive Analysis

- The predictive was bult with the imported libraries. The model was improved by testing and training the data until the optimum accuracy was achieved.

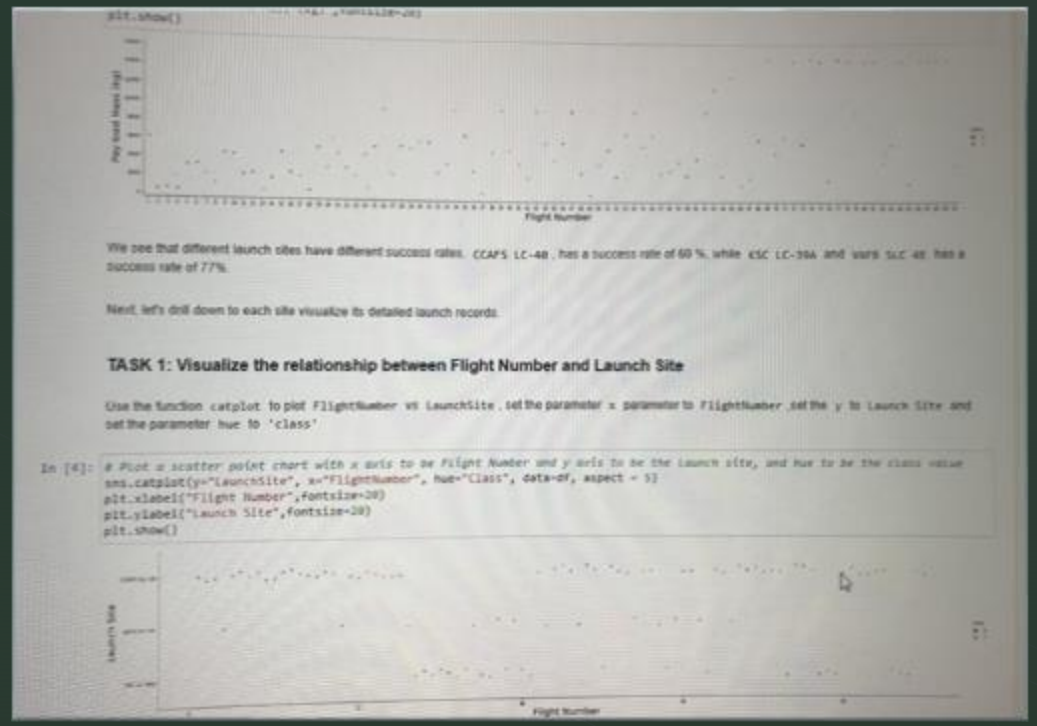
- https://github.com/MikeDataCourse/Data-Science-Final.git



# EDA with Visualization

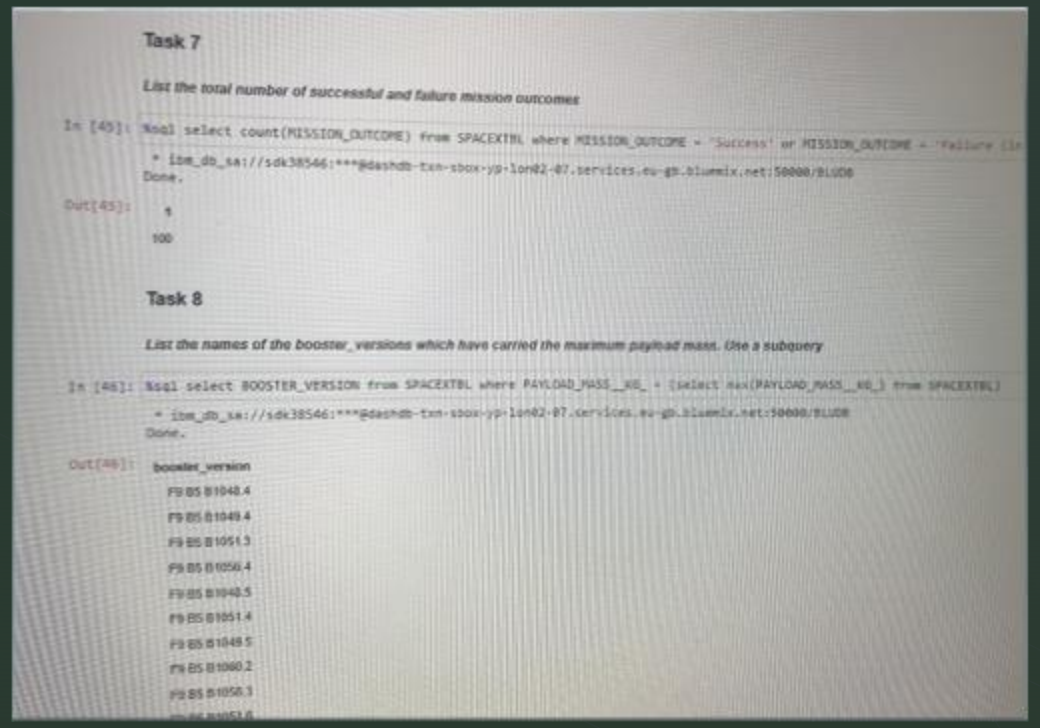
- The scatter plots allow visualization of the data collected and provide interactive manipulation

- https://github.com/MikeDataCourse/Data-Science-Final.git



# EDA with SQL

- Queries used in this EDA
- In order to extract the data results sought, the following Queries were employed
  - Distinct
  - Where
  - Like
  - Sum
  - Avg
  - Count



**Task 7**

*List the total number of successful and failure mission outcomes*

```
In [45]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
Out[45]:
1
100
```

**Task 8**

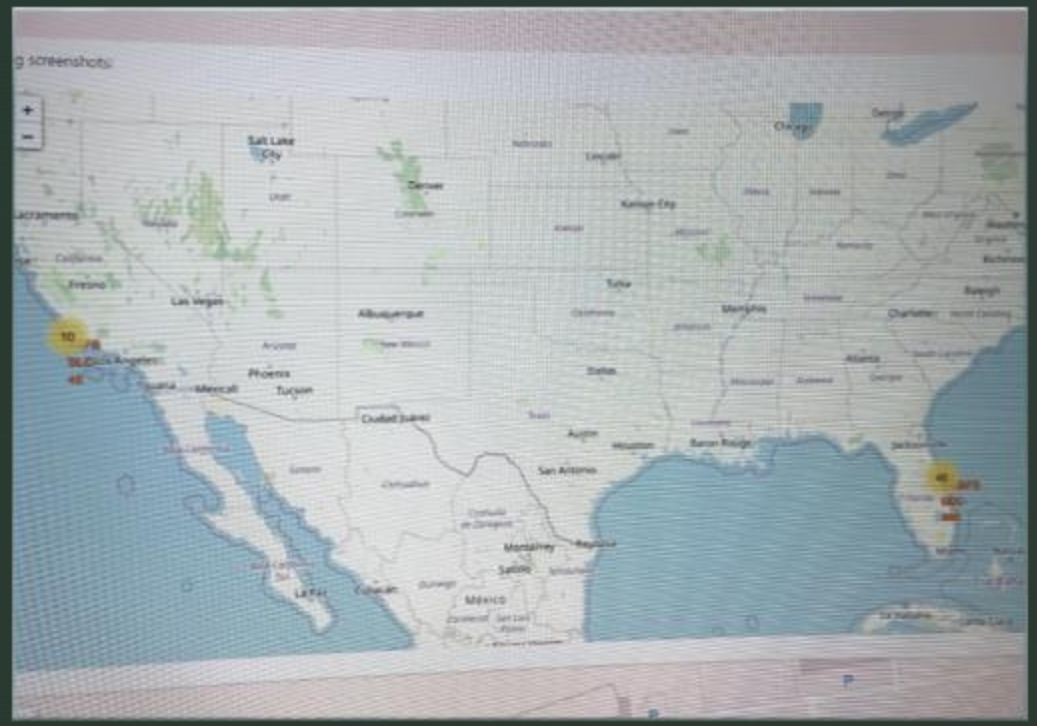
*List the names of the booster_versions which have carried the maximum payload mass. Use a subquery*

```
In [46]: %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
Out[46]:
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
```

# Interactive Map with Folium

- The folium map was created with circles, lines and markers in order to more clearly show the relationship, proximity and number of competing companies to each launch site.

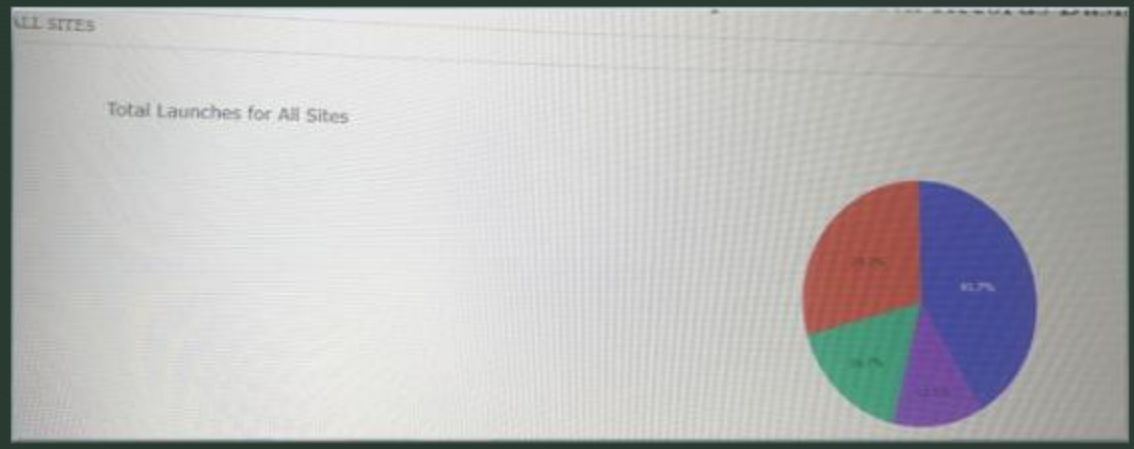
- https://github.com/MikeDataCourse/Data-Science-Final.git



This is the interactive dashboard of the launch records.

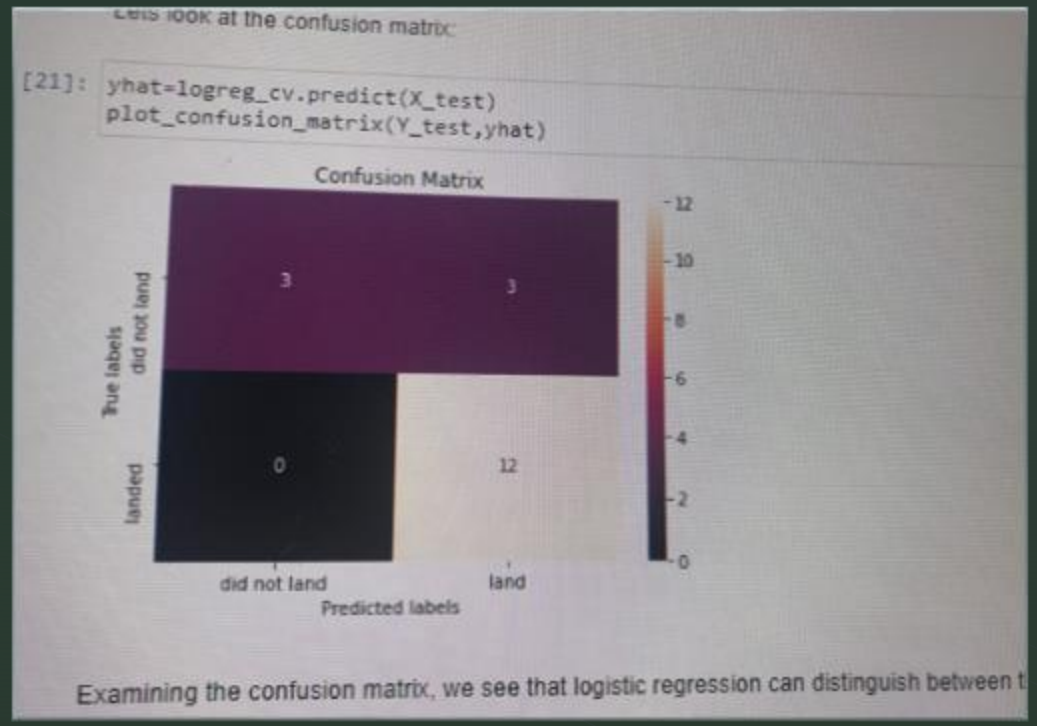
# Dashboard with Plotly Dash

https://github.com/MikeDataCourse/Data-Science-Final.git



# Predictive analysis (Classification)

- The confusion Matrix is a predictor of the success of a launch using regression
- https://github.com/MikeDataCourse/Data-Science-Final.git



# Conclusion

- As reflected in the results of the data, extracted, from the SpaceX API, there are several factors that can guide to a more successful first stage landing. As reflected, with the flight number increase, the first stage is more likely to land successfully, but when attempting to carry a massive payload will decrease the first stages successful landing. The data shows that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60% while KSC LC-39A and VAFB 4E have a success rate of 77%. Launch site CCAFS SLC 40 is more suitable for launching rockets with payload mass varying from low to very high and launch site VAFB SLC 4E is preferred for medium payload mass. The data reflects that Orbit is also a factor with heavy payloads having negative influence on the GTO orbits and positive on GTO and Polar LEO(ISS) orbits.
- The combined factors of payload mass , orbit type and launch location , if properly matched, can predict a successful first stage landing.

# Creativity to Improve the Presentation beyond the template

Perhaps, I do not fully understand how the various style sides function, but, I would like to have been able to add more than one screenshot to a slide. I think it would have allowed for more in depth explanation of the methods used to extract the data.

# Innovative insights

- The line chart plus a Pie or Bar chart in the same frame would allow me more expression of my presentation.