Winning Space Race with Data Science

Dr. Alaa Temimy

1st. Oct. 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Data about SpaceX - Falcon 9 missions are gathered , analyzed , plotted , and ML models are built to predict the rocket-stage#1 re-use ability.

- The results show that, the successful landing of the rocket-stage#1 are depends on many variables such as : Launch site , Orbit , Payload mass , Reused Count ..etc.

Introduction

- Space X 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 Space X 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 space X for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

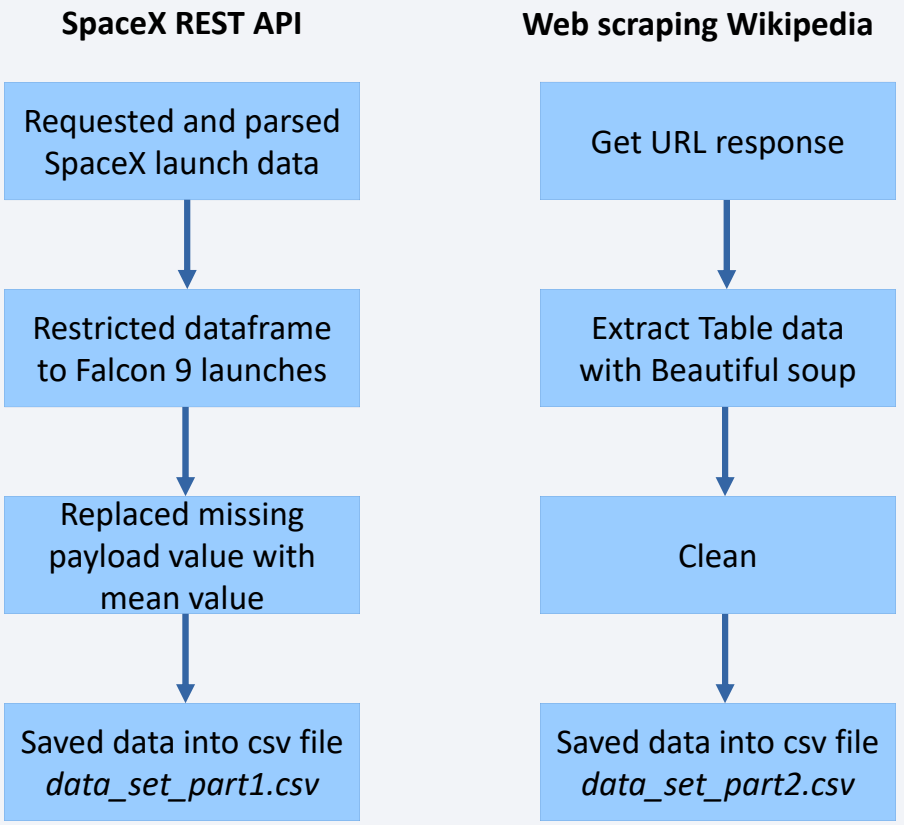
- Data collection methodology:
  - Data are collected from www. spacexdata.com and from Wikipedia en.wikipedia.org using API procedures.
- Perform data wrangling
  - Data are wrangled by using some of Python packages such as Pandas and Numpy.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Optimization procedure with Grid-Search model for different models hyper-parameters are used to tune the ML models to get best accuracy.

Data Collection

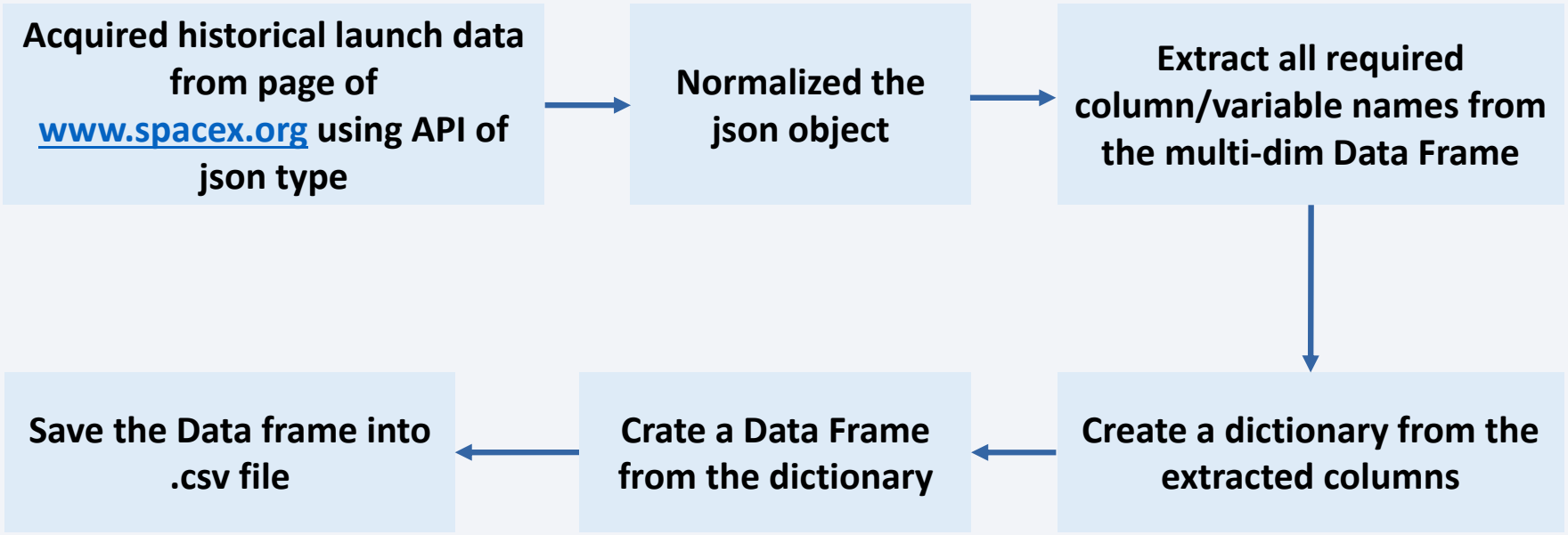
Data was collected from:

- Open source SpaceX REST API

- Web-scraping Falcon9 launch data in Wikipedia

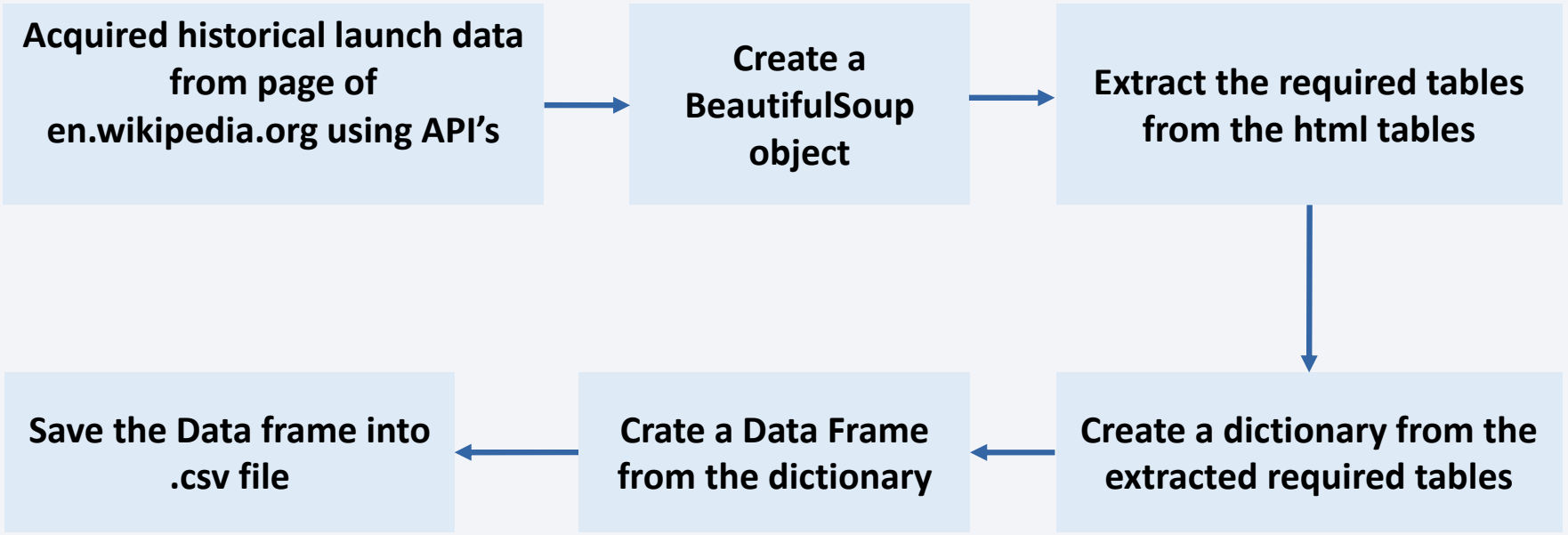


Data Collection – SpaceX API



https://github.com/Temimy/Data_Science_Cap_Stone

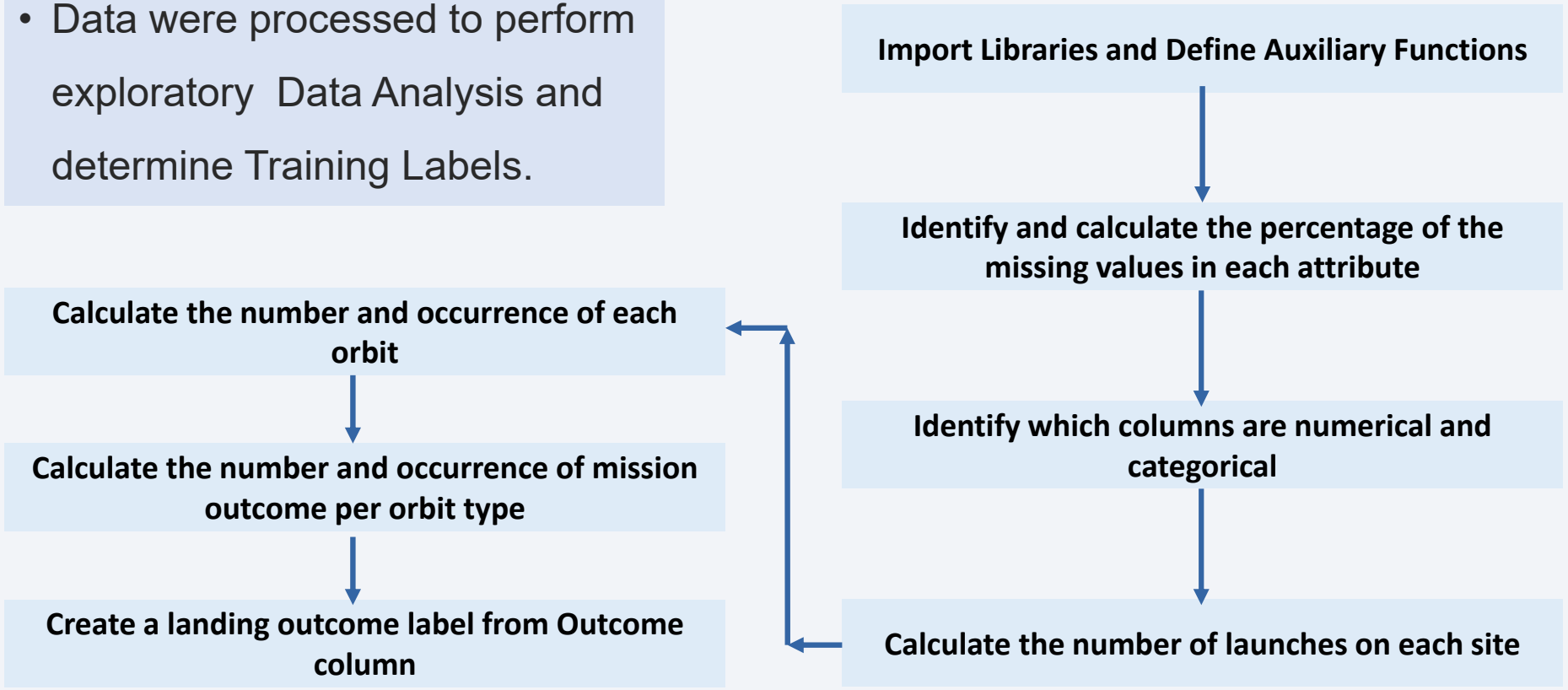
Data Collection - Scraping



https://github.com/Temimy/Data_Science_Cap_Stone

Data Wrangling

- Data were processed to perform exploratory Data Analysis and determine Training Labels.



https://github.com/Temimy/Data_Science_Cap_Stone

EDA with Data Visualization

- To get insights from the data , the following charts are plotted :

Chart Title	Notes
Relationship between Flight Number and Pay load Mass	**To observe the effect of Flight Number and Pay load Mass on successful landing**
Relationship between Flight Number and Launch Site	**To observe the effect of Flight Number and Launch Site on successful landing**
Relationship between Payload and Launch Site	**To observe the effect of Payload and Launch Site on successful landing**
Relationship between success rate of each orbit type	**To observe the effect of the Orbits on successful landing**
Relationship between FlightNumber and Orbit type	**To observe the effect of Flight Number and Orbit type on successful landing**
Relationship between Payload and Orbit type	**To observe the effect of Payload and Orbit type on successful landing**
The launch success yearly trend	**To observe yearly progression rates of the project**

https://github.com/Temimy/Data_Science_Cap_Stone

EDA with SQL

The performed SQL queries are :

- **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 successful landing outcome in ground pad was achieved.**

(to be continue)

https://github.com/Temimy/Data_Science_Cap_Stone

EDA with SQL

The performed SQL queries are :

- **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 failed landing_outcomes in drone ship, their booster versions, and launch site names for 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**

https://github.com/Temimy/Data_Science_Cap_Stone

Build a Dashboard with Plotly Dash

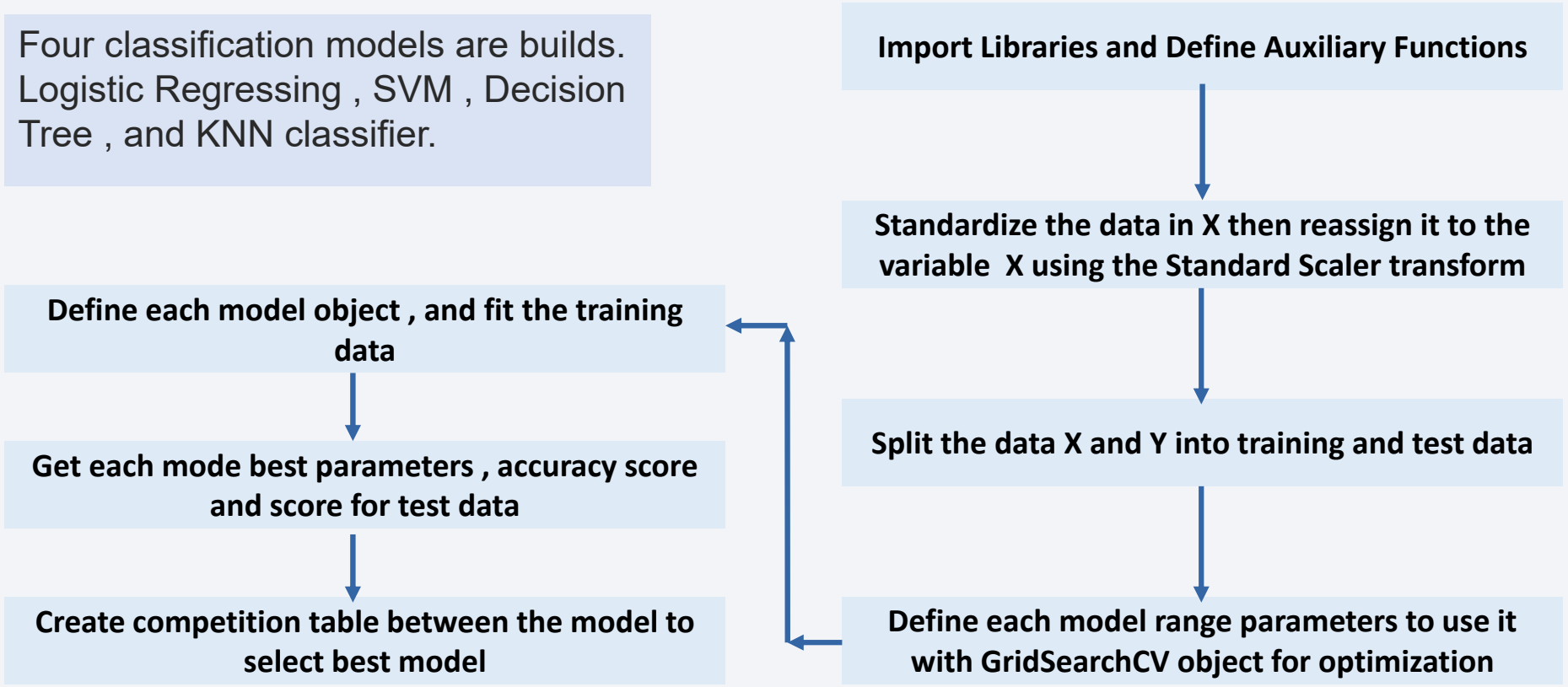
The dashboard plots/graphs and interactions are :

- **Dropdown list to enable Launch Site selection.**
- **Pie chart to show the total successful launches count for all sites.**
- **Slider to select payload range.**
- **Scatter chart to show the correlation between payload and launch success.**

https://github.com/Temimy/Data_Science_Cap_Stone

Predictive Analysis (Classification)

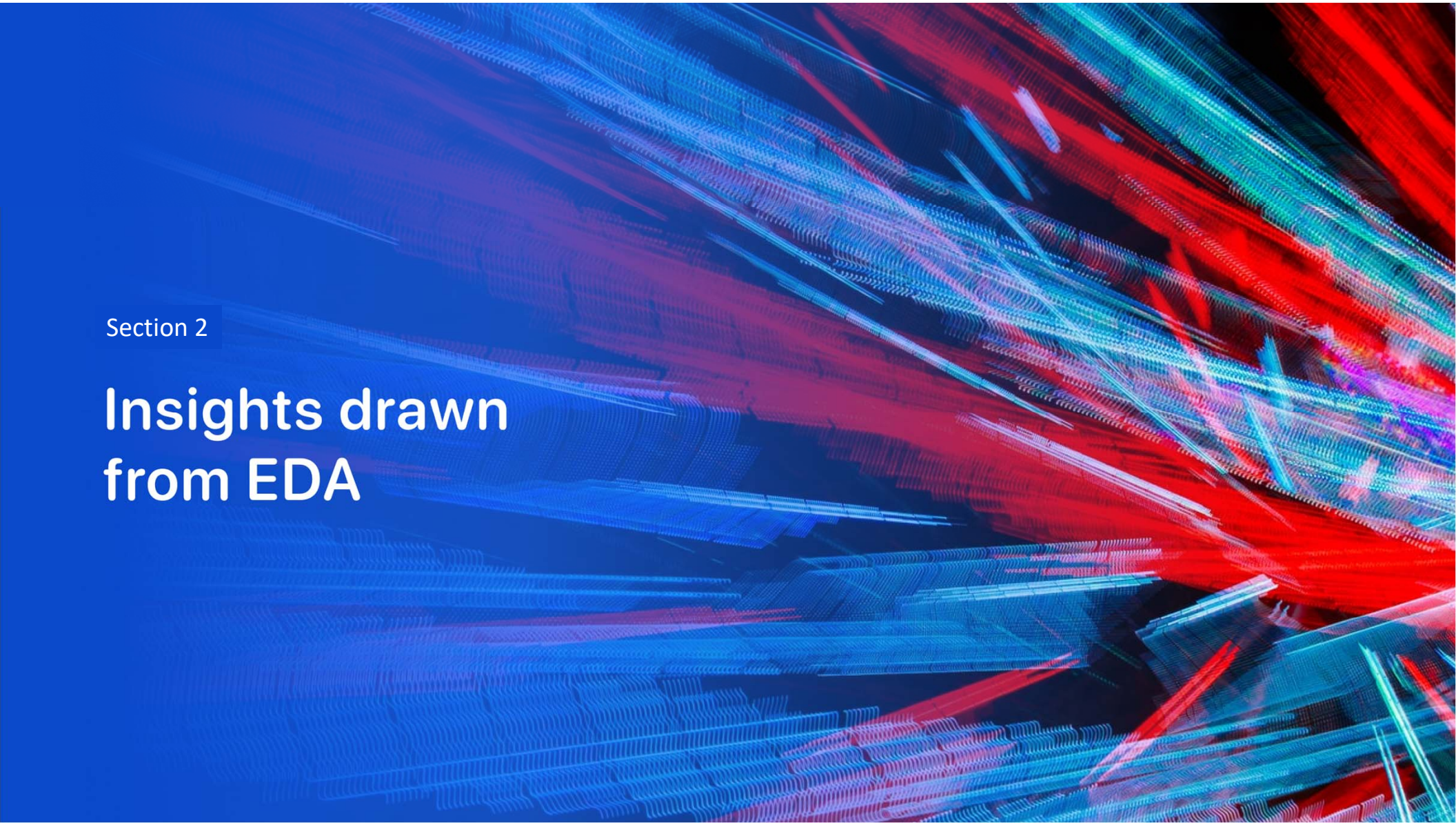
Four classification models are builds. Logistic Regressing , SVM , Decision Tree , and KNN classifier.



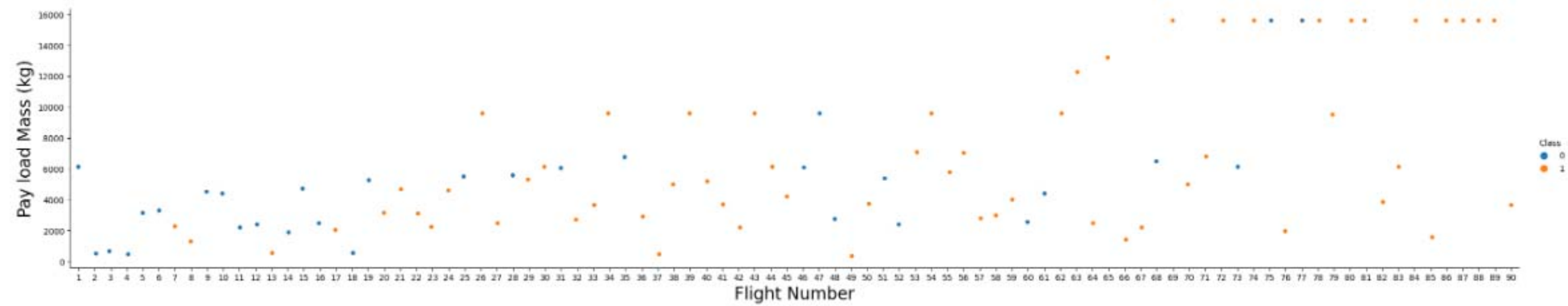
https://github.com/Temimy/Data_Science_Cap_Stone

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn
from EDA

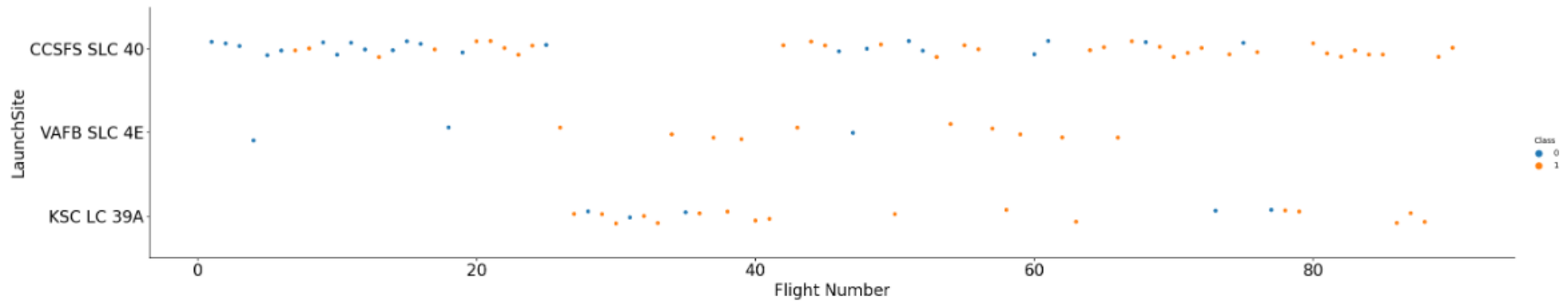
Flight Number vs. PayLoadMass



As the flight number increases, the first stage is more likely to land successfully

https://github.com/Temimy/Data_Science_Cap_Stone

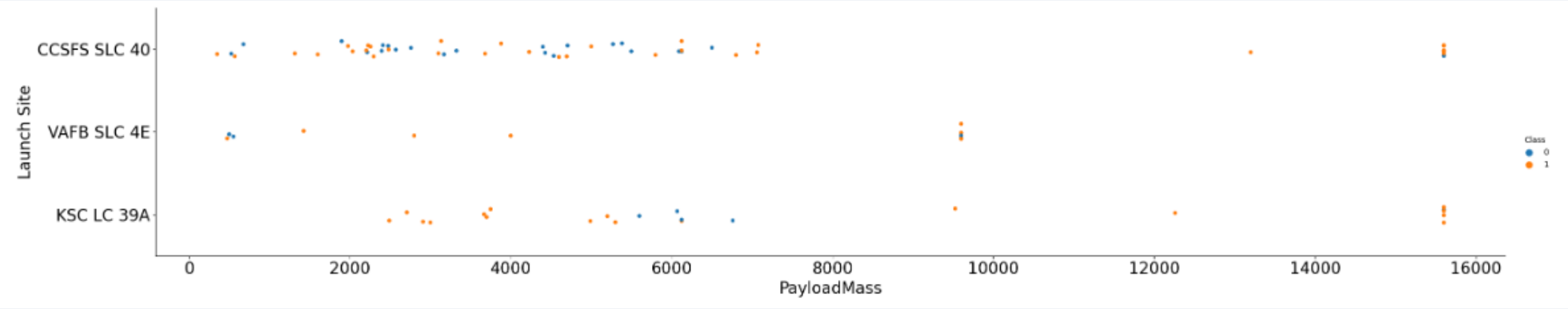
Flight Number vs. Launch Site



As the flight number increases, the first stage is more likely to land successfully

https://github.com/Temimy/Data_Science_Cap_Stone

Payload vs. Launch Site

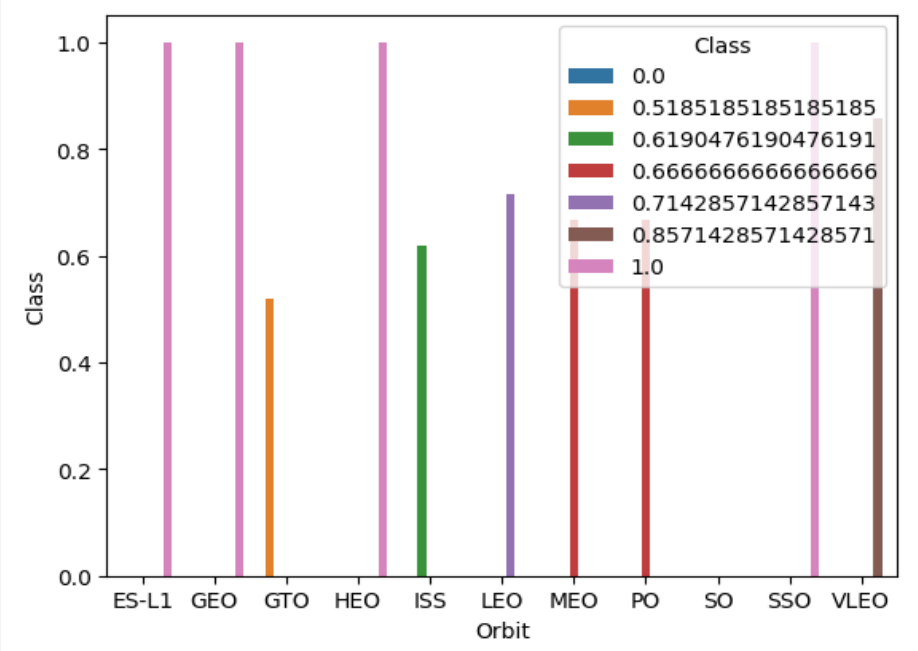


For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)

https://github.com/Temimy/Data_Science_Cap_Stone

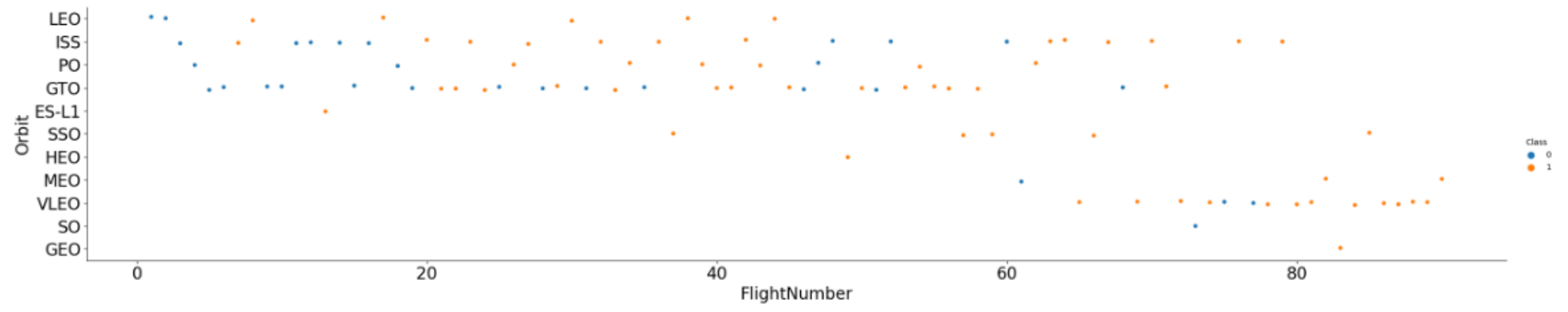
Success Rate vs. Orbit Type

The success landing rates are high for the orbits :
ES-L1 , GEO,HEO,SSO



https://github.com/Temimy/Data_Science_Cap_Stone

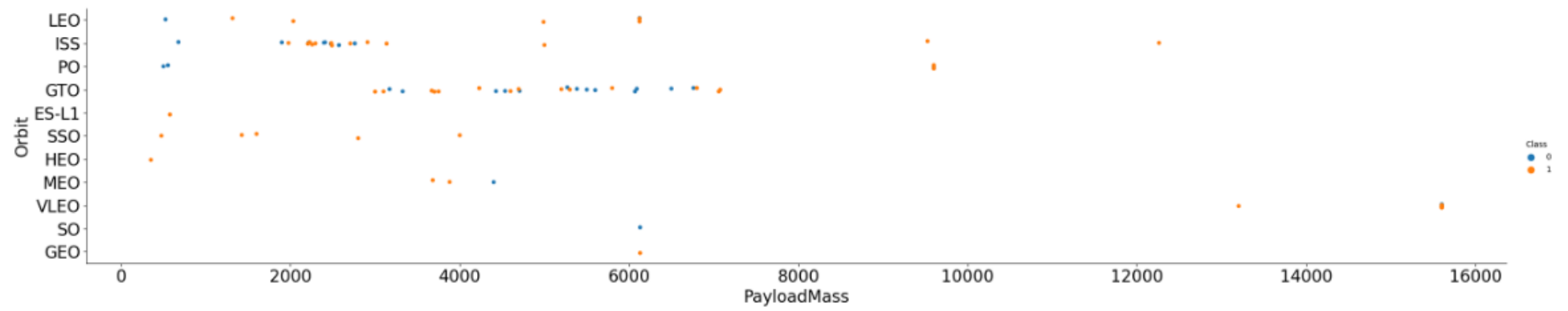
Flight Number vs. Orbit Type



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

https://github.com/Temimy/Data_Science_Cap_Stone

Payload vs. Orbit Type



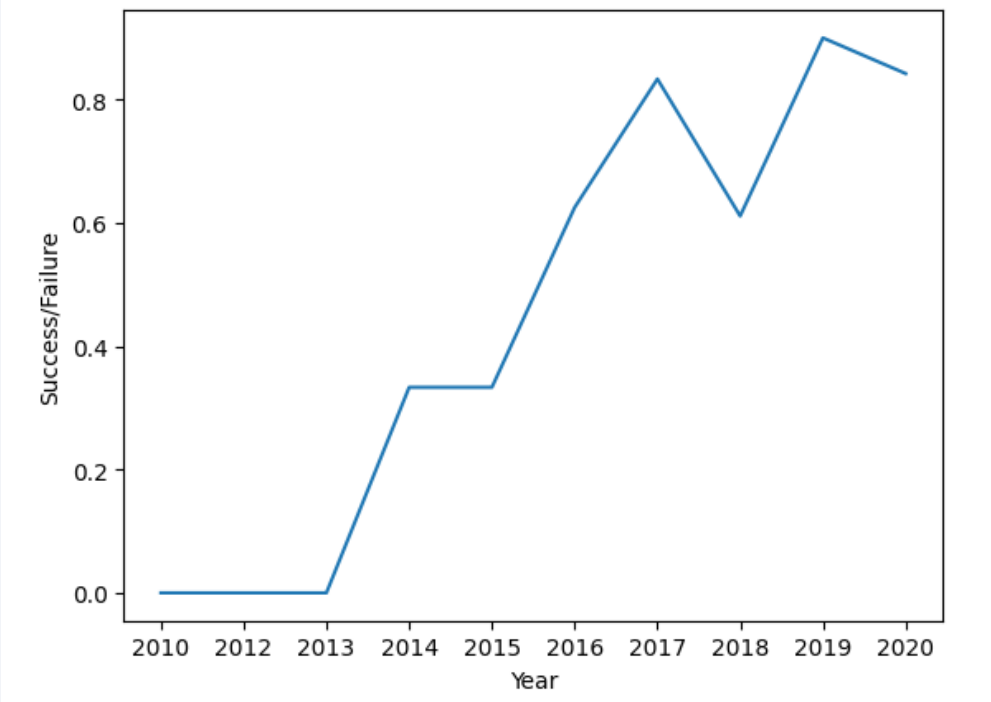
With heavy payloads the successful landing or positive landing rate are more for Polar , LEO and ISS.

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

https://github.com/Temimy/Data_Science_Cap_Stone

Launch Success Yearly Trend

We can observe that the success rate since 2013 kept increasing till 2020



https://github.com/Temimy/Data_Science_Cap_Stone

All Launch Site Names

- 4 launch sites are used for spacex missions

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA` (sample of query's EDA)

[12]:

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload carried by boosters from NASA about 46 ton

sum(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 , it’s about 3 ton.

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad recorded in 1/5/2017

Min(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- A list of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are :

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are :

successful missions

	COUNT(Mission_Outcome)
0	100

failed missions

	COUNT(Mission_Outcome)
0	1

Boosters Carried Maximum Payload

- A list of the names of the booster which have carried the maximum payload mass are :

	Booster_Version	PAYLOAD_MASS__KG_
0	F9 B5 B1048.4	15600
1	F9 B5 B1049.4	15600
2	F9 B5 B1051.3	15600
3	F9 B5 B1056.4	15600
4	F9 B5 B1048.5	15600
5	F9 B5 B1051.4	15600
6	F9 B5 B1049.5	15600
7	F9 B5 B1060.2	15600
8	F9 B5 B1058.3	15600
9	F9 B5 B1051.6	15600
10	F9 B5 B1060.3	15600
11	F9 B5 B1049.7	15600

2015 Launch Records

- A list of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are :

	Booster_Version	Launch_Site	Landing_Outcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Section 3

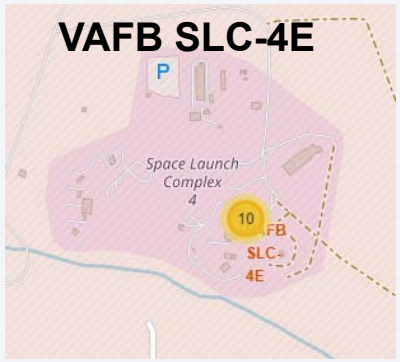
Launch Sites Proximities Analysis

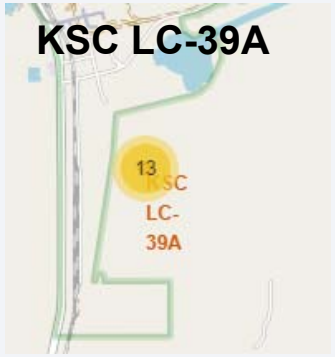
SpaceX Launch Records Dashboard

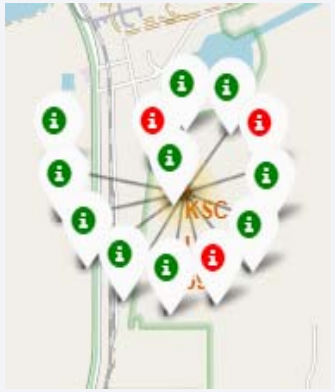
- All launch sites are shown in the map

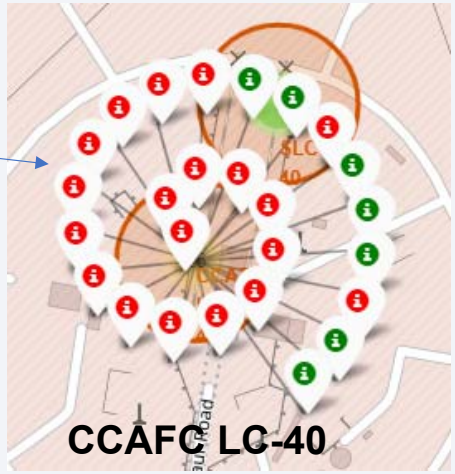
- SpaceX uses a total of 4 facilities;
  - California:
    - VAFB SLC-4E
  - Florida:
    - KSC LC-39A
    - CCAFS LC-40
    - CCAFS SLC-40

Color Labeled Launches Outcomes







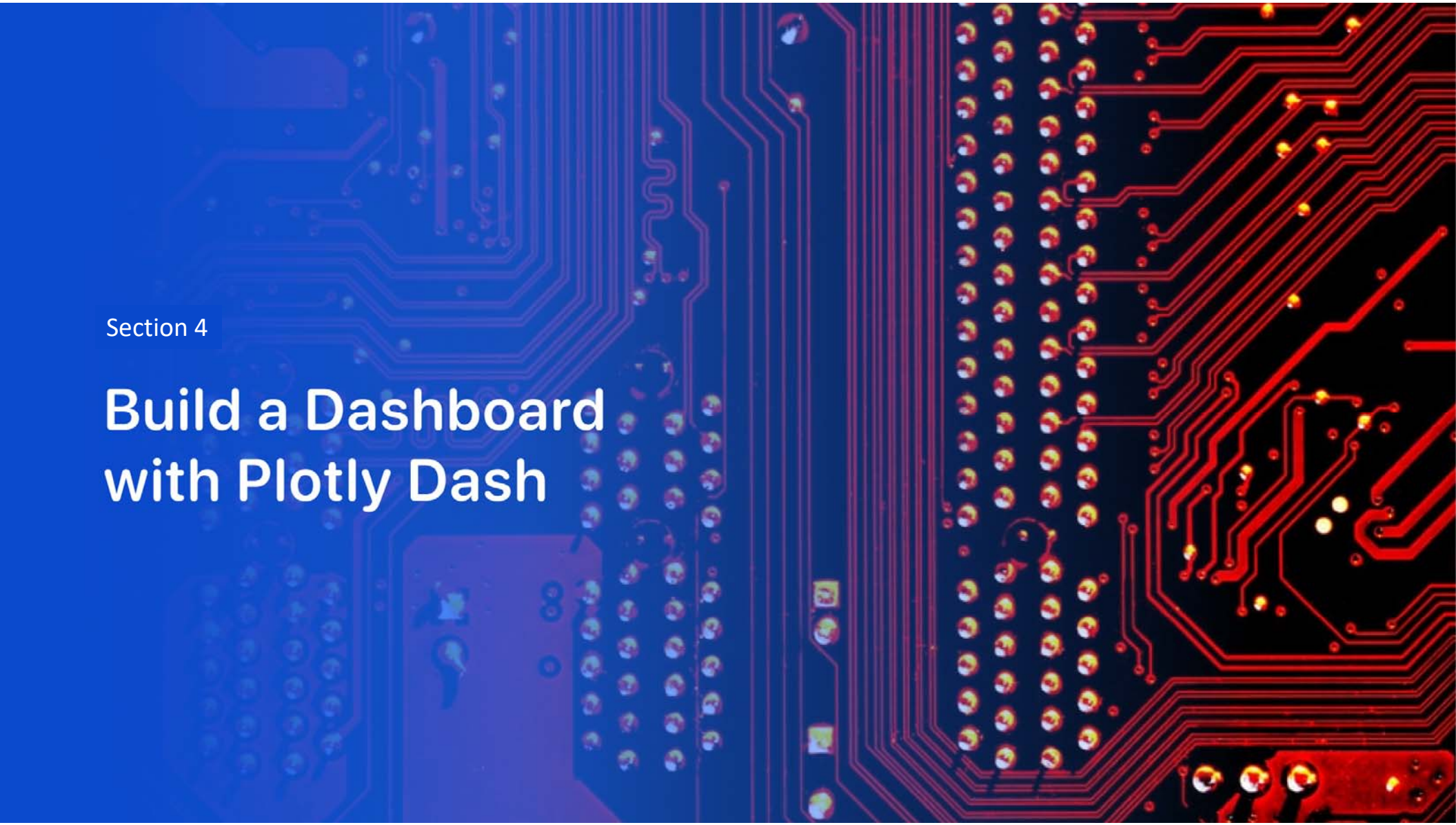


GREEN MARKERS
Successful landings

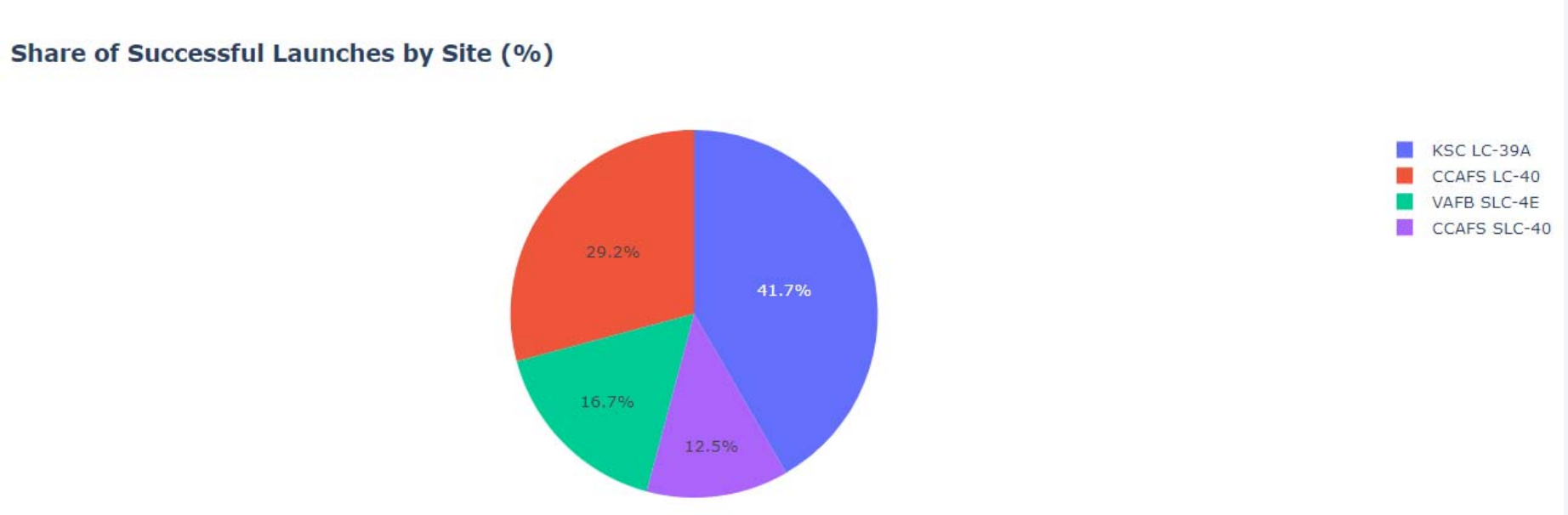
RED MARKERS
Unsuccessful landings

Selected Launch Site to its Proximities

- Distance CCAFS_SLC40 to Florida East Coast Railway: ABOUT 22 km
- Distance CCAFS_SLC40 to Highway I95: 26.8 km
- Distance CCAFS_SLC40 to Orlando: 78.8 km


Section 4
Build a Dashboard
with Plotly Dash

Launch Success Count for all Sites

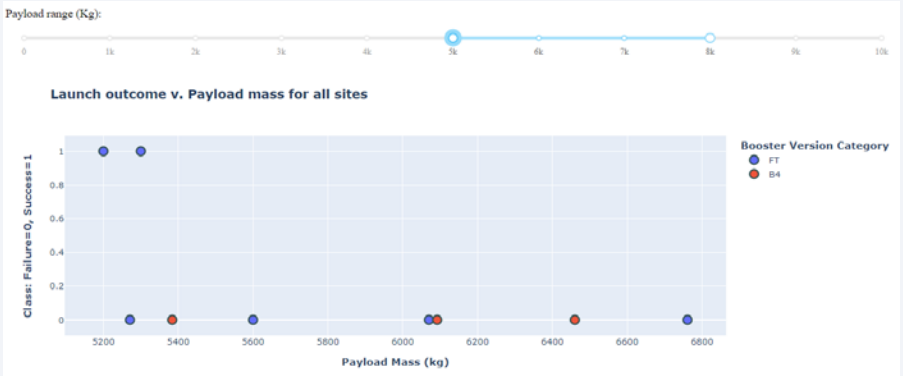


It’s clear that KSC LC-39A is the launch site with the highest success rate of 41.7%

Launch Success rates for KSC LC-39A

Launch attempts outcome (%) for site KSC LC-39A
Total number of attempts= 13
Number of success= 10
Number of failures= 3

< Payload vs. Launch Outcome scatter plot for all sites



- Most missions are in the range of PayLoadMass of (2-5) k.
- For the PayLoadMass range (5-8)K most booster fails to land successfully.

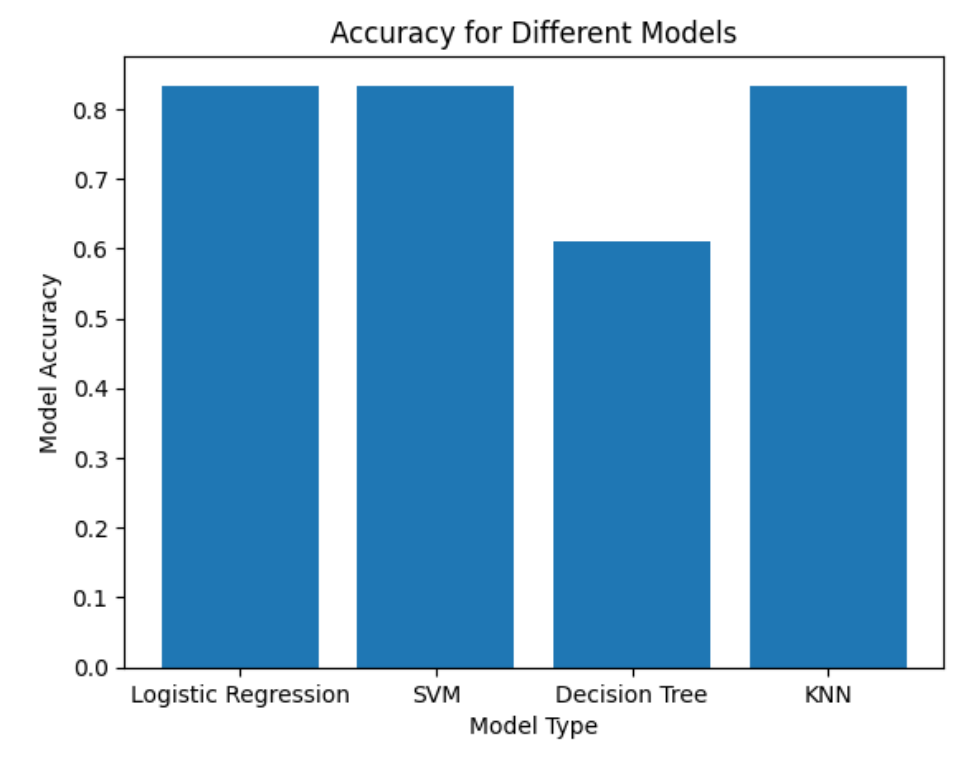
Section 5

Predictive Analysis (Classification)

Classification Accuracy

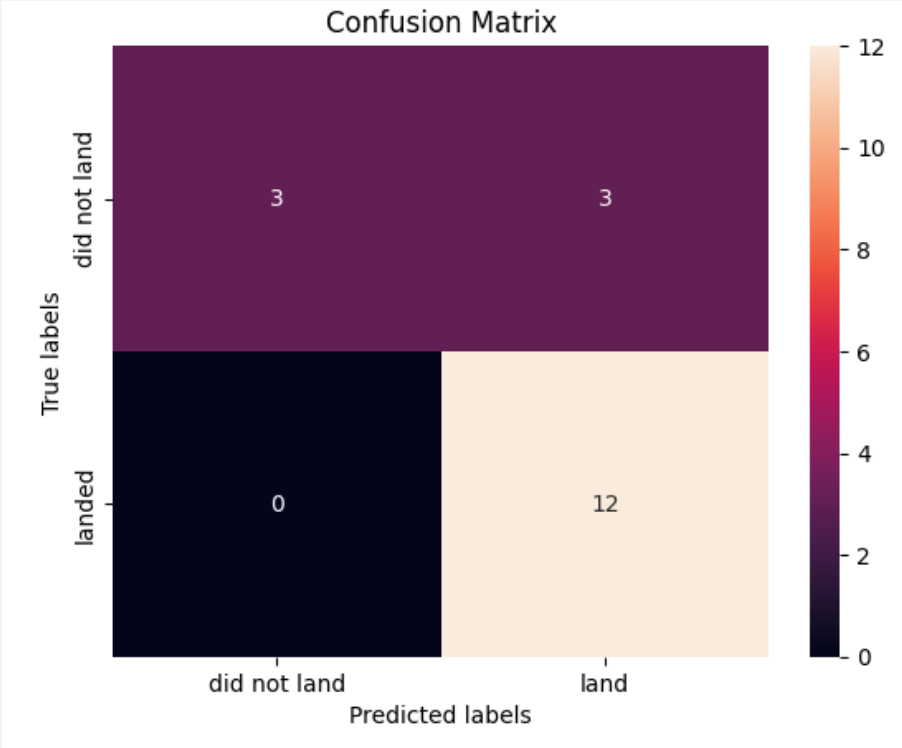
The used Classification Models are :

- Logistic Regressions
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbor model



Confusion Matrix

- The Confusion Matrix for the Decision Tree Classifier is shown.
- Since the classifier give a score of 0 for the prediction of (FN) , i.e. the False prediction for the non successful landing is equal to zero. So that , the classifier will predict with good apportinuty the bad landing , and this will help to prevent of wasting money on a non successful landing , i.e. prevent accident from occurs.



Conclusions

1. Orbits GEO, ES-L1 , GEO,HEO,SSO have the best success rate.
2. SpaceX has come a long way, as their landing success rates have steadily increased over the years and show a rising trend going forward.
3. While CCAFS SLC-40 site is used for most launches ,KSC LC-39A in Florida, is the site with the most successful landing outcomes overall.
4. Most missions are in the range of PayLoadMass of (2-5) k.
5. The Decision Tree algorithm is the best pick for our predictive model using this dataset.

Thank you!