IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yap Kim Seong
30 March 2025

Outline

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

Executive Summary

Methodologies:

- Data Collection using SpaceX API and Web Scrapping using BeautifulSoup
- Exploratory Data Analysis methods such as data wrangling, visualization and interactive dashboards
- Various predictive machine learning algorithms

Summary of all results

- Data Collection from various public sources is viable
- EDA and feature engineering are essential to successful predictions
- Choosing the best machine learning model is essential to predict outcomes accurately

Introduction

- The main goal is to help Space Y to compete with Space X

- How to achieve the main goal?
    - What are the best and strategic location to launch rockets
    - Estimate the cost for all launches
    - Predicting sucessful landings during the first stage of rocket launches

Section 1

Methodology

Methodology

Executive Summary

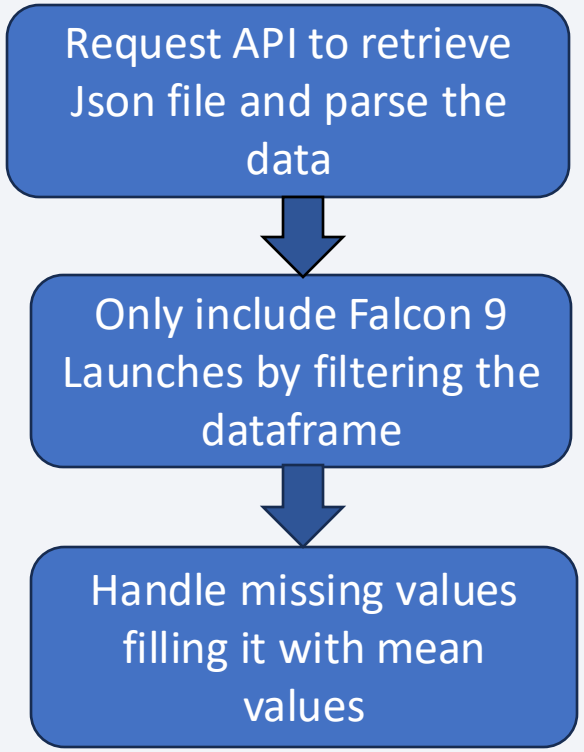
- Data collection methodology:
  - Data was collected via SpaceX API and WebScraping of Falcon Launches
- Perform data wrangling
  - The outcome data was modified to a 0 and 1 classification label for better understanding of each variable's contribution to the prediction output
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was standardized and normalized
  - Data was split into training and test sets
  - Four classification models were trained and evaluated using different parameters

Data Collection – SpaceX API

- Github link for SpaceX API

Data Collection:

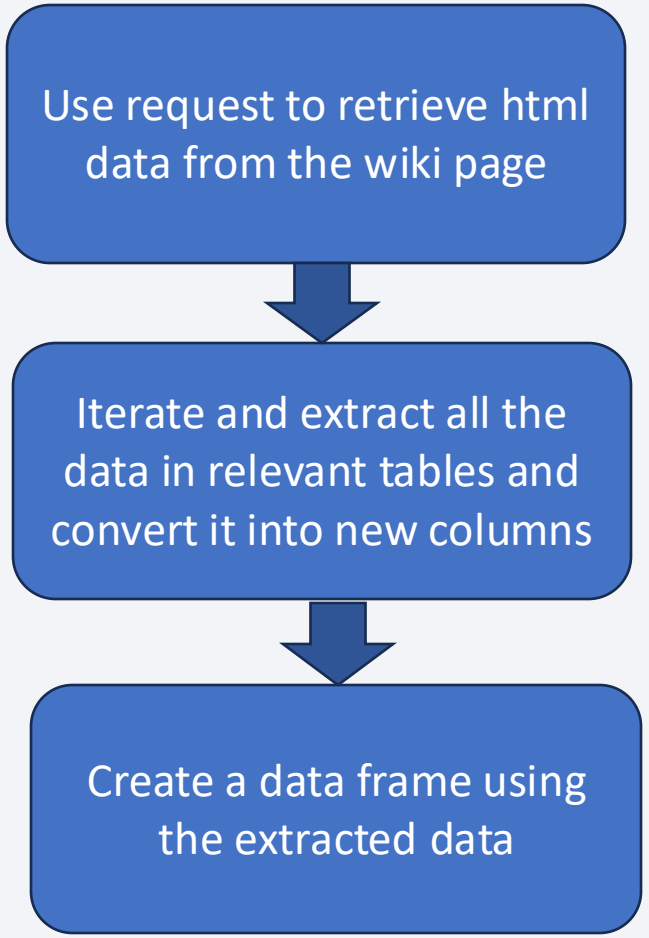
https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20Data%20Collection%20API.ipynb



Data Collection - Scraping

- Github Link for WebScraping:

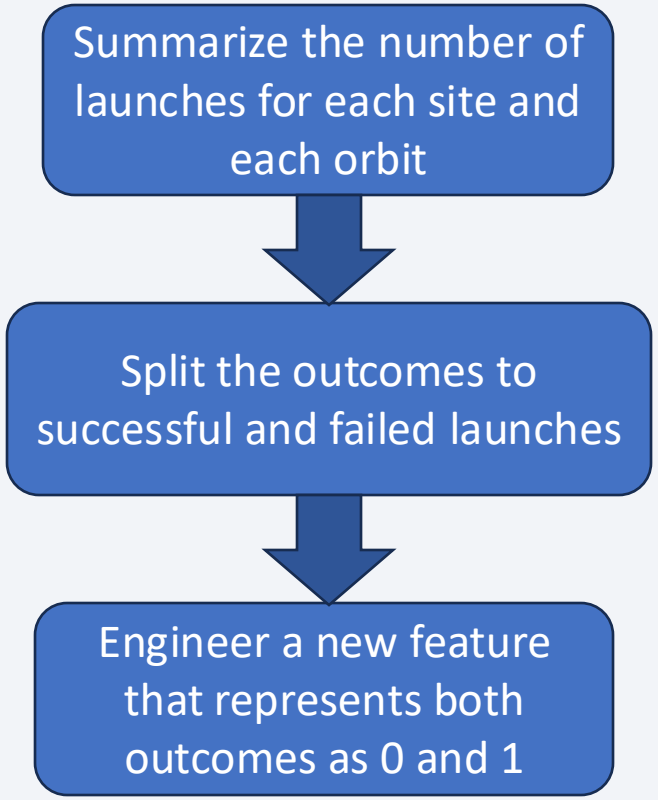
https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20Webscraping.ipynb



Data Wrangling

Source code:

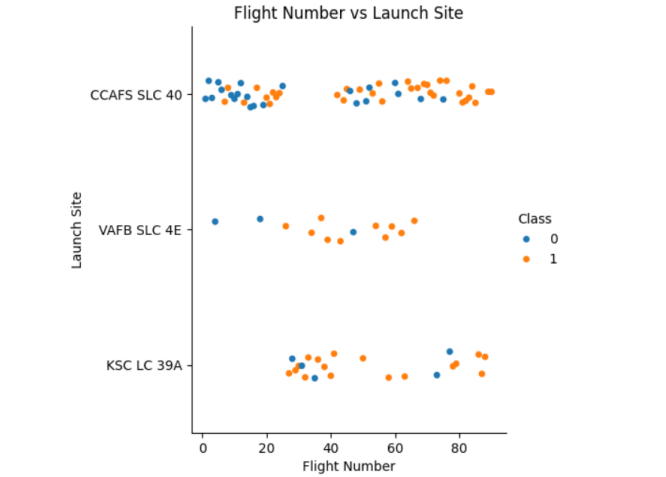
https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20Data%20Wrangling.ipynb

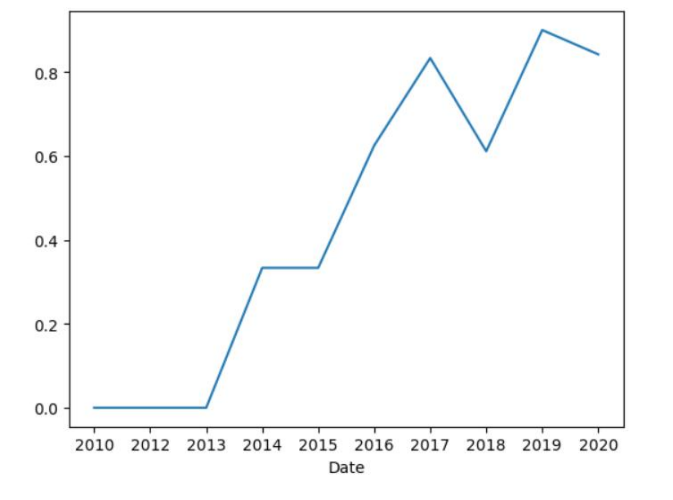


EDA with Data Visualization

- Scatter Plots were used to understand the relationship between the target variable and the independent variable
- Line Plots were used to analyze the trend of successful launches as the year increases
- Source code:

https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20EDA%20Visualization.ipynb





EDA with SQL

- Query the unique names of launch sites
- Query the total payload mass carried by booster launched by NASA (CRS)
- Query the first successful landing outcome in ground pad
- Query the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Query the total number of successful and failed mission outcomes
- Query the months on failed landing in drone in the year 2015
- Rank the count of landing outcomes between a given time in descending order
- Source code:

https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20EDA%20SQL.ipynb

Build an Interactive Map with Folium

- Markers are used to represent launch sites
- Circles represent highlighted areas around the markers
- Marker clusters represent the group of events that occurred in various locations, such as launches that happens in each launch sites
- Lines represents the distances between two locations

- Source code:

https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Pie chart is used to visualize the success percentage for each launch sites
- The scatterplot is used to visualize the relationship between different ranges of payloads with launch sites
- Both graphs/plots helps to identify which launch site is the best place to launch according to the payload range.

- Source code:

https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

- Models used:

Logistic Regression

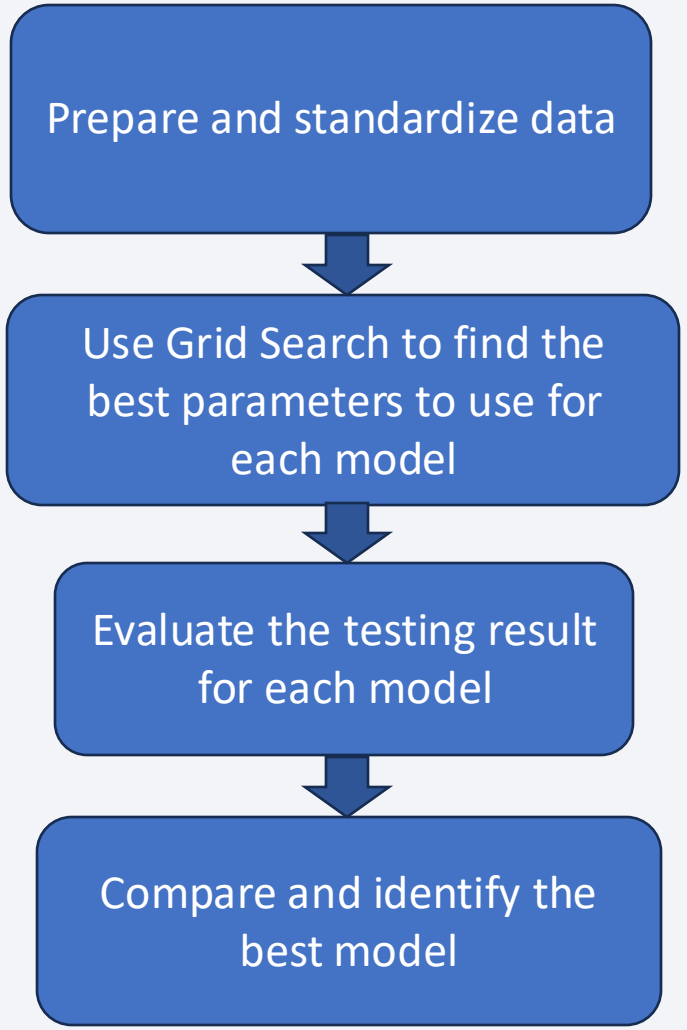
Support Vector Machine (SVM)

Decision Tree

K-Nearest Neighbour (KNN)

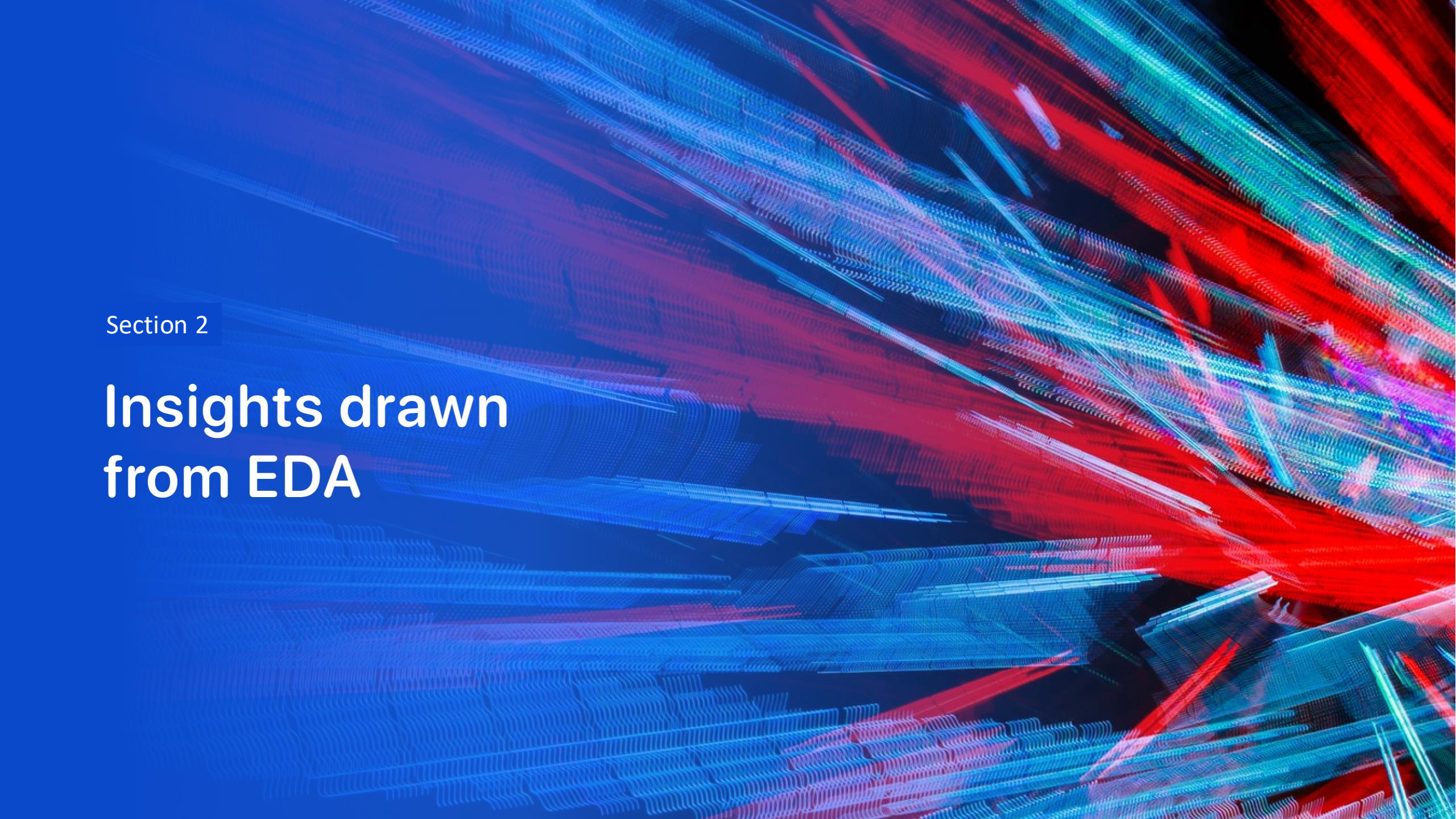
Source code:

https://github.com/KimSeong22/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

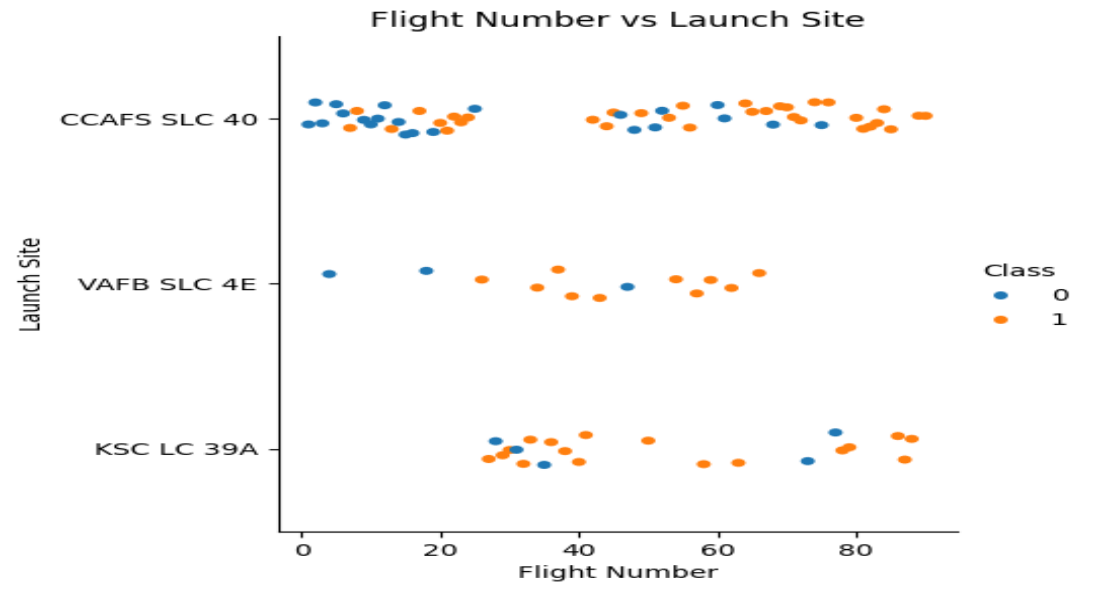


Results

- There are 4 unique launch sites
- The first successful launch was on 2015-12-22
- Falcon 9 booster versions have successful landings in drone ships when the payload is above the mean
- The number of successful landing increases as the year increases
- KSC LC-39A has the highest successful launches ratio with a 76.9% success rate
- Most launch sites were near the coastal areas
- Most launch sites are far away from cities to ensure safety
- All the machine learning models have the same accuracy of 83.33%

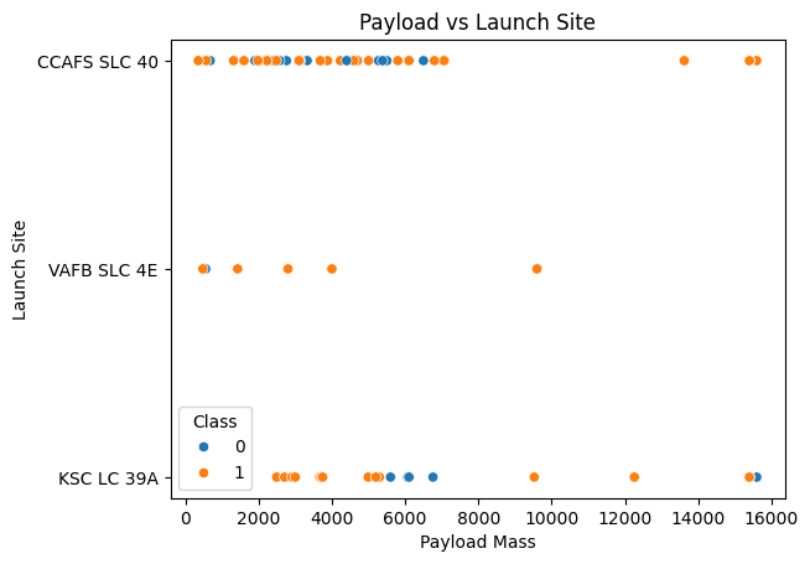

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



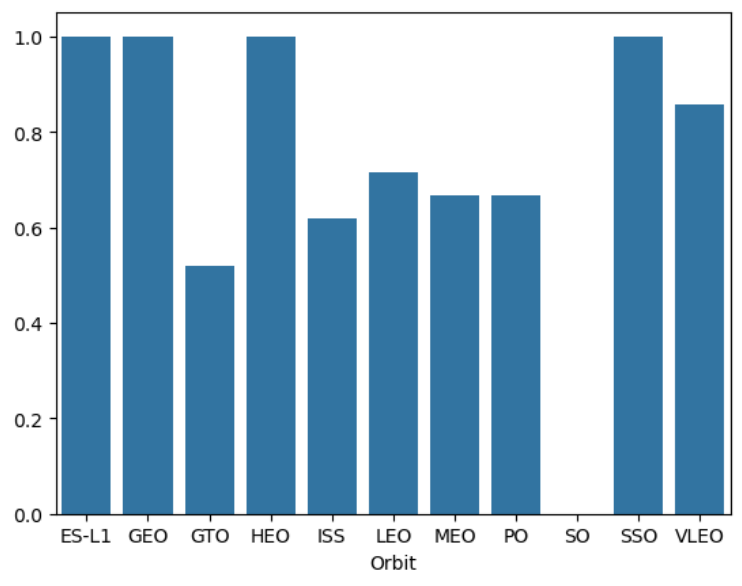
- The success rate increases as the flight number increases for CCAFS SLC 40.
- The correlation between the flight number and success rate for KSC LC 39A is weak, as the success rate is consistent as flight number increases
- The best launch site among the three sites is CCAFS SLC 40 as the success rate is high in the most recent times.

Payload vs. Launch Site



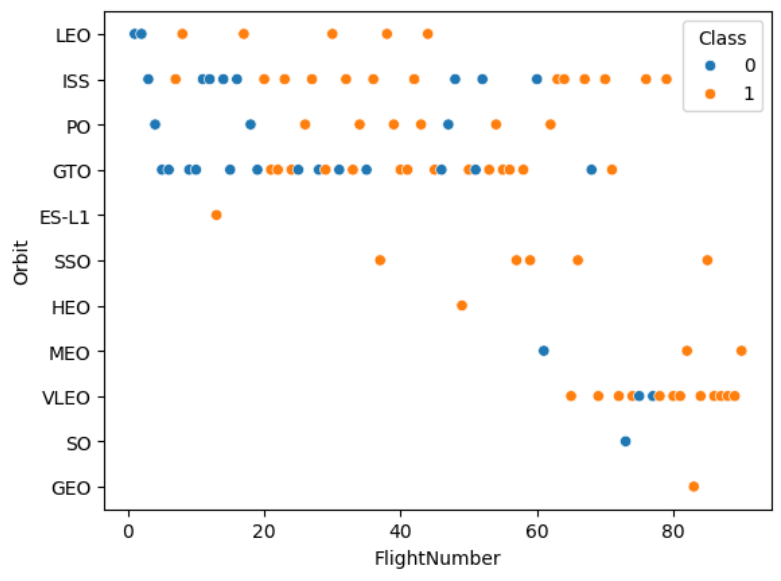
- There are no payload above the weight 10000kg for VAFB SLC 4E
- CCAFS SLC 40 is have the most successful launches above the payload weight of 10000kg
- KSC LC 39A is have more successful launches with payload mass under 6000kg and above 7000kg

Success Rate vs. Orbit Type



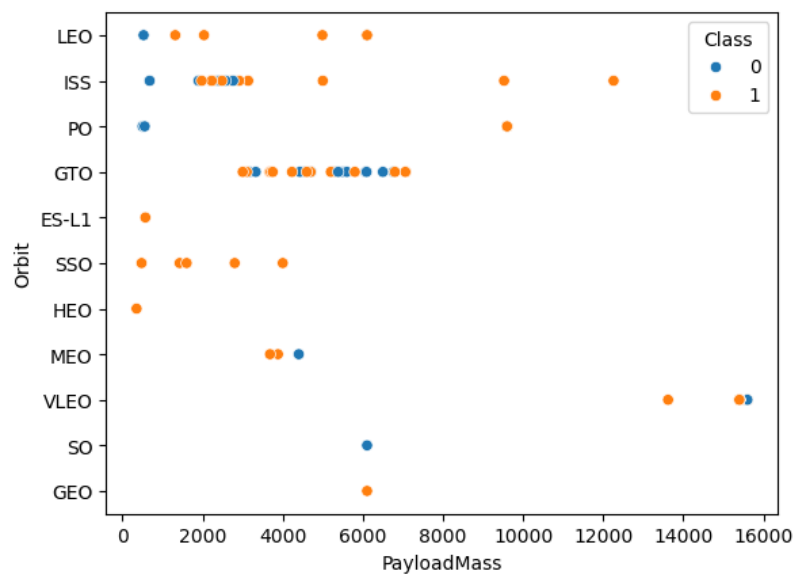
- ES-L1, GEO, HEO, and SSO orbits have a 100% launch success rate
- VLEO has a relatively high success rate of around 90%
- SO orbit has the lowest success rate of 0%

Flight Number vs. Orbit Type



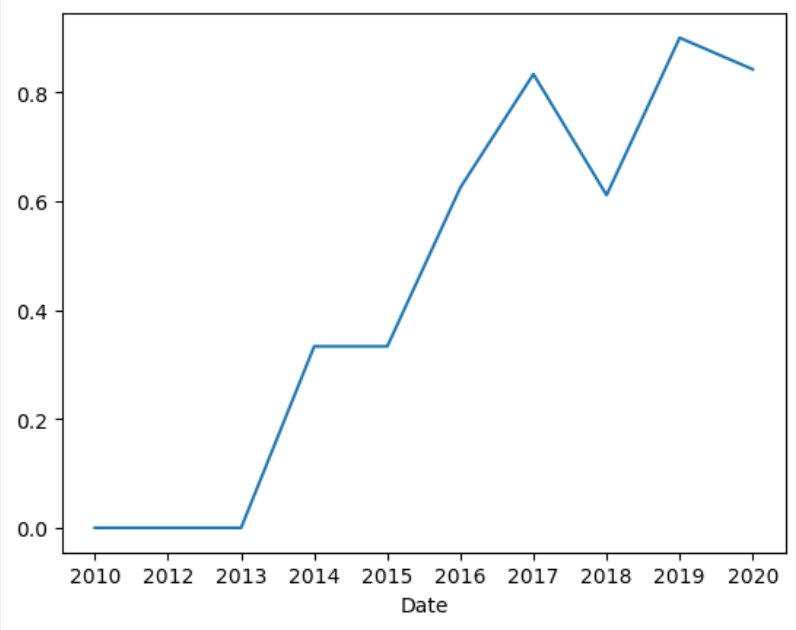
- LEO orbit has a high correlation between flight number and success rate
- There is no correlation between flight number and success rate for GEO orbit

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for PO, LEO and ISS.
- There is no correlation for successful landing and payload mass for GTO

Launch Success Yearly Trend



- The launch success rate remains stagnant between the year 2010 and 2013 as it could be a period of preparation for launches
- The launch success rate increases as the year increases until the year 2017
- There is a huge dip in success rate during the year 2018

All Launch Site Names

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

- There are 4 launch site names. The data was queried using a distinct function

Launch Site Names Begin with '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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 attemp

- These are the 5 examples of data with launch sites names that begins with CCA

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
45596

- The total payload mass from the NASA (CRS) is 45596kg

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

- The average payload mass by F9 v1.1 is 2928.4kg

First Successful Ground Landing Date

MIN(Date)
2015-12-22

- The first successful ground landing date is 22 December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- These are the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

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

- The data is grouped together to find the total count of each mission outcome

Boosters Carried Maximum Payload

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

- These are the booster versions that carries the maximum payload mass

2015 Launch Records

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- There are two booster versions with their respective launch sites that failed their landing in the drone ship

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

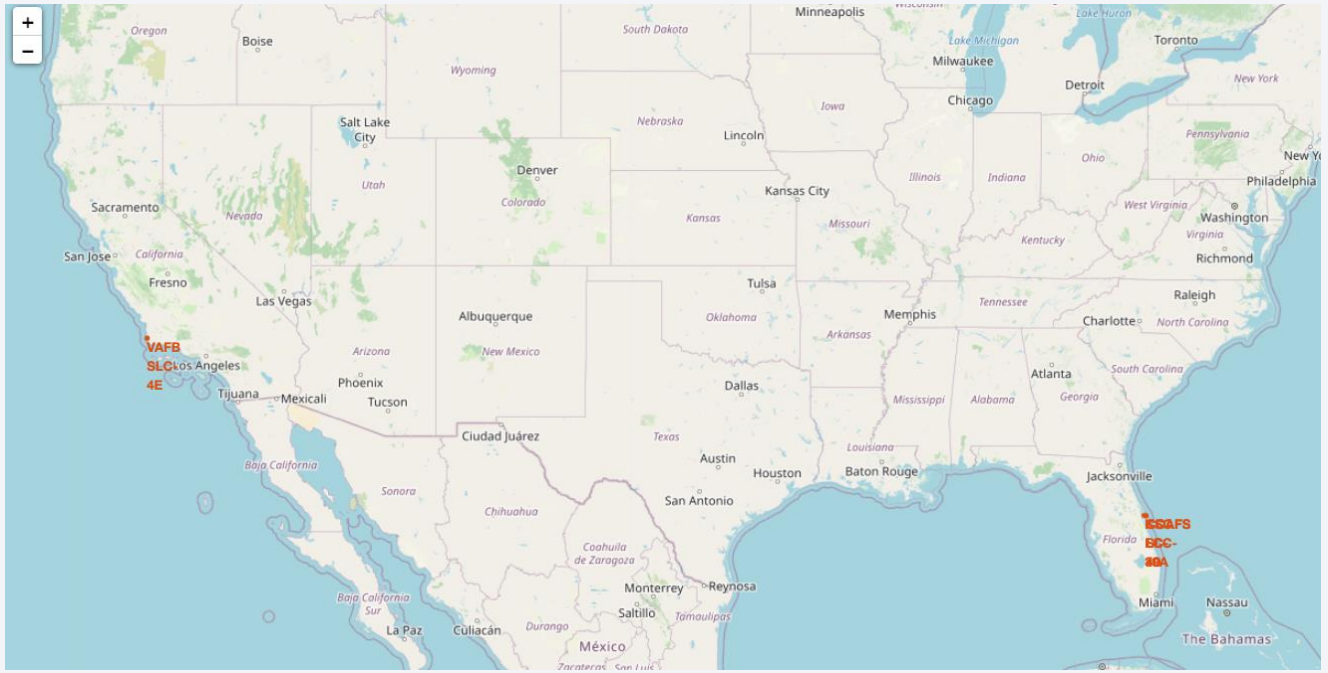
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This is the ranking of 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

Section 3

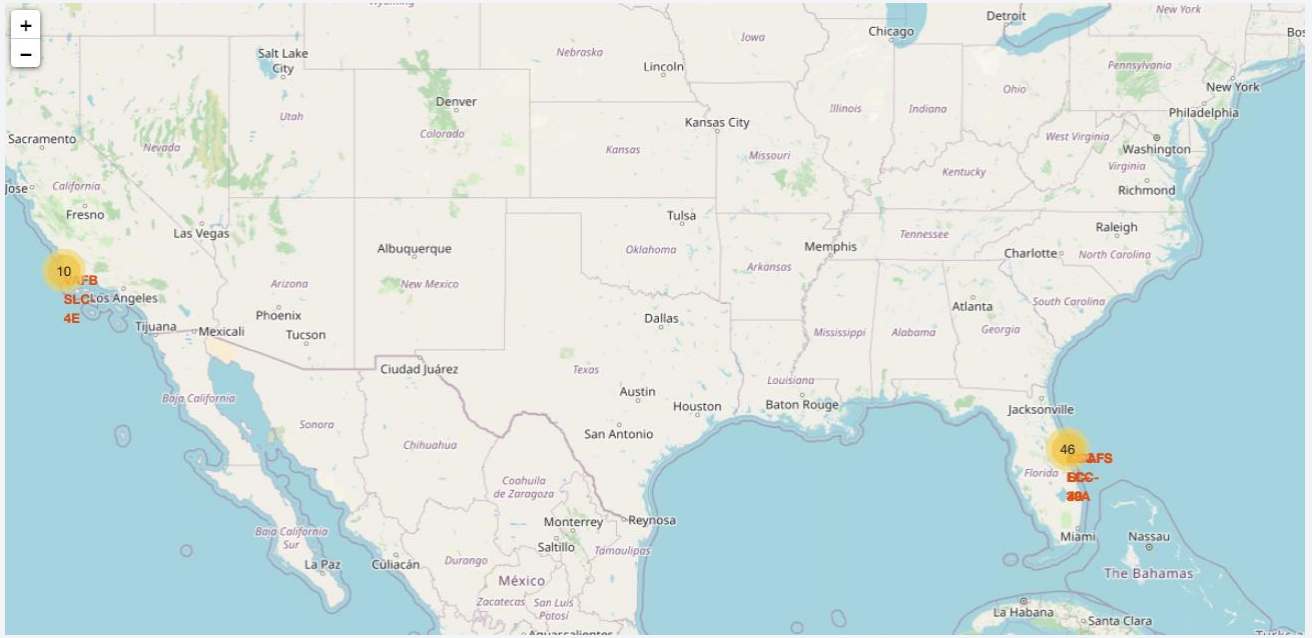
Launch Sites Proximities Analysis

Locations of All Launch Sites

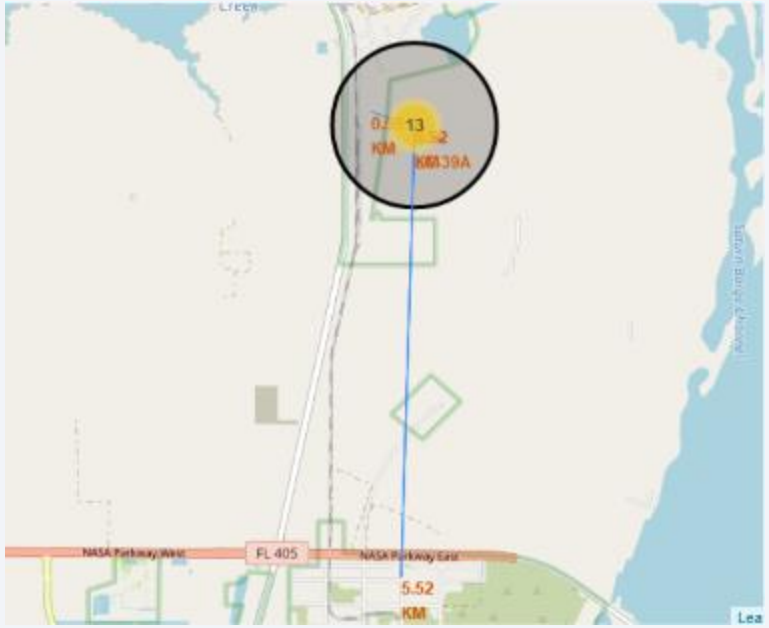


- The markers represent the location of each of the launch sites.
- It can be seen from the map above, all the launch sites are near the coastal area

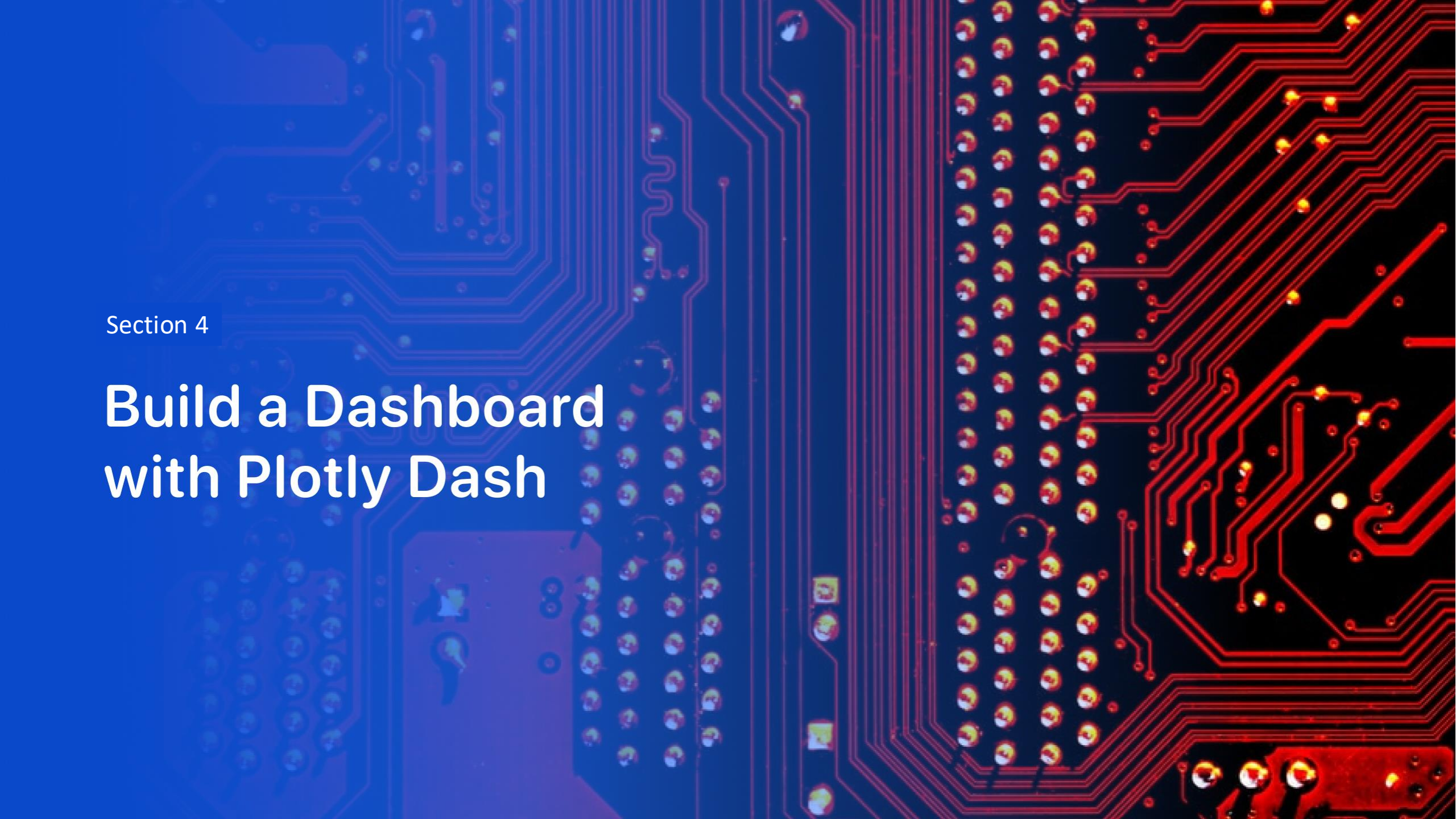
Launch Outcome for Launch Sites



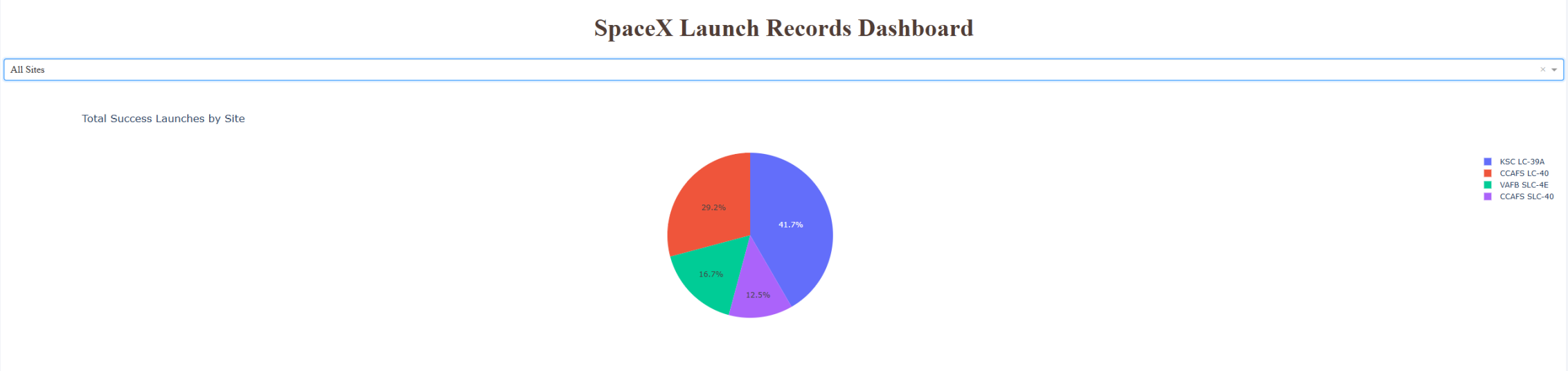
Distance between Launch Sites and Railway



- The blue line represents the distance between the railway and launch site KSC LC-39A

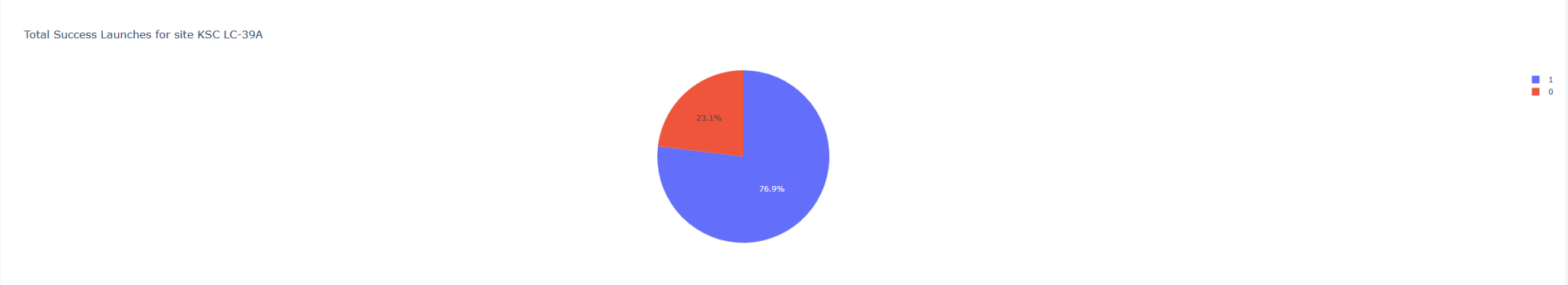

Section 4
Build a Dashboard with Plotly Dash

Success Count for All Launch Sites



- The pie chart displays the total success launches for each launch sites
- KSC LC-39A has the highest count of successful launches

Highest Launch Success Ratio

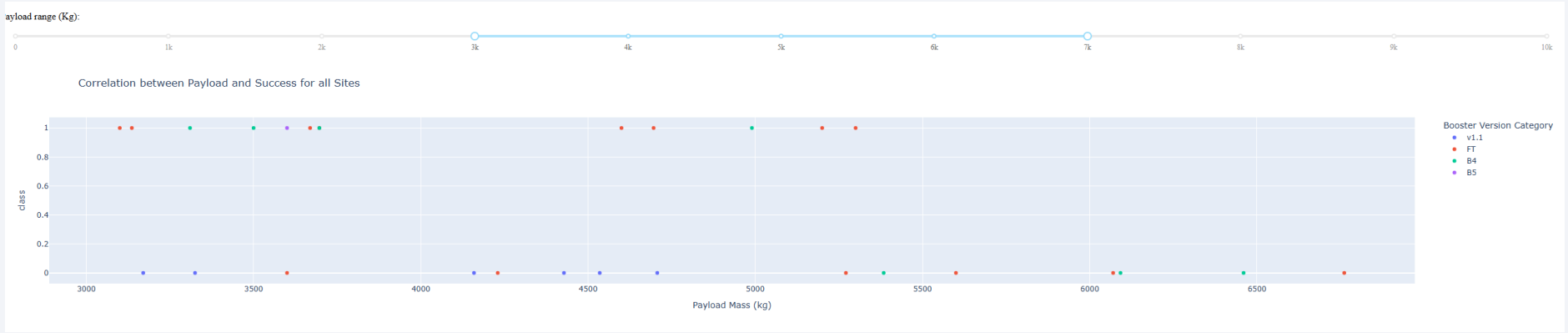


- KSC LC-39A has the highest launch success ratio
- KSC LC-39A has a 76.9% success rate to a 23.1% failure rate
- This shows KSC LC-39A is an amazing launch site

Payload vs Launch Outcome

- FT has a strong correlation between payload mass and the success of a launch

Payload vs Launch Outcome



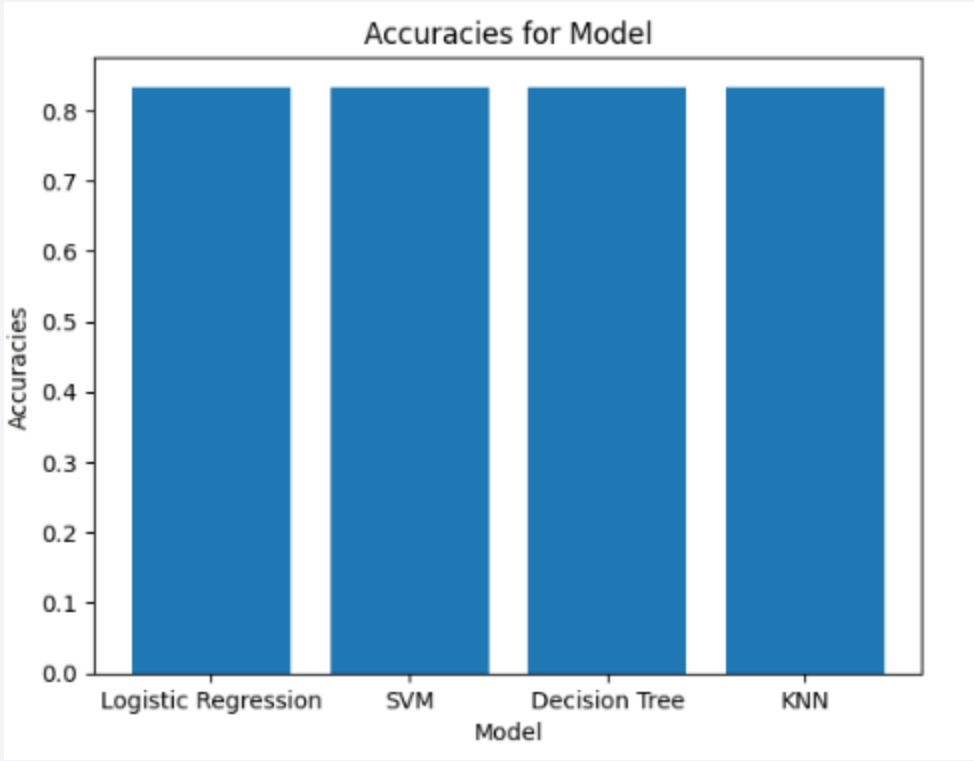
- FT also shown a strong correlation between the payload mass and the success of a launch between the payload mass of 3000 to 7000

Section 5

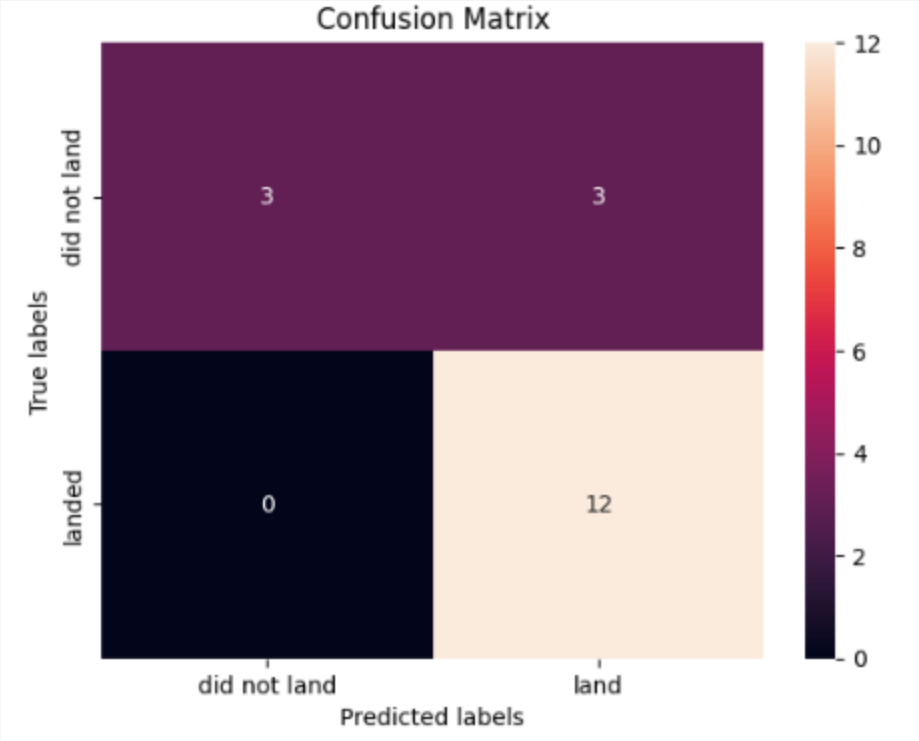
Predictive Analysis (Classification)

Classification Accuracy

- All the models have an accuracy of 83.33%



Confusion Matrix



- All the model have a precision score of 83.33%
- All models correctly predict 15 positive outcomes but falsely predicted 3 positive outcomes

Conclusions

- Collecting data from the correct sources is crucial for a successful data analysis and outcome prediction
- The success rate for launches improves as the years increases. This shows constant improvement in rocket engineering
- Coastal area are the best location to establish a launch site due to safety reasons
- KSC LC-39A is the most successful launch sites due to high successful rate with various payload mass
- All the models used are suitable and efficient to predict successful landings, which will lead to higher cost savings and increased safety

Thank you!