WINNING *space race*

WITH DATA SCIENCE

Ho Hsu Hsia / 2 May 2024

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

Background. SpaceY wants to predict the success rate of Falcon 9 first stage landing using various sources of data, machine learnings and testing models.

Methodology. Data was collected from the SpaceX public API and publically available data on Wikipedia. Data collected are flight number, payloas mass, launch dates, launch sites, mission outcomes and orbit types. Data wrangling included extracting launch outcome information to serve as the dependent variable in the Machine Learning models. SQL and data visualizations dashboards are used to discover insights and answer questions. Predictive analysis were conducted using machine learning such as Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) .

Results. With different visualisation and machine learning models, SpaceY will have higher success rate in the future for Falcon 9.

Introduction

Purpose of project.

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

Common problems.

- Factors impact on the successful rocket landing
- Conditions that SpaceY require for success landing rate
- Accuracy of data from different courses
- Relevancy of machine learning methods

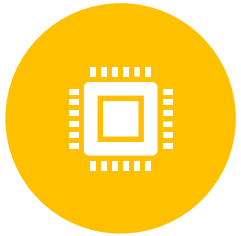
Section 1

Methodology

Methodology

2 sources of data collected: SpaceX API and Wikipedia (BeautifulSoup) table data.

Data was cleaned up before any machine learning model is applied.

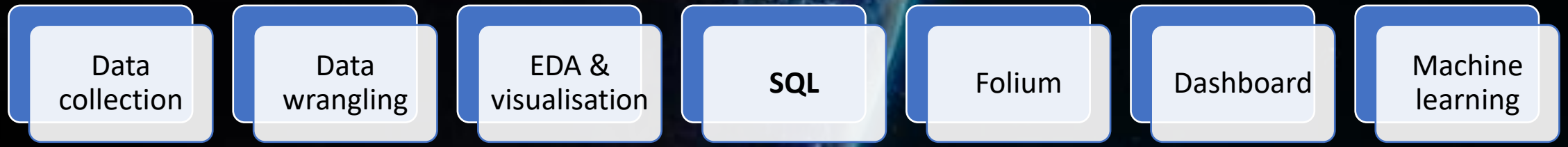


EDA and SQL performed to check on relationships of various factors for successful landings.

Folium and PlotlyDash were created for visualisation purposes

Predictive analysis using classification models

Data Collection



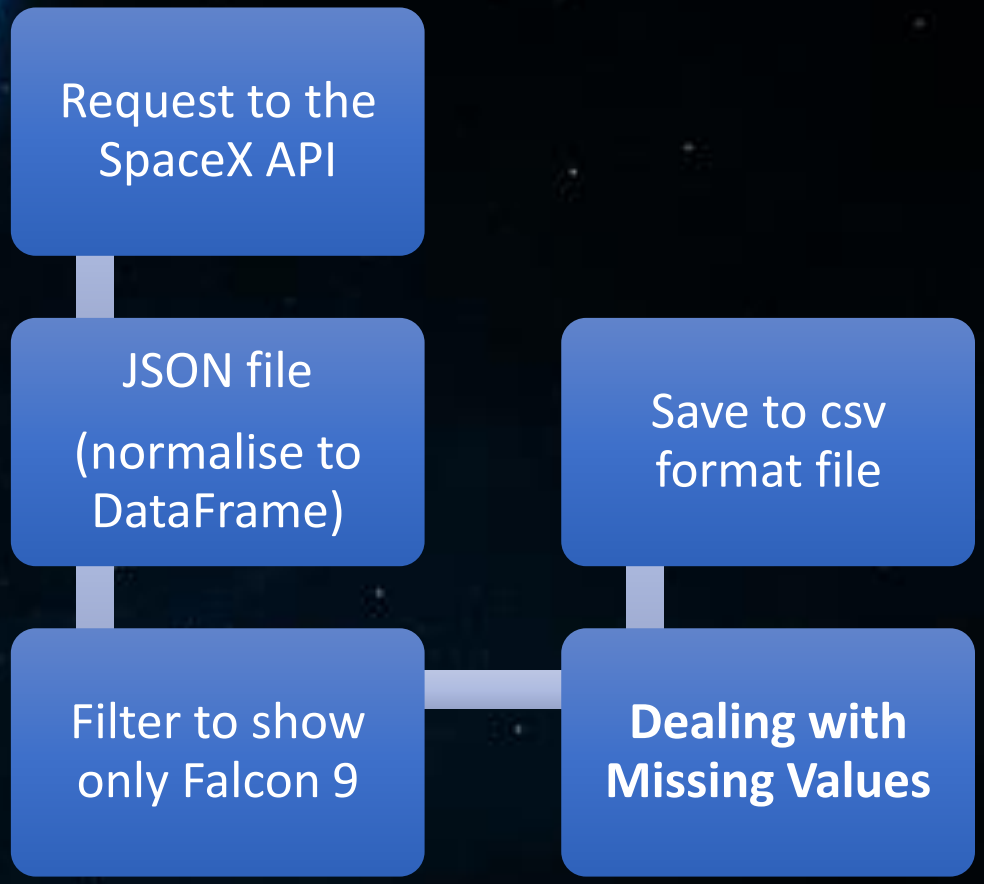
Data collection from sources such as SpaceX REST API requests web scrapping data from BeautifulSoup Wikipedia

Data Columns used include Flight Number, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, Version Booster and Booster landing

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%201/Wk1-Data%20Collection%20API%20Lab.ipynb

Data Collection – SpaceX API

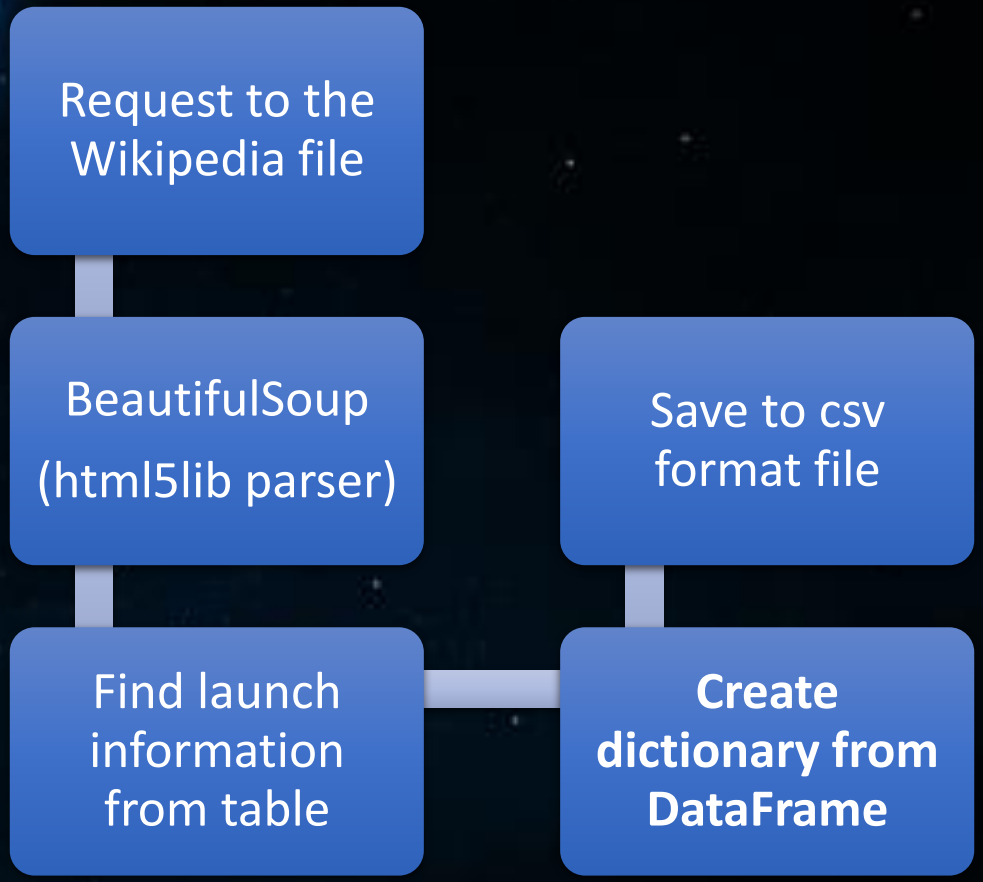
"After Importing Libraries and Defining Auxiliary Functions by APIs, these data requested will be stored in lists and will be used to create a new dataframe."



Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%201/Wk1-Data%20Collection%20API%20Lab.ipynb

Data Collection - Scraping

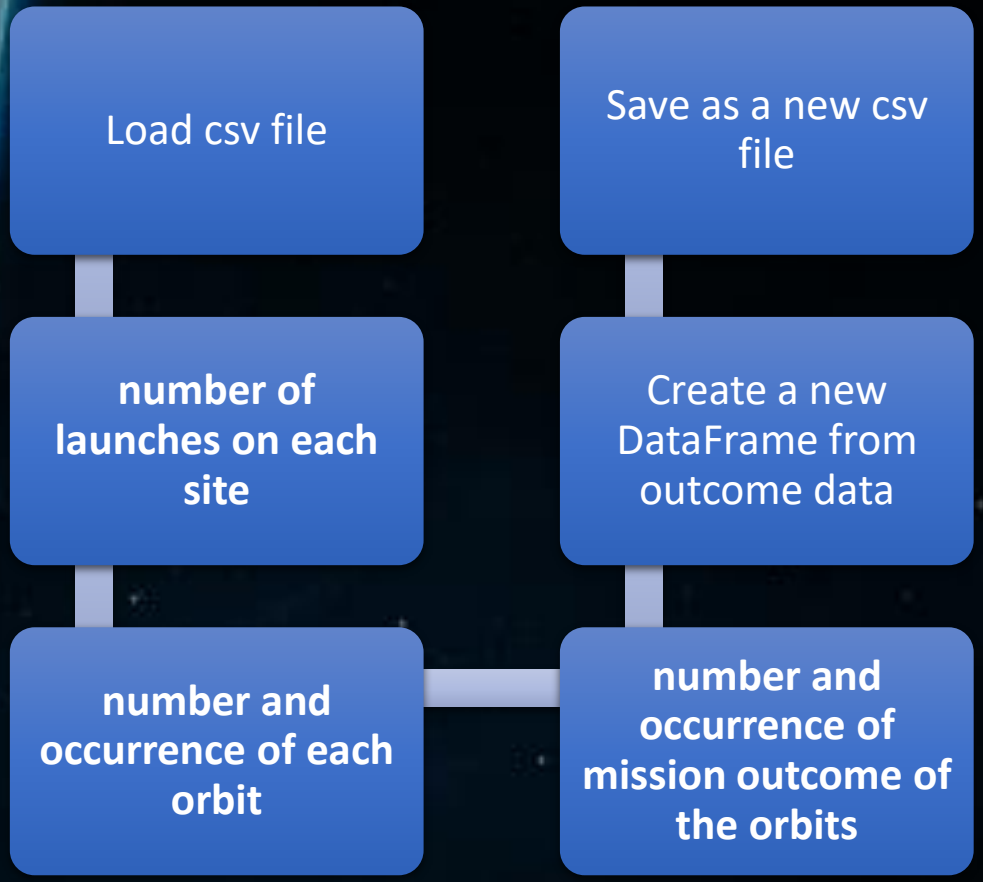
"Creating an object from HTML response and data frame by parsing the launch HTML tables."



Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%201/Wk1-webscrap.ipynb

Data Wrangling

"Loading SpaceX dataset to analyse important information such as number of launches, occurrence of mission outcome of orbits and landing outcome."



Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%201/wk1-wrangling.ipynb

EDA with Data Visualization

Chart type	Relationship
Scatterplot	Launch Site and Flight Number
Scatterplot	Launch Site and Payload
Bar chart	Orbit Type
Scatterplot	Orbit Type and Flight Number
Scatterplot	Orbit Type and Payload
Line plot	Year

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-eda%20visual.ipynb

EDA with SQL

Queries including:

- Finding the unique launch sites begin with the string 'CCA'
- total payload mass carried by boosters launched by NASA (CRS)
- Calculating the average payload mass carried by booster
- Finding those boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Calculate the total number of successful and failure outcomes
- Finding those boosters carrying the maximum payload mass
- Rank the count of landing outcomes from 2010 to 2017

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Build an Interactive Map with *Folium*

- Markers added to the various launch sites
- Marker also include NASA Johnson Space Center
- Lines to indicate proximity to nearby railway, highway coast and city

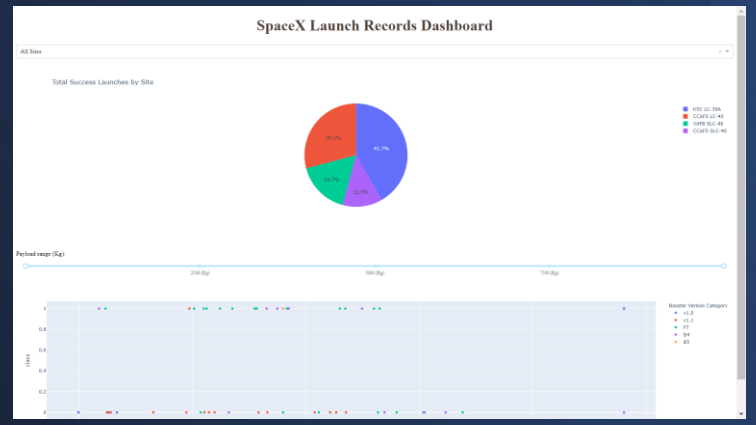
Github link:
https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk3-folium.ipynb

Build a Dashboard with *Plotly Dash*

Dashboard comprises of a pie chart and a scatter plot.

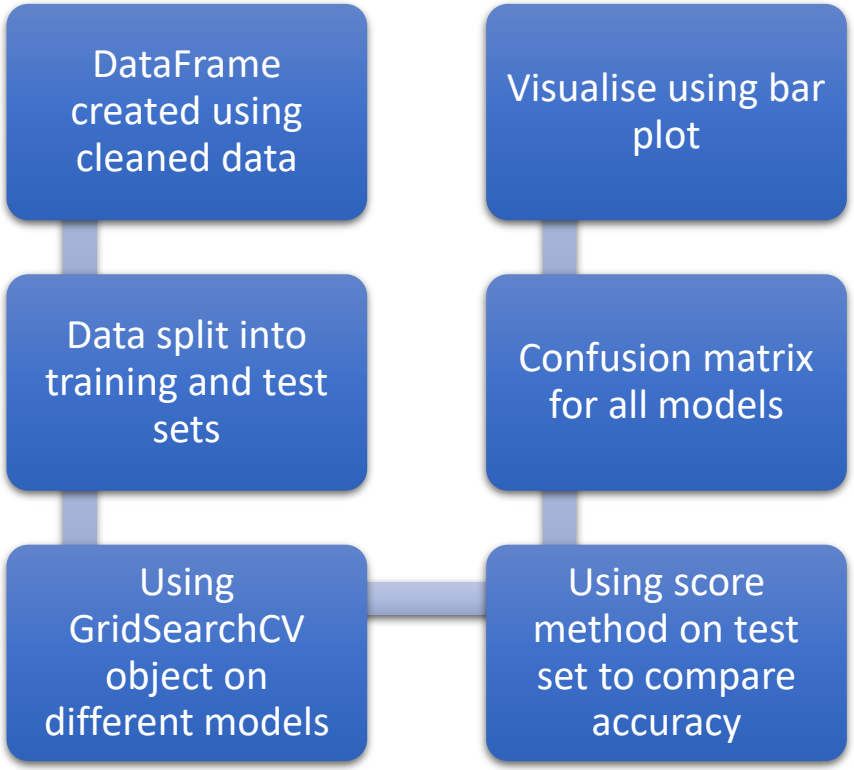
Pie chart is used to show distribution of successful Falcon 9 first stage landings across all launch sites. It is to visualize success rate of the sites.

Scatter plot has 2 inputs slider: payload mass and distribution to show the success rate across variables such as payload mass, booster version and launch sites.



Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk%203%20-dash_app.py

Predictive Analysis (Classification)



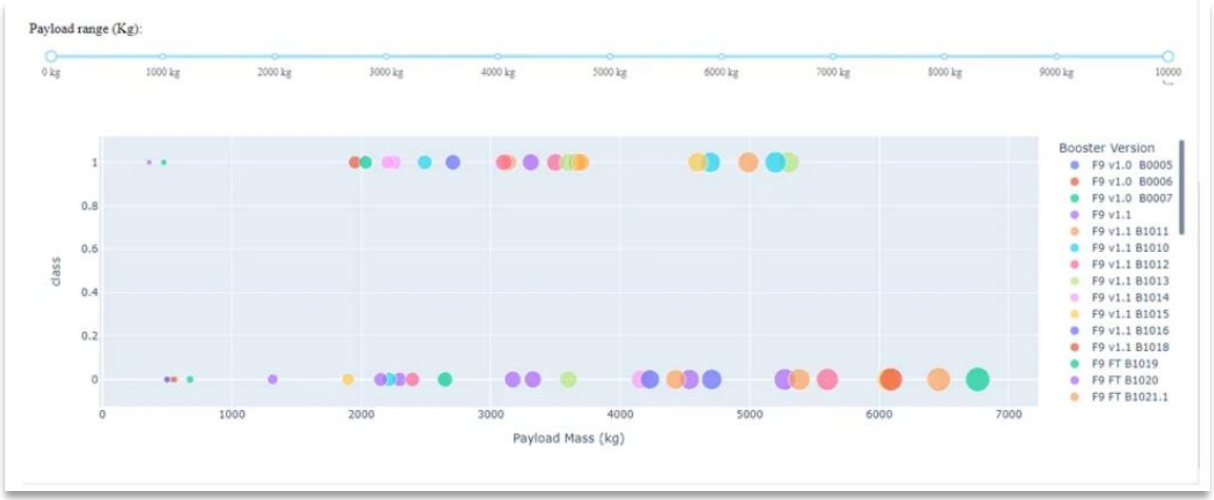
2 sets of data used for training and testing which was split from the dataset.

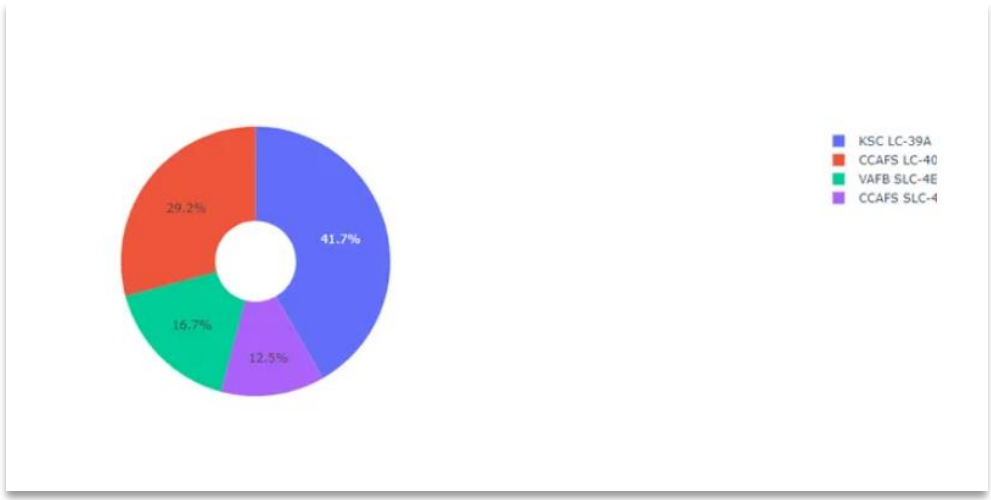
Various machine learning models are applied on the training set: Decision Tree, Logistic Regression, Support Vector Machine, k-Nearest Neighbours

GridSearchCV() used for evaluating on test set for the scoring on accuracy to decide the best method.

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb

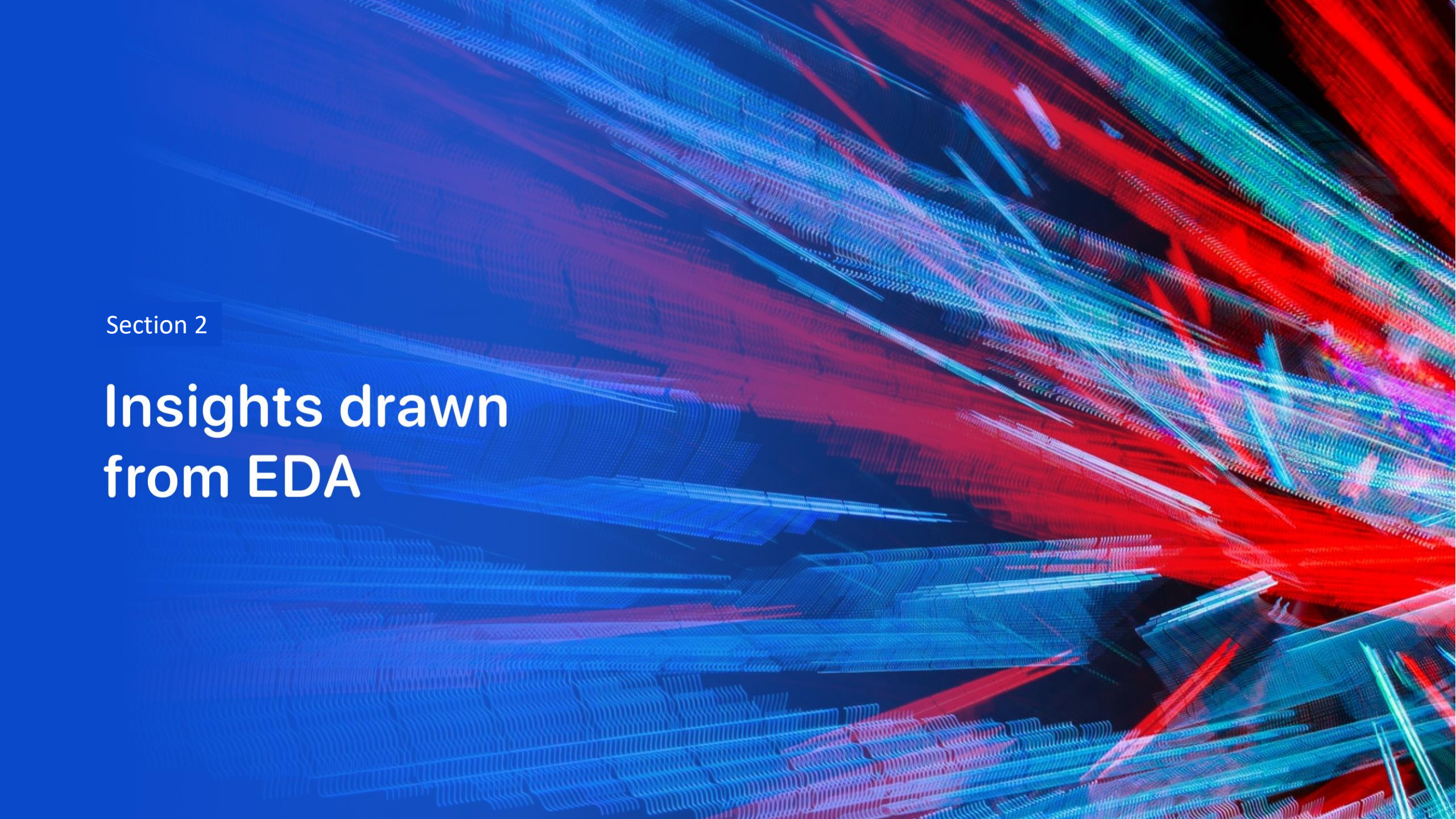
Results



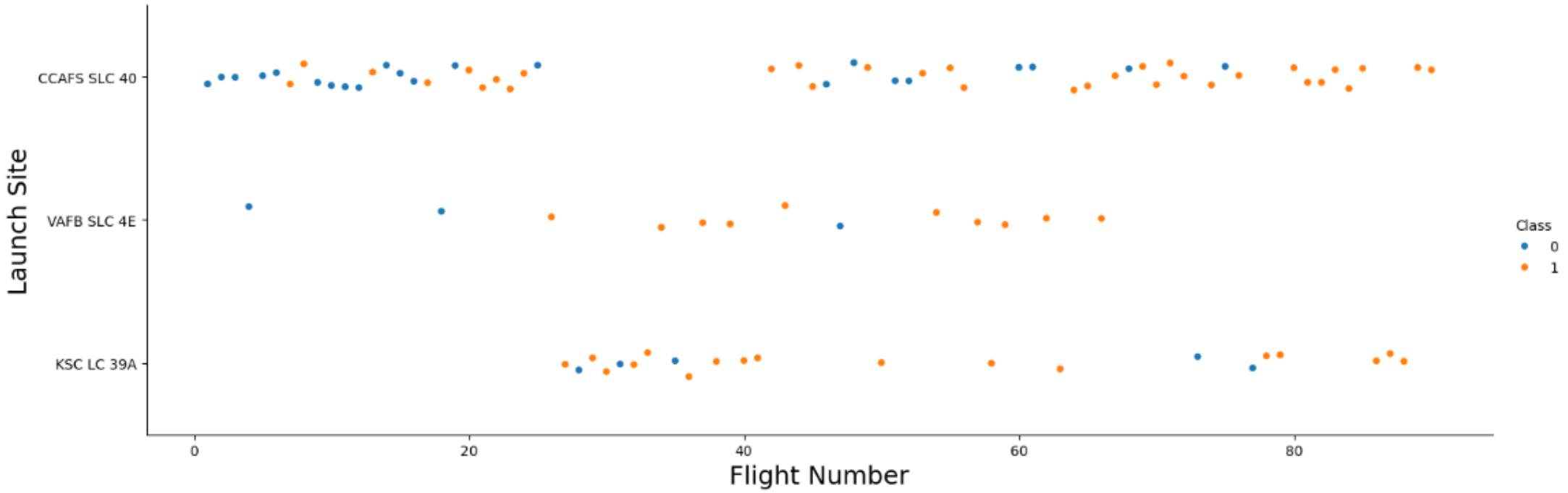


In conclusion, the decision tree model show ***satisfactory*** results. The F1-score is at 94%. Based on the above confusion matrix, the machine learning model is slightly better at predicting if the first stage will land s***uccessfully***.

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb

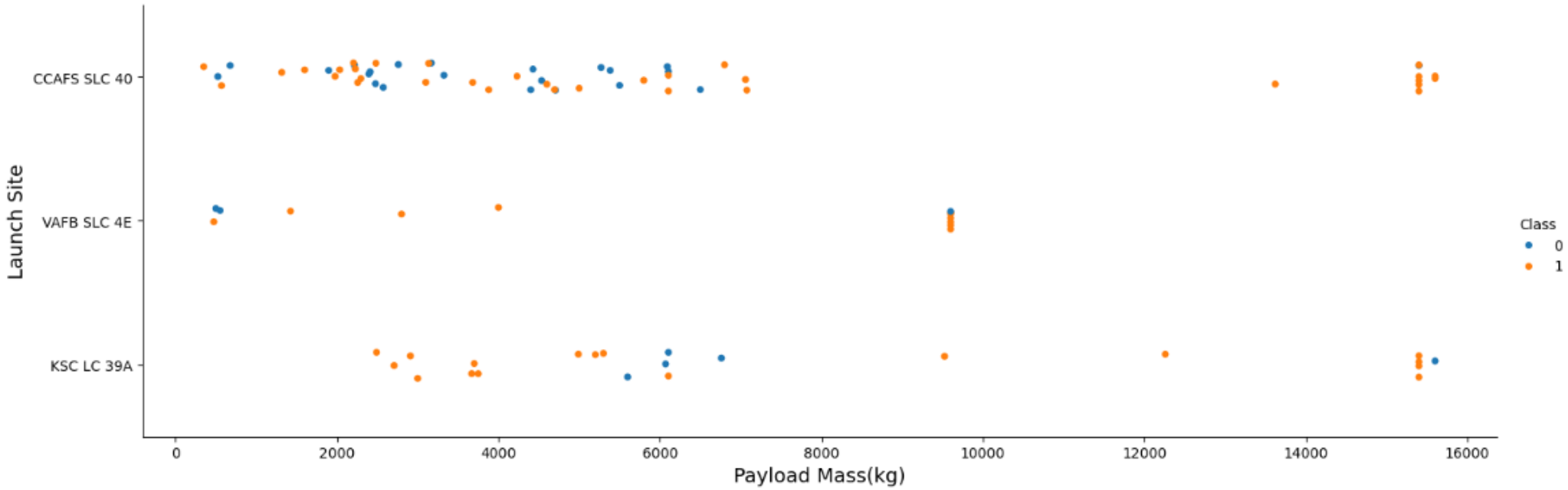

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- *Success rate depends on launch site*
- *Falcon 9 landings are more successful as the flight number increases*

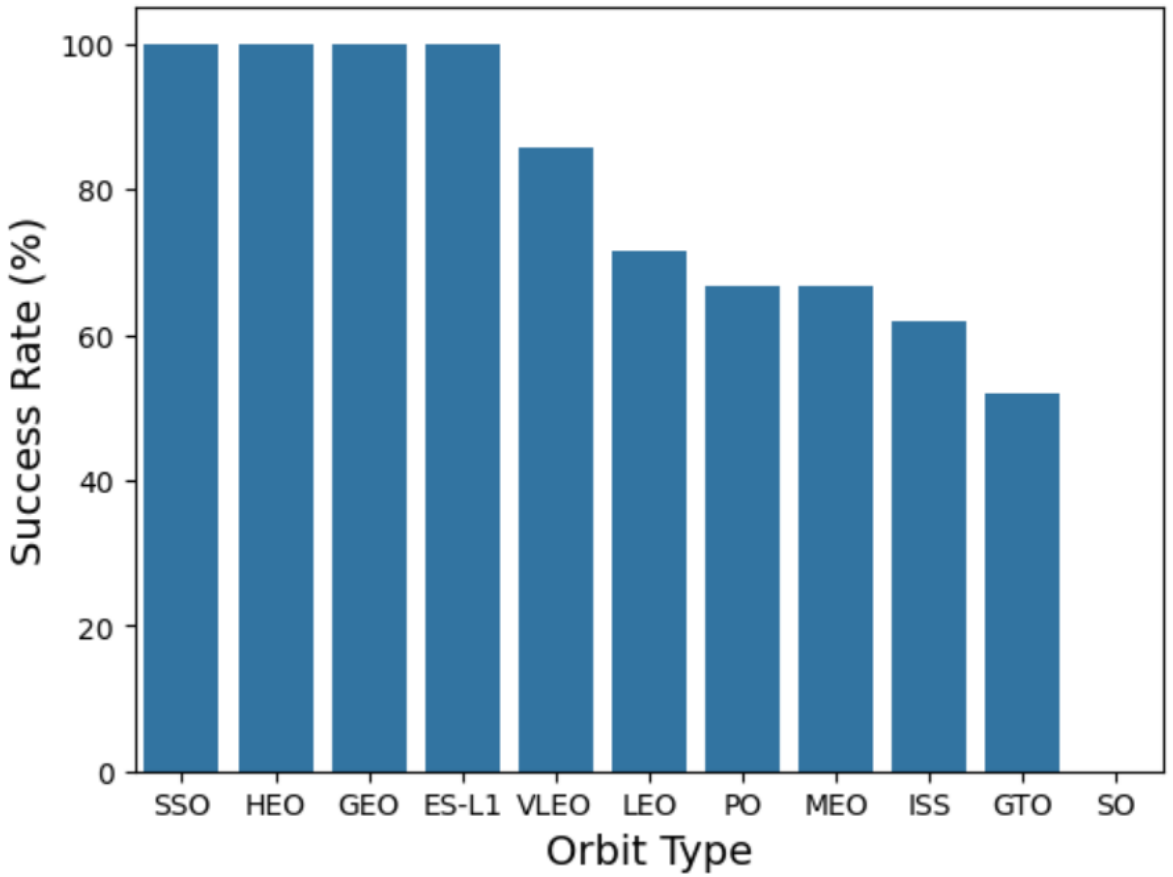
Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb



- *Correlation between launch site and payload for CCAFS SLC 40 is stronger than the others.*
- *KSC LC 39A has more failed landings when the payload is less than 6000.*

Payload vs. Launch Site

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

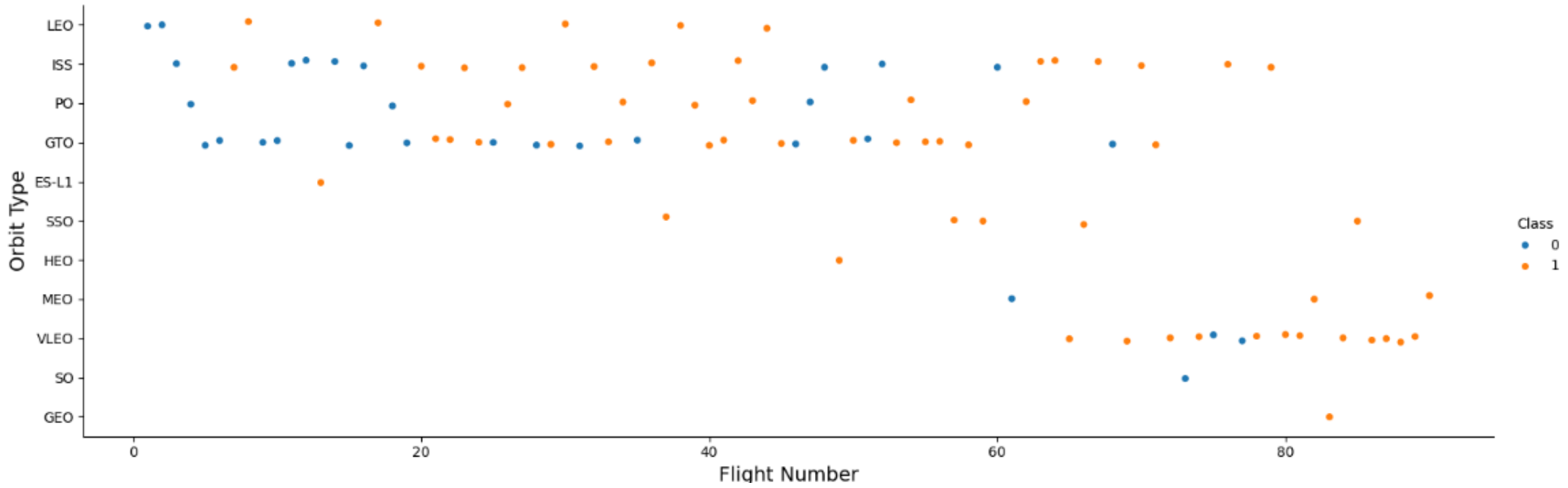


- *SSO, HEO, GEO and ES-11 has 100% success rate*
- *GTO is the lowest*
- *SO has zero*

Success Rate vs. Orbit Type

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

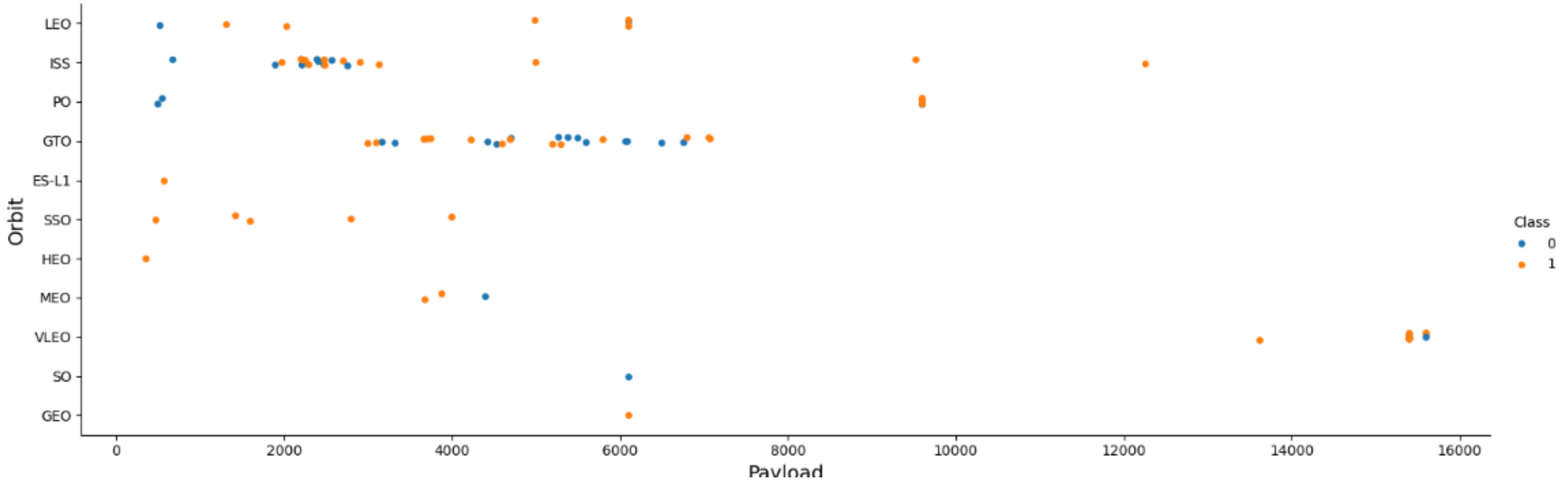
Flight Number vs. Orbit Type



- *Flight number is strongly correlated with orbit type*
- *With a larger flight number, it has a higher success rate*

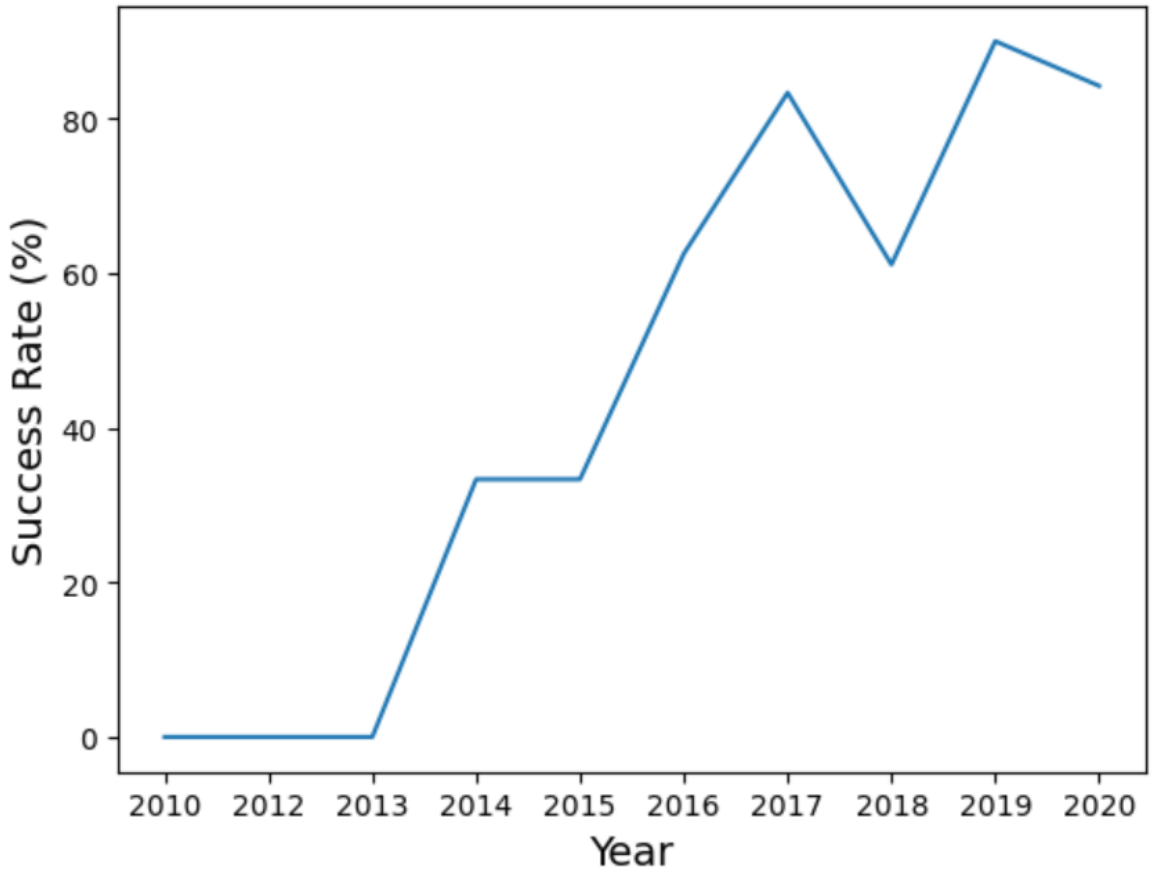
Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Payload vs. Orbit Type



- *Payload has little correlation with orbit*
- *VLEO is more successful with the higher payload*

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb



- *Success rate increases over time though it was increasing only after 2013*
- *There is a dip in 2018*

Launch Success Yearly Trend

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Launch Site Names Begin with 'CCA'

This query will return the top 5 records of the launch sites starting with CCA

```
%%sql
SELECT LAUNCH_SITE
FROM SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Total Payload Mass

This query is calculating the total payload in kilogram by the client NASA which is 45596kg.

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Customer = 'NASA (CRS)';
```

SUM(PAYLOAD_MASS__KG_)
45596

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Average Payload Mass by F9 v1.1

This query is calculating the average payload in kilogram carried by booster version F9 v1.1 which is 2928kg.

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.1';
```

AVG(PAYLOAD_MASS__KG_)
2928.4

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

First Successful Ground Landing Date

This query showed the first successful landing which is in Year 2015, though landing has been launched since 2013.

```
%%sql
SELECT MIN(Date)
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
```

MIN(Date)
2015-12-22

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Successful Drone Ship Landing with Payload between 4000 and 6000

This query returned 14 booster version that are successful drone ships landings with a payload of between 4000 to 6000 kg.

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (drone ship)'
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Total Number of Successful & Failure Mission Outcomes

- This query returned the number of outcomes if it is successful or failure.
- The success rate is high at 98 number of mission outcomes with 1 failed mission.
- However, there is an unknown payload status but with successful mission outcome.

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

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

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Boosters Carried Maximum Payload

- This query returned those boosters that carried the highest payload.
- Total of 12 booster.

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

```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL);
```

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

2015 Launch Records

- Total of 2 failed outcomes in 2015 which is in Jan and Apr.
- Both located in the same site of CCAFS LC-40

```
%%sql
select substr(Date, 6,2) as Month,
Landing_Outcome, Booster_Version, Launch_Site
from SPACEXTABLE where Landing_Outcome =
'Failure (drone ship)' and substr(Date,0,5) = '2015'
```

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

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

Rank Landing Outcomes (2010-06-04 ~ 2017-03-20)

- The highest outcome was “no attempt” at 10
- The success outcome of “drone ship” is higher than “ground pad”

```
%%sql
SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

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

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%202/Wk2-sql.ipynb

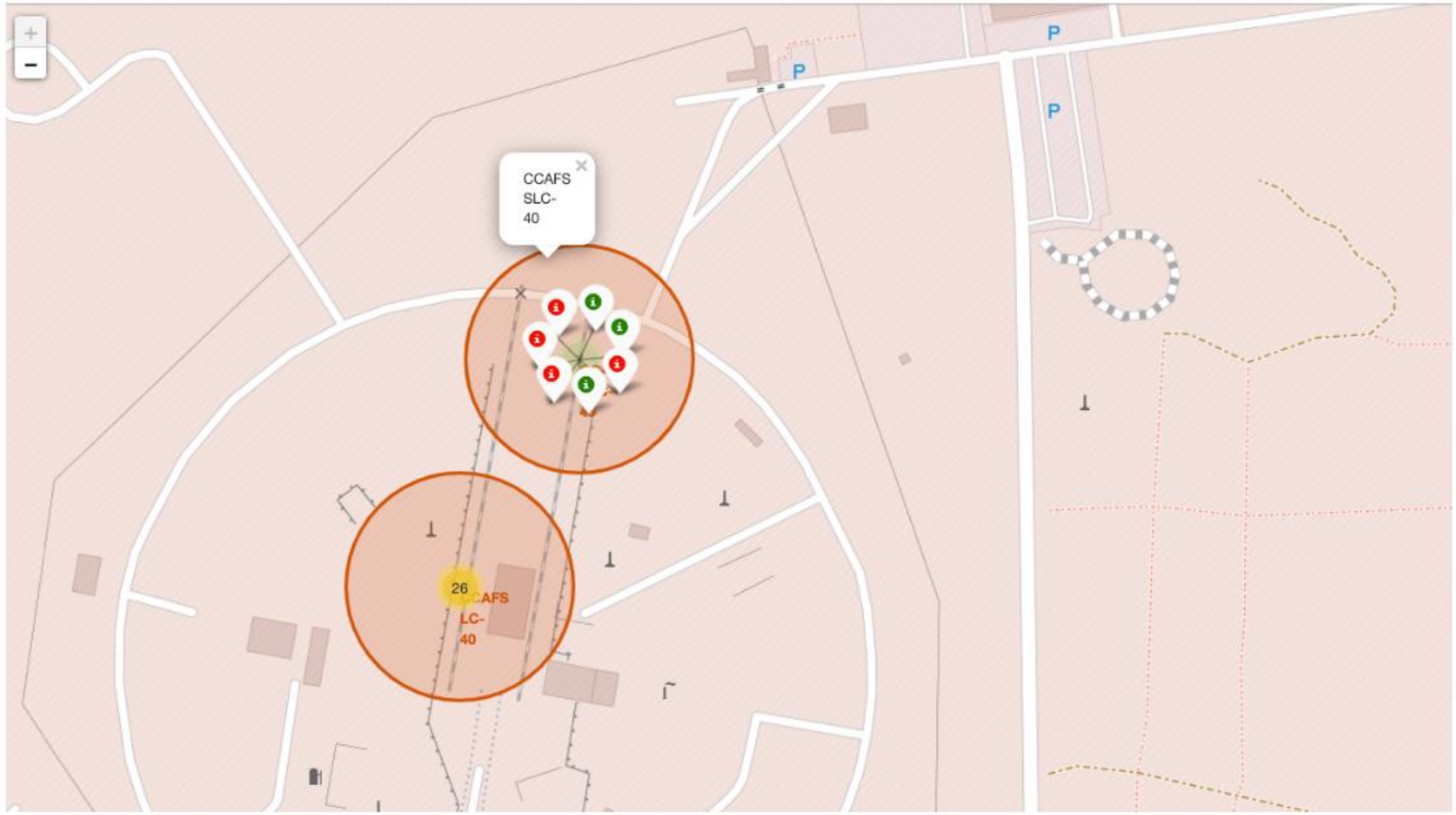
Section 3

Launch Sites Proximities Analysis

Location of Launch Sites

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk3-folium.ipynb

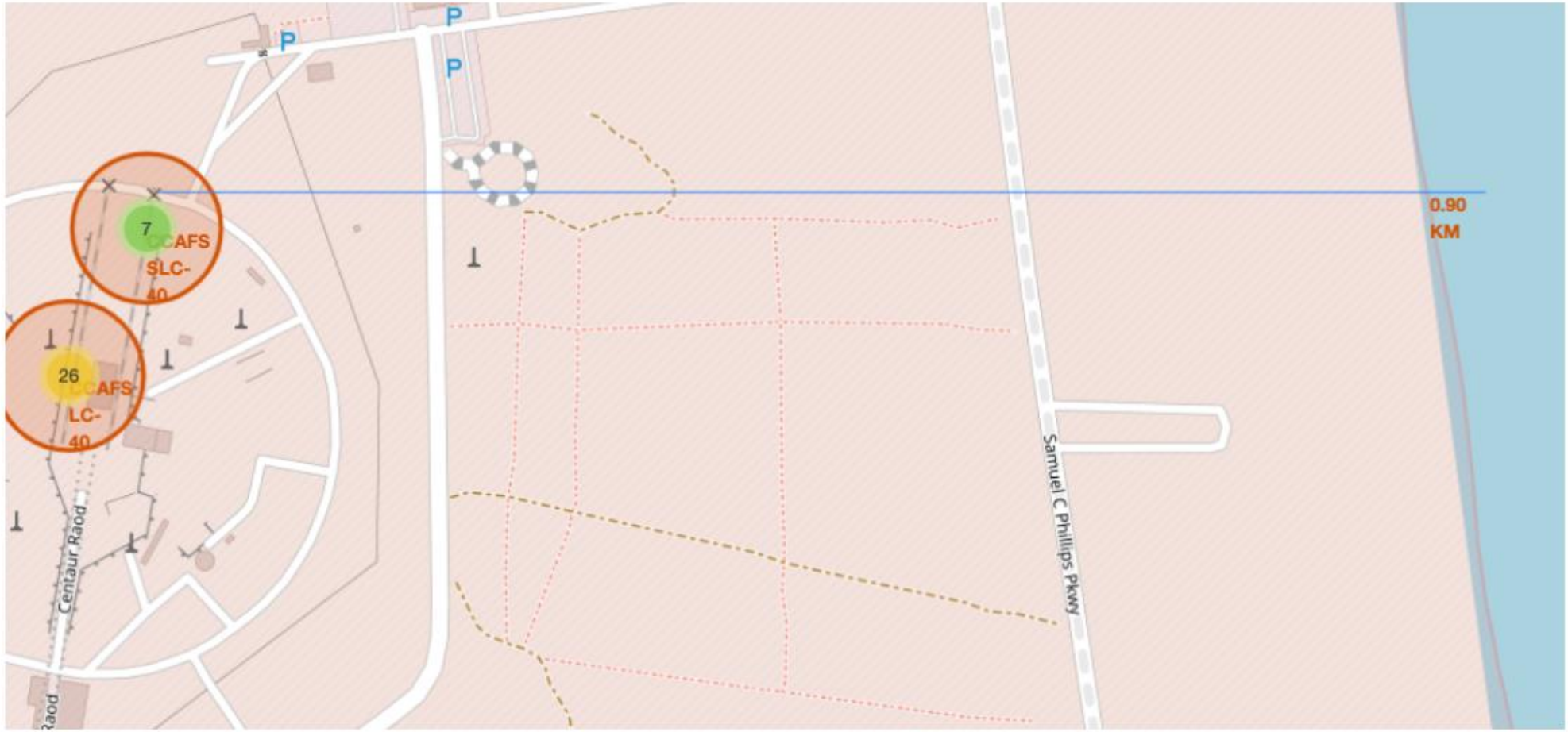
This map shows the location on the globe which can be zoom in and out.



Indication of success or failed outcomes by locations

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk3-folium.ipynb

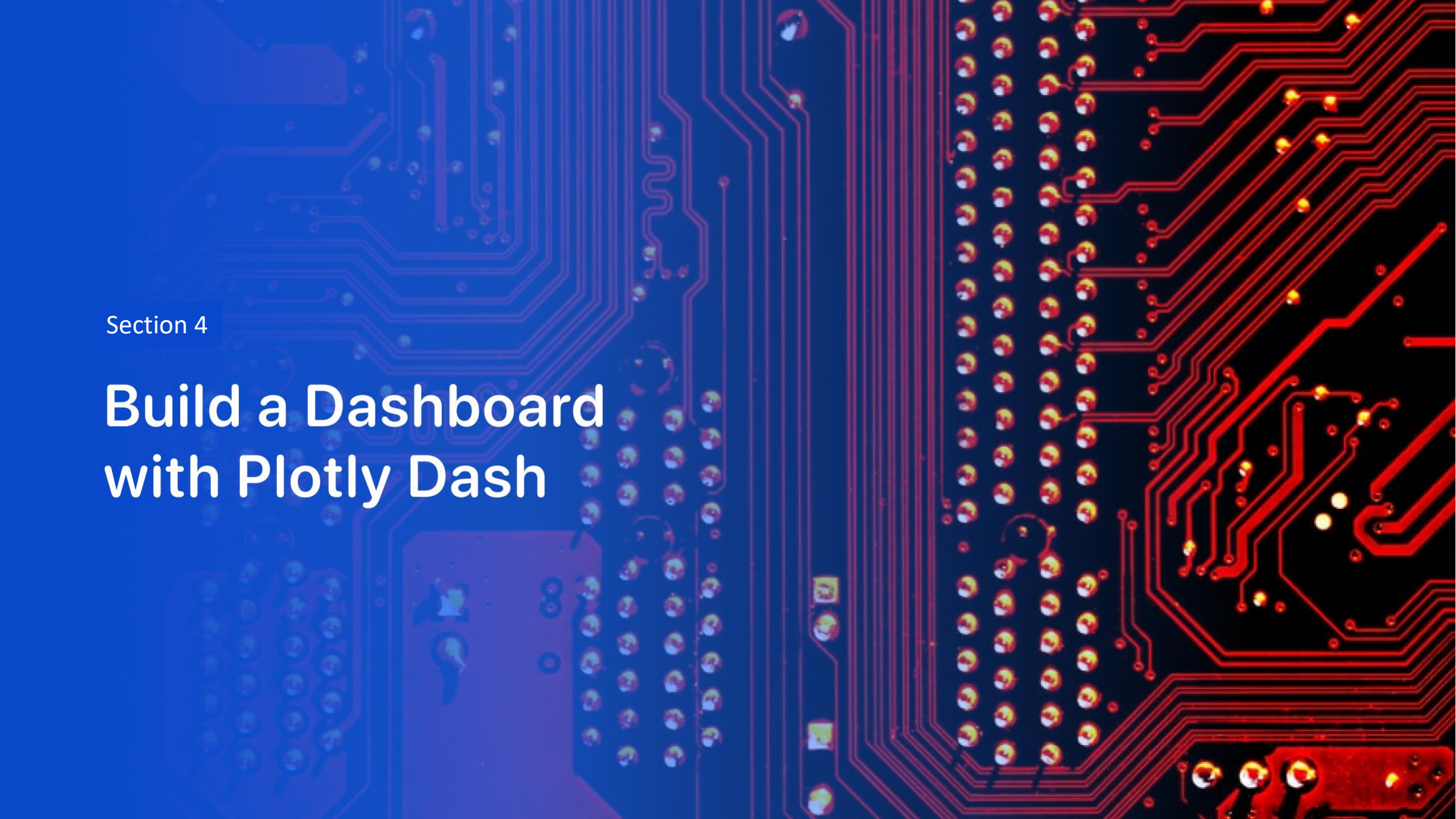
This map shows the launch site with success and failed outcomes in different colours.

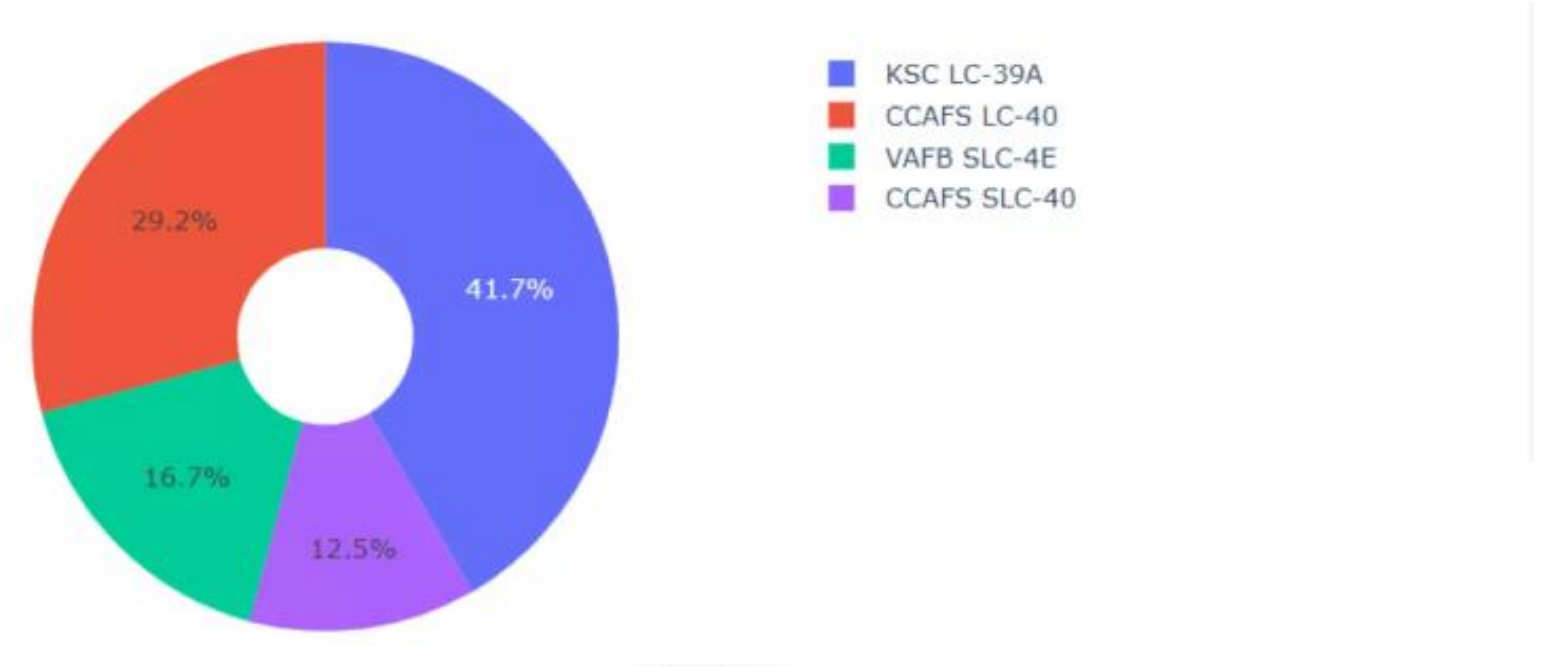


Proximity to landmarks

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk3-folium.ipynb

This map shows the distance from main landmark such as railway or highway coast.

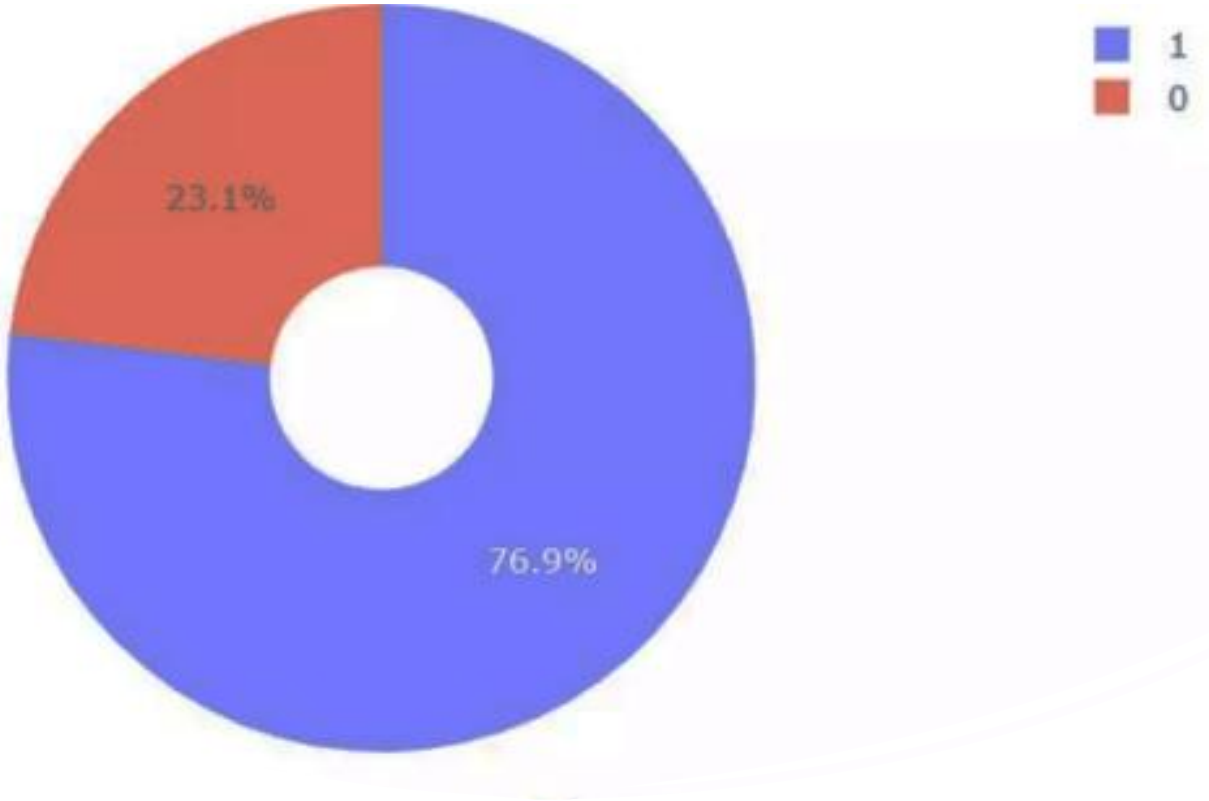

Section 4
Build a Dashboard with Plotly Dash



Successful Launches

- From the donut plot, it is clear that KSC LC-39A has the most successful launches from all the sites (41.7%).
- On the contrary, CCAFS SLC-40 has the lowest successful launches which is not even 30% of KSC LC-39A.

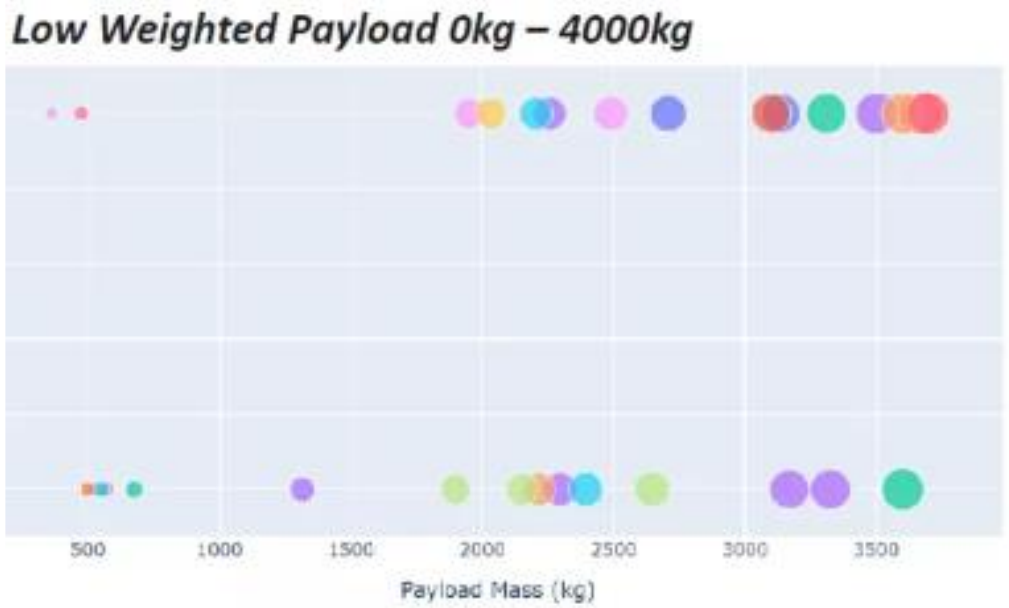
Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk%203%20-dash_app.py

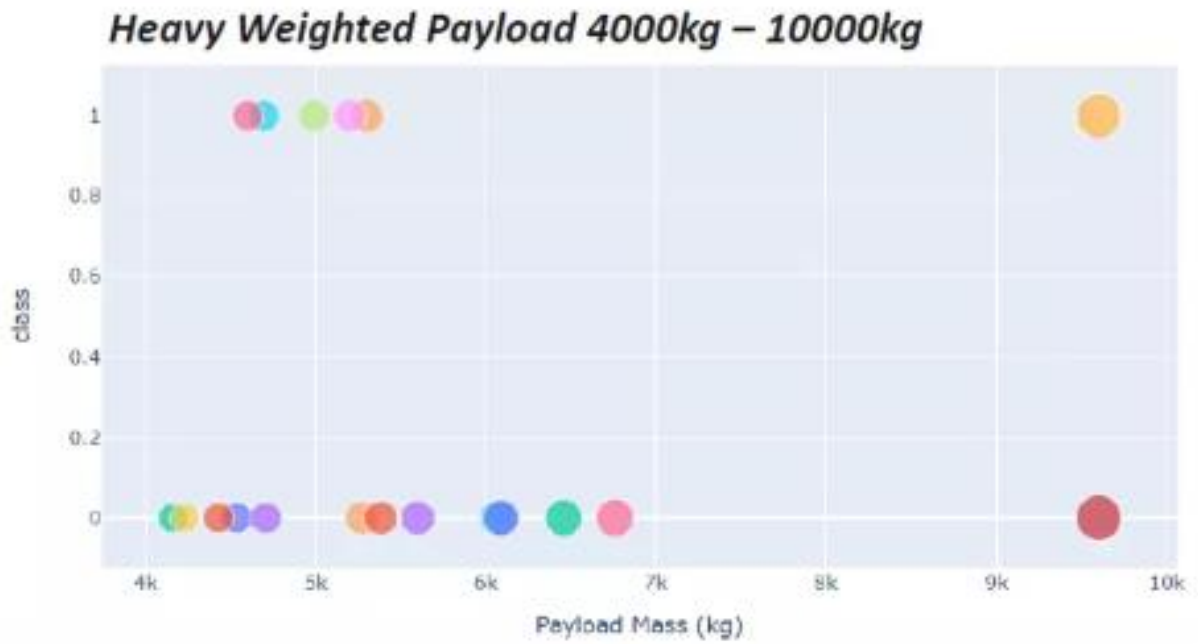


Success rate of launches

- The legend indicate “successful landings” by “1” in blue.
- KSC LC-39A has a high ratio of 76.9% success rate of all landings.

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk%203%20-dash_app.py





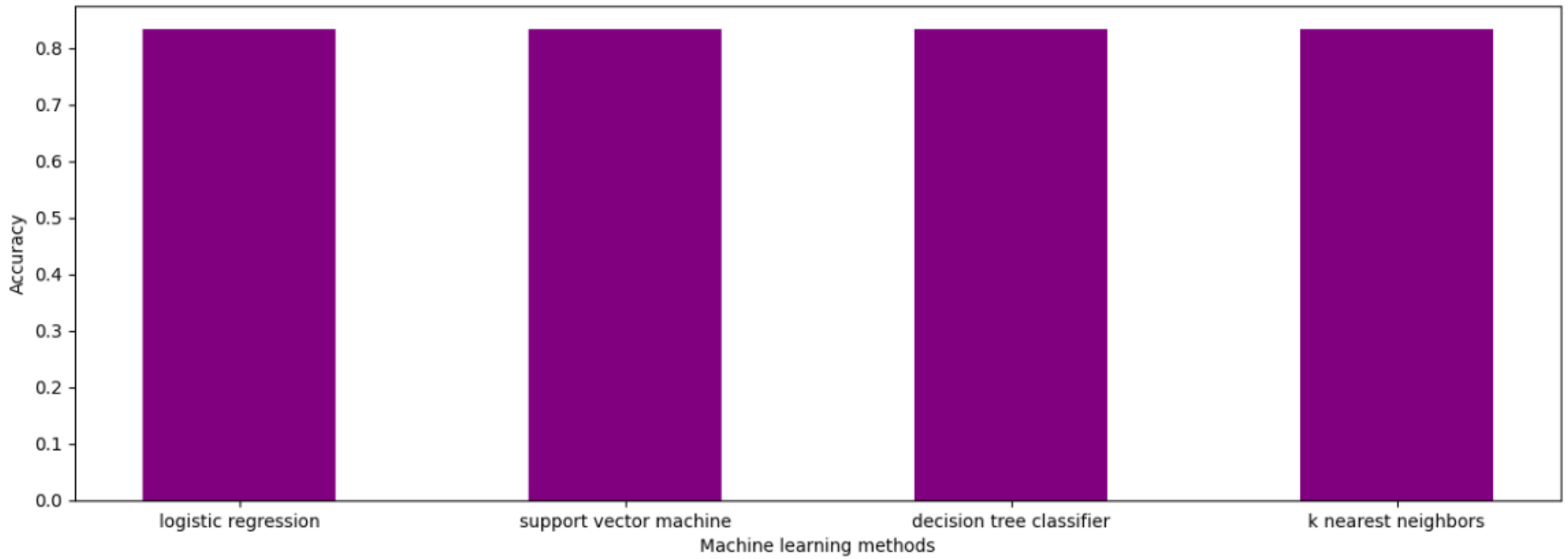
Payload by class

- The low weighted payload is up to 4000kg, while the heavier payload is from 4000kg onwards.
- By comparing the 2 charts, the low payload has higher successful landings than the heavy weighted payload

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%203/Wk%203%20-dash_app.py

Section 5

Predictive Analysis (Classification)



Classification Accuracy

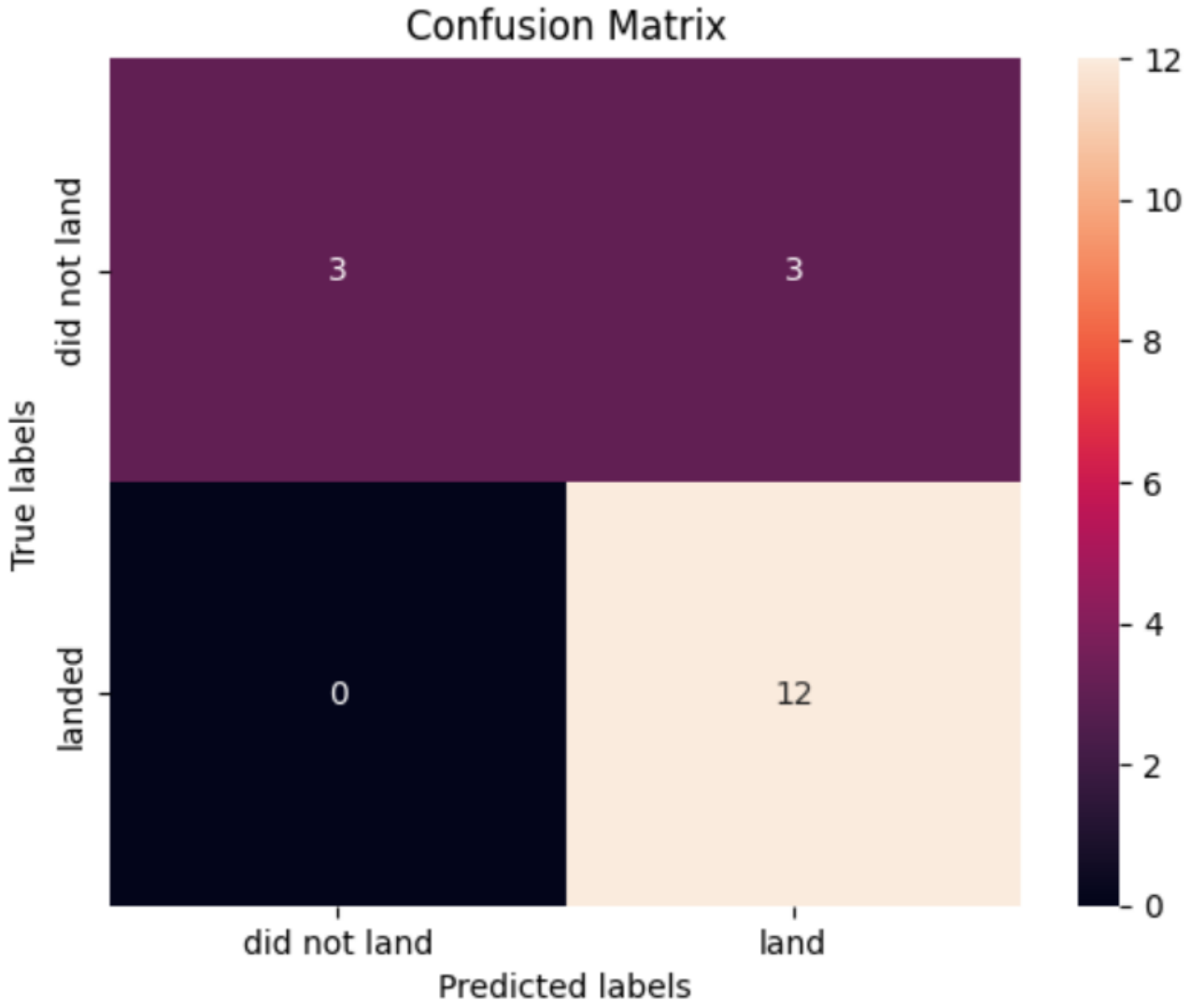
The 4 machine learning methods has equally the same accuracy of ***84%***

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb

Confusion Matrix

All 4 methods have the same results”

- 12 true positive and 3 true negative
- 3 false positive and 0 false positive



Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb

Conclusions

KSC LC-39A has the most successful launches

Falcon 9 first stage landing outcomes are getting higher successful rates over the time in future

Any machine learning models are accurate enough to predict the accuracy

Spacey will be able to predict its costing more accurate in future

Github link: https://github.com/HoHsuHsia/final-project/blob/main/Week%204/Wk4-prediction.ipynb

Appendix

SpaceX API https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Wikipedia https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

SpaceX (CSV) https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01

Launch Dash (CSV) https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv