IBM APPLIED DATA SCIENCE CAPSTONE

WINNING SPACE RACE WITH DATA SCIENCE

TYLER LEE ANG REN

7TH NOVEMBER 2024

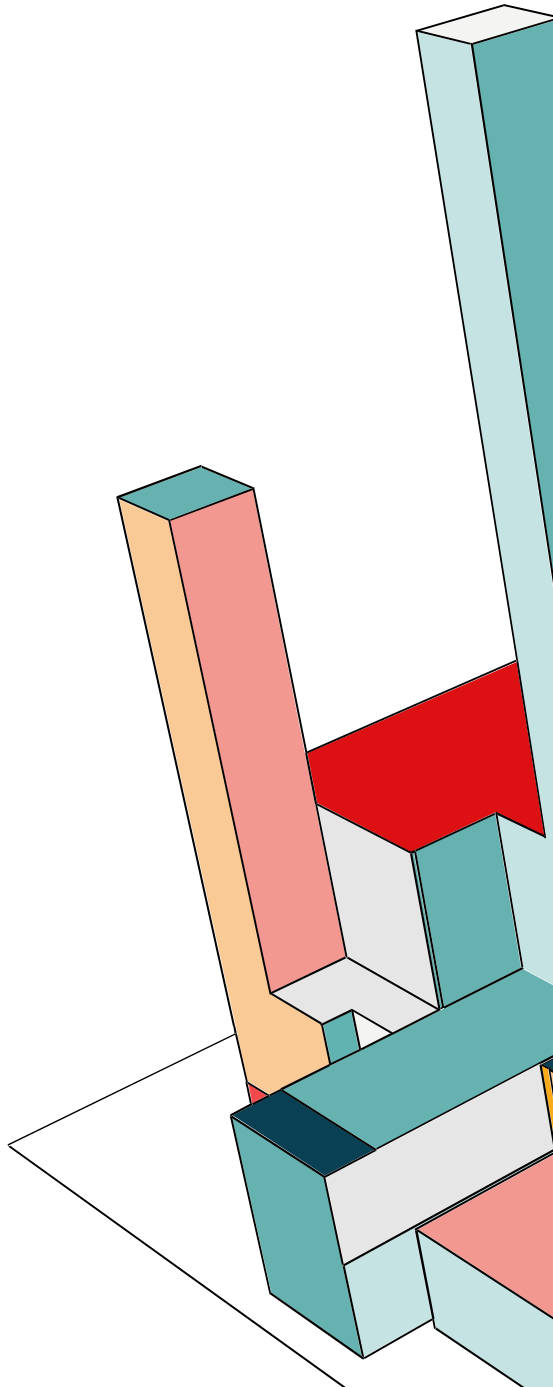
HTTPS://GITHUB.COM/TYLERKON

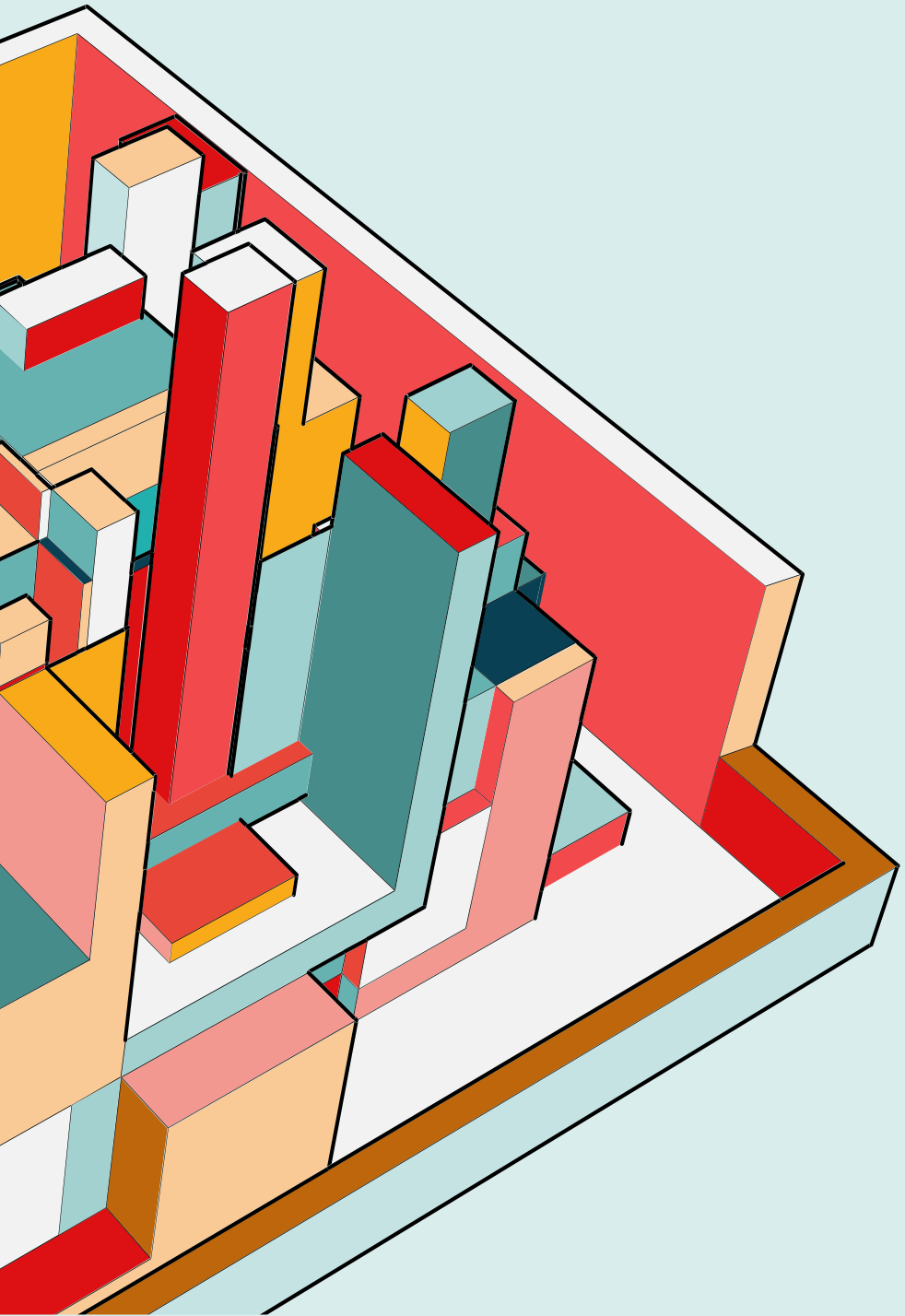
OUTLINE OF PROJECT

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

EXECUTIVE SUMMARY

- The current project has utilized the public SpaceX API and web-scraping from the SpaceX Wikipedia for data collection. The column 'class' was created to indicate the general outcome of landing. The data were explored with visualizations from python libraries, SQL queries, Folium maps and a Plotly Dash Dashboard. Machine Learning were then performed on the data with GridSearchCV for the best parameters, standardizing data, and measured accuracies of each models for the best model. The accuracies of models were then visualized
- Four machine learning models (Logistic Regression, Support Vector Machine, Decision-Tree Classifier, and K-Nearest Neighbors) were produced to perform classification prediction on the dataset. All models have produced similar results of 83.33% of model accuracy. The confusion matrix has shown a similar prediction across all model with only 3 false positives prediction. The results have suggested that more samples are needed for a better model accuracies.





INTRODUCTION

Project Background and context

- As the commercial space age arrives, more companies are making travel in space possible and affordable for everyone.
- SpaceX being the most successful one of all, was able to reuse their first stage landings, and save up costs of over 100 millions as compared to other providers.
- Space Y, plans to compete with SpaceX

Problem Statement

- As the first stage are the most expensive, we are to predict the success rate of the first stage
- To determine the price of each launch from the success rate of the first stage
- To predict whether if SpaceX will reuse the first stage

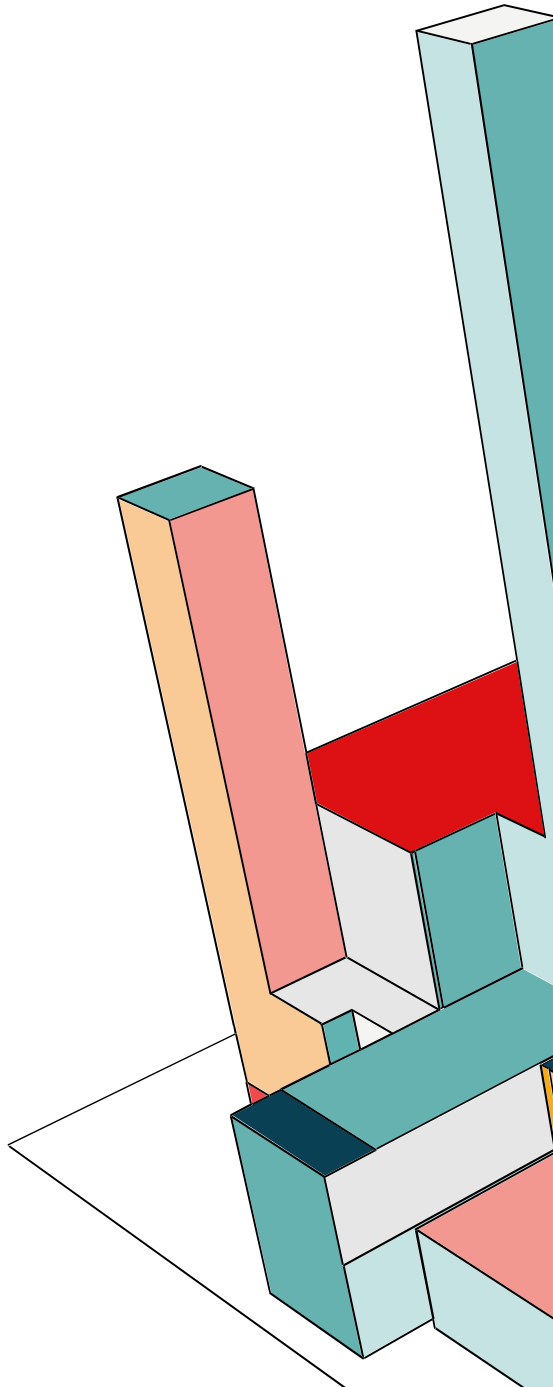
SECTION 1

METHODOLOGY

METHODOLOGY

Executive Summary

- Data collection methodology:
  - Data was collected and combined from SpaceX public API and Web-scraped from SpaceX Wikipedia page
- Perform data wrangling
  - A new column labelled "class" was added to classify successful and unsuccessful landings
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Models were built upon GridSearchCV after standardizing the data to locate the best parameters of each models
  - The models' Accuracy, Jaccard-score, and F1-score were measured to find the best models



DATA COLLECTION (OVERVIEW)

The current project has utilized the request of SpaceX API and Web-Scraping from SpaceX Falcon 9 Launch Wikipedia page as the methods of data collection.

The step-by-step method of each data collection process will be presented in a Flow Chart format in the following pages

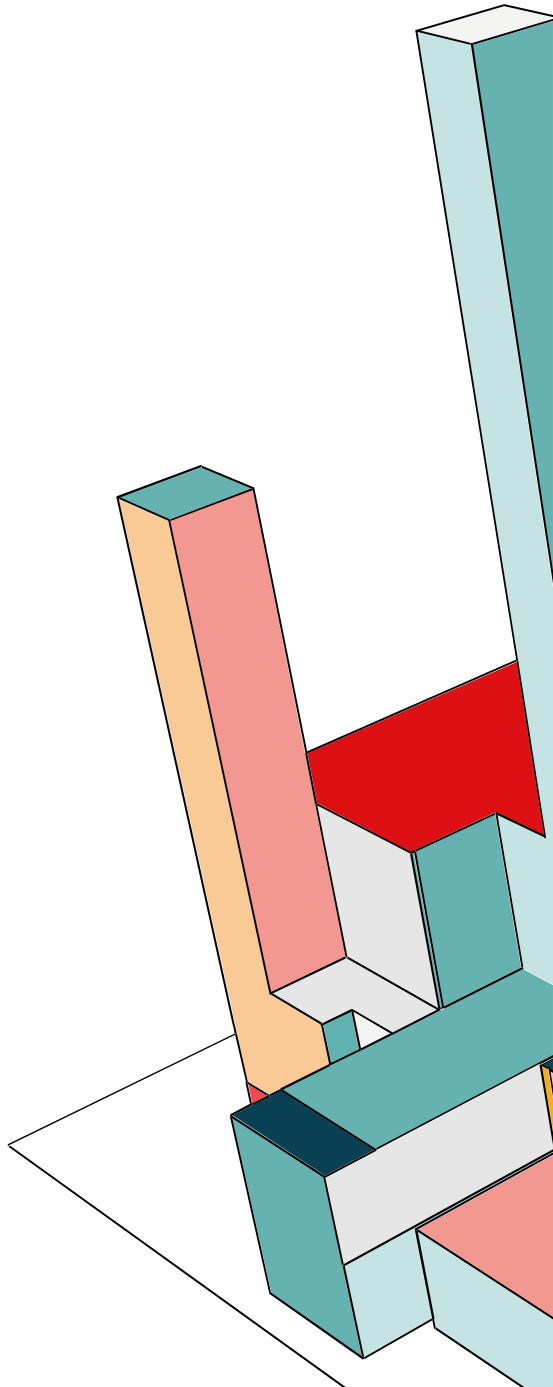
The following are the data columns extracted from the data collection processes:

SpaceX API request Data:

Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, FlightNumber

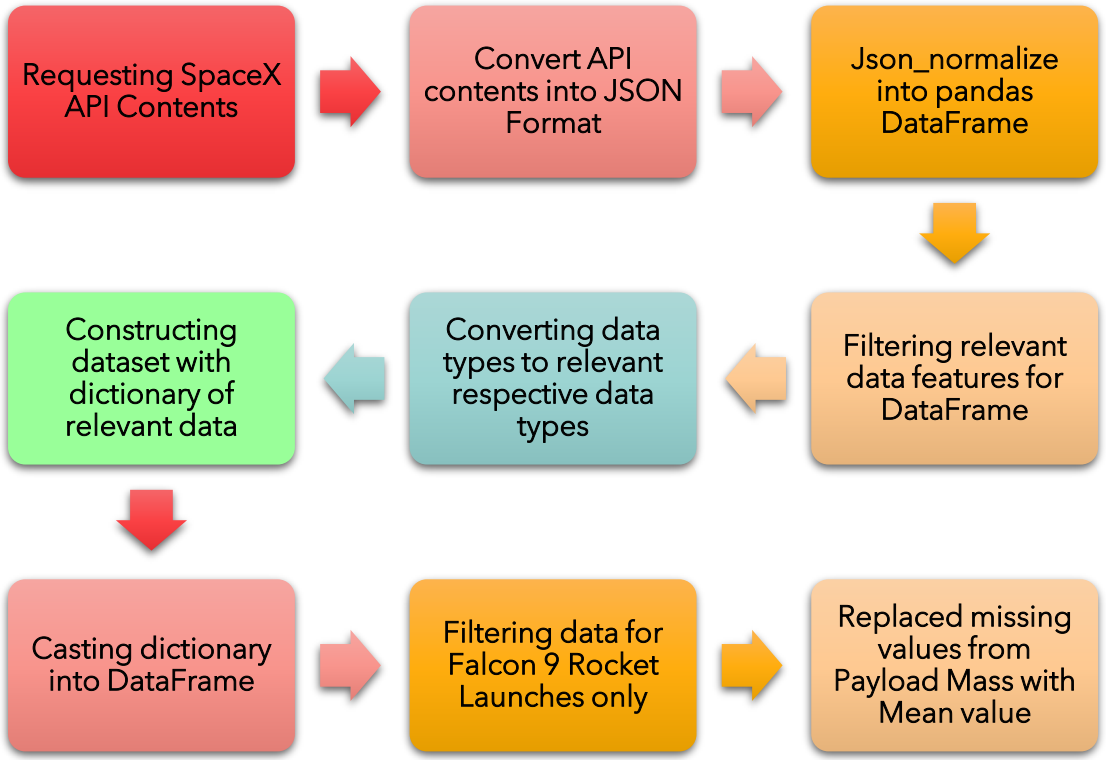
Web-Scraping from Falcon 9 Launch Wikipedia

Flight No., Date, Time, Version Booster, Launch site, Payload, Payload mass, Orbit, Customer, Launch outcome, Booster landing.



DATA COLLECTION – SPACEX API

Flow Chart of the Data Collection process via SpaceX public API

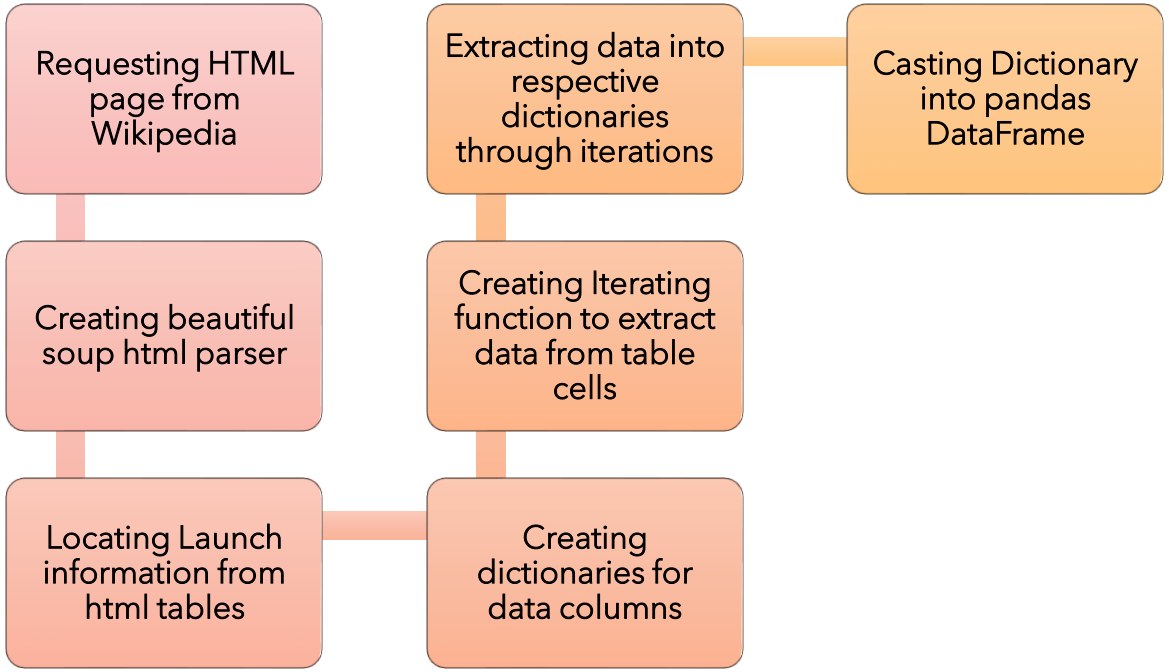


Github URL

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

DATA COLLECTION – WEB SCRAPING

Flow Chart of the Data Collection process via Web-Scraping of SpaceX Wikipedia page



Github URL

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Data%20Collection/WebScraping.ipynb

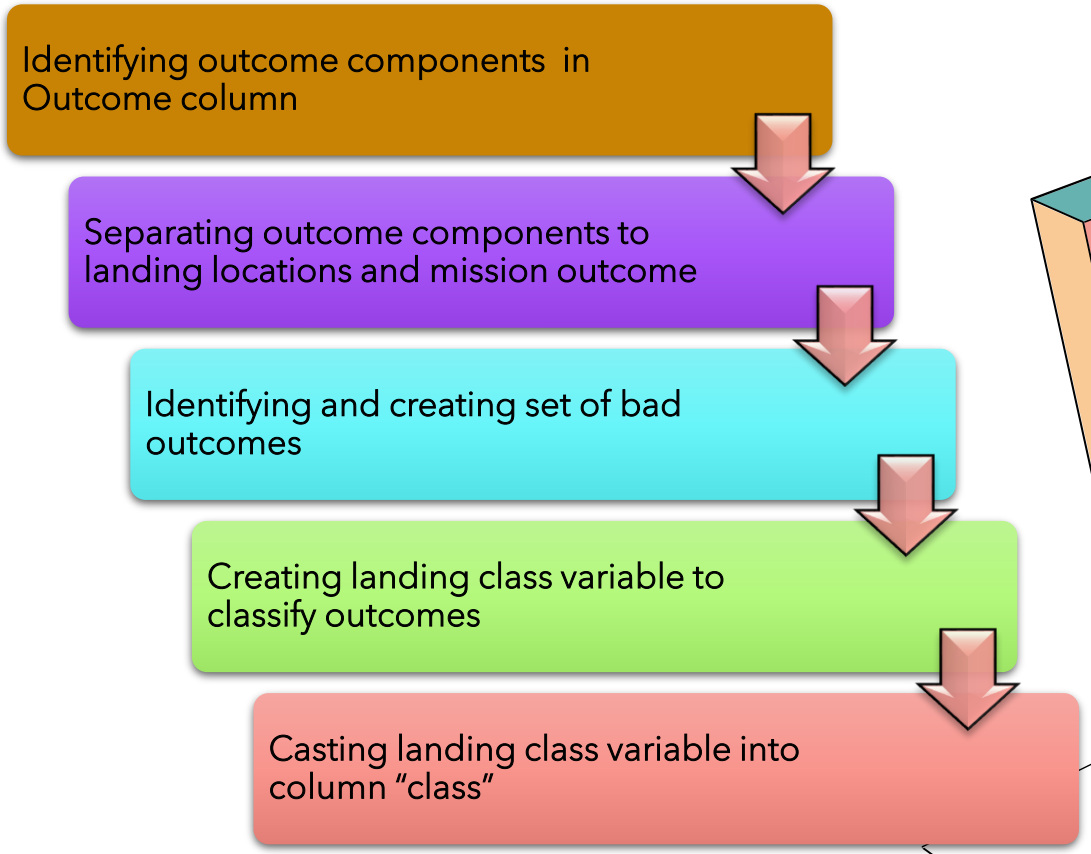
DATA WRANGLING

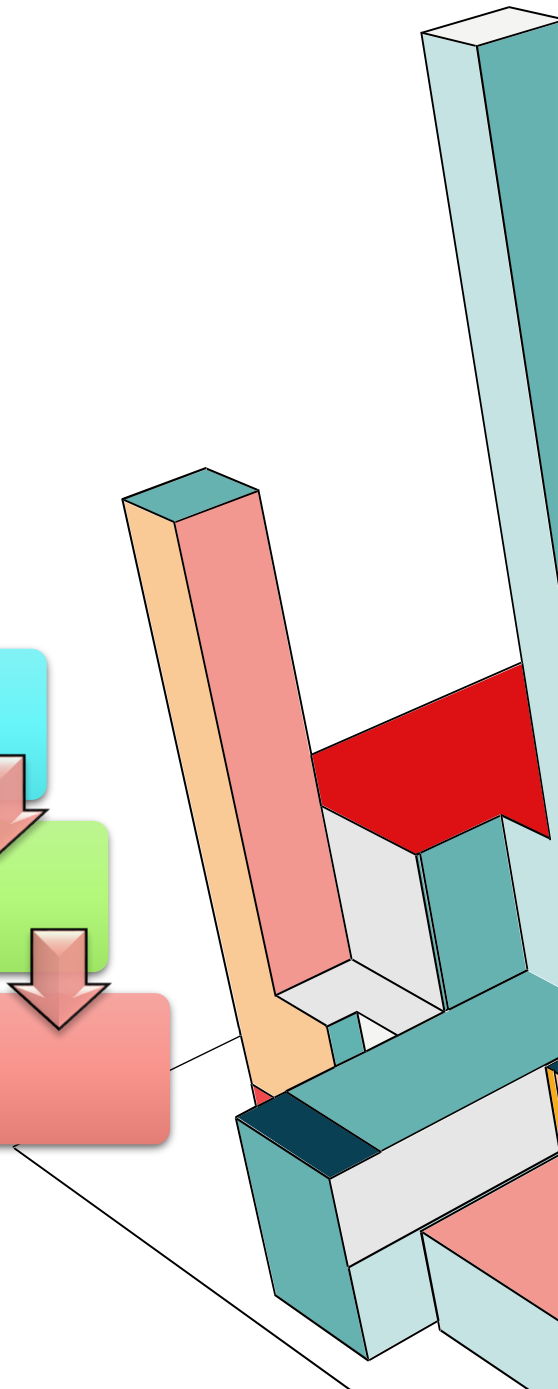
Flow Chart of the Data Wrangling process.

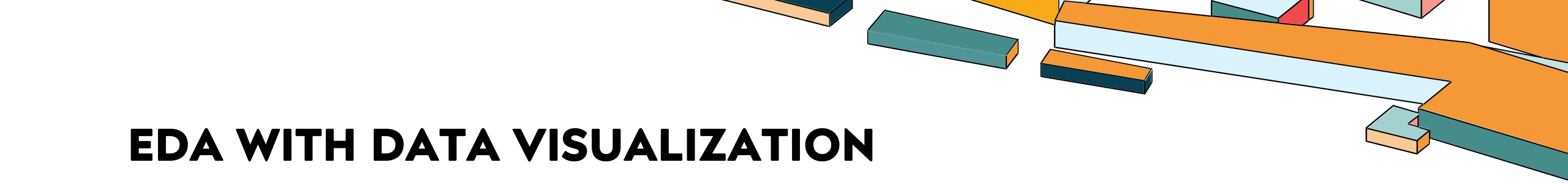
1. Outcome columns are separated into two columns
2. Column "class" are created for better classification of outcome
3. Value Mapping:
   - ❖ True ASDS, True RTLS, True Ocean = 1
   - ❖ None None, False ASDS, None ASDS, False ocean, False RTLS = 0

Github URL

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Data%20Collection/Data%20Wrangling.ipynb







EDA WITH DATA VISUALIZATION

Exploratory Data Analysis was performed on the Dataset, with variables such as Payload Mass (kg), Flight number, Launch site, Orbit, Date, and Class. The plots used in this project are:

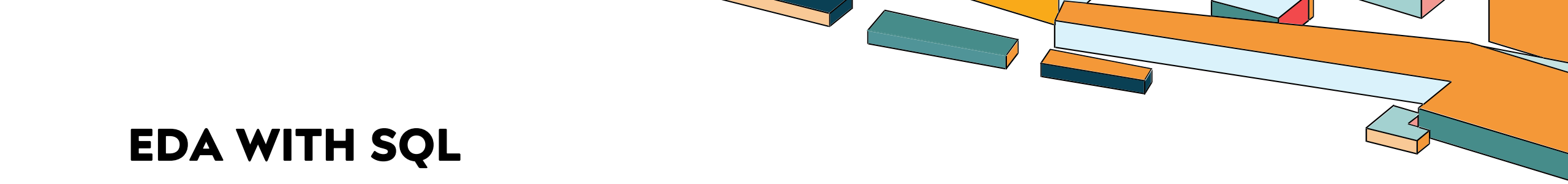
- 4x Scatter Plots
- 1x Bar Chart
- 1x Line Chart

To measure the relationship between the variables in question for further training in machine learning model such as:

1. Flight Number vs. Launch Site
2. Payload Mass vs. Launch site
3. Success rate of each Orbit Type
4. Flight Number vs. Orbit Type
5. Payload Mass vs. Orbit Type
6. Launch Success Yearly Trend

Github URL

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/EDA/EDA%20with%20Data%20Viz.ipynb



EDA WITH SQL

- Loaded SQL extension and established a connection with the Database
- Exploratory Data Analysis were performed with SQL queries with Python integration
- Dataset were filtered for better understanding with SQL queries, showing information about:
  - Total Payload Mass carried by Boosters launched by NASA (CRS)
  - Average Payload mass carried by booster version F9 v1.1
  - Date of first successful ground pad landing
  - Total number of successful and failure mission outcome
  - Booster Versions that have carried the maximum Payload Mass
  - And more

GitHub URL

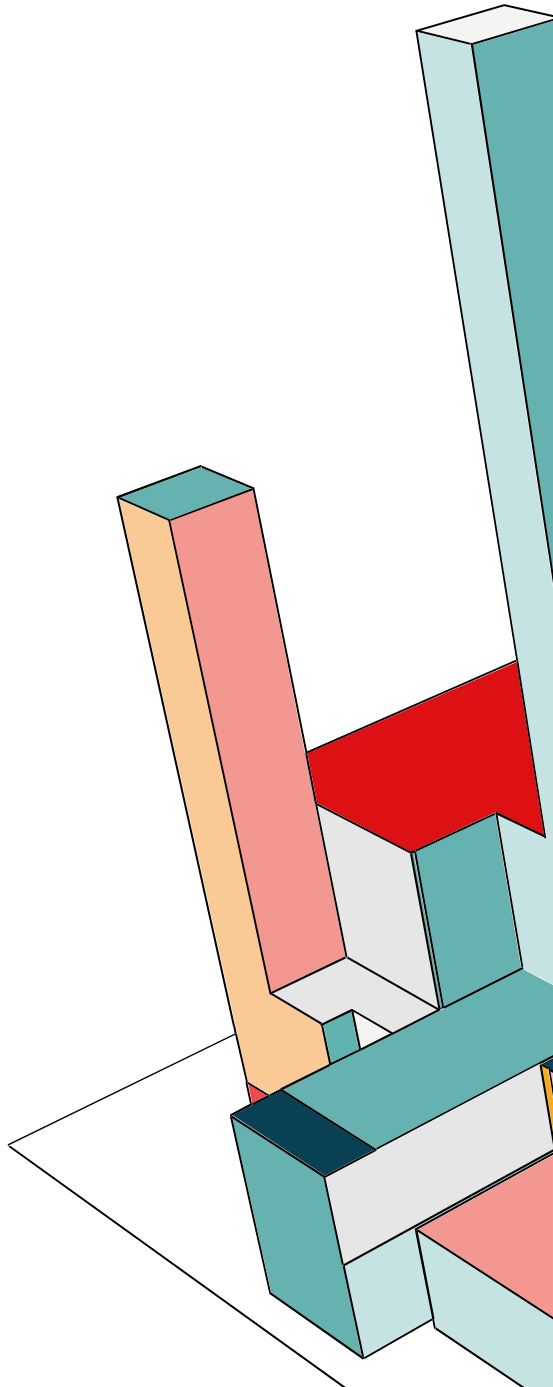
https://github.com/TylerKoN/Data-Science-Capstone/blob/main/EDA/EDA%20with%20SQL.ipynb

BUILDING INTERACTIVE MAP WITH FOLIUM

- ❑ Markers have been added into the Folium map indicating the location of the 4 launch sites, alongside with colored circles that indicate the launch outcomes.
- ❑ Lines were also included for estimating distance between Launch sites and key locations such as Railway, Highway, Cities, and Coastline.
- ❑ These specific elements were included as it will allow us to understand better on the specific landmarks and criteria of Launch Sites,
- ❑ Understanding the factors behind the best Rocket Launching locations, and potentially predict landings' success ratio from its location.

GitHub URL

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Visual%20Analytics%20and%20Dashboard/Launch%20site%20location%20Folium.ipynb

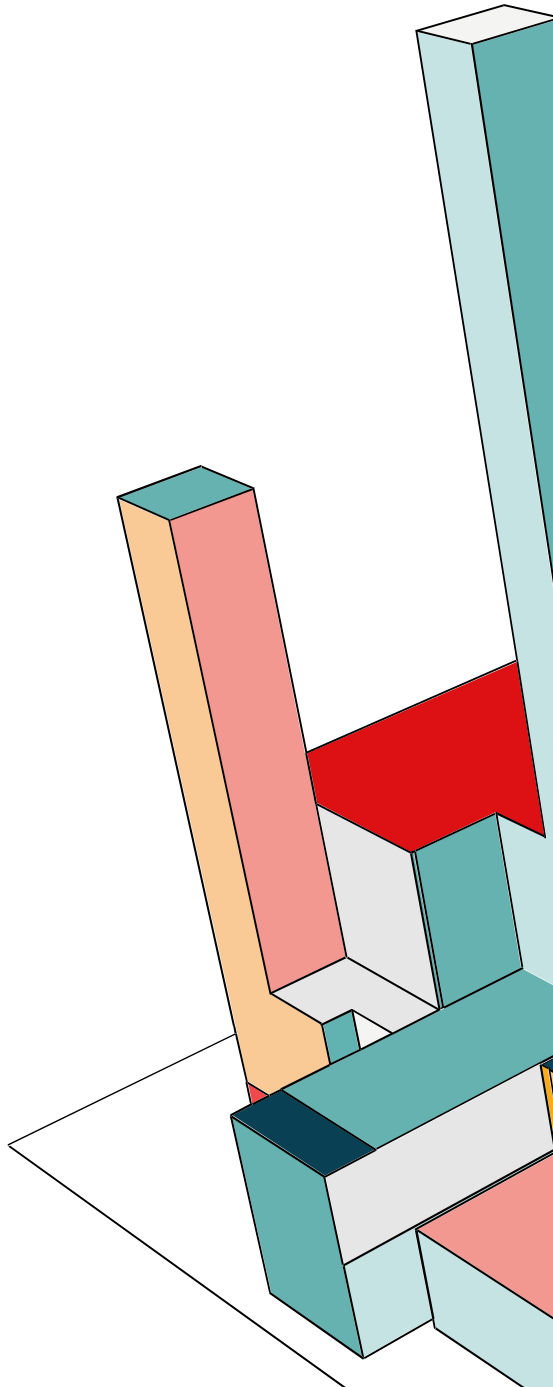


BUILDING A DASHBOARD WITH PLOTLY DASH

- The Plotly Dash Dashboard were created with two charts for visualization and analysis, being a Scatterplot and a Pie Chart.
- The Pie Chart from the Dashboard would display the distribution of successful landing across all launch sites and were given the ability to display the success ratio of each individual launch sites.
- The Scatterplot from the Dashboard would display the success ratio based on two inputs:
  - Payload Mass with a Range Slider from 0kg to 10000kg
  - Each individual Launch Sites or All launch sites
- Both Charts in this case helps visualizing the success ratio of each Launch Sites, with the Scatterplot being able to display success ratio across various variables such as Payload Mass, Booster Version Categories, and Launch sites.

GitHub URL

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Visual%20Analytics%20and%20Dashboard/spacex_dash_app.py

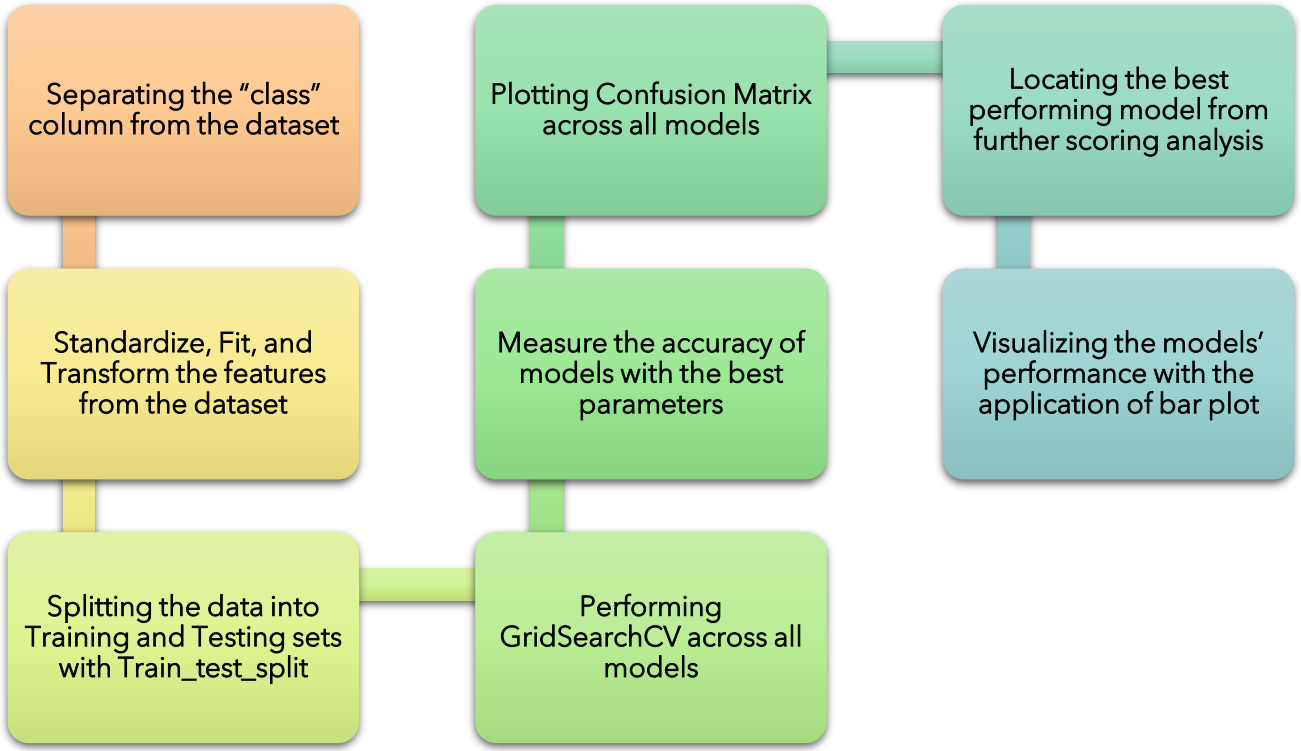


PREDICTIVE ANALYSIS (CLASSIFICATION)

Flow Chart of the Classification

Predictive Analysis Process:

1. All models performed similarly with similar score on each scoring categories

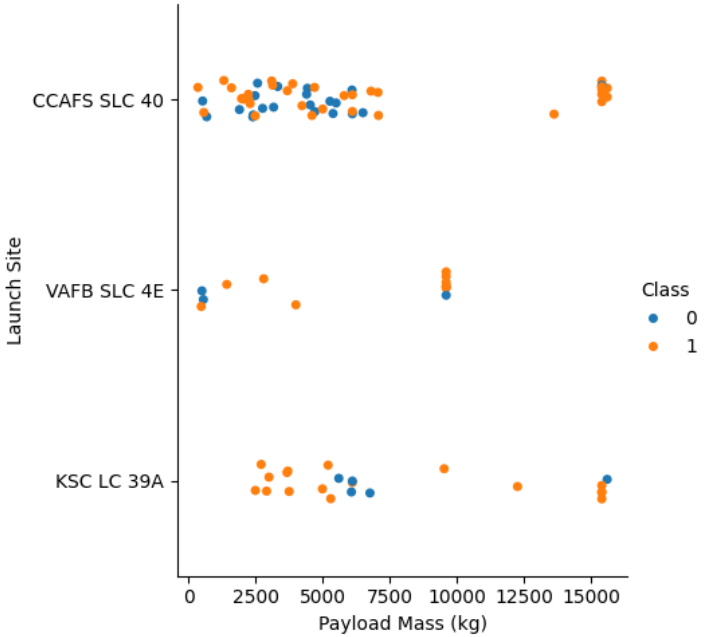


GitHub URL

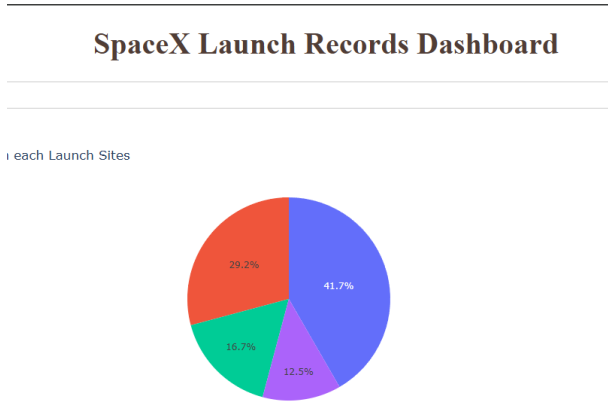
https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Predictive%20Analysis/Classification%20Predictive%20Analysis.ipynb

RESULTS

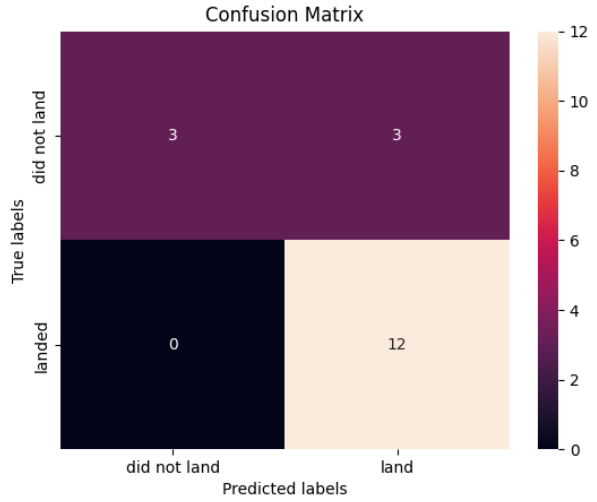
Below is the examples of results displayed from each section.



Exploratory data analysis



SpaceX Launch Records Interactive Dashboard

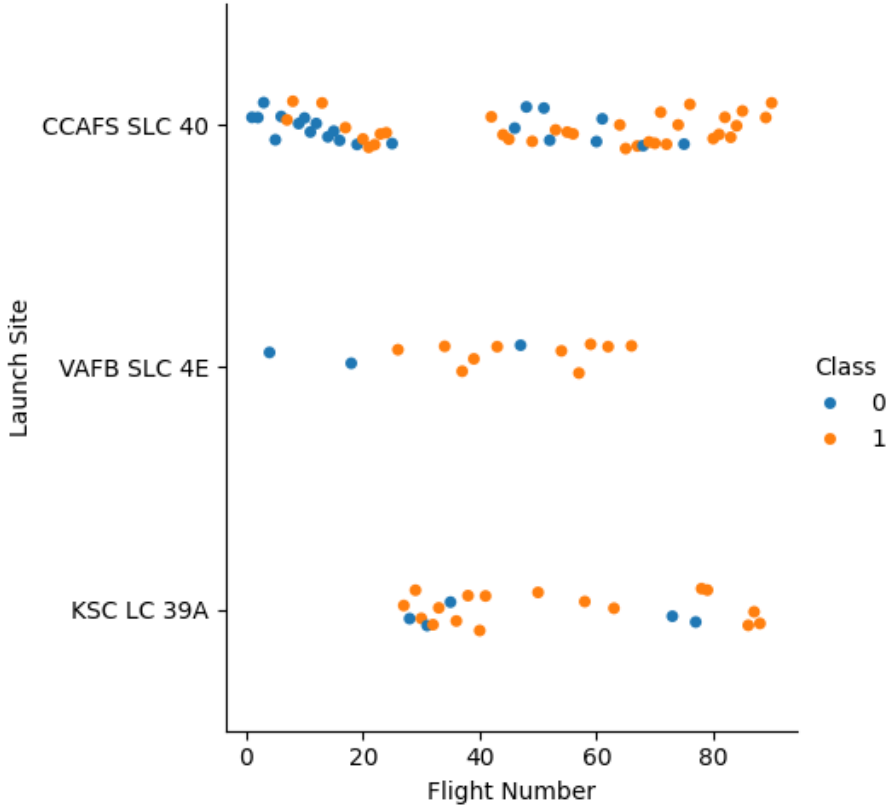


Predictive Analysis Confusion Matrix

SECTION 2

INSIGHTS DRAWN FROM EDA

FLIGHT NUMBER VS. LAUNCH SITES

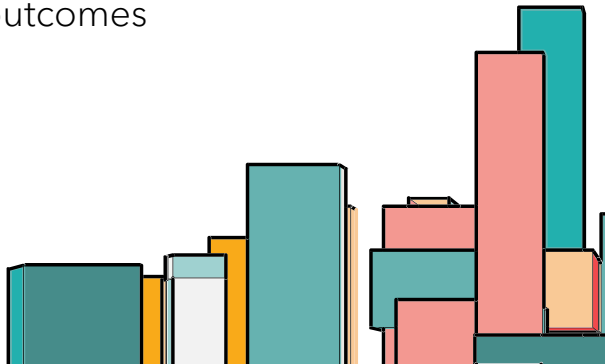


Class 0 (blue) suggests Unsuccessful landings

Class 1 (orange) suggests Successful landings

Explanation

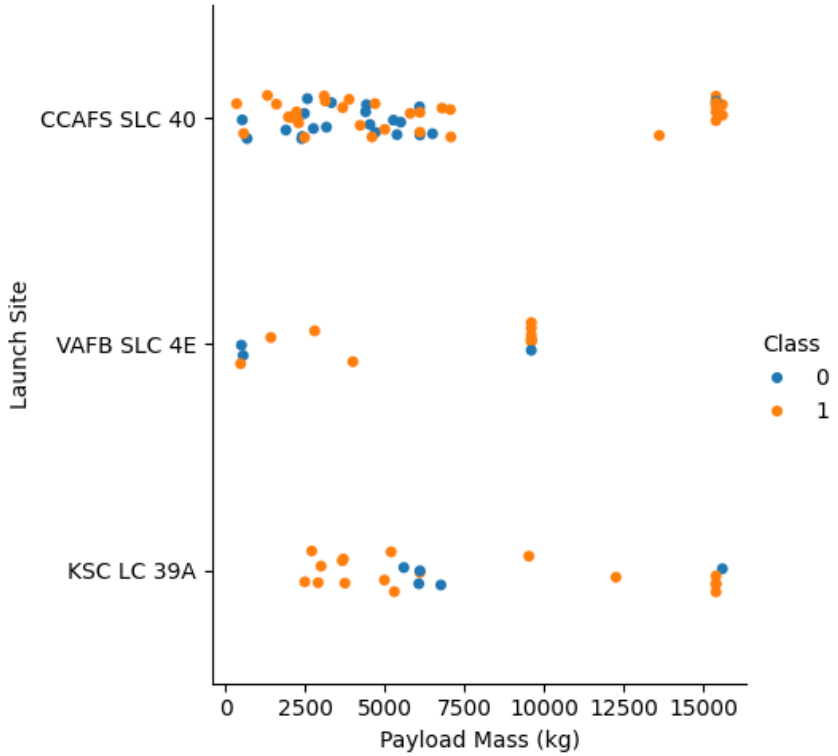
- CCAFS SLC 40 Launch site has the most landings across the other launch sites
- It is shown that CCAFS SLC 40 launch site has more failure in landing outcomes at an early stage of flight number, with its success rate increasing after 80 flights
- The rest of the launch sites has shown more successful landing outcomes than unsuccessful outcomes



PAYLOAD VS. LAUNCH SITE

Class 0 (blue) suggests Unsuccessful landings

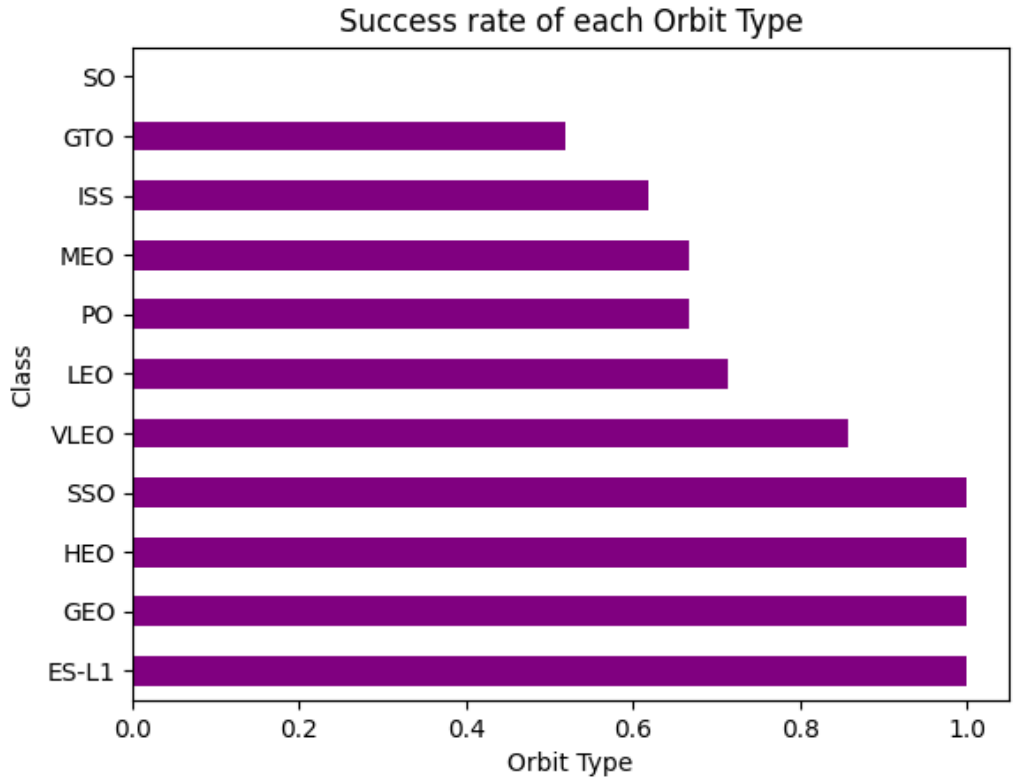
Class 1 (orange) suggests Successful landings



Explanation

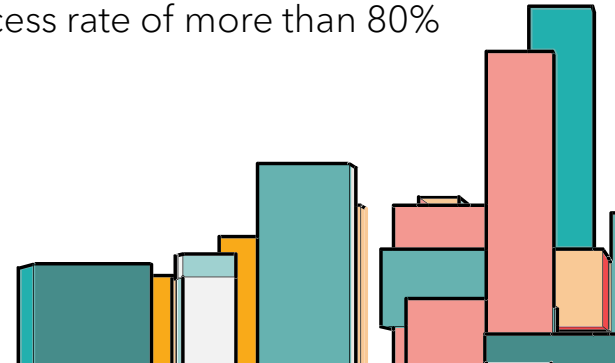
- KSC LC 39A Launch Site appears to land successfully with payload mass are within 2500 to 5000kg
- Both KSC LC 39A and CCAFS SLC 40 Launch Sites have a higher success rate when it comes with higher Payload Mass at 15000kg+
- It seems that VAFB SLC 4E Launch site has a higher landing success ratio with payload mass of below 10000kg
- Yet, VAFB SLC 4E Launch site were the launch site that has the least amount of landing as compared to other Launch sites

SUCCESS RATE VS. ORBIT TYPE



Explanation

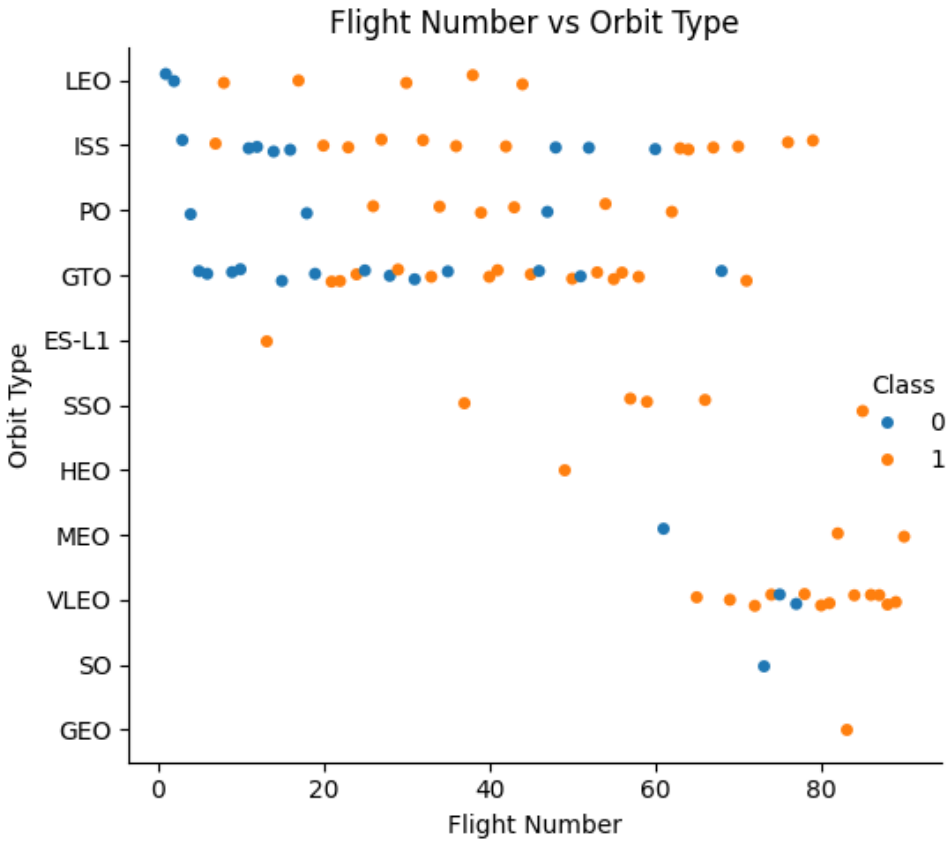
- The bar chart displayed has shown a perfect success rate of landing outcomes on the orbits (ES-L1, GEO, HEO, and SSO)
- Besides, almost all but except SO orbit has equal and above 50% rate of successful landing
- SO are the only orbit type that has displayed 0% rate of success in landing outcome
- GTO, being the orbit type with the most sample (27), has the success rate of 50%
- VLEO with 14 samples, second only to GTO being the most sample, have displayed a great success rate of more than 80%



FLIGHT NUMBER VS. ORBIT TYPE

Class 0 (blue) suggests Unsuccessful landings

Class 1 (orange) suggests Successful landings



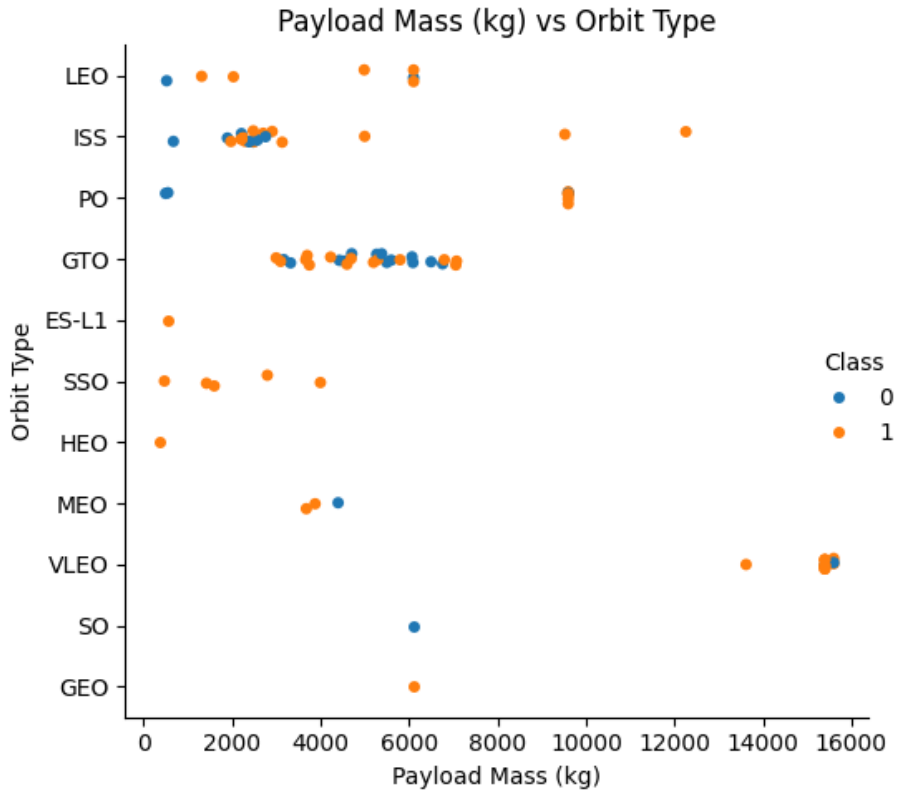
Explanation

- According to the scatter plot, it shows that when flight number increases, most successful landings appears to be within the VLEO orbit
- There is also a pattern of unsuccessful landings around the orbit type of LEO, ISS, PO, and GTO during the early stages of Falcon 9 Launches
- ES-L1, SSO, and HEO orbits have also showcased a 100% success ratio of landings despite having a relatively smaller sample size.
- GTO orbit was also shown to be the orbit type with the most unsuccessful landings (50%)

PAYLOAD VS. ORBIT TYPE

Class 0 (blue) suggests Unsuccessful landings

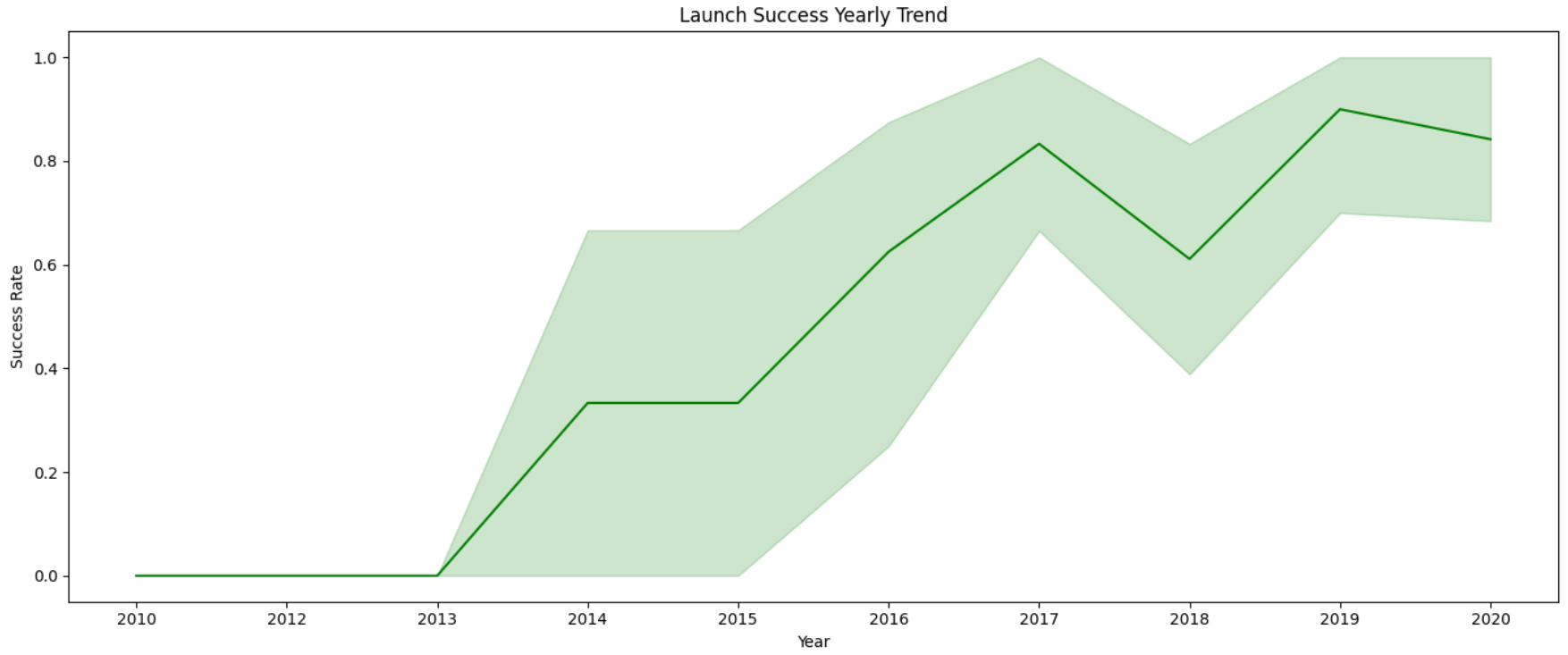
Class 1 (orange) suggests Successful landings



Explanation

- The scatterplot has shown the VLEO orbit were often utilized for launches with larger payload mass
- ES-L1, SSO, and HEO orbits were shown to have more successful landings with payload mass of 0kg to 6000kg
- It is also shown that LEO, ISS, and PO orbit were able to land successfully when payload mass are between 5000kg to 10000kg
- The GTO orbit were generally utilized for payload mass of 2000 to 8000kg, with only a 50% rate of successful landing

LAUNCH SUCCESS YEARLY TREND



Based on the line graph, the launch success ratio started to increase from the year 2013 onwards and has gradually improved until 2017, where the performance dropped slightly in 2018 before increasing in 2019 again.

The year 2019 can be seen as the year with the highest launch success ratio, with over 80% of launch success ratio.

ALL LAUNCH SITES NAMES

- The query below displayed all the unique values within the “Launch_site” column
- 4 names of launch sites can be identified from the column, being:
  - CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40

Display the names of the unique launch sites in the space mission

```
[11]: %sql SELECT DISTINCT "Launch_site" FROM SPACEXTBL
 * sqlite:///my_data1.db
Done.
```

[11]:

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

LAUNCH SITE NAMES BEGIN WITH ‘CCA’

- The query below displayed the top 5 records of launch sites name beginning with “CCA”

Display 5 records where launch sites begin with the string 'CCA'

```
[15]: %sql SELECT * FROM SPACEXTBL WHERE "Launch_site" LIKE 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

[15]:

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

TOTAL PAYLOAD MASS

- The query below has displayed the sum for total payload mass that was carried by boosters launched by NASA (CRS)
- The total amount of Payload Mass that is carried by NASA (CRS)'s Boosters towards the International Space Station (ISS) was shown to be 45596kg

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

```
%sql select Customer, sum(PAYLOAD_MASS__KG_) as "Total_Payload_Mass" from SPACEXTBL where Customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

Customer	Total_Payload_Mass
NASA (CRS)	45596

AVERAGE PAYLOAD MASS BY F9 V1.1

- The query below has displayed the average payload mass that was carried by booster version F9 v1.1 with the usage of average function.
- The average amount of Payload Mass that is carried by Booster version F9 v1.1 were shown to be 2928.4kg, which was relatively low based on the average value of payload mass from the entire dataset (6138.29kg)

Display average payload mass carried by booster version F9 v1.1

```
%sql select Booster_Version, AVG(PAYLOAD_MASS__KG_) as "Average_Payload_Mass" from SPACEXTBL where "Booster_Version" = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Average_Payload_Mass
F9 v1.1	2928.4

FIRST SUCCESSFUL GROUND LANDING DATE

- The query below has displayed the date of the first successful landing outcome on a ground pad with the utilization of the “min” function
- The first successful ground pad landing date were shown to be on 22nd of December in the year 2015.

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

Hint:Use min function

```
%sql SELECT min(Date) as "First_GroundPad_Landing", Landing_Outcome FROM SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

First_GroundPad_Landing	Landing_Outcome
2015-12-22	Success (ground pad)

BOOSTER VERSIONS OF SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

- The query below has displayed the list of successful drone ship landing booster versions with the payload mass of between 4000kg to 6000kg
- It is shown that the booster versions of successful drone ship landing were F9 FT B1022, B1026, B1021.2, and B1031.2, all of which have carried payload that weighted between 4000kg and 6000kg

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

```
%sql select Booster_Version, PAYLOAD_MASS__KG_, Landing_Outcome from SPACEXTBL where "Landing_Outcome" = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

- The query below has displayed total number of successful missions and failure mission outcomes with the “sum” function that locates any values that associated with “success” and “failure”
- It is shown that SpaceX has only had one mission failure outcome, with the indicator of the mission failure being “Failure in flight”, showcasing a 99% rate of mission success ratio.

List the total number of successful and failure mission outcomes

```
%sql select sum(Mission_Outcome LIKE '%Success%') as "Total_Successful_Missions" , sum(Mission_Outcome LIKE '%Failure%') as "Total_Failure_Missions" from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

Total_Successful_Missions	Total_Failure_Missions
100	1

BOOSTERS CARRIED MAXIMUM PAYLOAD

- The query below has displayed the list of Boosters Version that has carried the maximum payload mass of 15600kg

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

```
%sql select DISTINCT Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 LAUNCH RECORDS

- The query below has displayed launch records in 2015 where the landing outcomes were failure in drone ship, along with its Launch site, booster version, and date.
- It is shown only two records of such outcome were recorded in 2015, in which both has occurred in the same Launch Site during January and April of 2015.

```
%sql select substr(Date,0,5) as Year, substr(Date, 6, 2) as month, Booster_Version, Launch_site, Landing_Outcome from SPACEXTBL WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)'
 * sqlite:///my_data1.db
Done.
```

Year	month	Booster_Version	Launch_Site	Landing_Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

- The query below has displayed the count of landing outcomes ranked from most to least during 4th of June 2010 to 20th of March 2017.
- The table below shows that most landing outcomes were counted as No attempt, in which showcased that most landing of launches were not even attempted (10) by SpaceX

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.

```
%sql select Landing_Outcome, COUNT(Landing_Outcome) as "Count_of_Outcomes" from SPACEXTBL where Date Between '2010-06-04' and '2017-03-20' GROUP BY Landing_Outcome ORDER BY "Count_of_Outcomes" DESC
```

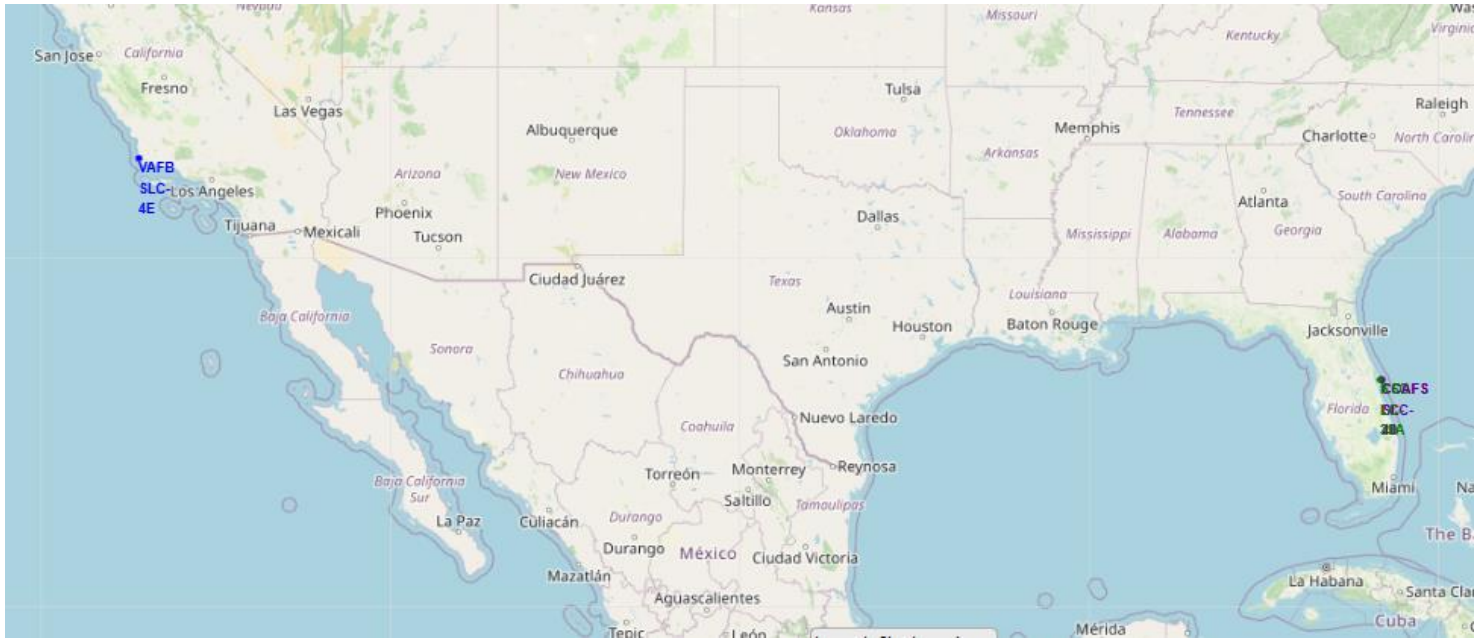
```
 * sqlite:///my_data1.db
Done.
```

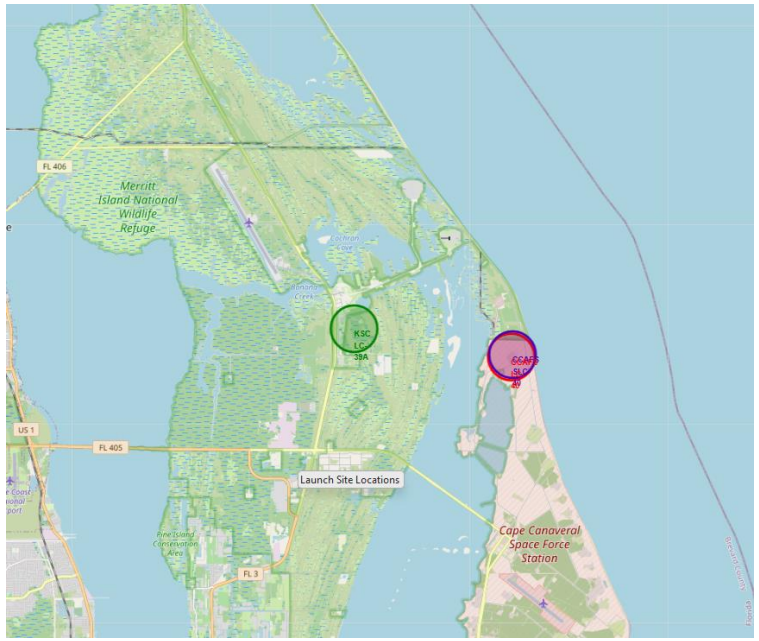
Landing_Outcome	Count_of_Outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

SECTION 3

LAUNCH SITES PROXIMITIES ANALYSIS

LOCATION OF LAUNCH SITES



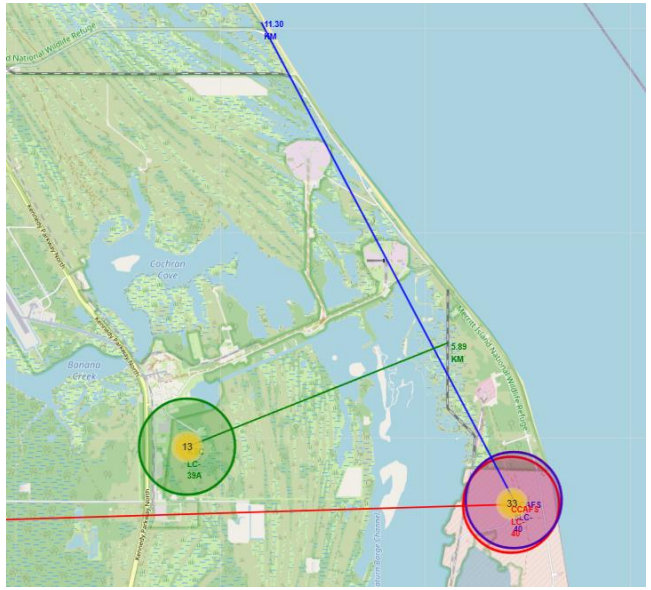


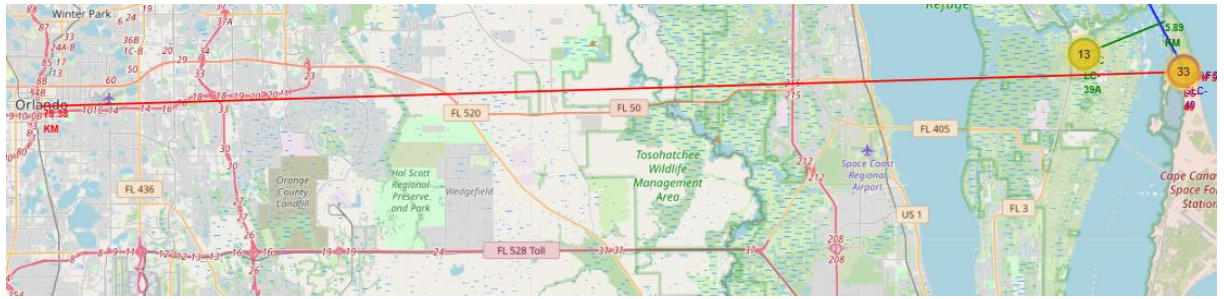
- The folium map above showcases the locations of all Launch sites across the US, where only VAFB SLC-4E can be noticed as the only Launch Site that is on the west side of the country, while the other launch sites were relatively close to each other on the east side of the country.
- All launch sites can be seen as closer to the coastline and isolated from large cities

LAUNCH MARKERS WITH COLOR INDICATORS

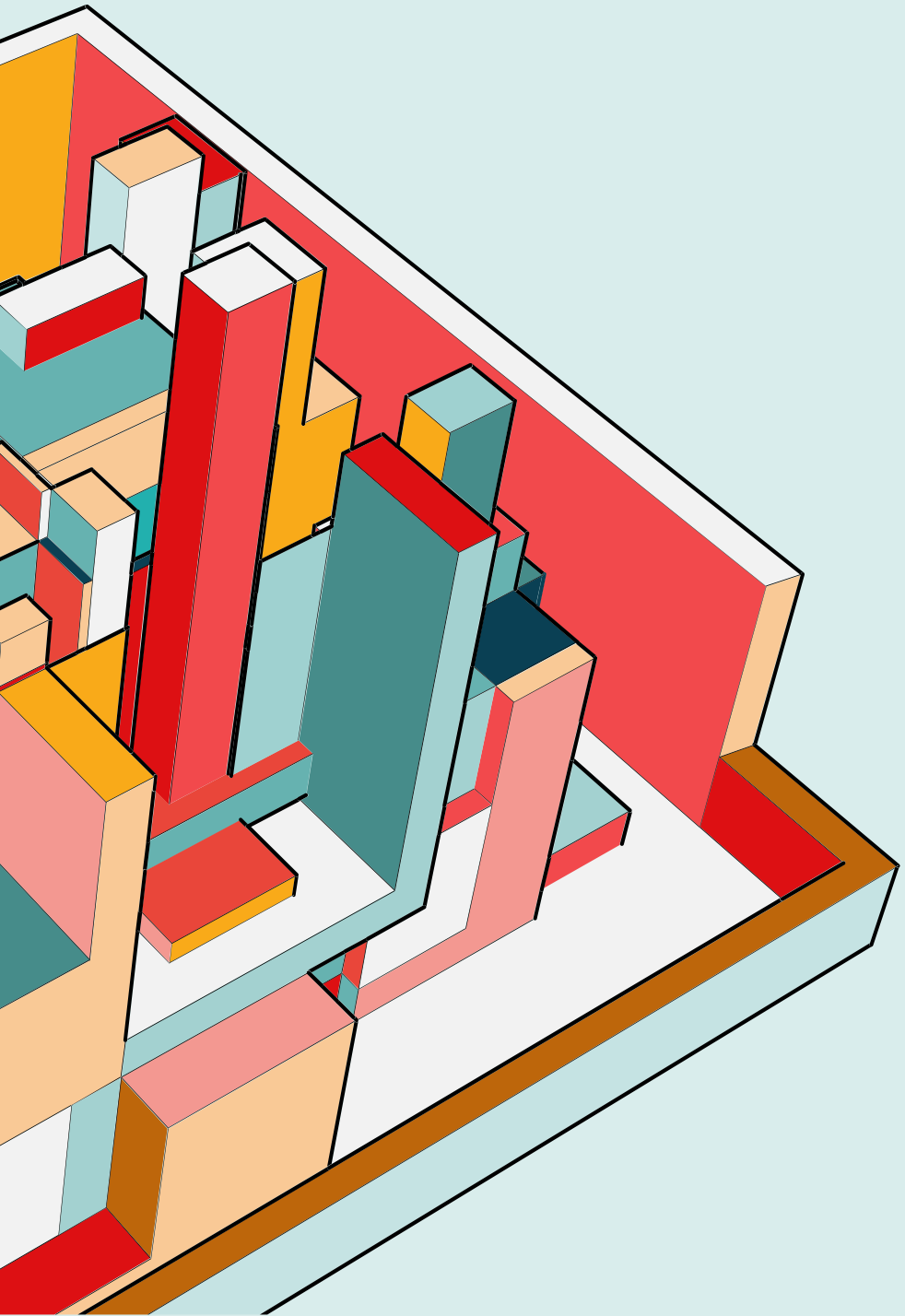
- The folium map above showcases the markers for the locations of all successful and unsuccessful launches from each launch sites with colors indicating its outcome.
- We can see the launch site VAFB SLC-4E only had 10 launches, with most landing outcomes being unsuccessful as shown from the image above.

DISTANCE BETWEEN LAUNCH SITES AND KEY LOCATIONS





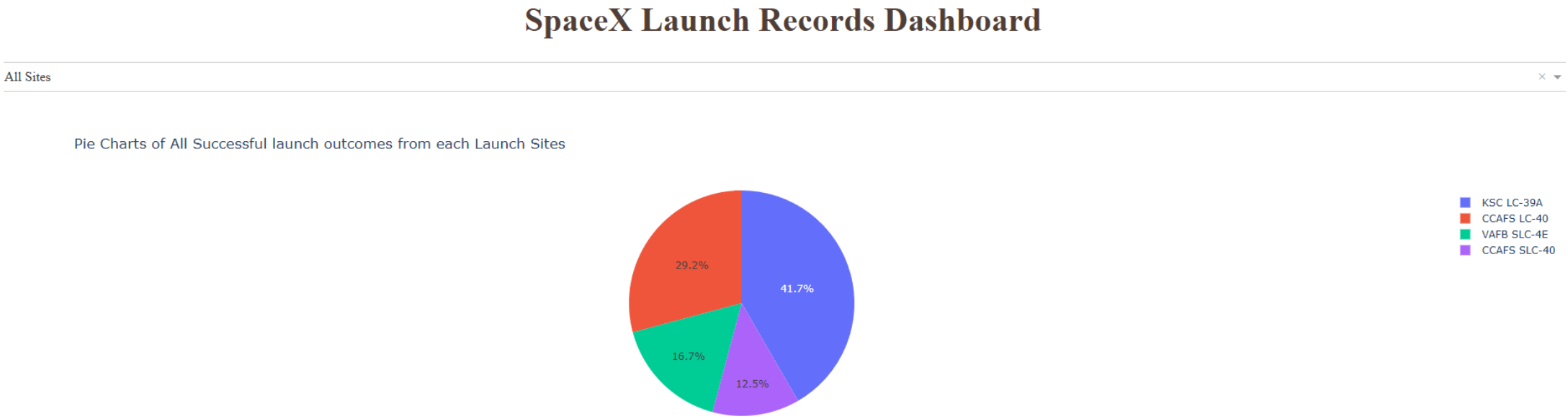
- The folium map above showcases the distance between launch sites and some key locations as shown from a line with a remark that indicates the total distance between the launch sites and key locations.
- It is noticeable that the Launch Sites are all in close proximity to Railways, highways, and coastlines (Except KSC LC-39A). Possibly to ease the transportation for each launch missions especially when there is payload.
- All Launch Sites can be seen as keeping quite a distance away from cities.



SECTION 4

BUILDING A DASHBOARD WITH PLOTLY DASH

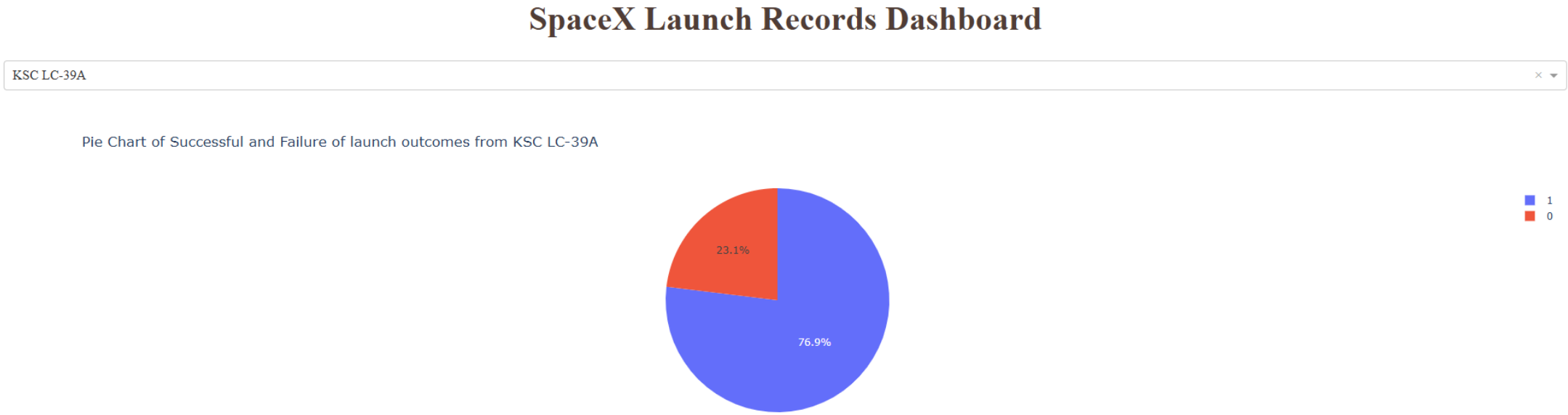
SUCCESS RATIO OF EACH LAUNCH SITES



The Pie chart above showcases the landing success outcome ratio for each launch sites, with KSC LC-39A being the highest rated successful landing outcome launch site with 41.7%.

However, it should be noted that CCAFS LC-40 and CCAFS SLC-40 could potentially be a similar launch site that should be compared along each other due to their close proximity with one another, making their total success ratio being 41.7%, same as KSC LC-39A

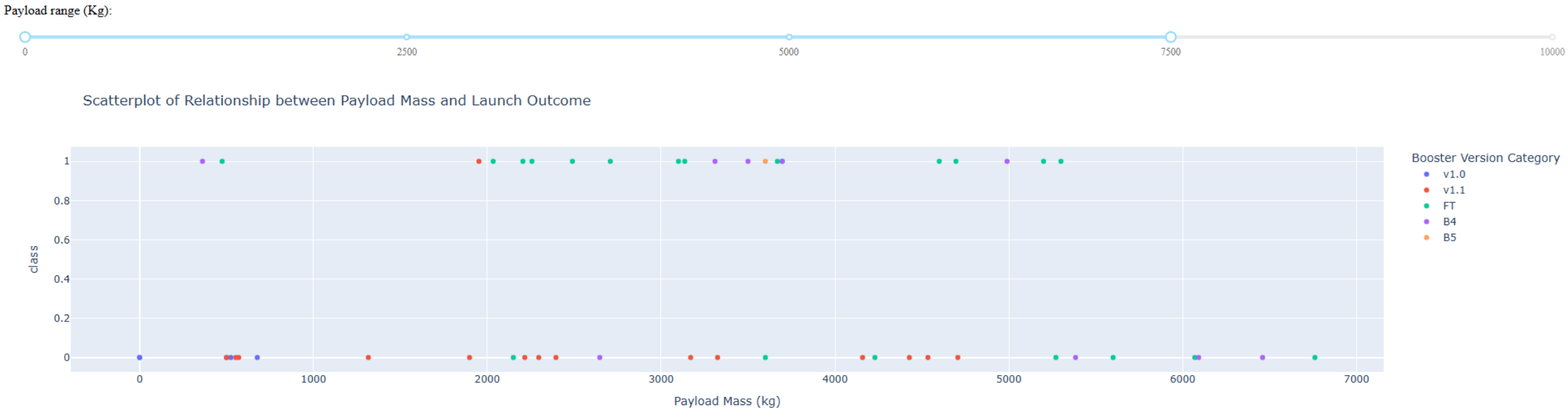
LAUNCH SITE WITH HIGHEST SUCCESS RATIO



The Pie chart above showcases the landing success outcome ratio for KSC LC-39A launch site, with 76.9% of its landing outcomes being successful, making it the launch sites with the highest success ratio in launch outcomes.

KSC LC-39A achieved the highest successful landing outcome with a total of 13 launches, in which 10 of the launches have landed successfully, and only 3 of the launches have failed its landing.

RELATIONSHIP BETWEEN PAYLOAD MASS AND LAUNCH OUTCOMES



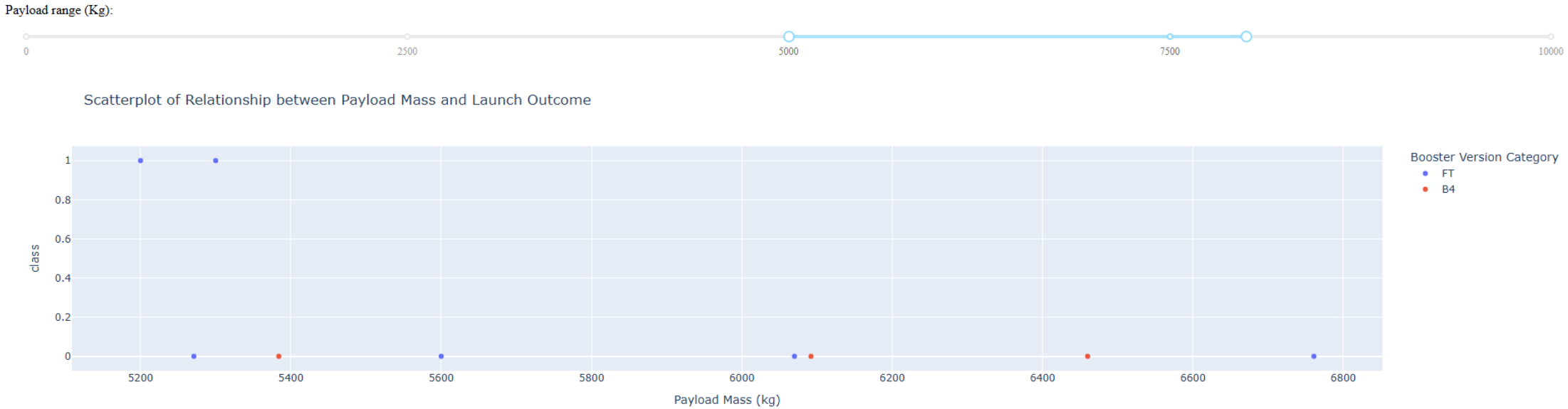
The scatterplot above showcases the relationship between payload mass and landing outcomes from all launch sites along with their booster version categories.

The payload range within the Dashboard can be adjusted with the range slider for better interpretation of the variables' relationship.

It can be seen from the plot above that most successful launches are associated with the Booster Version Category of "FT", and most failure are associated with "v1.1".

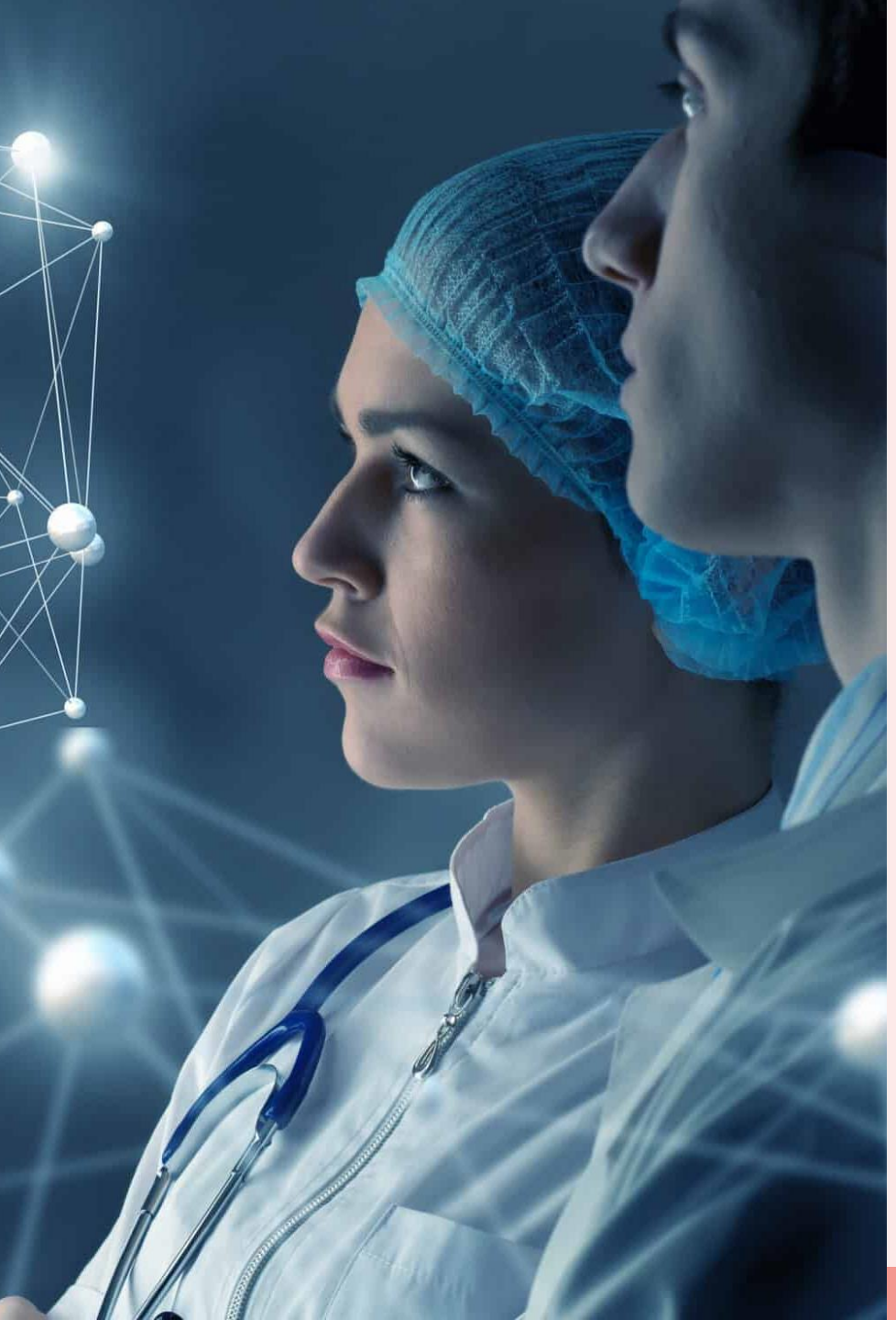
The Range between 2000kg to 6000kg can also be seen as the range with the highest success ratio according from the scatterplot

LOWEST SUCCESS RATIO PAYLOAD RANGE



The scatterplot above showcases the relationship between payload mass and landing outcome from each launch within the payload mass range of 5000kg to 8000kg.

It can be seen from the plot above that most launches fail when it comes to the payload mass range between 5000kg to 8000kg, making this range as the lowest success ratio for launch outcomes.



SECTION 5

PREDICTIVE ANALYSIS (CLASSIFICATION)

CLASSIFICATION ACCURACY

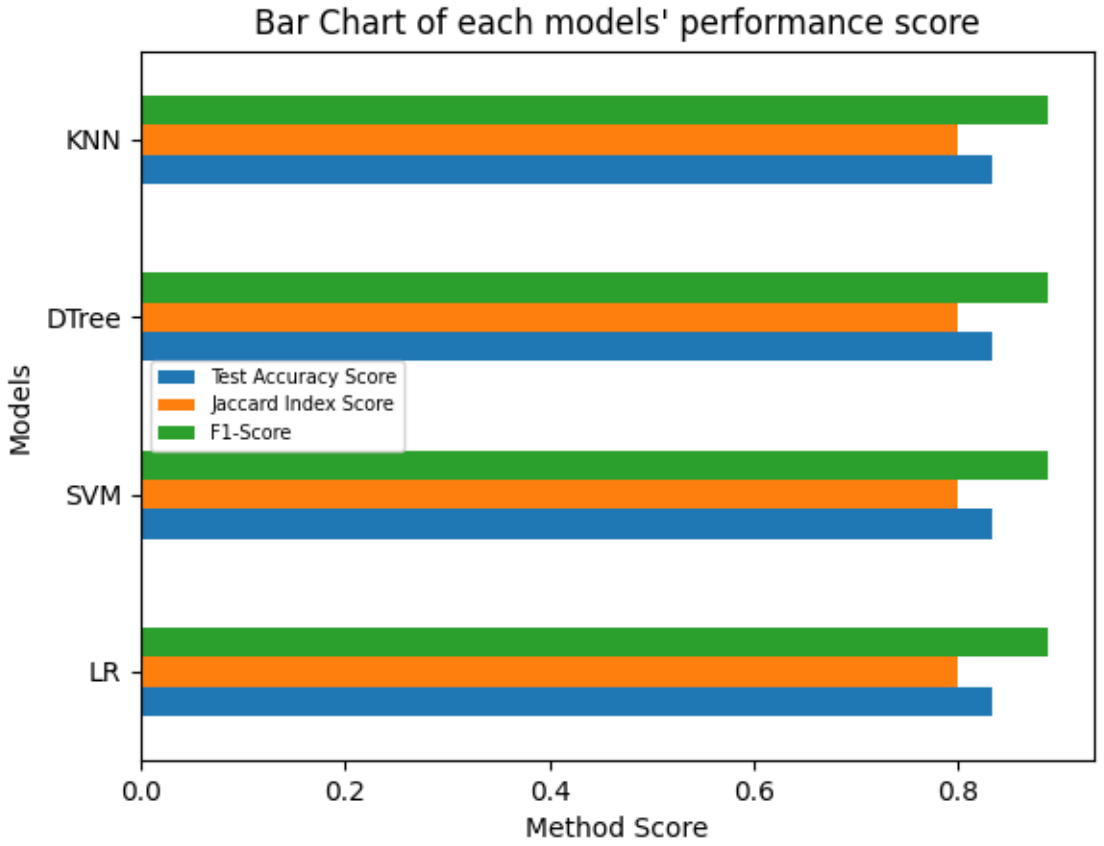
Based on the horizontal bar chart as shown,

- All models displayed similar performance
- With accuracy being 83.33% across all models.

However, as the sample size of the testing dataset was only 18, the result should be interpreted with cautions due to its small test sample size.

Such an issue may cause larger variances within the accuracy results, causing inaccurate value to be produced from overfitting or underfitting models

It is suggested that more testing sample should be collected for further analysis on model testing



CONFUSION MATRIX

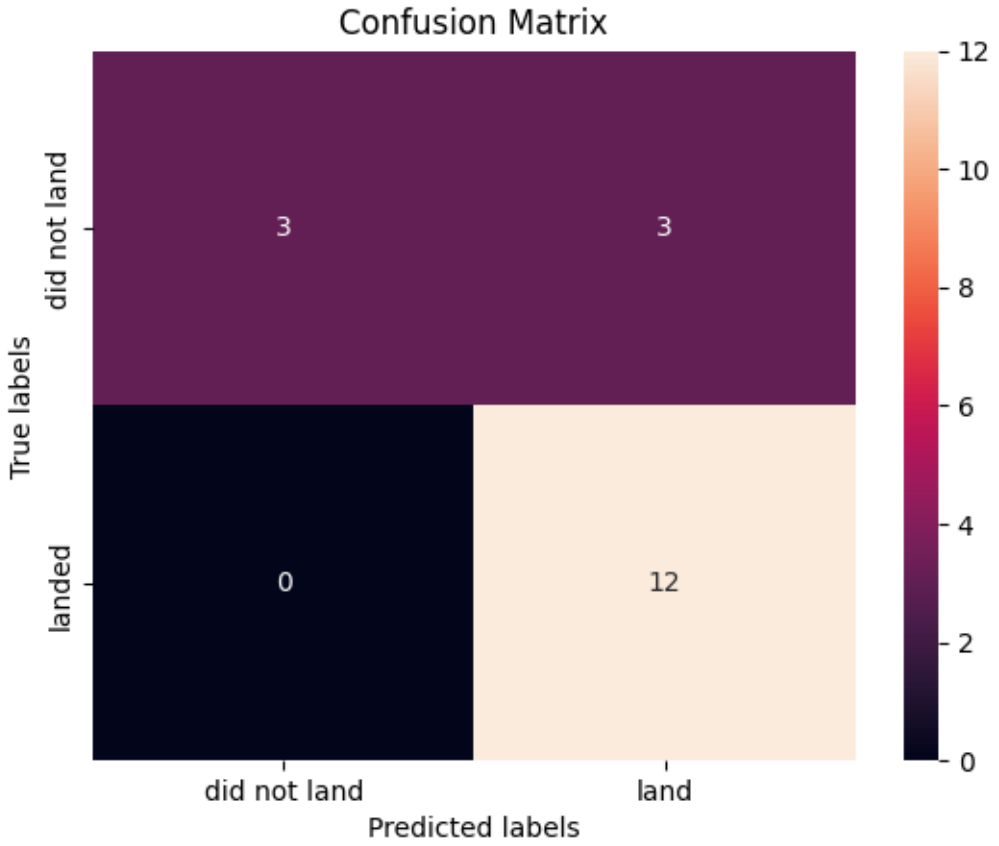
Since all models have the exact same performance based on the confusion matrix, all models are interpreted similarly.

- All models were able to successfully predict 12 successful landings
- All models were able to successfully predict 3 unsuccessful landings
- All models had 3 false positive predictions where the models predicted 3 successful landings with their true label being unsuccessful landings

This shows that all models had a high accuracy in terms of predicting the success ratio of landing outcomes.

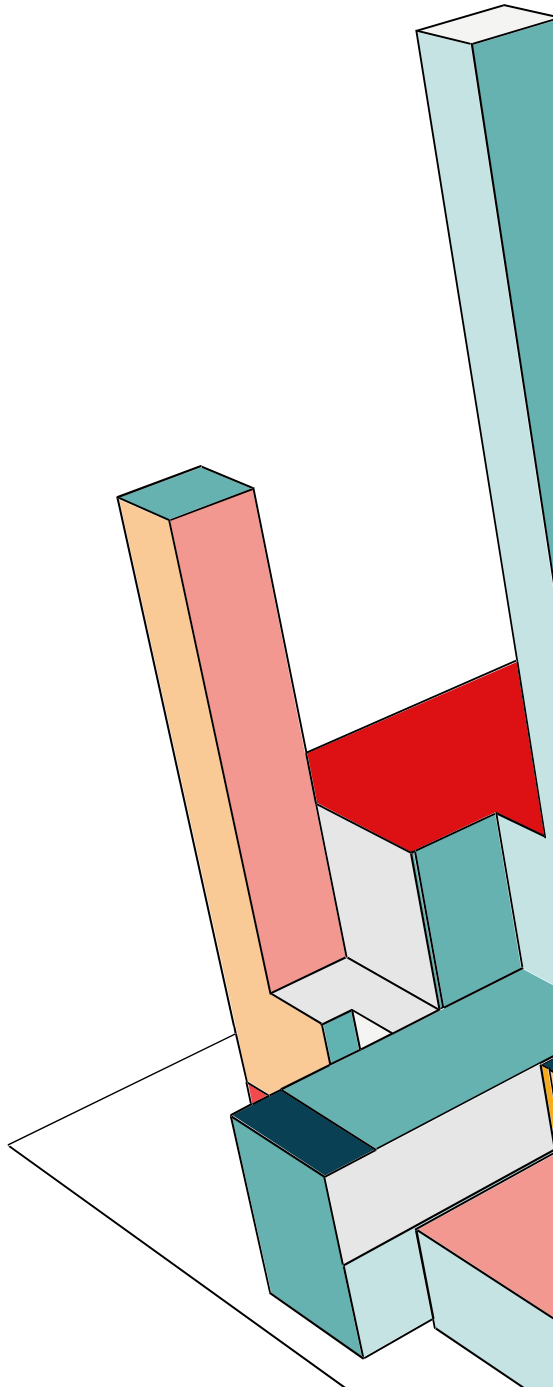
Though the previous cautions mentioned should also be taken into consideration for better interpretation of the models.

Correct predictions are on a diagonal from top left to bottom right.



CONCLUSION

- The initial goal of this current project was to develop machine learning models to predict the landing outcomes of the First Stage for Space Y
- Data was collected from SpaceX API and Web-Scraped off SpaceX Falcon 9 Wikipedia.
- The data were then used to perform EDA, where specific patterns in variables can be noticed and were collected as training data in training the machine learning models for during the EDA, such as:
  - ✓ Payload Mass, Flight Numbers, Orbit types and etc.
- Dashboard and Folium maps were created for better visualization of further analysis
- The final machine learning model building showcased 4 types of models with the same performance and accuracy of 83%, which in theory, any of the 4 models can be used for predicting landing outcomes of the First stage by Space Y.
- However, Space Y should remain cautious while interpreting the models' result as per mentioned, the model were tested on a small sample size, in which the data could potentially be inaccurate due to its large variances.
- More sample testing data should be collected for further model building to determine the best performing models.



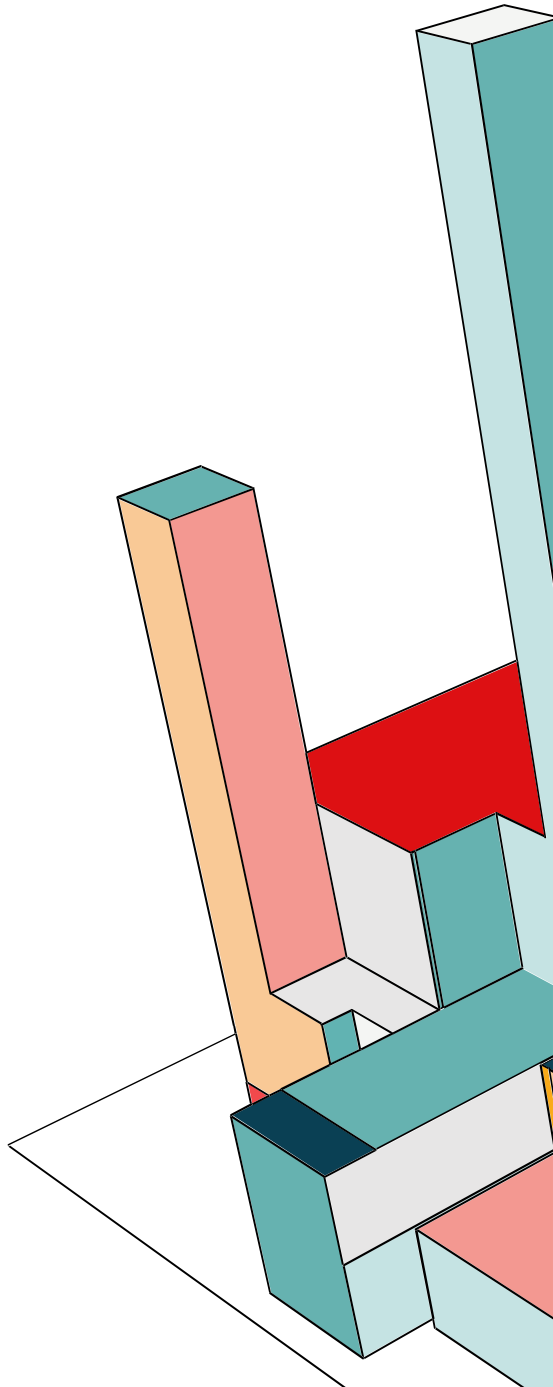
APPENDIX

Github URL for Full Capstone Project

https://github.com/TylerKoN/Data-Science-Capstone

Code Snippets in one URL:

https://github.com/TylerKoN/Data-Science-Capstone/blob/main/Appendix/Appendix.ipynb



THANK YOU

CAPSTONE PROJECT BY:
TYLER LEE