IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : Advent fel R. Banez
Date : 4/30/2023

Outline

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

Executive Summary

- The use of Falcon 9 rockets have been increasing and have test from data that it is the most reliable rocket type than other version of falcon 9. with this data and other key insight on the factors that affect the success of the launch site will be test using key tools in data science such as EDA and data visualization. The following results have been derived that payload mass, target orbit , flight number and location of launch site have a significant effect on the success of the launch recovery of falcon 9 rocket. Further more the tree decision model can predict the success launch with accuracy of > 90%. There preliminary studies will serve as future reference for further in deep study of the factors that may in addition affect the launch success.

Introduction

- A start up company plans to compete with the Space X company for a reusable rocket to lessen its expense since Space X advertise its rocket launch cost of 62 Million Dollars compare to other which cost 160 Million Dollars a good as a start up company this preliminary study will serve as precedence for further study on the feasibility of building a rocket that would compete with the space X.

- This Project aims to predict whether the first stage of Falcon 9 rocket will land successfully. being able to predict whether the launch would be successful or not is crucial to determine the cost per launch of the rocket so that a reasonable amount of resources and time will be allocated to develop a reusable rocket that can significantly lower the launch cost and to bid against the space X in the future.

Section 1

Methodology

Methodology

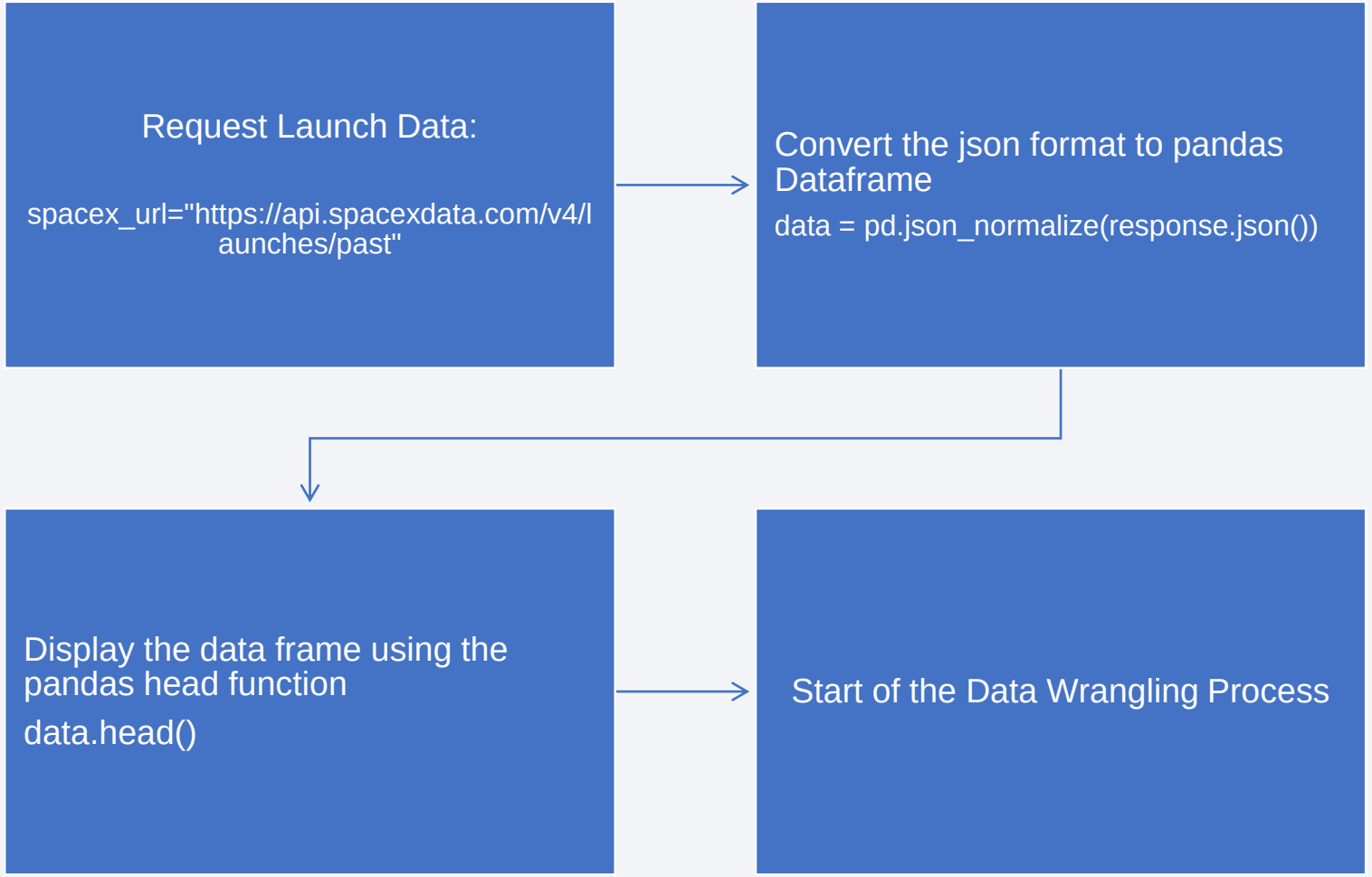
Executive Summary

- Data collection methodology:
  - Data collection process is through API and Data Webscraping method.
- Perform data wrangling
  - From raw data in json format it is converted to pandas.json_normalize function so that a uniform file format is created after that it is coverted to pandas dataframe and crucial factors in the table are extracted
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Since the prediction is based on the classification type then a series of model in machine learning such as K-neares neighbor, decision trees and SVM model are deploy to determine which model gives a more accurate prediction.

Data Collection

- Data is collected in the following manner.

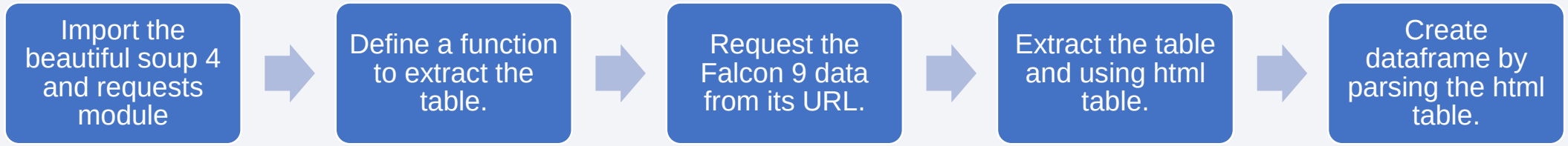
Data Collection – SpaceX API



Github URL

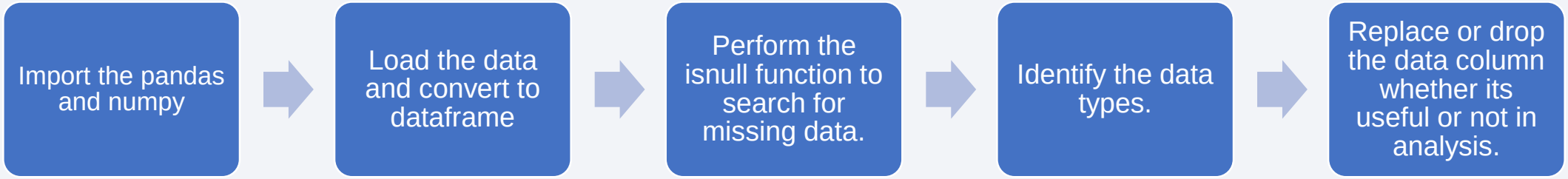
- https://github.com/advent-developer/IBM-Data-Science-Course/blob/main/Course%2010%20Applied%20Data%20Science%20Capstone/Lab%201%20Spacex%20data%20collection-api.ipynb

Data Collection - Scraping



- https://github.com/advent-developer/IBM-Data-Science-Course/blob/main/Course%2010%20Applied%20Data%20Science%20Capstone/Lab%202%20Jupyter%20Webscraping.ipynb

Data Wrangling



https://github.com/advent-developer/IBM-Data-Science-Course/blob/main/Course%2010%20Applied%20Data%20Science%20Capstone/Lab%203%20Data%20Wrangling%20Local%20Run.ipyn

EDA with Data Visualization

- The EDA Visualization technique use are scatter plot , lineplot and barplot this graphs are specifically selected due to the reason that it clearly shows the relationship of different variables that is easy to understand and provides a preliminary analysis of which factors are significant.

- https://github.com/advent-developer/IBM-Data-Science-Course/blob/main/Course%2010%20Applied%20Data%20Science%20Capstone/Lab%205%20EDA%20and%20Visualization.ipynb

EDA with SQL

- Download the data set and store it in the database
- Connect the IBD DB2 to the databse.
- Based on the result of the EDA Further perform data manipulation to create tables that reflect the important variables in the EDA visualtiozatin
- Save the data to the database for further analysis.
- Extract the data.

Build an Interactive Map with Folium

- A folium map is created to determine the optimum location of the launch site and to determine whether the location has a significant effect on the success of the first stage rocket launch of Falcon 9 the rocket launch are plotted in dots with radium and straight line between different address.
- Circles and Marks are added to the folium map to make the identification of launch site easy to locate and infer from the launch result.

- https://github.com/advent-developer/IBM-Data-Science-Course/blob/main/Course%2010%20Applied%20Data%20Science%20Capstone/Lab%206%20Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

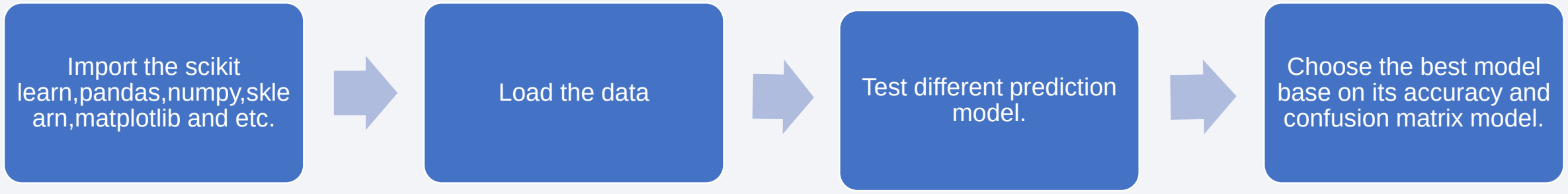
- An Interactive Dash board is created which contains the following graphs and chart namely a pie chart that plot the list of location of launch site the have a high success rate and a scatter plot of payloadmass vs success.

- This chart are included in the dashboard because location and payloadmass have a significant factor in the determination of the success of Falcon 9 launch.

- https://github.com/advent-developer/IBM-Data-Science-Course/tree/main/Course%2010%20Applied%20Data%20Science%20Capstone/Dash%20board%20Screenshot

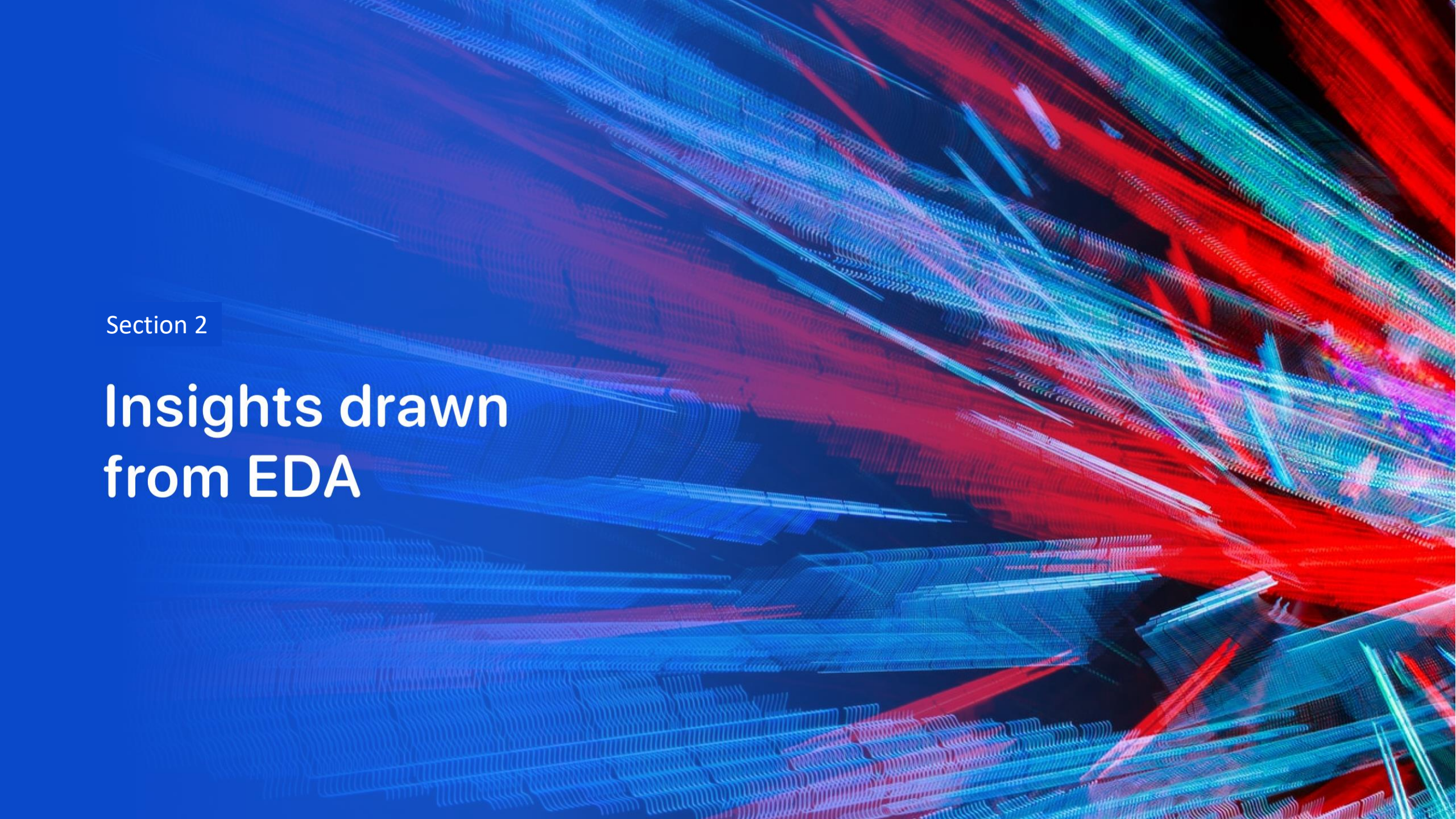
Predictive Analysis (Classification)

- Selection of Machine Learning Model in classification model and test each algorithm to determine the most accurate and reliable type of machine learning model

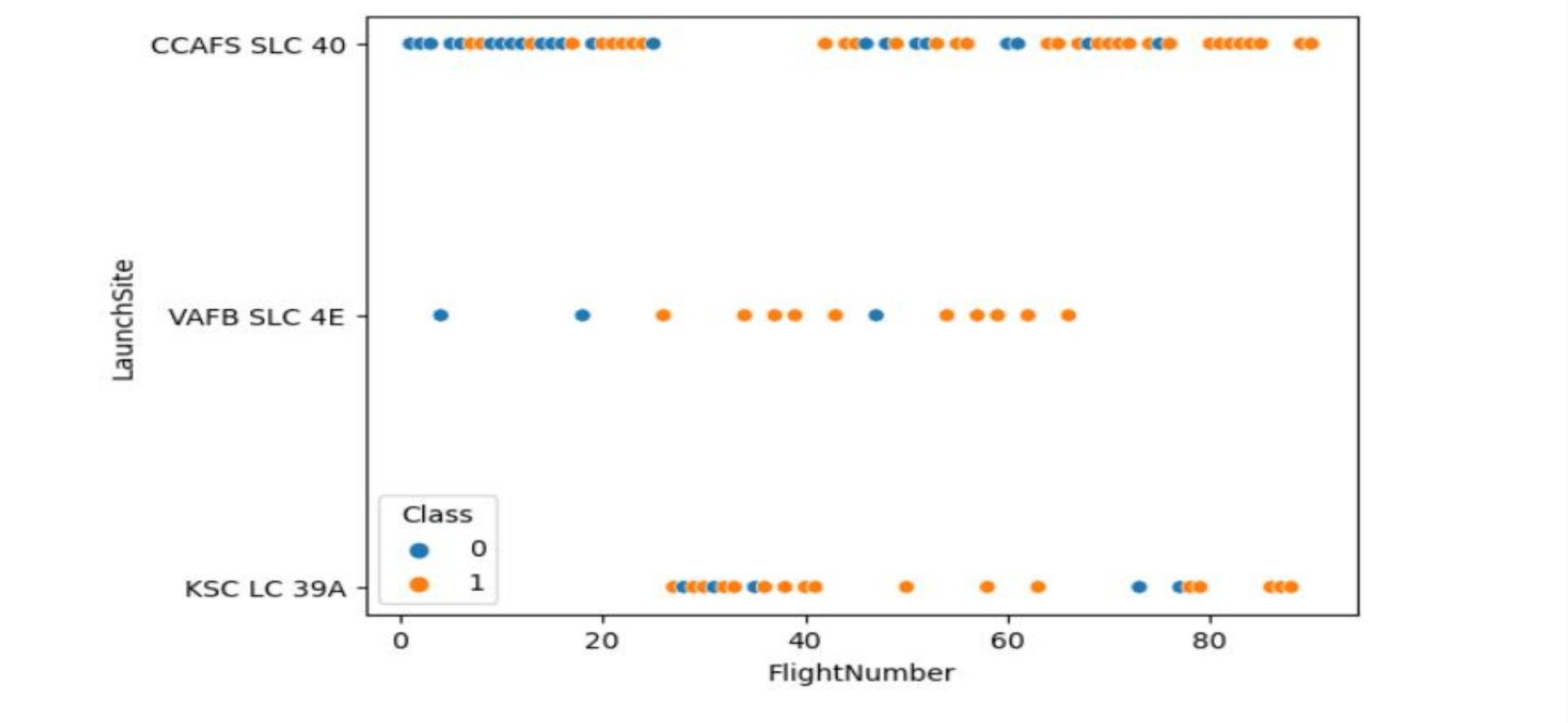


Results

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

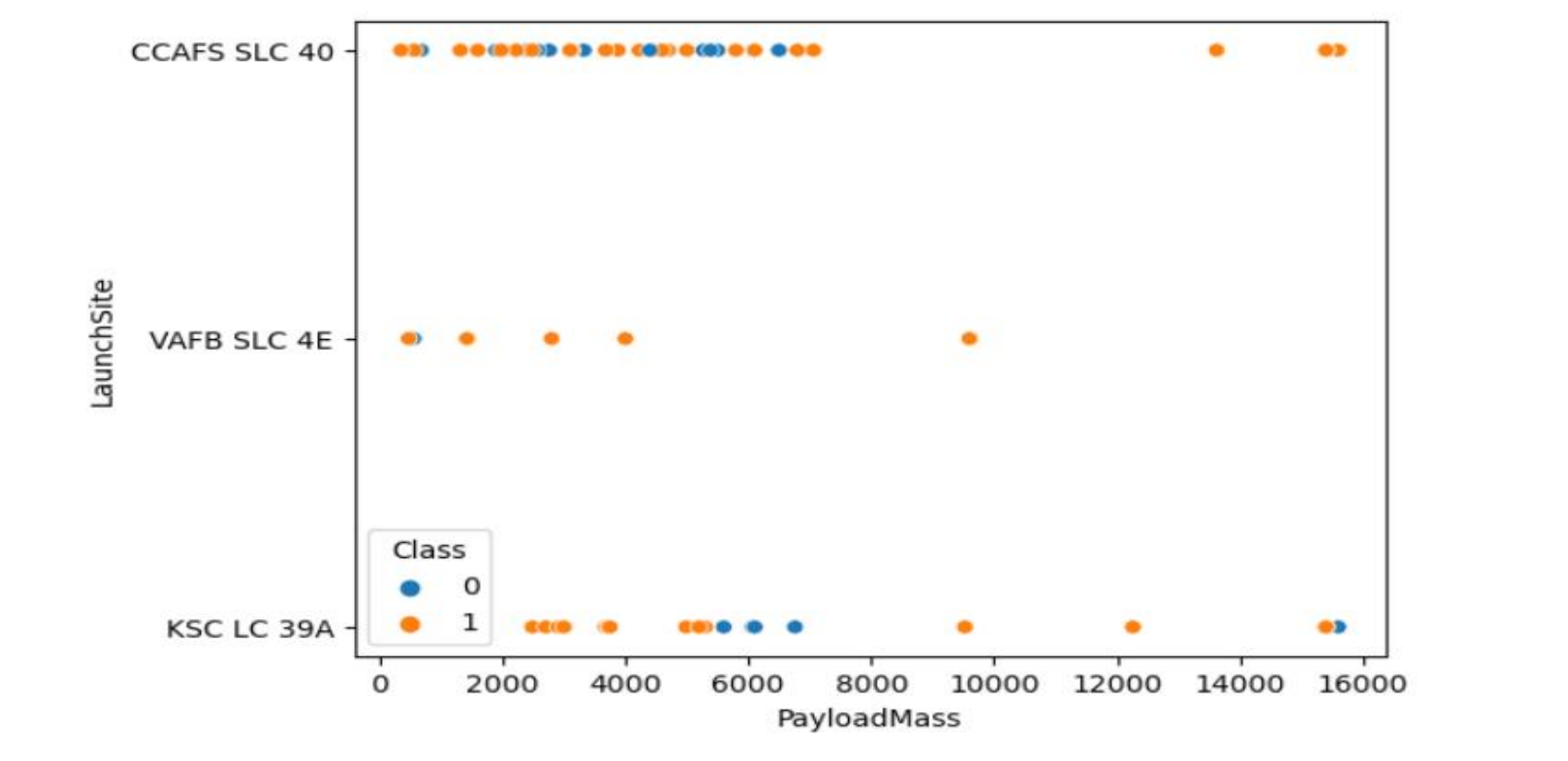

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



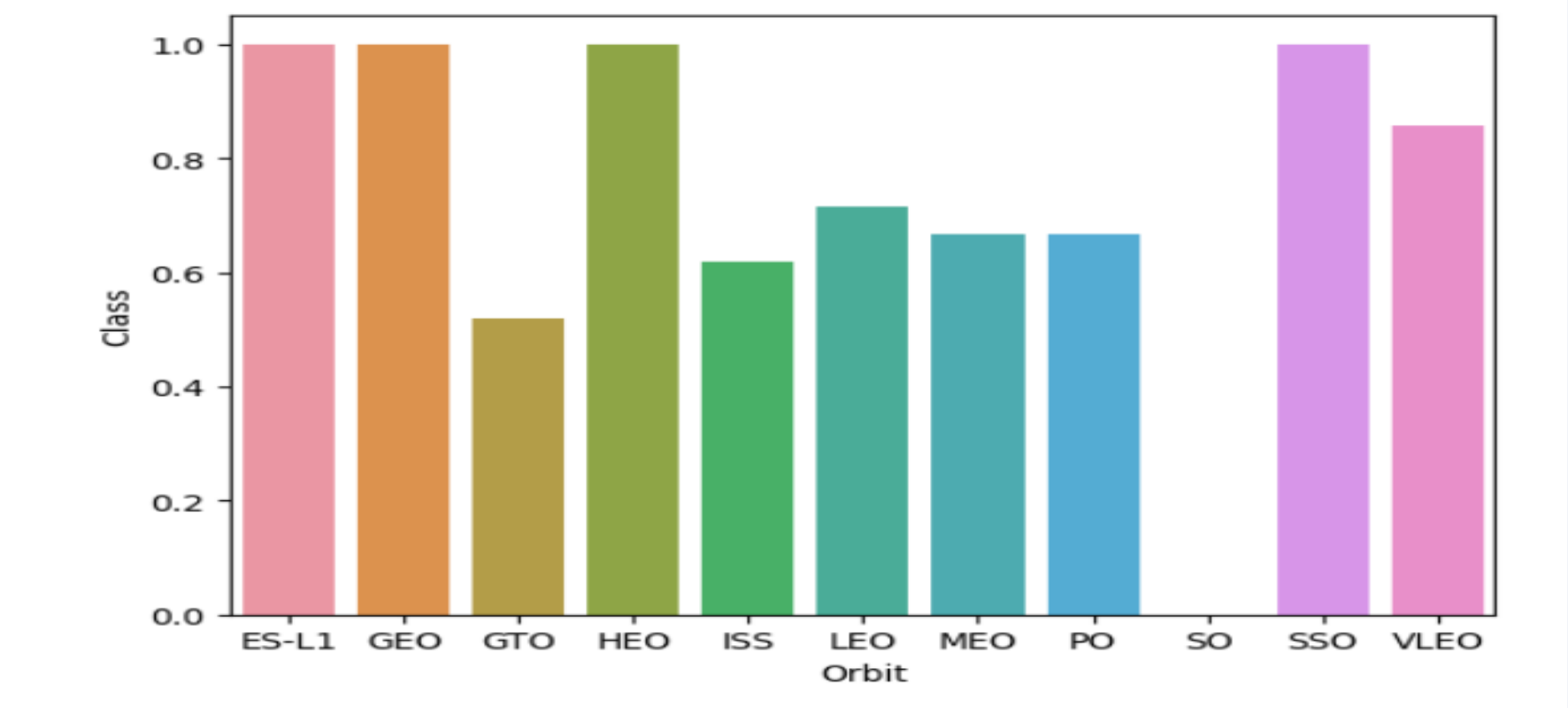
- Figure 1 shows that the rate of success and flight number increases in the location of SLC 40 and LC 39A with flight number of 23 and above the success is high.

Payload vs. Launch Site



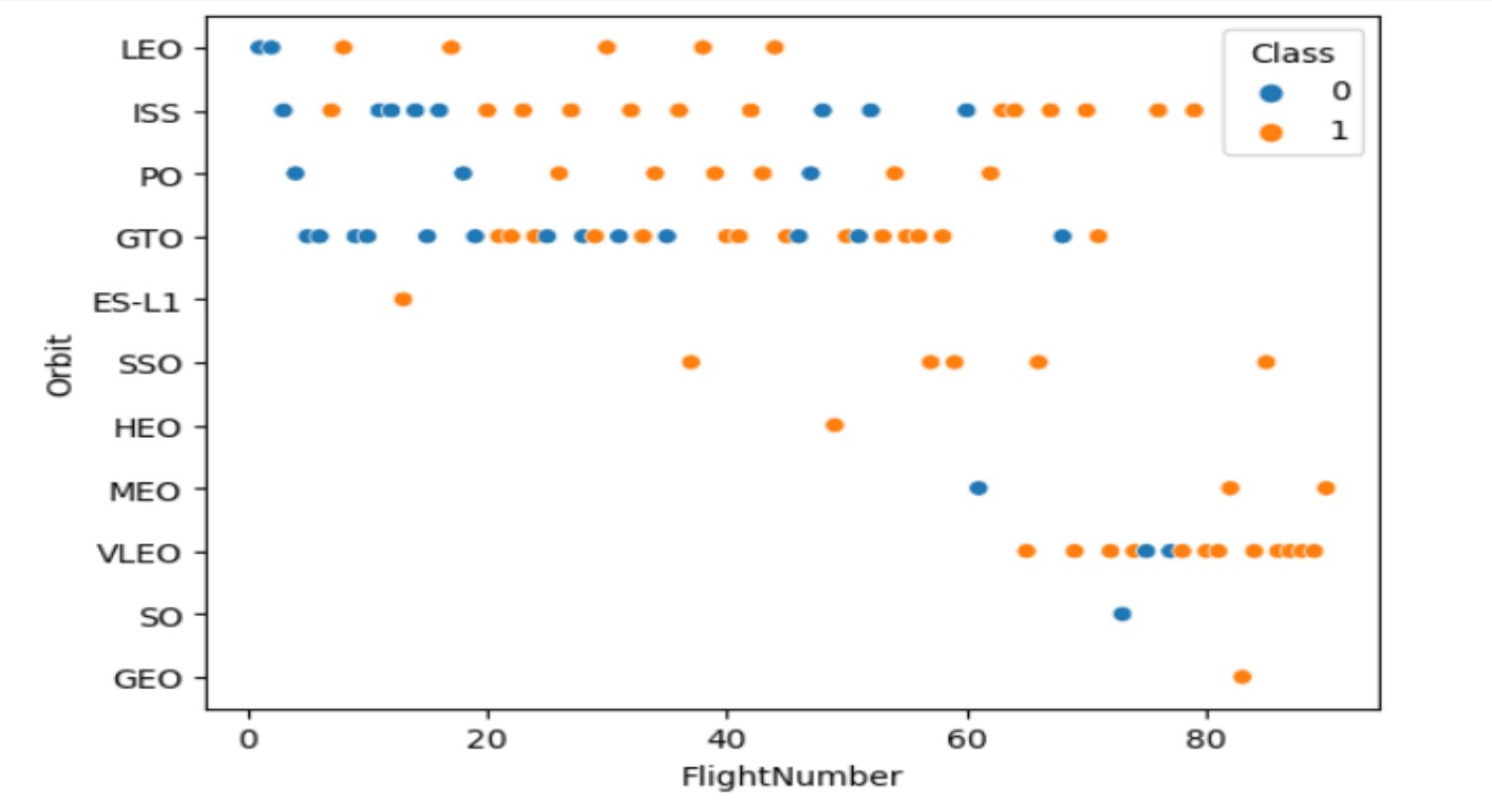
- Fig 2 shows that payload with a range of 500 to 5000 have a high chance of success and payload ranging to 6000 have a high chance of failure.

Success Rate vs. Orbit Type



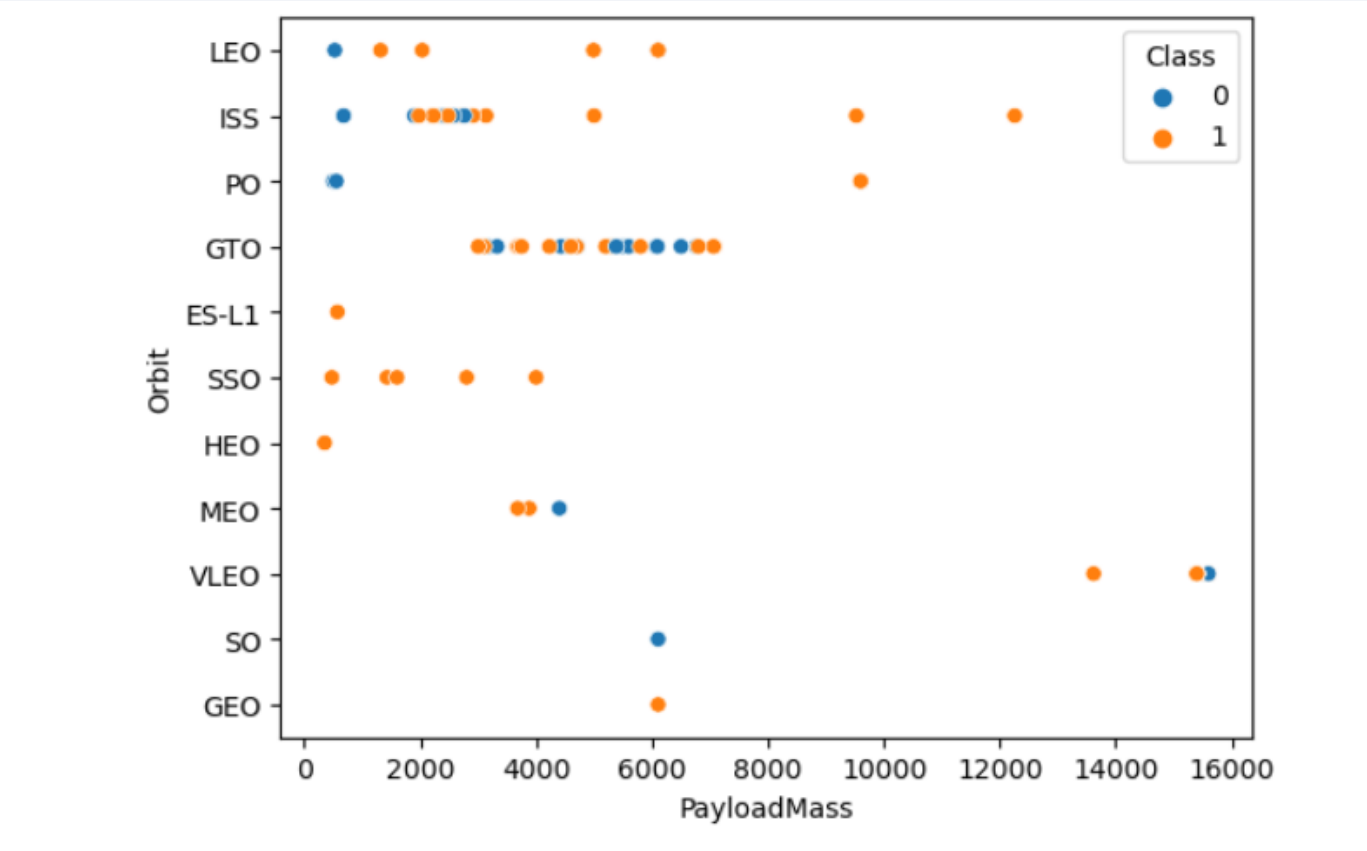
- Fig 3 shows the orbit in ES-L1,GEO,HEO,SSO and VLEO have a successful launch with the lowest in GTO orbit.

Flight Number vs. Orbit Type



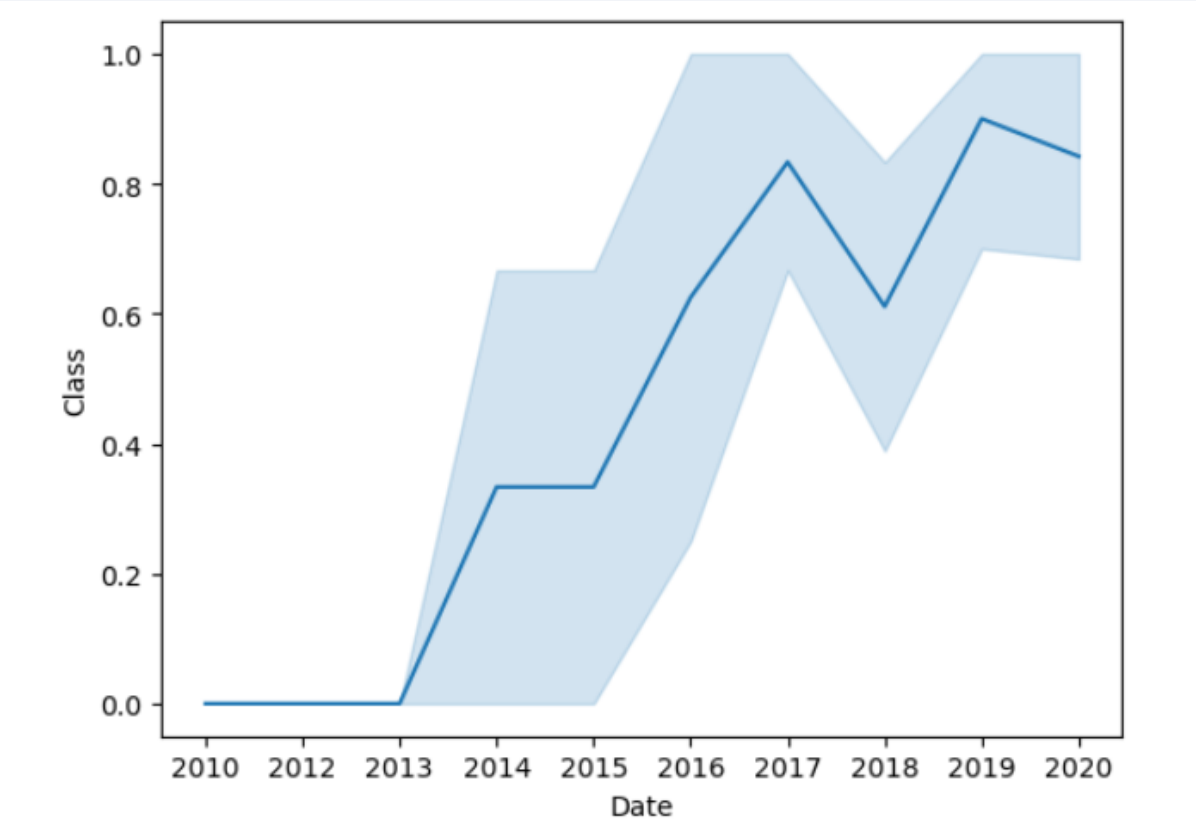
- Fig 4 from the figure above it can be inferred that orbit ISS and VLEO have a corresponding high rate of success as the flight number increase because of more experience in fine tuning the required parameter for a successful flight.

Payload vs. Orbit Type



- Fig. 5 shows that SSO orbit with a payload range of 500 – 4000 have a high percentage of successful launch and recovery of the rocket.

Launch Success Yearly Trend



- Fig 6 shows the upward trend of successful launch starting from the year 2013 onwards with a slight down trend in the year 2017 – 2018.

All Launch Site Names

```
In [5]: %sql SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXTBL;
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[5]:

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

- Using the Count function in the SQL query it returns a list of unique location site.

Launch Site Names Begin with 'CCA'

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

In [6]:
```
%sql SELECT LAUNCH_SITE FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[6]:

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

- The code above instruct the Database to extract launch site with a name in CCA from Launch site table.

Total Payload Mass

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

In [12]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[12]:

total_payload_mass
45596

- The code instruct to show the total payload of the database from SPACEXTBL.

Average Payload Mass by F9 v1.1

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

In [13]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[13]:

average_payload_mass
2928

- The program instruct the database to retrieve the average payload from SPACEXTBL.

First Successful Ground Landing Date

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

Hint:Use min function

In [15]:
```
%sql SELECT MIN(DATE) AS FIRST_SUCCESSFUL_GROUND_LANDING FROM SPACEXTBL \
    WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[15]:

first_successful_ground_landing
2015-12-22

- The code instruct to Retrieve the first successful landing from the SPACEXTBL table.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

In [36]:
```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL \
    WHERE (LANDING__OUTCOME = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000);
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[36]:

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

- The code instruct the data base to retrieve the names of 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

List the total number of successful and failure mission outcomes

In [49]:
```
%sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[49]:

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- The code select the column mission outcome and count it to list successful and failed launch in the SPACEXTBL table.

Boosters Carried Maximum Payload

```
In [54]: %sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEXTBL \
             WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[54]:

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- The code extract from the database a list of booster with a maximum payload in carry.

2015 Launch Records

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [58]:
```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL \
    WHERE (LANDING__OUTCOME = 'Failure (drone ship)') AND (EXTRACT(YEAR FROM DATE) = '2015');
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[58]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- A the code retrieves the booster version where the result is a failed landing from drone ship from the date 2015.

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

In [66]:
```
%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;
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[66]:

landing__outcome	total_number
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

- The code retrieves the landing outcome and counts the respective result store it in the total number between the data 2010 – 2017.

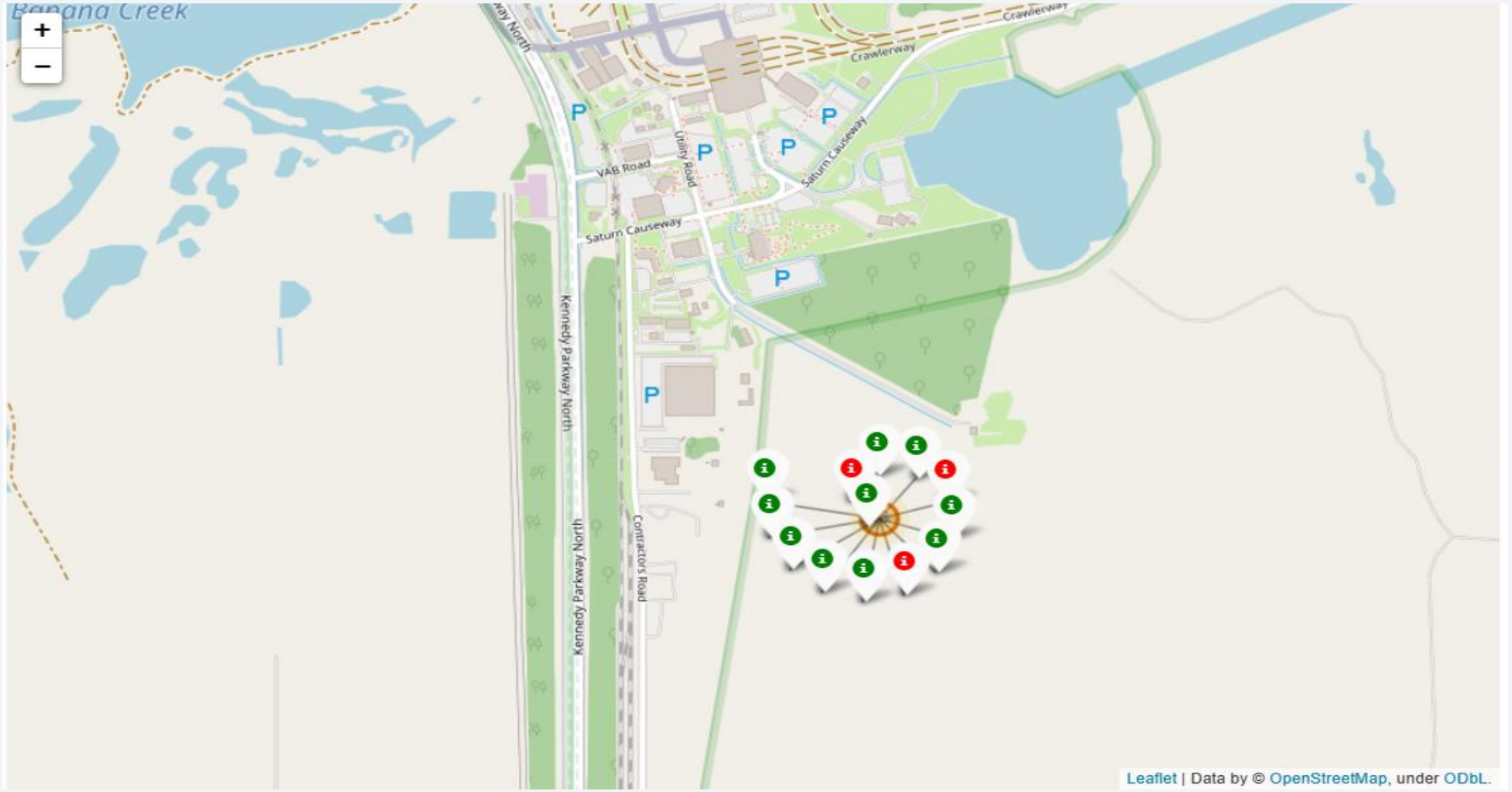
Section 3

Launch Sites Proximities Analysis

Launch Site Location

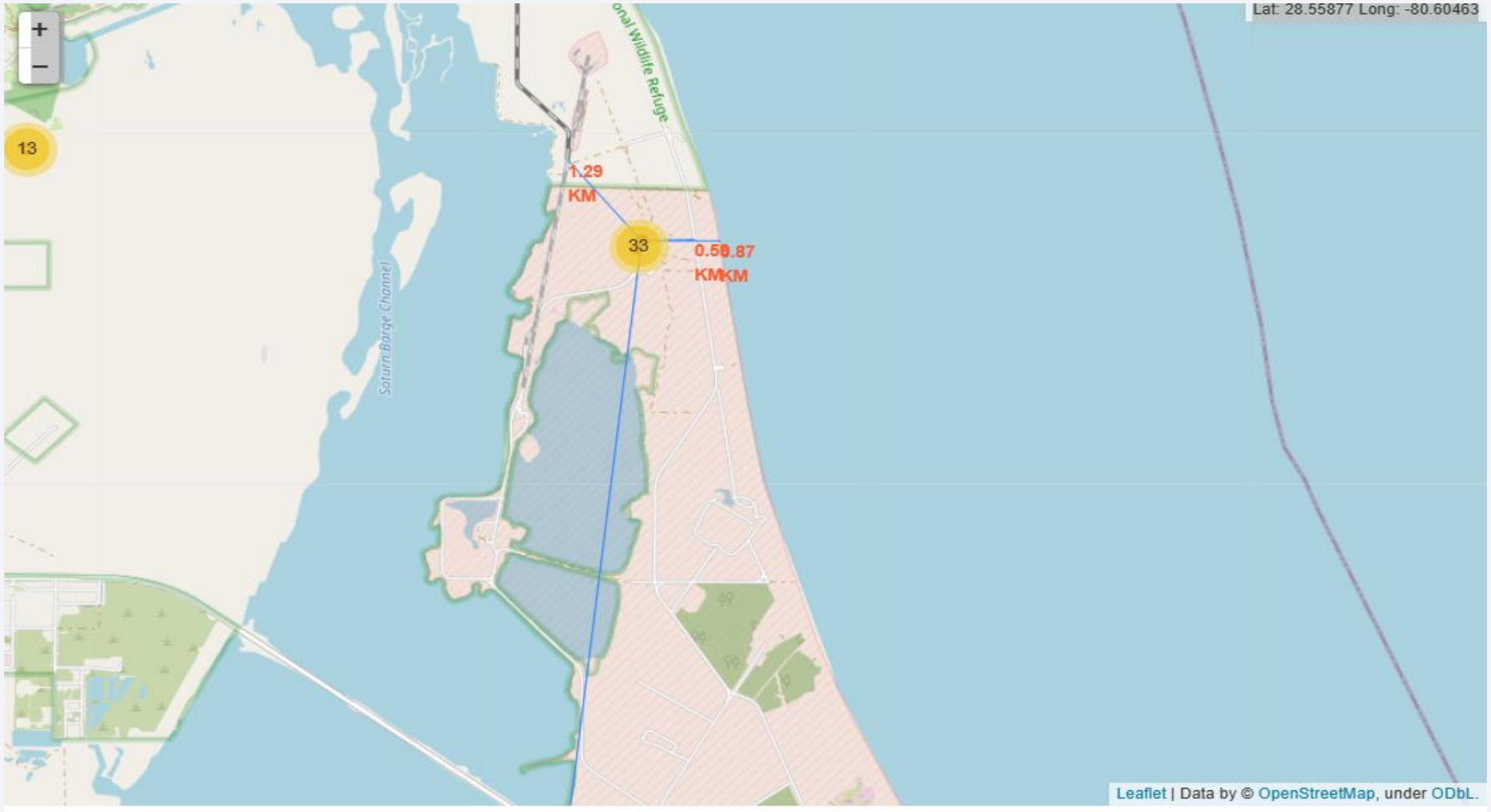
- The red dot indicate the launch site.from the hypothesis that success of the location can have a factor in the determination of whether the launch is successful or not. It is important to show the location of this launch site in folium map

Location of Successful launch site

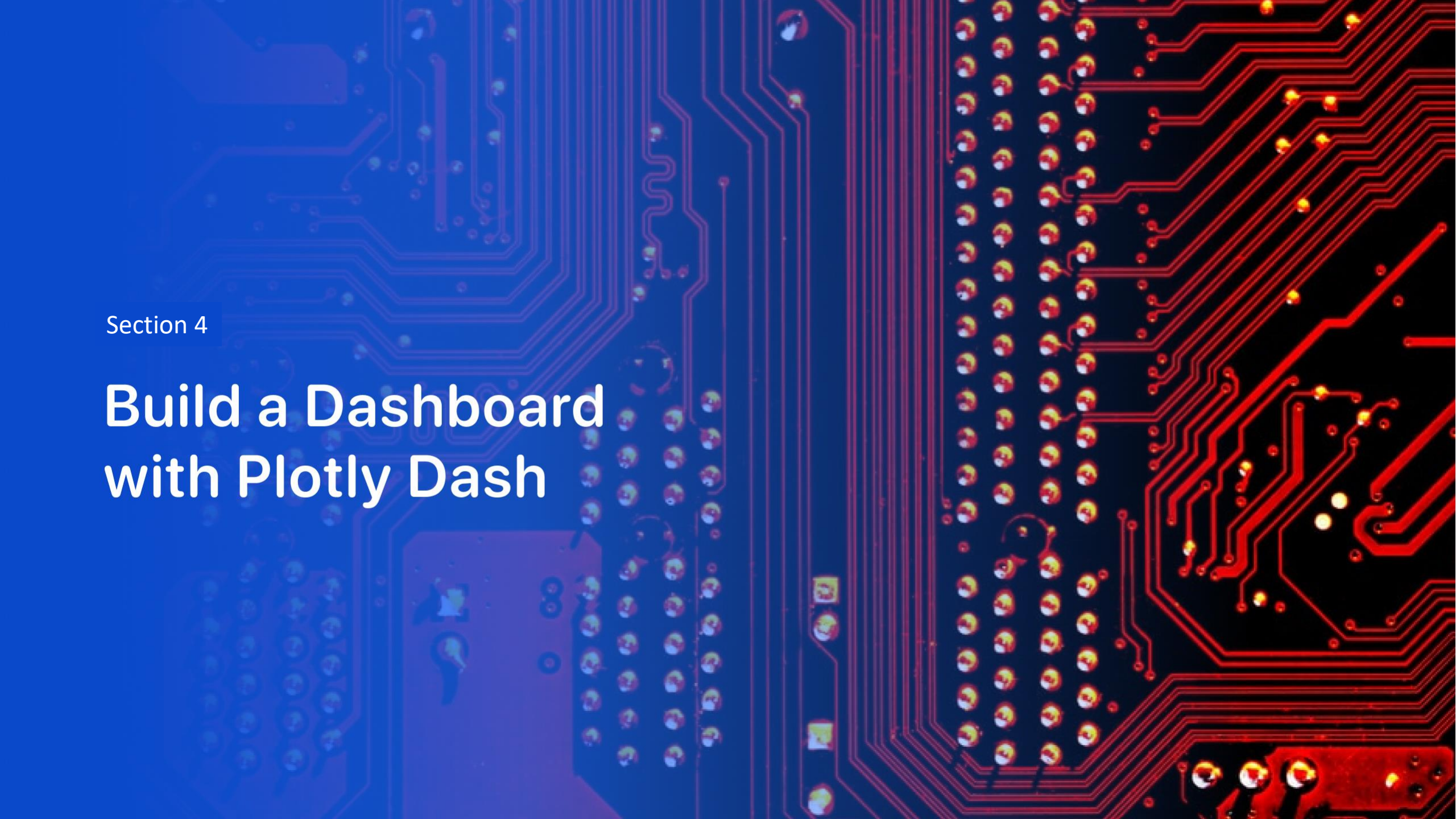


- From the figure above it shows that the Cape Canoveral Space center has the most number of successful launching site than its east coast counterpart the Vanderberg Space Center.

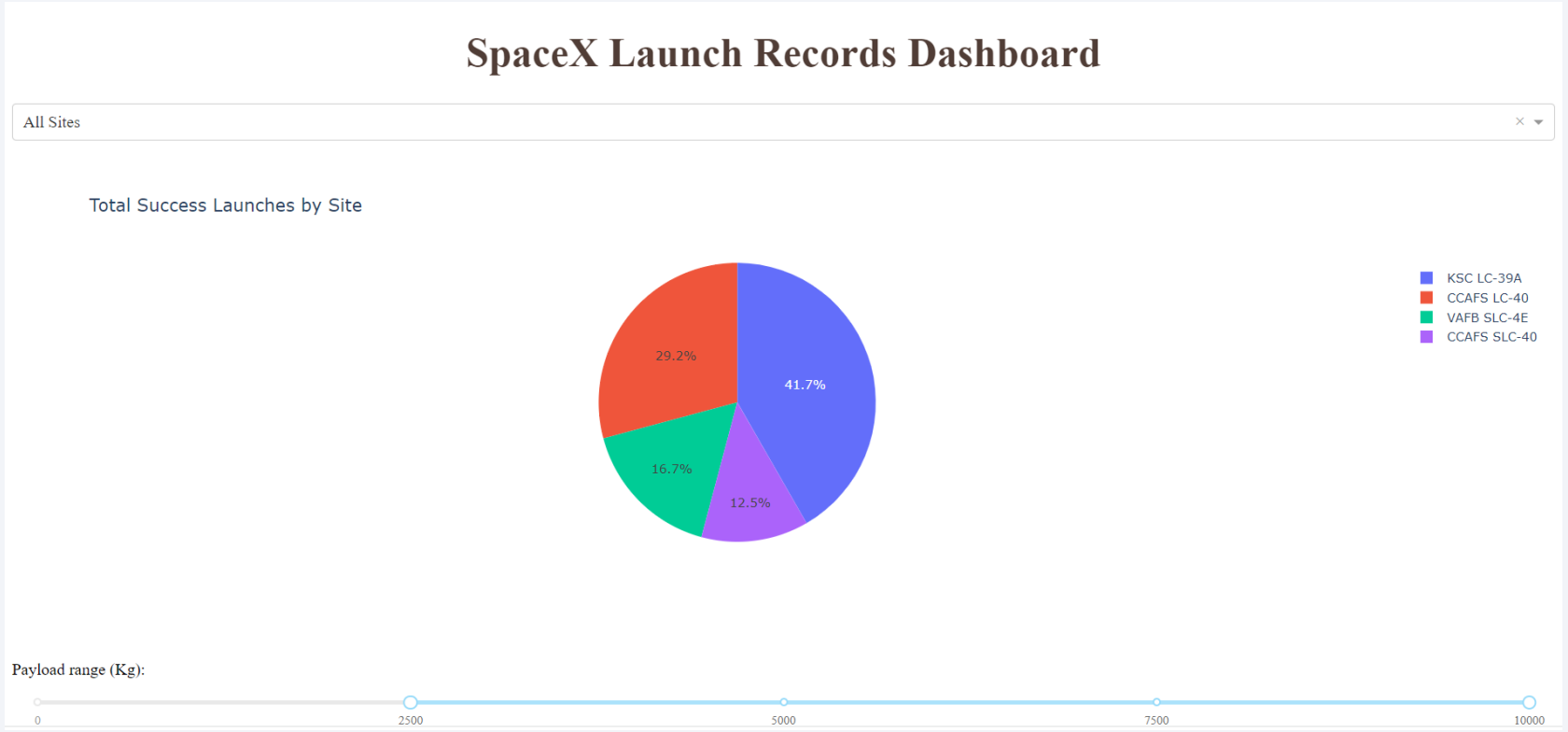
Approximate Distance from coastline and urban area.



- The distance from its nearest road is 1.29 km while from coastal area is 0.87 km. this area is less dangerous to the populace due to the danger of rocket failure that can cause debris to fall that may result in severe casualty if the site is near in the city or crowded place.

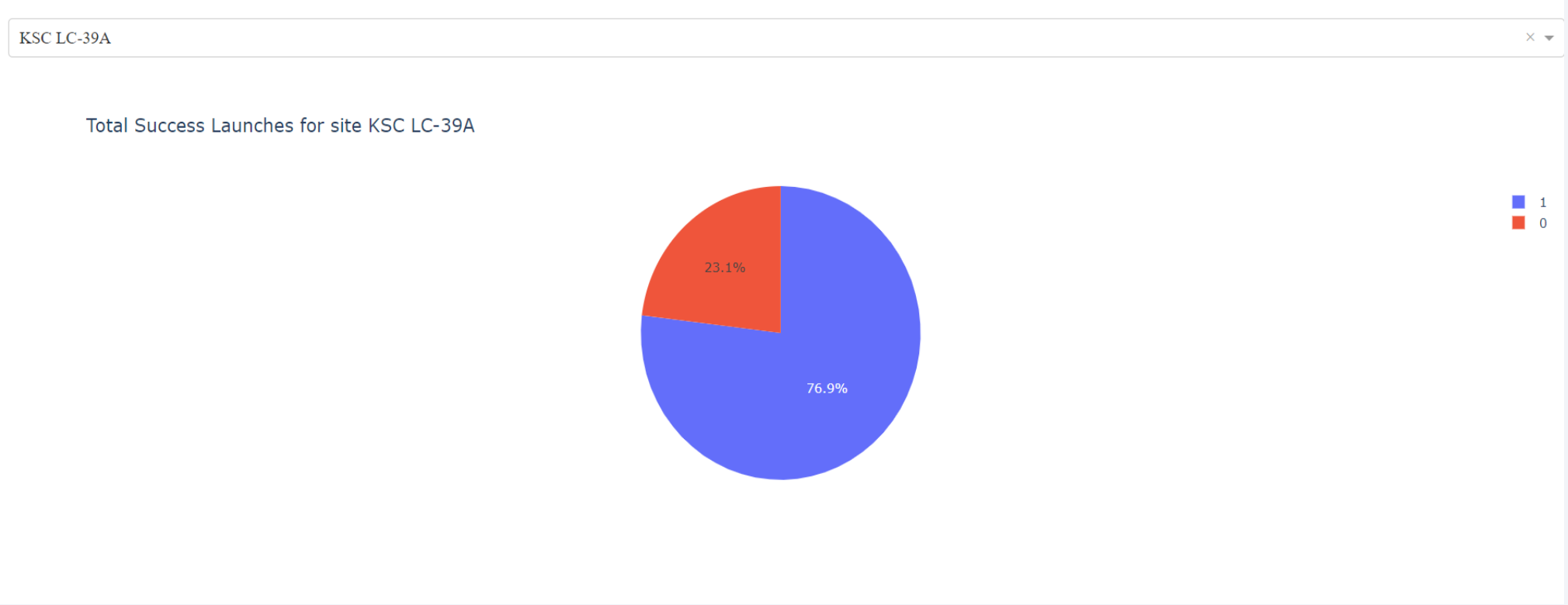

Section 4
Build a Dashboard with Plotly Dash

Interactive Visualization App



- From the pie chart above the location which shows the most successful launch is the KSC LC-39A with a total success of 41.7% than the rest of the area.

Percentage of Successful Launch site area.



- The percentage of the successful launch in the KSC LC-39-1 Is 76.9%

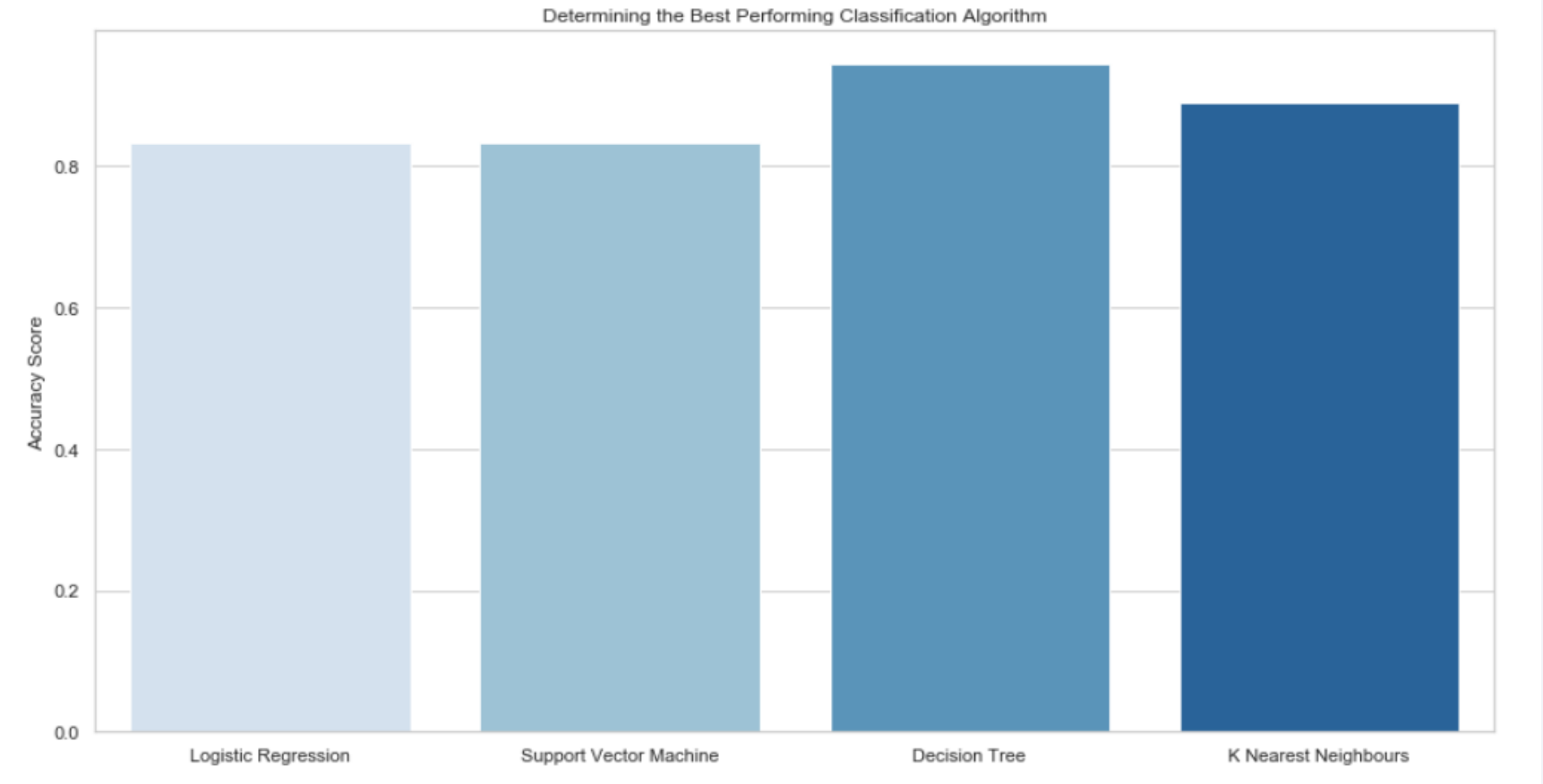
Payload Mass vs Successful launch site

- From the graph above it shows that the payload range of 3k to 5.5k produce results into successful launch with Falcon 9 v1.1 type of rocket

Section 5

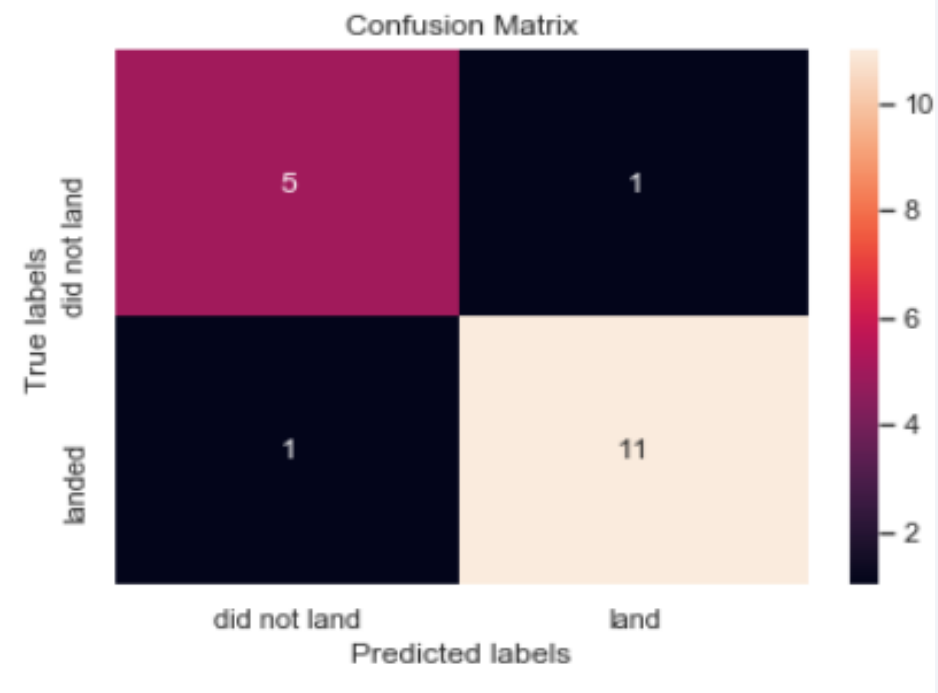
Predictive Analysis (Classification)

Classification Accuracy



- The figure above shows that the decision tree model is more accurate in predicting the success the launch with accuracy of 94%.

Confusion Matrix



- The model predicted 5 True negative and 11 true positive with 1 false negative and 1 true negative compared to other classification model with gives more error in the table.

Conclusions

- The decision tree model is the most accurate among the K-nearest neighbor and SVM model.
- The Payload mass of 500 to 5,500 kg with a rocket type of falcon 9 v1.1 has a higher percentage of successful landing than other rocket type.
- The Launch site KSC LC 39A has a 41.7% success rate than the rest of the launch site.
- The higher the flight number the higher the success rate of the launch as shown in the flightnumber Vs Launch Site.
- A launch with VLEO as the target orbit has a high chance of success.

Thank you!