IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Melissa Howard
22 June 2024

Outline

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

Executive Summary

Summary of methodologies

- Data Collection using API
- Data collection with web scraping
- Data Wrangling
- Exploratory data analysis with SQL
- Exploratory data analysis with Data Visualization
- Interactive Visual Analytics with Folium
- Machine learning prediction

Executive Summary

To calculate and visualize the accuracy of the SpaceX landing score I used a variety of models to highlight the accuracy score of the landings.

Using the exploration of data using SQL, visualization and folium maps and dashboards while updating the categories used in standardized data to find the best machine learning parameters, to visualize the accuracy score of all the models created.

This collection of data from public sources including Wikipedia SpaceX and the and the public access to the SpaceX API to as used to calculate the accuracy score of the landings.

Introduction

- I would like to determine the price of each launch of SpaceX.
- Using my dashboards, I will share with the team the information I have gathered to calculate of SpaceX will reuse the first stage.
- The calculations will display the relevant data required to calculate the reuse of the first stage.

Section 1

Methodology

Methodology

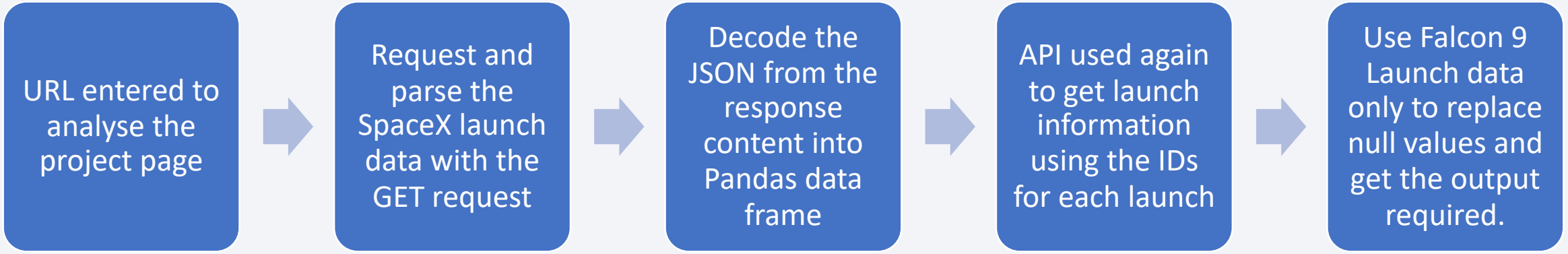
Executive Summary

- Data collection methodology:
  - Using Wikipedia web scraping and SpaceX API Get requests (to the API).
- Perform data wrangling.
  - Cleaning the data.
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.
- Perform predictive analysis using classification models.
  - By developing the most accurate Machine Learning Model.

Data Collection

Data sets are collected by

- Web Scraping (Wikipedia).
- SpaceX API request.



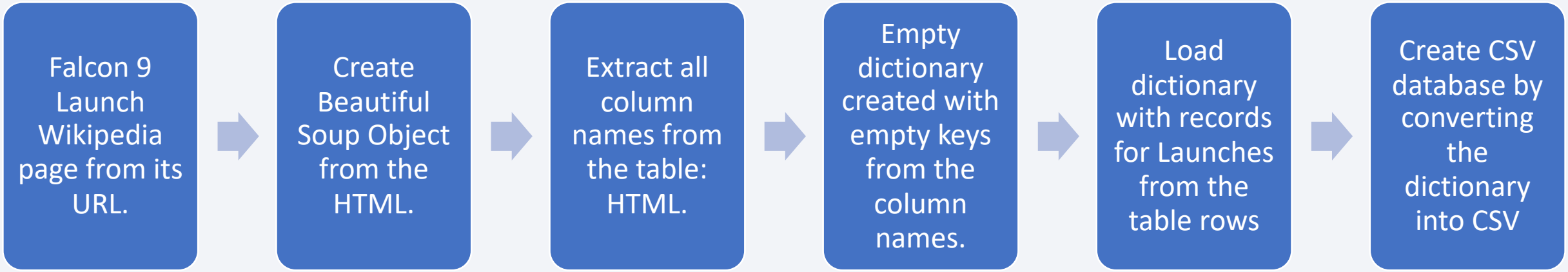
Data Collection – SpaceX API

- Data collection with SpaceX flowchart

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Data Collection by Web Scraping flowchart



GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/jupyter-labs-webscraping%20(2).ipynb

Data Wrangling

Calculated numbers of launches on each site.

Calculated the number of occurrences of each orbit.

Calculated the number and occurrence of each mission outcome per orbit type.

Created a landing outcome label from Outcome Column

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/EDA%20Notebook.ipynb

EDA with Data Visualization

Charts used:

- Scatter Plot, Flight Number vs Payload Mass, Flight Numbers Launch Sites, Payload and launch sites, Flight number and Orbit Type, Payload and Orbit Type.
- Success of each orbit with a Bar Chart
- Success rate and date with a Line Plot

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

SQL queries performed:

- 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 mas 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.
- 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. (Using a subquery).
- List the records that will display the month names, faliure_landing_outcomes in drone ship, booster versions, launch_site for the months in 2015.
- Task 10
- Rank the count of successful landing _outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium markers in the SpaceX launch sites were used to display areas affected by important transportation routes surrounding the area like:
- Railways
- Highways
- Cities
- Coastlines

Red represents rocket launch failures.

Green represents the rocket launch successes.

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

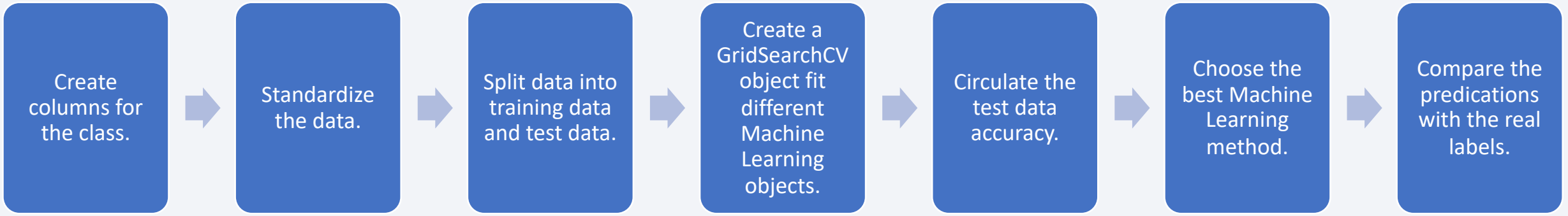
Build a Dashboard with Plotly Dash

- To visualize the launch of the SpaceX, the following charts were used:
  - Pie Charts
  - Scatter Charts
- From the charts that displayed the success rate of rocket per launch site, we could further understand the underlying factors affecting the success rate per site.
  - Successful launches were represented by 1
  - Failures (unsuccessful launches were represented by 0)

GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

Predictive Analysis (Classification)

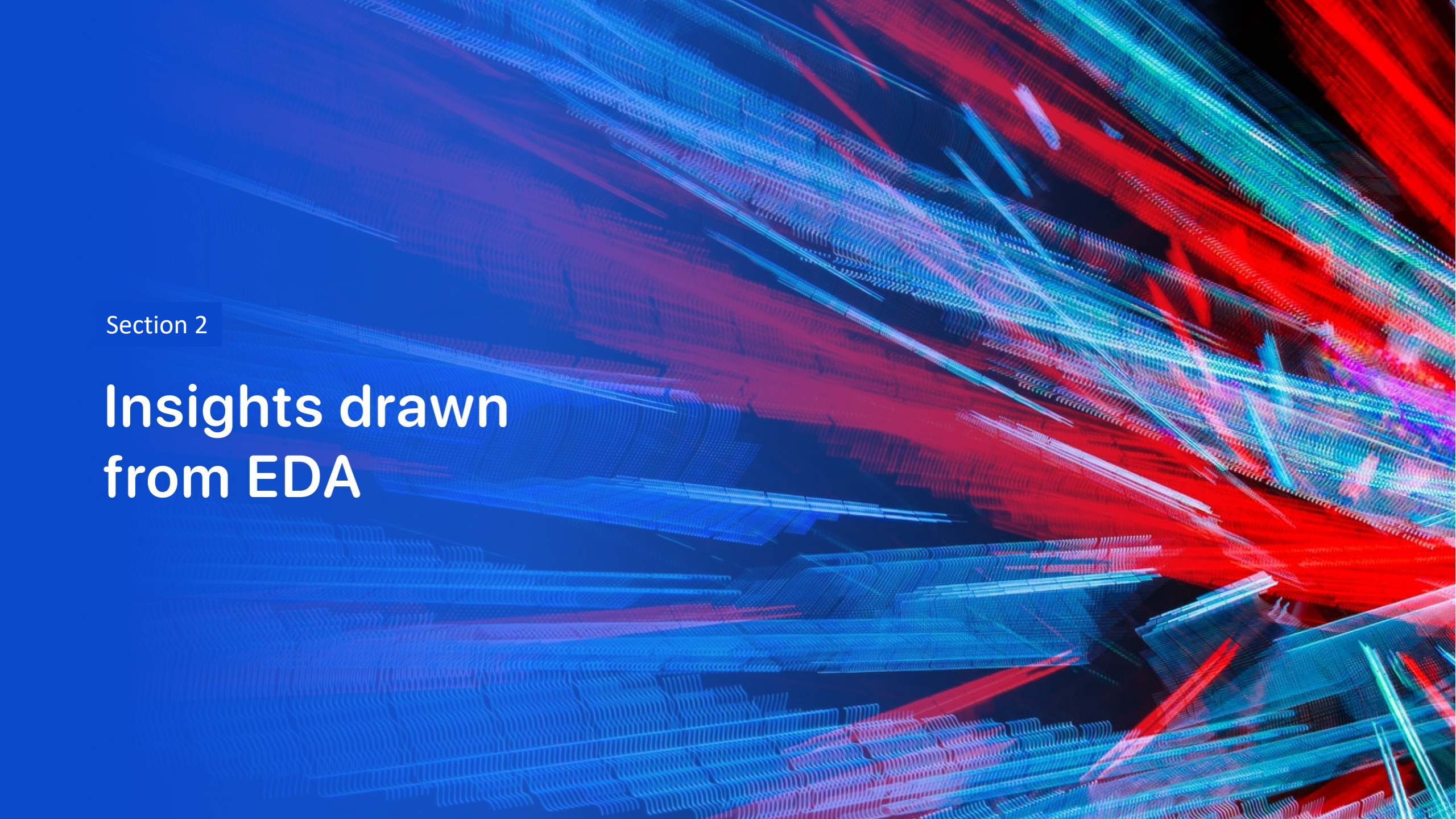
- The Machine Learning library that was used: Scikit-learn, for our predictive analysis.
- A machine learning pipeline was created to predict if the first land will occur, given the data.



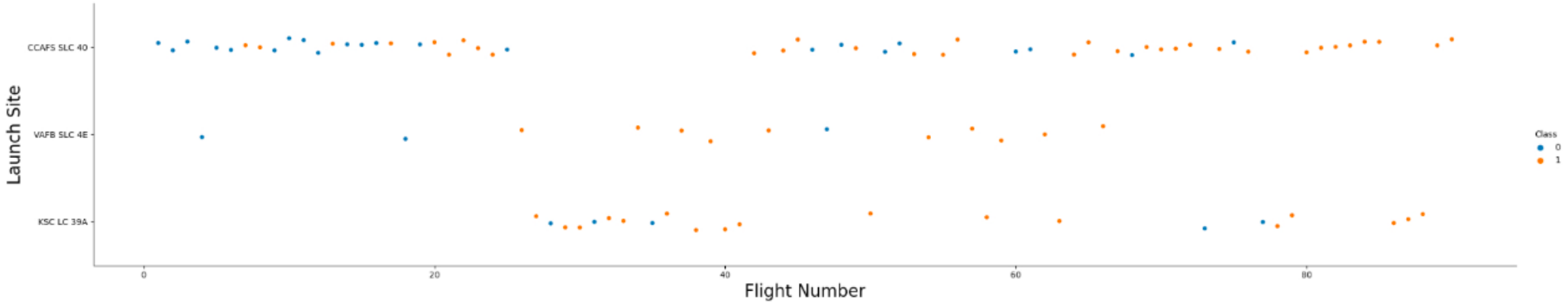
GitHub URL https://github.com/MelissaHowardza/CapstoneSpacexIBMDataScience/blob/main/module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

Results

- Exploratory data analysis results show us that successful landing outcomes have had a big increase since 2015.
- Successful landings somehow correlate with the flight number.
- Sites launched close to the coast lines could prove it it is easier to find test rocket landings without affecting any existing transportation routes in the area.
- Transportation routes closest to the railways and highways could make it quicker to access equipment and research material.
- The success rate Accuracy Score from the Machine Learning models displayed an accuracy score of 83.33%.

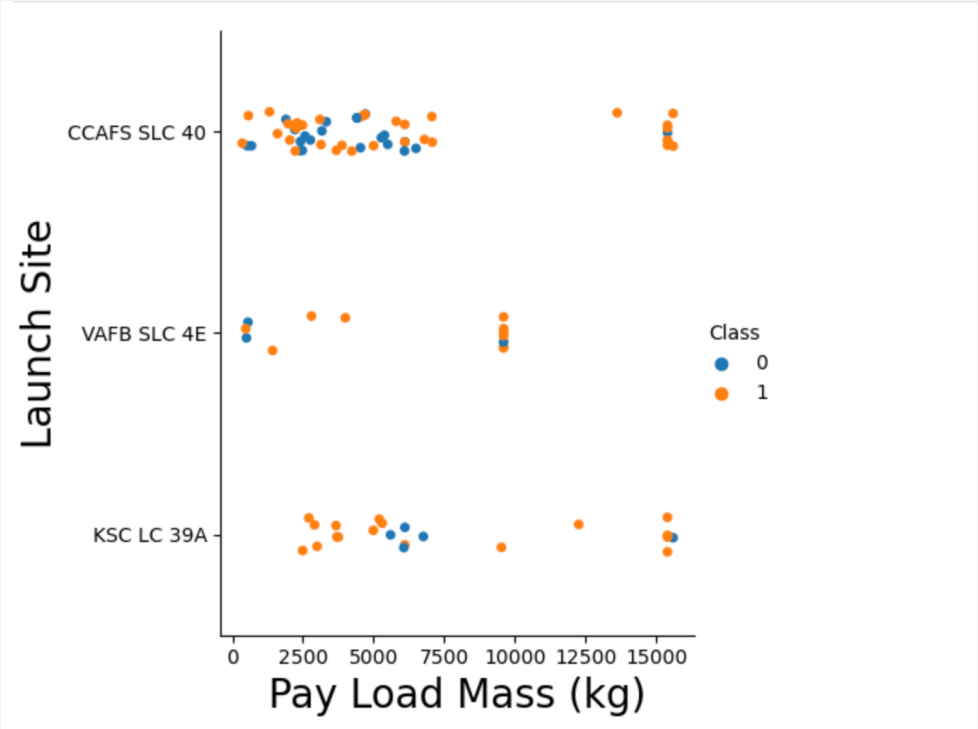

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



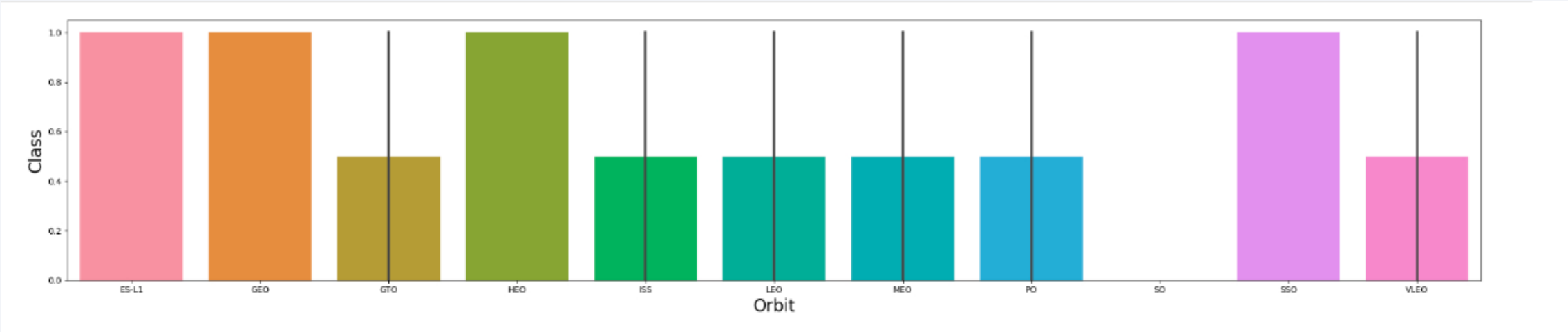
- Successful landings increased as the flight numbers increased.
- CCAFS SLC 40 had the highest number of landings.

Payload vs. Launch Site



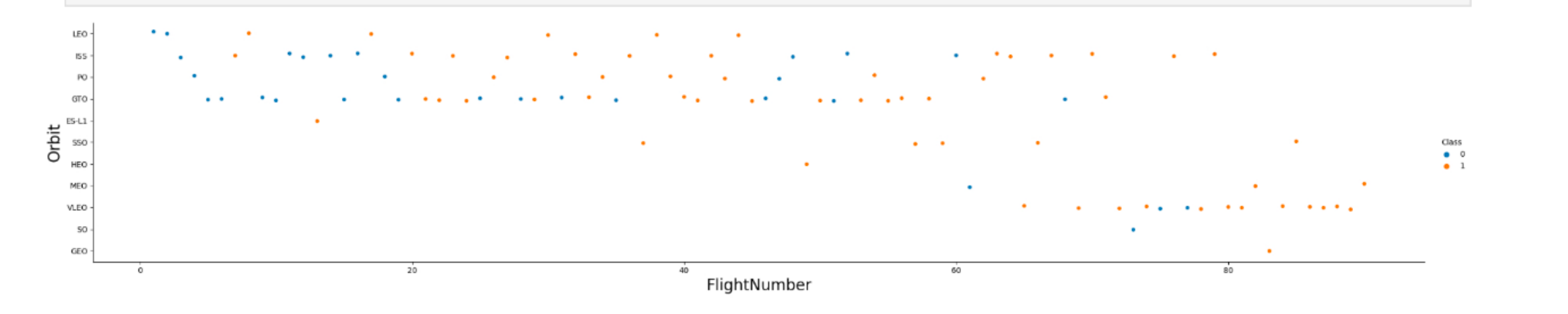
- The Scatter Point Chart shows that the VAFB SLC launch site has no rockets launched with a heavy payload mass, greater than 10000.

Success Rate vs. Orbit Type



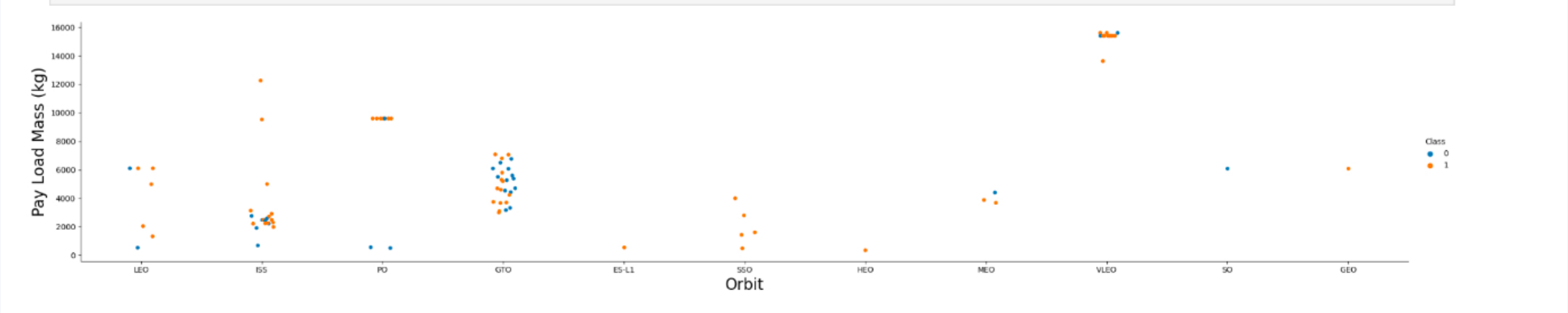
- ORBITS with the highest success rates are
  - ES-L1
  - GEO
  - HEO
  - SSO

Flight Number vs. Orbit Type



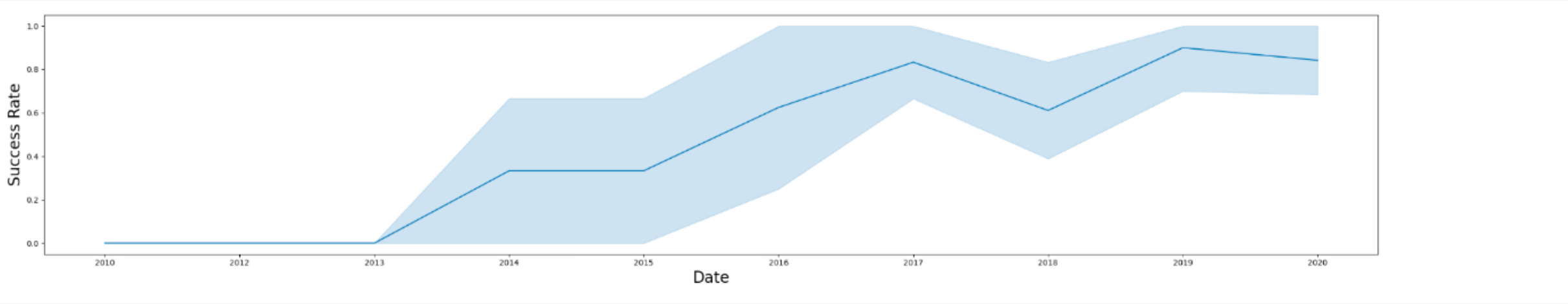
- There is no relationship between Flight Number and Orbit Type.

Payload vs. Orbit Type



- Polar, LEO and ISS with heavy payloads have had a successful landing rate.

Launch Success Yearly Trend



- There has been an increase in success rates between 2013 and 2020.

All Launch Site Names

- The data shows the landings were attempted at the following launch sites:
  - CAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

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

```
%sql select distinct (Launch_Site) from SPACEXTBL
```

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

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

Launch Site Names Begin with 'CCA'

In [33]:
```
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
```

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

Out[33]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	La
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	F
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	F
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	

5 records where launch sites begin with `CCA`.

Total Payload Mass

The total payload carried by boosters from NASA calculated at 45596kg:

```
In [47]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
 * sqlite:///my_data1.db
Done.
Out[47]: sum(PAYLOAD_MASS__KG_)
                        45596.0
```

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is calculated at 2928.4kg.

```
In [48]: %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
 * sqlite:///my_data1.db
Done.
Out[48]: avg(PAYLOAD_MASS__KG_)
         2928.4
```

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad was on the 1 August 2018.

In [52]:
```
%sql select min("Date") from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)'
```

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

Out[52]:

min("Date")
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

```
In [45]: %sql select BOOSTER_VERSION from SPACEXTBL where "Landing_Outcome" = 'Success (drone ship)' and PAYLOAD_MASS__KG_
 * sqlite:///my_data1.db
Done.
```

Out[45]:

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

Total Number of Successful and Failure Mission Outcomes

There are a total number of 99 successful and failure mission outcomes

```
In [61]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure
 * sqlite:///my_data1.db
Done.
Out[61]: count(MISSION_OUTCOME)
                              99
```

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass are:

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

```
In [60]: %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTB
 * sqlite:///my_data1.db
Done.
Out[60]:
```

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

2015 Launch Records – Failed Landing Outcomes

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are:

Failure outcomes for 2015

0. Booster Version: F9 v1.1: B1012

Launch site: CCAFS LC -40

1. Booster Version: F9 v1.1: B1015

Launch site: CCAFS LC -40

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

- There have been 5 landing outcomes (such as Failure (drone ship) or Success (ground pad) between the date 2010-06-04 and 2017-03-20

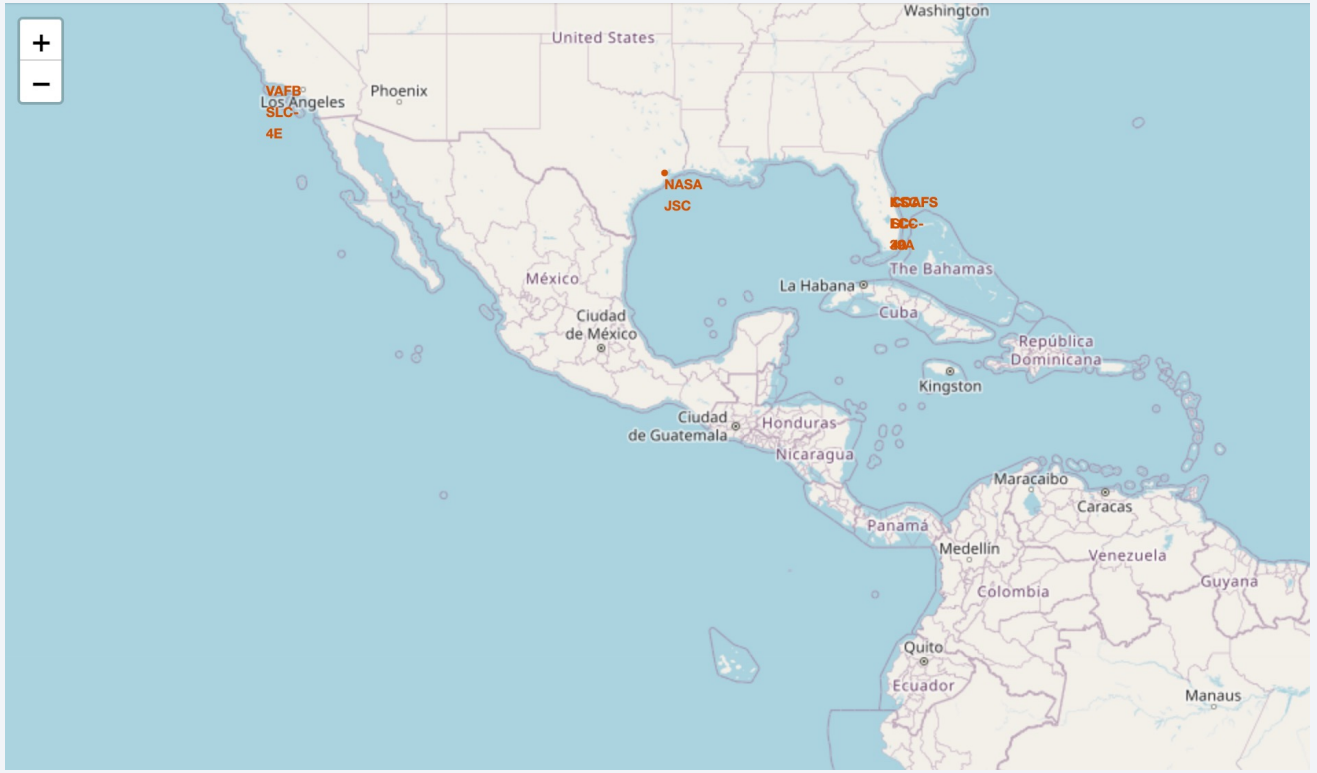
```
In [72]: %sql select count("Landing_Outcome") from SPACEXTBL where ("Landing_Outcome")=("Failure (drone ship)")  or ("Land
 * sqlite:///my_data1.db
Done.
Out[72]: count("Landing_Outcome")
                                5
```


Section 3
Launch Sites
Proximities Analysis

Launch Site Locations

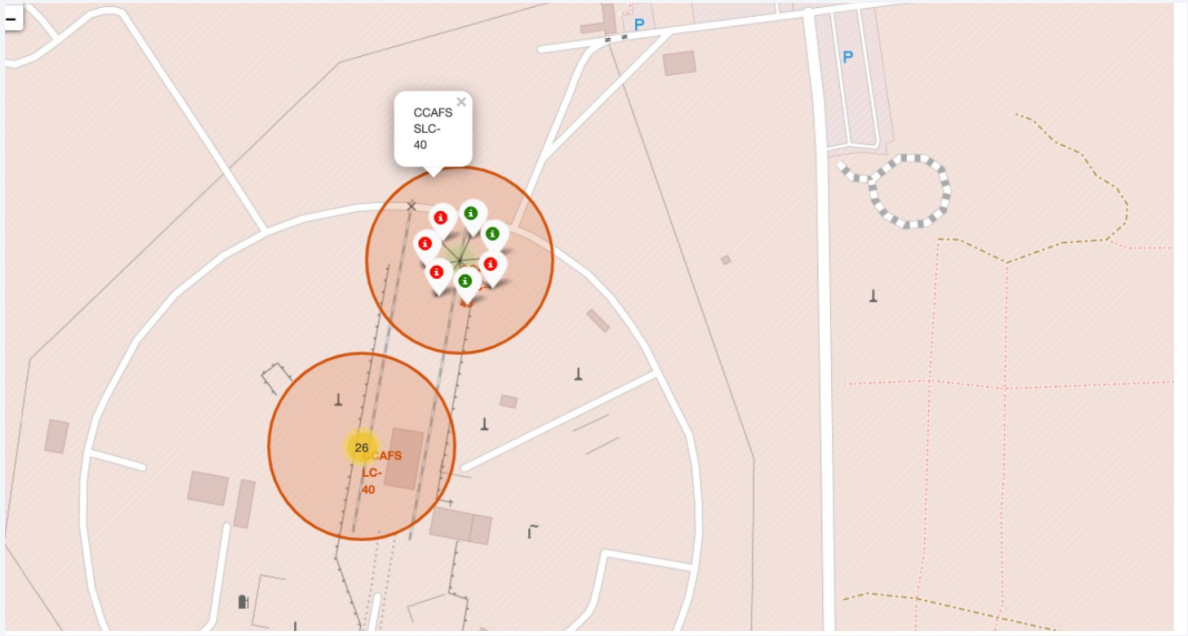
There are 3 Launch site locations marked in red on the map:

1. VAFB SLC-4E
2. NASA JSE
3. KSE LC-39A



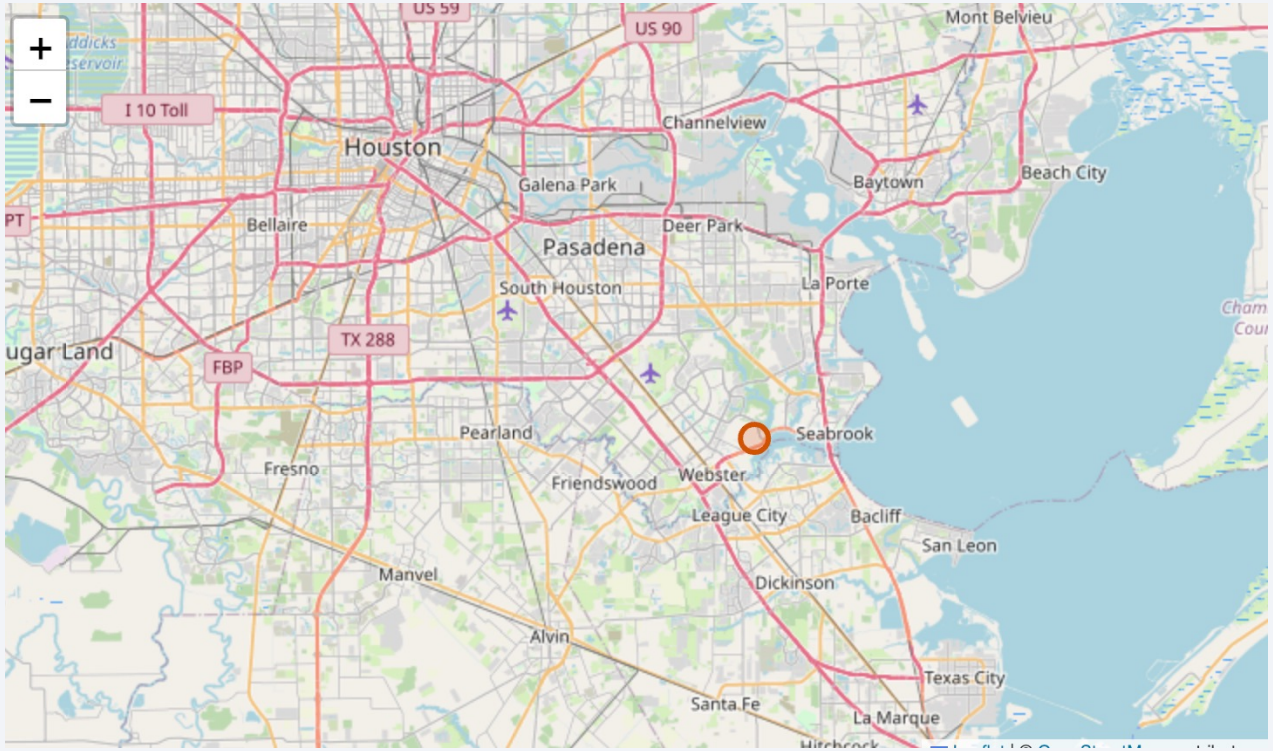
Colour display map indicators of success and failure rates

- Successful landings are displayed in Green.
- Unsuccessful landings are displayed in Red.

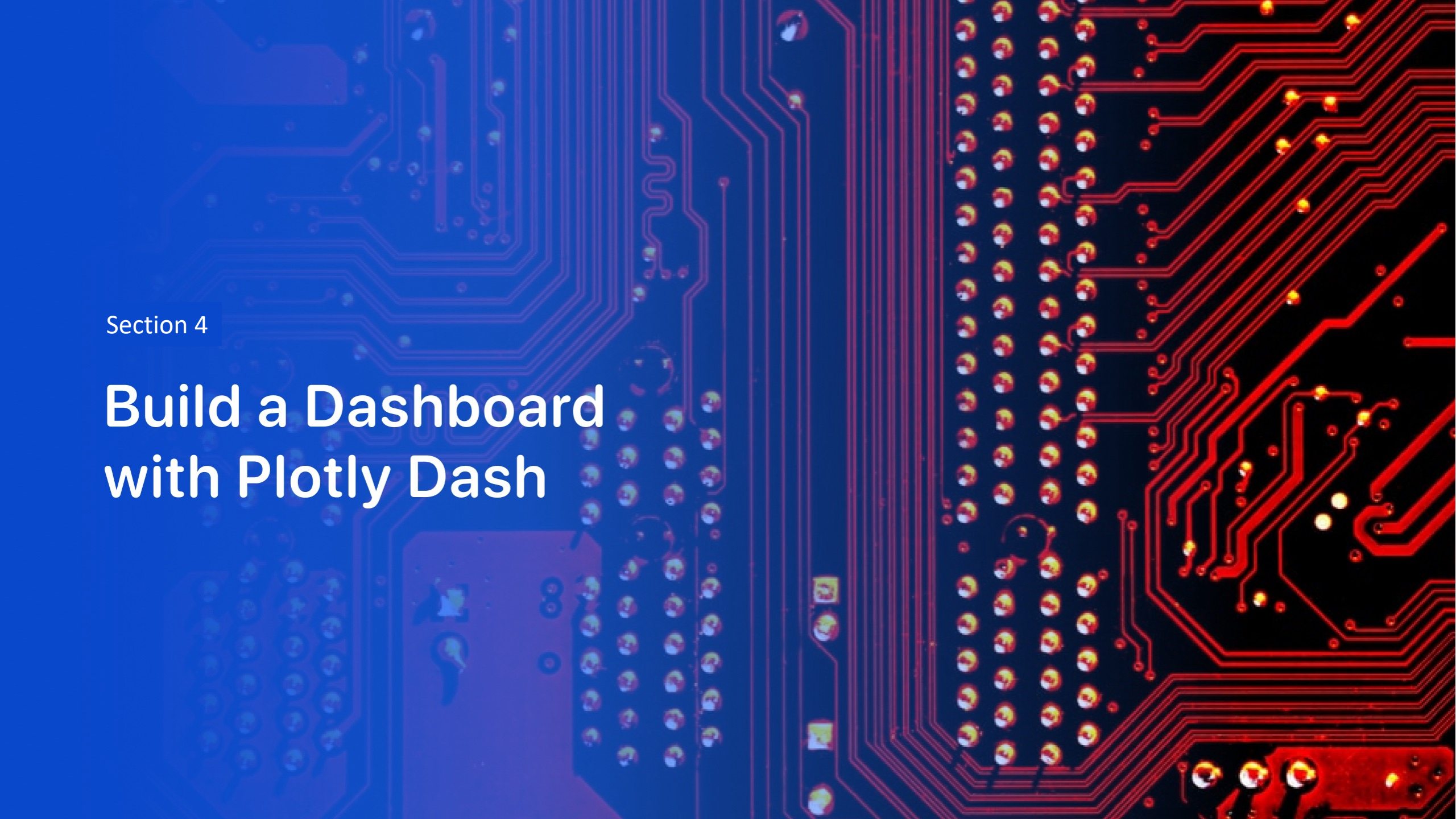


Launch sites near transportation areas

Successful launch sites are near proximity to railway and transportation areas.

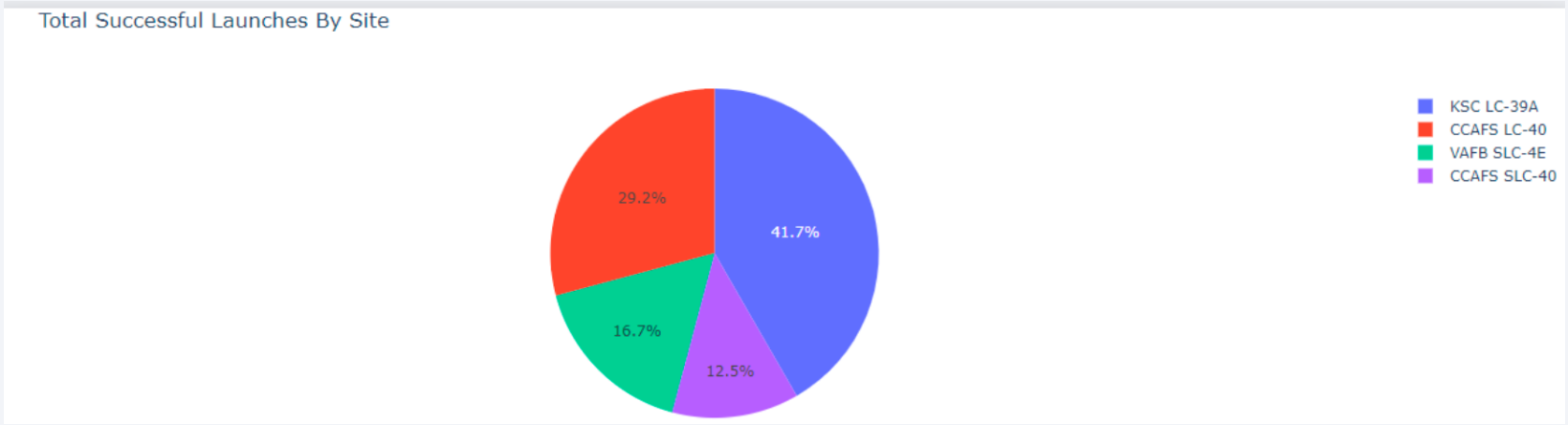


Here are displayed railway lines and airports close to the transportation areas and to the ocean should any launch fails happen.


Section 4
Build a Dashboard with Plotly Dash

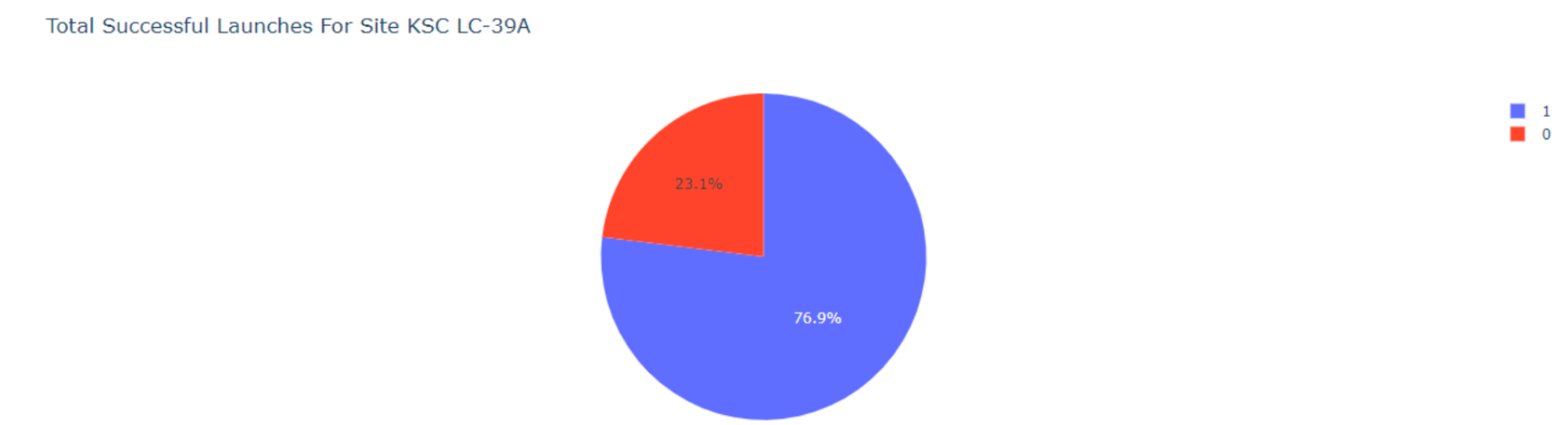
Total successful launches by Site

- KSC LC-39A has the most successful launches and the highest success rate.



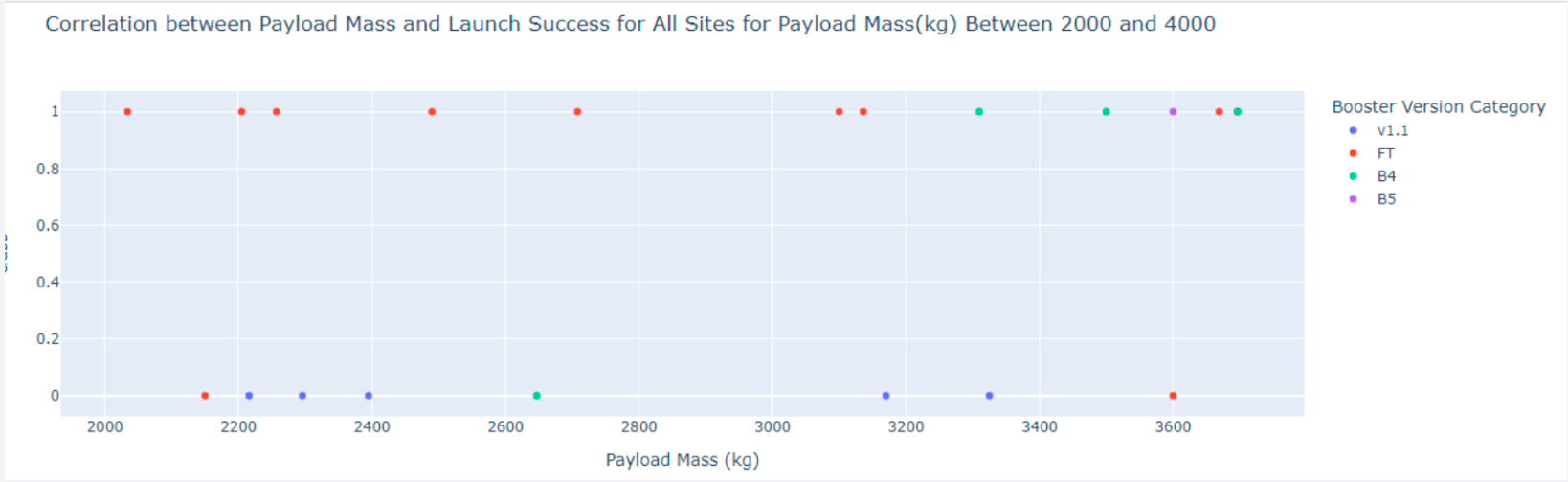
Total Successful Launches for Site KSC LC-39A

- 76.9% of the KSC LC-39A were successful. Making it the highest rate of successful launches.



Payload Mass vs. Launch Success for All Sites

- The payload mass between 2000 and 4000 has the highest success rate.



Section 5

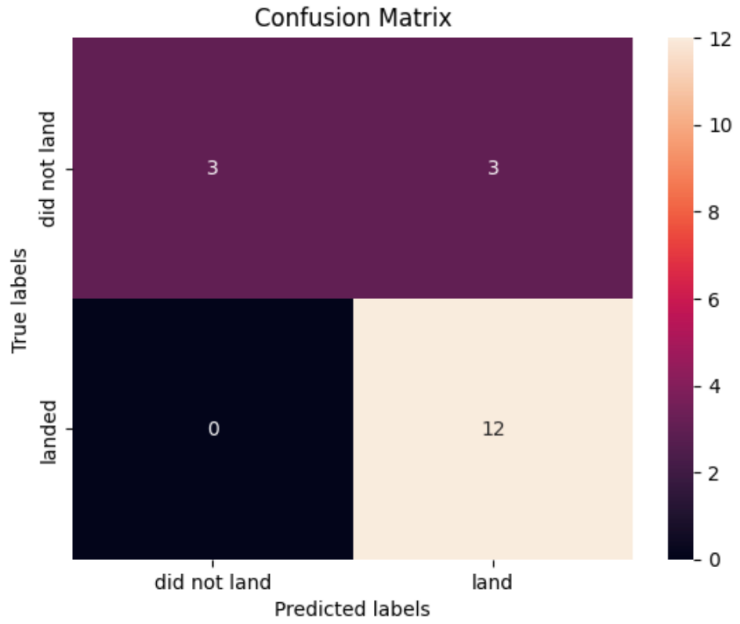
Predictive Analysis (Classification)

Classification Accuracy

Find the method performs best:

In [127…

```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



- The K Nearest Neighbors using logistic regression performed the best.

Confusion Matrix

- After using various techniques including Logistic Regression the following techniques did not give the best result:
  - Method Score
  - Grid Search
  - Object Tree_CV

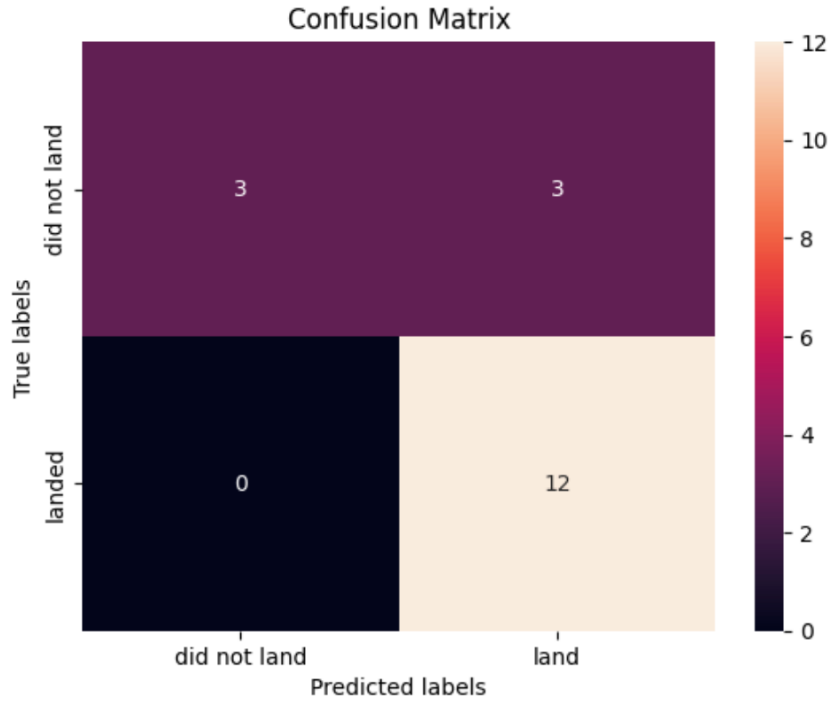
  However using the KNN with the Confusion Matrix method we can see the difference between
  - How many landed
  - How many did not land
  - The Predicted labels
  - The true labels.

  This gives us a great overview of all possibilities overlapping true and predicted results.

Find the method performs best:

In [127…

```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

An over view of general success methods used to get a general picture of the results are:

1. Site KSC LC-39A had the highest success rates of all the launch sites.
2. After 2015 success rates increased.
3. The landing success rates increased with the flight number.
4. Costal sites were the most favorable away from the city. This was assisted by testing rocket landings without much interference and easy to transport equipment quickly and efficiently.
5. The Machine Learning model was able to predict the landing outcome of 83.3% accuracy for the rocket launches.

Thank you!