IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Imisi J
22/11/2022

Outline

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

Executive Summary

This project is based on predicting whether the SpaceX Falcon 9 rocket first stage will launch successfully. The following tasks were carried out:

- Data Collection via SpaceX API & Wikipedia Page
- Exploratory Data Analysis – SQL, Visualisation using Seaborn and Folium Maps
- Visualisation using Plotly Dash
- Classification models for predictive analysis - Logistic Regression, SVM, K-Nearest Neighbours and Decision Tree
- Results from these models
- Conclusions

Introduction

Background

- 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.
- Through prediction of Falcon 9 first stage landing successfully, the cost of a launch can be determined
- This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Problems to find answers

- Understand which features impact the success of landing
- Select the best performing Machine Learning Algorithm to predict if the first stage will land successfully

Section 1

Methodology

Methodology

- Data collection methodology:
  - Data collected from SpaceX Wikipedia site and SpaceX Public API
- Perform data wrangling
  - Data processed so the landing outcomes could be classified as successful or unsuccessful
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was split into training & testing, normalised and tuned to retrieve the best hyperparameters via GridSearchCV for each classification model

Data Collection

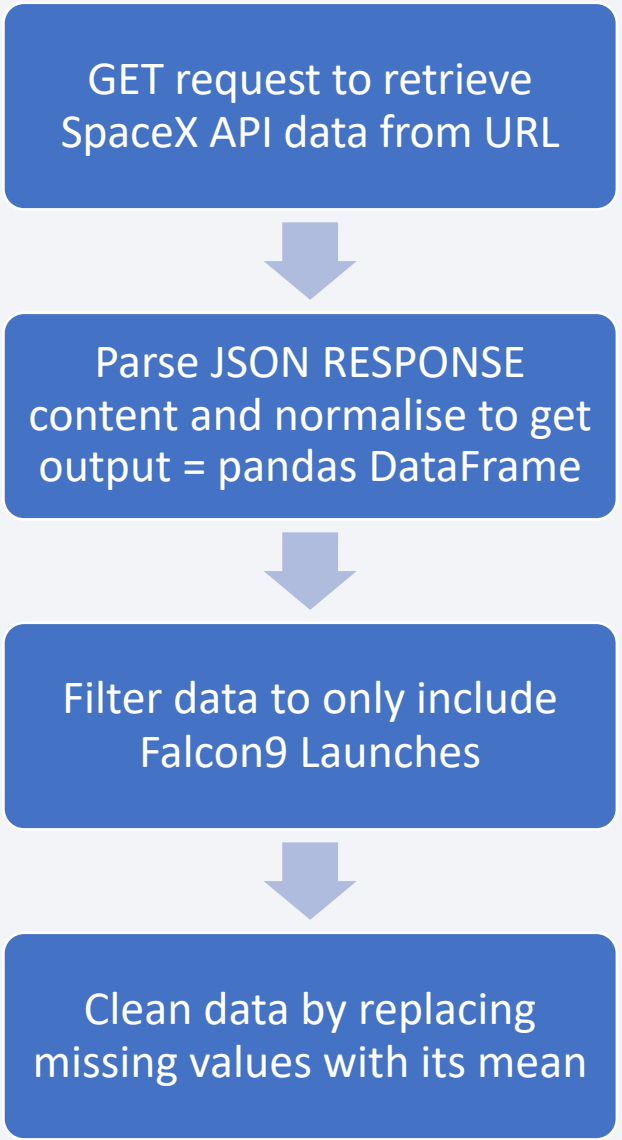
Data collected was completed through:

- SpaceX Public API Requests
- Space X Wikipedia entry via Webscraping

Data Collection – SpaceX API

- SpaceX REST API Flowchart

GitHub URL:
https://github.com/ij1214/IBM-Data-Science-Capstone/blob/7063ca1cf99338896d06adb4b552d38fe5a77a9d/Week1%20-%20Data%20Collection%20%E2%80%93%20SpaceX%20API.ipynb



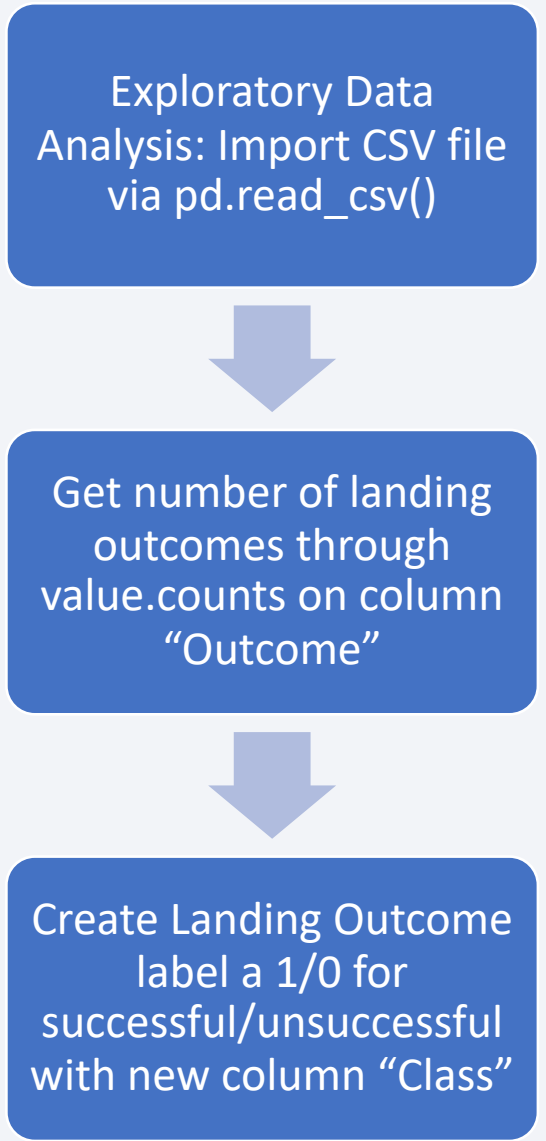
Data Collection - Scraping

- SpaceX Webscraping flowchart

Github URL:
https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week1%20-%20Data%20Collection%20-%20Scraping.ipynb

Data Wrangling

- Data processed flowchart
- GitHub URL: https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week1%20-%20Data%20Wrangling.ipynb



EDA with Data Visualization

Charts plotted

- Flight Number vs Payload Mass
- Flight Number vs Launch Site
- Payload vs Launch Site
- Success rate of each Orbit type
- Flight Number vs Orbit type
- Payload vs Orbit type
- Launch Success Yearly Trend

Reason

- Charts above plotted to understand relationship between the various input features and their outcome in landing success
- Mixture of scatter, bar and line plots to clearly visualise the relationship

GitHub URL : https://github.com/ij1214/IBM-Data-Science-Capstone/blob/2034a2778ad12d031bc2dc865e85ed9c8304722b/Week2%20-%20EDA%20Visualisation.ipynb

EDA with SQL

SQL Queries performed:

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites begin with the string 'CCA’
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved.
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster_versions which have carried the maximum payload mass using a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Ranked the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

GitHub URL: https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week2%20-%20EDA%20Visualisation.ipynb

Build an Interactive Map with Folium

Map markers added to folium map:

- Launch sites on a map as markers and circles
- Success/failed launches for each site on the map as clusters, colour coded
- Distances between a launch site to its proximities as lines

Reason

- Understand the areas of high success rates for launches
- Understand if there's a pattern with the successful launches and its proximities eg coastal line, railway point etc

GitHub URL: https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week3%20-%20EDA%20-%20Folium.ipynb

Build a Dashboard with Plotly Dash

Dashboard contains

- **Pie chart**
  - To show the launch success counts
- **Scatter chart**
  - To show the launch outcome dependent on the launch site(s) and the payload mass
- **Drop down menu interaction**
  - To select the launch site and view their pie & scatter charts
- **Range slider**
  - To view the impact of changing the payload mass for each site on the scatter chart

GitHub URL: https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week3%20-%20Plotly%20Dash.py

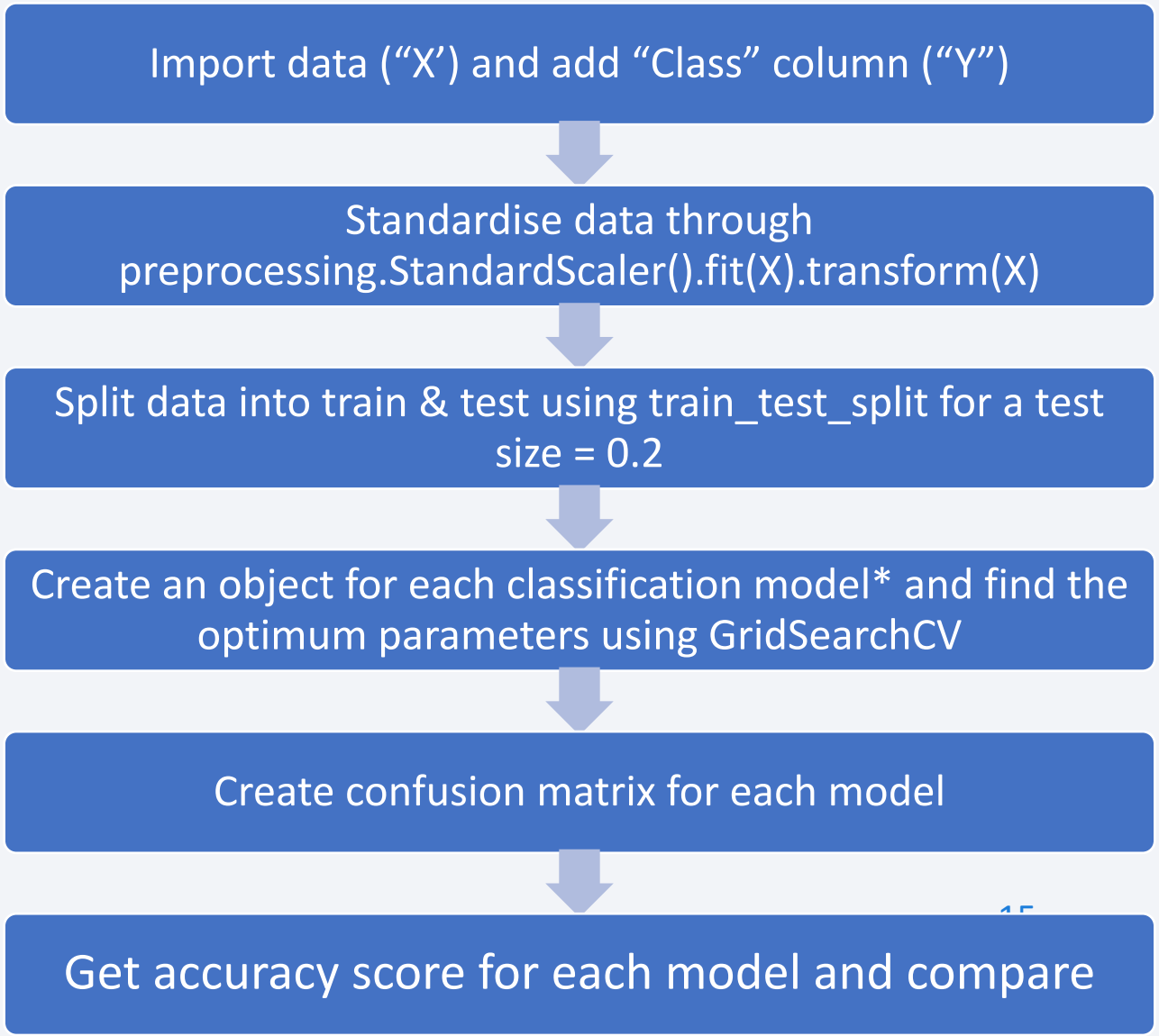
Predictive Analysis (Classification)

Classification model summary

* Models used are:

- Logistic Regression
- SVM
- Decision Tree
- K-Nearest Neighbours

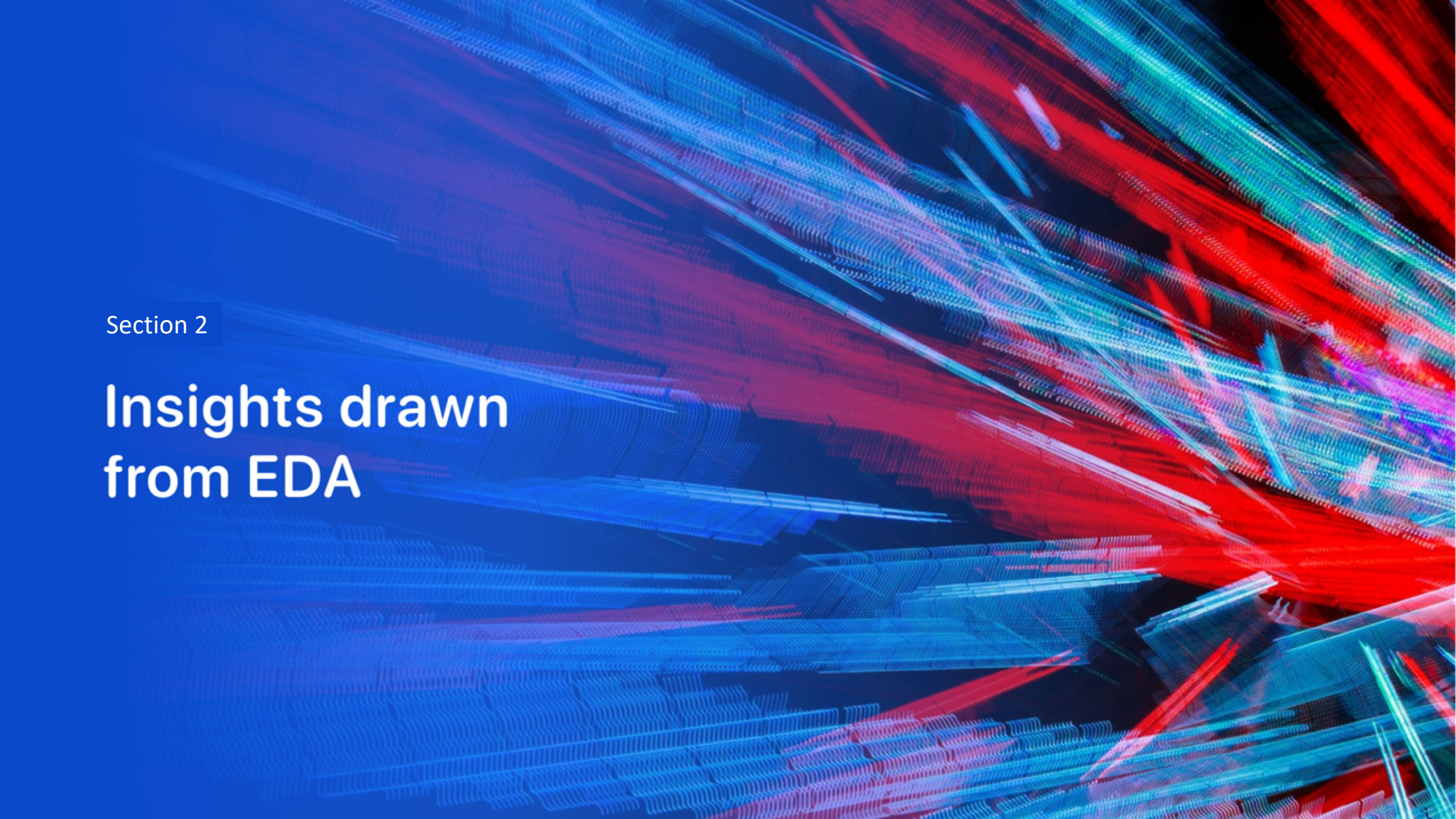
GitHub URL:
https://github.com/ij1214/IBM-Data-Science-Capstone/blob/main/Week4%20-%20Predictive%20Analysis.ipynb



Results

The following slides will show the following results:

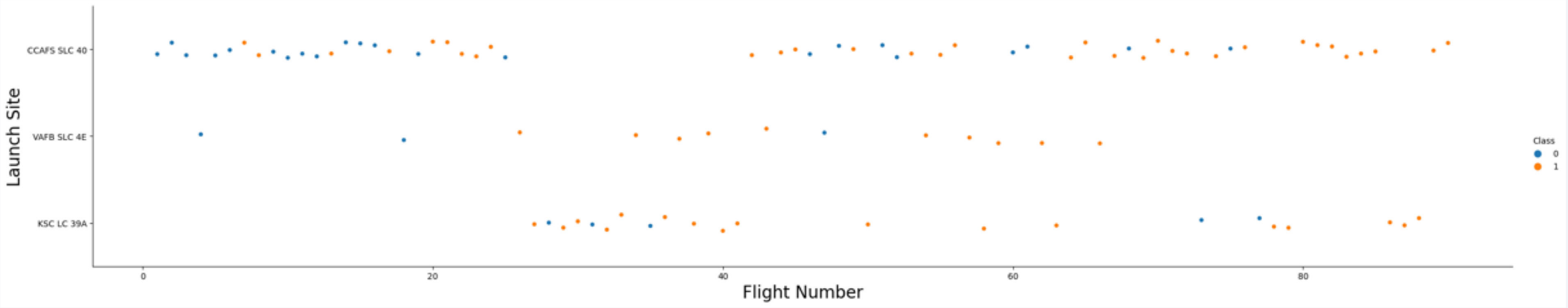
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

EDA VISUALISATION

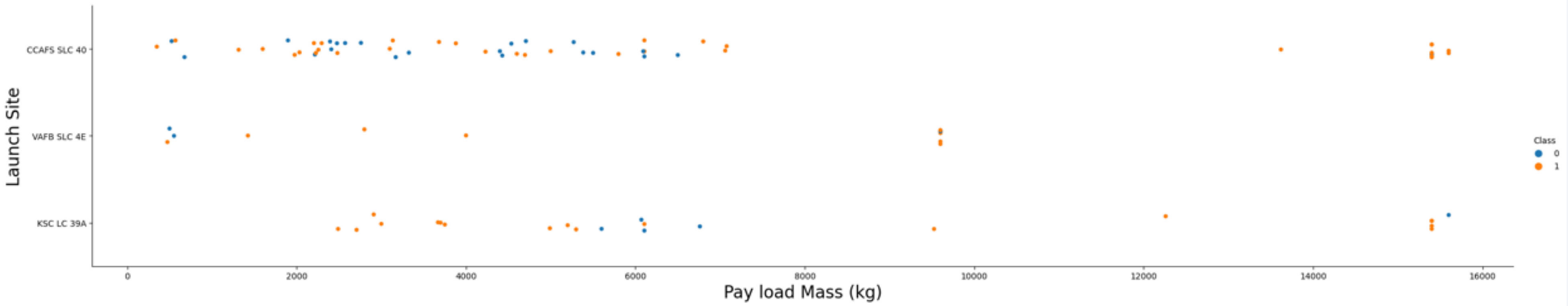
ANALYSIS USING CHARTS

Flight Number vs. Launch Site



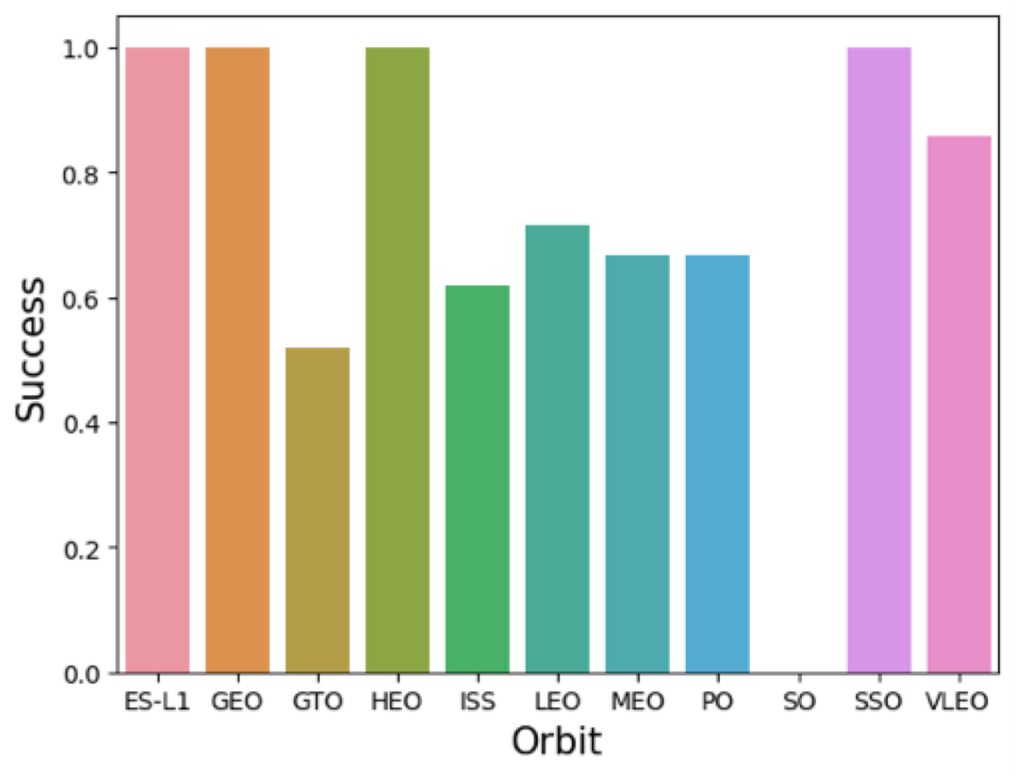
- Scatter chart shows a greater number of success for each site as the flight number increases
- Launch Site **CCAFS SLC 40** had the largest number of flights

Payload vs. Launch Site



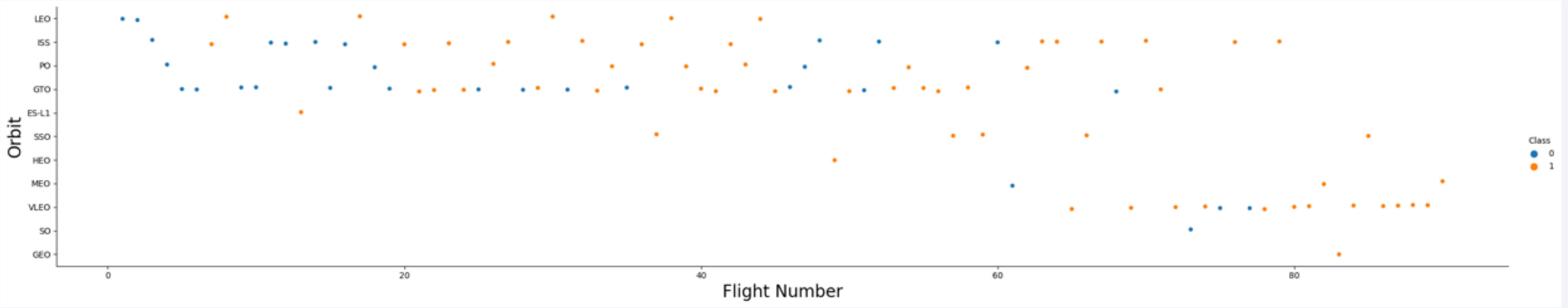
- Scatter chart shows the success rate increases for a payload of around 9000kg
  - VAFB-SLC launch site had no launches for a payload mass of >10,000kg

Success Rate vs. Orbit Type



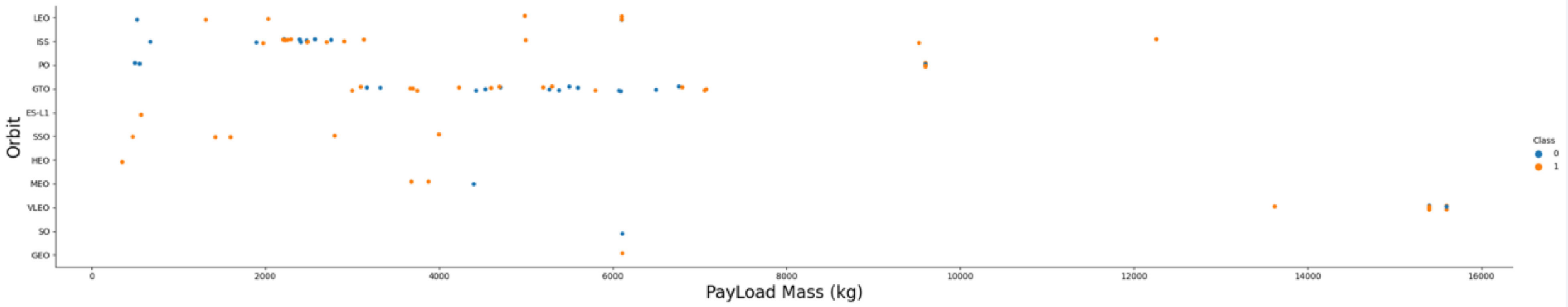
- Bar chart shows Orbit type SO had no success
- ES-L1, GEO, HEO, SSO and VLEO all had a 100% success rate

Flight Number vs. Orbit Type



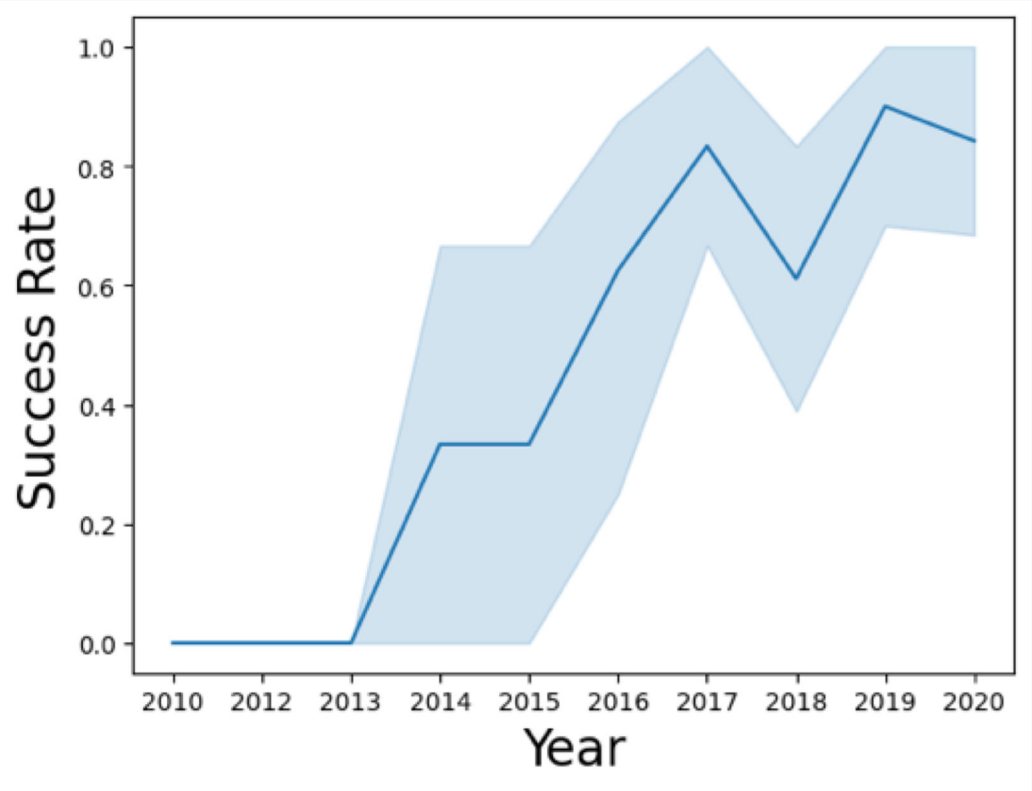
- Scatter chart shows the majority of success occurred around 65+ flights
- Orbit types used changed with increasing flight numbers, as VLEO was mostly used from flight number 65+

Payload vs. Orbit Type



- Scatter plot shows the success rate is higher for greater payloads for PO, LEO and ISS.
  - No clear effect for payload mass for GTO
  - SEO and GEO have much less orbits than the rest
  - The majority of launches were below 8000kg

Launch Success Yearly Trend



The success rate increased from 2013 till 2020

No success before 2013

EDA USING DATABASE (SQL)

ANALYSIS USING SQL

All Launch Site Names

```
%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

SELECT DISTINCT to only get unique names

There are 4 Launch Sites

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE "CCA%" LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

o expand output; double click to hide output

LIMIT 5 to only retrieve 5 records

LIKE clause to have only those starting with CCA

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as total_payload_mass FROM SPACEXTBL WHERE Customer="NASA (CRS)"
```

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

total_payload_mass
45596

SUM clause used to retrieve total payload mass

WHERE clause to only retrieve for NASA (CRS)

Total Payload Mass = **45596**

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as avg_payload_mass FROM SPACEXTBL WHERE Booster_Version LIKE "F9 v1.1%"
```

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

avg_payload_mass
2534.6666666666665

AVG clause used to retrieve Average

WHERE clause to only retrieve Booster Version F9 V1.1 and its similarities using LIKE clause

Average payload mass carried by booster version F9 v1.1 = 2534.67kg (rounded to 2dp)

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) as first_success_groundpad FROM SPACEXTBL WHERE `Landing _Outcome`="Success (ground pad)"
```

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

first_success_groundpad
22-12-2015

MIN clause used to retrieve earliest date

WHERE clause to only retrieve success for Ground Pad

First successful Ground landing date is **22-12-2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Booster_Version from SPACEXTBL WHERE (Mission_Outcome LIKE "Success%") AND (PAYLOAD_MASS__KG_ B
```

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

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

DISTINCT clause used to get unique booster version names

WHERE clause to only retrieve successful landing for drone ship

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT DISTINCT Mission_Outcome, COUNT(Mission_Outcome) AS total from SPACEXTBL WHERE Mission_Outcome IN ("Succ
```

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

Mission_Outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

COUNT to sum total number of success & failures for each mission outcome

GROUP BY used for Mission outcome due to aggregation used (COUNT)

Only 1 failure

Boosters Carried Maximum Payload

```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ in
(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDER BY Booster_Version
```

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

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

Subquery contains selection of Booster Versions with the maximum payload

12 Booster Versions found

2015 Launch Records

```
%sql SELECT SUBSTR(DATE,4,2) as Month,`Landing _Outcome`, Booster_Version, Launch_Site from SPACEXTBL
WHERE SUBSTR(DATE,7,4)='2015' AND `Landing _Outcome`='Failure (drone ship)'
```

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

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

NB MONTHNAME is not supported here as using SQLLITE not IBM SQL, hence month number as used

2 Boosters Found for Months January & April with Failures, both on Launch Site CCAFS LC-40

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

```
%sql SELECT `Landing _Outcome`, COUNT(`Landing _Outcome`) as count_success from SPACEXTBL
WHERE `Landing _Outcome` LIKE ("Success%") AND DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY `Landing _Outcome`
ORDER BY count_success DESC
```

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

Landing _Outcome	count_success
Success (ground pad)	5
Success (drone ship)	5

COUNT clause to sum the number of successful landing outcomes

The total success was 10 from ground pad & drone ship

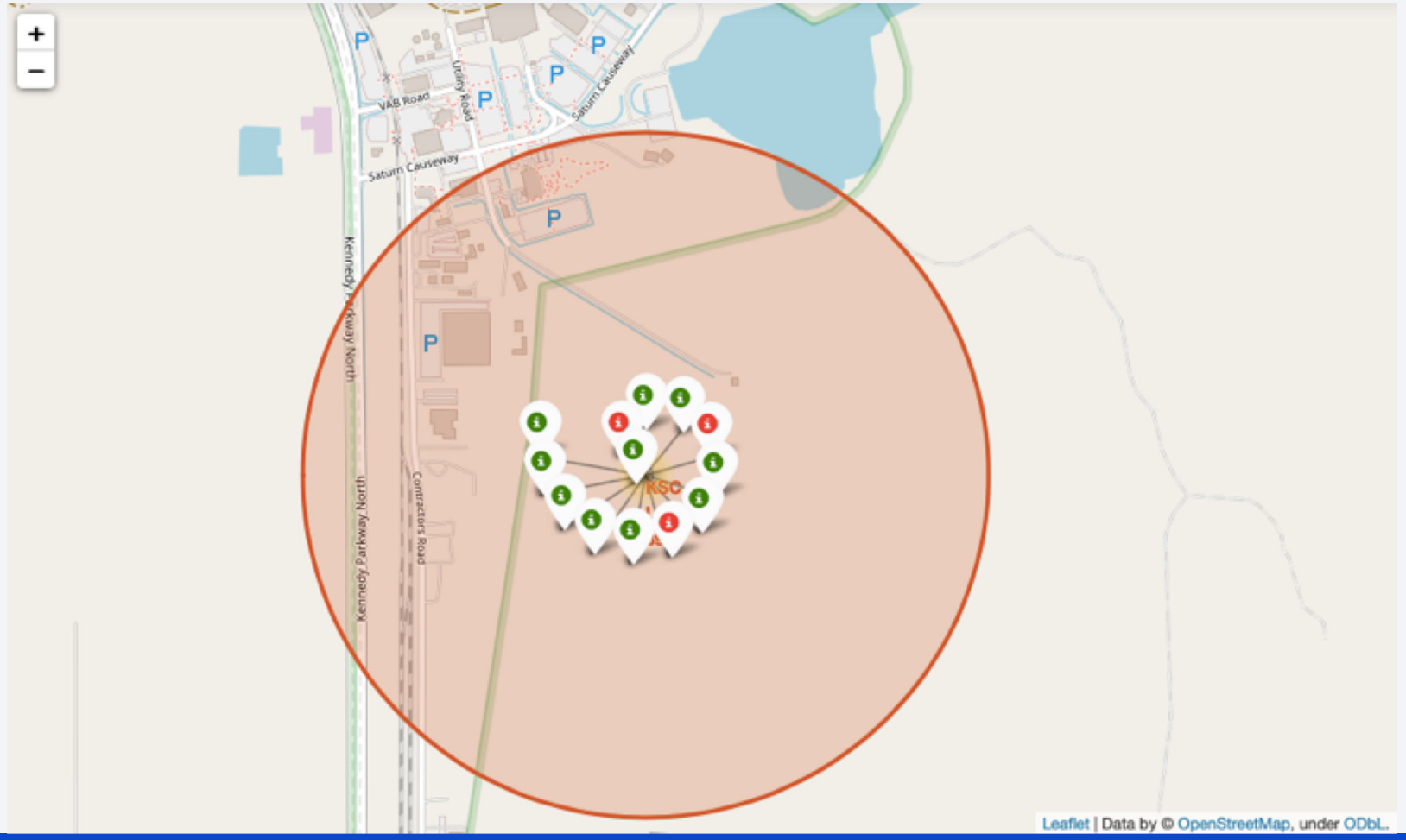
Section 3

Launch Sites Proximities Analysis

Launch Sites Locations

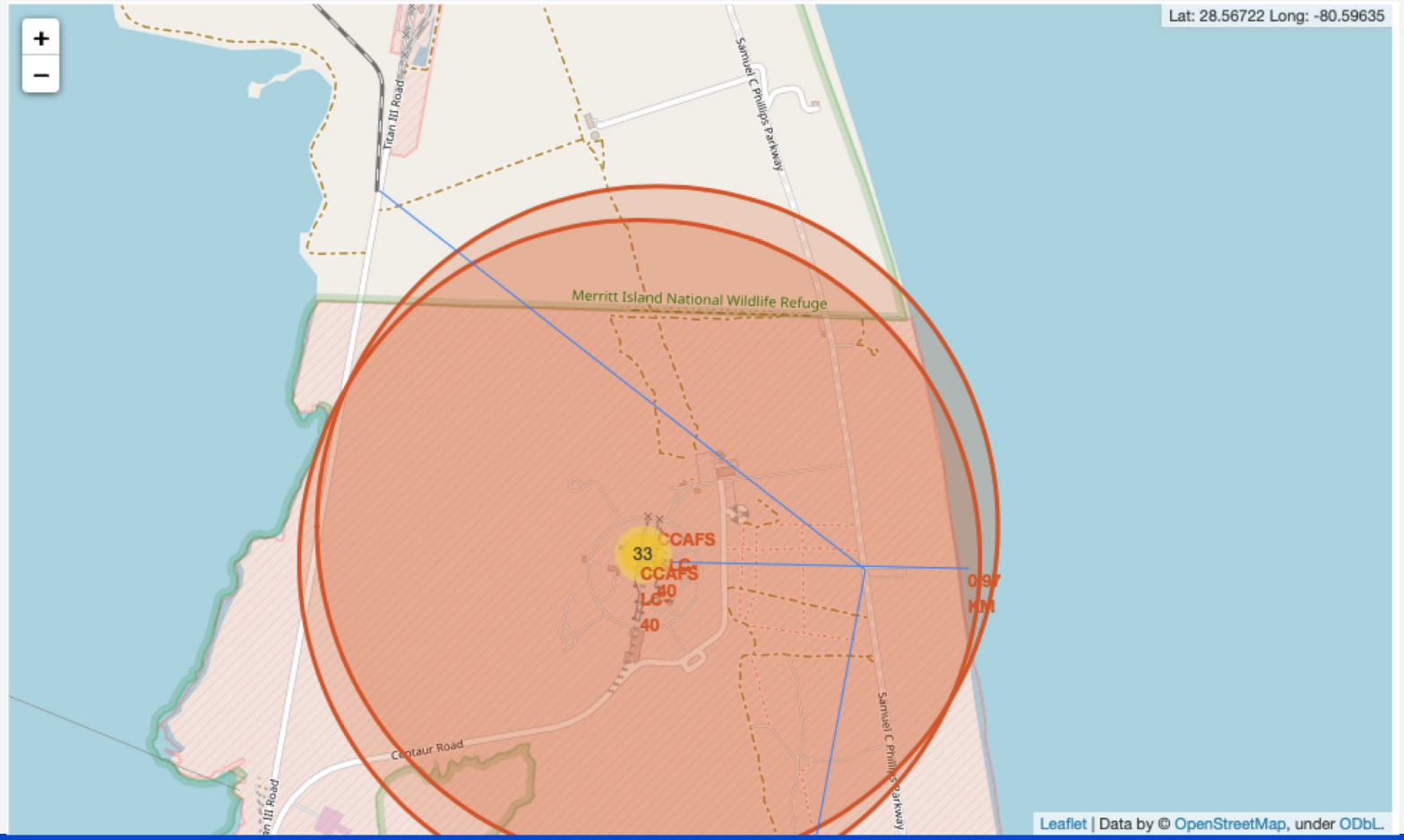
All launch sites are near an ocean

Colour-Labelled Launch Outcomes

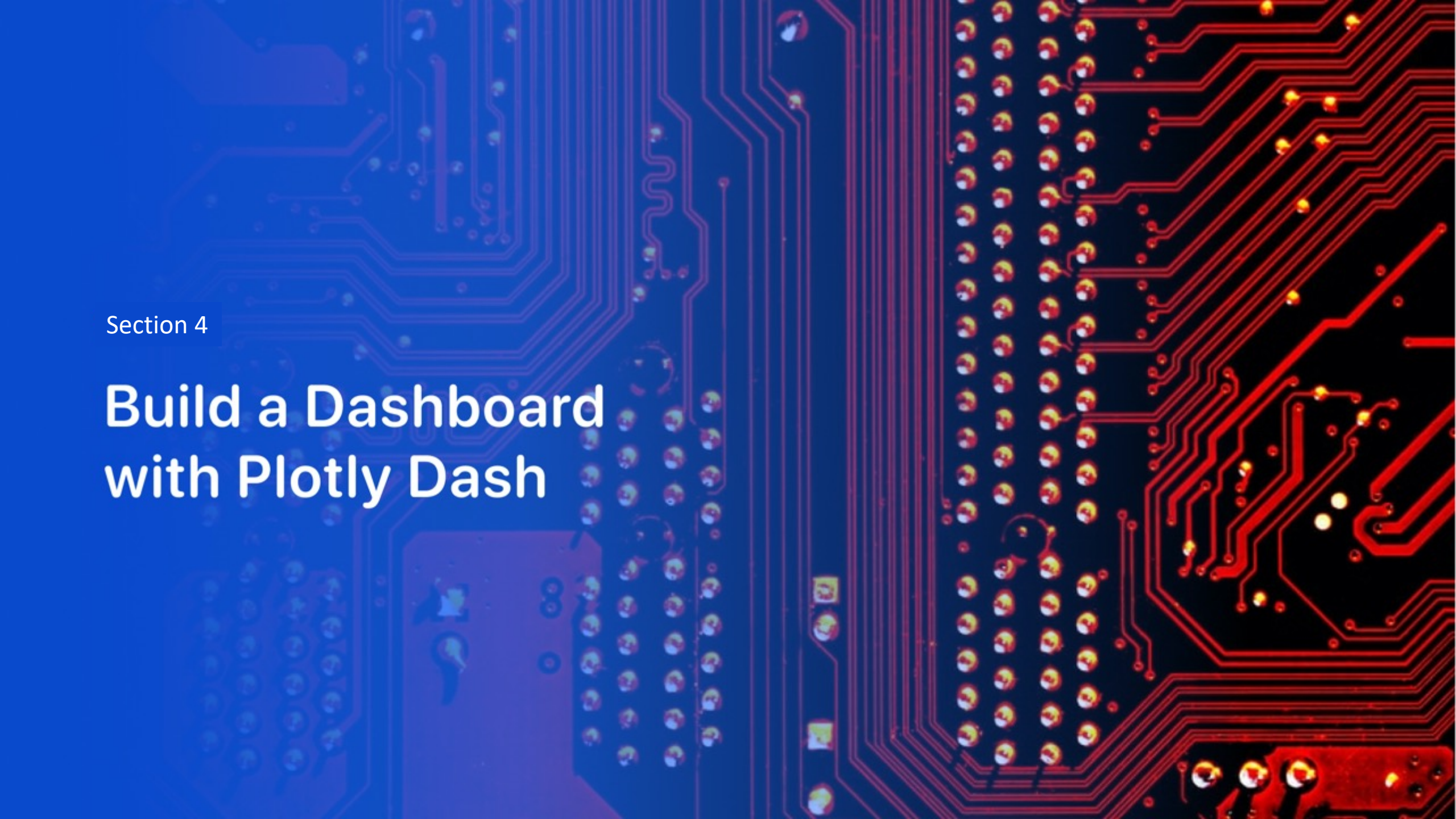


Failed (red) and Successful (green) launches shown in this cluster on the Folium map for Launch Site KSC LC-39A

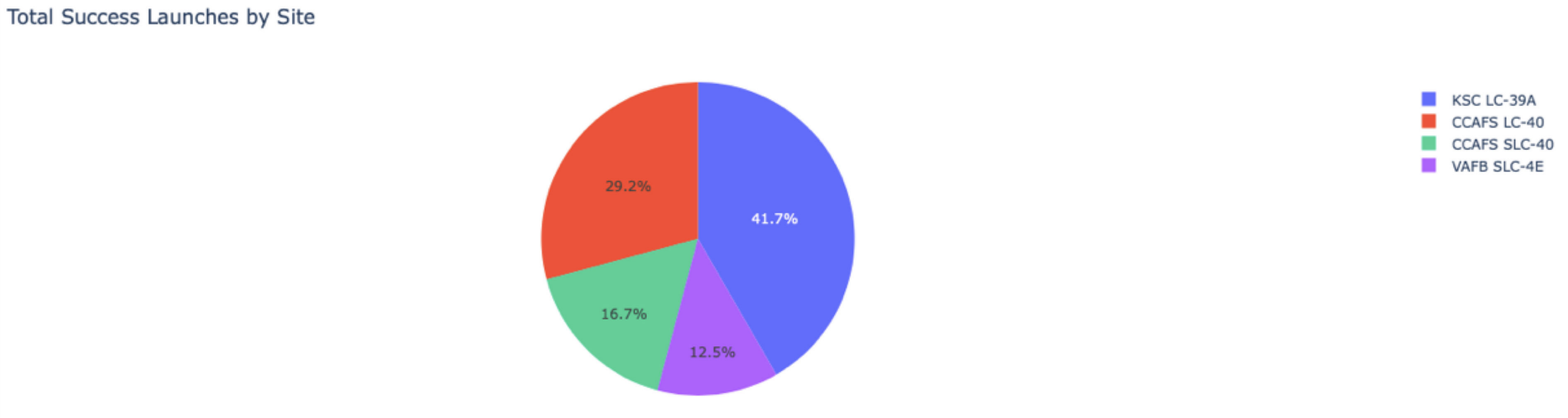
Launch site proximities



The closest proximity was the coastline, followed by the railway track and then the city. This would be to ensure the launches would not cause any damage and affect inhabitants in nearby cities.

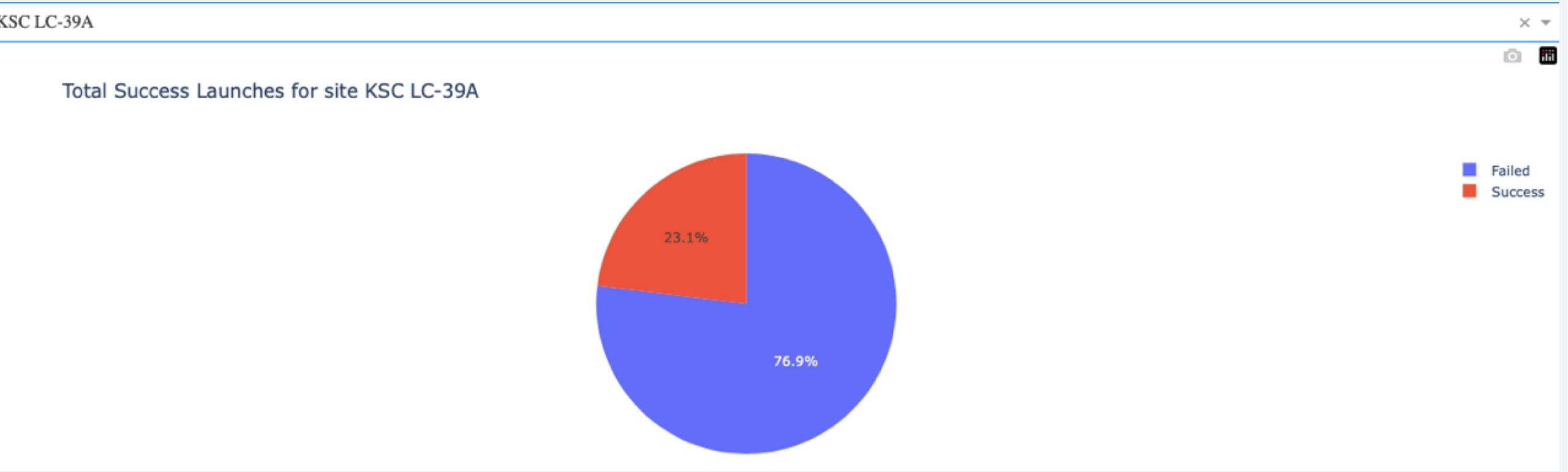

Section 4
Build a Dashboard with Plotly Dash

Launch Success across all launch sites



KSC LC-39A Had the highest success launch rate across all sites, with VAFB SLC-4E having the least

Most successful launch site



KSC LC-39A had a success rate of 23.1% out of all its launches, which equates to 3 successful launches

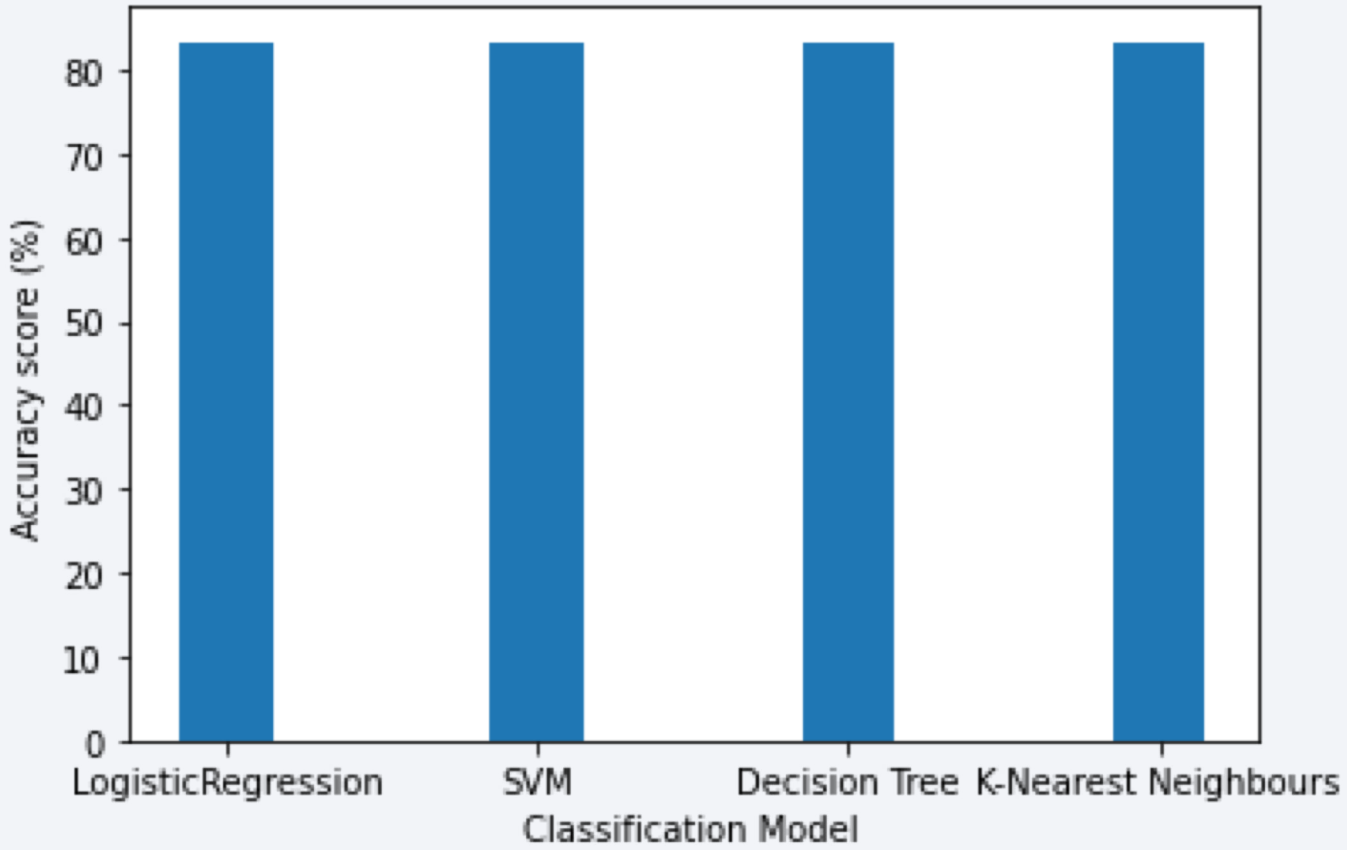
Payload vs. Launch Outcome

Scatter plot shows booster version FT had the highest success rate. However, when the payload range is from 5000kg+, there are much less launches - only 3 were successful. The booster versions that are successful from 5000kg+ are FT & B4

Section 5

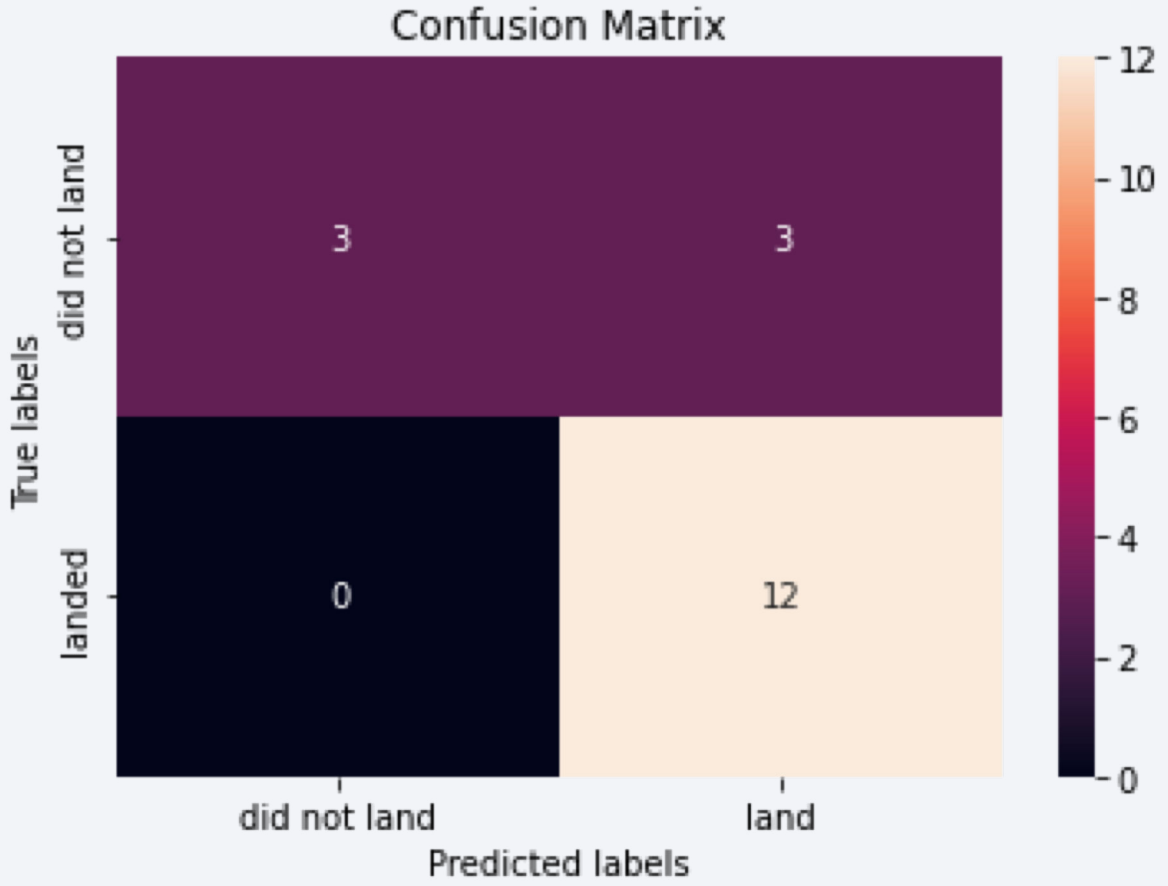
Predictive Analysis (Classification)

Classification Accuracy



All classification models have the same accuracy, hence more data is required to determine the best model to use

Confusion Matrix



Confusion matrix is the same for all classification models as the accuracy score is the same. We can see the models had a 100% correct prediction for landed labels but 50% correct prediction for did not land labels, as there are 3 False Positive results

Conclusions

- Launches started to have success rate of >0 from 2013
- Greater success rate for each site as the flight number increases
- Launch Site **CCAFS SLC 40** had the largest number of flights
- Launch site **KSC LC-39A** had the highest success launch rate across all sites
- Launches are all done in close proximity to the coast
- The success rate is higher for payloads <5000kg
- All classification models analysed had the same accuracy score of 83.33%

Appendix

- GitHub repo link for all codes, Jupyter notebooks etc: https://github.com/ij1214/Data-Science-Capstone

Thank you!