IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Juan Miguel Gutiérrez García

31/12/2023

Outline

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

Executive Summary

- Summary of methodologies
    - Data collection
    - Data wrangling
    - Exploratory Data Analysis with Data Visualization
    - Exploratory Data Analysis with SQL
    - Building an interactive map with Folium
    - Building a Dashboard with Plotly Dash
    - Predictive analysis (Classification
- Summary of all results
    - Exploratory Data Analysis results
    - Interactive analytics demo in screenshots
    - Predictive analysis results

Introduction

- Project background and context

SpaceX is asuccessful company that advertises Falcon

9 rocket launches on its website, with a cost of 62 million dollars;

other 165 million dollars each. Much of this savings comes from reusing the first stage. We will predict if SpaceX will reuse the first stage.

- Problems you want to find answers

How do variables affect the success of the first stage?

How the rate of landings ichange over the years?

What is the best algorithm for binary classification?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using Space X Rest API
  - Using Web Scrapping
- Perform data wrangling
  - Filter the data
  - Deal with missing values and preparation of data to a binary classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning and evaluating classification models for best possible result.

Data Collection

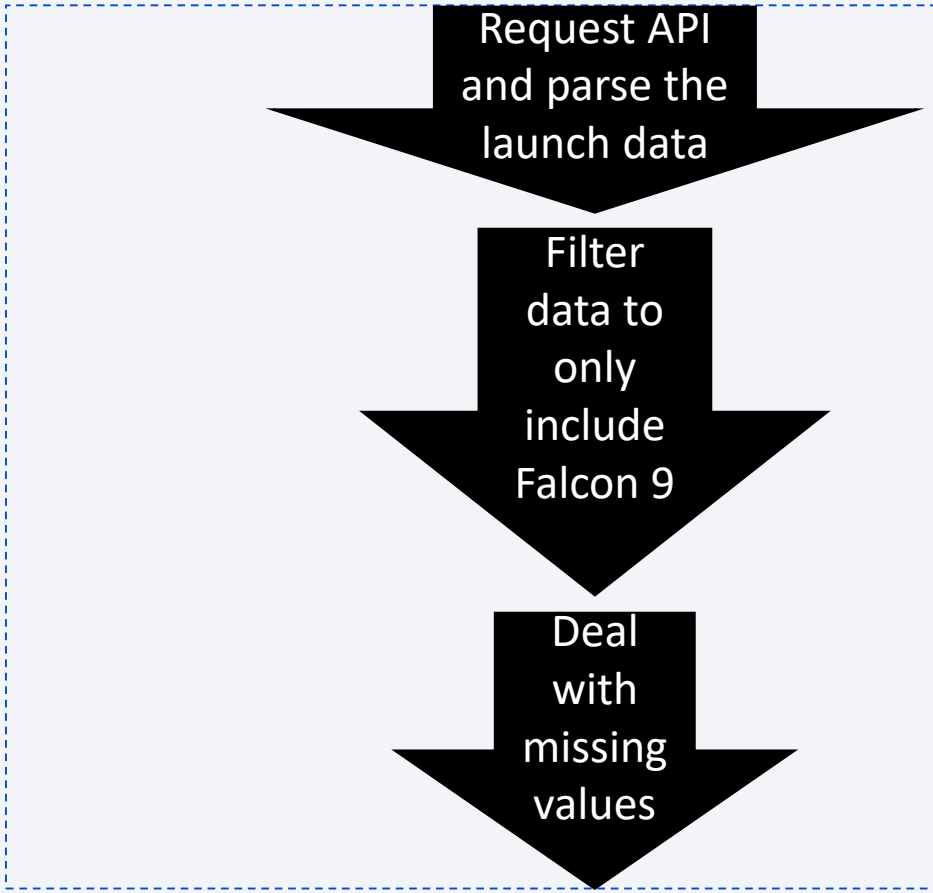
To obtain the Data Columns it was achieved by using SpaceX REST API:

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

We also used Wikipedia Web Scraping for: Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time.

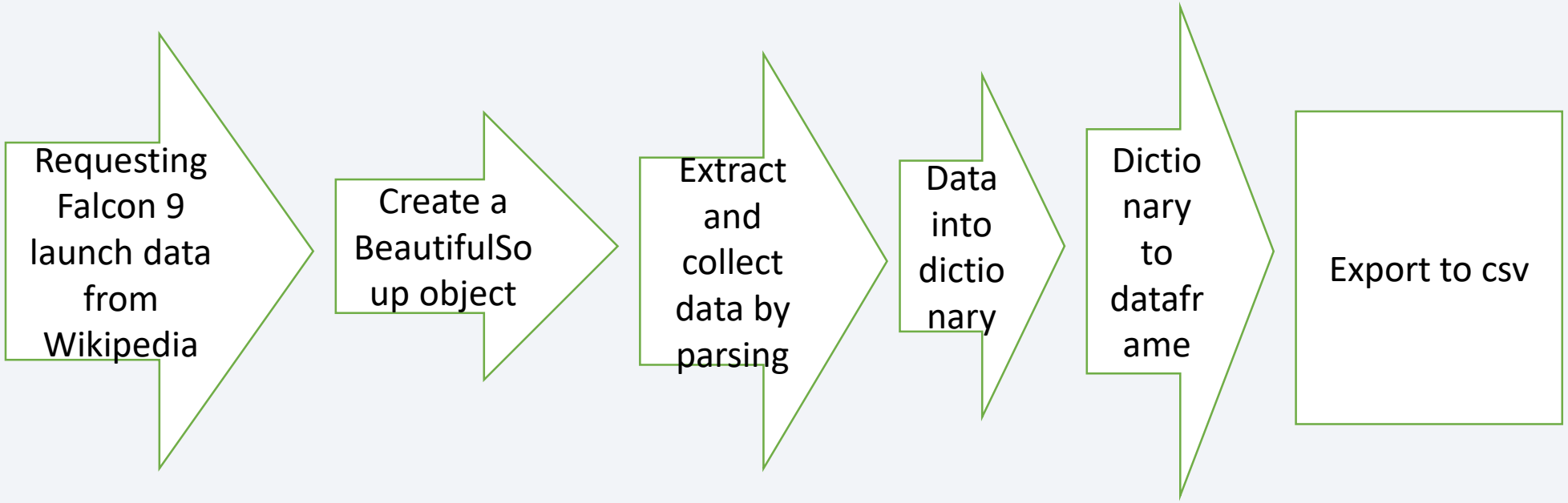
Data Collection - SpaceX API

- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/data-collection-api.ipynb



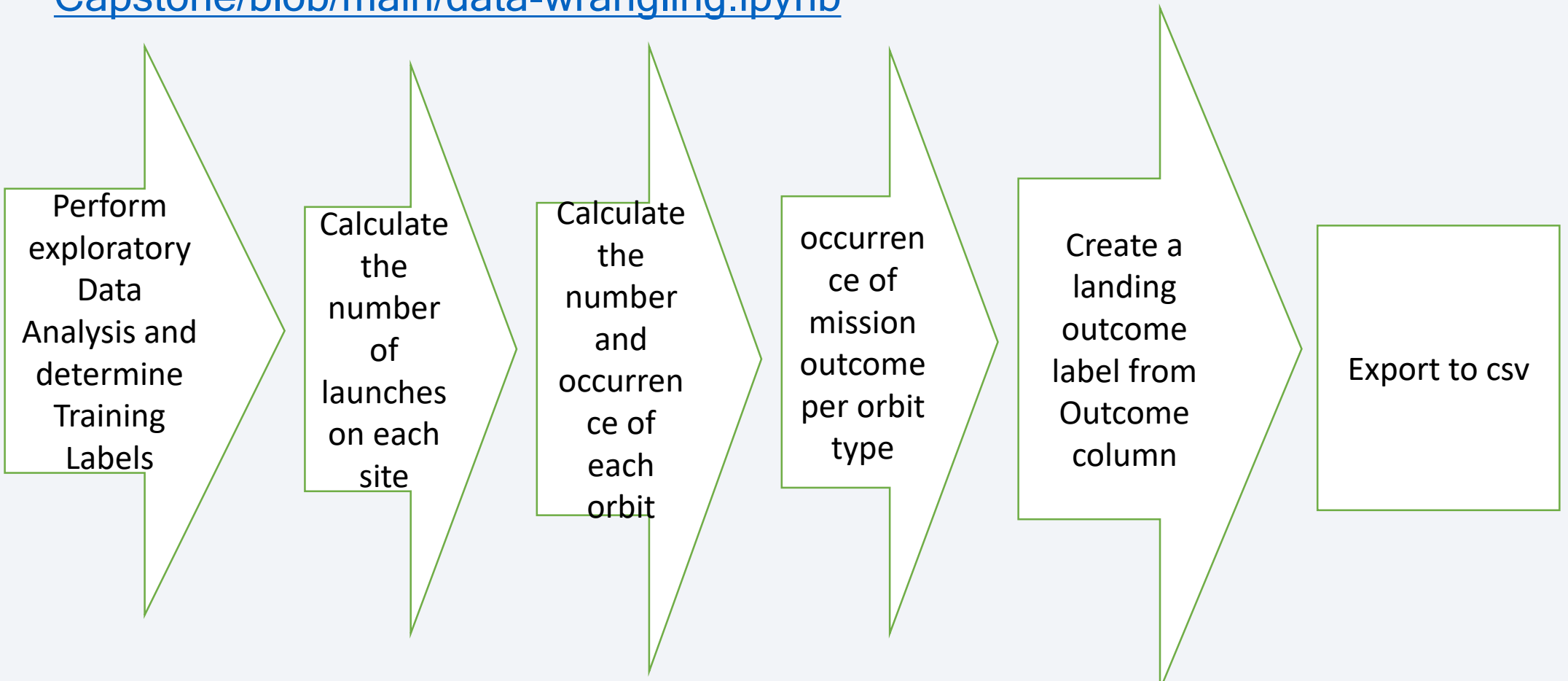
Data Collection - Scraping

- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/data-collection-webscraping.ipynb



Data Wrangling

- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/data-wrangling.ipynb



EDA with Data Visualization

- Charts that were plotted

Scatter plot: shows relationship between variables. If a relationship exists, they could be used in machine learning model.

Bar chart: show comparisons among discrete categories. Can show the relationship between the specific categories being compared and a measured value.

Line chart: shows trend over time.

- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/eda-dataviz.ipynb

EDA with SQL

- Names of unique launch sites
- Top 5 launch sites begin with the string ‘CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome
- Names of boosters with success in drone ship and 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
- Failed landing outcomes in drone ship, booster versions and launch site for the months in year 2015
- Ranking the count of landing outcomes
- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/eda-sql.ipynb

Build an Interactive Map with Folium

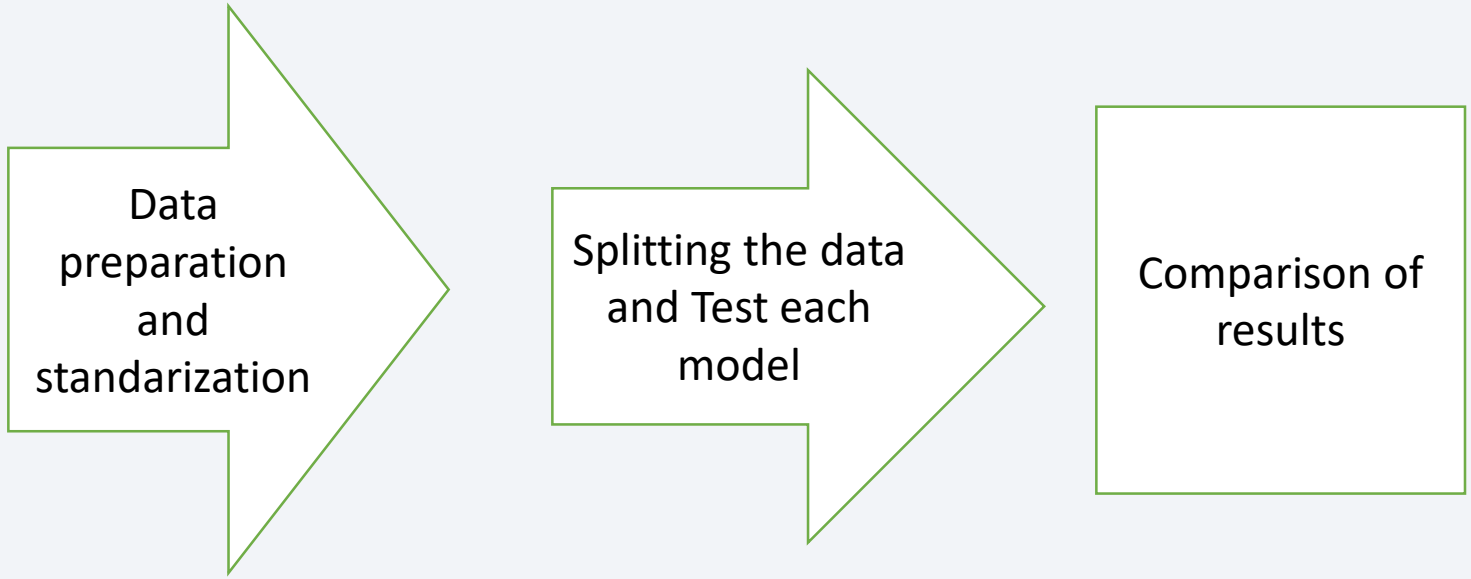
- Markers of all Launch Sites:
- Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center.
- Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.
- Added coloured Markers of success and failed using Marker Cluster to identify high success launches rates.
- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/interactive-map-with-Folium.ipynb

Build a Dashboard with Plotly Dash

- Dropdown list to enable Launch Site selection.
- Pie Chart showing Success Launches
- Added a scatter chart to show the correlation between Payload and Launch Success.
- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

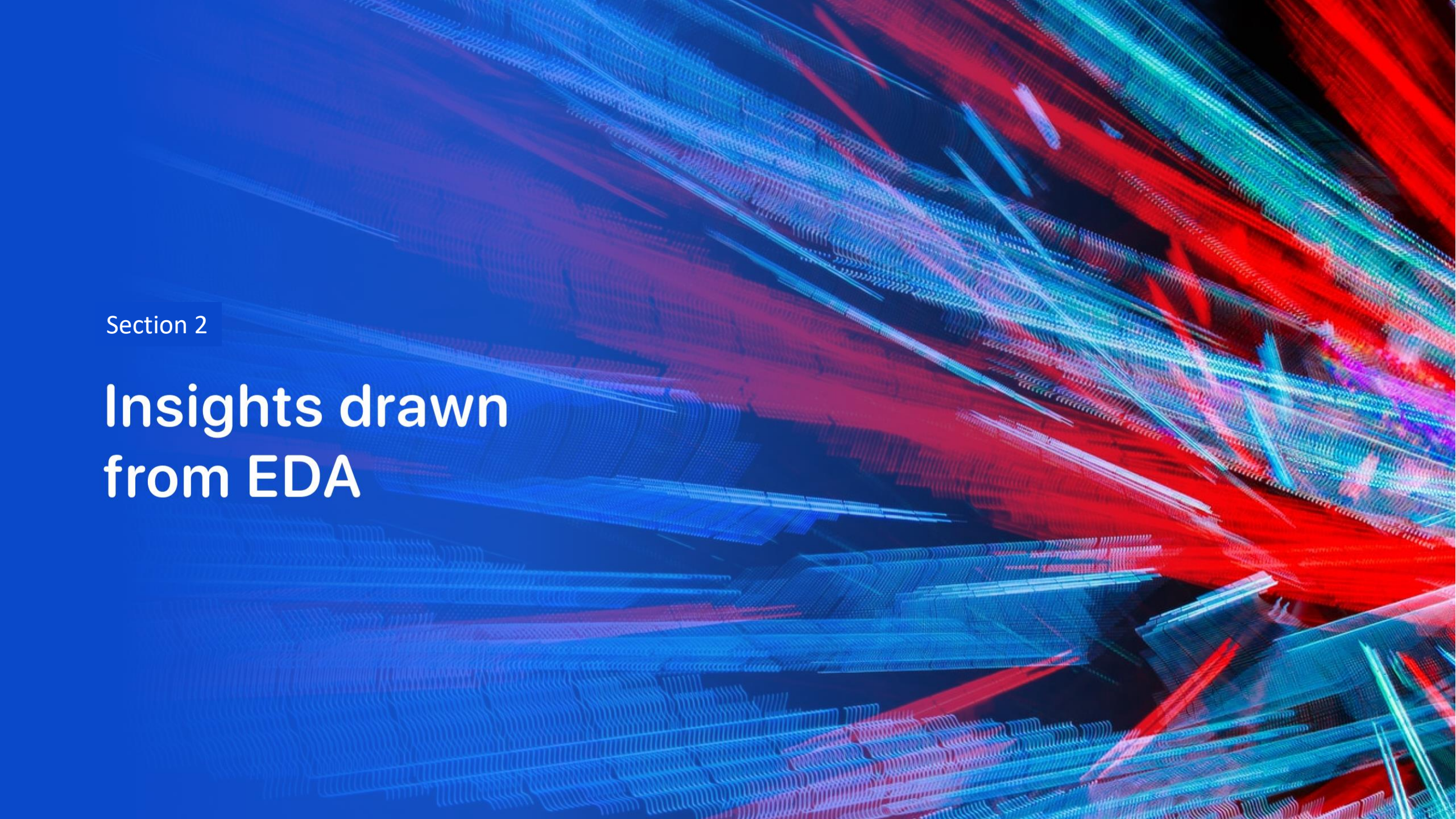
- Four classification models were compared: regression, support vector machine, decision tree and nearest neighbors.



- Github URL: https://github.com/juan-miguel-1989/IBM-Data-Science-Capstone/blob/main/Machine-Learning-Prediction.ipynb

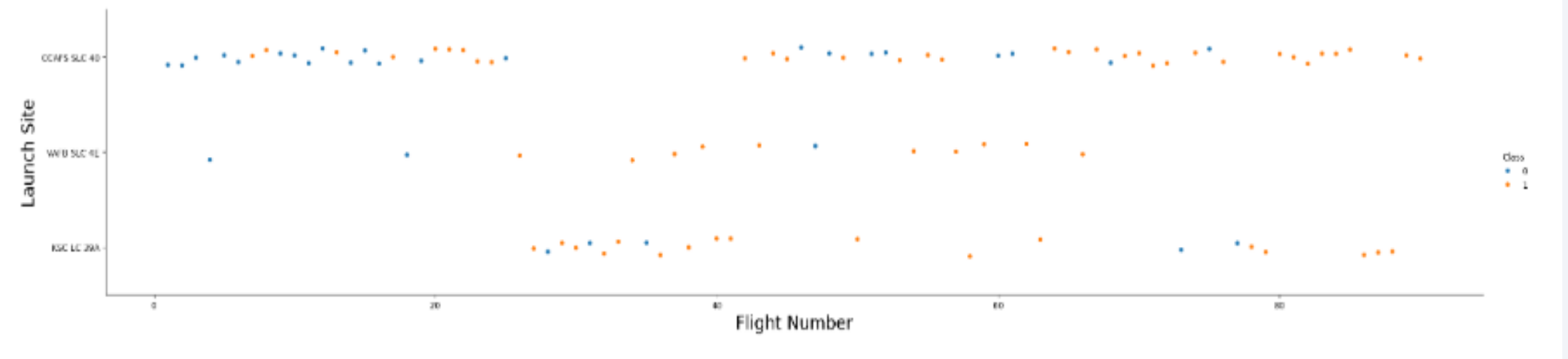
Results

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


Section 2
Insights drawn from EDA

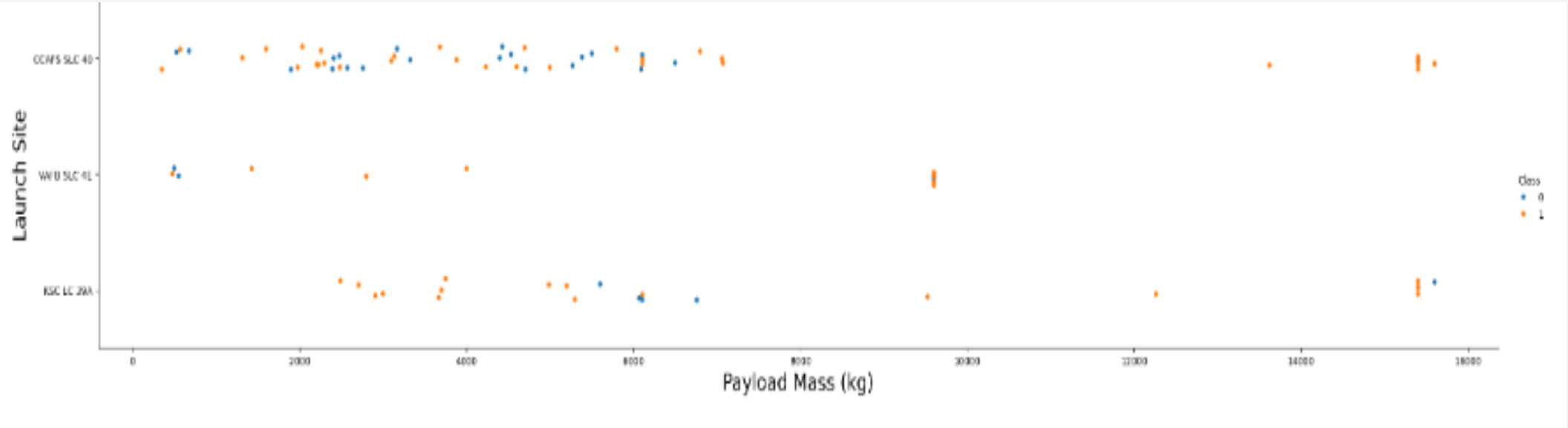
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site



- The CCAFS SLC 40 is the best launch stie.
- VAFB SLC 4E and KSC LC 39A have second and third success rates.
- It can be assumed that each new launch has a higher rate of success.

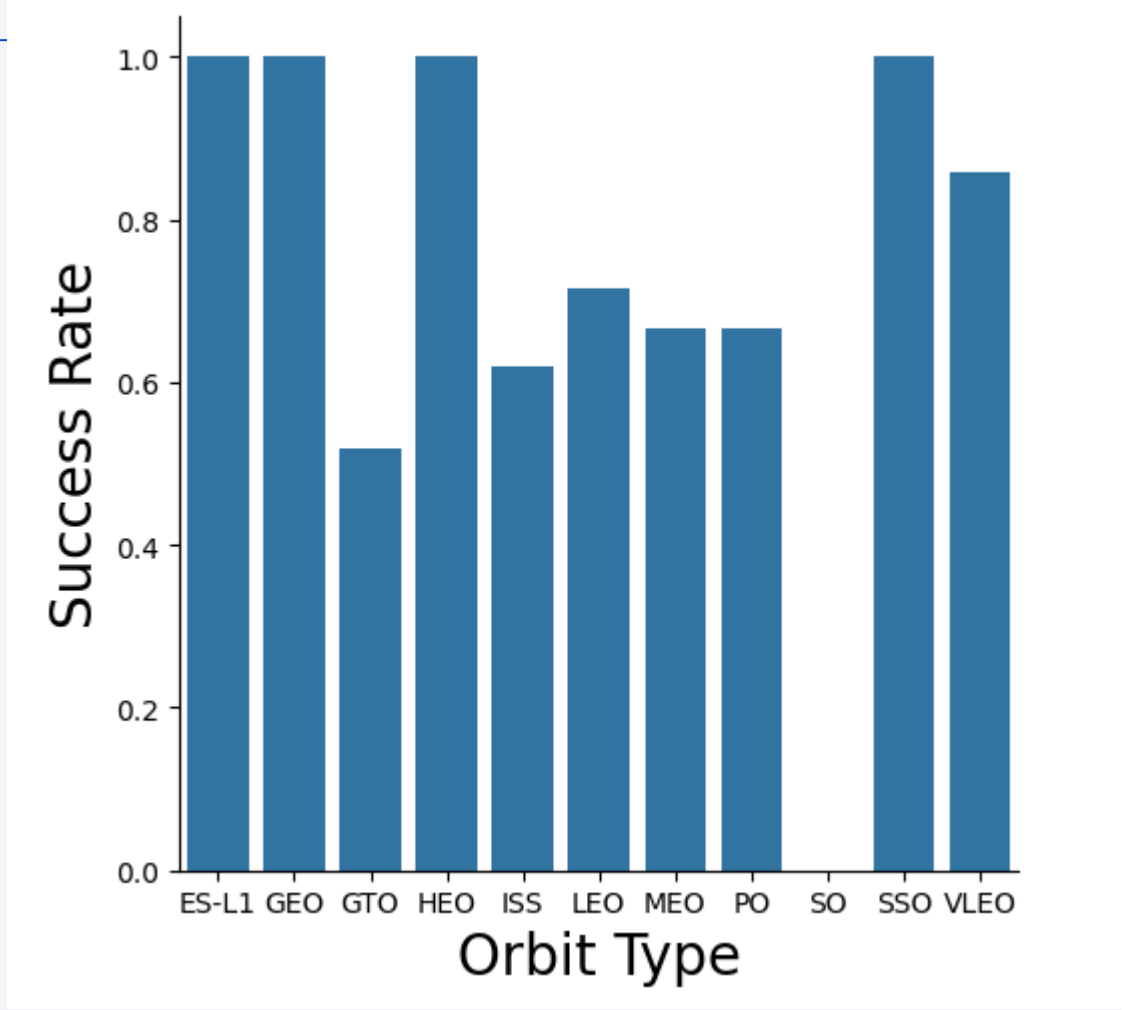
Payload vs. Launch Site



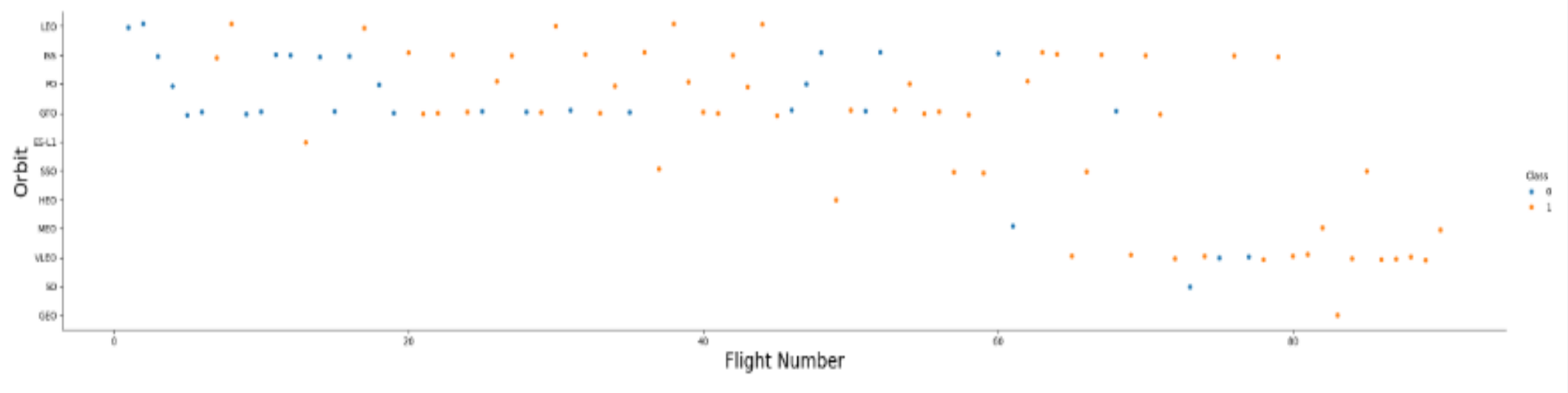
- For every launch site the higher the payload, the higher the success rate.
- launches with payload mass over 7000 kg were successful.

Success Rate vs. Orbit Type

- Worst success rates:
  - SO
- Biggest success rates:
  - ES-L1
  - GEO
  - HEO
  - SSO

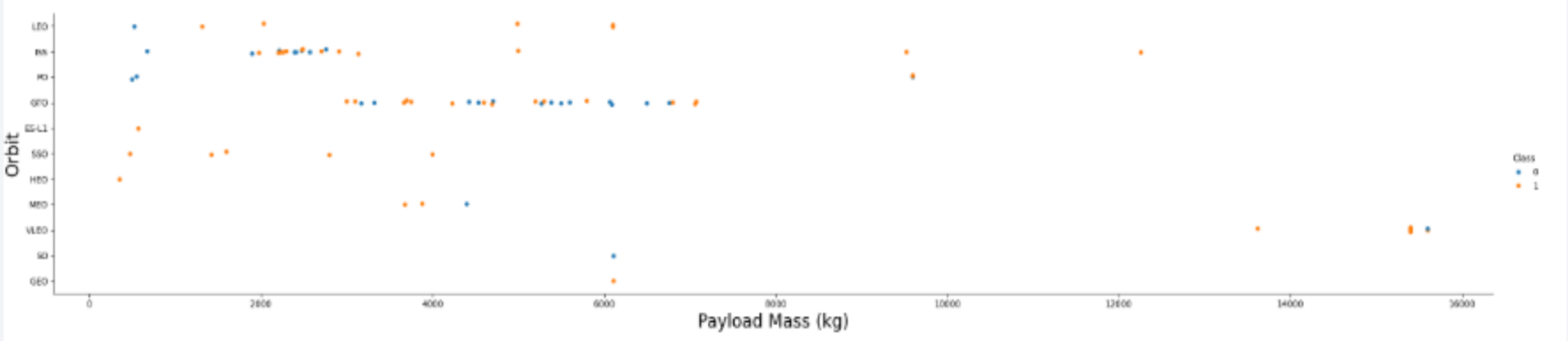


Flight Number vs. Orbit Type



- In LEO, Success seems to be related to the number of flights
- Success rate improve over time

Payload vs. Orbit Type



- Heavy payload seems to have a negative influence on GTO and positive on GTO and Polar LEO.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

Names of the unique launch sites

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'

- Find 5 records where launch sites begin 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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

45596 kg

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

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

340.4 kg

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.0%';
```

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

AVG(PAYLOAD_MASS__KG_)
340.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

22/12/2015

In [22]:
```
%%sql
SELECT MIN(Date)
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
```

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

Out[22]:

MIN(Date)
2015-12-22

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

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

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Success 99

Failure 1

Present your query result with a short explanation here

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

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

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

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

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

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

2015 Launch Records

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

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

- 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 TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

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

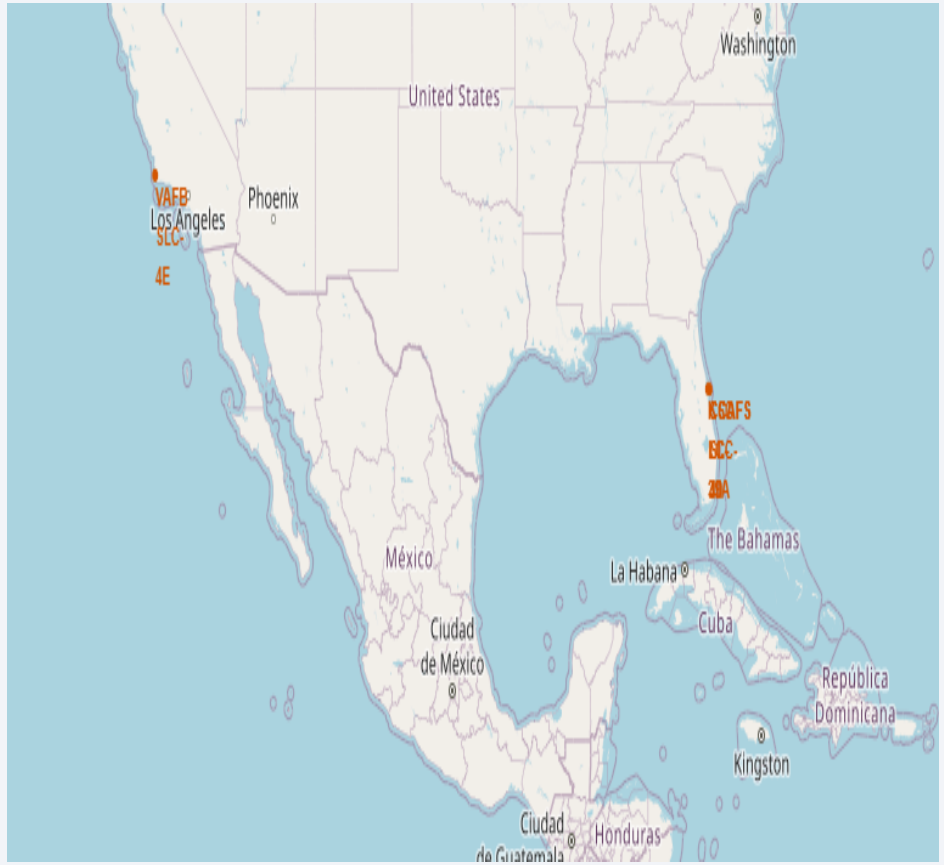
Landing_Outcome	TOTAL_NUMBER
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Global map

- Launch sites near the Equator where land moves faster. If a ship is launched from here, inertia will help the spacecraft keep up the speed to stay in orbit.
- Launch sites close to the coast as to minimize the risk of debris falling near people



Colored success and failure

- This map allowed us to see easily which launch sites have relatively high success rates.

  Green = Successful

  Red = Failed

Launch CCAFS SLC-40

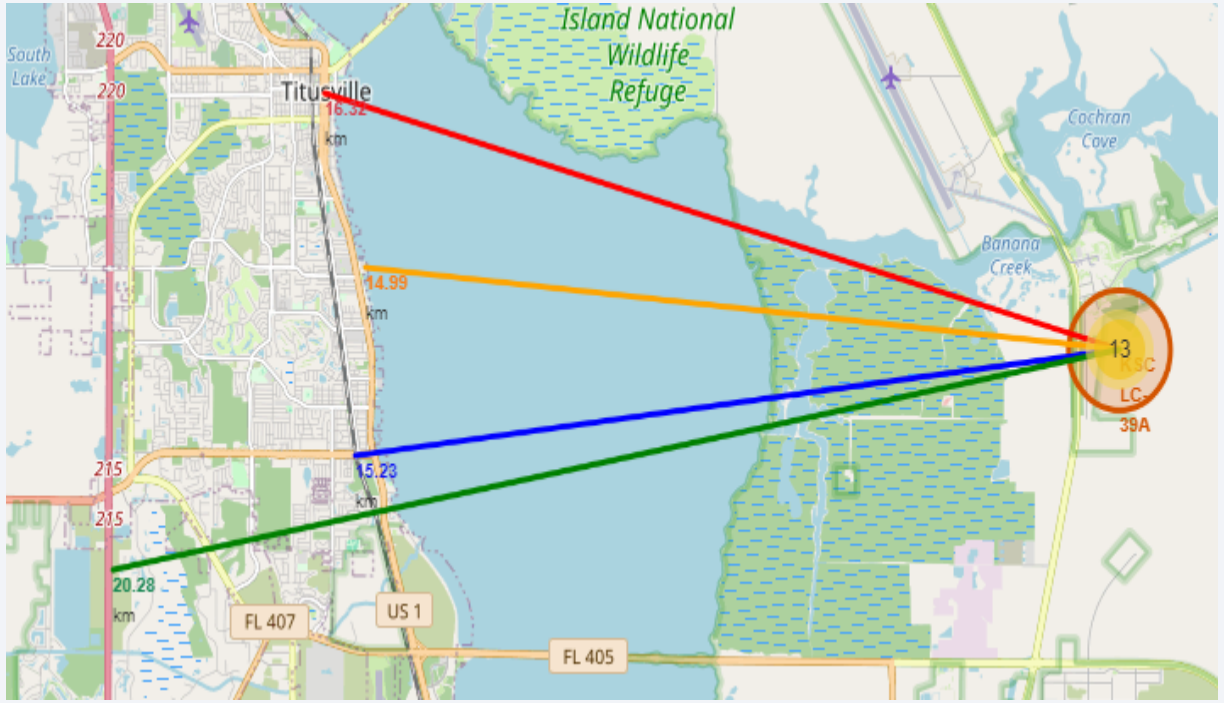
Distances

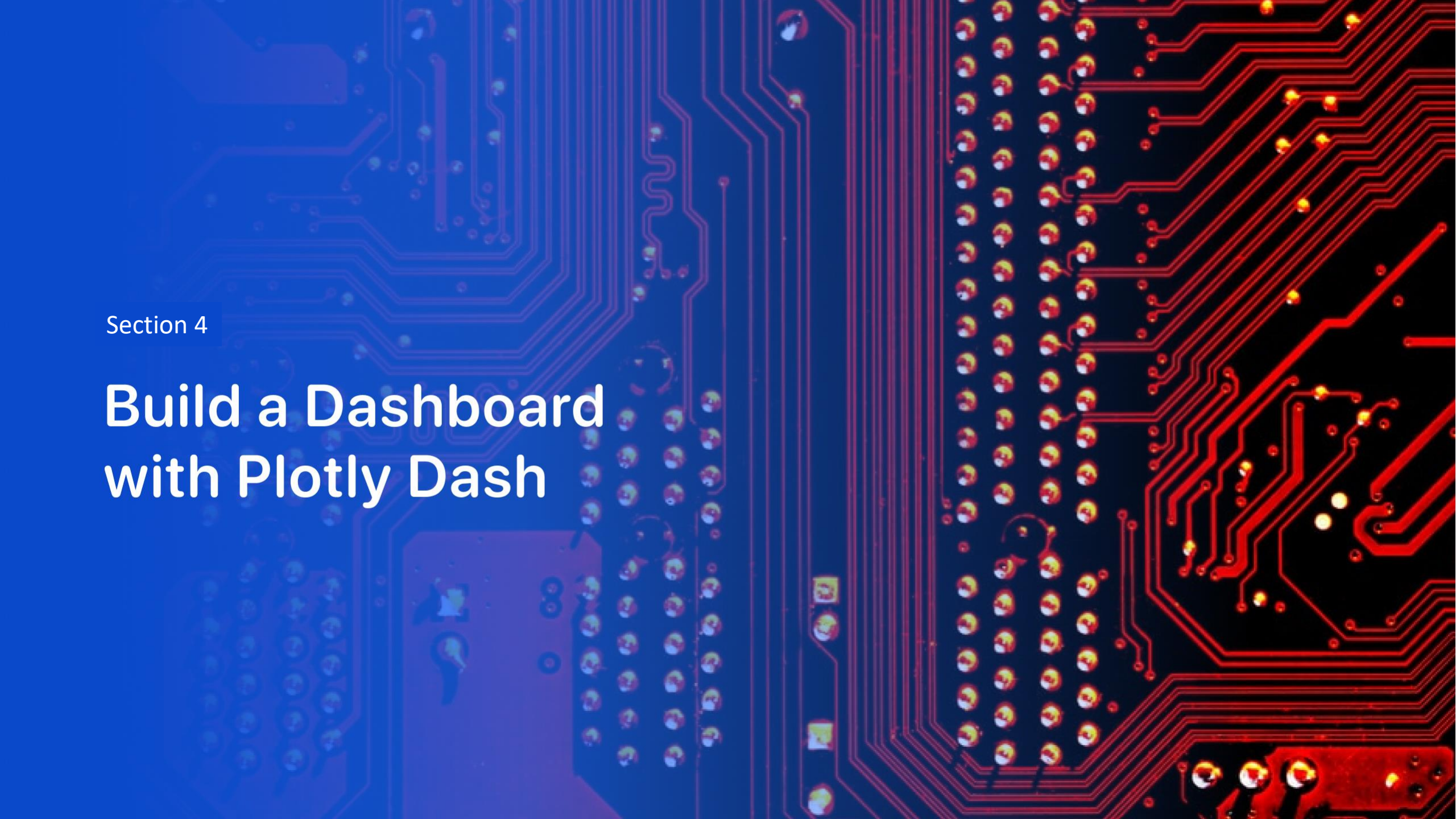
This map shows some distances from the launching site to:

relative close to railway (15.23 km)

relative close to highway (20.28 km)

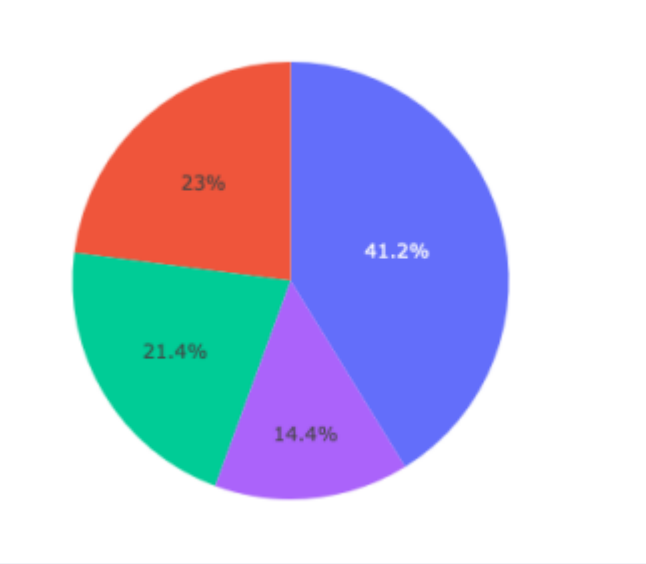
relative close to coastline (14.99 km)




Section 4
Build a Dashboard with Plotly Dash

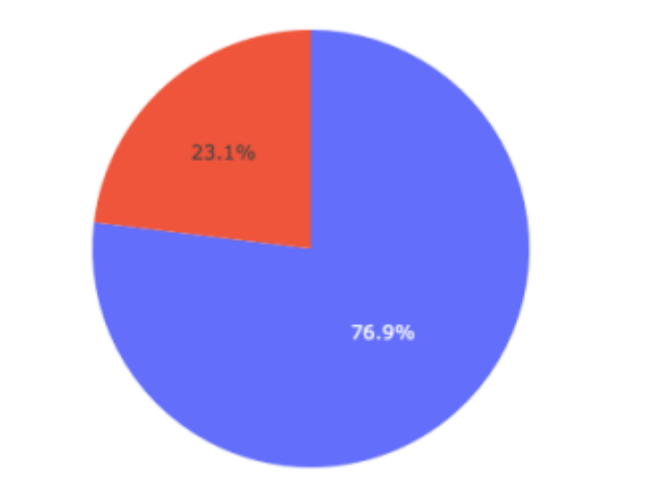
Launch success count

- KSC LC-39A has the most successful launches



Highest launch success ratio

- KSC LC-39A has the highest launch success rate



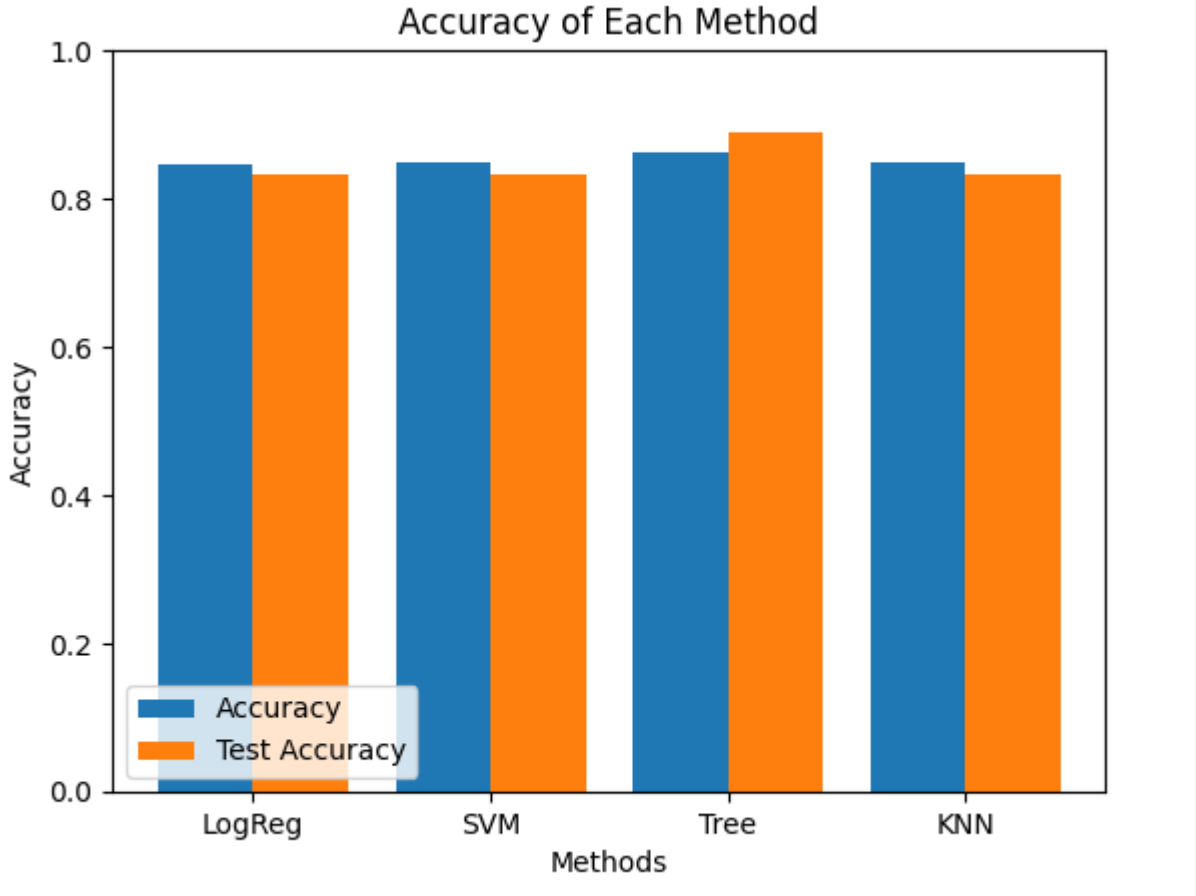
Scatter plot

Section 5

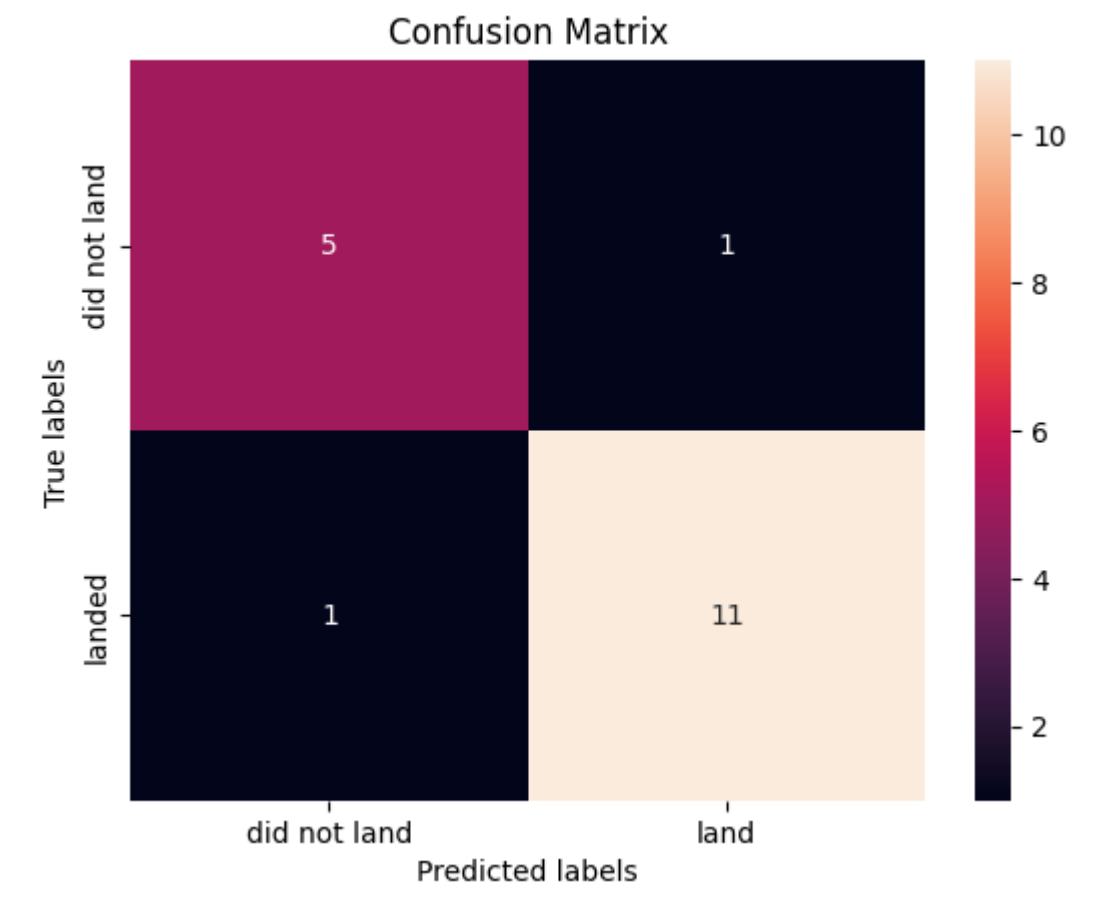
Predictive Analysis (Classification)

Classification Accuracy

- Bar chart thar visualizes the built model accuracy for all built classification models.
- The model that has the highest classification accuracy is the decision tree



Confusion Matrix



Conclusions

- Decision Tree Model best algorithm this time.
- Launches with a low payload show better results.
- Launch sites are near the Equator and the coast.
- The success rate of launches increases over time.

Appendix

- Thanks to courser and my peers for reviewing

Thank you!