IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Shoaib Ahmad Sajid
24-March-2030

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection Through API and Web Scraping
  - Exploratory Data Analysis (EDA)
    - Data Wrangling
    - EDA with SQL
    - EDA with Visualization
  - Building Map with Folium
  - Dashboard with Plotly Dash
  - Prediction with Machine Learning
- Summary of all results
  - EDA Results
  - Interactive Analysis
  - Predictive Analysis

# Introduction

- Project background and context

  Space X is the company of the commercial space age, they launches with very affordable cost. Space X 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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

- Problems you want to find answers

  - Find the factors, features for successful landing of stage 1.
  - Find the best place where landing will be successful.
  - Can successful landing increase and with what conditions.

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - From Space X API
  - From Web scraping from wikipedia page (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01)
- Perform data wrangling
  - Fill the null values where required
  - Filter the data
  - Get the summary

# Methodology

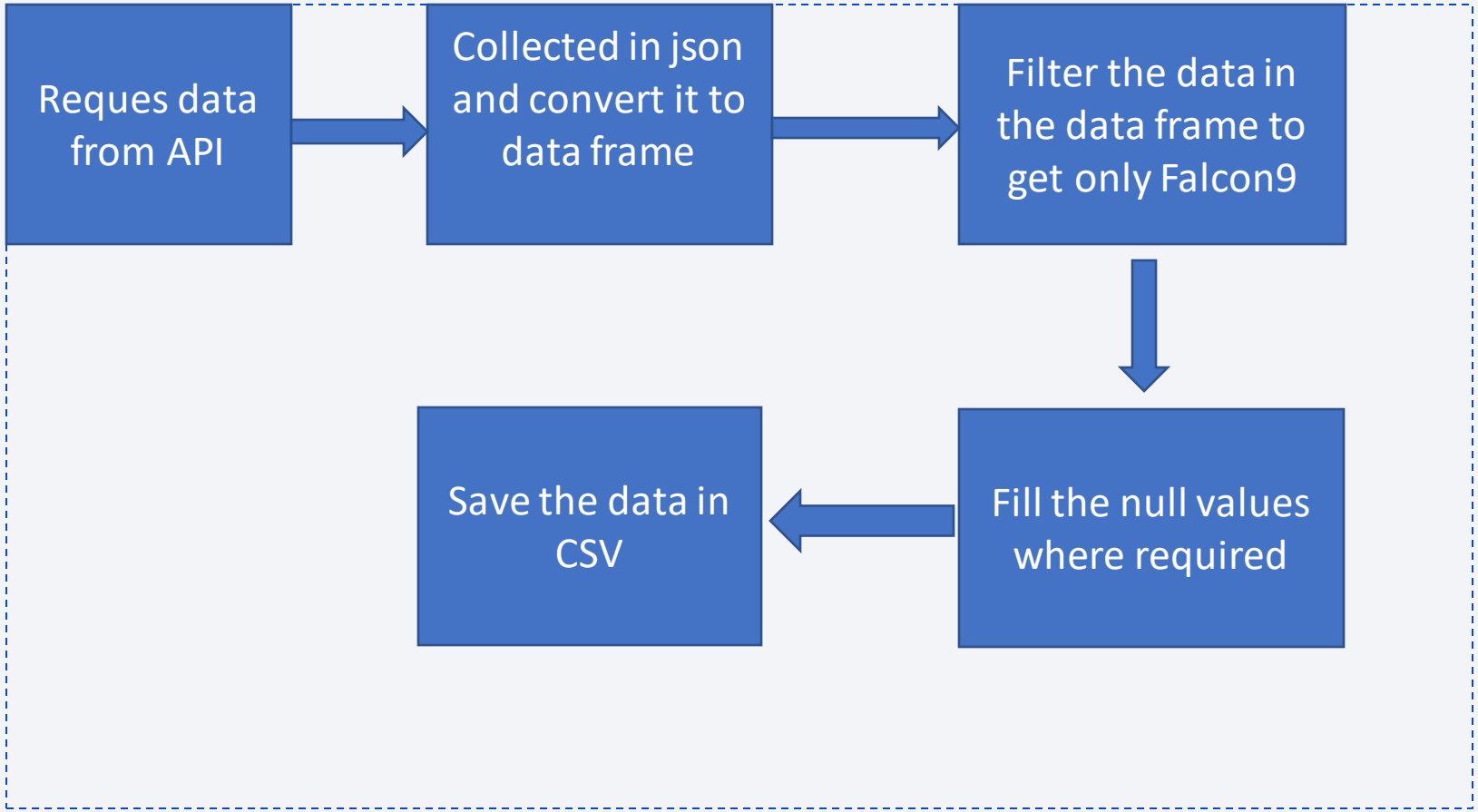
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Normalize the data
  - Devide into 2 parts training and testing
  - Apply the different models to get the best results.

# Data Collection

- Describe how data sets were collected.
  - Collected the data sets from Space X API
  - Web Scraping Wikipedia page.

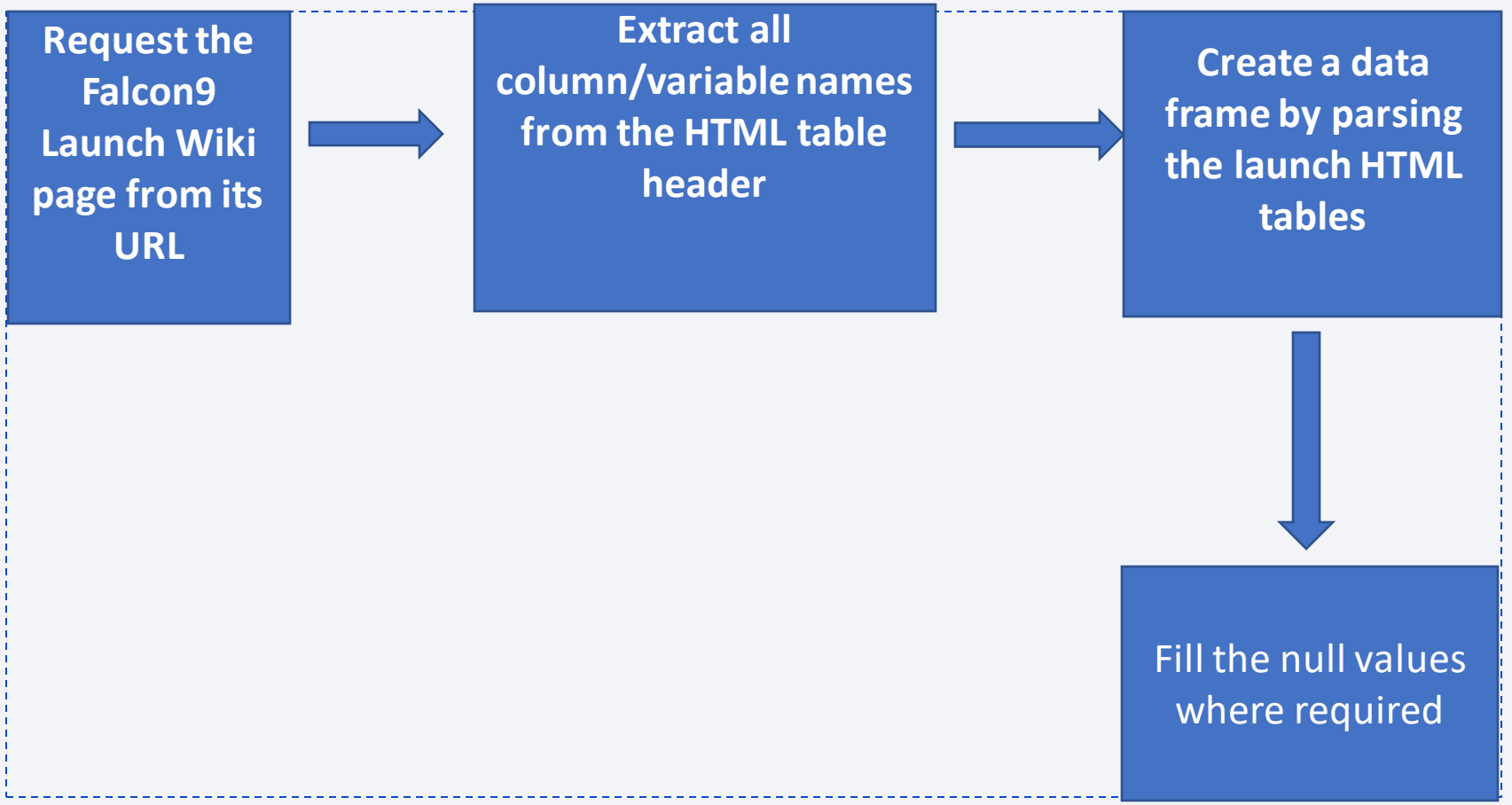
# Data Collection – SpaceX API

- GitHub URL of the completed SpaceX API calls notebook



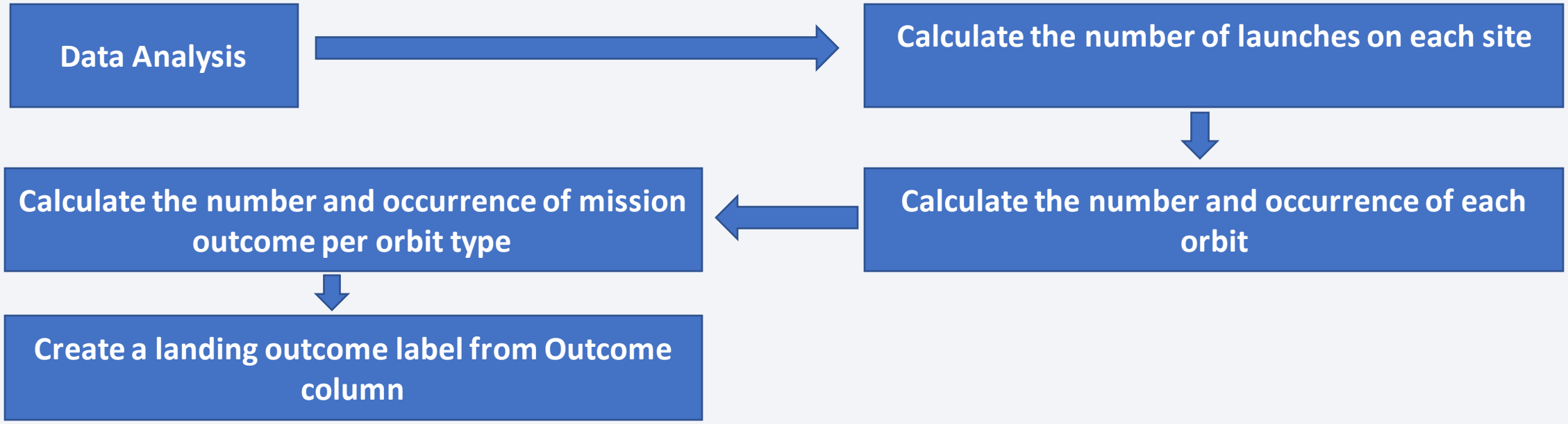
# Data Collection - Scraping

- GitHub URL of the completed web scraping notebook



# Data Wrangling

- I performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.




- GitHub URL of data wrangling notebooks

# EDA with Data Visualization

- Used Scattered Plot to show the relationships between the following
  - FlightNumber and PayloadMass
  - FlightNumber and LaunchSite
  - Payload and LaunchSite
  - FlightNumber and OrbitType
  - Payload and OrbitType
- Bar chart to Visualize the relationship between success rate of each orbit type
- Line chart to Visualize the launch success yearly trend
- GitHub URL EDA with data visualization notebook

# EDA with SQL

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in ground pad was achieved.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster versions which have carried the maximum payload mass. Use a subquery
9. List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. 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

- GitHub URL of EDA with SQL notebook

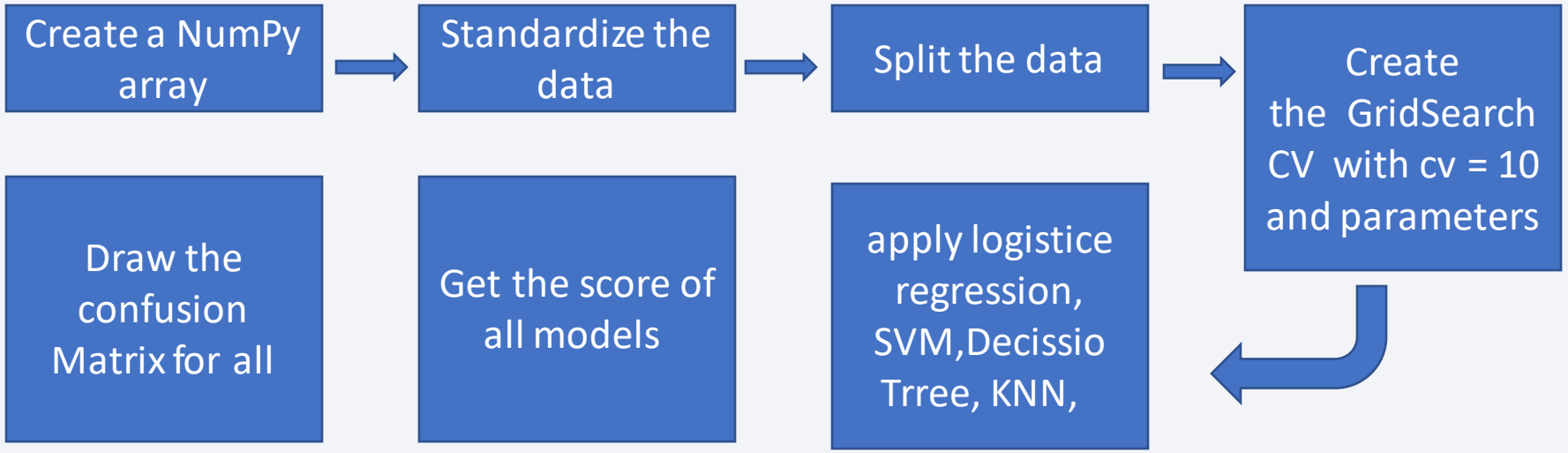
# Build an Interactive Map with Folium

- Added marker , circle, line, popup and marker cluster
- Line  to show the distance between to point.
- Circle to highlight the area
- Marker to indicate the launch sites
- GitHub URL of interactive map with Folium map

# Build a Dashboard with Plotly Dash

- Added dropdown list to select the launch site
- Added pie chart for all sites and also for selected site to show the successful launches counts vs failed launches counts.
- Added the slider to select pay load mass range.
- Added a scatter plot to show the relationship between pay load mass and class
- GitHub URL Plotly Dash lab

# Predictive Analysis (Classification)



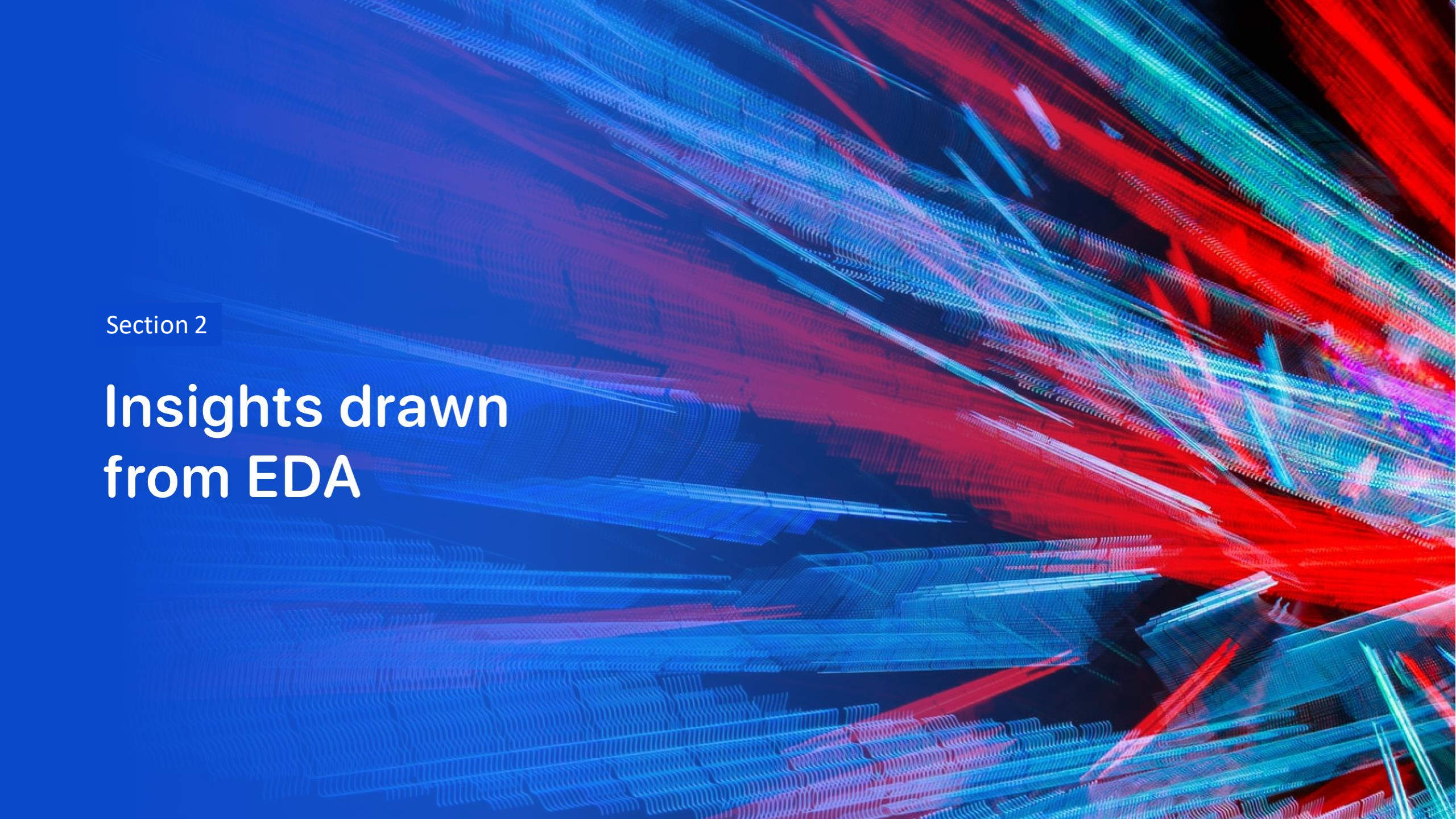

- GitHub URL predictive analysis lab

# Results

- Exploratory data analysis results
  - Different launch sites have different success rates. CAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4Ehas a success rate of 77%.
  - there are no rockets launched for heavy payload mass(greater than 10000).
  - 4 orbits have highest success rate
- Interactive analytics demo in screenshots
- Predictive analysis results
  - Score and confusion plots for all models are same, so we can use any one.

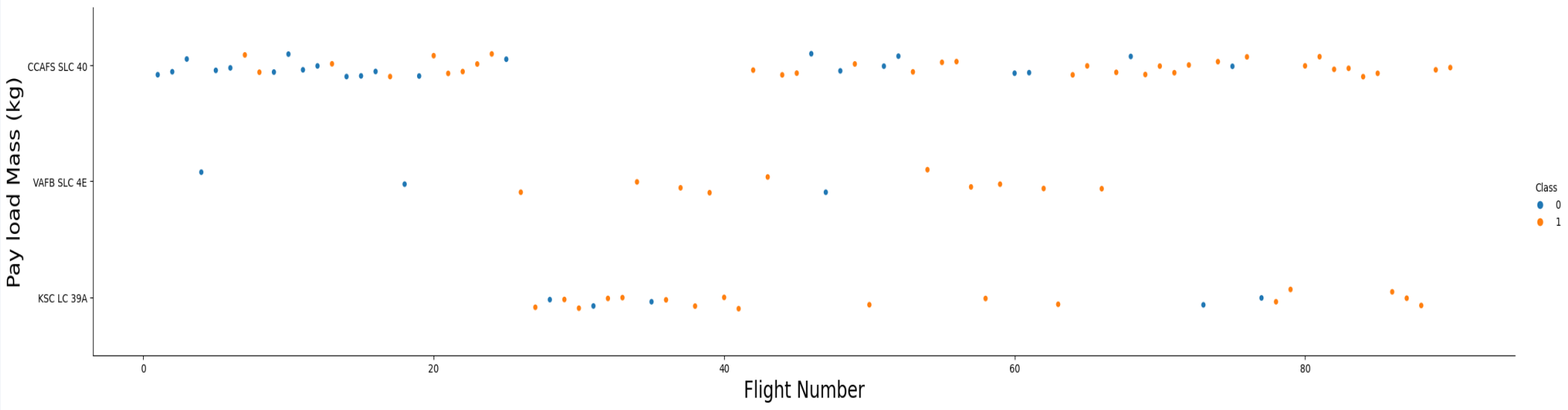
Section 2

# Insights drawn from EDA



# Flight Number vs. Launch Site

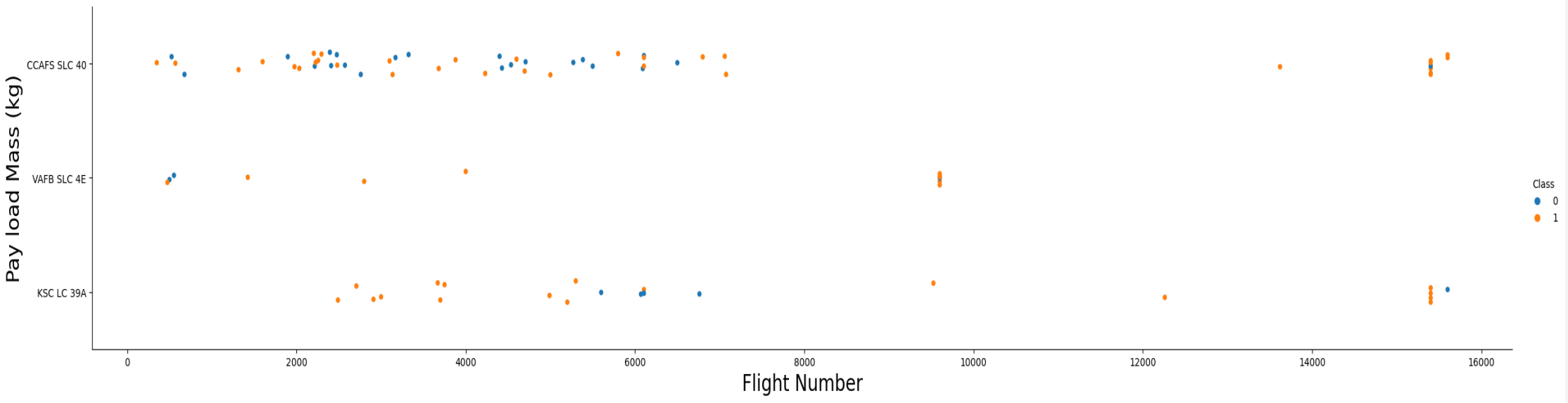
- Show a scatter plot of Flight Number vs. Launch Site




- CCAFS SLC 40 has more launches.
- VAFB SLC 4E has higher success rate.

# Payload vs. Launch Site

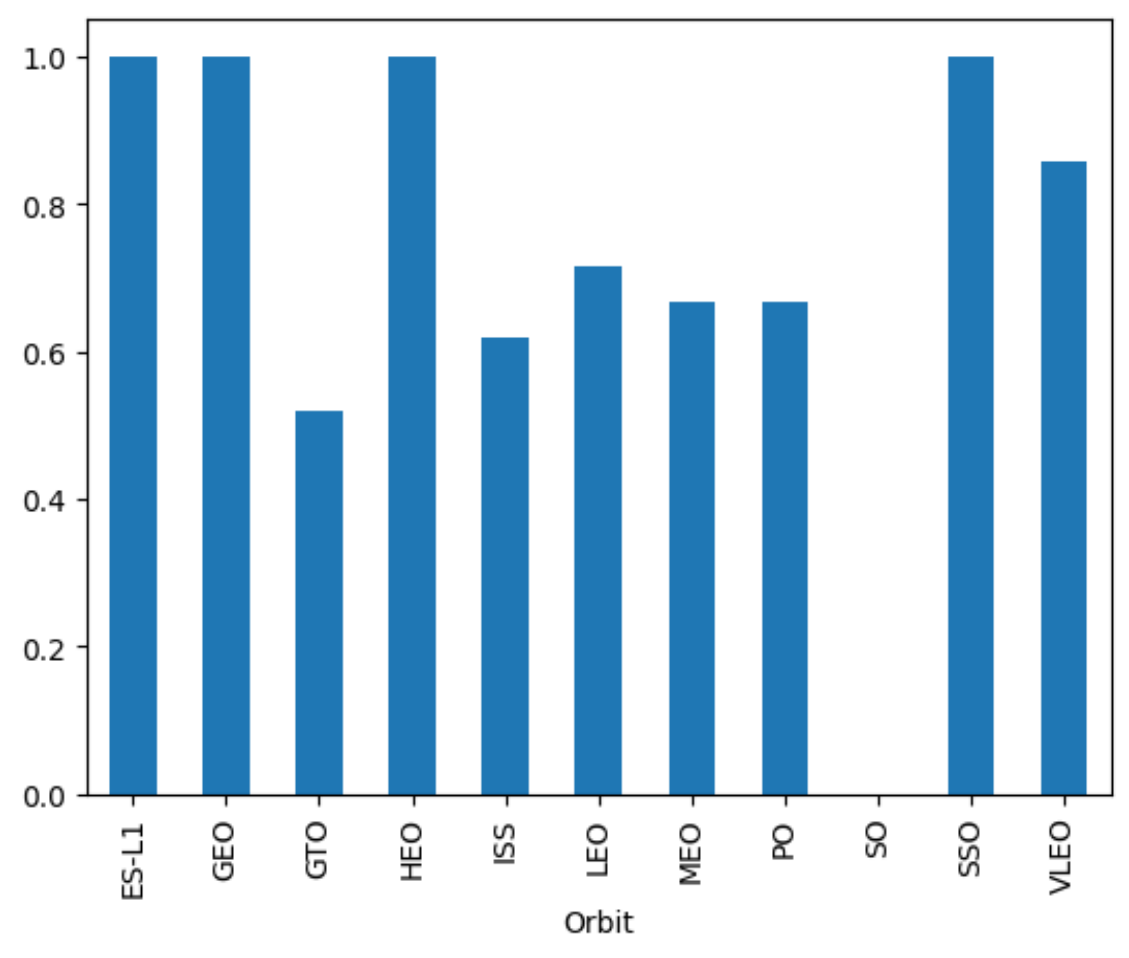
- Show a scatter plot of Payload vs. Launch Site




- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- The higher the payload mass the higher success rate

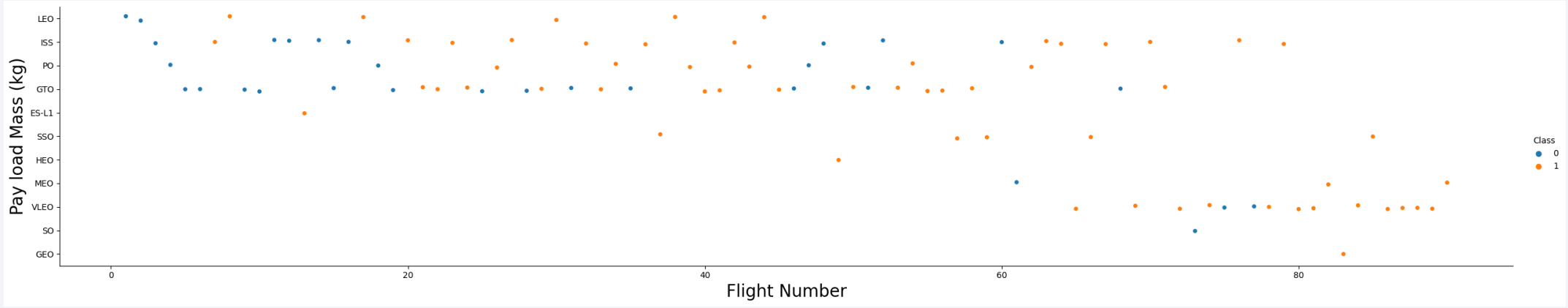
# Success Rate vs. Orbit Type

- ES-L1, GFO, ISS, SSO with highest success rate
- SO is 0% success
- 4 orbits between 40 to 80 percent success rate



# Flight Number vs. Orbit Type

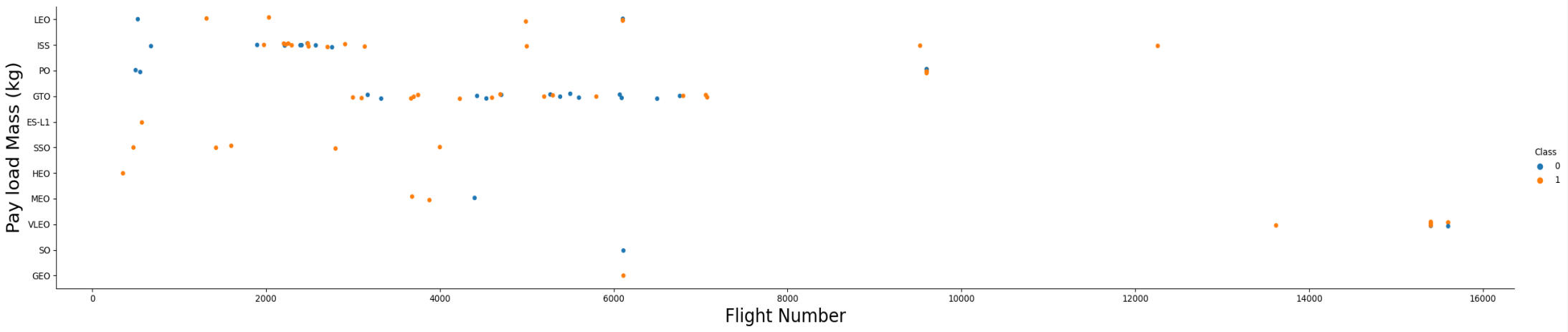
- Show a scatter point of Flight number vs. Orbit type




- The LEO orbit the Success appears related to the number of flights
- no relationship between flight number when in GTO orbit

# Payload vs. Orbit Type

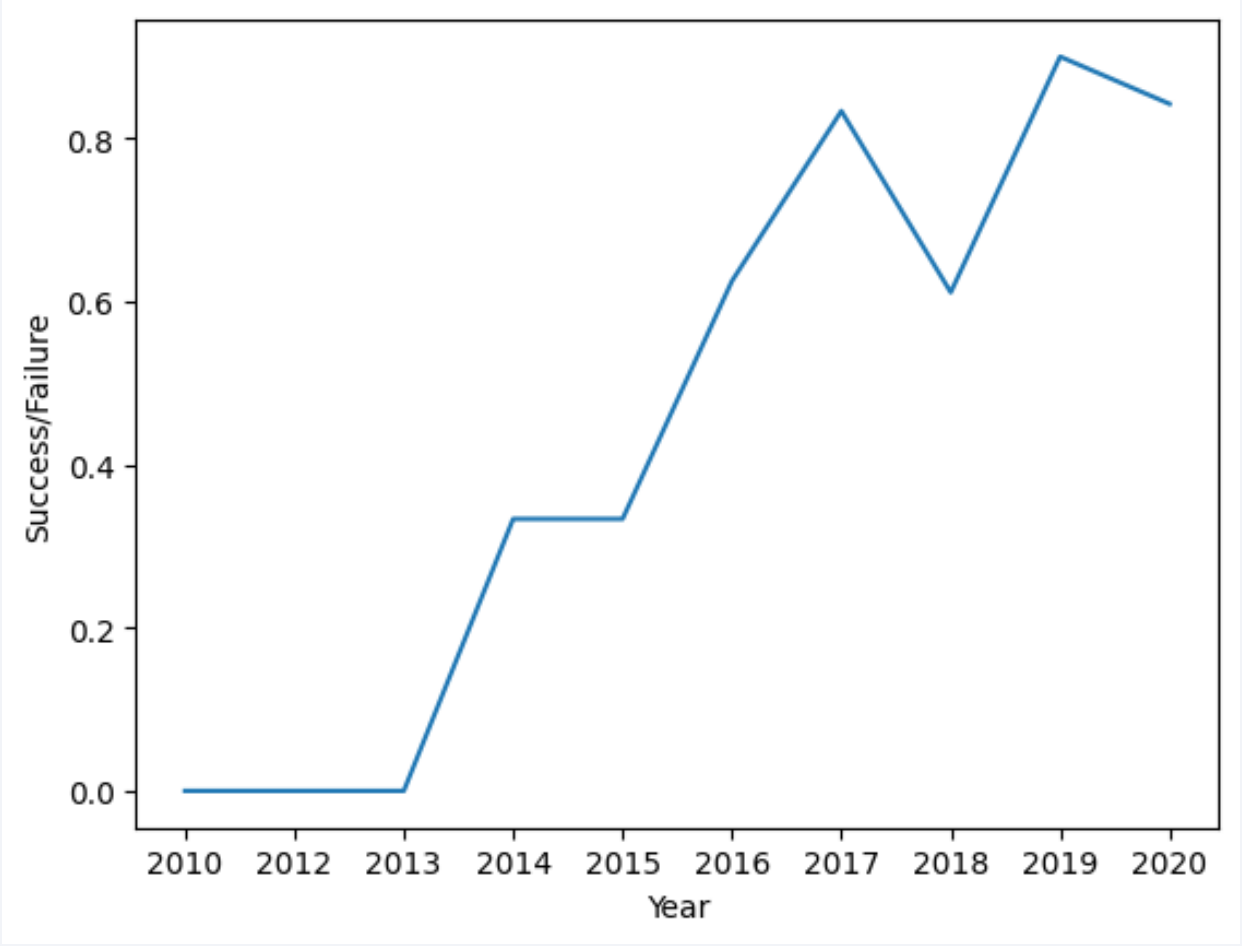
- Show a scatter point of payload vs. orbit type




- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- for GTO we cannot distinguish this well as both positive landing rate and negative landing

# Launch Success Yearly Trend

- Success rate kept increasing since 2013 till 2020



# All Launch Site Names

- Use the distinct keyword to get the unique names

```
%sql select distinct launch_site from spacex
✓ 0.1s
```

| LAUNCH_SITE |
| --- |
| CCAFS LC-40 |
| CCAFS SLC-40 |
| KSC LC-39A |
| VAFB SLC-4E |

# Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Use the like keyword with % sign in where clause and limit the rows to 5

```
%sql select * from spacex where launch_site like 'CCA%' limit 5
✓  0.0s
```

| DATE | TIME_UTC | BOOSTER_VERSION | LAUNCH_SITE | PAYLOAD | PAYLOAD_MASS_KG_ | ORB |
|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO |
| 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 |
| 2012-05-22 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO |
| 2012-10-08 | 0:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO |

# Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Select the 'nasa (crs) only and count the pay load mass

```
%sql select count(payload_mass_kg_) as totalpayload from spacex where customer = 'NASA (CRS)'
✓ 0.0s
```

| TOTALPAYLOAD |
| --- |
| 20 |

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Filter the data by booster version and find the average by using avg function

```
%sql select avg(payload_mass_kg_) as avgpayload from spacex where booster_version = 'F9 v1.1'
✓ 0.0s
```

| AVGPAYLOAD |
| --- |
| 2928 |

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Find the minimum date from table where landing outcome is success

```
%sql select min(date) from spacex where landing_outcome = 'Success (ground pad)'
```

ecute Cell (^Enter)

| 1 |
| --- |
| 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
- Filtered by landing outcome and payload mass ranges between 4000 to 6000

```
%sql select * from spacex where landing_outcome = 'Success (drone ship)' and (payload_mass_kg_ > 4000 and payload_mass_kg_ < 6000)
```

✓ 0.0s

| DATE | TIME_UTC | BOOSTER_VERSION | LAUNCH_SITE | PAYLOAD | PAYLOAD_MASS_KG_ | ORBIT | CUSTOMER | N |
|---|---|---|---|---|---|---|---|---|
| 2016-05-06 | 5:21:00 | F9 FT B1022 | CCAFS LC-40 | JCSAT-14 | 4696 | GTO | SKY Perfect JSAT Group | S |
| 2016-08-14 | 5:26:00 | F9 FT B1026 | CCAFS LC-40 | JCSAT-16 | 4600 | GTO | SKY Perfect JSAT Group | S |
| 2017-03-30 | 22:27:00 | F9 FT B1021.2 | KSC LC-39A | SES-10 | 5300 | GTO | SES | S |
| 2017-10-11 | 22:53:00 | F9 FT B1031.2 | KSC LC-39A | SES-11 / EchoStar 105 | 5200 | GTO | SES EchoStar | S |

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Count the mission out come with group by

```
%sql select mission_outcome, count(mission_outcome) as aa from spacex group by mission_outcome
✓ 0.0s
```

| MISSION_OUTCOME | AA |
|---|---|
| Failure (in flight) | 1 |
| Success | 99 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- get maximum payload mass and filter the main query by this.

```
%sql select booster_version, payload_mass_kg_ from spacex where payload_mass_kg_ in (select max(payload_mass_kg_) from spacex group by booster_version)
```

✓ 0.0s

| BOOSTER_VERSION | PAYLOAD_MASS_KG_ |
|---|---|
| F9 v1.1 B1014 | 4159 |
| F9 v1.1 B1015 | 1898 |
| F9 v1.1 B1016 | 4707 |
| F9 v1.1 B1018 | 1952 |
| F9 FT B1019 | 2034 |
| F9 v1.1 B1017 | 553 |
| F9 FT B1020 | 5271 |
| F9 FT B1021.1 | 3136 |
| F9 FT B1022 | 4696 |
| F9 FT B1023.1 | 3100 |
| F9 FT B1024 | 3600 |

# 2015 Launch Records

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

- Present your query result with a short explanation here

```
%sql select * from spacex where landing_outcome = 'Failure (drone ship)' and year(date) = '2015'
✓ 0.0s
```

| DATE | TIME_UTC | BOOSTER_VERSION | LAUNCH_SITE | PAYLOAD | PAYLOAD_MASS_KG_ | ORBIT | CUSTOMER | MISSION_OUTCON |
|---|---|---|---|---|---|---|---|---|
| 2015-01-10 | 9:47:00 | F9 v1.1 B1012 | CCAFS LC-40 | SpaceX CRS-5 | 2395 | LEO (ISS) | NASA (CRS) | Success |
| 2015-04-14 | 20:10:00 | F9 v1.1 B1015 | CCAFS LC-40 | SpaceX CRS-6 | 1898 | LEO (ISS) | NASA (CRS) | Success |

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

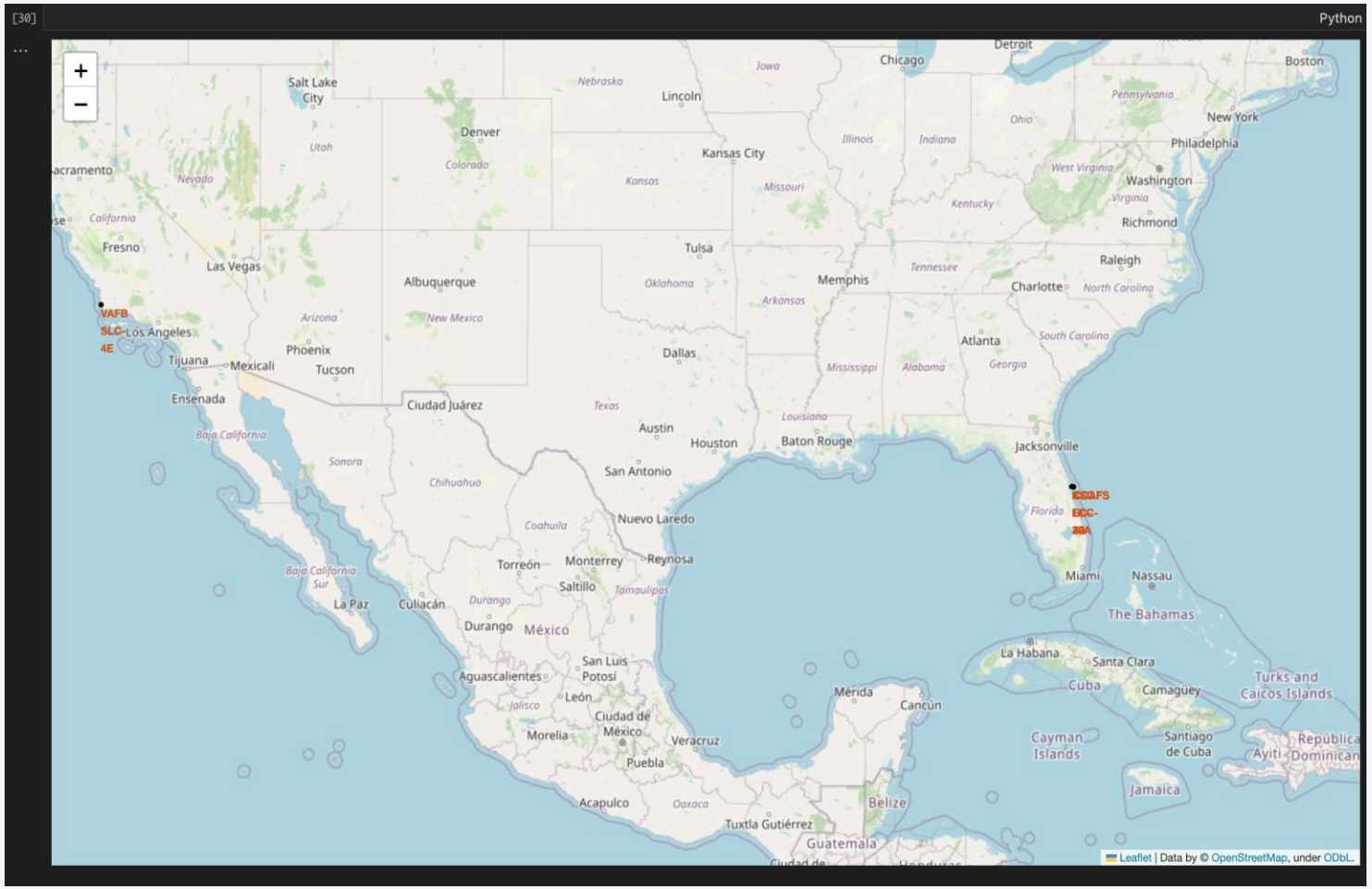
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- Filtered by dates , group by landing outcome and then sorted by count in descending order

```
%%sql SELECT LANDING_OUTCOME, COUNT(*) AS COUNT_LAUNCHES FROM SPACEX
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY COUNT_LAUNCHES DESC
✓ 0.0s
```

| LANDING_OUTCOME | COUNT_LAUNCHES |
|---|---|
| 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 |

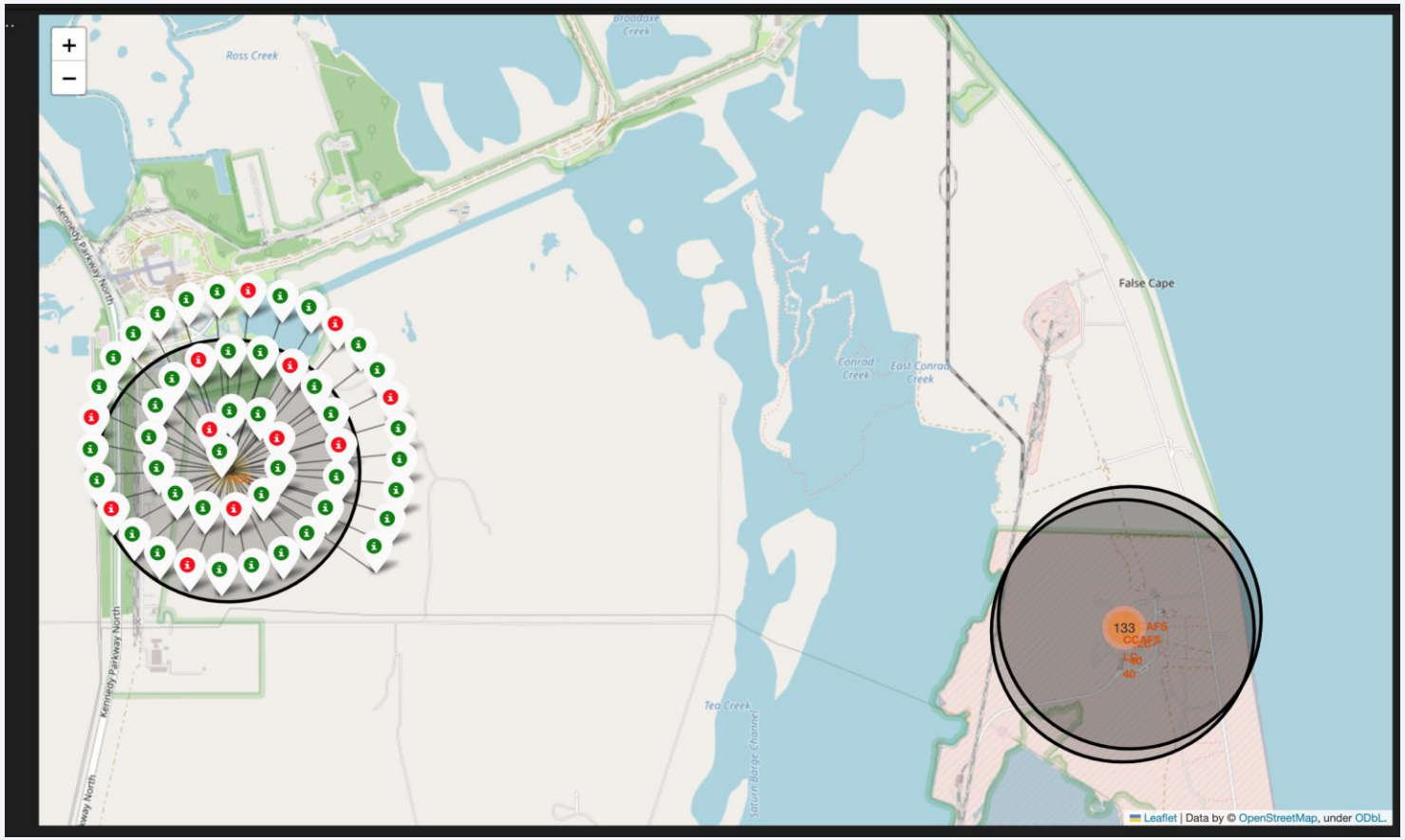

Section 3
Launch Sites
Proximities Analysis

# All Launches Sites



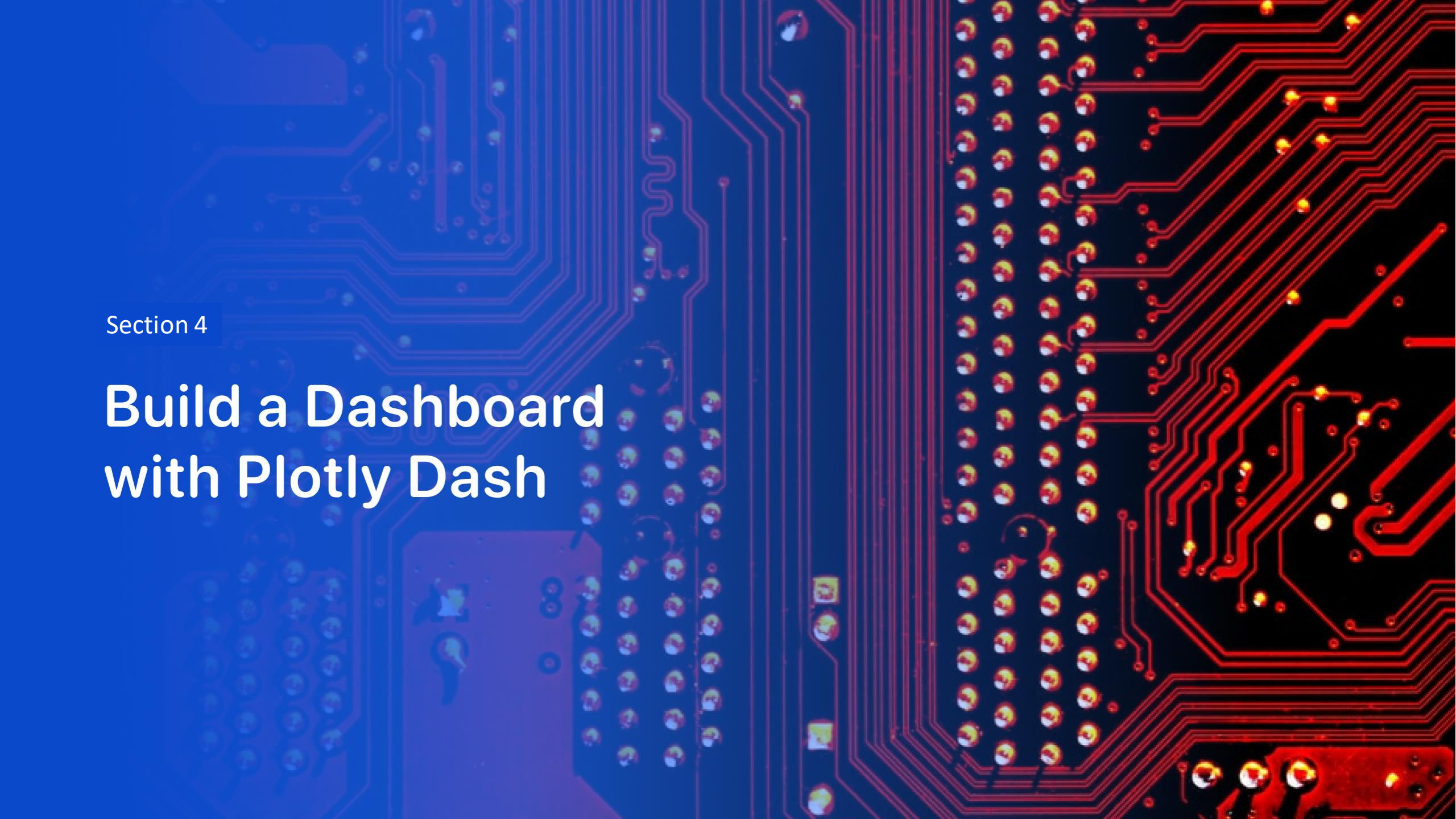
# Success and Failed Launches

- The Red shows the failed launched and green shows the success launches.
- When click on the any highlighted circle it shows like this.

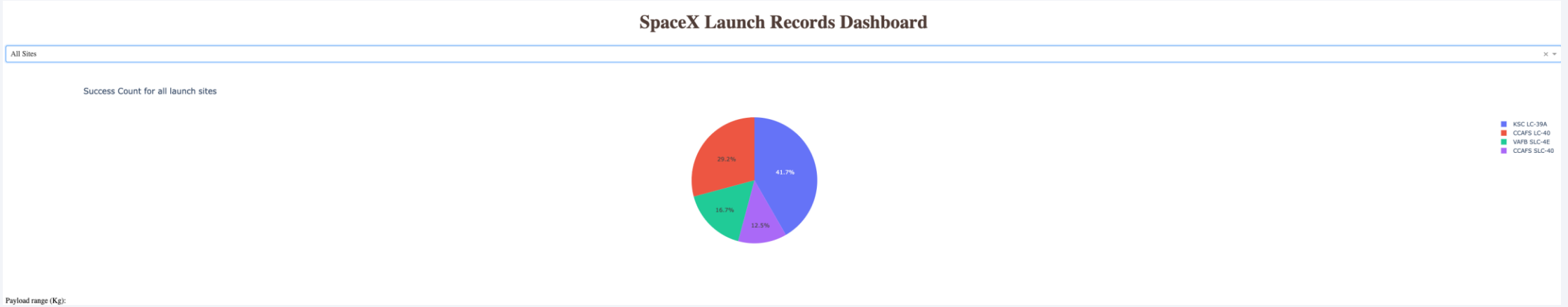


# Launch Sites Distance

- Show the distance between the sites.

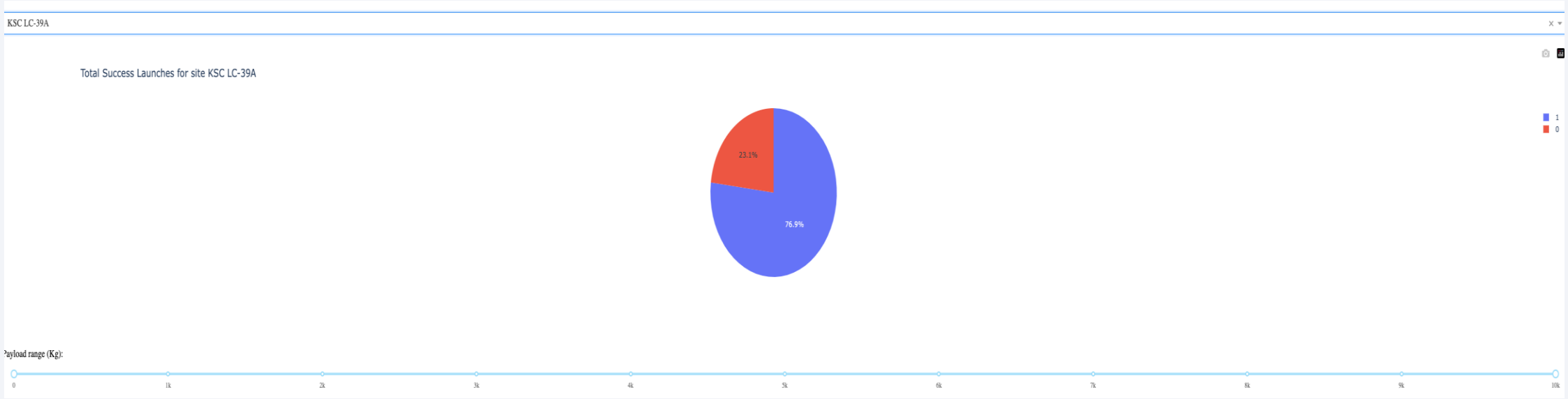

Section 4
Build a Dashboard with Plotly Dash

# Launch Success Count All Sites



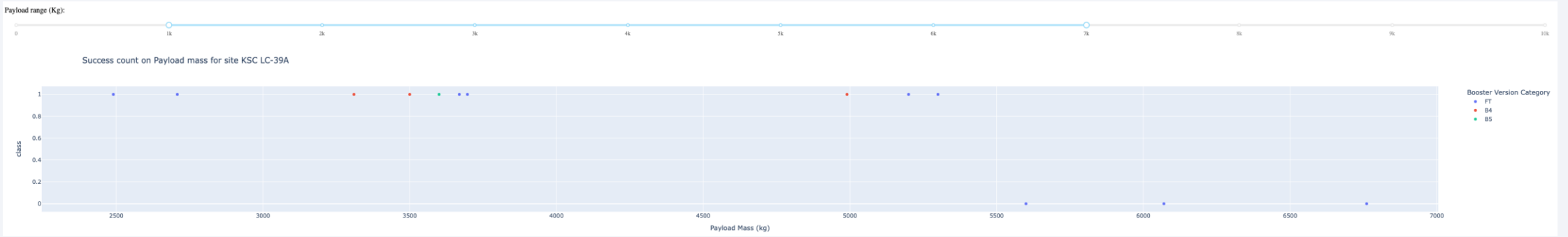

- KSC LC 39A has the highest success rate.
- CCAFS SLC 40 has the minimum success rate.

# Highest Launch Success Ratio




- 76.9% success ratio

# Payload Mass Vs Launch Outcomes

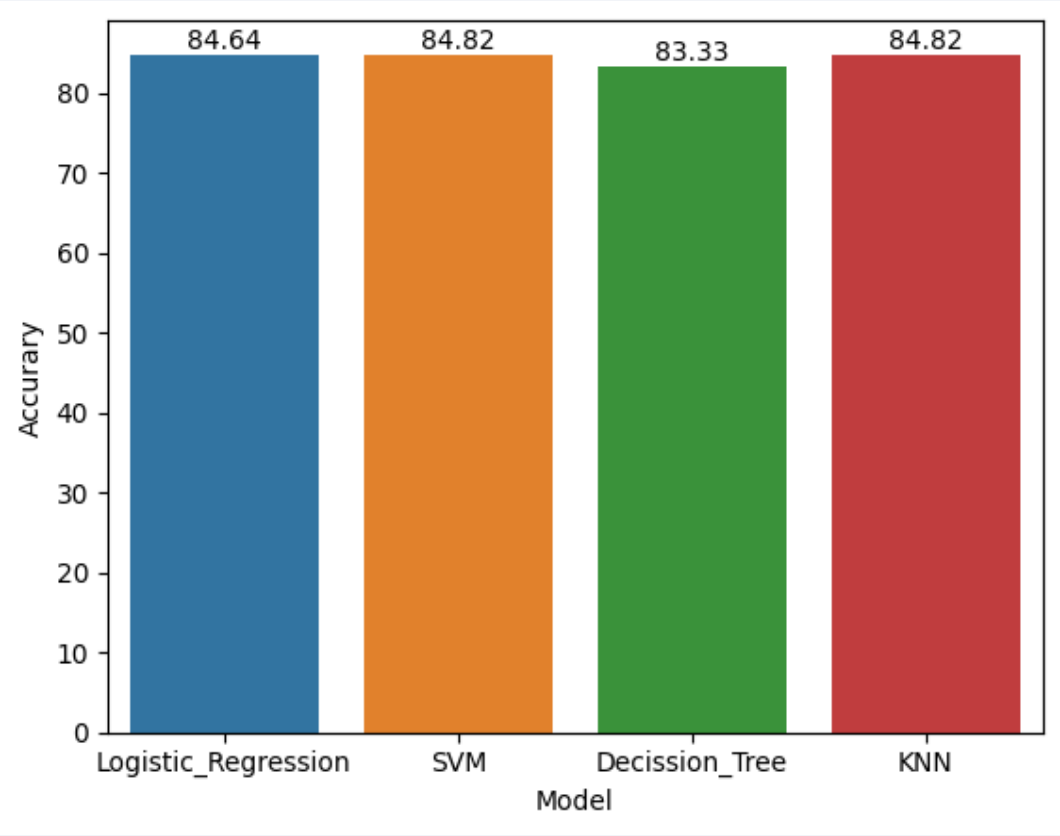



Section 5

# Predictive Analysis (Classification)

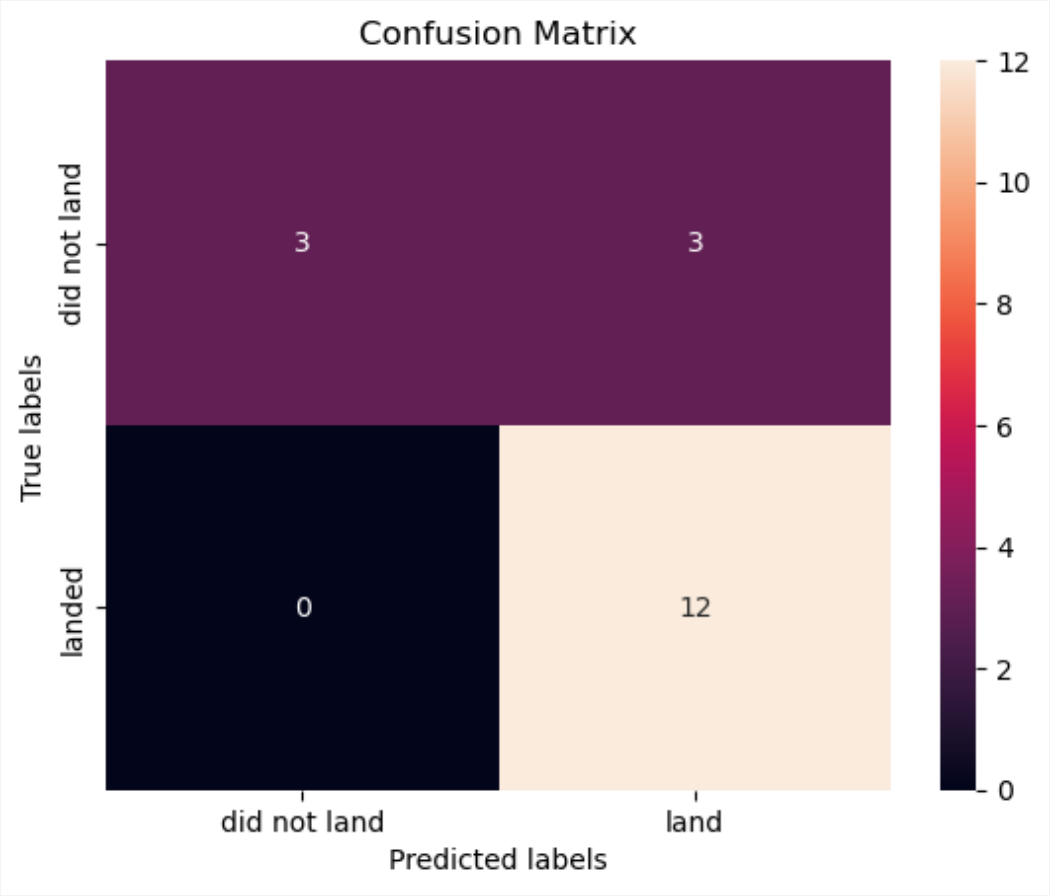
# Classification Accuracy

- SVM and KNN have the highest accuracy 84.82%



# Confusion Matrix

- The major issue is in the false positive as we can see there failed land in the chart which is not correct.



# Conclusions

- The Higher the payload mass the higher success
- Launches success rate started increasing from 2013 till 2020
- Decision Tree is more suitable algorithm for this dataset.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.
- KSC LC-39A has the highest success rate of the launches from all the sites.

# Thank you!