



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

*Victor Hugo  
Kawano Monteiro  
06/02/2025*



# Outline

---

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

# Executive Summary

---

- API integration, Web Scrapping, Data Preprocessing
- EDA, Visualization with Maps and Dynamic Charts
- Machine Learning approches comparison
- Machine Learning surpass 80% accuracy

# Introduction

---

- SpaceX's revolutionary approach to space launches has disrupted the aerospace industry, primarily through their innovation in rocket reusability. The Falcon 9 rocket, their flagship vehicle, represents a significant advancement in space technology and commercial spaceflight economics.
- When the Falcon 9 will Land succesful?
- Which Launch Site have more sucessful landings?



Section 1

# Methodology

# Methodology

---

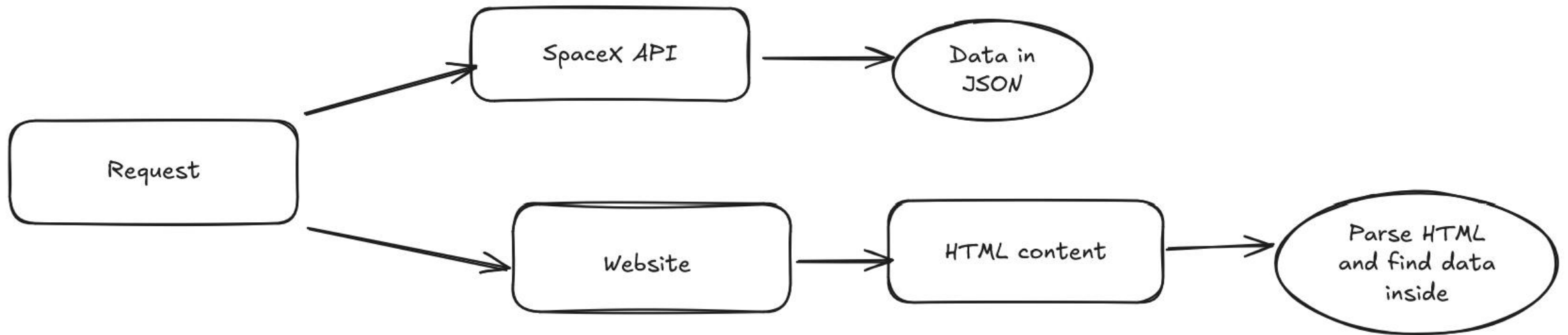
## Executive Summary

- Data collection methodology:
  - By SpaceX API and Wikipedia Scrapping
- Perform data wrangling
  - The data are loaded and processed in pandas
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The models is tuned and evaluated with sklearn abstractions

# Data Collection

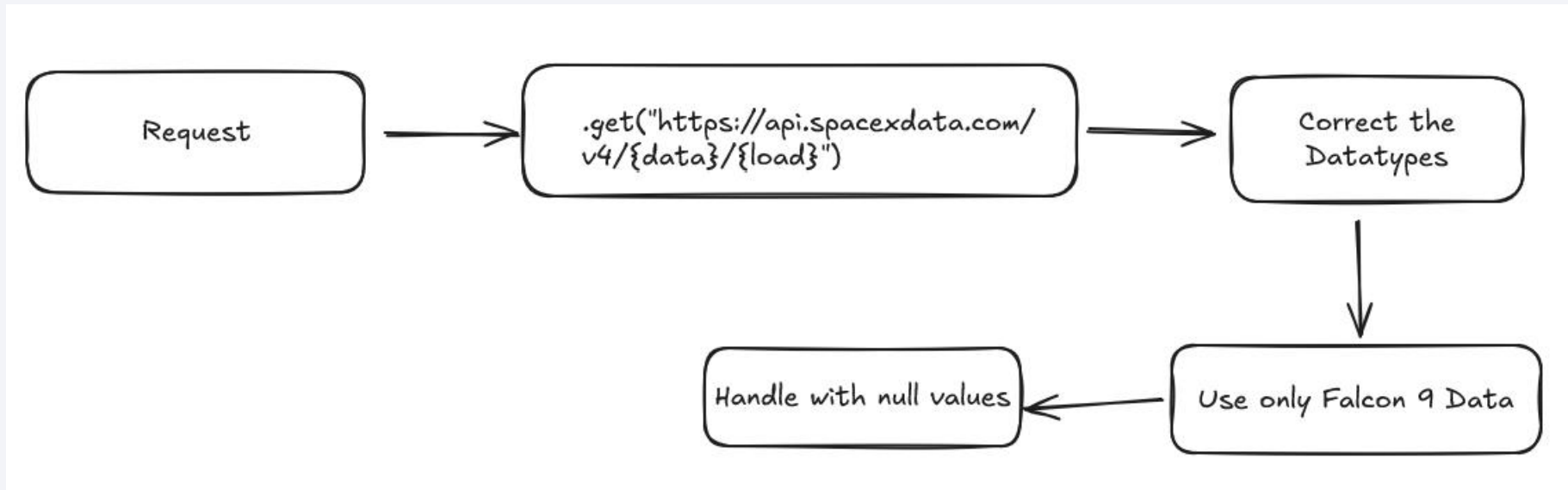
---

- The data get collected using the Requests to get response from SpaceX API and BeautifulSoup to scrap HTML Tables from Wikipedia



# Data Collection – SpaceX API

---

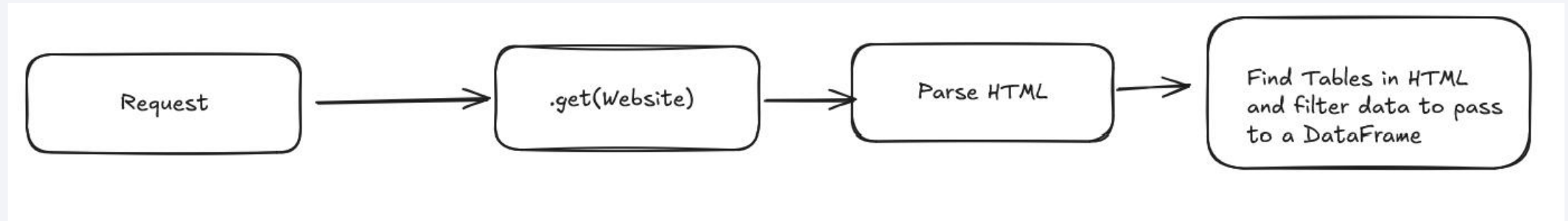


<https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb>



# Data Collection - Scraping

---



<https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-webscraping.ipynb>

# Data Wrangling

---



<https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb>

# EDA with Data Visualization

---

- During EDA we analyze relationship between Flight Number and Payload Mass and see as the flight number increases often return successfully.
- Visualize relationship between Flight Number and Launch Site to see if the Launch Site impact in number of successful landing also visualize the relationship between Payload Mass and Launch Site.
- Check the success rate in each Orbit, relationship between Payload Mass and Orbit
- See the Success Rate in general increase since 2013

*<https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/edadataviz.ipynb>*

# EDA with SQL

---

- Show distincts values of Launch Site
- Show 5 records with Launch Site starting with CCA
- Show total Payload Mass carried by NASA CRS
- Show average Payload carried by F9 v1.1
- Show the date of first succesful landing in ground
- Show boosteres which succes drone ship with payload between 4000 and 6000
- Show all succesful and failure mission outcomes
- Show boosters which carried the maximum payload mass
- Show the number of each landing outcome between 2010-06-04 and 2017-03-20

[https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-eda-sql-coursera\\_sqllite.ipynb](https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb)

# Build an Interactive Map with Folium

---

- Markers and Circles are add for each Launch Site to locate them
- Marker Cluster to group the Launch Records in each Launch Site and they outcome in marker color red ou green
- Line to Coastline and nearest city

[https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/lab\\_jupyter\\_launch\\_site\\_location.ipynb](https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/lab_jupyter_launch_site_location.ipynb)



# Build a Dashboard with Plotly Dash

---

- In this dash app is possible to select in a dropdown section the Launch Site or include all to filter in a Pie Chart with Successful and Failure Outcomes, additionally there's a Scatter chart to see correlation between Payload and Launch Outcome with a Range Slider to filter the Payload values

[https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/spacex\\_dash\\_app.py](https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/spacex_dash_app.py)

# Predictive Analysis (Classification)

---

- First the dataset is splitted in 80/20 for Train and Test, after that each algorithm is initialized and use GridSearchCV to adjust the parameters. In the end the scores is compared and all models perform 0.83 score.

[https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/SpaceX\\_Machine\\_Learning\\_Prediction\\_Part\\_5.ipynb](https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.ipynb)

# Results

---

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



The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is dynamic and technological.

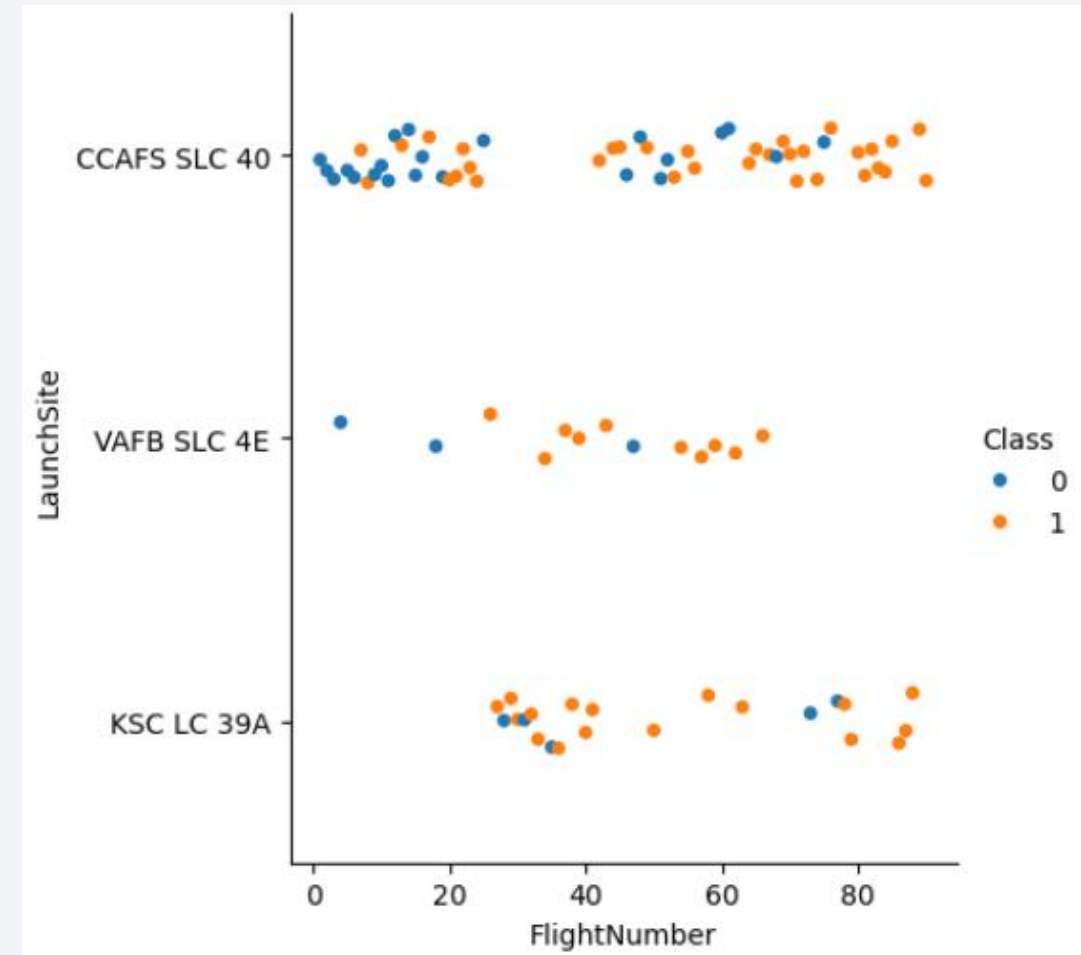
Section 2

# Insights drawn from EDA



# Flight Number vs. Launch Site

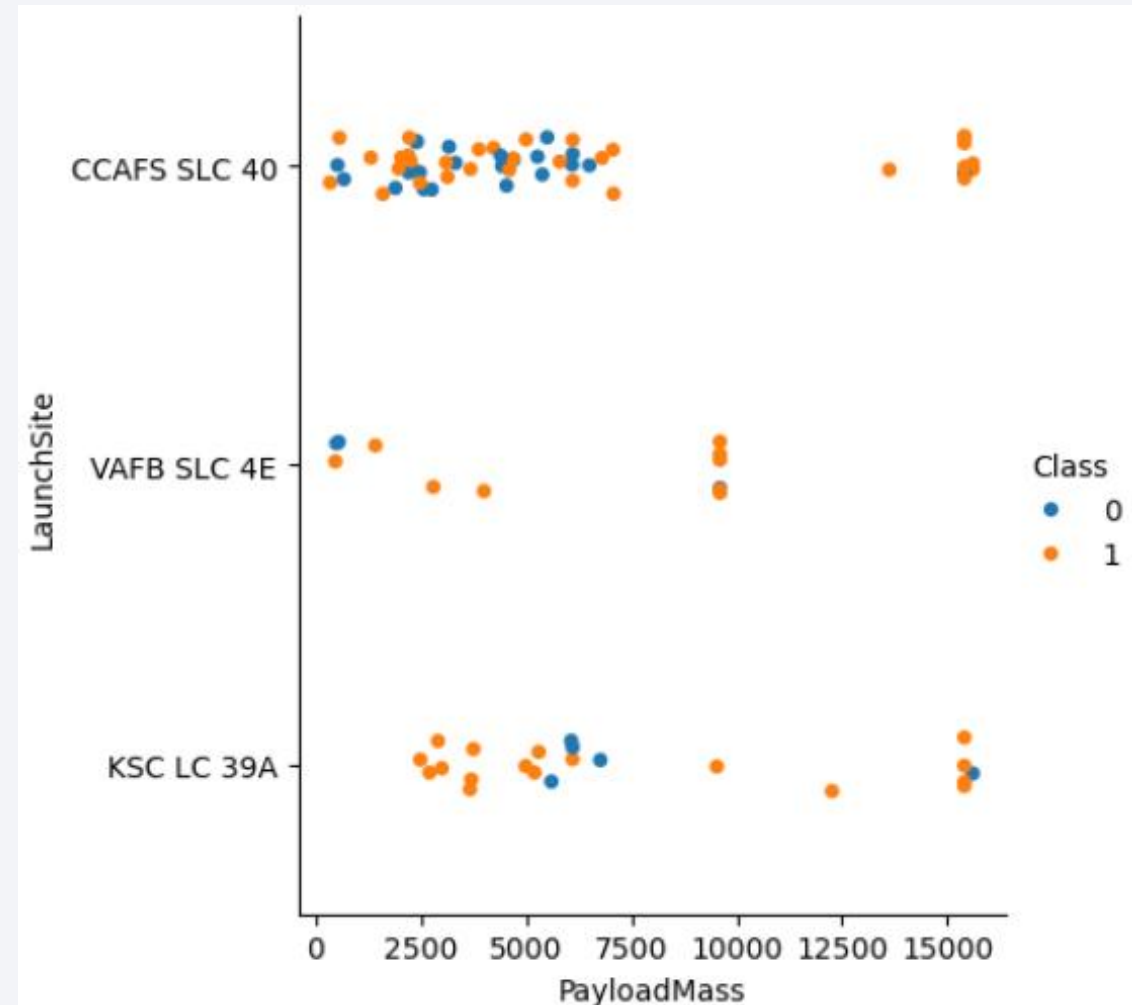
- As the FlightNumber increase the often succes is





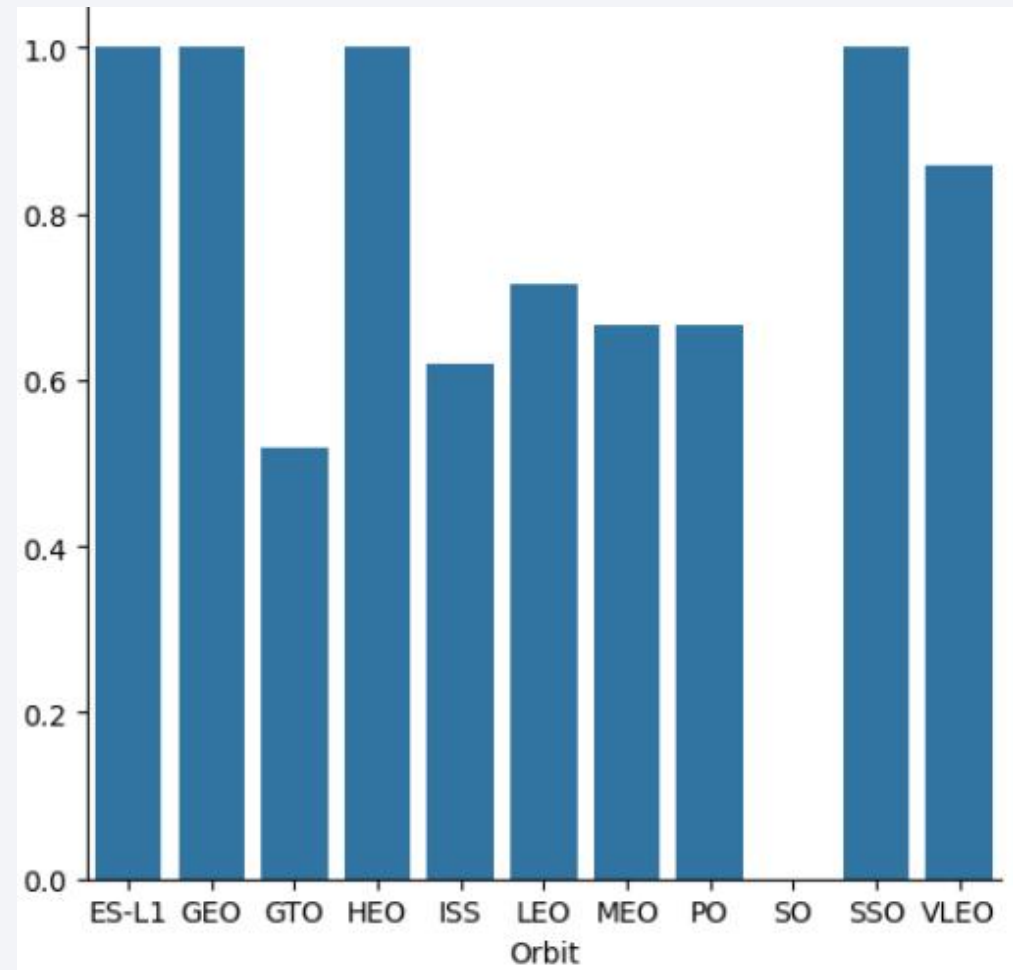
# Payload vs. Launch Site

- Higher Payload have more success and VAFB SLC 4E doesn't have rockerts launched with more than 10k Kg



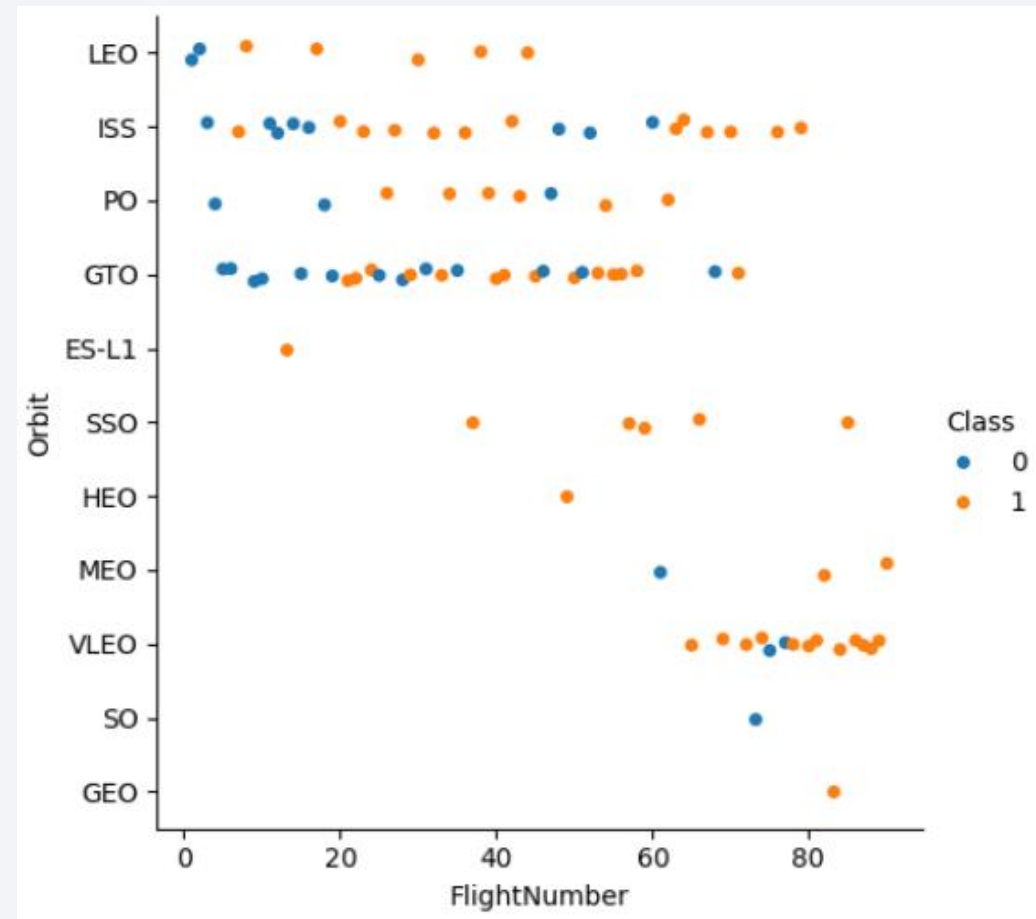
# Success Rate vs. Orbit Type

- *ES-L1, GEO, HEO and SSO has a perfect success rate and SO doesn't has a single success outcome.*



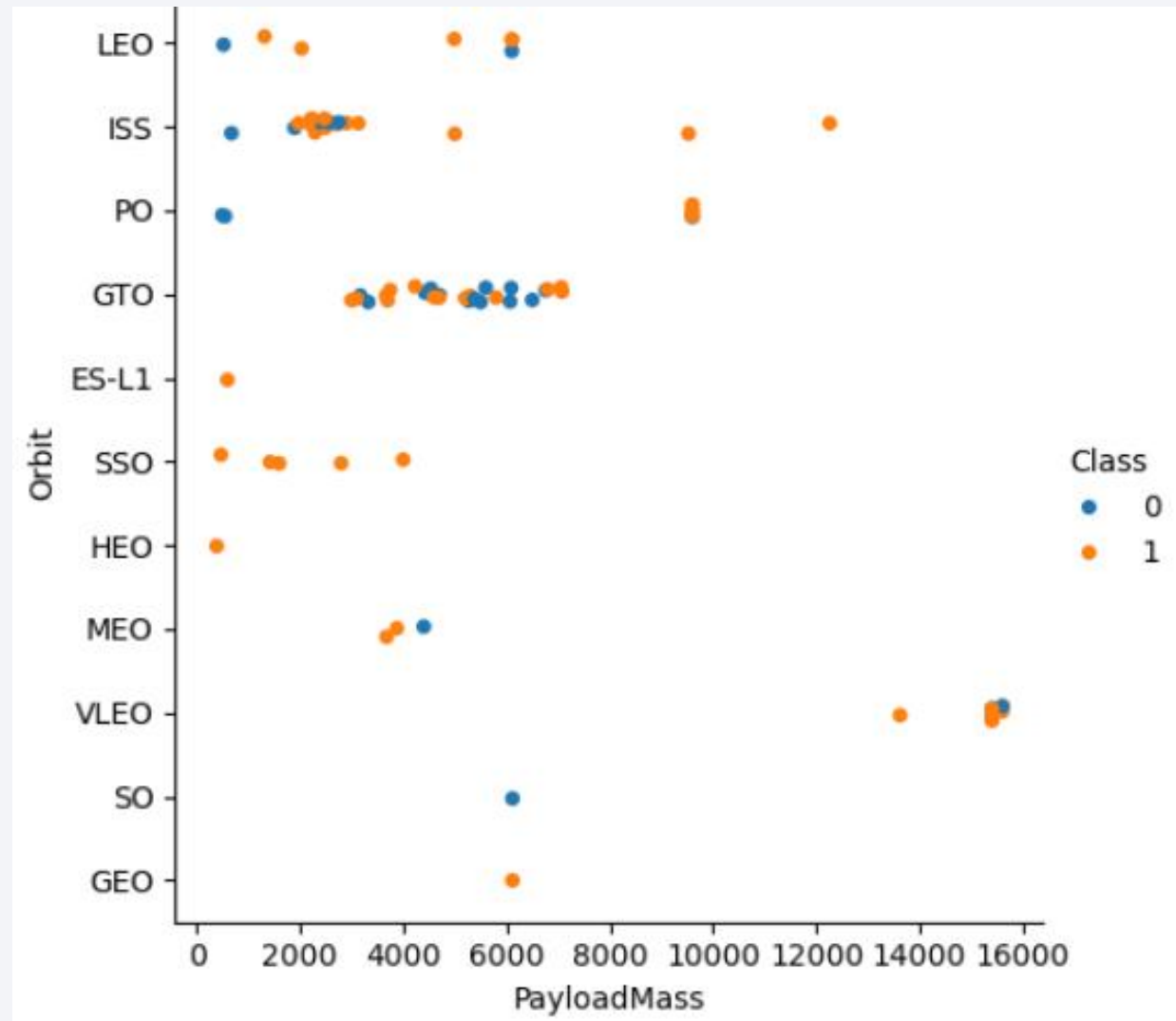
# Flight Number vs. Orbit Type

- Some Orbits hasn't lower values of FlightNumber
- LEO has a relationship between FlightNumber and Succes



# Payload vs. Orbit Type

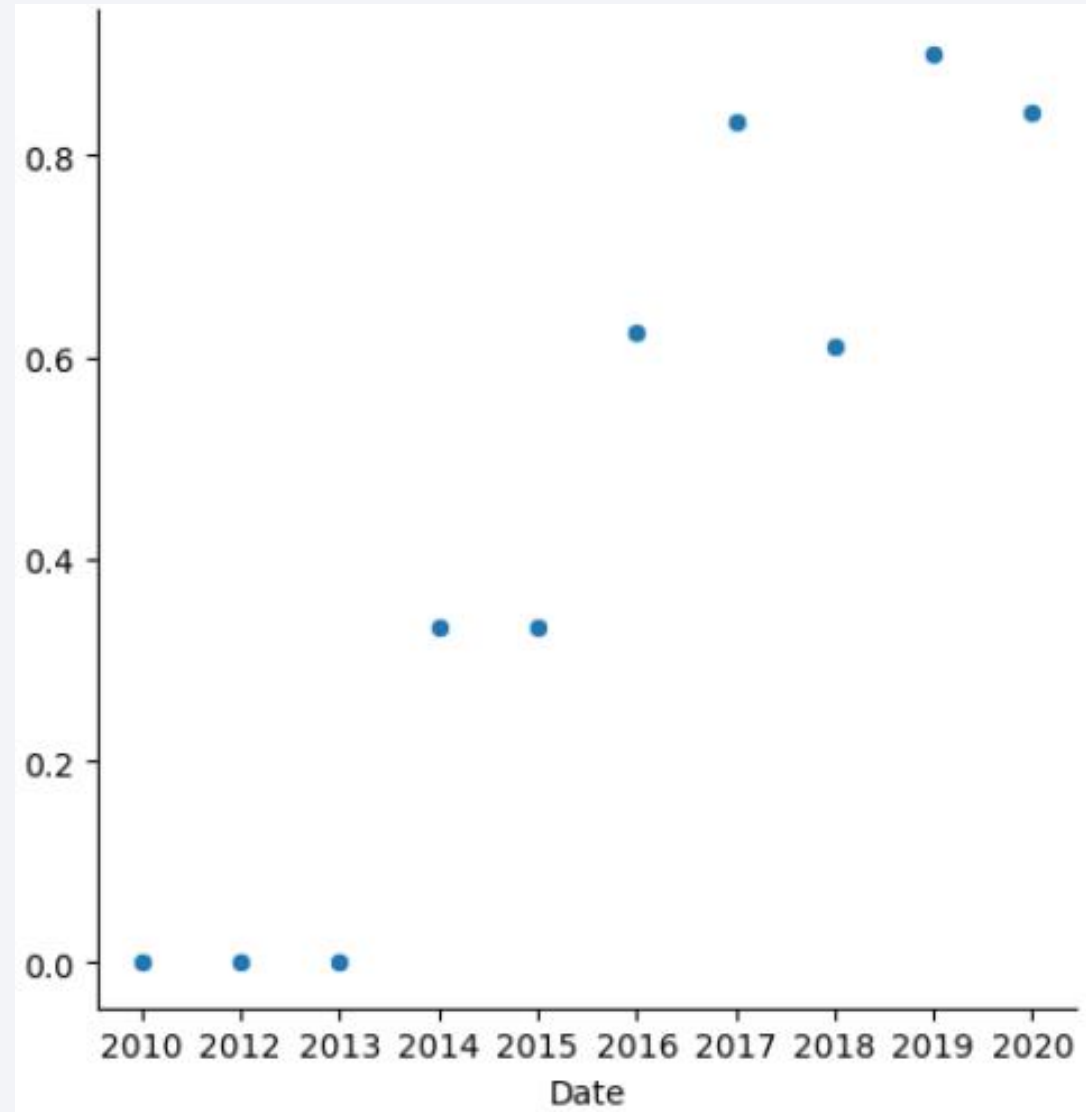
- Like in FlightNumber the trends differ from orbit to orbit



# Launch Success Yearly Trend

---

- As the year pass the succes rate tend to increase





# All Launch Site Names

---

- *Select unique values from Launch Site*

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

# Launch Site Names Begin with 'CCA'

---

- Select Launch Site with CCA and some other pattern and limit the results to only 5 records

| Date       | Time (UTC) | Booster_Version | Launch_Site | Payload                                                       | PAYLOAD_MASS_KG_ | Orbit     | Customer        | Mission_Outcome | Landing_Outcome     |
|------------|------------|-----------------|-------------|---------------------------------------------------------------|------------------|-----------|-----------------|-----------------|---------------------|
| 2010-06-04 | 18:45:00   | F9 v1.0 B0003   | CCAFS LC-40 | Dragon Spacecraft Qualification Unit                          | 0                | LEO       | SpaceX          | Success         | Failure (parachute) |
| 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 (ISS) | NASA (COTS) NRO | Success         | Failure (parachute) |
| 2012-05-22 | 7:44:00    | F9 v1.0 B0005   | CCAFS LC-40 | Dragon demo flight C2                                         | 525              | LEO (ISS) | NASA (COTS)     | Success         | No attempt          |
| 2012-10-08 | 0:35:00    | F9 v1.0 B0006   | CCAFS LC-40 | SpaceX CRS-1                                                  | 500              | LEO (ISS) | NASA (CRS)      | Success         | No attempt          |
| 2013-03-01 | 15:10:00   | F9 v1.0 B0007   | CCAFS LC-40 | SpaceX CRS-2                                                  | 677              | LEO (ISS) | NASA (CRS)      | Success         | No attempt          |

# Total Payload Mass

---

- Sum all the values in Payload when the customer is NASA CRS

```
SUM(PAYLOAD_MASS_KG_)
45596
```

# Average Payload Mass by F9 v1.1

---

- Calculate the average value of Payload from all records with booster version F9 v1.1

```
AVG(PAYLOAD_MASS_KG_)
```

```
2534.66666666666665
```

# First Successful Ground Landing Date

---

- Select the minimum value of Date where the Mission Outcome is Succes

```
MIN(Date)  
2010-06-04
```



## Successful Drone Ship Landing with Payload between 4000 and 6000

---

- Show all boosters with succes in drone ship and the payload is between 4000 and 6000

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

# Total Number of Successful and Failure Mission Outcomes

---

- Count the number of rows with each Mission Outcome

| Mission_Outcome                  | COUNT("Mission_Outcome") |
|----------------------------------|--------------------------|
| Failure (in flight)              | 1                        |
| Success                          | 98                       |
| Success                          | 1                        |
| Success (payload status unclear) | 1                        |

# Boosters Carried Maximum Payload

---

- *Show all Boosters which queried the Highest Payload*

| 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

---

- Show month, booster and launch site for failures in drone ship in 2015

| 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 |

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

---

- Order in descending each landing outcome between 2010-06-04 and 2017-03-20

| Landing_Outcome        | COUNT("Landing_Outcome") |
|------------------------|--------------------------|
| 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                        |

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

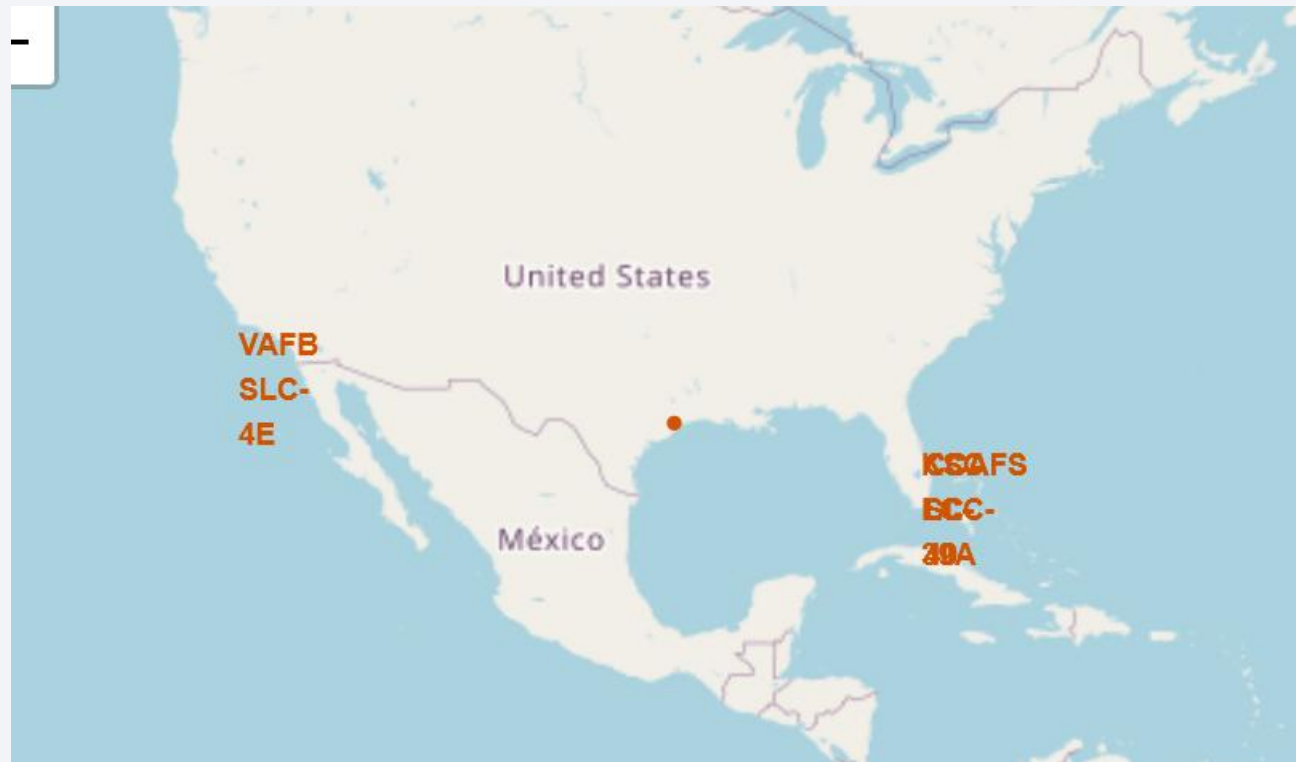
Section 3

# Launch Sites Proximities Analysis

# <Folium Map Screenshot 1>

---

- Is possible to see the Launch Sites in the map

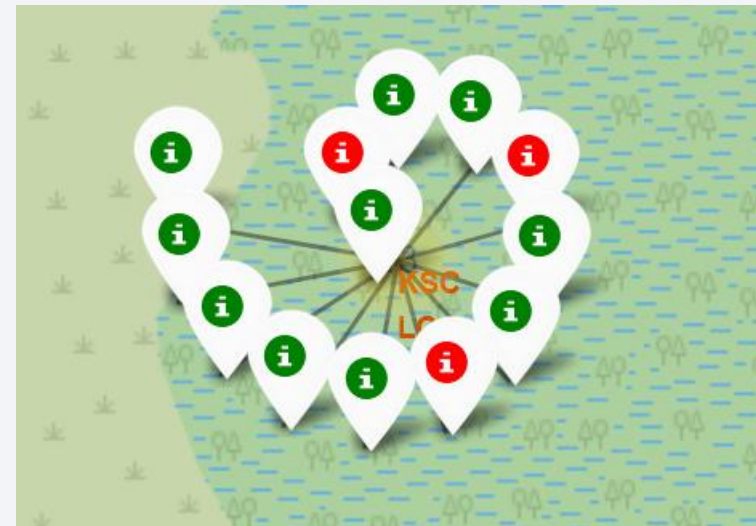
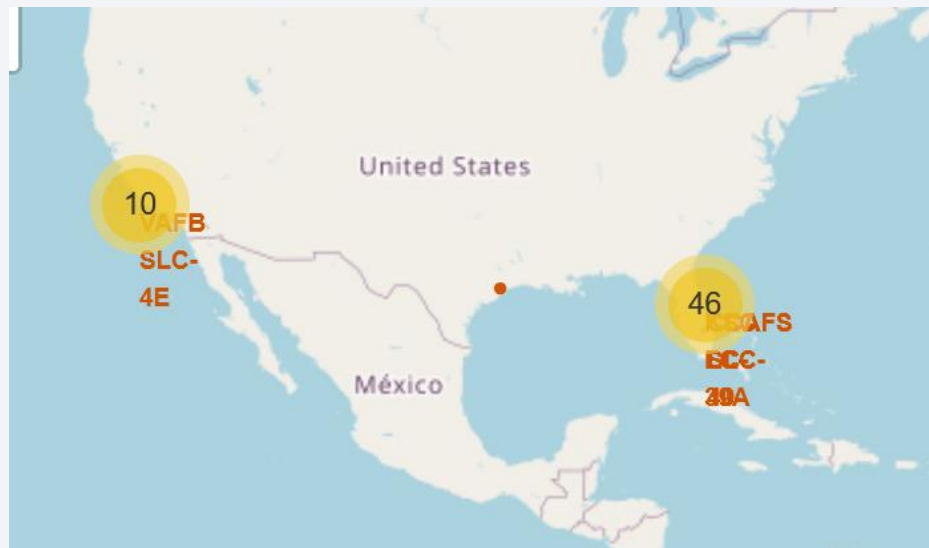




# Marker Clusters and labels

---

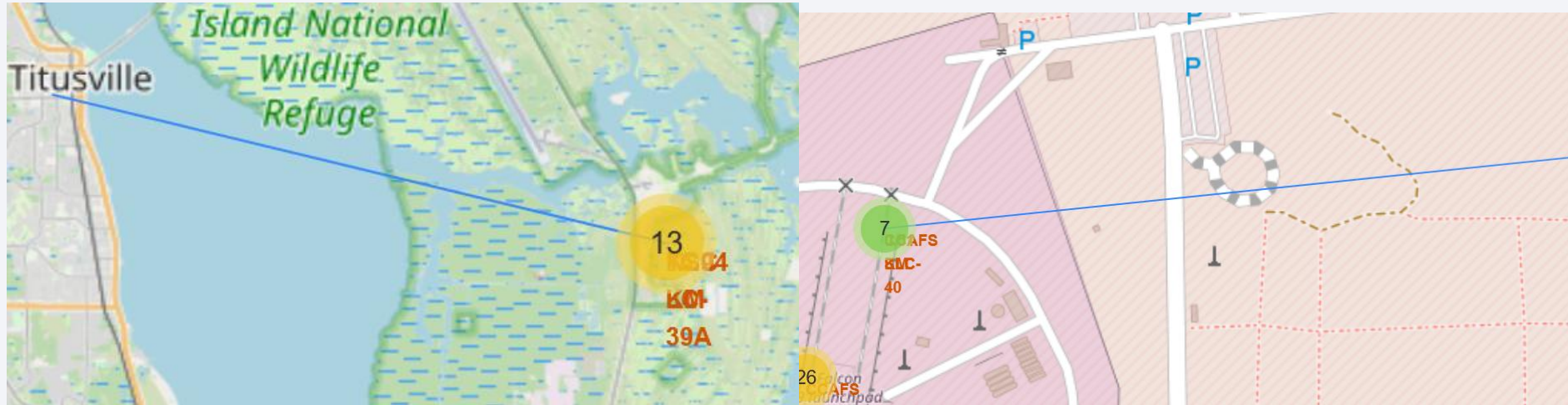
- Now more than the names its showing records in each launch site with markercluster and each record with marker to show their outcome



# Distances and lines

---

- A line connecting to the nearest city, railway and highway and other line connecting to the coastline







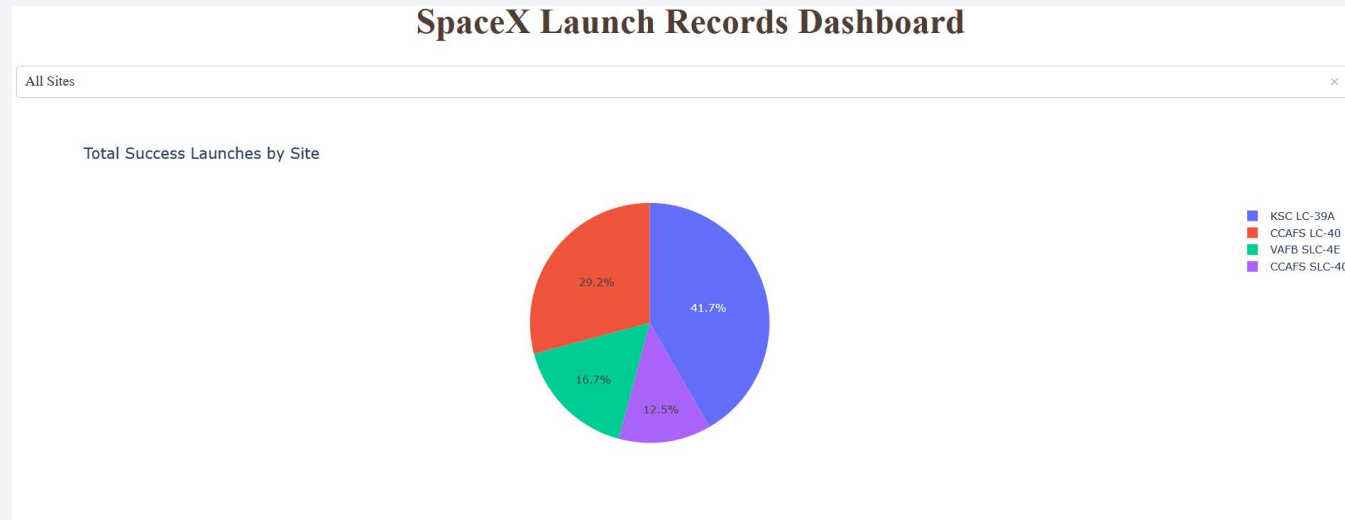
Section 4

# Build a Dashboard with Plotly Dash

# All Sites

---

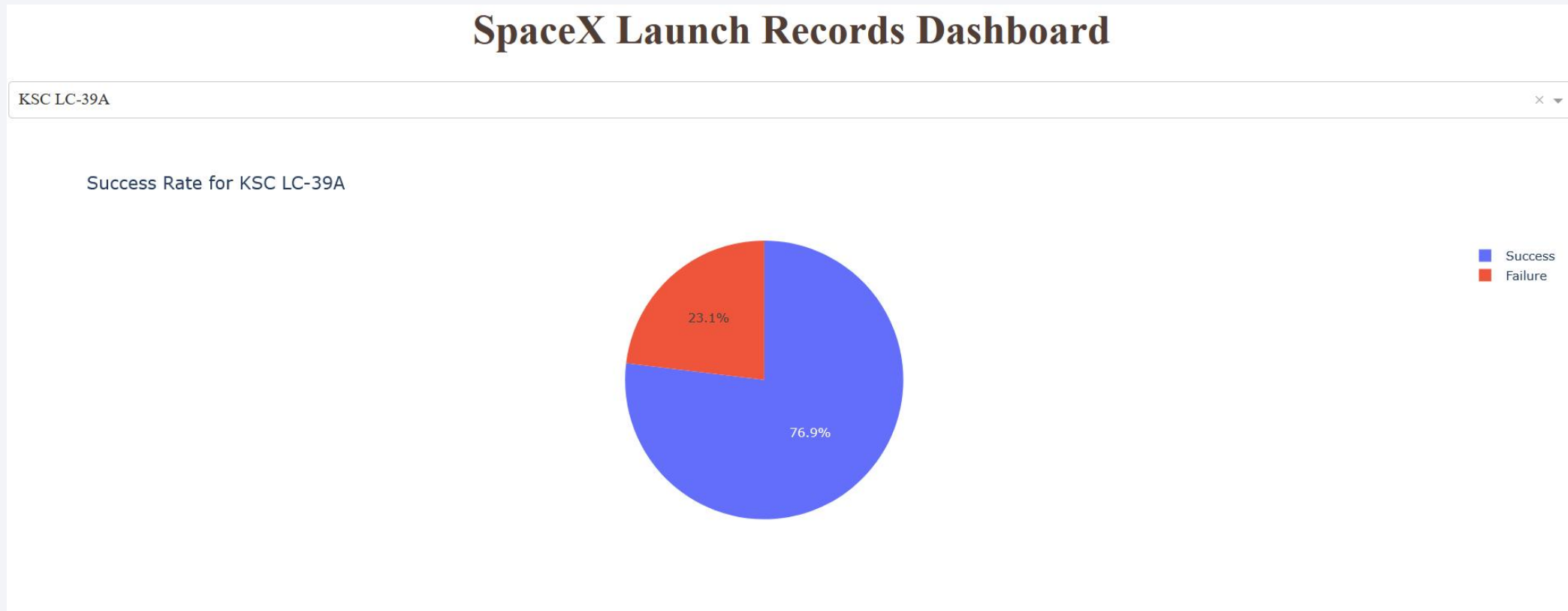
- Title, Dropdown Section and Pie Chart following the dropdown



# Highest Success Rate

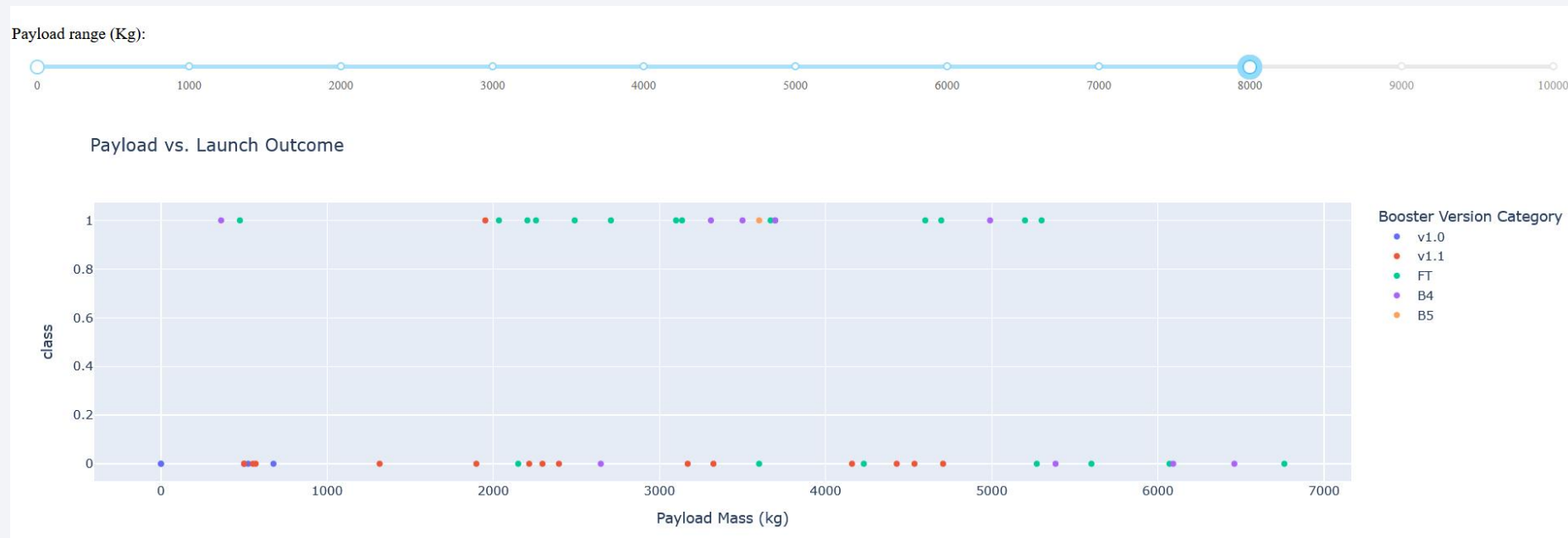
---

- Title, Dropdown Section selecting one site and Pie Chart Filtering to this one Launch Site



## <Dashboard Screenshot 3>

- RangeSlider to filter the Payload Mass range and Scatter Plot showing in this range the FT is often the booster version in succes outcomes





Section 5

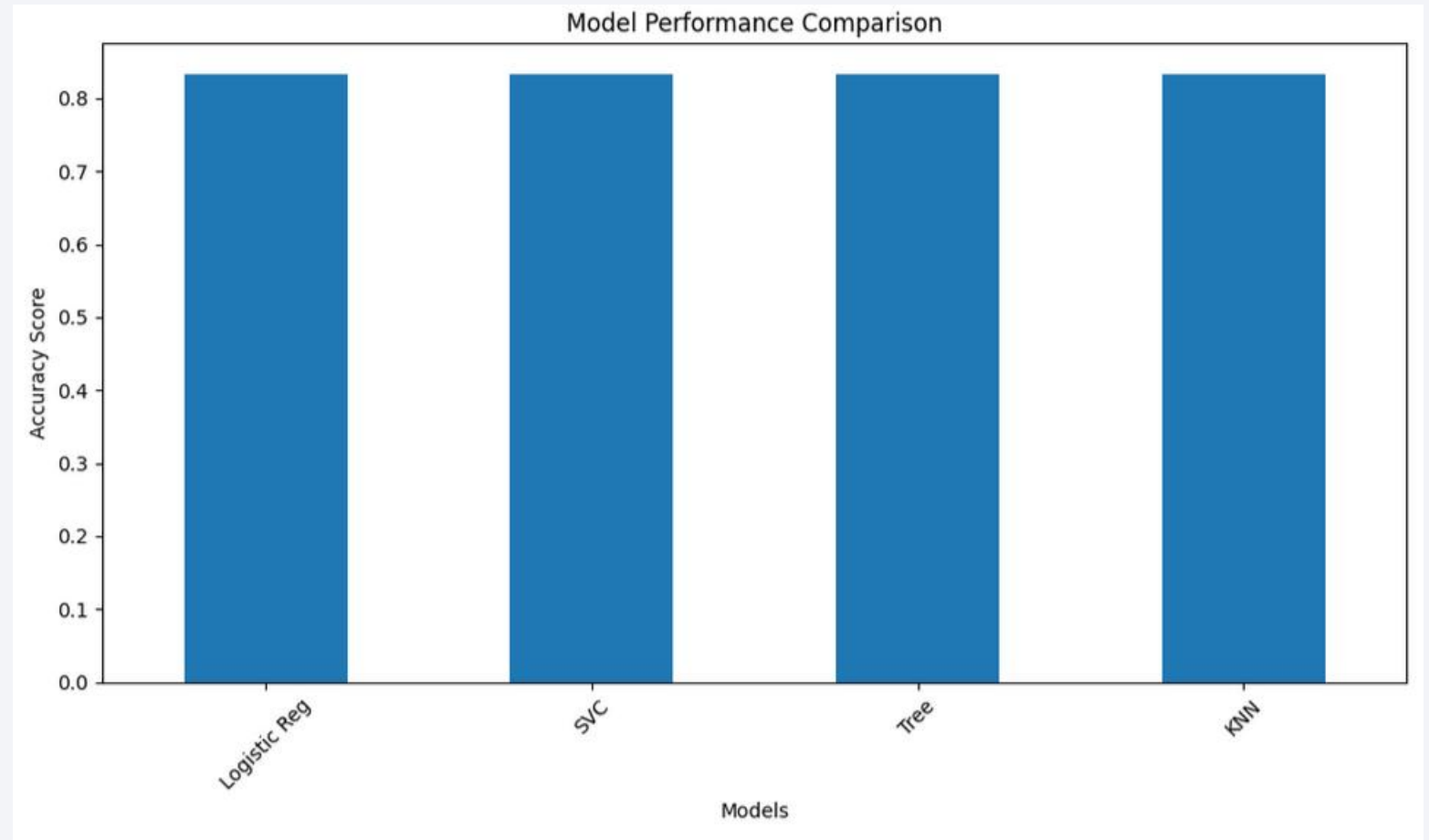
# Predictive Analysis (Classification)



# Classification Accuracy

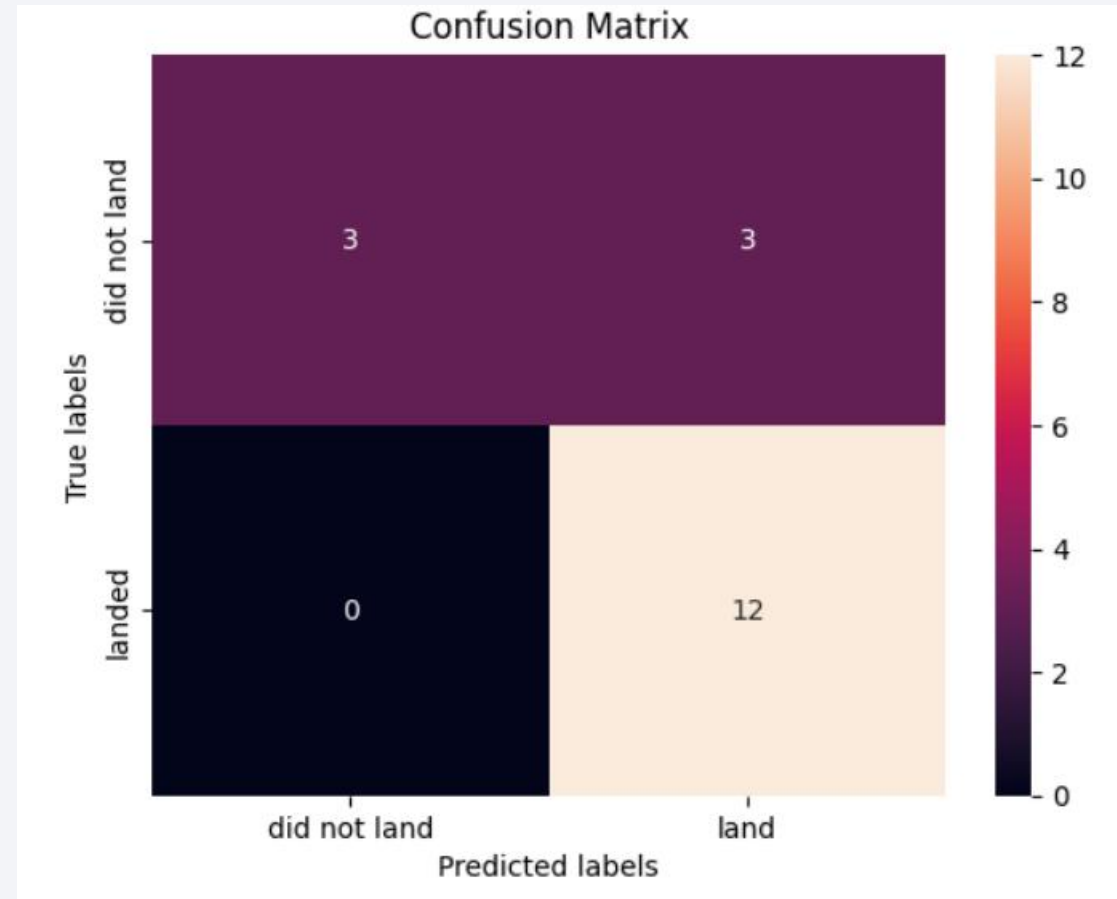
---

- They had all the same result



# Confusion Matrix

- They had all the same result



# Conclusions

---

- Independent on the model in this dataset the score value is the same and the confusion matrix too

# Appendix

---

- All the notebooks and pdf will be available in:  
<https://github.com/VicVald/Applied-Data-Science-Capstone-IBM/>

Thank you!

