IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

<Name>
<Date>

# Outline

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

# Executive Summary

- Summary of methodologies
- Summary of all results

# Introduction

- We will predict if the Falcon 9 first stage will land successfully : 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.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

# Methodology

# Methodology

## Executive Summary

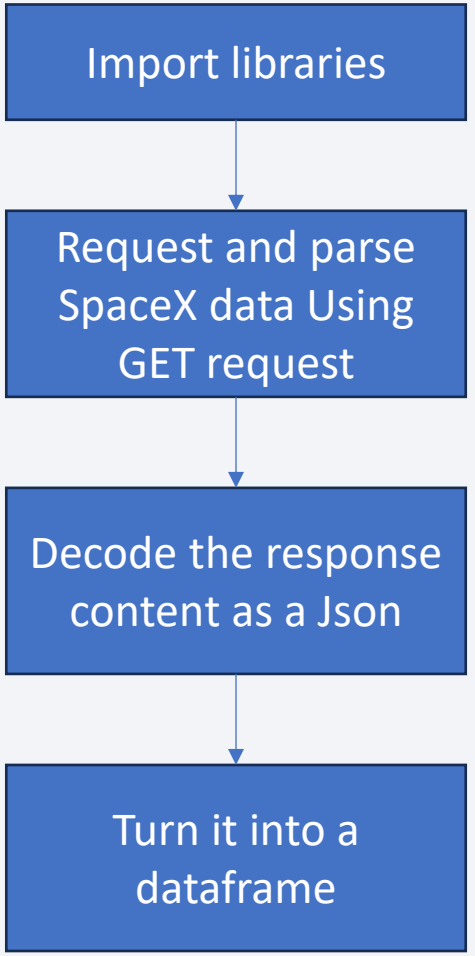
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

# Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

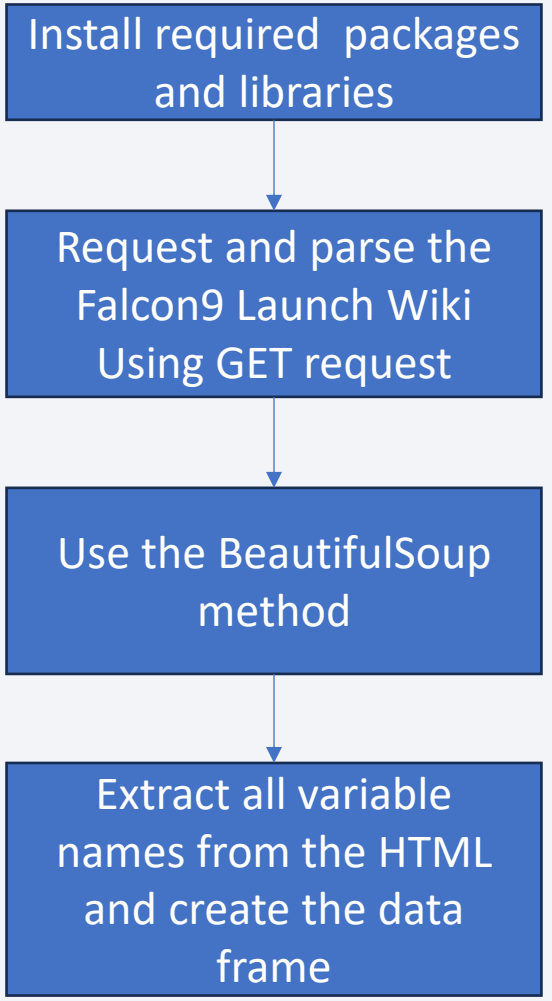
# Data Collection – SpaceX API

Capstone-project-module/jupyter-labs-spacex-data-collection-api.ipynb at main · YouennMARTIN/Capstone-project-module



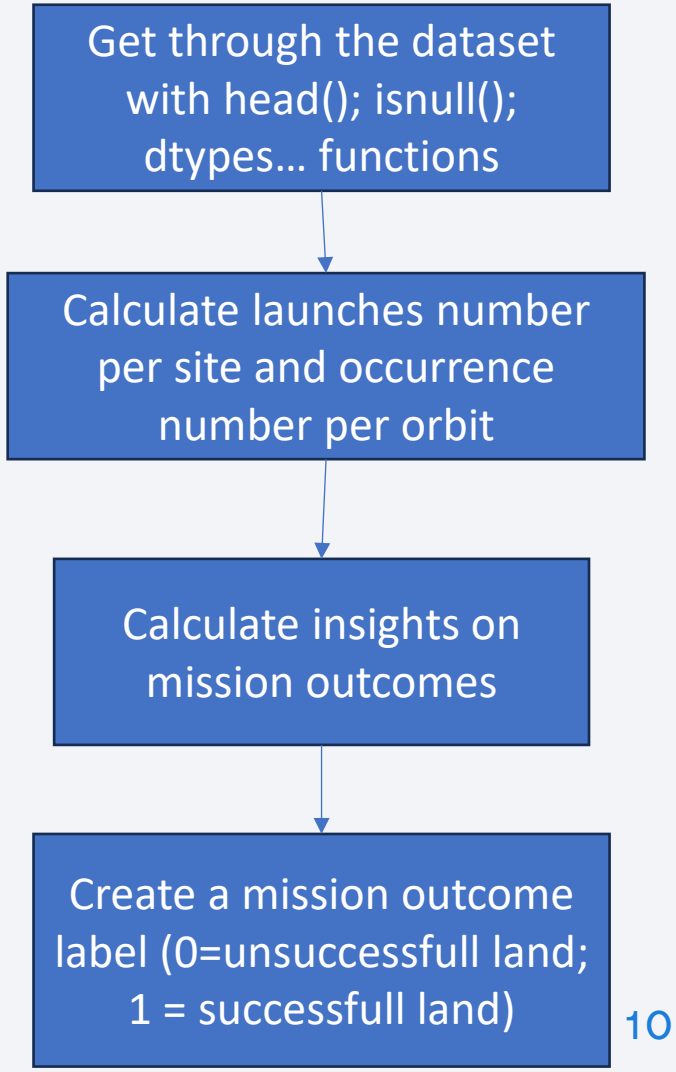
# Data Collection - Scraping

Capstone-project-module/jupyter-labs-webscraping.ipynb at main · YouennMARTIN/Capstone-project-module



# Data Wrangling

[Capstone-project-module/labs-jupyter-spacex-Data wrangling.ipynb at main · YouennMARTIN/Capstone-project-module]



# EDA with Data Visualization

- Summary of the relationship we assess by applying several plots:
  - FlightNumber VS PayloadMass
  - FlightNumber VS LaunchSite
  - PayloadMass VS LaunchSite
  - Orbit VS Class
  - FlightNumber VS Orbit
  - PayloadMass VS Orbit
  - Success Rate throughout years

Capstone-project-module/edadataviz (2).ipynb at main · YouennMARTIN/Capstone-project-module

# EDA with SQL

- Summary of the SQL queries:
  - Distinct
  - Limit 5
  - Sum(), Avg(), Min(), Max()
  - Between … and …
  - Count() and Groupby

- Capstone-project-module/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · YouennMARTIN/Capstone-project-module

# Build an Interactive Map with Folium

- The launch success rate may depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.
- To answer this question, we marked launch sites; success/failed launches for each site and calculate the distance between launch site to its proximities with markers; circles…

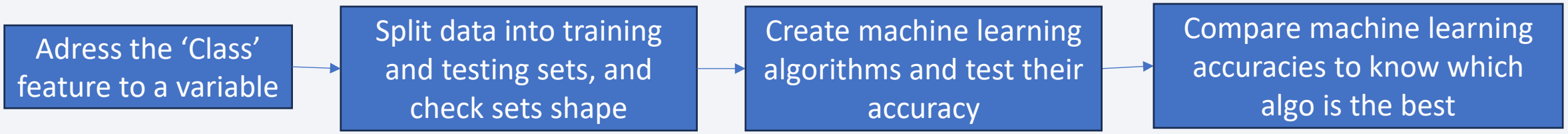
Capstone-project-module/lab_jupyter_launch_site_location.ipynb at main · YouennMARTIN/Capstone-project-module

# Build a Dashboard with Plotly Dash

- We first plotted a pie chart of the launch success depending on the launch sites. Then we created a scatterplot of the launch success rate per payload mass for each booster category. Those 2 interactive charts were plotted as we saw in the EDA work launch success rate varies among the launch site; payload mass and booster category variables.

- Capstone-project-module/spacex_dash_app.py at main · YouennMARTIN/Capstone-project-module

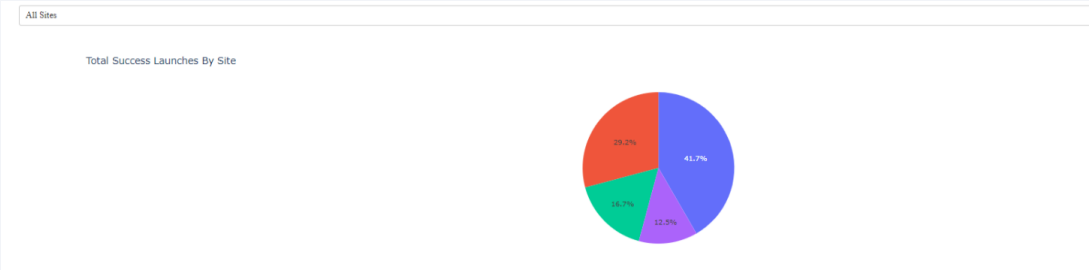
# Predictive Analysis (Classification)

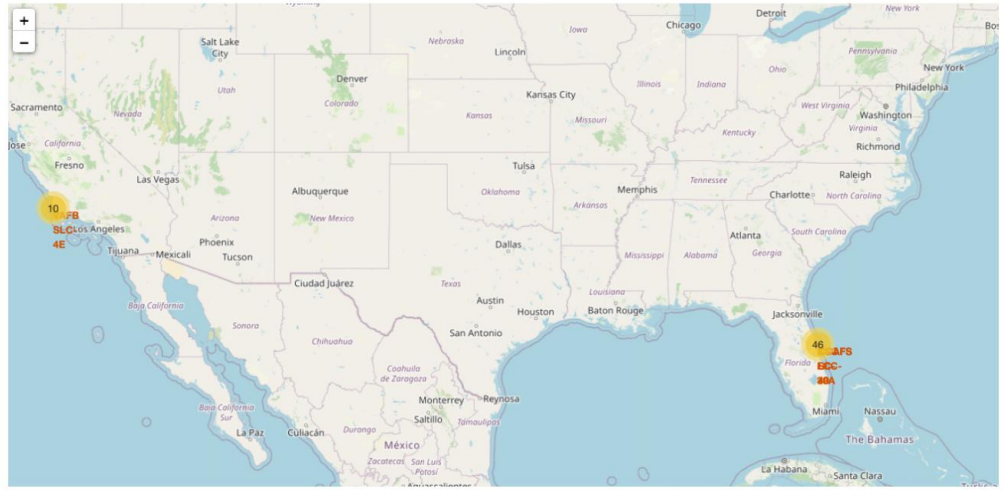


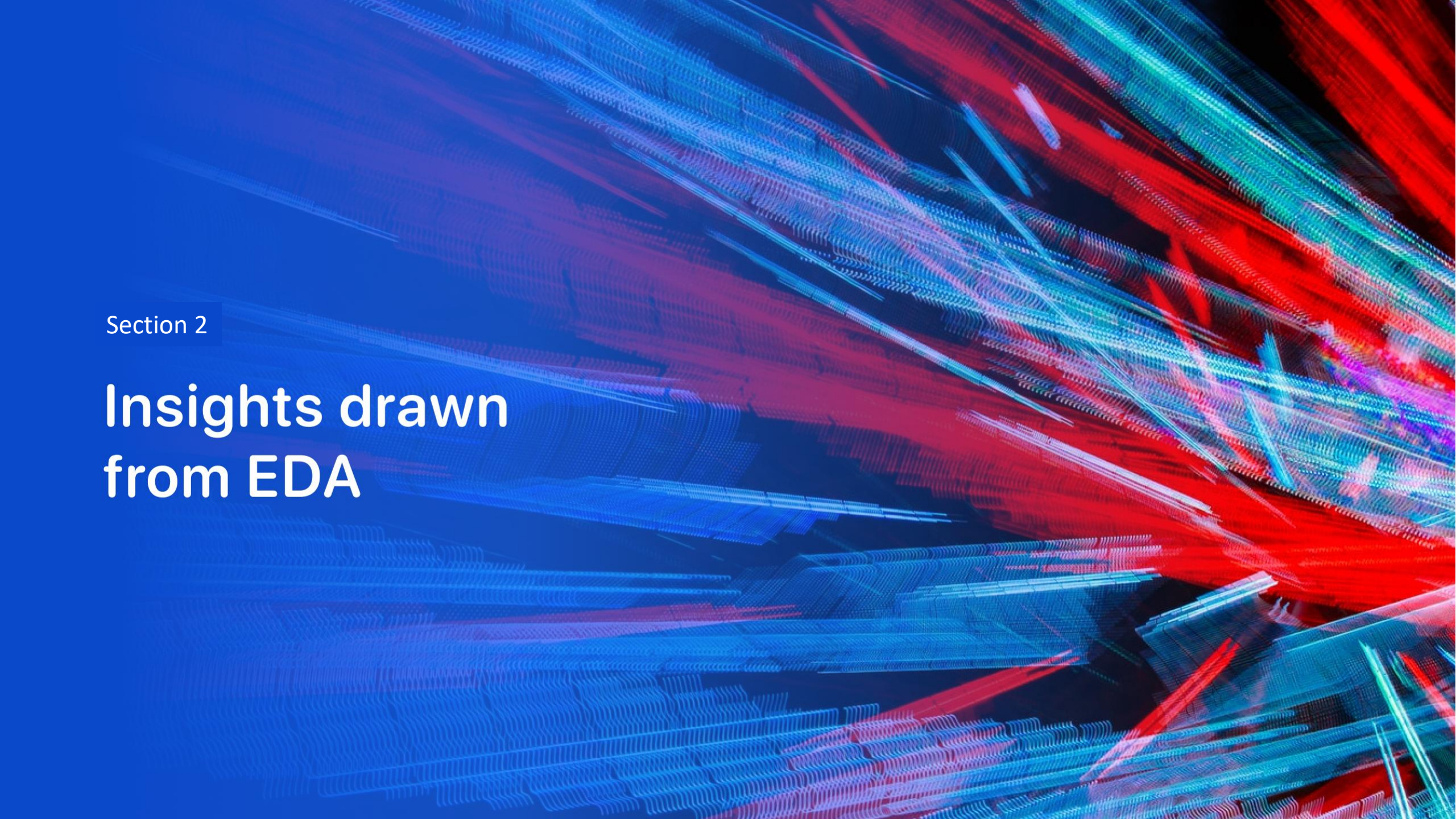

- Capstone-project-module/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb at main · YouennMARTIN/Capstone-project-module

# Results

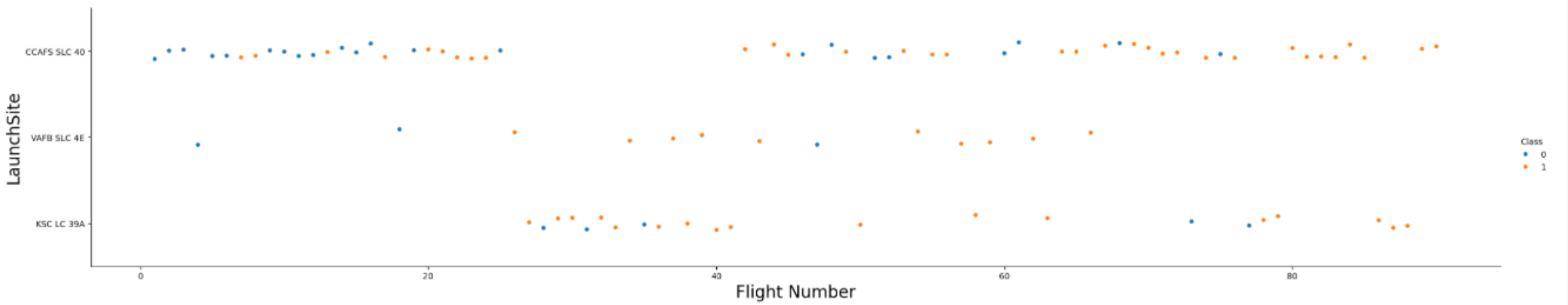
- Exploratory data analysis results : yep
- Interactive analytics demo in screenshots
- Predictive analysis results : algorithms had the same accuracy




Section 2
Insights drawn from EDA

# Flight Number vs. Launch Site



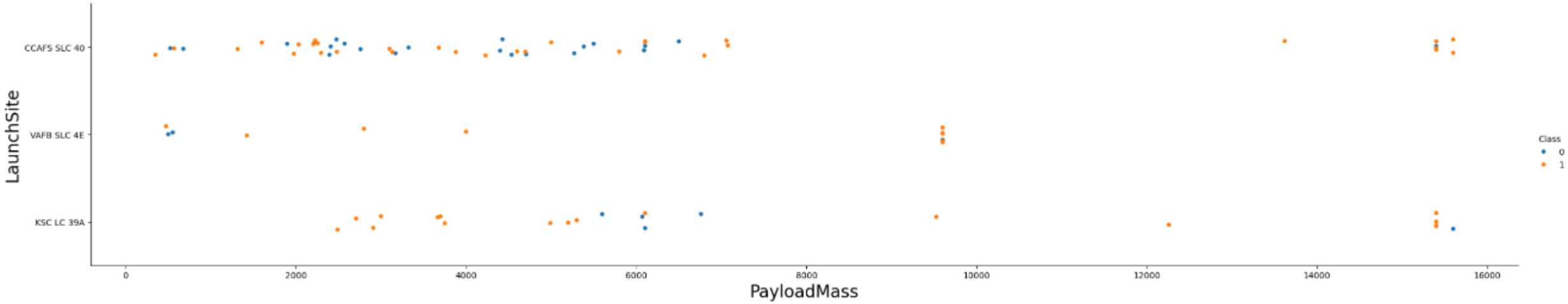

Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

Most of the first flights took place is CCAFS, and were a failure. There is an absence of flights amongh the mid section of total flights. The trend of successive flights continue after that middle section, but with a high success ratio this time for flights happening in CCAFS.

For the other two locations, the flights are very spread away, with the only exception of the mid section of total flights, most of which happened in the KSC location. KSC had any flight among the first group of flights.

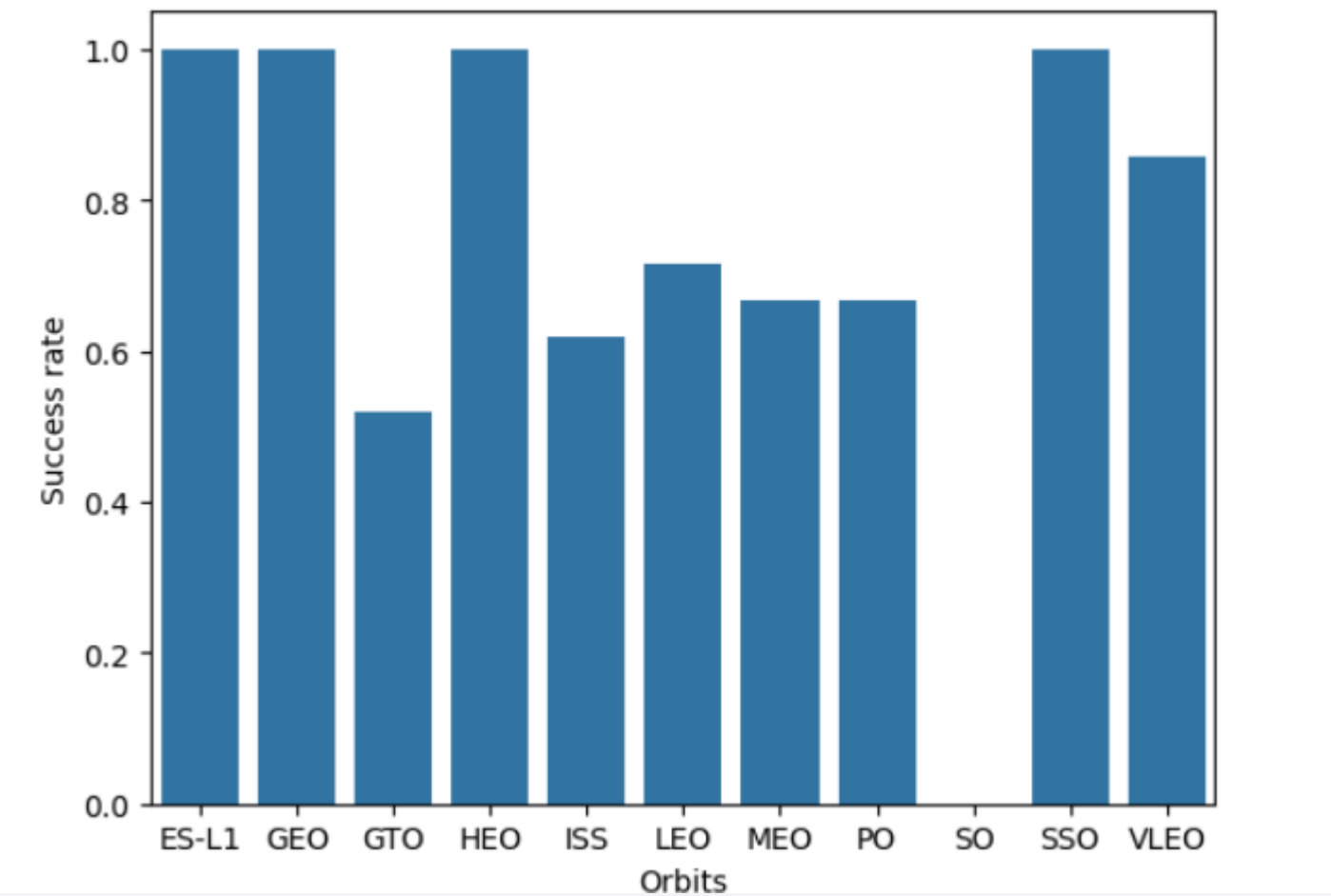
The success ratio for VAFB and KSC is similar, though with a low number of flights.

# Payload vs. Launch Site




Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

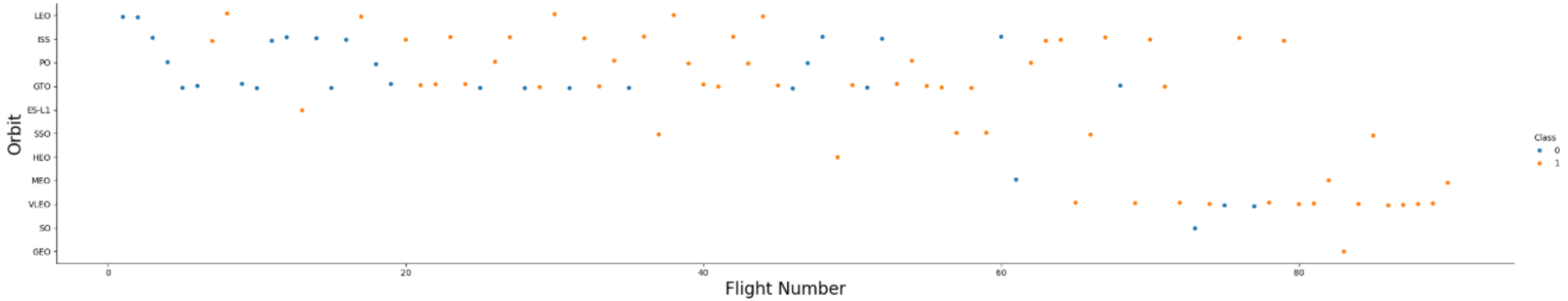
# Success Rate vs. Orbit Type




4 orbits had only successfull launches, where 5 orbits were close to 50% of success

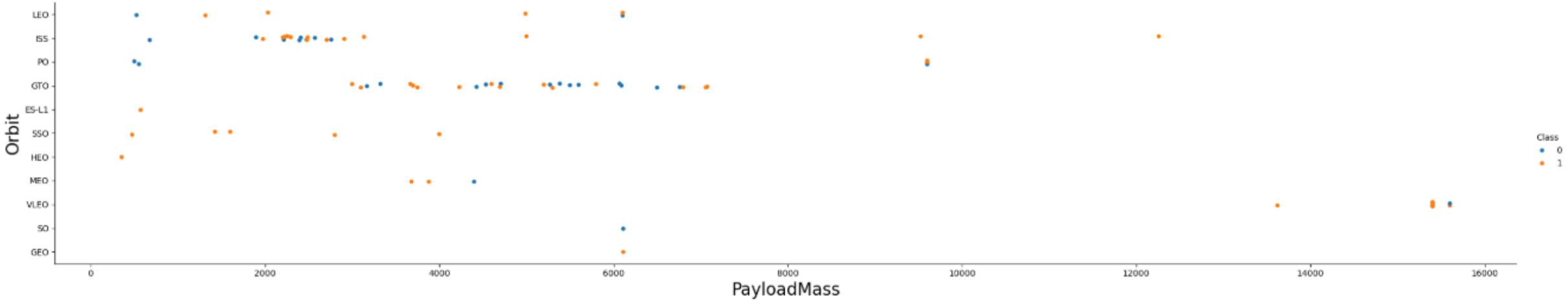
# Flight Number vs. Orbit Type

```
]: # Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```




You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

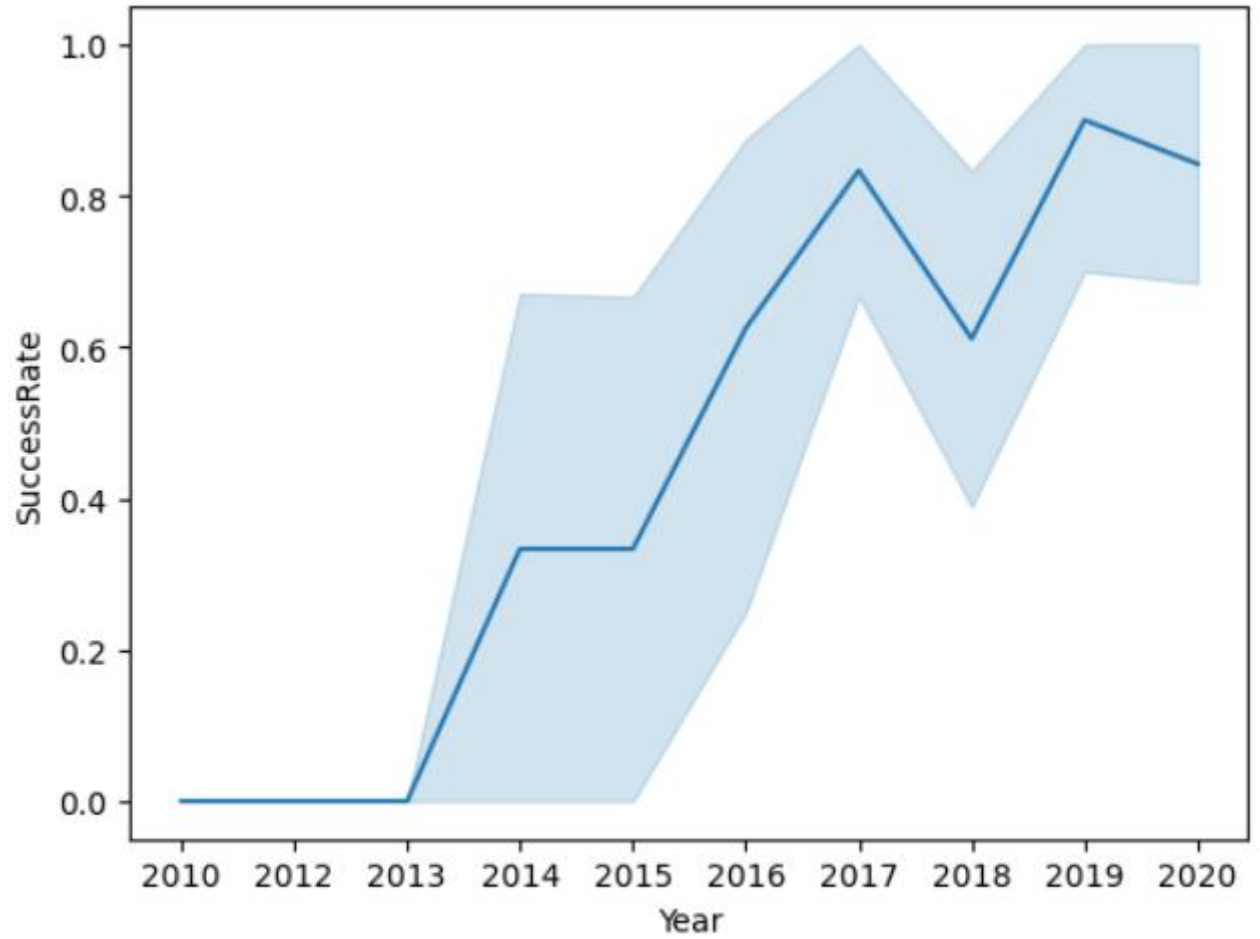
# Payload vs. Orbit Type




With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

# Launch Success Yearly Trend




you can observe that the sucess rate since 2013 kept increasing till 2020

# All Launch Site Names

## Task 1

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

```
12]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE
```

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

12]:

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

# Launch Site Names Begin with 'CCA'

## Task 2

Display 5 records where launch sites begin with the string 'CCA'

[72]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%%'  LIMIT 5
```

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

[72]:

| 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

## Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[27]: %sql SELECT SUM("PAYLOAD_MASS__KG_") AS PAYLOAD_MASS_TOTAL FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)'
```

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

[27]:

| PAYLOAD_MASS_TOTAL |
|---|
| 45596 |

# Average Payload Mass by F9 v1.1

### Task 4

Display average payload mass carried by booster version F9 v1.1

```
[28]: %sql SELECT AVG("PAYLOAD_MASS__KG_") AS F9v11_PAYLOAD_MASS_TOTAL FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1'
```

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

[28]:

| F9v11_PAYLOAD_MASS_TOTAL |
|---|
| 2928.4 |

# First Successful Ground Landing Date

## Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

*Hint:Use min function*

[92]:
```
%sql SELECT MIN("Date"), "Booster_Version","Launch_Site", "PAYLOAD_MASS_KG" FROM SPACEXTABLE WHERE ("Mission_Outcome" = 'Success')
```

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

[92]:

| MIN("Date") | Booster_Version | Launch_Site | "PAYLOAD_MASS_KG" |
|---|---|---|---|
| 2010-06-04 | F9 v1.0 B0003 | CCAFS LC-40 | PAYLOAD_MASS_KG |

# Successful Drone Ship Landing with Payload between 4000 and 6000

## Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[69]: %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE ("PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000) AND ("Mission_Outcome" = 'Success') AND ("Landing_Outcome" LIKE '%
```

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

[69]:

| Booster_Version |
|---|
| F9 FT B1020 |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

# Total Number of Successful and Failure Mission Outcomes

## Task 7

List the total number of successful and failure mission outcomes

70]:
```
%sql SELECT COUNT("Mission_Outcome") FROM SPACEXTABLE GROUP BY "Mission_Outcome"
```

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

70]:

| COUNT("Mission_Outcome") |
|---|
| 1 |
| 98 |
| 1 |
| 1 |

# Boosters Carried Maximum Payload

## Task 8

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

```
87]: %sql SELECT "Booster_Version", "PAYLOAD_MASS__KG_"  FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = 15600
```

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

87]:

| Booster_Version | PAYLOAD_MASS__KG_ |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1060.3 | 15600 |
| F9 B5 B1049.7 | 15600 |

# 2015 Launch Records

## Task 9

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.

**Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.**

4]:
```
%sql SELECT substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site FROM Spacextbl WHERE Landing_Outcome = "Failure (drone ship)" AND substr(Da
```

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

4]:

| 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

## Task 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.

[89]:
```
%sql SELECT COUNT("Landing_Outcome") FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY "Date" DESC
```

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

[89]:

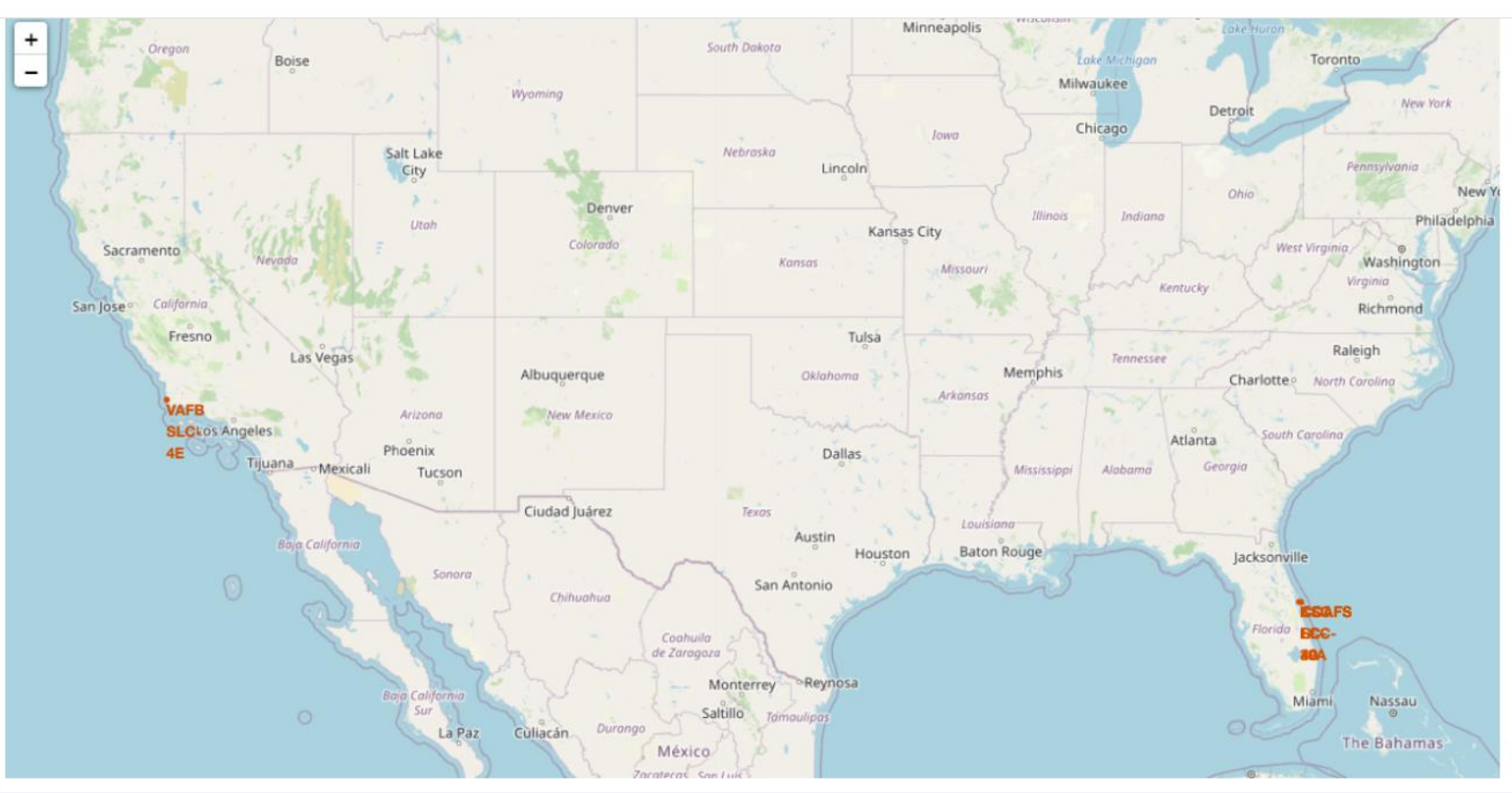
| COUNT("Landing_Outcome") |
|---|
| 31 |

Section 3

# Launch Sites Proximities Analysis

# US Map with launch sites location

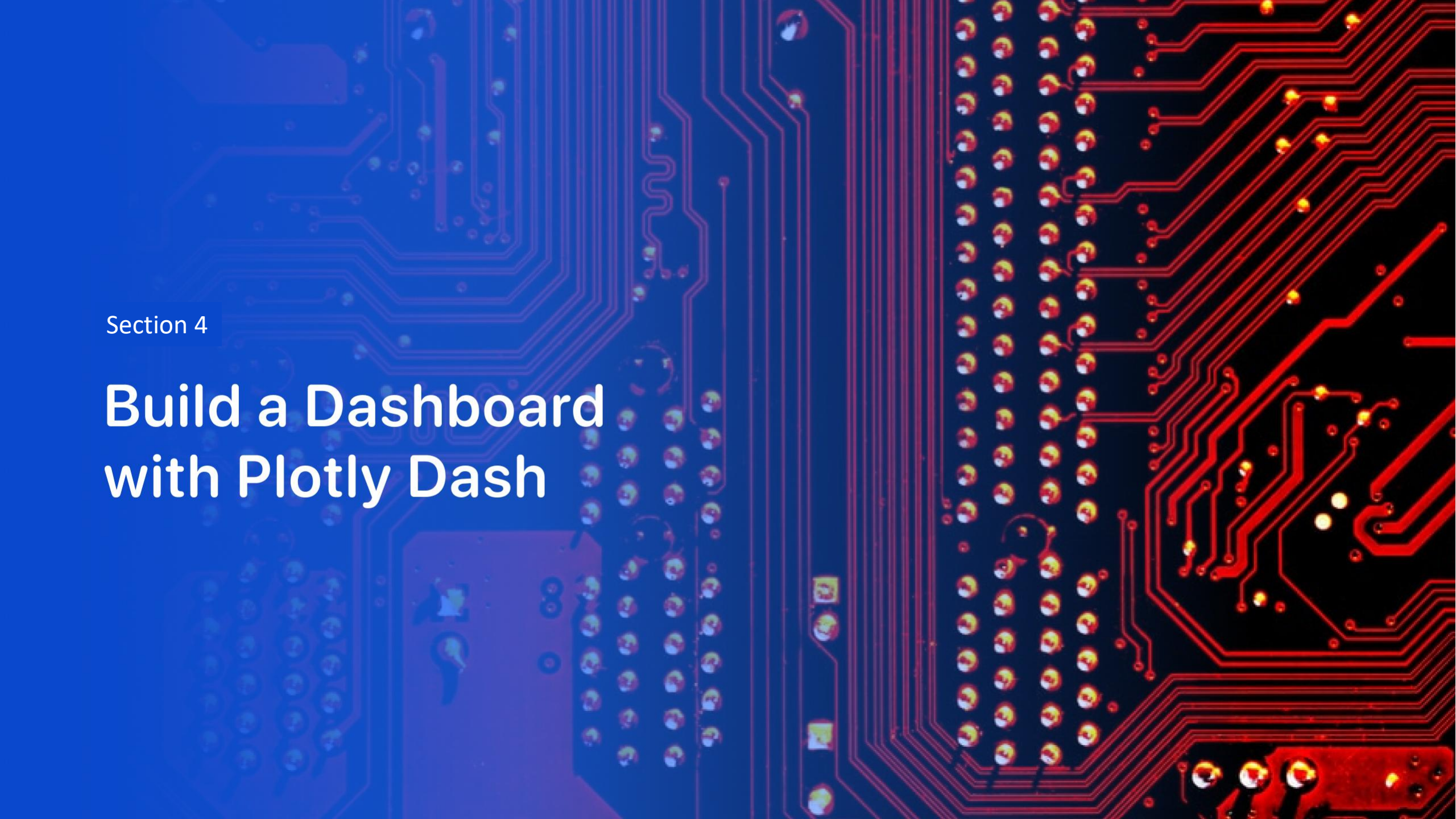
Launch sites found in 2 different places, Florida and California



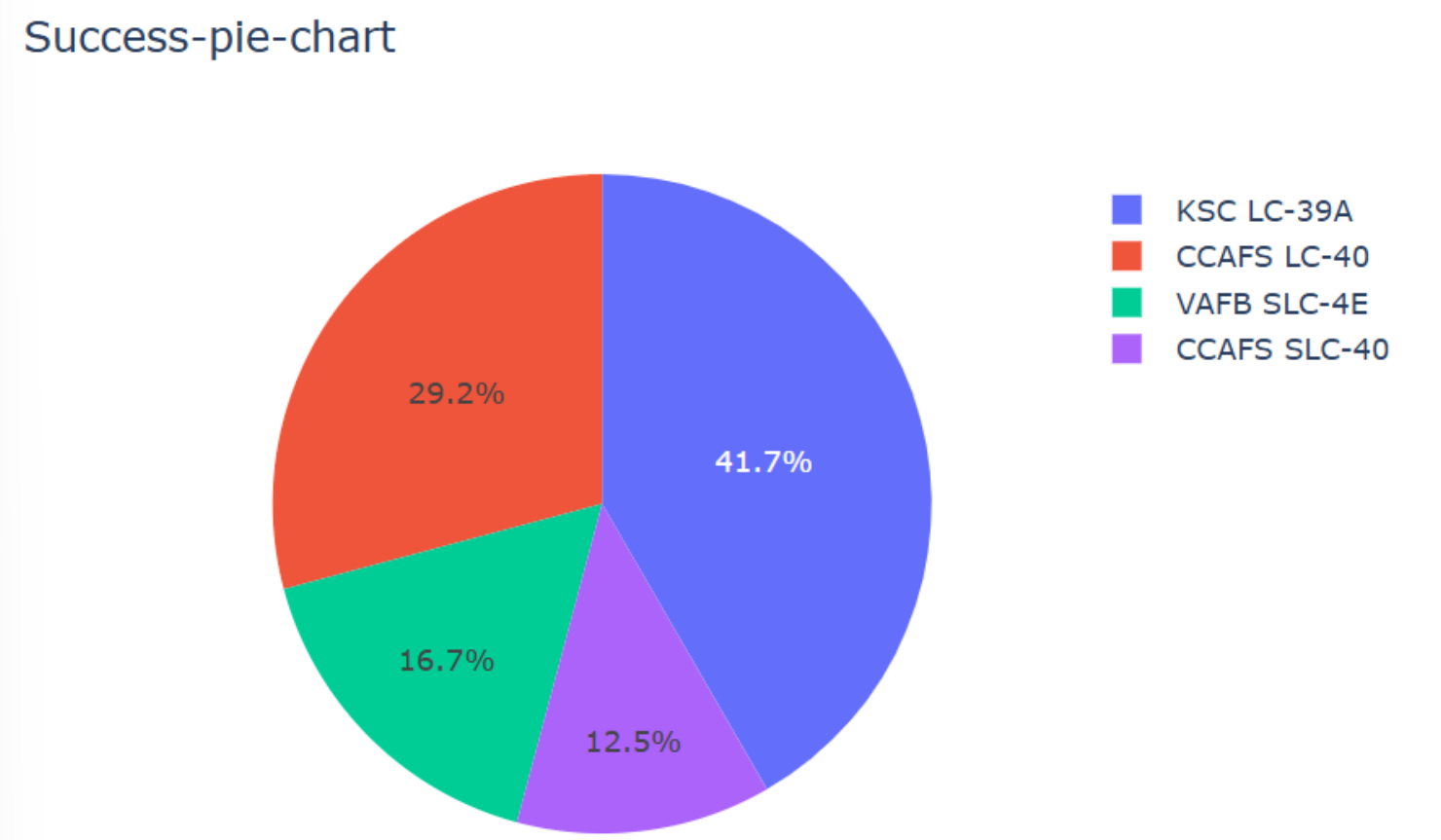
# US Map with color labeled sites location

Launch site success rate printed as green for successful launch and red for failed ones.

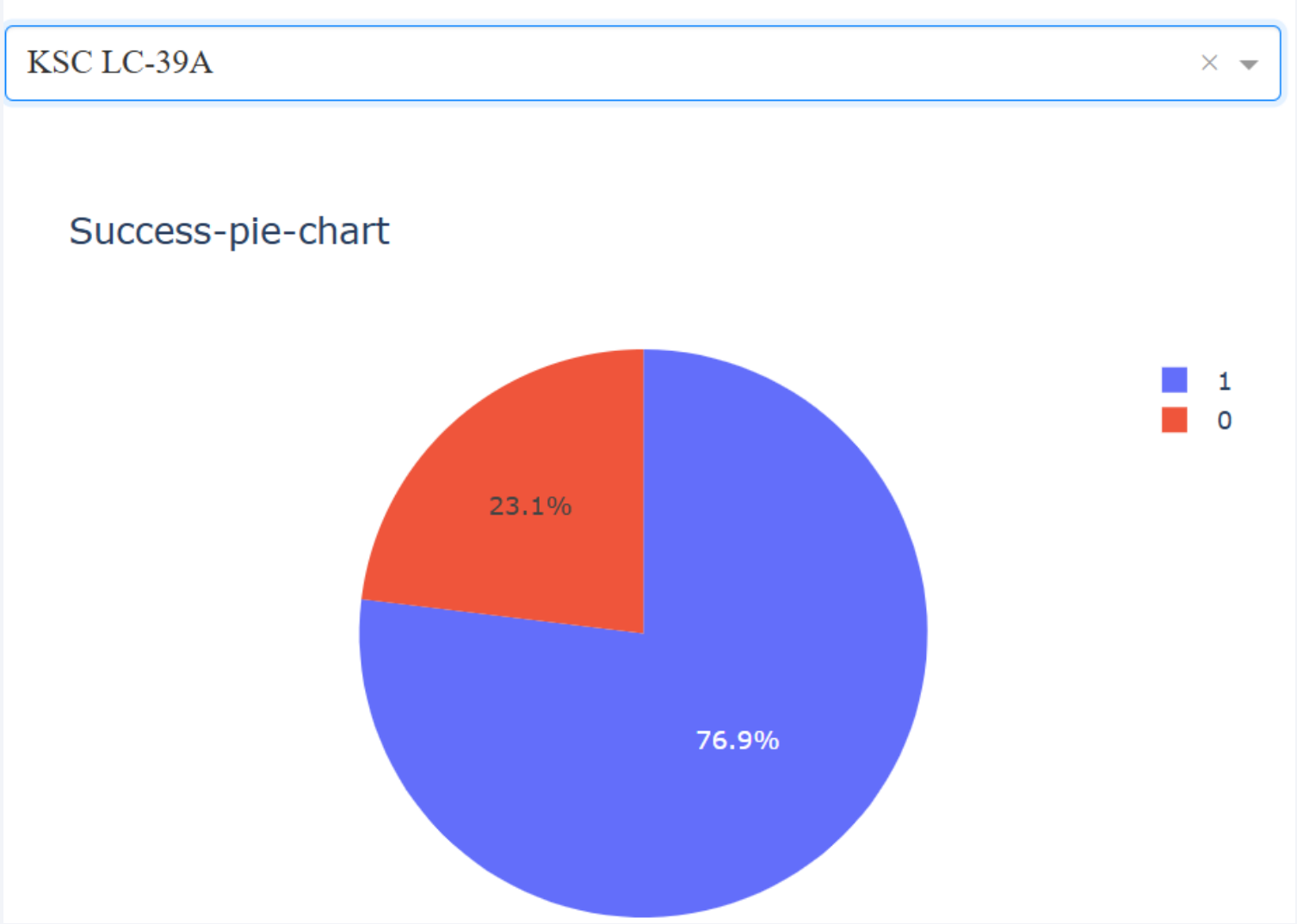
# Launch sites proximities


Section 4
Build a Dashboard with Plotly Dash

# Launch success count for all sites



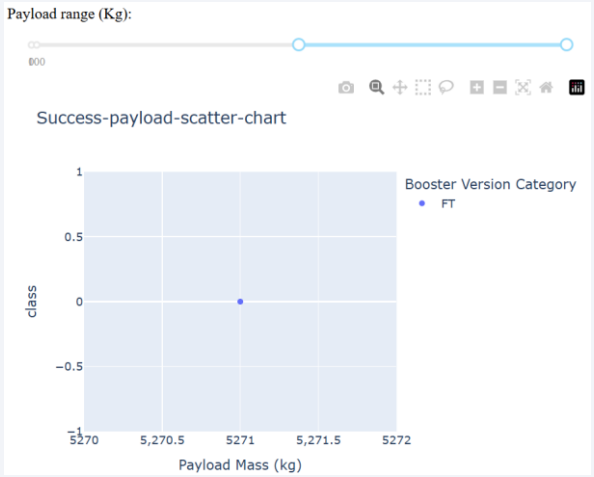
# Most successful launch site



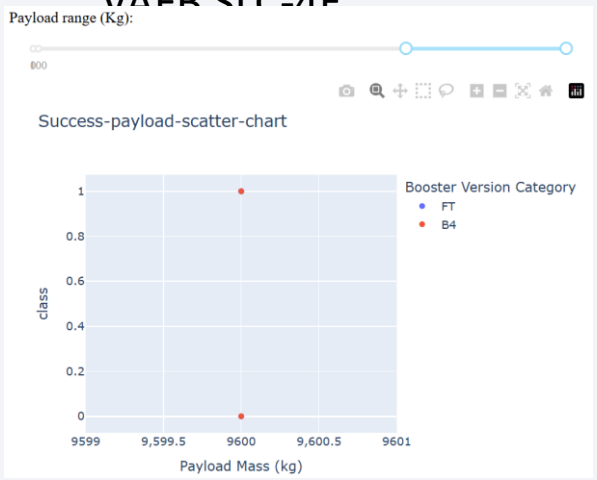
# Payload vs Launch Outcome plot

KSC LC-39A

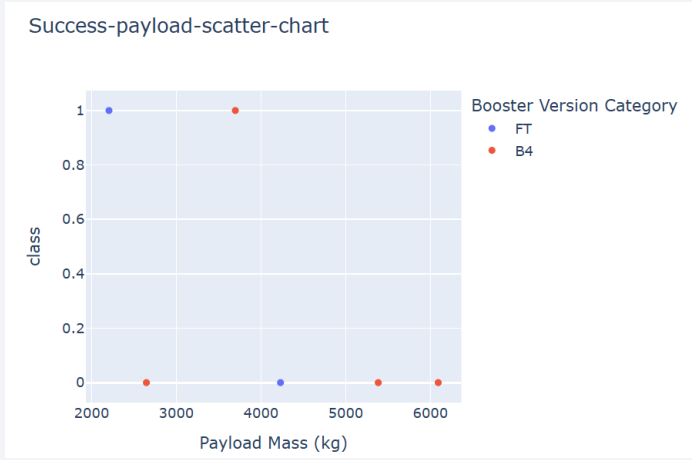

CCAFS LC-40




VAFB SLC-4E




CCAFS SLC-40

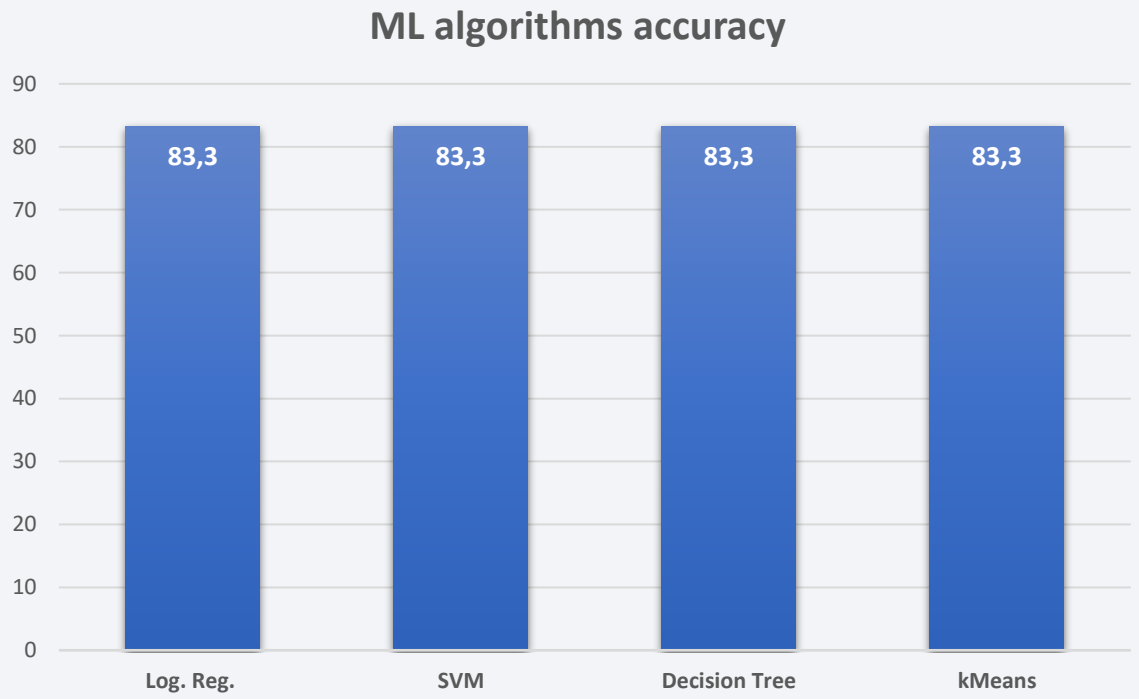


Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

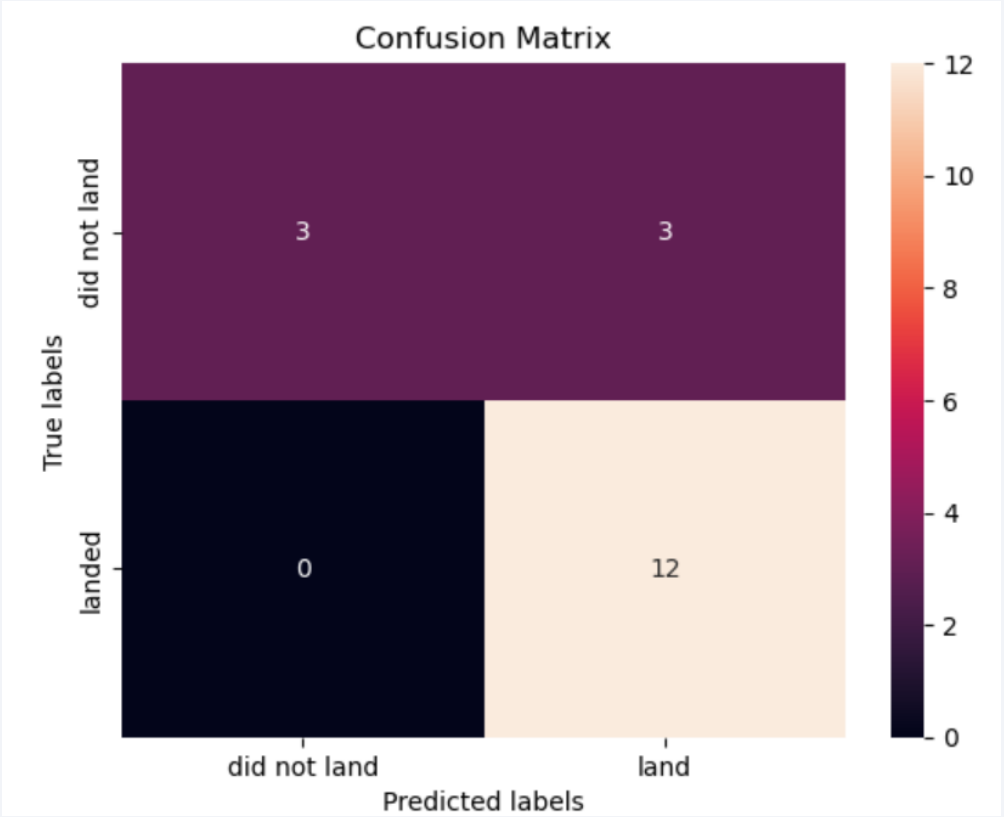
(Made on Excel because the notevook just stopped working without any explanation)



# Confusion Matrix

- In this DT algorithm, 50% of the unsuccessful launch were actually classified as unsuccessful launches, where all successful launches were well classified

Decision tree confusion matrix



# Conclusions

- This long work highlighted the dependency of launch site success rate with their location; payload mass; booster category, as illustrated during the EDA; Dash; Folium tasks.

# Thank you!