IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

James Goedmakers
11/28/2025

# Outline

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

# Executive Summary

- Summary of methodologies: SpaceX Launch Data was collected via SpaceX api and webscraping. Launch Data organized into Pandas dataframe, visualized with Seaborn, geospatial visualizations with Folium, web dashboard visuals with Plotly Dash. Classification models trained to predict launch stage 1 landing outcome using SciKit Learn.
- Summary of all results: General trend is higher landing success rate at higher launch numbers. Launch outcome prediction with Logistic Regression, Support Vector Machines, and K-Nearest Neighbor share similar test accuracy performance around 83%, with Decision Tree demonstrating lower performance at 78% test accuracy.

# Introduction

- Project background:

  SpaceX is able to provide commercial space launches for lower prices by landing and reusing their stage 1 booster rockets. Not all of these launches land their stage 1 boosters successfully, and there are many launch variables that may influence whether or not the booster landing will be successful.

- Problems to answer:

  1) What launch variables correlate with a successful/failed landing?

  2) Can we predict a succesful/failed landing using past and present launch data?

Section
1

# Methodology

# Methodology

## Executive Summary

Data collection methodology:

- SpaceX Launch Data was collected using a combination of the SpaceX API and Webscraping (Python BeautifulSoup)

Perform data wrangling

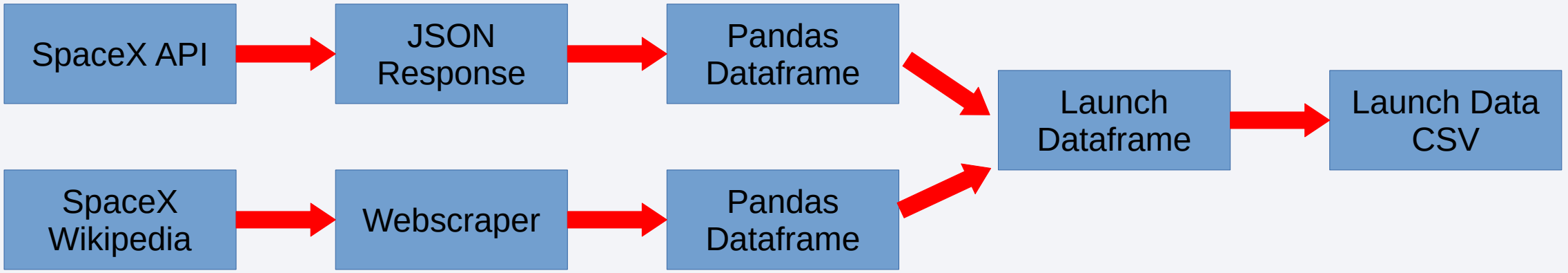
- Launch data was organized into dataframes and cleaned using Pandas

Perform exploratory data analysis (EDA) using visualization and SQL

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Launch data was divided into train/test splits, four classification models were tested (Logistic Regression, SVC, Decision Tree, KNN). Parameter tuning was performed using GridSearchCV

# Data Collection

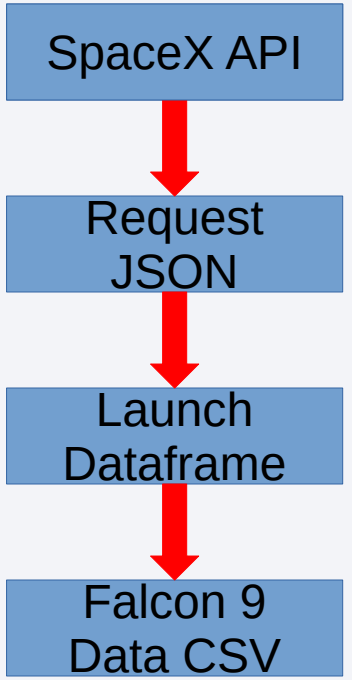
SpaceX Launch Data was collected using a combination of the SpaceX API and Webscraping (Python BeautifulSoup):



# Data Collection – SpaceX API

- JSON response received by SpaceX API, converted into Pandas dataframe, and filtered to only include Falcon 9 launches.
- GitHub URL of the completed SpaceX API calls notebook:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/1_SpaceX_Data_Collection_api.ipynb

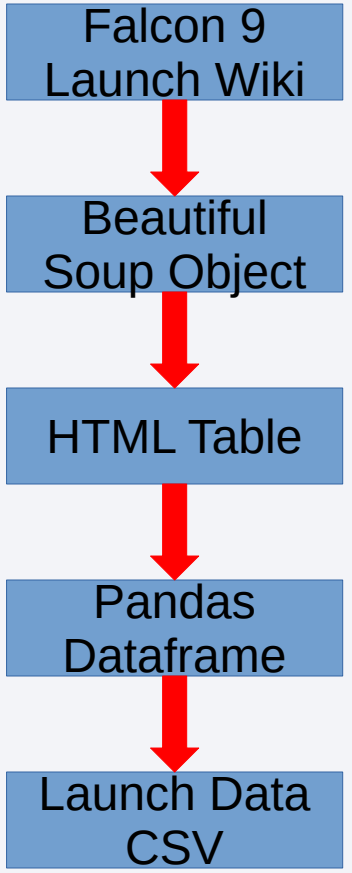


# Data Collection - Scraping

- The Falcon 9 Launch Data is retrieved from the Wikipedia page containinga  Launch Data HTML table. Python BeautifulSoup library is used to extract this webscraped HTML table, save it to a dataframe, and export to a csv file

- GitHub URL of the completed web scraping notebook:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/2_SpaceX_Webscraping.ipynb



# Data Wrangling

- ‘Class’ column added based off of launch outcome. 0 for failed landing, 1 for successful landing

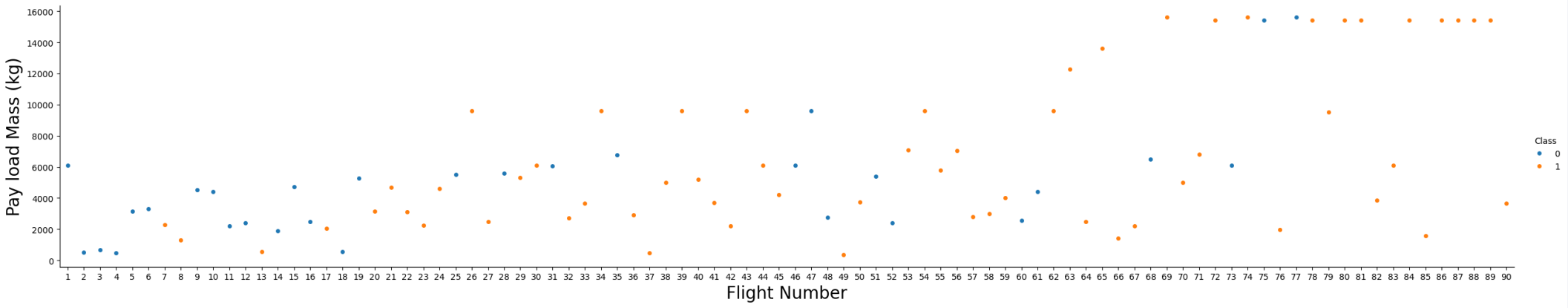
- GitHub URL of your completed data wrangling related notebooks: https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/3_SpaceX_Data_Wrangling.ipynb

| Date | BoosterVersion | PayloadMass | Orbit | LaunchSite | Outcome | Flights | GridFins | Reused | Legs | LandingPad | Block | ReusedCount | Serial | Longitude | Latitude | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | Falcon 9 | 6104.959412 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0003 | -80.577366 | 28.561857 | 0 |
| 2012-05-22 | Falcon 9 | 525.000000 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0005 | -80.577366 | 28.561857 | 0 |
| 2013-03-01 | Falcon 9 | 677.000000 | ISS | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0007 | -80.577366 | 28.561857 | 0 |
| 2013-09-29 | Falcon 9 | 500.000000 | PO | VAFB SLC 4E | False Ocean | 1 | False | False | False | NaN | 1.0 | 0 | B1003 | -120.610829 | 34.632093 | 0 |
| 2013-12-03 | Falcon 9 | 3170.000000 | GTO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B1004 | -80.577366 | 28.561857 | 0 |

# EDA with Data Visualization

- Charts plotted to visualize relationships between independent launch variables (Flight Number, Payload Mass, Orbit type, Date, …) and landing outcome (‘Class’, 0=failed landing, 1=successful landing)
- GitHub URL of your completed EDA with data visualization notebook:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/5_SpaceX_Data_Visualization.ipynb



# EDA with SQL

- Using bullet point format, summarize the SQL queries you performed:
  - Display names of unique launch sites
  - Display launch sites beginning with ‘CCA’
  - Display total payload mass from NASA (CRS)
  - Display average payload mass from booster F9 v1.1
  - Select the date of first successful ground pad landing
  - Select boosters successfully landing on drone ship
  - List total number of successful and failure mission outcomes
  - List booster versions that have carried max payload mass
  - List launch records from a specific date
  - List launch records from a date range
- GitHub URL of your completed EDA with SQL notebook:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/4_SpaceX_EDA_SQL.ipynb

# Build an Interactive Map with Folium

- Shaded circles plotted to represent unique SpaceX launch sites, colored markers used to represent individual launch attempts (green=success, red=failed landing). Lines plotted to represent distance from launch sites to notable landmarks (cities, coastline, roads, ...)
- GitHub URL of your completed interactive map with Folium map:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/6_SpaceX_Launch_Sites_Folium.ipynb

# Build a Dashboard with Plotly Dash

- Dashboard dropdown list added to select all or one of the SpaceX launch sites. Pie chart added to visualize the number of successful launches. Scatter plot added to visualize the relationship between payload and launch success.
- GitHub URL of your completed Plotly Dash lab:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/7_spacex-dash-app.py

# Predictive Analysis (Classification)

- Launch data divided into training and testing data. Four classification models tested (Logistic Regression, SVM, Decision Tree, KNN) with GridSearchCV to optimize model parameters. Model accuracy evaluated on test data.
- GitHub URL of your completed predictive analysis lab:

  https://github.com/jgoedmakers/IBMDataScienceProject/blob/main/8_SpaceX_Machine_Learning.ipynb

# Results

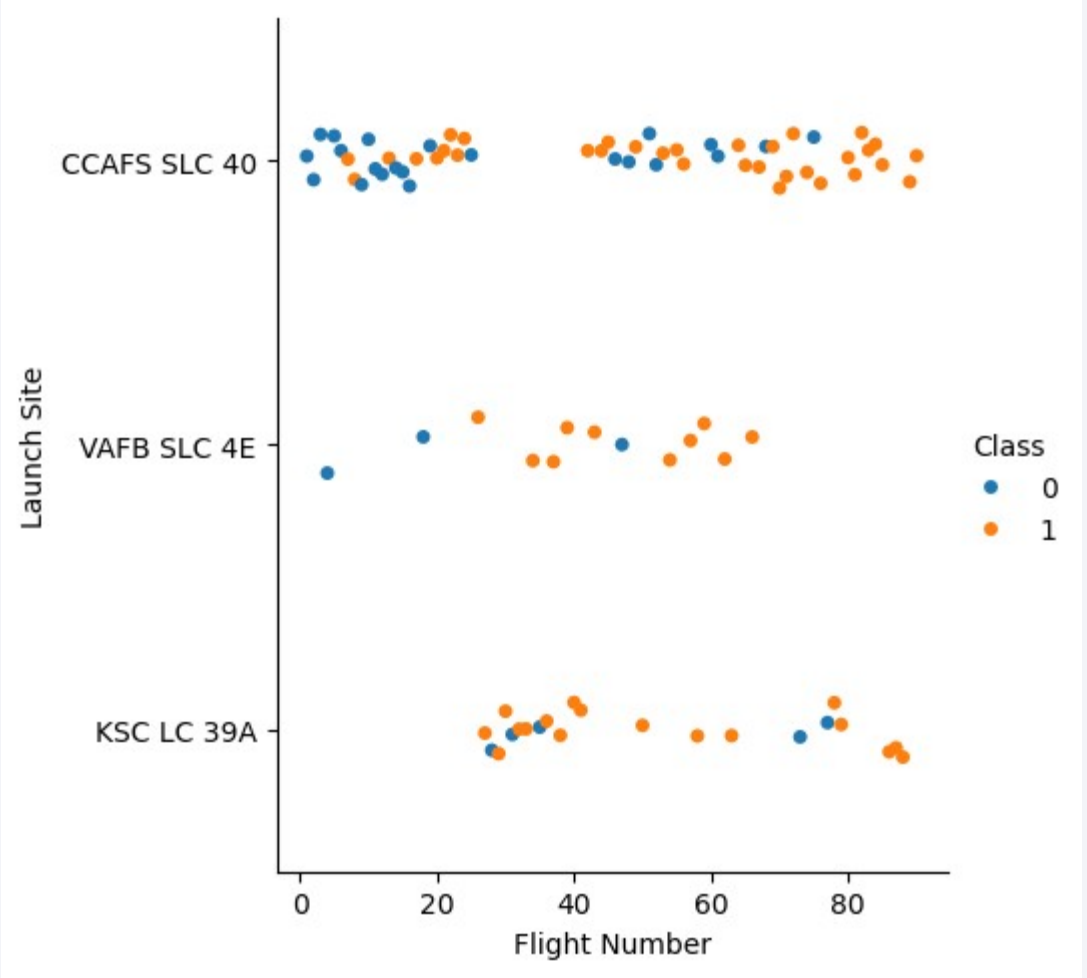
- Exploratory data analysis results: Flight number correlated strongly with launch success. Higher flight numbers had more successful launches.
- Interactive analytics demo in screenshots
- Predictive analysis results: LR, SVC, and KNN all achieved 83% test accuracy in predicting launch success, Decision Tree achieved 78% test accuracy

Section 2

# Insights drawn from EDA

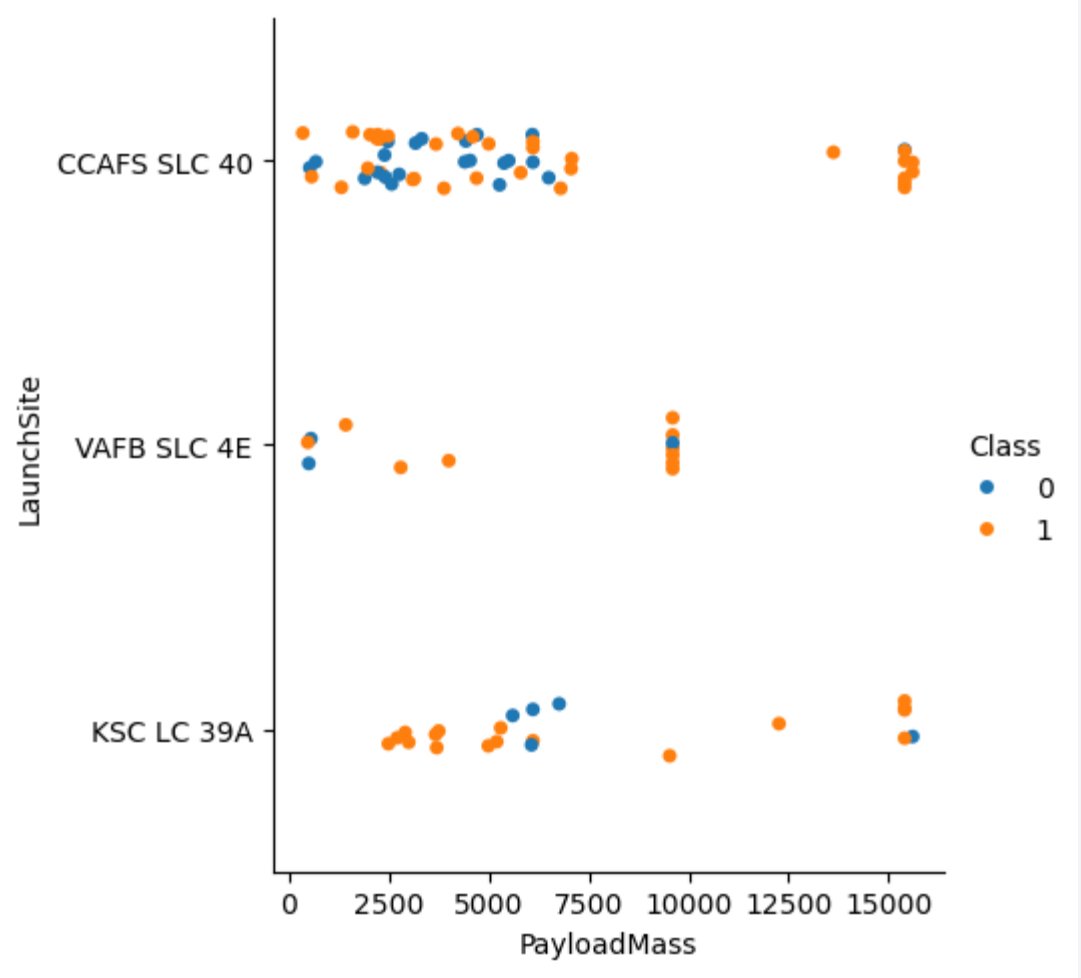
# Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site
- CCAFS had more early (low flight number) launches, KSC had more later launches
- Launch success appears more strongly linked to flight number, not launch site



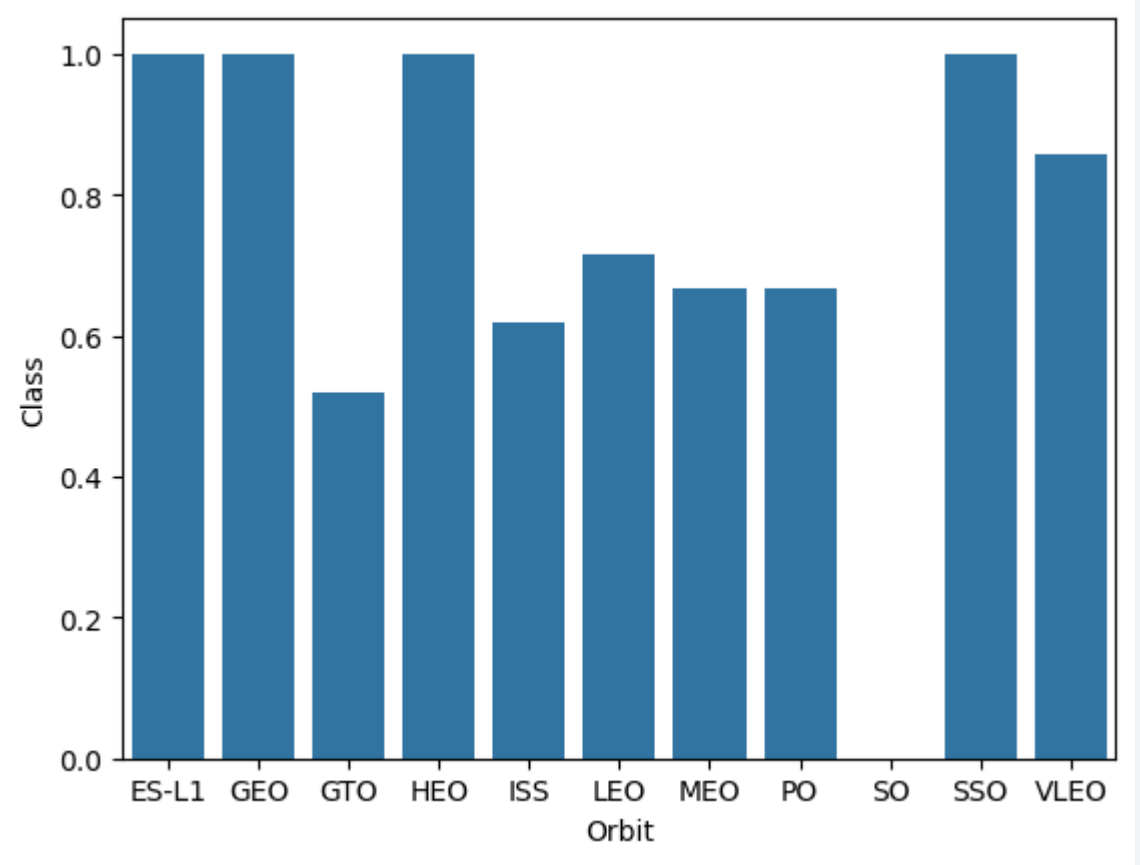
# Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- Payload mass doesn’t appear to correlate strongly with launch success
- VAFB launches appear to have a lower max payload



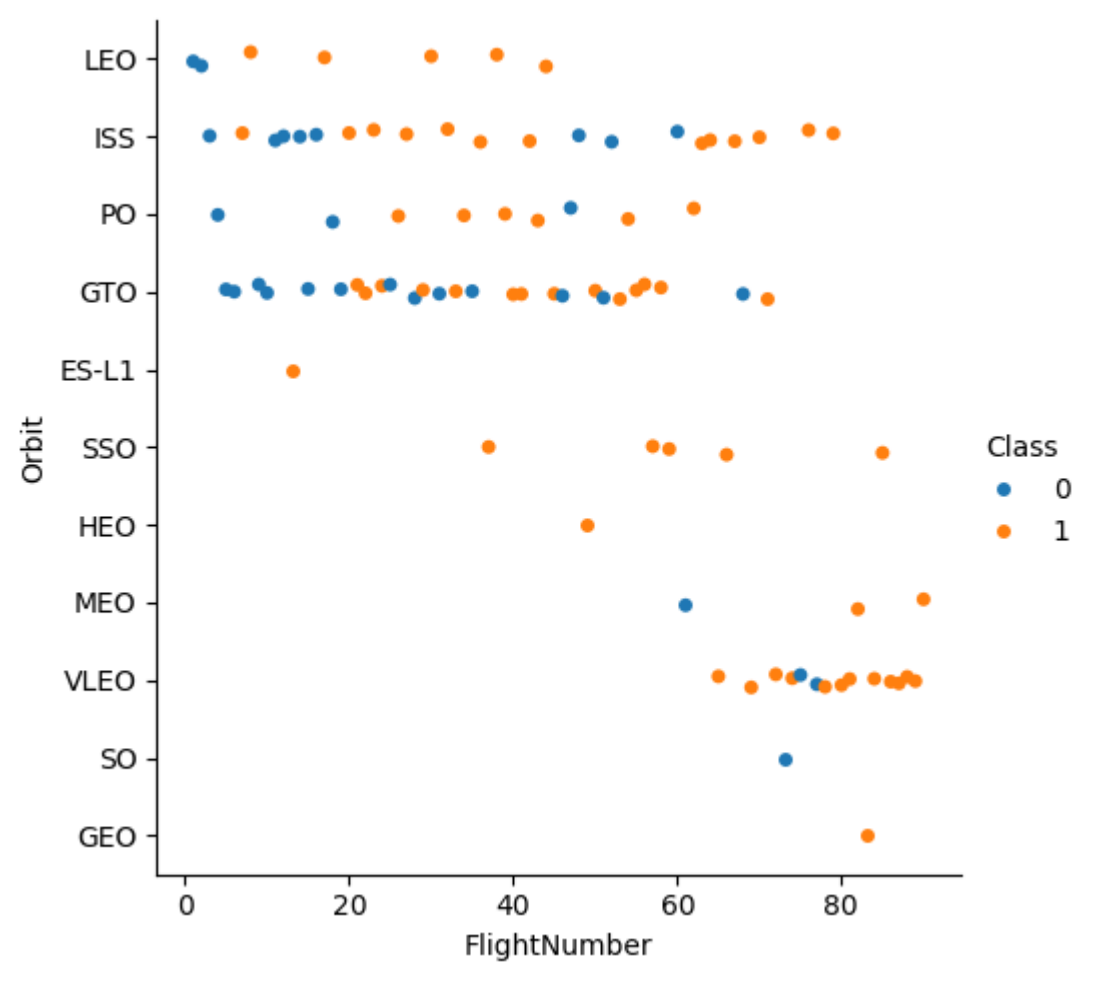
# Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
- Some launch orbit types (ES-L1,GEO,HEO,SSO) have higher landing success rates than others (SO,GTO,ISS,...)



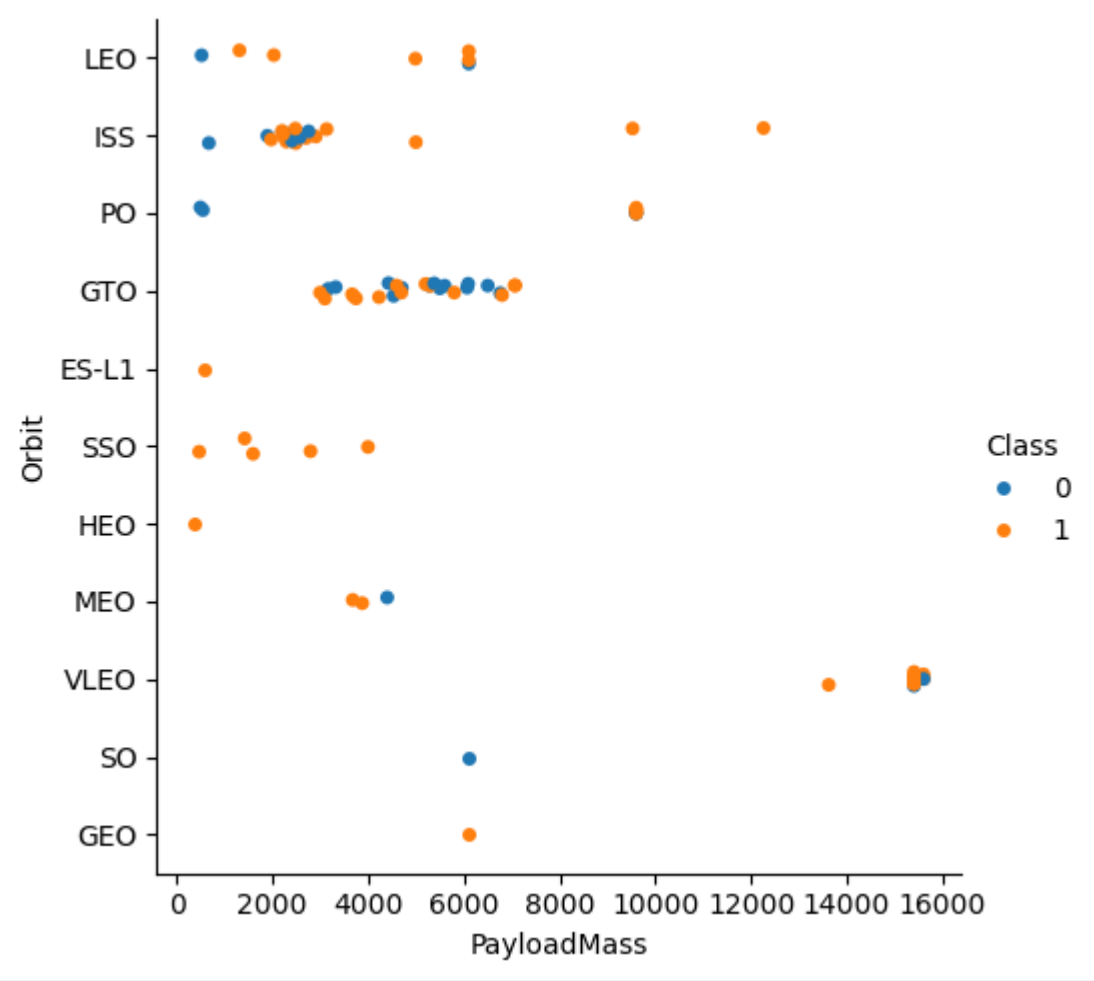
# Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type
- Some orbits (GTO,ISS) have significantly more launch attempts than others
- Some orbit types (SO,GEO) were not attempted until later flight numbers



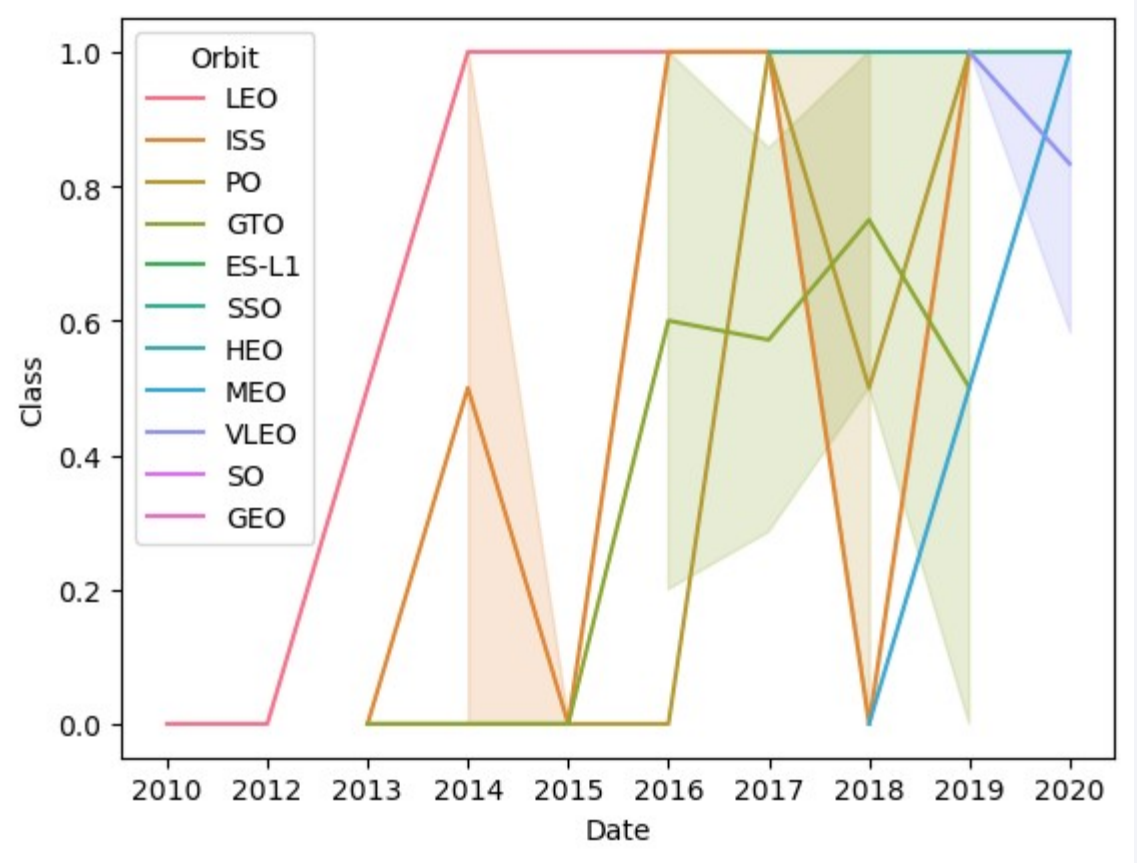
# Payload vs. Orbit Type

- Scatter point of payload vs. orbit type
- Orbit types tend to have distinct payload ranges
- ie. VLEO tends to have a higher payload range than GTO



# Launch Success Yearly Trend

- Line chart of yearly average success rate
- Launch success rate trends upward over time



# All Launch Site Names

Find the names of the unique launch sites

### Task 1

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

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

```
 * 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`

Task 2

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

```
%sql SELECT * FROM SPACEXTABLE 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 | Landin |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure |
| 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 |
| 2012-05-22 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | |
| 2012-10-08 | 0:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | |

# Total Payload Mass

Calculate the total payload carried by boosters from NASA

### Task 3

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

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

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

| SUM(PAYLOAD_MASS__KG_) |
|---|
| 45596 |

# Average Payload Mass by F9 v1.1

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

Task 4

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

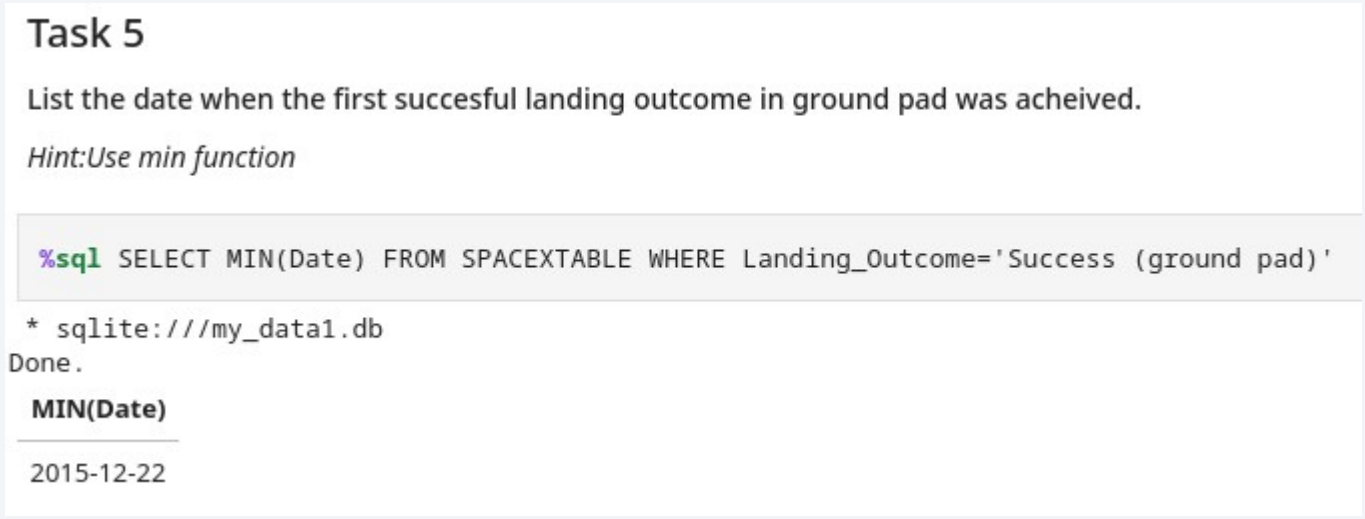
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version='F9 v1.
```

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

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

# First Successful Ground Landing Date

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



Task 5

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

*Hint:Use min function*

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome='Success (ground pad)'
```

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

| 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

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

```
%sql SELECT * FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

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

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2016-05-06 | 5:21:00 | F9 FT B1022 | CCAFS LC-40 | JCSAT-14 | 4696 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 2016-08-14 | 5:26:00 | F9 FT B1026 | CCAFS LC-40 | JCSAT-16 | 4600 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 2017-03-30 | 22:27:00 | F9 FT B1021.2 | KSC LC-39A | SES-10 | 5300 | GTO | SES | Success | Success (drone ship) |
| 2017-10-11 | 22:53:00 | F9 FT B1031.2 | KSC LC-39A | SES-11 / EchoStar 105 | 5200 | GTO | SES EchoStar | Success | Success (drone ship) |

# Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome,COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome
```

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

| Mission_Outcome | COUNT(*) |
|---|---|
| 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

Task 8

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

```
%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

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

| 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

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

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.**

```
%sql SELECT SUBSTR(Date,6,2) AS 'Month',Booster_Version,Launch_Site,Landing_Outcome FROM SPACEXTABLE WHERE Date LIKE '2015%' AND Landing_Outcome='Failure (drone ship)'
```

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

| Month | Booster_Version | Launch_Site | Landing_Outcome |
|---|---|---|---|
| 01 | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| 04 | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

# 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

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.

```
%sql SELECT Landing_Outcome,COUNT(*) FROM SPACEXTABLE WHERE Date > '2010-06-04' AND Date < '2017-03-20' GROUP BY Landing_Outcome ORDER BY Date DESC
```

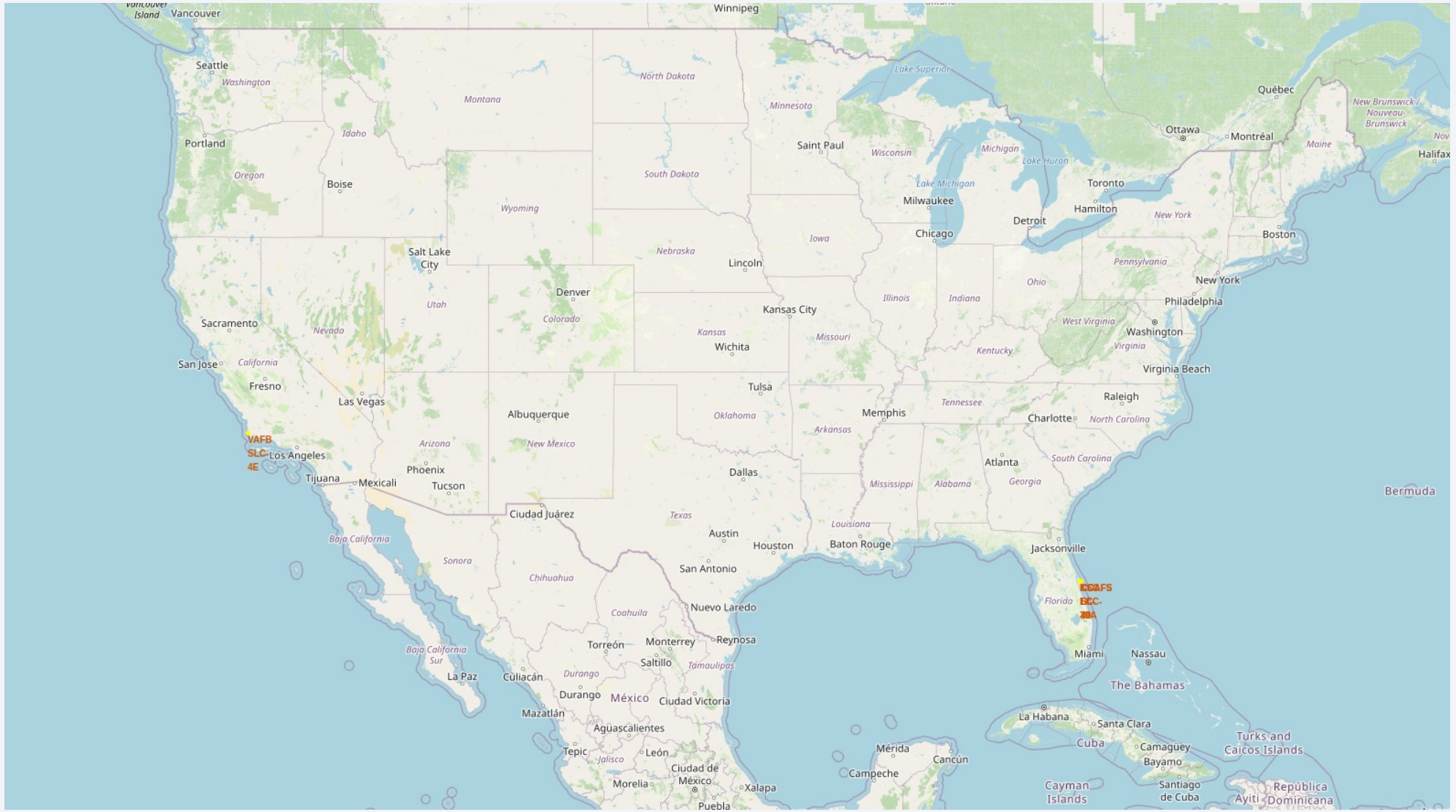
```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

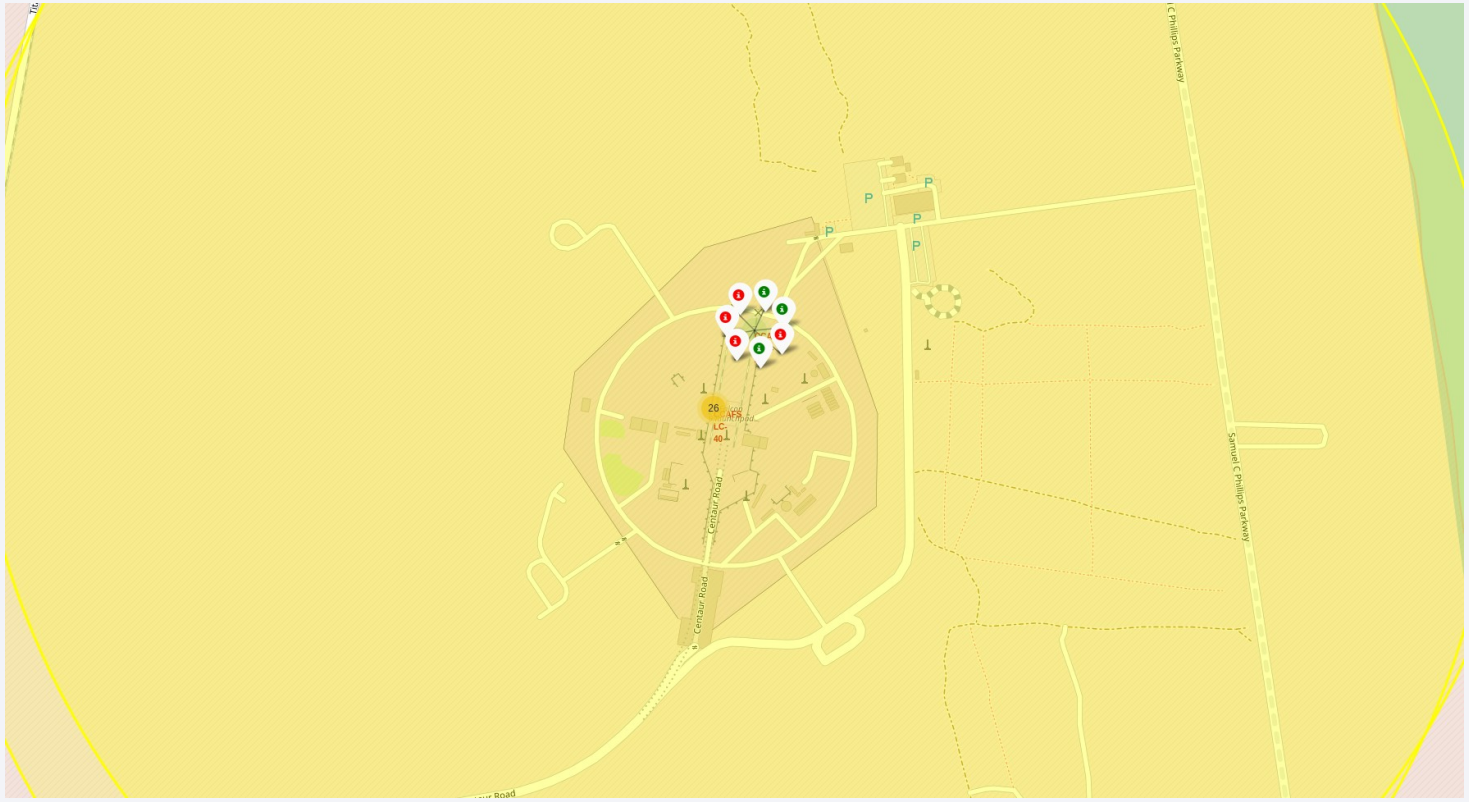
# Launch Sites Proximities Analysis

# Launch Site Folium Map

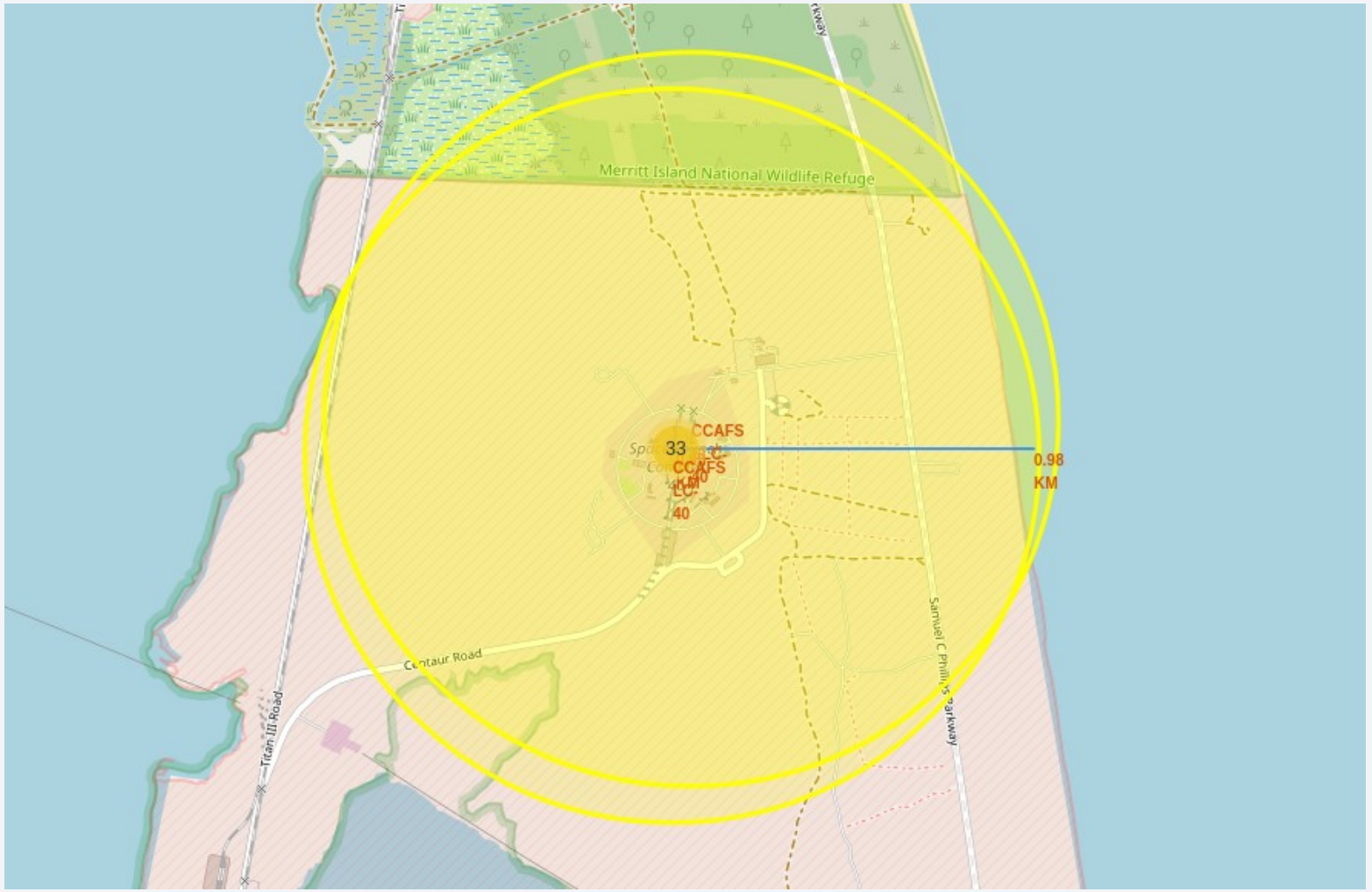


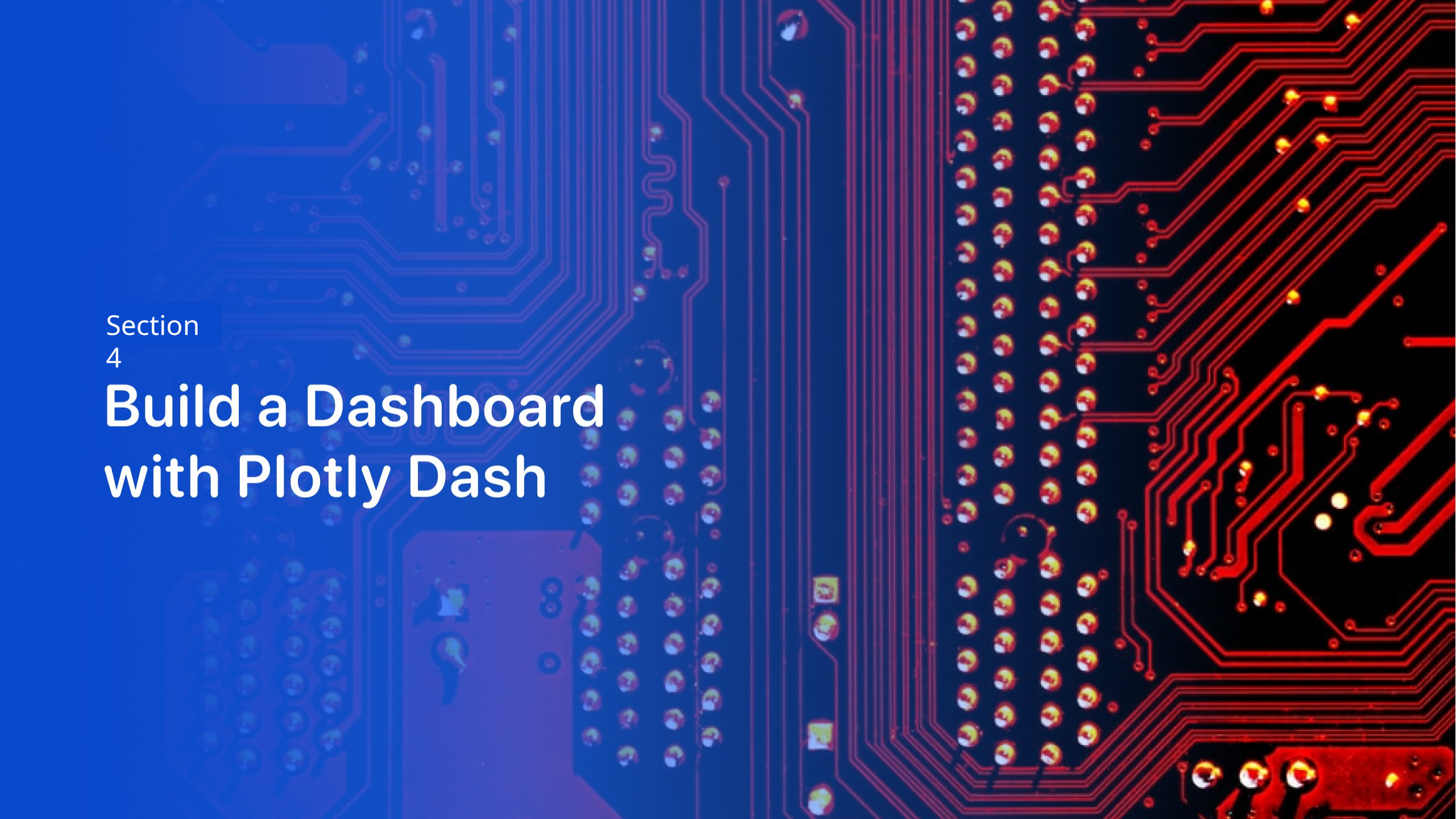
# Launch Outcomes Folium Map

Yellow Circle: Launch Site, Green Icon: Successful Landing, Red Icon: Unsuccessful Landing

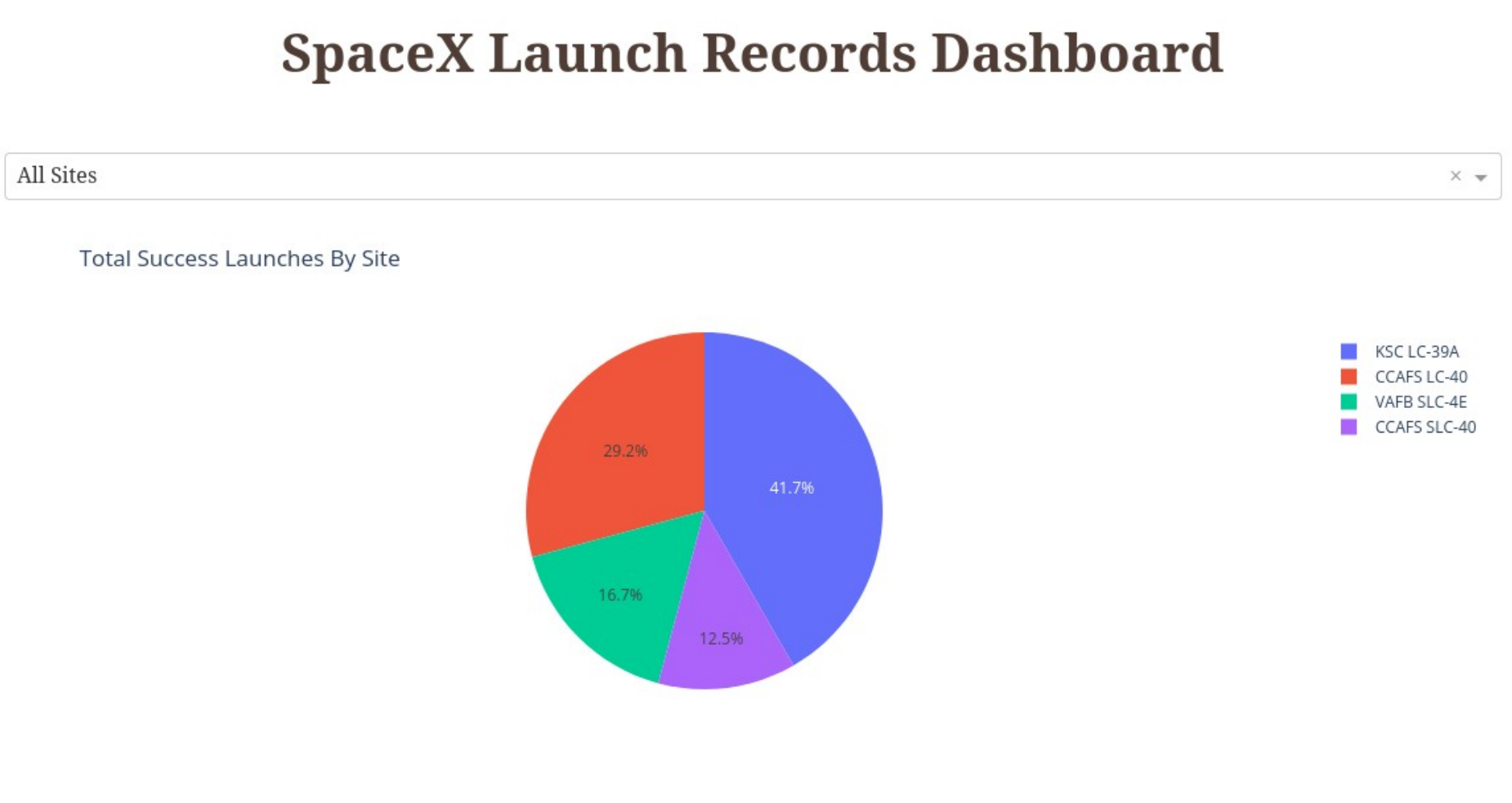


# Launch Proximities Folium Map

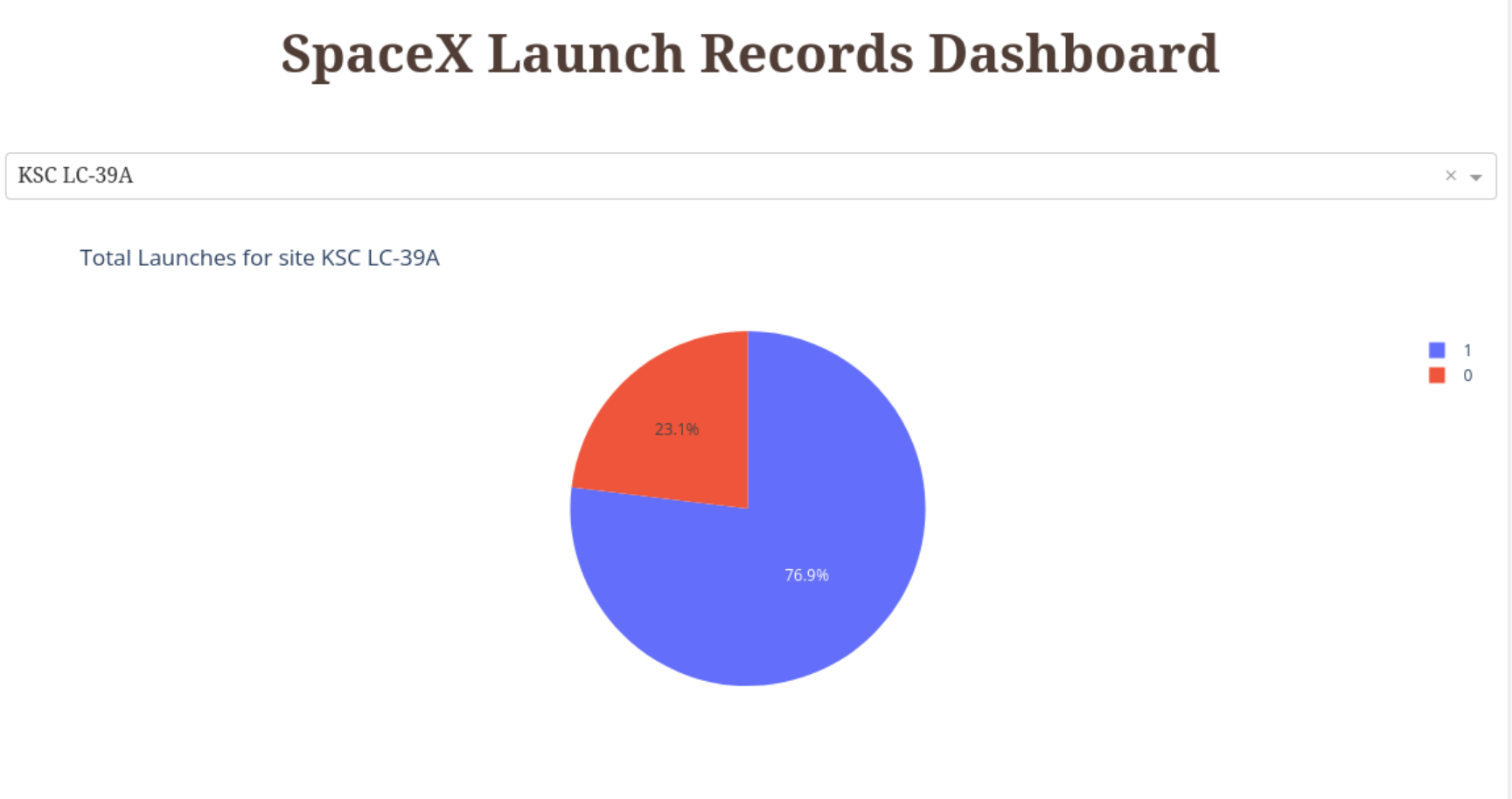



Section
4
Build a Dashboard
with Plotly Dash

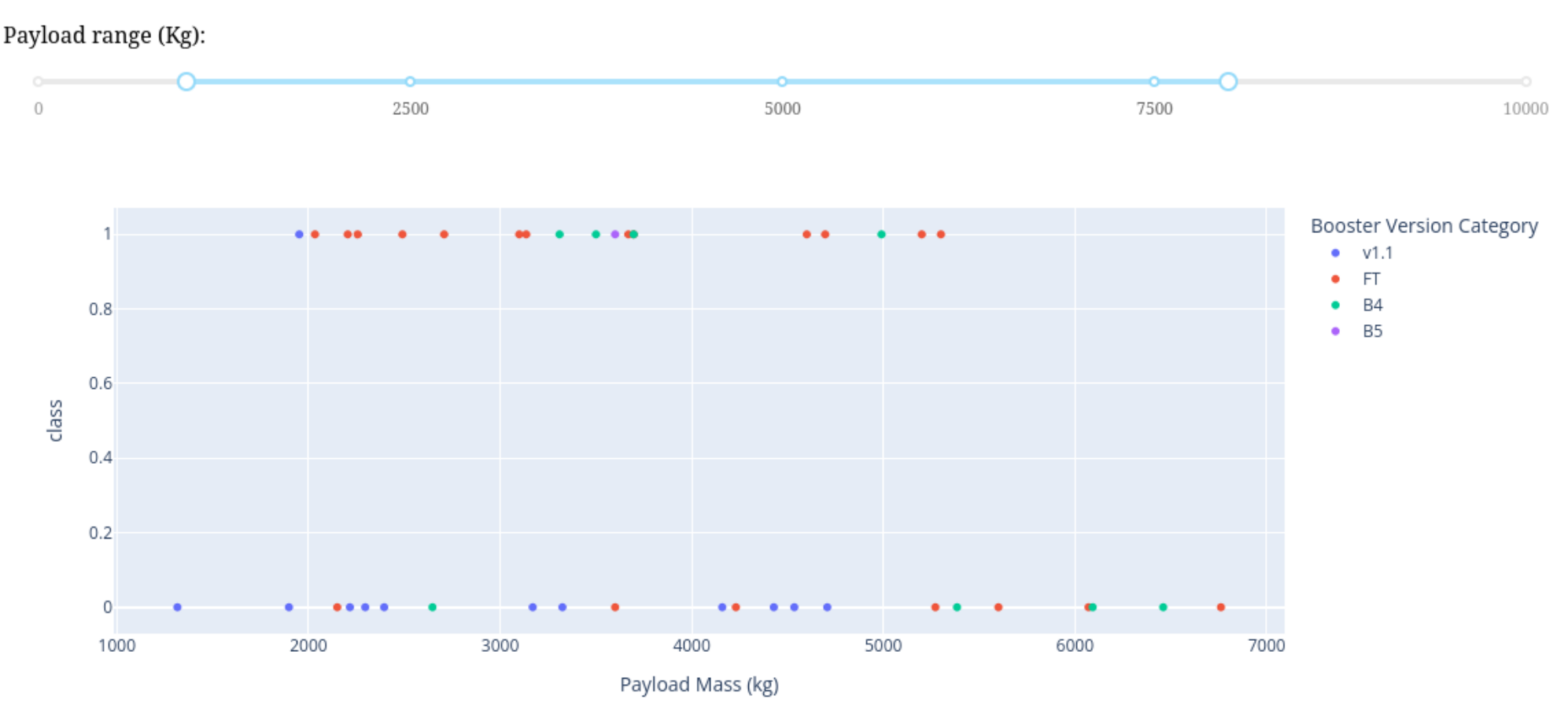
# Launch Success Count Piechart Dashboard



# Launch Site With Highest Success Rate



# Payload vs. Launch Outcome for all Sites

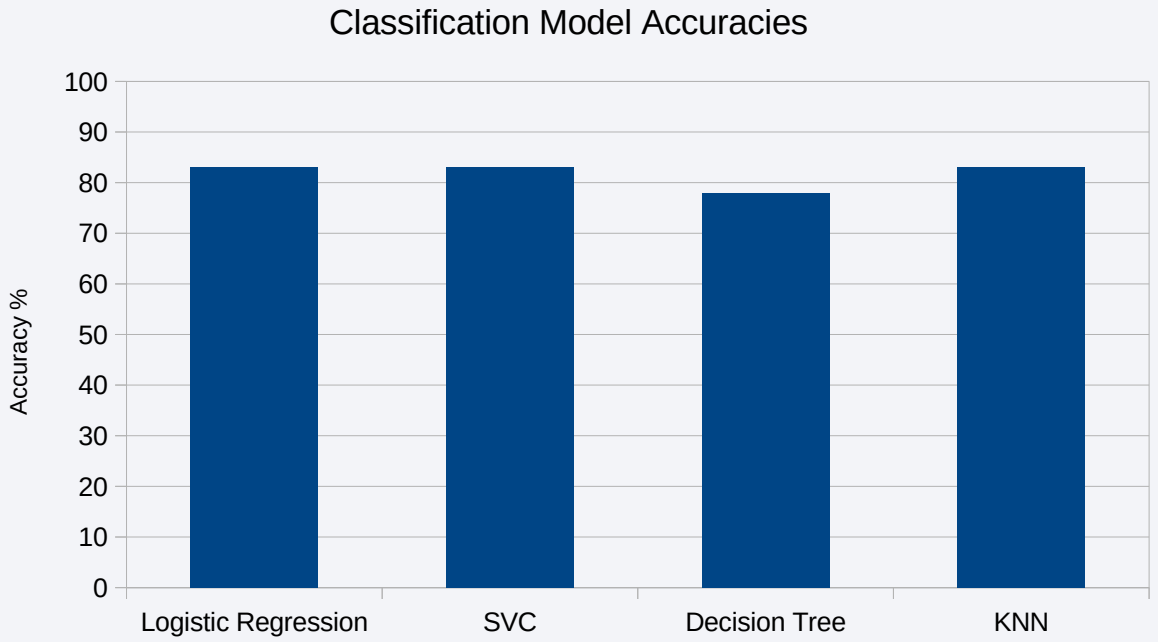


Section
5

# Predictive Analysis (Classification)

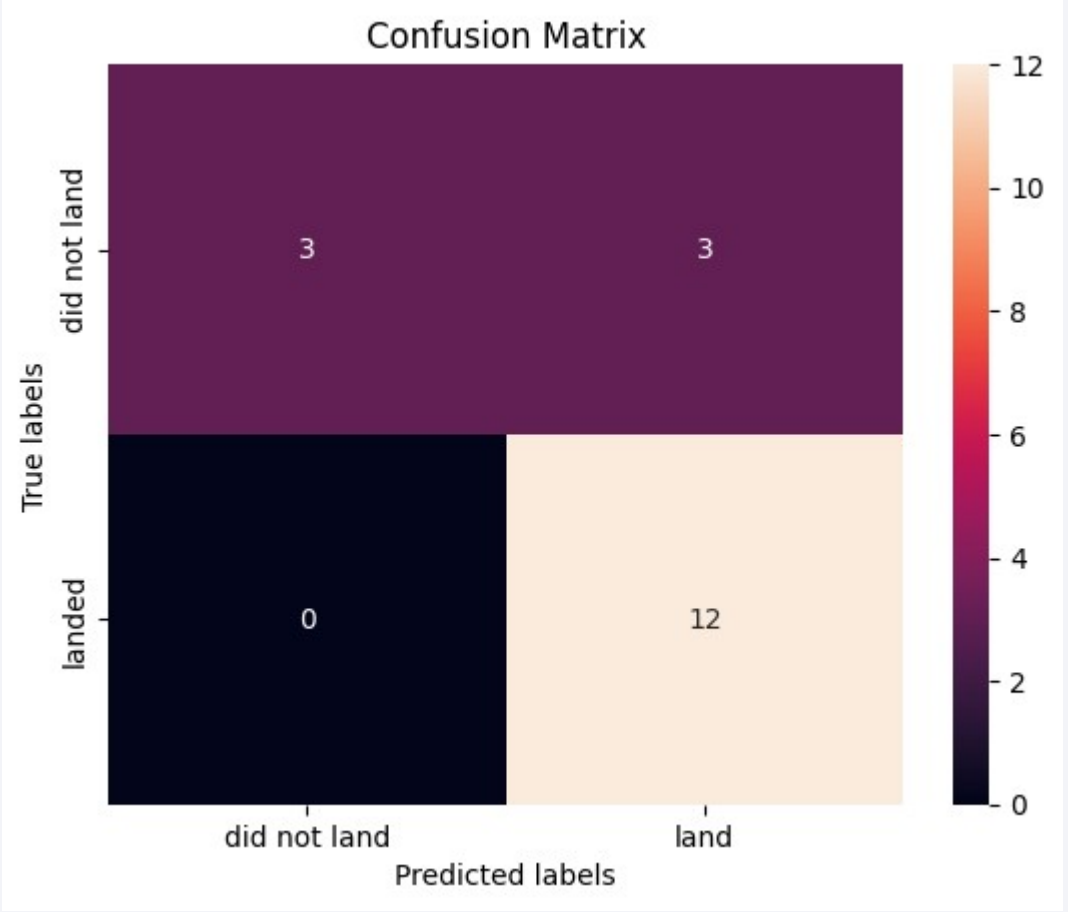
# Classification Accuracy

- Logistic Regression, SVC, and KNN all had equal test accuracy at 83%. Decision Tree was slightly lower at 78%



# Confusion Matrix

- Log Reg, SVC, and KNN had equivalent confusion matrices. SCV had a slightly longer runtime than Log Reg and KNN



# Conclusions

- Success rate of SpaceX booster landings generally increased over time
- Higher flight number launches had higher rates of success
- Certain launch orbit types like VLEO have more launches and higher rates of landing success than others such as SO
- Logistic Regression, SVC, and KNN all achieved equal classification accuracy predicting launch outcomes, around 83%

# Appendix

Capstone Project Github link:

https://github.com/jgoedmakers/IBMDataScienceProject/tree/main

# Thank you!