
# Applied Data Science Capstone

Gavin Liburd

12/30/2021

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

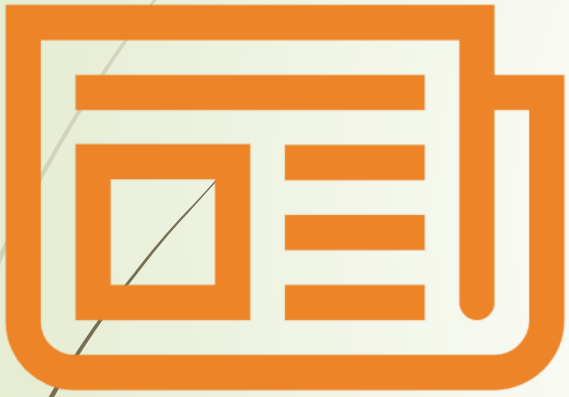
# EXECUTIVE SUMMARY

- SpaceY plans to join the space race and compete with SpaceX.
- SpaceY will compete with SpaceX's Falcon9 rockets.
- The Falcon9 rockets are special because SpaceX reuses the first stage.
- SpaceY will create a machine learning algorithm to determine when SpaceX will be able to reuse the first stage of their rocket.
- Data collection, data wrangling, data analysis, and interactive dashboards were utilized prior to the machine learning algorithm.

# INTRODUCTION

- SpaceX's historical Falcon9 rocket data will be collected and analyzed.
- SpaceX's rockets see success because their Falcon9 rockets have a cheaper price point compared to their competitors.
- The cheaper price point is due to the fact that Falcon9 can reuse the first stage of their rockets.
- Following collection of SpaceX's data, questions to answer are:
  - Which elements of a Falcon9 rocket have the biggest effect on if the first stage is successfully retrieved?
  - When will SpaceX be able to successfully retrieve the first stage of their rockets?

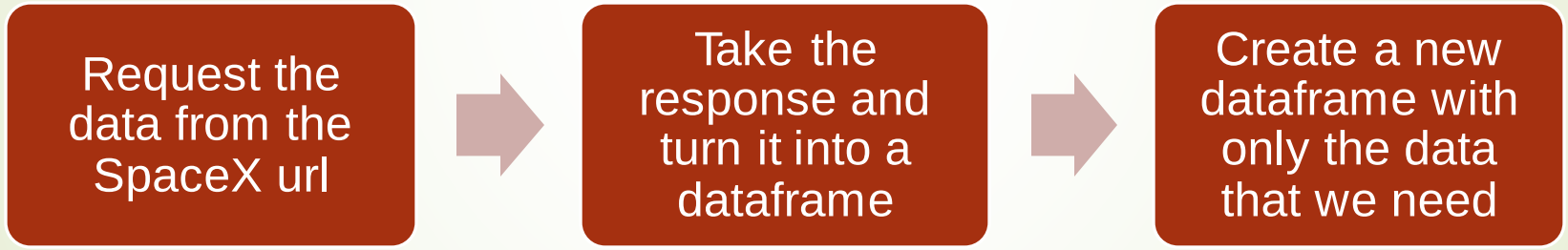
# METHODOLOGY



- Data Collection via Web Scraping
- Data Wrangling
- Data Analysis via SQL
- Data Analysis via Pandas and MatPlotLib
  - Line Plots, Scatter Plots, and Bar Plots
  - One-hot Encoding
- Interactive Analytics Dashboard
- Predictive Analysis
  - Logistic Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors

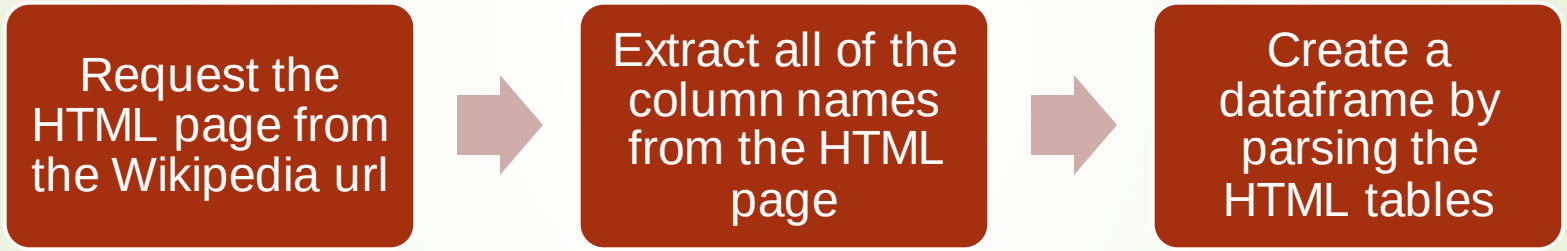
# Data Collection-SpaceX API

To begin the project, we need to collect and clean our data




https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Data%20Collection.ipynb

# Data Collection-Web Scraping




https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

# Data Wrangling Methodology


https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Data%20Wrangling.ipynb

# EDA with Data Visualization

Scatter Plots

- Payload Mass vs. Flight Number
- Launch Site vs. Flight Number
- Launch Site vs. Payload Mass
- Orbit vs. Flight Number
- Orbit vs. Payload Mass

These plots were all used to attempt to discern the relationships of the independent variables and if they have any bearing on whether the first stage successfully reaquired.

Bar Graphs

- Success Rate vs. Orbit

This graph was used to determine which orbit types have the highest success rate.

https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Exploratory%20Data%20Analysis%20using%20Pandas%20and%20Matplot.ipynb

# EDA with SQL

https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Exploratory%20Data%20Analysis%20Using%20SQL.ipynb

List of SQL Queries

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

# Interactive Visual Analytics

Folium

- We visualized the launch data in an interactive map.
- Added circles and popups on the map in order to clearly display launch sites.
- Added markers to indicate the outcome on the map.
- Created a line to show the distance between a launch site and the selected coastal point.

https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Interactive%20Visual%20Analytics%20Dashboard%20with%20Folium.ipynb

Plotly Dashboard

- Created a pie chart to show the success rates for all of the respective launch sites to see which sites were the most successful.
- Created a scatter plot that showed the relationship between payload mass and landing outcome.

https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/spacex_dash_app.py

# Predictive Analysis

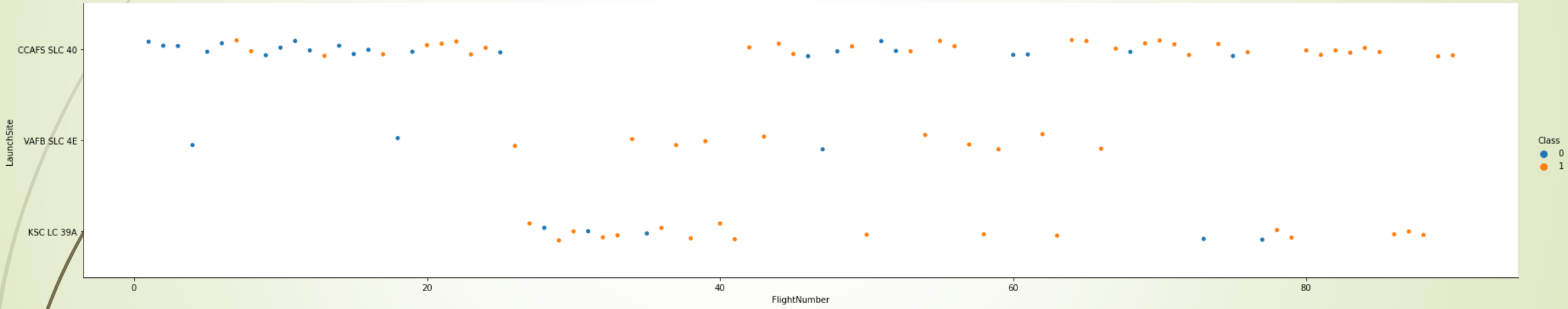
Standardized the x variables

Split the x and y variables into training and test sets

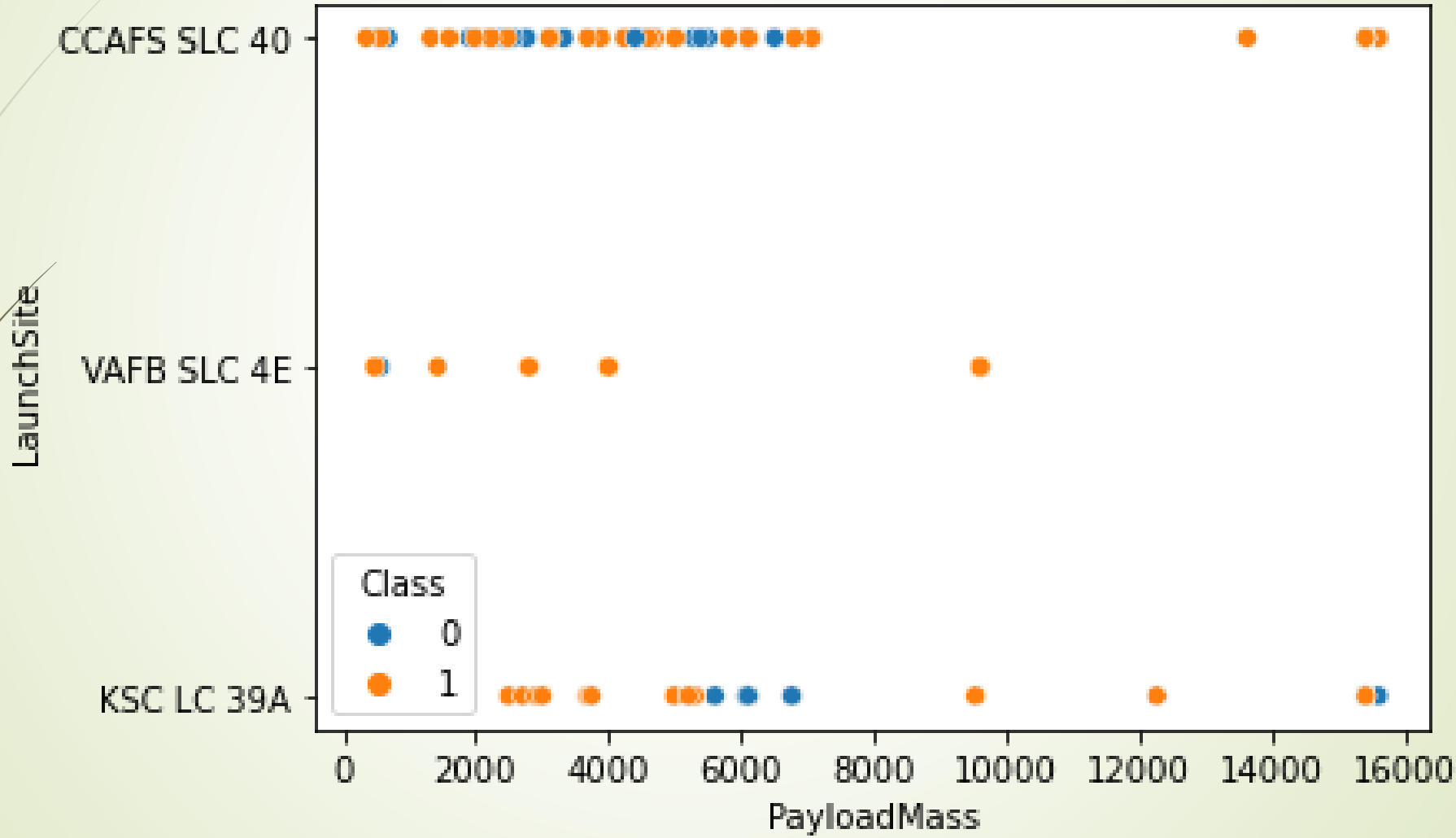
Used logisitic regression, support vector machine, decision tree classifier and k nearest neighbors methods

https://github.com/Lasspeng/SpaceX_Capstone_Project/blob/main/Predictive%20Analysis.ipynb

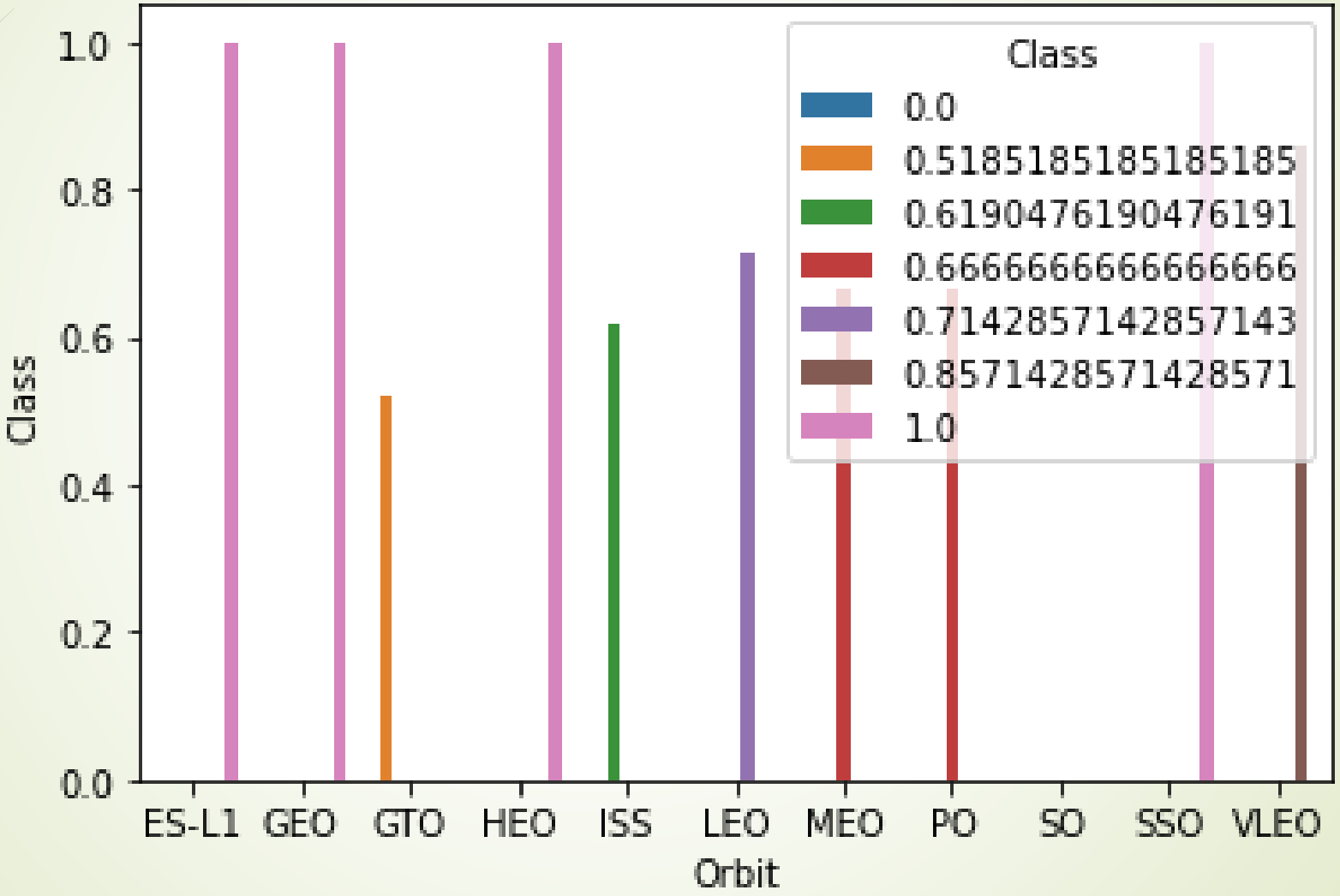
# EDA with Visualization Results



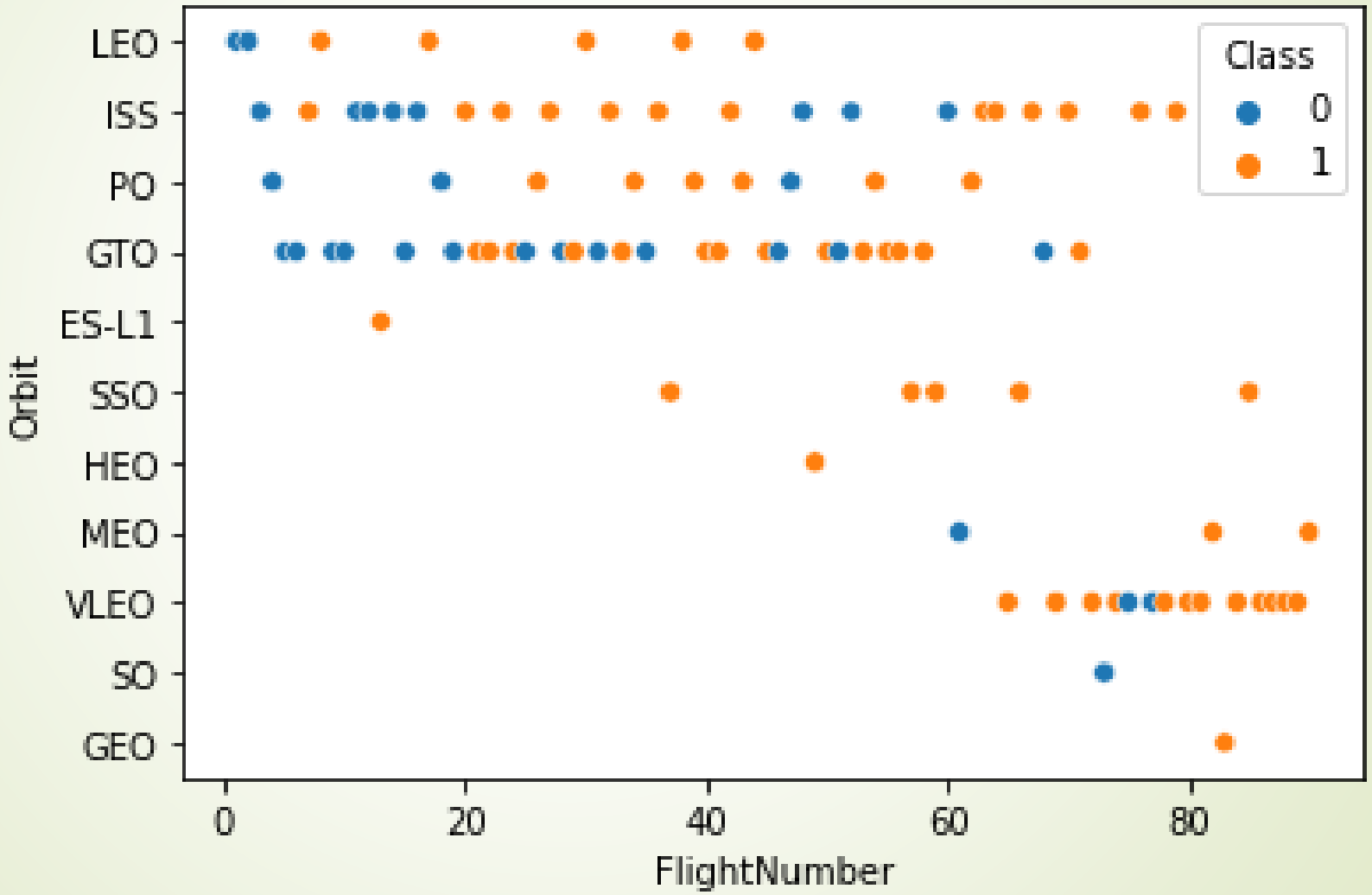
# EDA with Visualization Results



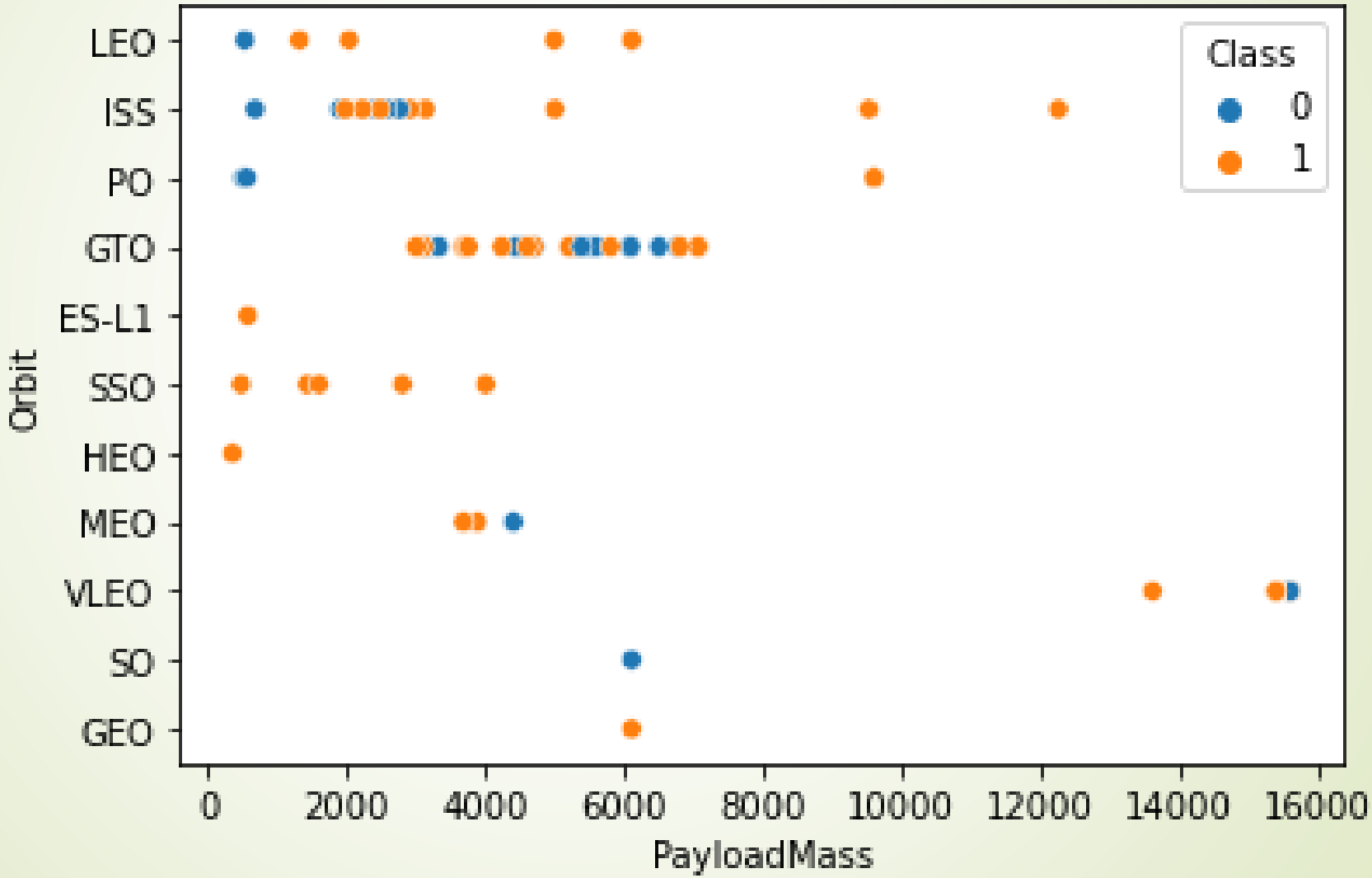
# EDA with Visualization Results



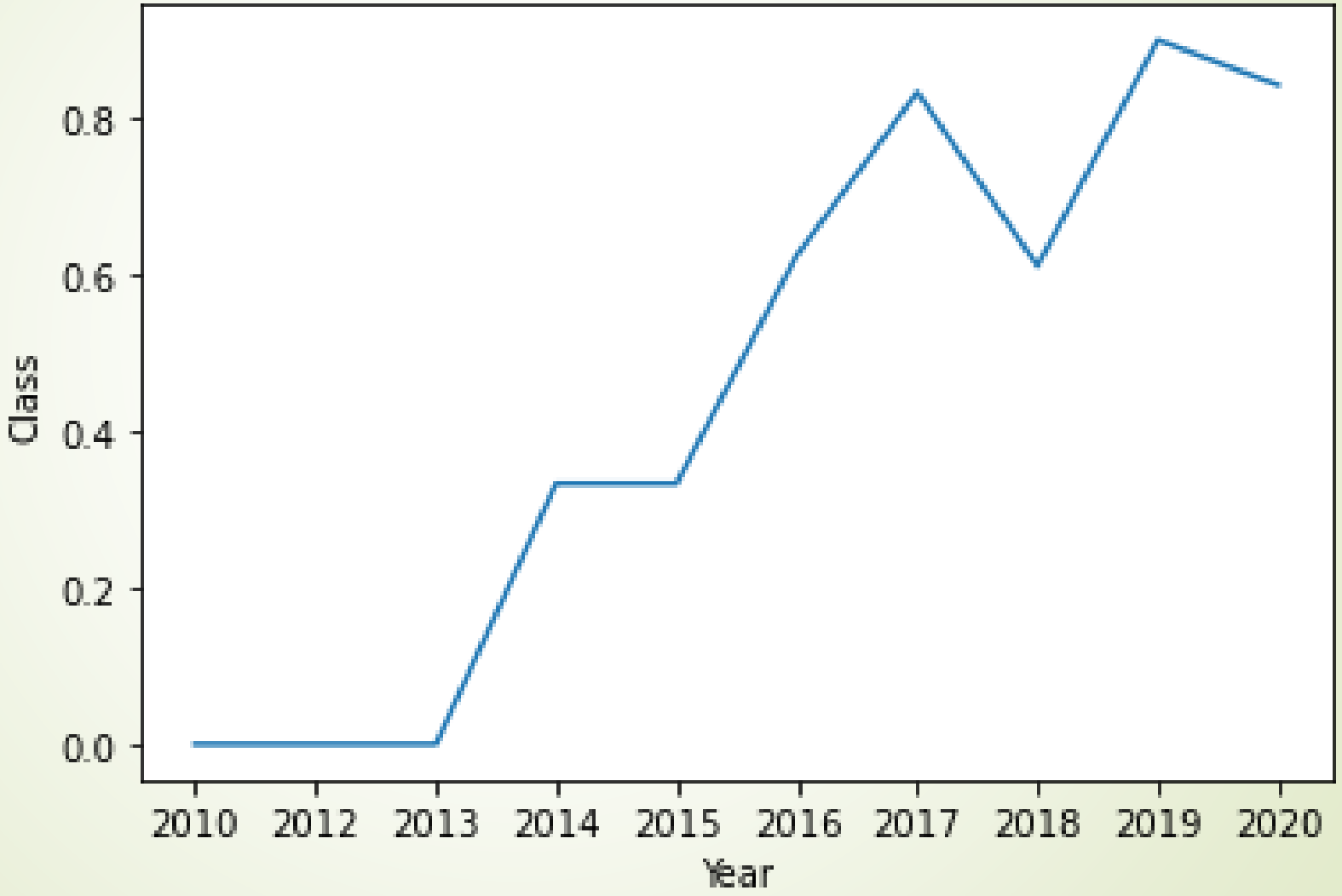
# EDA with Visualization Results



# EDA with Visualization Results



# EDA with Visualization Results



# EDA with SQL

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

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

# EDA with SQL

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

| 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 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 00:35:00 | 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 |

# EDA with SQL

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

| 1 |
| --- |
| 45596 |

# EDA with SQL

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

# EDA with SQL

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

| 1 |
| --- |
| 2015-12-22 |

# EDA with SQL

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

| booster_version |
| --- |
| F9 B4 B1040.2 |
| F9 B4 B1040.1 |
| F9 B4 B1043.1 |
| F9 B5 B1046.2 |
| F9 B5 B1046.3 |
| F9 B5 B1047.2 |
| F9 B5 B1048.3 |
| F9 B5 B1051.2 |
| F9 B5 B1058.2 |
| F9 B5B1054 |
| F9 B5B1060.1 |
| F9 B5B1062.1 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |
| F9 FT B1032.2 |
| F9 FT B1020 |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1030 |
| F9 FT B1032.1 |
| F9 v1.1 |
| F9 v1.1 B1011 |
| F9 v1.1 B1014 |
| F9 v1.1 B1016 |

# EDA with SQL

List the total number of successful and failure mission outcomes

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

# EDA with SQL

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

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

# EDA with SQL

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

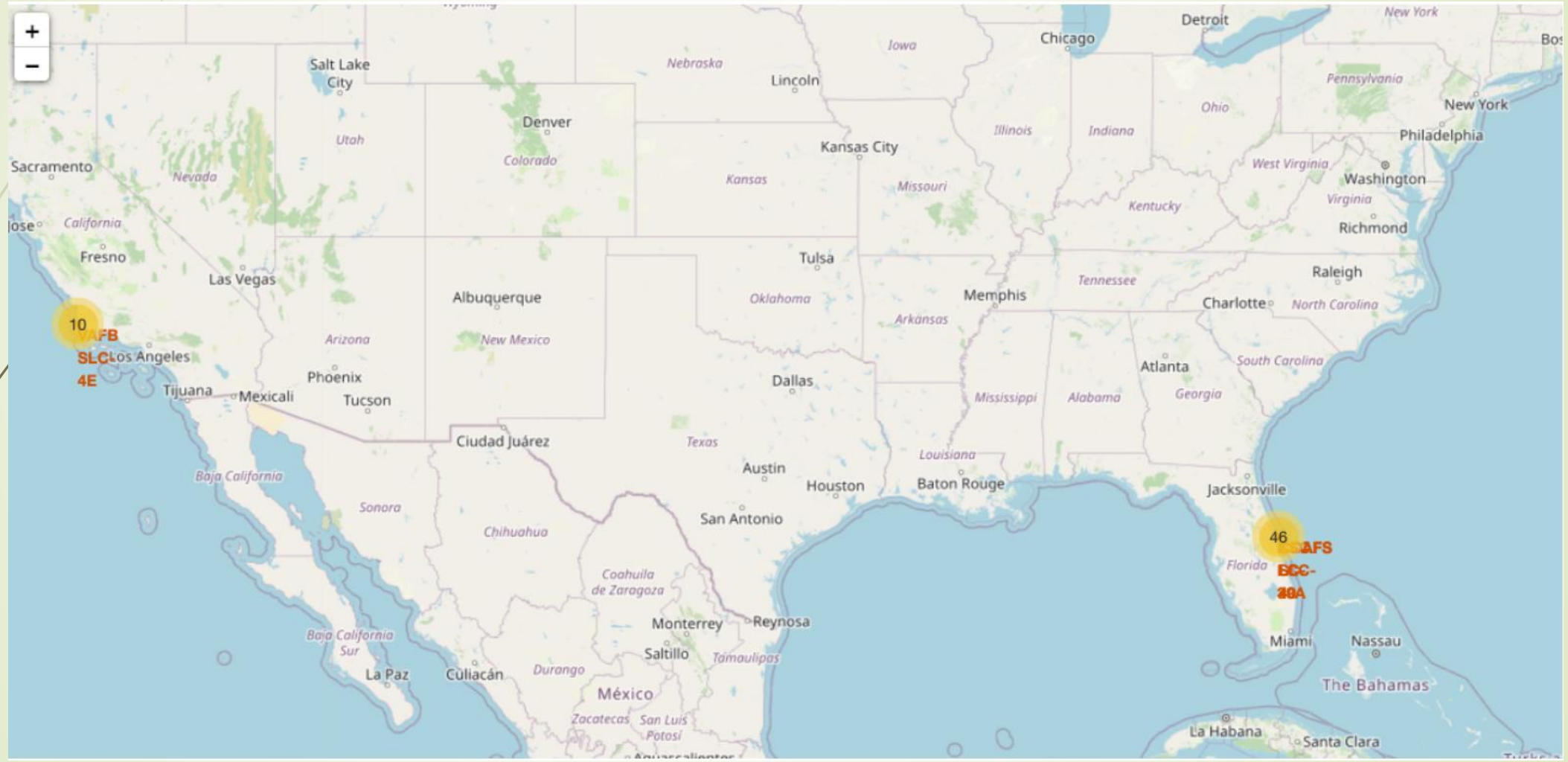
| landing__outcome | booster_version | launch_site |
|---|---|---|
| Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

| landing__outcome |
| --- |
| Uncontrolled (ocean) |
| Uncontrolled (ocean) |
| Success (ground pad) |
| Success (ground pad) |
| Success (ground pad) |
| Success (drone ship) |
| Success (drone ship) |
| Success (drone ship) |
| Success (drone ship) |
| Success (drone ship) |
| Precluded (drone ship) |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| No attempt |
| Failure (parachute) |
| Failure (parachute) |
| Failure (drone ship) |
| Failure (drone ship) |
| Failure (drone ship) |
| Failure (drone ship) |

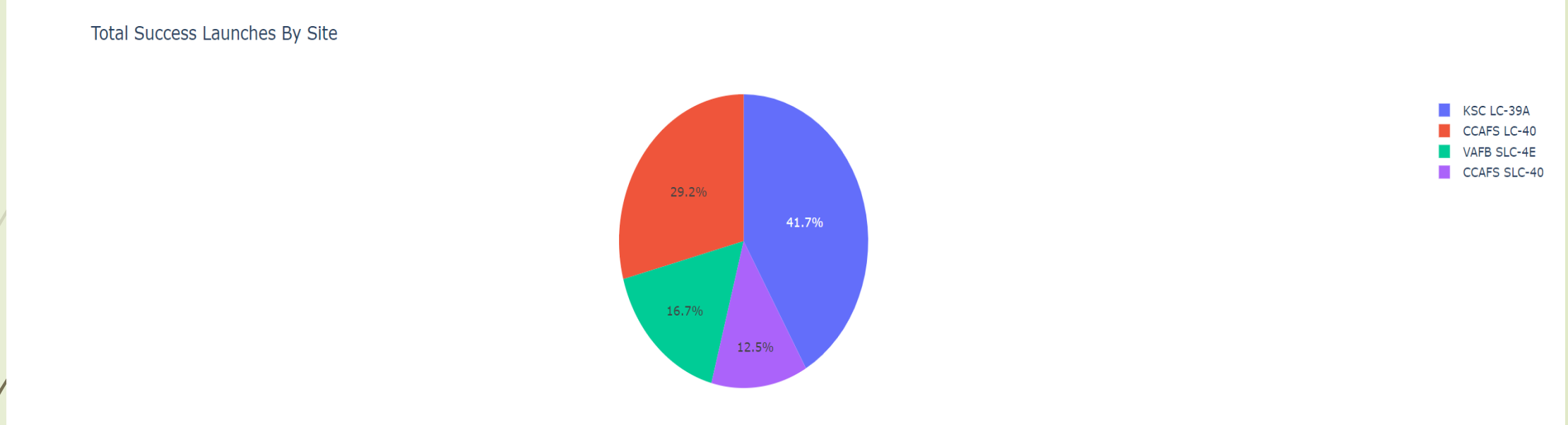
# EDA with SQL

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

# Folium Map

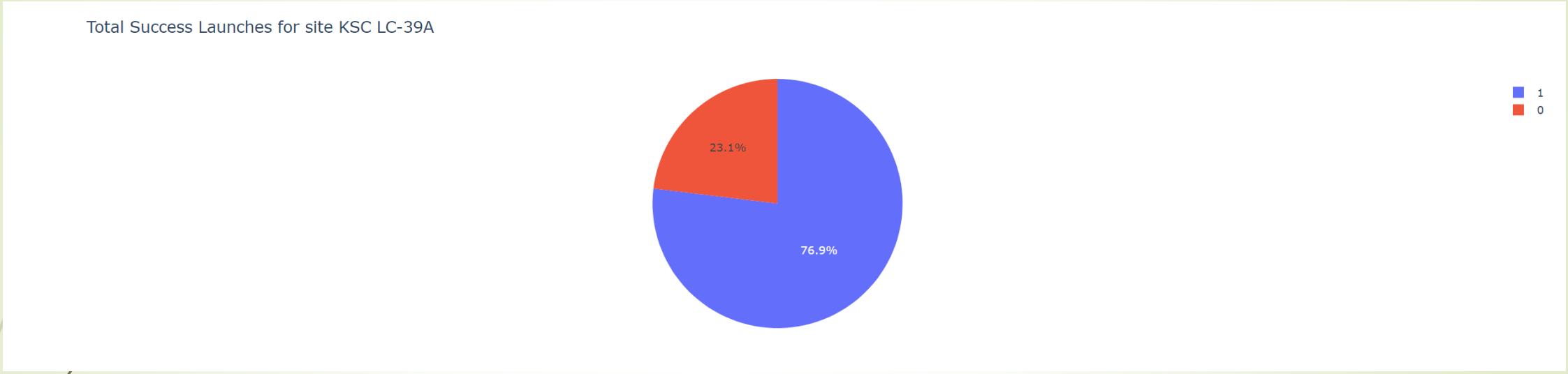

+
−
Salt Lake City
Nebraska
Iowa
Chicago
Detroit
New York
Lincoln
Pennsylvania
Denver
Ohio
Utah
Illinois
Indiana
New York
Philadelphia
Sacramento
Colorado
Kansas City
West Virginia
Washington
Nevada
Kansas
Missouri
Virginia
Kentucky
California
Richmond
Fresno
Tulsa
Las Vegas
Tennessee
Raleigh
Albuquerque
Oklahoma
Memphis
Charlotte
North Carolina
10
Arkansas
Arizona
New Mexico
Los Angeles
Atlanta
South Carolina
Phoenix
Dallas
Tijuana
Mexicali
Tucson
Mississippi
Alabama
Georgia
Ciudad Juárez
Texas
Louisiana
Austin
Houston
Baton Rouge
Jacksonville
Baja California
Sonora
San Antonio
46
Chihuahua
Florida
Coahuila de Zaragoza
Monterrey
Reynosa
Miami
Nassau
Baja California Sur
Saltillo
Tamaulipas
La Paz
Culiacán
Durango
The Bahamas
México
Zacatecas
San Luis Potosí
La Habana
Santa Clara

# Plotly Dash




KSC LC-39A has the most successful launches

# Plotly Dash

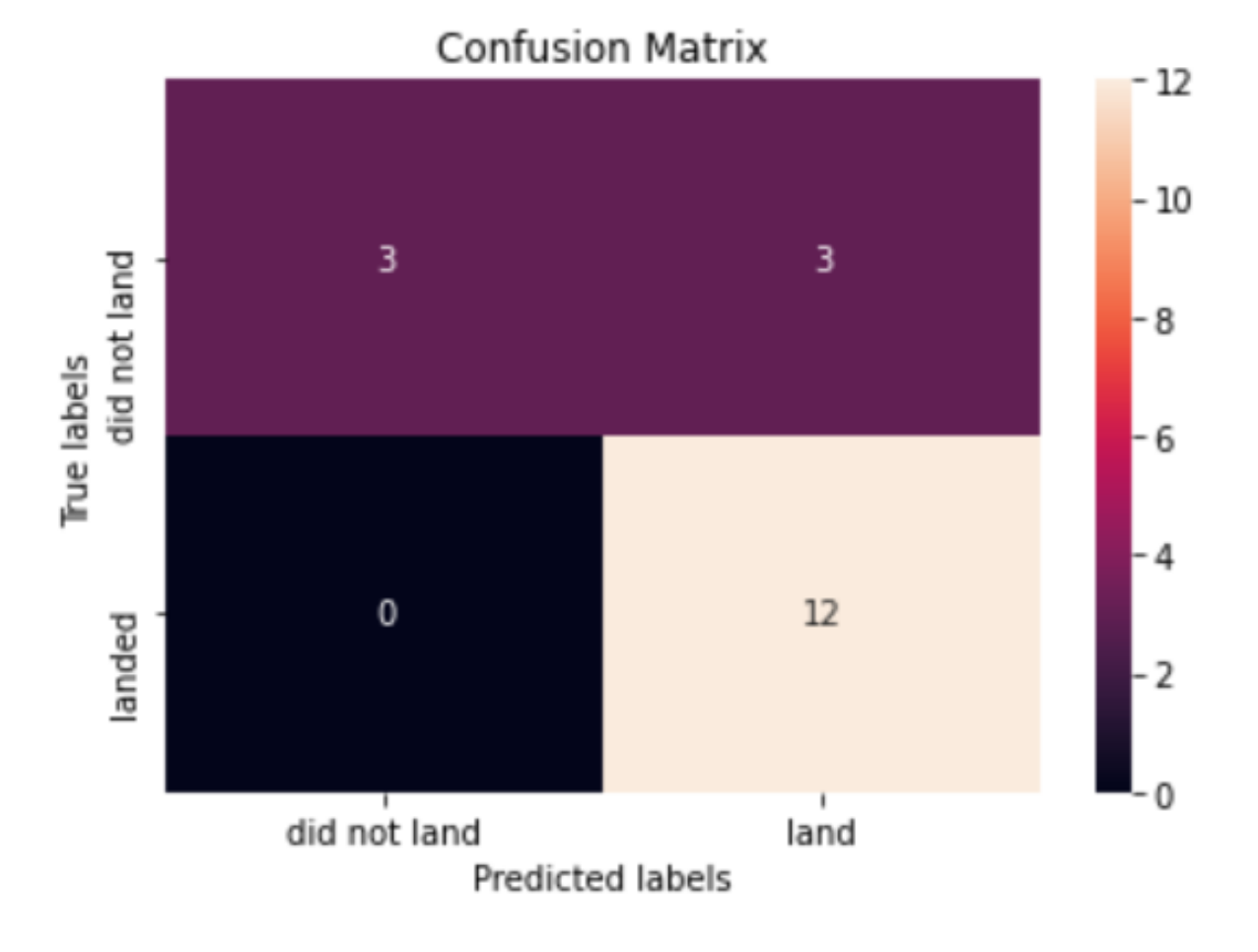



KSC LC-39A has the highest proportion of successful launches

# Predictive Analysis (Decision Tree)

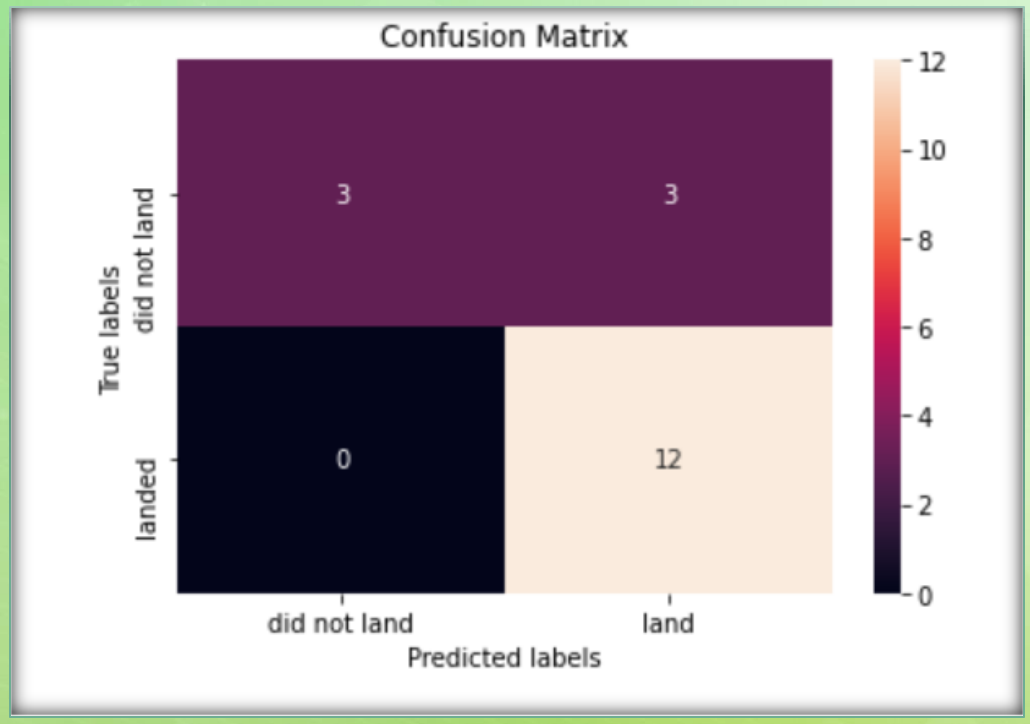
```
In [33]: tree_cv.score(X_test, Y_test)

Out[33]: 0.8333333333333334
```

# Predictive Analysis (Decision Tree)



# Predictive Analysis

- All of the methods gave identical accuracy scores of 83.33%. The methods also all gave identical confusion matrices.



# CONCLUSION

- Any of the predictive analysis methods can be used for this dataset.
- The success rates for retrieving the Falcon9's first stage has increased with every passing year.
- Heavier-weighted payloads are more likely to have a first stage that is successfully retrieved and reused
- Launch site KSC LC-39A has historically had the most success with Falcon9 launches