IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Miguel Mares
January 27th, 2025

# Outline

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

# Executive Summary

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 Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Florida is a safer place to launch and land rockets.

Heavier payloads are more likely to succesful outcome.

# Introduction

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 Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

# Data Collection – SpaceX API and Wikipedia

- Request to the SpaceX API

https://api.spacexdata.com/v4/*keywords*

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/01_jupyter-labs-spacex-data-collection-api-v2.ipynb

- Request the Falcon9 Launch Wiki

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/02_jupyter-labs-webscraping.ipynb

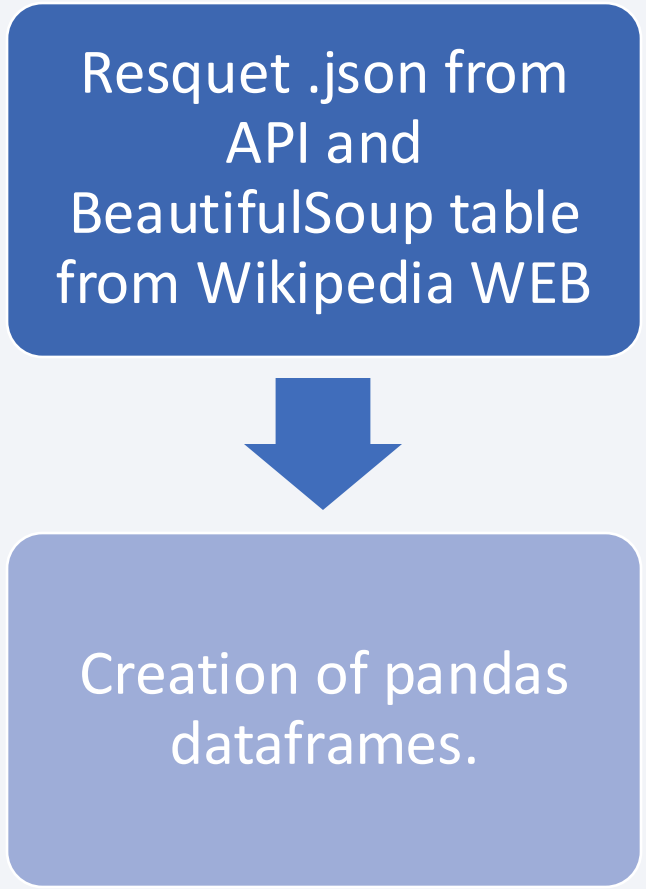
# Data Collection - Scraping

- Appending columns to dataframe from SpaceX API specific information.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/01_jupyter-labs-spacex-data-collection-api-v2.ipynb

- BeautifulSoup HTML extraction from Wikipedia Falcon 9 Launch to dataframe.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/02_jupyter-labs-webscraping.ipynb



# Data Wrangling

- Identification data types.
- Identification of missing values.
- Identification unexpected annotations.
- Fix of missing values with mean.
- Fix annotations and types.
- Numbering and booling necessary object types for numerical analysis.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/03_labs-jupyter-spacex-Data%20wrangling-v2.ipynb

# EDA with Data Visualization

- Scatter plots from several relations.
- Histograms from ‘success rate’.
- Line plot from ‘success rate’.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/04_jupyter-labs-eda-dataviz-v2.ipynb

# EDA with SQL

Magic SQL for:

- Connecting and setting database.
- Queries and subqueries to get insights of data.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/05_jupyter-labs-eda-sql-coursera_sqllite.ipynb

# Build an Interactive Map with Folium

- Marks to launching sites to observe geographical conditions.
- Marks on successful and failed launches related each site.
- Features in proximities of launching sites (how far are structures to help operations).

Run in your JupyterLab to see folium map.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/06_lab-jupyter-launch-site-location-v2.ipynb

# Build a Dashboard with Plotly Dash

- Pie chart of success launches count by site.

- Scatter chart based in success count on payload mass.

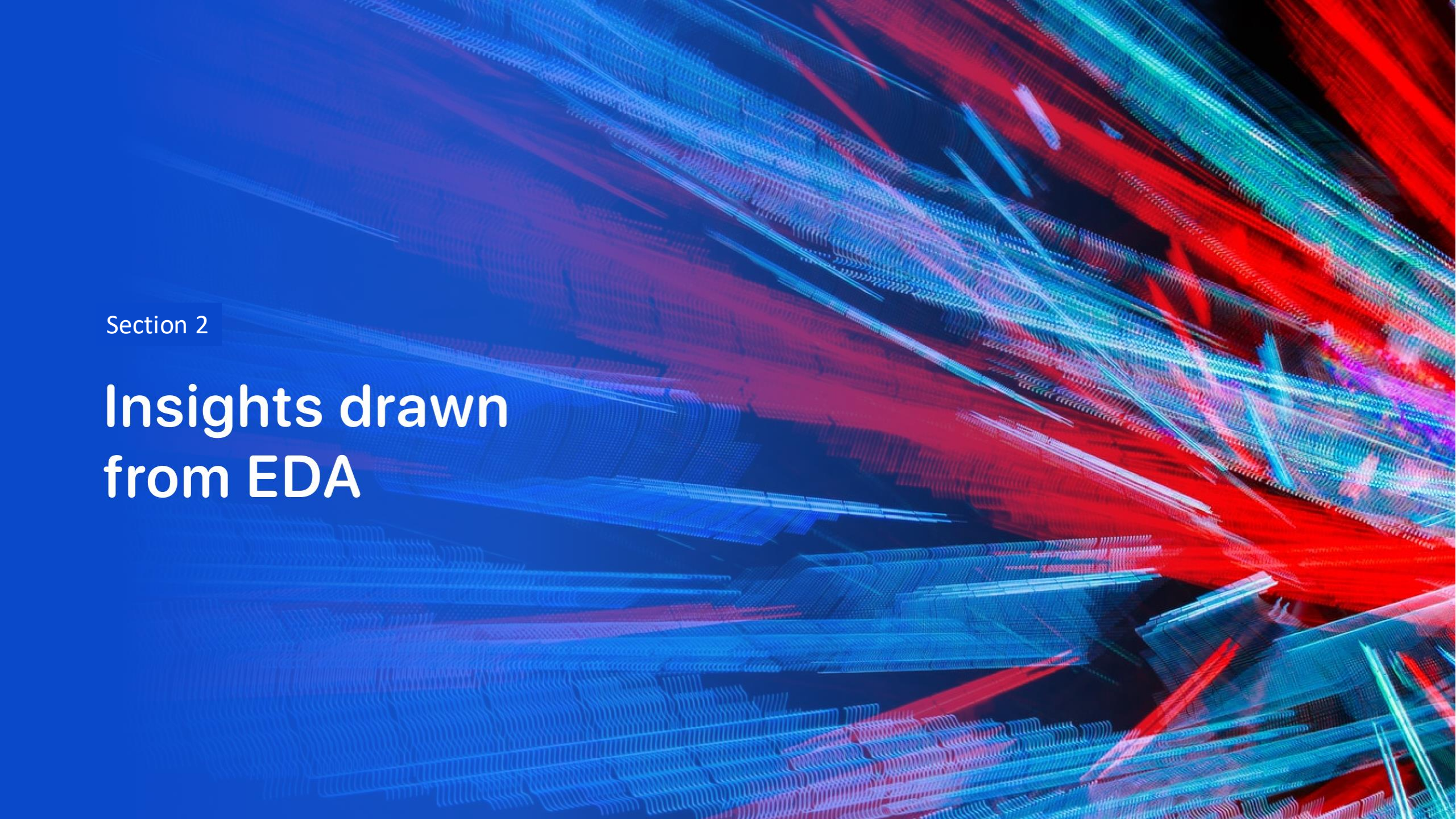
Run in your JupyterLab to see the dynamic chart.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/07_Hands-on%20Lab_%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb

# Predictive Analysis (Classification)

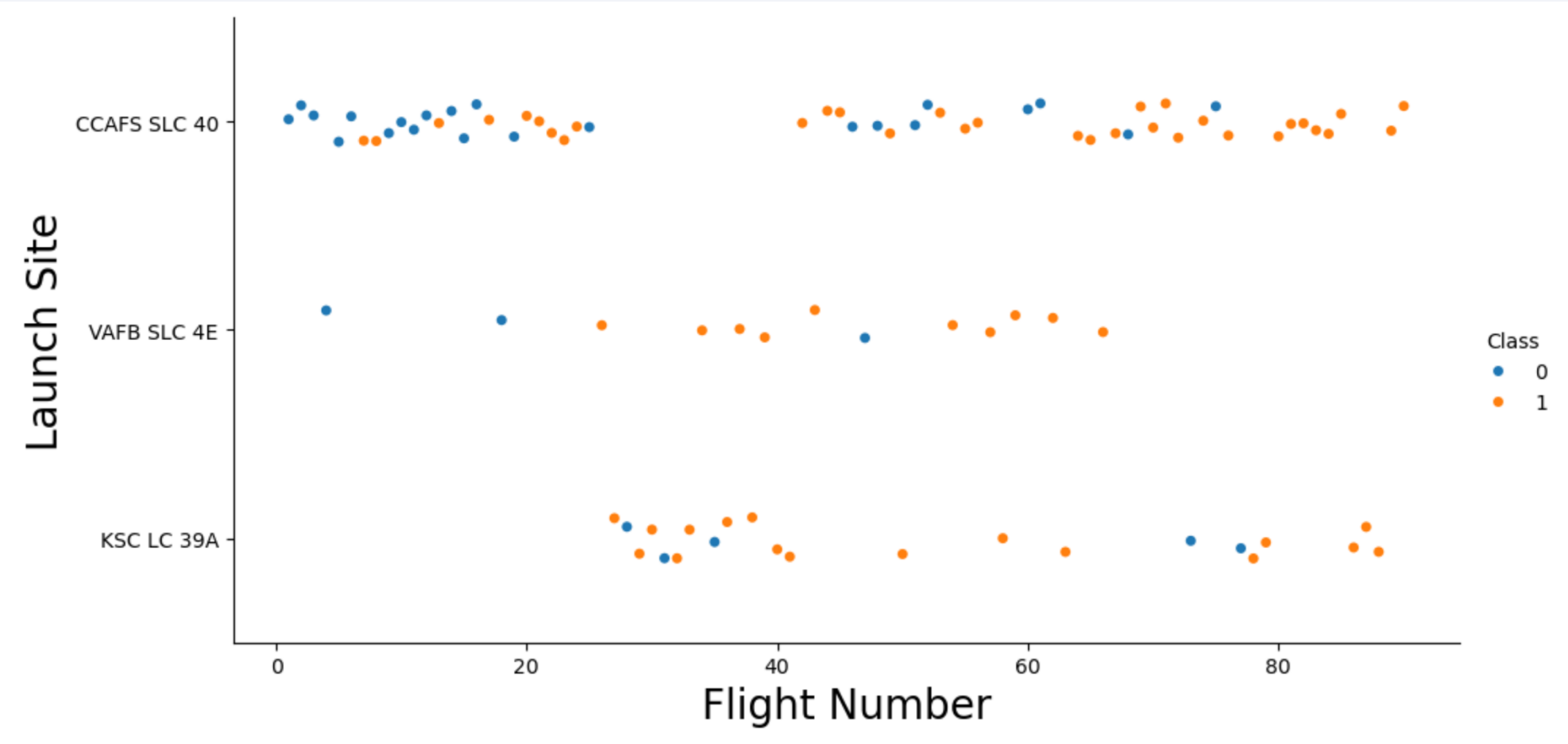
- Split data train.
- Use of GridSearchCV for getting best parameters for each Classification algorithm.
- Get best score (accuracy).
- Plot confusion matrix for each classifications to be aware of false positives.
- Report of all classifiers.

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM/blob/main/08_SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb


Section 2
Insights drawn from EDA

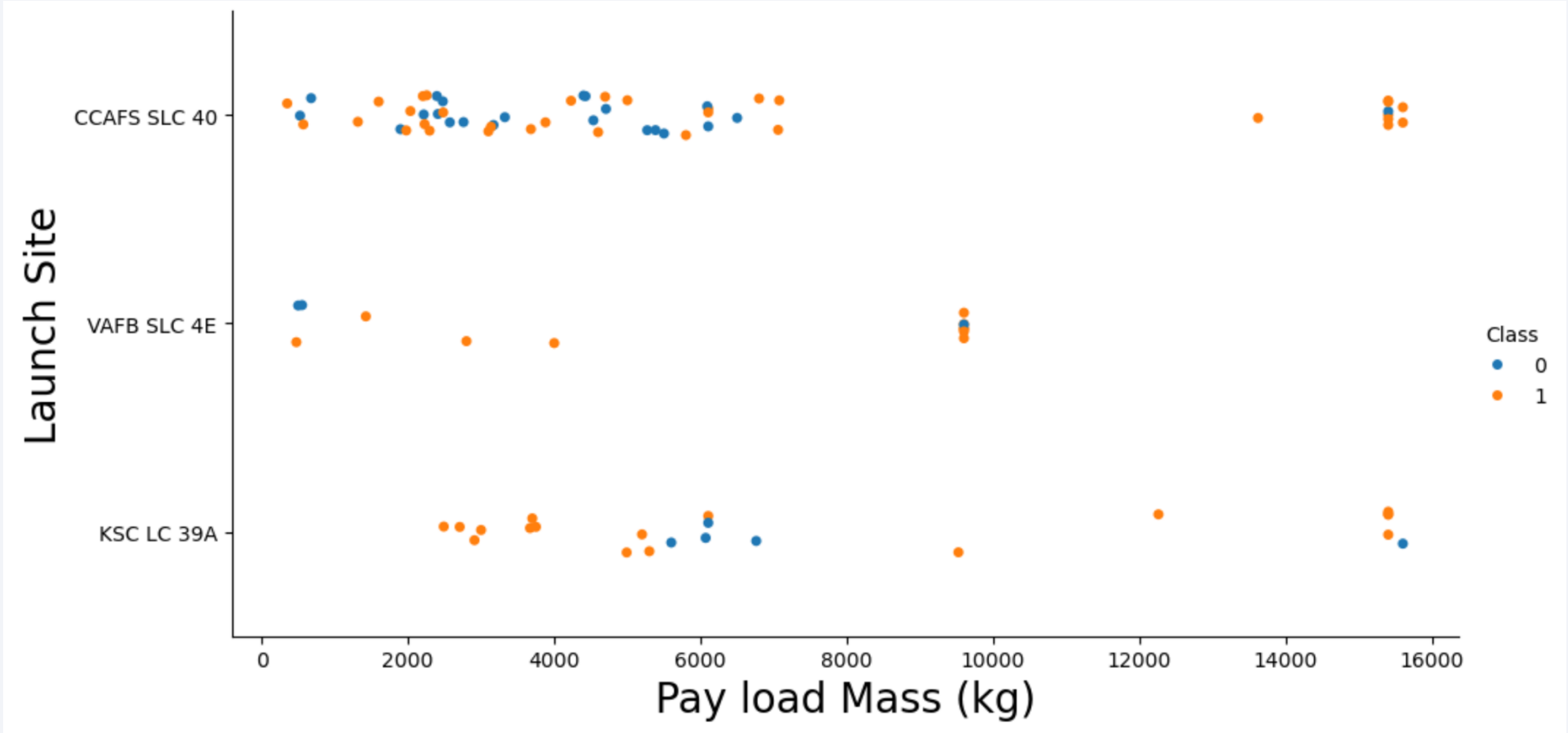
# Flight Number vs. Launch Site

- Most of launches are made in Florida, but California has better success rate.

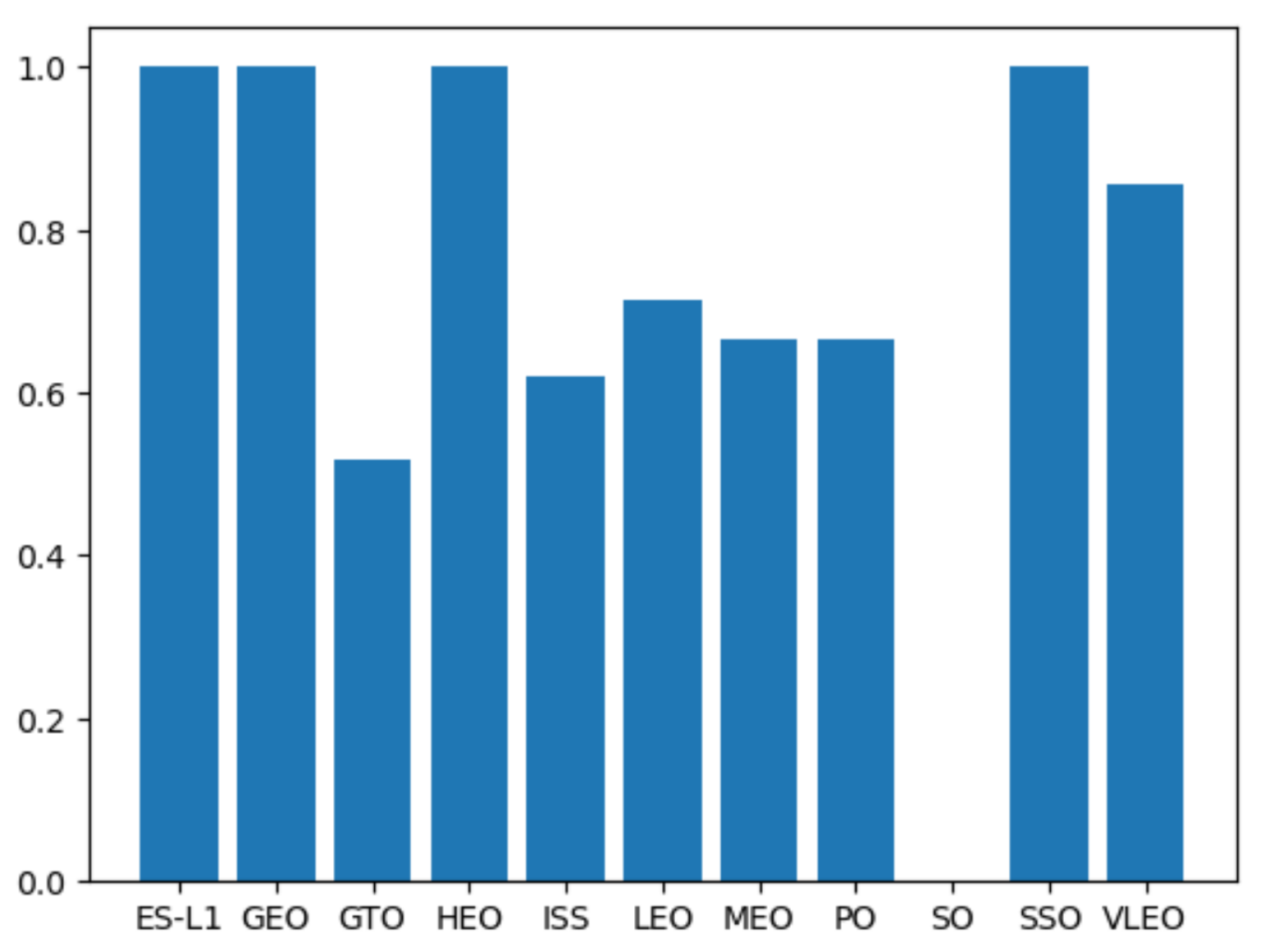


# Payload vs. Launch Site

- Heavier payload are more like to successful outcomes.



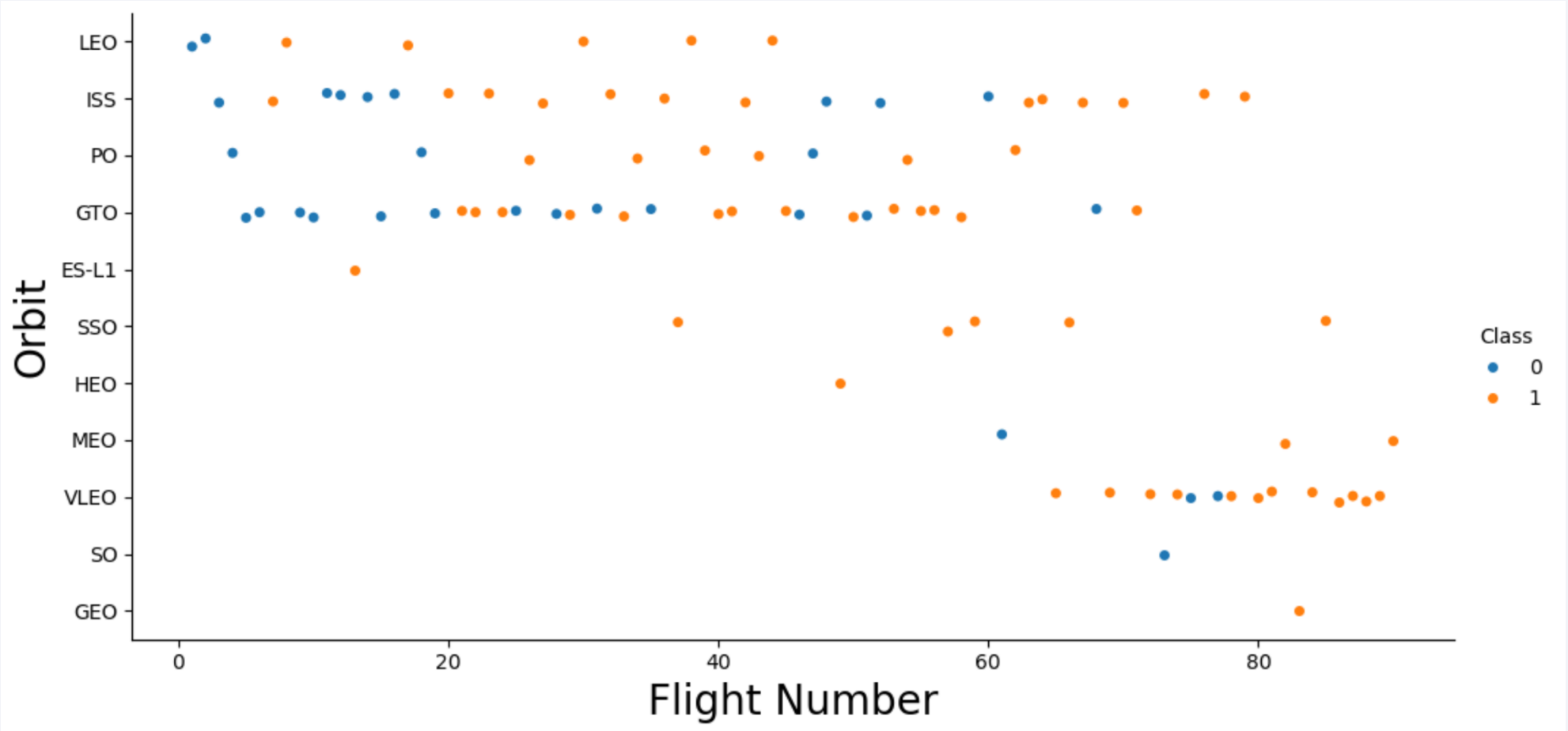
# Success Rate vs. Orbit Type




- Higher orbits have more success rate.

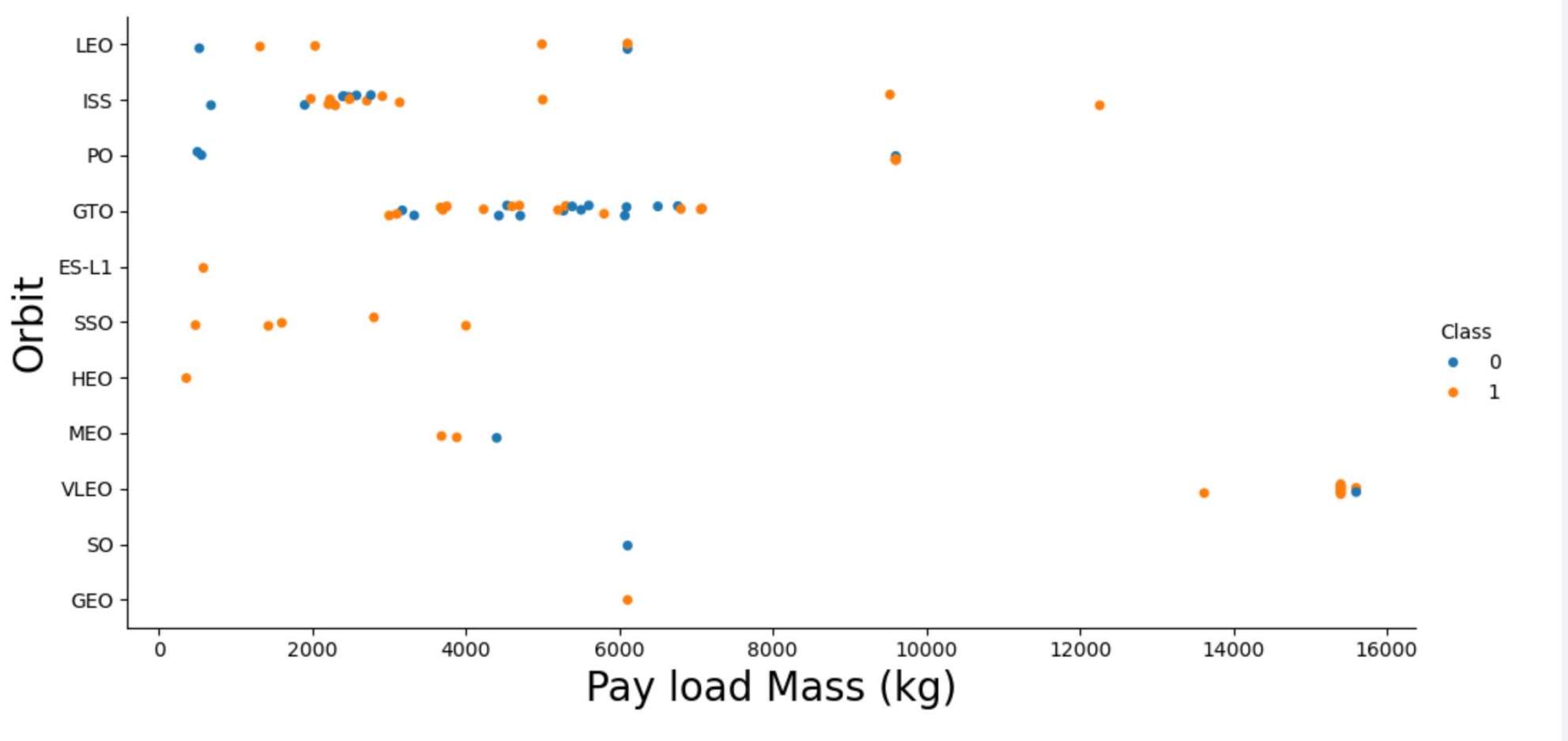
# Flight Number vs. Orbit Type

- Confirms at first glance what pointed in previous bar chart.

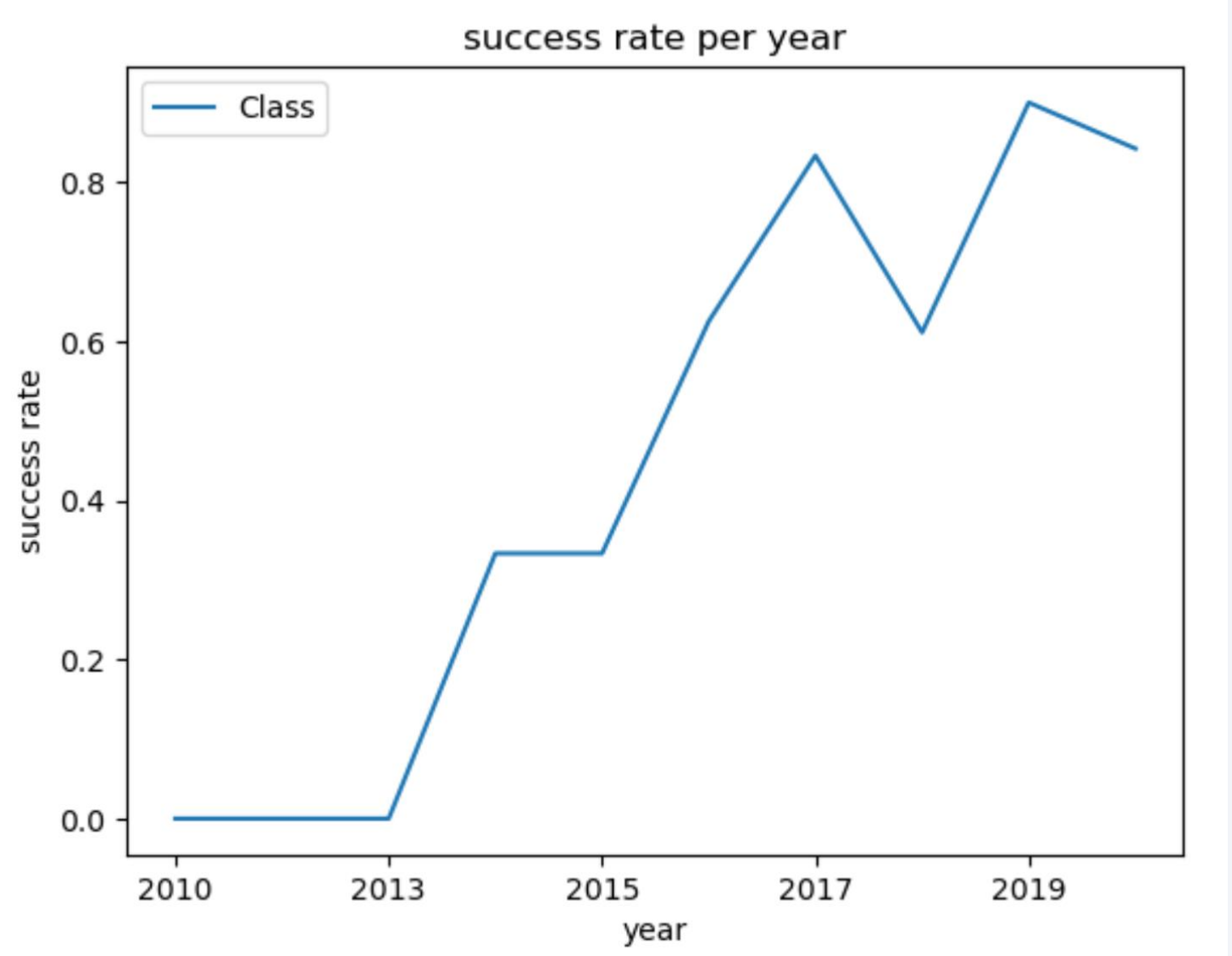


# Payload vs. Orbit Type

- Lower orbits normally have lower payloads.

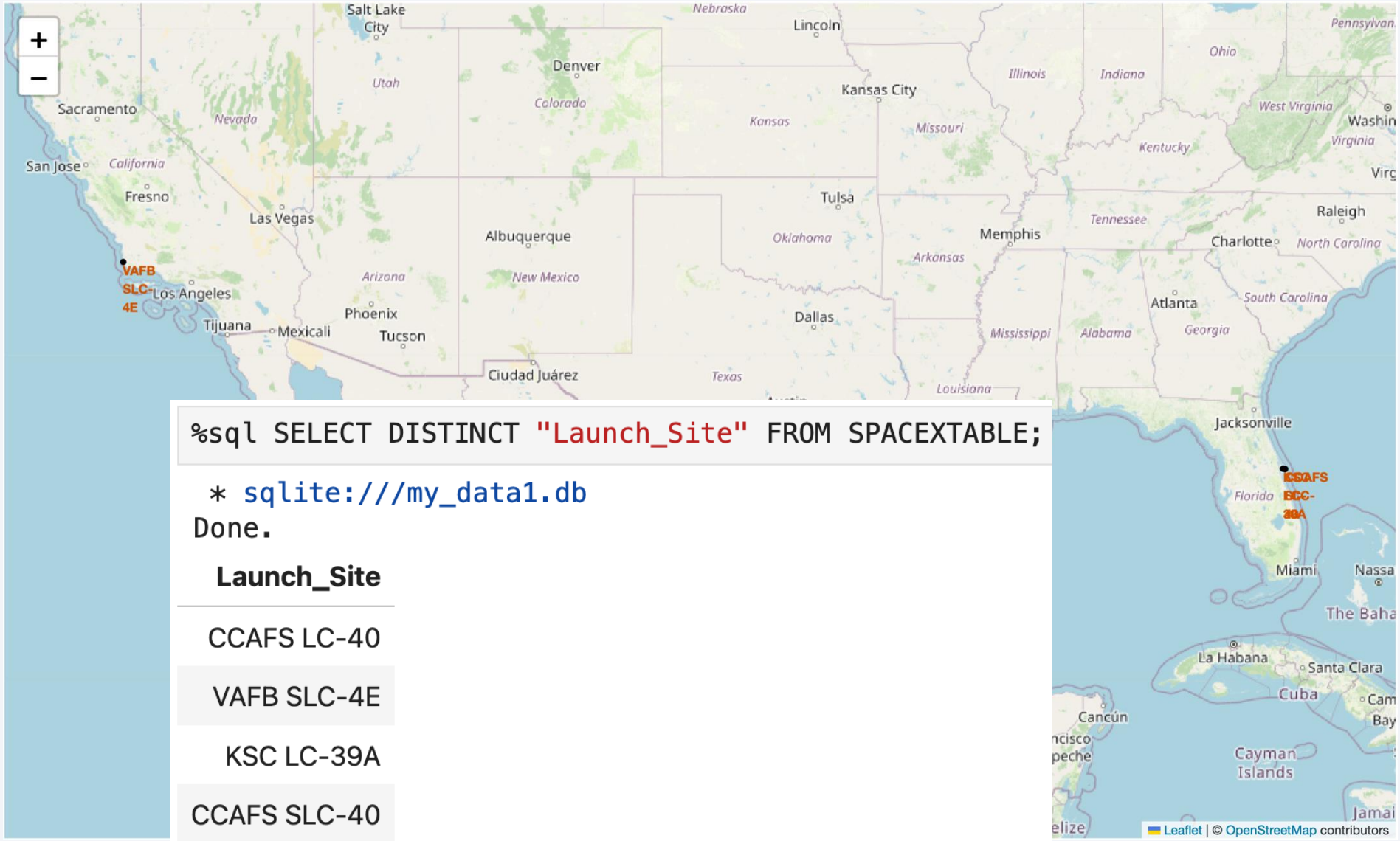


# Launch Success Yearly Trend




- Success rate improved every year from 2013 to 2017 when it started to fluctuate but keeping tendency.

# All Launch Site Names




```
%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 |

- Two main areas for launching site.

# Launch Site Names Begin with 'CCA'

- Querying by launching site.

```
%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 | 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

- Querying by customer.

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "Total payload mass NASA(CRS)" FROM SPACEXTABLE WHERE "Customer" LIKE "NASA (CRS)";
```

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

| Total payload mass NASA(CRS) |
|---|
| 45596 |

# Average Payload Mass by F9 v1.1

- Querying booster by version.

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "Total payload mass F9 v1.1" FROM SPACEXTABLE WHERE "Booster_Version" LIKE "F9 v1.1 %";
```

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

| Total payload mass F9 v1.1 |
|---|
| 2337.8 |

# First Successful Ground Landing Date

- Querying type of outcome and ordered date.

```
%sql SELECT * FROM (SELECT * FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)") AS TBL2 \
     ORDER BY "Date" ASC LIMIT 1;
```

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

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2015-12-22 | 1:29:00 | F9 FT B1019 | CCAFS LC-40 | OG2 Mission 2 11 Orbcomm-OG2 satellites | 2034 | LEO | Orbcomm | Success | Success (ground pad) |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- Sub-querying booster by type of landing between range of payloads.

```
%sql SELECT DISTINCT "Booster_Version" FROM (SELECT * FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)") AS TBL3 \
    WHERE "PAYLOAD_MASS__KG_">4000 AND "PAYLOAD_MASS__KG_"<6000;
#
##or
#
#%sql SELECT DISTINCT "Booster_Version" FROM (SELECT * FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_">4000 AND "PAYLOAD_MASS__KG_"<6000) AS TBL3
#    WHERE "Landing_Outcome" = "Success (drone ship)";
#
##or
#
#%sql SELECT * FROM (SELECT * FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)") AS TBL3 \
#    WHERE "PAYLOAD_MASS__KG_">4000 AND "PAYLOAD_MASS__KG_"<6000;
```

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

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

# Total Number of Successful and Failure Mission Outcomes

- Querying count grouped by itself.

```
%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

- Sub-querying booster by maximun payload.

```
%sql SELECT "Booster_Version" FROM (SELECT * FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_"=15600) AS TBL4 ;
```

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

| 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

- Sub-querying booster failure drone ship by month in 2015.

```
%sql SELECT substr("Date",0,5) AS "Year", substr("Date", 6,2) AS "Month", "Landing_Outcome","Booster_Version","Launch_Site" \
     FROM (SELECT * FROM SPACEXTABLE WHERE "Landing_Outcome"="Failure (drone ship)") AS TBL5 \
     WHERE "Year" = '2015';
```

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

| Year | Month | Landing_Outcome | Booster_Version | Launch_Site |
|---|---|---|---|---|
| 2015 | 01 | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| 2015 | 04 | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

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

- Sub-querying lading outcome count in range of dates in descendant order.

```
%sql SELECT "Landing_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" \
     GROUP BY "Landing_Outcome" ORDER BY "Total" DESC;
```

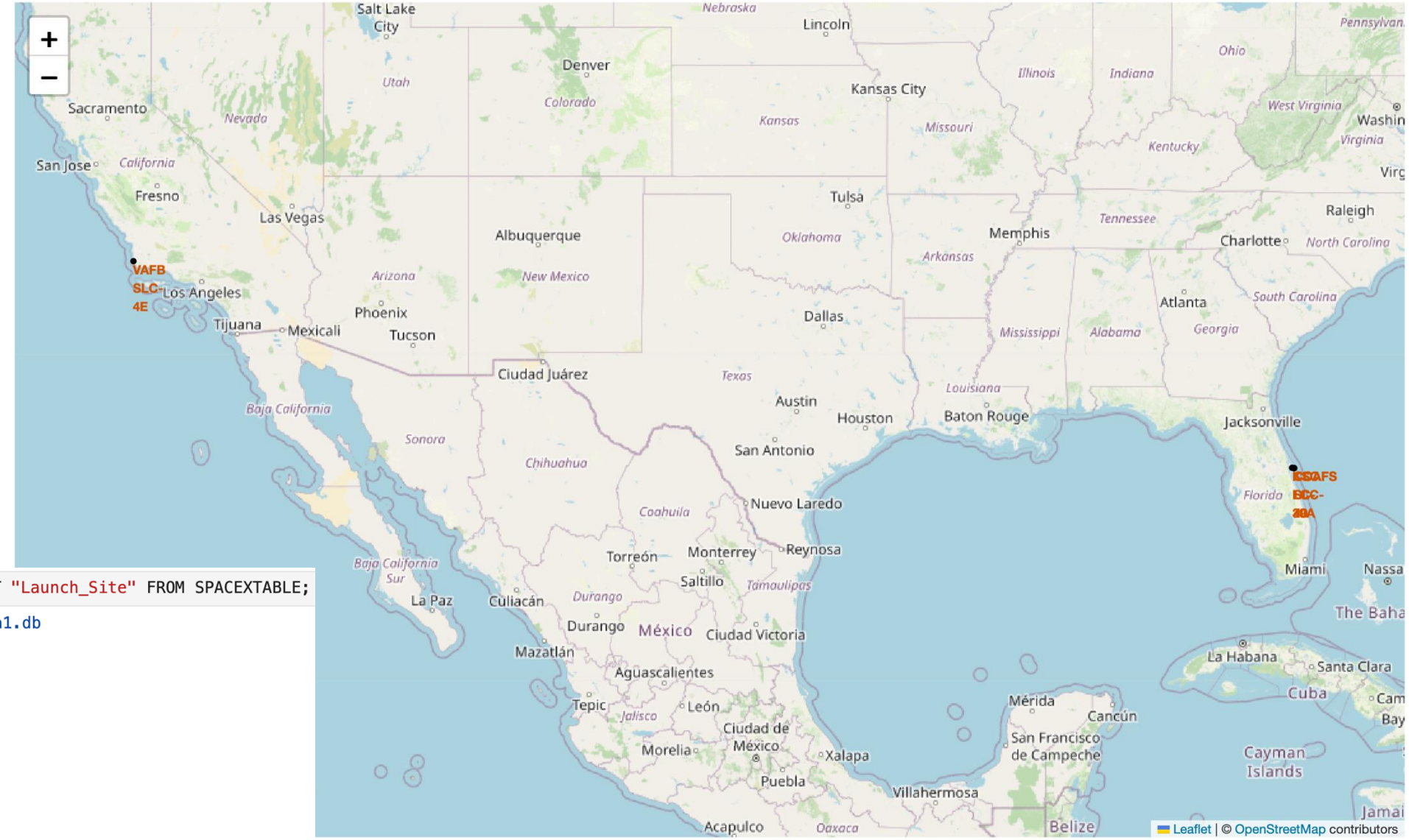
```
 * sqlite:///my_data1.db
Done.
```

| Landing_Outcome | Total |
|---|---|
| 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 |

Section 3

# Launch Sites Proximities Analysis

# Main areas for launching sites: California y Florida



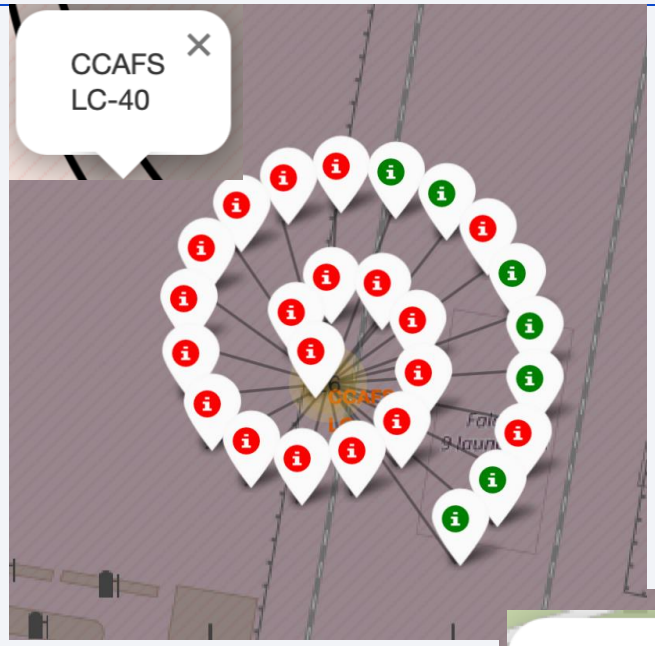

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

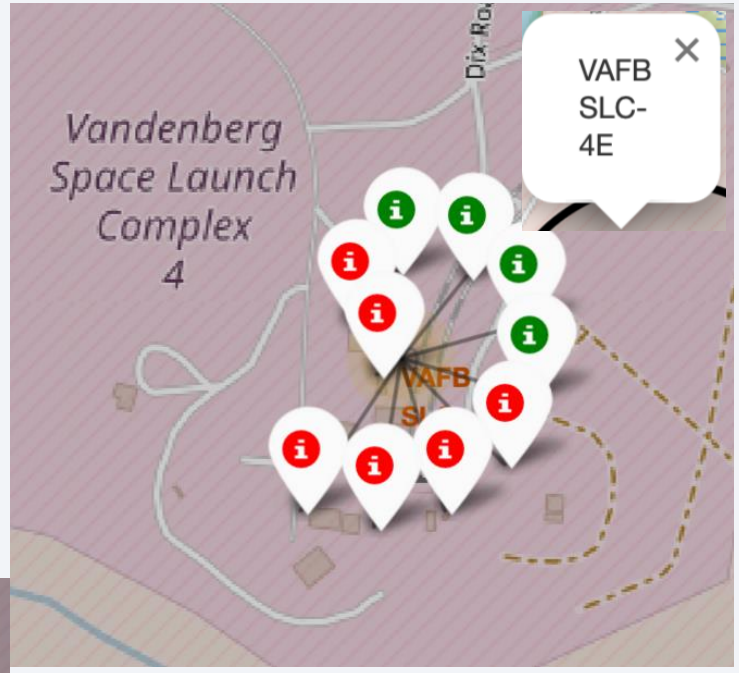
```
 * sqlite:///my_data1.db
Done.
```

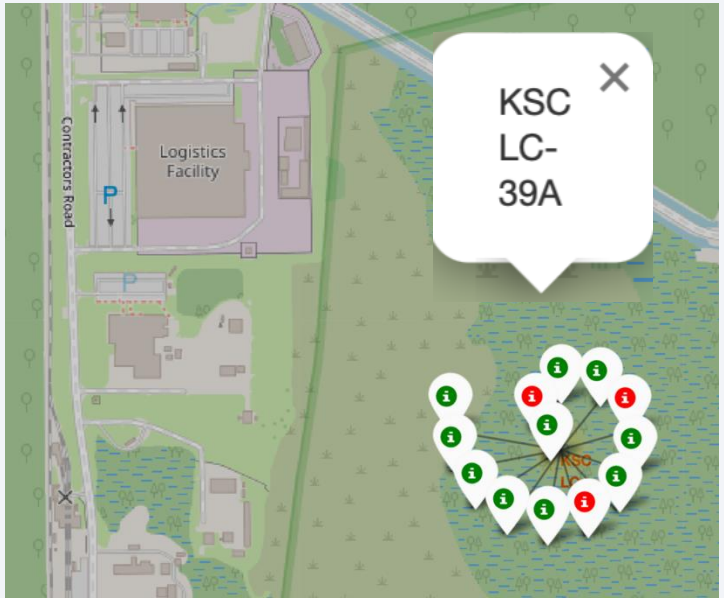
| Launch_Site |
|---|
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |

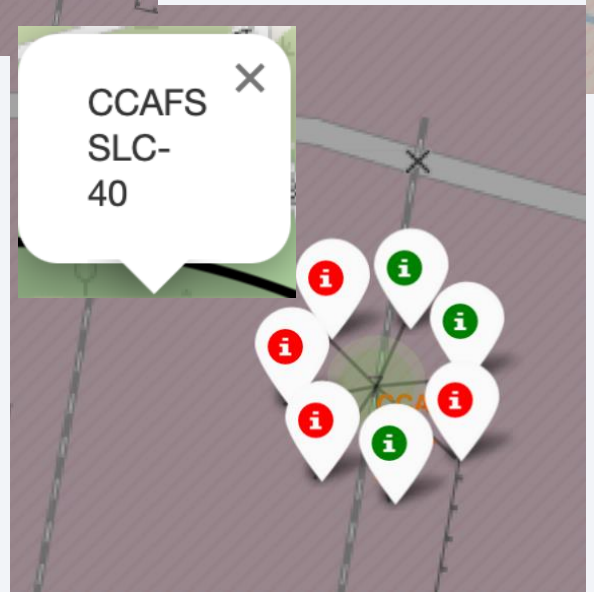
# Launch status.

- Successful: green.
- Failure: red.



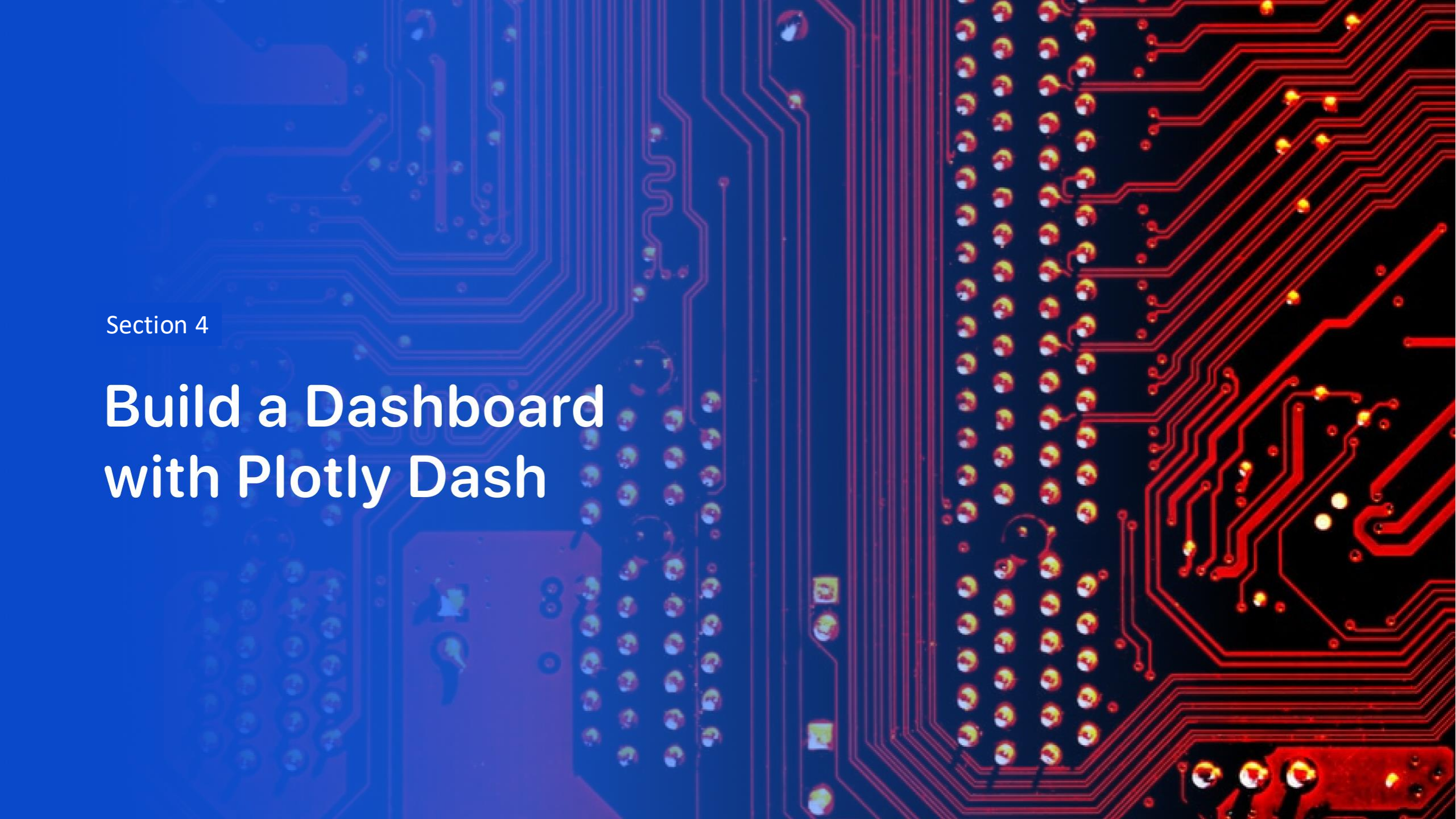



# Proximities to launching site CCAFS SLC 40

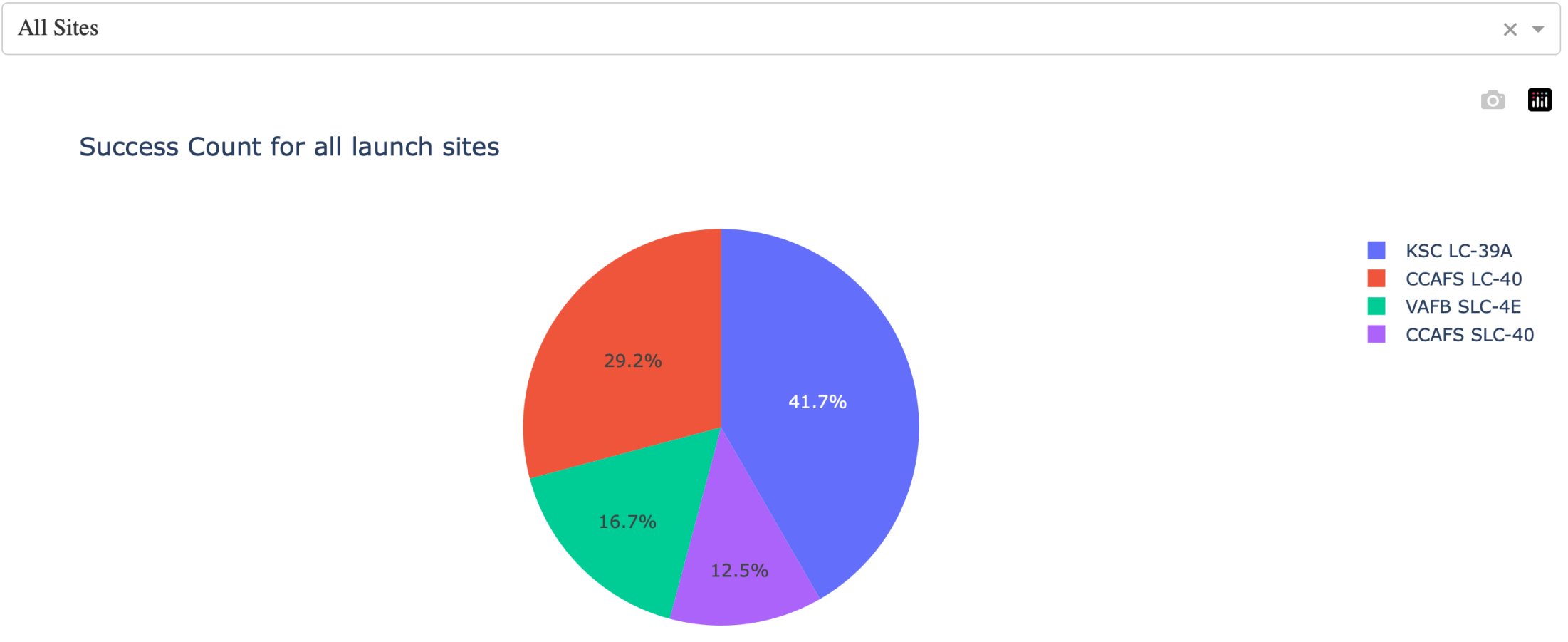


- Near highway, 17.16 km
- Near city, 17.41 km.
- Near coast, 0.87 km
- Near railroad, 0.998 km


Section 4
Build a Dashboard with Plotly Dash

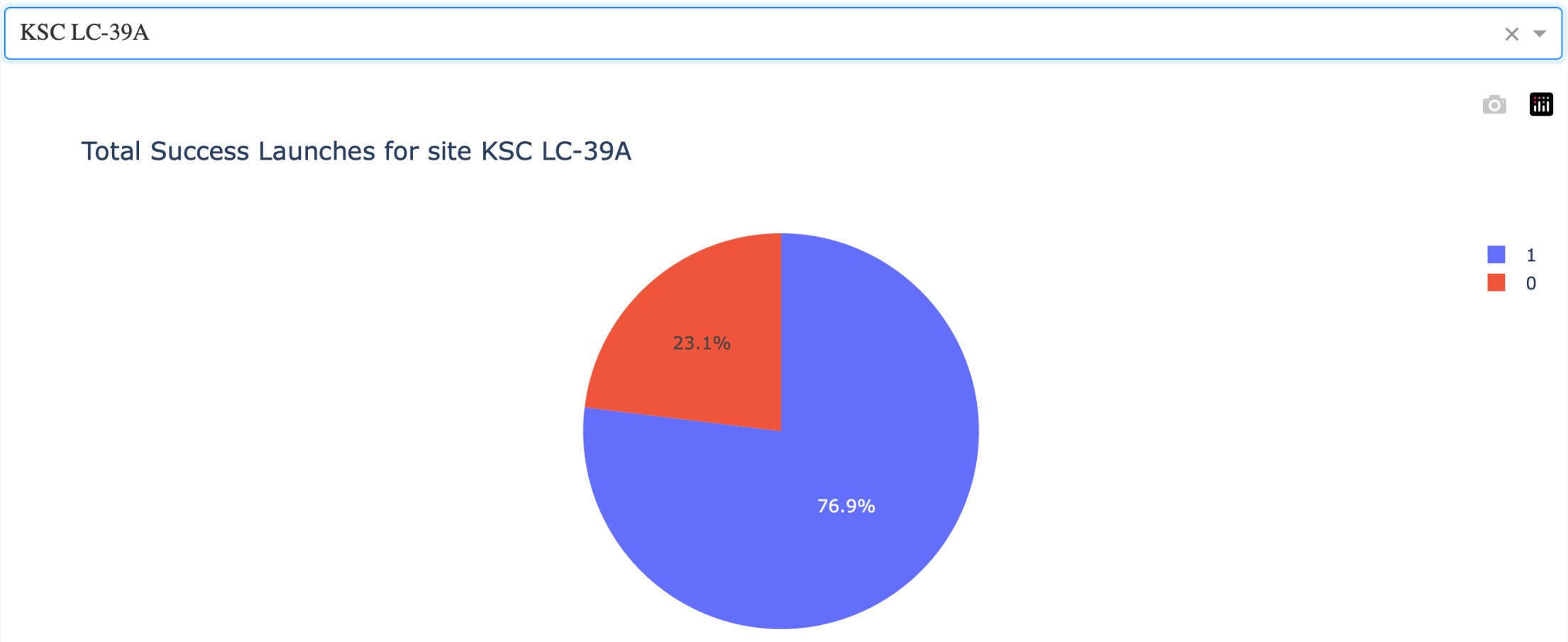
# Success rate for all sites

- Total percentage of all successful launches by site.



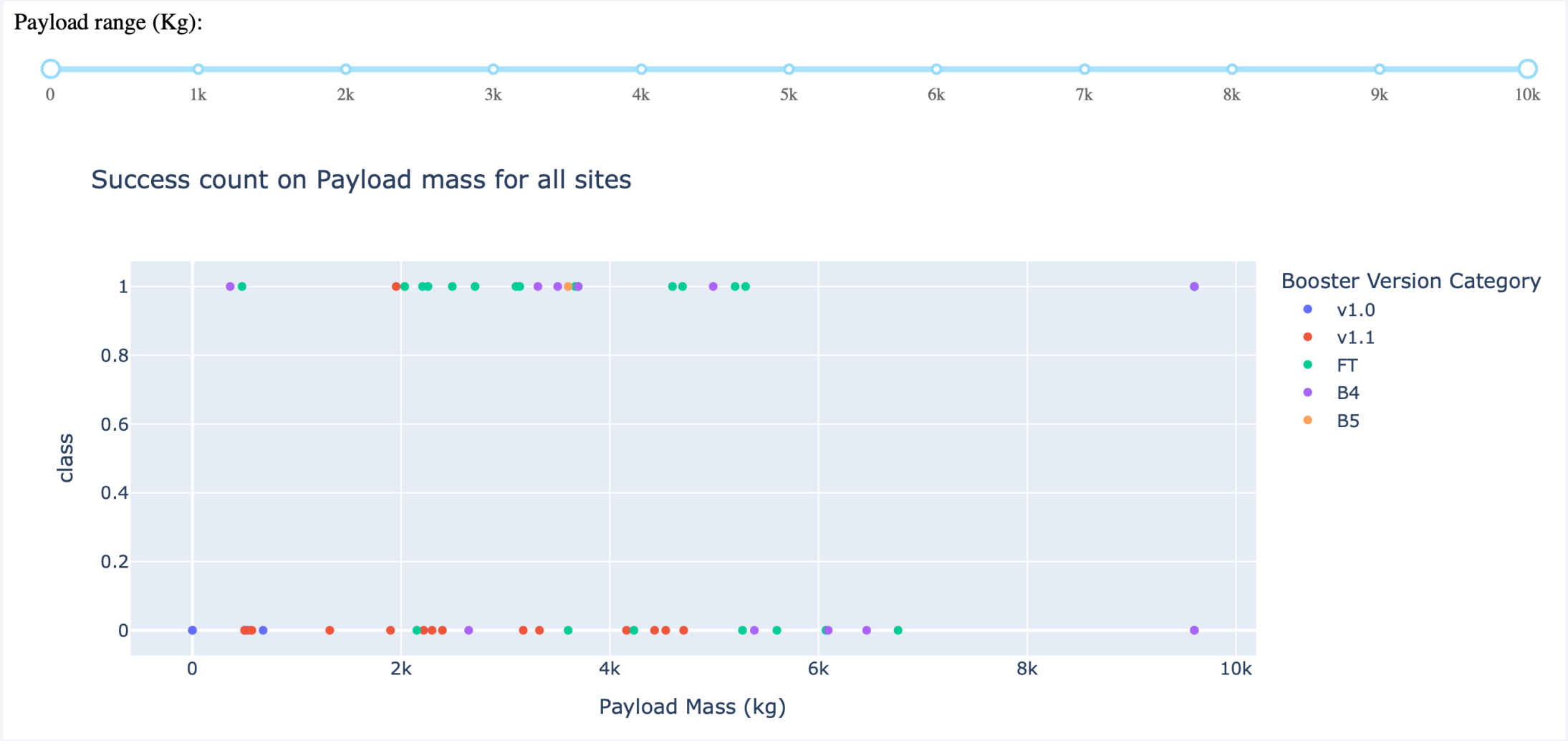
# Site with highest success launching rate

- Kennedy Space Center Launch Complex 39A



# Success count on Payload mass for all sites

- FT booster holds the best success rate by payload mass while v1.1 is the worst.

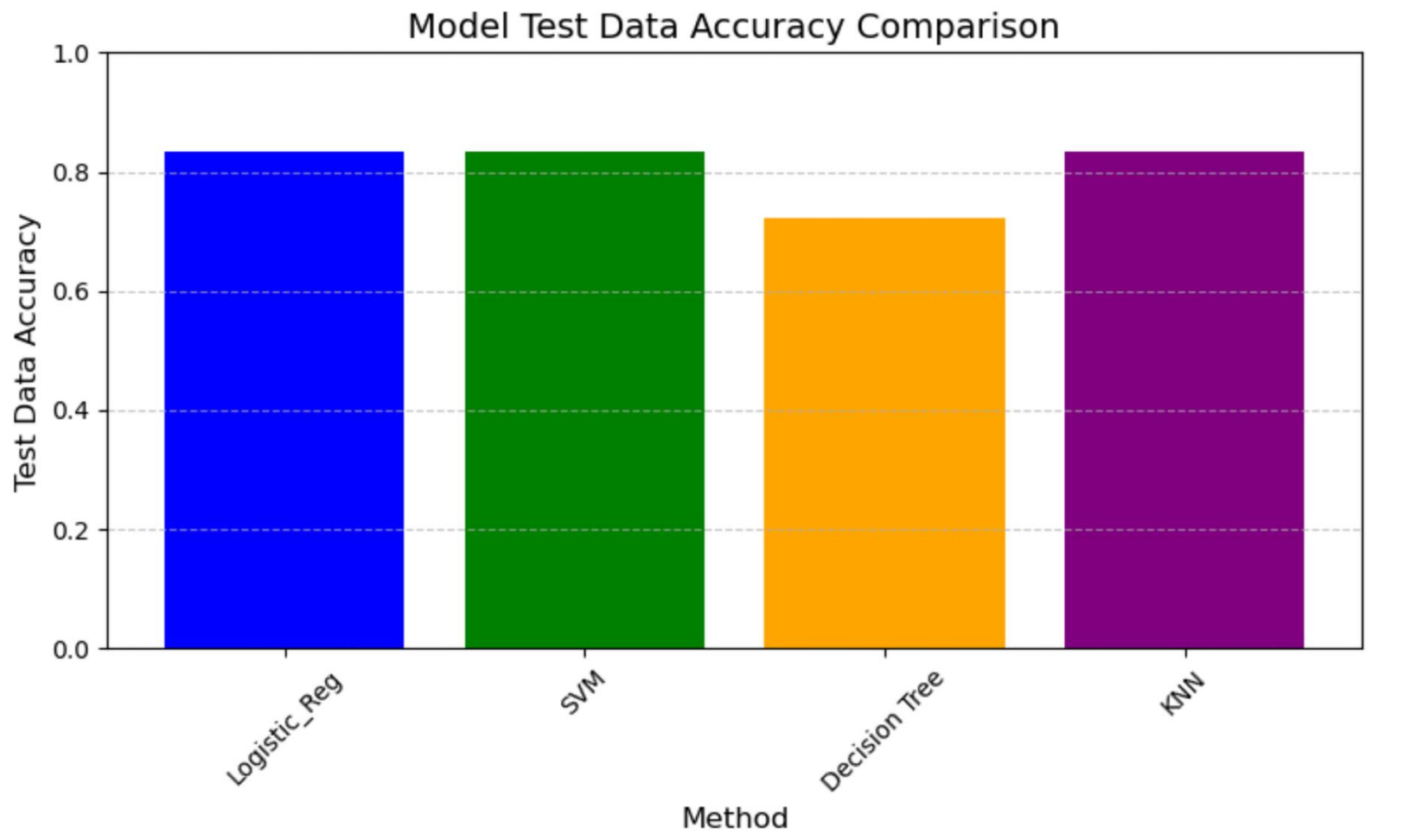


Section 5

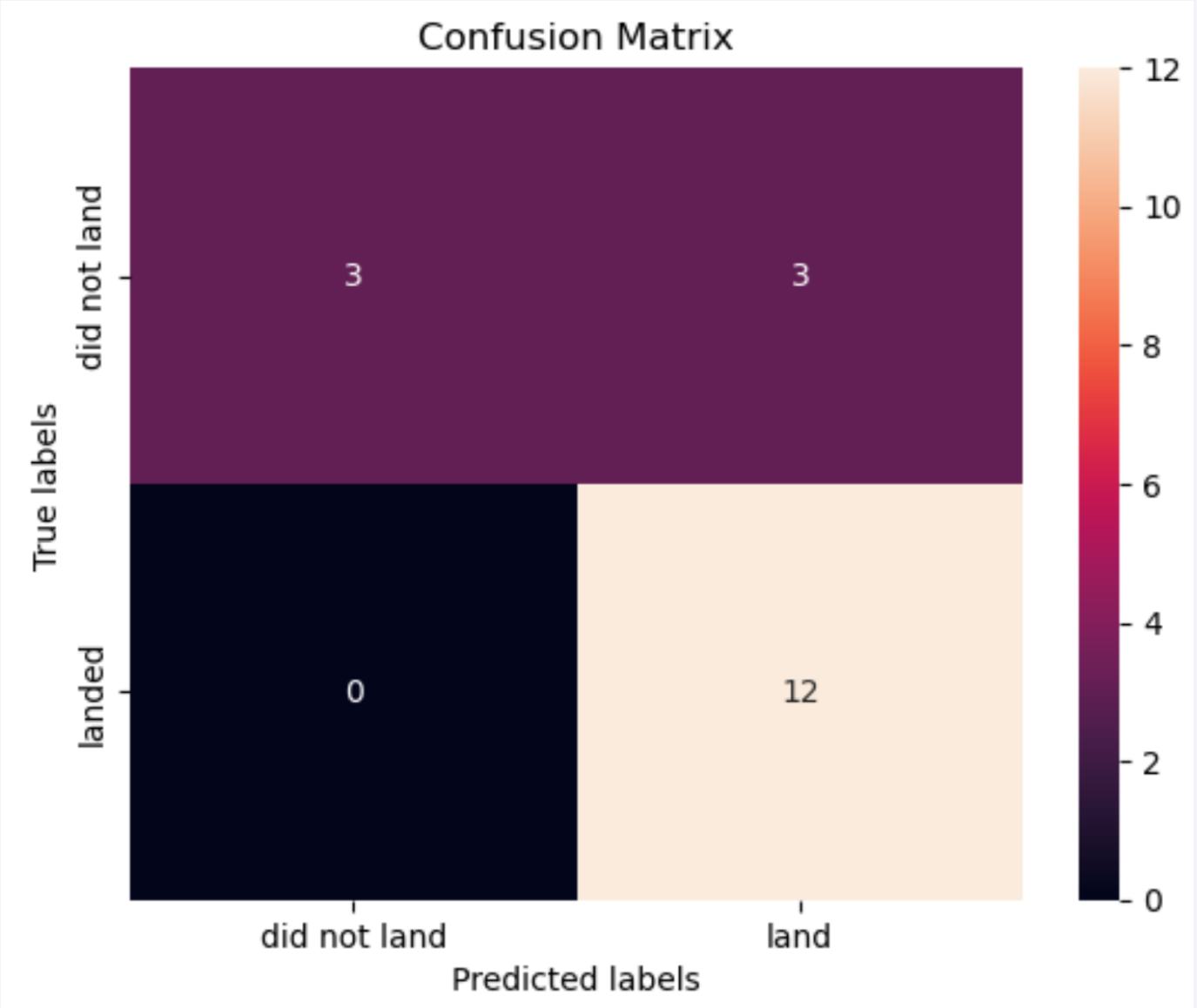
# Predictive Analysis (Classification)

# Classification Accuracy

- Running GridSearchCV change parameters, ideally all of them coincide in .83



# Confusion Matrix




- All of them show similar performance, only Decision Tree have lower performance sometimes

# Conclusions

- Florida Kennedy Space Center Launch Complex 39A is a safer place to launch rockets.
- Heavier payloads are more likely to successful outcome.
- 0.83 the first stage will land successfully based on given data.
- Lower orbits outcomes tend have bad outcomes.
- Success rate launch trend is improving through years.

# Appendix

- All notebooks available in:

https://github.com/MigMares/Applied-Data-Science-Capstone-IBM

# Thank you!