
# Winning Space Race with Data Science

Jesus R Rosila Mares
12 October 2022

# Outline

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

# Executive Summary

Summary of Methodologies

- Web scraped using SpaceX API and python packages such as scikit-learn.
- Cleaned and formatted data for data visualization and machine learning.
- Split data into train and test.

Summary of all results

- KSC launch site had highest success rate at 76.9% out of 13 launches.
- Decision Tree method had highest accuracy at 88.9% with only 2 misclassifications
- First Booster Stage landing on May 1, 2017

# Introduction

Project background and context

- Investigate SpaceX Falcon 9 First-stage Booster Recovery
- Determine if flight will be a success or failure
- Implement Machine Learning to Classify future launches successes

Problems you want to find answers

- What type of Orbits and Payload Mass are being used? What is the success rate?
- Where and when are they launching these missions?
- How has years of experience increased SpaceX success
- How well does trained models predict success

Section 1

# Methodology

# Methodology

## Executive Summary

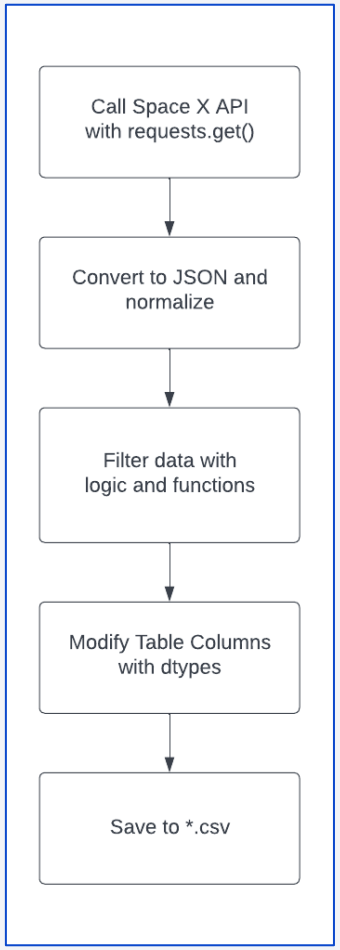
- Data collection methodology:
  - Data Scraping using SpaceX API
- Perform data wrangling
  - Filtered and Used Regex to collected data from JSON
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Compare with: SVM, Tree, Log Reg, and KNN with GridSearchCV

# Data Collection

- SpaceX API Requests
- Clean data
- Down select to hone-in only to Falcon 9 launches
- Fix null and missing values
- Use in following data analysis

# Data Collection – SpaceX API

- Use SpaceX API to scrape data for Statistical Data Analysis and Machine Learning

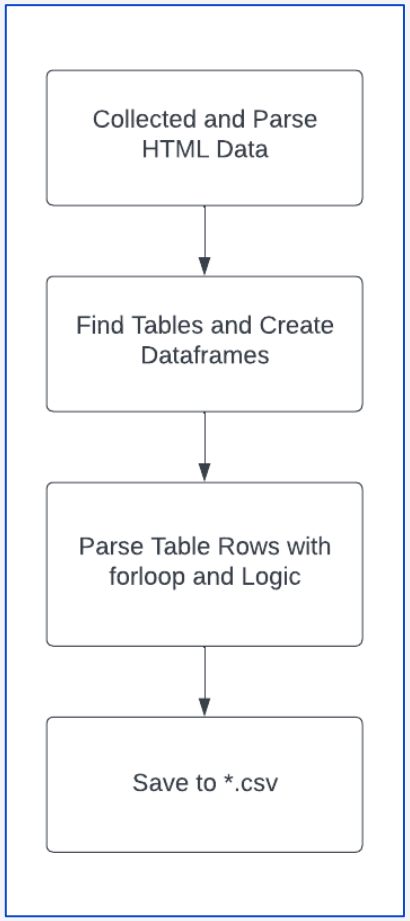



- Notebook @ Git

# Data Collection - Scraping

- Scrape data from generated CSV files
- Apply filters to get desired output

```
{'Flight No.': 121,
'Launch site': 121,
'Payload': 121,
'Payload mass': 121,
'Orbit': 121,
'Customer': 121,
'Launch outcome': 121,
'Version Booster': 121,
'Booster landing': 121,
'Date': 121,
'Time': 121}
```




- Notebook @ Git

# Data Wrangling

- Classify Success and Failures from scraped data
- True ASDS, None None, True RTLS, False ASDS, True Ocean, False Ocean, None ASDS, False RTLS
- Collect success and failures to ‘Class’ column
- Prep for Data Analysis

- Notebook @ Git

# EDA with Data Visualization

- Success Rates for each Orbit type
- Payload Mass per each Orbit type
- Payload Mass per Flight Number
- Flights per Launch Site and Classified Success

- Notebook @ Git

# EDA with SQL

- Query database for DISTINCT site launch locations
- Find how many NASA missions there were
- Average Payload Mass for F9 v1.1 Booster
- First Ground Success
- Sub-Query for Booster landing failures

- Notebook @ Git

# Build an Interactive Map with Folium

- Mark all Falcon 9 Launch Sites
- Use ‘Class’ Label to define success and failures
- Find distances of Points of Interests to Launch Sites

- Notebook @ Git

# Build a Dashboard with Plotly Dash

- Build Dash dashboard for manger use
- Inputs: Launch Sites and Payload Mass slider
- Populate with Success or Success Rate Pie Charts
- Display Launch Site specific ‘Class’ within Payload Mass slider

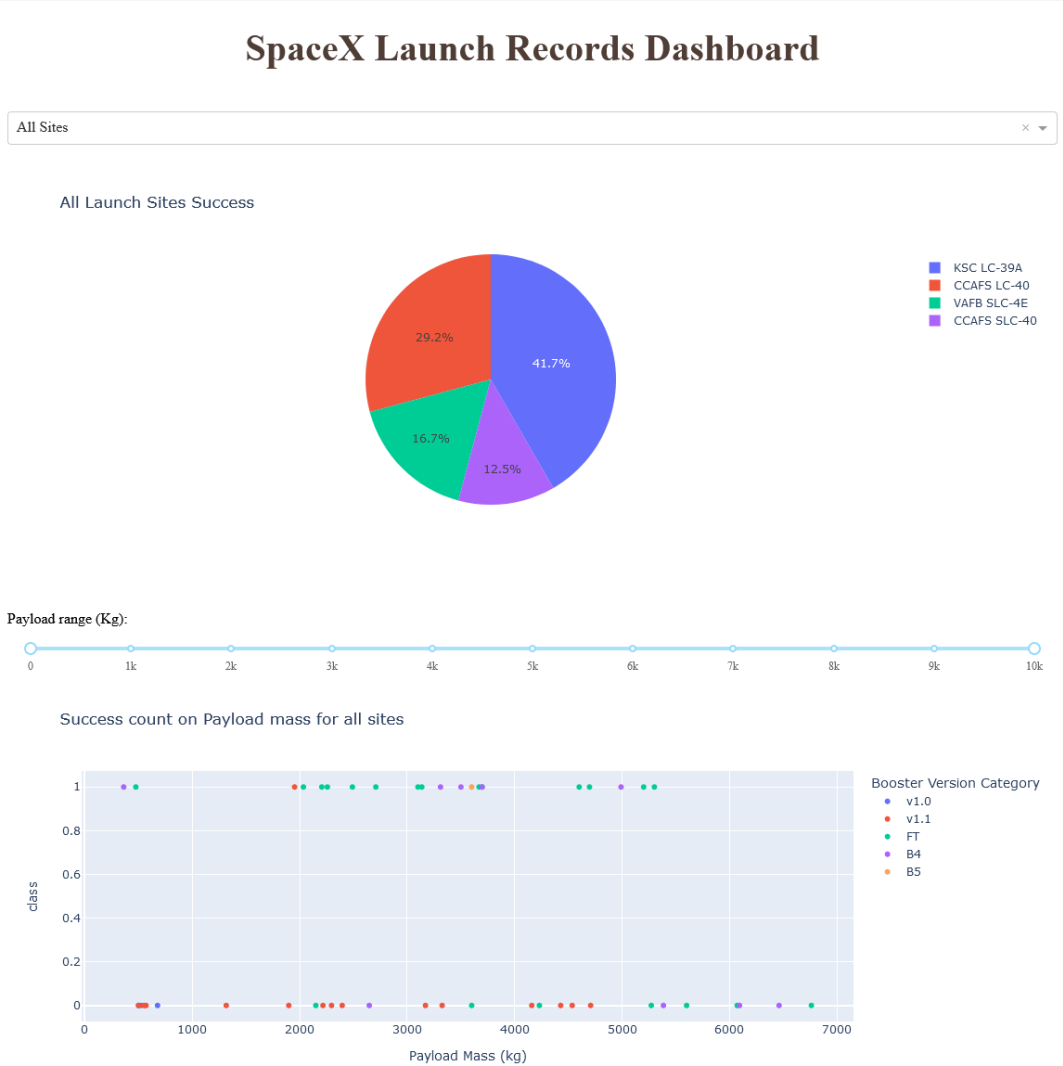
- Notebook @ Git

# Predictive Analysis (Classification)

- One Hot Encoding
- Machine Learning Methods
  - SVM, Tree, Log Reg, and KNN with GridSearchCV
  - Compare with accuracies
- Confusion Matrices

- Notebook @ Git

# Results

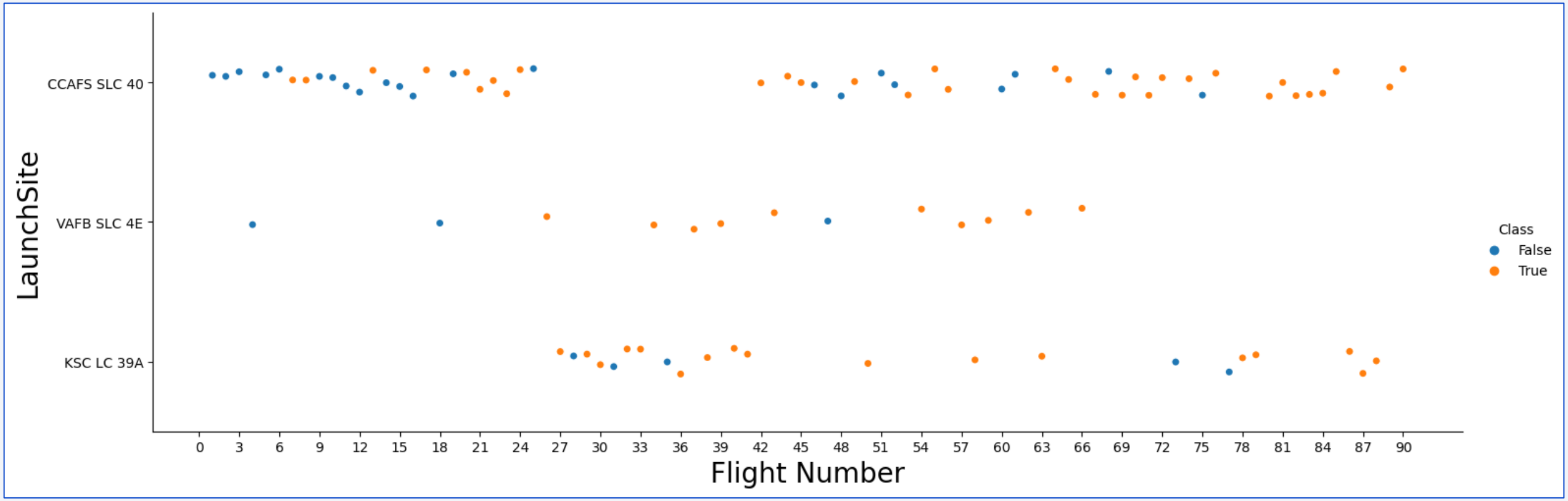



| | Orbit | success | attempts | SuccessRate |
|---|---|---|---|---|
| 0 | ES-L1 | 1 | 1 | 1.000000 |
| 1 | GEO | 1 | 1 | 1.000000 |
| 2 | GTO | 14 | 27 | 0.518519 |
| 3 | HEO | 1 | 1 | 1.000000 |
| 4 | ISS | 13 | 21 | 0.619048 |
| 5 | LEO | 5 | 7 | 0.714286 |
| 6 | MEO | 2 | 3 | 0.666667 |
| 7 | PO | 6 | 9 | 0.666667 |
| 8 | SO | 0 | 1 | 0.000000 |
| 9 | SSO | 5 | 5 | 1.000000 |

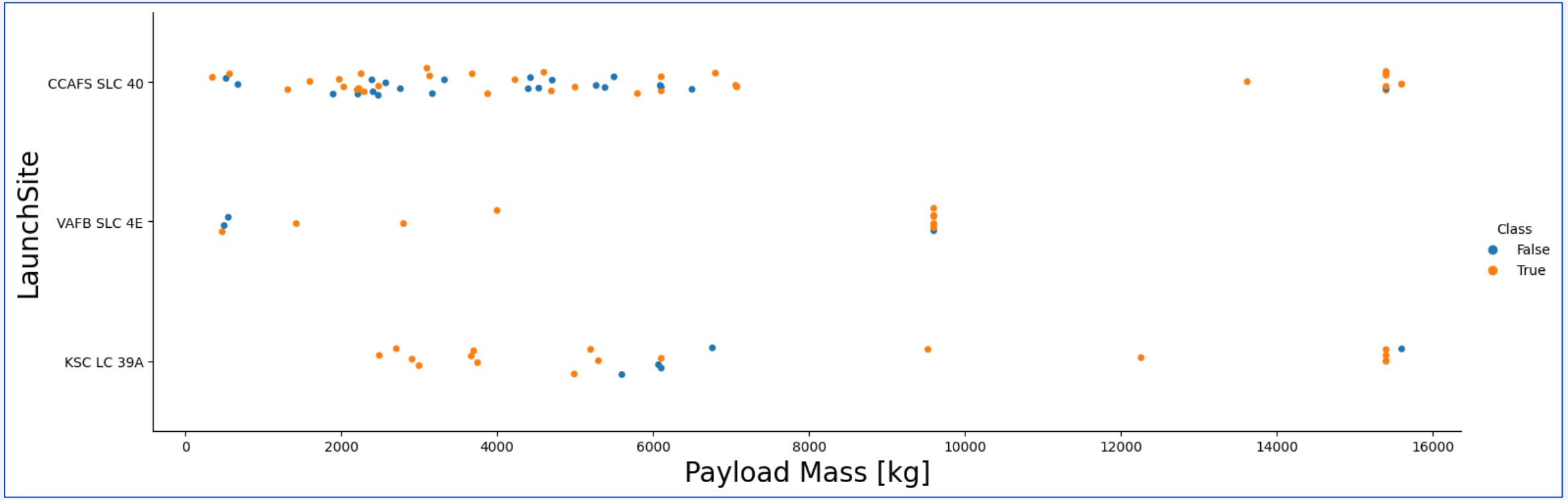
Section 2

# Insights drawn from EDA

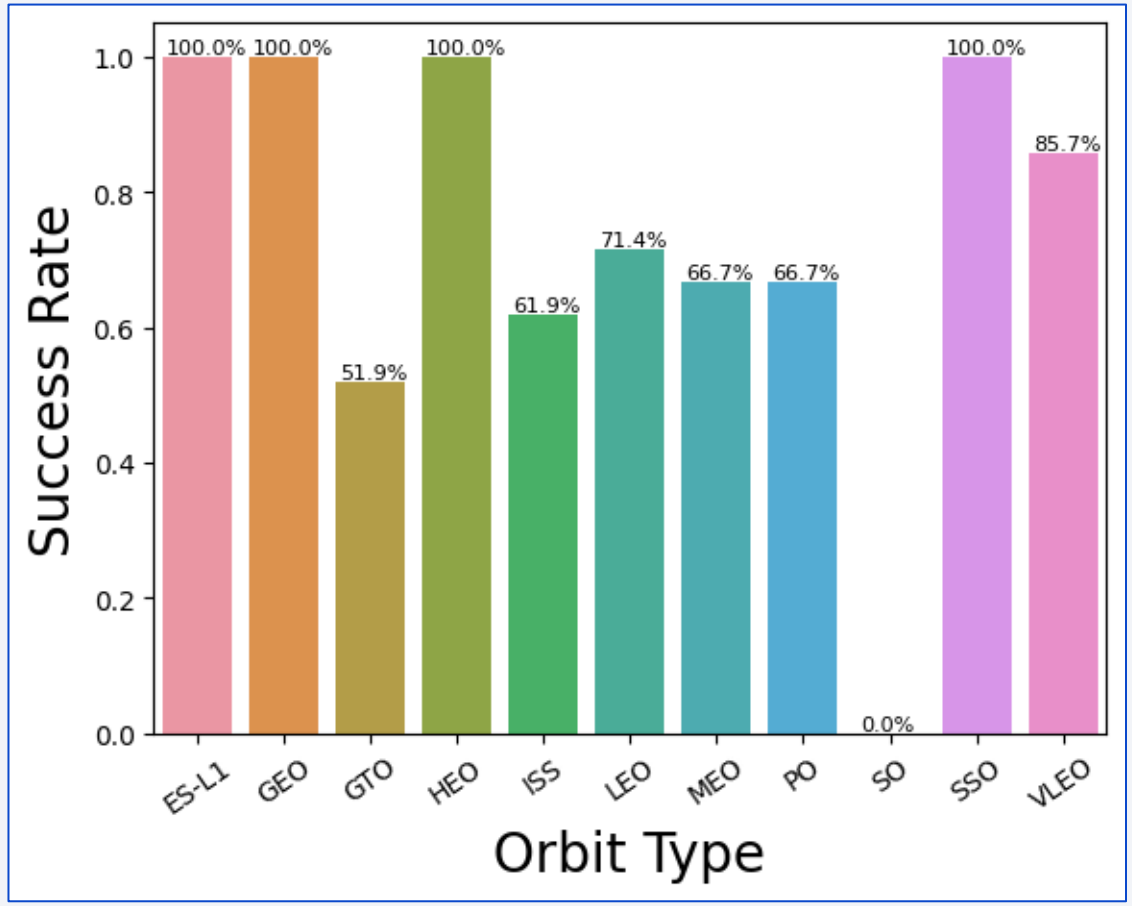
# Launch Site vs. Flight Number



# Launch Site vs. Payload

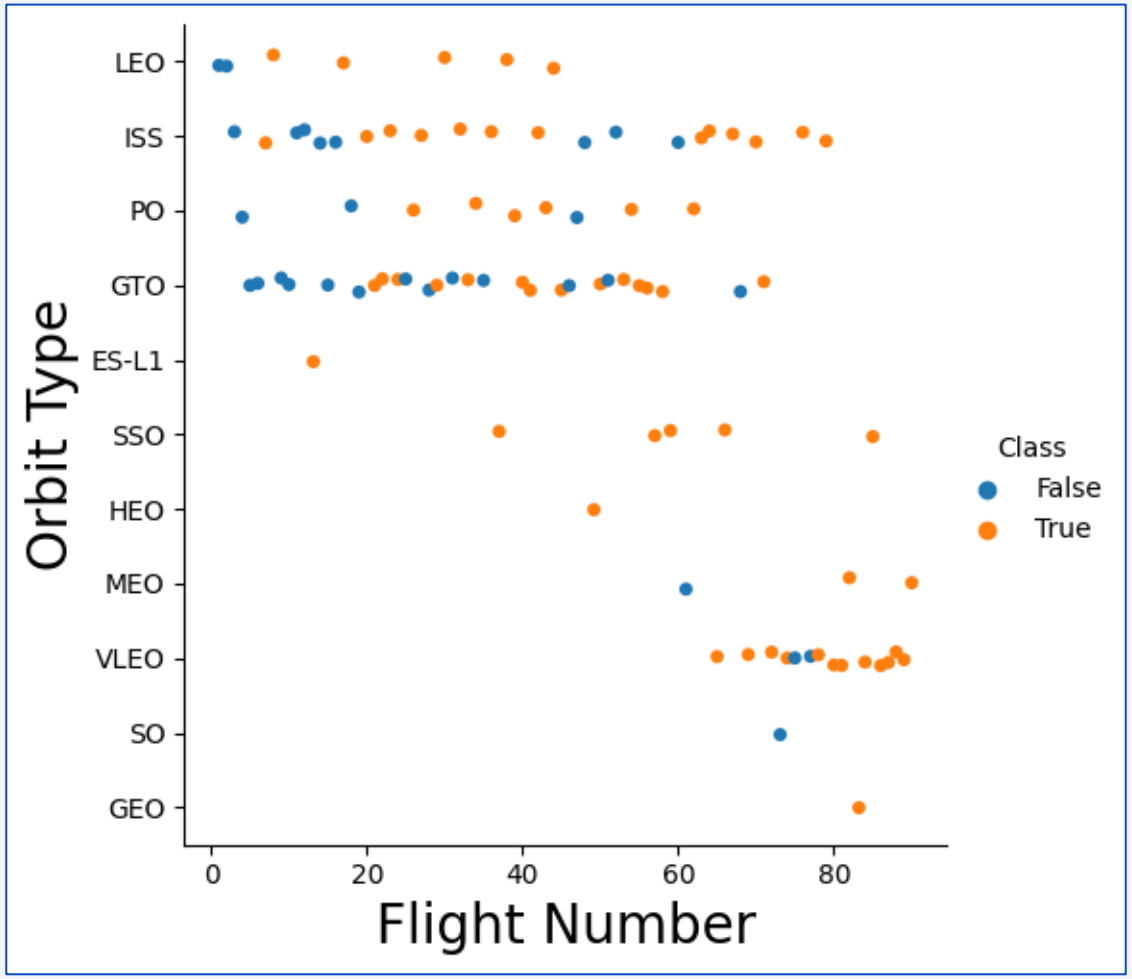


# Success Rate vs. Orbit Type

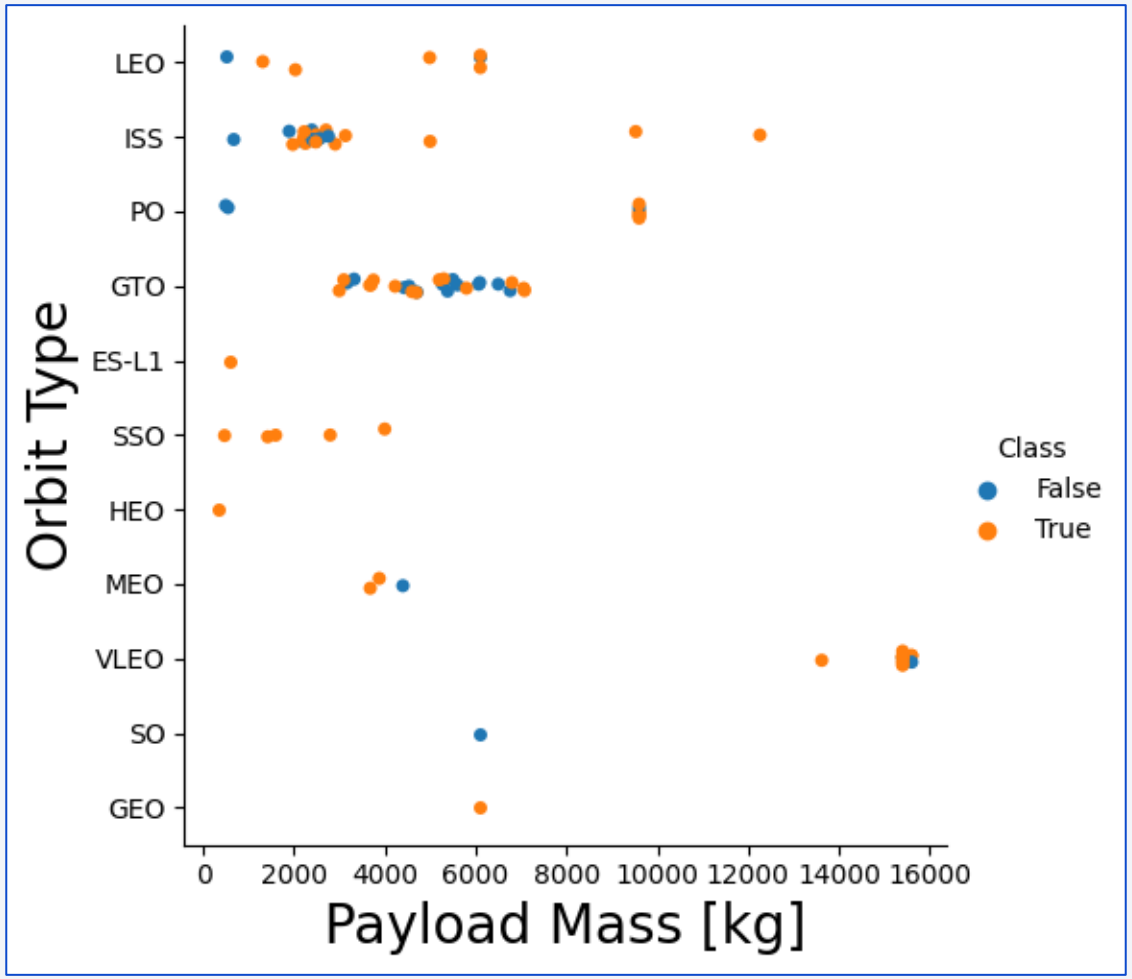



| Orbit Shorthand | Orbit Name |
|---|---|
| ES-L1 | Heliocentric Lagrange Point 1 |
| GEO | Geocentric Orbit |
| HEO | Heliocentric Orbit |
| ISS | LEO to International Space Station |
| LEO | Low Earth Orbit |
| MEO | Medium Earth Orbit |
| PO | Polar Orbit |
| SO | Sub-Orbital |
| SSO | Sun-Synchronous Orbit |
| VELO | Very Low Earth Orbit |

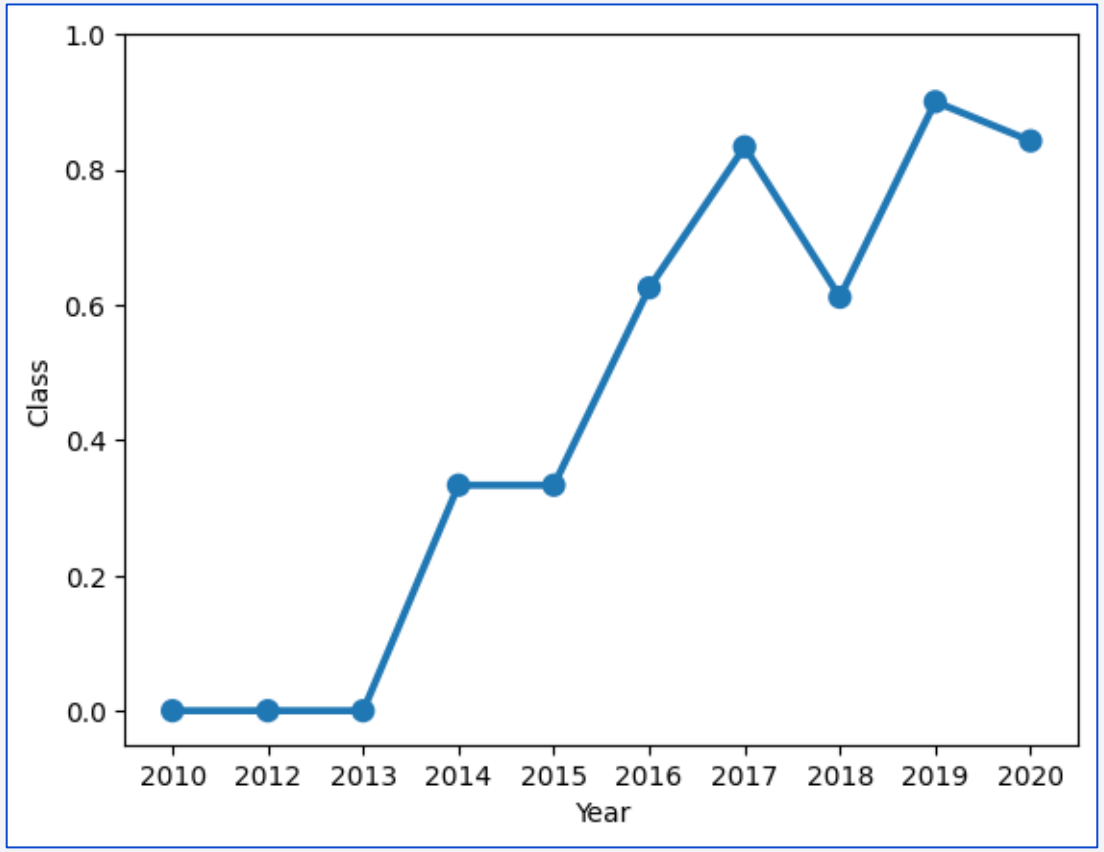
# Orbit Type vs. Flight Number



# Orbit Type vs. Payload



# Launch Success Yearly Trend



# All Launch Site Names

```
%%sql
SELECT DISTINCT Launch_Site
FROM spacex_data
ORDER BY Launch_Site;
```

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

# Launch Site Names Begin with 'CCA'

```
%%sql
SELECT *
FROM spacex_data
WHERE Launch_Site like 'CCA%'
LIMIT 5;
```

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 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) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

# Total Payload Mass

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) as 'Total Payload Mass kg'
FROM spacex_data
WHERE UCASE(customer) LIKE '%NASA%';
```

| Total Payload Mass kg |
| --- |
| 107010 |

# Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) as avg_mass
FROM spacex_data
WHERE UCASE(Booster_Version) LIKE '%F9 V1.1%';
```

| avg_mass |
| --- |
| 2534.6667 |

# First Successful Ground Landing Date

```
%%sql
SELECT MIN(Date) as 'First Success Ground'
FROM spacex_data
WHERE LCASE(Landing_Outcome) LIKE 'success%ground%';
```

| First Success Ground |
| --- |
| 01-05-2017 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT Booster_Version
FROM spacex_data
WHERE LCASE(Landing_Outcome) LIKE 'success%drone%'
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

# Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) as Mission_Freq
FROM spacex_data
GROUP BY Mission_Outcome;
```

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

# Boosters Carried Maximum Payload

```
%%sql
SELECT DISTINCT Booster_Version
FROM spacex_data
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)
                           FROM spacex_data);
```

| 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

```
%%sql
SELECT DATE_FORMAT(Date,"%M") as Month, Landing_Outcome, Booster_Version, Launch_Site
FROM (Select STR_TO_DATE(Date, '%d-%m-%Y') as Date, Landing_Outcome, Booster_Version, Launch_Site
      FROM spacex_data) as spacex_data
WHERE LCASE(Landing_Outcome) LIKE 'failure%drone%'
AND YEAR(Date) = 2015;
```

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

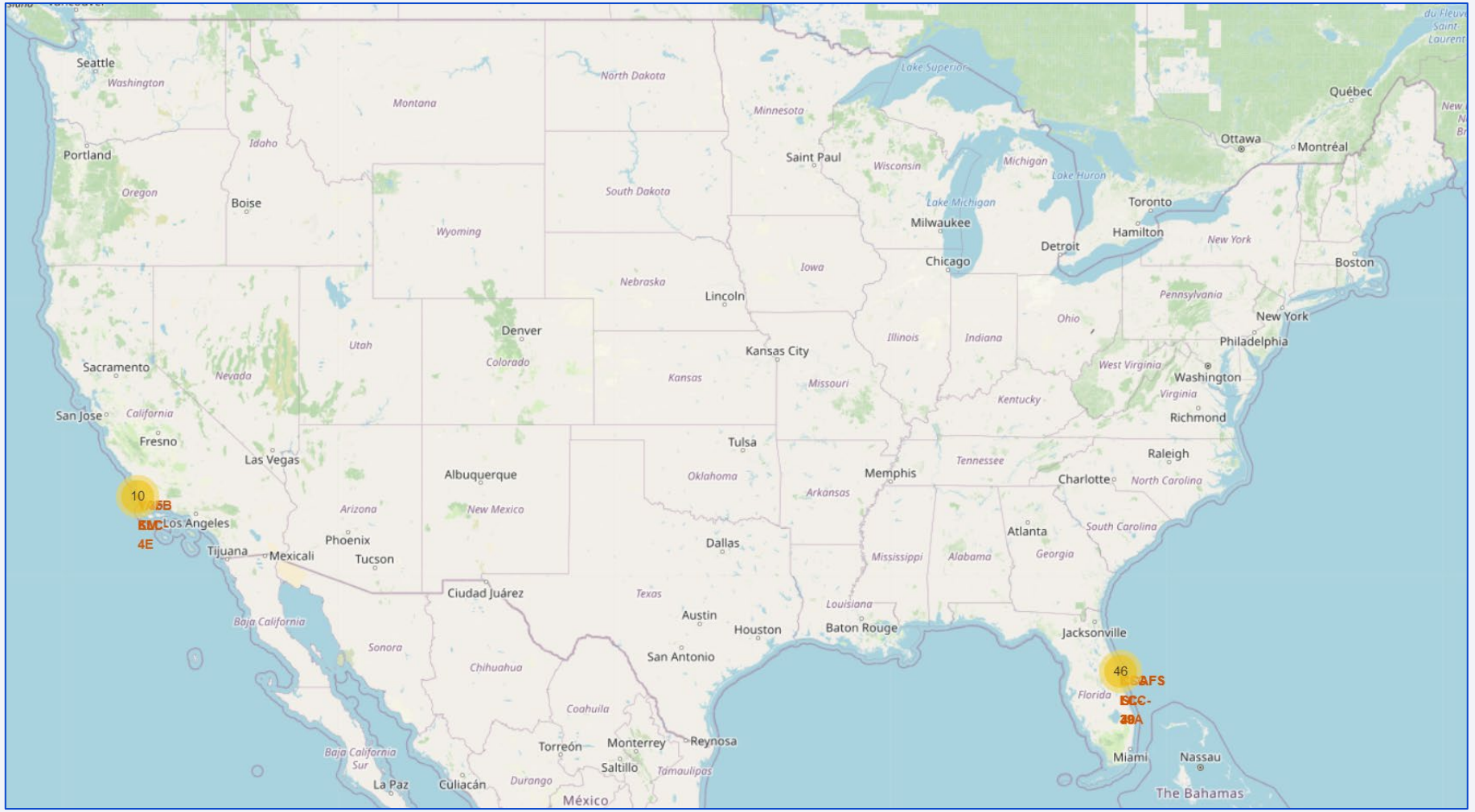
# Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
%%sql
SELECT Landing_Outcome, COUNT(Landing_Outcome) as Landing_Outcome_Freq
FROM (Select STR_TO_DATE(Date, '%d-%m-%Y') as Date, Landing_Outcome
      FROM spacex_data) as spacex_data
WHERE LCASE(Landing_Outcome) LIKE '%success%'
AND Date BETWEEN CAST('2010-06-04' AS DATE) AND CAST('2017-03-20' AS DATE)
GROUP BY Landing_Outcome
ORDER BY Date DESC;
```

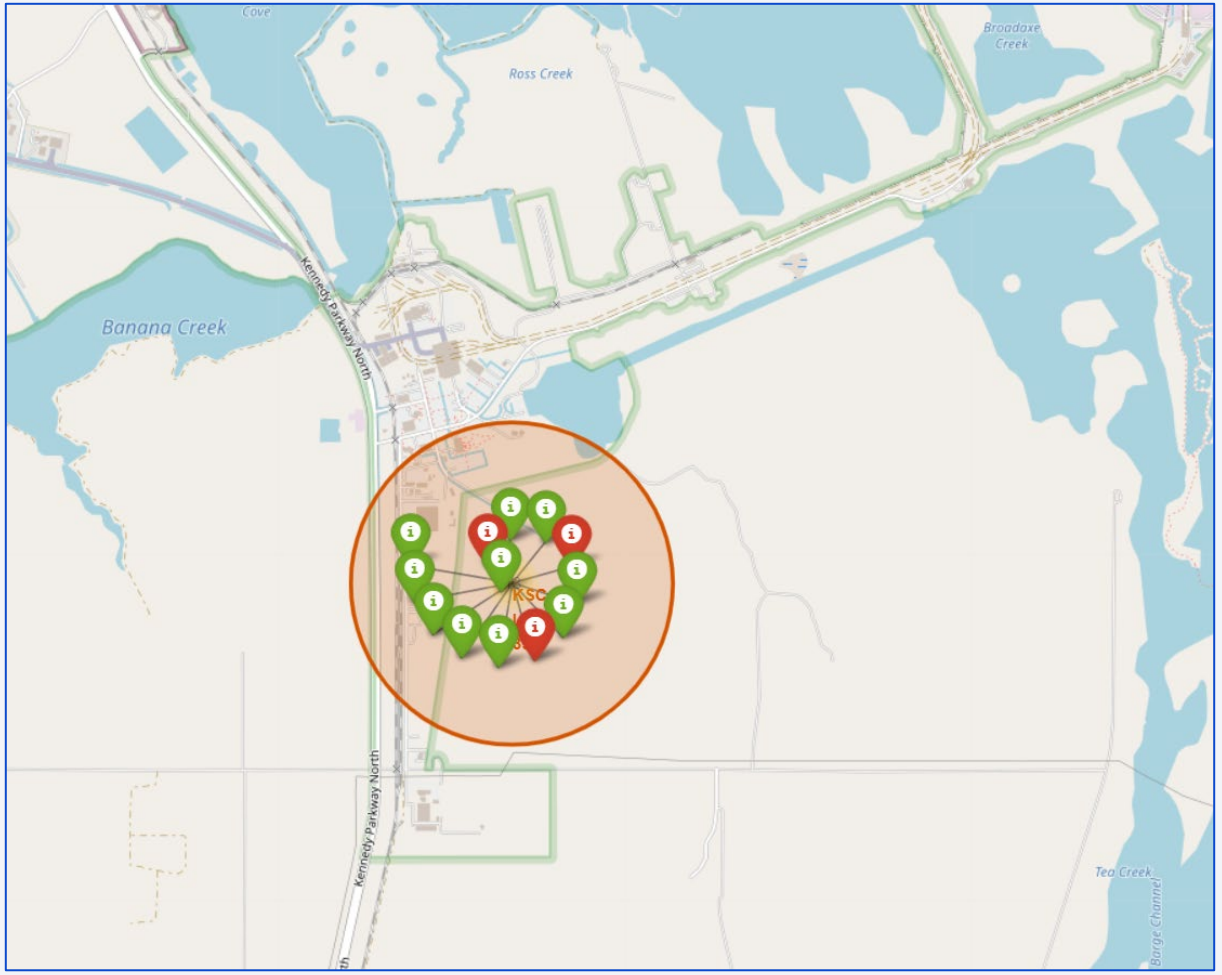
| Landing_Outcome | Landing_Outcome_Freq |
|---|---|
| Success (drone ship) | 5 |
| Success (ground pad) | 3 |


Section 3
Launch Sites
Proximities Analysis

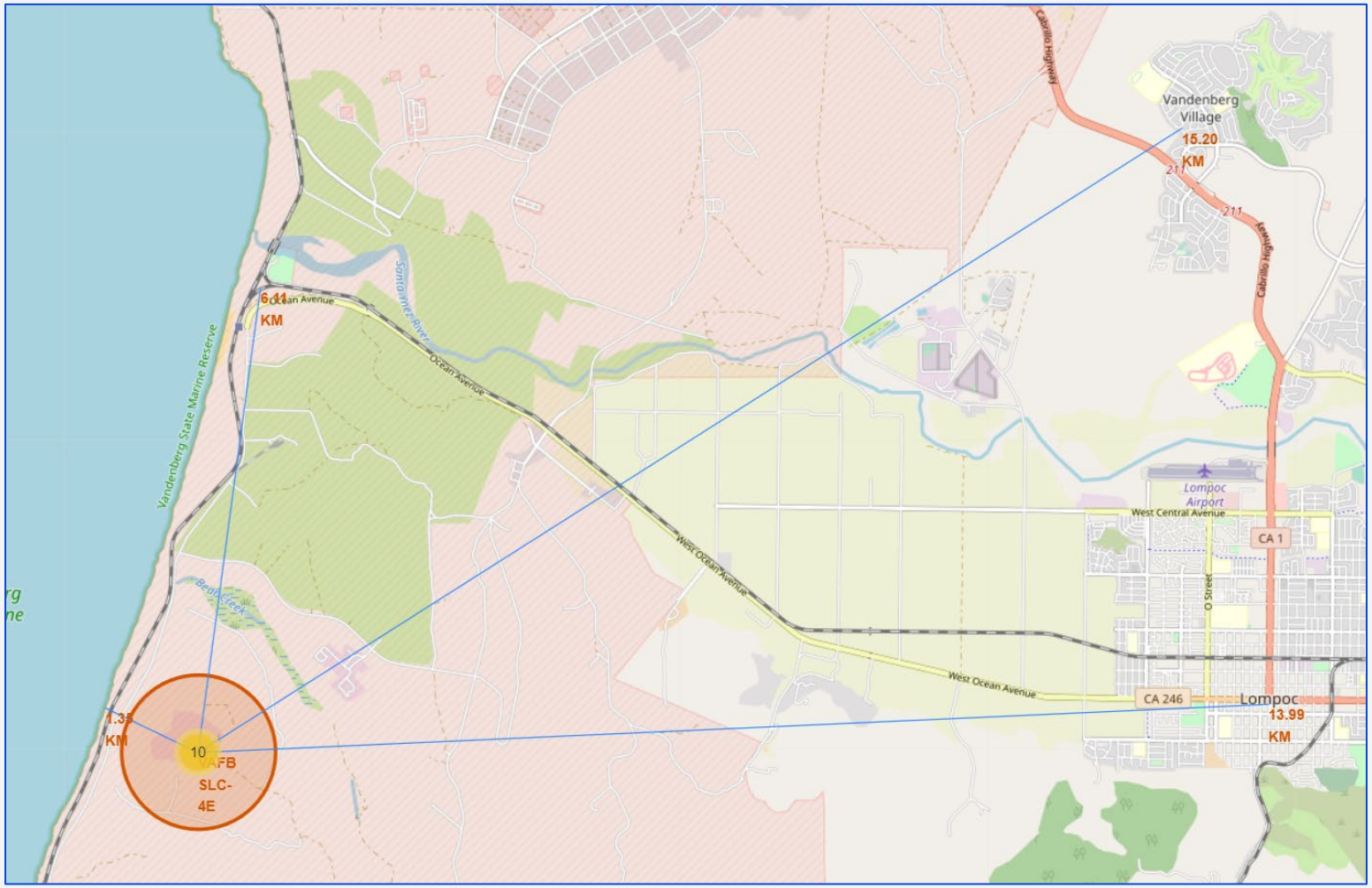
# Map of All Space X Launches

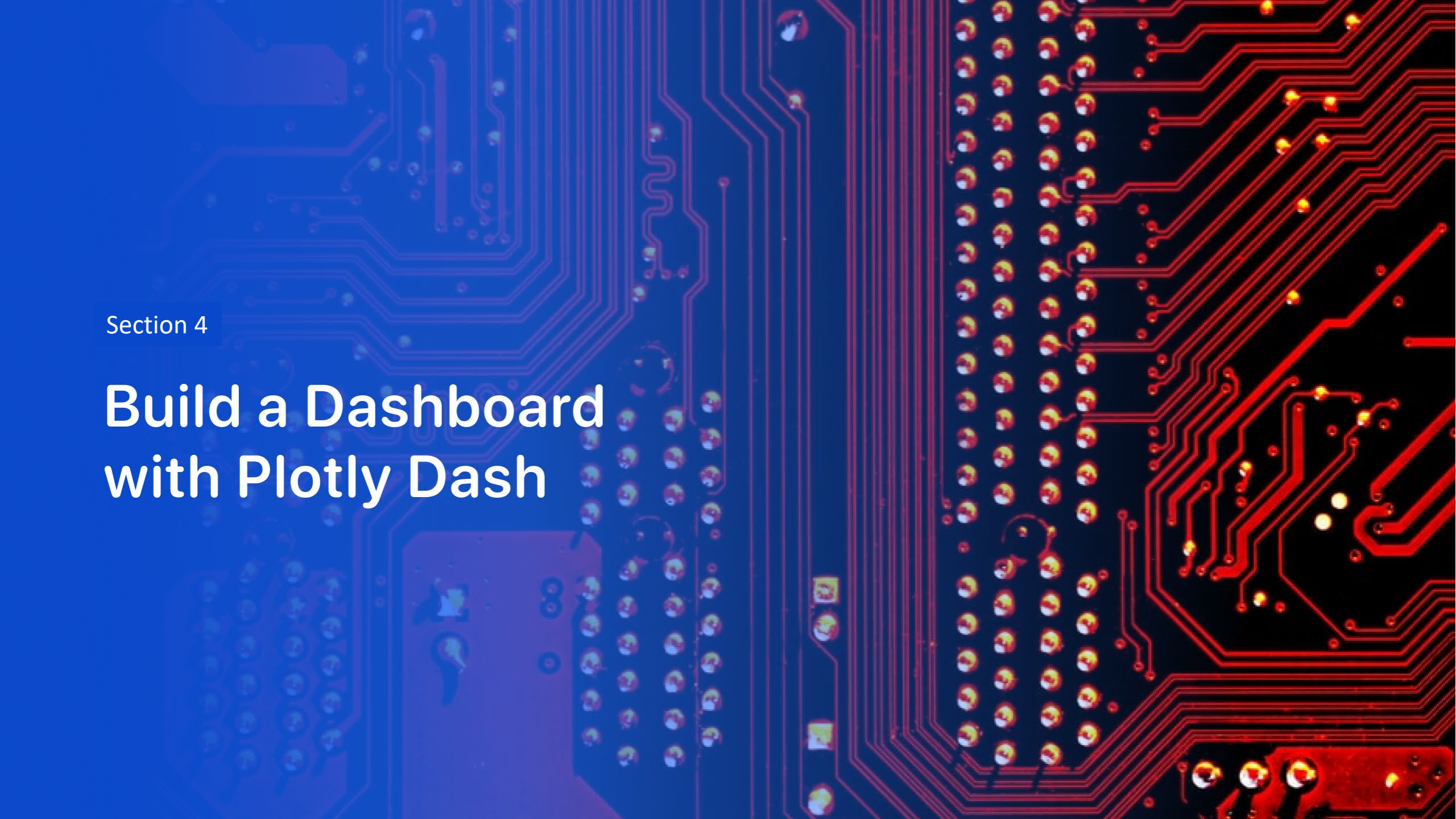


# Attempted Launches at KSC

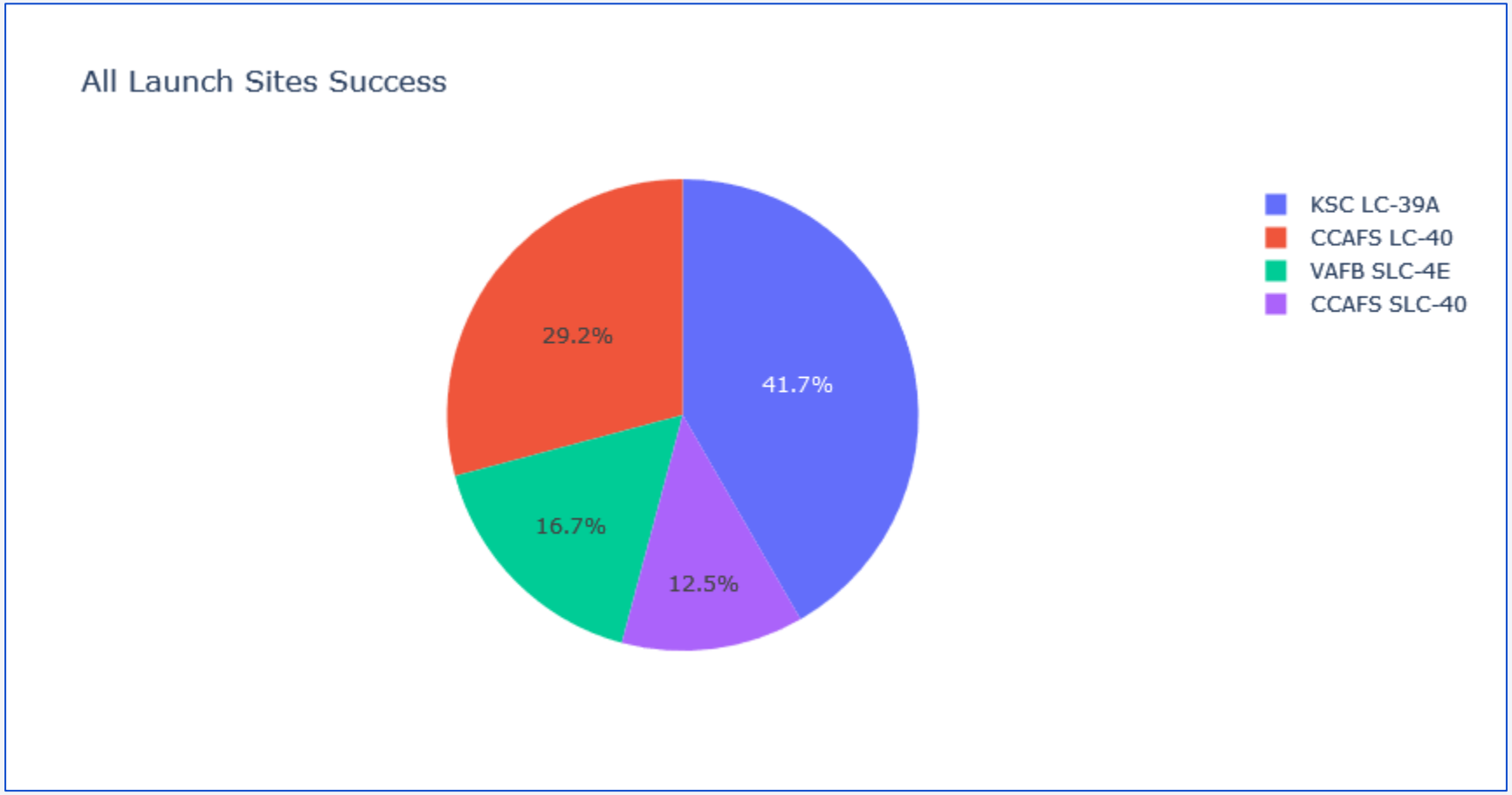


# Key Points of Interest near VAFB

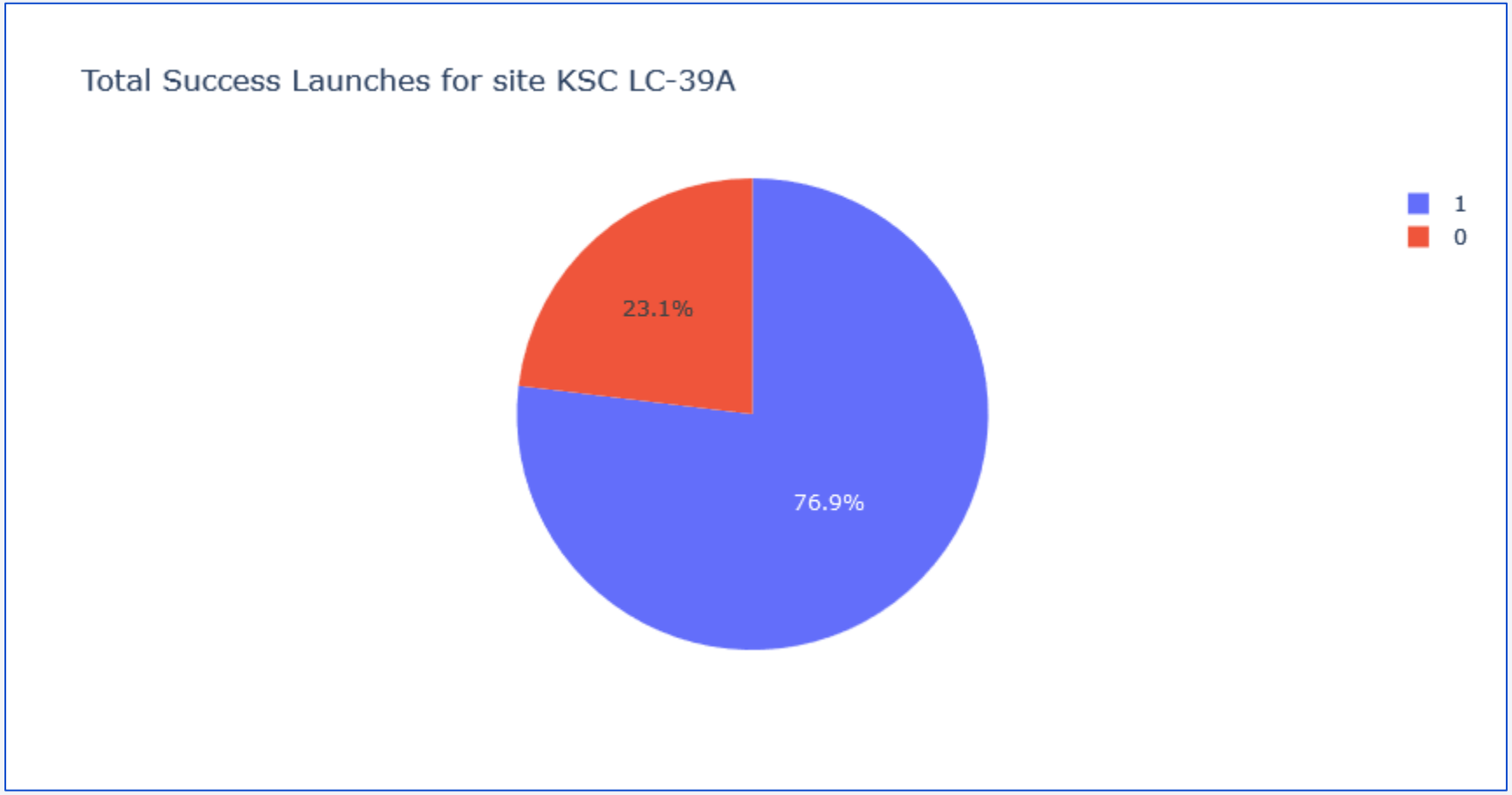



Section 4
Build a Dashboard with Plotly Dash

# Success Rate Share per Site

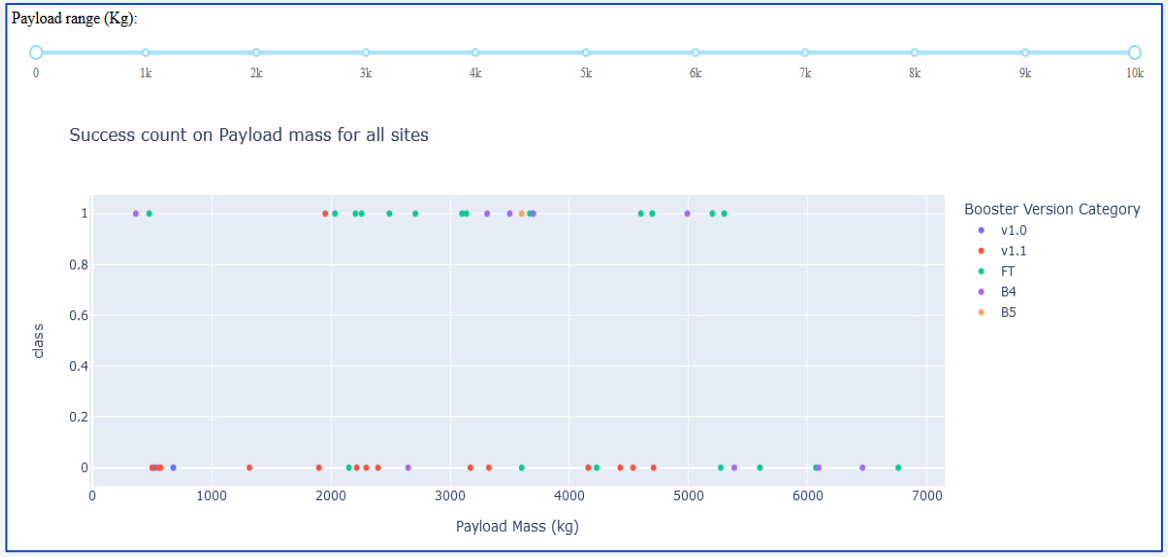


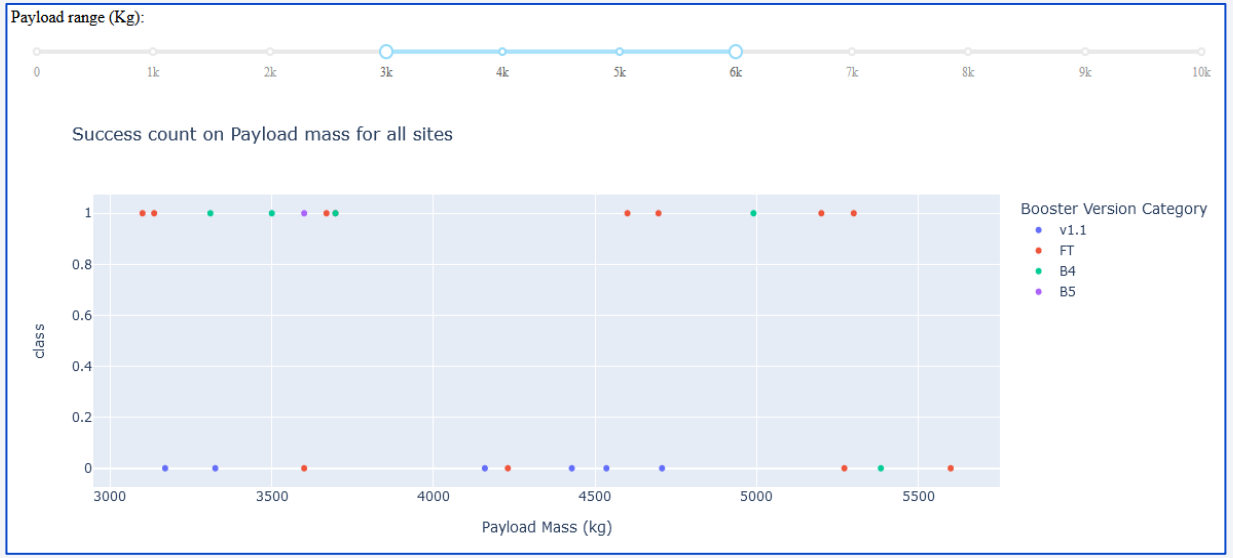
# Highest Success Percentage



# Successes in Payload Mass Range

- All Sites

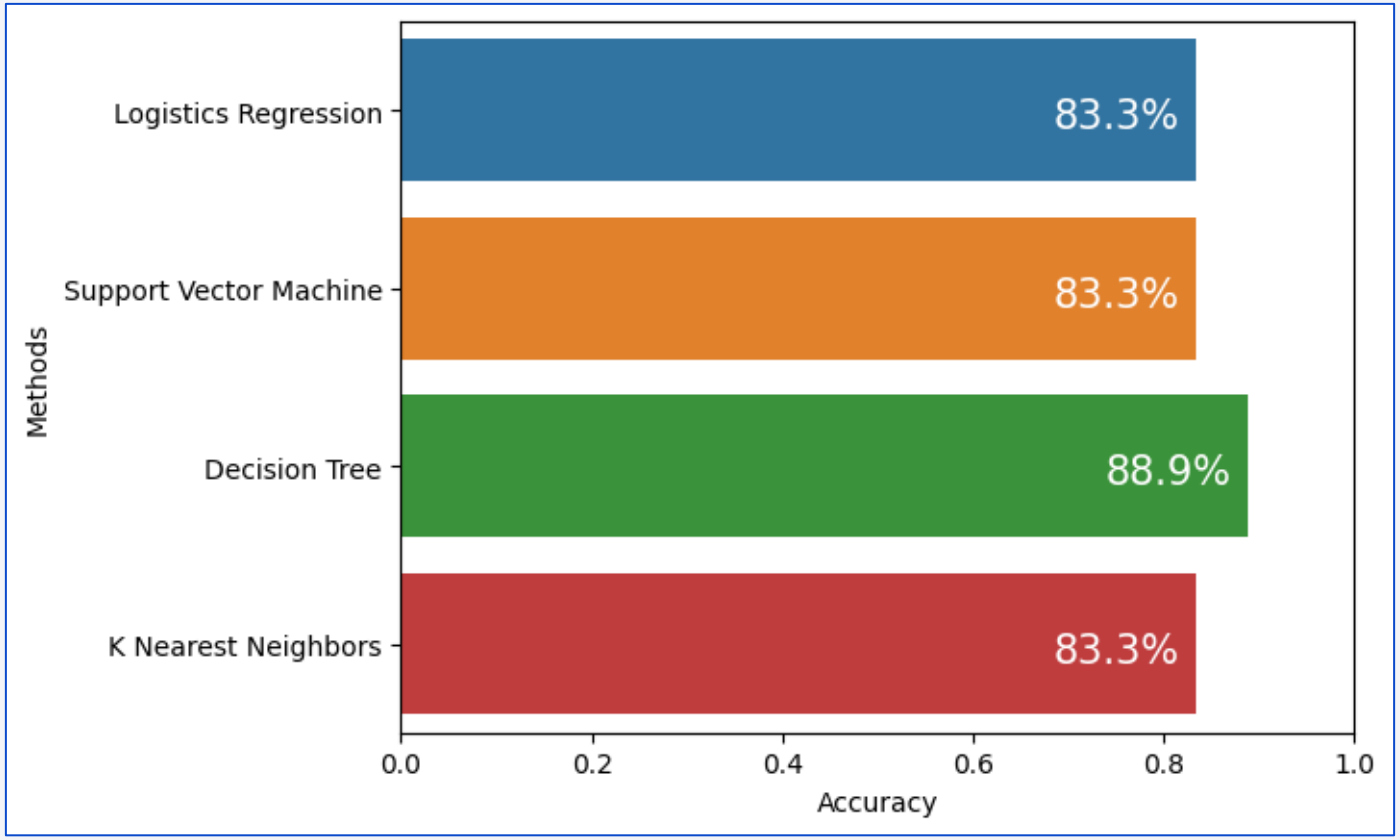


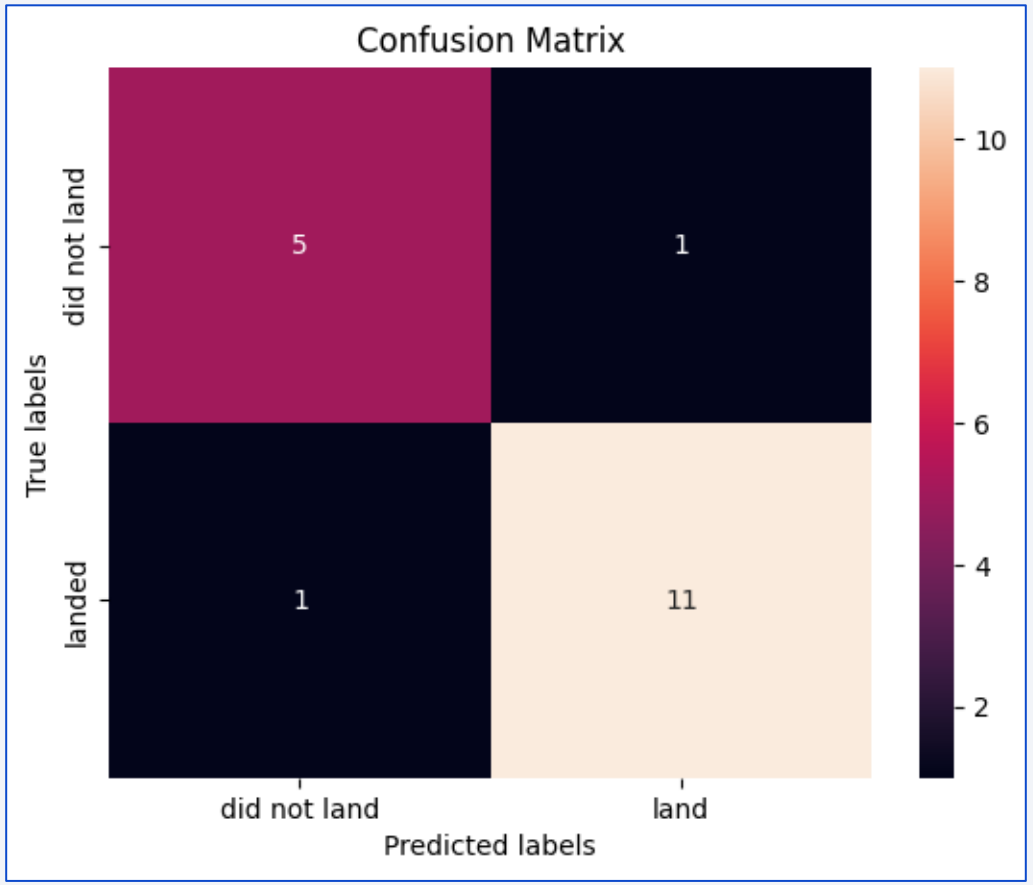
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy



# Confusion Matrix



# Conclusions

- Tree Decision Machine Learning
- Highest Success Rate at KSC
- Success Rate increasing consistently over time
- First Booster Successful Landing on 01-05-201
- Learning

# Thank you!