
# Winning Space Race with Data Science

Mª Kira Jaén Fernández
August 2022

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data collection using **SpaceX API** and web scraping with **Beautiful Soup**
  - Exploratory Data Analysis (**EDA**) in **Python**, including **data visualization**
  - EDA with **SQL**
  - **Interactive maps** with Folium
  - **Interactive dashboard** with Plotly Dash
  - Prediction using **Machine Learning models**
- Summary of all results
  - EDA has allowed to find out which variables are more likely to influence a positive outcome of each launch
  - ML models predict a successful recovery of the booster of **83.3%** of attempts

# Introduction

- Project background and context
  - Company SpaceY wants to compete with SpaceX, so a prediction of the best launch characteristics is needed in order to save costs by recovering the launch booster.

- Problems you want to find answers
  - What parameters allow for a successful launch and recovery of the booster?

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - SpaceX API
  - Web scraping with Beautiful Soup of the data in this Wikipedia page
- Perform data wrangling
  - SpaceX API
  - Web scraping with Beautiful Soup of the data in this Wikipedia page

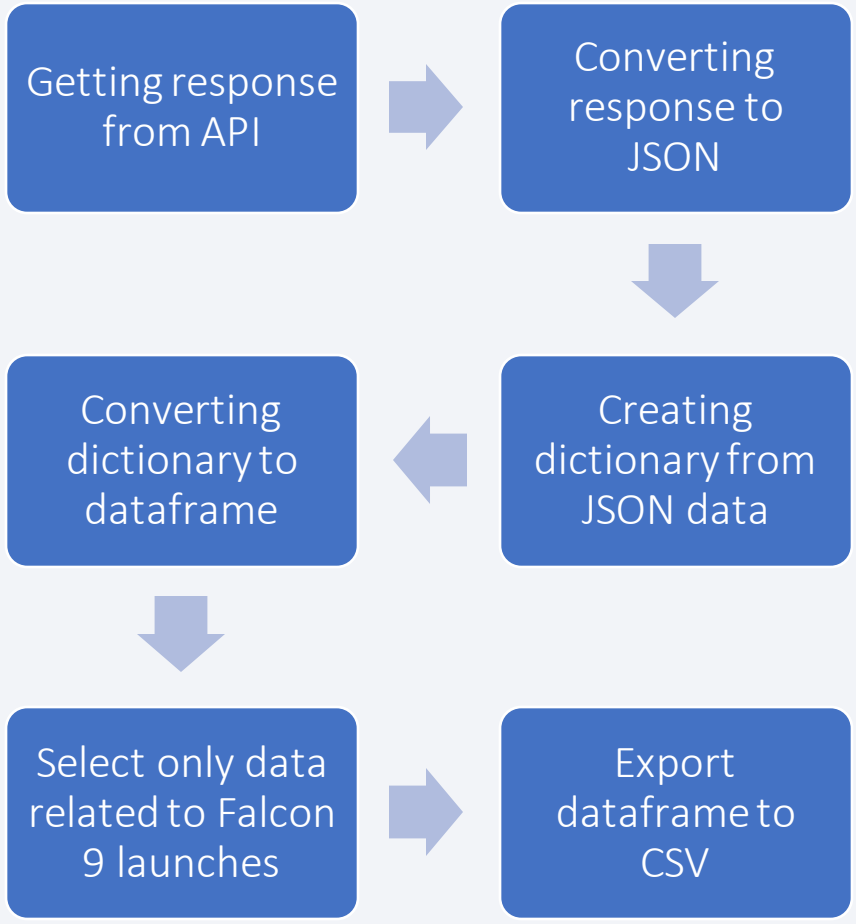
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Four different models were evaluated to check which one showed a better accuracy.

# Data Collection – SpaceX API

By using SpaceX API we have been able to obtain data regarding launches, booster and payload.

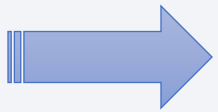
***Link to code:***

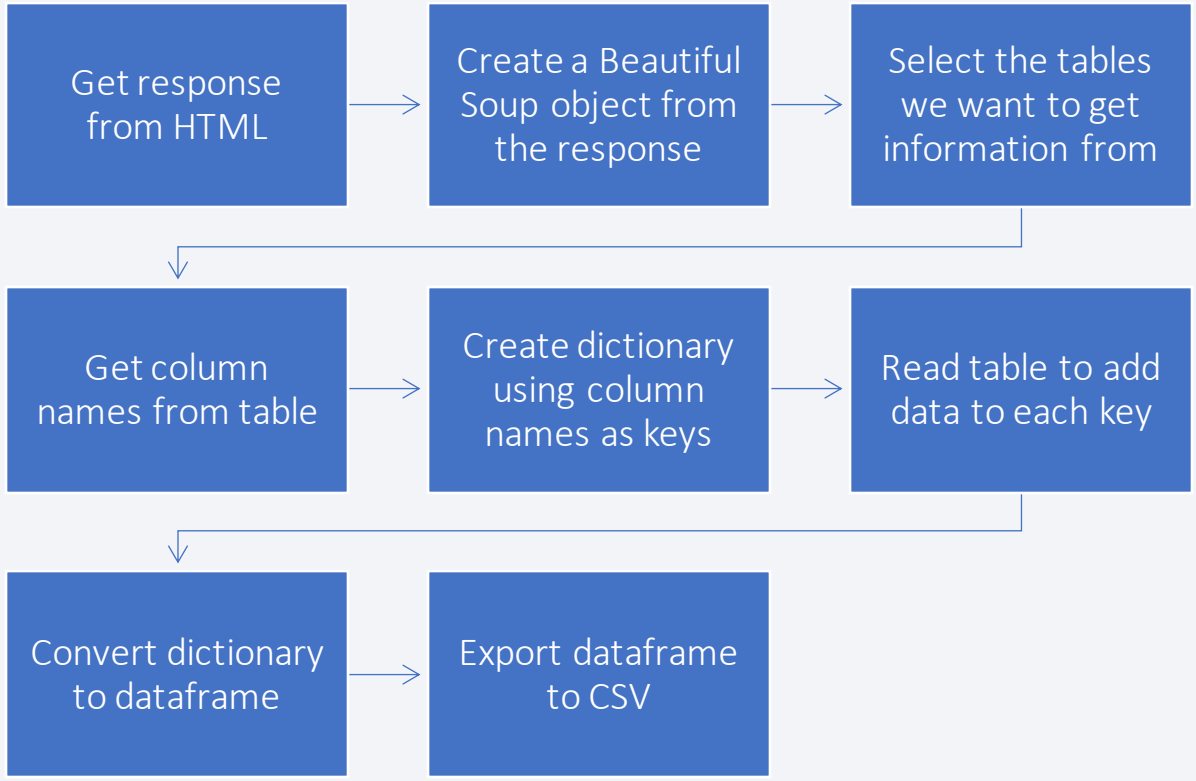
https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/Data%20Collection%20API.ipynb



# Data Collection - Scraping

Web scraping has allowed us to add more data to our collection






***Link to code:***

https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

# Data Wrangling

In order to determine which launches have been successful (i.e. booster was recovered) or not, we need to find out the outcome of each launch for each place and orbit.


***Link to code:***

https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/Data_Wrangling.ipynb

# EDA with Data Visualization

## Scatter Graphs

*Scatter plots show the correlation between numeric variables.*

- Flight Number x Payload Mass
- Flight Number x Launch Site
- Payload x Launch Site
- Orbit x Flight Number
- Payload x Orbit Type
- Orbit x Payload Mass

## Bar Graphs

*Bar graphs show the relationship between numeric and categorical variables.*

- Success Rate x Orbit

## Line Graphs

*Line graphs show the trend for numerical variables.*

- Success Rate x Launch Year

***Link to code:***

https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/EDA_with_Data_Visualization.ipynb

# EDA with SQL

## SQL Queries performed:

***Link to code:***

https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/EDA%20with%20SQL.ipynb

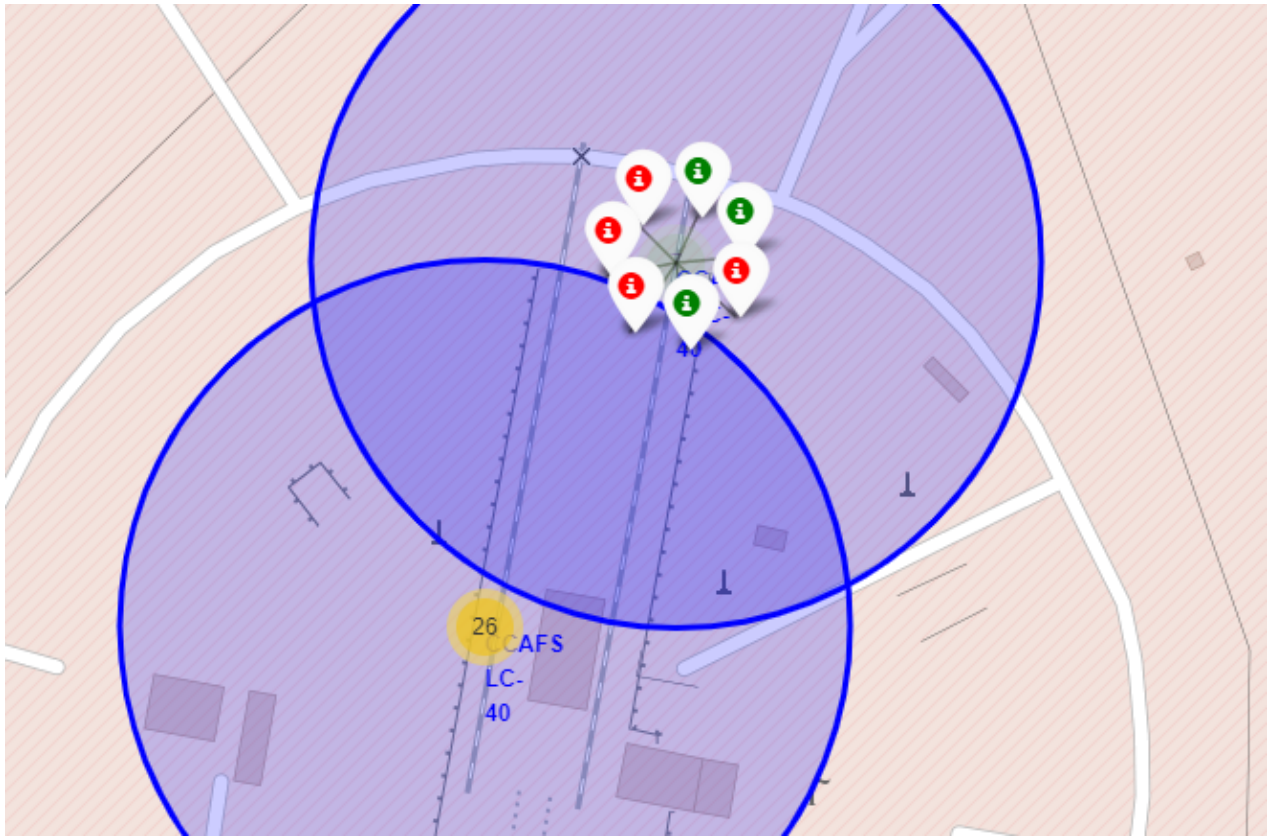
- Names of unique launch sites
- Five records where the launch site begin with "CCA"
- Total payload mass launched by NASA (CRS)
- Average payload mass carried by booster Falcon 9 v1.1
- Date of first successful landing in ground pad
- Boosters successfully landed on drone ship whose payload mass is between 4000 and 6000 kgs
- Total number of successful and failed missions
- Booster versions which have carried the maximum payload mass
- Failed missions landing in drone ship, booster versions and launch sites in 2015
- Ranking the count of successful landing_outcomes between the dates 04-06-2010 and 20-03-2017 in descending order

# Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

***Link to code:***

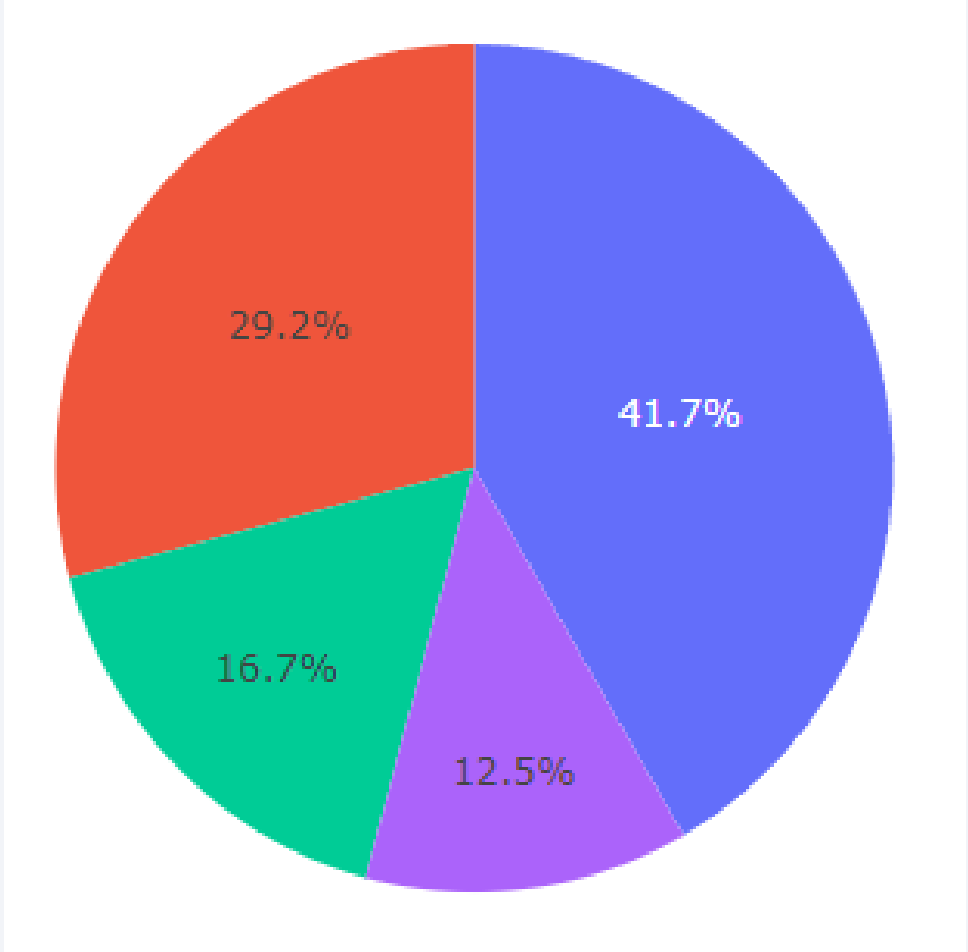
https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/Launch_site_location.ipynb



# Build a Dashboard with Plotly Dash

Dashboard includes

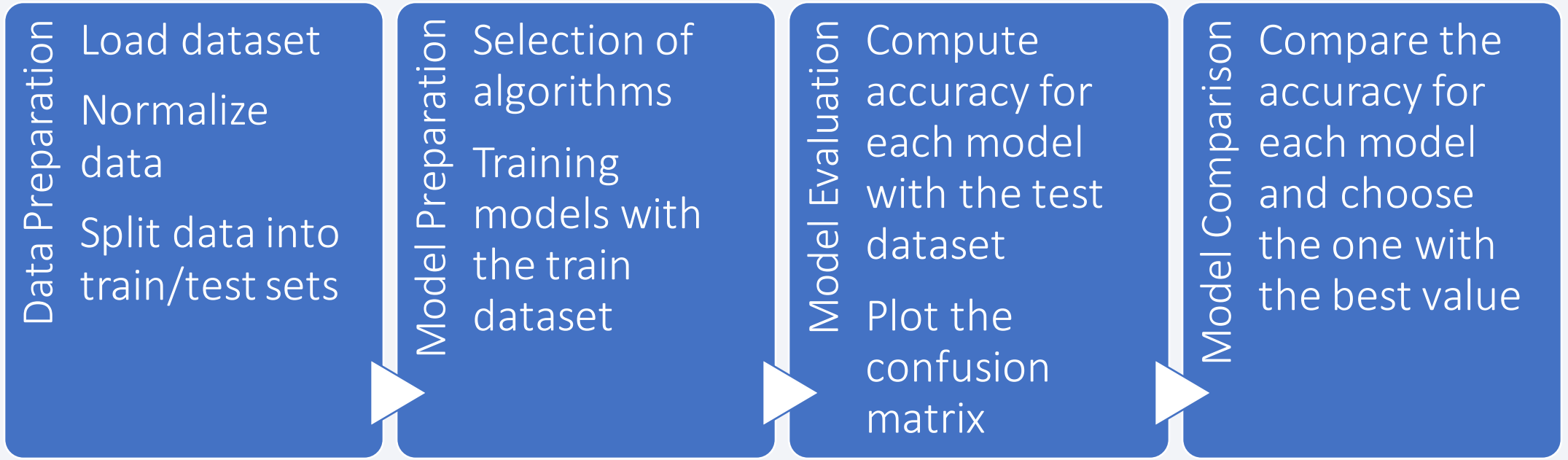
- Dropdown to choose between the different launch sites (of all of them)
- Pie chart showing the success/failure rates for each site, or percentage of success for all of the launch sites
- Rangeslider to select a payload mass
- Scatter chart showing relationship between Successful Launch and Payload Mass




***Link to code:***

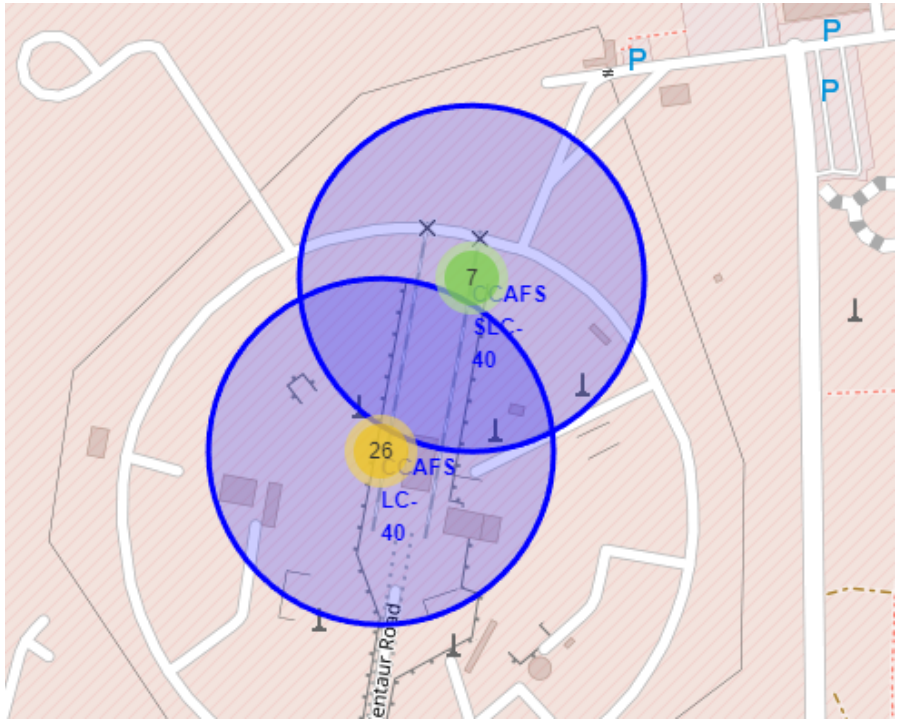
https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/Spacex%20Dash%20App.ipynb

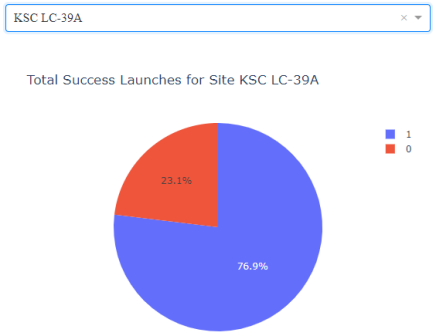
# Predictive Analysis (Classification)

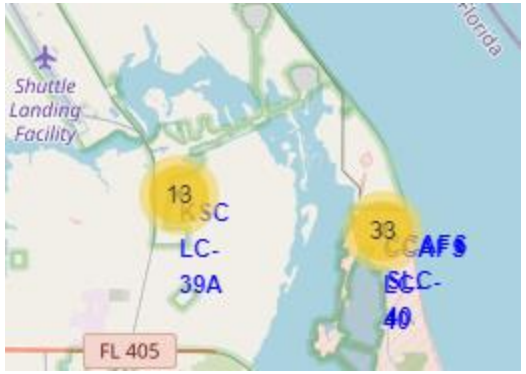


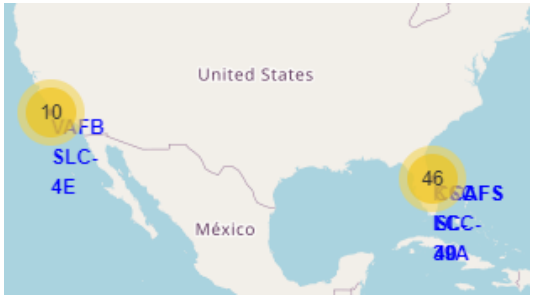

***Link to code:***

https://github.com/Fdez99/Applied-Data-Science-Capstone/blob/master/ML%20Prediction%20lab.ipynb



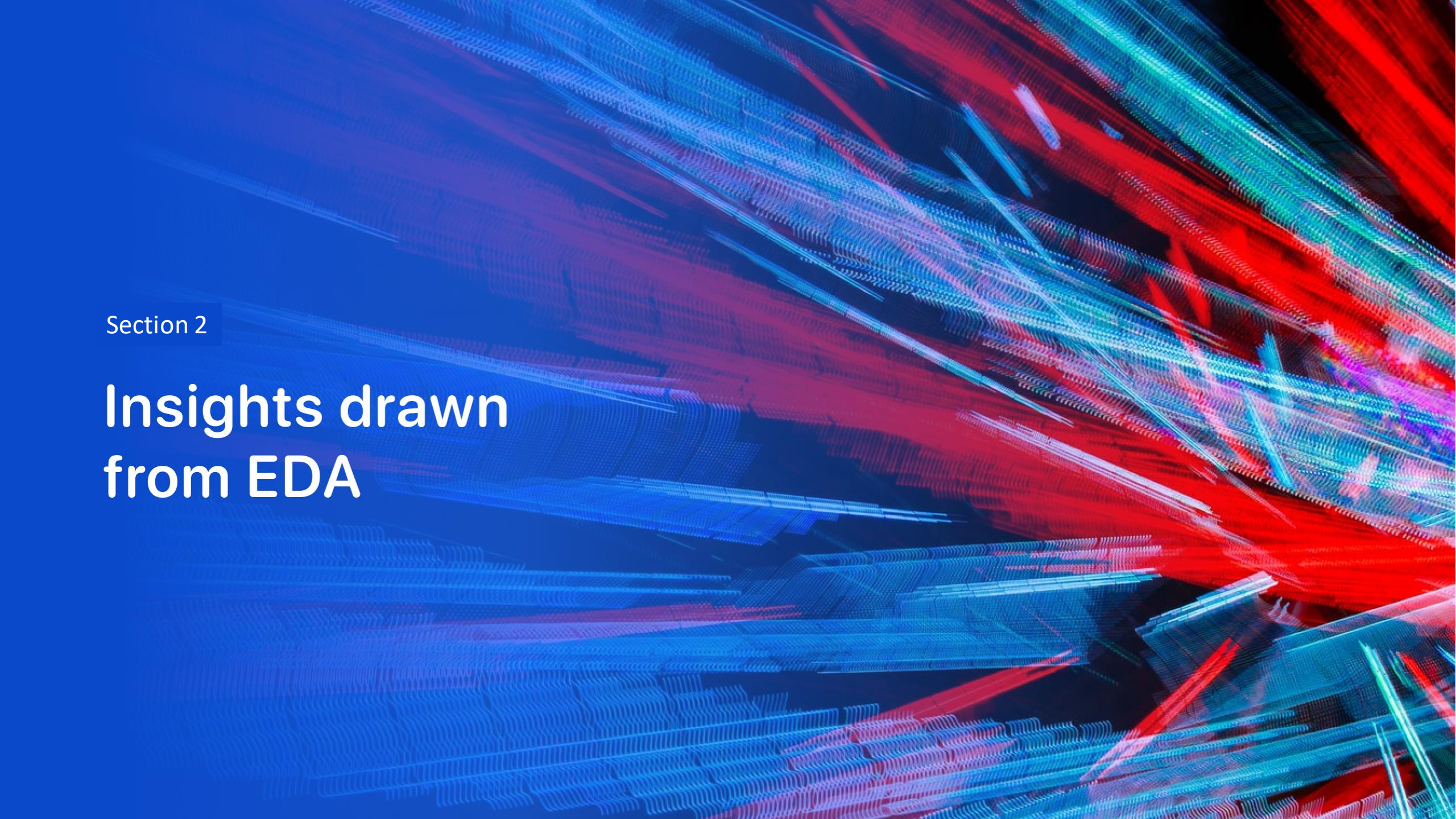

```
LogReg:  0.8333333333333334
SVM:  0.8333333333333334
Decission Tree:  0.9444444444444444
KNN:  0.8333333333333334
```

# Results

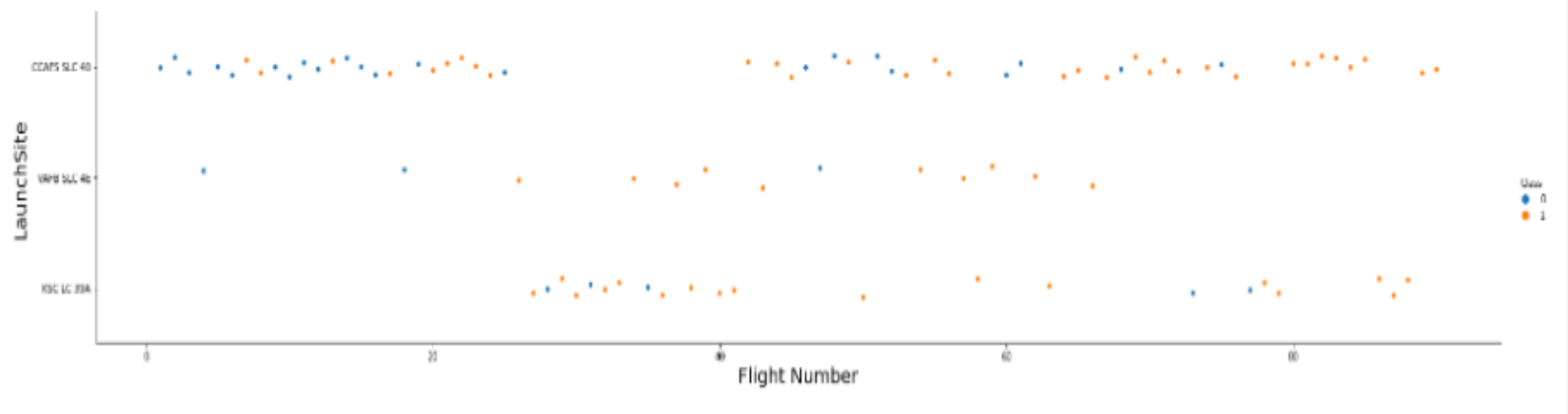
- SpaceX uses 4 different launch sites: CCAFS LC-40 VAFB SLC-4E KSC LC-39A CCAFS SLC-40
- Several orbits, the 3 most common GTO (geosynchronous orbit), ISS (International Space Station in low Earth orbit) and LEO (Low Earth Orbit)
- Average payload 2928.4 kgs
- First time a landing success was achieved: 01/05/2017
- 66.67% of success recovering the booster
- Most launch sites are near the sea and not far from railways and highways
- 12 different Falcon 9 booster versions
- Decision Tree is the best predictive model

Section 2

# Insights drawn from EDA

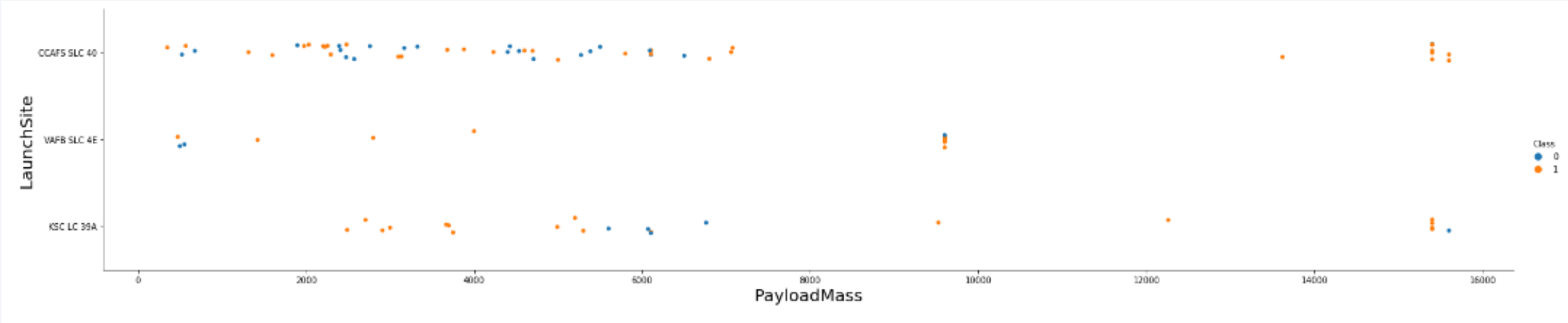


# Flight Number vs. Launch Site



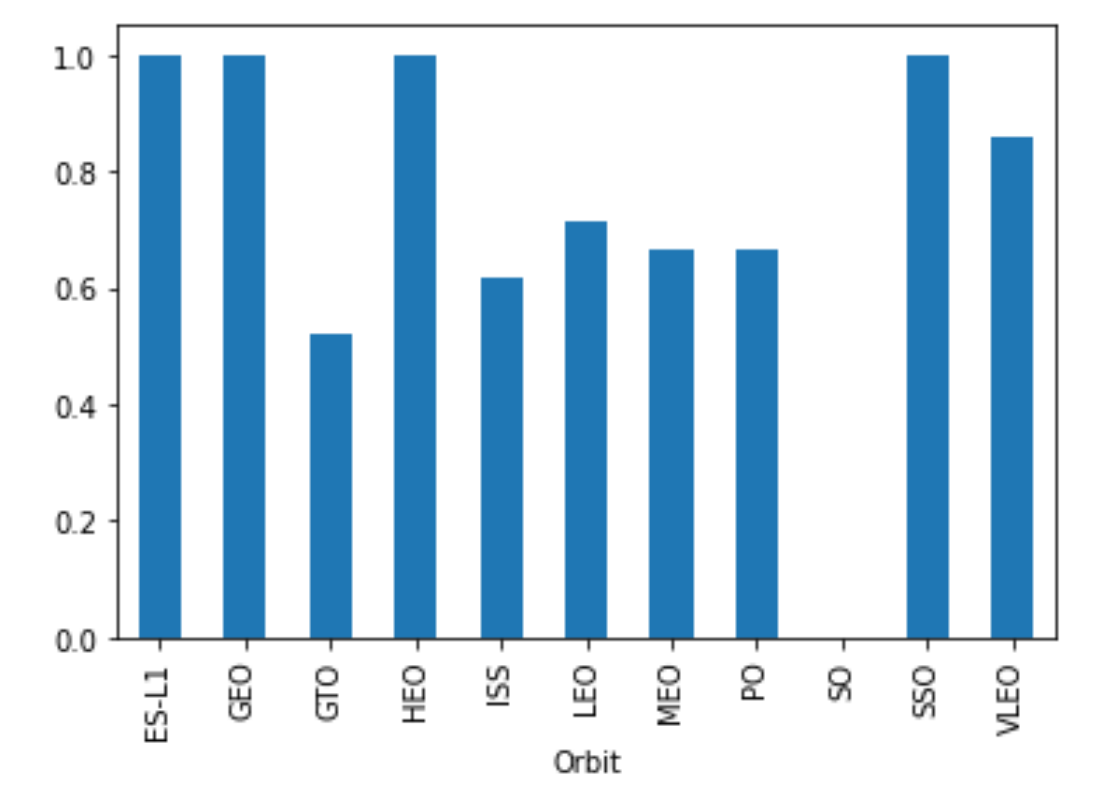

- Best launch site seems to be CCAF5, followed by VAFB SLC 4E and KSC LC 39A.
- Success rate improves over time

# Payload vs. Launch Site




- Payloads above 8000 kgs have a much higher success rate
- Heavier payloads (> 12000 kgs) can be only launched on CCAFS SLC 40 and KSC LC 39A

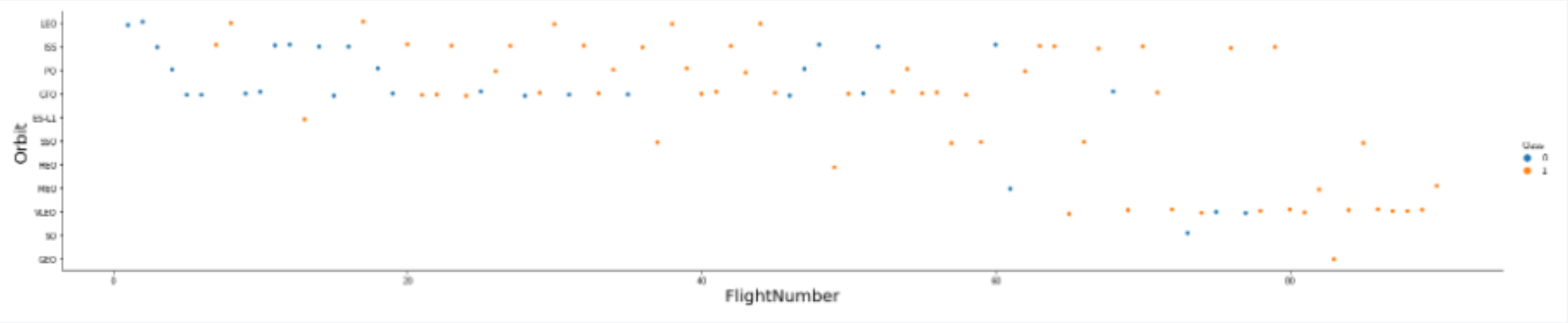
# Success Rate vs. Orbit Type




Four most succesful rates for orbits

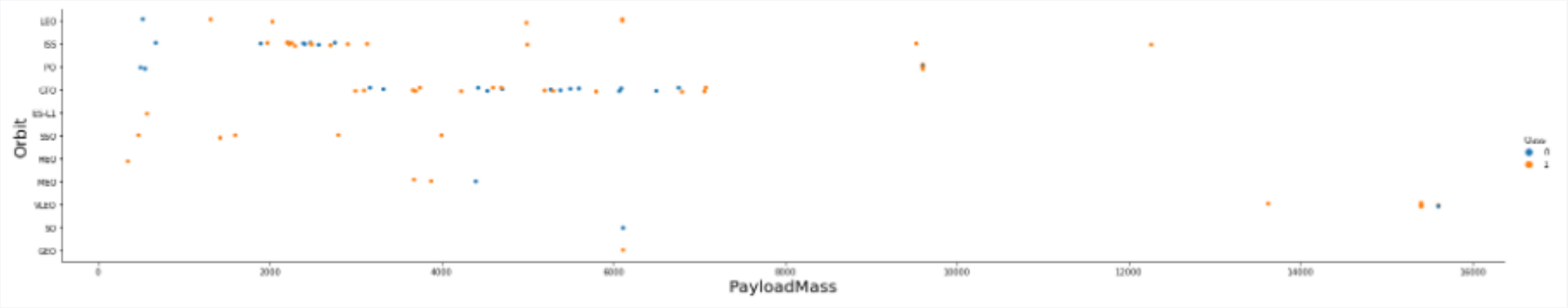
- ES-L1
- GEO
- HEO
- SSO

# Flight Number vs. Orbit Type



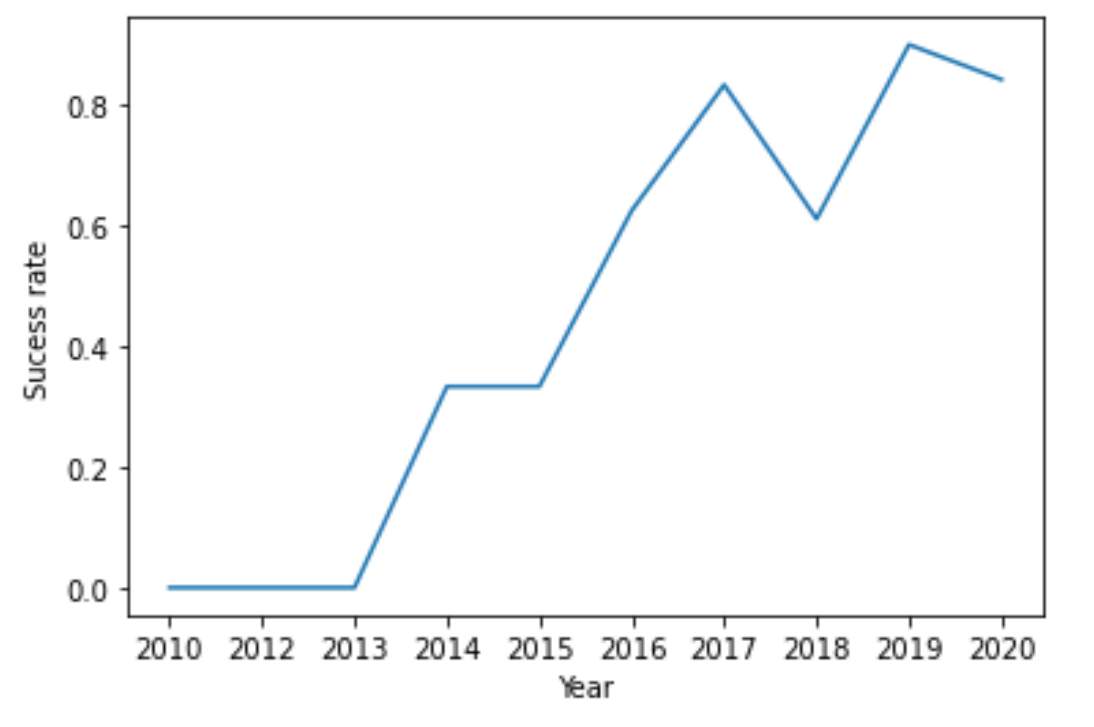

- Improvement over time seems common for all orbits
- More frequent orbits in later successful launches are ISS and VLEO

# Payload vs. Orbit Type




- Most successful orbits for heavier payloads are Polar, LEO and ISS
- Lighter payloads are mosts successful in ES-L1, SSO and MEO orbits

# Launch Success Yearly Trend




- Years from 2010 to 2013 were mostly unsuccessful
- Rate of success increases from 2013 onwards, with a decline in 2018

# All Launch Site Names

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

- There are four launch sites, shown on the table on the left
- Obtained by a SQL query, selecting only unique Launch Site values from the launch table

# Launch Site Names Begin with 'CCA'

| 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 |

The five Cape Canaveral launching sites

# Total Payload Mass

- Total payload carried by NASA boosters

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

- Obtained by adding up all payloads from the same client (NASA)

# Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

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

- After filtering the data to match only with version 1.1 of Falcon 9, the average payload mass obtained is the one shown above.

# First Successful Ground Landing Date

| min("DATE") |
| --- |
| 01-05-2017 |

- Filtering the results for landing place (Ground Pad) and querying for the smallest value for Date, we obtained this value.

# Successful Drone Ship Landing with Payload between 4000 and 6000

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

- Filtering data so that successful landing on drone ships that carried a payload between 4000 and 6000 kgs, the booster version obtained are those on the left.

# Total Number of Successful and Failure Mission Outcomes

- Filtering by mission outcome and counting the results for each, we obtained the following:

| Mission_Outcome | No_missions |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

| 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 |

- Names of the booster which have carried the maximum payload mass

# 2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

| Month | Booster_Version | Launch_Site |
|---|---|---|
| 01 | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | F9 v1.1 B1015 | CCAFS LC-40 |

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

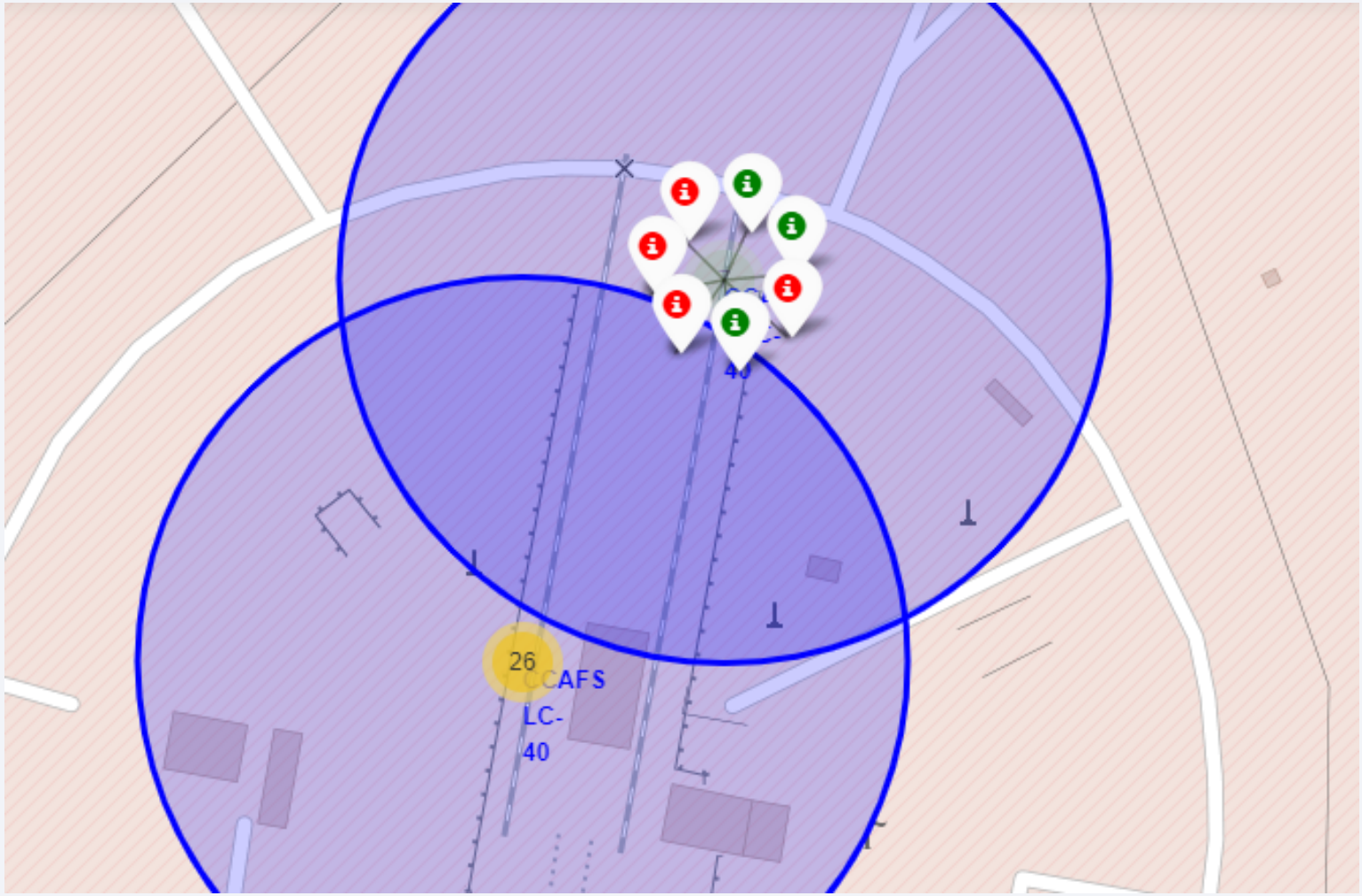
| Landing _Outcome | COUNT("LANDING _OUTCOME") |
|---|---|
| Success | 20 |
| Success (drone ship) | 8 |
| Success (ground pad) | 6 |


Section 3
Launch Sites
Proximities Analysis

# All the launch sites


- All launch sites are near the sea.

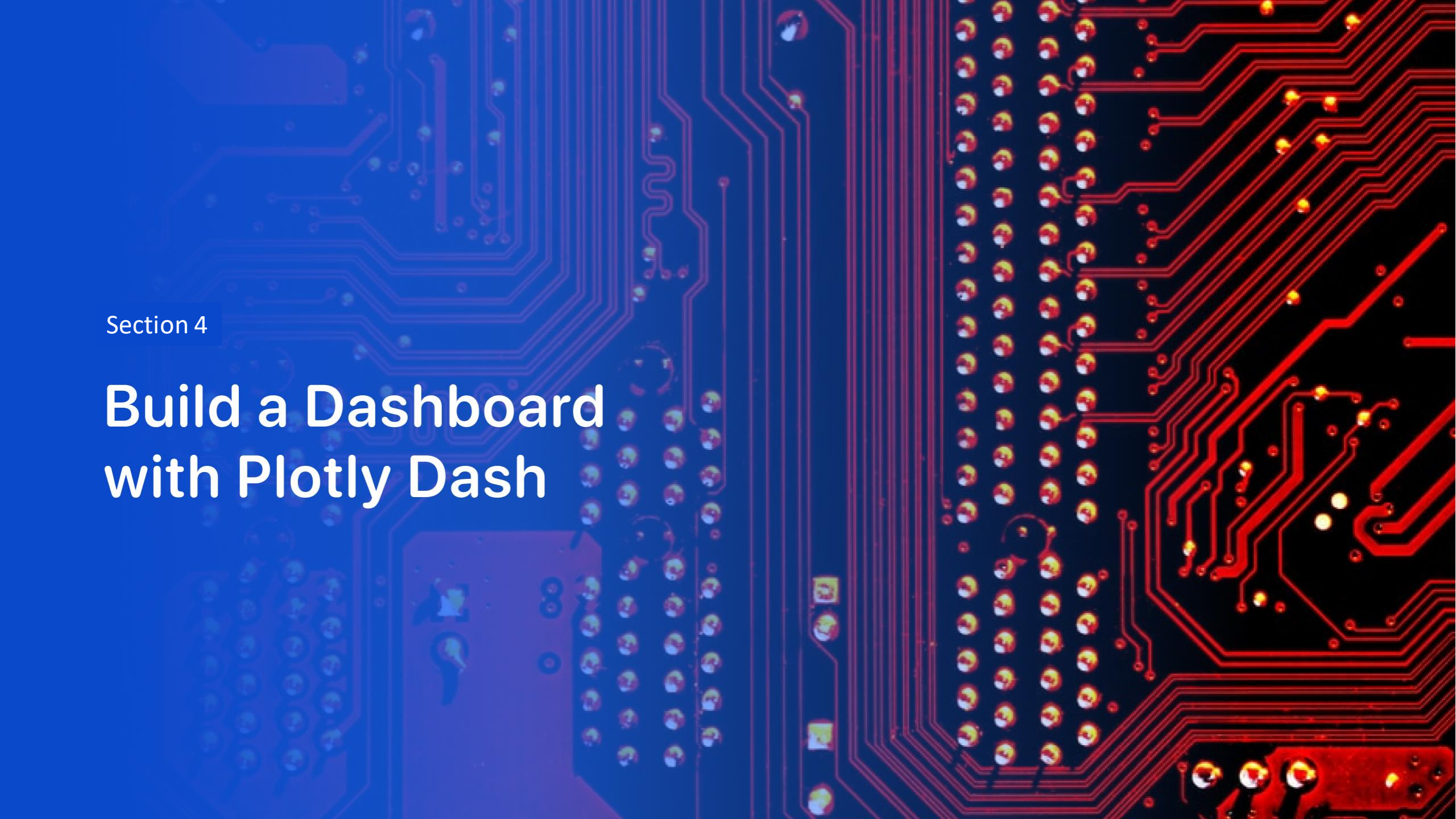
# Launch site outcomes



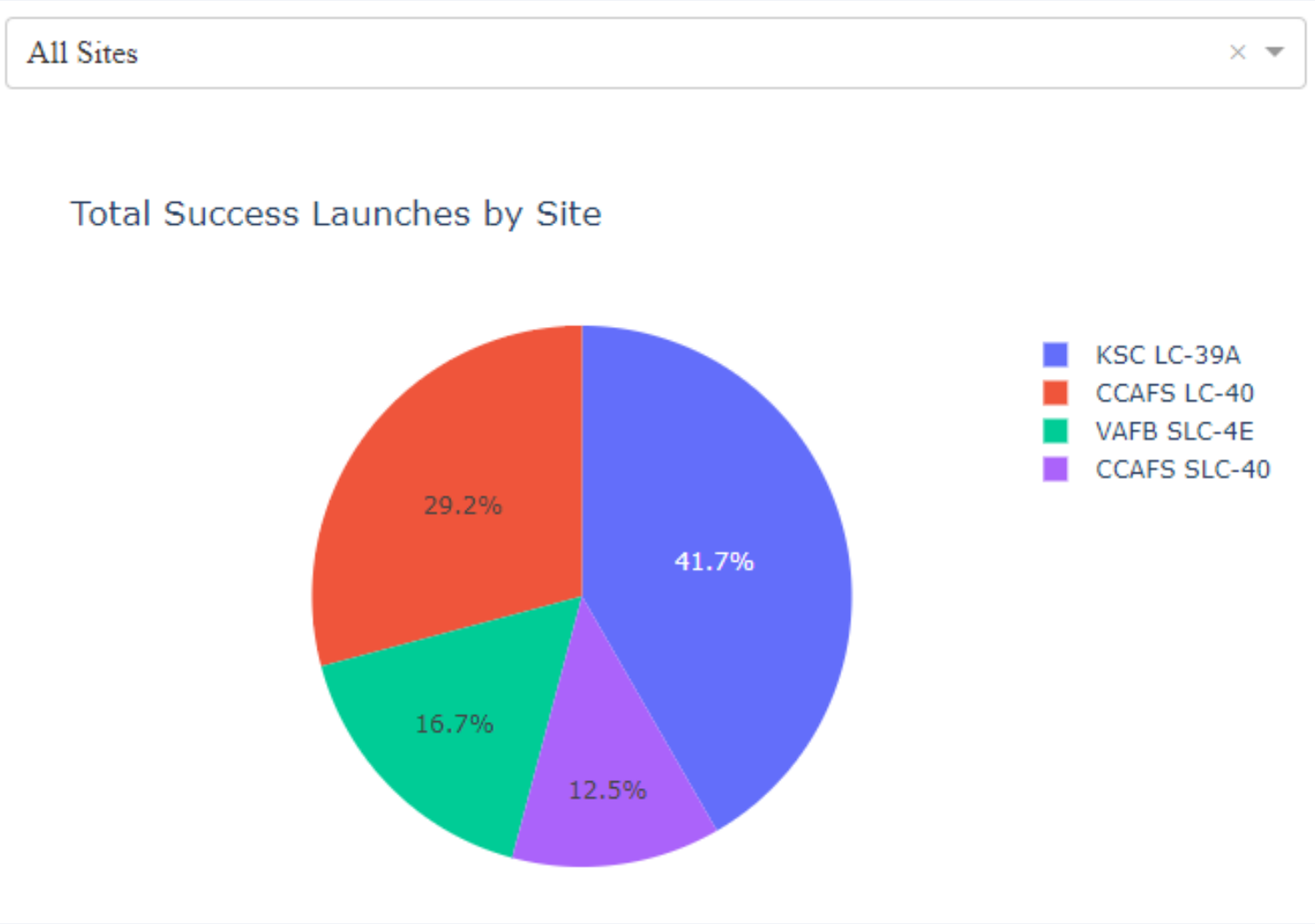

- Numbers indicate how many launches have been made on each site.
- Green markers mean the launch was successful; red, failed.

# Placement


Site KSC LC-39A close to the sea and a road, and at the same time far from cities.

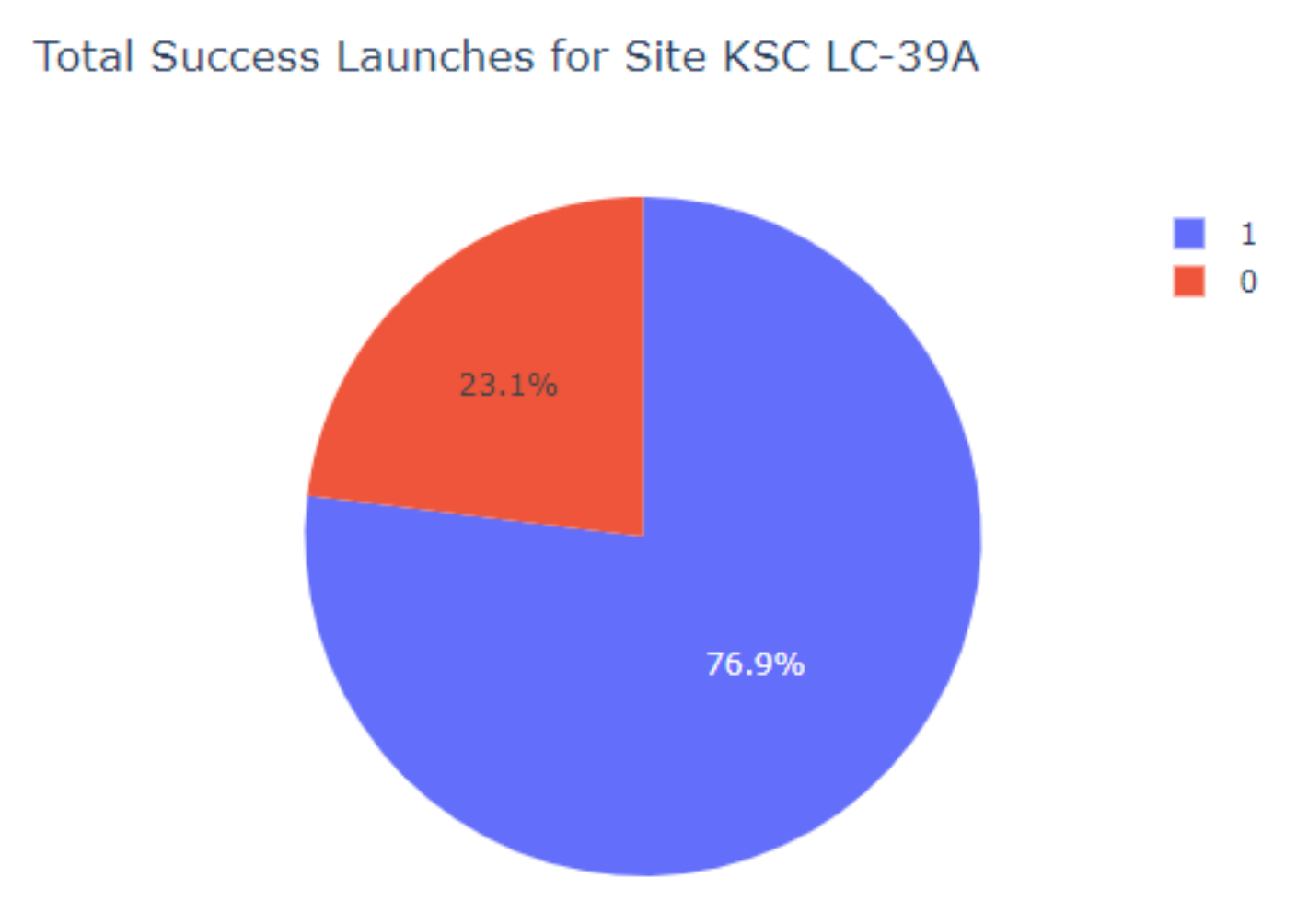

Section 4
Build a Dashboard with Plotly Dash

# Successful launches by site




**KSC LC-39A** seems to be the launch site with the highest success rate.

# Success rate for KSC LC-39A




KSC LC-39A has a 76.9% of successful launches.

# Payload x Launch outcomes

FT boosters are most successful for payloads under 4000 kgs.

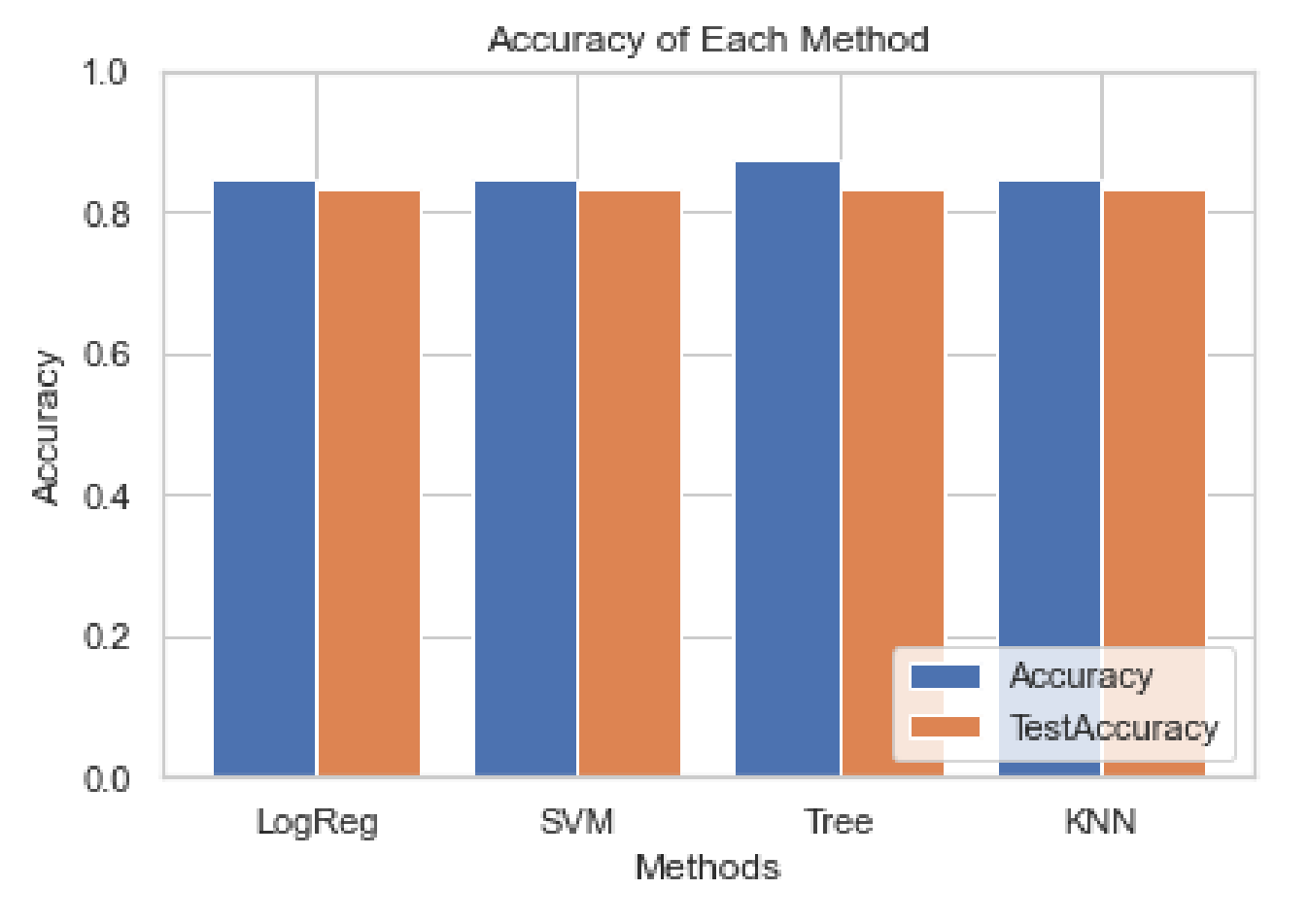


B4 boosters are most successful with heavier payloads (although data seem to be insufficient).

Section 5

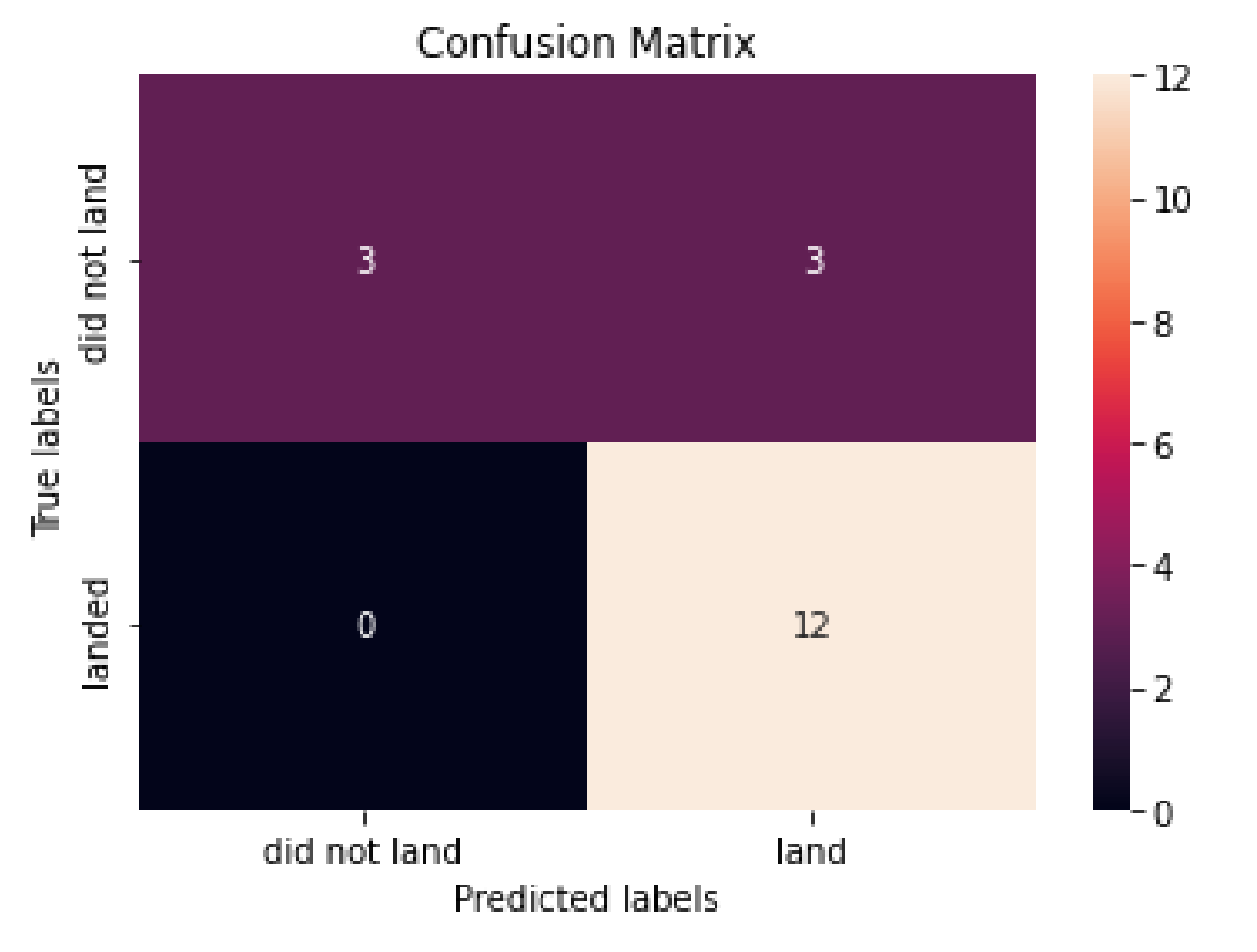
# Predictive Analysis (Classification)

# Classification Accuracy




Of the four models considered, the Decission Tree is the model which showed a higher accuracy.

# Confusion Matrix




In this model, we can see there are no false negatives (predicting that the booster would not land, but it landed), and that it correctlly predicted the landing of most successful missions.

# Conclusions

- The success of a launch seems to depend on the launch site, the orbit and the payload mass.
- Launch site with most success rate is KSC LC-39A.
- Orbits with most successful launches are GEO, HEO, SSO and ES-L1.
- Lower payloads lead to a greater success rate of missions.
- The best algorithm to predict outcomes for this dataset seems to be the Decision Tree model.

# Appendix

- GitHub doesn't show the maps, so screenshots from the actual notebook were taken.

# Thank you!