IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Daniel
03/01/2025

# Outline

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

# Executive Summary

- Summary of methodologies
- Summary of all results

# Introduction

- Project background and context
- Problems you want to find answers

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - The data was collected using the diverse public SpaceX APIs
- Perform data wrangling
  - The outcomes of the landings were converted to 0 and 1 for easier analysis and predictions.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The models were trained on different parameters, and their metrics were reviewed to find the best predictive model which can be built.

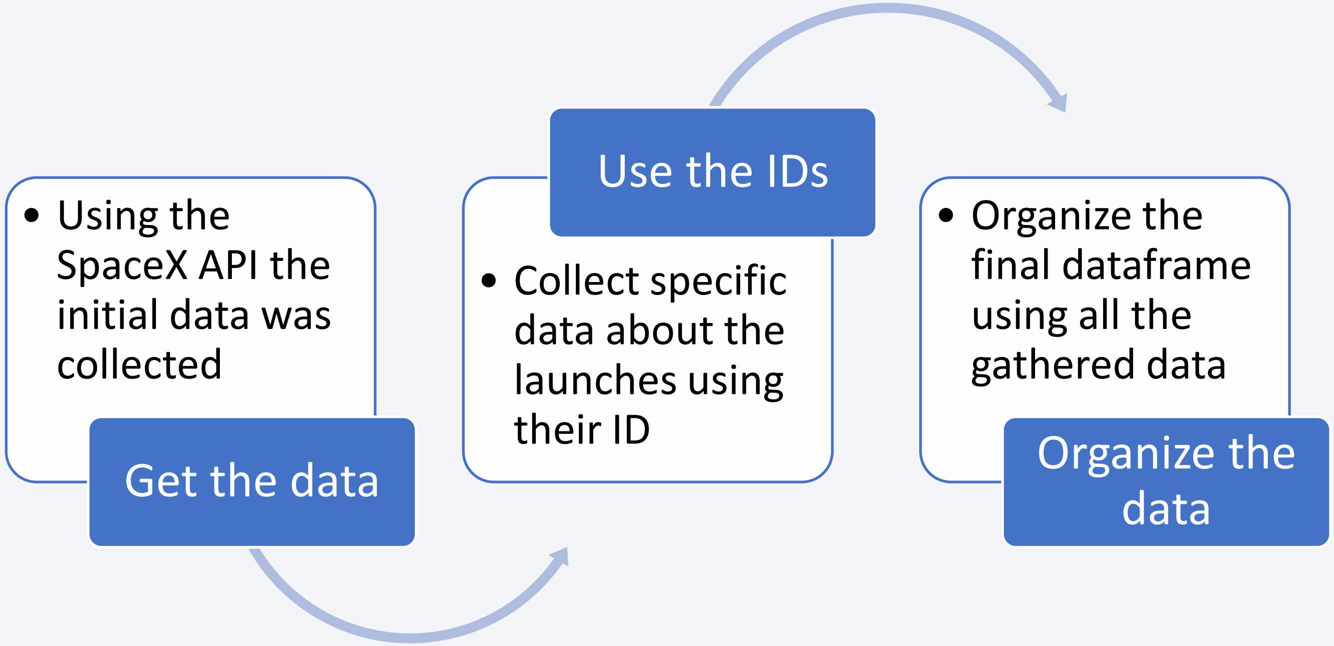
# Data Collection

- Describe how data sets were collected.
  - The data was collected using the SpaceX public APIs and the Wikipedia webpage.
  - The data was organized and cleaned after its collection to make sure the prediction model is going to perform according to the real-world findings.
- You need to present your data collection process use key phrases and flowcharts

# Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

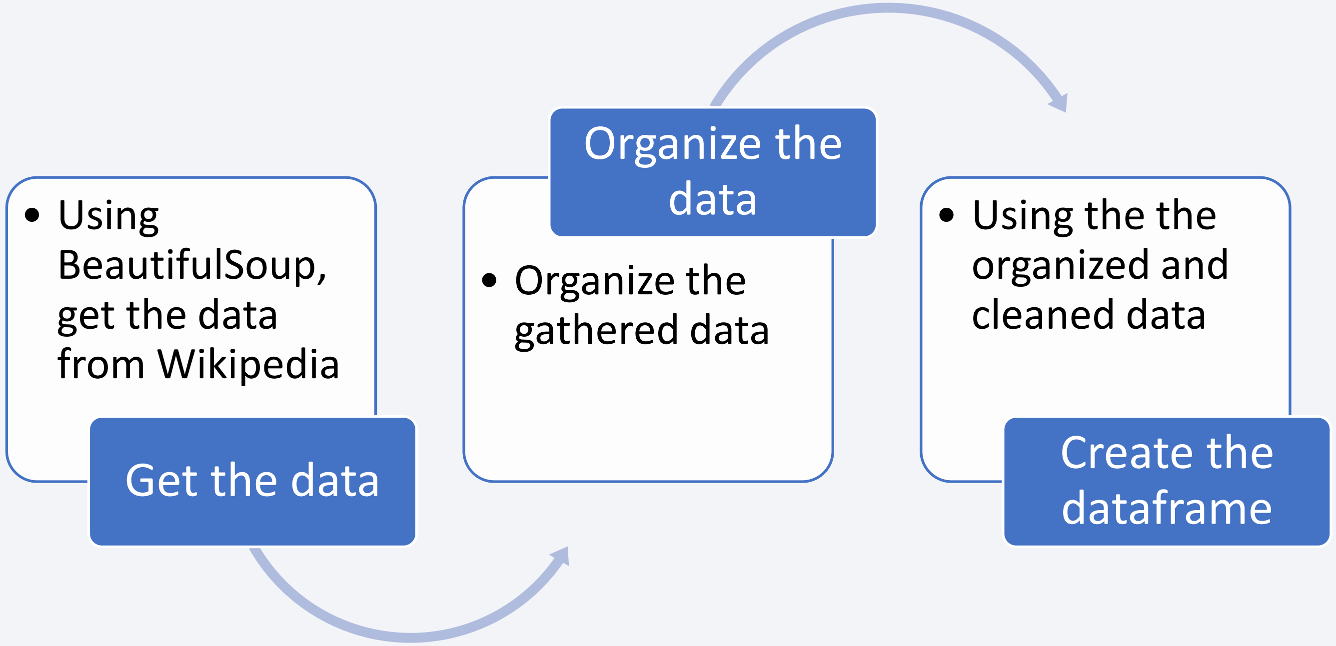
- GitHub URL: [Data Collection - SpaceX API (GitHub)]()



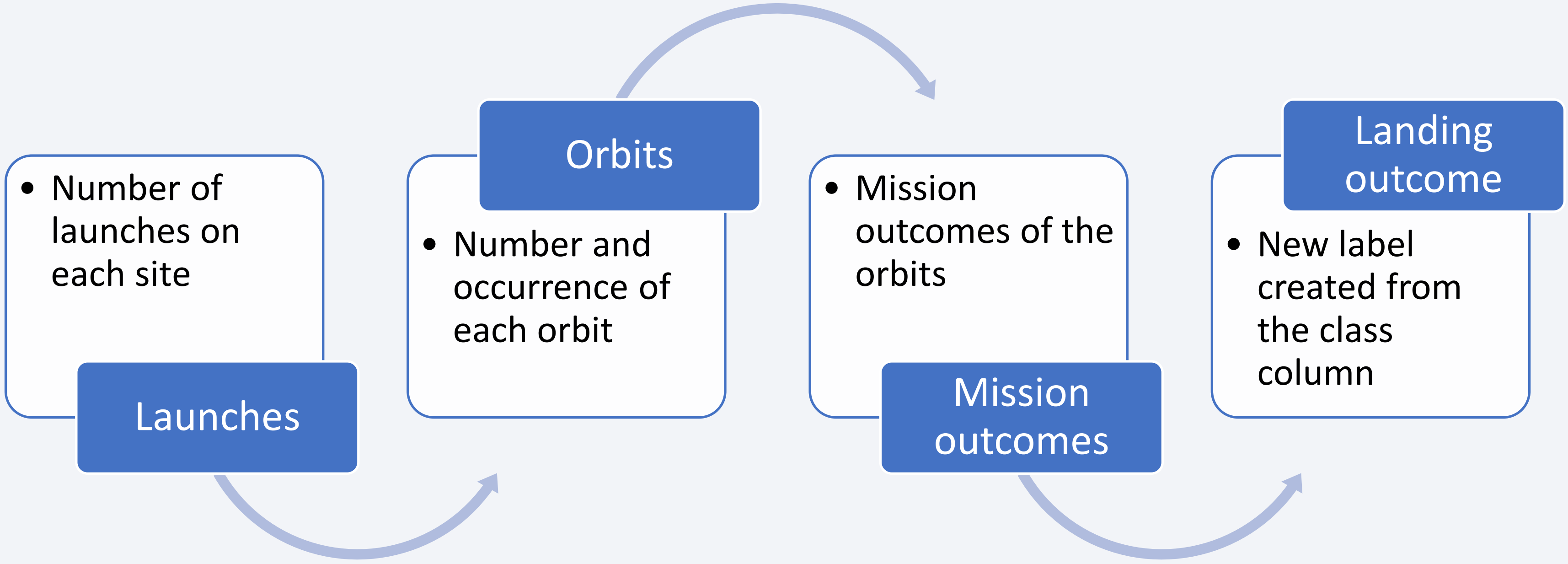
# Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL: [Data Collection - Scraping (GitHub)]()



# Data Wrangling




- GitHub URL: [Data Wrangling (GitHub)]

# EDA with Data Visualization

- Relationship between Flight number and launch site:
    - See the possible relationship between those two variables using the landing outcome as hue to try finding if the newer flights have more or less success.
- Relationship between payload mass and launch site:
    - Scatterplot using the outcome of the landing as color. Look for possible relationship between the payload mass of the launch site and the success or fail of the landing
- Relationship between success rate of each type of orbit:
    - Barplot to look for the success rate of the landings by each type of orbit. Helps finding possible orbits where the landings are more successful.
- Relationship between flight number and orbit:
    - Scatterplot to look for relationship between flight number and orbit

# EDA with Data Visualization

- Relationship between Payload Mass and Orbit type:
  - Scatterplot to see the relationship between the payload mass and the orbit type
- Success rate over the years:
  - Line plot to see the success rate evolution over the years
- GitHub URL: EDA - Data Visualization (GitHub)

# EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Display the names of the unique launch sites
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display the avg. payload mass carried by booster version F9 v1.1
  - Get the date of the first successful landing in ground pad
  - List the count of successful and failed mission outcomes
  - Rank the count of landing outcomes between 2010-06-04 and 2017-03-20
- GitHub URL: [EDA - SQL (GitHub)]

# Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - NASA Space Center
  - Launch sites (allows to know where the launch sites are located)
  - Markers with the landings' outcomes (fast and clear way to see the outcomes of that launch site)
  - Line to the nearest coast (easy way to tell where the nearest coast is)
  - Line between two launch sites (allows to know the distance between two launch sites)
- GitHub URL: Interactive Map with Folium (GitHub)

# Build a Dashboard with Plotly Dash

- The dashboard includes a pie chart and a scatter plot.
- The pie chart shows the success launches for all the launch sites or the success rate for each launch site.
- The scatter plot shows the payload mass in kg and the landing outcome, colored by the booster version category.
- The dashboard allows an easy way to find success rate per launch site and the success and failed landings per booster version category and payload mass
- GitHub URL: Dashboard with plotly

# Predictive Analysis (Classification)

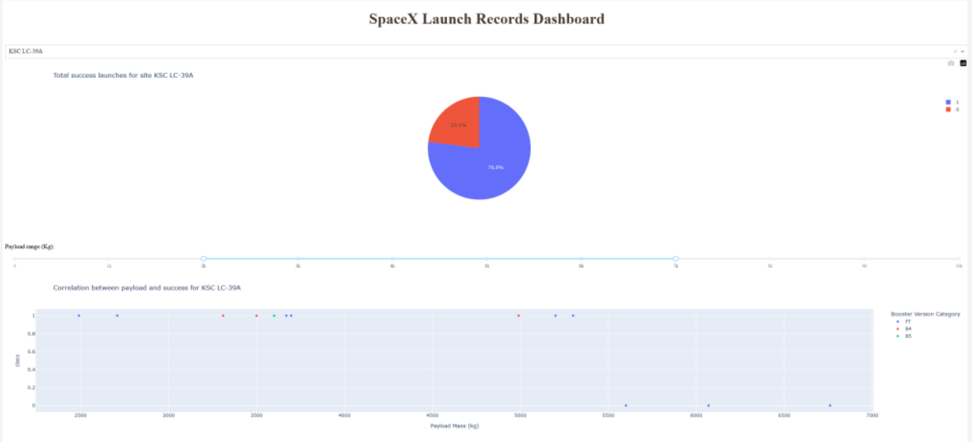
- Summarize how you built, evaluated, improved, and found the best performing classification model:
  - Using various parameters, each model was built and then the performance was tested on new unseen data.
  - The confusion matrix was used to see how it classified the unseen data.
- To find the best model, the next steps where done:
  - Train various classification models
  - Test the models and get the performance metrics to find how they perform on unseen data
  - Retrain the best model with all the data
- GitHub URL: Predictive Analysis (Classification) (GitHub)

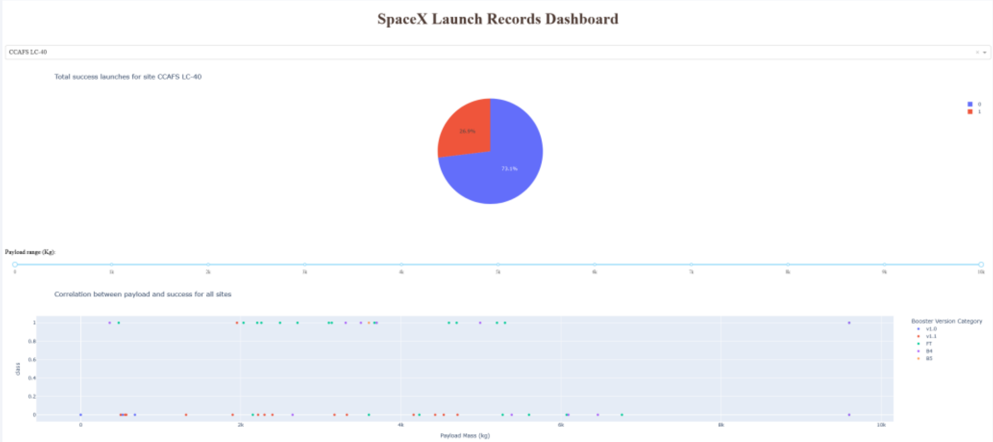
# Results

- Exploratory data analysis results
    - The max payload mass for VAFB-SLC launch site is less than 10000. The greater the mass, the greatest is the number of successful landings.
    - ES-L1, GEO, HEO and SSO are the orbits which have a 1.0 rate of successful landings, while GTO, ISS, LEO, MEO, PO and VLEO have mixed landing success. SO is the only orbit where all the landings have failed.
    - There is not a clearly relationship between the flight number and the orbits.
    - For LEO, ISS and PO orbits, we can see that the heavier the mass, the greatest is the success rate.
    - The success rate has clearly increased with the years, being very close to an average all success in the last years. The only problematic year is 2018, where the success rate is going down.

# Results

- Interactive analytics demo in screenshots



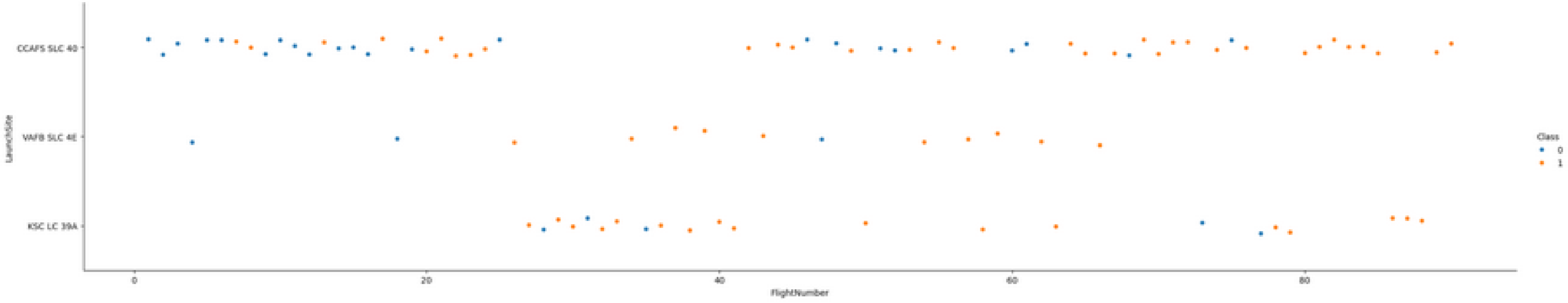

- Predictive analysis results
  - After testing various parameters in the models and testing them with unseen data, Logistic Regression, SVM and KNN perform the same, so we can use one of them with very good prediction rates.

Section 2

# Insights drawn from EDA

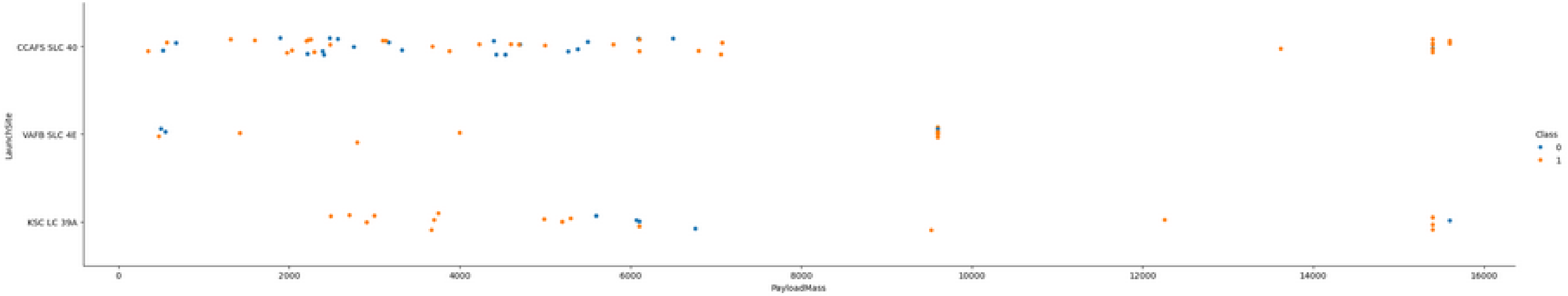
# Flight Number vs. Launch Site

- The last flights in CCAFS and KSL tend to have a more successful landing, while the first flights tend to fail their landing. In VAFB is not very clear that relation as almost all the flights have had successful landings.



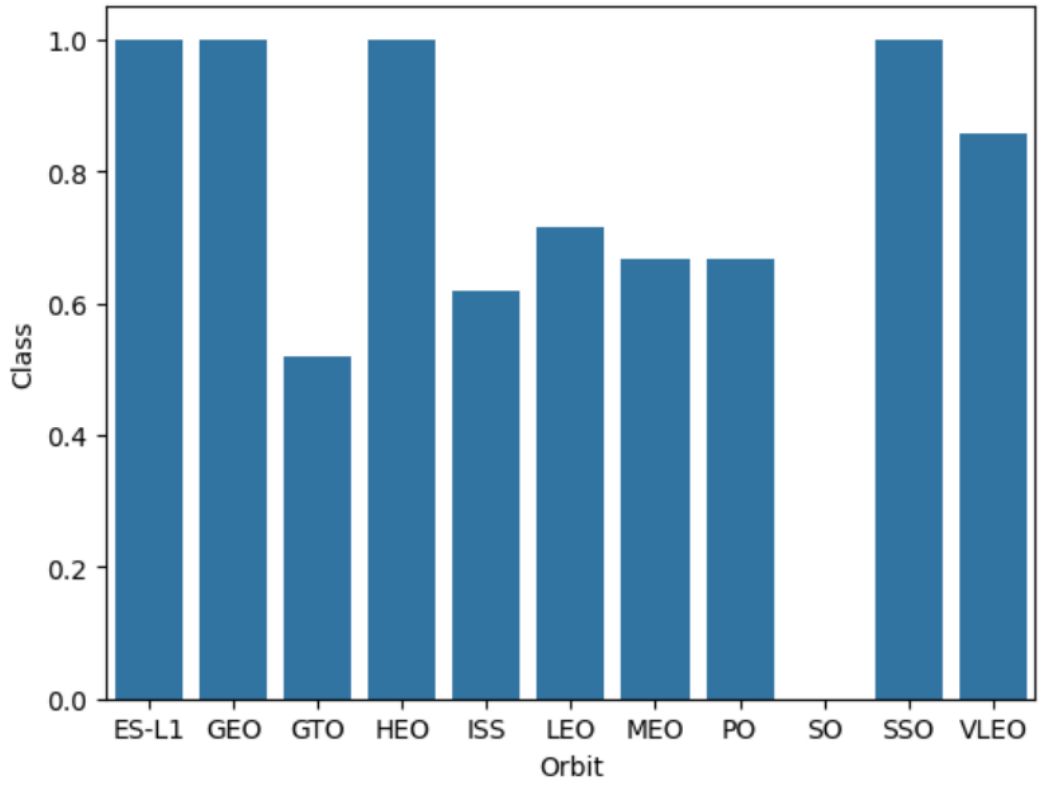
# Payload vs. Launch Site

- In VAFB the maximum payload mass is less than 10000 kg, while in the other two launch sites the heavier ones have around 15000 kg.
- The heavier the mass, the greatest the success rate is for the landing in CCAFS launch site
- In KSC launch site there is not a clearly relationship between the payload mass and the success rate of the landing.



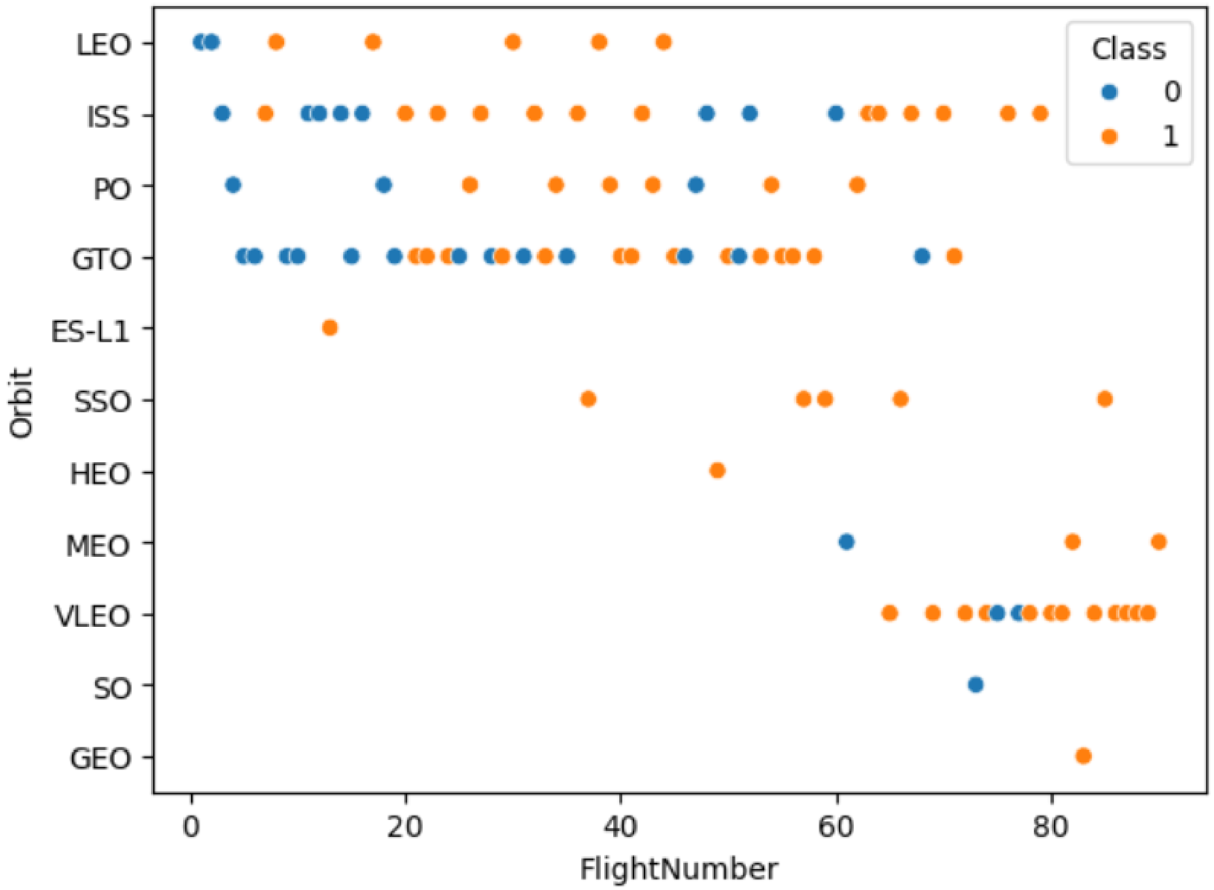
# Success Rate vs. Orbit Type

- For ES-L1, GEO, HEO and SSO we can see that all the landing were successful.
- This is not the case for SO orbit, where the only launch had a failed landing.
- For GTO, ISS, LEO, MEO, PO and VLEO there is a mix between failed and successful landings.



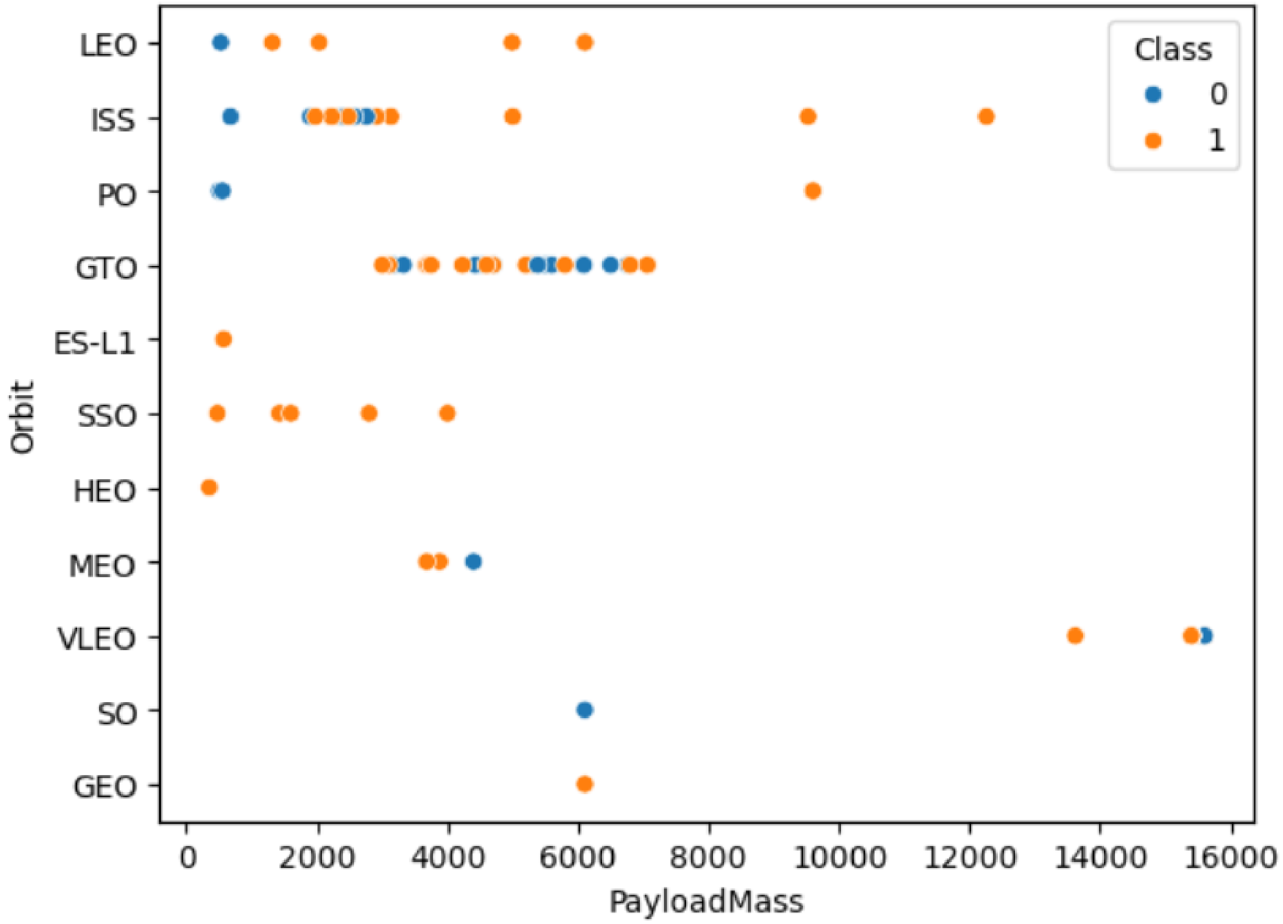
# Flight Number vs. Orbit Type

- For LEO, ISS and PO there is a relationship between the flight number and the success rate.
- For the other orbits there is not a clear relationship, so we cannot state anything with this chart.



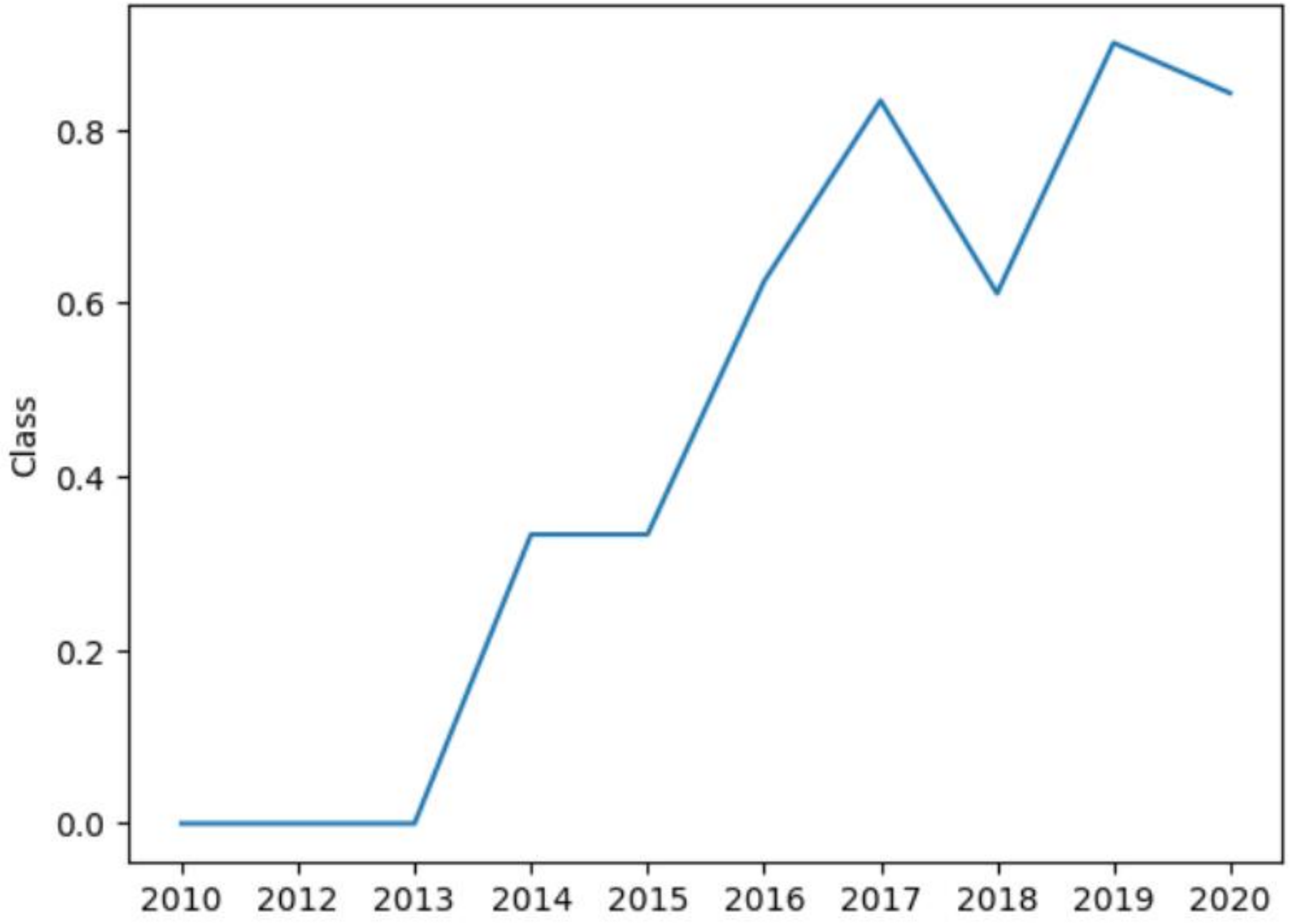
# Payload vs. Orbit Type

- For LEO, ISS and PO there is a relationship between the mass and the success of the landing. The heavier the payloads, the greatest the success rate.
- For GTO there is not a clear relationship as the payload is very similar between the flights and there are both failures and success in the landings.



# Launch Success Yearly Trend

- The successful rate is clearly increasing with the years.
- The past years were almost entire successful.
- There was a problem in the year 2018 which made the success rate to drop



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

# All Launch Site Names

Those are the launch sites that were used in this study.

| 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 |

# Site Names Begin with 'CCA'

Those are 5 records where the site name begins with ‘CCA’

| sum(payload_MASS__KG_) |
| --- |
| 45596 |

# Total Payload Mass

The total payload mass carried by the booster launched from NASA is 45596 KG

| Average payload mass (KG) for booster version 1.1 |
| --- |
| 2928.4 |

# Average Payload Mass by F9 v1.1

The average payload mass for booster F9 v1.1 is 2928.4 Kg

| min(date) |
| --- |
| 2015-12-22 |

# First Successful Ground Landing Date

The first successful landing on ground was in 2015.

# Successful Drone Ship Landing with Payload between 4000 and 6000

- All the booster version which carried between those payload mass were F9 FT version.

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

# Total Number of Successful and Failure Mission Outcomes

- The outcomes of the mission are presented as failures, and success counts for all the missions studied.
- We can see that only 1 mission outcome was failed, while 100 were successful.

| Failure count |
|---:|
| 1 |

| Success count |
|---:|
| 100 |

# Boosters Carried Maximum Payload

- Those are the booster, with their specific version, which carried the maximum payload.
- All the boosters are F9 B5 version.

| 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

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

- In 2015, there were two landing failures on drone ships, the first one in January and the second one in April.
- Both were done with the F9 booster v1.1, but different subversion.

| Landing_Outcome | count |
|---|---|
| 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 |

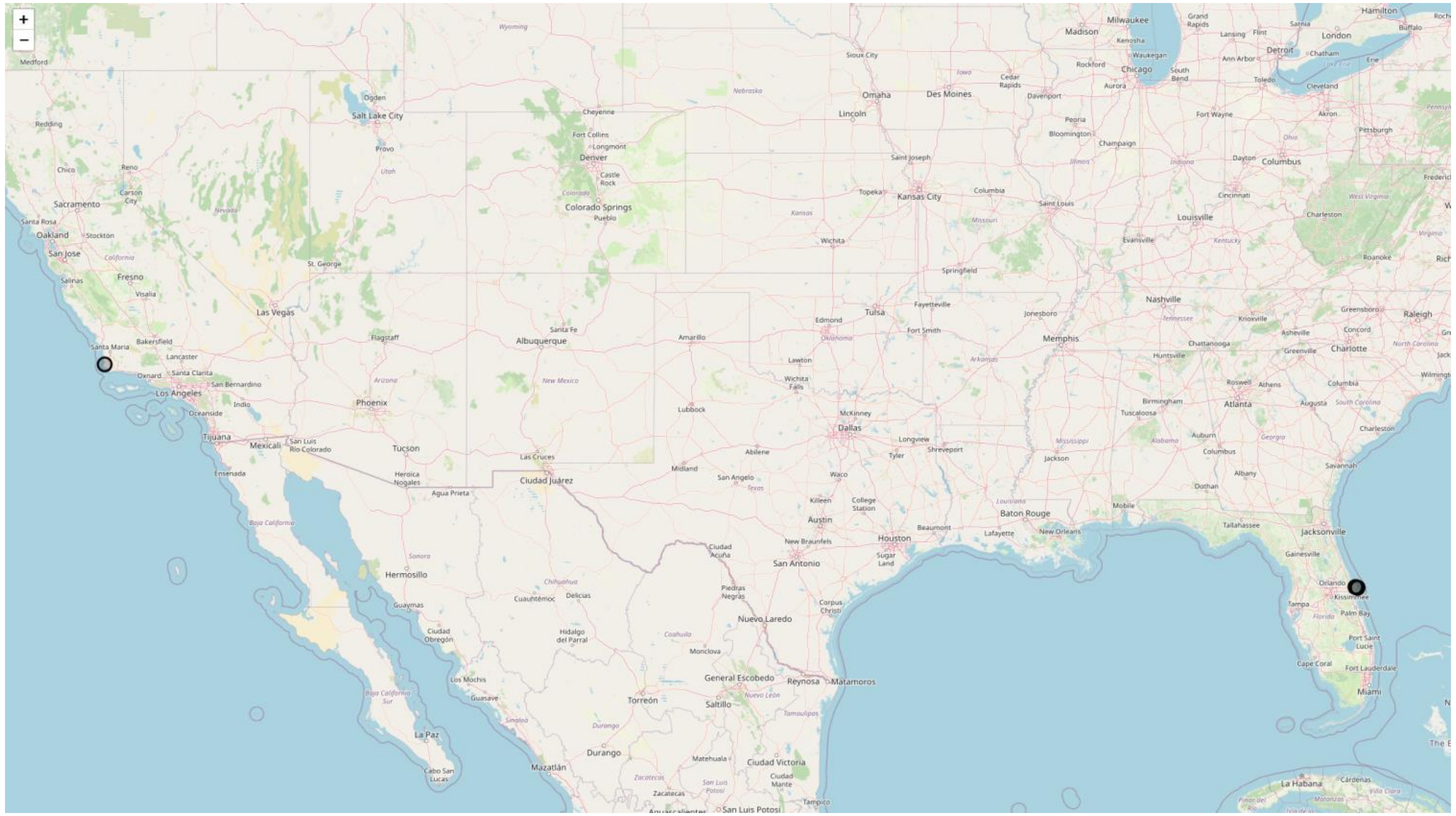
# Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Between those years, the success count was 8, on drone ship and ground pad.
- There were 7 failures in total.
- 10 times there was no attempt and 1 time was precluded
- We also have 3 controlled and 2 uncontrolled, both on the ocean

Section 3

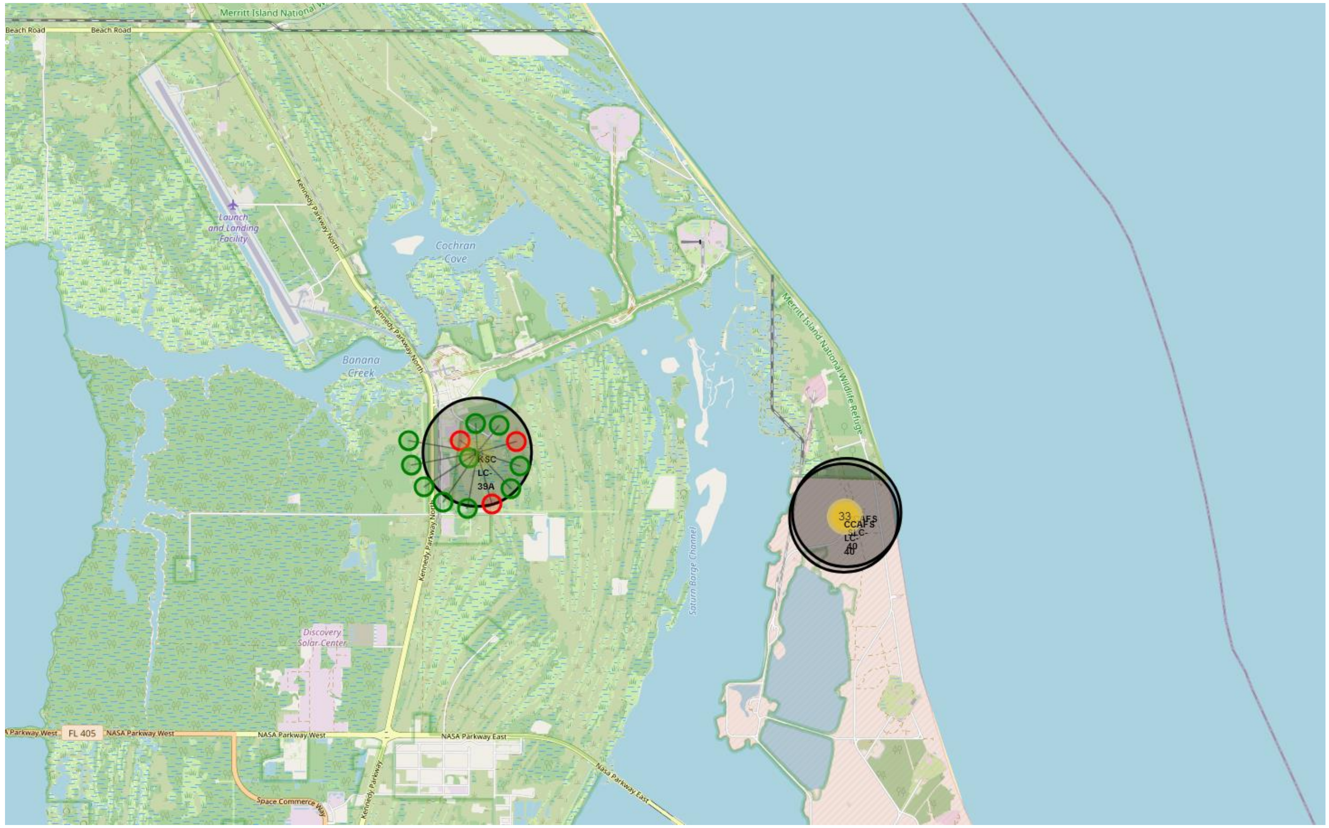
# Launch Sites Proximities Analysis

# Launch sites' location



- The black circles represent the location of the launch sites.
- All the launch sites are located near the coast.
- There are 3 launch sites which are very near between them, while 1 is in the other coast of USA.

# Launch outcomes




- Each launch site has its own outcomes in the presented map.
- Green for the success and red for the failed outcomes, it is easy to know the rate per site watching the colored circles.

# Nearest highway

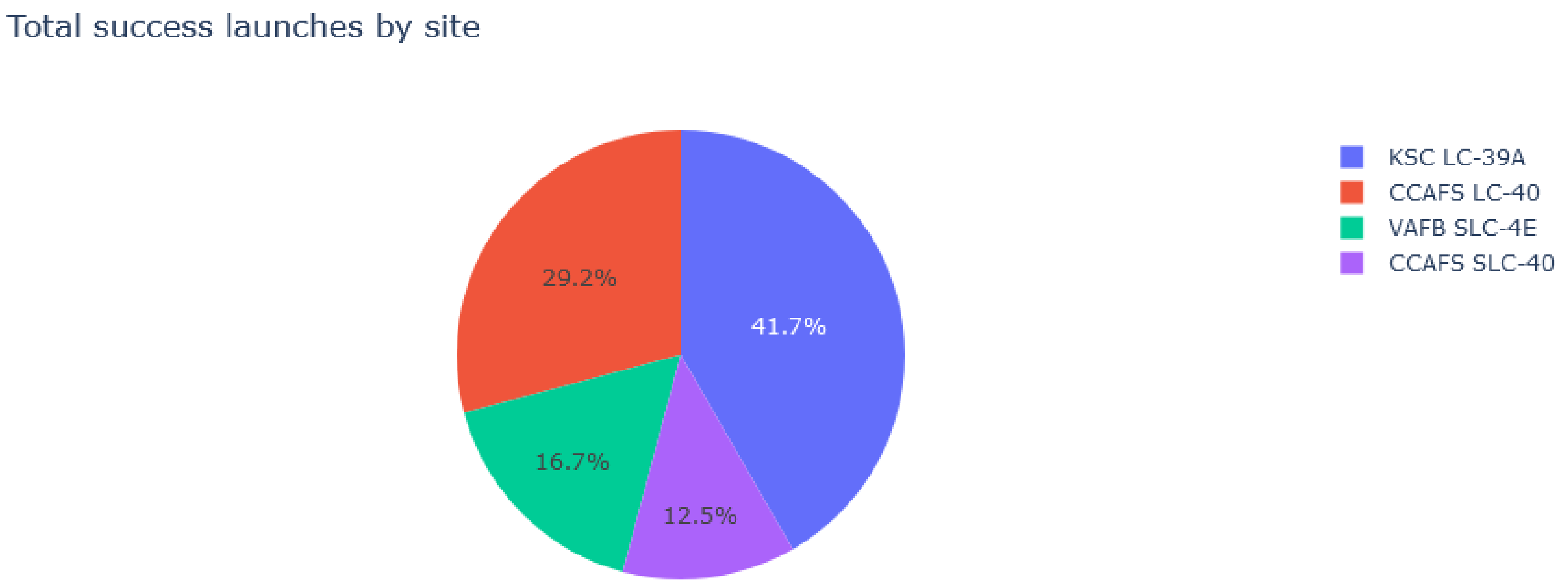

- The blue line shows the distance between one launch site and the nearest highway, in km.
- It helps locate the near access to the site.

Section 4

# Build a Dashboard with Plotly Dash

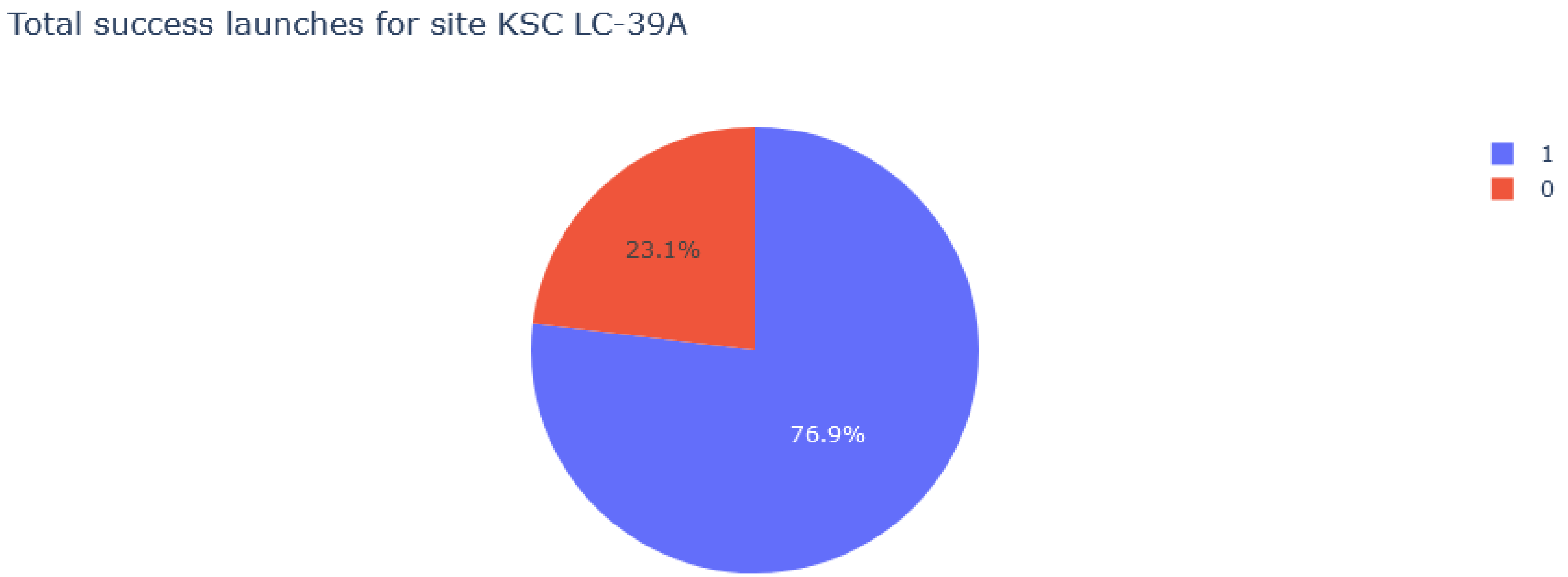
# Total success launches for all sites

- KSC launch site is the one with the most successful launches in comparation with the other launch sites.
- On the other hand, CCAFS SLC is the one with the less successful launches, with only 12.5%.



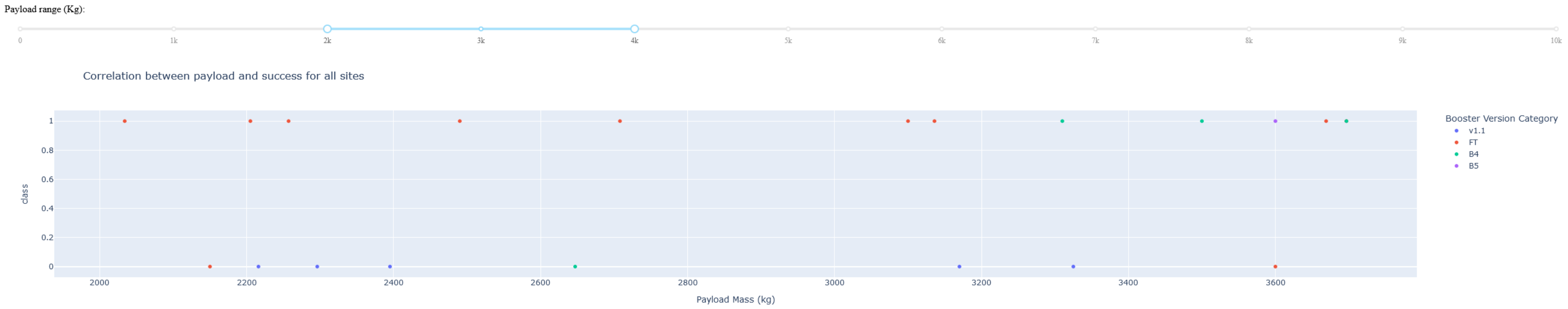
# KSC's landings success rate

- From KSC launches, almost 77% had successful landings
- Only 23% had failed landings.
- KSC has the most successful landing rate



# Payload Mass per Booster Version Category

- FT booster version is the most successful (green dots in the first image).
- V1.1 is the most failed one, with just 1 success (red dots in the first image)
- The most successful booster versions have between 2000 and 4000 kg as payload mass.
- The second image shows the versions between 2000 and 4000 payload mass, being the B5 the one with more failures, while FT is still the one with the most success rate.

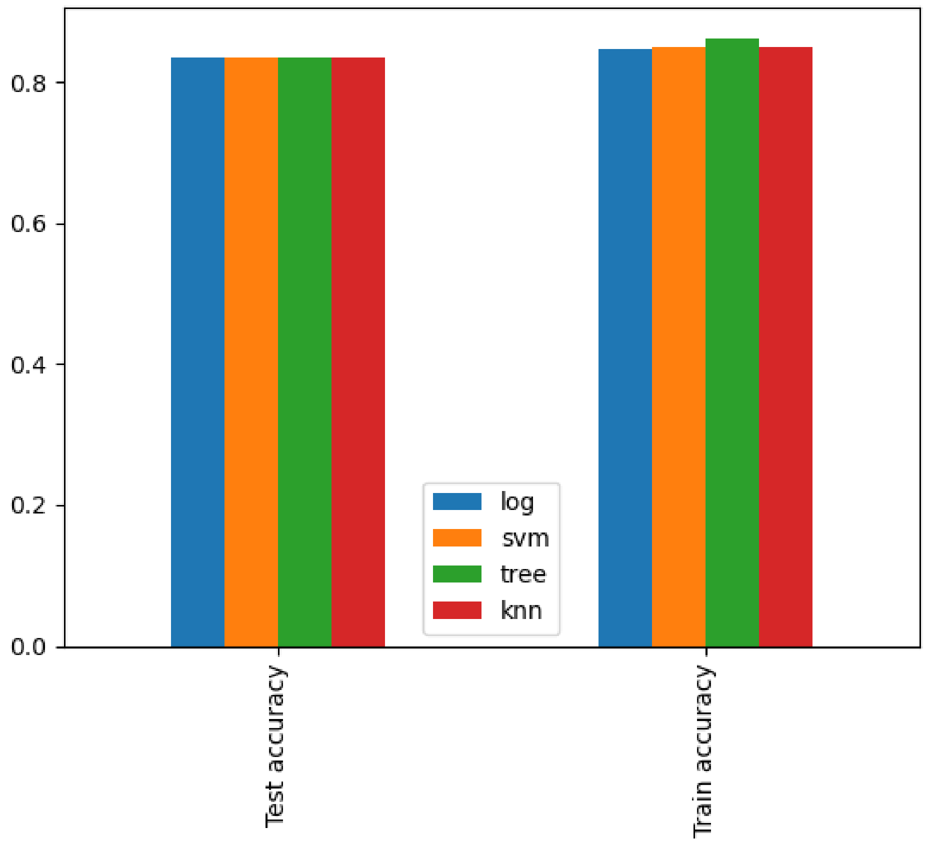



Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

- In this bar plot we can find the accuracy on the trained and tested data for all the models.
- The test accuracy on unseen data is the same for all the models, but we can see an increase in the accuracy of the tree model in the trained data, so we can use it as the chosen model



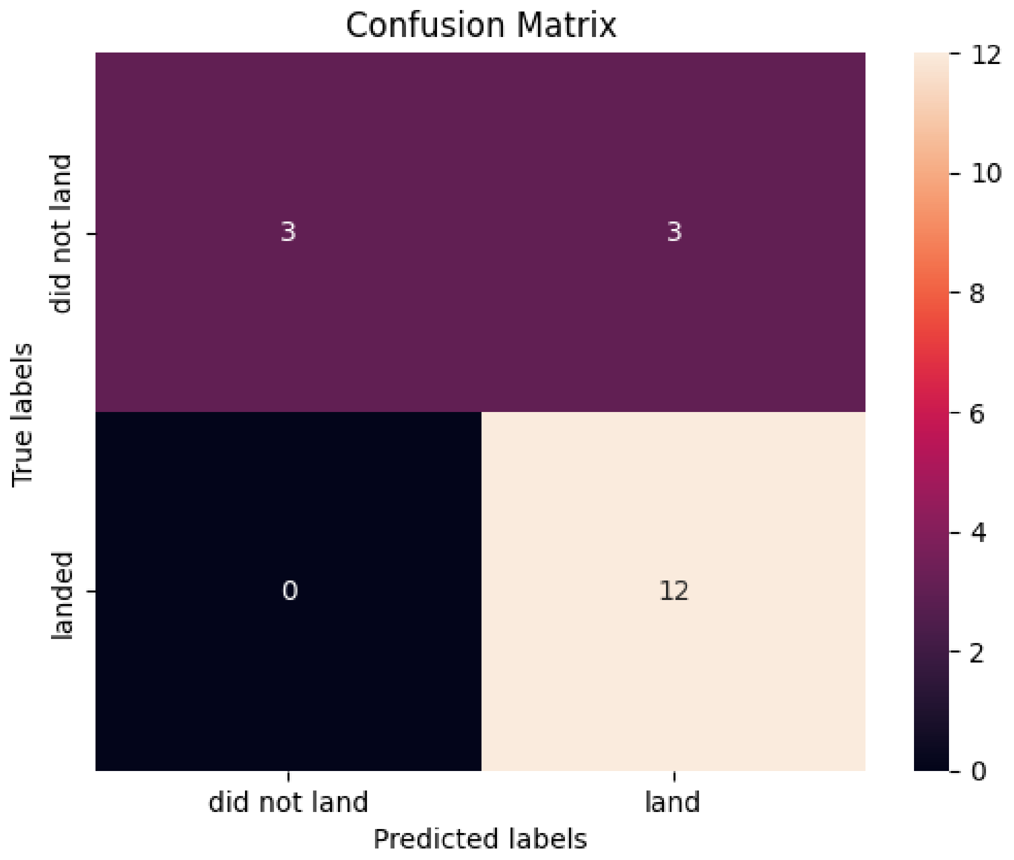
# Confusion Matrix

The confusion matrix represents the predicted labels of the model vs the real labels of the data.

We can see that the model predicted a total of 15 outcomes as landed, from which 12 of those were correct.

On the other hand, it predicted 3 outcomes as not landed and all of them were correctly predicted.

We can see that it performs very good as it only failed 3 predictions out of 18.



# Conclusions

- The EDA helped to define some relations based on the payload mass and the launch site.
- There were relationships found between the different versions of the F9 and the max. payload mass they can carry, as well as which of them failed the landing on drone ships.
- The best model to predict the next outcomes of the landings is the tree classifier.

# Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

# Thank you!