IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdolvahab Daylami
October 2023

Outline

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

Executive Summary

- Summary of methodologies
- Data Collection
- Data Wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with plotly Dash
- Predictive Analysis

Summary of all results

- Exploratory Data Analysis results
- Interactive analytics result
- Predictive analysis result

Introduction

- SpaceX has gained worldwide attention for a series of historic milestones.

  It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas 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 SpaceX for a rocket launch.

- In this project, we will create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collection with SpaceX REST
  - Web Scrapping from Wikipedia
- Perform data wrangling

  performing some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, Decision Tree Classifier, K Nearest Neighbors and Support Vector Machine, models have been built and evaluated for the best classifier

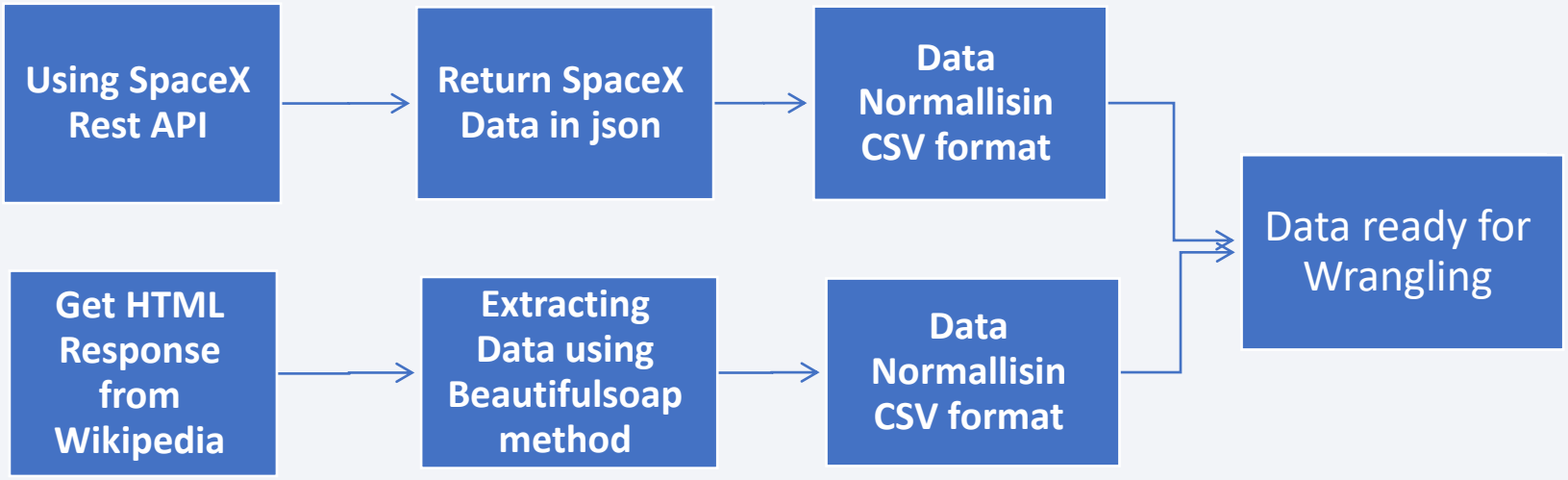
Data Collection

1. Understanding the SpaceX Data:
    - API gives data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

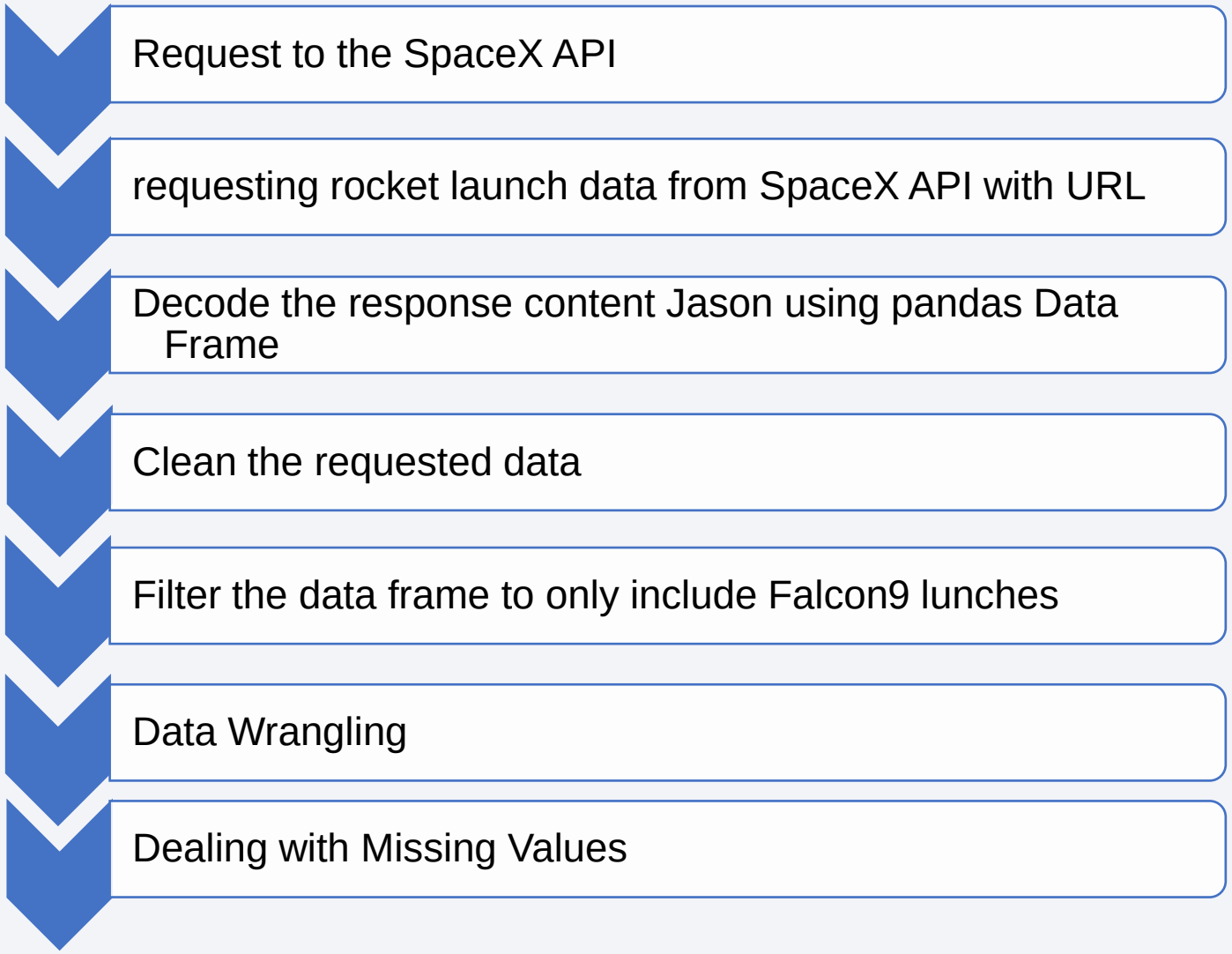
2. Load the dataset in corresponding table:
    - The SpaceX REST API endpoints, or URL, starts with api.spacex.com/v4/.
    - Datasets SpaceX launch data is collected from the SpaceX REST API.

1. Execute SQL queries
    - Another popular source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.

Data Collection Flowchart

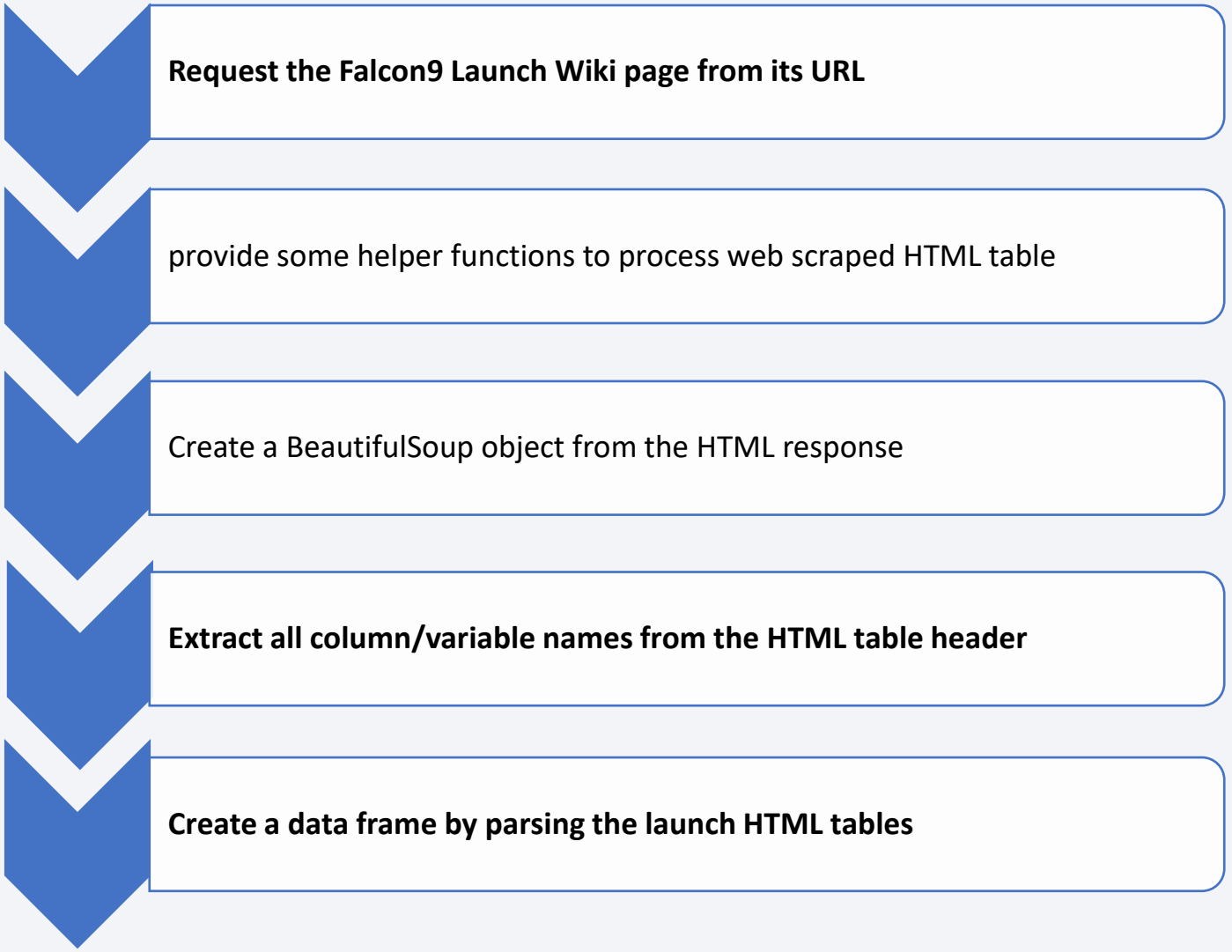


Data Collection – SpaceX API



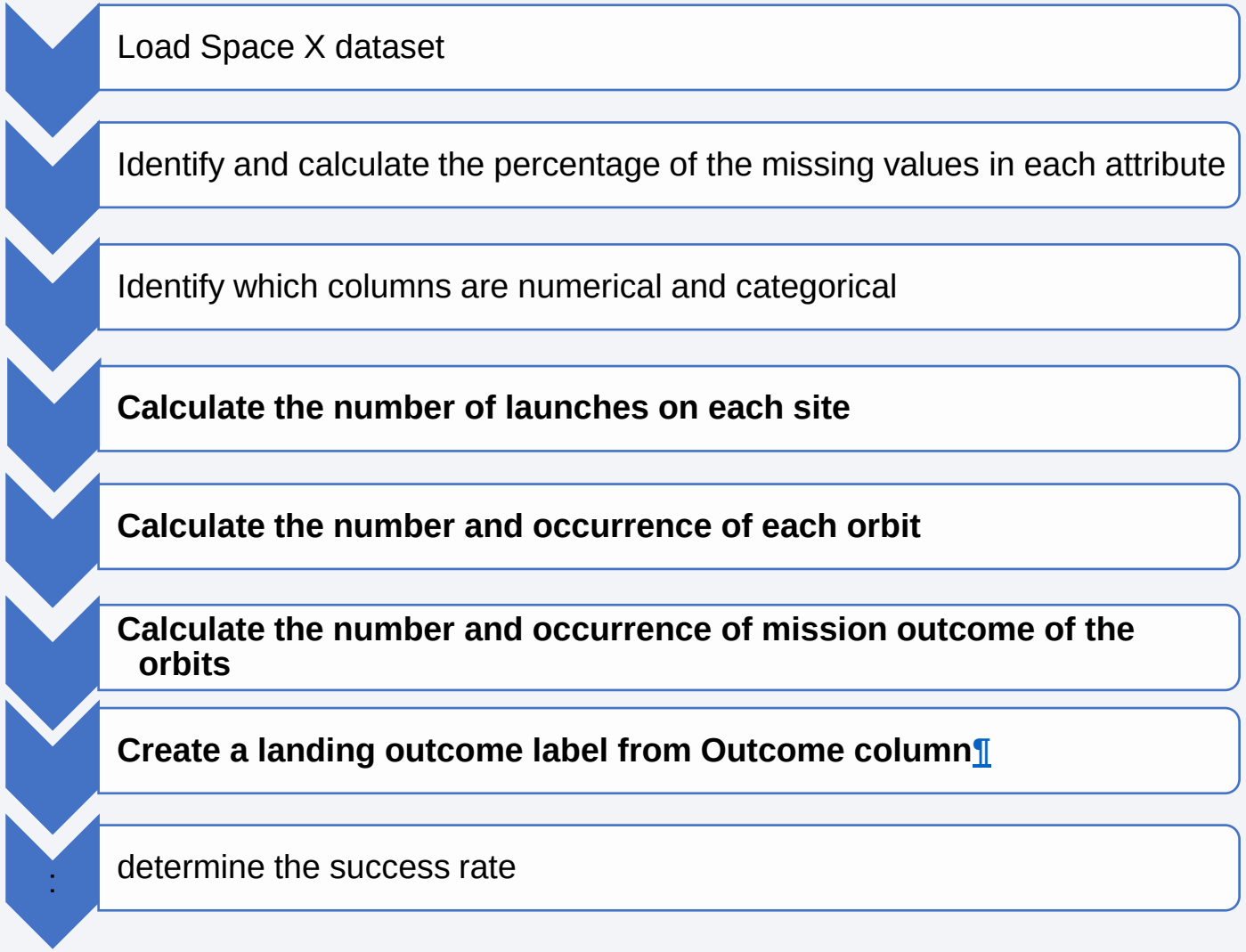
https://github.com/daylami1400/Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/daylami1400/Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/daylami1400/Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Using Cat plot chart to Visualize the relationship between Flight Number and Payload Mass: to overlay the outcome of the launch
- Using Cat plot chart to Visualize the relationship between Flight Number and Launch Site: to observe if there is any relationship between launch sites and Flight Number
- Using Scatterplot chart to Visualize the relationship between Payload and Launch Site: to observe if there is any relationship between launch sites and their payload mass.
- Using Bar plot chart to Visualize the relationship between success rate of each orbit type: to visually check if there are any relationship between success rate and orbit type
- Using Scatter plot to Visualize the relationship between Flight Number and Orbit type: to see if there is any relationship between Flight Number and Orbit type.
- Using Scatter plot to Visualize the relationship between Payload and Orbit type: to reveal the relationship between Payload and Orbit type
- Using Line plot to Visualize the launch success yearly trend: to get the average launch success trend per years

- https://github.com/daylami1400/Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Loading the SQL extension and establish a connection with the database
- Writing and execute SQL queries to *Display the names of the unique launch sites in the space mission*
- *Displaying 5 records where launch sites begin with the string 'CCA'*
- *Displaying the total payload mass carried by boosters launched by NASA (CRS)*
- *Displaying average payload mass carried by booster version F9 v1.1*
- *Listing the date when the first successful landing outcome in ground pad was achieved.*
- *Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
- *Listing the total number of successful and failure mission outcomes*
- *Listing the names of the booster versions which have carried the maximum payload mass. Use a subquery*
- *Listing the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015*
- *Ranking 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*
- **https://github.com/daylami1400/Capstone/blob/main/labs-eda-sql-coursera..ipynb**

Build an Interactive Map with Folium

- adding each site's location on a map using site's latitude and longitude coordinates.
- creating a folium Map object, with an initial center location to be NASA Johnson Space Center at Houston, Texas.
- using `folium Circle` to add a highlighted circle area with a text label on a specific coordinate.
- adding a circle for each launch site in data frame Launch sites.
- Creating and add `folium Circle and folium Marker` for each launch site on the site map.
- Creating a new column in `launch sites` data frame called `marker color` to store the marker colors based on the `class` value.
- Adding a `folium Marker to marker cluster,` for each launch result in `space X` data frame.
- Adding a `Mouse Position` on the map to get coordinate for a mouse over a point on the map. (such as railway)
- Calculating the distance between two points on the map based on their Lat and Long.
- Mark down a point on the closest coastline using Mouse Position and calculate the distance between the coastline point and the launch site.
- Drawing a `Poly Line` between a launch site to the selected coastline point

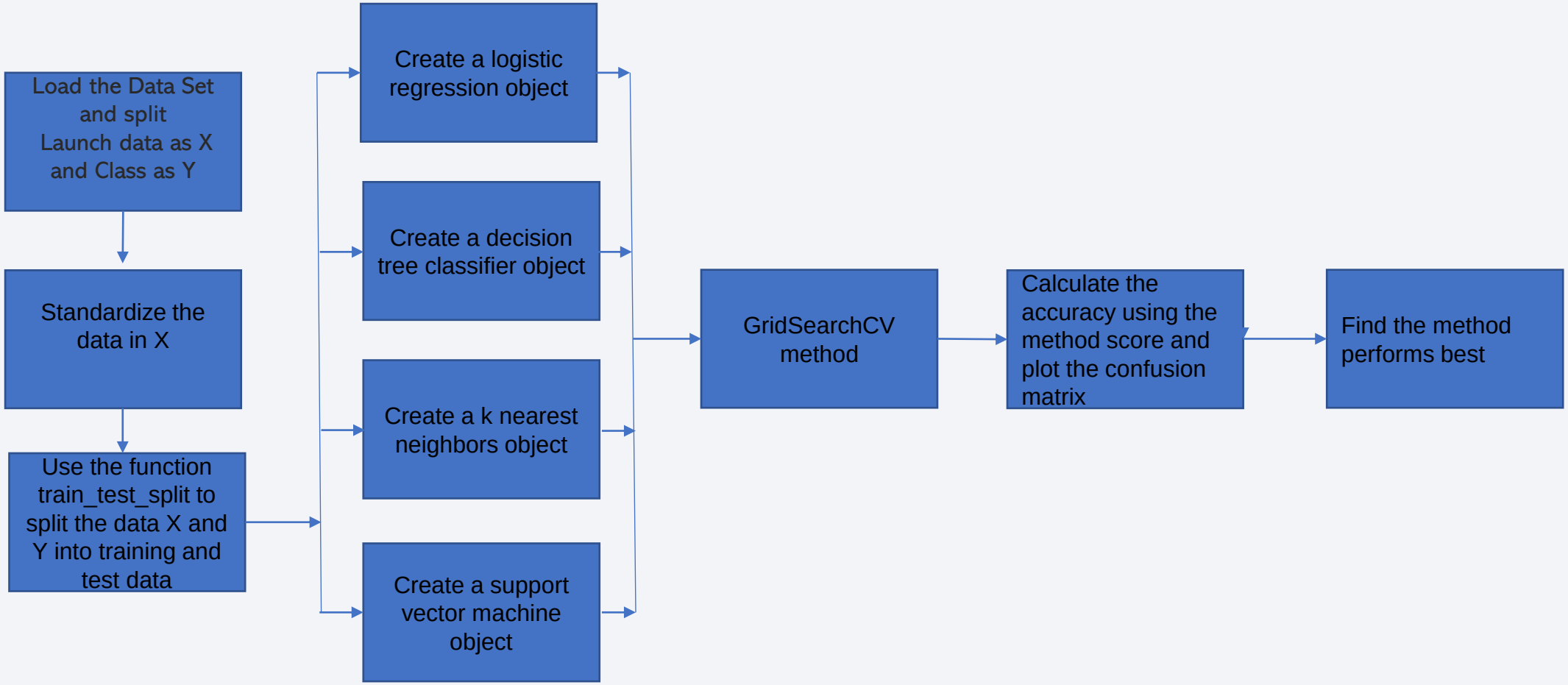
- https://github.com/daylami1400/Capstone/blob/main/Launch_Sites_Locations_Analysis_with_Folium.ipynb

Build a Dashboard with Plotly Dash

- Creating a dropdown menu to let us select different launch sites.
- Adding a callback function for `site-dropdown` as input and `success-pie-chart` as output to select one specific site and check its detailed success rate
- Adding a Range Slider to Select Payload
- we want to find if variable payload is correlated to mission outcome. From a dashboard point of view, we want to be able to easily select different payload range and see if we can identify some visual patterns.
- Adding a callback function for `site-dropdown` and `payload-slider` as inputs, `success-payload-scatter-chart` as output
- plotting a scatter plot with the x axis to be the payload and the y axis to be the launch outcome. As such, we can visually observe how payload may be correlated with mission outcomes for selected site.
- to color-label the Booster version on each scatter point and observe mission outcomes with different boosters.

- https://github.com/daylami1400/Capstone/blob/main/Build_a_Dashboard_Application_with_Plotly_Dash.ipynb

Predictive Analysis (Classification) Flowchart



Results

- ## Exploratory data analysis results

- We see that different launch sites have different success rates. CCAFS LC-40 has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- Launch Site CCAFS SLC 40 has more flights while VAFB SLC 4e has more success flights.
- Observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- The success rate since 2013 kept increasing till 2020

Results

- ## Interactive analytics demo in screenshots:

- Launch Site KSC LC 39A achieved a 76.9% success rate, getting a 23.1% failure rate
- 41.7% of launches don in KSC LC 39A launch site
- Payload Mass in range 2600 to 3800 has more success rate

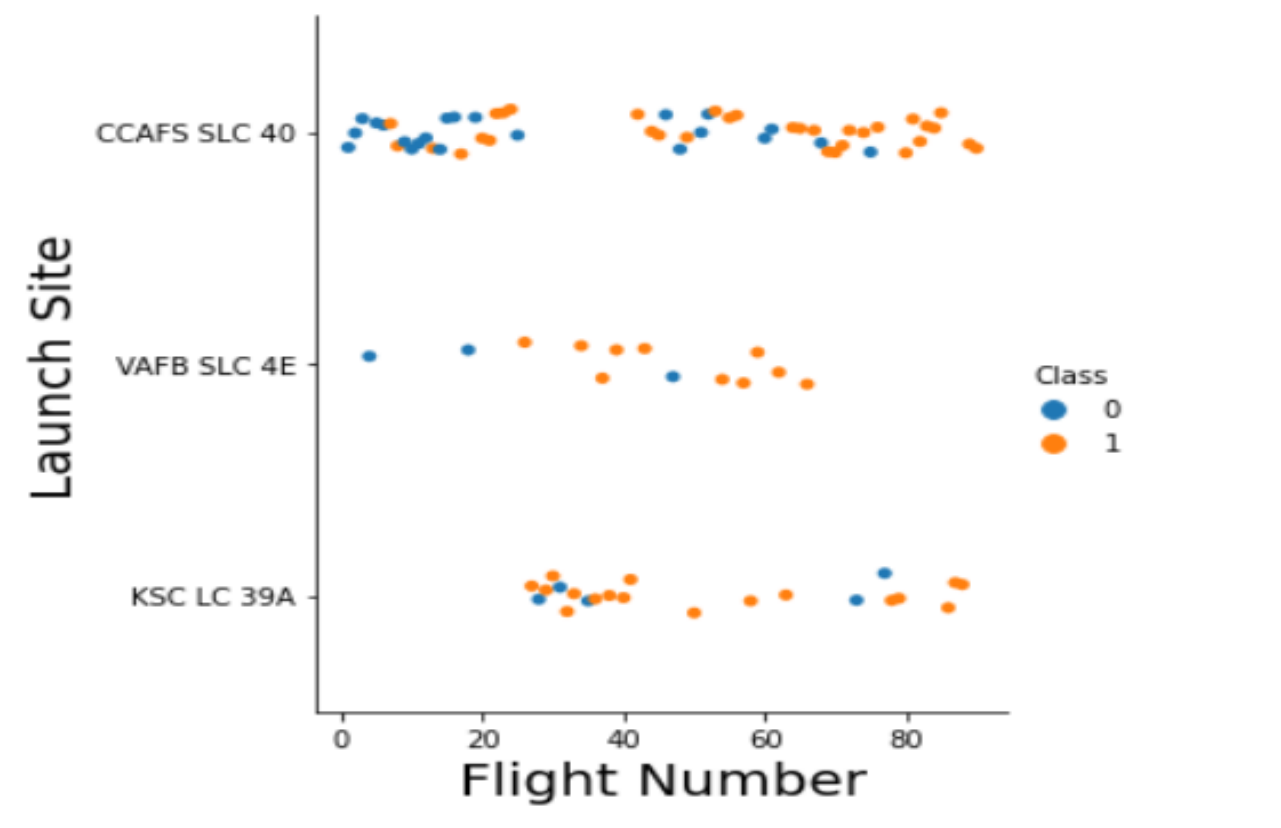
- ## Predictive analysis results:

- The best methods from logistic regression, decision tree classifier, k nearest neighbors and support vector machine is decision tree classifier by 0.8767 accuracy.

Section 2

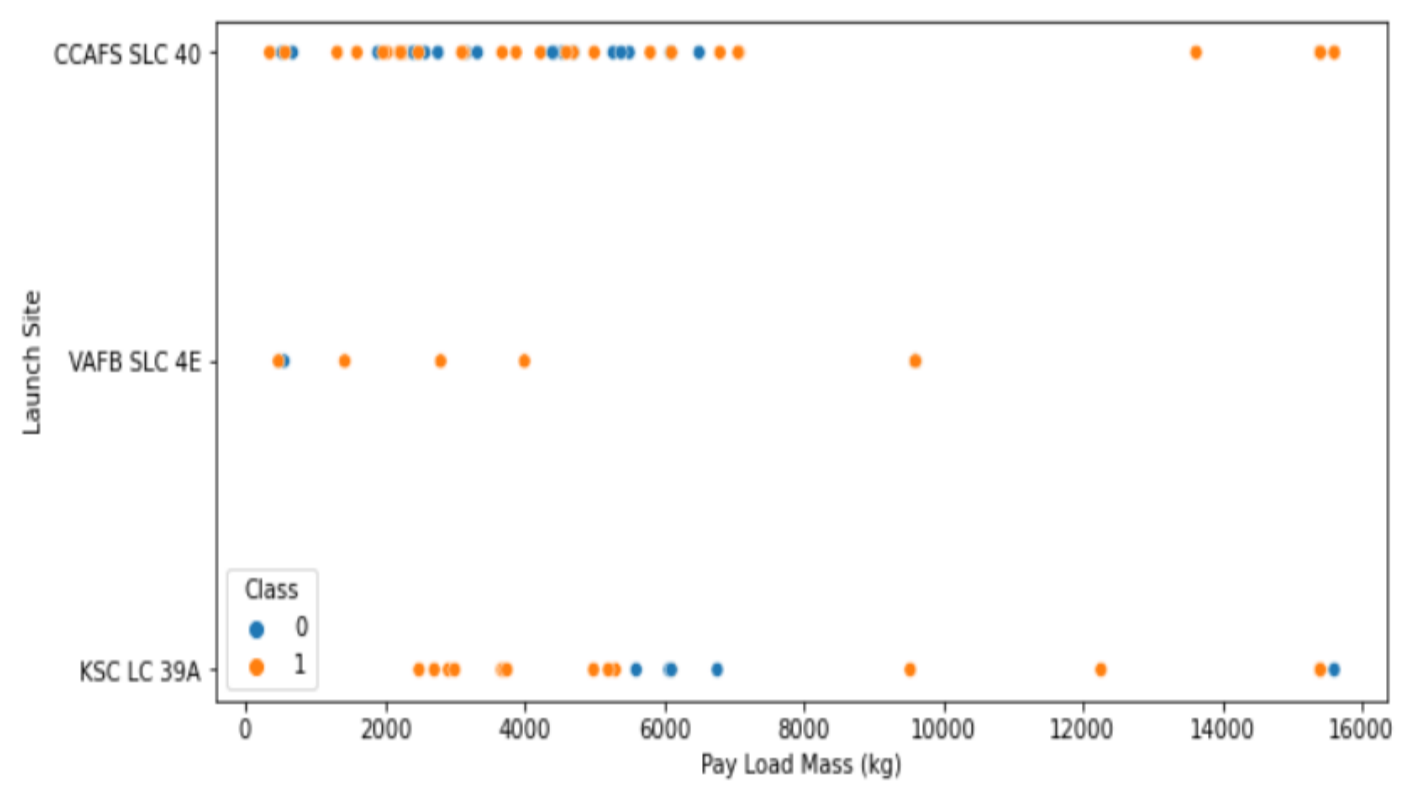
Insights drawn from EDA

Flight Number vs. Launch Site



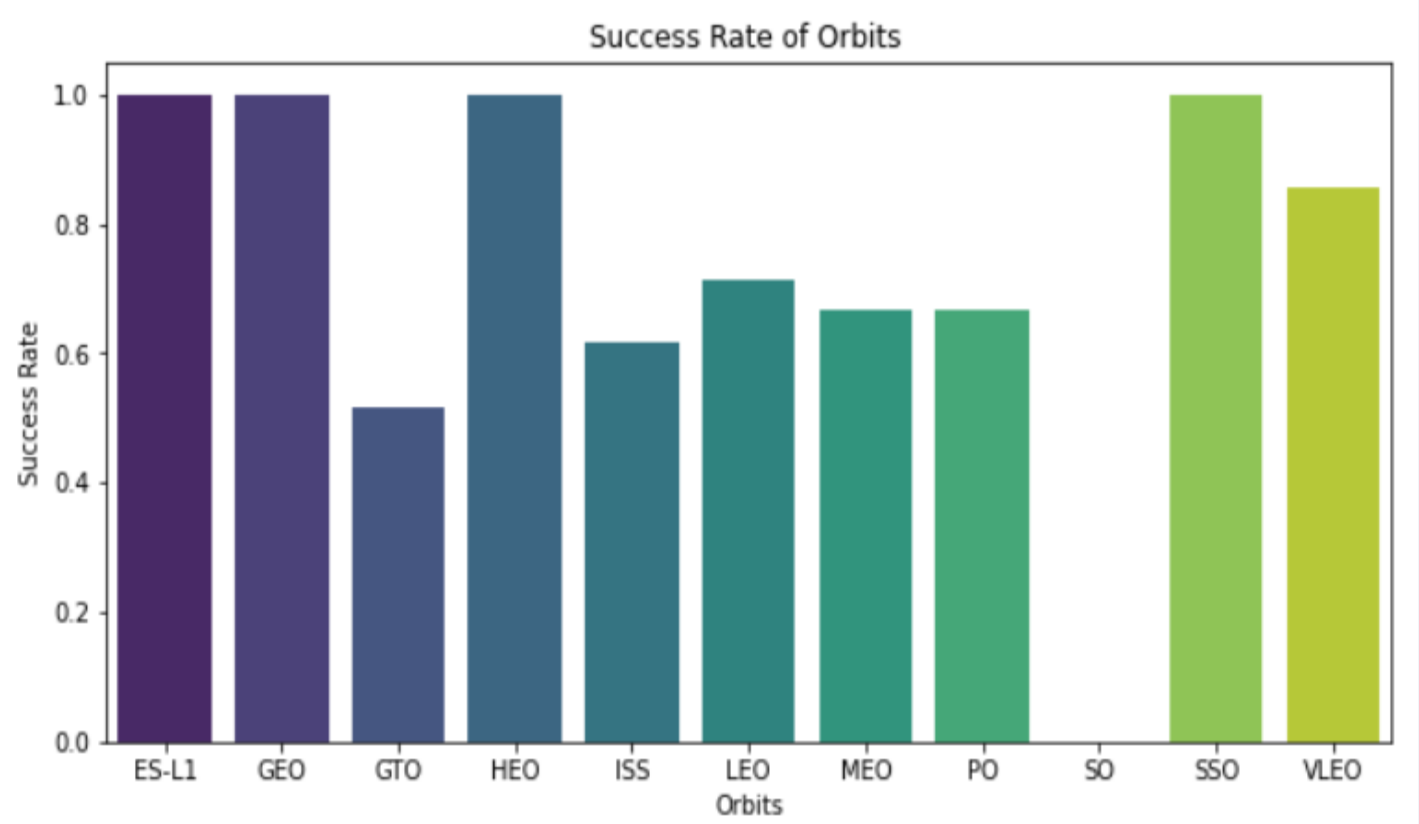
- Launch Site CCAFS SLC 40 has more flights while VAFB SLC 4E has more success flights.

Payload vs. Launch Site



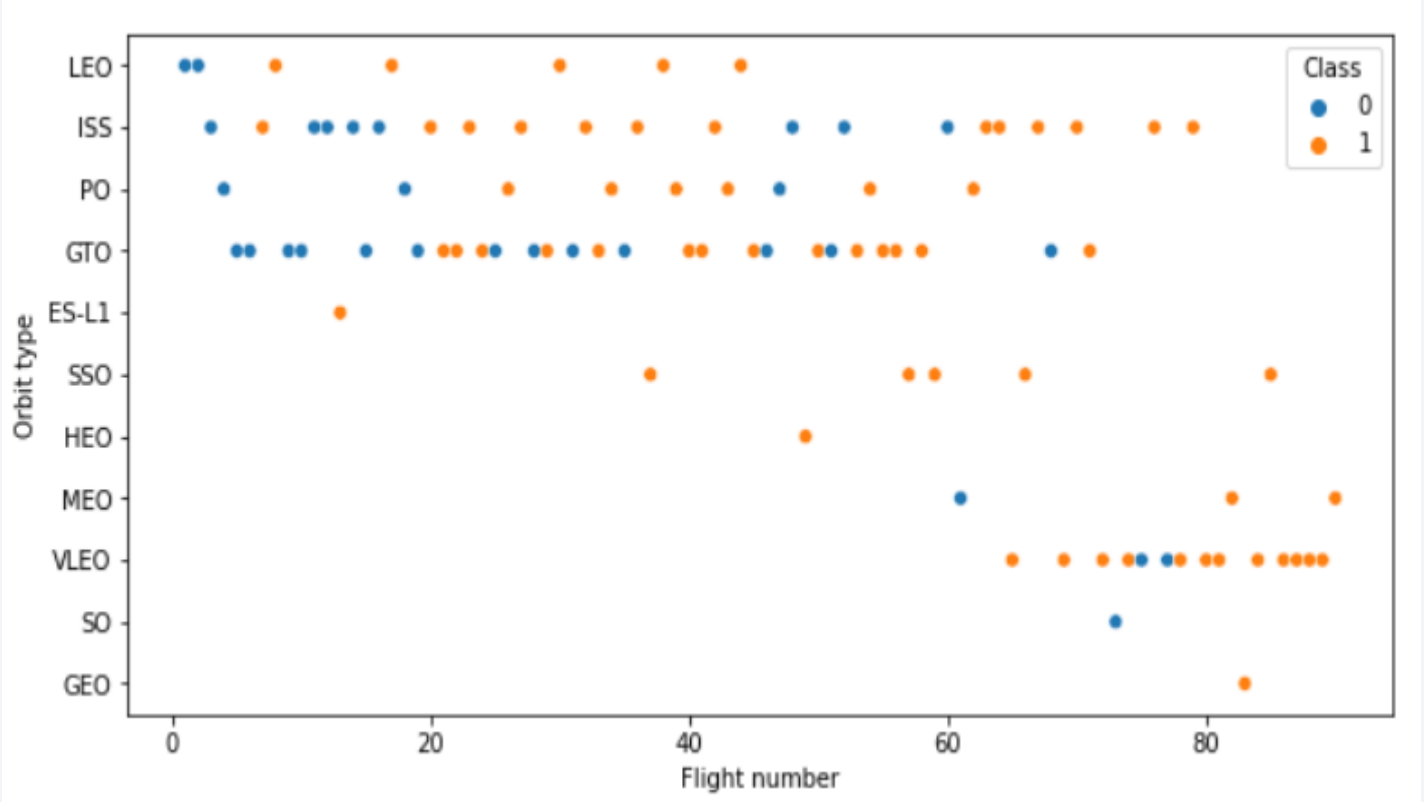
We observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



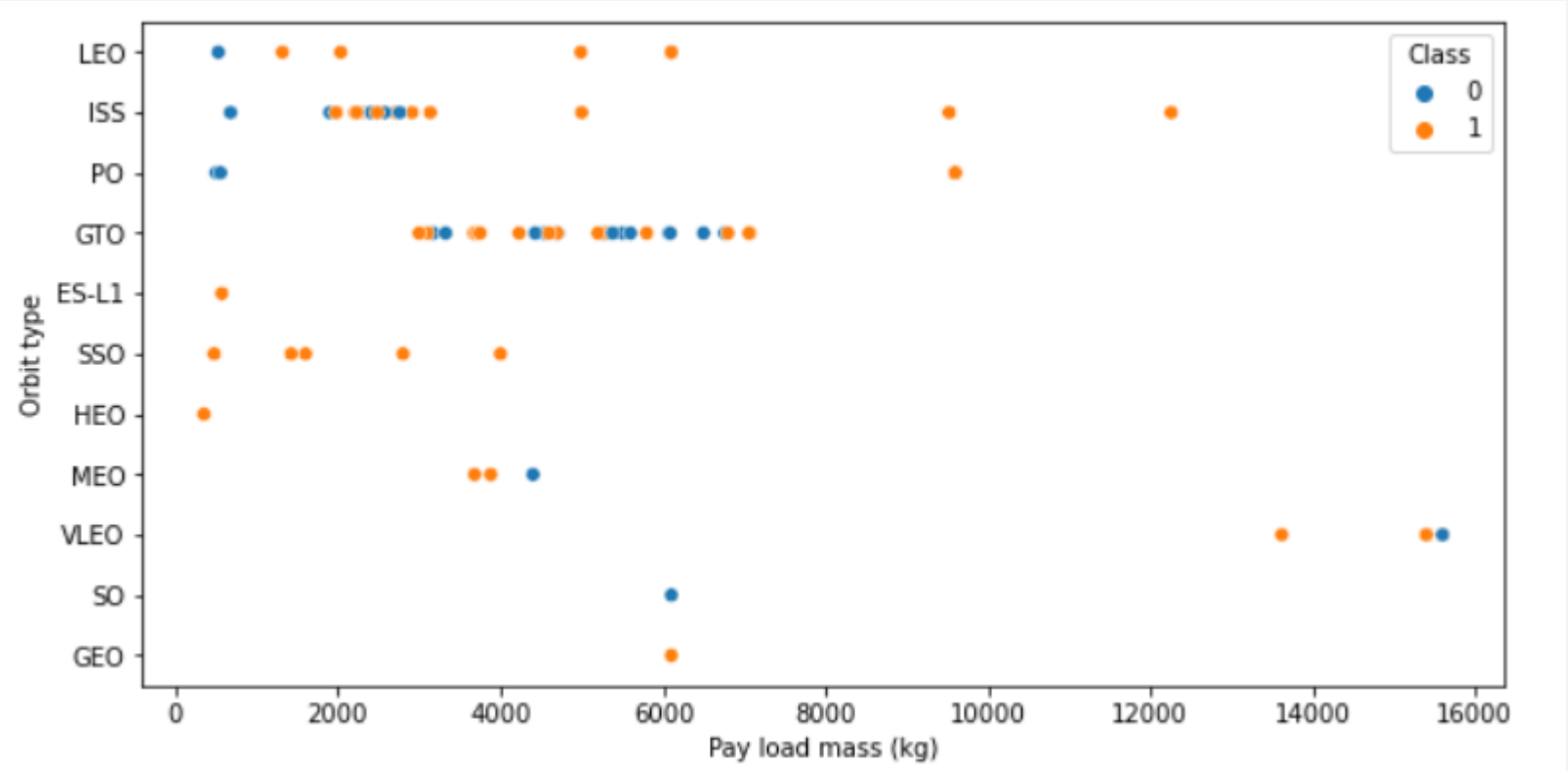
- The orbit types of ES-L1, GEO, HEO, and SSO are the highest success rate.

Flight Number vs. Orbit Type



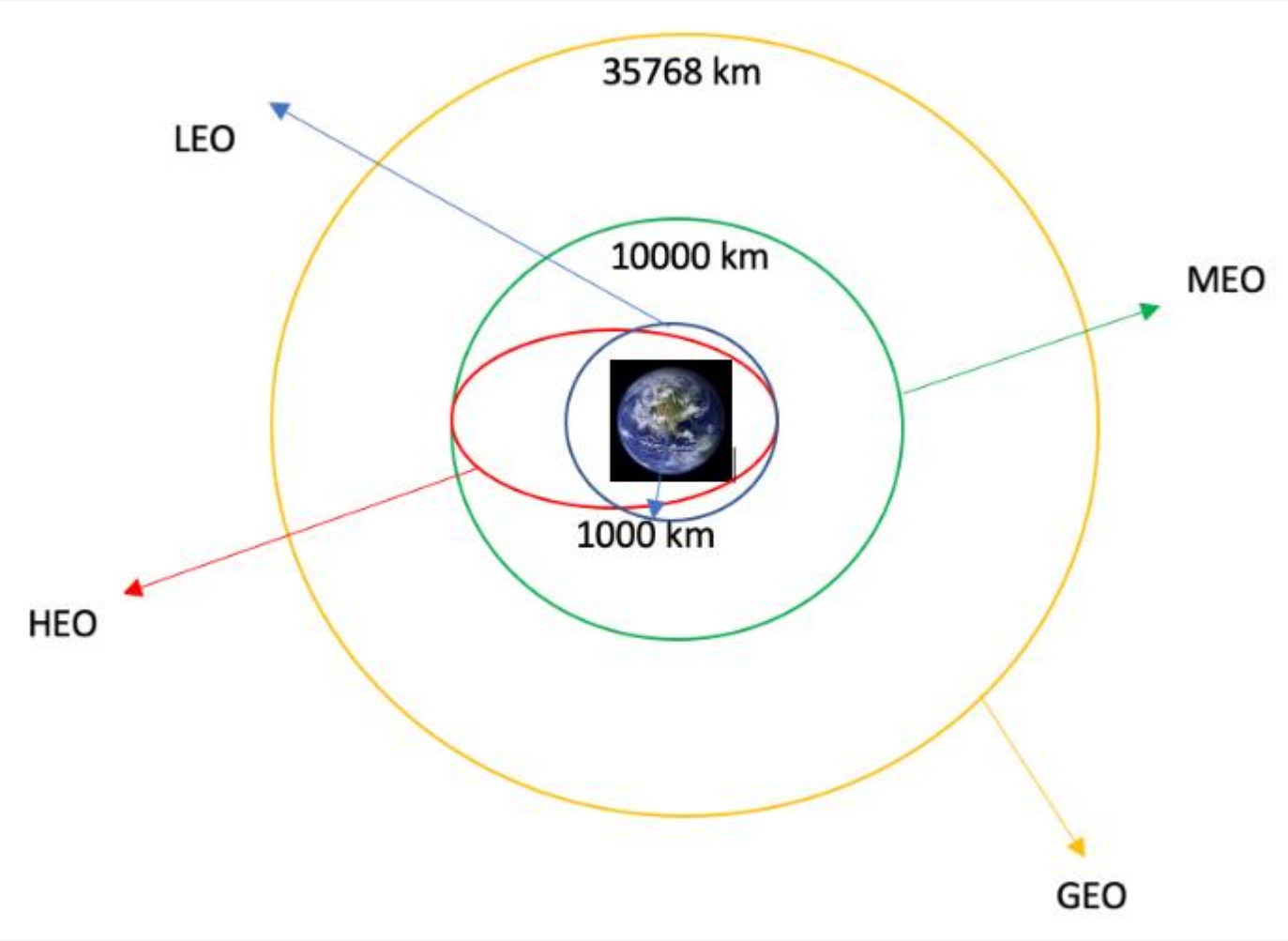
- The chart shows that for LEO orbit the success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

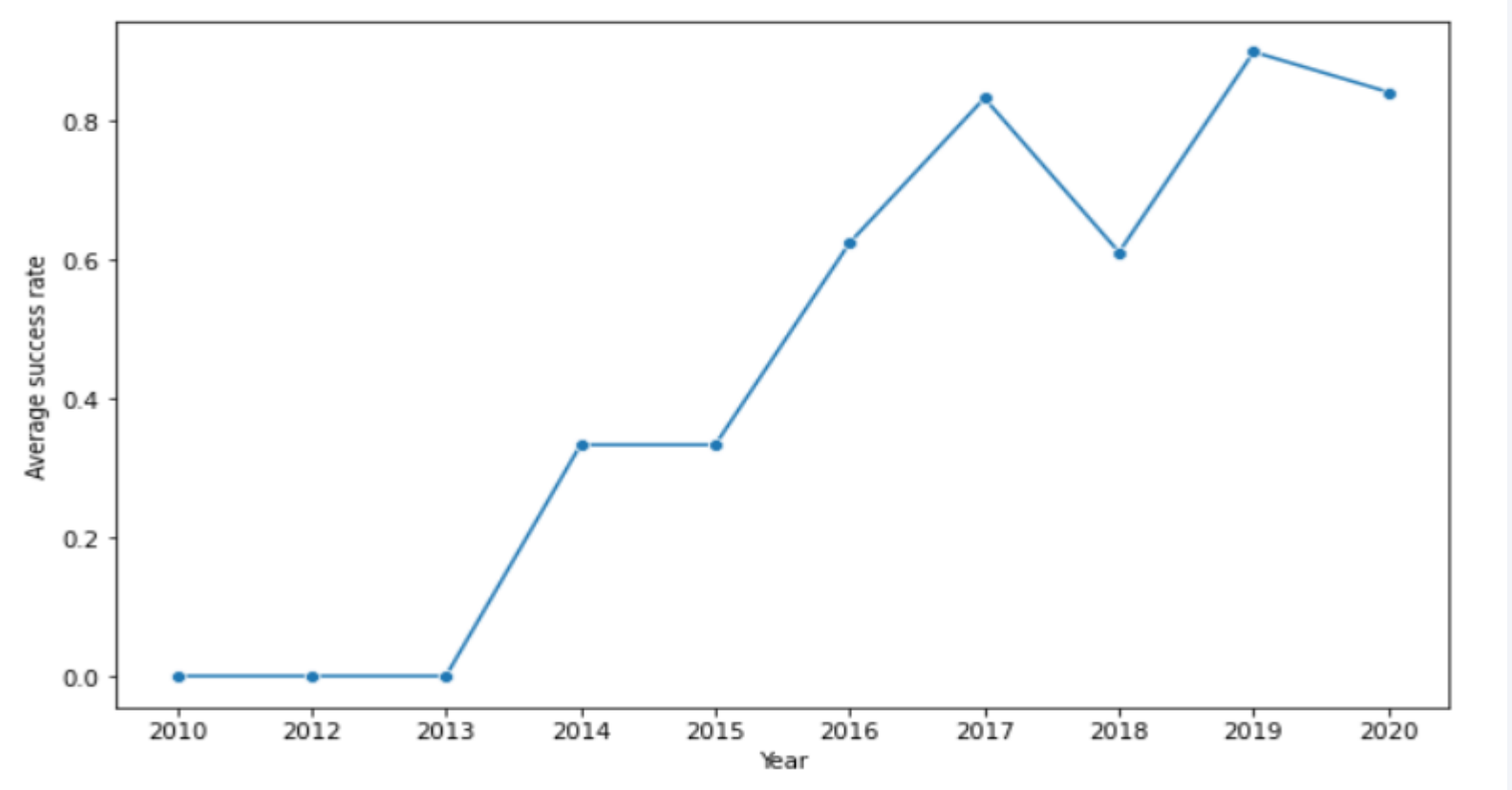


- With heavy payloads the successful landing or positive landing rate are more for Polar LEO and ISS.
- However for GTO we cannot distinguish this well, as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Orbit Types



Launch Success Yearly Trend



- The chart shows that the success rate since 2013 kept increasing till 2020

All Launch Site Names

%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL:

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5:

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

Total Payload Mass

- *total payload mass carried by boosters launched by NASA (CRS)*
- %sql select sum(PAYLOAD_MASS__KG_) as Total_Payload_Mass from SPACEXTBL where Customer = 'NASA (CRS)' :
- Total Payload Mass = 45596 kg

Average Payload Mass by F9 v1.1

- *Display average payload mass carried by booster version F9 v1.1*
- %sql select avg(PAYLOAD_MASS__KG_) as Average_Payload_Mass from SPACEXTBL where Booster_Version = 'F9 v1.1’
- Average Payload Mass by F9 v1.1 = 2928.4 kg

First Successful Ground Landing Date

- %sql select min(Date) as First_Successful_Landing from SPACEXTBL where Landing_Outcome = 'Success (ground pad)’
- First successful ground landing date = 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select Booster_Version from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ >4000 and PAYLOAD_MASS__KG_ <6000

Booster Version:

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

Total Number of Successful and Failure Mission Outcomes

- %sql select Mission_Outcome, count(*) as Count_of_Mission from SPACEXTBL group by Mission_Outcome

Mission Outcome	Count of Mission
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%%sql select Booster_Version

from SPACEXTABLE

where PAYLOAD_MASS__KG_ = (select
MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE);
```

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

- from SPACEXTBL where Date like'%2015%' and Landing_Outcome like 'Failure (drone ship)

Month	Failure_Landing_Outcome	Booster Version	Launch Site
October	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

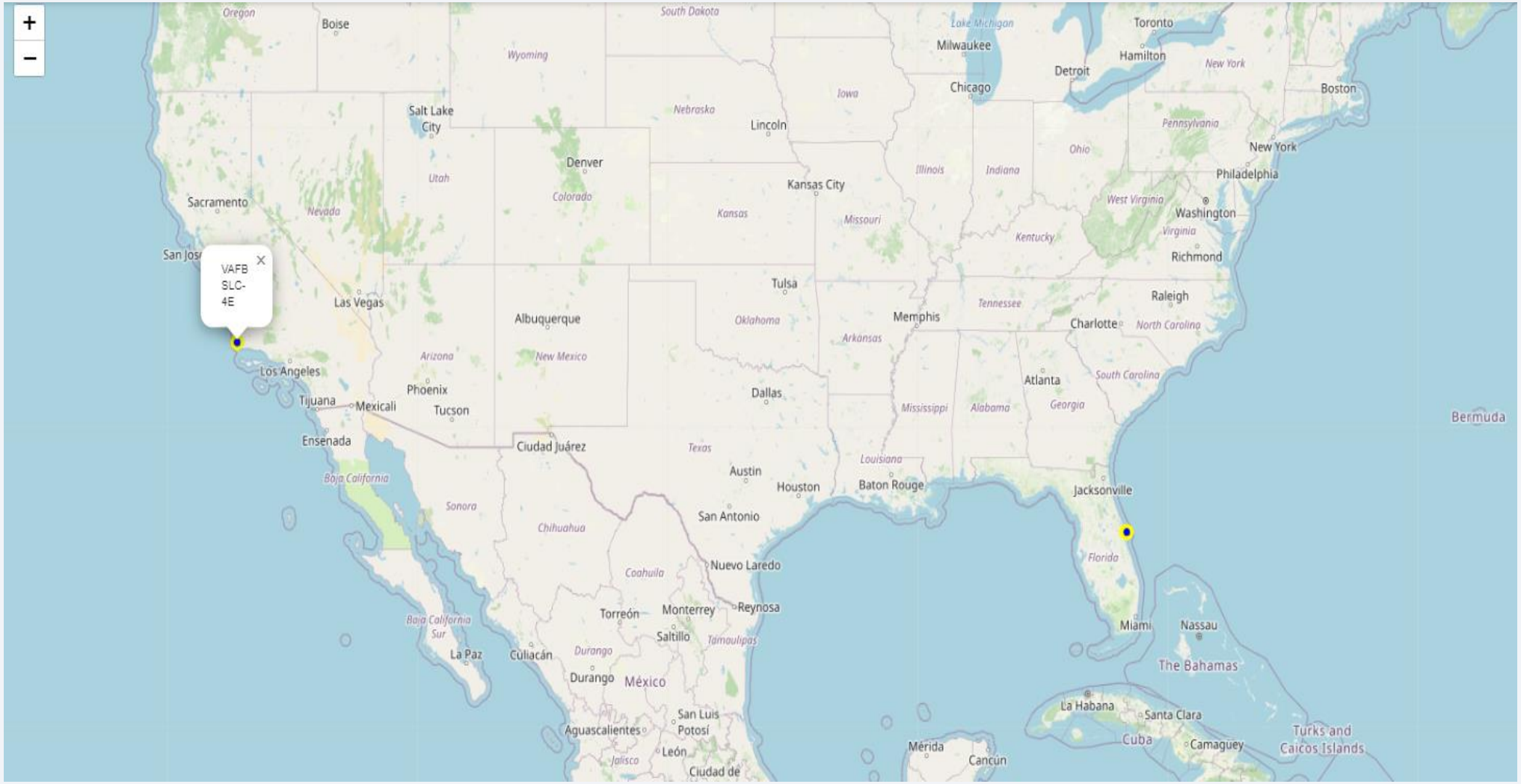
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- %%sql select Landing_Outcome, count(*) as Outcome_Count from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20’ group by Landing_Outcome order by Outcome_Count desc;

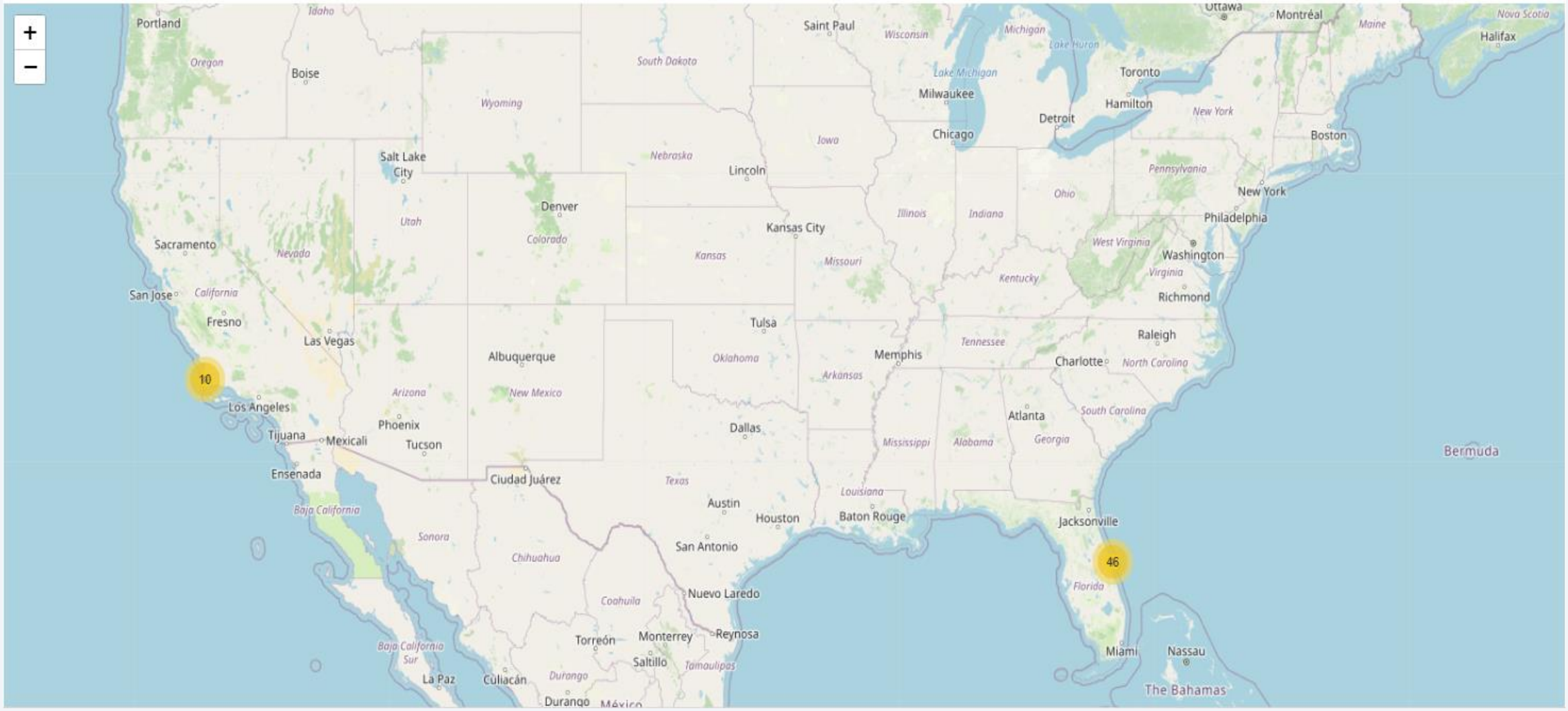
Landing Outcome	Outcome Count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1


Section 3
Launch Sites
Proximities Analysis

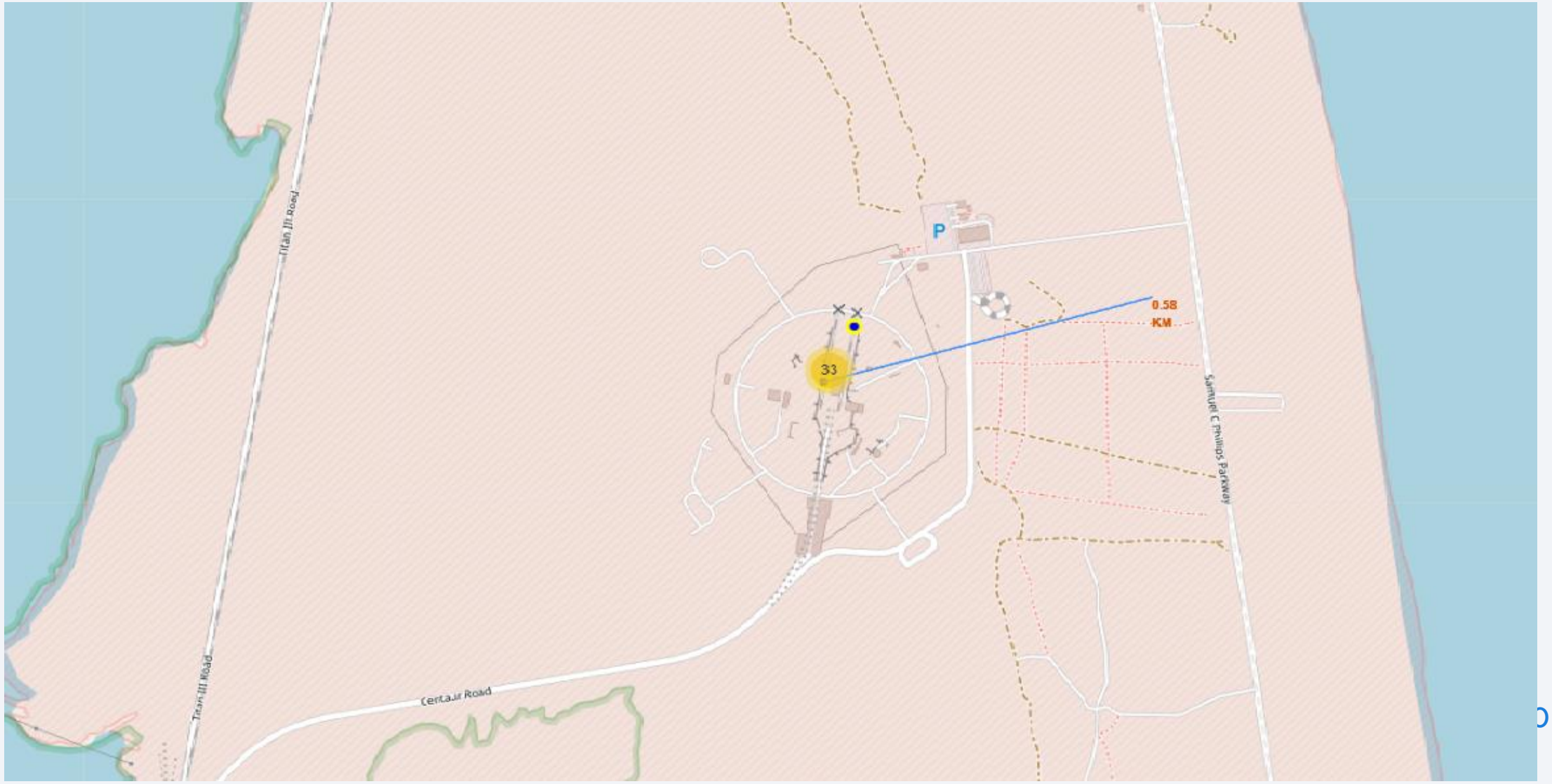
All launch sites marked on the map

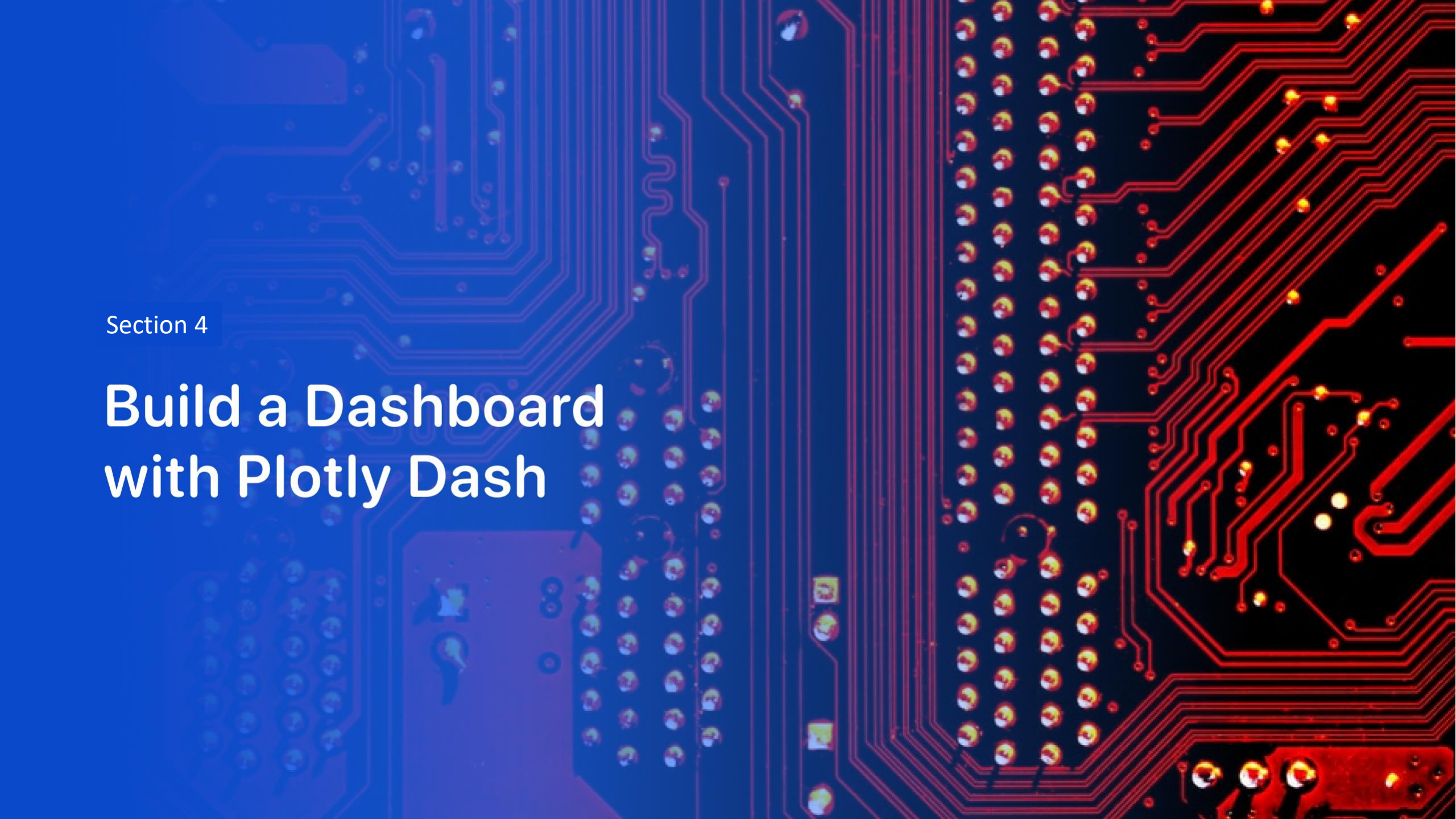


Color-labeled launch outcomes



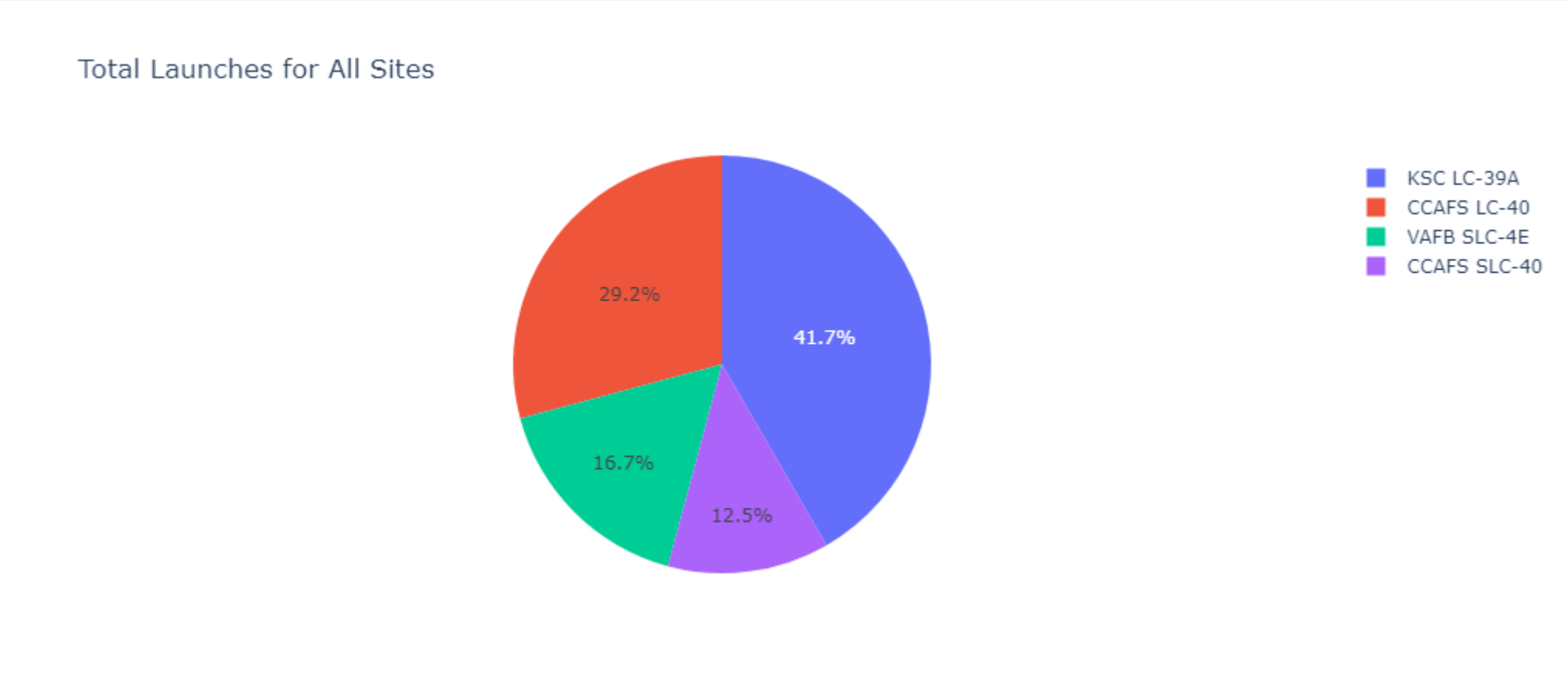
Distances between a launch site to its proximities




Section 4
Build a Dashboard with Plotly Dash

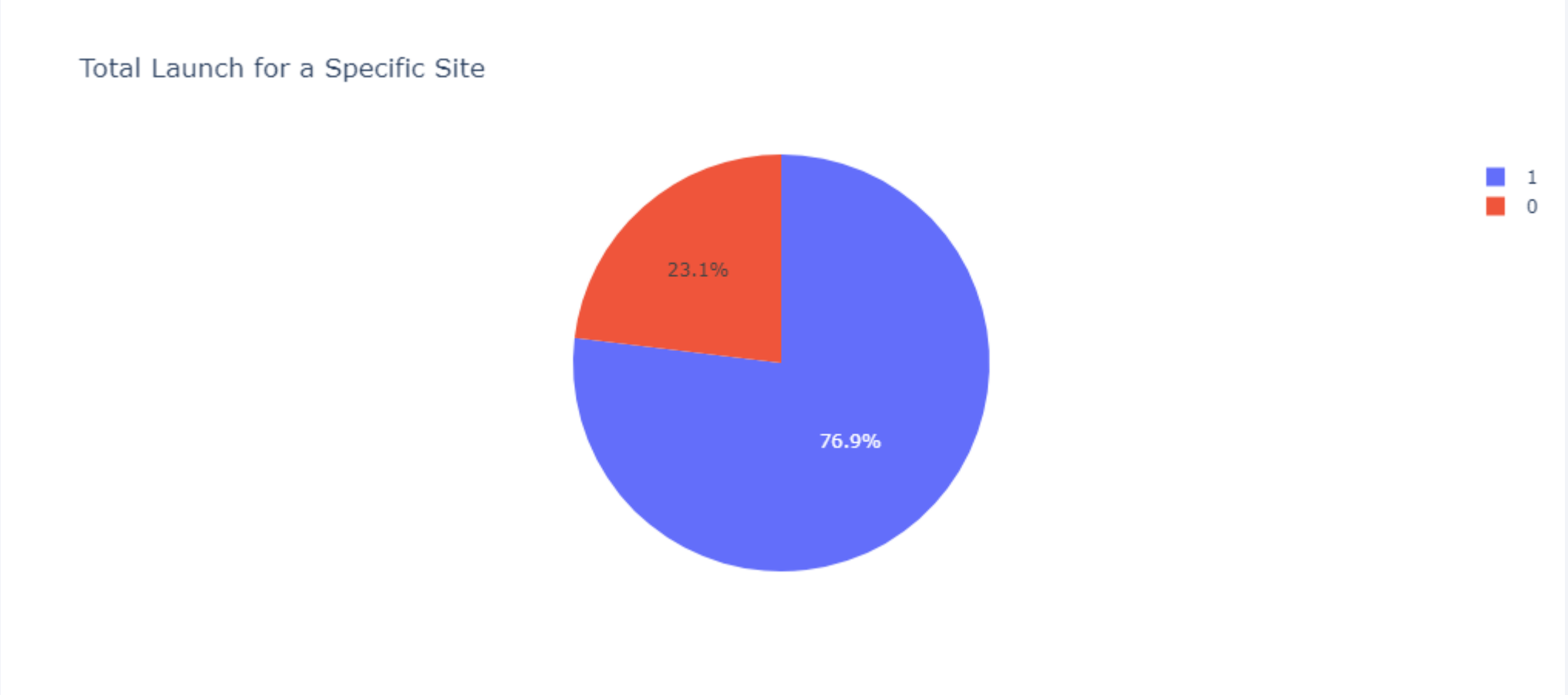
Total success launches for all sites

- **SpaceX Launch Records Dashboard**



KSC LC 39A has the most successful launches

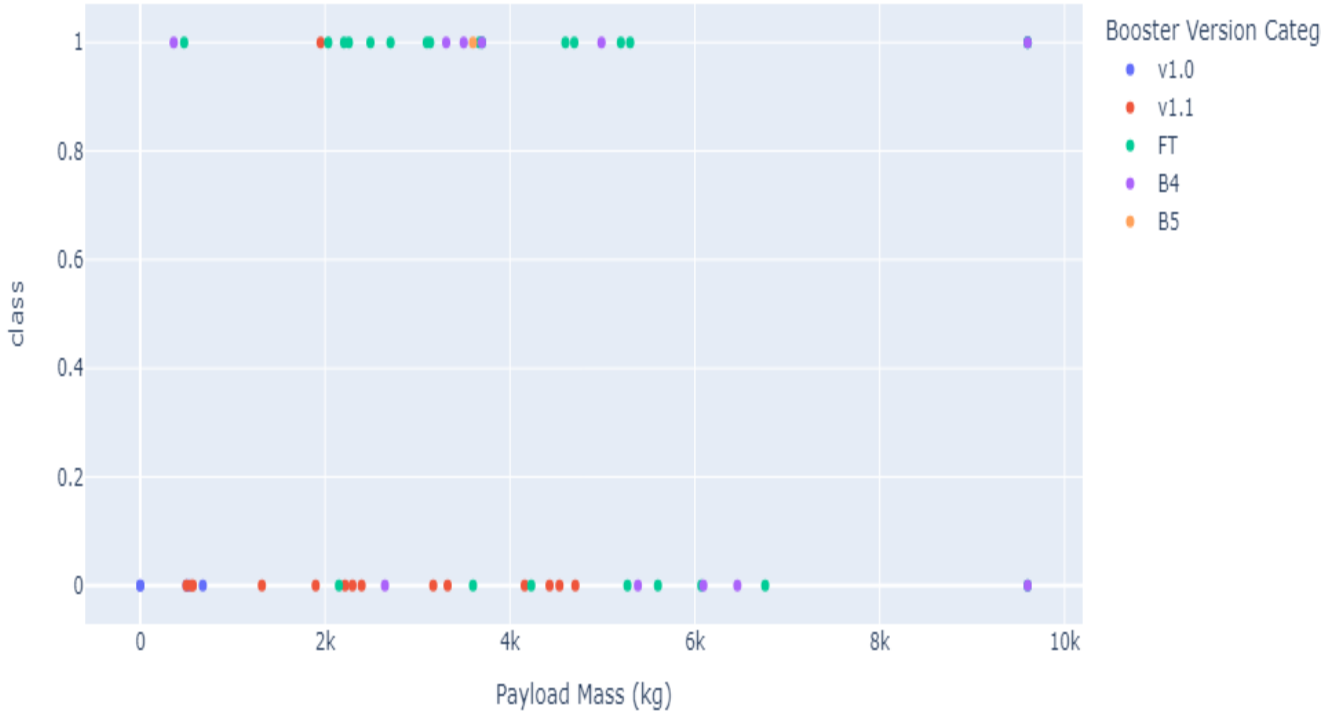
Success rate by site KSC LC 39A

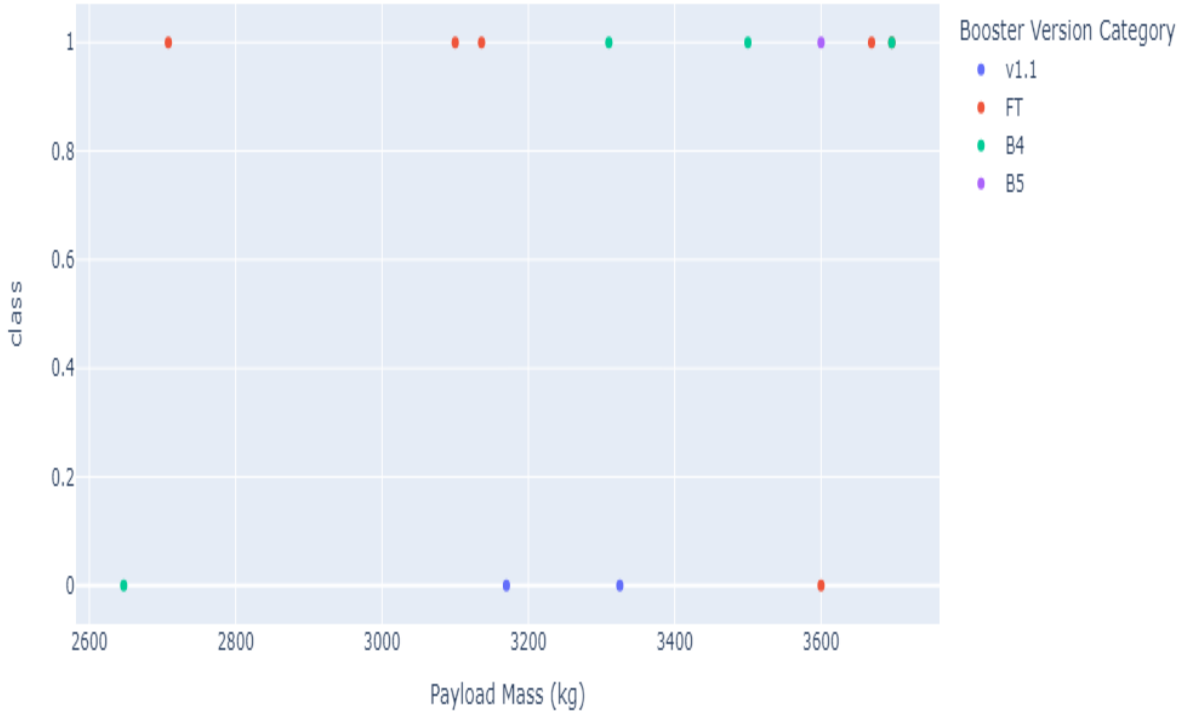


- KSC LC 39A achieved a 76.9% success rate getting a 23.1% failure rate

Payload vs. Launch Outcome scatter plot for all sites

- Booster Version FT has more success rate
- Payload Mass in range 2600 to 3800 has more success rate

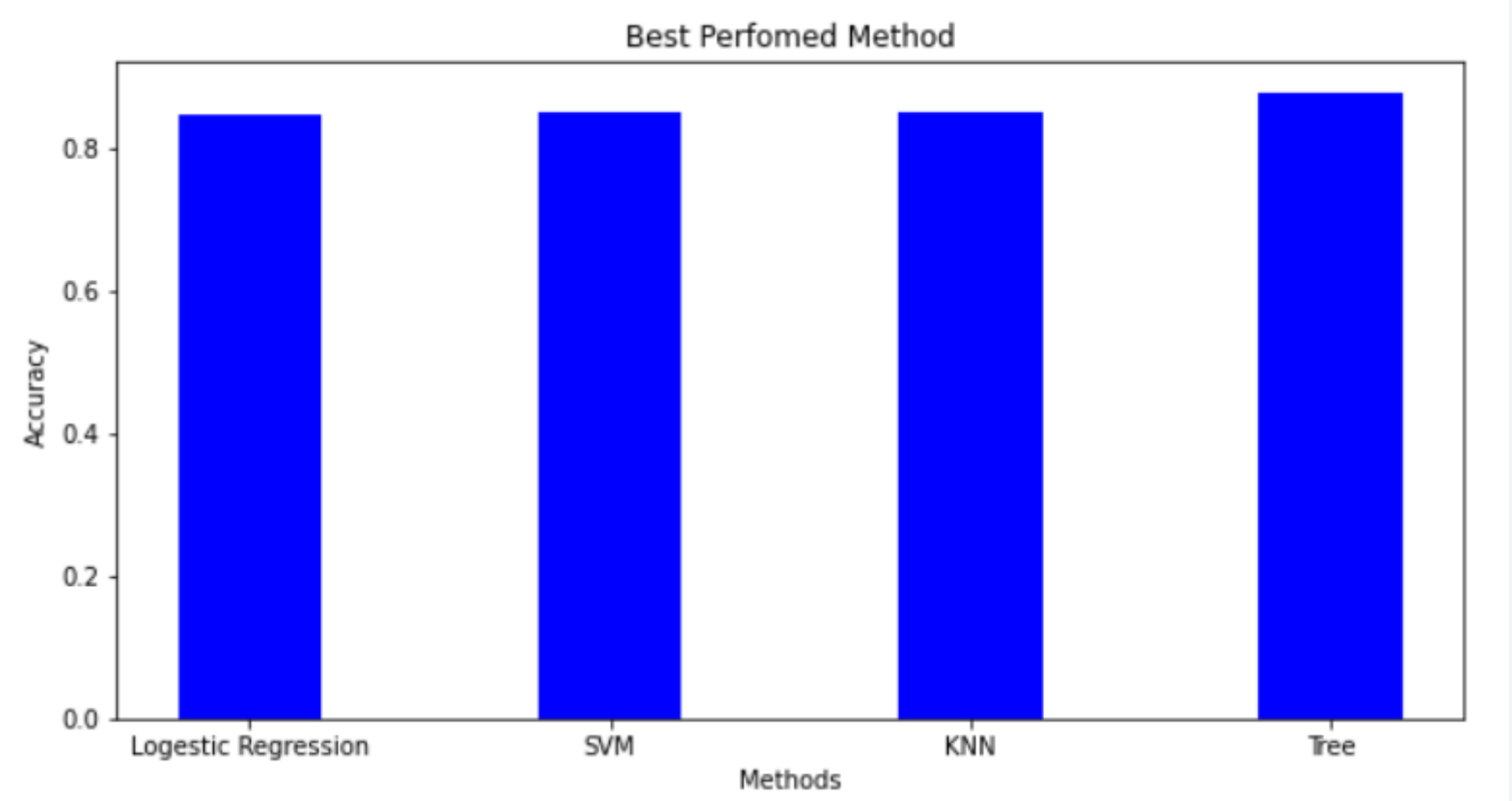




Section 5

Predictive Analysis (Classification)

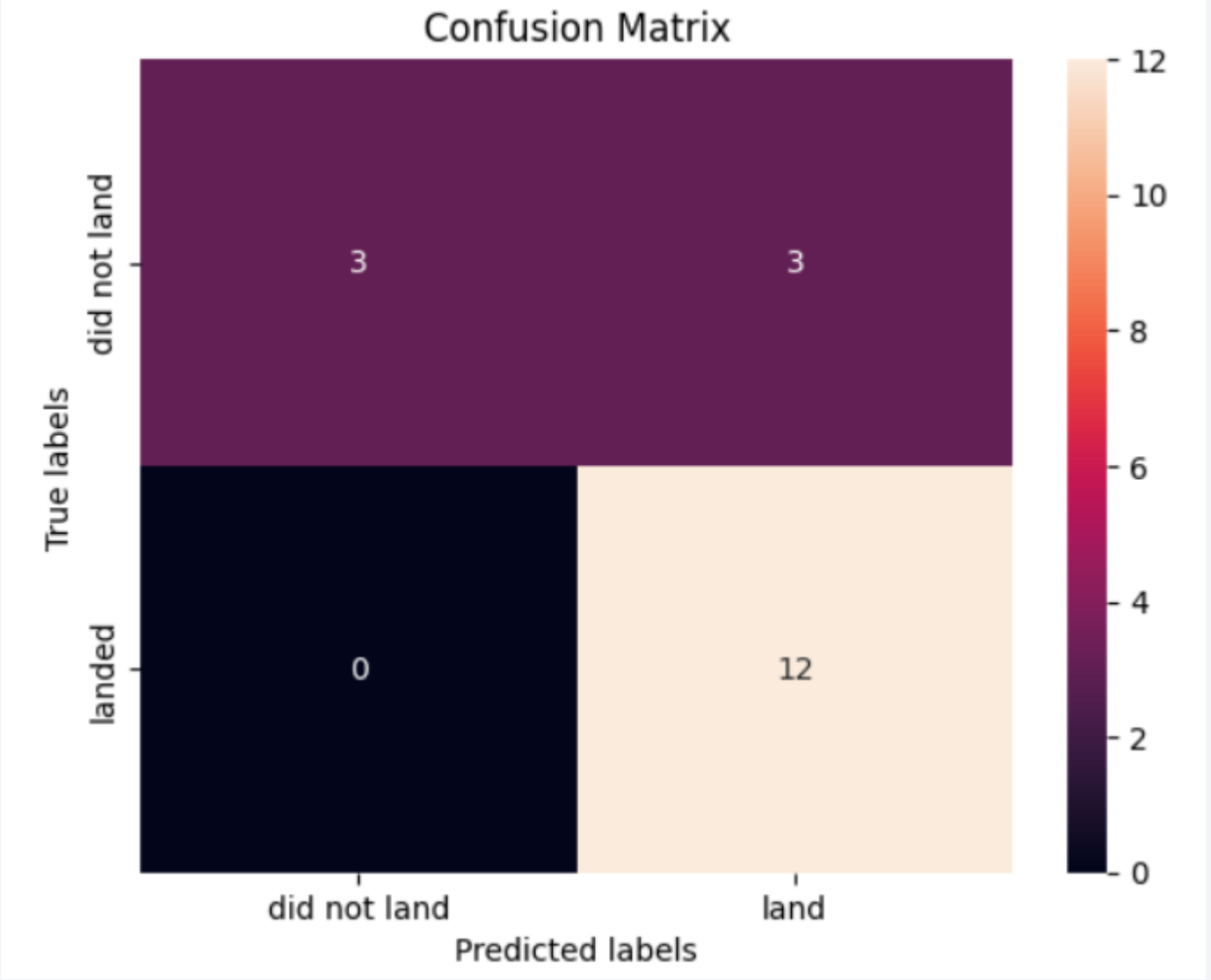
Classification Accuracy



Decision Tree model has best classification accuracy

Confusion Matrix

- Decision Tree is the best performing model
- Confusion Matrix results are the same for all four models



Conclusions

- First successful ground landing date is 2015-12-22
- All Models have the same accuracy and confusion matrix.
- Decision tree classifier is the best methods from logistic regression, decision tree classifier, k nearest neighbors and support vector machine, by 0.8767 accuracy.
- Lower weight payloads perform better than heavier weight payloads.
- The success rates for SpaceX launches are directly proportional time in years they will eventually perfect the launches.
- Launch site KSC LC 39A has the most successful launches from all the sites.
- The orbit types of ES-L1, GEO, HEO, and SSO are the highest success rate

Appendix

- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- https://api.spacexdata.com/v4/launches/past
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
- Hands-on Lab : String Patterns, Sorting and Grouping
- Hands-on Lab: Built-in functions
- Hands-on Lab : Sub-queries and Nested SELECT Statements
- Hands-on Tutorial: Accessing Databases with SQL magic
- Hands-on Lab: Analyzing a real World Data Set
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv

Thank you!