IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Guido Marinelli
22-04-2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- The era of commercial space has begun. SpaceX is the most successful company, that's because the launch of Falcon 9 rockets cost $62 million compared to other vendors whose costs are over $165 million. This success is determined by the reuse of the first stage of the rocket.
- Is it possible to use the launch data of a ScaceX rocket to develop an alternative offering?

Section 1

Methodology

Methodology

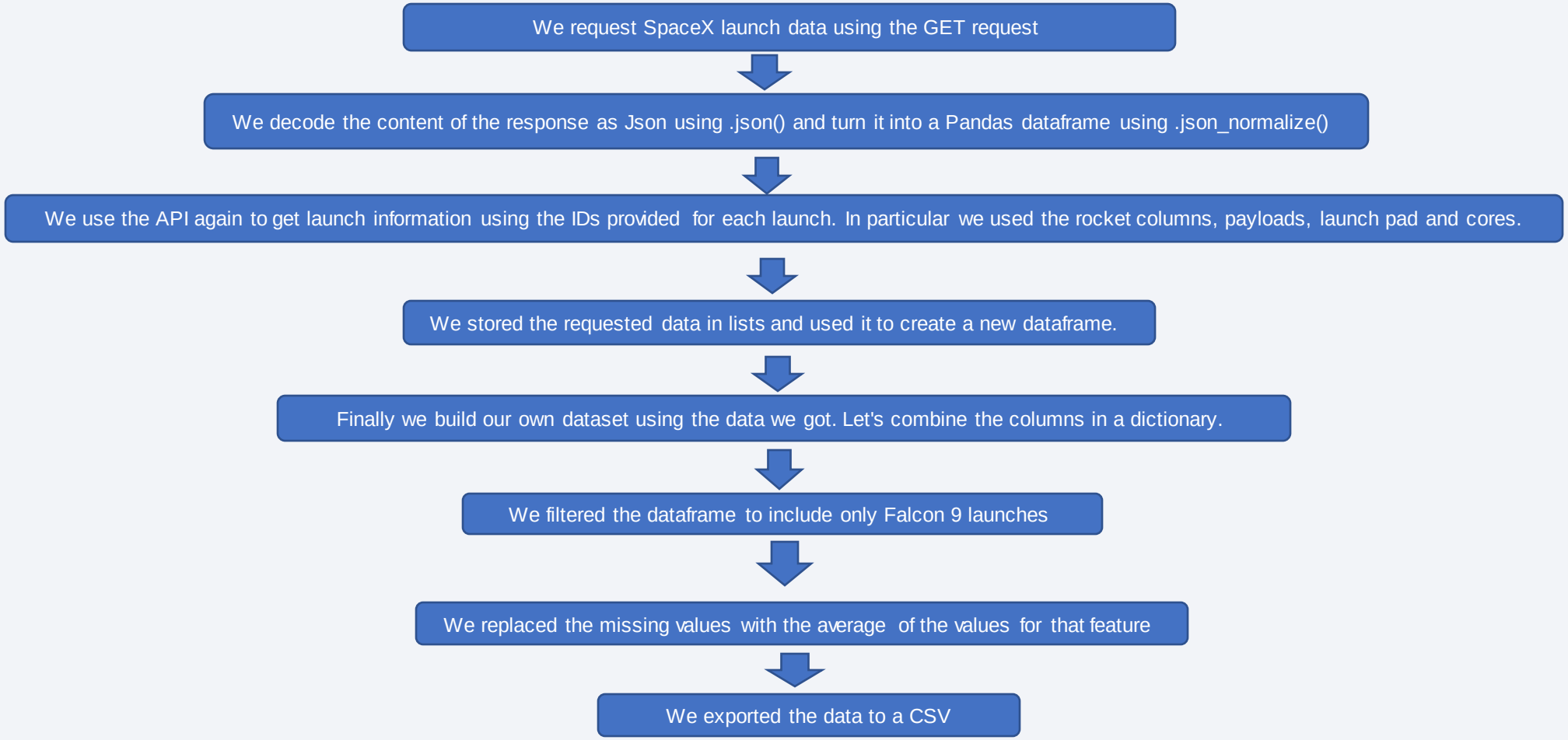
Executive Summary

- Data collection methodology:
  - Request to the SpaceX API
  - Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Perform data wrangling
  - The missing values in the PyloadMass column have been replaced with the average of the values present
  - Creating a landing_class column to transform the categorical values in the Outcome column into the 0 bad outcome and 1 success outcome classes
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using the SVM, Classification Trees and Logistic Regression models in the Sciiti-learn library, tuning the best hyperparameter with the GridSearchCV function and evaluating them with the score function

Data Collection

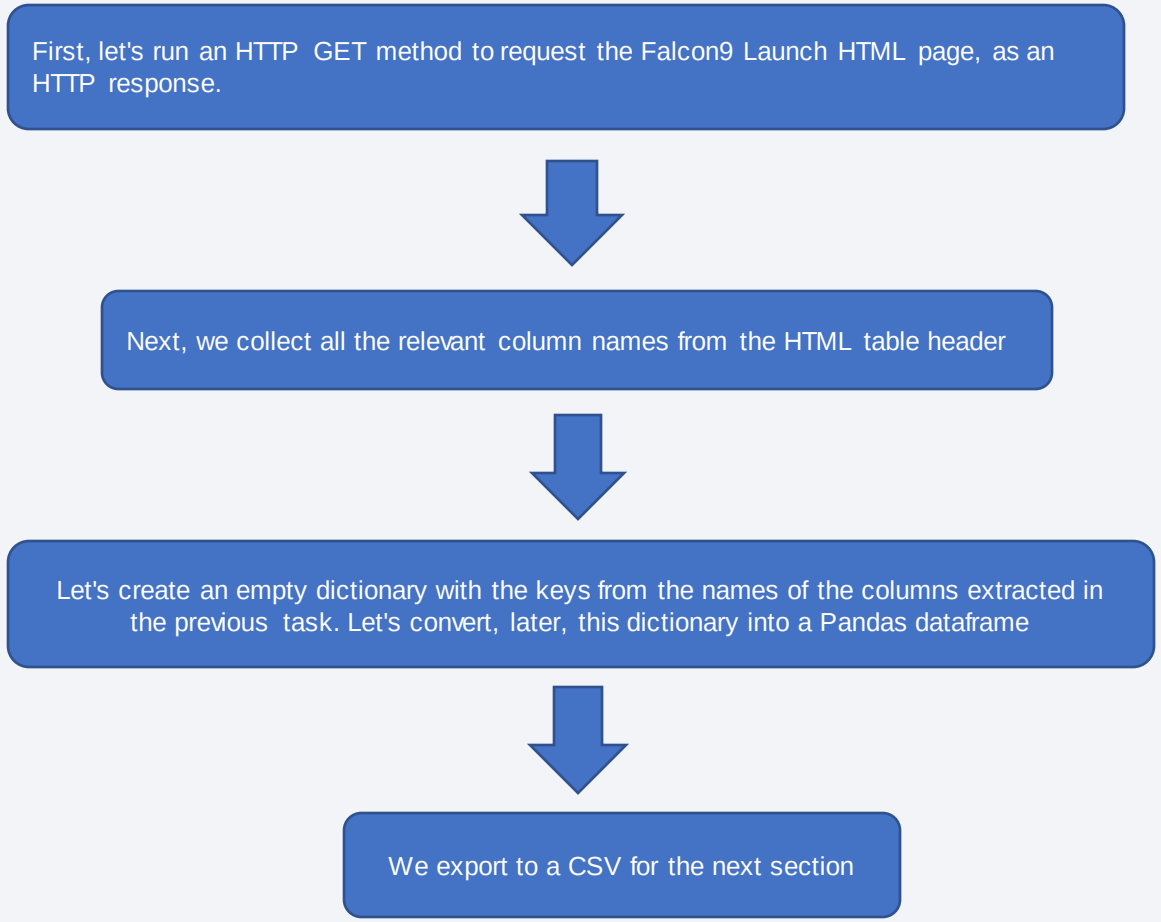
- The data was collected through the SpaceX API and from the Wikipedia page on List of Falcon 9 and Falcon Heavy launches

Data Collection – SpaceX API



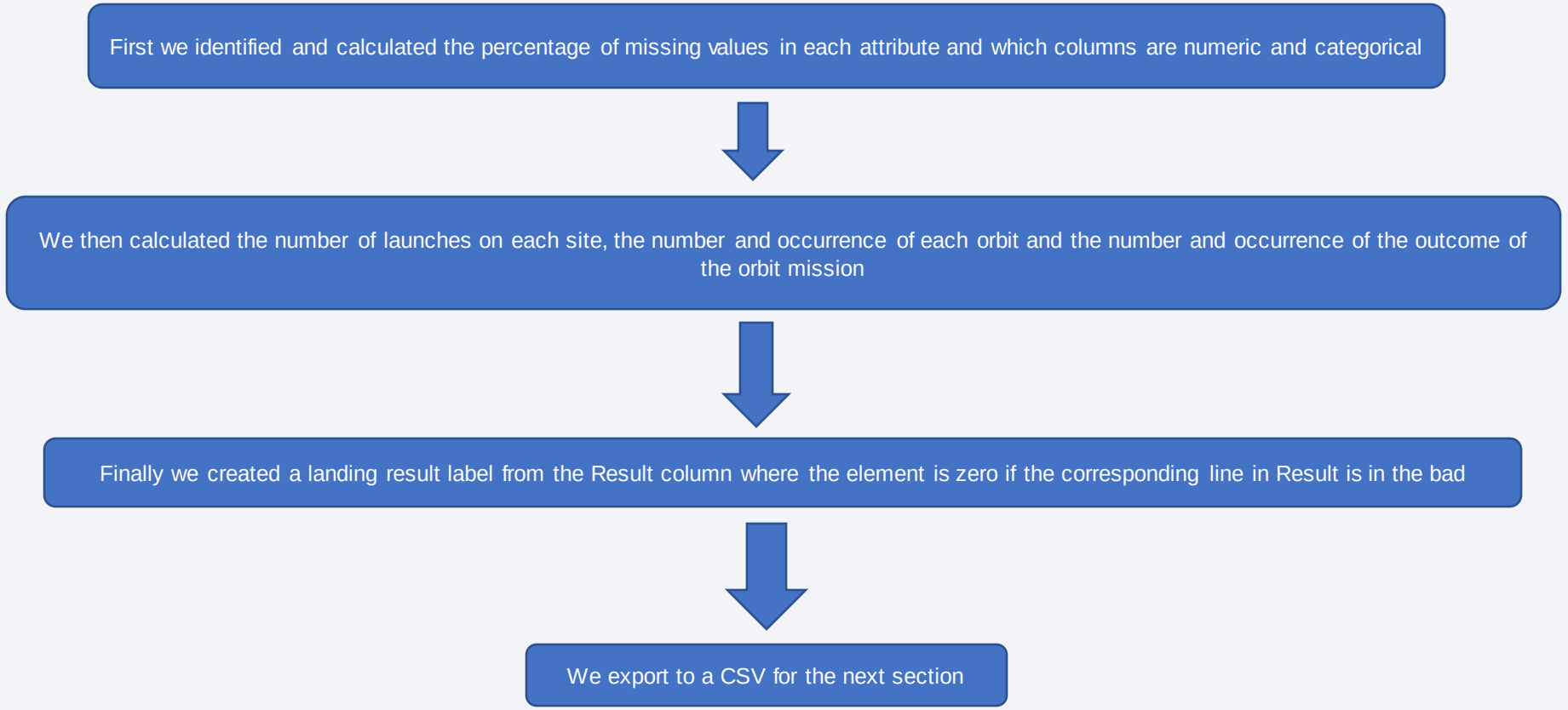
https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter point chart to view the relationship between Flight Number and Launch Site
- Scatter point chart to view the relationship between Payload and Launch Site
- Bar chart to view the relationship between the success rate of each type of orbit
- Scatter point chart to view the relationship between FlightNumber and Orbit type
- Line chart to view the annual trend of launch success

https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- The names of unique launch sites in the space mission
- 5 records where launch sites begin with the string "CCA"
- The total payload mass carried by NASA-launched boosters (CRS)
- The average payload mass carried by the F9 booster version v1.1
- The date the first successful landing result in the ground pad was achieved
- The names of the boosters that are successful in the drone ship and have a payload mass greater than 4000 but less than 6000
- The total number of successful and failed mission results
- The names of the booster_versions that carried the maximum payload mass.
- The records that will display the names of the months, the failure landing_outcomes in the drone ship, the booster versions, lunch_site for the months of the year 2015
- The count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Types of objects created and added to the folium map:
  - folium.Circle
  - folium.Marker
  - folium.PolyLine

- These objects were created for:
  - Mark all launch sites on the map
  - Mark successful/failed launches for each site on the map
  - Show the distance between a launch site and its proximity to the railway, highway, coastline, city

https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Graphs and interactions added to the dashboardpie chart
  - scatter chart
  - dropdown

- These plots and interaction were inserted to show intuitively
  - Which site has the largest successful launches?
  - Which site has the highest launch success rate?
  - Which payload range(s) has the highest launch success rate?
  - Which payload range(s) has the lowest launch success rate?
  - Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest

https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/spacex_dash_app.py

Predictive Analysis (Classification)

- Using the SVM, Classification Trees and Logistic Regression models in the Sciiti-learn library, tuning the best hyperparameter with the GridSearchCV function and evaluating them with the score function

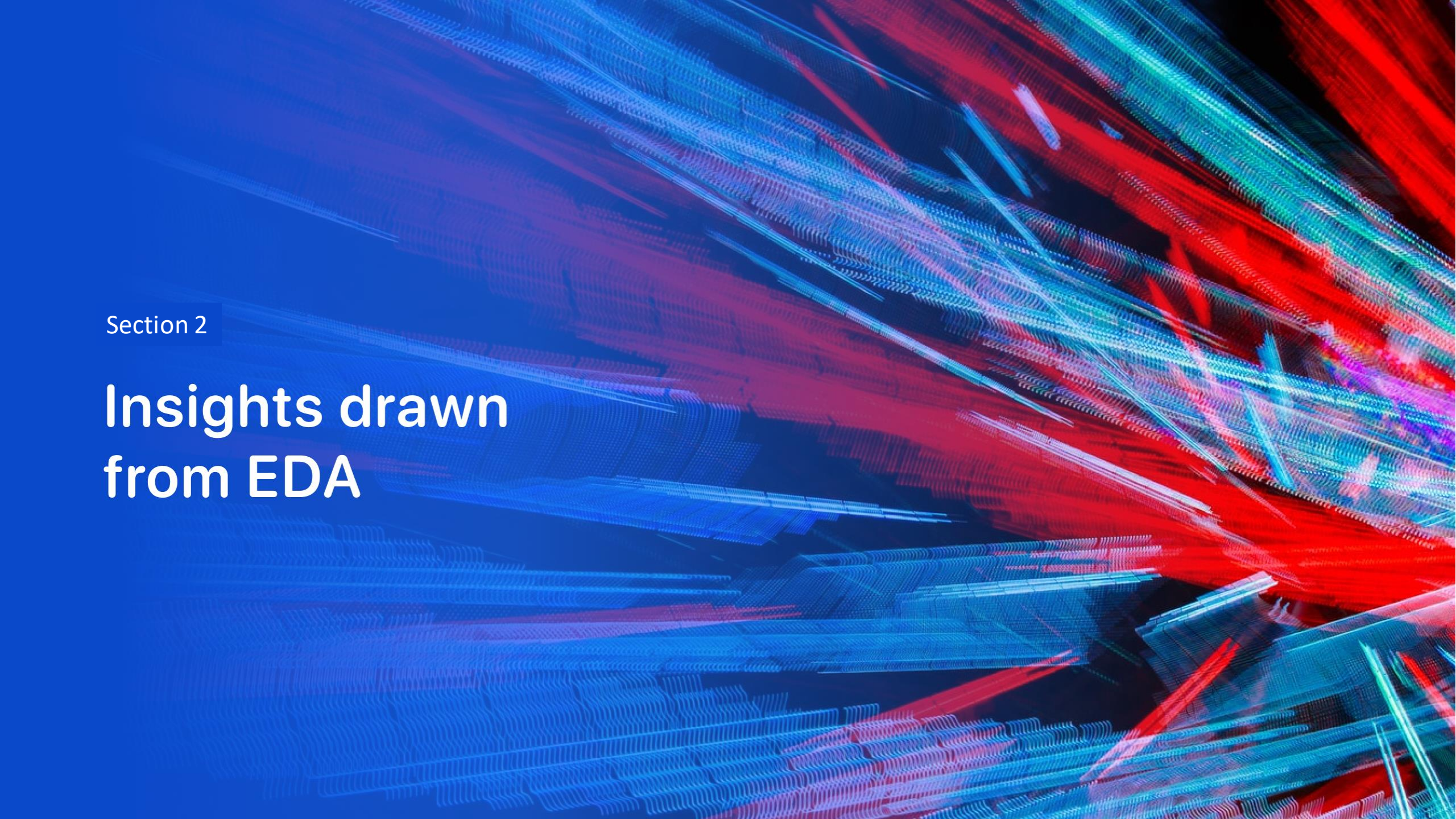
- https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

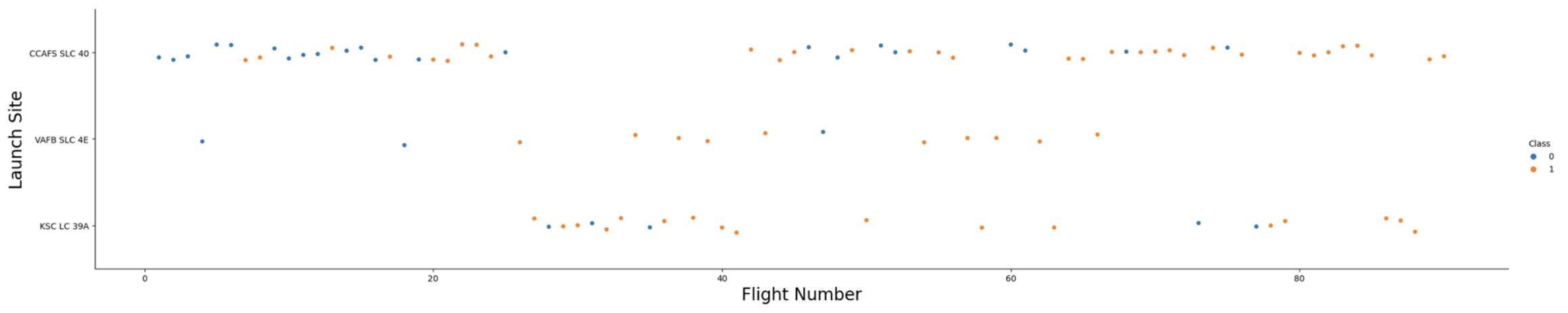
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

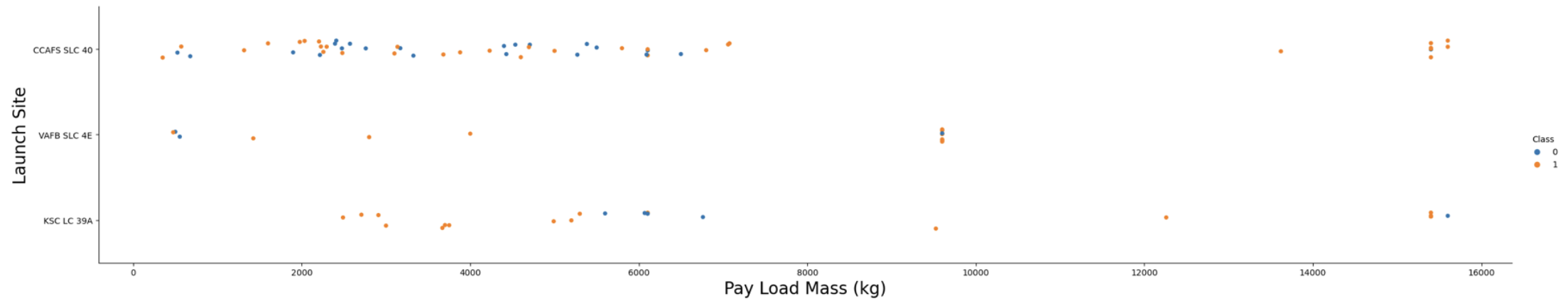


Flight Number vs. Launch Site



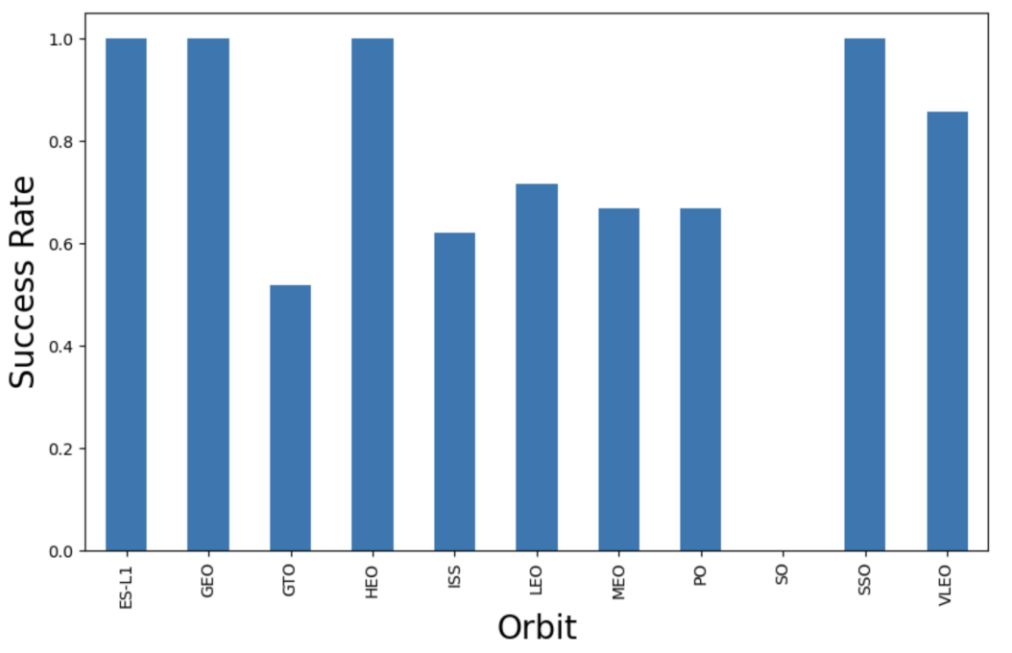
We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Payload vs. Launch Site



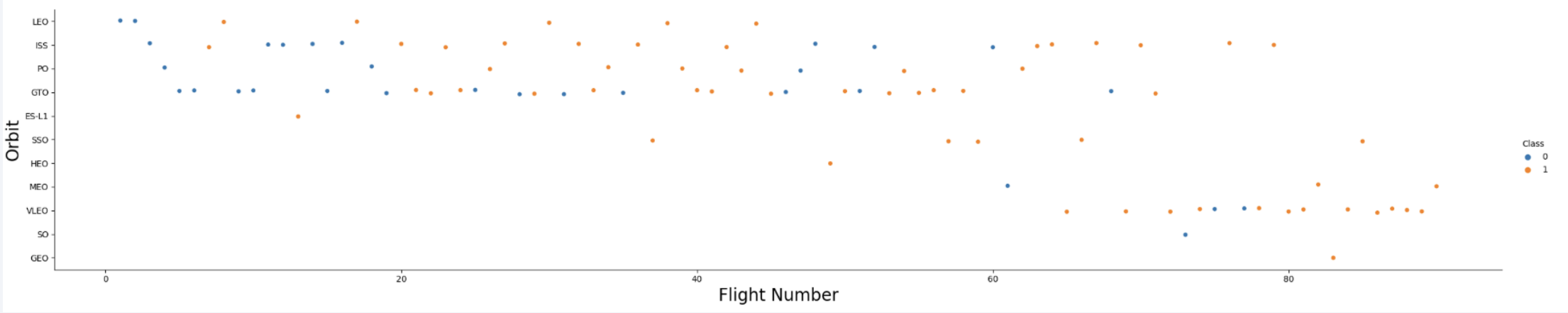
We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

Success Rate vs. Orbit Type



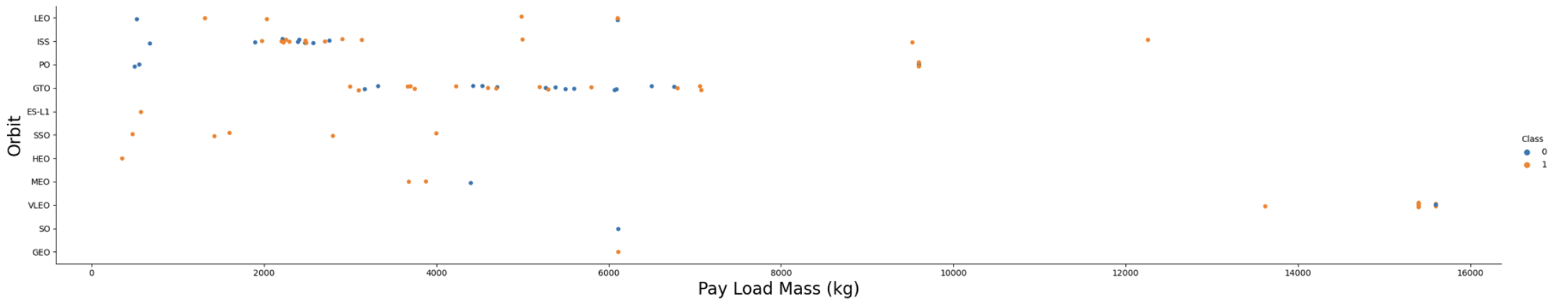
We note that for the ES-L1, GEO, HEO and SSO orbits the Success Rate is 100% while for the SO orbit it is 0%.

Flight Number vs. Orbit Type



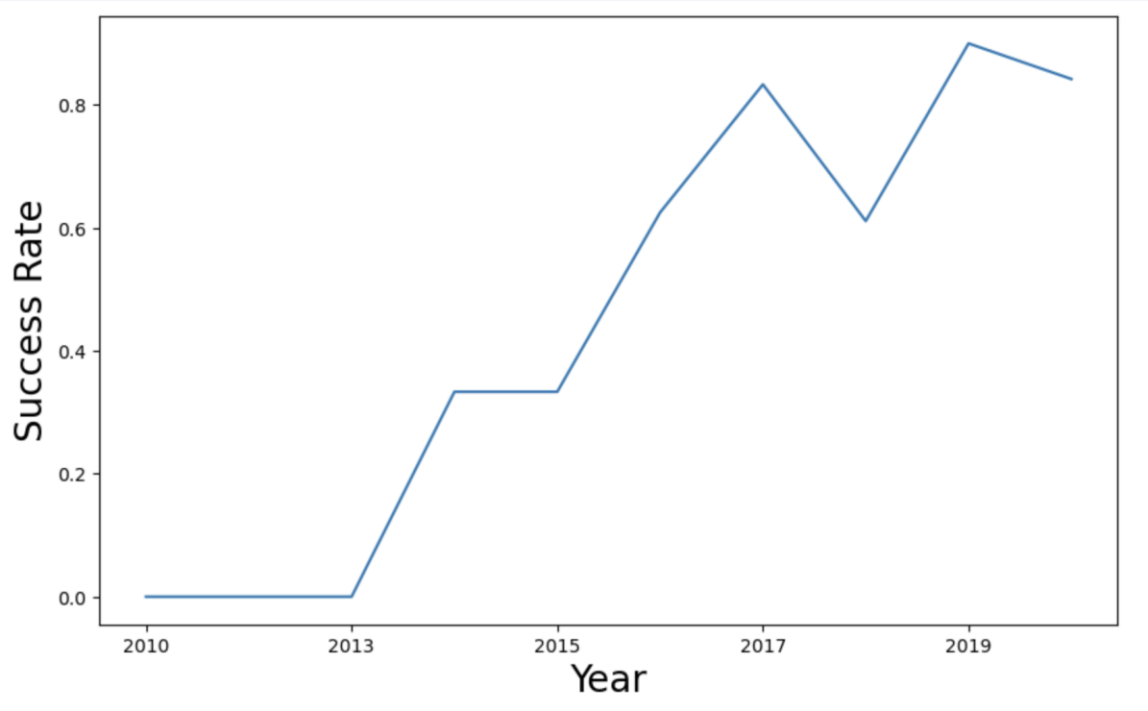
We note that in the LEO orbit the Success appears to be related to the number of flights; on the other hand, there does not seem to be any relationship between the flight number when it is in GTO orbit.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

Launch Success Yearly Trend



Let's note how the success rate since 2013 has continued to increase until 2020

All Launch Site Names

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

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

Launch Site Names Begin with 'CCA'

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
48213

Average Payload Mass by F9 v1.1

AVG_PAYLOAD_MASS__KG_
2534.6666666666665

First Successful Ground Landing Date

Date	Landing _Outcome
22-12-2015	Success (ground pad)

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

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Mission_Outcome_Count
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

2015 Launch Records

Month	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

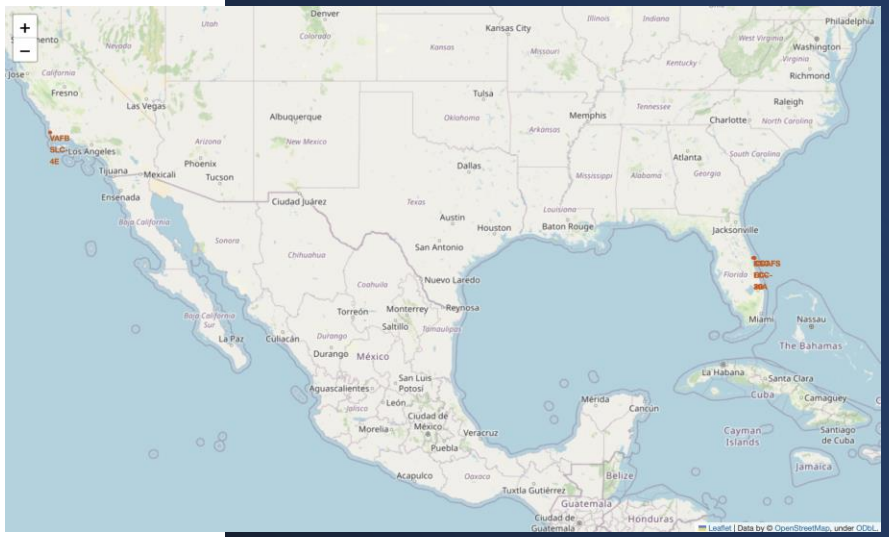
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

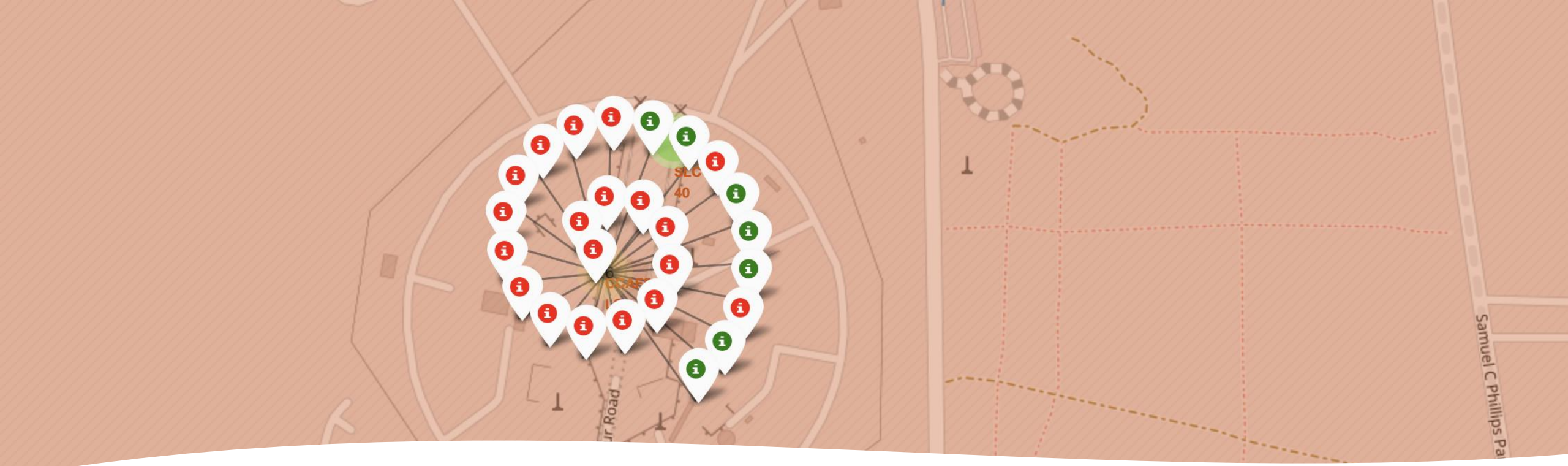
Landing _Outcome	Landing_Outcome_Count
Success	20
Success (drone ship)	8
Success (ground pad)	6


Section 3
Launch Sites
Proximities Analysis

All Launch Sites on Map

- We note that all the launch sites are close to the Equator line and are very close to the coast

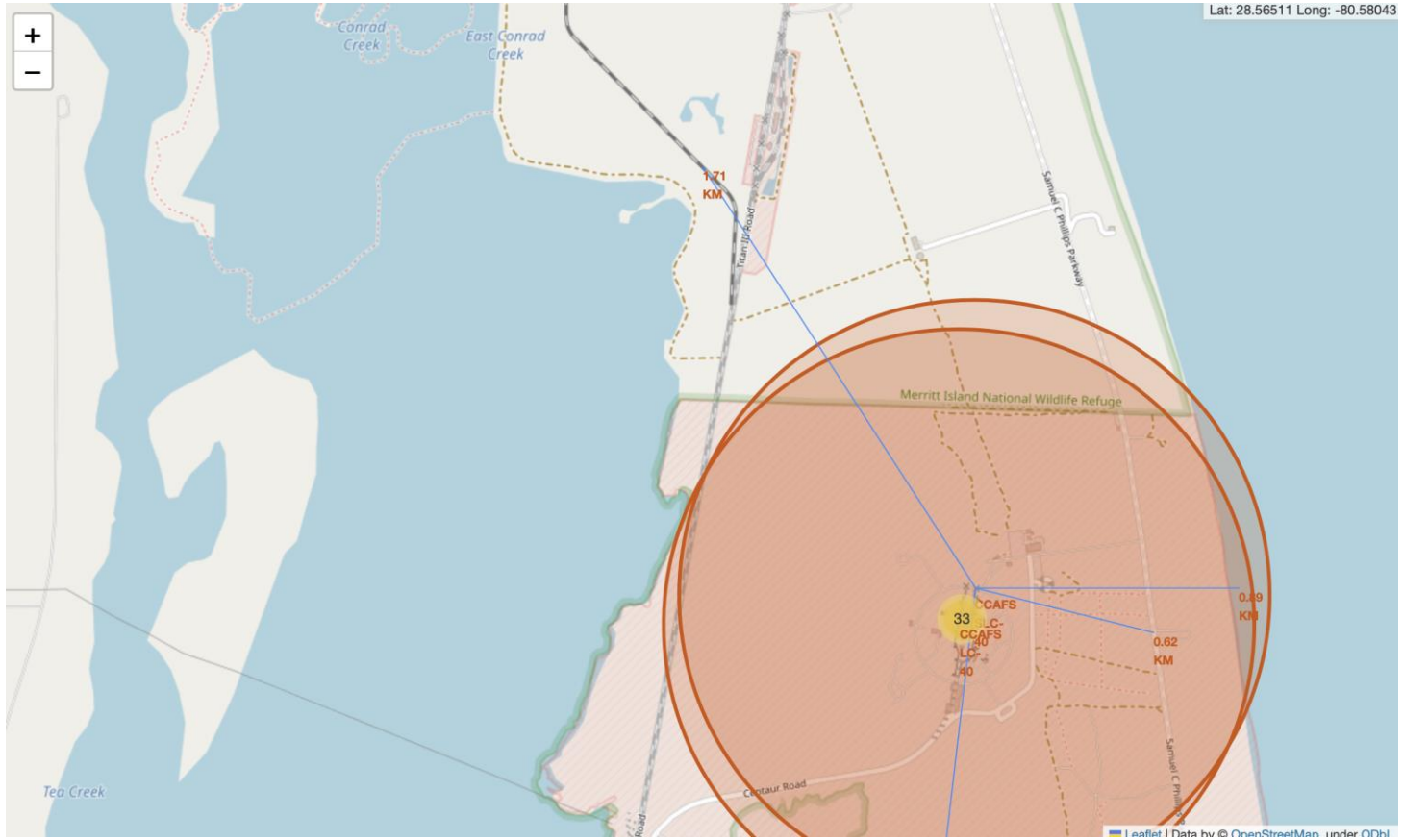


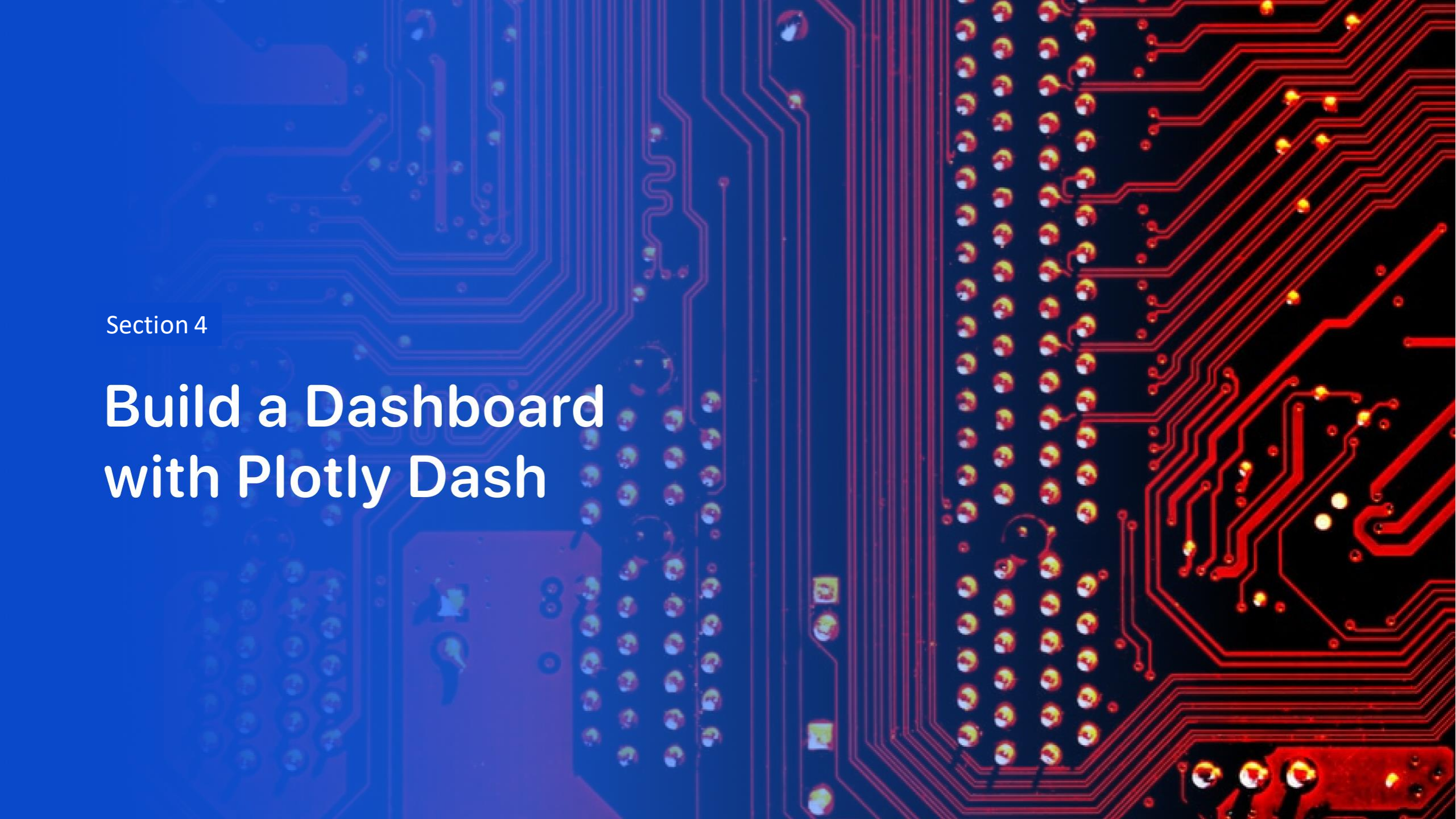


The Success/Failed Launches for each Site on the Map

- From color-labeled markers in marker clusters we are able to easily identify which launch sites have relatively high success rates.

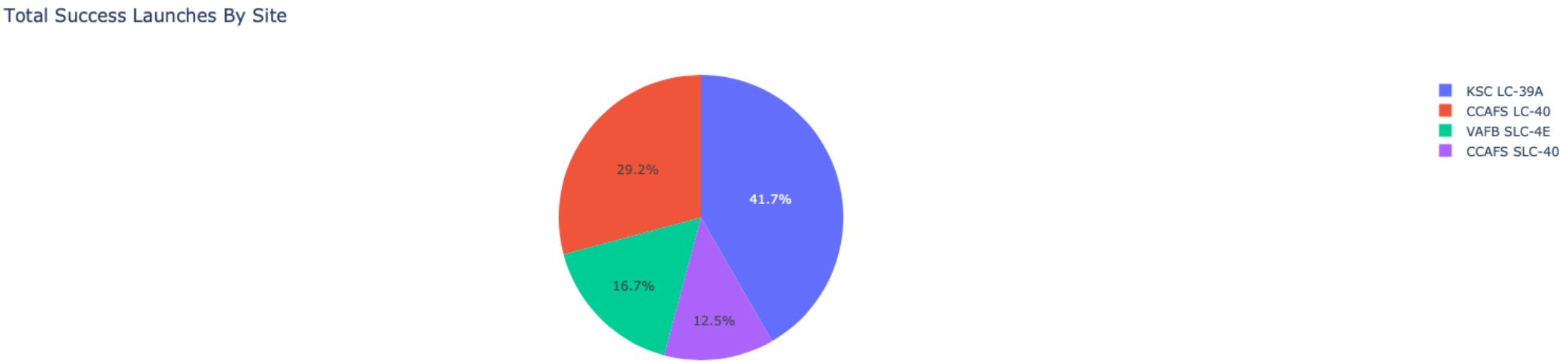
Distance between a Launch Site to its Proximities



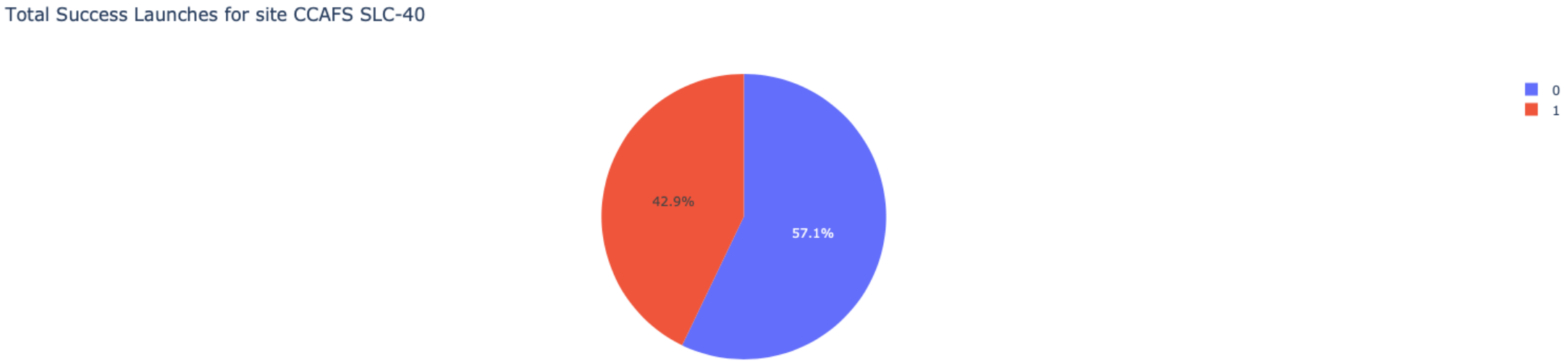

Section 4
Build a Dashboard with Plotly Dash

Piechart for the Launch Success Count for All Sites

- We note that the most successful launch site is the KSC LC-39A

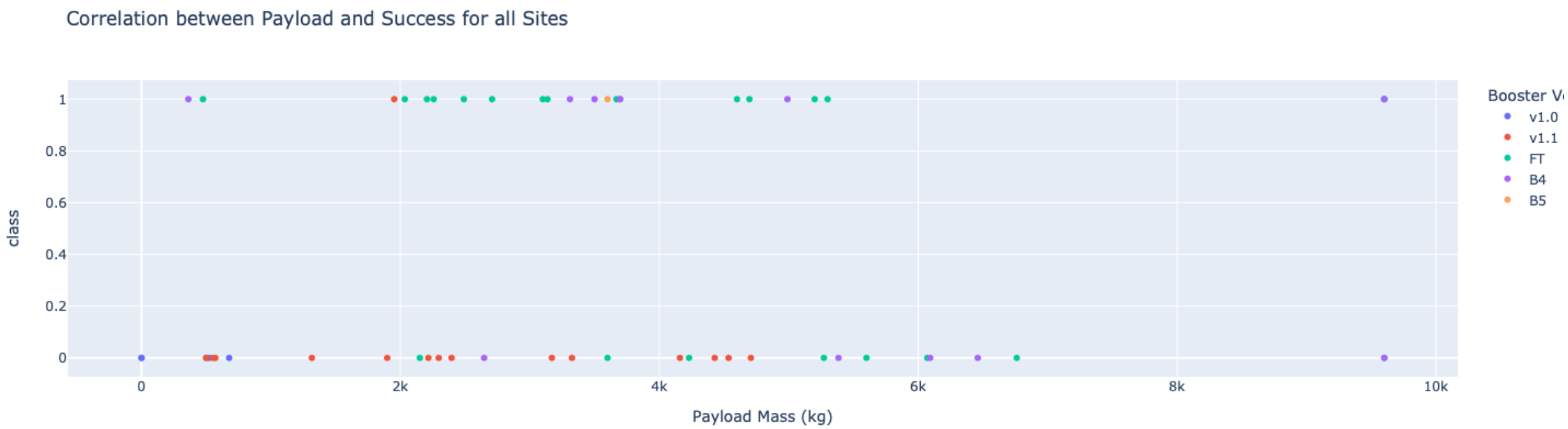


Piechart for the Launch Site with Highest Launch Succces



Correlation between Payload and Success for all Sites

- It is noted that FT Boosters with Payload Mass between 2k and 6k have the highest success rate

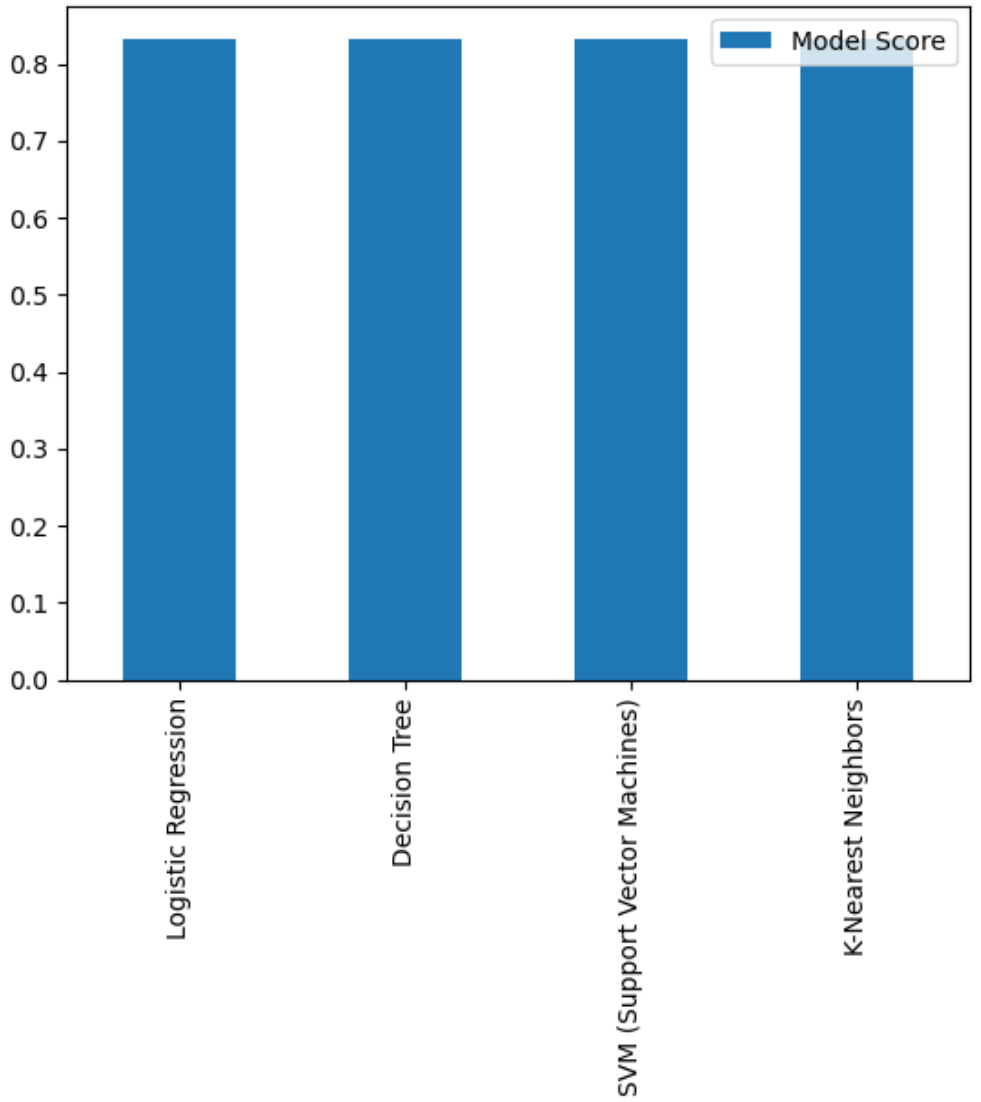


Section 5

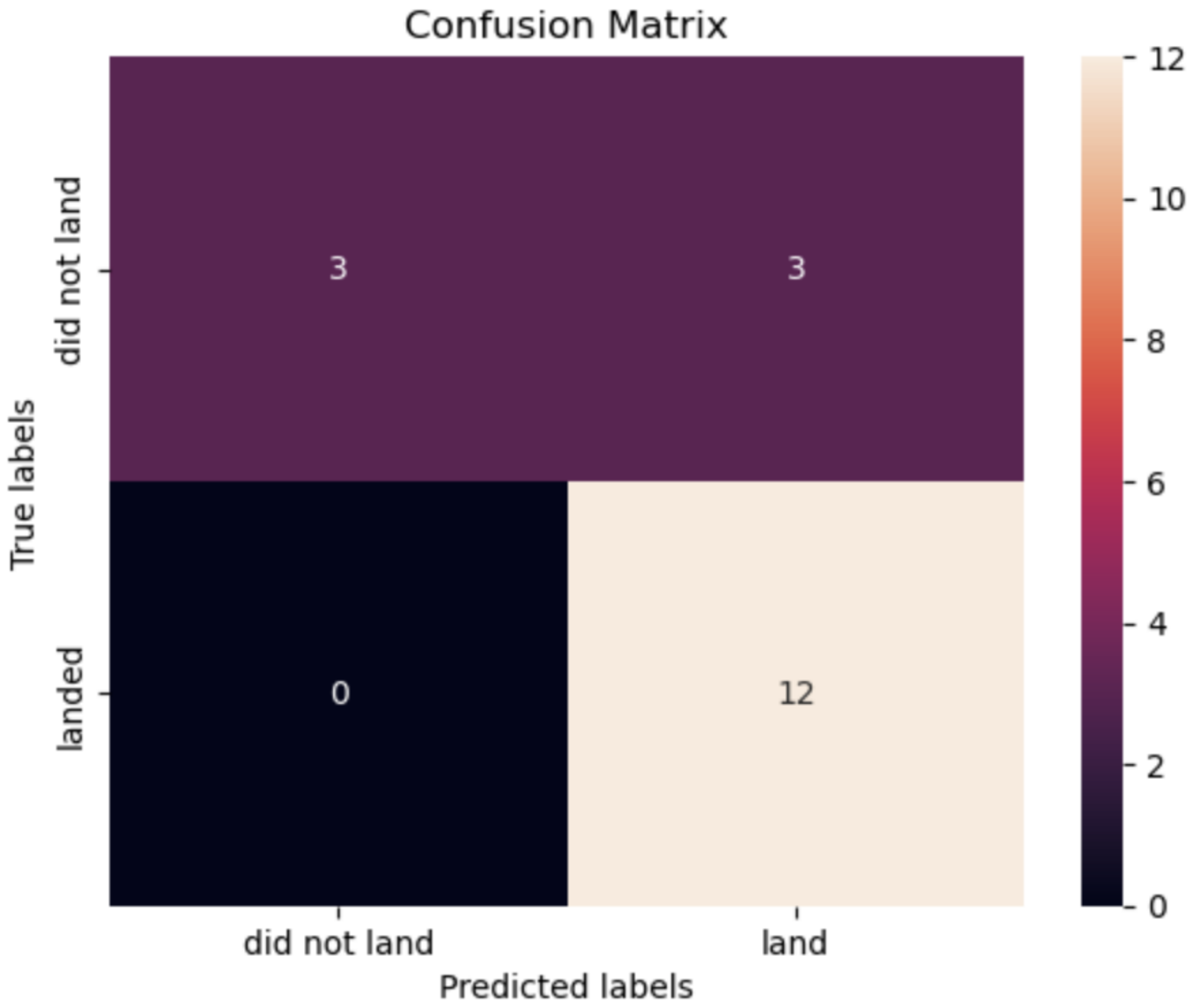
Predictive Analysis (Classification)

Classification Accuracy

- Each models have the same accuracy:



Confusion Matrix



Conclusions

- To our initial question “Is it possible to use the launch data of a ScaceX rocket to develop an alternative offering?” We can answer with a yes!
- Choose the KSC LC-39A and VAFB SLC 4E sites that are those sites with the highest success
- To minimize failures, especially initial failures, reduce the payload in order to increase the chances of the first stage of the rocket returning
- Aim for ES-L1, GEO, HE and SSO orbits

Appendix

https://github.com/GuidoMarinelli/IBM-Data-Science-Professional-Certificate/tree/main/SpaceX%20%20Falcon%209%20first%20stage%20Landing%20Prediction%20Project

Thank you!