IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rodrigo Antúnez
17/6/2024

Outline

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

Executive Summary

- In the follow capstone project we will use some machine learning algorithms to predict if the SpaceX Falcon 9 can successfully land.
- The main steps followed in this project include:
  - Data collection, wrangling and formatting.
  - Exploratory data analysis.
  - Interactive data visualization.
  - Machine learning prediction.
- We can summarize that our graphics show that exists a correlation in the features of the rocket launches and the outcome of the launches.

Introduction

- The main idea of the project is to predict if the SpaceX Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- The main question that we are trying to answer is, through a set of features of the SpaceX Falcon 9 rocket which include it payload mass, orbit type, launch site, landing pad, etc. ¿The first stage of the Falcon 9 land will be successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collection via a request to the SpaceX API - Web Scraping.
- Perform data wrangling
  - Data wrangling and elimination of missing values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune and evaluation of classification models

Annotation

The URLs to each notebook are found as hyperlinks on their corresponding slide. If you have any problems accessing them, you can find in the appendix the URL to access the repository where all the notebooks are located.

Data Collection

- Data collection process involved a combination of API request from the SpaceX API and web scraping data from a SpaceX Wikipedia page.
- The next slide will show more about the sequence of processing the data from SpaceX API and the sequence of processing the data from web scraping.
- Also, the slides will include some pictures for a visualization of the results.

Data Collection – SpaceX API

- The API used is the following https://api.spacexdata.com/v4/rockets/
- The API provides data about different types of rocket launches done by SpaceX. Therefore the data is filtered to only include Falcon 9 launches.
- Missing values in the data were treated.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Notebook on GitHub

Data Collection - Scraping

- The data was scraped from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- The website contains data only about the Falcon 9 launches.

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	.mw-parser-output .plainlist ol,.mw-parser-out...	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Notebook on GitHub

Data Wrangling

- The data is processed to ensure that are no missing entries and categorical features are encoded using one-hot encoding.
- An extra column called “Class” is also added to the data frame. The column contains 0 if a given launch is failed and 1 if the launch is successful.

Notebook on GitHub

EDA with Data Visualization

- Were performed exploratory data analysis on variables flight number, payload mass, launch site, orbit, class and year.
- Scatter plots, line charts, and bar plots were used to compare the relationship between the different variables to decide if exist some relationship so that they could be used in the training of the machine learning model.

Notebook on GitHub

EDA with SQL

- The SQL queries performed are the following:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.
  - Rank 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.

Notebook on GitHub

Build an Interactive Map with Folium

- Folium maps mark Launch sites, successful and unsuccessful landings and a proximity example to key locations how can be the railway, highway, coast, city and so on.
- This allow us to understand where launch sites may be located and why. Also this give us a visualization of the successful landings relative to the location.

Notebook on GitHub

Build a Dashboard with Plotly Dash

- Dashboard with Plotly contains a pie chart and a scatterplot which allow us to visualize the total success launches from each launch site and the correlation between payload mass and mission outcome for each launch site.
- Those plots are utilized to generate an interactive site where we can toggle the input using a dropdown menu and a range slider.

Python file on GitHub

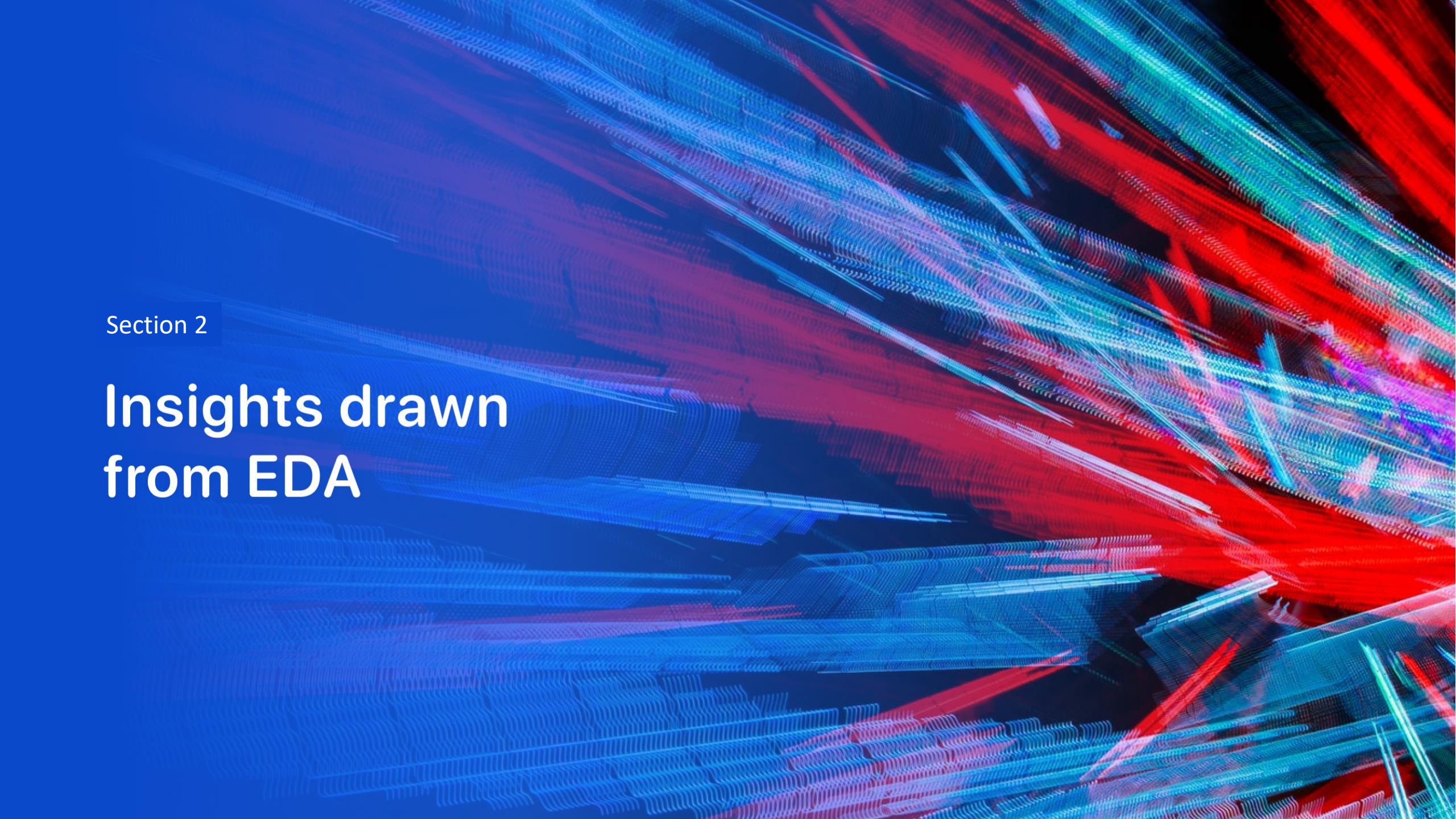
Predictive Analysis (Classification)

- The machine learning prediction phase was carried out taking into account the following steps:
  - Data standardization
  - Data split (train – test)
  - Creation of machine learning model:
    - LR (Logistic Regression)
    - SVM (Support Vector Machine)
    - Tree (Decision Tree Classifier)
    - KNN (K Nearest Neighbors)
  - Model fit with train data
  - Search of best combination for each model
  - Evaluation of the models

Notebook on GitHub

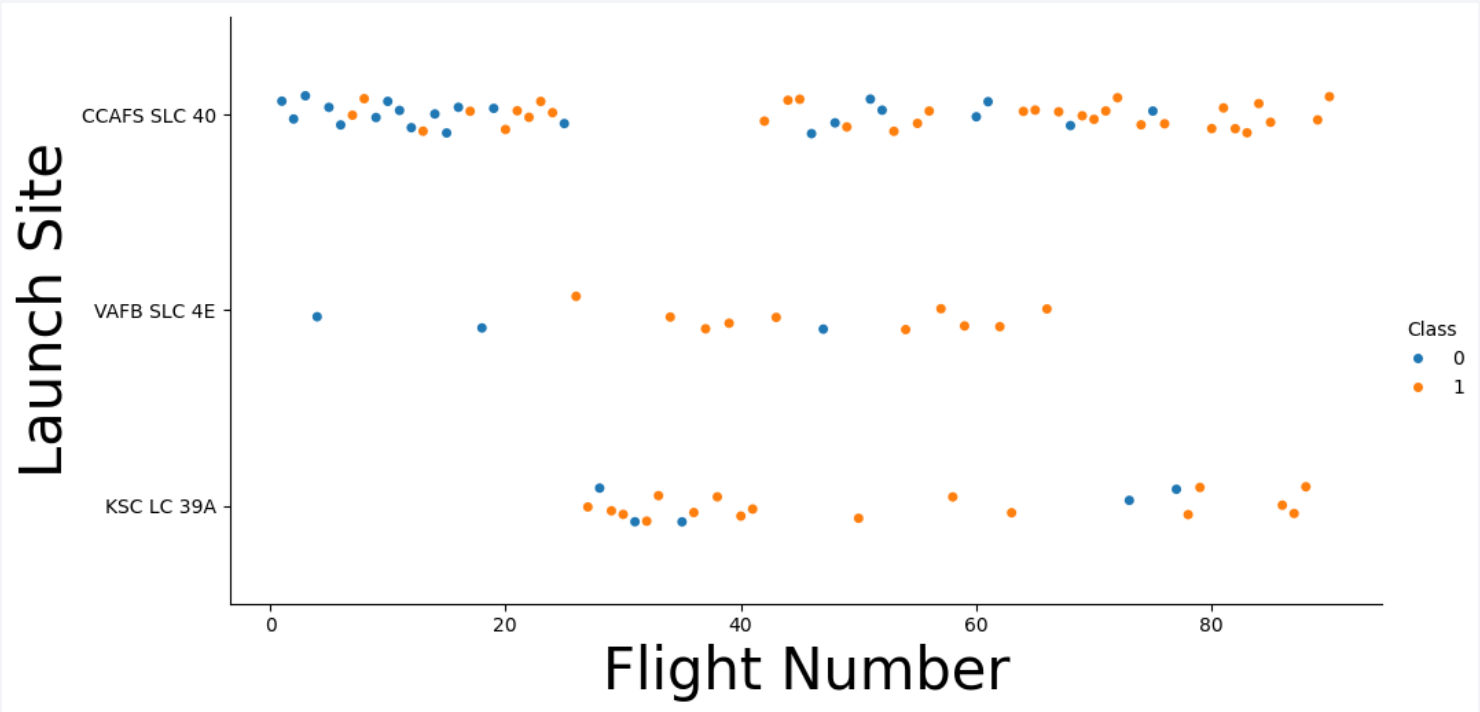
Results

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


Section 2
Insights drawn from EDA

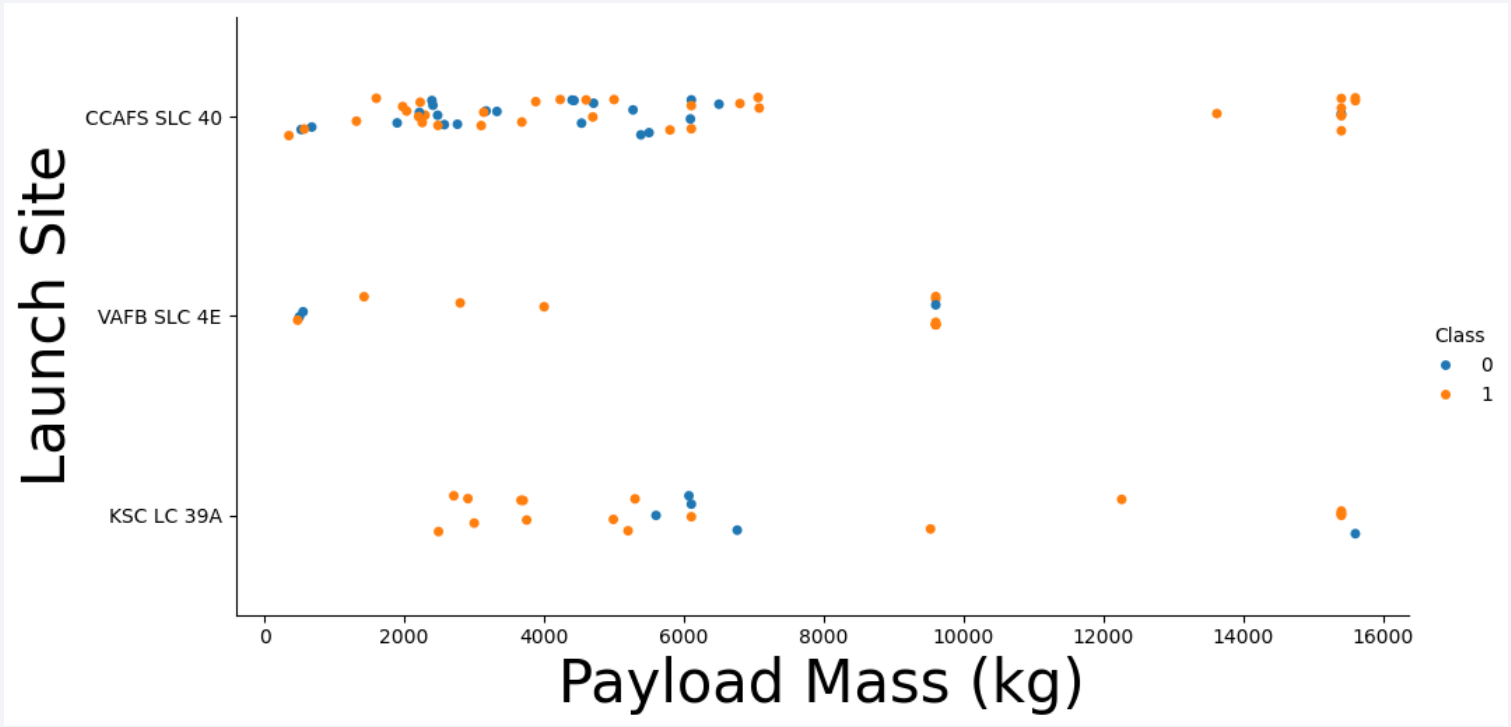
Flight Number vs. Launch Site

The scatter plot suggest that CCAFS SLC 40 appears to be the main launch site according to the volume of launch.



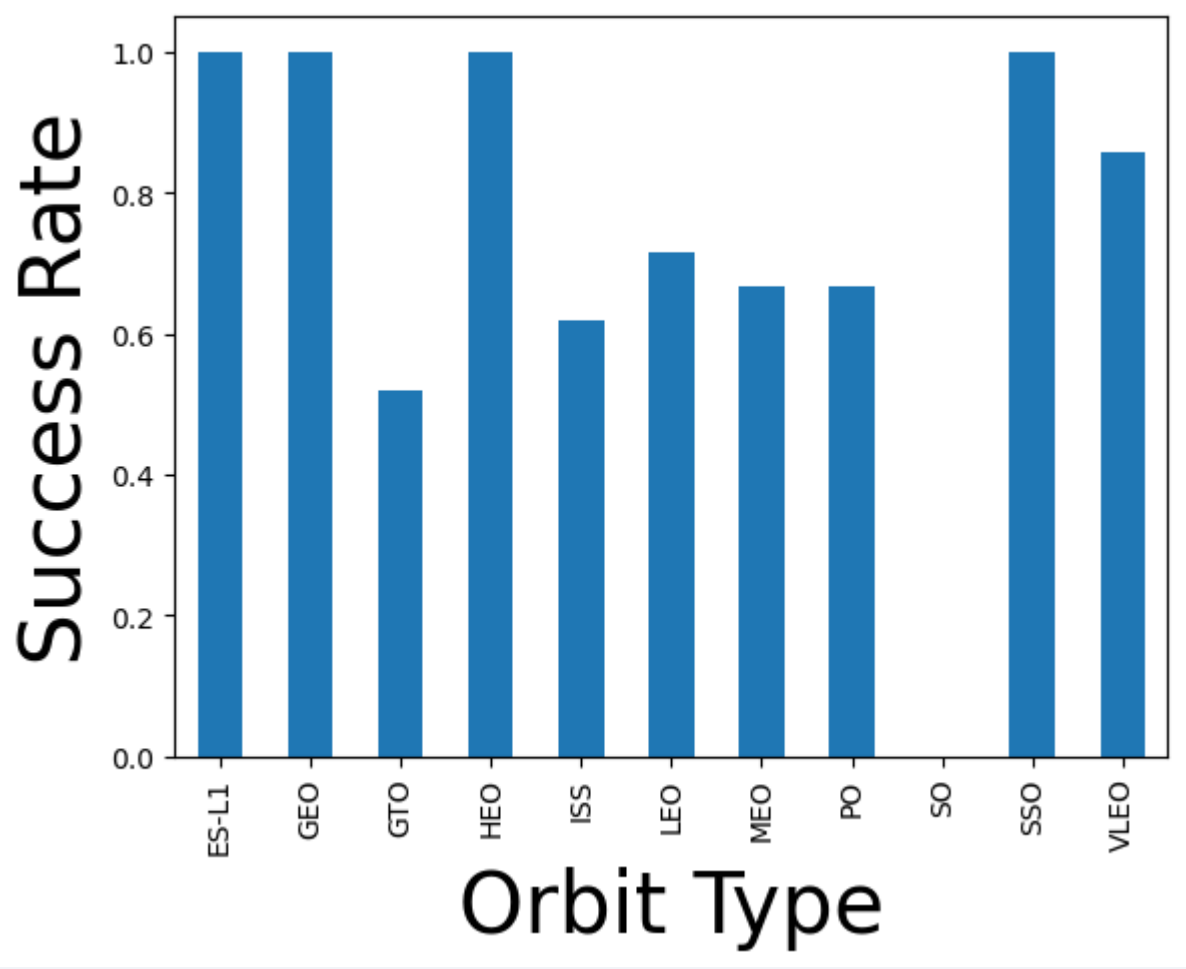
Payload vs. Launch Site

The scatter plot suggests that for the VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000).



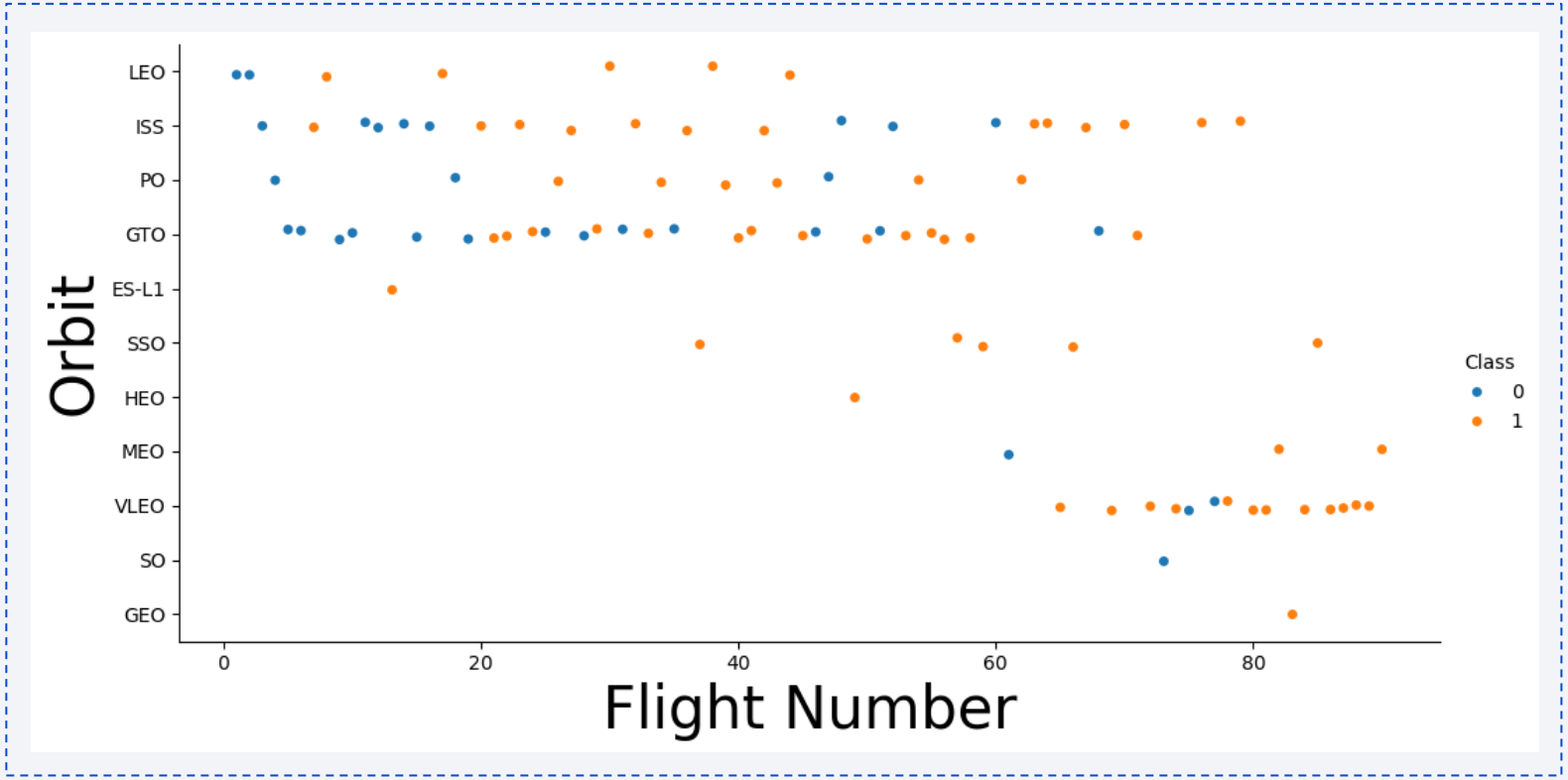
Success Rate vs. Orbit Type

From the scatter plot we can address that ES-L1, GEO, HEO and SSO orbit have a successful rate of the 100%. Also, is possible to address that the orbit SO have a 0% successful rate.



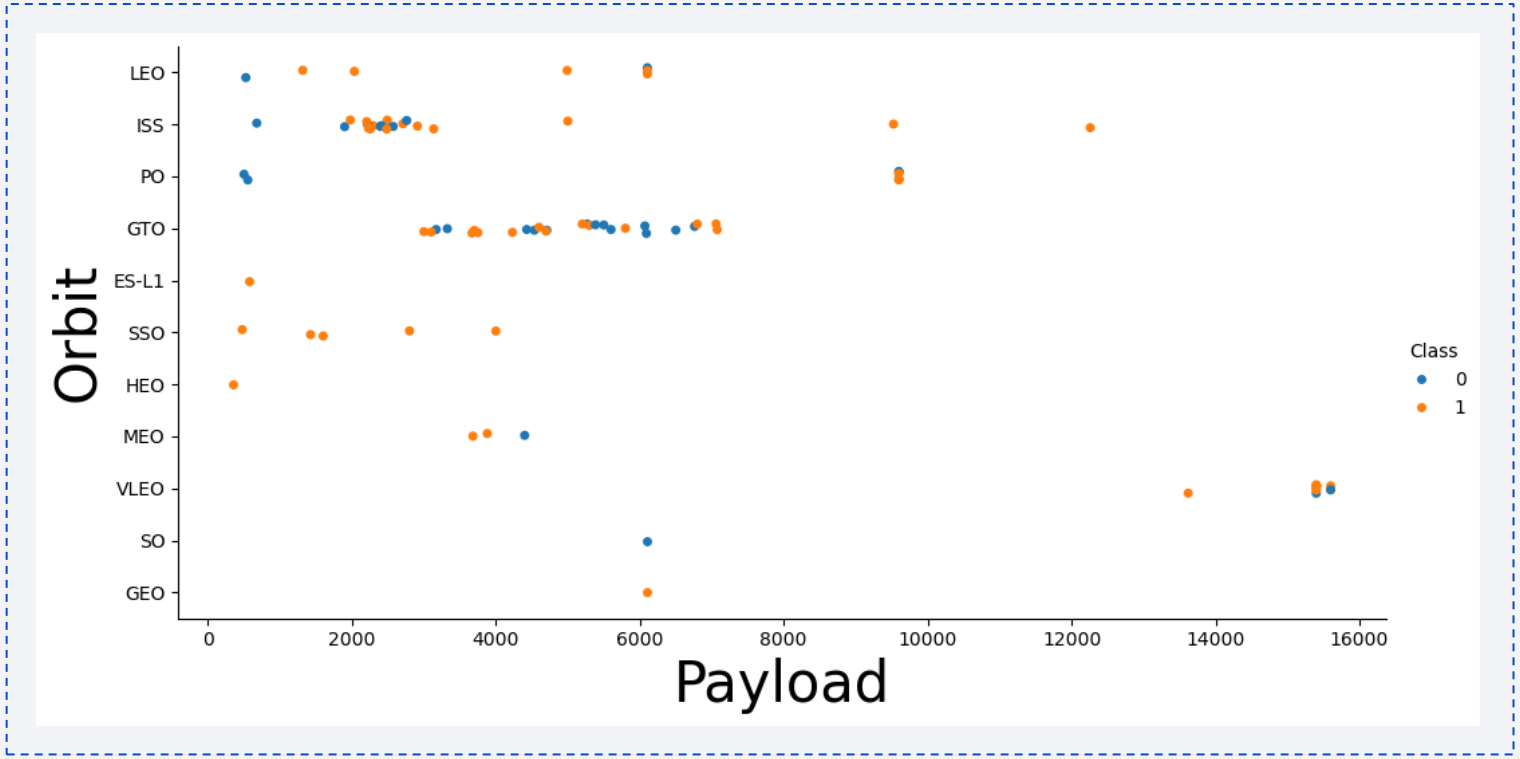
Flight Number vs. Orbit Type

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



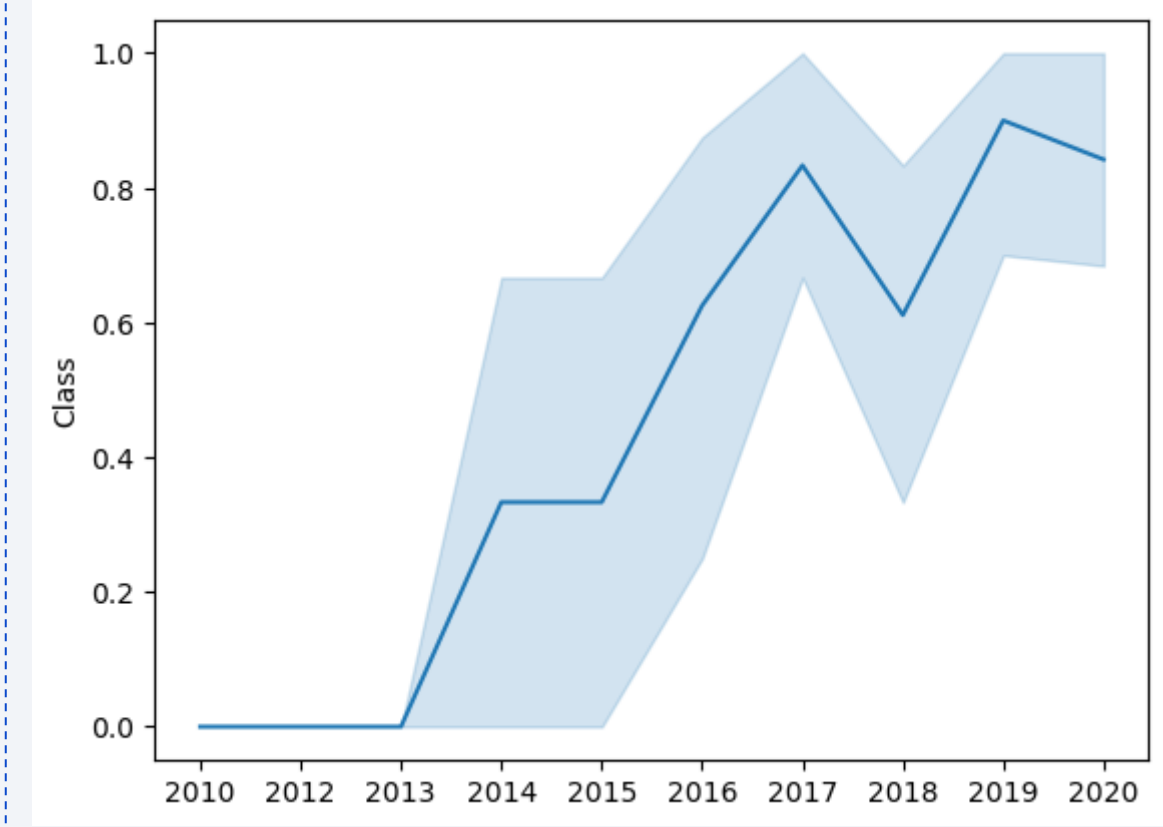
Payload vs. Orbit Type

From the scatter plot we can address that with heavy payloads the successful landing or positive landing rate are more for 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.



Launch Success Yearly Trend

The scatter plot show that the success rate since 2013 kept increasing till 2020.



All Launch Site Names

The query result present all the launch sites names:

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

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

Launch Site Names Begin with 'CCA'

First 5 records where the launch site begin with `CCA`

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

Total Payload Mass

This query result present the sum of the total payload mass in kg where NASA was the customer.

Total payload mass (NASA (CRS))
45596

Average Payload Mass by F9 v1.1

This query result present the average payload mass or launches which used booster version F9 v1.1

Average payload mass (booster version F9 v1.1)
2534.6666666666665

First Successful Ground Landing Date

Bellow we can see the query result for the first successful landing achieved.

First successful landing acheived
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The query return the four booster versions that had successful drone ship landings and a payload mass 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

This query returns a count of each mission outcome, including success and failure both. Also the query include the sum of both outcomes.

Outcome	Count
Success	61
Failure	40
(All)	101

Boosters Carried Maximum Payload

This query returns the booster versions that carried the highest payload mass.

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

The follow query result present the list of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.

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

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

This query returns the rank 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.

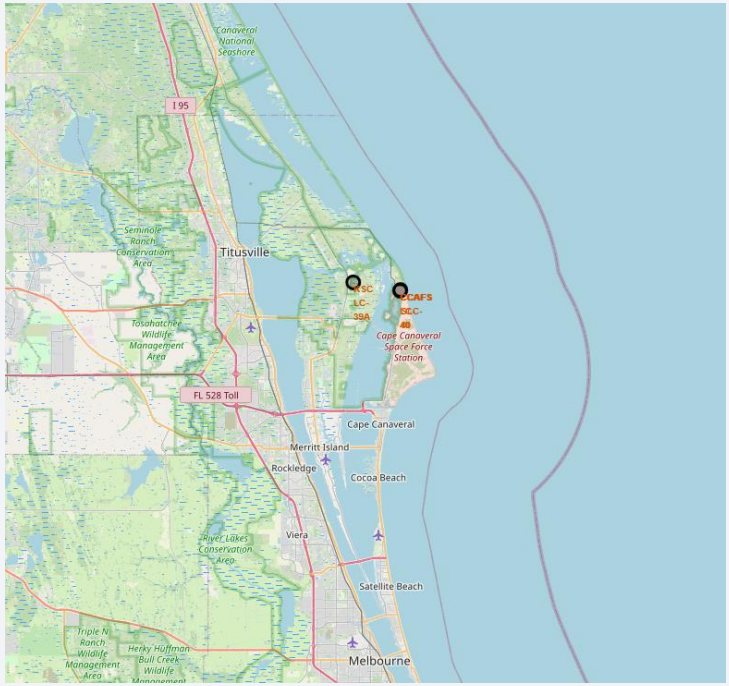
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

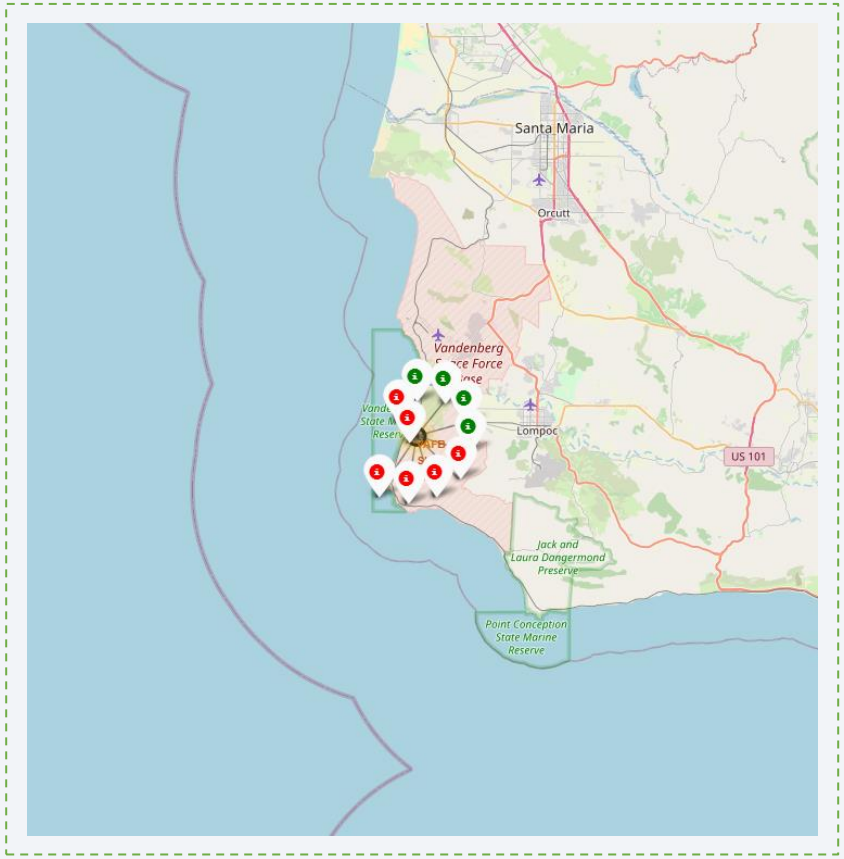
Launch Site Locations

The left map shows all launch sites in the US map. In other hand, the right map shows the specifics launch sites in Florida.



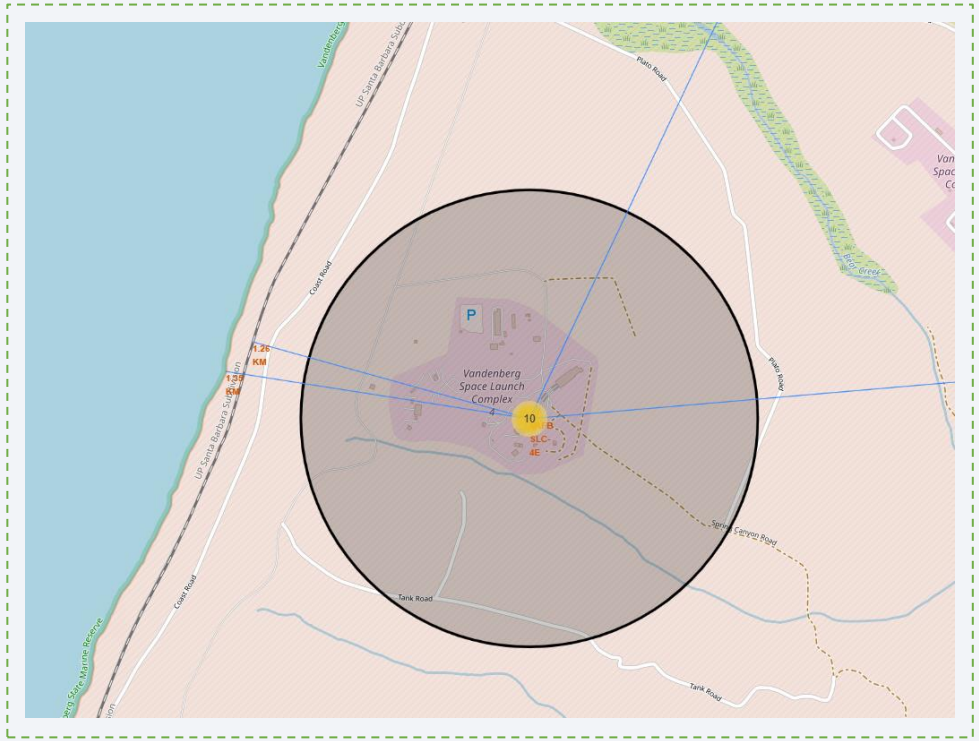
Launch Markers

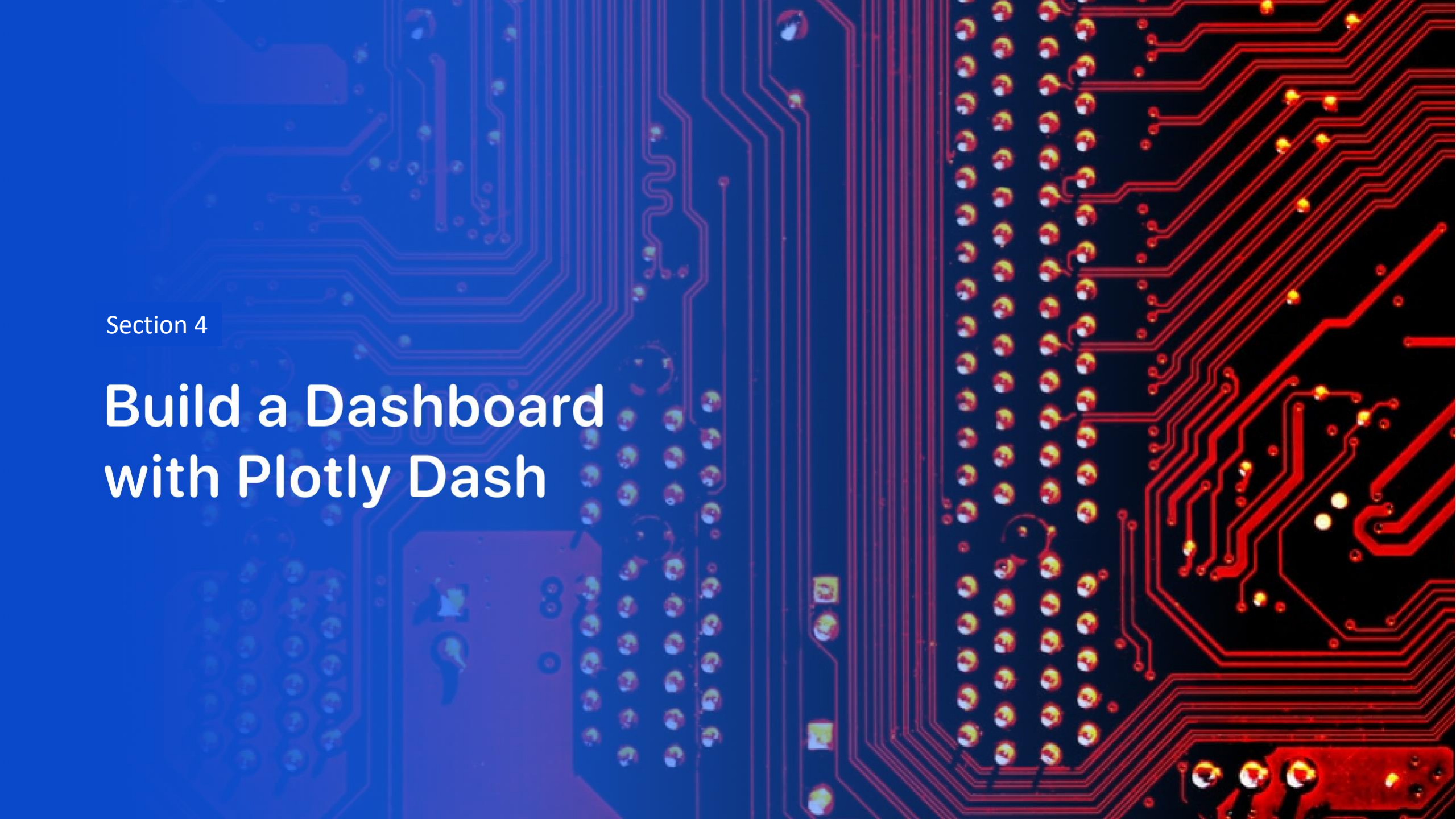
Clusters can be clicked on to display each successful landing or failed landing. In the picture we can see that VAFB SLC-4E have 4 successful landings and 6 failed landings.



Locations Proximities

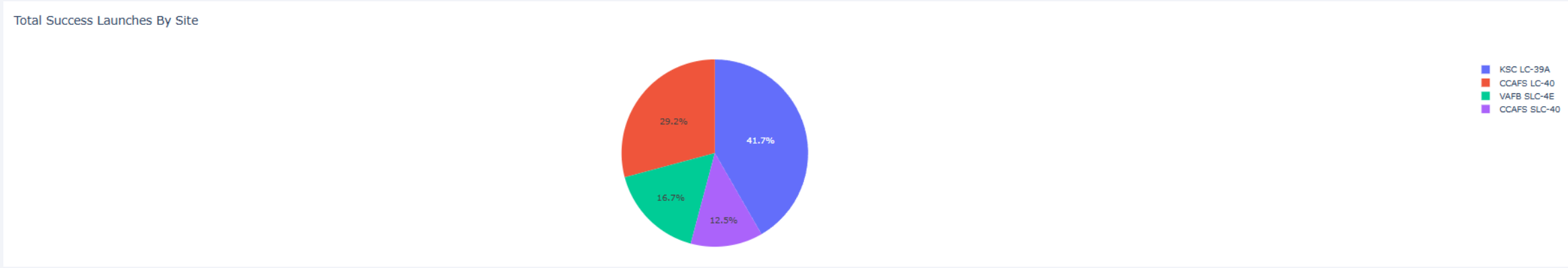
The picture shows the distance between the VAFB SLC-4E launch site and the nearest coastline. From this we can address that the launch sites are to nearest to the coastlines.




Section 4
Build a Dashboard with Plotly Dash

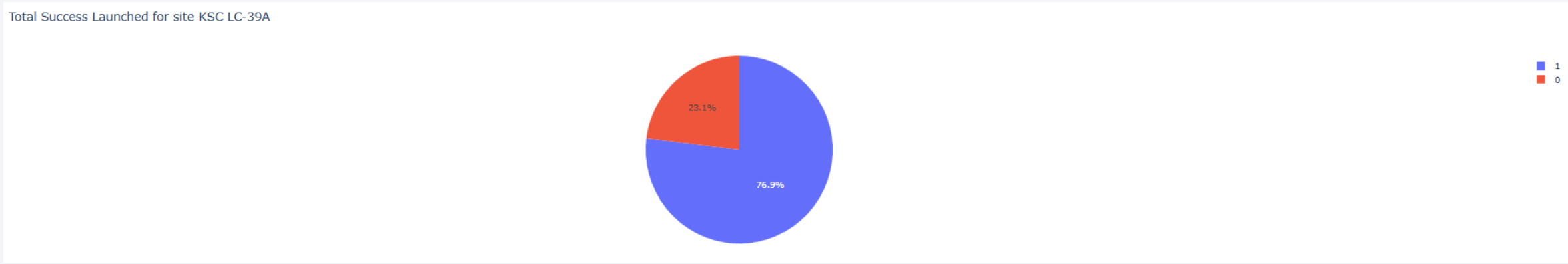
Successful Launches Across Launch Sites

This distribution shows the successful landings across all launch sites. We can infer that the KSC LC-39A have the majority of the successful landings. VAFB SLC-4E and CCAFS SLC-40 share an almost equal percentage of successful landings.



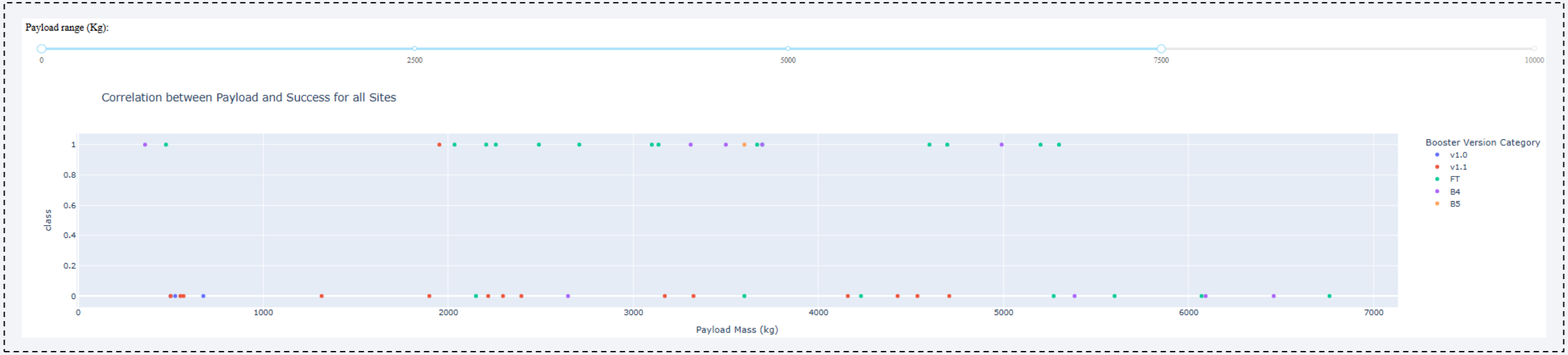
Highest Success Rate Launch Site

The picture below shows that the KSC LC-39A has the highest success rate.



Payload Mass vs. Launch Outcome for All Sites

The Dashboard shows what occurs when you select all the launch sites and a Payload range between 0 and 7500.

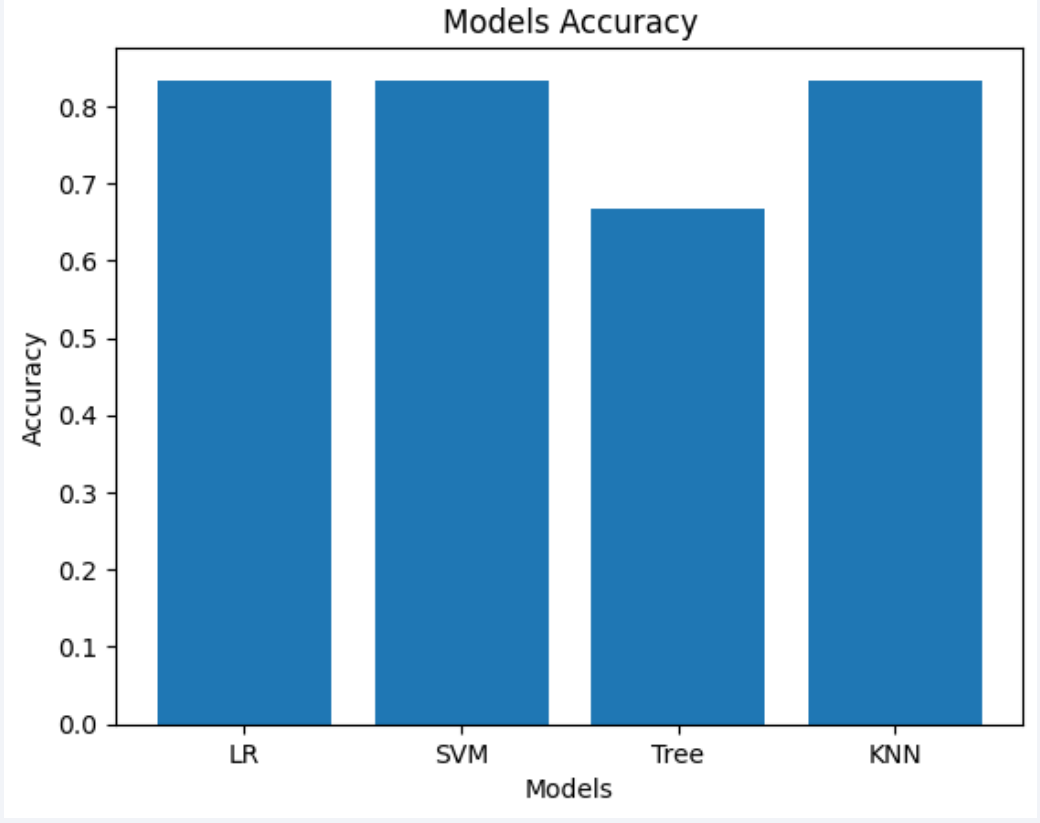


Section 5

Predictive Analysis (Classification)

Classification Accuracy

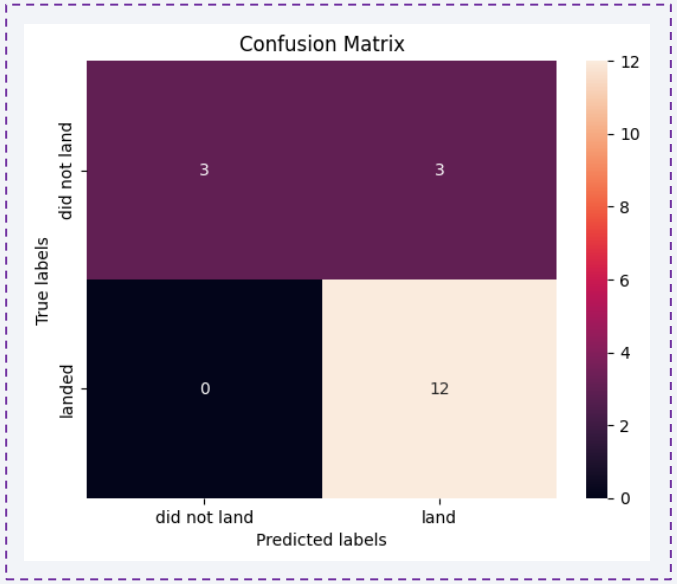
In essence, the models have almost exactly the same precision (83.33%), except for the decision tree classifier (72.23%). We can say that is necessary more data to determine which is the best model.



Confusion Matrix

Since most models yielded the same precision, the confusion matrix is largely the same across all models.

- The models predicted 12 successful landings when the true label was successful landing.
- The models predicted 3 unsuccessful landings when the true label was unsuccessful landing.
- The models predicted 3 successful landings when the true label was unsuccessful landings.
- Our models over predict successful landings.



Conclusions

- This project was carried out with the main idea of trying to predict if the first stage of a given Falcon 9 launch could work, with the intention of determining the cost of the launch.
- Some features of the Falcon 9 launch, such as the payload mass, the launch location or the type of orbit, may affect the outcome of the mission in some way.
- The machine learning algorithms used to study the Falcon 9 launch allowed us to create predictive models in order to determine the final result of a Falcon 9 launch.
- The predictive model produced by Logistic Regression, Supportive Vector Machine, and K Nearest Neighbors performed the best among the 4 machine learning algorithms employed.

Appendix

You can access to the GitHub repository from the following URL

atwncito/Capstone-final-project: Capstone final project from Coursera (github.com)

Special thanks to Coursera

platform for giving the course c:

Thank you!