IBM Developer
SKILLS NETWORK

Winning Space Race With Data Science

Zinat Nawfar

October 2022

Outline

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

Executive Summary

Methodologies

- Data Collection: SpaceX API & Web Scraping.
- Data Wrangling.
- Exploratory Data Analysis (EDA): SQL & Visualization (Matplotlib, Seaborn, Folium).
- Machine Learning Prediction: (build, tune and evaluate classification models).

Results

- Creating a decision Tree Classifier model that recorded 94% accuracy on variable data.
- A series of SQL and graphing queries provided us with an excellent framework to support the finding.
- An interactive map with launch integrated launch information allows the user to explore the data on their own.

Introduction

Project context

- SpaceX is probably most successful company that will make space travel affordable for everyone.
- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
- Other providers cost upward of 165 million dollars each.
- Much of the savings is because Space X can reuse the first stage
- 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 space X for a rocket launch.

Our questions

- Determining the price of each launch by gathering information about SpaceX and creating dashboards.
- Training a machine learning model and using public information to predict if SpaceX will reuse the first stage.

Methodology

- Data collection methodology (SpaceX Rest API & Web Scrapping from Wikipedia)
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

Requesting & Parsing the SpaceX launch data using the GET request

Get information about the launches using the IDs given for **each** launch (Columns: rocket, payloads, launchpad, and cores.)

Filter the dataframe to only include Falcon 9 launches

API returns SpaceX data in .JSON

Normalize data into flat data file such as .csv

[GitHub URL]()

Data Collection - Scraping

- Requesting the Falcon9 Launch Wiki page from its URL
- Extracting all column/variable names from the HTML table header
- Creating a data frame by parsing the launch HTML tables
- Normalize data into flat data file such as .csv

[GitHub URL]

Data Wrangling

In the data set, there are several cases where the booster did not land successfully.

Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed in a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed in a specific region of the ocean.

True RTLS means the mission outcome was successfully landed on a ground pad False RTLS means the mission outcome was unsuccessfully landed on a ground pad.

True ASDS means the mission outcome was successfully landed on a drone ship False

ASDS means the mission outcome was unsuccessfully landing on a drone ship.

I mainly convert those outcomes into Training Labels with 1 meaning the booster successfully landed 0 meaning it was unsuccessful.

Data Wrangling

- Performing EDA to find some patterns in the data and determine what would be the label.
- Calculating number of launches on each site, number and occurrence of each orbit, mission outcome per orbit type.
- Creating a new column with the launch outcome as a binary variable.
- Working out success rate for every landing in dataset.

[GitHub URL]

EDA with Data Visualization

Visualizng the SpaceX launch dataset using `matplotlib` and `seaborn` and discovering some preliminary correlations between the launch site and success rates:

Scatter plots						Bar Graph	Line Graph
Flight Number VS. Payload Mass	Flight Number VS. Launch Site	Payload VS. Launch Site	Orbit VS. Flight Number	Payload VS. Orbit Type	Orbit VS. Payload Mass	Mean VS. Orbit	Success Rate VS. Year

GitHub URL

EDA with SQL

Performing SQL queries to gather information about SpaceX data set

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC’
- 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 where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad 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.
- Listing the records which will display the month names, successful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
- Ranking the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

GitHub URL

Build an Interactive Map with Folium

The launch success rate may depend on many factors such as payload mass, orbit type, and so on.

It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories.

Finding an optimal location for building a launch site certainly involves many factors.

We have discovered some of the factors by analysing the existing launch site locations.

Build an Interactive Map with Folium

We have taken the Latitude and Longitude Coordinates at each launch site and added a Circle Marker around each launch site with a label of the name of the launch site to visualize the Launch Data into an interactive map.

Success and failed launches were added as markers with different colours within clusters for each launch site.

Distances from launch site to risk locations were calculated and marked with lines.

[GitHub URL]

Build a Dashboard with Plotly Dash

To allow an interactive exploration of launch data a dashboard with a dropdown option for launch sites and a range slider to select payload has been created.

A pie chart with the successful launch outcome and a scatter plot with the launch outcome for the payload have been created.

Predictive Analysis (Classification)

We have created **a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.**

Building Model, Evaluating Model, Improving Model, Finding the Best Performing Classification Model

To predict the launch outcome four different classification models have been evaluated: Logistic Regression, Support Vector Machine, Decision Tree Classifier, k Nearest Neighbors

GitHub URL

Results

Calculating number of launches per site

Calculating number and occurrence of each orbit

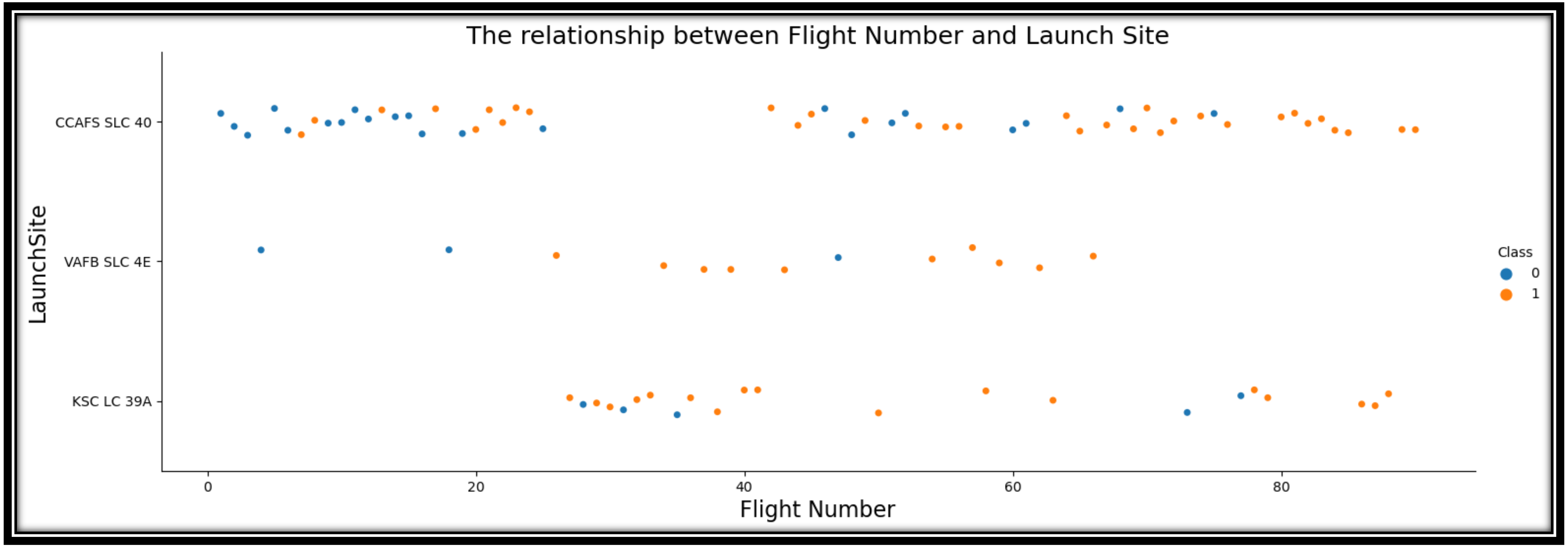
Calculating the number and occurrence of mission outcome per orbit type

Plotting flight characteristics (payload, flight number, launch site) vs. success rate

Predictive analysis results: Decision tree classifier performed best & Accuracy rating of 94.4%.

Insight drawn from EDA

Flight Number vs. Launch Site

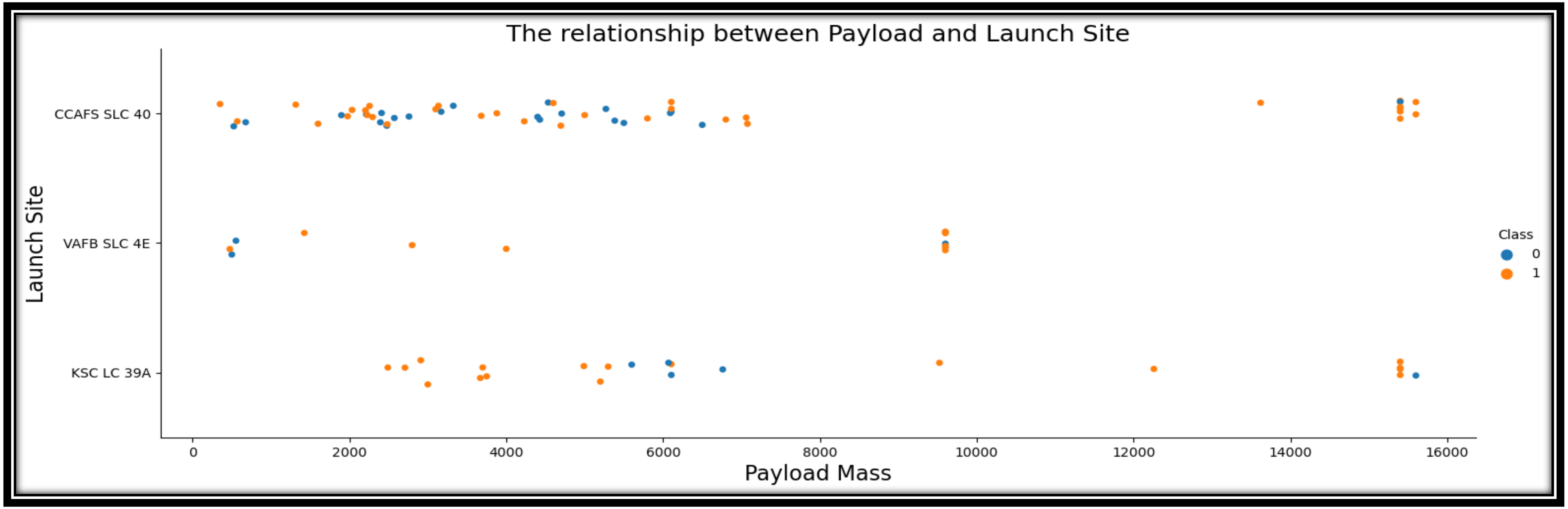


Higher Flight Numbers have more success in launch outcomes for all launch sites.

WAFB SLC 4E has a lower Flight Number than the other launch sites.

Low Flight Numbers have a high number of unsuccess launch outcomes.

Payload vs. Launch Site



Higher Flight Numbers have more success in launch outcomes for all launch sites.

WAFB SLC 4E has a lower Flight Number than the other launch sites.

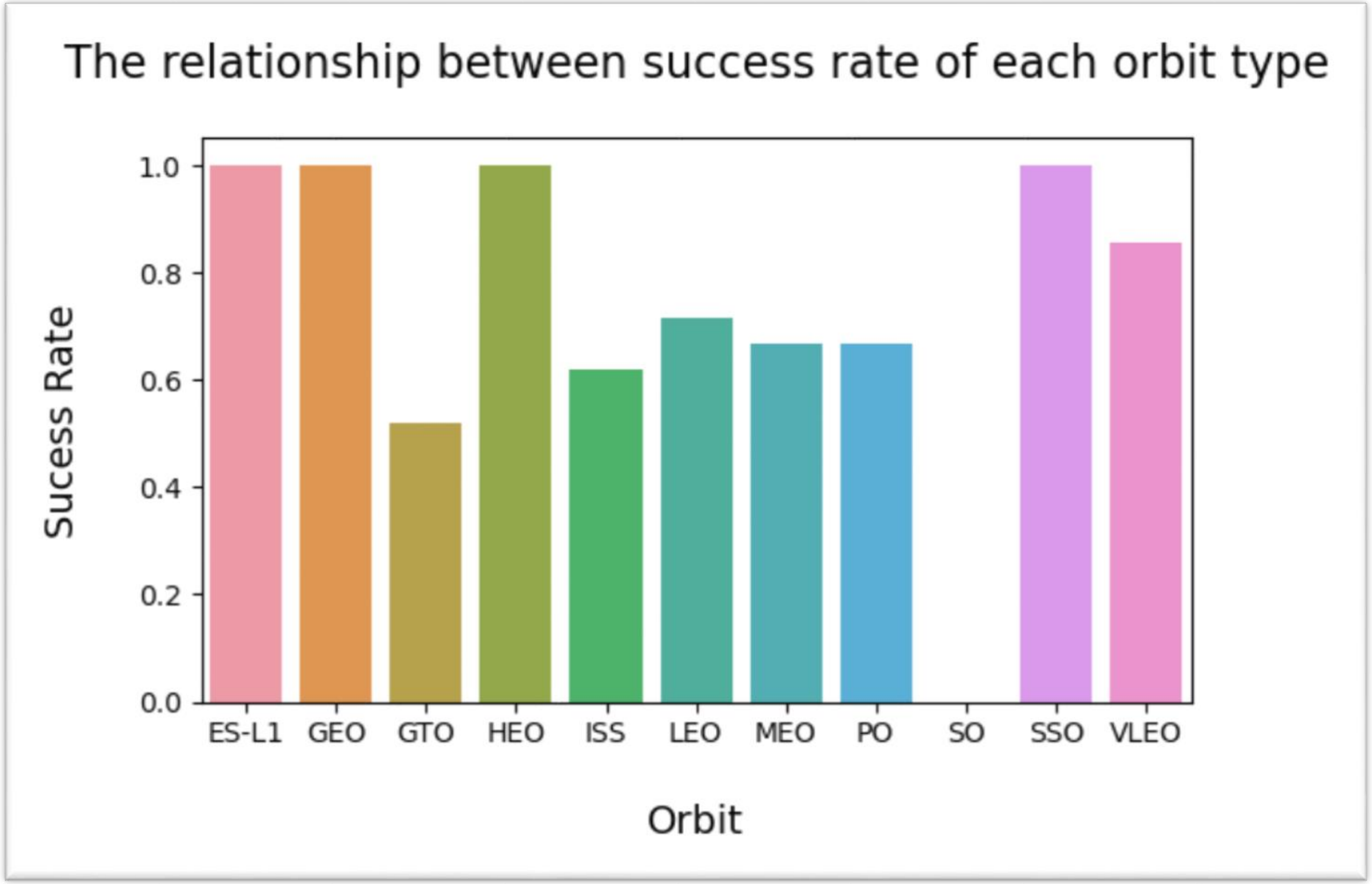
Low Flight Numbers have a high number of unsuccess launch outcomes.

For the VAFB-SLC launch site, there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

Four Orbits have a total higher success rate of launch outcome (ESL1, GEO HEO, and SSO).

GTO is the Orbit with a lower success rate of launch outcome.

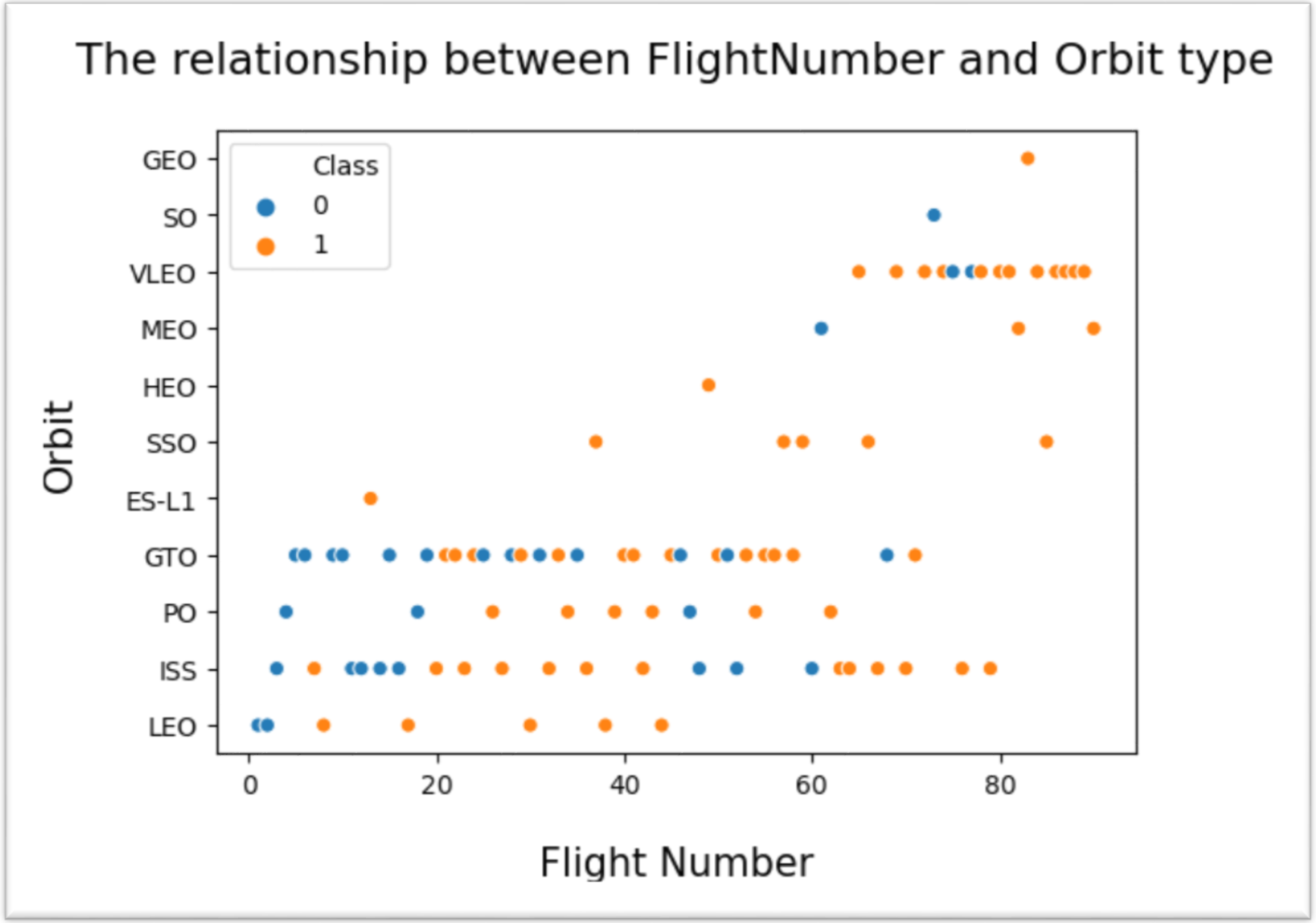


Flight Number vs. Orbit Type

Low Flight Numbers only occur in 5 Orbits

Success is higher when the Flight Number is higher for most Orbits

No relationship between flight numbers when in GTO orbit.

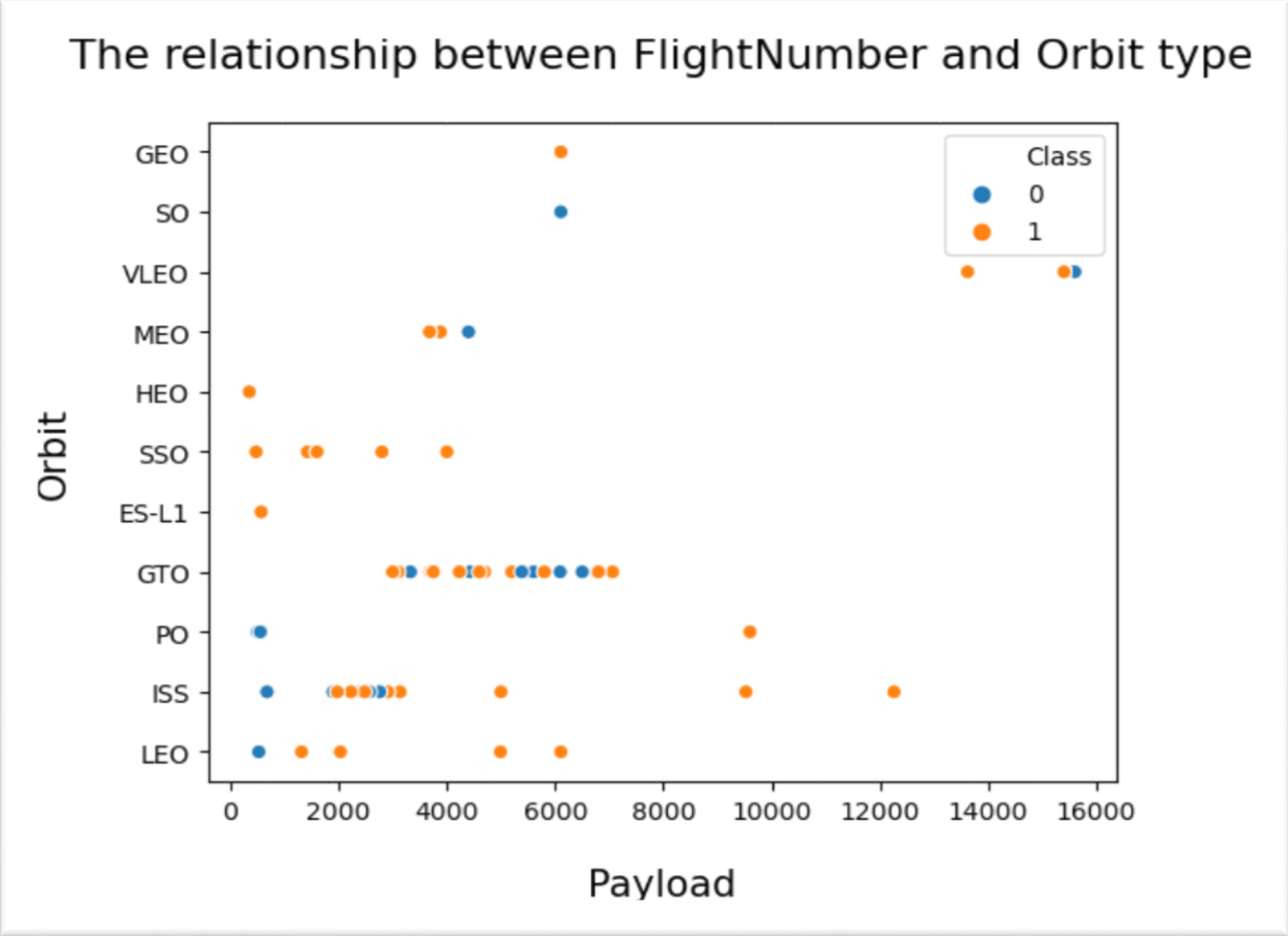


Payload vs. Orbit Type

Payloads higher than 8000 Kg only occur in 3 Orbits (ISS, PO, VLEO).

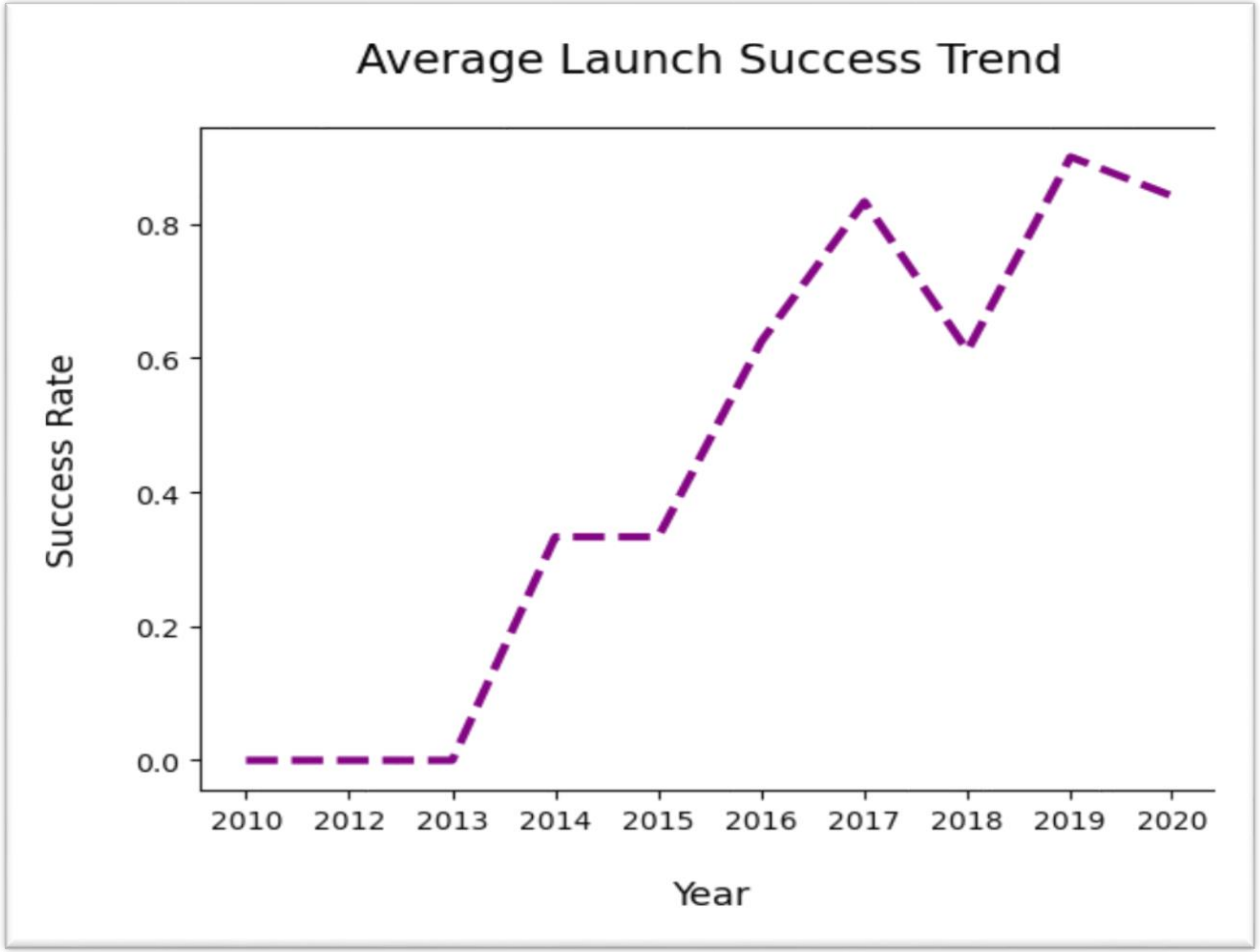
For higher Payloads, only VLEO Orbit has unsuccess launches.

For GTO Cannot distinguish either a positive landing rate or a negative landing.



Launch Success Yearly Trend

Launch's outcome success rate has been going up from 2013 to 2020.



All Launch Site Names

There are FOUR unique names for launch sites

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

Launch Site Names Begin with 'CCA'

First five records of launch sites that start 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

Total Payload Mass

Total payload mass (Kg) carried by boosters launched by NASA

Total Payload Mass by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

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

Average Payload Mass by Booster Version F9 v1.1
2928.4

First Successful Ground Landing Date

First successful landing outcome on ground pad was achieved on 2015

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters' versions with payload mass between 4000 and 6000 Kg that has successfully landed on the drone ship

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

Total Number of Successful and Failure Mission Outcomes

Most missions have been successful.

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

The twelve boosters have carried the maximum payload mass.

Booster Versions which carried the Maximum Payload Mass
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 booster versions and launch site for the two failed landing outcomes in drone ships in 2015.

booster_version	launch_site	DATE
F9 v1.1 B1012	CCAFS LC-40	2015-01-10
F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

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.

Landing Outcome	Total Count
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Launch Sites Proximities Analysis

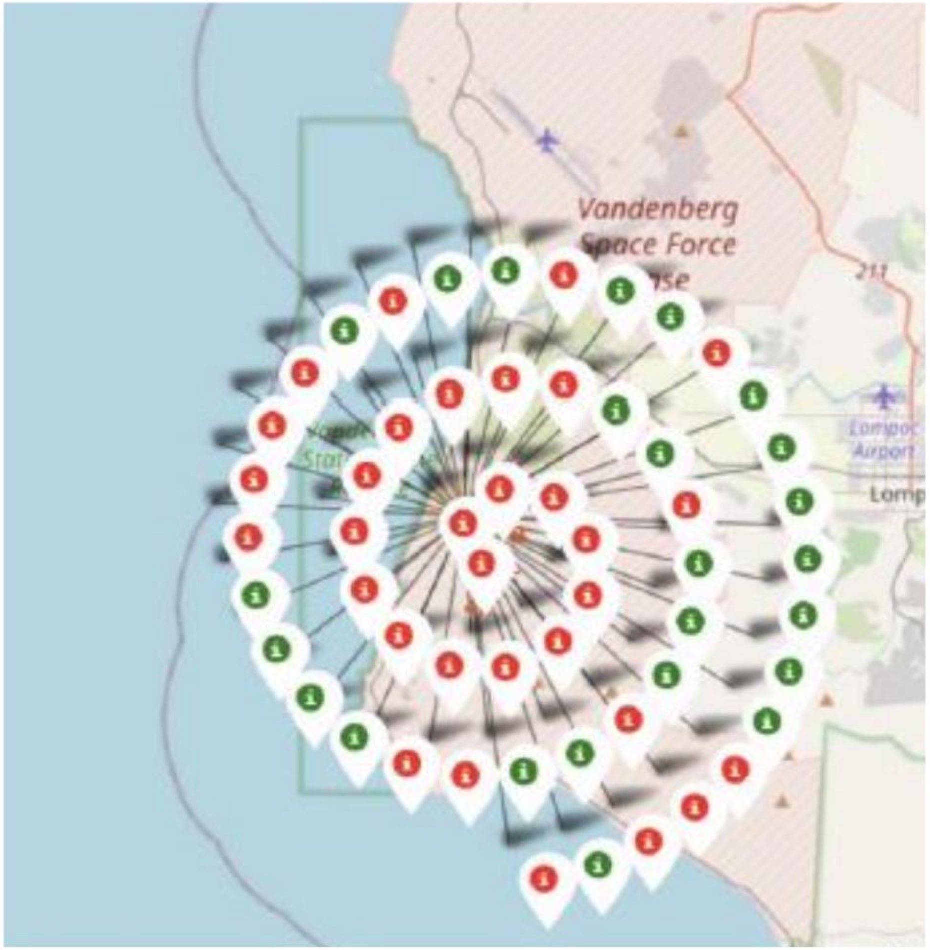
Mark all launch sites

- **The SpaceX launch sites are on the United States of America coasts.**
- **Three are located on the east coast, and only one is on the west coast.**

Mark the success/failed launches for each site

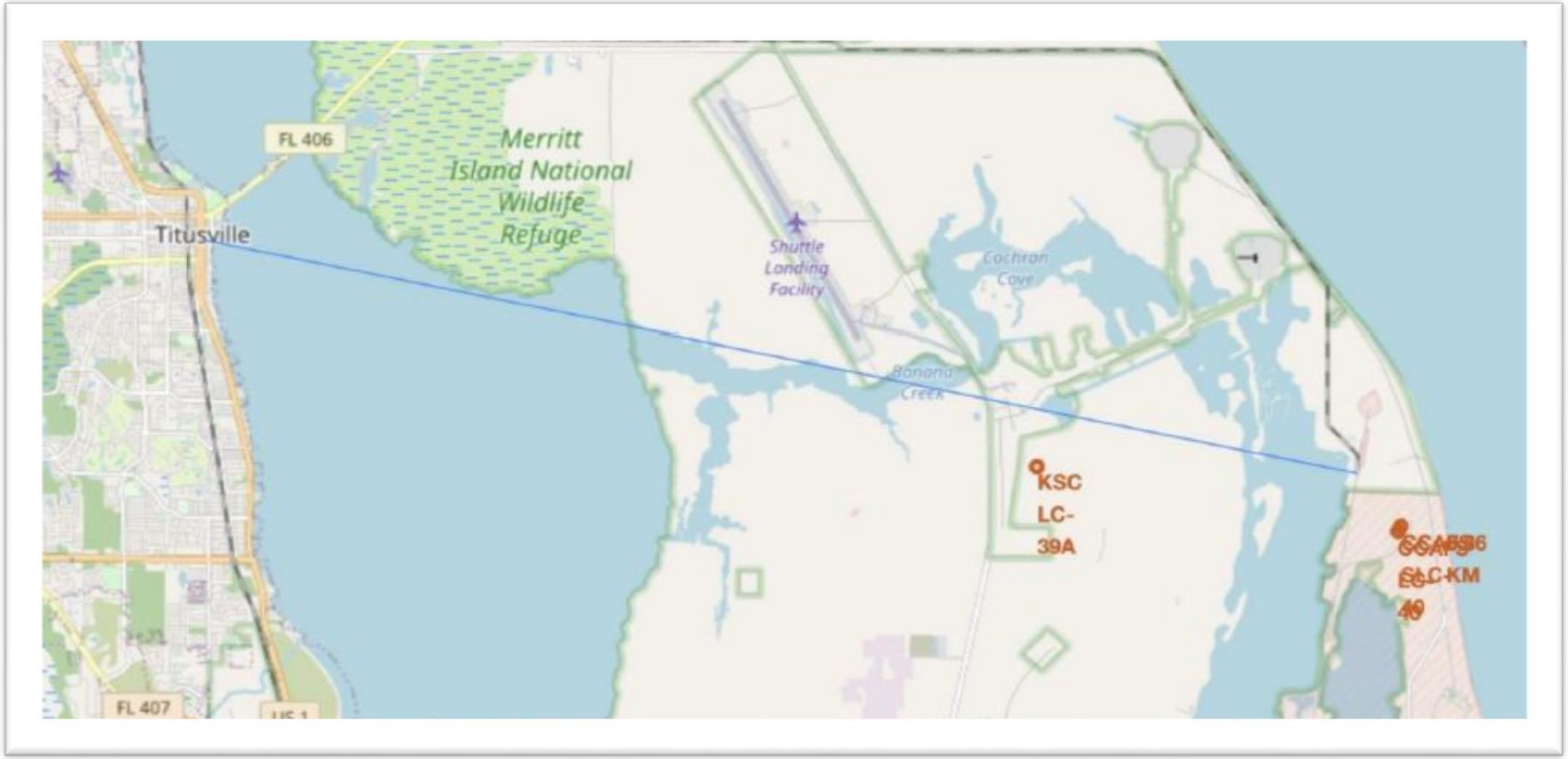
- Green Marker demonstrates successful Launches, and Red Marker demonstrates Failures.

- From the seven launches from this site, only three had first-stage recovery success.



The distances between a launch site to its proximities

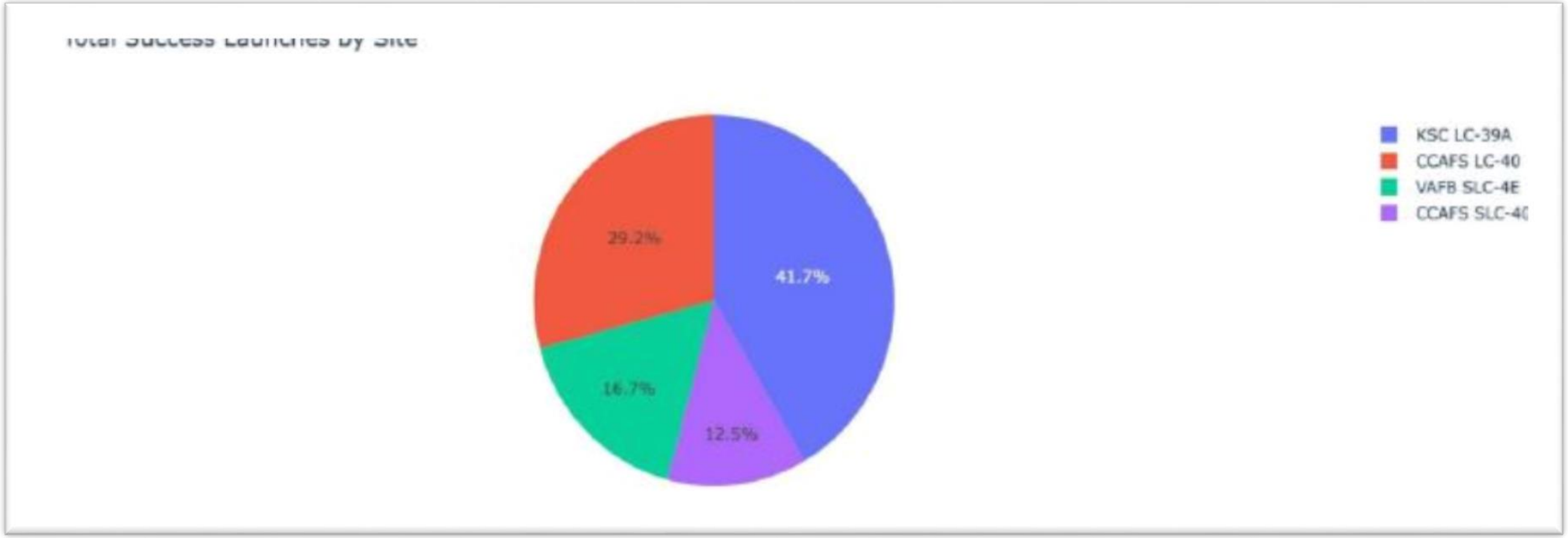
- **Launch sites have been located near the coastline and have highways and railways in a short distance.**



- **They keep certain distances from cities.**

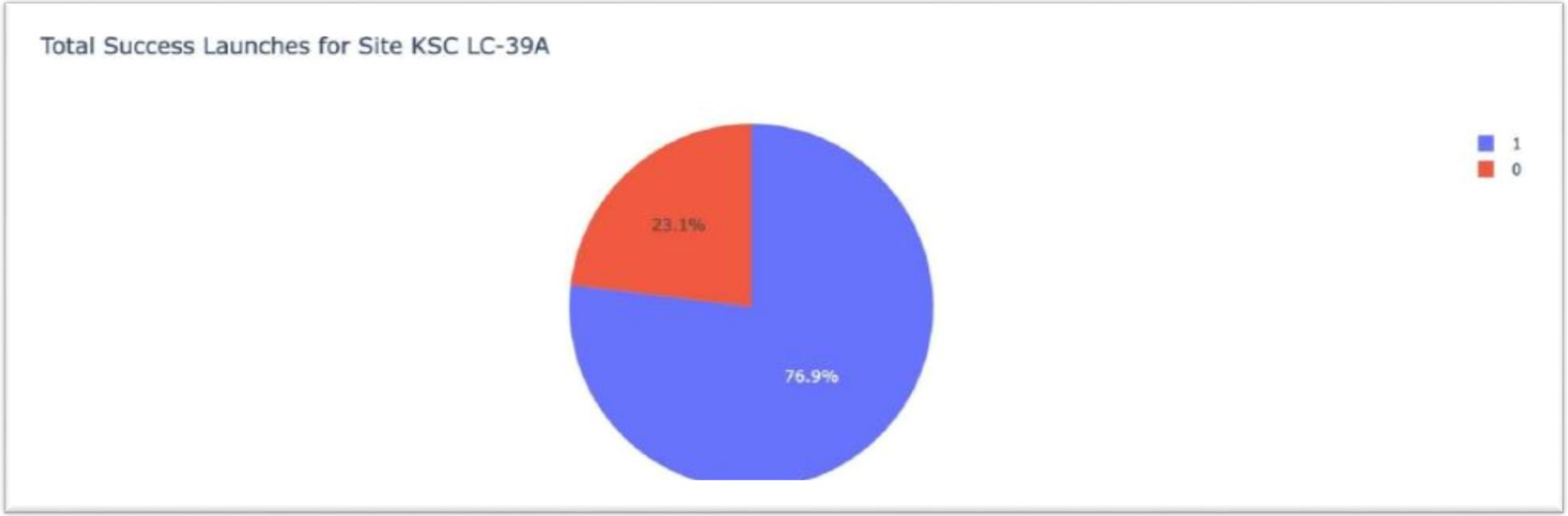
Dashboard with Plotly Dash

The launch success ratio for all sites



- **KSCL LC 39A with the higher successful launches number**
- **CCAFS SLC 40 with the lower successful launches number**

Launch sites with the highest launch success ratio



KSC LC-39A has achieved a higher success ratio with 76.9% of launch outcomes.

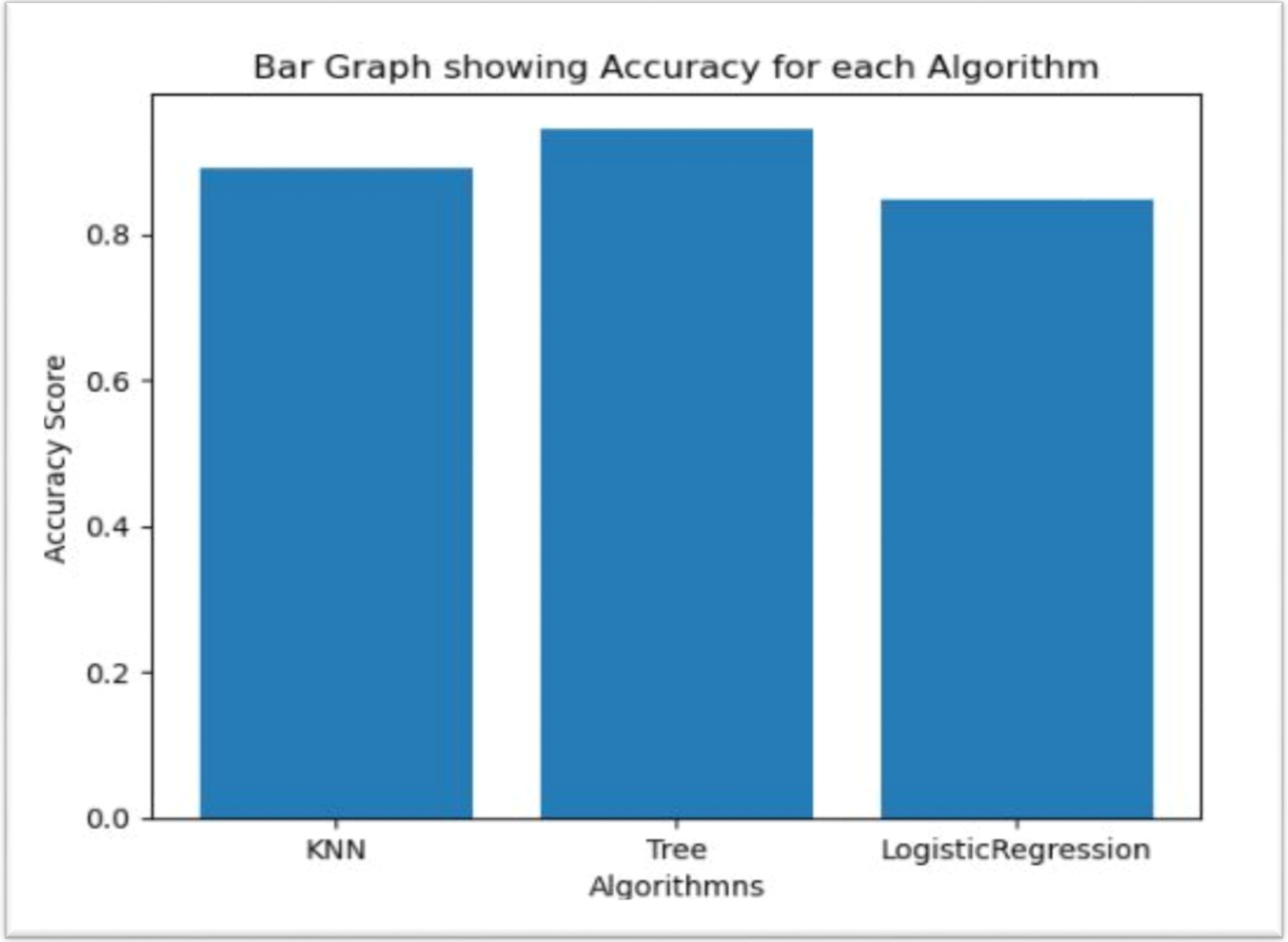
Payload vs. Launch Outcome scatter plot

With different Payload Mass, Booster version FT has a high success rate.

Predictive Analysis

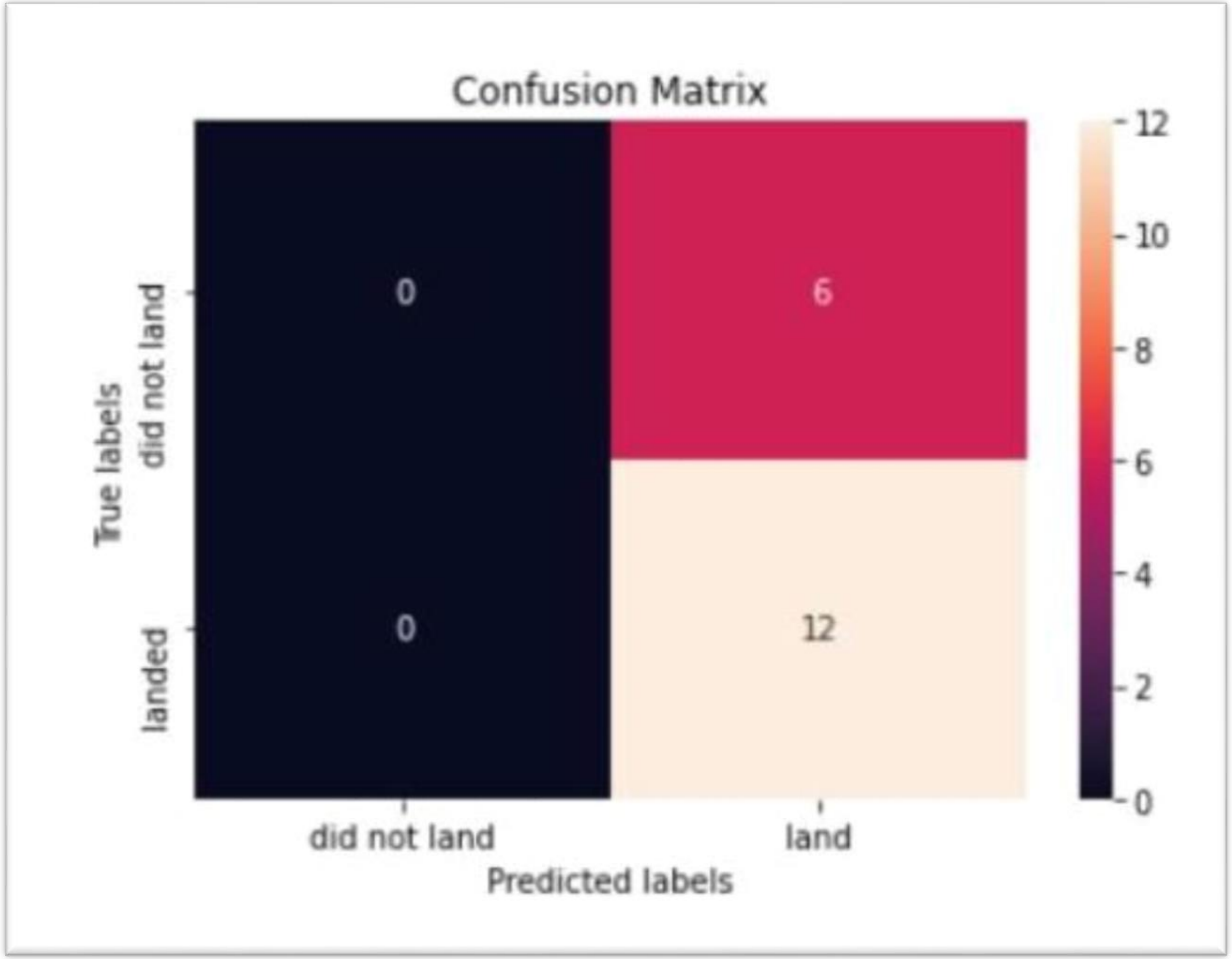
Classification Accuracy

The tree algorithm has reached the best accuracy score.



Confusion Matrix

It is clear that Tree can distinguish between the different classes. We see that the major problem is false positives.



Conclusions

Payload mass, Orbit, and Launch Site have been significant variables to predict the launch outcome success.

Best Launch Site: KSCL LC 39A .

Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate

The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches

Totally, successful outcomes have been higher than failed ones.

Appendix

- Git repository Link

Thank You!