IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Introduction

Project background and context:

Advancement in technology is accelerating over time. Subsequently, launching rockets to space has had a drastic price decrease. This decrease mainly comes from recovering the rocket's first stage, the bulk of the expense

Problems we want to find answers to:

To better understand the cost of launching a rocket to space, we need to determine if the first stage can be recovered. Our main objective is to find a method to label whether the first stage will be recovered or not

Executive Summary

Summary of methodologies:

- Data Collection
- Data Wrangling
- Exploratory Data Analysis

Summary of all results:

- Presenting Preliminary Findings (using interactive visuals and dashboards)
- Predictive Analysis with Machine Learning

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using APIs and Web Scraping (BeautifulSoup)
- Perform data wrangling
  - The Data was manipulated, processed, and cleaned using pandas on Python
- 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 predictive analysis was optimized by comparing the different methods: Decision Tree, k-Nearest Neighbors, Support Machine Vector, Logistic Regression and choosing the one providing the least error

Data Collection

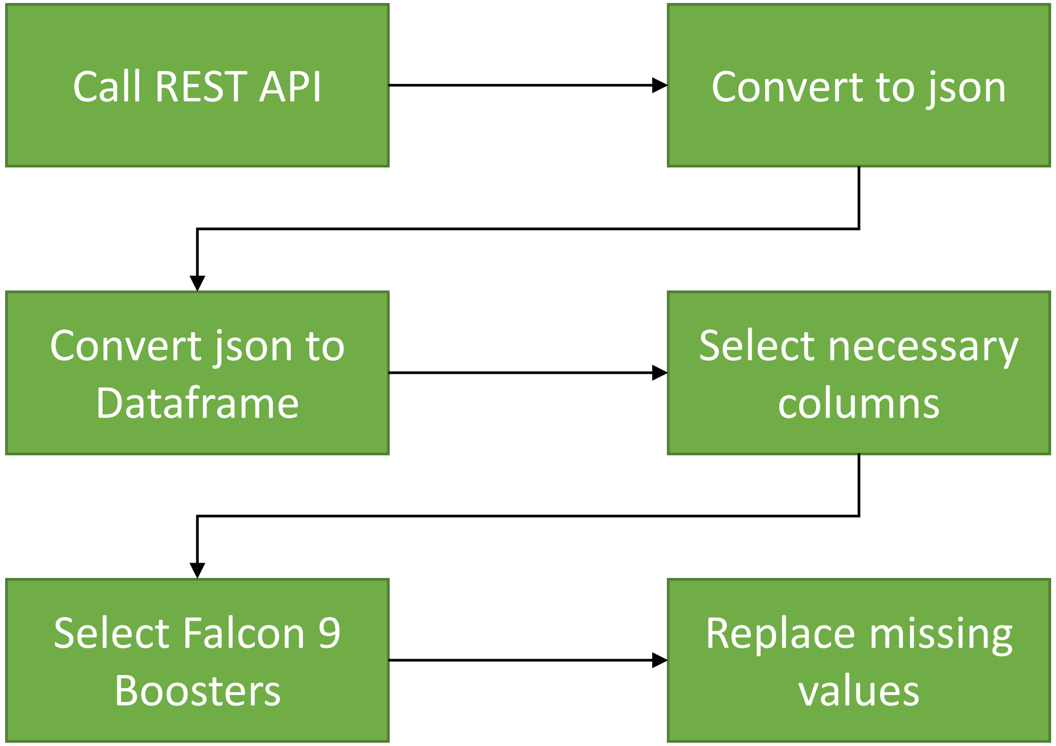
- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Call the REST API from:

  https://api.spacexdata.com/v4/launches/past

- Convert the REST API to a readable json file using the .json() method
- Turn the json file into a panda dataframe using the .json_normalize() method
- Select the necessary columns from the table
- Select only the Falcon 9 Booster Version
- Replace all missing values in the Payload Mass column with the column average and leave the rest as missing



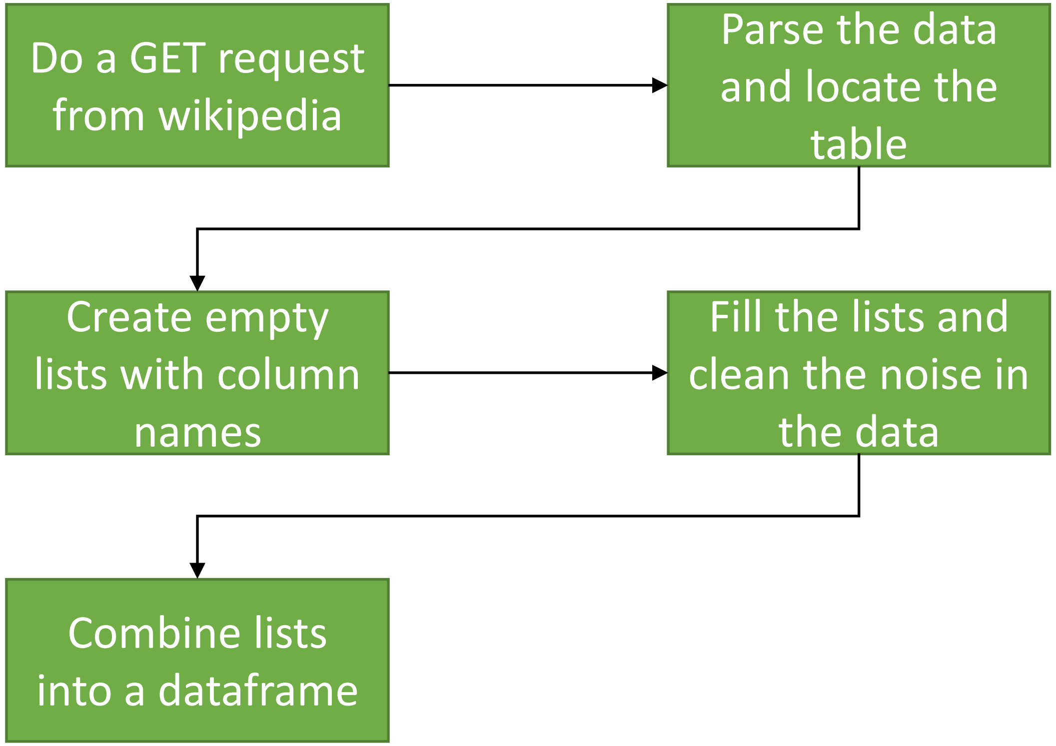
GitHub Data Collection Permalink

Data Collection - Scraping

- Call GET the request from:

  https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

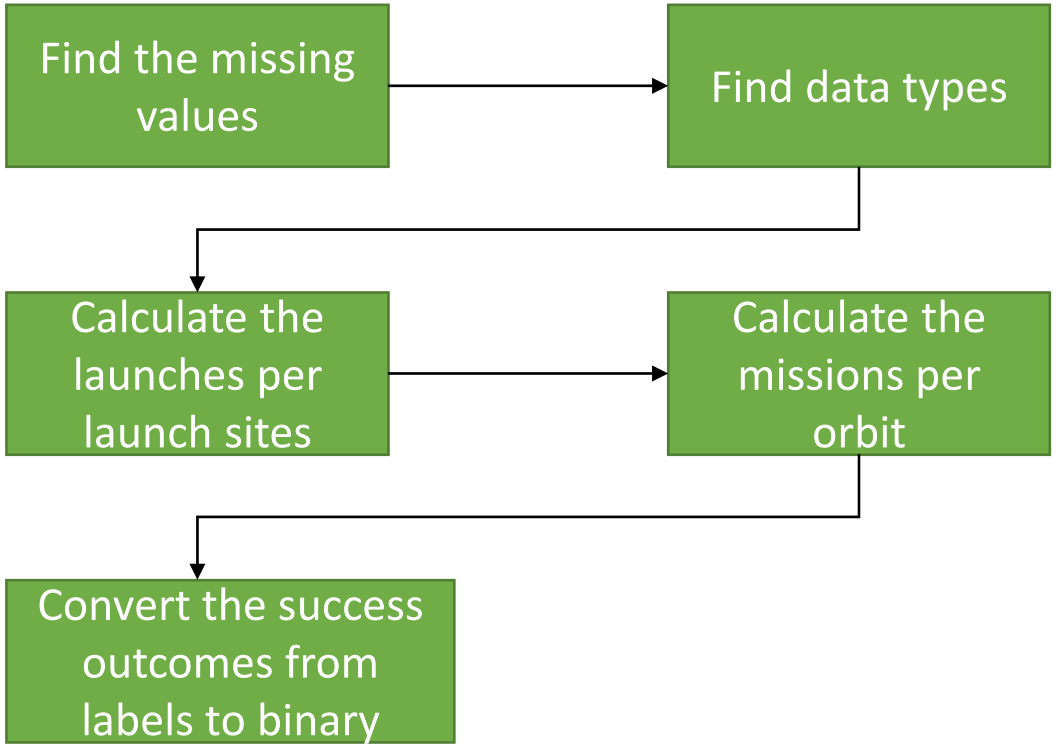
- Parse the request using BeautifulSoup and find the table by using the find_all() method
- Select the necessary columns from the table by creating empty lists with the targeted columns
- Fill the lists and while filling clean the noise in the data, since HTML from Wikipedia comes with a lot of noise and citations
- Create a dataframe by combining the lists



GitHub Data Collection Permalink

Data Wrangling

- Find where the missing values are and how abundant they are
- Find out the data types
- Get the number of launches per launch site
- Get the number of missions per orbit
- Create a binary column depending if the outcome was a success or not



GitHub Data Collection Permalink

EDA with Data Visualization

Scatterplot charts

- Payload mass and flight number
- Launch site and flight number
- Payload mass and launch site
- Flight number and orbit type
- Payload and orbit type

To understand the how the change in payload mass and orbit type, payload mass and flight number, launch site and flight number, launch site and flight number, and orbit types and flight number affects landing success

Bar chart

- Number of successful landings and orbit type

To understand the relationship between the success rate and the orbit type

Line chart

- Success rate and year

To better represent the improvement in successful landing over the course of many years

[GitHub Data Collection Permalink]

EDA with SQL

- Displayed the names of unique launch sites
- Displayed some records with 'CCA' launch sites
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved
- Listed the names of the boosters which had success in drone ship landing and had a payload mass greater than 4000 and less than 6000
- Listed the total number of successes and failures in mission outcomes
- Listed the names of the booster versions which have carried the maximum payload mass
- Listed the failed landing outcomes in drone ship, in year 2015, their booster versions, and launch site names
- Ranked the count of landing outcomes between 2010-06-04 and 2017-03-20

[GitHub Data Collection Permalink]

Build an Interactive Map with Folium

- Added all launch sites and locations on the map
- Created a marker cluster to visualize the success and failures for each launch site

  Green was the success and red was the failure

- Created a line between VAFB SLC-4E and its nearest coastline, and calculated that distance

  I did all of that to get a better grasp of the launch site environment and why one might be more successful than the other

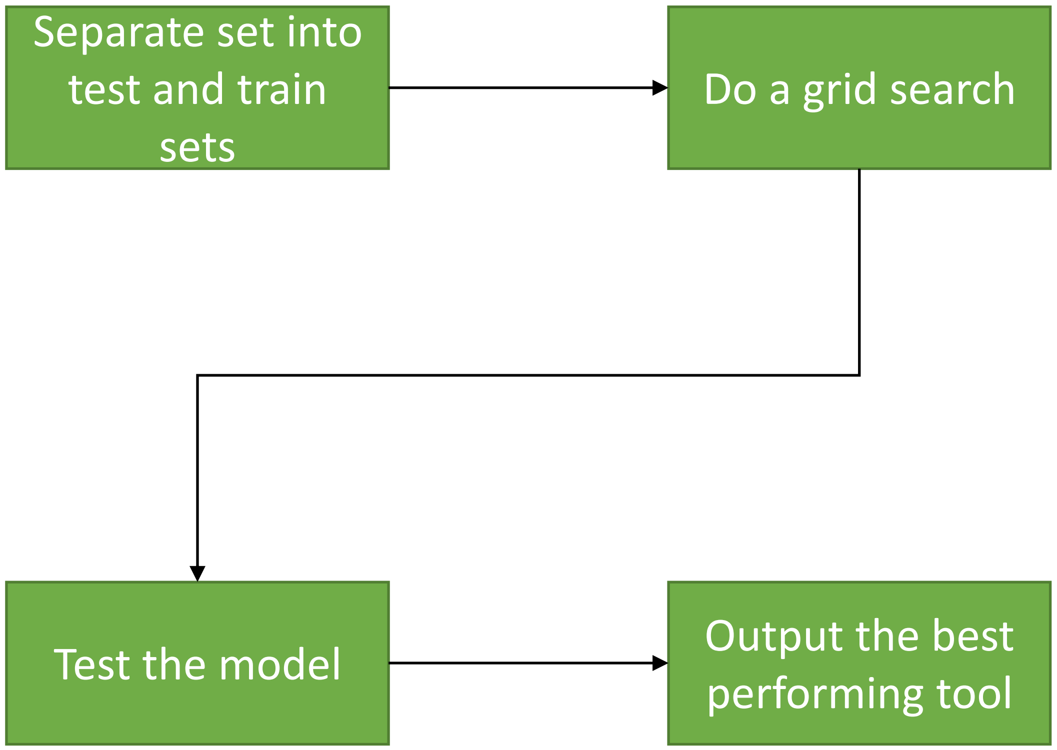
Build a Dashboard with Plotly Dash

- Added a pie chart comparing successful and failed landings
- Added a scatterplot visualizing success as a function of payload
- Added dropdown menus to choose different launch sites or all launch sites
- Added a slider to choose the payload range

I added all these to better visualize the effect of payload and launch site on the ratio of success to failed landing

Predictive Analysis (Classification)

- Break the dataset into train and test sets, with test set being 20% of the dataset
- Create a grid search for:
  - SVM, KNN, LR, and Decision Tree
  - Iterate with many different parameters
  - Cross validate over 10 folds
- Run the test set on the model and calculate the scores
- Iterate over the results to find the best performing prediction tool



Results

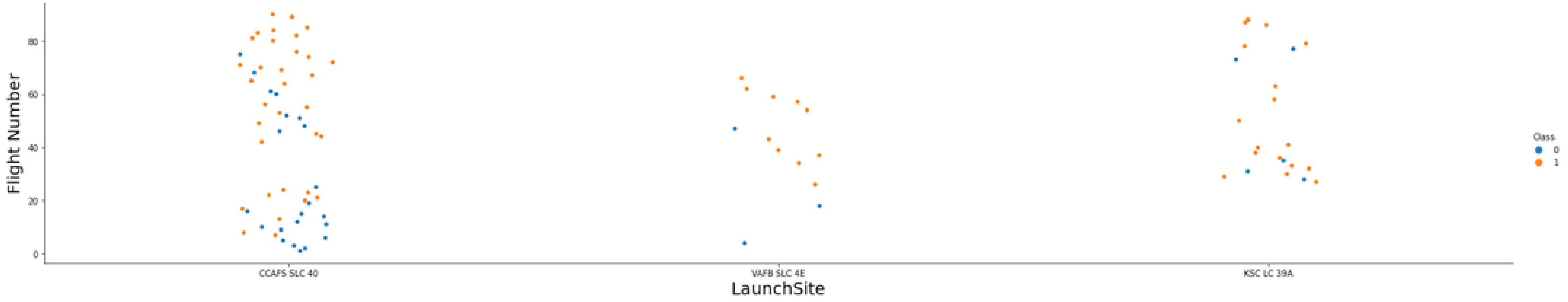
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

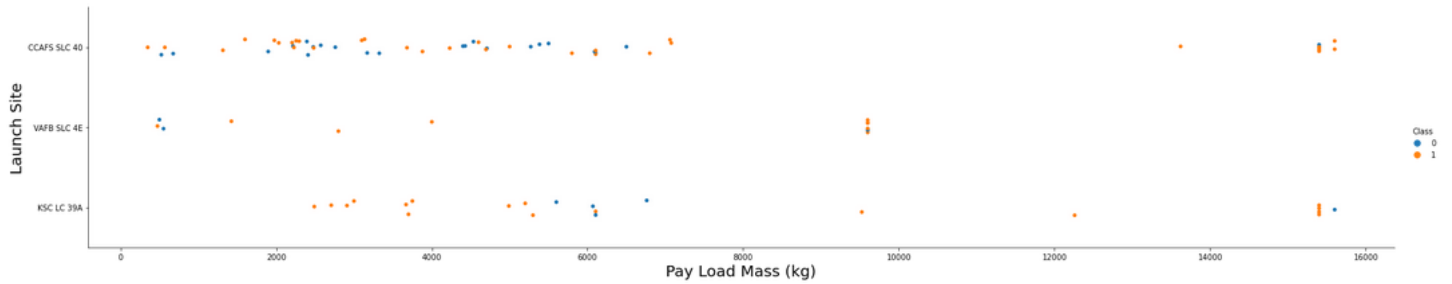
Flight Number vs. Launch Site

We can see a that different launch sites have a different pattern in success rate when compared to flight numbers

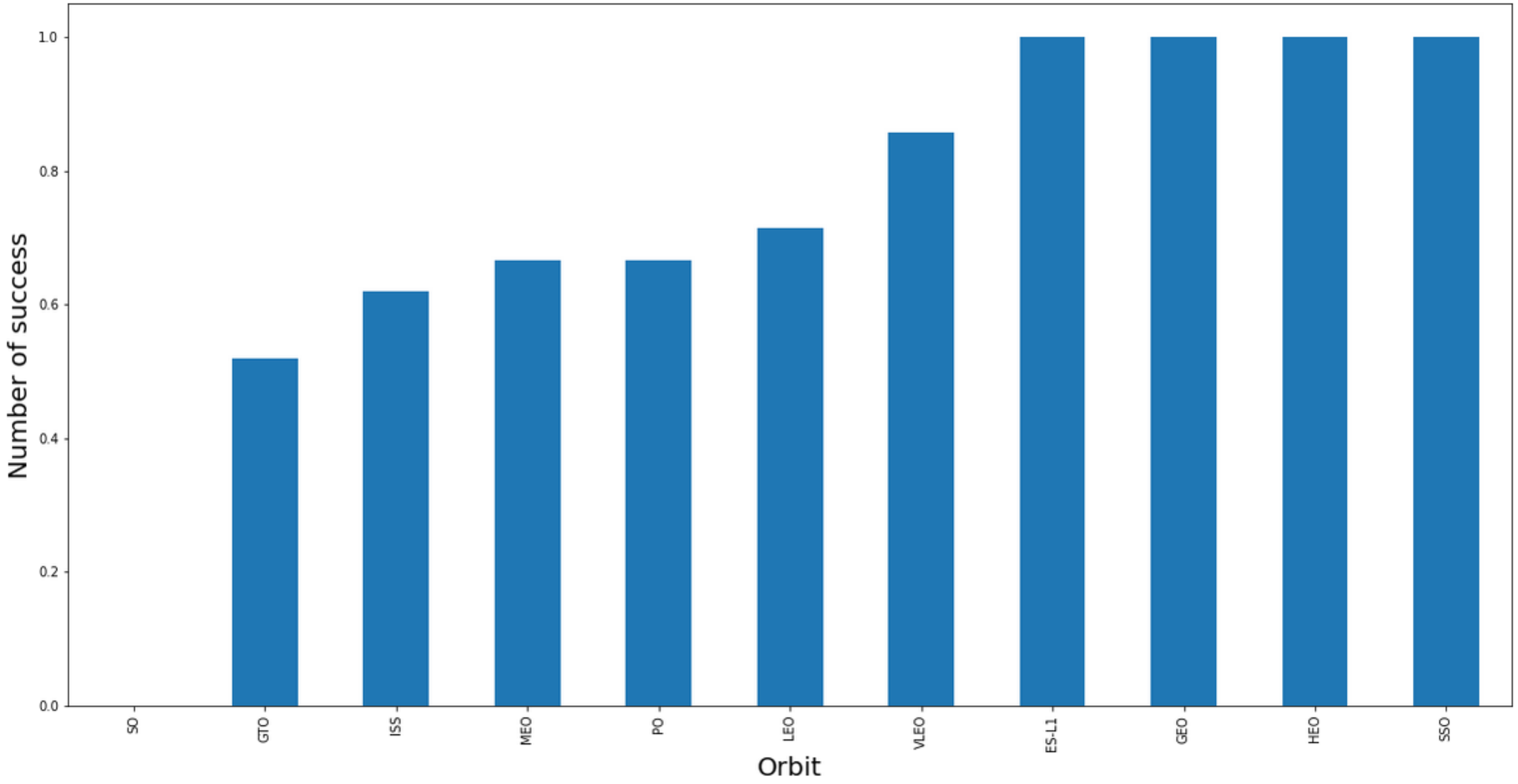


Payload vs. Launch Site

We can see a high success rate at either very high payloads or at load payloads in launch site KLC LC 39A

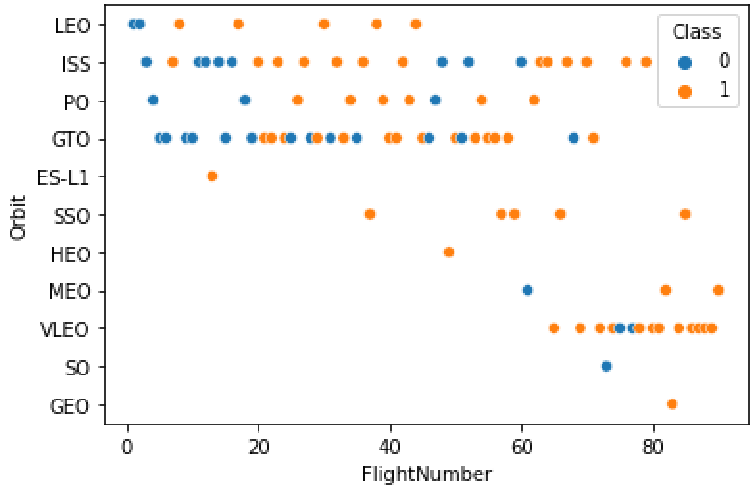


Success Rate vs. Orbit Type



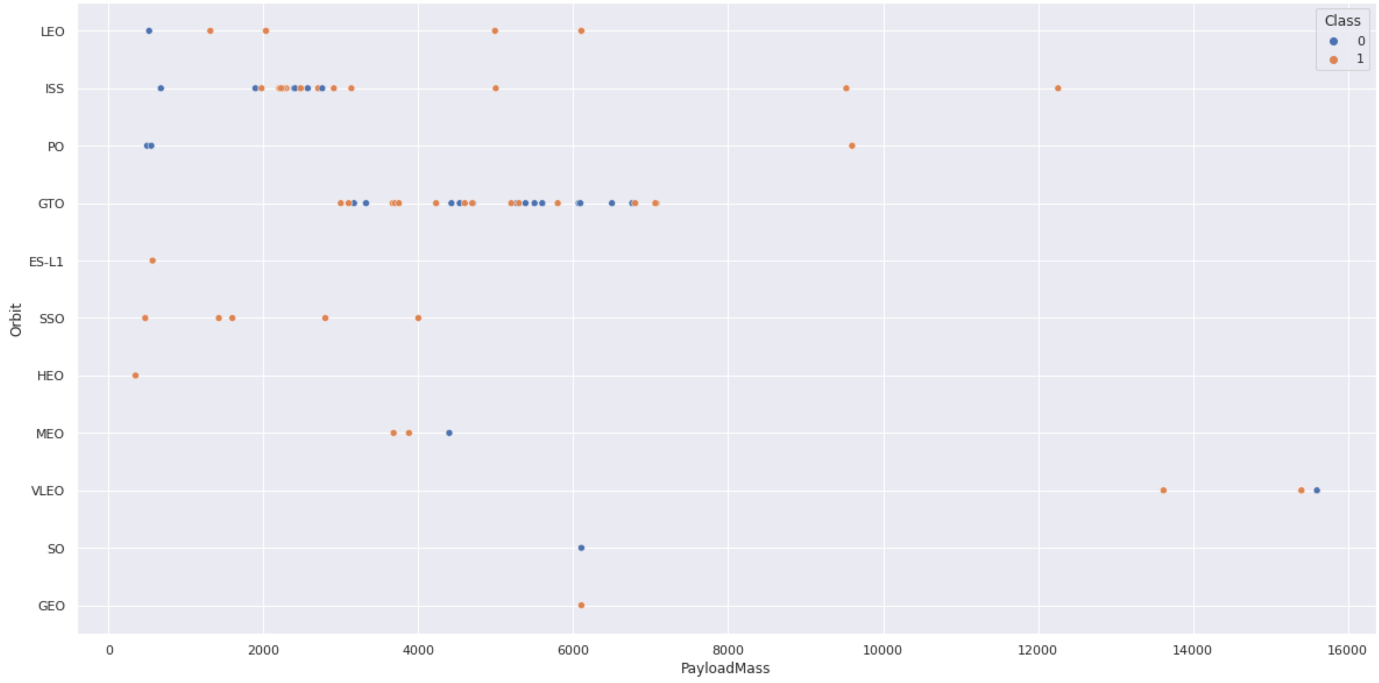
Flight Number vs. Orbit Type

We can see a that different orbits have a different pattern in success rate when compared to the flight number



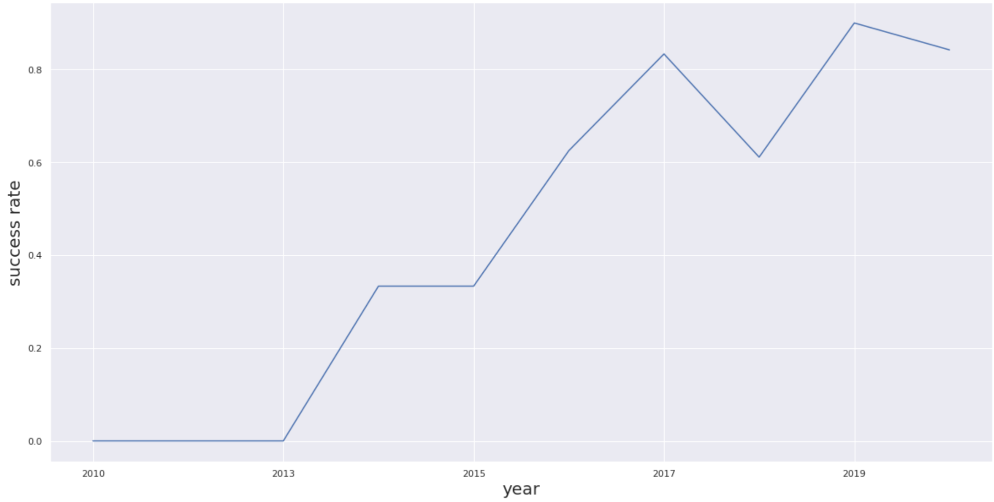
Payload vs. Orbit Type

For every orbit, the high success rates are dependent on payload. A pattern is visible



Launch Success Yearly Trend

As time advances so does the success rate. But there are some hiccups that are visible and would be interesting to look into



All Launch Site Names

In total there are 4 launch sites, with 2 starting with CCA and it will be seen later that they are indeed in the same area

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

Launch Site Names Begin with 'CCA'

These are 5 launch information that were initiated in a launch site 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 table represents the total payload by NASA

Sum of payload by NASA
99980

Average Payload Mass by F9 v1.1

The average payload carried by an F9 V1.1 booster

Average payload by booster version F9 V1.1
2534

First Successful Ground Landing Date

Here is displayed the first successful ground landing

First successful ground landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

This table represents the boosters that had a successful drone ship landing with a 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

This table represents the total success and failures of missions, I kept one of the successes as it is since it might be useful with the fact that the payload is unclear

Mission outcome	**Number per mission outcome**
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

These are all the boosters that carried the maximum payload

Booster version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

Failed 2015 Launch Records

Here are all the booster versions and their launch site that have had a failure

Booster version	Launch site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40
F9 v1.1 B1018	CCAFS LC-40

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

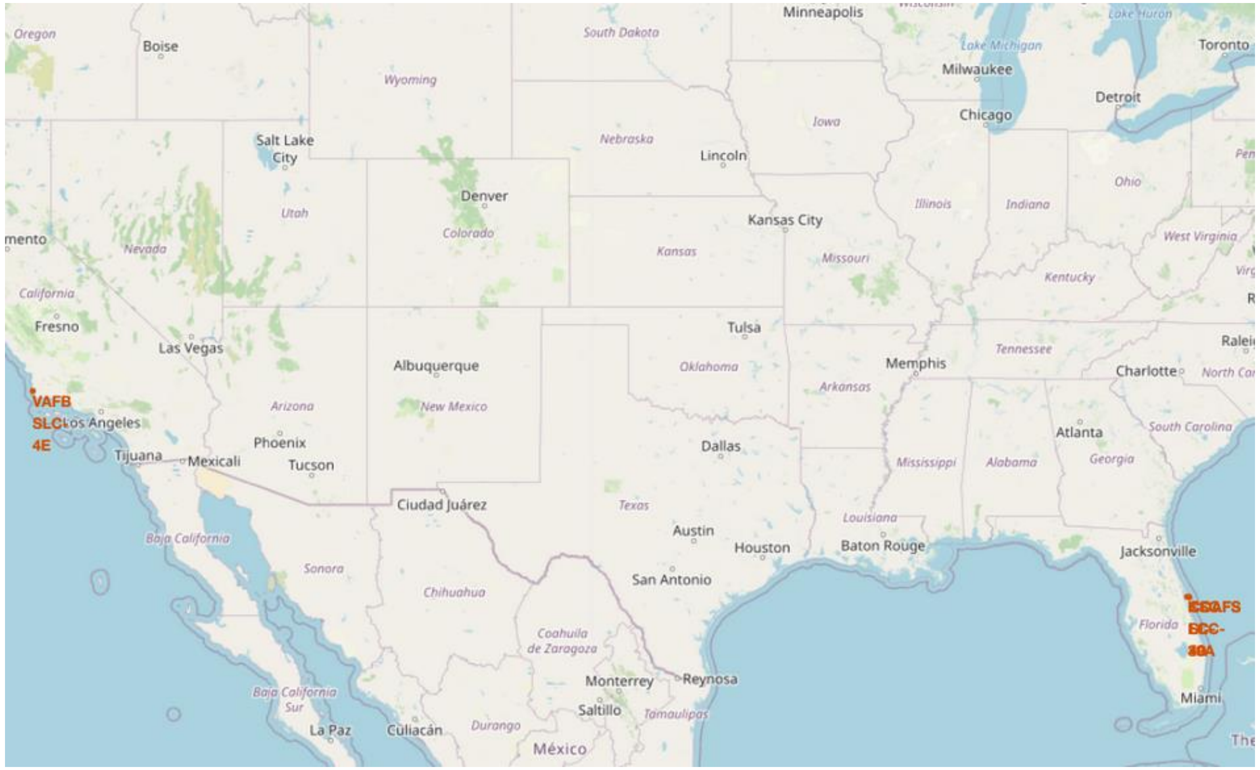
Here is listed all the landing outcomes and how many times each happened

Landing outcome	Number of that outcome
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

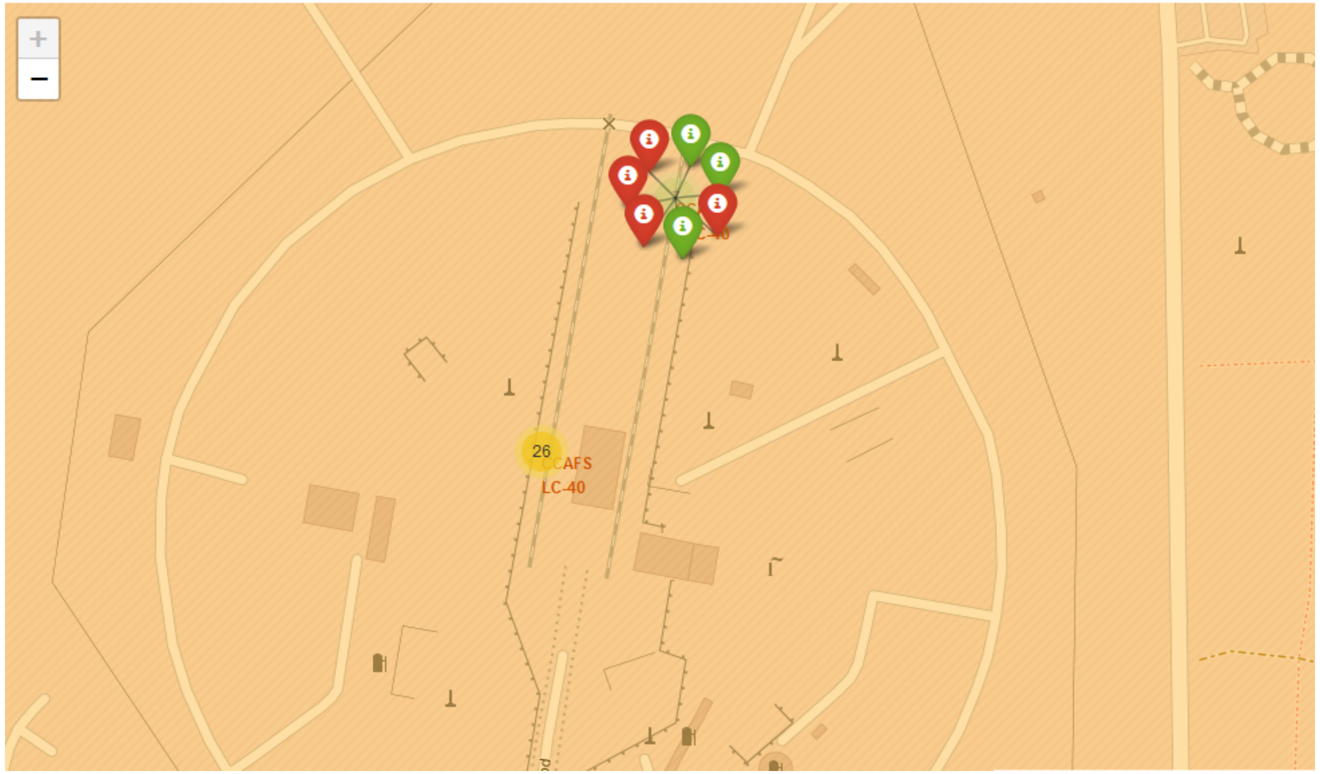
Launch Sites Proximities Analysis

All launch sites



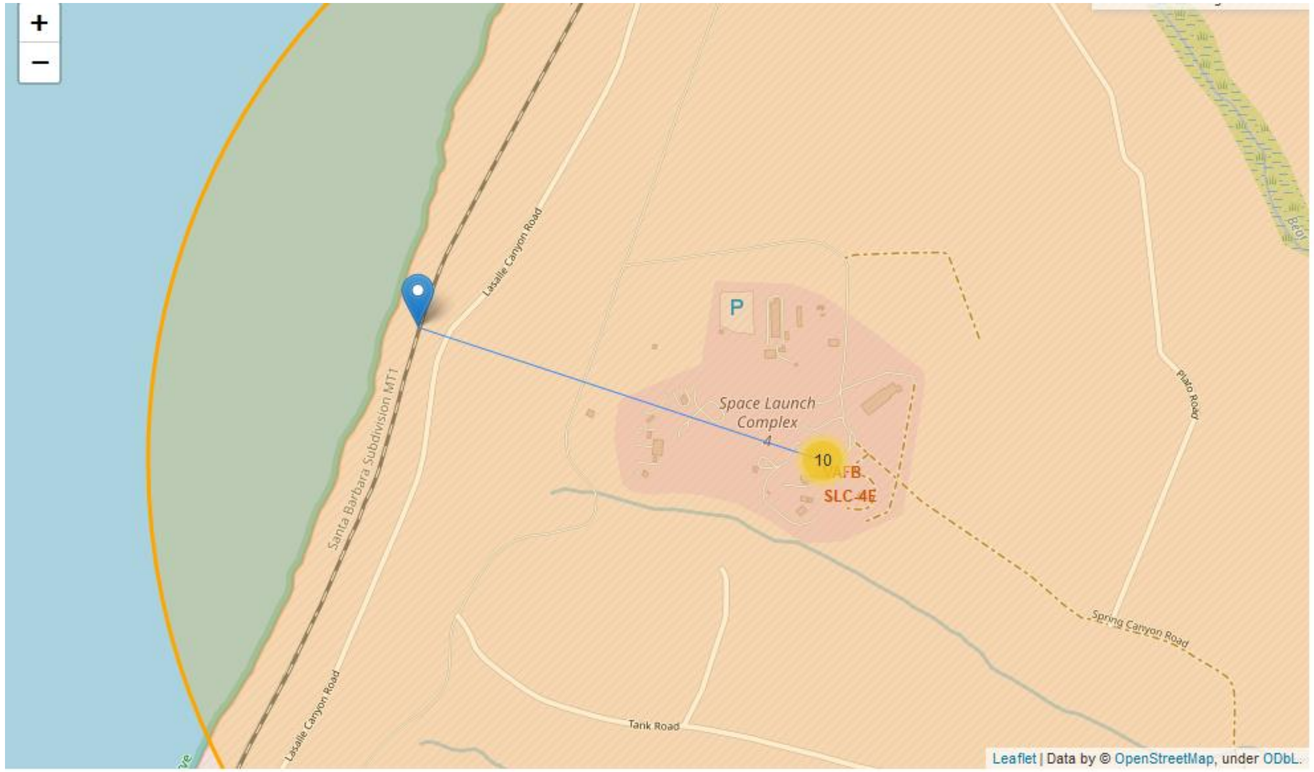
- We can see that all launch sites are situated in the united states of America
- They are only located in 2 area and there are 4 launch sites, which means that some are nearby

Success-failure marker clusters



- Red popups represent failure
- Green popups represent success
- For more details refer to the notebook and navigate through the map

Distance to other locations



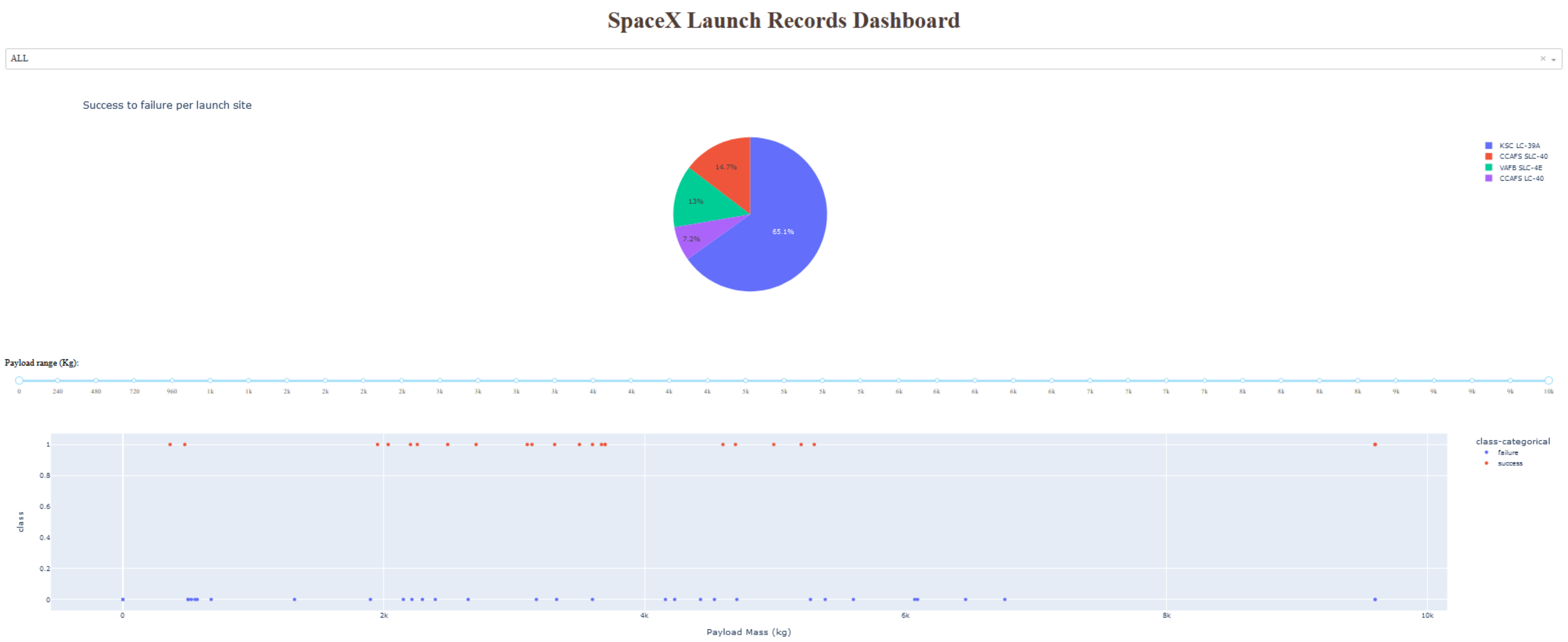
- VAFB SLC 4E is quite close to the cost, and that cannot be said for all other launch sites
- A better understanding of the impact of being on a coast might be large and should be looked into

Section 4

Build a Dashboard with Plotly Dash

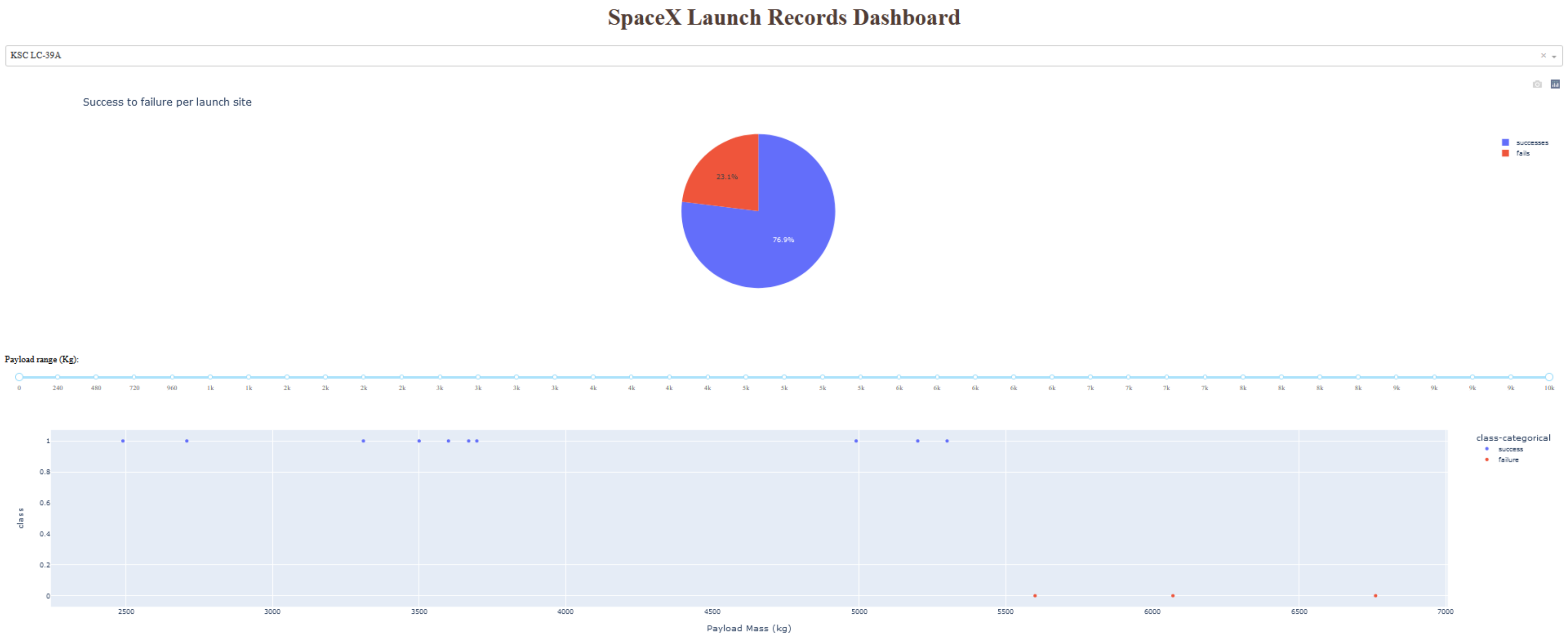
Launch success count for all sites

Most successful launches were launched from KSC LC 39A. Yet this can be due to the most launches being launched from that launch site



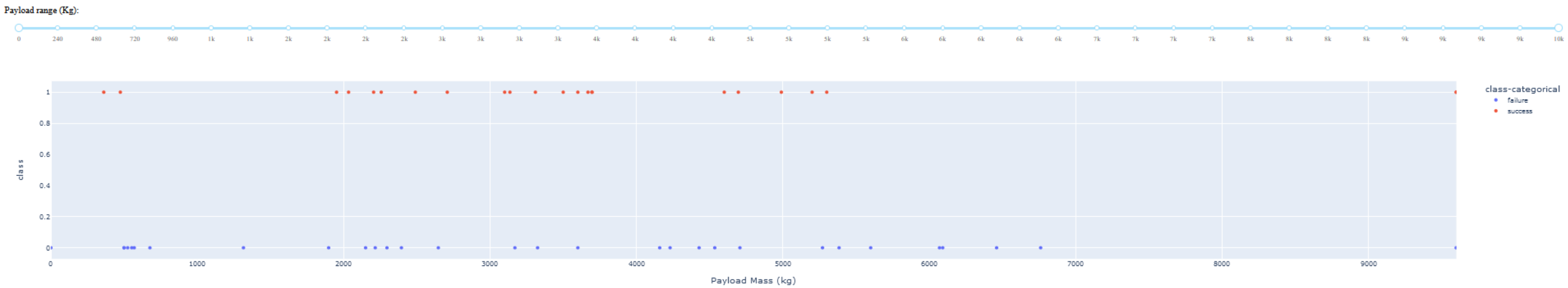
Launch site with highest the launch success ratio

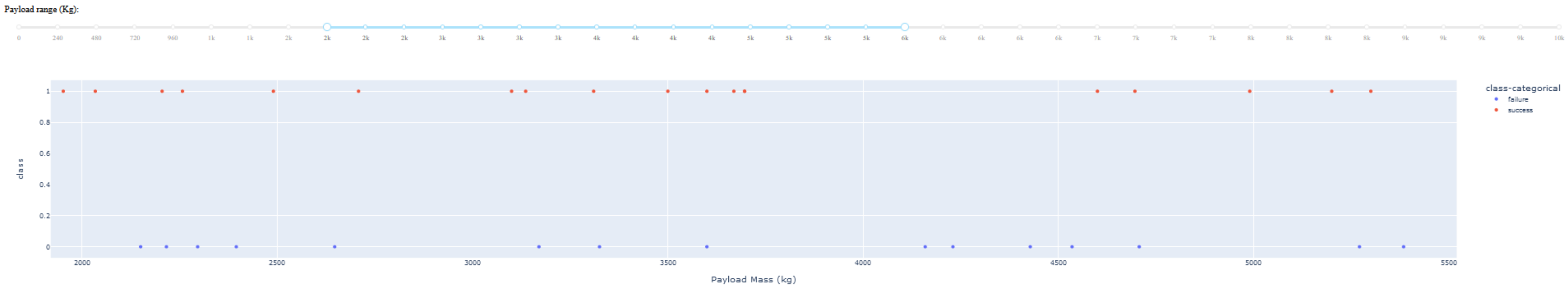
This launch proves our original assumptions that KSC LC 39A has the largest amount of success and success ratio



Payload vs. Launch for different payload ranges

- In the first picture we see that the payload is set over the entire range
- As we concentrate the range where the success is optimal, we see that we concentrate ourselves towards the middle



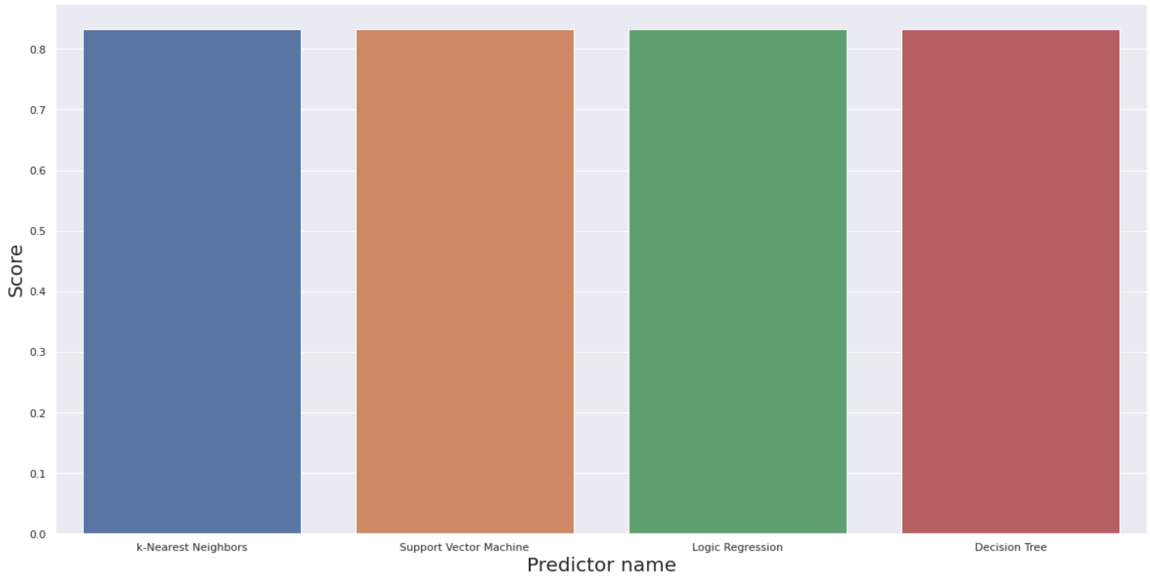


Section 5

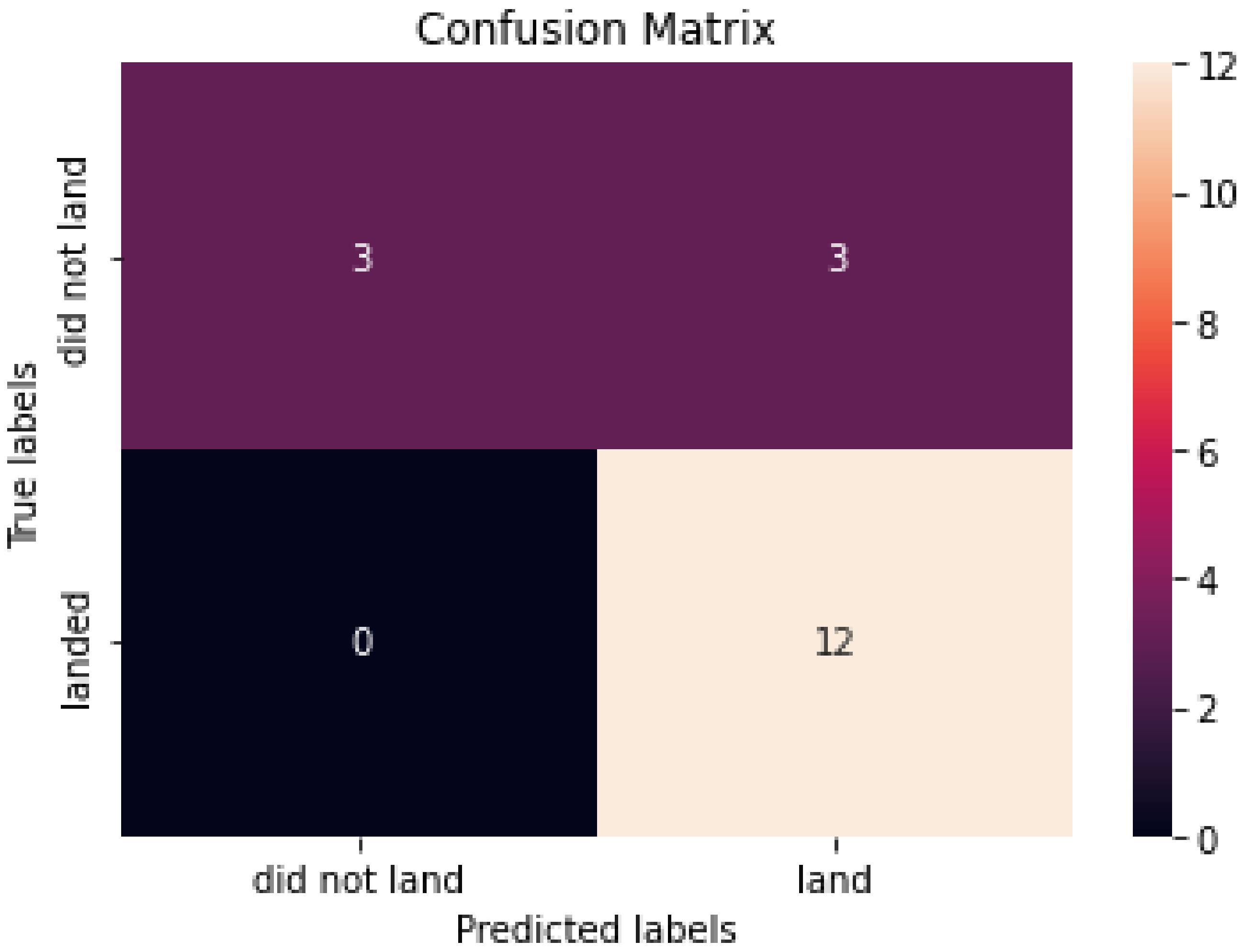
Predictive Analysis (Classification)

Classification Accuracy

- It seems regardless of the predictor, the value is the same for all



Confusion Matrix



Conclusions

When considering the optimal conditions for a successful recoup one must:

- Work with a payload that is low to medium optimally
- Mission to orbits ES-L1, GEO, HEO, SSO
- Use the launch site KSC LC 39A
- For the prediction method, it seems all methods return equal prediction scores
- Testing for the optimal prediction method should be continued as it might change in the future

Thank you!