IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexe Razvan
6/19/2023

Outline

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

Executive Summary

- Methodologies:

  Data collection via webscraping and SpaceX API calls

  Data wrangling, visualization and EDA

  Machine Learning Predictions

- Results:

  The best attributes were determined that predict the desired outcome.

  A predictive model was conceived.

Introduction

- The purpose of this project is to take information from SpaceX and attempt to use it to predict which attributes influence launch success outcomes.
- What are the most favourable locations for those outcomes?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected by making calls to the SpaceX REST API and by scraping the wikipedia launches page.
- Perform data wrangling
  - The column Outcome was used as the classification variable (1 for successful landing, 0 otherwise).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different models were scored based on the predictive accuracy they displayed.

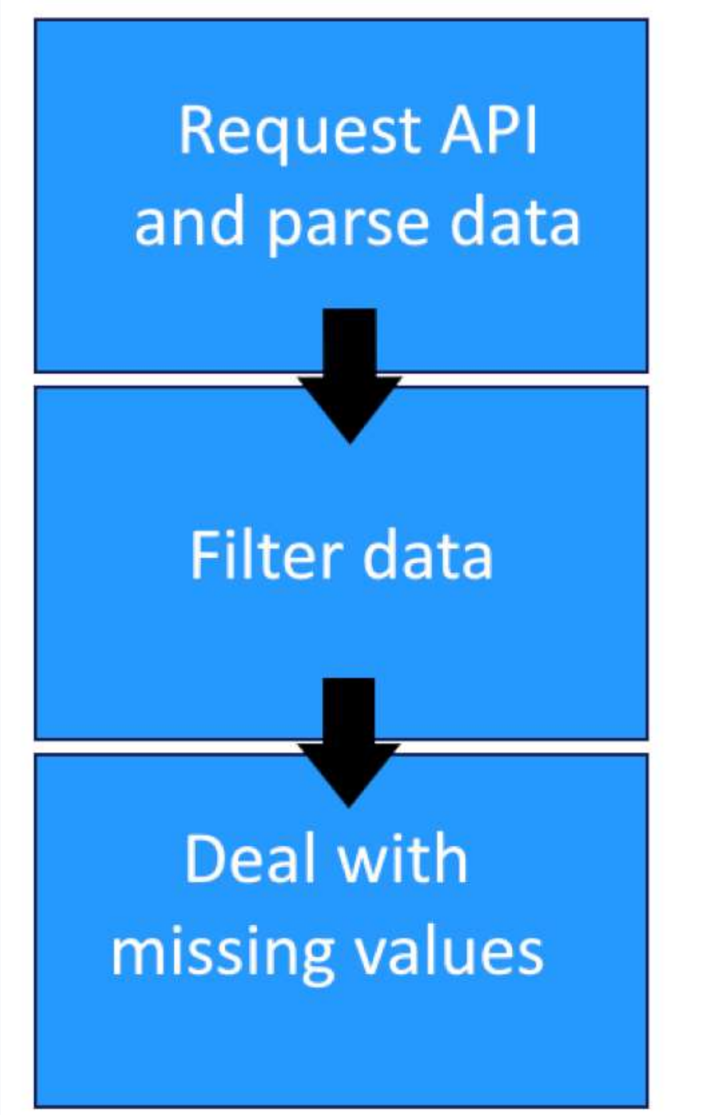
Data Collection

- Data was collected by making calls to the SpaceX REST API and by scraping the wikipedia launches page.

Data Collection – SpaceX API

- The presented flowchart presents the process of obtaining data with calls to the SpaceX API.

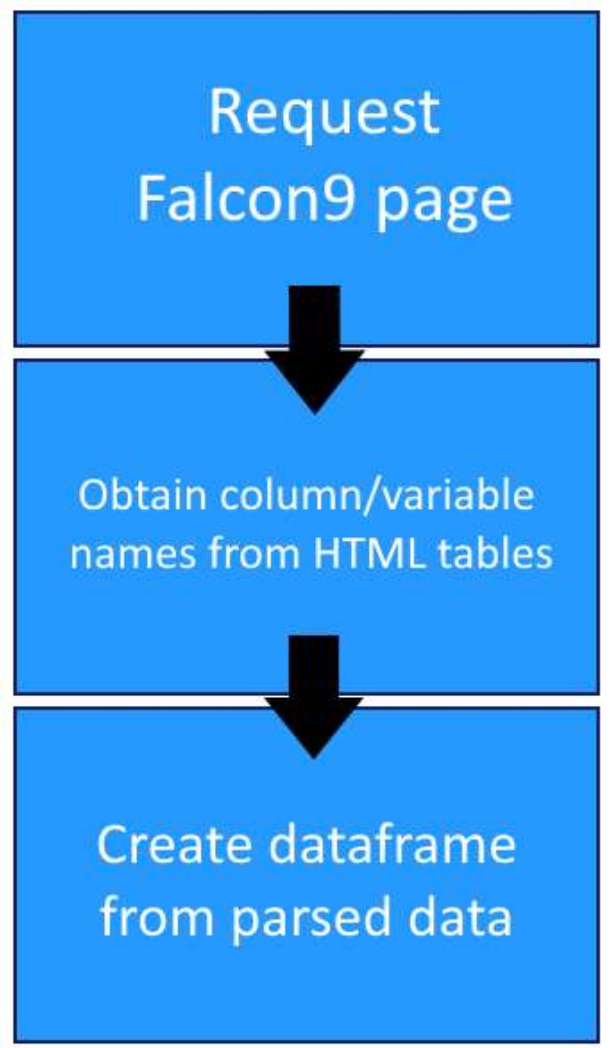
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-collection.ipynb



Data Collection - Scraping

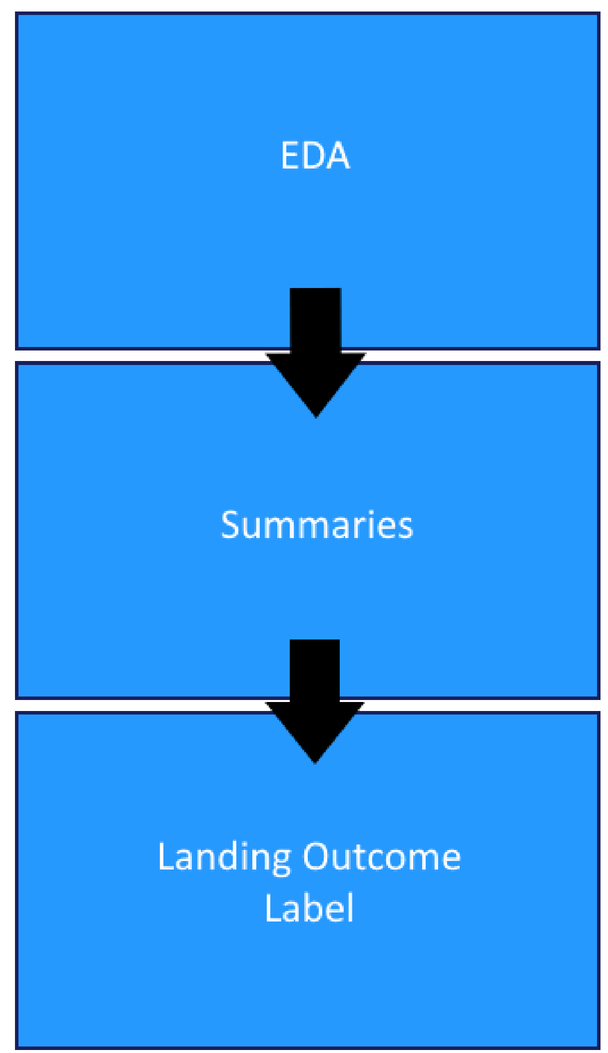
- Via webscraping we obtained information about SpaceX launches from the launches wikipedia page.

- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-webscraping.ipynb



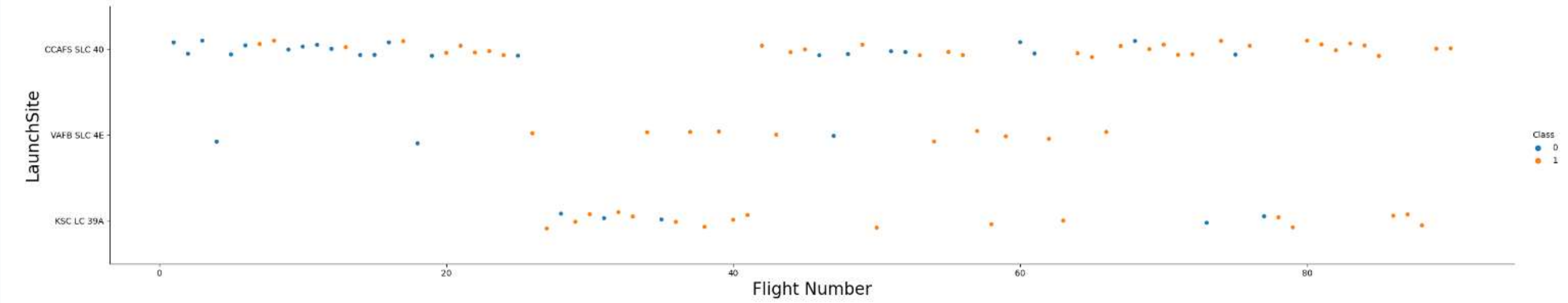
Data Wrangling

- Initial EDA
- Summaries were conceived
- The Landing Outcome Label is created from the Outcome column
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-wrangling.ipynb



EDA with Data Visualization

- We wanted to see a visual representation of the relatioship between certain features of the the dataset.
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-edav.ipynb



EDA with SQL

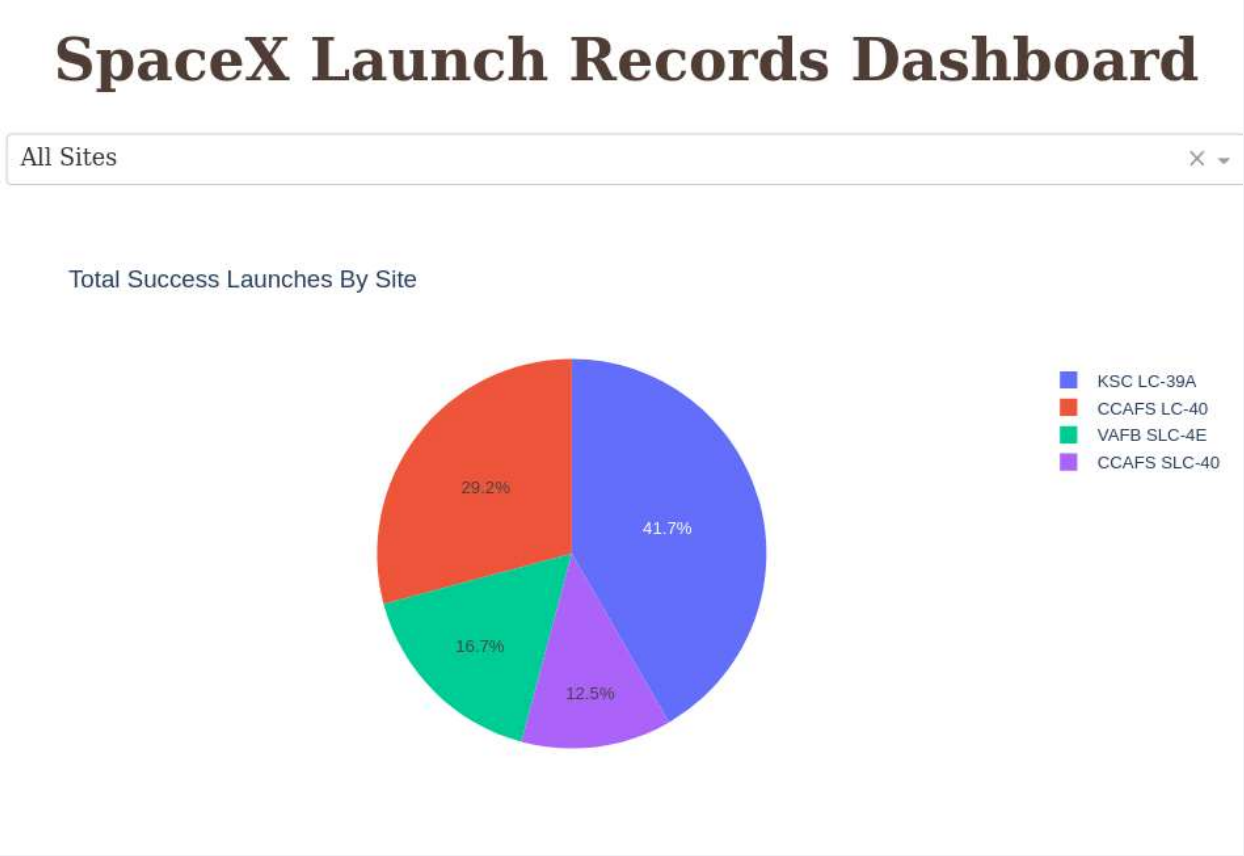
- SQL Queries Executed:
  - Names of unique launch sites
  - Top 5 launch sites that begin with “CCA”
  - Total payload mass
  - Average payload mass
  - First successful landing outcome date
  - Names of boosters which have success in drone ship and within a certain payload mass
  - Total number of failed and successful mission outcomes
  - Names of boosters which carried the maximum amount of payload mass
  - Failed landing outcomes
  - Descending ordering of the count of landing outcomes
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-eda.ipynb

Build an Interactive Map with Folium

- With Folium we were able to create circles and markers on certain locations.
- We used this in order to visualize the geographical importance of certain locations.
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-folium.ipynb

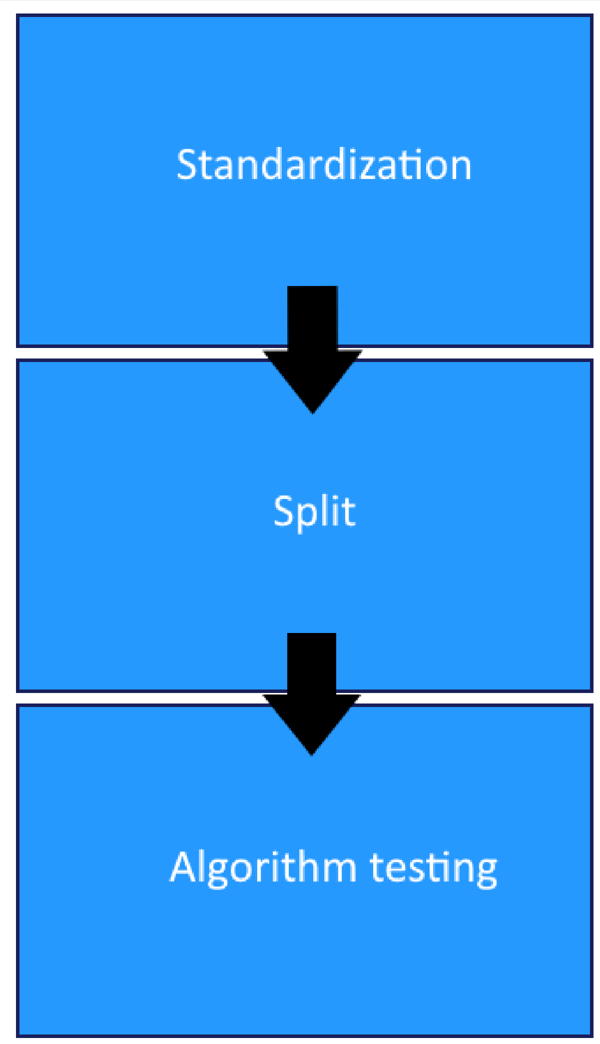
Build a Dashboard with Plotly Dash

- Plotly was used in order to generate interactible pie charts.
- We use these interactible plots in order to visualize specific data that can influence the importance of certain outcomes.
- Source code: https://github.com/RazvanAlexe/SpaceX-IBM-Capstone/blob/main/spacex-data-plotly.py



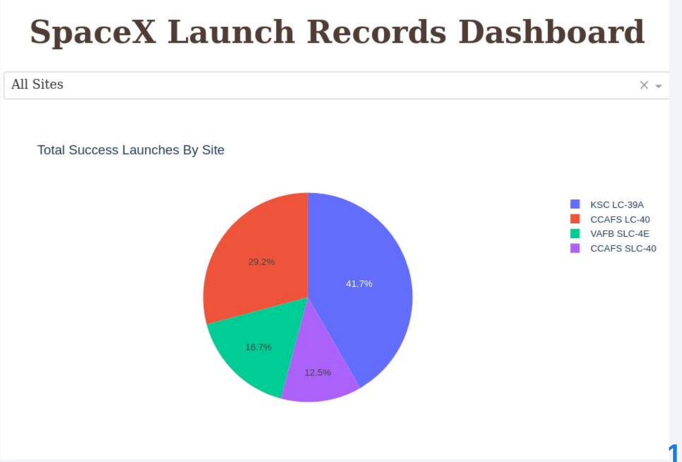
Predictive Analysis (Classification)

- The required data was standardized
- The data was split into training and test sets
- Regression, decision trees, vector machines and K nearest neighbor algorithms were tested to score predictive accuracy.



Results

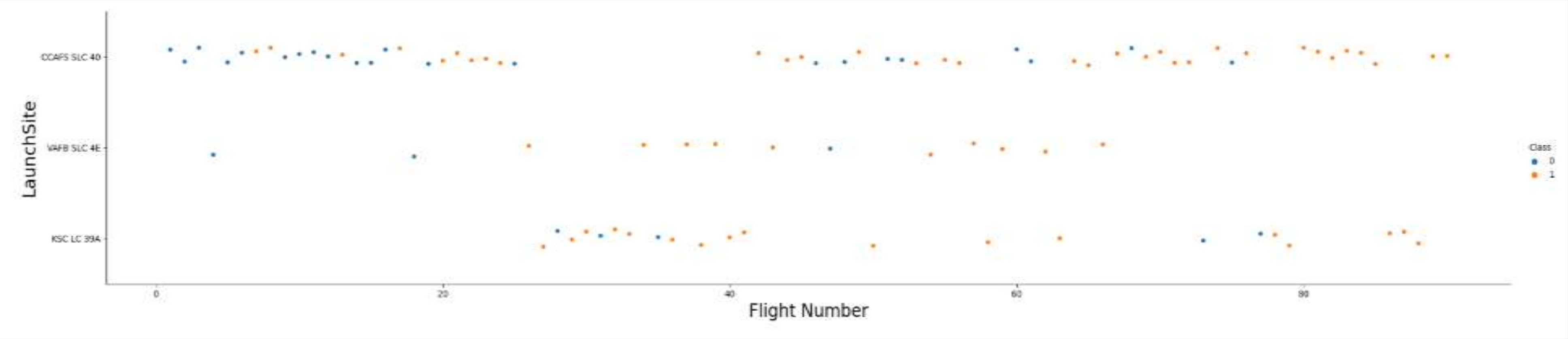
- EDA assisted us with identifying from the given data the names of the launch sites, the payload intervals, the booster versions and how they influence the ratios of success.
- Graphical representations further solidified our understanding of ideal conditions for a launch.
- Decision classifiers are ideal for predicting the desired outcomes.



Section 2

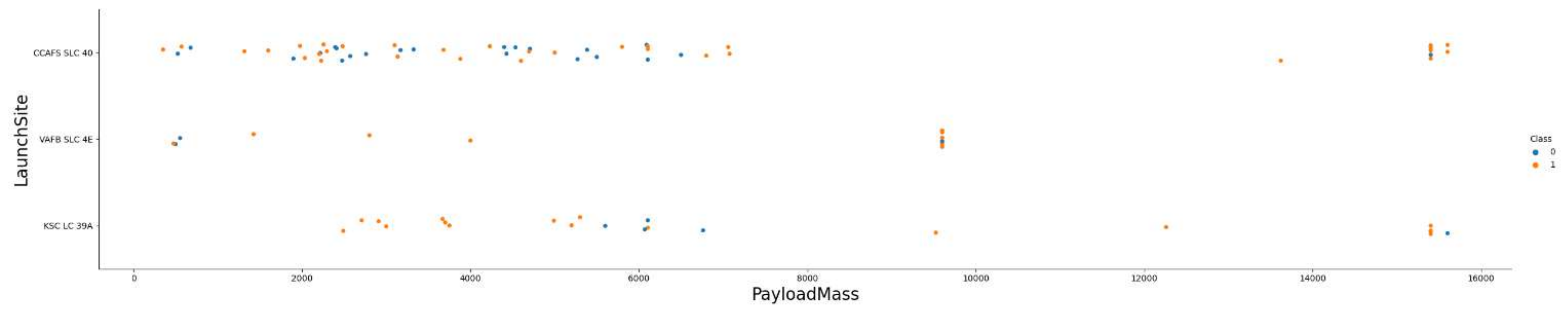
Insights drawn from EDA

Flight Number vs. Launch Site



- Based on the Flight Number versus Launch Site graph we can identify the ideal launch site: CCAFS SLC 40

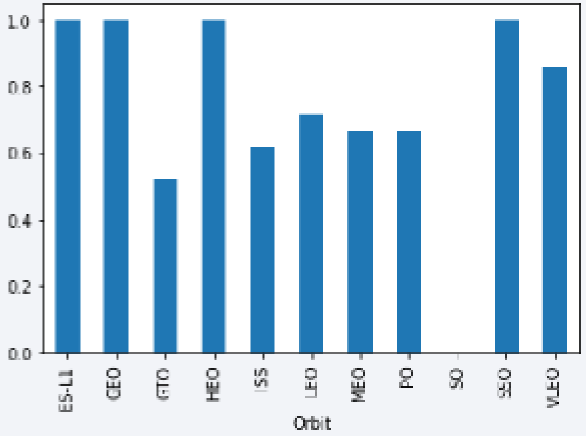
Payload vs. Launch Site



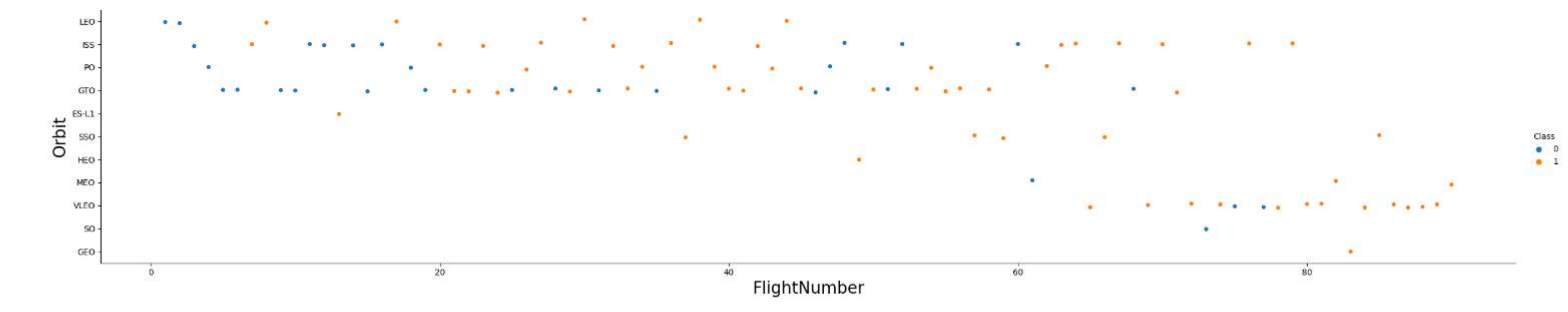
- Based on the Payload versus Launch Site graph we can identify the most suited sites for high payload launches: CCAFS SLC 40.

Success Rate vs. Orbit Type

- Looking at the Orbit Type versus Success Rate bar chart we can identify the ideal Orbit destinations.



Flight Number vs. Orbit Type



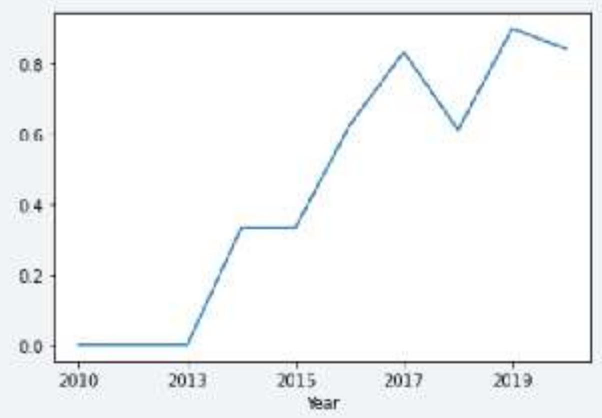
- Based on the Flight Number versus Orbit Type graph we can more clearly determine ideal Orbit for successful flights: VLEO.

Payload vs. Orbit Type

- Looking at the Payload versus Orbit Type graph we have additional confirmation of VLEO being the ideal Orbit.

Launch Success Yearly Trend

- Graphing the Launches using the Year attribute we can see the increase in success with the passage of time as well as identifying the drop in success ratio in the year 2018.



All Launch Site Names

- Using EDA we can identify all launch sites.

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

Launch Site Names Begin with 'CCA'

- Using EDA we can explore information about the CCAFS launch sites.

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 attemp

Total Payload Mass

- A total payload EDA can help us better understand the requirements for a successful launch.

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- Using EDA we can determine the mean payload for a F9 booster.

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- Using EDA we can determine the date of the first successful launch. Having a better historic understanding of launches can help us understand what it takes for success.

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Using EDA we can determine the ideal boosters for the desired payload intervals.

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Using EDA we can keep track of launch success ratios.

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanatic

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

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Using EDA we can determine the failed launch booster types as well as launch sites.

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- Using EDA we can analyze historic data regarding the success and failures of launches between given dates.

Landing Outcome	Occurrences
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

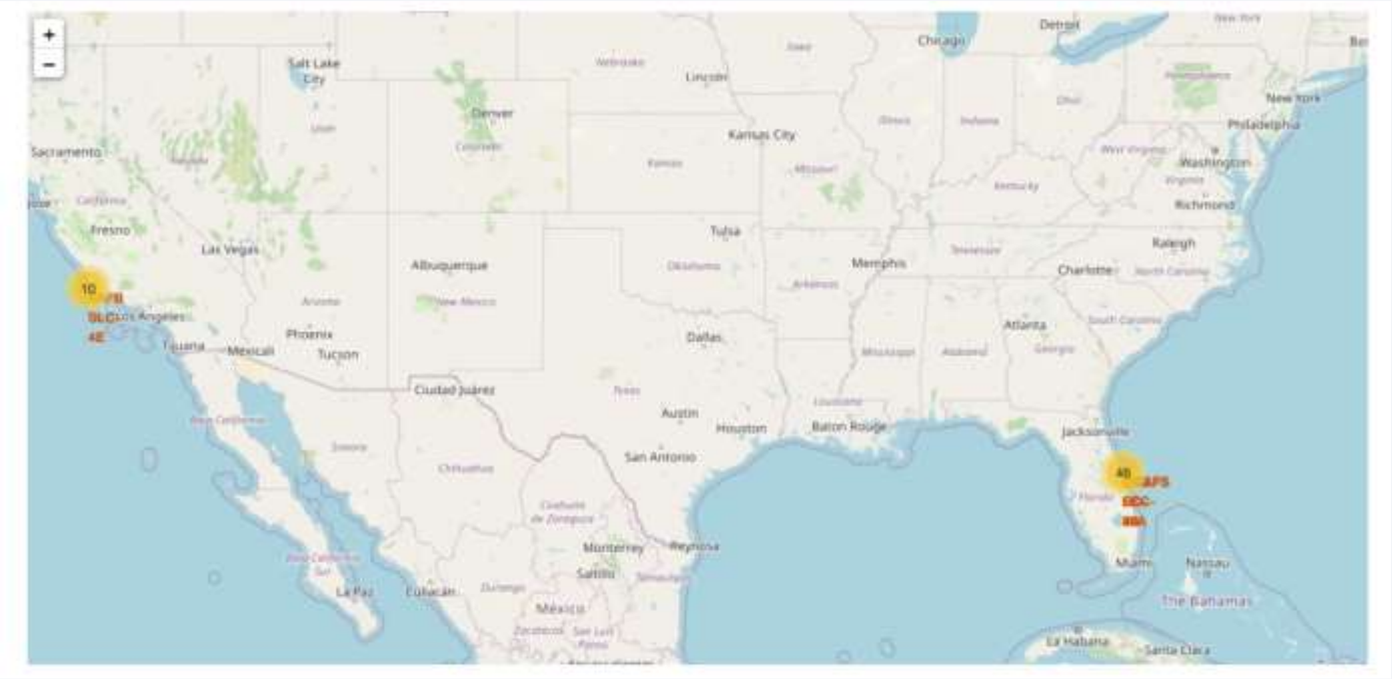
Launch Sites Proximities Analysis

Folium launch site map

- With Folium we can generate a interactible map of all the launch sites.

Folium launch numbers

- We can display, along the location of the sites, the number of launches that took place in said locations



Folium Proximity Maps

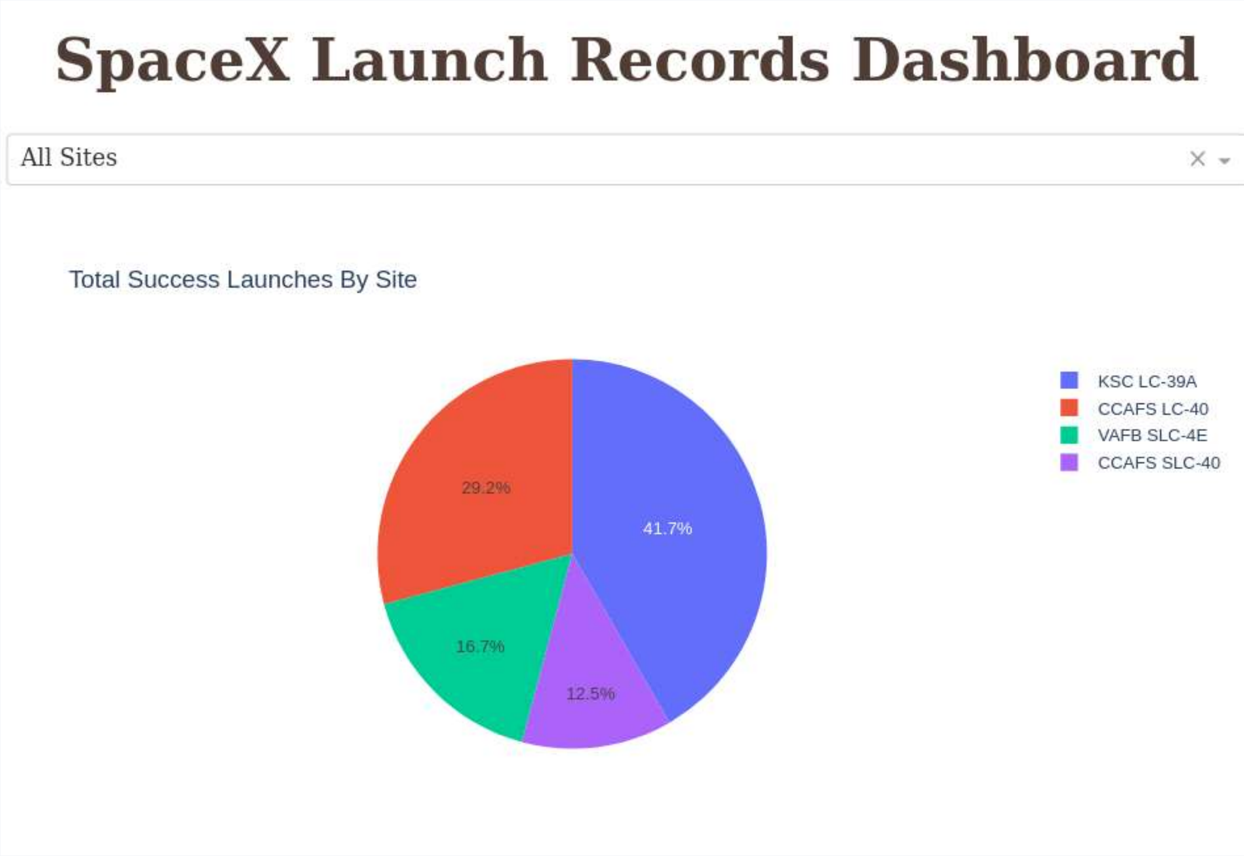
- The generated maps also provide proximity to logistical elements that are key to determining optimal outcomes (next to roads, railways etc.)

Section 4

Build a Dashboard with Plotly Dash

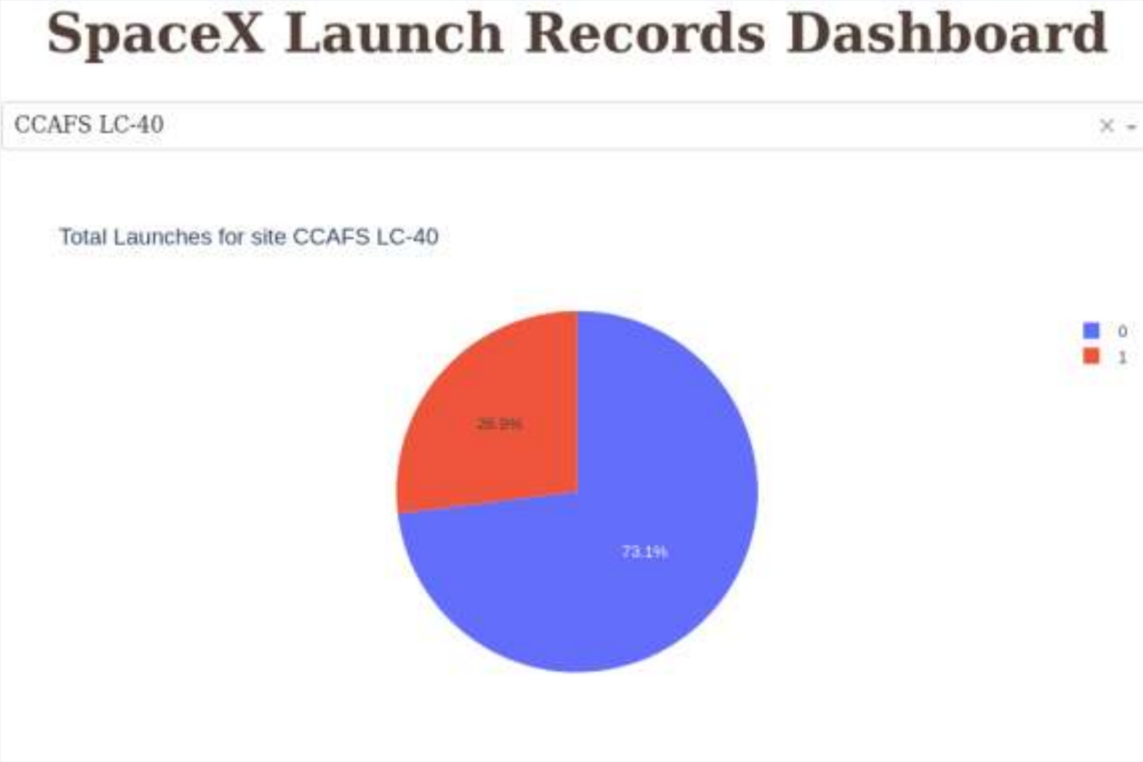
SpaceX Launch Records Dashboard

- Visualizing the LaunchSite/Outcome pie chart we can easily identify the most well suited launch site for succesful launches.



SpaceX Total Launches Dashboard

- From the adjacent pie chart we can showcase how the launch site CCAFS LC-40 has a very high percentage of successful launches out of total launches.



Payload and booster dashboard

- Given the adjacent graph we can identify the ideal payload interval (<6k kg) as well as the ideal booster version (FT).

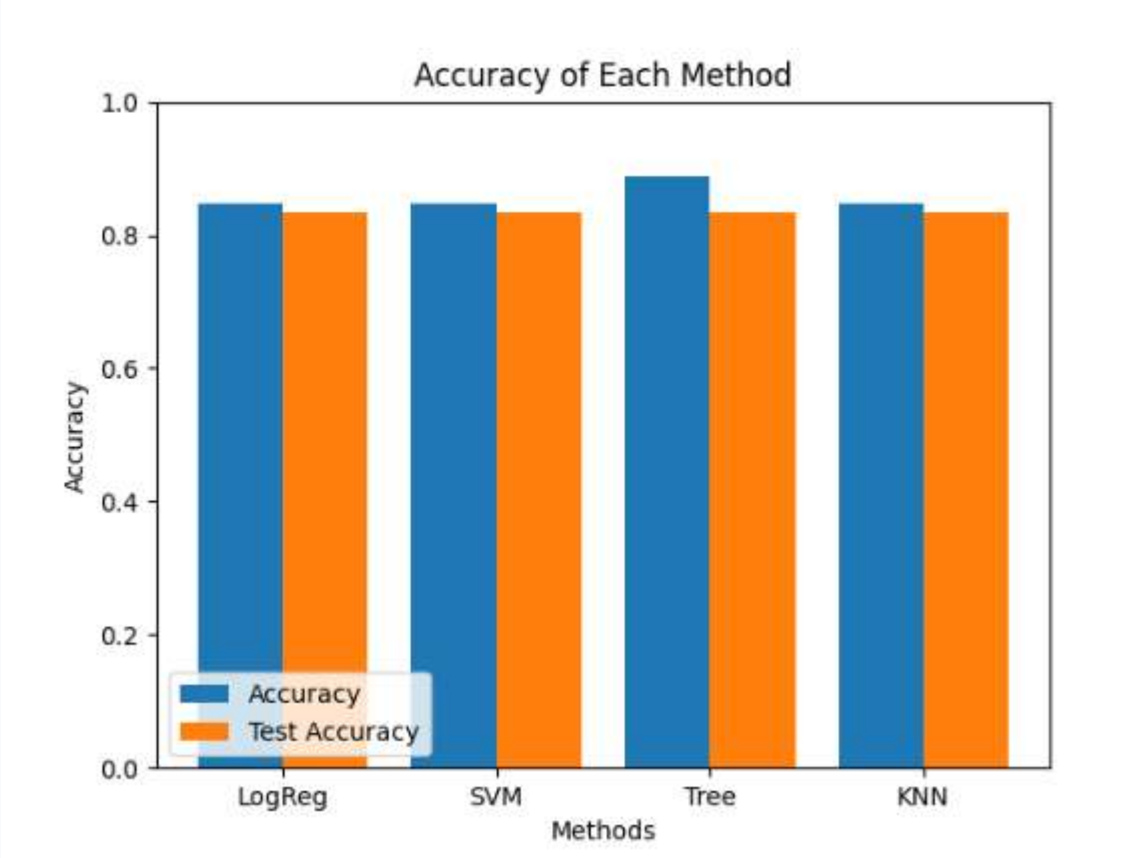
Section 5

Predictive Analysis (Classification)

Classification Accuracy

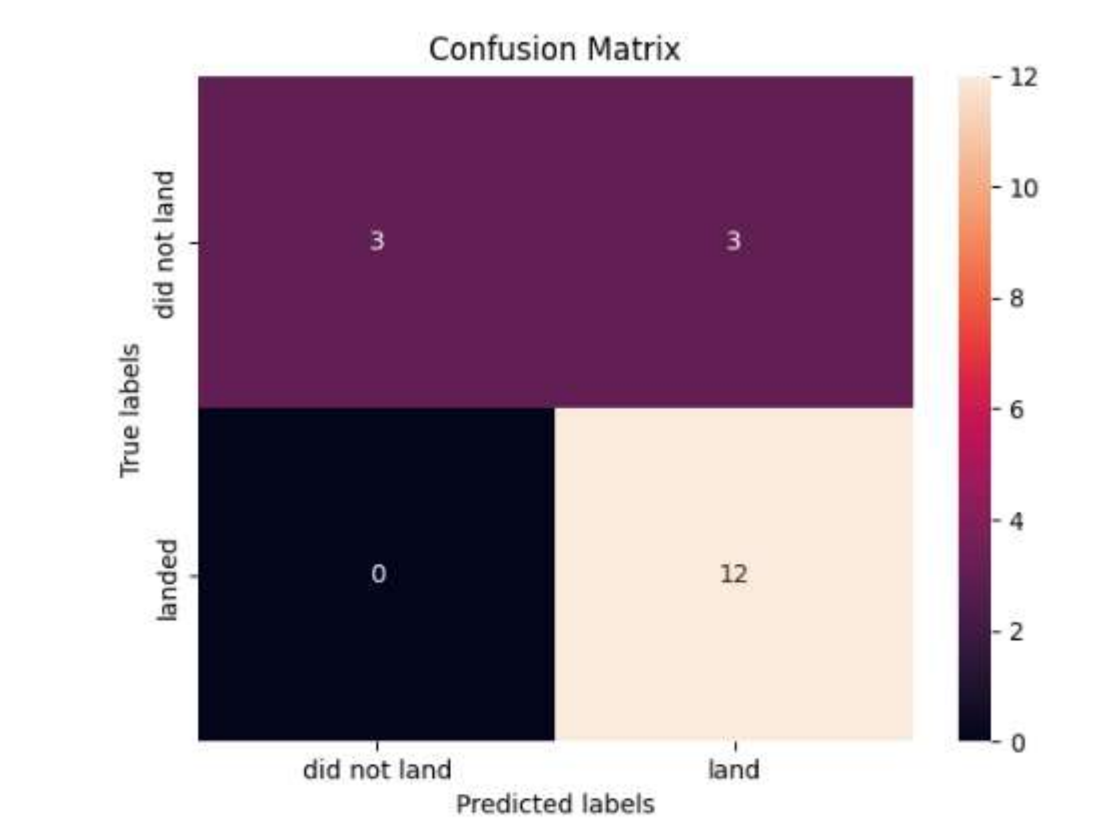
- A bar chart was conceived to showcase the accuracies of each type.

- The decision tree algorithm showed the best results when it came to the prediction accuracy.



Confusion Matrix

- The confusion matrix for the Decision Tree classification shows the amount of true positives and true negatives obtained when attempting to predict. Since it is the best performing algorithm these numbers are the highest.



Conclusions

- While Decision Tree classifiers are well suited for predictions further development of models would be ideal.
- We have plenty of data that support that the certain variables that influence the success ratios of a launch site are logistical and geographical.
- Historic analysis of data offer us a optimistic view on the progress of launches.

Appendix

Thank you!