IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Veronika Ruppert
2023-02-26

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using web scraping and SpaceX API
  - Explanatory Data Analysis including data wrangling, data visualization and interactive visual analytics
  - Machine Learning Prediction
- Summary of all results
  - It was possible to collect valuable data from public sources
  - Explanatory Data Analysis allowed us to identify which features are the best to predict success of launchings
  - Machine Learning Prediction showed the bet model to predict which characteristics are important to drive this opportunity by the best way using all collected data

Introduction

- Project background and context
  - The objective is to evaluate the viability of the new company SpaceY to compete with SpaceX
- Problems you want to find answers
  - Where is the best place to make the launches
  - Find the best way to estimate the total cost for launches by predicting successful landings of the first stage of rockets

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from two sources:
    - SpaceX API
    - WebScraping
- Perform data wrangling
  - The collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - The collected data was normalized, split in training and test datasets and then evaluated by four classification models, and then using different combinations of parameters to evaluate the accuracy of each model

Data Collection

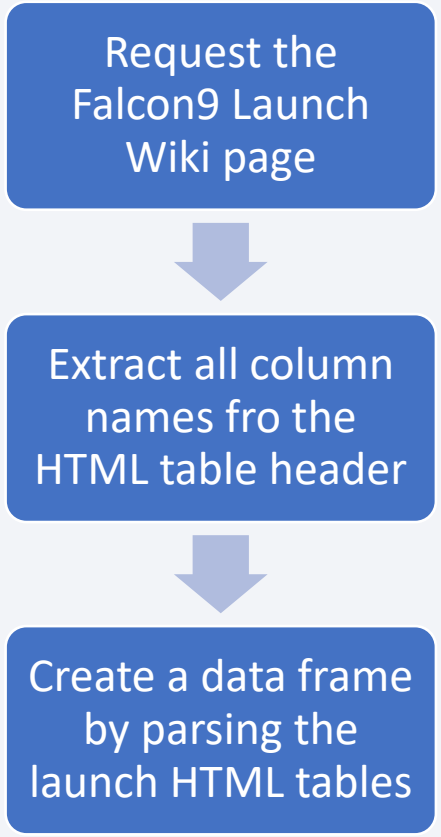
- Data sets were collected from SpaceX API and from Wikipedia with web scraping technics.

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and used
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/3775a8723097c3b3659b3758ab64d85e1b76ebd2/jupyter-labs-spacex-data-collection-api.ipynb

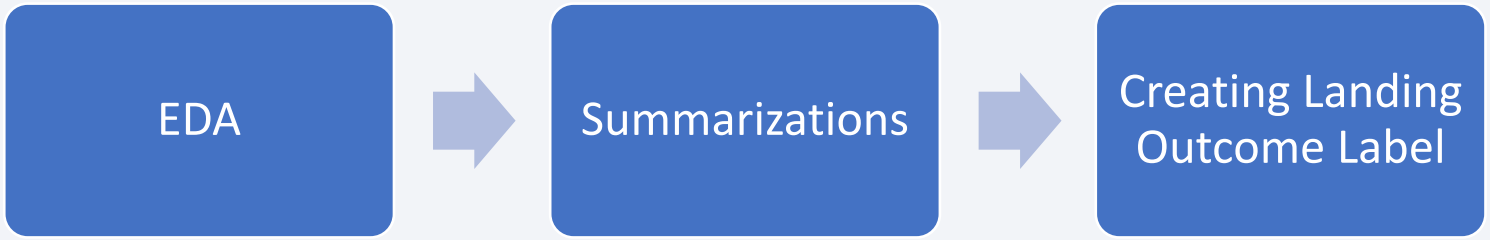
Data Collection - Scraping

- Data was extracted from Wikipedia as the second data source according to the flowchart
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/3775a8723097c3b3659b3758ab64d85e1b76ebd2/jupyter-labs-webscraping.ipynb



Data Wrangling

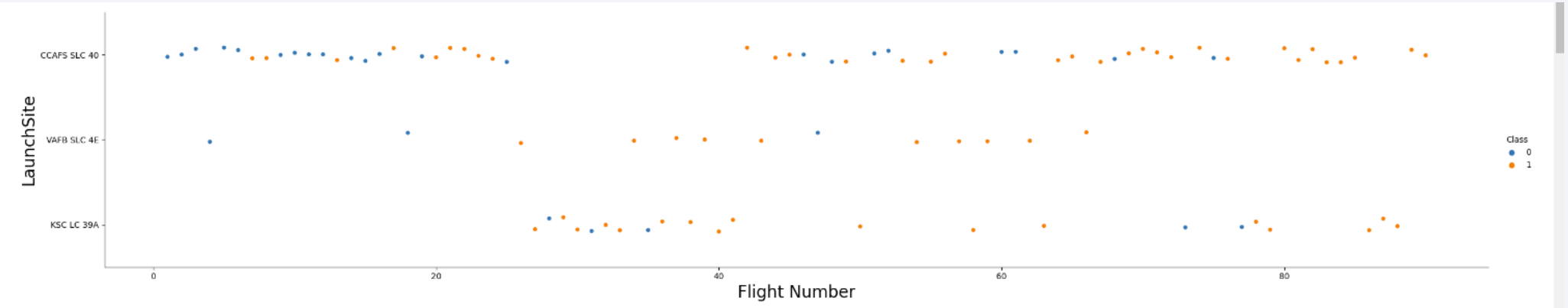
- Initially some Explanatory Data Analysis was performed on the dataset
- Then summaries launcher per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated



- https://github.com/froniR/ToolsForDataScienceCertificate/blob/50a7edb22bf827e9fb7d515144caab901928370b/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pairs of features



- https://github.com/froniR/ToolsForDataScienceCertificate/blob/2486bbe1c082ec4731206c277fc76bb7333bde4d/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- 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 succesful landing outcome in ground pad was acheived.
- 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/d96d8cd52b96b8f3cc26b14926c19457501d1bc2/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb

Build an Interactive Map with Folium

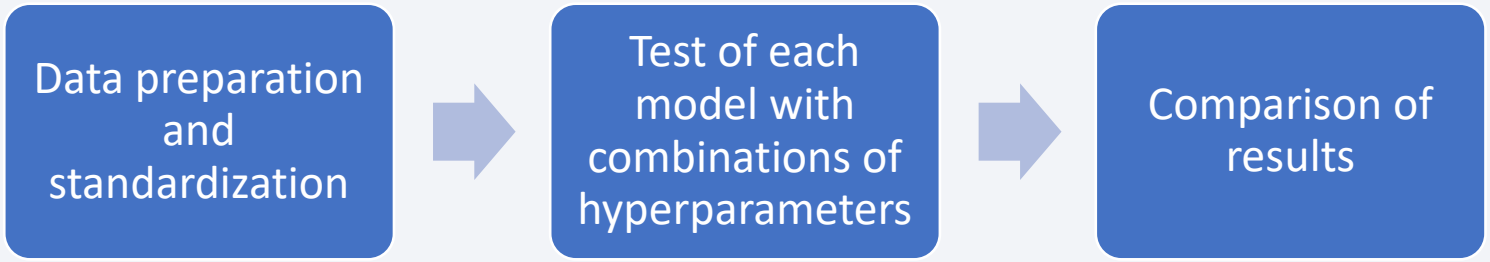
- Markers, circles, lines and marker clusters were used with Folium Maps
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
  - Marker clusters indicate groups of events in each coordinate like launch sites
  - Lines are used to indicate distances between two coordinates
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/8dca80a12edc6d75c6a20a7cee7472067871ff5f/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- I used the following graphs and plots to visualize data
  - Percentage of launches by site
  - Payload range
- This combination helped me to quickly analyze the relation between payloads and launch sites and to identify where the best place to launch according to payloads is
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/9a81a96305db8f8233481bc99e33896cb3d9f26a/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared; logistic regression, support vector machine, decision tree and k-nearest-neighbours



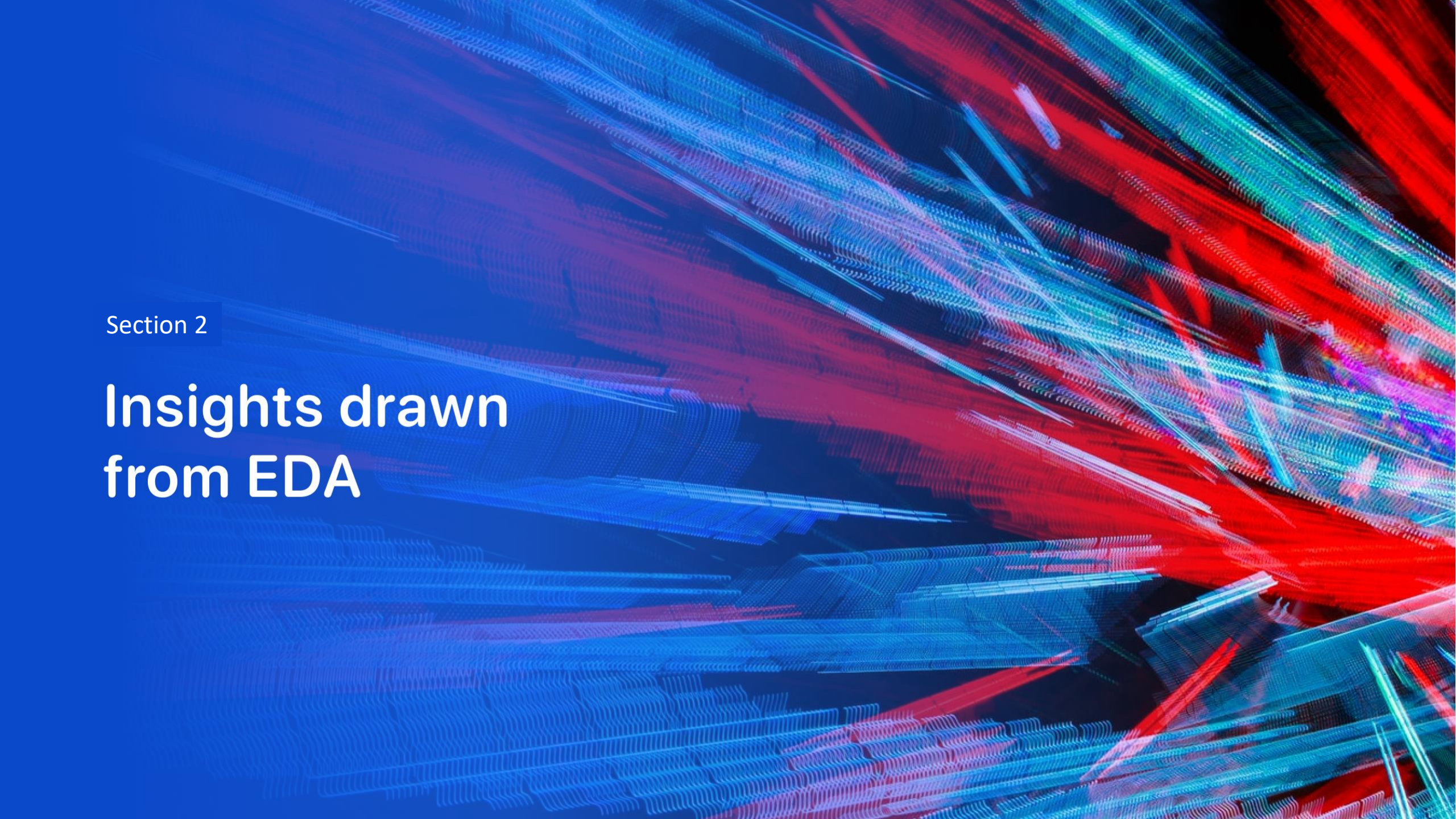
- https://github.com/froniR/ToolsForDataScienceCertificate/blob/5b3d8bf700247c3e851949d08f9e6dd35d0a2cbd/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite%20(2).ipynb

Results

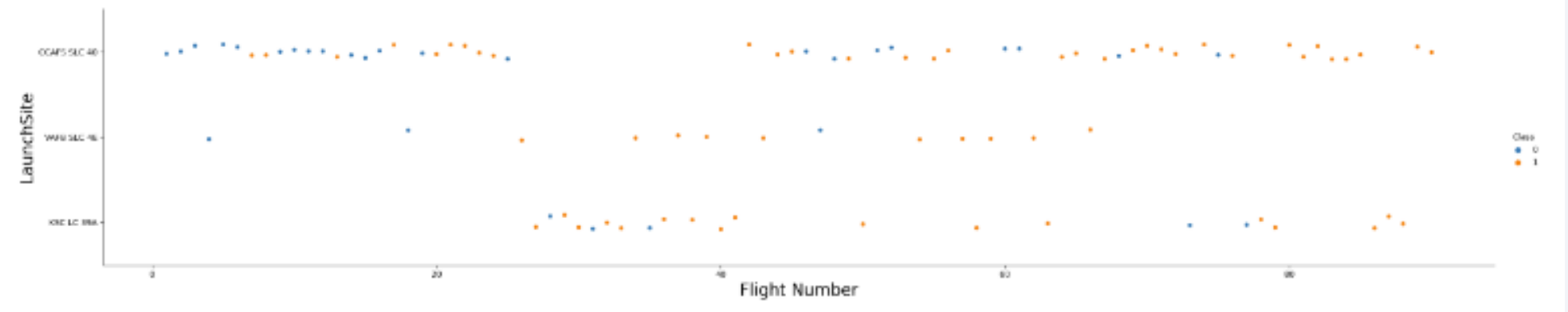
- Exploratory data analysis results
  - SpaceX uses 4 different launch sites
  - The first launches were doneby SpaceX and NASA
  - The average payload of the Falcon9 was 2.928 kg
  - The first success landing outcome happened in 2015; years after the first lauch
  - Many Falcon9 booster versions were successful at landing in drone ships having payload above average
  - Almost 100% of mission outcomes were successful
  - The number of landing outcomes became better with years passed

Results

- Interactive analytics demo in screenshots
  - Using interactive analytics it was possible to identify that launch sites are in safety places, e.g. near the sea and have good logistic infrastructure
  - Most launches happened at east coast launch sites
- Predictive analysis results
  - Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings with an accuracy over 94%

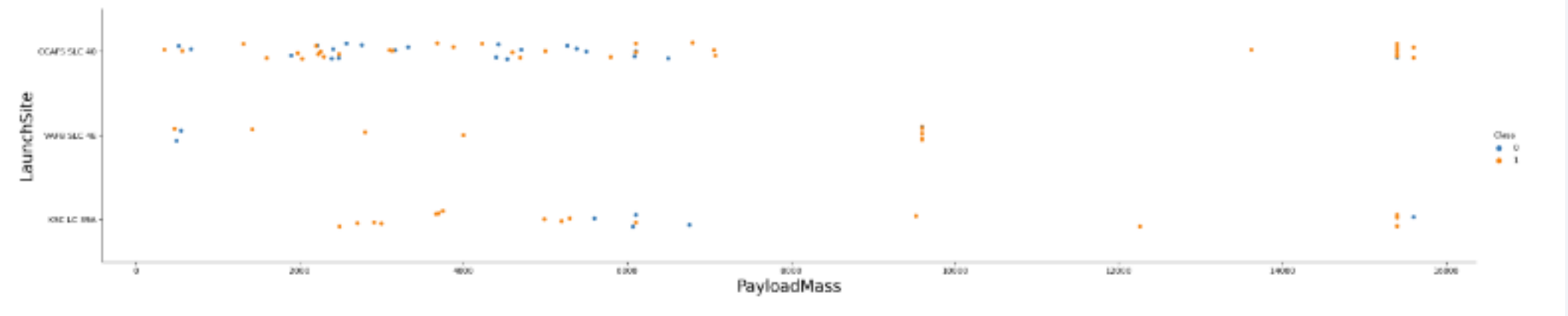

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot, we can verify that the best launch site now is CCAF5 SLC40, where the most recent launches were successful
- The second place is VAFB SLC4E and third place is KSC LC39A
- We can see that the general success rate improved over time

Payload vs. Launch Site

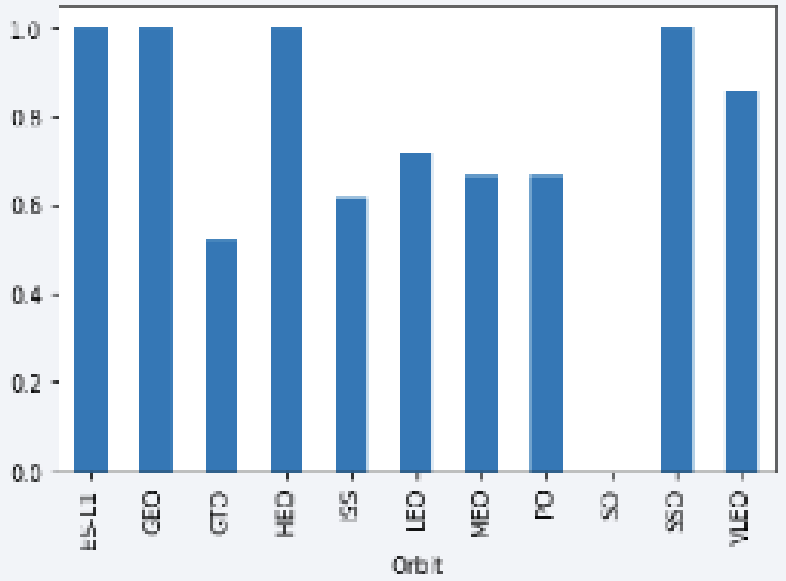


- Payloads over 9000 kg have a excellent success rate

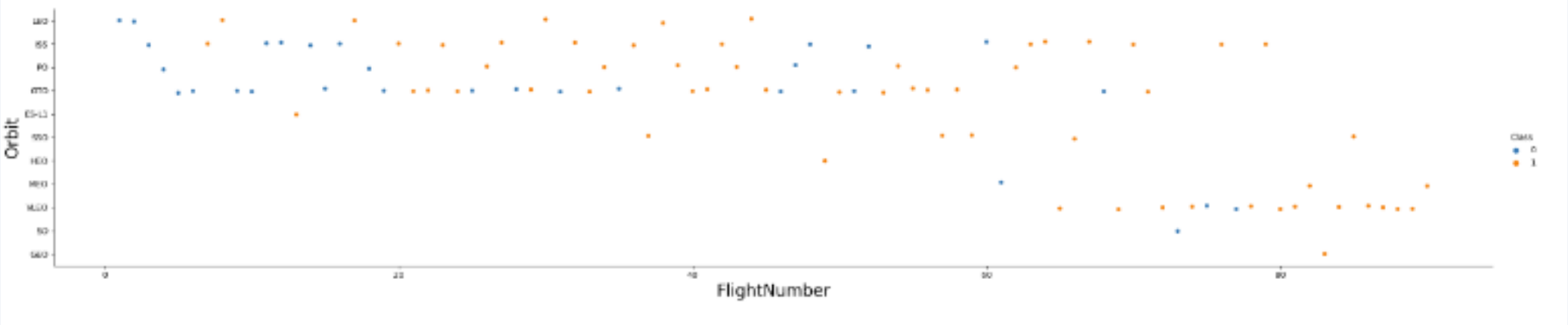
- Payloads over 12000 kg are only possible with CCAFS SLC40 and KSC LC39A

Success Rate vs. Orbit Type

- The biggest success rates with 1 had the orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed with ca. between 0.7 and 0.8 by:
  - VLEO
  - LFO



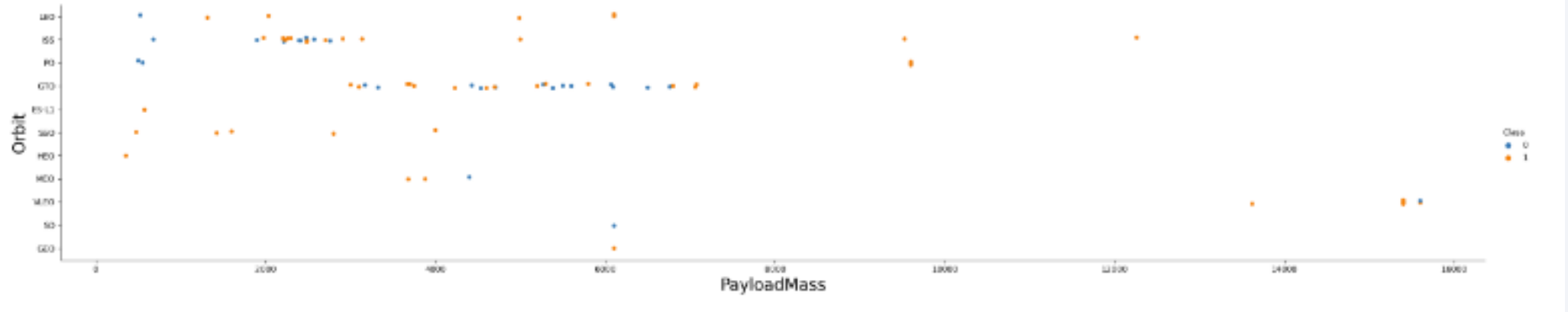
Flight Number vs. Orbit Type



- The success rate improved over time

- VLEO increased it’s frequency recently

Payload vs. Orbit Type

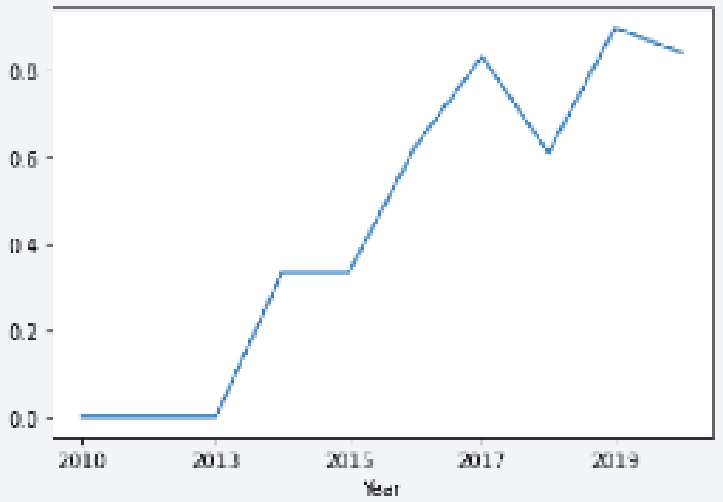


- GTO does not show much of a relation between success rate and payload

- ISS has the widest range of payload and a good success rate

Launch Success Yearly Trend

- Success rate started to increase in 2013
- It decreased 2020 and increased again until 2019
- Before 2013 the line stays at 0



All Launch Site Names

- There are four launch sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- 5 CCAFS LC-40 launches

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landi _Outcor
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failu (parachu
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	Failu (parachu
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	I atten
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	I atten
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	I atten

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- A total payload of 111.268 for codes that contain CRS

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Average payload of 2928 kg

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- We found out that the date of the first successful landing outcome was on 22. December 2015

	firstsuccessfull_landing_date
0	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- These are the boosters which have the described conditions

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- We used '%' to filter wheter there was a success or a failure

The total number of successful mission outcome is:

	successoutcome
0	100

The total number of failed mission outcome is:

	failureoutcome
0	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 explanation here

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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- We filtered for failed landing outcomes in drone ship, their boosters and launch sites from 2015

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

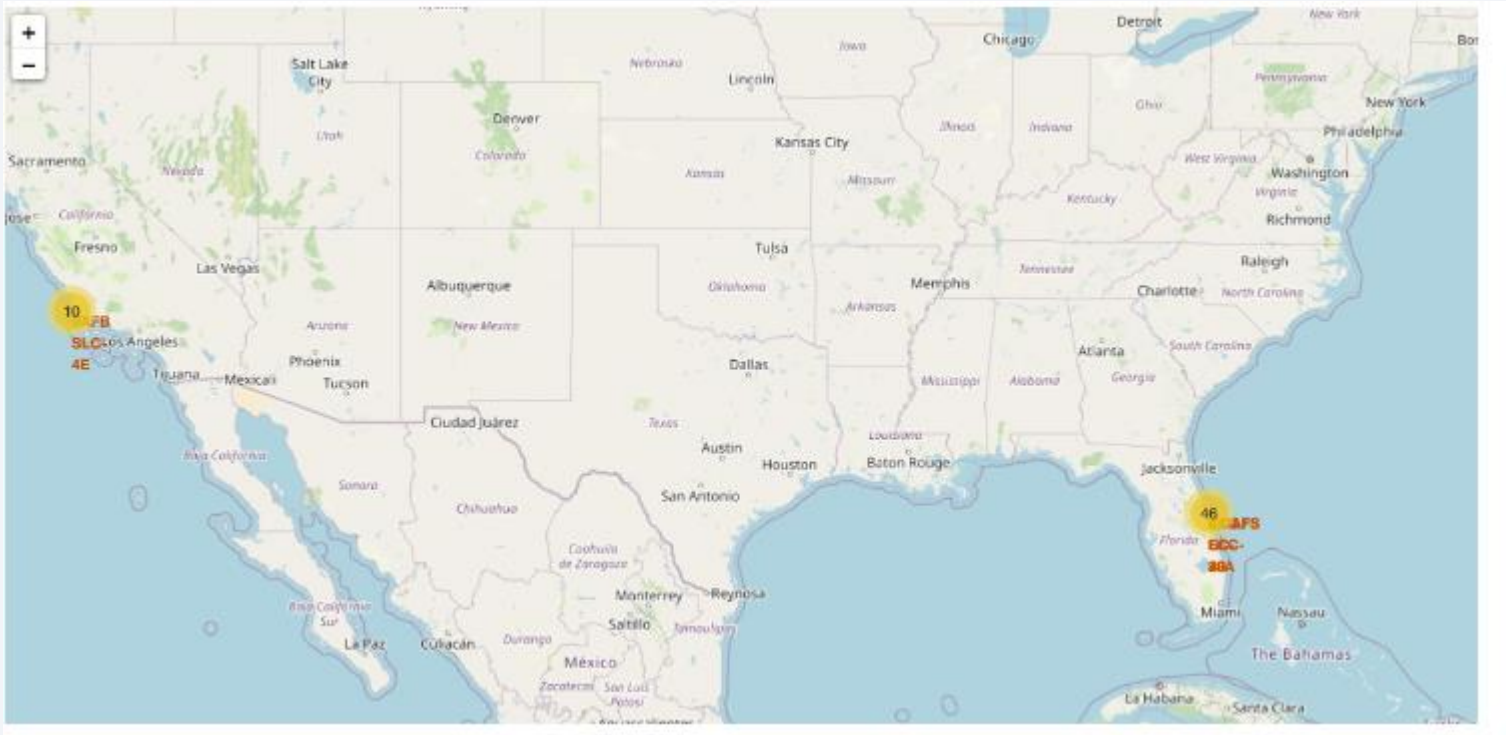
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- We counted the landing outcomes and filtered the timespan
- We then applied GROUP BY landin outcomes and ordered in descending order

Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Section 3

Launch Sites Proximities Analysis

Global map markers



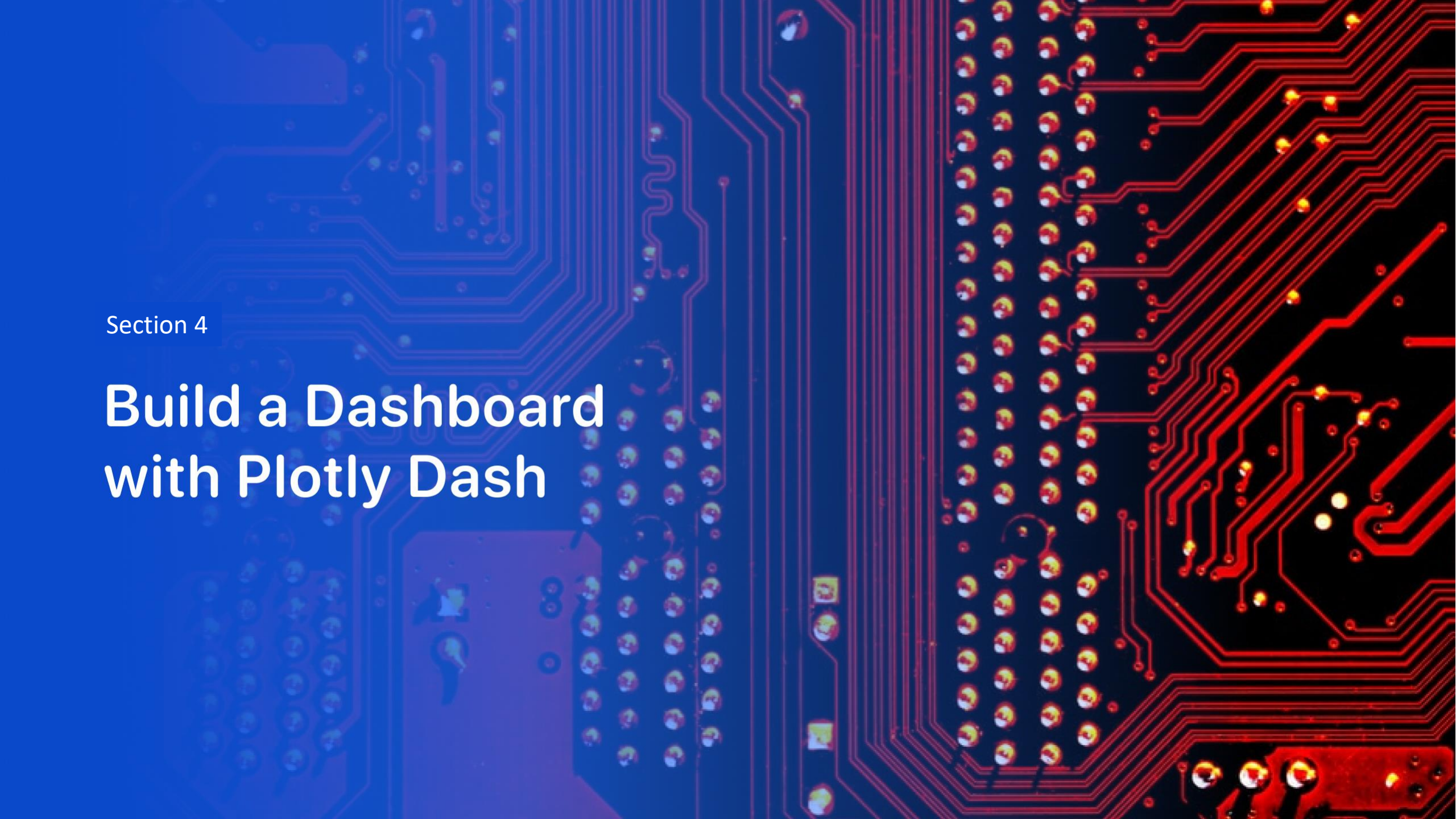
- We see the spaceX launch sites in the United States coasts in Florida and California

Markers showing launch sites success

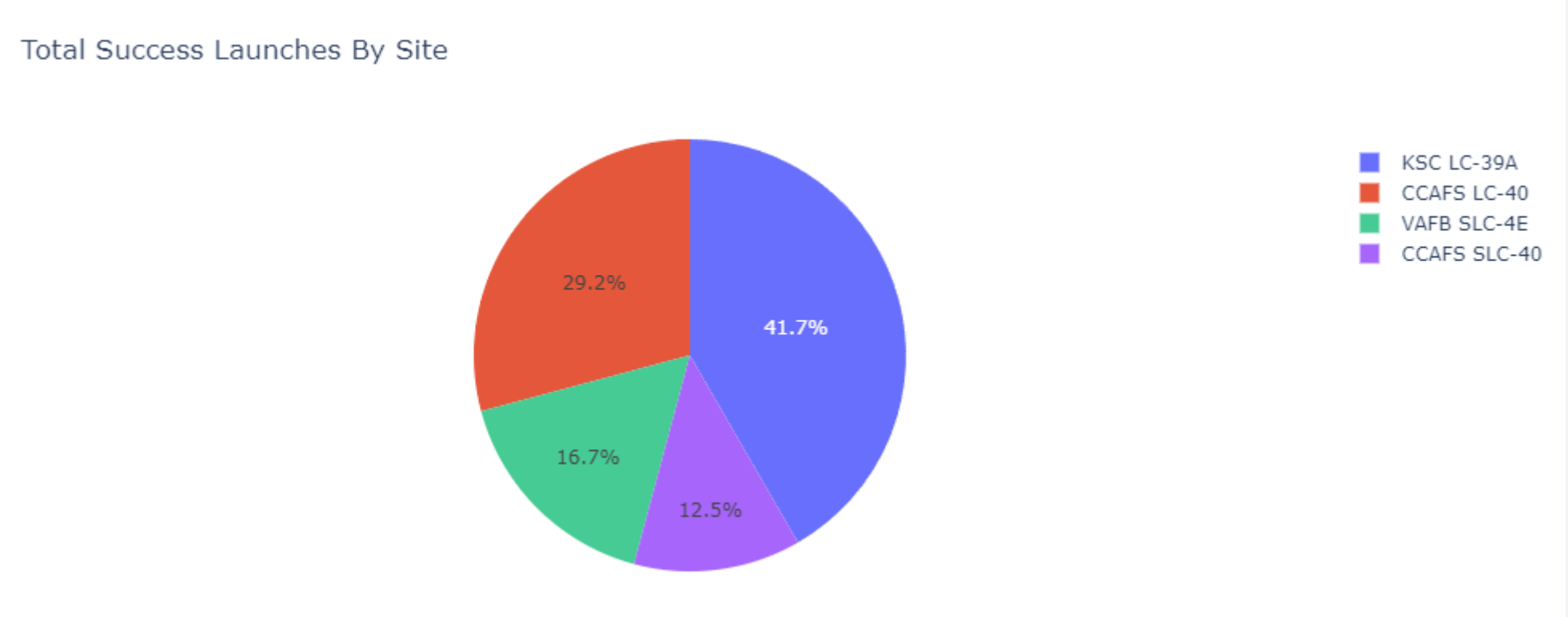
- Florida Launch Site, we see successful launches in green and failures in red

Launch Site distance to landmarks

- Launch sites are not in close proximity to the coastline and cities

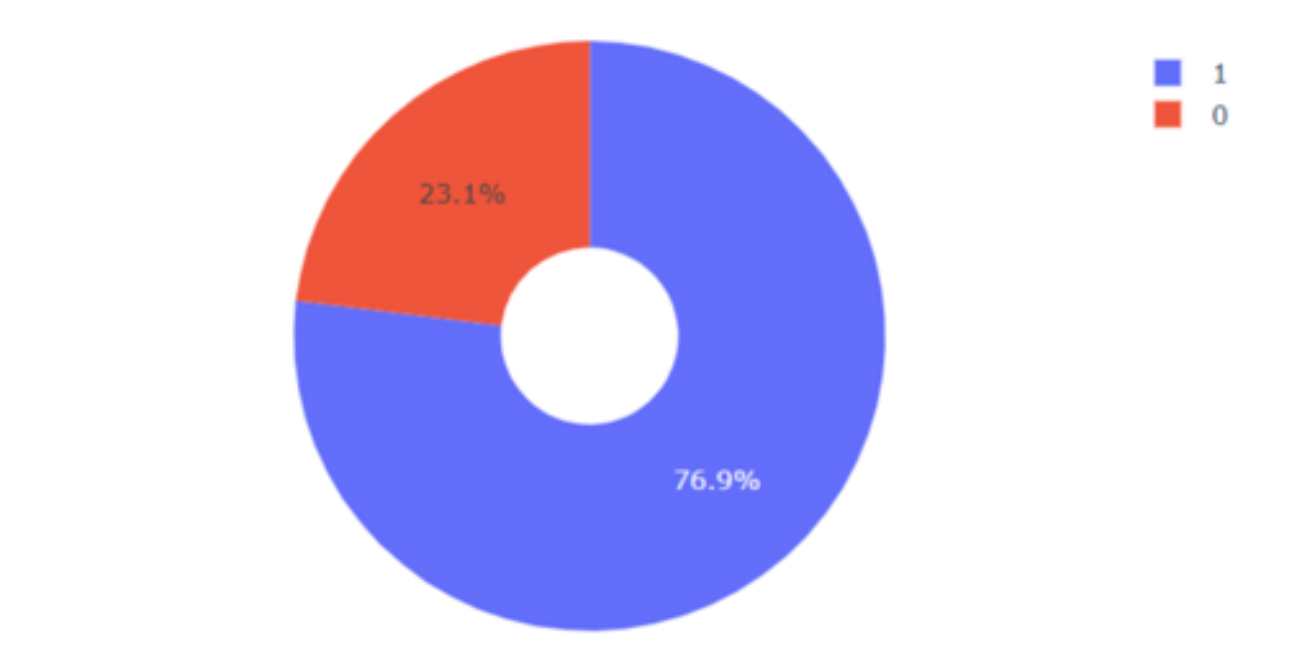

Section 4
Build a Dashboard with Plotly Dash

Pie Chart total success lauches



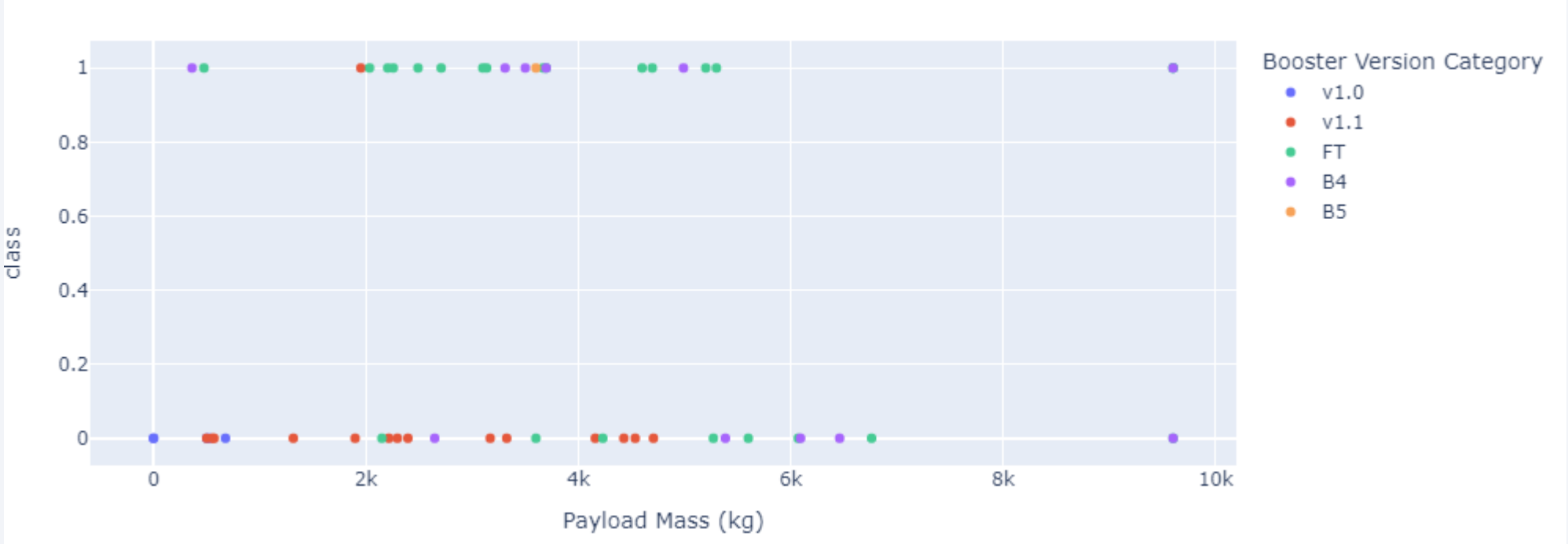
- We see that KSC LC-39A had the most successful launches

Pie chart launch with highest launch success ratio



- KSC LC-39A achied 76.9% success and 23.1% failure

Scatter plot of Payload vs Launch Outcome



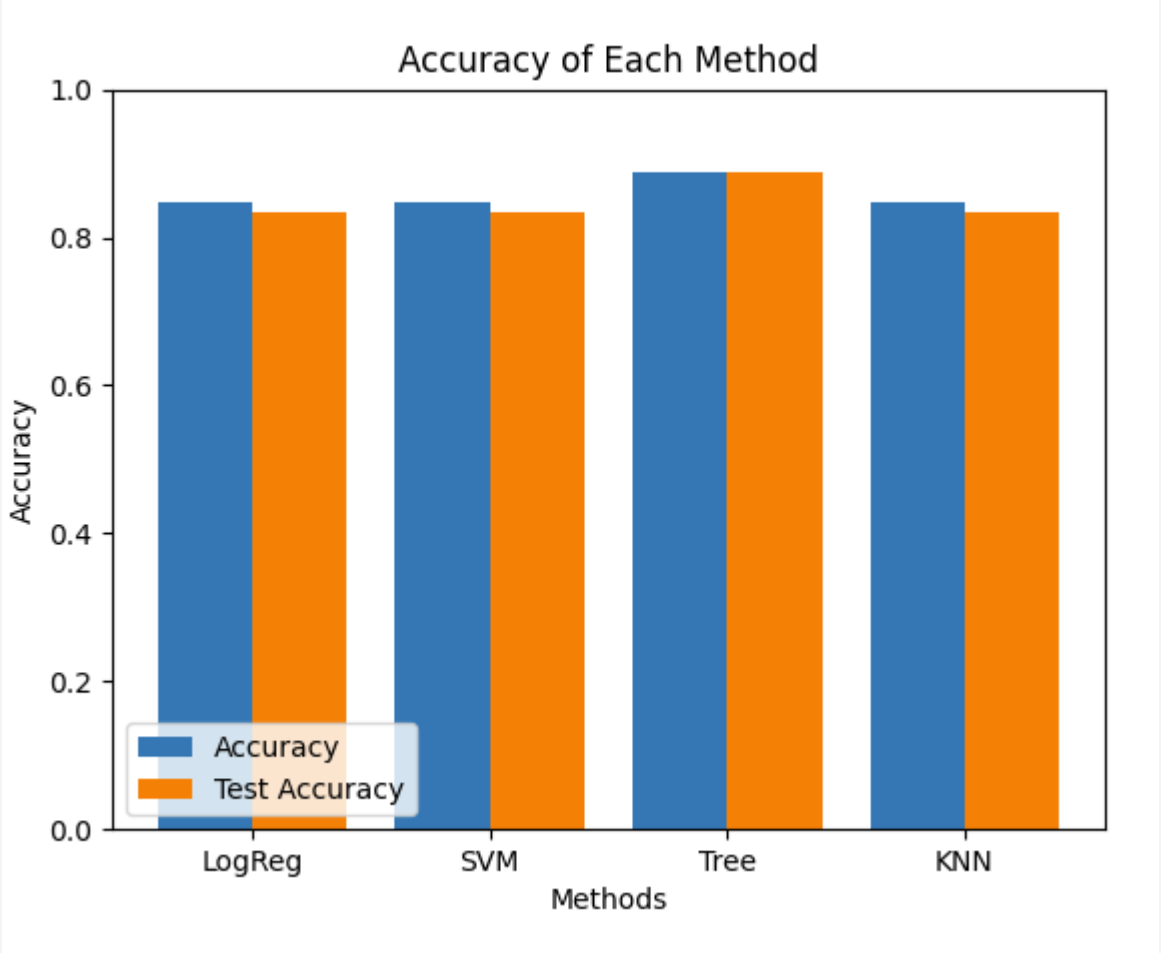
- We see success rates for low weighted payloads

Section 5

Predictive Analysis (Classification)

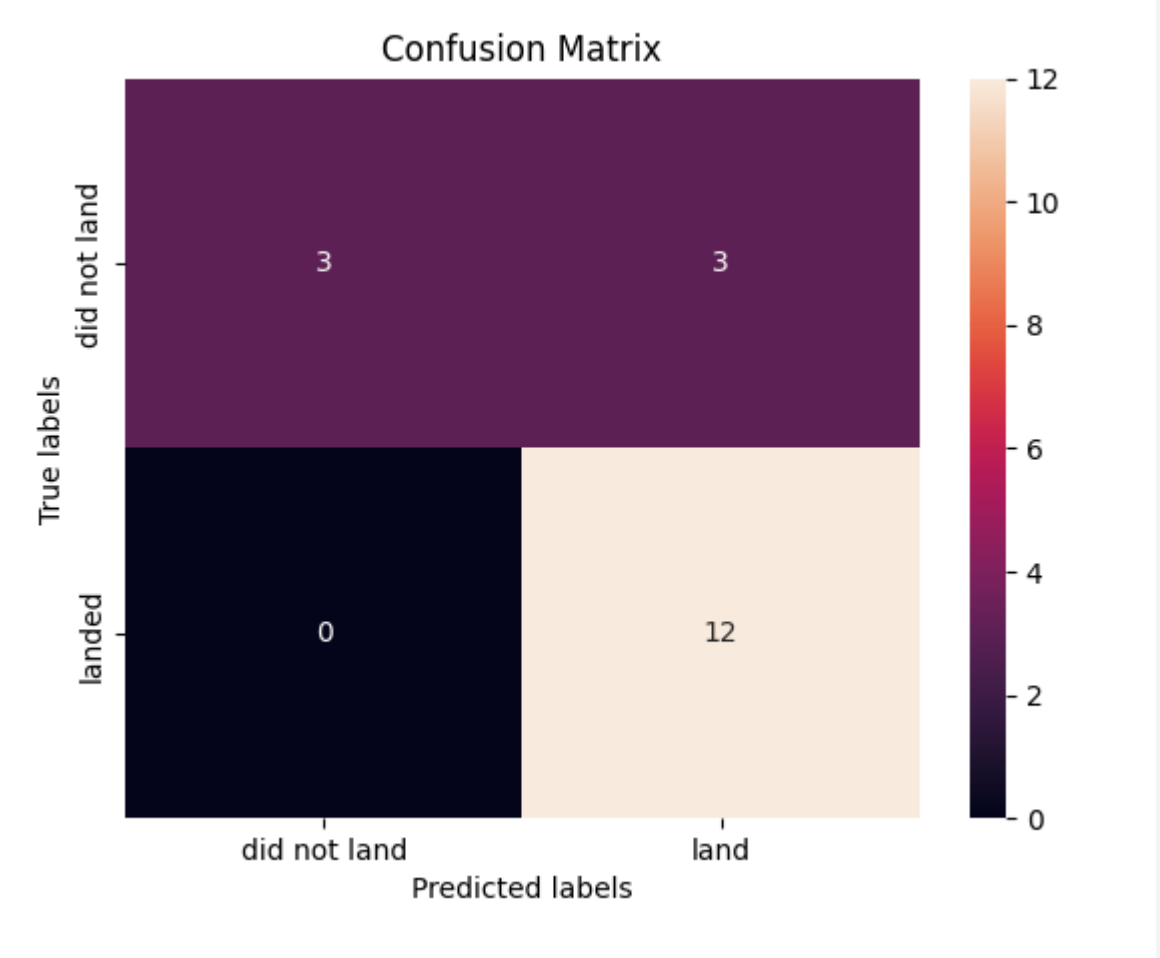
Classification Accuracy

- Decision trees have the highest classification accuracy



Confusion Matrix

- The confusion matrix of the best performing model is for decision trees, it shows the four different classes with the biggest problem of false positives



Conclusions

- The Orbits ES-L1, GEO, HEO, SSO, VLEO had the highest success rate
- Launch success rates started to increase in 2013
- The decision tree classifier is the best algorithm for this task
- KSC LC-39A had the most successful launches

Thank you!