IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Khaled Bouabdallah
17 Mars 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using REST API calls and Web scraping
  - Data Wrangling by cleaning and transforming the data and prepare it for a classification task
  - Exploratory Data Analysis using SQL and Visualization packages in Python
  - Interactive dashboard Web App using Plotly
  - Exploring Launch Sites using interactive Folium Maps
  - Predictive landing outcome given a set of variables

- Summary of all results
  - Exploratory Data Analysis Results
  - Predictive Analysis Results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website for $62 million were other companies charge upward of $165 million per launch. SpaceX can save costs because it can reuse the first stage of the rocket.

- If the first stage successfully lands, it helps determine the overall cost of a launch, so our goal is to predict if rocket will successfully make the first stage land

Section 1

Methodology

Methodology

Executive Summary

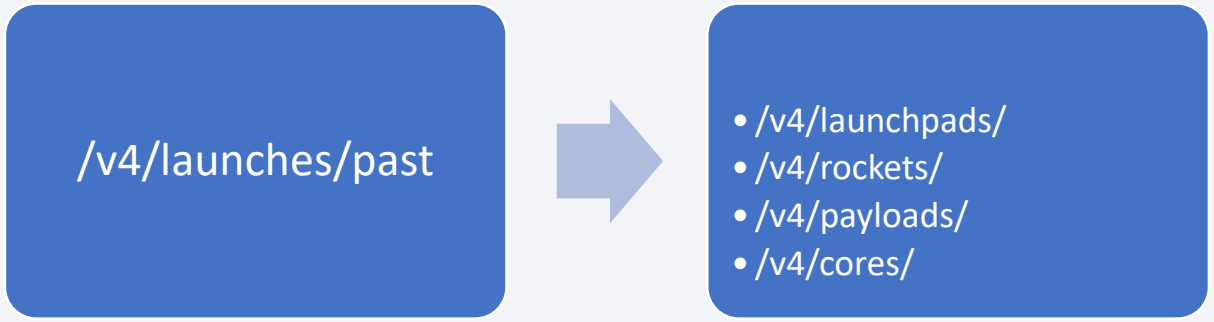
- Data collection methodology:
  - Data was collected from multiple sources: SpaceX API and Web Scraping
- Perform data wrangling
  - Data was cleansed, transformed and prepared.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data was collected from two different sources:
    1. SpaceX REST API.
    2. Web scraping from Wikipedia.

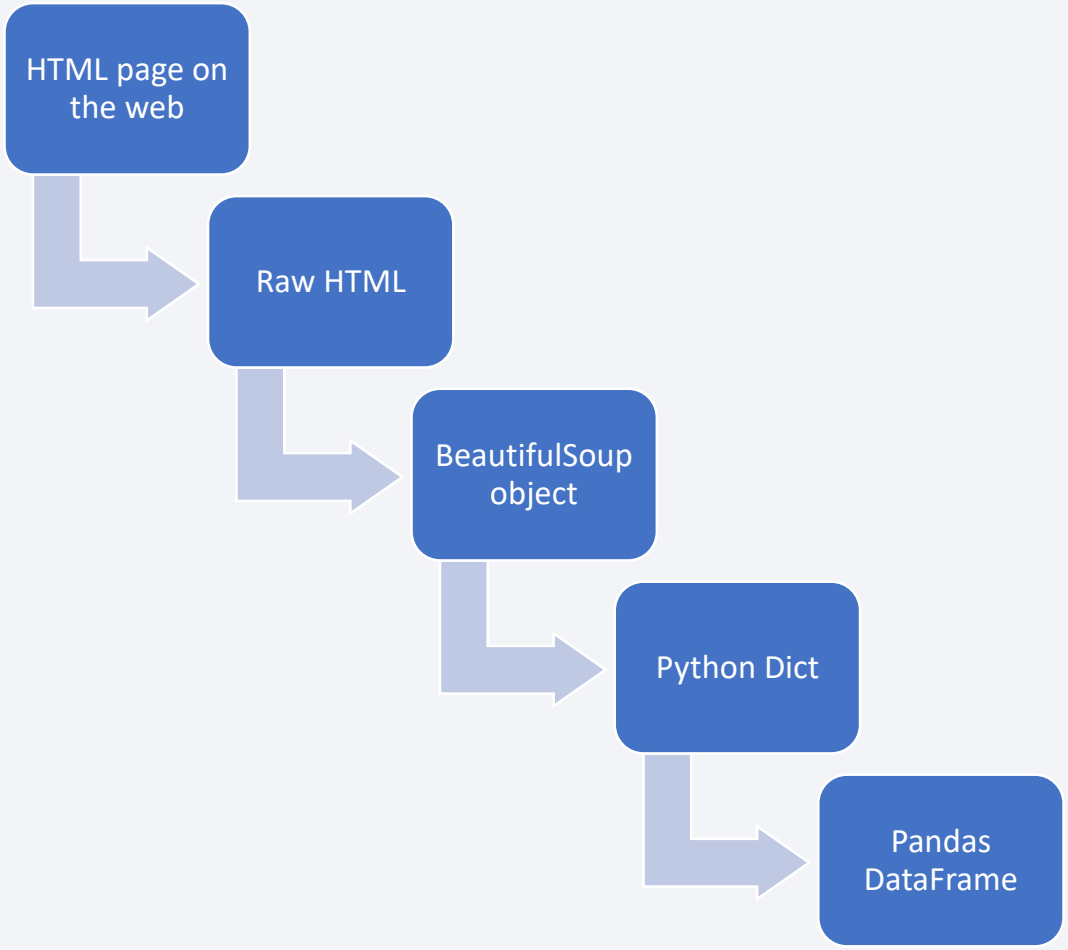
Data Collection – SpaceX API

- SpaceX API is a Rest API that offers multiples endpoints, each endpoint can be used to extract useful information about the SpaceX company.
- The following flowchart show the different endpoints called.
- All the details about the process of data collection for the SpaceX API can be found in this notebook.



Data Collection - Scraping

- Web scraping was done on a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches.
- The page contain details about Falcon 9 launches.
- All the details about the process of web scraping can be found in this notebook.



Data Wrangling

- Data was processes first by dealing with missing values.
- The outcome of the landing was converted into binary variable which we will use as our dependent variable.
- All the details of the data wrangling step can be found in this notebook.

EDA with Data Visualization

- Different Charts and plots were used to visualize and understand relationships between variables

- Scatter plots and line charts are the charts that were primarily used.

- All the details of the data visualization step can be found in this notebook.

EDA with SQL

- A number of SQL queries where preformed to understand the data and answer some questions like:
  - All Launch Site Names
  - Total Payload Mass sent to space
  - Total Number of Successful and Failure Mission Outcomes

- All the details of the data visualization step can be found in this notebook.

Build an Interactive Map with Folium

- Each launch site got marked on the map to know the proximity between them.
- All launches attempts got marked on the map, where successful ones got marked in green and failed attempts in red.
- Distances line were drawn from launch sites to nearby cities, coastlines, railroads to see if a launch site position must follow some specifications.
- All the details can be shown in this notebook.

Build a Dashboard with Plotly Dash

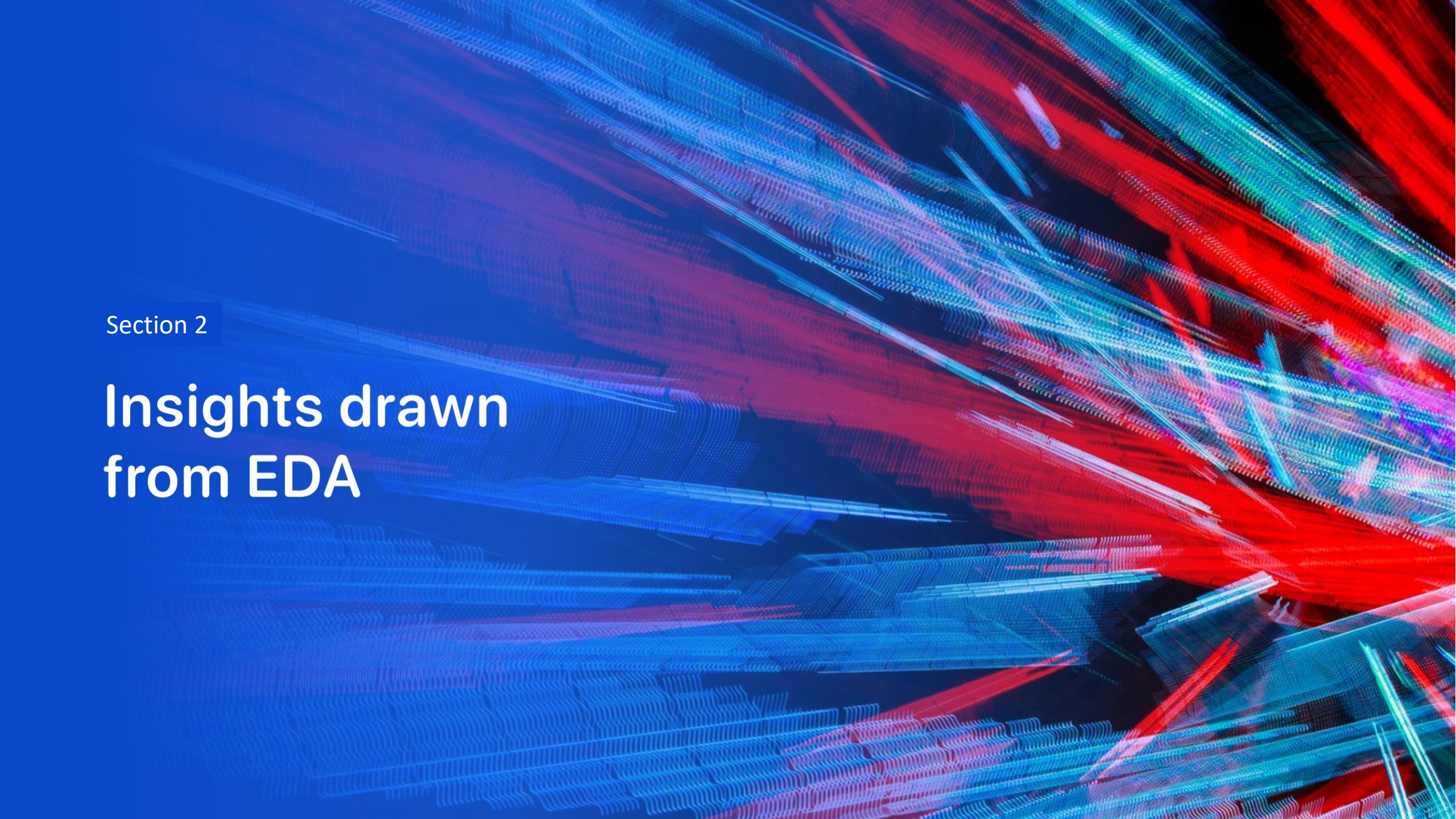
- Pie charts where used to show the percentage of success and failure for each site and for all sites combined, this visualization can help us see which launch sites where the most successful.
- A scatter plot was used to show case the correlation between the payload weight and the result of the landing and the user can set the max and min of the payload weight.
- The rest of the details can be seen here.

Predictive Analysis (Classification)

- Four algorithms were used for classification:
  - Logistic regression
  - SVM
  - Decision Tree
  - KNN

- All four models had 83.33 % test accuracy

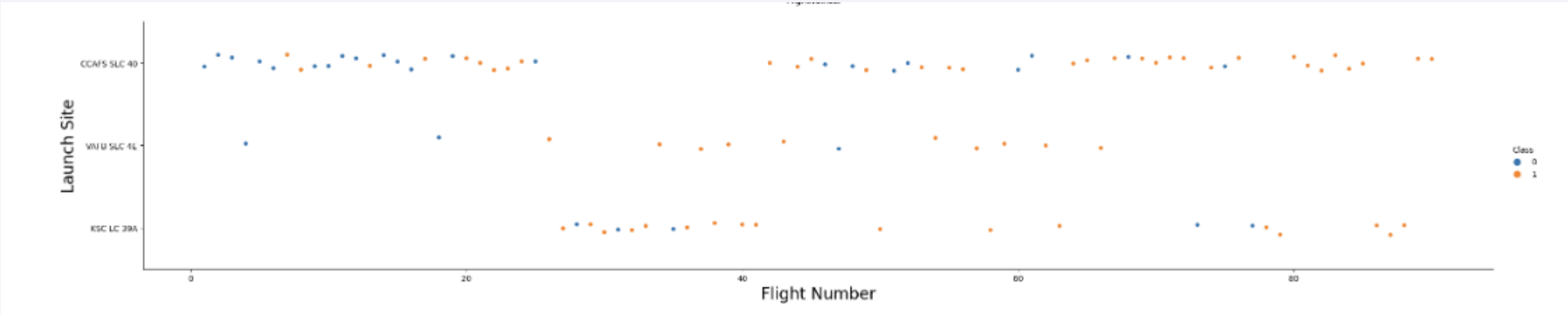
- All the details of this step can be seen in this notebook


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- SpaceX used primarily CCAFS SLC-40 launch site for the first 25 launch then KLC LC-39A was used primarily for the next 15 launch

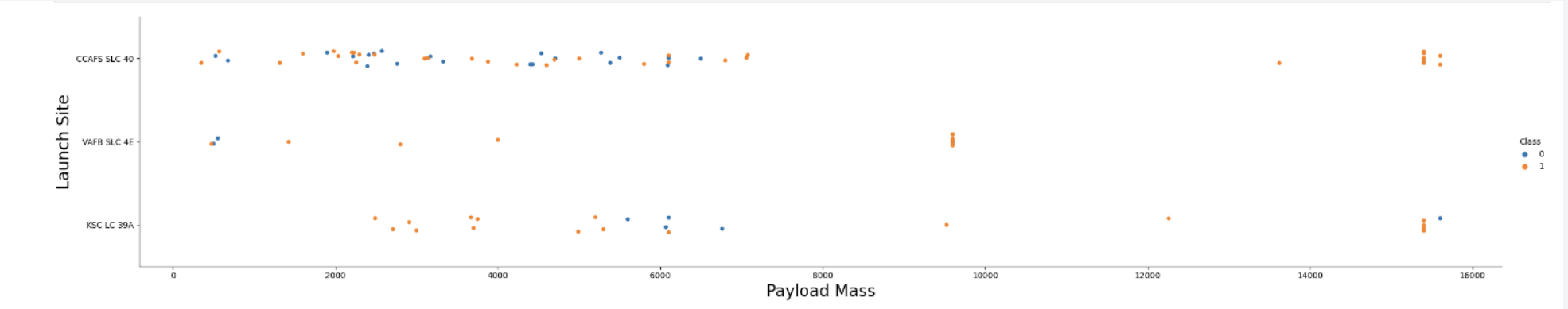
- CCAFS SLC-40 was used again as the primary launch site while VAFB SLC-4E and KLC LC-39A were used occasionally



Payload vs. Launch Site

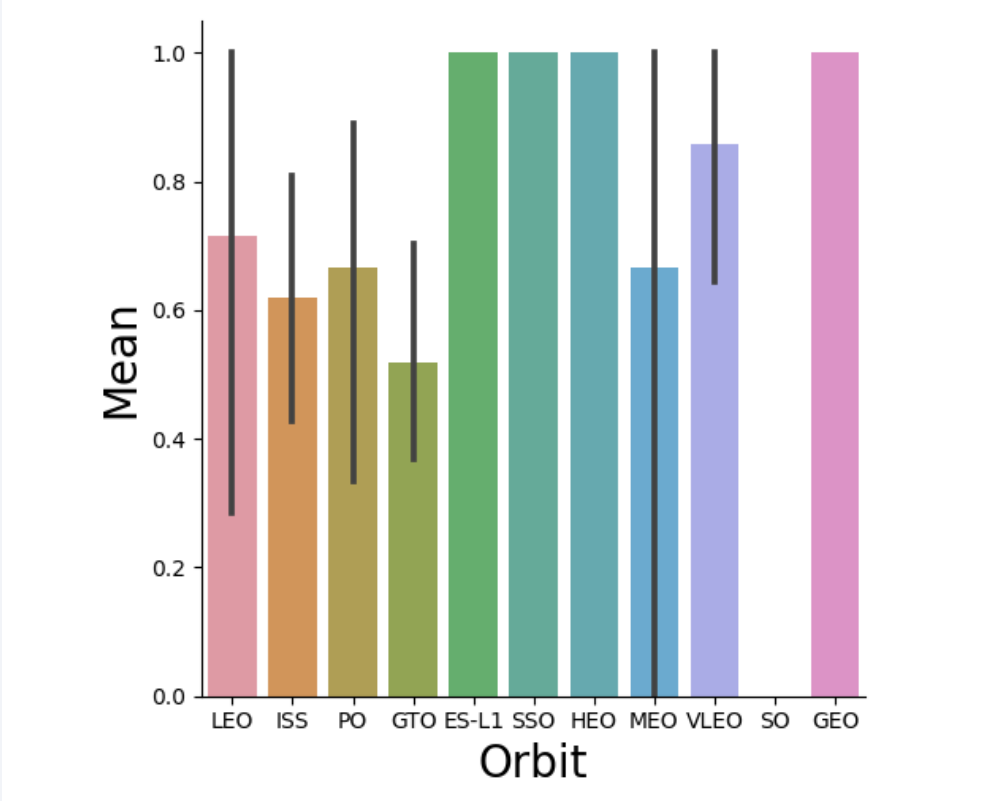
- VAFB SLC-4E launch site was not used with payload greater than 10,000 kg

- All launch sites were mostly used with payload lighter than 8000 kg



Success Rate vs. Orbit Type

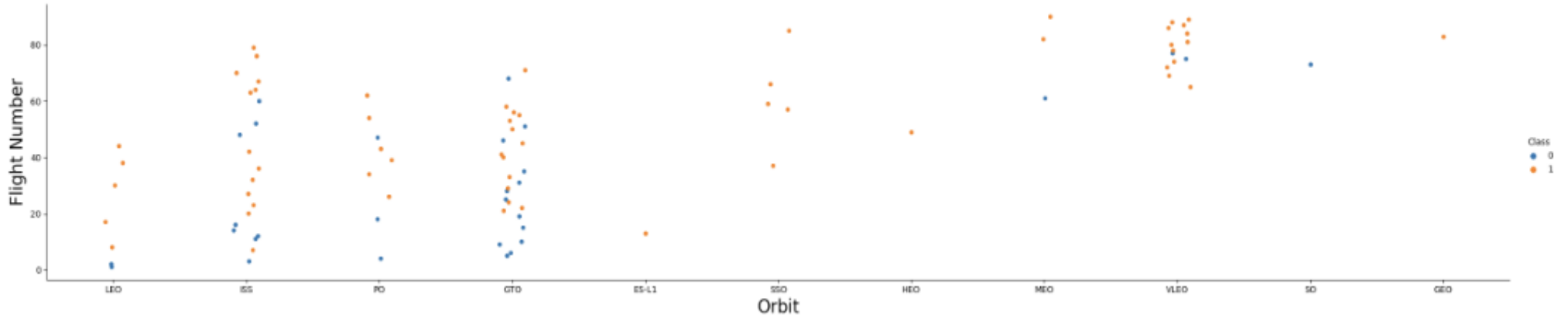
- GEO, HEO, SSO and ES-L1 are all orbits that have 100% success rate.
- GTO has the lowest success rate at almost 50%



Flight Number vs. Orbit Type

- Most of the early flights were done on LEO, SS, PO and GTO orbits

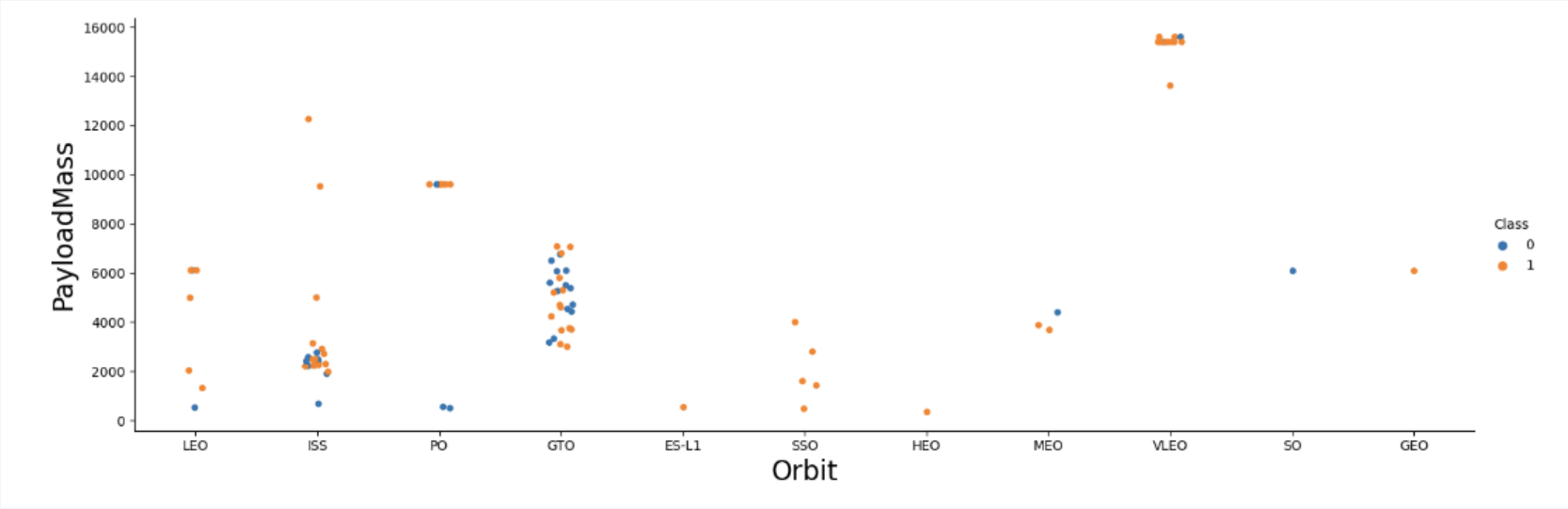
- The rest of the orbits got used just a few times except for VLED which got used quite a few times at the end.



Payload vs. Orbit Type

- Most of the orbits were used with a payload less than 10,000 kg

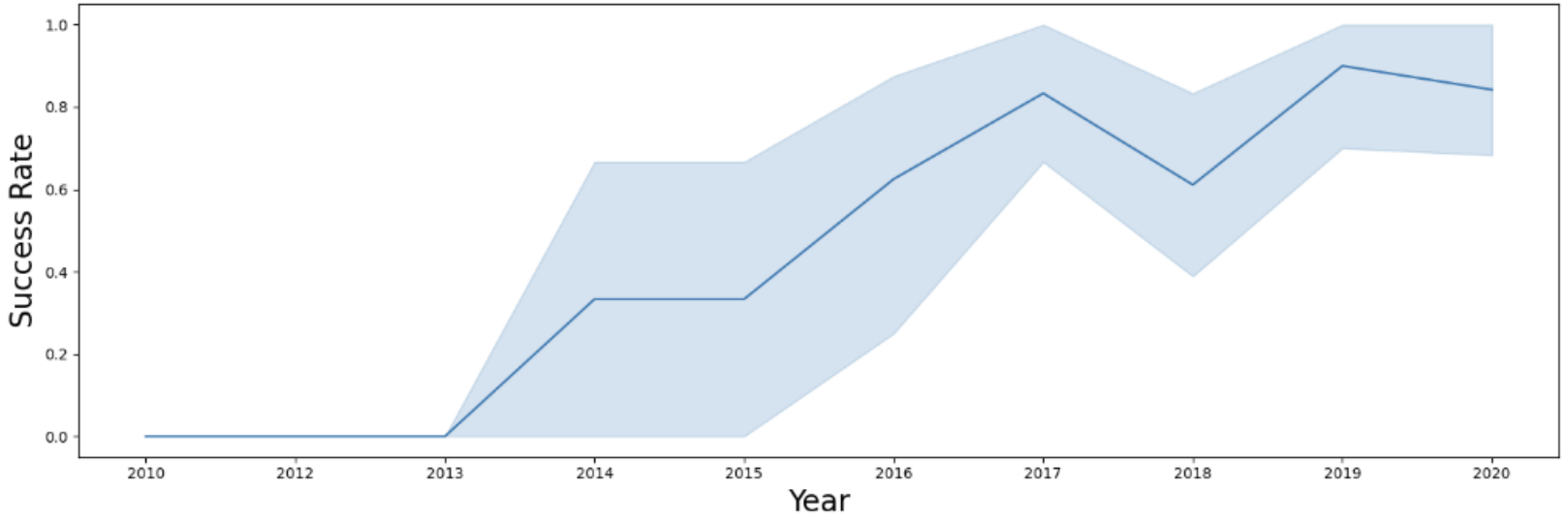
- VLEO got used only with a payload greater than 10,000 kg



Launch Success Yearly Trend

- In general, the success rate increases as the years goes by

- The success rate decreasing in the 2017-2018 period is probably duo to the fact that SpaceX started to test heavier payloads in that period.



All Launch Site Names

- From the query we can see that we have four launch sites, two of them are in the same complex

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin 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

- In total, NASA has sent a payload of 45,596 KG into space

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- In average payload mass carried by booster version F9 v1.1 is 2928.4 KG

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- On 01-05-2017, the first successful landing outcome on ground pad was archived.

MIN("Date")
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- In total, SpaceX had 71 successful and failure mission outcomes

COUNT(*)
71

Boosters Carried Maximum Payload

- List of the names of the booster which have carried the maximum 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

- In 2015, SpaceX had two landing failures, the first in January and the second in April.

substr("Date",4,2)	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

- Rank of 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

2]:

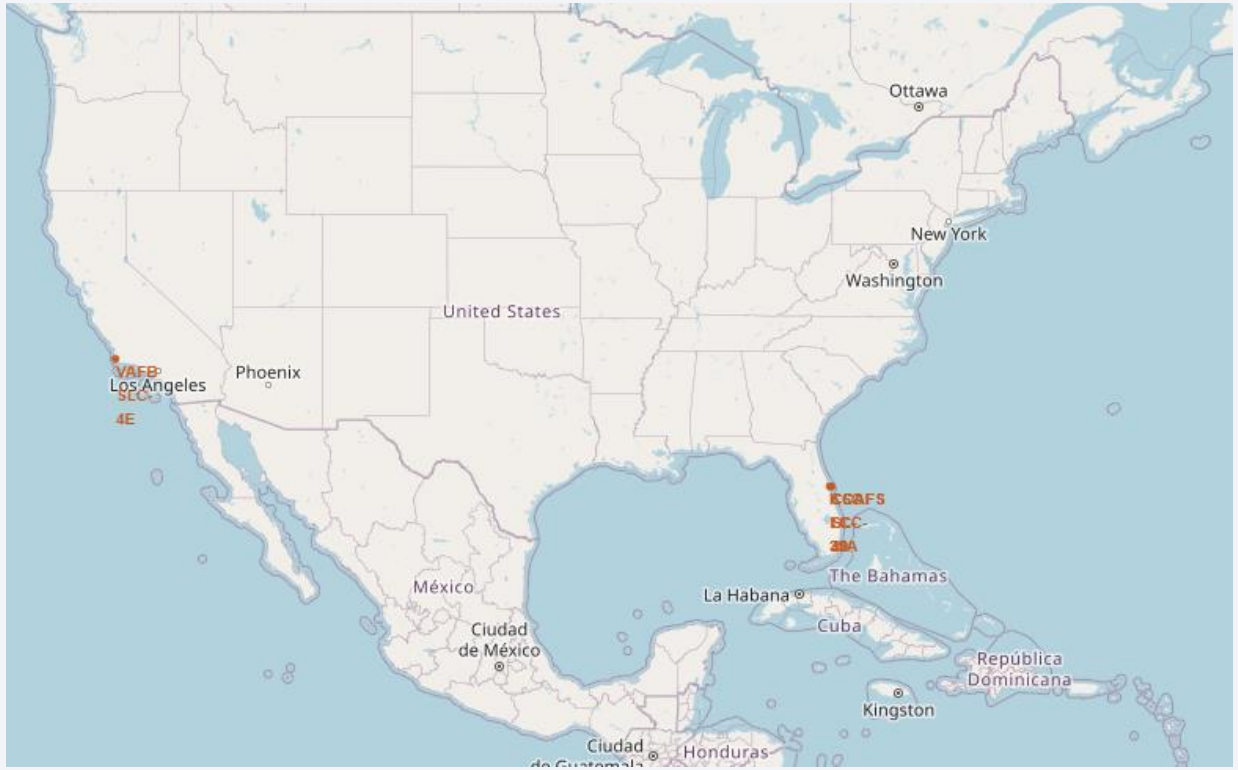
Landing _Outcome	count
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

Launch Sites Proximities Analysis

SpaceX launch Sites Locations

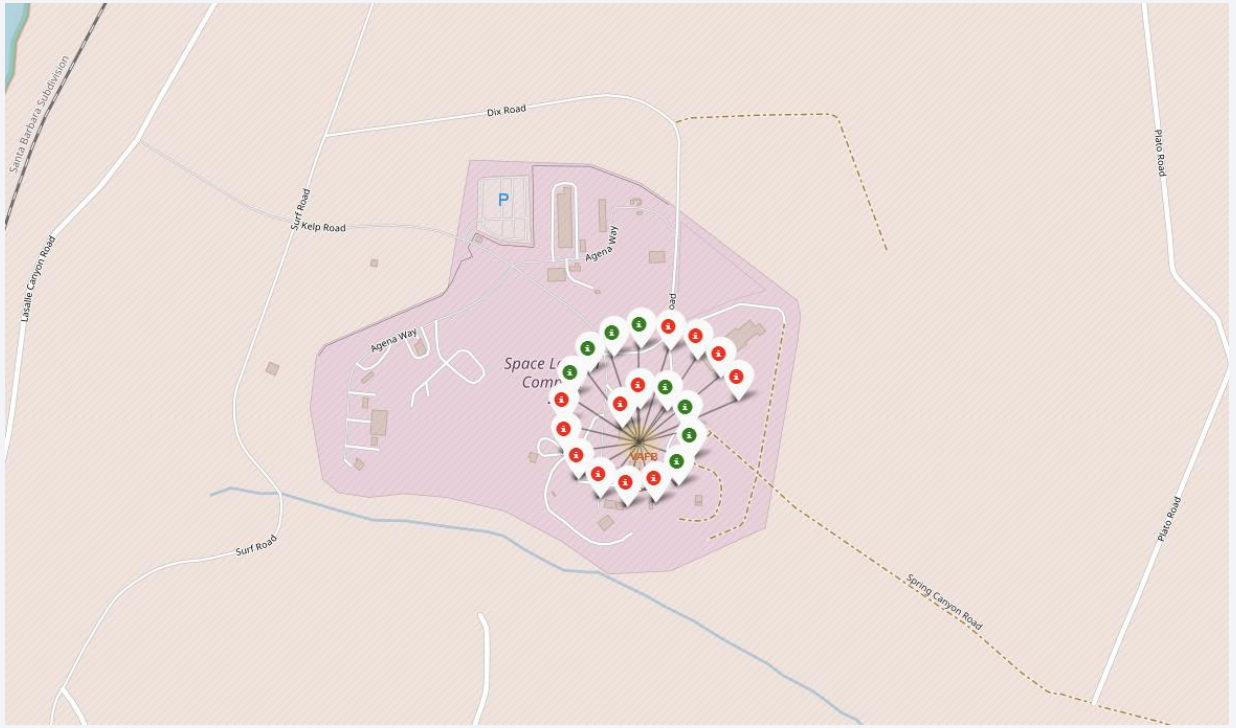
- All launch sites are located near the coastline
- Three of them are on the east coast while the last one is on the west coast (so, the booster stage doesn't fall on a populated area)
- All site launches were located as south as possible near the equator line (The Bahamas sea zone prevented the eastern launch sites to be located more in the south)



Launch Site Attempts

- 20 launches were made in VAFB SLC-4E launch site

- 8 of them were successful and 12 resulted in failure.

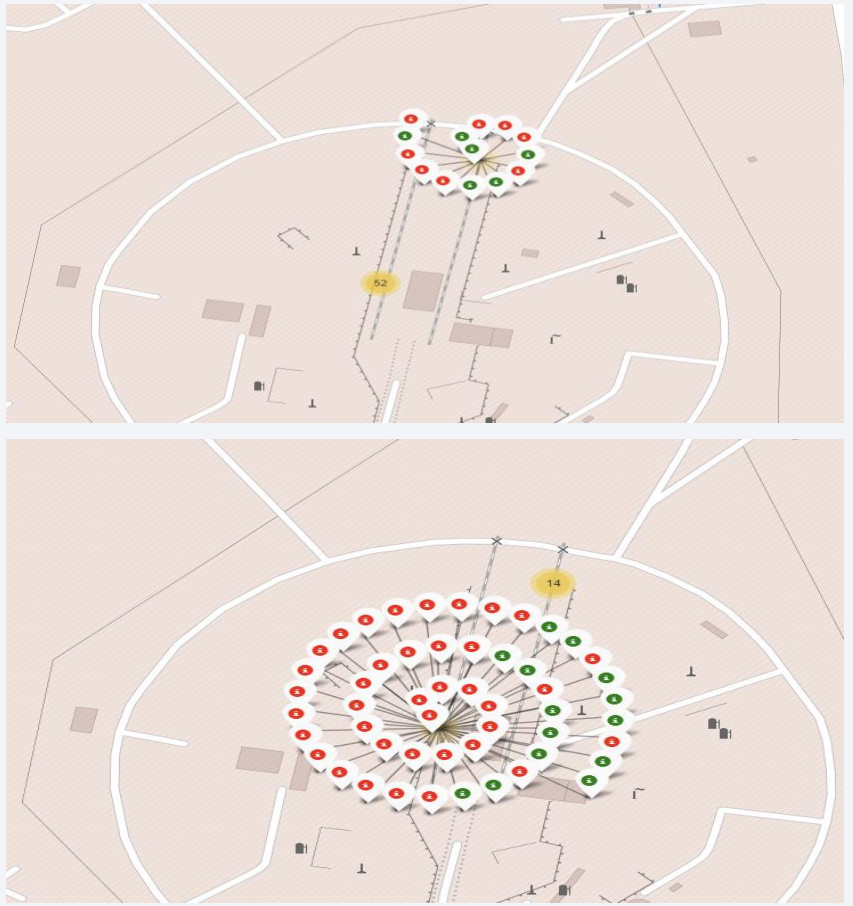


Launch Site Attempts

- 52 launches were made in the CCAFS LC-40 launch site while only 14 were made in CCAFS SLC-40

- CCAFS LC-40 is the main launch site in this location

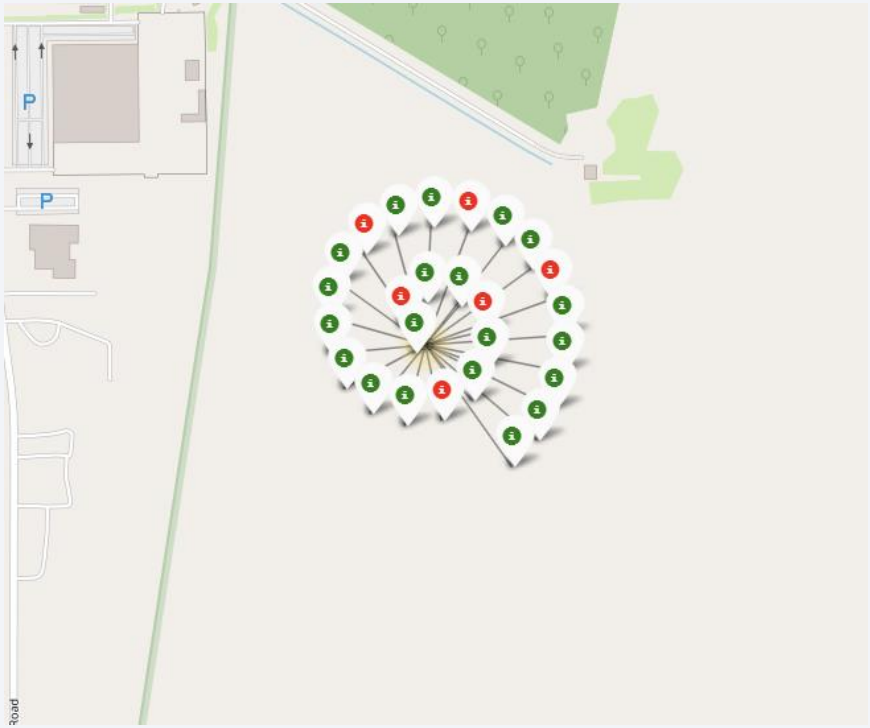
- Most of the launches failed in this two launch sites

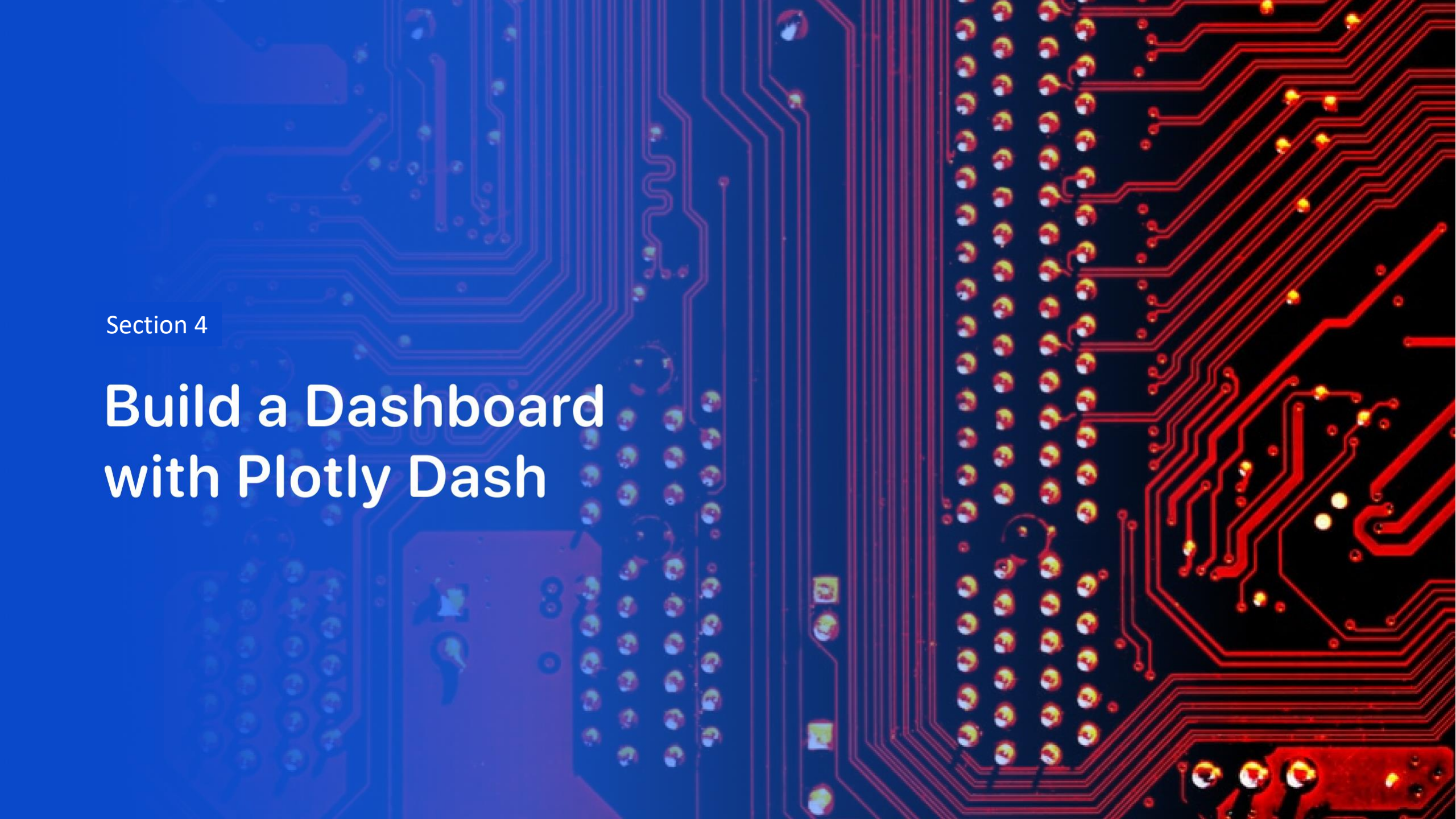


Launch Site Attempts

- 26 launches were made in KSC LC-39A launch site

- 20 of them were successful were only 6 were not.

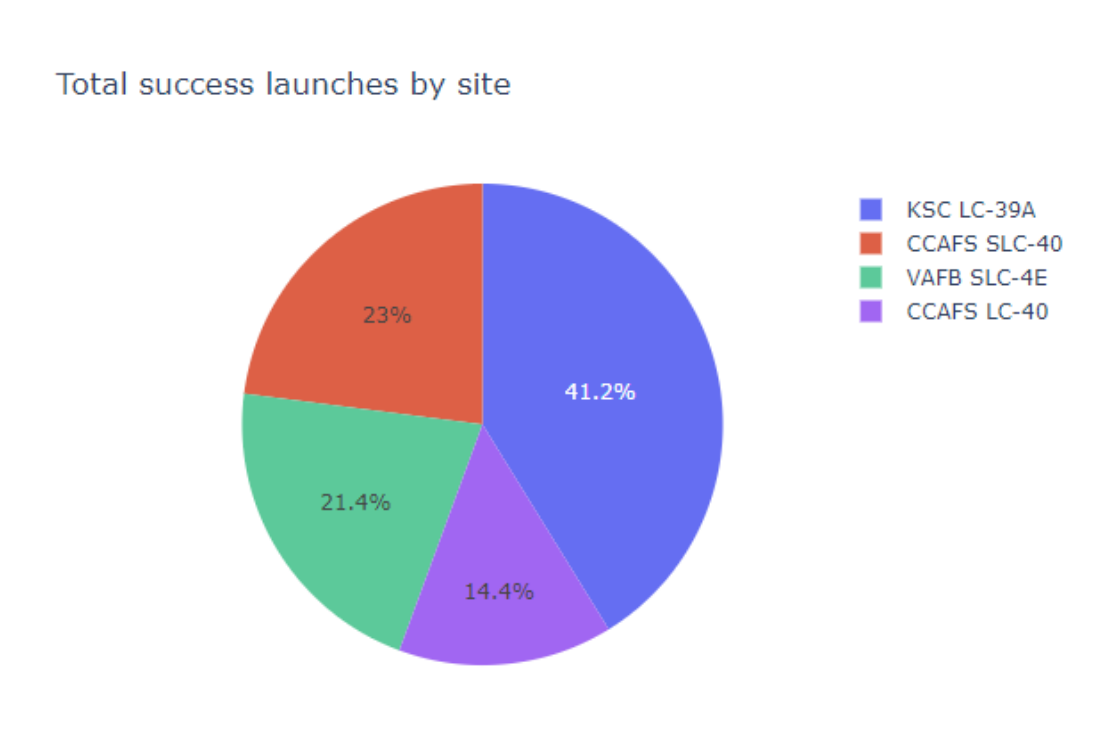



Section 4
Build a Dashboard with Plotly Dash

Landing Sites Overview

- From the pie chart is may seem that the launch site effects the outcome of the launch but, each site was used in a different time with different parameters.

- KSC LC-39A has the most success while CCAFS LC-40 has the least



Most Successful Landing Site

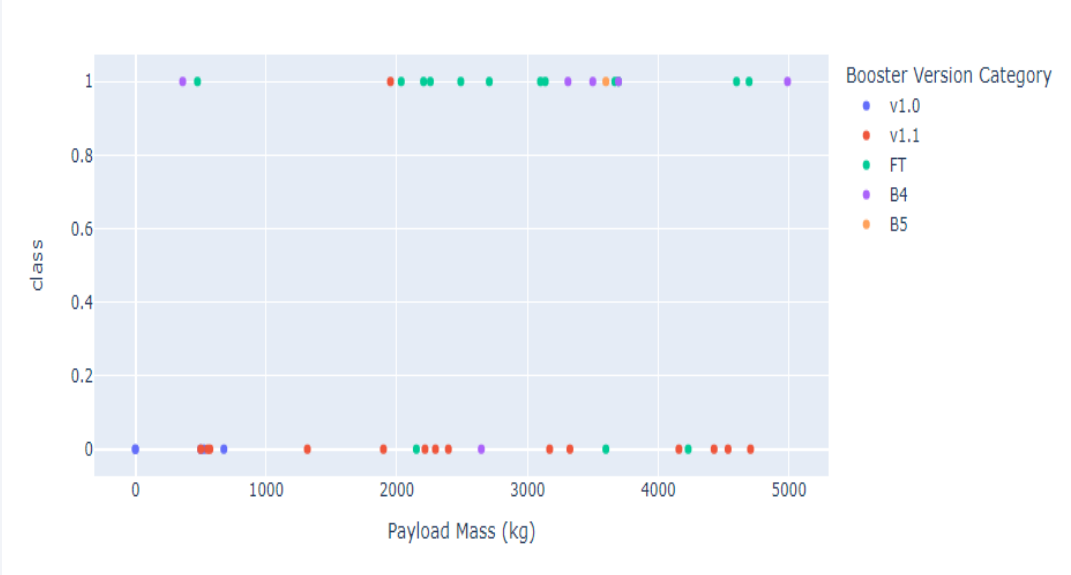
- KSC LC-39A has a 76.9% success rate which is higher than the other landing sites

- This high success rate is duo the fact that this site was used after the other sites.

Payload Vs Launch Outcome

- Most of the launches were done with a payload heavier than 2000k kg.

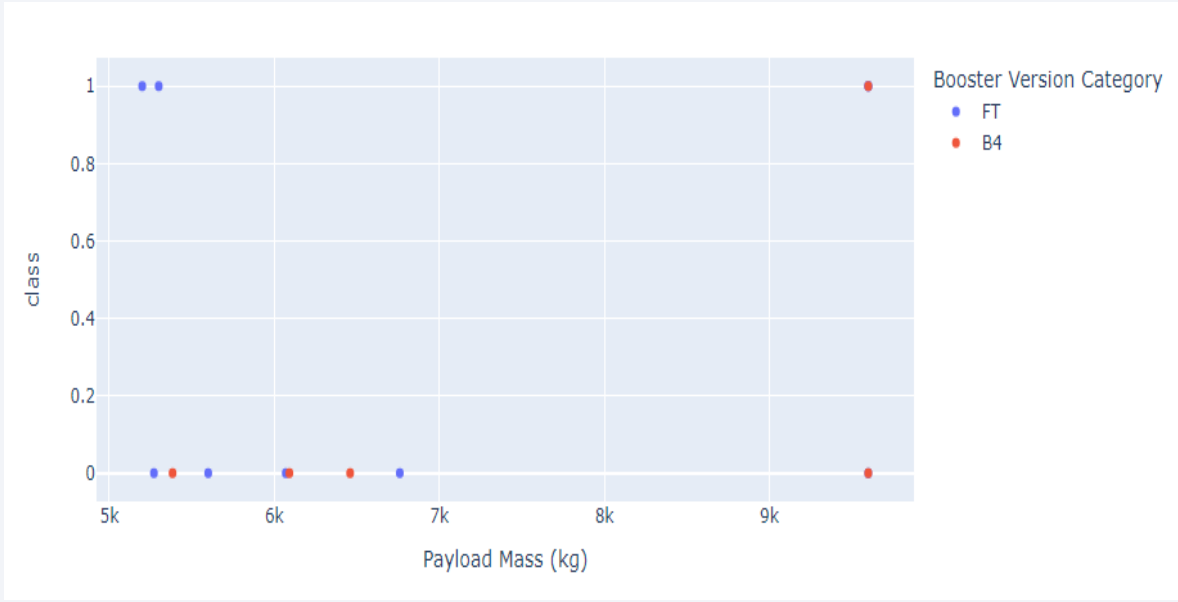
- FT and B4 have the highest successful landings while B5 have the most failures.



Payload Vs Launch Outcome

- Heavy payload (>5000k kg) are more were more challenging with only 3 successful launches.

- FT and B4 are the only two Booster versions used for payloads heavier than 5000k kg



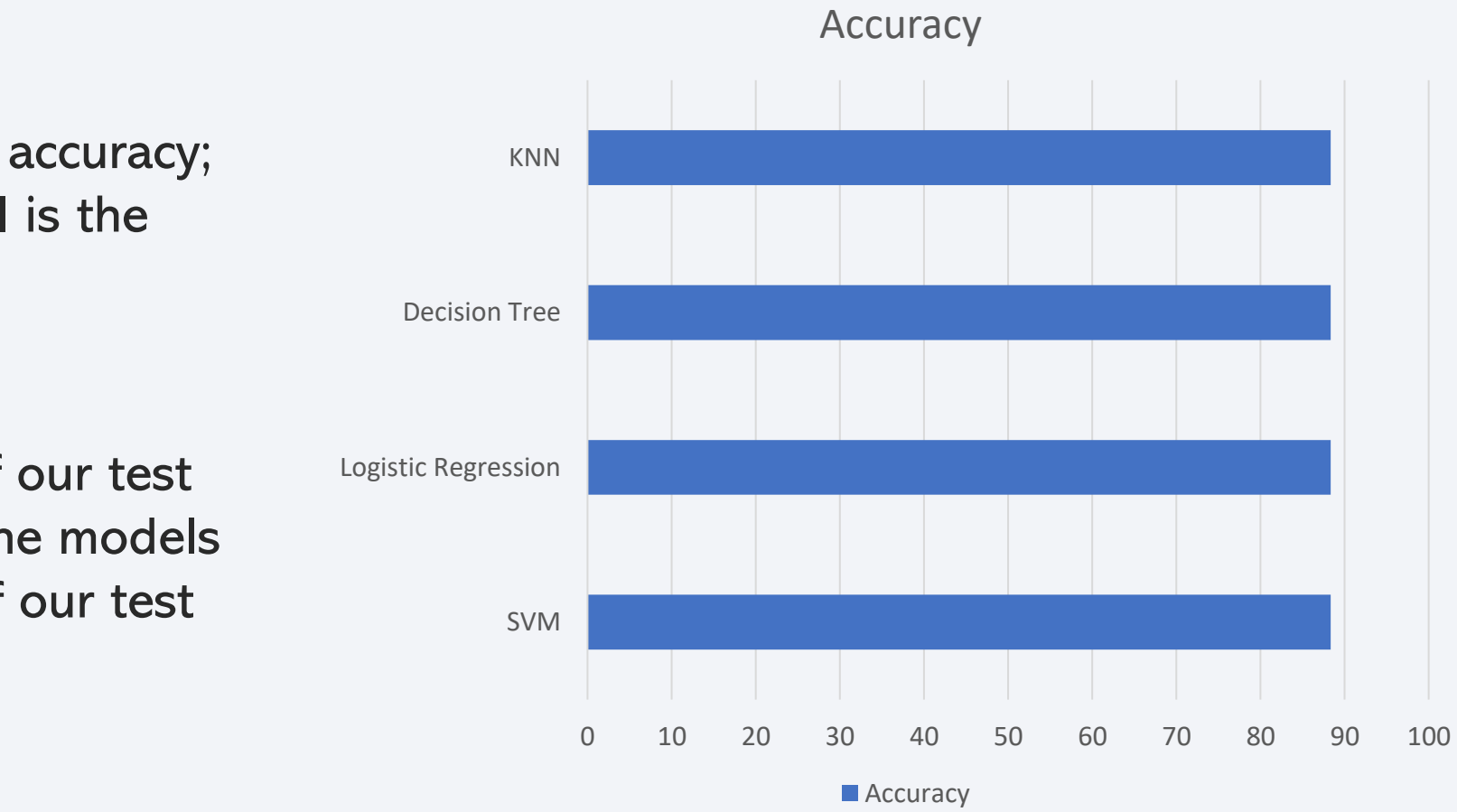
Section 5

Predictive Analysis (Classification)

Classification Accuracy

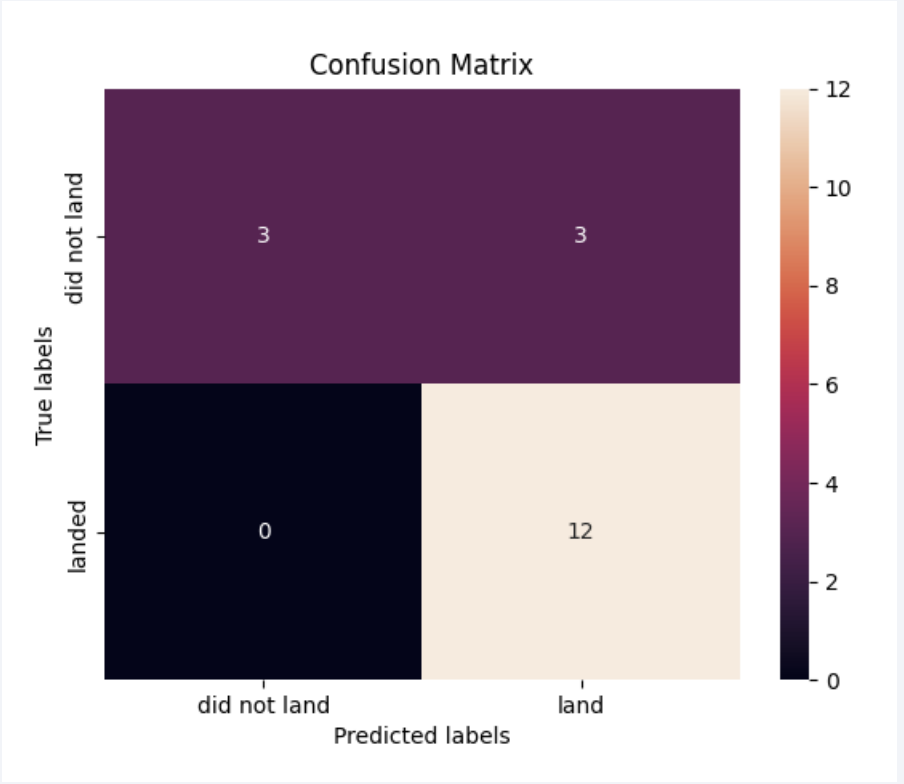
- All models have the same test accuracy; thus, we can't tell which model is the best one.

- This is due to the small size of our test set, to fix this we can retrain the models and increase the proportion of our test set.



Confusion Matrix

- Every model have the same confusion matrix on the test set.
- On the diagonal, we can see the test instances that were classified correctly
- We can also see that we have 3 false positives.



Conclusions

- Orbit type can heavily impact the result of the landing stage; some orbits are more difficult than others
- The more the payload mass increases the more the landing will be more difficult thus increasing the chance of a failure
- The number of flights is very limited which impacted our preposed models' performances
- As the times goes by, SpaceX is getting better thus we should retrain our models on newer data

Thank you!