IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdalwhab Bakheet Mohamed Abdalwhab
08-06-2-24

Outline

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

Executive Summary

- This project aimed to answer the possibility of predicting the success of landing of the rocket’s first stage to reduce the overall cost of launching rocket.
- The analysis was based on SpaceX historical launch data
- After data cleaning and testing different predictive models, we were able to reach a more that 80% accuracy of prediction

Introduction

Reusability of the rocket parts is a major factor in reducing launching costs.

The first stage of the rocket is the most expensive part, and hence the ability to retrieve it and reuse it a crucial factor for cost reduction.

In this project we aim to answer, is it possible to predict the success or failure of landing of the first stage of the rocket based on the flight information, like the payload mass and launching site.

Section 1

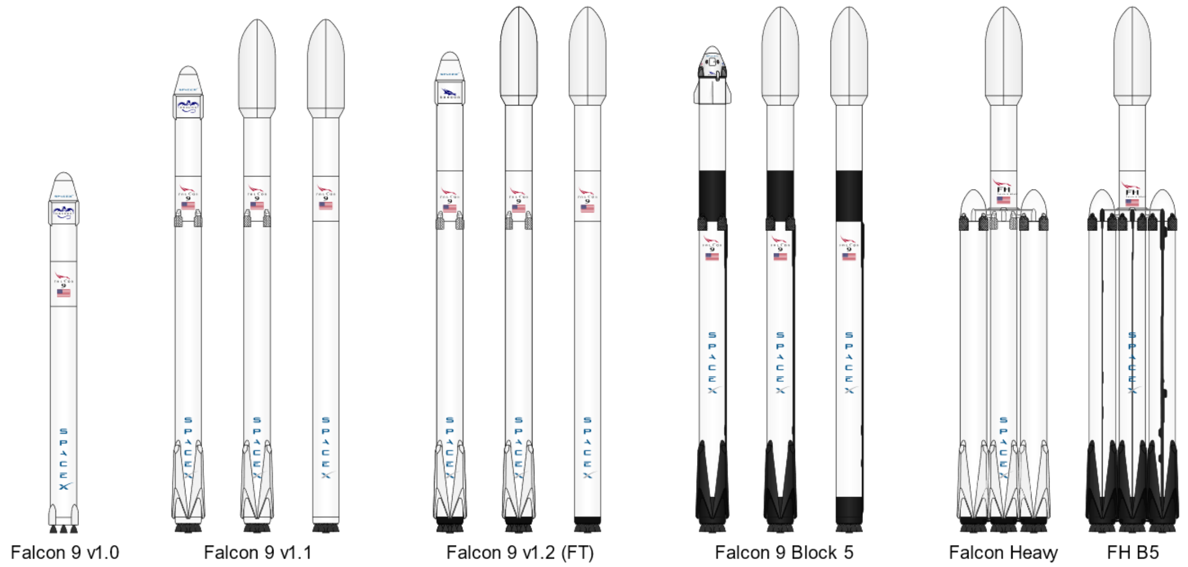
Methodology

Methodology: Executive Summary

- Data collection methodology:
  - The data was collected from a publicly available SpaceX API, and web scraping of wiki pages about SpaceX rocket flights information
- Performed data wrangling
  - Data was preprocessed to replace missing values
- Performed exploratory data analysis (EDA) using visualization and SQL to better understand the data
- Performed interactive visual analytics using Folium and Plotly Dash to better relate the data to the geographical locations
- Performed predictive analysis using classification models
  - Used Skit-learn to build, tune, evaluate classification models to find the best one.

Data Collection

- Data was collected from a publicly available [SpaceX API]
- In addition to web scraping of [wiki pages] SpaceX rocket flights information



Data Collection – SpaceX API

- URL for a notebook explaining the detailed steps and code
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/1.%20data-collection-api.ipynb

Data Collection - Scraping

- A URL for a notebook explaining more details and the code
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/2.%20webscraping.ipynb

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Wrangling

- Data was processed to replace missing data
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/3.%20Data%20wrangling.ipynb

EDA with Data Visualization

- Exploratory data analysis was performed using sqllite
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/4.%20EDA%20with%20sqllite.ipynb.ipynb
- In addition to Exploratory data analysis with data visualization
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/5.%20EDAwithDataViz.ipynb

Build an Interactive Map with Folium

- To better understand the geolocation relationships Folium was used to show the data on the map with markers for various launch sites
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/6.%20GeolocationAnalysis.ipynb

Build a Dashboard with Plotly Dash

- To better visualize the data an interactive a dashboard was built using Plotly Dash library.
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/7.%20Plotly_Dash_app.py

Predictive Analysis (Classification)

- Different ML predictive models were evaluated to select he best one.
- https://github.com/Abddo9/Data-Science-Projects/blob/main/SpaceX-Project/8.%20Machine%20Learning%20Prediction.ipynb

Results

EXPLORATORY DATA ANALYSIS RESULTS

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

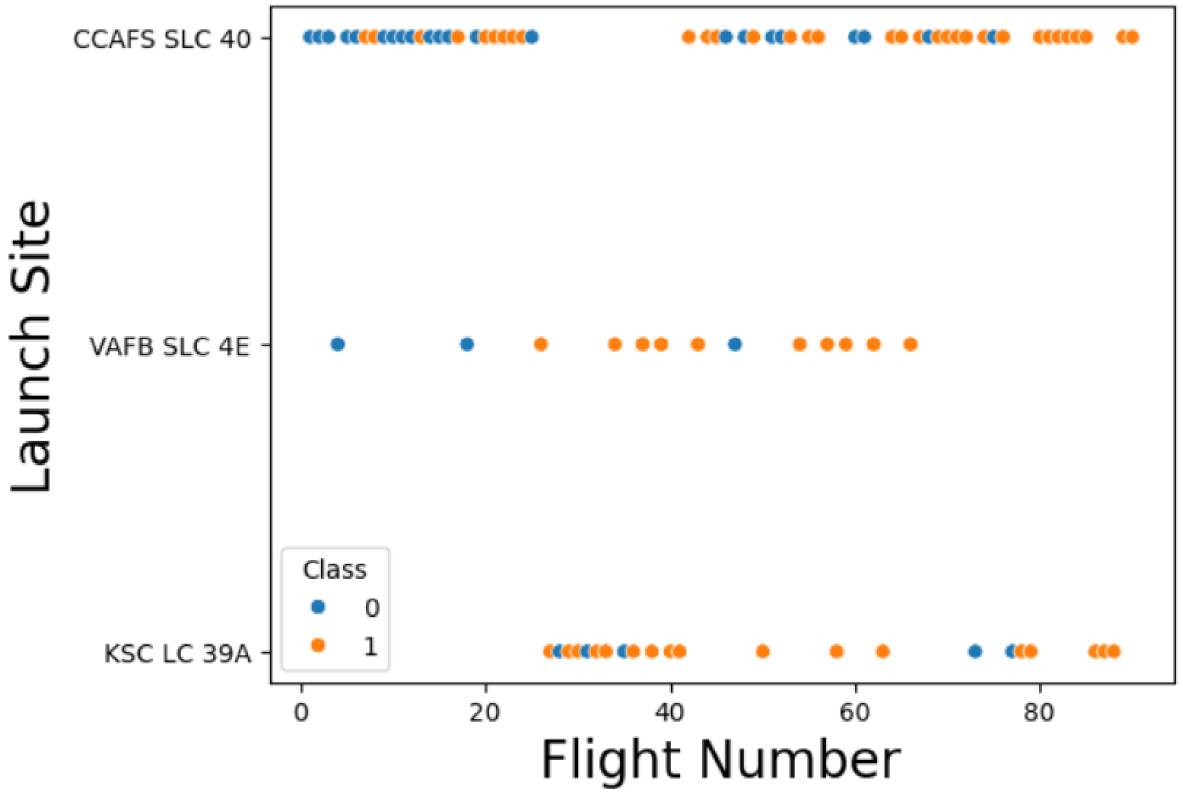
PREDICTIVE ANALYSIS RESULTS

Section 2

Insights drawn from EDA

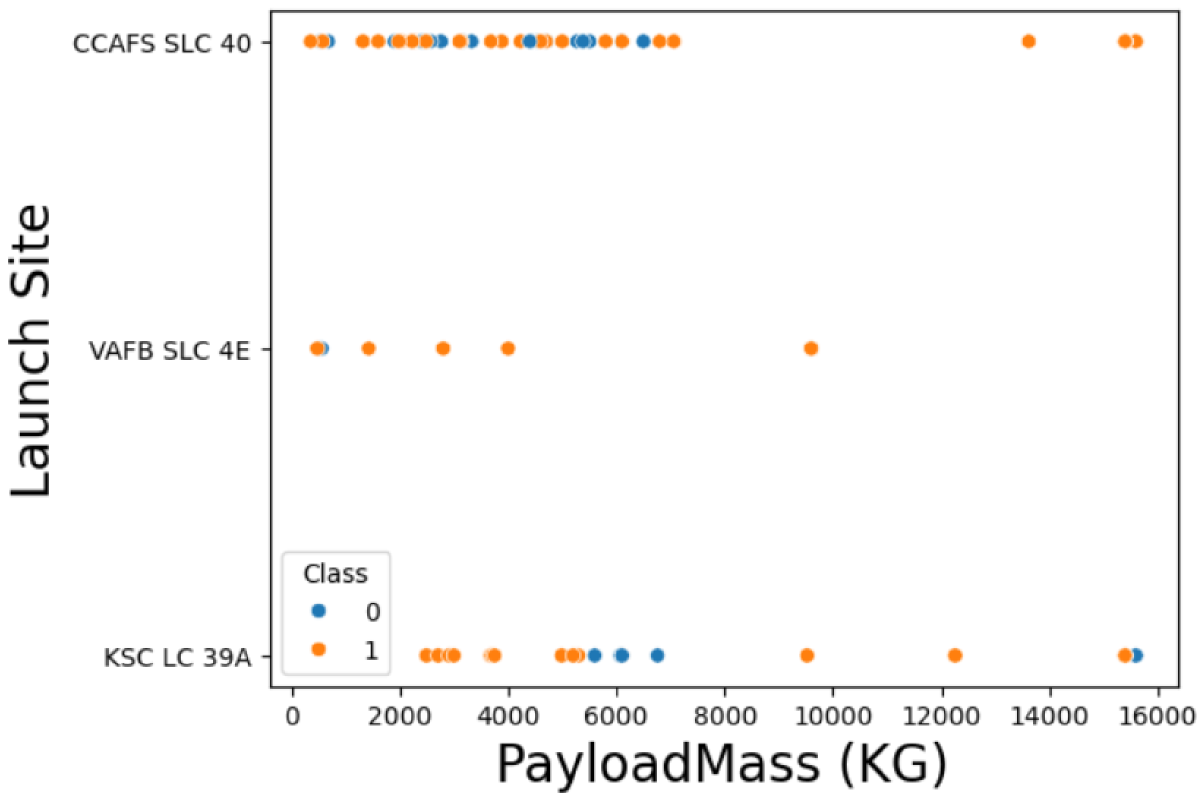
Flight Number vs. Launch Site

- For flight numbers after 40 the success rate is very high, however, it is always a success in the VAFB site after flight number 50, and after 80 for the other two sites
- No flight number for VAFB site after around flight number 70



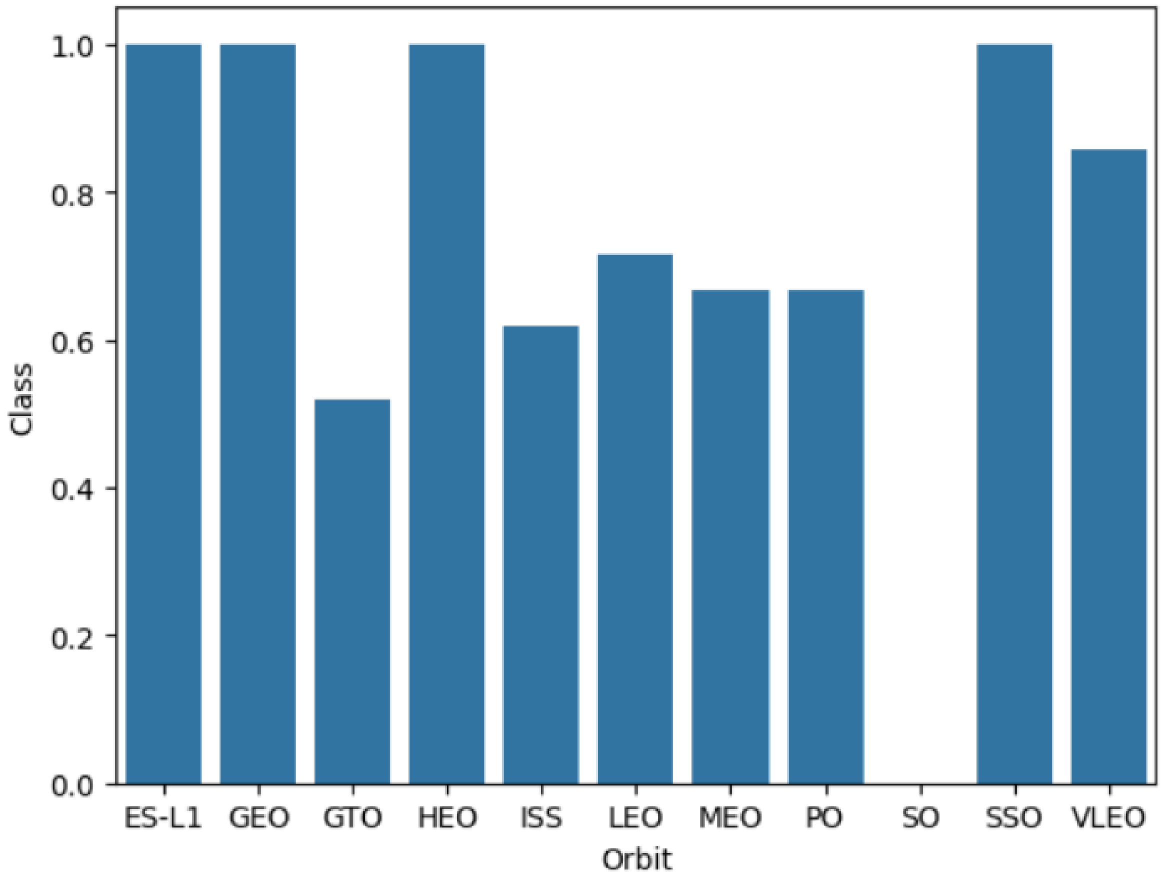
Payload vs. Launch Site

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



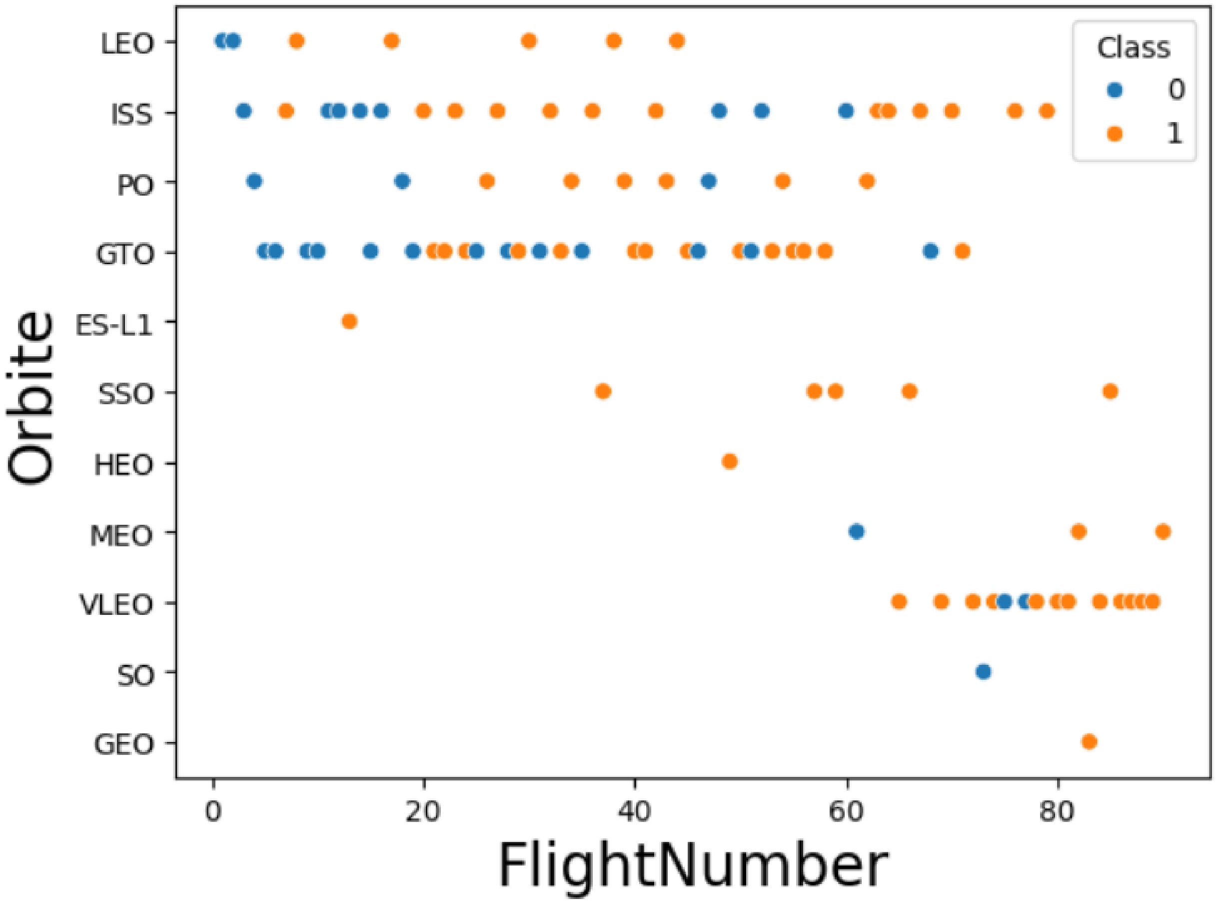
Success Rate vs. Orbit Type

ES-L1, GEO,HEO and SSO have the highest success rate of 100%



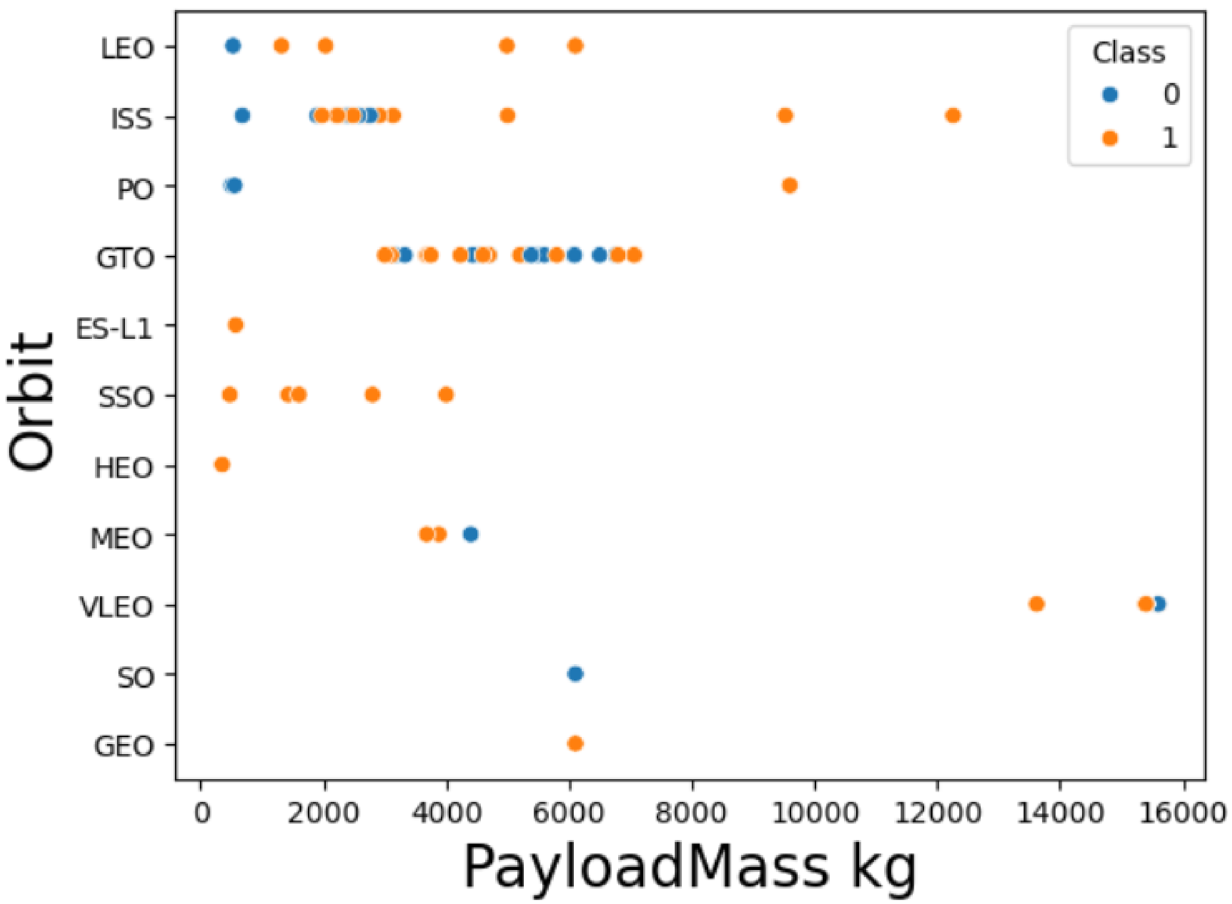
Flight Number vs. Orbit Type

- The LEO orbit the Success appears related to the number of flights;
- On the other hand, there seems to be no relationship between flight number when in GTO orbit



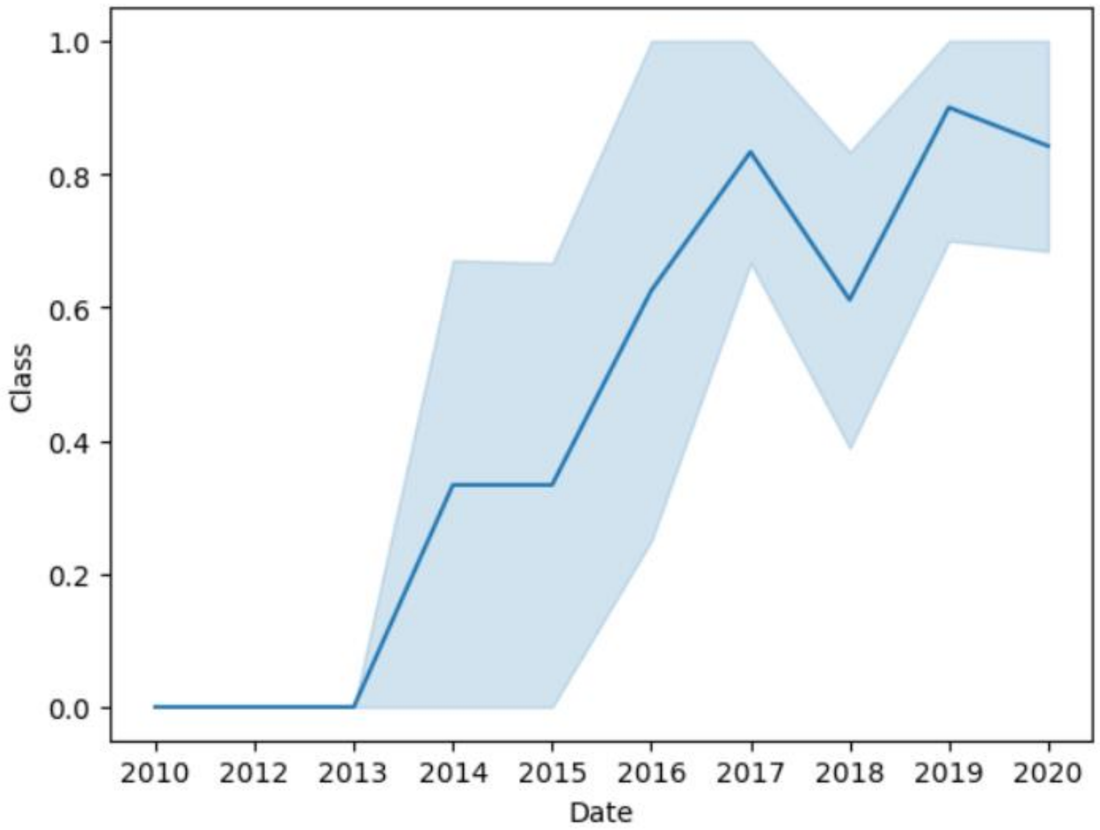
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here



Launch Success Yearly Trend

Sucess rate since 2013 kept increasing till 2020



All Launch Site Names

4 different launching sites

```
%sql select distinct Launch_Site From SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Here is the result, here we noticed that no explicit order was induced by the query itself.

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

sum(PAYLOAD_MASS__KG_)
45596

- The total payload carried by boosters from NASA
- Around 50 thousand KG in total

avg(PAYLOAD_MASS__KG_)
2928.4

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1
  - Around 3 Thousand KG.

min(Date)
2015-12-22

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad
  - Near the end of 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- 4 booster versions

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
- Different categories of success and failures are present in the data

Landing_Outcome	COUNT(*)
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- 12 boosters 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

month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

2015 Launch Records

- List of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Only 2 and both at CCAFS LC-40 site

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

- The ranking og 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
- No Attempt has the heist count of 10.

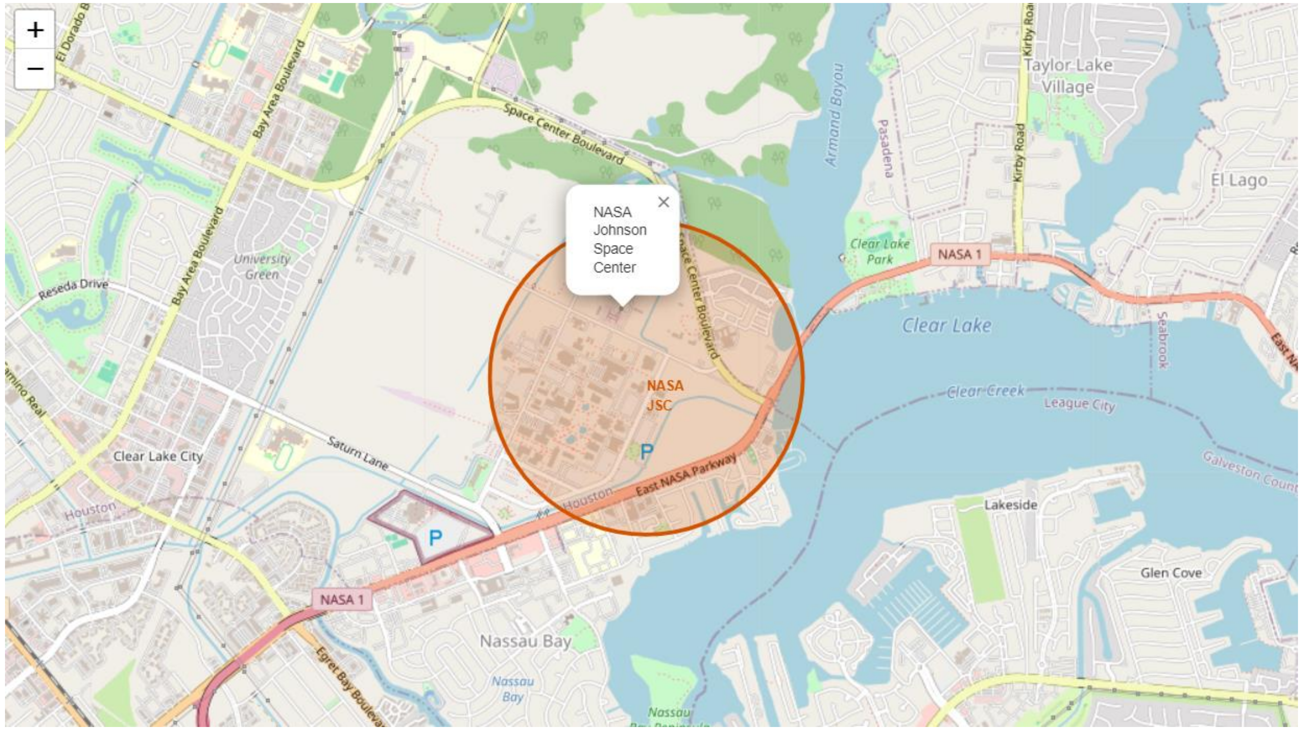
Landing_Outcome	COunt
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

The map centered at NASA Johnson Space Center

- The location is nicely shown with a circle marker, a popup and text.



Launching Sites

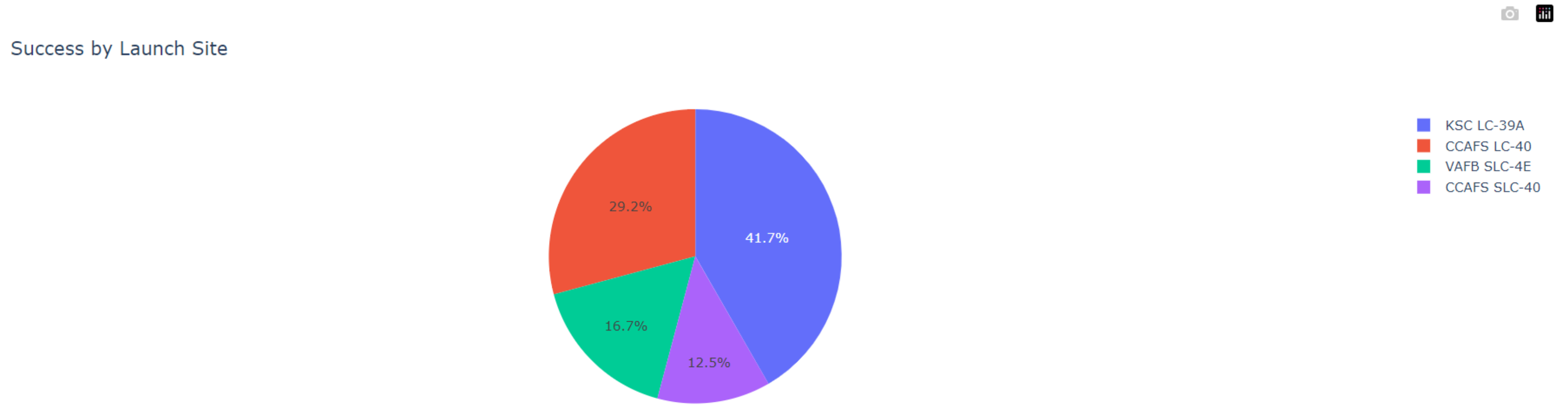
- The map is zoomed in it show a couple of launching sites

Marking of Success and Failure Missions

- Marking of all success and failure missions in the map
- Ability to see them in a summary top view or drill down to the details

Section 4

Build a Dashboard with Plotly Dash

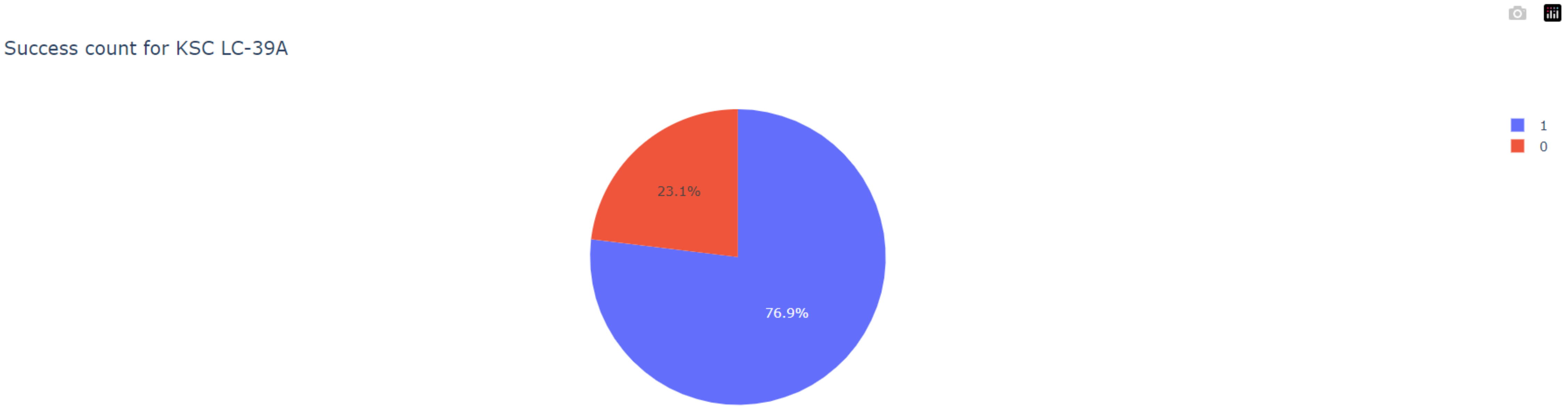


Launch Success Rates For All Sites

- We can see that KSC LC-39A has the highest contribution of successful missions (41.7% of all successful missions)

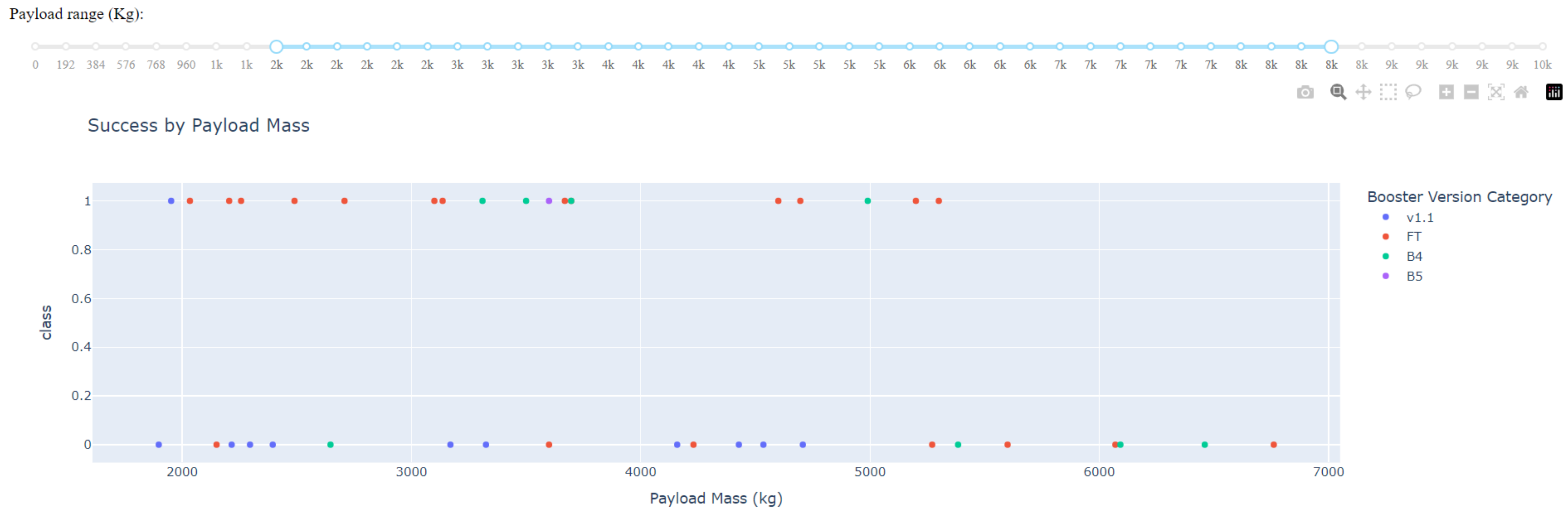
KSC LC-39A Mission Success and Failure

KSC LC-39A has around 75% successful missions



Payload VS Mission Success

A slider control to control the range of payload to disply



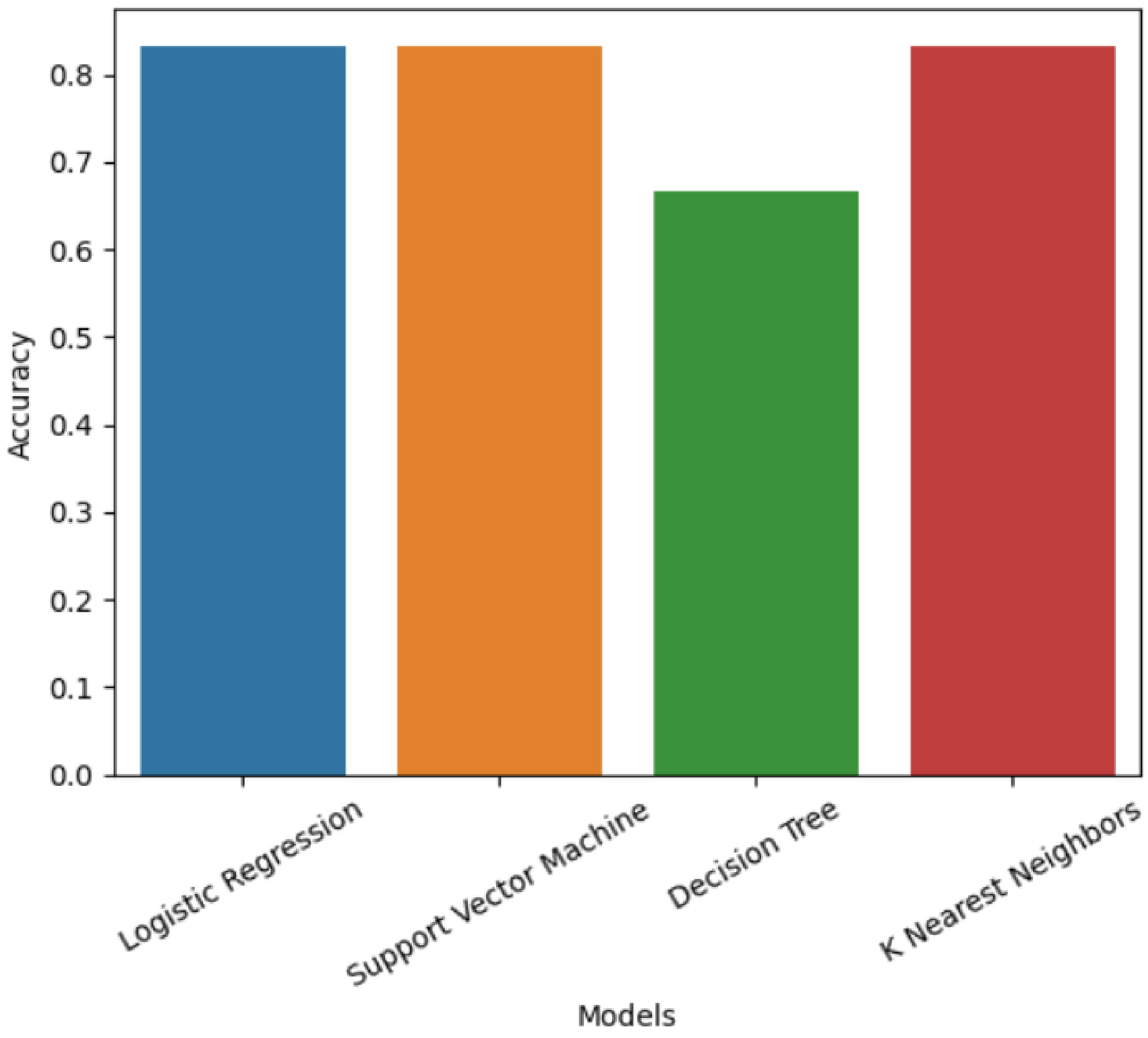
Section 5

Predictive Analysis (Classification)

Classification Accuracy

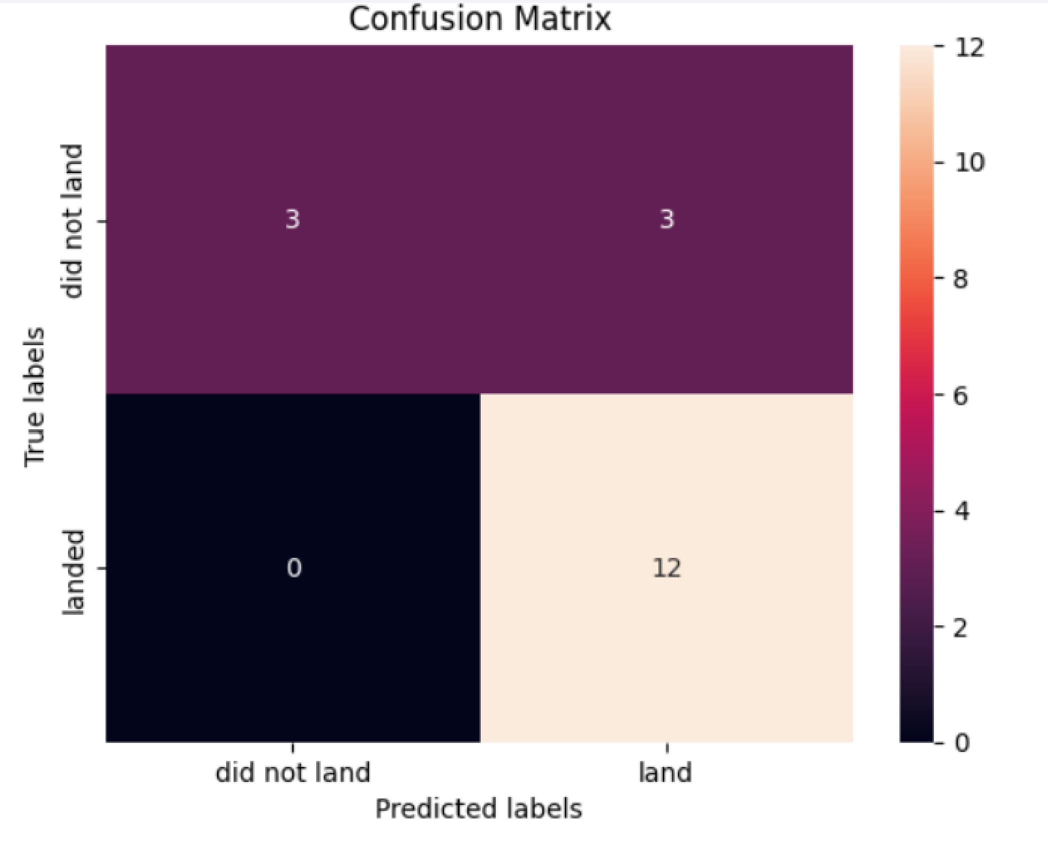
- Different classification models accuracy in the test data

- Decision Trees perfumed less than the others who all scored around 84% accuracy



Confusion Matrix For The Best Model

- The model can correctly predict all the correctly landed missions.
- However, it classified half of the failed landing missions as landed and the other half as not landed
- The model is not good in detecting negative cases



Conclusions

- In this work the data was collected from publicly available API and through web scraping and then cleaned and transformed.
- Exploratory data analysis and visualization was performed
- Various models were evaluated to find the best models.
- It was found that Logistic Regression, SVM and KNN were able to predict the success or failure of landing of the first stage with more that 80% accuracy
- More insight is that I think more data is needed to get better accuracy and to notice the difference between different models.

Appendix

- The link include an intermediate datasets during the cleaning and transformation
- https://github.com/Abddo9/Data-Science-Projects/tree/main/SpaceX-Project/data

Thank you!