IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Arvind Kumar Sharma
2023-09-06

Outline

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

Executive Summary

- SpaceX has been launching many rockets in minimal cost. This is due to their capability to reuse the first stage of the rocket. There has been many factor for this decisions. In this report we have analyze all the data and build a ML model to predict if Falcon 9's first stage will be reuse or not.

Introduction

- SpaceX has decreased the cost of sending something in the space very affectively. It is due to the reusability of first stage of Falcon 9. But how does anyone decides that if first stage is going to be reused or not. There could be many factors for it.
- We are trying to find the answer of this question. What are these factors? How they affect the decision and most importantly using a machine learning model to predict the decision for the future sinario.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - By REST api and web scrapping
- Perform data wrangling
  - Processed to form proper dataset
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build the linreg, svm, tree and knn model and evaluate on 18 sample.

Data Collection

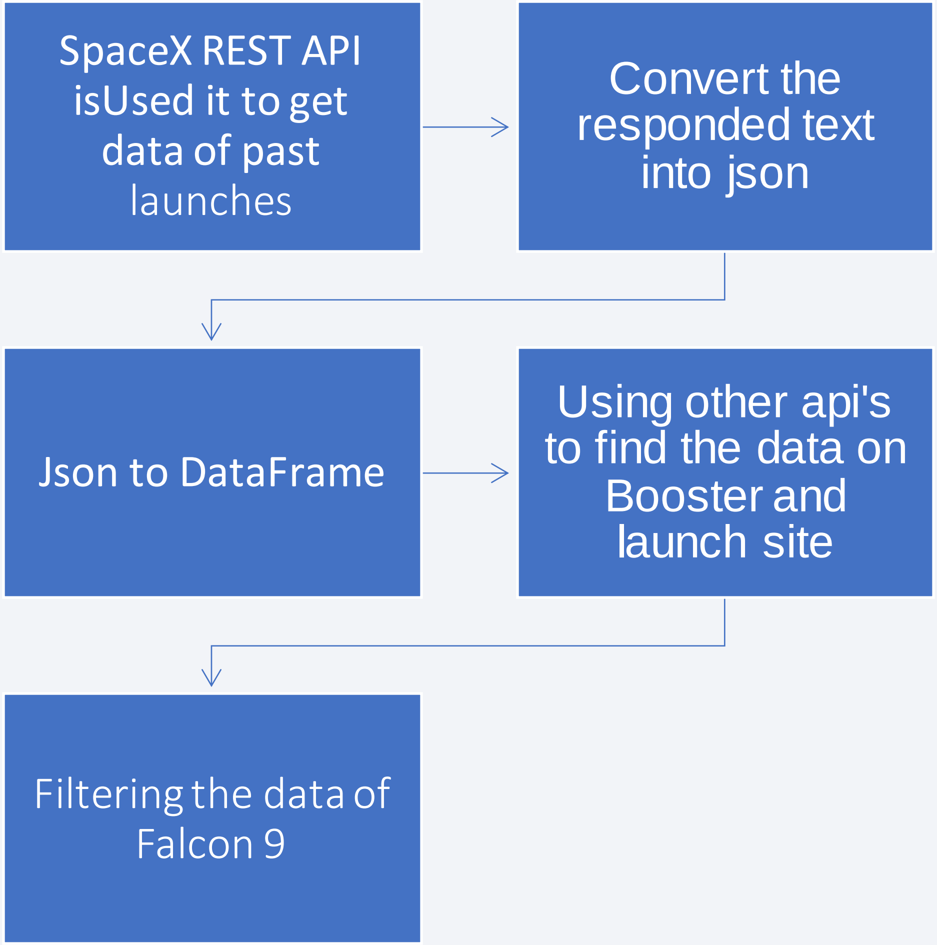
- Data Collection is done from two sources:

1. SpaceX REST API
2. Web scraping Wikipedia

The data of these two sources are then combined into one dataset.

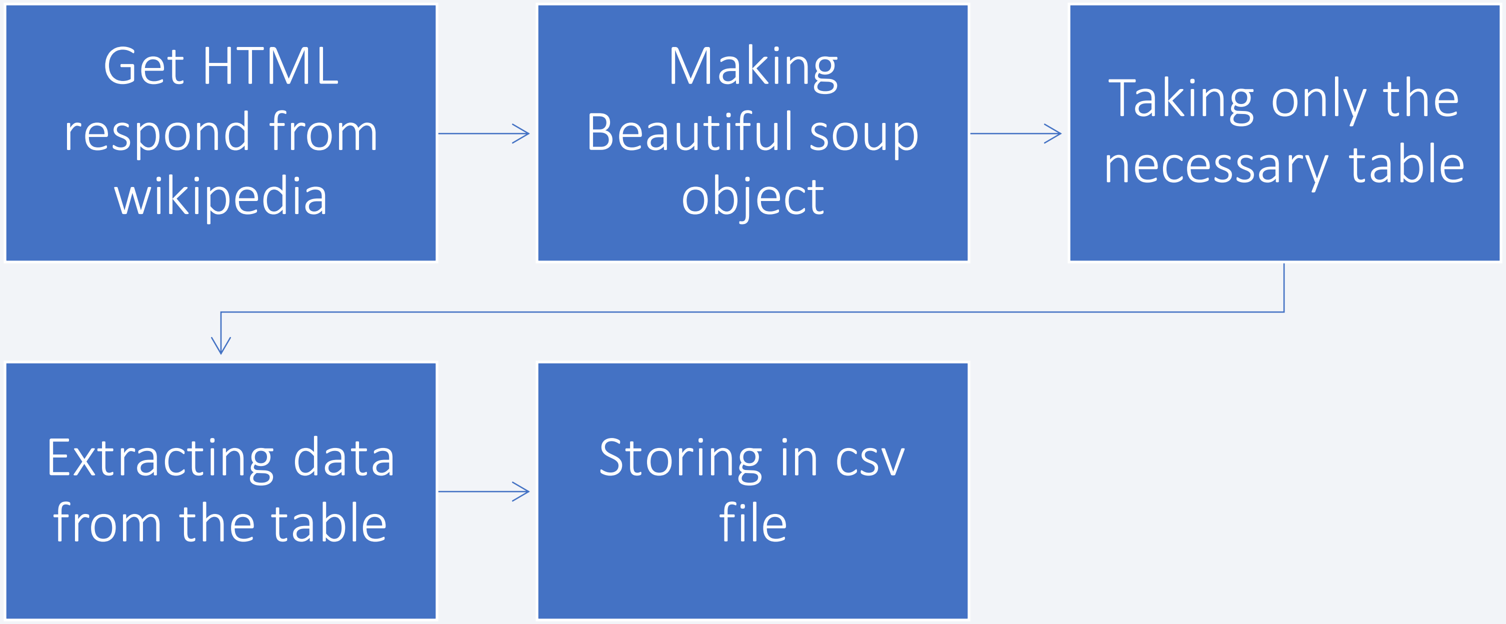
Data Collection – SpaceX API

https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb



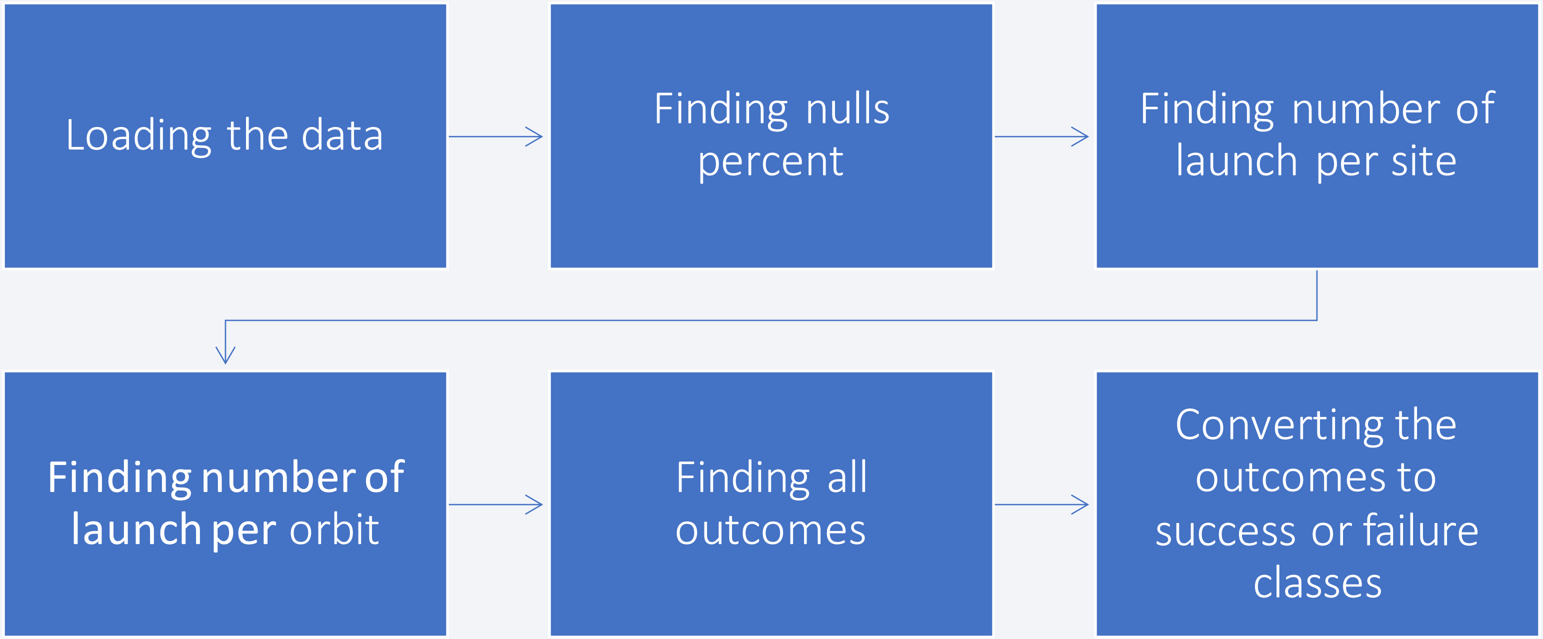
Data Collection - Scraping

- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb



EDA with Data Visualization

- We have used multiple charts and graph to identify the relation between different factors.
- This helped us to understand the relations between many events.
- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite(1).ipynb

EDA with SQL

- Explore the multiple point of the database.
  - Finding all the launch sites
  - Finding total payload mass for NASA
  - Average payload for booster F9 V1.1
  - Many other things
- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Dashboard save us from running and modifying the code repeatedly for different graphs.
- It help us to identity different relationship between attribute.
- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- We have trained multiple models on this dataset.
- Among all of them decision tree performs the best.
- This is due to the catagorical nature of decision tree
- https://github.com/ArvindSharma126/Applied-data-science-capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

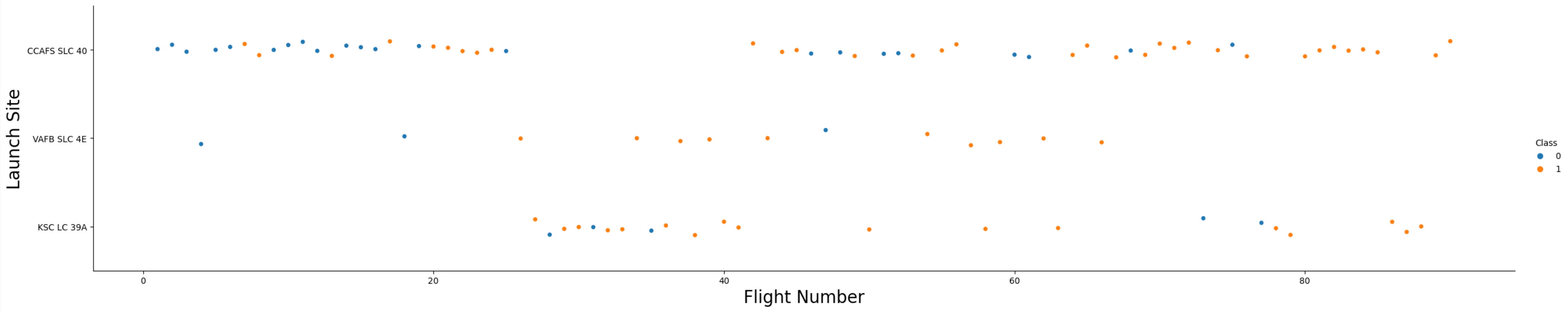
Results

- The EDA of the data shows some interesting fact.
- Some facts about the location and some other.
- Out of all decision tree perfoms the best

Section 2

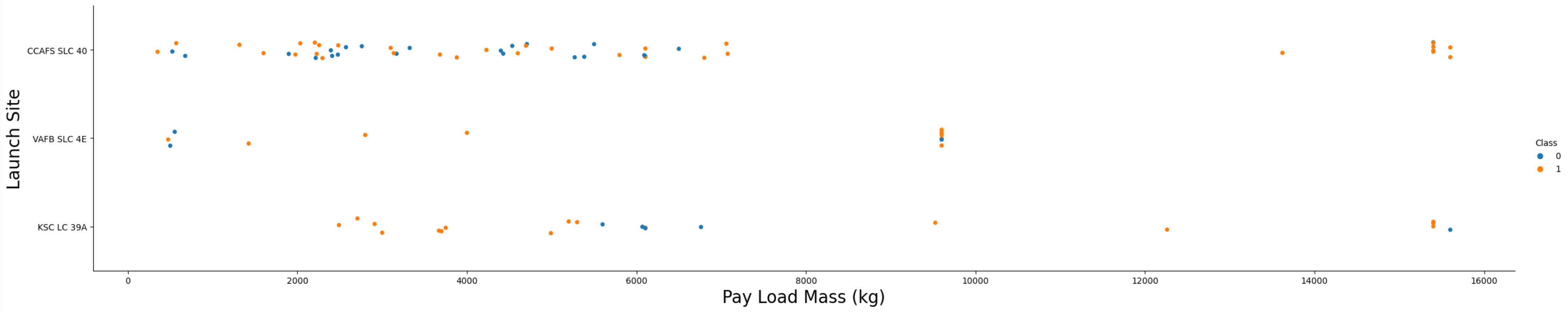
Insights drawn from EDA

Flight Number vs. Launch Site



- As the plot shows the number of launches are more from CCAFS SLC-40, and last of few of the launches from this site has more success rate compare to its beginning.

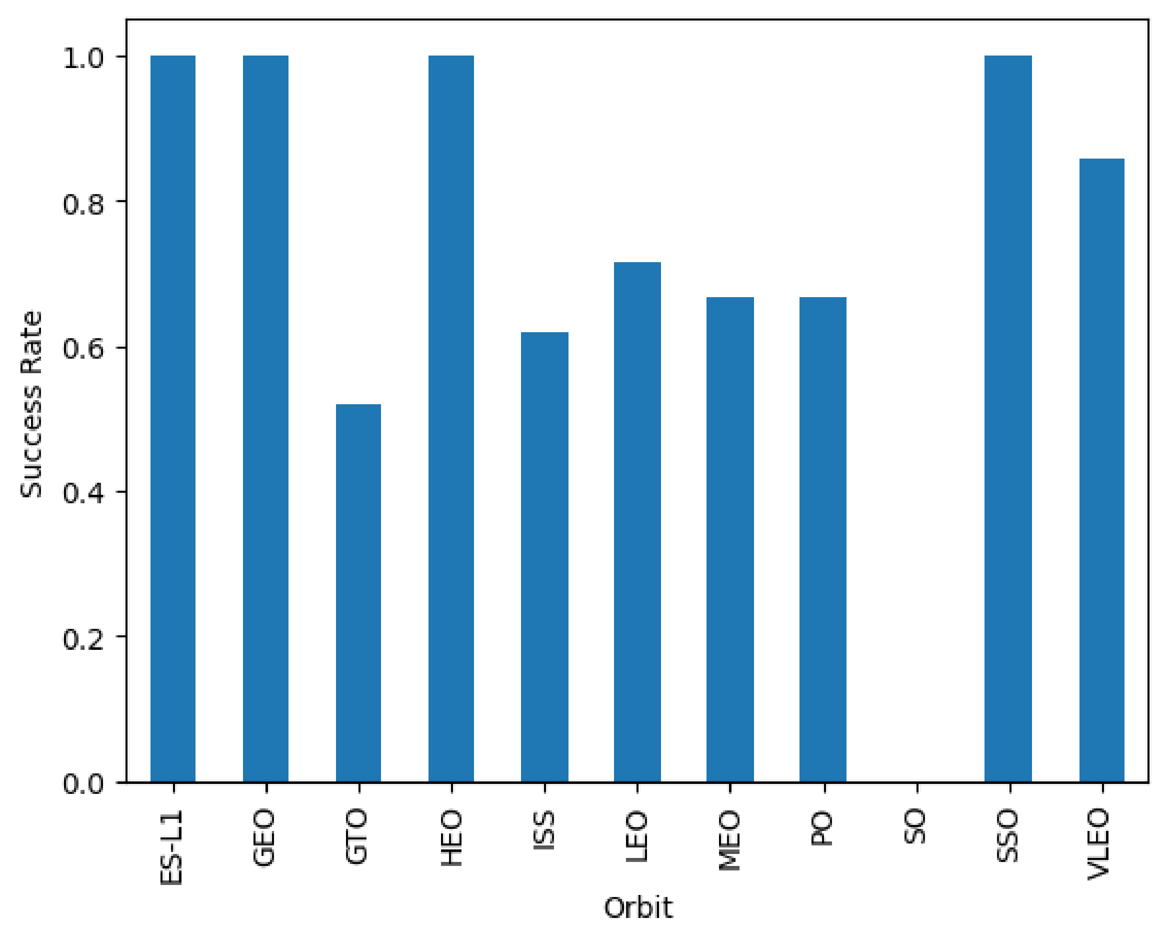
Payload vs. Launch Site



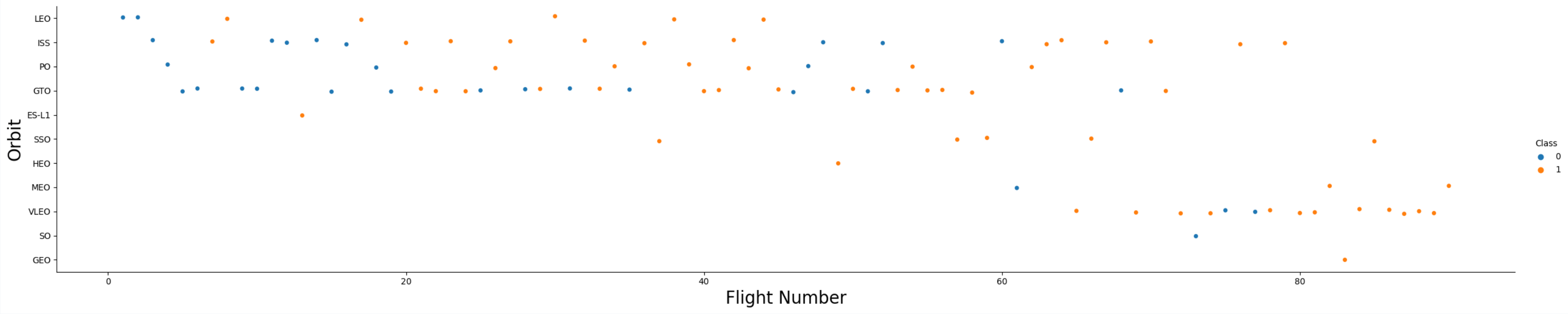
- The plot shows that the more number of launches are less than 8000 kg payload.
- A particular payload of nearly 9500 kg is from VAFS SLC-4E

Success Rate vs. Orbit Type

- The bar graph shows that ES-L1, GEO, HEO and SSO orbit has 100% success rate while SO orbit has 0% success rate.

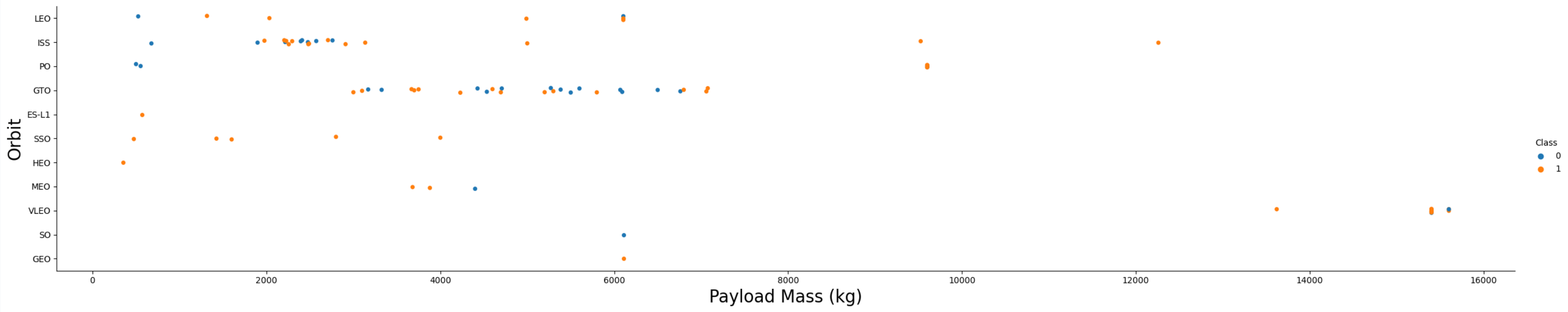


Flight Number vs. Orbit Type



- 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.

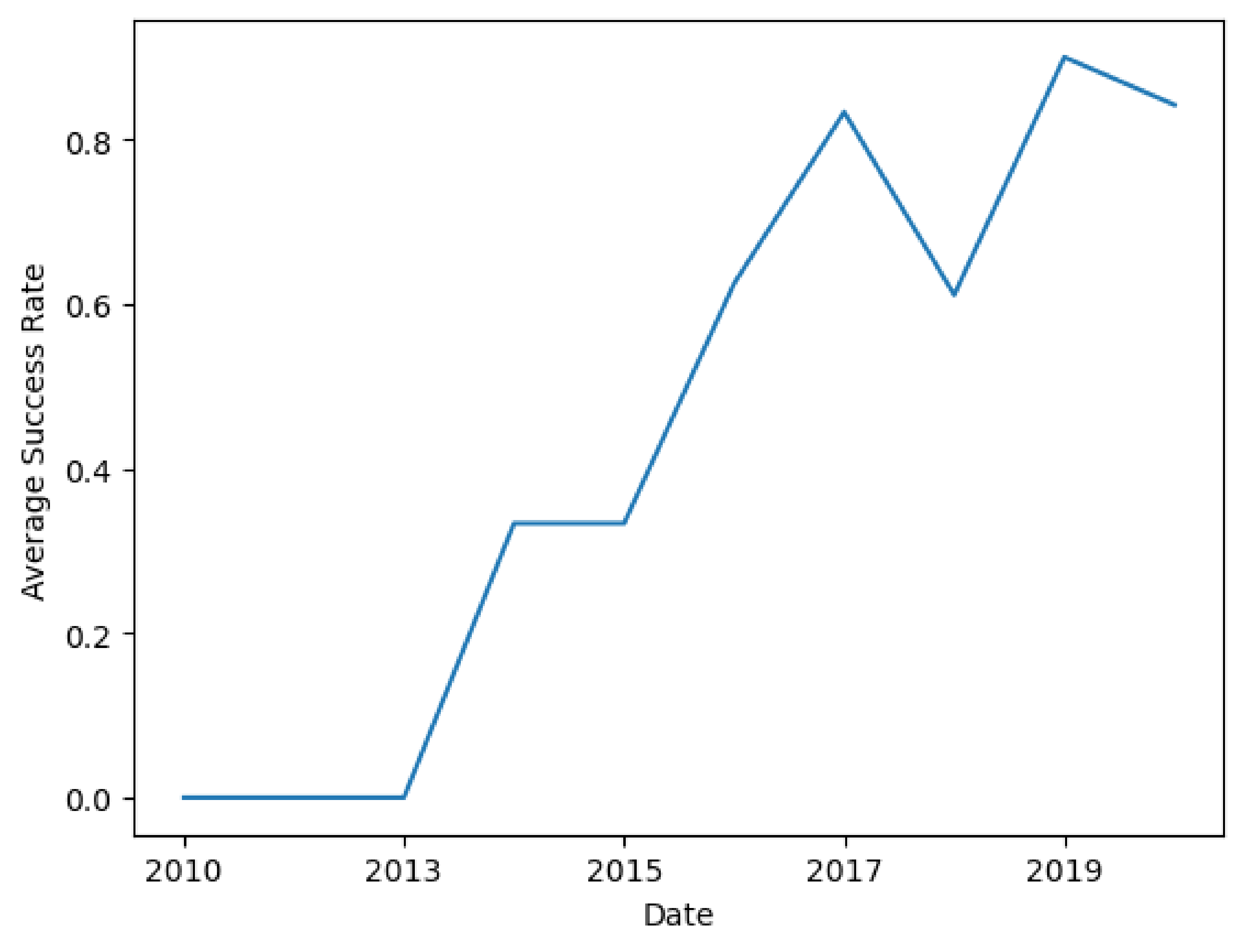
Payload vs. Orbit Type



- 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

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- There are basically 4 launch sites

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- First 5 launch site Names begin with 'CCA' are these.
- 4 of them is for NASA and all of them were success

Total Payload Mass

- The total payload carried by boosters from NASA is 45596 kg.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg
- Which see in the lower region the range of payload.
- It shows that F9 v1.1 is mostly used for lower payload missions.

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad 2015-12-22
- While the success rate start increasing in 2013 the first successful ground pad landing was done in late 2015.
- This is expected as before 2013 landing were unsuccessful and it would be better if these testing happened in remote areas like ocean.

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 are:

1. F9 FT B1022
2. F9 FT B1026
3. F9 FT B1021.2
4. F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Expect than one most of them were success.
- This shows the capability of SpaceX.

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

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass ->

- Most of them are F9 B5 series this mean that this booster is used mostly for the heavy payload mission.

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

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

- There has been only two flight failure in 2015 with drone ship.
- Both from the same booster and launching site.

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

- During the duration of Between 2010-06-04 and 2017-03-20 most of the landing outcome was 'No attempt' while least of the landing outcome was 'Precluded (drone ship)'.

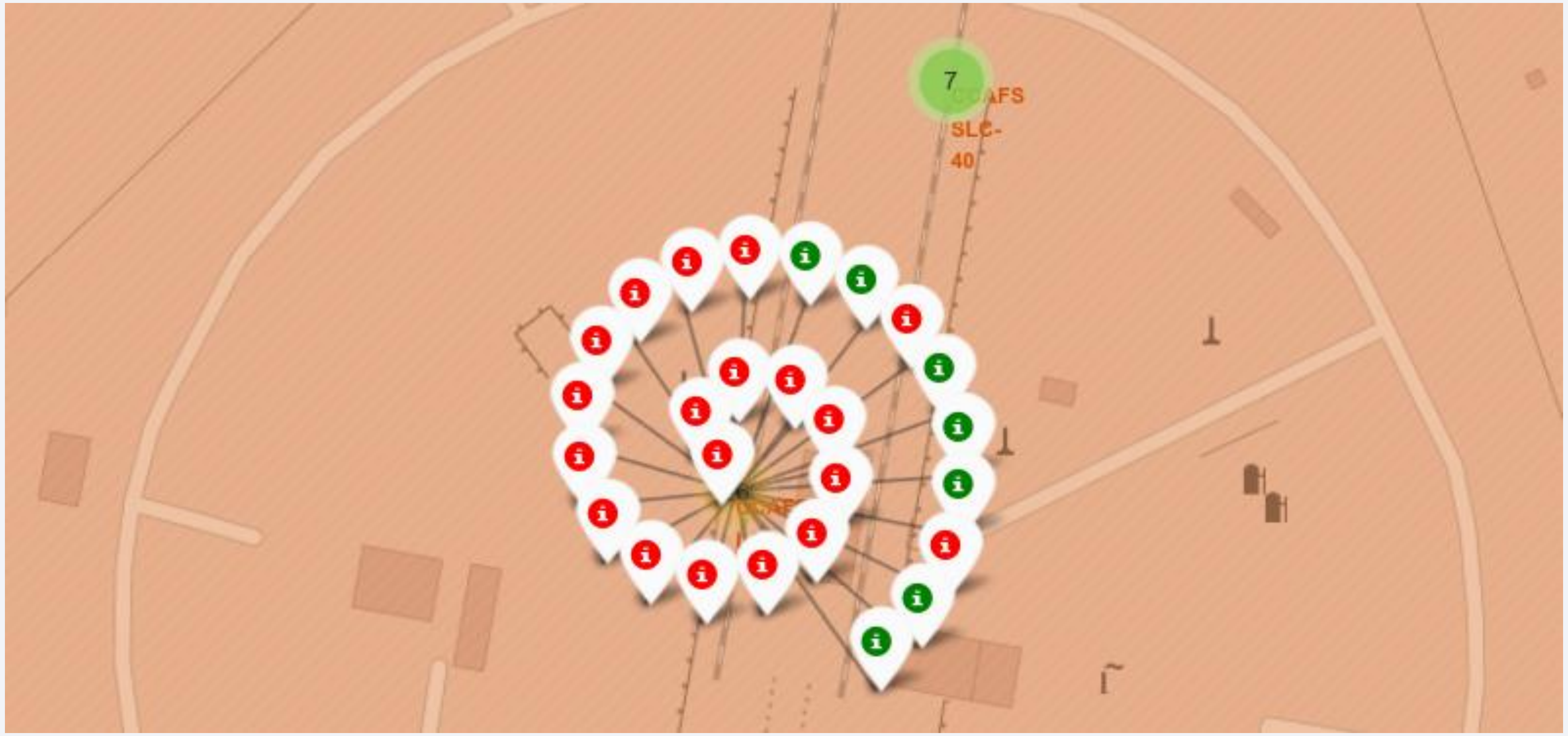
Landing_Outcome
No attempt
Failure (drone ship)
Success (drone ship)
Success (ground pad)
Controlled (ocean)
Uncontrolled (ocean)
Failure (parachute)
Precluded (drone ship)

Section 3

Launch Sites Proximities Analysis

Location of launch sites

Label of different launches from the same launch site



<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

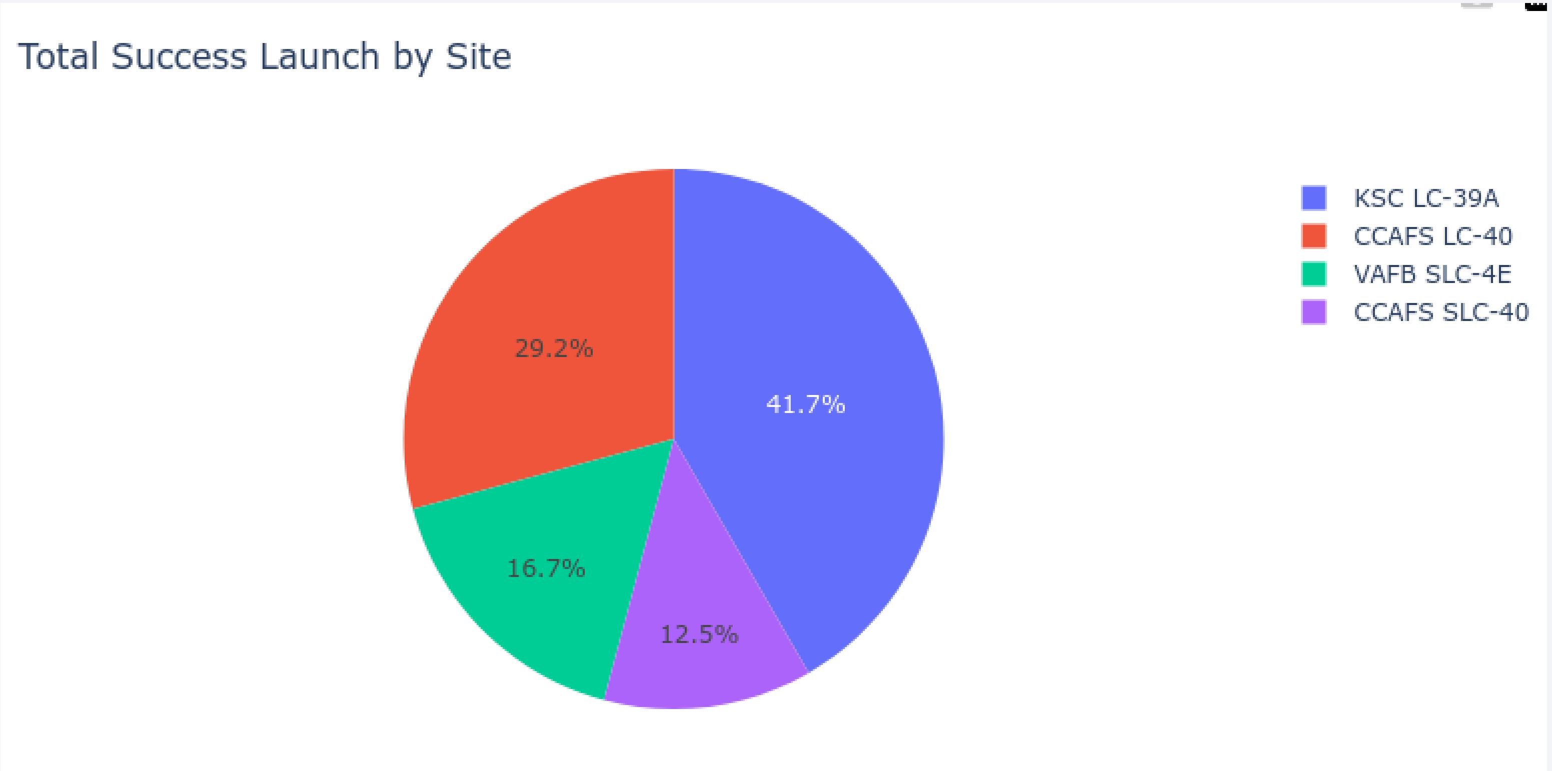
- Explain the important elements and findings on the screenshot

Section 4

Build a Dashboard with Plotly Dash

Success ratio based on launch site

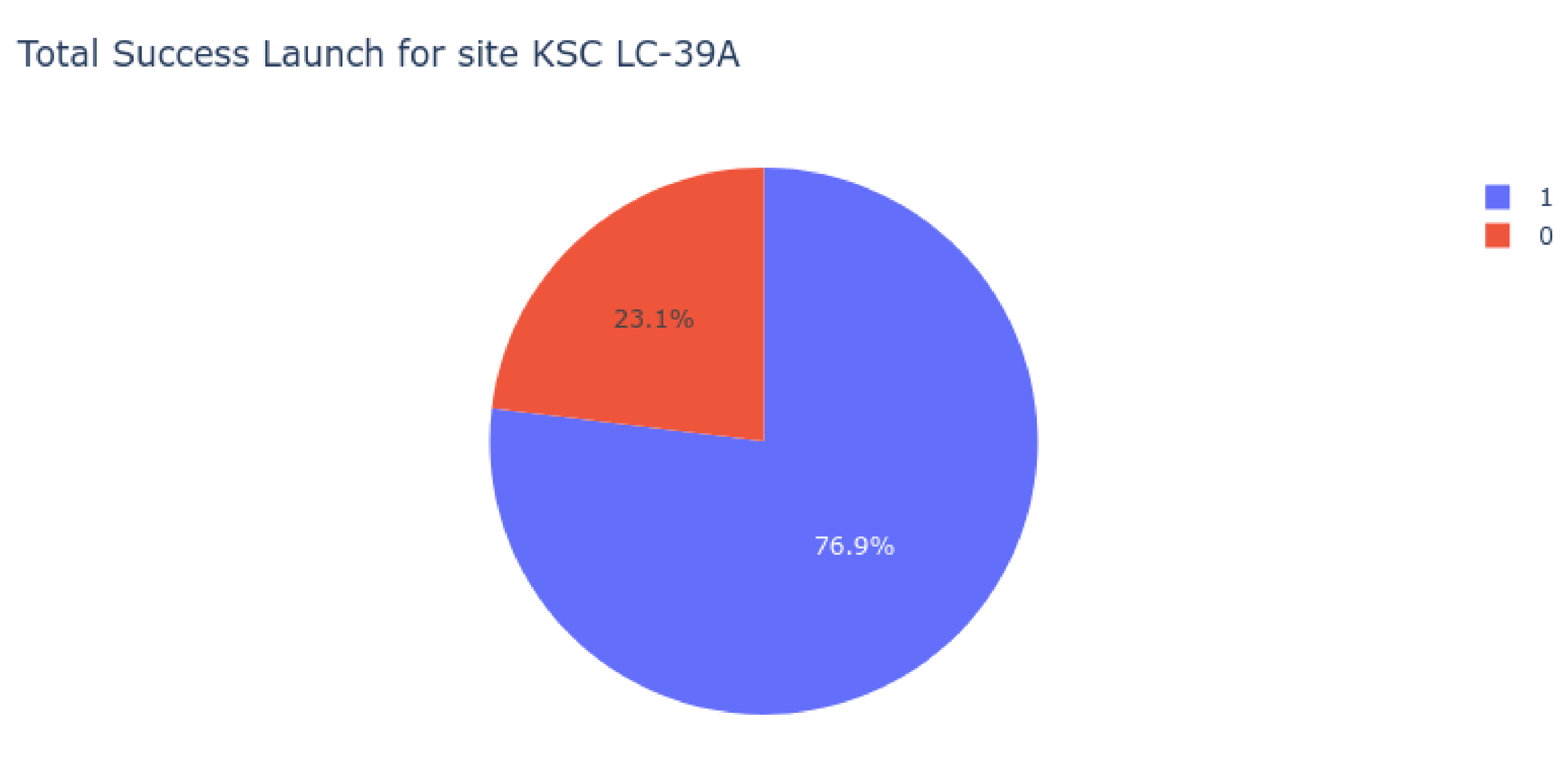
Success ratio is given in this pie chart. This show the ratio of Successful landing with respect to the sites. KSC LC-39A has more Successful landing then all others.



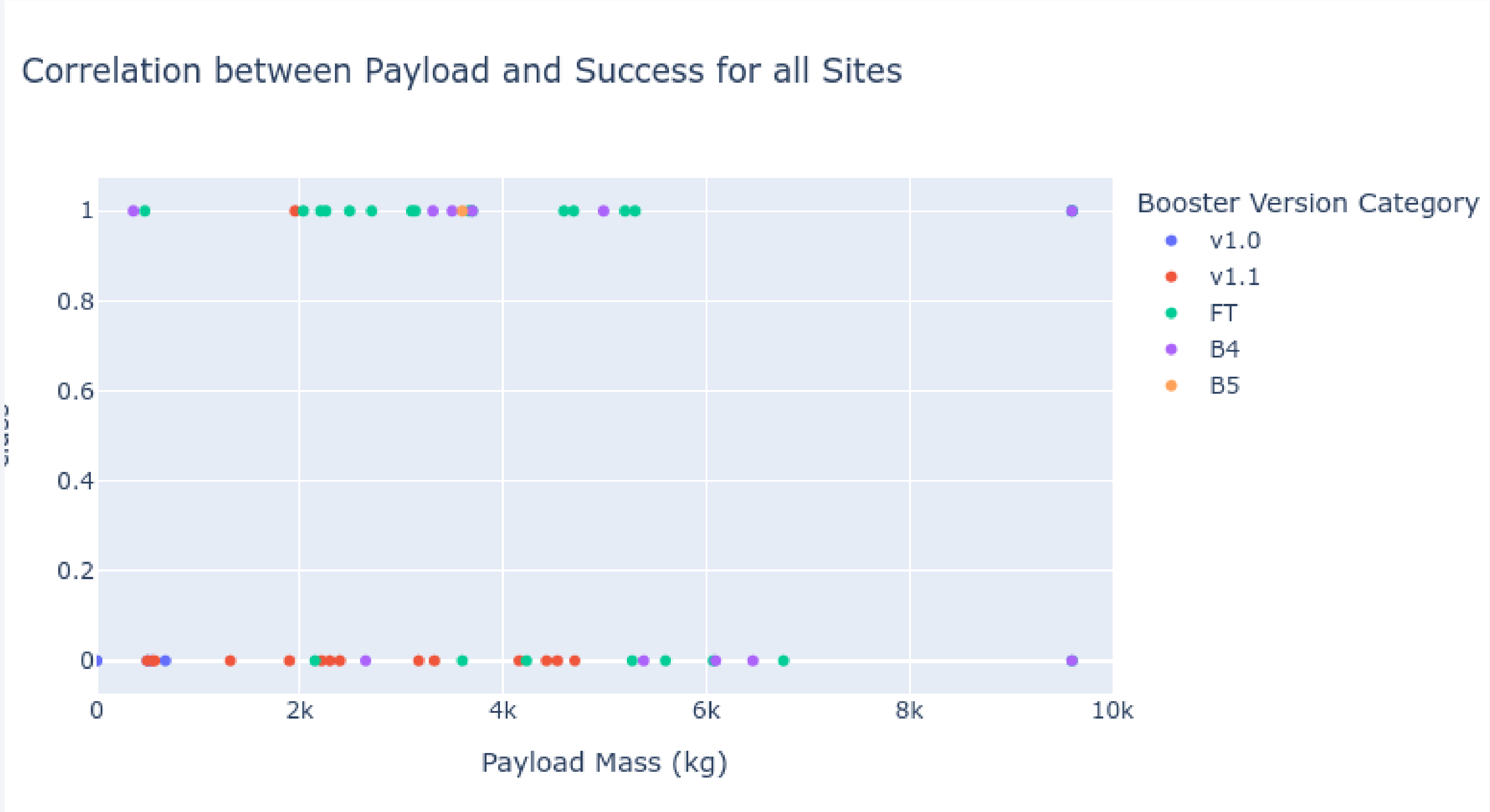
Success rate from site KSC LC-39A

The success rate from site

KSC LC-39A is highest among all four site. The success rate is 76.9%



Correlation between Payload and Success

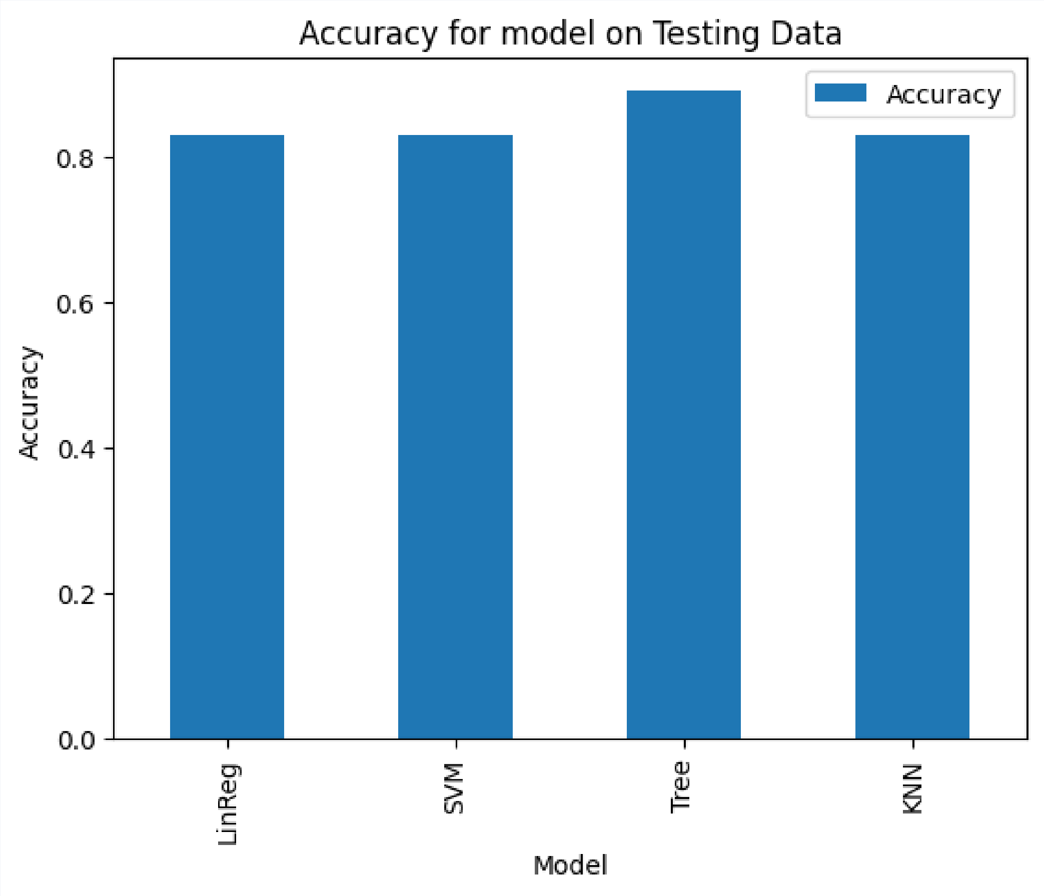


Section 5

Predictive Analysis (Classification)

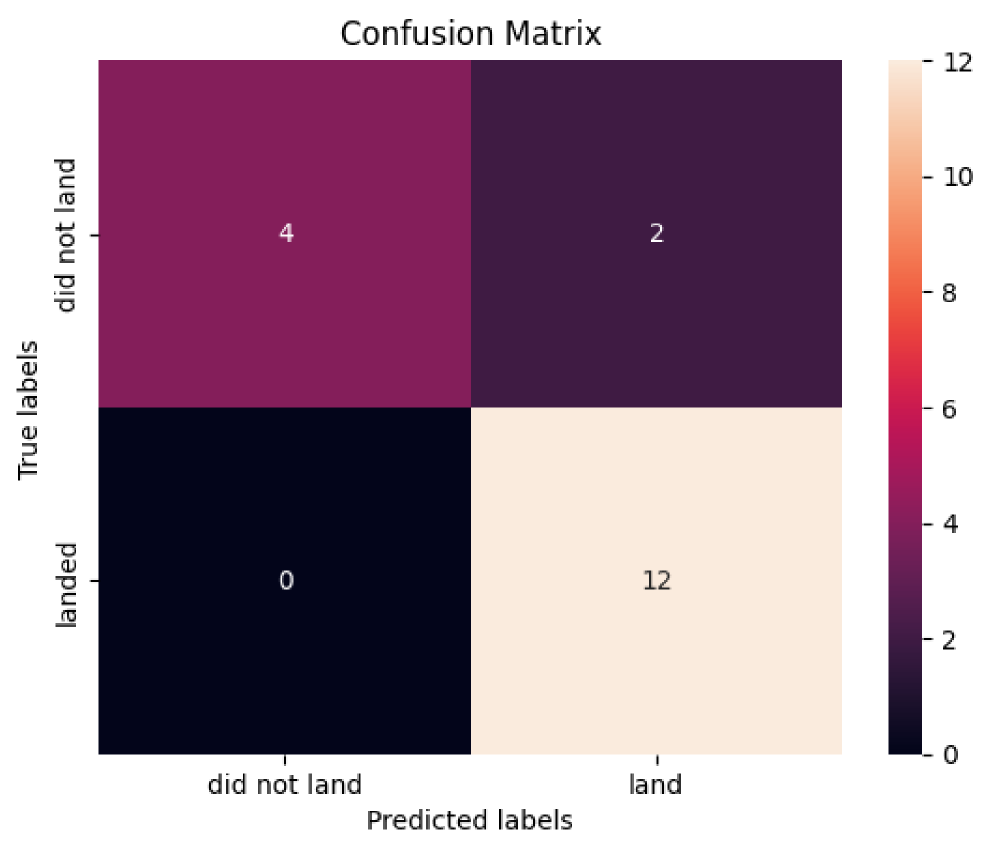
Classification Accuracy

- Most of the models give output based on the values but Decision Tree performs well because of higher number of categorical data.
- Due to this Decision Tree is the best fit model for this.



Confusion Matrix

- As shown in the confusion matrix the label "landed" has higher accuracy as this is always planned by the SpaceX but in the case of "did not land" the model is having less accuracy because this could be due to the client who does not want the first stage to be reused which cannot be predicted by out model as there is not data for it.



Conclusions

- SpaceX is very efficient in the planning the type of landing they want.
- For particular orbits landing is completely predictable.
- We can predict about a landing outcome with given feature with almost 90% accuracy.

Thank you!