IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dennis
2024.11.18

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

SpaceX Falcon9 rocket launches usually cost less than other competitors. The reason of this is because of the reuse of the first stage of the launch.

This project is to estimate the SpaceX will reuse the first stage or not and use these information to determine the prices of launches.

Section 1

Methodology

Methodology

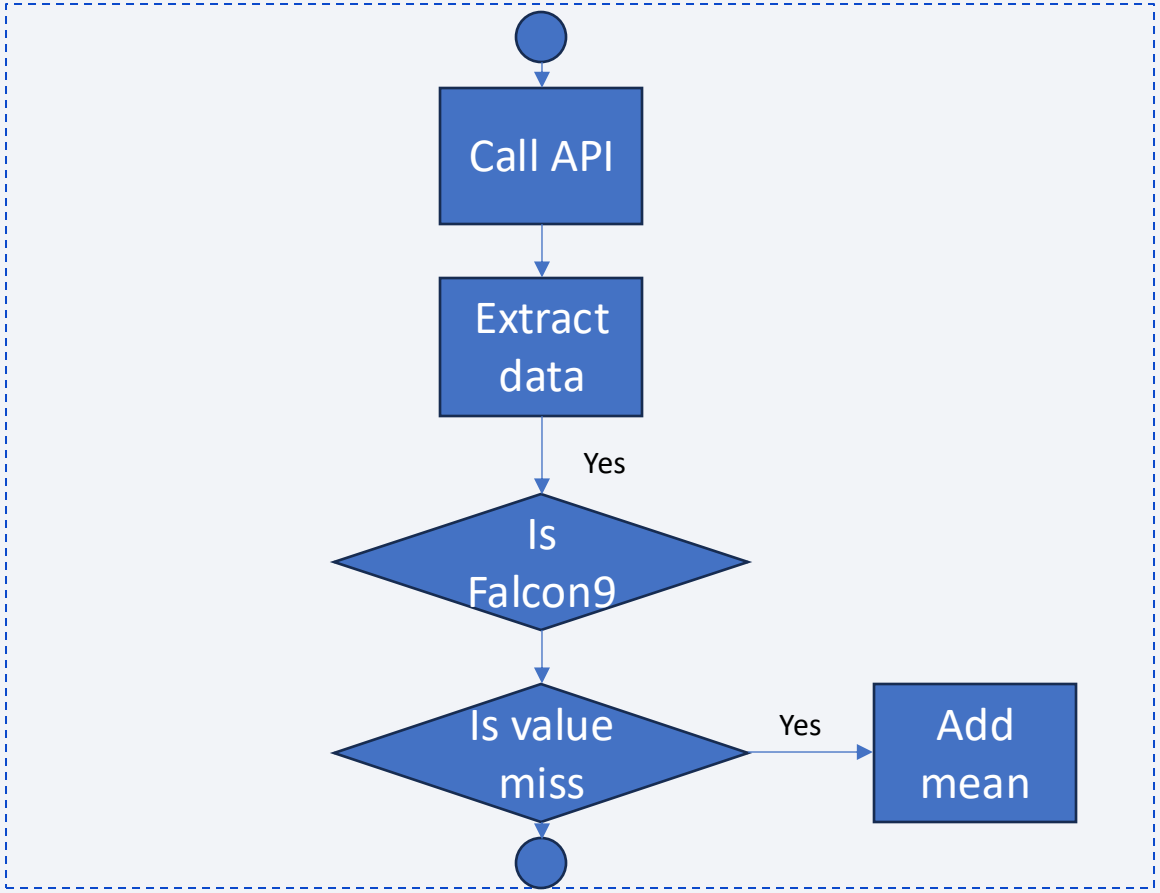
Executive Summary

- We use python to call the SpaceX API and the Wikipedia to collect data about the Falcon 9 launches.
- After the collect stage, we use the BeautifulSoup parse the information and transfer to the pandas dataframe for the latter usage.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

- Call SpaceX API to get the data
- Extract data from column
- Filter Falcon9 launches
- Handling the missing value

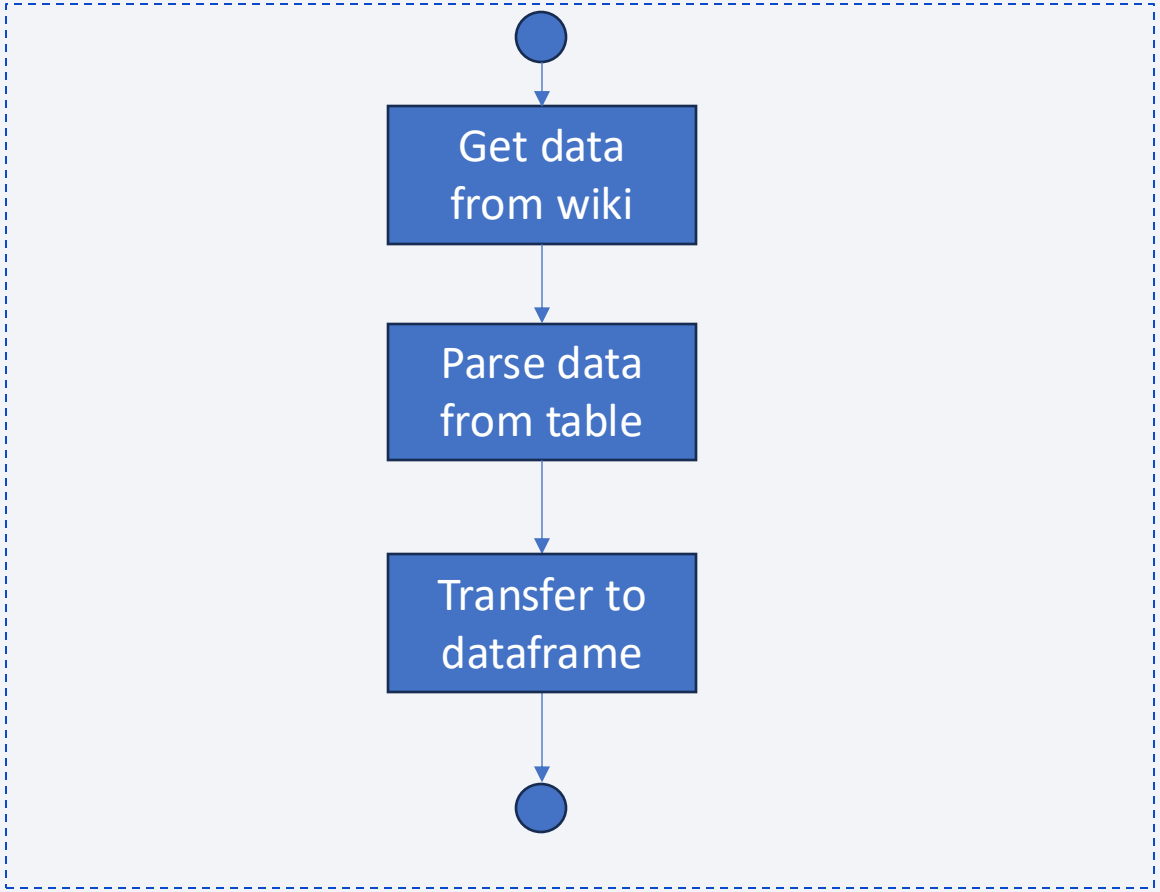
https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Get data from wikipage URL
- Parse data from table's column and row
- Transfer data to Pandas DataFrame object

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Using Scatter plot to find the relationship between two variables
- Using bar plot to find the highest successful rate with variables.

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/edadataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Add the scatter and pie chart for analysis.
- Using pie chart can see the total successful rate in which site and can go deeper to find out the percentage of successful rate in each site.
- Using the scatter plot can find out the relationship between payload mass and launch site.

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/dashboard.py

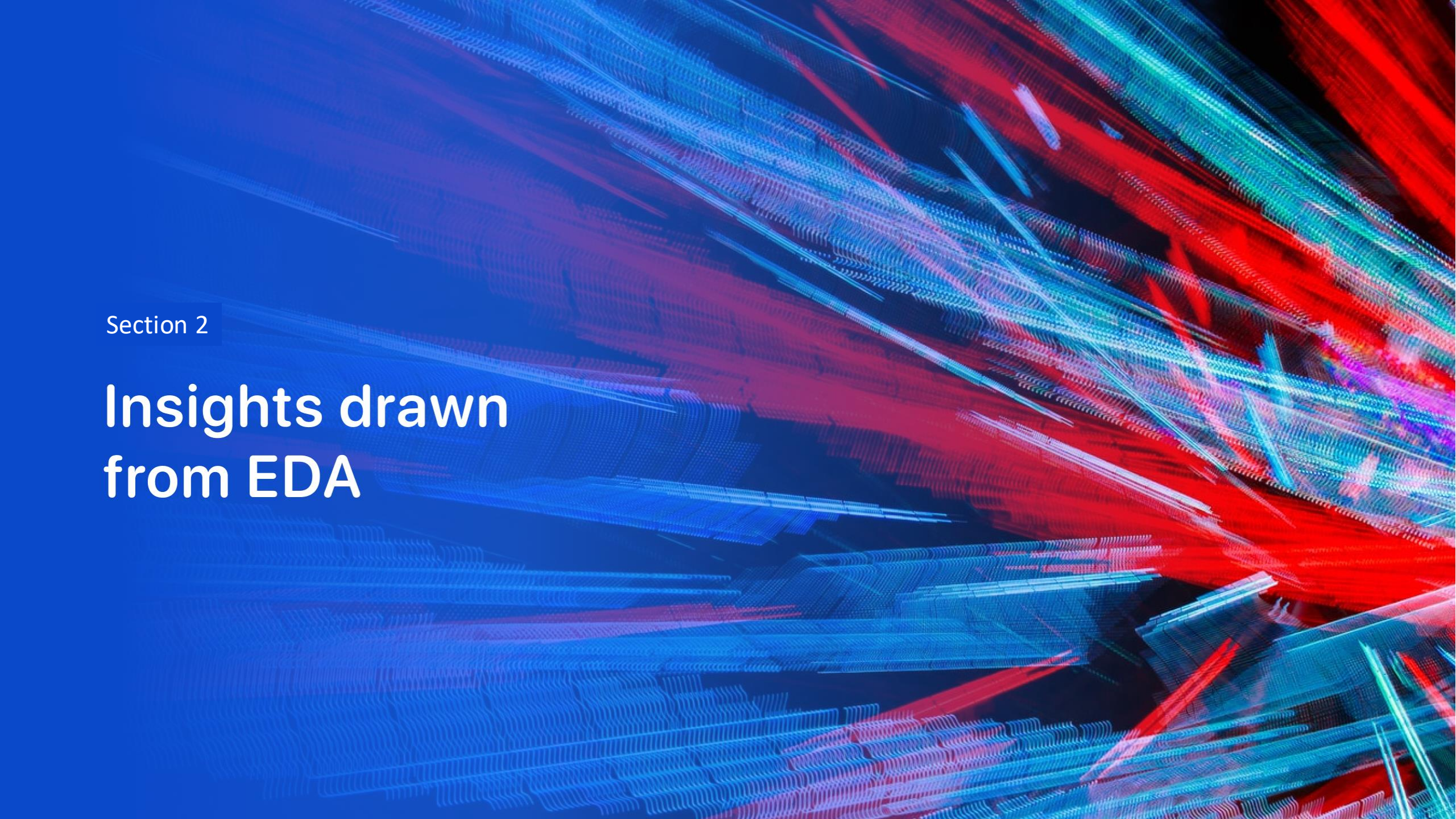
Predictive Analysis (Classification)

- Compare Logistic Regression, Decision Tree, KNN, SVM models' accuracy to choose the best model
- Using GridSearchCV to find the best parameter for all models
- The outcome shows the Decision Tree do the best job.

https://github.com/CCChou/HW-for-IBM-DS/blob/main/capstone/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

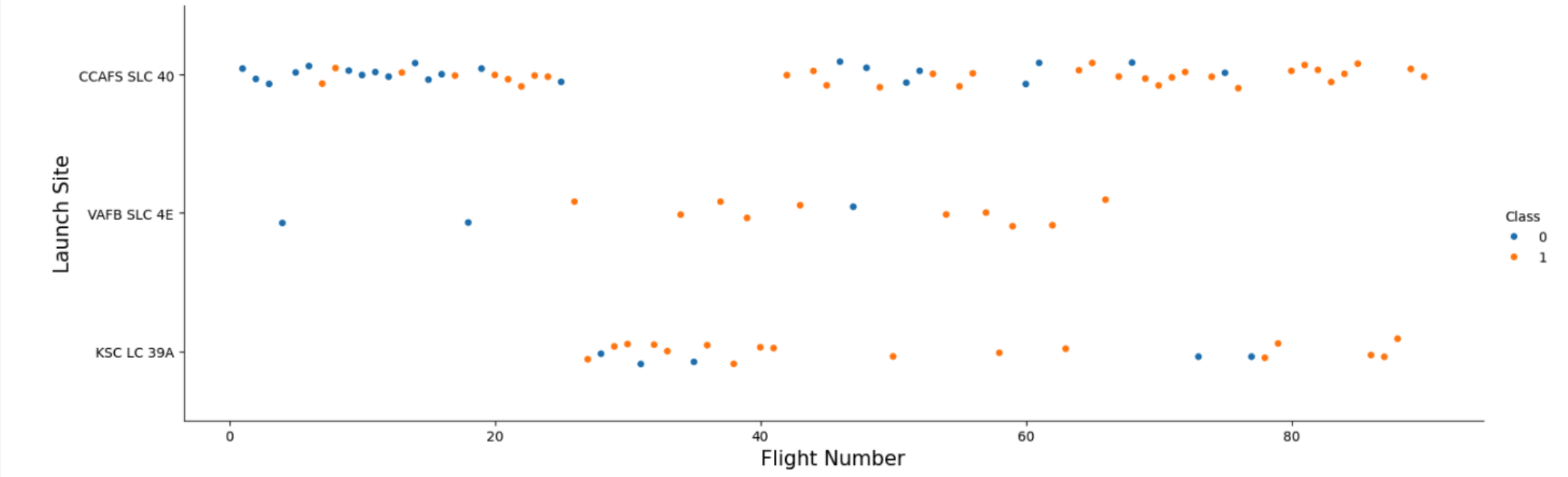
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

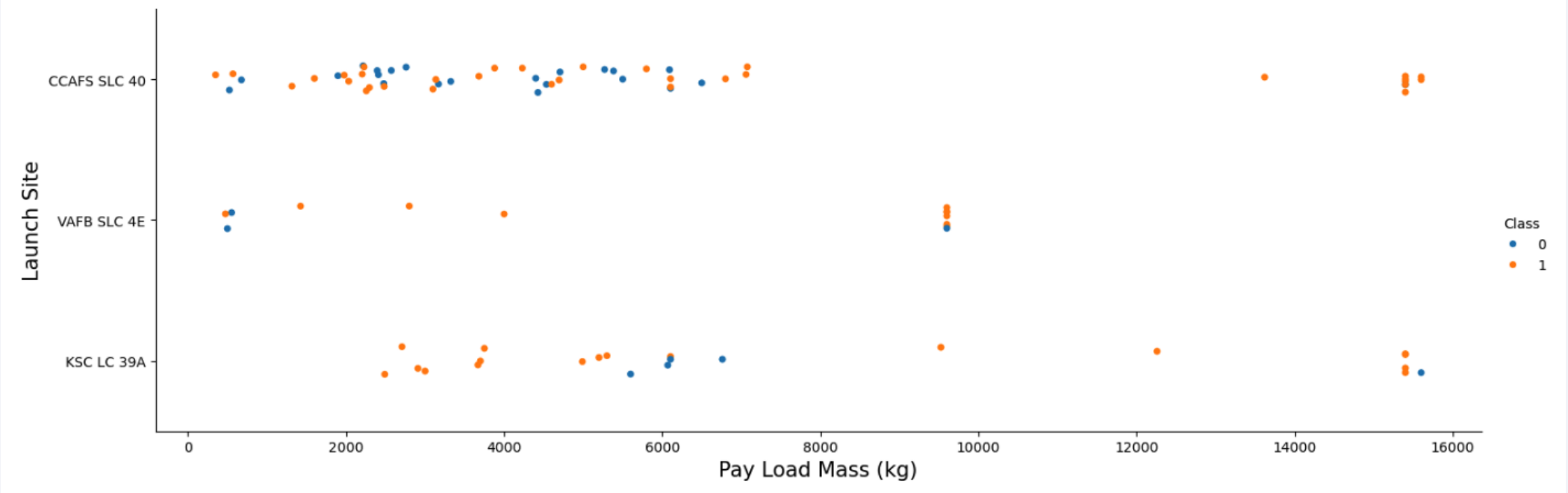
Flight Number vs. Launch Site

- More flight number comes with more success in every launch site



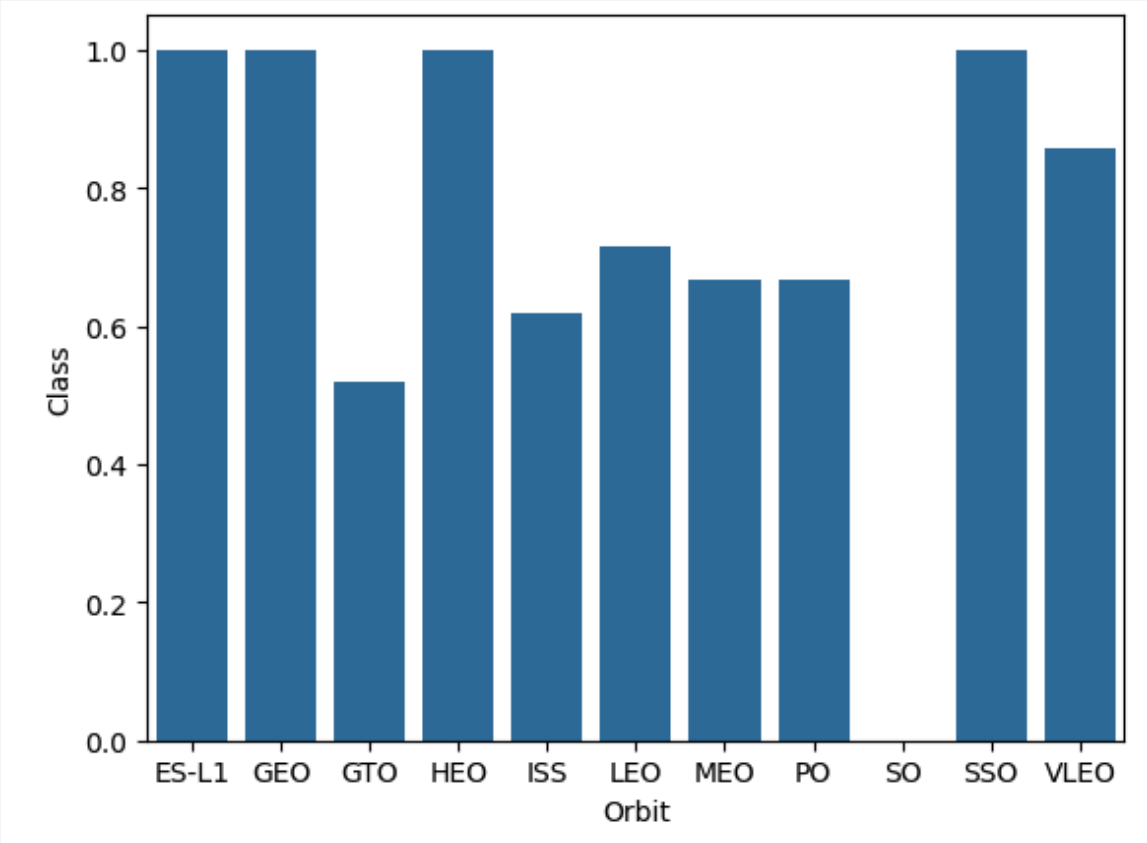
Payload vs. Launch Site

- The VAFB-SLC launch site has no rockets launched for heavy payload
- The KSC-LC launch site has more success for the light payload



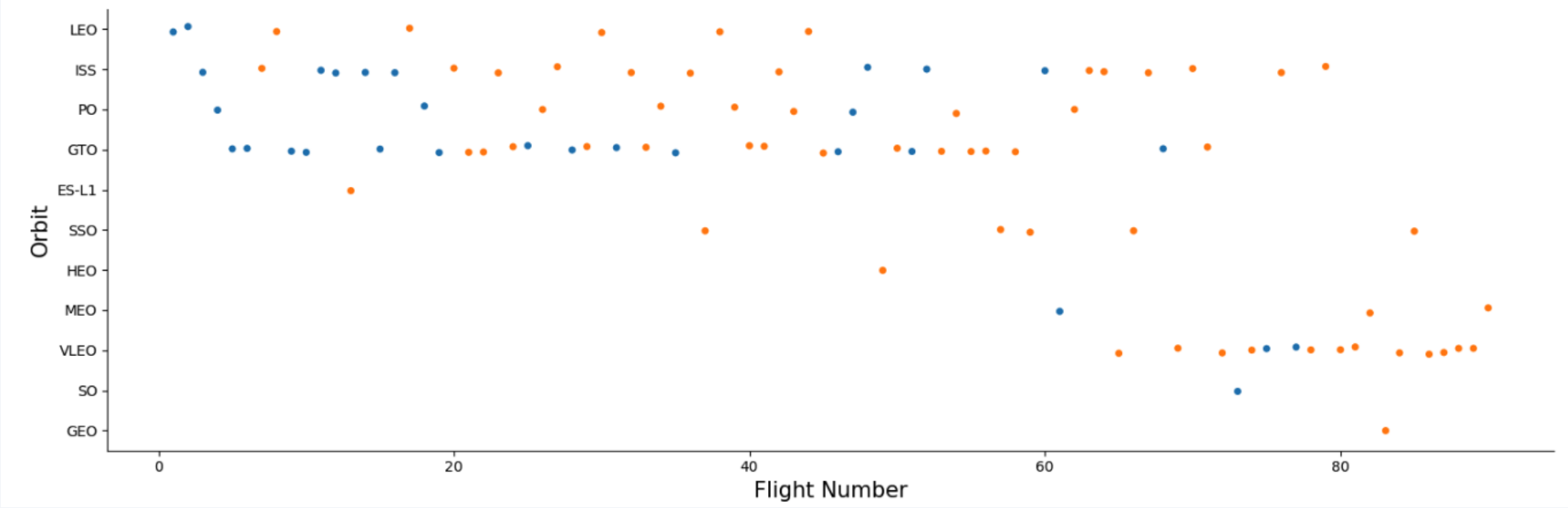
Success Rate vs. Orbit Type

- The SO orbit never success
- In ES-L1, GEO, HEO, SSO have more successes.



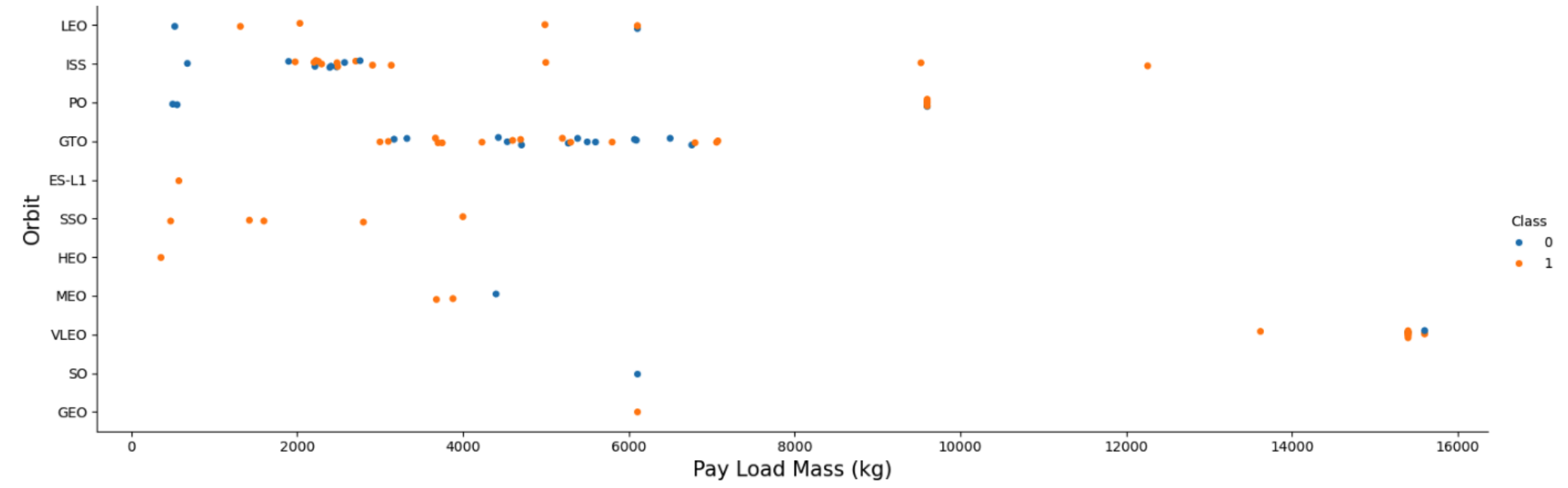
Flight Number vs. Orbit Type

- It's hard to find out there's a relationship between orbit and flight number
- However, in LEO orbit seems higher flight number comes with more successes



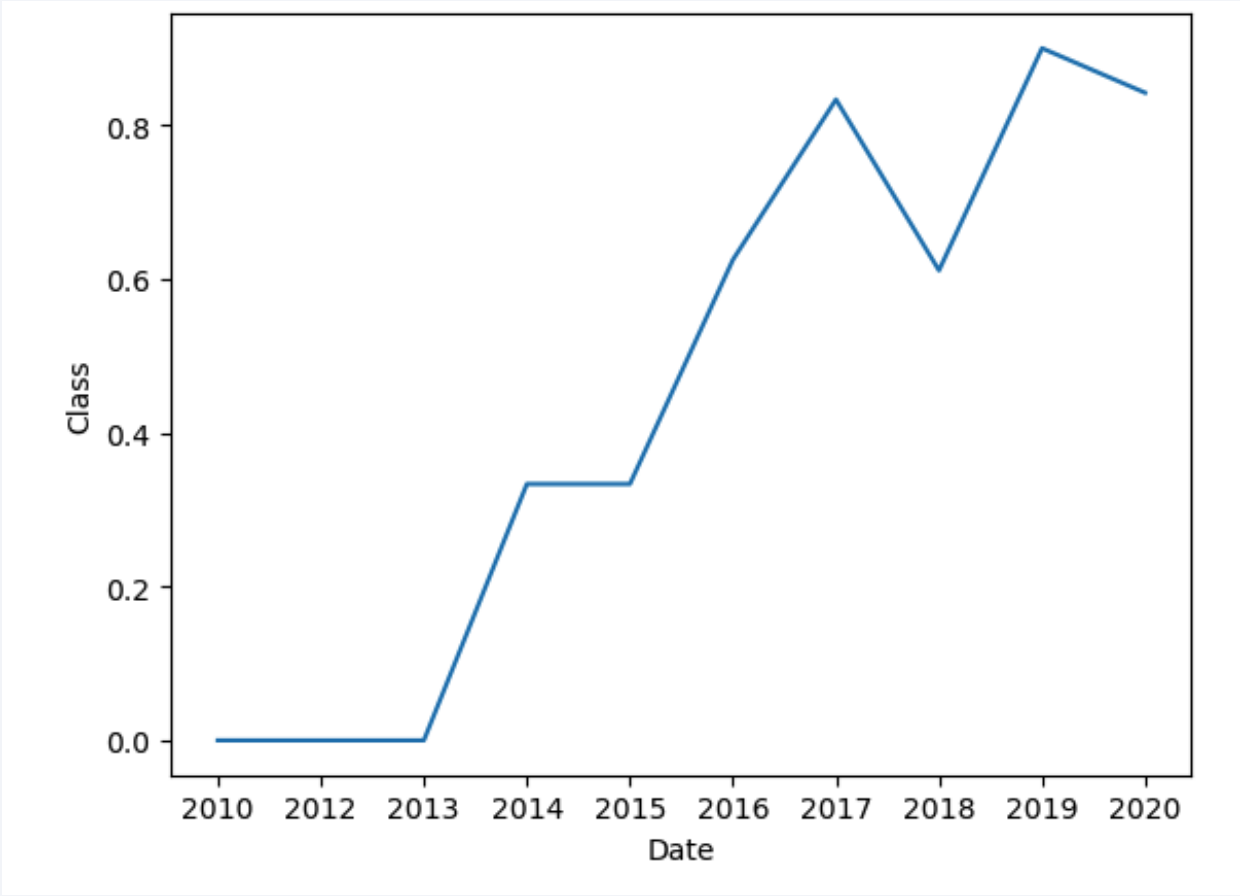
Payload vs. Orbit Type

- For the scatter plot, it shows there seems no relationship between payload and orbit type



Launch Success Yearly Trend

- Through the line chart you can say the successful rate increase when the year growth.
- However, there's some reason makes the successful rate fall in 2018



All Launch Site Names

- As the figure shows there's four launch site for Falcon 9.

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

Launch Site Names Begin with 'CCA'

- You can see the first 5 record shows that most of launch site's orbit are LEO.
- The launch records shows these 5 records are success for mission but landing usually not.

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

- The below figure shows the summary of payload mass is 99980 KG.

sum(PAYLOAD_MASS__KG_)
99980

Average Payload Mass by F9 v1.1

- The below figure shows the average of payload mass is 2534.67 KG for booster version which is F9 v1.1.

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The first successful ground landing is in 2015-12-22.

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The figure shows the successful drone ship landing with payload between 4000 to 6000.
- You could see the F9 FT series booster version is the most common version in this range with successful drone ship landing

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 the figure you can find out there's 101 launches and most of mission are success.
- There four mission outcome type (perhaps there's typo because the success show twice)

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

Boosters Carried Maximum Payload

- There's 12 booster version that carried maximum payload

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 there's two landing outcome are failure.
- Both the failure records were launched from CCAFS LC-40

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

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

- The maximum landing outcome between 2010-06-04 and 2017-03-20 is No attempt with the count 10.
- The minimum landing outcome is Preclude (drone ship) with the count 1.

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

Section 3

Launch Sites Proximities Analysis

Map for Launches

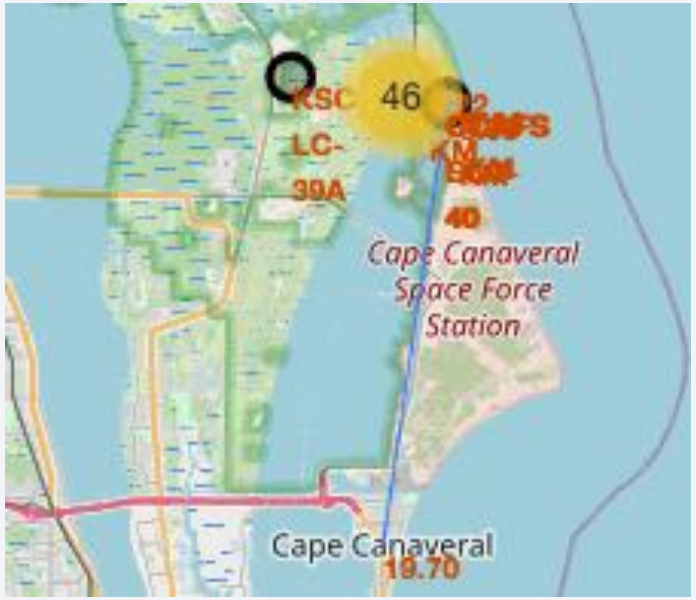
- As the map shows, we can find out the launch sites are all locate at the coast.
- The launch sites located at both east and west coast.

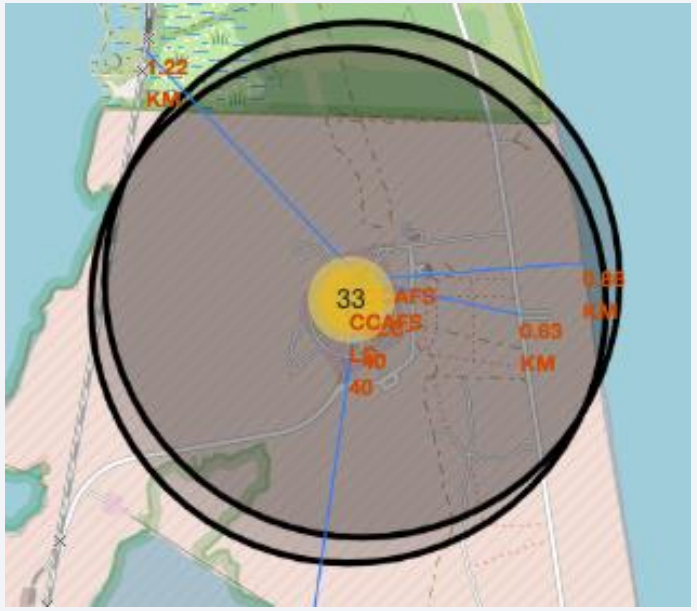
Map for Launches Successful Rate

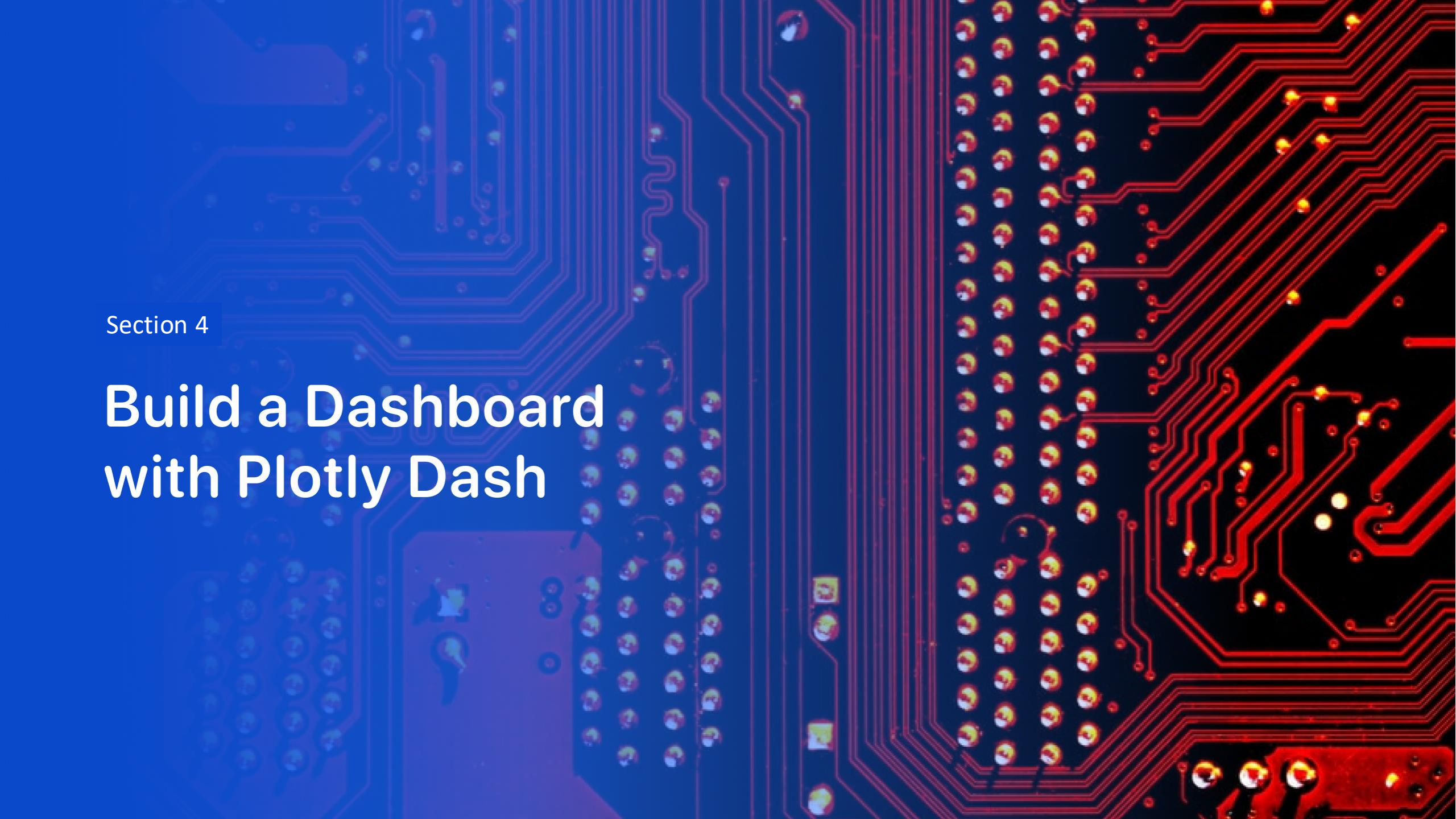
- As the Map can see we have total 56 launches on both east and west sites
- For the most successful sites is the KSC LC-39A which has 10/13 successful rate

Map for the distance between launch site

- As you can see the map shows the site is far from the city side but near coast.
- This might show the best launch suppose take place at coast and should be away from city.

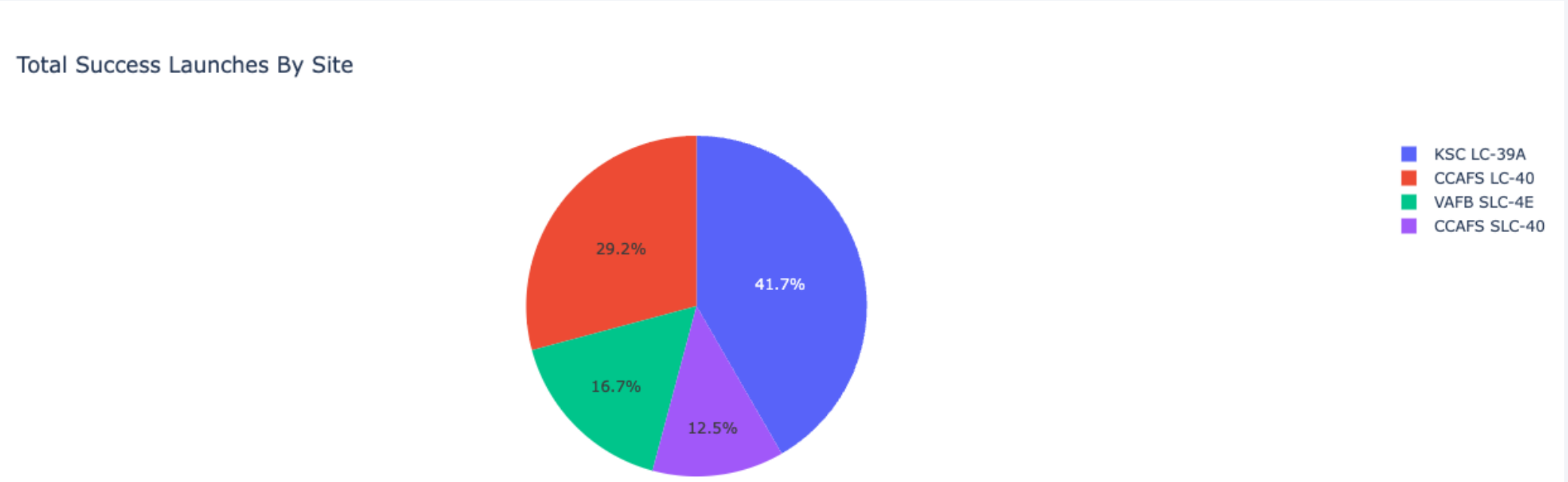





Section 4
Build a Dashboard with Plotly Dash

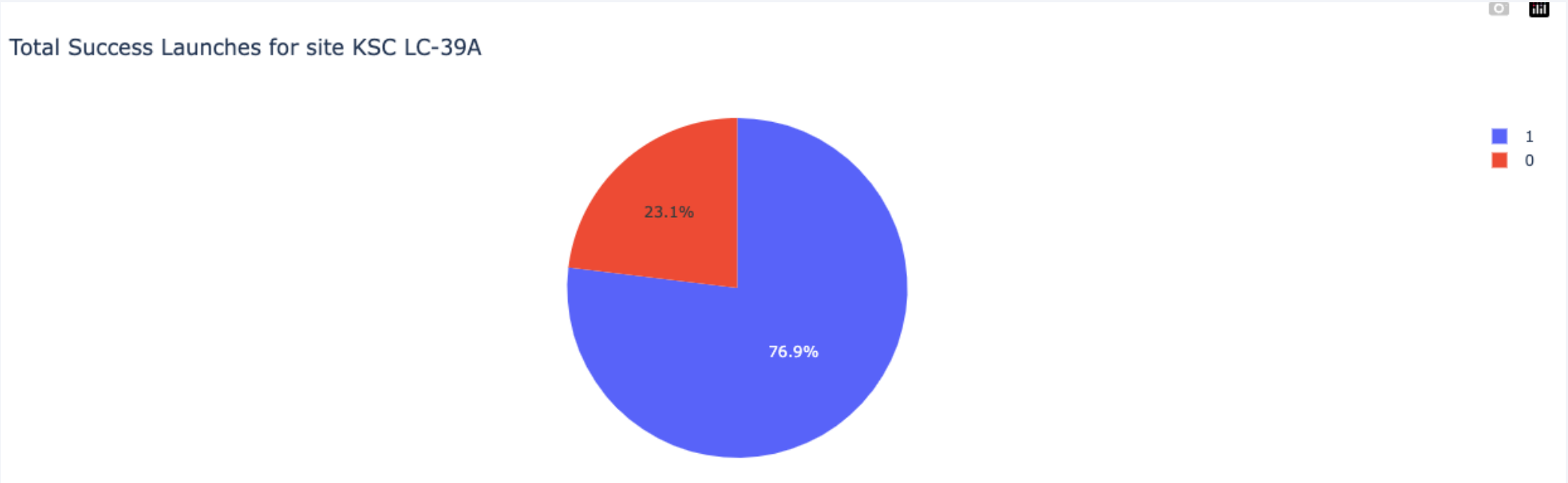
Total Success Launches By Site

- Through the pie chart you can see the KSC LC-39A launch site has the highest successful rate.



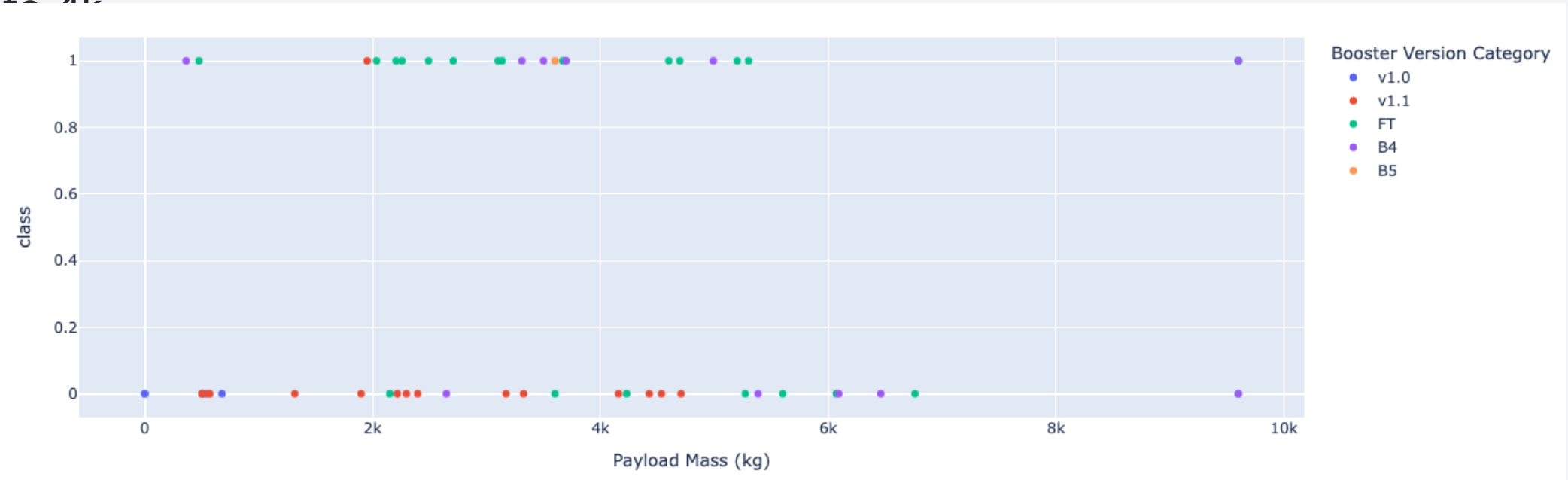
Total Success Launches for site KSC LC-39A

- The KSC LC-39A launch site has 76.9% successful rate and only 23.1% failure.



Scatter plot for all site with Payload Mass and Class

- In the plot you can find out the booster version v.1.1 has the lowest successful rate.
- The booster version FT has the highest successful rate in payload mass between 2k to 4k

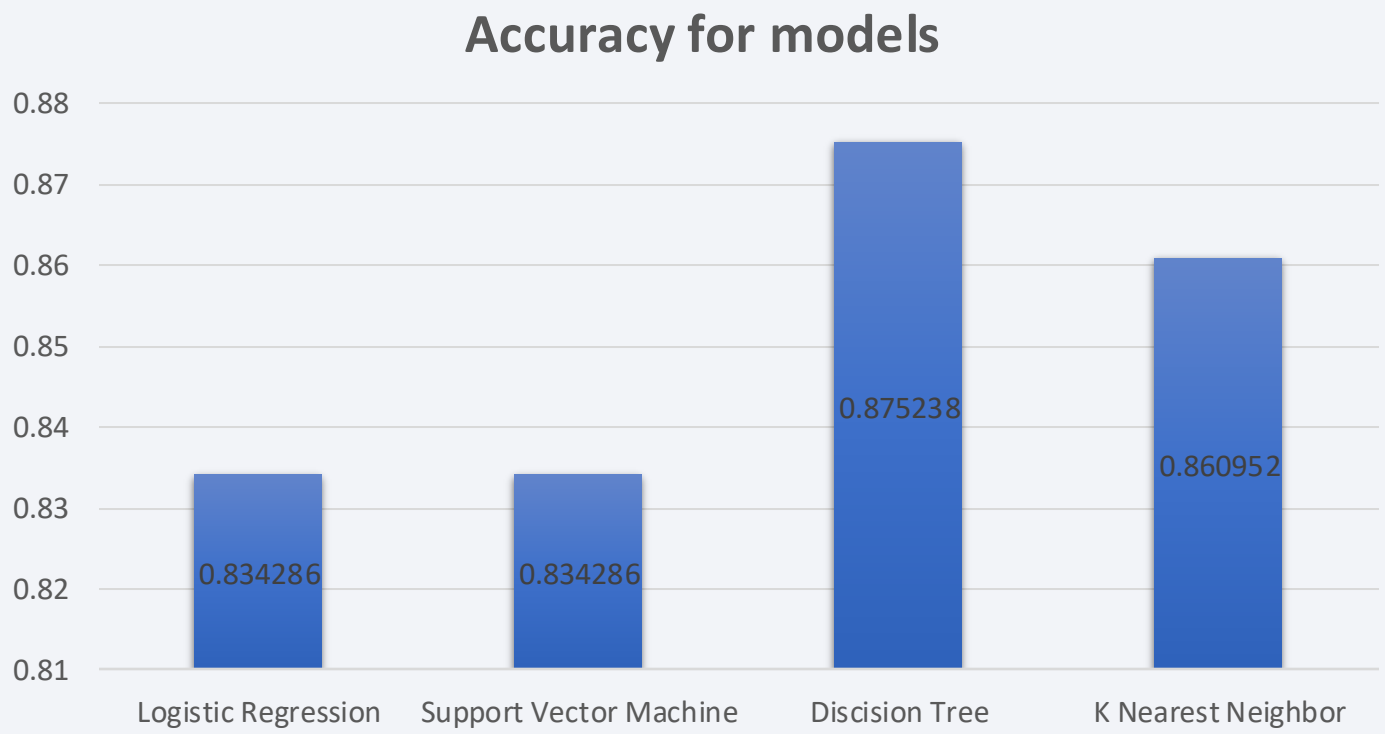


Section 5

Predictive Analysis (Classification)

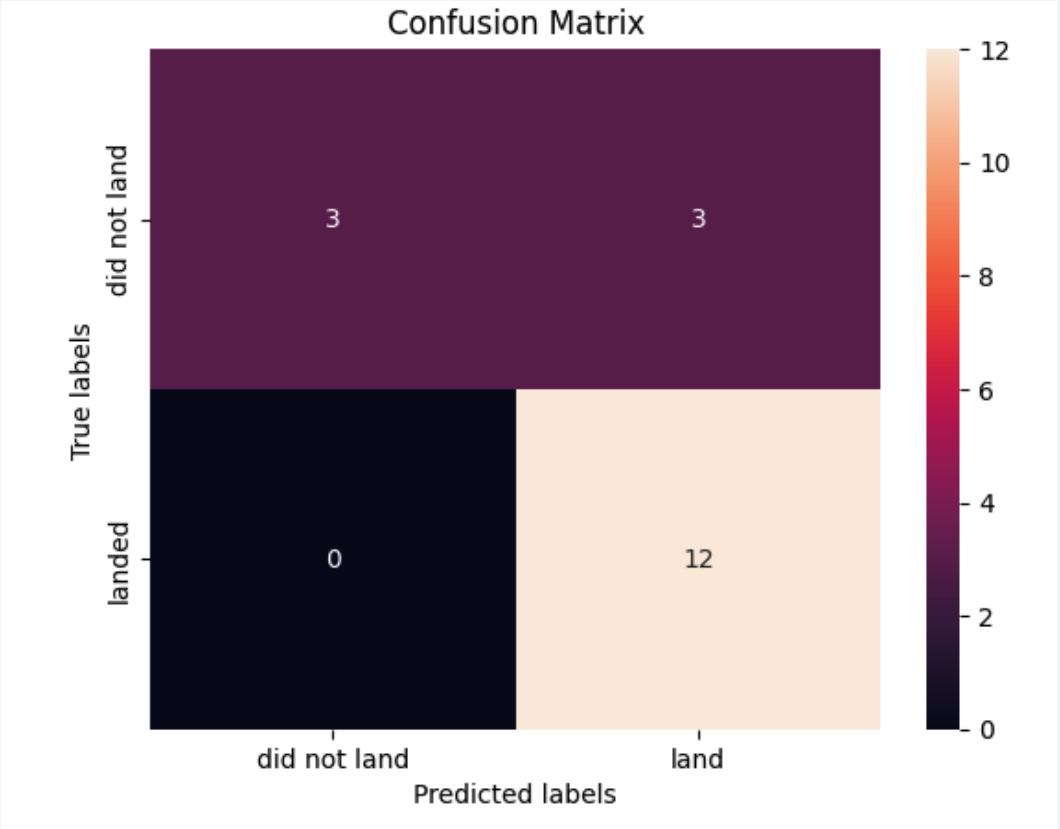
Classification Accuracy

- In the table you can see the highest accuracy model is decision tree



Confusion Matrix

- The confusion matrix shows the models has 3 false negative errors.



Conclusions

- The launch success or not is related to booster version, payload mass, orbit.
- The Decision Tree can do the best prediction.

Appendix

- All the resource could be found in https://github.com/CCChou/HW-for-IBM-DS/tree/main/capstone

Thank you!