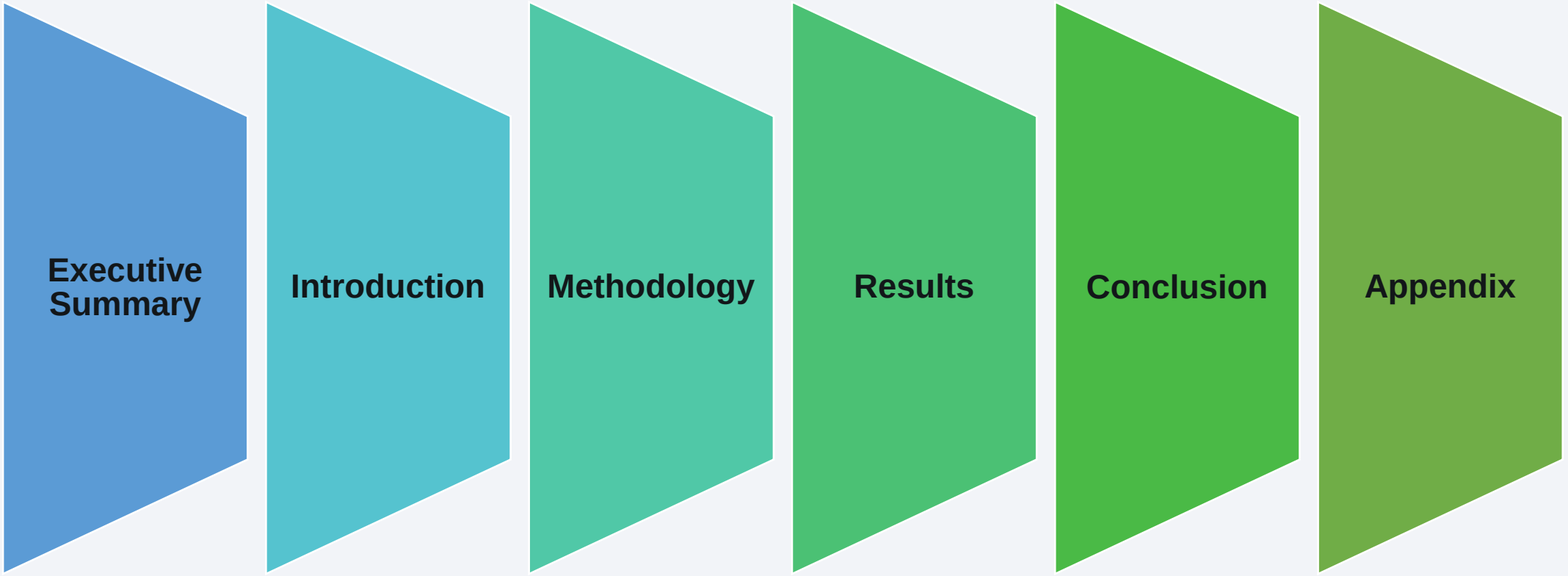
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Komal Patel
October 24, 2021

Outline


Executive Summary
Introduction
Methodology
Results
Conclusion
Appendix

Executive Summary

- Summary of methodologies
- The commercial space age is here, companies are making space travel affordable for everyone. In this race perhaps the most successful is SpaceX. As Space Y would like to compete with SpaceX, I would like to determine the price of each launch. SpaceX's Falcon 9 rocket launches cost is much lower than other providers, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Also would like to determine if SpaceX will reuse the first stage. To determine if the first stage will land successfully, trained a machine learning model and used public information to predict if SpaceX will reuse the first stage.
- Summary of all results
- SpaceX has continued to return various reusable Falcon 9 rocket in both ground and sea landings. The landing success rate observed was 66%. There are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident. The launch success rate may depend on many factors such as payload mass, orbit type, and etc. It may also depend on the location. Overall, the success rate since 2013 kept increasing till 2020. Also, different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Introduction

- Project background and context
- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Problems you want to find answers
- If we can determine if the first stage will land, we can determine the cost of a launch. The launch success rate may depend on many mission parameters like payload mass, orbit type, and etc. This information can be used if SapceY wants to bid against SpaceX for a rocket launch.

Section 1

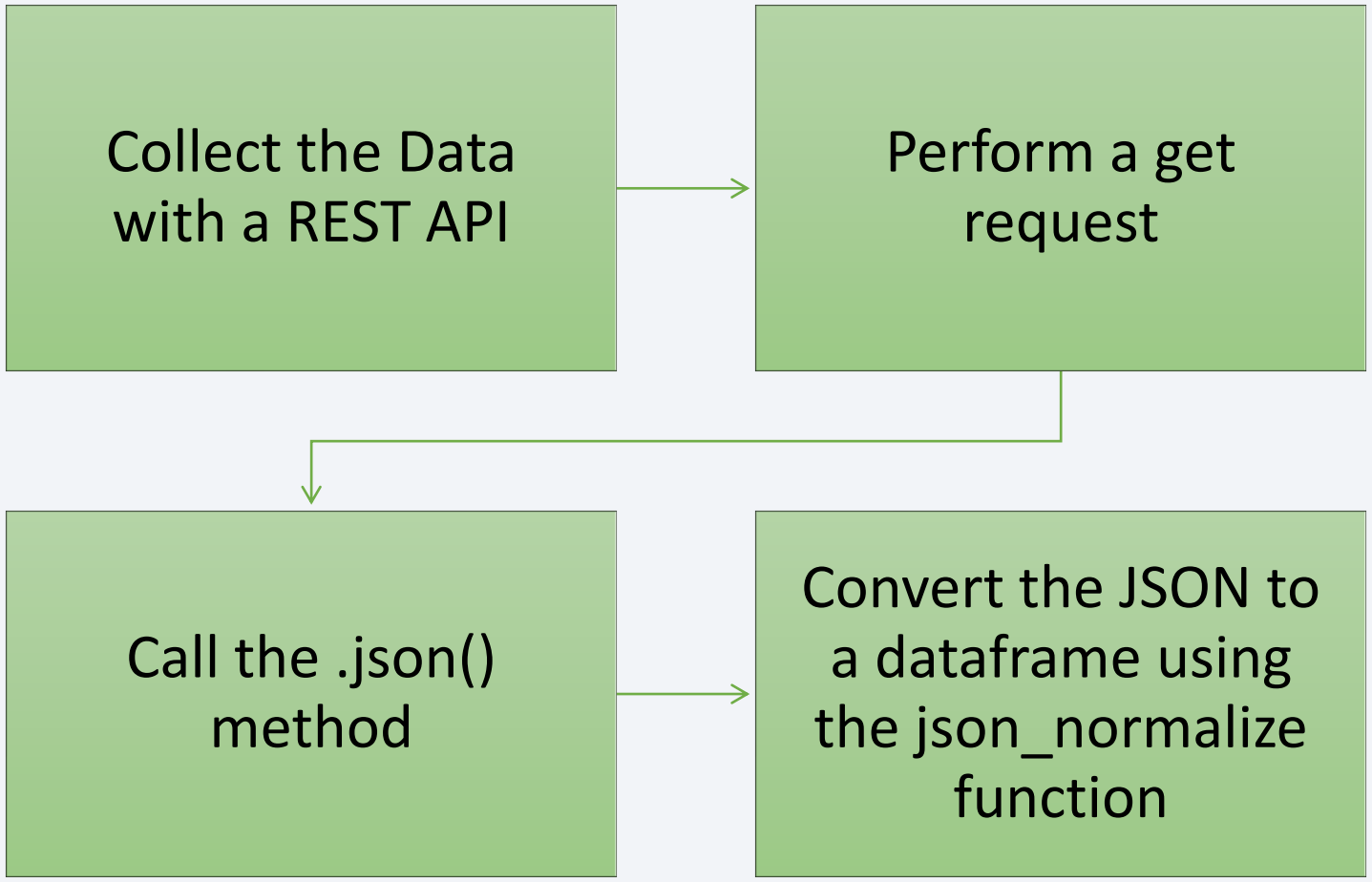
Methodology

Methodology

Executive Summary

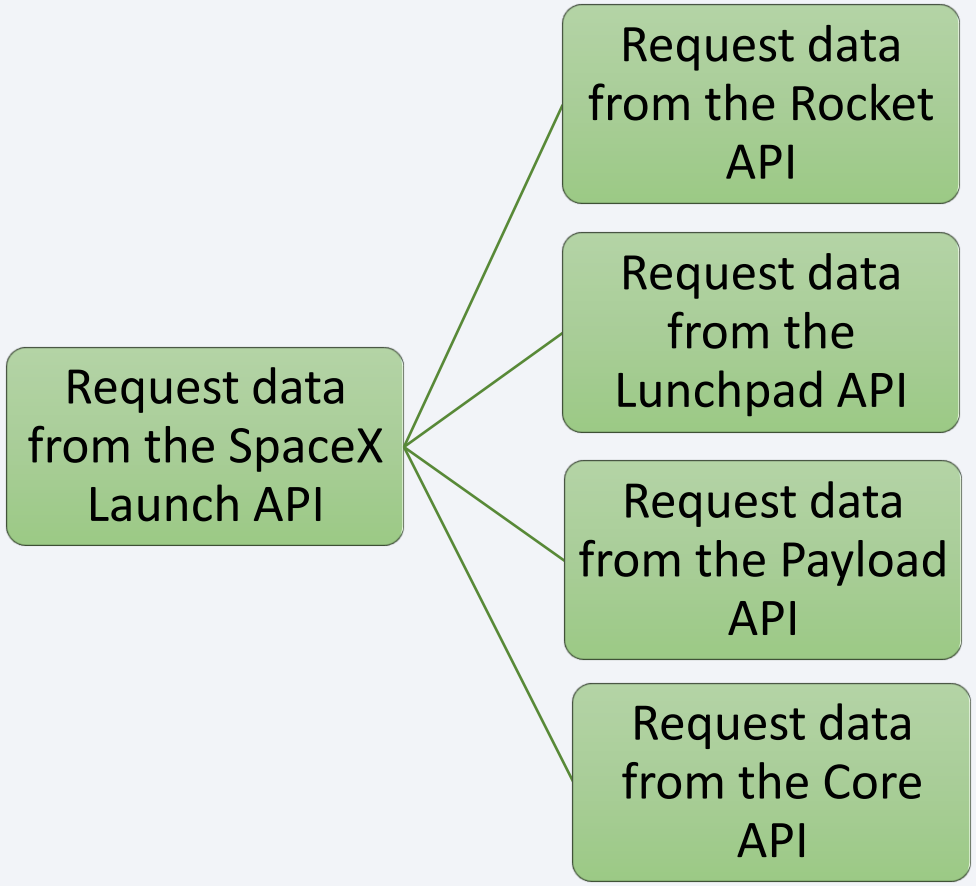
- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection



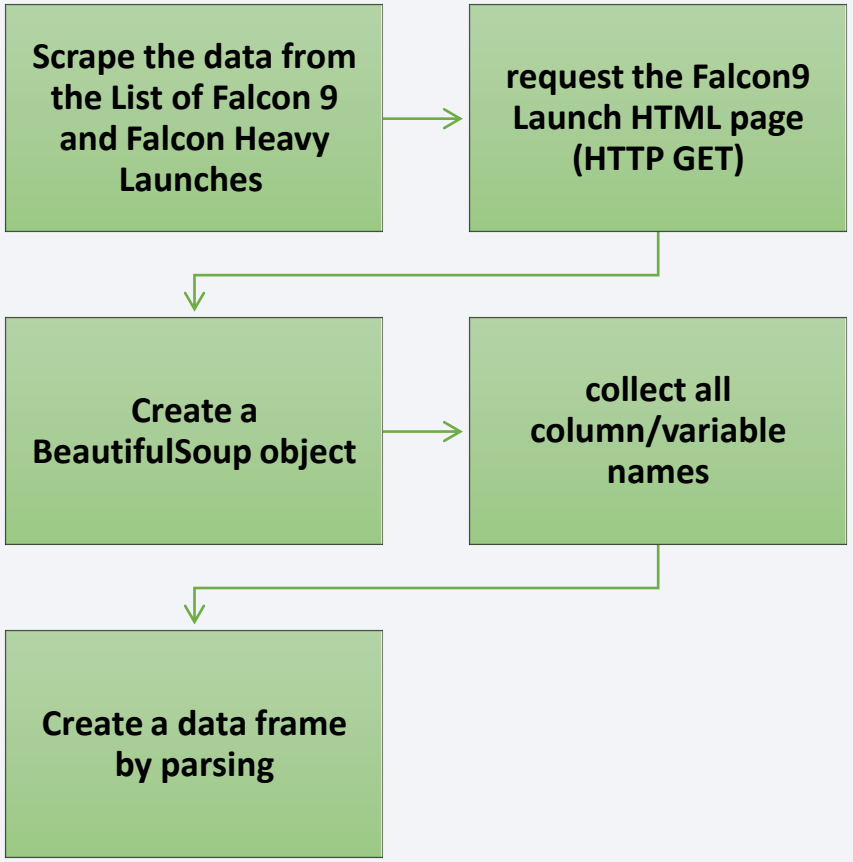
Data Collection – SpaceX API

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN1%20Data%20Collection%20REST%20API.ipynb



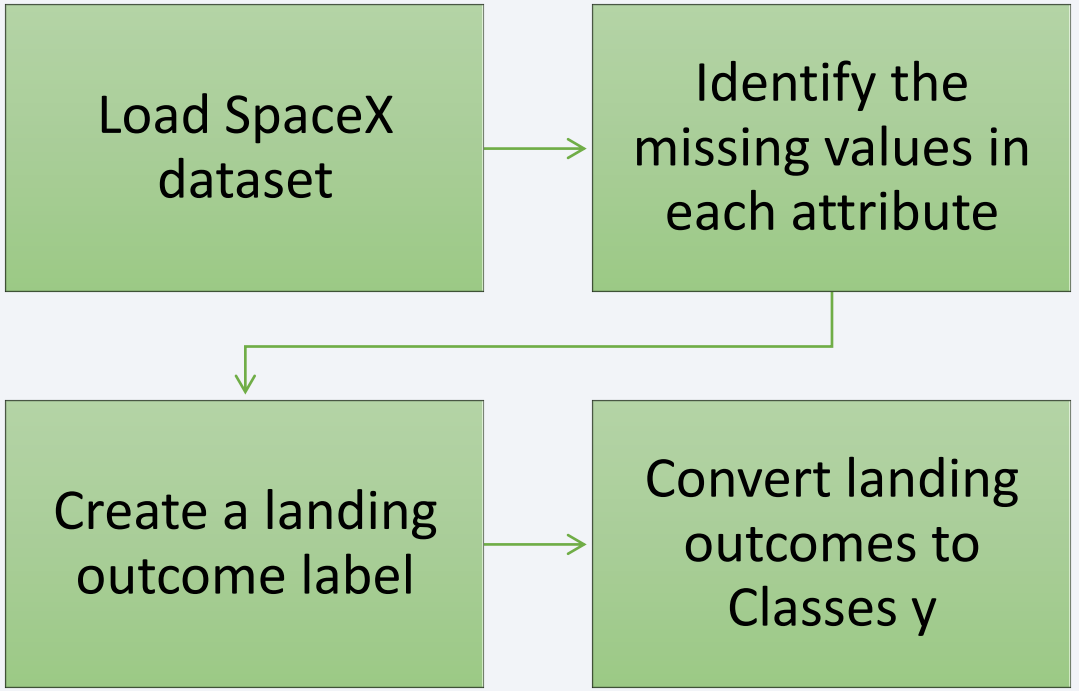
Data Collection - Scraping

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN2%20Data%20Collection%20Web%20Scraping.ipynb



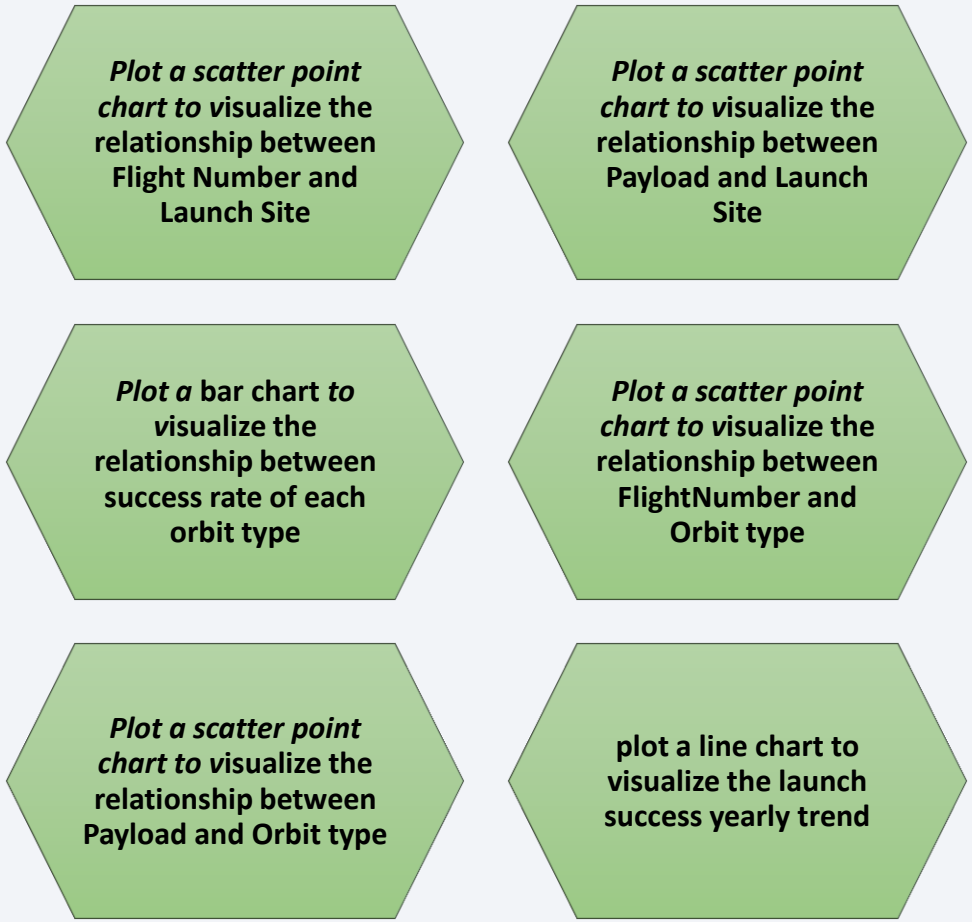
Data Wrangling

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN3%20The%20EDA.ipynb



EDA with Data Visualization

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN5%20The%20EDA%20with%20Visualization.ipynb

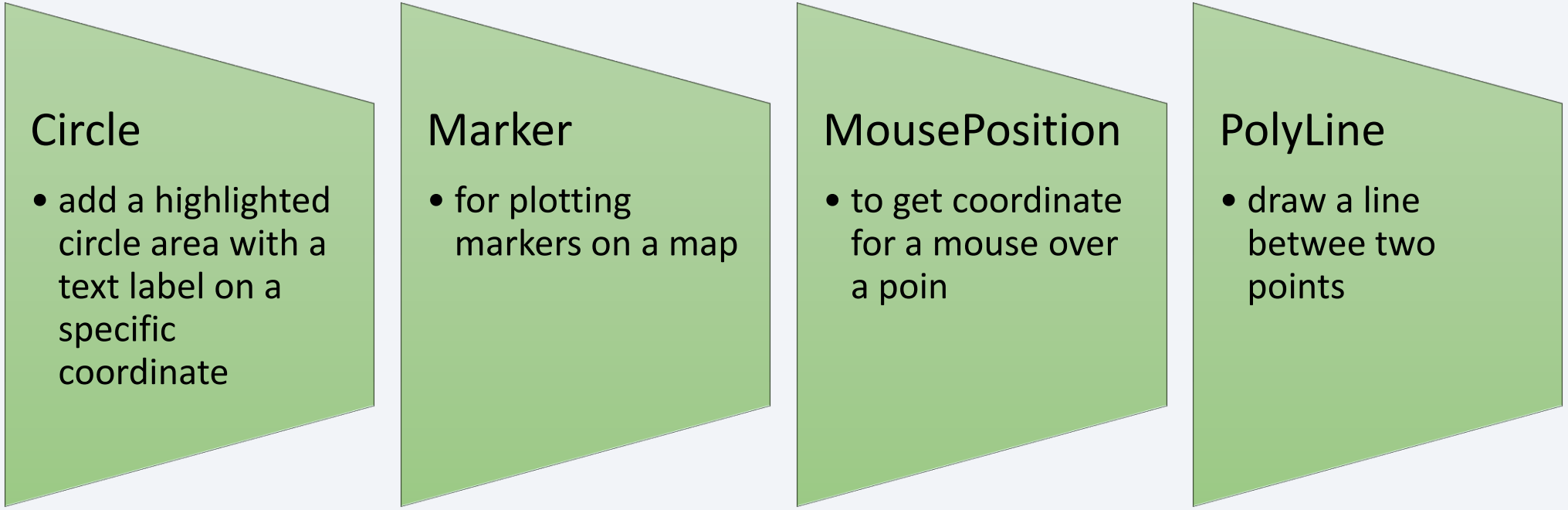


EDA with SQL

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN4%20EDA%20with%20SQL.ipynb

- **load the SQL extension; establish a connection with the database**
- ***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 successful 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.***
- ***List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015***
- ***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***

Build an Interactive Map with Folium

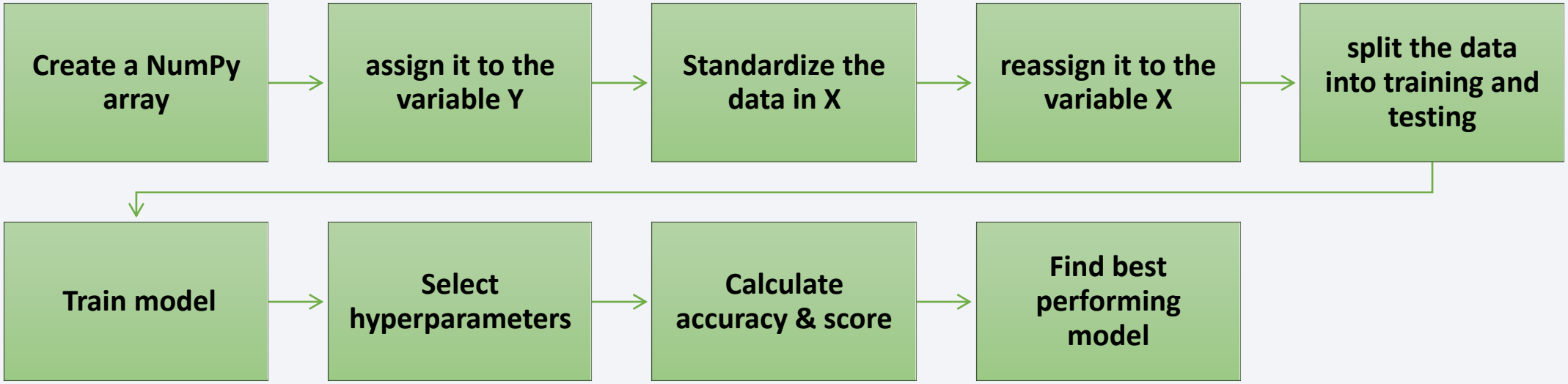


- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN6%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN7%20Interactive%20Dashboard%20with%20Ploty%20Dash.py

Predictive Analysis (Classification)

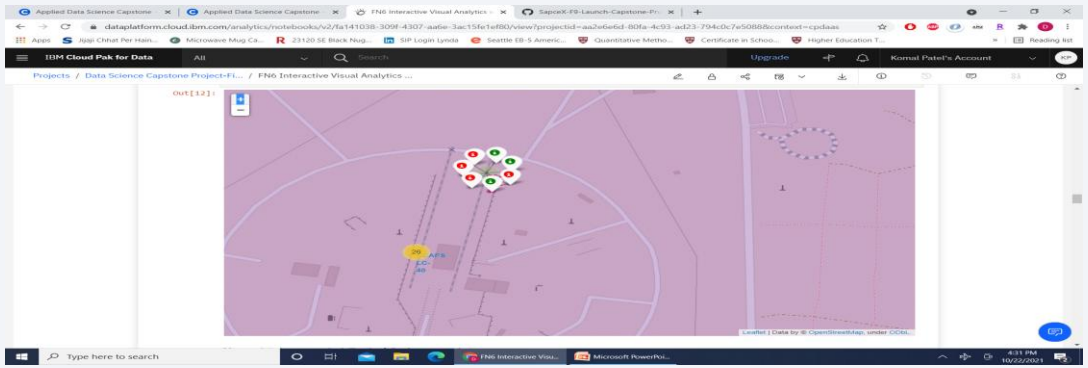


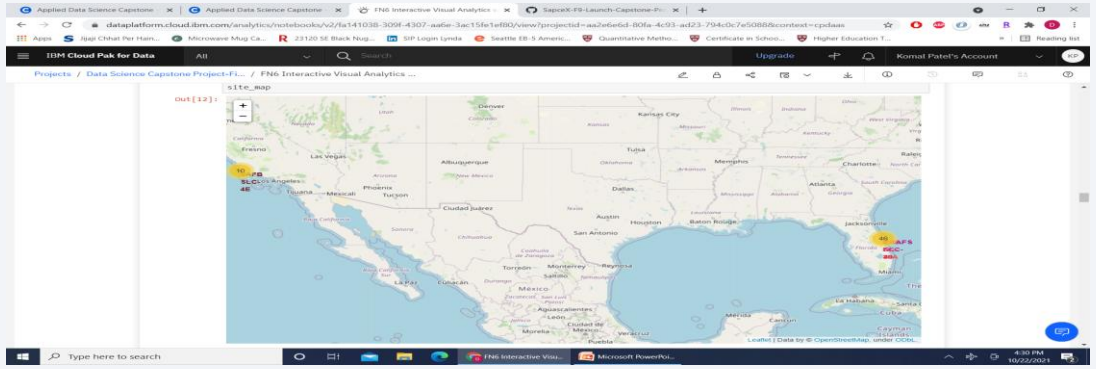
- https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project/blob/master/FN8%20the%20Machine%20Learning%20Prediction.ipynb

Results

Exploratory data analysis

- the success rate is 66%



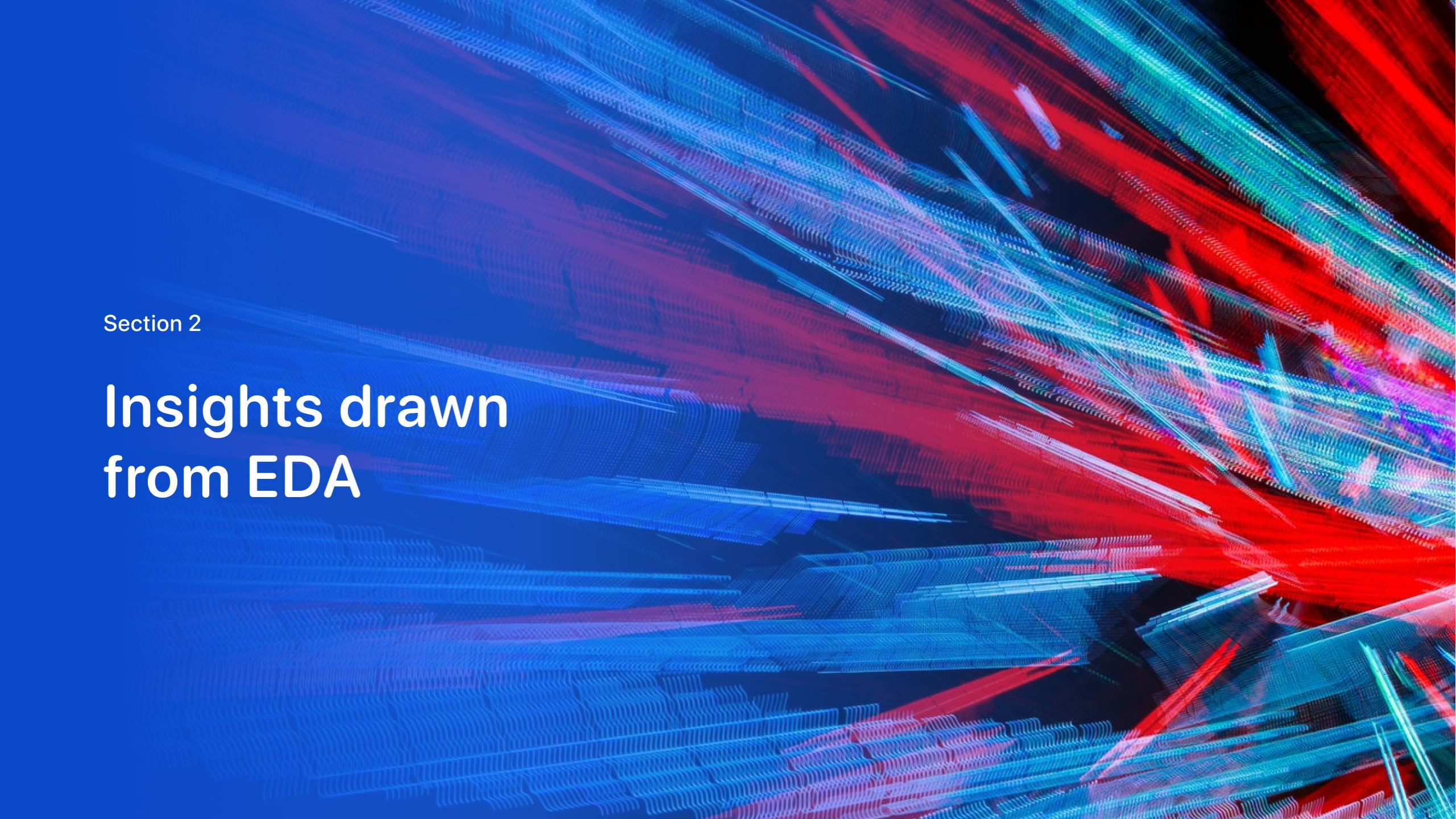


logistic regression, SVM, Decision Tree, KNN models

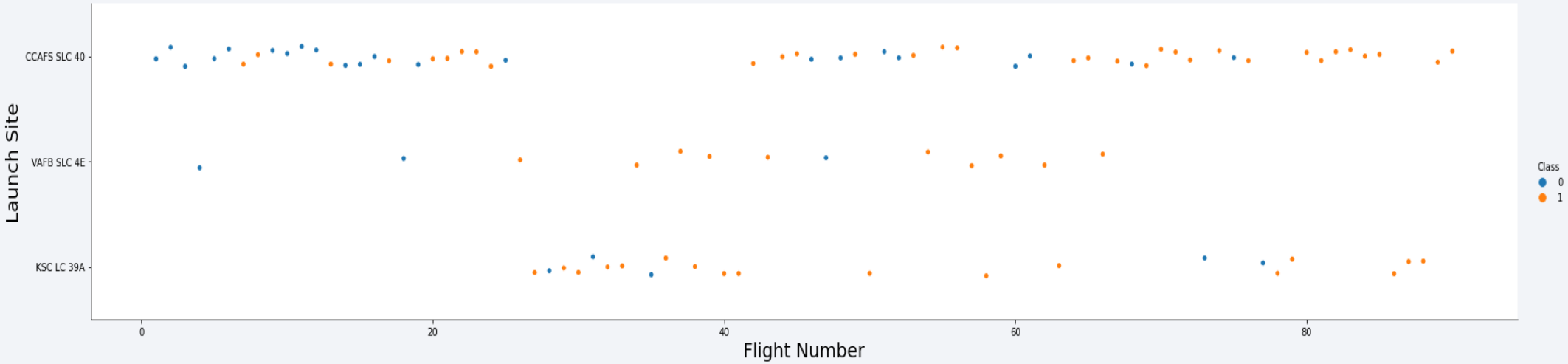
- accuracy score was 83% for all
- In Confusion Matrix major problem is false positives results

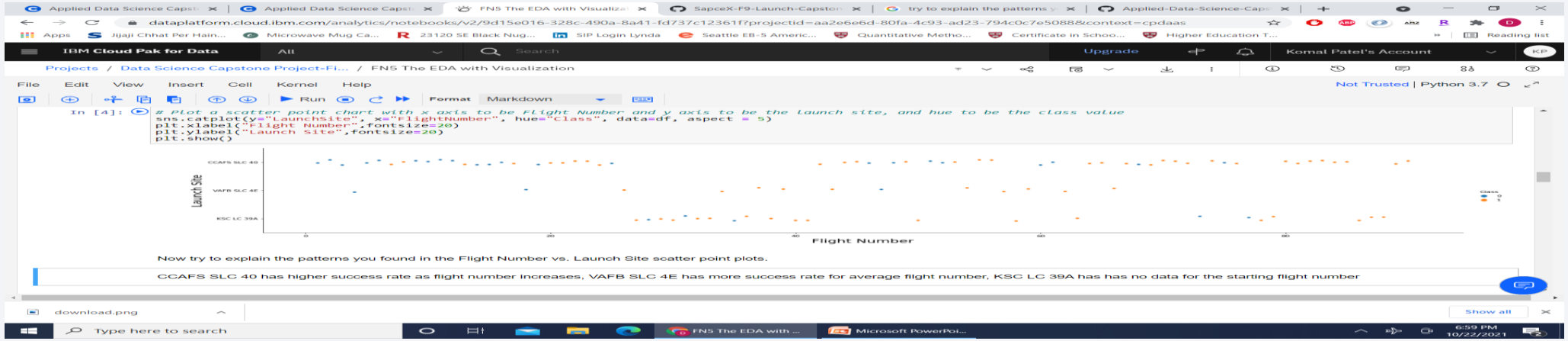
Section 2

Insights drawn from EDA



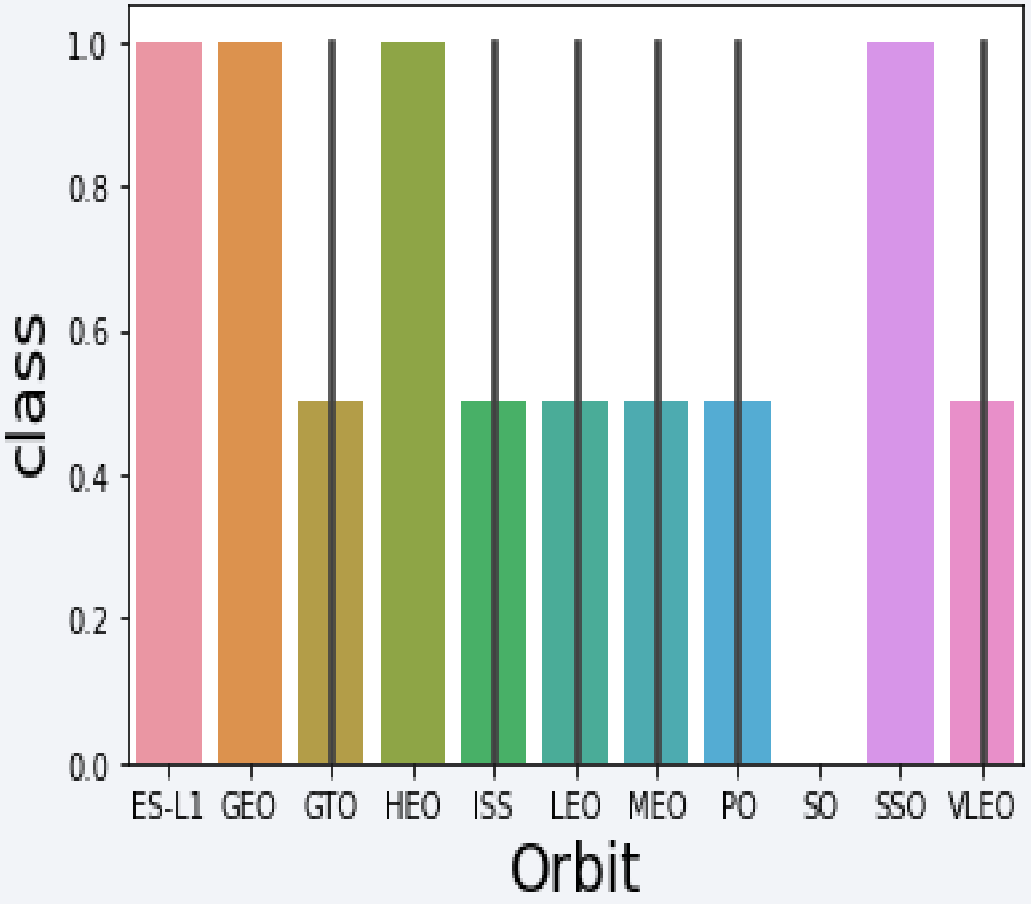
Flight Number vs. Launch Site

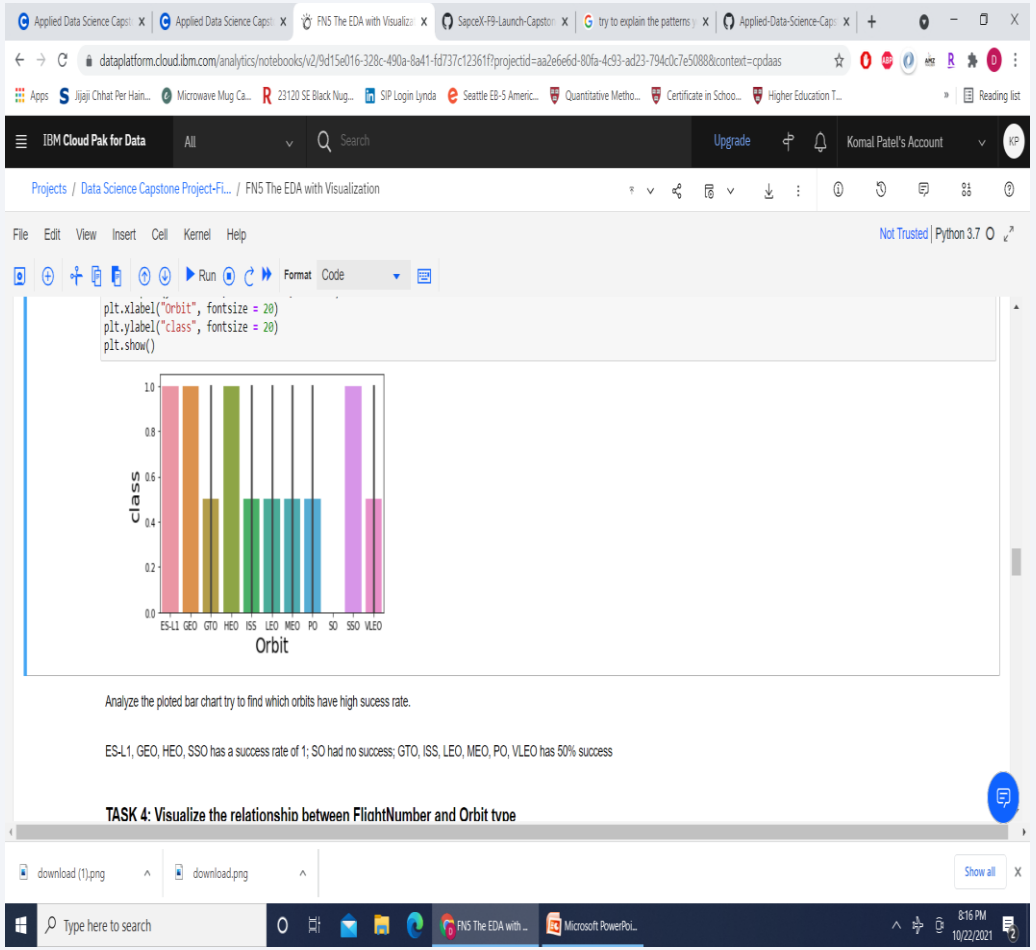




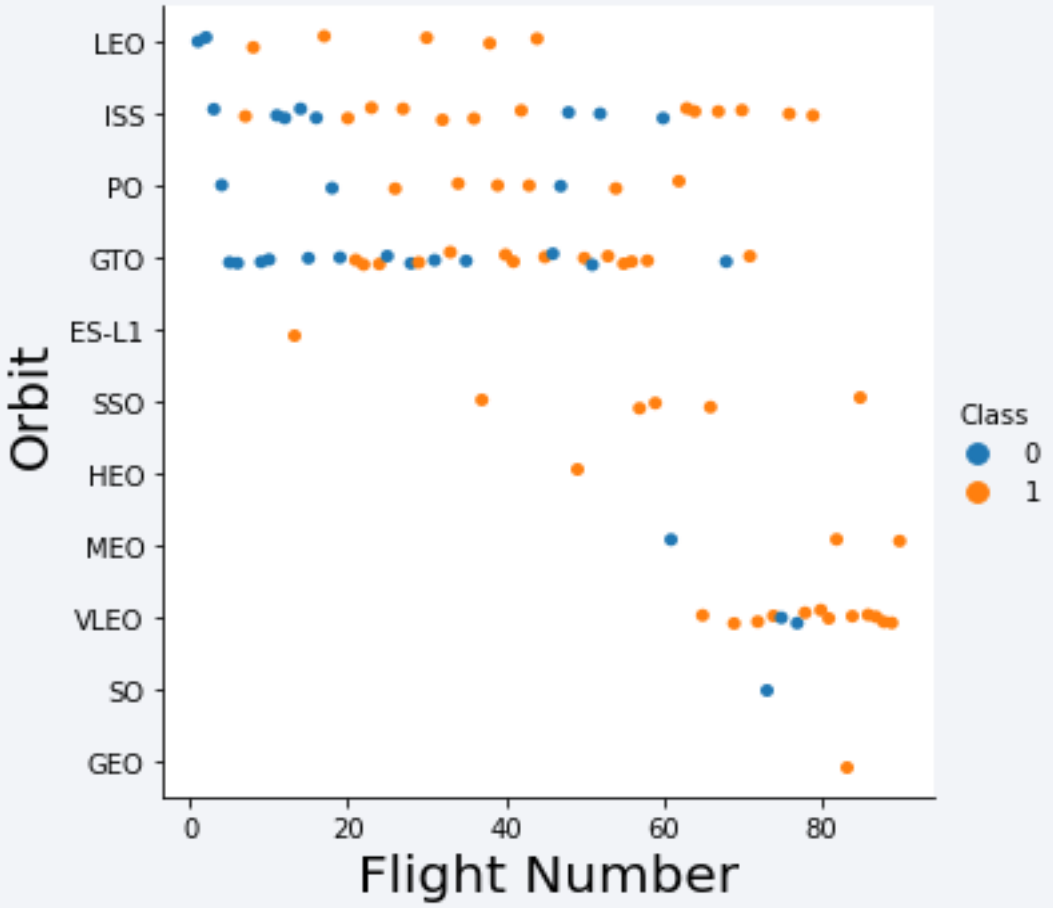
Payload vs. Launch Site

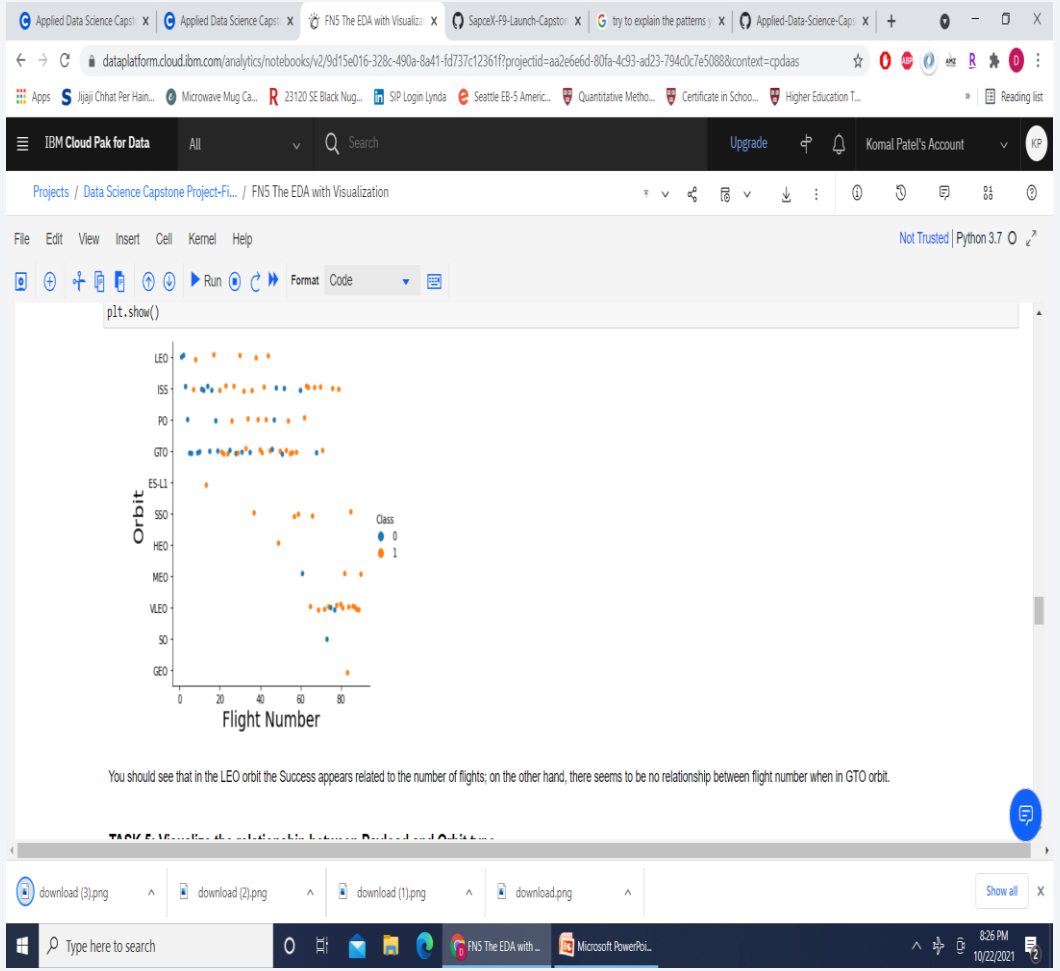
Success Rate vs. Orbit Type



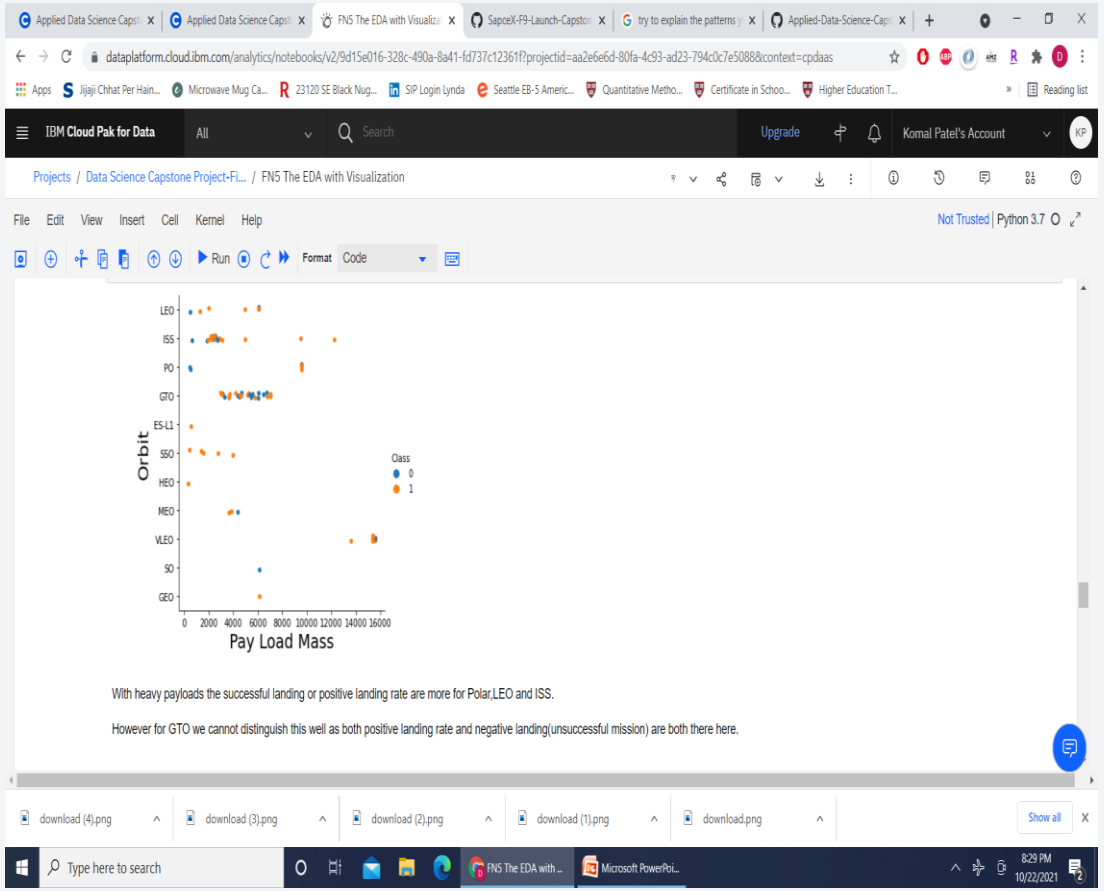


Flight Number vs. Orbit Type

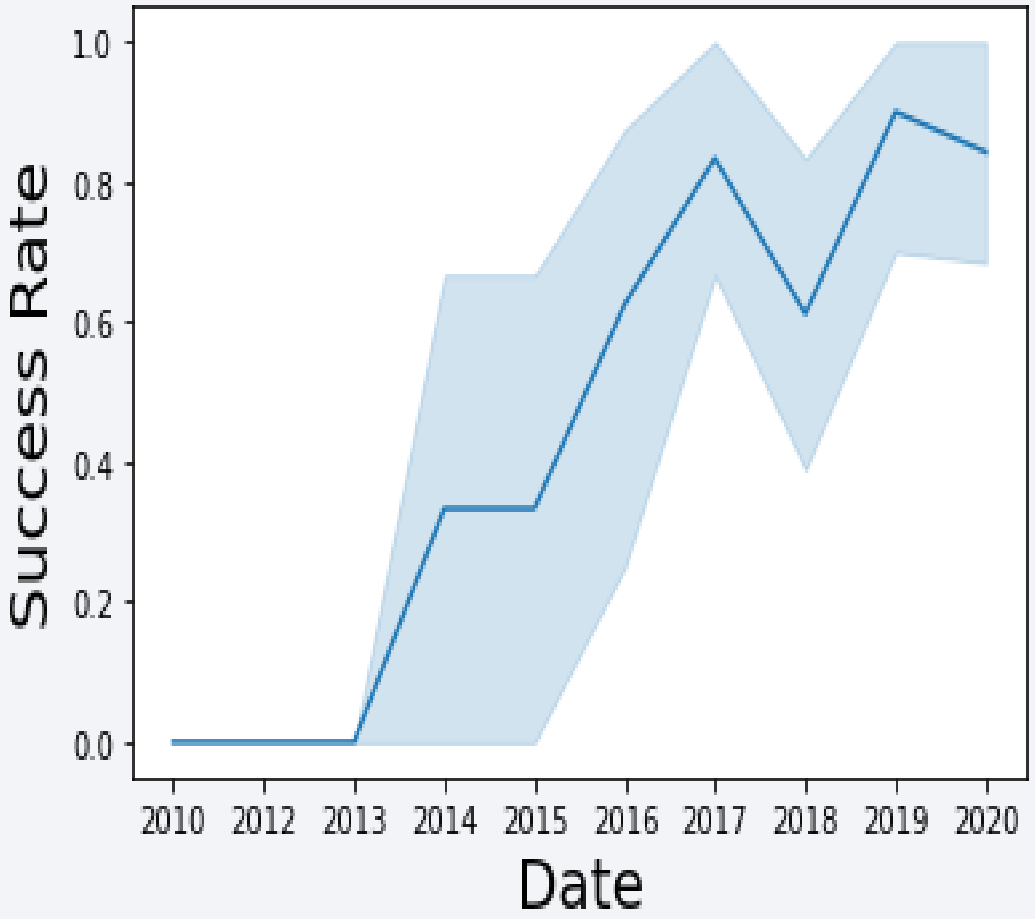


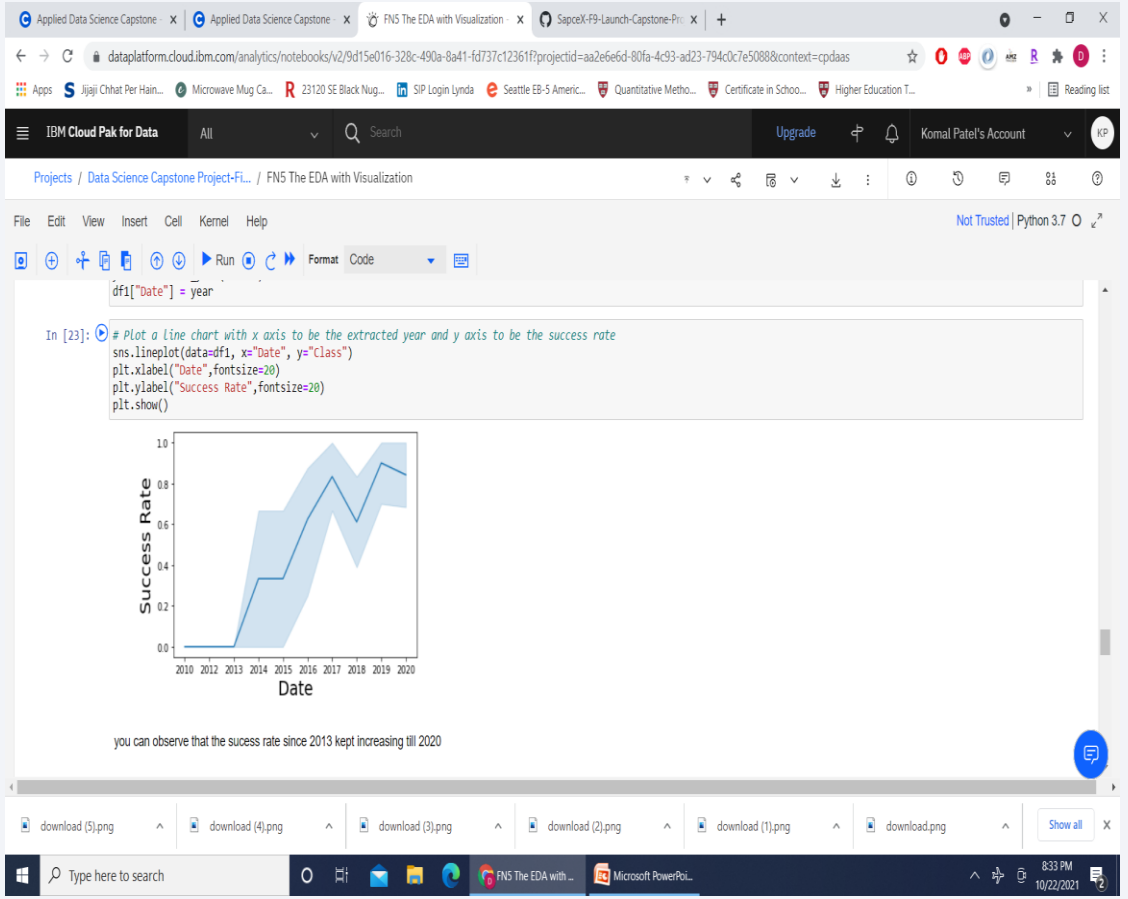


Payload vs. Orbit Type



Launch Success Yearly Trend





All Launch Site Names

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Used the distinct statement to find unique launch sites names in the space mission dataset

Launch Site Names Begin with 'CCA'

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Used the where and like statement to find launch sites begin with the string 'CCA'

Total Payload Mass

1
45596

Used the sum and where statement to calculate the total payload carried by boosters from NASA

Average Payload Mass by F9 v1.1

1
2928

Used the avg and where statement to calculate the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

1
2015-12-22

Used the min and where statement to find the dates of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Used the ‘and’ and where statement to find the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

1
100

Used count to Calculate the total number of successful and failure mission outcomes

Boosters 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

Used a subquery to List the names of the booster which have carried the maximum payload mass

2015 Launch Records

Booster_Version	Launch_Site	Date
F9 v1.1 B1012	CCAFS LC-40	10-01-2015
F9 v1.1 B1015	CCAFS LC-40	14-04-2015
F9 v1.1 B1018	CCAFS LC-40	28-06-2015

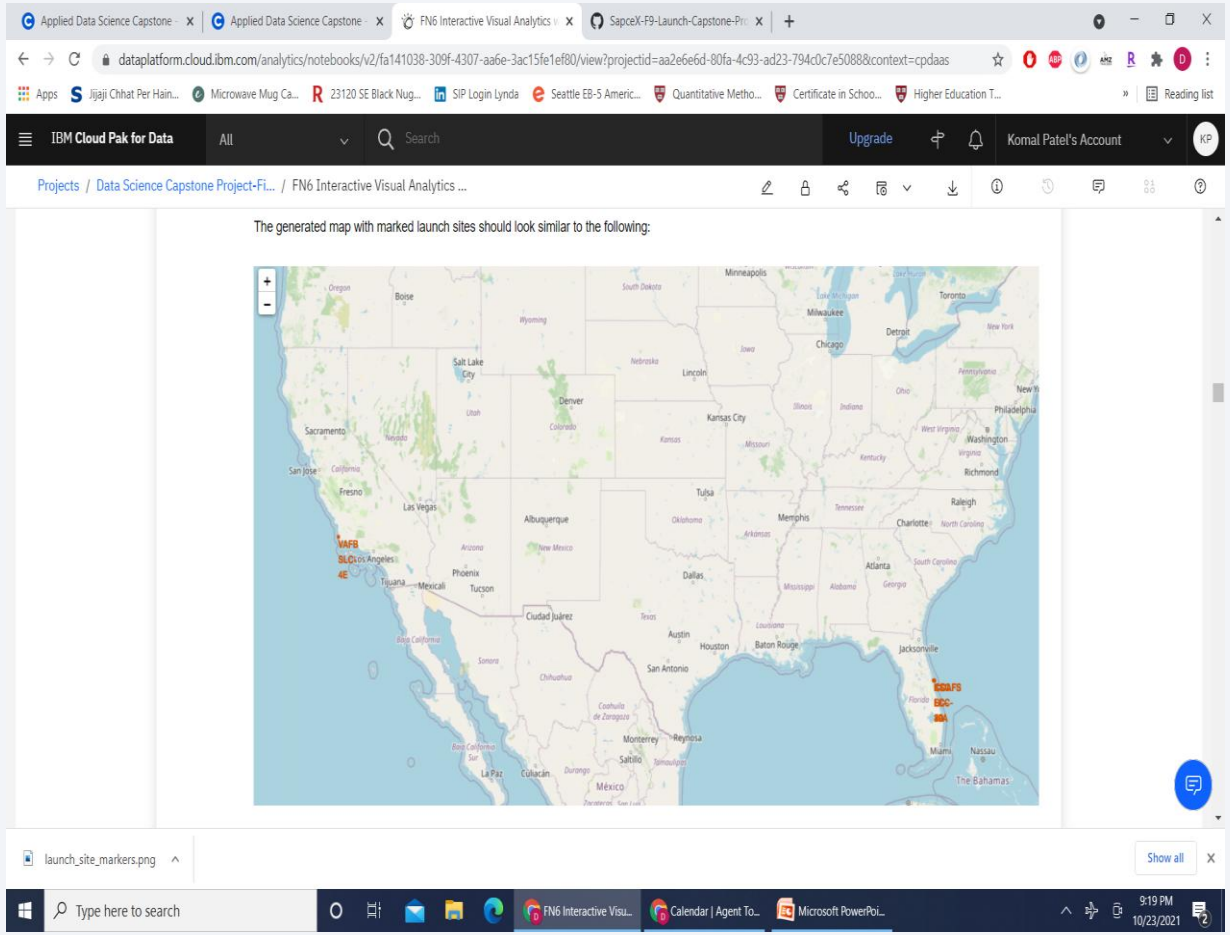
Used the like, ‘and’ and where statement to list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

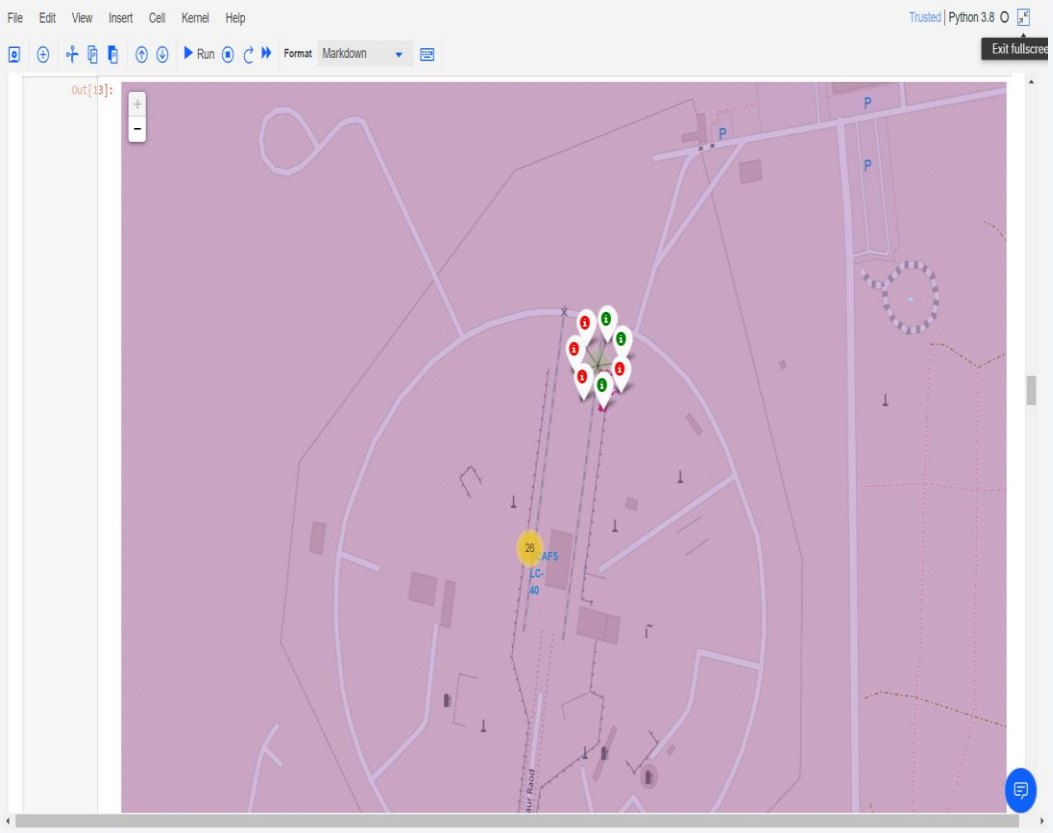
Used the like, ‘and’, order by and where statement to rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Launch Sites' Location Markers



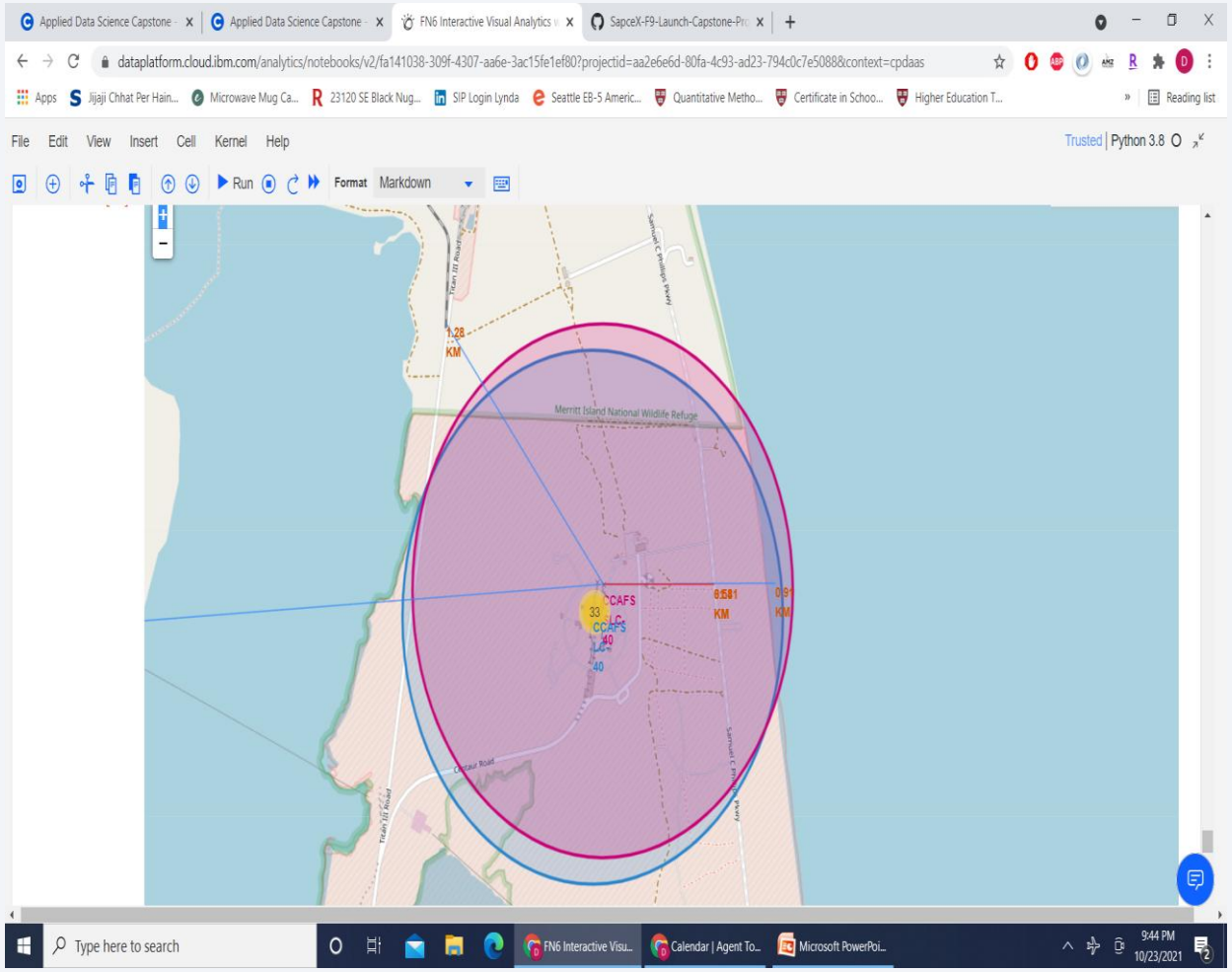
All launch sites are in very close proximity to the coast

The Color-labeled Launch Outcomes



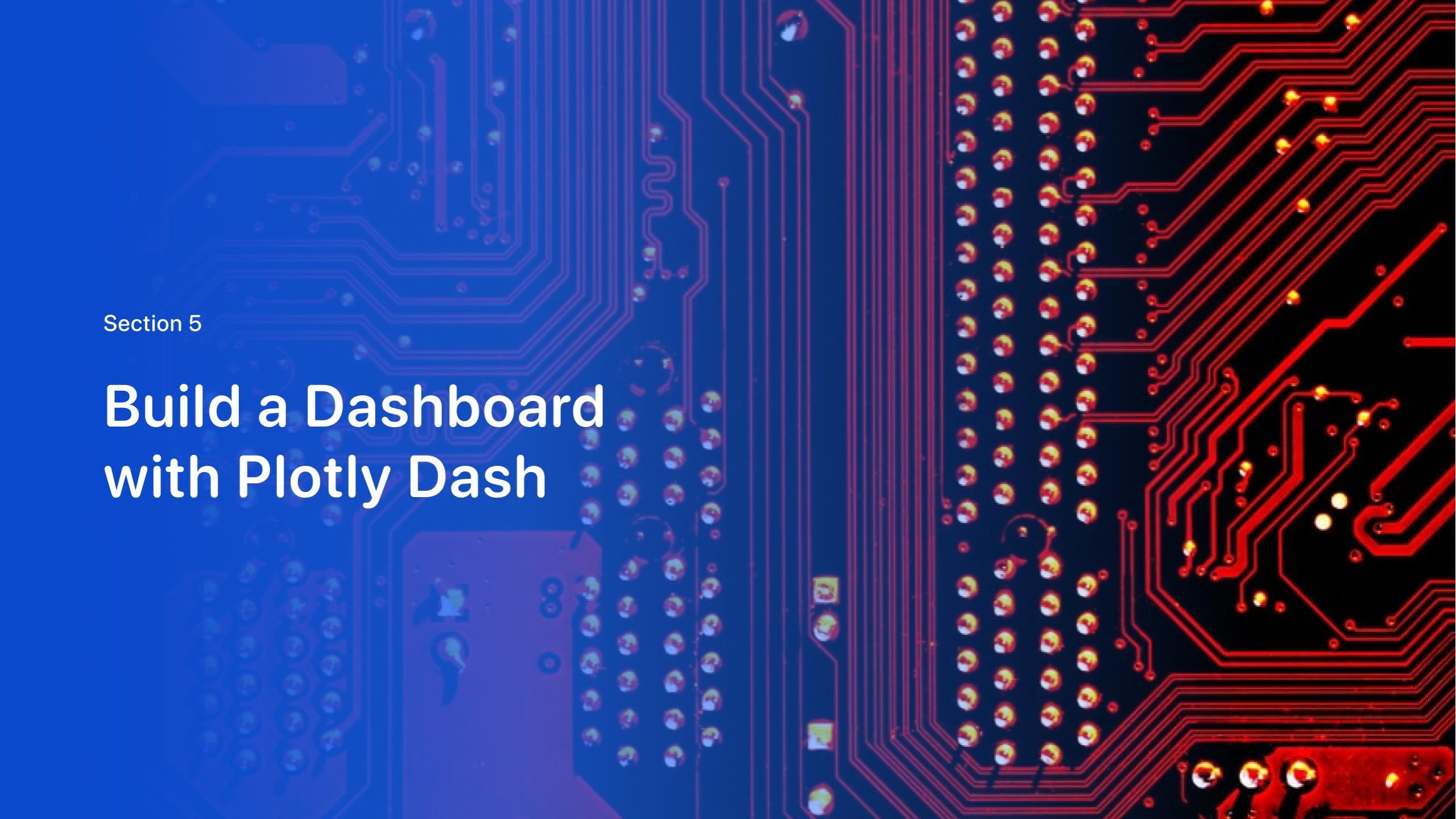
KSC LC-39A launch site have relatively high success rate

Launch Site to Its Proximities

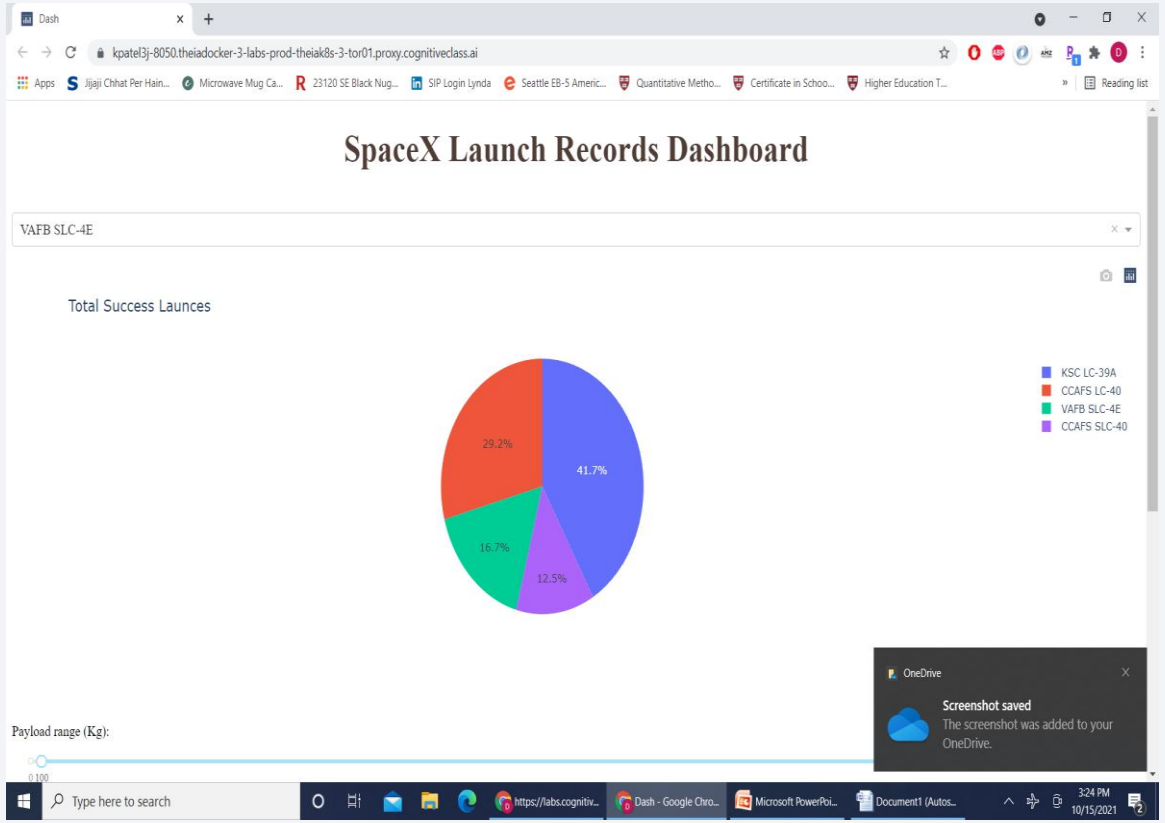


Launch sites in close proximity to railways, highways and coastline.

Launch sites are away from cities

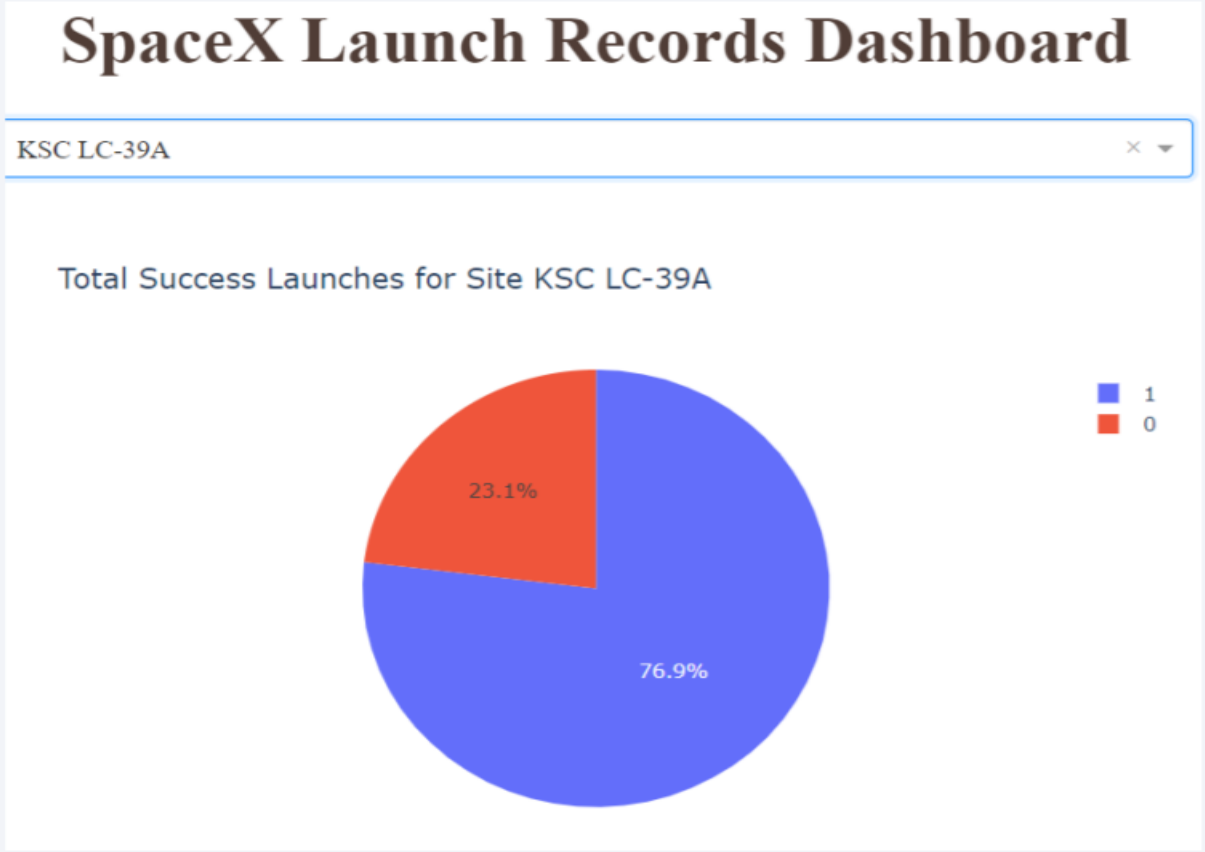

Section 5
Build a Dashboard with Plotly Dash

Launch Success Count



KSC LC-39A, CCAFS LC-40 had higher launch success count

The Launch Site with Highest Success Ratio



KSC LC-39A had highest launch success ratio of 76.9%

Payload vs. Launch Outcome

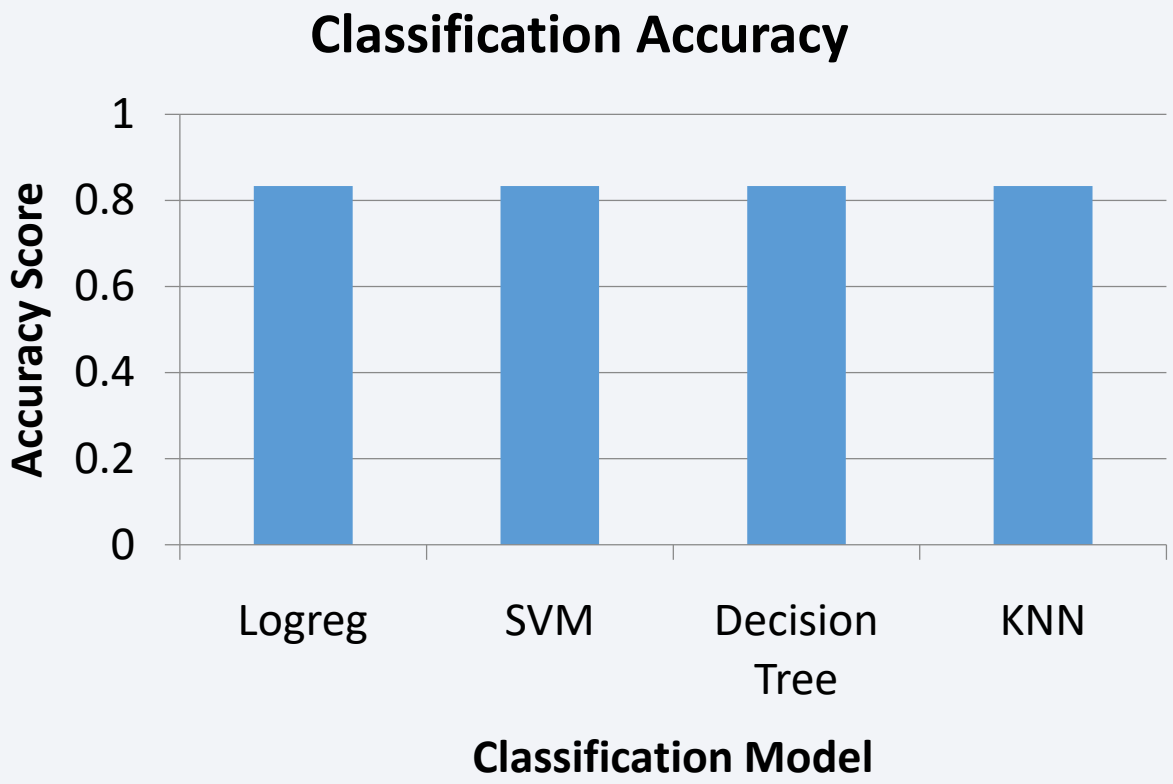
Booster version FT has more success rate between 2K to 6K payloadmass

Booster V1.1 has very less rate for most payloadmass

Section 6

Predictive Analysis (Classification)

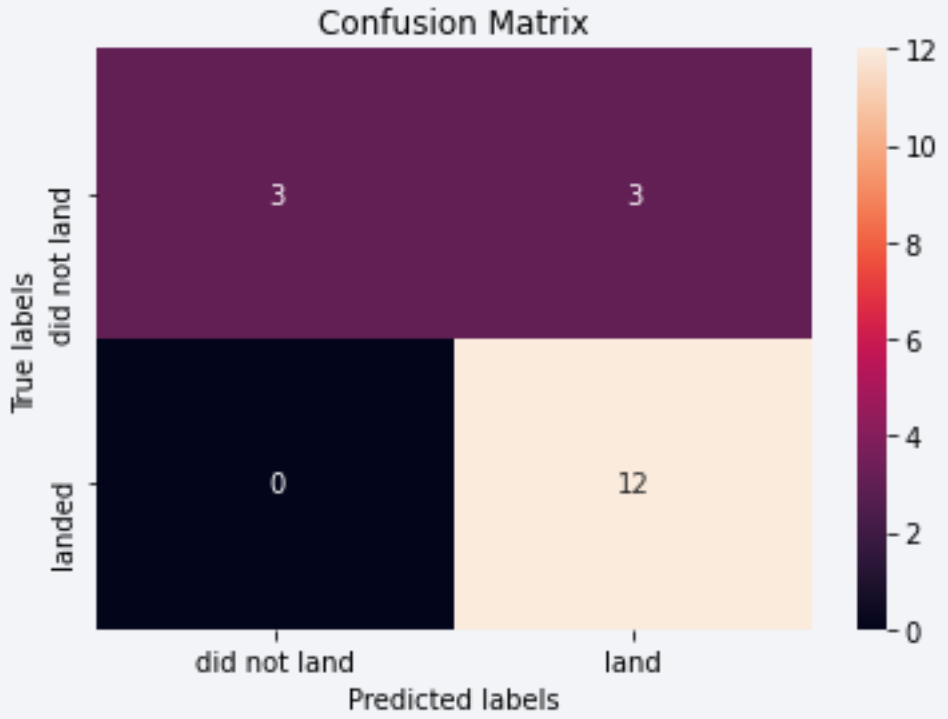
Classification Accuracy



All model has the same classification accuracy score

Confusion Matrix

All of the performing model had same confusion matrix



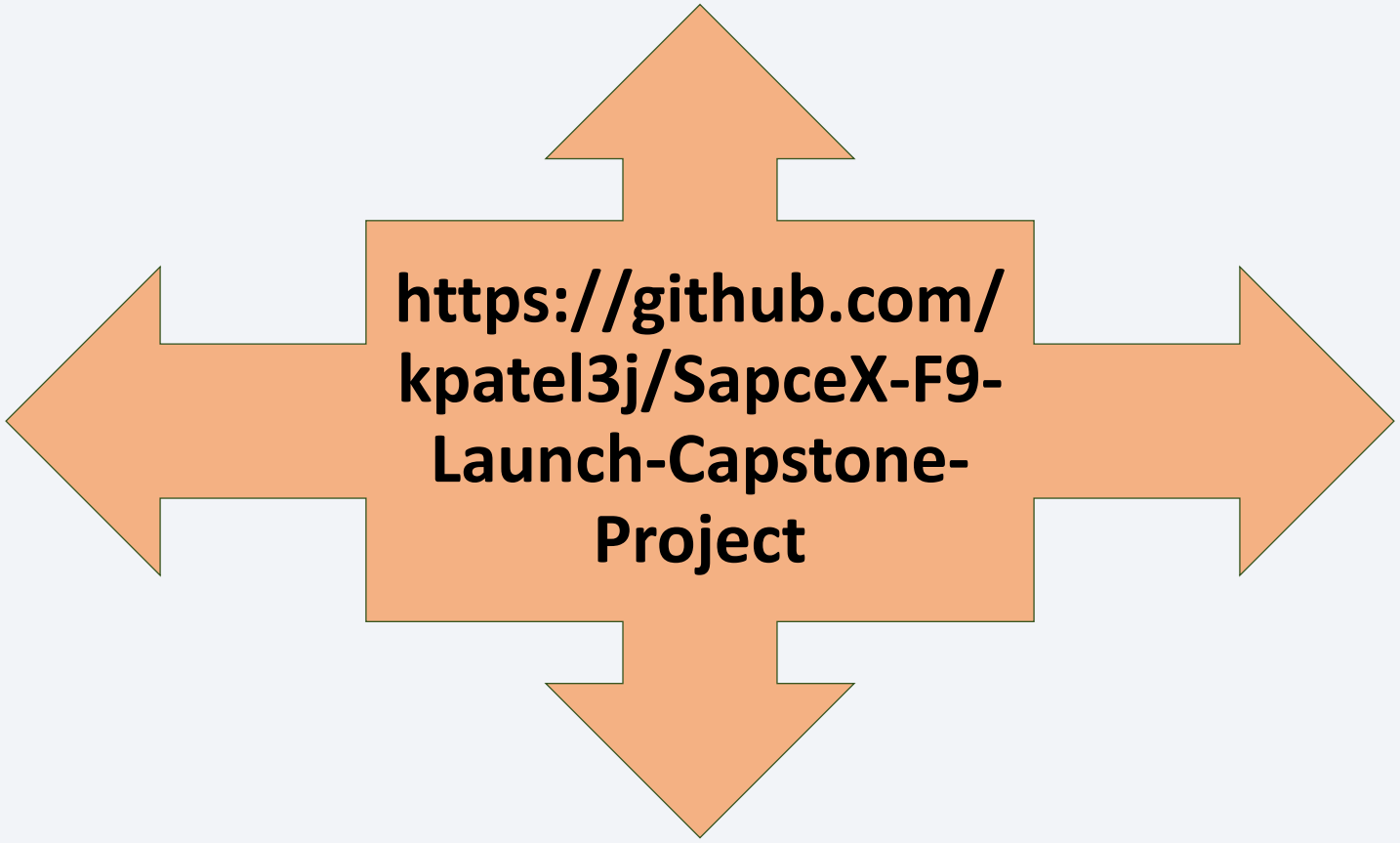
Conclusions

The success rate is 66% in exploratory data analysis

The success rate is 66% in exploratory data analysis

The success rate is 66% in exploratory data analysis

Appendix


https://github.com/kpatel3j/SapceX-F9-Launch-Capstone-Project

Thank you!