IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Karshni Mitra
27/11/2022

Outline

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

Executive Summary

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- EDA with Folium
- Predictive analysis (Classification)
- Exploratory data analysis results
- Predictive analysis results

Introduction

- Project background and context
- Analyse SpaceX Launch data to improve their rate of success in missions and find out key factors that directly increase success rate
- Problems you want to find answers
- Find the factors that affect successful landing
- Find the features of rockets to predict their mission success rate
- Find the desirable conditions for SpaceX to increase landing chances

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX REST API requests and WebScraping
- Perform data wrangling
  - Landing outcome label created, null values taken care of, one hot encoding of dummy variables
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built Several classifications models (KNN,LR,SVM) and compared them using accuracy score metrics on the test dataset.

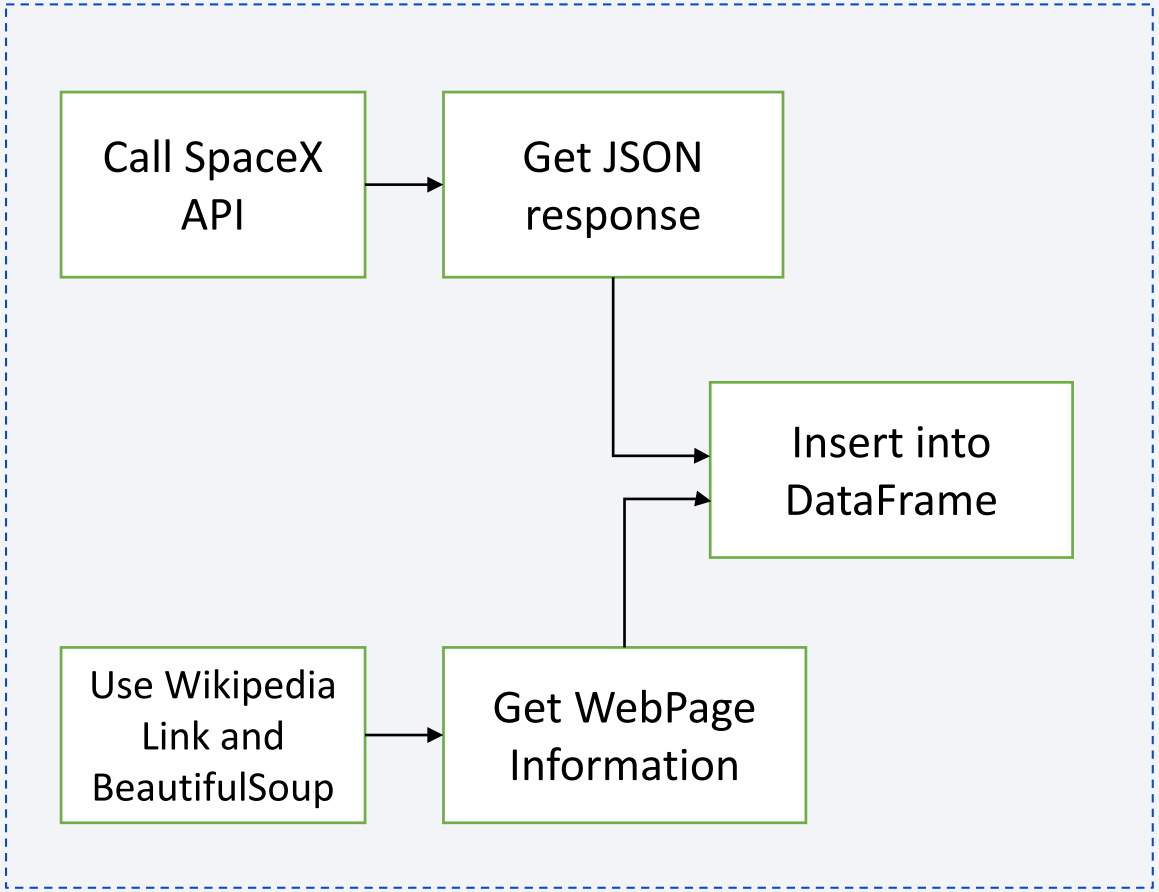
Data Collection

- Data was collected using BeautifulSoup which is a webscraping library. Leveraged this to scrape Wikipedia data.

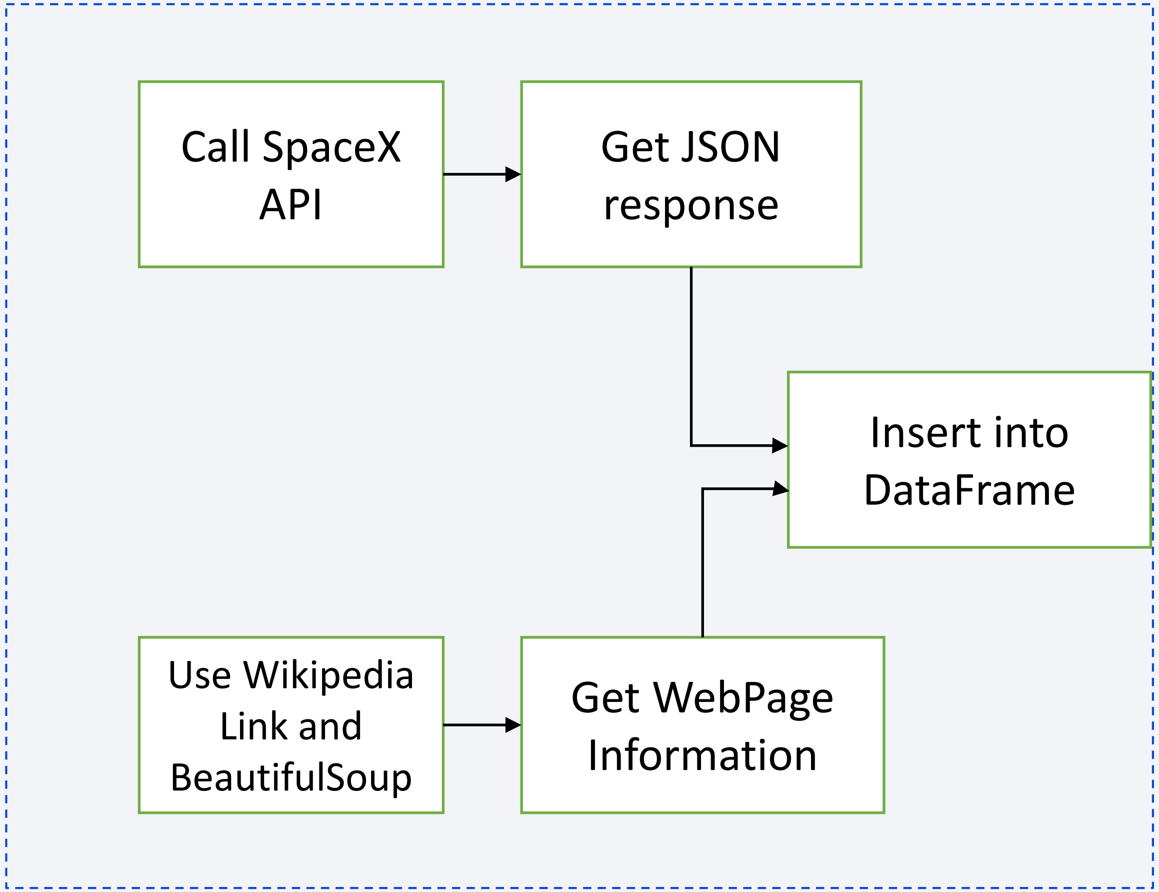
- Data was also collected from SpaceX REST API endpoints.

Data Collection – SpaceX API

- https://github.com/karshnimitra/IBM-DS-Capstone/blob/7fc5bf55d8cfef928ea912e296ee038b963fc7b8/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping



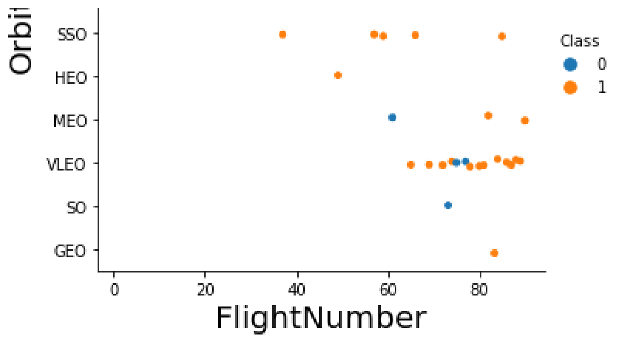
Data Wrangling

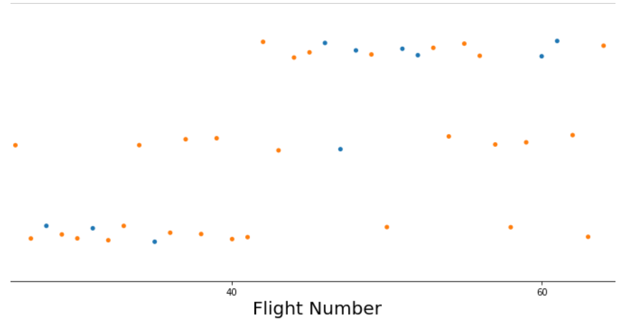
- Checked for null values and handled
- Computed launches for each site
- Created values from data about number of orbit occurrences and their types
- Analyzed mission outcome for each type
- Created a landing outcome label column
- https://github.com/karshnimitra/IBM-DS-Capstone/blob/7fc5bf55d8cfef928ea912e296ee038b963fc7b8/labs-jupyter-spacex-Data%20wrangling.ipynb

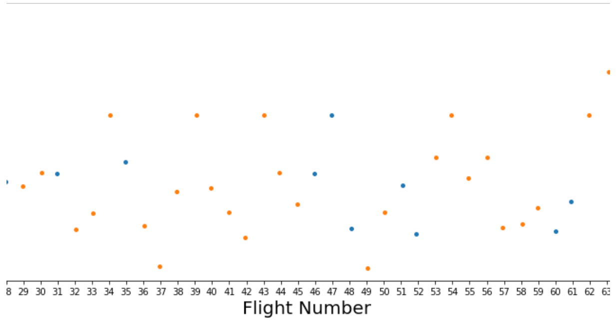
EDA with Data Visualization

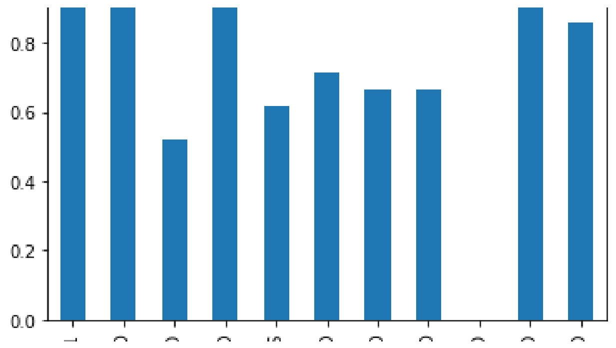
- https://github.com/karshnimitra/IBM-DS-Capstone/blob/7fc5bf55d8cfef928ea912e296ee038b963fc7b8/jupyter-labs-eda-dataviz.ipynb

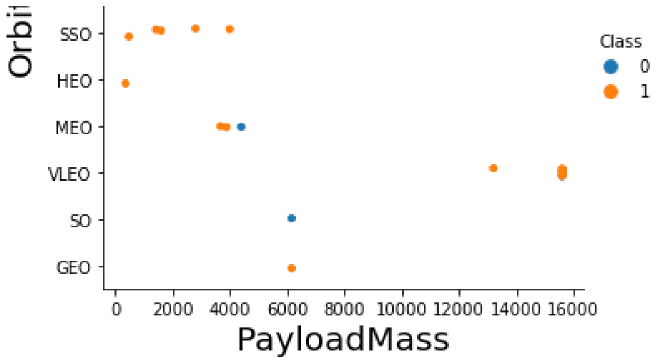
- Visualized the relationship between various features of the data.

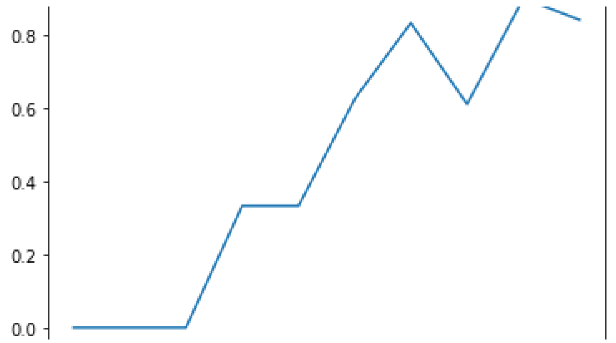












- Chart 1 shows the relationship between orbit type and success.
- Chart 2 shows relationship between flight success and launch site.
- Chart 3 shows the effect of payloadmass and launchsite on mission success.
- Chart 4 shows success rate per orbit type.
- Chart 5 shows the yearly trend for mission success.
- Chart 6 shows relationship between orbit and payload type.

EDA with SQL

- 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 succesful 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 using a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

https://github.com/karshnimitra/IBM-DS-Capstone/blob/7fc5bf55d8cfef928ea912e296ee038b963fc7b8/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

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

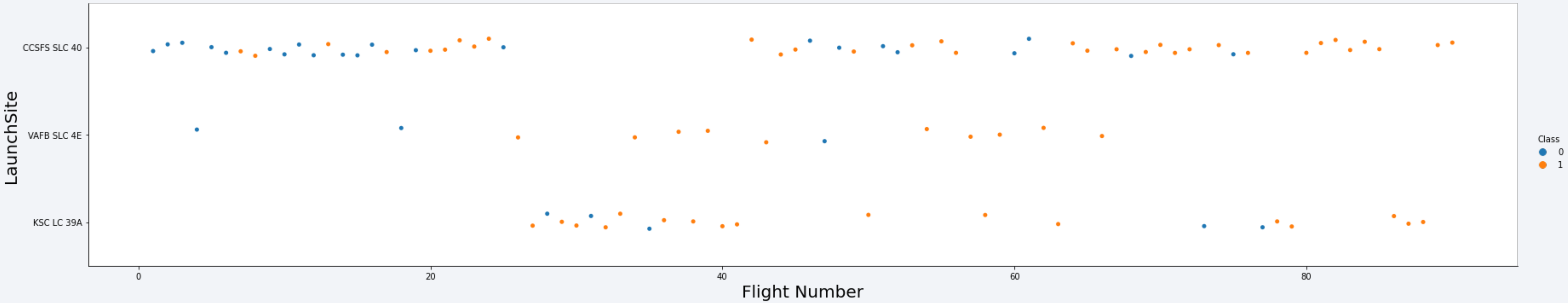
Results

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

Section 2

Insights drawn from EDA

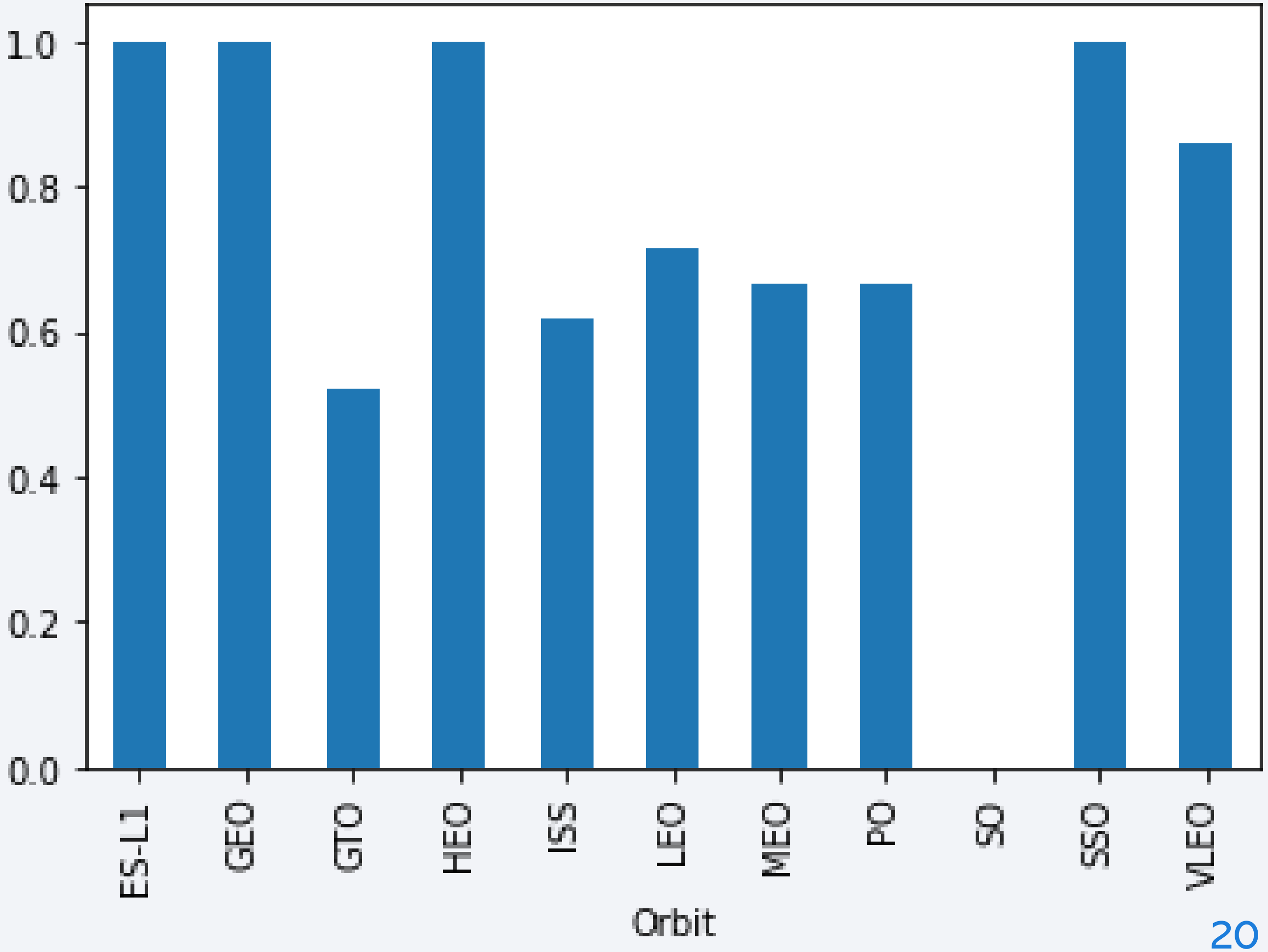
Flight Number vs. Launch Site



CCSFS SLC 40 has the most launches, but KSC LC 39A seems to have the higher success rate.

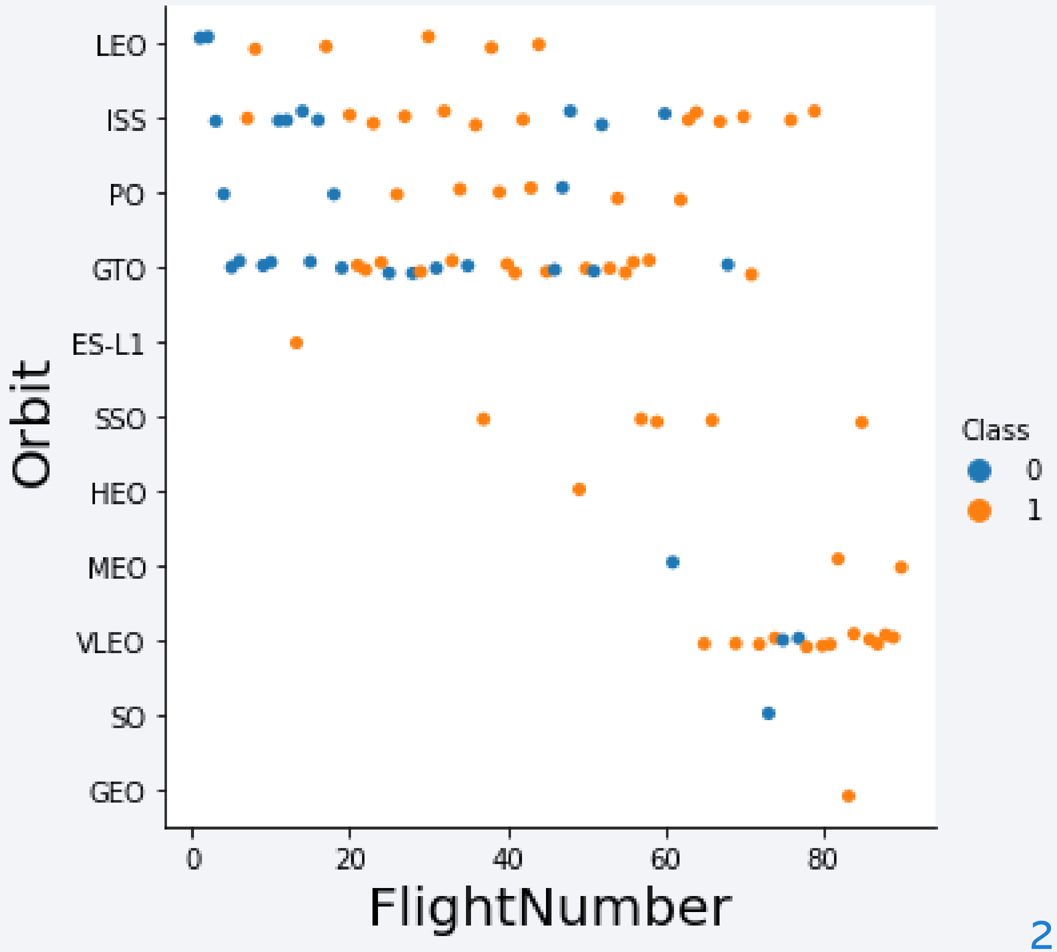
Success Rate vs. Orbit Type

- HEO, ES-L1, HEO, SSO seem to have perfect success rates.
- GTO has the lowest, closely followed by ISS, LEO, PO



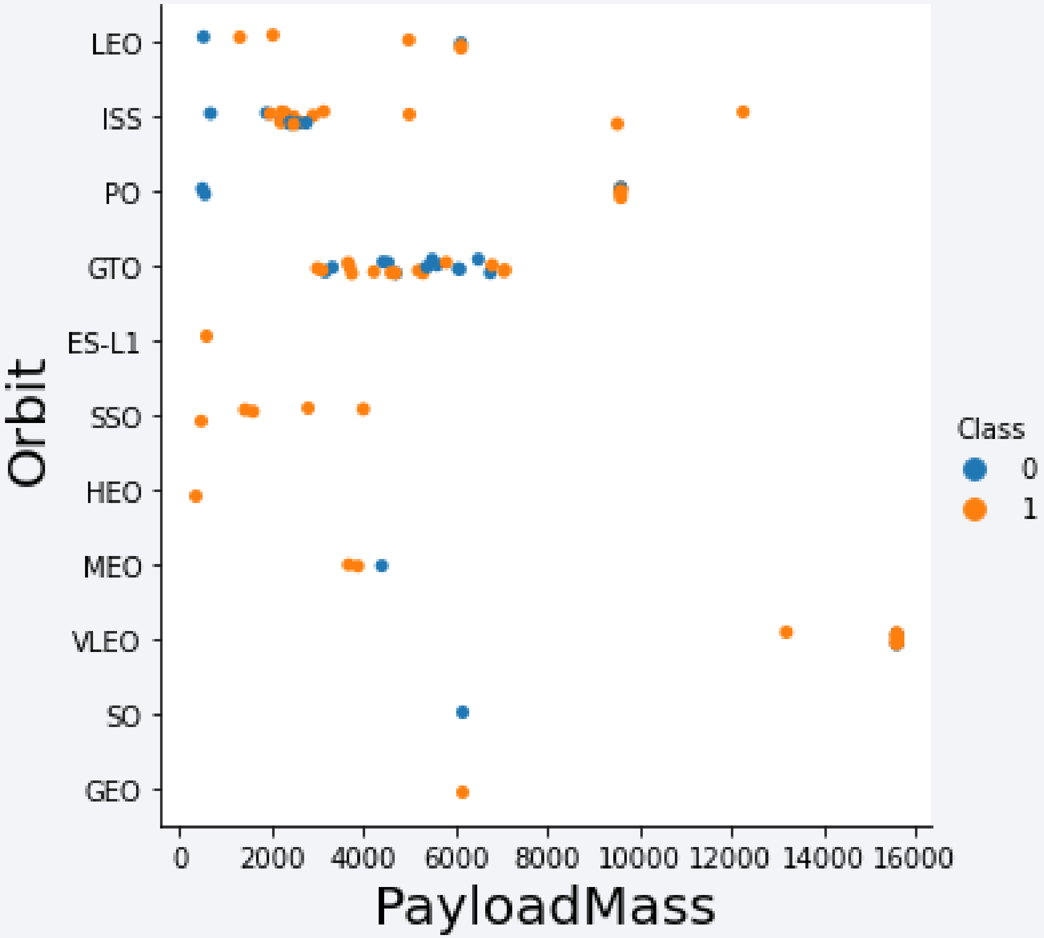
Flight Number vs. Orbit Type

- VLEO seems to have a lot of flights/launches recently as opposed to the others that are nearly consistent.



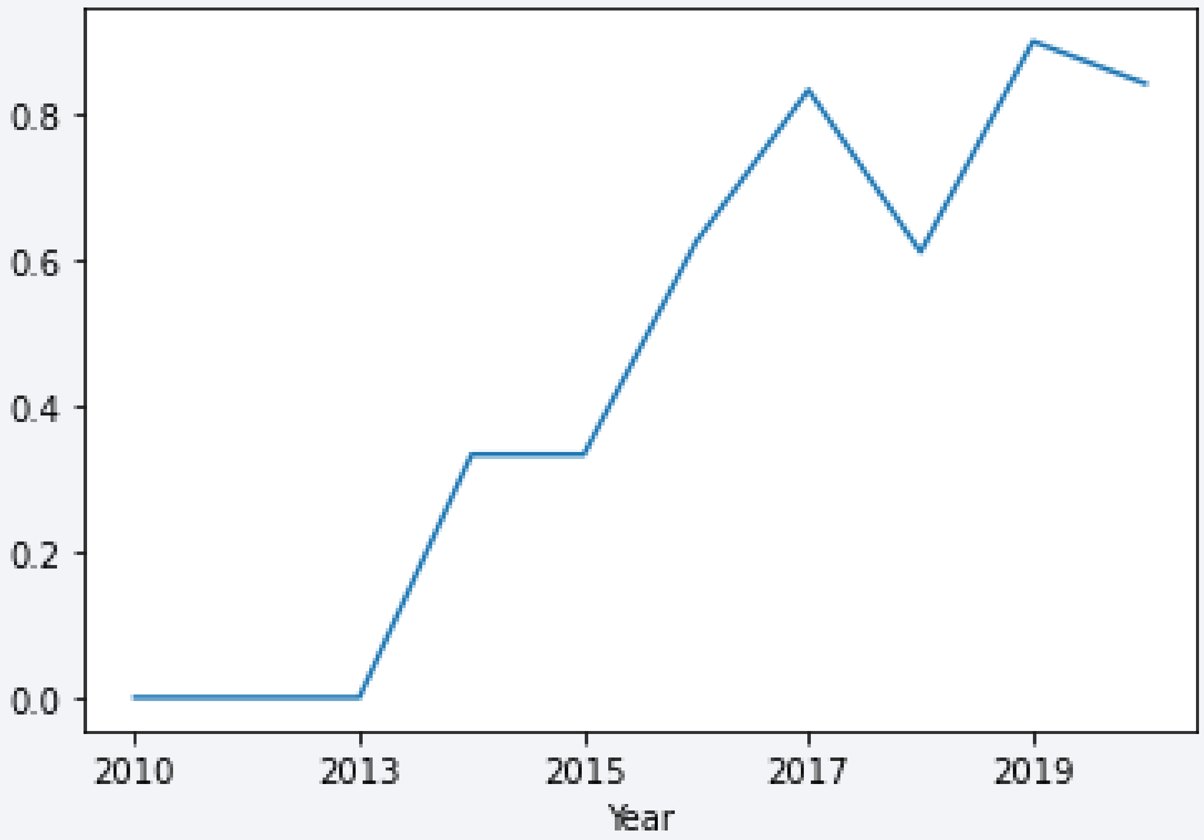
Payload vs. Orbit Type

- Lower payloads have more flights as it is cheaper and they seem to have a better success rate.



Launch Success Yearly Trend

- The success yearly trend is positive hence number of successes are increasing year by year on an average.



All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL
- The various launch sites are:

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where "Launch_Site" like 'CCA%' limit 5

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

- %sql select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" LIKE 'NASA (CRS)’

- Total payload for all NASA (CRS) in kg:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- %sql select avg("PAYLOAD_MASS__KG_") from SPACEXTBL where "Booster_Version" like 'F9 v1.1’

- Average payload in KG for F9 boosters is:

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- %sql select min("Date") from SPACEXTBL where "Landing_Outcome" LIKE 'Success (ground pad)’

min(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select "Booster_Version" from spacextbl where "Landing_Outcome" LIKE 'Success (drone ship)' and "PAYLOAD_MASS__KG_" between 4000 and 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

- %sql select "Mission_Outcome", count(*) as Count from spacextbl group by "Mission_Outcome“

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

Boosters Carried Maximum Payload

- %sql select "Booster_Version" from spacextbl where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from spacextbl)

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

- %sql SELECT "Booster_Version","Landing_Outcome","Launch_Site", substr("Date", 4, 2) as "Month", FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)' and substr("Date",7,4) = '2015’

Booster_Version	Landing_Outcome	Launch_Site	Month	Date
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40	01	10-01-2015
F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40	04	14-04-2015

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

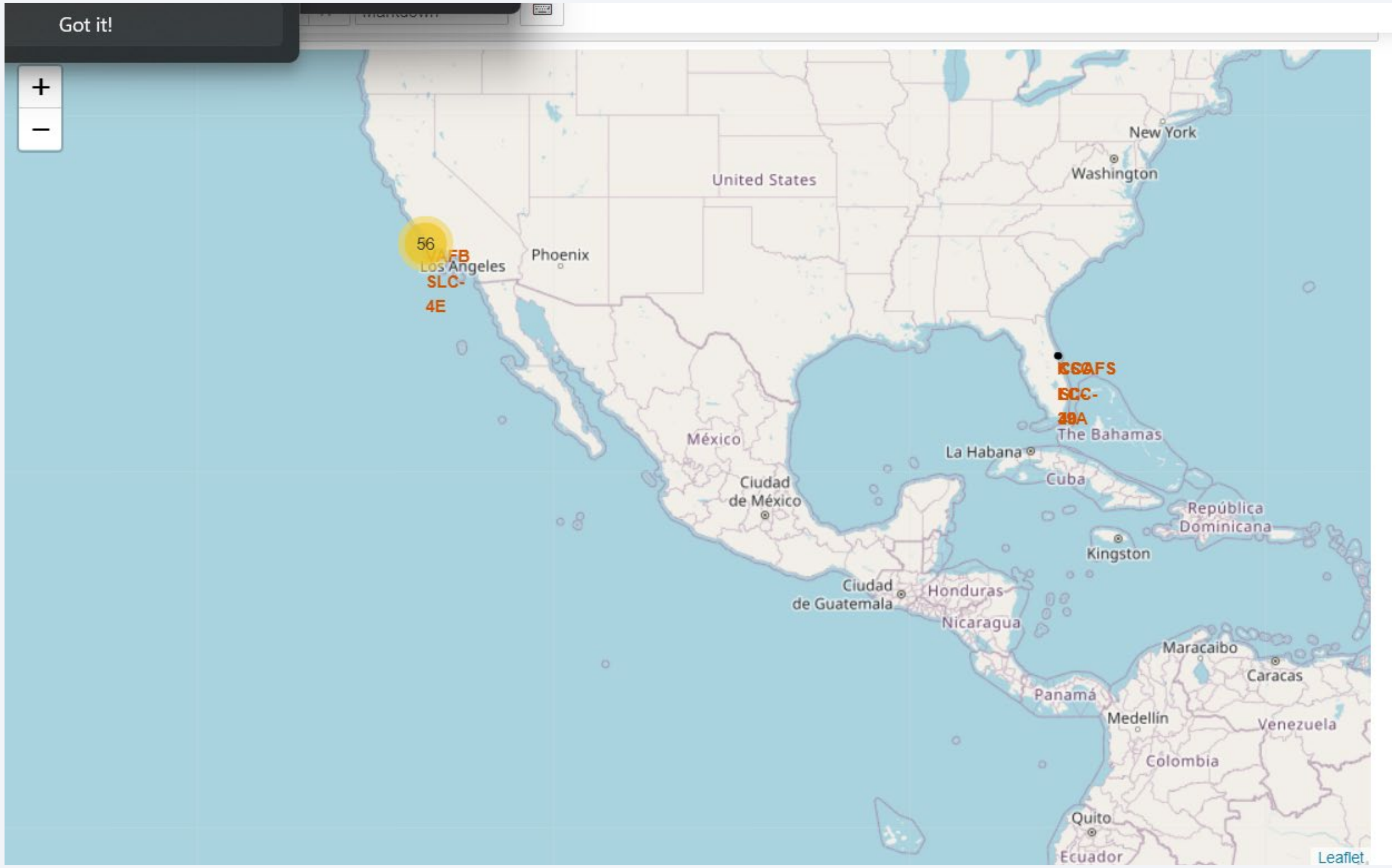
- %sql select * from spacextbl where "Landing_Outcome" like 'Success%' and "Date" between '04-06-2010' and '20-03-2017' order by "Date" desc

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)

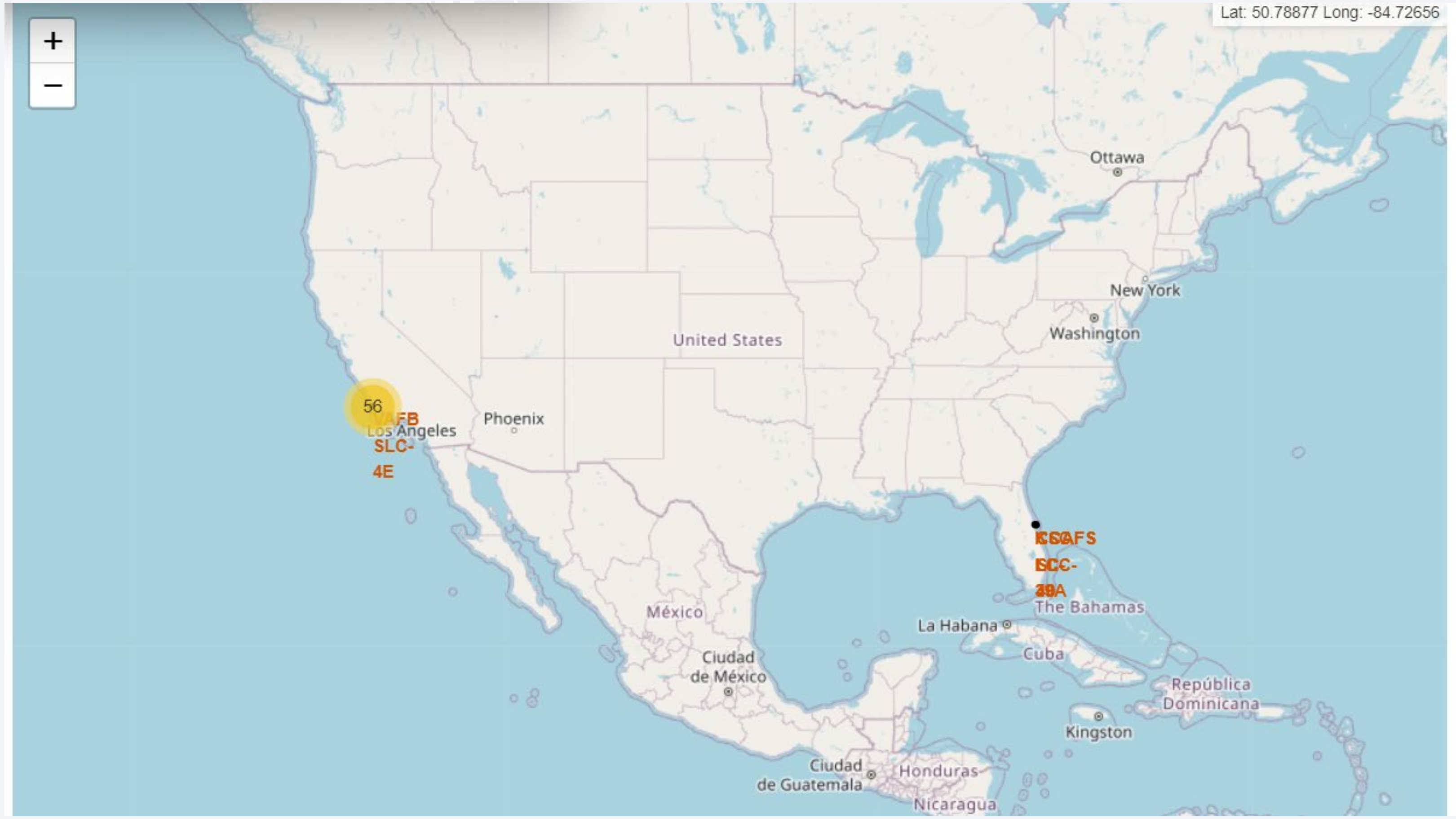
Section 3

Launch Sites Proximities Analysis

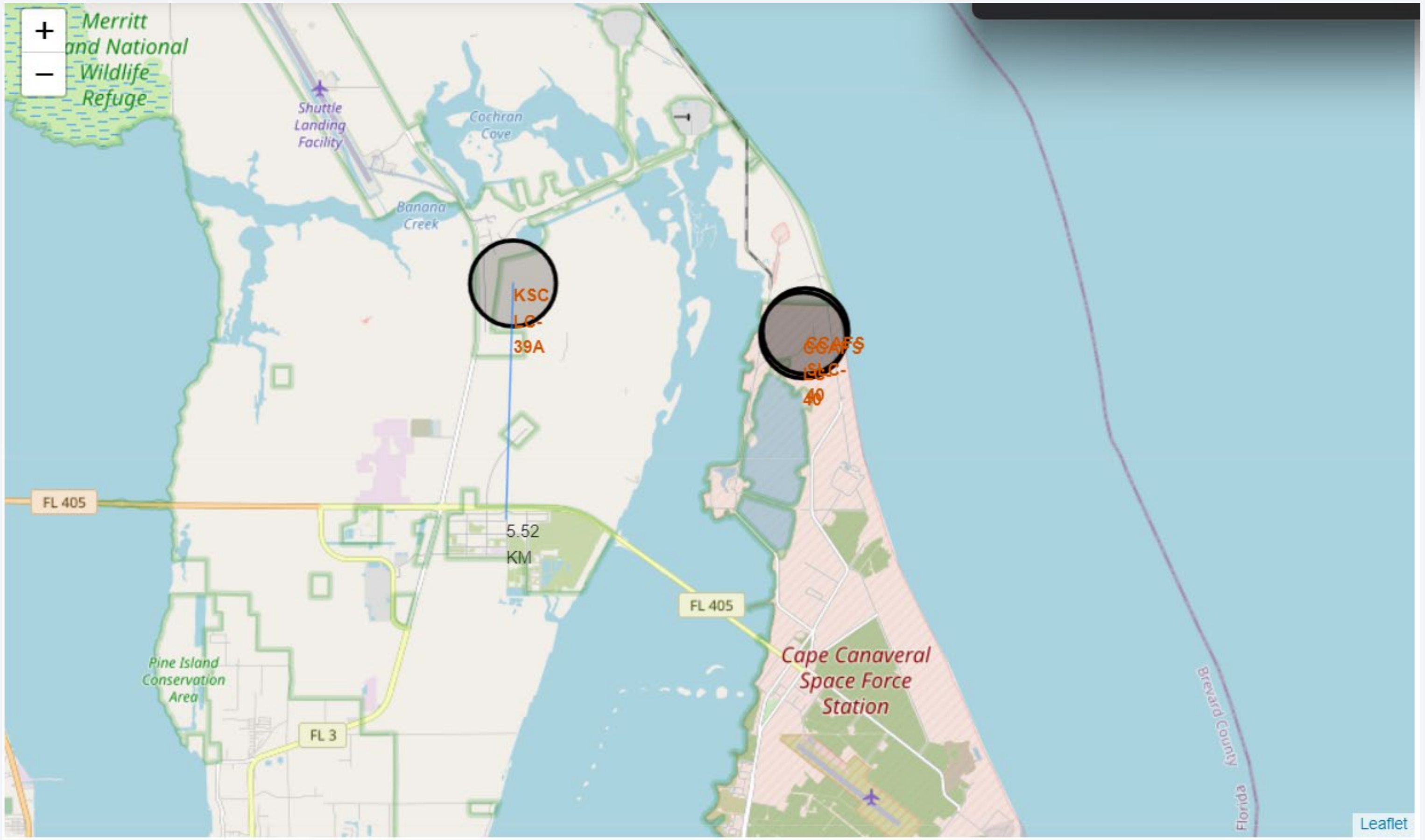
Launch Locations Folium Map



Launch Sites Folium Map



Proximities Folium Map

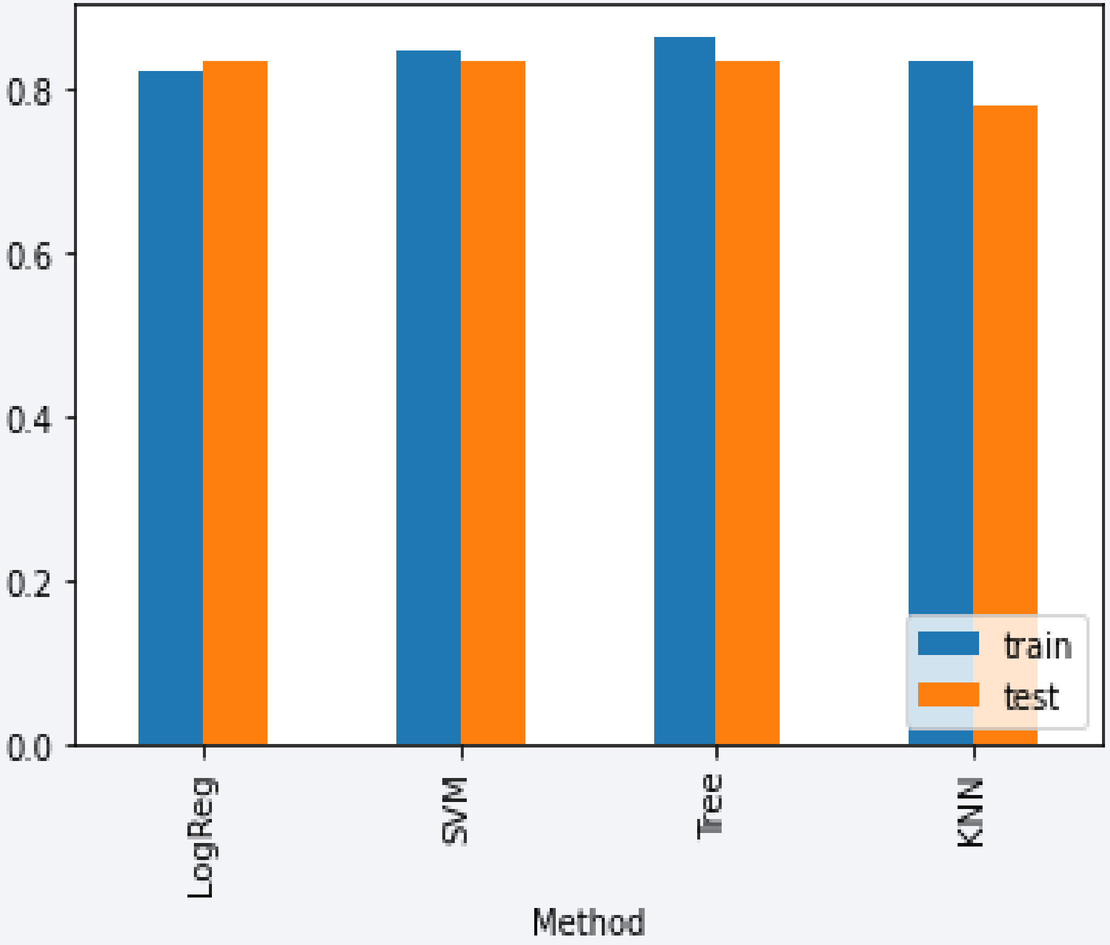


Section 5

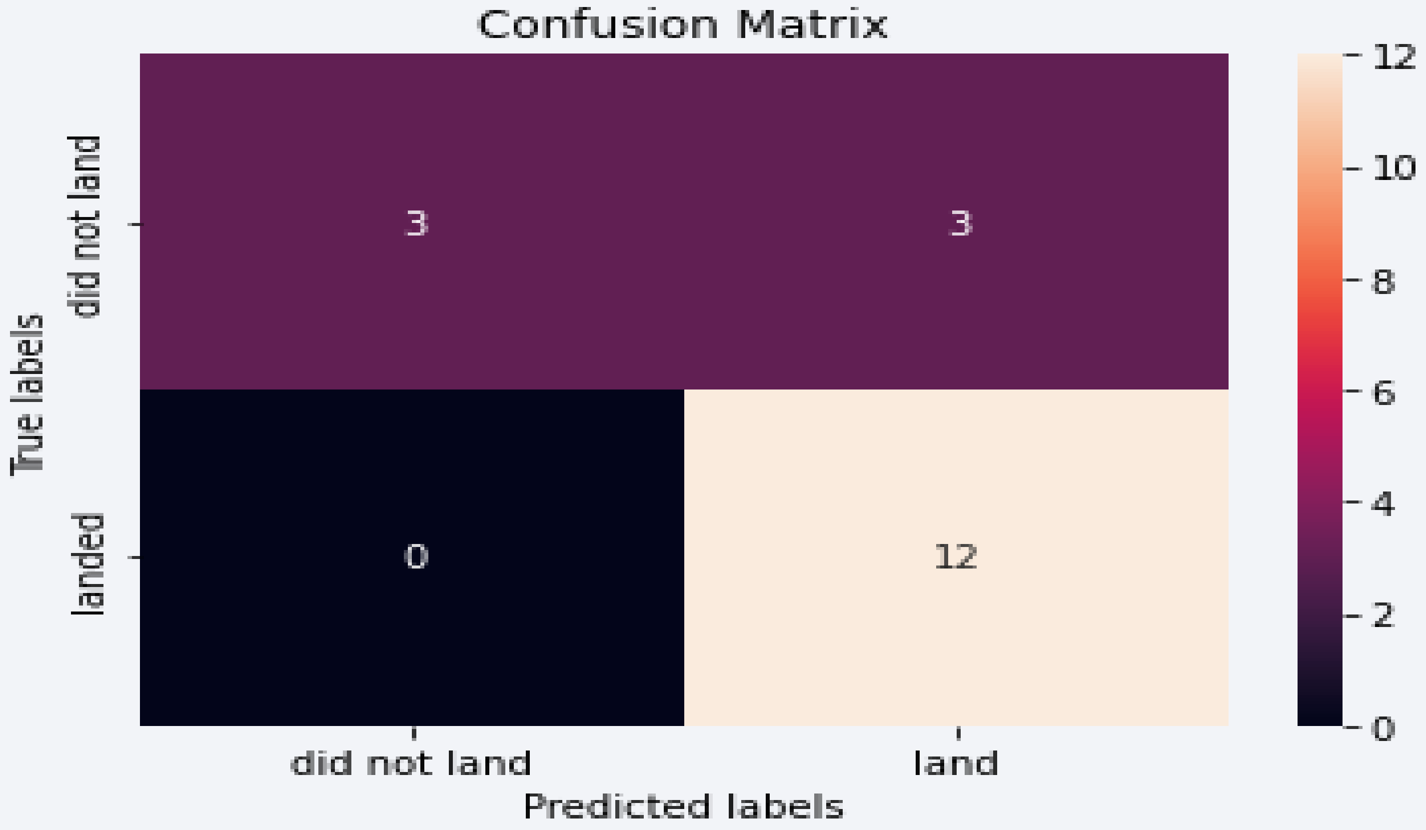
Predictive Analysis (Classification)

Classification Accuracy

- From the chart data, we can see that the decision tree has the highest accuracy rate



Confusion Matrix



- The confusion matrix of the Decision Tree is as follows: 15/18 accuracy

Conclusions

- The Decision Tree with the optimal parameters seems to be the best classifier model for this dataset.
- On average, the success rate of SpaceX is increasing annually.
- Lower Payload flights seem to perform better.
- KSC LC 39 A is the most successful launch site in terms of proportion of successful flights in comparison to the other launch sites.

Thank you!