IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Nyan Htet Aung>
<2022-Jul-16>

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection using API and Web Scraping of Space X
  - Exploratory Data Analysis (EDA) with SQL
  - Data Wrangling, Data Visualization and Interactive Data Visual Analytics
  - Machine Learning Prediction
- Summary of all results
  - Collection of valuable data from public sources
  - Data analytics methods were used to get result of launching prediction data.
  - Machine Learning prediction along with data collection was used to identify the characteristics necessary to success of project Space Y.

Introduction

- The main objective of this project is to evaluate viability of New Company “Space Y” to compete with “Space X”.

- It is necessary to study the success stories Space X so that Space Y will have lower failure rate.
- It is also necessary to study the best possible launch sites.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from REST API: https://api.spacexdata.com/v4/rockets
  - Data was collected from web scraping of Wikipedia: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Perform data wrangling
  - Data collected were cleaned and changed into format that can be summarized into as outcome data.

Methodology

Executive Summary (continued)

- 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 collected were normalized, split into training and test data sets, evaluated by different classification models, accuracy of each model is evaluated by using combination of parameters.

Data Collection

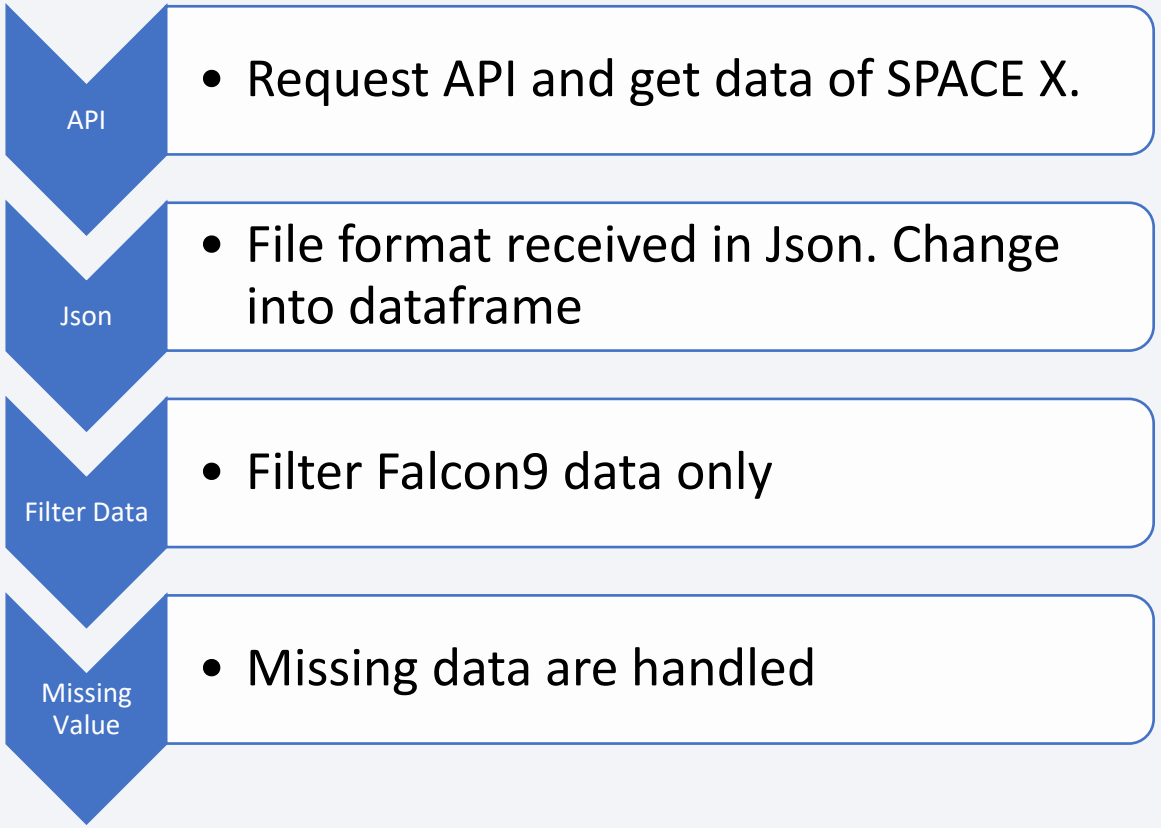
- First data collection were done from SPACE X's API using Rest API and collected data are in Json format. Then data are normalized

- Data collection from Wikipedia was done with Web Scraping technic using Beautiful Soap. Then data are normalized.

Data Collection – SpaceX API

- Data are collected with REST API from SPACE X website and followed into sequence as per the workflow.

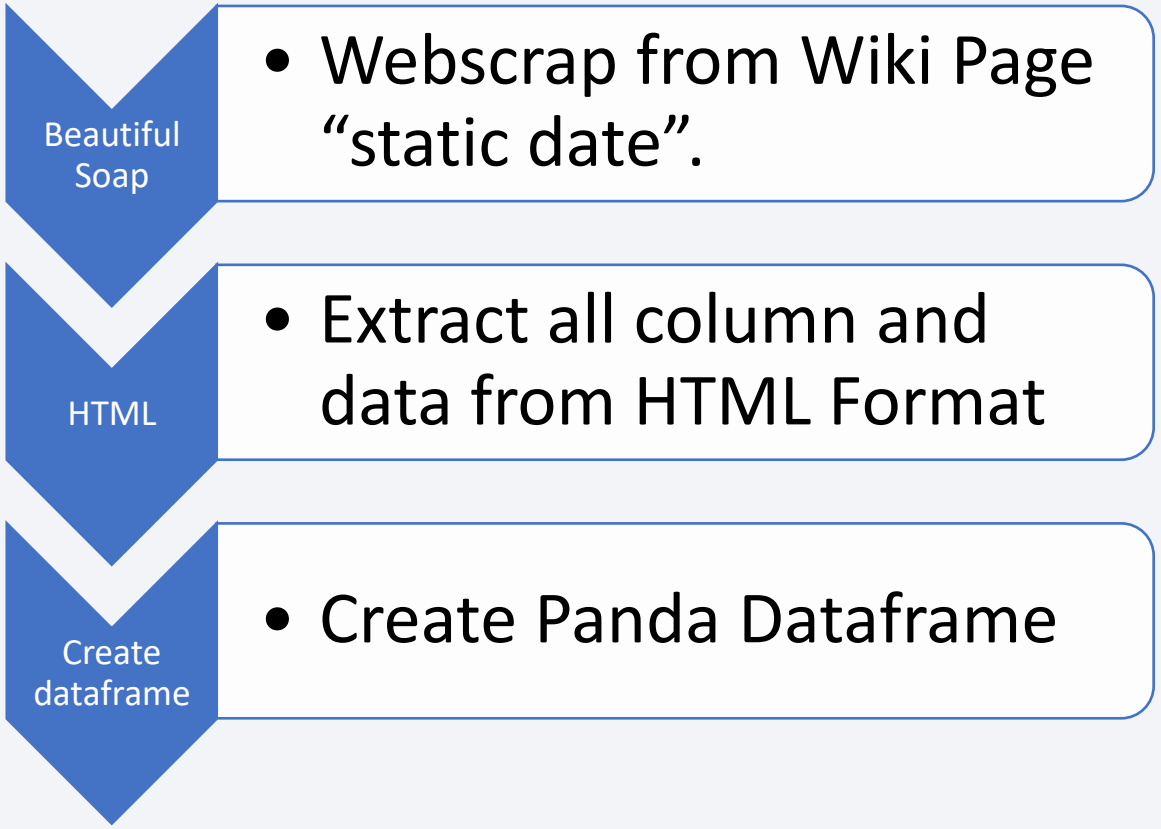
- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/Week%201%20Capstone.ipynb



Data Collection - Scraping

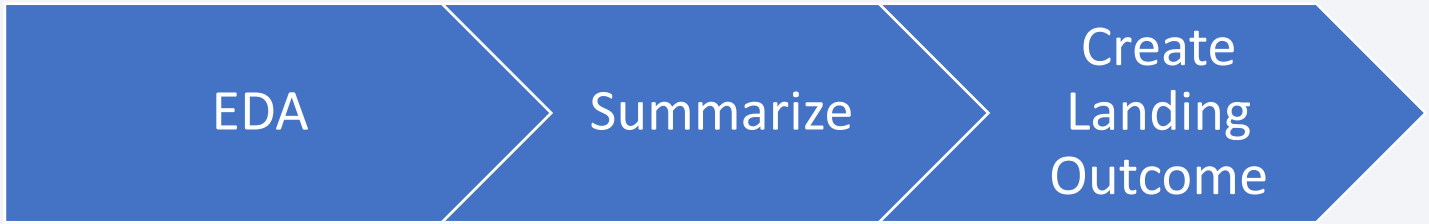
- Data are collected from Wikipedia using Beautiful soap and followed into sequence as per the workflow.

- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/webscrap.ipynb



Data Wrangling

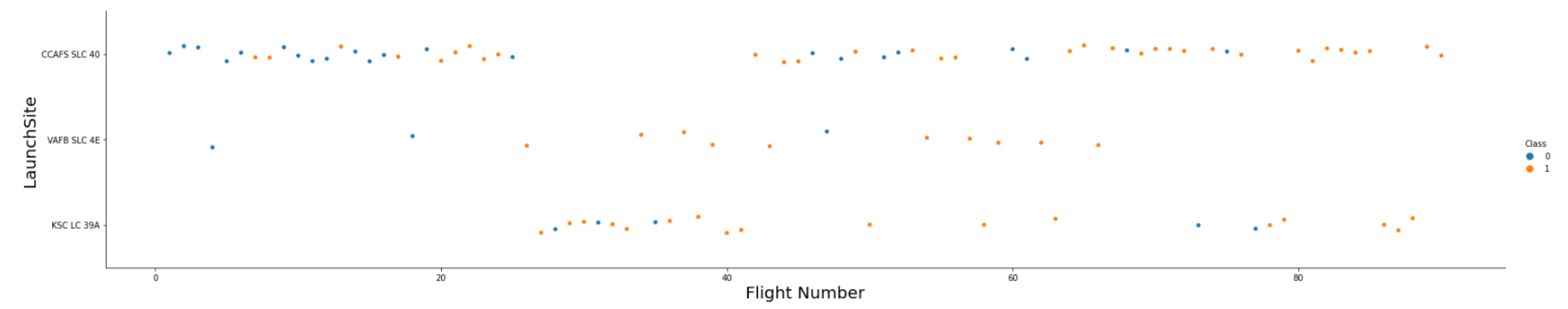
- EDA was performed on the data set.
- Summary of launches per site, Orbit, Payload, Mission Outcomes per orbit, etc were calculated.
- Lastly, Landing Outcome column was created and success rate was calculated.

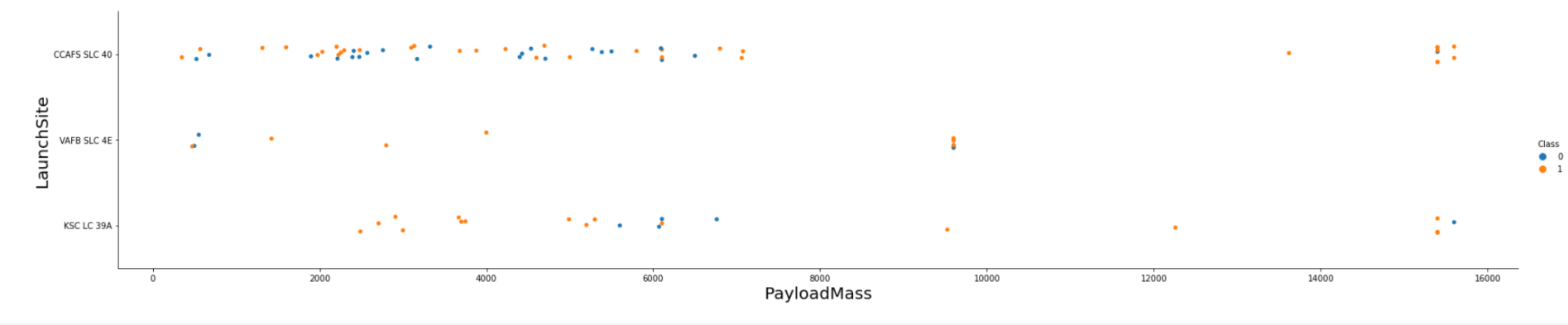


- GitHub: https://github.com/nyan-htet/Cousera/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

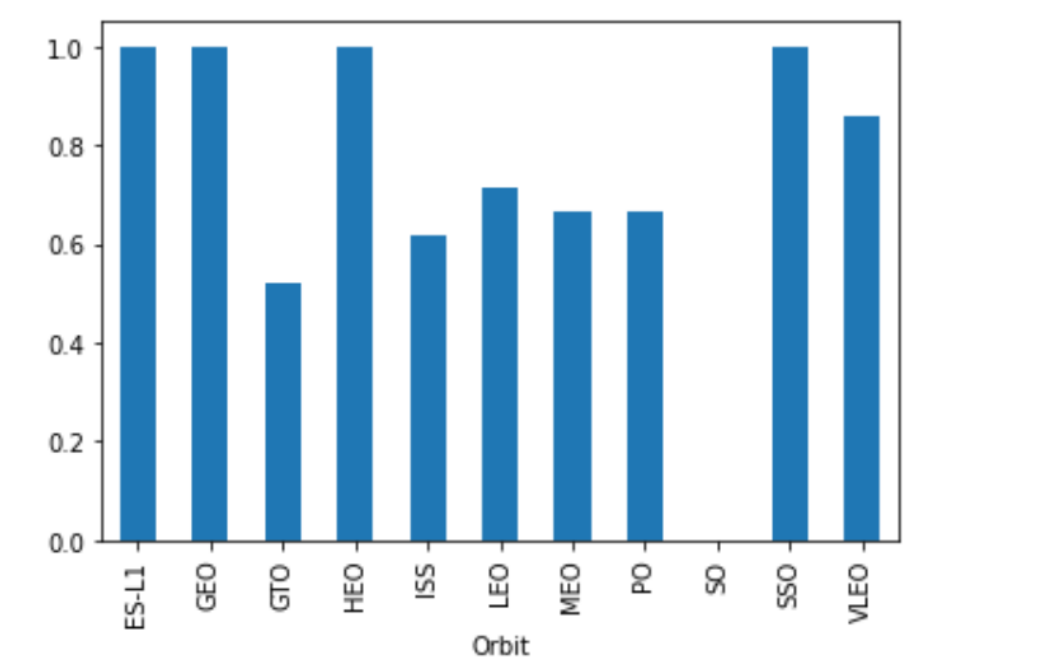
EDA with Data Visualization

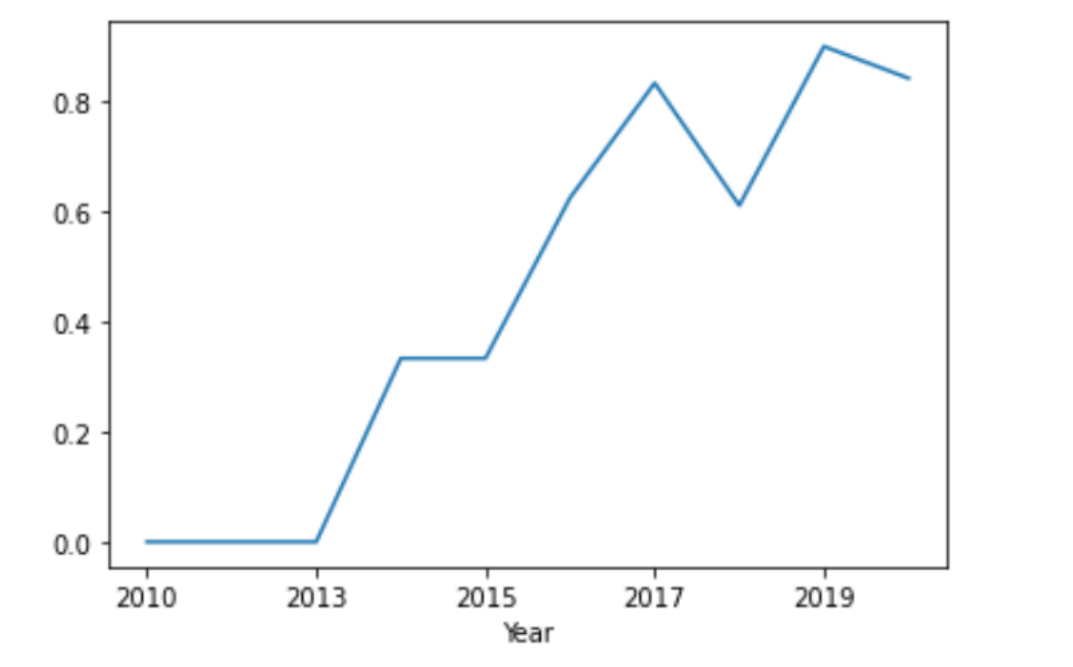
- Scatter plots and bar plots were used for launchsite, payload, flight number, Orbit, and then find relation among them. Also identified the success rate progression year over year.





EDA with Data Visualization





- GitHub: https://github.com/nyan-htet/Cousera/blob/master/EDA%20with%20DataVisual.ipynb

EDA with SQL

- Get unique launch site names.
- Query Top 5 launch sites
- Summarized total payload mass
- Average payload mass by F9 v1.1
- Total successful and failure missions
- Booster version that carried out max payload
- Failure outcomes in year 2015
- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/EDA1.ipynb

Build an Interactive Map with Folium

- Launchsites were indicated by Markers.
- Circles were used in highlighting coordinates.
- Marker clusters are used in groups of events, launches in launchsite.
- Lines are used to indicate among coordinates.

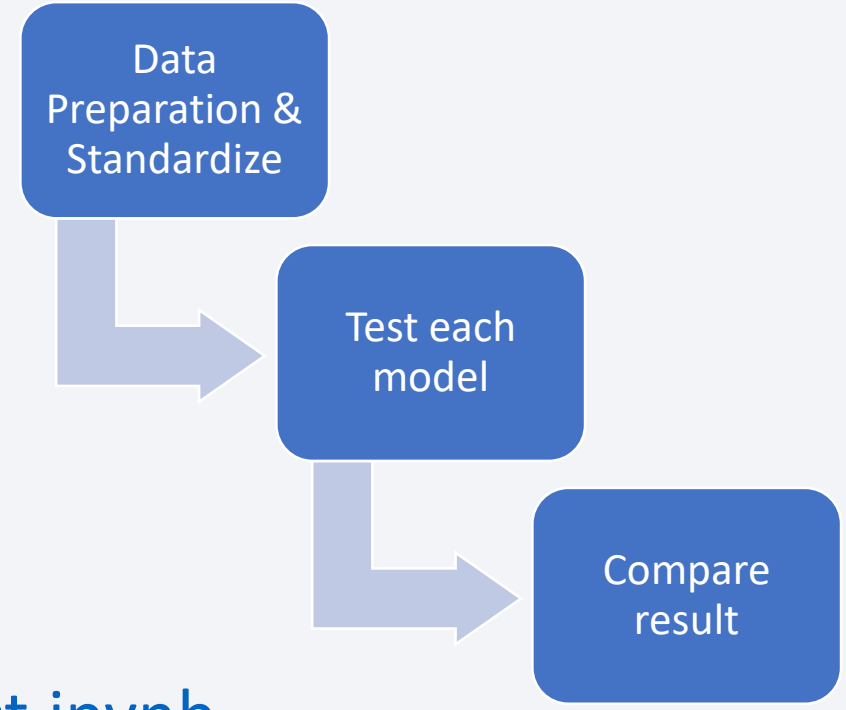
- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/Interative%20Visual.ipynb

Build a Dashboard with Plotly Dash

- In Summary, a graph and plot were used in visualize data.
  - Percentage of launches and payload range.
- The combination of payload and launchsites quickly help identify the best possible place to launch relating to payloads.
- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/Interative%20Visual.ipynb

Predictive Analysis (Classification)

- Total of 4 classifications model were studied.
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K nearest neighbours.



- GitHub URL: https://github.com/nyan-htet/Cousera/blob/master/ML%20Predict.ipynb

Results

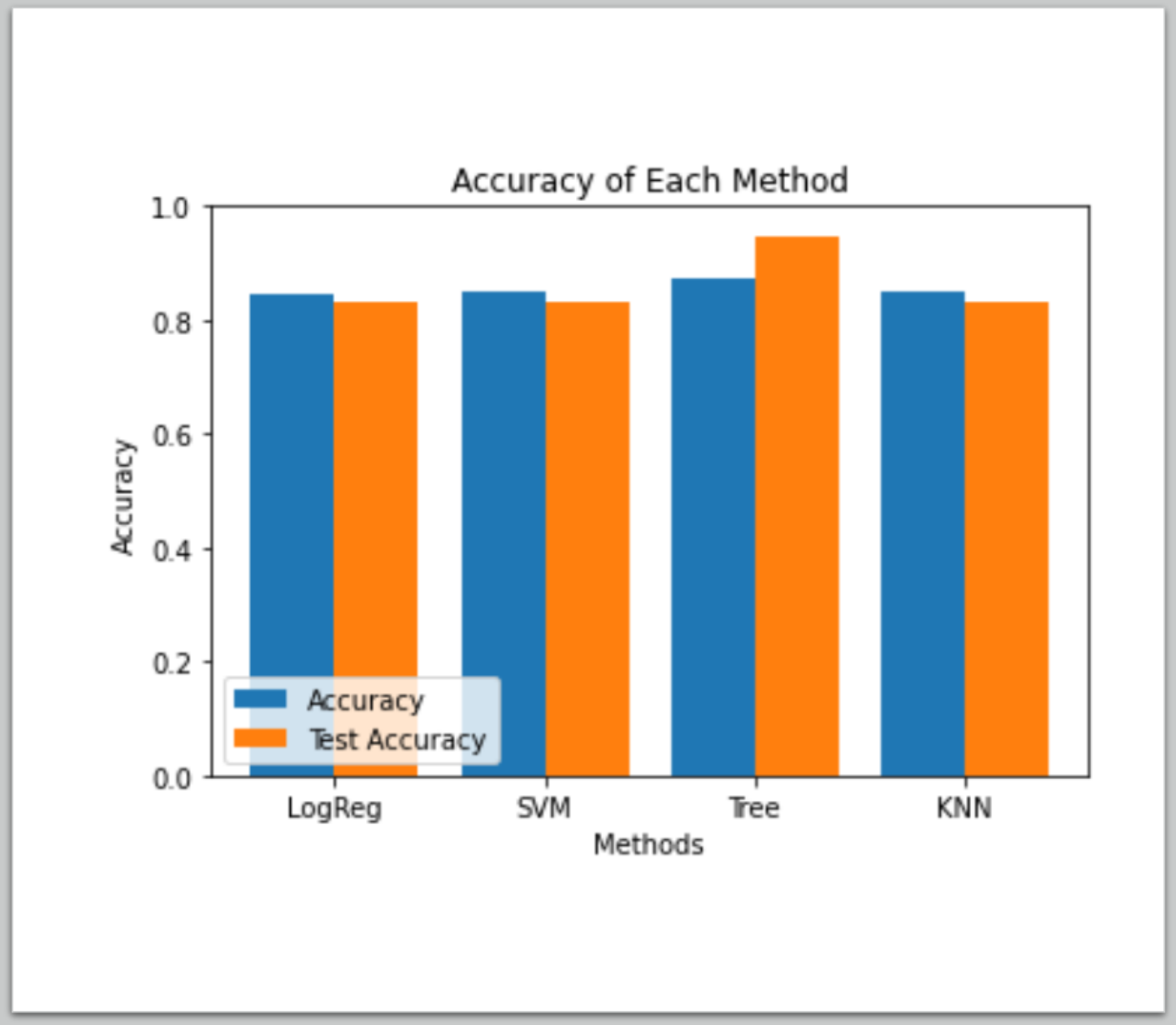
- Exploratory data analysis results
  - Space X has 4 launch sites.
  - Average payload F9 v1.1 = 2928kg.
  - First success landing – 2015
  - Falcon 9 booster with above average payload are successful at landing in drone ships (almost 100%)
  - Two booster versions fail: B1012, B1015.
  - Progress of success increase as years go by.

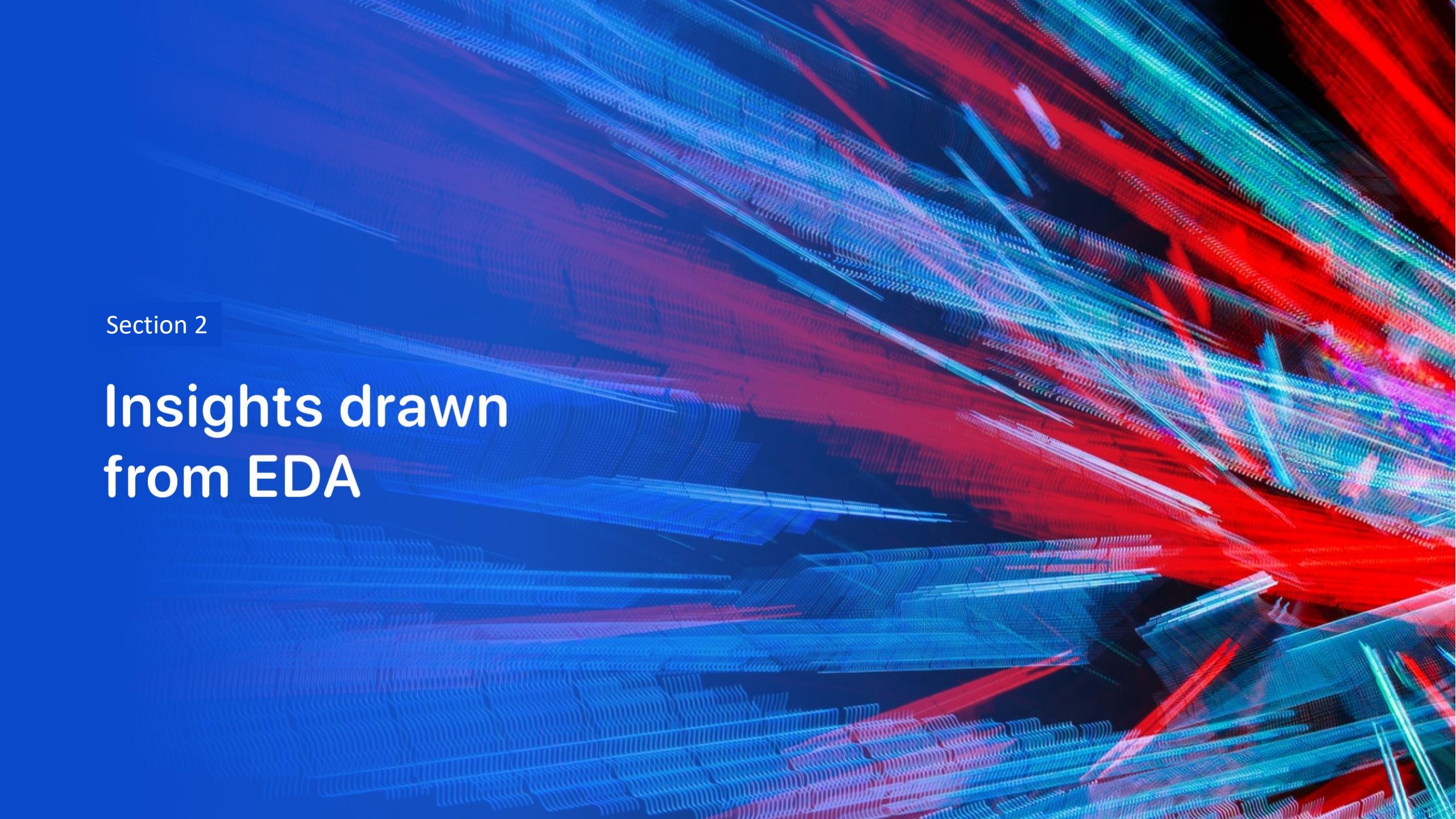
Results

- Interactive analytics demo in screenshots

Results

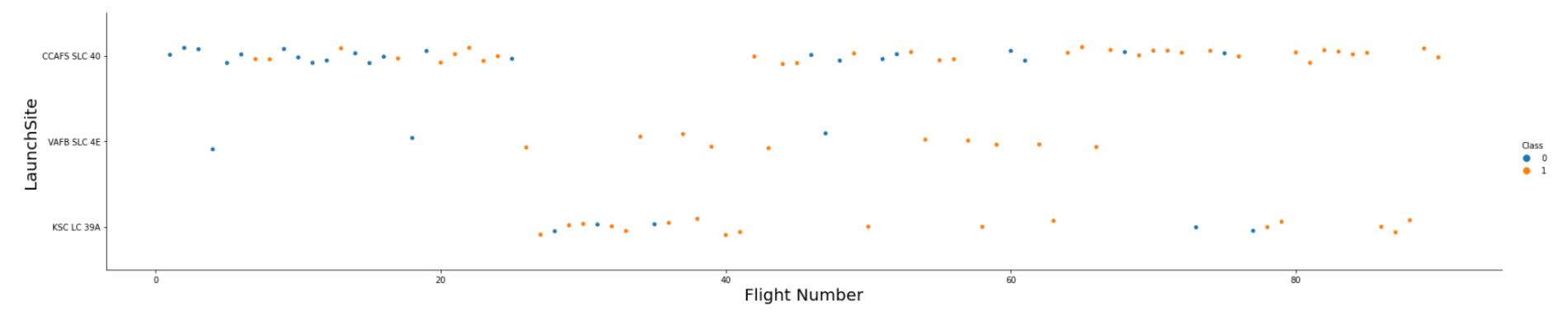
- Predictive analysis results
- The results show that decision tree classifier is the best model.
- It has predicted accuracy over 87% and test data accuracy over 94%.




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

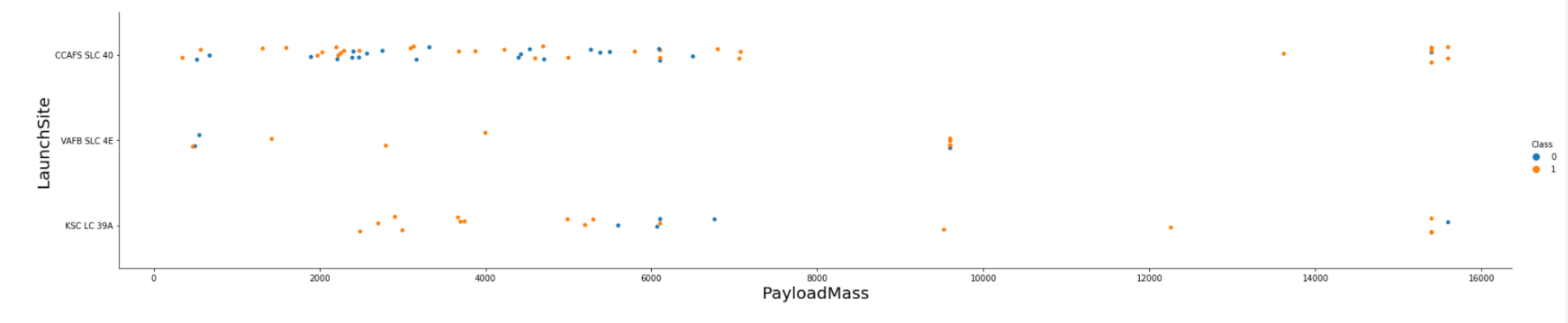
- Show a scatter plot of Flight Number vs. Launch Site



- Explanations
  - CCAF5 SLC 40 – most recent successful launches
  - VAFB SLC 4E ranks 2nd.
  - KSC LC 39A ranks 3rd

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

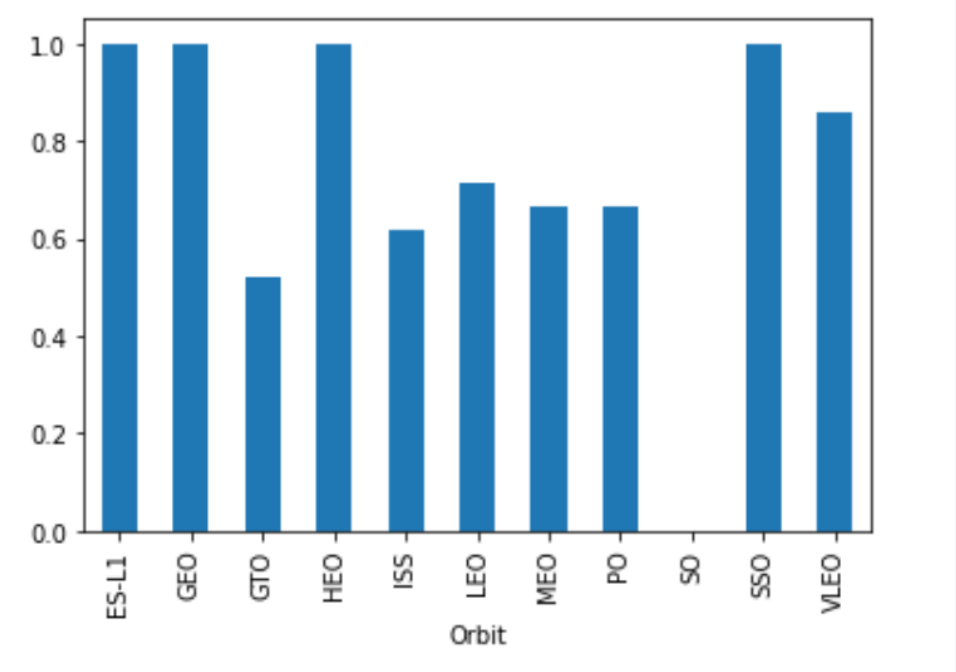


- Explanations
  - Payload over 9000 kgs show good success rate
  - Payload above 12,000kg are possible to CCAFS SLC 40 and KSC LC 39A.

Success Rate vs. Orbit Type

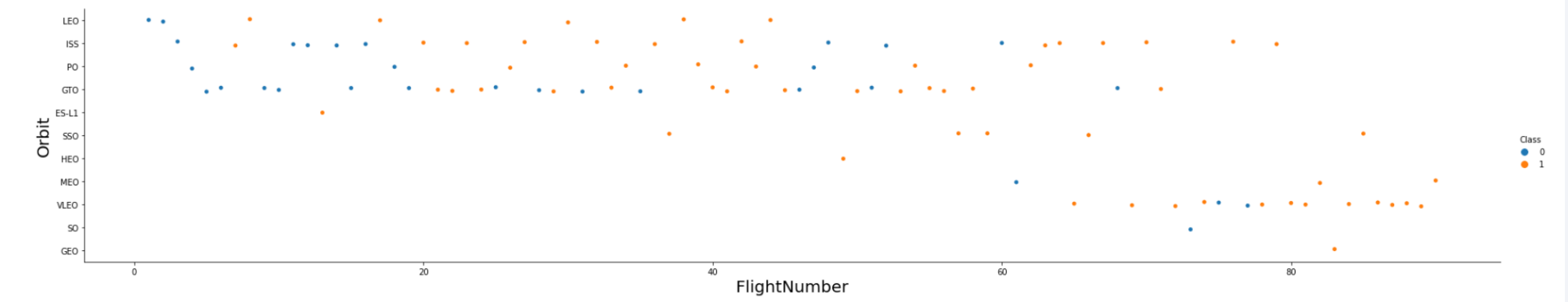
- Show a bar chart for the success rate of each orbit type

- Best success rates:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Least success rate
  - GTO



Flight Number vs. Orbit Type

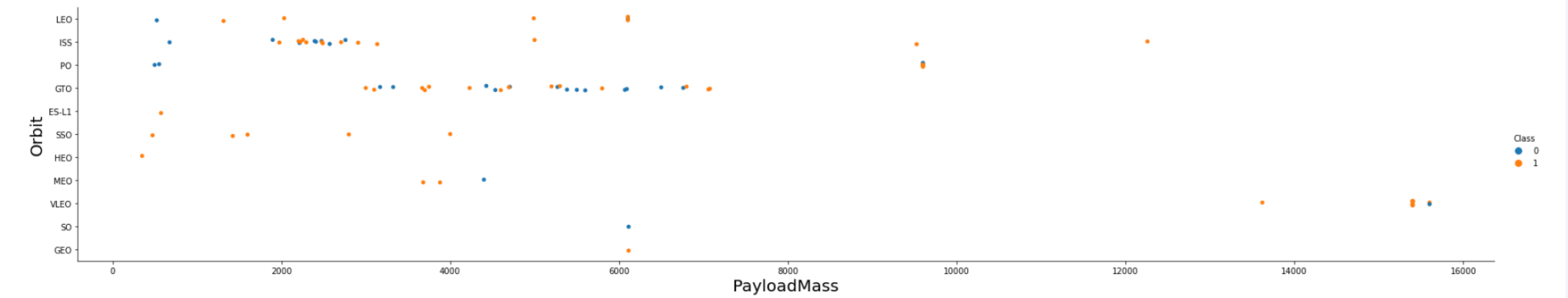
- Show a scatter point of Flight number vs. Orbit type



- Explanations
  - With increasing flight number, (as progress by time), more success rate is observed.

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

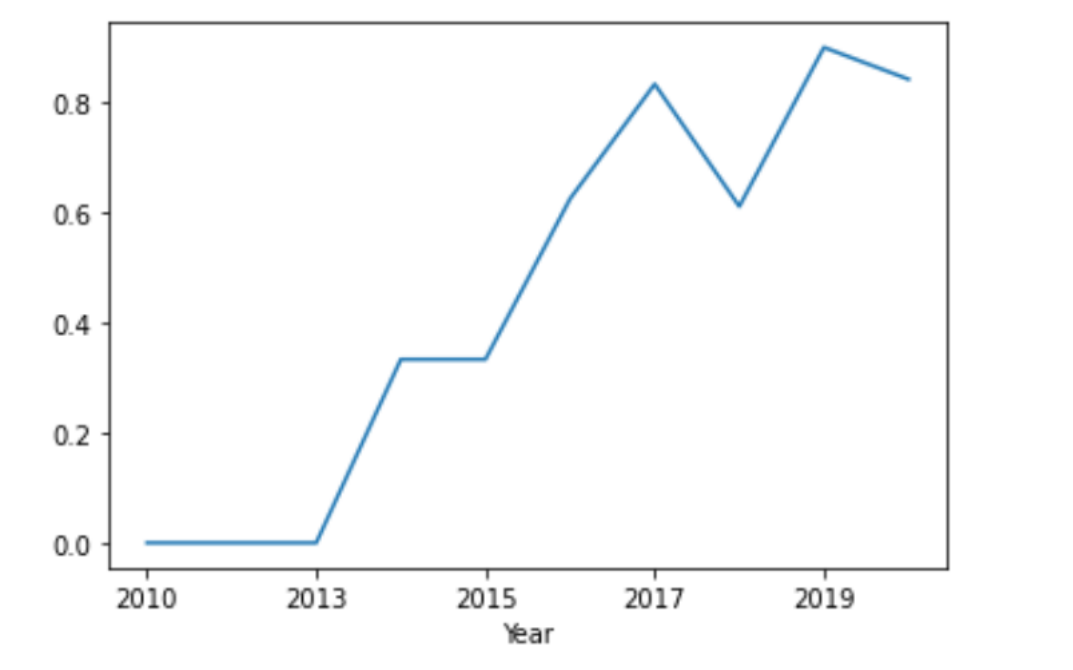


- Explanations
  - There does not seem to have relation between payload and success rate for GTO orbit.
  - Few launches to SO and GEO orbit.
  - Majority of less than 10,00 payload mass are found on Orbit like LEO, ISS PO, GTO

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Explanations
  - Success rate increase over the year.
  - 2010 to 2013 does not have success rate.



All Launch Site Names

- Names of the unique launch sites

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

- Used Select query of Launch_site from data.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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

- Query with %CAA and used limit 5.

Total Payload Mass

- The total payload carried by boosters from NASA

total_payload
111268

- Select Sum(payload mass kg) from data set and filter for NASA.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

avg_payload
2928

- Uses Average(payload_mass_kg) from dataset and filter by F9 using booster version

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad
  - 22-Dec-2015
- Select Min date from first success column.

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

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

- Query using distinct booster version and filter the payload mass kg between 4000 to 6000

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- Group the mission outcomes and count the records.

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass
- Select the query with max (payload kg mass)

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- Two filters are added in the query
  - Landing outcoe – failure
  - Year - 2015

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

- 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

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

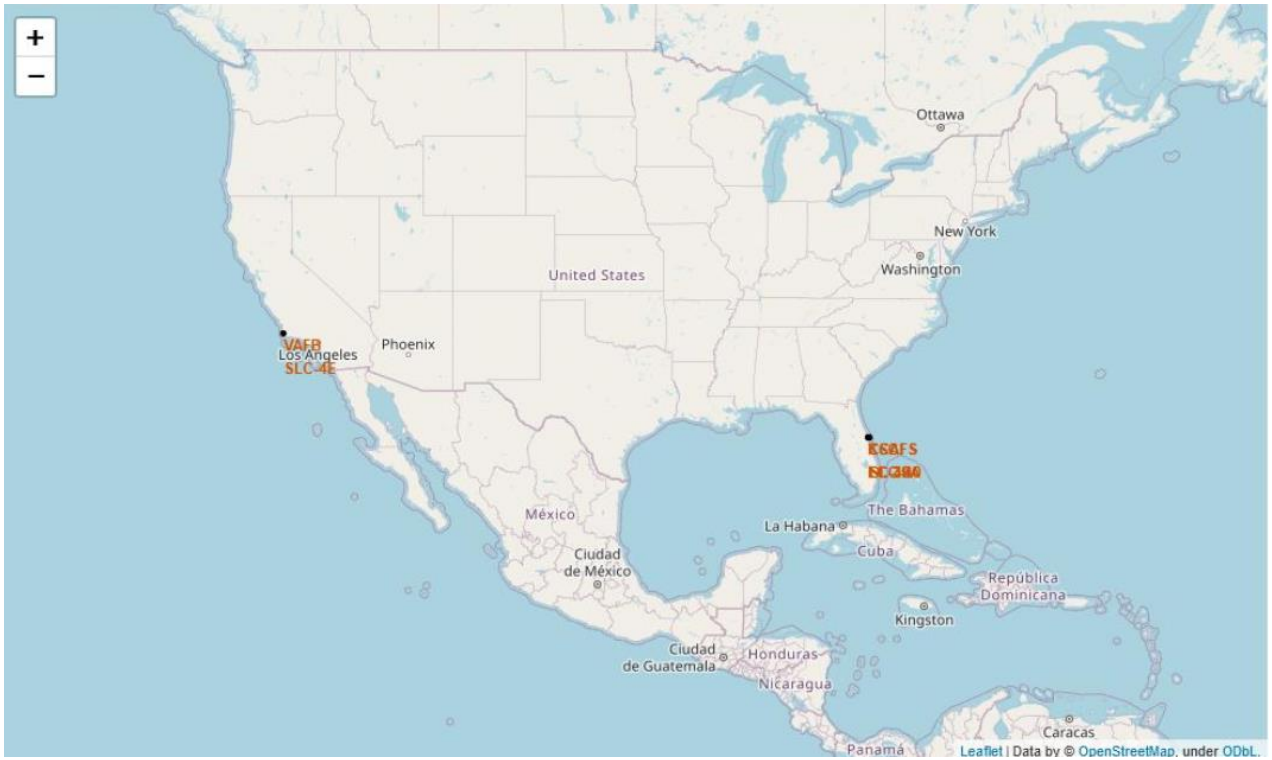
- Query using the count and filter the data as for date between and then used group by and order by.
- Also need to be aware of no attempt.

Section 3

Launch Sites Proximities Analysis

All Launch Sites

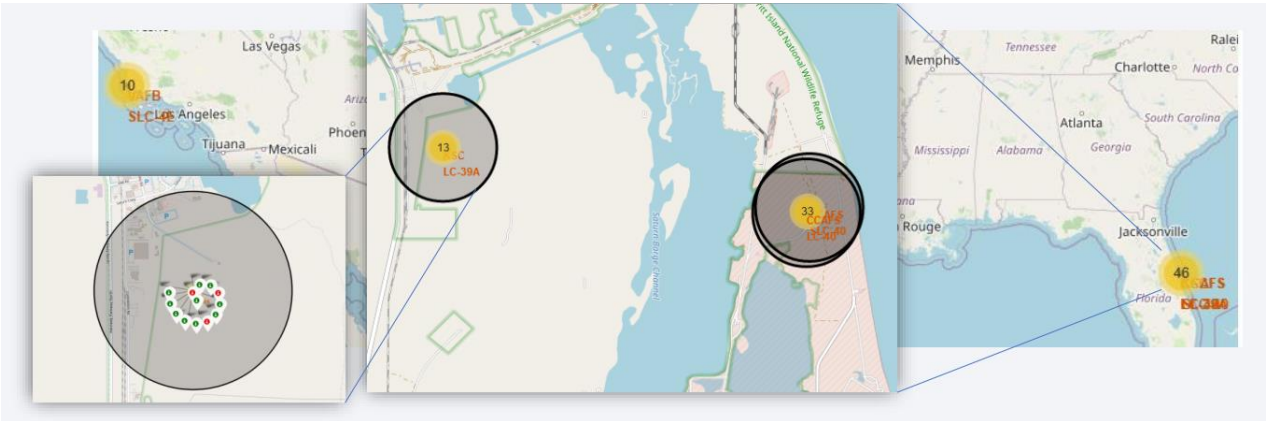
Launch sites are costal area.



Launch outcome by sites

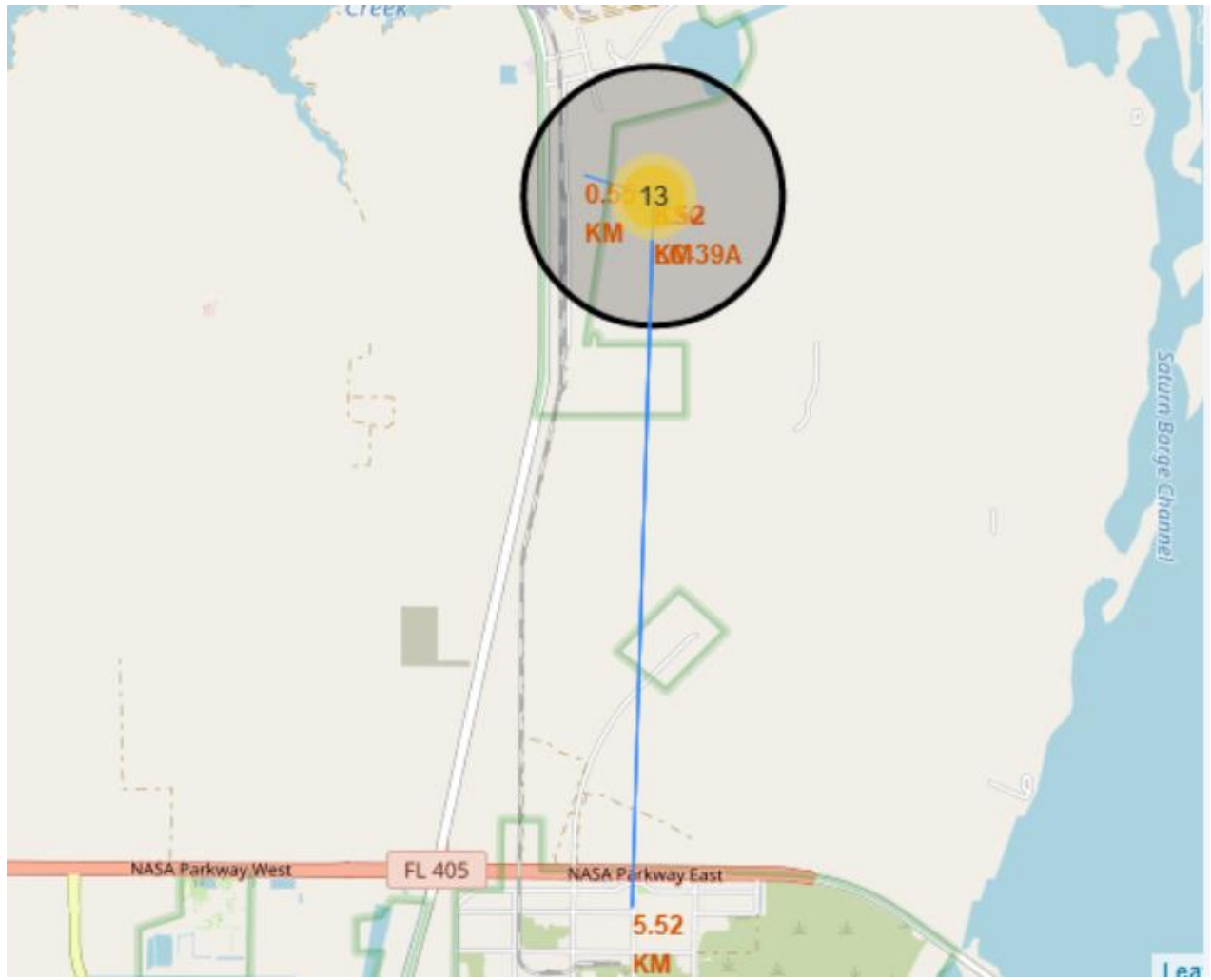
Green indicate success

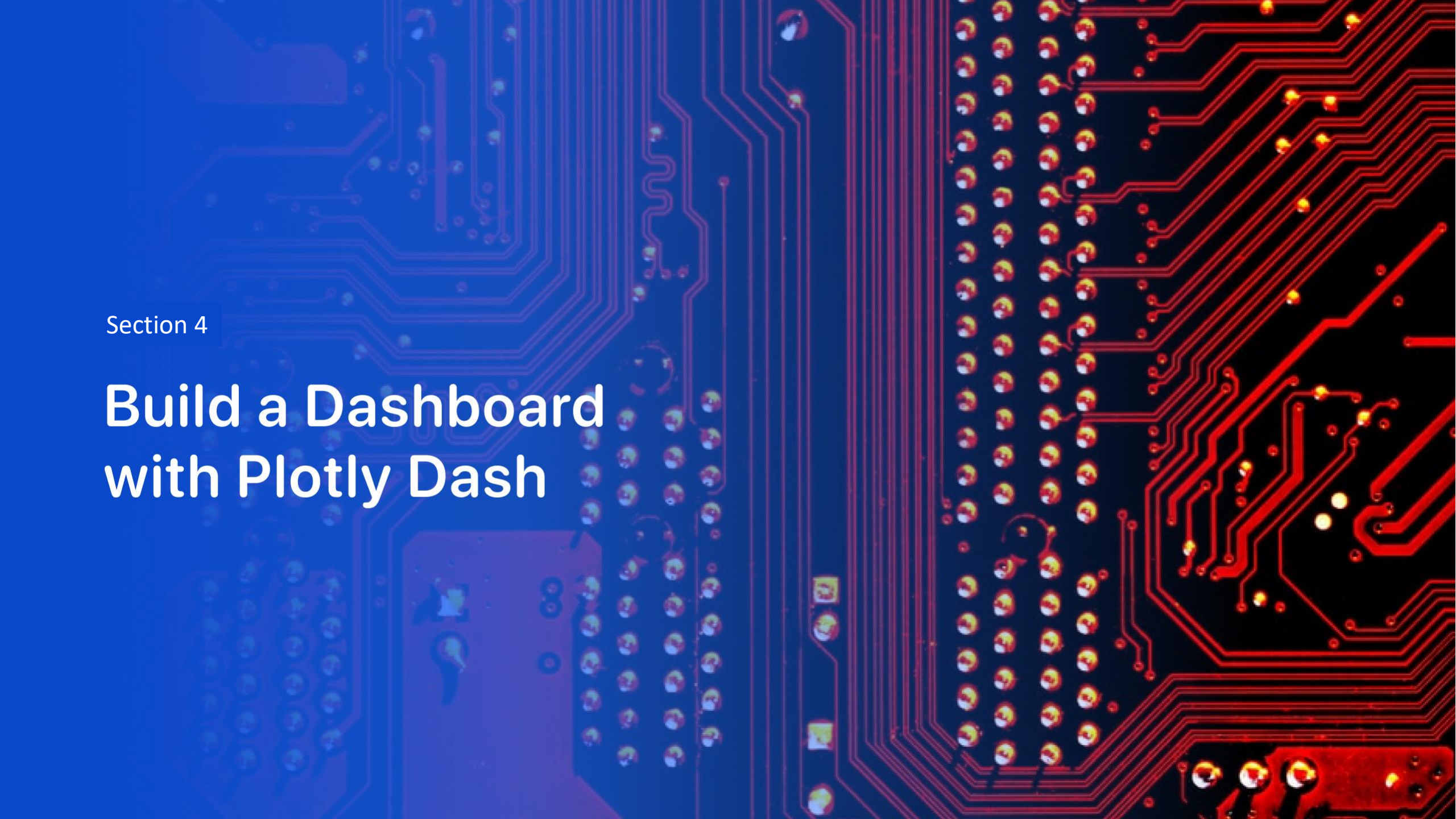
Red indicate failure



Logistic and Safety

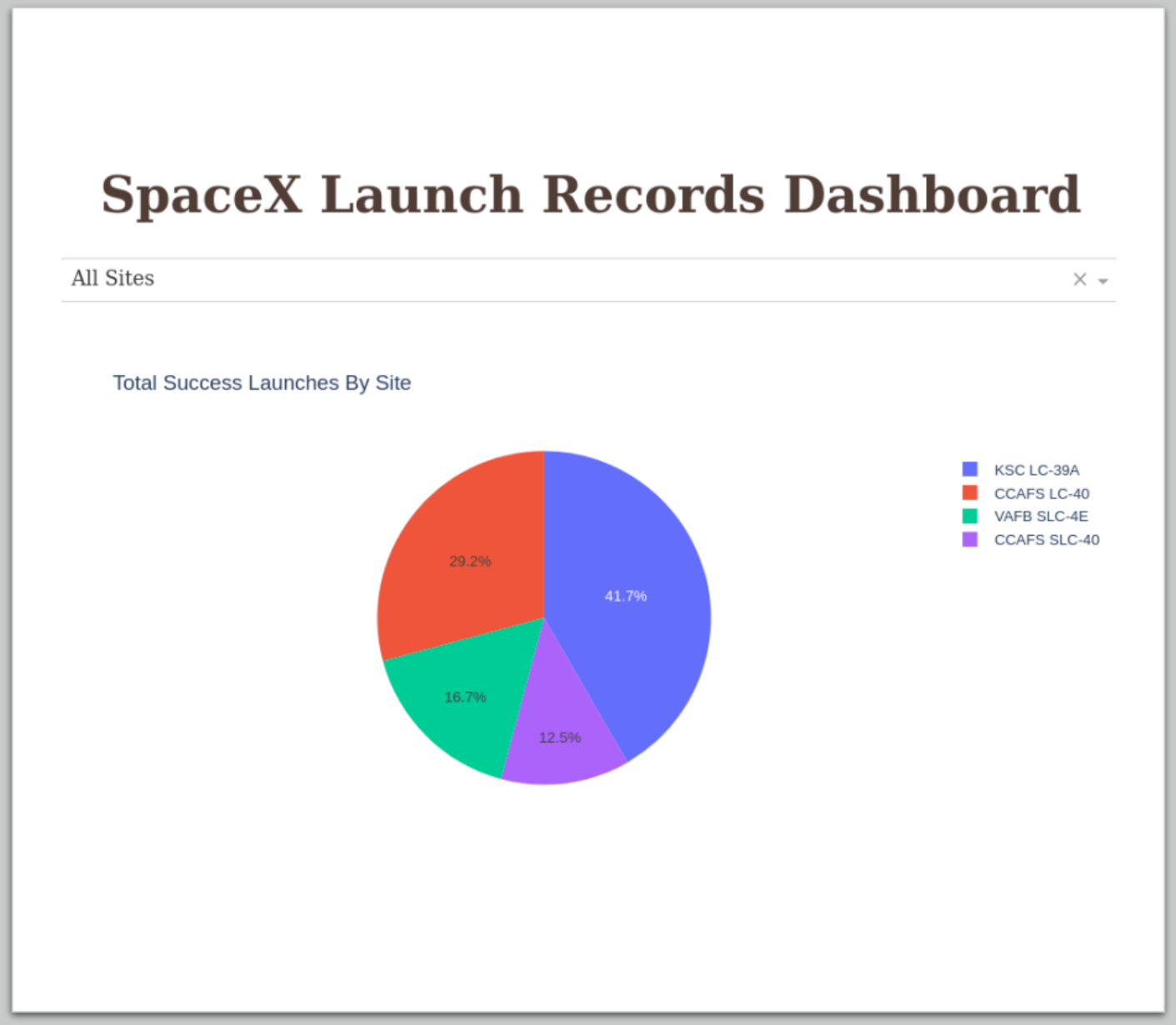
KSC LC 39A has good logistic




Section 4
Build a Dashboard with Plotly Dash

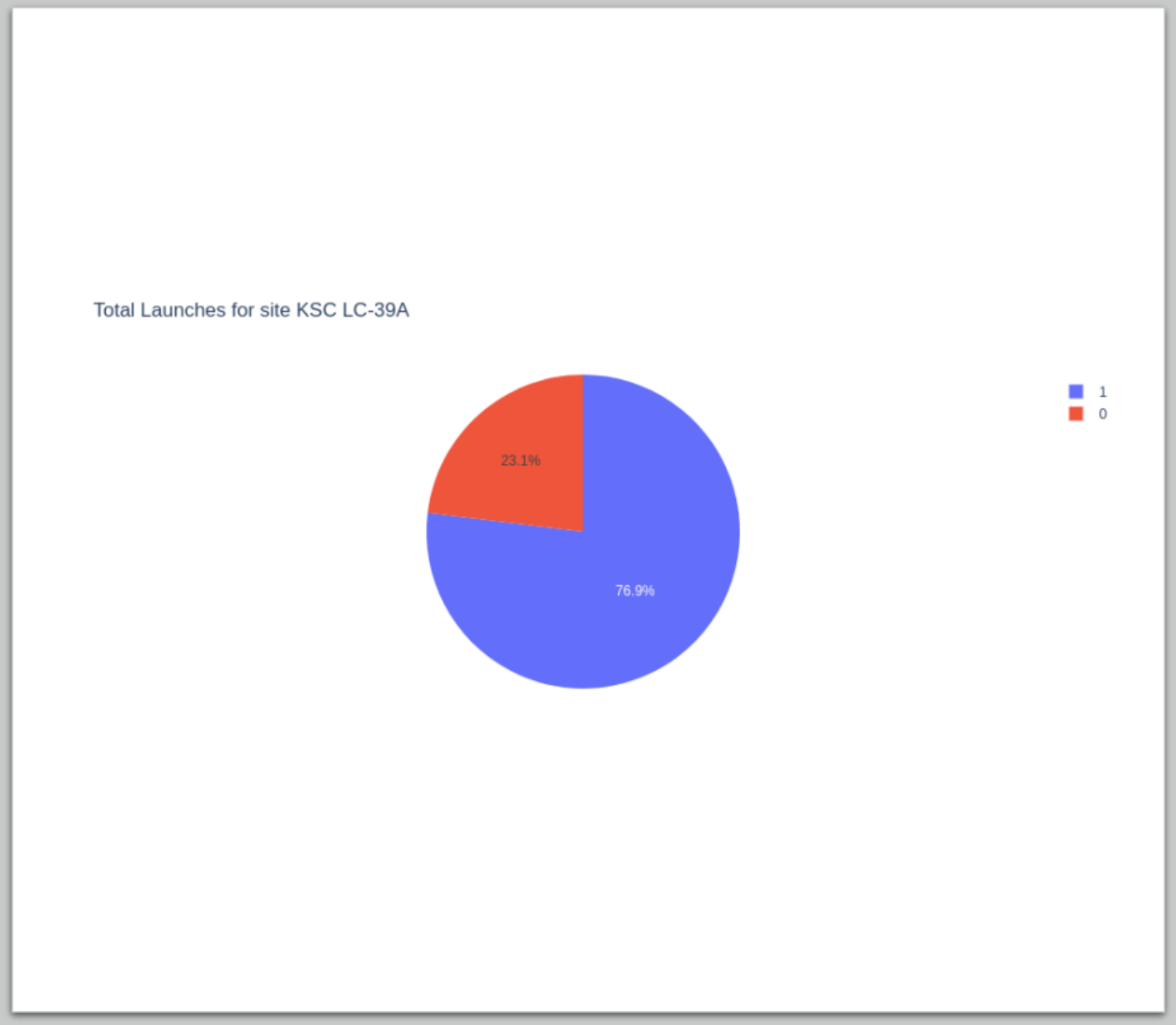
Successful Launches by Sites

- It is important to see launch sites that has success rate



Launch success pie chart

KSC has more success rate than failure.



Payload vs Launch Outcome

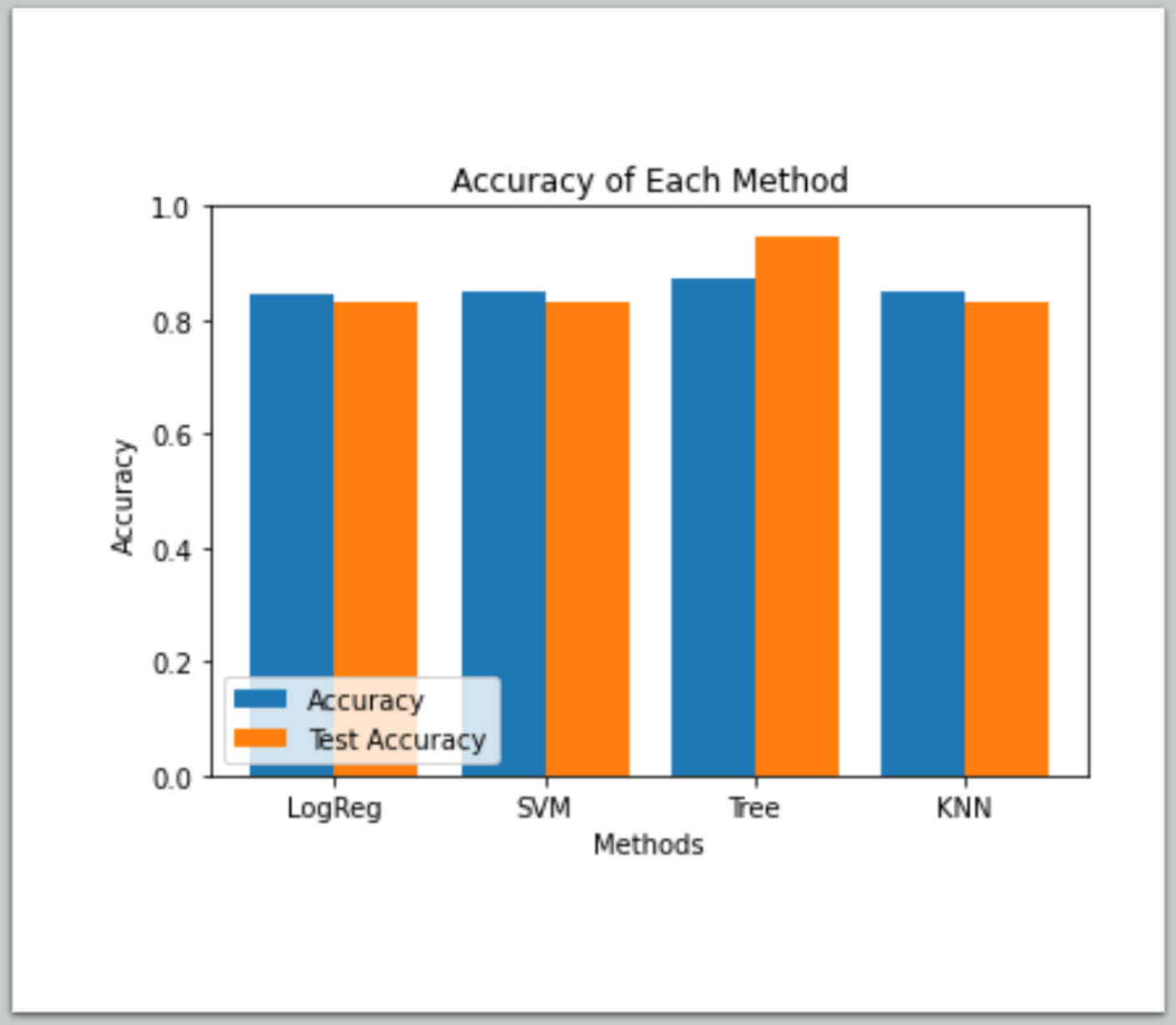
- Payload under 6000kg & FT Boosters has good success rate

Section 5

Predictive Analysis (Classification)

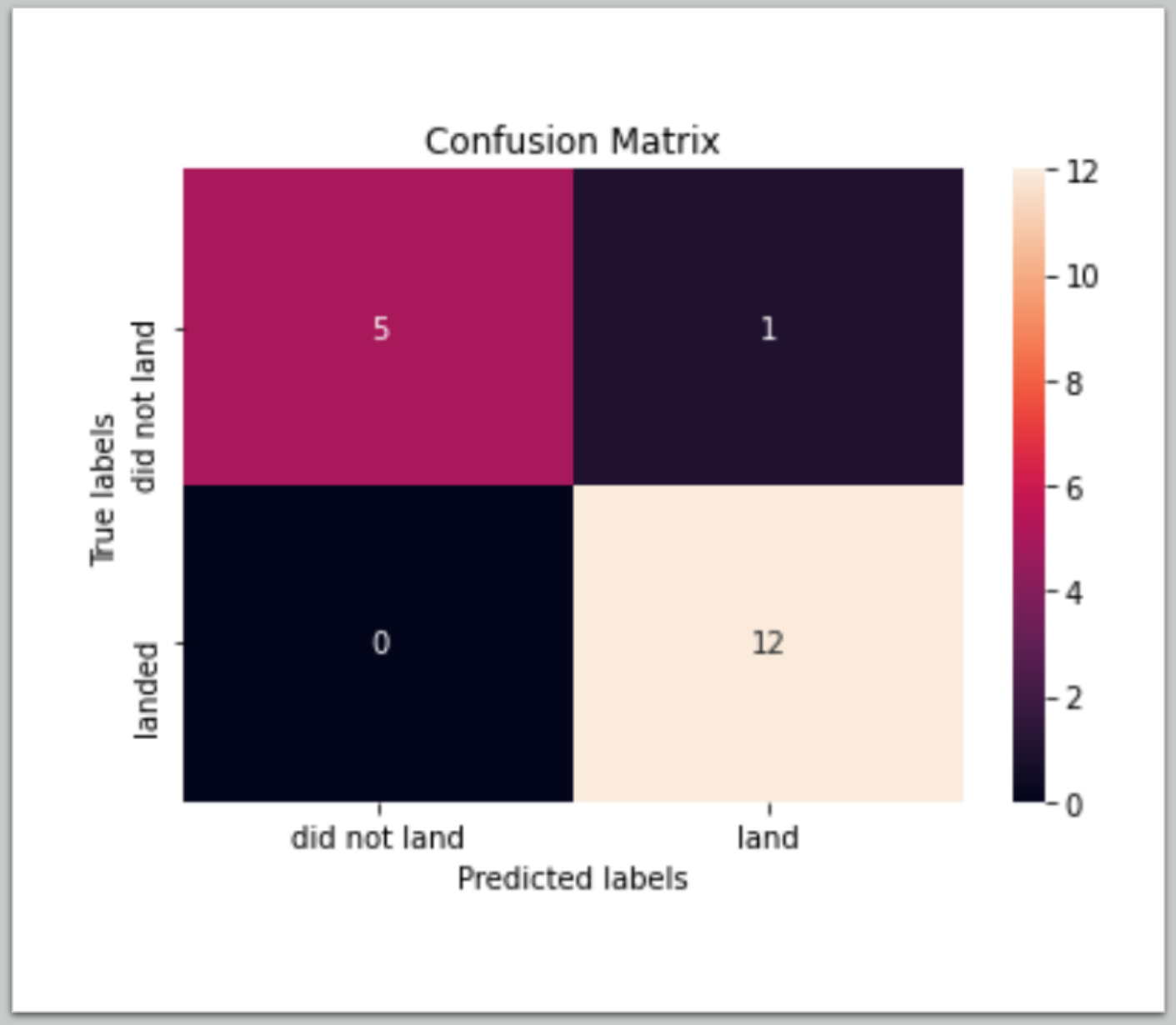
Classification Accuracy

- Decision tree has the best accuracy with above 87%



Confusion Matrix

- The confusion matrix of the best performing model is decision tree.
- Big number in Land and did not land.



Conclusions

- Different data sources were analyzed and refined to get conclusion
- Best launch site is KSC LC 39A
- Launches with payload of 7000kg has good success rate.
- As year goes by, the success rate increases.
- Decision tree classifier can predict well for successful landing and can be used for future models.

Appendix

- IBM account has reached to maximum runtime within a month. Hence, remaining works are done on jupyter notebook.
- Some codes are difficult and would prefer to get some more exercise.
- This PDF and presentation is too long. No senior executives would read it.

Thank you!