IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nikhil Anil Prakash
May 2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data of Space X can be obtained two sources:
    - SpaceX API(https://api.spacexdata.com/v4/launches/past)
    - Webscraping(https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The collected data was modified by introducing a landing outcome label based on outcome data.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.
- Perform predictive analysis using classification models.

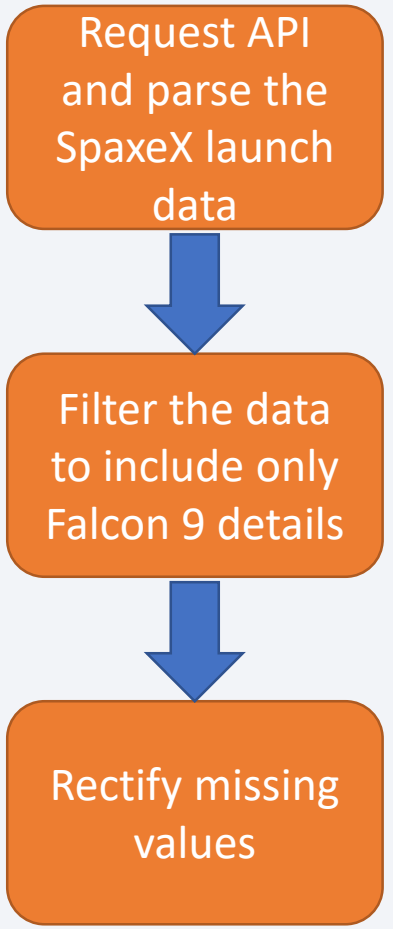
Data Collection

- Datasets were collected from Space X API (https://api.spacexdata.com/v4/launches/past)
- And from Wikipedia using webscraping https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- Space X offers a public API from which everyone can access and obtain data

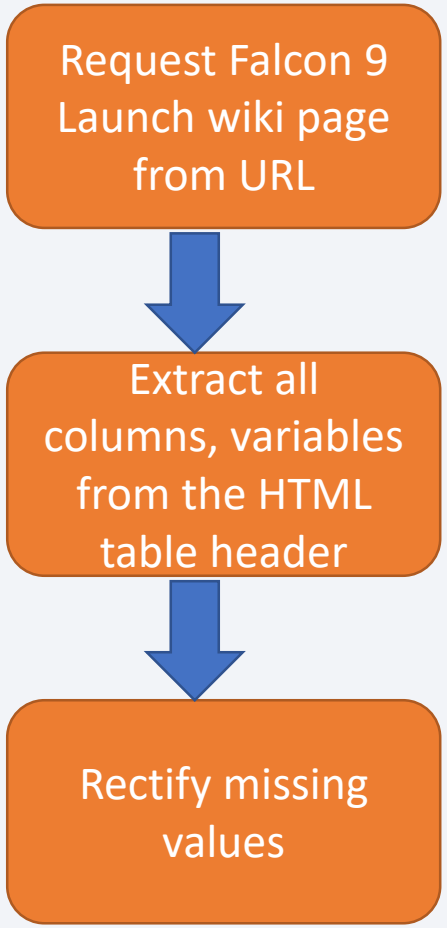
- GitHub URL for reference: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- In web scraping, the data is collected from the wiki page provided https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- GitHub URL for reference: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Initially exploratory data analysis(EDA) was performed on the obtained dataset.
- Then the summaries of launches per site, number and occurances of mission outcome for each orbit were found.
- Then, the landing outcome label was created from outcome.

GitHub link: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To find out how attributes in our dataset are related, data visualization was performed using pandas, matplotlib and seaborn.
- Scatter plots, bar charts and line plots were plotted to express the above visualization.
- GitHub reference: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- EDA with SQL were performed using the following queries:
  - Unique launch sites in the mission
  - Top 5 launch sites whose name starts with ‘CCA’
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - The date when the first succesful landing outcome in ground pad was acheived.
  - The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - The total number of successful and failure mission outcomes
  - The names of the booster_versions which have carried the maximum payload mass.
  - The records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - The count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- GitHub Link: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, Circles, lines and marker clusters were used in Folium to create an interactive map.
  - Markers indicate launch sites.
  - Circles indicate areas highlighted around specific coordinates.
  - Marker clusters indicate groups of events.
  - Lines to indicate distance between two coordinates.

GitHub link: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/lab_jupyter_launch_site_location.ipynb

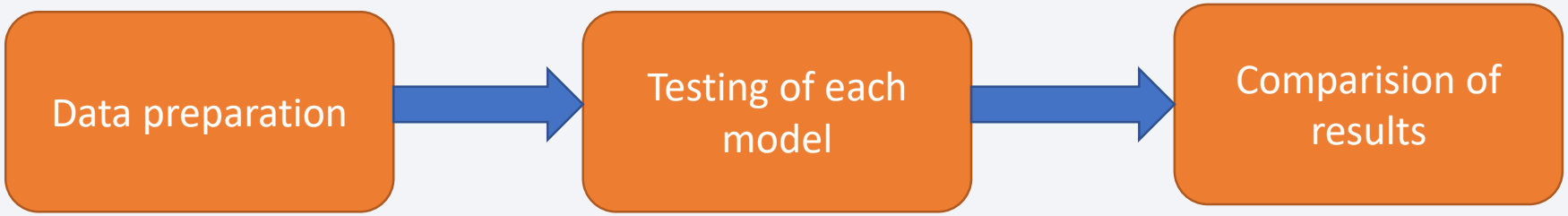
Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches
  - Payload
- This resulted in quick analysis of relationship between payloads and launch sites, identifying which is the best place to launch those payloads.

- GitHub link: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were performed: logistic regression, support vector machines, decision tree and k nearest neighbours.
- GitHub link: https://github.com/NikAP9260/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



Results

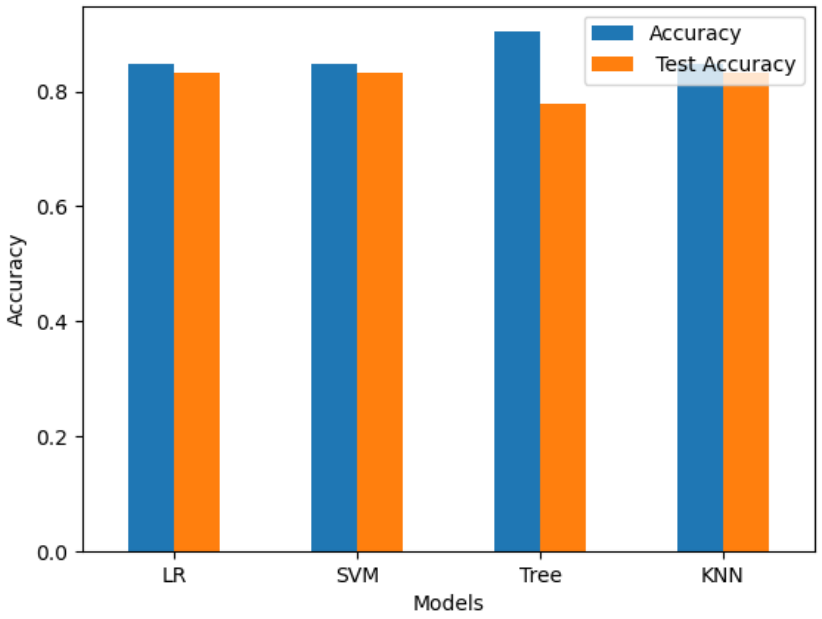
- Exploratory data analysis results
  - SpaceX uses 4 different launch pads.
  - The first launches were at SpaceX and NASA.
  - The average payload of F9 v1.1 booster is 2928.4kg
  - There was 1 mission failure, and almost a 100 success missions.

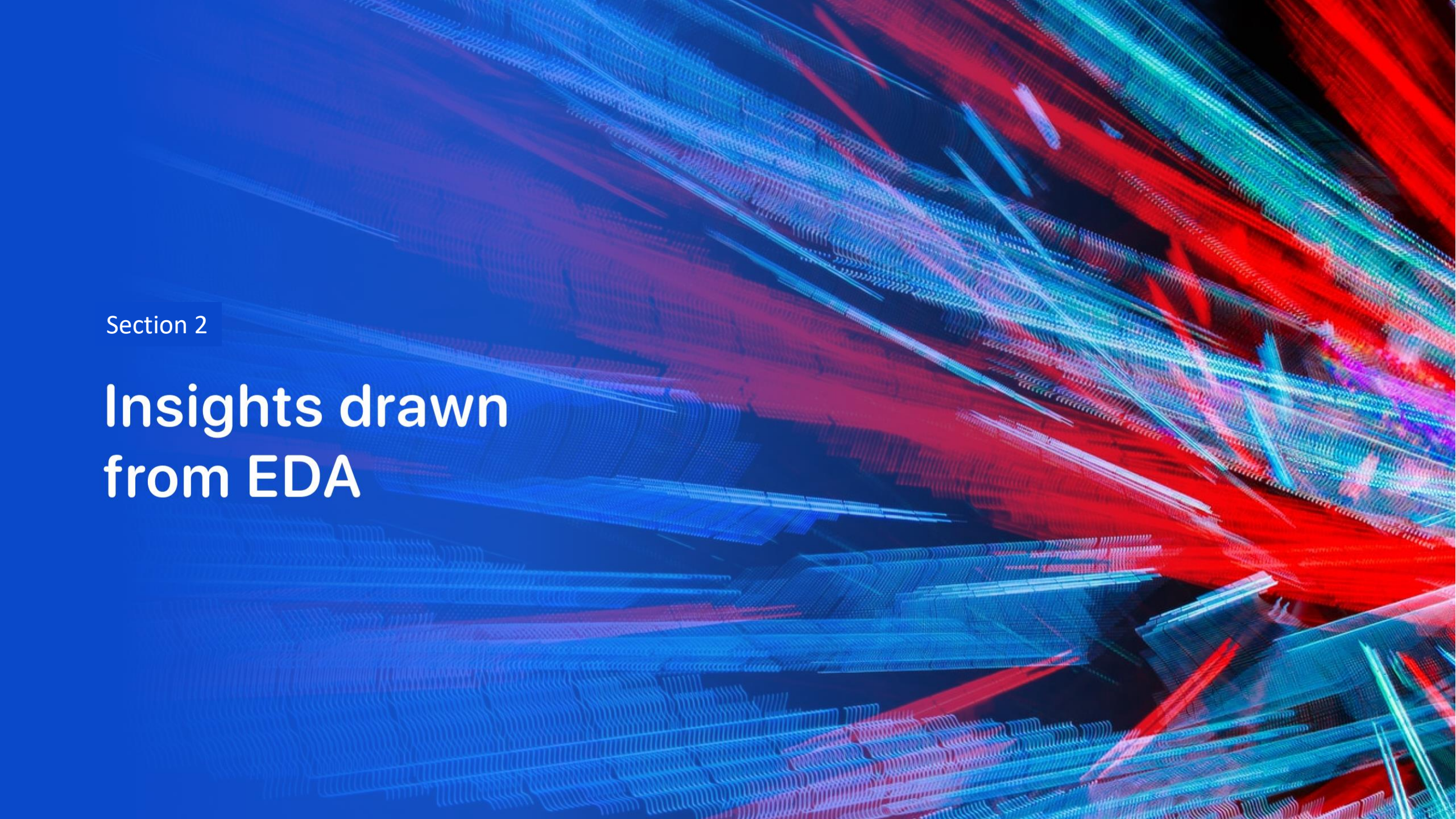
Results

- Interactive analytics demo in screenshots
  - A few results obtained on using folium maps to represent the data is shown below.

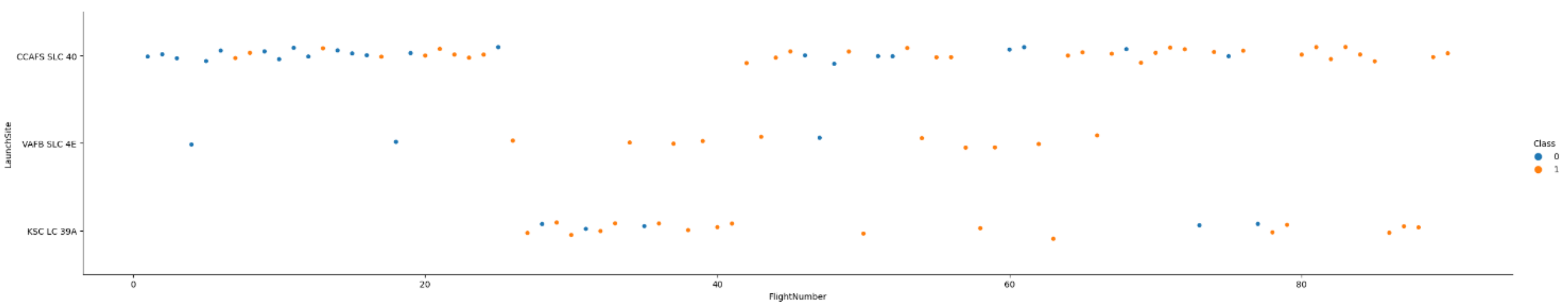
Results

- Predictive analysis results
  - Scores from the four different models were implemented and compared.



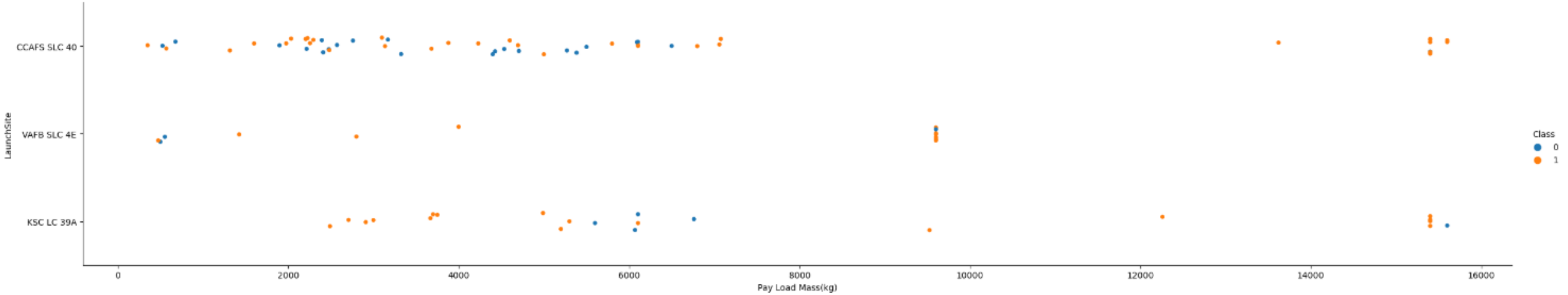

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- The best launch site according to the plot above is CCAF5 SLC 40.
- In second place VAFB SLC 4E, in third place KSC LC 39A

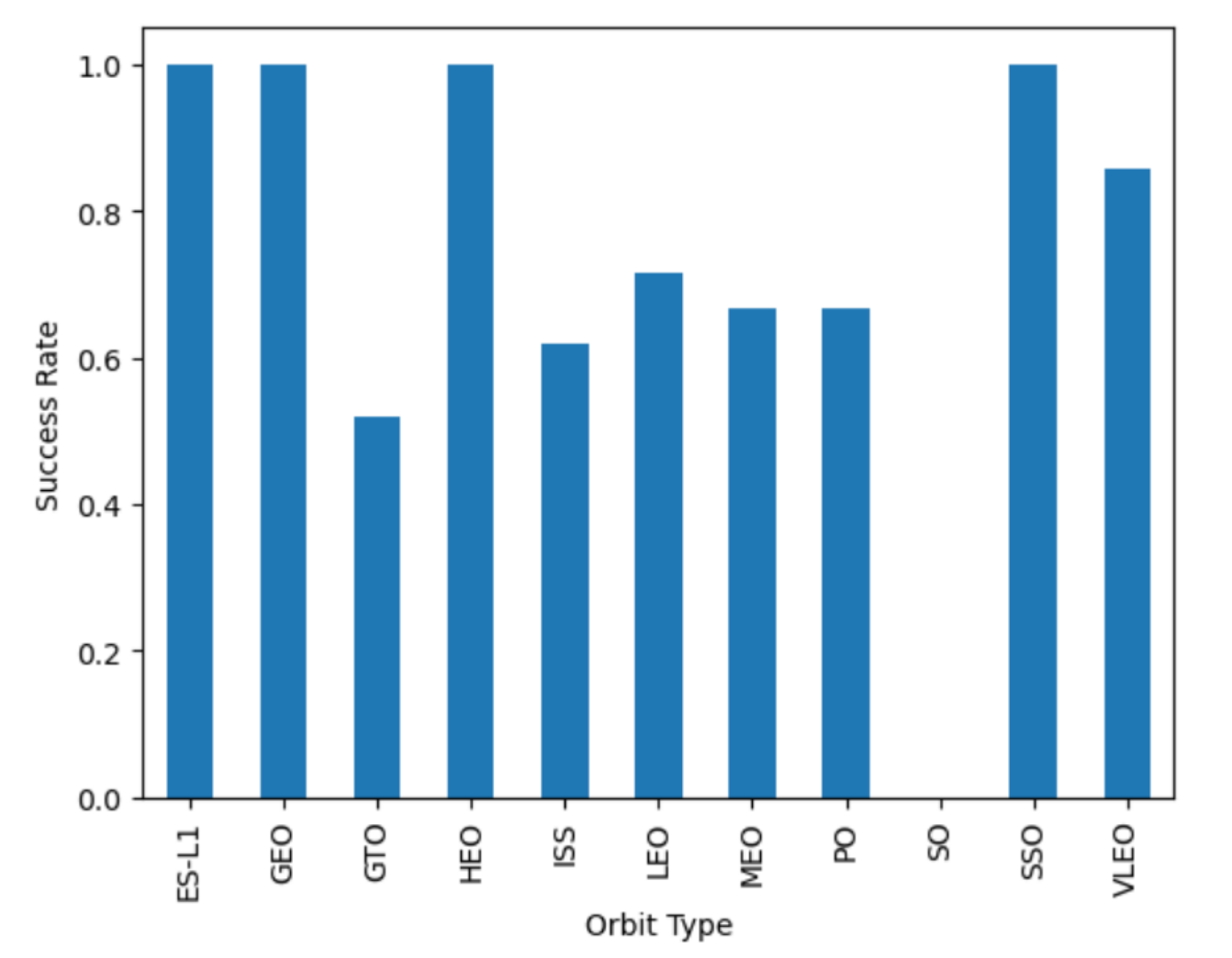
Payload vs. Launch Site



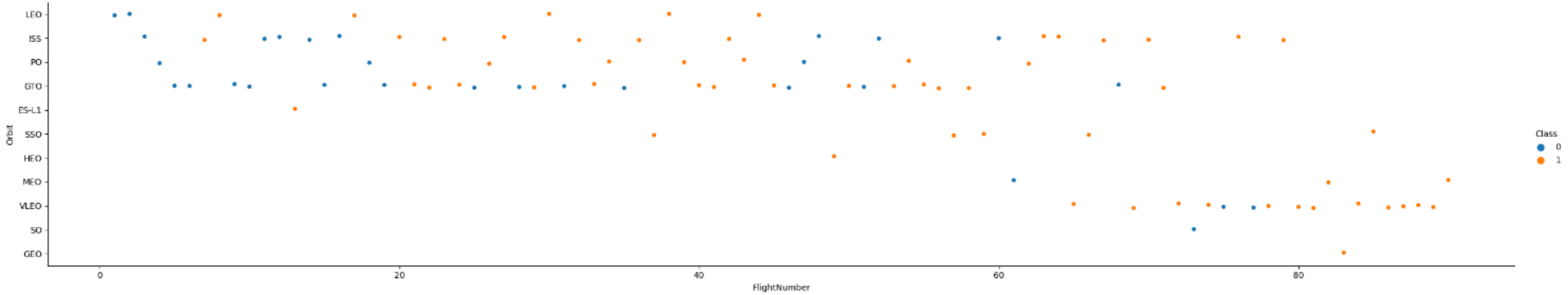
- Excellent launch rate for payload above 9000 kg at the launch site CCAF5 SLC 40.
- Whereas there is no launch at the VAFB SLC 4E for payloads above 10000 kg.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO orbit types have the highest success rate of 1.
- VLEO orbit types has a success rates above 80% and less than 100%.

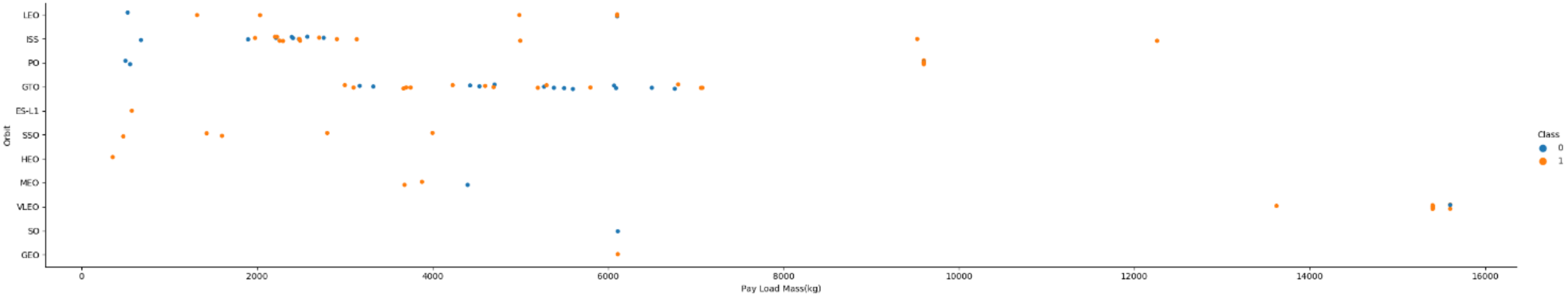


Flight Number vs. Orbit Type



- It can be inferred that in 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.

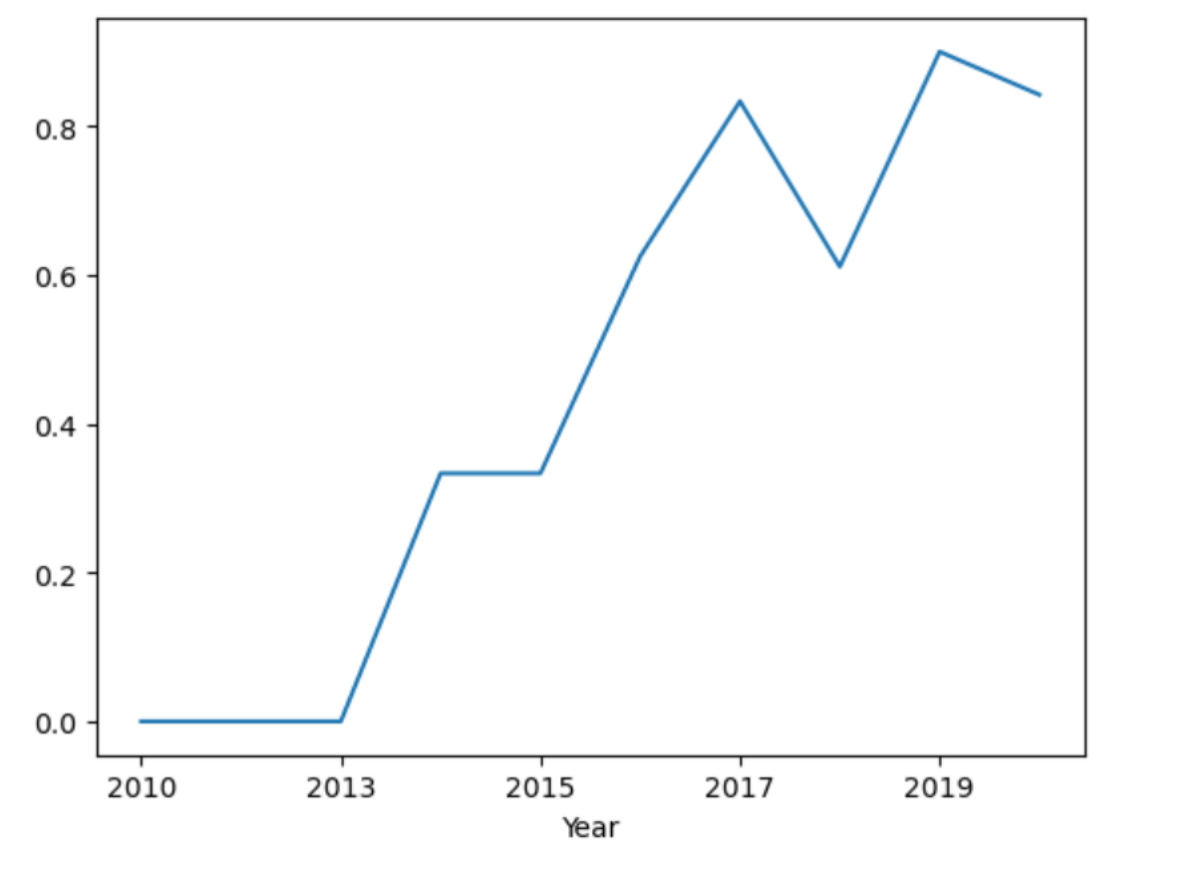
Payload vs. Orbit Type



- 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.
- There are few launches to SO and GEO.

Launch Success Yearly Trend

- We can observe that the success rate since 2013 has been increasing till 2020



All Launch Site Names

- According to the data:

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

- There are four unique launch sites as shown in the screenshot above.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA` are:

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

- The total payload carried by boosters from NASA:

TOTAL_PAYLOAD
111268

- The total payload was calculated from codes which contained CRS which belongs to NASA.

Average Payload Mass by F9 v1.1

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

AVG_PAYLOAD
2928.4

- Filtering data by booster versions, we were able to obtained the average payload mass carried by the booster version F9 v1.1.

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad

Min Date
2015-12-22

- Filtering the data, lead us to a date mentioned above as the day in which the first successful landing outcome on ground pad took place.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List 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

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	98
Success	1
Success (payload status unclear)	1

- Grouping mission outcomes lead us to the above results.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass are shown on the right.

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

- List 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

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

- Ranking of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20:

Landing Outcome	Occurances
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

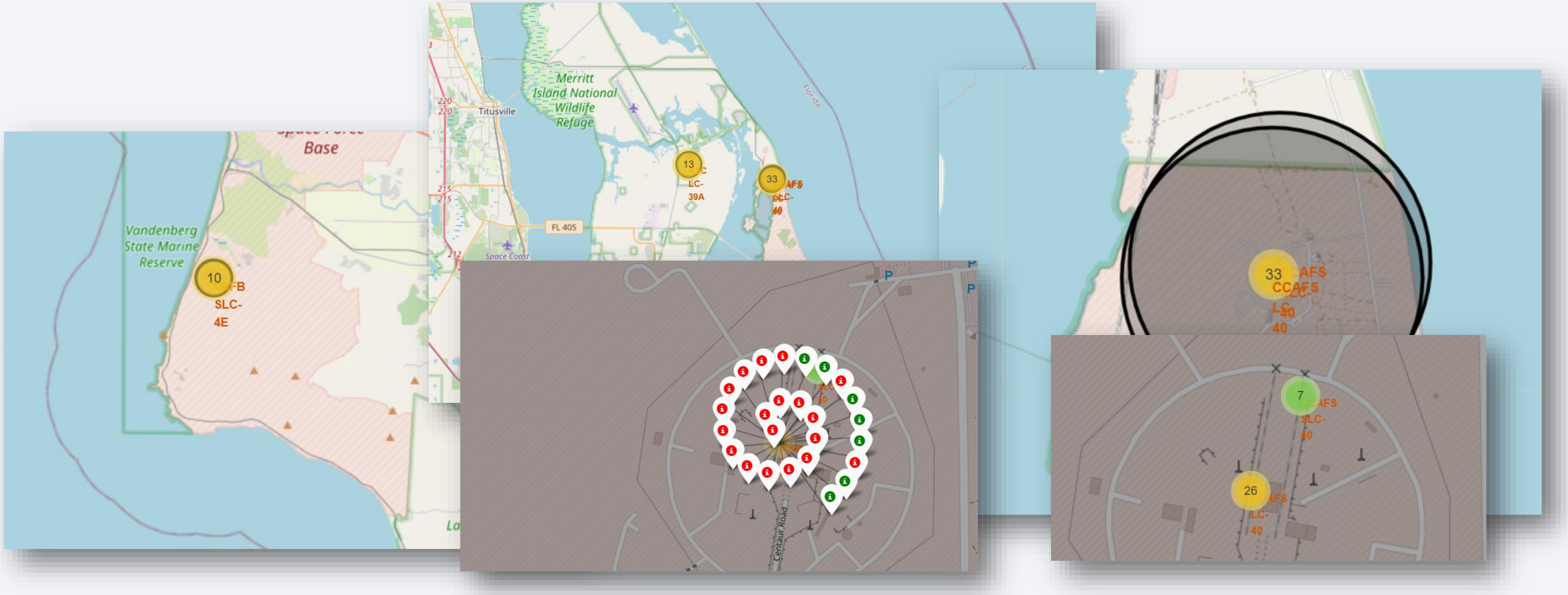
Section 3

Launch Sites Proximities Analysis

All Launch Sites

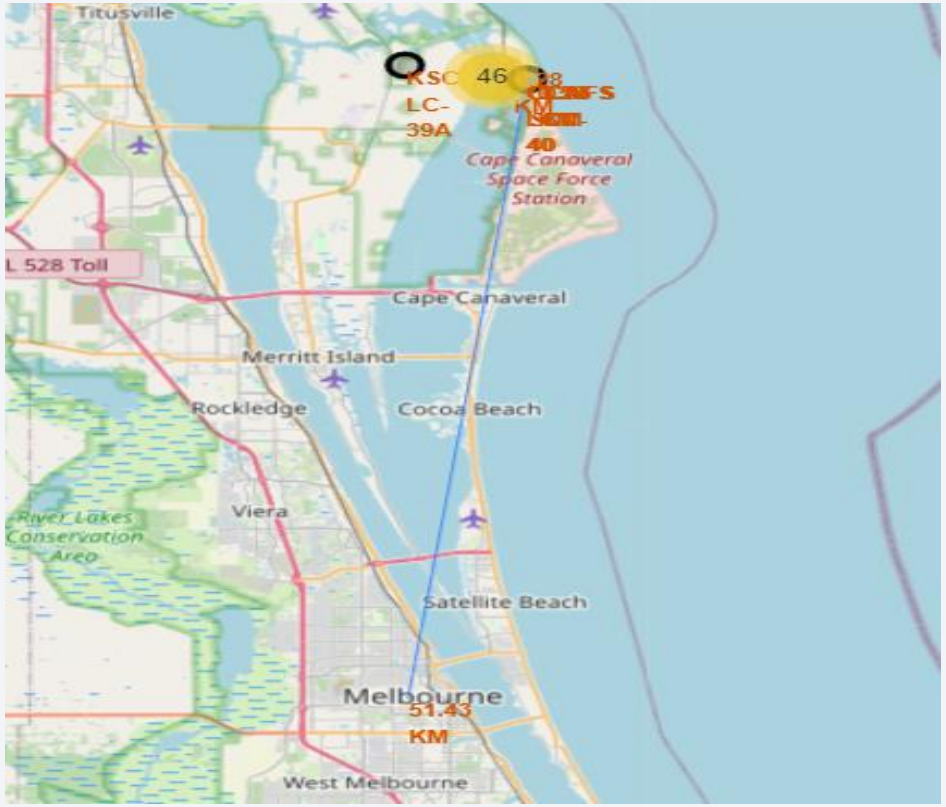
- Launch sites are near the ocean, probably due to safety reasons.

Launch Outcomes on the Map

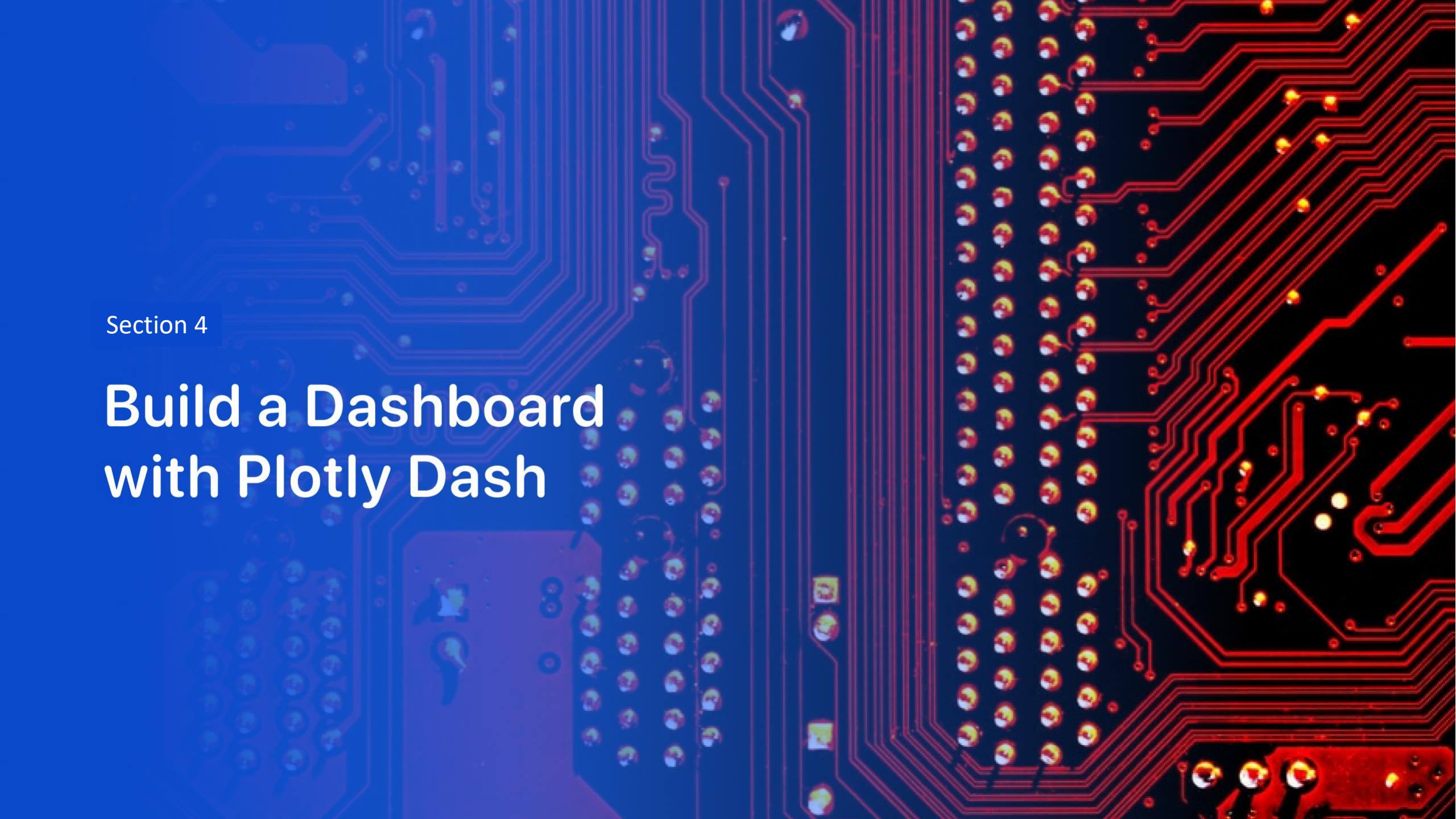


- Green markers indicate success. Red indicate failure.

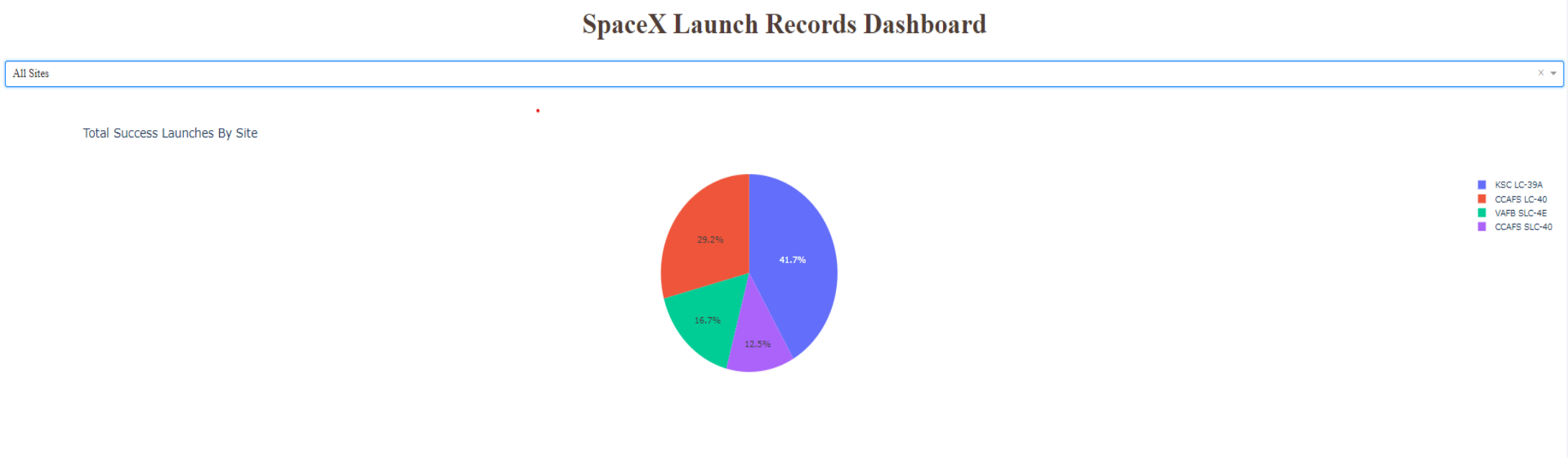
Proximity to Public Facilities



- The distance to public facilities such as roads, railways coastline also play an important role.

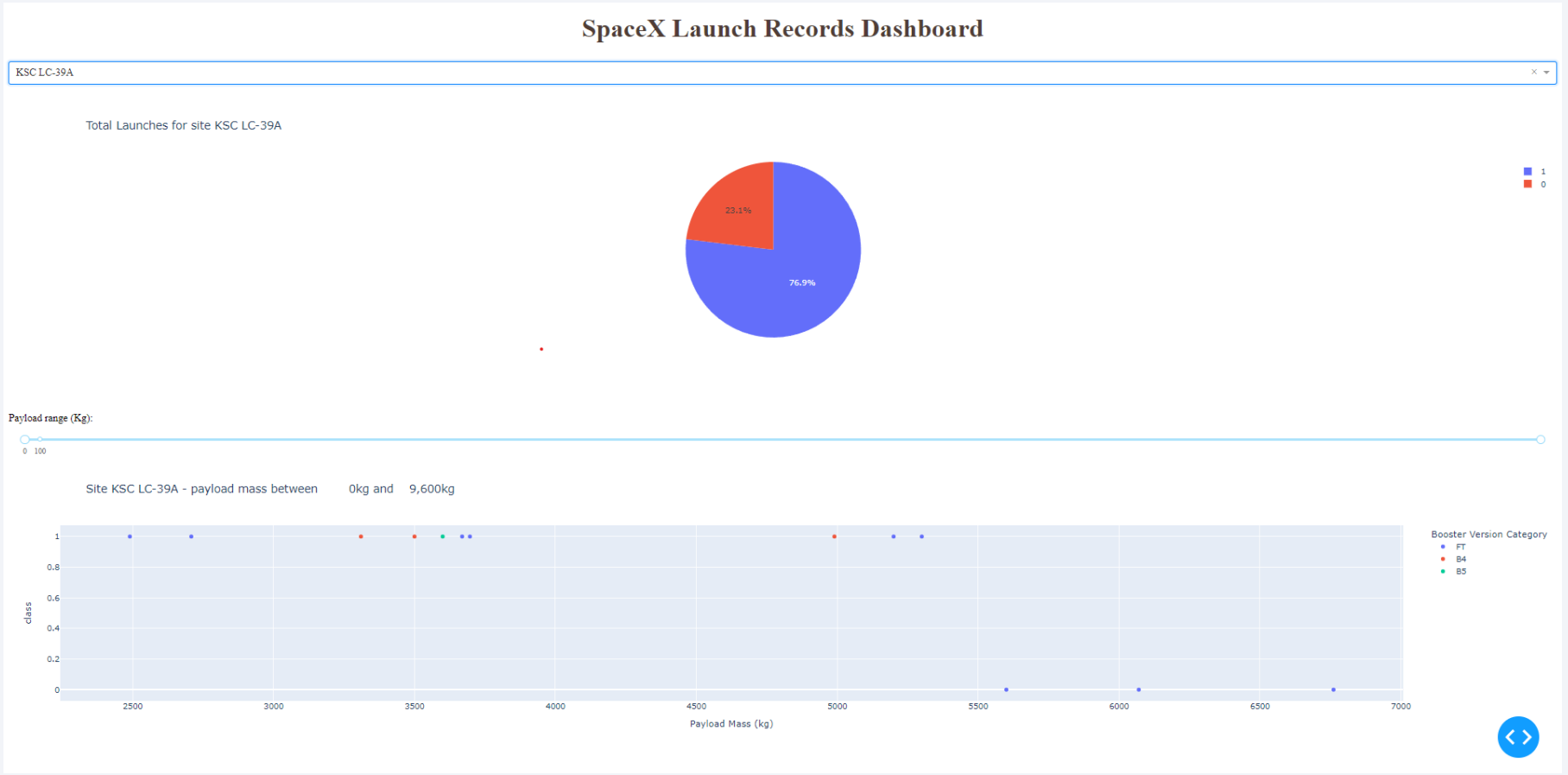

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



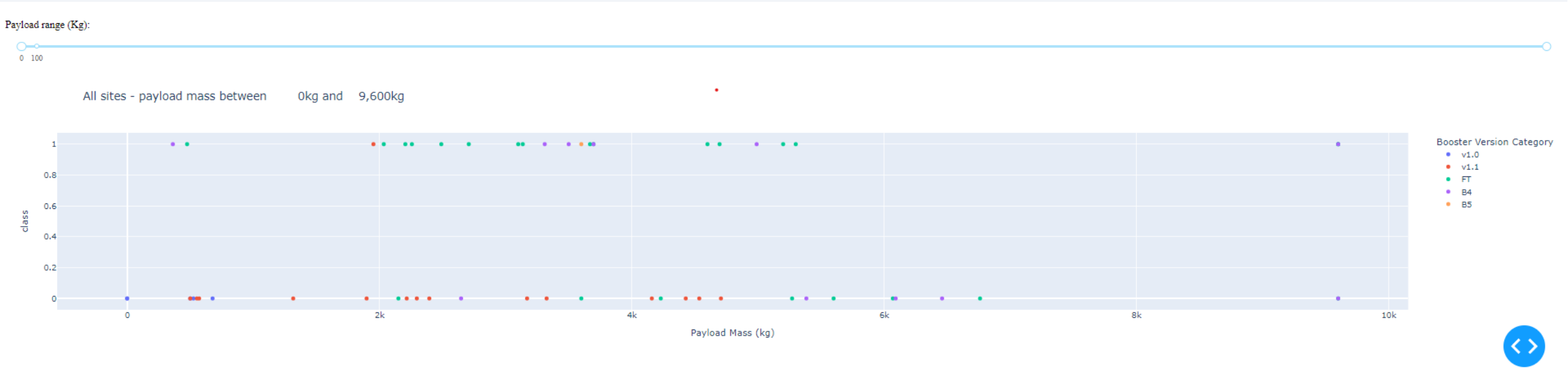
- Launch sites are shown to be an important factor.

Launch Site with Highest Launch to Success Ratio



- KSC LC 39A is shown to have highest success rate of 76.8%

Payload vs Launch Outcome



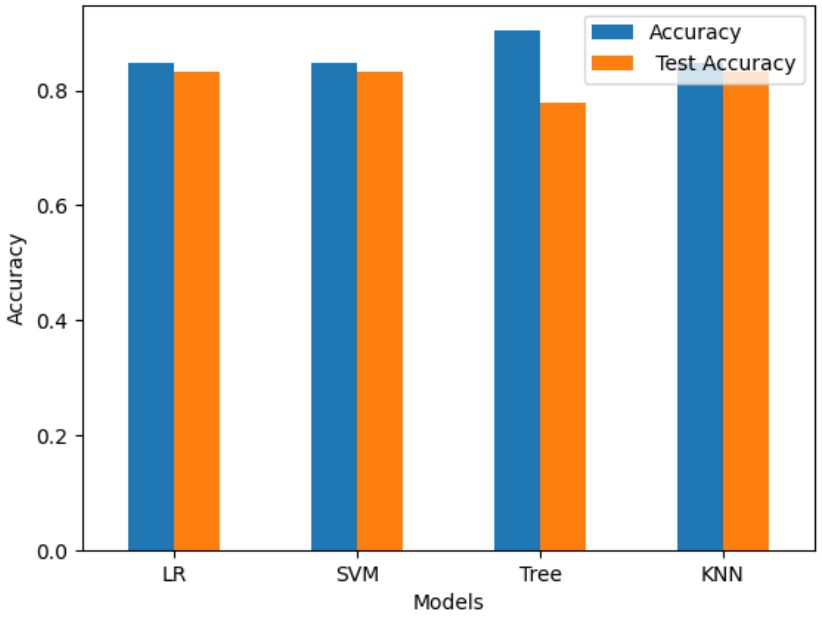
- Payload under 6000 kgs and FT boosters are a good combination for higher success rate.

Section 5

Predictive Analysis (Classification)

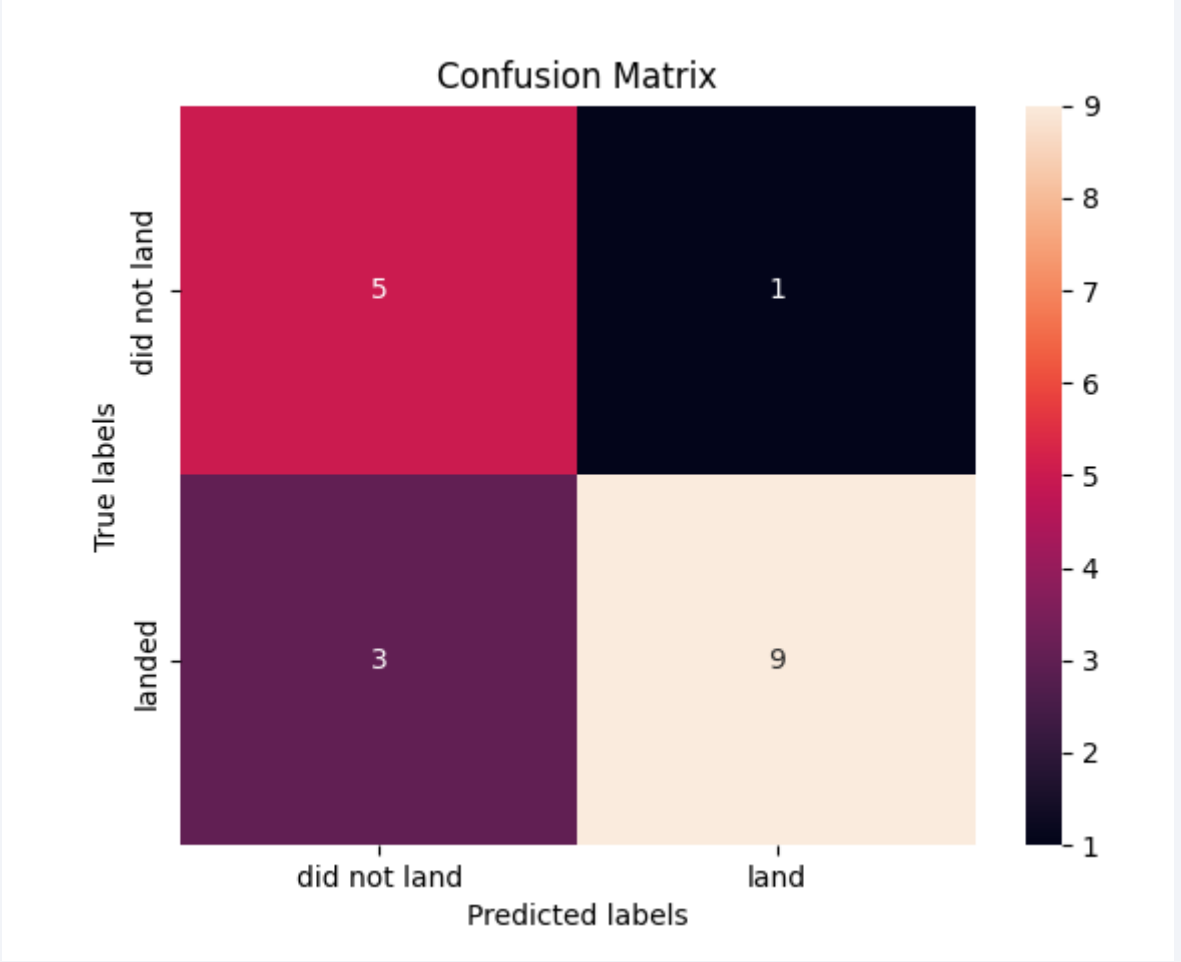
Classification Accuracy

- The classification models were implemented and visualized.
- The test scores for each model was calculated and compared.
- Decision Tree is shown to have a higher accuracy than the rest of the models.



Confusion Matrix

- Decision Tree is shown to have the best accuracy.



Conclusions

- Different methods were applied and refined conclusions were made in the process;
- The best launch site was found to be KSC LC-39A.
- Mission outcomes are mostly successful but successful landing seems to have improved over time.
- Decision tree can be used to help increase the chances of future success in regards to launches.

Thank you!