IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Aalian Naviwala>
<11/20/2024>

Outline

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

Executive Summary

- Summary of methodologies: The main programming language used was Python in Jupyter Notebook. Multiple data analysis and data visualization libraries were used matplotlib, seaborn, beautiful soup, etc.
- Summary of all results: The results show the success and failures of spacex rocket launches and the location of launches as well. It also shows the price for each launch and whether or not spacex reused the first stage rocket for that launch. A model was trained to see if a rocket will reuse the first stage. The results are promising, and will be further explained in later slides.

Introduction

- Project background and context: SpaceY founded by Allon Mask would like to compete with SpaceX. Information will be gathered from public sources and formed into dashboards for team members to use.
- Problems you want to find answers: We want to find out the price of each SpaceX launch, the success rate, when the first stage rocket will be reused and if the first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

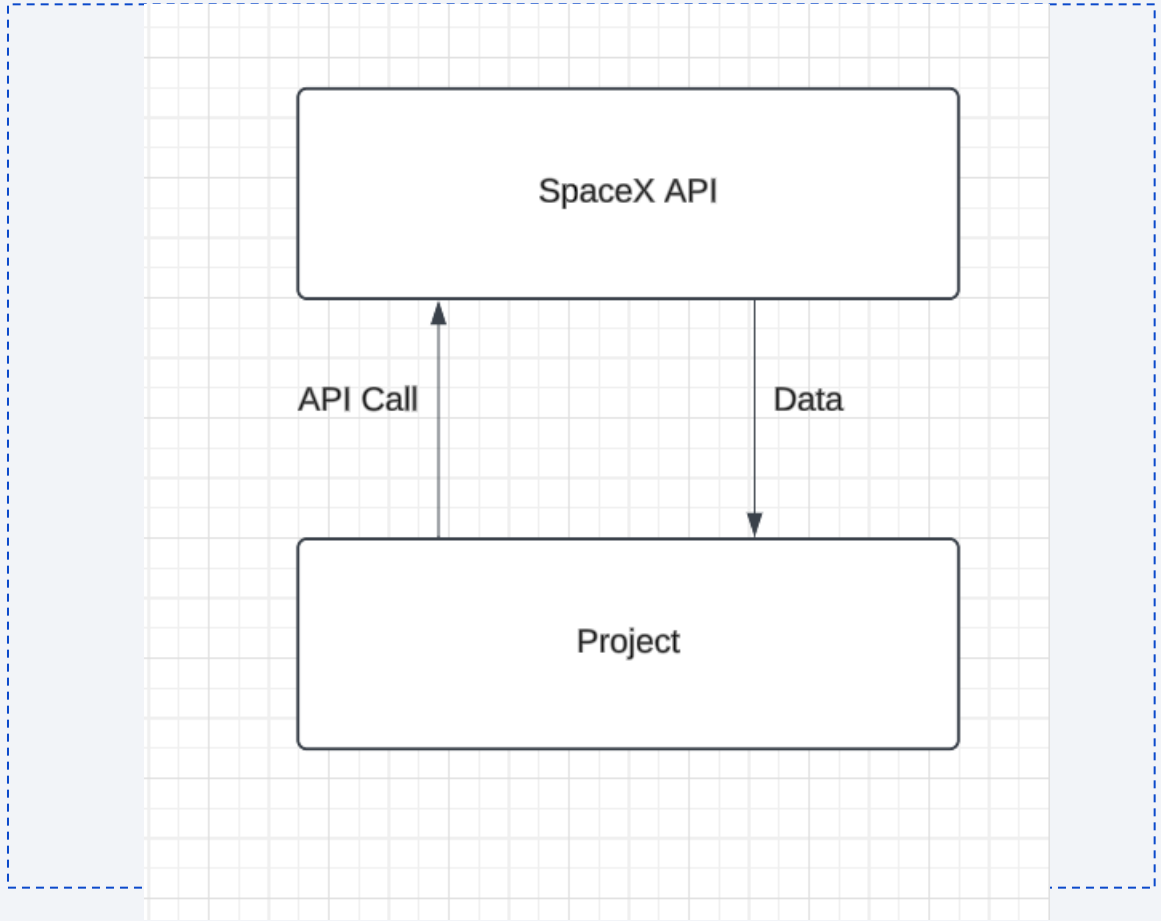
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using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

Data Collection

- Describe how data sets were collected: The data was collected through the SpaceX API. It was organized into charts.
- You need to present your data collection process use key phrases and flowcharts: We first installed the python libraries needed and imported the data from the SpaceX API. We then normalized the data and did a bit of testing to see if everything was working fine. We then dealt with any missing values. Finally, we exported the collected data into a csv file for further use in this project.

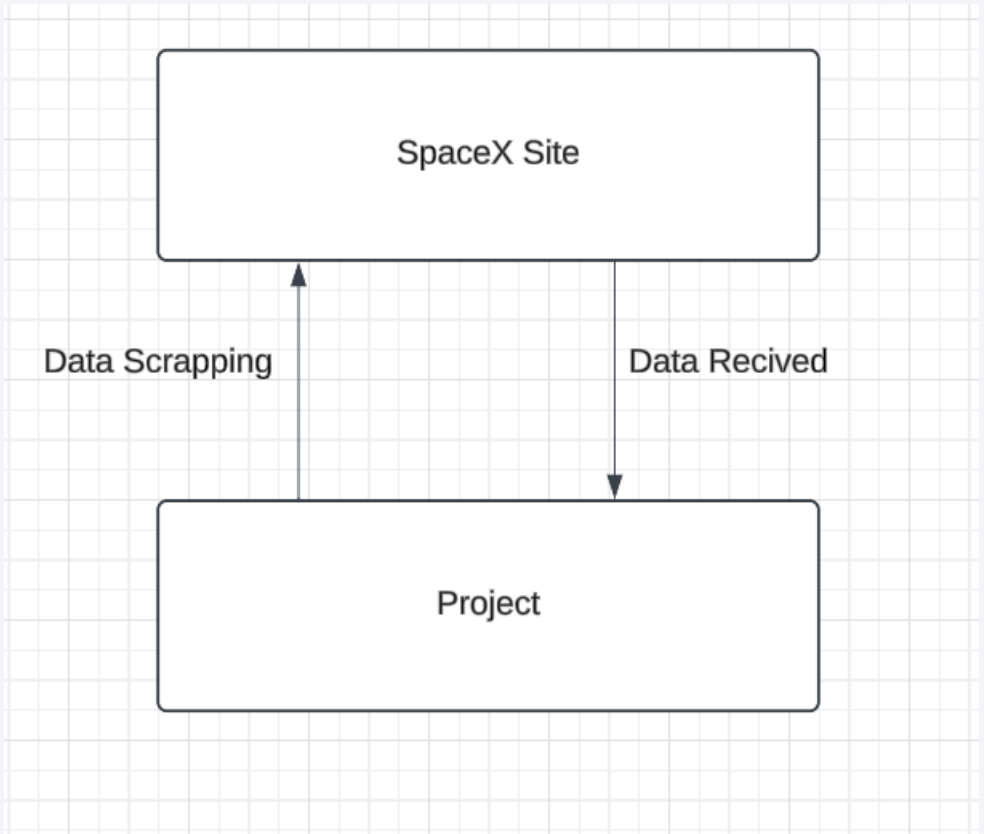
Data Collection – SpaceX API

- Data was collected using the api calls to the spacex api. The flowchart shows the basic ide of it.
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose. The link is below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Using the beautiful soup library, we were able to get data on the falcon 9.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose Link is below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Describe how data were processed: The data was processed using multiple libraries and functions to simplify the data.
- You need to present your data wrangling process using key phrases and flowcharts: The outcome of the data was changed to show whether the booster landed successfully or not.
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose. Link below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts: The charts showed the relationships between different variables and their success rates.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose. Link below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/edadataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- We were able to find the launch sites and what rockets were launched from there. We also found the total payload and the average payload. We also found the amount of successful and failed missions.
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose. Link below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).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: We were able to find the locations of the launch sites. The circles were the launch sites, any nearby coastline, nearby railway, nearby highway, and nearby city. The lines showed their distance to the launch site. The color of the circle on the launch site also shows whether it was a success or a failure.
- Explain why you added those objects: These were to show the results on a map and their relative distance to any nearby populated place.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose. Link below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

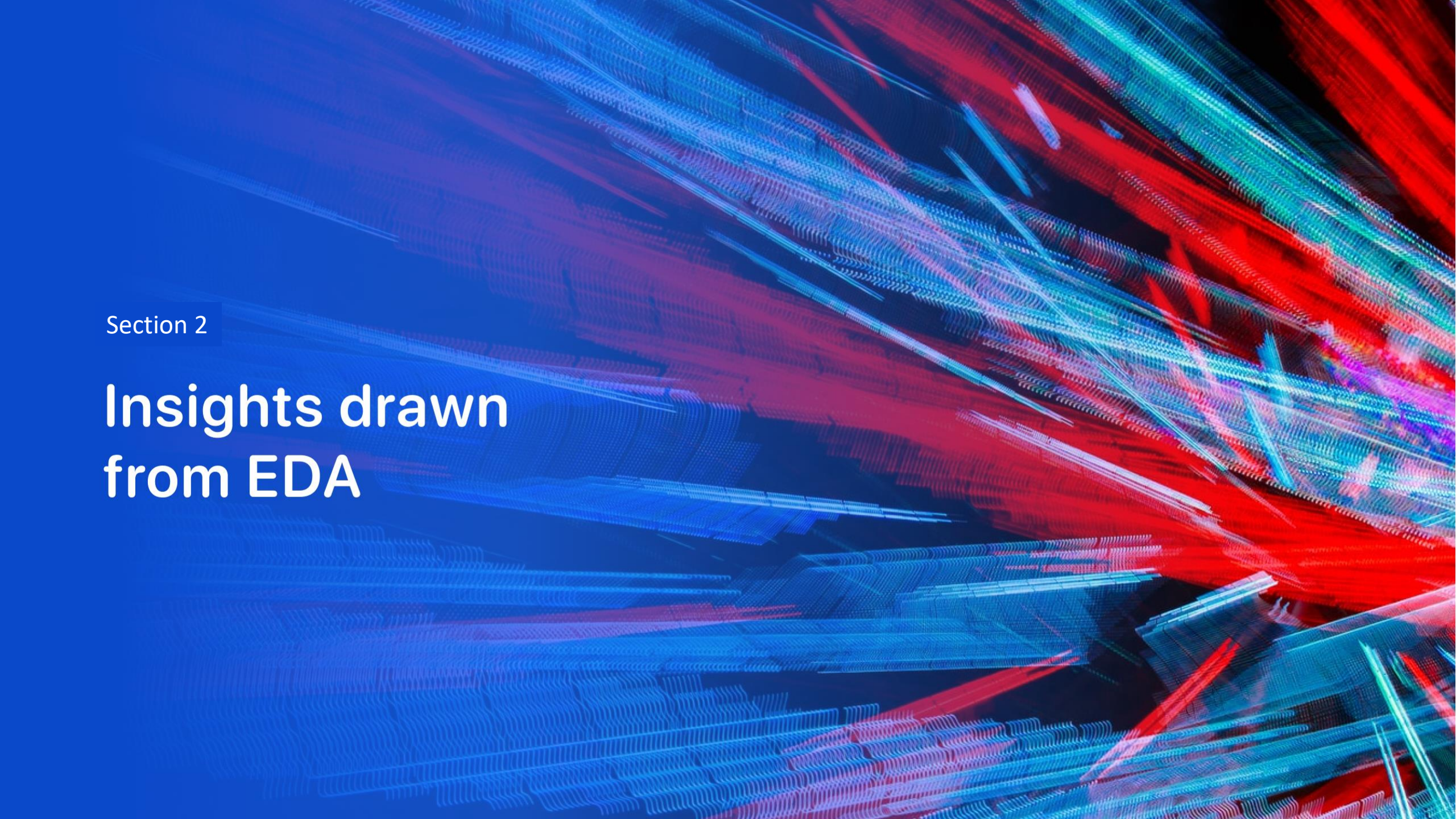
- Summarize what plots/graphs and interactions you have added to a dashboard. The is a pie chart showing the percentage of successful missions at each launch site. There is also a scatter plot showing the payload of each rocket. They both can be altered to show the info for different sites and mass.
- Explain why you added those plots and interactions. These were added to show the data in a visual way.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose. The link will take you to the code file, not the resulting dashboard.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model: The appropriate libraries were installed, and the data was received from its location. The data was split and used to train the model. Parameters were set and the model was trained.
- 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. Link below.
- https://github.com/4PathLeaf/Coursera-Capstone-Project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

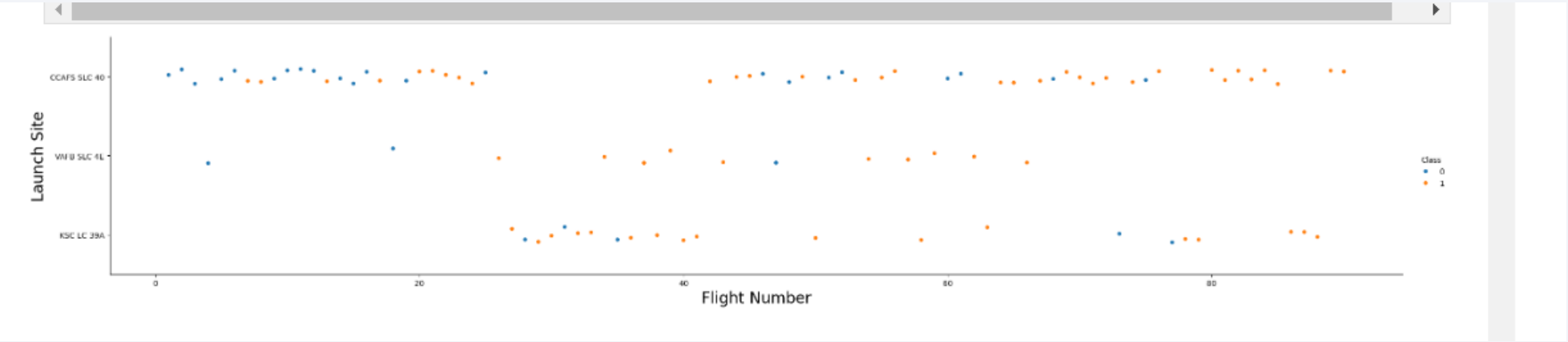
- Exploratory data analysis results: They were interesting.
- Interactive analytics demo in screenshots: They'll be shown later on
- Predictive analysis results: The results show a model with a high degree of accuracy


Section 2
Insights drawn from EDA

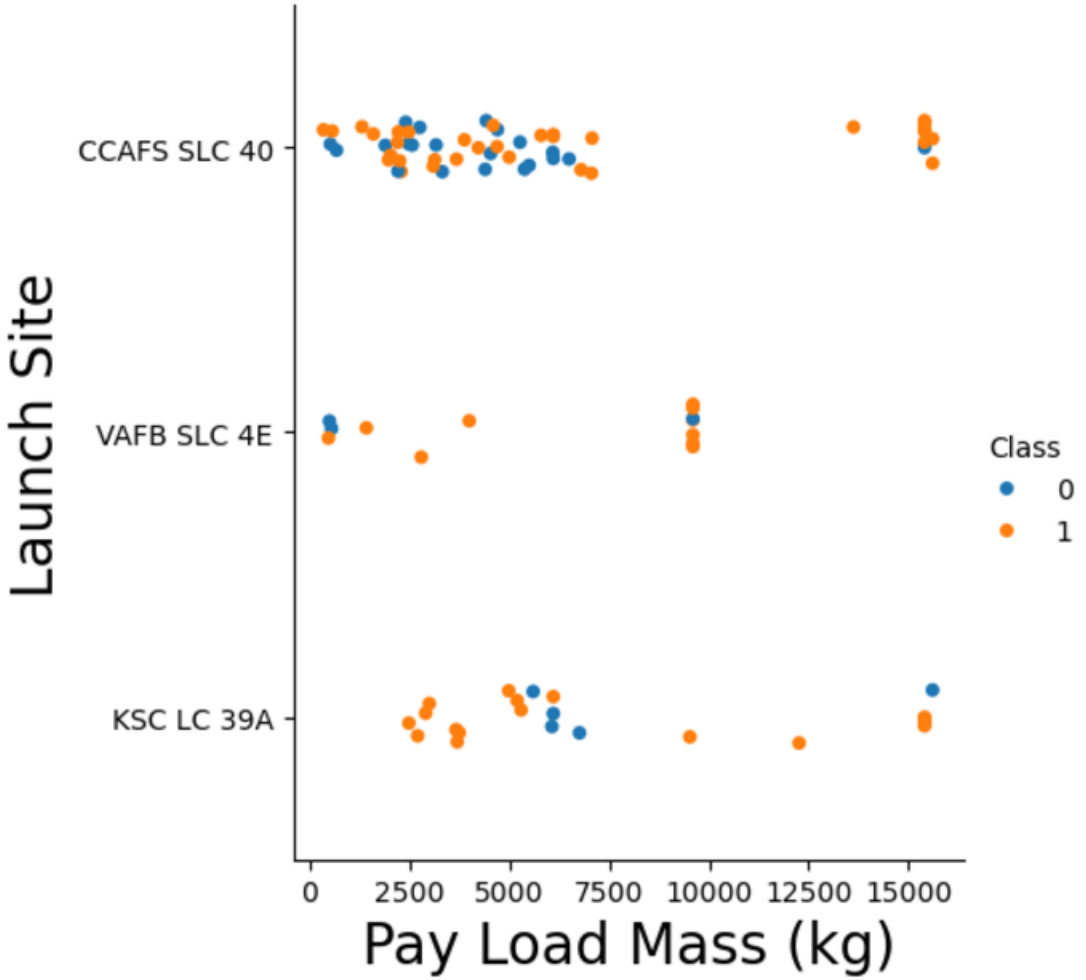
Flight Number vs. Launch Site

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

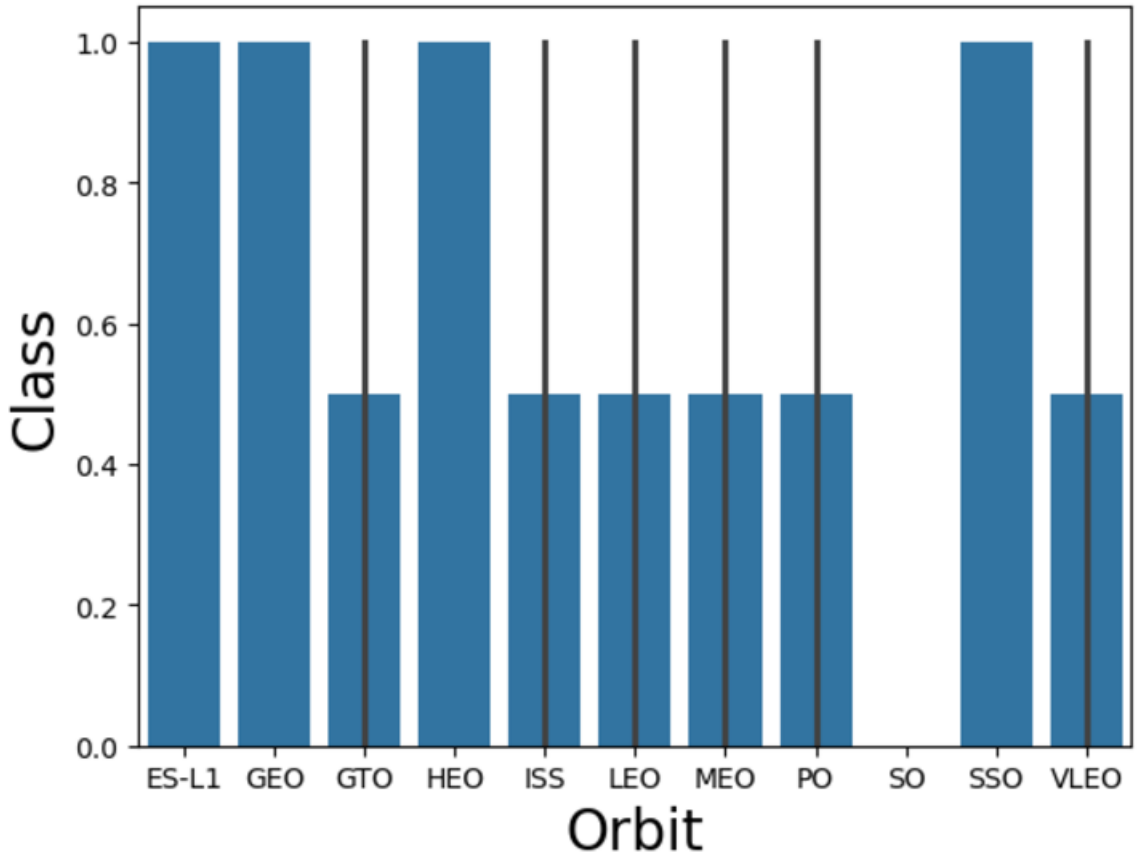
- Show the screenshot of the scatter plot with explanations



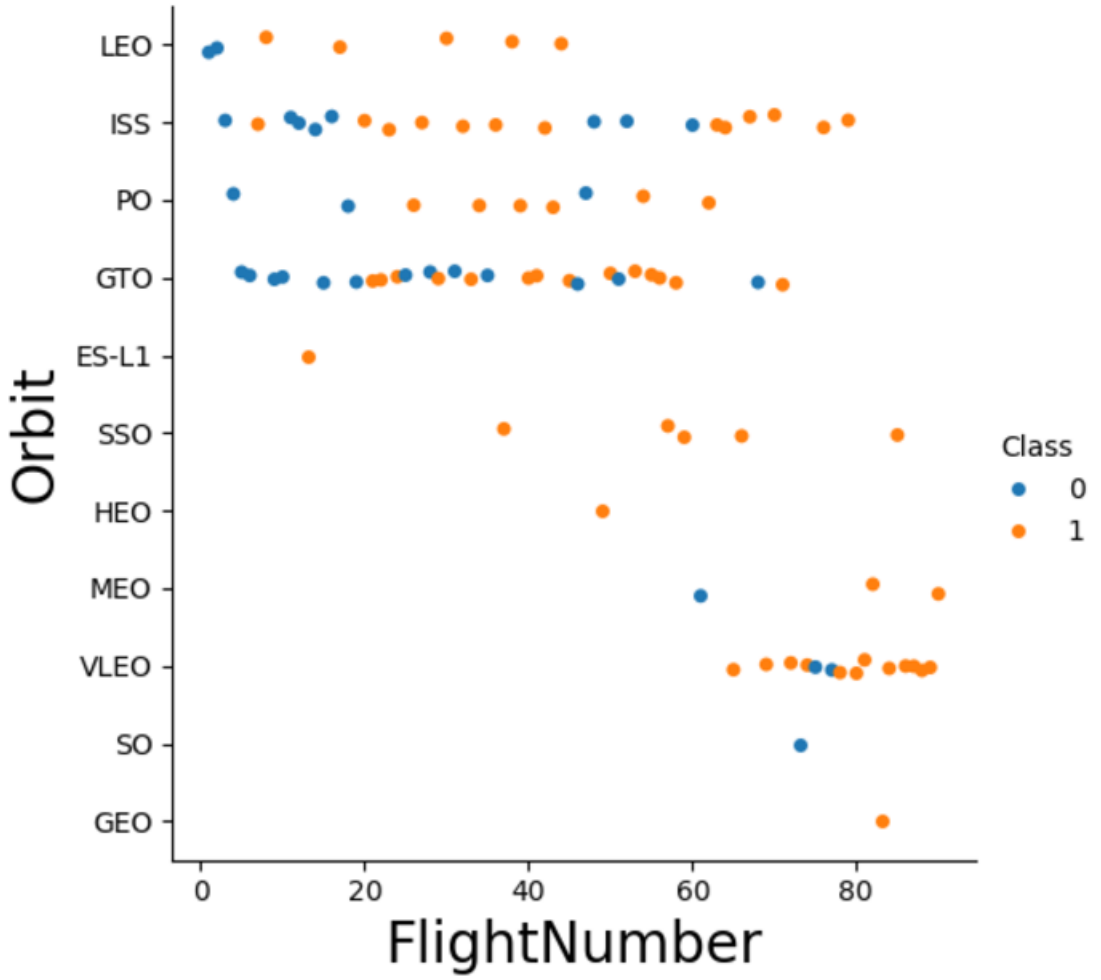
Payload vs. Launch Site



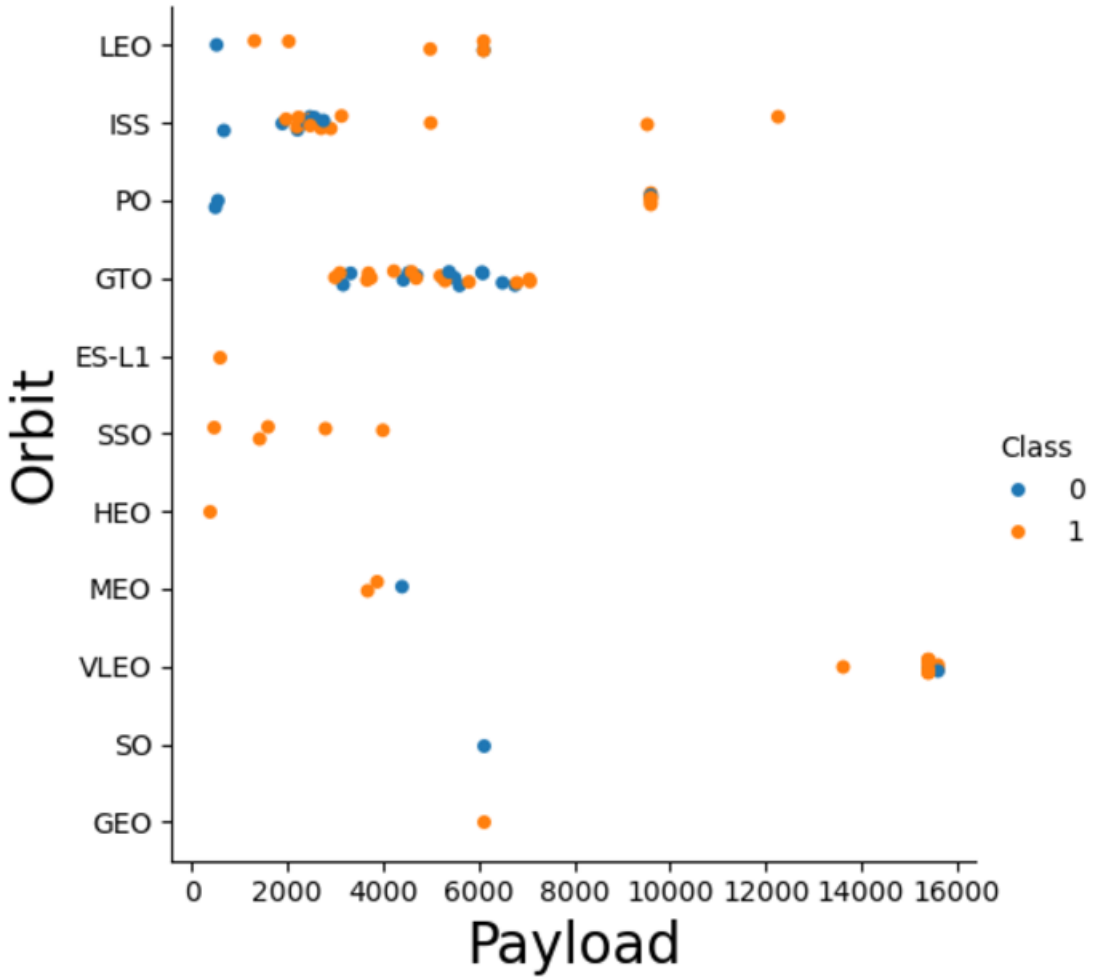
Success Rate vs. Orbit Type



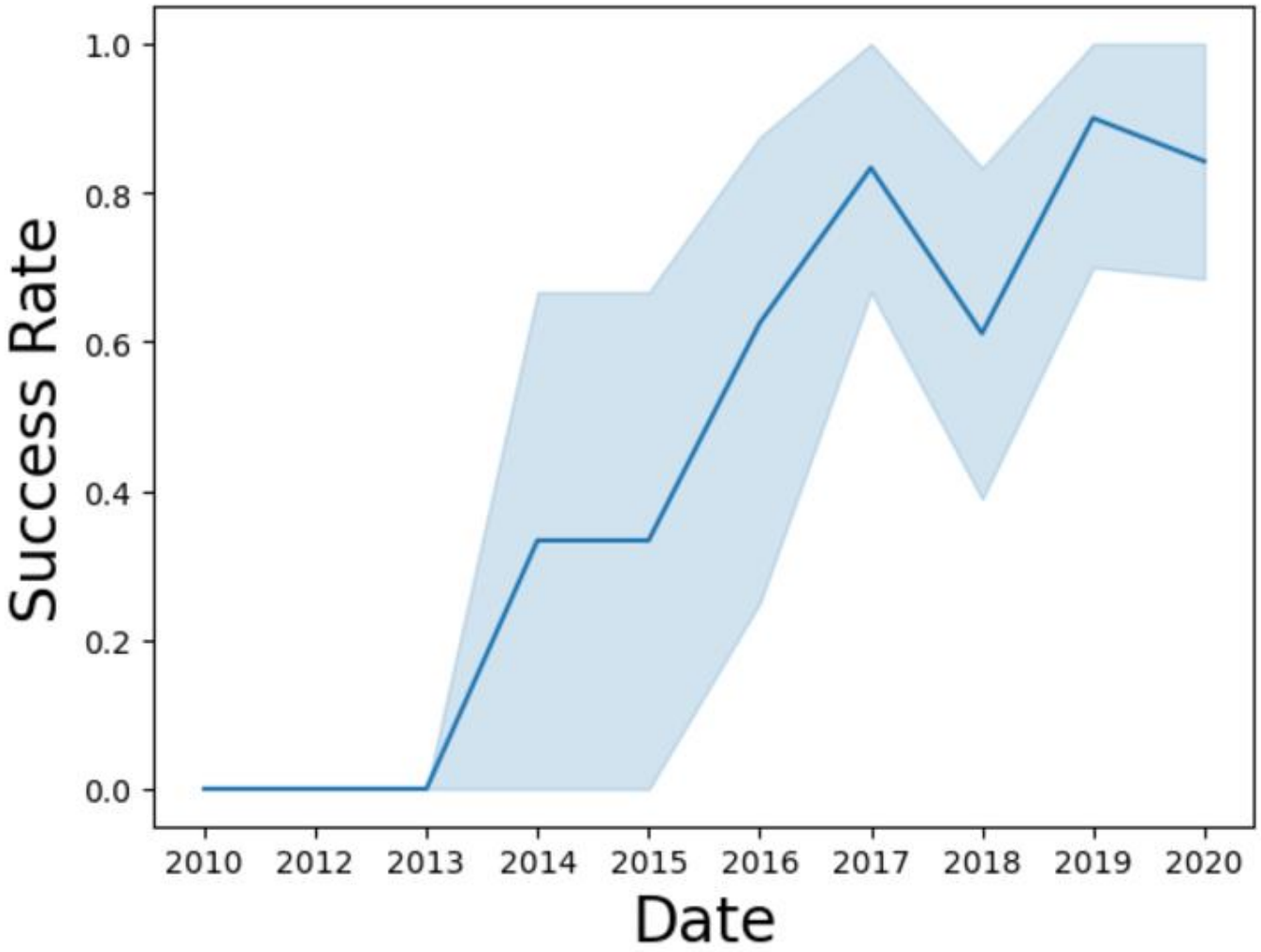
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here
- There are four unique site names.

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

Launch Site Names Begin with 'CCA'

Out[14]:

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

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here
- The total payload mass is 111286 KG.
- This shows that there has been many launches in the past.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here
- The average mass is 2928.4 KG.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here
- The date of the first successful landing was on 2015-12-22.

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

- Present your query result with a short explanation here

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here
- This shows a low failure rate.

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here
- This shows that many boosters carried the max payload.

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

- Present your query result with a short explanation here. This shows a low failure rate.

]:

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

- 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

- Present your query result with a short explanation here
- This shows a variety of results.

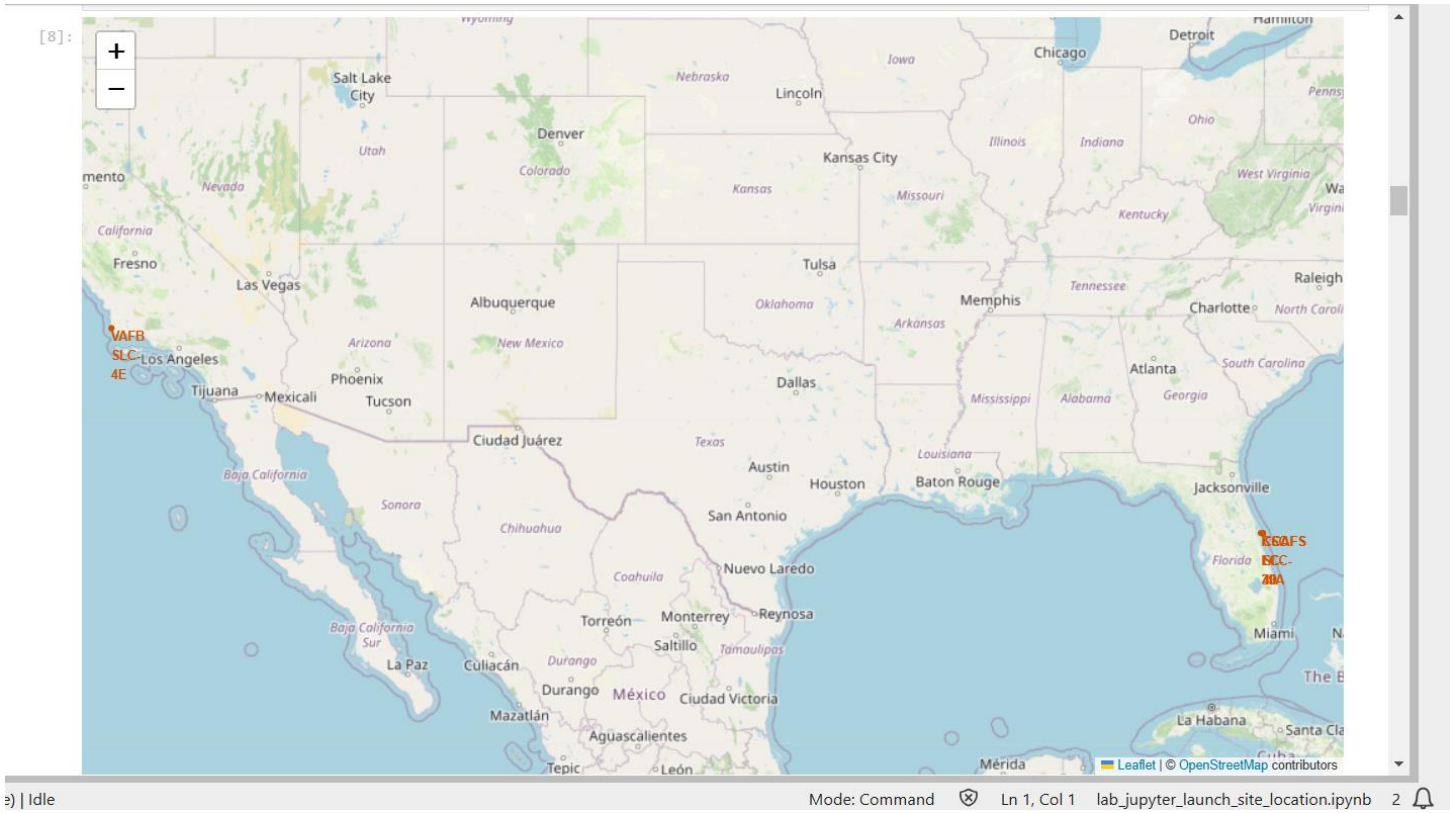
Landing_Outcome	QTY
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

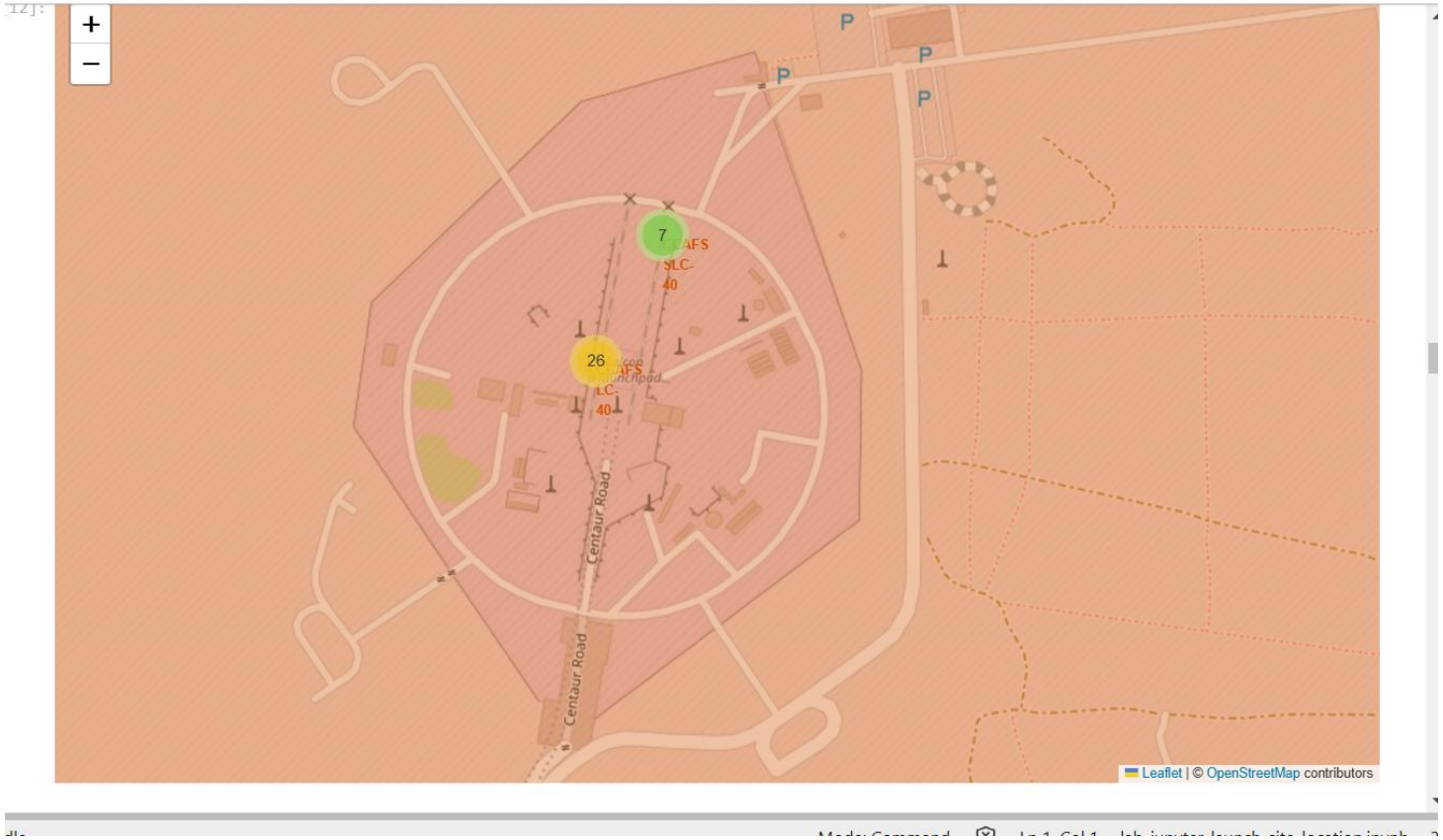
Launch Site Map

- This map shows where the launch sites are.



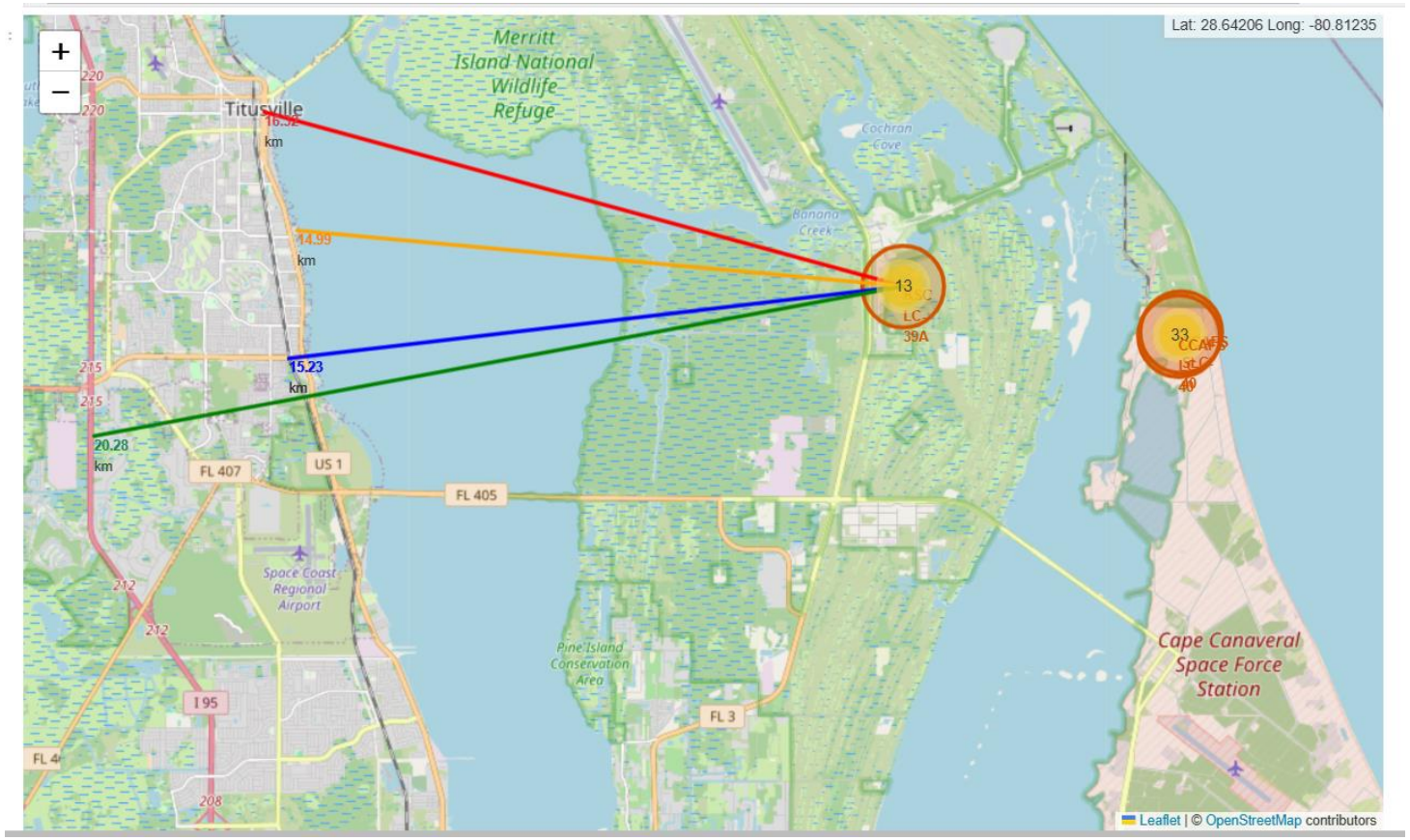
Launch Success Map

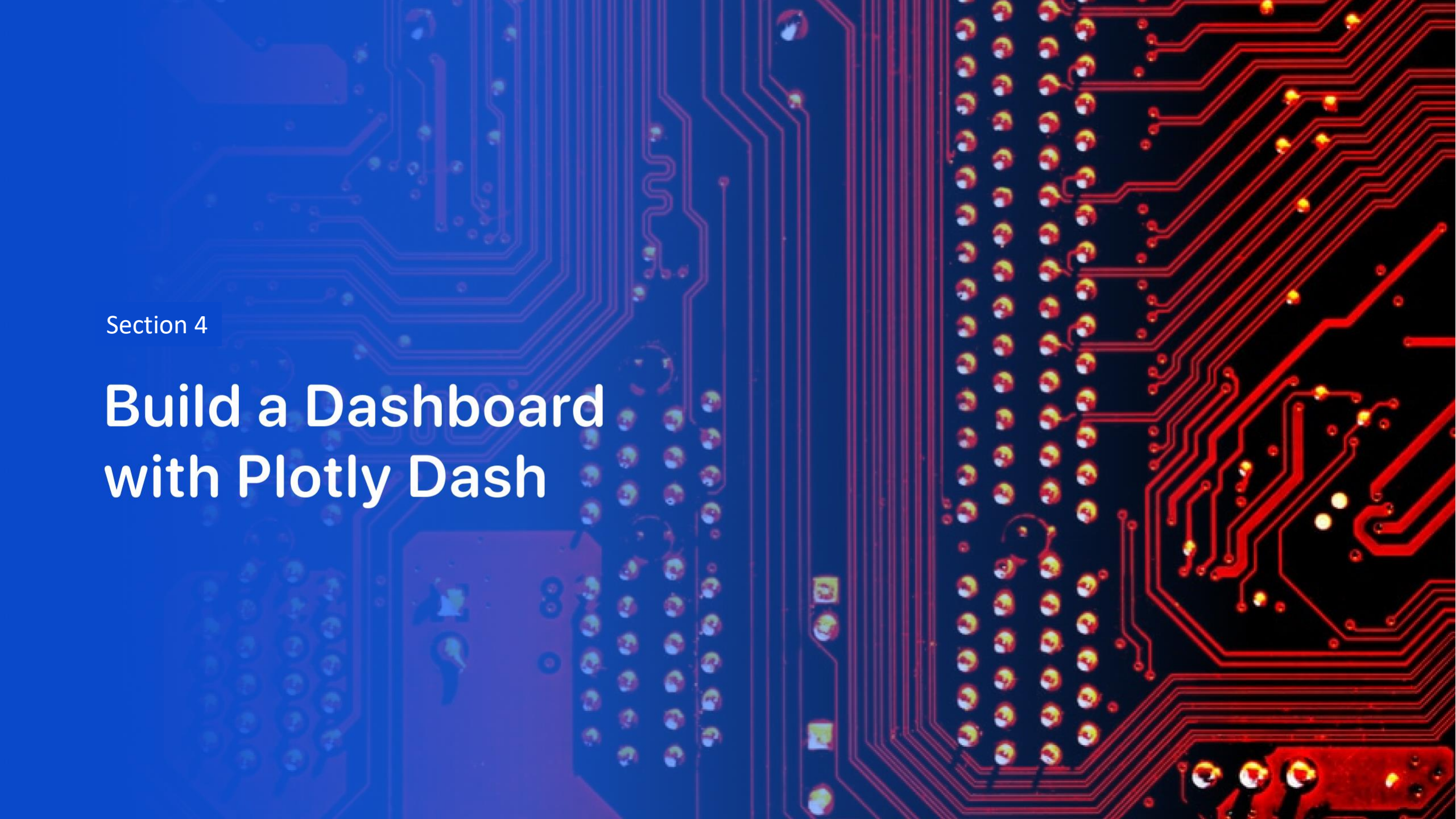
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot
- This shows the types of launches and success.



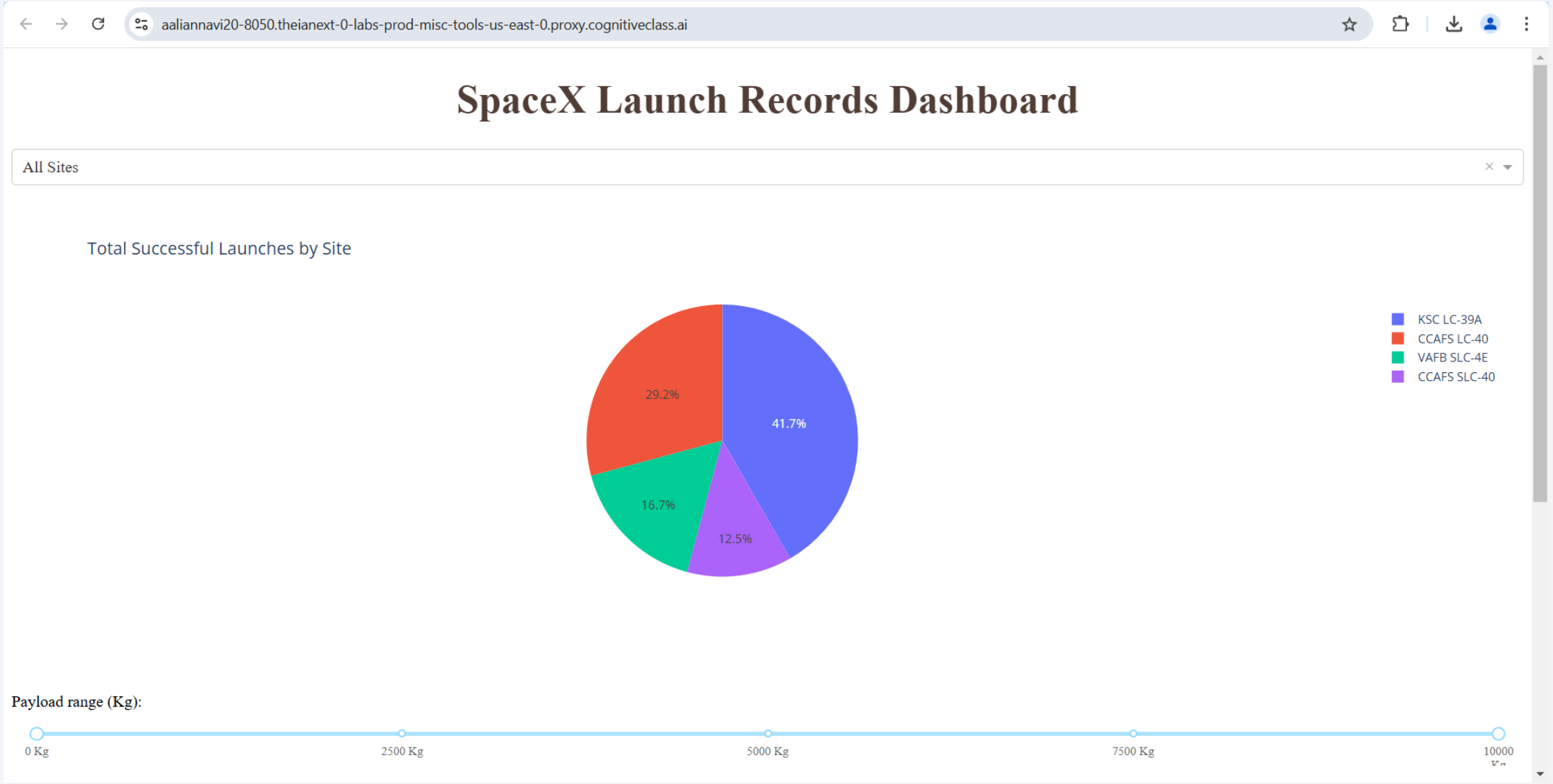
Launch Site Distance

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot
- The map image shows the distance from the launch site to the nearby city, coastline, highway and railroad.



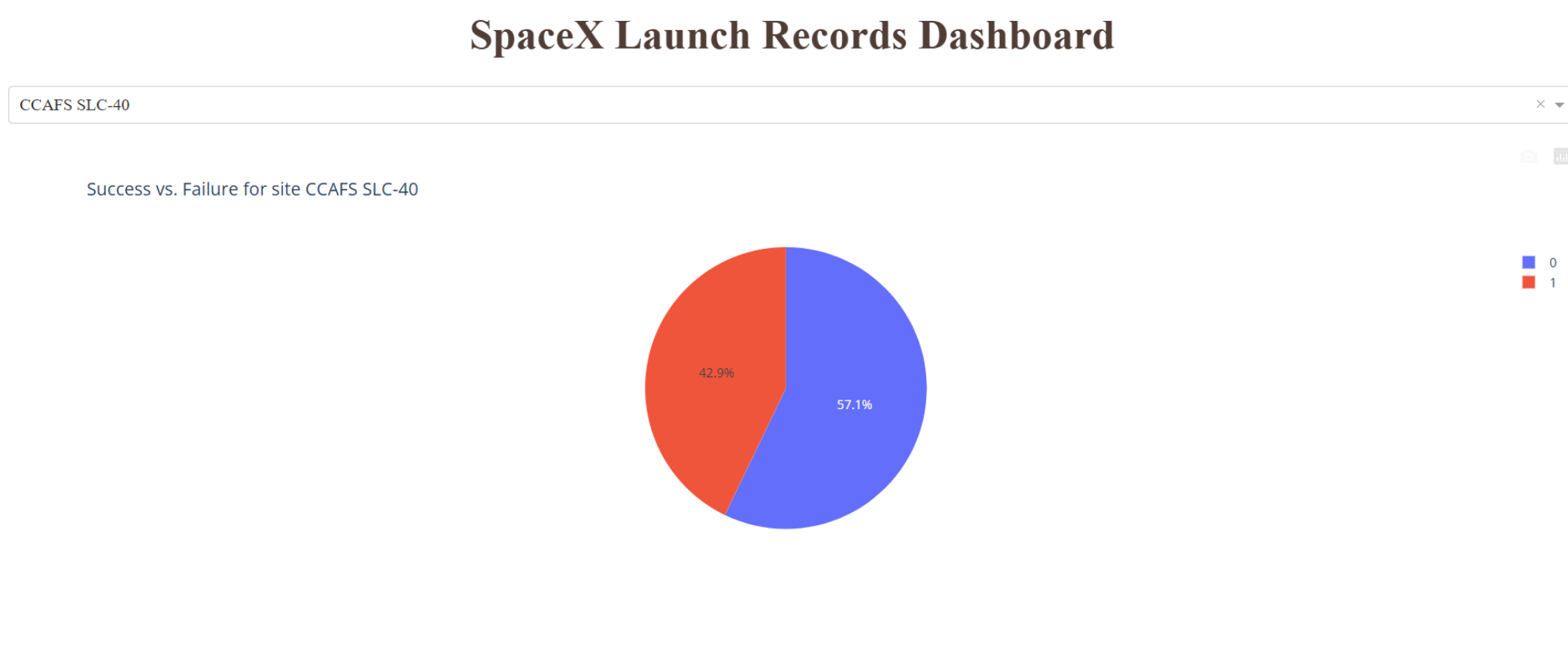

Section 4
Build a Dashboard with Plotly Dash

Pie Chart


aaliannavi20-8050.theianext-0-labs-prod-misc-tools-us-east-0.proxy.cognitiveclass.ai
SpaceX Launch Records Dashboard
All Sites
Total Successful Launches by Site
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
0 Kg
2500 Kg
5000 Kg
7500 Kg
10000

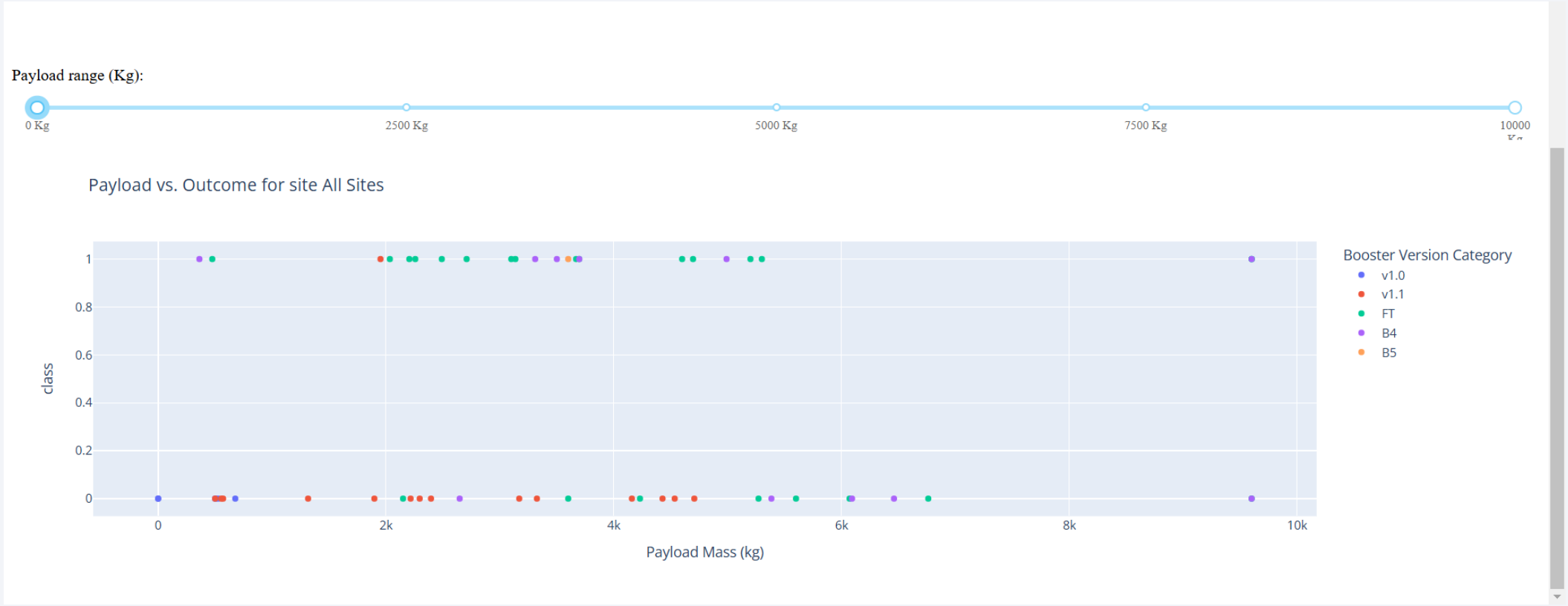
Highest Success Rate Pie Chart

The red is success, blue is failure.



Payload Scatter Plot

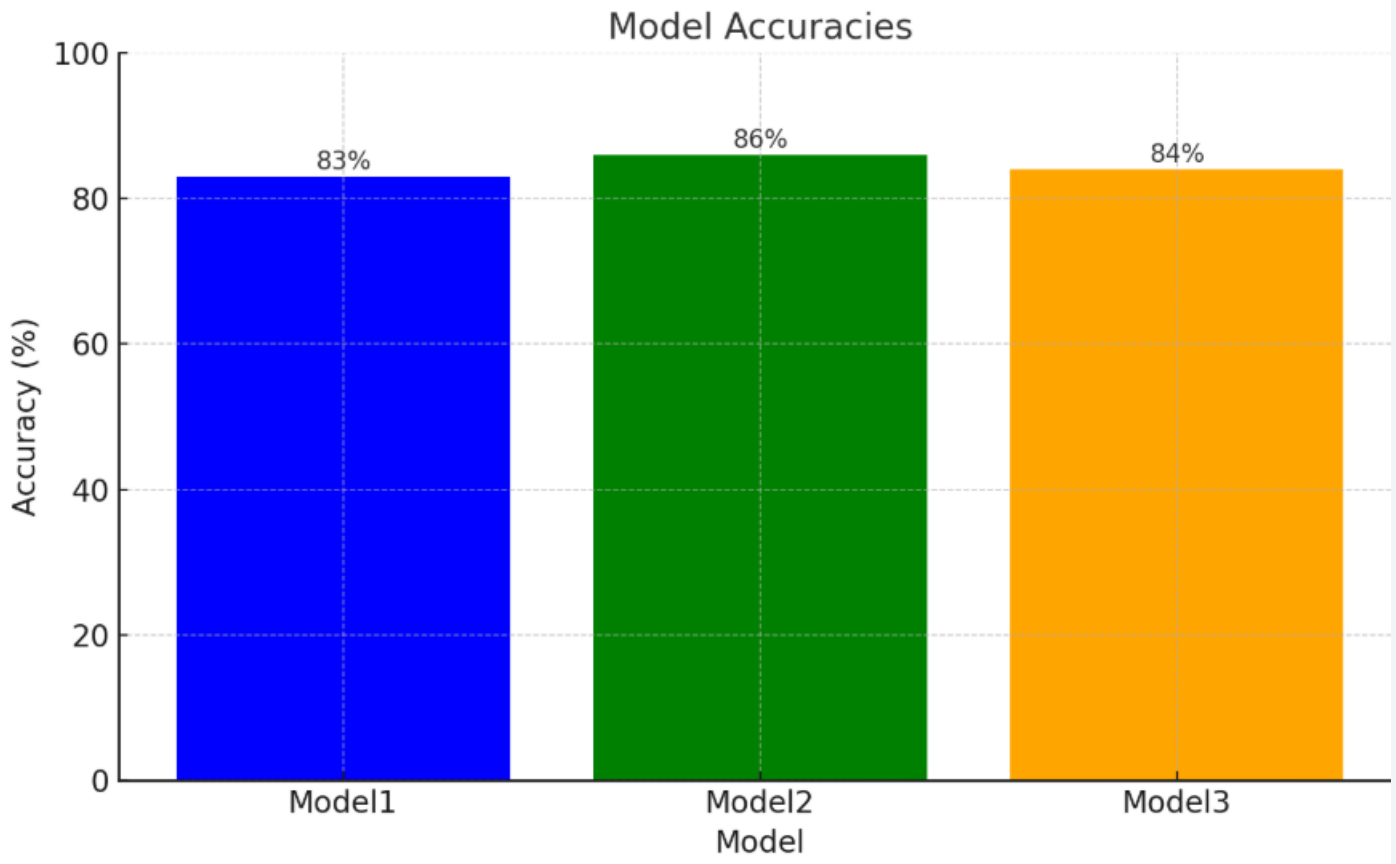
- This shows that the boosters carry various masses.



Section 5

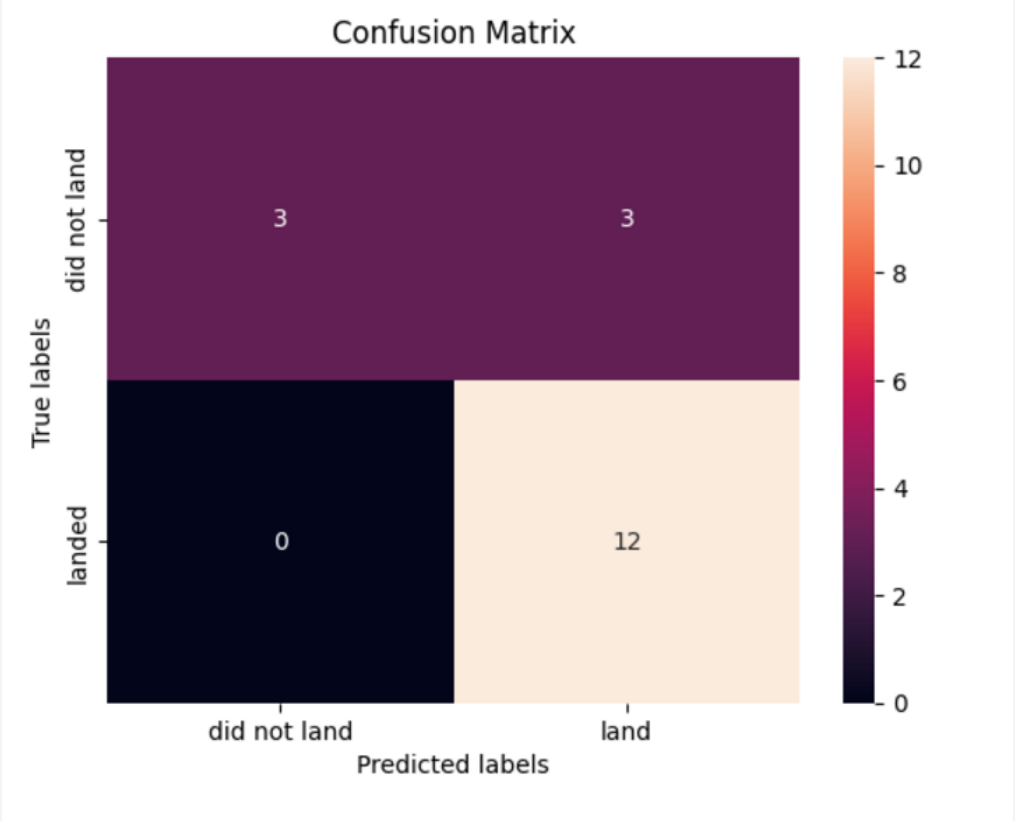
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- This confusion matrix shows that the model performs well for predicting "landed" but makes some errors in predicting "did not land."



Conclusions

- Point 1: SpaceX has a generally high success rate, and the failures are mostly planned by SpaceX.
- Point 2: SpaceX has four launch sites with varying degrees of success.
- Point 3: Boosters often carried the max payload.
- Point 4: The launch success rate increases every year.
- Point 5: The launch sites are kept at a distance away from any major population center, in case a launch goes wrong.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project
- The code are in the GitHub file links provided earlier in the power point

Thank you!