IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdullah Khaled Alqulaiti
2023/12/30

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:
    - Using SpaceX Rest API
    - Using Web scrapping from Wikipedia
- Perform data wrangling
    - Filtering the data and handling the missing values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Building, training and evaluation the model to get the best resalt

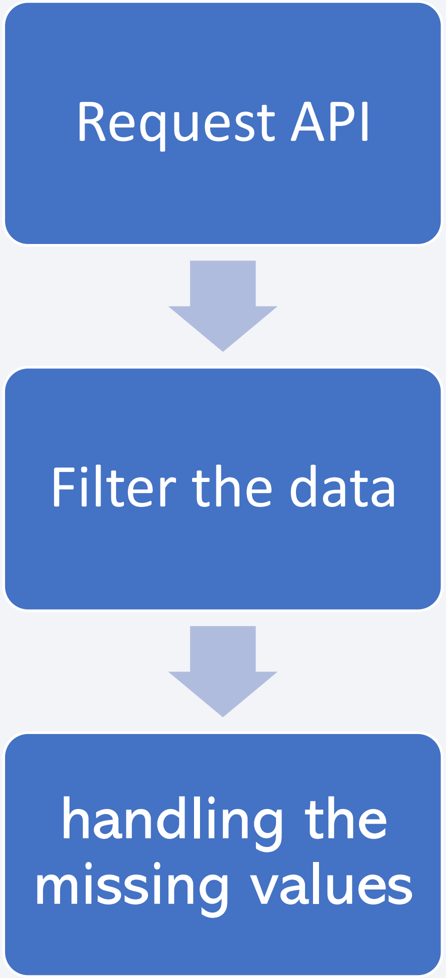
Data Collection

- data sets were collected from SpaceX API and form Wikipedia.
- Methods will be presented in next slides.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

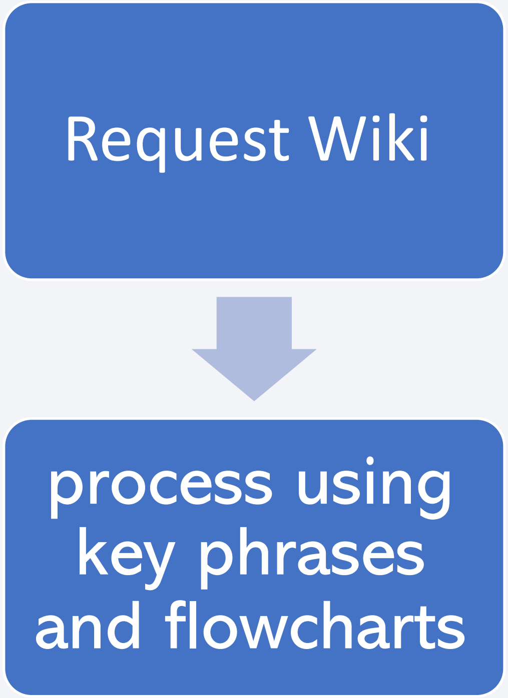
- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/jupyter-labs-webscraping.ipynb

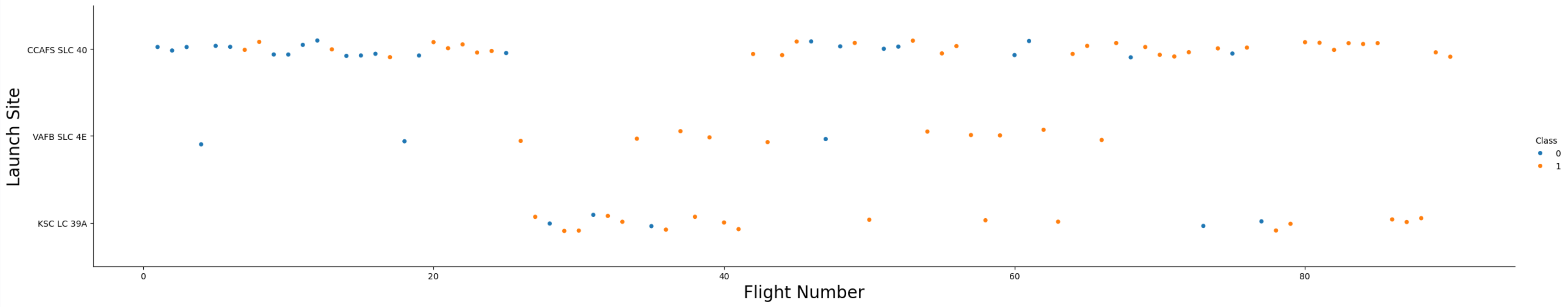


Data Wrangling

- First EDA was performed on the dataset then summarizations and creation of landing outcome label.
- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- There was different chart used to visualize the relationship like scatterplots and barplots.
- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb
- **Visualize the relationship between Flight Number and Launch Site**



EDA with SQL

- These are most of the SQL queries that were executed:
  - names of the unique launch sites in the space mission
  - 5 records where launch sites begin with the string 'CCA'
  - the total payload mass carried by boosters launched by NASA (CRS)
  - average payload mass carried by booster version F9 v1.1
  - date when the first succesful landing outcome in ground pad was acheived.
  - total number of successful and failure mission outcomes
  - names of the booster_versions which have carried the maximum payload mass.
  - the count of landing outcomes
  - Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps.
  - Markers indicate points like launch sites;
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site;
  - Lines are used to indicate distances between two coordinates.

Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/spacex_dash_app.py

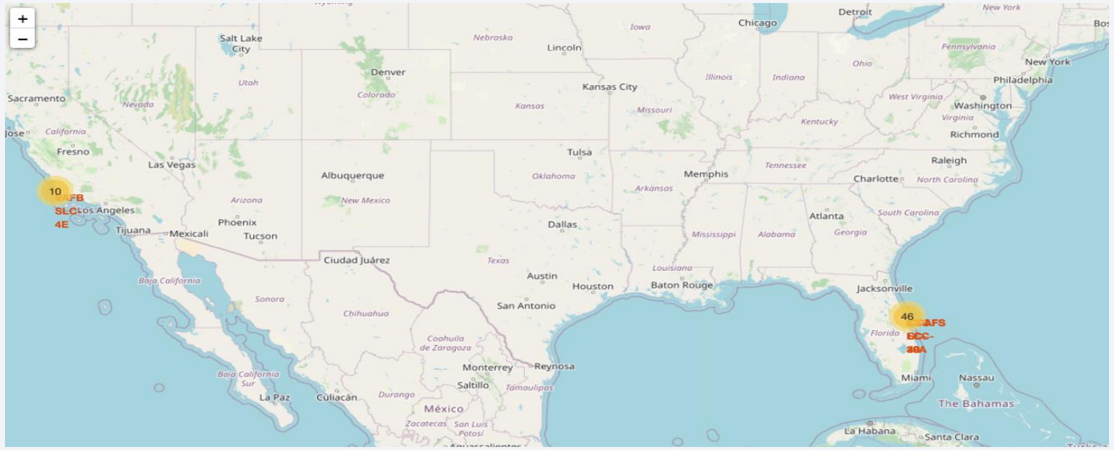
Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

- Source code: https://github.com/abdullah-KQ/final-project-for-IBM-Data-Science-Professional-Certificate/blob/149c715c5bf033365df0620287ebddbe136cbf45/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results
  - Exploratory data analysis results:
  - Space X uses 4 different launch sites
  - The first launches were done to Space X itself and NASA
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Predictive Analysis showed that Decision Tree



Classifier is the best model to predict successful

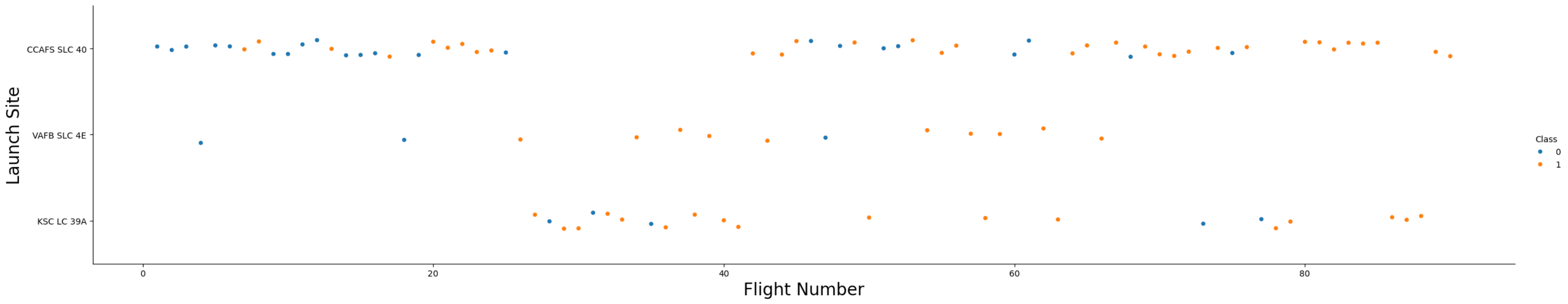
landings, having accuracy over 87% and accuracy

for test data over 94%.

Section 2

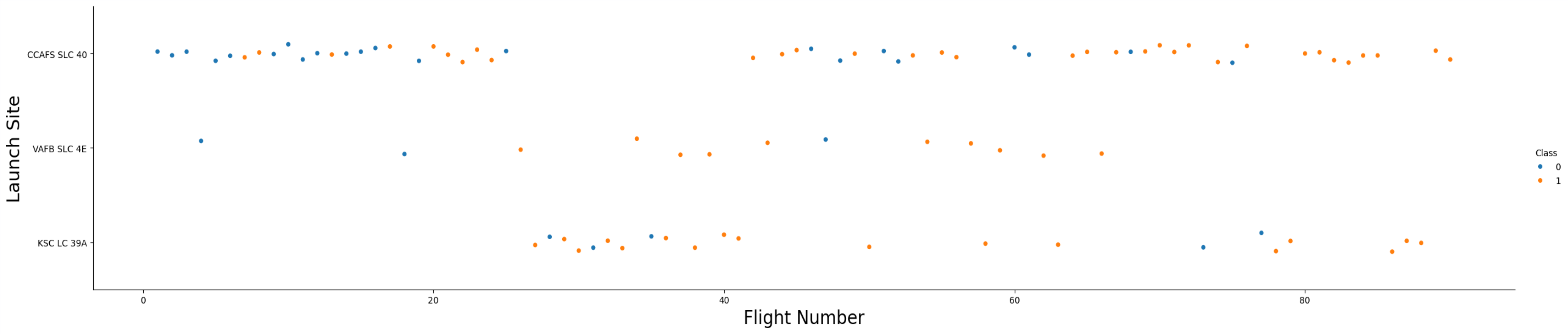
Insights drawn from EDA

Flight Number vs. Launch Site



- Show the screenshot of the scatter plot with explanations

Payload vs. Launch Site



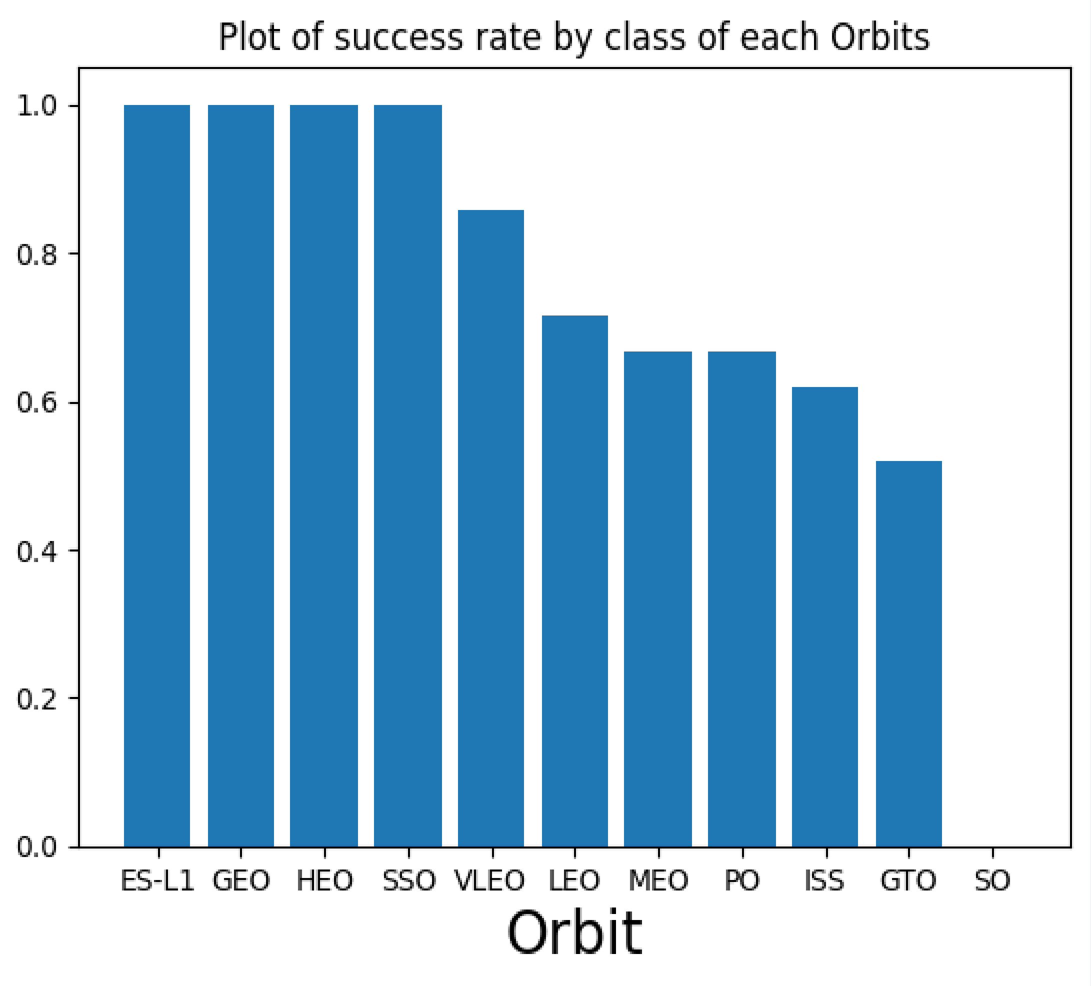
- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

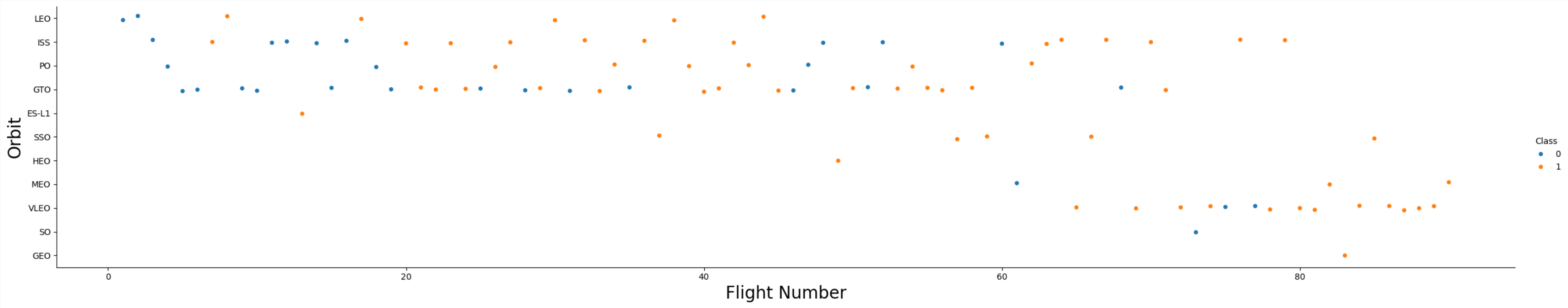
Success Rate vs. Orbit Type

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

- Show the screenshot of the scatter plot with explanations



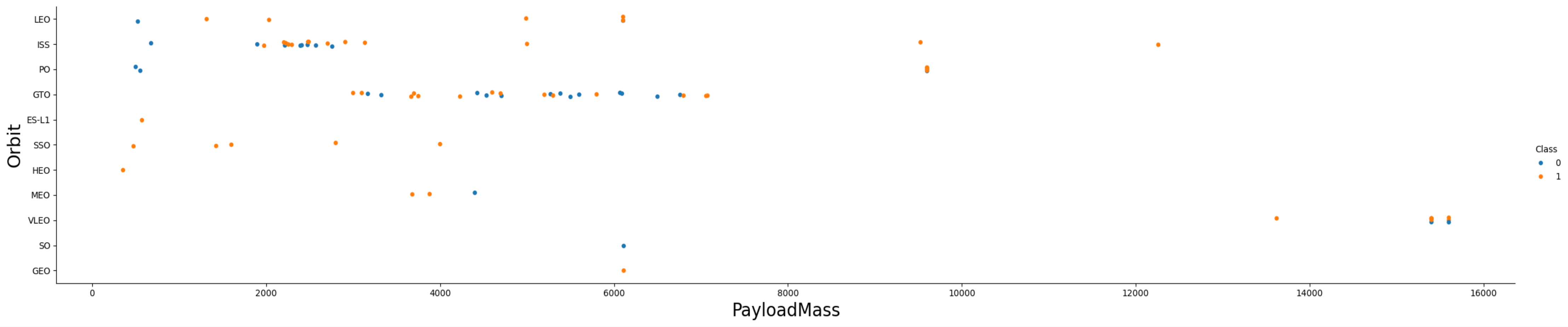
Flight Number vs. Orbit Type



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

- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type



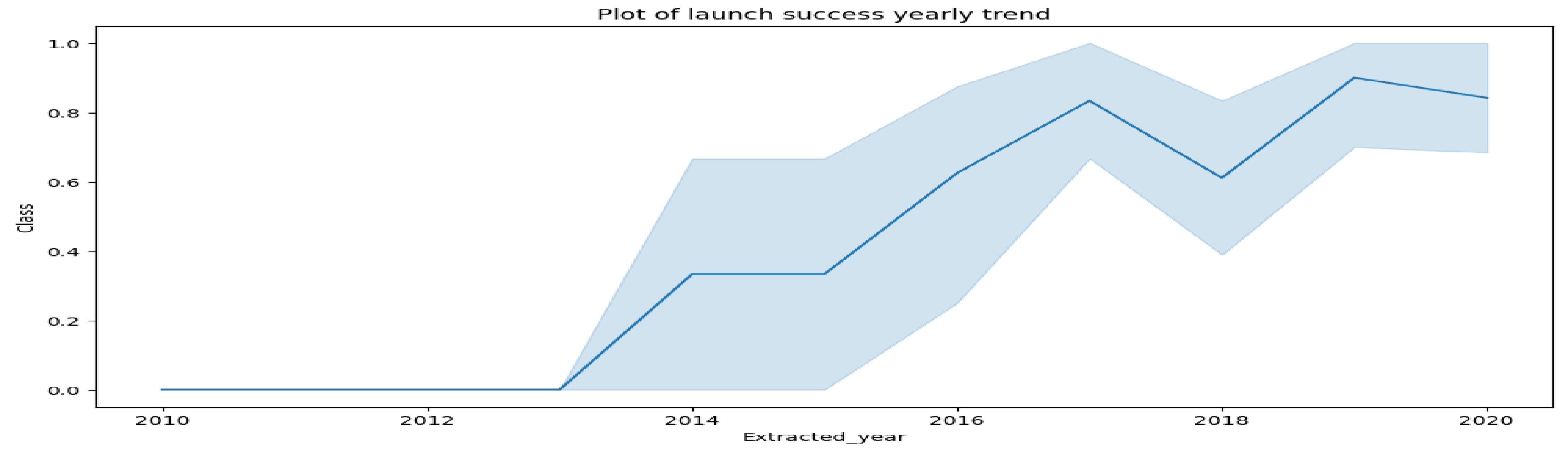
- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- According to data, there are four launch sites:

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

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	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

- Total payload carried by boosters from NASA:

SUM(PAYLOAD_MASS__KG_)
45596

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

- the average payload mass we obtained the value of 2,928 kg

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- number of successful mission outcomes

SUCCESSFUL_MISSIONS
61

Boosters Carried Maximum Payload

- booster which have carried the maximum payload mass

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

Landing_Outcome	Booster_Version	Launch_Site	DATE_YEAR
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	01
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04

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

- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	COUNT
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

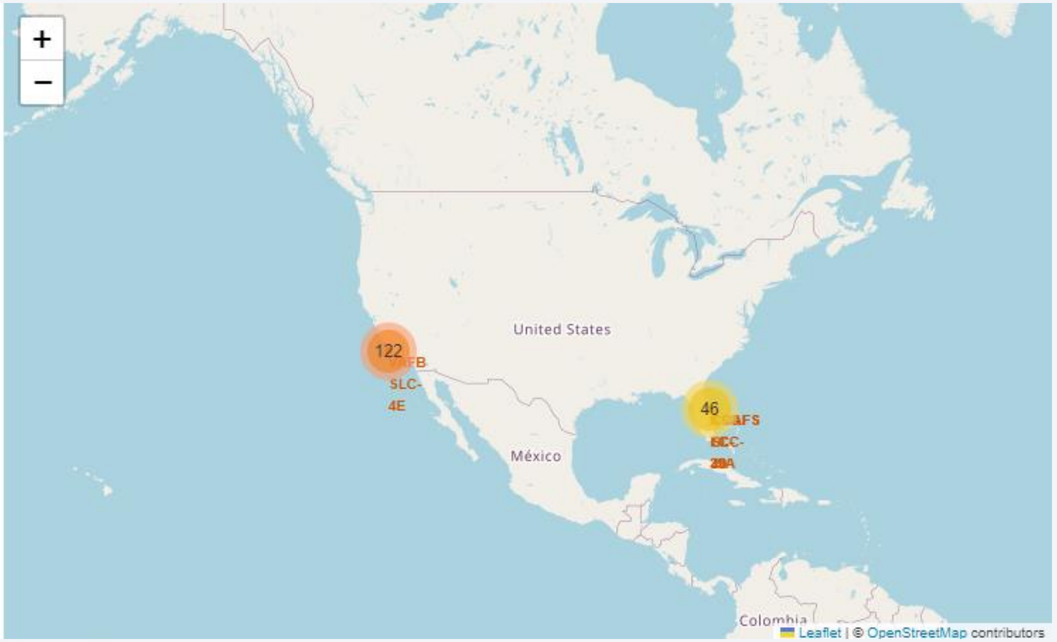
All launch sites

All launch sites

- Launch sites are near sea, probably by safety, but not too far from roads and railroads

Launch Outcomes by Site

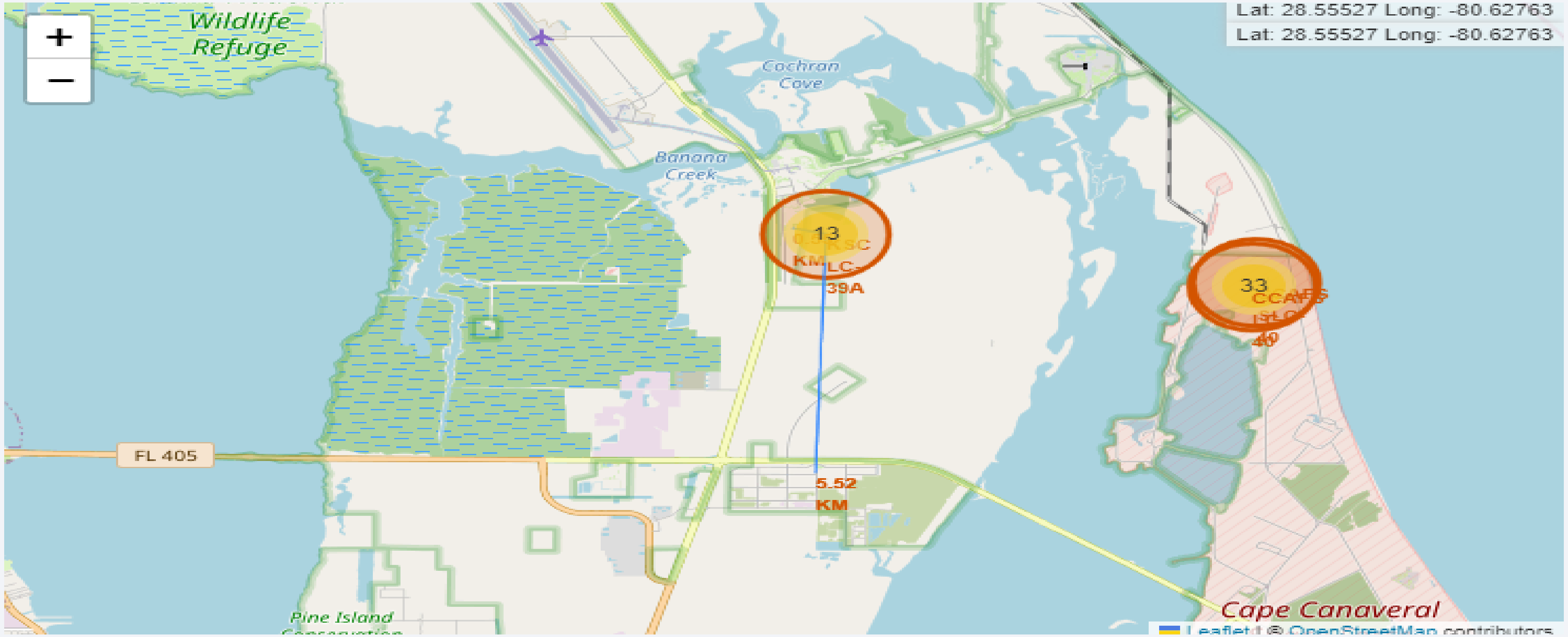
Launch Outcomes by Site



- Green markers indicate successful and red ones indicate failure.

Logistics and Safety

Logistics and Safety

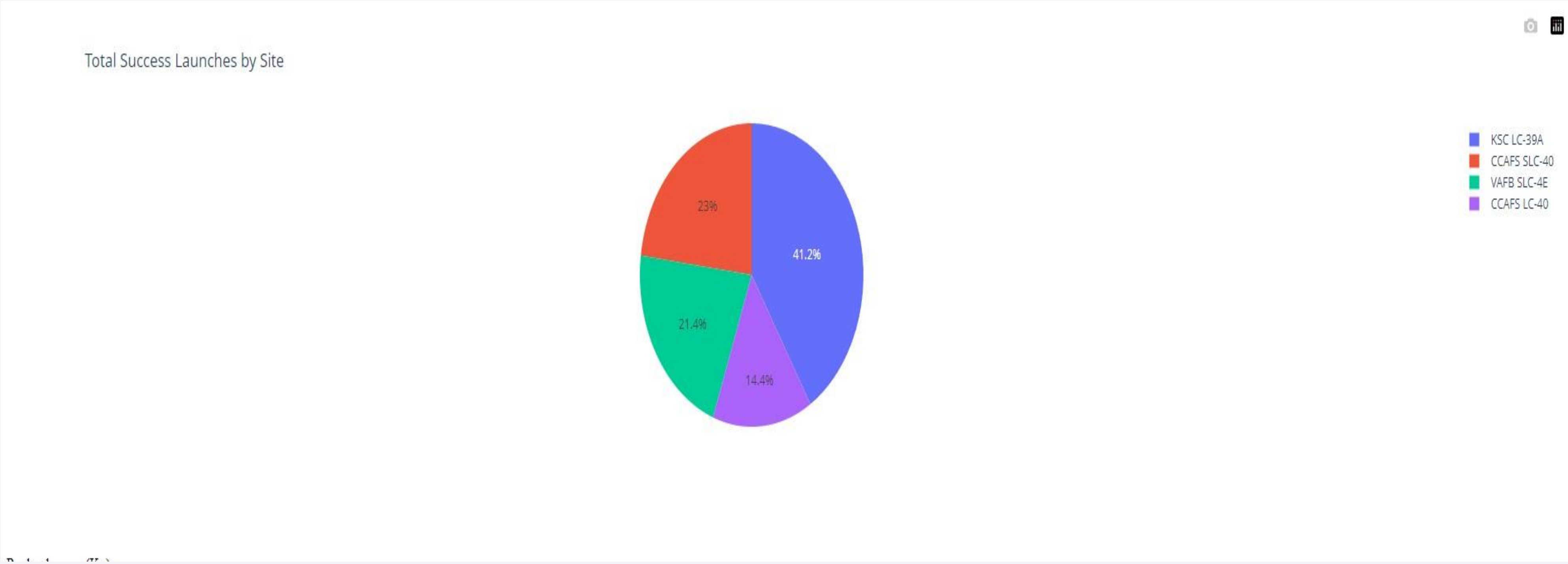


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas

Section 4

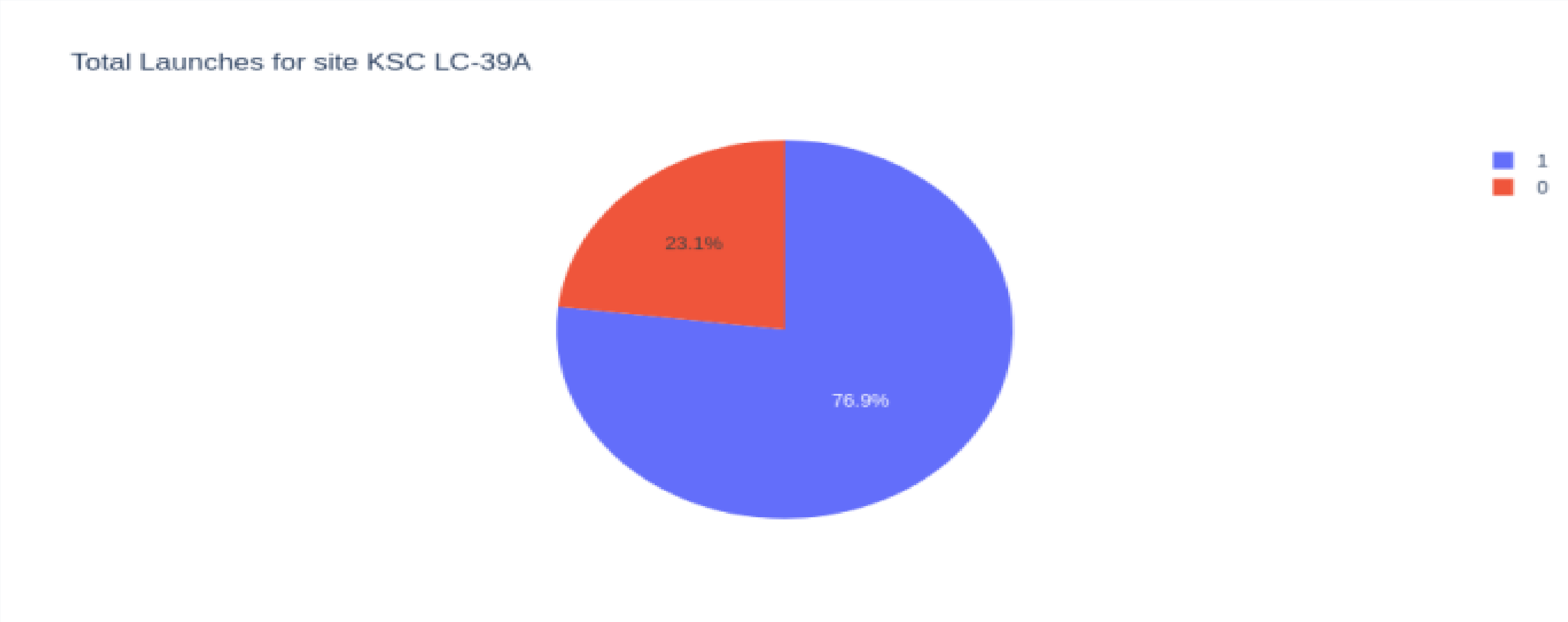
Build a Dashboard with Plotly Dash

Successful Launches by Site



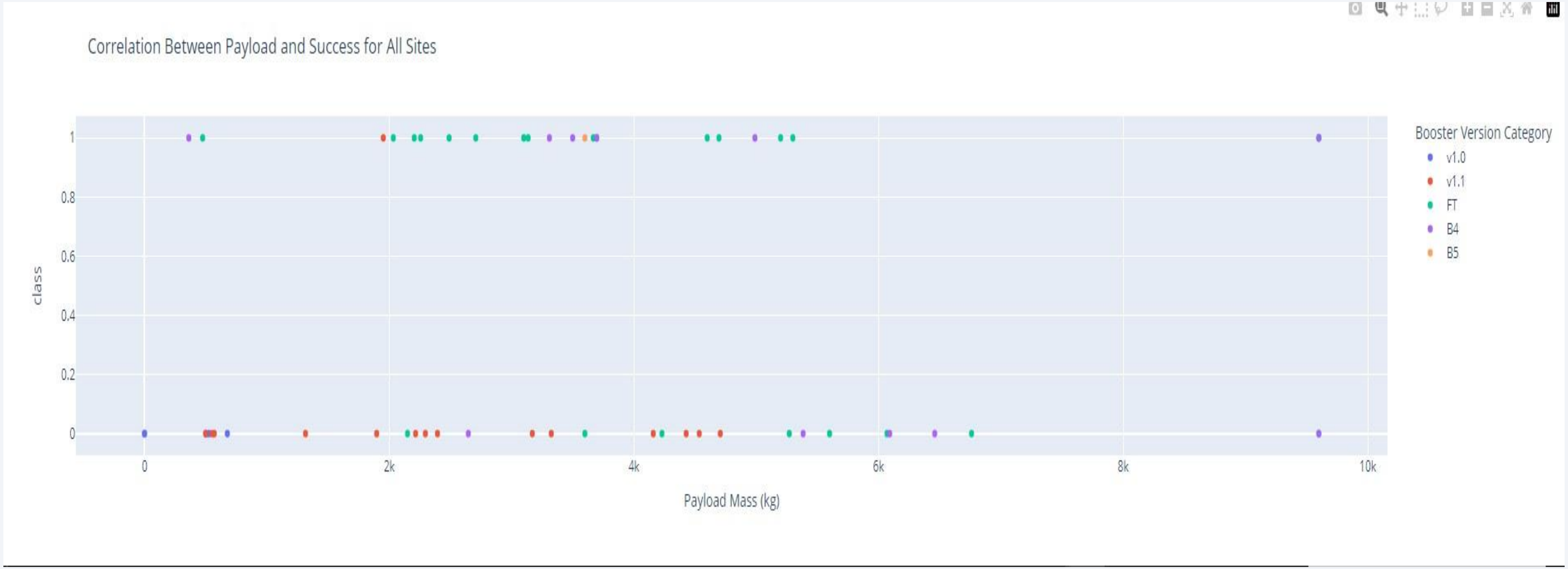
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

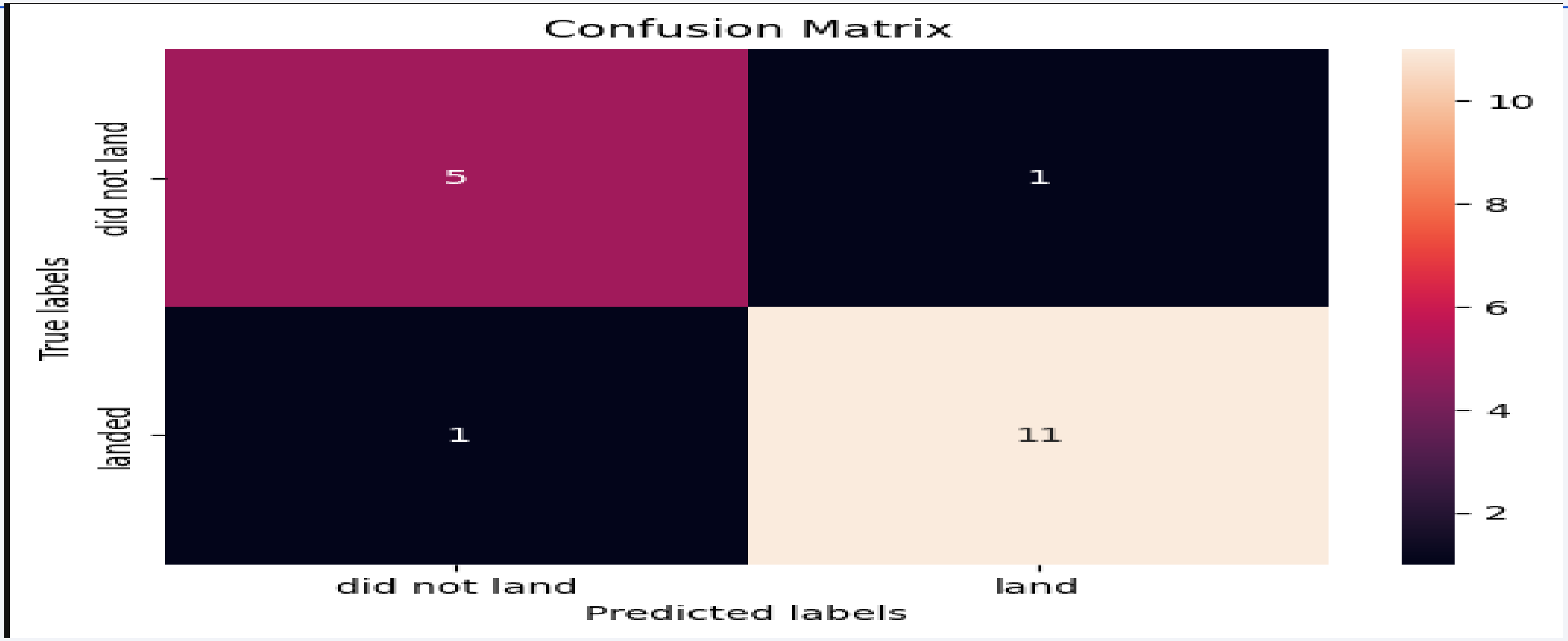
Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%

```
LogReg: Accuracy:0.84643   TestAccuracy:0.83333
SVM:    Accuracy:0.84821   TestAccuracy:0.83333
Tree:   Accuracy:0.875     TestAccuracy:0.88889
KNN:    Accuracy:0.84821   TestAccuracy:0.83333
```

Confusion Matrix : Decision Tree



- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- most of mission outcomes are successful, successful landing outcomes seem to improve over time
- Decision Tree Classifier can be used to predict successful landings and increase profits

Appendix

- Folium didn't show maps on GitHub

Thank you!