Winning Space Race with Data Science

Mariangela Santorsola
March 4th 2022

Outline

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

Executive Summary

- Summary of methodologies
  - SpaceX Data Collection
    - Web scraping
    - API
  - Exploratory Data Analysis (EDA)
  - Machine Learning (ML) Predictions
  - Visualization of key results
- Summary of all results
  - Prediction of the total cost of launches by best ML model

Introduction

- Project background and context
  - Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- Problems you want to find answers
  - SpaceX Falcon 9 First Stage Landing Prediction
  - Cost of launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API (api.spacexdata.com/v4/rockets)
  - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Perform data wrangling
  - Data summarized and landing outcome label created.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Collected data were normalized. Dataset split in test and training sets. Performed four different ML models, evaluation of accuracy.

Data Collection

- Data collection methodology:
  - SpaceX API (api.spacexdata.com/v4/rockets)
  - Data scraping

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

- SpaceX API (api.spacexdata.com/v4/rockets)

https://github.com/msantorsola/python_project_data_science/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

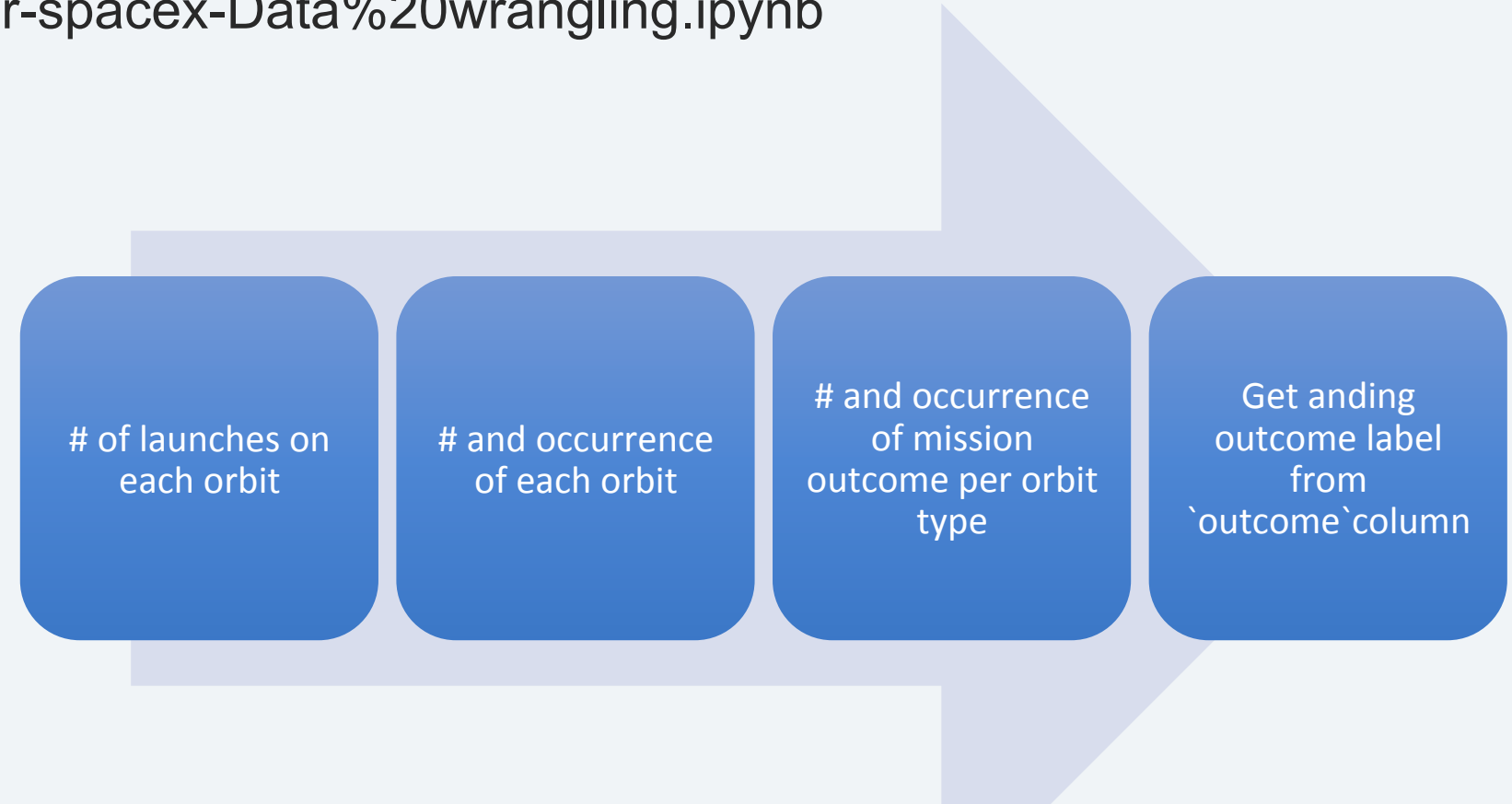
Data Collection - Scraping

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

https://github.com/msantorsola/python_project_data_science/blob/main/jupyter-labs-webscraping.ipynb

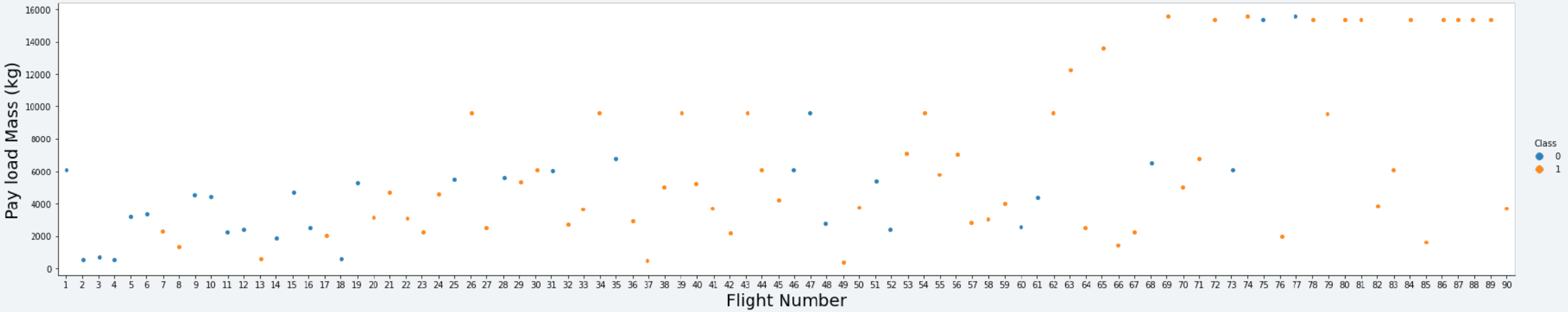
Data Wrangling

- EDA to determine the labels for training supervised models
- https://github.com/msantorsola/python_project_data_science/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatterplot and barplot to visualize relationship between each feature pair



- https://github.com/msantorsola/python_project_data_science/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- SQL queries:
  - Total payload mass by booster F9v1.1
  - 5 records where launch sites begin with the string 'CCA'
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - date when the first successful landing outcome in ground pad was acheived.
  - names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - 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

- https://github.com/msantorsola/python_project_data_science/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters added to a folium map
  - Markers for launch sites
  - Circles for areas around defined coordinates
  - Marker clusters for events in each coordinates
  - Lines for distances between coordinates

- https://github.com/msantorsola/python_project_data_science/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie chart of launch success
- Scatter plot for correlation between payload and launch success

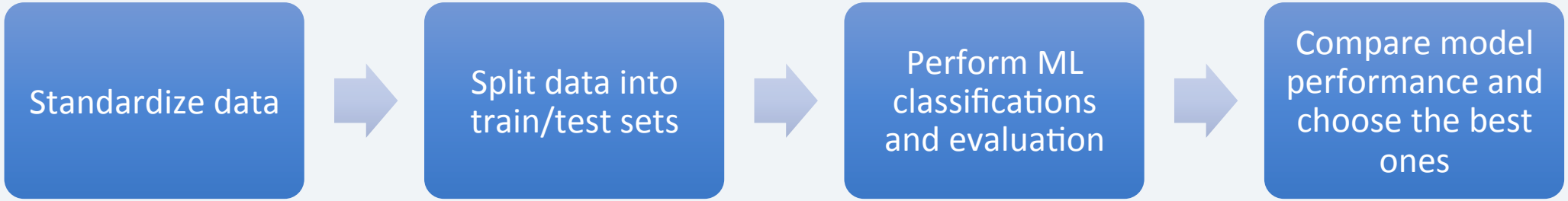
- AIM: Identification of the best place to launch according to payload

- https://github.com/msantorsola/python_project_data_science

Predictive Analysis (Classification)

ML classification models used:

- Logistic regression
- support vector machine
- decision tree
- k neighrest neighbors



github.com/msantorsola/python_project_data_science/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

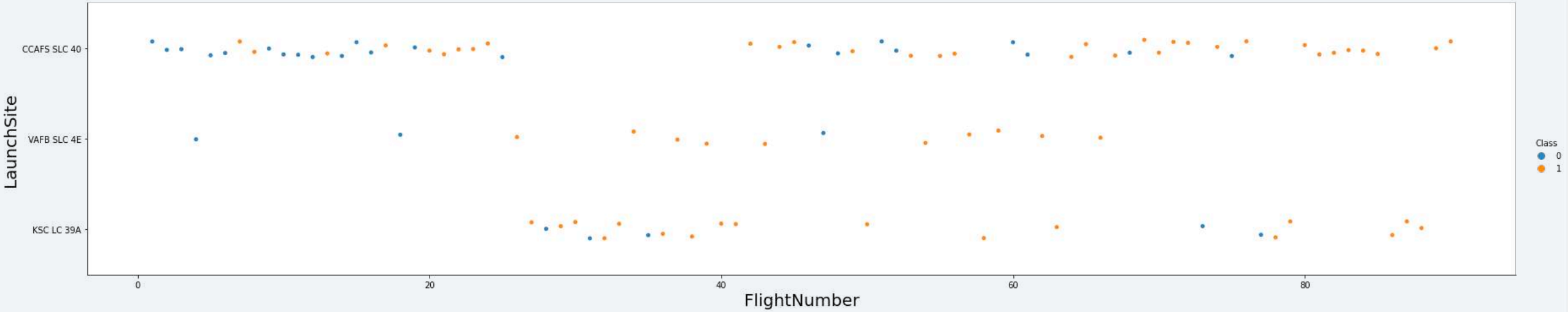
Results

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

Section 2

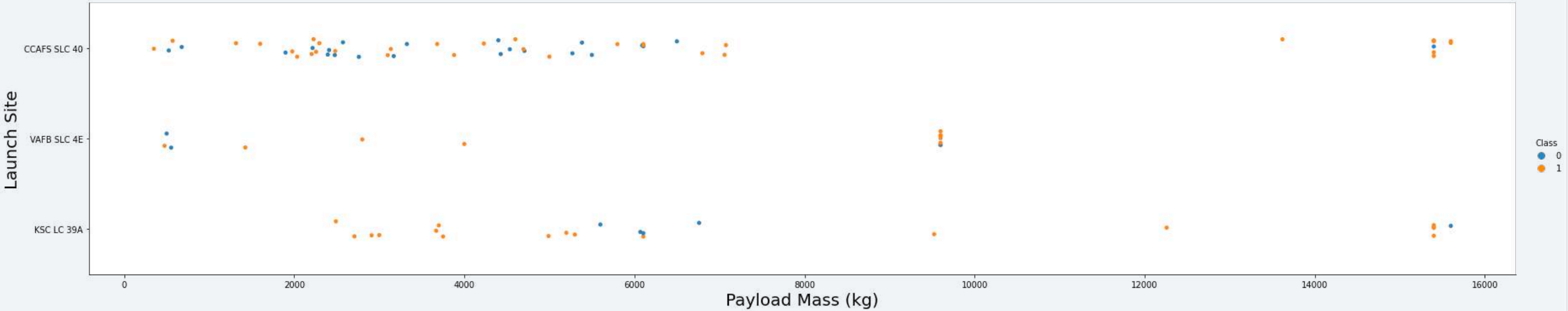
Insights drawn from EDA

Flight Number vs. Launch Site



- The best launch site for number of success is CCAF5 SLC 40

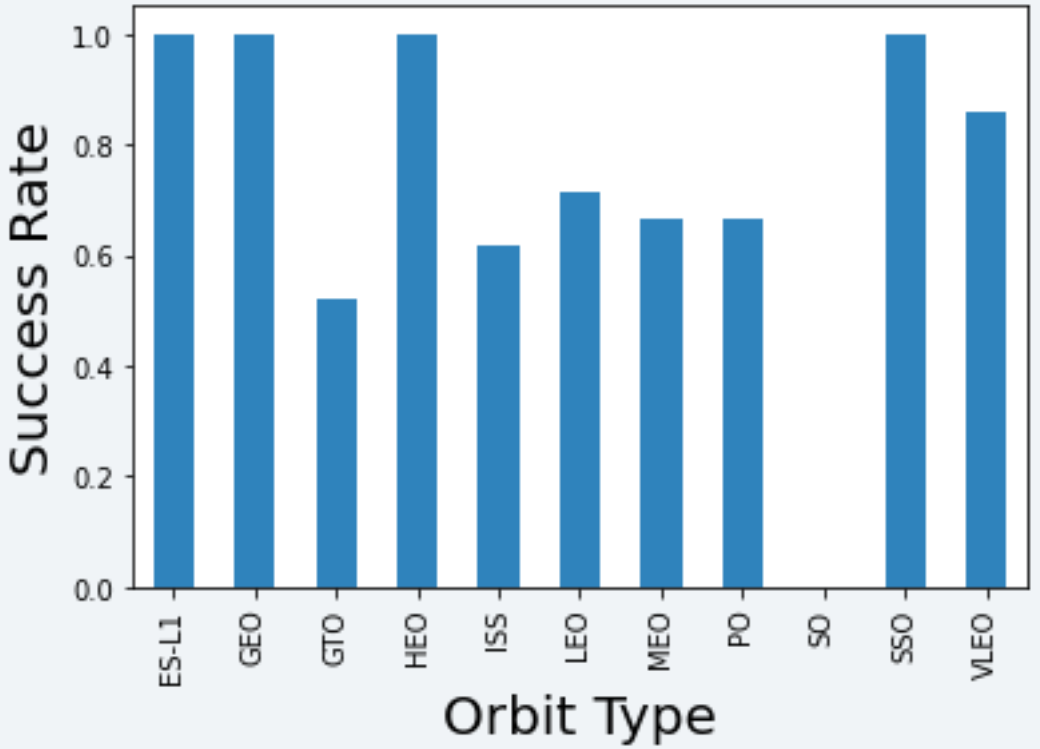
Payload vs. Launch Site



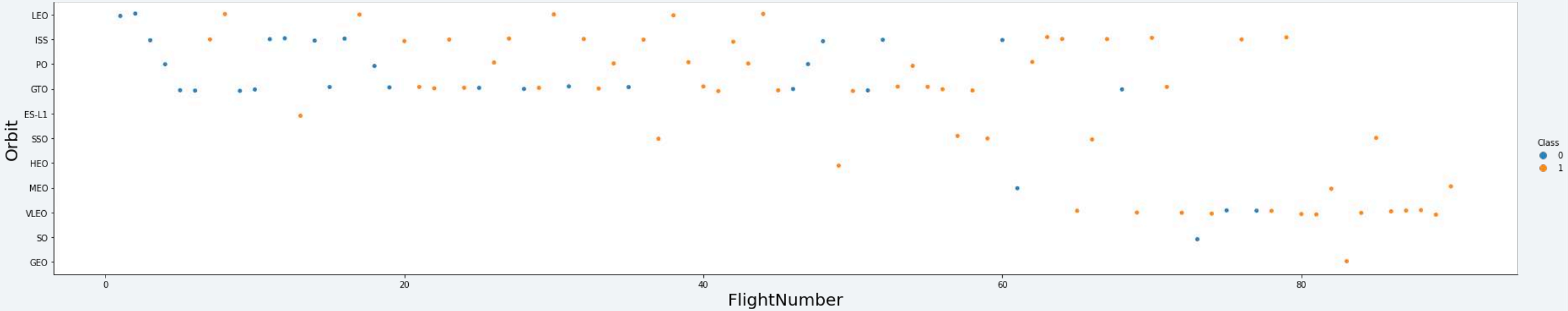
- For CCAF5 SLC 40 and KSC LC 39A lauchsites, payload mass less than 9,000 kg have a higher success rate

Success Rate vs. Orbit Type

- High success rate in orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
  - VLEO

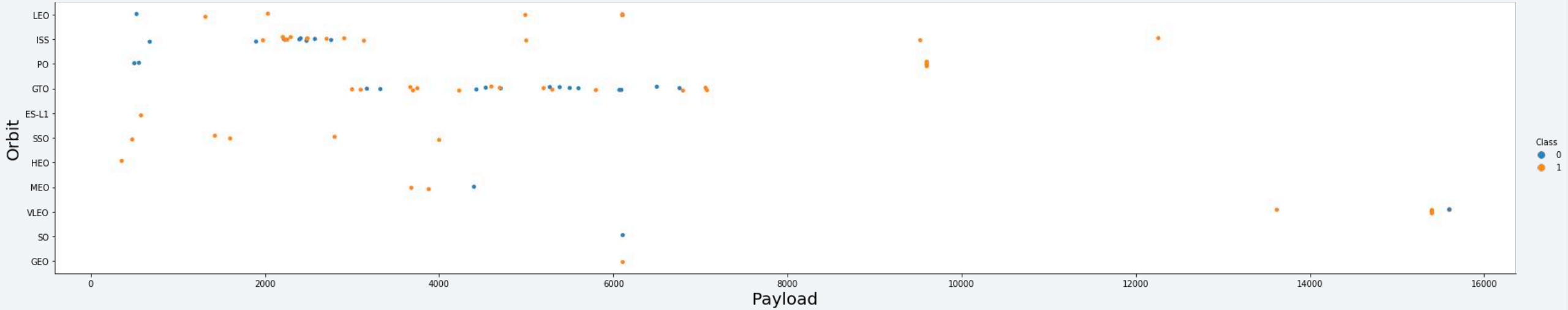


Flight Number vs. Orbit Type



- Lauch success rate correlates with flight number for LEO orbit.
- VLEO orbit has the most recent increasing of launch success

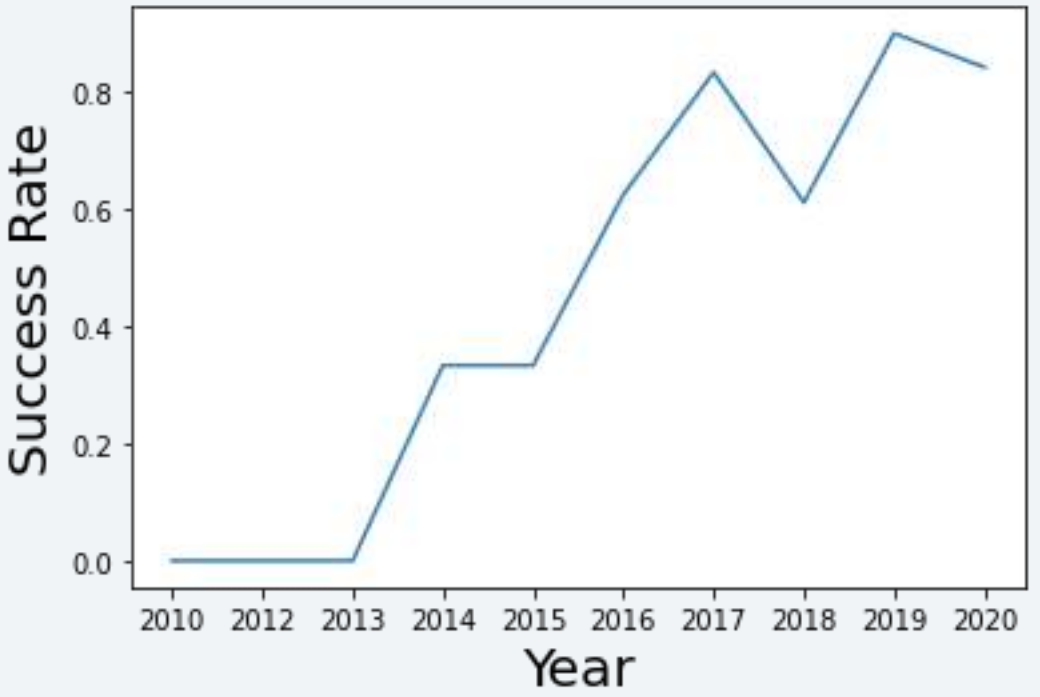
Payload vs. Orbit Type



- No positive or negative correlation for GTO orbit.
- With the successful landing rate for heavy payloads is higher in LEO and ISS orbits

Launch Success Yearly Trend

- Sucess rate increased since 2013 till 2020



All Launch Site Names

- Unique launch sites
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SDLC-4E

- obtained with `features['LaunchSite'].unique()`

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
In [16]: features.loc[features['LaunchSite'].str.startswith('CCA')].head()
```

Out[16]:

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004
5	6	3325.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1005

Total Payload Mass

- Total payload carried by boosters from NASA

total_payload
111268

Average Payload Mass by F9 v1.1

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

avg_payload
2928

First Successful Ground Landing Date

- First successful landing outcome on ground pad

first_success_gp
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship, with 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

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- Booster which have carried the maximum payload 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

- Failed landing_outcomes in drone ship 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

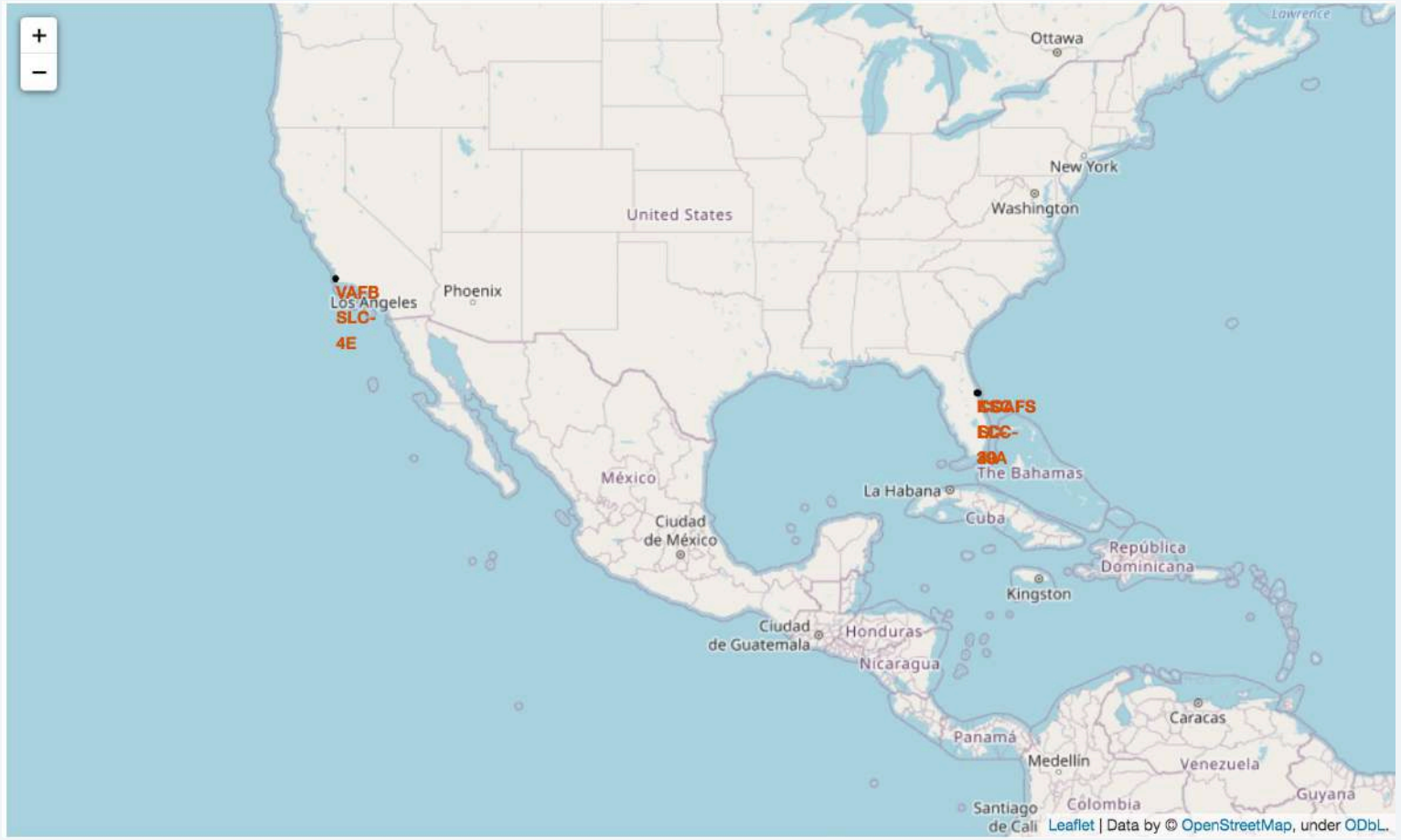
- Count of landing outcomes (Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

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

Section 3

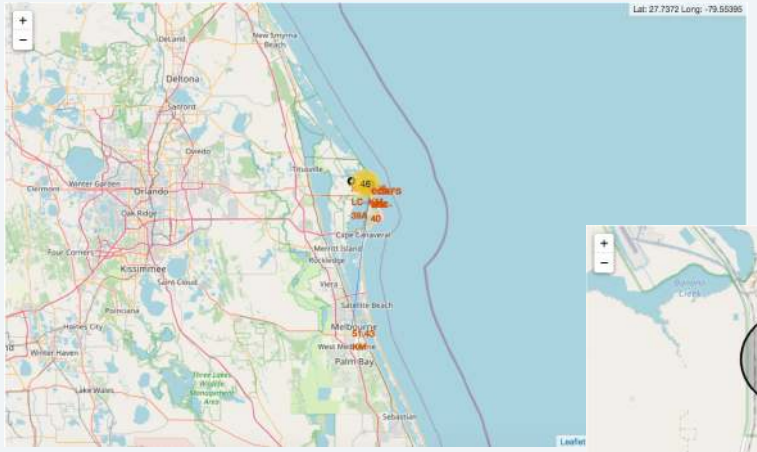
Launch Sites Proximities Analysis

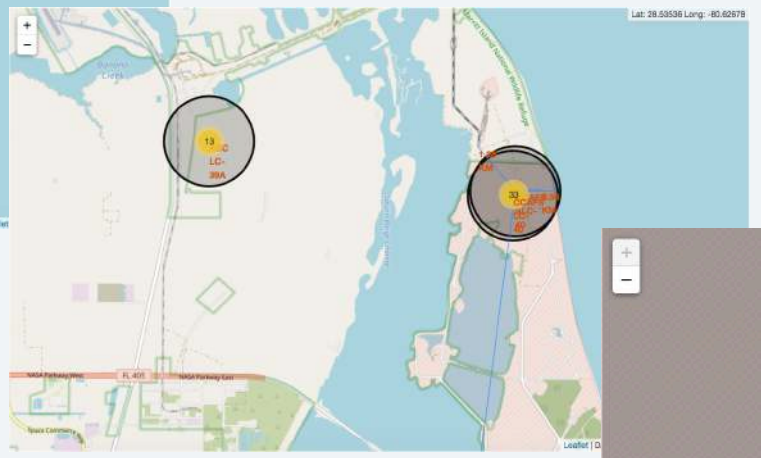
Launch sites on map

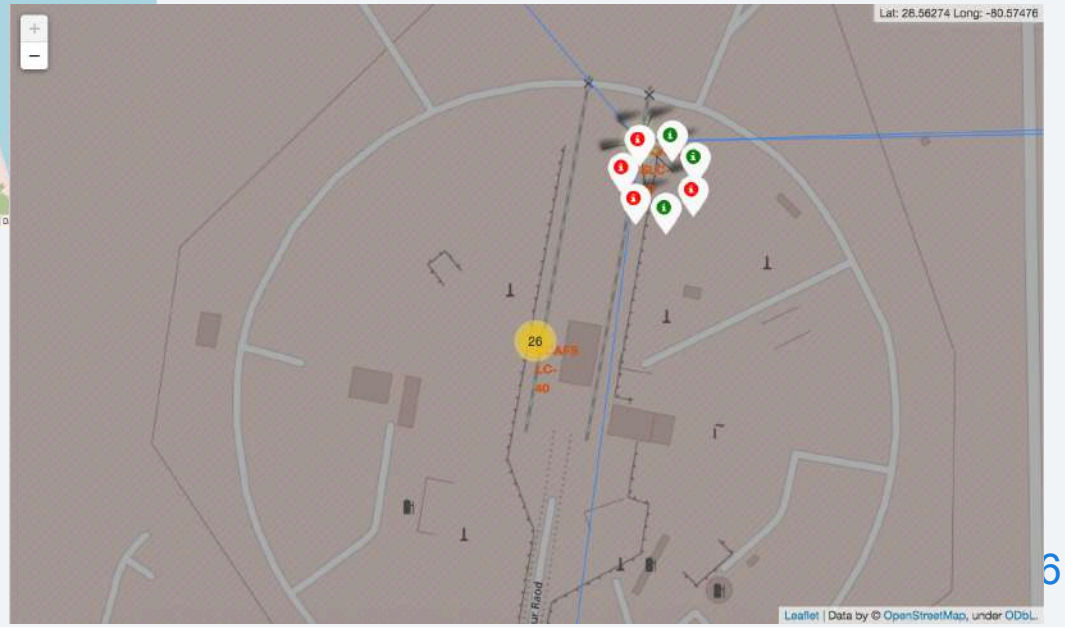


Launch sites are located on coastlines

Launch outcomes on Map



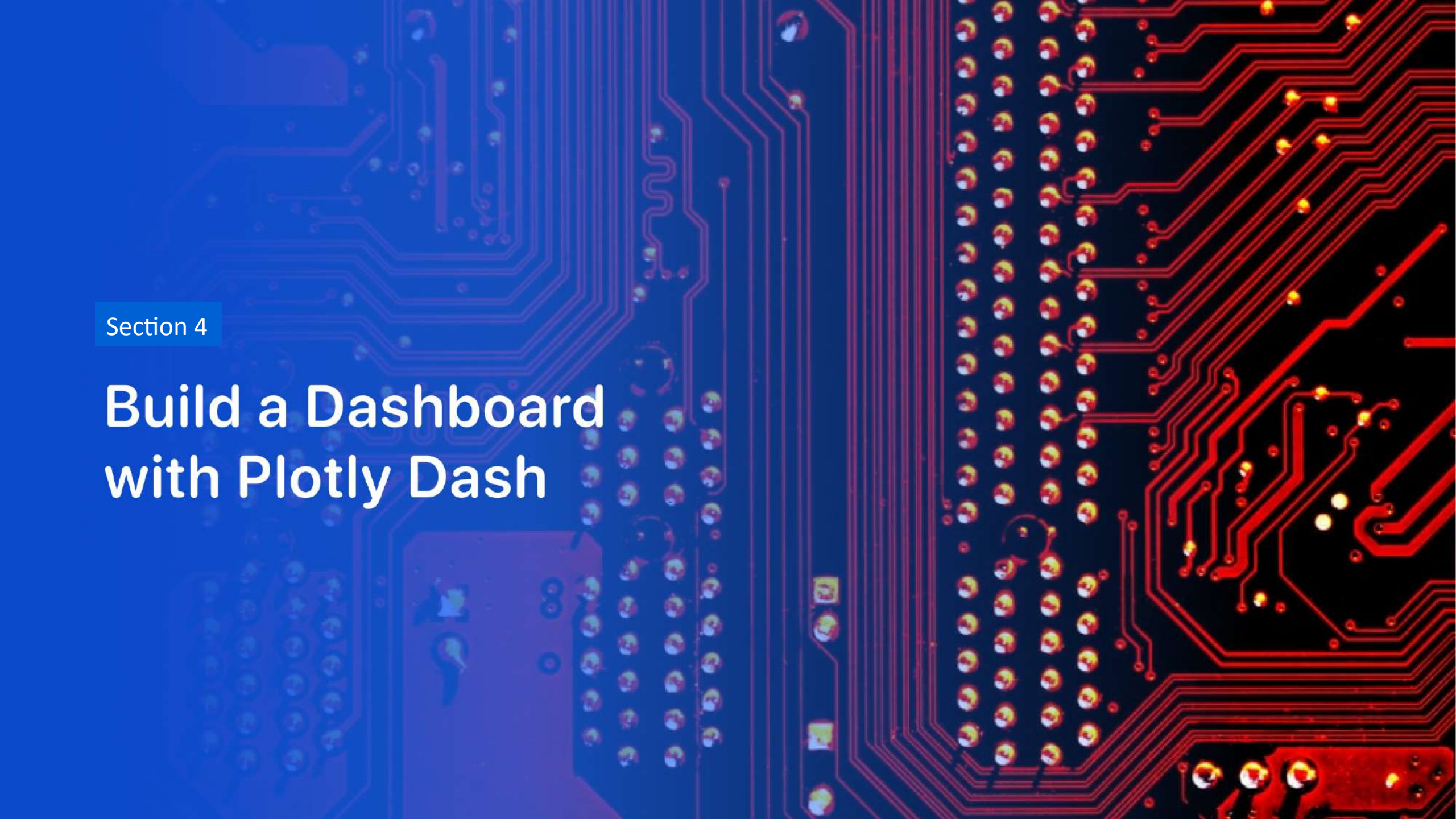




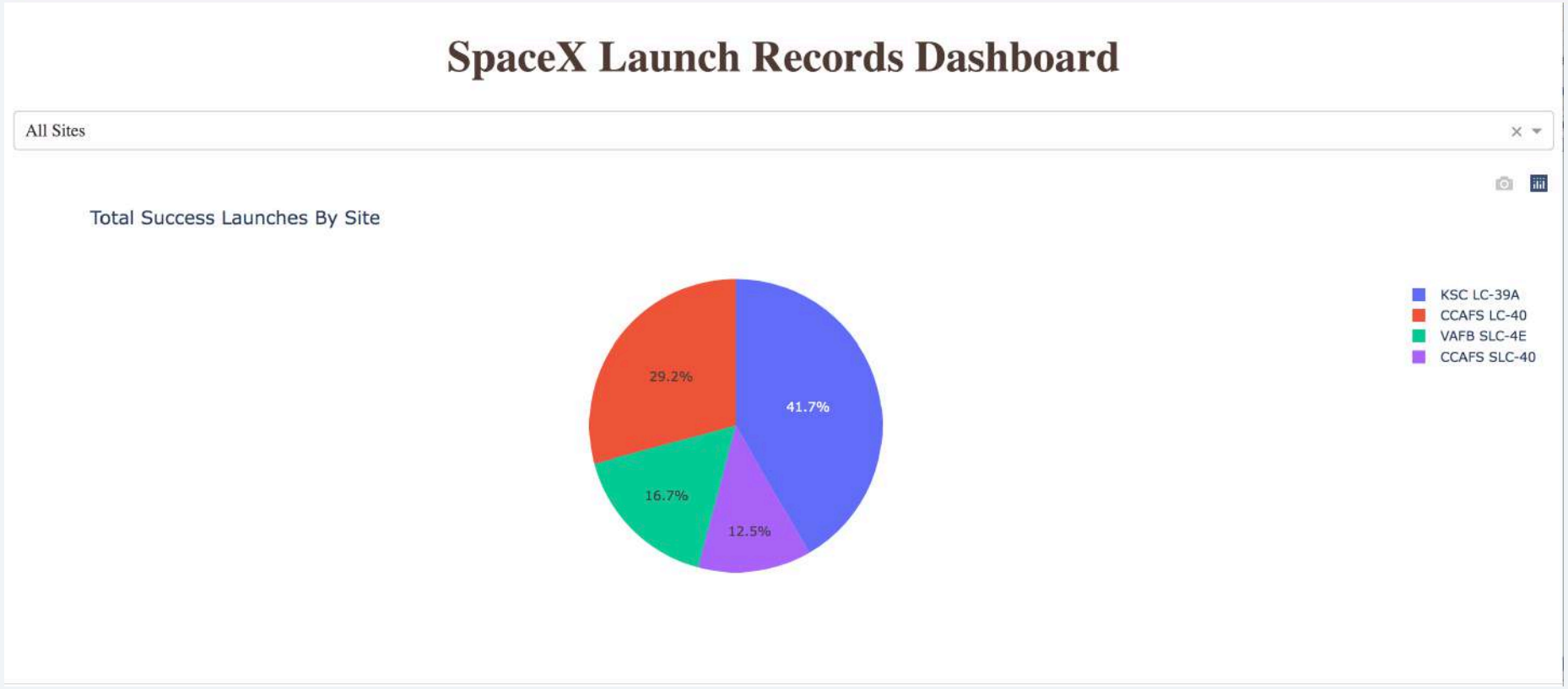
Green marks success and red failure launches

<Folium Map Screenshot 3>

- CCAFS site is less than one kilometer from coastline

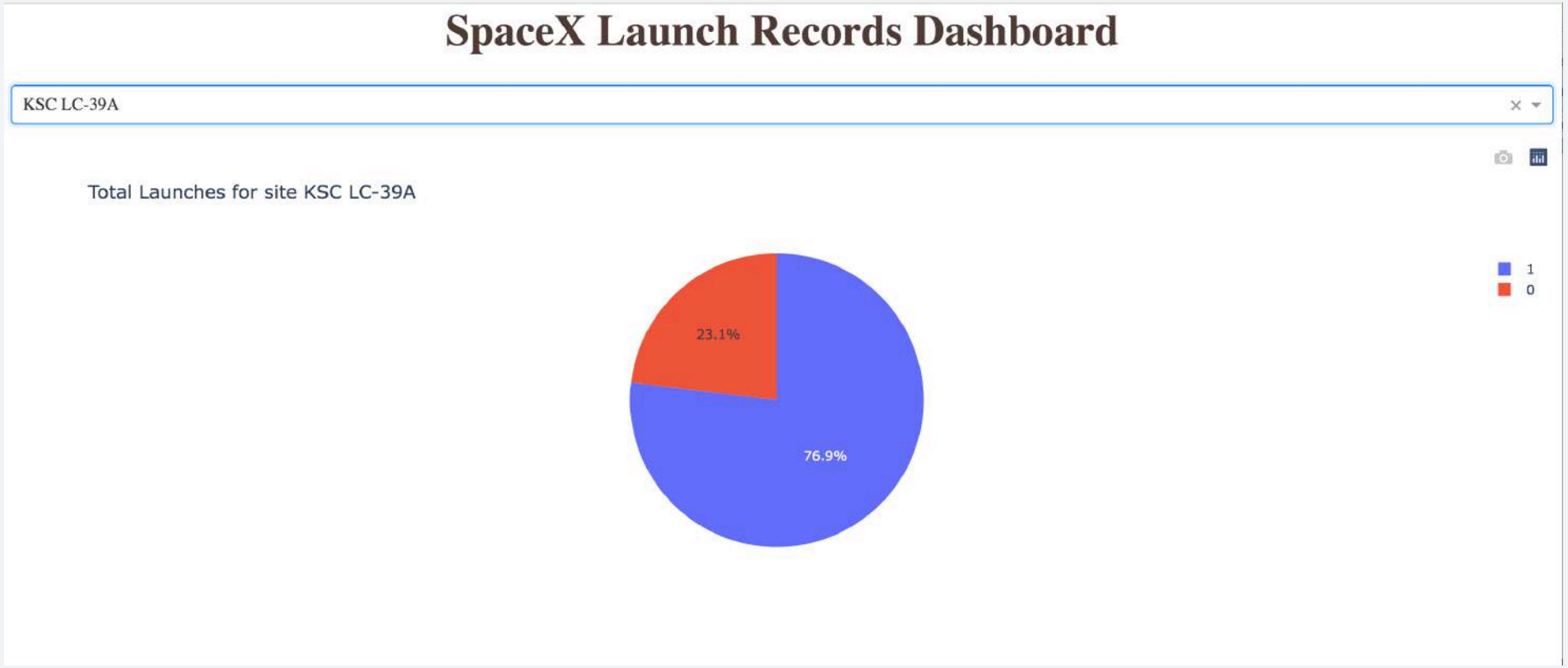

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all sites



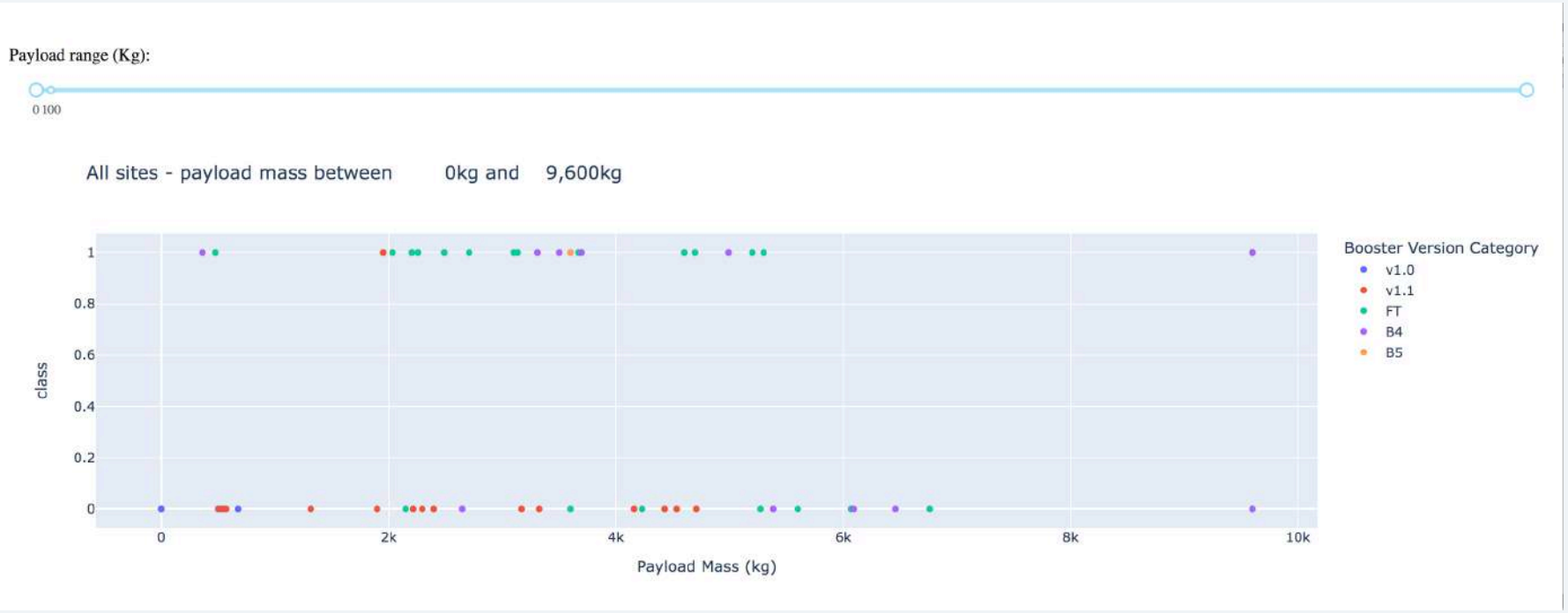
- Successful launches vary are related to launch sites

Launch site with highest launch success ratio



- KSC LC-39A is the site with highest success rate

Payload vs. Launch Outcome for all sites



- FT booster and payload less than 6.000 Kg is the most successful combination

Section 5

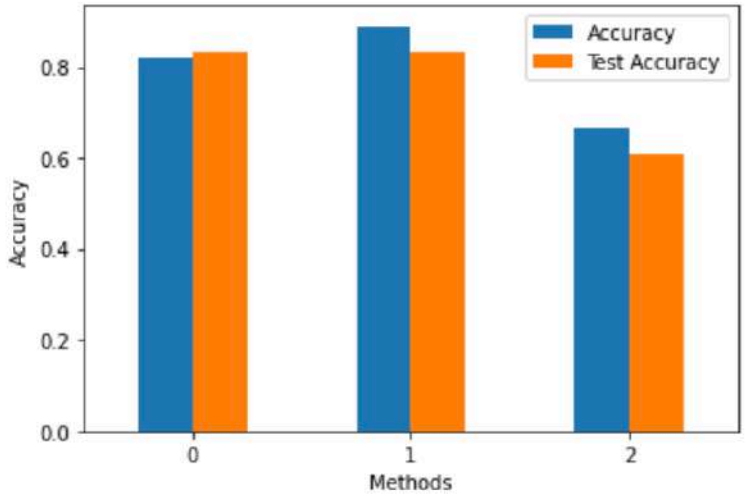
Predictive Analysis (Classification)

Classification Accuracy

Decision Tree model has the highest classification accuracy

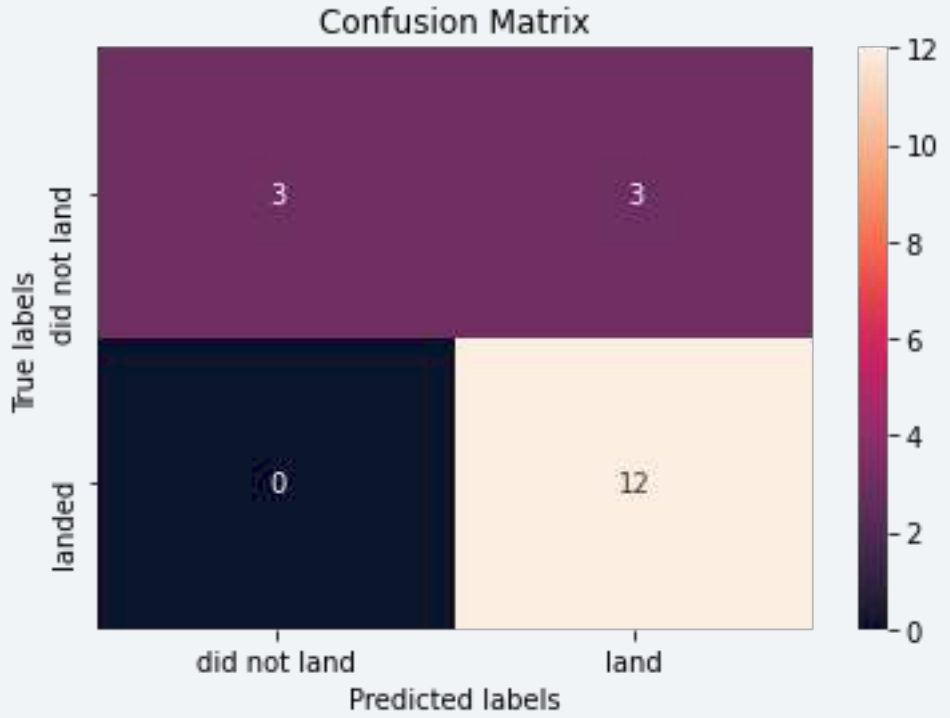
	Method	Accuracy	Test Accuracy
0	LogReg	0.820000	0.833
1	Tree	0.889286	0.833
2	KNN	0.664000	0.611

```
df.plot.bar(rot=0)
plt.xlabel("Methods")
plt.ylabel("Accuracy")
plt.show()
```



Confusion Matrix

- Confusion matrix of Decision Tree Classifier showing the higher number of True Positive and True Negative



Conclusions

- KSC LC 39A is the best launch site
- Success rate is improving over the time
- Decision Tree Classifier is the best ML model for prediction of successful landing

Appendix

- I was not able to perform SVM on my PC

Thank you!