Winning Space Race with Data Science

Md Ahsan Habib
09/01/2021

Outline

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

Executive Summary

- Data Collection Methodology
  - using get request to collect API data
  - parsing data and to json file
  - converting to pandas dataframe
- Perform data wrangling
  - desired columns were extracted from dataframe
  - finding missing value
  - filling in missing values with mean value of respective columns
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash
- Performing predictive analysis
  - building predictive model
  - train model model with train data
  - test model accuracy with test data

Introduction

- In this project we are trying to find out cost of SapceX falcon 9.
- Cost of this project depends on first landing since spacex reuse first stage.
- We want to build predictive model
- To predict whether this first stage landing will be successful

Section 1

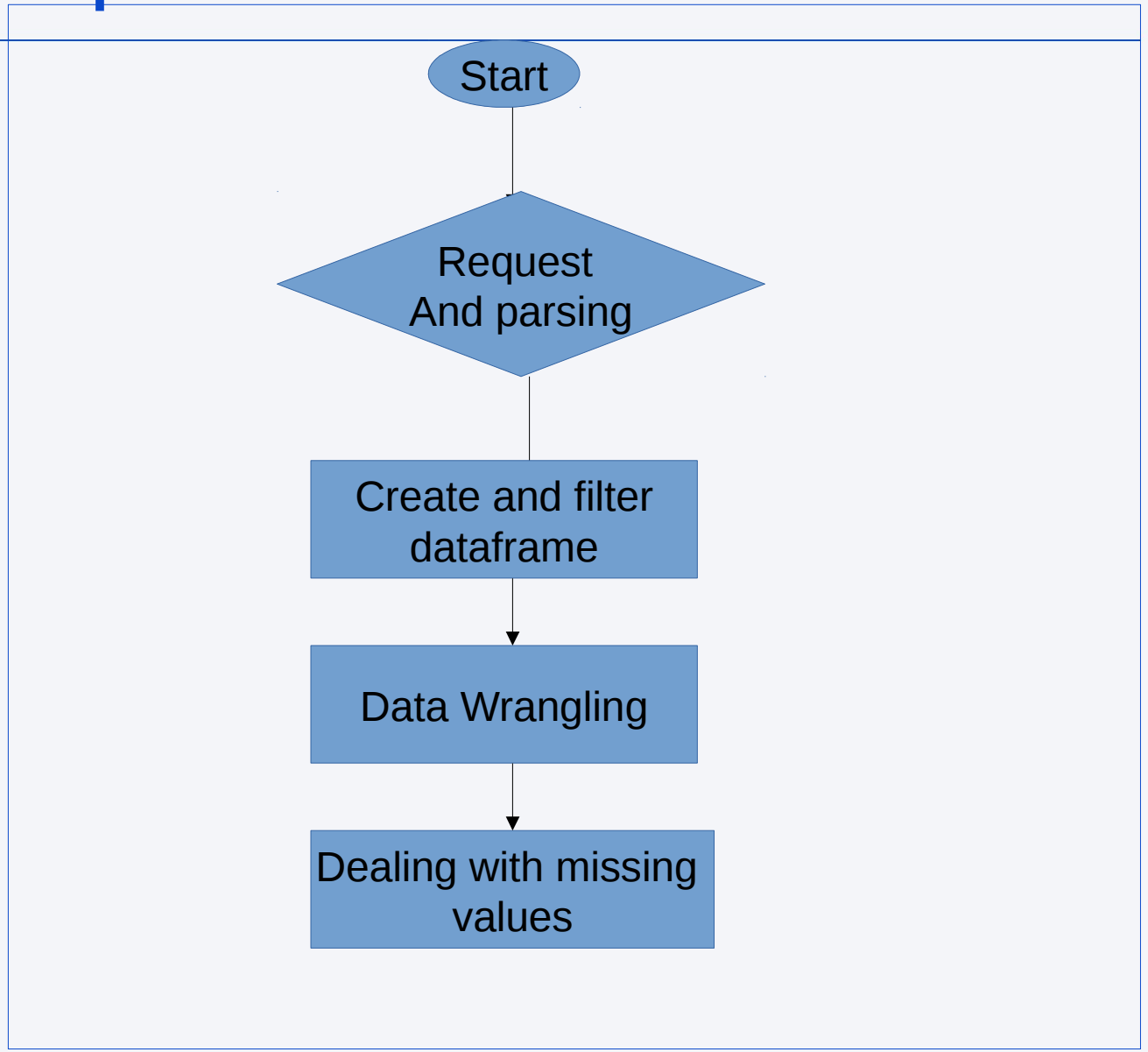
Methodology

Data Collection

- Using webscraping method we request and parse data.
- Convert response to json text.
- Convert this jason text to pandas dataframe.
- Filter dataframe to get on Falcon 9 data.
- Data wrangling.
- Filling in missing data with mean value

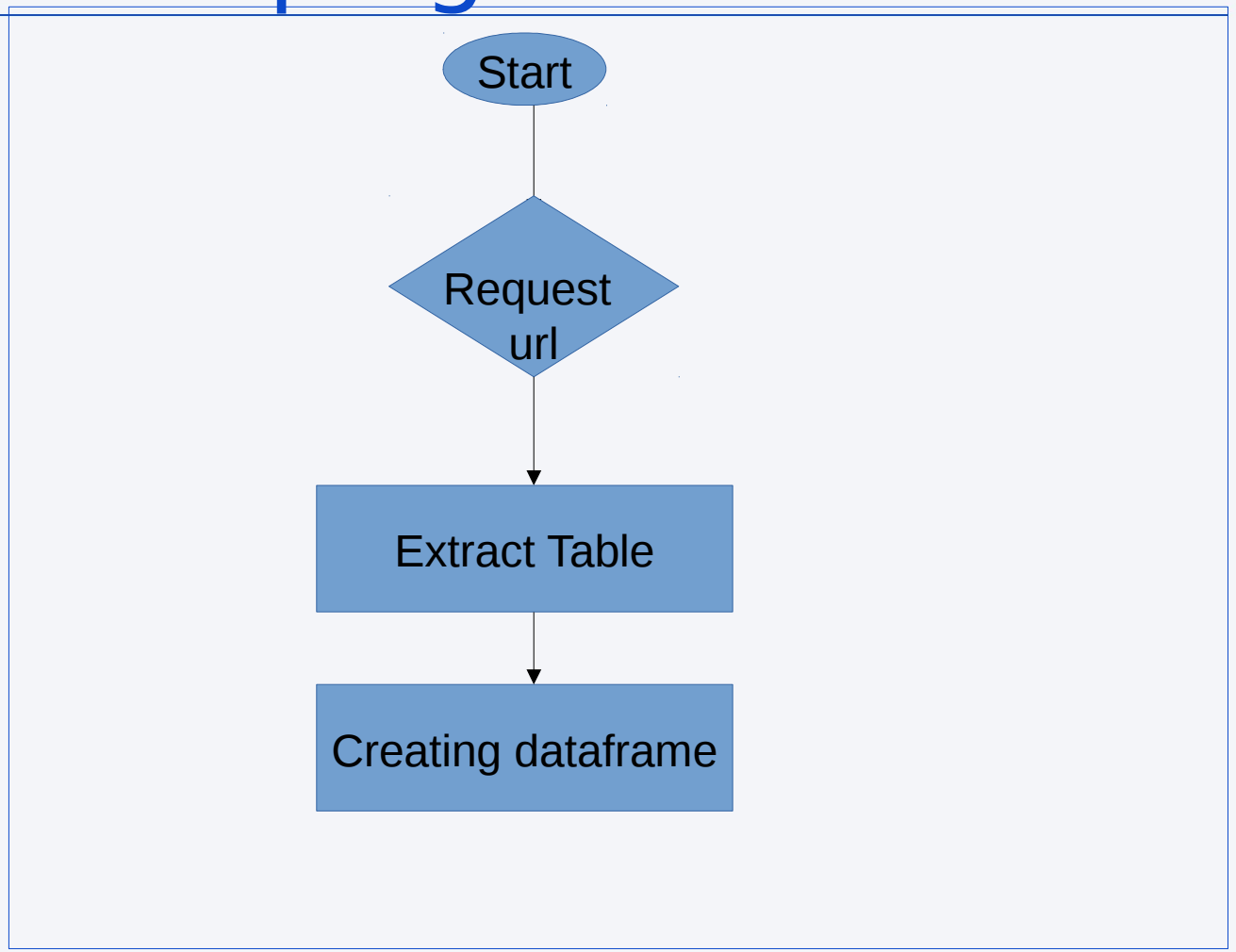
Data Collection – SpaceX API

- GitHub URL for this script
- https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- GitHub URL for the project
- https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/jupyter-labs-webscraping.ipynb

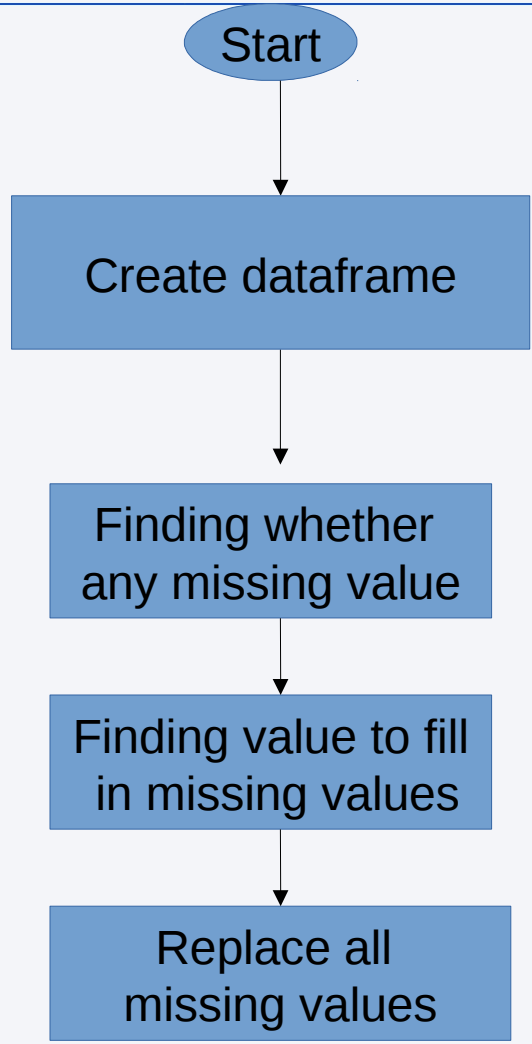


Interactive map with folium

- Interactive map with folium
- Map initiated
- All launch sites were marked
- Successful and failed launches for each site
- Distance of nearby from launch sites
-
- Github url for script https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/lab_jupyter_launch_site_location.ipynb

Data Wrangling

- GitHub URL for this script
- https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/Data%20wrangling.ipynb



EDA with Data Visualization

Charts plotted
- payload vs flight number
- launch site location vs flight number
- launch site vs payload mass
- orbit vs success rate
- orbit vs payload mass
- orbit vs flight number
- success rate vs year

Github url for the script
https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Sql command performed
- Unique
- Sum, Avg, Min
- Where
- Like
- Between
- Subquery
- Count
- And
- Group by
- Order by
- Desc
- Github url for script https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/spacex.sql

Interactive map with folium

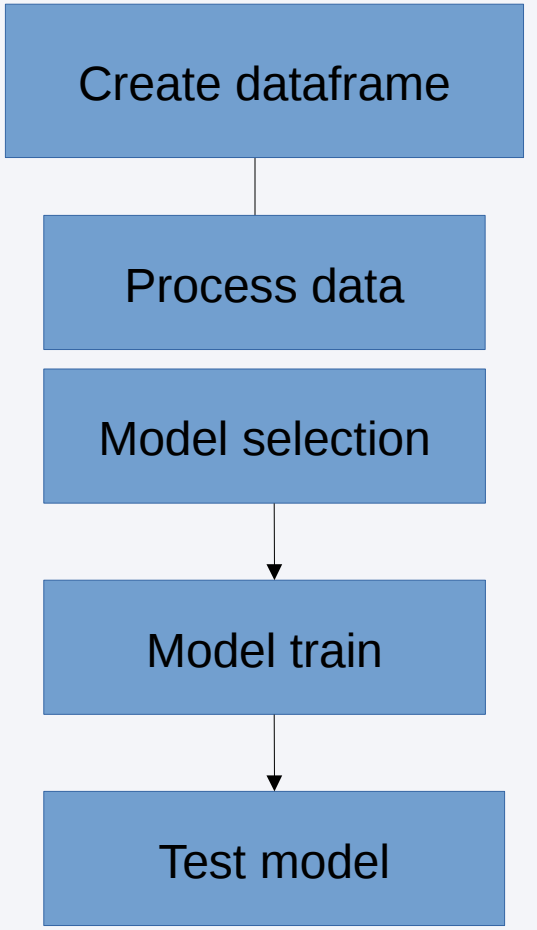
- Interactive map with folium
- Map initiated
- All launch sites were marked
- Successful and failed launches for each site
- Distance of nearby from launch sites
-
- Github url for script https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dash board with plotly
- Drop down for launch site was created
- Pie chart graph for success rate was added
- Range slider was added to select payload range
- Scatter figure for payload mass added
- Callback function added for launch site as an input and pie chart as output
- Another call back was added payload mass
-
- Github url for script https://github.com/ahsan07habib22/IBM_DataScience_Capstone_project/blob/main/capstone_dash.ipynb

Predictive Analysis (Classification)

- Predictive analysis
- Data imported and converted to pandas dataframe
- Required columns extracted
- Data were standardized
- Train test data were splitted
- Classification models were built with train data
- Built models were applied to test data for finding accuracy
- Model with best accuracy chosen
- Confusion matrix plotted for best model
- Github url for script http://localhost:8889/notebooks/qualifier/jobs/capstone_project/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

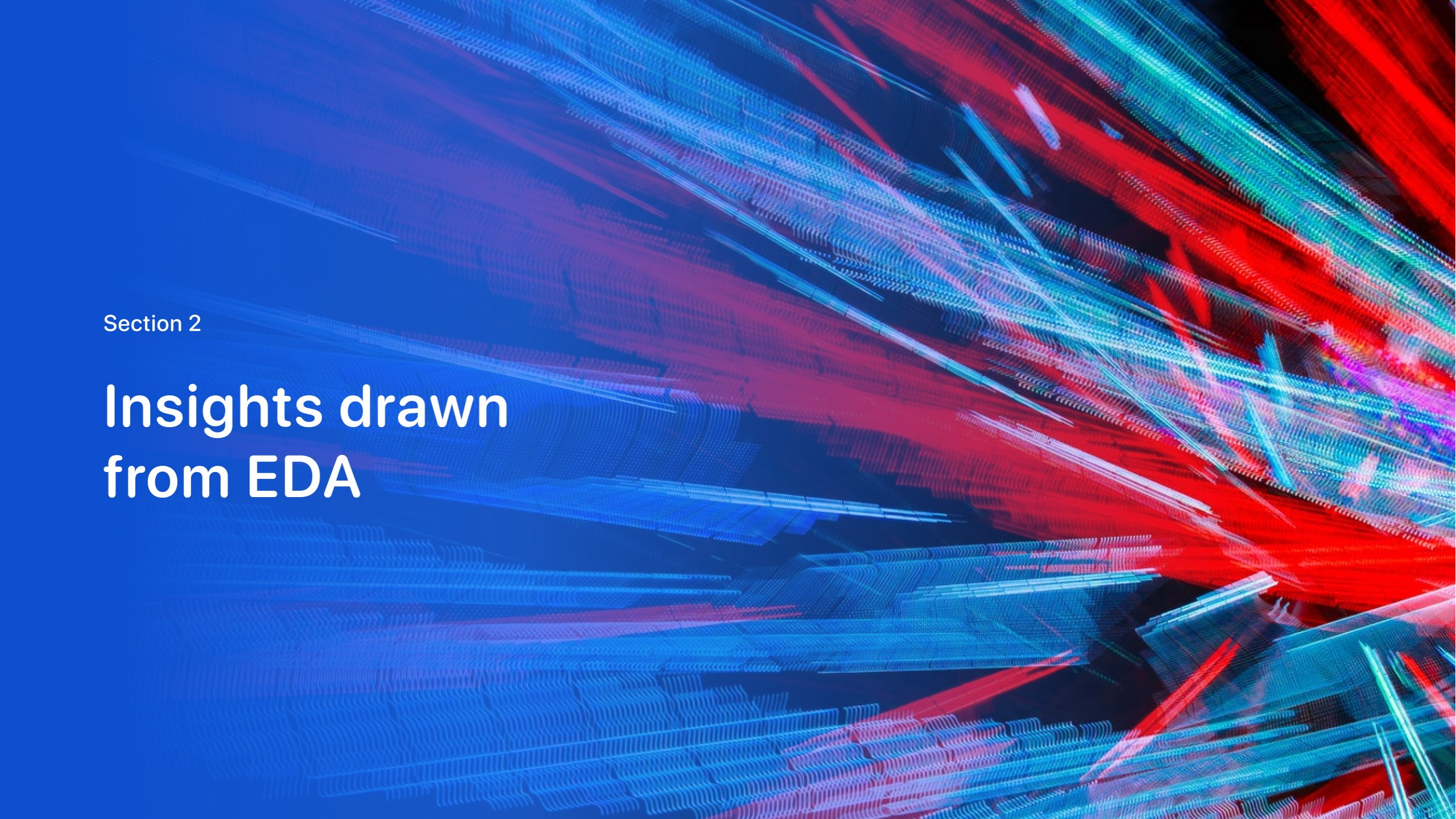


Results

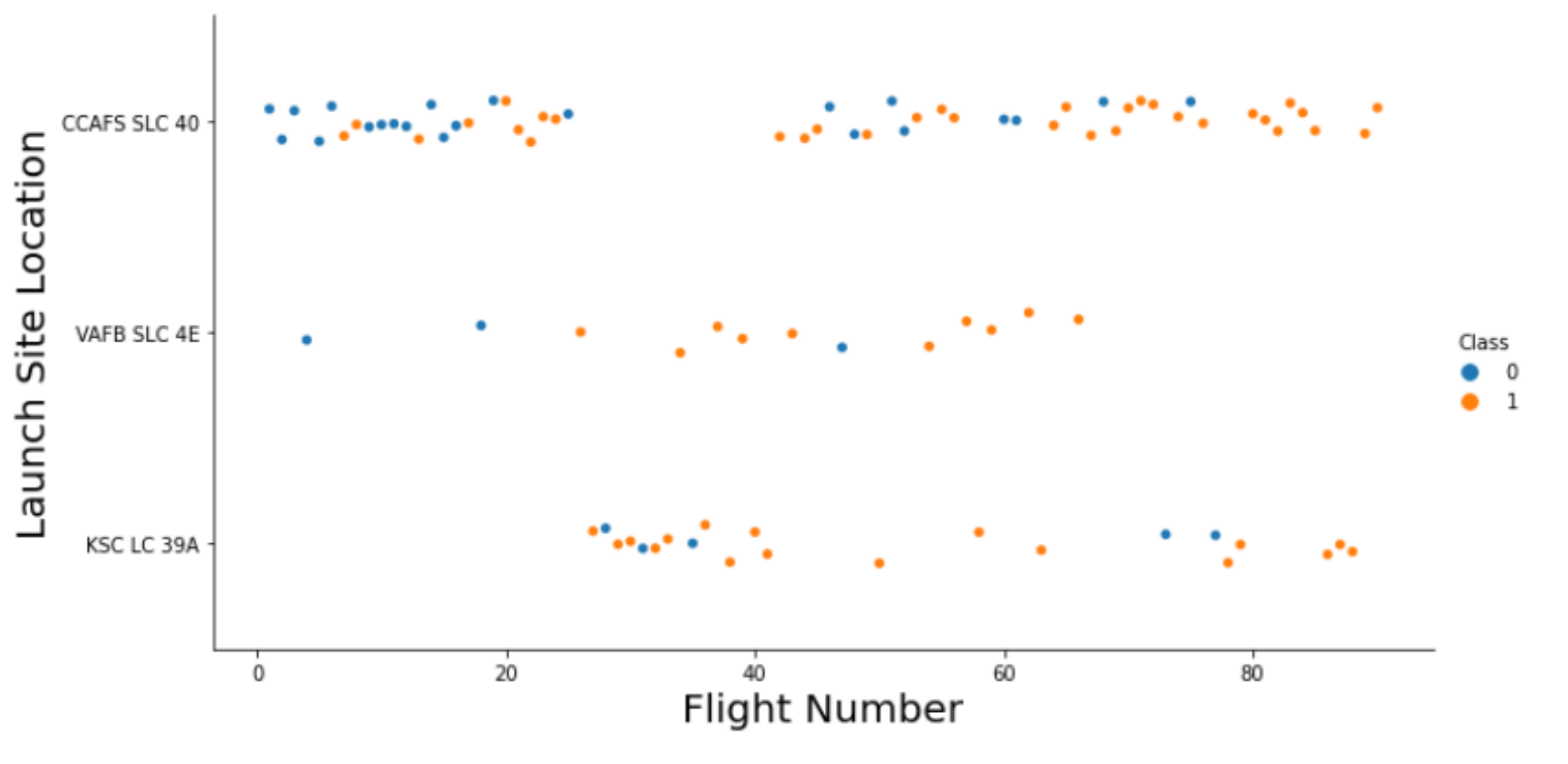
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

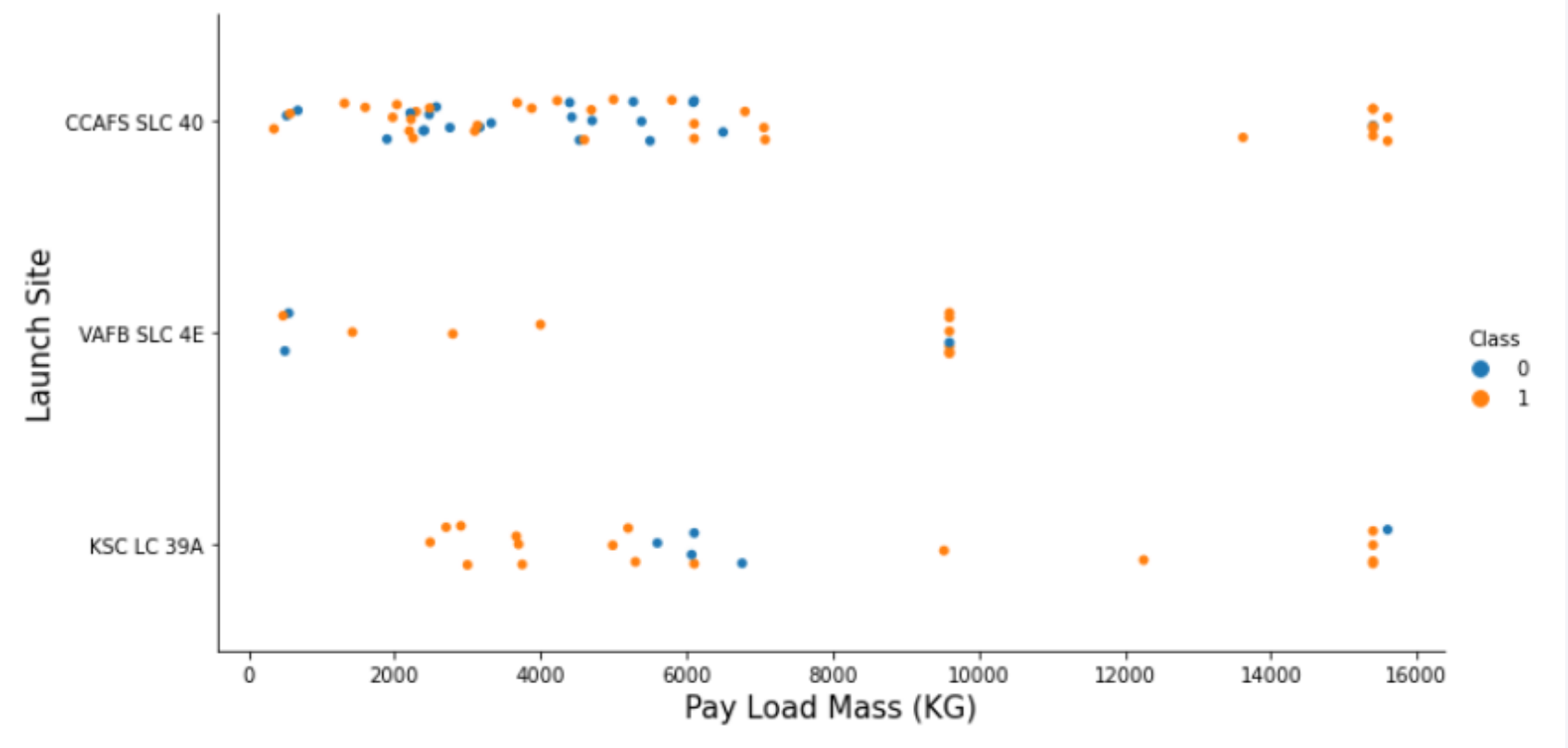
Insights drawn from EDA



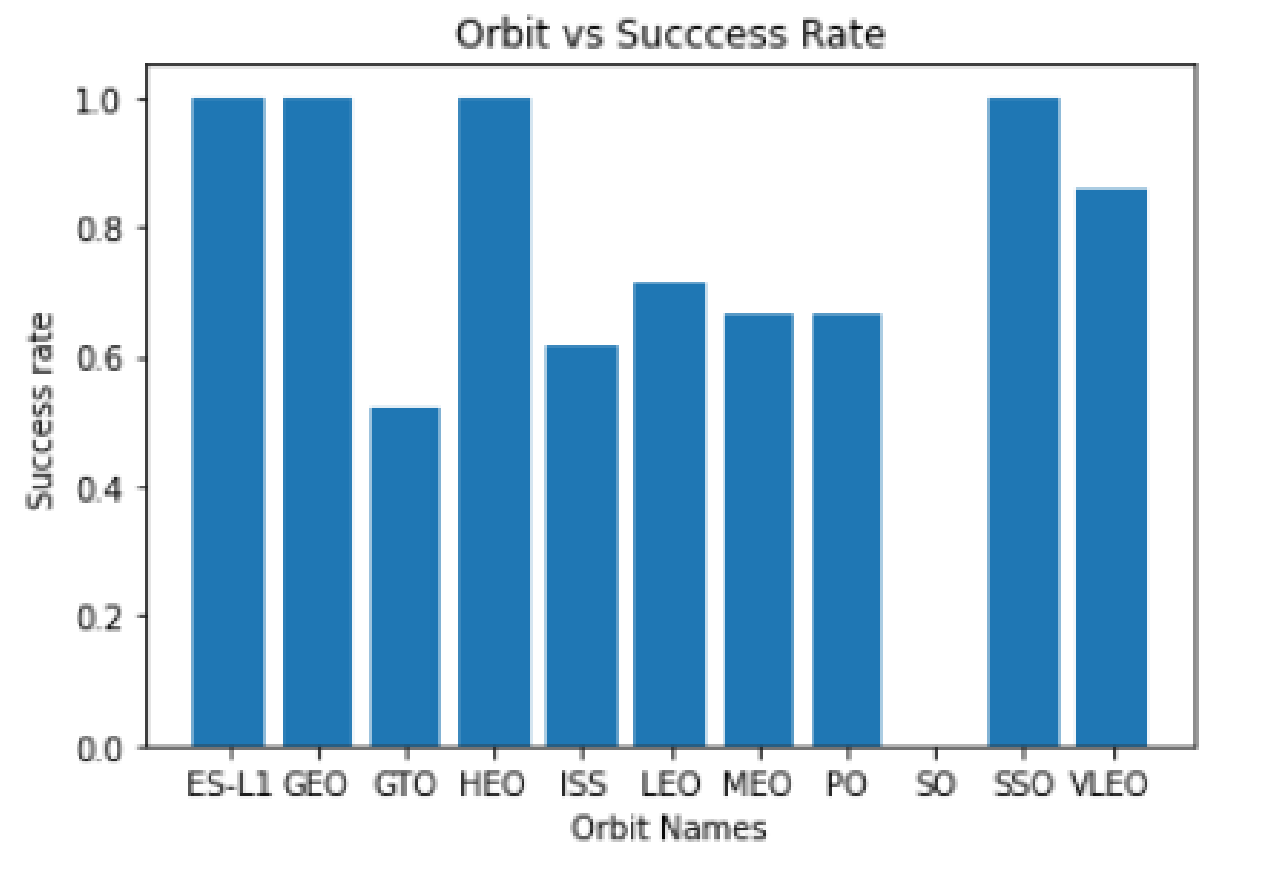
Flight Number vs. Launch Site



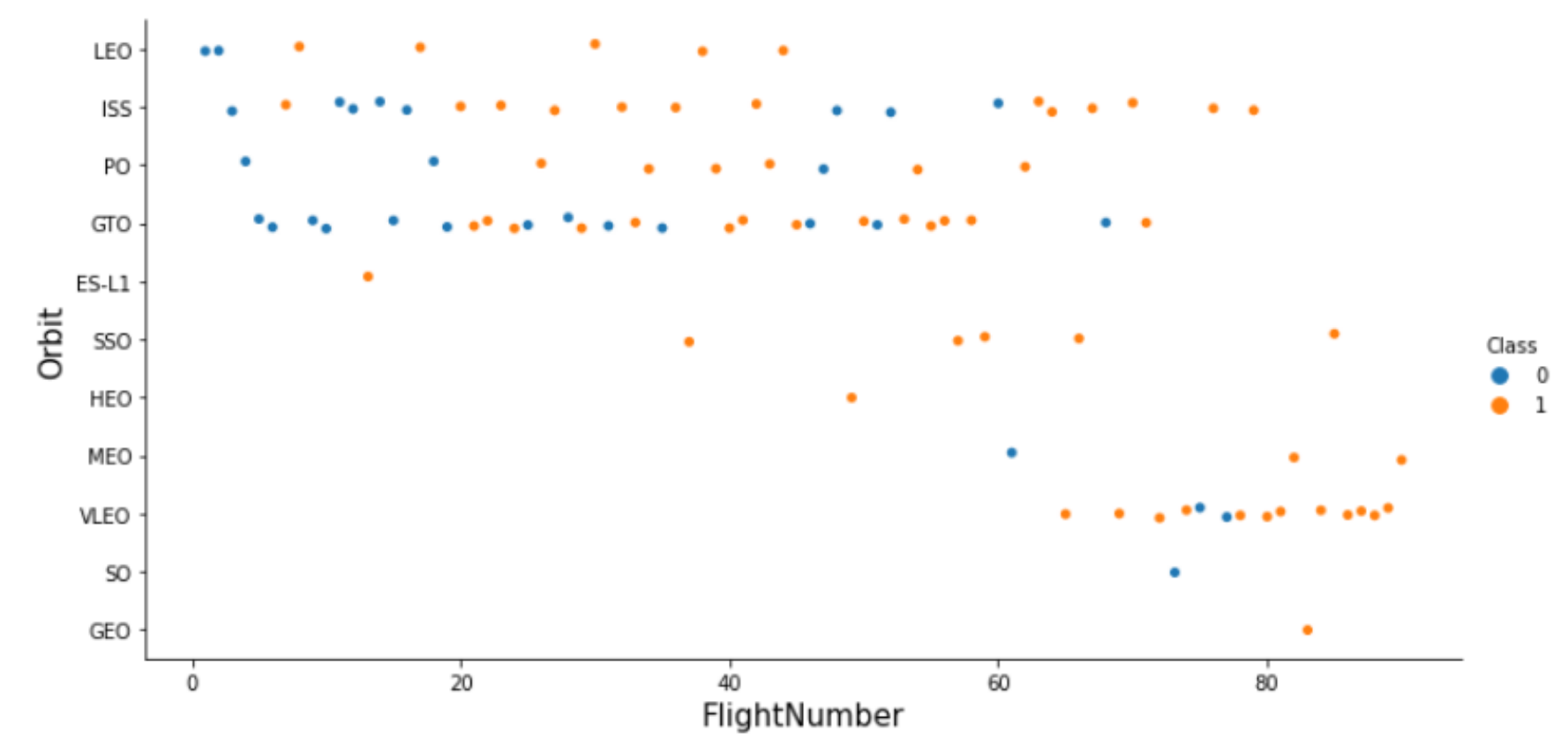
Payload vs. Launch Site



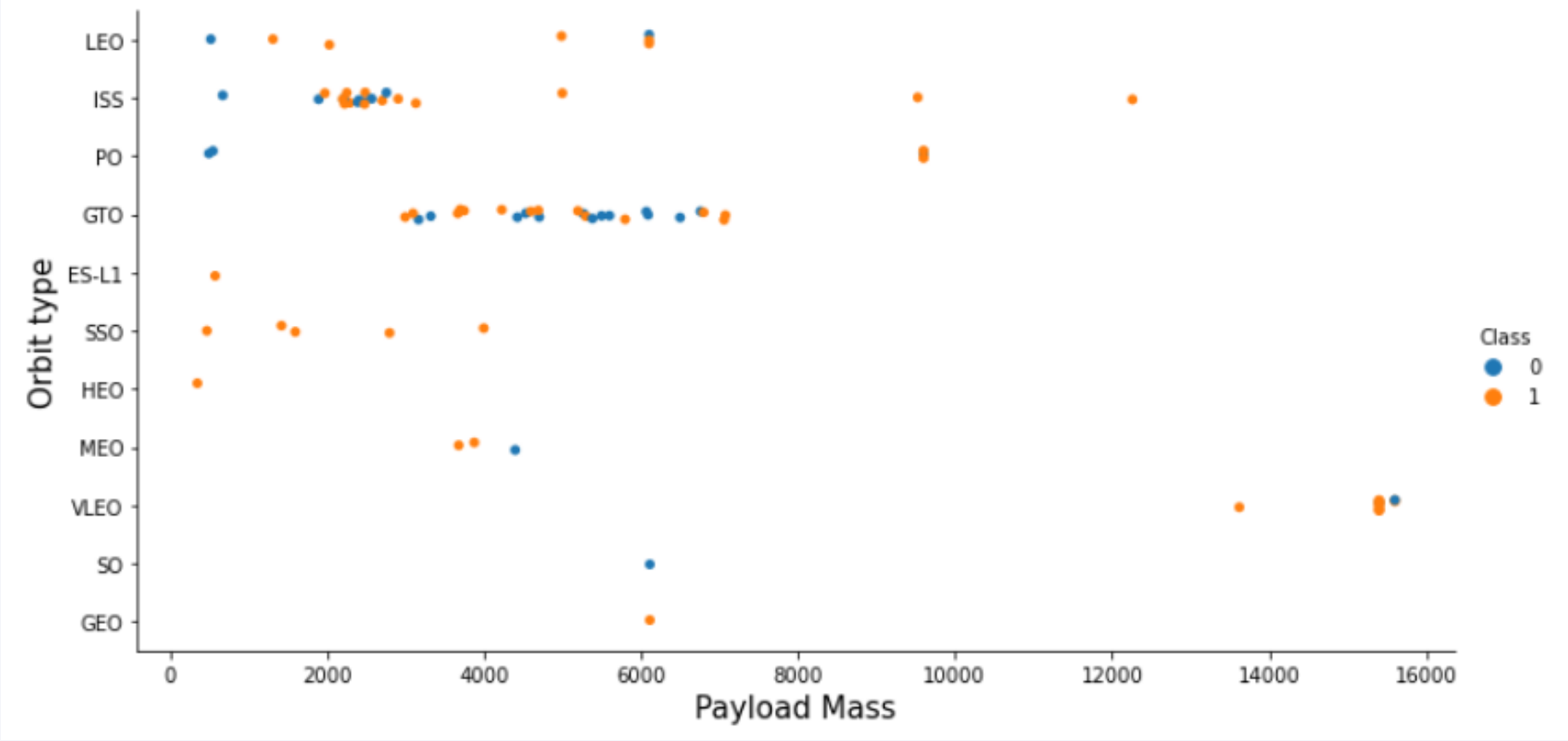
Success Rate vs. Orbit Type



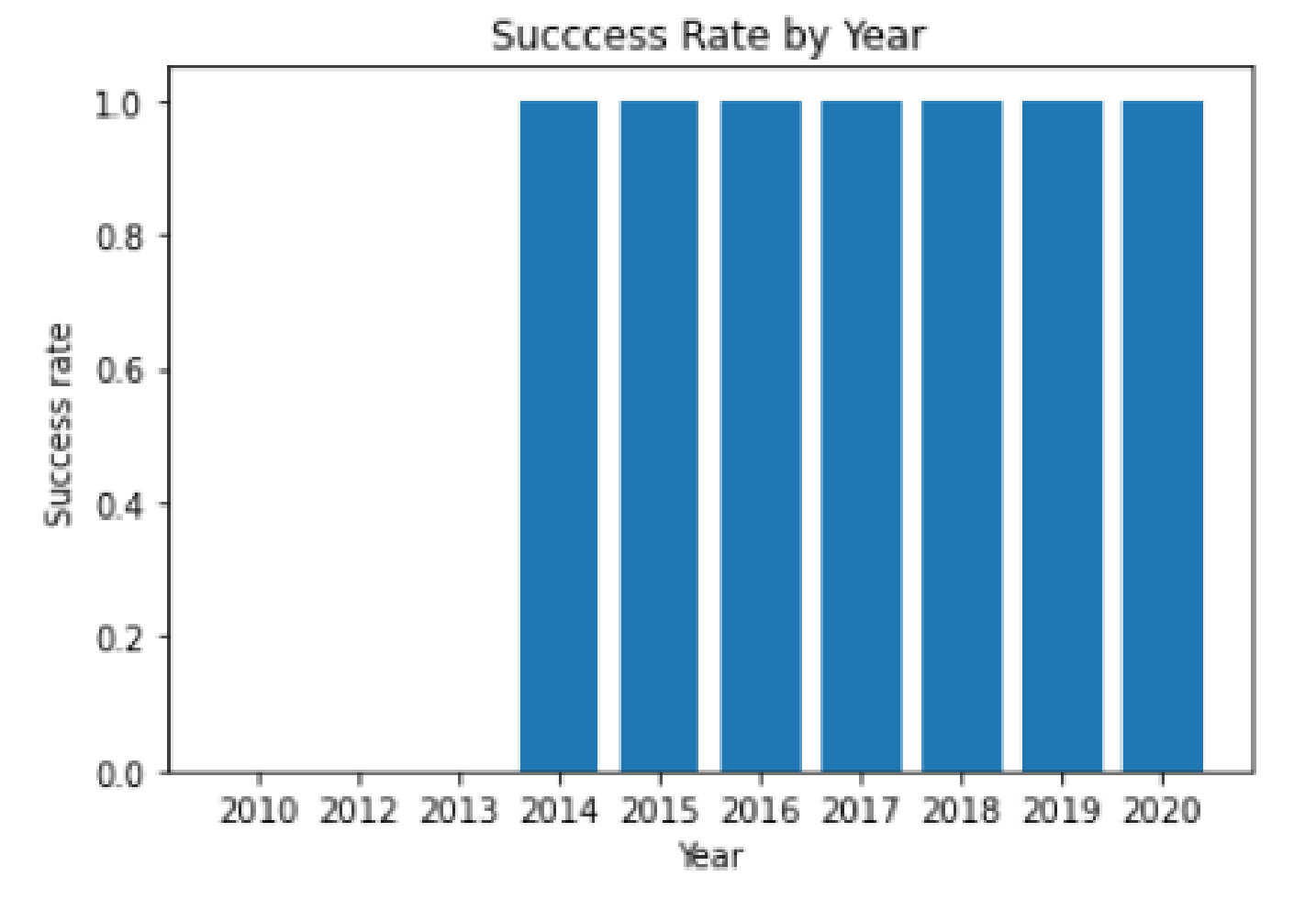
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



EDA with SQL

All Launch Site Names

%%sql
select unique(launch_site) from spacextable;
Present your query result with a short explanation here.

Result set 1

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

Launch site names begin with `CCA`

- Find all launch sites begin with `CCA`

```
%%
select launch_site from spacextable
where launch_site like 'CCA%' limit 5;
```

LAUNCH_SITE
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total payload mass

Calculate the total payload carried by boosters from NASA

- %%sql

```
select sum(payload_mass__kg_) as sumpayload
from spacextable
where payload like '%CRS%';
```

Result set 1

SUMPAYLOAD
56479

Average payload mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

- %%sql

```
select avg(payload_mass__kg_) from spacextable
where booster_version like 'F9 v1.1';
```

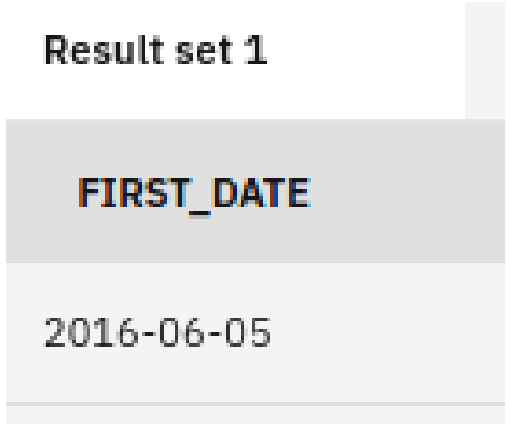
Result set 1

AVG_MASS
3676

First successful ground landing date

Find the date when the first successful landing outcome in ground pad.

```
%%sql
select unique(mission_outcome) from spacextable;
select min(date) from spacextable
where landing__outcome like '%Success%';
```



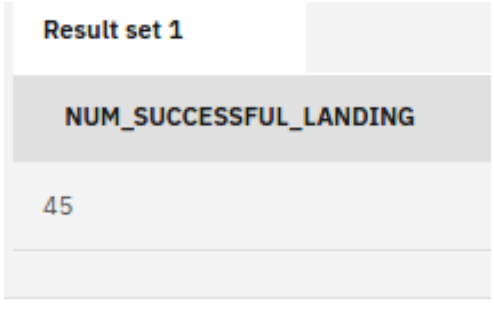
Successful drone ship landing with payload between 4000 and 6000

```
%%sql
select booster_version from spacextable
where payload_mass__kg_ between 4000 and 6000
and  landing__outcome = 'Success (drone ship)';
```

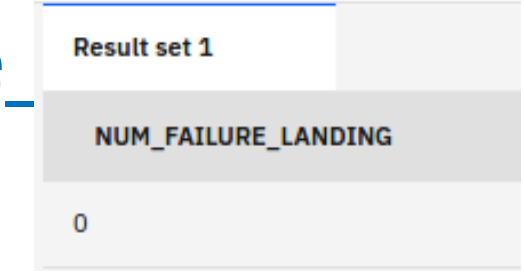
BOOSTER_VERSION	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600

Total number of successful and failure mission outcomes

```
select count(mission_outcome) as
num_successful_landing
from spacextable
where mission_outcome like '%Success%';
```




```
select count(mission_outcome) as num_failure_
from spacextable
where mission_outcome like '%Failure%';
```



Boosters carried maximum **payload**

```
select booster_version, payload_mass__kg_ from spacextable
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextable);
```

-
-

BOOSTER_VERSION	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600

2015 launch records

```
select landing__outcome, booster_version, launch_site,
DATE
from spacextable
where landing__outcome = 'Failure (drone ship)'
and date like '%2015%';
```

LANDING__OUTCOME	BOOSTER_VERSION	LAUNCH_SITE	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-10-01

Rank success count between 2010-06-04 and 2017-03-20

```
select count(*) as landout, date from spacextable
where date between '2010-06-04 ' and '2017-03-20'
and landing__outcome = 'Success (ground pad)'
or landing__outcome = 'Failure (drone ship)'
group by date
order by date
desc;
```

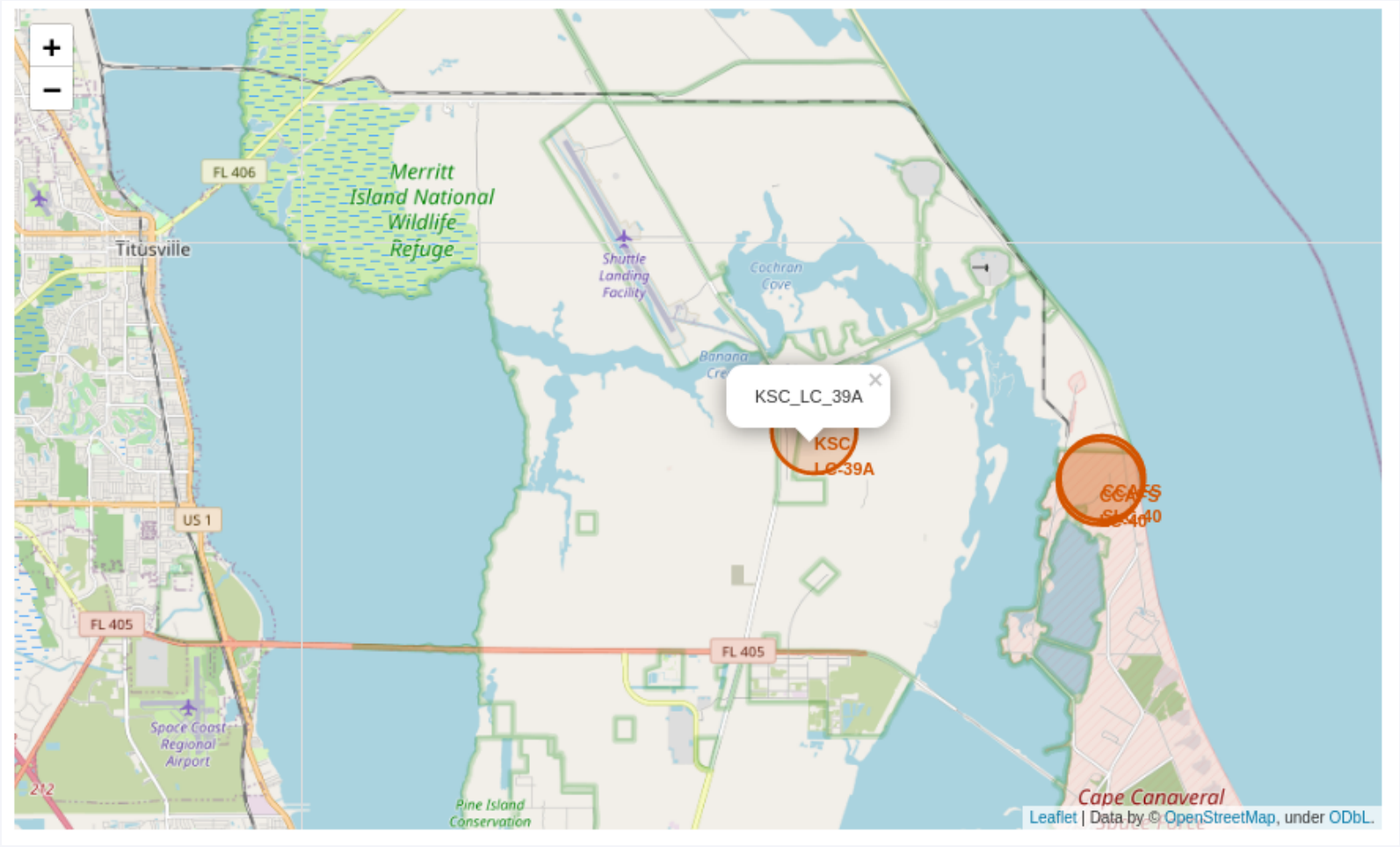
LANDOUT	DATE
1	2017-03-06
1	2017-01-05
1	2016-04-03
1	2015-10-01

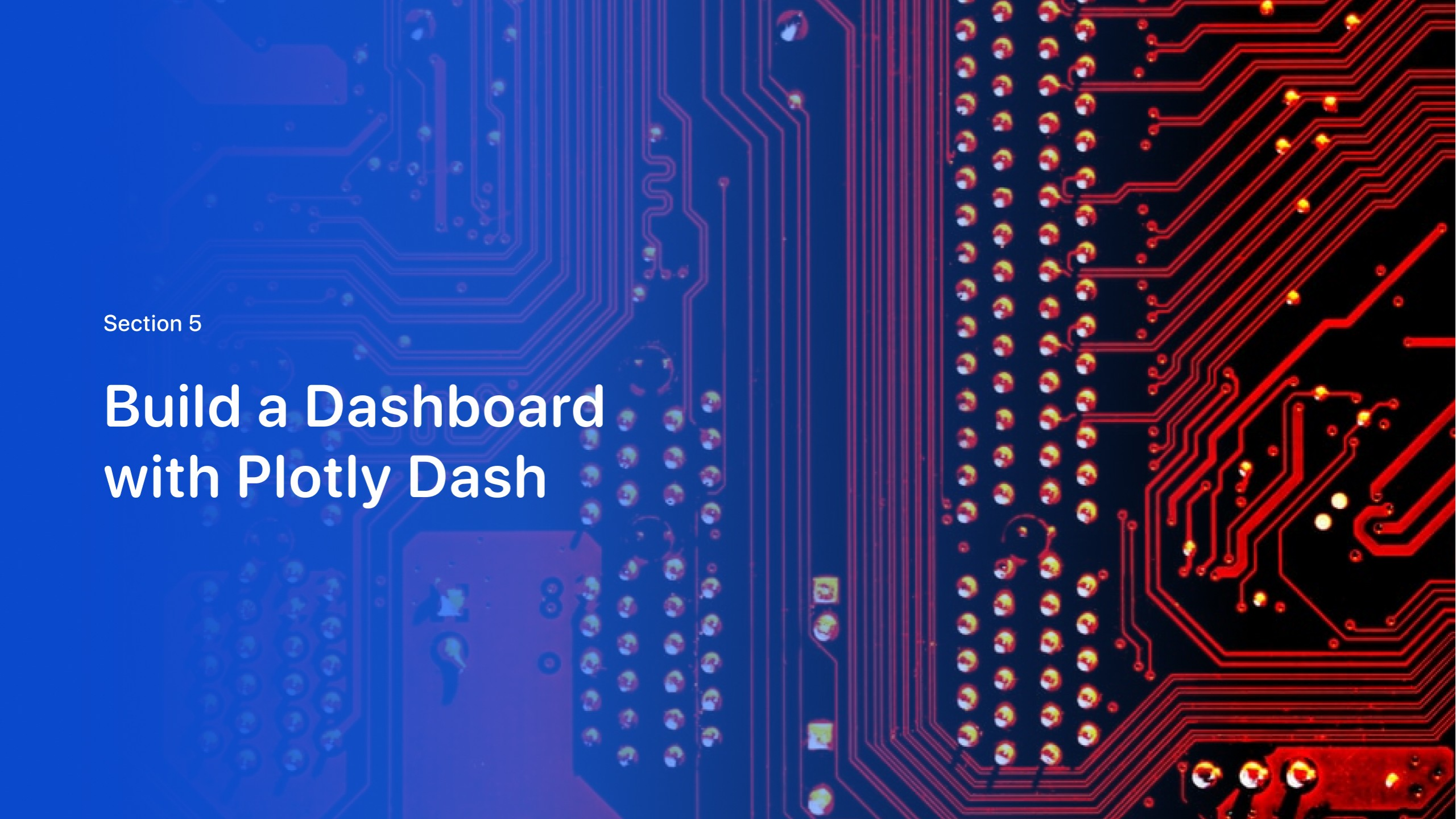
Section 4

Launch Sites Proximities Analysis

Folium Map Launch Site

Folium Map

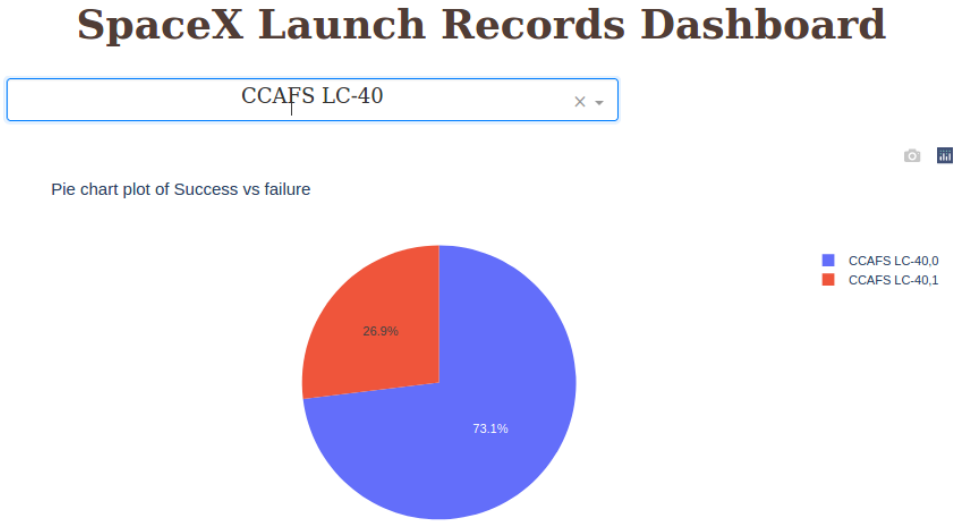

KSC_LC_39A
Merritt Island National Wildlife Refuge
Titusville
FL 406
US 1
FL 405
FL 405
Shuttle Landing Facility
Cochran Cove
Banana Cre
KSC LC-39A
CCAFS SLC-40
Space Coast Regional Airport
Pine Island Conservation
Cape Canaveral
Leaflet | Data by © OpenStreetMap, under ODbL.

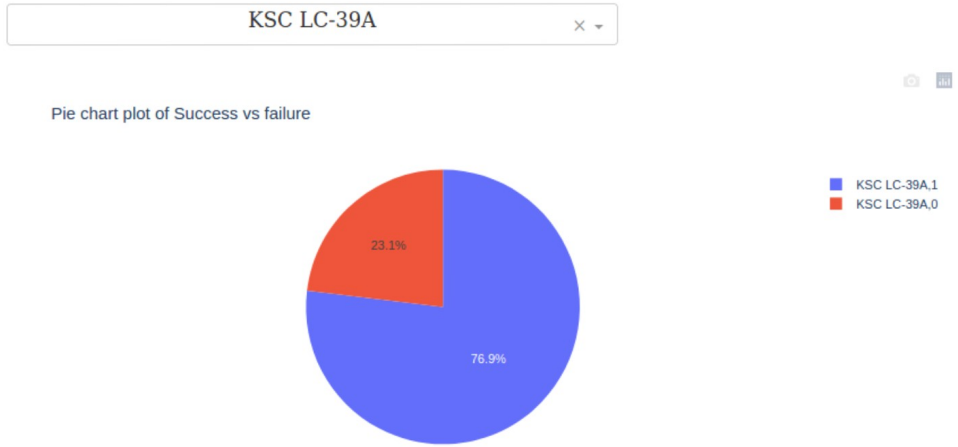


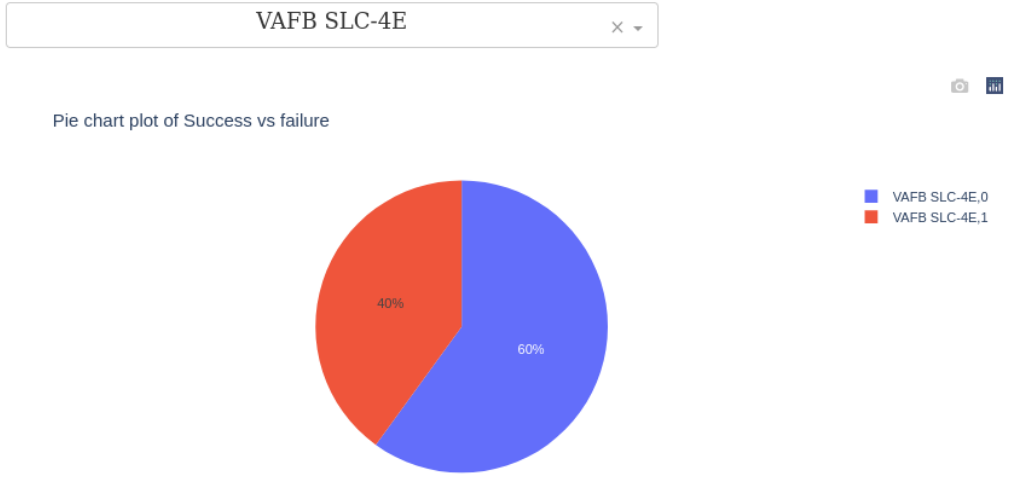
Section 5

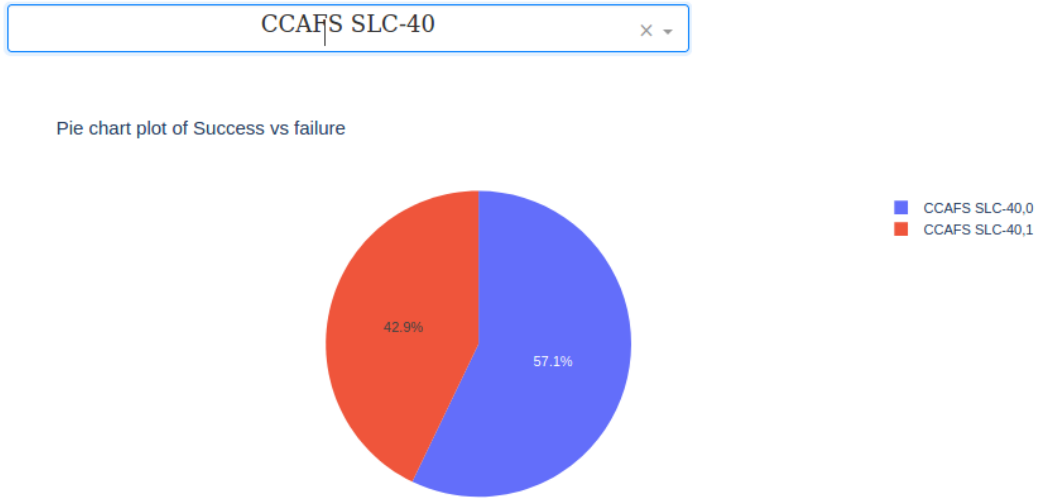
Build a Dashboard with Plotly Dash

Launch station success rate

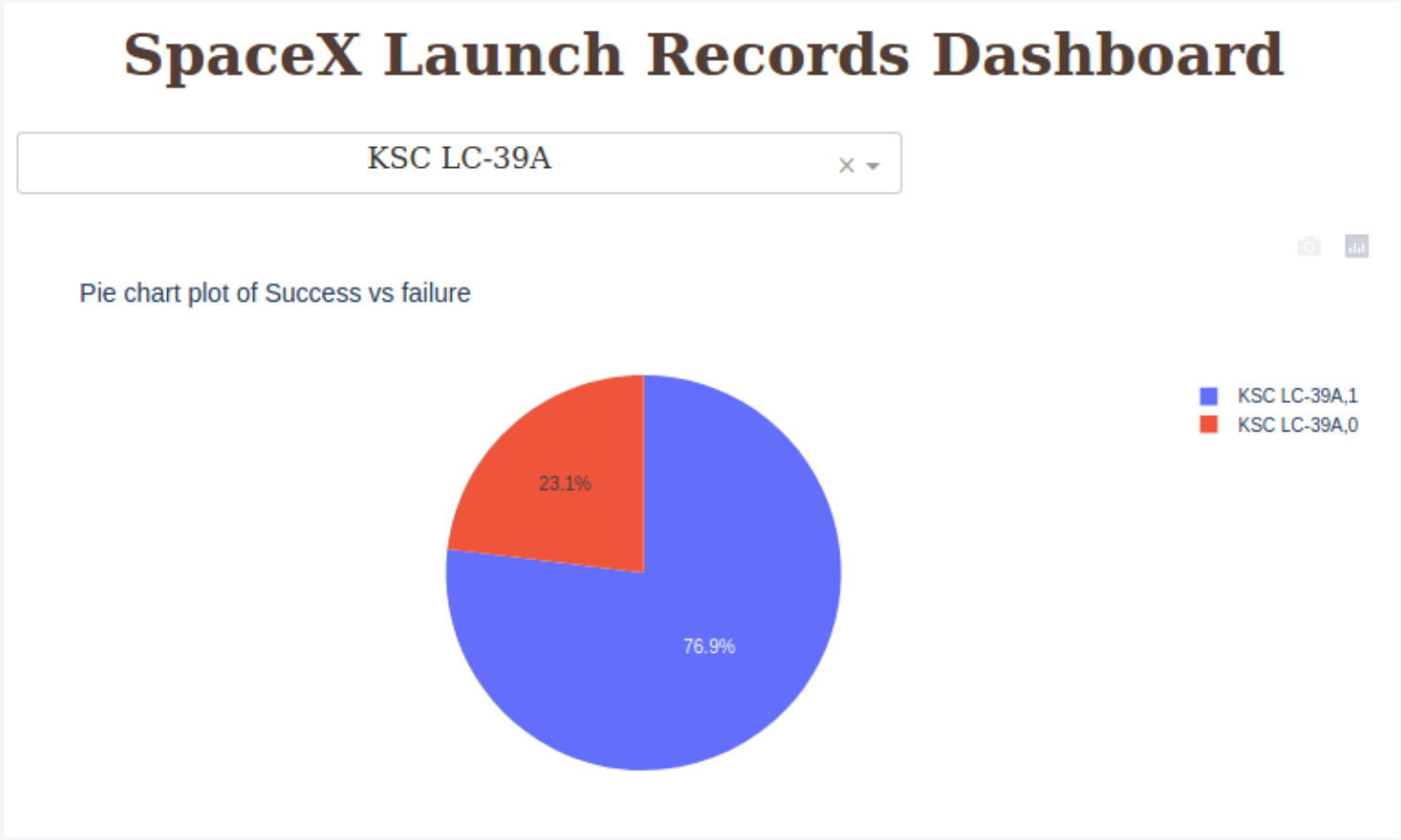




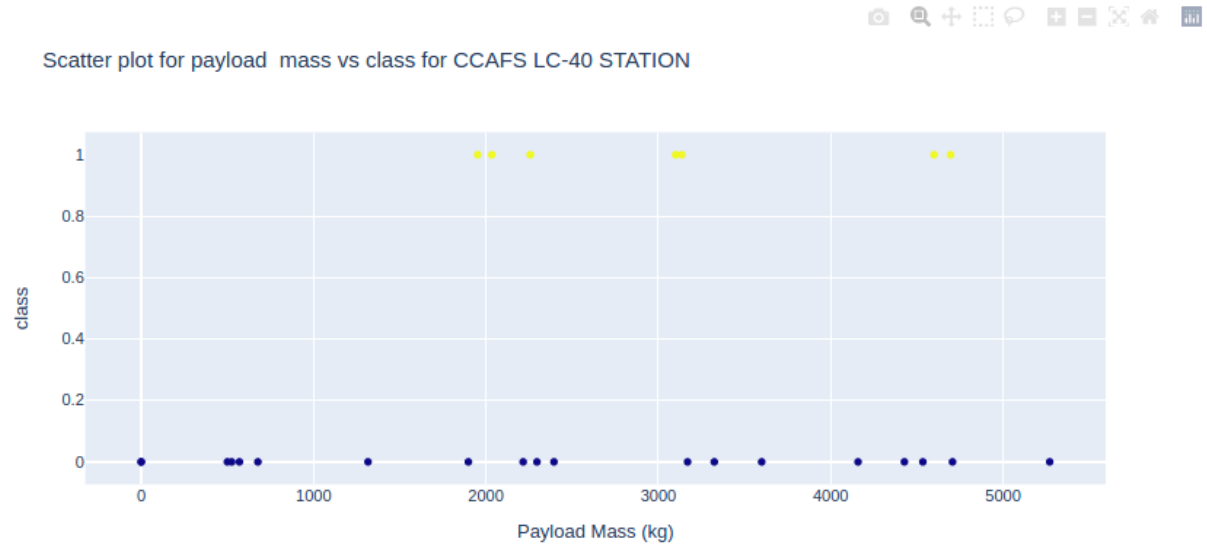




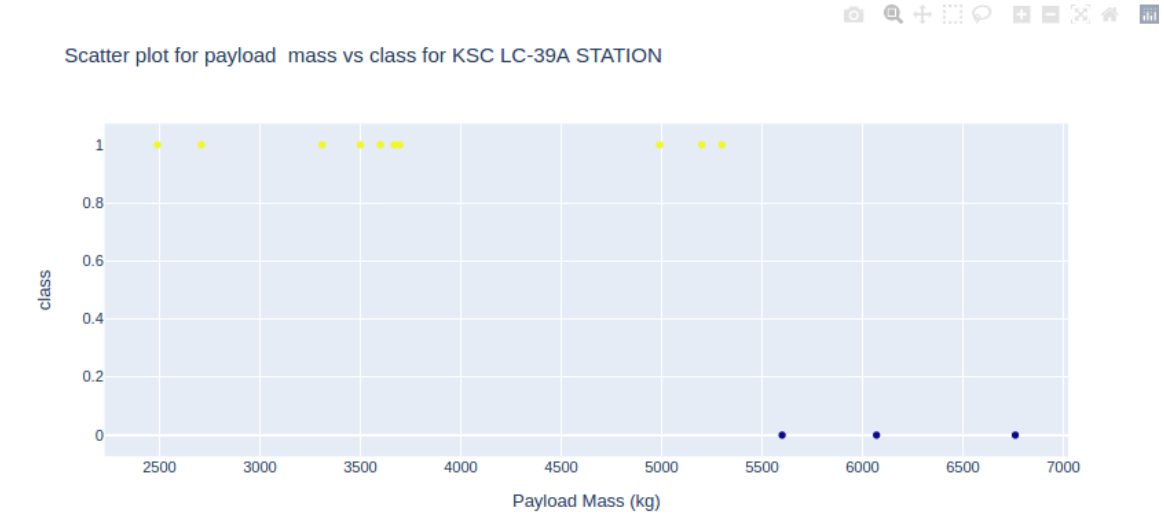
Launch station with highest success rate



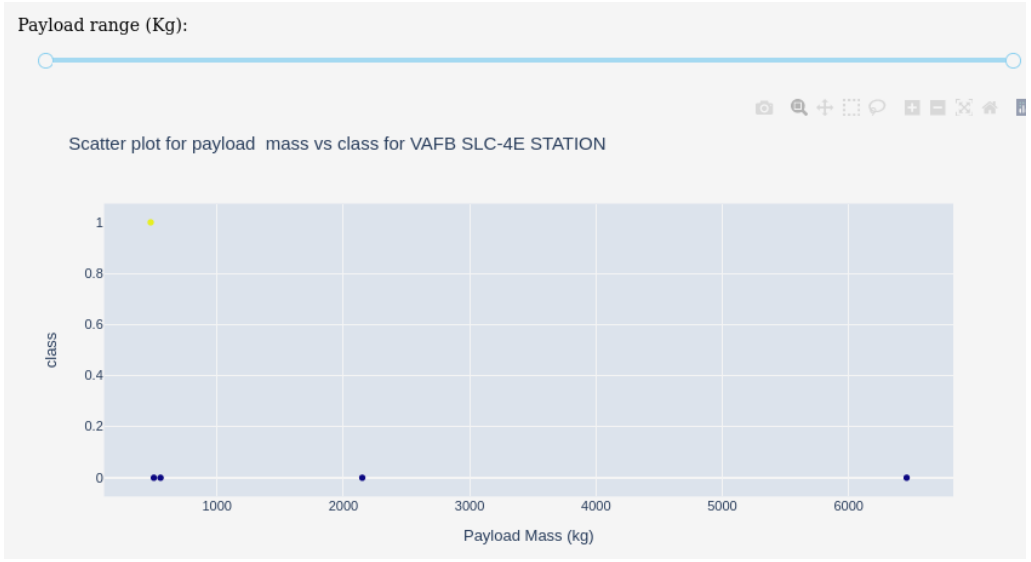
Payload range (Kg):



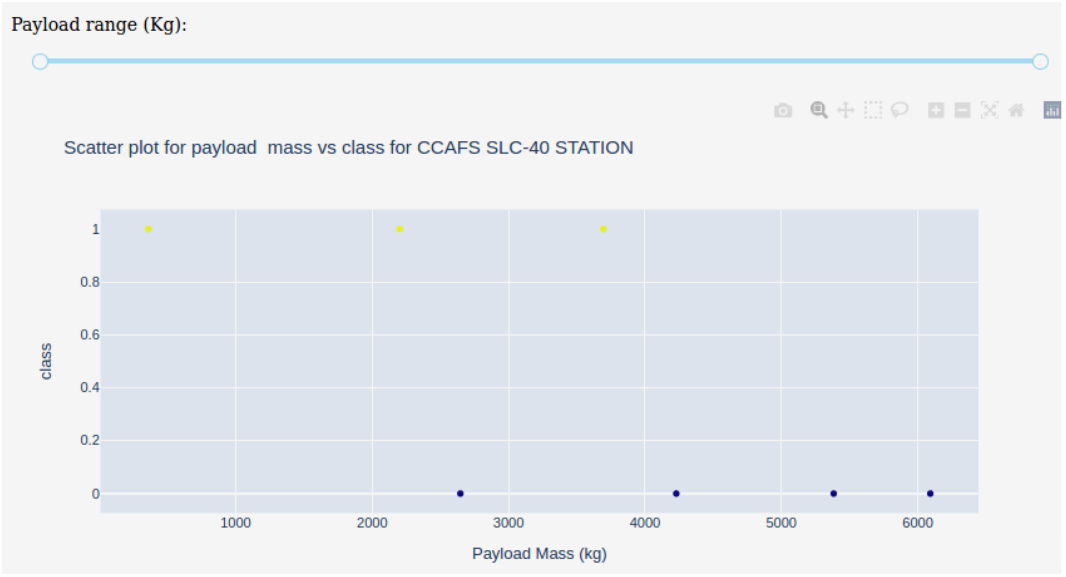
Payload range (Kg):



Payload range (Kg):



Payload range (Kg):

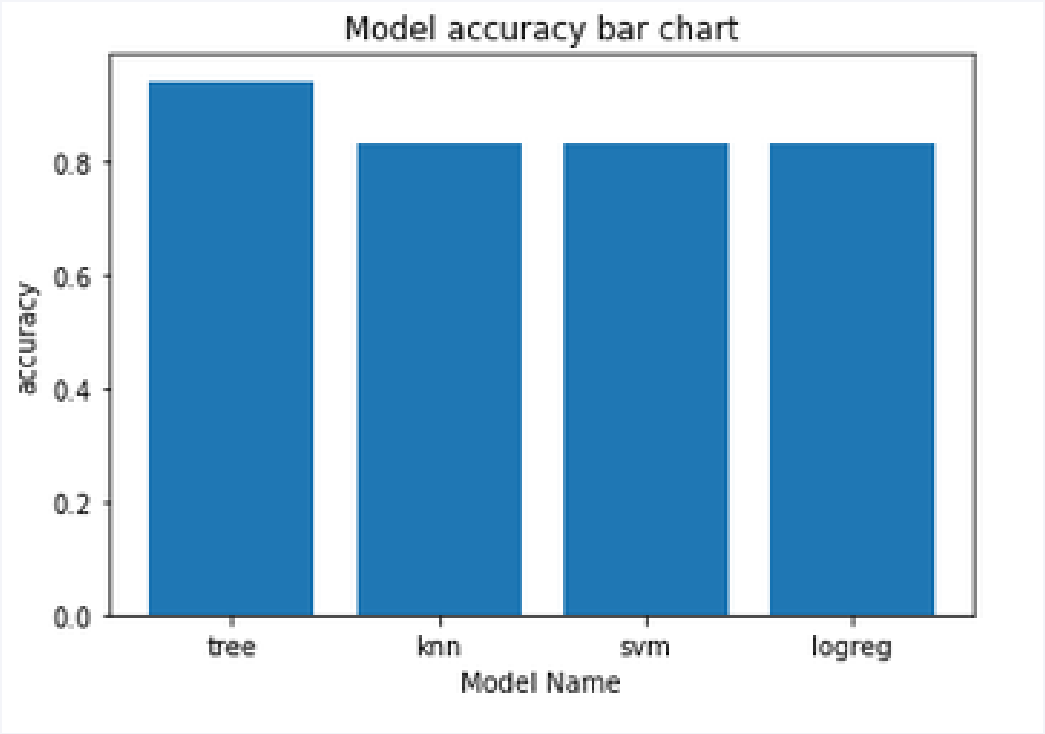


Section 6

Predictive Analysis (Classification)

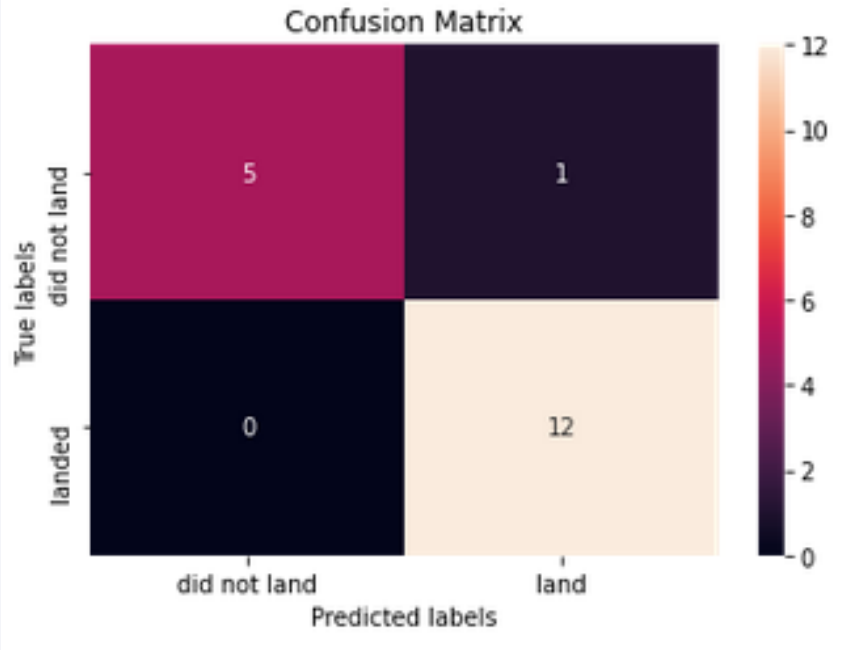
Classification Accuracy

- Tree model has the highest accuracy



Confusion Matrix

- Confusion matrix of the best performing classification tree model.
- Successfully predicts all landing.
- Successfully predicts 5 out of 6 unsuccessful landing.



Conclusions

- Data science methodology was performed for project.
- First project goal was determined. Then data were collected, processed, visualized, model was built and tested.
- With due procedure and careful programming with help of visuals good predictive model could be built.
- Our best predicted model has 94% accuracy on test data.
- Its prediction on successful landing is 100%.
- Prediction of failure missed 1 out of 6.
- For further development focus could be on launch failure prediction.
- With this model we can calculate cost of landing since we could predict first launch out outcome.

Thank you!