IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jackson Currin
March 22, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Wrangling
  - Data collection through API
  - Data collection with Web Scraping
  - Data exploration with SQL
  - Data exploration with visualization
  - Visual analytics with Folium
  - Machine learning predictions
- Summary of all results
  - The collected data and machine learning tools were successful in predicting which characteristics are pertinent in predicting future launch outcomes.

Introduction

- This project encapsulated an effort to collect, wrangle, and analyze data valuable in predicting the outcome of future SpaceX launches, specifically whether the first stage of Falcon 9 launches will successfully land.

Section 1

Methodology

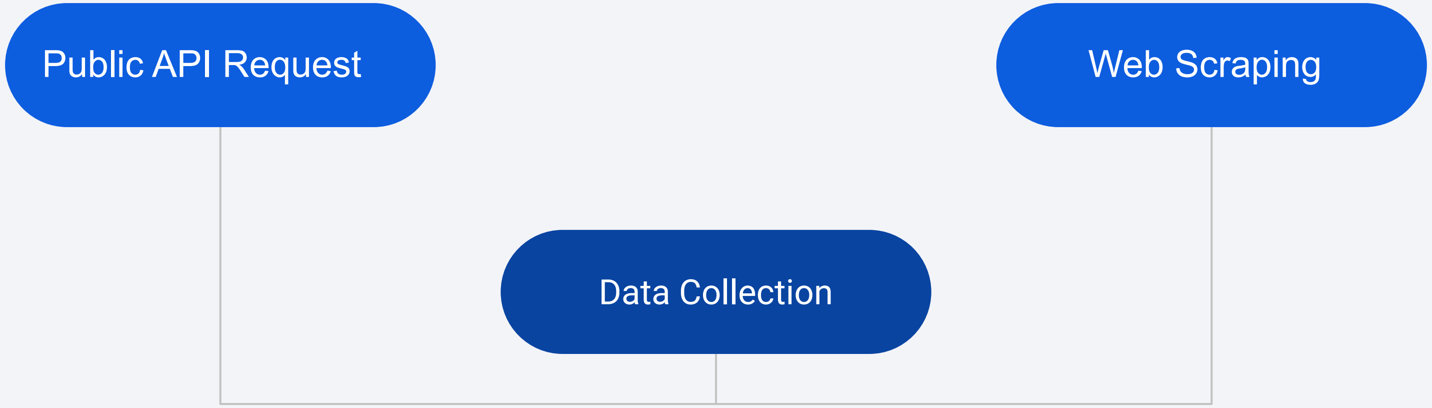
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from public API requests.
- Perform data wrangling
  - Data was processed using Python
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - A number of classifications models were tested in this project, notably KNN and RBF.

Data Collection

- Data sets were collected via web scraping with beautiful soup and public API JSON file requests. Once the raw data was collected, it was cleaned and formatted to csv files.



Data Collection – SpaceX API

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

- 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

Request Data from public API > Create dataframe using requested data > clean and filter collected data

https://github.com/Jcurrin27/ibmproject/blob/master/notebook_Data_Collection_API_Lab_mBqvsdORq.ipynb

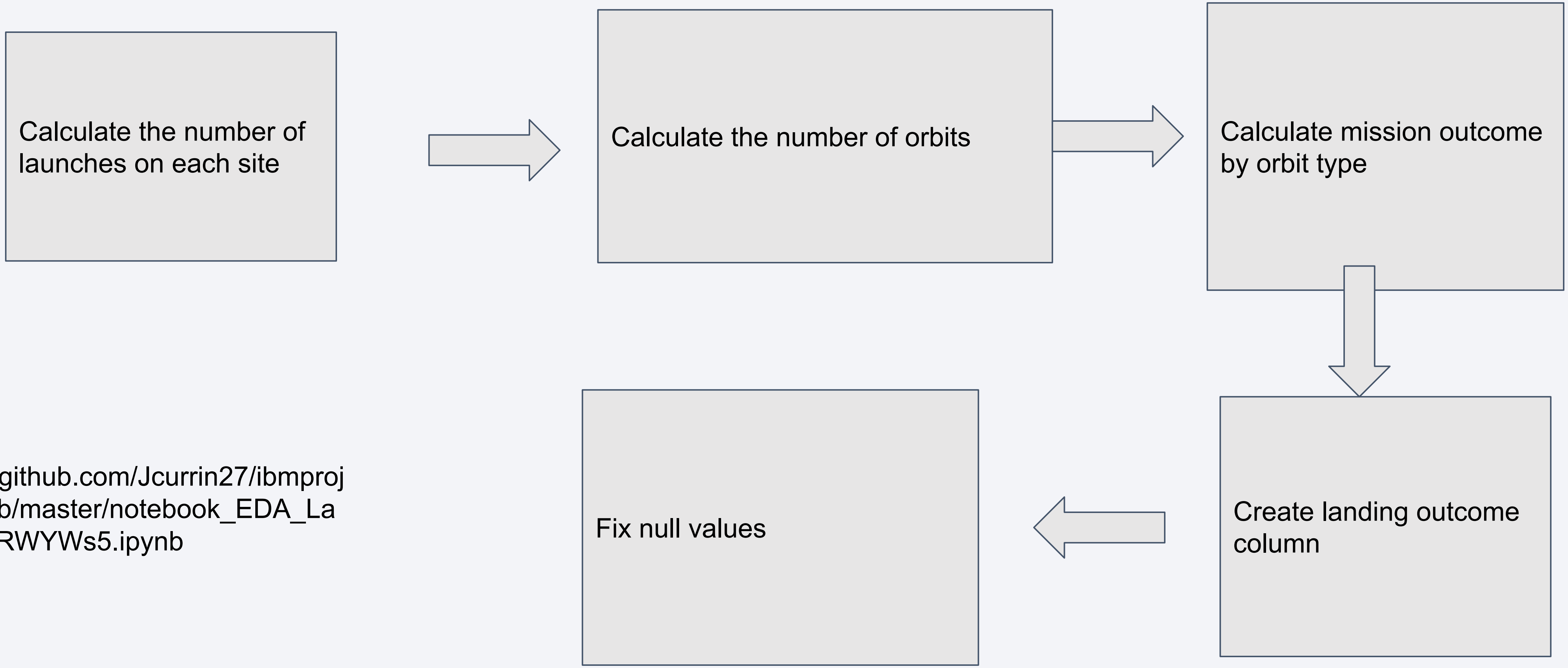
Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Create soup object > find column names > fill with data from wiki table

Data Wrangling



https://github.com/Jcurrin27/ibmproject/blob/master/notebook_EDA_Lab_75BRWYWs5.ipynb

EDA with Data Visualization

- We used geographic, folium charts in order to visualize success rates according to geographic locality
- https://github.com/Jcurrin27/ibmproject/blob/master/notebook_Interactive_Visual_Analytics_with_Folium_lab_Wa9DkPc60.ipynb

EDA with SQL

- Using like clause, query for launch sites beginning with KSC
- Using sum aggregation, query for total payload mass
- Using avg aggregation, query for average payload mass
- Using min function query for first successful launch
- Query for list names of which have success in ground pad and have payload mass greater than 4000 but less than 6000
- Query for mission success and failures
- Query for booster versions which have carried maximum payload mass
- Query for the records which will display the month names, successful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
- Query for the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- https://github.com/Jcurrin27/ibmproject/blob/master/ibm_dashnotes_3-14-23.txt

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- 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
- https://github.com/Jcurrin27/ibmproject/blob/master/Machine%20Learning%20Prediction%20lab.ipynb

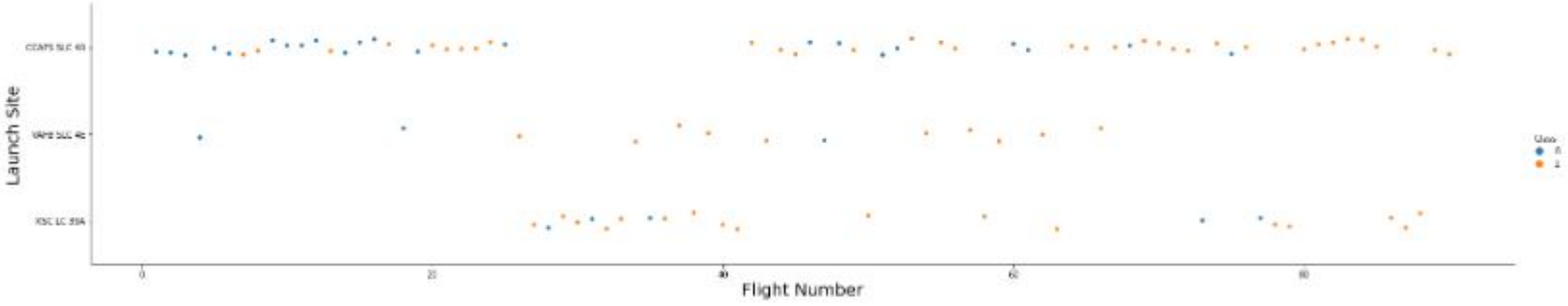
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

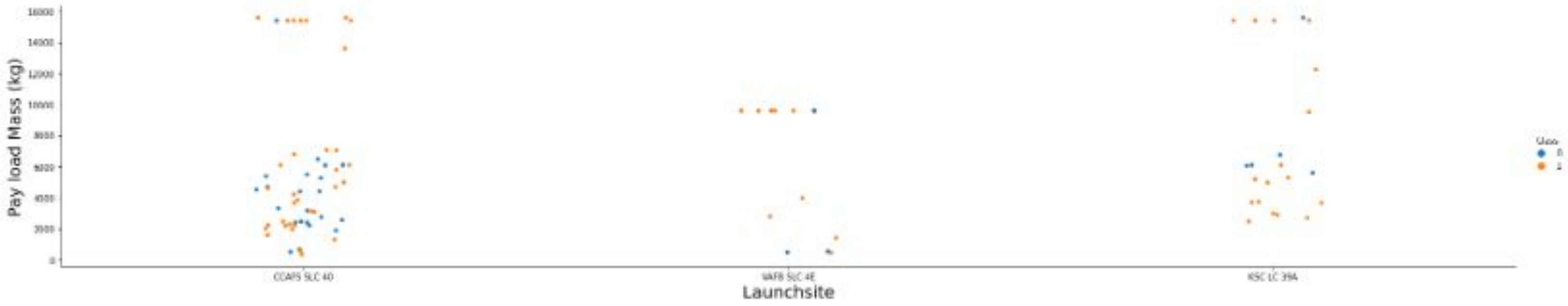
In [16]:

```
# Plot a scatter point chart with x axis to be Flight Number and y axis
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, asp
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



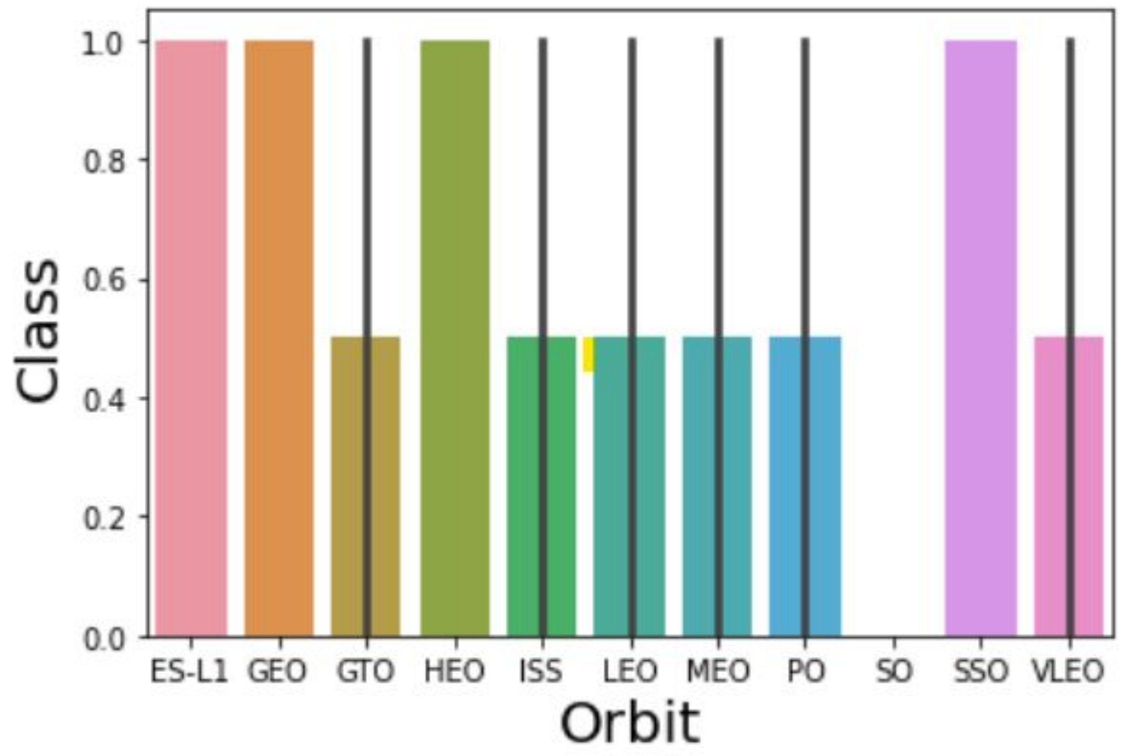
Payload vs. Launch Site

```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y
sns.catplot(y="PayloadMass", x="LaunchSite", hue="Class", data=df, aspe
plt.xlabel("Launchsite",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```

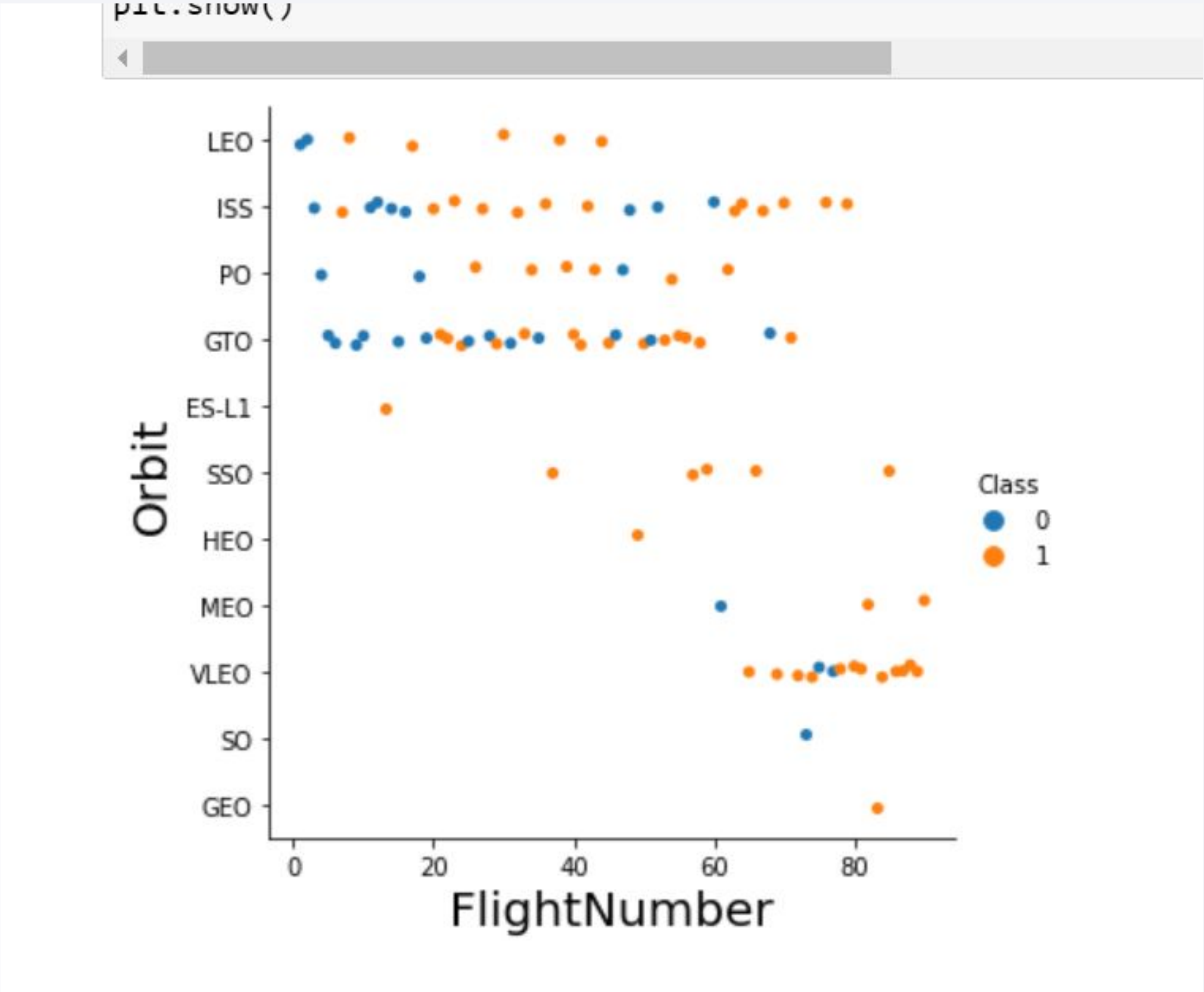


Success Rate vs. Orbit Type

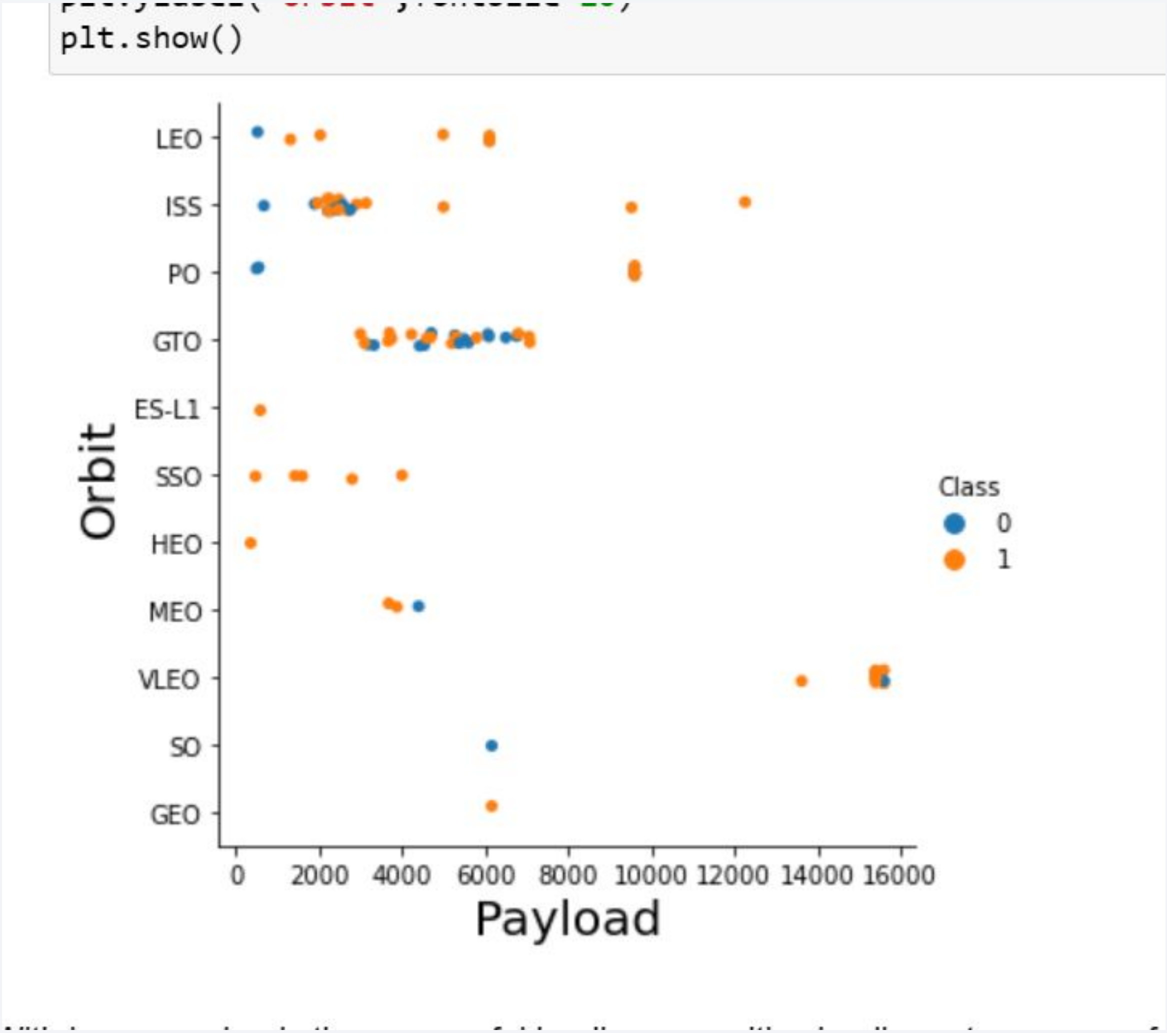
```
plt.ylabel("Class",fontsize=20)
plt.show()
```



Flight Number vs. Orbit Type



Payload vs. Orbit Type



All Launch Site Names

- Find the names of the unique launch sites:

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

- Present your query result with a short explanation here

  Select distinct(launch_site) from spacex

Launch Site Names Begin with 'KSC'

- Find 5 records where launch sites' names start with `KSC`
- Present your query result with a short explanation here

```
select *
from spacex
where launch_site like "KSC%"
limit 5
```

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

```
select sum(payload_mass_kg)
from spacex
where customer like “NASA%”
```

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

  select avg(payload_mass_kg)

  from spacex

  where booster_version = “F9 v1.1”

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on drone ship. Present your query result with a short explanation here

```
select min(date)
from spacex
where landing_outcome = “Success (drone ship)”
```

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

select distinct(booster_version)

from spacex

where landing_outcome= “Success (drone ship)”

and payload_mass_kg > 4000

and payload_mass_kg < 6000

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

select distinct(mission_outcome), count(mission_outcome)

from spacex

group by distinct(mission_outcome)

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

select booster_version

from spacex

where payload_mass_kg = (select max(payload_mass_kg) from spacex)

2015 Launch Records

- List the records which will display the month names, successful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

- Present your query result with a short explanation here

select date, landing_outcome, booster_version, launch_site

from spacex

where landing_outcome = “Success (ground pad)”

and month like “%2017”

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

- Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

- Present your query result with a short explanation here

select count(landing_outcome)

from spacex

where date between “2010-06-04” and “2017-03-20”

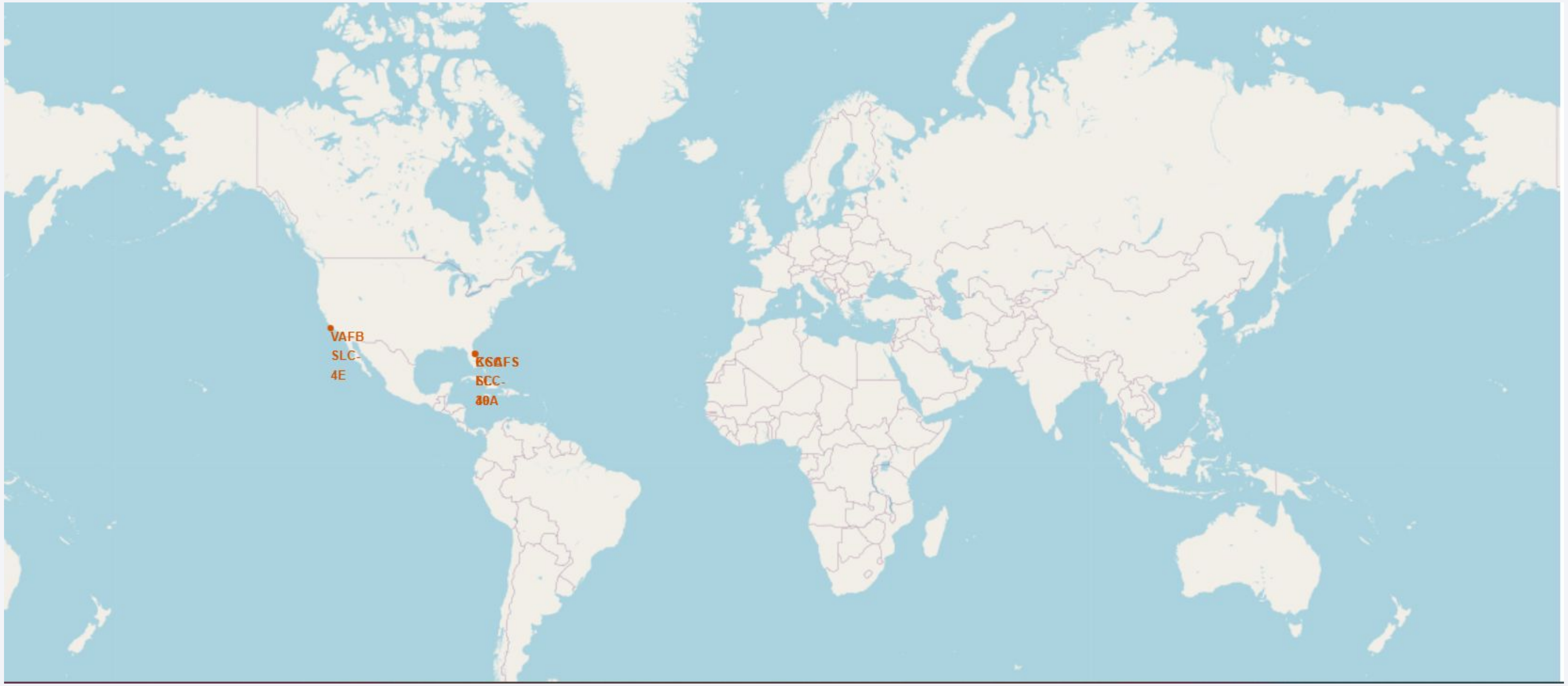
order by date desc

Section 3

Launch Sites Proximities Analysis

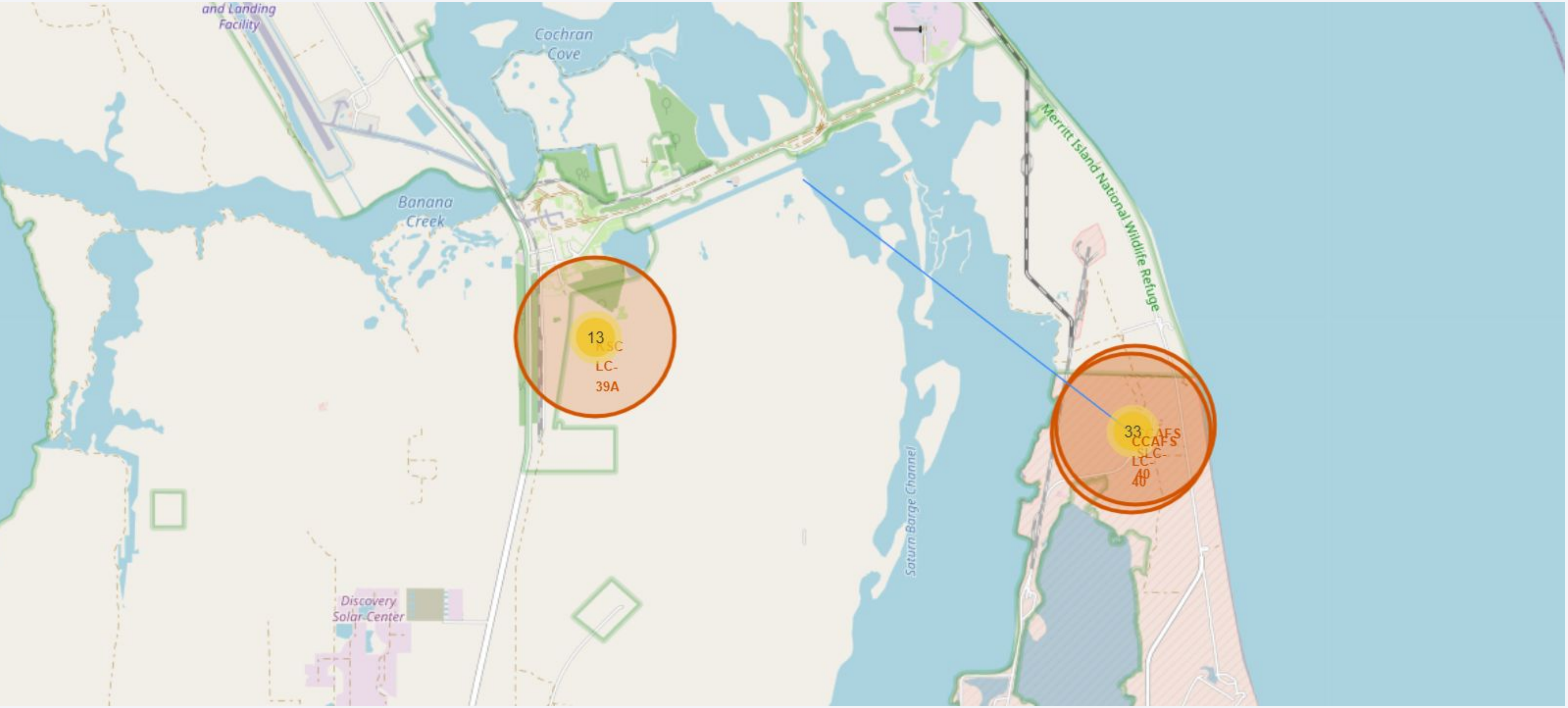
Folium Map

- Replace <Folium map screenshot 1> title with an appropriate title



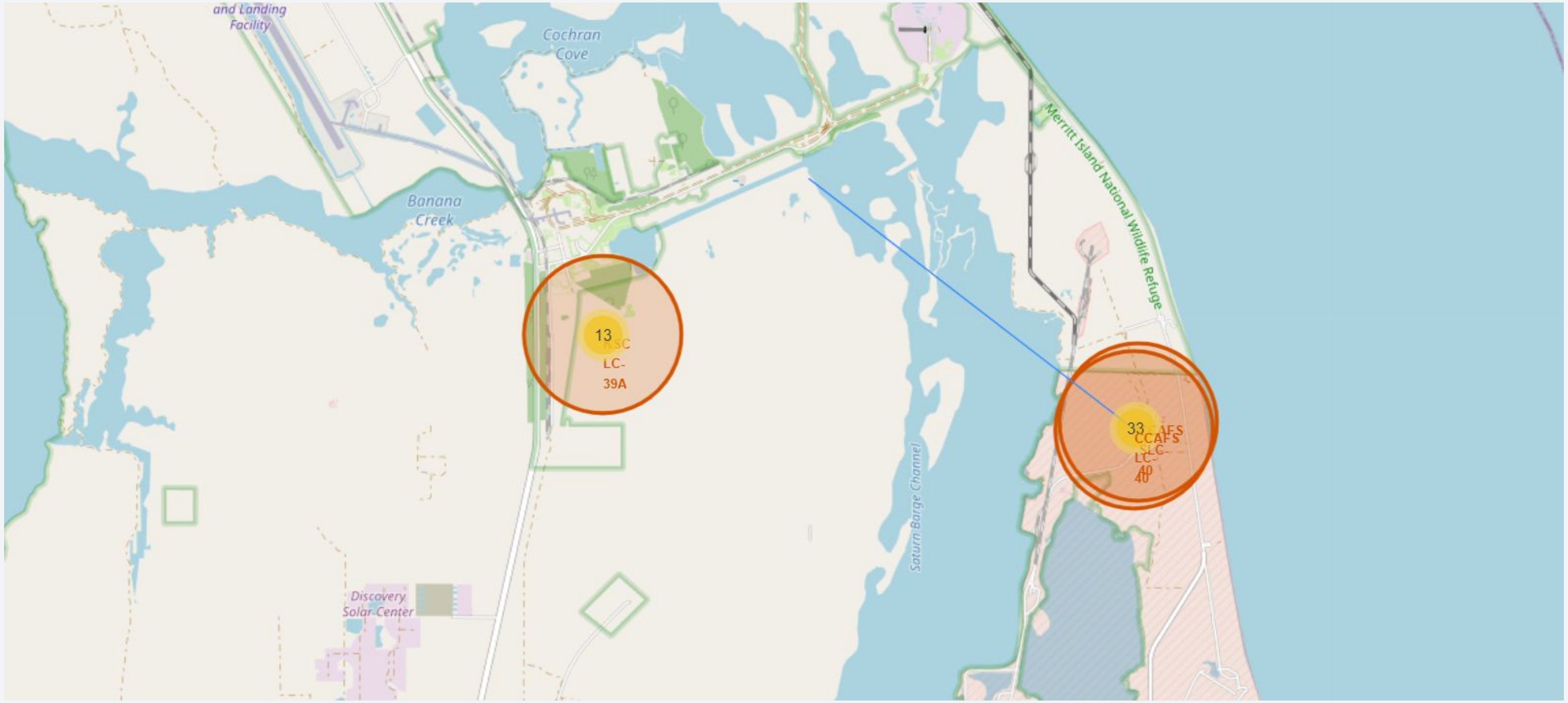
- e

Folium Map 2



and Landing Facility
Cochran Cove
Banana Creek
Merritt Island National Wildlife Refuge
13
KSC LC-39A
33
CCAFS SLC-40
CCAFS LC-40
Saturn Barge Channel
Discovery Solar Center

Folium Map3


and Landing Facility
Cochran Cove
Banana Creek
Merritt Island National Wildlife Refuge
13
KSC LC-39A
33
CCAFS SLC-40
Saturn Barge Channel
Discovery Solar Center

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- the more launches at a particular site, the greater the chance of future success
- Orbits ES-LI, GEO, HEO, SSO had the greatest rates of success
- CCAFS SLC-40 had the largest number of successful stage 1 landings
- launch success rates are increasing over time. this indicates the potential for greater success in the future.

Thank you!