IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmed Al Abed
28th September 2021

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

- Collecting Data through an API and through web scrapping with beautiful soup.
- Wrangling data with Pandas.
- EDA with SQL, Pandas and Seaborn.
- Data Visualization with Folium, Dash and Plotly.
- Used Machine learning algorithms for prediction.

Introduction

Project background and context

- SpaceX is an American aerospace manufacturer, known for its Falcon 9 rockets.
- Falcon 9 rockets saves a lot of money because they can land after their mission.
- We wanted to know what features influence the landing success rate.

Problems you want to find answers

- What features influence the success rate the most.
- How can we make sure that the landing will be successful.

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- We Acquired the data using a REST API and by web scrapping a Wikipedia page.

Data wrangling

- Dealt with missing values and created a landing outcome label.

Exploratory data analysis (EDA) using visualization and SQL.

- We studied the different relationships between features and how they impact the final outcome

Methodology

Executive Summary

Interactive visual analytics using Folium and Plotly Dash

Created a map with the locations of launch sites.

Predictive analysis using classification models

How to build, tune, evaluate classification models

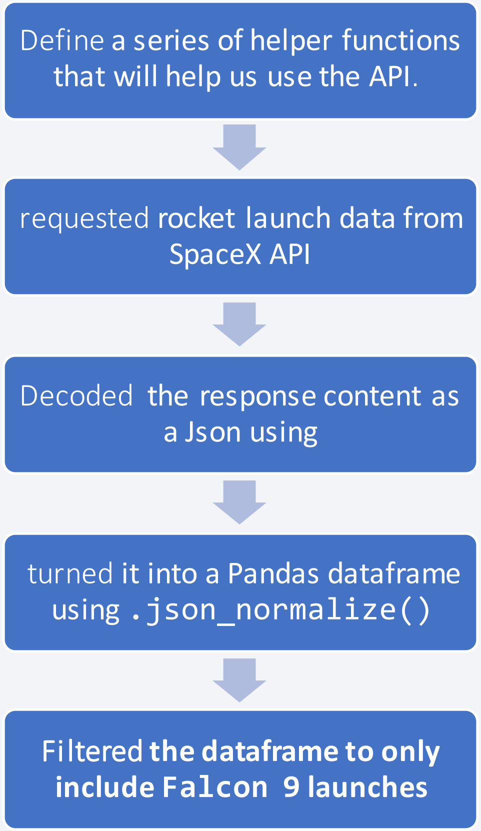
Data Collection

- How data sets were collected.
  - We requested the SpaceX API.
  - Extracted Falcon 9 launch records HTML table from Wikipedia.
  - We cleaned the Requested Data to match our objective.
  - Parsed the table and converted it into a Pandas data frame.

Data Collection – SpaceX API

- We Used
  - `spacex_url="https://api.spacexdata.com/v4/launches/past"`
  - `response = requests.get(spacex_url)`
  - `data = pd.json_normalize(response.json())`
- GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/Data%20Collection%20API.ipynb

Click to add text



Data Collection - Scraping

- We used :
  - `r = requests.get(static_url)`
  - `soup = BeautifulSoup(r.text, "html.parser")`
  - `html_tables = soup.find_all('table')`
- GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Requested the Falcon9 Launch Wiki page from its URL

↓

Created a `BeautifulSoup` object

↓

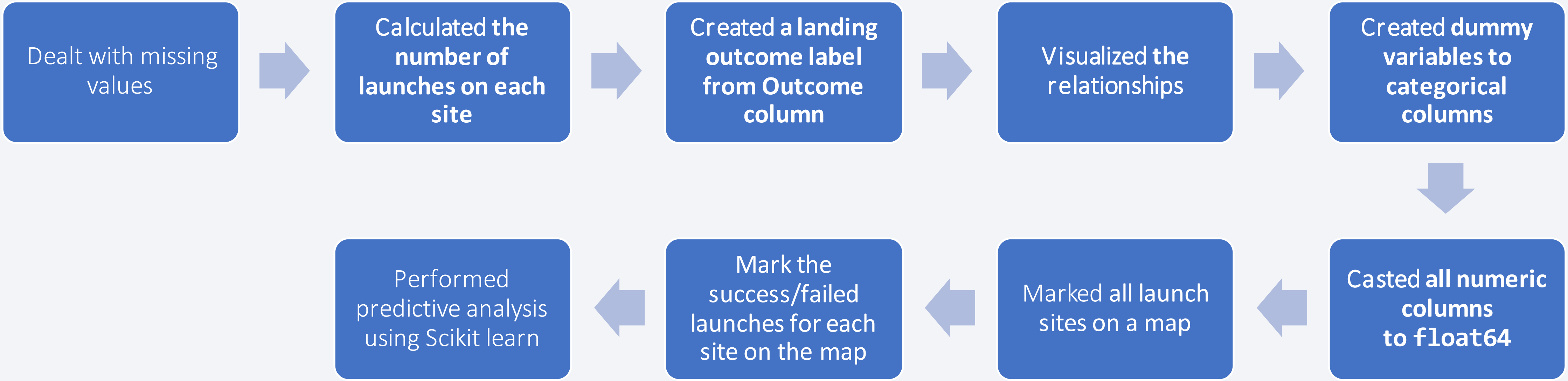
Extracted all column/variable names

↓

Created a data frame

Data Wrangling

GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/EDA.ipynb



EDA with Data Visualization

- We Plotted these Charts to Incept The different relationships and gather insight about how they affect the outcome.
- Plotted Charts :
  - **Flight Number and Launch Site**
  - **Payload and Launch Site**
  - **the relationship between success rate of each orbit type**
  - **between FlightNumber and Orbit type**
  - **Payload and Orbit type**
  - GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/EDA%20with%20SQL.ipynb

- SQL queries :
    - *the names of the unique launch sites in the space mission*
    - *ecords 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*
    - *the date when the first successful landing outcome in ground pad was acheived.*
    - *the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
    - *the total number of successful and failure mission outcomes*
    - *the names of the booster_versions which have carried the maximum payload mass.*
    - *the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015*

Interactive Map with Folium

- **Marked all launch sites on a map**
- **Marked the success/failed launches for each site on the map**
- **The distances between a launch site to its proximities**
- **We got a good insight on the locations of the launch sites and how close they are to other landmarks**

- GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

A Dashboard with Plotly Dash

- Pie chart that shows the total number of launches from each site
- A pie chart for each launch site success rate
- A scatter graph for the relationship between the outcome and pay load mass.
- GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/Dash%20Plotly.ipynb

Predictive Analysis (Classification)

- Transformed the Data to fed it into the machine learning algorithm
- Split the data into training set and testing set
- Tried out deffirent machine learning algorithms like SVM and Disicion trees
- Tuned the Hyperparameters for each model
- Evaluated each model and the selected the best one
-]GitHub URL = https://github.com/a7madalabed/Ahmed-Al-Abed-DS-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

Results

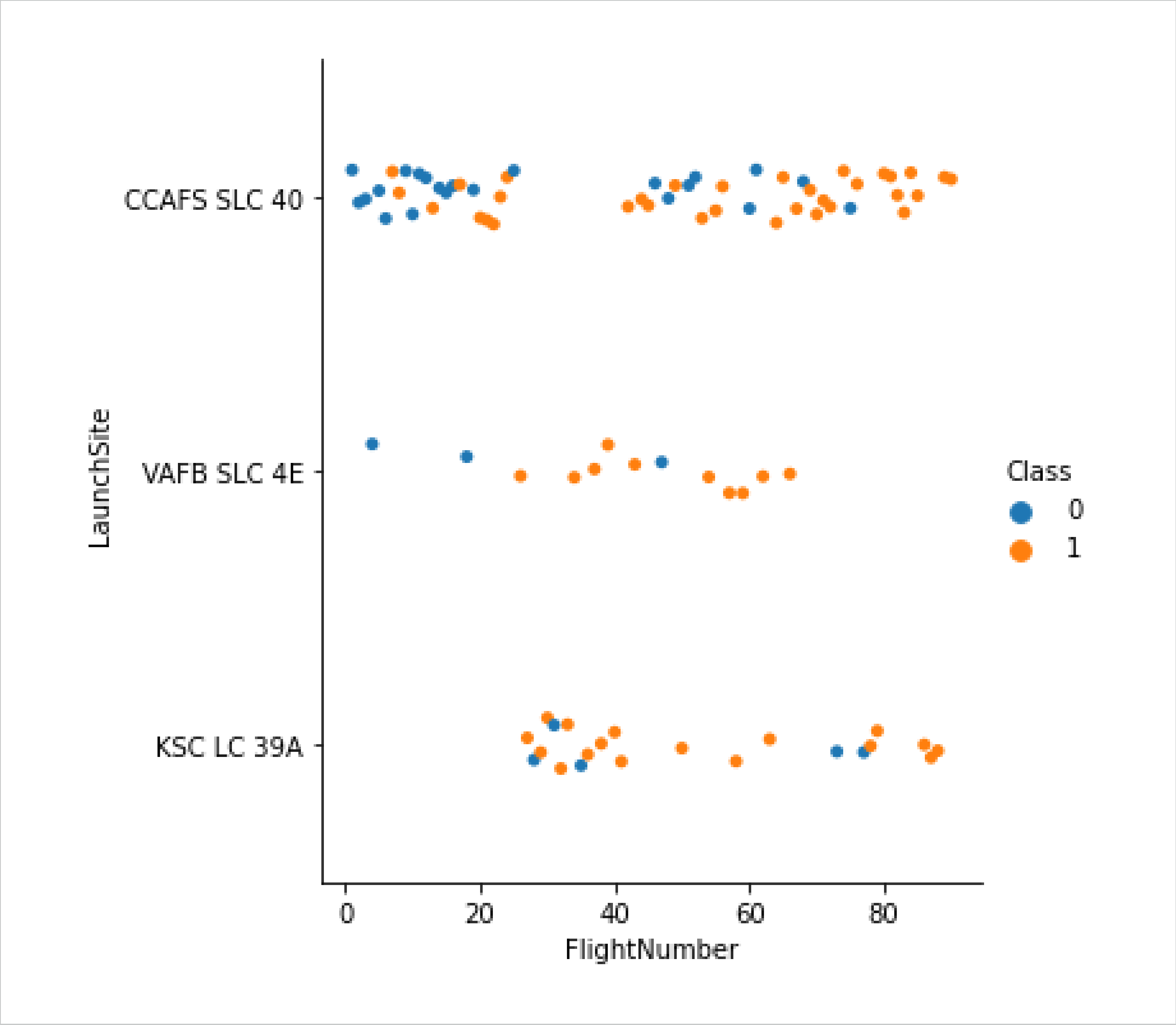
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

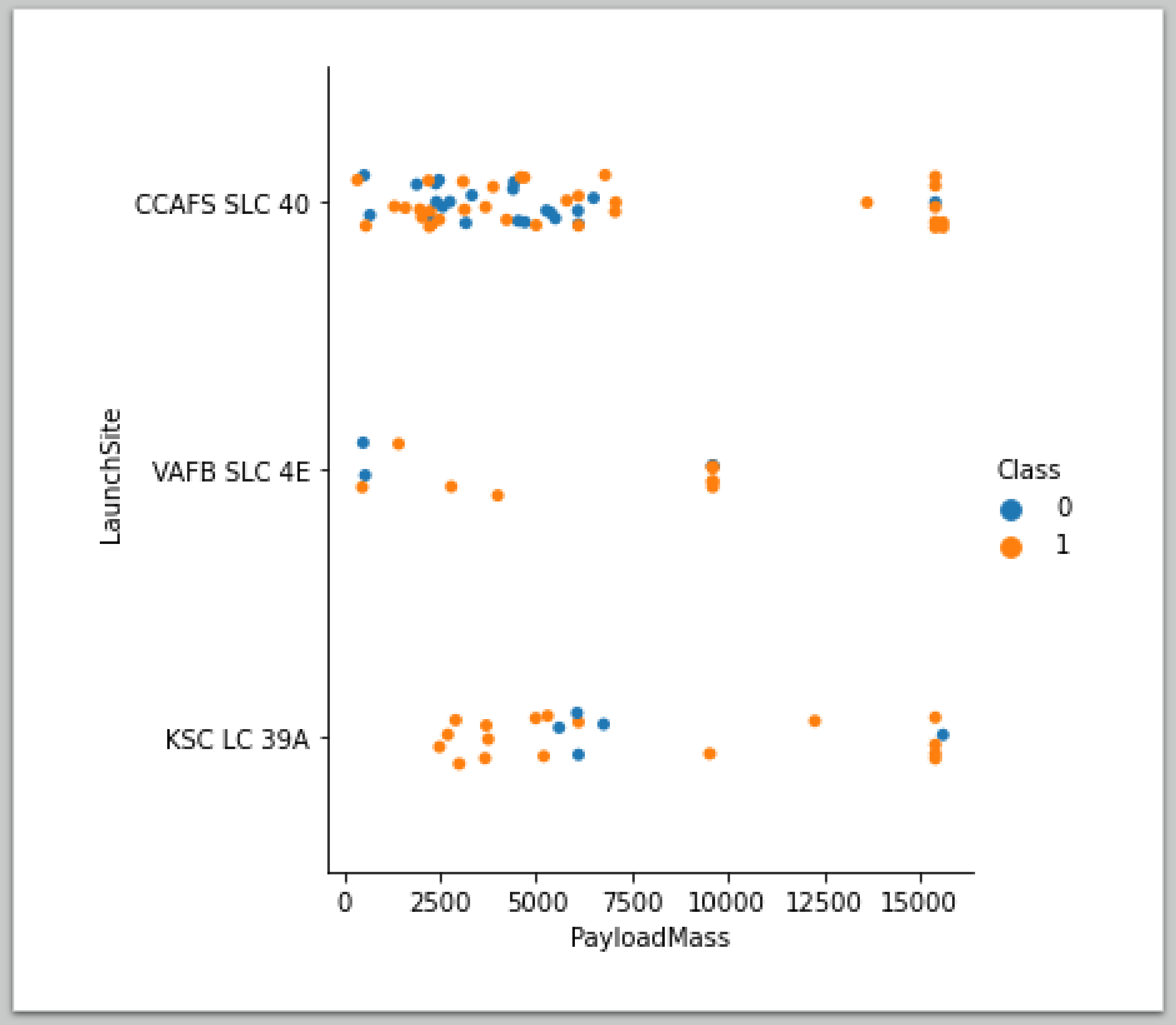
Flight Number vs. Launch Site

success rate increases as number of flights increases



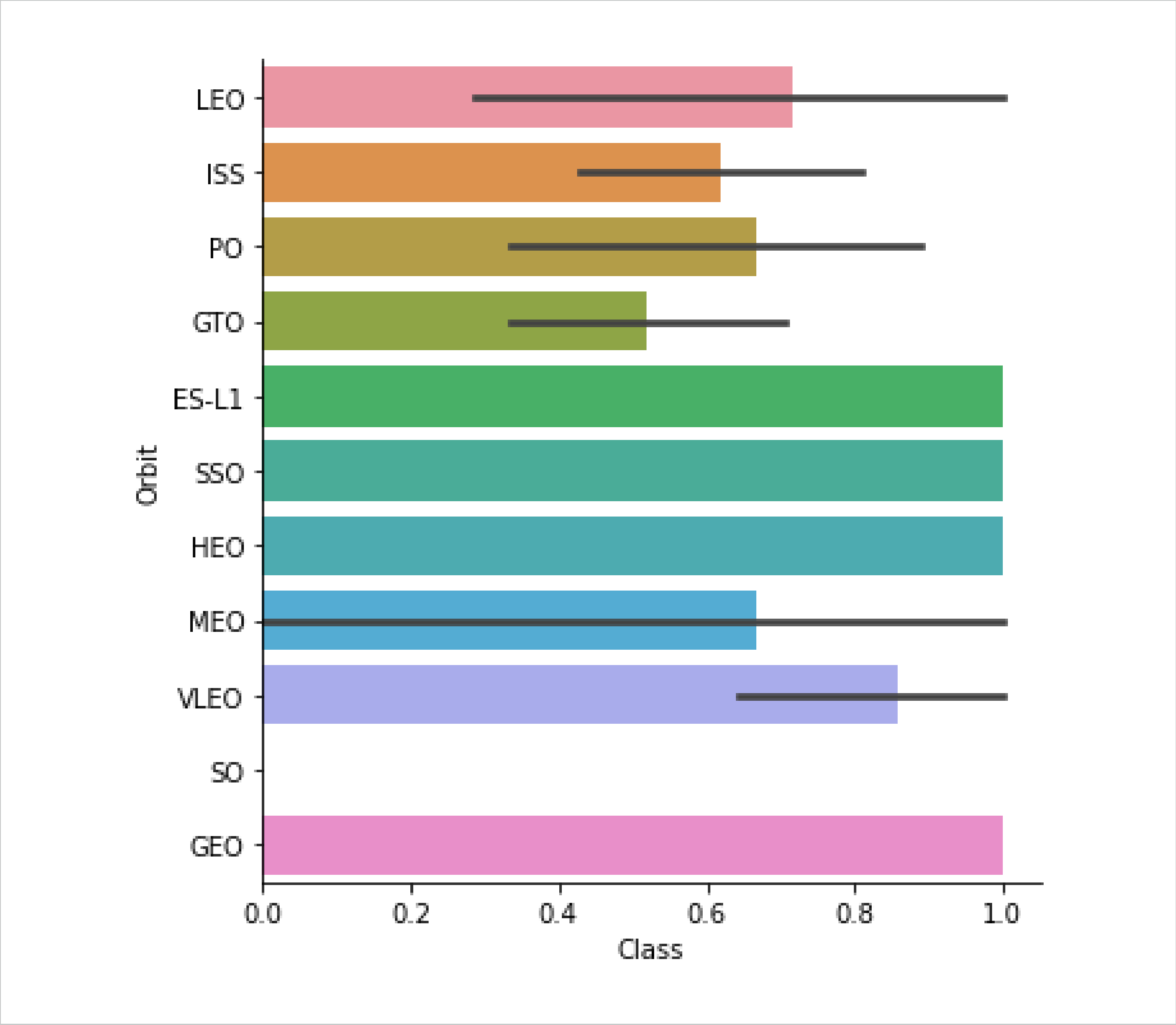
Payload vs. Launch Site

the higher the payload mass is the better the success chances



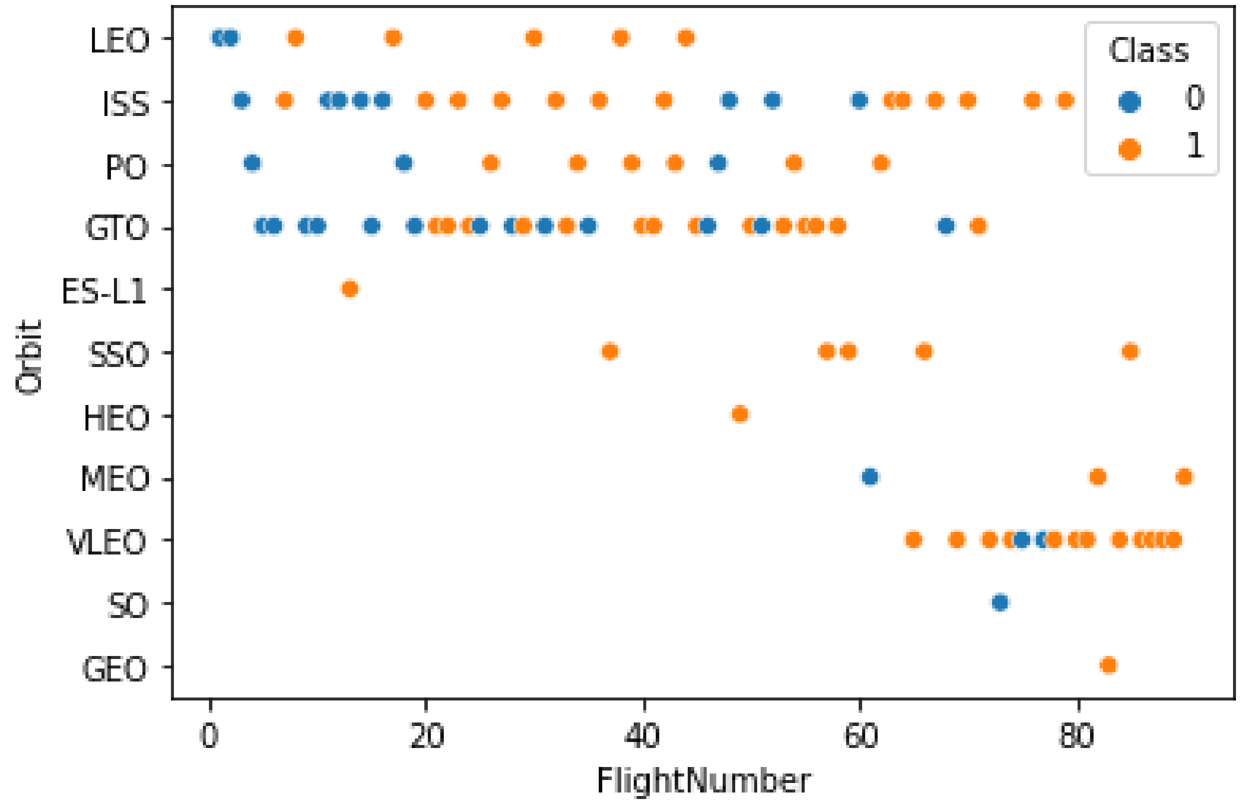
Success Rate vs. Orbit Type

Orbits GEO, HEO, SSO and ES-L1 has the best Success Rate



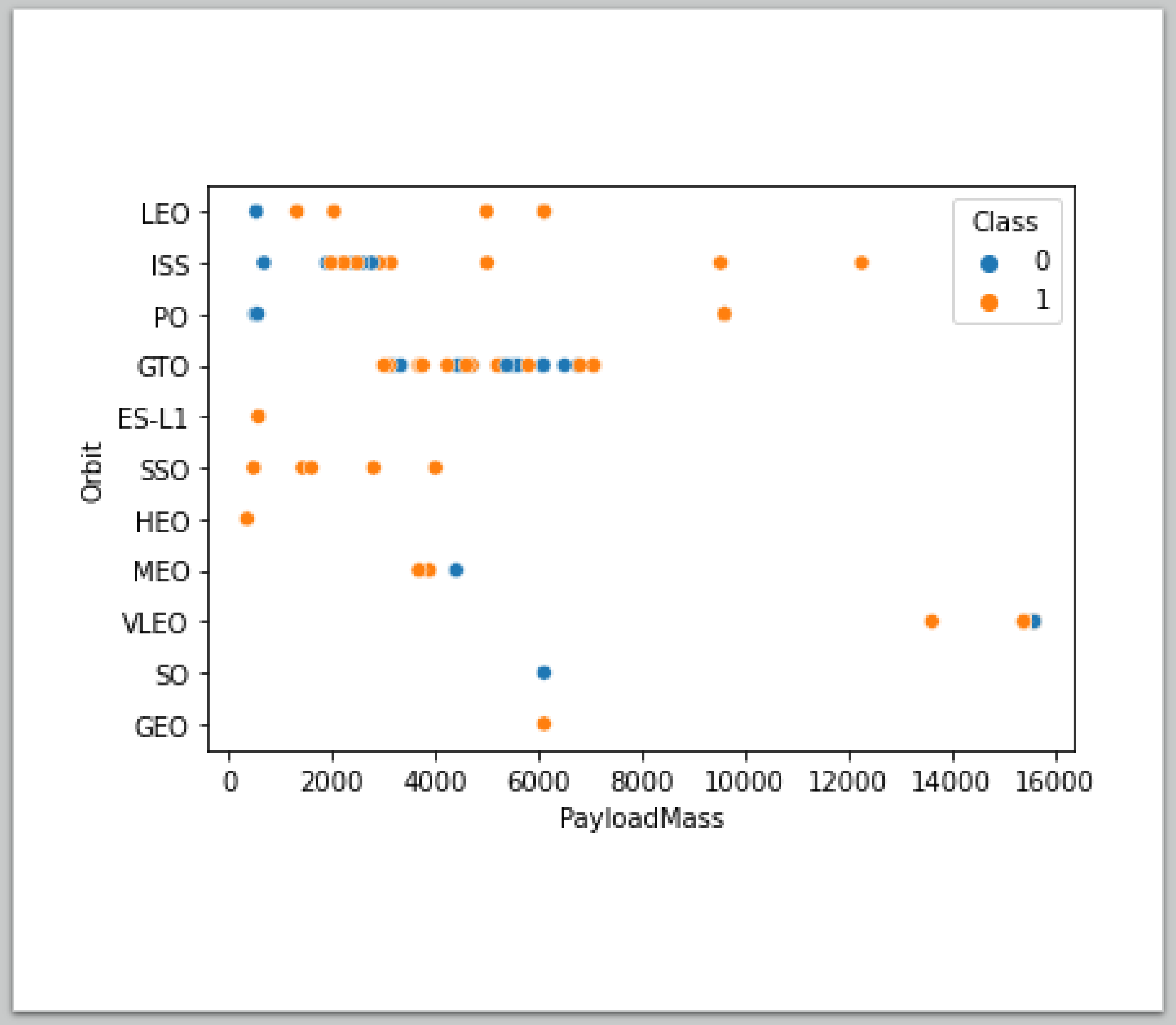
Flight Number vs. Orbit Type

The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



All Launch Site Names

- `select DISTINCT Launch_Site from SPACEXDATASET`

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

Launch Site Names Begin with 'CCA'

- `select * from SPACEXDATASET WHERE Launch_Site LIKE 'KSC%' limit 5`

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-01-05	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-03-06	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-05-07	23:38:00	F9 FT B1037	KSC LC-39A	Intelsat 35e	6761	GTO	Intelsat	Success	No attempt
2017-07-09	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (ground pad)
2017-11-10	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Payload Mass

- SELECT SUM(PAYLOAD_MASS__KG_) as total_payload from SPACEXDATASET where Customer = 'NASA (CRS)'

total_payload
22007

Average Payload Mass by F9 v1.1

- select AVG(PAYLOAD_MASS__KG_) as AVG_PAYLOAD from SPACEXDATASET where Booster_Version = 'F9 v1.1'

avg_payload
3676

First Successful Ground Landing Date

- select MIN(Date) from SPACEXDATASET where Landing__Outcome like 'Success (drone ship)'

1
2016-06-05

Successful Drone Ship Landing with Payload between 4000 and 6000

- select Booster_Version from SPACEXDATASET where Landing__Outcome = 'Success (ground pad)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000

booster_version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

- `SELECT Count(Mission_Outcome) as success from SPACEXDATASET where Mission_Outcome like '%Success%'`
- `SELECT Count(Mission_Outcome) as failure from SPACEXDATASET where Mission_Outcome LIKE '%Failure%'`

failure
0

success
45

Boosters Carried Maximum Payload

- SELECT DISTINCT Booster_Version, MAX(PAYLOAD_MASS__KG_) AS Max FROM SPACEXDATASET GROUP BY Booster_Version ORDER BY Max DESC LIMIT 5

booster_version	MAX
F9 B5 B1048.4	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1049.5	15600
F9 B5 B1049.4	15600

2015 Launch Records

- `SELECT Date, Booster_Version, Launch_Site FROM SPACEXDATASET WHERE Landing__Outcome LIKE '%Failure%'`

2015-10-01	F9 v1.1 B1012	CCAFS LC-40

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

- SELECT COUNT(Landing__Outcome) FROM SPACEXDATASET WHERE Landing__Outcome LIKE '%Success%'

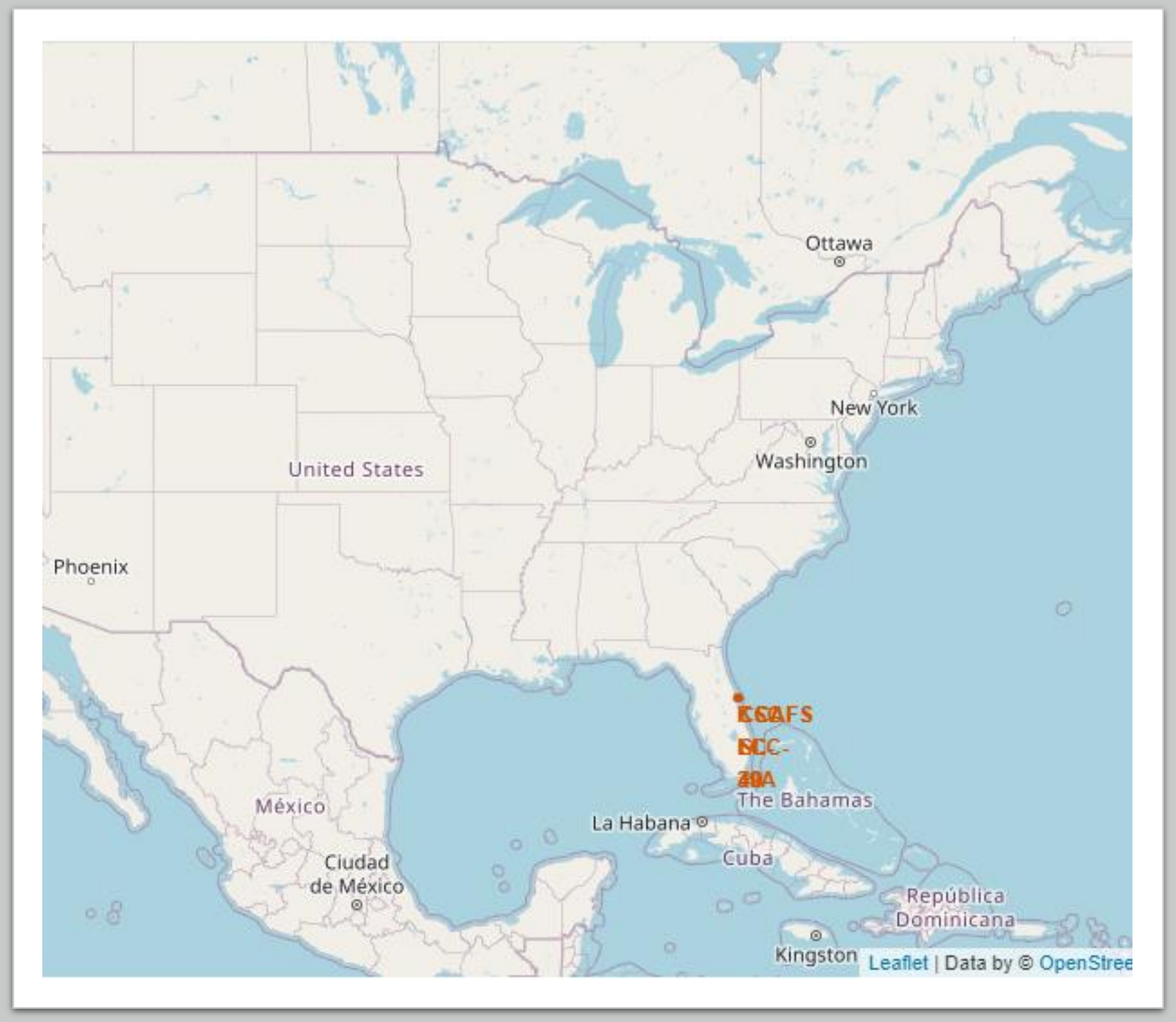
1
27

Section 4

Launch Sites Proximities Analysis

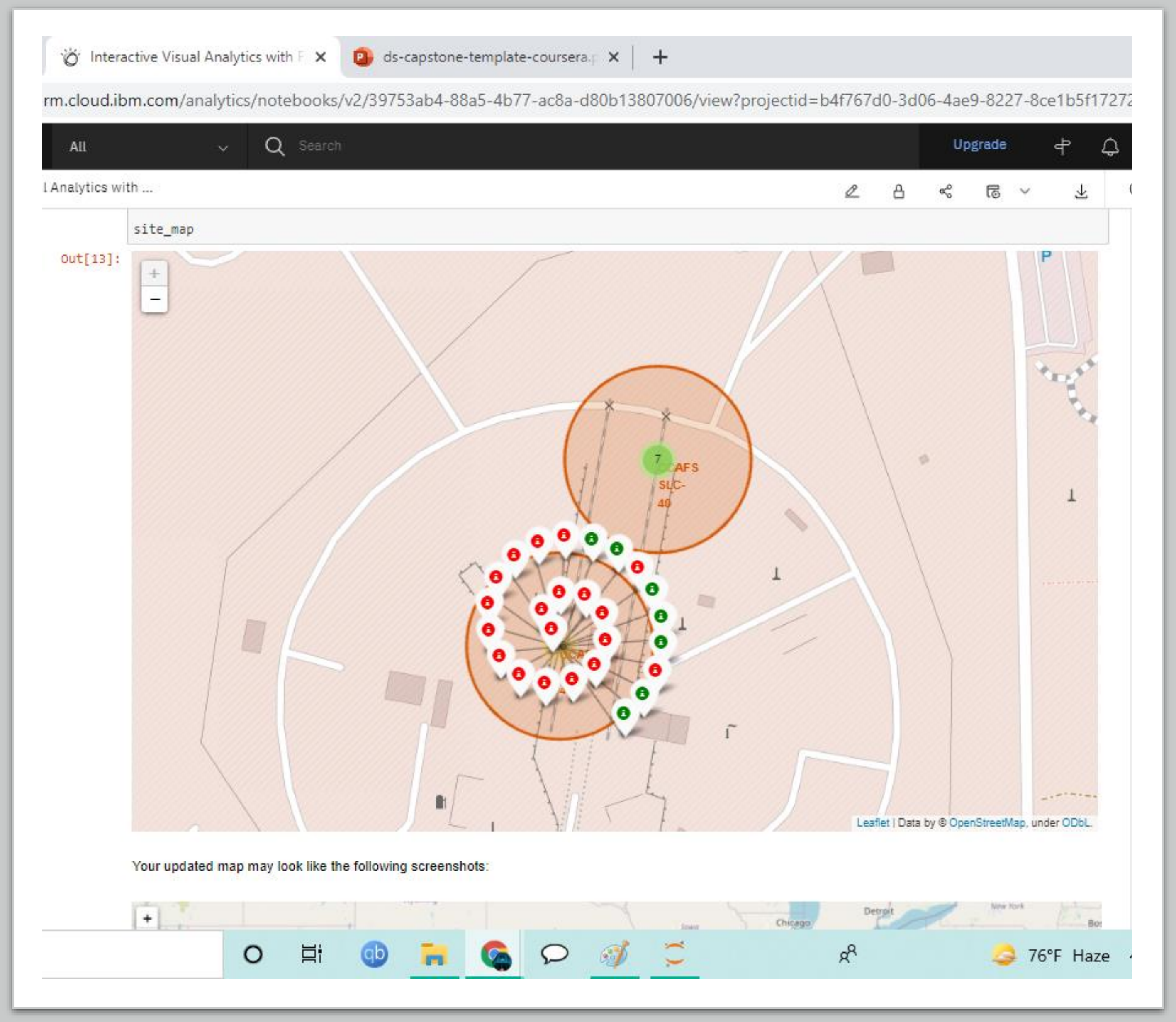
Launch Sites

- The launch sites are located in the United States more spacificaly in California and Florida coasts.



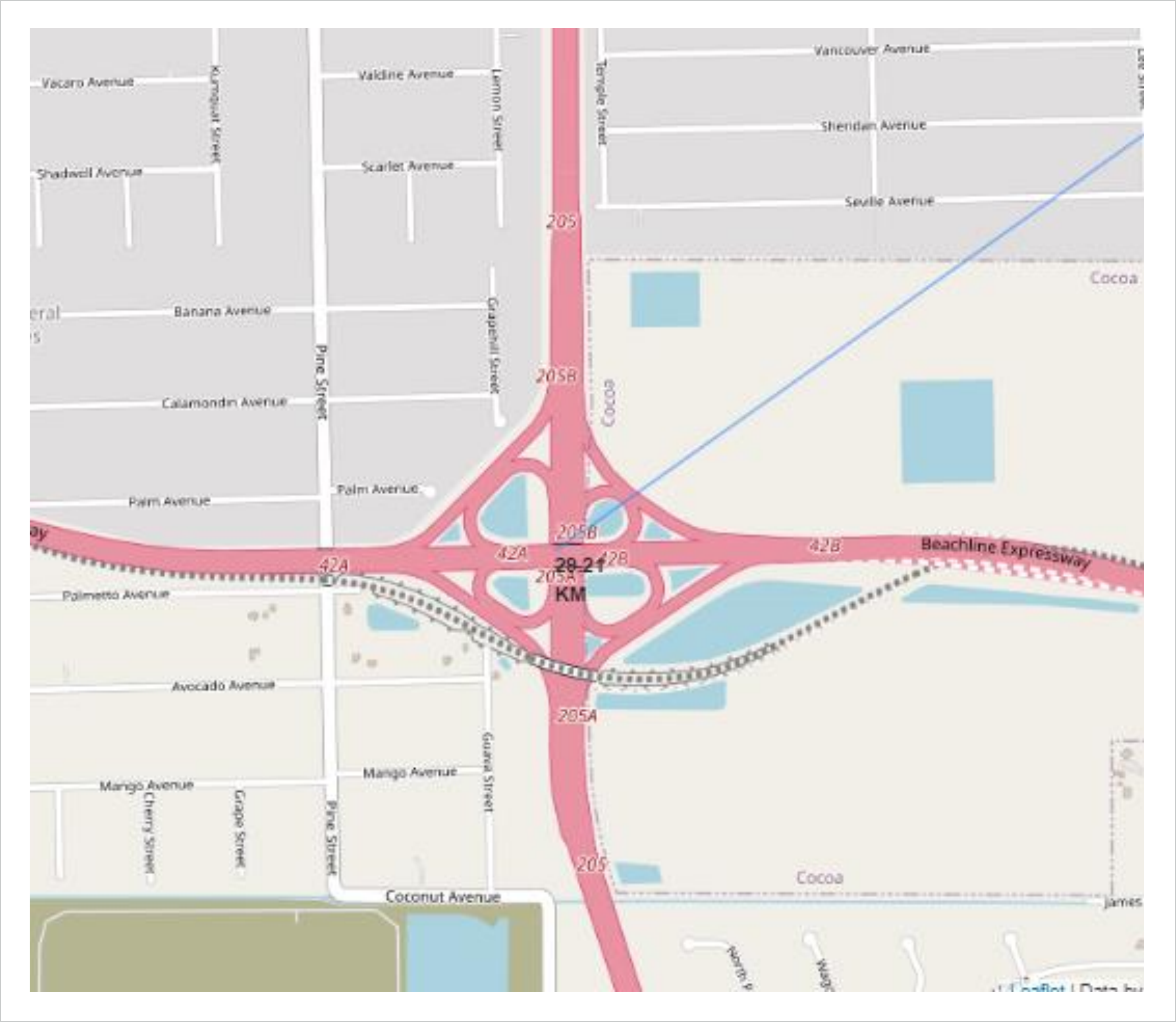
Success rate of each location

- Green markers indicates a successful landing and Red indicates otherwise



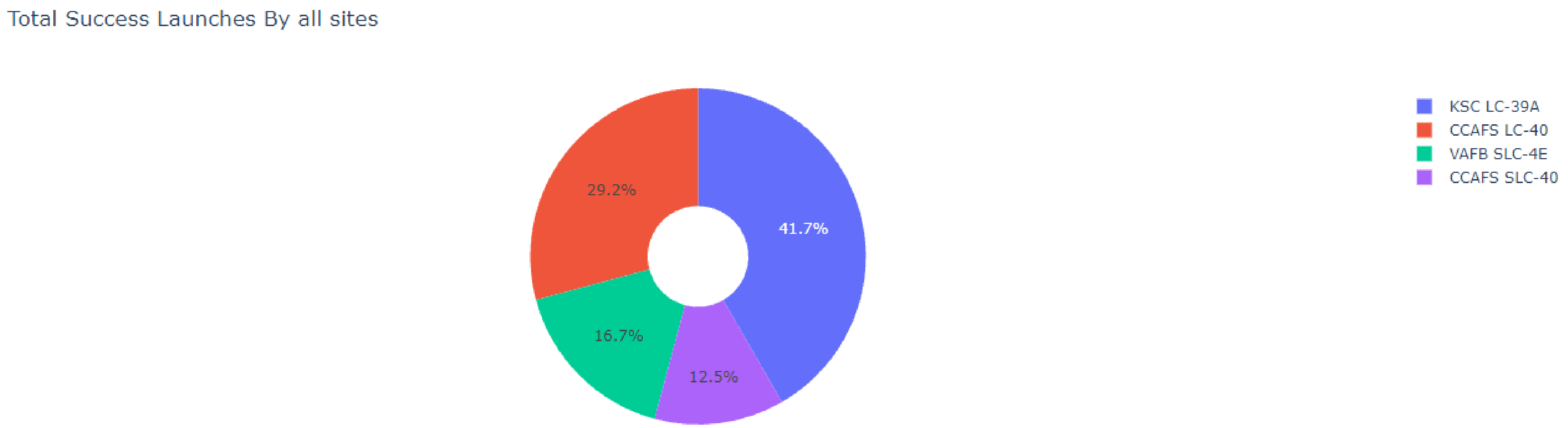
Distance to nearest Highway

- We also calculated the distance for several other locations souch as nearest coast



Section 5

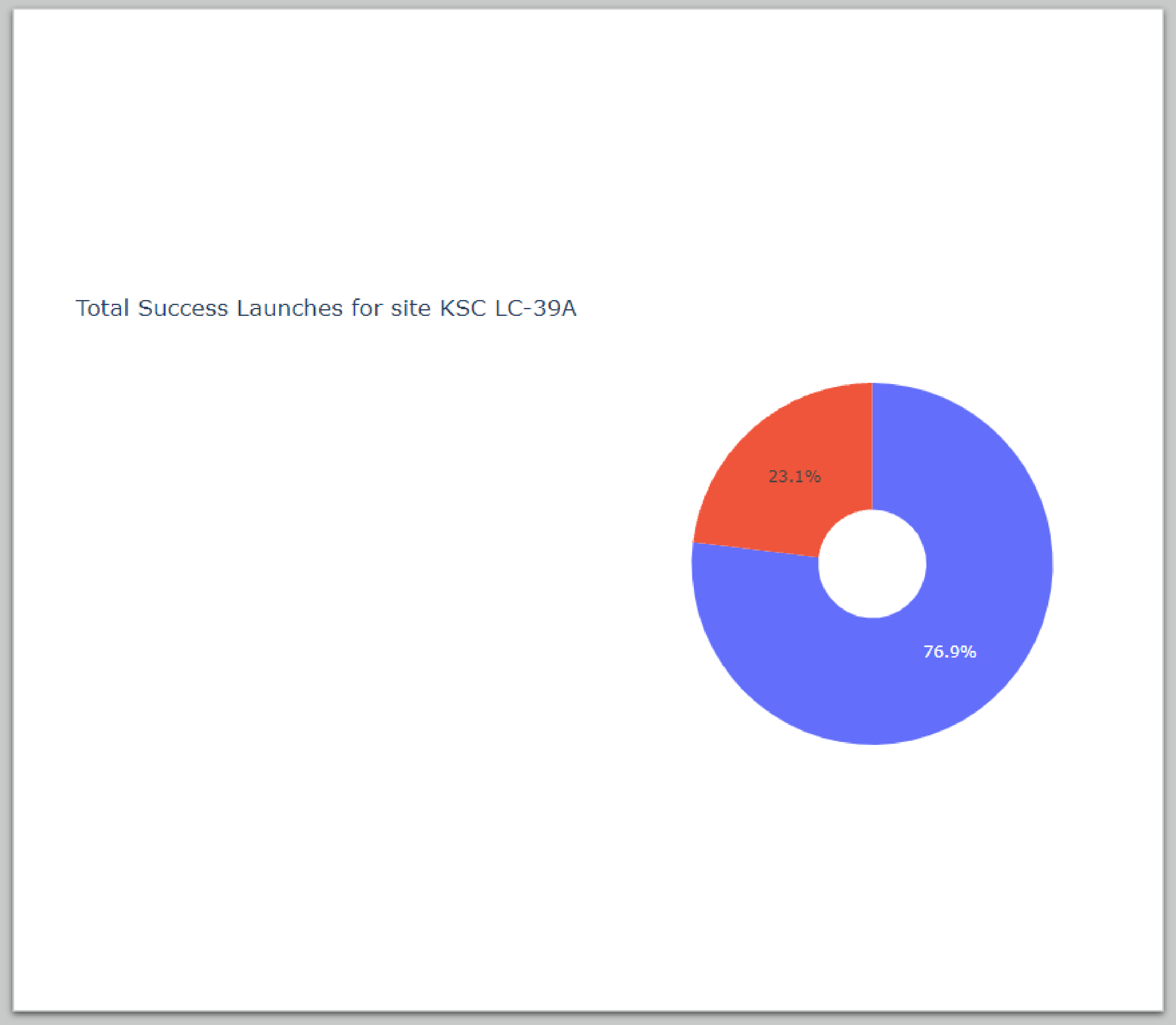
Build a Dashboard with Plotly Dash

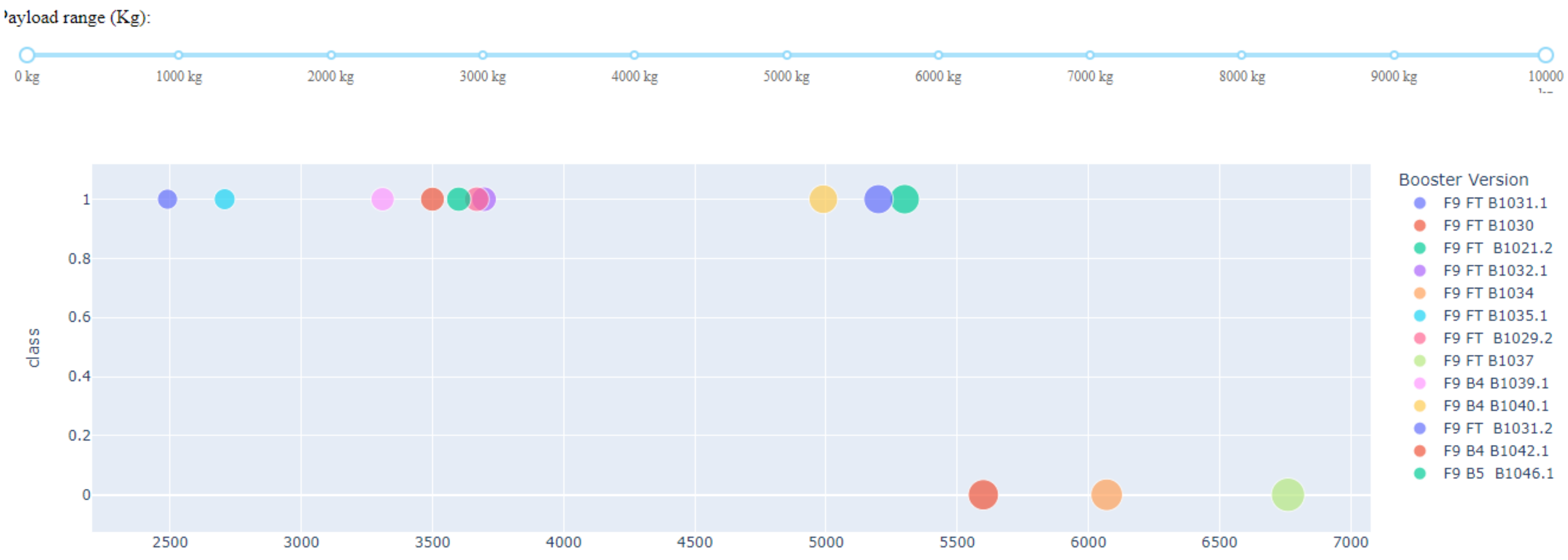


total number of launches from each site

Launch site success rate

- KSC LC-39A Have a Success rate of over 76%





the relationship between the outcome and pay load mass.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

```
Best Algorithm is Tree with a score of 0.8767857142857143
Best Params is : {'criterion': 'gini', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 5,
'splitter': 'random'}
```

As We can see after testing all the algorithms and comparing them we found out the decision trees are best classifier

Confusion Matrix

- Our classifier get's everything right exept for one FP

Conclusions

- From EDA we concluded that :
  - success rate increases as number of flights increases
  - the higher the payload mass is the better the success chances
  - Orbits GEO, HEO, SSO and ES-L1 has the best Success Rate
  - KSC LC-39A Have a Success rate of over 76%

Conclusions

- Using Machine learning we concluded that :
    - Best classifier is the Decision tree classifier
    - We can predict the future launches outcome with an accuracy around 88%

Thank you!