IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX

Executive Summary

The purpose of this project was to determine whether Falcon 9 first stage will land successfully or not using all skills and methodologies learnt in the Data Science course.

SpaceX 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 SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Methodologies used in this project included Data collection and data cleaning in python. Data wrangling was performed using SQL and python libraries such as pandas. Initial insights were gained about the data by performing visualization in python using libraries like matplotlib and seaborn. Finally, data analysis was performed by using regression techniques for predictive analysis using libraries like numpy and scikitlearn.

Results of analysis were studied to determine the best suited models for predictive analysis and the accuracy of results. This tells which method could be used to best predict the landing chances of the first stage launch pad so that it could be reused.

Introduction

PROJECT BACKGROUND AND CONTEXT

PROBLEMS YOU WANT TO FIND ANSWERS

Section 1

Methodology

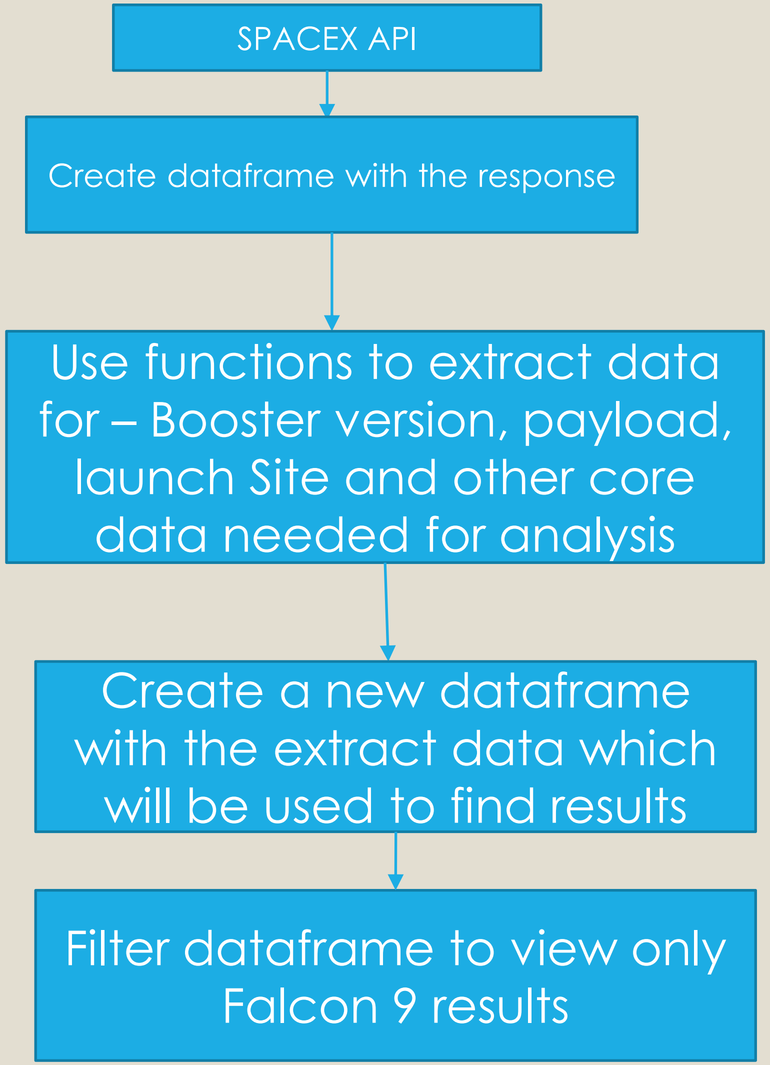
Methodology

- Executive Summary
- Data collection methodology:
  - Data collection was done by requests using the SpaceX API
- Perform data wrangling
  - The response from the api was then read into a datagram format for better viewing and access to data wrangling.
  - The necessary fields and values were then extracted and a new final pandas dataframe was generated which was used to do all analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The datasets were split into training and testing sets and various predictive models were used to determine which model was the best suited for the SpaceX dataset.
  - We analyzed the best Hyperparameter for SVM, Classification Trees and Logistic Regression

Data Collection – SpaceX API

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

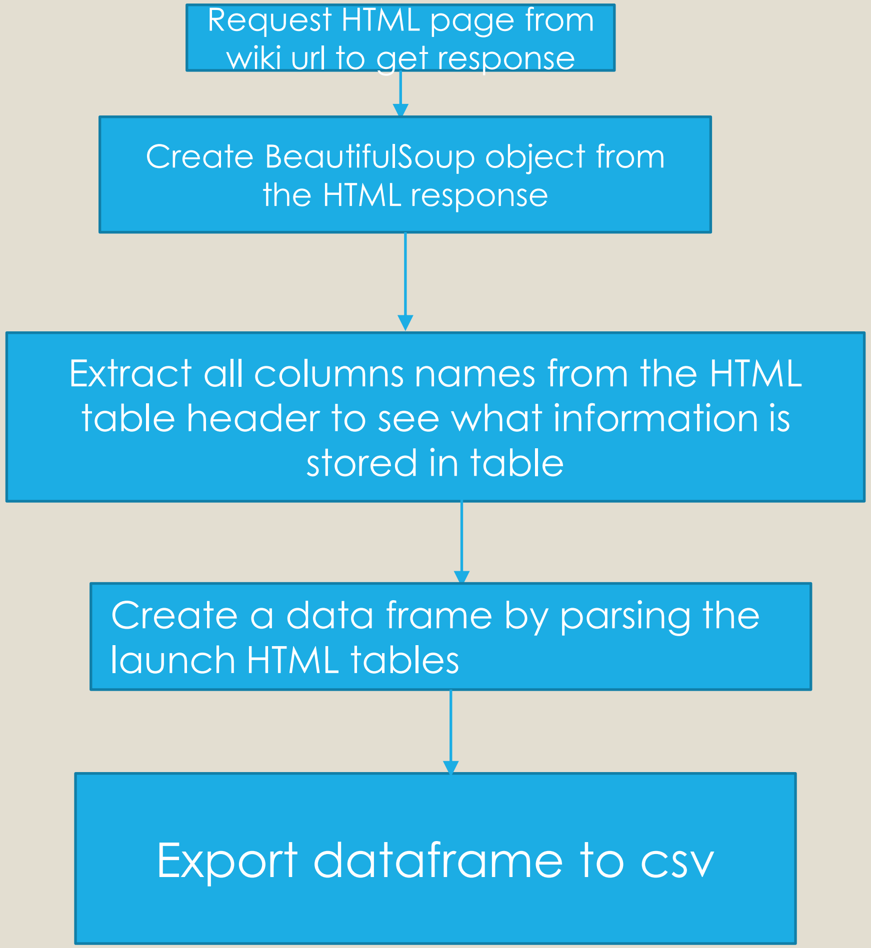
- https://github.com/vilasini29/Submissions/blob/main/Data%20Collection%20and%20web%20scraping%20-%20Capstone.ipynb



Data Collection - Scraping

- BeautifulSoup was used to scrape data from Wikipedia to extract launch resords

- https://github.com/vilasini29/Submissions/blob/main/Data%20Collection%20and%20web%20scraping%20-%20Capstone.ipynb



Data Wrangling

- Data wrangling is mainly the transforming and viewing of data in the early data analytics process
- Data wrangling in our project was done by performing some exploratory data analysis on the SpaceX Falcon 9 datasets that were generated in the previous steps.
- This process helps in getting a better understanding of the data we are working with and to decipher what type of further analysis would be needed to gain insights. Mainly we did some wrangling to understand the number of Launch site and the associated landing outcomes for those.
- https://github.com/vilasini29/Submissions/blob/main/Data%20wrangling.ipynb

EDA with Data Visualization

- Data visualization is great method to do initial EDA to gai insights on the dataset we are working with.
- This helps understand the relationship between the variables that we are working with which will help us decide with variable to emphasis on during our predictive analysis.
- Libraries used for this included pandas, numpy, Matplotlib and Seaborn
- Relationship between the Launch Sites and Flight number, Orbit Type and Payload mass was studied by creating a scatter plot
- The success rate for each orbit was studies using a bar chart
- Lastly the launch success over the years was understood by a line plot.
- https://github.com/vilasini29/Submissions/blob/main/Data%20Visualization%20-%20Capstone.ipynb

EDA with SQL

- Get unique Launch site names
- Filter Launch Sites beginning with 'CCA'
- Get Total payload mass carried by boosters launched by NASA(CRS)
- Avg payload mass carried by booster version
- Date of first successful landing outcome in ground pad
- 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
- Names of the booster versions which have carried the maximum payload mass
- Core details about failed landing outcomes
- Ranking the successful and failed landing outcomes in a period
- https://github.com/vilasini29/Submissions/blob/main/Data%20Analysis%20using%20SQL.ipynb

Build an Interactive Map with Folium

- We used Folium to generate interactive visual analytics
- After the initial folium map was generated, we used Circle and Marker functions to add launch sites to the map.
- MarkerCluster was used to mark the successful and failed launches at each ot the launch sites
- This visualization helps identifying the site locations and the proximity to other significant amenities including railways, highways and coastlines.
- https://github.com/vilasini29/Submissions/blob/main/Data%20visualization%20Folium%20-%20Capstone.ipynb

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
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- We created a machine learning pipeline to predict it he first stage will land or not. Various predictive models were used to determine the best model based on the accuracy of results
- Hyperparameters for SVM, classification trees and regression were determined to conclude the best models
- We split the data into training and testing sets with a 20% testing split
- We trained the various models and determined the best accuracy by analyzing the confusion matrix and the parameters
- https://github.com/vilasini29/Submissions/blob/main/Predictive%20analysis%20-%20Capstone.ipynb

Results

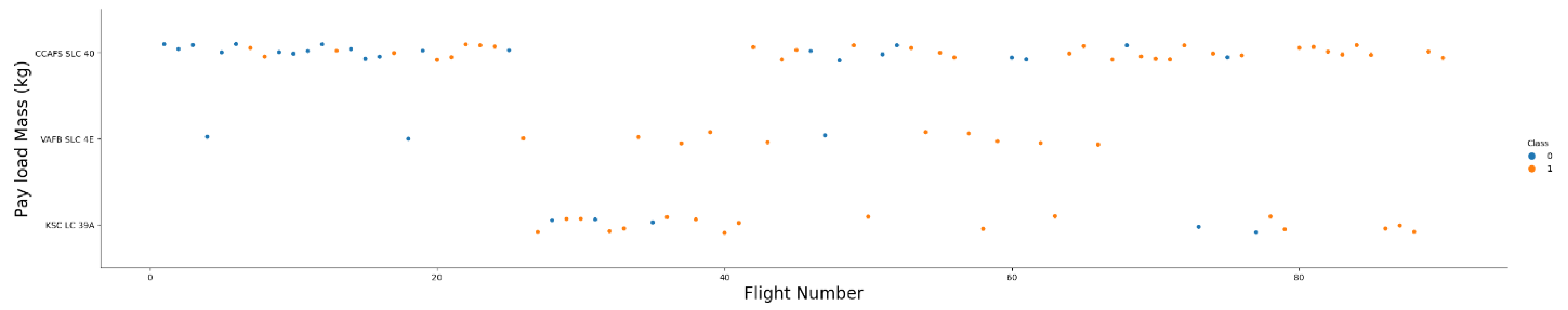
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

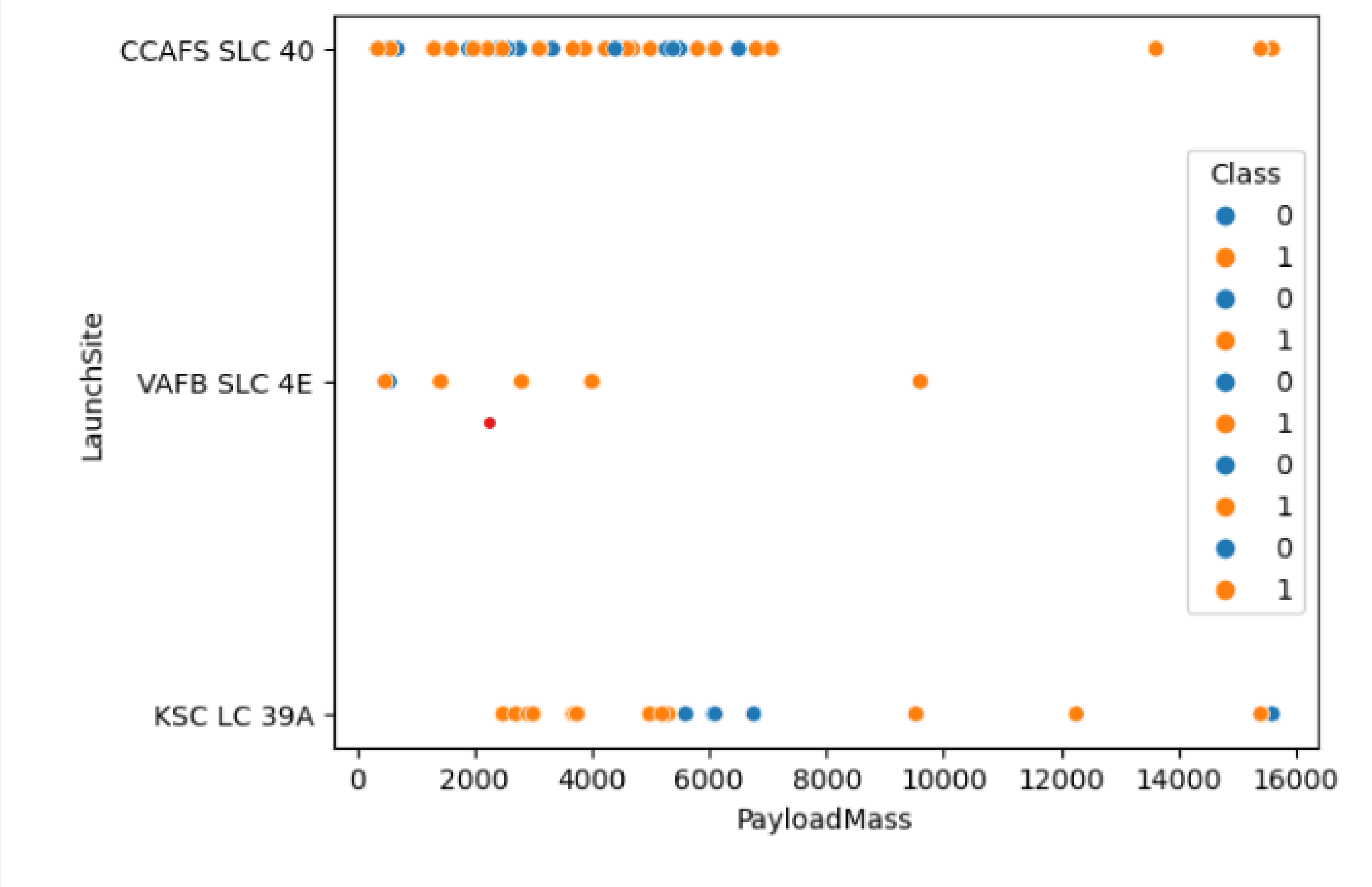
Flight Number vs. Launch Site

- The Class represents the success and failure class that were assigned to the datasets.
- The greater the pay load mass showed a more successful result.



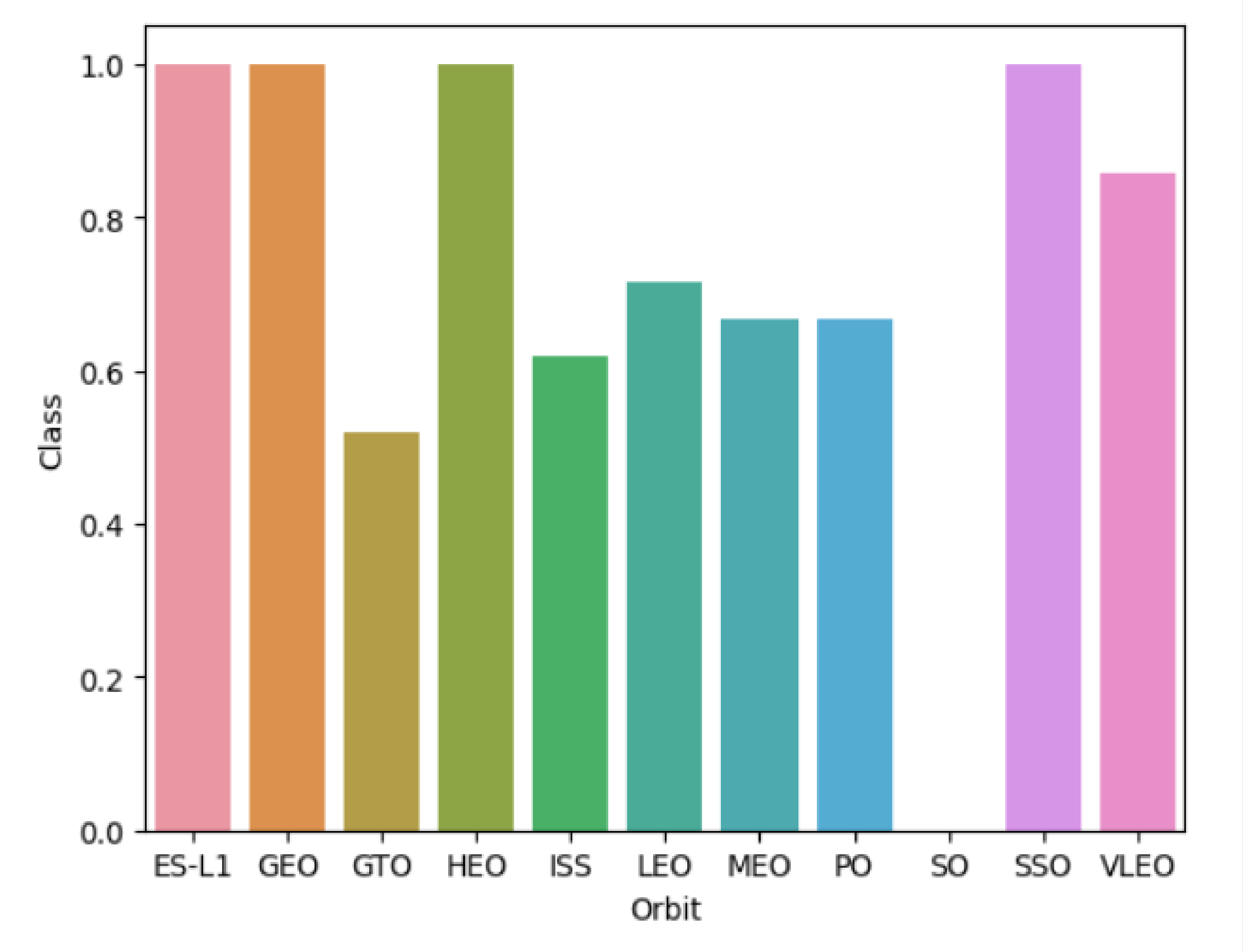
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)



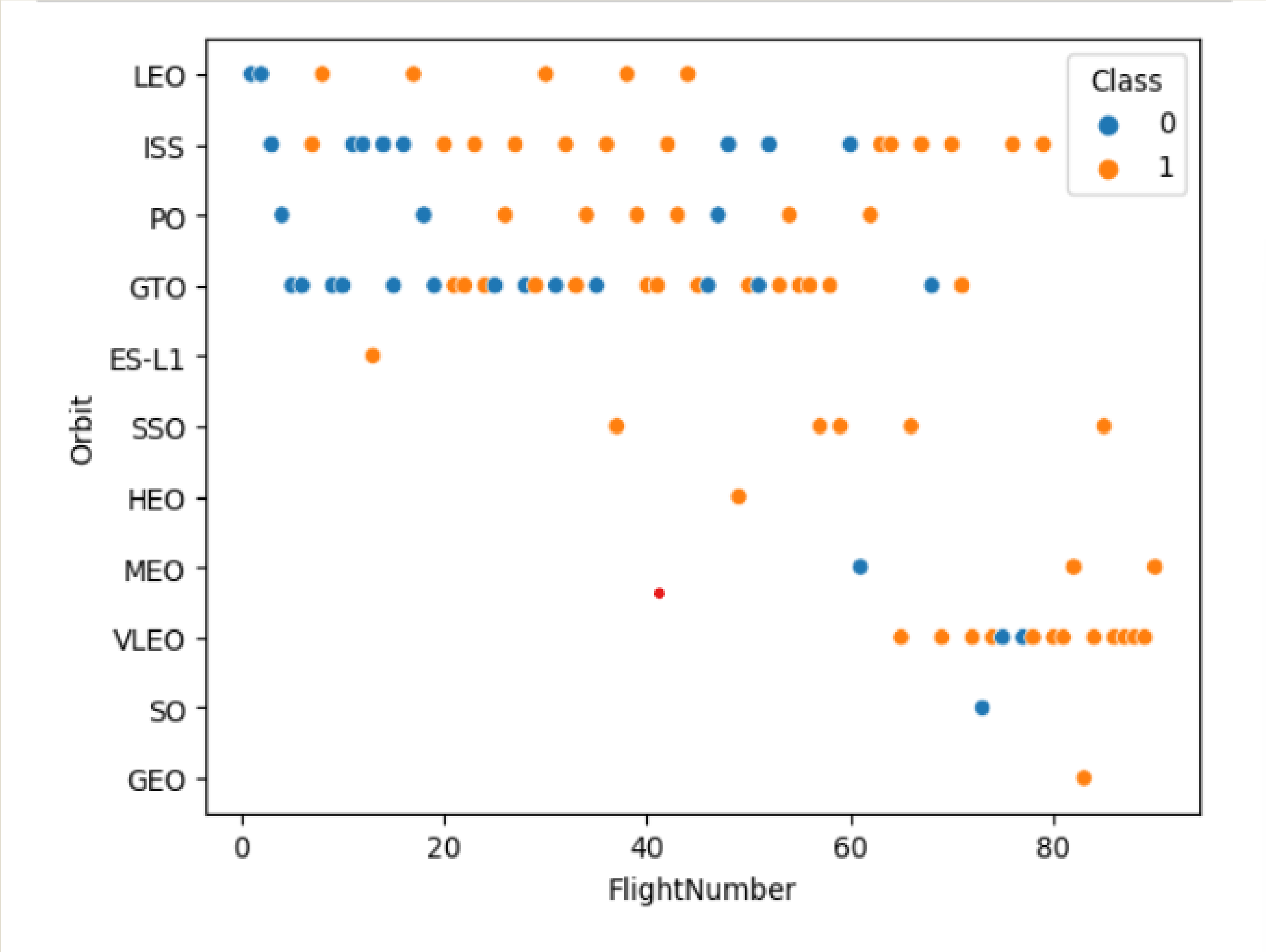
Success Rate vs. Orbit Type

- Class represents the success rate of each launch record which we are relating against the Orbit type.
- The bar chart shows the orbits that have the most successes which includes ES-L1, GEO, HEO, SSO and VLEO.



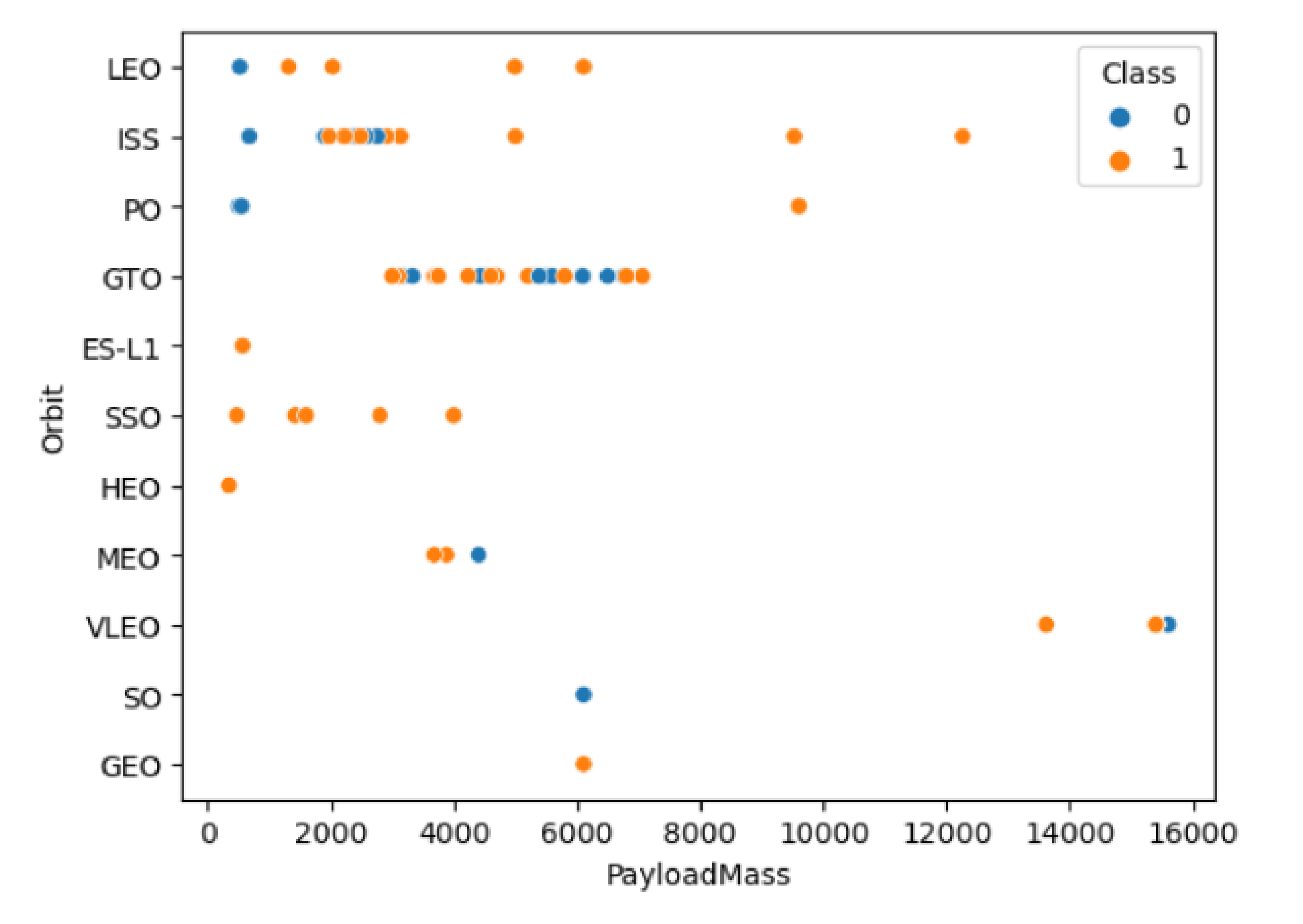
Flight Number vs. Orbit Type

- The following scatter plot shows the success rate with increasing flight numbers for each orbit.
- Analysis the results we can see the some of the orbits show a clear relationship however some, eg. GTO orbit doesn't clearly define the relationship with the increasing flight numbers.
- However, we can see the that VLEO and LEO has better success rate with increase in flight numbers.



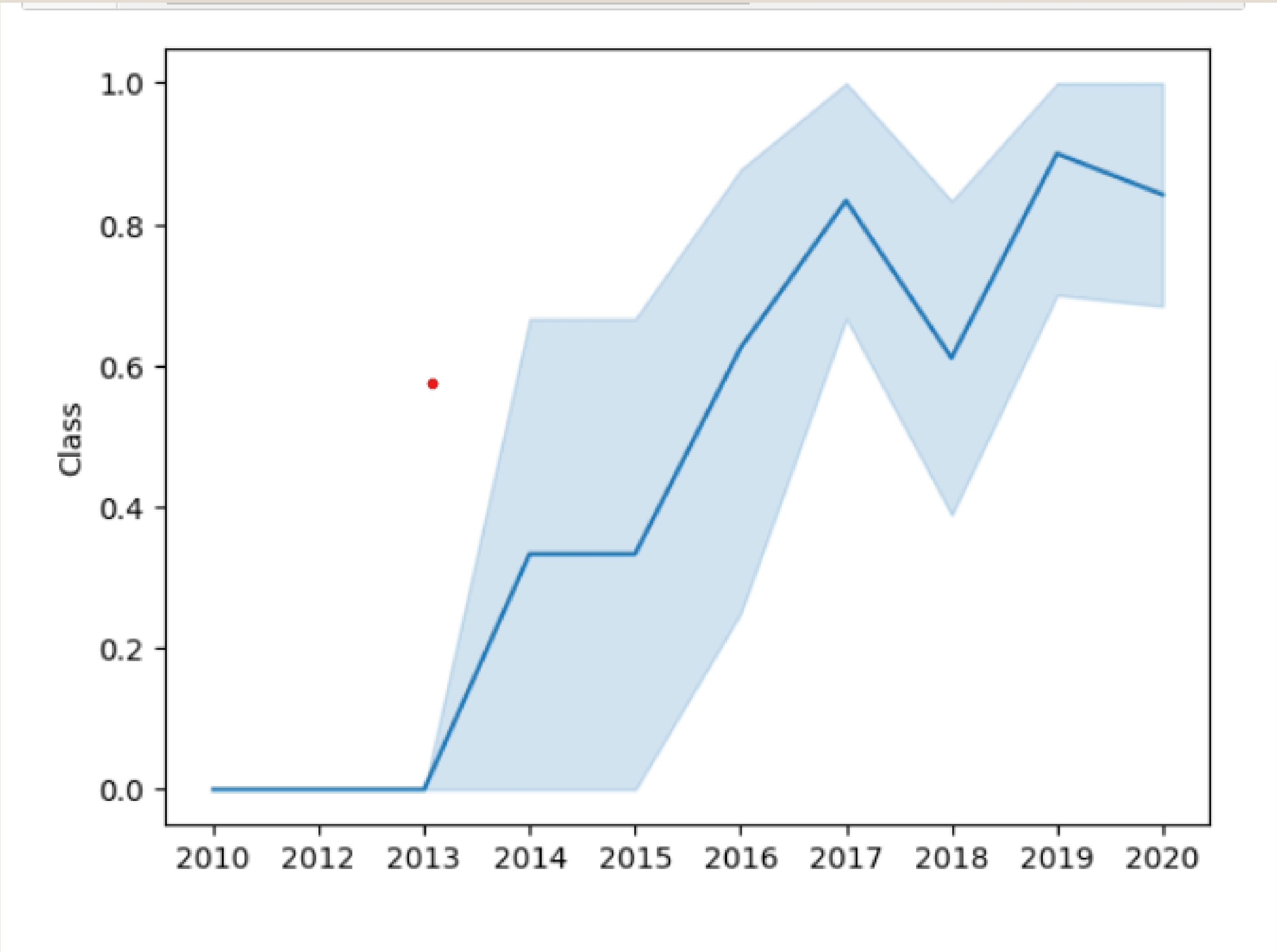
Payload vs. Orbit Type

- The scatter plot of orbit vs payload mass shows what combination has a more successful landing
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here



Launch Success Yearly Trend

- We plotted the success rate over the years
- Success rate since 2013 kept increasing till 2020



All Launch Site Names

- Find the names of the unique launch sites
- Presen

In [41]:

```
results = %sql SELECT UNIQUE(launch_site) FROM SPACEX_DATA
results
```

[redacted]
Done.

Out[41]:

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

Launch Site Names Begin with 'CCA'

- Fir
- Pr

```
%sql SELECT * FROM SPACEX_DATA WHERE launch_site LIKE 'CCA%' LIMIT 5
```

* [redacted]

Done.

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocean)
2014-09-21	05:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216	LEO (ISS)	NASA (CRS)	Success	Uncontrolled (ocean)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

Total Payload Mass

- Calc
- Pres

```
1 %sql SELECT sum(payload_mass__kg_) FROM SPACEX_DATA WHERE customer = 'NASA (CRS)'
```

* [redacted]
Done.

]:

1
23589

Average Payload Mass by F9 v1.1

- Calc
- Pres

```
%sql SELECT AVG(payload_mass__kg_) FROM SPACEX_DATA WHERE booster_version = 'F9 v1.1'
```

* [redacted]

Done.

]:

1
1806

First Successful Ground Landing Date

- Fin
- Pres

```
%%sql
SELECT min(DATE) FROM SPACEX_DATA
WHERE mission_outcome = 'Success'
```

```
* [redacted]
Done.
```

1
2012-05-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT booster_version FROM SPACEX_DATA
WHERE payload_mass__kg_ BETWEEN 4000 AND 6000
AND landing__outcome = 'Success (drone ship)'
```

```
 * [redacted]
Done.
```

booster_version
F9 FT B1026
F9 FT B1021.2

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT mission_outcome,COUNT(mission_outcome) FROM SPACEX_DATA
GROUP BY mission_outcome
```

```
* [redacted]
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	55

Boosters Carried Maximum Payload

```
%%sql
SELECT booster_version,payload_mass__kg_ FROM SPACEX_DATA
WHERE payload_mass__kg_ = (SELECT max(payload_mass__kg_) FROM SPACEX_DATA)
```

* [redacted]

Done.

]:

booster_version	payload_mass__kg_
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

```
%%sql
SELECT landing__outcome,booster_version,launch_site FROM SPACEX_DATA
WHERE landing__outcome LIKE 'Failure%'
AND YEAR(DATE) = '2016'
```

```
* [redacted]
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
Failure (drone ship)	F9 FT B1024	CCAFS LC-40

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

```
%%sql
SELECT landing__outcome,COUNT(landing__outcome) FROM SPACEX_DATA
WHERE DATE Between '06-04-2010' AND '03-20-2017'
GROUP BY landing__outcome
```

* [redacted]

Done.

landing__outcome	2
Controlled (ocean)	2
Failure (drone ship)	3
No attempt	3
Precluded (drone ship)	1
Success (drone ship)	3
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

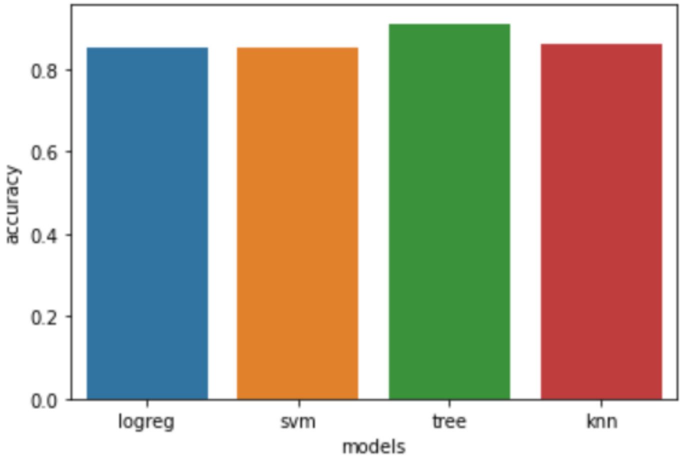
<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

Section 5

Predictive Analysis (Classification)

```
sns.barplot(data =acc_df,x='models',y='accuracy')
plt.show()
```

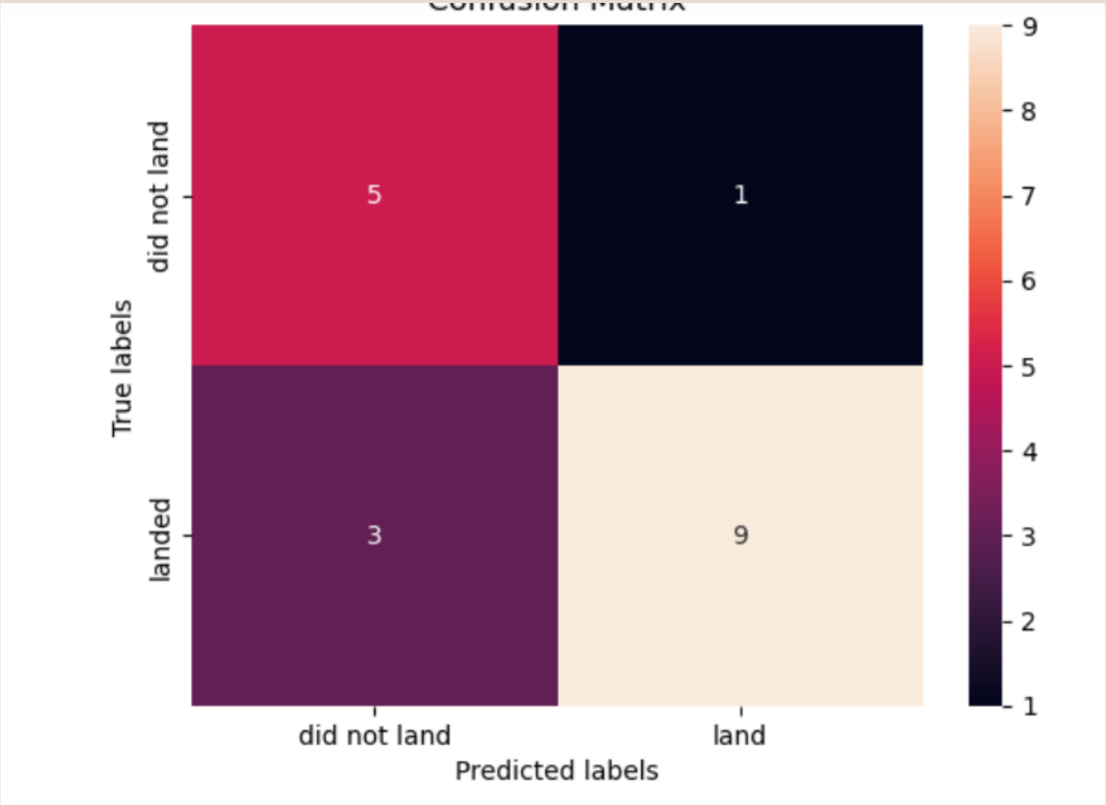


Classification Accuracy

- If we visualize the accuracies of all the models we can see that the tree plot has the best results as based ono the accuracy.

Confusion Matrix

- According to the accuracy plotting via confusion matrix for the Tree plot we can see that 5 out 6 were predicted correctly as did not land.
- We can also see that 3 out of 12 were predicted incorrectly for landed which is a 25% incorrect prediction.



Conclusions

- This capstone project helped us use different techniques for data science and analysis to determine results for real life dataset.
- We utilized various languages including SQL and python to predicted whether a Falcon9 rocket launcher first stage will have successful landing.
- We determined the combination of Orbit and Launch site, booster version and the launch site to know the best possible combination to get successful results.
- We also analyzed the predictive models to use in this analysis however, we noticed that not all models have the same level of accuracy.
- These conclusions will help an alternate company determine how it wants to bid against SpaceX for a rocket launcher.

Appendix

- All jupyter notebooks have been shared to the github repository
- https://github.com/vilasini29/Submissions/commits?author=vilasini29

Thank you!