IBM Developer
SKILLS NETWORK

Haifa Al-Nasi
02-5-2023

Outline

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

- ***In this final project, we will use various machine learning classification techniques to forecast whether the first stage of the SpaceX Falcon 9 rocket will land successfully.***
- ***The main steps in this project include:***
  - ✓***Data collection, wrangling, and formatting***
  - ✓***Exploratory data analysis***
  - ✓***Interactive data visualization***
  - ✓***Machine learning prediction***
- ***According to our charts, there is a relationship between certain characteristics of the rocket launches and whether they succeed or fail.***
- ***We have also determined that the decision tree algorithm may be the most effective machine learning method for predicting the successful landing of the Falcon 9 first stage.***

Executive Summary

Introduction

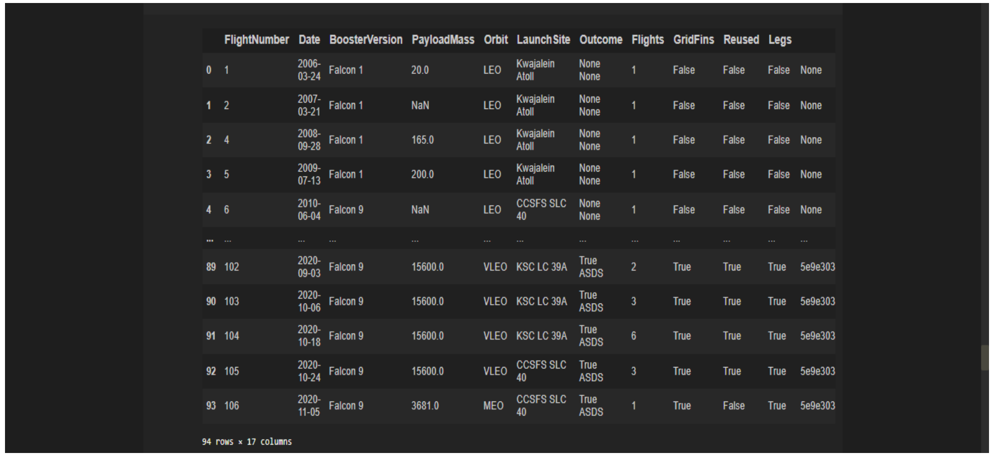
- *In this capstone , we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage . 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.*
- *The majority of unsuccessful landings are intentional. On occasion, SpaceX will execute a controlled landing in the ocean.*
- *The main question that we are trying to answer is , for a given set of features about Falcon 9 rocket launch which includes its payload mass, orbit type, launch site, and so on, will the first stage of the rocket land successfully?*

Methodology

METHODOLOGY

- ***1 Data collection, wrangling, and formatting***

- SpaceX API
  - The API used is https://api.spacexdata.com/v4/rockets/
  - The API provides data about many types of rocket launches done by SpaceX, the data is therefore filtered to include only Falcon9 launches.
  - Every missing value in the data is replaced by the mean of the column that the missing value belongs to .
  - We end up with 90 rows or instances and 17 columns or features. The picture below shows the first few rows of the data:



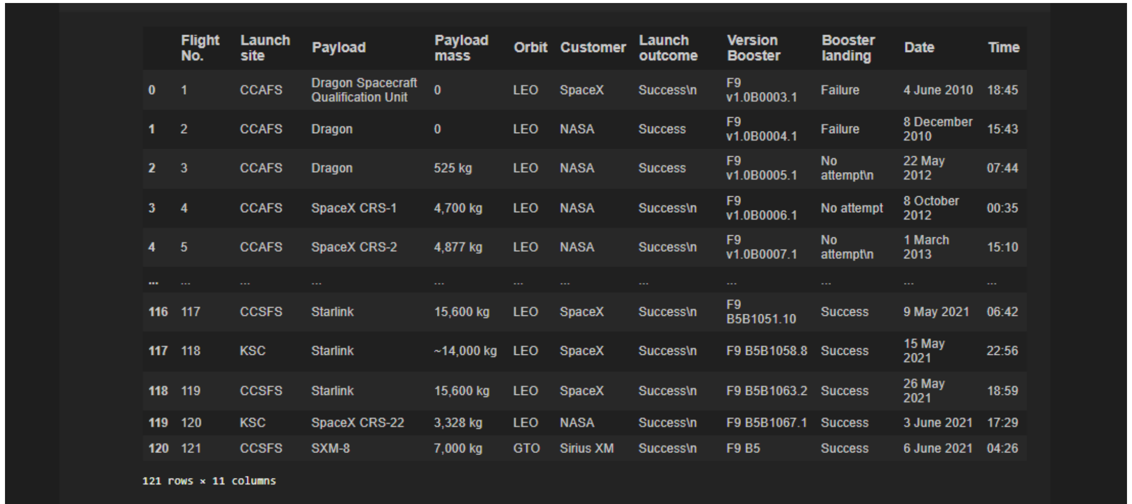
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None
...	...	...	...	...	...	...	...	...	...	...	...	...
89	102	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e303
90	103	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e303
91	104	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e303
92	105	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e303
93	106	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e303

94 rows × 17 columns

METHODOLOGY

- *1 Data collection, wrangling, and formatting*

- Web scraping
  - The data is scraped from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Heavy_launches&oldid=1027686922
  - The website contains only the data about Falcon9 launches.
  - We end up with 121 rows or instances and 11 columns or features. The picture below shows the first few rows of data:



	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
116	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	06:42
117	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8	Success	15 May 2021	22:56
118	119	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1063.2	Success	26 May 2021	18:59
119	120	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA	Success\n	F9 B5B1067.1	Success	3 June 2021	17:29
120	121	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	F9 B5	Success	6 June 2021	04:26

```
121 rows × 11 columns
```

METHODOLOGY

- *1 Data collection, wrangling, and formatting*

- The data is later processed so that there are no missing entries and categorical features are encoded using on-hot encoding.
- An extra column called 'Class' is also added to the data frame . The column 'Class' contains 0 if a given launch is failed and 1 if it is successful.
- In the end , we end up with 90 rows or instances and 83 columns or features

METHODOLOGY

2 Exploratory Data Analysis (EDA)

- Pandas and NumPy
  - Functions from the Pandas and NumPy libraries are used to derive basic information about the data collected, which includes :
    - The number of launches on each launch site
    - The number of occurrence of each orbit
    - The number and occurrence of each mission outcome
  - SQL
    - The data is quired using SQL to answer several questions about the data such as:
      - The names of the unique launch sites in the space mission
      - The total payload mass carried by boosters launched by NASA (CRS)
      - The average payload mass carried by booster version F9 v1.1

METHODOLOGY

3 Data Visualization

- Matplotlib and Seaborn
  - Functions from the Matplotlib and Seaborn libraries are used to visualize the data through scatterplots, bar charts, and line charts.
  - The plots and charts are used to understand more about the relationships between several features, such as:
    - The relationships between flight number and launch site
    - The relationship between payload mass and launch site
    - The relationship between success rate and orbit type
  - Folium
    - Functions from the Folium libraries are used to visualize the data through interactive maps.
      - The Folium library is used to :
        - Mark All launch sites on a map
        - Mark the successeded launches and failed launches for each site on the map
        - Mark the distance between a launch site to its proximities such as the nearest city, railway, or highway

METHODOLOGY

3 Data Visualization

- Dash
  - Functions from Dash are used to generate an interactive site where we can toggle the input using a dropdown menu and a range slider .
  - Using a pie chart and a scatterplot, the interactive site shows:
    - The total success launches from each launch site .
    - The correlation between payload mass and mission outcome (success or failure) for each launch site.

METHODOLOGY

4 Machine Learning Prediction

- Functions from Scikit-learn library are used to create our machine learning models.
- The machine learning prediction phase includes the following steps:
  - Standardizing the data
  - Splitting the data into training and test data
  - Creating machine learning models, which include:
    - Logistic regression
    - Support vector machine (SVM)
    - Decision tree
    - K nearest neighbours (KNN)
  - Fit the models on the training set
  - Find the best combination of hyperparameters for each model
  - Evaluate the models based on their accuracy scores and confusion matrix

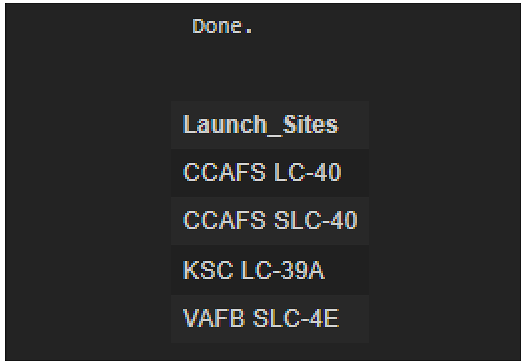
RESUTLS

- The results are split into 5 sections:
  - SQL(EDA with SQL)
  - Matplotlib and seaborn(EDA with Visualization)
  - Folium
  - Dash
  - Predictive Analysis
- In all of the graphs that follow, class 0 represents a failed launch outcome while class 1 represents a successful launch outcome.

RESULTS

- *1 SQL (EDA with SQL)*

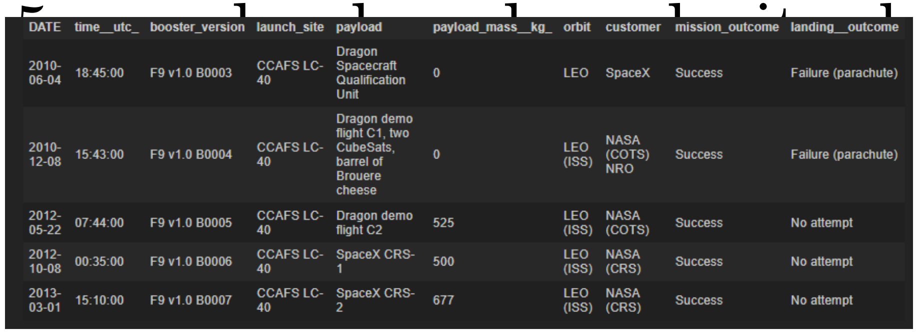
- The names of the unique launch sites in the space mission:



```
Done.
```

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

- 5 records where launch sites begin with 'CCA'

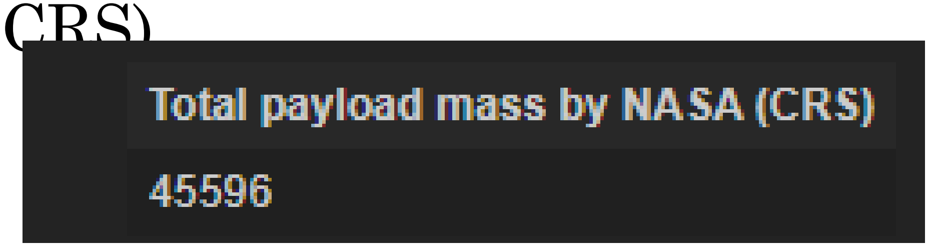


DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
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
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

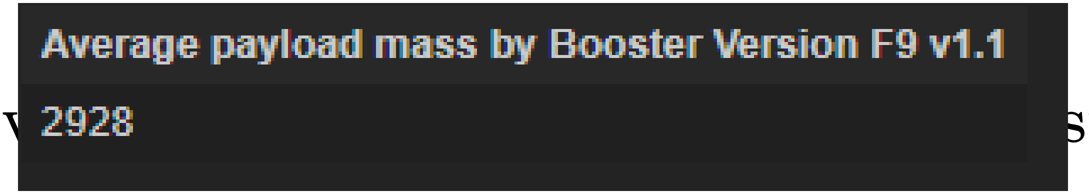
RESULTS

- *1 SQL (EDA with SQL)*

- The total payload mass carried by boosters launched by NASA(CRS)

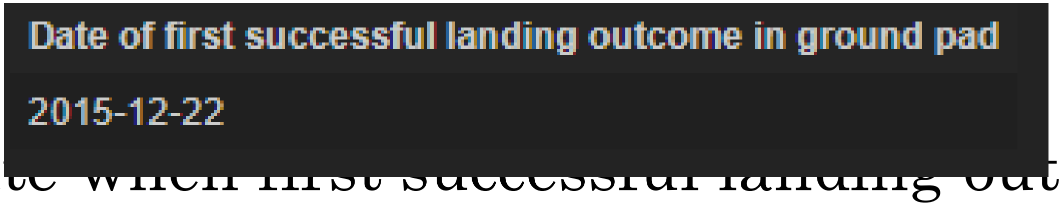


Total payload mass by NASA (CRS)
45596



Average payload mass by Booster Version F9 v1.1
2928

- The av ster version F9 v1.1



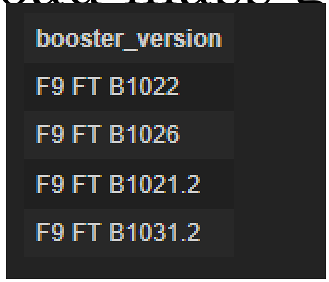
Date of first successful landing outcome in ground pad
2015-12-22

- The date when first successful landing outcome in ground pad was

RESULTS

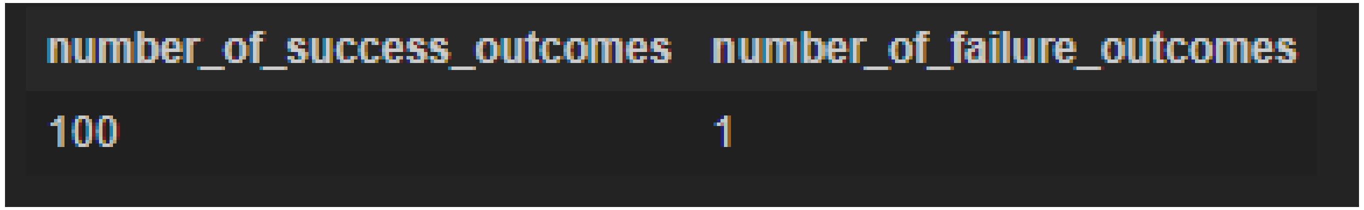
- *1 SQL (EDA with SQL)*

- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000



booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- The total number of successful and failure mission outcomes

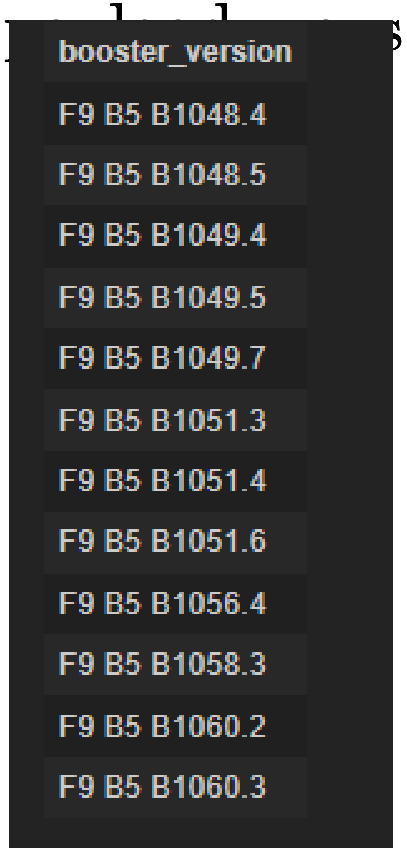


number_of_success_outcomes	number_of_failure_outcomes
100	1

RESULTS

- *1 SQL (EDA with SQL)*

- The names of the boosters which have carried the maximum payload mass

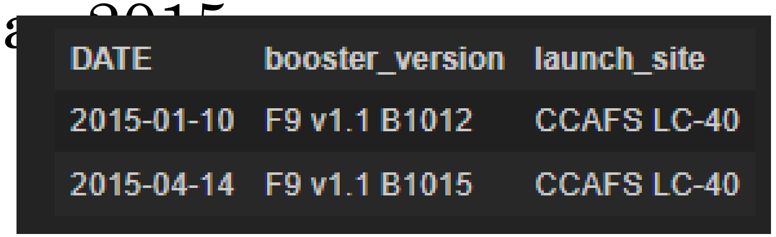


booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

RESULTS

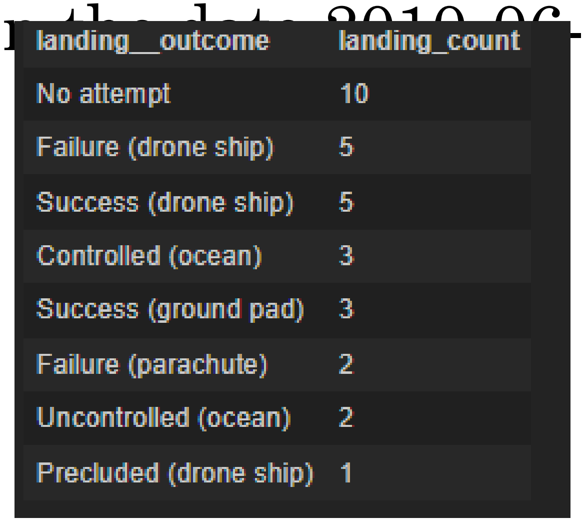
- *1 SQL (EDA with SQL)*

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015



DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

- The count of landing outcomes between the date 2010-06-04 and 2017-04-20, in descending order

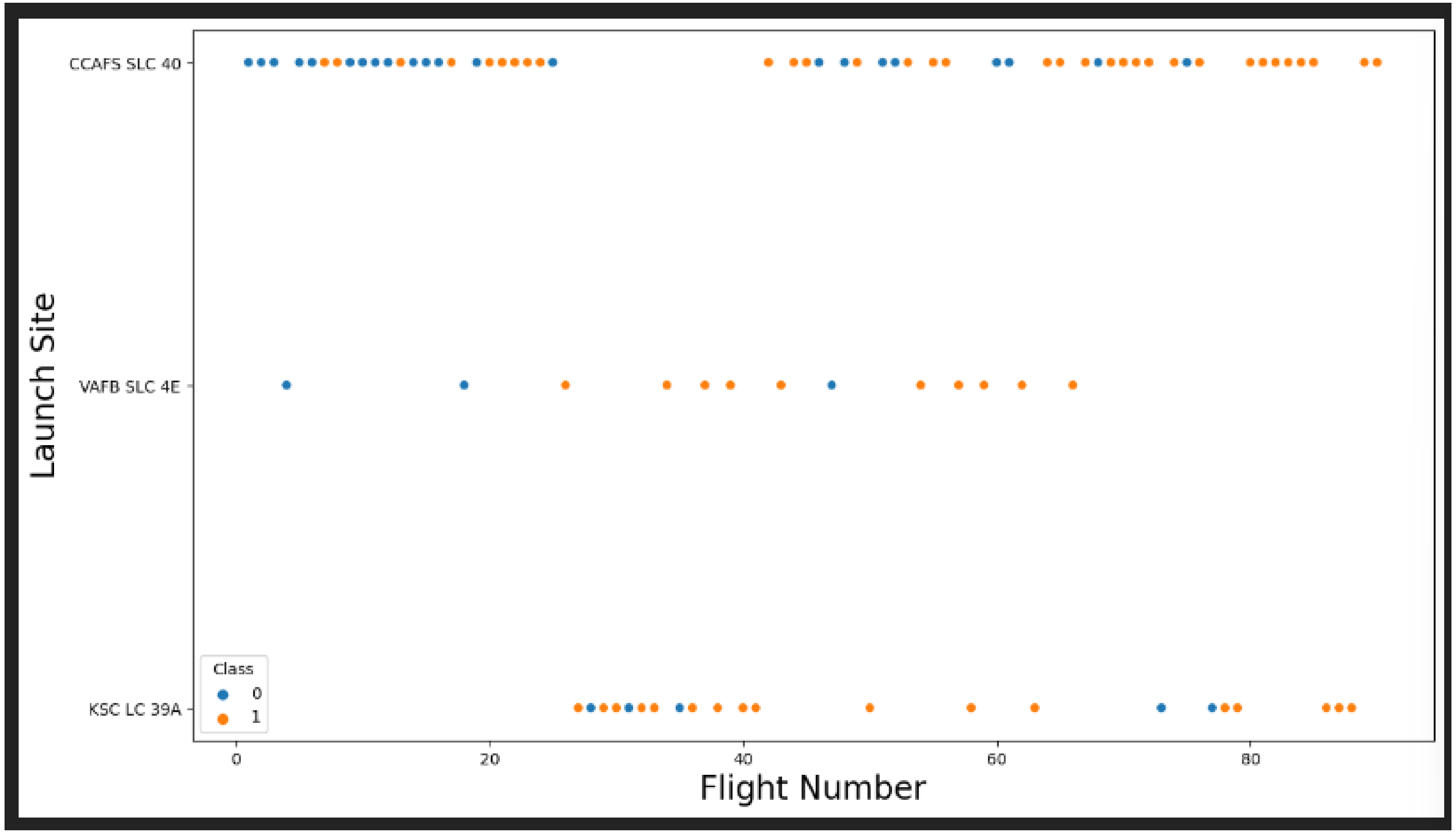


landing__outcome	landing_count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

RESUTLS

- **2 Matplotlib and Seaborn (EDA with Visualization)**

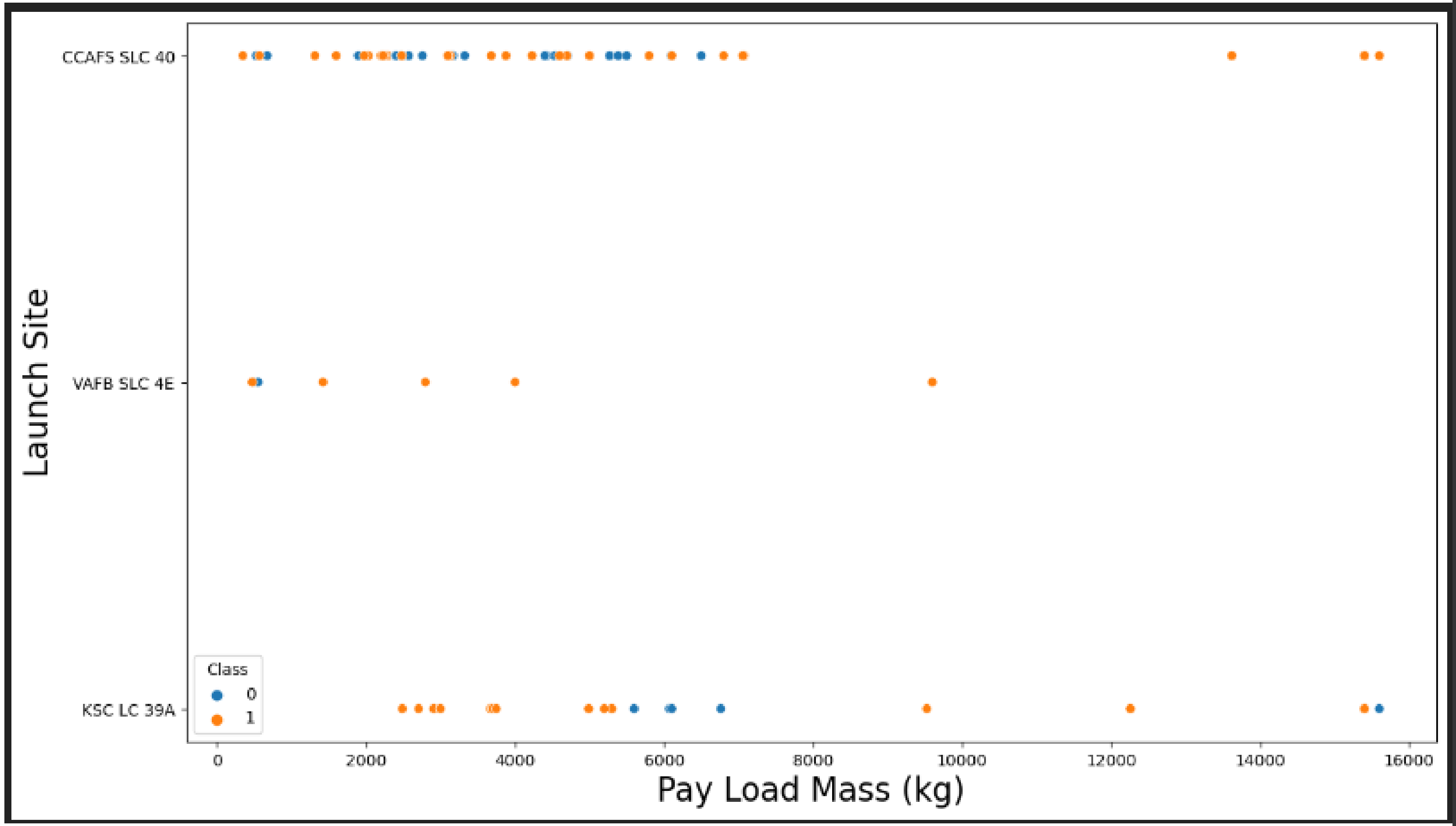
- The relationship between flight number and launch site



RESUTLS

- **2 Matplotlib and Seaborn (EDA with Visualization)**

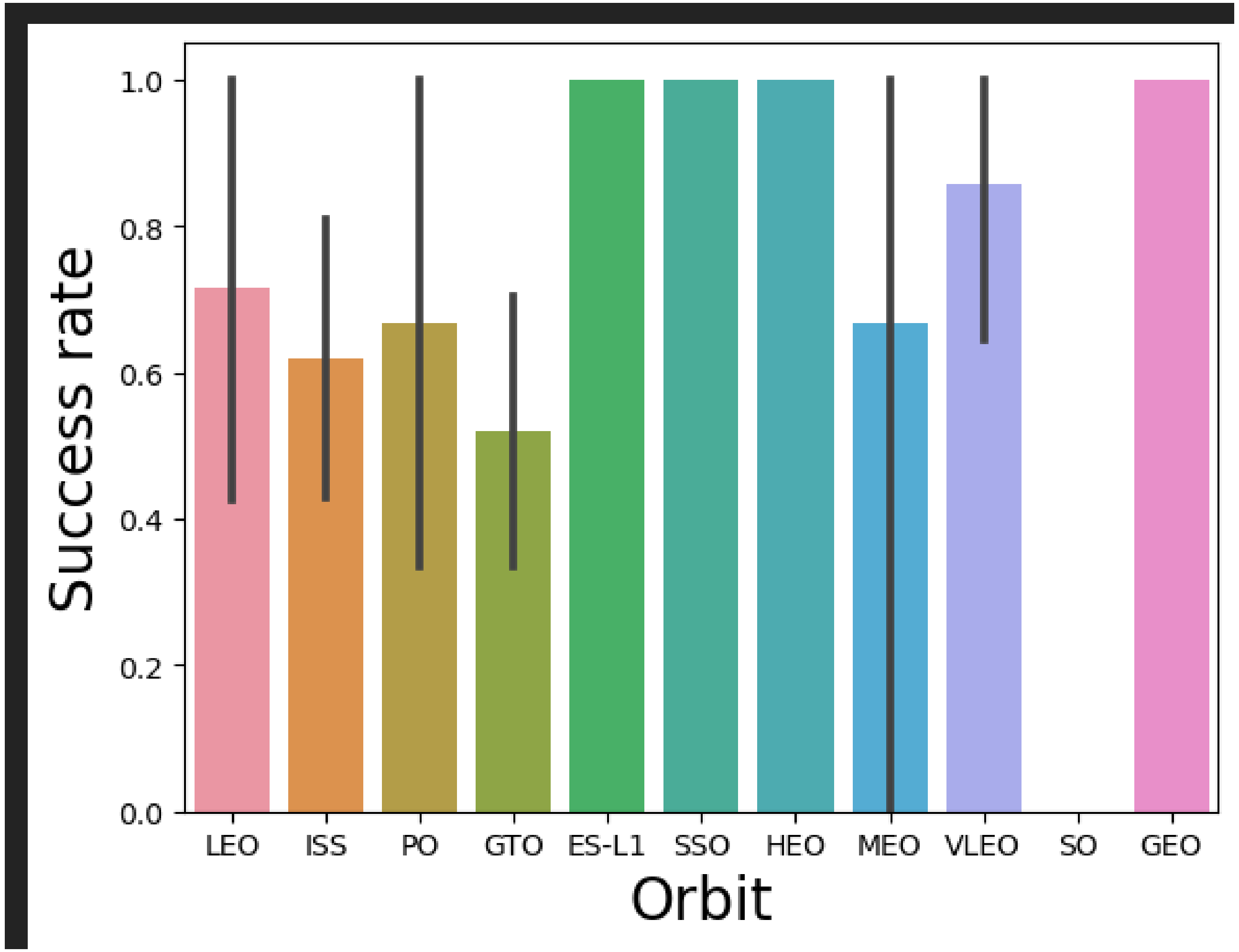
- The relationship between payload mass and launch site



RESUTLS

- **2 Matplotlib and Seaborn (EDA with Visualization)**

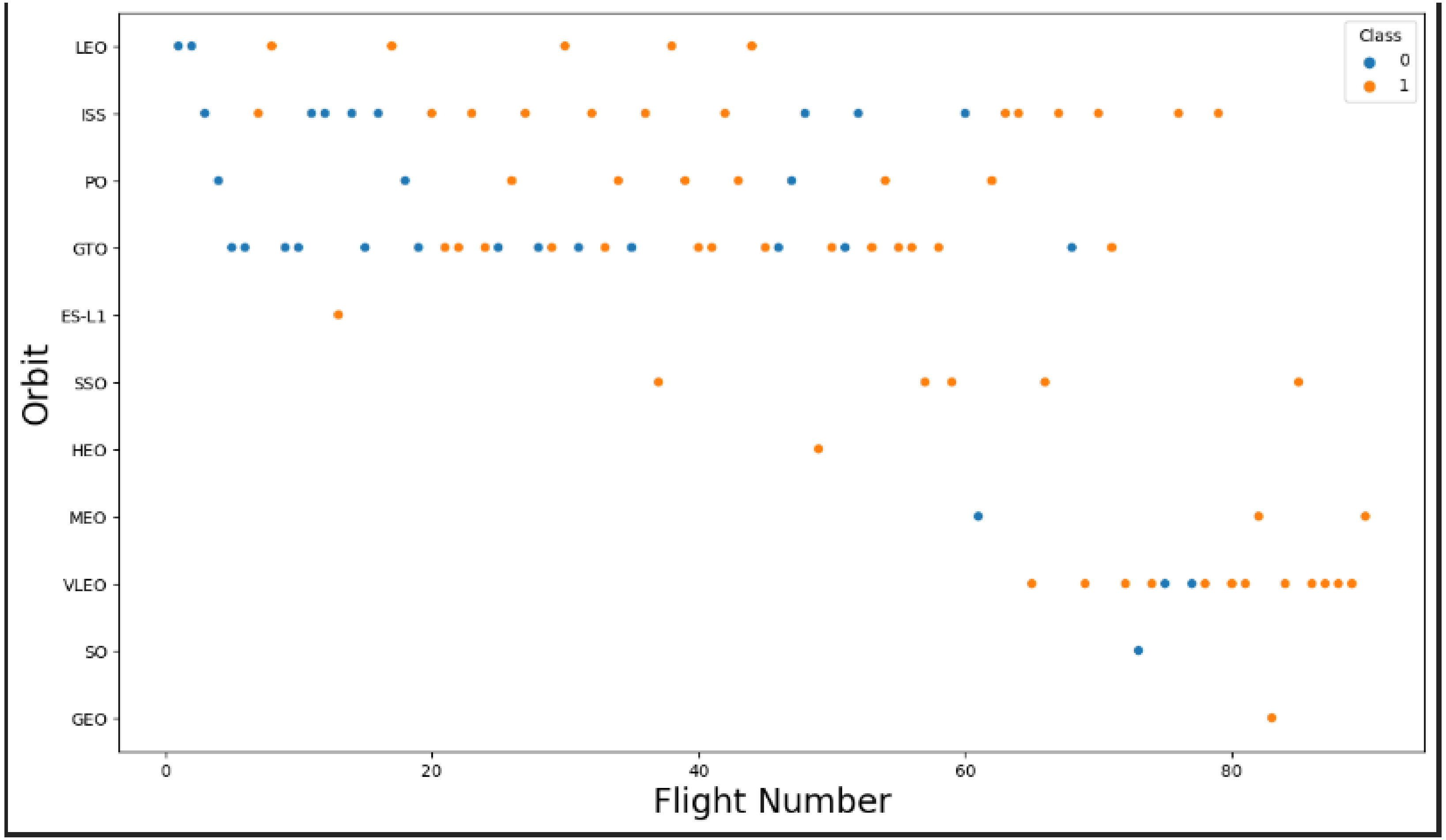
- The relationship between success rate and orbit type



RESUTLS

- # 2 Matplotlib and Seaborn (EDA with Visualization)

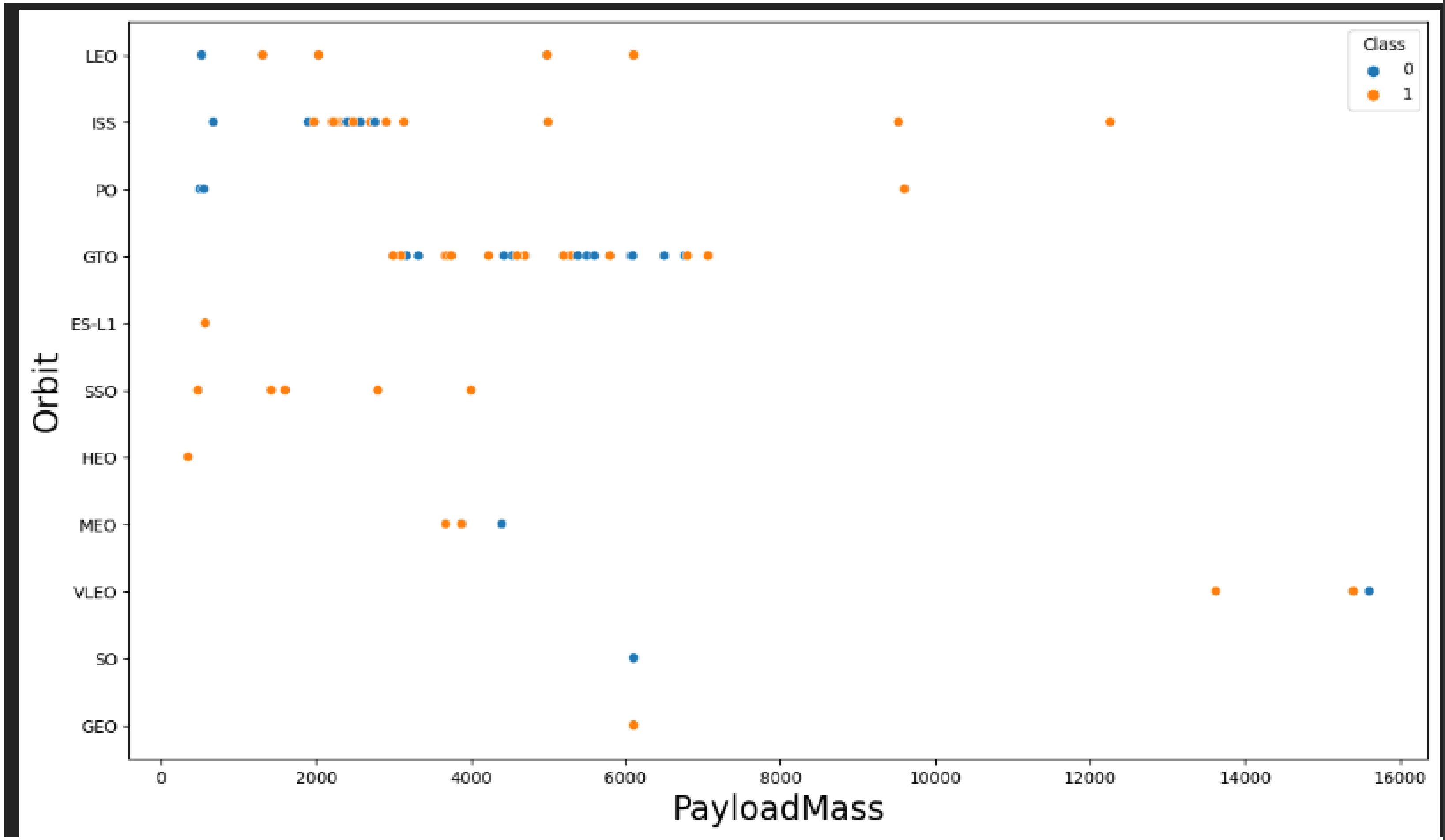
- The relationship between flight number and orbit type



RESUTLS

- # 2 Matplotlib and Seaborn (EDA with Visualization)

- The relationship between payload mass and orbit type

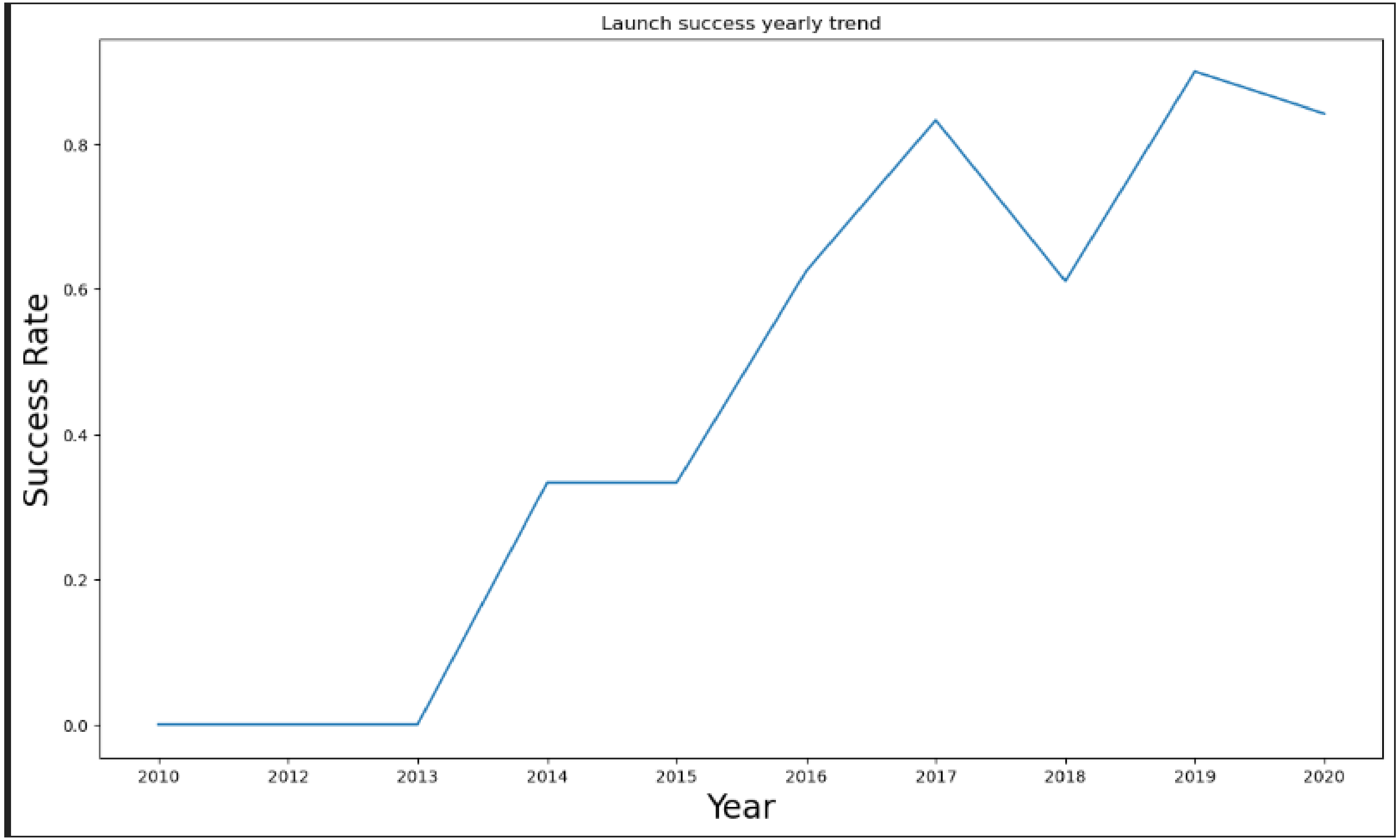


RESUTLS

-

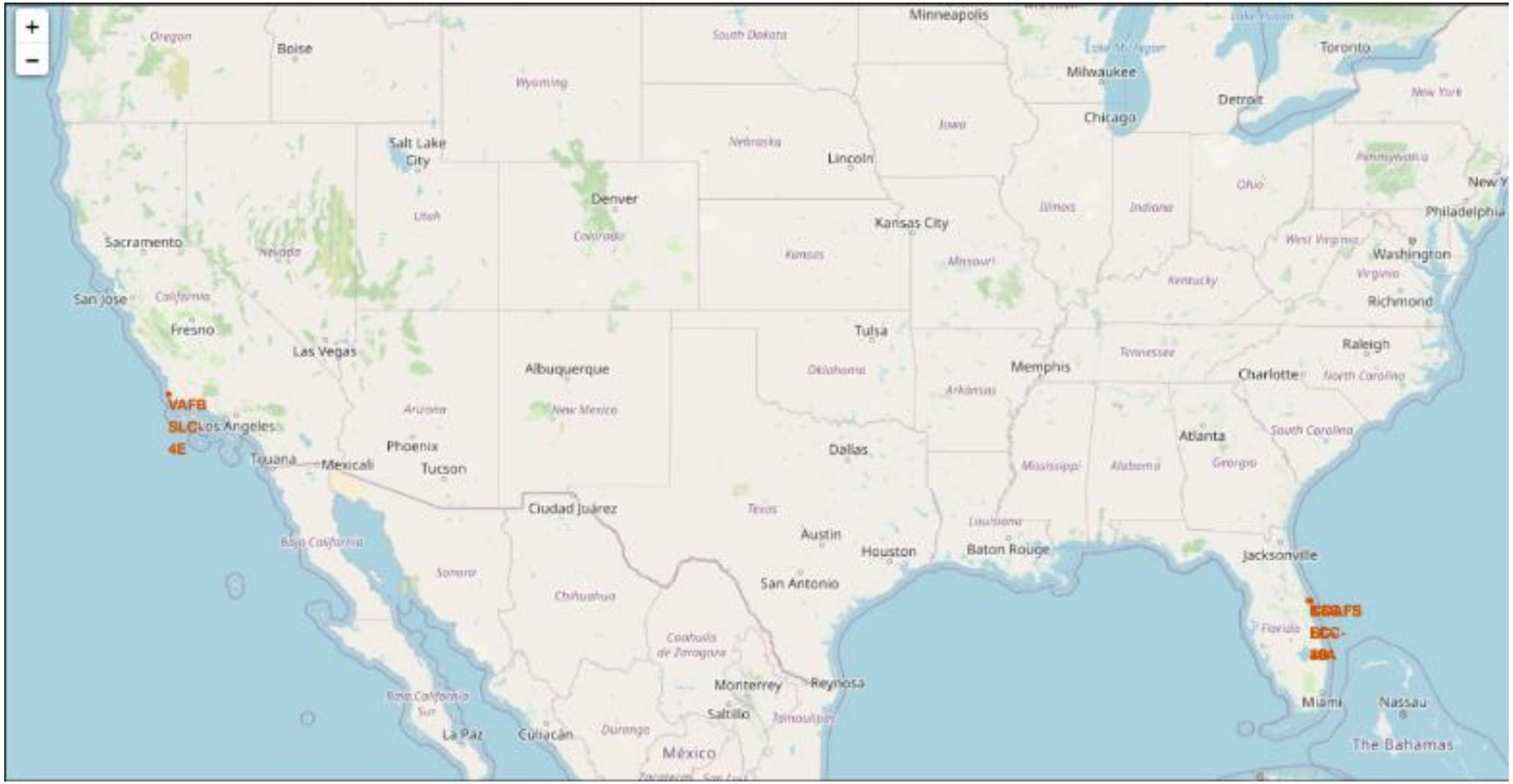
2 Matplotlib and Seaborn (EDA with Visualization)

- The launch success yearly trend



Folium

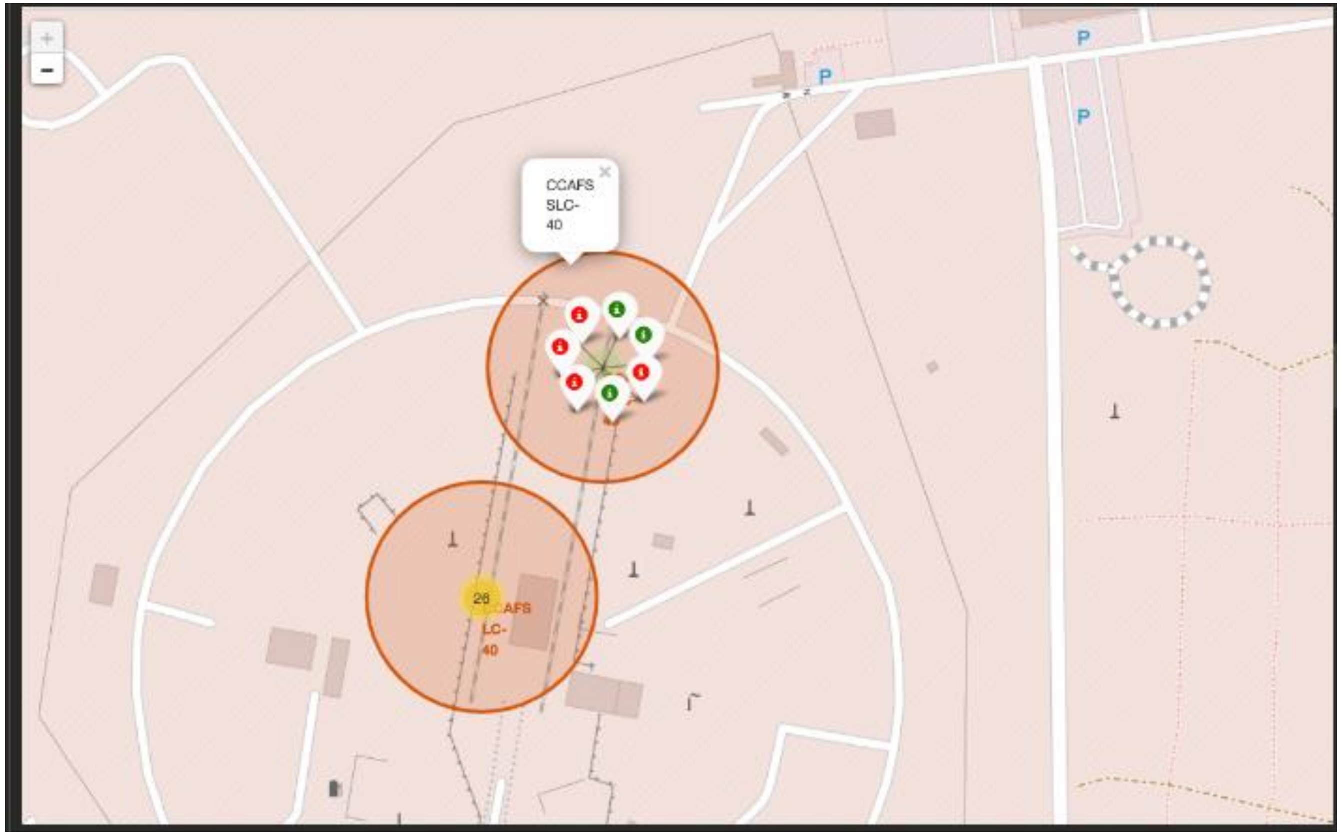
- All launch sites on map



RESUTLS

- 3 Folium

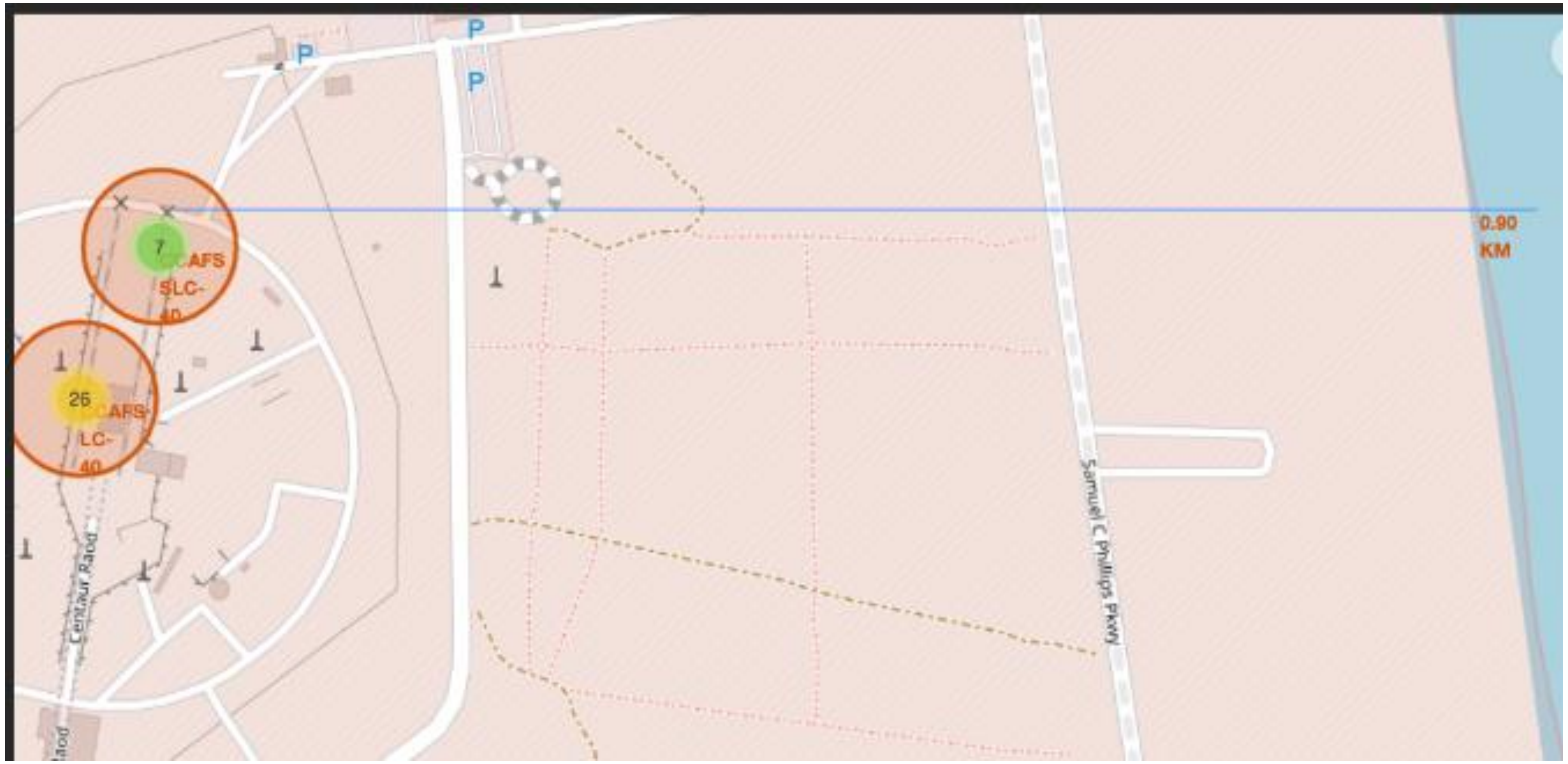
- The succeeded launches and failed launches for each site on map
  - If we zoom in one of the launch site, we can see green and red tags . Each tag represents a successful launch while each red tag represents a failed launch



RESUTLS

- 3 Folium

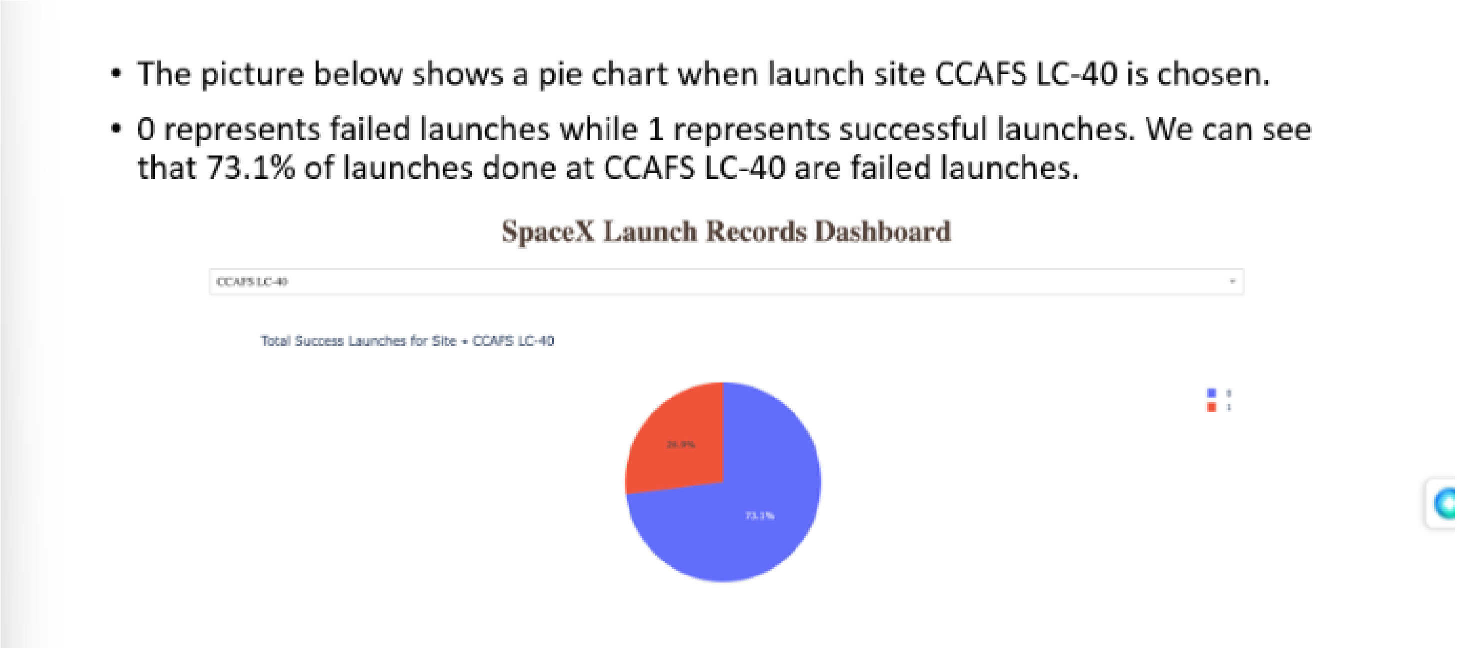
- The distance between a launch site to its proximities such as the nearest city, railway, or highway
  - The picture below shows the distance between the VAFB SLC-4E launch site and the nearest coastline



RESUTLS

- 4 Dash

- The picture below shows a pie chart when launch site CCAFS LC-40 is chosen.
- 0 represents failed launches while 1 represents successful launches . We can see that 73.1% of launches done at CCAFS LC-40 are failed launches.



- The picture below shows a pie chart when launch site CCAFS LC-40 is chosen.
- 0 represents failed launches while 1 represents successful launches. We can see that 73.1% of launches done at CCAFS LC-40 are failed launches.

SpaceX Launch Records Dashboard

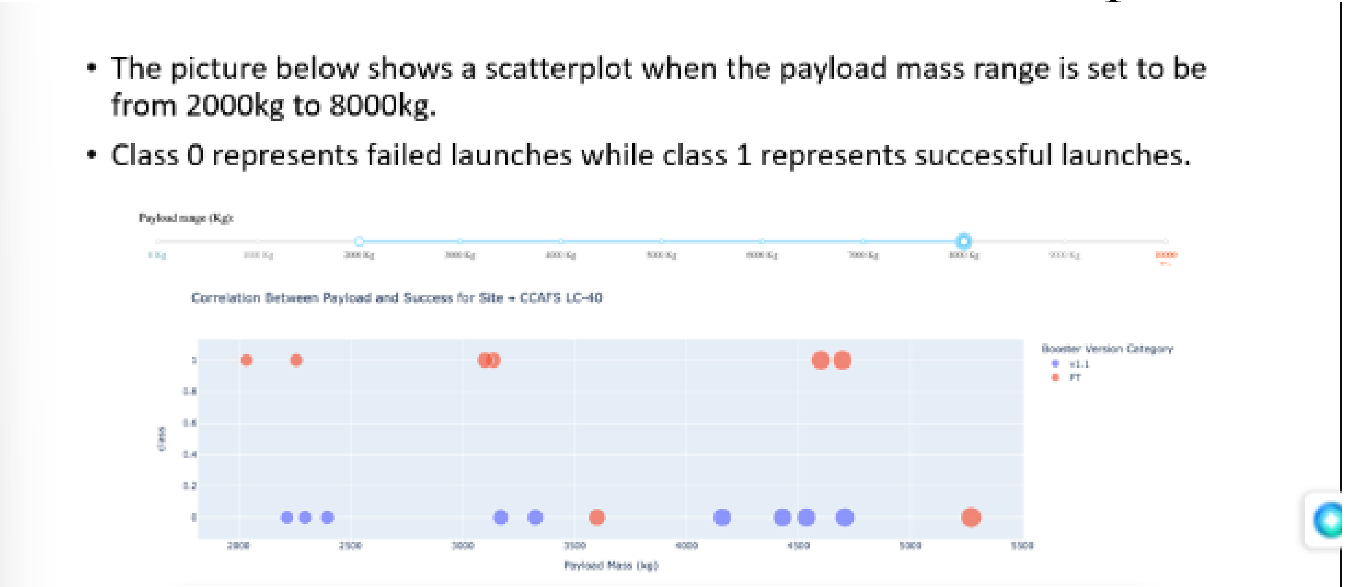
CCAFS LC-40

Total Success Launches for Site = CCAFS LC-40

RESUTLS

- 4 Dash

- The image below displays a scatterplot with the payload mass range set between 2000kg and 8000kg.
- Message received. The picture below shows a scatterplot where the payload mass range is set from 2000kg to 8000kg.
- Class 0 represents failed launches while class 1 represents successful launches .

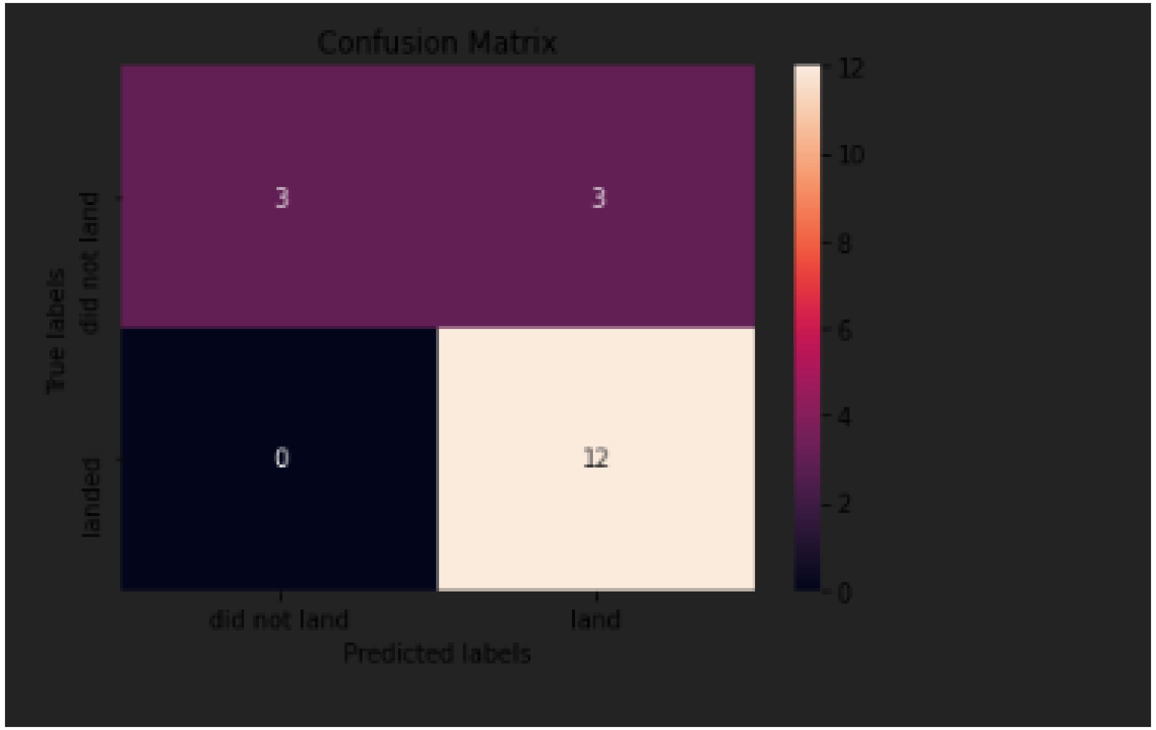


- The picture below shows a scatterplot when the payload mass range is set to be from 2000kg to 8000kg.
- Class 0 represents failed launches while class 1 represents successful launches.

RESUTLS

- 5 Predictive Analysis

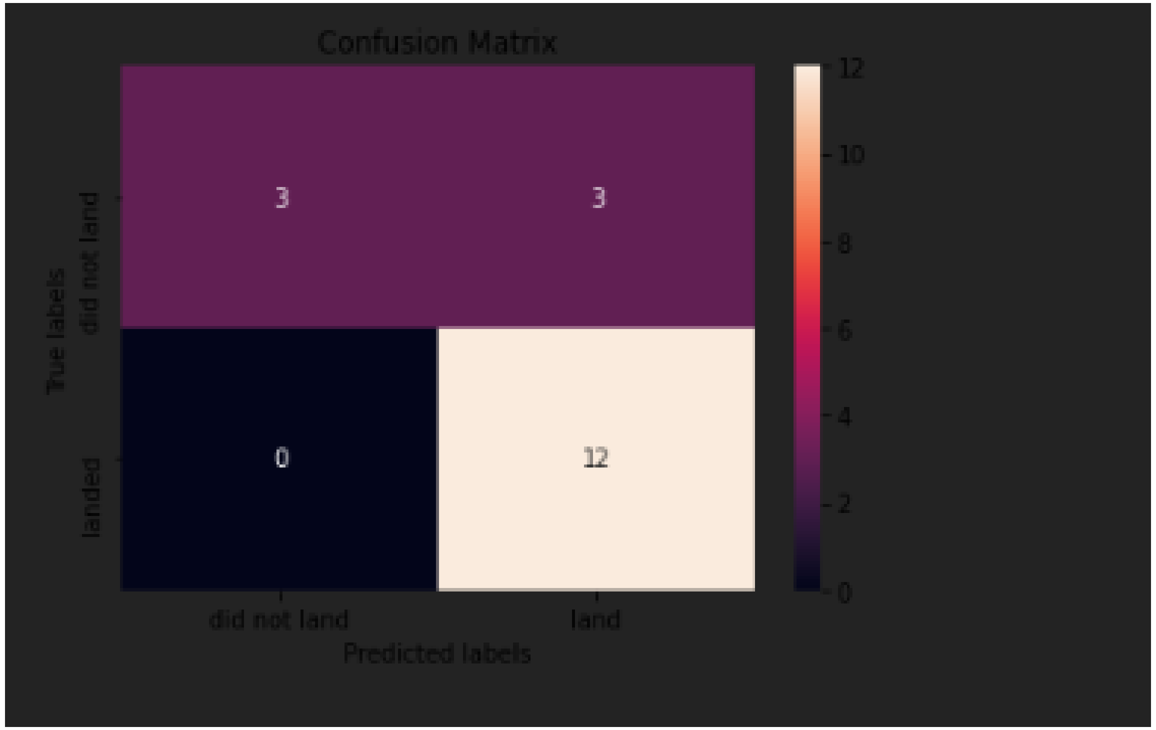
- Logistic regression
  - GridSearchCV best score: 0.8464285714285713
  - Accuracy score on test set: 0.833333333333333334
  - Confusion matrix:



RESUTLS

- 5 Predictive Analysis

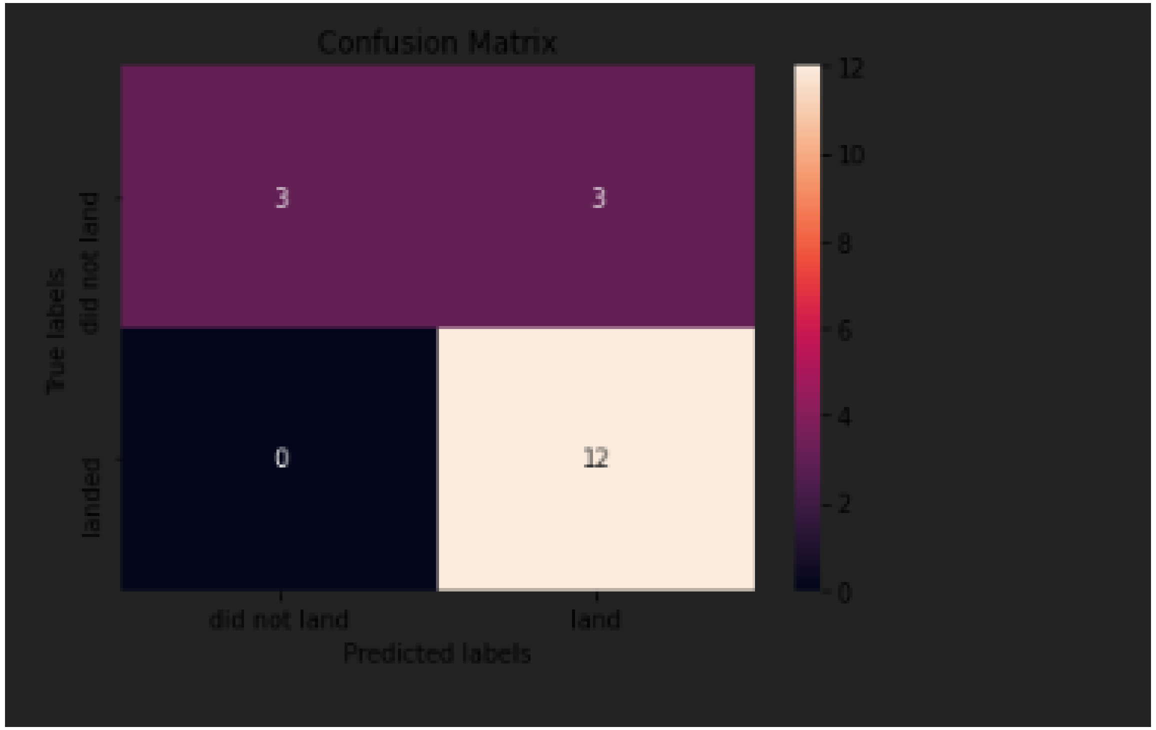
- Support vector machine (SVM)
  - GridSearchCV best score: : 0.8482142857142856
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:



RESUTLS

- 5 Predictive Analysis

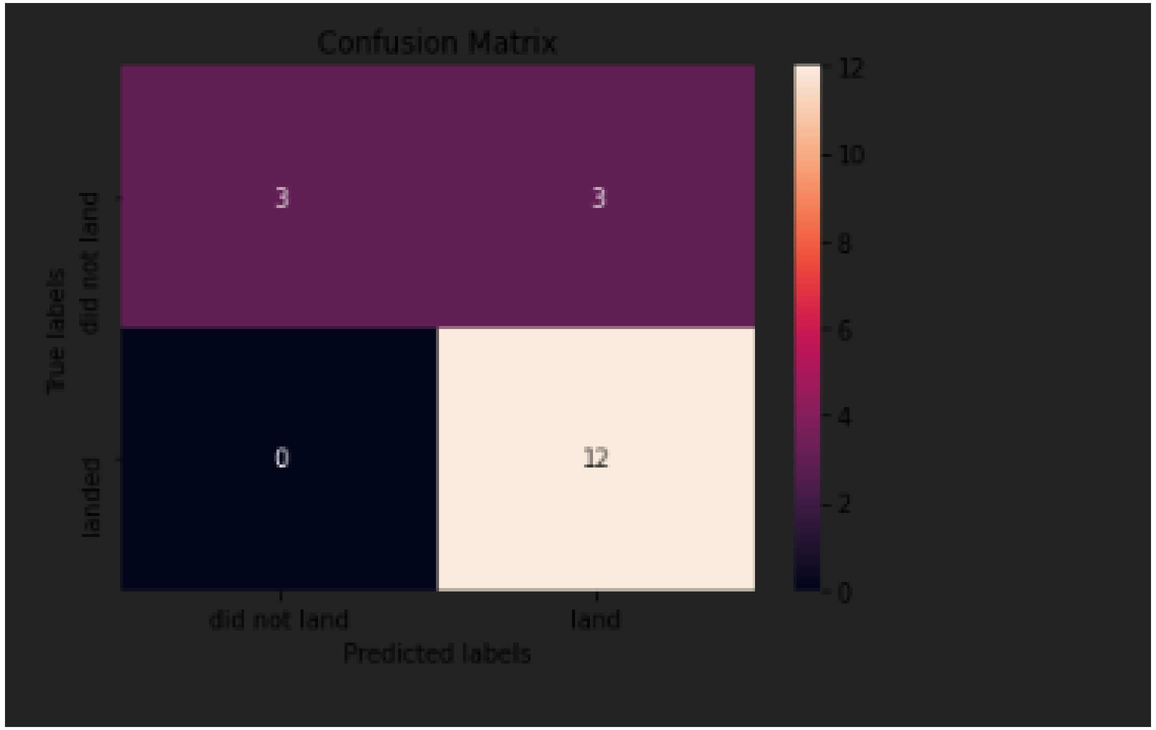
- Decision tree
  - GridSearchCV best score: : 0.8892857142857142
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:



RESUTLS

- 5 Predictive Analysis

- K nearest neighbours (KNN)
  - GridSearchCV best score: : 0.8482142857142858
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:



RESULTS

5 *Predictive Analysis*

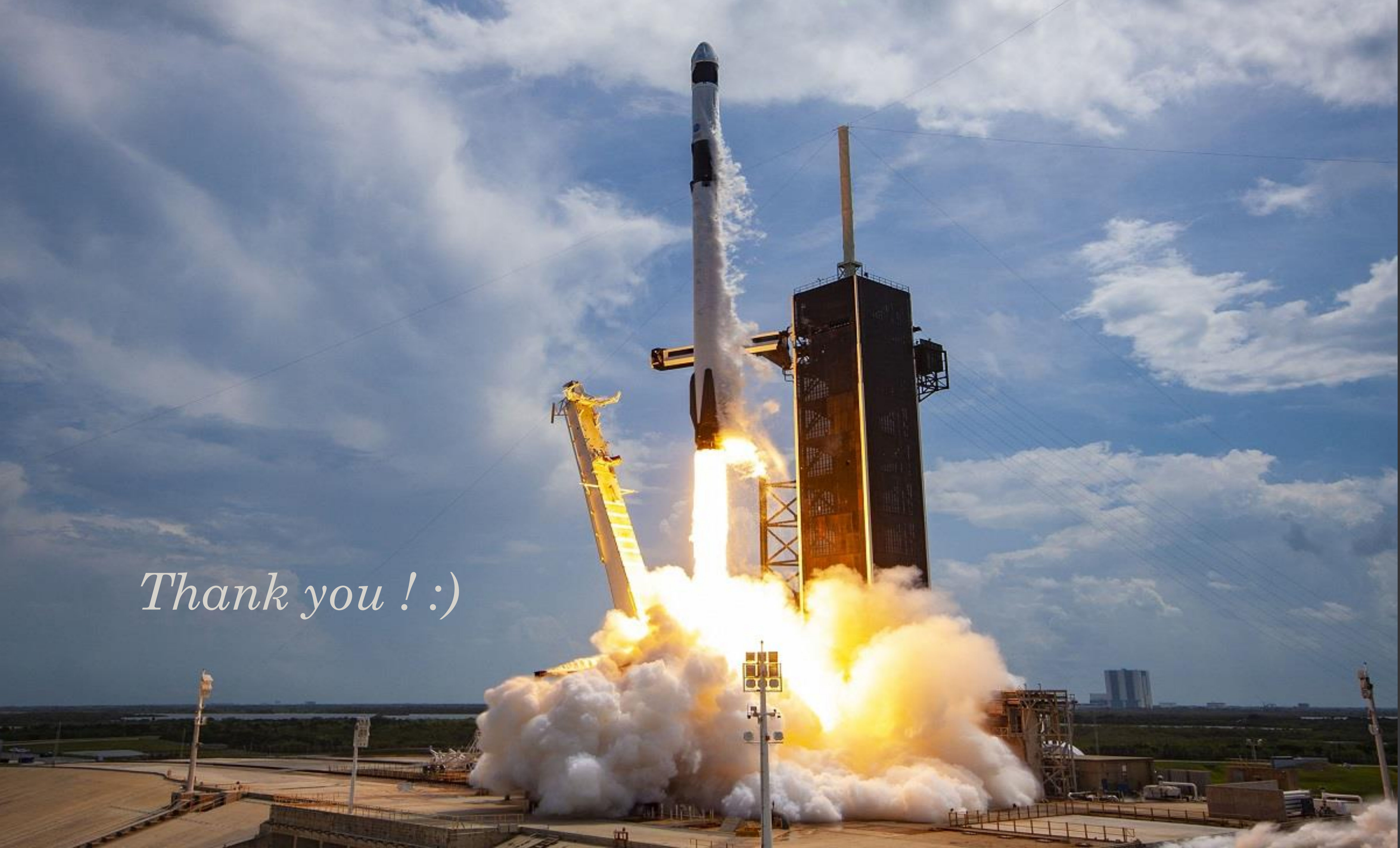
- When we compare the results of all four models, we can see that they all have the same accuracy score and confusion matrix when evaluated on the test set.
- As a result, we use their GridSearchCV best scores to rank them. Based on these scores, the models are ranked in the following order, with the first being the best and the last being the worst:
  1. Decision tree (with a GridSearchCV best score of 0.8892857142857142).
  2. K nearest neighbours, KNN (GridSearchCV best score: 0.8482142857142858)
  3. Support vector machine, SVM(GridSearchCV best score : 0.8482142857142856)
  4. Logistic regression (GridSearchCV best score : 0.8464285714285713)

DISCUSSION

- From the data visualization section, it is apparent that certain features may be correlated with the mission outcome in various ways. For instance, for heavy payloads, the success rate of landings or positive landing rate is higher for orbit types such as Polar, LEO, and ISS. However, for GTO orbit type, this distinction is not as clear since both positive and negative landing rates (unsuccessful missions) are present.
- Each feature may have a specific impact on the final outcome of the mission. However, it is difficult to determine exactly how each of these features affects the mission outcome. Nonetheless, we can use some of these features to predict whether a mission will be successful or not.

CONCLUSION

- In this project, our goal is to forecast whether the first stage of a given Falcon 9 launch will land successfully in order to estimate the cost of the launch.
- Each characteristic of a Falcon 9 launch, such as its payload mass or orbit type, may have a specific impact on the success of the mission.
- We use a variety of machine learning techniques to analyze patterns in historical Falcon 9 launch data and create predictive models that can forecast the outcome of a Falcon 9 launch.
- Of the four machine learning algorithms we used, the decision tree algorithm produced the most accurate predictive model. .



Thank you ! :)