Will the first stage of a Falcon 9 rocket launch successfully land?

Mohammed Abu Jaafar

Date: 2/13/2023

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

We will predict whether the SpaceX Falcon 9 first stage will successfully land in this capstone project.

We can calculate the cost of a launch if we know if the first stage will land. This will be accomplished through the application of various machine learning classification algorithms.

Data Collection, Data Wrangling and Preprocessing, Exploratory Data Analysis, Data Visualization, and finally Machine Learning Prediction will be the methodology used

.During our investigation, the results of our analysis show that there are some characteristics of rocket launches that are related to the failure or success of the launch.

Finally, we conclude that the Decision Tree is possibly the best machine learning algorithm for this problem.

INTRODUCTION

The primary goal of this capstone project is to forecast whether the Falcon 9 first stage will successfully land. SpaceX is so proud of its ability to reuse the first stage of a rocket launch that they advertise on their website that their rocket launches cost 62 million while other providers charge up to 165 million. The reusability of the first stage accounts for a large portion of these savings. We can calculate the cost of a launch if we know if the first stage will land. This data can be used if another company wants to compete with SpaceX for a rocket launch.

This takes us to our major question, which is: Will the first stage of a Falcon 9 rocket launch successfully land?

METHODOLOGY

Data was collected in two ways: by requesting information from the SpaceX API and by web scraping launch data from a Wikipedia page. The data was then transformed and cleaned using Python’s Pandas library.

Exploratory data analysis (EDA) was performed on the clean data using visualization tools like Python's matplotlib and seaborn libraries, as well as SQL queries to answer questions. Python's interactive visualization packages were used to answer some analytical questions.

Plotly Dash was used to create interactive data visualizations and Folium was used to create maps.

For the predictive analysis, four different machine learning classification models were used. Logistic regression, support vector machines, k-nearest neighbor, and decision tree classifier were the models used. To find the best model, each one was trained, tuned, and evaluated.

Completed labs : Data wrangling & Data Collection

• Data Collection :

- Request and parse the SpaceX launch data using the GET request
- Normalize JSON response into a dataFame
- Extract only useful columns using auxiliary functions
- Create new Pandas data Frame from dictionary
- Filter data Frame to only include Falcon 9 launches
- Handle missing values
- Export to CSV file

• Data Wrangling :

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column using one-hot encoding
- Export to CSV

EDA with Data Visualizations

Visualize the relationship between :	**Flight Number and Launch Site-** scatter plot	**Visualize the relationship between :**	**Flight Number and Orbit type -**scatter plot
	Payload and Launch Site- scatter plot		**Payload and Orbit type-** scatter plot
	success rate of each orbit type -bar plot		**launch success yearly trend –** line plot
	Click here		

EDA with SQL

The following are some of the SQL queries operated on the dataset:

Displaying

- the names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1

Listing

- the date when the first successful landing outcome in ground pad was achieved.
- 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.

EDA with SQL(Continued)

The following are some of the SQL queries operated on the dataset:

Ranking

- the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Listing

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

- Click here

Launch Sites Locations Analysis with Folium

- Marked all launch sites on a map
- Marked the success/failed launches for each site on the map
- Calculated the distances between a launch site to its proximities

- Click here

By combining these objects, the following geographical trends about launch locations are discovered:

- Yes, launch sites close proximity to **railways**
- Yes, launch sites in close proximity to **highways**
- Yes, launch sites in close proximity to **coastline**
- Yes, launch sites keep certain distance away from **cities**

Build a Dashboard with Plotly Dash

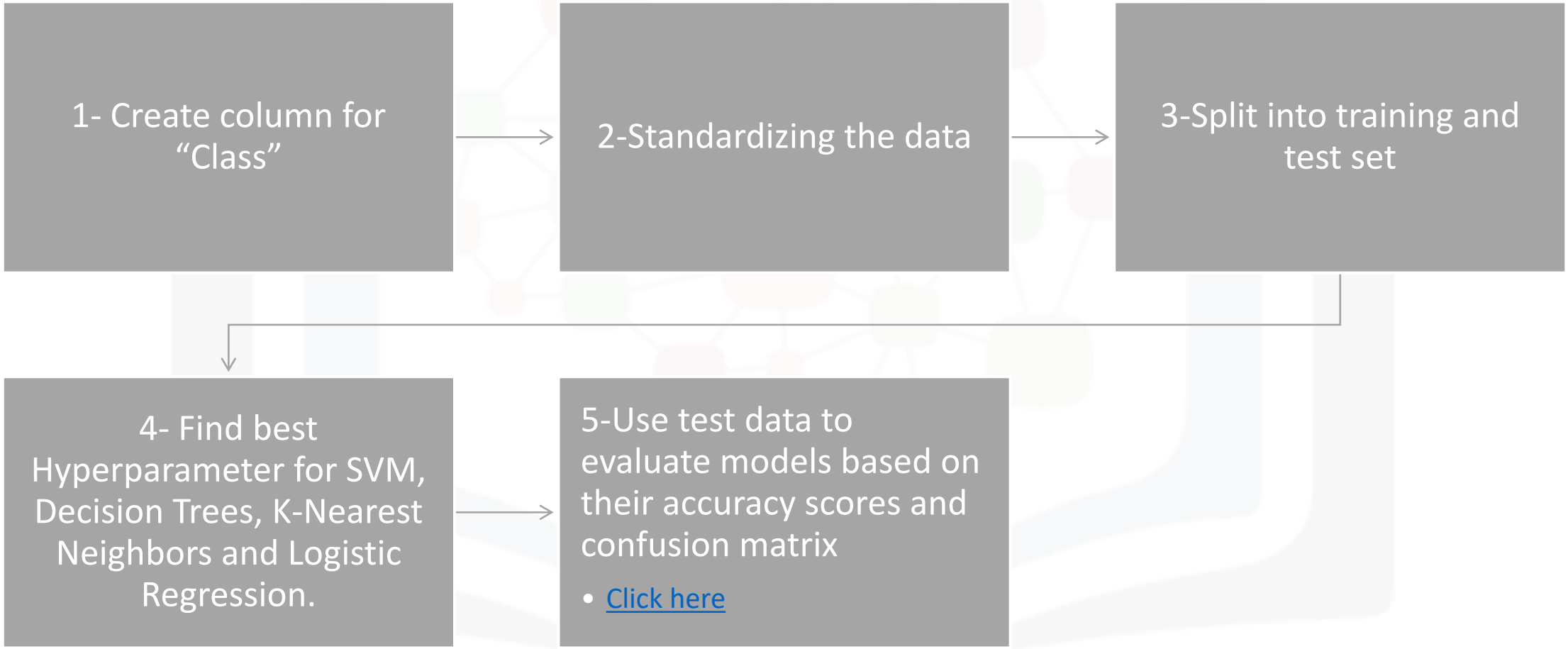
The dashboard will shows us :

1- a pie chart displaying each site's successful launch. This graph is helpful since it allows you to see the success rate of launches on specific sites or view the distribution of landing results across all launch locations.

2- a scatter diagram showing the link between landing success and the mass of various boosters.

Click here

Predictive Analysis Methodology

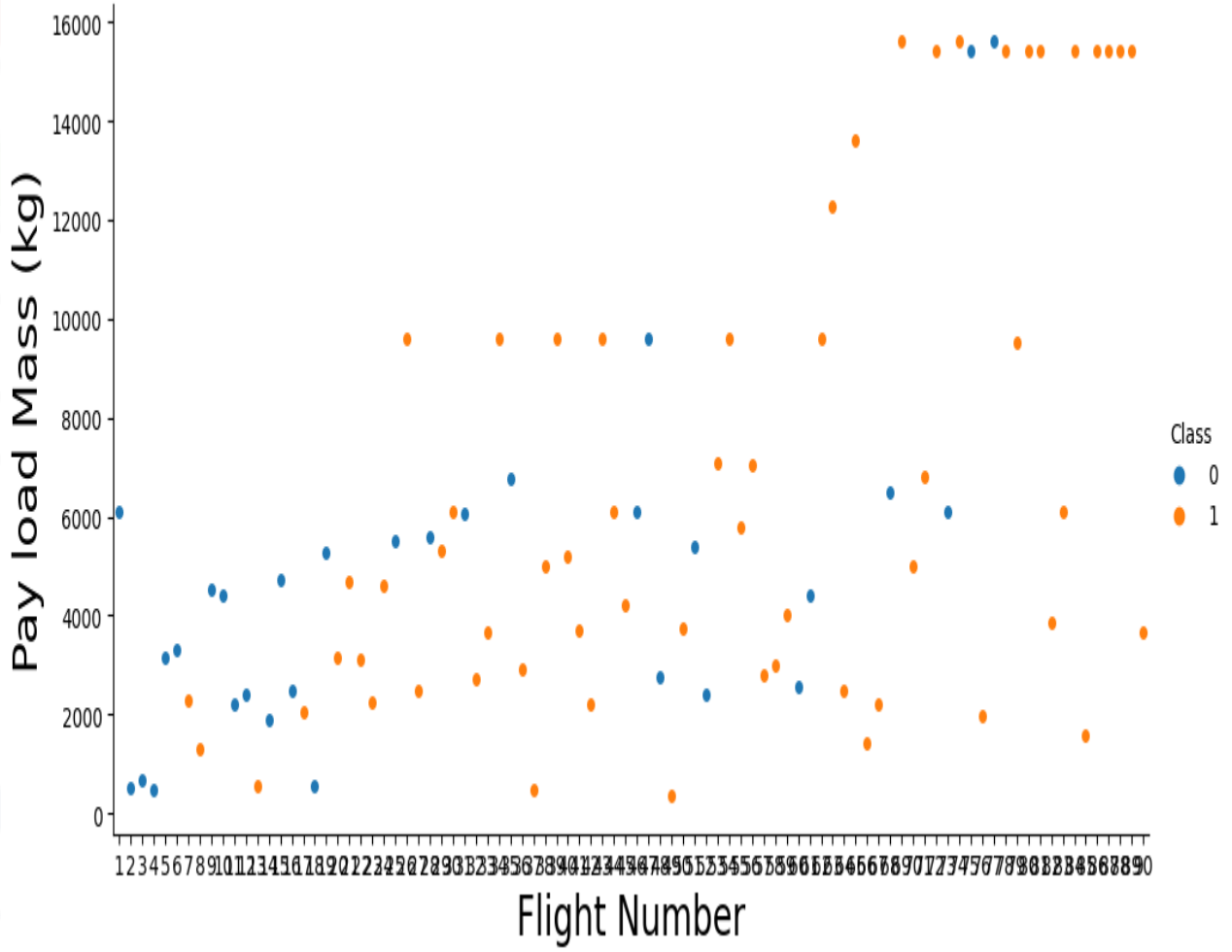


RESULTS

Flight Number vs Pay load Mass(Kg)

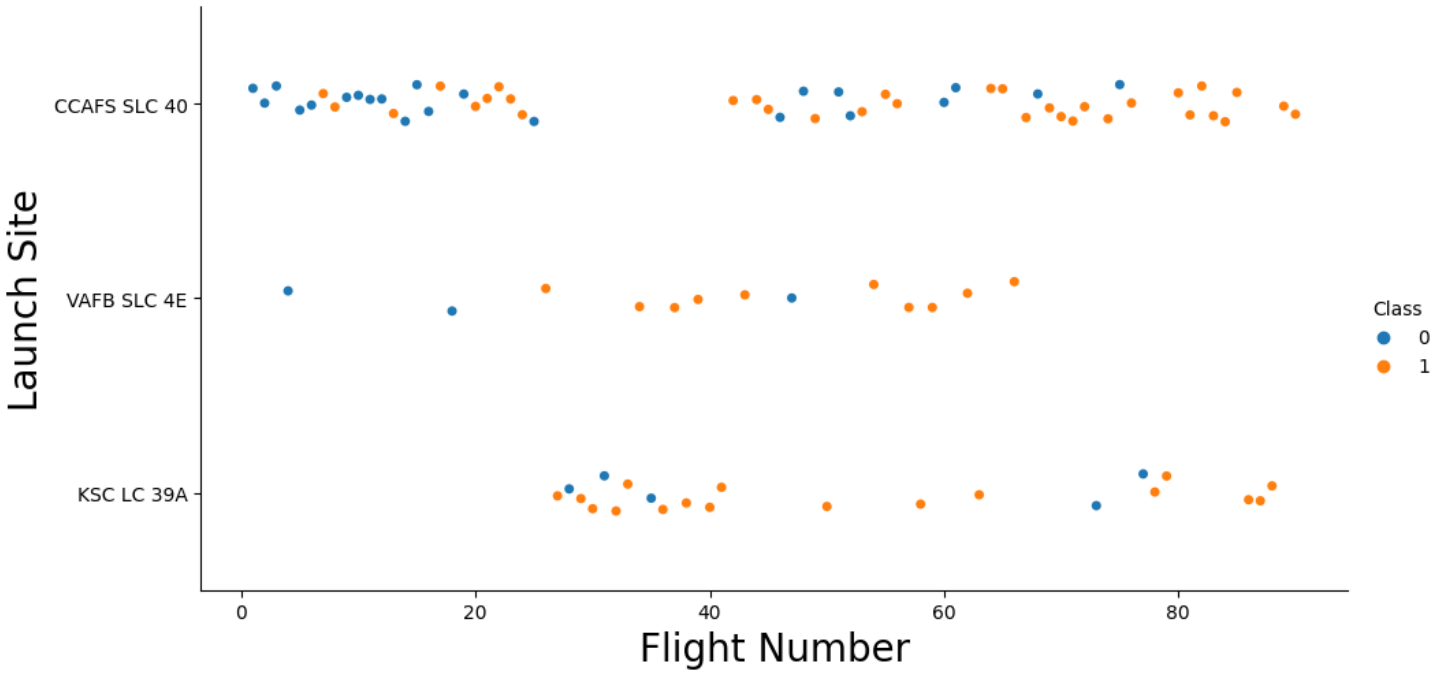
flight number increases, the first stage is more likely to land successfully.

more massive the payload, the less likely the first stage will return



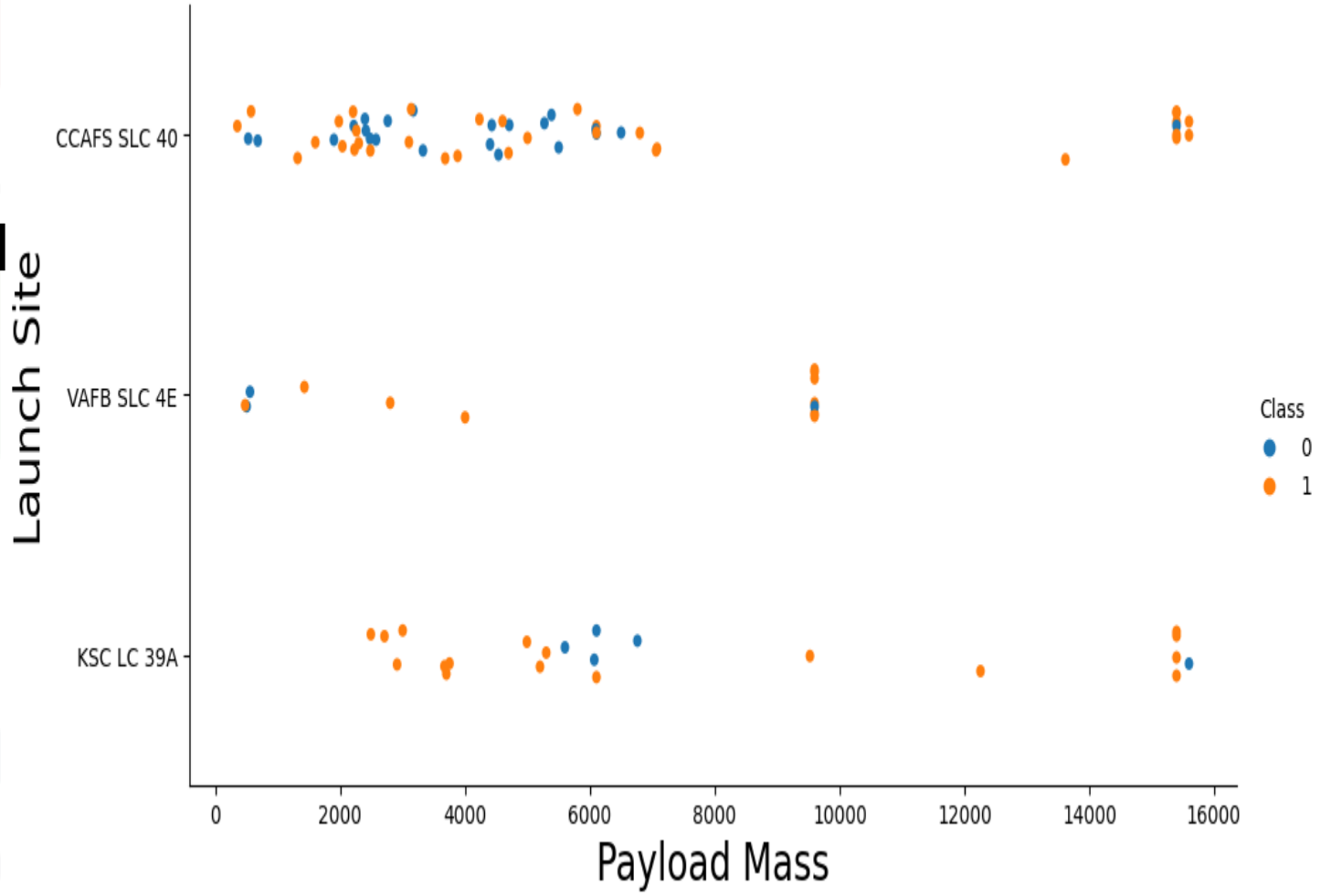
relationship between Flight Number and Launch Site

- different launch sites have different success rates.
- **CCAFS LC-40**, has a success rate of 60 %, while **KSC LC-39A** and **VAFB SLC 4E** has a success rate of 77%.



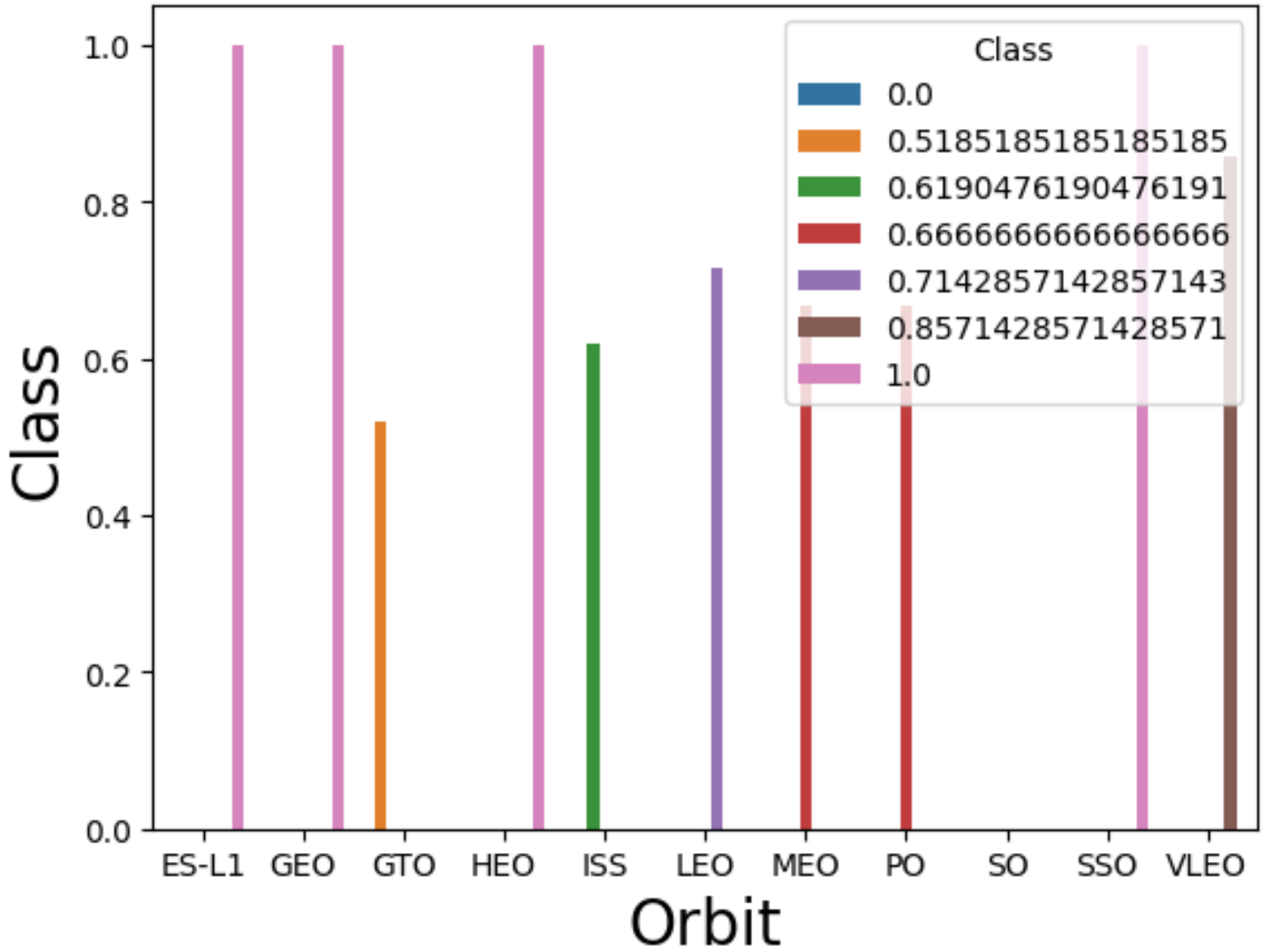
relationship between Payload and Launch Site

- the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



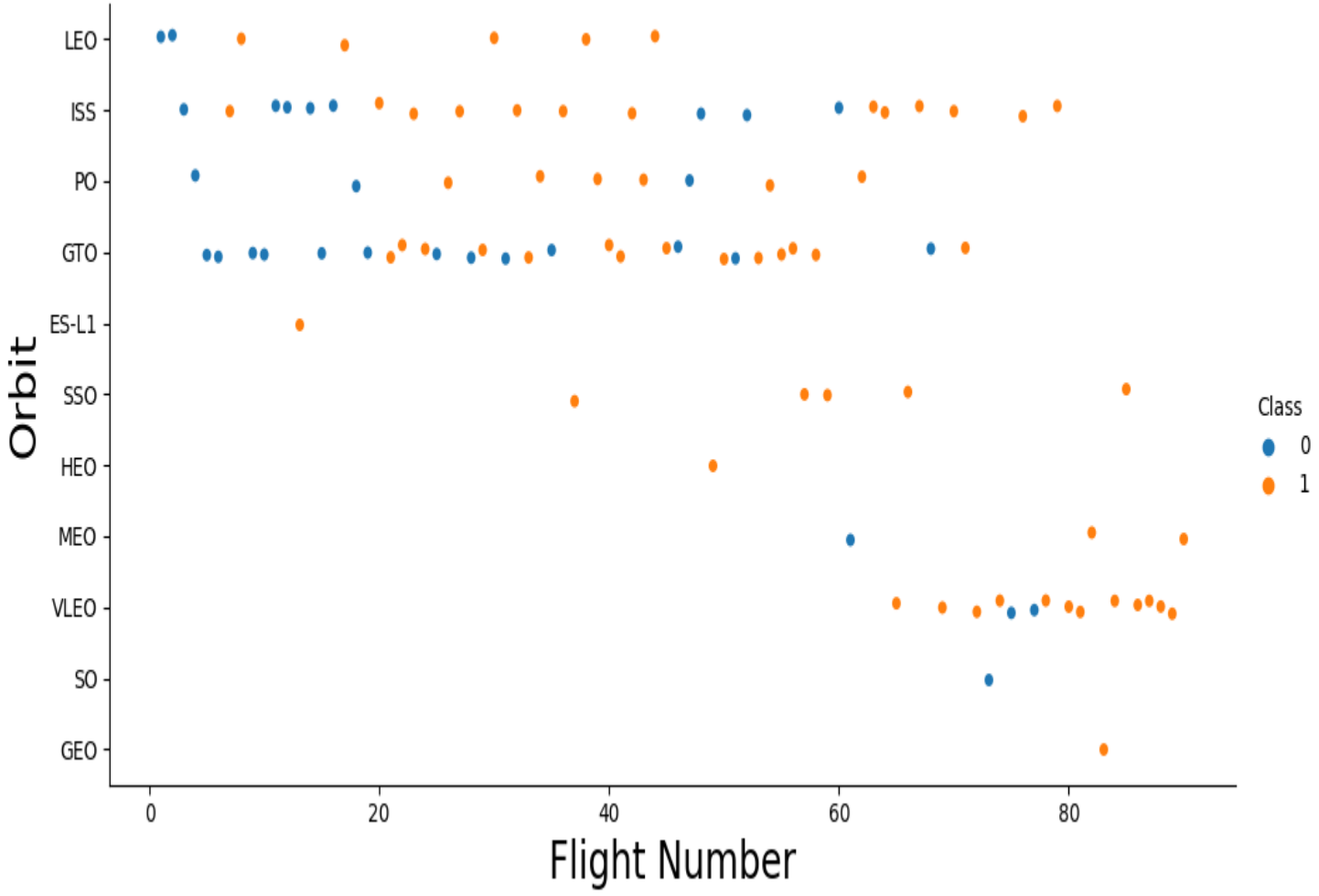
relationship between success rate of each orbit type

- Successful Orbits are ES-L1,HEO, SSO,GEO

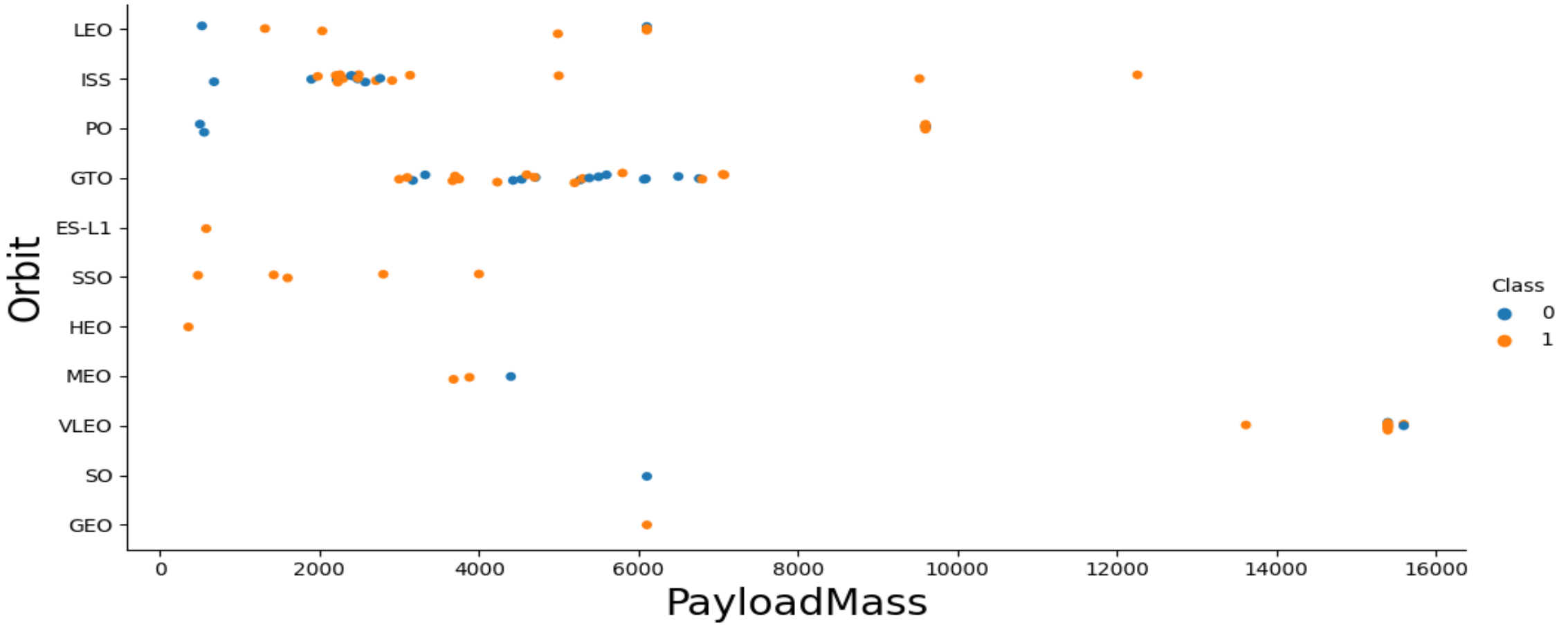


relationship between Flight Number and Orbit type

- LEO orbit the Success appears related to the number of flights
- seems to be no relationship between flight number when in GTO orbit.



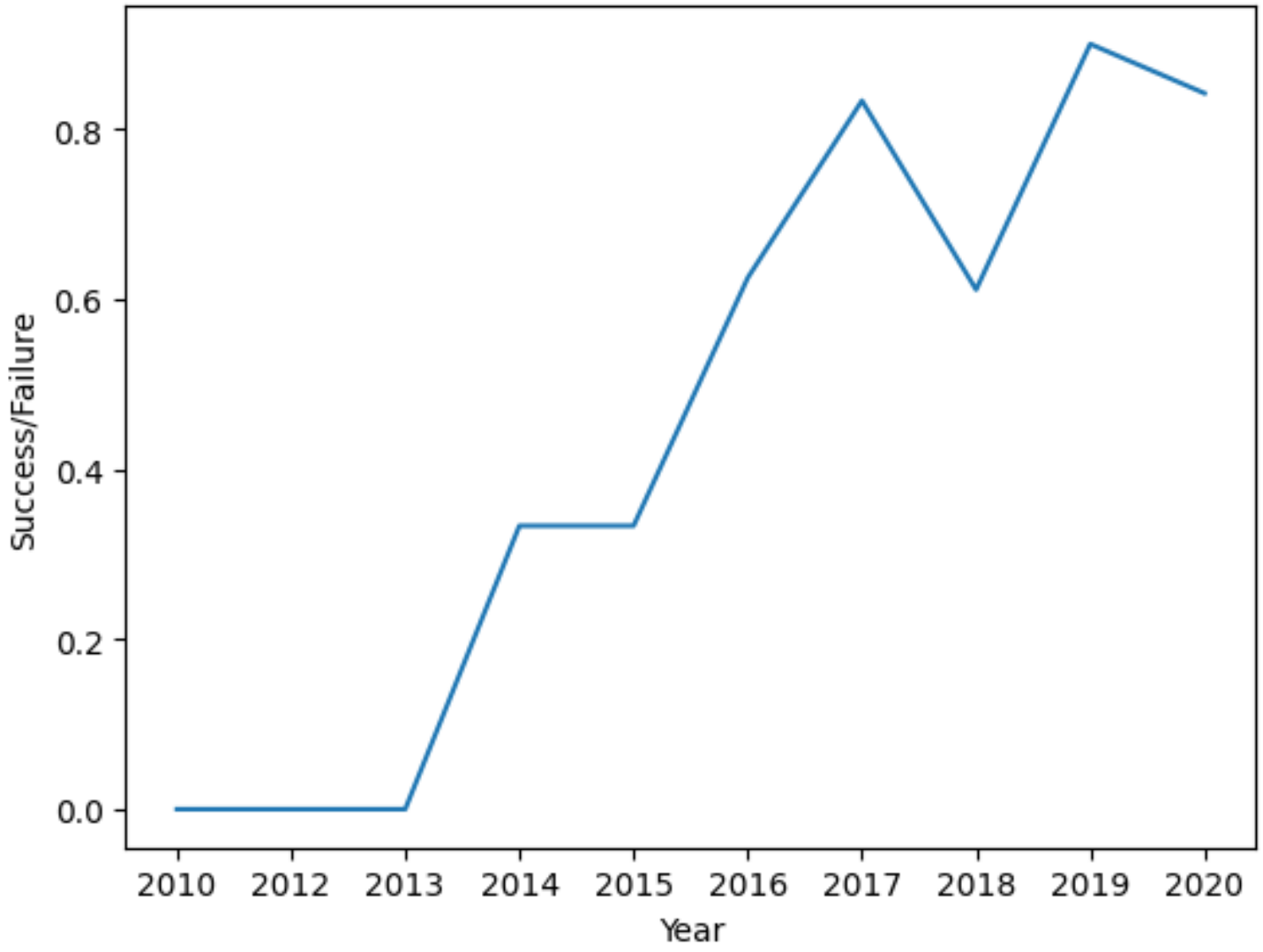
relationship between Payload and Orbit type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS

launch success yearly trend

- success rate since 2013 kept increasing till 2020



Results – EDA with SQL : displaying

- 1-unique launch sites in the space mission

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

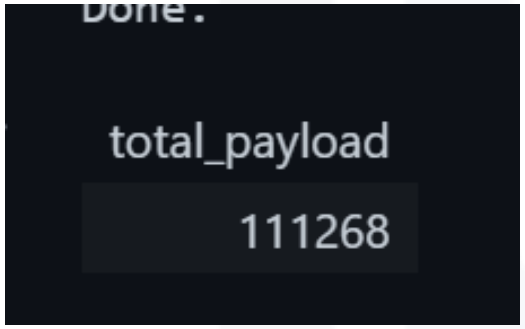
Results – EDA with SQL : displaying

2- 5 records where launch sites begin with the string 'CCA'

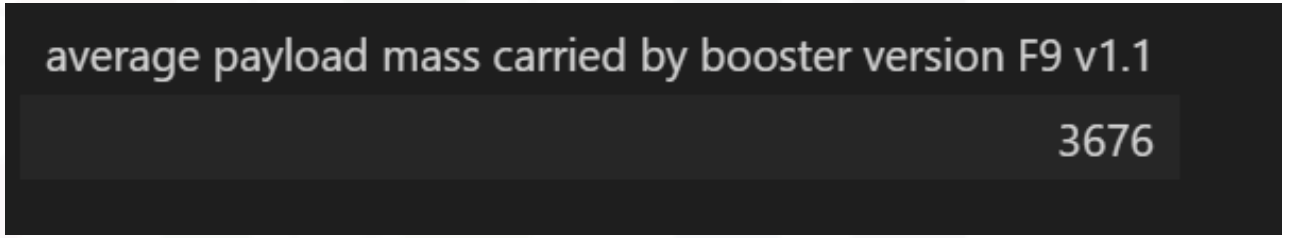
DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight 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 – EDA with SQL : displaying

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



4- average payload mass carried by booster version F9 v1.1



Results – EDA with SQL : listing

5-date when the first successful landing outcome in ground pad was achieved

first_success_gp
2015-12-22

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

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

Results – EDA with SQL : listing

7- total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

8- names of the booster_versions 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 – EDA with SQL : listing

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

1	mission_outcome	booster_version	launch_site
10	Success	F9 v1.1 B1012	CCAFS LC-40
11	Success	F9 v1.1 B1013	CCAFS LC-40
2	Success	F9 v1.1 B1014	CCAFS LC-40

Results – EDA with SQL : Ranking

landing__outcome	qty
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

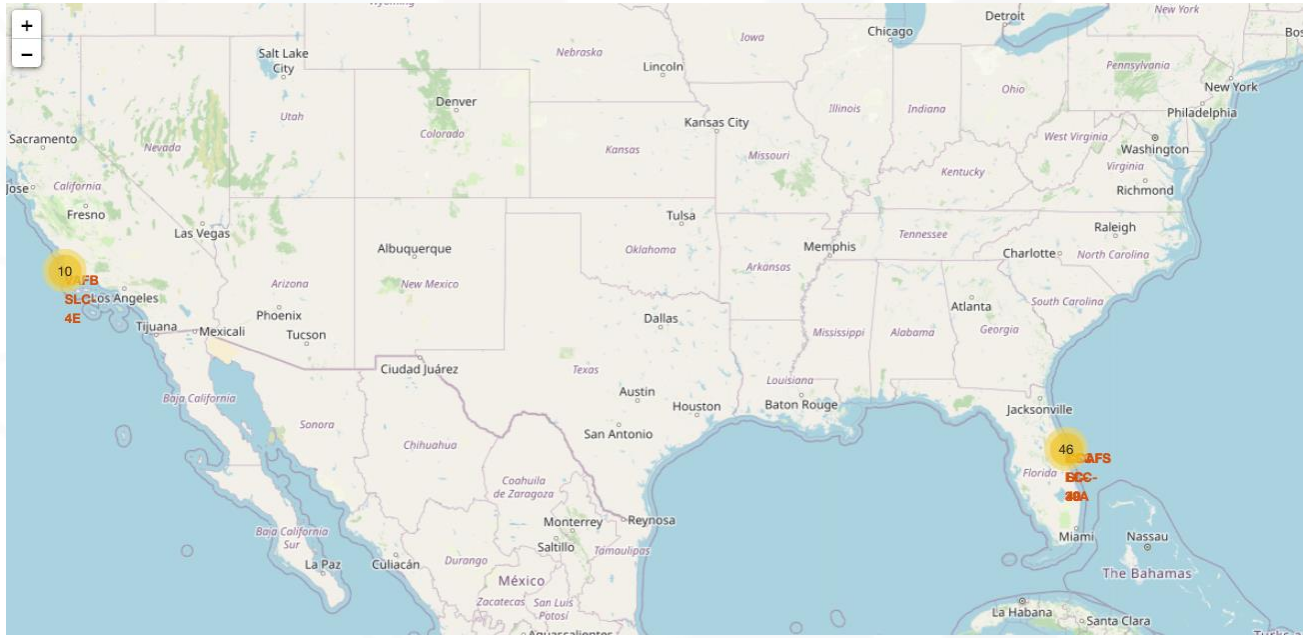
- 10- count of landing outcomes (such as Failure (drone ship) or Success

- between the date 2010-06-04 and 2017-03-20, in descending order

SpaceX Launch Sites Locations

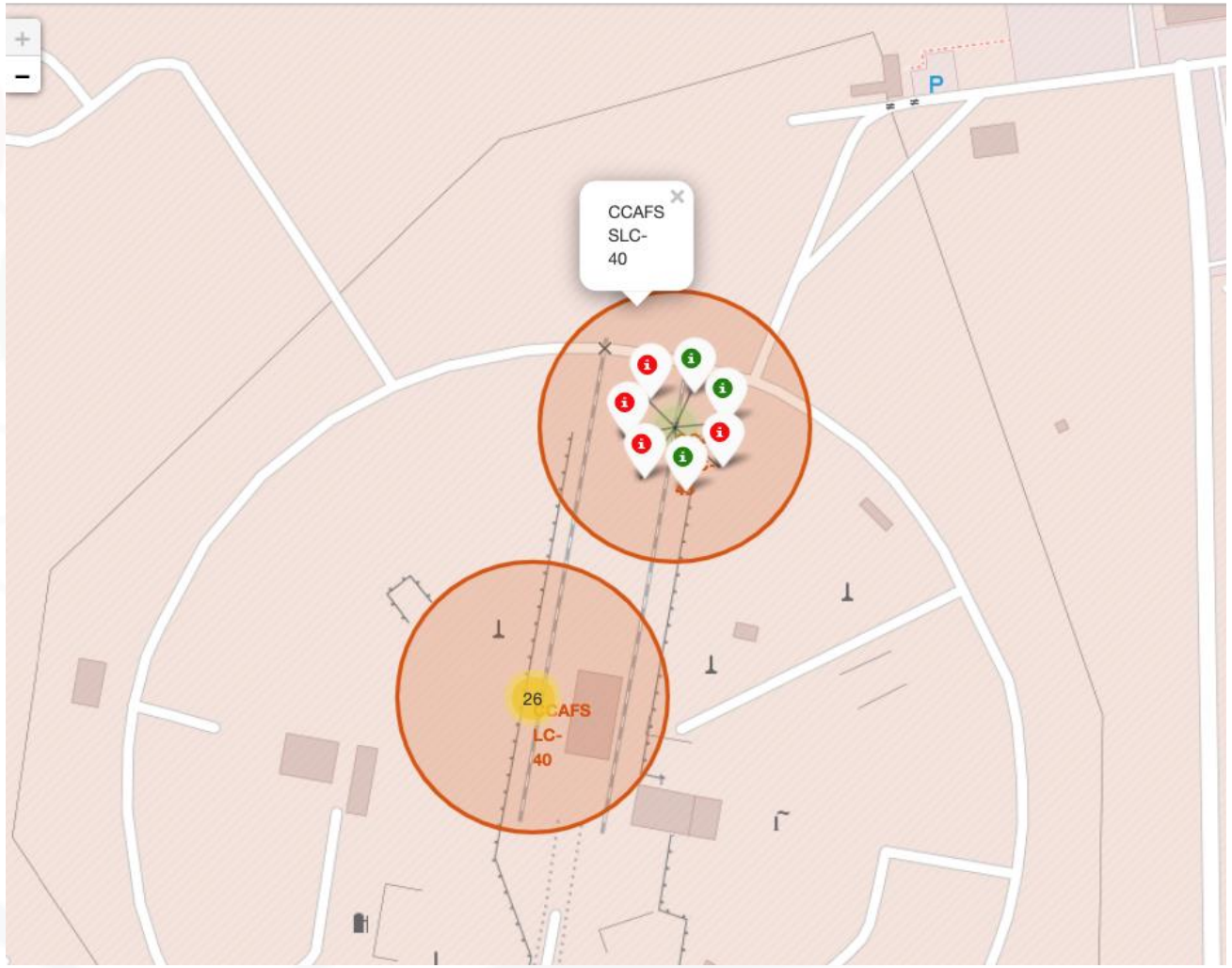
The locations of all the SpaceX launch sites in the US are shown by the yellow markers.

The launch locations have been situated strategically along the coast.



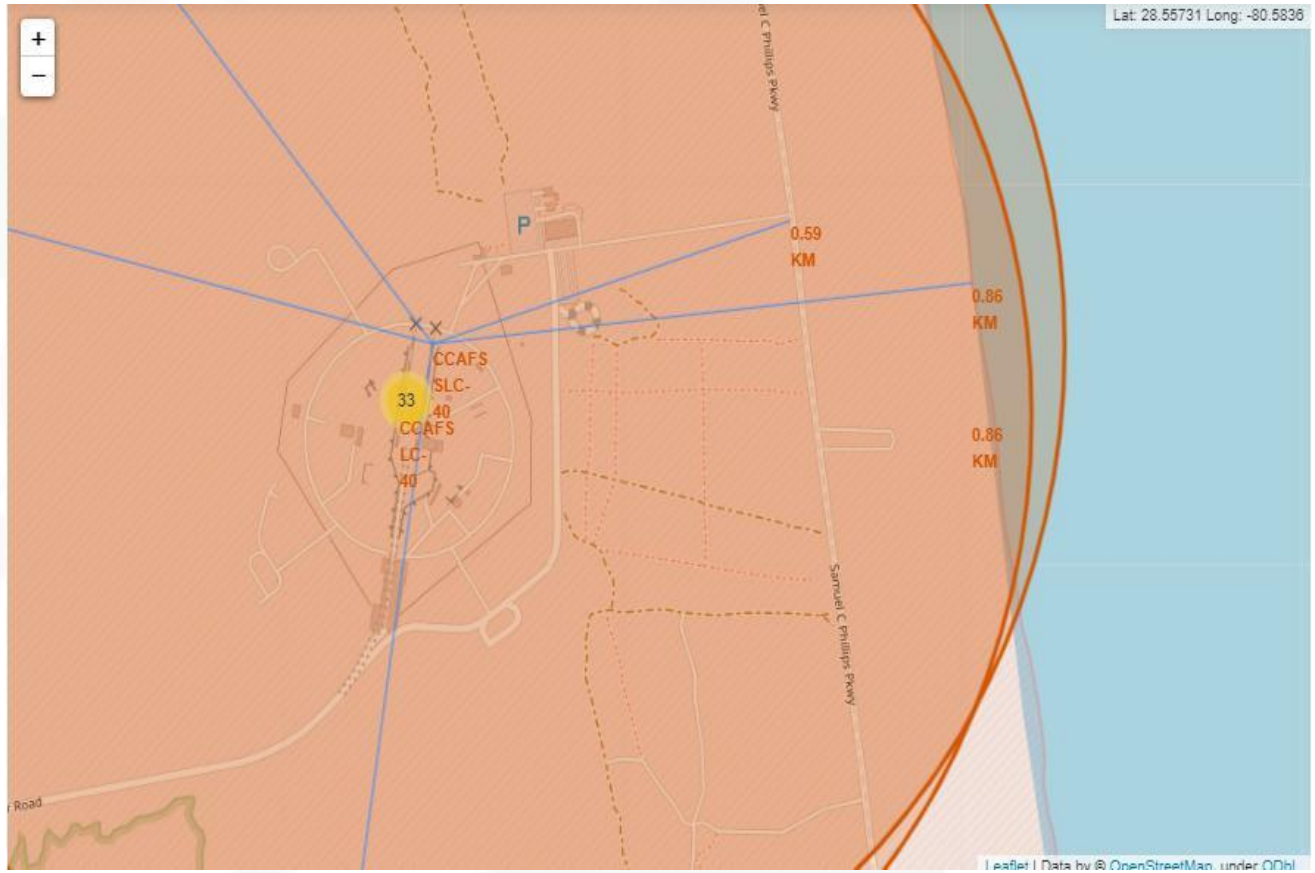
Does it succeed or fails?

The launch site show marker clusters of successful landings (green) or failed landings as we zoom in on a launch location (red).



Launch Site Distances

The produced map reveals that the chosen launch point is near to a highway for people and equipment transportation. For launch failure testing, the launch location is also near to the coast.



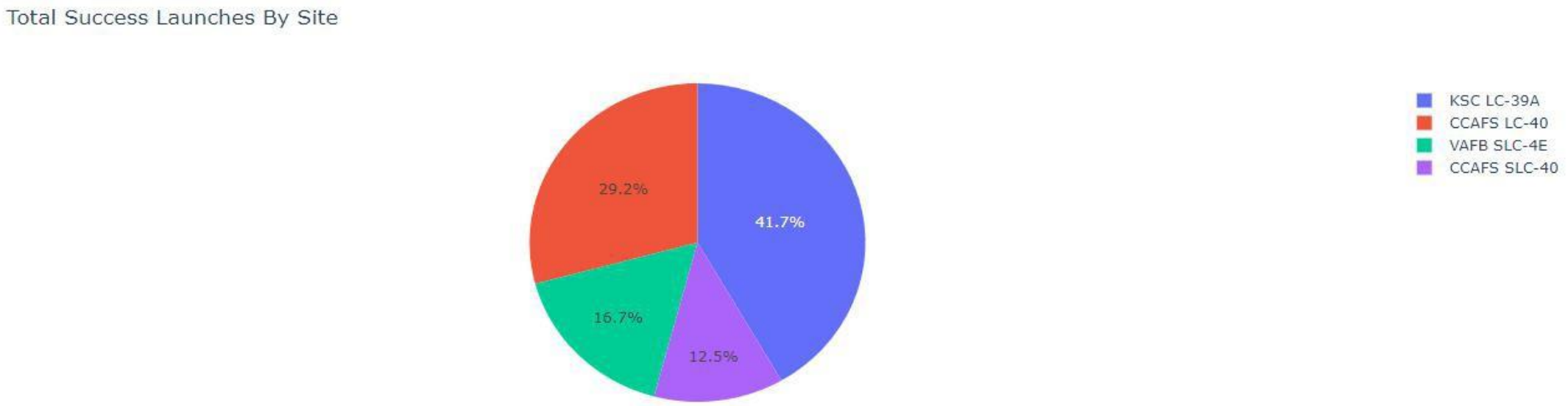
DASHBOARD

Dashboard with Plotly Dash

IBM Developer

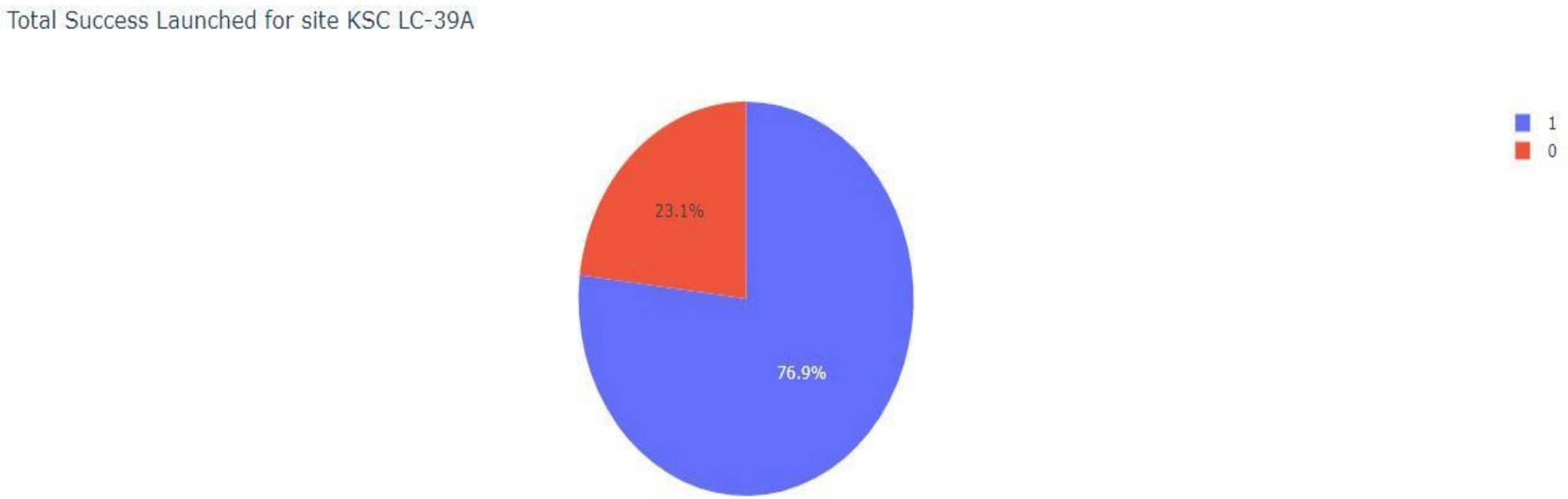
SKILLS NETWORK

Total Successful Launches By Site



The KSC LC-39A Launch site has the most successful launches with 10 in total.

Launch Site With Highest Success Ratio



The KSLC-39A has the highest success rate with 76.9%.

Payloads vs Launch Outcome

- The v1.1 booster offers the highest percentage of effectiveness across both weight ranges.
- Payloads between 0 and 2500 kg have a somewhat lower launch success percentage than payloads between 250 and 5000 kg. Actually, there isn't much of a distinction between the two.

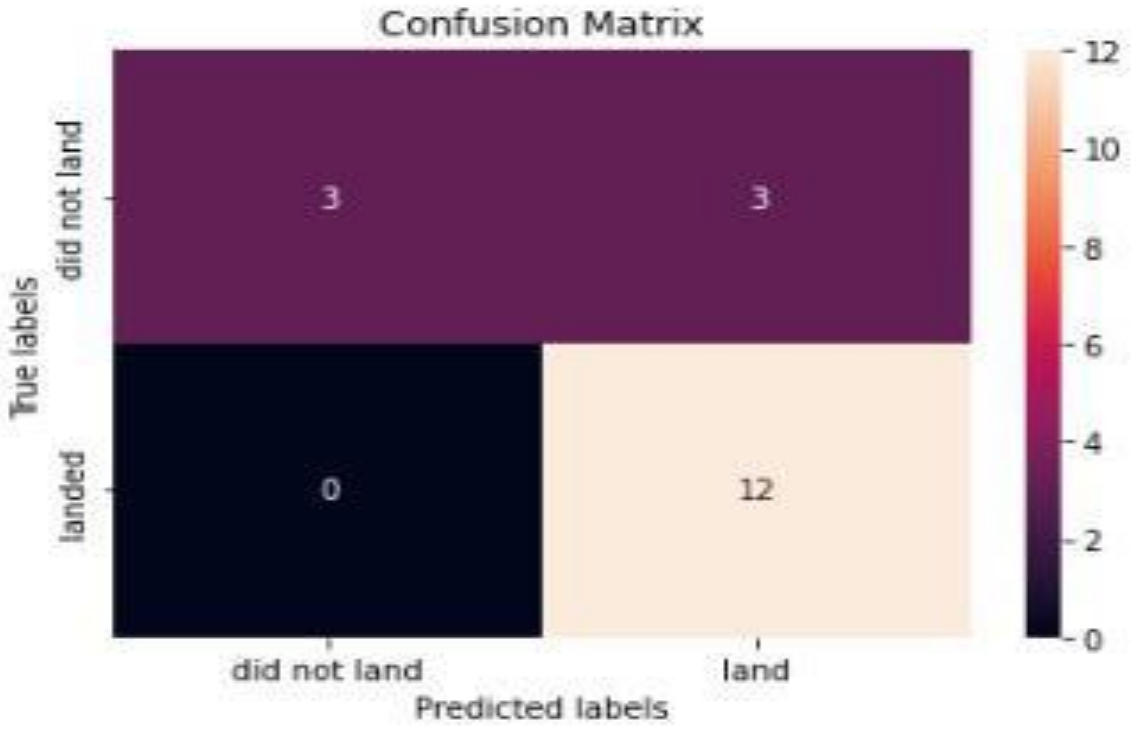
Classification Accuracy

The Decision Tree classifier had the best accuracy at 94%.

	method	accuracy
0	Logistic regression	0.833333
1	Support vector machine	0.833333
2	Decision tree classifier	0.944444
3	K nearest neighbors	0.833333

Confusion Matrix

- When the True label was success (True Positive), the model predicted 12 successful landings, and when it was failure, it predicted 3 unsuccessful landings (True Negative).
- When the True label was an unsuccessful landing, the algorithm also predicted three successful landings (False Positive).
- Successful landings were frequently anticipated by the model.



Conclusions

- The study revealed that as the success rate has increased over time, there is a positive link between the number of flights and success rate.
- The most successful launches occurred in orbits like SSO, HEO, GEO, and ES-L1.
- Payload mass can be related to success rate since lighter payloads have often had better results than bigger payloads.
- The launch sites are strategically located near highways and railways for transportation of personel and cargo, but also far away from cities for safety.

- The Decision Tree Classifier is the most accurate prediction model for this dataset, with a 94% accuracy rate.

Appendix

For all files you can enter the GitHub Repository

GitHub Repository: https://github.com/abujafam/ibm-data-science/tree/main/Final%20project