IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Luis Felipe Beltran
July 21, 2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection through API
  - Data Collection through web scrapping
  - Data Wrangling
  - EDA with SQL
  - EDA with data visualization
  - Interactive maps with folium
  - Dashboard with plotly
  - Machine learning prediction
- Summary of all results
  - EDA result
  - Interactive analytics result
  - Predictive analytics result

Introduction

Rocket engineering and sustainable space flights have become the 21st century version of the 1970s arms race. With billions of dollars worth of investment and the entire world excited for new results, space fight has become quite a famous subject.

Unfortunately, it's viability relies on how successful flights and there so many variables in place that it seems daunting just to think about how to solve this.

Here's where data comes in.

The aim for this project is to use Data Science in order to analyze all of the variables at play in order to accurately predict success probability as well as identify the most important variables at play.

This will provide with valuable information which will prove to be quite useful for space programs in the near future.

Questions we want to answer:

- Determine which variables have the most weight determining if the rocket will successfully land
- The operating conditions that need to be in place for a successful landing
- The interaction between variables in order to create an accurate prediction model for future landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected mostly via web scrapping from several websites. Though most of the information comes directly from the flight results page from Wikipedia.
- Perform data wrangling
  - After acquiring data we went through the entire process of preparing the data such as data wrangling, exploratory data analysis and basic visualizations to understand more about the data at play.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

Data was collected via web scrapping using a get request on Wikipedia's page for Spacex past landings. After that we used that response to create a beautiful soup object in order to properly read through the text file and finally use that object to create a panda's data frame for better data handling.

Data Collection – SpaceX API

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

- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/SpaceX_API_Calls.ipynb

Now let's start requesting rocket launch data from SpaceX API with the following URL:

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
print(response.content)
```

b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/94/f2/NN6Ph45

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

```
response=requests.get(static_json_url)
```

```
response.status_code
```

```
200
```

```
data = response.json()
data = pd.json_normalize(data)
```

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub Link:
- https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcon9_web_scrapping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

```python
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

```python
response=requests.get(static_url)
```

```python
soup=BeautifulSoup(response.text, 'html.parser')
```

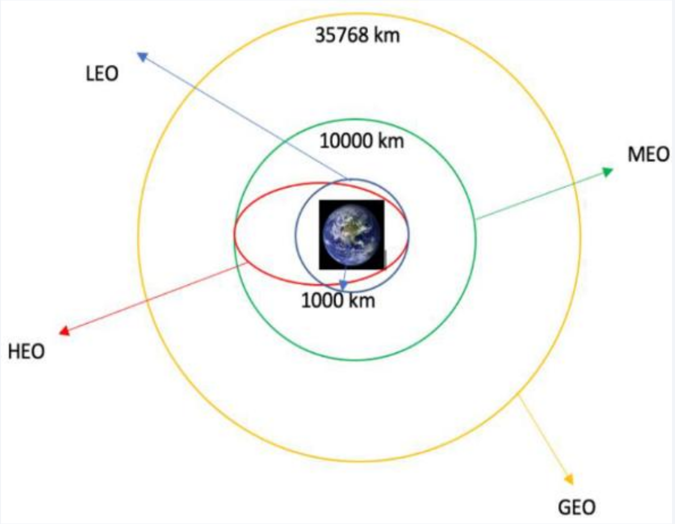
```python
if soup.title:
    print(soup.title.string)
else:
    print("Title tag not found in the HTML.")
```

```
List of Falcon 9 and Falcon Heavy launches - Wikipedia
```

```python
import requests
from bs4 import BeautifulSoup

static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0 Safari/537.36"
}

response = requests.get(static_url, headers=headers)

soup = BeautifulSoup(response.text, 'html.parser')

```

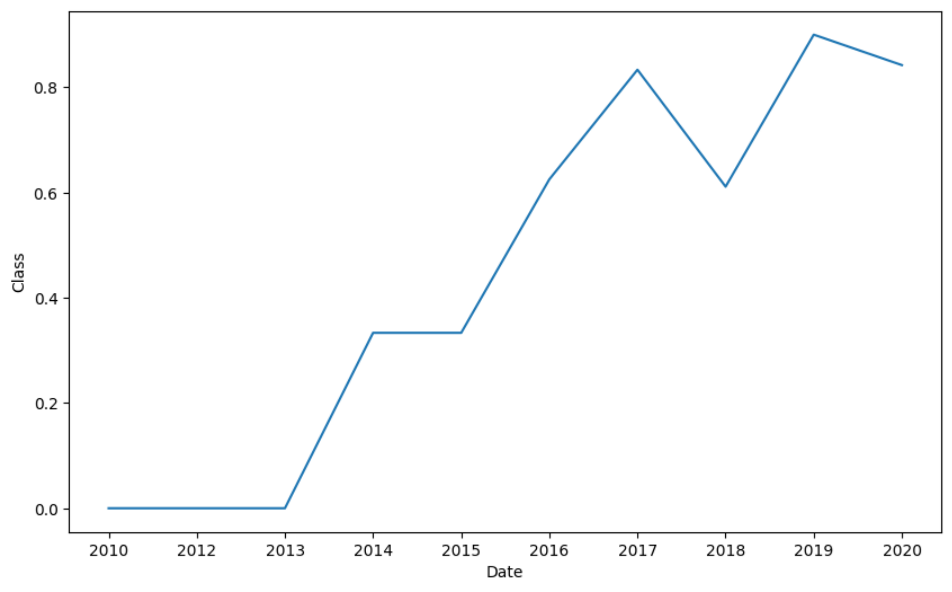
Data Wrangling

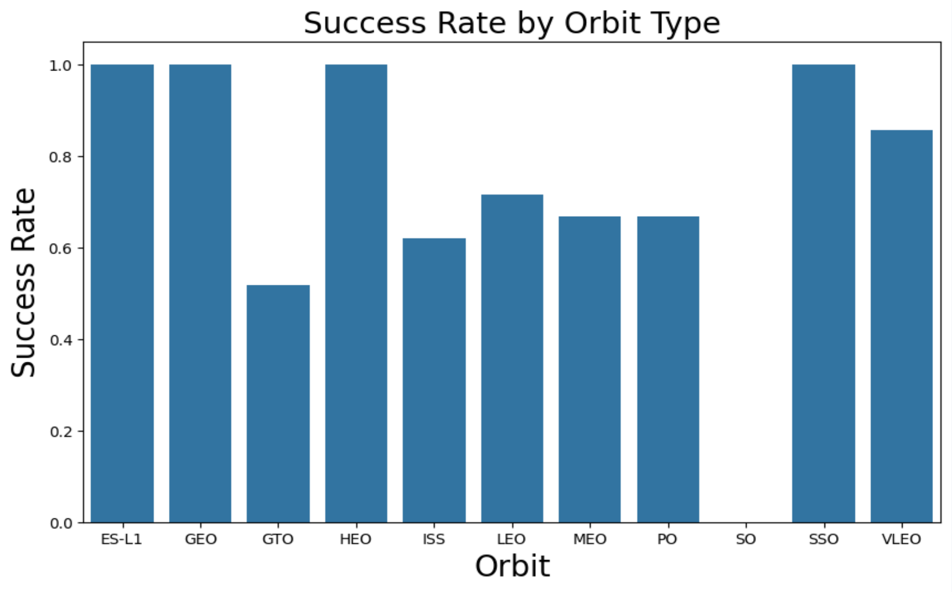
- We performed EDA to determine the testing and training labels
- We calculated the number of occurrence in each orbit as well as the mission outcome
- We also created a landing outcome label
- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcomn9_lan2_datawrangling.ipynb



EDA with Data Visualization

- I used this charts to illustrate the success rate in order to understand the relationship between success and orbit as well as understand how the success rate has increased ever since 2013 which means there's also a trend regarding new rockets and their success.
- GitHub Link: 07/Data





EDA with SQL

- We used SQL lite via python in order to retrieve some important information such as:
  - The distinct launch sites
  - The total payload mass launched by nasa
  - Average payload for a specific booster
  - The date of the first successful landing
  - Booster names with a specific payloads
- We used all of this information to gain even more insight about the data we're using.
- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcon9_SQL_assignment.ipynb

Build an Interactive Map with Folium

- Using python library's folium we built an interactive map in order to have a better visual understanding about the launch's location as well as a visual representation of the success and failures on each site.
- Using this color-labeled outcomes we can better understand the rate of success of each site.

- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcon9_InteractiveVisuals_Folium.ipynb

Build a Dashboard with Plotly Dash

- I created a Dashboard using python's library Plotly in order to better present the data.
- I built pie charts which serve as a great visual representation of the percentage of launches per site
- Also there's a scatter graph that shows the existing relationship between Outcome and Payload Mass.
- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcon9_Dashboard.ipynb

Predictive Analysis (Classification)

- In order to find the best model I used several prediction models which specific parameters with each. This way by using GridSearch I was able to find the best parameter combinations for each model and ultimately see each's accuracy and score to see which was more efficient at predicting a specific outcome.
- GitHub Link: https://github.com/Luipebeltran3107/Data_Science_Capstone/blob/main/Falcon9_Predictive_Analysis.ipynb

Results

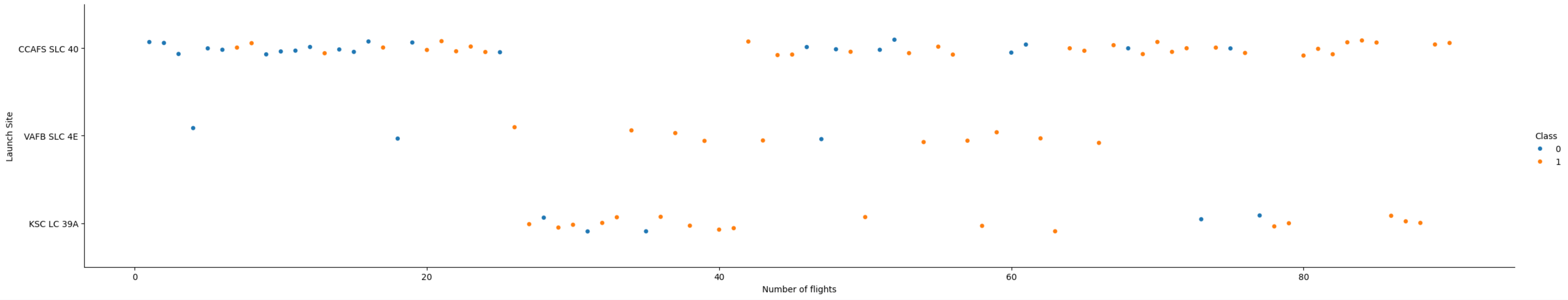
- EDA Results: Here we found that there a strong relationship between PayloadMass and Launch site. Also between number of flights and orbit type. And ultimately a strong relationship between date and success rate.
- Predictive analysis results: After running GridSearch an all models with different parameters, we found that the best performing by a small difference was the Decision tree, specifically due to it’s accuracy.

Section 2

Insights drawn from EDA

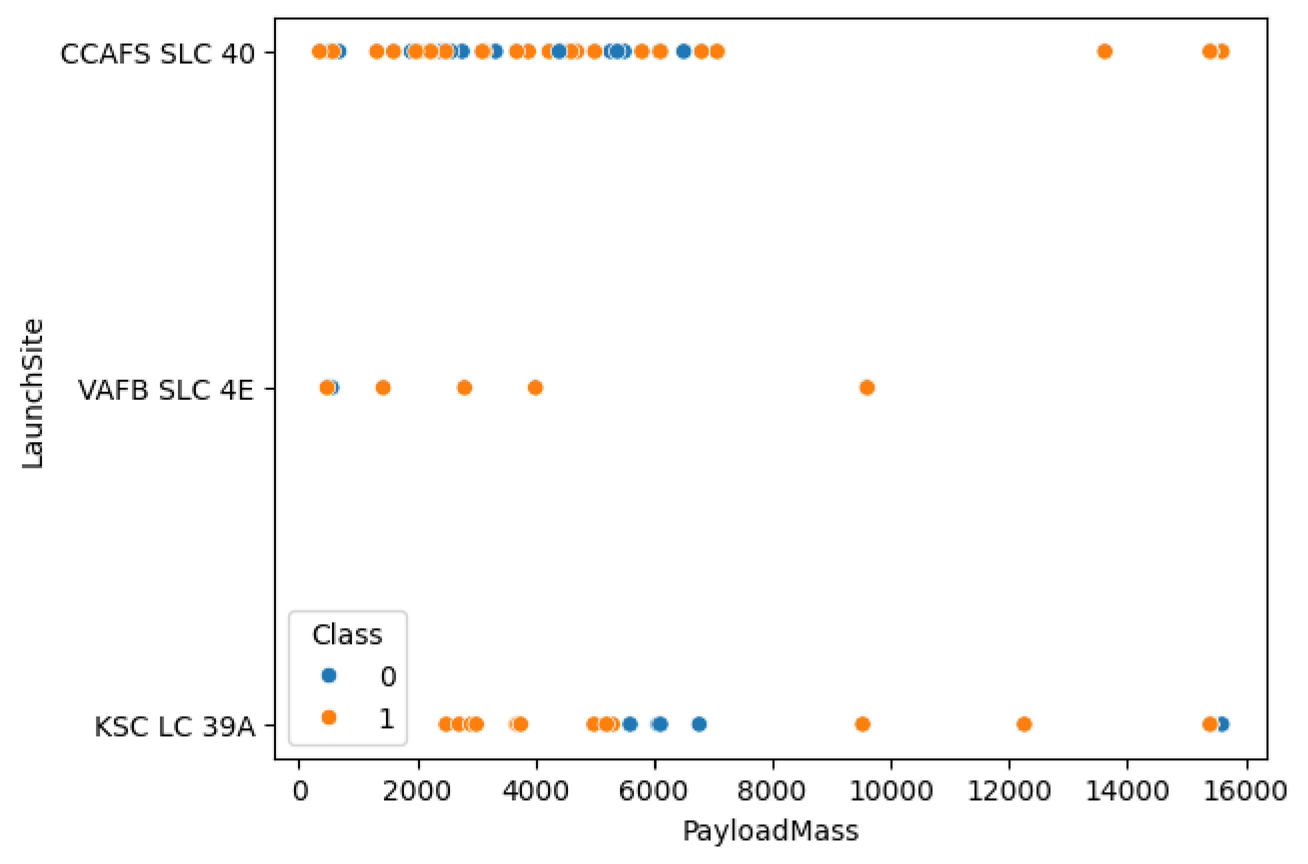
Flight Number vs. Launch Site

- From this data we can infer that the more flights there are at a specific launch site, higher is the chance for success.



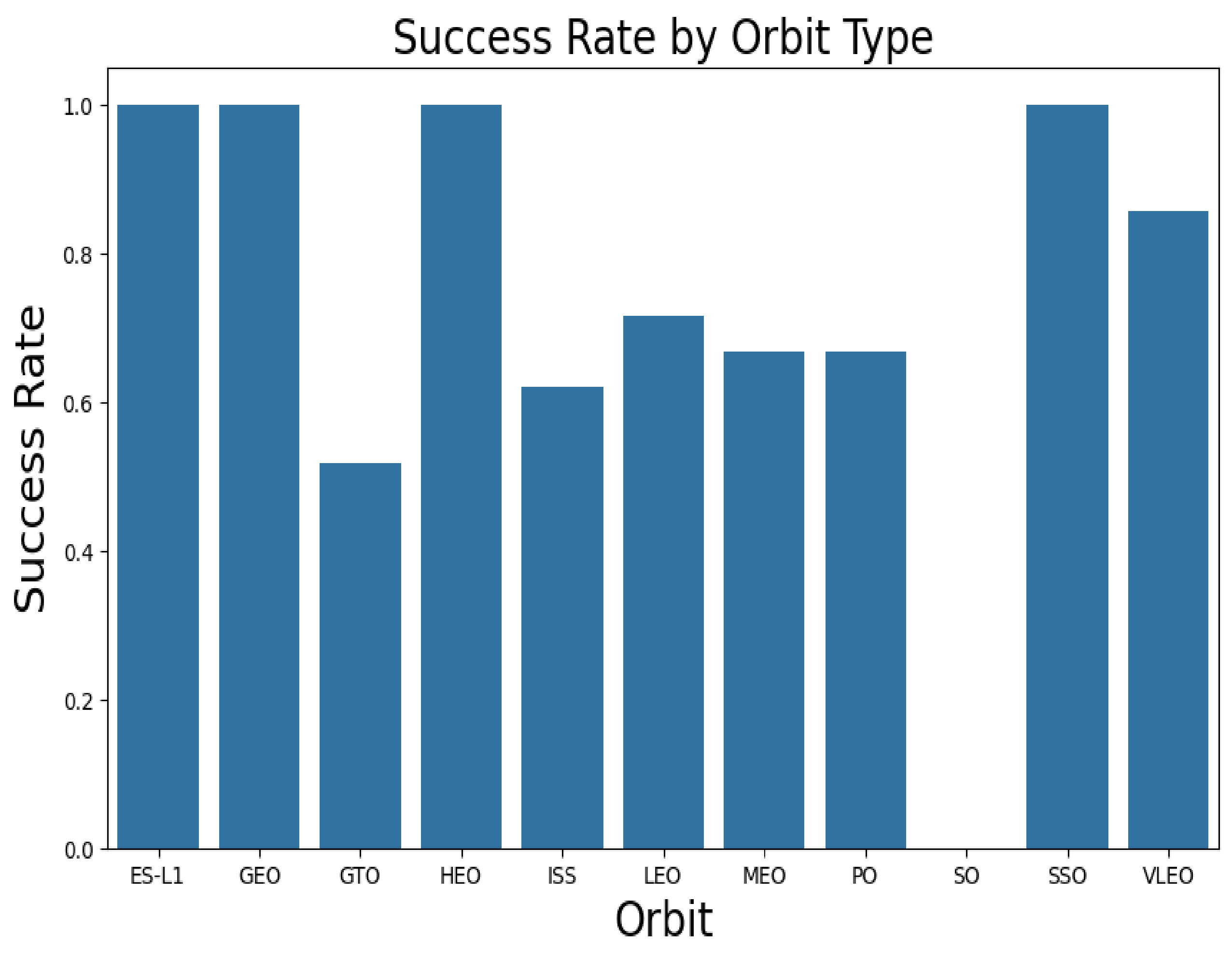
Payload vs. Launch Site

- For the CCAFS SLS 40, the bigger the payload mass, the higher the success rate.
- For the other launch sites, there seems to be close to 0 relationship between both launch site and payload mass.



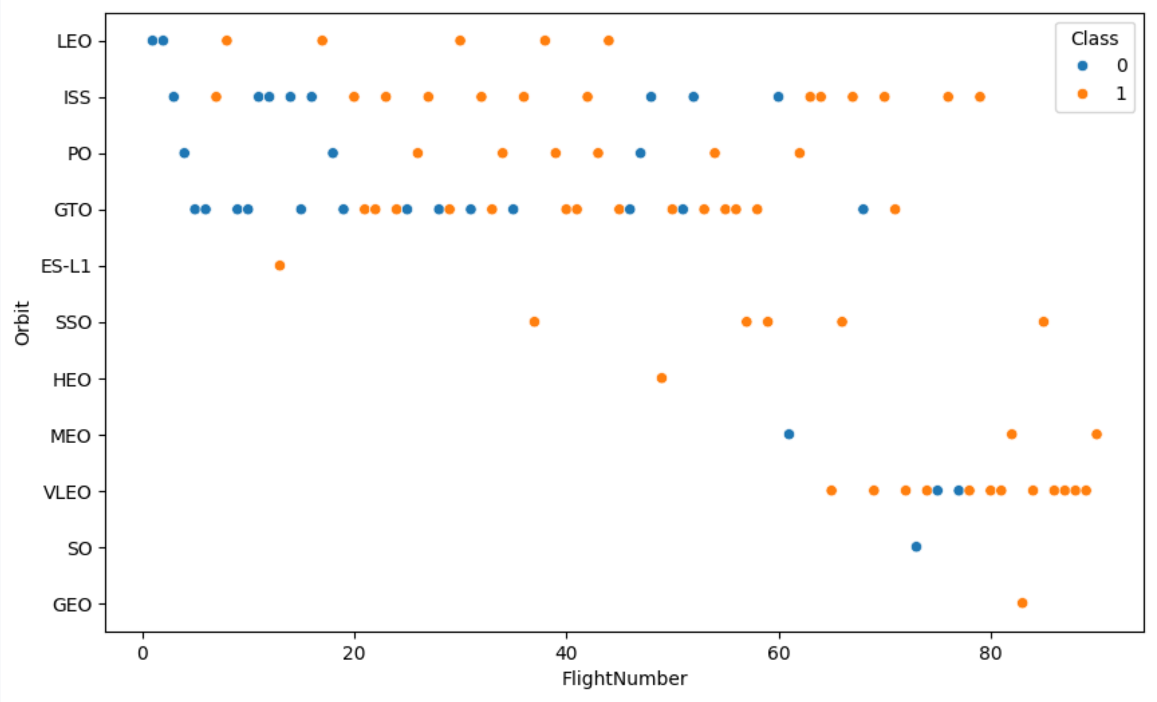
Success Rate vs. Orbit Type

- From this graph we can see which orbits have a higher degree of success rate .



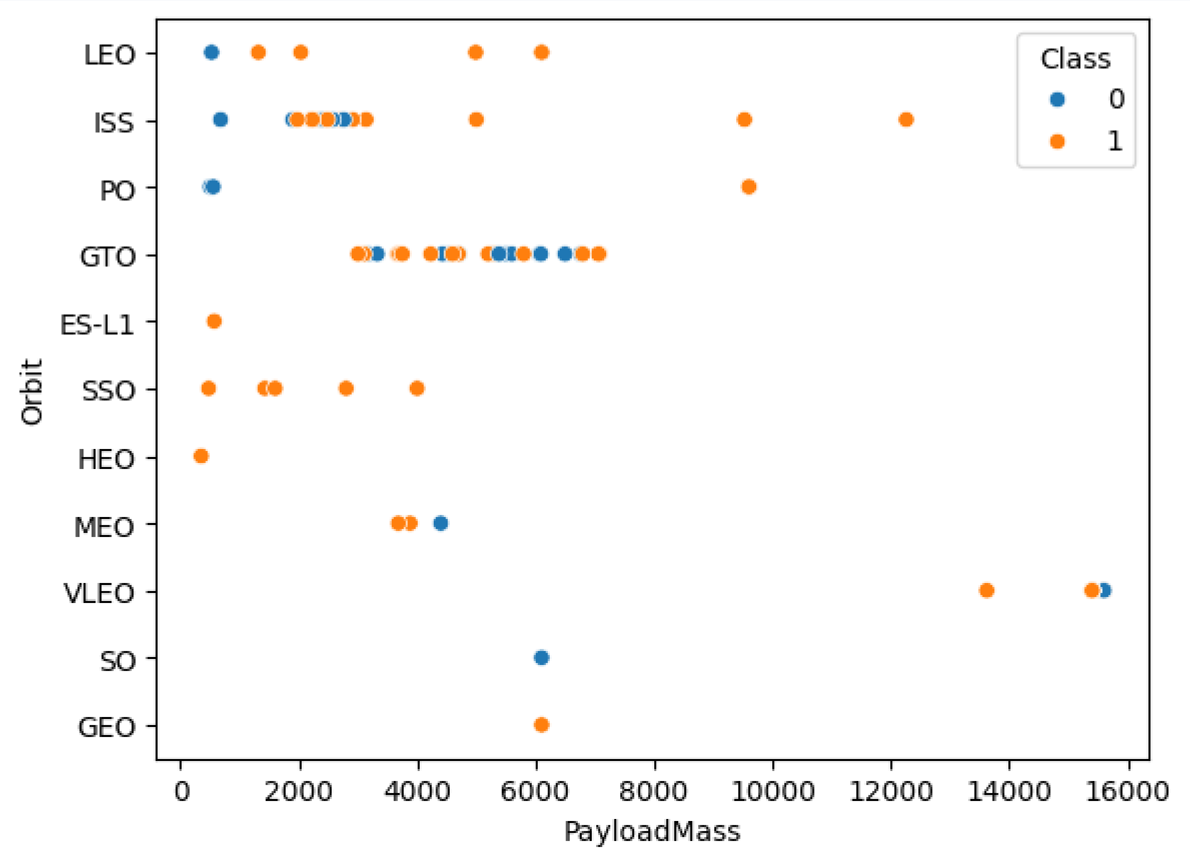
Flight Number vs. Orbit Type

- For orbits such as LEO and PO, there seems to be a relationship between more flights and a better outcome.



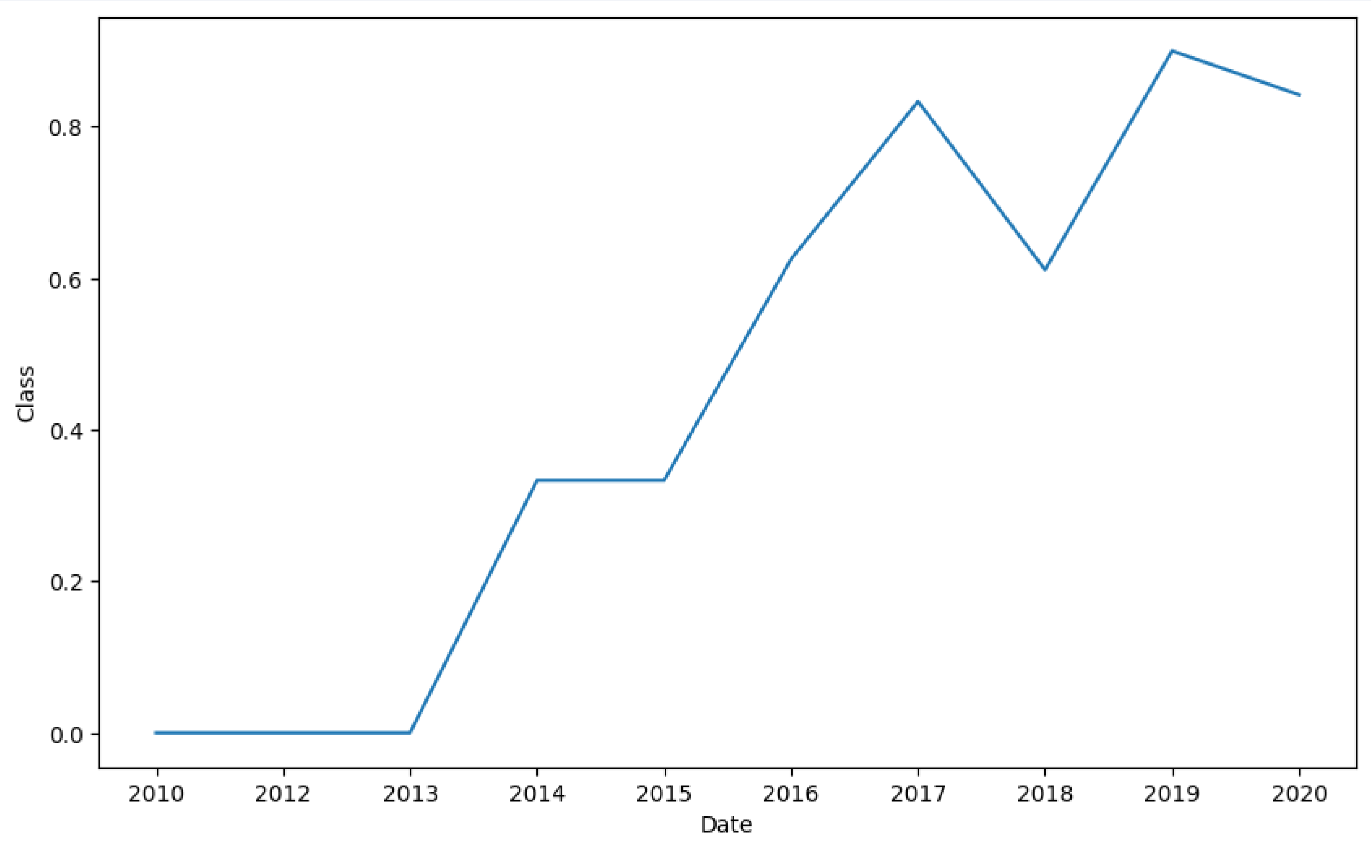
Payload vs. Orbit Type

- We can see that orbits such as Polar, LEO and ISS have more success with heavier payloads.



Launch Success Yearly Trend

- We can see that ever since 2013, launch success has been increasing, though there's a slight dip in 2017.



All Launch Site Names

- There's 4 unique launch sites in the data frame.

Task 1: Display the names of the unique launch sites in the space mission

```
[ ]  %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- This are all the launch sites that begin with the string 'CCA'

Task 2: Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

Total Payload Mass

- As you can see, the total was 45.596 KG

Task 3: Display the total payload mass carried by boosters launched by NASA (CRS)

+ Code + Text

```
[ ]  1 %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload was 2,928.4 KG

Task 4: Display average payload mass carried by booster version F9 v1.1

+ Code + Text

```
[ ]  %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The date for the first successful landing was December 22, 2015

Task 5: List the date when the first succesful landing outcome in ground pad was acheived.

```
[ ]   1 %sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Below you can see the name of the successful booster with a payload between 4000 and 6000 kg.

Task 6: List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)'  AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is 101.

Task 7: List the total number of successful and failure mission outcomes

```
[ ]  1 %sql SELECT COUNT(*) FROM SPACEXTABLE WHERE MISSION_OUTCOME LIKE 'Success%' OR MISSION_OUTCOME LIKE 'Failure%';
```

```
 * sqlite:///my_data1.db
Done.
```

COUNT(*)
101

Boosters Carried Maximum Payload

- Below you can see the boosters that carried the maximum payload mass.

Task 8: List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

+ Code + Text

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

```
 * sqlite:///my_data1.db
Done.
```

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

2015 Launch Records

- The following are the failed landing outcomes using a drone ship in the year 2015

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
1 %sql SELECT * FROM SPACEXTABLE WHERE Landing_Outcome = 'Failure (drone ship)' AND Date LIKE '2015%';
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
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)

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

- Here you can see the ranking of the landing outcomes with the specific variables.

Task 10: Rank 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.

+ Code + Text

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

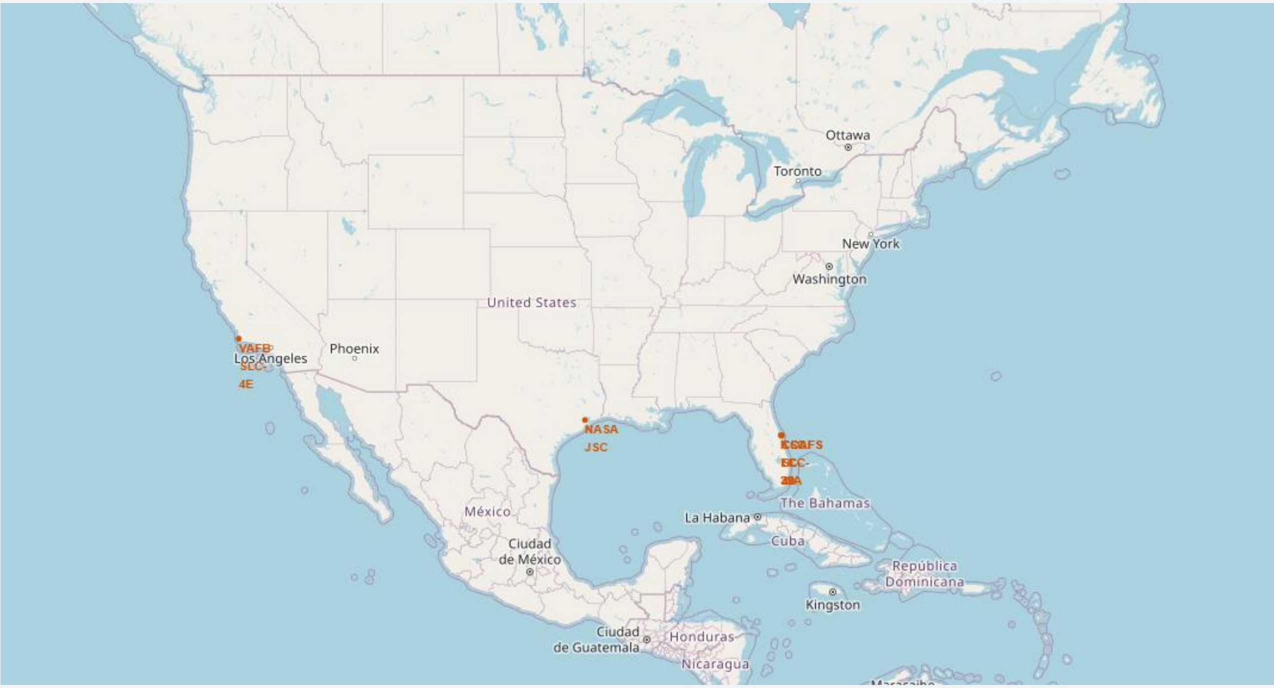
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

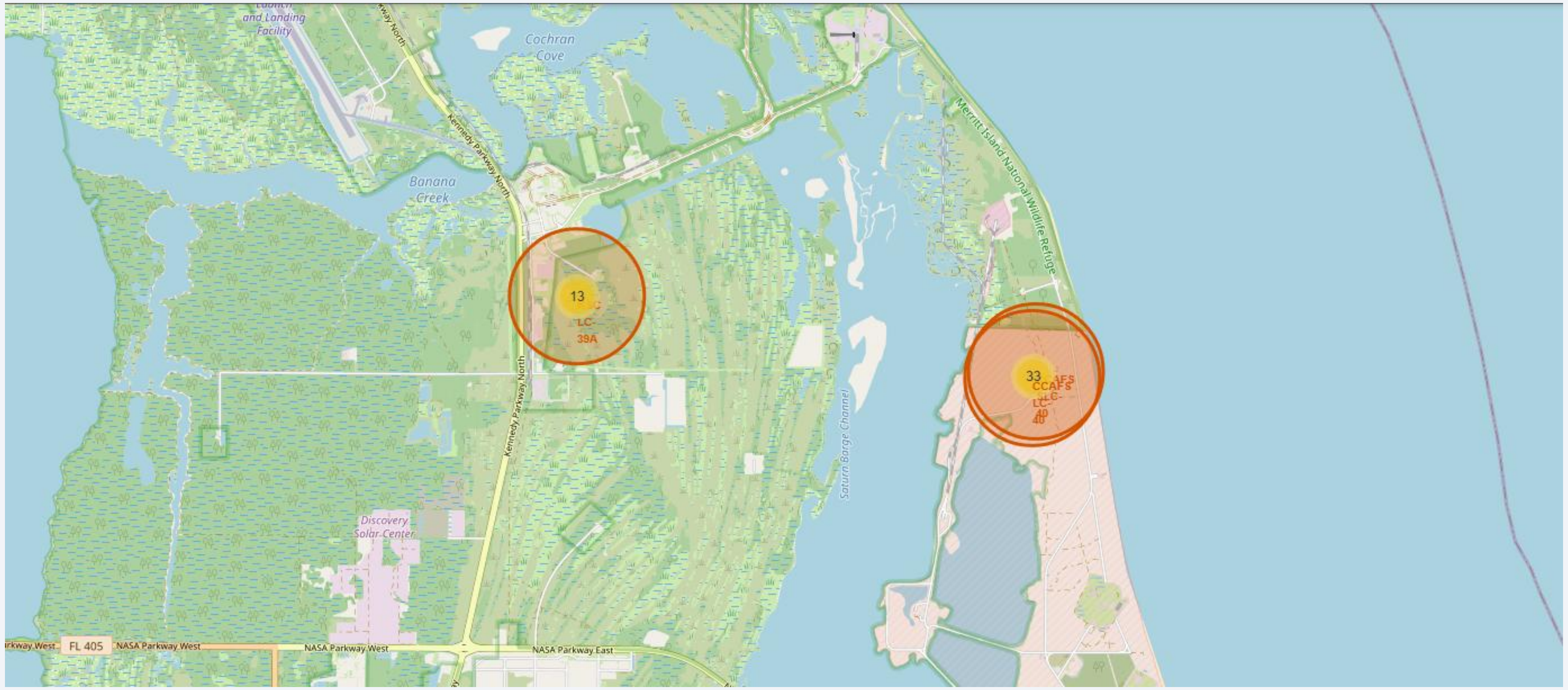
Launch site locations in US map

- In the following map you can see the launch sites in the US and how distant they are.



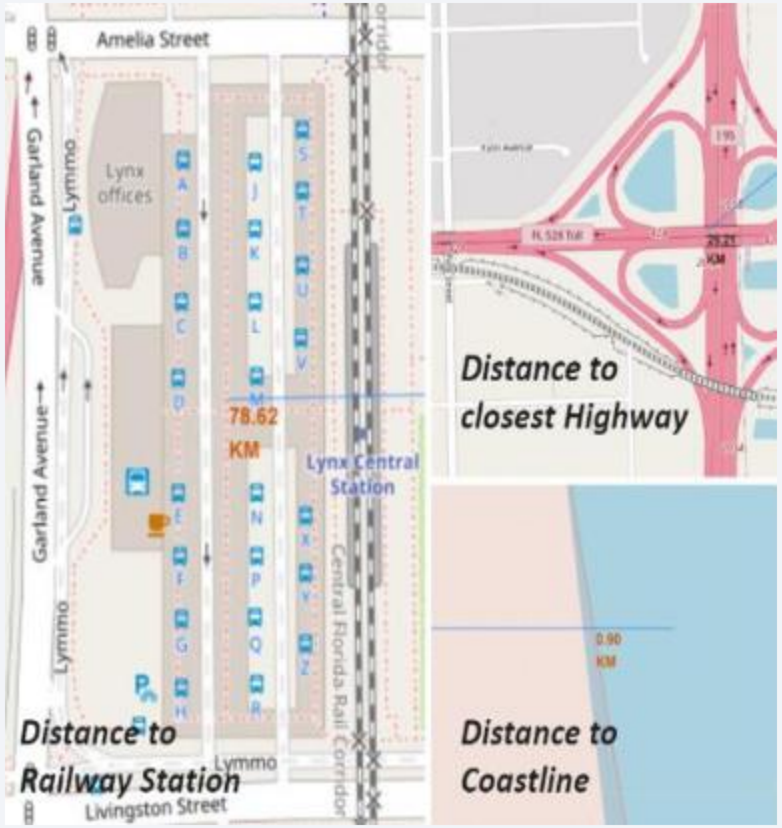
Amount of landings per Launch site in map

- Here you can see the amount of launches per site showing hwo big the difference between each launch site is and how that's going to affect the rate of success.



Landing site proximity to landmarks

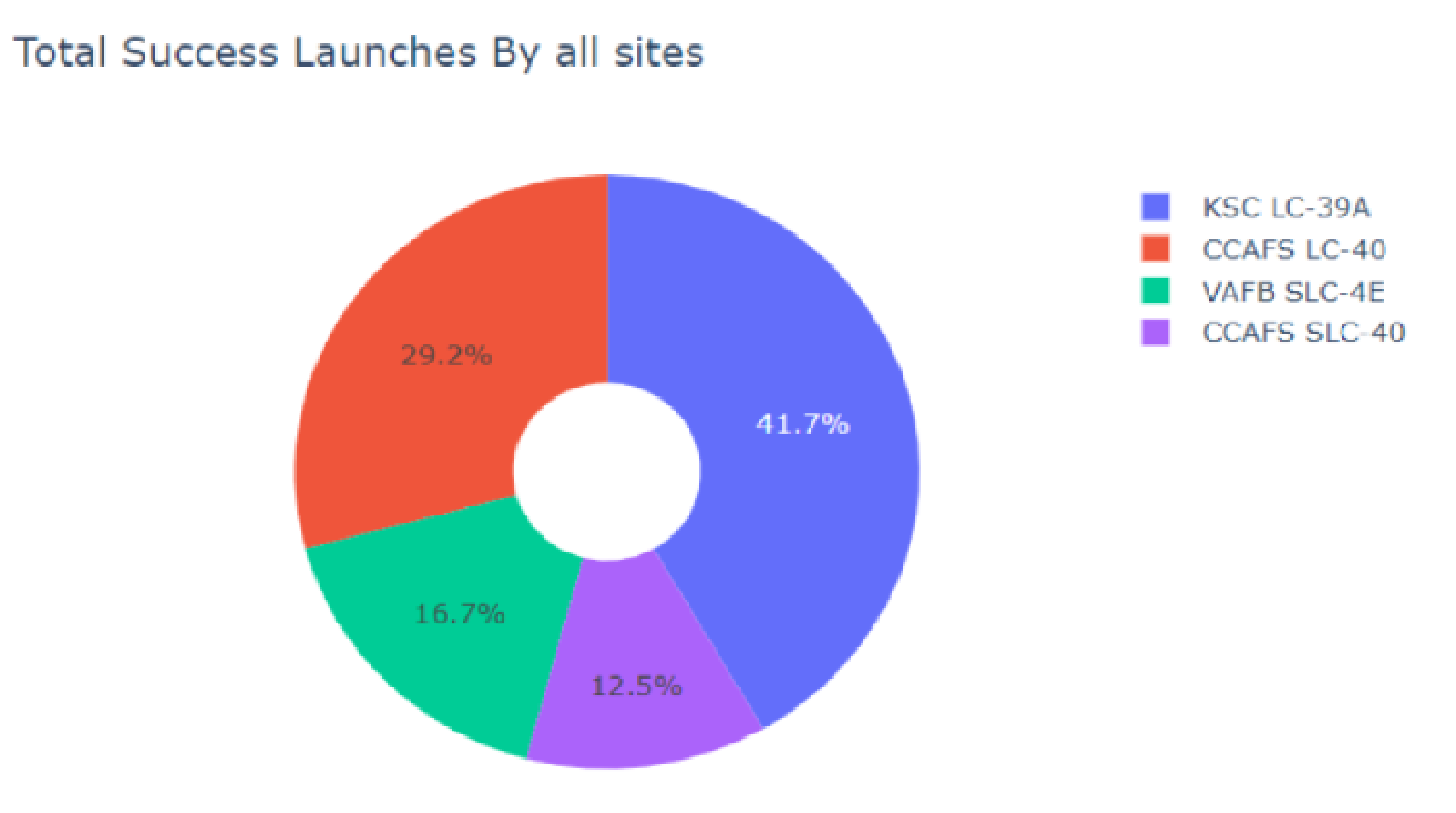
- In the following screenshot we can see a brief analysis of what's in proximity to each launch site to find possible important information that might affect landings success rate.



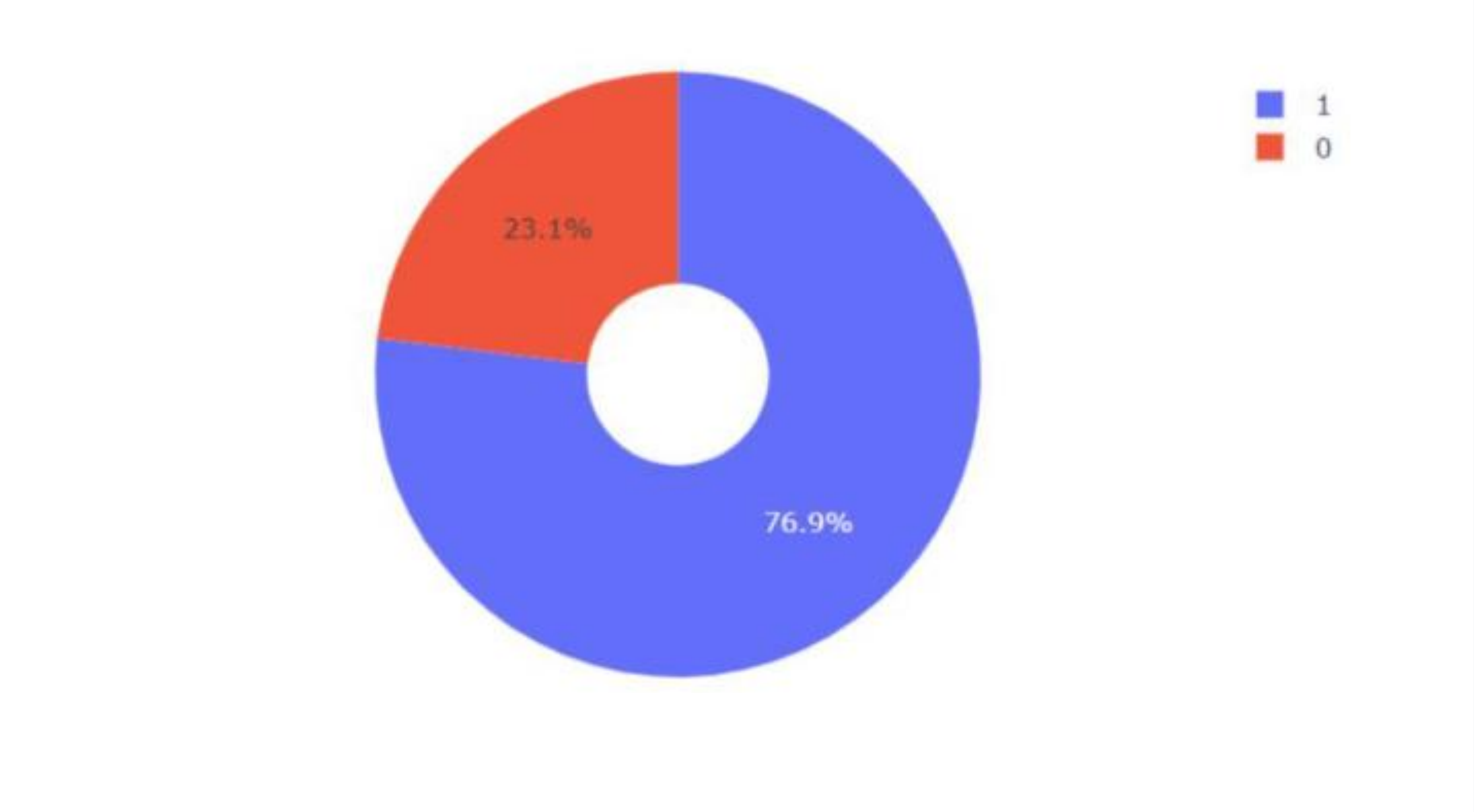
Section 4

Build a Dashboard with Plotly Dash

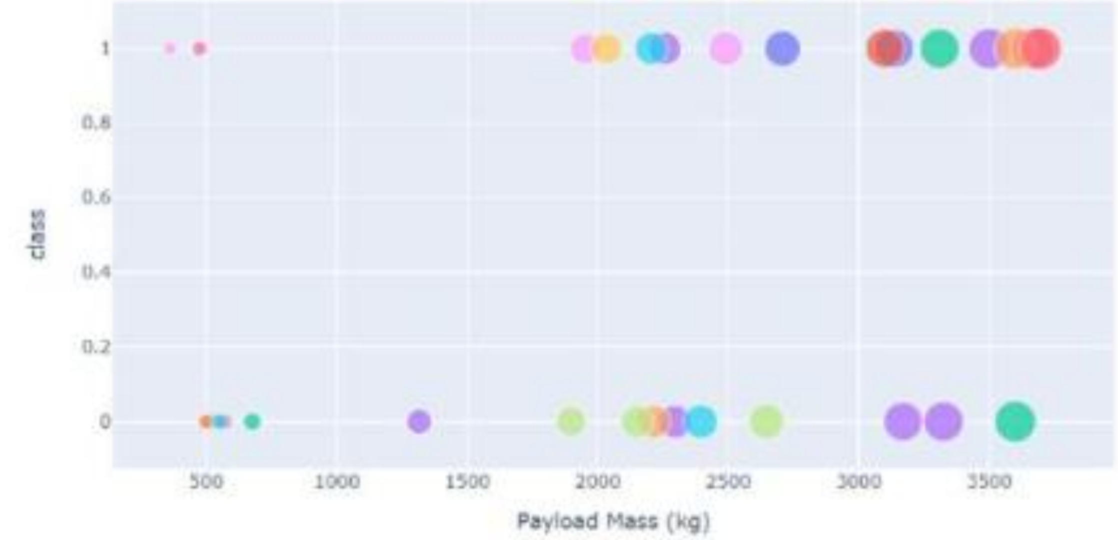
Success percentage per launch site

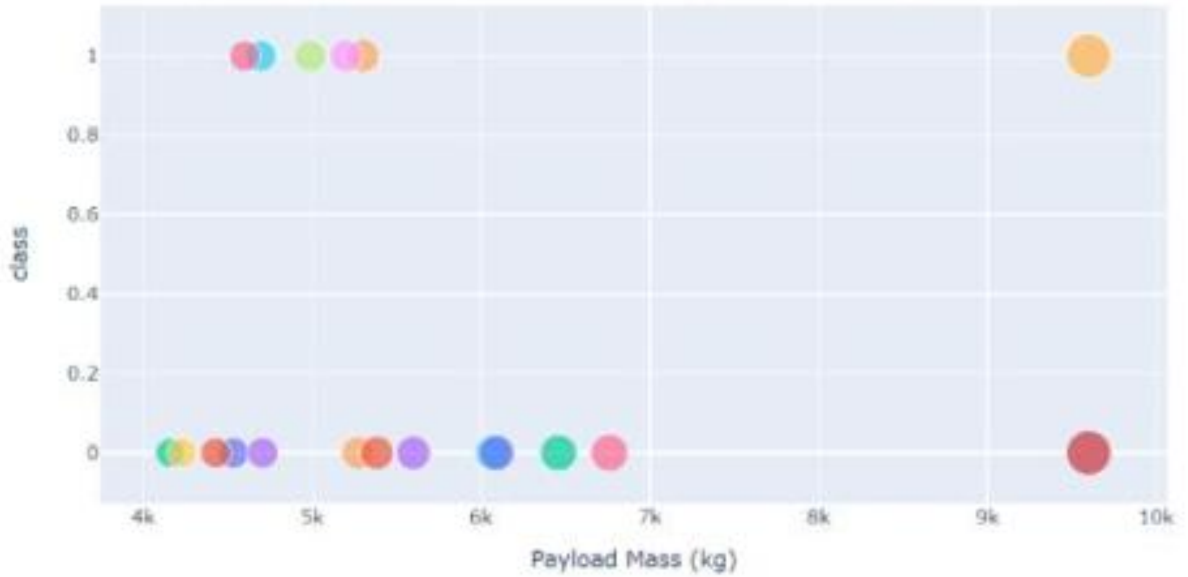


Launch site with highest success ratio: KSC LC39-A



Payload vs. Launch Outcome





Section 5

Predictive Analysis (Classification)

Classification Accuracy

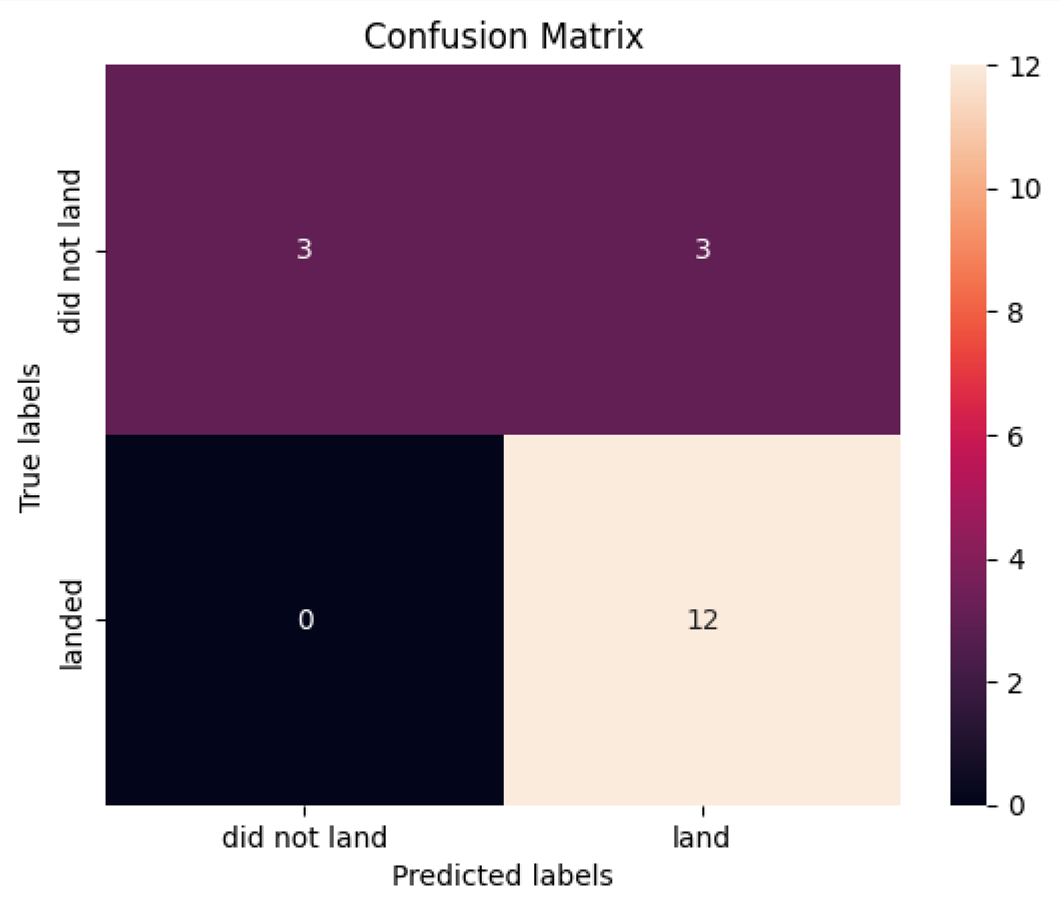
- As the picture shows it, the model with the highest accuracy is the Decision Tree with an astounding 90% accuracy.

```
1 print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
2 print("accuracy :",tree_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5, 'splitter': 'random'}
accuracy : 0.9
```

Confusion Matrix

- This is the decision tree confusion matrix and as you can see it achieved some incredible results considering it has 3 true negatives and 12 true positives.



Conclusions

- From all the data presented we can conclude that:
  - The more flights been done, the higher the success rate
  - Theres some launch sites that have better success with heavier payloads
  - Launch success rate has constantly increased since 2013
  - KSC LC-39A is the launch site with the most amount of success
  - The decision tree classifier is the best classifier model of all three.

Thank you!