IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Mina Edwar>
<7/8/2024>

Outline

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

Executive Summary

IN THIS PROJECT WE DID DATA COLLECTION, DATA WRANGLING, EXPLORATORY DATA ANALYSIS, INTERACTIVE VISUAL ANALYTICS AND PREDICTIVE ANALYSIS.

WE ALSO DID SQL QUERIES THAT SHOWED GREAT RESULTS IN EDA.

WE HAVE FOCUSED ON FOUR MODELS AND ESPECIALLY ON THE BEST MODEL OUT OF THEM THAT GAVE US 88.8% ACCURACY.

Introduction

- we want to 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.

Section 1

Methodology

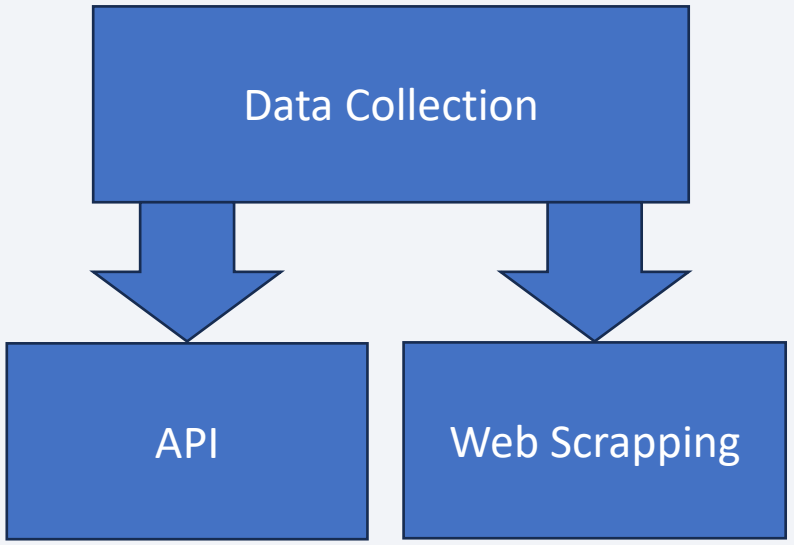
Methodology

Executive Summary

- Data collection methodology:
  - Collected data in two ways: API, Scraping.
- Perform data wrangling
  - Landing outcomes mapped to two classes 0(Failure) and 1(Successful).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We have built 4 models, tuned them and evaluated them with Decision Tree achieving the highest accuracy but the other models also performed well.

Data Collection

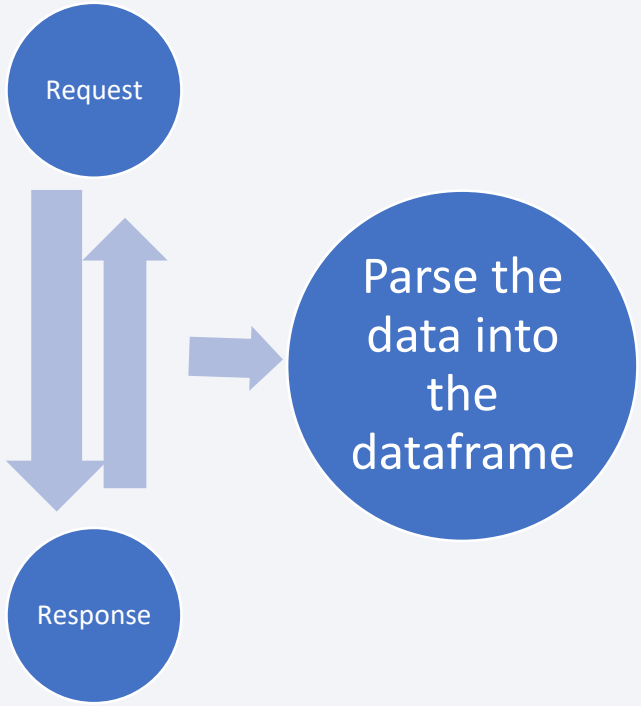
- In different two ways, we collected the data.



Data Collection – SpaceX API

- We got the url so we requested to get the data with this url, then we got a response with the json file that we started parsing the content into a dataframe.

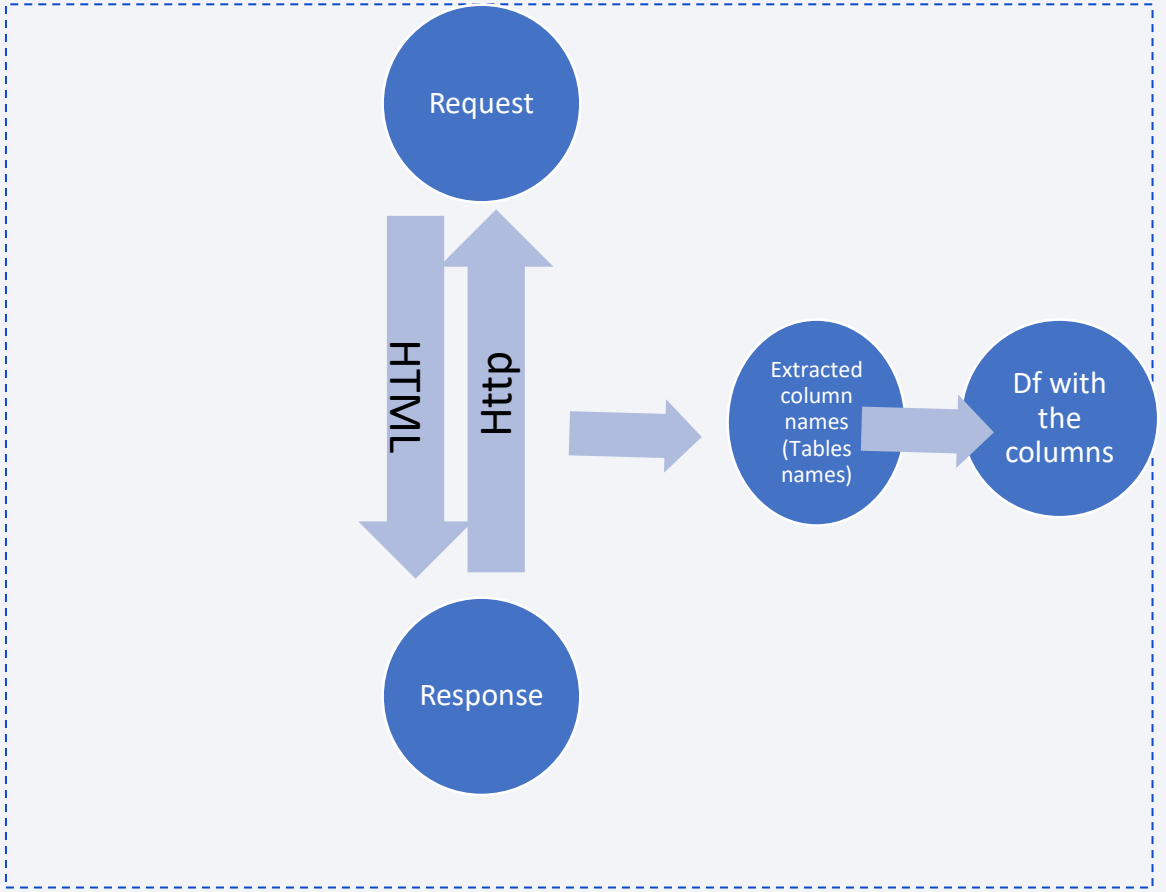
- Click to see the notebook with the code:



Data Collection - Scraping

- We requested with the URL : Falcon9 Launch HTML page then we got the http response and started getting the text of the content into an object of BeautifulSoup.

- Click here to see the notebook with the code:



Data Wrangling

- Calculated the number of launches on each site, then mapped the landing outcomes into two classes of landing 0 , 1.
- Class 0 for unsuccessful landing (bad), 1 for successful landing(good).

- Click here to see the notebook with the code:

```
    print(i,outcome)
```

```
0 True ASDS
1 None None
2 True RTLS
3 False ASDS
4 True Ocean
5 False Ocean
6 None ASDS
7 False RTLS
```

We create a set of outcomes where the second stage did not land successfully:

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
bad_outcomes
```

```
{'False ASDS', 'False Ocean', 'False RTLS', 'None ASDS', 'None None'}
```

TASK 4: Create a landing outcome label from Outcome column

Using the `Outcome`, create a list where the element is zero if the corresponding row in `Out` one. Then assign it to the variable `landing_class`:

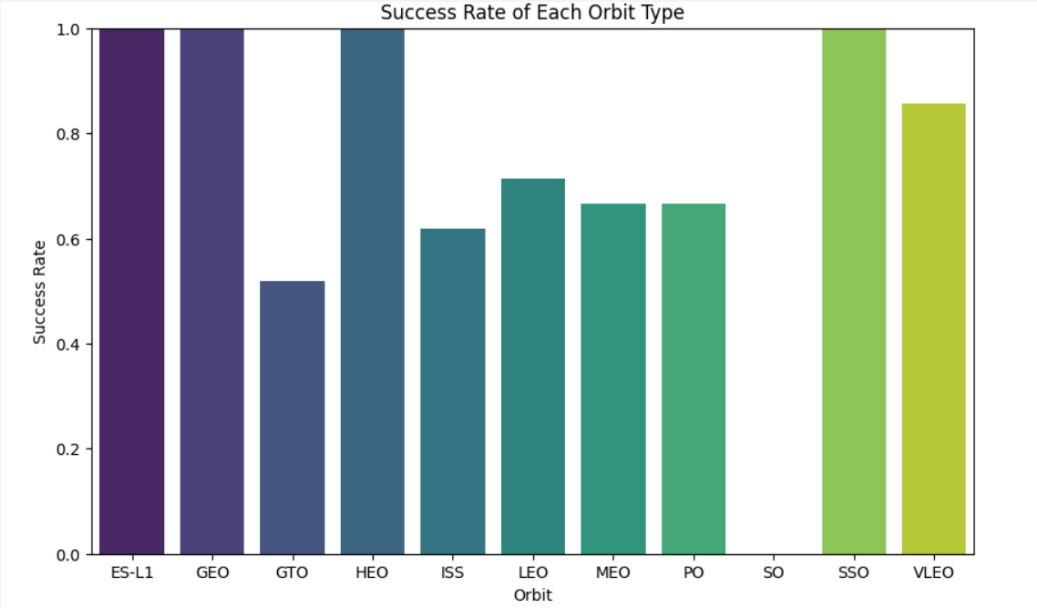
```
landing_class= df['Outcome'].apply(lambda x: 0 if x in bad_outcomes else 1)
```

This variable will represent the classification variable that represents the outcome of each lau land successfully; one means the first stage landed Successfully

```
df['Class']=landing_class
df[['Class']].head(8)
```

EDA with Data Visualization

- We did scatter, line, cat and bar plots.
- Check them here in the notebook:

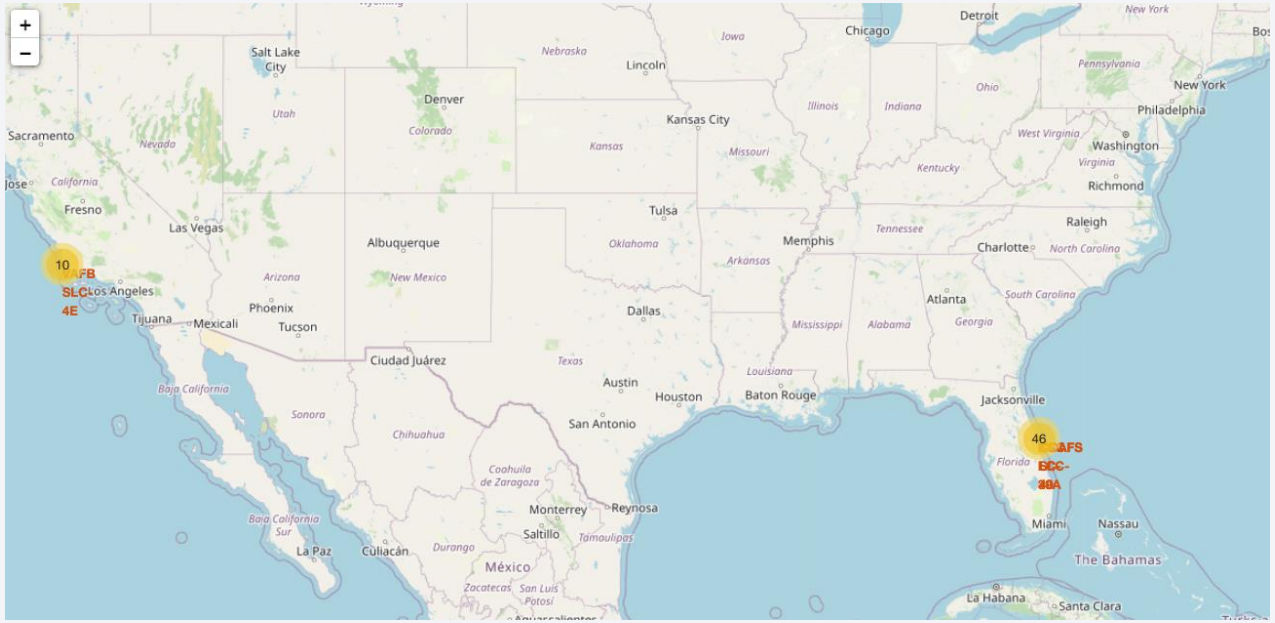


EDA with SQL

- First query displayed the name of the unique launch sites.
- Second query displayed records where launch sites begin with the string 'CCA'.
- Third query displayed the total payload mass carried by boosters launched by NASA (CRS).
- Fourth query displayed the average payload mass carried by booster version F9 v1.1.
- Fifth query listed the date when the first successful landing outcome in ground pad was achieved.
- Sixth query listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- Seventh query listed the total number of successful and failure mission outcomes.
- Eighth query using subquery listed the names of the booster versions which have carried the maximum payload mass.
- Nineth query List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015.
- Tenth query ranked 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..
- Check the notebook here <<<<<

Build an Interactive Map with Folium

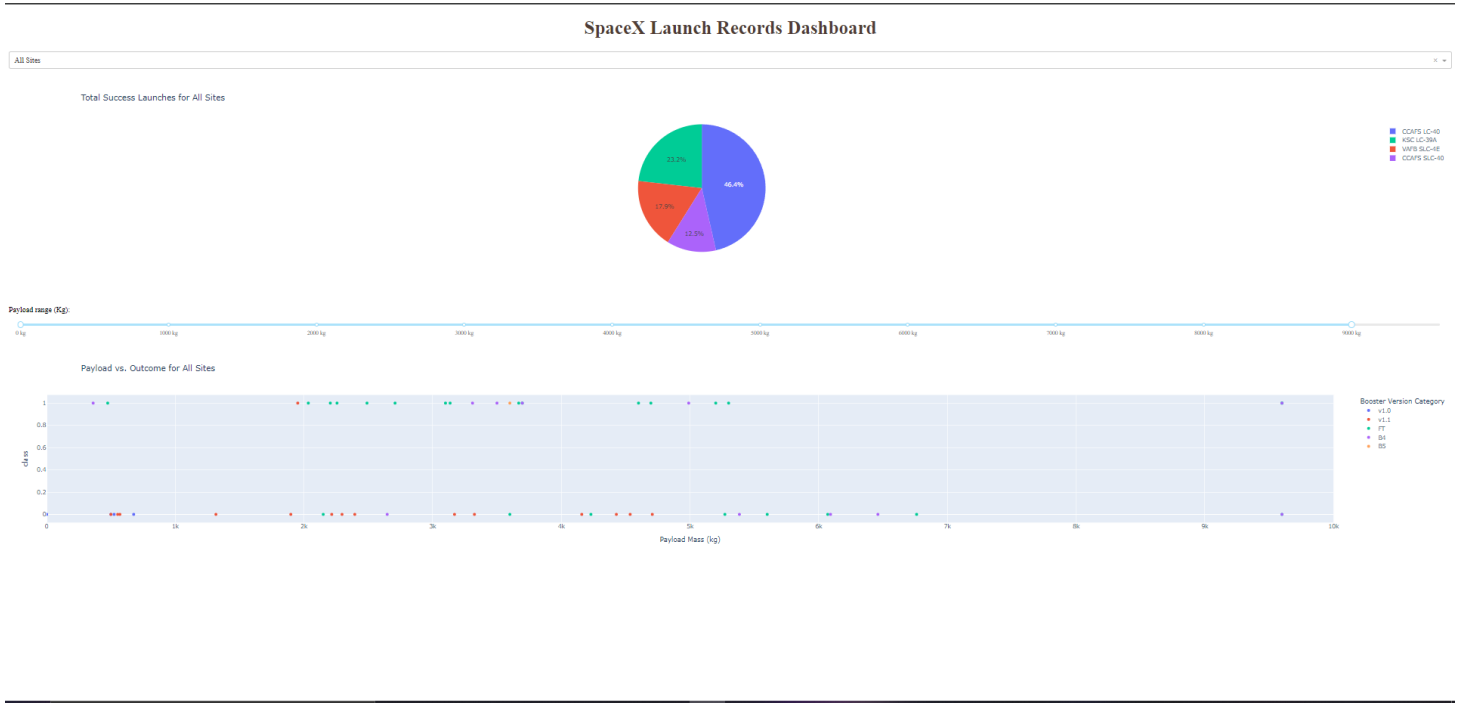
- We created map objects like Circle, Marker, to mark all launch sites on the map then we used markerCluster object with green color if successful landing and red one if failed landing.



- >>> Notebook here <<<

Build a Dashboard with Plotly Dash

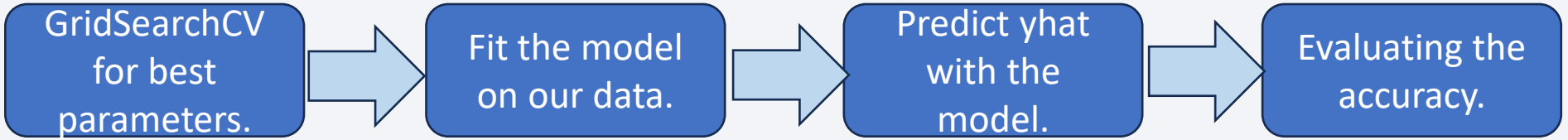
- Did a dashboard with pie chart and scatter plots.
- As u can see u can check the percentage of successful landings in each site and in all sites together so we can gain some useful insights.



- >>> Check the code here <<<

Predictive Analysis (Classification)

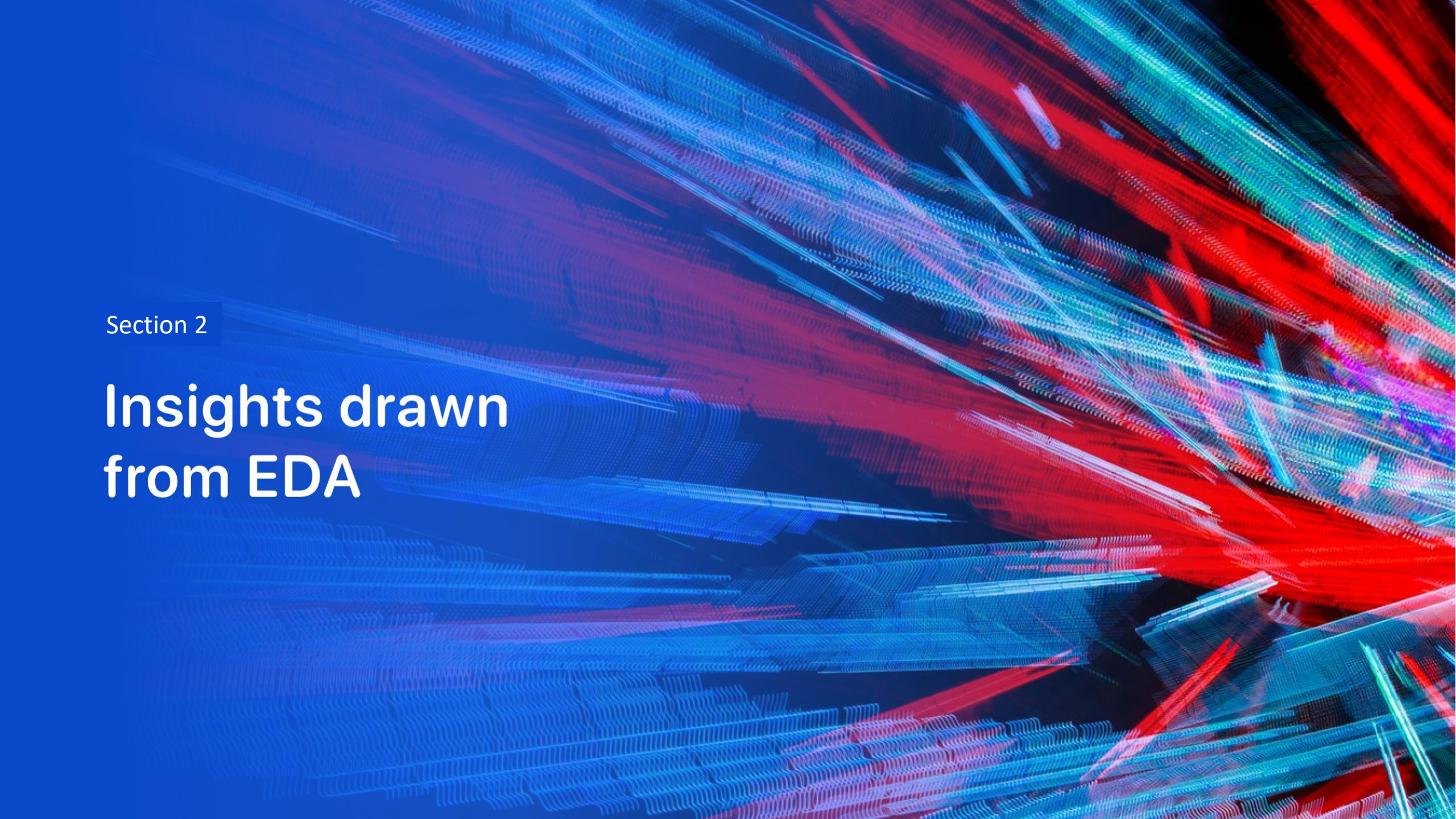
- Using the built-in scikit-learn library we used the built in models to fit on our data.
- We used 4 different models to predict if the falcon 9 will land or no.
- SVM, Decision Tree, KNN, Logistic Regression.
- After splitting the data the process went like this.



- >>> Check notebook here <<<

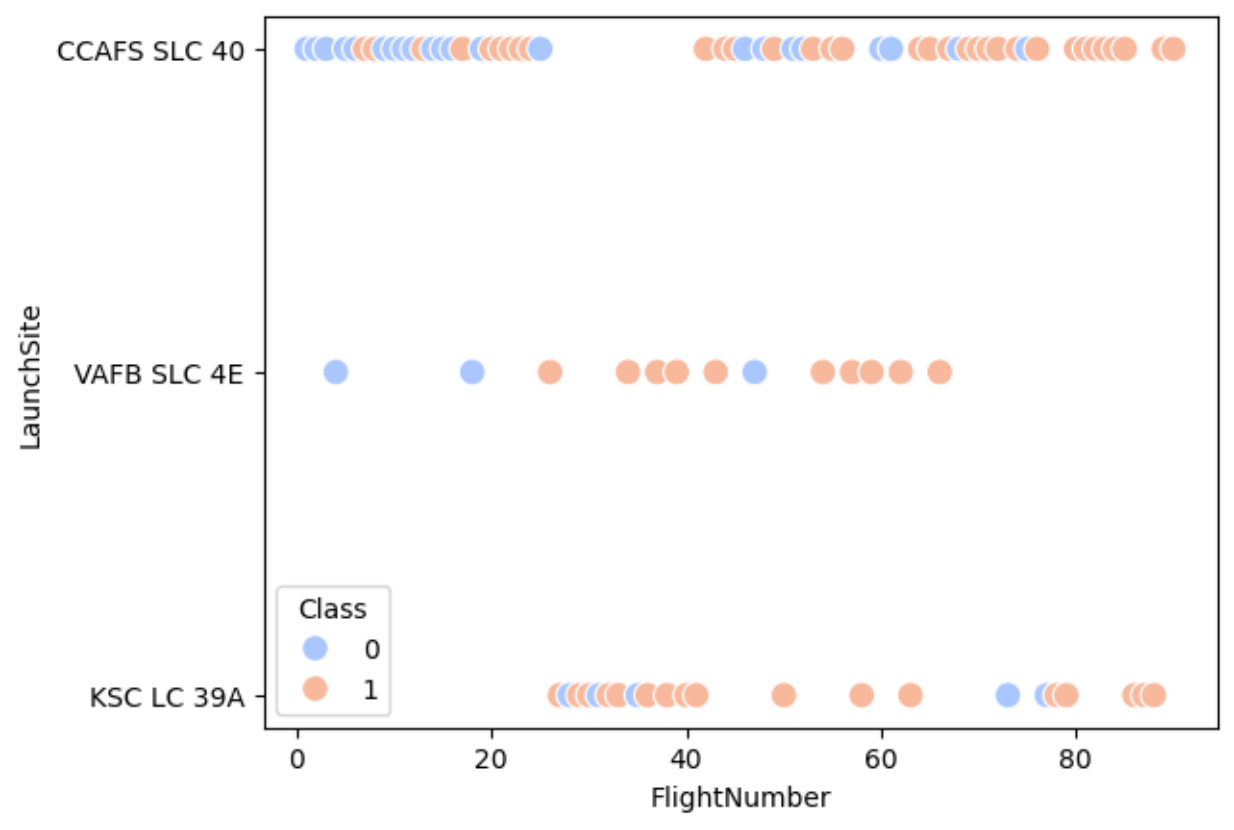
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

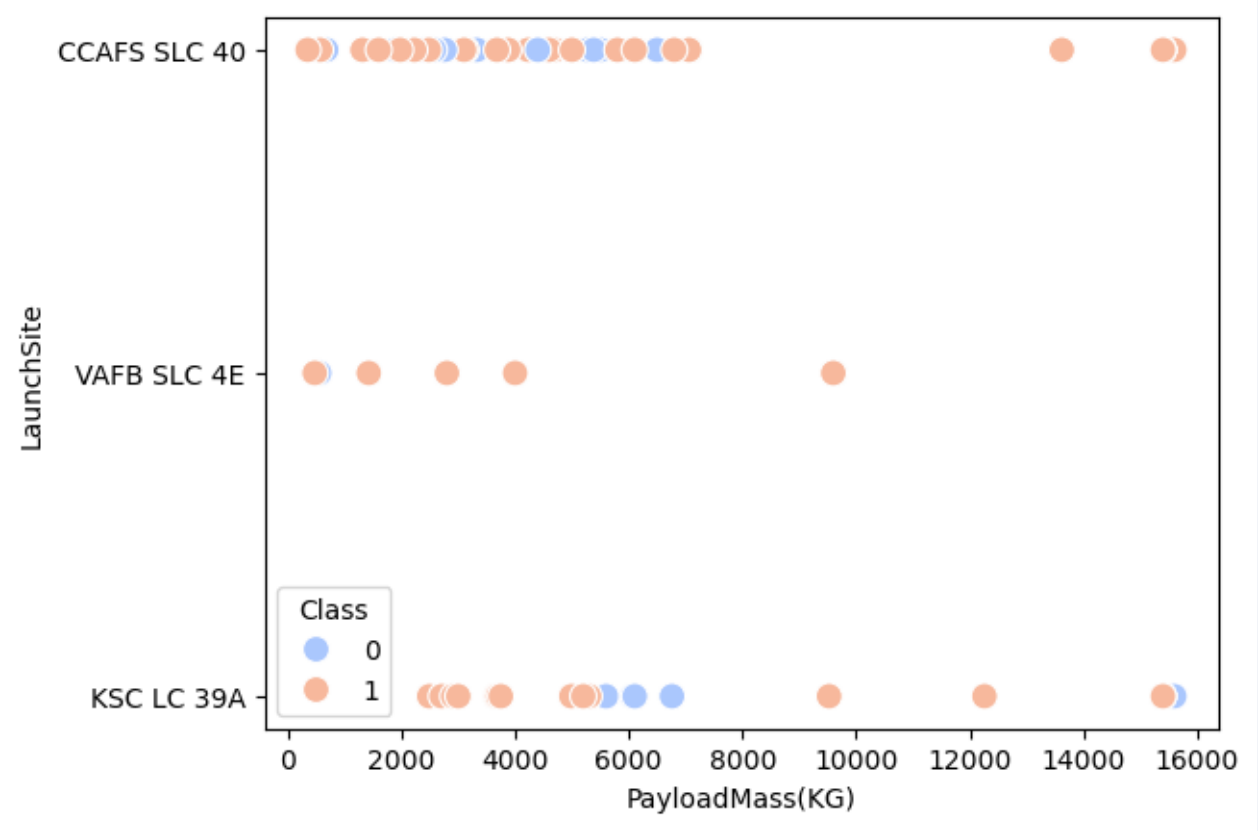
- (x-axis) Flight Number vs. (y-axis) Launch Site scatter plot.
- As you can see for VAFB SLC 4E if the flightNumber > 50 it is completely class 1!



Payload vs. Launch Site

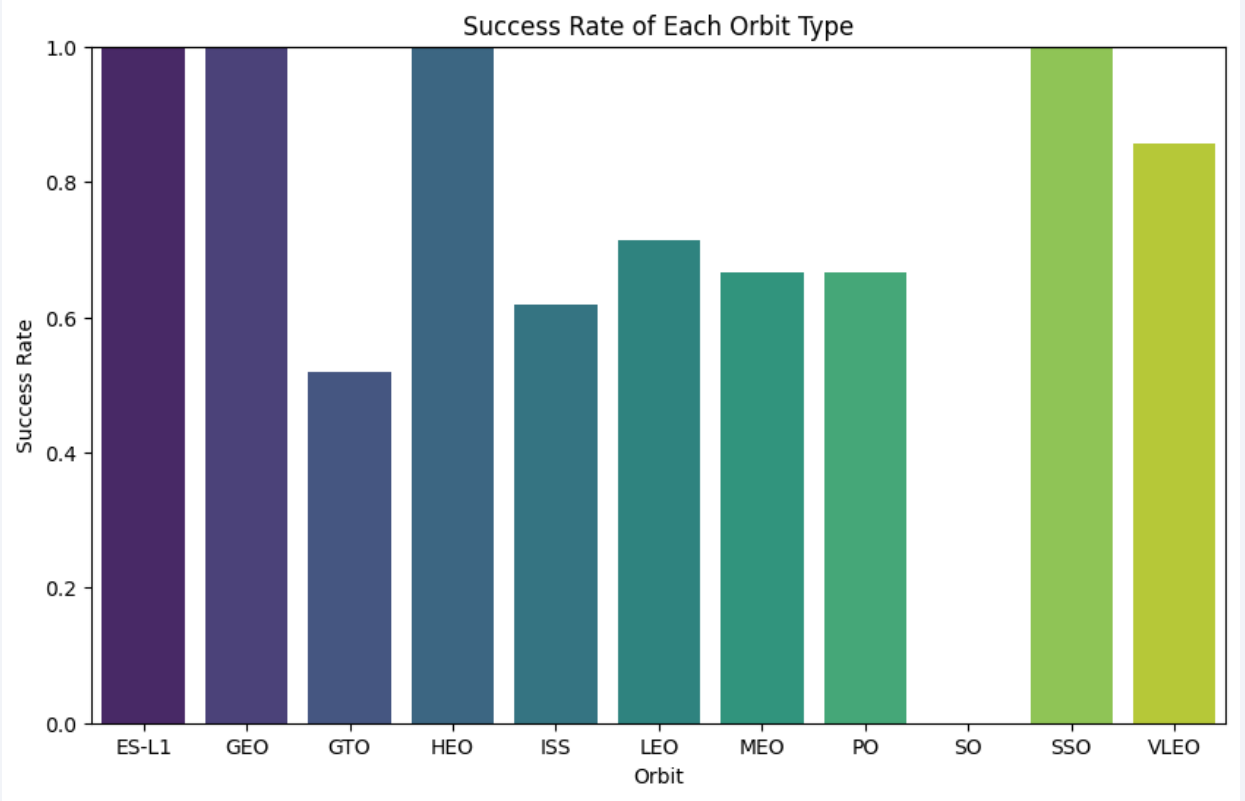
- (x-axis) Payload vs. (y-axis) Launch Site

- If the mass > 7000kg for CCAFS SLC 40 it is completely class 1 also if mass > 1000kg for VAFB SLC 4E.



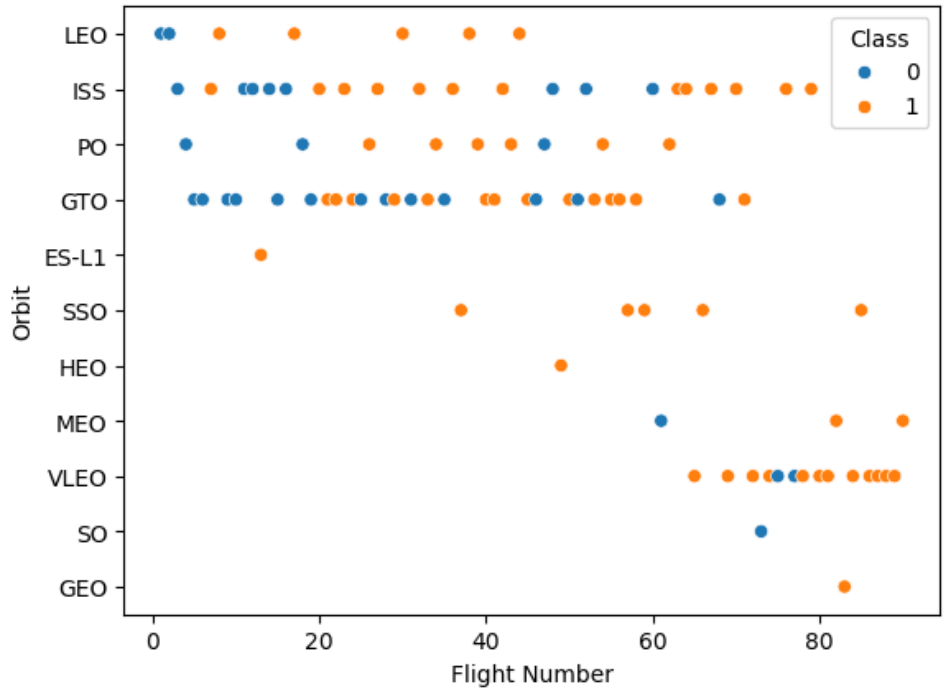
Success Rate vs. Orbit Type

- We can conclude that the success rate for ES-L1, GEO, HEO, SSO orbits 100%.
- Success Rate for SO orbit is 0%.



Flight Number vs. Orbit Type

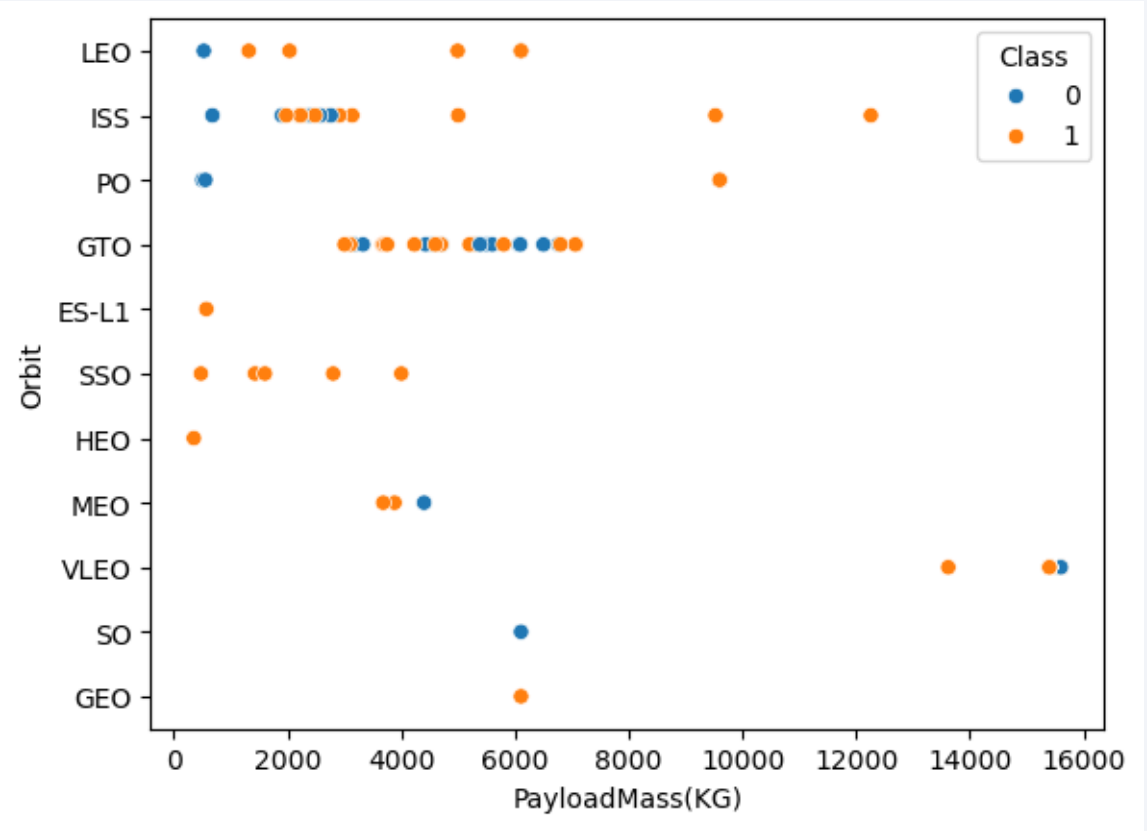
- (x-axis) Flight number vs. (y-axis) Orbit type
- SO is always class 0 independent on FlightNumber (always fails).
- While GEO, SSO, HEO, ES-L1 are all of class 1 which is always successful.



Payload vs. Orbit Type

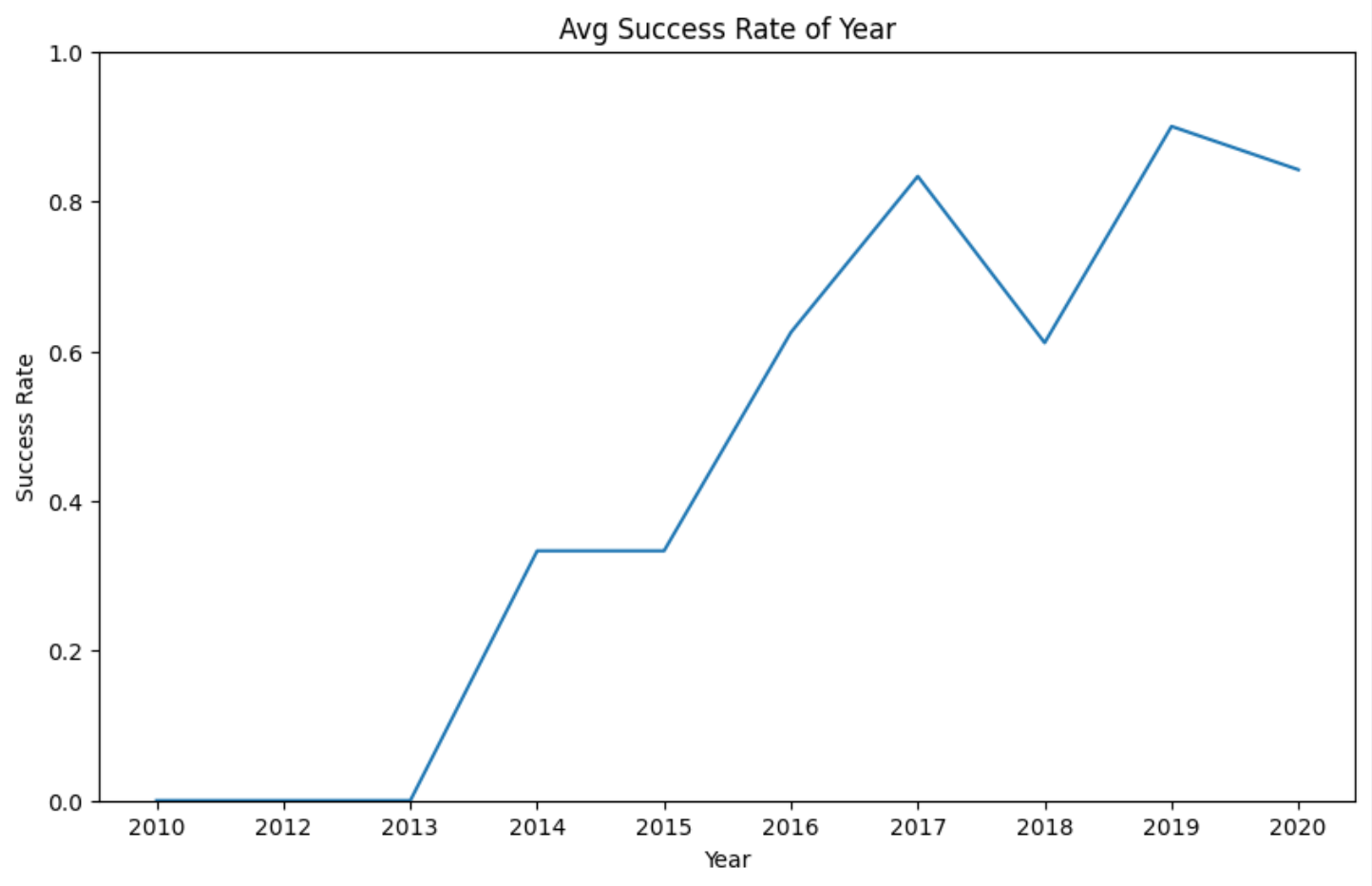
- (x-axis) payload vs. (y-axis) orbit type

- As mentioned before the 5 orbits ES-L1, SSO, HEO, SO and GEO are independent on any of the features.



Launch Success Yearly Trend

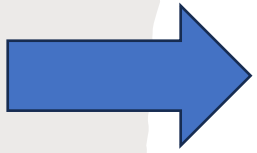
- Yearly average success rate.
- As we can see the interval 2015-2017 showed us increase in the success rate!
- 2019 is the most year with avg success rate.



All Launch Site Names

- The following are the names of the unique launch sites!

- There are 4 distinct launch sites.



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'

- 5 records where launch sites begin with `CCA`

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

- The total payload carried by boosters from NASA is 45596kg.

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

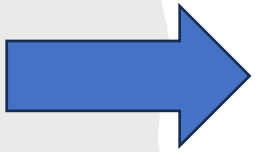
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer='NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4kg.



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

```
%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 of the first successful landing outcome on ground pad is 2015-12-22

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

Hint:Use min function

```
%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

- The list of names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000!

- There are 4 booster versions to do so.

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
```

```
* 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.

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(Mission_Outcome) from SPACEXTABLE
```

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

COUNT(Mission_Outcome)
101

Boosters Carried Maximum Payload

- The list of names of the booster which have carried the maximum payload mass:

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT 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 failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in 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.

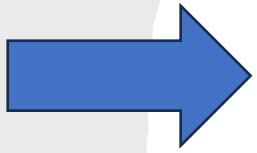
```
%sql SELECT SUBSTR(Date, 6, 2), Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE Landing_Outcome = 'Fai
```

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

SUBSTR(Date, 6, 2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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



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.

```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) as Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-0
```

```
 * 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

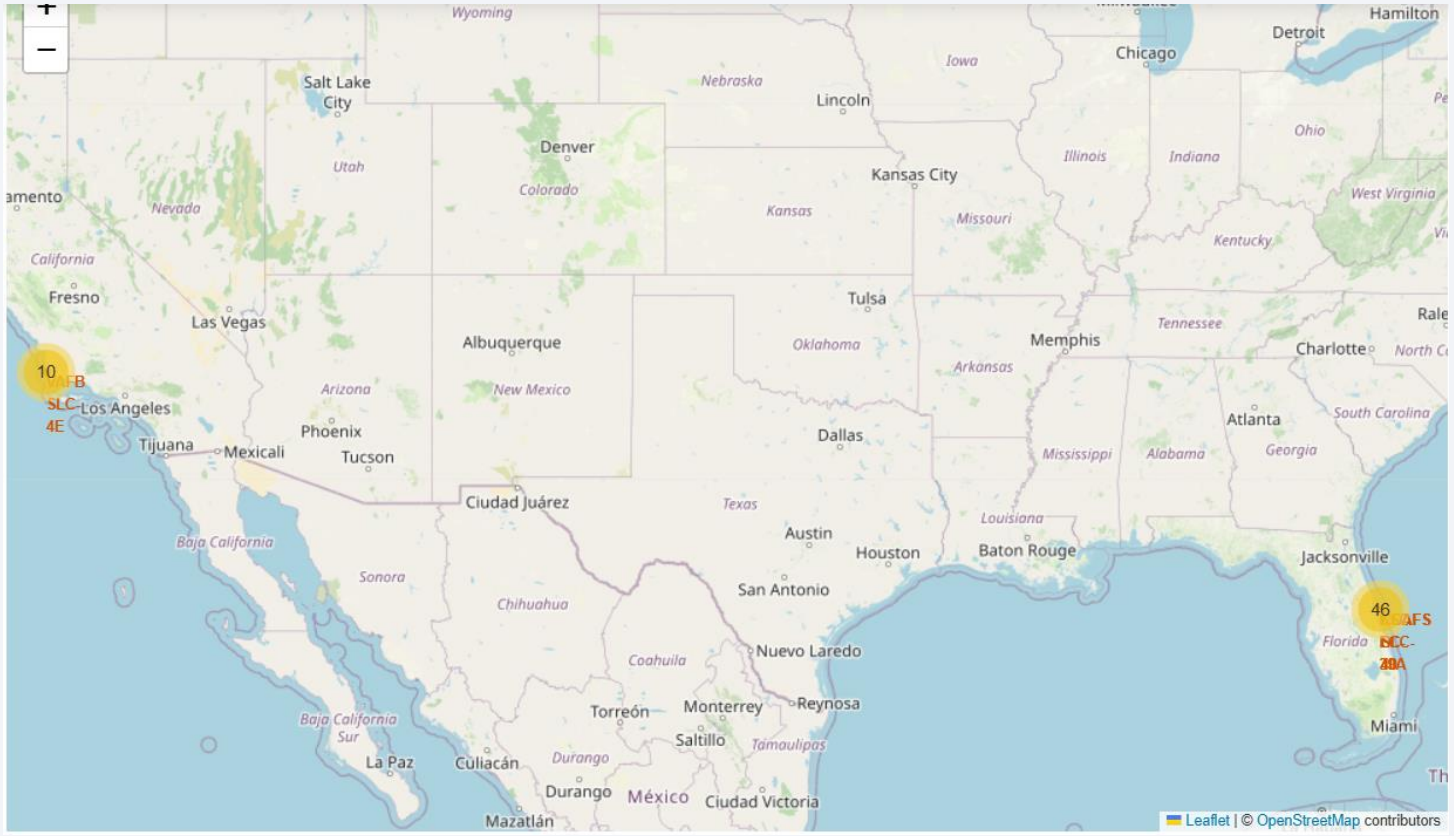
Launch Sites Proximities Analysis

Launch Sites Map

- Locations marked on the map with red color.

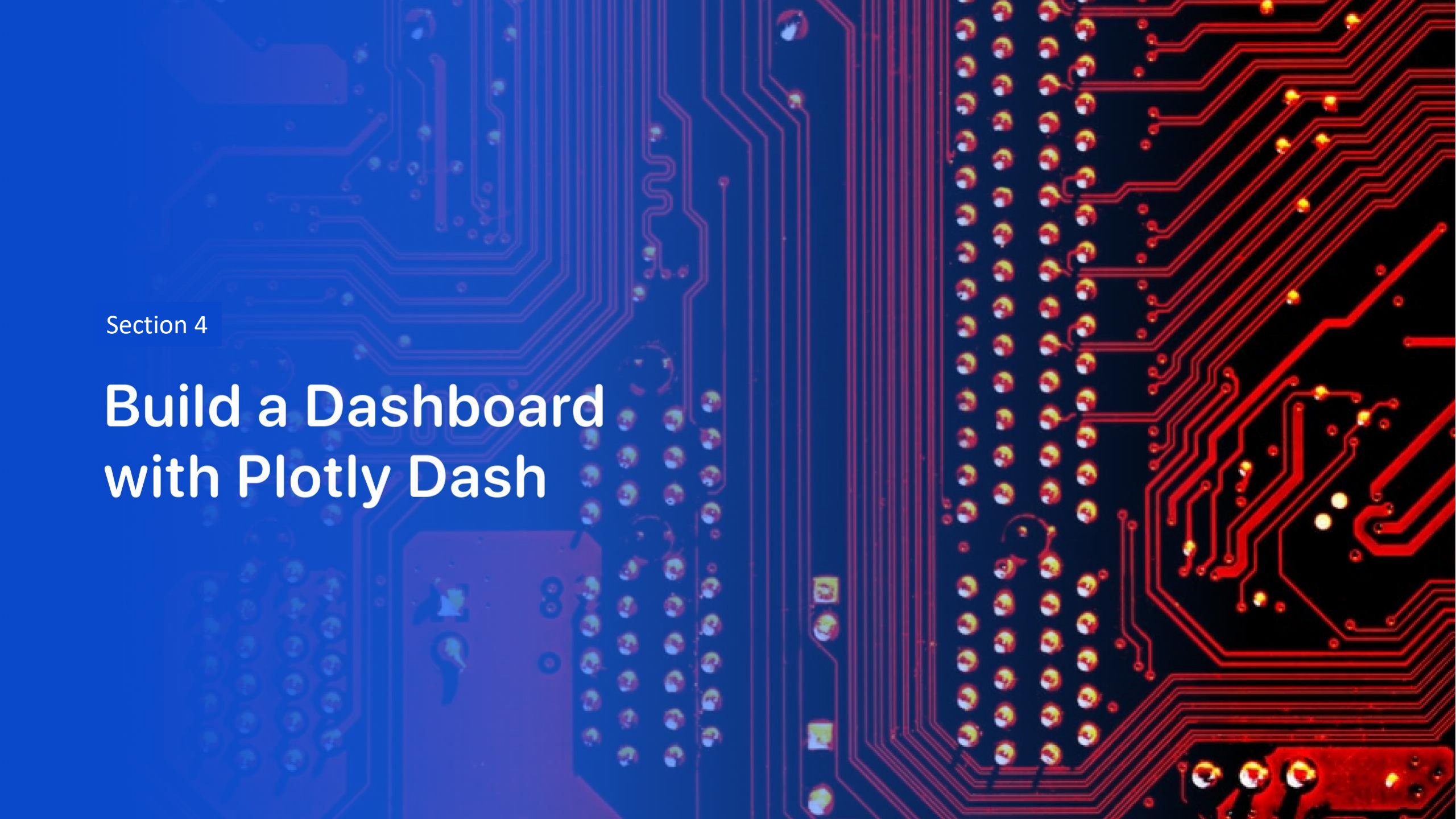
Launch Outcomes

- As we can see the numbers on the launch sites:



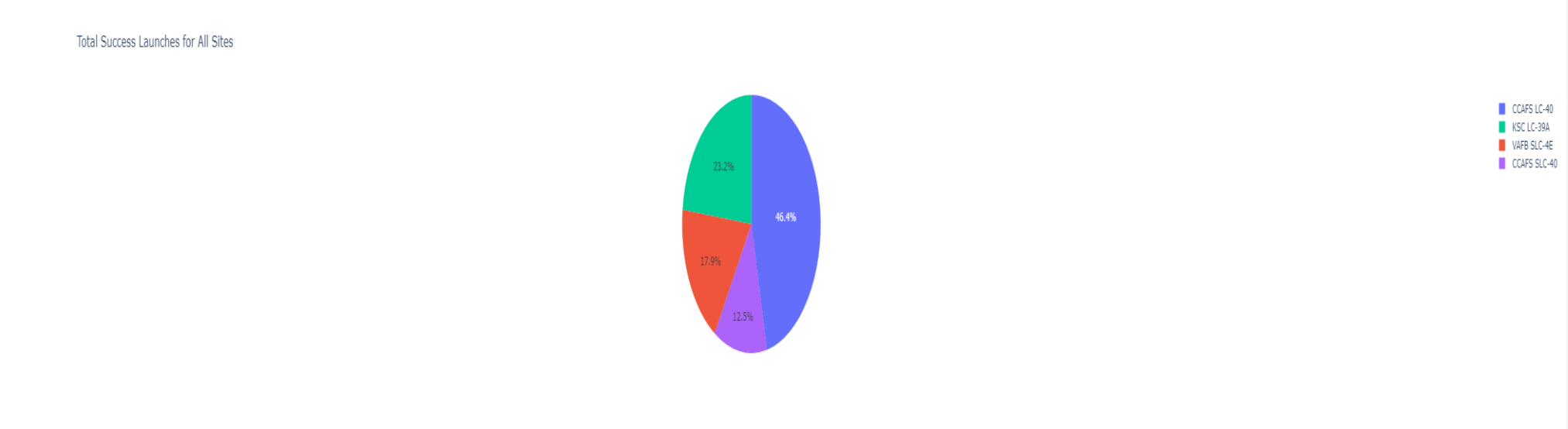
Distance Calculated

- As u can see the distance from the launch site to the ocean is 0.90 km!


Section 4
Build a Dashboard with Plotly Dash

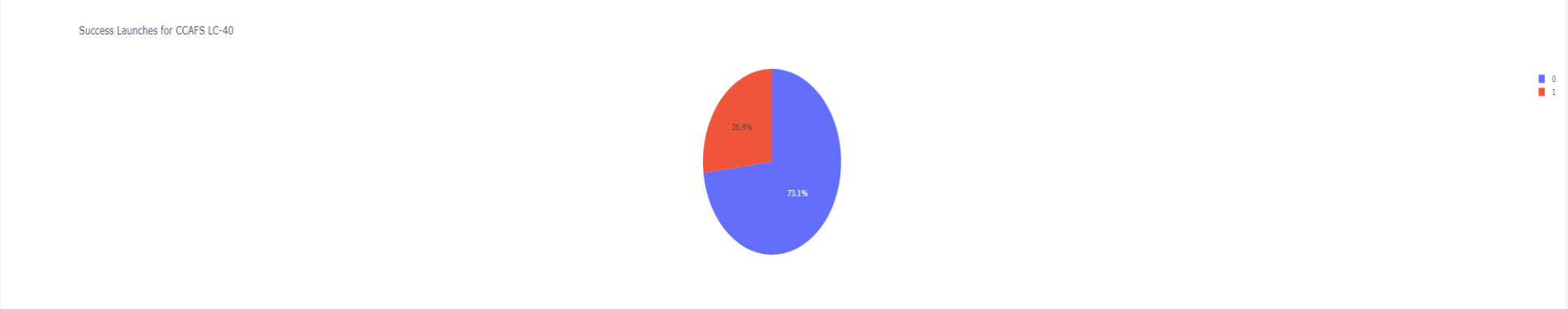
Launch Sites Success count

- For CCAFS LC-40 the success count is 46%, and for others we can see them in the following pie chart.



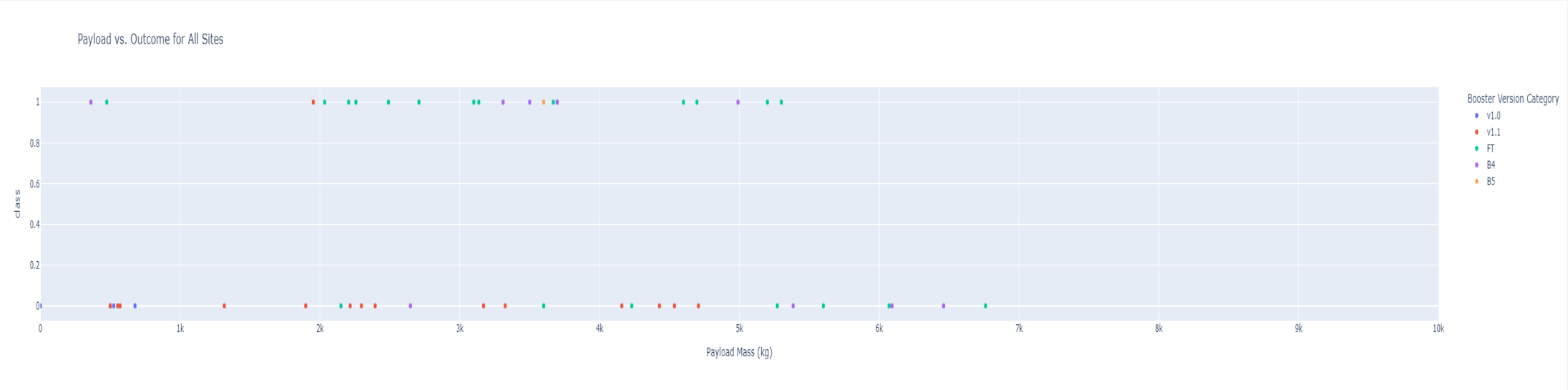
Highest Launch Site Success Ratio

- As we can see the most successful launch site is CCAFS LC-40 with Success rate 73.1%



Launch Sites Payload vs Outcome

- Booster Versions with different Payload Mass and different classes!
- v1.0 has 0 success rate! v1.1 has a very big failure rate!

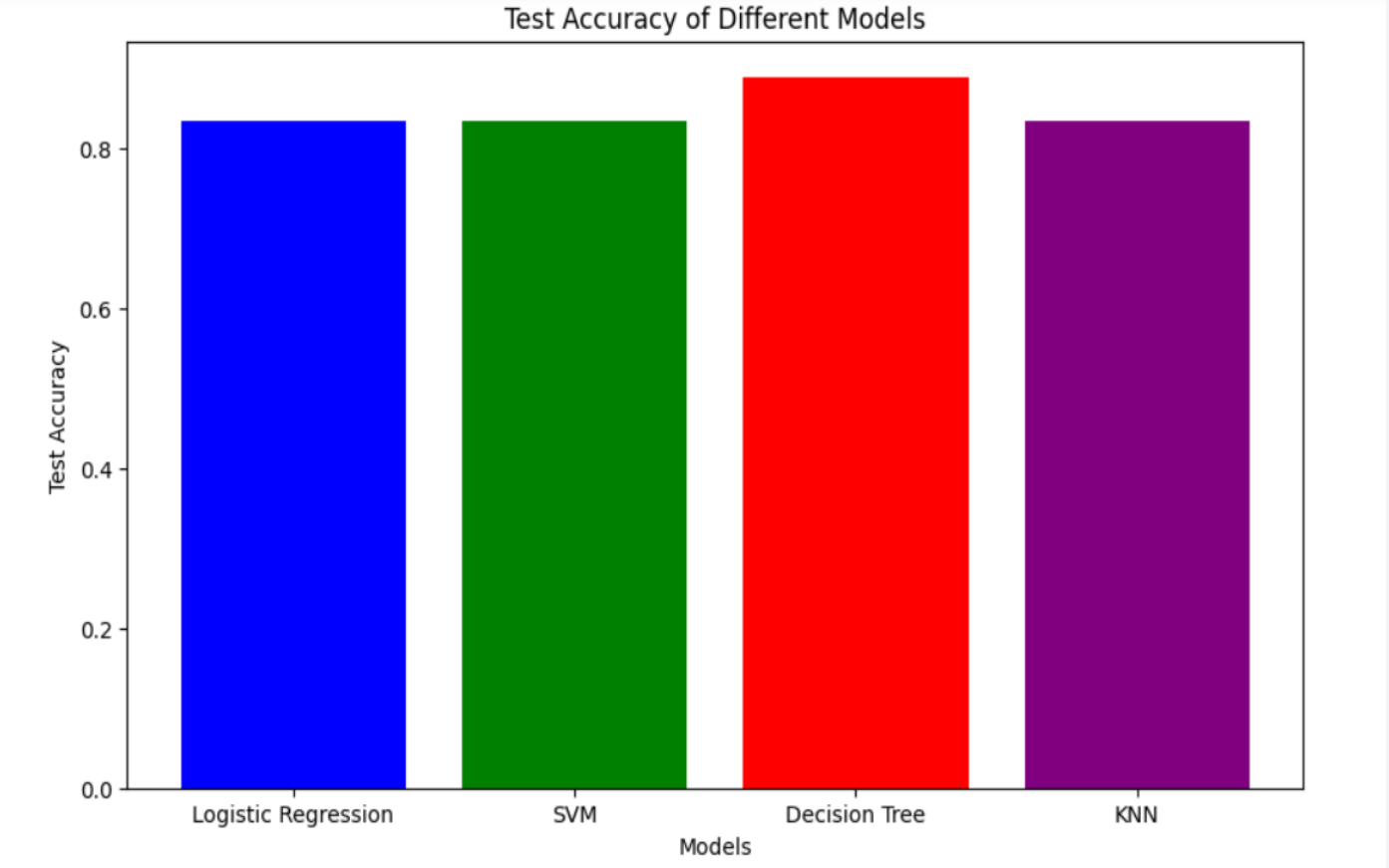


Section 5

Predictive Analysis (Classification)

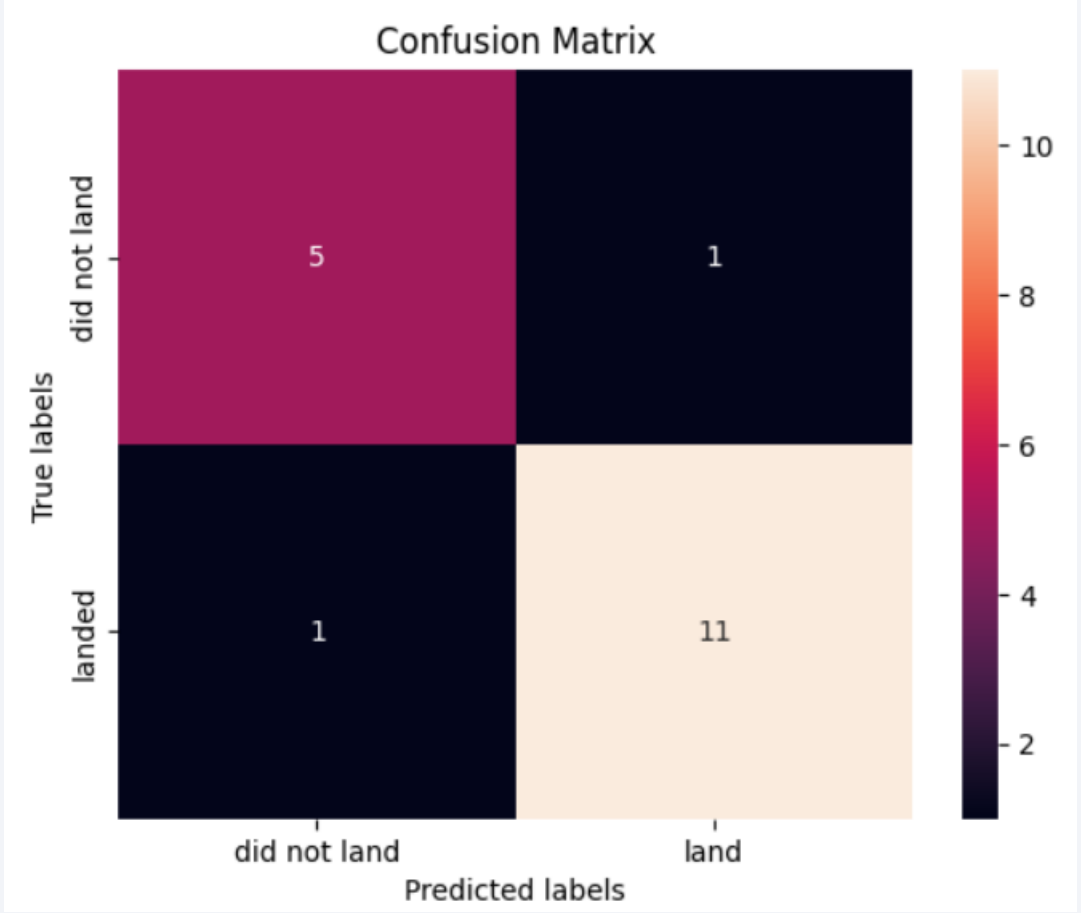
Classification Accuracy

- The highest model with accuracy is the Decision Tree model with accuracy 88.8%



Confusion Matrix

- The following is the Confusion Matrix of the decision tree model.
- It had made strong TP and FN compared to FP and TN.



Conclusions

- Landing outcomes into two classes of landing 0 (Failure), 1 (Successful).
- VAFB SLC 4E if the flightNumber > 50 it is completely class 1!
- We can conclude that the success rate for ES-L1, GEO, HEO, SSO orbits 100%, SO orbit 0%.
- The most successful launch site is CCAFS LC-40 with Success rate 73.1%
- Decision tree is the best classification model for this problem with accuracy 88.8%.

Appendix

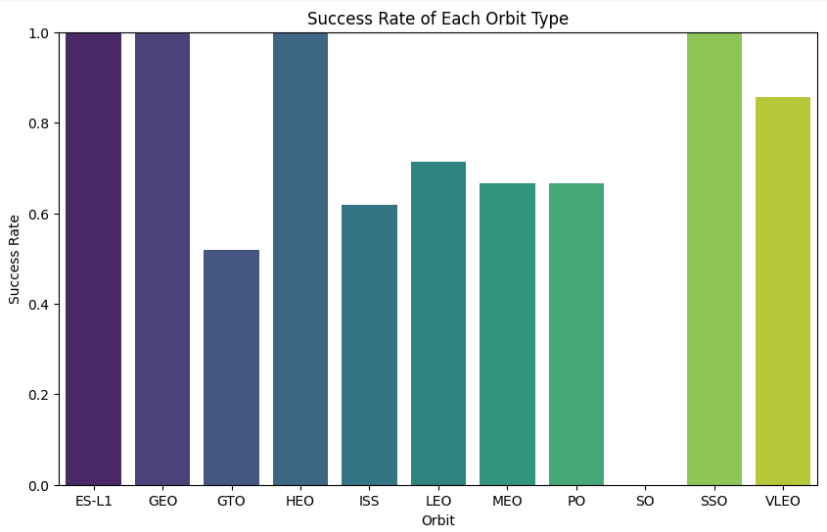
- You can find the full repo here : >>>Click<<<
- Some snippets:

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



Thank you!