IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Abdulmelik Dihye Zaim>
<01.11.2023>

Outline

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

Executive Summary

- The data obtained from SpaceX REST API and using Web Scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia. After that successful and failure class column labeled with data wrangling methods.
- Exploratory data analysis made with SQL, several visualizations, Folium maps, and Plotly dashboards. According to the insights gathered from EDA, features are selected. 4 machine learning models are created with GridSearchCV to find best model. Models are evaluated and compared with barplots.
- All models had same %83 test accuracy.

Introduction

- A company named Space Y wants to compete with Space X, so we are examining the success and failure of the launches executed by Space X.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected from SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - Class labels are created from Outcome column.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - GridSearchCV used to try different parameters.

Data Collection

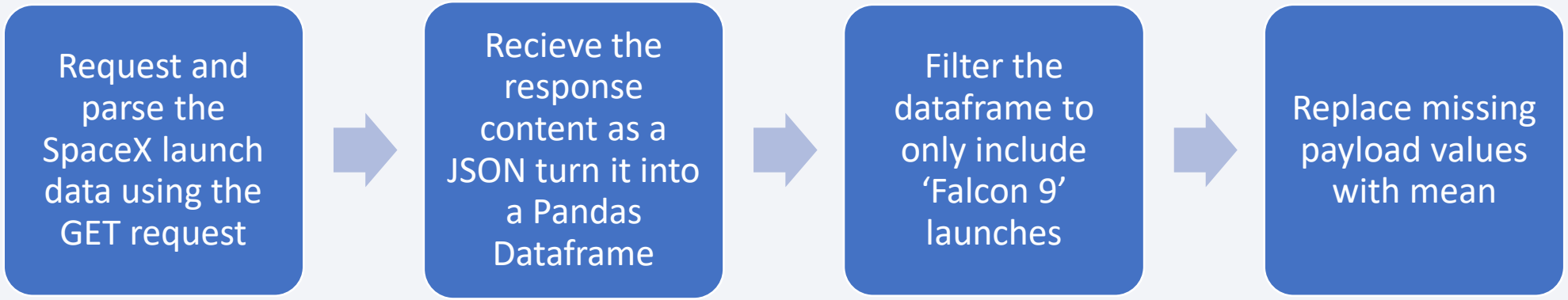
- Data collected from 2 different sources:

1) SpaceX REST API

2) Web Scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

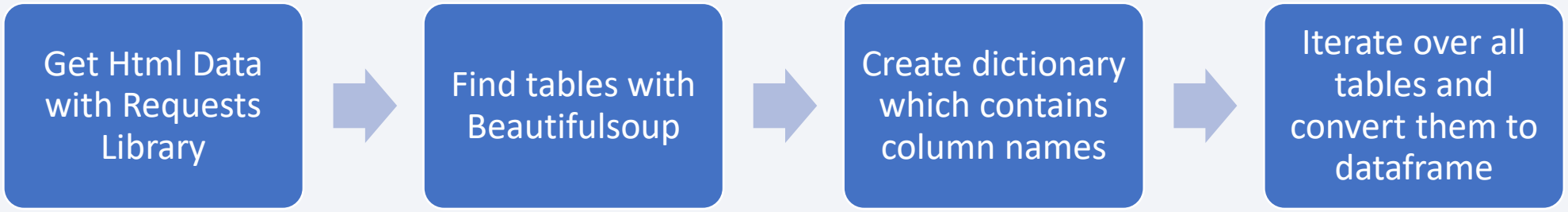
Data Collection – SpaceX API

- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/1_1_jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/1_2_jupyter_labs_webscraping.ipynb



Data Wrangling

- Class labels are created from **Outcome** column which includes the knowledge of successfully or unsuccessfully landing.
- False ASDS, False Ocean, False RTLS, None ASDS, None None are labeled as 0.
- True ASDS, True RTLS, True Ocean labeled as 1.
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/1_3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/2_2_jupyter-labs-eda-dataviz.ipynb
- Important features and relation between features are examined with EDA.

EDA with SQL

- With several queries, data examined in different aspects and it fascilitates to easier to understand and evaluate data.

- Queries are in the link below:
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/2_1_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

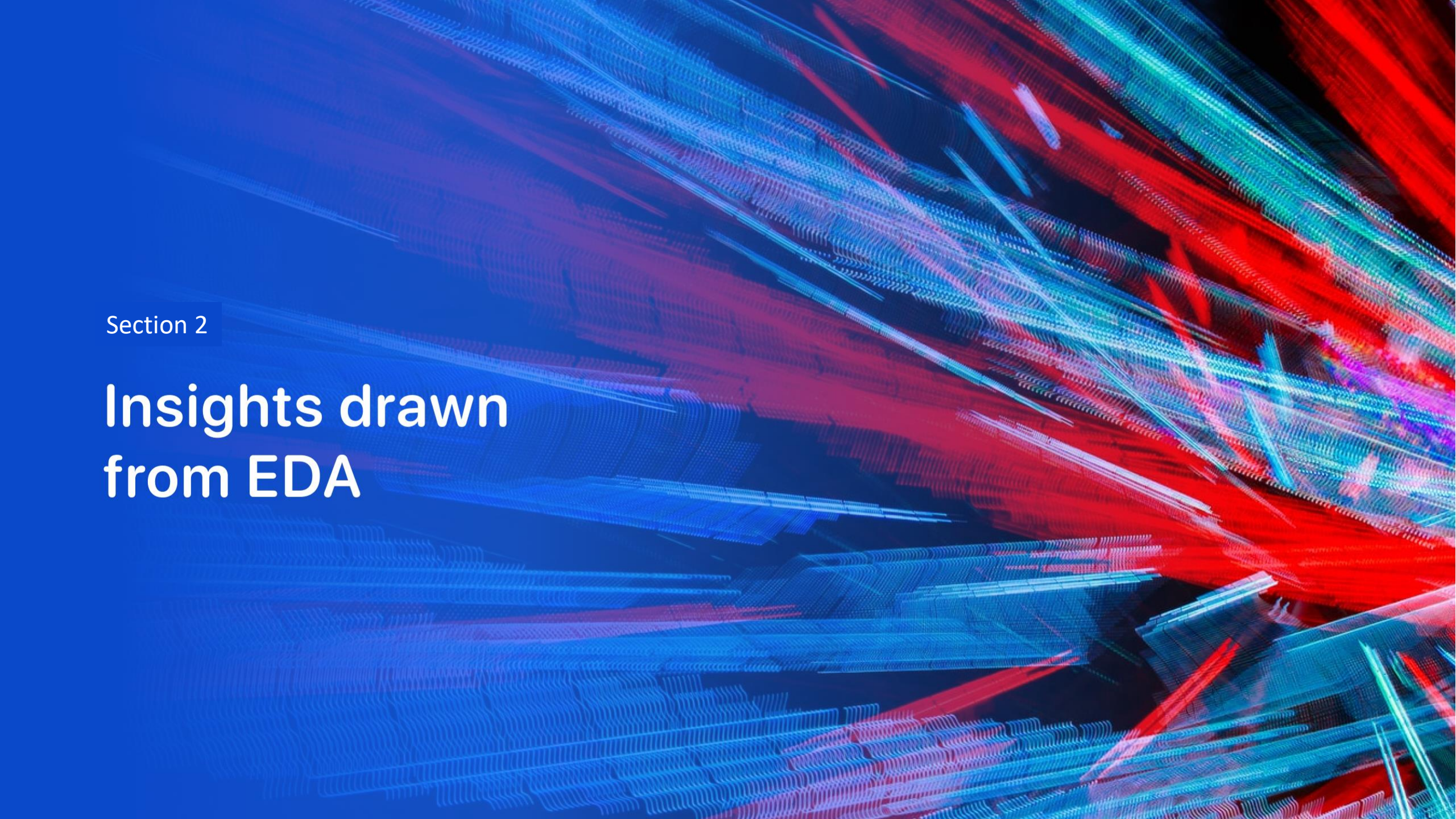
- Launch sites are circled and for each launch sites, successful and unsuccessful launches are marked.
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/3_1_launch_site_location.ipynb

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
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/3_2_spacex_dash_app.py

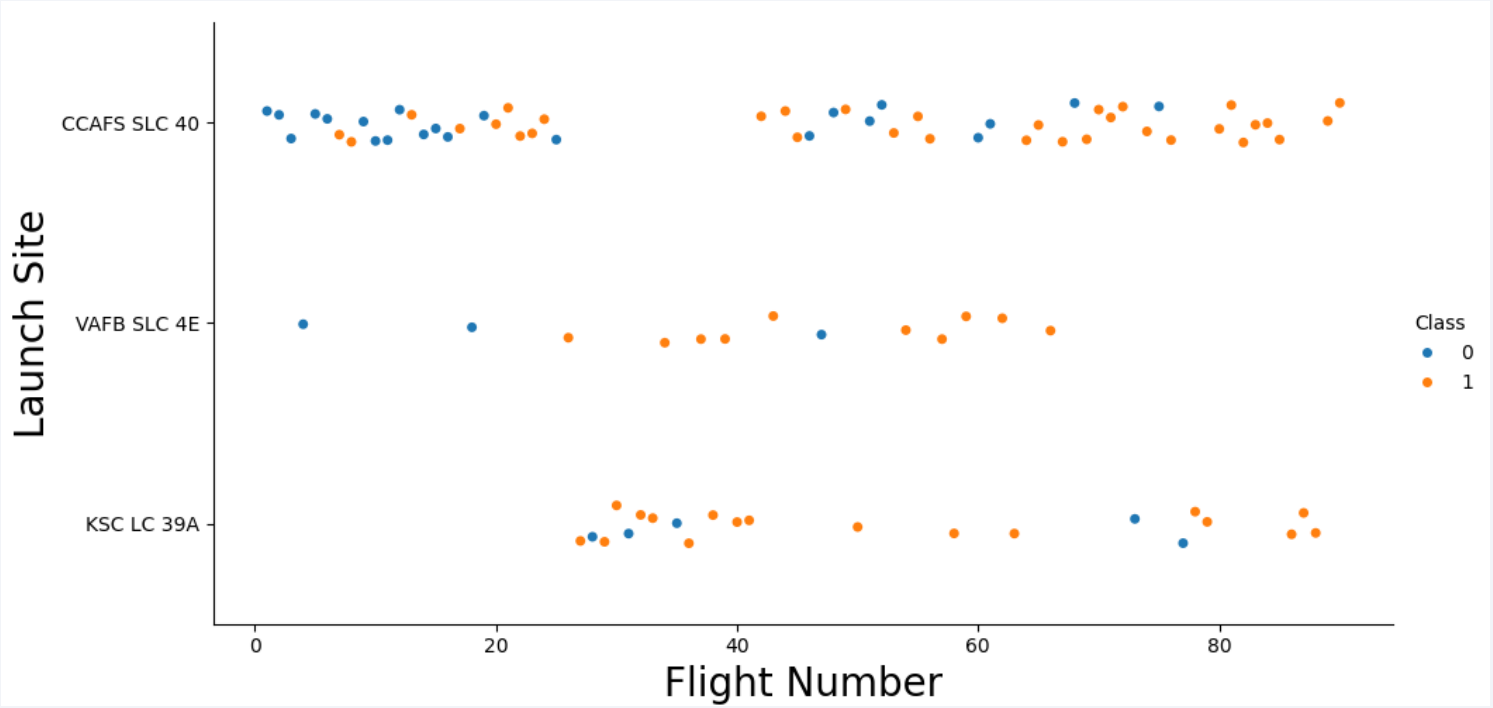
Predictive Analysis (Classification)

- 4 models (Logistic Regression, SVM, Decision Tree, KNN) are created with GridSearchCV. Best models are selected.
- Models are compared according to their test score.
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/4_SpaceX_Machine_Learning_Prediction.ipynb


Section 2
Insights drawn from EDA

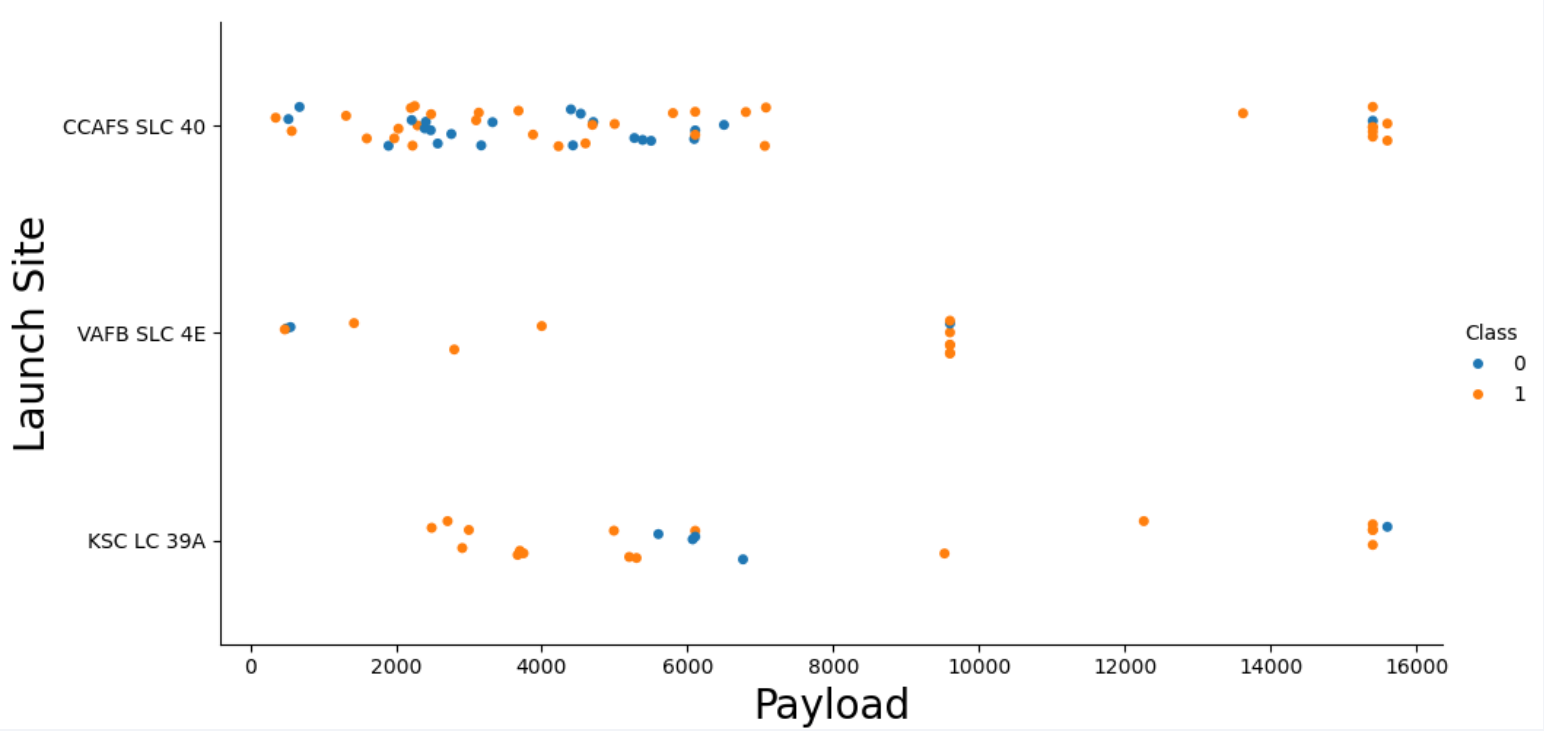
Flight Number vs. Launch Site

- Figure suggests that as flight number increases, the successful landings increases.



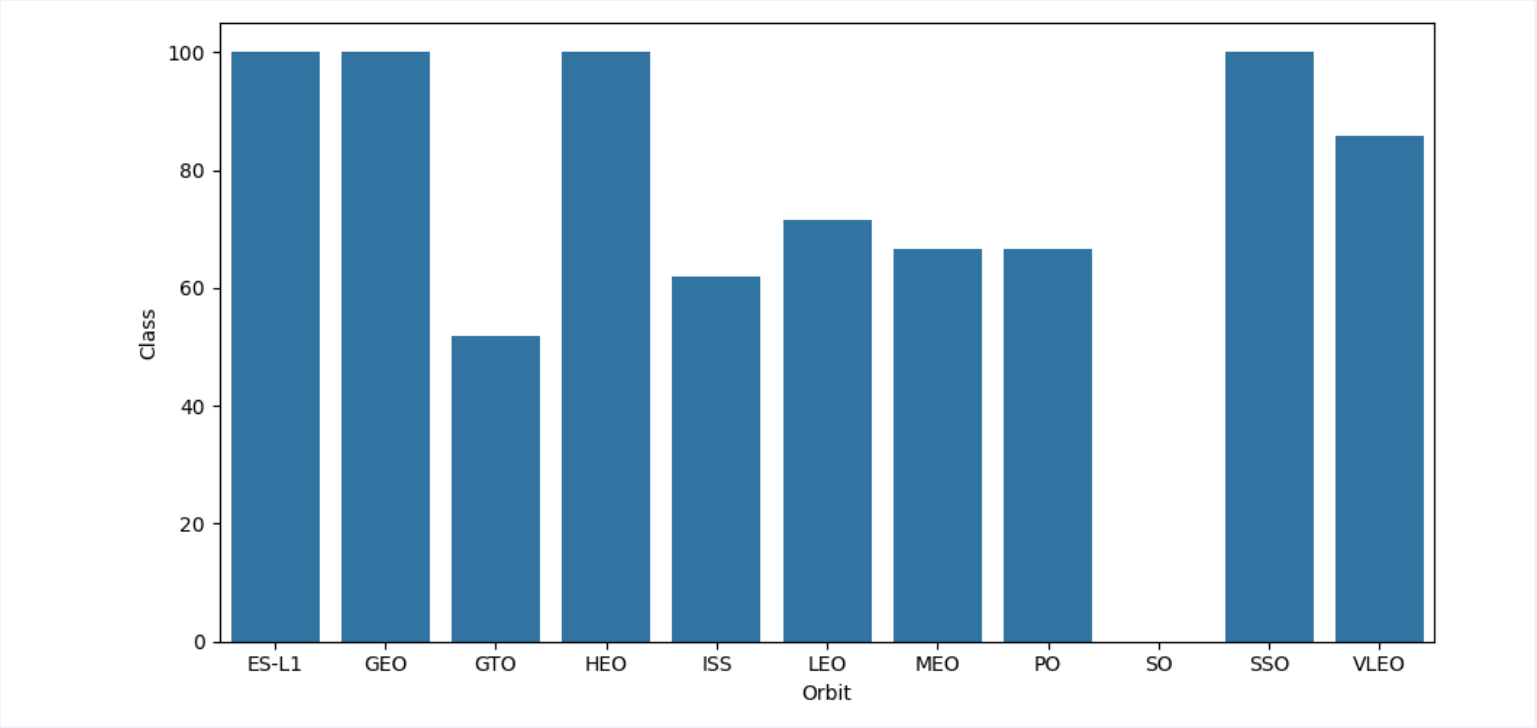
Payload vs. Launch Site

- When payload increases, successful landings increase.
- VAFB-SLC launchsite there are no rockets launched for mass greater than 10000.



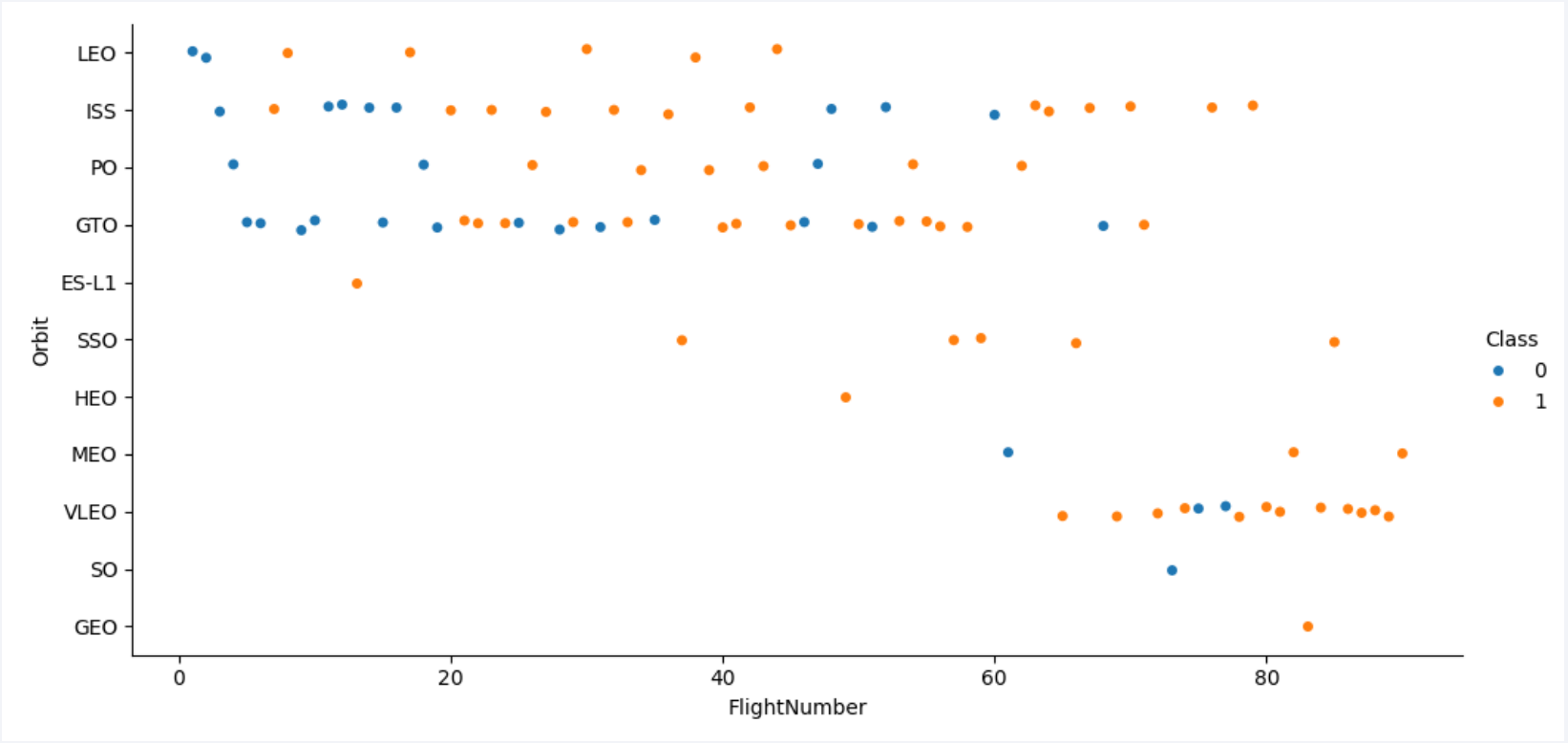
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO are the most successful orbits.
- SO is the least.



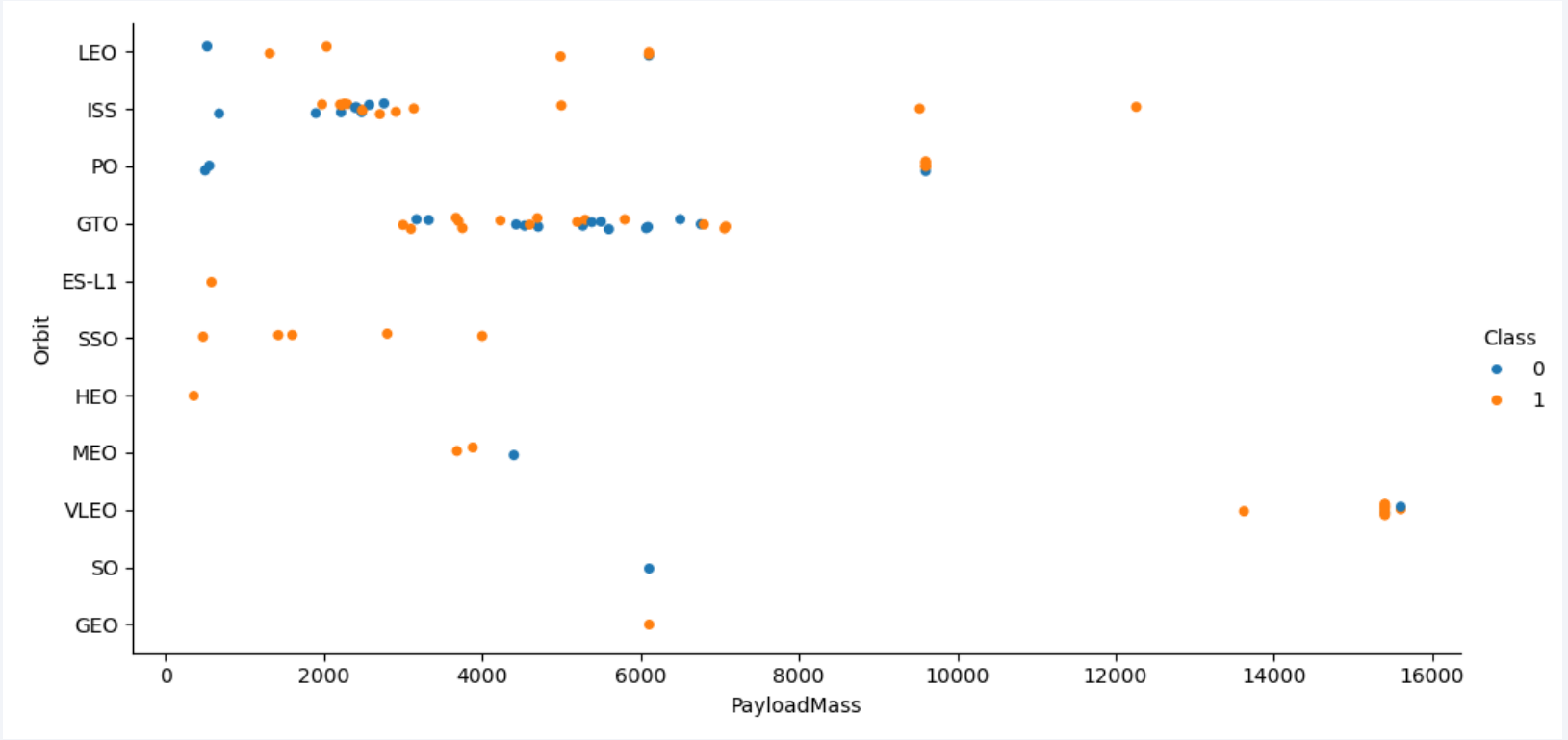
Flight Number vs. Orbit Type

- Until 60th flight, some orbits didn't targeted.
- As flight number increase, successful landing rate also increase.



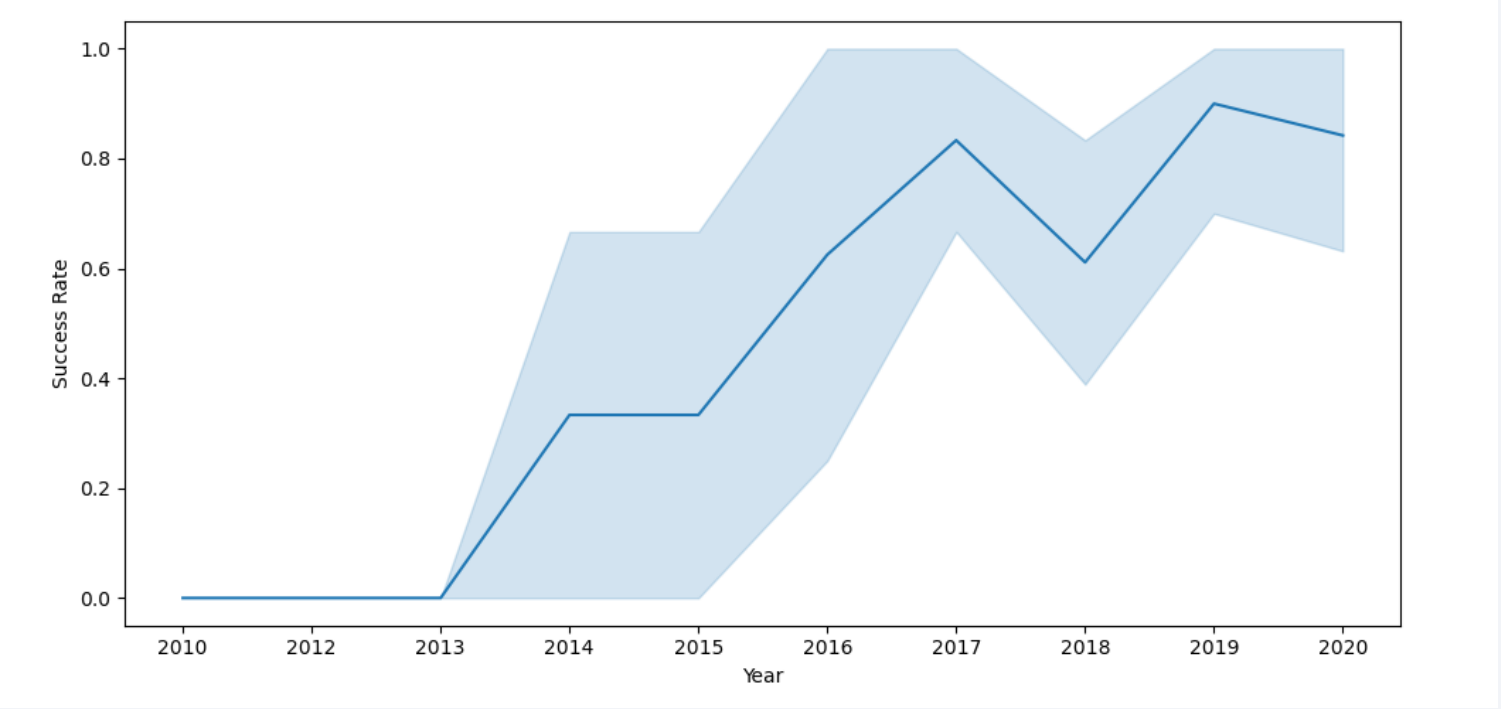
Payload vs. Orbit Type

- Some orbits have increased success rate according to increased mass, as an example VLEO.
- ES-L1, SSO, HEO, MEO had low pay load mass.



Launch Success Yearly Trend

- Success rate increased dramatically in years between 2013-2014, 2015-2017, 2018-2019.
- Figure suggest that there are 2 decline in 2017-2018 and 2019-2020



All Launch Site Names

- We obtain each unique launch site.

Task 1

Display the names of the unique launch sites in the space mission

In [28]:
```
%sql select distinct LAUNCH_SITE from SPACEXTABLE
```

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

Out[28]:

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

Launch Site Names Begin with 'CCA'

- We get 5 records where launch sites begin with `CCA`

Task 2

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

In [10]:
```
%sql select * from SPACEXTABLE where launch_site like 'CCA%' limit 5
```

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

Out[10]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	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 calculated.

Task 3

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

In [11]:
```
%sql select sum(payload_mass__kg_) as sum from SPACEXTABLE where customer like 'NASA (CRS)'
```

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

Out[11]:

sum
45596

Average Payload Mass by F9 v1.1

- Here the average payload mass carried by booster version F9 v1.1 calculated.

Task 4

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

In [12]:
```
%sql select avg(payload_mass__kg_) as Average from SPACEXTABLE where booster_version like 'F9 v1.1%'
```

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

Out[12]:

Average
2534.6666666666665

First Successful Ground Landing Date

- Here there is the dates of the first successful landing outcome on ground pad

Task 5

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

Hint:Use min function

In [14]:
```
%sql select min(date) as Date from SPACEXTABLE where mission_outcome like 'Success'
```

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

Out[14]:

Date
2010-04-06

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

In [19]:
```
%sql select booster_version from SPACEXTABLE where (mission_outcome like 'Success') AND (payload_mass__kg_ BETWEEN 4000 AND
```

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

Out[19]:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

In [20]:
```
%sql SELECT mission_outcome, count(*) as Count from SPACEXTABLE GROUP by mission_outcome ORDER BY mission_outcome
```

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

Out[20]:

Mission_Outcome	Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the 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

In [22]:
```
maxm = %sql select max(payload_mass__kg_) from SPACEXTABLE
maxv = maxm[0][0]
%sql select booster_version from SPACEXTABLE where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTABLE)
```

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

Out[22]:

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

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

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

- 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

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.

In [27]:
```
%sql select landing_outcome, count(*) as count from SPACEXTABLE where Date >= '2010-06-04' AND Date <= '2017-03-20' GROUP by
```

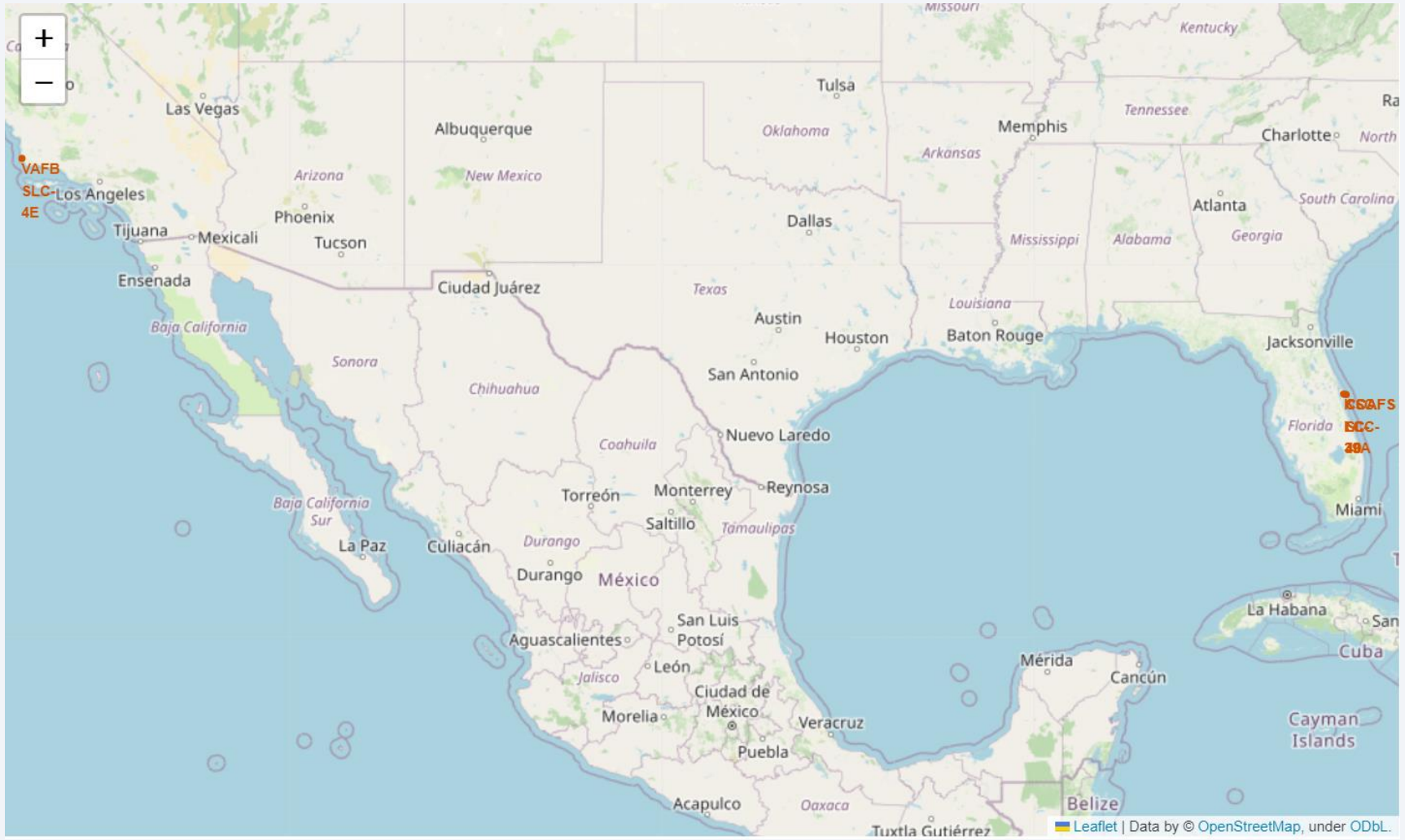
```
* sqlite:///my_data1.db
Done.
```

Out[27]:

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

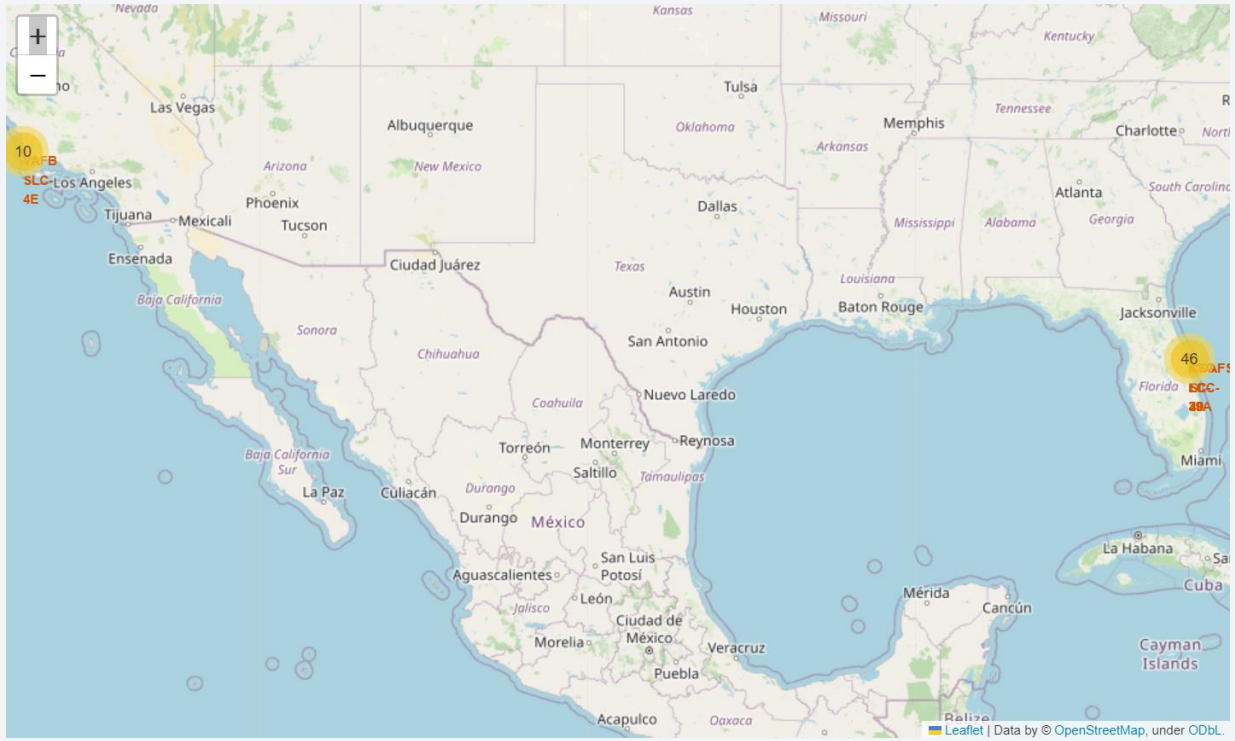

Section 3
Launch Sites
Proximities Analysis

Launch Site Locations

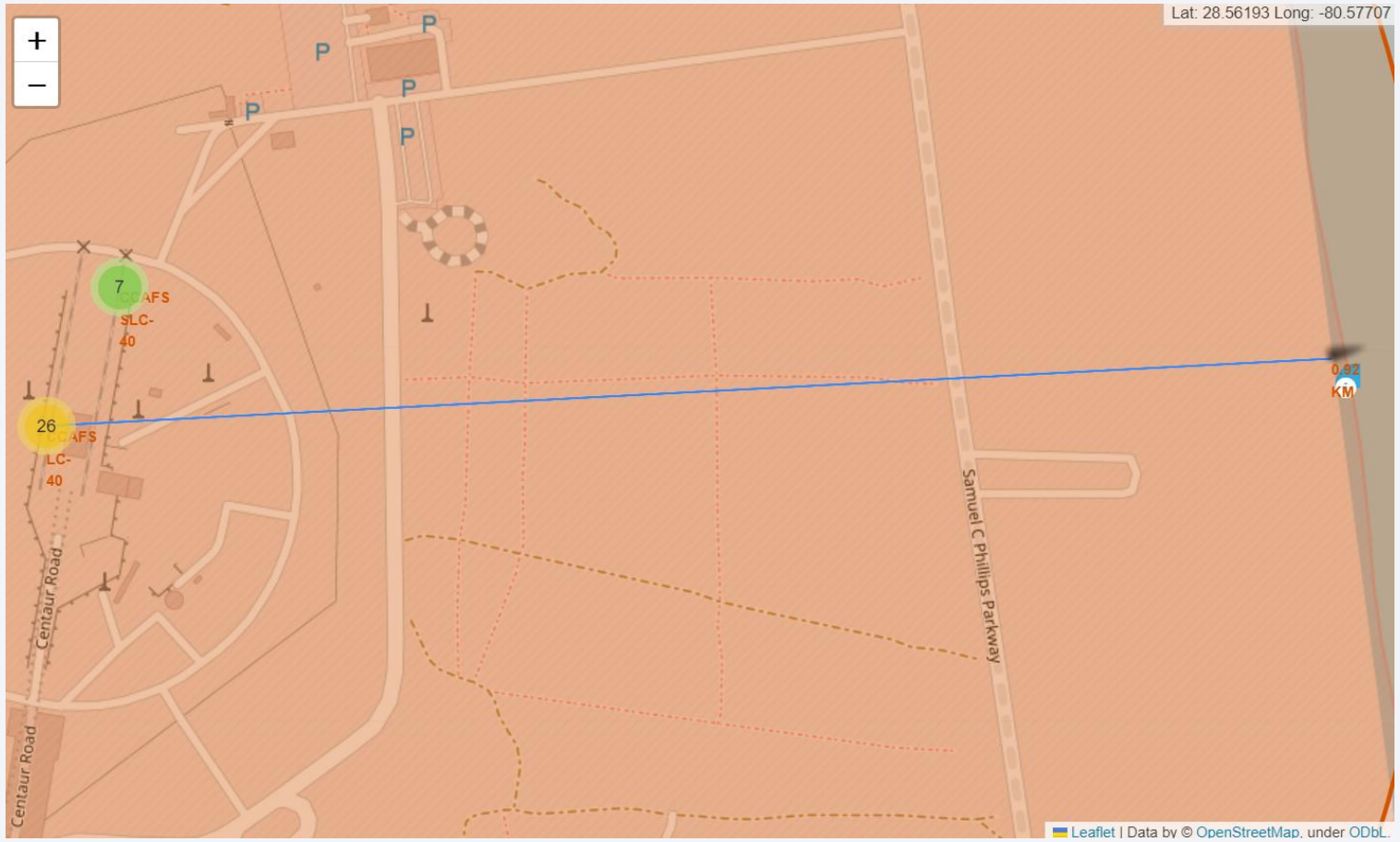


- Launch site locations are given in the map.
- https://github.com/d-zaim/Applied-Data-Science-Capstone/blob/main/3_1_launch_site_location.ipynb

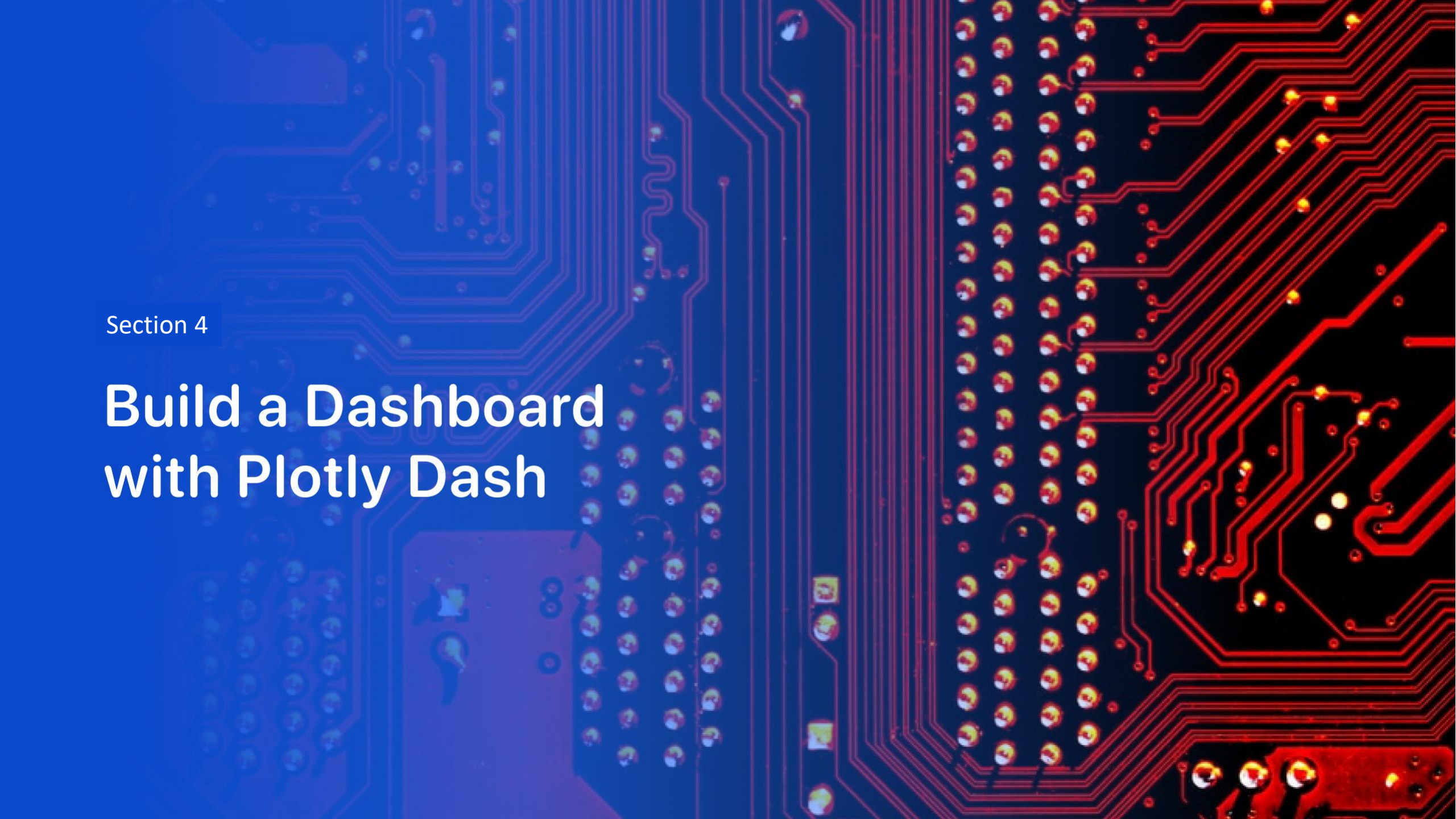
Successed/Failed Launches



Location Proximities

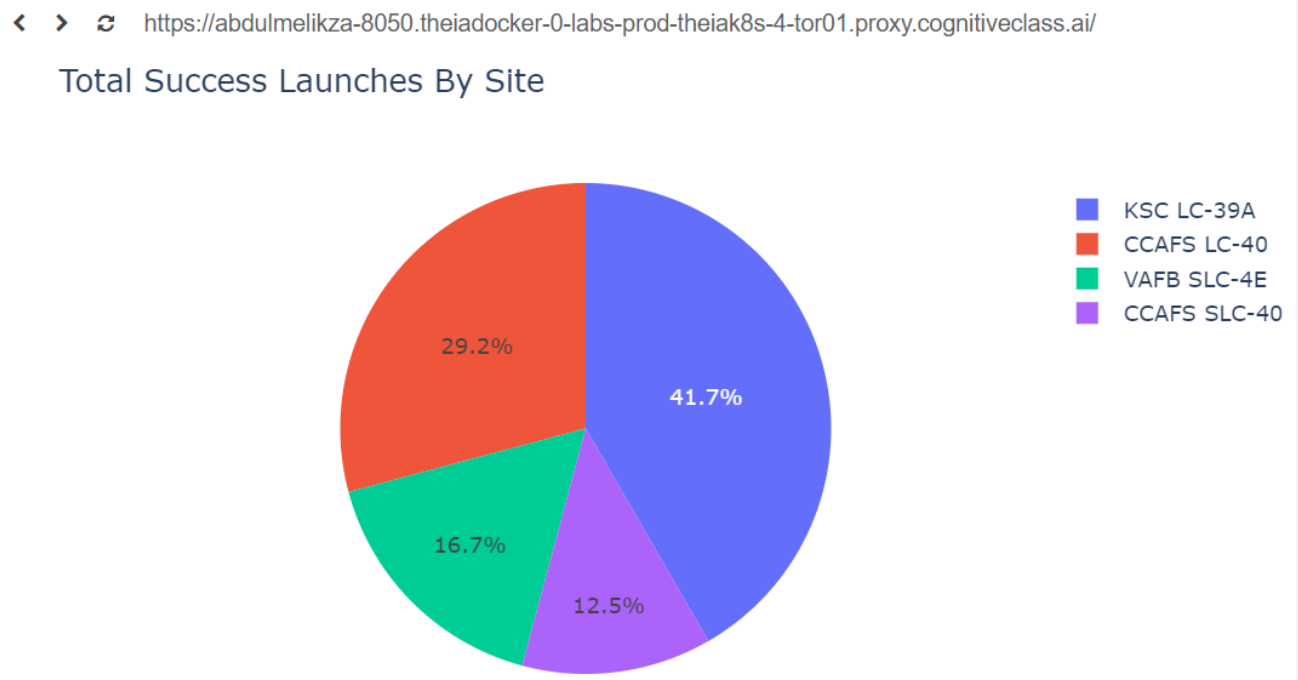


- In this map, the distance between launch site and coastline demonstrated and measurement added.


Section 4
Build a Dashboard with Plotly Dash

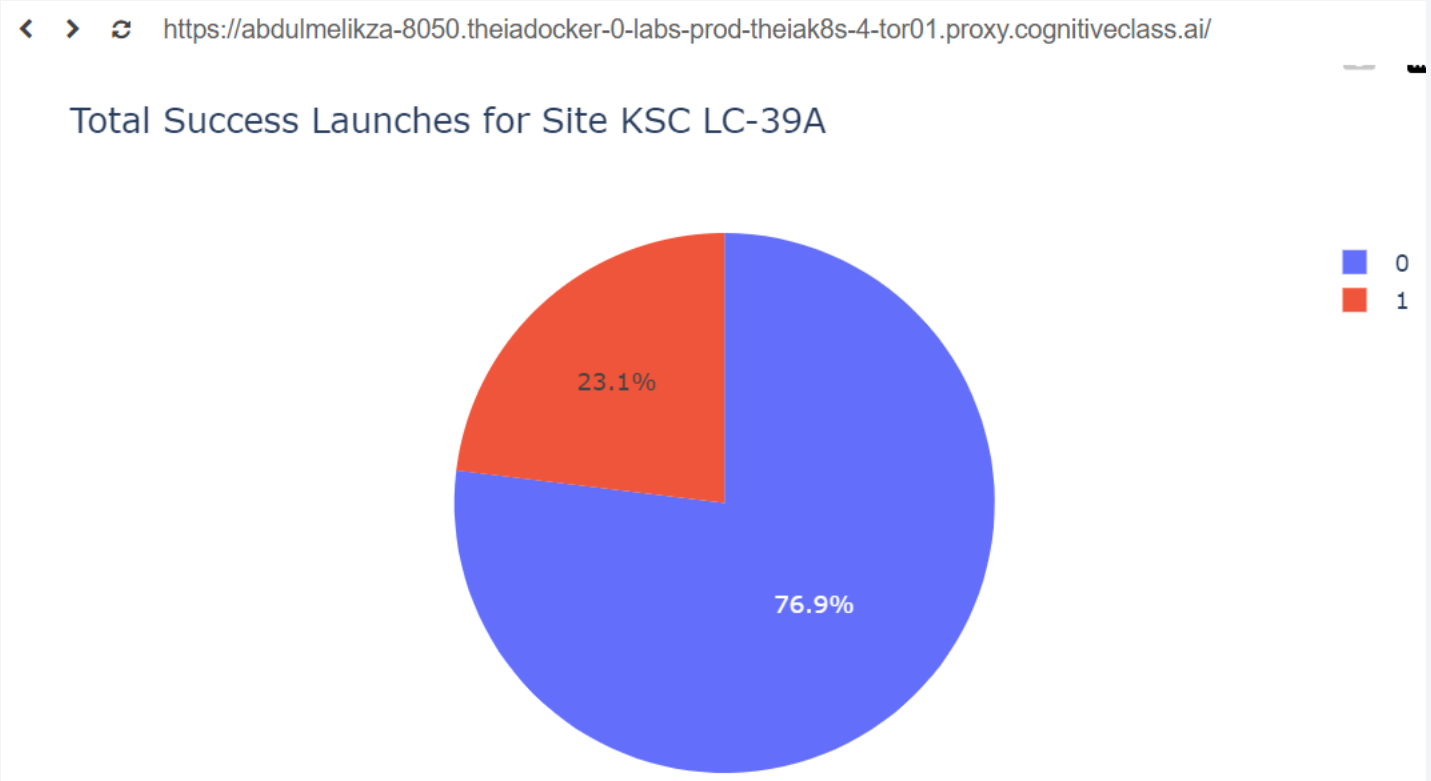
Total Success Launches By Site

- The most successful launches are at KSC LC-39A, and the least successful launch site is CCAFS SLC-40.



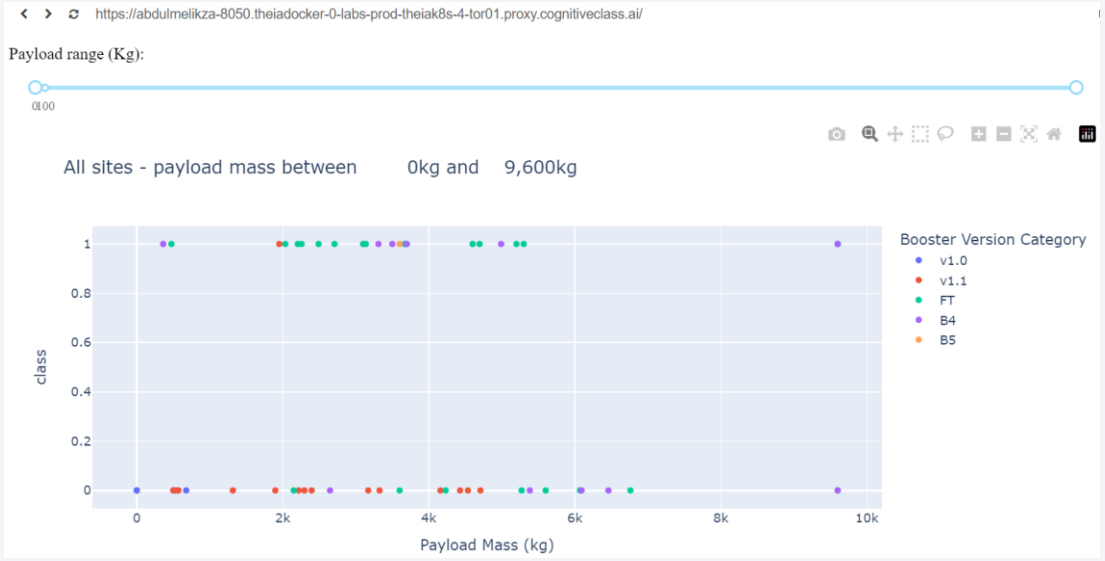
The Site Which Has Highest Launch Success Ratio

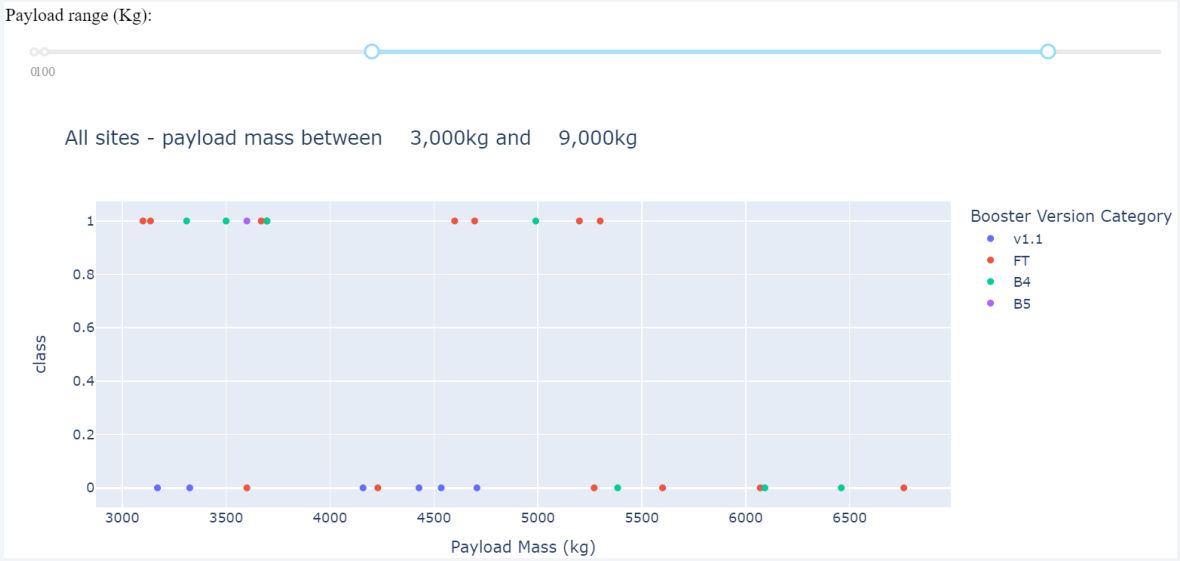
- As demonstrated in previous slide, the highest launch success ratio is at KSC LC-39A launch site.



Payload Mass vs Launch Outcome

- Payload mass vs launch outcome are demonstrated in below with different payload mass parameters.



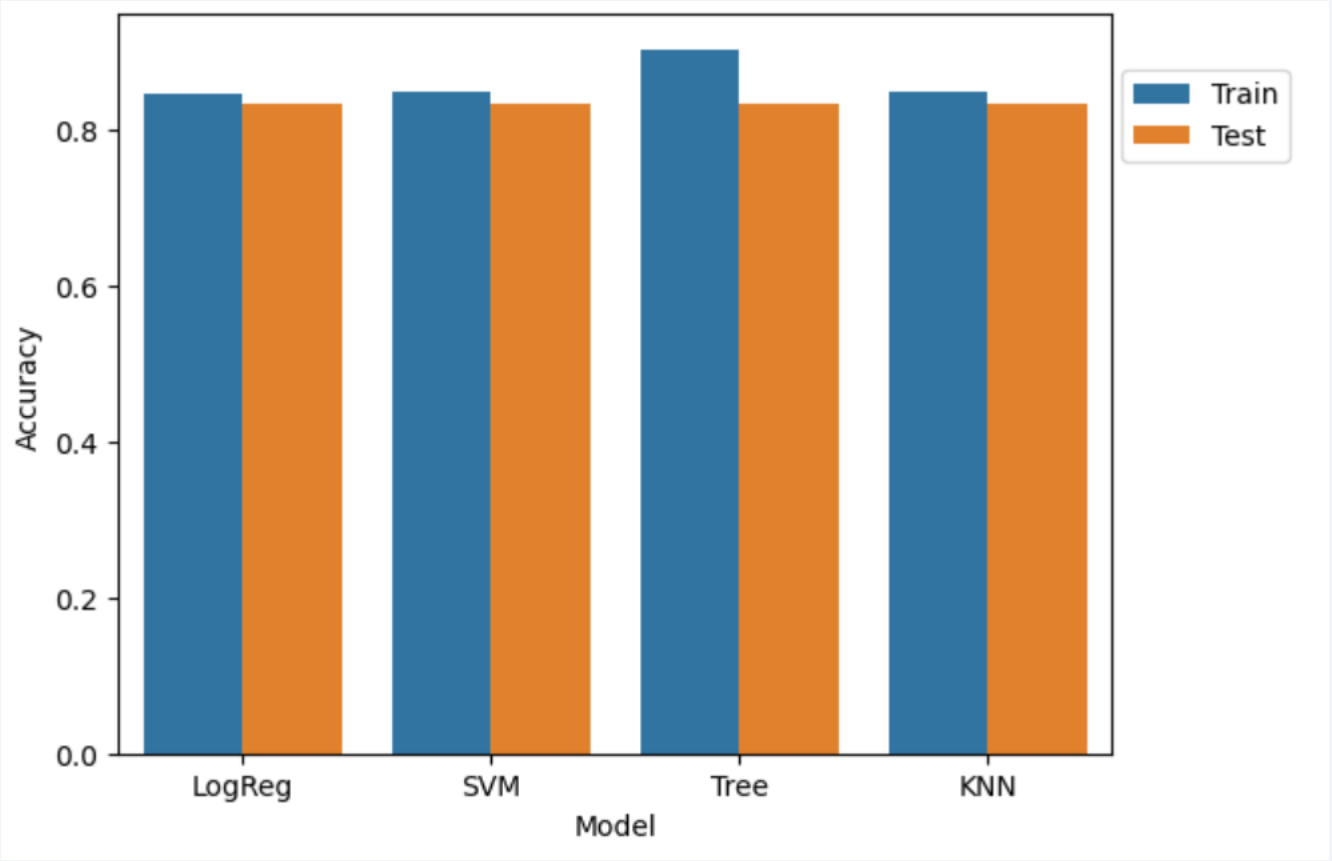


Section 5

Predictive Analysis (Classification)

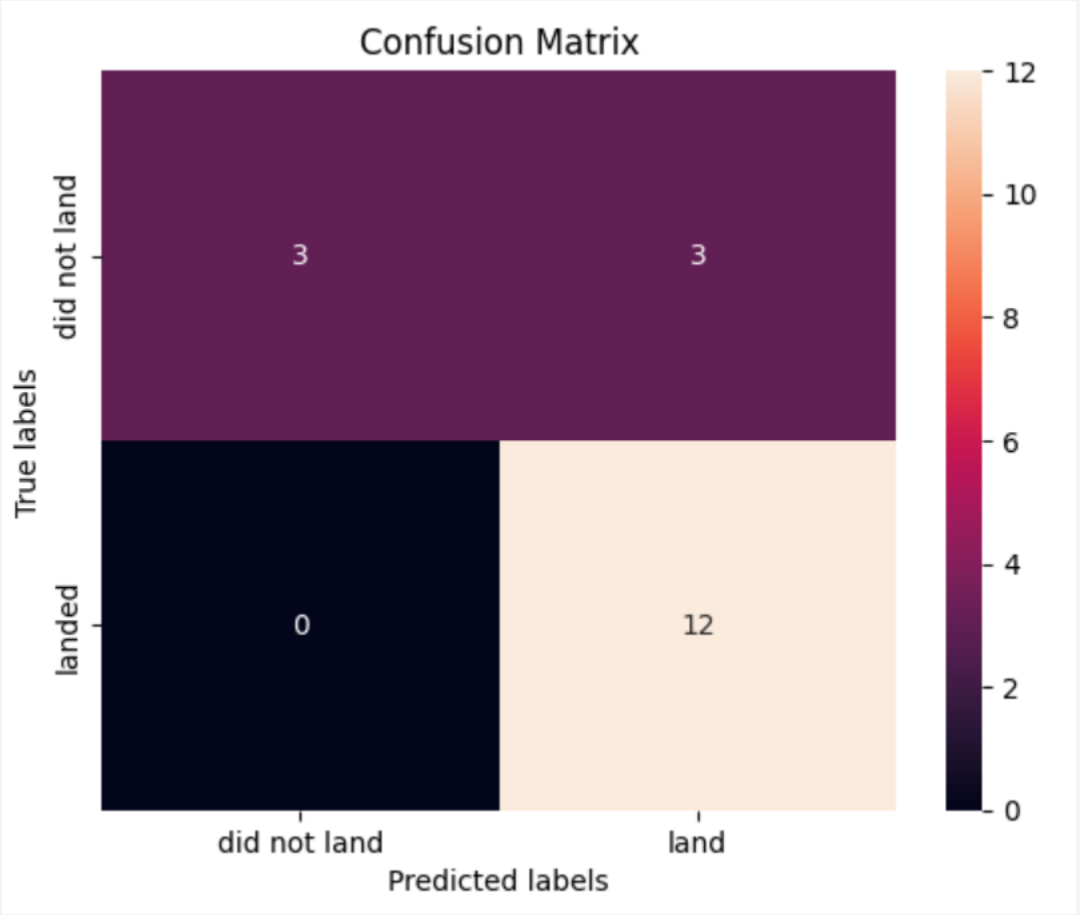
Classification Accuracy

- Figure demonstrates that Decision Tree classifier has higher train accuracy, on the other hand all of the models have the same test accuracy which is %83.



Confusion Matrix

- As it is explained the previous slide, every model has the same confusion matrix. It may be caused by the data.



Conclusions

- Decision Tree classifier has higher train accuracy.
- All models have the same test accuracy.

Appendix

- https://github.com/d-zaim/Applied-Data-Science-Capstone/tree/main

Thank you!