IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Punam Patil>
<29-03-2022>

Outline

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

Executive Summary

- Summary of methodologies

1. Data Collections
2. Data Wrangling
3. EDA
4. Interactive Map with Folium
5. Plotly Dashboard via Python
6. Predictive Analysis

- Summary of all results

Introduction

- Project background and context

Several companies, including SpaceX have been making space travel possible for individuals. We believe an increasing number of commercial space travel will arrive in the near future. SpaceX have advertised its Falcon 9 rocket launches on its website. They claim that the Falcon 9 cost is much lower compared with other providers, since the first stage of them can be reused.

We will calculate the real cost of a successful launch based on the successful rate, average cost in specific situations when the first stage is reused.

- Problems you want to find answers

1. Variable Engineering: right features for successful rate;
2. Prediction: best condition for a successful launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

1. Describe how data sets were collected.

- Data Collection includes both API requests and web-scraping from a Wikipedia page.

- For SpaceX API, data includes: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused,Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

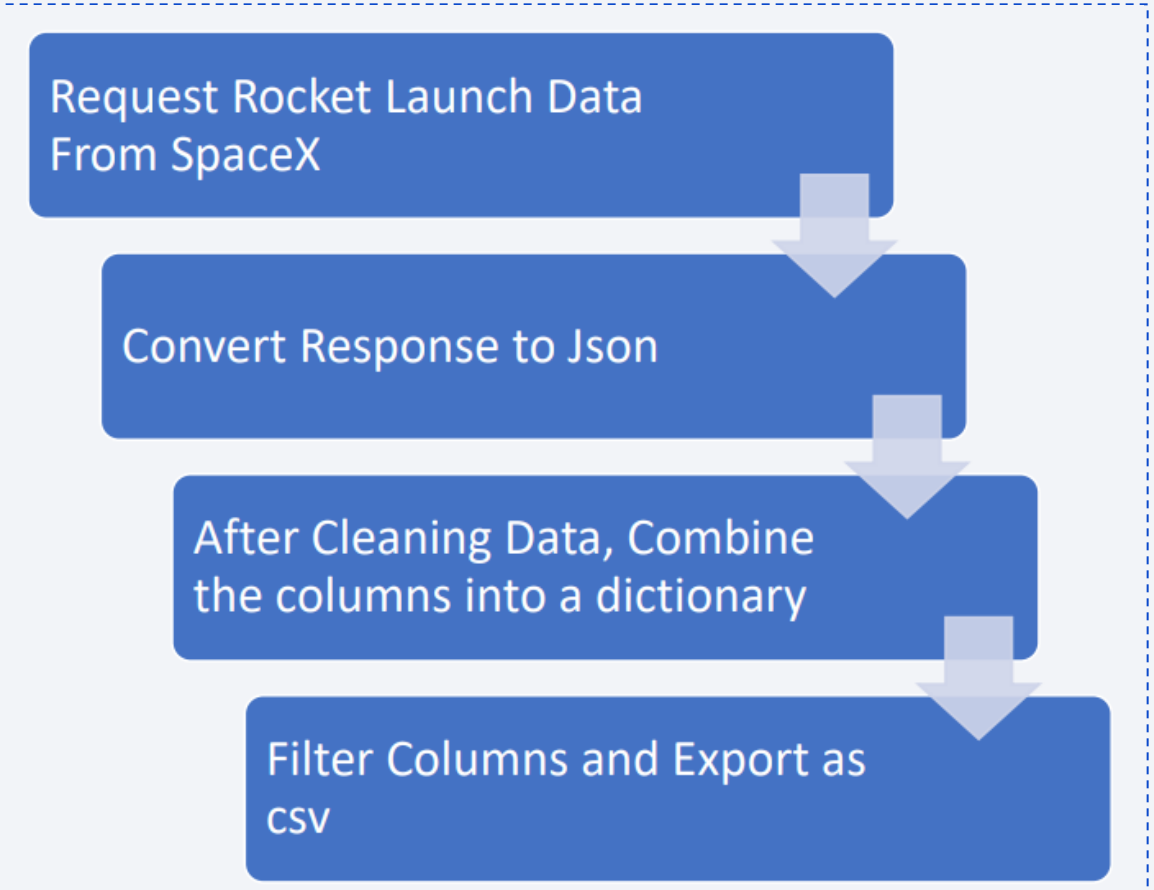
- For the web-scarping on Wikipedia page, data contains: Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

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

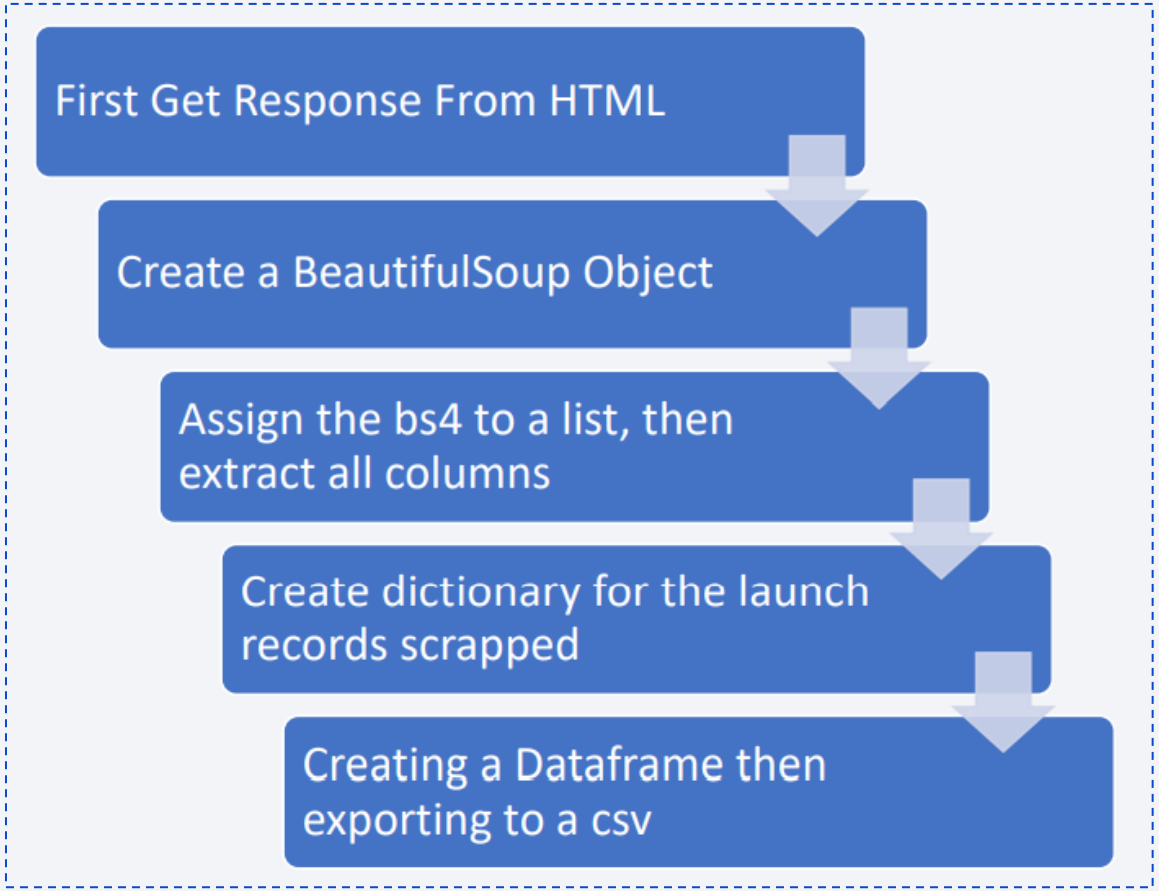
- Project/Space X Falcon 9 First Stage Landing Prediction.ipynb at main · Punampatil25/Project (github.com)



Data Collection - Scraping

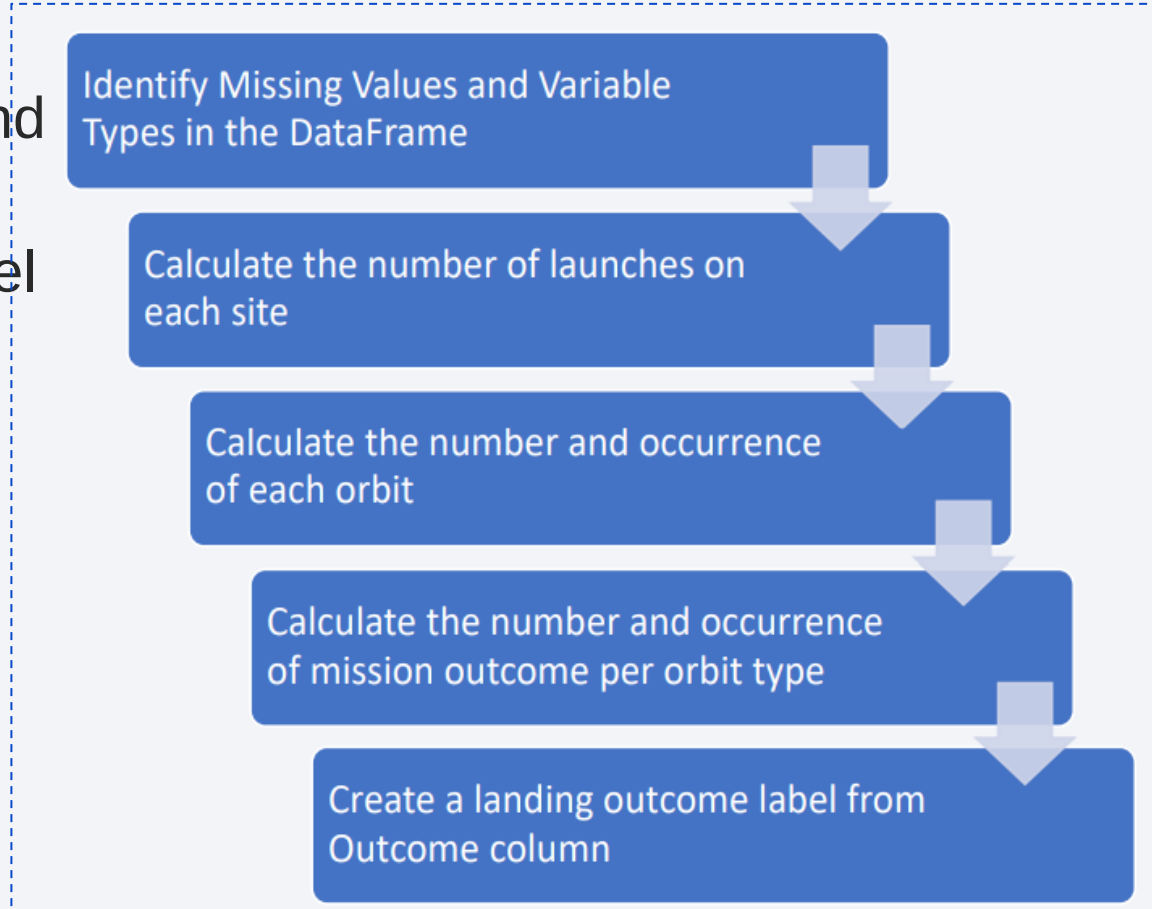
- Present your web scraping process using key phrases and flowcharts

- Space X Falcon 9 First Stage Landing Prediction Web scraping Falcon 9 and Falcon Heavy Launches Records - Jupyter Notebook



Data Wrangling

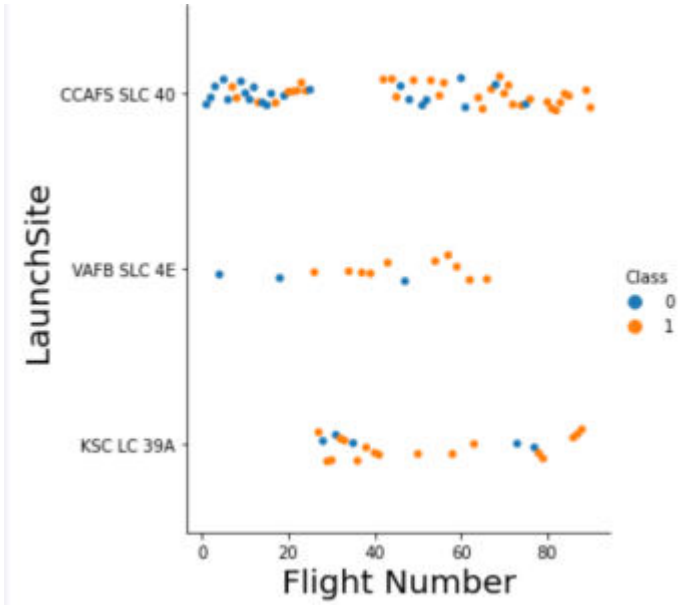
- Data wrangling is done with Pandas and NumPy.
- The values of the landing outcome label 'Class' are worked out with Python list comprehension.
- [Project/Space X Falcon 9 First Stage Landing Prediction (Lab 2 Data wrangling).ipynb at main · Punampatil25/Project (github.com)]

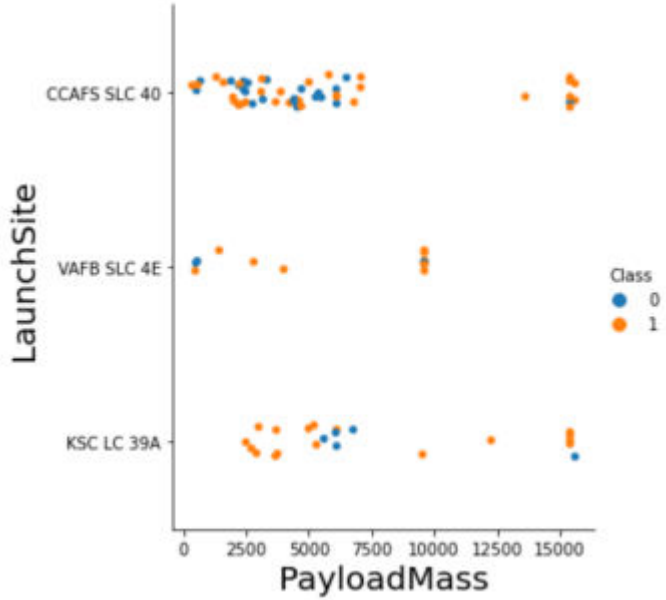


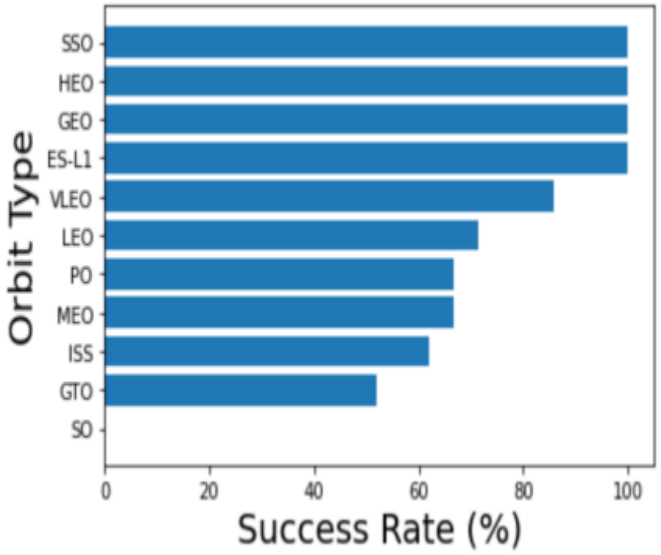
EDA with Data Visualization

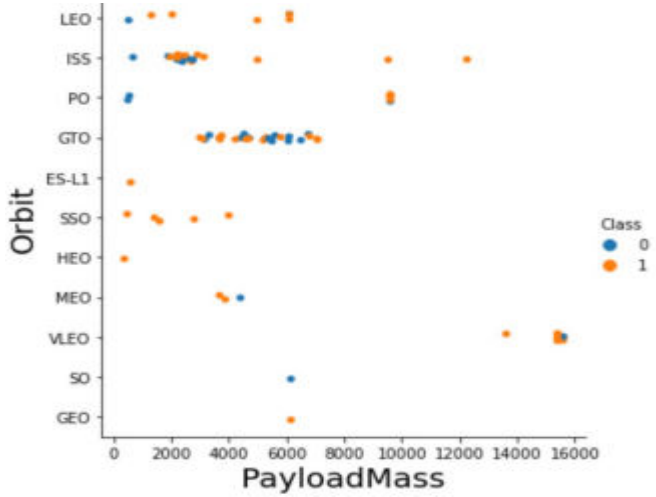
- We visualize the relationship between: Flight Number & Payload and Launch Site; success rate and orbit type; Flight Number & Payload and Orbit type; launch success yearly trend
- Project/SpaceX Falcon 9 First Stage Landing Prediction Assignment Exploring and Preparing Data.ipynb at main · Punampatil25/Project (github.com)

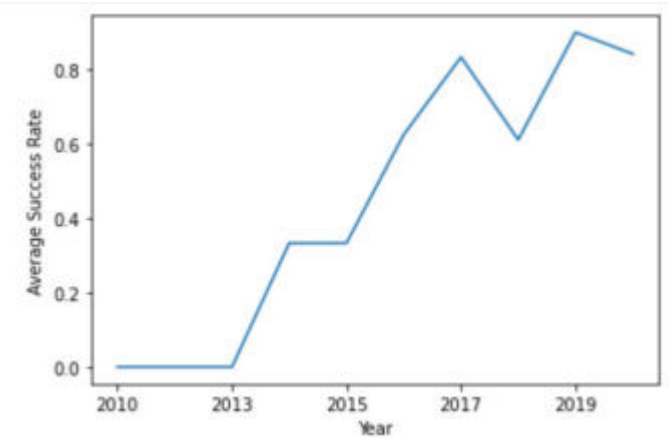
EDA with Data Visualization











EDA with SQL

- Loading the dataset into the corresponding table in a Db2 database, we execute:

1. Displaying the names of the unique launch sites in the space mission; 5 records where launch sites begin with the string ‘CCA’; the total payload mass carried by boosters launched by NASA (CRS); average payload mass carried by booster version F9 v1.1
2. Listing the date when the first successful landing outcome in ground pad was achieved
3. Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000, and the total number of successful and failure mission outcomes
4. Listing the names of the booster_versions which have carried the maximum payload mass, and the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
5. 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

- Project/SQL_EDA Notebook.ipynb at main · Punampatil25/Project (github.com)

Build an Interactive Map with Folium

- Objects created and added to a folium map:

1. Markers that show all launch sites on a map
2. Markers that show the success/failed launches for each site on the map
3. Lines that show the distances between a launch site to its proximities

- By adding these objects, following geographical patterns about launch sites are found:

1. Launch sites are in close proximity to railways
2. Launch sites are in close proximity to highways
3. Launch sites are in close proximity to coastline
4. Launch sites keep certain distance away from cities

- Project/Launch Sites Locations Analysis with Folium.ipynb at main · Punampatil25/Project (github.com)

Build a Dashboard with Plotly Dash

- The dashboard application contains a pie chart and a scatter point chart.
- *Pie chart*

1. For showing total success launches by sites
2. This chart can be selected to indicate a successful landing distribution across all launch sites or to indicate the success rate of individual launch sites.

- *Scatter chart*

1. For showing the relationship between Outcomes and Payload mass(Kg) by different boosters
2. Has 2 inputs: All sites/individual site & Payload mass on a slider between 0 and 10000kg
3. This chart helps determine how success depends on the launch point, payload mass,and booster version categories.

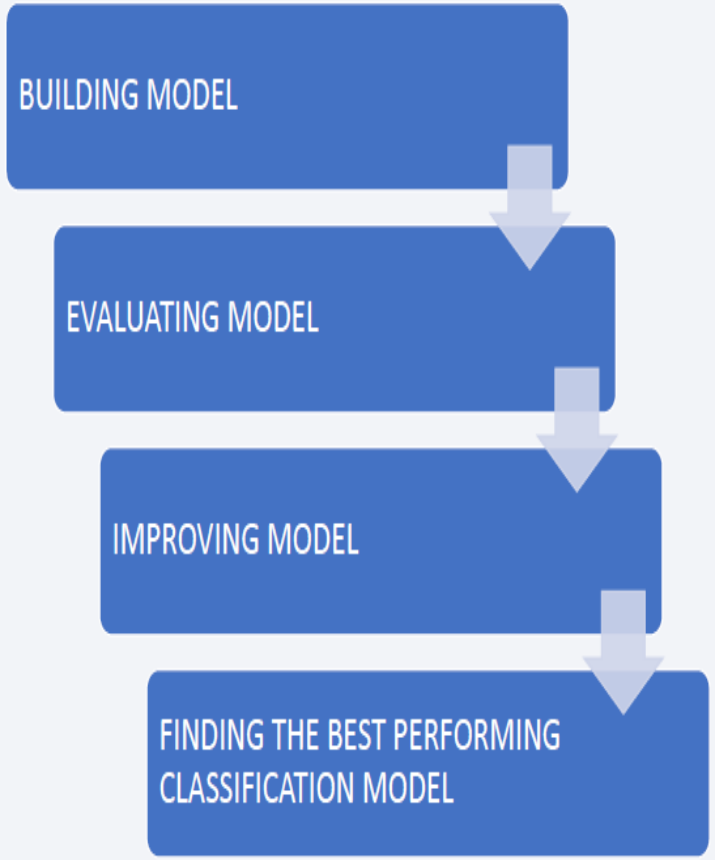
- Project/Build a Dashboard with Plotly Dash.ipynb at main · Punampatil25/Project (github.com)

Predictive Analysis (Classification)

- Perform exploratory Data Analysis anddetermine Training Labels

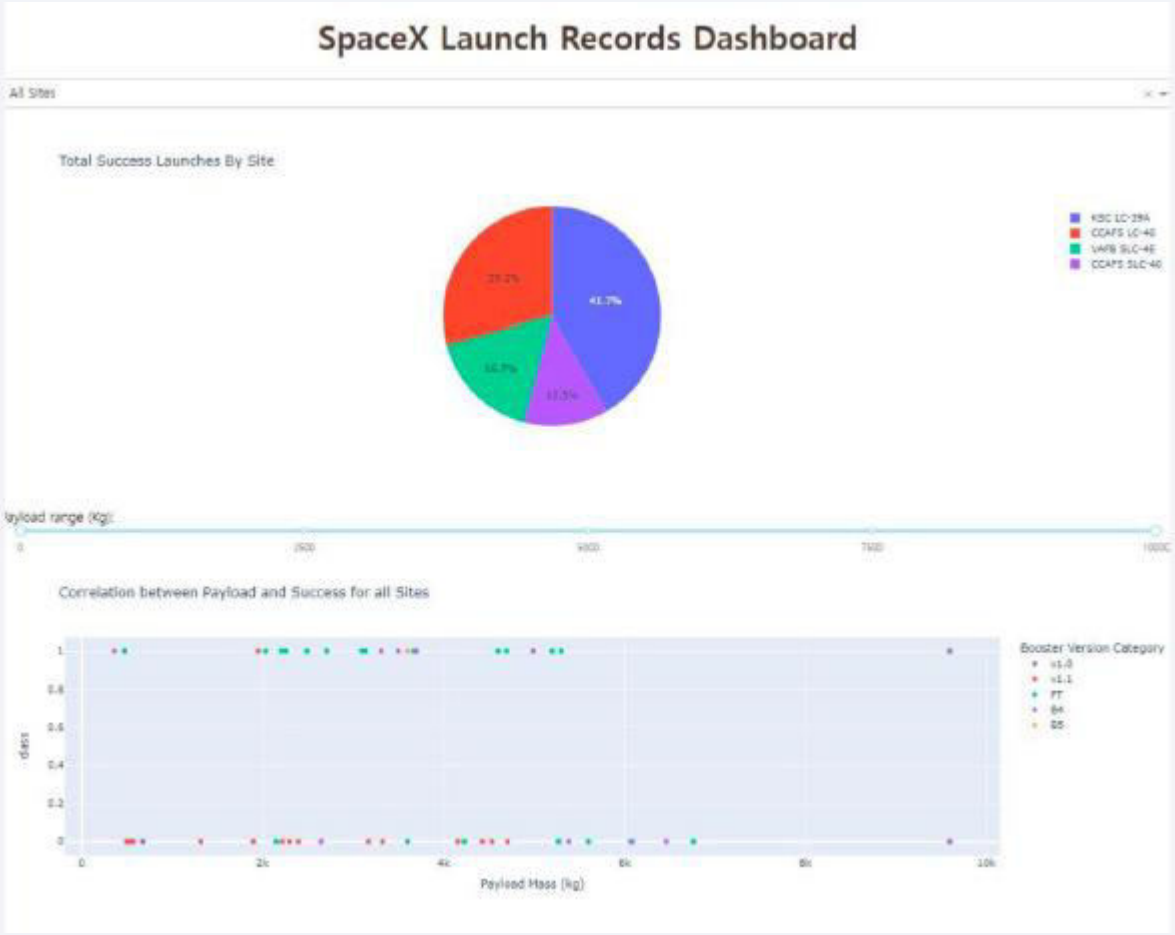
1. Create a column for the class
2. Standardize the data
3. Split into training and test dataset

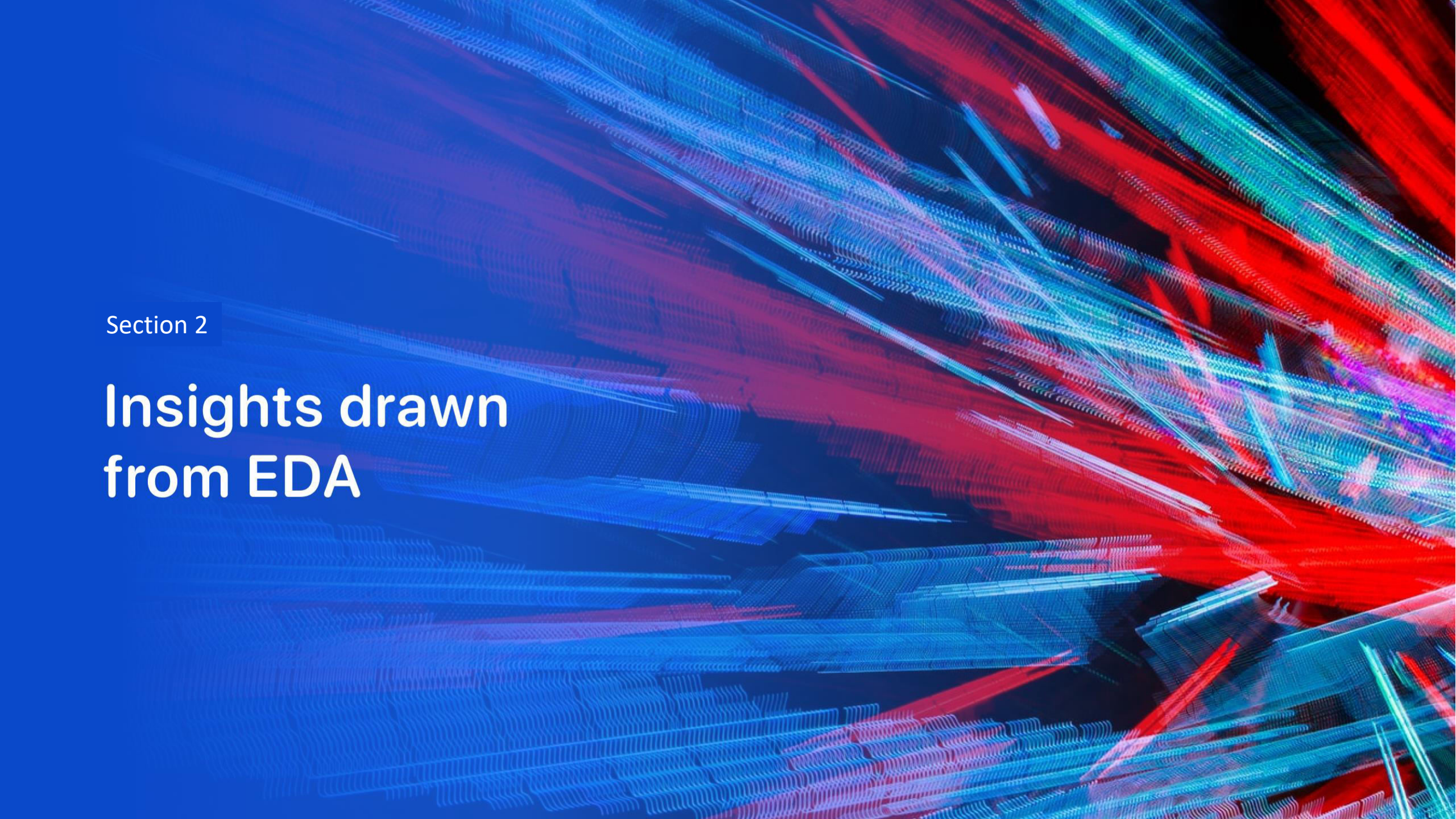
- Find best Hyper parameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using thegiven test data
- [Project/Predictive Analysis (Classification).ipynb at main · Punampatil25/Project (github.com)]



Results

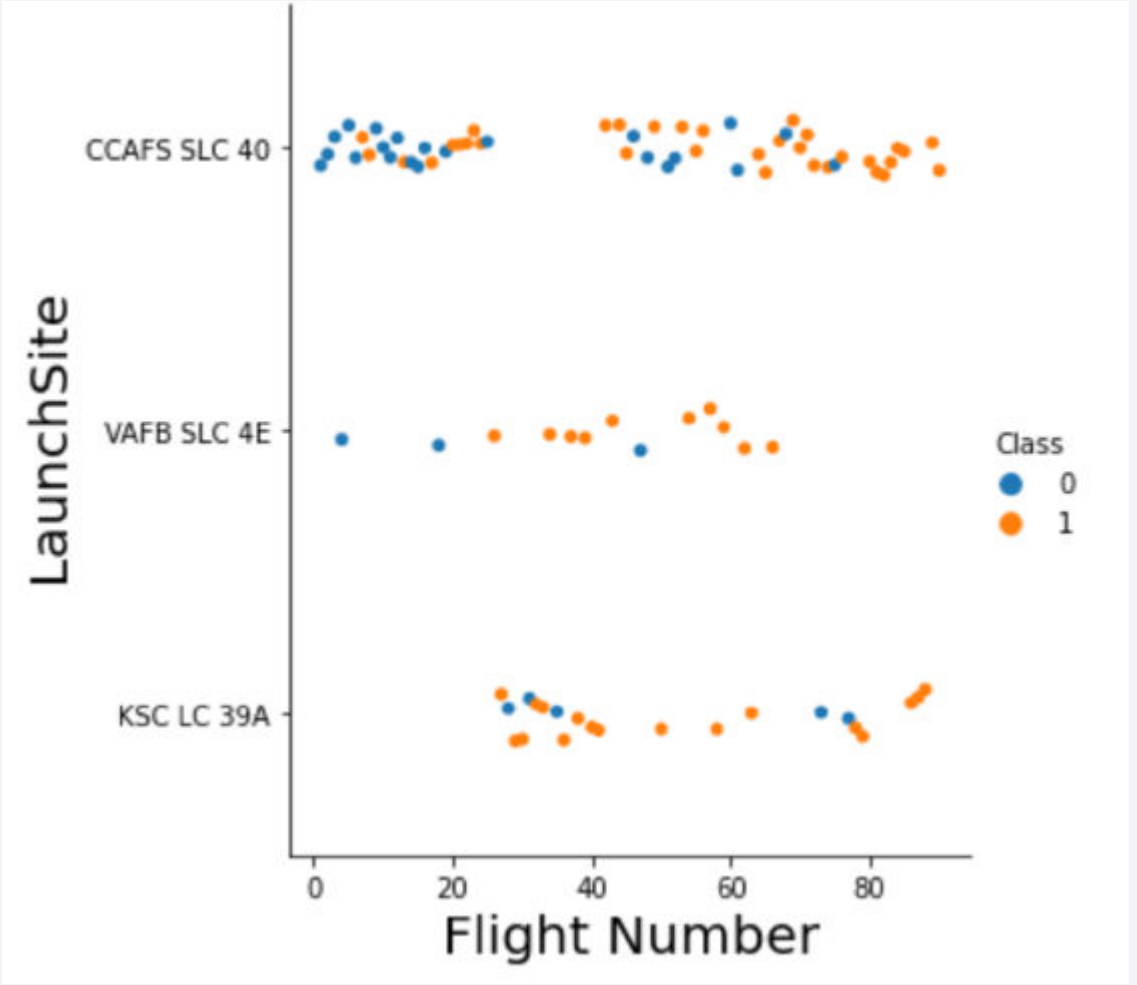
- See a preview of the Plotly Dash.
- The results of EDA with visualization, EDA with SQL, Interactive Map with Folium, and Interactive Dashboard will be shown in following slides.
- Comparing the accuracy of the four methods, all return the same accuracy of about 83% for test data.




Section 2
Insights drawn from EDA

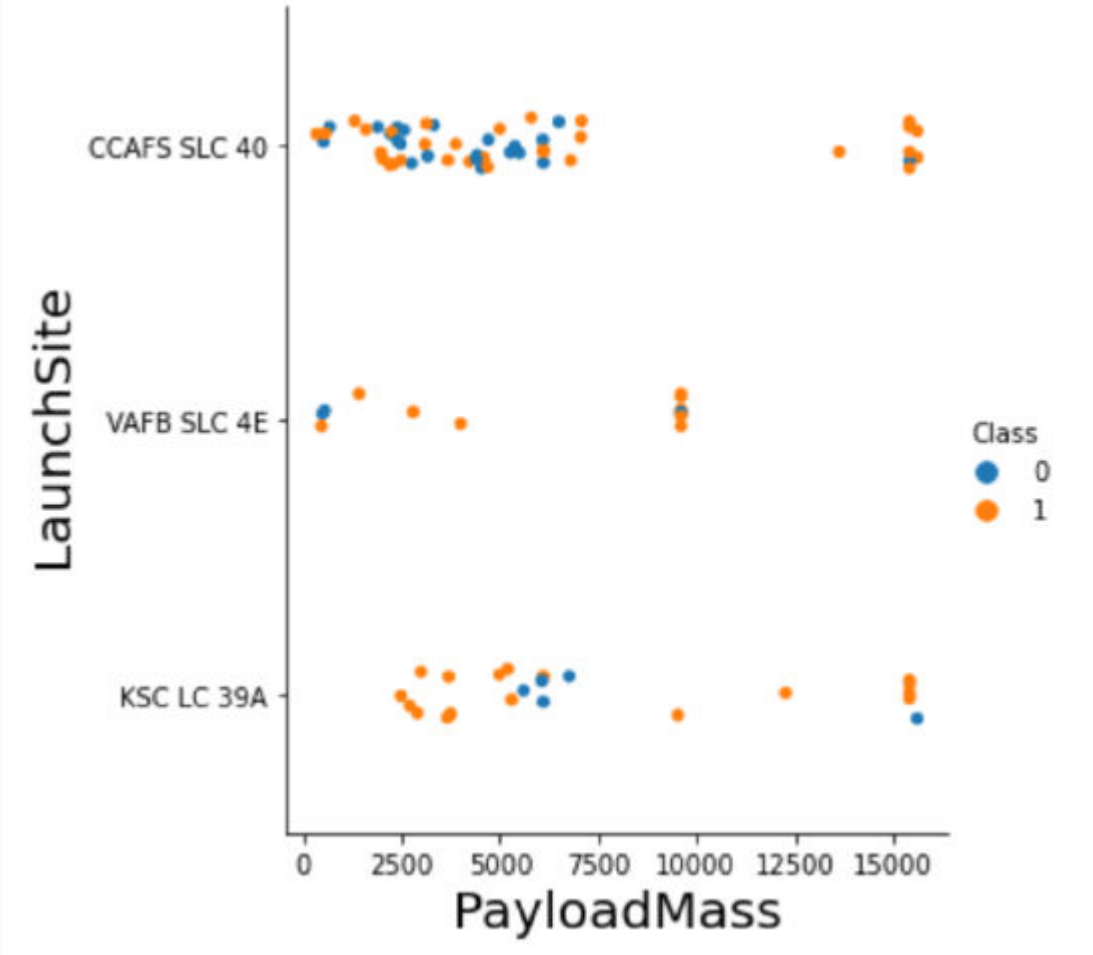
Flight Number vs. Launch Site

- Class 0 (blue) represents unsuccessful launch, and Class 1 (orange) represents successful launch.
- The figure shows that the success rate increased as the number of flights increased.
- As the success rate has increased considerably since the 20th flight, this point seems to be a big breakthrough.



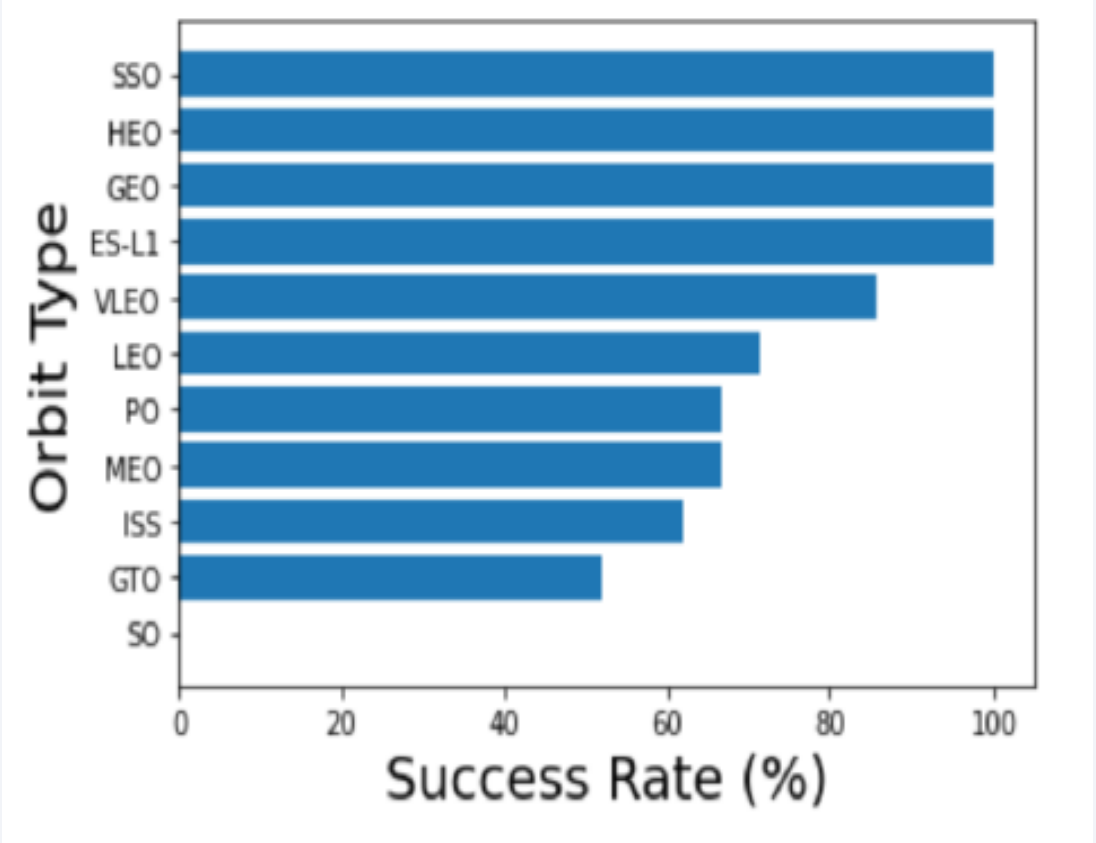
Payload vs. Launch Site

- Class 0 (blue) represents unsuccessful launch, and Class 1 (orange) represents successful launch.
- Based on the pic, the larger pay load mass, the higher the rocket‘s success rate. However, it is difficult to make decisions based on this figure, since no clear pattern are tested between successful launch and Pay Load Mass.



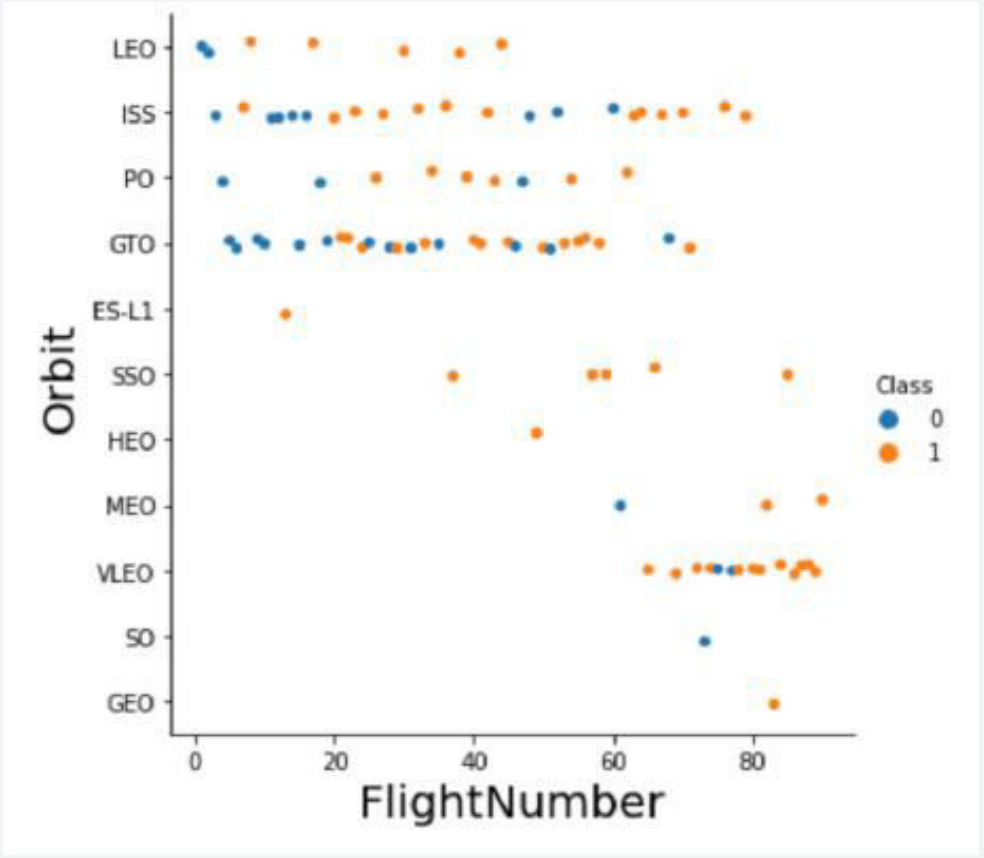
Success Rate vs. Orbit Type

- Orbit types SSO, HEO, GEO, and ES-L1 have the highest success rates (100%).
- On the other hand, the success rate of orbit type GTO is only 50%, and it is the lowest except for type SO, which recorded failure in a single attempt.



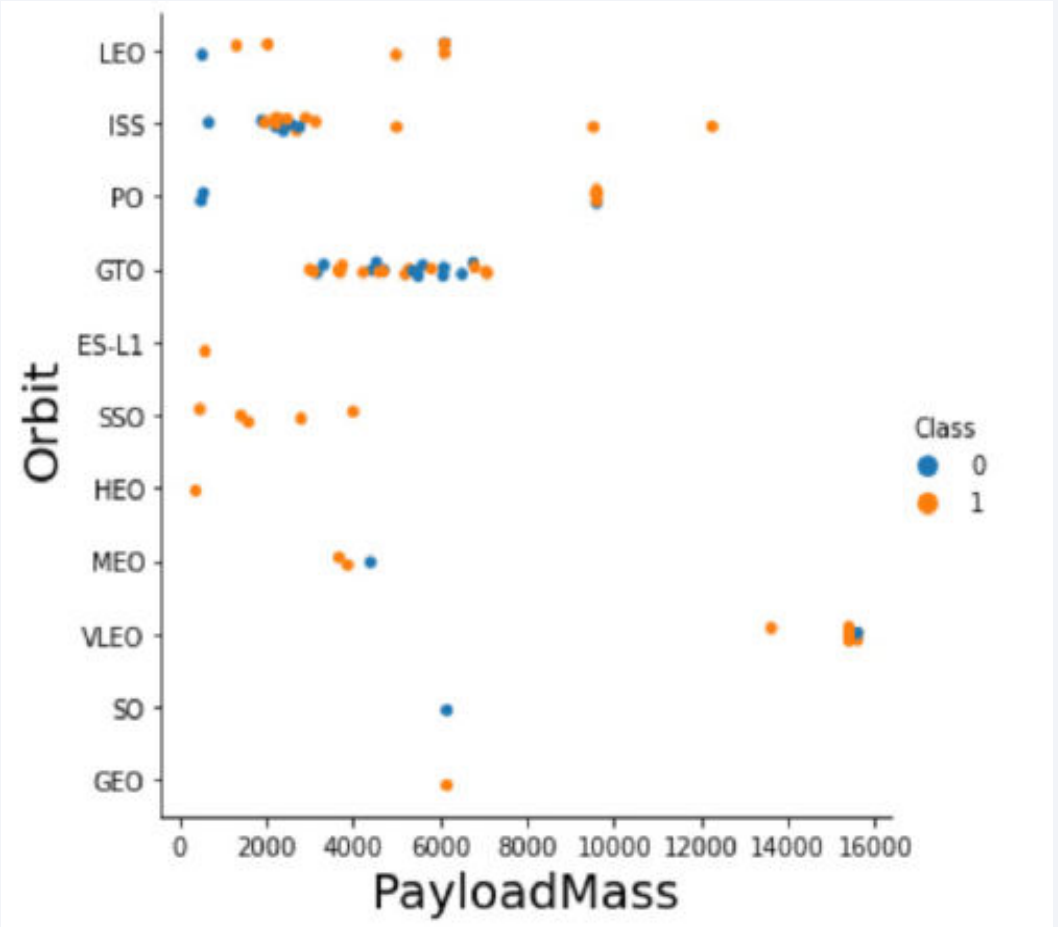
Flight Number vs. Orbit Type

- In most cases, the launch outcome seems to be correlated with the flight number.
- Meanwhile, in GTO orbit, there seems to be no correlation between flight numbers & success rate.
- SpaceX starts with LEO with moderate success rate, and it seems that VLEO, which has a high success rate, is used the most in recent launches.



Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for LEO and ISS.
- However, in the case of GTO, it is hard to distinguish the positive landing rate with the negative landings.



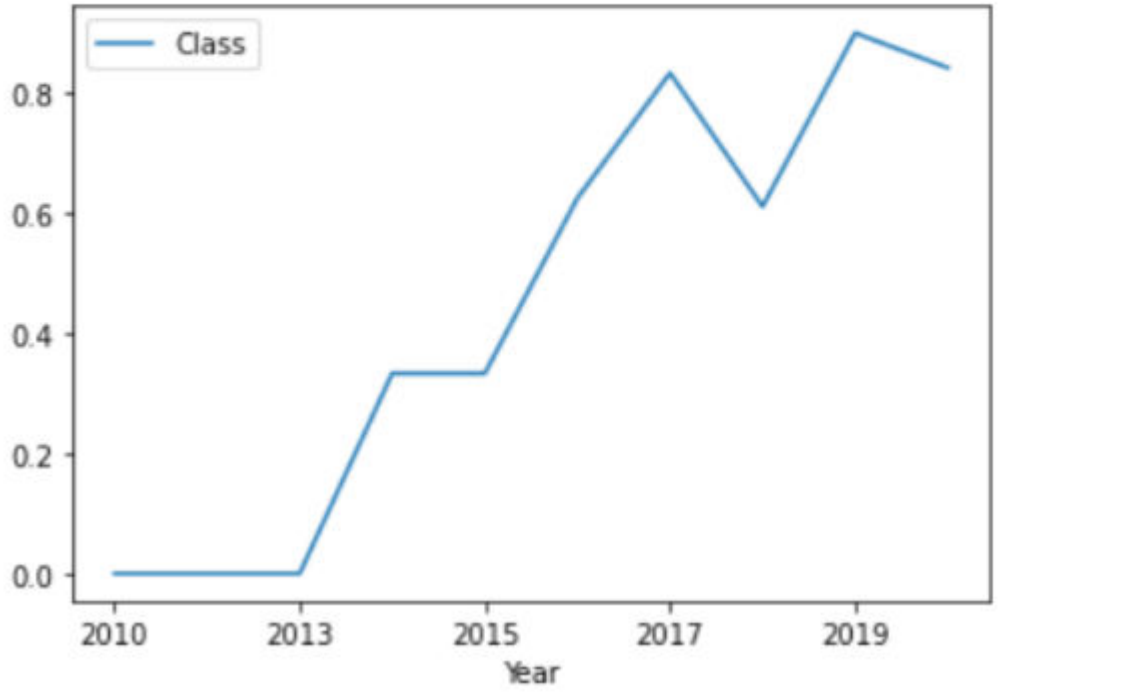
Launch Success Yearly Trend

- On the right is a line chart of yearly average success rate.
- From the chart we can see the following:

1. SpaceX did not manage to land any first stage successfully between 2010 and 2013.
2. Since their first successful landing in 2014, SpaceX managed to increase the success rate to >80% in 2019 and 2020.

```
df[['Year', 'Class']].groupby('Year').mean().plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7844f0f350>
```



All Launch Site Names

Task 1

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

```
%sql SELECT UNIQUE launch_site FROM eda_with_sql
```

```
 * ibm_db_sa://ytj70292:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

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

The names of the 4 unique launch sites are shown above.
SQL keyword “UNIQUE” helps to remove repeating values from the query result.

Launch Site Names Begin with 'CCA'

Task 2

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

```
%sql SELECT * FROM eda_with_sql WHERE launch_site LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://ytj70292:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
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	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

- The above query finds 5 records where launch sites begin with `CCA`
- SQL keyword “LIKE” helps match the launch_site name with the required string pattern, and “LIMIT 5” establishes the max number of results to be returned.

Total Payload Mass

Task 3

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

```
%sql SELECT SUM(payload_mass__kg_) FROM eda_with_sql WHERE UCASE(customer)='NASA (CRS)'
```

```
 * ibm_db_sa://ytj70292:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

1
45596

- The above query calculates the total payload carried by boosters from NASA.
- SQL function SUM() specifies the total payload mass to be returned, and “WHERE UCASE(customer)=’NASA (CRS)’” identifies the customer of interest and makes it insensitive to case.

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql SELECT AVG(payload_mass__kg_) FROM eda_with_sql WHERE booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://ytj70292:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

1
2534

- The above query calculates the average payload mass carried by booster version F9 v1.1
- SQL function AVG() specifies the average payload mass to be returned, and “WHERE booster_version LIKE ‘F9 v1.1%’” pattern-matches to include all F9 v1.1 boosters.

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

```
%sql SELECT MIN(date) FROM eda_with_sql WHERE landing__outcome='Success (ground pad)'
```

```
 * ibm_db_sa://ytj70292:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

1
2015-12-22

- The above query finds the date of the first successful landing outcome on ground pad.
- SQL function MIN() ensures the first or smallest date is to be returned, and “WHERE landing_outcome=’’Success (ground pad)” narrows down the landing target and outcome

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following query is used to list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Successful Drone Ship Landing with Payload between 4000 and 6000

```
SELECT BOOSTER_VERSION

FROM spacexlab

WHERE LANDING__OUTCOME = 'Success (drone ship)’

AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

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 following query is used to calculate the total number of successful and failure mission outcomes

```
SELECT MISSION_OUTCOME, COUNT(*) AS total_number

FROM spacexlab

GROUP BY MISSION_OUTCOME
```

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

Boosters Carried Maximum Payload

- The following query is used to list the names of the booster which have carried the maximum payload mass

```
SELECT DISTINCT BOOSTER_VERSION,
PAYLOAD_MASS__KG_ FROM spacexlab
WHERE PAYLOAD_MASS__KG_ = (
SELECT MAX(PAYLOAD_MASS__KG_)
FROM spacexlab
);
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

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

```
SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM spacexlab WHERE LANDING__OUTCOME = 'Failure (drone ship)'
AND YEAR(DATE) = '2015'
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

```
SELECT LANDING__OUTCOME,
count(LANDING__OUTCOME) as cnt
FROM spacexlab
WHERE DATE between '2010-06-04' and '2017-03-20'
group by LANDING__OUTCOME
order by 2 desc
```

landing__outcome	cnt
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


Section 3
Launch Sites
Proximities Analysis

Map of All Launch Sites

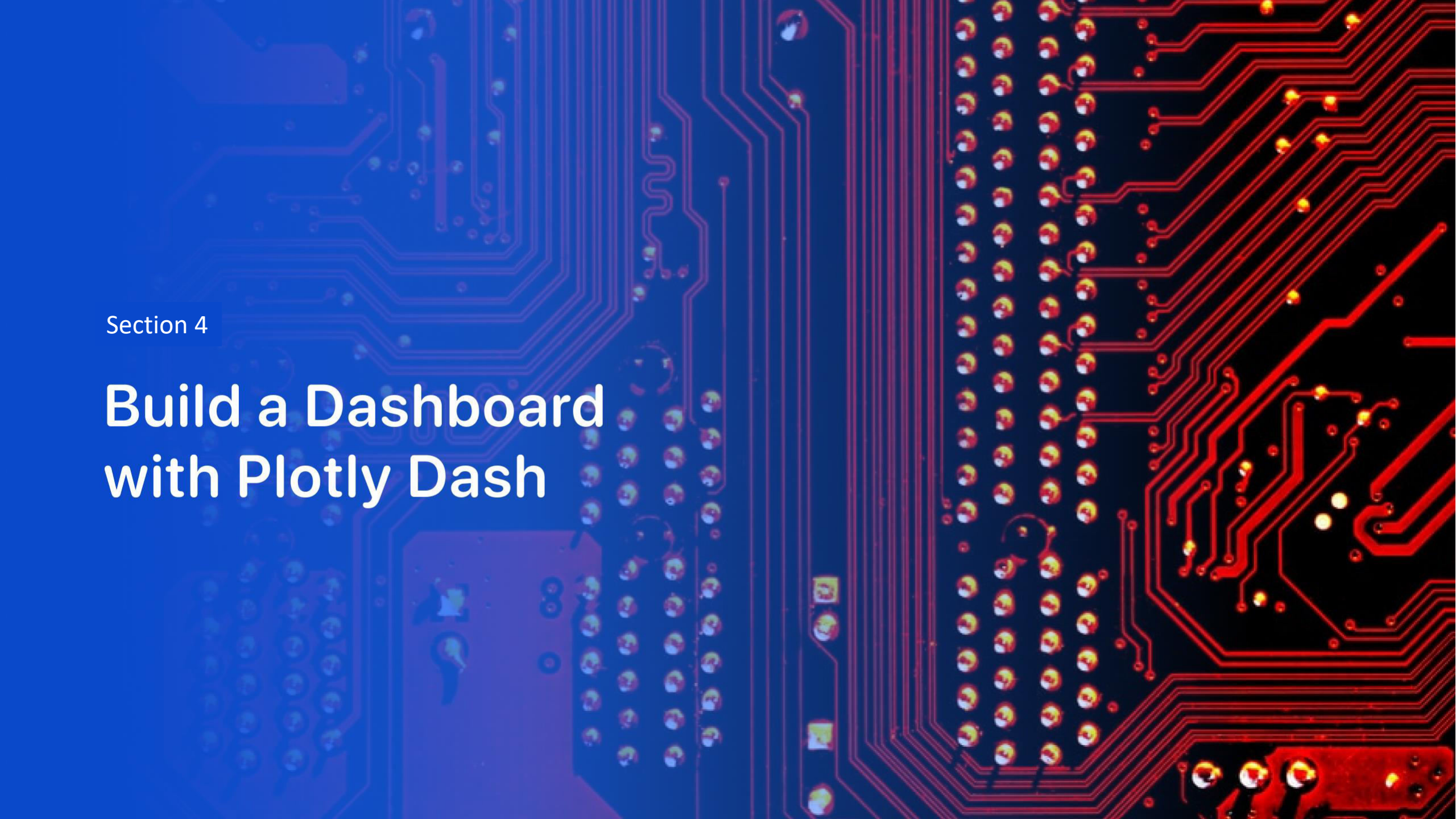
- See left map to find all SpaceX launch sites, where mostly located in California and Florida.
- As shown, all launch sites are near coastline for safety reason.

Color-labeled Launch Outcomes

- We can explore the folium map and make a clear deep dive in successful landing (green) and failed ones (red).
- This is used to elaborate between different sites.

Proximities to Coastline

- Explore the generated folium map and show the proximities of coastline
- After dive deeper into the visualized map, we can also conclude that the launch sites are far from cities centers.

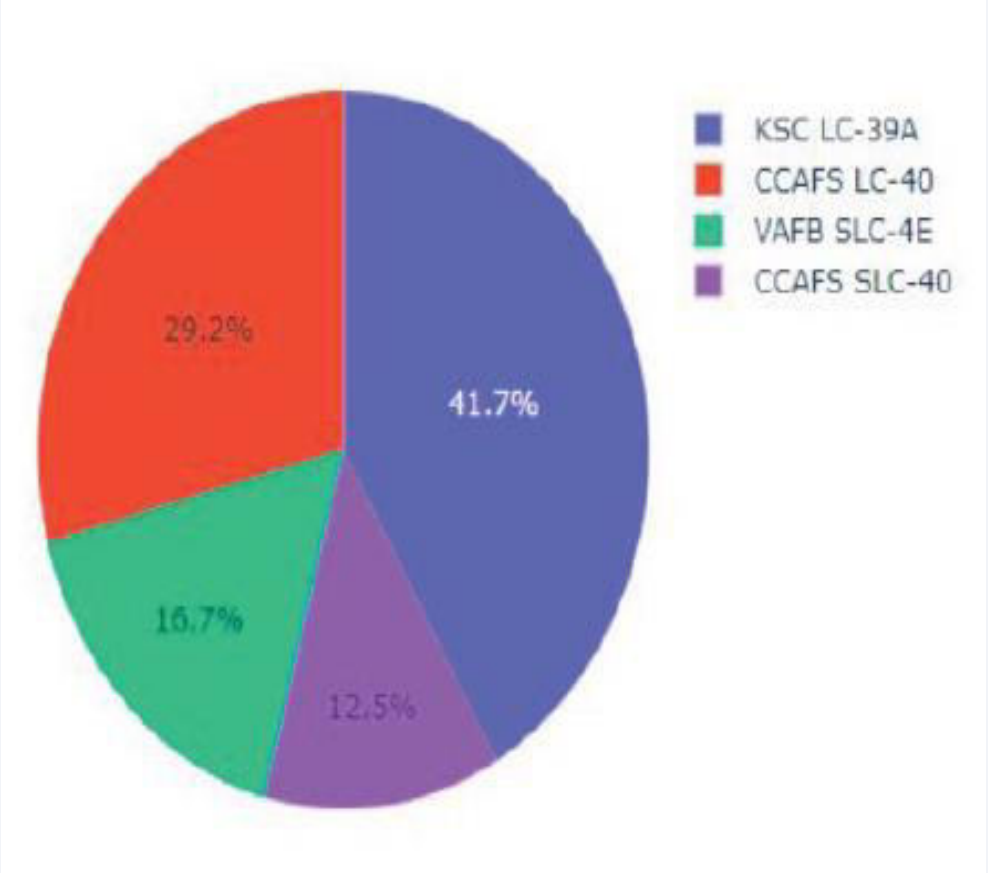

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches by Launch Sites

- The site KSLC-39A records the most launch success among all sites.
- The VAFB SLC-4E site has the fewest launch success. Possibly because

i. The data sample is not large enough to compared

ii. Since it's located in California, the launch difficulty on the west coast is higher than on the east coast.



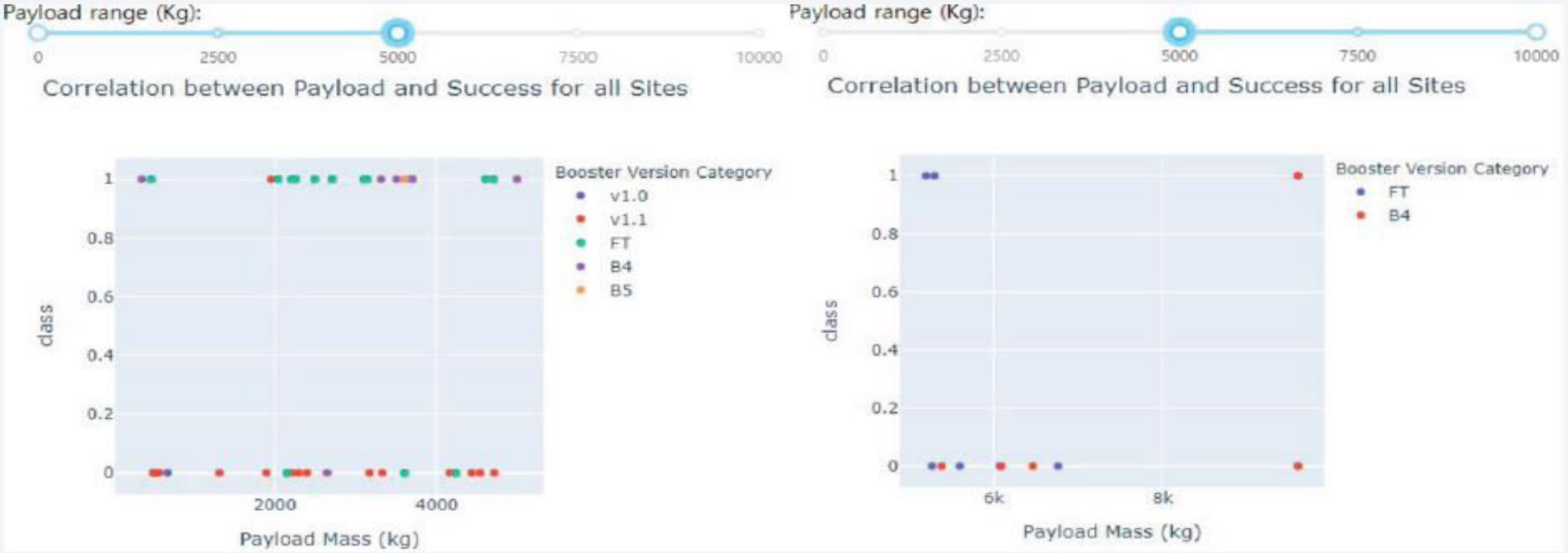
Launch Sites with Highest Launch Success Rate

- SLC-39A has the highest success rate with:

i. 10 landing successes (76.9%)

ii. Only 3 landing failures (23.1%)

Launch Success Rate by Payload



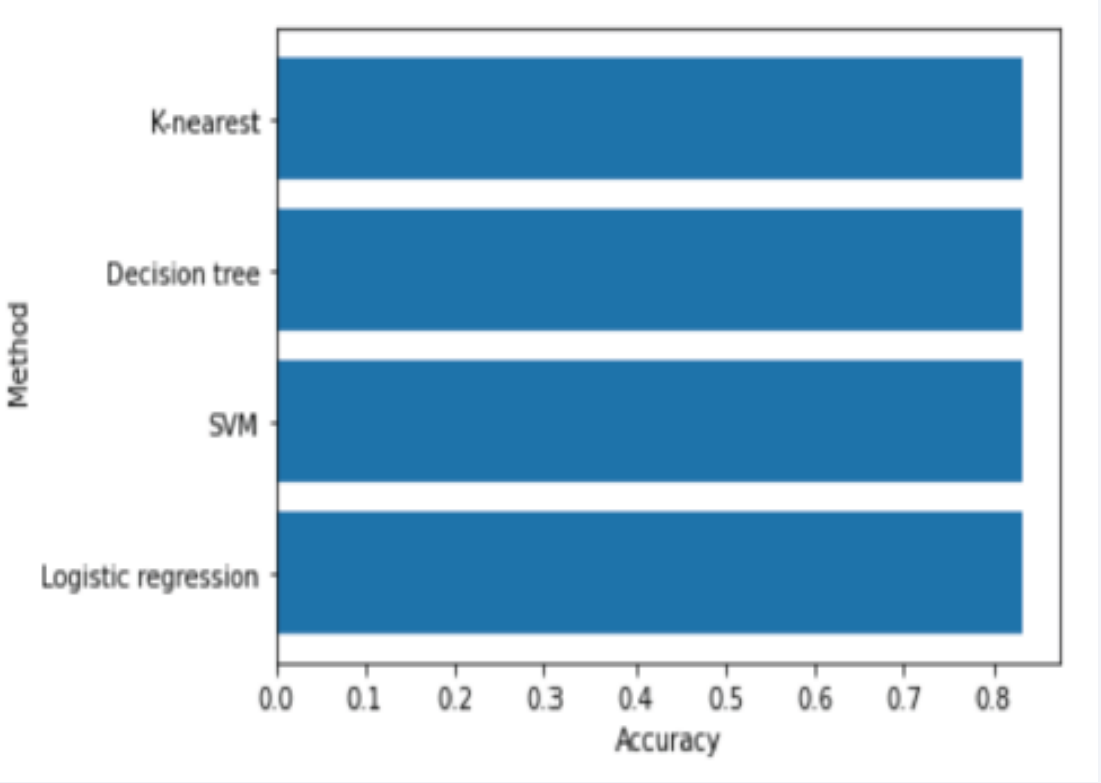
- Launch success rates (class 1) for low weighted payloads (0-5000 kg) are higher than heavy weighted payloads (5000-10000 kg).

Section 5

Predictive Analysis (Classification)

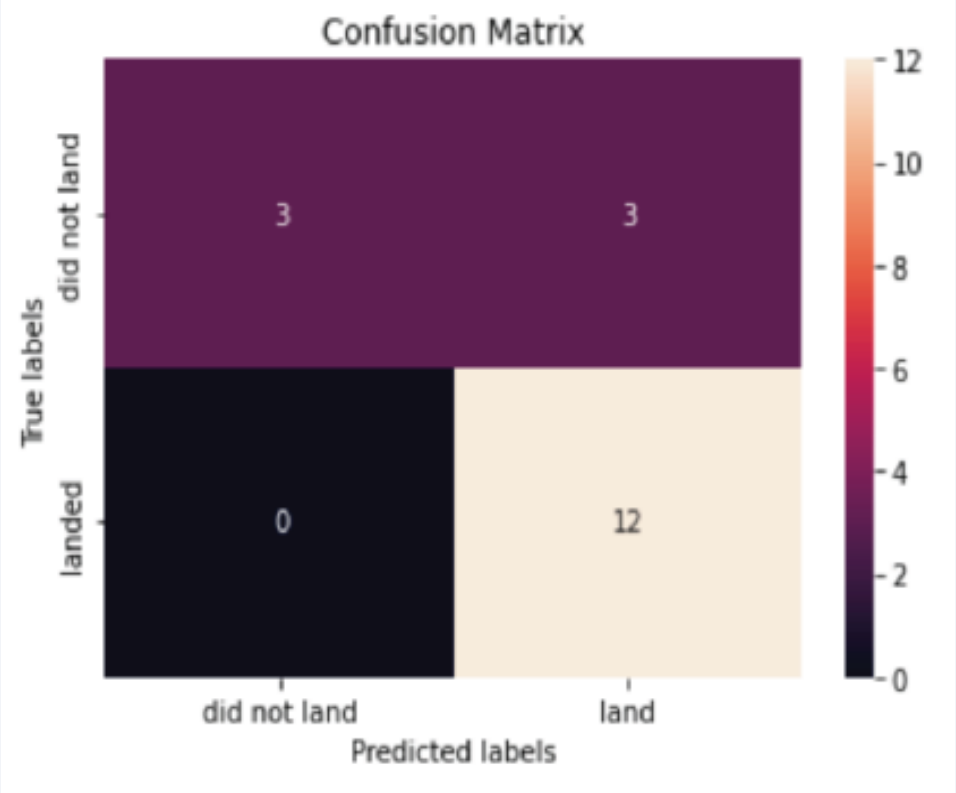
Classification Accuracy

- The accuracy of all models was virtually the same at 83.33%.
- It should be noted that the test size was small at 18.
- Therefore, we need more data to determine using a Machine Learning optimal model.



Confusion Matrix

- The confusion matrix is the same for all models, since all models performed with the same test and train set.
- The models predicted 12 successful landings when the true label was successful and 3 failed landings when the true label was failure. There are also 3 predictions that said successful landings when the true label was failure (false positive).



Conclusions

- As the number of flights increased, the success rate increased. Recent records can exceed 80%.
- Orbital types SSO, HEO, GEO, and ES-L1 have the highest success rate (100%).
- According to sites on map, the launch site is close to coastline, but far from cities.
- KSLC-39A has the highest number of launch successes and the highest success rate among all sites.
- The launch success rate of low weighted payloads is higher than that of heavy weighted payloads.
- The dataset size is so small that all models have the same accuracy (83.33%). More data is needed to determine the optimal model.

Appendix

- Github: https://github.com/Punampatil25/Project/upload/main

Thank you!