IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Azhagan Avr
21/11/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory Data Analysis with data visualization and SQL
  - Building an interactive map with Folium
  - Building a Dashboard with PlotlyDash
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis results
  - Dashboard Interaction results
  - Predictive analysis results

Introduction

BACKGROUND

- 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.

Questions answered in this project

- If the Falcon 9 first stage will land successfully or not ?
- What’s the secret behind successful launches?

Section 1

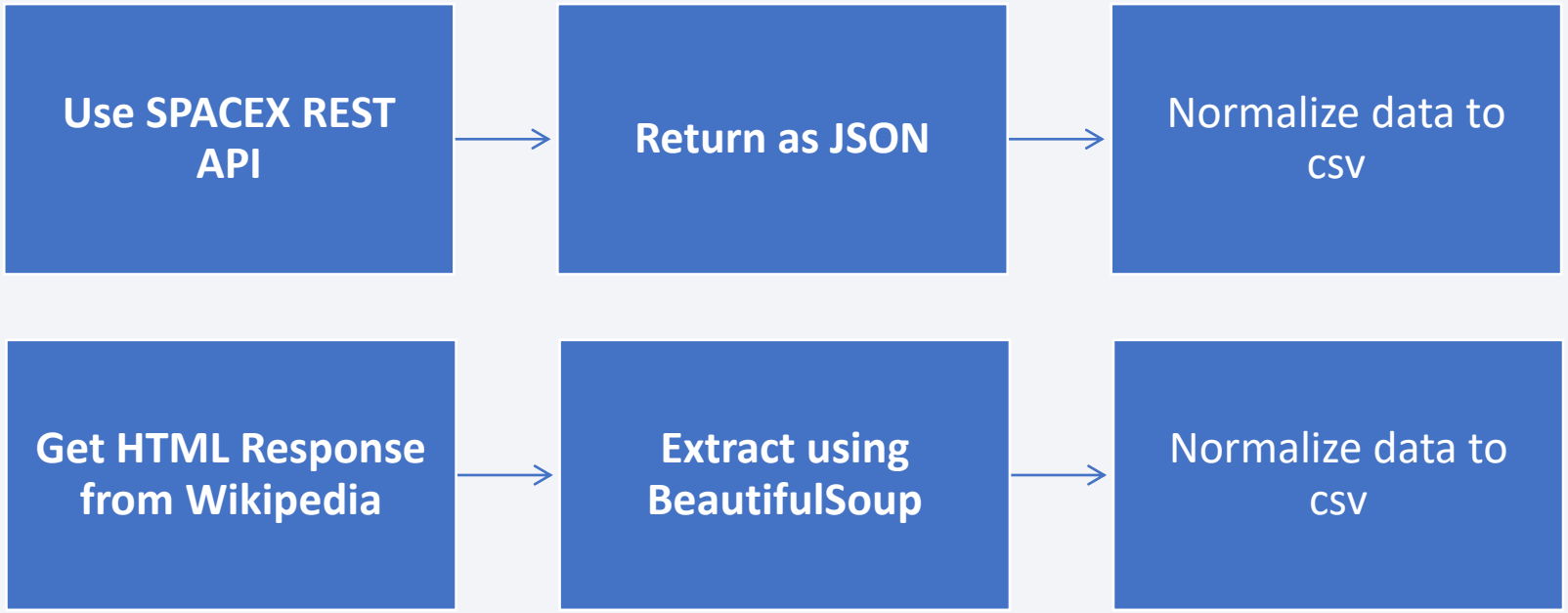
Methodology

Methodology

Executive Summary

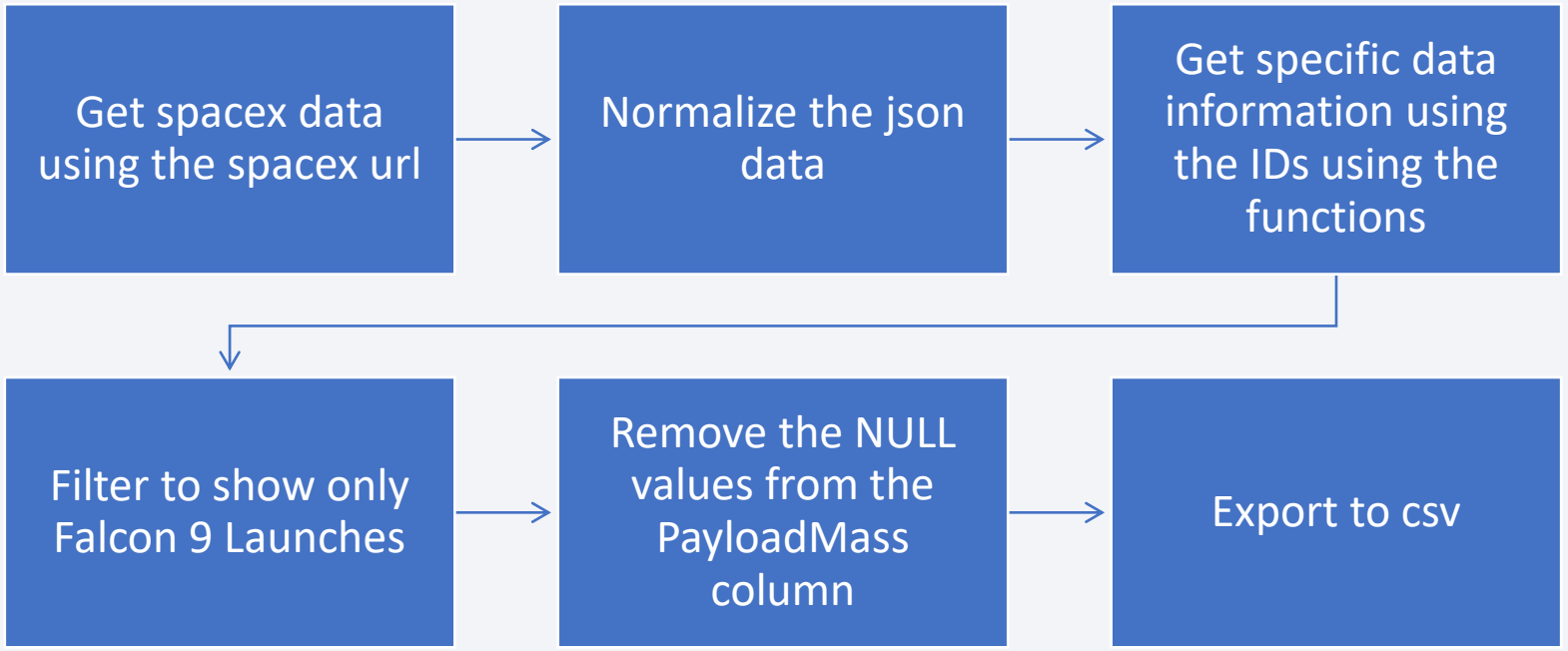
- Data collection methodology:
  - SpaceX launch data that is gathered from an API, specifically the SpaceX REST API and web scrapping from Wikipedia
- Perform data wrangling
  - One hot encoding data fields for machine learning and Dealing with Nulls to clean the dataset.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, K-Nearest Neighbors, Support Vector Machine and Decision Tree models were used to determine the best classifier model to predict the outcome

Data Collection



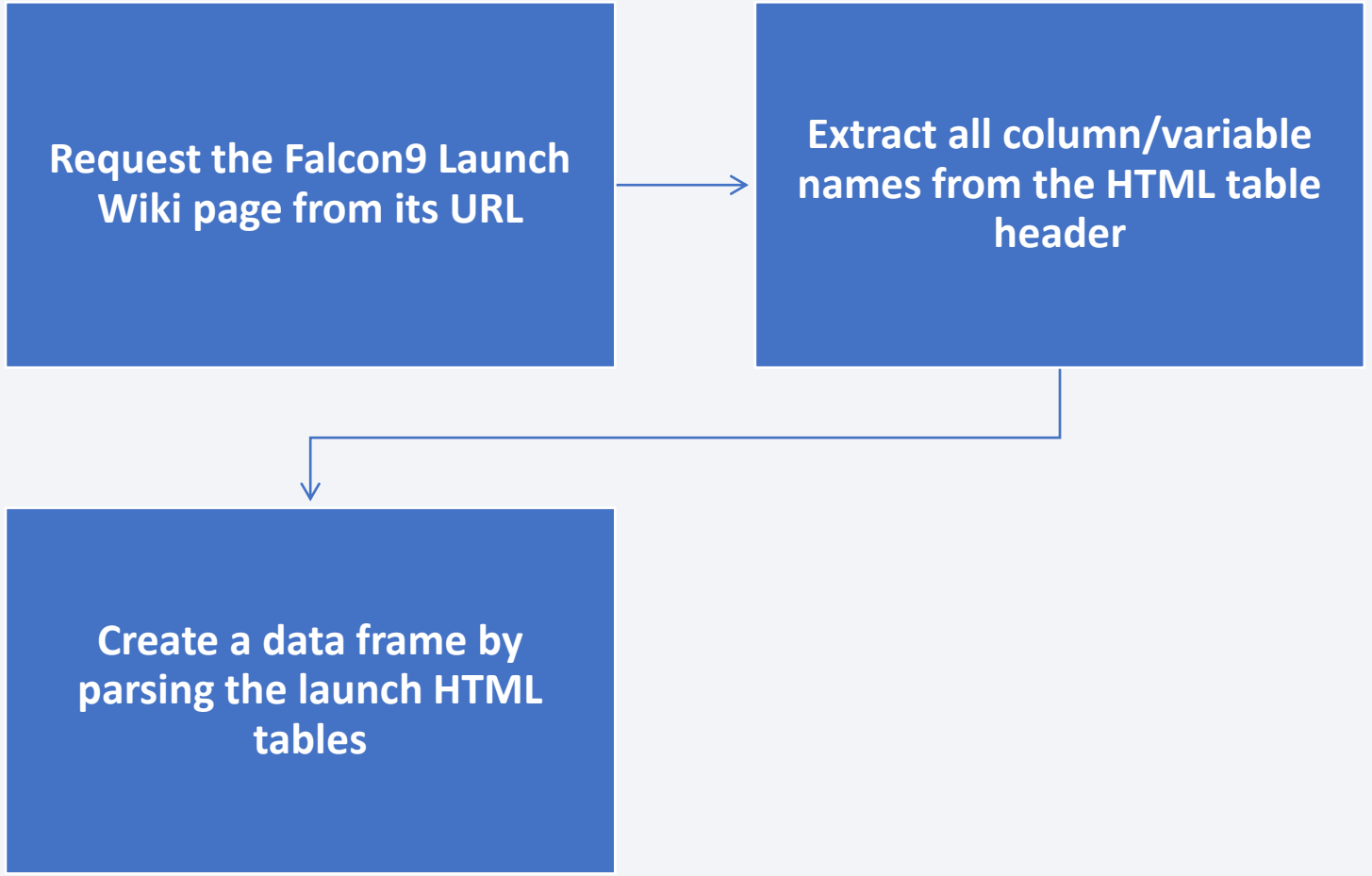
- SpaceX launch data that is gathered from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Also, BeautifulSoup library was used to webscrape Falcon 9 launch records from Wikipedia.

Data Collection – SpaceX API



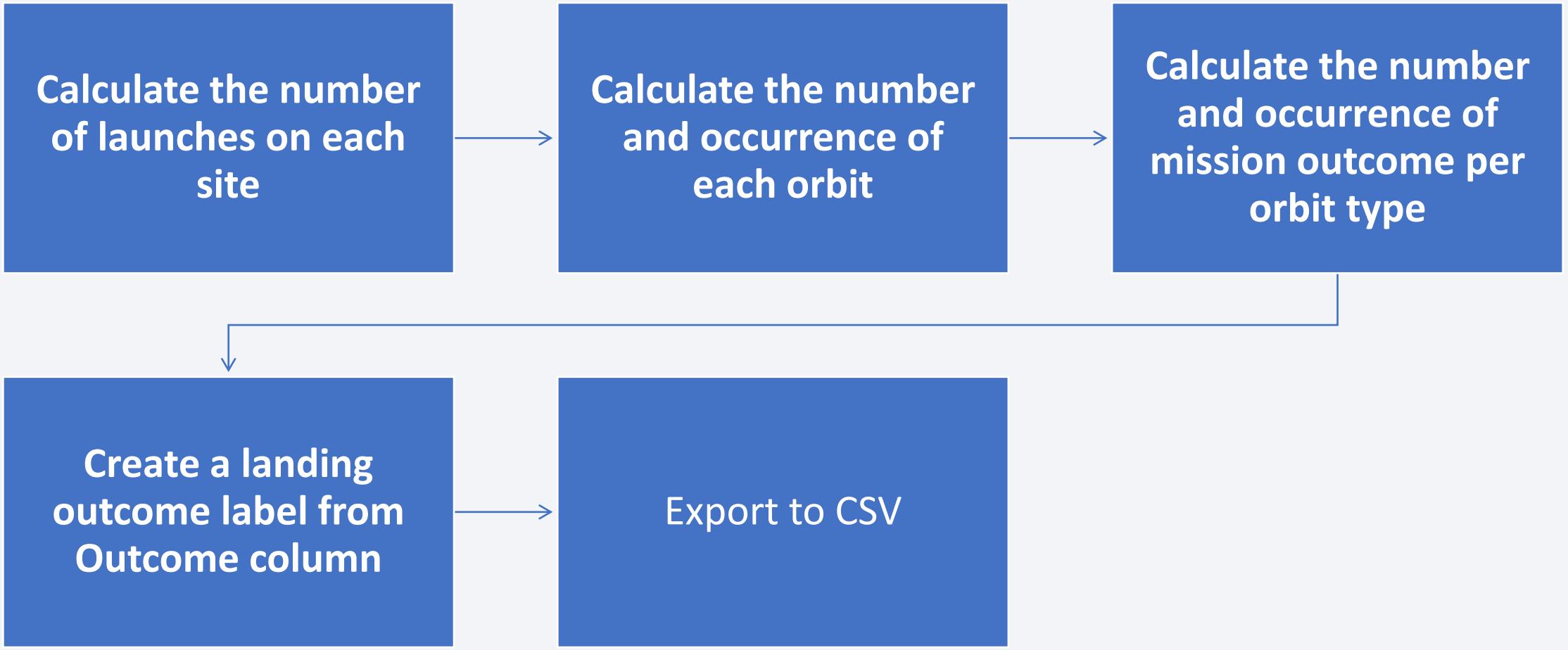
- https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/Data%20Collection%20API%20Lab%20-%20Week%201.ipynb

Data Collection - Scraping



- https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/Data%20Collection%20with%20Web%20Scraping%20lab%20-%20Week%201.ipynb

Data Wrangling



- https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/EDA%20lab-Data%20wrangling.ipynb

EDA with Data Visualization

Visualize the relationship between Flight Number and Launch Site	Visualize the relationship between Payload and Launch Site	Visualize the relationship between success rate of each orbit type
Visualize the relationship between Flight Number and Orbit type	Visualize the relationship between Payload and Orbit type	Visualize the launch success yearly trend

- https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/EDA%20with%20Visualization%20lab%20-%20Week%202.ipynb

EDA with SQL

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

https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/EDA%20with%20SQL%20lab%20-%20Week%202.ipynb

Build an Interactive Map with Folium

- Circle markers were added to the Follium map to get an understanding of the location of the launch sites
- Line markers were used to get an understanding of the proximities of nearby services

https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab-Week3_final.ipynb

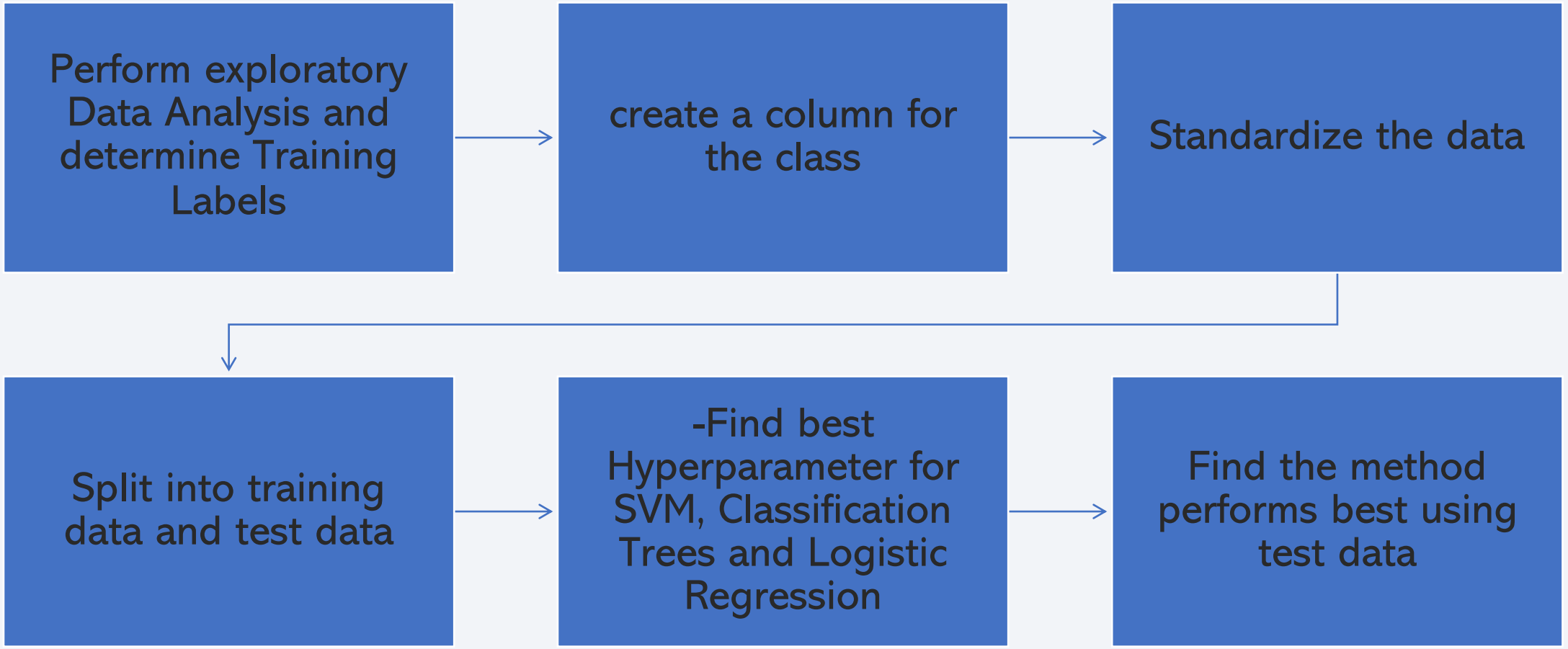
Copy paste the github link in this website if the interactive maps don't load
https://nbviewer.org/

Build a Dashboard with Plotly Dash

- A dropdown list to choose a launch site or all launch sites at once
- A pie chart to summarize the success of launches by site.
- A range slider to adjust the payload weight
- A scatter plot to correlate between Success and Payload weight
- The dropdown list controls both the charts and range slider controls the scatter plot.
- These charts gives us an understanding of successful launch sites and their corresponding payload weights

https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/spacex_dash_app.py

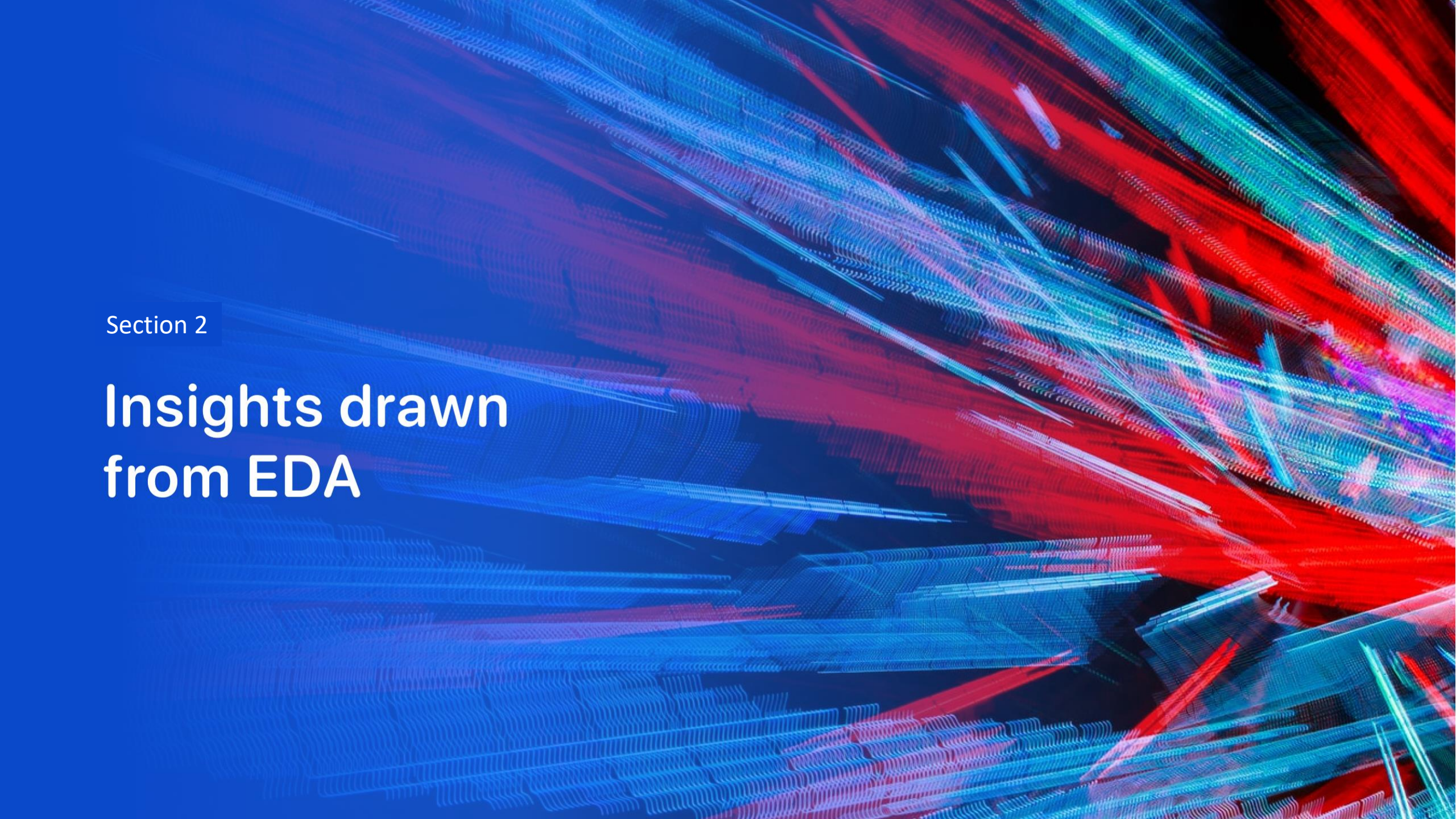
Predictive Analysis (Classification)



https://github.com/azhaganavr/IBM-Data-Science-Capstone-Project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5_final.ipynb

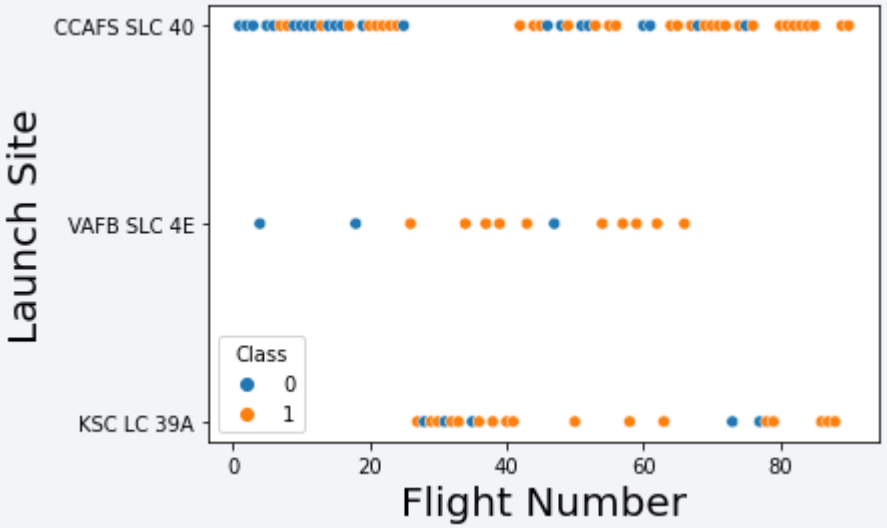
Results

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


Section 2
Insights drawn from EDA

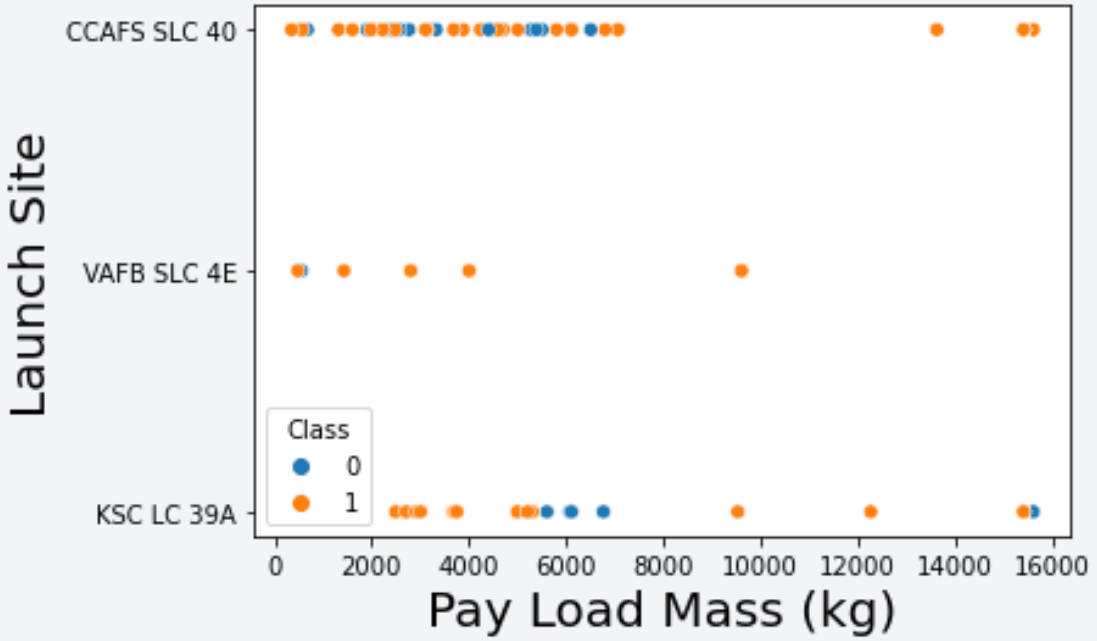
Flight Number vs. Launch Site

- CCAFS SLC 40 has the highest number of launches
- VAFB SLC 4E has the least number of launches



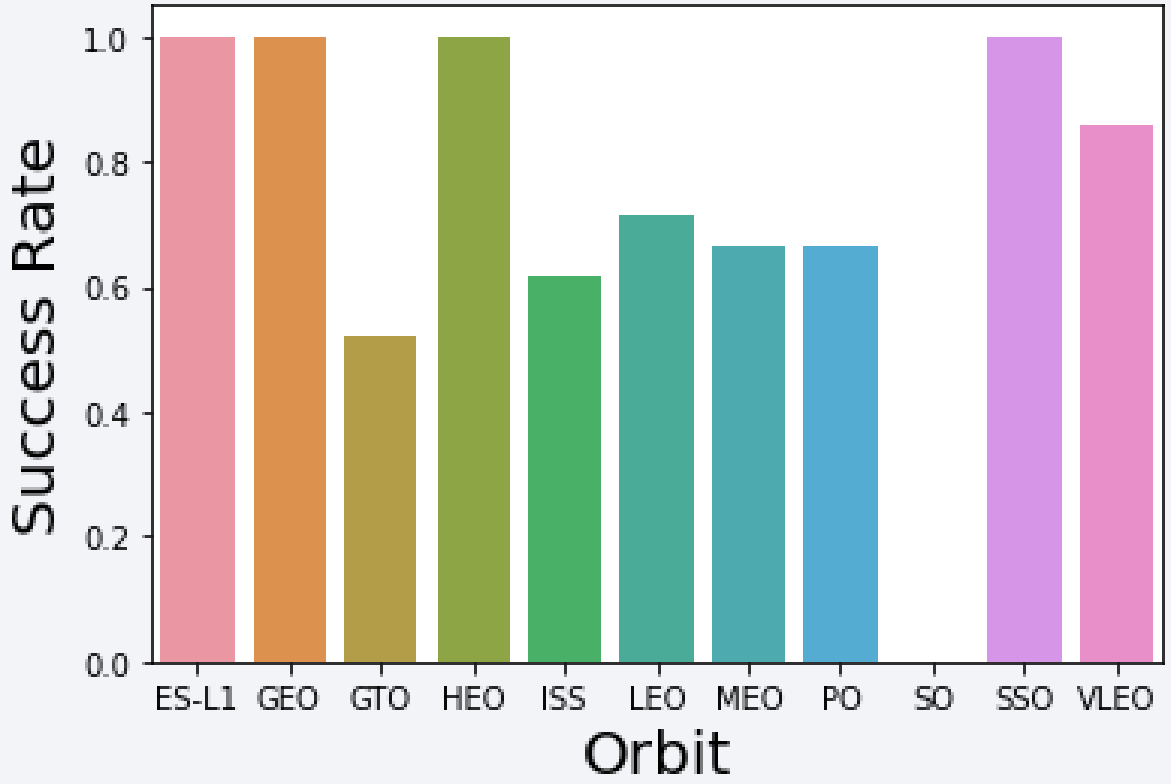
Payload vs. Launch Site

- VAFB SLC 4E has the least number of heavy payload launches
- CCAFS SLC 40 has the highest number of small payload launches



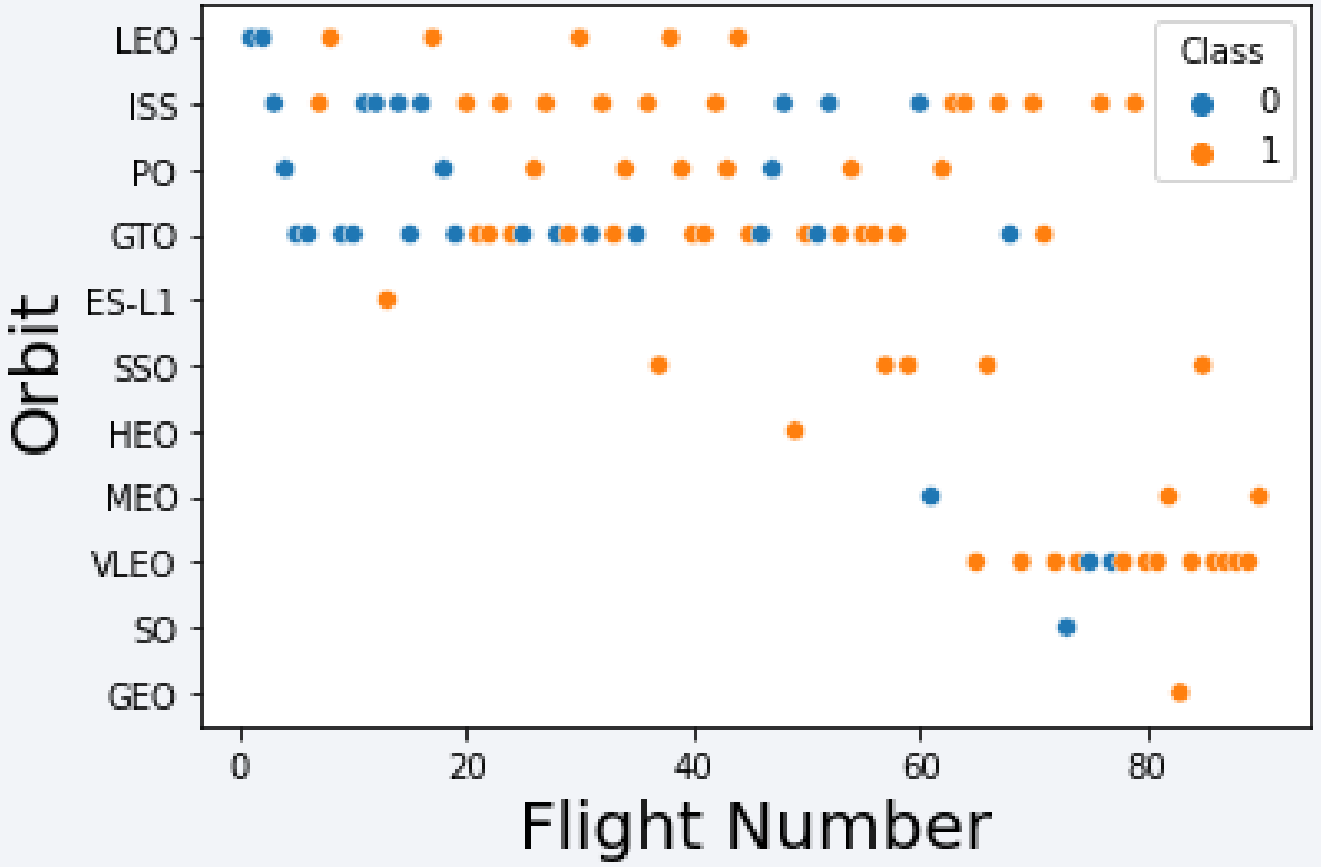
Success Rate vs. Orbit Type

- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.
- GTO has the lowest success rate



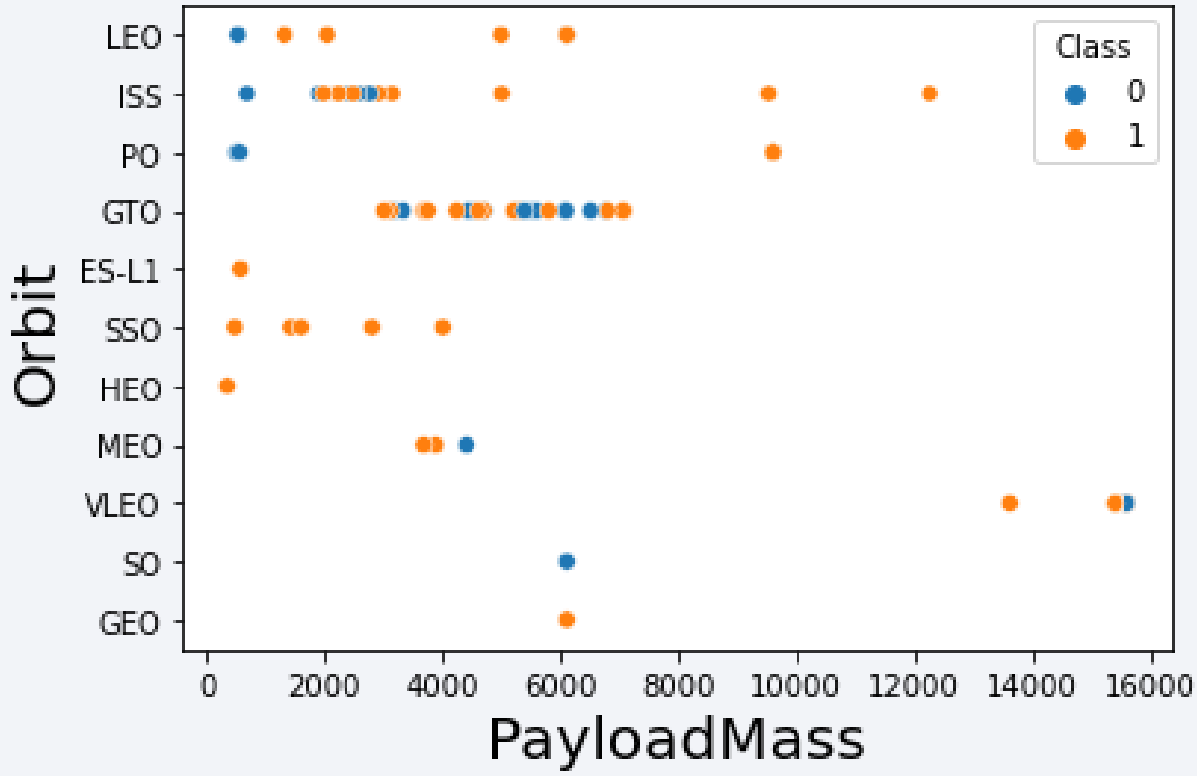
Flight Number vs. Orbit Type

- VLEO Orbit launches are more frequent in the recent years and have been successful



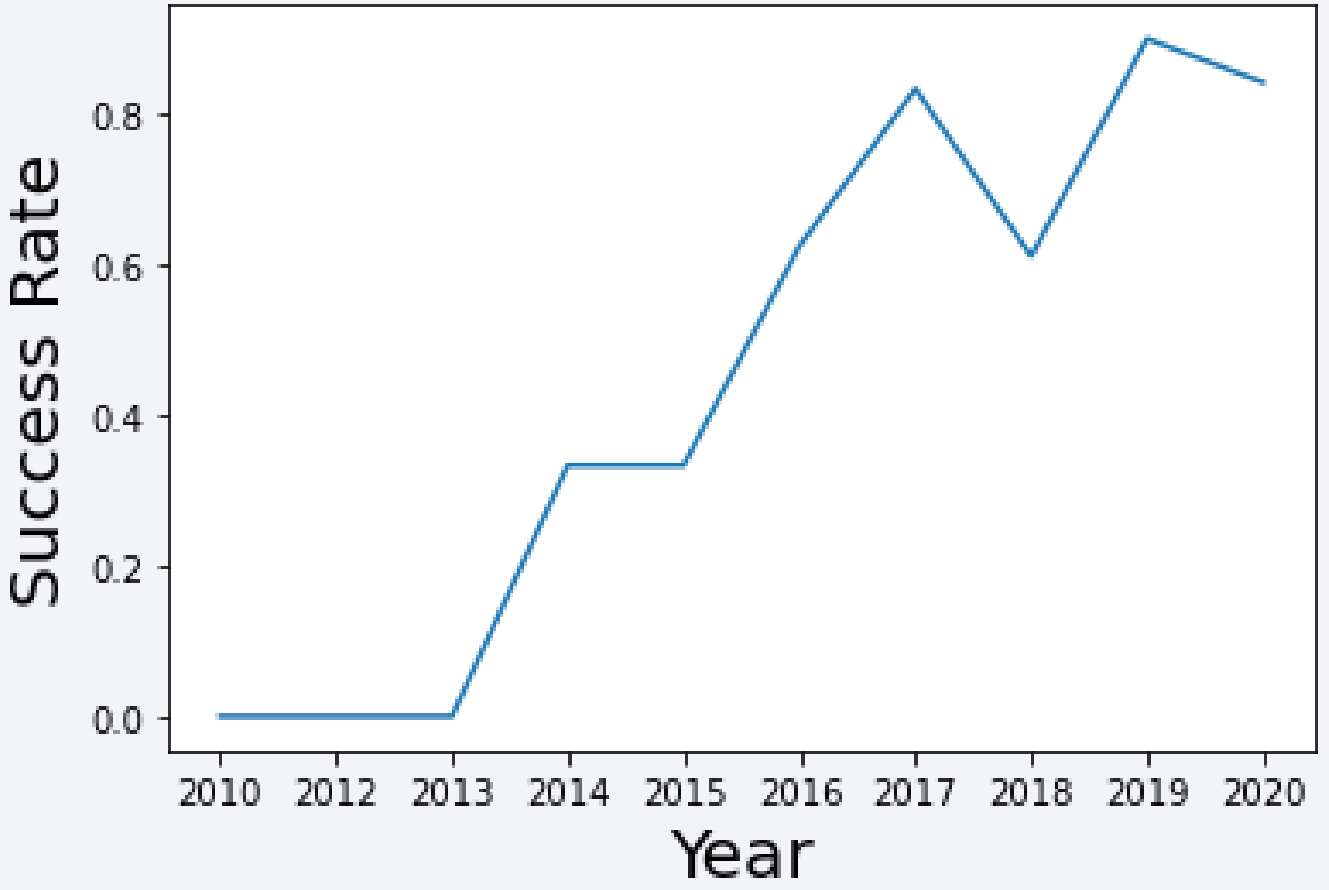
Payload vs. Orbit Type

- ISS Orbit launches have a payload around 2000Kg
- GTO Orbit launches have a payload ranging between 4000 – 8000 kg



Launch Success Yearly Trend

- Sucess rate since 2013 kept increasing gradually till 2020



All Launch Site Names

Task 1

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

```
%sql select DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

Launch Site Names Begin with 'CCA'

Task 2

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

```
%sql SELECT * FROM SPACEXTBL WHERE (LAUNCH_SITE LIKE '%CCA%') LIMIT 5;
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.data
bases.appdomain.cloud:32733/bludb
Done.
```

DATE	time_utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_ou
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	

Total Payload Mass

Task 3

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

```
%sql select SUM(payload_mass__kg_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.data
bases.appdomain.cloud:32733/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql select DISTINCT booster_version FROM SPACEXTBL;
%sql select * FROM SPACEXTBL WHERE booster_version = 'F9 v1.1';
%sql select AVG(payload_mass__kg_) FROM SPACEXTBL WHERE booster_version = 'F9 v1.1';
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.data
bases.appdomain.cloud:32733/bludb
Done.
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.data
bases.appdomain.cloud:32733/bludb
Done.
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.data
bases.appdomain.cloud:32733/bludb
Done.
```

1
2928

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

In [28]:
```
%sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdom
Done.
```

Out[28]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 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 [34]:
```
%sql SELECT DISTINCT booster_version FROM SPACEXTBL WHERE (Landing__Outcome = 'Success (drone ship)') AND (payload_mass__kg_ BETWEEN 4000 AND 6000);
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[34]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [36]:
```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)';
# %sql SELECT DISTINCT mission_outcome FROM SPACEXTBL;
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[36]:

1
100

Boosters Carried Maximum Payload

Task 8

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

In [12]:
```
%sql SELECT DISTINCT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL);
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[12]:

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

2015 Launch Records

Task 9

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

In [41]:
```
# %sql SELECT DISTINCT landing__outcome FROM SPACEXTBL; # Failure (drone ship)
%sql SELECT MONTH(DATE),MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXTBL where EXTRACT(YEAR FROM DATE)='2015' and landing__outcome = 'Failur
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[41]:

1	mission_outcome	booster_version	launch_site
1	Success	F9 v1.1 B1012	CCAFS LC-40
4	Success	F9 v1.1 B1015	CCAFS LC-40

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

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 [57]:

```
# %sql SELECT landing_outcomes FROM SPACEXTBL WHERE (Date >= '2010-06-04' and Date <='2017-03-20') AND (ORDER BY DATE DESC);


%sql SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) as "Count" FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'\
GROUP BY LANDING__OUTCOME\
ORDER BY COUNT(LANDING__OUTCOME) DESC;

# %sql SELECT landing__outcomes, COUNT(landing__outcomes) FROM SPACEXTBL;
```

```
 * ibm_db_sa://vjv72326:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[57]:

landing__outcome	Count
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

Mark all launch sites on a map

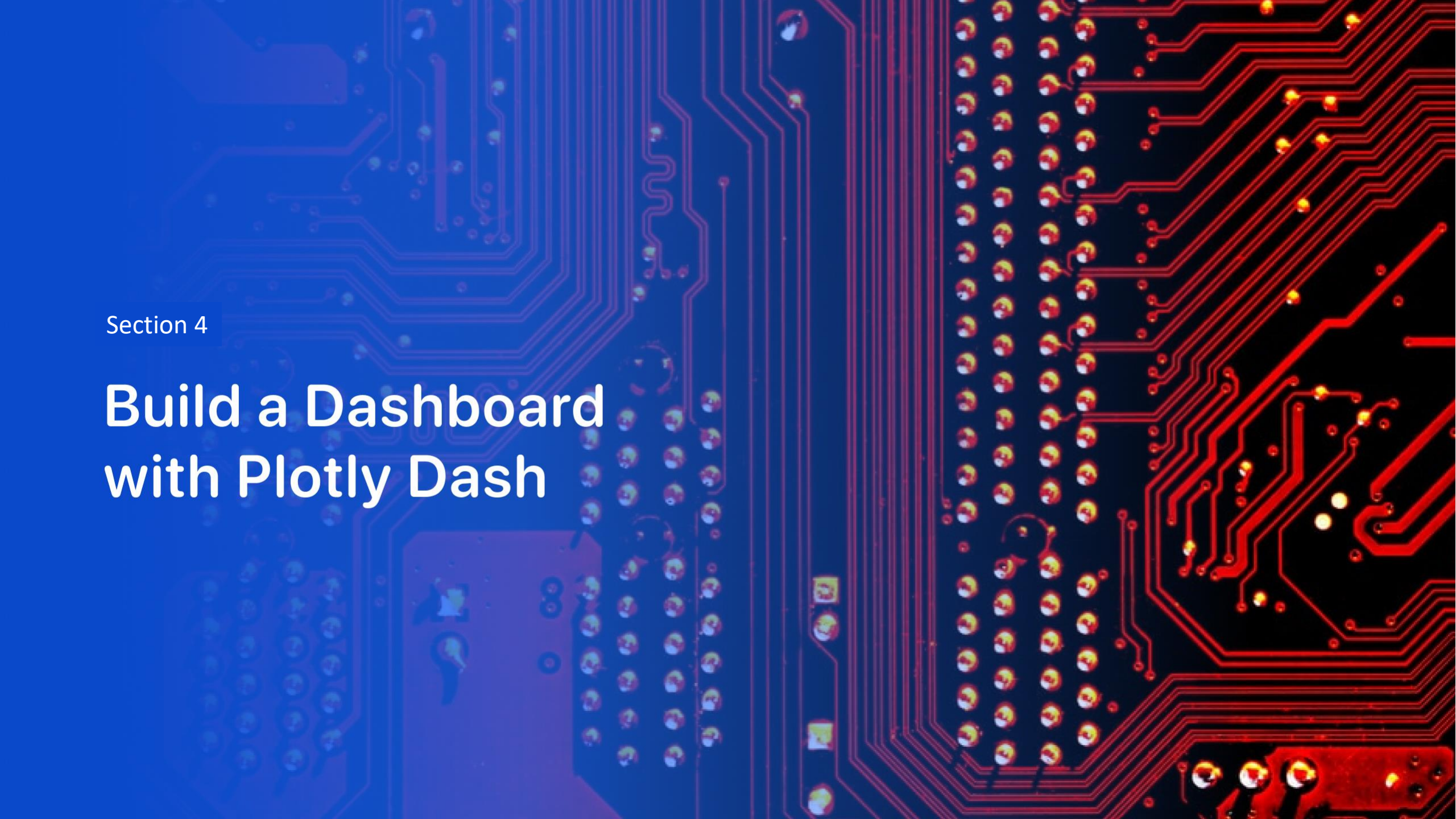
- All Launch Sites are marked on the map

Mark the success/failed launches for each site on the map

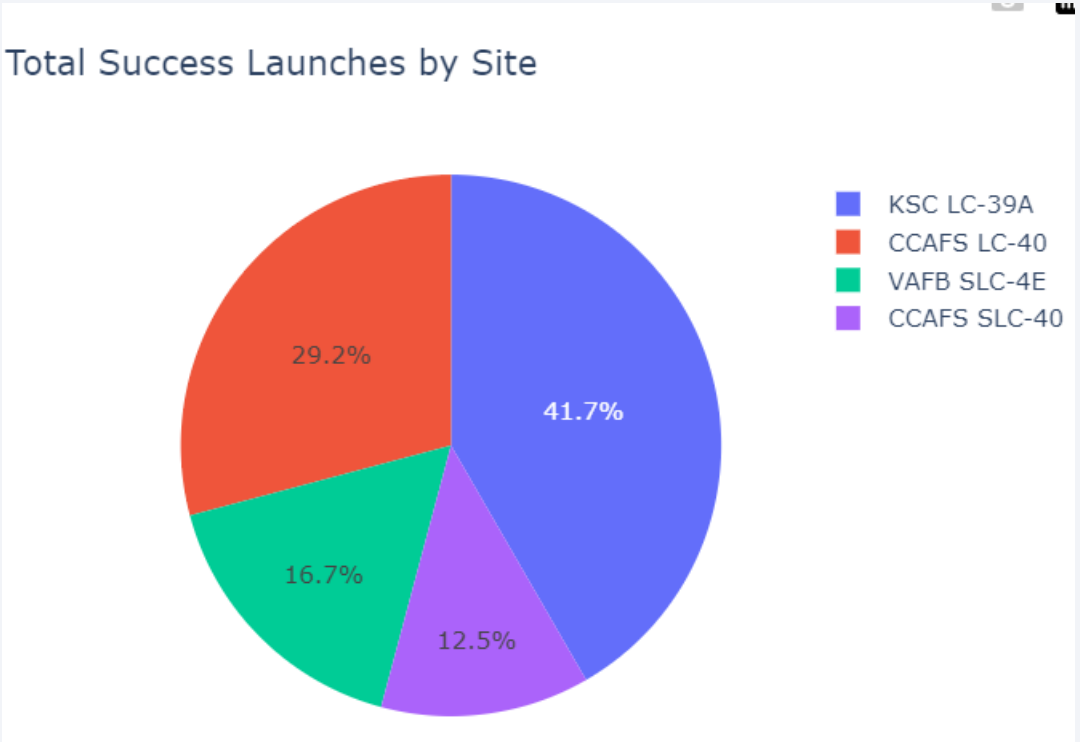
- color-labeled launch outcomes are plotted on the map

Calculate the distances between a launch site to its proximities

- Launch site to its proximities such as railway, highway, coastline, with distance calculated are displayed

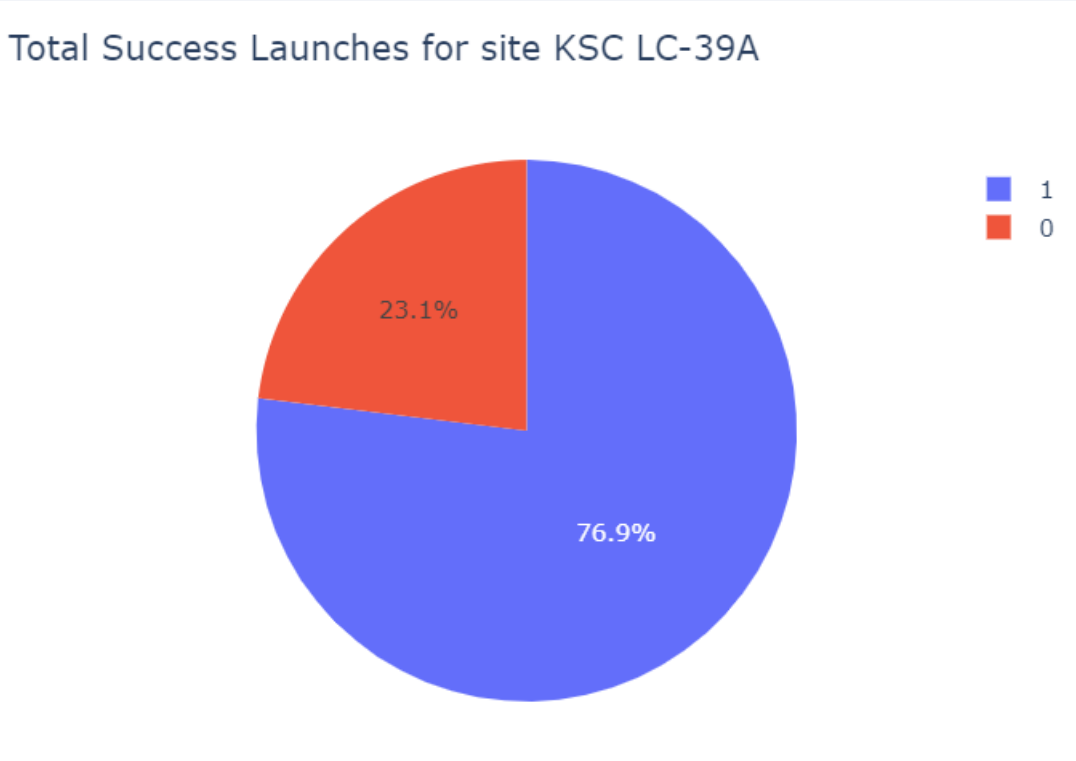

Section 4
Build a Dashboard with Plotly Dash

Total Successful launches by site



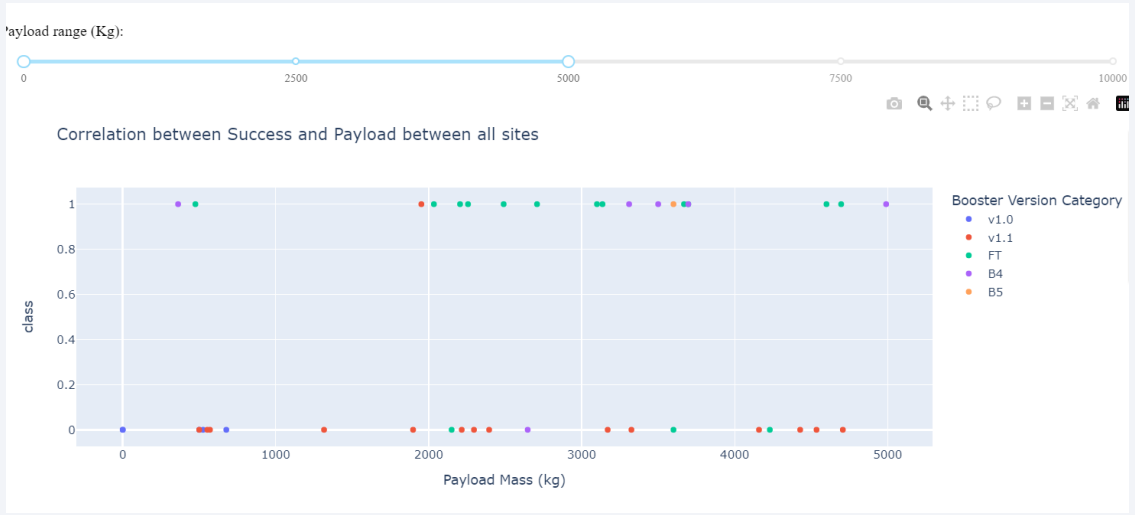
- KSC LC-39 A has the highest success rate in launches

Highest Success Ratio



- KSC LC 39A has the highest success ratio

Payload vs launch outcome

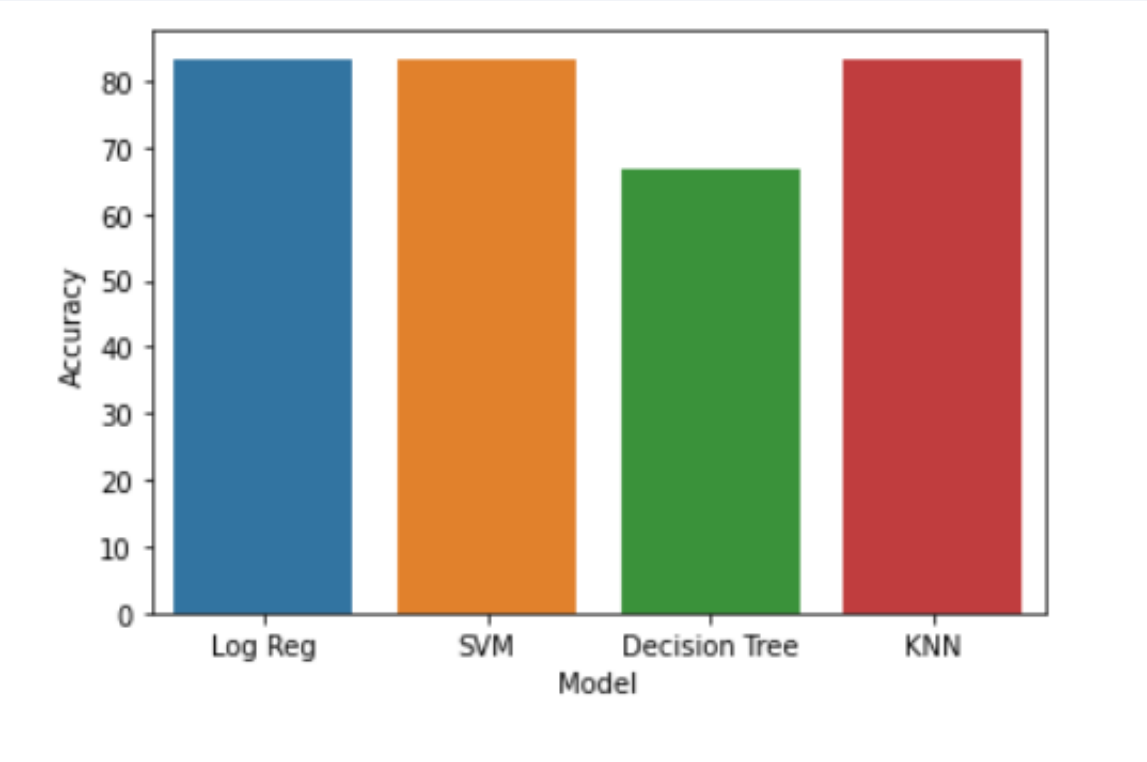


Smaller payloads are more successful than heavier payloads

Section 5

Predictive Analysis (Classification)

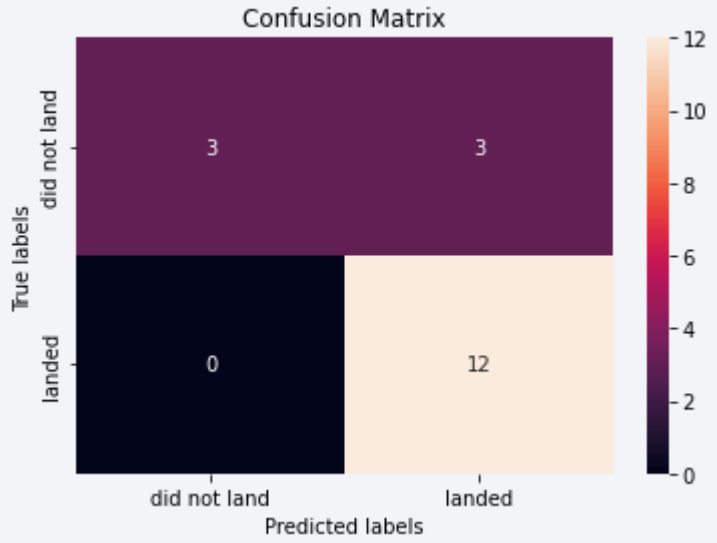
Classification Accuracy



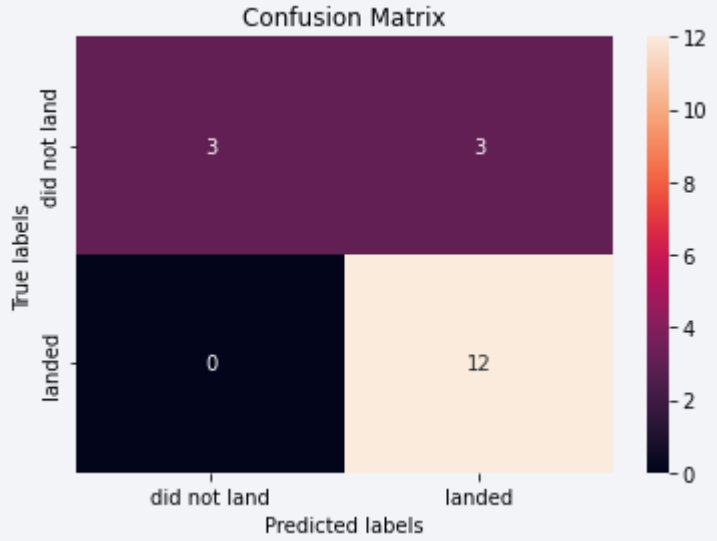
SVM, Logistic Regression and KNN have the highest accuracy

Confusion Matrix

LR

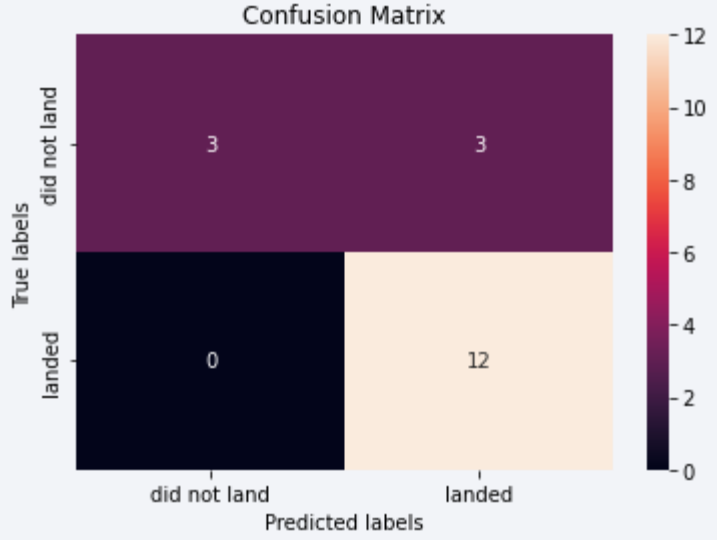


SVM



DT

KNN



Conclusions

- CCAFS SLC 40 has the highest number of launches and majority of the small payload launches.
- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.
- VLEO Orbit launches are more frequent in the recent years and have been
- Success rate increases with time
- KSC LC-39 A has the highest launch success rate and success ratio.
- Smaller payloads are more successful than heavier payloads
- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.

Thank you!