IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Parvathi Mullapudi>
<11/09/2024>

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection Methodology
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Interactive Visual Analytics and Dashboards
  - Predictive Analysis

Introduction

- Project background and context
  - Sending spacecraft to the International Space Station
  - a satellite internet constellation providing satellite Internet access
  - Sending manned missions to Space.
- Problems you want to find answers
  - determine the cost of a launch
  - bring the payload to orbit
  - You will do this by gathering information about Space X and creating dashboards for your team

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:** SpaceX API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- **Perform data wrangling:** Review the attributes in first stage like Flight Number, Date, Booster version. The column Outcome indicates if the first stage successfully landed.Landing outcomes to be converted to Classes (either 0 or 1). 0 is a bad outcome, that is, the booster did not land. 1 is a good outcome, that is, the booster did land.
- **Perform exploratory data analysis (EDA) using visualization and SQL:** Exploratory Data Analysis is the first step of any data science project. In the first lab, you will perform some Exploratory Data Analysis using a database. In the second lab, you will see if the data can be used to automatically determine if the Falcon 9's first stage will land.

Methodology

- **Perform interactive visual analytics using Folium and Plotly Dash:** we are using Folium and Plotly Dash to build an interactive map and dashboard to perform interactive visual analytics. The first part of this module will be focused on analyzing launch site geo and proximities with Folium. We first mark the launch site locations and their close proximities on an interactive map. Then, we can explore the map with those markers and try to discover any patterns from them. Finally, we should be able to explain how to choose an optimal launch site. Next, you will be building a dashboard application with the Python Plotly Dash package.
- **Perform predictive analysis using classification models:** Preprocessing, allow to standardize our data, and Train_test_split, allow to split our data into training and testing data, We will train the model and perform Grid Search, allowing us to find the hyperparameters that allow a given algorithm to perform best. Using the best hyperparameter values, we will determine the model with the best accuracy using the training data. You will test Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors. Finally, we will output the confusion matrix.

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - %sql select distinct(LAUNCH_SITE) from SPACEXTBL
  - %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
  - %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
  - %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
  - %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'%sql select Booster_Version from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
  - %sql select count(Mission_Outcome) from SPACEXTBL WHERE Mission_Outcome = 'Success' or Mission_Outcome = 'Failure (in flight)'

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - %sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
  - %sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,7,4) = '2015'
  - %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'%sql select Booster_Version from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
  - %sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN '04-06-2010' AND '20-03-2017' GROUP BY Landing_Outcome ORDER BY Numbers DESC;
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-09ab93de324a8be291b8904715d3d5eec692a4c8

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

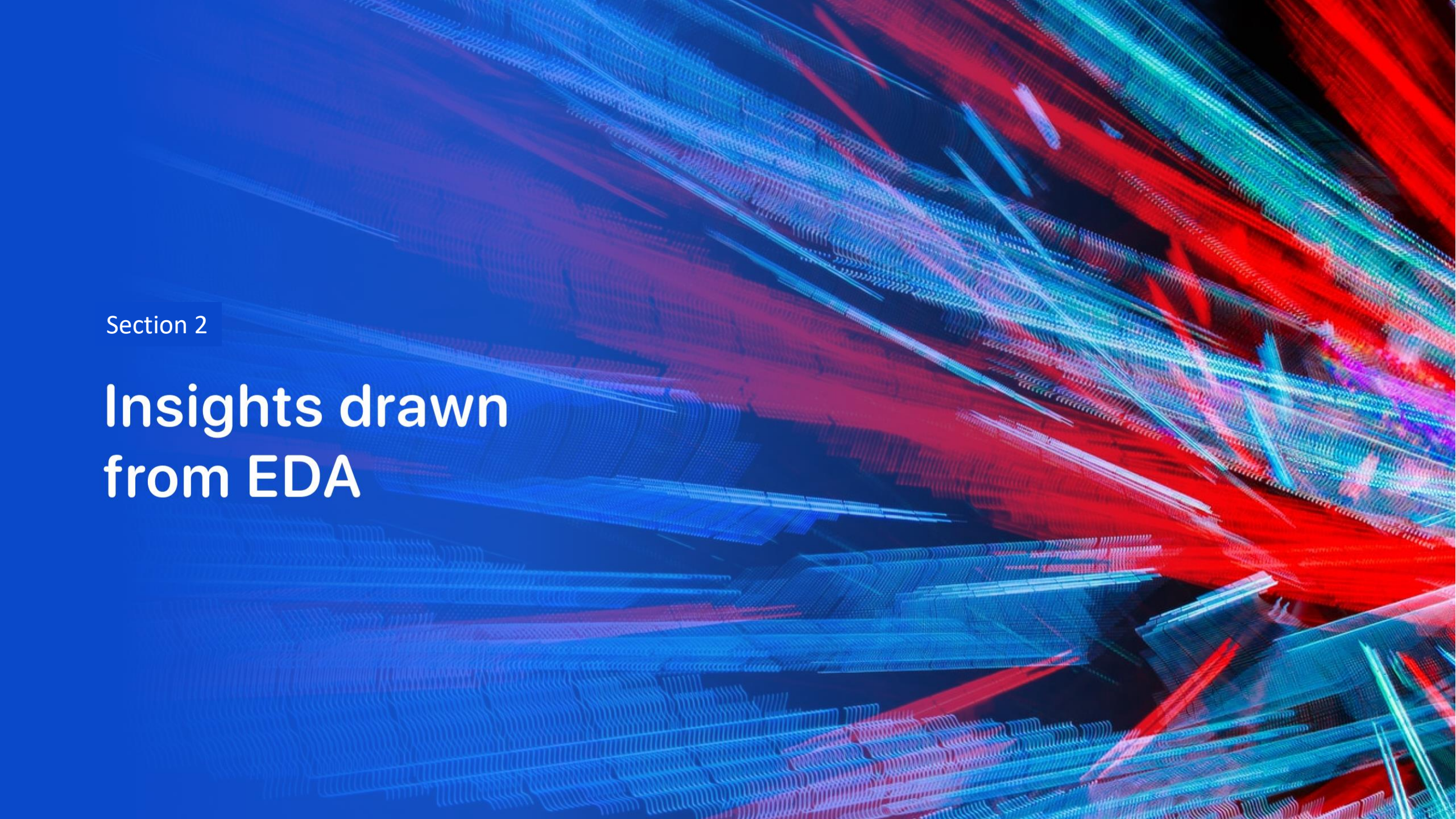
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

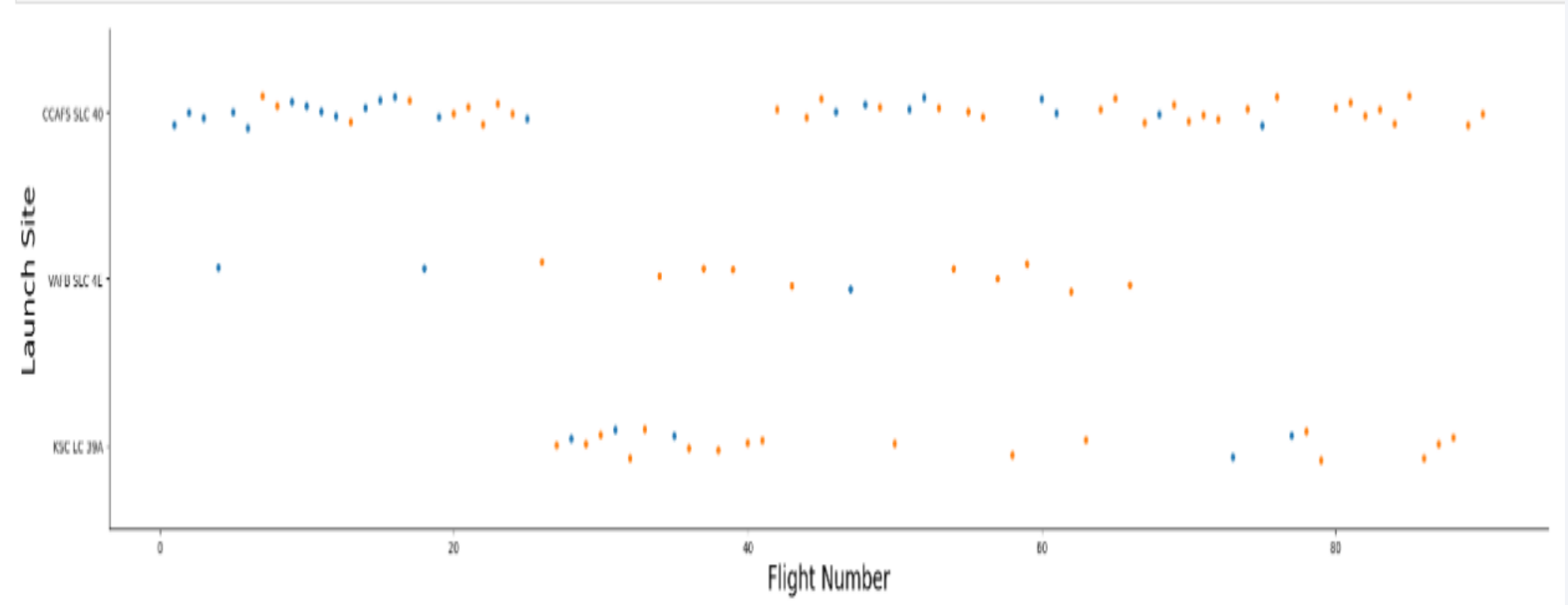
Results

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


Section 2
Insights drawn from EDA

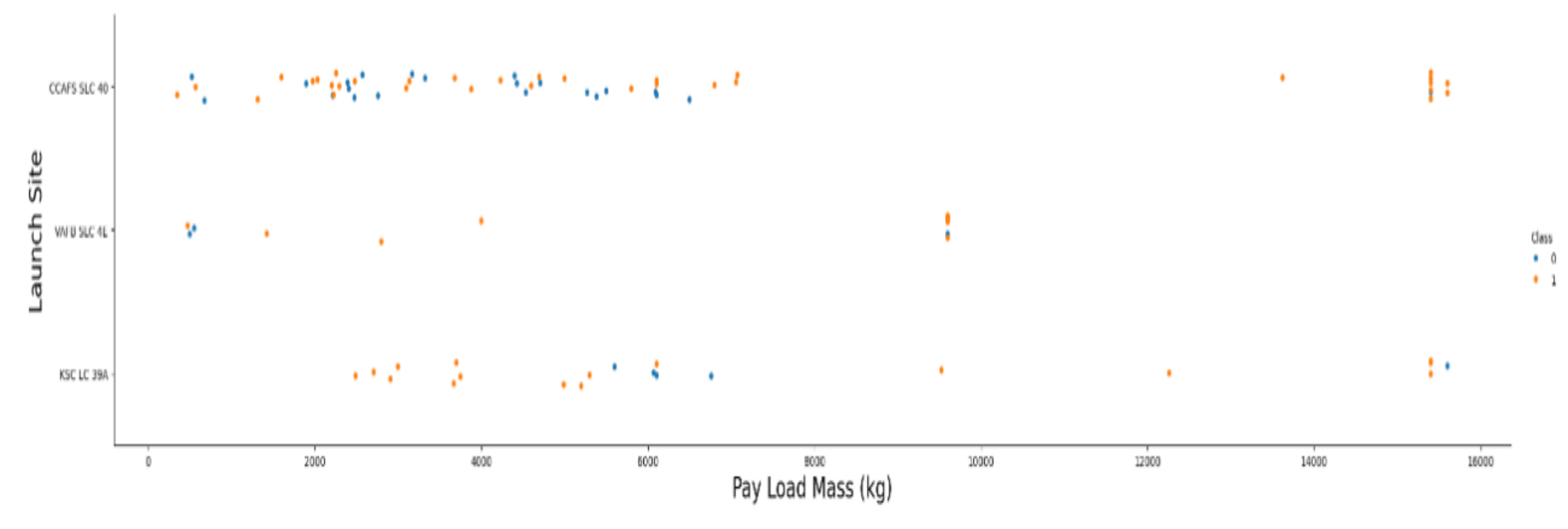
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site



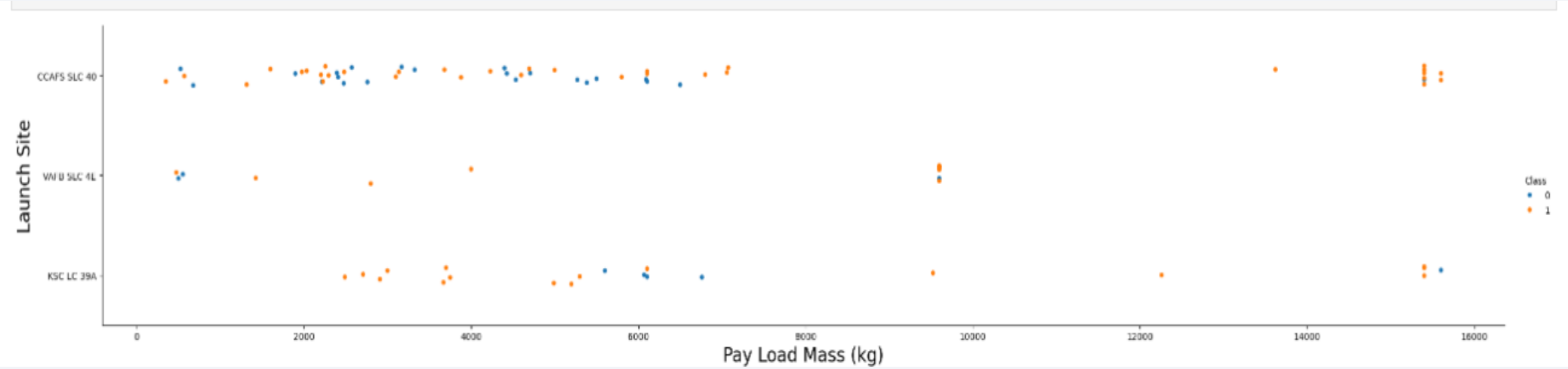
Flight Number vs. Launch Site

Show the screenshot of the scatter plot with explanations



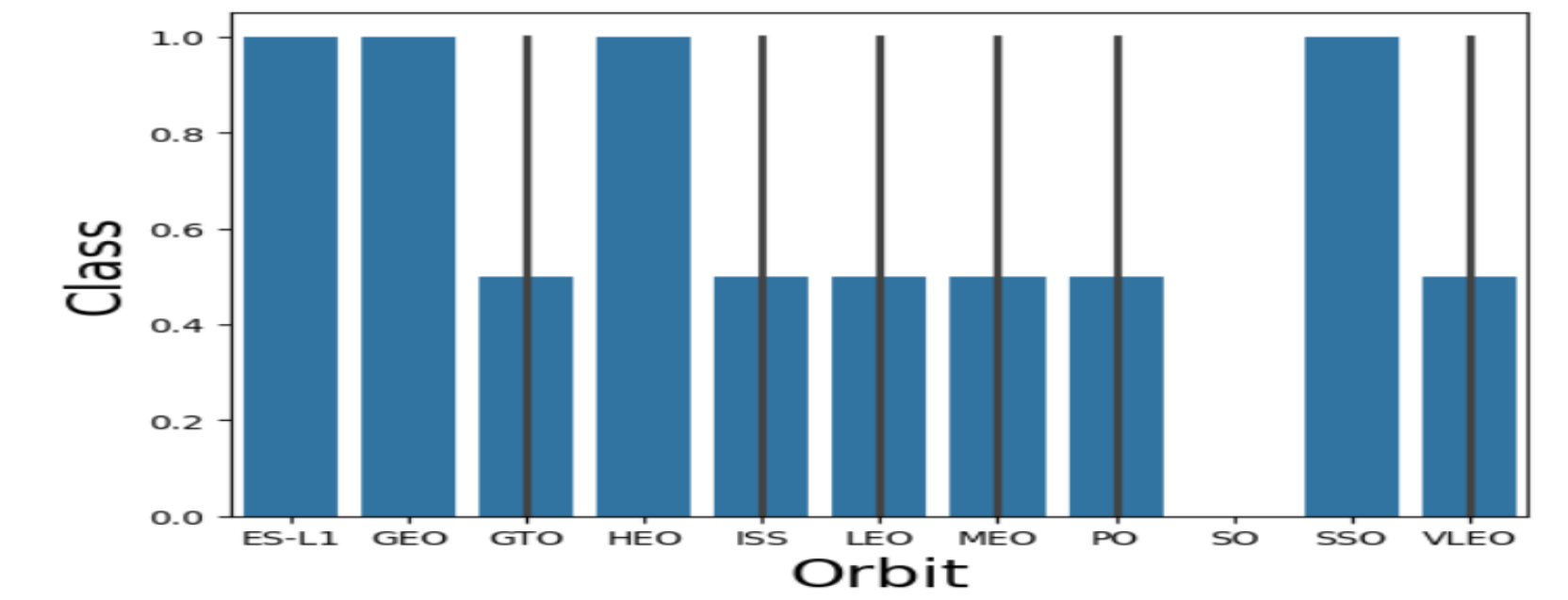
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



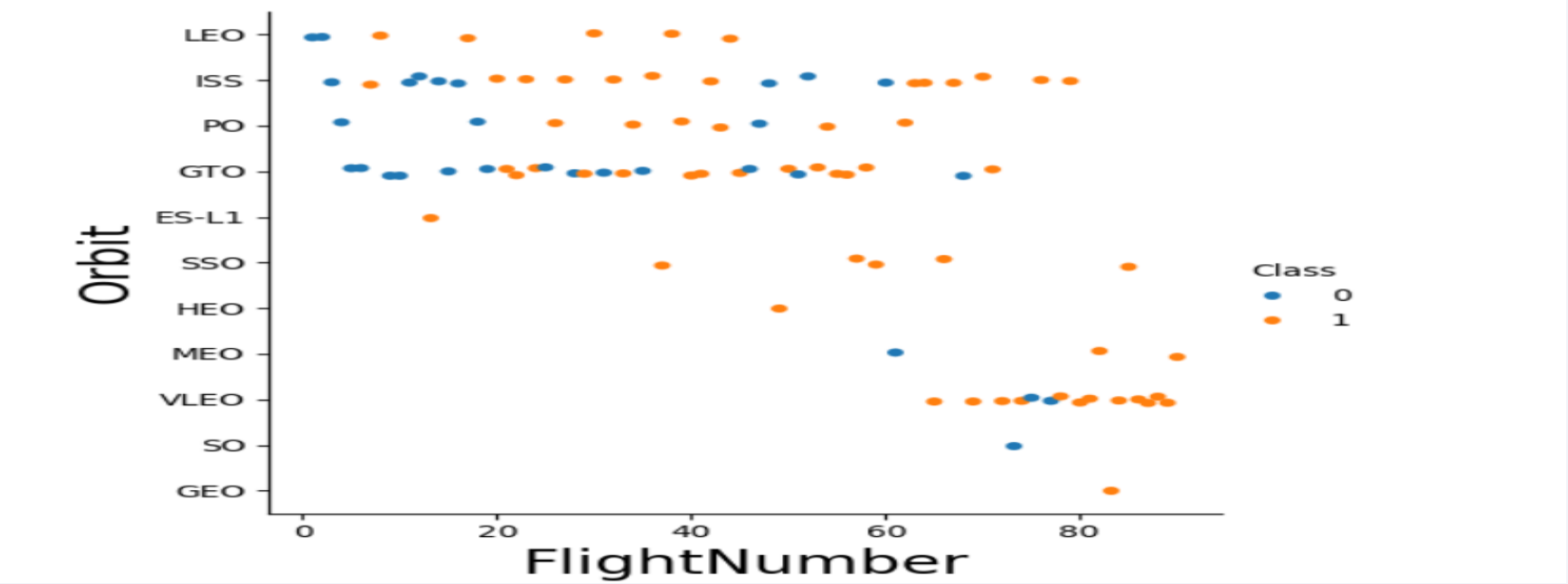
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type



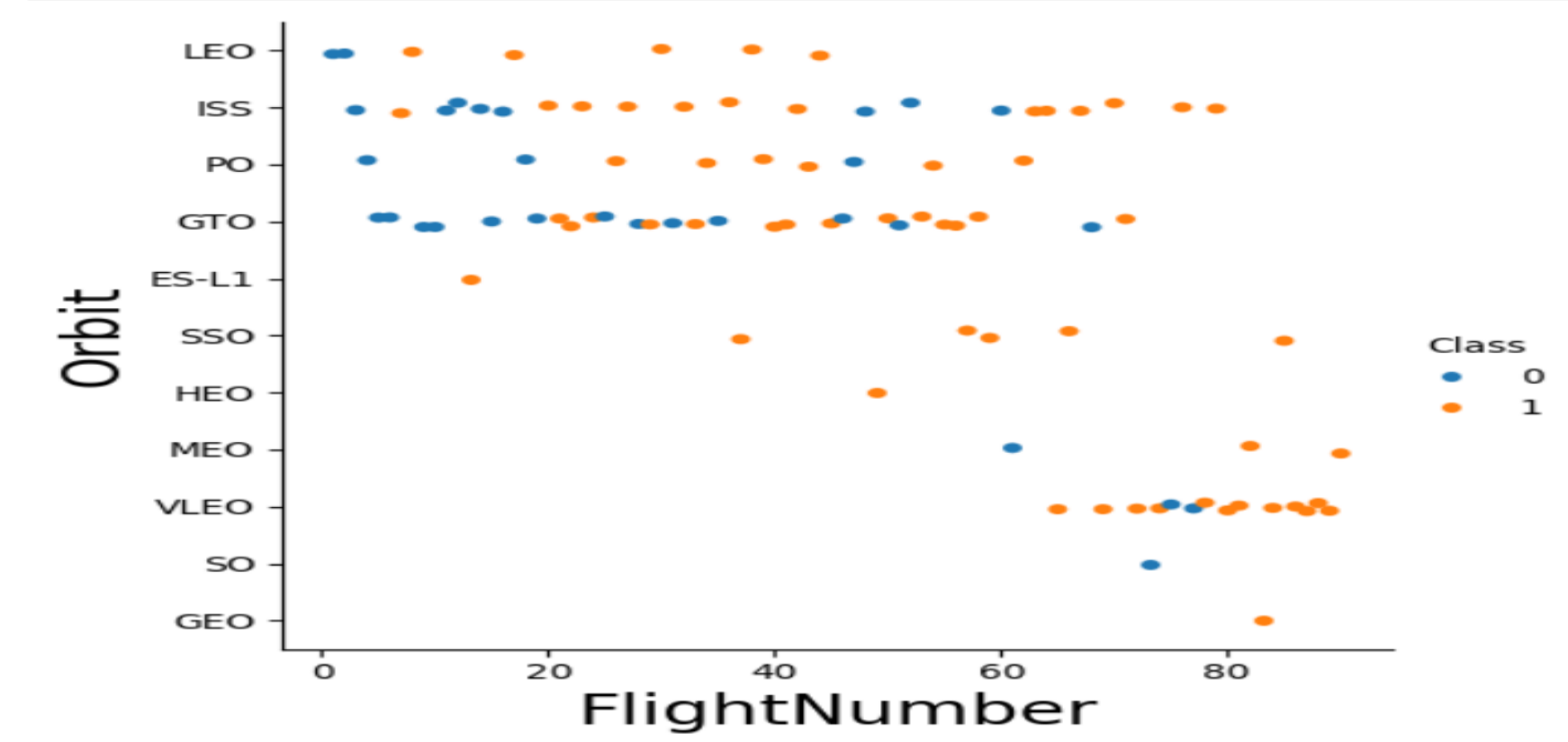
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type



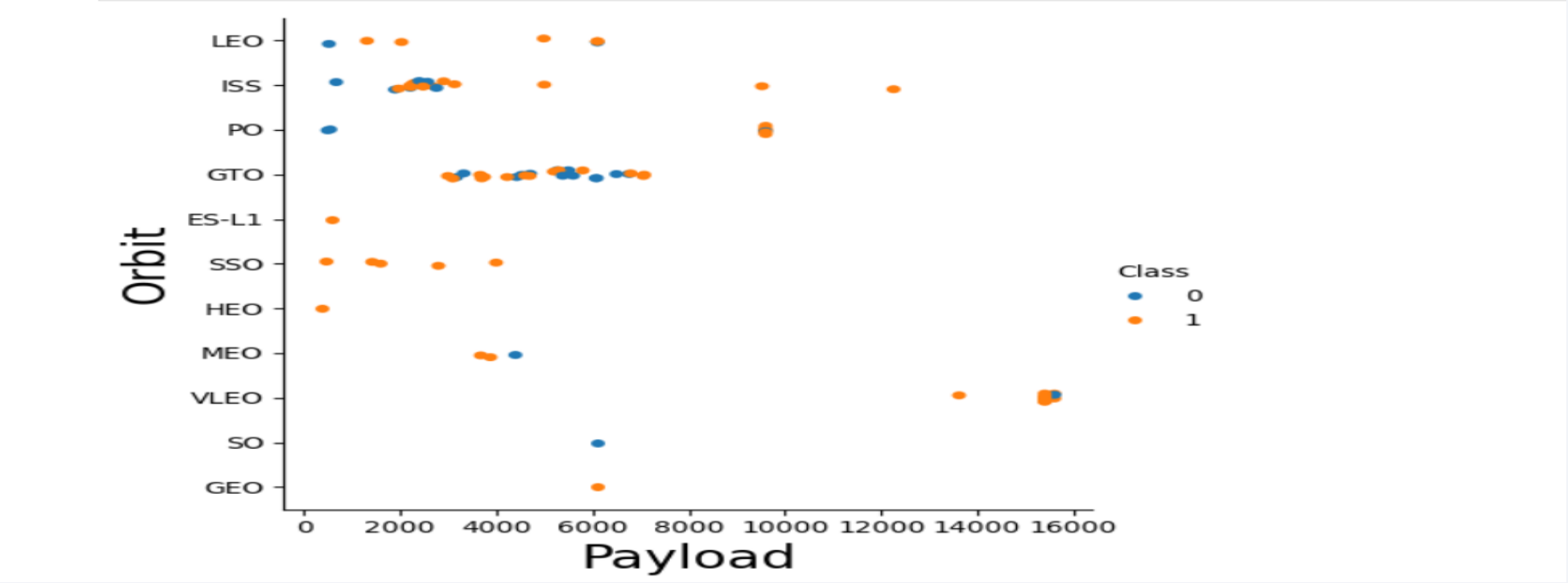
Flight Number vs. Orbit Type

Show the screenshot of the scatter plot with explanations



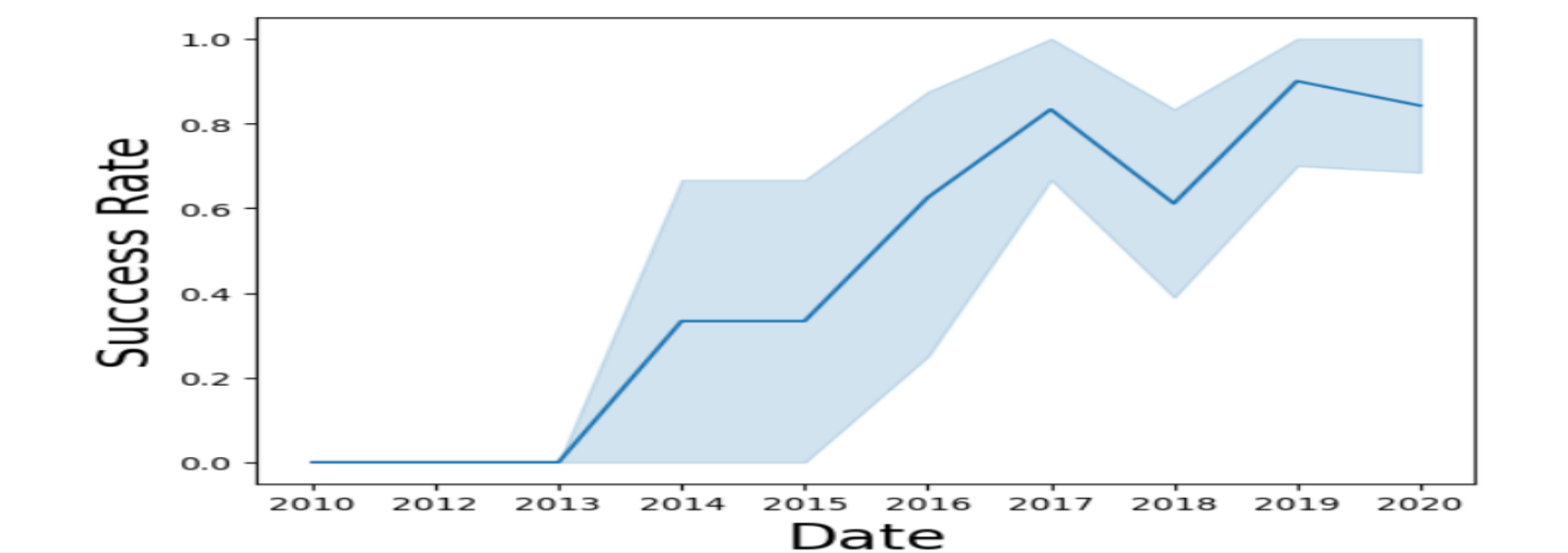
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Show a line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch sites

one.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

- Present your query result with a short explanation here

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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

- Present your query result with a short explanation here

```
%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

sum(PAYLOAD_MASS__KG_)
45596

- Present your query result with a short explanation here

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
 * sqlite:///my_data1.db
Done.
```

Average Payload Mass by F9 v1.1

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

avg(PAYLOAD_MASS__KG_)
2928.4

- Present your query result with a short explanation here

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
 * sqlite:///my_data1.db
Done.
```

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

min(DATE)
2015-12-22

- Present your query result with a short explanation here

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

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

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

- Present your query result with a short explanation here

```
[14]: %sql select Booster_Version from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ <
 * sqlite:///my_data1.db
```

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

15]:

count(Mission_Outcome)
99

- Present your query result with a short explanation here

```
%sql select count(Mission_Outcome) from SPACEXTBL WHERE Mission_Outcome = 'Success' or Mission_Outcome = 'Failure (in flight)'
 * sqlite:///my_data1.db
```

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- Present your query result with a short explanation here

Task 6

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

[16]:
```
%sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

2015 Launch Records

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

Month	Booster_Version	Launch_Site
10	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- Present your query result with a short explanation here

```
%sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date
```

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

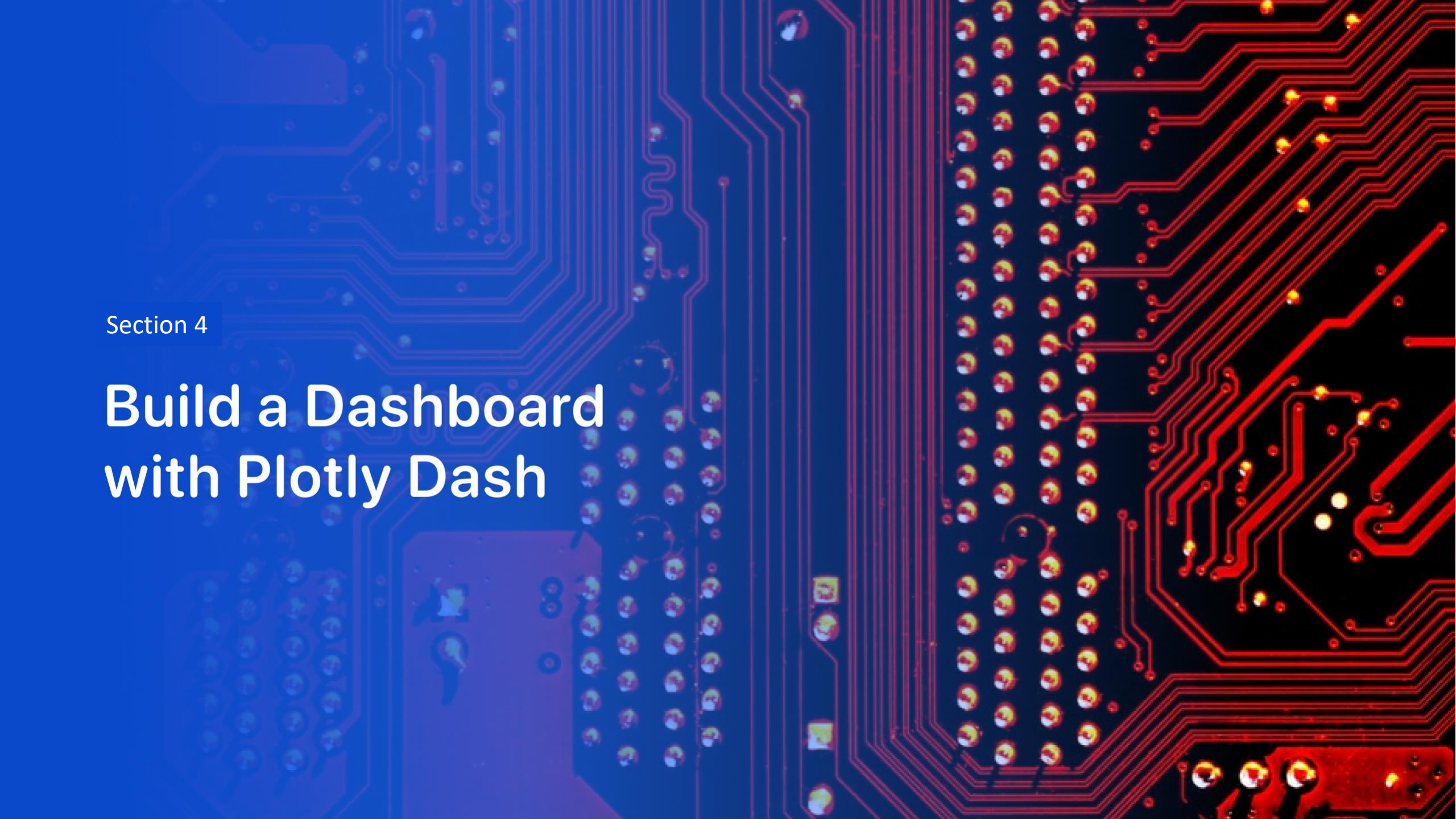
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Done.

Landing_Outcome	Numbers
Success	20
Success (drone ship)	8
Success (ground pad)	7

- Present your query result with a short explanation here

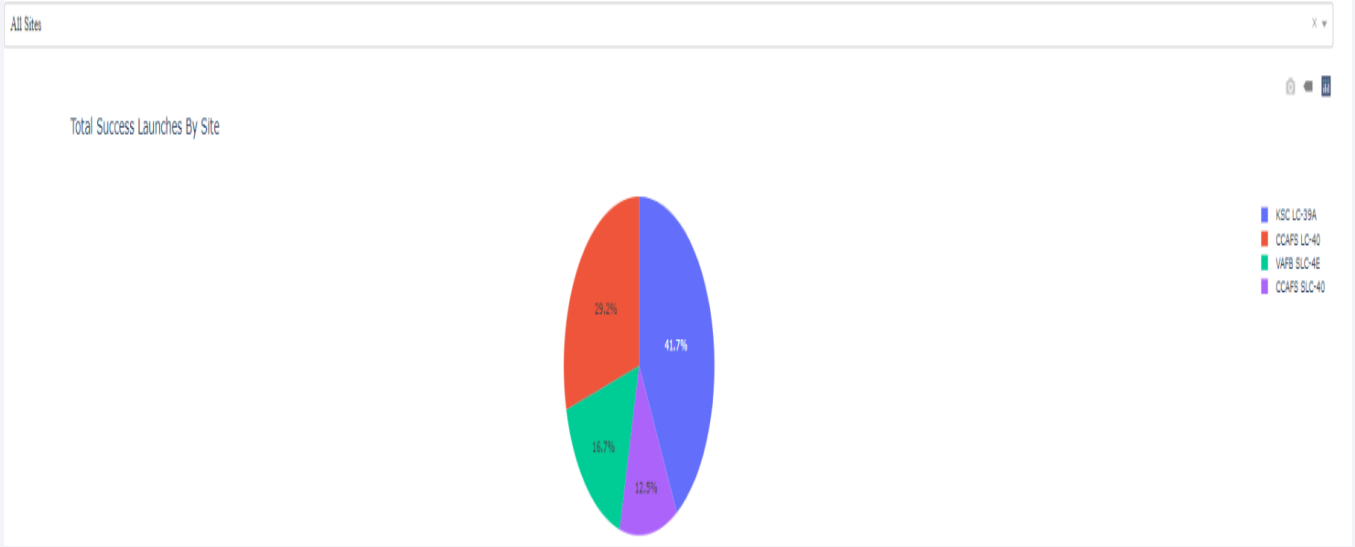
```
%sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN '04-06-2010' AND '20-03
```


Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

- Replace SpaceX Launch Records Dashboard title with an appropriate title

- Show the screenshot of launch success count for all sites, in a piechart



Section 5

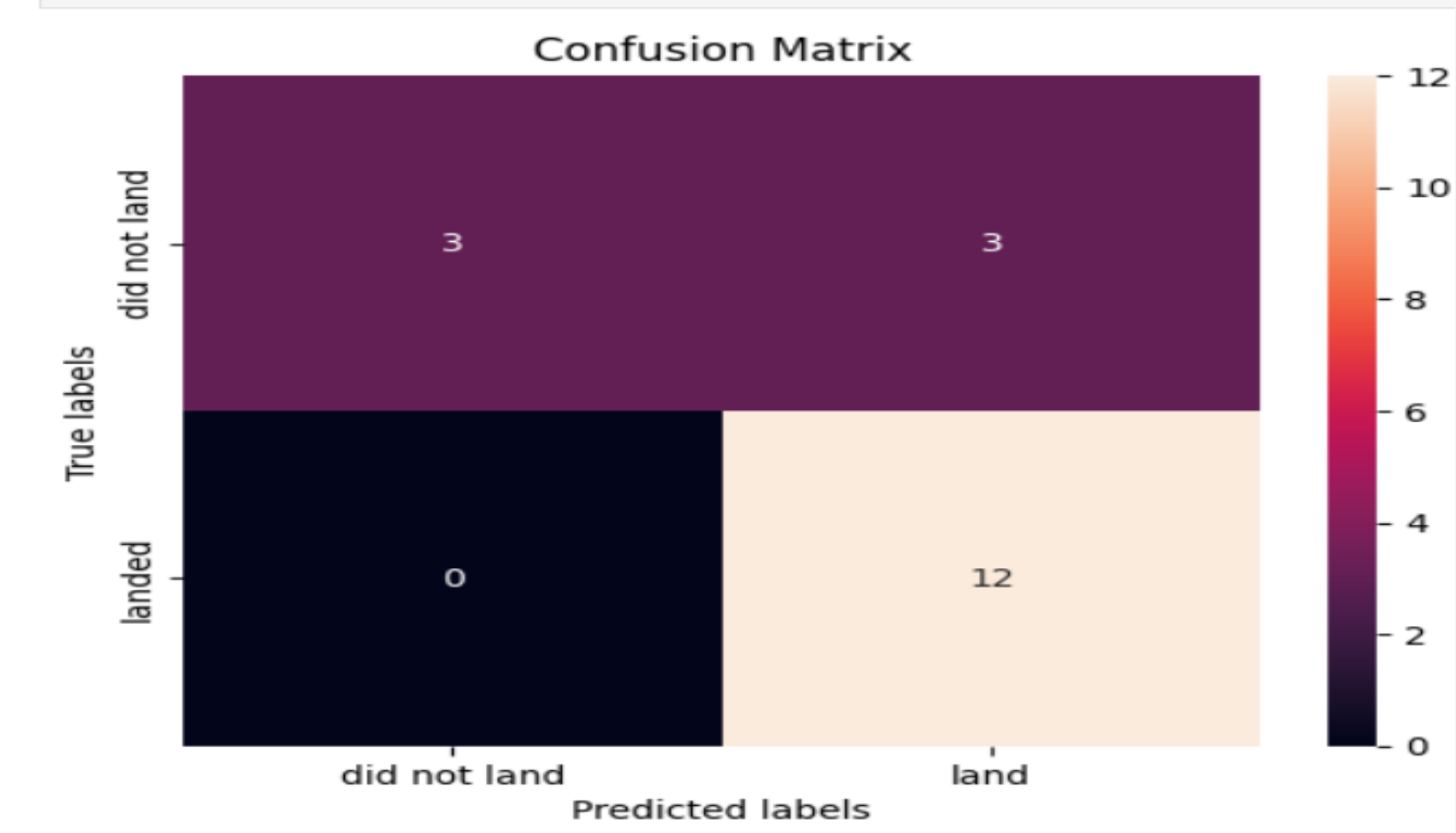
Predictive Analysis (Classification)

Classification Accuracy

- Find which model has the highest classification accuracy

  Seems all models gives same result

Confusion Matrix



Thank you!