IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- The project analyzes data on spacecraft landings
- We want to investigate what factors influence the success of spacecraft landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was downloaded from the number open sources from net
- Perform data wrangling
  - Data was cleaned and grouped. We made some feature engineering
- 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 used logistic regression, svm, KNN, decision trees algorithms. All models were tuned with cv=10. All scores were getting from test sets

Data Collection

- Data was collected from different spacex apis.
  - `https://api.spacexdata.com/v4/rockets/"`
  - `"https://api.spacexdata.com/v4/payloads/"`
  - `"https://api.spacexdata.com/v4/cores/"`
  - `https://api.spacexdata.com/v4/launches/past`
- Also we used data scrabbing from wiki
- `https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches`

Data Collection – SpaceX API

- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/jupyter-labs-webscraping.ipynb

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- We made Outcome column and clean the data
- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- We plot a several scatter charts to understand data distribution
- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- select distinct Launch_Site from SPACEXTABLE
- select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
- select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer='NASA (CRS)'
- select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version='F9 v1.1'
- select Landing_Outcome,min(date) from SPACEXTABLE group by 1
- select distinct Booster_Version from SPACEXTABLE where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_>4000 and PAYLOAD_MASS__KG_<6000
- select Mission_Outcome,count(*) from SPACEXTABLE group by 1
- select distinct Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
- select substr(Date, 6,2), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015'
- select Landing_Outcome ,count(*) from SPACEXTABLE where date(Date) between '2010-06-04' and '2017-03-20'
- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We build map of landing sites with the marker of success
- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We add piecharts for diff sites
- https://github.com/Andrii-study/Assignment/tree/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/dash

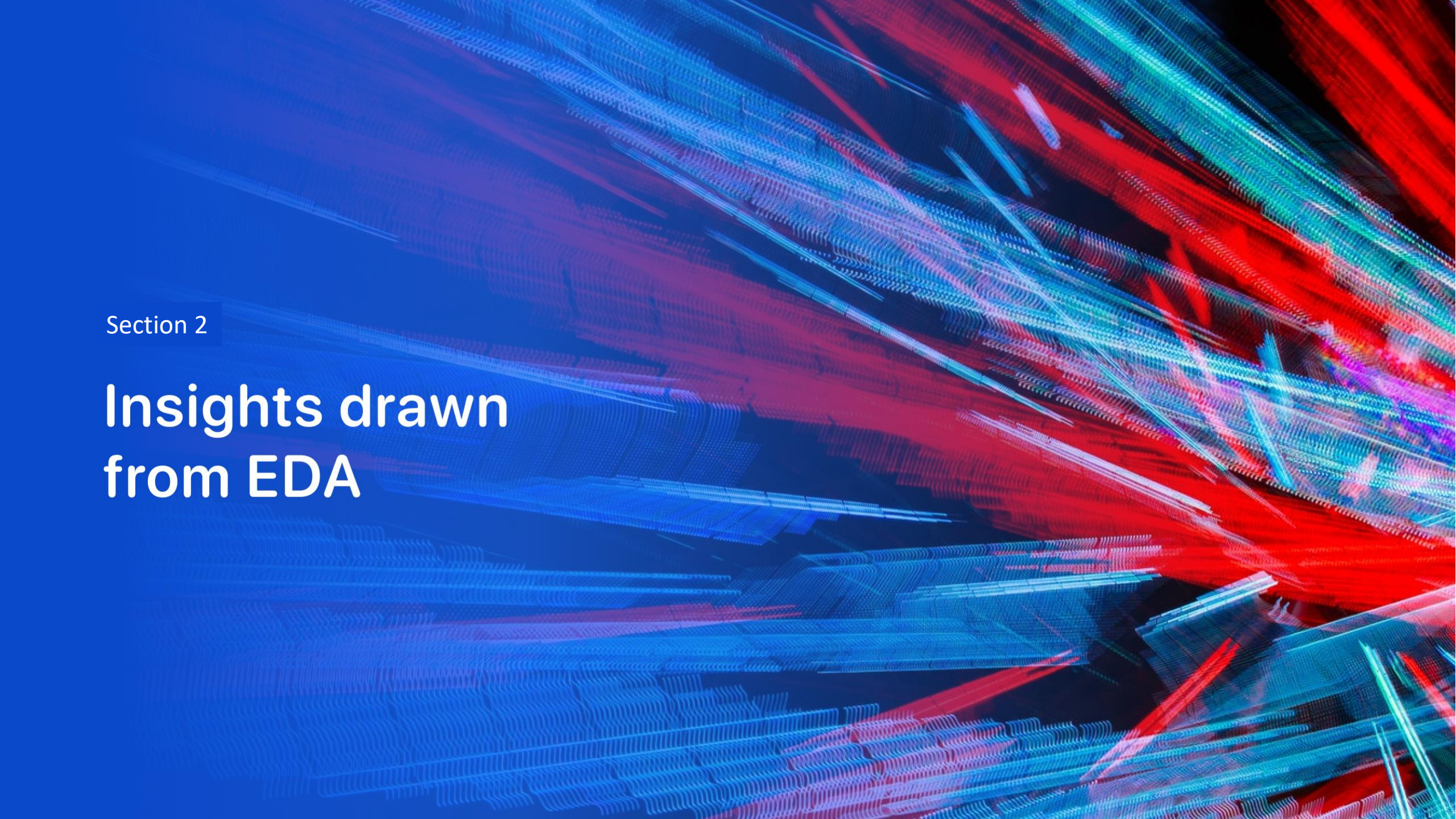
Predictive Analysis (Classification)

- We used 4 different types of models

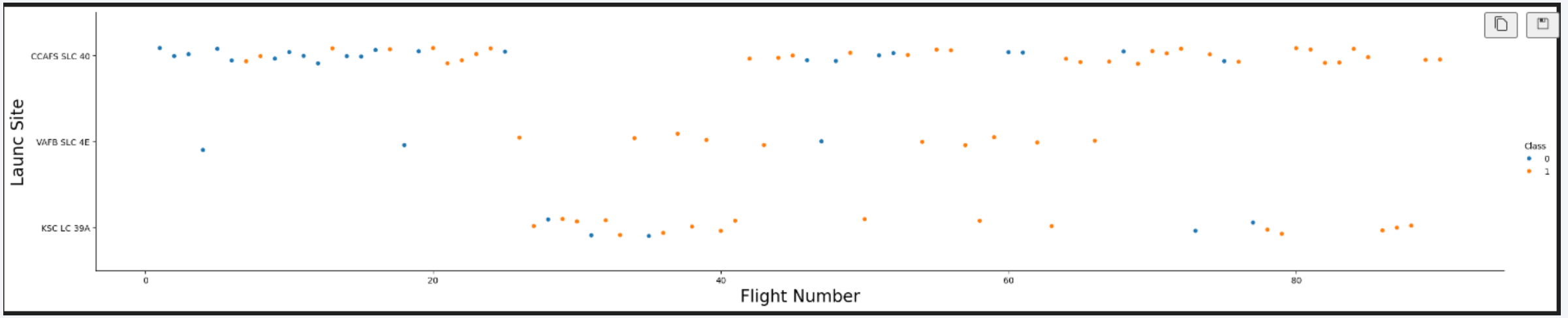
- https://github.com/Andrii-study/Assignment/blob/9476c7f980ab4b188badbf2d7902b46b3f7ec1ad/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

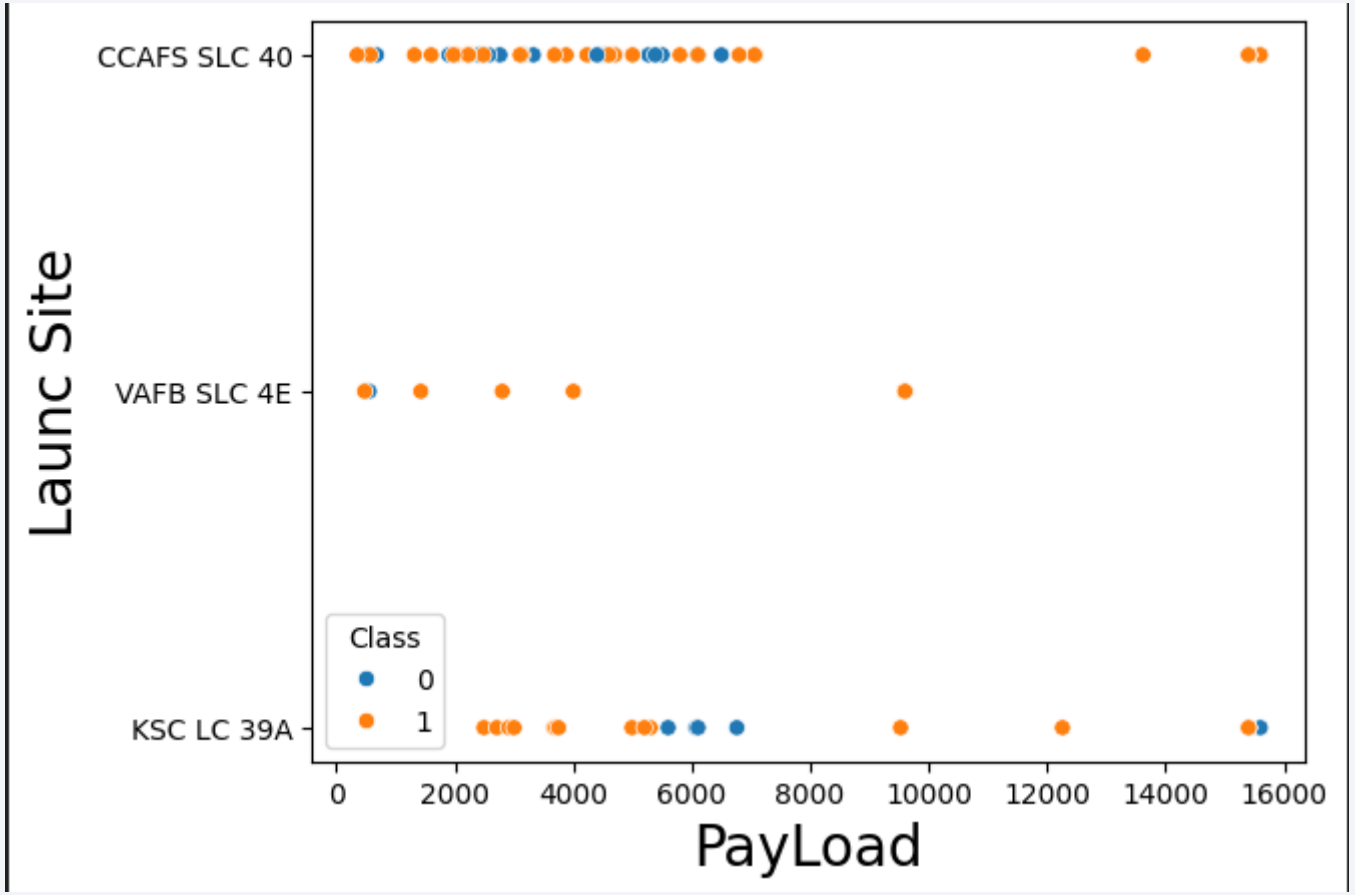
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

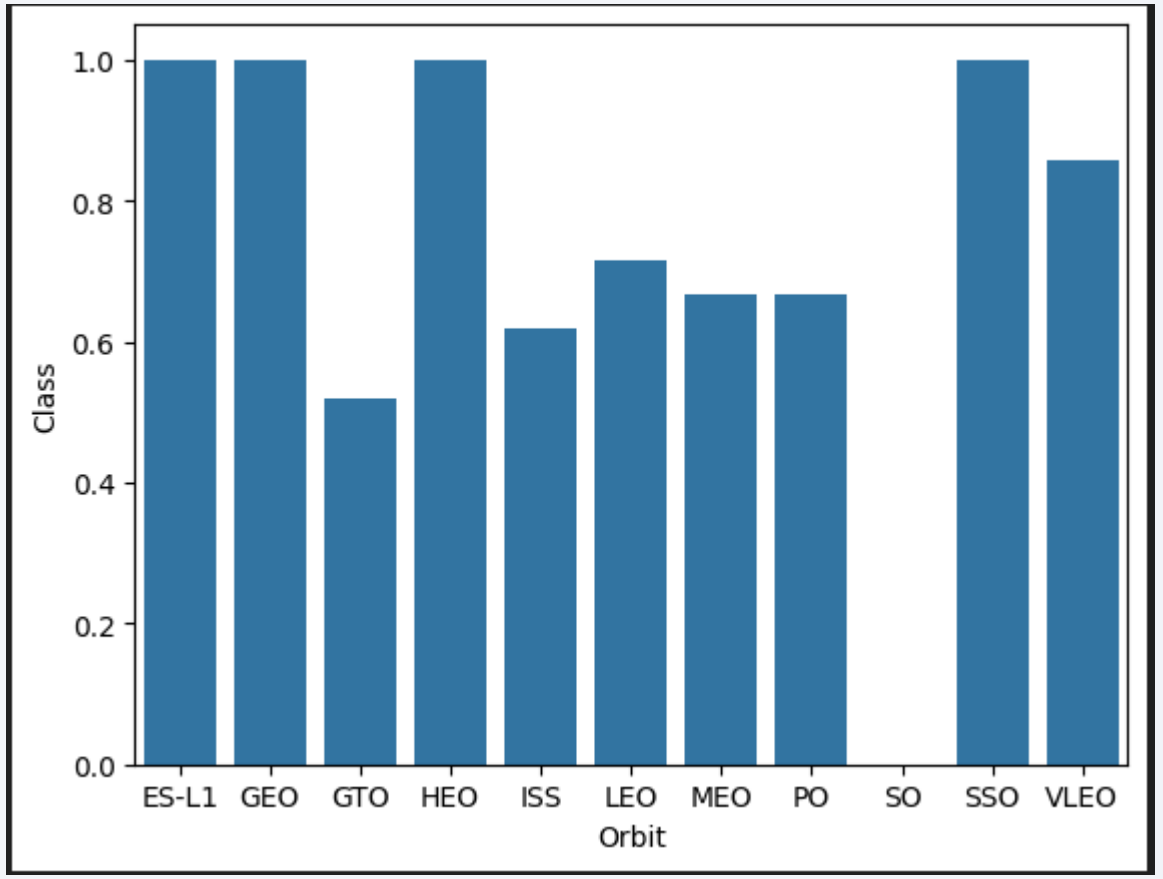
Flight Number vs. Launch Site



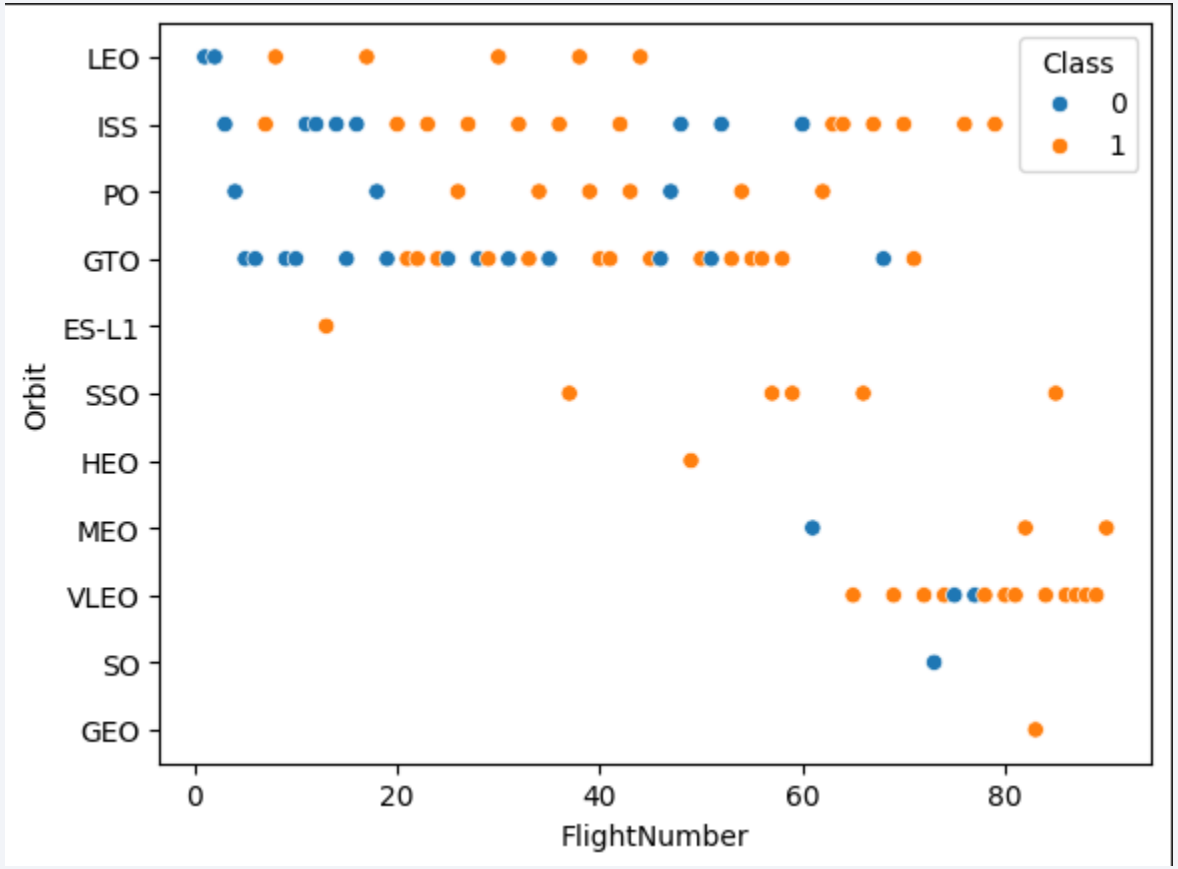
Payload vs. Launch Site



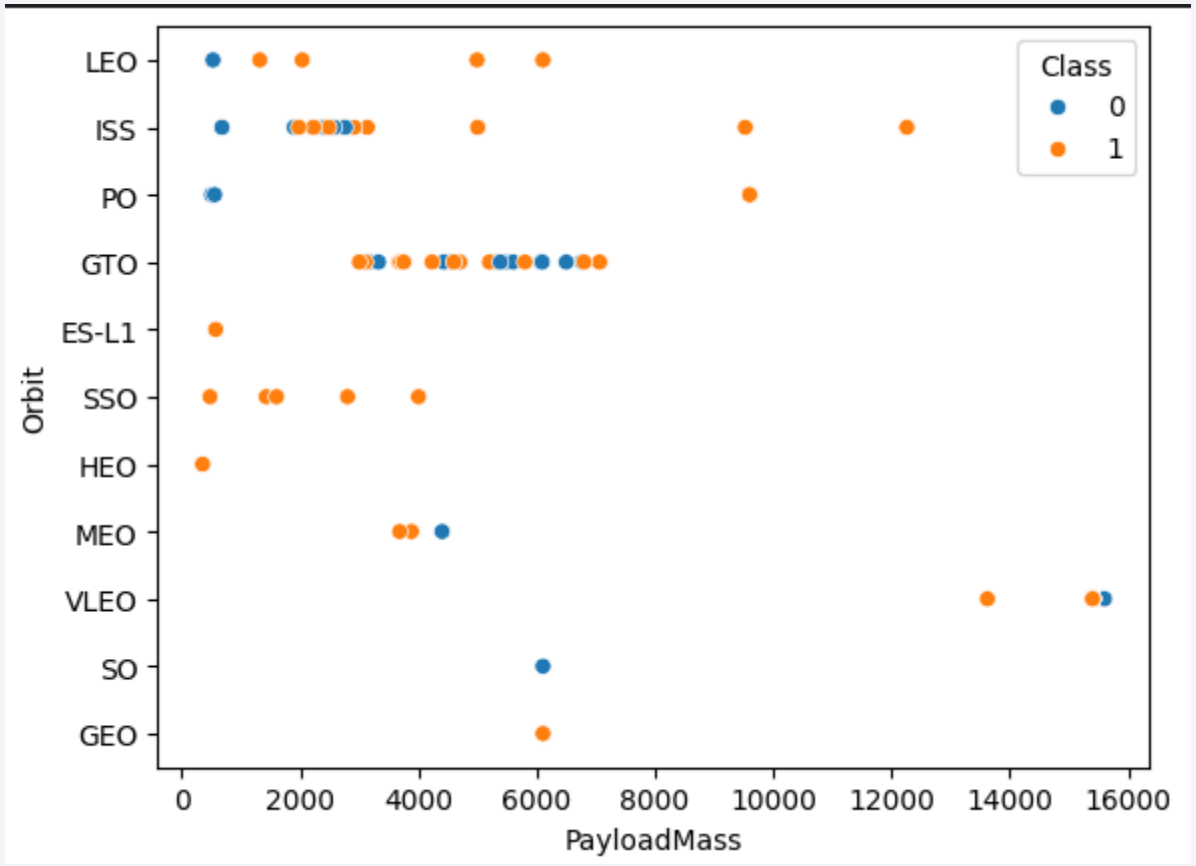
Success Rate vs. Orbit Type



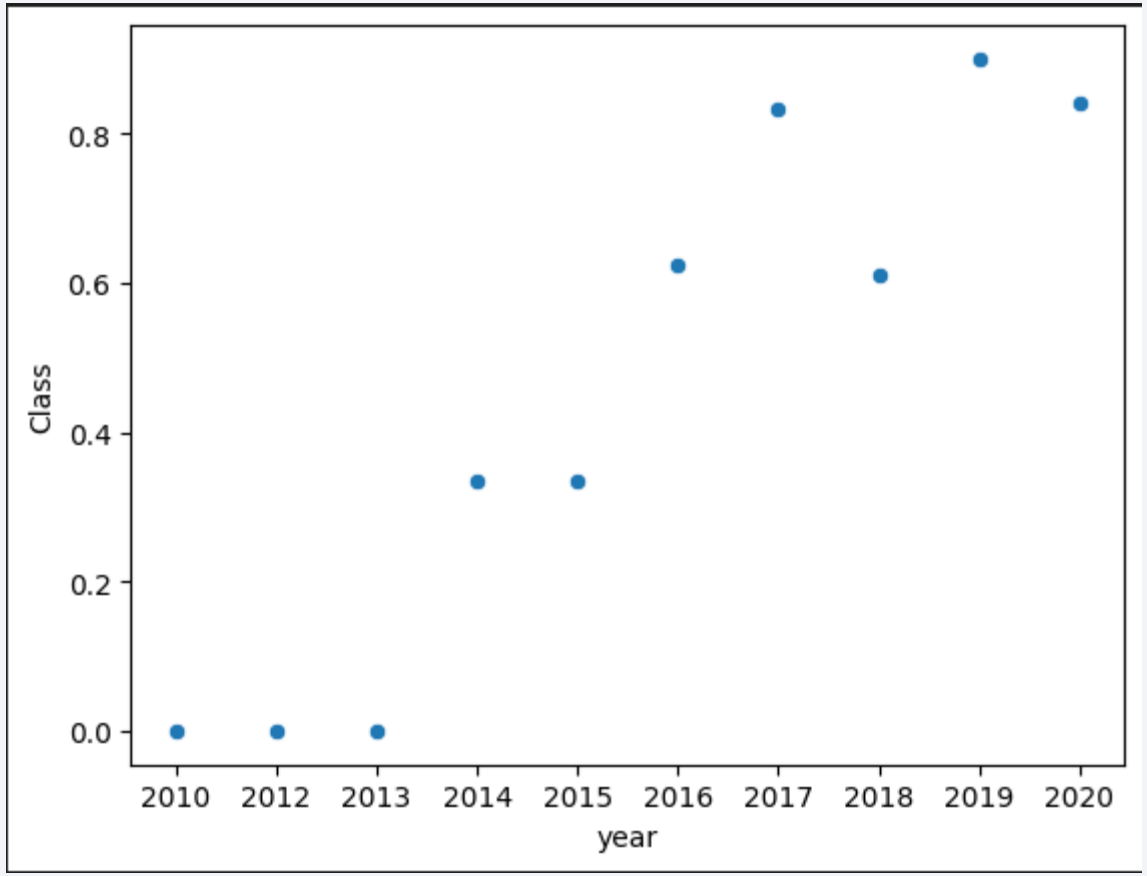
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

	Launch Site
0	CCAFS LC-40
1	CCAFS SLC-40

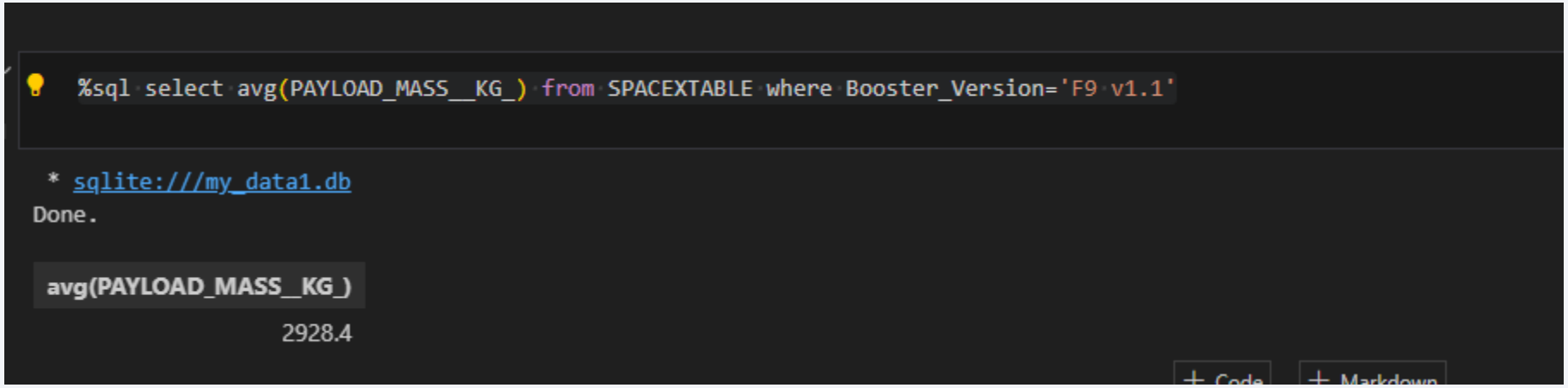
Total Payload Mass

```
%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



First Successful Ground Landing Date

```
%sql select Landing_Outcome,min(date) from SPACEXTABLE group by 1
```

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

Landing_Outcome	min(date)
Controlled (ocean)	2014-04-18
Failure	2018-12-05
Failure (drone ship)	2015-01-10
Failure (parachute)	2010-06-04
No attempt	2012-05-22
No attempt	2019-08-06
Precluded (drone ship)	2015-06-28
Success	2018-07-22
Success (drone ship)	2016-04-08
Success (ground pad)	2015-12-22
Uncontrolled (ocean)	2013-09-29

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

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome,count(*) from SPACEXTABLE group by 1
```

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

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

Boosters Carried Maximum Payload

```
[27] %sql select distinct 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

```
%sql select substr(Date, 6,2), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015'
```

```
 * 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
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

```
%sql select Landing_Outcome ,count(*) from SPACEXTABLE where date(Date) between '2010-06-04' and '2017-03-20'
```

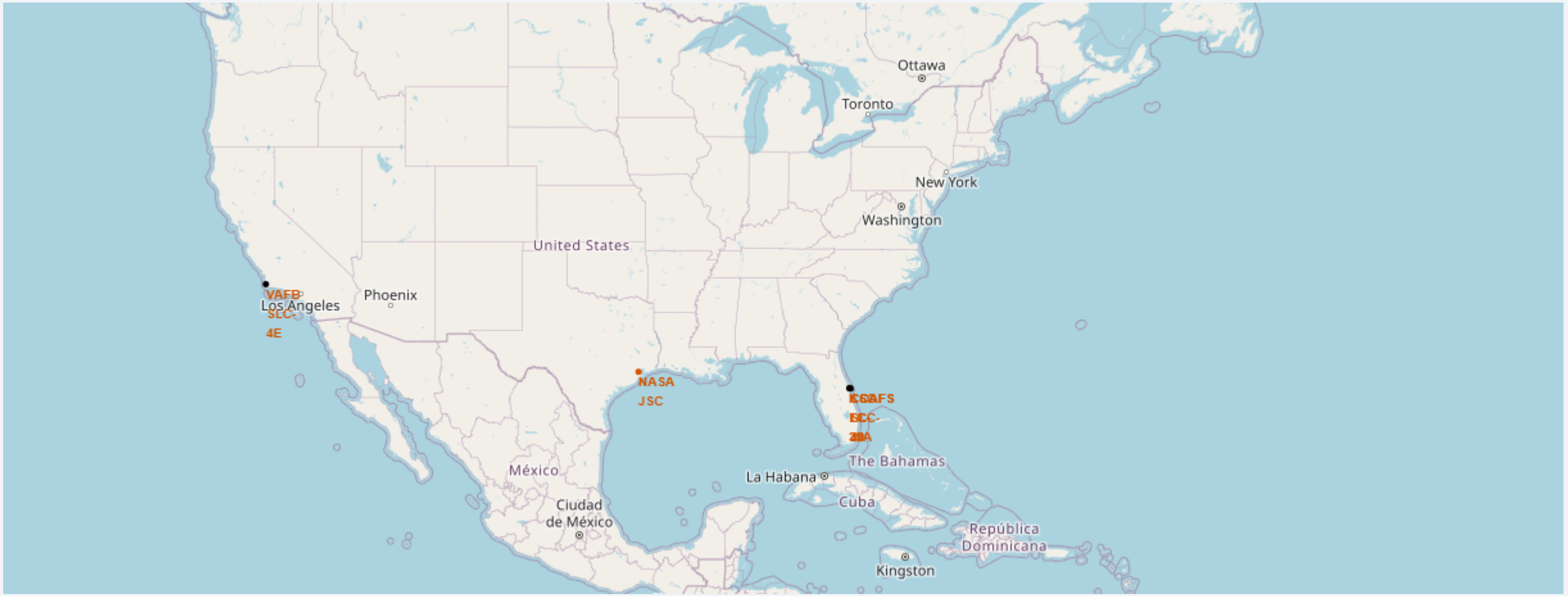
```
 * sqlite:///my_data1.db
Done.
```

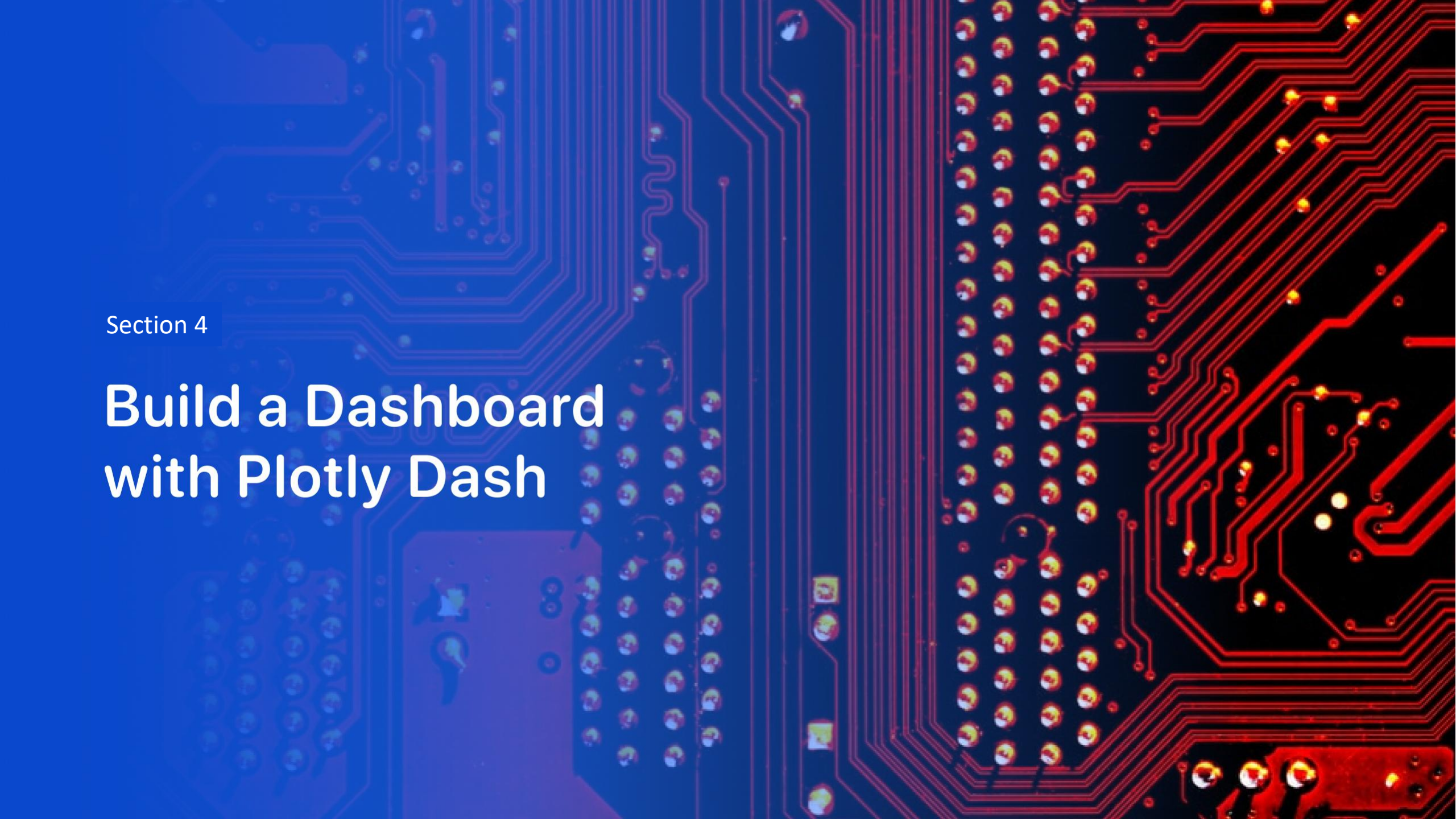
Landing_Outcome	count(*)
Failure (parachute)	31

Section 3

Launch Sites Proximities Analysis

<All launch sites>

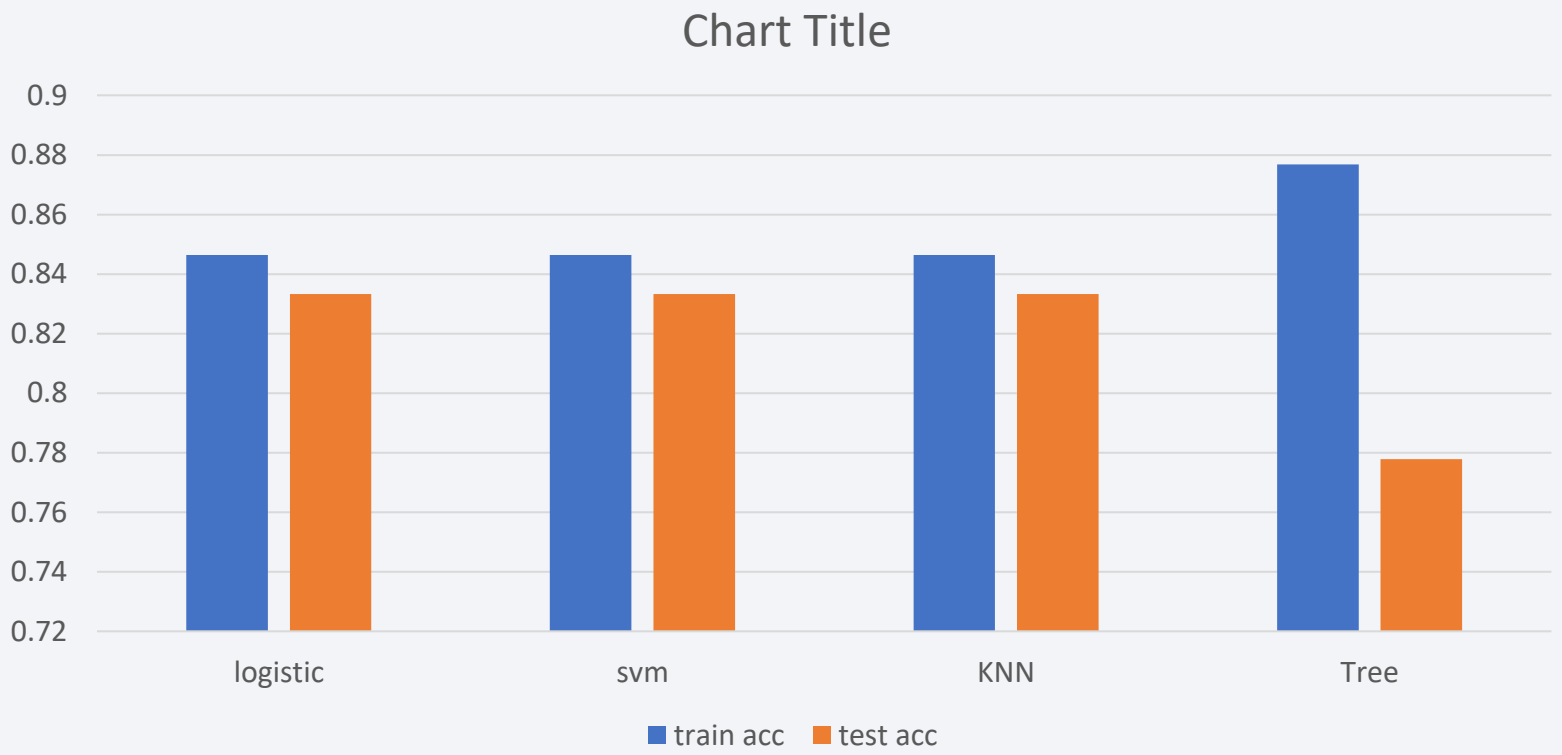



Section 4
Build a Dashboard with Plotly Dash

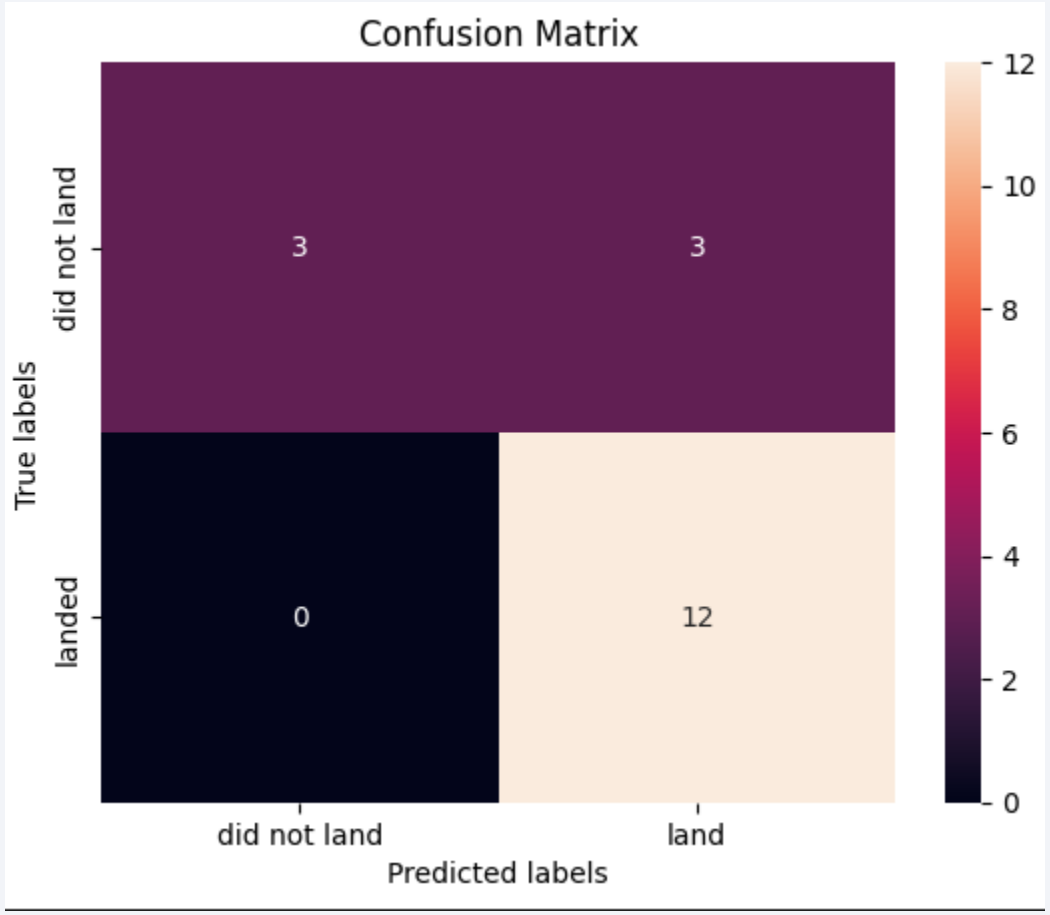
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!