IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrés Tasende Caamaño
18/09/2022

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

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

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

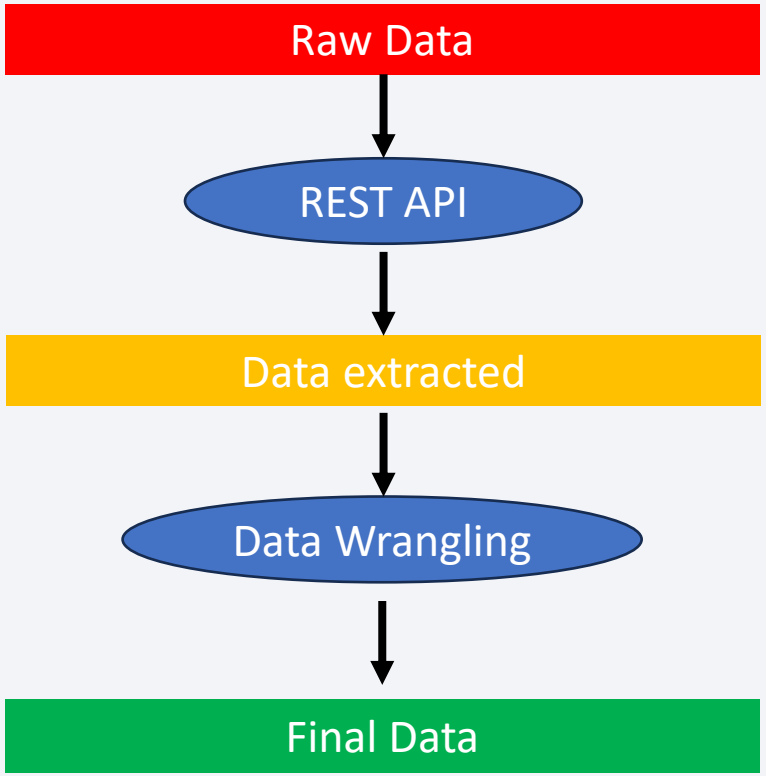
Data Collection – SpaceX API

In this part, the data will be downloaded and prepared for further analysis.

URL:

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-labs-spacex-data-collection-api.ipynb

Flowchart of SpaceX API calls here:



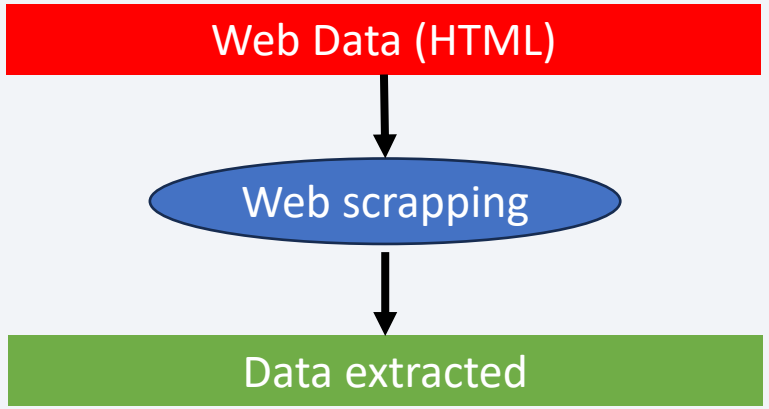
Data Collection - Scraping

In this part, the data will be downloaded directly from the website.

URL:

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-labs-webscraping.ipynb

Flowchart of web scraping:



Data Wrangling

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Flight Number vs Launch Site
- Flight Number vs Payload Mass
- Success for each orbit type
- Flight Number vs Orbit
- Payload Mass vs Orbit

URL:

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- select *
- select distinct
- select count
- where

URL:

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Added markers to localize each launch site
- Added markers for each flight in the different sites (through MarkerCluster).
- Added icons to indicate if a flight ended with success or failure.

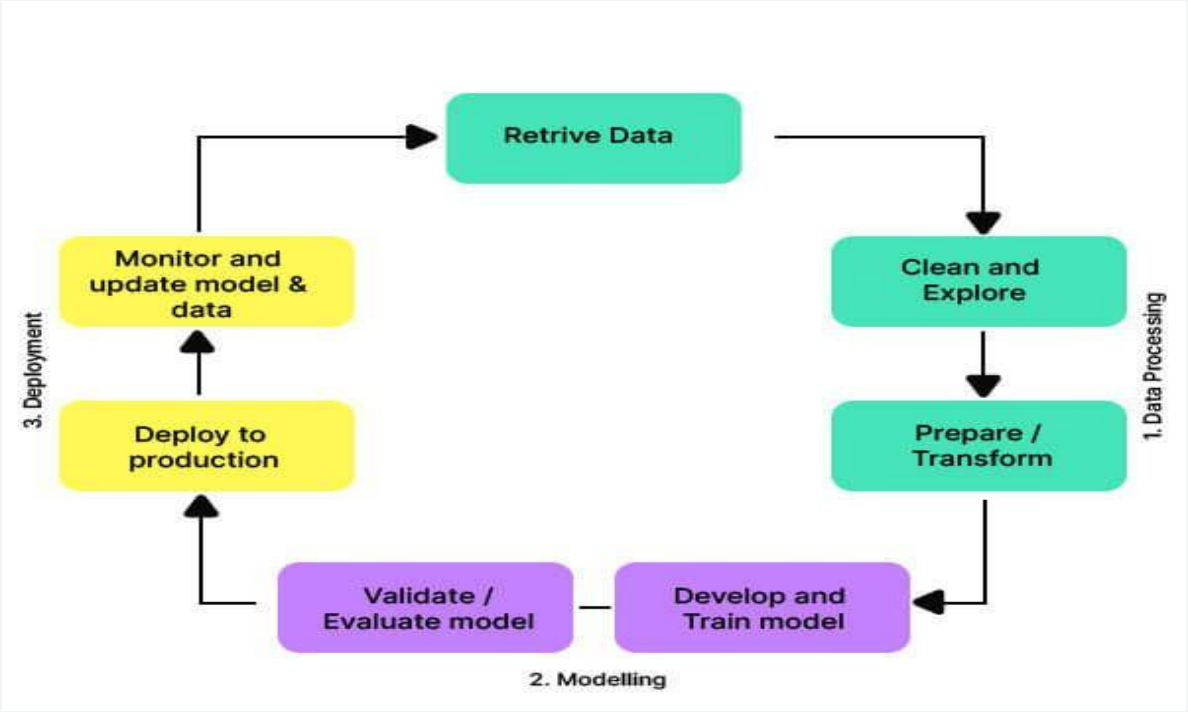
URL:

- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/jupyter-labs_launch_site_location.jupyterlite.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
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

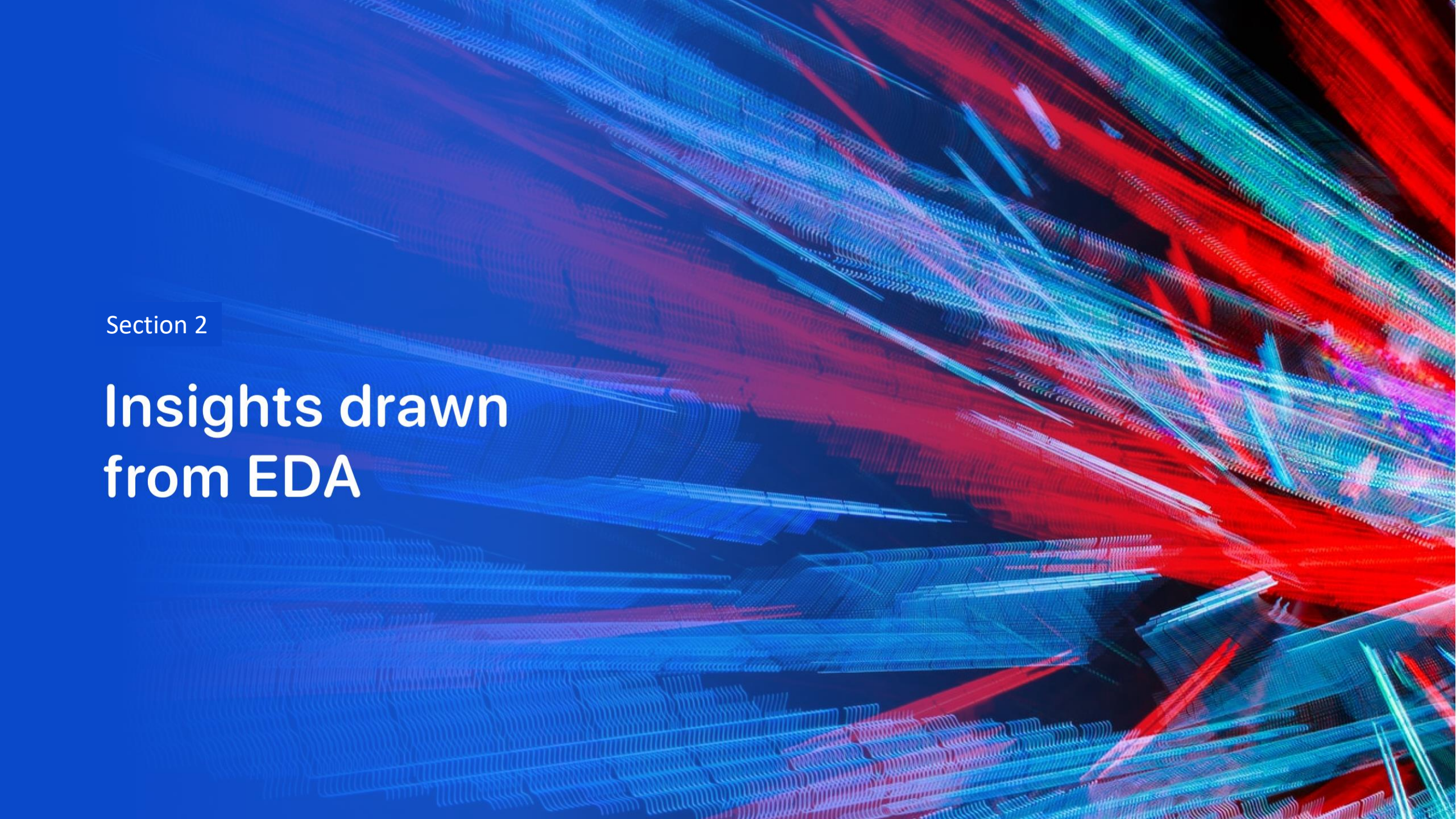


URL:

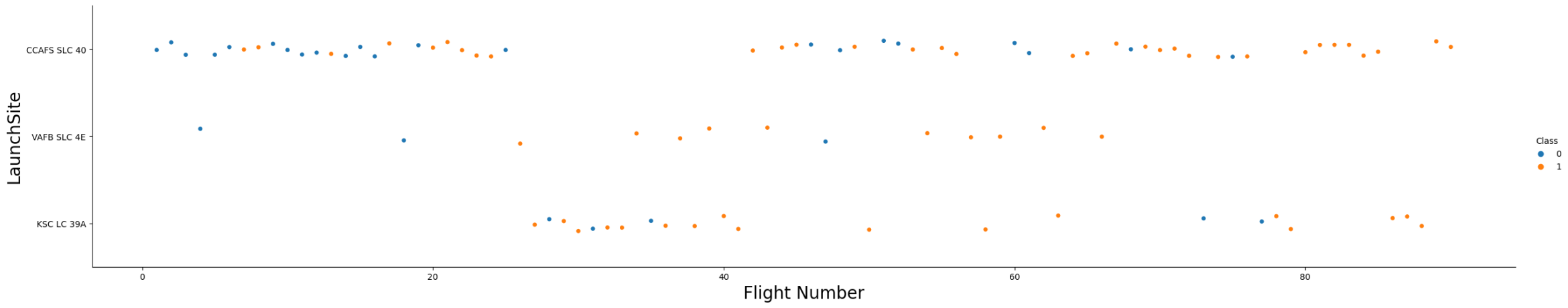
- https://github.com/andrus444/DataScienceCourse_Public/blob/main/DataScience_Capstone/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

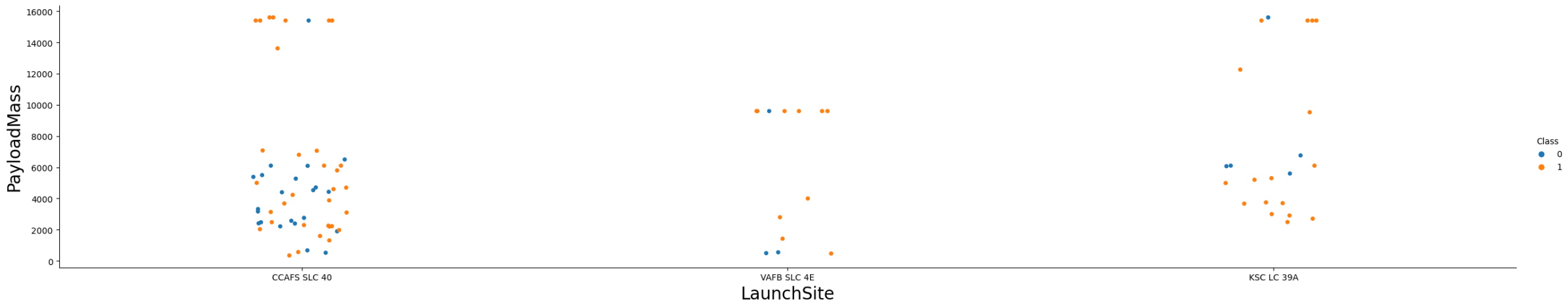
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

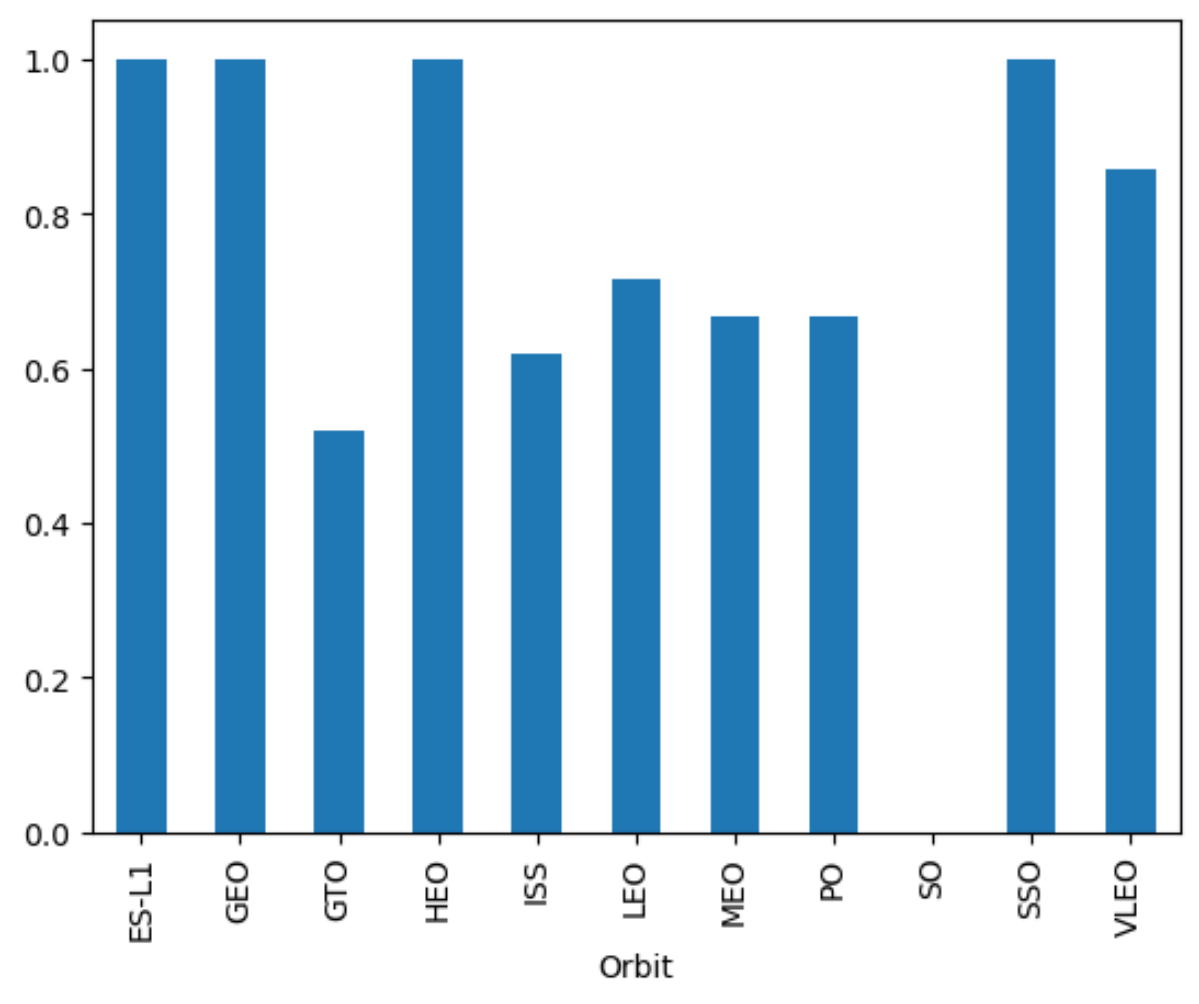
Flight Number vs. Launch Site



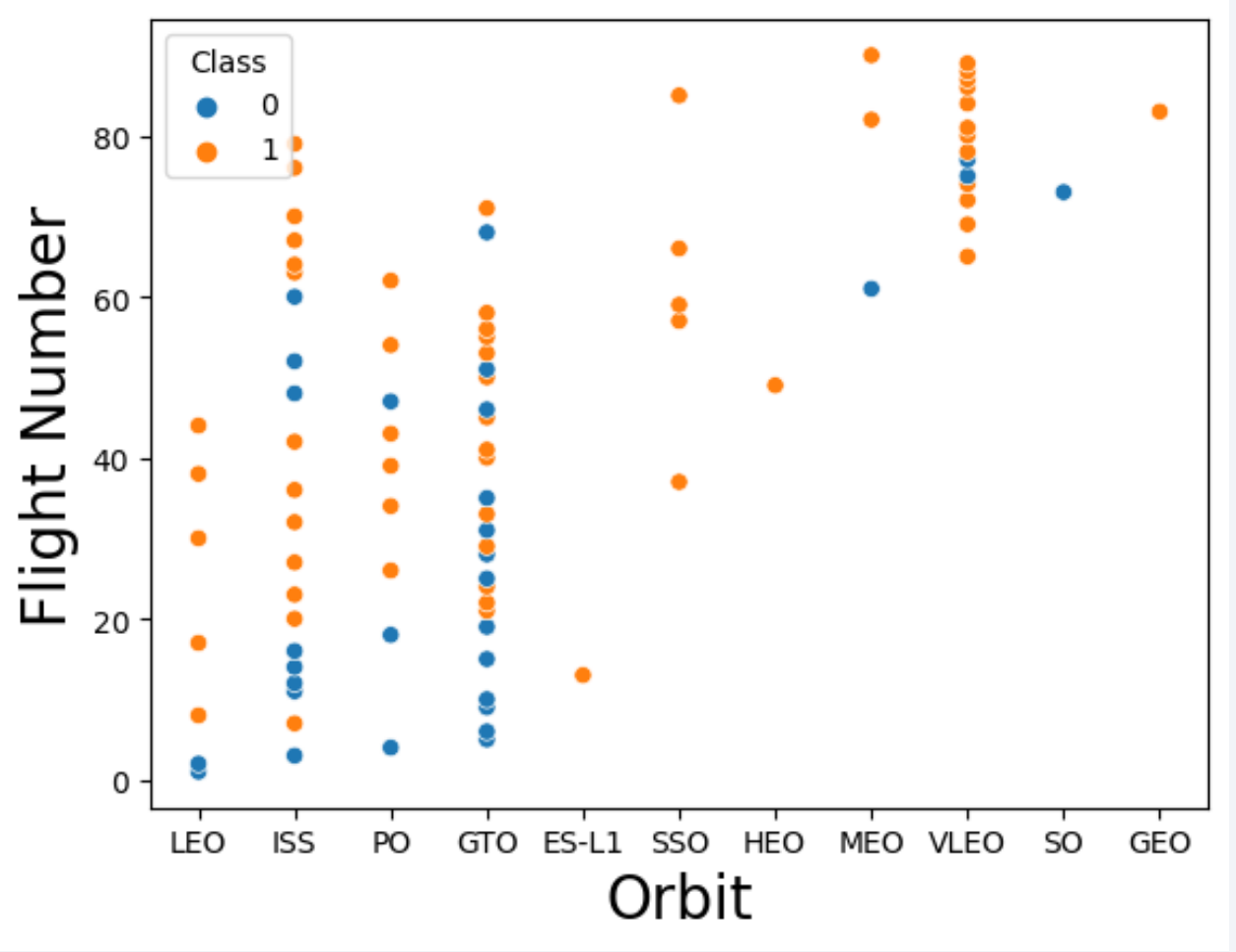
Payload vs. Launch Site



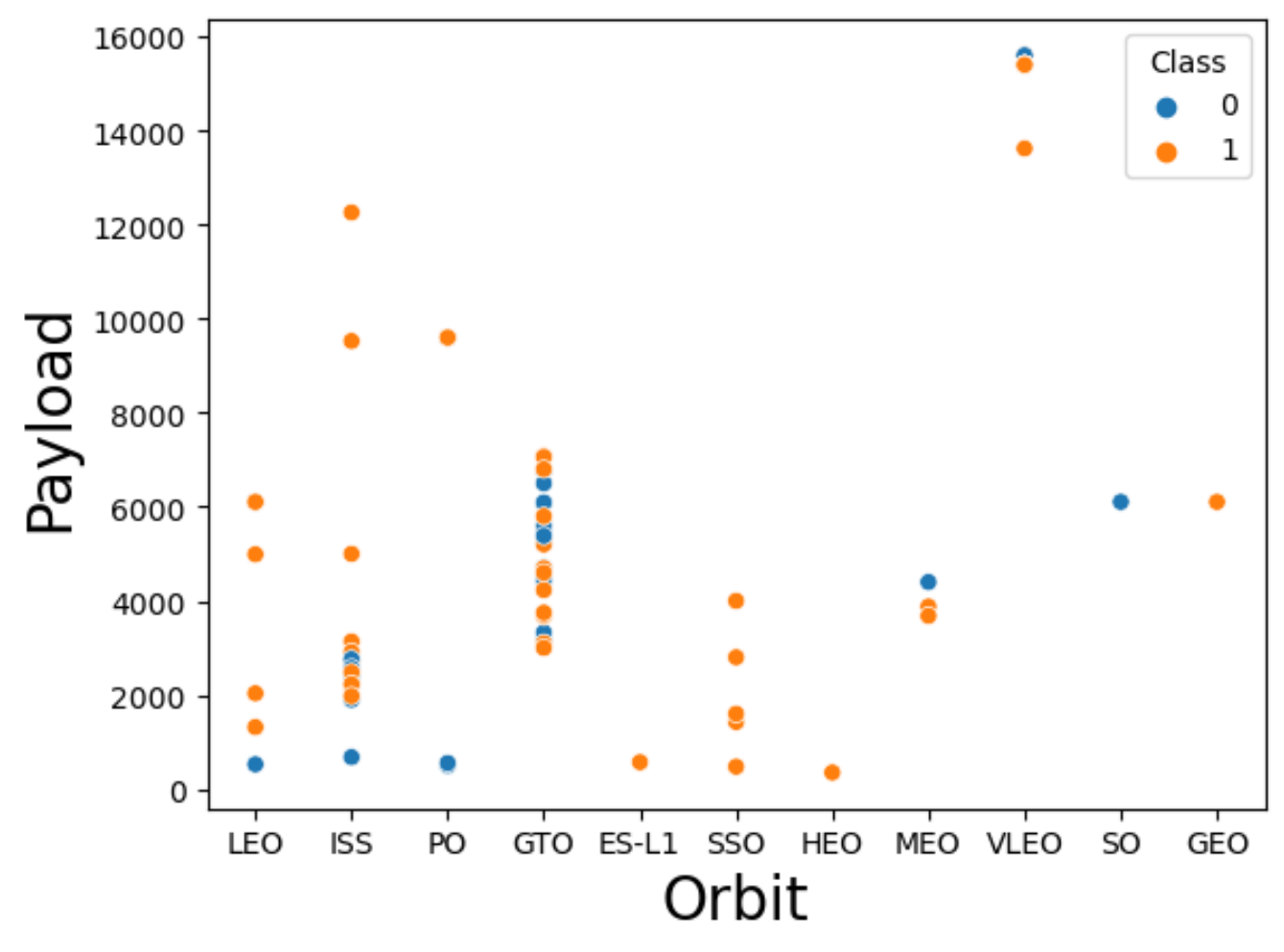
Success Rate vs. Orbit Type



Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend

All Launch Site Names

Task 1

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

In [15]:
```
%sql select distinct Launch_Site from SPACEXTABLE
```

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

Out[15]:

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

Launch Site Names Begin with 'CCA'

Task 2

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

In [39]:
```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

Out[39]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Ou
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-01-03	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)

In [53]:
```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer like 'NASA (CRS)'
```

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

Out[53]:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Task 4

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

In [54]:
```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.1'
```

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

Out[54]:

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

In [62]:
```
%sql PRAGMA table_info(SPACEXTABLE);
```

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

Out[62]:

cid	name	type	notnull	dflt_value	pk
0	Date	TEXT	0	None	0
1	Time (UTC)	TEXT	0	None	0
2	Booster_Version	TEXT	0	None	0
3	Launch_Site	TEXT	0	None	0
4	Payload	TEXT	0	None	0
5	PAYLOAD_MASS__KG_	INT	0	None	0
6	Orbit	TEXT	0	None	0
7	Customer	TEXT	0	None	0
8	Mission_Outcome	TEXT	0	None	0
9	Landing_Outcome	TEXT	0	None	0

In [71]:
```
%sql select min(Date) from SPACEXTABLE
```

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

Out[71]:

min(Date)
2010-04-06

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 [92]:
```
%sql select distinct Booster_Version from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)'
```

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

Out[92]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [85]:
```
success = %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome='Success'
print(success)
```

```
 * sqlite:///my_data1.db
Done.
+------------------------+
| count(Mission_Outcome) |
+------------------------+
|           98           |
+------------------------+
```

In [87]:
```
failure = %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome!='Success'
print(failure)
```

```
 * sqlite:///my_data1.db
Done.
+------------------------+
| count(Mission_Outcome) |
+------------------------+
|           3            |
+------------------------+
```

Boosters Carried Maximum Payload

Task 8

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

In [99]:
```
%sql select distinct Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_
```

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

Out[99]:

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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

In [112…
```
#%sql select substr(Date,1,4),substr(Date,6,2),Date from SPACEXTABLE
```

In [127…
```
%sql select substr(Date,6,2) as Month,Landing_Outcome,Booster_Version,Launch_site from SPACEXTABLE
```

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

Out[127…

Month	Landing_Outcome	Booster_Version	Launch_Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	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 [130…

```
%sql select Landing_Outcome, COUNT(*) AS Outcome_Count from SPACEXTABLE GROUP BY Landing_Outcome
```

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

Out[130…

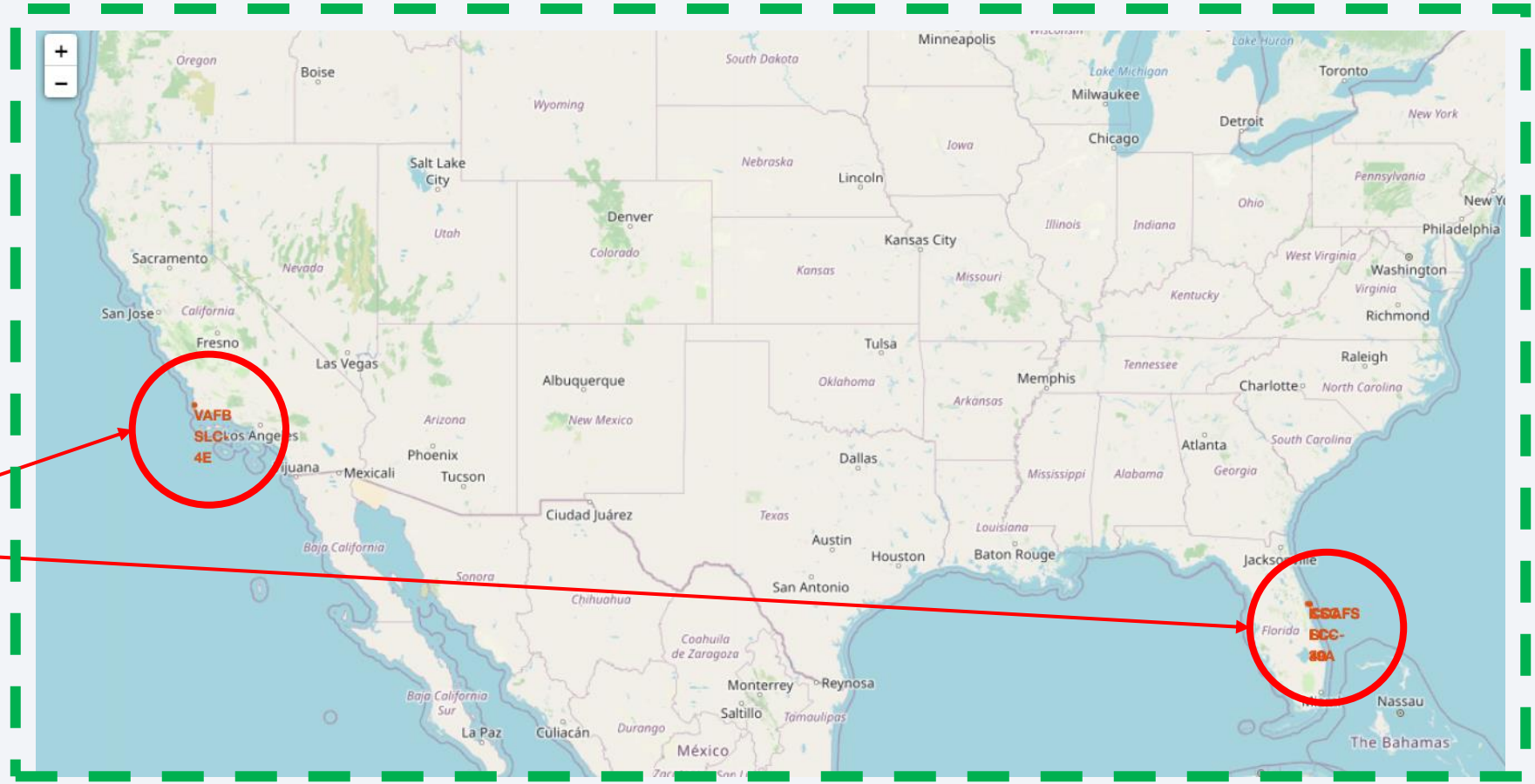
Landing_Outcome	Outcome_Count
Failure (parachute)	101

Section 3

Launch Sites Proximities Analysis

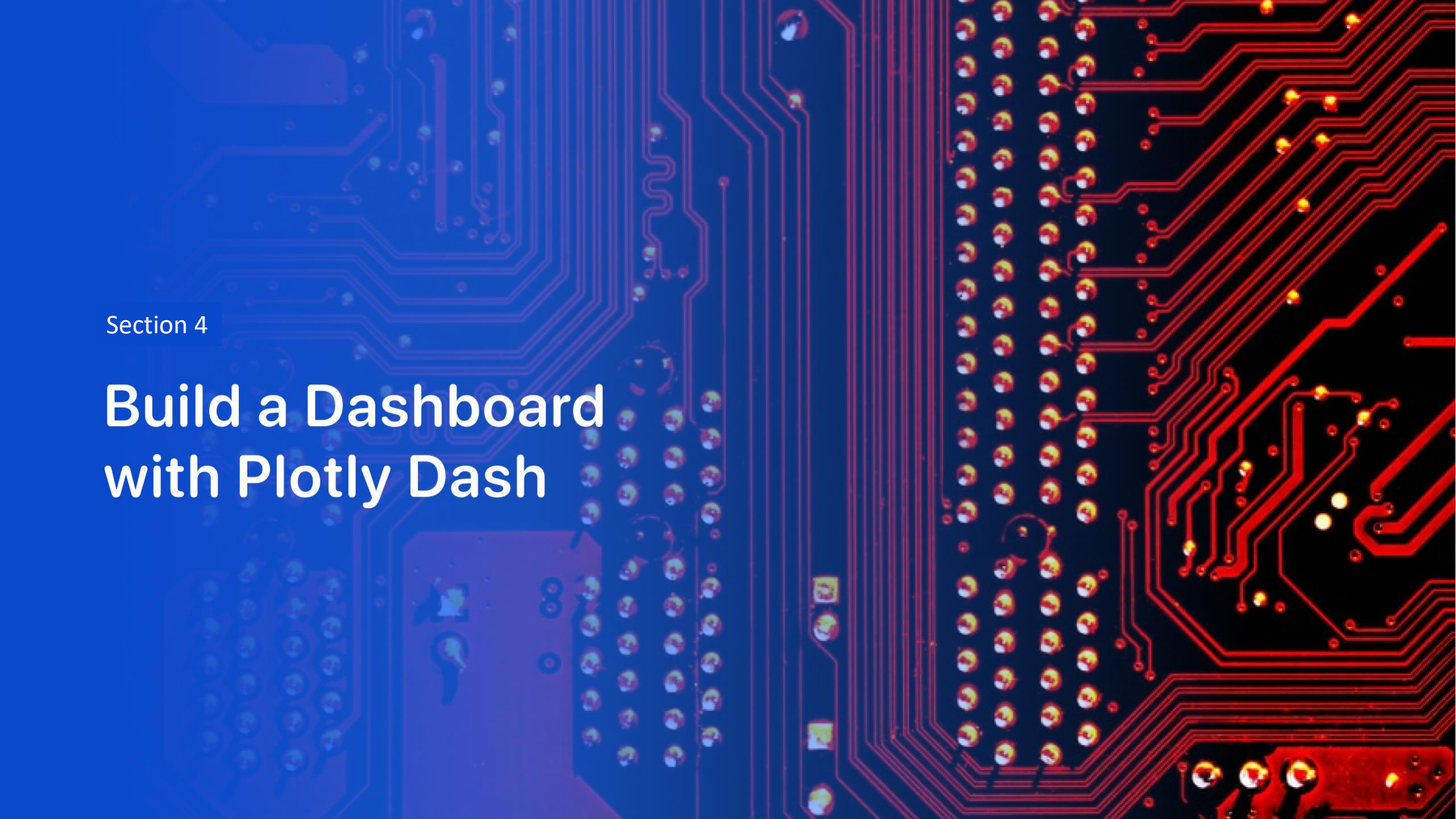
Folium Task 1!

Map



Folium Task 2!

Folium Task 3!


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

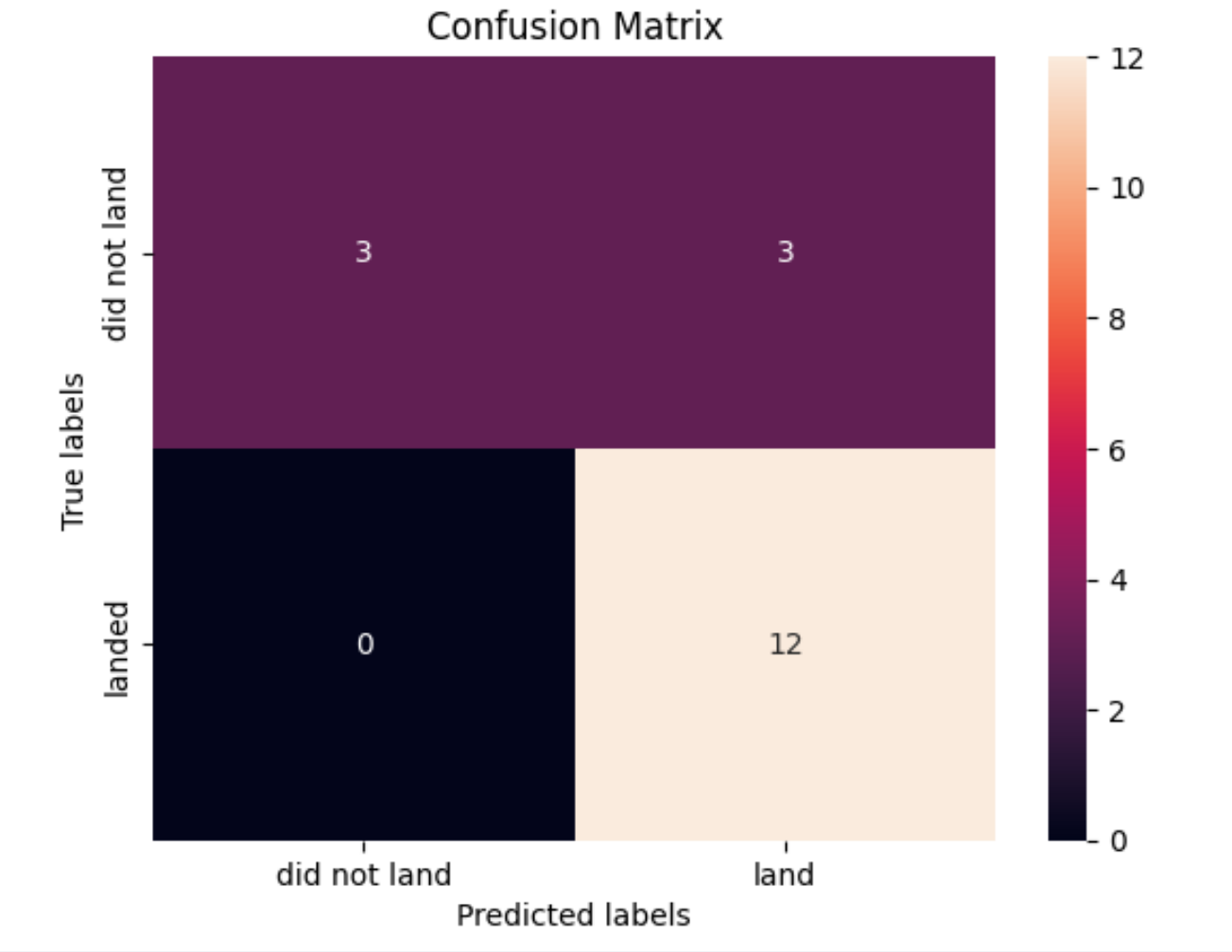
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

- Data Science has proved to be fundamental for analyzing raw data and extracted results.
- SQL queries became a powerful tool to interact with data bases.
- Flight Success can be predicted through several variables.
- The main results presented in graphs allow to determine the behavior of the system.

Appendix

Thank you!