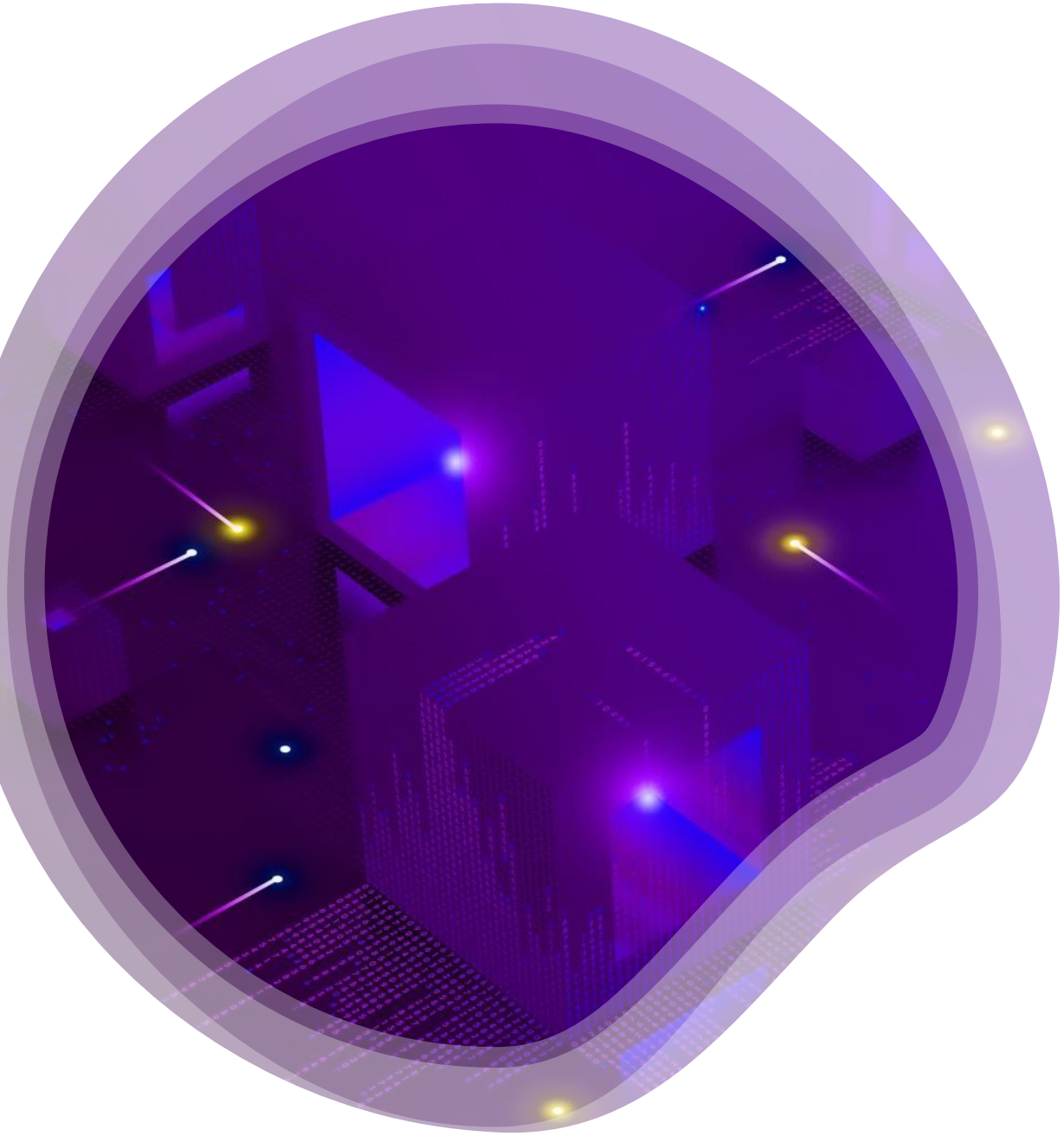
IBM Developer
SKILLS NETWORK

SpaceX

Winning Space Race with Data Science

ANKIT KUMAR

August – 08 - 2023



Outline

Executive Summary

- Summary of methodologies

- Data Collection using API

- Data Collection with web scraping

- Data Wrangling

- Exploratory Data Analysis with SQL

- Exploratory Data Analysis with Data Visualization

- Interactive Visual Analytics with Folium

- Machine Learning Prediction

- Summary of all results

The exploratory data analysis has shown us that successful landing outcomes are somewhat correlated with flight number. It was also apparent that successful landing outcomes have had a significant increase since the year 2015. All launch sites are located near the coast line. Perhaps, this makes it easier to test rocket landings in the water. sites are also located near highways and railways. This may facilitate transportation of equipment and research material. The machine learning were able to predict the landing success of rockets with an accuracy score of 83.33%.

Introduction

- Project background and context

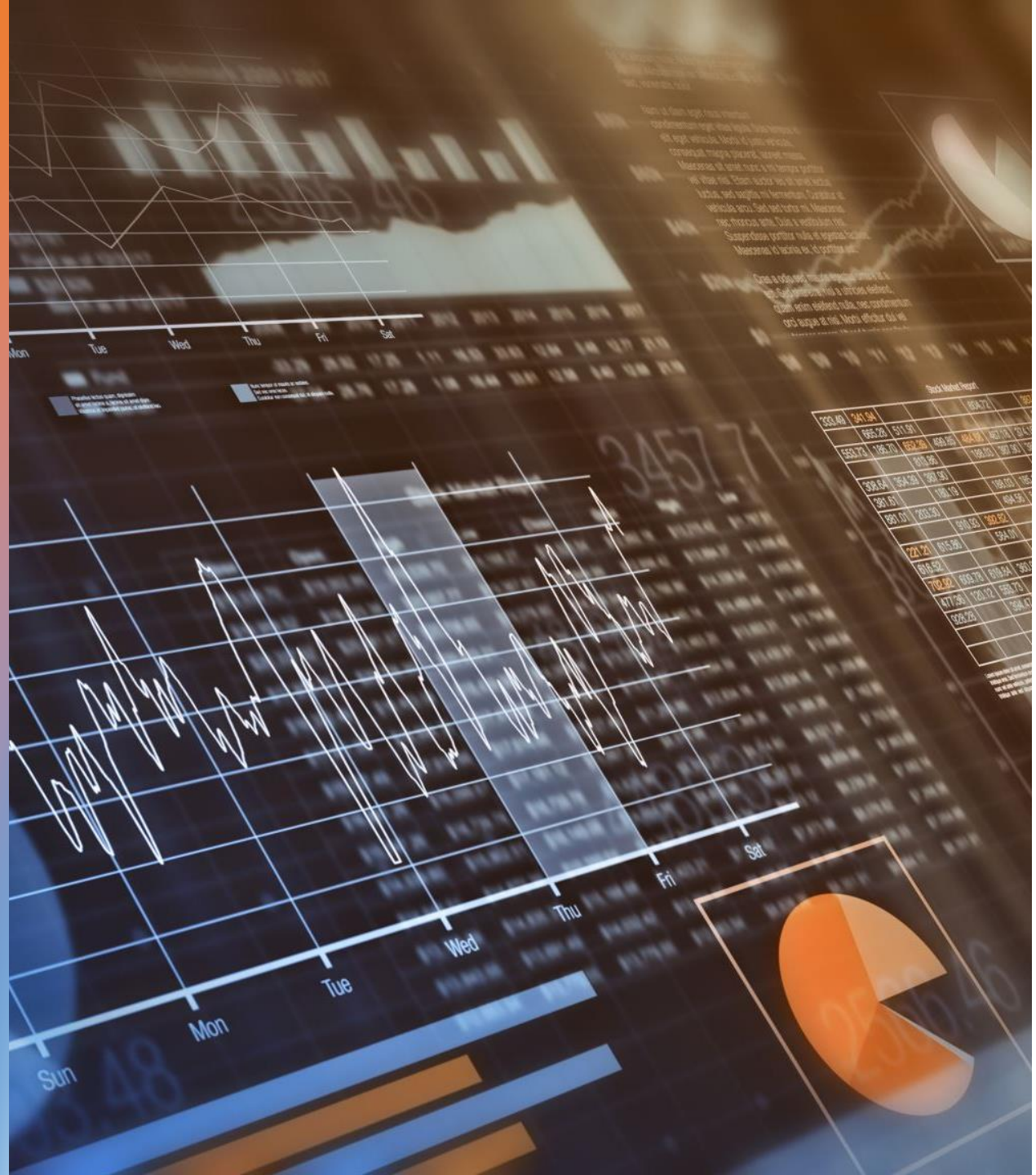
In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. 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.

- Problems you want to find answers

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 you can accurately predict the likelihood of the first stage rocket landing successfully, you can determine the cost of a launch. With the help of your Data Science findings and models, the competing startup you have been hired by can make more informed bids against SpaceX for a rocket launch.

Section 1

Methodology



Methodology

- Executive Summary
- Data collection methodology:
  - Get requests to the Space X API and web scraping from wikipedia
- Perform data wrangling
  - Clean the Data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Creating best Machine Learning model.

Data Collection

- The data sets were collected by
  - SpaceX API request.
  - Web Scraping

Enter URL of the page to be analyzed

Request and parse the SpaceX launch data using the GET request

Decode the response content as a JASON and turn it in a Pandas data frame

Now the API again to get information about the launches using IDs given for each launch

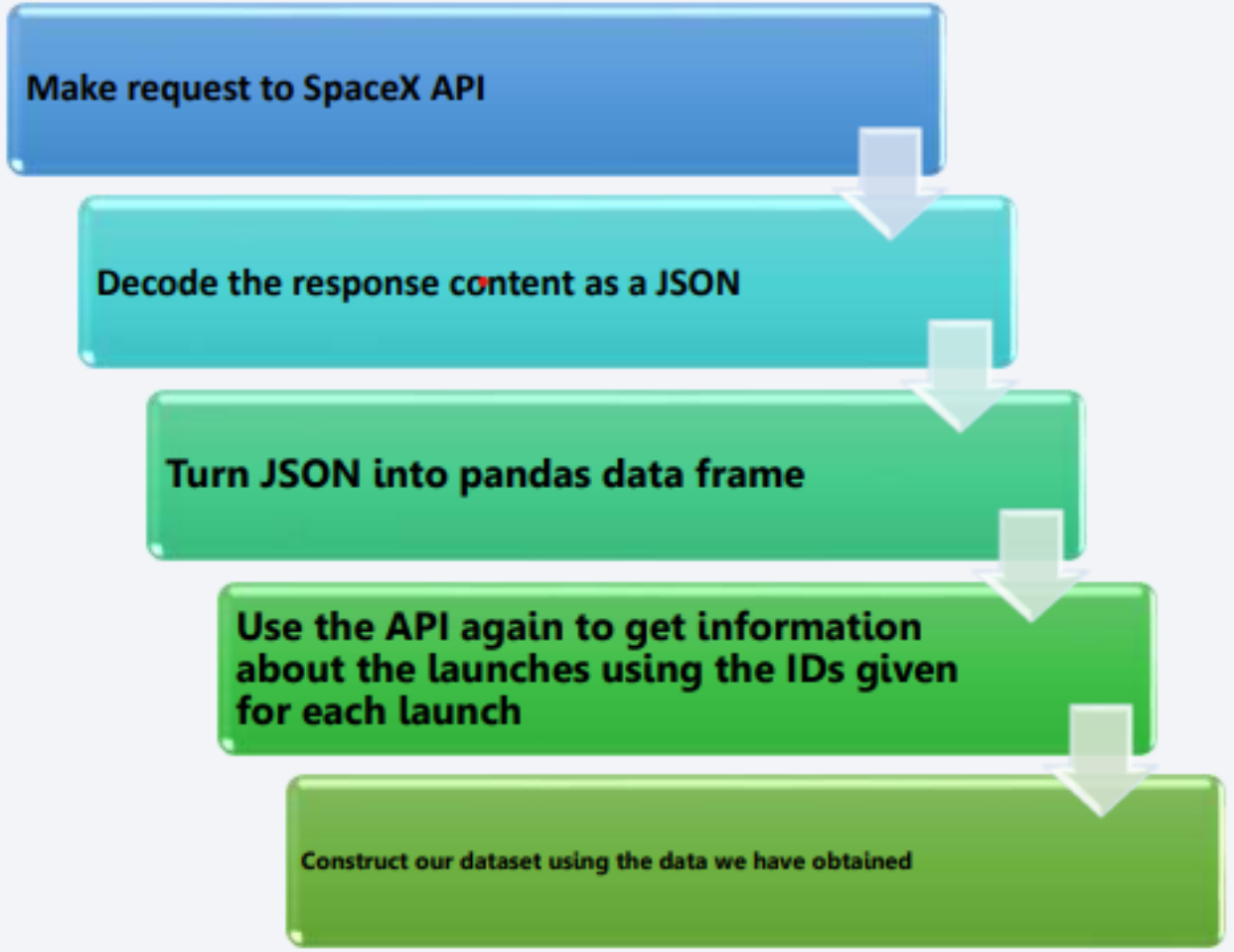
Filter the data frame to only include Falcon 9 launches and replace null values and get required output

Data Collection – SpaceX API

- How the data collection has been done using SpaceX API is presented in the form of a flow chart.
- For complete notebook link is given below.

- The GitHub URL link :

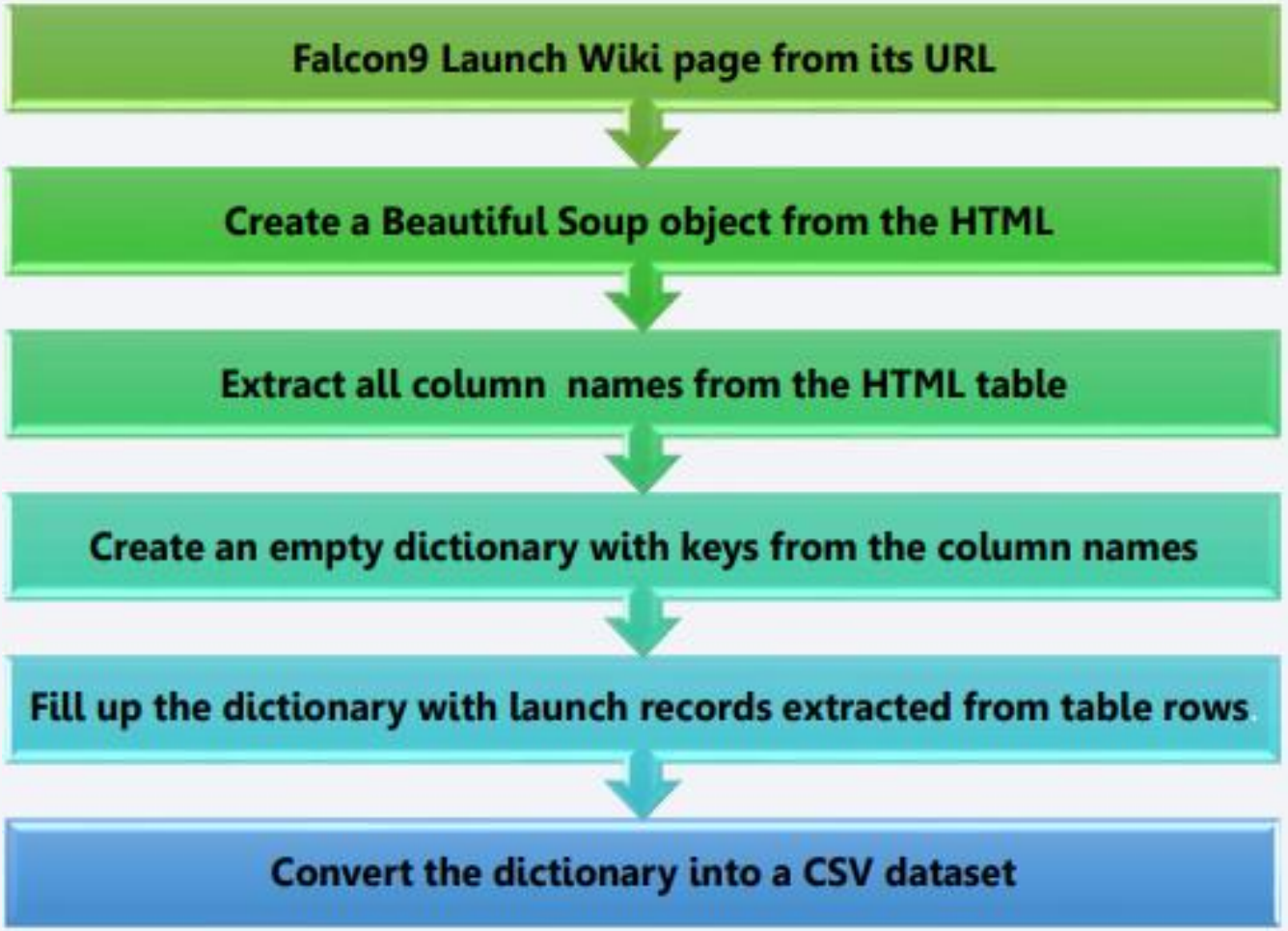
[IBM-Applied-Data-Science-Capstone-Project/Lab-1-jupyter-labs-spacex-data-collection-api.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)]



Data Collection - Scraping

- How the data collection has been done using WebScraping is presented in the form of a flow chart.
- For complete notebook link is given below.
- The GitHub URL link :

IBM-Applied-Data-Science-Capstone-Project/Lab-2-jupyter-labs-webscraping-data-collection.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)

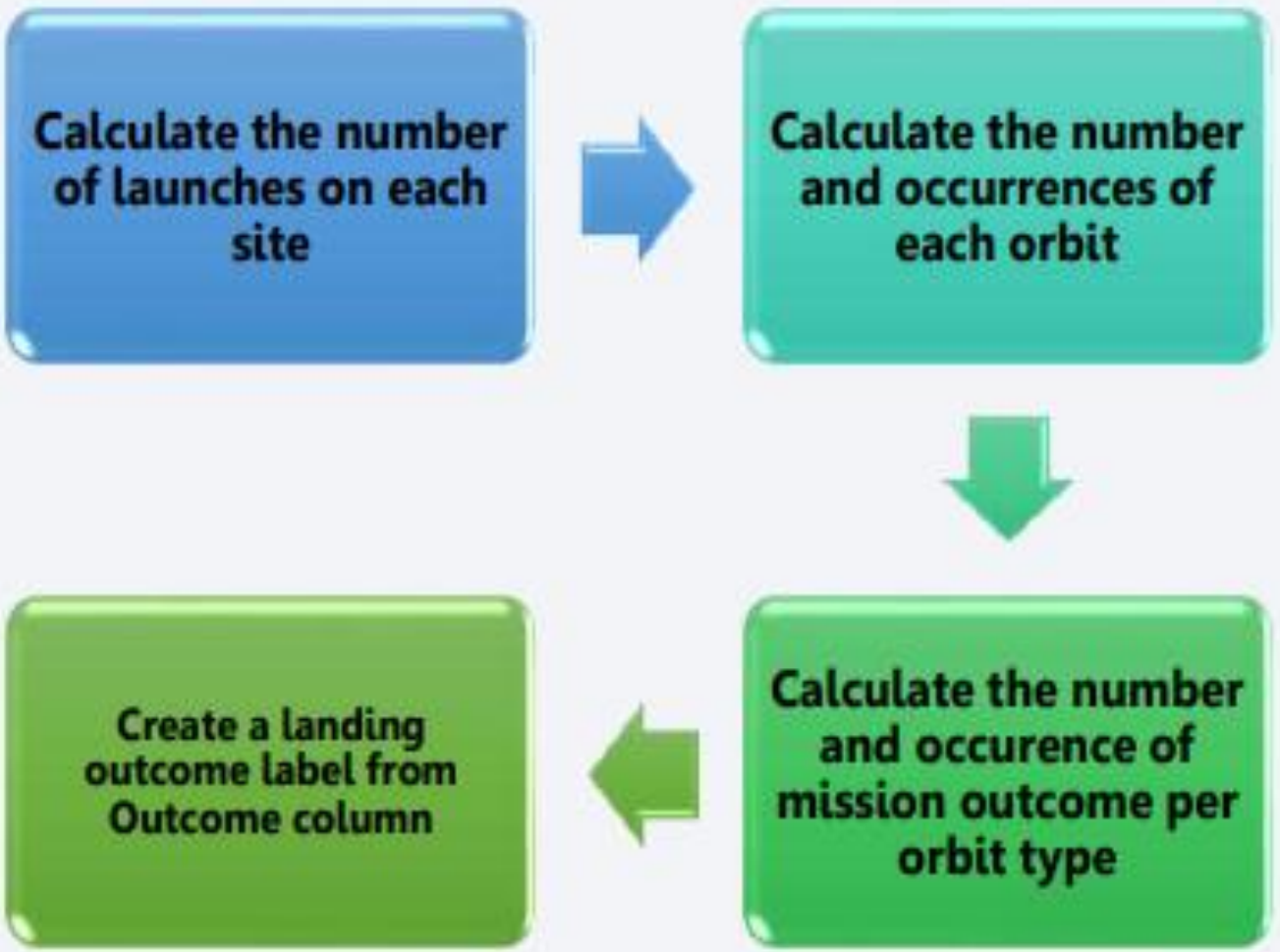


Data Wrangling

- How the data collection has been done using WebScraping is presented in the form of a flow chart.
- For complete notebook link is given below.

- The GitHub URL link :

IBM-Applied-Data-Science-Capstone-Project/Lab-3-jupyter-spacex-Data wrangling.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)



EDA with Data Visualization

- Types of Charts Used :
  - Scatter plot - Flight Number vs Payload Mass , Flight Number vs Launch Sites , Payload and Launch Sites , Flight Number and Orbit Type , Payload and Orbit Type
  - Bar chart – Success rate of each orbit
  - Line plot – success rate and Date
- For complete notebook link is given below.

- The GitHub URL link :
- IBM-Applied-Data-Science-Capstone-Project/Lab-5-jupyter-eda-dataviz.ipynb.jupyterlite.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)

EDA with SQL

- Summary of SQL queries that were used:

- 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 acheived
- 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.
  - 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

- For complete notebook the GitHub URL link :

- [IBM-Applied-Data-Science-Capstone-Project/Lab-4-jupyter-eda-sql-coursera_sqllite.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)]

Build an Interactive Map with Folium

- Folium Markers were used to show the Space X launch sites and their nearest important landmarks like railways, highways, cities and coastlines. Polylines were used to connect the launch sites to their nearest landmarks.

  Red represents rocket launch failures

  Green represents the successes.

- For complete notebook the GitHub URL link :
- [IBM-Applied-Data-Science-Capstone-Project/Lab-6_jupyter_launch_site_locations_analysis_with_folium.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)]

Build a Dashboard with Plotly Dash

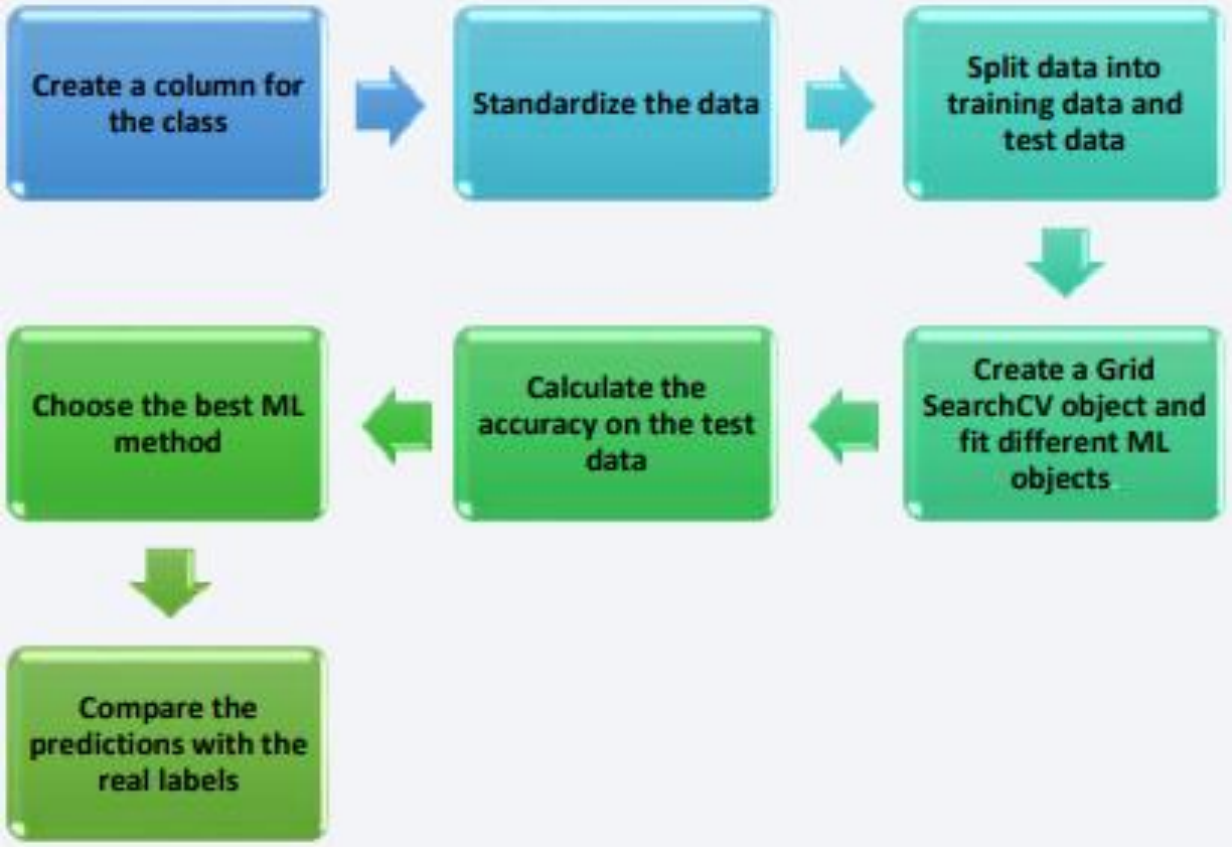
- Pie charts and scatter charts were used to visualize the launch records of Space X.
- These charts displayed the rocket launch success rate per launch site. We are were able to get an understanding of the factors that may have been influencing the success rate at each site. Such as the payload mass and booster versions.
- Successful launches were represented by 1 while failures were represented by 0.

Predictive Analysis (Classification)

Scikit-learn is Machine Learning library that was used for predictive analysis.

The following took place:

- Created a machine learning pipeline to predict if the first stage will land given the data

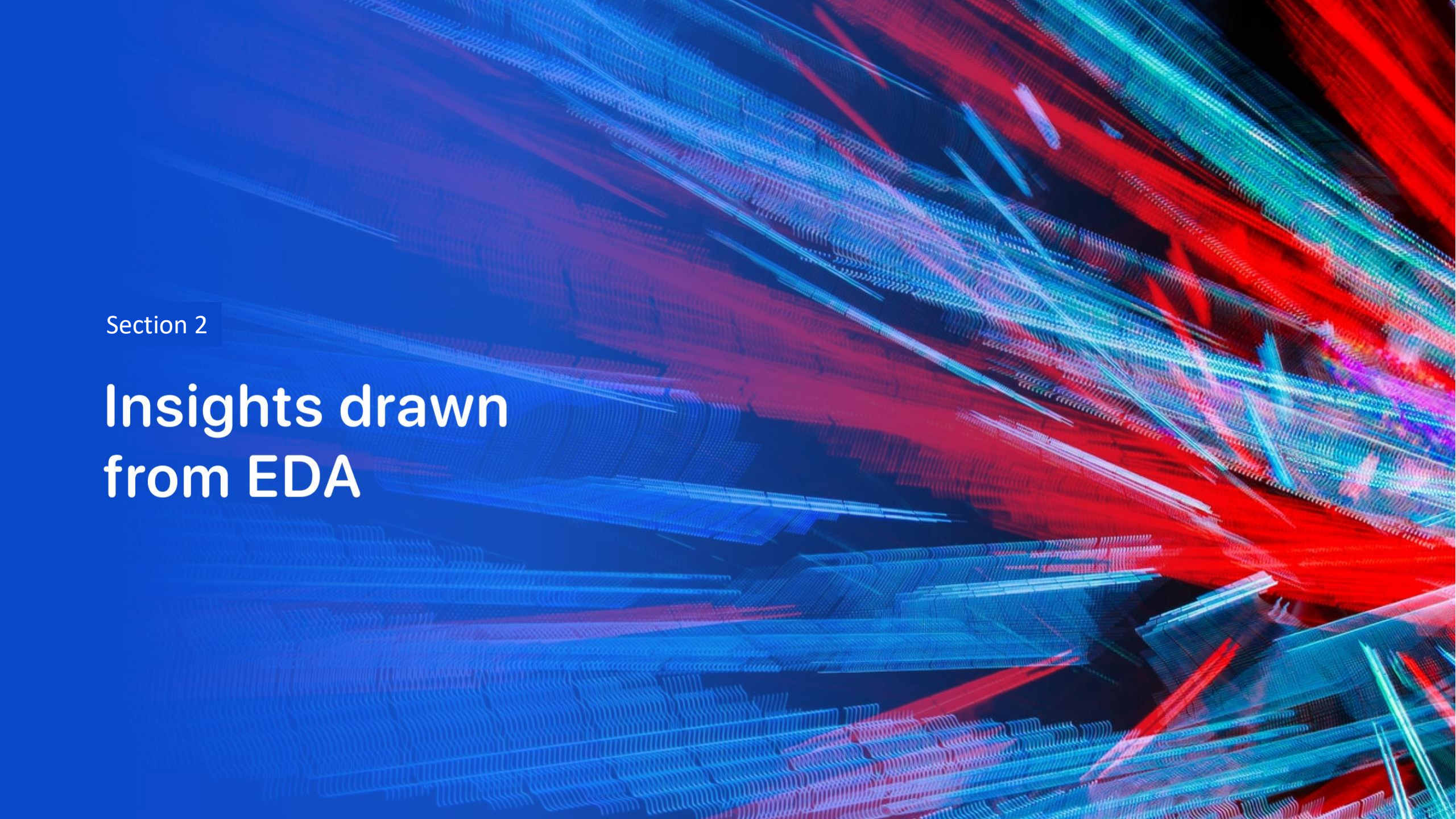


- The GitHub URL Link :

IBM-Applied-Data-Science-Capstone-Project/Lab-7_SpaceX_Machine_Learning_Prediction.jupyterlite.ipynb at main · Duhankit4012/IBM-Applied-Data-Science-Capstone-Project (github.com)

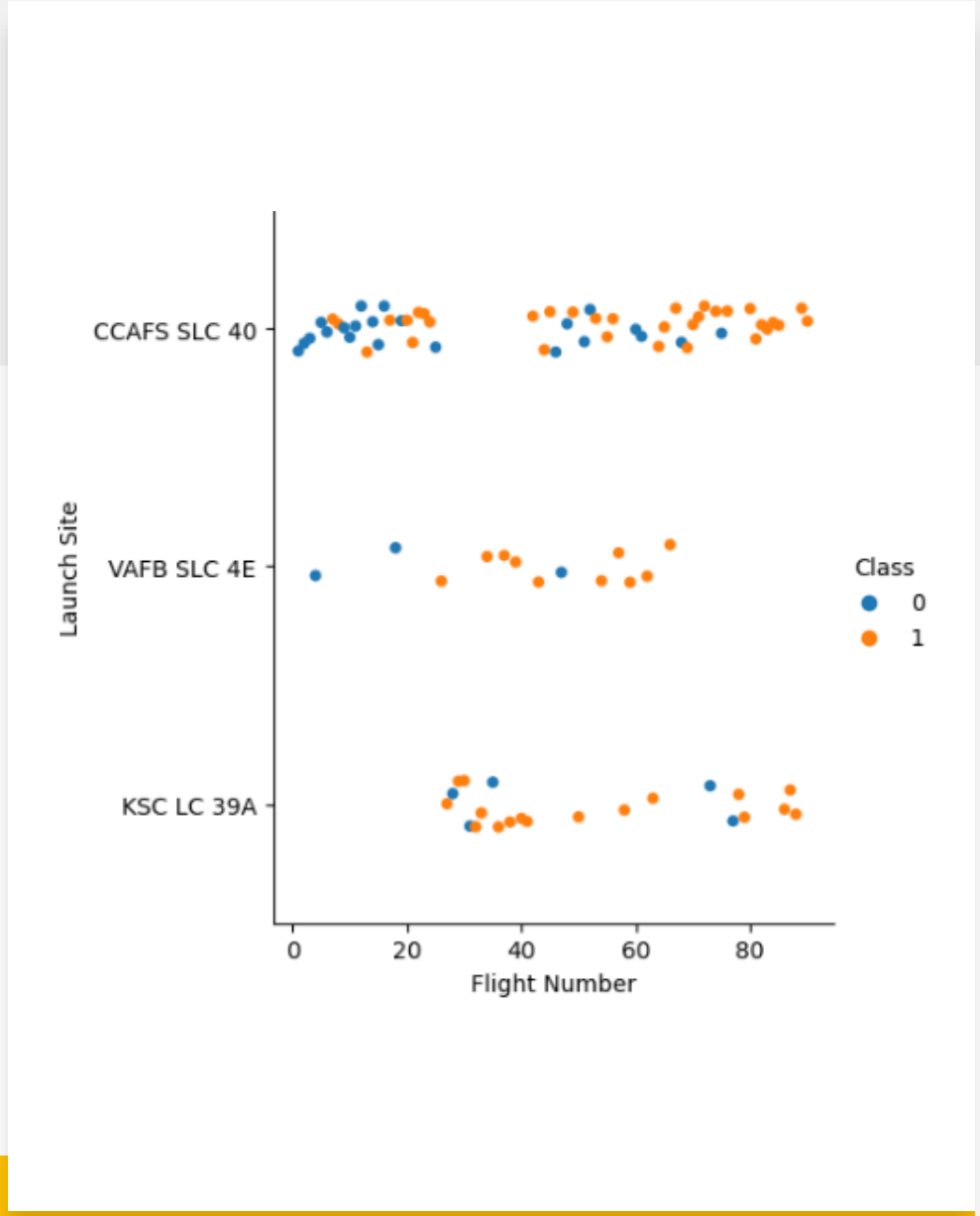
Results

- The exploratory data analysis has shown us that successful landing outcomes are somewhat correlated with flight number. It was also apparent that successful landing outcomes have had a significant increase since the year 2015.
- All launch sites are located near the coastline. Perhaps, this makes it easier to test rocket landings in the water. sites are also located near highways and railways. This may facilitate transportation of equipment and research material.
- The machine learning were able to predict the landing success of rockets with an accuracy score of 83.33%.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- From the chart beside it appears that there were more successful landings as the flight numbers increased. launch site **CCAFS SLC 40** had the most number of landing.

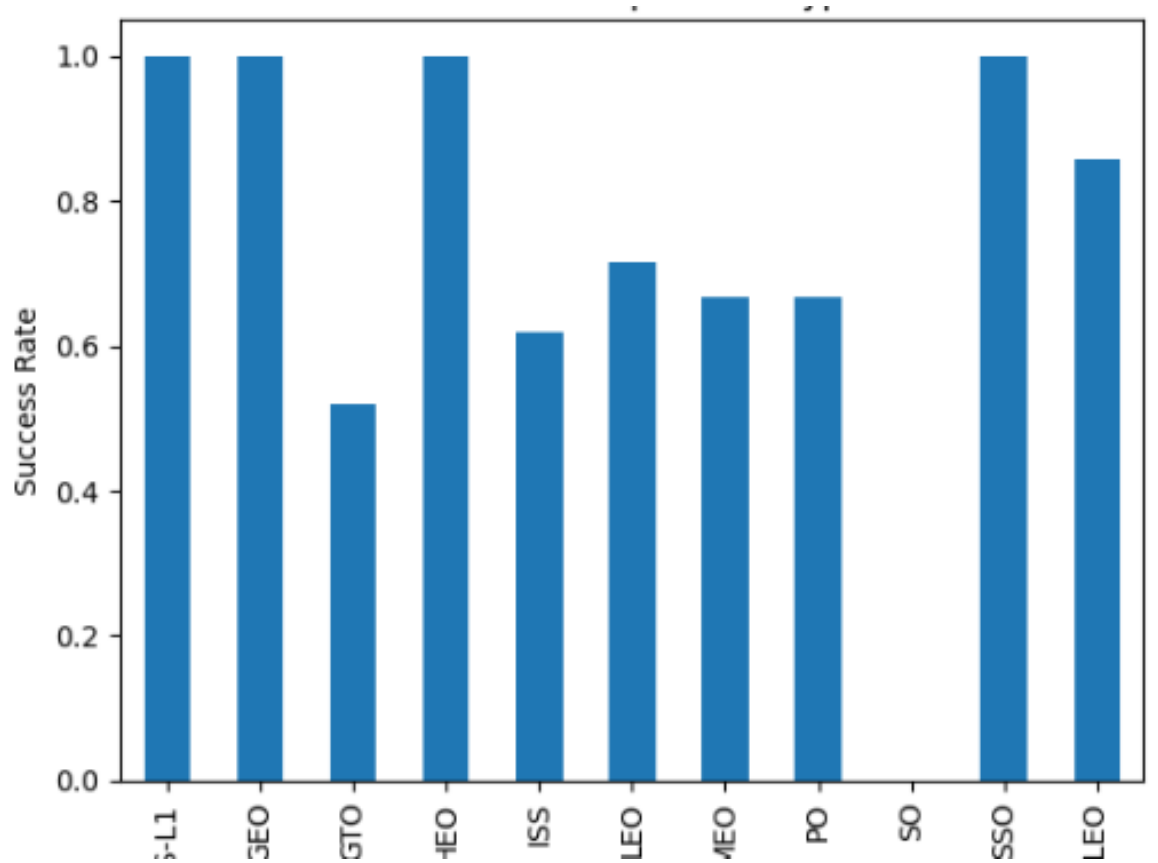


Payload vs. Launch Site

- Now we can observe the scatter point chart and find for the **VAFB-SLC** launch site there are no rockets launched for heavy payload mass(greater than 10000).

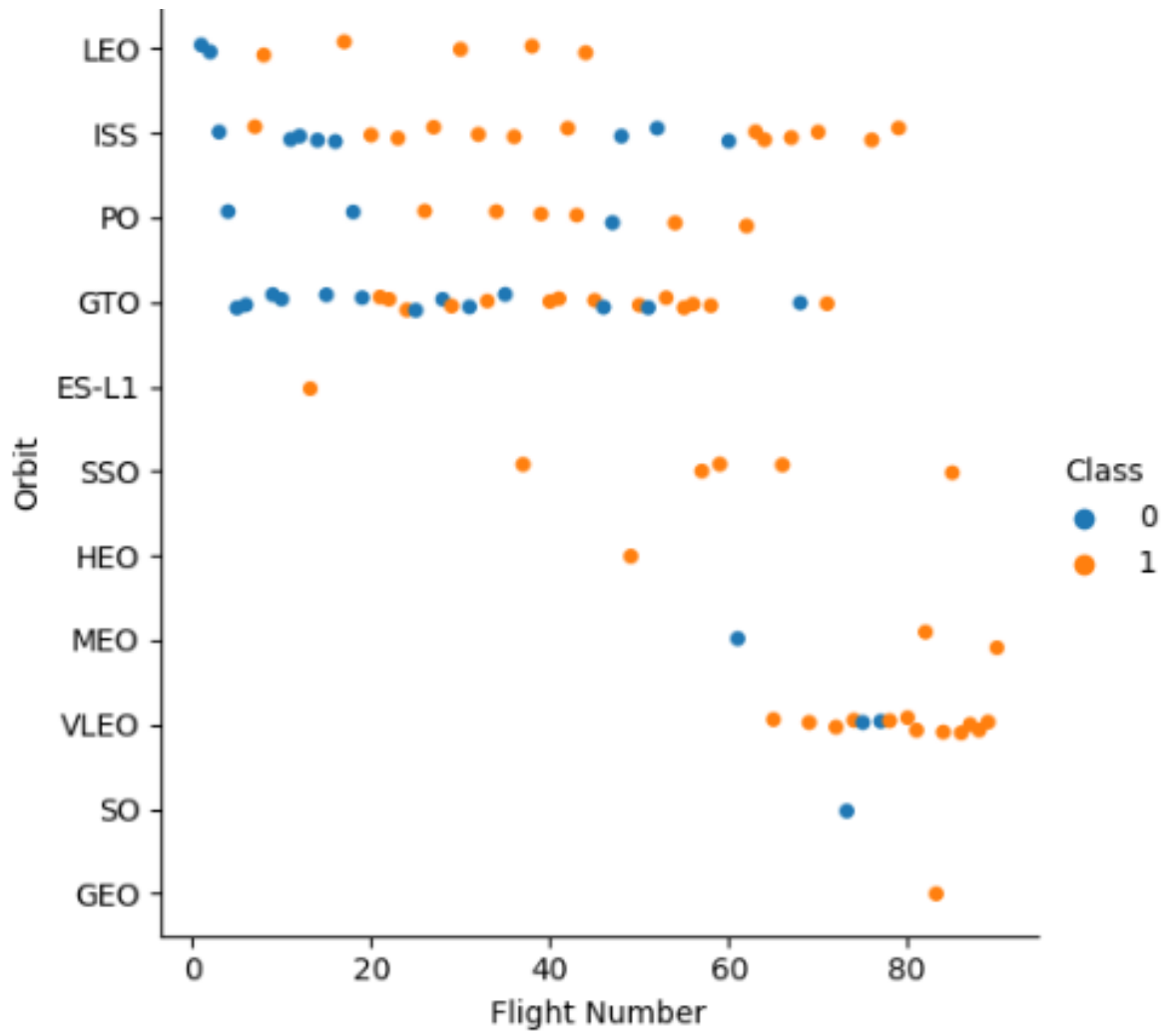
Success Rate vs. Orbit Type

- The highest success rate ORBITS are for
- - ES-L1
- - GEO
- - SSO
- - HEO



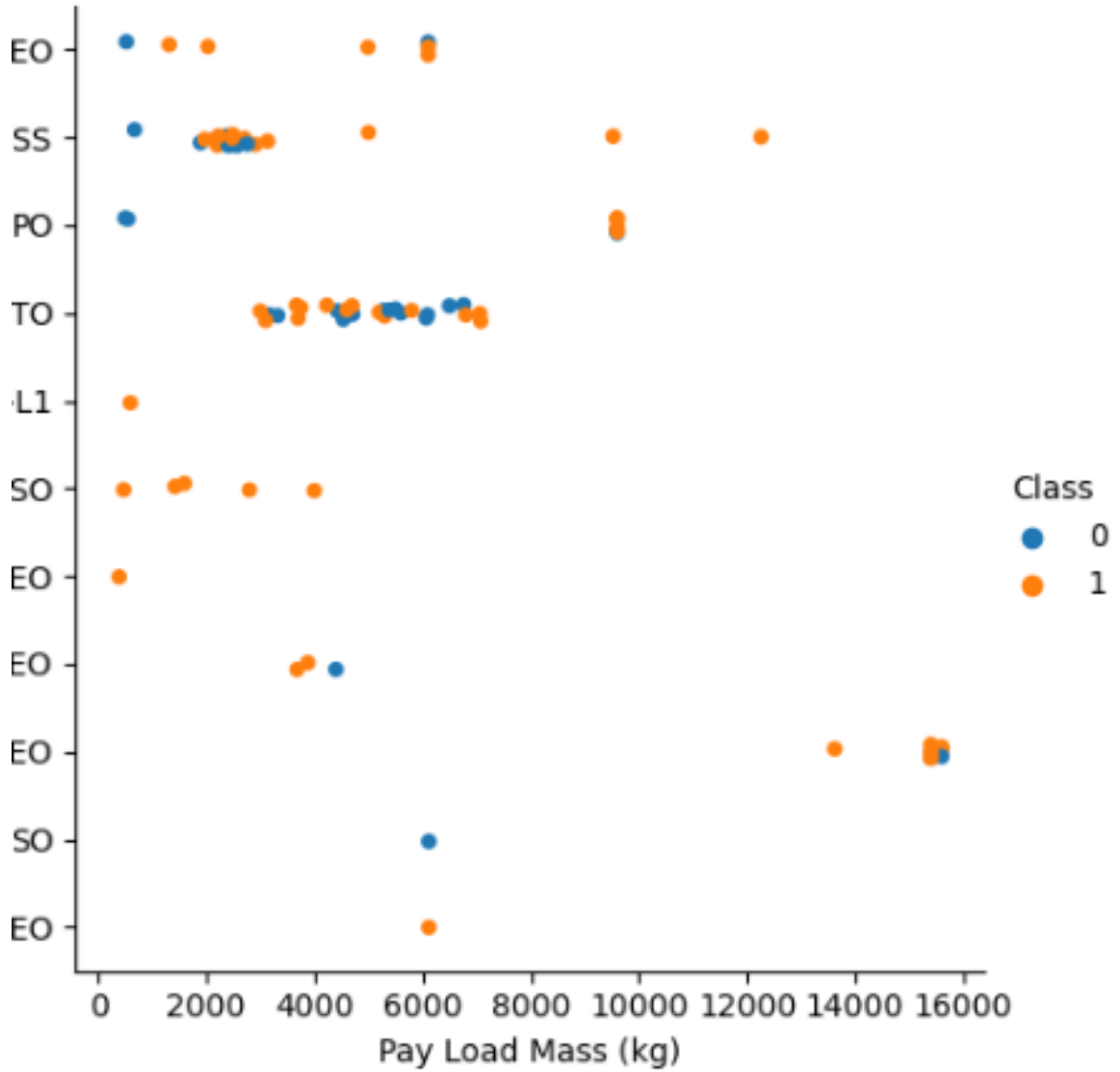
Flight Number vs. Orbit Type

- You can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



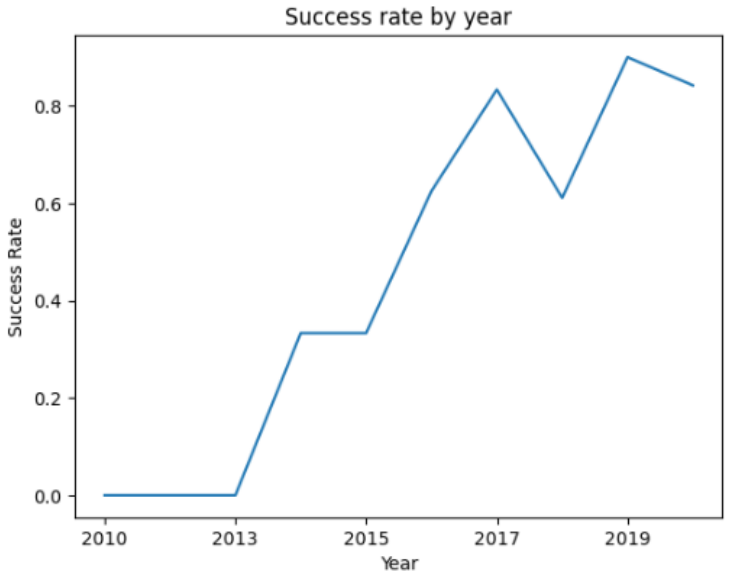
Payload vs. Orbit Type

- • With heavy payloads the successful landing or positive landing rate are more for **Polar, LEO and ISS**.
- • However for **GTO** we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there.



Launch Success Yearly Trend

- It is apparent that the success rate has significantly increased from 2013 to 2020.



All Launch Site Names

Given the data, these are the names of the launch sites where different rocket landings where attempted:

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

- These are 5 records where launch sites begin with the letters 'CCA'. As we can see, there are other organizations besides Space X that were testing their rockets.

Total Payload Mass

- The information in the picture displays the total payload mass carried by boosters launched by NASA

```
%%sql
SELECT sum(payload_mass__kg_) AS "Total payload mass (NASA (CRS))" FROM SPACEXTBL WHERE customer = 'NASA (CRS)';
```

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

Total payload mass (NASA (CRS))
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_) AS "Average payload mass (booster version F9 v1.1)" FROM SPACEXTBL WHERE booster_version LIKE 'F9 v1.1%';
```

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

Average payload mass (booster version F9 v1.1)
2534.6666666666665

- The average payload mass carried by F9 v1.1 was 2928.4 kg.

First Successful Ground Landing Date

- From the picture given above you can see that the first successful ground pad was in 22 December 2015.

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

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

min(DATE)
2015-12-22

```
%%sql
SELECT BOOSTER_VERSION from SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ <6000;
```

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

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

Successful Drone Ship Landing with Payload between 4000 and 6000

It appears that there only 4 Boosters with a payload mass between 4000 and 6000 they are

- • F9 FT B1022
- • F9 FT B1026
- • F9 FT B1021.2
- • F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- The Above picture show the total number of successful and failure mission outcomes

```
%%sql
SELECT 'Success' AS "Outcome", count(*) AS "Count" FROM SPACEXTBL WHERE landing_outcome LIKE 'Success%'
UNION ALL
SELECT 'Failure' AS "Outcome", count(*) AS "Count" FROM SPACEXTBL WHERE landing_outcome NOT LIKE 'Success%'
UNION ALL
SELECT '(All)' AS "Outcome", count(*) AS "Count" FROM SPACEXTBL;
```

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

Outcome	Count
Success	61
Failure	40
(All)	101

Boosters Carried Maximum Payload

From the picture it shows that 12 boosters have carried the maximum payload mass of 15600 kg

```
%%sql
SELECT DISTINCT booster_version
FROM SPACEXTBL
WHERE payload_mass__kg_ = (
    SELECT max(payload_mass__kg_)
    FROM SPACEXTBL
)
```

```
 * 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

- 2 boosters
- F9 v1.1B1012_CCAFS LC-40
- and
- F9v1.1B1015 CCAFS LC-40
- failed to land at 2015

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

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

The number of successful landings have increased since 2015.

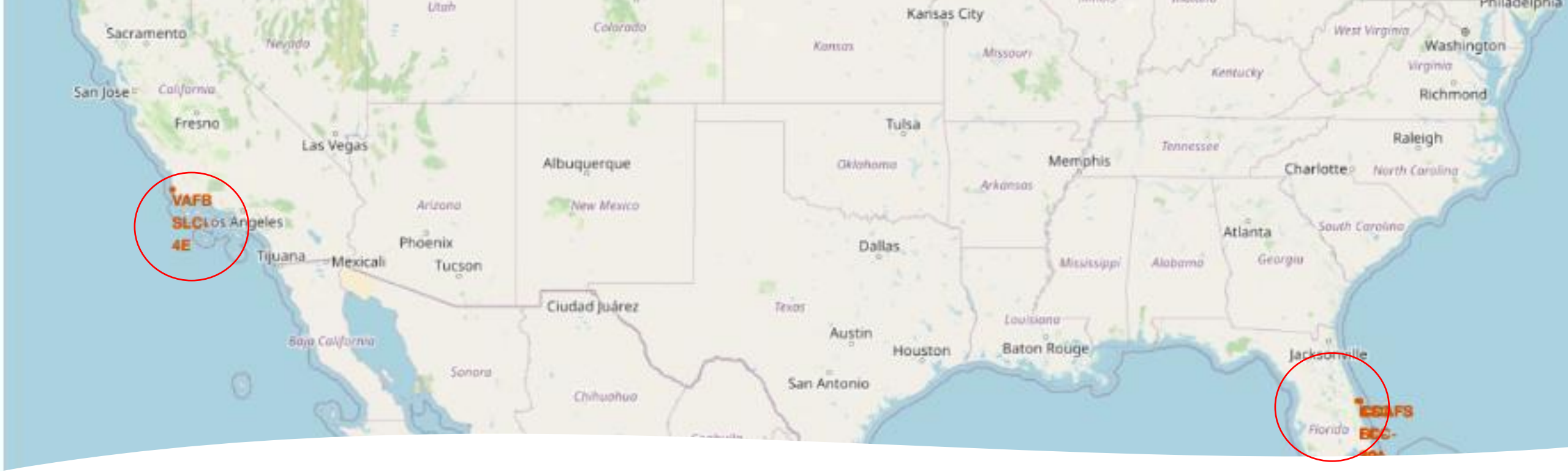
```
%%sql

SELECT landing_outcome, COUNT(*) AS "Count"
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY landing_outcome
ORDER BY Count DESC
;
```

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

Landing_Outcome	Count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

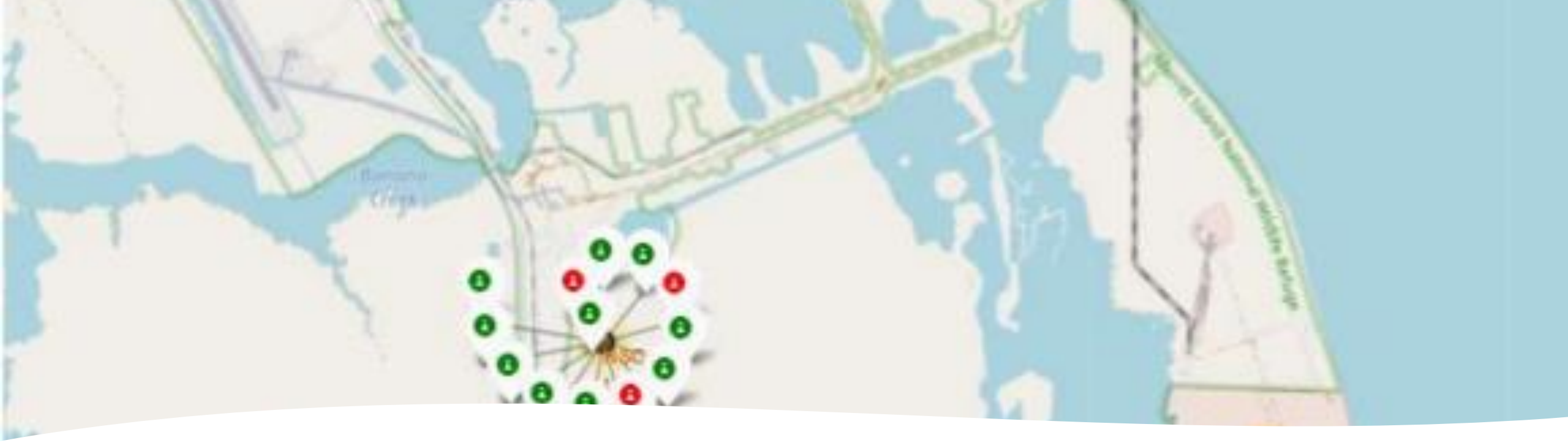
Section 3

Launch Sites Proximities Analysis



Launch Site Locations

- All launch sites are in very close proximity to the coast and they are also a couple thousand kilometers away from the equator line.

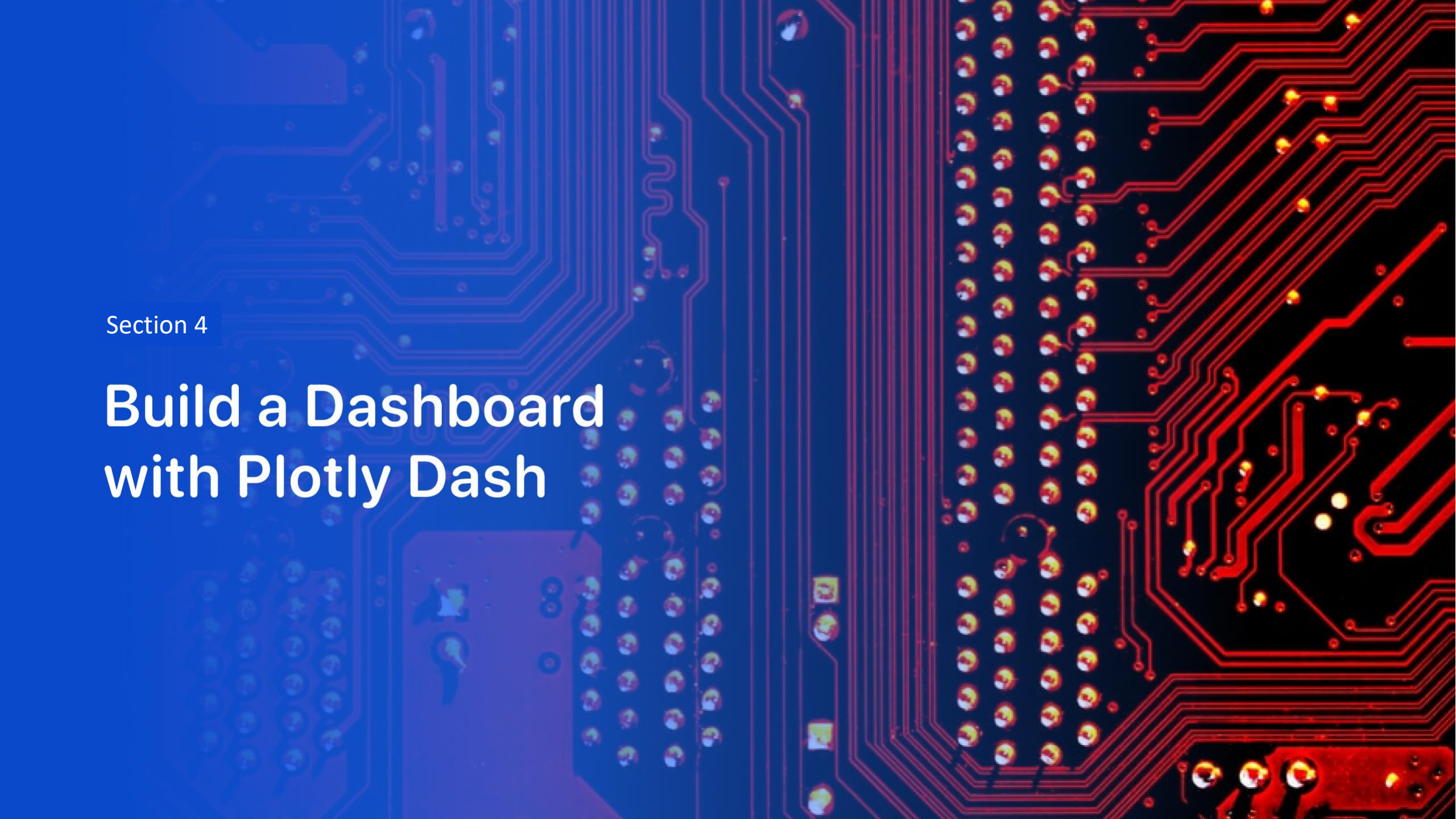


Success Rate of Rocket Launches

- The successful launches are represented by a green marker while the red marker represents failed rocket launches

Surrounding Landmarks

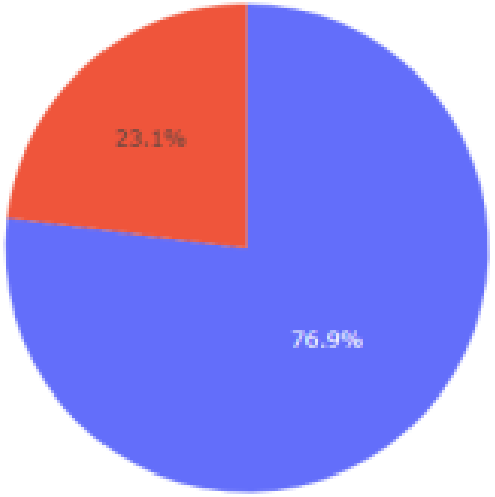
- It appears that launch sites are usually set up at least 18 km away from cities. This may be because of the desire to prevent any crashes near populated areas.
- It is also apparent that launch sites are in very close proximity to railways and highways. Perhaps, due to the necessary transportation requirements for rocket parts.
- The sites are close the coast line. This is evident with the many rocket landing tests on water bodies like the ocean.


Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site

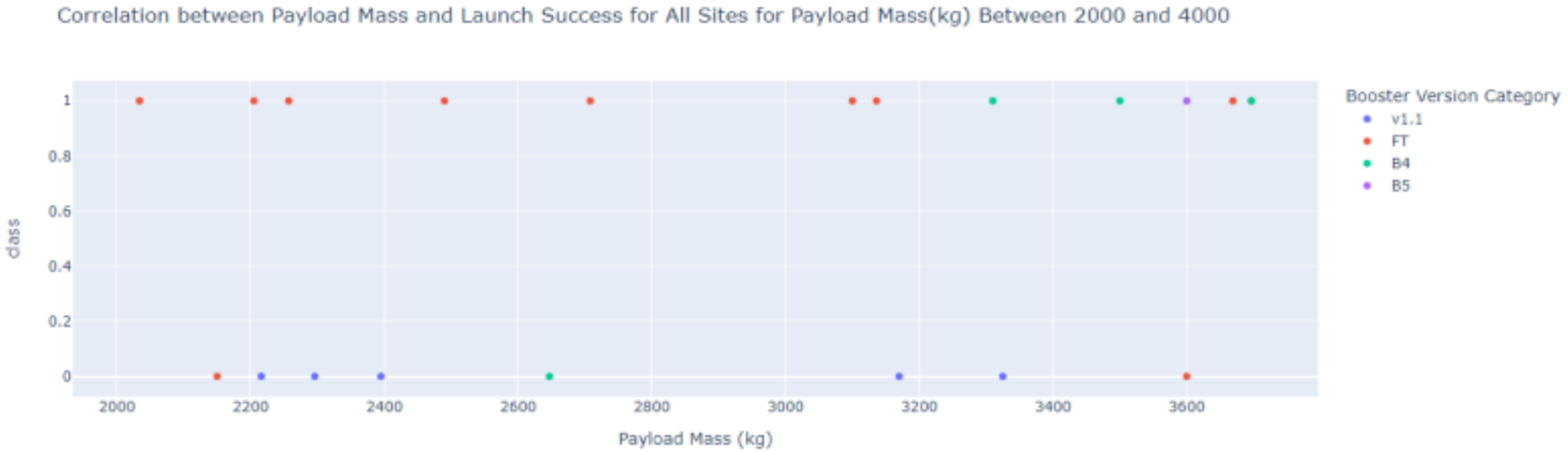
- You can see from the plot that Site KSC LC-39A has the largest successful launches as well the highest launch success rate.

Total Successful Launches for Site KSC LC-39A



- You can see that 76.9% of the total launches at site KSC LC-39A were successful. This is a the highest success rate of all the different launch sites.

Payload Mass vs. Launch Success for All Sites



It appears that the payload range between 2000 kg and 4000 kg has the highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- You can see that All the methods have an identical accuracy score of 83.33%, so we decided to use Logistic Regression for the classification

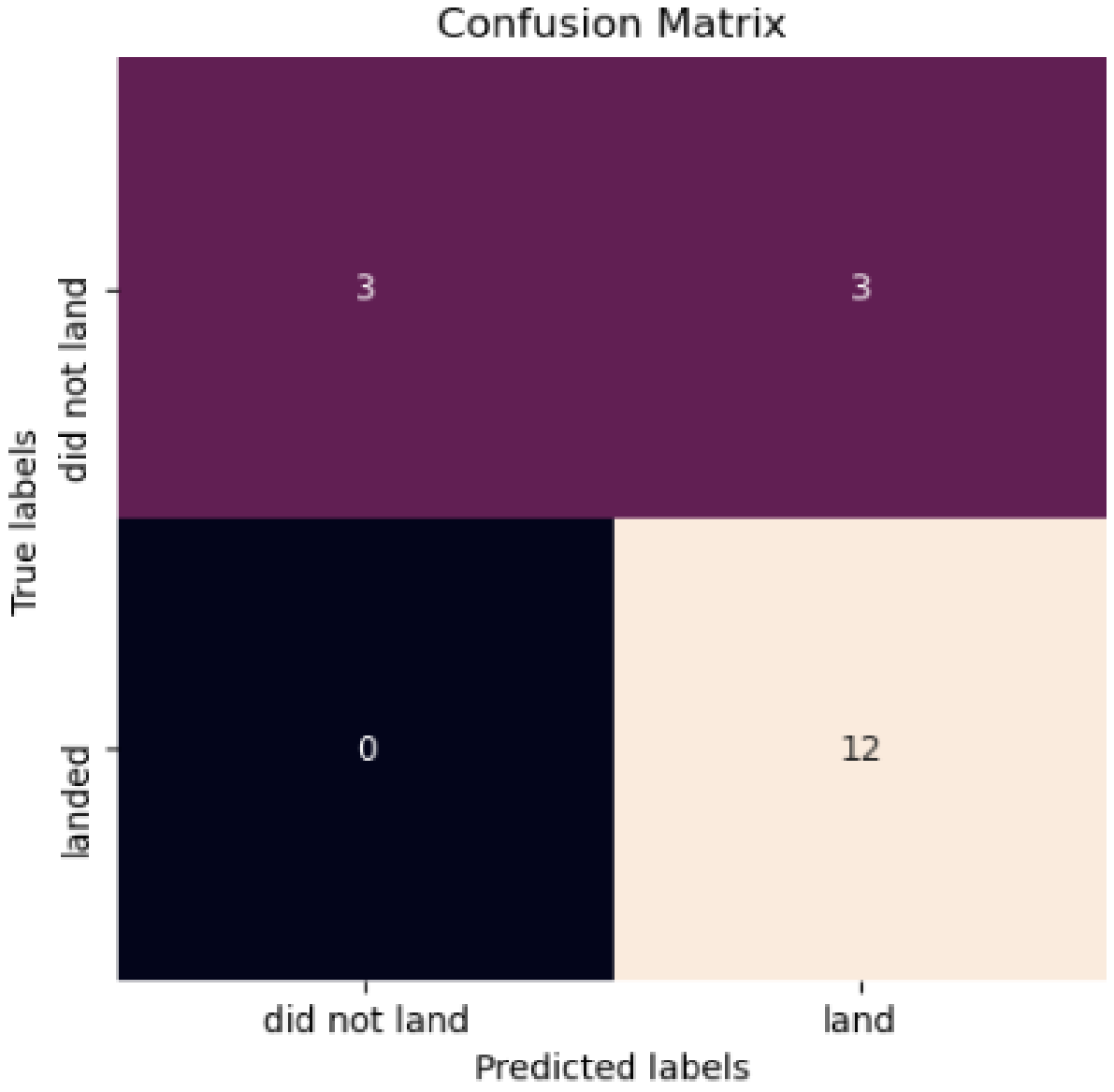
Find the method performs best:

```
accuracy = [svm_cv_score, logreg_score, knn_cv_score, tree_cv_score]
accuracy = [i * 100 for i in accuracy]

method = ['Support Vector Machine', 'Logistic Regression', 'K Nearest Neighbour', 'Decision Tree']
models = {'ML Method':method, 'Accuracy Score (%)':accuracy}

ML_df = pd.DataFrame(models)
ML_df
```

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix

- The chart shows the confusion matrix of the Logistic Regression model that was chosen.
- The model only failed to accurately predict 3 labels.



Conclusions

- The exploratory data analysis has shown us that successful landing outcomes are somewhat correlated with flight number. It was also apparent that successful landing outcomes have had a significant increase since the year 2015.
- All launch sites are located near the coast line. Perhaps, this makes it easier to test rocket landings in the water. sites are also located near highways and railways. This may facilitate transportation of equipment and research material.
- The machine learning were able to predict the landing success of rockets with an accuracy score of 83.33%.

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!