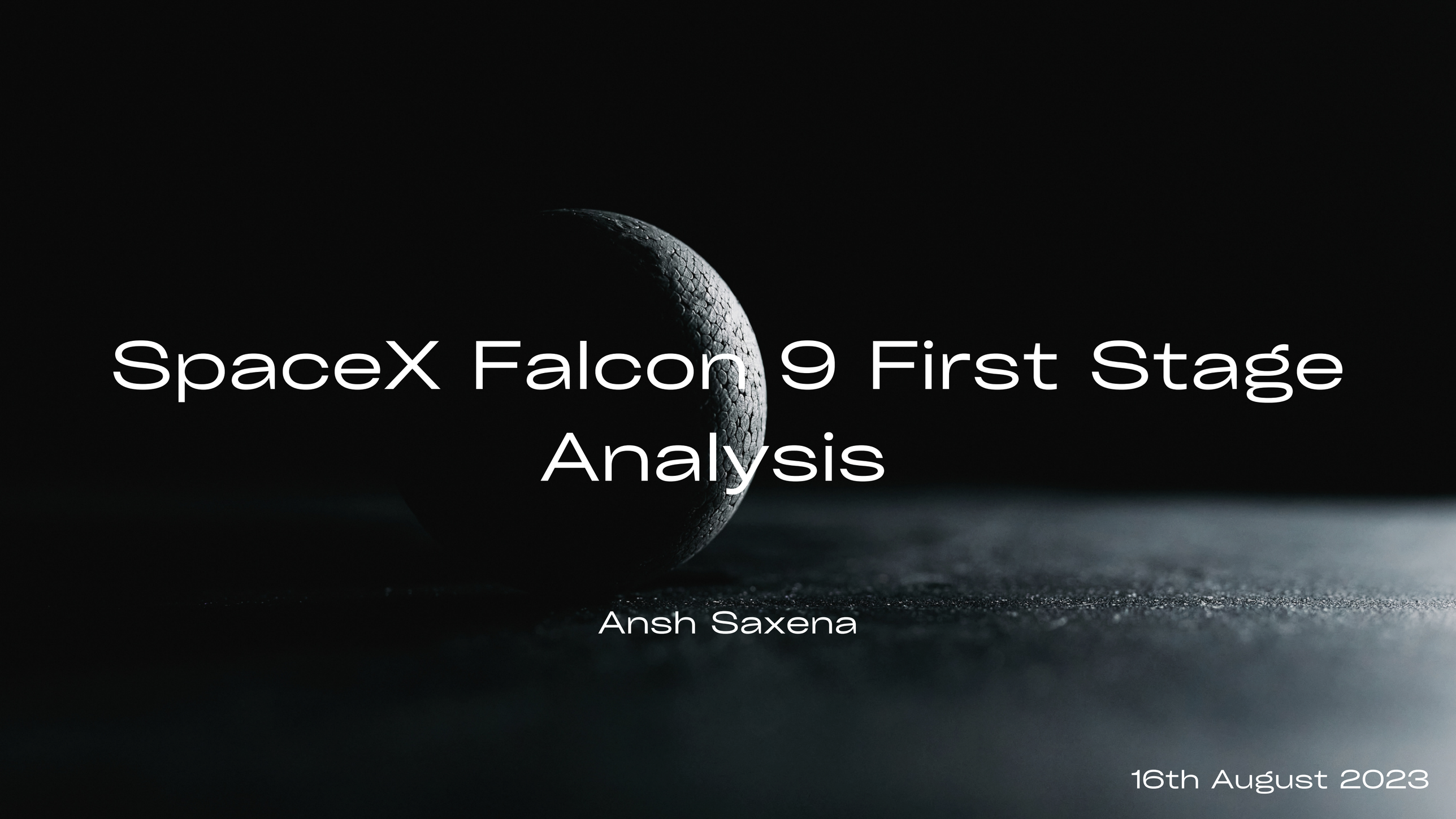


SpaceX Falcon 9 First Stage Analysis

Ansh Saxena

16th August 2023

EXECUTIVE SUMMARY

This executive summary offers a succinct overview of the pivotal insights and conclusions from an in-depth data analysis of SpaceX's Falcon 9 first-stage landing. SpaceX launches rockets at one-third cost compared to others as they reuse the first stage. The analysis aimed to extract valuable patterns, trends, and actionable intelligence on SpaceX's first-stage landing.

Follows a classic method for this analysis includes:

- **Data Collection** - Requesting the SpaceX API and cleaning it.
- **Web Scraping** - Extracting a Falcon 9 records HTML table from Wikipedia and parsing it.
- **Data Wrangling** - Perform EDA and determine training labels.
- **Exploratory Data Analysis** - Using SQL, Matplotlib, and Folium libraries.
- **Plotly Dashboard** - Creating a Dashboard.
- **Predictive Analysis** using Machine learning algorithms.

TABLE OF CONTENTS

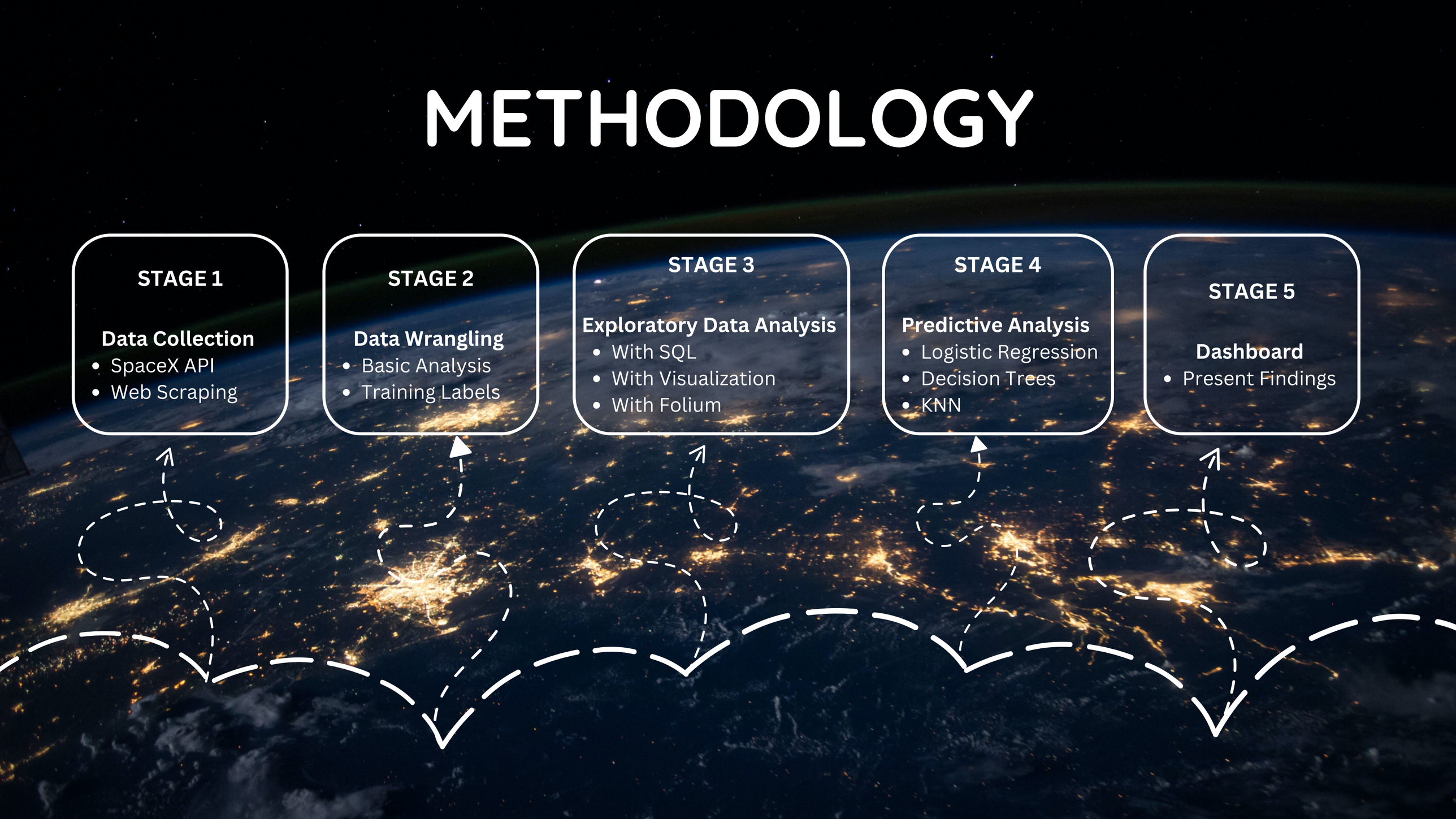
INTRODUCTION

In this analysis, we try to predict if SpaceX's Falcon 9 first stage will land successfully or not. For this, we collect the previous first landing data, clean it, normalize it, analyze and visualize it to get more insights about the data, and then build a predictive model to predict.

SpaceX advertises Falcon 9 launches with a cost of 62 million dollars, and other providers cost upward of 165 million dollars, this difference in the cost is because SpaceX can reuse the first stage.
Therefore we focus on determining "If the first stage will land?", and use this analysis to determine the cost of a launch.

This information will be useful for an alternate company if they want to bid against SpaceX for a rocket launch.

METHODOLOGY



STAGE 1 : DATA COLLECTION

SpaceX API

For Notebook: Click Here

- Request & parse the SpaceX launch data using GET request.
- Filter data that includes Falcon 9.
- Export it to a .csv file.

Web Scraping

For Notebook: Click Here

- Request the Falcon 9 launch Wikipedia page from its URL.
- Extract all column/variable names and create a data frame by parsing the launch HTML tables.
- Export it to a .csv file.

STAGE 2 : DATA WRANGLING

Basic Analysis

- Dealing with the missing values.
- Calculate the number of launches from each site, occurrence of each orbit.

For Notebook: Click Here

Training labels

- Calculate the number and occurrence of the mission outcome of the orbits.
- Create a landing outcome label from the Outcome column.
- Export it to .csv file.

STAGE 3 : EDA

SQL

For Notebook: Click Here

- Perform Exploratory data analysis using SQL queries.

Visualization

For Notebook: Click Here

- Perform Exploratory data analysis using a cat plot, scatterplot, and line chart.
- Apply data feature Engineering.

Folium

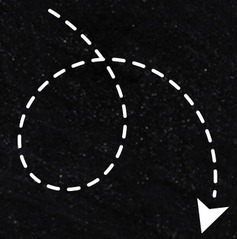
For Notebook: Click Here

- Mark all launch sites on a map.
- Mark the success/failed launches for each site on the map.
- Calculate the distances between a launch site to its proximities.

STAGE 4 : PREDICTIVE ANALYSIS

Preparing data

- Create a column for the class.
- Standardize the data.
- Split into training data and test data.



For Notebook: Click Here

Training and Testing Models

- **Find best** Hyperparameters and accuracy for -
  - Logistic Regression
  - Decision Tree
  - Support Vector Machine
  - KNN
- Find algorithm performs best using test data.

STAGE 5 : DASHBOARD

Making Dashboard that shows

- All sites with launches done from each site and their contribution in overall launches.
- A Scatter graph of Class Vs Payload Mass with Booster version category.
- A Payload Range setter above Scatter graph to set the range.

For Notebook: Click Here

RESULTS

Exploratory Data Analysis

- With SQL
- With Visualization
- With Folium

Plotly Dashboard

Predictive Analysis

EDA RESULTS

With SQL

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

```
%%sql

SELECT DISTINCT Launch_Site FROM SPACEXTBL
```

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

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

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

Hint:Use min function

```
%%sql

SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome='Success (ground pad)';
```

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

MIN(Date)
2015-12-22

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

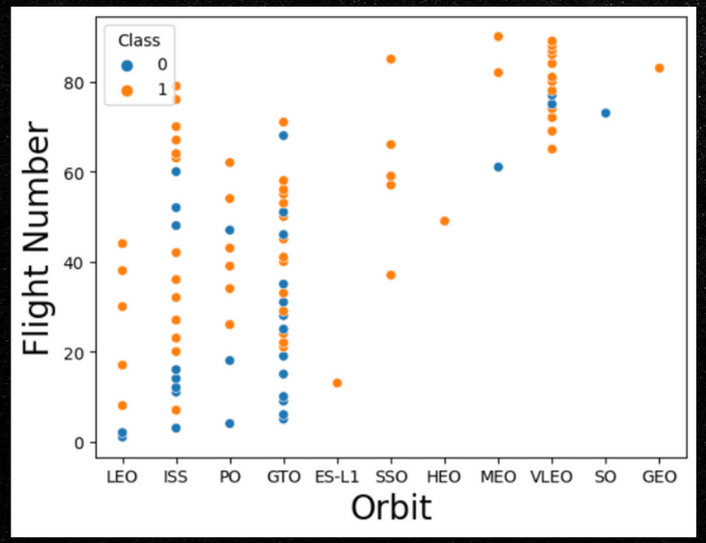
```
%%sql

SELECT AVG(payload_mass__kg_) FROM SPACEXTBL WHERE booster_version='F9 v1.1';
```

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

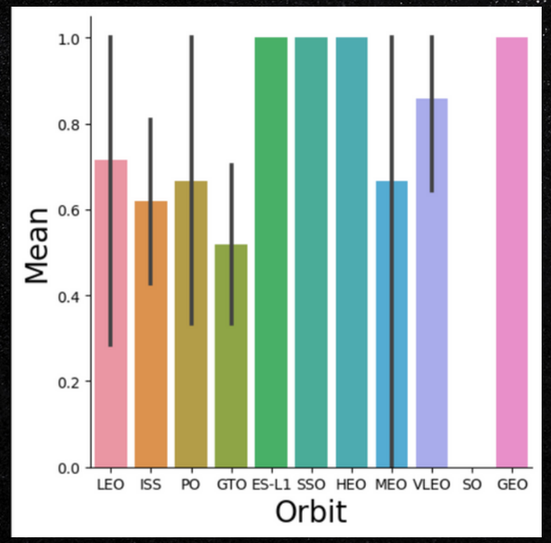
AVG(payload_mass__kg_)
2928.4

EDA RESULTS

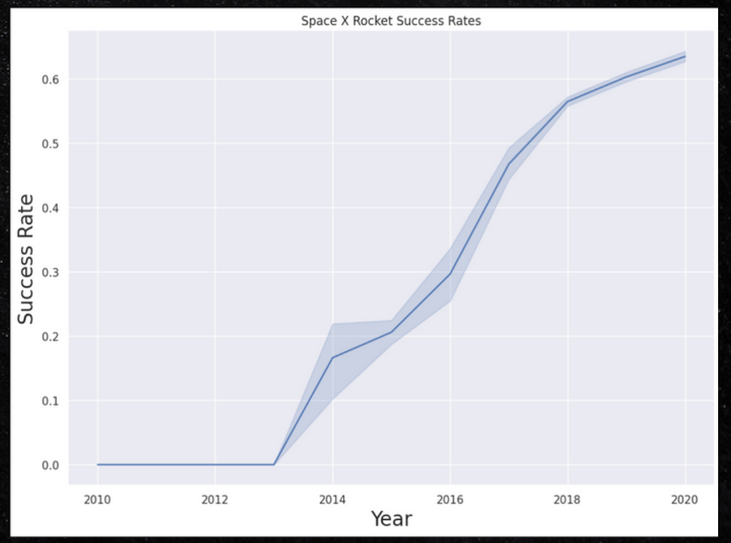
With Visualization



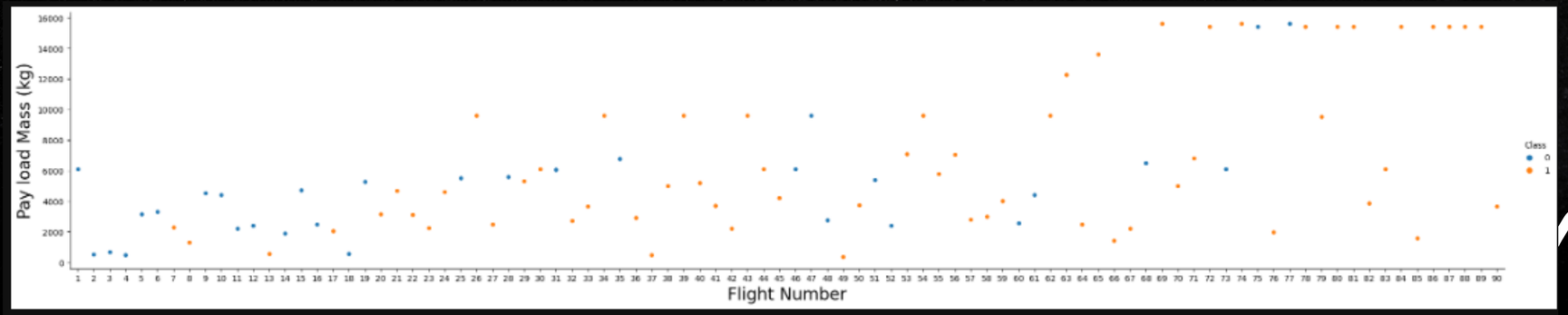
Flight Number Vs Orbit (Scatter Plot)



Success rate of each Orbit (Bar Chart)



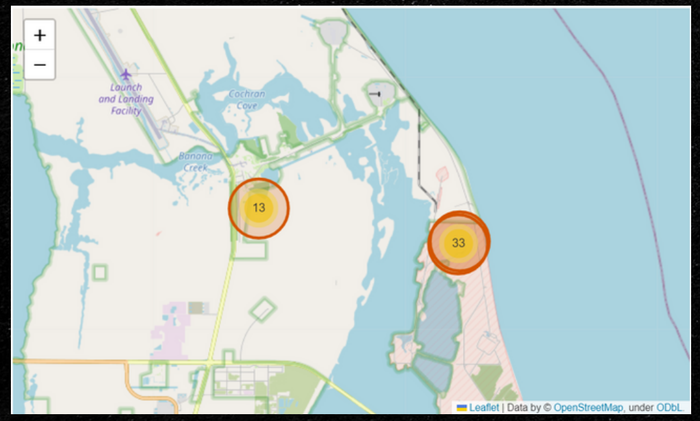
Yearly Success Trend (Line Plot)



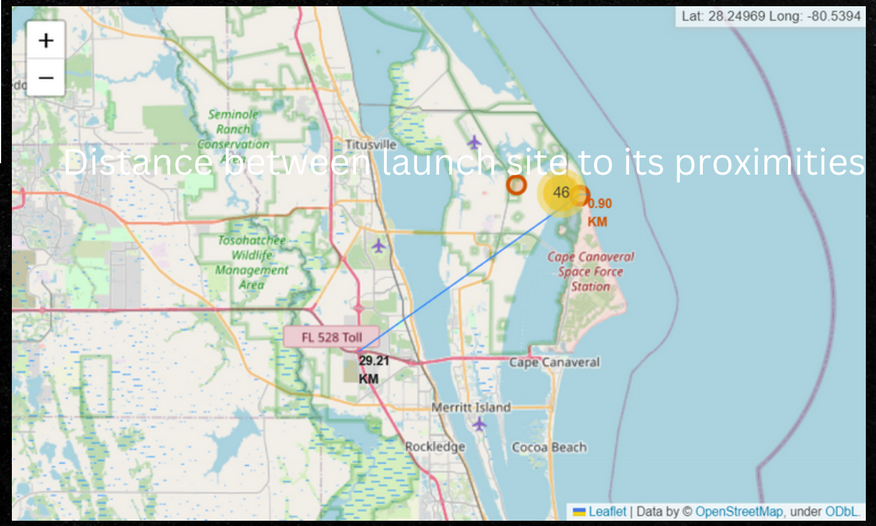
Payload Mass Vs Flight Number (Cat Plot)

EDA RESULTS

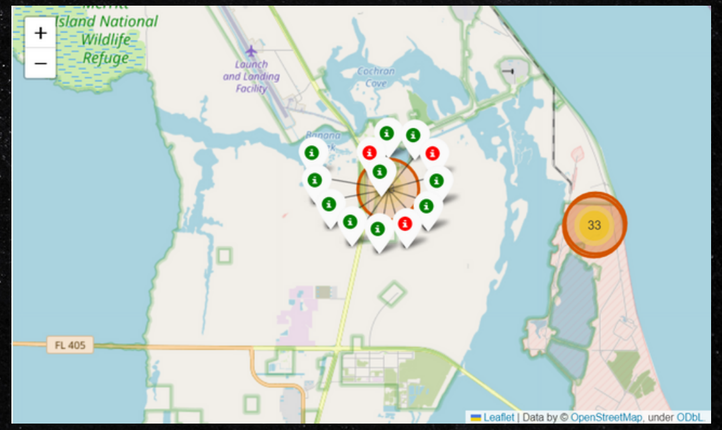
With Folium



Mark all launch sites

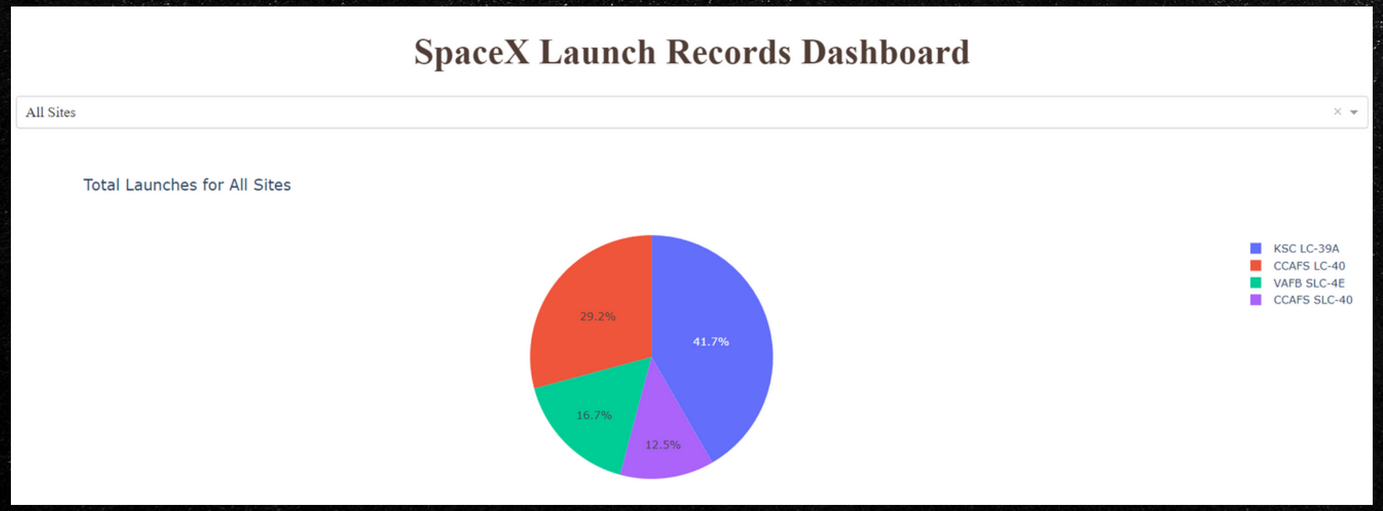


Distance between launch site to its proximities



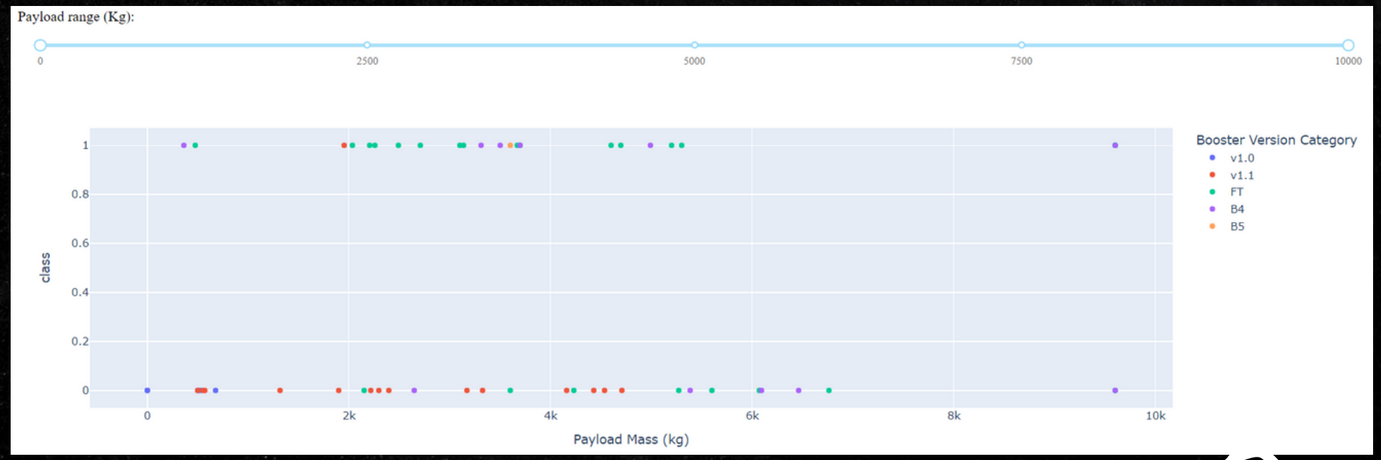
Mark success/failed launches for each site

PLOTLY DASHBOARD



Dashboard shows launches from all sites, and also from each particular site as we choose site name from the dropdown menu.

This part contains Payload range which is used to update the scatter plot between Class and payload mass based on Booster version category.



PREDICTIVE ANALYSIS

Find the method performs best:

```
algorithms = {'LogisticRegression':logreg_cv.best_score_, 'SVM':svm_cv.best_score_, 'Tree':tree_cv.best_score_, 'KNN':knn_cv
bestalgorithm = max(algorithms, key=algorithms.get)
print('Best Algorithm is',bestalgorithm,'with a score of',algorithms[bestalgorithm])
if bestalgorithm == 'LogisticRegression':
    print('Best Params is :',logreg_cv.best_params_)
if bestalgorithm == 'SVM':
    print('Best Params is :',svm_cv.best_params_)
if bestalgorithm == 'Tree':
    print('Best Params is :',tree_cv.best_params_)
if bestalgorithm == 'KNN':
    print('Best Params is :',knn_cv.best_params_)
```

```
Best Algorithm is Tree with a score of 0.875
Best Params is : {'criterion': 'entropy', 'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split':
5, 'splitter': 'random'}
```

By comparing all algorithms we get accuracy of each as follows -

- Logistic Regression - 84.6%
- Decision Tree - 87.5% **(Best Accuracy)**
- Support Vector Machine - 84.8%
- KNN - 84.8%

CONCLUSION

- We get to know that KSC-LC-39A and VAFB-SLC-4E launch sites has high success rate of 77%.
- VAFB-SLC launch site has no rockets launched for payload greater than 10000 kgs.
- Success rate is best for ES-L1, SSO, HEO and GEO orbit types.
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- SpaceX first stage successful landing rate kept increasing since 2013 and reached 65% in 2020.
- All launch sites in very close proximity to the coast.
- All launch sites keep certain distance away from cities for safety purpose.


THANK YOU
VERY MUCH!