# Data science project for Space X

By Codifi215

Oct 22, 2024

# Outline

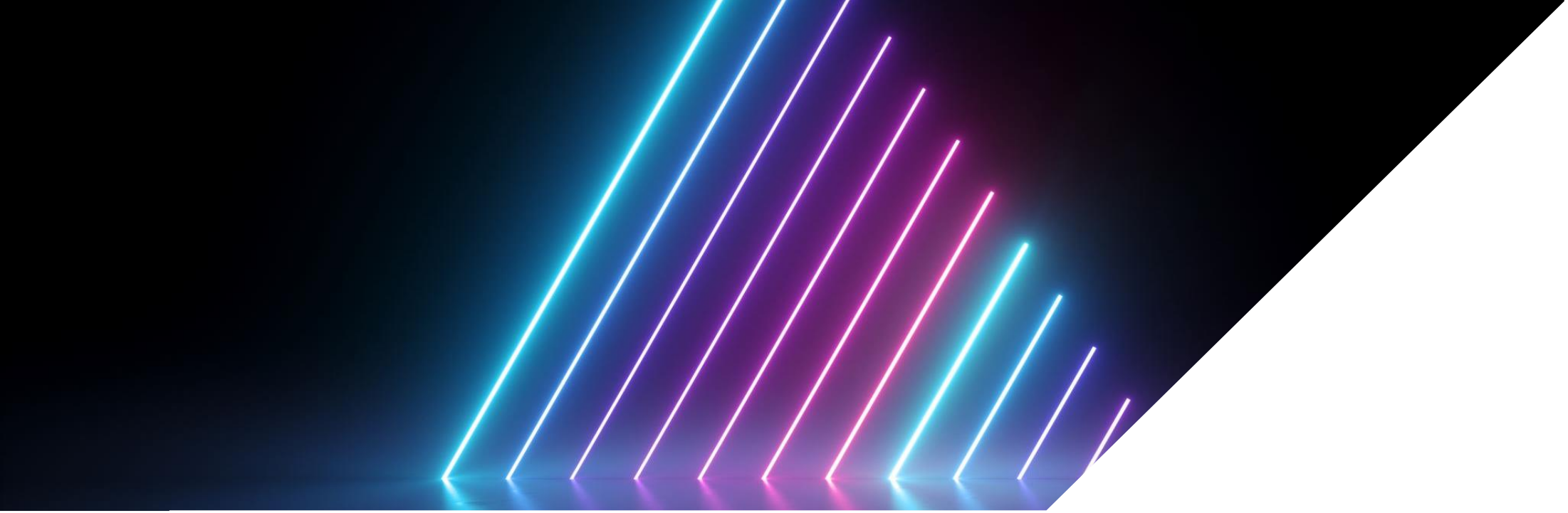
- Summary
- Introduction
- Methodology
- Results
- Conclusion

# summary

- methodologies
- Collection through API
- Collection with Web Scraping
- Wrangling
- Analysis with SQL
- Data Visualization
- Machine Learning
- screenshots

# Introduction

- SpaceX is an American aerospace company founded in 2002 by Elon Musk. Its mission is to make spaceflight affordable and accessible, ultimately enabling human life on Mars.
- Reusable Rockets: SpaceX pioneered the concept of reusable rockets, significantly reducing the cost of spaceflight. Their Falcon 9 and Falcon Heavy rockets have successfully landed and been reused multiple times.
- Commercial Spaceflight: SpaceX was the first private company to launch, orbit, and recover a spacecraft. They have also sent cargo and astronauts to the International Space Station (ISS).
- Starlink: SpaceX is developing Starlink, a satellite internet constellation that aims to provide high-speed internet to people around the world, especially in remote areas.
- Key achievements and ongoing projects include:
- The development of the Starship, a fully reusable spacecraft designed to transport humans and cargo to the Moon, Mars, and beyond.
- The development of advanced propulsion technologies.



# Methodologies

# methodologies

- Executive Summary
- • Data collection methodology:
- • Data was collected using SpaceX REST API and web scrapping from Wikipedia
- • Perform data wrangling
- • Data was processed using one-hot encoding for categorical features
- • 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

Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. As mentioned, the dataset was collected by REST API and Web Scrapping from Wikipedia

For REST API, its started by using the get request. Then, we decoded the response content as Json and turn it into a pandas dataframe using json_normalize(). We then cleaned the data, checked for missing values and fill with whatever needed.

For web scrapping, we will use the BeautifulSoup to extract the launch records as HTML table, parse the table and convert it to a pandas dataframe for further analysis

# Collection through API

**Request data through API**

**Convert JSON result to data frame**

**Cleaning data**

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

```
# Lets take a subset of our dataframe keeping only the features we want a
nd the flight number, and date_utc.
data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number',
'date_utc']]

# We will remove rows with multiple cores because those are falcon rocket
s with 2 extra rocket boosters and rows that have multiple payloads in a
single rocket.
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]

# Since payloads and cores are lists of size 1 we will also extract the s
ingle value in the list and replace the feature.
data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])

# We also want to convert the date_utc to a datetime datatype and then ex
tracting the date leaving the time
data['date'] = pd.to_datetime(data['date_utc']).dt.date

# Using the date we will restrict the dates of the launches
data = data[data['date'] <= datetime.date(2020, 11, 13)]
```

# Scraping

**Request data from WIKI url**

**Create a BeautifulSoup from the HTML response**

**Extract all column/variable names from the HTML header**

# Wrangling

Data Wrangling is the process of cleaning and unifying messy and complex data sets for easy access and Exploratory Data Analysis (EDA).

First calculate the number of launches on each site, then calculate the number and occurrence of mission outcome per orbit type.

This makes it easier for further analysis

# EDA with Data Visualization

We first started by using scatter graph to find the relationship between the attributes such as between:

- Payload and Flight Number.
- Flight Number and Launch Site.
- Payload and Launch Site.
- Flight Number and Orbit Type.
- Payload and Orbit Type.

Scatter plots show dependency of attributes on each other. Once a pattern is determined from the graphs. It's very easy to see which factors affecting the most to the success of the landing outcomes.

# EDA with SQL

Using SQL, we had performed many queries to get better understanding of the dataset, Ex:

- Displaying the names of the launch sites.

- Displaying 5 records where launch sites begin with the string ‘CCA’.

- Listing the failed landing_outcomes in drone ship, their booster versions, and launch sites

names for in year 2015.

- Rank the count of landing outcomes

# Predictive Analysis (Classification)

Building the Model

•Load the dataset into

NumPy and Pandas

•Transform the data and

then split into training and

test datasets

•Decide which type of ML to

use

•set the parameters and

algorithms to

Evaluating the Model

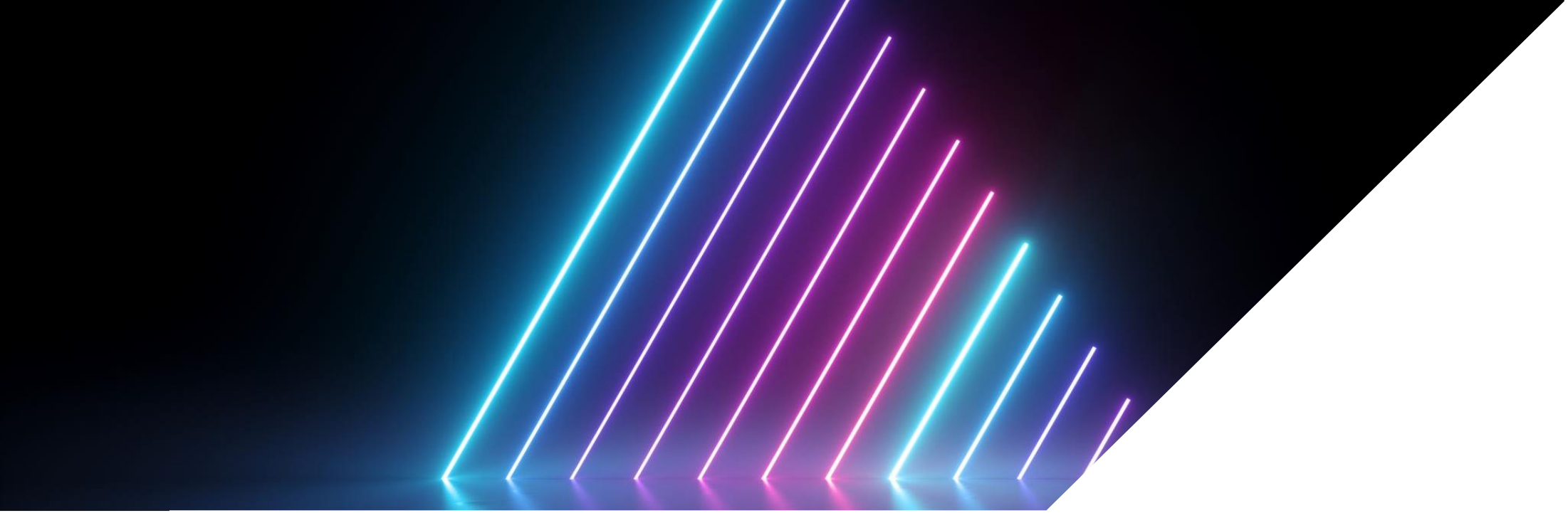
•Check the accuracy for each
model

•Get tuned hyperparameters for each type of algorithms.

•plot the confusion matrix.

Improving the Model

•Use Feature Engineering
and Algorithm Tuning



# Conclusion

# Conclusion

There are 3 main result as below.

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