Table of Contents

[Project Scenario and Overview 1](#_Toc100265829)

[Data preparation: 2](#_Toc100265830)

[Data collection 2](#_Toc100265831)

[Data wrangling 3](#_Toc100265832)

[Web scraping from Wikipedia 3](#_Toc100265833)

[Extract a Falcon 9 launch records HTML table from Wikipedia 3](#_Toc100265834)

[Parse the Table and convert it into a pandas data fram 4](#_Toc100265835)

[Data wrangling 5](#_Toc100265836)

[Exploring and Preparing Data 7](#_Toc100265837)

[Exploratory Data Analysis 7](#_Toc100265838)

[Preparing Data Feature Engineering 8](#_Toc100265839)

[Launch Sites Locations Analysis with Folium 8](#_Toc100265840)

[Mark all launch sites on a map 8](#_Toc100265841)

[Mark the success/failed launches for each site on the map 9](#_Toc100265842)

[Calculate the distances between a launch site to its proximities 9](#_Toc100265843)

[Build a Dashboard Application with Plotly Dash 9](#_Toc100265844)

[Result 9](#_Toc100265845)

# Project Scenario and Overview

Your job is to determine the price of each launch.

You will do this by gathering information about Space X and creating dashboards for

your team.

You will also determine if SpaceX will reuse the first stage.

Instead of using rocket science to determine if the first stage will land successfully,

you will train a machine learning model and use public information to predict if SpaceX

will reuse the first stage.

Our goal is to use this data to predict whether SpaceX will attempt to land a rocket or not

1. Objective:

* Determine the price of launch – by determining if the first stage will land
* Train ML model to predict if SpaceX will reuse the first stage.

1. Classification

0 – the first stage did not land successfully

1 – the first stage landed successfully

FlightNumber (indicating the continuous launch attempts.)

# Data preparation:

## Data collection

The rocket launch data is collected from SpaceX API. The initial data contains too much massive information and some are irrelevant.

Graphical user interface

Description automatically generated with low confidence

* From the rocket we would like to learn the booster name
* From the payload we would like to learn the mass of the payload and the orbit that it is going to
* From the launchpad we would like to know the name of the launch site being used, the longitude, and the latitude.
* From cores we would like to learn the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, whether the core is reused, whether legs were used, the landing pad used, the block of the core which is a number used to seperate version of cores, the number of times this specific core has been reused, and the serial of the core.

Table

Description automatically generated

Table

Description automatically generated

## Data wrangling

Remove null values (missing values)

The missing values of payloadmass is replaced by the mean

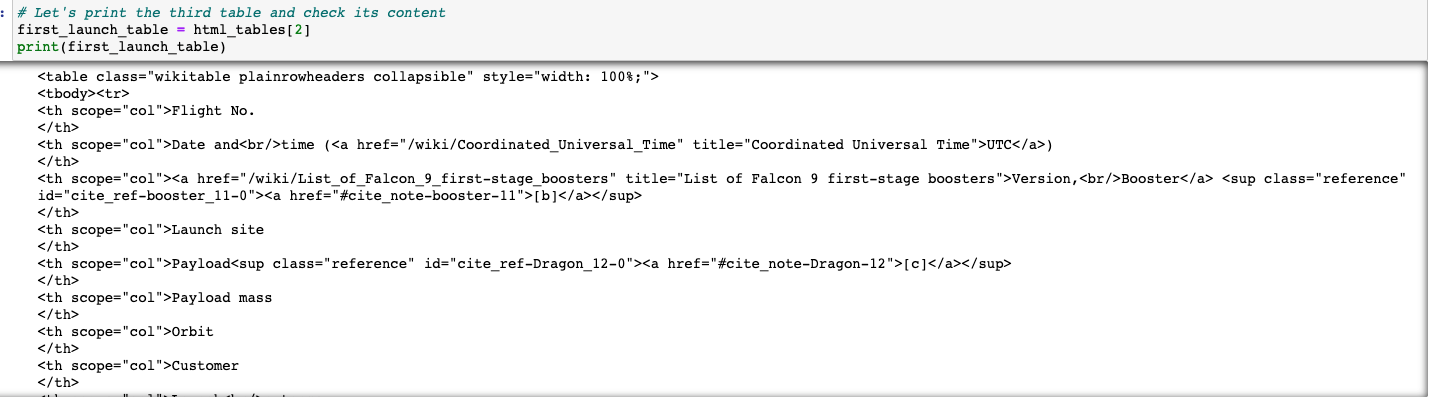
# Web scraping from Wikipedia

## Extract a Falcon 9 launch records HTML table from Wikipedia

Graphical user interface, text, application

Description automatically generated

The default format of data derived from wikipidia



## Parse the Table and convert it into a pandas data fram

Graphical user interface

Description automatically generated

# Data wrangling

Graphical user interface, table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Calculate the number and occurrence of each orbit

A picture containing table

Description automatically generated

Calculate the number and occurrence of mission outcomes per orbit type

A picture containing table

Description automatically generated

True Ocean means the mission outcome was successfully landed in a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed in a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed to a drone ship False ASDS means the mission outcome was unsuccessfully landed to a drone ship. None ASDS and None None these represent a failure to land.

Create class representation for the outcome of each launch.

# SQL

Graphical user interface, text, application, email

Description automatically generated

# Exploring and Preparing Data

## Exploratory Data Analysis

Python package: seaborn

* Using scatter plot to display the relationship between:
  + Payload mass vs. flight Number

Chart, scatter chart

Description automatically generated

* + Launch site vs. flight number

Chart, scatter chart

Description automatically generated

* + Launch site vs. payload mass

Chart, application, scatter chart

Description automatically generated with medium confidence

* + Flight number vs. orbit type

Chart, scatter chart

Description automatically generated

* + Payload mass vs. orbit

Chart

Description automatically generated with medium confidence

* Bar plot to display the success rate of each orbit

Chart, bar chart

Description automatically generated

* Line plot to display the success rate of each year

Chart, line chart

Description automatically generated

## Preparing Data Feature Engineering

* Using pd.get\_dummies to create a new dataframe

# Launch Sites Locations Analysis with Folium

Interactive visual tool - Foliun

## Mark all launch sites on a map

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Mark the success/failed launches for each site on the map

MarkerCluster

Marker clusters can be a good way to simplify a map containing many markers having the same coordinate.

## Calculate the distances between a launch site to its proximities

# Build a Dashboard Application with Plotly Dash

# Result

Graphical user interface, chart, application, Excel

Description automatically generated

Graphical user interface, chart, table, Excel

Description automatically generated

Chart

Description automatically generated

Chart, Excel

Description automatically generated

Chart, table

Description automatically generated

# Machine Learning Prediction