

# DATA SCIENCE PROJECT

◇ Bruno Paulillo 29/07/2024

# Executive Summary

## Summary of methodologies

- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

## Summary of all results

- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results





# Introduction

## Project background and context

SpaceX is the most successful company of the commercial space age, making space travel affordable. The company 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. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

## Questions to be answered

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?



# Methodology

## Data collection methodology:

- Using SpaceX Rest API
- Using Web Scrapping from Wikipedia

## Performed data wrangling

- Filtering the data
- Dealing with missing values
- Using One Hot Encoding to prepare the data to a binary classification

## Performed exploratory data analysis (EDA) using visualization and SQL

## Performed interactive visual analytics using Folium and Plotly Dash

## Performed predictive analysis using classification models

- Building, tuning and evaluation of classification models to ensure the best results

