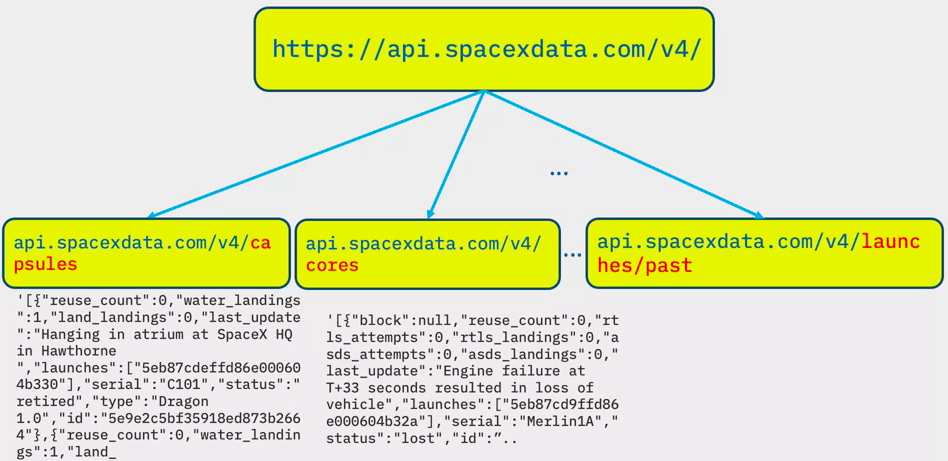
**APPLIED DATA SCIENCE CAPSTONE**

Data Collection Overview

1. Using a rest API to collect spaceX data



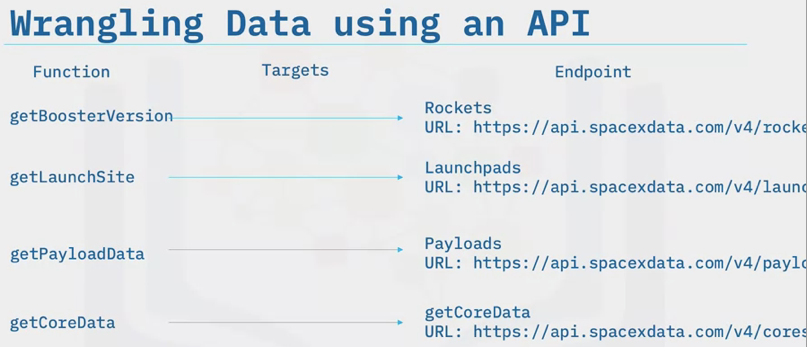
1. url = <https://api.spacexdata.com/v4/launches/past>
   1. Use this url to target a specific endpoint of the API and get past launch data
   2. Perform a get request using the requests library to obtain the launch data which will be used to get the data from the API
   3. This result can be viewed by calling the “response.json()” method
   4. Response will be in the form of a json, specifically a list of json objects

Wrangling Data Using an API

1. To convert the json to a dataframe we can use “pd.json\_normalize(response.json)))” function
   1. This will allow us to normalize the json data into a flat table

Web Scraping Falcon 9

1. Use BeautifulSoup package to scrape html tables that contain valable falcon 9 launch records
2. Parse the data and convert them into a pandas dataframe for further visualization and analyses



1. In some of the columns we have an identification number, rather than actual data
2. We will need to use the API again, targeting another endpoint gathering data for each ID number
3. These functions are already pre-created

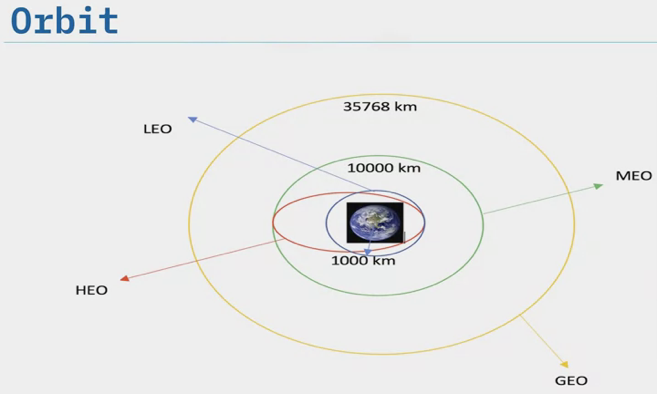
Sampling Data

1. The launch data also includes data for the Falcon 1 booster, but we only want Falcon 9
2. In this lab, you will figure out how to sample data and remove Falcon 1 launches

Dealing with Nulls

1. We will deal with the null values inside the payload mass
2. You will figure out a way to calculate the mean of the payload mass data and then replace the null values in the payload mass with the mean

Data Wrangling Overview



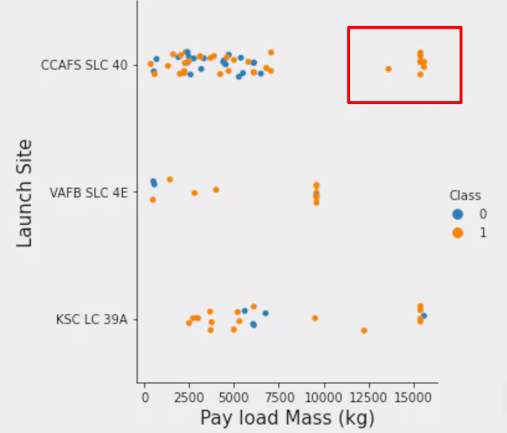
1. The column “orbits” are the different orbit of the payload
   1. LEO: Low earth orbit – altitude of 2000km
   2. GTO: Geosynchronous Orbit – high earth orbit allowing satellites to match earth’s rotation. Located at 35786km
2. Outcome: indicates if the first stage successfully landed
   1. True ASDS = booster successfully landed to a drone ship
   2. False ASDS = unsuccessfully landed

Outcome

1. Outcomes to be converted to classes *y* (either 0 or 1)
   1. 0 = bad outcome
   2. 1 = god outcome

Exploratory Data Analysis

1. The first step of any data science project
2. In the first lab, you will perform some exploratory data analysis using a database
3. In the second lab, you will see if the data can be used to automatically determine if the Falcon 9 second lab will lab
4. Some attributes can be used to determine if the first stage can be reused
5. This can be then used with machine learning to automatically determine if the first stage can land successfully
   1. Different launch sites have different success rates, as a result, they can be used to help determine if the first stage will land successfully



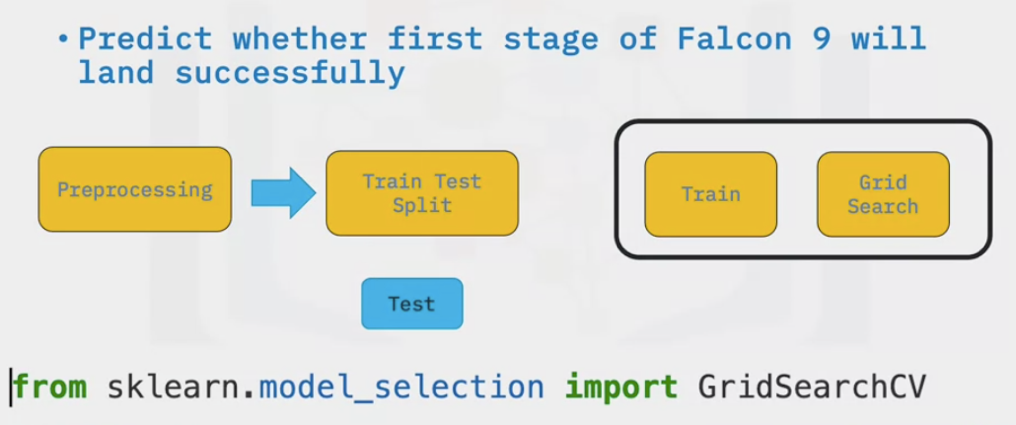
1. Categorically variables will be converted using one hot encoding preparing the data for a machine learning model that will predict if the first stage will successfully land

Interactive Visual Analytics

1. Enables users to explore and manipulate data in an interactive and real-time way
2. Common interactions include:
   1. Zoom
   2. Pan
   3. Filter
   4. Search
   5. Link
3. Users can find visual patterns faster and more effectively
4. Interactive visualization (dashboarding) – as opposed to static graphs – can tell a more appealing story
5. Folium and plotlydash

Predictive Analysis (Classification)

1. Build a machine learning pipeline
   1. Predict whether the first stage of Falcon 9 will land successfully
   2. Preprocessing
   3. Train, test, split (split data into training and testing





**How to Present your Findings**

Elements of a Successful Data Findings Report

1. Findings report: Answering the questions asked before the data is collected, ceaned and organized
2. Varies in structure
   1. Paper report
   2. Slideshow presentation
3. Create an outline
4. Structure towards the audience

Elements

1. Cover page
2. Executive summary
   1. Briefly explain he details
   2. Considered a stand-alone document
3. Table of contents
4. Introduction
   1. Explains the nature of the analysis
   2. States the problem
   3. States the questions for analysis
5. Methodology
   1. Explains the data sources
   2. Outlines the plan for the collected data
6. Results
   1. Details of the data collection
      1. How it was organized ana analyzed
   2. Charts and graphs
7. Discussion
   1. Engage the audience with the implications for the audience
8. Conclusion
   1. Reiterate problem in the introduction
   2. State the findings and if any steps are to be taken in the future
9. Appendix
   1. Contains information that didn’t fit in the main body of the repot but is still deemed important
   2. Locations where the raw data was collected
   3. References and acknowledgements

Best Practices for Presenting your Findings

* + - 1. Delivering your message
         1. Make sure charts and graph are not too small, and are clearly labeled
         2. Use the data only as supporting evidence
         3. Share only one point from each chart
         4. Eliminate data that does not support the key message
      2. Form the key messages – then go back and insert the data to support findings