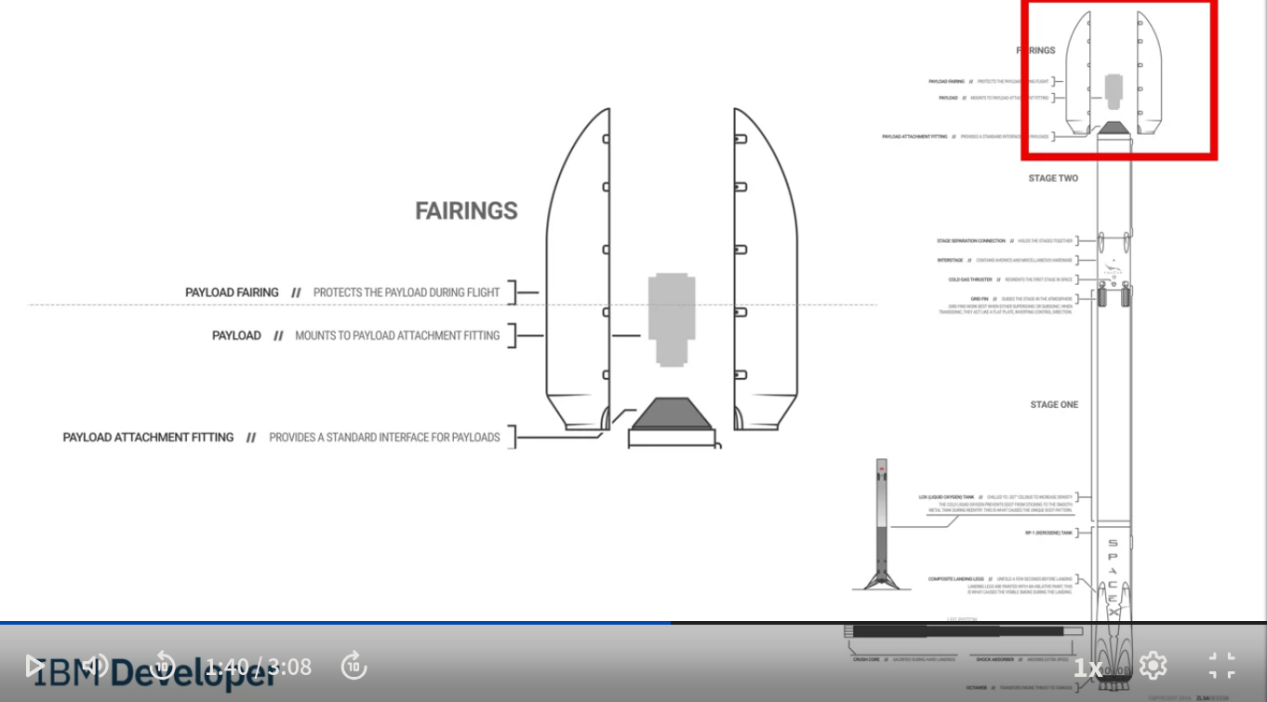
Project Scenario & Overview



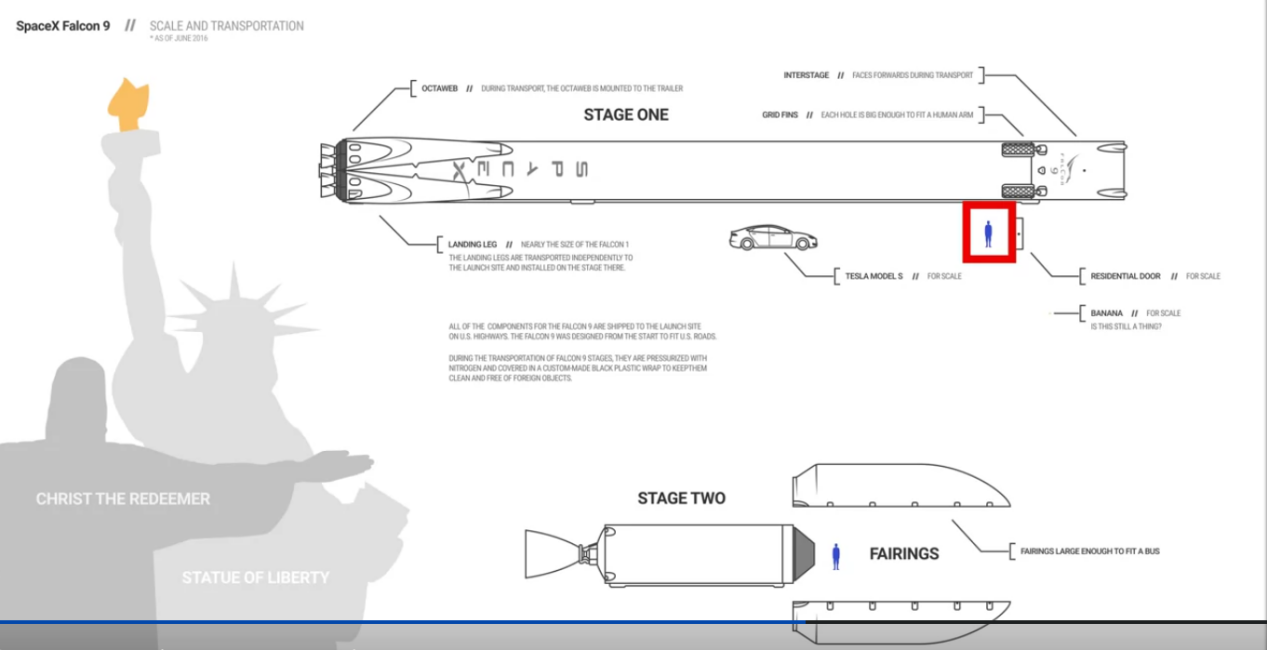
Space startups are emerging. Commercial space travel is starting. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets.



Most successful one is SpaceX. Projects Incude Sending spacecraft to the international space station. Starlink a satellite internet constellation providing satellite internet access. Sending manned mission to space. The project of relatively in expensive. Falcon 62mil other companies cost upwards of 165mil. This is because the reusable first stage of the spaceX.



lets look at the stucture of the Falcon 9. THe payload is inside the fairings. The 2nd stage helps to bring the payload to the orbit. but most of the work is done by the first stage.



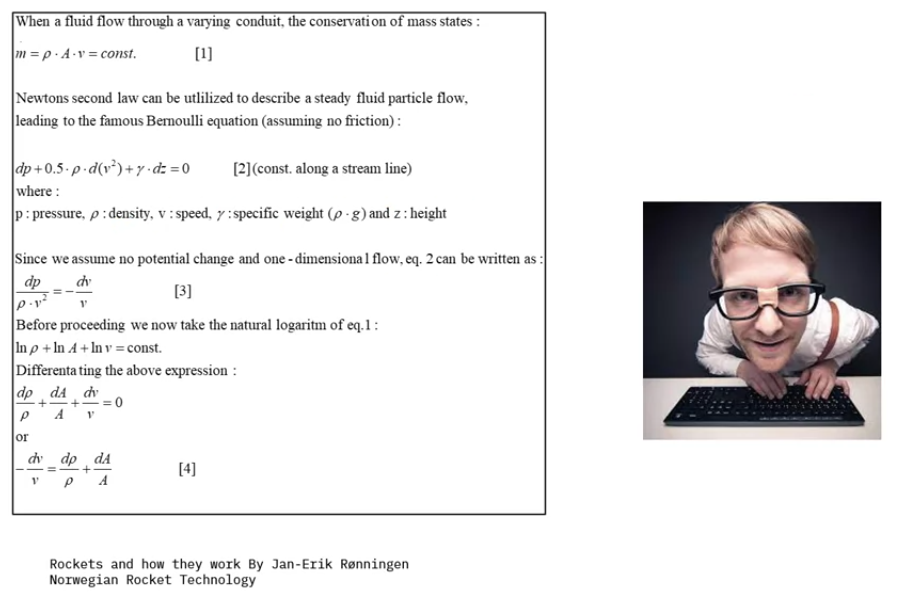
We can first stage of the rocket in the image above and its much larger than the second stage. We can compage the size of it again a person and a car. Spacex can reuse the first stage of the rocket aswell. Sometimes the first stage might blowup or SpaceX might sacrifice the first stage al together.



In this project we will work on a data science project called spaceY that competes with SpaceX.

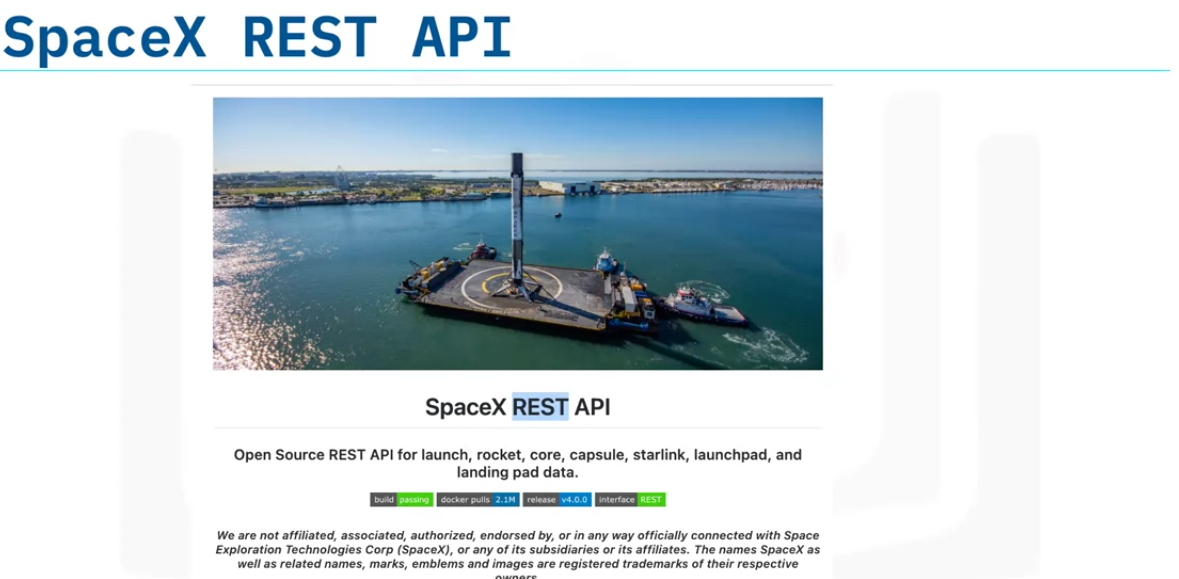
We will determine the price of each lauch we will do this by gatheting data, creating dashboards.

We will determine if SpaceX will reuse the first stage.



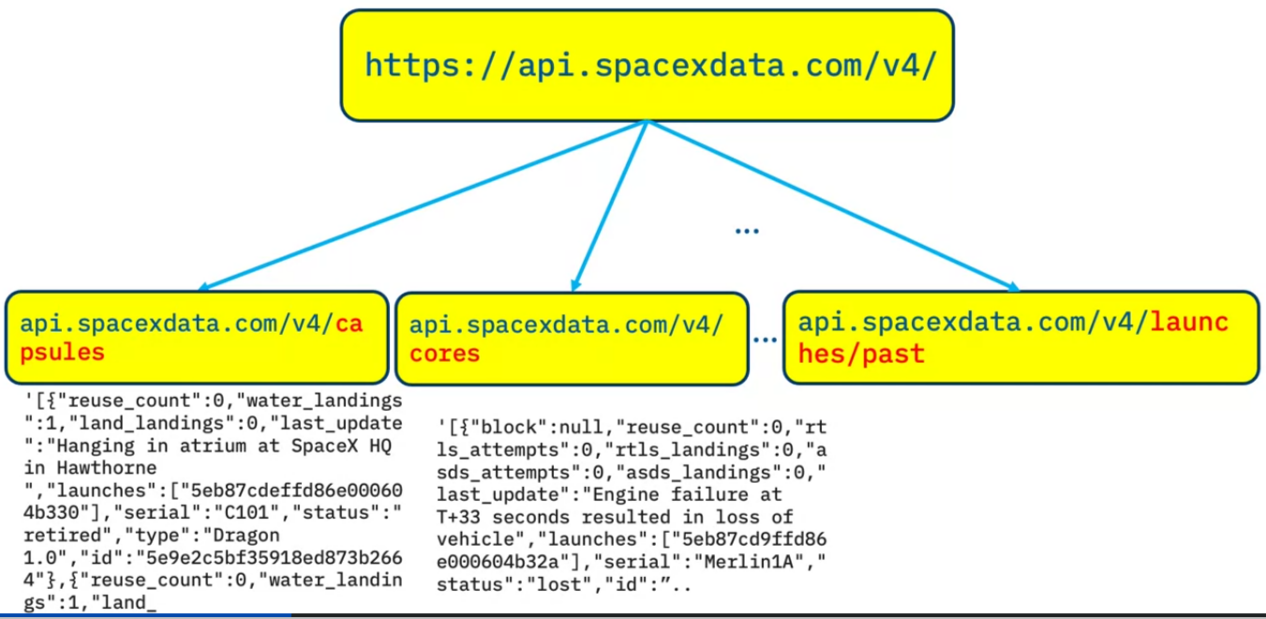
Instead of using Reocket Science , we will train a model that can predicts.

Data Collection Overview



We will use spaceX rest API to collect data.

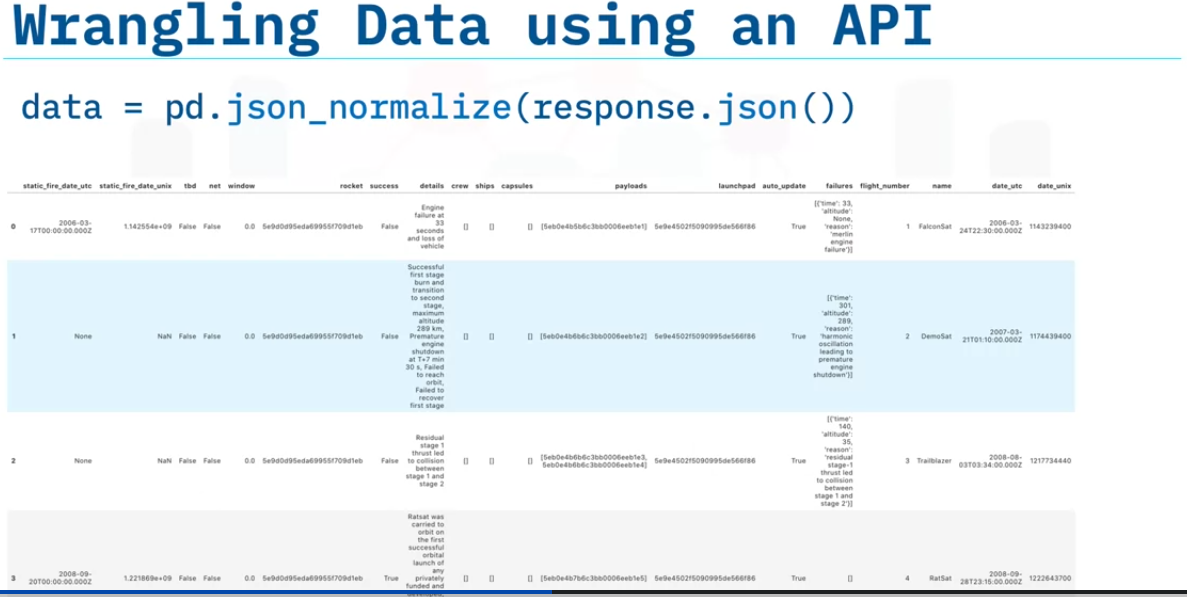
The API gives us data about launches, including information about the rocket used, payload delivered, launch specs, landing specs, and landing outcome. Our goal is to predict whether SpaceX will attempt to land a rocket or not.



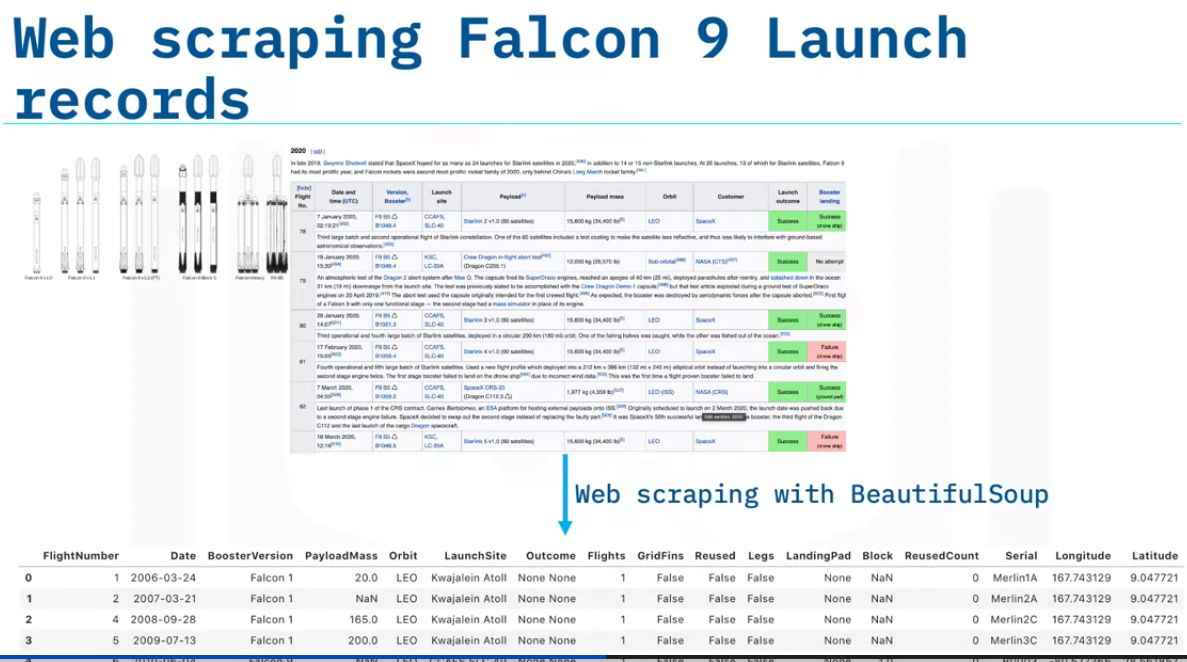
The above image shows the base url for the api and some of the different end points available. WE will use the launches/past



The above image shows a sample python code to grab data from the end point

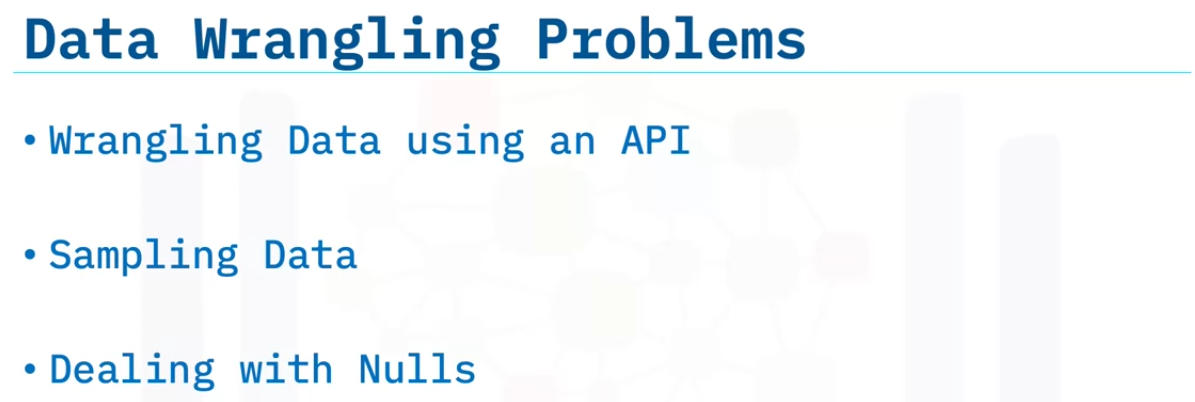


The above data shows loading the data into pandas DF using json\_normalize. It normalizes json by flattening it.

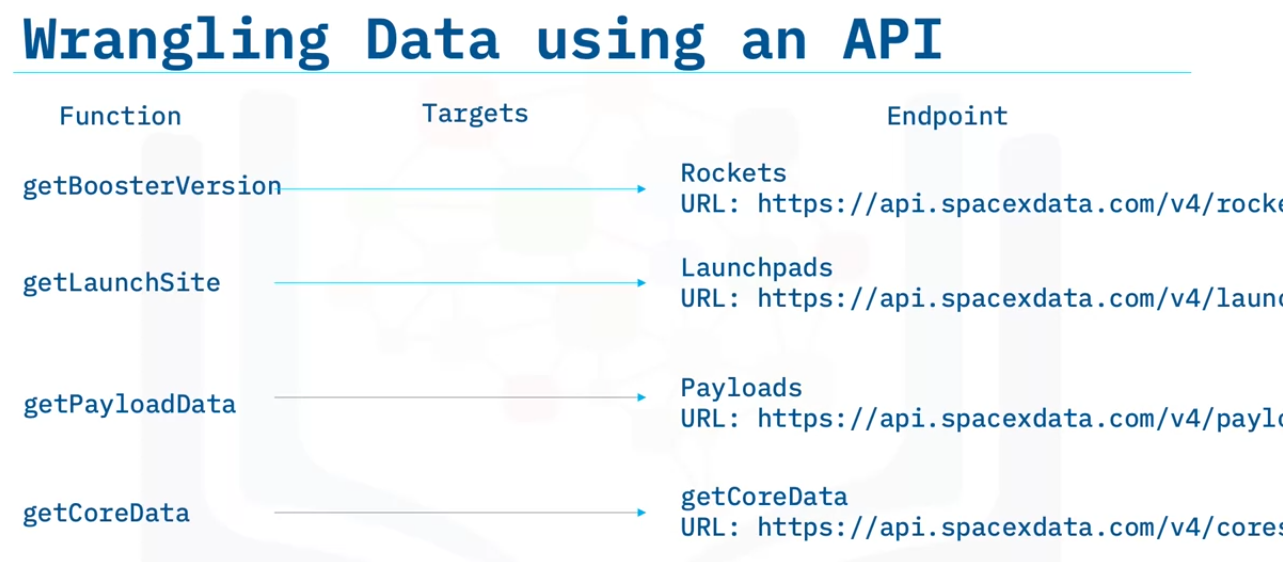


Another way to get the data is by using beatufulsoup on wikipages related to spaceX.

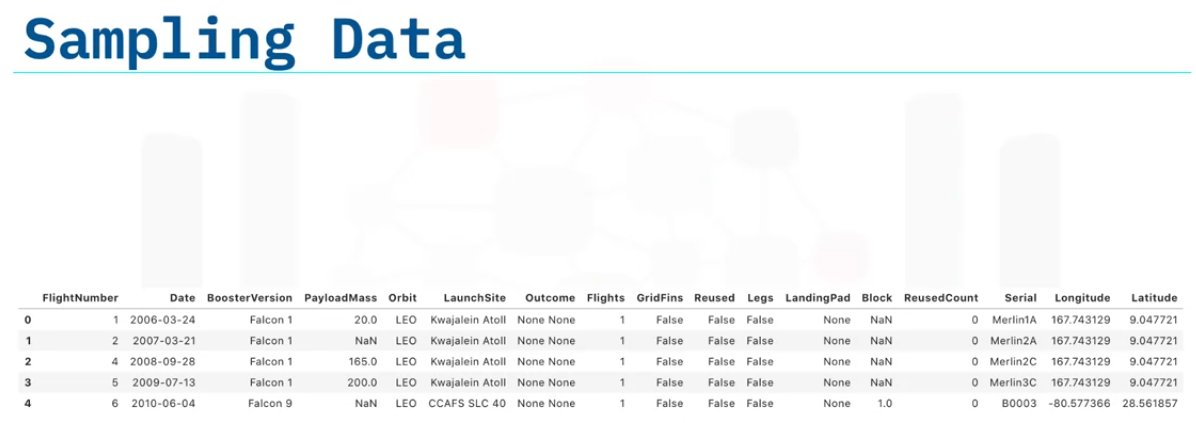
We will use bs4 for extracting some data in this way.



Once the data is collected, we need to perform data wrangling on it to make the data usable this can be by wrangling data using an API, Sampling Data, Dealing with nulls.



We can see in some columns, we have ids insead of real data. This is because data is distributed into different tables and we need to get that data too. We can use other endpoints to do this. In the exercise, the function in the above image are already created for us.

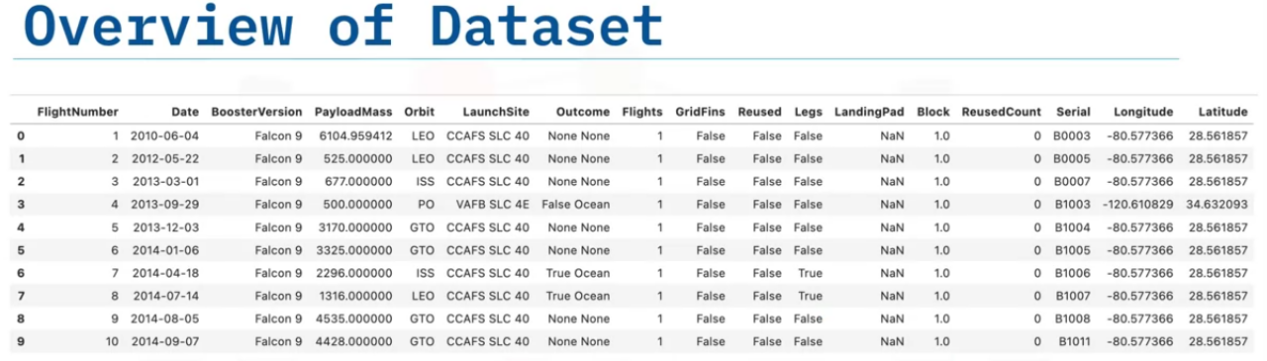


The data containse data related to falcon 1 booster. But we only need falcon 9 data. so we need to find a way to filter the data.

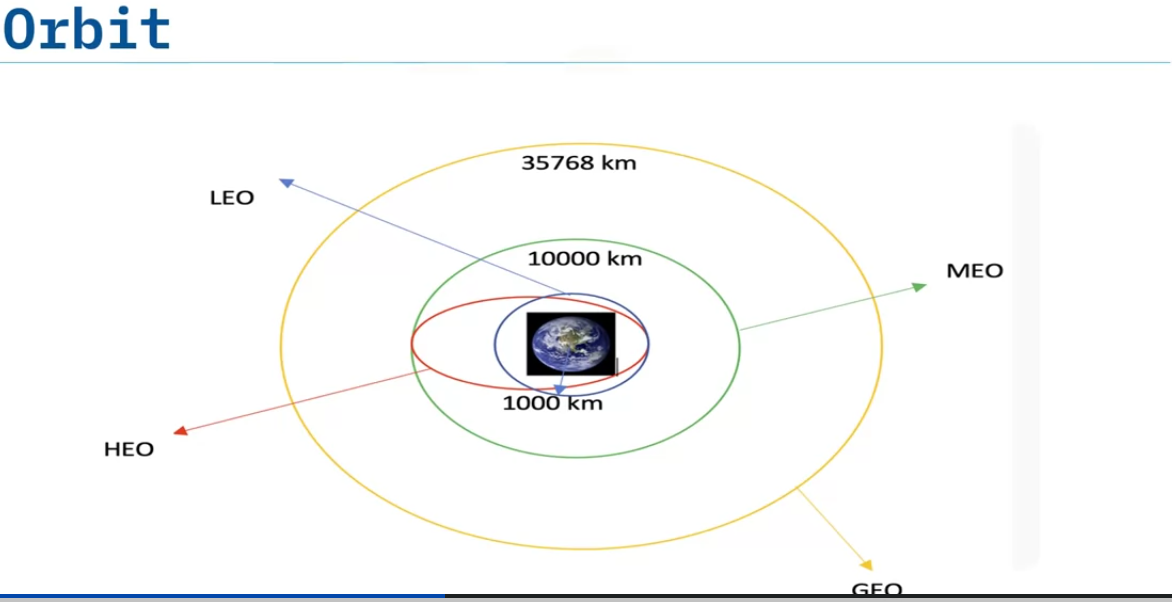
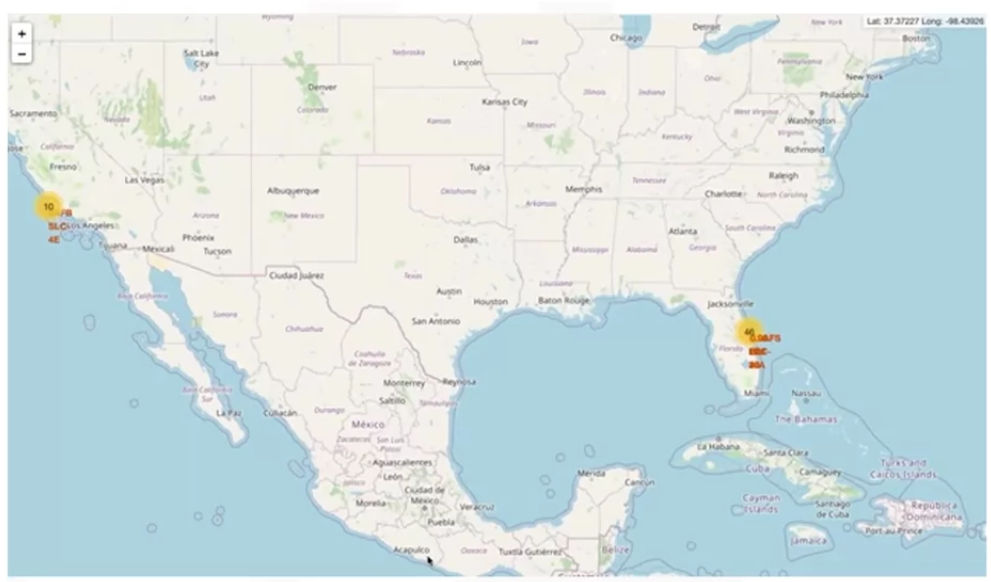


We may have null values. So we need to deal with them. In this case PayloadMass. WE cannot simply delete these rows. So we will fill them with the mean value. We will leave the landing pad with null values for now. We will use one hot encoding later.

Data Wrangling

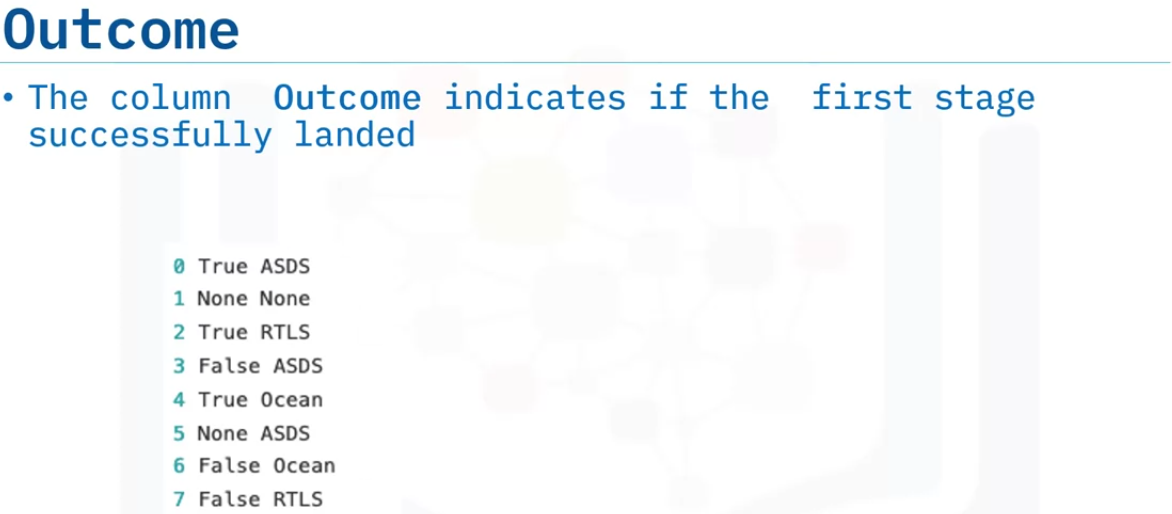


The column launchSite contains the fidderent launch sites, including: Vandenberg, AFB Space Launch, kennedy Space Center, CCAFS SLC 40



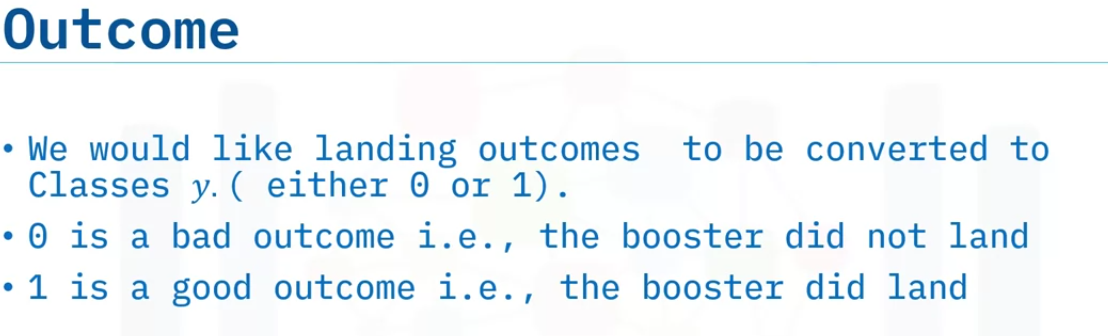
The column orbits are the differnt orbts of the payload. For example LEO: Low Earch orbit with an altitude of 2000 km.

GTO: Geo synchronous orbit is high earth orbit that allows satellites to march Earth’s rotation. It is located at 22236 miles (35,786 Kms) above earth’s equator.



True ASDS means that the booster landed successfully.

False ASDS means the landing failed (crash).



The outcome here is the target variable.