**Project Scenario and Overview**

In this video, we will provide the Project Scenario and Overview. The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX’s accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards 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. Spaces X’s Falcon 9 launch like regular rockets. To help us understand the scale of the Falcon 9, we are going to use these diagrams from Forest Katsch, at  zlsadesign.com. He is a 3D artist and software engineer. He makes infographics on spaceflight and spacecraft art. He also makes software. The payload is enclosed in the fairings. Stage two, or the second stage, helps bring the payload to orbit, but most of the work is done by the first stage. The first stage is shown here. This stage does most of the work and is much larger than the second stage. Here we see the first stage next to a person and several other landmarks. This stage is quite large and expensive. Unlike other rocket providers, SpaceX's Falcon 9 Can recover the first stage. Sometimes the first stage does not land. Sometimes it will crash as shown in this clip. Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer. In this capstone, you will take the role of a data scientist working for a new rocket company. Space Y that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk. 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.

**Data Collection Overview**

In this video, we will review Collecting the Data with an API. In this capstone assignment, we will be working with SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome. Our goal is to use this data to predict whether SpaceX will attempt to land a rocket or not. The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/. We have the different end points, for example: /capsules and /cores We will be working with the endpoint api.spacexdata.com/v4/launches/past. Let’s see how the API works. We will use this URL to target a specific endpoint of the API to get past launch data. We will perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API. This result can be viewed by calling the .json() method. Our response will be in the form of a JSON, specifically a list of JSON objects. Since we are using an API, you will notice in the lab that when we get a response it is in the form of a JSON. Specifically, we have a list of JSON objects which each represent a launch. To convert this JSON to a dataframe, we can use the json\_normalize function. This function will allow us to “normalize” the structured json data into a flat table. This is what your JSON will look like in a table form. Another popular data source for obtaining Falcon 9 Launch data is web scraping related Wiki pages. In this lesson, you will be using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records. Then you need to parse the data from those tables and convert them into a Pandas data frame for further visualization and analysis. We want to transform this raw data into a clean dataset which provides meaningful data on the situation we are trying to address: Wrangling Data using an API, Sampling Data, and Dealing with Nulls. You will notice, in some of the columns, like rocket, we have an identification number, not actual data. This means we will need to use the API again targeting another endpoint to gather specific data for each ID number. These functions are already created for you, and will use the following: Booster, Launchpad, payload, and core. The data will be stored in lists and will be used to create our dataset. Another issue we have is that the launch data we have includes data for the Falcon 1 booster whereas we only want falcon 9. In this lab, you will need to figure out how to filter/sample the data to remove Falcon 1 launches. Finally, not all gathered data is perfect. We may end up with data that contains NULL values. We must sometimes deal with these null values in order to make the dataset viable for analysis. In this case, we will deal with the NULL values inside the PayloadMass. In this lab, you must figure out a way to calculate the mean of the PayloadMass data and then replace the null values in PayloadMass with the mean. We will leave the column LandingPad with NULL values, as it is represented when a landing pad is not used. This will be dealt with using one hot encoding later on.

**Data Wrangling Overview**

In this video, we will review data wrangling. Let’s review some of the attributes: Flight Number, Date, Booster version, Payload mass Orbit, Launch Site, Outcome: this is the status of the first stage Flights, Grid Fins: these help with landing Reused, Legs: used in landing Landing pad, Block, Reused count, Serial, Longitude and latitude of launch Let’s take a look at some of these attributes. The column “LaunchSite” contains the different launch sites, including: Vandenberg AFB Space Launch Kennedy Space Center CCAFS SLC 40 The column orbits are the different orbits of the payload. For Example: LEO: Low Earth orbit (LEO)is an Earth-centered orbit with an altitude of 2,000 km GTO A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. It is located at 22,236 miles (35,786 kilometers) above Earth's equator. The column Outcome indicates if the first stage successfully landed. There are 8 of them, for example. True ASDS means the booster successfully landed to a drone ship as shown the following looped video. False ASDS means the mission outcome was unsuccessfully landed to a drone ship as shown in the video loop. Outcome We would like landing outcomes to be converted to Classes y. y. (either 0 or 1). 0 is a bad outcome, that is, the booster did not land. 1 is a good outcome, that is, the booster did land. The variable Y will represent the classification variable that represents the outcome of each launch.