Prompts from a data engineer ……

```
# Reading wine data from zipfile
# Show top 5 records
# Revising data types, amount of null values
# Dropping Nulls
# Stripping end whitespace
# Spliting data based on commas for the origin column
# Using the explode function from pandas in order to create different columns for each item in Origin_List
# define a function to split the list into columns
# apply the function to the DataFrame
# merge the new DataFrame with the original DataFrame
# Update the df3 dataframe
# Stripping whitespaces in the origin columns
# Checking the unique values of origin_1 column
# Defining a function to clean us territories
# Replacing values in the origin_1 column
# Apply cleaning function and renaming columns
# Verify if there are any repeated regions amongst Region and Origin 3
# Rename the columns accordingly to their content and dropping unneccessary ones
# Starting with changing the object columns to float64 datatype
# Checking missing values
# Creating a new feature called Price_Feature considering Price if available, if not use Price_Out-of-stock
# Checking the null values now that the Price_Feature was created
# Extracting the year of the wine from its name
# Replacing the numbers that do not correspond to a year
# Changing the data type to integer
# Stripping the year from the name column
# Check the number of nulls in this new column
# Replacing "COMMENTARY: " with an empty space
# Reorder the columns
# Getting the new numerical columns
# Check correlation between numerical columns
# Might need to 'bucket' values so we avoid it becoming a regression problem
# Getting all categorical columns
# Obtaining the top 10 varieties
# Creating a bar chart
# Obtaining the top 10 countries
# Creating a bar chart
# Average price for the top 10 countries
# Creating a bar chart
# Check the varieties most famuous in each top 10 country
# Dumping clean dataset into csv
# Dumping the clean dataset into a zip file
```