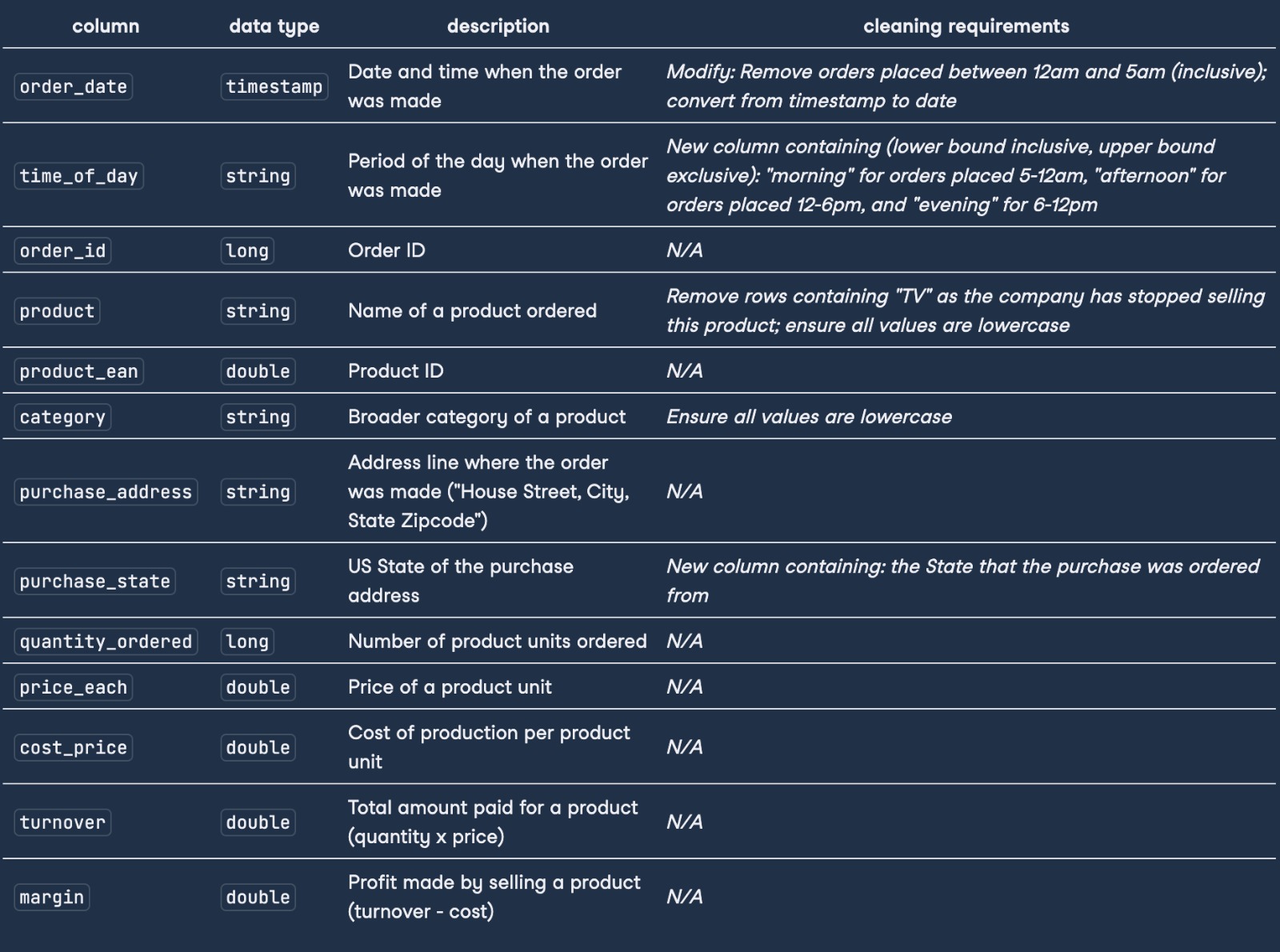
**Project Statement:**

As a Data Engineer at an electronics e-commerce company, Voltmart, you have been requested by a peer Machine Learning team to clean the data containing the information about orders made last year. They are planning to further use this cleaned data to build a demand forecasting model. To achieve this, they have shared their requirements regarding the desired output table format.

An analyst shared a csv file called "data.csv" for you to clean and preprocess.



In this picture you can see the existing schema of the file and there are new required columns you will be adding that are also mentioned here with all necessary details like data type etc.

- You are supposed to read the given csv file in PySpark.

- Then perform all necessary cleaning operation and process the data as mentioned clearly in above picture.

- Save the final result as a parquet file.  
  
  
Approach to this project:

1. Reading the data

2. Dealing with the order date column

3. Cleaning the product info columns

4. Extracting state from address and getting the number of states

5. Prepare and export the resulting table