**Interview Questions Related to above task:**

1. **What are missing values and how do you handle them?**

Missing values are the values of rows which are not available in the dataset.

We can replace/impute the missing values with measures of central tendency like mean, median, mode.

1. **How do you treat duplicate records?**

Generally we have to delete the duplicate records.

1. **Difference between dropna() and fillna() in Pandas?**

dropna(): It is a pandas method to remove missing values

fillna(): It is a pandas method to fill missing values

1. **What is outlier treatment and why is it important?**

Outliers are the data points which are deviated from normal distribution and can alter the measures of central tendency like mean, median, mode and may generate wrong results.

Outliers are skewed values of the distribution.

1. **Explain the process of standardizing data.**

Standardization is the process of making a every column dataset in a particular format like removing special symbols, space etcetera, so that a column can have same structure of data.

In the above example we removed

“ ₹  % and ',' symbol and converting the column to float”

1. **How do you handle inconsistent data formats (e.g., date/time)?**

It is better to convert them in date/time format using pandas datetime library.

Above example:

#4.Convert date formats to a consistent type (e.g., dd-mm-yyyy).

def convert\_date\_format(df,column):

    df[column] = pd.to\_datetime(df[column], format='%d-%m-%Y', errors='coerce') # convert to datetime format

    return df

1. **What are common data cleaning challenges?**
2. Finding null values using .isnull().sum() pandas method.
3. Replacing null values using .fillna() pandas method.
4. dropna(): It is a pandas method to remove missing values
5. removing outliers
6. handle inconsistent data formats (e.g., date/time)
7. standardizing data.

Etcetera..

1. **How can you check data quality?**
2. Finding null values using .isnull().sum() pandas method.
3. Replacing null values using .fillna() pandas method.
4. dropna(): It is a pandas method to remove missing values
5. removing outliers
6. handle inconsistent data formats (e.g., date/time)
7. standardizing data.
8. By data visualization

Etcetera.