DATA CLEANING STEPS

Tools used & steps taken:

The project followed a three phase methodology:

➢ Data cleaning & EDA(python)

➢ SQL Analysis

➢ Power bi dashboards

Data cleaning & EDA(python)

Data Cleaning and Preprocessing: This is often the most time-consuming step, involving handling inconsistencies and preparing data for analysis.

Handling Missing Values:

df.isnull().sum(): Count missing values per column.

df.dropna(): Remove rows/columns with missing values.

df.fillna(value): Fill missing values with a specific value (mean, median, mode, or a constant).

Handling Duplicate Values:

df.duplicated().sum(): Count duplicate rows.

df.drop_duplicates(): Remove duplicate rows.

Handling Outliers: (Often involves statistical methods or visualization to identify, then decide on removal or transformation).

➢ **SQL ANALYSIS**

Import all the cleaned csv files into sql by using bulk insert

Files imported into sql are products,customers,returns,stores,sales

Steps done in sql server:

Database : Create SQL database and load cleaned data

Index creation: Create relationships and indexes

Data verifications: check once the data successfully import in sql

Derived metrics: Write SQL to calculate derived metrics (profit, discount %)

10 business questions: Write SQL for business questions

➢ **Power bi dashboards**

Import cleaned csv files into power bi desktop and making dashboard based on the required business questions

Create dash board on

1. Sales Overview Report
2. Customer Insights Report
3. Product Performance Report
4. Store Analysis Report
5. Return Analysis Report

MOHAMMAD ASEEM

mohammadaseem197@gmail.com