

Project Report – Week 2

Overview

The aim of this project is to dive into the dataset containing information about e-commerce orders in Pakistan. There are various tasks in this project, and they involve calculating descriptive and inferential statistics using the datasets.

Steps

I started the project by loading the dataset into Jupyter Notebook to extract the first 1000 from the dataset. This is to make the work easier and also protect my mental health. After extracting using Python, I loaded the data in Excel and performed the necessary cleaning to check for missing values and format the dataset properly.

After cleaning the data, I calculated summary statistics for variables such as Price, Quantity, and Grand Total. I did this by inserting a pivot table. Using the pivot table, I calculated the sum, average, max, and min for each variable. This is to understand the distribution more. Next, I calculate the mean, median, standard deviation, and quartiles for relevant columns.

On the visualization sheet, I did various visualizations to show price, quantity order, and grand total. In addition, on another sheet, I identified the most popular payment method, and that's "COD." On the other hand, I also did visualization for the most common category, that's "Beauty and Grooming."

The next step should be to formulate hypotheses and test them out. So, I formulated a hypothesis that "there is a significant relationship between payment method and grand total." I tried to test this by calculating the standard deviation and average for the grand total; however, I got stuck trying to test it out.