Superstore Datastore Analysis

Exploratory Data Analysis and Visualization with Excel

Introduction

Aside every other project we have taken during our course of learning. This particular project is one among the project that will earn you a certificate of Data Science here at **Codar**. It is meant to check your knowledge on how analyze data and build an interactive dashboard using Excel

Overview of the Data

The Superstore Datastore is gotten from Kaggle, and it is a sales record detailing the categories, shipping records, products, sales, profits, and discounts. the items are categorized under Technology, office supplies and furniture with about 17 sub-categories with details of shipping to over 500 cities and various Regions such as Eastern Asia. the store serves three customer sets i.e., Consumers, corporate and home offices.

Data source: Global Super Store Data

Aim/Objectives

The aim of this project is to provide an in-depth analysis of this data and to answer all the following questions

Date of Submission

The project is meant to be publish on your **github account** with the necessary documentation on the README.md file.

Date of Presentation

Will be communicated.

Required Steps:

- Understanding the Data
- Data Cleaning
- Data Analysis
- Data Visualization

Data Cleaning

Before going ahead to perform any analysis on this data, you will need to clean up this data for any inconsistency or errors. Clean this data by using some of the Excel formulae such as Trim, Upper, lower, substitute (only in cases when needed) the features given by excel or using finding and replacing misspelled words, flash fill, deleting empty cells, finding and deleting duplicates, delete unwanted columns.

Data Analysis

Analysis of the superstore will be carried out based on the following.

1. Product Level Analysis
2. Customer Level Analysis
3. Region and Time Series Analysis

Product Level Analysis

- **What is the total number of products in the store?**
- **What is the number of categories and sub-categories?**
- **What is the number of products in each category and sub-categories?**
- **What categories of items sold have the most orders, generate the most profit and/or are the best-selling?**
- **What sub categories sell the best/least, generate least profit or losses?**
- **What shipping mode is the most used by customers?**
- **What is the Top 10 products by Cost?**

- **How many products from all categories is having 100% profits?**

To analyze further, we are going to calculate the cost and profit percentage variables

- **Calculating Cost = Sales – Profit**
- **Percentage Profit = (profit/cost * 100)**

Data Visualization

1. Show the distribution of Category by Sub-category using a Clustered Bar Chart
2. Show count of category by sub-category using a Clustered Bar Chart
3. Show Total profit and sales by sub-category using Column Chart
4. Show the distribution of the first 10 products in term of counts and revenue generated
5. Count of Categories with 100%.

Customer Level Analysis

- **What is the total number of customers?**
- **Which customer segment is the most profitable?**
- **Which ship mode is the most profitable?**
- **How the customers are distributed across the countries.**
- **Top 10 customers who ordered frequently from the store**
- **What is the Top 10 customers by profit and state?**
- **What is the Top 10 customers by sale and state?**

Region/Time Series Analysis

- **What is the number of products sold in each region?**
- **Which region is most profitable?**
- **Which city has the most sales?**
- **What countries generate the most profit?**
- **What year do we have the highest sales and profit**
- **Which region do we have the highest sales and profit**
- **Sales by Category for each year**
- **Profit and sales per year**
- **Profit and sales by year for each region**
- **Profit trend by category for each year**
- **Sales, profit and discount relationship**

To analyze further, we are going to calculate:

- **Shipment duration = ship_date – order_date**

Thank you for joining us at Codar