**Final Term Project week 12**

**Erwin Hammock**

**Bellevue University**

**06/01/2024**

**Outcomes of the EDA:**

* **Data Cleaning:** The data cleaning process involved addressing missing values (through imputation or removal), identifying and handling outliers, and ensuring data consistency across product names and categories.
* **Descriptive Statistics:** Central tendency measures like mean and median revealed average sales figures for products and departments. Standard deviation and range provided insights into the spread of sales data, highlighting which products had more consistent or variable sales patterns.
* **Visualizations:** Histograms and scatter plots were used to visualize the distribution of sales figures across product categories, days of the week, and time of day. This helped identify patterns such as higher sales for specific product categories on weekends or evenings.
* **Customer Segmentation:** By analyzing sales data alongside customer demographics (if available), we could potentially segment customers based on age, location. This could reveal variations in buying habits between different customer groups.

**Missed Aspects:**

* **Seasonality:** The analysis might not have captured seasonal trends if the data timeframe was limited. Analyzing sales data over a longer period could reveal how buying habits change throughout the year, influenced by holidays or seasonal products.
* **Marketing Promotions:** The impact of marketing campaigns or promotions might not be reflected in the sales data itself. Integrating data on marketing efforts with sales data can provide a more holistic view of how promotions influence customer behavior.

**Valuable Variables:**

* **Customer Demographics:** If available, data on customer demographics like age, location, and income can be incredibly valuable. This allows for customer segmentation and targeted marketing campaigns.
* **Product Features:** Information about product features (e.g., brand, size, organic) could be used to understand how these characteristics influence purchasing decisions.
* **Promotional Data:** Data on marketing campaigns, discounts, and promotions would allow us to assess their effectiveness and optimize future marketing strategies.

**Challenges Faced:**

* **Data Wrangling:** Cleaning and preparing messy or inconsistent sales data can be a significant challenge. Techniques like data imputation or outlier removal require careful consideration to avoid biasing the analysis.

**:**

* **Advanced Statistical Techniques:** While this EDA employed basic statistical methods and visualizations, more advanced techniques like time series analysis or machine learning algorithms could uncover deeper patterns and trends within the sales data.
* **Causal Relationships:** The analysis focused on identifying correlations between variables. However, establishing causal relationships between factors influencing sales requires more sophisticated statistical modeling.