**Tech Online Sales 2023 Analysis**

**Objective:**

The purpose of this project is to analyze potential trends in sales outcomes by examining detailed consumer behavior and spending patterns within a dataset from 2023. Our goal is to uncover insights into several key aspects of sales, including:​

* **Spending Trends:** Identifying whether consumers are actively spending and the overall direction of sales momentum.​
* **Categories of Interest:** Understanding which product categories are driving the most consumer engagement and revenue.​
* **Monetary Insights**: Analyzing the average and total amounts consumers are spending across different segments.​
* **Geographic Preferences**: Exploring regional or location-based differences in purchasing habits.​
* Brand and Product Popularity: Highlighting which specific products or brands are resonating most with consumers.​
* **Consumerism Patterns:** Examining broader consumer behavior to determine how economic and social factors might be influencing purchasing decisions.​

By diving into this data, we aim to provide actionable insights into what drives consumer purchases, the brands or products capturing the market’s attention, and how these trends might shape future sales strategies.​

**Content**

The dataset was scraped from publicly available results posted in Kaggle.

**Data Size**

Original data was over 302,010 lines. To be able to use it for the analysis we pre-processed it by excluding empty information for any transaction, then we filtered only USA transactions; this information was used for the initial analysis of total industry. Later we pre- processed for tech sales, in this case items such as AC, Fridges and TVs that were included in the electronics category were dropped. The final analysis was performed with 12,359 transactions. The information contained 30 columns that included descriptions of the customer, category, product, total sale, ratings, shipping mode and payment Method.

**Critical points that were answered in the analysis**

* Total industry composition and value
* Type of categories and products that were in high demand
* Highest consumption per type of consumer
* Correlations of different elements that might influence consumption
* Interactive HTML for mapping the information around the consumer insights
* And finally, the summary findings from the analysis

**Total industry composition and value**

* **Categories**: Total Sales 2023 $105MM largest categories are Groceries 27.3% and Electronics 21.2%.
* **Seasonality of sales**: highest months for total sales are July and March, while February is the lowest sales month. Grocery Sales are higher in January and lowest in September, while tech sales are higher in December and Back to School(Aug)
* **Purchase per transaction**: Purchases per transaction are very similar between categories the mean is around $260 per transaction.

**Type of categories and products that were in high demand in tech**

* **Mix and seasonality** Sales are clustered mainly in Smartphone and tablets​ representing 71.4% of the sales. Headphones and Laptop represent less than 30% of the industry, with in the largest categories Smartphones sales are higher in December and tablets on Back-to-School period ​
* **Products in high demand** Apple is the leading brand in the tech sales in value and transactions, Sony and Samsung with similar share​, but the gap is around 14pp.
* Apple's most relevant sales correspond to headphones, followed by smartphones and laptops being their smaller category. The higher sales are represented by this category which is unique to Apple.
* **Consumer**
  + The age group between 20-30​ is the largest purchaser of tech elements, followed by the groups of 30-40yrs and 40-50yrs. These three segments represent 81.4% of the sales.
  + 20-30 largest Product Type is Smartphones, followed by headphones
  + 40-50 largest Product Type is as well smartphone but the sales are ¼ of the sales of the 20-30 yrs consumer.
  + The highest total purchase in one transaction is done by the consumers between 40-50yrs. The group between 20-30 yrs has the largest dispersion. After 50 yrs purchases decline.
  + Women still under 50% of sales, could imply an opportunity given the increase in purchasing power of women. Women sales only represent 37% of the total, although their purchase per transaction is similar to the men’s. Consumption in terms of Brands is similar per gender, being the leading brand for purchases Apple. In terms of products Smartphones represent 42% of purchases of women and men, while headphones are 29%.
  + High and medium income consumers represent 73.7% of the purchases. However, high income purchase average is slightly below the medium income group, with high deviations. Definitely high income continues to be a target for sales of high value items​ and advertisement. While there might be an opportunity to develop affordable products for lower income to develop this segment.
  + Consumers value flexible offerings in shipping and payment as none is more relevant than the other it seems key to maintain these options available.

**Data Analysis and Correlations of different elements that might influence consumption** **:**

**Product Brand vs Gender:**

The heatmap highlights the distribution of product purchase by gender across three major brands; Apple, Samgsung, and Sony. Apple stands out with the highest overall count of purchase, dominated by males (2749), compared to females (1628). For Samsung, males (1859) also make up a larger share than females (1094), but the gap is smaller compared to Apple. The stacked bar chart reiterates the popularity of Apple in total purchases among both genders. While Samsung and Sony have smaller gender gaps.

**Total Purchases by Income Level and Gender**

The heatmap represents the total purchases across income levels (low, medium, high) in the Electronics product category. Males in the medium income level have the highest average purchase value. The second heatmap reflects a clear correlation between higher income levels and increased purchasing of electronics.

**Trends (Season & Holiday)**

The highest number of tech products were sold in the summer season. Indicating a peak in customer demand, possibly driven by back to school purchases. Winter records the lowest sales, possibly due to post holiday spending. Three holidays were used to track performance; Black Friday, Cyber Monday, and Christmas. The average daily sales for November and December were $47,394.99. Only Christmas exceeded this benchmark by a substantial margin.

**Correlation between Apple, Samsung, and Sony Purchases**

**Apple & Samsung:** A correlation of -0.51 indicates a negative relationship, suggesting that customers who purchase Apple products are less likely to purchase Samsung products.

**Apple & Sony:** A correlation of -0.52, suggests a similar trend. Customers buying Apple are less likely to purchase Sony.

**Samsung and Sony:** A weaker negative correlation of -0.38 suggests some overlap but still indicates a tendency for customers to favor one brand over the other.

**Customer Overlap Between Apple, Samsung, and Sony Purchases**

**Unique Customers:** Apple: 3991, Samsung: 2654, Sony: 2688

Shared Customers: Apple & Samsung: 131; Apple & Sony: 123; Samsung & Sony: 99; All three brands: 11

**CONCLUSIONS**

Apple emerges as the preferred choice among both genders, with male customers purchasing significantly more than females. In contrast, Samsung and Sony exhibit smaller gender gaps, showcasing a more balanced trend across genders. Overall, spending patterns are relatively balanced between genders, but income levels play a pivotal role in driving purchases. Higher-income individuals contribute significantly to total sales, underscoring their importance as a target demographic.

Seasonality also influences demand, with summer and spring accounting for the majority of sales. In comparison, fall and winter experience weaker sales, presenting an opportunity to introduce discounts or sales campaigns to boost revenue during these periods. Christmas stands out as the most lucrative holiday for tech product sales, likely driven by holiday promotions and last-minute gifting. Cyber Monday sales, however, fell below average, potentially reflecting less effective promotions or shifting consumer preferences. Black Friday slightly exceeded daily sales averages but failed to deliver a significant spike in demand.

The analysis of brand correlations reveals a competitive market characterized by strong brand loyalty. Customers primarily purchase from a single brand rather than diversifying across multiple brands. Apple dominates as the most favored brand, attracting the largest exclusive customer base. Samsung and Sony, with smaller exclusive customer pools, could benefit from targeted marketing campaigns aimed at converting customers from competing brands. Alternatively, they might explore strategic partnerships to tap into niche markets or enhance their competitive positioning.

**Limitations:**

**Limited Scope:** The dataset may not cover all relevant brands, product categories, or customer demographics. ​

**Sampling Bias:** The dataset might not represent the entire customer base, as it is limited to those who made specific purchases in specific product categories or from specific brands. Over representation or under representation of certain genders, income levels, or regions might lead to biased conclusions. ​

**Interpretation of Results:** Correlation results do not necessarily imply causation. External factors like marketing, promotions, or competitor actions could influence customer behavior. ​

**Market and Economic Conditions:** The dataset does not account for external influences such as economic downturns, product launches, or seasonal trends that could impact sales.

**Improvement for Analysis:**

To enhance the depth and accuracy of this analysis, several improvements can be made to the dataset and methodology. Expanding the dataset to include additional brands and product categories would provide a more comprehensive view of customer behavior.​

Incorporating time-related data, such as monthly or seasonal trends, would enable analysis of patterns like holiday sales spikes over time. ​

Capturing external factors such as promotions, economic conditions, or competitive actions would allow for more insight into causation rather than relying on correlation alone. ​

Including a broader geographic scope with more location data would help identify regional trends. ​

Incorporating additional data spanning multiple years would provide a more comprehensive view of trends and patterns. With multi-year data, we could explore an increase or decrease of sales for specific brands or product categories. Multi-year data could improve the accuracy to forecast future sales and to further understand the market.