**Data we used**

Our first dataset "Shopping Behavior" is from kaggle's Consumer Behavior and Shopping Habits Dataset, and this dataset provides the consumers' preferences, tendencies, and patterns during their shopping experiences. This dataset includes a diverse range of variables such as demographic information, preferences, purchase history, shopping frequency, and on/offline shopping behavior. We could know about the decision making process of consumers, and divide the purchase amount and frequency by age using this data.

Our second dataset "Ecommerce Consumer Behavior Analysis" is from kaggle's Ecommerce Consumer Behavior Analysis Data, and this dataset shows information on purchasing patterns, demographics, product preferences, customer satisfaction so that we can analyze the shopping behaviors and patterns by generation. Also, this dataset contains many variables such as income level, education level, gender, and living location, so we could check if there are some relationships that affect purchase amount and frequency.

**Methodology**

We used t-test, linear regression, and multiple linear regression to analyze the correlations among variables about the shopping behaviors and patterns by generation. The t-test is used to test which the mean difference between two groups is significant statistically. In our analysis, we could verify our hypothesis by comparing the mean difference between online shopping amount or frequency of middle aged people and young generations.

The linear regression is used for figuring out the impacts that one variable has on another variable. This can estimate the direction (positive or negative) and the degree of the effect of age that is a continuous variable on online shopping behavior. So we used this method to identify whether shopping amounts decrease or increase as age increases with a linear model.

However, the result of simple linear regression shows that there was no significant relationship between age and purchase amount. To deal with this problem, we used multiple linear regression which can broaden with another hypothesis that other psychological and behavioral variables

**Tools**

    **Python libraries**
- pandas: We used these libraries to deal with the data efficiently and convert to dataframe.
- matplotlib, seaborn: We used these libraries to visualize the results including linear regression, p-value, t-test, R squared, etc.

    **Visualization**
- Tableau: We used tableau as a visualization tool. We made graphs using this tool too.

- Streamlit: We used streamlit to make a data story webpage and distribute it to the public.

**Collaboration Environment**
- Google Colab: This was used to share our codes and results.
- Google Docs: This was used for documentation.
- Canva: This was used to create a visual poster for presentation.