**Research Question**

Assuming I’m a marketing analyst at a company that sells vacuums, I’m tasked with improving customer engagement. In the customer marketing field, “Customer Buying Journey" is an important concept. The two most important phases in the journey are moving from “Consideration” to “Purchase.” Understanding the influence of different sources of information on these two stages is essential for a marketing analyst. As a result, my research question is “**Which sources of information have the greatest influence on decision-making during the model consideration stage and the final model purchase stage?”**

**Data source**

The dataset for this idea comes from a website called “Texas Data Repository.” This is the link to the dataset.

The survey data contains 251 variables with responses from 1002 respondents. In the survey design, all questions are mandatory. Therefore no missing values exist, except for instances where respondents chose to respond with “prefer not to say” (in some sensitive demographic questions) or “I don’t know” (in some questions related to their social networks).

Although there are a lot of components, for the sake of the research question, I only use the raw survey data. However, I will not use all 251 variables. Below are the main variables that I am going to use: ResponseID, Gender, Age, Ethnicity, MaritalStatus, Education, Occupation, Income, TV\_Consideration, Billboard\_Consideration, OnlineAds\_Consideration, Reviews\_Consideration, SocialMedia\_Consideration, WordOfMouth\_Consideration, Salespeople\_Consideration, TV\_Purchase, Billboard\_Purchase, OnlineAds\_Purchase, Reviews\_Purchase, SocialMedia\_Purchase, WordOfMouth\_Purchase, Salespeople\_Purchase

**Data Analysis**

For the research question, I will merge the rating of each source of information of two stages into one column, for example, “The influence of TV” is a new column that contains ratings for both the consideration and purchase stages. Additionally, I will create a column called “Buying stage” that indicates whether the ratings are for consideration or purchase stage.

EDA: Understanding the influence of different sources of information for different types of demographics.Some visualizations like stacked bar charts would work great to present these findings.

Modelling:

1. The main model that I will use for this idea is the “Multivariate Analysis of Variance”. I will assess whether the influence of each source differs noticeably between the stages of consideration and purchase keeping control of the demographic elements.

Dependent variables:

* each respondent’s influence rating for all sources of information across both stages, which include the influence of TV, billboard, online ads, internet reports/reviews, social media/internet blogs/forums, family/friend/word of mouth, and salespeople
* all the demographic elements: gender, age, ethnicity, marital status, Education level, Occupation, Income range

Independent variable: column “stage” that indicates whether it is a consideration or purchase stage.

1. Paired t-test: Utilize paired t-tests (for each source) to assess whether there is a substantial distinction in the influence of the two stages.

**Application and Ethical Implication**

Understanding the influence of the shift from consideration to purchase is important. Firstly, when knowing what sources of information are the most influential, the social media team can develop several marketing strategies with these specific sources for people in the consideration stage. These strategies can then increase the conversion rate. Moreover, this research question can save the company a lot of money from investing in ineffective sources of information.

This data was collected from individuals. In the survey, there was a section called “Consent to Participate in a Research Study.” This section clearly explained the purpose of the research, and how the data was stored and protected.

The study could indicate how independent reports, social media, and internet reviews influence customer choices. These sources run the danger of being exploited by businesses to influence customers (e.g., by biased reporting or phony reviews). As a result, a marketing analyst needs to advocate for transparent marketing and advertising.