

How does social media influencer marketing affect consumer buying behavior?

A. Problem Definition & Objectives

One of the strongest marketing platforms can now be the social media. Influencers on Instagram, Tik Tok, YouTube and Facebook are used by brands to advertise their products. Shoppers, and particularly youths, are more likely to believe the suggestions by their influencers than the conventional advertising. Nevertheless, it is not yet clear how influencer marketing affects consumer buying behavior.

Objectives:

- To analyze whether social media influencers affect consumers' purchasing decisions.
- To identify which factors influence trust in influencers (credibility, engagement, followers, content quality).
- To compare general social media usage habits with actual buying decisions.
- To study differences in buying behavior based on age, gender, and platform (Instagram, TikTok, YouTube).
- To create a clean and structured dataset for further marketing analysis.

B. Data Sources

Dataset	Source & Link	File Type	Approx. Size	Key Variables
1. Influencer Marketing ROI Dataset	Kaggle (https://www.kaggle.com/datasets/tfisthis/influencer-marketing-roi-dataset)	CSV	400 data collections	Influencer Name/ID, Platform, Followers, Likes, Engagement Rate, Campaign Cost, Revenue, Product Type
2. Impact of Social Media Dataset	Kaggle (https://www.kaggle.com/datasets/lastman0800/impact-of-social-media-dataset)	CSV	300 survey responses	Age, Gender, Hours on Social Media, Number of Platforms, Purchased Because of Social Media (Yes/No), Type of Products Purchased

C. Data Management Plan

In Assignment 2, I will clean and prepare the selected datasets before analysis. The main steps are:

1. Handling Missing Values

- I will recognize the missing values. Median values will be used to fill in numerical gaps, whereas brand or platform and other categorical gaps will be filled in using the mode or the value of Unknown so as not to lose any data.

2. Detecting and Handling Outliers

- The IQR (Interquartile Range) method will be used to identify outliers in numerical columns (e.g. income, amount spent, followers). Rather than removing them, I will be capping extremes to keep things within realistic ranges.

3. Data Transformation and Standardization

- I will also change irregular text formats (e.g. Instagram, instagram) to lower case, Yes/No to 1/0 and make sure that all the types of numerical data are appropriate. New useful variables (e.g., ROI, purchase rate) can also be developed.

4. Merging and Dealing with Heterogeneous Data

- As the two datasets are dissimilar sources, the column names and formats will be adjusted. The comparison of influencer impact and consumer behavior will be made by them being logically combined based on a shared feature like social media platform

5. Maintaining Data Integrity and Consistency

- The records will be eliminated and clean data will be stored in different files as opposed to raw data. I will repeat the check of the consistency, absent values, and appropriate data type after the merging.

D. Expected Output

The end product of Assignment 2 will be a clean, well-organized and integrated data, that will be a combination of insights in the Influencer Marketing ROI dataset and the Social Media Consumer Behavior one.

The resulting cleaned and combined data will enable analysis of the following nature.

- ✓ Platform Comparison: Determine what platforms (Instagram, Tik Tok, YouTube) influence consumers to make purchases the most.
- ✓ Influencer Impact Analysis: Test to determine the relationship between increased influencer expenditure/salaries and increased consumer buying/use.
- ✓ Consumer Behaviour Insights: Understand how age, income, and time spent on social media affect buying behavior under influencer marketing.
- ✓ Data Visualization and Modeling: The resultant data set may be employed in developing graphs, pivot tables, statistical analysis of correlation, and perhaps predictive models to predict the pattern of buying.

A **structured and analysis-ready dataset** will be produced which is suitable for business decision-making, academic analysis, and answering the research question:

"How does social media influencer marketing affect consumer buying behavior?"