

Overview:

The goal of this assignment is to give candidates an opportunity to showcase their approach to a few typical situations and display the data analysis skills that will be necessary to be successful in this role.

You are required to submit your response in an email with relevant attachments **within 2 days of receiving** this assignment.

Feel free to ask us questions by writing to data@showtimeconsulting.in something is not clear.

Thank you. We appreciate your time and look forward to reading your responses!

Introduction:

To complete your application, you are required to submit an assignment as described below.

Responses to assignment should be submitted in one single PDF/PPT document.

- For assignment, you must attach a well-documented code file, any intermediate datasets generated, and any external data sources directly used in the analysis. For components that need narrative answers, tables, charts, etc., organize your responses in a Word file with clearly labelled questions and responses matching the numbering in this document.
 - You may choose between R or Python for this part, and preferably complete all tasks on the same platform. Solutions using only Microsoft Excel / Google Sheets will not be accepted.
 - There are often many ways to perform a given task so try to use relatively time-efficient and neat ways - ideally exploiting the strengths of the platform you have chosen.
 - Irrespective of the technical sophistication of your method, ensure you explain your methodology, results, and interpretation in simple language.
 - If you get stuck in one question and you can move on to the next one, go ahead. We encourage you to solve as many questions as you can and for incomplete and partial answers, please try to provide a narrative answer about your planned approach, constraints faced, data/information that could enable you to implement that.

You will be evaluated on the quality of your code and approach and not purely on the accuracy of results.

- Throughout, all sources used must be cited clearly, following any popular citation style. Plagiarism in text and data analysis - including from your current or past professional work - will lead to immediate disqualification. Follow the instructions for each part carefully.

Assignment: Exploratory Data Analysis on Facebook Utilization Data

Data Source link:

<https://docs.google.com/spreadsheets/d/1PHNL0qC1qCvVsN9vZMKEPo1cg5CmNB4JmYoCQcfVR3g/edit?usp=sharing>

Problem Statement:

The objective of the proposed framework is to study and analyse the differences in the way users are using Facebook based on their gender, age-group, etc. and Identify a pattern out of it.

Data Description:

No.	Column Name	Description
01	Userid	A numeric value uniquely identifying the user.
02	Age	Age of the user in years.
03	dob_day	Day part of the user's date of birth.
04	dob_year	Year part of the user's date of birth.
05	dob_month	Month part of the user's date of birth.
06	gender	Gender of the user.
07	tenure	Number of days since the user has been on FB.
08	friend_count	Number of friends the user has.
09	friendships_initiated	Number of friendships initiated by the user.

10	likes	Total number of posts liked by the user.
11	likes_received	Total Number of likes received by user's posts.
12	mobile_likes	Number of posts liked by the user through mobile app.
13	mobile_likes_received	Number of likes received by user through mobile app.
14	www_likes	Number of posts liked by the user through web.
15	www_likes_received	Number of likes received by user through web.

Data Exercise:

Using the provided dataset, please answer the questions below. Present your responses effectively using text, tables, and charts. While no word limit has been specified, you will be evaluated on your ability to highlight key insights concisely while responding to this part of the assignment. Feel free to support your responses with any relevant literature, data sources, and analysis of any other datasets. Remember to attach all working files relevant to your data analysis (well-documented code file, any intermediate datasets generated, any external data sources directly used in the analysis, etc.) to your response.

1) Load the data and impute missing values

Imputation of missing values:

- Replace the null values (NA) of gender column with its mode or median and explain why mode/median used to replace NA values
- Replace the null values (NA) of tenure column (numerical variable) with its median, and explain why mode/median used to replace NA values

2) Plot heatmap / correlation matrix on all the columns.**3) Analysis based on gender of the users**

- What is composition of male and female users?
- Which category of gender has more friends?
- Which category of gender initiated more friendships?
- What is the distribution of tenure across different categories of gender?

4) Analysis based on the least active users on Facebook

- How many users have no friends?

- How many users did not like any posts?
- How many users did not receive any likes?

5) Analysis based on the user accessibility (Mobile Devices vs. Web Devices)

- What is the average number of posts liked by users (based on gender) through web vs. mobile devices?
- What is the average number of likes received by users (based on gender) through web vs. mobile devices?