

Chapter 4 – Data Collection Plan

Instructions

1. Answer the below question in the boxes provided.
2. Please submit the assignment through TalentLabs Learning System.

Context

Imagine you are working in a luxury clothing company as a data analyst. As fast fashion is becoming more and more popular, your manager requested you to analyze the clothing purchasing habits of young adults. With the data insights, the company can then decide the right strategy to tackle the challenge from the fast fashion industry.

Before collecting any data in your city for this project, you need to make a data collection plan. Your task is to **write up the data collection plan** on how to carry out data collection to solve this problem.

We have already provided a framework (next page) for you to structure the plan. Although we provided a framework for you, please remember that you are creating a plan, not answering questions. Your plan should look like a proposal instead of exam paper.

Make use of the notes from the lectures to build your plan. Pay particular attention to privacy, bias and sampling. Explain the consideration you should make in this project regarding each of those things, and how each issue will be approached and avoided. (15 marks)

Data Collection Plan

Problem Statement (1 point)

Our goal is to analyze clothing purchasing habits of young adults and investigate what are the clothing features they prioritize when making purchases.

Data the needs to be collected & Scope of data (2 points)

Column	Data type	Description
Age	Integer	The age of the respondents, mainly includes the young adults between 18 – 40 years old
Gender	String – categorical	The gender identity of the respondents, whether it's male or female
Monthly Income Level	Float	Monthly income of the respondents
Monthly Expenditure	Float	Amount of money the respondents spend each month, which acts as a proxy to the respondent's spending patterns
Budget allocation on clothing	Float	Budget on clothing, expressed in % of monthly expenditure
Preferred Brands	String – categorical	The brand loyalty and style preferences of the respondents
Factors affecting recent purchases	String - categorical	The features and attributes the young adult consider when buying clothes (e.g., style trends, comfort, quality, pricing)

Timescale (1 point)

We propose to take 4 months to collect the data needed, and then another 1 month to clean and enrich our data.

Execution - How and Who will collection the data (2 points)

We will use a few ways to collect the data together with our staff:

- (i) The survey was conducted in our stores, where our sales staff can chat with customers to gather information, either by talking or through written responses. Yet, written response is preferred.
- (ii) The survey was done at the mall, where our team can approach shoppers in person for survey.
- (iii) Online survey via SurveyMonkeys to reach a broader and more diverse audience.

Data Storage Security (1 points)

To enhance data storage security, only managers or data analysts will have access to the complete data, to minimize the risk of data leakage issue.

Privacy Considerations and Remediation (2 points)

We have gathered information on respondents' income and expenses, which can be sensitive. To ensure privacy, we won't ask for respondents' names in the survey, instead we

may ask for their email address. It is to make sure the personally identifiable information is anonymous.

Bias Considerations and Remediation (2 points)

In this data analysis project, we need to be aware of potential confirmation bias. This means we should avoid seeking evidence to prove our initial assumptions right. For example, if we assume most young adults primarily prefer to follow the latest fashion trends, we should open the possibility that this isn't true.

Sampling Challenges and Remediation (2 points)

We are also likely to encounter a **sampling bias**. A survey of urban citizens analyzed clothing behaviours of young adults - doesn't account for those who live in rural areas.

What we do is that we try to collect the data outside the city areas via online surveys like Survey Monkey, etc to make sure the sample we have is more diverse to improve quality and accuracy of our data analysis project.

Limitations (1 point)

We would like to acknowledge some potential challenges in our data collection process. Verbal surveys or interviews can be challenging to record accurately, as people may tend to forget details and find it hard to summarize extensive information.

Back-up Plan (1 point)

We would like to propose to use some of these cloud backup plans such as SharePoint, and Google Workspace Drive. If possible, do some local backup as well.