</talentlabs>

Chapter 3 - Data Collection Mini-project

Instructions

1. Please submit the assignment through TalentLabs Learning System.

In this exercise you will work through the data collection process. An example solution is discussed in the final lesson of chapter 2. By the end of the exercise you will submit a zip file that will contain three files – the data file, the documentation file and **this word document** with your answers to previous questions.

Your challenge is to pick a topic of your choosing (something that interests you) and then **ask a question** that can be answered with data. Next, you need to state:
- The **aim** of your project
- The **type of data** that you require
- Whether primary or secondary data collection is better for the task and **why**
- With what **method** will the data be collected and **why**

All this information should be documented into the documentation file.

Next you should collect the data. Once the data is collected create a folder and add the data to it. In the same folder add this word document with your answers to previous questions in Part 1. To the folder, also add the documentation file and document your project within that file. When documenting, state the aim, type of data, method of data collection etc. Also explain the reasons behind your decisions and include anything else you think should be documented.

Finally, create a zip version of your folder that has the three files and submit the zip file to Talent Labs.

(9 marks)

visit out page:
www.talentlabs.org

contact us:
learn@talentlabs.org

19, Jalan USJ Heights 1/1B,
USJ Avenue, Subang Jaya, MY

</talentlabs>

Part 1 – Short-questions

Question 1.1:
What are the key steps of the data collection workflow? (2 marks)

1. Define the problem by asking questions
2. Making a plan for data collection
3. Collect the data
4. Document the process of collecting the data

Question 1.2:
Imagine you work as a data analyst for a clothing brand. You work in a team on a project trying to forecast future sales. From the list below, select if the person is a **stakeholder** in your project. Select **all** that apply. (1 mark)

- ☑ Your manager
- ☐ A customer of your brand
- ☑ The CEO of the company you work for
- ☑ A machine learning engineer building the forecasting model

Question 1.3:
If you are unsure if you have asked a good analytics question, what methodology could you refer to? (1 mark)

SMART – Specific, Measurable, Action-focused, Relevant, Time-bound

Question 1.4:
Discuss the advantages and disadvantages of carrying out your own data collection. (2 marks)

The advantages of carrying out your own data collection is the quality of data is good, trustworthy and has a better understanding of the data. However, the data collection will be time consuming and expensive.

Question 1.5:
Provide **four** examples of how data can be collected. (2 marks)

1. Web scraping
2.Automated
3.Questionnaires
4. Sensors

Question 1.6:

Give **one** reason why you should document your data collection? (1 mark)

It will provide context to the user who wants to use the data

.