Simulation exercise: Data Anonymization in Applied Research in VET

Simulation Exercise Title: Data Anonymization in Applied Research in VET

Objective: The objective is for learners to become familiar with anonymization techniques and methods by applying them in a real-life scenario. This simulation activity supports the overall aim of the module, which is to introduce ethics in research and provide practical techniques that researchers can use to ensure ethical conduct in their work.

Background: A local Institute in Limassol, Cyprus, in collaboration with a government employment agency, conducts a research study on the impact of training programs on the employement status of unemployed individuals, particular those who different type of health issues (chronic illnesses, non-chronic illnesses, mental health issues etc.). The research aims to assess the effectiveness of specific VET programs in improving employability for this vulnerable groups. The data gathered includes personal details, employment history, health conditions and training outcomes.

Data Provided: The dataset includes information from 10 participants. The following type of information is provided for each participant:

- ID Number
- Name (Full name of each participant)
- Age
- Gender
- Employment Status

- **Health Condition (Chronic/Non chronic condition/Mental Issues/None)**
- **Type of VET program they attended**
- **Location**
- **Post-training salary**
- **Duration of Unemployment**

ID	Name	Age	Gender	Employment Status	Health Condition	Training Program	Location	Salary (Post-training)	Duration of Unemployment
AI909091	Giorgos Ioannou	45	Male	Employed	Chronic	Cooking	Limassol	1.200	12
KI903236	Vasilis Papaiordanous	52	Male	Unemployed	Non-chronic	Coding	Nicosia	0	24
PO404042	Nikolaos Tzovenakis	46	Male	Employed	Mental Issues	Communication Practices	Nicosia	1.500	18
LI909789	Maria Onisiforou	38	Female	Employed	Mental Issues	Potting	Nicosia	1.100	8
QW343457	Dimitra Papagianni	37	Female	Unemployed	Chronic	Communication Practices	Larnaca	0	36
RR090201	Kypros Demosthenous	54	Male	Employed	Non-chronic	Coding	Paphos	1.550	6
YI245678	Elias Papatimotheou	35	Male	Employed	Non-chronic	Coding	Paphos	1.480	10
TO457906	Marios Polykarpou	39	Male	Unemployed	Mental-issues	Electrical	Paphos	0	20
GV345190	Iliana Vassou	41	Female	Employed	Non-chronic	Electrical	Nicosia	1.600	16

Task: A table including personal/sensitive information of 10 people is provided. Learners must apply various anonymization techniques, such as data-masking, data pseudonymization, and data generalization, in a real-

world VET context, emphasizing the importance of data protection while ensuring that data remain useful for analysis.

Guidelines:

Complete the table below, utilizing the following anonymization techniques. Ensure that data remain useful for analysis, while protecting the anonymity of the participants.

1. **Data Masking -** Apply data masking to anonymize the names and ID numbers in the dataset. For example, replace real names with pseudonyms like "Participant_001".
2. **Data Pseudonymization -** Replace the names with pseudonyms to ensure that the identity of the individuals cannot be directly tracked back to the dataset.
3. **Data Generalization -** Generalize the ages into broader categories (i.e. instead of specific ages, use categories like "25-30", "31-40", etc.). Similarly, aggregate locations into larger regions.
4. **Data Swapping:** Swap the health conditions and training programs between participants to further anonymize the dataset.

Deliverables:

- **Learners are required to complete the new table below, utilizing the anonymization techniques described earlier.**
- **Learners will be evaluated based on the accuracy and effectiveness if the anonymization techniques used.**
- **Learners must ensure that the data maintain their utility for analysis while privacy of the individual is being protected.**

ID	Name	Age	Gender	Employment Status	Health Condition	Training Program	Location	Salary (post-training)	Duration of Unemployment