Student Name: Shanikwa Haynes

Student ID:

Date: 03/13/2025

LFN1 Task 1: The Data Analytics Life Cycle

A: Phases of the Data Analytics Life Cycle

The Data Analytics Life Cycle consists of seven phases that guide analysts in solving business problems using data (Chapman et al., 2000). Below, I describe each phase and reflect on my expertise in each.

Business Understanding

Summary: This phase involves defining the business problem, identifying objectives, and determining the analysis scope. Stakeholders set clear goals to ensure data-driven decision-making (Chapman et al., 2000).

My Expertise: I have experience analyzing policies in criminal justice, which required defining objectives and identifying relevant data points. However, I have limited experience in defining business problems in a corporate setting.

Data Acquisition

Summary: This phase involves gathering data from various sources, such as databases, APIs, or direct data collection (Han et al., 2011).

My Expertise: I have experience retrieving data from structured reports, which translates to understanding how data sources can be leveraged for analysis. However, I have not yet worked with APIs or automated data collection.

Data Cleaning

Summary: Raw data often contains inconsistencies, missing values, or irrelevant information. This phase ensures the data is prepared for analysis (Rahm & Do, 2000).

My Expertise: I have experience cleaning manually recorded data and ensuring accuracy in reports, but I need more experience with automated data cleaning techniques such as Python's Pandas library.

Data Exploration

Summary: Analysts explore the dataset to identify trends, patterns, and anomalies, often using visualizations (Few, 2009).

My Expertise: I have created visual reports and graphs in Excel to analyze trends, but I need to strengthen my programming skills in data visualization tools like Python's Matplotlib or Tableau.

Data Modeling

Summary: Analysts develop predictive models using statistical and machine learning techniques (James et al., 2013).

My Expertise: While I have a basic understanding of regression models from statistics courses, I need hands-on experience with advanced predictive modeling.

Data Mining/Machine Learning

Summary: This phase involves applying algorithms to extract insights from data. Machine learning techniques may be used for predictive analysis (Bishop, 2006).

My Expertise: I have no direct experience with machine learning but have studied its applications. I aim to gain expertise through hands-on projects.

Reporting and Visualization

Summary: The final step is presenting insights to stakeholders through dashboards, reports, or presentations (Yau, 2013).

My Expertise: I have experience summarizing findings in structured reports, which is essential for effective communication of data insights.

A1: Proposal to Gain Expertise

To develop expertise in each phase, I plan to engage in the following activities:

1. **Business Understanding:** Taking *Google Data Analytics Professional Certificate* (Google, 2021) and participating in business analytics strategy workshops.
2. **Data Acquisition:** Practicing SQL queries and API data extraction from sources like *Kaggle* and *Google BigQuery* (Han et al., 2011).
3. **Data Cleaning:** Working with unstructured datasets in Python, using *Pandas* to remove missing values, correct inconsistencies, and prepare data (Rahm & Do, 2000).
4. **Data Exploration:** Using *Matplotlib* and *Power BI* to visualize trends and uncover insights (Few, 2009).
5. **Data Modeling:** Implementing regression and clustering techniques in *Scikit-learn* (James et al., 2013).
6. **Data Mining/Machine Learning:** Training a machine learning model on a classification task using *TensorFlow* (Bishop, 2006).
7. **Reporting and Visualization:** Building interactive dashboards in *Tableau* and *Power BI* (Yau, 2013).

A2: How the Goal and Mission Help the Analyst

1. **How an Organization's Goal Helps Identify Business Requirements:**

- A company's goal provides a clear direction for data analysis. For example, if a retailer aims to increase revenue by 20%, an analyst might focus on customer purchasing behavior and product profitability (Davenport & Harris, 2007).

2. **How an Organization's Mission Helps Identify Business Requirements:**

- A healthcare company aiming to improve patient outcomes would guide analysts to focus on treatment effectiveness rather than just financial data.

B: Applying SQL for Data Preparation

A retail company could use SQL to filter customer purchase data, allowing targeted marketing campaigns.

```
SELECT customer_id, COUNT(order_id) AS total_orders

FROM sales_data

GROUP BY customer_id

HAVING total_orders > 5;
```

B1: Three Risks of Using SQL

- **Data Security Issues:** Poorly secured databases can expose sensitive customer information (Rahm & Do, 2000).
- **Performance Bottlenecks:** Inefficient queries can slow database operations (Han et al., 2011).

- **Data Integrity Risks:** Incorrect joins or filtering can lead to misleading results (James et al., 2013).

B2: Organizational Problem

A retail company struggles with duplicate transactions and missing customer information. SQL is used to clean and structure the data before analysis.

C: Selecting the Appropriate Tool

SQL is the most efficient tool for handling structured databases, ensuring that data is properly cleaned and filtered before analysis (Han et al., 2011).

C2: Findings from SQL Usage

After applying SQL:

- **5% of duplicate records** were removed.
- **Transactions were categorized** based on customer purchase history.
- **Data was structured** for customer segmentation analysis.

C3: Ethical Risks of Using SQL

- **Data Privacy Violation:** If personal details aren't anonymized, it could lead to a data breach (Rahm & Do, 2000).
- **Bias in Data Filtering:** If incomplete data is used, biased insights may result (Few, 2009).
- **Unauthorized Data Access:** Weak SQL security settings could expose confidential business data (Davenport & Harris, 2007).

D: Sources

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM step by step data mining guide.

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business Press.

Few, S. (2009). Now you see it: Simple visualization techniques for quantitative analysis. Analytics Press.

Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques. Morgan Kaufmann.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin, 23(4), 3-13.

Yau, N. (2013). Data visualization and presentation with Microsoft Office. John Wiley & Sons.