Phase 1: Problem Definition and Design Thinking

Problem Definition

The project at hand involves harnessing the power of big data analysis using IBM Cloud Databases. The primary objective is to extract valuable insights from extensive datasets that encompass a wide range of domains, from climate trends to social patterns. This project encompasses various stages, including designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and finally, presenting the results in a manner conducive to business intelligence. Let's delve deeper into each aspect of the problem and outline the steps we will take to address them.

Understanding the Problem

Data Selection

The first step is to identify the datasets to be analyzed. The datasets can vary significantly in terms of size, format, and domain. Some potential datasets might include climate data, social media trends, financial data, or any other dataset that holds relevance for the problem at hand. It is essential to understand the nature of these datasets, their sources, and their potential impact on the project's objectives.

Database Setup

To effectively manage and analyze large datasets, we need to set up IBM Cloud Databases. This includes selecting the appropriate database type (e.g., relational, NoSQL), configuring database parameters, and ensuring data security and privacy compliance. The database setup phase is crucial as it forms the foundation for data storage and retrieval throughout the project.

Data Exploration

Once the datasets are stored in IBM Cloud Databases, the next step is to develop queries and scripts to explore these datasets. Data exploration involves understanding the structure of the data, identifying relevant variables, and formulating questions that can guide the analysis. It is essential to preprocess and clean the data to ensure its quality and suitability for analysis.

Analysis Techniques

The heart of this project lies in applying appropriate analysis techniques to uncover valuable insights. Depending on the nature of the datasets and the questions we want to answer, this may involve statistical analysis, machine learning, or a combination of both. It is crucial to choose the right tools and algorithms, perform feature engineering when necessary,

and validate the results rigorously.

Visualization

Data analysis is most effective when the results are presented in a visually understandable and impactful manner. Designing effective visualizations is a critical part of this project. Visualizations help in conveying complex patterns and trends to stakeholders and decision-makers. We must select the right visualization techniques and tools that suit the nature of our findings.

Business Insights

Ultimately, the goal of this project is to derive valuable business intelligence and actionable recommendations from the analysis. It's not enough to uncover patterns; we must interpret these findings in the context of the problem statement and the specific goals of the project. These insights will guide decision-making and strategic planning for the organization.

Design Thinking

Data Selection

- Collaborate with stakeholders to identify relevant datasets.
- Assess the quality, size, and availability of selected datasets.
- Document the data sources and any data preprocessing required.

Database Setup

- Choose the appropriate IBM Cloud Database service (e.g., Db2, Cloudant).
- Configure database settings, including security measures.
- Develop a data storage and retrieval strategy.

Data Exploration

- Explore data using SQL queries, Python scripts, or other tools.
- Perform data cleaning, handling missing values, and outlier detection.
- Document the data exploration process and findings.

Analysis Techniques

- Select suitable analysis techniques based on the nature of the data.
- Develop scripts or code for analysis, including any machine learning models.
- Implement validation and testing procedures.

Visualization

- Choose visualization tools such as Tableau, Matplotlib, or IBM Cognos Analytics.
- Design interactive and informative visualizations.
- Create a visualization plan to showcase key insights.

Business Insights

- Interpret analysis results in the context of the problem statement.
- Formulate actionable recommendations for stakeholders.
- Prepare a comprehensive report summarizing insights and recommendations.

Assignment Notebook Submission

File Naming Convention: CAD_Phase1

After completion, upload the project document to your private GitHub repository. Ensure that faculty evaluators from your college and industry evaluator (IndustryEvaluator@skillup.online) have access to your private GitHub repository for the evaluation process.