PROJECT OVERVIEW

Project Title: BIG DATA ANALYSIS

Domain : Cloud Application Development –Group 4

Assignment : PROJECT SUBMISSION PHASE 1

SUBMITTED BY

Name : K.JUBAITHA BEEVI

Mail id : jubaibawajon2003@gmail.com

COLLEGE NAME: P. R Engineering College

College code : 8212

Group 4 : Zone (13-16)

Problem Statement: Dive into the world of big data analysis with IBM Cloud Databases. Uncover hidden insights from vast datasets, from climate trends to social patterns. Visualize your findings and derive valuable business intelligence. Embark on data-driven adventure, exploring the endless possibilities of big data!

Problem Definition: The project involves delving into big data analysis using IBM Cloud Databases. The objective is to extract valuable insights from extensive datasets, ranging from climate trends to social patterns. The project includes designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and visualizing the results for business intelligence.

Design Thinking:

1.Data Selection

Objective: Identify the datasets to be analyzed, such as climate data or social media trends.

Analyzing big data often involves working with large and diverse datasets. Here are some examples of data sets.

1.Climate data:

- NASA ‘s Global Climate Change Data
- NOAA Climate Data
- European Climate Data

2.Social Media Trends:

- Twitter API Data
- Facebook Graph API Data
- Instagram API Data

Deliverable: The datasets of Climate Data or Social media trends are downloaded and analyzed.

2.Database Setup

Objective: Set up IBM Cloud Databases for solving and managing large datasets.

IBM offers several cloud database options suitable for storing and managing large datasets.

1. IBM Db2 on Cloud
2. IBM Db2 Warehouse on Cloud
3. IBM Cloudant
4. IBM Cloud Object Storage
5. IBM TimeSeries Database

Deliverable: These datasets can handle large datasets and offer various features like scalability, reliability, security, and ease of management, making them suitable for big data applications.

3.Data Exploration

Objective: Develop queries and scripts to explore the datasets, extract relevant information, and identify patterns.

- Understanding the Data
- SQL Queries for Data Exploration
- Identifying Relevant Information
- Data Extraction
- Pattern Identification
- Scripting Tools For Automation
- Statistical Analysis and Machine Learning

Deliverable: In this phase you gain insights and better understand the dataset.

4.Analysis Techniques

Objective: Apply appropriate analysis techniques, such as statistical analysis or machine learning, to uncover insights.

1.Statistical Analysis:

- Descriptive Statistics
- Correlation Analysis
- Hypothesis Testing
- ANOVA (Analysis of Variance)

2.Machine Learning:

- Data Preprocessing
- Exploratory Data Analysis
- Feature Engineering
- Model Selection
- Model Training and Evaluation
- Hyperparameter Tuning
- Interpretation of Results
- Ensemble Learning

Deliverable: The Both Statistical Analysis and Machine Learning techniques can provide valuable insights of Analysis Techniques. The choice of the technique depends on the nature of the data and the specific questions you aim to answer.

5.Visualization

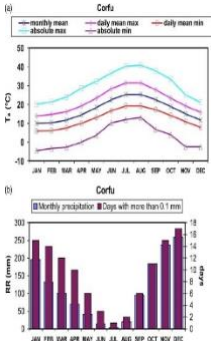
Objective: Design visualization to present the analysis results in an understandable and impactful manner.

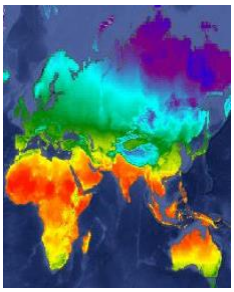
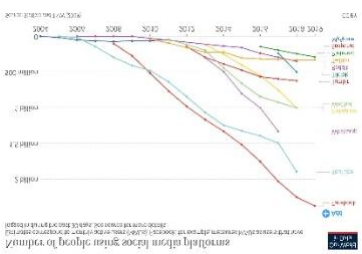
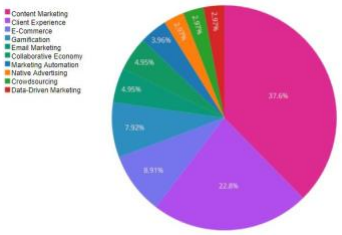
To visualize the analysis results we need to use some tools such as:

- Google Charts
- Tableau
- Infogram
- ChartBlocks

Types:

1. Column Chart
2. Area Chart
3. Pie Chart
4. Scatter Plot Chart
5. Bar Graph
6. Line Graph
7. Bullet Graph



Deliverable: To represent the analysis results in visualization tools can provide valuable insights in an understandable and impactful manner.

6.Business Insights

Objective: Interpret the analysis findings to derive valuable business intelligence and actionable recommendations.

- Finding new customers
- Increasing customer retention
- Improving customer service
- Better managing marketing efforts
- Tracking social media interaction

- Predicting sales trends

Deliverable: These are helps to provide insights that improve the way our society functions.

Conclusion

This project is aims to develop into big data analysis using IBM Cloud Databases. The objective is to extract valuable insights from extensive datasets, ranging from climate trends to social patterns. The project includes designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and visualizing the results for business intelligence. By following this structured approach, we will develop a highly effective and user-friendly virtual guide that meets the project's objectives.

THANK YOU