Big-Data-Project---Group-3

Deliverable 1

1) Team:-

a) Members

- Vaishnavi Baiken - 801316841
- Sushanth Reddy Boggula - 801305544
- Nishanth Dangethi - 801308165
- Yashaswini Golla - 801312123
- Amar Chowdary Gundapaneni - 801304875

b) Communication plan to include project artifact repository

- Github Repository link: https://github.com/sushanth-29/BigData_project_group3.git
- Project notes link: https://docs.google.com/document/d/1EthjbY5u5l4Zx97TPrZbQaXrV8dU14SIRiT1zm3SffY/edit?usp=sharinggZFE/edit
- As a team, we have decided to follow up virtually through Google Meet or Zoom every Thursday for analyzing ongoing project feature additions and solving issues that might surface during the development.
- Every Friday and Saturday, we have planned to discuss and assign the next project deliverables among the team.

Choice 1:

2) Selection of data to analyze from:

- Forest Cover Type Data Analysis
- Kaggle link:

https://www.kaggle.com/datasets/uciml/forest-cover-type-dataset?datasetId=308&sortBy=voteCount

3) Business Problem or Opportunity, Domain Knowledge (link to information on domain relative to data, problem or opportunity)

- **Business Problem or Opportunity:**

Forest Management: One business opportunity that can be derived from the Forest Cover Type Dataset is in the field of forest management. By analysing the various attributes of the dataset such as elevation, slope, soil type, and vegetation, forest managers can make informed decisions on forest planning, regeneration, and monitoring. This can result in more efficient and sustainable management of forest resources.

- **Domain Knowledge:**

Forestry and Environmental Science: The Forest Cover Type Dataset requires domain knowledge in forestry and environmental science. Understanding the different tree species, soil types, topography, and climatic conditions that influence forest cover is important to fully interpret the data.

Geographic Information Systems (GIS): The Forest Cover Type Dataset is a spatial dataset and requires knowledge of GIS. GIS is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. Understanding GIS is crucial for analyzing the spatial distribution and relationships of the forest cover types. The following link provides useful information on GIS: https://www.gislounge.com/what-is-gis/

4) Research Objectives and Question(s) (what you are trying to describe or predict with the data)

Objective:

The Forest Cover Data Analysis may be studied using Amazon services such as S3 for storage, Glue for data crawling, and Jupyter Notebook for data preparation and cleaning. KPIs can be developed extract meaningful insights like sales in different age groups. The results can presented on Amazon QuickSight a cloud-based business intelligence service. Designing detailed and engaging dashboards for visualizing our derived conclusions after analysis using Amazon QuickSight.

Question(s)

1. Do seasonal changes hill shade at noon stay relatively the same over years?
2. Is the relationship between soil type and horizontal distance to hydrology affected by horizontal distance to roadways?

3. Are specific tree types more likely to grow/are more affected by environmental factors such as soil type and distance to fire points?
4. Does the relationship between shade at different times of the day affect the soil type?
5. Which tree types can grow in more diverse environments? Are there certain tree types that are sensitive to an environmental factor, such as elevation or soil type?