MEETING NOTES

Date: 04/11/2024 - 05/05/2024

Communication:

The communication is done through teams, meeting in library. The detailed communication plan is provided in the table below.

Date and Time	Method	Tasks
April 11th	Library	Finalizing the use case
April 12th	Teams	Dataset Selection
April 13th	Library	Filter Dataset using R Studio and finalizing the dataset
April 15th	Teams	Progress Update
April 16th	Teams	Discussing the queries
April 22nd	Library	Establishment of bucket cluster
April 28-May 1st	Teams	Running Big Queries with finalized queries
May 2nd	Library	Creating tables and running queries in Hive and Spark
May 3rd	Library	Working on document and presentation
May 4th	Library	Recording Presentation
May 5th	Teams	Turn in Project/Finish Up

Individually Assigned Roles:

	Sumana Sree Nandala	Venu Gopalan Krishnagiri Tuppal	Rachana Bairi	Varun Kumar Chennuri	Mani Chandra Navuluri	Total
Data Filtering using R Studio	20%	30%	15%	20%	15%	100%
GCP	20%	20%	20%	20%	20%	100%
Hive and Spark	15%	25%	15%	25%	20%	100%
Documentation	20%	20%	20%	20%	20%	100%
Presentation	20%	20%	20%	20%	20%	100%
Participation	20%	20%	20%	20%	20%	100%

References:

8 steps in the data life Cycle | HBS Online. (2021a, February 2). Business Insights
    Blog. https://online.hbs.edu/blog/post/data-life-cycle

Google Cloud Documentation. (n.d.). Google Cloud. https://cloud.google.com/docs

Dataset URL: Consumer Financial Protection Bureau - Consumer Complaint
    Database https://catalog.data.gov/dataset/consumer-complaint-database

UNT Course Material (Module 6, 11 and 12)

A grammar of data manipulation. (n.d.). https://dplyr.tidyverse.org/ (For Processing Data)