Assessment task 2: Collaborative Development of an end to end project using Centralized Code Repositories + Github usage analysis and reflection of the project

Intent: Is to tie all the pieces taught in this Subject together. Do an end to end analysis and collaborate as a team using Github. To analyse the github usage of the entire group and reflection of your own project in terms of what worked and what didn't work.

Objective(s):

This task addresses the following subject learning objectives: 1 and 2

This assessment task contributes to the development of course intended learning outcome(s): 2.1, 2.2, 2.4, 3.3, 4.1 and 5.1

Type: Report/Code

Groupwork: Group, group assessed and Individual

Weight: 30% + 10%

Criteria:

1. Research on the effective data stores. And designing the data warehouse appropriately. (Group), 20%
2. Using SQL/R/Python to do basic analysis. (Group), 20%
3. Clarity and why a certain programming language was chosen. (Group), 10%
4. Appropriateness of commits and branches to collaborate within a team using Git, adhering to one of the documented workflows. (Group), 10%
5. Clarity and efficiency of content review and change negotiation using Pull Requests, and successful incorporation of individual changes into the team's master branch. (Group), 20%
6. Presenting and clearly communicating your findings as a report. (Group), 10%
7. Clarity on highlighting the individual and teams efforts on github usage. (Individual), 5%
8. Articulating what worked and didn't work in during the project. (Individual), 5%

Assessment 2

9. Form groups of 5. Make sure everyone has a github id.
10. Create a github repo where you will commit all your code.
11. Pick a topic that you want to research on (It can just be analysis doesn't have to be an ML solution).
12. Gather Data for that topic.
13. Start designing your Ingestion -> Storage -> Processing -> Visualization -> Reporting solution.
14. Then design your data store. So e.g. if you decide to use MYSQL which is a Relational database then design the Database as an OLTP system. However, if you decide to use

Redshift then design the database as an OLAP structure or you could just choose to access data directly in R or Python as files.

15. Now begin with your analysis. And you can do so using anyone (SQL/R or Python) or a combination of any languages like R and SQL or Python and SQL. Be clear to put in your document as to why you chose SQL/R/Python.
16. Visualizations have to be done in R and/or Python code. You cannot use Tools like Tableau etc. Not looking for anything fancy, just basic analysis.
17. All of the code should be regularly committed to Github over time. And there should be a lot of collaboration between team members.

This is the first aspect of your group assessment which is 30%.

The 10% of the report is individual analysis of your team's github data where we want you to analyse the team's github actions as a whole and write a reflection piece about this project.

Everyone needs to submit the assignment (Report + Code), so the first 30% will be the same for the 5 people in the team but just the last 10% bit will be different in every report.