# DURHAM COLLEGE

DATA 1202 - DATA ANALYSIS TOOLS ANALYTICS

ASSIGNMENT 5 – DATA ARCHIVE

Date – August 9, 2023

Name – Harsh Rawat

Id – 100852796



Submitted To –

Prof. Anthony Ridding

The course highlights various significant points and understandings related to data analysis tools and analytics:

1. Competitive Advantage Through Data: The summary underscores the increasing significance of data analysis tools in giving businesses a competitive advantage. It acknowledges that organizations can gain valuable insights from the vast amounts of data they generate and store.

2. Advancements in Analysis Tools: The course recognizes the importance of technological advancements in analysis tools, such as Hadoop and NoSQL databases. These tools enable the processing and analysis of large datasets, overcoming the limitations of traditional SQL databases.

3. Practical Implementation: The course suggests that merely adopting the latest data analysis tools does not guarantee success. It emphasizes the need to focus on practical implementation and how these tools can provide meaningful business insights.

4. Problem-Solving Approach: The course advocates for a problem-solving approach to data analytics. It implies that data analysis tools should be used as a means to solve specific business challenges, rather than being used in isolation.

5. Strategic Decision-Making: By highlighting the role of data analytics in solving business problems, the course suggests that data analysis tools are crucial for making informed and strategic decisions.

6. Balancing Technical and Business Knowledge: To effectively utilize data analysis tools, individuals need a combination of technical skills and business understanding. Understanding how to interpret data in a business context is as important as technical proficiency.

7. Critical Thinking: The course encourages critical thinking about data analysis tools and their application. It underscores the need to approach data analytics with a clear purpose and strategy.

8. Continuous Learning: The course implies that the field of data analysis is constantly evolving. Staying updated on the latest tools, techniques, and best practices is essential for ongoing success.

In conclusion, the key takeaways from this course include the importance of utilizing data analysis tools to solve real business problems, the need to align technical solutions with business challenges, and the recognition that technology alone cannot replace thoughtful analysis.

LEARNING FROM THE FINAL ASSIGNMENT

From the (FINAL ASSIGNMENT), there are multiple key takeaways that can be derived:

1. Data Extraction and Transformation: I learned how to extract data from a CSV file using the panda’s library in Python and perform transformations on the data using aggregation and JOIN operations. This involved understanding the data structure and making decisions on which columns to use for analysis.

2. SQL and Database Interaction: I gained experience in using SQL queries to create a new view in a MySQL database. This involved establishing connections to the database, executing queries, and understanding how to structure SQL statements for data manipulation.

3. Version Control with Git and GitHub: The project emphasized the significance of employing version control tools like Git and GitHub to maintain an organized repository for your project. I learned how to create a repository, add files, and create README to document your project.

4. Project Organization and Documentation: I learned how to structure a project by dividing it into distinct steps (extraction, transformation, loading), creating separate script files for each step, and documenting your code and process in a README file. Proper project organization and documentation are crucial for clarity and future reference.

5. Data Loading and ETL Process: I gained insights into the process of loading cleaned and transformed data back into a database. This included considerations for data types, creating tables or views, and ensuring the data is stored efficiently for future analysis.

6. Problem-Solving and Debugging: I encountered errors and learned how to troubleshoot and debug issues in your code. This process helps improve my problem-solving skills and taught me to think critically when facing challenges.

7. Reflection and Continuous Improvement: The project's reflection section allowed me to analyze your experience, challenges faced, and lessons learned. Reflecting on my work can help me identify areas for improvement and plan better strategies for future projects.

8. SQL and Database Skills: I acquired practical skills in SQL, including creating tables and views, using aggregation functions, and performing JOIN operations. These skills are valuable for working with relational databases and analyzing data.

9. Collaboration and Sharing: By using Git and GitHub, I learned about collaboration and sharing code with others. This is a crucial skill for teamwork and contributing to open-source projects.

10. Attention to Detail: The project required attention to detail, especially when working with data extraction, transformation, and loading. A small mistake can lead to incorrect analysis, highlighting the importance of being thorough in my work.