Reflective Report on Data Management and Exploratory Data Analysis Project

Description:

The data management and exploratory data analysis project for the CSC8631 module provided an invaluable opportunity to delve into real-life project methodologies. Spanning four weeks, the project involved a two-week timeline to conduct exploratory analysis on the Futurelearn MOOC cybersecurity course dataset. This endeavor required the utilization of various tools such as project templates, Git, R Studio, tidyverse, ggplot, and dplyr. Key concepts encompassed reproducibility, replicability, CRISP-DM, and data quality assessment.

Experience Overview:

This experience was a pivotal learning curve, providing insights into the practical approach required for project development. Embracing the CRISP-DM model enlightened me on the structured path to problem-solving. Adopting tools like tidyverse for the first time was challenging yet immensely rewarding. Project templates proved instrumental in ensuring project reproducibility, while Git facilitated version control effectively.

Learning Points:

Engaging with these tools instilled confidence in handling projects of considerable magnitude. Overcoming coding hurdles with perseverance bolstered my skills, aligning with the coding styles emphasized in the module. Crucially, formulating a pertinent business problem and thinking from a stakeholder's perspective honed my business acumen. Despite data quality issues, leveraging instincts led to compelling insights, highlighting the significance of problem formulation and critical thinking.

Key Takeaways:

The project duration illuminated the methodologies for managing extensive projects, emphasizing the vital role of defining precise business questions. The significance of comprehensive review cycles within CRISP-DM became evident, ensuring continuity and coherence in analysis. Confronting data quality challenges underscored the realism of such issues in practical scenarios.

Future Application:

Reflecting on this experience, a structured problem-solving approach, adherence to best coding practices, and a business-centric outlook will be integral in future projects. Moreover, overcoming the challenges of learning new tools offers a roadmap for approaching upcoming technological learning endeavors.

Conclusion:

The project within CSC8631 was a transformative experience, providing practical insights into project development, data analysis, and problem-solving methodologies. It equipped me with a holistic perspective on managing data-related challenges and establishing a strong foundation for future endeavors.