

# Reflective Report

Aulia Amirullah Zulkarneidi - 220606343

2022-11-16

## A. How I Approached the Task

The first thing I did to complete the task is:

1. Going through the CRISP-DM to structure the Data Analysis Pipeline;
2. Understanding Reproducibility and Replicability;
3. Getting a basic understanding of Data Project Life Cycle;
4. Reading Data Quality articles;
5. Going through some codes of dplyr and ggplot;
6. Understanding the assignment;
7. Create ProjectTemplate with the guidance of the tutorial on Canvas;
8. Making the analysis by reviewing dplyr and ggplot2 to get the techniques on points by highlighting the CRISP-DM and following the ProjectTemplate guidance;
9. Making the report by going through the RMarkdown guidance on Canvas;

What I've done is to produce plots that can feed into 2 cycles. Those plots are exploratory analysis that fits into the CRISP-DM approach. If I had more time, I might apply Logistic Regression. I'd probably try to explore more on how to make the reports more visually attractive with RMarkdown.

## B. The Tools and Techniques

In my opinion, CRISP-DM is a good framework to structure my analysis of a Data Mining Project so that I can make a project from scratch starting from identifying the user's needs and the objectives of the project that I'll be working on. It's a good framework to get a basic understanding of a problem that a customer wants or might be interested in. It will be really useful for the future when we get into a real project and we would know how to get started with a project by identifying a problem, list resources, risks, schedule as the requirement of the Business Understanding and we can move on to Data Understanding.

With Data Understanding itself, I get an understanding what to do with it starting from Legal Issues and Concerns, Data Procurement, What the Data tells me and even identify what I can get from a dataset before I get started with the project so that I won't encounter a problem later in the future.

In terms of Data Preprocessing, Modelling, Evaluation Deployment. It's a really good scheme for me to understand and probably I can directly apply the cycles and we can keep going back and forth through CRISP-DM as needed. Therefore, it's a good framework to start with project analysis and even identify risks and contingencies during the Business Understanding step.

ProjectTemplate is a really good framework for Data Analysts so that everyone can structure their project better. It may make the development of a project way easier since we've divided files based on their functionality. My experience based on the framework is splendid as it makes me easier organize and locate all my data files without any efforts. Even producing a standardized report is easier with RMarkdown from ProjectTemplate.

My Experience with the datasets is a little bit overwhelmed as there are many datasets I need to assess and eventually come up with two cycles. Even though, it doesn't take time to eventually come up with two cycles, we miss two datasets of the first two years of Video Statistics.

Dplyr and GGplot helps me so much to build a pipeline and plots I need for my analysis. It makes it easier and simpler for my analysis so that I can work more on other analysis without worrying how I can combine data selection, cleaning and munging within one process.

## C. The Task

The task is considered moderate as it is actually about following the CRISP-DM approach and make the project using ProjectTemplate as everything is already provided. However, it was a bit complicated at first as I was worried that my analysis might be wrong, or I'm doing not enough analysis. I was worried so much about the complexity of my analysis as I'm just doing an exploratory analysis not even come up with a prediction or more sophisticated analysis which made me overwhelmed. That's because I was afraid that I was not doing enough with my analysis. However, as I was explained that the task is not about how I come up with sophisticated analysis but rather following CRISP-DM to carry my analysis and starting a project with ProjectTemplate.

I wish the first and second year of Video Statistics datasets were provided. It would not make me feel guilty that I was not doing enough or trying to shift to other datasets but I was afraid I was not going to meet the deadline.