**Project Proposal**

The aim of our project is to use the ETL process to clean a dataset of a subject we were interested in. The topic for our project was to “uncover gender patterns in video games across multiple platforms relating to gender representation within modern day games”.

Before we started the technical work of cleaning the data and the ETL process we needed a dataset, in our group we initially held extensive discussion around the outline and scope and purpose of the project.

Firstly, we went away as individuals and bought back topics idea’s for this project and potential datasets. Then whilst reviewing the idea and dataset as a group we came to a unanimous decision by holding a vote for the project idea we wished to move forward with. The main focus of the vote was focusing on the criteria of,

1. Interesting topic
2. A wide variety of data sets
3. Needing transforming or cleaning

Graphical user interface, text, application, chat or text message

Description automatically generated

Once this was decided we as a group outlined the scope and purpose of our project.

Deciding who would take on what roles and tasks and what areas of the dataset needed changing to conform to our SQL database and tables.

Giving us focus and provide direction we create realistic short sharp targets to allow us to stay on track and also get all of us members to be on hand to help each other.

* The data we wanted to work with was preferably multiple csv files that we can join based on a common link within our agreed area of Gender representation within video games across multiple platforms.
* We’ll be examining the ways in which gender is represented by examining the relationships between Character id and Sexulixation ID, Game ID and character in game over the dataset which represented 638 character over 65 popular games across multiple platforms as well as other related relationships derived from the data.
* Source of our data is from Kaggle.com & diamondlobby.com where we found our data sets (3 datasets) in the form of csv file and relating to our topic of gender representation within video games. Also our final databased will be of a relational database structure containing data organised into tables.
* √ Gives relevant and succinct description of findings (2–3 sentences)