**Project Proposal**

The aim of our project is to uncover patterns in games across multiple platforms relating to gender representation within modern day games.

Before we start the technical work of cleaning the data and the ETL process, our group held extensive discussion around the outline and scope and purpose of the project.

Firstly, we went away as individual and bought back topics and idea’s for this project and then upon reviewing 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 your project.

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

… and purpose of your project???

This gave us focus and helped provide direction and focus and we had realist short shsrp targets to allow us to stay on track and also get all of us members to be on hand to help each other.

* The kind of data you’d like to work with and the field you’re interested in (finance, healthcare surveys, etc.) the data we would like to work with is 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.
* The questions you’ll ask of the data. 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, and other related relationships derived from the data.
* Source of our data is from Kaggle.com.

**Techical Report**

For Project 2, we worked as a team to complete the ETL process on a dataset.

**Extract**

**Transform**

**&**

**Load**

After considering multiple datasets our group decided on the topic of gender representation within modern day video games across multiple platforms.

With three datasets (638 character data across 65 modern day popular games)

* We used three datasets focused on PlayStation games info sources from Kaggle.com. These datasets were on games data, character data & sexualisation info.
* The types of data wrangling that you performed (such as cleaning & joining,). The types of functions we used as part of our transform element of the process is we cleaned select columns to ensure the Datatable in our jypyter is similar to our tables within SQL over the 3 datasets and data tables, we joined the data set on…….as well cleaned up column names to either simplify names or renaming columns to shorter names.
* The schemata that you used in the final production database.

You should focus your efforts on a specific industry. To help you choose, the following subsections provide examples of how ETL gets used in various possible specializations.

Attached is the scheme that we used in our final production database.

* Decide how you’re going to communicate with your group members when you begin. Create a Slack channel, exchange phone numbers, and ensure that the group knows each group member’s available working hours.
* Create internal milestones to ensure that your group is on track. Set due dates for these milestones so that you have a timeline for completing the project. Some of these milestones might include:
  + Project ideation
  + Data fetching/API integration
  + Data analysis
  + Testing
  + Creating documentation
  + Creating the presentation