**Project 3 Proposal:**

**Data Visualisation Track:**

Topic: (custom) 🡪 Steam Games Dataset analysis + visualisation

Source files:

1: Kaggle.com, **Martin Bustos, Steam Games Dataset**

url: <https://www.kaggle.com/datasets/fronkongames/steam-games-dataset>

2: Kaggle.com, **Nik Davis, Steam Store Games (Clean dataset)**

url: <https://www.kaggle.com/datasets/nikdavis/steam-store-games>

New Update: ( In the first url, please only download the ‘games.csv’ file. In the second url, please only download the ‘steam.csv’ file. As I merge two of the csv files together for this project.)

**Reason for choosing this topic:**

I am interested in providing analysis and visualisation of the world most widely used PC game platform: Steam. (which is a cross platform software that able to purchase, download, share, create, and publish game/videos/comments/even real trading.

I found this huge dataset on Kaggle which the author scrape the steam raw data using steamspy, the accessible file in the url is 1 csv file and 1 json file. Both providing the same raw data for analysis.

From the initial view of the dataset, I brought some question that can be used for this project using this data:

1: what are the 5 most popular games from 2020 to 2024?

2: what are the relationship between user review (positive or negative) and games’ price?

3: Which publishers have the highest positive ratings?

4: Do games with certain genre or descriptors tend to be more popular?

5: What factors are associated with longer average playtimes, and how do they impact a game's success and player engagement?

For all the questions above, the approach is below:

1: use pandas to read csv & perform data cleaning ->

2: segment dataframe into two/three tables (cleaned already), use the appID as foreign keys to create schema and load data to Postgres (this can be done either manually / using sqlalchemy + psycopg2) ->

3: now we can read data & perform visualisations (here I propose the NEW library for plotting: Seaborn)

4: for certain question above, we can visualise using javascript(d3.js + plotly) to create at least 1 interactive plot for user.

5: create flask API for all the plots above, then the user-interaction can be achieved via visiting urls and view all the graphs.