

COMP 3125 Data Science Fundamentals

Fall 2025

Topic of Individual Project

Due at 11:59 pm on Sunday, Nov 2

Based on the introductions of individual projects, please decide the topic of your individual project. Fill in the following questions and submit the filled PDF version before the deadline.

1. What's the topic? Add a short description here

The topic is about Steam, a video game hosting store, and using the data available on the platform to determine what makes a video game popular.

2. If the dataset exists, what's the link of the dataset/datasets? If not, how will you create this dataset?

Dataset: <https://www.kaggle.com/datasets/nikdavis/steam-store-games/data?select=steam.csv>

3. What's the three questions to solve/answer to this topic?

Questions:

1. What game features are most strongly associated with popularity on Steam?
2. Can we predict whether a game will be popular using only metadata available at release time?
3. How do pricing and release timing relate to popularity?

4. Among your questions, which one need quantitative analysis (Training a model) to get your answer. What model/method do you plan to use?

Among my questions, the one that would need quantitative analysis is question 2: can we predict whether a game will be popular using only metadata available at release time?

To answer this question, I will train a classification model that predicts whether a game is “popular” or “not popular” based on features such as price, release, year, genres, categories, and platforms.

I plan to use logistic regression as my main model because it is simple and interpretable. I may also compare it with a random forest classifier to see if a nonlinear model improves performance.

5. Are there any suggestions you want to provide for the instructor?

(None currently)