

Project Plan

Project Objective

The goal of this project is to explore visualization tools and techniques for a medium sized data set. In particular, our aim is to use this opportunity to examine video game sales data and slice and manipulate to see if any trends or bias emerge.

Project Approach

The data set will be sourced from kaggle.com, a source for datasets typically used for machine learning projects. The most promising data set looks to contain approximately 56,000 entries. Next we will research tools and techniques for arranging and examining the data. Our initial thought is to use python with numpy and pandas library for data analysis, but there are other packages and applications we'd like to explore (such as Tableau).

Once we have data in an environment we can easily manipulate, the next step is to determine if we can answer some of the following questions and similar:

- Publisher's with lowest review scores
- Platforms with best user scores
- Platforms with biggest discrepancy between critic and user scores
- Units sold versus marketing budget
- Biggest publishers for each platform
- Most popular genres by sales
- Revenue per genre each year

Team Structure

As a team of two, the plan is to largely work in a pair setting on the data analysis itself. Both of us have some degree of python experience and working with data imports.

Project Milestones

Main dataset identified

Additional sources identified for corroboration

Visualization techniques/tools explored and main tools chosen

Data sliced to support/answer proposed questions

Visualizations created for final presentation based on data slices

Final presentation assembled