|  |  |
| --- | --- |
| Project Title | Kickstarter Video Games Success and Popularity in Steam |
| Team | Peter Witwer |
| Project Description | Use several datasets from Kaggle to practice using ETL with.  The goal will be to end with a database that has detailed information for each game that started on kickstarter, made it to steam, and then how successful it was after making it to steam |
| Relational or non-relational database? | Relational –pgadmin4 |
| Datasets | * kickstarter csv dataset * kickstarter games that made it to steam csv * Steam csv dataset that includes a variety of interesting columns for each game * Potentially other datasets depending on time constraint? |
| Concerns | * Does this meet project requirements? * Should I narrow/expand the scope of my project? (I am working alone) * Is there anything else I need to be considering here? This project seems deceptively simple. We’re not looking for anything special with the data, just practicing using ETL, correct? |

Kickstarter Video Games Success and Popularity in Steam

ETL Technical Process

**Extract:**

I used several databases downloaded from Kaggle.com formatted as csv files.

1. <https://www.kaggle.com/trolukovich/steam-games-complete-dataset?select=steam_games.csv>
   1. Data type of all columns set as “object”
   2. Only need name, recent reviews, all reviews, release date, publisher, game details, genre, and original price
   3. Columns contain non-ascii characters.