Names: Moshe Dego 315044511, Tzach Fleischer 305152969

Research question:

for regression: we intend to find out if the numerical value of video game critic score may be predicted based on metacritic data

background/motivation: personal interest in the gaming industry

Features (number and names of columns of data and what they represent):

See at the end .

Instances (number of rows, and what each of them represent): 200,000 approximately, each representing a different game or DLC(expansion packs).

Data sources (websites, what kind of edge they give you over publicly open sources):

Site: <https://www.metacritic.com/browse/games/score/metascore/all/all/filtered>

the site is an aggregator of game reviews and contains lots of entries & info

Data mining methods (API and/or crawling, which libraries and how deep do you need to go to get your data): crawling, beautiful soup libraries

Planned visualizations (2D/3D/ND, dimension reduction methods etc.): 2D

Planned models (appropriate models and loss for the task): supervised learning - linear regression

Validation methods: cross validation

DATASET columns(13): subject to change

game\_name – name of the game

score – critic score (what we want to predict)

user score – score by the users

developer – name of the company that developed the game

genre – the genre of the game

number\_of\_players – max number of player simultaneously

rating – the age group that is allowed to play this game

developer score – the company aggregator median score of all the games they developed

number of critic reviews – number of critic reviews

number of user reviews – number of user reviews

release date – the released date.

Platforms – what platforms the game was released to

Description – description about the game

if we'll have the time we want to do text processing on reviews to see if it will help predict the score