This exercise is designed to introduce you to some of the data we work with at DeviantArt, and to provide a starting point for further discussion during your upcoming interview.

First, some terminology:

- a **user** is a registered user of DeviantArt (aka a "deviant")
- An **item** is a piece of user-generated artwork on DeviantArt (aka a "deviation")
- There are 4 main types of user-item interaction on DeviantArt:
  - users can **submit** new items
  - users can **comment** on items
  - users can **favourite** items (but not an item they submitted)
  - users can **view** items

Consider browsing or searching for items on current homepage of deviantart.com. The two most-used sort orders are:

- **Newest**: orders items in descending order of submission time
- **Popular 24h**: orders items in decreasing order of number of interactions (with some weighting of different interaction types).

Neither of these metrics is personalized to the interests of the user viewing the content.

Instructions

The CEO of DeviantArt has tasked your team with releasing a personalized homepage by the end of the next quarter, in support of a broader company goal to increase the number and quality of user-item interactions. The CTO has given your team a week to decide on a high-level technical approach.

You've been asked to help the team make an informed decision. Your task has two parts:

1. Prototype an approach for making personalized item rankings. Use any tools you deem appropriate. Evaluate your approach as compared to a reasonable popularity baseline on the below dataset.
2. Write a high-level "executive summary" of your recommended approach. In this summary, describe the advantages and disadvantages of your proposal as well as the engineering work needed to put your prototype into production. Illustrate the behaviour of your prototype as compared to a popularity baseline in terms of the top 10 recommendations made to a few example users.

Please send us your prototype code and executive summary at least 24 hours before your next scheduled interview.

Dataset

The following dataset contains a random sample of user-item interactions. The interactions were sampled as follows:

1. Randomly sample a subset of users **U**
    (97156 users)
2. Identify all items **I** submitted by users **U** during a 30 day period
    (224858 items)
3. Identify all user-item interaction data involving users **U** and items **I** during that 30 day period
    (3637045 interactions)

The interaction data can be downloaded here:
https://s3.amazonaws.com/devartanalytics/exercise.interactions.csv.gz [73MB]

Thumbnail images for the items in the interactions dataset can be downloaded here:
https://s3.amazonaws.com/devartanalytics/exercise.thumbnails.tar [3.2GB]

You have permission to use this dataset only for the purposes of completing this hiring exercise. In particular, you may not use the dataset, or any model trained using the dataset, for any purpose other than communicating with employees of DeviantArt regarding your potential employment with DeviantArt. The dataset has been anonymized, and any attempt to de-anonymize the dataset is expressly prohibited.