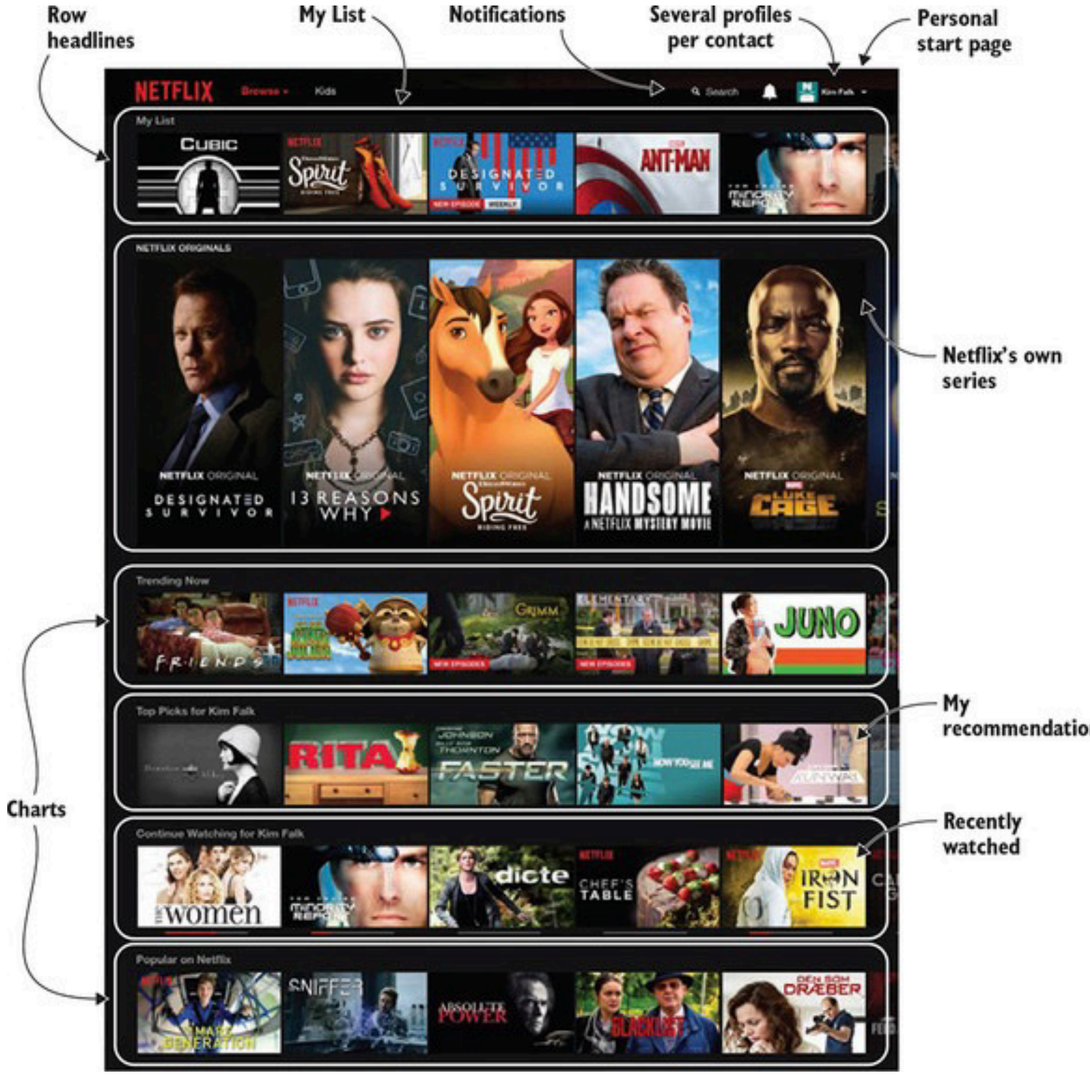
Recommender System

- A RS can recommend specific movies [Netflix has 3600 movies (vs 14 genres)]
- Users arrive at random and the RS observe some partial information about the user (e.g. movies watched, geographical location, personal information, etc.) so-called contextual information
- The RS recommends a movie to the user
- The feedback is whether the user watched and liked the movie or not
- Objective: Design a RS that maximizes the number of movies watched



Contextual Bandits

- Treat each movie as an arm, can be extremely large
- For each round t:

  observe a context x_t for the user arrived

  pick an arm a_t i.e. recommend a movie

  receive a reward $r_t(x_t, a_t)$ which is higher when a user likes a movie

Goal: Given any sequence of context $\{x_1, x_2 \ldots, x_T\}$ or a sequence of users, find an algorithm which maximises the total reward $\sum_{t=1}^{T} r_t(x_t, a_t)$