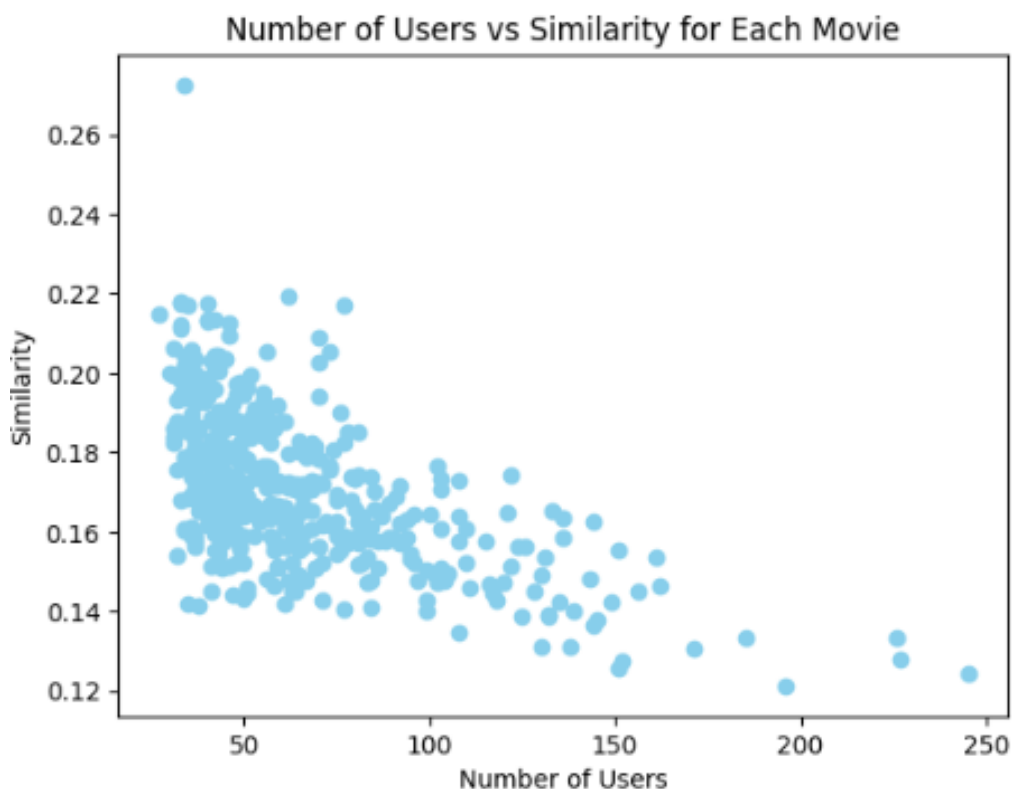


## Progress Meeting 6

이번 주에는 인기도 편향을 보완하는 방법을 조사하고, 문제점을 살펴 가설을 보충하기로 했다.

인기도 편향을 줄이기 위해서 단순히 상호작용한 횟수만 고려하는 것이 아니라, 해당 아이템을 사용한 유저들의 유사도도 고려해야 한다고 생각한다. 현재 CF 기반 추천시스템의 정적/동적 편향 제거 방법론은 대부분 상호작용한 횟수만을 고려해 그에 맞는 가중치를 학습하여 인기도 편향을 줄인다. 그러나 인기있는 아이템 중에도 특정 사용자 집단의 특성을 반영하는 아이템이 있을 것이고 그렇지 않은(취향에 관계없이 인기있는 아이템)이 있다고 생각한다. 이에 같은 상호작용 횟수를 가지더라도 다른 가중치를 부여하는 방법이 필요하다고 생각한다. 이를 확인하기 위해 코사인 유사도의 평균치를 이용해서 사용자 수와 유사도 간 산점도 그래프를 확인해보았다. 그 결과 상호작용한 사람의 수가 같더라도 코사인 유사도의 평균치에 많은 차이가 있음을 확인할 수 있었고 사용자 수가 늘어남에 따라 유사도가 점차 감소하는 경향을 확인할 수 있었다.



## [Hypothesis 1]

- > Popularity bias: 해당 아이템과 상호작용한 유저가 많은 경우를 인기있음으로 분류
- > 기존 debias method: 상호작용횟수(클릭 수)를 고려해 모델이 예측한 점수를 재스케일링
- > 취향이 유사한 사람들이 많이 보는 아이템과 그렇지 않은 아이템이 상호작용한 횟수가 유사하다는 이유로 같은 debias method가 적용되어야 하는가?

## [Challenges]

### 1. 취향에 관계없이 인기있는 콘텐츠에 대한 정의를 어떻게 할 것인가?

- > 특정 아이템과 상호작용한 유저를 추출해서 유저 간의 유사도를 유저-아이템 행렬을 이용한 코사인 유사도 지표를 활용해 구한다. 그 유사도가 낮을 때 취향에 관계없는 아이템이라고 할 수 있다.

### 2. 추출한 유저 수가 많을수록 유사도가 낮아지지 않을까?

- > 그렇다. 즉 사용자의 수와 유사도를 복합적으로 고려해서 debias를 해야 한다.
- 왜? '사용자 수가 많고 유사한 집단이 상호작용한' 콘텐츠와, '사용자 수가 많고 유사하지 않은 집단이 상호작용한' 콘텐츠는 서로 다른 의미를 가지게 될 것이다.

### 3. '대중적인 취향' 을 가진 유저는 어떻게 알아낼 것인가?

- > 총 2단계로 분류한다.

[1단계]

1번 과정에 따라 취향에 관계없이 인기있는 콘텐츠를 라벨링한다.

[2단계]

해당 아이템들이 상호작용 이력의 많은 비율을 차지하는 유저들을 라벨링한다.

### 4.. 아이템 소비 = 자기 취향 + 대중 취향, 기존 연구와 무엇이 다른가?

- > 기존 연구는 단순히 상호작용한 사람의 수가 많은 것을 'Popular item'으로 지정했지만 해당 아이템과 상호작용한 사람 간의 취향 유사도도 함께 고려하고 싶다. 즉 특정 유저의 취향에 '대중적인(취향에 관계없는 아이템을 좋아하는) 취향'이 포함될 수 있는 것이다.

### 5. 어떻게 평가할 것인가?

- > 기존의 정확도에 대한 개선이 가능할 것이다. 대중적인 취향을 가진 사람들에게는 대중적인 취향을 고려한 추천이 이루어질 것이고, 그렇지 않은 사람들에게는 대중적인 아이템의 영향을 줄일 수 있을 것이다.



## **[Hypothesis 2]**

- 협업 필터링에서 popularity가 낮은(상호작용한 사람의 수가 적은) 아이템에 대한 보강이 필요하다. (collaborative filtering에서 인기있는 아이템이 ideal value에 비해 높게 평가되기 때문)

1. 개인의 취향을 정의하기 위해 취향에 관계없는 아이템을 제거해야할 것이다.
2. 낮은 상호작용 횟수를 가진 아이템 중 유사한 아이템끼리 상호작용 내역을 공유한다.

