

Business Intelligence

Proposal for recommendation System

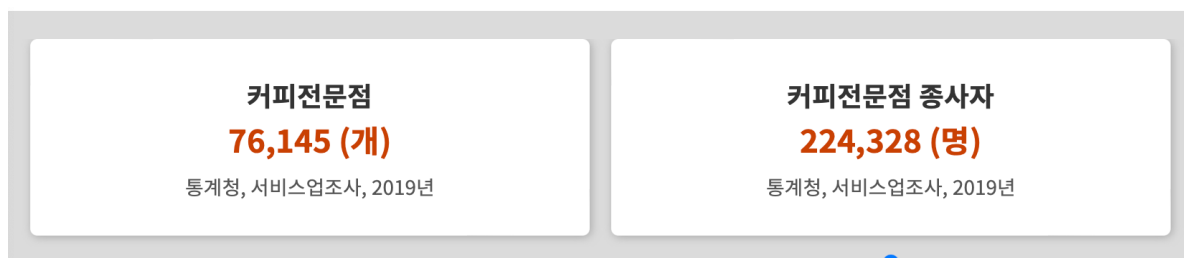
20172861 김재민

Business Domain : Cafe

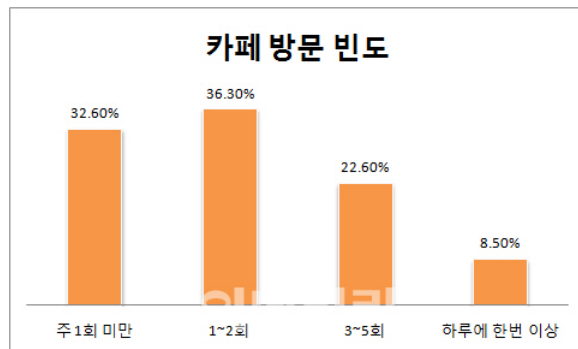
⇒ Café Recommendation System

Motivation for Proposal

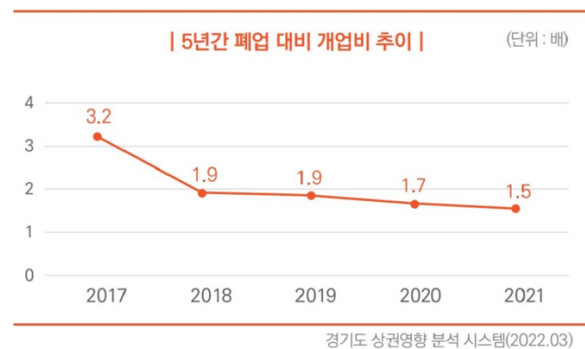
- Statics



Referece -KOSIS(국가통계포털)



Reference – OpenSurvey



Explanation for motivation

2022년, 카페는 더 이상 커피를 마시는 공간이 아닌, 생활문화적 공간이라 칭해도 어색하지 않다. 위와 같은 통계를 통해 카페 수는 오히려 늘어나고 있으며, 사람들의 방문빈도 수는 나날이 증가하고 있다. 본래의 기능인 단순히 커피를 즐기는 사람들, 사회문화적 모임, 데이트, 심지어 공부나 단순히 사진을 찍기 위해서 등등 여러가지 이유로 카페는 대한민국 생활에 있어 빠질 수 없는 요인이 되었다.

하지만 대부분 카페를 검색하거나, 별점(평점), SNS를 통해서 찾게되는 경우가 80%이상에 달한다. 따라서 사용자의 취지에 맞는 카페를 추천할 수 있도록 하는 카페 추천시스템을 제안하게 되었다.

Process for Café Recommendation System

1. Collect data

카페에 대한 정보를 받아오고, 별점이나 평가에 대한 데이터도 받아와야 하므로, 네이버지도, 카카오맵 등에서 데이터를 크롤링한다.

2. Describe a data model

크롤링 데이터의 사용자 수가 불분명하고, 모든 사용자가 평가를 하지는 않으므로, 중복데이터를 제거하고, 사용자 수를 충분한 수(1000명)로 고정한다.

40000000(자국민 수) : 1000이므로, root값인 200으로 사용자당 1명의 평가수로 계산한다.

Training set : Test set = 4 : 1 로 하여 1인당 160개평가 : 40개의 평가를 확인한다.

3. Develop environment



4. Learning

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$\text{pearson_sim}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \mu_u) \cdot (r_{vi} - \mu_v)}{\sqrt{\sum_{i \in I_u} (r_{ui} - \mu_u)^2} \cdot \sqrt{\sum_{i \in I_v} (r_{vi} - \mu_v)^2}}$$

- Cosine similarity vs Pearson Similarity
- Number of NearestNeighbor, Processed Crawling data(map data, Evaluation)
- NearestNeighbor이므로, Collaborative filtering 방식으로 진행한다.

5. Conformity Criteria

- 사용자만족도(대중적인 평점), 다른 사용자들과의 유사도(정확성)
- input값에 따라서, 사용자가 이용하고자 했던 취지의 카페종류에 부합한지 확인한다.
- 스타벅스와 같이, 대형커피숍뿐만 아니라, 테마카페, 개인카페 등도 포함하여 범용성을 갖는다.
(단, 스터디카페는 독서실과 같은 기능을 하므로 기준에 포함하지 않는다)

6. Expected Result

- 사용자가 그때 그때 필요에 적합한 카페를 찾기 용이해진다.
- 사용자 입장에서 광고와 같은(SNS) 매개체를 통한 단순히 유명한 카페를 찾아가는 것보다, 신뢰성이 높다.
- 사용자 수가 많아질 경우, 같은 목적성을 띤 사용자그룹이 생성되므로, 기조에 의한 각 그룹의 만족도가 높아질 확률이 비교적 높다.

7. System Improving

- NLP등을 활용하여 후기와 같은 데이터가 쌓일 수록 시스템 정확도가 올라갈 수 있다.
- 큰 목적성을 갖고 있으므로, 부가적인 알고리즘 추가가 가능하다.
- 학습에 있어서, Jacard Similarity도 추가가 가능하다.