

# 빅데이터 프로젝트 제안서



조민수 20161851  
이상진 20162191  
김재민 20172861

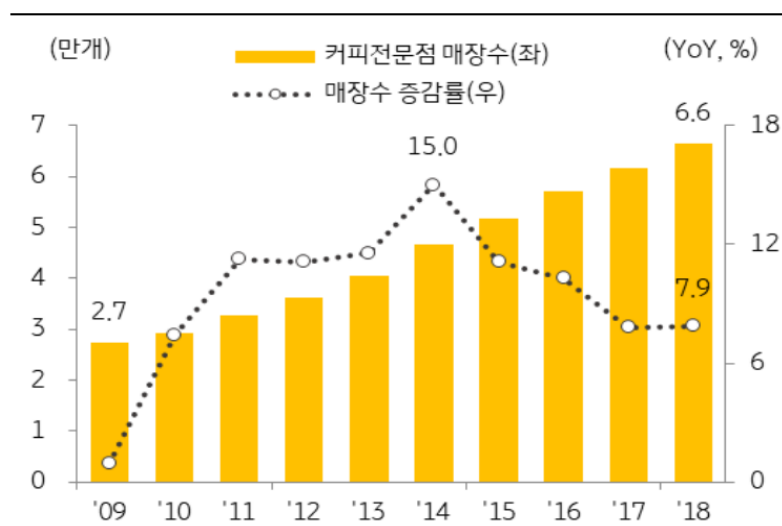
## 주제 요약

카페 방문 목적이 다양화된 지금, 사용자의 요구에 맞는 사용자 맞춤 카페 추천 시스템

## 주제 기획동기 및 필요성



[fig1. 국내 커피 산업 현황, 2019, KOSIS(국가통계포털)]



[fig2. 국내 커피 전문점수 추이, 2019, KB금융지주, 커피전문점 현황 및 시장여건 분석]

2022년, 카페는 더 이상 커피를 마시는 공간이 아닌, 생활문화적 공간이라 칭해도 어색하지 않다.

2019년 기준 8만개에 달하는 커피 전문점들이 국내에서 영업을 하고 있으며, 종사자의 수도 20만명을 능가할 정도로 커피 산업의 양적 성장이 이루어졌다(fig1). 양질의 생활 문화공간을 원하는 수요의 증가로 카페의 수는 계속하여 늘어나고 있다(fig2).

### 카페(커피전문점) 이용 빈도

	전체	성별		연령대			
		남성	여성	20대	30대	40대	50대
(Base)	(1349)	(662)	(687)	(336)	(357)	(344)	(312)
일주일에 1잔 미만	0.5	0.2	0.9	-	0.3	0.3	1.6
일주일에 1~3잔	70.1	68	72.1	66.7	64.7	72.7	76.9
일주일에 4~6잔	21.7	23.4	20.1	25.3	25.5	17.4	18.3
일주일에 7~13잔 (하루 1~2잔)	6.8	7.6	6.1	7.4	7.8	8.4	3.2
일주일에 14잔 이상 (하루 2잔 이상)	0.9	0.9	0.9	0.6	1.7	1.2	0.0
평균 이용 빈도 (일주일 기준)	3.1잔	3.2잔	3.0잔	3.4잔	3.4잔	3.1잔	2.5잔

[Base: (스크리닝 조사) 최근 1개월 내 카페(커피전문점) 이용한 자, N=1349, 단수응답, %]  
\*하늘색 음영: 전체 평균 대비 +3%P 이상인 데이터

[fig3. 카페 방문 통계, 2020, OpenSurvey, 카페 이용 트렌드 리포트 (20~50대 표본 800인)]

### 카페(커피전문점) 방문 목적

■ 1순위 □ 1+2+3순위	전체	성별		연령대				주 이용 브랜드				
		남성	여성	20대	30대	40대	50대	스타벅스	이디야	투썸 플레이스	할리스	파스쿠찌
(Base)	(1000)	(500)	(500)	(250)	(250)	(250)	(250)	(550)	(176)	(101)	(23)*	(21)*
맛있는 커피를 마시기 위해	<div><div></div></div> 41.6 63.3	63.0	63.6	56.8	68.8	69.2	58.4	68.7	54.5	50.5	65.2	47.6
대화하기 위해 (친목/업무 목적 등 모두 포함)	<div><div></div></div> 27.9 53.3	53.2	53.4	48.8	50.4	50.4	63.6	48.2	61.9	66.3	47.8	66.7
시간을 보낼 장소가 필요해서	<div><div></div></div> 14.3 41.2	40.2	42.2	44.4	41.2	38.4	40.8	40.9	40.3	41.6	60.9	42.9
맛있는 빵/케익/디저트를 먹기 위해	<div><div></div></div> 5.4 19.5	13.8	25.2	22.4	16.8	20.8	18.0	22.0	11.4	33.7	13.0	4.8
공부/일을 하기 위해	<div><div></div></div> 6.3 16.0	14.4	17.6	33.6	13.6	8.8	8.0	16.4	11.9	26.7	21.7	4.8
그 브랜드에서만 판매하는 굿즈 구매를 위해	<div><div></div></div> 2.8 7.5	7.2	7.8	6.4	9.6	8.0	6.0	11.8	2.8	0.0	4.3	4.8
사진을 찍기 위해 (SNS 공유 등)	<div><div></div></div> 0.7 1.4	2.0	0.8	2.0	2.4	1.2	0.0	1.5	1.7	1.0	0.0	0.0

[Base: (본 조사) 전체 응답자, N=1000, 순위형 응답, %]

\*하늘색 음영: 전체 평균 대비 +4%P 이상인 데이터 / \* 배너분석은 1+2+3순위 데이터임

\* Small base

[fig4. 카페 방문 목적, 2020, OpenSurvey, 카페 이용 트렌드 리포트 (20~50대 표본 800인)]

20 ~ 50대를 대상으로 진행한 2020년의 설문에서 확인할 수 있듯, 설문모집단의 91.8%는 일주일에 최대 6잔의 커피를 카페에서 구매하는 것으로 나타난다(fig3).

위와 같은 커피 산업, 다시말해 커피 전문점의 양적 성장은 본래의 기능인 '커피'를 즐기는 것을 넘어, 모임, 데이트, 공부, SNS포스팅 등 여러 사회적 내인들을 배경으로 이루어졌음을 알 수 있다(fig4).

카페 방문 전 정보 탐색 방법

	전체	성별		연령				
		남성	여성	20대	30대	40대	50대	
Base	(1000)	(472)	(528)	(226)	(235)	(281)	(258)	
가족/친구/지인 추천	55.9	50.4	60.8	56.2	55.3	52.3	60.1	
인터넷 커뮤니티/ 블로그/카레	53.0	46.2	59.1	59.7	61.7	52.3	39.9	
지도 앱	52.5	53.8	51.3	64.6	55.3	47.0	45.3	
SNS	41.5	35.2	47.2	61.9	54.0	32.7	21.7	
유튜브	16.5	19.1	14.2	22.6	16.6	15.3	12.4	
다이닝 앱	5.9	7.0	4.9	8.4	8.5	4.6	2.7	
정보 탐색하지 않음	15.9	20.3	11.9	10.2	12.3	20.6	19.0	

카페 방문 시 탐색 정보

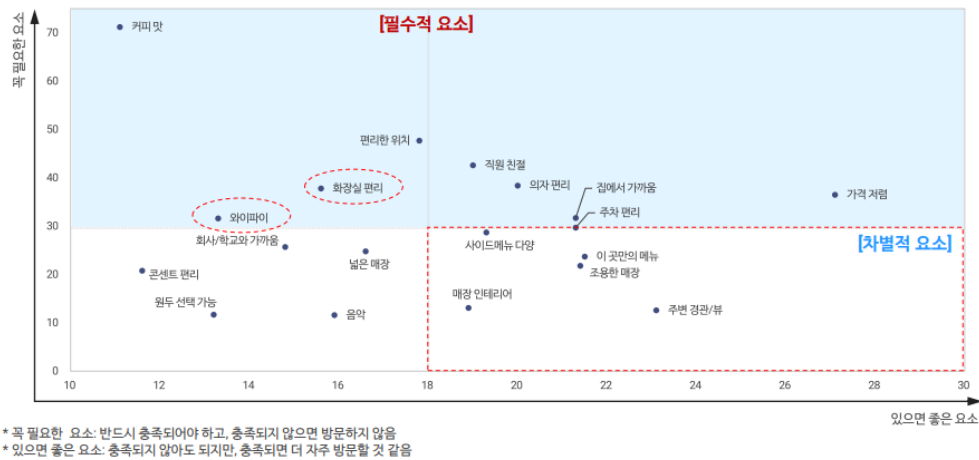
	전체	연령			
		20대	30대	40대	50대
Base	(841)	(203)	(206)	(223)	(209)
음료의 맛/퀄리티	18.8 43.3	42.9	45.1	42.6	42.6
음식(디저트 등)의 맛/퀄리티	13.7 34.7	36.0	32.0	32.7	38.3
방문자 후기/리뷰	14.0 33.7	36.9	40.3	32.7	24.9
거주지/약속장소와의 거리	11.5 29.3	21.7	29.6	30.0	35.4
인테리어	9.6 28.7	27.6	33.0	27.4	26.8
주차 가능 여부	8.1 27.0	8.4	27.7	34.5	36.4
가격	6.2 23.3	27.1	20.4	24.7	21.1
매장 시설	2.7 14.0	17.7	9.7	12.1	16.7
음식(디저트 등) 다양성	3.4 11.9	13.3	11.7	13.0	9.6
음료/음식의 비주얼	2.6 10.8	17.2	12.6	9.4	4.3
음료의 다양성	2.7 9.4	9.4	9.2	10.3	8.6
방문자 평점	2.4 9.2	8.9	7.3	8.5	12.0
영업 시간	1.5 6.4	11.3	6.3	4.0	4.3
매장 크기	1.4 6.2	7.9	8.3	4.5	4.3
음료 커스텀 가능 여부	0.8 3.0	3.9	0.5	3.6	3.8
환경보호 실천 여부	0.2 1.4	1.5	1.0	0.9	2.4

[fig5. 카페 방문을 위한 정보 탐색 경로, 2022, OpenSurvey, 20~50대 표본 800인]

위의 통계에서 확인할 수 있듯이 카페는 대한민국 생활에 있어 빠질 수 없는 요소가 되었다. 카페가 하나의 생활문화적 공간으로 대표되며, 카페에서 기대하는 요구 사항의 다변화가 이루어졌다. 때문에, 기존의 브랜드 커피를 주로 소비하는 시장에서 자신의 요구에 부합하는 카페를 찾는 경우가 많아졌으며, 2022년의 통계자료를 통해 이용자 주도적인 카페 방문을 위한 다변화된 정보 탐색이 이루어지고 있음을 확인할 수 있다(fig5). 특히 온라인 상에서 검색하여 SNS상의 리뷰, 포털 사이트에서의 평점 정보, 유튜브 등의 매체를 통해 찾게 되는 경우가 다수이다.

하지만 현재 사용자의 세부적 요구에 부합하는 특정한 카페를 추천해 줄 수 있는 추천시스템이 구축되어 있지 않다. 따라서 사용자의 목적에 맞는 카페를 추천할 수 있도록 하는 카페 추천시스템을 제안한다.

카페(커피전문점) 방문을 위한 꼭 필요한 요소 X 있으면 좋은 요소



[fig6. 카페 방문시 판단 항목 설문, 2020, OpenSurvey, 20~50대 표본 800인]

## 프로젝트 개요

1. 카페 데이터들을 네이버 지도로부터 크롤링 후, 데이터 내 평가지표를 활용하여 새로운 label을 새긴다.
2. 초기 데이터에 따른 지표가 좋은 카페들을 선정하고, 설정한 Labeling 기준에 따라 각 label별로 카페들을 분류한다. (조용한 분위기 카페, 분위기 좋은 카페, 메뉴 많은 카페 등등)
3. 사용자에게 이용한 카페들에 대한 리뷰를 요구해 데이터를 수집한다.
4. 새롭게 얻은 데이터들을 통해 시스템의 정확성을 높인다.
5. 사용자 맞춤 카페를 추천해주고 사용자의 데이터를 저장하여 추후 collaborative filtering에 활용한다.

< LITITCOFFEE

홈

메뉴

리뷰

사진

이런 점이 좋았어요 ②

✓ 55회 (47명 참여)

나도 참여

 "커피가 맛있어요"

26

 "집중하기 좋아요"

18

 "매장이 청결해요"

15

 "친절해요"

14

 "음료가 맛있어요"

10

 "인테리어가 멋져요"

10

 "디저트가 맛있어요"

9

 "가성비가 좋아요"

7

 "대화하기 좋아요"

7

 "뷰가 좋아요"

3

 "사진이 잘 나와요"

2

 "특별한 메뉴가 있어요"

2

 "화장실이 깨끗해요"

2

 "좌석이 편해요"

1

 "주차하기 편해요"

1

[fig7. 활용할 라벨링 데이터의 예시, 네이버 지도, 상도동 리릿커피]

---

## 사용할 툴/기술 소개

- 기본개발도구 python

크롤링을 위한 beautifulsoap, 데이터처리를 위한 pandas 학습을 위한 scikit-learn

데이터시각화를 위한 seaborn,matplotlib

- 학습 모델링

Classic statistics, K-means, PCA, Logistic Regression, KNN, SVM, Gaussin Naïve Bayes, Random Forest, Decision Tree, Gradient Boost, MLP, Stochastic Gradient Descent 등 가장 적합한 모델을 간추릴 것이다.

## 기존 혹은 연관된 접근법

KNN collaborative filtering.

content-based filtering

---