빅데이터 분석 결과 보고서

2021년 12월 03일.

과목 : 빅데이터

담당 : 허준 교수

학번 : 201744075

이름 : 김정욱

목차

1. **분석 개요**
   1. 분석 대상
   2. 분석 목적
   3. 활용 데이터
   4. 분석 수행절차
   5. 분석 결과 요약
2. **분석 절차**
   1. 활용 데이터 수집 대상 및 방법
   2. 활용 데이터 가공 방법
   3. 활용 데이터 분석 도구 및 방법
3. **분석 방법**
   1. 분석 도구
   2. 분석 방법
   3. 분석 알고리즘
   4. 분석 결과 시각화 및 설명
4. **시사점, 한계점 및 소감**
   1. 시사점
   2. 한계점
   3. 본 프로젝트 수행 후 소감

# 분석 개요

## 분석 대상

(이디야 커피 로고) (스타벅스 커피)

이 프로젝트의 분석 대상은 카페들입니다. 그 중 처음에는 서울의 이디야 커피와 스타벅스의 위도와 경도를 분석 대상으로 정했습니다. 분석 중 서울만 분석하기에는 표본이 적어 인천과 경기도를 포함해 수도권 전체의 이디야 커피와 스타벅스의 위도와 경도를 분석 대상으로 정했습니다.비교를 위해 메가커피, 공차, 할리스커피도 분석 대상으로 추가하였습니다.

## 분석 목적

<http://www.iconsumer.or.kr/news/articleView.html?idxno=10133>

- 스타벅스를 따라다닌 이디야’, 이디야 문창기 대표가 펼친 특별한 전략

이디야 커피는 주로 스타벅스 커피에 분점을 내는 스타벅스 옆자리 전략으로 성장했다는 유명한 이야기가 있습니다. 이 얘기를 들었을 당시에는 스타벅스가 가장 큰 카페 프렌차이즈니까 상업적으로 좋은 위치에 분점을 내기도 하고, 스타벅스보다 이디야 커피가 가격이 더 싸기 때문에 그럴듯하다고 생각했었습니다. 그러나 이디야 커피가 정말 옆자리 전략을 사용했는지 알아보고 싶었습니다. 그래서 이디야 커피와 스타벅스 사이의 거리를 분석하고 옆자리 전략이 효과가 있는지 알아내는 것을 분석 목적으로 정했습니다.

## 활용 데이터

<https://www.data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15083033>

- 소상공인시장진흥공단\_상가(상권)정보

텍스트이(가) 표시된 사진

자동 생성된 설명  
공공데이터포털에서 소상공인시장진흥공단이 제공한 상가(상권)정보에 지역별로 카페들의 위도, 경도 데이터가 포함되어 있어서 이 csv데이터를 활용하였습니다.

수도권이 분석 대상이라 전 지역 중 서울, 경기, 인천 3개를 사용했습니다.

## 분석 수행절차

활용할 데이터 수집 -> 분석할 카페 결정 -> 스타벅스와 가까운 거리 기준 결정 ->

각 카페 별 스타벅스와 기준 거리 안에 있는 지점 수 확인 ->

각 카페 별 전체 지점에서 스타벅스가 근처에 있는 지점 비율 계산

## 분석 결과 요약

True : 300m내에 스타벅스가 있는 지점, False : 300m내에 스타벅스가 없는 지점

**서울**

**테이블이(가) 표시된 사진

자동 생성된 설명**

**경기**

**테이블이(가) 표시된 사진

자동 생성된 설명**

**인천**

**텍스트, 테이블이(가) 표시된 사진

자동 생성된 설명**

**수도권**

테이블이(가) 표시된 사진

자동 생성된 설명

# 분석 절차

## 활용 데이터 수집 대상 및 방법

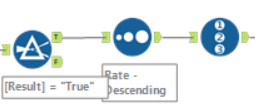
이 프로젝트의 수집 대상 데이터는 서울, 경기, 인천의 모든 스타벅스와 이디야, 메가커피, 공차, 할리스 커피의 위치입니다. 그래서 정확한 위도와 경도가 나와있는 데이터를 수집하기 위해 공공데이터포털에서 이 프로젝트에 활용할 만한 데이터를 검색했고, 결과적으로 소상공인진흥공단에서 제공한 상가정보에서 분석 대상 카페들만 필터링하여 수집하였습니다.

## 활용 데이터 가공 방법

Alteryx에서 위도와 경도를 받아 Point를 만드는 Create Point블록을 사용했습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명만들어진 Point를 Report Map블록을 사용해 지도로 나타냈습니다.

Formula로 구한 Result가 True인 행들의 비율을 높은 순서대로

정렬하여 순위를 매겼습니다.

## 활용 데이터 분석 도구 및 방법

스타벅스와 해당 카페의 데이터를 Affend Fields해서 하나의 이디야 커피에 모든

스타벅스를 비교하도록 Join했습니다.

Distance블록을 사용해서 스타벅스와 해당 카페의 Point 사이 거리를 구했습니다.

Summarize를 사용하여 각 이디야 커피별로 Group By를 해 Distance의 최소값만을

나타냈습니다.

테이블이(가) 표시된 사진

자동 생성된 설명

스타벅스가 근처에 있는지 없는지 나타내는 Result를 Group By하여 각 고유번호인

상가업소번호를 Count해 Result별 지점 개수를 구했습니다.

테이블이(가) 표시된 사진

자동 생성된 설명

각 고유번호인 상가업소번호를 Count해 전체 지점 개수를 구했습니다.



텍스트이(가) 표시된 사진

자동 생성된 설명가까운 거리의 기준을 300m로 잡고 스타벅스와의 300m보다 작거나 같으면 True, 크면

False값을 가지는 Result열을 생성했습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

Summarize로 구한 각 카페의 True, False에 해당하는 개수에 전체 지점의 개수를 나누고

100을 곱한 후 “%”를 붙혀 비율을 구했습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

# 분석 방법

## 분석 도구



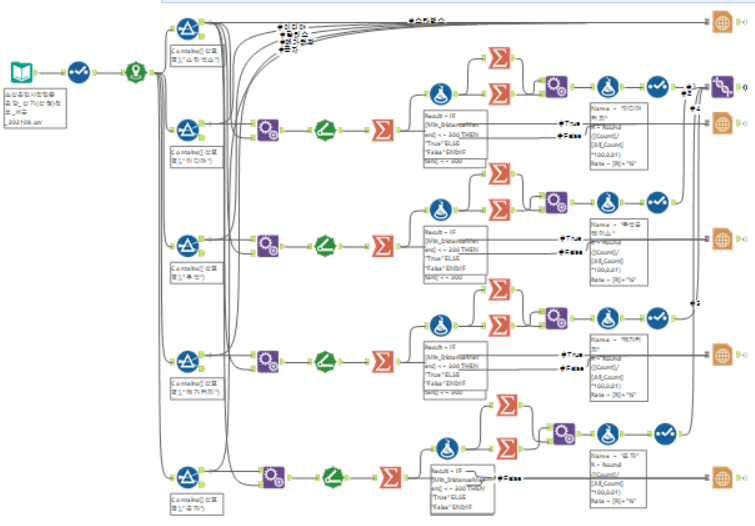
Alteryx Designer 2021.3 x64를 사용하여 분석했습니다.

## 분석 방법

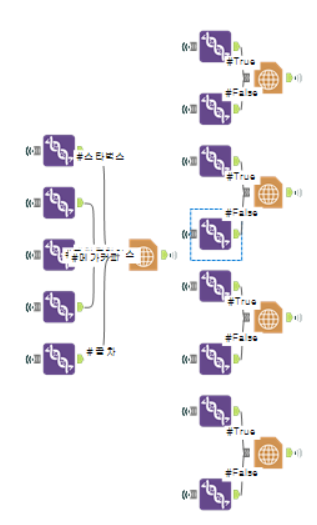
Map을 사용하여 전체 카페들을 시각화하고, 300미터 내에 스타벅스가 있는 카페와 없는 카페를 구분하여 시각화했습니다. 그리고 스타벅스가 있는 카페와 없는 카페 수를 구해 비율을 계산했습니다.

## 분석 알고리즘

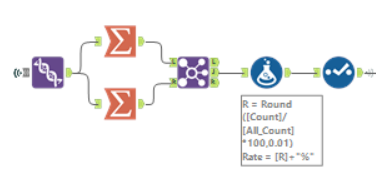
**Row data 가공**



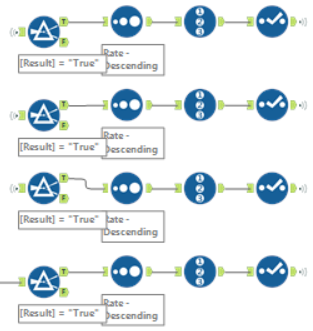
**서울/경기/인천 데이터 전체/카페별 지도로 통합**



**서울/경기/인천 결과 테이블 수도권으로 통합**



**서울/경기/인천/수도권별 비율 순위 생성**

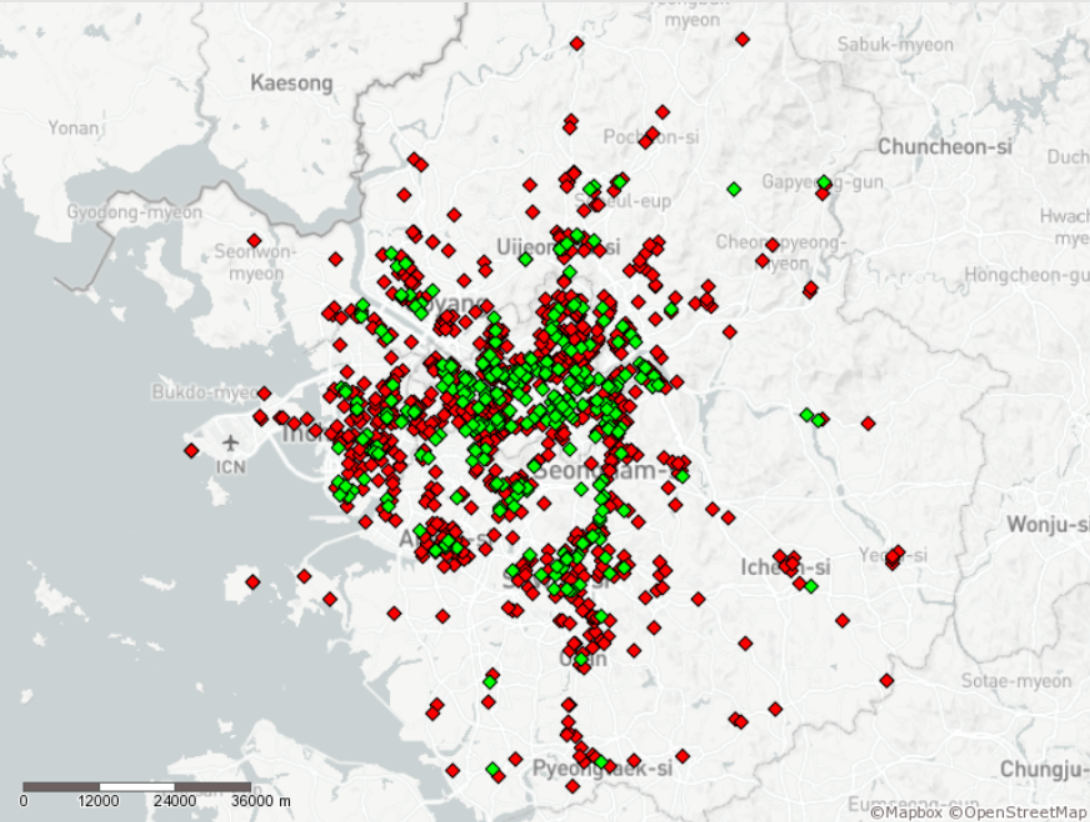


**지도, 분석 결과 테이블, 순위 Browse**



## 분석 결과 시각화 및 설명

**이디야 커피 투썸플레이스**

지도이(가) 표시된 사진

자동 생성된 설명

**메가커피 공차**

지도이(가) 표시된 사진

자동 생성된 설명지도이(가) 표시된 사진

자동 생성된 설명

예상과는 다르게 이디야 커피는 다른 카페에 비해 스타벅스가 근처에 없는 지점이 많았습니다. 주로 경기 외각지역에서는 거의 대부분의 지점이 스타벅스와 떨어져 있고, 인천 지역에도 스타벅스와 가까이 있는 지점이 많지 않았습니다. 그나마 서울에는 스타벅스 근처에 있는 지점이 많았지만 이 부분은 다른 카페에서도 나타났습니다.

# 시사점, 한계점 및 소감

## 시사점

결과에 비율을 보면 이디야 커피가 스타벅스 근처에 지점을 낸다는 것은 틀리다고 보여집니다. 하지만 기사에서도 보시다시피 이디야커피 대표가 직접 스타벅스 옆자리 전략 이야기를 하였고, 스타벅스 옆자리 전략을 사용한 것으로 생각됩니다.

그렇다면 왜 이디야 커피는 다른 카페에 비해 서울, 경기, 인천 그리고 다 통합한 수도권에서 꼴지의 비율이 나왔을까요? 저는 지점 수의 차이로 인해 생긴 결과라고 보여집니다. 이디야 커피는 스타벅스를 제외한 다른 카페에 비해 지점수가 제일 많습니다. 그래서 초기에 이디야 커피의 지점이 많지 않을 시기에는 특히 서울지역에서 이디야 커피 대표의 말처럼 스타벅스 근처에 있는 지점의 비율이 많았을 것이라 생각됩니다. 현재 분석한 네 카페 중 분점이 제일 적은 공차의 경우 수도권 전체 69.26%의 비율을 보이며 다른 카페에 비해 가장 많은 스타벅스 근처 지점을 갖고 있습니다. 발달된 상권에는 이미 스타벅스가 상주해 있을 것이고, 활발히 분점을 내야 하는 초기에는 발달된 상권에 분점을 주로 내기 때문에 분점이 적을수록 근처에 스타벅스가 있게 될 확률이 더 높다고 생각됩니다. 그러나 지점이 늘어나다 보면 매우 발달한 상권에는 이미 분점을 다 냈고, 그렇게 분점 수가 많아지면 이미 활발히 발달된 상권이 아닌 다른 지역도 노리면서 분점을 내게 되니 분점이 많아지면서 자연스레 비율이 줄어든 것으로 보여집니다.

그리고 서울이 아닌 지역에서는 스타벅스 옆자리 전략이 도움이 되지 않을 수 있다고 생각되고, 그래서 경기 외각지역의 이디야 커피는 거의 대부분이 스타벅스와 거리가 있도록 지점을 두었다고 생각합니다. 다른 카페들은 아직 서울 안에서도 분점을 내고 노릴 지역이 많지만 이디야 커피는 서울에 이미 많은 곳에 분점을 냈기 때문에 다른 지역에서도 분점을 늘리려고 할 것입니다. 스타벅스 옆자리 전략은 가격적인 면도 있겠지만, 서울은 유동인구가 우리나라에서 가장 많은 곳이고 그로 인해 스타벅스에 자리가 없어 밀리게 되면서 근처에 있는 다른 카페를 가려는 사람들을 잡기 위해서도 사용하는 전략이라고 생각합니다. 그런데 경기 그중에서도 외각지역은 유동인구가 서울과는 급격히 차이가 나고, 스타벅스에 자리가 꽉 차는 일도 드물다 보니 오히려 스타벅스 그리고 다른 카페와 거리를 두어 그 지역의 유일한 카페가 되는 것이 더 전략적으로 이득이라고 분석되었기 때문에 근처에 스타벅스 커피가 없는 것으로 보여집니다.

## 한계점

기간이 길지 않고, 시험기간이다 보니 시간적 여유가 없어 전국이 아닌 서울, 경기, 인천 지역만 분석하게 되었고, 카페도 4개밖에 비교하지 못하다 보니 스타벅스가 근처에 있는 지점 비율과 총 지점과의 관계점을 찾기엔 표본이 부족했다고 생각합니다.

## 본 프로젝트 수행 후 소감

Alteryx를 사용해보고 나니 데이터를 분석하는 과정이 생각보다 간단했고, 지도에도 나타낼 수 있어서 시각적으로 분석 결과를 보기에도 좋았습니다.

내가 평소에 궁금했던 점이 사실인지 분석하는 부분에서도 흥미를 느꼈고, 분석 결과에 대해 왜 이렇게 결과가 나왔는지 생각해보는 것도 좋았습니다.

본인은 2021년 2학기 빅데이터 과목의 빅데이터 분석 결과보고서를 제출합니다.