4차 6기 세미프로젝트

|  |
| --- |
| 날씨 영향에 따른 백화점 유인 효과 분석 |

**2019년 12월 20일**

**서비스 산업 데이터를 활요한 머신러닝 분석 과정(A)**

**HAPPY BEAN**

신승현

배윤성

이종현

**목 차**

**1. 프로젝트 개요**  **1**

1.1 프로젝트 기획 배경 및 목표 1

1.2 구성원 및 역할 2

1.3 프로젝트 추진 일정 3

**2. 프로젝트 결과**  **7**

2.1 데이터 수집 7

2.2 데이터 분석 8

2.3 데이터 분석 결과 9

**3. 기대 효과**  **10**

3.1 향후 개선 사항 10

3.2 기대 효과 11

**4. 개발 후기**  **12**

**1. 프로젝트 개요**

1.1 프로젝트 기획 배경 및 목표

국내 극심한 미세먼지는 오프라인 유통업계 매출을 증가와 깊은 상관관계를 보이고 있다. 실내 복합쇼핑몰을 대표하는 백화점의 경우, 미세먼지 기승이 심한 매년 3월 매출과 방문객이 급증하는 양상을 보여주었다. 금년 롯데백화점의 경우, 3월에 발령된 미세먼지 비상저감 조치 기간 5일 동안 전국지점의 매출이 9.1% 상승하였고 구매 고객은 18.8% 증가하였다. 본 프로젝트에서는 미세먼지를 포함한 기상 조건의 영향이 백화점을 방문하는 고객의 수에 어떠한 영향을 미치는지 분석했다.

데이터는 서울시가 공개한 데이터를 가장 중요하게 사용했다. KT 통신사의 LTE 사용 고객의 신호를 통해서 1시간 단위마다의 생활인구 집계 자료를 통해, 날짜와 시간에 따른 백화점 주변 실기간 인구를 파악했다. 또한, 서울시의 생활인구 데이터, 기상청과 서울시에서 제공한 기상, 대기질 데이터를 이용하여 백화점과 기상과의 상관성을 알아내고자 했다. 기상 요인에 따른 백화점 방문객 수와 고객 세그먼트 분석은 백화점의 전략적인 의사결정에 도움을 줄 것이다.

1.2 구성원 및 역할

|  |  |  |  |
| --- | --- | --- | --- |
| 이름 | 전공 | 역할 | 구현 부분 |
| 이종현 | 인지과학 | 팀장 | 프로젝트 관리 |
| 신승현 | 국문 & 경영 | 팀원 | 데이터 수집  데이터 전처리 |
| 배윤성 | 교정보호 | 팀원 | 시나리오 작성  지도 시각화 |
| 강자연 | 경제 | 팀원 | 데이터 수집  데이터 전처리 |

1.3 프로젝트 추진 일정

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | 기간 | 활동 | 비고 |
| 사전  기획 |  | 프로젝트 기획 및 팀 구성 |  |
|  | PJT주제 선정, 팀(PM/팀원) 구성 | 3~5 인/팀 |
|  | 프로젝트 멘토링  [프로젝트 방향 설정 및 현업프로젝트 소개] | 현업 멘토 참여 |
| PJT  수행  /  완료 |  | 프로젝트 수행 |  |
|  | 프로젝트 설계 |  |
|  | 프로젝트 멘토링  [프로젝트 점검 및 기술자문 ] | 현업 멘토 참여 |
|  | 구현 및 테스트 |  |
|  | 팀별 최종 발표 (구축 완료 보고) | 최우수 한 팀 선발  멘토 평가 |

**2. 프로젝트 개발 결과**

2.1 데이터 수집

1. 서울열린데이터광장
   1. 서울특별시 일별 평균 대기오염도 정보 (미세먼지 및 초미세먼지)[[1]](#footnote-1)
      1. 2017년 1월 1일 ~ 2019년 11월 30일
   2. 서울시 강수량 및 강우일수 통계 (강수여부, 강수량)[[2]](#footnote-2)
      1. 2017년 1월 1일 ~ 2019년 11월 30일
   3. KT 기지국 신호 기반 내국인 생활인구 데이터[[3]](#footnote-3)
      1. 2017년 1월 1일 ~ 2019년 11월 30일
2. 통계청 통계지리정보서비스
   1. 행정구역 집계구역[[4]](#footnote-4)
      1. 2016년 기준
3. 국가공간정보포털
   1. 전수집계구 경계[[5]](#footnote-5)
      1. 2016년 기준
4. QGIS & Vworld open API
   1. 오픈 소스 지리 정보 시스템[[6]](#footnote-6)
   2. 국가 공간 정보[[7]](#footnote-7)
5. 백화점 소재지
   1. 주요 백화점 업체 3사
   2. 롯데 (13), 현대(10), 신세계 (2)
      1. 2019년 11월 30일 기준

2.2 데이터 분석

1. 백화점 소재지 파악

정확한 백화점 소재지 파악을 통한 집계인구 분석은 프로젝트에서 가장 중요한 과정이다. 서울시에서 제공하는 생활인구 데이터는 각 공간별, 일자별, 시간별 생활인구 정보가 제공된다. 공간별 생활인구 데이터는 2가지의 방법으로 집계되었다. 일반적인 ‘행정동’ 구역에 따른 집계방법과 통계청에서 사용하는 집계구역에 따른 방법이 존재했다. 프로젝트의 분석에서는 집계구역에 따른 방식으로 인구를 집계하였다. 그 이유는 다음과 같다. 먼저 행정동 구역은 서울시를 기준으로 구별 24개소, 동별 424개소로 구분되어 있다. 반면 집계구역은 19,500개소로 서울시를 구분되어, 백화점 내 인구를 더 세밀하게 파악할 수 있었다. 또한 집계구역은 실 거주인구를 기준으로 약 거주인구 500명당 하나의 집계구역을 배정한다. 따라서 해당 구역 거주민인지 혹은 외부 방문인구인지를 결정함에 있어서 더욱 정확성이 높았다.

백화점이 속한 집계구역을 파악을 위해 Google의 geo-coding API를 이용하여 백화점의 주소를 경도와 위도로 변환했다. 또한 SHP 포맷으로 제시된 서울특별시의 집계구역 지도 역시 기존 중부원점 기준에서 경, 위도 기준으로 변환하였다. 이를 바탕으로 각각의 집계구역의 중심 경도, 위도를 추출하였다. 백화점과 집계구역 중심과의 최소거리를 구하고, 이 수치를 바탕으로, 총 25개 백화점이 각각 소속된 집계구역을 정했다. 이후, 실제 백화점이 올바른 집계구역으로 배정되었는지 사후검정을 실시하였다. 25개의 해당 집계구역을 찾는 과정에서 집계구역 내 역이나 공항과 같은 중요 대중교통 시설이 존재하는 경우, 분석대상에서 제외했다.

1. 유동인구 및 거주인구 정의

생활인구 데이터는 거주인구와 유동인구가 분리되어 있지 않다. 생활인구 데이터는 해당 집계구역에서 발생한 LTE 신호와 기지국의 위치를 통해서 0시부터 23시까지 매시 한번 생활인구를 추정하는 자료이다. 따라서 우리 프로젝트에서는 유동인구와 거주인구를 다음과 같이 정의하였다.

*거주인구* : 화요일, 수요일, 목요일 /

02시 ~ 05시에 백화점 포함 집계구역에서 발생한 생활인구 (1년) 평균

*유동인구* : 토요일, 일요일 /

백화점 개장 시간 (10시 ~ 22시)에 백화점 포함 집계구역에서 발생한 생활인구

– 해당 집계구역 거주인구

거주 인구의 경우, 이사 등의 요인으로 변동될 가능성이 있기 때문에 1년 단위로 평균을 갱신했다. 유동인구는 해당 집계구역의 거주 인구를 제외하는 방법으로 순수 유입만을 유동인구로 특정하였다. 또한 백화점의 개장 시간에 발생하는 유동 인구를 파악하기 위해서 백화점 개장 시간으로 제한하였다.

1. 날씨 효과 정의

백화점 방문에 영향을 미치는 날씨의 요인을 좋은 날씨와 악천후의 방법으로 binary coding했다. 기상 조건은 기온, 강수 여부, 미세먼지로 특정했다.

기온의 경우 …

강수 여부의 경우, …

미세먼지는 미세먼지와 초미세먼지 두 종류로 구분되어 있다. 일반적으로 미세먼지, 초미세먼지 심각성은 기존에 정의된 기준치와 비교하여 새로이 등급을 매기었다. 특히 본 프로젝트에서는 WHO 기준을 참조했다. WHO 등급기준은 미세먼지, 초미세먼지의 심각성에 따라 8단계로 구분되어 있다. 미세먼지와 초미세먼지를 하나의 기준으로 정의하기 위해, 미세먼지의 등급과 초미세먼지의 등급 중 하나의 기준이라도 4등급을 초과하는 경우를 ‘미세먼지 기상 조건 나쁨’으로 정의하였다. 이는 미세먼지 혹은 초미세먼지 하나의 기준이 나쁨 이상에 해당되는 경우를 의미한다.

1. 통계 모델링

기상 조건에 따른 백화점 유동 인구의 변화를 살피기 위해서 각각의 백화점이 소재한 집계구역별로 다중 회귀 분석을 진행하였다.

1. 시각화

2.3 데이터 분석 결과

..... [본문 : 10pt / 맑은 고딕]

**3. 기대 효과 [대제목 : 16pt / 맑은 고딕]**

3.1 향후 개선 사항 [소제목 : 12pt / 맑은 고딕]

..... [본문 : 10pt / 맑은 고딕]

3.2 기대 효과..... [본문 : 10pt / 맑은 고딕]

**4. 개발 후기**

|  |
| --- |
| 팀 사진 첨부 |

|  |  |
| --- | --- |
| 성명 | 후기 |
|  |  |
|  |  |
|  |  |
|  |  |

1. https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-2218&srvType=S&serviceKind=1&currentPageNo=1 [↑](#footnote-ref-1)
2. https://data.seoul.go.kr/dataList/datasetView.do?infId=10672&srvType=S&serviceKind=2 [↑](#footnote-ref-2)
3. https://data.seoul.go.kr/dataList/datasetView.do?infId=OA--4979&srvType=F&serviceKind=1&currentPageNo=1 [↑](#footnote-ref-3)
4. https://sgis.kostat.go.kr/contents/shortcut/shortcut\_05.jsp [↑](#footnote-ref-4)
5. http://data.nsdi.go.kr/dataset/20171206ds00004 [↑](#footnote-ref-5)
6. https://qgis.org/ko/site/ [↑](#footnote-ref-6)
7. http://www.vworld.kr/dev/v4api.do [↑](#footnote-ref-7)