**산학공동프로젝트 연구원 주간 업무일지**

| 학과 | 컴퓨터공학과 | 학번 | 2016253046 | 성명 | 정두용 |
| --- | --- | --- | --- | --- | --- |

|  |  |
| --- | --- |
| 프로젝트명 | 코인세탁소 신규 창업시 위치기반 입점 추천 분류기 제작 |
| 참여업체 | (주) 에스크린 |
|  |  |

사용 API 선정

**1.1 Jupyter notebook**

- ‘주피터 노트북’은 오픈소스 기반의 웹 애플리케이션 으로, 파이썬을 비롯하여 40여개의 프로그래밍 언어로 코드를 작성하고 실행하는데 개발환경을 제공해준다.

- 특히 메모리 버퍼에 배열 데이터를 저장하고 처리하는 효율적인 인터페이스를 제공한다.

**1.2 Numpy**

- 파이썬 기반 데이터 환경에서 Numpy 는 행렬 연산을 위해 핵심 라이브러리입니다. Numpy 는 Numerical Python 의 약자로 대규 다차원 배열과 행렬 연산에 필요한 다양한 함수를 제공한다.

특히 메모리 버퍼에 배열 데이터를 저장하고 처리하는 효율적인 인터페이스를 제공한다. 파이썬 list 객체를 개선한 numpy nd array 객체를 사용하면 더 많은 데이터를 더 빠르게 처리할 수 있습니다.

**1.3 Folium plugin**

- python에서 제공해주는 라이브러리로서 지도를 다루는 대표적인 라이브러리다.

- 이것을 사용하기 위해서는 위도 경도를 알아야한다.

- Map 이라는 함수에서 location 으로 처음 화면이 나오는 부분을 설정해준다.

- folium 은 다양한 형식의 마커 특정위치를 표시하는 표식과 마커를 클릭했을 때 나타나는 정보를 지정할 수 있다

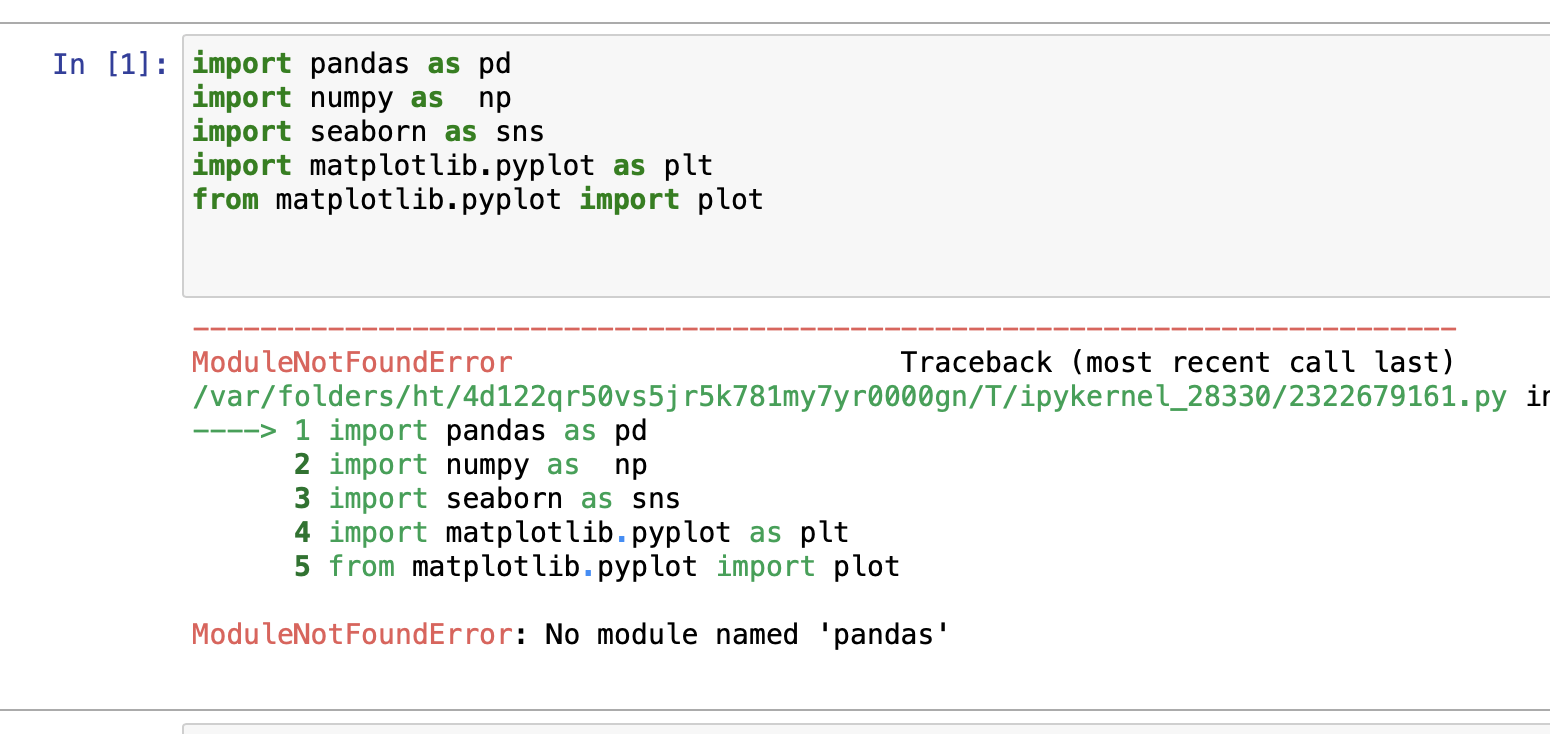
**1.4 Tensorflow**

- 구글에서 오픈소스로 공개한 기계학습 라이브러리이다.

- 여러 언어를 지원한다.

- Tensorflow에 있는 Sequential 모델 사용

- 여러 레이어를 간단하게 추가하고 컴파일, 학습 시킬 수 있다.



데이터셋 구축

2.0 데이터셋을 준비하기 전에 데이터셋에 넣을 특성들을 먼저 정해야 한다.

참여기업은 여타의 코인 세탁소와 공통점과 차별점이 존재한다.

공통점: 비대면 의류 세탁서비스 제공

차이점: 드라이클리닝과 같은 더 다양한 의류 관리 서비스를 제공

비대면 조작을 위해 코인세탁소보다 비교적 사용하기 어려운 키오스크를

이용해야 함

-> 이러한 특징들을 통해 참여기업에 매출에 관련될 수 있는 특성들을 파악해본다.

2.1 데이터셋을 구성할 특성들

설문조사에서 비대면을 선호하는 연령대는 20대 30대가 다른 연령대보다 높게

관찰되고 있다.

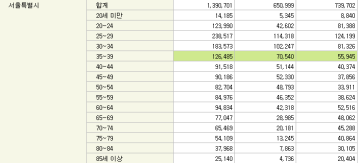
또한 무인 매장이용 의향도 20대가 61.7%로 압도적으로 높게 조사되었다.

따라서 입점할 위치에 주변 **활동 인구수의 연령대**가 좋은 특성이 될 수 있겠다.

(출처: 메조미디어, 2020유통업종분석리포트)

세탁기와 건조기를 모두 구비하기 힘들고 1인가구의 연령대가 20대 30대의

비율이 높기에 **가구당 사람 수**도 특성에 추가했다.



<그림2. 1인가구의 연령대 자료>

출처:https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\_1PL1502&conn\_path=I2

세탁서비스를 제공하는 기업이므로 주변 **가구의 평균소득** 또한 좋은 특성이

될 수있다.

비대면 세탁소이기 때문에 지출되는 비용의 대부분을 매장임대료가 차지한다.

따라서 안정적인 매출 을 위해서 지역평균 **평균 임대료**도 특성으로 추가했다.

세탁소와 기업의 특성과는 상관없이 기본적으로 매출과 관련이 큰 특성으로

**지역 인구수**를 추가했다.

5개의 특성으로 **활동 인구수의 연령대, 가구당 사람 수, 가구의 평균소득, 임대료, 지역 인구수** 구성했다.

분류기를 학습시킬 타겟 데이터로는 매출로 설정해 준다. 이렇게 예측된 매출을

평균임대료로 나눠 다시 추천 등급으로 분류한다.

2.2 설정한 특성으로 데이터셋 구성

구성할 특성들은 위치정보를 기반으로 찾을 수 있다.

따라서 세탁소의 위치정보를 코드화한 행정동코드와 법정동코드를 이용해 특성들의 데이터를 연결할 수 있다.

**법정동코드 행정동코드 연동**

0) 공공 데이터 포털에서 소상공인시장진흥공단\_상가(상권)정보.csv를 불러온다.

1) 상권업종소분류명을 세탁소/빨래방으로 설정하여 세탁소만을 구해준다.

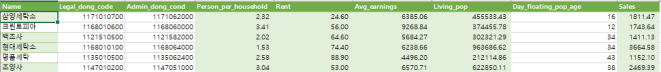
2) 위 데이터에서 상호명, 행정동코드와 법정동코드를 이용해 새로운 데이터를 만든다.



<그림 3 . 상호명 행정동코드 법정동코드>

2.3 행정동코드와 법정동코드를 이용해 여러 특성들을 연결시켜 데이터셋을 완성한다.

정했던 특성들 (**가구당 사람 수, 평균 임대료, 가구 평균 소득, 생활인구, 주간활동인구 연령대**)을 세탁소의 법정동, 행정동 코드에 연결하여 하나의 데이터를 완성한다.

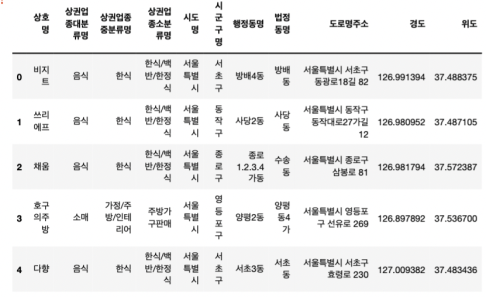


<그림 4 . 데이터셋 사진>

2.4 공공데이터를 통해 현재 서울시에 있는 세탁소들의 위치를 시각화 해서 보여 준다

0) 공공 데이터 포털에서 소상공인시장진흥공단\_상가(상권)정보.csv를 불러온다.

1) json 형태로 되어 있으므로 데이터에서 . 형태로 구분해준다. 2) 데이터에서 "상호명", "상권업종대분류명","상권업종중분류명","상권업종소분류명","시도명","시군구명","행정동명","법정동명","도로명주소","경도","위도 만 추출해서 column 형태로 넣는다.



<그림 5 . seoul\_open.csv 파일에서 추출한 사진>

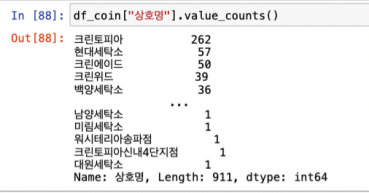


**3.코인 세탁소 데이터 추출**

3.1.상호명을 통한 코인 세탁소 및 경쟁업체 데이터 추출

-“월드크리닝|워시테리아|크린하우스|크린스피드|워시테리아|빨래방|세탁소|크린에이드|크린스피드|세탁을위한사람들|크린위드|테라스크린|크린토피아” 키워드 추출한다.

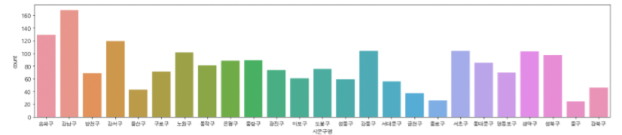
- "상권업종대분류명" ,에서만 “소매” 만 추출한다.

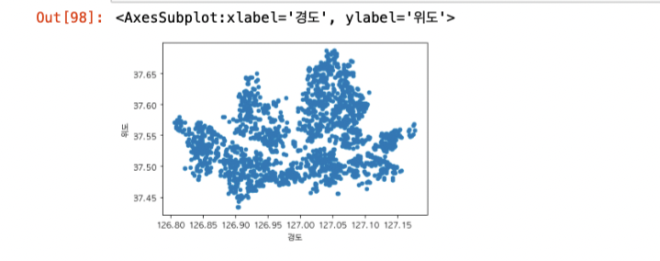


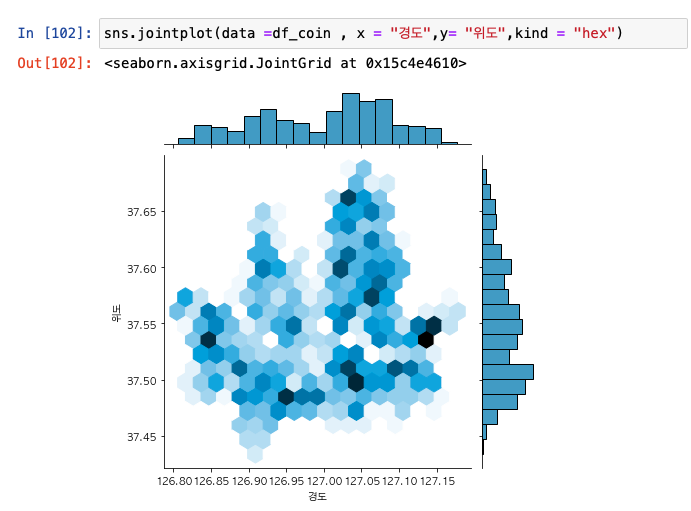
<그림 6. 세탁소 빈번하게 나타나는 횟수>

- 시군구명을 “시군구명” 따로 추출해서 시각화를 진행했다.

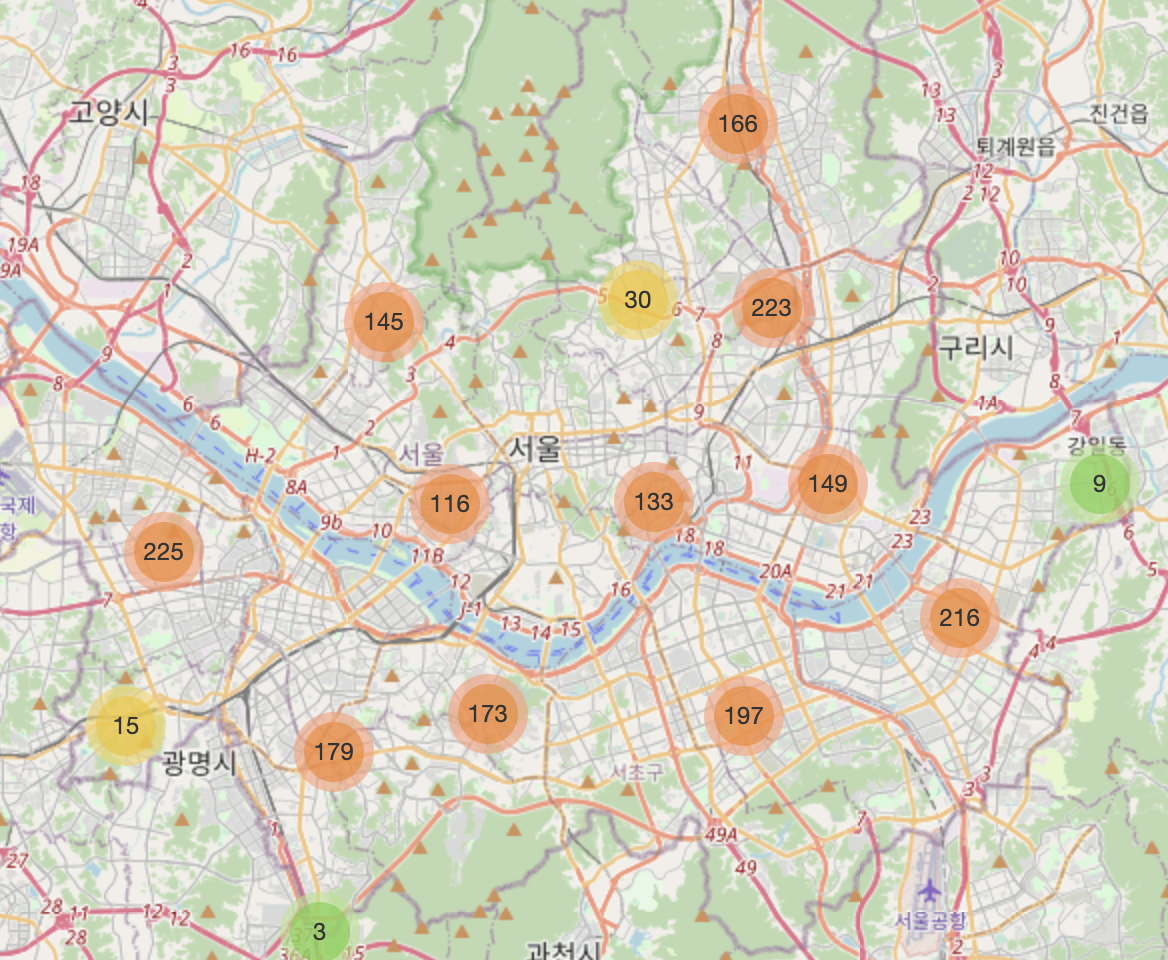
- 위도 경도를 서울시 지도에서 나타냈다.







지도데이터를 갖는 인터넷 사이트를 제작하였다.



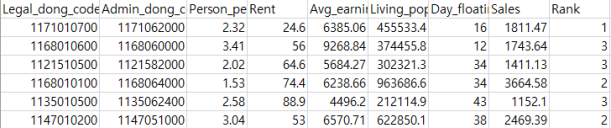
분류기 구축

- Colab환경에서 tensorflow를 import하여 분류기를 구축한다.

- 분류기는 5개의 특성을 통해 3개의 등급 중 하나의 등급을 출력해준다.

- 학습시킬 데이터셋의 전처리를 진행해준다.

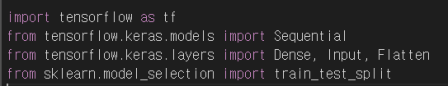
1) 엑셀파일로 되어있는 데이터를 한글명을 없애고 csv파일로 변환한다.



2) 학습데이터와 테스트데이터를 나눈다.

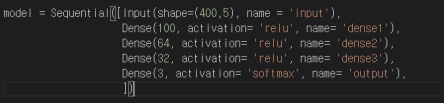
- Tensorflow에 Sequential Model을 이용해서 분류기를 구축한다.

1) 모듈 Import



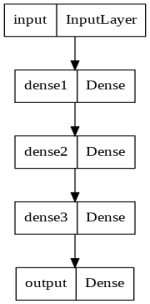
2) Sequential Model 생성

입력층, 은닉층, 출력층을 추가시켜 분류기를 구성한다.



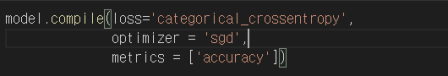
은닉층 3개의 노드 수를 100 64 32로 낮추며 활성화 함수는 relu함수를 사용한다.

마지막 출력층은 등급 3개를 분류하므로 3개의 출력과 분류를 위해 활성화 함수는 softmax를 이용 한다.

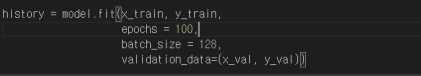


<그림7. 분류기 층별 구성도>

- 모델 컴파일 및 학습



loss함수는 categorical\_crossentropy를 사용하고 확률적 경사하강법을 사용한다.



반복횟수 100번 배치크기는 128로 모델을 학습시킨다. 검증세트도 추가해준다.

- 분류기의 특성을 예측해보면 특성별로

| **가구당 사람 수** | **평균 임대료** | **가구 평균 소득** | **생활인구** | **주간활동인구 연령대** |
| --- | --- | --- | --- | --- |
| 1에 가까울수록  높은 등급 | 낮을 때 높은 등급 | 높을수록 높은 등급 | 많을수록 높은 등급 | 많을수록 높은 등급 |

- 예비 창업자는 창업하려는 곳의 위 다섯 특성을 입력하면 매출이 예측되고 매출을 비용의 가장 큰 부분인 임대료를 나눠 추천 등급을 알 수 있다.

결과 보고서 작성 및 판넬 작성 및 피피티 발표

