

## 데이터 분석을 통해 알아보는 치킨집 매출과 인구의 특성의 관계

1669063 국제 이송

### 1. 주제: 인구의 특성과 치킨집 매출에는 관계가 있을까?

어느 동네를 가든, 또 누구든 좋아하는 치킨에 대해서 데이터분석을 해보고자 하였습니다. 사는 지역, 나이 등에 따라서 달라지는 치킨집 매출을 분석하고자 소주제를 2 개를 잡고 프로젝트를 진행했습니다. 각 소주제에 대하여 가설을 걸고, 그 가설이 맞는지 확인해보았습니다.

#### (1) 소주제 1: 지역별(구별) 인구와 치킨집 통화량의 관계를 파악

구체적으로는 서울시의 25 개 구별로 분석을 하여, 인구수가 많을 수록 치킨집의 통화량이 많다는 가설을 검증해보고자 합니다.

#### (2) 소주제 2: 연령별 인구와 치킨집 주문의 관계

구체적으로는 서울시의 10 대-60 대 연령대를 분석하여 연령대가 어릴 수록 치킨 주문량, 즉 통화량이 많다는 가설을 검증해보고자 합니다.

### 2. 사용한 데이터:

저는 인구의 특성에 따른 치킨집의 매출을 알아보하고자 하여 인구 데이터와 치킨집 데이터 두 가지를 사용하였습니다.

#### (1) SK 콘텐츠 허브의 19 년 9 월 서울시 치킨 판매업종 이용 통화량

소주제 1 을 위해서는 위 데이터에서 서울시 각 구별 통화량을 추출하여 매출의 지표로 사용하였습니다.

소주제 2 를 위해서는 위 데이터에 서울시 각 연령대의 통화량을 추출하였습니다

#### (2) 수업시간에 활용한 agv.csv 데이터

치킨 매출과 연령 데이터를 함께 사용하여 더 풍부한 분석을 하고자 하였습니다.

소주제 1에서는 서울시 구별 인구수를 사용하였고, 소주제 2에서는 서울시 연령별 인구수를 사용했습니다.

### 3. 프로그램 구조

소주제 1 과 소주제 2 는 큰 틀에서 같은 구조를 띄고 있습니다.

특히 #1. 치킨데이터 가공하기의 일부는 Function()을 만들어서 변수를 사용해 코드를 같이 사용하고 있습니다.

#### # 1. 치킨 데이터 가공하기:

- PANDAS 를 사용하여 '서울시 피킨 판매업종 이용 통화량' 데이터를 불러오기
- 변수 입력값에 따라 NUMPY 등을 사용하여 '시군구'열과 '전화량' 열을 추출하고, 이를 가지고 새 데이터프레임을 만들기
- 그리고 새로운 데이터프레임으로 '시군구'별 리스트와 '전화량' 리스트를 만들기

#### #2. 인구 데이터 추출하기:

- PANDAS 를 사용하여 age.csv 파일을 DF 로 불러오기
- 구별/연령별로 행열 슬라이싱하여 리스트로 저장

#### #3. 치킨집 전화량과 구별/연령별 인구수 비교 및 계산:

(소주제 1) -구별 인구수에 비해 전화량이 얼마나 되는지 비율을 구함: 25 개의 구별 전화량/구별 인구수를 구하여 1 명당 전화량이 얼마나 되는지 알아보고, 숫자를 더 쉽게 다루기 위해 10000 을 곱하여 처리하였다.

(소주제 2) -연령별 인구수에 비해 전화량이 얼마나 되는지 비율을 구함: 인구 데이터에서 연령별 인구수를(10 대, 20 대, 30 대, 40 대, 50 대, 60 대) 더하여 6 개의 값을 리스트에 넣고, 이걸 바탕으로 연령별 전화량/연령별 인구수 계산을 하여 각 연령대별로 1 명당 전화량이 얼마나 되는지 알아보았다, 숫자를 더 쉽게 다루기 위해 10000 을 곱하여 처리하였다.

#### #4. 그래프 그리기:

(소주제 1)-구별 데이터 분석에서 구별 인구수에 비한 구별 전화량이 각 구마다 비슷한지 plot 그래프로 표현함: 위에서 다룬 데이터와 그리고 그래프 바의 높이를 비교할  $y=a$  상수 그래프를 그려서 표현하였다.

(소주제 2)-연령별 데이터 분석에서는 연령대별 비율을 비교하기 위해 pie 그래프로 표현: 10 대, 20 대가 타 연령층에 비해 몇 퍼센트나 차지하는지를 보기 위해서 pie 그래프를 선택하였다.

#### 4. 프로그램 실행 방법

파이썬 파일을 작동시키면 1. 구별 인구수와 치킨집 주문의 관계 혹은 2. 연령대와 치킨집 주문의 관계를 고르라는 메뉴가 뜬다. 1 과 2 중 하나를 선택하면 그 번호에 해당하는 프로젝트를 실행하여 분석 과정을 보여주고 plot 혹은 pie 차트를 정리하여 보여준다.

#### 5. 사용한 라이브러리

-Matplotlib

1. 구별 프로젝트를 표현하기 위해 plot 그래프 사용하였습니다. (line 103)
2. 연령별 프로젝트를 표현하기 위해 pie 그래프 사용하였습니다. (line 170)

-Numpy

1. 치킨 데이터에서 구별, 연령별 데이터를 얻기 위해 중복된 성분을 제외한 특정 값을 추출해주는 np.unique()함수 사용하였습니다.(line 37)
2. np.sum()함수를 np.unique()와 함께 사용하여 추출한 특정값이 반복되는 횟수를 저장하여 이를 가지고 새로운 데이터프레임을 만들었습니다. (line 40)
3. np.round()함수 사용하여 소수점 자리가 너무 긴 숫자들을 올림하였습니다. (line 164)

-Pandas

1. 두 csv 데이터를 불러올 때 모두 판다스로 불러왔습니다. (line 33, 69)
2. df.iloc, loc, df 등 데이터 슬라이싱하기 (line 58, 59, 73, 76, 83, 133)
3. DataFrame 만들기: Function()에서 여러 처리 후 필요한 요소만 모아서 볼 수 있도록 데이터프레임을 새로 만들었습니다.(line 42)
4. del df 로 필요 없는 열을 삭제하였습니다. (line 134)
5. df.reset\_index()로 새로운 인덱스를 만들어서 기존의 index 를 행값으로 가져왔다. (line 77)
6. df.head()로 df 을 5 줄까지만 보여준다. (line 44, 79)

## 6. 추가한 함수 및 코드들

- (1) Np.unique()함수는 중복된 성분을 제외하고 유니크한 값만 반환받아 사용하는데, 데이터에서 줄마다 읽으면서 유니크한 '시군구'값을 저장하는데 사용되었다.
- (2) np.round() 함수는 소숫점자리가 너무 길어서 리스트를 반올림하기 위해서 사용되었다.
- (3) Plt.axhline(y=b, color='brown') 코드로 plot 그래프에 y=상수 그래프를 그렸습니다. 이는 각 구별 높이를 더 잘 비교할 수 있도록 하기 위해서 넣은 코드이다.
- (4) 기존의 matplotlib 에서 한글 깨짐을 방지해주는 코드를 찾지 못해서 구글링해서 이 코드를 사용하였다.

```
import matplotlib.font_manager as fm
```

```
font_name =
```

```
fm.FontProperties(fname='C:\Users\WW\Downloads\NanumBarunGothic.ttf',  
size=50).get_name()
```

```
plt.rc('font', family=font_name)
```

(5) `df.reset_index()`로 새로운 인덱스를 추가한다. 기존의 인덱스값을 뽑아서 쓰고싶었기 때문에 이를 활용했다.

## 7. 실행결과

(1) 소주제 1: 가설은 구별 인구수와 치킨 주문량이 비례하므로, plot 그래프가 모두 비슷한 높이로 나왔어야 했습니다. 하지만 그래프에서는 구별로 큰 차이가 있음이 나타났습니다. 빨간색으로  $y=a$  상수 그래프를 그려서 더 비교가 용이하도록 하였습니다. 그리하여 강남구와 강서구는 매우 수치가 높은 반면, 도봉구, 동작구, 종로구 등은 매우 낮게 나옴을 알 수 있습니다. 이를 통하여 인구수와 치킨 주문량이 꼭 비례하는 것이 아님을 알 수 있고, 또 더 자세한 분석을 위해서는 구별 특성(구별 연령별 분포, 구민 평균 소득)등을 더 고려해보아야 함을 깨달을 수 있었습니다.

```
C:\Users\user\PycharmProjects\PAD\Scripts\python.exe C:/Users/user/PycharmProjects/PAD/Chick66.py
1669063  국제 이슬
인구와 치린집 주문량의 관계를 파악하는 프로젝트입니다.
어떤 메뉴를 선택하시겠습니까?

1. 인구수(age data)와 치린집 주문의 상관관계
2. 연령대와 치린집 주문의 상관관계: 1

1. 인구수와 치린집 주문의 상관관계를 실행하겠습니다

# 1. 치린 데이터 가공하기
*치린 데이터에서 시군구과 전화량 열로 만든 데이터프레임
시군구 전화량
0 강남구 2149
1 강동구 1099
2 강북구 851
3 강서구 1600
4 관악구 530

*인구 데이터에서 추출한 개별 인구수 리스트
['과남구', '과동구', '과북구', '과서구', '관악구', '관전구', '구로구', '금천구', '노원구', '도봉구', '동대문구', '동작구', '마포구', '서대문구', '서초구', '성동구', '성북구', '송파구', '양천구',
```

\*인구 데이터에서 추출한 전화량 리스트

[2149, 1099, 851, 1600, 530, 418, 1081, 653, 878, 298, 827, 553, 1193, 1324, 1429, 818, 986, 1035, 646, 1254, 1069, 749, 656, 1362, 659]

#2. 인구 데이터 활용하기

\*인구 데이터에서 추출한 구별 인구수 데이터프레임

index	행정구역	2019년02월_계_총인구수
0	1 서울특별시 종로구	152880.0
1	19 서울특별시 중구 (1114000000)	125995.0
2	35 서울특별시 용산구 (1117000000)	229279.0
3	52 서울특별시 성동구 (1120000000)	307064.0
4	70 서울특별시 광진구 (1121500000)	355387.0

\*인구 데이터에서 추출한 구별 총 인구수 리스트

[152880, 125995, 229279, 307064, 355387, 347459, 402616, 435754, 318268, 338373, 541800, 484070, 311700, 375372, 463103, 596602, 403668, 234157, 368437, 398685,

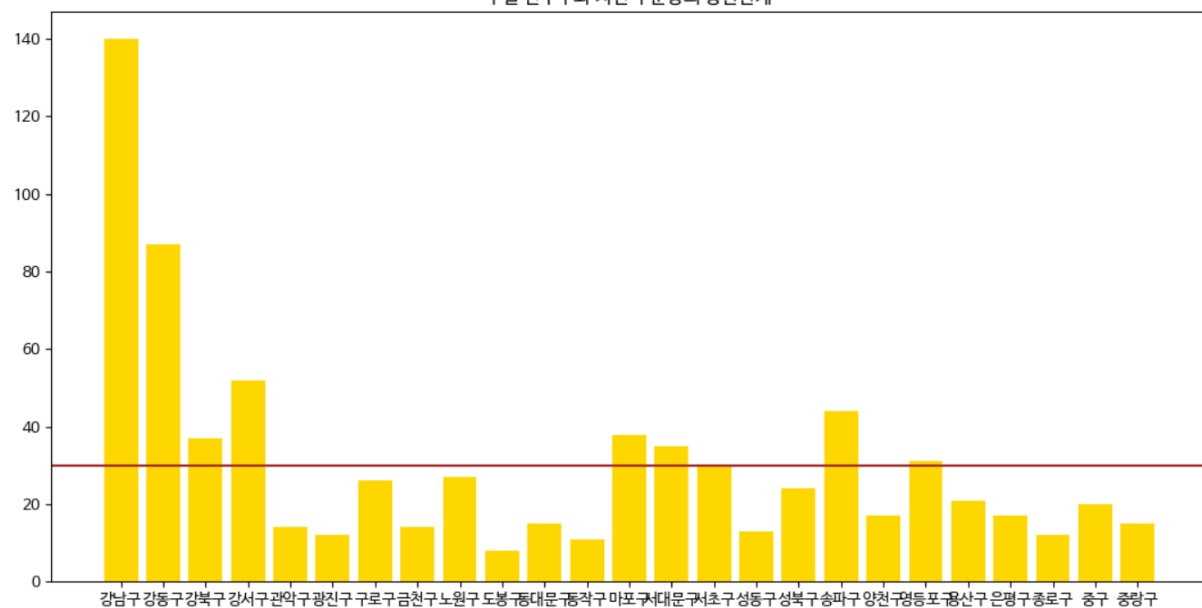
#3. 치킨 전화량과 구별 인구수 비교 및 계산

\*총 인구수에 따른 치킨 주문 비율 리스트

[140, 87, 37, 52, 14, 12, 26, 14, 27, 8, 15, 11, 38, 35, 30, 13, 24, 44, 17, 31, 21, 17, 12, 20, 15]

#4. 치킨 전화량과 구별 인구수의 plot그래프 그리기

구별 인구수와 치킨 주문량의 상관관계



#4. 치킨 전화량과 구별 인구수의 plot그래프 그리기

결론: 치킨 주문 횟수는 인구수에 비례하지 않는다.

만약 주문 횟수가 인구수에 비례한다면 바 그래프가 모두 같은 높이를 취해야 한다.  
하지만 높이가 차이가 있는 것으로 보아 주문 횟수는 인구수에 비례하지 않는다고 할 수 있다.

Process finished with exit code 0

(2)소주제 2: 가설은 연령이 낮을 수록 치킨 주문량이 많다는 것이었습니다. pie 그래프로 10 대-60 대의 비율을 비교해보았을 때, 10, 20, 30 대 순으로 비율이 가장 높음을 알 수 있었습니다. 이를 통해 가설이 옳았음을 알 수 있었습니다. 특히나 10 대가 과반수 이상을 차지하고 있고, 10 대와 20 대가 2/3 을 차지하고 있습니다. 이를 통해서 추후 치킨 회사에서 마케팅 계획 등을 세울 때 우선순위를 10 대, 20 대로 두어야 한다는 아이디어도 낼 수 있습니다.

```
C:\Users\user\PycharmProjects\PAD\venv\Scripts\python.exe C:/Users/user/PycharmProjects/PAD/Chicken6.py
```

```
1669063 극제 이승
```

```
인구와 치킨집 주문량의 관계를 파악하는 프로젝트입니다.
```

```
어떤 메뉴를 선택하시겠습니까?
```

1. 인구수(age data)와 치킨집 주문의 상관관계
2. 연령대와 치킨집 주문의 상관관계: ☒
2. 연령대와 치킨집 주문의 상관관계를 알려드리겠습니다.

```
*치킨 데이터에서 연령대와 전화량 열로 만든 데이터프레임
```

```
연령대   전화량
```

```
0   10대   3137
1   20대   4530
2   30대   4703
3   40대   4837
4   50대   3735
```

```
#2. 인구 데이터 활용하기
```

```
*인구 데이터에서 추출한 10대-60대까지 연령대별 인구수 리스트
```

```
[850988.0, 2301044.0, 3843502.0, 5433707.0, 6983610.0, 8134176.0]
```

```
# 3. 치킨 전화량과 연령별 인구수 비교 및 계산
```

```
*연령대별 인구 대비 전화량의 비율을 구하고 10000을 곱한 리스트
```

```
[36 19 12  8  5  3]
```

```
# 4. 치킨 전화량과 연령별 인구수의 pie그래프 그리기
```

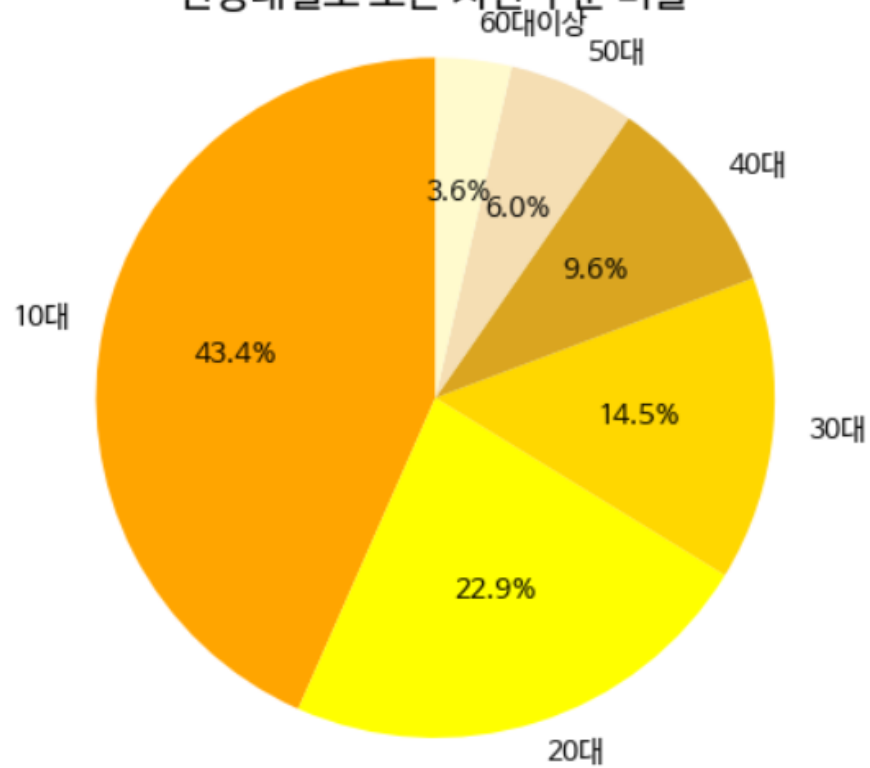
```
결론: 연령대가 낮을 수록 치킨을 많이 시켜먹는다.
```

```
10대-60대까지 내림차순으로 비율이 감소하는 것을 보았을 때 치킨 소비량은 나이가 어릴 수록 많다는 것을 알 수 있다.
```

```
특히 10-20대의 비율이 과반수를 넘었다는 것을 알 수 있다.
```

```
Process finished with exit code 0
```

연령대별로 보는 치킨 주문 비율



```
# 4. 치킨 전화량과 연령별 인구수의 pie그래프 그리기
```

```
-----  
결론: 연령대가 낮을 수록 치킨을 많이 시켜먹는다.
```

```
10대-60대까지 내림차순으로 비율이 감소하는 것을 보았을 때 치킨 소비량은 나이가 어릴 수록 많다는 것을 알 수 있다.  
특히 10-20대의 비율이 과반수를 넘었다는 것을 알 수 있다.  
-----
```

```
Process finished with exit code 0
```