

서울시 행정구별 커피분야 소비인구 변화 추이 분석

이예원, 임형주, 설진철, 조해영





기획배경

타겟층

커피 프렌차이즈를
꿈을 꾸는 창업가

제공할 데이터

서울시의 행정구별
커피시장의 가공된 자료

해결 방안 제시

코로나 전과 후의
행정동별 커피분야
소비인구 변화 정보를
분석한 자료를 제공



해결 방안



데이터 탐색과 전처리

카페 창업시 참고할
데이터를 수집
/
서울시로 한정해서
데이터 전처리 수행



데이터 분석과 시각화

행정구별, 연령별, 일자별
커피소비인구 분석
/
plot 으로 시각화



의미 도출

데이터 의미 분석
/
카페 위치선정 및
고객층 선정에 도움 제공



데이터 정보

gov_dn_cd	brtc_nm	signgu_nn	adstrd_nm	de	sex_se	year_se	cnsmr_popltn_co
1111051500	서울특별시	종로구	청운효자동	20200301	F	20	91.44571764
1111051500	서울특별시	종로구	청운효자동	20200301	F	25	83.82524117
1111051500	서울특별시	종로구	청운효자동	20200301	F	30	53.34333529
1111051500	서울특별시	종로구	청운효자동	20200301	F	35	60.96381176
1111051500	서울특별시	종로구	청운효자동	20200301	F	45	30.48190588
1111051500	서울특별시	종로구	청운효자동	20200301	M	20	45.72285882
1111051500	서울특별시	종로구	청운효자동	20200301	M	25	182.8914353
1111051500	서울특별시	종로구	청운효자동	20200301	M	30	129.5481
1111051500	서울특별시	종로구	청운효자동	20200301	M	35	60.96381176
1111051500	서울특별시	종로구	청운효자동	20200301	M	40	30.48190588
1111051500	서울특별시	종로구	청운효자동	20200301	M	45	45.72285882
1111051500	서울특별시	종로구	청운효자동	20200301	X	xx	30.48190588

2020년 3월~5월

행정동별 커피 소비 인구

row : 1048575

column : 8

- 데이터 : 행정동별_커피분야_소비인구(2020.03_2020.05).csv
- 출처 : https://www.bigdata-environment.kr/user/data_market/detail.do?id=306627b0-2f03-11ea-bccd-b704c648ae09



데이터 전처리(EDA)

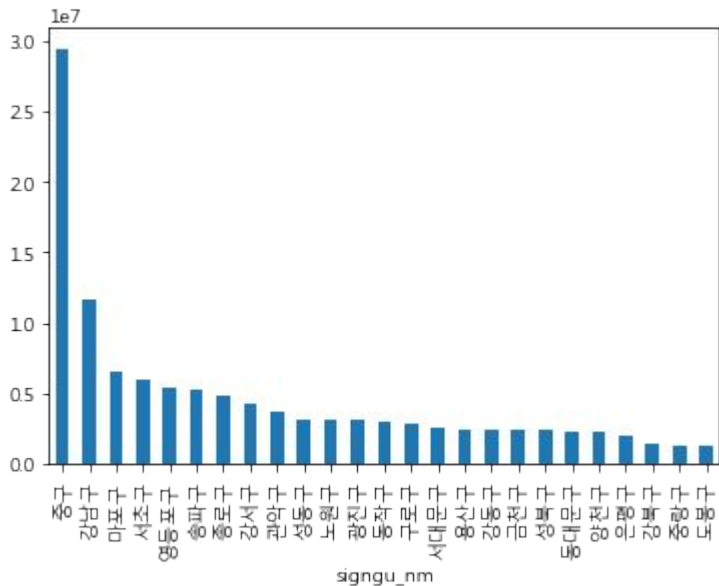
```
idx_=df[df['sex_se']=='X'].index
```

```
df.drop(idx_, inplace=True) //외국인 데이터 생략
```

```
df_seoul=df[df['brtc_nm']=='서울특별시'] //서울시 데이터만 고려
```



데이터 분석



3개월 동안의 서울시 내 구별 커피
소비인구의 합을 구하여 나타낸
그래프입니다.

서울시 **중구**의 커피 소비인구가 월등히 높은 것으로 파악됩니다. 중구 다음으로 강남구, 마포구, 종로구 순으로, 중구의 커피 소비인구가 다음 세 구의 커피 소비인구의 합보다 많음을 알 수 있습니다.

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [ ]: df = pd.read_csv('C:/Users/aaaaa/Downloads/행정동별_커피분야_소비인구(2020.03_2020.05).csv')
```

```
In [ ]: # 데이터 확인
df.shape
df.info()
df.columns
df.head()
df.tail()
df.describe
```

```
In [ ]: # 데이터 전처리 : 법인 및 외국인(X) 제외, 서울특별시만 추출
idx=df[df['sex_se']=='X'].index
df.drop(idx_) #외국인 데이터 생략
df_seoul=df[df['brtc_nm']=='서울특별시'] #서울시 데이터만 고려
```

```
In [ ]: # 'de' 컬럼 데이터 타입 변환
import datetime
df_seoul['de'] = pd.to_datetime(df_seoul['de'].astype(str), format='%Y-%m-%d')
```

```
In [ ]: # 한글 깨지는 현상 방지
from matplotlib import font_manager, rc
f_path = "C:/windows/fonts/malgun.ttf"
font_name = font_manager.FontProperties(fname= f_path).get_name()
rc('font', family = font_name)
```

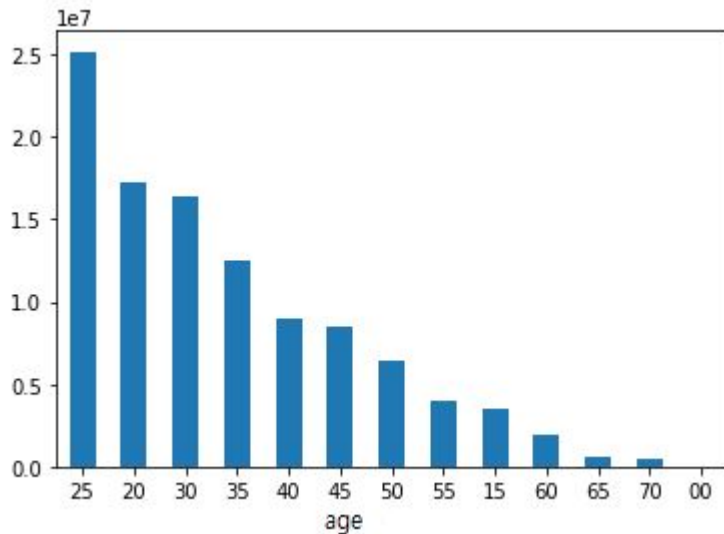
```
In [ ]: df_seoul.groupby('signgu_nm')['cnsmr_popltn_co'].mean().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby(['year_se'])['cnsmr_popltn_co'].sum().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby('de')['cnsmr_popltn_co'].sum().plot(kind='line')
```

```
In [ ]: df_seoul.pivot_table(index='signgu_nm', columns='year_se', values='cnsmr_popltn_co', aggfunc=np.mean)
```


고객층 선정



고객층 선정에 관한 자료로 연령대별로 커피 소비인구의 3개월간 총 합계를 나타내는 그래프입니다.

이로써 **25~29세** 연령대가 가장 커피를 많이 소비하는 것을 알 수 있습니다.

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [ ]: df = pd.read_csv('C:/Users/aaaaa/Downloads/행정동별_커피분야_소비인구(2020.03_2020.05).csv')
```

```
In [ ]: # 데이터 확인
df.shape
df.info()
df.columns
df.head()
df.tail()
df.describe
```

```
In [ ]: # 데이터 전처리 : 법인 및 외국인(X) 제외, 서울특별시만 추출
idx=df[df['sex_se']=='X'].index
df.drop(idx_) #외국인 데이터 생략
df_seoul=df[df['brtc_nm']=='서울특별시'] #서울시 데이터만 고려
```

```
In [ ]: # 'de' 컬럼 데이터 타입 변환
import datetime
df_seoul['de'] = pd.to_datetime(df_seoul['de'].astype(str), format='%Y-%m-%d')
```

```
In [ ]: # 한글 깨지는 현상 방지
from matplotlib import font_manager, rc
f_path = "C:/windows/fonts/malgun.ttf"
font_name = font_manager.FontProperties(fname= f_path).get_name()
rc('font', family = font_name)
```

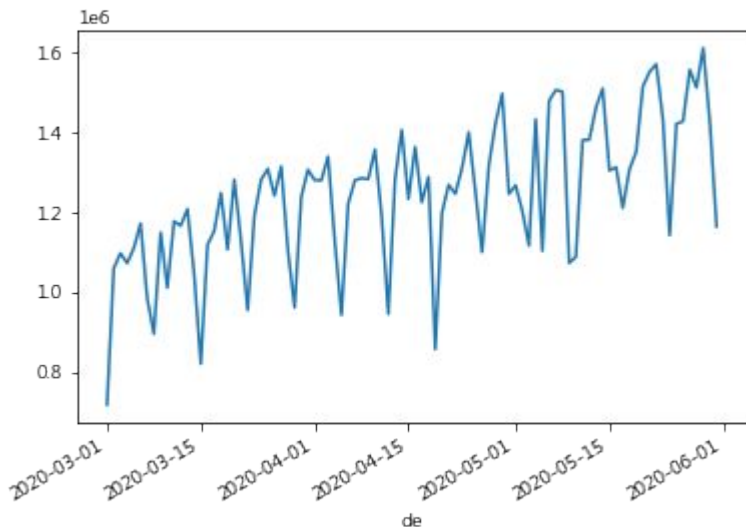
```
In [ ]: df_seoul.groupby('signgu_nm')['cnsmr_popltn_co'].mean().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby(['year_se'])['cnsmr_popltn_co'].sum().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby('de')['cnsmr_popltn_co'].sum().plot(kind='line')
```

```
In [ ]: df_seoul.pivot_table(index='signgu_nm', columns='year_se', values='cnsmr_popltn_co', aggfunc=np.mean)
```

고객층 선정



총 3개월 기간(3월~5월)동안의 커피 소비 인구 변화를 보여줍니다. 전반적으로 소비인구가 증가하는 추세를 알 수 있습니다.

커피 소비 인구의 증가 이유에는 여러 요소가 있을 수 있지만, 봄에서 여름으로 넘어가는 날씨 변화 및 코로나 장기화에 따라 소비자들의 활동량 증가로 인해 소비량이 증가한 것으로 추측됩니다. 또한 배달 어플 발달 등으로 코로나가 커피 소비 인구 양에 큰 영향을 주지는 않았다고 예측합니다.

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [ ]: df = pd.read_csv('C:/Users/aaaaa/Downloads/행정동별_커피분야_소비인구(2020.03_2020.05).csv')
```

```
In [ ]: # 데이터 확인
df.shape
df.info()
df.columns
df.head()
df.tail()
df.describe
```

```
In [ ]: # 데이터 전처리 : 법인 및 외국인(X) 제외, 서울특별시만 추출
idx=df[df['sex_se']=='X'].index
df.drop(idx_) #외국인 데이터 생략
df_seoul=df[df['brtc_nm']=='서울특별시'] #서울시 데이터만 고려
```

```
In [ ]: # 'de' 컬럼 데이터 타입 변환
import datetime
df_seoul['de'] = pd.to_datetime(df_seoul['de'].astype(str), format='%Y-%m-%d')
```

```
In [ ]: # 한글 깨지는 현상 방지
from matplotlib import font_manager, rc
f_path = "C:/windows/fonts/malgun.ttf"
font_name = font_manager.FontProperties(fname= f_path).get_name()
rc('font', family = font_name)
```

```
In [ ]: df_seoul.groupby('signgu_nm')['cnsmr_popltn_co'].mean().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby(['year_se'])['cnsmr_popltn_co'].sum().sort_values(ascending=False).plot(kind='bar')
```

```
In [ ]: df_seoul.groupby('de')['cnsmr_popltn_co'].sum().plot(kind='line')
```

```
In [ ]: df_seoul.pivot_table(index='signgu_nm', columns='year_se', values='cnsmr_popltn_co', aggfunc=np.mean)
```

발전방향 및 결론



한계 및 발전방향

- 한정적인 연령대
- 접근 가능한 데이터의 한계로 시계열이 좁음
- 각 지역 카페가 몇 개인지 나와있는 데이터와 결합시 더 좋은 분석 가능할 것으로 전망
- 파일 손상으로 인해 코로나와 커피 판매량 사이 상관관계를 입증할 만한 데이터셋의 범위 한도가 좁음





결론

이번 프로젝트는

카페 창업을 고려하는 이들에게 매출 및 성별, 연령별 소비자 정보를
제공했습니다.

특히 코로나 이후로 다양한 변수들의 변화 추이를 분석하여, 향후에도
지속가능한 운영이 가능한 입지를 고려할 때 참고할 수 있는 정보를 제공해
줍니다.

