

나는 돈을 어디에 썼을까?

코로나 발생 전후로 살펴보는 소비 데이터 분석

By.광인사 데이터 파이터

목차

1. 주제 및 목적

2. 팀원 역할

3. 프로세스

4. 데이터 소개

5. 데이터 전처리

6. EDA 및 시각화

7. 프로젝트 결과

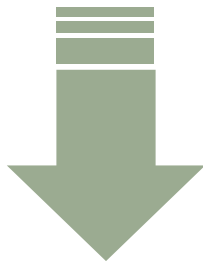
8. 결론



주제 및 목적

월말 신용카드 이용가능 회원의 업종별 거래데이터 분석

부제: 코로나 발생 전후 비교



우리의 목표

코로나 발생 전후로 신용카드 거래량 변화를 파악하여
코로나가 거래에 어떤 영향을 미쳤는지 살펴보자





팀원 역할

김채란

데이터 파악
데이터 분석 방향 설정

고석영

데이터 파악
발표자료 준비

봉명선

데이터 파악
데이터 분석 방향 설정

최선영

데이터 파악
발표자료 준비



프로세스



13일

데이터 파악
데이터 분석 방향 설정
역할 및 업무 분담



14일

데이터 분석 진행



15일

데이터 분석 보완
시각화 진행



16일

시각화 진행
코드 취합 및 정리
발표자료 준비



75개의 열 워드클라우드

- 254394행, 75열
- Int: 70개의 열 / object: 5개의 열
- 결측치 없음, 단 10미만의 레코드 0 처리됨을 고려

데이터 훑아보기





데이터 전처리

1 새로운 열 생성

- '소비년월' 열에서 year, month 열로 분할
- '고객거주지역' 열에서 앞 두 글자만 추출

2 클러스터링

- 유사업종끼리 통합
48개 업종에서 9개 업종으로 축소

3 불필요한 열 제거

- 클러스터링 이후 기존의 열 제거



EDA 및 시각화

▶ 파이 그래프



```
def pie_chart(data1, data2, title, ncol=3):
    f, ax = plt.subplots(1, 2, figsize=(15, 15))
    labels = data1.columns
    colors = sns.color_palette('hls', len(labels))

    # 고객비율
    data1.T.plot.pie(labels=data1.T.index, autopct='%.0f%%', ax=ax[0], colors=colors, subplots=True,
    fontsize=15)
    ax[0].set_title( title + ' 고객비율', position=(0.5, 0.5+0.42), fontsize=20)
    ax[0].set_ylabel('')
    ax[0].legend(loc=(0.0, 1.0), ncol=ncol, fontsize=10)

    # 거래비율
    data2.T.plot.pie(labels=data2.T.index, autopct='%.0f%%', ax=ax[1], colors=colors, subplots=True,
    fontsize=15)
    ax[1].set_title( title + ' 거래비율', position=(0.5, 0.5+0.42), fontsize=20)
    ax[1].set_ylabel('')
    ax[1].legend(loc=(0.0, 1.0), ncol=ncol, fontsize=10)

    plt.show()
```

시각화 주요 함수





EDA 및 시각화

➤ 막대 그래프

```
def bar_graph(data, title, xlabel, ylabel):  
    data.plot(kind='bar', figsize=(10, 6))  
  
    plt.title(title, fontsize=18)  
    plt.xlabel(xlabel, fontsize=15)  
    plt.ylabel(ylabel, fontsize=15)  
    plt.xticks(rotation=45, fontsize=12)  
    plt.yticks(fontsize=12)  
    plt.axhline(y=0, xmin=0.0, xmax=1.0, color='black')  
  
    plt.show()
```

➤ 누적 막대 그래프

```
def barh_graph(data, title, xlabel, ylabel):  
    data.plot(kind='barh', stacked=True, figsize=(10, 8))  
  
    plt.title(title, fontsize=25)  
    plt.xlabel(xlabel, fontsize=15)  
    plt.ylabel(ylabel, fontsize=15)  
    plt.xticks(rotation=45, fontsize=12)  
    plt.yticks(fontsize=12)  
    plt.legend(loc='best', ncol=5, fontsize=12)  
  
    plt.show()
```

시각화 주요 함수

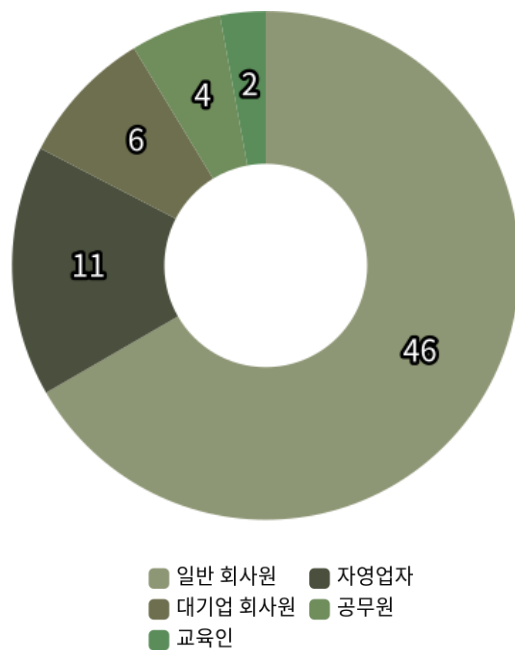




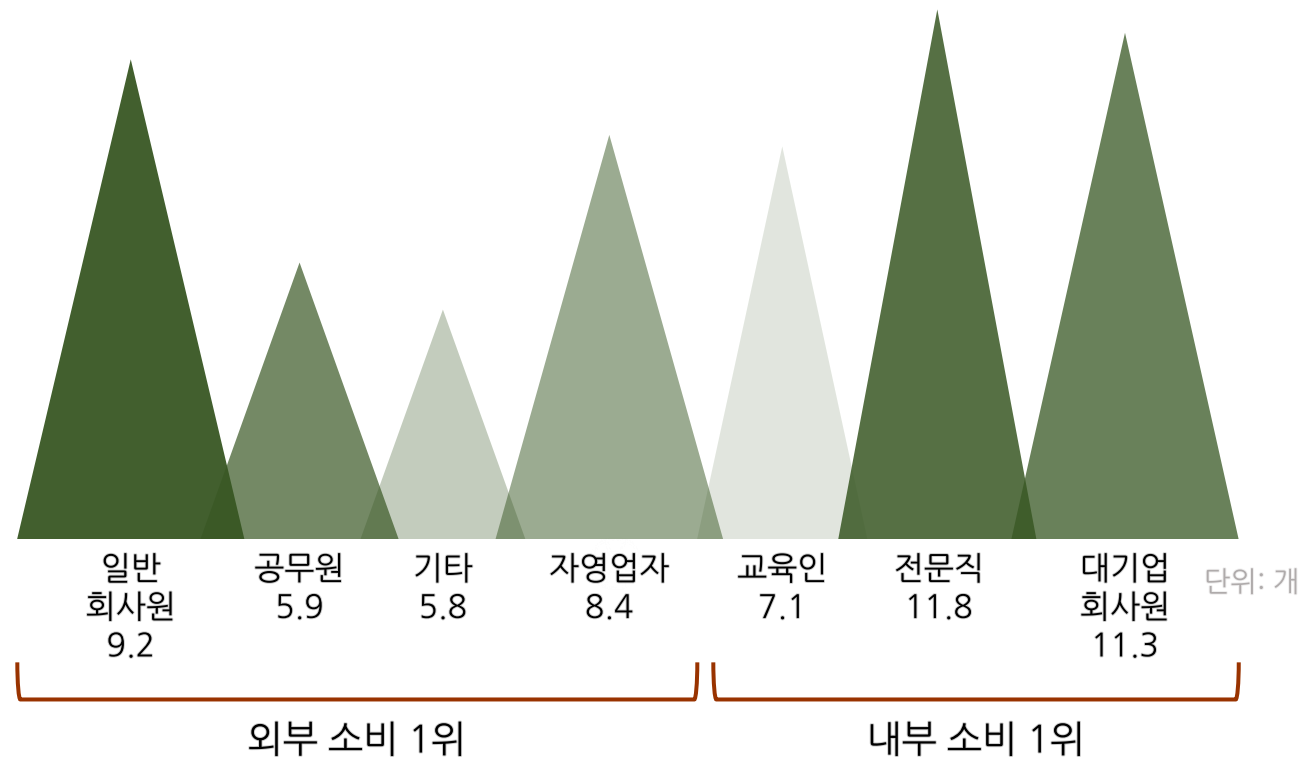
EDA 및 시각화

업종 집중 분석 코로나 발생 전후 비교 코로나 발생 장단기 비교

➤ 직업별 업종 거래비율



➤ 직업별 1인당 거래수



직업 특성에 따라 내/외부 소비량이 다름

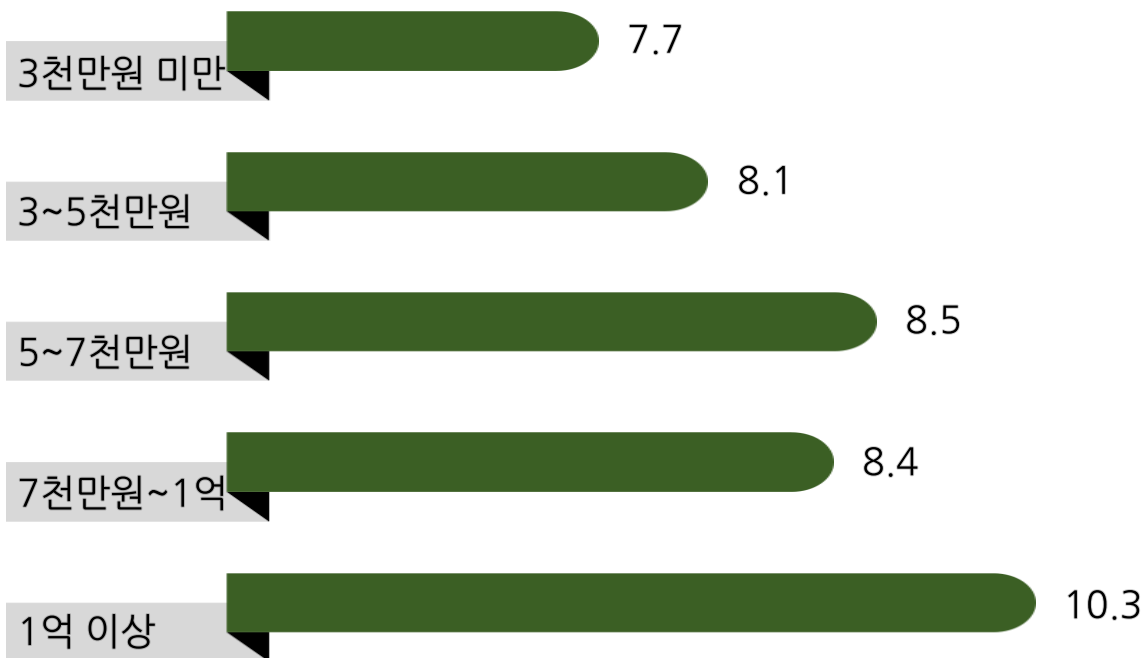




EDA 및 시각화

업종 집중 분석 코로나 발생 전후 비교 코로나 발생 장단기 비교

➤ 추정소득별 1인당 거래수



소득이 많을수록 거래수도 많은 경향





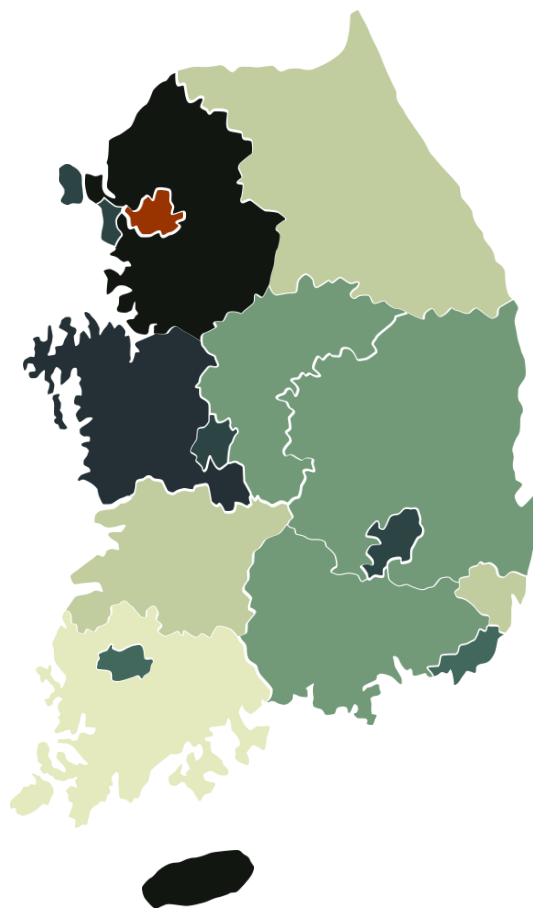
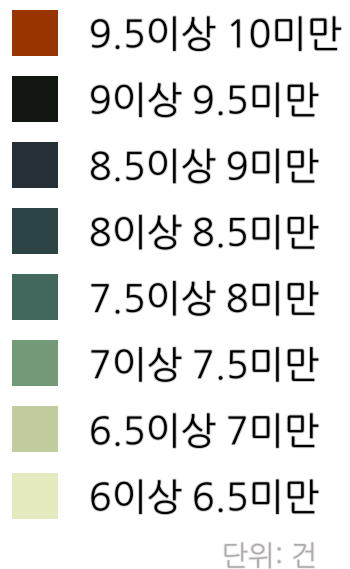
EDA 및 시각화

업종 집중 분석

코로나 발생 전후 비교

코로나 발생 장단기 비교

➤ 고객 거주 시도별 1인당 거래수



인구수가 많을수록 거래량이 많은 경향





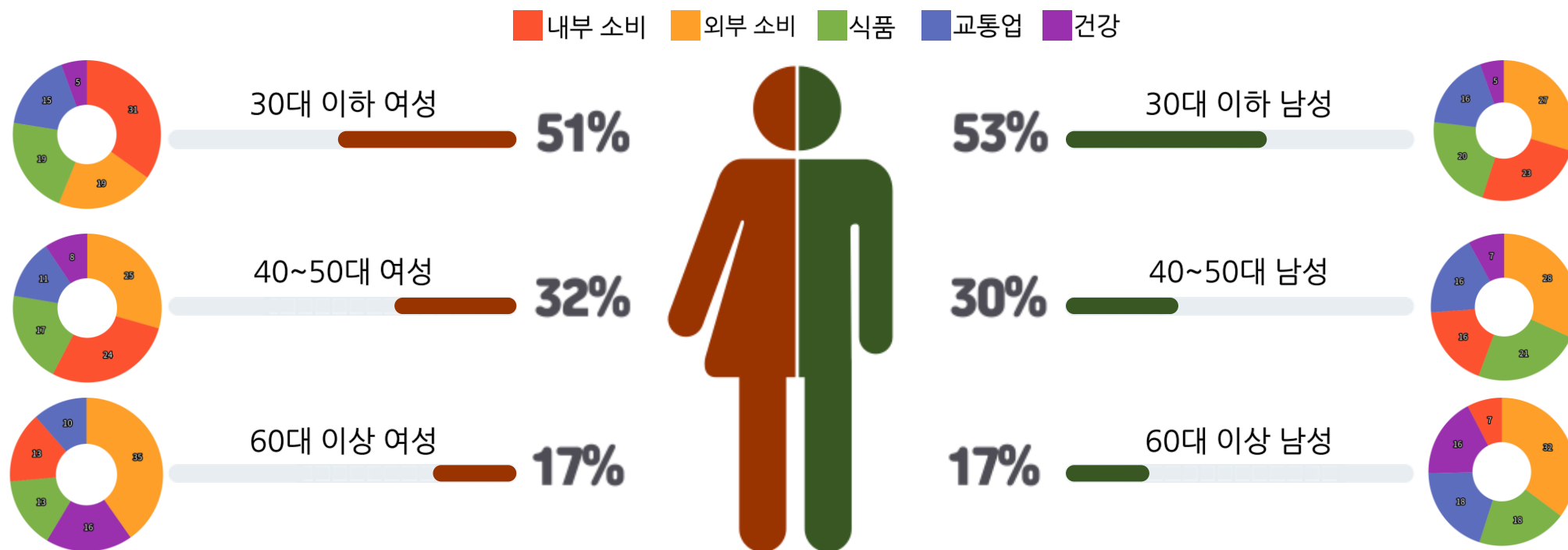
EDA 및 시각화

업종 집중 분석

코로나 발생 전후 비교

코로나 발생 장단기 비교

➤ 성별연령별 1인당 거래수



성별에 의한 차이는 미미, 연령별 소비 영역은 크게 다름





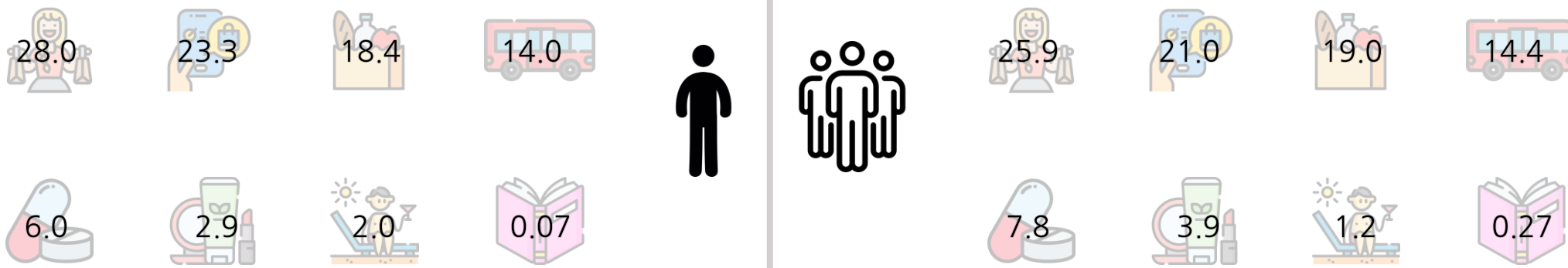
EDA 및 시각화

업종 집중 분석

코로나 발생 전후 비교

코로나 발생 장단기 비교

➤ 1인/다인가구별 업종 거래 비율



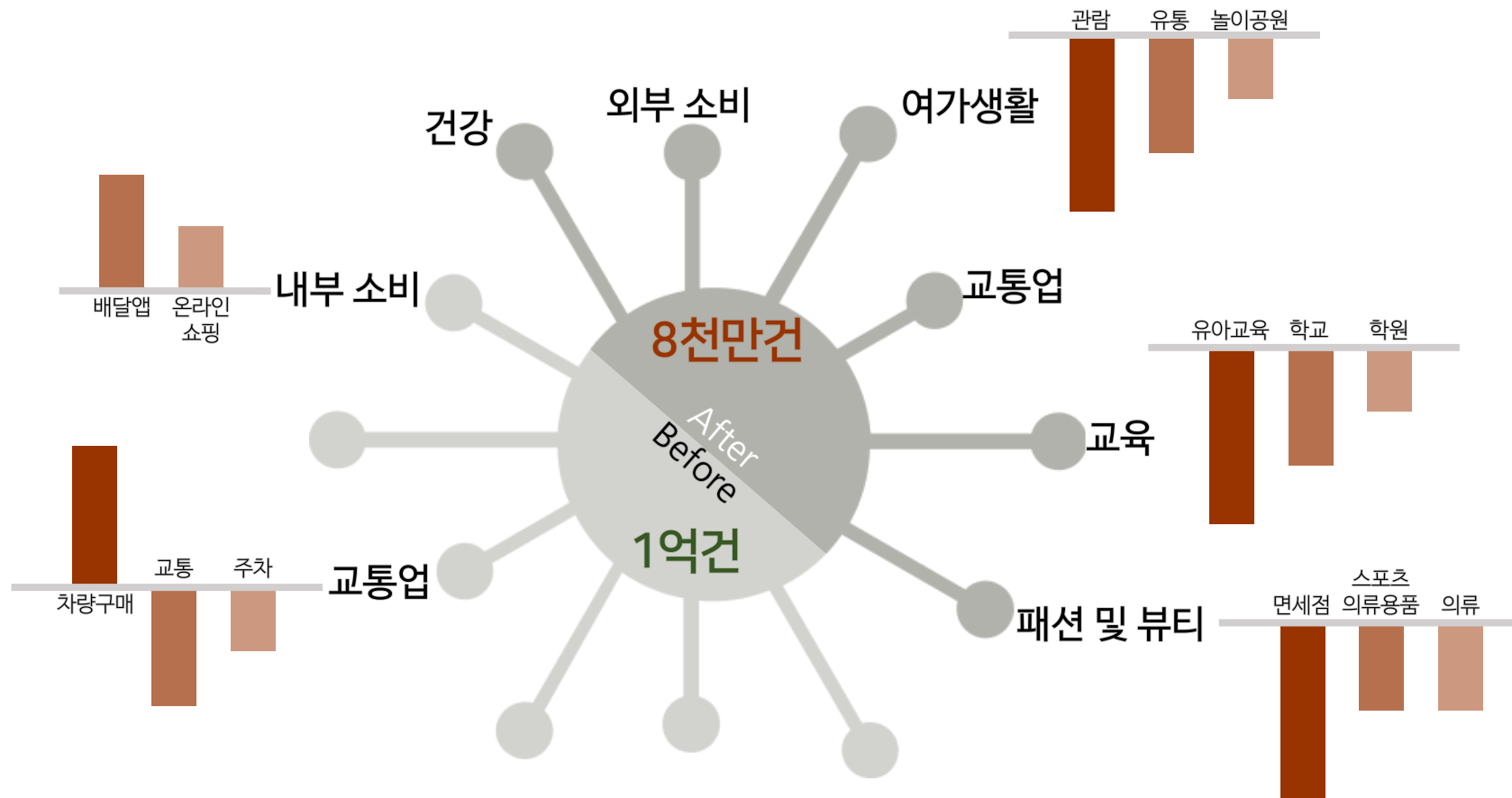
비슷한 1인가구와 다인가구의 소비패턴





EDA 및 시각화

업종 집중 분석 코로나 발생 전후 비교 코로나 발생 장단기 비교



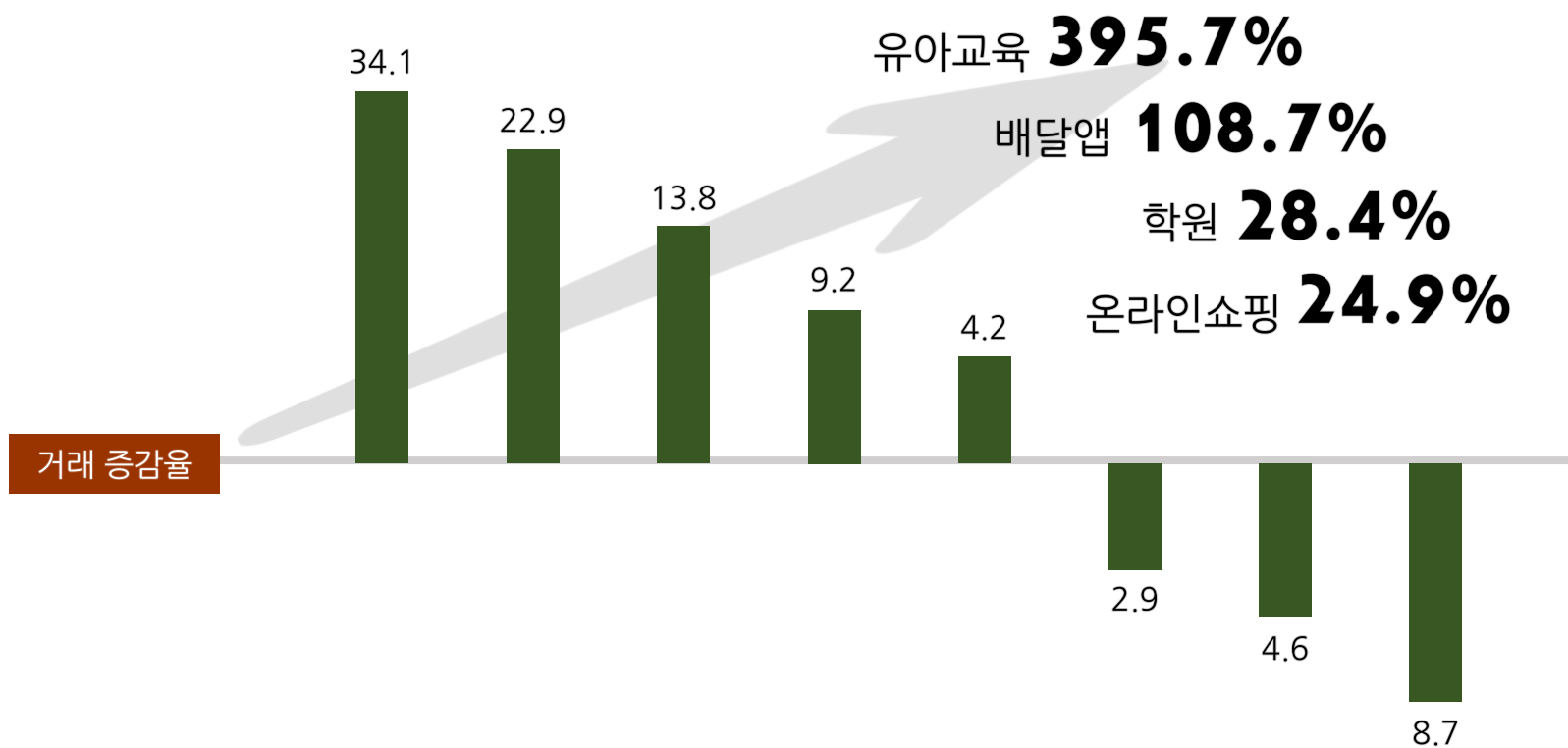
코로나 이후 변화하는 소비 트렌드





EDA 및 시각화

업종 집중 분석 코로나 발생 전후 비교 코로나 발생 장단기 비교



코로나 장기화를 기회로 소비회복을 넘어 성장하는 업종들





프로젝트 결과

POINT 1

소비 건수 전체적으로 감소



소비 회복으로 거래 전체 건수 10% 회복

POINT 2

그 중에서도
관람, 유흥, 놀이공원, 취미 등의
여가생활관련 소비가 크게 감소



코로나로 크게 경제적으로 위축되었으나
일부 항목을 제외하고 코로나의 장기화와 함께
소비 심리가 크게 회복한것으로 보임

POINT 3

배달과 온라인 쇼핑만
유일하게 증가



스포츠, 화장품 등의 산업은 거리두기와 마스크
착용으로 경제적인 회복이 어려워 보임



결론

