

# 빅데이터 시각화 팀

---

이륜차와 배달원 상관관계의 데이터

참여자: 한지희, 고유진, 김지연

# 목차

## 주제 선정 이유

전 세계 적으로 팬데믹이 유행하게 되면서 여러 방면에서 많은 변화가 일어나게 되었습니다.

그 중에서도 우리나라에서 배달을 이용하는 상황이 많아짐에 따라 매년 점차 증가하는 추세이기 때문에 이를 다루어 보기 위해 주제로 선정하였습니다.

## 사용한 데이터&목표

배달에서 주로 사용되어 지는 교통 수단은 이륜차, 즉 오토바이이기 때문에 늘어나는 배달원 수와 이륜차 등록 수의 변화가 연관이 있을 것이라고 유추할 수 있습니다.

그래서 각각의 연도별 데이터를 수집하였습니다.

그리고 이를 이용해서 **둘의 상관관계를 시각화**하여 나타낼 것입니다.

## 시각화 방식

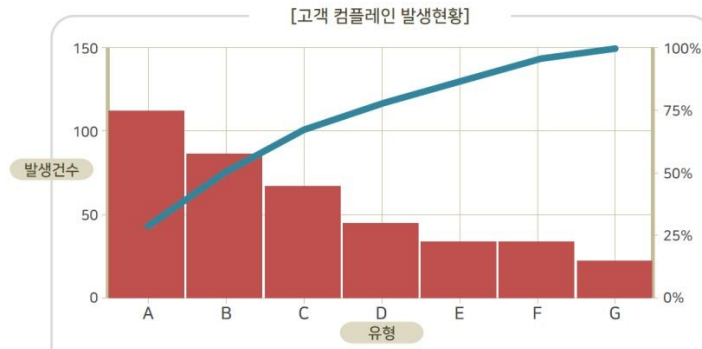
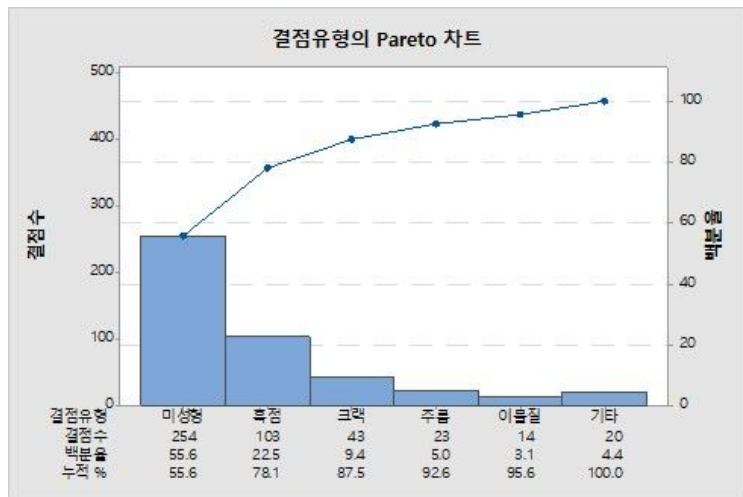
각 연도별 이륜차 등록 수=>히스토그램

공식적으로 취업에 등록된 배달원 수 =>꺾은 선 그래프

즉, 파레토 차트를 이용하여 시각화할 것입니다.

# 시각화 방식(시각화 디자인)-파레토 차트

저희 팀이 나타내고 싶은 관계를 이 유형의 차트가 잘 표현해 줄 것 같아서 결정했습니다.



컴플레인 유형 A, B, C, D가 전체 컴플레인의 80% 가량을 차지함












출처: <https://stat.molit.go.kr/portal/main/portalMain.do> 국토교통  
통계누리에서 제공하는 표를 CSV파일로 저장하여 참고

# 데이터 가공 방법

배달원 csv의 경우 : 원본 pdf파일에서 배달원의 수만 찾아서 작성. 원본에서는 천명 단위로 둔채 나와있었지만 구현할 때는 배달원의 수를 확실하게 보여주기 위해 명단위로 바꾸어서 csv 파일에 작성함.

출처: [통계청 \(kostat.go.kr\)](http://kostat.go.kr) 에서 지역별 고용조사 자료 참고

	2019년 상반기 지역별고용조사 취업자...	2022-11-30 오후 2:57	Microsoft Edge P...	4,356KB
	2019년 하반기 지역별고용조사 취업자...	2022-11-30 오후 2:56	Microsoft Edge P...	3,038KB
	2020년 상반기 지역별고용조사 취업자...	2022-11-30 오후 2:54	Microsoft Edge P...	3,757KB
	2020년 상반기 지역별고용조사 취업자...	2022-11-30 오후 2:54	한컴오피스 한글 ...	4,814KB
	2020년 하반기 지역별고용조사 취업자...	2022-11-30 오후 2:54	Microsoft Edge P...	2,999KB
	2021년 상반기 지역별고용조사 취업자...	2022-11-30 오후 2:53	Microsoft Edge P...	2,841KB
	2021년 하반기 지역별고용조사 취업자...	2022-11-30 오후 2:43	Microsoft Edge P...	2,854KB
	2021년 하반기 지역별고용조사 취업자...	2022-11-30 오후 2:43	한컴오피스 한글 ...	3,613KB
	2022년 상반기 지역별고용조사 취업자...	2022-11-30 오후 2:40	Microsoft Edge P...	2,922KB

## 1. 직업소분류별 취업자 현황

### 가. 전체 취업자

- 직업소분류(153개, 군인제외)별 취업자 규모를 살펴보면,
  - '경영관리사무원' 235만 5천명(8.6%), '매장판매종사자' 194만 7천명(7.1%), '작물재배종사자' 127만명(4.6%) 순으로 많았음
- 전년동기대비 취업자 증감을 살펴보면,
  - '청소원및환경미화원'에서 7만 3천명, '작물재배종사자'에서 5만 6천명 증가하였고, '건설및광업단순종사자'에서 3만 1천명, '영업종사자'에서 3만명 감소하였음

< 직업소분류별 상위 취업자 현황 >

직업소분류	2018년 하반기	구성비	2019년 하반기	구성비	증감	증감률
경영 관리 사무원	2,366	8.7	2,355	8.6	-11	-0.5
매장 판매 종사자	1,940	7.2	1,947	7.1	7	0.4
작물 재배 종사자	1,215	4.5	1,270	4.6	56	4.6
자동차 운전원	1,009	3.7	1,012	3.7	3	0.3
조리사	961	3.5	996	3.6	35	3.6
청소원 및 환경미화원	880	3.2	953	3.5	73	8.3
회계 및 경리 사무원	814	3.0	836	3.0	22	2.7
영업 종사자	699	2.6	668	2.4	-30	-4.4
식음료 서비스 종사자	621	2.3	658	2.4	36	5.8
문리·기술 및 예능 강사	681	2.5	656	2.4	-26	-3.7
행정 사무원	539	2.0	509	1.9	-30	-5.6
돌봄 및 보건 서비스 종사자	454	1.7	509	1.8	55	12.0
사회복지 관련 종사자	424	1.6	436	1.6	12	2.9
음식 관련 단순 종사자	390	1.4	419	1.5	29	7.4
건설 및 광업 단순 종사자	449	1.7	419	1.5	-31	-6.8
학교 교사	405	1.5	411	1.5	7	1.7
감정·기술 영업 및 중개 관련 종사자	420	1.5	396	1.4	-24	-5.7
제조 관련 단순 종사자	406	1.5	390	1.4	-16	-4.0
비품 관리 서비스 종사자	318	1.2	352	1.3	34	10.6
배달원	346	1.3	349	1.3	3	0.7

\* 구성비는 전체 취업자(2019. 10월 기준 27,500천명) 대비 비중임

# 코드

```
# 한글 문제 해결
!sudo apt-get install -y fonts-nanum
!sudo fc-cache -fv
!rm ~/.cache/matplotlib -rf

Reading package lists... Done
Building dependency tree
Reading state information... Done
fonts-nanum is already the newest version (20170925-1).
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'sudo apt autoremove' to remove it.
0 upgraded, 0 newly installed, 0 to remove and 20 not upgraded.
/usr/share/fonts: caching, new cache contents: 0 fonts, 1 dirs
/usr/share/fonts/truetype: caching, new cache contents: 0 fonts, 3 dirs
/usr/share/fonts/truetype/humor-sans: caching, new cache contents: 1 fonts, 0 dirs
/usr/share/fonts/truetype/liberation: caching, new cache contents: 16 fonts, 0 dirs
/usr/share/fonts/truetype/nanum: caching, new cache contents: 10 fonts, 0 dirs
/usr/local/share/fonts: caching, new cache contents: 0 fonts, 0 dirs
/root/.local/share/fonts: skipping, no such directory
/root/.fonts: skipping, no such directory
/var/cache/fontconfig: cleaning cache directory
/root/.cache/fontconfig: not cleaning non-existent cache directory
/root/.fontconfig: not cleaning non-existent cache directory
fc-cache: succeeded
```

```
[ ] import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np

CHART_NAME = 'seabornTest'
cnt, PNG, UNDERBAR = 0, '.png', '_'
filename = 'bike2.csv'
```

```
[ ] bike = pd.read_csv(filename, encoding='euc-kr')

print(bike.columns)

print('-'*30)

print(bike.head(10))
print('-'*30)

Index(['시기', '오토바이', '배달원'], dtype='object')
-----
   시기  오토바이  배달원
0  2019-1H   11851   343000
1  2019-2H   11731   349000
2  2020-1H   11781   371000
3  2020-2H   11919   390000
4  2021-1H   12008   423000
5  2021-2H   11742   428000
-----
```

```
[ ] df = pd.DataFrame(bike)
```

```
# 1. 기본 스타일 설정
plt.style.use('default')
plt.rcParams['figure.figsize'] = (8, 6)
plt.rcParams['font.size'] = 12

# 2. 데이터 준비
x = np.arange(2020, 2027)
y1 = np.array([1, 3, 7, 5, 9, 7, 14])
y2 = np.array([1, 3, 5, 7, 9, 11, 19])

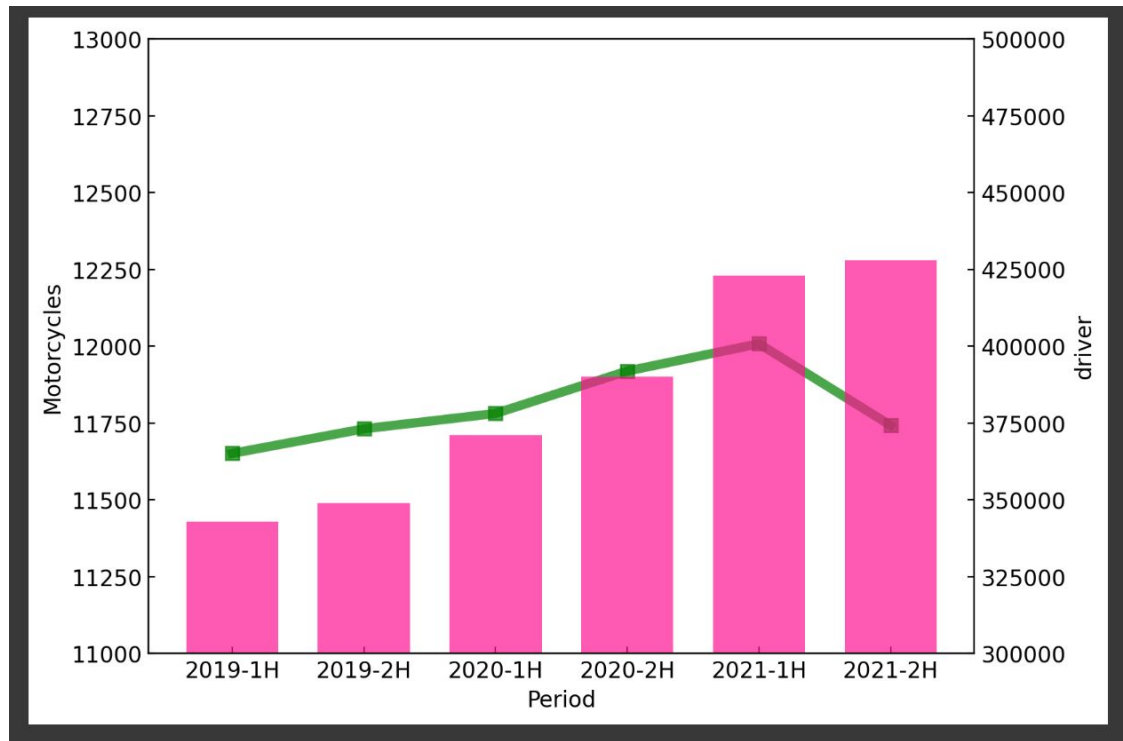
# 3. 그래프 그리기
fig, ax1 = plt.subplots()

ax1.plot(df['시기'], df['오토바이'], 's', color='green', markersize=7, linewidth=5, alpha=0.7, label='Price')
ax1.set_ylim(11000, 13000)
ax1.set_xlabel('Period')
ax1.set_ylabel('Motorcycles')
ax1.tick_params(axis='both', direction='in')

ax2 = ax1.twinx()
ax2.bar(df['시기'], df['배달원'], color='deeppink', label='Demand', alpha=0.7, width=0.7)
ax2.set_ylim(300000, 500000)
ax2.set_ylabel(r'driver')
ax2.tick_params(axis='y', direction='in')

plt.show()
```

# 시각화 구현한 것



표현하고자 했던 것으로 구현한  
결과입니다.

상반기는 20xx-1H,  
하반기는 20xx-2H  
로 나타내었습니다.

# 시각화 결과에 대한 생각

막대그래프 및 꺾은선 그래프를 사용했고, 예상과는 다르게 2021년 하반기에는 오토바이의 수가 줄어드는 것을 알게 되었다. 반면 배달원의 수는 예상과 같게 늘어난 것을 알 수 있게 되었다. 특히 년도가 바뀌면서 격차가 커지는 것을 알게 되었다.

비록 우리조가 알아보려던 오토바이의 수와 배달원의 관계에서는 2021년 하반기에는 맞지 않게 되었지만, 전반적으로 배달원의 수가 늘면서 오토바이의 수가 늘었다는 것을 알게되었다.

이에 대한 그래프를 보고 생각하게 된 점은 점점 오토바이 등록수는 유지 되거나 줄어들 것이고 배달원의 수는 늘어날 것이라고 생각된다. 왜냐하면 한 개인이 다양한 업체에 등록하게 되면 등록 수는 늘어나지만 소유하고 있는 오토바이의 수가 늘어나는 것은 아니기 때문에 이를 통해서 여러 배달업을 종사하고 있다는 것을 유추할 수 있다. 이는 실제 배달업을 하고 있는 지인에게서 알게 된 사실을 기반으로 유추한 것이다.



감사합니다