

# AI 사무업무활용 5주차 과제

2021204077 노은선

# 목차

- 데이터 설명
- 분석 목표
- 시각화
- 분석
- 결과 해석 및 인사이트

# 데이터 설명

## 데이터 개요

- 총 200개(명)
- 6개의 컬럼 존재
  - 이름 (식별용)
  - 성별 (남/여)
  - 나이 (정수형)
  - 구매금액 (만원 단위)
  - 가입연도 (정수형, 2015~2024)
  - 회원등급 (일반/VIP)

## 데이터 특징

- 20세부터 60세까지의 연령대
- 5만원부터 294만원까지의 구매금액

1	이름	성별	나이	구매금액(만원)	가입연도	회원등급
2	최우훈	남	51	131	2024	VIP
3	정준아	여	58	159	2024	VIP
4	박서영	남	51	277	2023	VIP
5	조영기	남	23	108	2020	VIP
6	이수영	남	49	250	2019	VIP
7	장준희	여	56	180	2020	VIP
8	장영희	남	42	43	2021	일반
9	조민진	남	58	174	2020	VIP
10	박서희	남	34	251	2020	VIP
11	최수아	여	48	30	2023	일반
12	조준정	남	55	17	2023	일반
13	장수아	남	32	40	2023	일반
14	최우기	남	51	177	2024	VIP
15	김우아	남	26	24	2021	일반
16	김지진	여	41	268	2023	VIP
17	윤영기	남	47	146	2019	VIP
18	정현희	여	21	147	2024	VIP
19	김민희	여	25	96	2022	일반
20	박수진	여	47	292	2019	VIP
21	조준영	남	47	219	2019	VIP
22	박우아	여	39	55	2023	일반
23	조수영	남	49	157	2022	VIP
24	정수철	여	30	190	2022	VIP
25	이민정	여	47	67	2024	일반
26	최현기	여	44	194	2022	VIP
27	윤민진	여	58	129	2021	VIP
28	윤민희	여	52	154	2023	VIP
29	이지희	여	20	62	2022	일반
30	김수아	여	46	184	2024	VIP
31	윤민정	여	32	174	2021	VIP
32	정현희	남	60	275	2020	VIP
33	윤민진	남	22	58	2020	일반
34	최수진	여	58	105	2021	VIP
35	조지아	여	25	268	2024	VIP
36	박민정	여	27	57	2021	일반
37	장지훈	남	46	64	2023	일반
38	최민아	여	28	112	2023	VIP
39	조우희	남	56	9	2020	일반

# 데이터 설명

📌 데이터 기본 정보:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   이름            200 non-null    object
1   성별            200 non-null    object
2   나이            200 non-null    int64
3   구매금액(만원)  200 non-null    int64
4   가입연도        200 non-null    int64
5   회원등급        200 non-null    object
dtypes: int64(3), object(3)
```

📌 데이터 요약 통계:

	나이	구매금액(만원)	가입연도
count	200.000000	200.0000	200.000000
mean	41.610000	155.1150	2021.565000
std	12.168329	83.4633	1.761046
min	20.000000	5.0000	2019.000000
25%	31.000000	89.5000	2020.000000
50%	43.000000	154.5000	2022.000000
75%	52.000000	227.0000	2023.000000
max	60.000000	294.0000	2024.000000

📌 결측치 확인:

```
이름            0
성별            0
나이            0
구매금액(만원)  0
가입연도        0
회원등급        0
dtype: int64
```

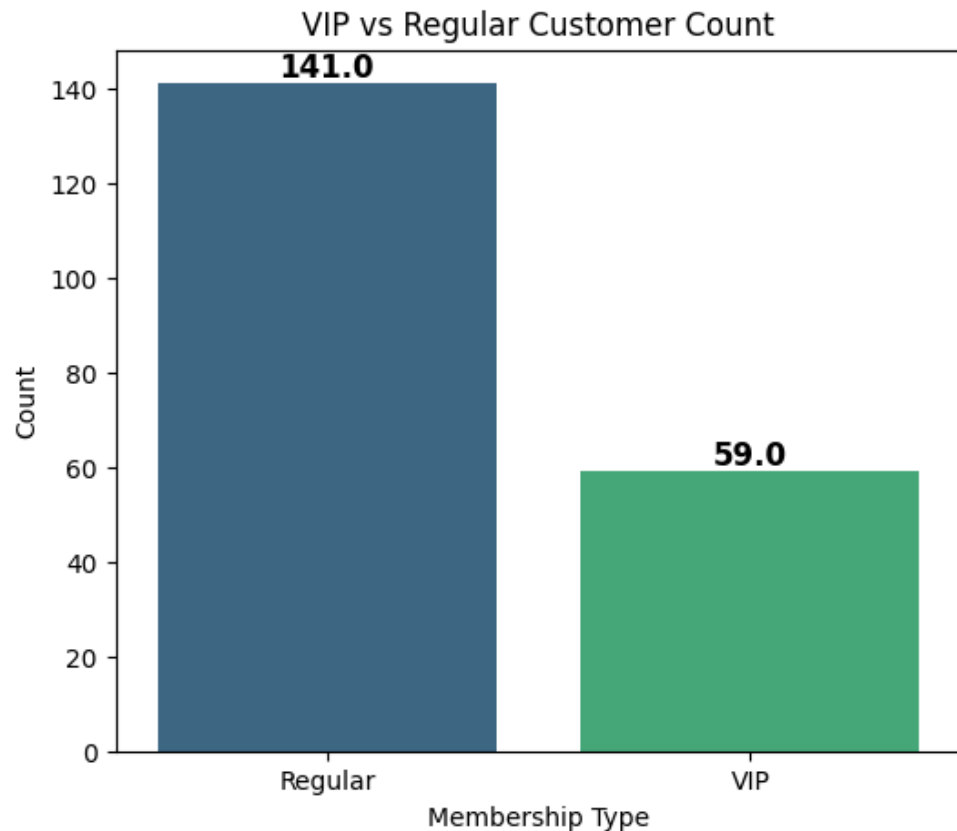
- 나이: 평균 41.61세 (20~60세)
- 구매금액: 평균 155.12만원 (5~294만원)
- 가입연도: 2019~2024년
- 회원등급: VIP 및 일반 고객 존재
- 결측치 X

# 분석 목표

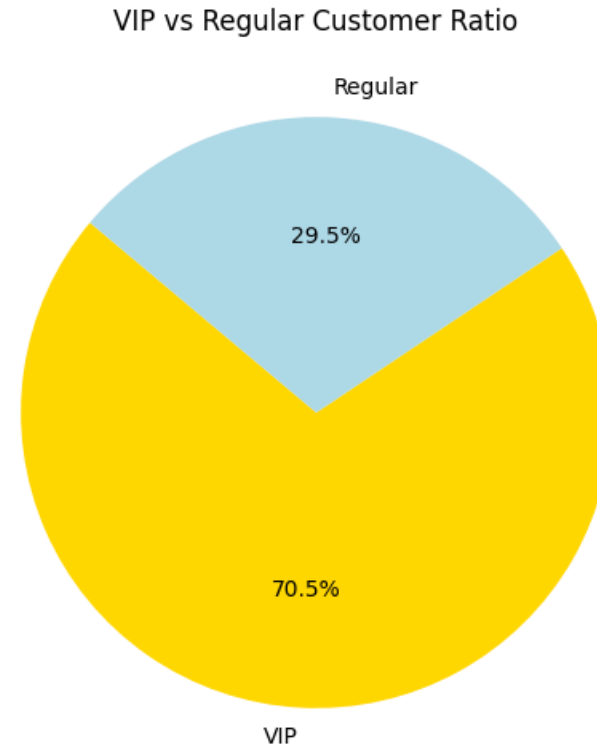
- 가입연도 별 고객 행동 변화

# EDA : VIP 와 일반 고객 비교

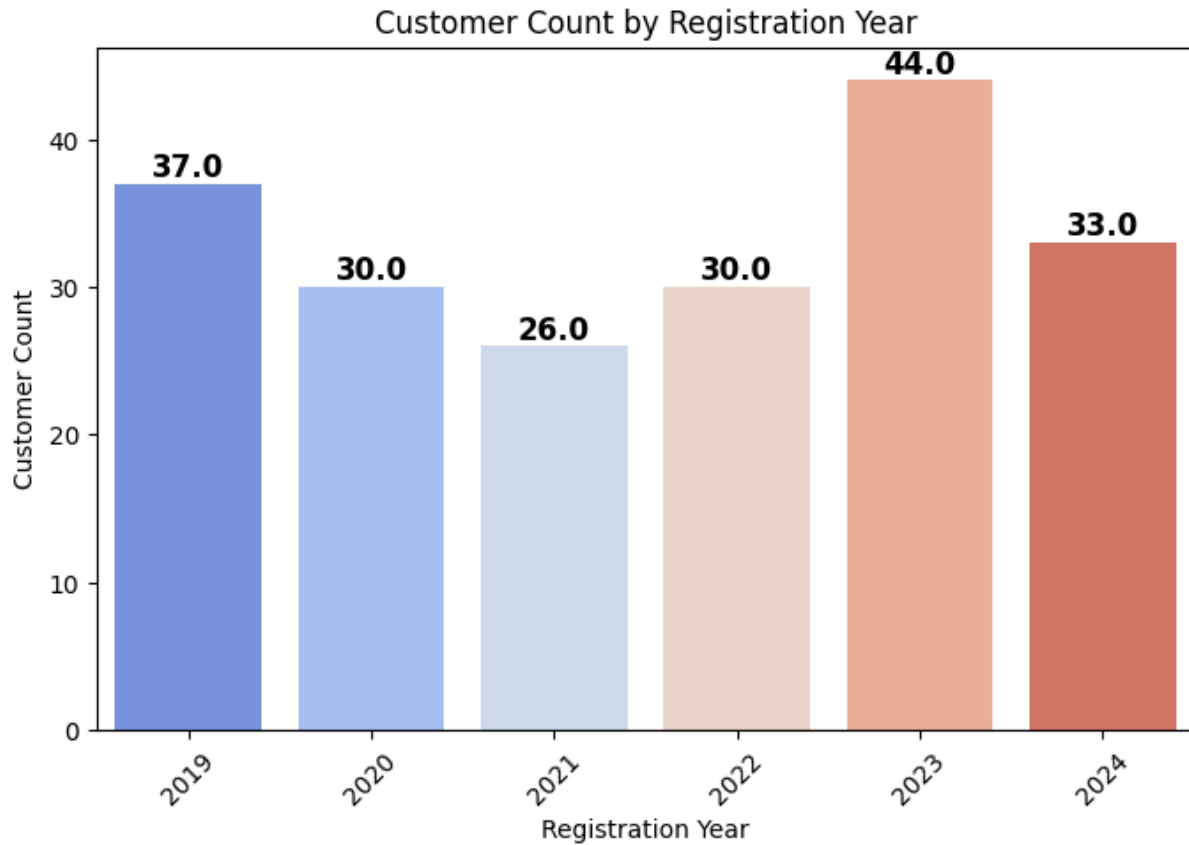
VIP와 일반 고객의 수



VIP와 일반 고객의 비율



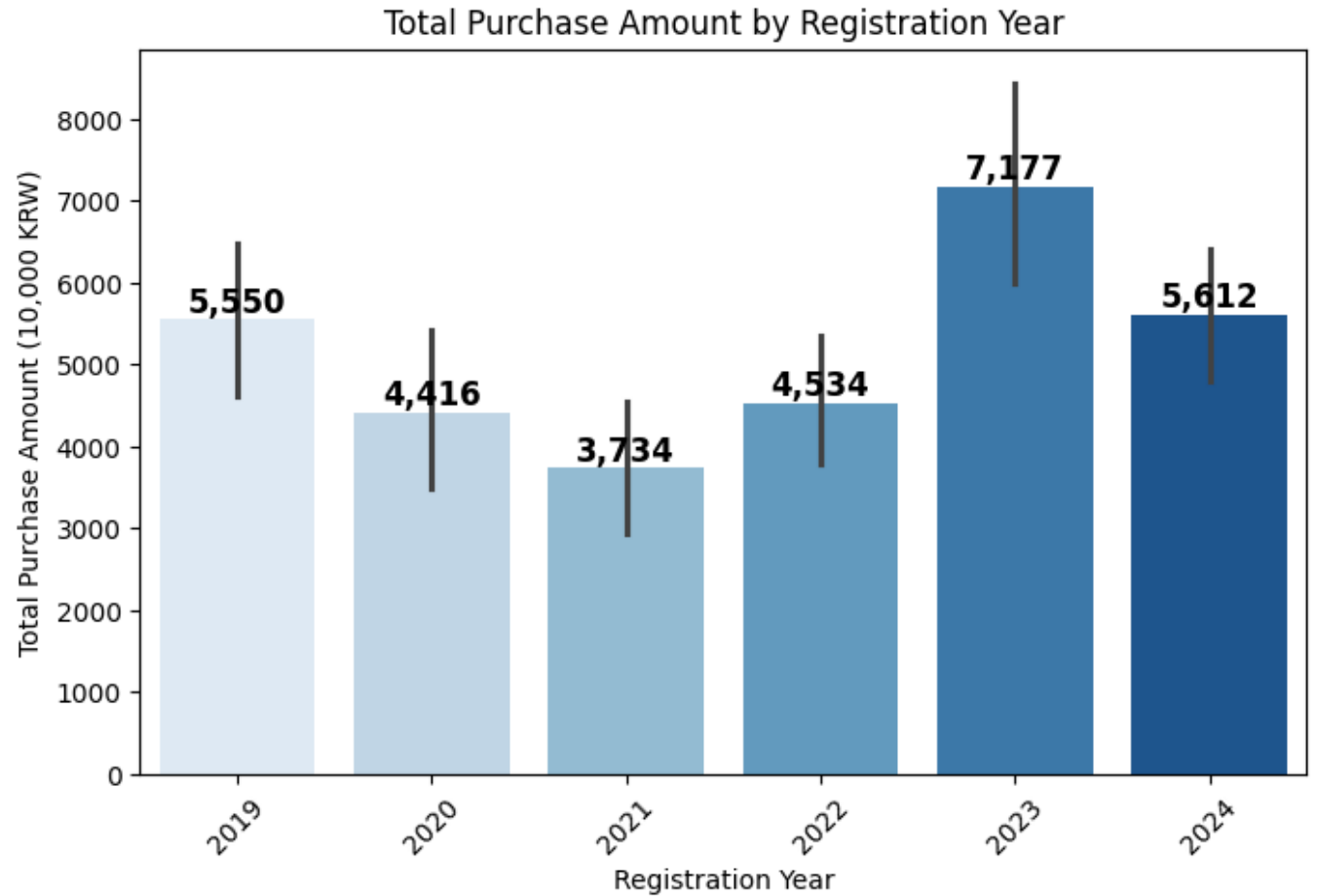
# EDA: 연도별 신규가입 고객의 수



➡ 2023년에 가입자가 가장 많았으며, 2021년에 가장 적었음.

# EDA:가입연도별 구매금액

→ 2023년이 총 구매금액 최대,  
2021년 최소에 해당함.



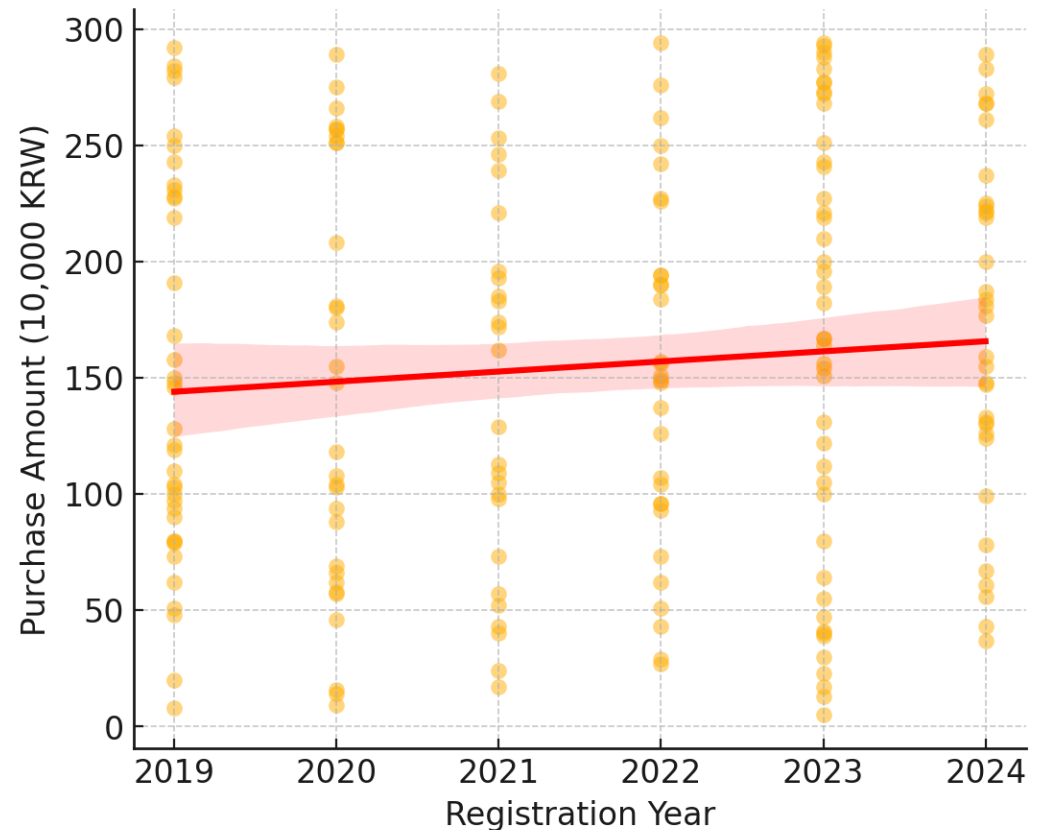


# 가입연도별 구매금액의 상관관계

- 가입 연도와 구매 금액 간 상관계수는 **0.09**로 거의 상관관계가 없는 수준

- 가입 시점보다는 다른 요인이 구매 금액에 더 큰 영향을 줄 가능성이 높음.

Correlation between Registration Year and Purchase Amount  
(Correlation Coefficient: 0.09)



# 가입 연도와 구매 금액 간 상관관계 인사이트

- 상관계수: 0.09

- 거의 무시할 수 있는 수준의 약한 양의 상관관계
- 가입 연도가 구매 금액에 유의미한 영향을 미치지 않음

- 전 연도에 걸쳐 구매 금액은 고르게 분포

- 특정 연도에 가입했다고 해서 높은 구매 금액을 보장하지 않음
- 다양한 연도에서 고액 및 저액 구매자 혼재

- 추세선은 미세한 우상향

- 가입 연도가 최근일수록 다소 구매 금액이 늘어나는 경향은 있으나
- 신뢰구간 넓고, 확신하기 어려움

- 마케팅 시사점

- 가입 연도 기반 타겟팅은 효과 낮음
- 대신, 구매 행동/금액 기반 세분화가 더 효과적

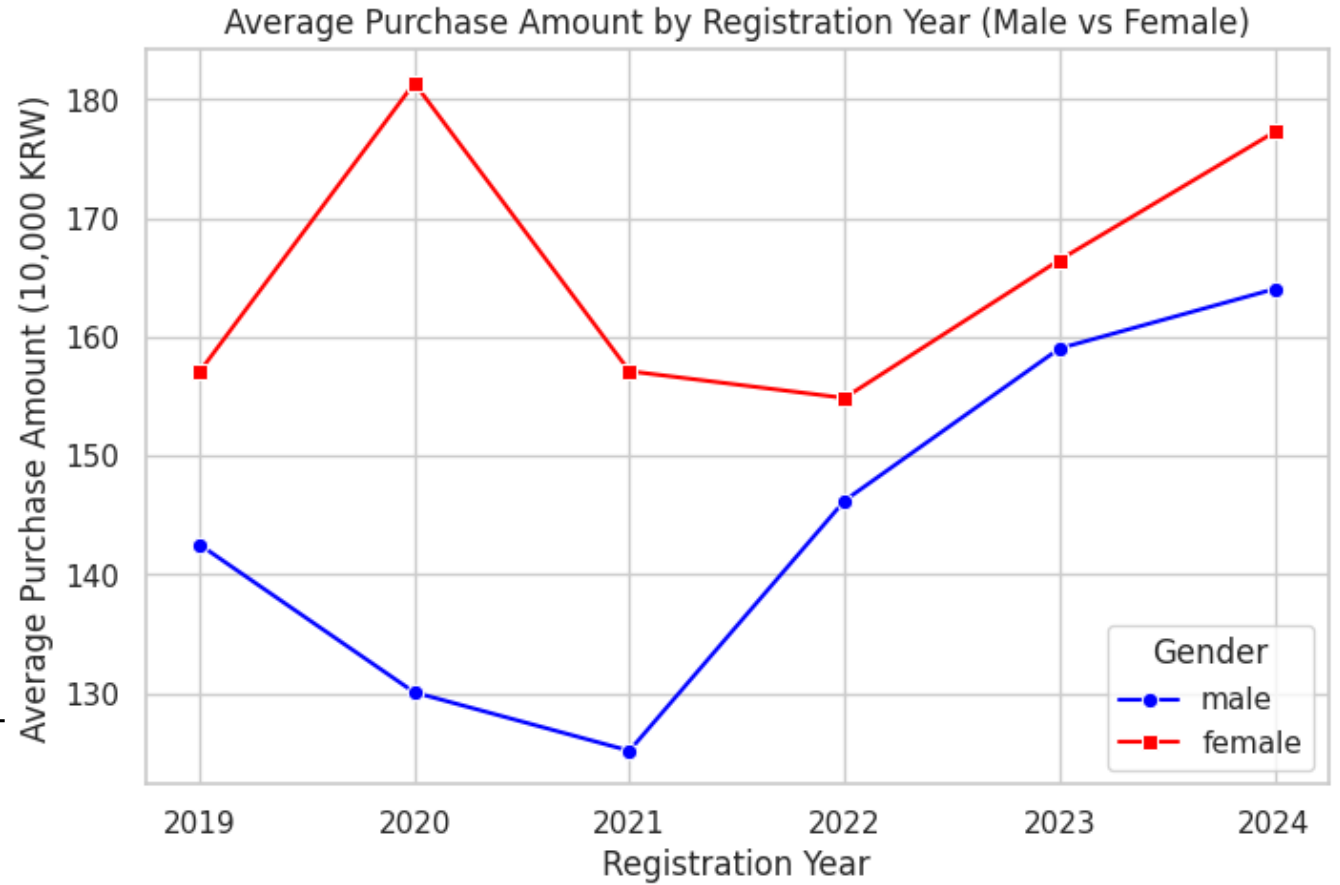
# 가입연도별 평균 구매금액 변화

## 여성 구매 패턴

- 2020년에 최고 구매금액 기록, 이후 감소
- 2021~2022년까지 감소 후 2023년부터 회복세
- 전반적으로 남성보다 높은 구매금액 유지

## 남성 구매 패턴

- 2019년부터 2021년까지 평균 구매금액 감소
- 2021년 최저점 도달 후 2022년부터 꾸준한 상승
- 2023년부터 여성과의 구매금액 격차가 크게 줄어들음



# 가입연도별 평균 구매 금액 비교 인사이트

- **2021년, 남녀 모두 평균 구매 금액 최저점 기록**
  - → COVID-19 팬데믹 장기화 영향으로 소비 심리 위축
  - 외출 감소 및 경제 불안정이 비필수 소비 축소로 이어진 것으로 해석
- **2022년 이후 구매 금액 회복세**
  - 사회적 거리두기 완화 및 소비 심리 회복 반영
  - 특히 남성은 2021~2022년 사이 큰 폭의 증가 → 소비 회복 성향 가능성
- **여성은 전반적으로 높은 구매 금액 유지**
  - 팬데믹 중에도 비교적 안정적인 소비 활동 지속
- **마케팅 시사점**
  - 팬데믹 같은 외부 요인은 소비 패턴에 직접적 영향
  - 향후 유사 위기 상황 시 성별에 따른 소비 반응 차이 고려 필요

QnA

감사합니다.