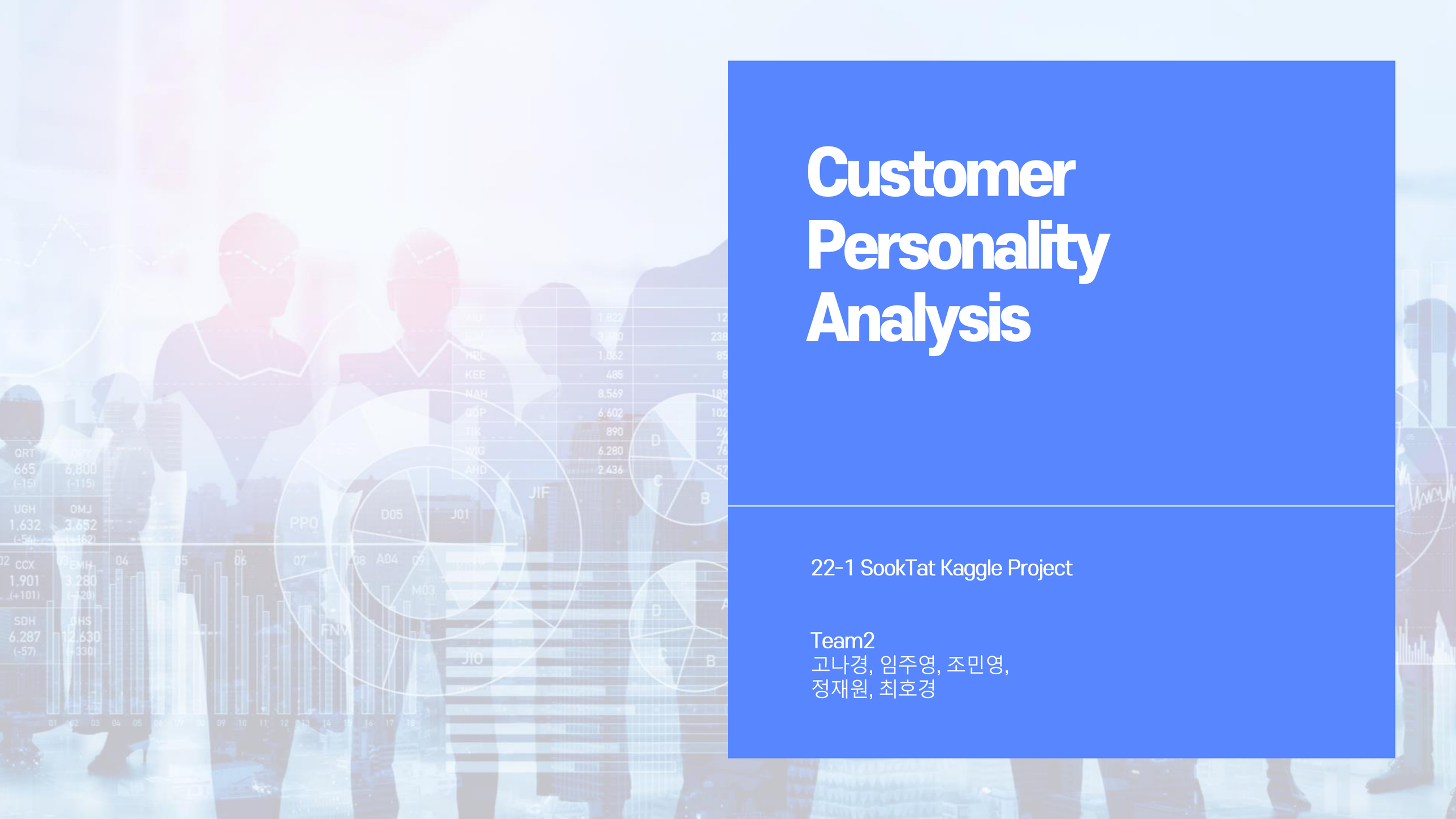
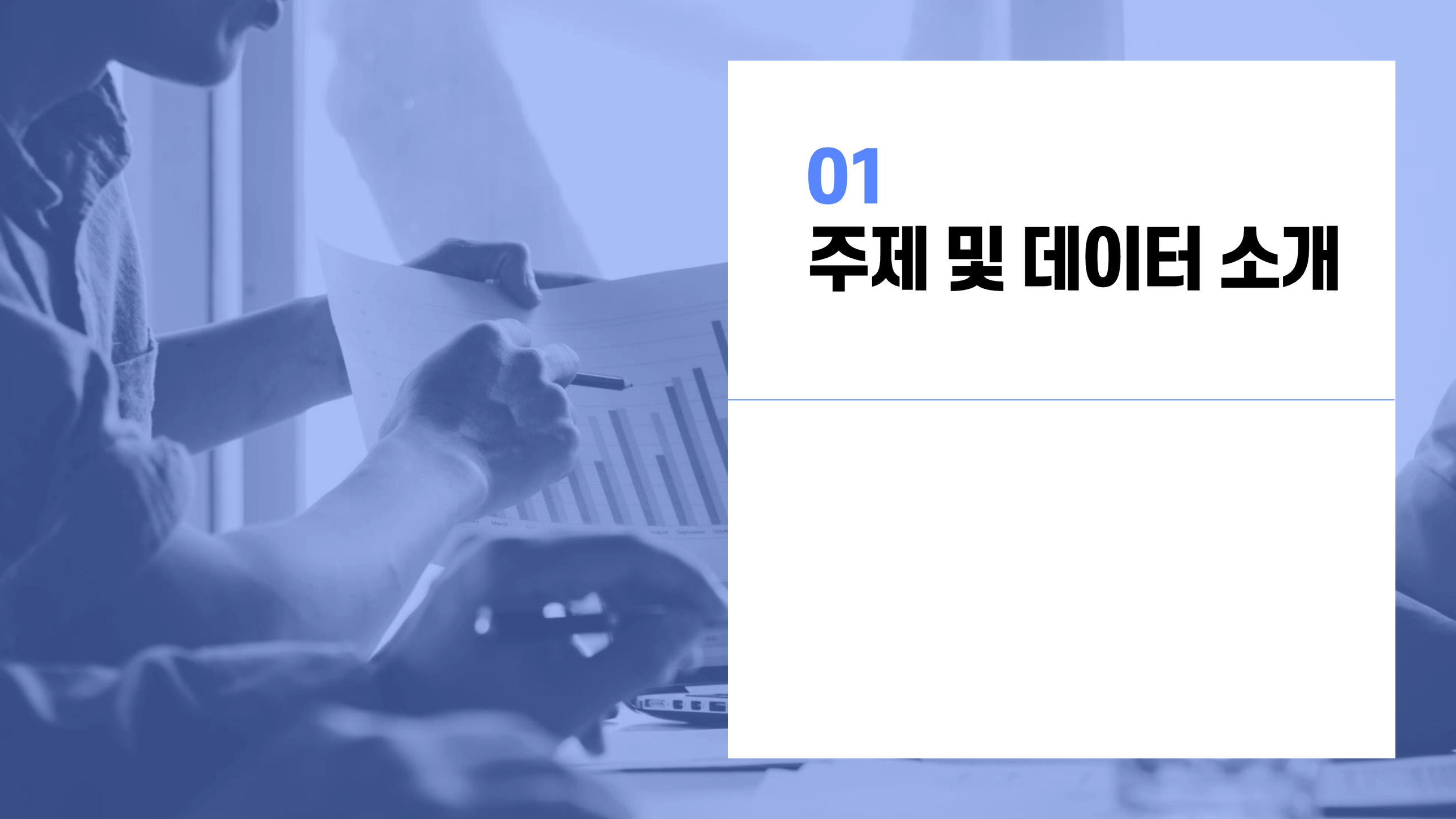


Customer Personality Analysis

22-1 SookTat Kaggle Project

Team2
고나경, 임주영, 조민영,
정재원, 최호경



01
주제 및 데이터 소개

Customer Personality Analysis

■ 고객 특성 분석 (분류)

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

고객 특성 분석은 회사의 이상적인 고객에 대한 자세한 분석입니다. 기업이 고객을 더 잘 이해하고 다양한 유형의 고객의 특정 요구, 행동 및 우려 사항에 따라 제품을 더 쉽게 수정할 수 있도록 도와줍니다.

고객 성격 분석은 기업이 다양한 유형의 고객 세그먼트의 대상 고객을 기반으로 제품을 수정하는 데 도움이 됩니다. 예를 들어, 회사 데이터베이스의 모든 고객에게 신제품을 마케팅하는 데 돈을 쓰는 대신 회사는 제품을 구매할 가능성이 가장 높은 고객 세그먼트를 분석한 다음 해당 특정 세그먼트에서만 제품을 마케팅할 수 있습니다.

Customer Personality Analysis

■ ‘marketing_campaign.csv’

```
data = pd.read_csv("marketing_campaign.csv", sep="₩t")
data
```

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	7	0
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	5	0
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4	0
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6	0
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2235	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	46	709	...	5	0
2236	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	56	406	...	7	0
2237	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	91	908	...	6	0
2238	8235	1956	Master	Together	69245.0	0	1	24-01-2014	8	428	...	3	0
2239	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	40	84	...	7	0

2240 rows × 29 columns

Customer Personality Analysis

People

- ID : Customer's unique identifier
- Year_Birth : Customer's birth year
- Education : Customer's education level
- Marital_Status : Customer's marital status
- Income: Customer's yearly household income
- Kidhome : Number of children in customer's household
- Teenhome : Number of teenagers in customer's household
- Dt_Customer : Date of customer's enrollment with the company
- Recency : Number of days since customer's last purchase
- Complain : 1 if the customer complained in the last 2 years, 0 otherwise

Product

- MntWines: Amount spent on wine in last 2 years
- MntFruits: Amount spent on fruits in last 2 years
- MntMeatProducts: Amount spent on meat in last 2 years
- MntFishProducts: Amount spent on fish in last 2 years
- MntSweetProducts: Amount spent on sweets in last 2 years
- MntGoldProds: Amount spent on gold in last 2 years

Promotion

- NumDealsPurchases: Number of purchases made with a discount
- AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
- AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
- AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
- AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
- AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
- Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

Place

- NumWebPurchases: Number of purchases made through the company’s website
- NumCatalogPurchases: Number of purchases made using a catalogue
- NumStorePurchases: Number of purchases made directly in stores
- NumWebVisitsMonth: Number of visits to company’s website in the last month

02
프로젝트 진행 순서

- 데이터 전처리 (EDA)
- 모델 공부
- 모델 적용
- 하이퍼파라미터 조정

Customer Personality Analysis

■ GitHub을 통해 코드 공유

EDA2	Add files via upload	4 hours ago
README.md	Update README.md	8 days ago

README.md

Customer_Personality_Analysis

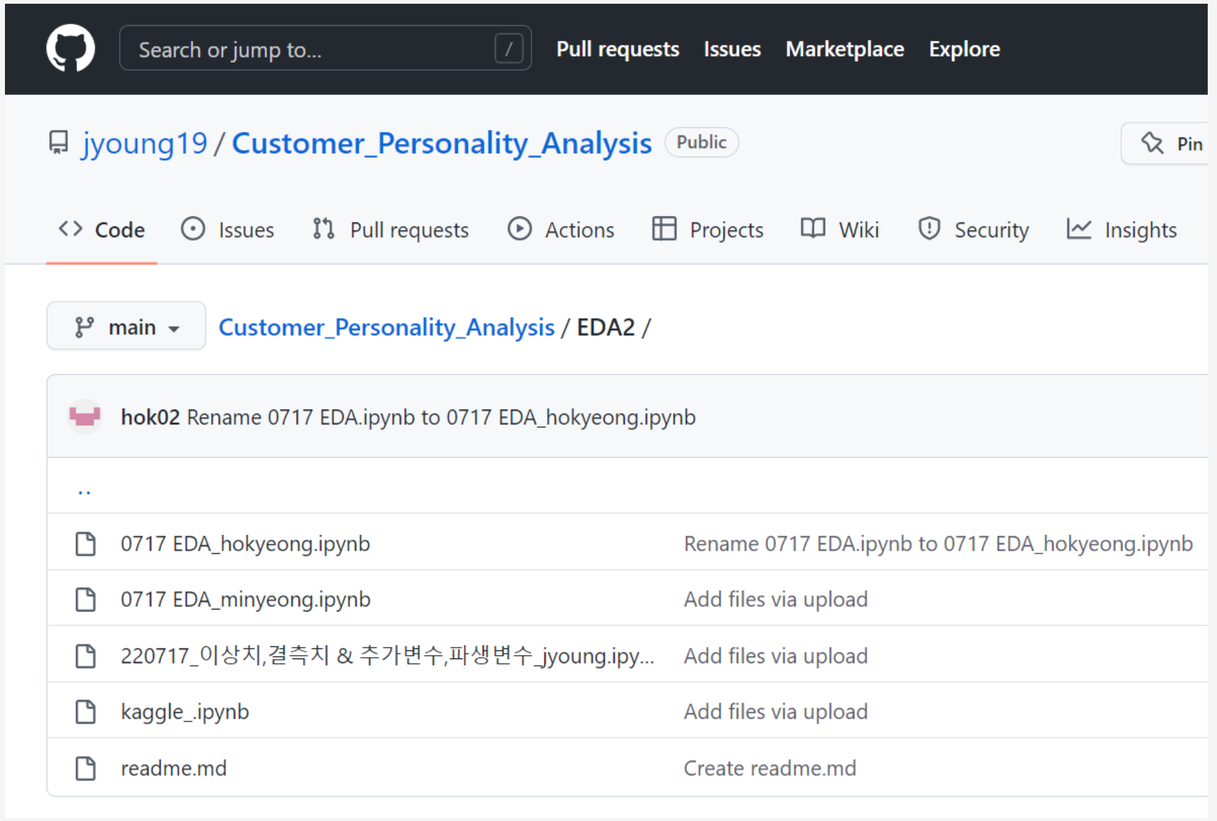
- 📊 22-1 SookTat Kaggle Project 📊

Subject

- 🛍️ 고객 특성 분류 🛍️
- https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

Members

- 2기 : 고나경, 임주영, 조민영
- 3기 : 정재원, 최호경



03
EDA

- 데이터 확인
- 이상치, 결측치 처리
- 추가변수, 파생변수 생성
- 변수간 상관관계 파악

Customer Personality Analysis

■ 데이터 확인

In [23]:
```
df.describe().T
```

Out[23]:

	count	mean	std	min	25%	50%	75%	max
ID	2240.0	5592.159821	3246.662198	0.0	2828.25	5458.5	8427.75	11191.0
Year_Birth	2240.0	1968.805804	11.984069	1893.0	1959.00	1970.0	1977.00	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	35303.00	51381.5	68522.00	666666.0
Kidhome	2240.0	0.444196	0.538398	0.0	0.00	0.0	1.00	2.0
Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
Recency	2240.0	49.109375	28.962453	0.0	24.00	49.0	74.00	99.0
MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0
MntFishProducts	2240.0	37.525446	54.628979	0.0	3.00	12.0	50.00	259.0
MntSweetProducts	2240.0	27.062946	41.280498	0.0	1.00	8.0	33.00	263.0
MntGoldProds	2240.0	44.021875	52.167439	0.0	9.00	24.0	56.00	362.0
NumDealsPurchases	2240.0	2.325000	1.932238	0.0	1.00	2.0	3.00	15.0
NumWebPurchases	2240.0	4.084821	2.778714	0.0	2.00	4.0	6.00	27.0
NumCatalogPurchases	2240.0	2.662054	2.923101	0.0	0.00	2.0	4.00	28.0
NumStorePurchases	2240.0	5.790179	3.250958	0.0	3.00	5.0	8.00	13.0
NumWebVisitsMonth	2240.0	5.316518	2.426645	0.0	3.00	6.0	7.00	20.0

In [22]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 29 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   ID                   2240 non-null   int64  
 1   Year_Birth           2240 non-null   int64  
 2   Education            2240 non-null   object 
 3   Marital_Status       2240 non-null   object 
 4   Income               2216 non-null   float64
 5   Kidhome              2240 non-null   int64  
 6   Teenhome             2240 non-null   int64  
 7   Dt_Customer          2240 non-null   object 
 8   Recency              2240 non-null   int64  
 9   MntWines             2240 non-null   int64  
 10  MntFruits            2240 non-null   int64  
 11  MntMeatProducts      2240 non-null   int64  
 12  MntFishProducts      2240 non-null   int64  
 13  MntSweetProducts     2240 non-null   int64  
 14  MntGoldProds         2240 non-null   int64  
 15  NumDealsPurchases    2240 non-null   int64  
 16  NumWebPurchases      2240 non-null   int64  
 17  NumCatalogPurchases  2240 non-null   int64  
 18  NumStorePurchases    2240 non-null   int64  
 19  NumWebVisitsMonth    2240 non-null   int64  
 20  AcceptedCmp3         2240 non-null   int64  
 21  AcceptedCmp4         2240 non-null   int64  
 22  AcceptedCmp5         2240 non-null   int64  
 23  AcceptedCmp1         2240 non-null   int64  
 24  AcceptedCmp2         2240 non-null   int64  
 25  Complain             2240 non-null   int64  
 26  Z_CostContact        2240 non-null   int64  
 27  Z_Revenue            2240 non-null   int64  
 28  Response             2240 non-null   int64  
dtypes: float64(1), int64(25), object(3)
memory usage: 507.6+ KB
```

2440 rows x 29 columns

Customer Personality Analysis

■ 이상치, 결측치 처리

In [61]:
```
customer_df.isnull().sum()
```

Out[61]:

ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	24
Kidhome	0
Teenhome	0
Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
NumDealsPurchases	0
NumWebPurchases	0
NumCatalogPurchases	0
NumStorePurchases	0
NumWebVisitsMonth	0
AcceptedCmp3	0
AcceptedCmp4	0
AcceptedCmp5	0
AcceptedCmp1	0
AcceptedCmp2	0
Complain	0
Z_CostContact	0
Z_Revenue	0
Response	0

dtype: int64

In [62]:
```
customer_df.dropna(inplace=True)
```

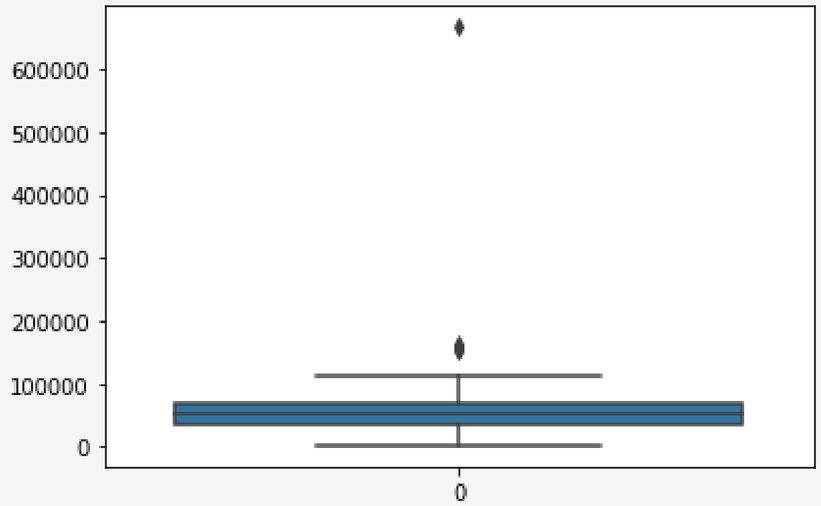
In [9]:
```
# 4분위수, IQR 계산

q1 = customer_df['Income'].quantile(0.25)
q2 = customer_df['Income'].quantile(0.5)
q3 = customer_df['Income'].quantile(0.75)

iqr = q3 - q1
iqr
```

Out[9]: 33219.0

In [10]:
```
# outlier cutoff, lower/upper bound 계산

cut_off = iqr * 1.5

lower = q1 - cut_off
upper = q3 + cut_off

print(lower)
print(upper)
```

```
-14525.5
118350.5
```

In [11]:
```
# 1사분위와 4사분위에 속해있는 데이터 각각 저장

data1 = customer_df[customer_df['Income'] > upper]
data2 = customer_df[customer_df['Income'] < lower]
```



In [21]:
```
customer_df[customer_df['Income'] >= 153924]
```

Out[21]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4
164	8475	1973	PhD	Married	157243.0	0	1	01-03-2014	98	20	...	0	0	0
617	1503	1976	PhD	Together	162397.0	1	1	03-06-2013	31	85	...	1	0	0
655	5555	1975	Graduation	Divorced	153924.0	0	0	07-02-2014	81	1	...	0	0	0
687	1501	1982	PhD	Married	160803.0	0	0	04-08-2012	21	55	...	0	0	0
1300	5336	1971	Master	Together	157733.0	1	0	04-06-2013	37	39	...	1	0	0
1653	4931	1977	Graduation	Together	157146.0	0	0	29-04-2013	13	1	...	1	0	0
2132	11181	1949	PhD	Married	156924.0	0	0	29-08-2013	85	2	...	0	0	0
2233	9432	1977	Graduation	Together	666666.0	1	0	02-06-2013	23	9	...	6	0	0

8 rows × 29 columns

Customer Personality Analysis

■ 추가변수, 파생변수 생성

```
#제품에 지출된 총액
mk['TotalMntSpent'] = mk['MntWines'] + mk['MntFruits'] + mk['MntMeatProducts'] + mk['MntFishProducts'] + mk['MntSweetProducts'] + mk['MntGoldProducts']
```

```
#총 구매 건수
mk['TotalNumPurchases'] = mk['NumWebPurchases'] + mk['NumCatalogPurchases'] + mk['NumStorePurchases'] + mk['NumDealsPurchases']
```

```
# 승인된 총 캠페인 수
mk['Total_Acc_Cmp'] = mk['AcceptedCmp1'] + mk['AcceptedCmp2'] + mk['AcceptedCmp3'] + mk['AcceptedCmp4'] + mk['AcceptedCmp5'] + mk['Response']
```

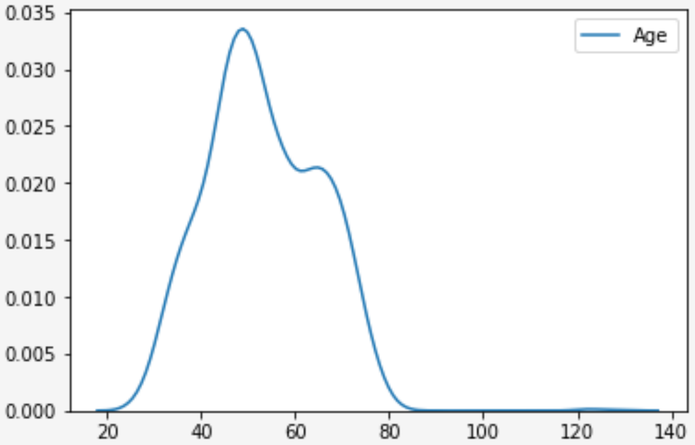
```
# 결혼상태에 따른 파트너 유무
mk["Partner"]=mk["Marital_Status"].replace({"Married":"Yes", "Together":"Yes", "Absurd":"No", "Widow":"No", "YOLO":"No", "Divorced":"No", "Single":"No","Alone
```

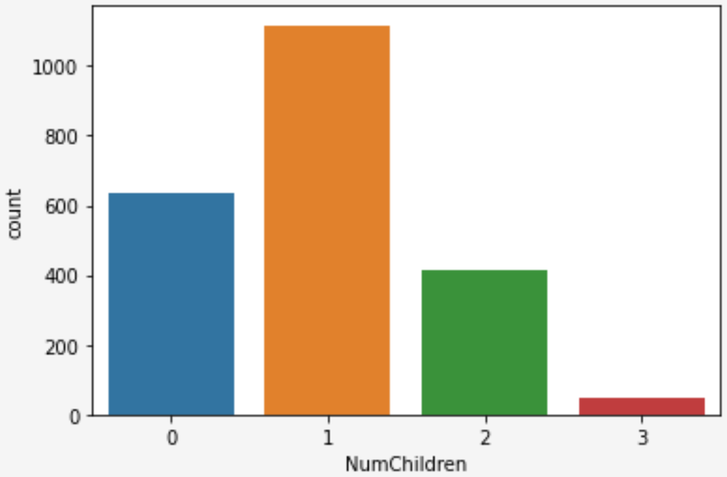
```
# 가구 내 총 자녀 수
mk["Children"] = mk["Kidhome"] +  mk["Teenhome"]
```

```
# 가족수
mk["Family_Size"] = mk["Partner"].replace({"No": 1, "Yes":2})+ mk["Children"]
```

```
#부모인지 아닌지
mk["Is_Parent"] = np.where(mk.Children> 0, 1, 0)
```

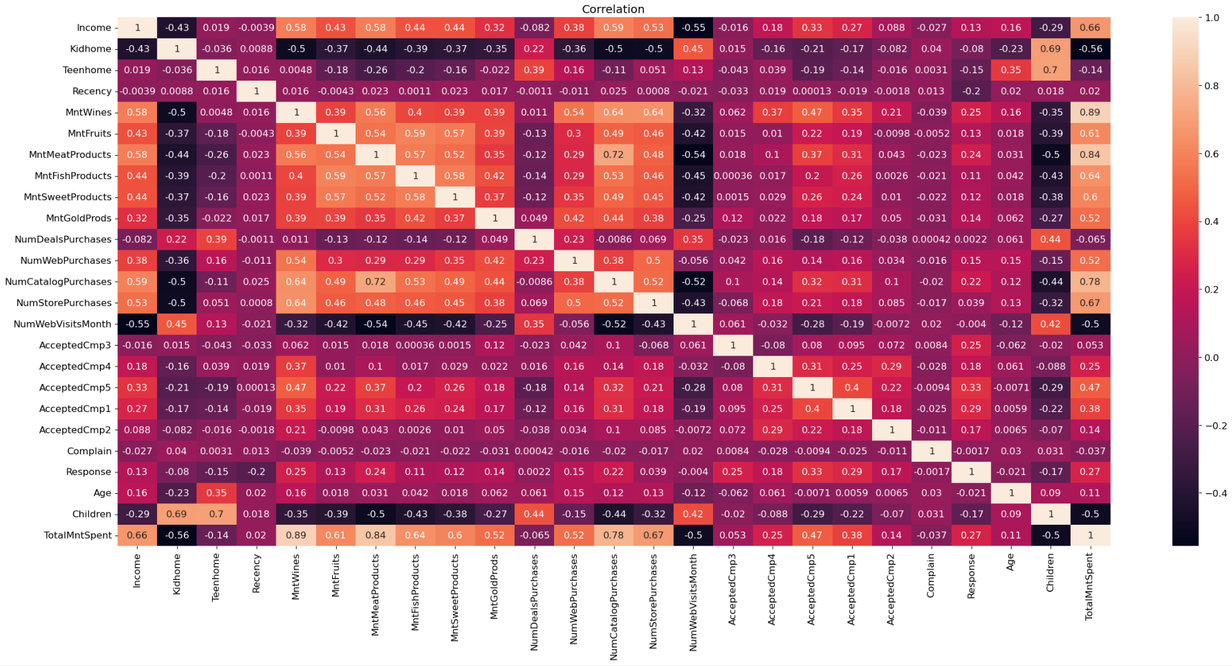
```
# 교육 수준을 세 그룹으로 세분화
mk["Education_Level"]=mk["Education"].replace({"Basic":"Undergraduate","2n Cycle":"Undergraduate", "Graduation":"Graduate", "Master":"Postgraduate", "PhD":"Po
```

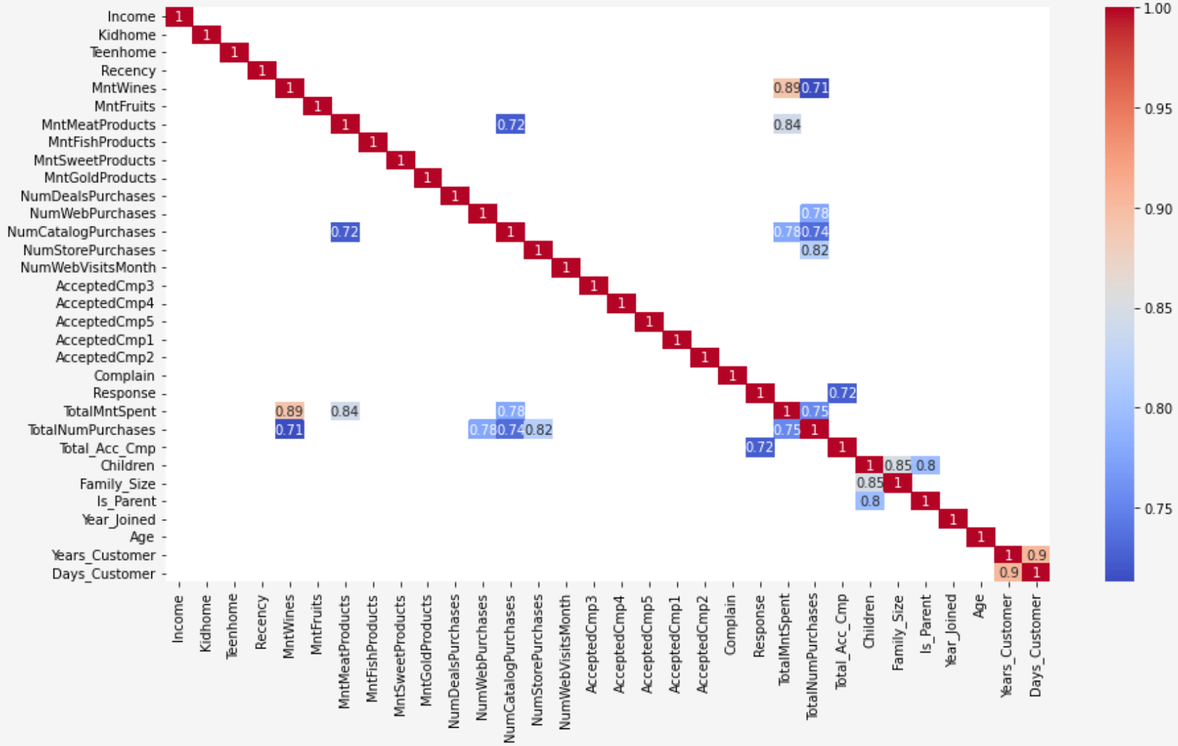




Customer Personality Analysis

■ 상관관계 분석





THANK YOU

https://github.com/jyoung19/Customer_Personality_Analysis