



1 유형 서비스이력 분석 - 정답 및 코드 (최종 수정본)

```
import numpy as np
import pandas as pd

file = '/Users/yeongjunjeon/python/집중강화/1 유형_서비스데이터.csv'

df = pd.read_csv(file)
print(df)

...

문제 1. 결측치 처리
• `사용개월수`의 결측치는 전체 평균값으로 채워라.
• `서비스만족도`의 결측치는 `고객성별`과 `제품군`별 평균 만족도로 채워라.
• `고객연령` 결측치는 같은 `지역`과 `고객성별` 그룹의 중앙값으로 채워라.
...
df['사용개월수'] = df['사용개월수'].fillna(df['사용개월수'].mean())
satis_mean = df.groupby(['고객성별', '제품군'])['서비스만족도'].transform('mean')
df['서비스만족도'] = df['서비스만족도'].fillna(satis_mean)
age_mid = df.groupby(['지역', '고객성별'])['고객연령'].transform('median')
df['고객연령'] = df['고객연령'].fillna(age_mid)

...

문제 2. 파생변수 생성
• `서비스까지_소요일수` : `생산일시`부터 `서비스접수일시`까지의 일수
• `고객연령대` : `고객연령`을 기준으로 10 단위 연령대 구간 생성 (예: 20 대, 30 대...)
• `서비스시간대` : `서비스접수일시` 기준, 시간(hour)을 추출하여
* 06~11 시는 '오전',
* 12~17 시는 '오후',
* 그 외는 '야간'으로 분류
...
df['생산일시'] = pd.to_datetime(df['생산일시'])
df['서비스접수일시'] = pd.to_datetime(df['서비스접수일시'])
df['서비스까지_소요일수'] = (df['서비스접수일시'] - df['생산일시']).dt.days

customer_bins = [19, 29, 39, 49, 59, 69, 79, float('inf')]
customer_labels = ['20 대', '30 대', '40 대', '50 대', '60 대', '70 대', '80 대이상']
df['고객연령대'] = pd.cut(df['고객연령'], bins = customer_bins, labels = customer_labels)
# print(df[['고객연령', '고객연령대']])
```

```

service_hour_bins = [6,11,16, float('inf')]
service_labels = ['오전','오후','야간']
df['서비스시간대'] = pd.cut(df['서비스접수일시'].dt.hour, bins = service_hour_bins,
labels = service_labels)
# print(df[['서비스접수일시','서비스시간대']])

print(df.columns)

...

```

문제 3. 서비스 이력 집계

- `고객성별`, `제품군`, `고객연령대` 별로
- * 평균 `서비스만족도`
- * 평균 `서비스까지_소요일수`
- * 총 `재발횟수`
- 위 집계 결과에서 `서비스만족도` 평균이 가장 낮은 1 건을 출력하시오.

```

...

```

피벗 테이블 생성

```

df_pivot = pd.pivot_table(df, index=['고객성별', '제품군', '고객연령대'],
                           values=['서비스만족도', '서비스까지_소요일수', '재발횟수'],
                           aggfunc={'서비스만족도': 'mean', '서비스까지_소요일수':
                           'mean', '재발횟수': 'sum'}).reset_index()

```

```

print(df_pivot)

```

```

df_group = df.groupby(['고객성별', '제품군', '고객연령대']).agg({'서비스만족도' :
'mean', '서비스까지_소요일수' : 'mean', '재발횟수' : 'sum'})
print(df_group)

```

```

print(df_group['서비스만족도'].idxmin())

```

```

...

```

문제 4. 불량 집중 고객군 탐색

- 다음 조건을 만족하는 고객 수를 구하시오.
- * `재발횟수`가 2 회 이상
- * `서비스만족도`가 전체 평균보다 낮음
- * `리콜조치유무`가 'N'
- * `사용개월수`가 6 개월 이상

```

...

```

```

cnt = len(df[(df['재발횟수'] >=2) & (df['서비스만족도'] < df['서비스만족도'].mean()) &
& (df['리콜조치유무'] == 'N') & (df['사용개월수'] >= 6)])
print(f'fail_focus_customer : {cnt}명')

```

```

...
문제 5. 이상치 및 통계 분석
• `서비스까지_소요일수`가 3 $\sigma$  이상인 값을 이상치로 판단하고, 해당 건수는 몇 건인가?
• `사용개월수`와 `서비스만족도` 간의 피어슨 상관계수를 소수점 셋째자리까지
구하시오.

...
std = df['서비스까지_소요일수'].std()
mean = df['서비스까지_소요일수'].mean()
upper = mean + 3*std

outlier = len(df[(df['서비스까지_소요일수'] >= upper)])
print(f'이상치 서비스건수 : {outlier}')

corr = round(df[['사용개월수','서비스만족도']].corr()['사용개월수']['서비스만족도'],3)
print(f"'사용개월수'와 `서비스만족도` 간의 피어슨 상관계수 : {corr}")

...
문제 6. 최종 분석
• `고객성별`, `제품군`, `고객연령대` 별로 `서비스만족도`가 가장 낮은 조합을 구하고,
해당 조합의 다음 정보를 표 형태로 출력하시오:
• | 고객성별 | 제품군 | 연령대 | 평균만족도 | 평균사용개월수 | 평균소요일수 |
재발횟수합계 |

...
df_gr_min = df.groupby(['고객성별','제품군','고객연령대'],
observed=True)['서비스만족도'].min().idxmin()
print(df_gr_min)
gender = df_gr_min[0]
product = df_gr_min[1]
age = df_gr_min[2]

df_final = df[(df['고객성별'] == gender) & (df['제품군'] == product) &
(df['고객연령대'] == age)]
df_result = pd.DataFrame({'고객성별' : gender, '제품군' :
product,'고객연령대' : age,
'평균만족도' : [df_final['서비스만족도'].mean()],
'평균사용개월수' : [df_final['사용개월수'].mean()],
'평균소요일수' : [df_final['서비스까지_소요일수'].mean()],
'재발횟수합계' : [df_final['재발횟수'].sum()]})


print(df_result.to_string(index=False))

print(df_result.to_markdown(index=False))

```

고객성별 제품군 고객연령대

남 TV 20 대 5.000000 306.000 0

30 대 NaN NaN 0

40 대 2.600000 220.000 9

50 대 2.500000 49.500 2

60 대 3.500000 198.500 8

...

여 청소기 40 대 3.240000 29.500 8

50 대 3.520000 178.500 2

60 대 3.244444 257.375 4

70 대 3.250000 268.250 4

80 대이상 2.000000 15.000 2

[84 rows x 3 columns]

('남', '건조기', '30 대')

fail_focus_customer : 23 명

이상치 서비스건수 : 2

'사용개월수'와 '서비스만족도' 간의 피어슨 상관계수 : 0.065

('남', 'TV', '40 대')

고객성별 제품군 고객연령대 평균만족도 평균사용개월수 평균소요일수 재발횟수합계

남 TV 40 대 2.6 11.2 220.0 9

| 고객성별 | 제품군 | 고객연령대 | 평균만족도 | 평균사용개월수 | 평균소요일수 |
재발횟수합계 |

|:-----|:-----|:-----|:-----|:-----|:-----|

| 남 | TV | 40 대 | 2.6 | 11.2 | 220 | 9 |