

CONTENTS

Chapter→

I 요구사항 정의서

Chapter→

II 프로젝트 수행 내역

Chapter→

III Lessons & Learned

1. 요구사항 정의서

1 2

미션별 Action Planning

각 미션을 수행하기 위한 데이터를 확인하고, 분석 columns을 선정하는 방향으로 요구사항을 충실히 이행할 수 있도록 사전 작업을 진행하였음.

프로젝트 수행 과제

1 : **성별/나이에 따른 사용금액 분석**

2 : **수입에 따른 사용금액 분석**

3 : **수입에 따른 카드한도 분석**

데이터의 확인 및 분석 columns 선정

```
df = pd.read_csv('data/데이터.csv', encoding='EUC-KR')  
df.tail(10)
```

	Cstno	Age	Sex	Income	USE_Month	Credit_Limit	Total_Trans_Amt	Total_Trans_Ct
10117	712503408	57	M	6~8천만 원미만	40	17925000	17498000	111
10118	713755458	50	M	6~8천만 원미만	36	9959000	10310000	63
10119	716893683	55	F	Unknown	47	14657000	6009000	53
10120	710841183	54	M	4~6천만 원미만	34	13940000	15577000	114
10121	713899383	56	F	2천만원 미만	50	3688000	14596000	120
10122	772366833	50	M	2~4천만 원미만	40	4003000	15476000	117
10123	710638233	41	M	2~4천만 원미만	25	4277000	8764000	69
10124	716506083	44	F	2천만원 미만	36	5409000	10291000	60
10125	717406983	30	M	2~4천만 원미만	36	5281000	8395000	62
10126	714337233	43	F	2천만원 미만	25	10388000	10294000	61

1. 요구사항 정의서

1 2

미션별 Action Planning

각 프로젝트 미션을 수행하기 위하여 필요한 사전 작업을 탐색하고, 세부 진행 방향을 설계하고 해당 작업을 진행하였음.

○ 프로젝트 수행 과제 분류

1 : 성별/나이에 따른 카드 사용금액 분석

2 : 수입에 따른 사용금액 분석

3 : 수입에 따른 카드한도 분석

Action Plan 설계

- 데이터 Segmentation

- 나이(Age) 데이터의 Segmentation

연령대를 10 단위로 구분하여 카드 사용금액을 분석

- 카드 사용금액의 Segmentation

카드 사용금액을 150만원 미만, 300만원 미만, 450만원 미만, 700만원 미만, 700만원 초과 군집으로 구분

- 카드 한도의 Segmentation

카드 한도를 250만원 미만, 500만원 미만, 800만원 미만, 1,200만원 미만, 1,900만원 미만, 2,800만원 초과 군집으로 구분

- dict 타입 생성 관리

Dict 타입의 key / values 형태로 생성하여 데이터 조합

2. 프로젝트 수행 내역

* Mission 1 Segmentation 진행

- 나이 : 10살 간격으로 seg 분류 (ex : 10대, 20대, 30대, 40대, 50대 ...)
- 카드 사용금액 : 5개 사용량 seg 로 분류 (ex : 150만원 미만, 300만원 미만, 450만원 미만, 700만원 미만, 700만원 초과)

- 나이 seg 분류

```
age_series = df.Age.dropna().apply(lambda age: math.floor(age / 10) * 10)
age_series.name = "Age_Group"
age_series
```

```
0      40
1      40
2      50
3      40
4      40
...
10122   50
10123   40
10124   40
10125   30
10126   40
```

Name: Age_Group, Length: 10127, dtype: int64

	Sex	Total_Trans_Amt	Age_Group
0	M	1144000	40
1	F	1291000	40
2	M	1887000	50
3	F	1171000	40
4	M	816000	40
...
10122	M	15476000	50
10123	M	8764000	40
10124	F	10291000	40
10125	M	8395000	30
10126	F	10294000	40

- 카드 사용량 seg 분류

```
for i in np.arange(len(df)):
    if inform1_df['Total_Trans_Amt'].values[i] < 1500000:
        inform1_df['Use_Amt'].values[i] = '1'
    elif inform1_df['Total_Trans_Amt'].values[i] < 3000000:
        inform1_df['Use_Amt'].values[i] = '2'
    elif inform1_df['Total_Trans_Amt'].values[i] < 4500000:
        inform1_df['Use_Amt'].values[i] = '3'
    elif inform1_df['Total_Trans_Amt'].values[i] < 7000000:
        inform1_df['Use_Amt'].values[i] = '4'
    # elif inform1_df['Total_Trans_Amt'].values[i] < 10000000:
    #     inform1_df['Range'].values[i] = 'E'
    else: inform1_df['Use_Amt'].values[i] = '5'
```

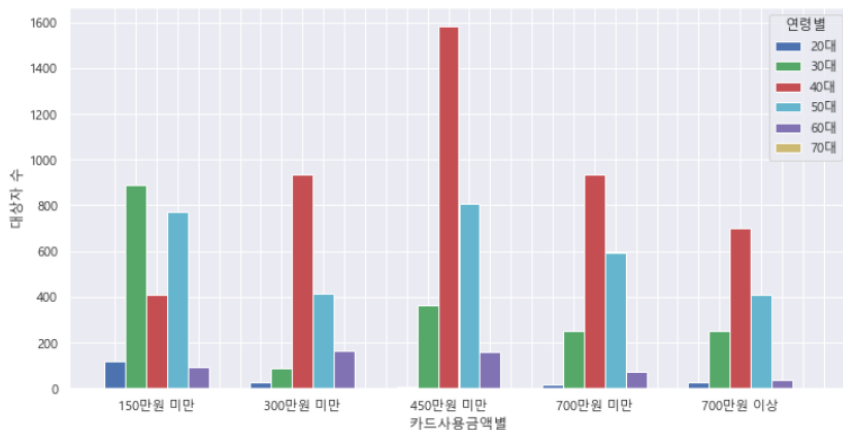
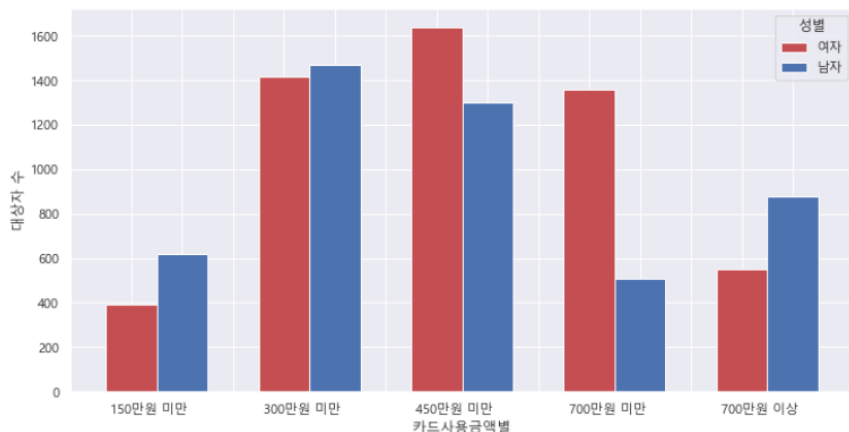
inform1_df

	Sex	Total_Trans_Amt	Use_Amt
0	M	1144000	1
1	F	1291000	1
2	M	1887000	2
3	F	1171000	1
4	M	816000	1
...
10122	M	15476000	5
10123	M	8764000	5
10124	F	10291000	5
10125	M	8395000	5
10126	F	10294000	5

2. 프로젝트 수행 내역

* Mission 1 차트 작성 및 간략 분석

- 카드 사용금액 seg 를 x축으로 하여 성별 / 연령별 대상자 수를 y축에 표현



Mission 1

카드 사용금액별 남/녀 사용자 수 비교

1. 여자의 경우 카드 사용금액 구간별 정규분포를 이루고 있음. 가장 많은 사용자 그룹은 450만원 미만이며 가장 적은 그룹은 150만원 미만으로 대부분의 사용자가 150 ~ 700만원의 카드를 사용하고 있는 것으로 나타남.
2. 450이상 700만원 미만 사용자 그룹은 남자에 비해 여자 사용자 수가 2배 이상 많은 것으로 확인.

3. 남자의 경우 450만원 미만 사용자가 대부분의 비중을 차지하나 700만원 이상의 초고액 사용자그룹은 여자에 비해 높음.

1. 전 연령 중, 40대의 카드 사용의 주축을 이룸
2. 30대, 50대가 그 다음으로 많은 사용자 그룹을 보유

2. 프로젝트 수행 내역

* Mission 2 소득구간별 사용금액 분석

- 소득구간별(Income), 사용금액 구간별(Use_Amt) 대상자 수 집계(count)를 위하여 dict key/values 형태로 구현

- 사용금액 구간별 대상자수 dict 타입화

```
from functools import *  
  
def count_columns(result_dic, element) :  
    if result_dic.get(element) :  
        result_dic[element] += 1  
    else :  
        result_dic[element] = 1  
  
    return result_dic  
  
def get_group_count(df_column) :  
    return reduce(count_columns, df_column, dict())
```

```
result = get_group_count(inform1_df['Use_Amt'])  
result
```

```
{ '150만원미만': 1010,  
  '300만원미만': 2885,  
  '450만원미만': 2938,  
  '700만원미만': 1869,  
  '700만원이상': 1425}
```

```
dict_element = result['150만원미만']  
dict_element
```

```
{ '4~6천만원미만': 173,  
  '2천만원미만': 269,  
  '2~4천만원미만': 175,  
  '8천만원이상': 100,  
  '6~8천만원미만': 209,  
  'Unknown': 84}
```

```
dict_values = [dict_element[key] for key in dict_element.keys()]  
dict_values
```

```
[173, 269, 175, 100, 209, 84]
```

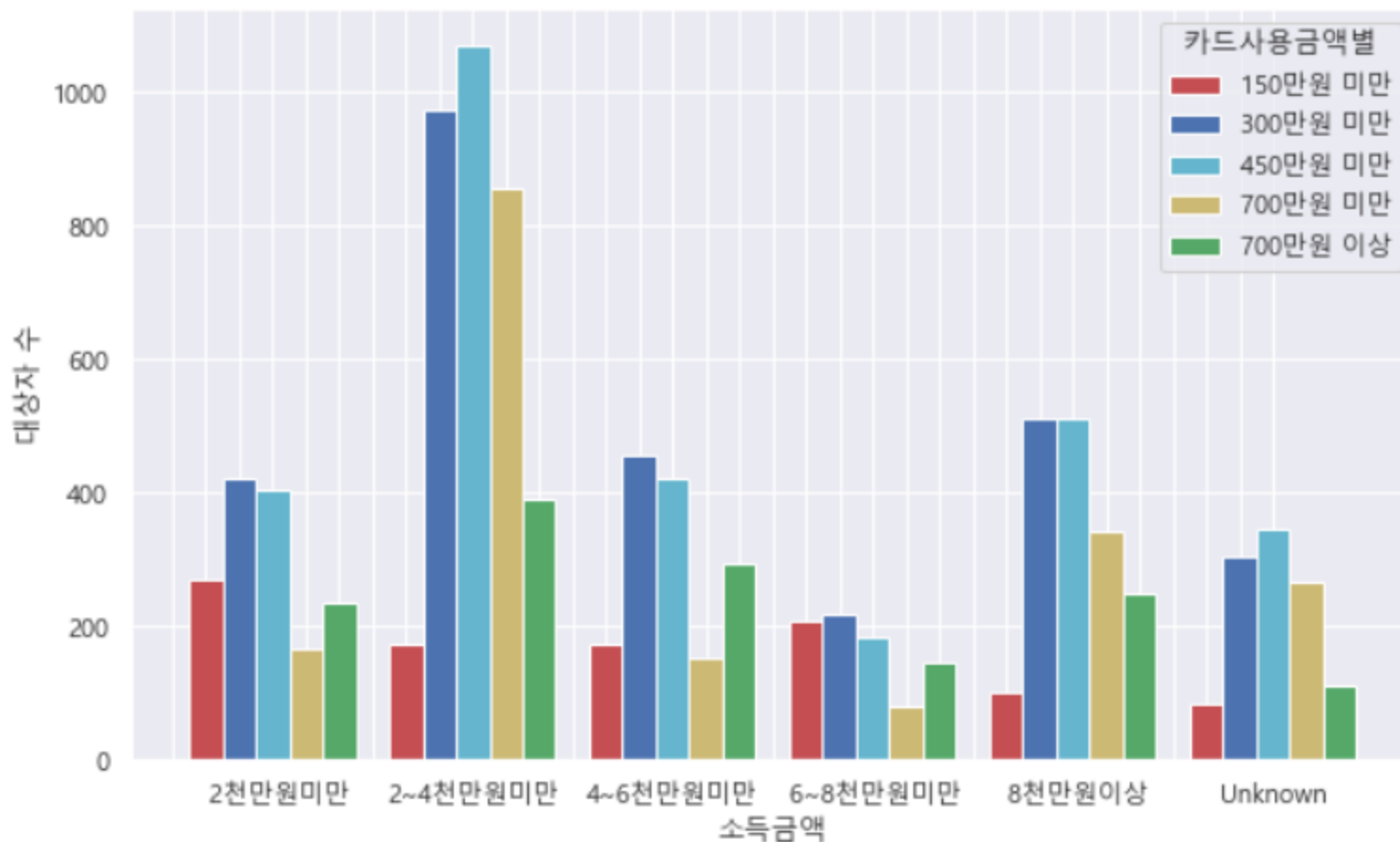
- 사용금액 구간별 {소득구간 : 대상자 수}를 Dict key / values 형태로 구현

```
def get_group_count(inform1_df_columns) :  
  
    print('len(inform1_df_columns) :', len(inform1_df_columns))  
  
    def count_columns(result_dic, element_tuple) :  
        element1, element2 = element_tuple  
        # print(element1, ': ', element2)  
  
        if not (result_dic.get(element1)) :  
            result_dic[element1] = {}  
  
        if result_dic[element1].get(element2) :  
            result_dic[element1][element2] += 1  
        else :  
            result_dic[element1][element2] = 1  
  
        return result_dic  
  
    result = reduce(count_columns, inform1_df_columns, dict())  
    return result
```

2. 프로젝트 수행 내역

* Mission 2 차트 작성 및 간략 분석

- 각 소득 구간별 카드 사용량 seg를 막대 차트로 표현하여 소득에 따라 카드 사용량이 어떠한 차이를 보이는지 확인하고자 함.



Mission 1

카드 사용금액별 남/녀 사용자 수 비교

1. 2~4천만원 소득자의 카드 사용이 제일 활발
2. 전반적으로 소득구간별 카드 사용금액도 정규분포를 가짐.
3. 150~450만원 미만 사용자 그룹은 모든 소득 구간에서 가장 많이 사용하는 핵심 그룹임.

* Mission 3 소득구간별 카드한도 분석

- 소득구간별(Income), 카드 한도 구간별(Limit_Range) 대상자 수 집계(count)를 위하여 dict key/values 형태로 구현

- 카드 한도 구간 설정(Limit_Range) 설정

```
for i in np.arange(len(df)):
    if inform1_df['Credit_Limit'].values[i] < 2500000:
        inform1_df['Limit_Range'].values[i] = '250만원미만'
    elif inform1_df['Credit_Limit'].values[i] < 5000000:
        inform1_df['Limit_Range'].values[i] = '500만원미만'
    elif inform1_df['Credit_Limit'].values[i] < 8000000:
        inform1_df['Limit_Range'].values[i] = '800만원미만'
    elif inform1_df['Credit_Limit'].values[i] < 12000000:
        inform1_df['Limit_Range'].values[i] = '1천2백만원미만'
    elif inform1_df['Credit_Limit'].values[i] < 19000000:
        inform1_df['Limit_Range'].values[i] = '1천9백만원미만'
    elif inform1_df['Credit_Limit'].values[i] < 28000000:
        inform1_df['Limit_Range'].values[i] = '2천8백만원미만'
    else: inform1_df['Limit_Range'].values[i] = '2천8백만원이상'
```

inform1_df

	Income	Credit_Limit	Limit_Range
0	4~6천만원미만	12691000	1천9백만원미만
1	2천만원미만	8256000	1천2백만원미만
2	6~8천만원미만	3418000	500만원미만
3	2천만원미만	3313000	500만원미만

- Dict 타입 key / values 형태로 구간 : 대상자 수 생성

```
result = get_group_count(inform1_df['Limit_Range'])
result
```

```
{'1천9백만원미만': 963,
'1천2백만원미만': 1087,
'500만원미만': 2939,
'2천8백만원이상': 749,
'2천8백만원미만': 609,
'800만원미만': 1361,
'250만원미만': 2419}
```

```
result = get_group_count(inform1_df['Income'])
result
```

```
{'4~6천만원미만': 1402,
'2천만원미만': 3561,
'6~8천만원미만': 1535,
'2~4천만원미만': 1790,
'8천만원이상': 727,
'Unknown': 1112}
```

```
def get_group_count(inform1_df_columns) :

    print('len(inform1_df_columns) :', len(inform1_df_columns))

    def count_columns(result_dic, element_tuple) :
        element1, element2 = element_tuple
        # print(element1, ' ', element2)

        if not (result_dic.get(element1)) :
            result_dic[element1] = {}

        if result_dic[element1].get(element2) :
            result_dic[element1][element2] += 1
        else :
            result_dic[element1][element2] = 1

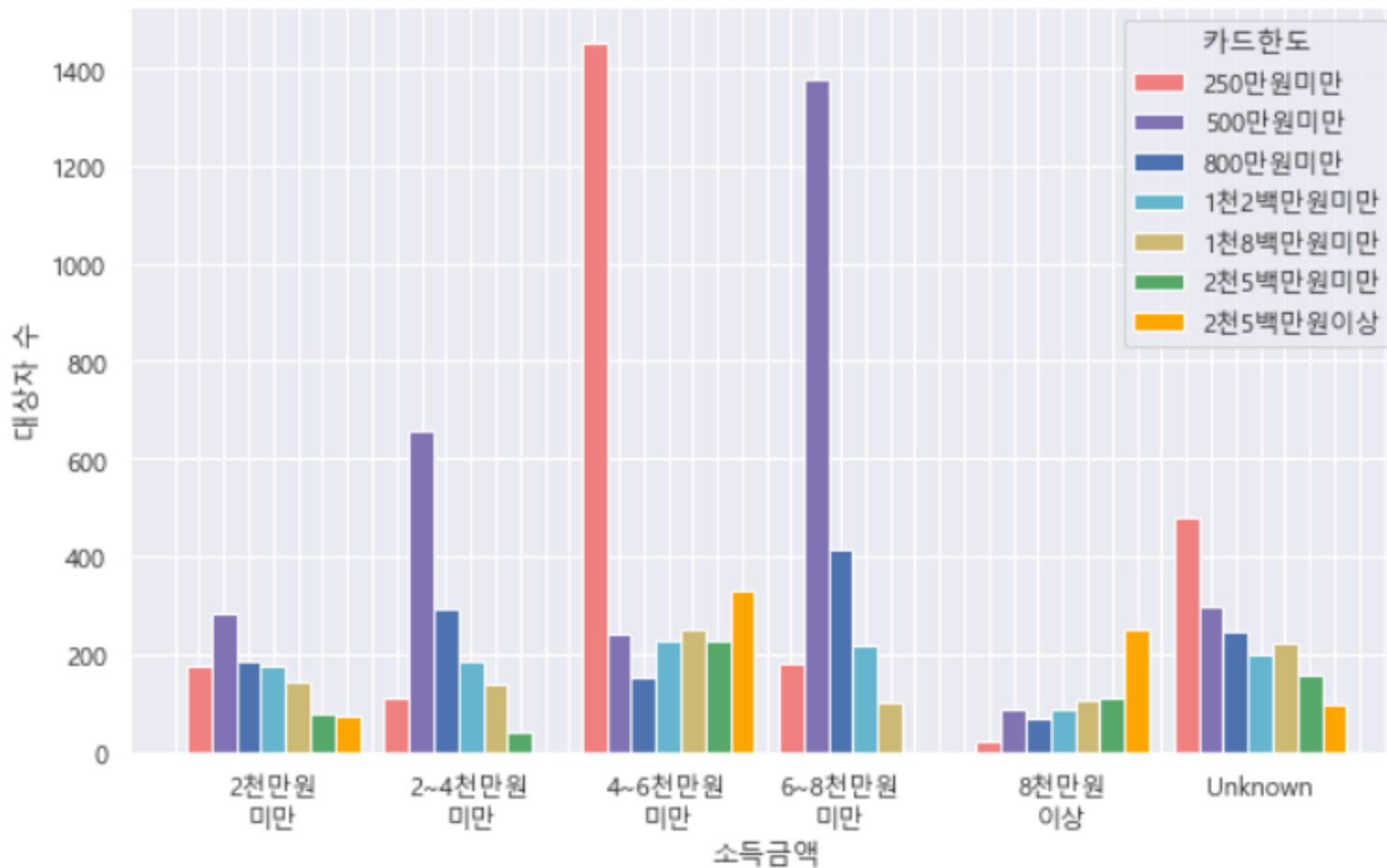
    return result_dic

result = reduce(count_columns, inform1_df_columns, dict())
return result
```


2. 프로젝트 수행 내역

* Mission 3 차트 작성 및 간략 분석

- 각 소득 구간별 카드 한도 seg를 막대 차트로 표현하여 소득에 따른 카드 한도 대상자 수를 막대차트로 표현하였음.



3. Lessons & Learned

요청자가 요구하는 사항에 따른 데이터의 선택 / 가공하는 작업이 무엇보다 필요한 작업이었음.

...

데이터 분석
요청자의
니즈 파악이
가장 중요!!
(구체화)

Lessons & Learned

박동현

- 데이터 분석 전 분석 대상 데이터에 대해 정확한 숙지가 필요.
- 분석 대상 데이터의 정합성 확인을 통해 신뢰할 수 있는 데이터 모수 구성 필요
- 다양한 케이스가 있기 때문에 데이터의 양도 중요.

오승현

- 데이터 분석에 있어 요청자의 요구사항이 무엇인지 정확히 이해하고 구체화 시키는 것이 가장 중요하다는 점을 알게 되었음.
- 미션 수행 과정에서 발생한 여러 세그먼트의 분류가 적절한지의 타당성 및 수행 방법에 대한 의구심이 분석을 진행하는데 가장 많은 시간을 소요