# OA기말프로젝트 보고서

8조

## A안

### 데이터 준비 및 전처리

주어진 데이터는 Education, Marital Status, Income, MntPurchases로 우선 Randomforest를 사용하기전에 Education과 Marital을 각 항목의 마케팅 비용으로 Mapping진행하고, Income수익의 범위내로 범주화 시킨다. 그 다음 Education_M_cost, Marital_M_cost,Income_M_cost 의 새로운 컬럼을 생성해서 3개를 groupby 해준다. 이 세가지 비용의 그룹을 이용해서 MntPurchases 를 예측해보려고 한다.

<표1>은 각 항목마다 드는 마케팅 비용으로 인코딩을 하고,
세가지 마케팅 비용이 같으면 한 그룹으로 보고,MntPurchases를 평균으로 그룹핑 처리한 데이터 모습이다.
이 그룹 데이터는 실제로는 다른 세그먼트라도 마케팅비용들과, 수익의 범주가 같은 데이터는 같은 그룹이라고 본다.

| | Education_M_cost | Marital_M_cost | Income_M_cost | MntPurchases |
|---|---|---|---|---|
| 0 | 10 | 10 | 100.0 | 76.500000 |
| 1 | 10 | 10 | 200.0 | 70.833333 |
| 2 | 10 | 10 | 300.0 | 82.923077 |
| 3 | 10 | 10 | 400.0 | 93.125000 |
| 4 | 10 | 10 | 500.0 | 505.555556 |
| ... | ... | ... | ... | ... |
| 116 | 50 | 50 | 600.0 | 534.937500 |
| 117 | 50 | 50 | 700.0 | 1128.137931 |
| 118 | 50 | 50 | 800.0 | 1368.263158 |
| 119 | 50 | 50 | 900.0 | 1492.642857 |
| 120 | 50 | 50 | 1000.0 | 1556.000000 |

**<표1>**

### 데이터 구분을 잘하고 모델을 구축했는가?

1) 그룹핑한 데이터(121개)의 features(교육수준 마케팅비용, 결혼여부마케팅비용, 수입범주 마케팅비용)을 변수로 두고, target(MntPurchases)을 구하고자하는 값으로 둔다.
2) 변수들을 train_test_split 을 통해서 나눈다. 사이즈는 0.2 or 0.3 정확한 성능차이를 알아보기위해 random_state는 31로 고정

3) X_train과 Y_train으로 학습시킨다.

4) 학습시킨 모델로 X_test의 결과를 예측한다.

### 사용한 모델에 대한 이해도는 충분한가?

랜덤포레스트는 결정트리의 단점을 보완해서 만든 머신러닝기법으로 여러 예측을 진행하고 다수결로 통해 모델을 선정해서 과적합을 줄이고 일반화를 강화시킬수있다. 또 변수의 중요도 기반으로 원하는 특성에 맞게 모델을 개선할 수 있으며, 변수 선택과 Feature Engineering에 도움을 줄 수 있다. 또한 여러가지 데이터 유형을 학습에 이용할 수 있어 유용하다.
하지만 모델의 해석에 있어서 많은 트리들이 복잡하게 얽힐수있기때문에 개별 트리수준에서 설명하기 힘들다. 앞서 말했듯 여러 트리의 결합으로 인해서 예측 시간과 계산비용이 많이 소모되고 트리수가 증가할수록 많은 비용이든다. 또한 간단한것에 대해서는 빠르고 간단하지만 많은 데이터를 사용한다면 적합한 방법은 아니라고 볼수있다 그만큼 성능이 저하될 수 있고 트리를 생성하면서 처리해야할 데이터가 늘어난다.

### **Hyper-parameter-Tuning** 진행하였는가?

테스트 세트 비율 **0.2**일 때 **)**

처음에는 지정한 모든 파라미터를 확인하며 시간이 많이드는 GridsearchCV를 진행하기보다 여러가지 항목들을 랜덤하게 빨리 확인할 수 있는 RandomsearchCV로 중점적으로 확인할 파라미터를 찾는데 집중했다.

```
Best Parameters: {'max_depth': 10, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 10, 'random_state': 35}
Best MSE score: 30646.23737203522
MSE: 4046.647150674442
```

최적화를 진행하면서 Max_depth =10, n_estimators=10인 파라미터를 이용했을때 MSE가 약 4046으로 가장 낮게 나왔다. 하지만 파라미터를 설정할때 주의할점은 각 속성마다 너무 작거나 크거나하면 과적합의 위험이 있다는것입니다. 이 파라미터의 깊이는 너무 깊고, 트리의 갯수는 너무 적어서 과적합의 위험이 있다고 생각된다.

max_depth: 트리의 최대 깊이를 지정 : 값이 클수록 과적합 가능성 증가

min_samples_leaf: 잎 노드가 되기 위해 필요한 최소 샘플 수 지정 : 작을수록 모델 복잡해지고 과적합 가능성 증가

min_samples_split: 내부 노드를 분할하기 위해 필요한 최소 샘플 수 지정 : 작을수록 모델 복잡해지고 과적합 가능성 증가

n_estimators: 앙상블에 사용할 트리의 개수 지정 : 값이 클수록 모델 성능 좋아지지만 계산 비용 증가

다시 과적합을 주의하면서 적당한 수치로 RandomizedSearchCV를 이용하여 최적 파라미터를 탐색했다.

```
Best Parameters: {'max_depth': 5, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 50, 'random_state': 35}
Best MSE score: 27969.294566568908
MSE: 4926.584114976422
```

이후 MSE는 증가하였지만 max_depth=5,n_estimators=50 으로 비교적 과적합의 위험이 적은 파라미터로 결정했습니다.

### **FeatureEngineering**을 진행하였는가**?**

FeatureEngineering에서는 Education,Marital Status등을 마케팅비용으로 매핑하여 속하는 항목에 상관없이 마케팅비용과 수익범위에 따라서 그룹화를 진행해서 세부적인 항목을 상관없이 마케팅비용으로 비교해서 매출을 최대화 할 수 있는 모델링을 하였습니다. 마케팅비용이 같더라도 세부적인 항목을 구분 가능한 모델은 다음 B안 모델링에서 확인가능합니다.

# B안

### 데이터 준비 및 전처리

주어진 데이터로 머신러닝 데이터분석을 하기위해서 Education, Marital Status 의 값을 0~n의 숫자로 인코딩을 하고 Income의 경우 0~10000 처럼 각 범위별로
분류를 진행하기위해서 그룹화를 진행하고 0~10000은 100,10001~20000은 200 으로 지정했다.
다음으로 항목별 마케팅비용을 확인해서 Income 범위별 마케팅비용과 더해서 MarketingCost라는 열을 추가했다.

이후 분석단위를 지정하기위해서 Education,Marital_Status, Income,MarketingCost를 그룹화하였다.

| | Education | Marital_Status | Income | marketingCost | MntPurchases |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 200.0 | 220.0 | 48.000000 |
| 1 | 0 | 2 | 300.0 | 320.0 | 56.000000 |
| 2 | 0 | 2 | 400.0 | 420.0 | 51.250000 |
| 3 | 0 | 2 | 500.0 | 520.0 | 641.333333 |
| 4 | 0 | 2 | 600.0 | 620.0 | 644.000000 |
| ... | ... | ... | ... | ... | ... |
| 178 | 4 | 6 | 600.0 | 660.0 | 655.285714 |
| 179 | 4 | 6 | 700.0 | 760.0 | 799.333333 |
| 180 | 4 | 6 | 800.0 | 860.0 | 1141.333333 |
| 181 | 4 | 6 | 900.0 | 960.0 | 1409.000000 |
| 182 | 4 | 7 | 500.0 | 560.0 | 424.000000 |

<표2>

이를 통해서 <표2>는 실질적으로 다른그룹인데 같은 마케팅비용이어서 한그룹으로 묶이는 상황을 없앴다.

데이터 구분을 잘하고 모델을 구축했는가?

1) 그룹핑한 데이터(183개)의 features(교육, 결혼여부, 수입, 총마케팅비용)을 변수로 두고, target(MntPurchases)을 구하고자하는 값으로 둔다.
2) 변수들을 train_test_split 을 통해서 나눈다. 사이즈는 0.2 or 0.3
3) X_train과 Y_train으로 학습시킨다.
4) 학습시킨 모델로 X_test의 결과를 예측한다.

**Hyper-parameter-Tuning** 진행하였는가?

**A**안 에는 Education,Marital Status항목을 바로 비용으로 매핑한뒤 Groupby해서 121개의 그룹으로 진행했고, 이는 마케팅비용기준으로 groupby을 해서 Marital Status에서 마케팅비용이 10달러인것들은 서로 구분하지 못하는상태다.

```
✓ 0.2s
Best Parameters: {'max_depth': 10, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 10, 'random_state': 35}
Best MSE score: 29097.41721732124
MSE: 4046.647150674442
```

평가기준이 MSE에 있기때문에 최소화를 목표로만 진행을 했는데, Hyper-parameter-Tuning에 대해 더 알아보고 어떤 기준으로 사용해야할지 알게되면서 n-estimators (트리의 갯수)가 10이고 max-depth가 10이어서 과적합의 가능성이 매우 크다는 것을 느꼈다.

183개 그룹으로 학습을 진행하는 RandomForestRegressor의 최적 파라미터를 다시 찾기 위해서 생각한것은 우선 MSE를 무작정 낮추기보다 트리수는 너무적지않게, 최대깊이가 너무 깊지않게 하는 파라미터를 찾아내는것을 목표로 하였습니다. train_test_split의 random-state는 23으로 고정한 뒤 진행하였고

max_depth: 트리의 최대 깊이를 지정 : 값이 클수록 과적합 가능성 증가

min_samples_leaf: 잎 노드가 되기 위해 필요한 최소 샘플 수 지정 : 작을수록 모델 복잡해지고 과적합 가능성 증가

min_samples_split: 내부 노드를 분할하기 위해 필요한 최소 샘플 수 지정 : 작을수록 모델 복잡해지고 과적합 가능성 증가

n_estimators: 앙상블에 사용할 트리의 개수 지정 : 값이 클수록 모델 성능 좋아지지만 계산 비용 증가

과적합을 주의하면서 적당한 수치로 RandomizedSearchCV를 이용하여 최적 파라미터를 탐색했습니다.

RandomizedSearchCV를 이용하면서 cv 항목을 추가하여 Cross-Validation을 동시에 진행함.
GridSearchCV를 이용할 수 있었지만 모든 경우의수를 확인하는것보다 랜덤하게 원하는 갯수만 확인해서 모델찾는시간을 단축하기위함

```
Best Parameters: {'random_state': 9, 'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 5, 'max_depth': 3}
Best MSE: 47439.27592373047
mse: 15587.611274129822
```

결과적으로 test size가 0.2일때의 mse는 15587.611274129822

```
Best Parameters: {'random_state': 9, 'n_estimators': 130, 'min_samples_split': 3, 'min_samples_leaf': 5, 'max_depth': 5}
Best MSE: 48621.552591553576
mse: 22348.630404549305
```

test size가 0.3일때 mse는 22348.630404549305 로 테스트 사이즈 별로 최적의 파라미터는 약간의 변동이있고 다른 파라미터보다 비교적 가장 낮은 MSE를 도출할 수 있었습니다.

**FeatureEngineering**을 진행하였는가**?**

FeatureEngineering에서는 Education,Marital Status등을 더미변수화 하지않고 인코딩을 이용하여서 0~n가지의 숫자로 설정함으로 써 모델의 복잡성을 줄이려는 방안을 이용하였습니다. 또 분류한뒤 각각의 마케팅비용을 따로 추가하여 학습하지않고 그룹별 총 마케팅비용을 계산해서 기존에 존재하는 특성을 필요한 데이터로 추가해서 이용했습니다. 모든 마케팅비용을 따로 입력해서 학습하는것보다 각 개별적인 그룹마다의 관계를 잘 확인할 수 있습니다.

## LpProblem

var_i : 고객_i 의 vip여부 , Binary(0또는 1)

feature :

$$TotalMarketing_i = EducationM_i + MaritalStatus_i + Income_i$$

target :

$$PredictedMntPurchase$$

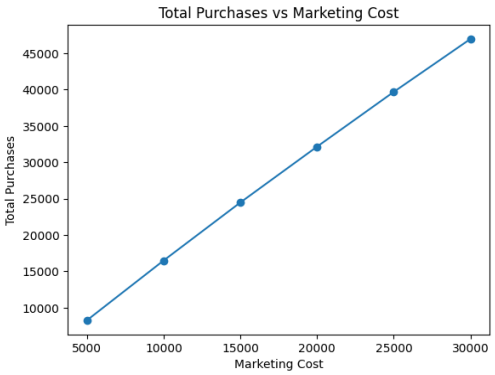
objective function :

$$maximize(\Sigma_{i=0}^{n}(var_i * target_i))$$

constraints :

$$\Sigma_{i=0}^{n}(var_i * feature_i) \leq 5000$$

(5000, 10000, 15000, 20000, 25000, 30000 순서대로 대입후 vip list와 매출액 비교)




x = [5000, 10000, 15000, 20000, 25000, 30000]
y = [8260, 16488, 24471, 32152, 39690, 46974]

마케팅비용이 5000씩 늘어나도 총 매출액이 거의 일정하게 늘어나기 때문에, 마케팅비용에 더 투자를 하는 것도 매출 최대화에 도움이 될 수 있다.

## 경영학적 **Implication**

1) 마케팅 예산 최적화: 세그먼트 우선순위를 기반으로 마케팅 예산을 최적화할 수 있다. VIP 고객을 대상으로 예상 매출액을 최대화하기 위해 더 많은 마케팅 예산을 해당 세그먼트에 투자할 수 있다. 이는 한정된 마케팅 예산을 효율적으로 분배하여 최대한의 결과를 얻을 수 있도록 도와준다.

2) 고객 세분화 및 개인화: 세그먼트 우선순위를 통해 VIP 고객을 정의하고 해당 고객을 위한 개인화된 마케팅 전략을 구축할 수 있다. 각 세그먼트에 대해 특정한 마케팅 활동과 메시지를 개발하여 VIP 고객의 관심을 높일 수 있다.

A 안에서 마케팅 비용 5000제한시 고객리스트

| | Education_M_cost | Marital_M_cost | Income_M_cost | marketingCost | MntPurchases | PredictedMnt |
|---|---|---|---|---|---|---|
| 44 | 30 | 10 | 1000.0 | 1040.0 | 2181.666667 | 1762.987588 |
| 71 | 40 | 10 | 800.0 | 850.0 | 1386.384615 | 1366.711099 |
| 72 | 40 | 10 | 900.0 | 950.0 | 1570.000000 | 1505.833544 |
| 82 | 40 | 30 | 1000.0 | 1070.0 | 1998.000000 | 1812.563485 |
| 91 | 40 | 50 | 1000.0 | 1090.0 | 2219.000000 | 1812.563485 |

A 안에서 마케팅 비용 10000제한시 고객리스트

| | Education_M_cost | Marital_M_cost | Income_M_cost | marketingCost | MntPurchases | PredictedMnt |
|---|---|---|---|---|---|---|
| 9 | 10 | 10 | 1000.0 | 1020.0 | 1263.000000 | 1654.744992 |
| 44 | 30 | 10 | 1000.0 | 1040.0 | 2181.666667 | 1762.987588 |
| 54 | 30 | 30 | 1000.0 | 1060.0 | 1789.666667 | 1786.391421 |
| 71 | 40 | 10 | 800.0 | 850.0 | 1386.384615 | 1366.711099 |
| 80 | 40 | 30 | 800.0 | 870.0 | 1624.500000 | 1366.711099 |
| 82 | 40 | 30 | 1000.0 | 1070.0 | 1998.000000 | 1812.563485 |
| 91 | 40 | 50 | 1000.0 | 1090.0 | 2219.000000 | 1812.563485 |
| 99 | 50 | 10 | 800.0 | 860.0 | 1304.454545 | 1366.711099 |
| 101 | 50 | 10 | 1000.0 | 1060.0 | 1638.000000 | 1767.773257 |
| 110 | 50 | 30 | 1000.0 | 1080.0 | 1685.000000 | 1791.177090 |

A 안에서 마케팅 비용 15000제한시 고객리스트

| | Education_M_cost | Marital_M_cost | Income_M_cost | marketingCost | MntPurchases | PredictedMnt |
|---|---|---|---|---|---|---|
| 9 | 10 | 10 | 1000.0 | 1020.0 | 1263.000000 | 1654.744992 |
| 44 | 30 | 10 | 1000.0 | 1040.0 | 2181.666667 | 1762.987588 |
| 54 | 30 | 30 | 1000.0 | 1060.0 | 1789.666667 | 1786.391421 |
| 64 | 30 | 50 | 1000.0 | 1080.0 | 1801.571429 | 1786.391421 |
| 71 | 40 | 10 | 800.0 | 850.0 | 1386.384615 | 1366.711099 |
| 80 | 40 | 30 | 800.0 | 870.0 | 1624.500000 | 1366.711099 |
| 82 | 40 | 30 | 1000.0 | 1070.0 | 1998.000000 | 1812.563485 |
| 91 | 40 | 50 | 1000.0 | 1090.0 | 2219.000000 | 1812.563485 |
| 99 | 50 | 10 | 800.0 | 860.0 | 1304.454545 | 1366.711099 |
| 100 | 50 | 10 | 900.0 | 960.0 | 1395.230769 | 1507.867122 |
| 101 | 50 | 10 | 1000.0 | 1060.0 | 1638.000000 | 1767.773257 |
| 108 | 50 | 30 | 800.0 | 880.0 | 1323.000000 | 1366.711099 |
| 109 | 50 | 30 | 900.0 | 980.0 | 1549.285714 | 1531.270955 |
| 110 | 50 | 30 | 1000.0 | 1080.0 | 1685.000000 | 1791.177090 |
| 120 | 50 | 50 | 1000.0 | 1100.0 | 1556.000000 | 1791.177090 |

A 안에서 마케팅 비용 20000제한시 고객리스트

| | Education_M_cost | Marital_M_cost | Income_M_cost | marketingCost | MntPurchases | PredictedMnt |
|---|---|---|---|---|---|---|
| 8 | 10 | 10 | 900.0 | 920.0 | 1267.000000 | 1403.150697 |
| 9 | 10 | 10 | 1000.0 | 1020.0 | 1263.000000 | 1654.744992 |
| 29 | 20 | 30 | 100.0 | 150.0 | 110.000000 | 171.488302 |
| 43 | 30 | 10 | 900.0 | 940.0 | 1564.387097 | 1463.147829 |
| 44 | 30 | 10 | 1000.0 | 1040.0 | 2181.666667 | 1762.987588 |
| 53 | 30 | 30 | 900.0 | 960.0 | 1419.571429 | 1486.551662 |
| 54 | 30 | 30 | 1000.0 | 1060.0 | 1789.666667 | 1786.391421 |
| 64 | 30 | 50 | 1000.0 | 1080.0 | 1801.571429 | 1786.391421 |
| 71 | 40 | 10 | 800.0 | 850.0 | 1386.384615 | 1366.711099 |
| 72 | 40 | 10 | 900.0 | 950.0 | 1570.000000 | 1505.833544 |
| 80 | 40 | 30 | 800.0 | 870.0 | 1624.500000 | 1366.711099 |
| 81 | 40 | 30 | 900.0 | 970.0 | 1653.000000 | 1529.237378 |
| 82 | 40 | 30 | 1000.0 | 1070.0 | 1998.000000 | 1812.563485 |
| 90 | 40 | 50 | 900.0 | 990.0 | 1424.700000 | 1487.773984 |
| 91 | 40 | 50 | 1000.0 | 1090.0 | 2219.000000 | 1812.563485 |
| 99 | 50 | 10 | 800.0 | 860.0 | 1304.454545 | 1366.711099 |
| 100 | 50 | 10 | 900.0 | 960.0 | 1395.230769 | 1507.867122 |
| 101 | 50 | 10 | 1000.0 | 1060.0 | 1638.000000 | 1767.773257 |
| 109 | 50 | 30 | 900.0 | 980.0 | 1549.285714 | 1531.270955 |
| 110 | 50 | 30 | 1000.0 | 1080.0 | 1685.000000 | 1791.177090 |
| 120 | 50 | 50 | 1000.0 | 1100.0 | 1556.000000 | 1791.177090 |

이런식으로 A안에서 30000까지 Vip list를 마케팅비용별로 추출할 수 있다. 마케팅금액의 총합이 작은부분에 있던 고객의 리스트들이 마케팅 금액의 총합이 점점 커질수록 조금씩 바뀌기도 한다. 마케팅 비용이 작았을때는 교육수준이 grad, master인 고객들의 리스트가 대부분이고 커져가면서 phd의 수준을 갖는 고객들이 증가하는 추세이다.

결혼상태에 있어서는 others, single의 범주가 더 차지하는비율이 크다. 결혼한 사람들보다 더 가중치를 두고 마케팅을 진행해야 함을 알 수 있다.

소득수준에 있어서는 예상한 바와 같이, 소득수준이 큰 고객들에게 더 가중치를 두고 마케팅을 진행해야함을 알 수 있다. vip list중 대부분이 70000-100000사이의 소득 수준을 갖고있다.

종합하자면 마케팅을 진행할때, 교육수준이 grad이상이면서, married상태가 아니면서, 소득수준이 70000-100000사이에 있는 고객을 위주로 마케팅을 진행해야한다.

B안에서 마케팅 비용 5000 제한시 고객리스트

| | Education | Marital_Status | Income | marketingCost | MntPurchases | PredictedPurchases |
|---|---|---|---|---|---|---|
| 67 | 2 | 3 | 900.0 | 980.0 | 1403.285714 | 1544.605736 |
| 78 | 2 | 4 | 1000.0 | 1060.0 | 1789.666667 | 1790.466823 |
| 85 | 2 | 5 | 800.0 | 840.0 | 1279.650000 | 1323.409647 |
| 146 | 4 | 2 | 1000.0 | 1060.0 | 1779.000000 | 1776.149036 |
| 174 | 4 | 5 | 1000.0 | 1060.0 | 1497.000000 | 1776.439870 |

B안에서 마케팅 비용 10000 제한시 고객리스트

| | Education | Marital_Status | Income | marketingCost | MntPurchases | PredictedPurchases |
|---|---|---|---|---|---|---|
| 68 | 2 | 3 | 1000.0 | 1080.0 | 1801.571429 | 1790.175990 |
| 78 | 2 | 4 | 1000.0 | 1060.0 | 1789.666667 | 1790.466823 |
| 85 | 2 | 5 | 800.0 | 840.0 | 1279.650000 | 1323.409647 |
| 87 | 2 | 5 | 1000.0 | 1040.0 | 2181.666667 | 1737.905733 |
| 112 | 3 | 3 | 1000.0 | 1090.0 | 2219.000000 | 1790.175990 |
| 122 | 3 | 4 | 1000.0 | 1070.0 | 1998.000000 | 1790.466823 |
| 129 | 3 | 5 | 800.0 | 850.0 | 1209.000000 | 1334.263115 |
| 135 | 3 | 6 | 800.0 | 850.0 | 1812.500000 | 1334.263115 |
| 146 | 4 | 2 | 1000.0 | 1060.0 | 1779.000000 | 1776.149036 |
| 174 | 4 | 5 | 1000.0 | 1060.0 | 1497.000000 | 1776.439870 |

B안에서 마케팅 비용 15000 제한시 고객리스트

| | Education | Marital_Status | Income | marketingCost | MntPurchases | PredictedPurchases |
|---|---|---|---|---|---|---|
| 67 | 2 | 3 | 900.0 | 980.0 | 1403.285714 | 1544.605736 |
| 68 | 2 | 3 | 1000.0 | 1080.0 | 1801.571429 | 1790.175990 |
| 78 | 2 | 4 | 1000.0 | 1060.0 | 1789.666667 | 1790.466823 |
| 85 | 2 | 5 | 800.0 | 840.0 | 1279.650000 | 1323.409647 |
| 87 | 2 | 5 | 1000.0 | 1040.0 | 2181.666667 | 1737.905733 |
| 103 | 3 | 2 | 800.0 | 850.0 | 1428.800000 | 1334.263115 |
| 111 | 3 | 3 | 900.0 | 990.0 | 1424.700000 | 1544.605736 |
| 112 | 3 | 3 | 1000.0 | 1090.0 | 2219.000000 | 1790.175990 |
| 122 | 3 | 4 | 1000.0 | 1070.0 | 1998.000000 | 1790.466823 |
| 129 | 3 | 5 | 800.0 | 850.0 | 1209.000000 | 1334.263115 |
| 135 | 3 | 6 | 800.0 | 850.0 | 1812.500000 | 1334.263115 |
| 146 | 4 | 2 | 1000.0 | 1060.0 | 1779.000000 | 1776.149036 |
| 156 | 4 | 3 | 1000.0 | 1100.0 | 1556.000000 | 1776.149036 |
| 165 | 4 | 4 | 1000.0 | 1080.0 | 1685.000000 | 1776.439870 |
| 174 | 4 | 5 | 1000.0 | 1060.0 | 1497.000000 | 1776.439870 |

B안에서 마케팅 비용 20000 제한시 고객리스트

| | Education | Marital_Status | Income | marketingCost | MntPurchases | PredictedPurchases |
|---|---|---|---|---|---|---|
| 15 | 0 | 3 | 800.0 | 860.0 | 1003.888889 | 1334.263115 |
| 22 | 0 | 4 | 800.0 | 840.0 | 1307.500000 | 1323.409647 |
| 33 | 0 | 5 | 1000.0 | 1020.0 | 1263.000000 | 1575.970141 |
| 48 | 2 | 0 | 800.0 | 840.0 | 1216.000000 | 1323.409647 |
| 57 | 2 | 2 | 800.0 | 840.0 | 1229.470588 | 1323.409647 |
| 68 | 2 | 3 | 1000.0 | 1080.0 | 1801.571429 | 1790.175990 |
| 76 | 2 | 4 | 800.0 | 860.0 | 1230.448276 | 1334.263115 |
| 78 | 2 | 4 | 1000.0 | 1060.0 | 1789.666667 | 1790.466823 |
| 85 | 2 | 5 | 800.0 | 840.0 | 1279.650000 | 1323.409647 |
| 87 | 2 | 5 | 1000.0 | 1040.0 | 2181.666667 | 1737.905733 |
| 93 | 2 | 6 | 800.0 | 840.0 | 1432.000000 | 1323.409647 |
| 103 | 3 | 2 | 800.0 | 850.0 | 1428.800000 | 1334.263115 |
| 112 | 3 | 3 | 1000.0 | 1090.0 | 2219.000000 | 1790.175990 |
| 122 | 3 | 4 | 1000.0 | 1070.0 | 1998.000000 | 1790.466823 |
| 129 | 3 | 5 | 800.0 | 850.0 | 1209.000000 | 1334.263115 |
| 135 | 3 | 6 | 800.0 | 850.0 | 1812.500000 | 1334.263115 |
| 146 | 4 | 2 | 1000.0 | 1060.0 | 1779.000000 | 1776.149036 |
| 156 | 4 | 3 | 1000.0 | 1100.0 | 1556.000000 | 1776.149036 |
| 165 | 4 | 4 | 1000.0 | 1080.0 | 1685.000000 | 1776.439870 |
| 174 | 4 | 5 | 1000.0 | 1060.0 | 1497.000000 | 1776.439870 |
| 180 | 4 | 6 | 800.0 | 860.0 | 1141.333333 | 1334.263115 |

B안도 이런식으로 마케팅비용별로 Vip list를 추출할 수 있다. A안과 마찬가지로 마케팅비용이 바뀌면서 고객 리스트들이 변하기도 한다. 마케팅 비용이 달라져도 교육수준은 grad, master와 phD가 골고루 분포하는 것을 볼 수 있다.

결혼상태에서는 A안과 달리 married 비율이 많아진 것을 볼 수 있고, married, single, together가 대부분을 차지하고 있다.

소득수준에서는 대부분 80000-100000 사이의 소득을 보여주고 있어 소득이 큰 고객들을 중심으로 마케팅을 진행해야한다는 것을 알 수 있다.

종합하면 마케팅을 진행할 때, 교육수준이 grad 이상이고, 결혼상태는 married, single, together이며 소득수준은 80000-100000인 고객을 위주로 마케팅을 해야한다.

3) 수익 증대: VIP 고객을 대상으로 예상 매출액을 최대화하는 것은 전체적인 수익을 증대시킬 수 있다. VIP 고객은 일반 고객보다 높은 가치를 가지며, 이들에게 집중적인 마케팅 활동을 통해 매출을 증가시킬 수 있다. 이는 수익성을 향상시키고 기업 가치를 높이는 데 도움이 된다.

4) 경쟁 우위 확보: 세그먼트 우선순위를 통해 경쟁사와의 경쟁에서 우위를 확보할 수 있다. VIP 고객을 대상으로 한 개인화된 마케팅 전략을 통해 경쟁사보다 더 많은 VIP 고객을 유치하고 유지할 수 있다. 이는 시장 점유율과 경쟁력을 향상시키는 데 도움이 된다.