
서울 병원
신환 및 재환 고객
FACT 및 분석 보고서

PART I

집단 별 이탈 비율 FACT

신환 이탈 고객

VS

재환 고객

- 19년 이전에 지방흡입수술을 받고 그 이후에 병원에 한 번도 방문하지 않은 고객
- 신환 이탈 고객의 특성
- 신환 이탈 고객이 보내는 신호를 잡아서 재환 고객으로 전환하면 병원 수익에 긍정적인 영향

- 재환 고객: 1기에 지방흡입수술을 받고 병원에 다시 방문한 고객.
- 재환 이탈 고객의 특성

고객들의 1기 행동 데이터 수집

- 지방 흡입 수술 비용, 내원 시간대 등의 데이터 수집

Clustering 알고리즘을 통해서 유사 집단 발굴

- 집단 예시
 - 집단1: 직장인으로 추정되는 30대 초반 약통통 고객
 - 집단2: 주로 오후에 방문하는 40대 중반 통통 고객

집단별 이탈 비율

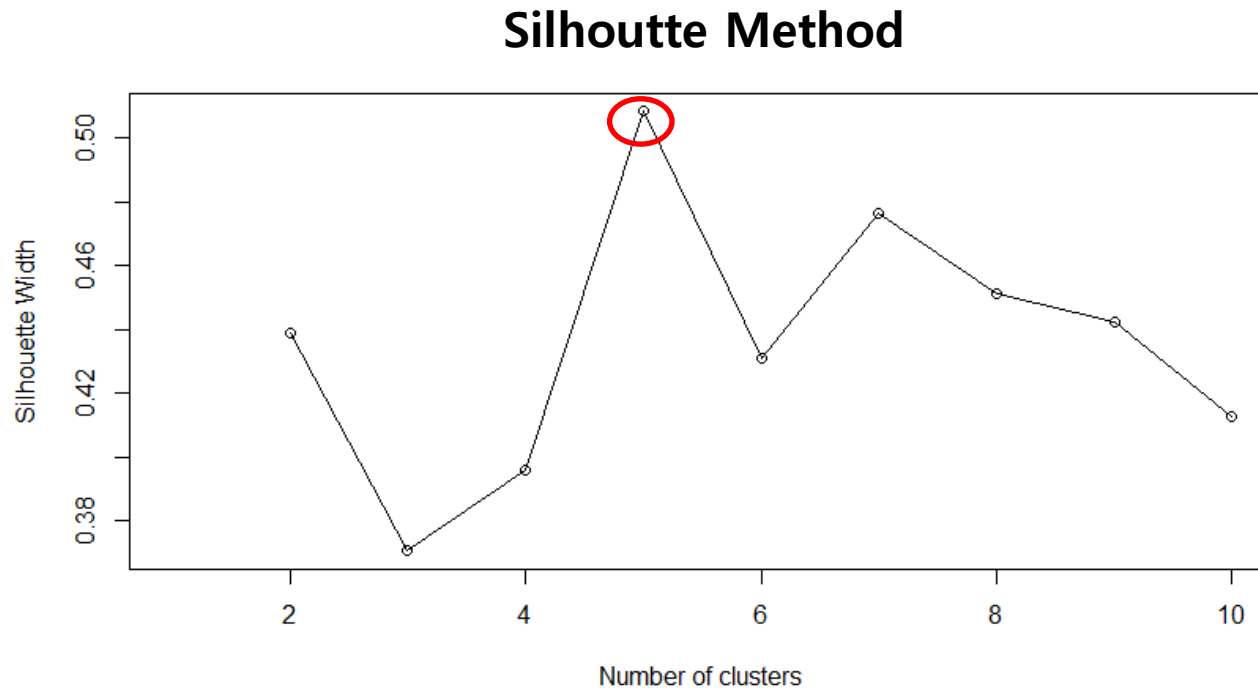
- 일종의 Fact 보고서, 집단1이 집단2보다 높은 이탈 비율을 보임을 확인

사용 변수 (모두 1기까지의 행동 데이터)

- 지방흡입수술 결제값
- 평균 병원 방문 간격
- 1기일 때, 총 병원 방문 횟수
- 가장 많이 방문한 시간대(오전 / 오후 / 직장인 시간대)
- 체성분 변수(체지방 / 체중 / 신장 / bmi / 기초대사량 / 근골격근량)
- 수술 전, 후 체지방 / 체중 / bmi 변화량(수술전 - 수술후)
- 신상정보(나이 / 결혼 여부)
- 감량 지방량

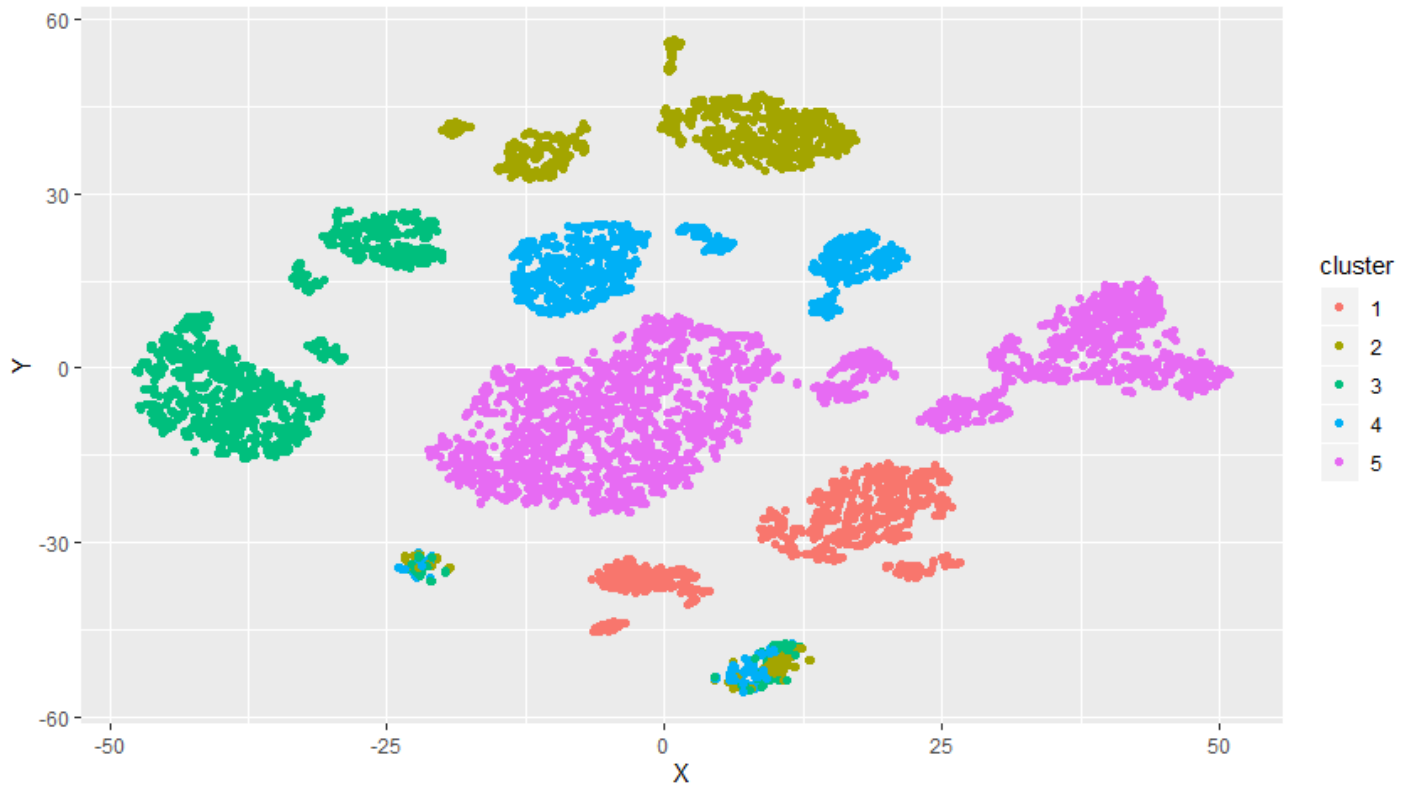
Clustering Algorithm: Partitioning Around Medoids

- 연속형, 범주형 변수가 혼재된 데이터에 대한 클러스터링
- 거리 행렬: Gower Matrix



- Silhouette width가 가장 큰 5를 집단의 개수로 설정

Clustering 결과 확인: TSNE를 통한 2차원 축소

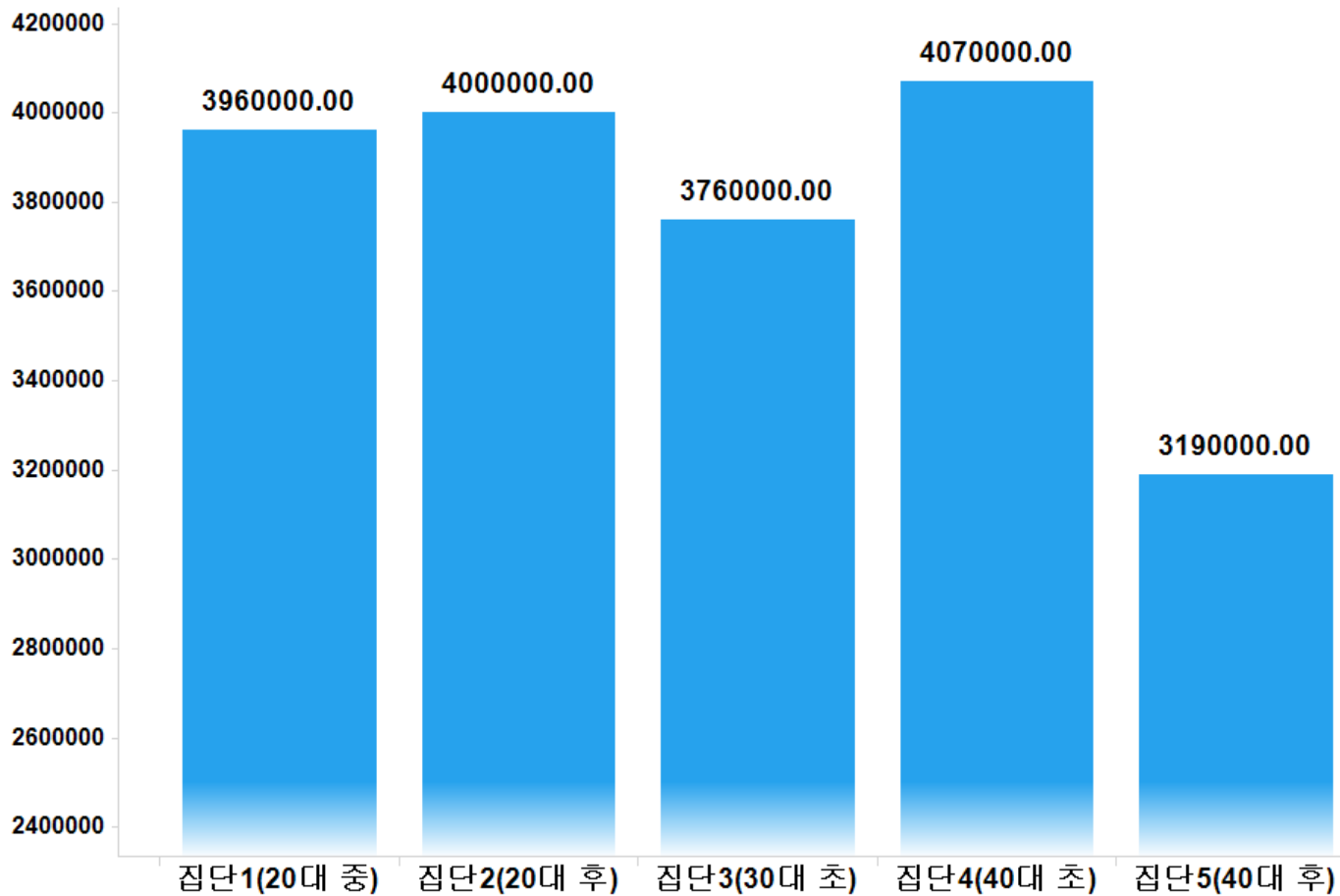


- 집단별로 분류가 잘 되었음을 확인할 수 있다.

Clustering Medoids을 통한 집단별 기본 신상 정보

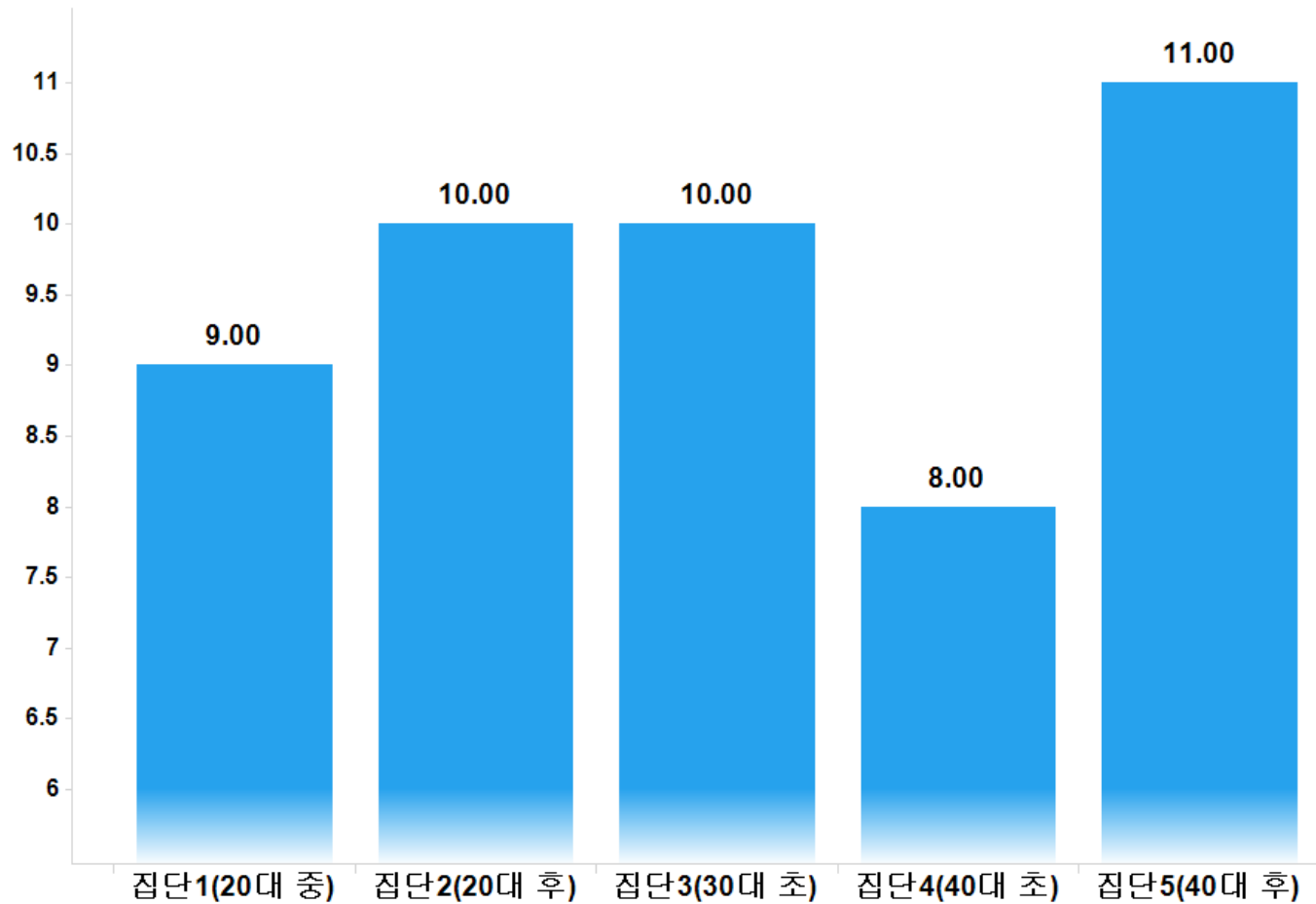
- 집단1: 오전에 방문하는 미혼인 20대 중반 고객
- 집단2: 오후에 방문하는 미혼인 20대 후반 고객
- 집단3: 직장인으로 추정(주로 12시~1시, 6시 이후 방문)되는 미혼인 30대 초반 고객
- 집단4: 오후에 방문하는 기혼인 40대 초반 고객
- 집단5: 오전에 방문하는 기혼인 40대 후반 고객

Clustering Medoids를 통한 집단별 지흠 비용 비교



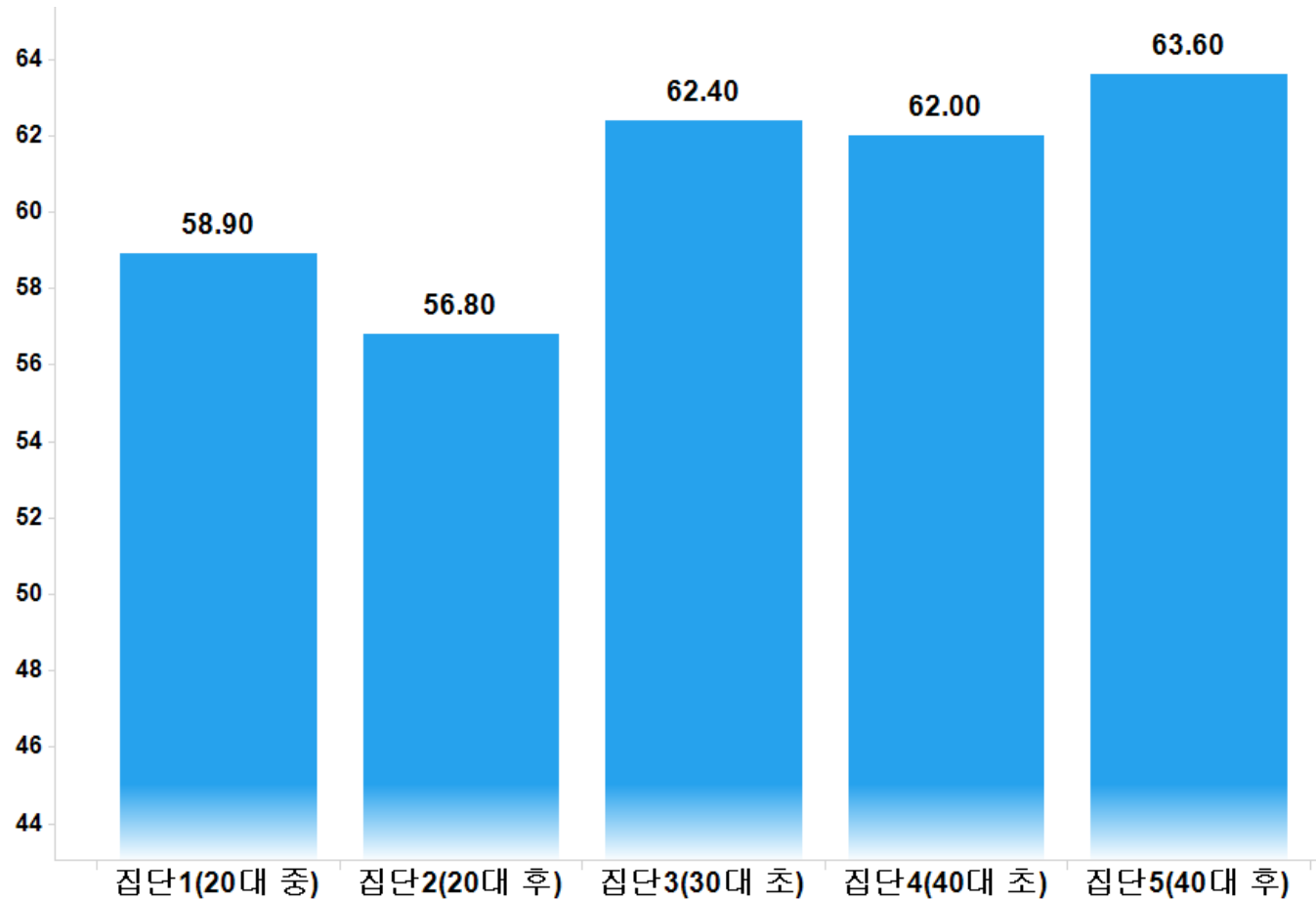
- 집단별로 상이한 지흠 비용을 보이며 집단5의 지흠 비용이 가장 낮다.

Clustering Medoids를 통한 집단별 내원 횟수 비교



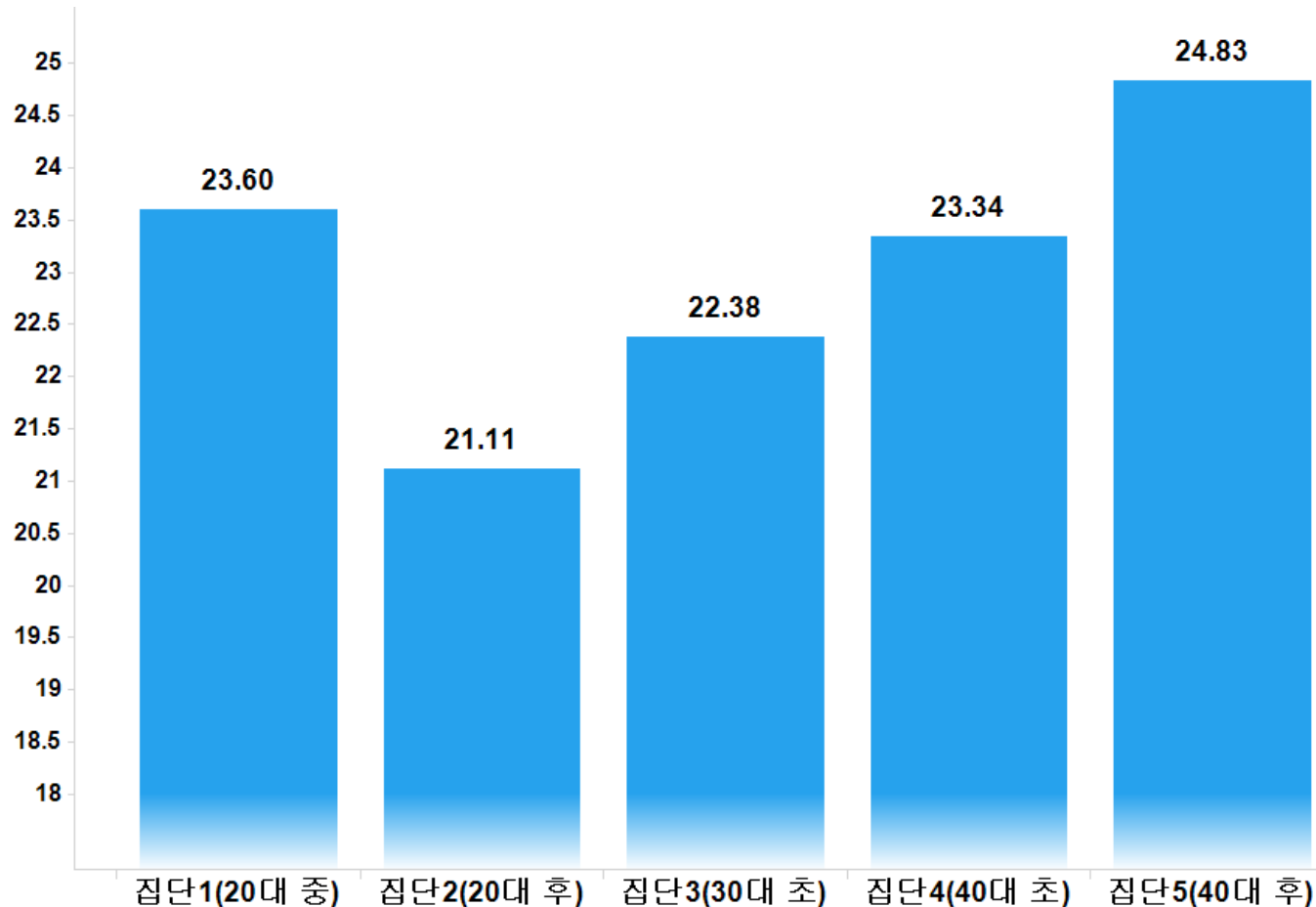
- 집단4, 집단5가 각각 최소, 최대값을 가진다.

Clustering Medoids를 통한 집단별 체중 비교



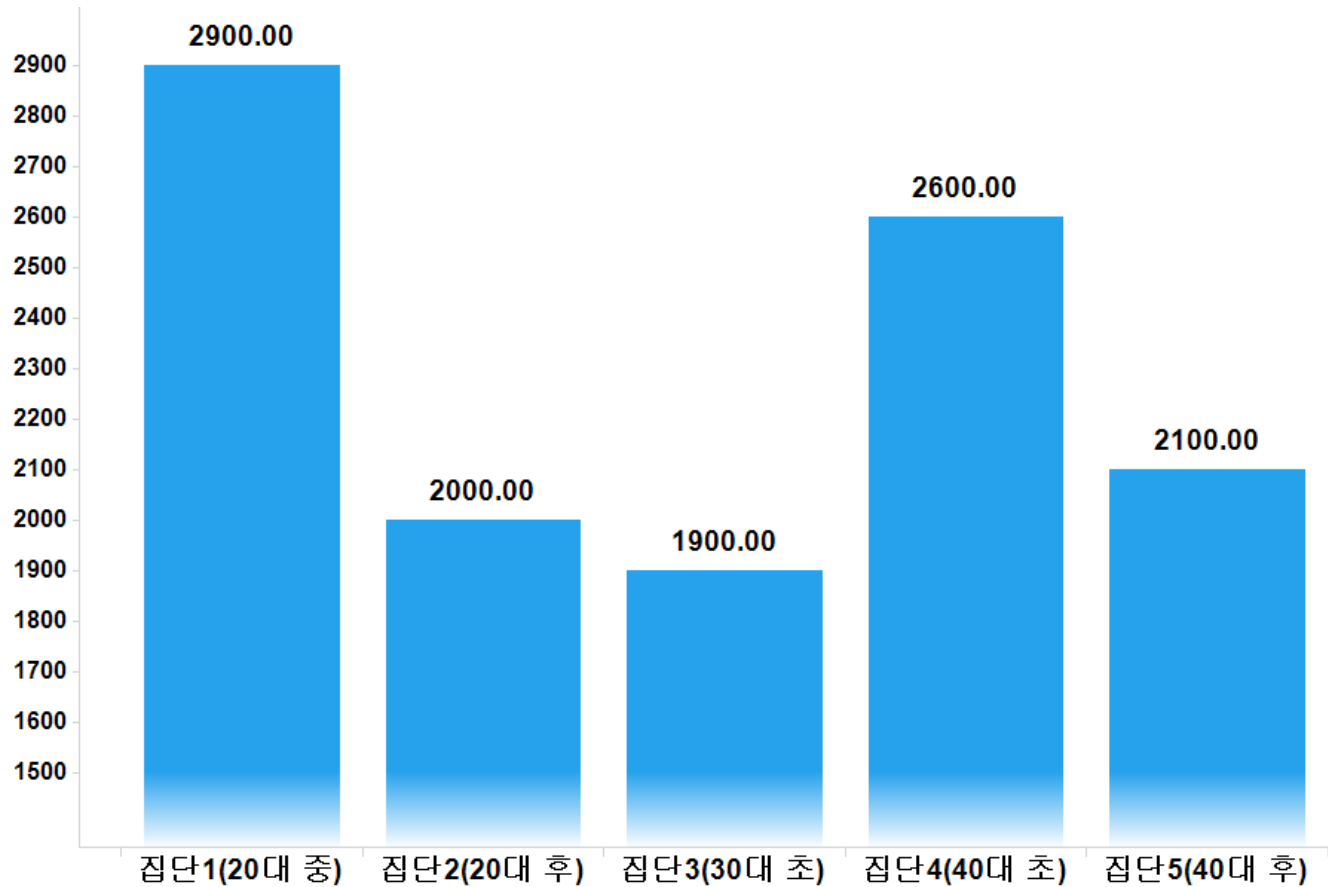
- 집단5의 체중이 가장 높고 20대인 집단1, 집단2는 체중이 작다.

Clustering Medoids를 통한 집단별 BMI 비교



- 체중이 높은 집단1, 집단2에서 BMI지수로는 집단2가 현저히 낮고 집단5의 BMI지수는 거의 과체중에 가깝다.

Clustering Medoids을 통한 집단별 감량 지방량 비교



- 감량 지방량에서는 집단1이 가장 높고 집단4가 그 뒤를 잇는다.

- 집단1: 약통통(BMI 지수가 23.6) 20대 중반 고객. 지흡 비용과 감량 지방량 모두 높다.
- 집단2: 가장 마른(BMI 지수가 21.11) 20대 후반 고객. 지흡 비용이 높지만 감량 지방량은 가장 적다.
- 집단3: 직장인 추정 보통 (BMI 지수가 22.38) 30대 초반 고객. 지흡 비용과 감량 지방량이 적은 편이다.
- 집단4: 약통통(BMI 지수가 23.34) 40대 초반 고객. 병원에 가장 적게 오지만 지흡 비용과 감량 지방량이 큰 편이다.
- 집단5: 가장 통통(BMI 지수가 24.83)한 40대 후반 고객. 병원에 가장 많이 오지만 지흡 비용은 가장 적게 사용한다.

집단 이탈 비율

| Cluster | 이탈 비율 |
|-------------|-------|
| 집단1 (20대 중) | 84.6% |
| 집단2 (20대 후) | 83.2% |
| 집단3 (30대 초) | 82.6 |
| 집단4 (40대 초) | 81% |
| 집단5 (40대 후) | 84.5% |

- 전체적으로 이탈 비율이 높다.
- 병원에 많이 오며 간을 보는 40대 후반의 집단 5와 20대 중반의 집단 1의 이탈 비율이 가장 높다.
- 반면에, 지흡 비용과 감량 지방량이 높은 40대초 집단의 이탈 비율이 가장 낮다.

PART II

고객 별로 이탈, 재환에 영향을 미치는 요인 분석

분석 로드맵

고객들의 1기 행동 데이터 수집

- 지방 흡입 수술 비용, 내원 시간대 등의 데이터 수집

모델링 결과를 통해 신환 이탈 및 재환 확률 계산 및 요인 분석

- 계산된 확률에 따라 class를 배정

분석 결과를 토대로 고객 맞춤 대응

- class 배정 요인 분석을 통한 고객 맞춤 대응

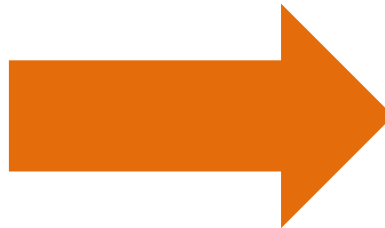
<사용 모델>

- RF Classifier
- XG Boost
- Cat Boost
- Gradient Boost

**Hyper Parameter
Tuning with
GridSearch and
CV**

※ Score Function: roc_auc

Hyper Parameters Selected



Under – Sampling during CV

**Train Model with
Selected Hyper
Parameters and
Predict Label with
Test data**

모델링 결과

| | Specificity | Precision | Accuracy |
|----------------|-------------|-----------|----------|
| RF | 0.64 | 0.73 | 0.68 |
| XG Boost | 0.68 | 0.58 | 0.61 |
| Cat Boost | 0.6 | 0.55 | 0.57 |
| Gradient Boost | 0.59 | 0.55 | 0.57 |

- **Specificity:** 모델이 신환 이탈 고객이라고 예측한 것들 중 실제로 신환 이탈인 비율
 - **Precision:** 모델이 재환 고객이라고 예측한 것들 중 실제로 재환인 비율
 - RandomForest 모델이 Precision, Accuracy에서 높은 지표를 보이므로 해당 모델을 선택한다.
-

RandomForest 모델의 중요 변수

| Feature | Score |
|----------|-------|
| 지흡 금액 | 0.21 |
| 감량 지방량 | 0.12 |
| 수술 후 bmi | 0.088 |
| 기초대사량 | 0.073 |
| 수술 후 체지방 | 0.072 |
| 수술 후 체중 | 0.059 |
| 나이 | 0.047 |
| Bmi 변화량 | 0.046 |

- 지방흡입수술 결제 금액과 감량 지방량이 신환 이탈 고객인지, 재환 고객인지 결정짓는 가장 중요한 두 변수이다.
- 수술 후 측정한 체성분 변수의 중요도가 그 뒤를 잇는다.
- 수술 후 자신의 변화한 몸의 상태(감량 지방량, 수술 후 체성분 데이터)가 재환 여부에 중요함을 알 수 있다.

- 고객들의 1기 행동 패턴을 통해 재환 확률을 계산할 수 있다.

고객 예시

| | 지흡 결제 비용 | 감량 지방 량 | 수술 후 bmi | 수술 후 체지방 | Bmi 변화량 | 재환 확 률 |
|------|-------------|------------|-------------|-------------|------------|-----------|
| 고객 A | 209만원 | 550 | 20.39 | 18.8 | -0.13 | 7.5% |
| 고객 B | 730만원 | 5500 | 31.76 | 41 | 0.24 | 94% |

tree interpreter를 통한 고객 분석 예시1

- 예시) 재환 확률이 낮게 나온 고객 A에 대한 분석.



데이터 수집

- 1기에 지방흡입수술을 한 고객 대상으로 처음 내원 부터 수술 후 체성분 검사 까지 데이터 수집

확률 계산

- RF 모델을 통해 고객 A의 재환 확률이 7.5%로 계산 됨.

요인 분석

- 고객 A의 재환 확률에 가장 큰 영향을 끼친 변수
 - 내원 횟수
 - 지흡 비용
 - 수술 후 BMI
 - 수술 후 체지방

tree interpreter를 통한 고객 분석 예시2

- 예시) 재환 확률이 높게 나온 고객 B에 대한 분석.

