텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

제목

20180860 최보영, 20180835 정지윤

1. 서론

충성도 높은 고객 식별, 고객 세분화, 타깃 마케팅 및 기타 마케팅 산업의 사용 사례에 사용할 수 있는 고객 신용카드 정보 데이터 세트이다.

고객의 신용카드 정보 데이터를 이용하여 마케팅 업계에 충성도가 높은 고객을 알아보고자 한다. (모든 결과값은 소수점 셋째자리에서 반올림 하였다.)

2. 고객의 신용카드 정보 데이터

2.1 데이터 출처 및 변수 설명

본 논문에 사용된 데이터는 Kaggle(https://www.kaggle.com/)에서 제공한 데이터로 고객의 신용카드 데이터를 나타낸 것이다. 분석에 사용될 변수는 다음의 표 2.1과 같다.

표2.1 변수명

|  |  |  |
| --- | --- | --- |
| 변수 | 변수이름 | 변수설명 |
| id | Sl\_No | 고객 일련 식별 번호 |
| X1 | Age\_Credit\_Limit | 고객의 평균 신용카드 한도 |
| X2 | Total\_Credit\_Customer | 고객이 소유한 총 신용카드 |
| X3 | Total\_visits\_bank | 고객의 총 은행 방문 수 |
| X4 | Total\_visits\_online | 고객의 총 온라인 은행 이용 수 |
| X5 | Total\_calls\_made | 고객의 총 전화 은행 이용 수 |

2.2 기초통계량

본 분석인 다변량 분석에 들어가기 전, 변수의 기초통계량을 정리하여 변수의 특성에 대해 알아보았다. 자료의 관측개수는 각 변수별로 결측값이 없이 모두 660개로 동일하다.

변수 X1인 최솟값과 최댓값의 차이가 매우 크므로 고객의 평균 신용카드 한도의 차이가 심한 것을 알 수 있다. 변수 X3,X4,X5를 통해 고객 중 은행을 전혀 이용하지 않는 고객과 온라인 은행을 전혀 이용하지 않는 고객과 전화 은행을 전혀 이용하지 않는 고객이 있음을 확인할 수 있다. 또한 고객은 평균적으로 5개의 신용카드를 가지고 있음을 알 수 있다.

나무상자 그림에서 X1 변수의 값이 나머지 변수들의 값과 차이가 많이 나서 제대로 된 확인이 불가능했기에 단위가 동일한 변수들은 따로 묶어서 나무상자그림을 그렸다. X1(신용카드 한도)와 X4(온라인 은행 이용 수)는 이상값을 가지고 있음을 볼 수 있다. 또 3가지 은행 이용 방법 중에 X5(전화 이용)가 평균이 가장 높고 데이터의 퍼짐 정도도 가장 큰 것을 확인할 수 있다.

표 2.2 연속형 변수들의 기초통계량

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 변수 | 관측개수 | 최솟값 | 중앙값 | 평균 | 최댓값 | 표준편차 |
| X1 | 660 | 3000 | 18000 | 34574 | 200000 | 37625.49 |
| X2 | 660 | 1.00 | 5.00 | 4.71 | 10.00 | 2.17 |
| X3 | 660 | 0.00 | 2.00 | 2.40 | 5.00 | 1.63 |
| X4 | 660 | 0.00 | 2.00 | 2.61 | 15.00 | 2.94 |
| X5 | 660 | 0.00 | 3.00 | 3.58 | 10.00 | 2.87 |

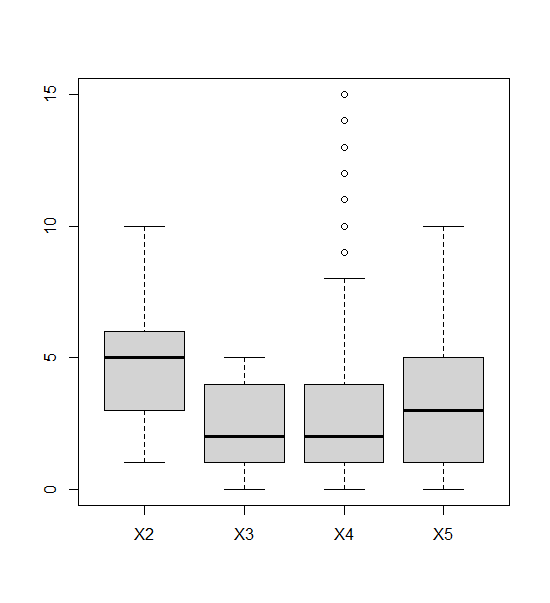
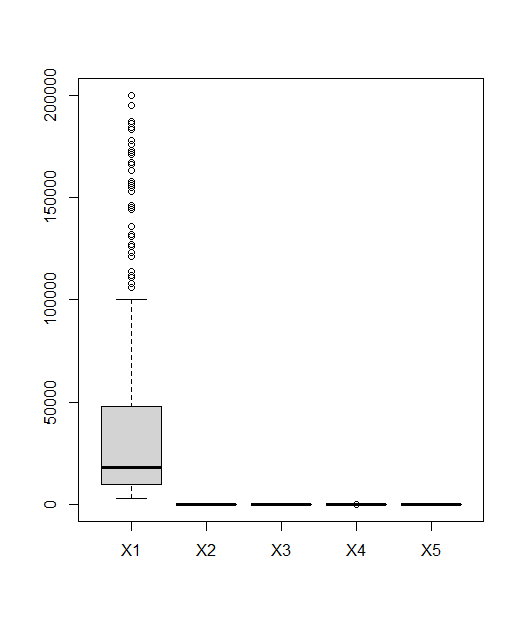


그림2.1 연속형 변수들에 대한 boxplot

3. 다변량 통계분석

3.1 상관분석

상관분석(correlation analysis)이란 두 변수 간에 선형관계를 파악하기 위한 통계적 기법으로 상관계수는 변수간 관계의 정도 혹은 방향에 대한 척도이다. 이때 두 변수간의 관계의 강도를 상관관계(Correlation, Correlation coefficient)라한다. 상관계수 는 값이 –1에 가까울수록 두 변수가 강한 음의 상관관계를 가지며, 1에 가까울수록 두 변수가 강한 양의 상관관계를 갖는다. 또한 상관계수가 0에 가까우면 두 변수 사이에는 선형의 상관관계가 없다는 것을 의미한다.

X1과 X2의 상관계수는 0.61로 가장 높은 양의 상관관계를 보이며 X1과 X4의 상관계수는 0.55로 두번째로 높은 양의 상관관계를 보인다. X1과 X3의 상관관계가 -0.10으로 가장 약한 것을 알 수 있다. 그림3.1은 변수들 간의 상관관계를 보다보기 편하게 하기 위하여 시각화한 것이다.

표 2.3 변수간 상관행렬

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| X1 | 1.00 | 0.61 | -0.10 | 0.55 | -0.41 |
| X2 | 0.61 | 1.00 | 0.32 | 0.17 | -0.65 |
| X3 | -0.10 | 0.32 | 1.00 | -0.55 | -0.51 |
| X4 | 0.55 | 0.17 | -0.55 | 1.00 | 0.13 |
| X5 | -0.41 | -0.65 | -0.51 | 0.13 | 1.00 |

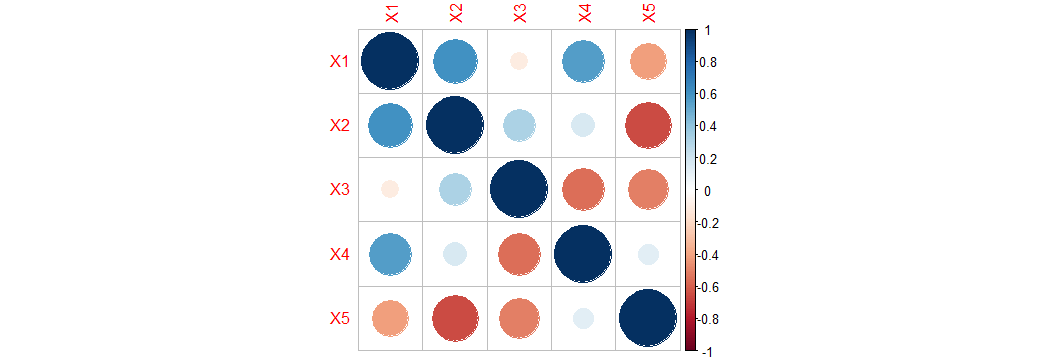
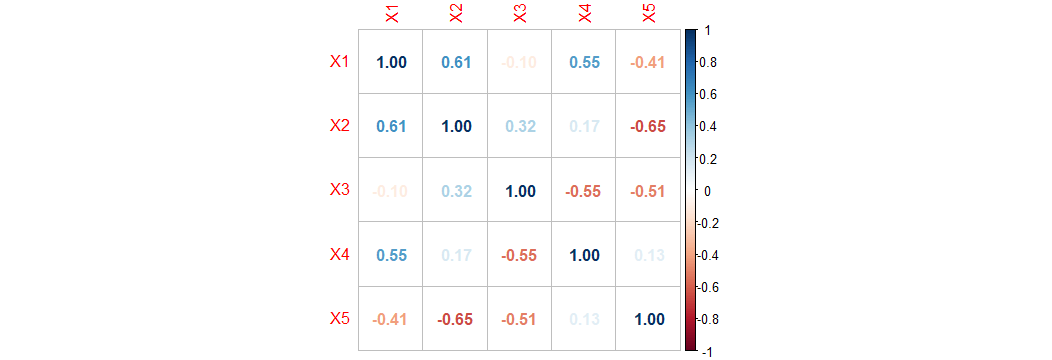


그림3.1 상관행렬의 시각화