**온라인 쇼핑몰 판매금액에 영향을 미치는 요소 분석**

3(팀)

이름 김영규, 김솔미, 이현희

**요약** : 온라인 쇼핑몰의 증가와 코로나19로 인한 비대면 상황이 증가함에 따라 온라인 쇼핑몰의 판매금액 증대에 대한 관심이 주요 관심사라고 할 수 있다. 최근 온라인 쇼핑몰 시장의 성장으로 인해 시장의 흐름이 바뀌며, 이 흐름에 맞는 전략을 세우기 위한 움직임이 시작되고 있다. 그러나, 온라인 쇼핑몰은 인터넷을 기반으로 시장의 크기가 큰 만큼, 판매금액 증대에 미치는 요소를 분석하는 것에 복잡성이 있다. 이에 따라 온라인 쇼핑몰에서는 기업에서 사용하는 RFM분석을 바탕으로 판매금액 증대 전략에 대한 고민을 이어 나가며, 이를 위해 RFM 분석이 판매금액과 연관성을 가졌는지를 분석한 후 월, 연도별 판매금액 및 순수익에 대한 분석, 결제 방법과 판매금액간의 연관성에 대해 분석하여 판매금액과의 연관성뿐 아니라 판매금액 증대 추후 전략에 대해서도 함께 고민하며 이후 판매금액 증대를 실현하는데 기여할 것으로 기대된다.

**약어 :** RFM : 최근성(Recency), 최빈성(Frequency), 총구매액(Monetary)

**Ⅰ. 서론**

최근 온라인 쇼핑몰의 시장이 급격하게 성장하면서 기업 간의 경쟁이 치열해졌다. 많은 기업은 시장 점유율을 높이기 위해서 다양한 분석 방법을 활용 중이다. 이 중 고객에 대한 분석이 중요해지고 있으며 고객 기여도에 대한 분석을 기반으로 한 판매금액 증대에 대한 관심이 높아지고 있다. 본 보고서는 다양한 분석 방법 중에서 3가지 방법을 사용할 것이다.

첫째 RFM분석을 통해 고객 기여도(본 데이터에서는 '업체명' 기준)를 알아내어 온라인 쇼핑몰 판매금액에 끼치는 영향력에 대해 알아보고자 한다.

둘째, 판매금액 시각화를 통한 판매금액 및 순수익을 분석하고자 한다. 판매금액의 기준은 기업에 따라 다르므로, 판매금액의 기준을 세워 이에 따른 월별, 연도별 판매금액을 분석하고 할부와 처리상태를 고려한 순수익도 함께 분석하여 할부가 월별 판매금액에 주는 영향력에 대해 알아보고자 한다.

셋째, 결제 방법과 판매금액이 가지는 연관성에 대해 알아보고자 한다. 다양한 결제 방법 중 어떤 결제 방법이 가장 많이 사용되었으며 판매금액과는 어떠한 관련성이 있는지 분석하려 한다.

위와 같은 3가지 분석을 활용해 온라인 쇼핑몰 판매금액에 영향을 미치는 요소들을 알아보고자 한다. 이를 위해 주제를 '온라인 쇼핑몰 판매금액에 영향을 미치는 요소들'로 잡고 다음과 같은 3가지 가설을 세웠다.

1) RFM 등급이 판매금액과 연관성이 있다.

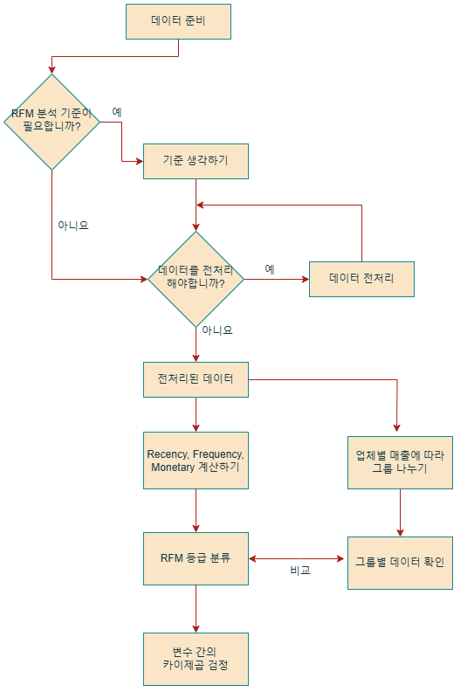
2) 월, 연도별 판매금액의 차이가 있다.

3) 결제 방법과 판매금액이 연관성이 있다.

3가지 가설의 검증을 끌어냄에 따라 판매금액 증대에 미치는 다양한 요소를 분석하고 '온라인 쇼핑몰 판매금액 극대화 방법은 무엇일까?'에 대한 결론을 내리고자 한다.

**1. 입점 기업별 RFM 분석**

**1-a. 워크 플로우**

****

**1-b. RFM 계산**

[[1]](#footnote-1)RFM은 최근성(Recency), 최빈성(Frequency), 총구매액(Monetary)의 약자로써 세 가지 요소로서 각 개체의 자료를 점수화하여 고객들을 세분화하는 기법이다. RFM에 대해서 상품 구매의 측면에서 부연하면 다음과 같다.

(1) Recency: 고객이 얼마나 최근에 구매했는가.

(2) Frequency: 고객이 얼마나 자주 구매했는가.

(3) Monetary: 고객이 얼마나 많이 지출했는가.

일반적으로 RFM의 개념은 R, F, M 변수들 각각에 대해서 점수를 부여하는 것이다.

첫째, R에 대한 관련 점수는 다음과 같이 기준일로부터 주문일이 가까울수록 높은 점수를 부여한다. 2022년 11월 8일까지의 데이터를 가지고 있기 때문에 기준 날짜를 2022년 12월 1일로 설정하였다. 각 주문에 대해 기준일로부터 시간이 얼마나 지났는지 구하고 등급을 적용하고 데이터에 등급 칼럼을 추가하였다.

(1) 3점 - 기준일로부터 1년 이내

(2) 2점 - 기준일로부터 2년에서 1년 사이

(3) 1점 - 기준일로부터 2년 초과

둘째, F에 대한 관련 점수는 주문 빈도수가 높을수록 높은 점수를 부여한다. 각 업체에 대해 주문이 들어온 빈도수를 구한 후 등급을 적용하고 데이터에 등급 컬럼을 추가하였다. 업체별 빈도수를 알아봤을 때 가장 높은 Frequency는 85553이고 가장 낮은 Frequency는 1이 나왔다.

(1) 3점 - Frequency가 10,000 이상

(2) 2점 - Frequency가 1,000 이상 10,000 미만

(3) 1점 - Frequency가 0보다 크고 1,000 미만

셋째, M에 대한 관련 점수는 업체별 총 판매 금액이 높을수록 높은 점수를 부여하였다. 각 업체에 대해 주문이 들어온 모든 판매 금액을 합한 후 등급을 적용하고 데이터에 등급 칼럼을 추가하였다.

(1) 3점 - 판매 금액 합이 100,000,000 이상

(2) 2점 - 판매 금액 합이 10,000,000 이상 100,000,000 미만

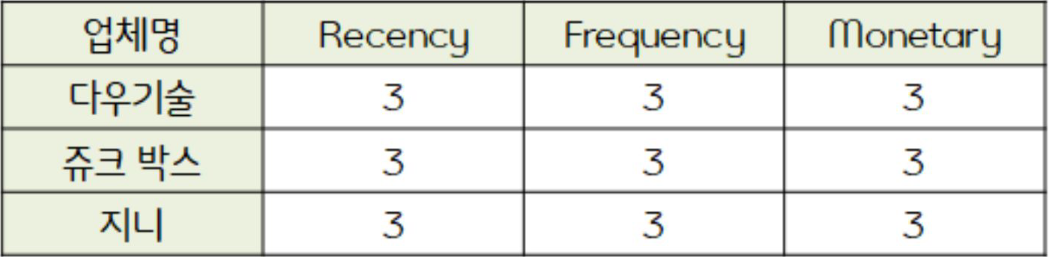
(3) 1점 - 판매 금액 합이 10,000,000 미만

Recency, Frequency, Monetary 모두 3개의 구간으로 나누었고, 해당하는 각 집단의 점수는 3, 2, 1점으로 부여하고 있다.

**1-c. 판매금액에 따른 분류**

업체별 총 판매 금액 합에 따라 100,000,000 이상을 billion\_group, 10,000,000 이상 100,000,000 미만을 ten\_million\_group, 1,000,000미만을 below\_group으로 묶었다.

billion\_group에 속한 랜덤한 업체 3개의 Recency, Frequency, Monetary의 점수를 확인하였다.



<1.1> billion\_group 그룹에 속한 업체 3개의 Recency, Frequency, Monetary의 점수

이와 같이 Recency, Frequency, Monetary 모두 3점으로 점수가 높은 것을 확인할 수 있다.

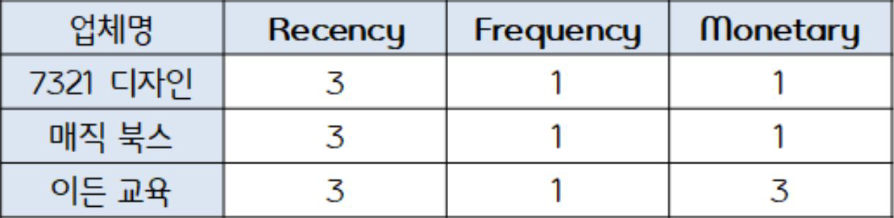
이어서 ten\_million\_group에 속한 랜덤한 업체 3개의 Recency, Frequency, Monetary의 점수를 확인하였다.



<1.2> ten\_million\_group 그룹에 속한 업체 3개의 Recency, Frequency, Monetary의 점수

Recency, Frequency, Monetary의 점수가 모두 높거나 모두 낮지 않게 다양하게 나온다.

마지막으로 below\_group을 확인하였다.



<1.3> below\_group 그룹에 속한 업체 3개의 Recency, Frequency, Monetary의 점수

모두 Recency 점수는 높았지만, Frequency, Monetary의 점수가 1점인 것을 확인할 수 있다. 이러한 사실을 통해 Recency, Frequency, Monetary의 총합이 높아야 판매금액이 잘 나온다는 것을 알 수 있다.

**1-d. Recency, Frequency, Monetary 카이제곱 검정**

Recency, Frequency, Monetary 간의 연관성이 있는지 알아보기 위해 카이제곱 검정을 사용하였다.

 두 변수 간의 연관성이 없다.

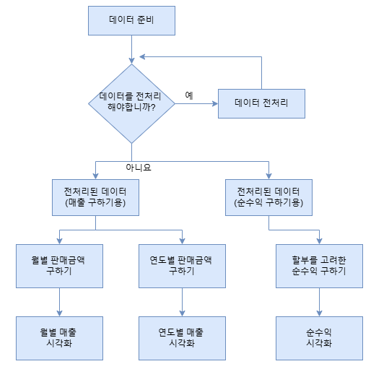
 두 변수 간의 연관성이 있다.

|  |  |
| --- | --- |
| Recency, Frequency 간의  카이제곱 검정 | 카이제곱 통계량: 6646.998535445041  p-value: 0.0 |
| Recency, Monetary 간의  카이제곱 검정 | 카이제곱 통계량: 4512.236474974696  p-value: 0.0 |
| Frequency,Monetary 간의  카이제곱 검정 | 카이제곱 통계량: 177125.0592794641  p-value: 0.0 |

3개의 카이제곱 검정 모두 p-value의 값이 0.05보다 작으므로 H0, 즉 귀무가설을 기각한다. 따라서 각 변수 간의 연관성이 있다.

**2. 판매금액 시각화를 통한 판매금액 및 순수익 분석**

**2-a. 워크 플로우**

****

**2-b. 월별 및 연도별 판매금액**

판매금액(sales)은 기업에서 생산한 제품이나 서비스를 내다 파는 일 즉, 판매하는 행위를 이야기한다. 온라인 쇼핑몰의 월별 및 연도별 판매금액을 알기 위해 데이터를 이용하여 꺾은선 및 막대그래프로 표현하였다.

이때 사용한 데이터의 특징은 다음과 같다.

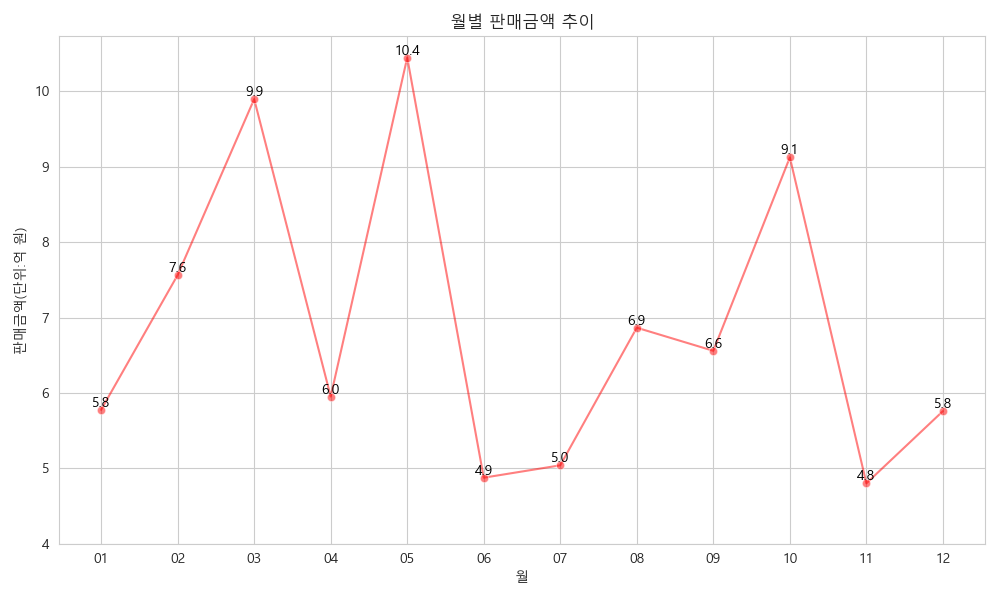
1) 2019년 12월부터 2022년 11월의 데이터로 각 월은 3개의 연도를 포함하고 있다.

2) 판매금액의 기준은 '결제 시점'으로 보고, 이에 따라 온라인 쇼핑몰 데이터의 '처리상태' 17가지 중 다음과 같은 항목 5가지 '구매확정', '배송완료', '배송중', '배송준비', '결제완료'로 설정하였다.

3) 월별 판매 금액을 모두 합친 금액은 8,266,485,118원로 약 83억이다.

4) 판매 금액의 합을 판매금액으로 보았다.

먼저, 위의 항목을 기준으로 계산한 월별 판매 금액 추이는 다음과 같다.

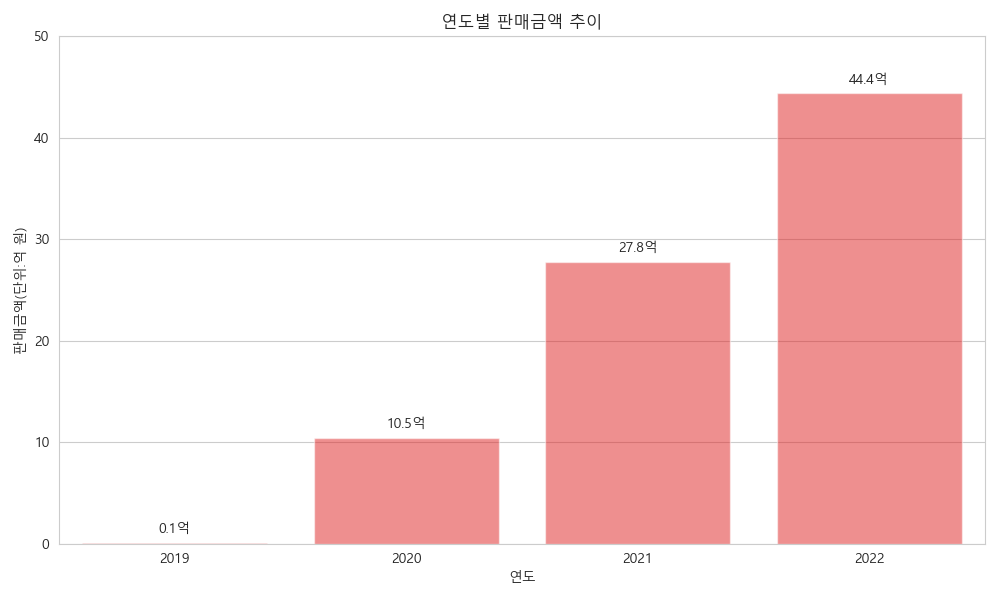


<2.1> 월별 판매 금액 추이

2019년 12월부터 2022년 11월까지의 데이터를 '월'을 기준으로 나누어 '판매금액'을 합친 결과, 3월, 5월, 10월에 판매 금액이 높게 나오는 것을 확인 할 수 있다.

판매 금액이 높은 월을 빼고 나머지 월은 비슷한 판매 금액의 추세를 보인다.

다음으로, 위와 동일한 기준을 바탕으로 연도별 판매 금액 추이는 다음과 같다.



<2.2> 연도별 판매 금액 추이

위 그래프 중 2019년(2019년 12월) 데이터는 그래프 설정상 소수점 두 번째 자리에서 반올림하였으며, 반올림한 값이 0.1억으로 2019년도의 판매 금액은 500만 원이다. 위 그래프에 따르면, 매년 판매 금액이 증가하는 추세임을 확인할 수 있다.

**2-c. 월별 순수익 (할부 및 처리상태 고려)**

월별 순수익은 위에서 제시한 판매금액액과 다르게 본 온라인 쇼핑몰에서 실제로 얻은 순수익만을 포함한 데이터를 바탕으로 분석하였다.

이 데이터의 특징은 다음과 같다.

1) 2019년 12월부터 2022년 11월의 데이터로 각 월은 3개의 연도를 포함하고 있다.

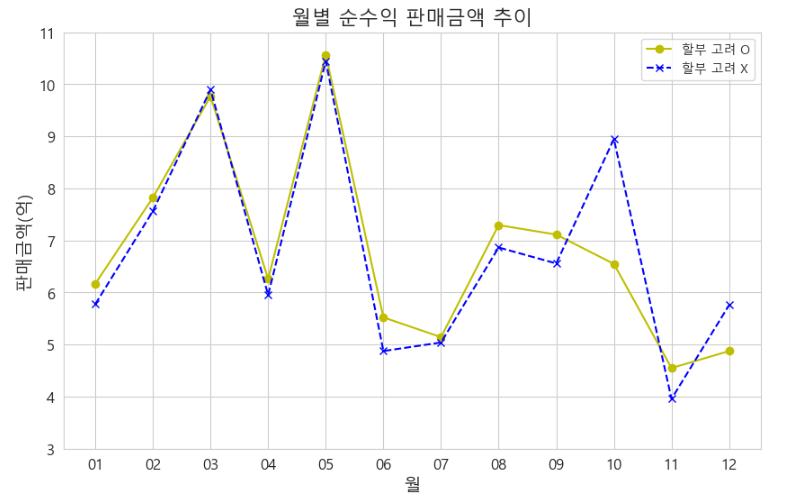
2) 판매금액의 기준은 '구매 확정'으로 보고, 이에 따라 '처리상태' 중 '구매 확정'에 해당하는 판매 금액만을 사용하였다.

3) 할부는 1개월, 6개월, 12개월, 18개월, 24개월이 존재하였으며, 1개월은 일시불로 보고, 6개월, 12개월, 18개월, 24개월만 할부로 보고 데이터를 설정하였다.

4) 할부 원리에 따라 월별 할부 금액은 '결제된 월' 기준 그 음 달로 넣어 계산하였다.

5) 월별 순수익을 모두 합친 금액은 8,164,569,838원으로 약 82억이다.

다음 그래프는 할부를 고려했을 때와 고려하지 않았을 때 판매 금액 추이를 나타낸 그래프이다.



<2.3> 월별 순수익 판매 금액 추이

다음 그래프에서 월별 판매금액과 월별 순수익의 차이는 101,915,280원으로 공통된 구매 확정을 제외한 '배송완료', '배송중', '배송준비', '결제완료' 중 101,915,280원은 취소 및 환불되어 순수익으로 계산되지 못했음을 알 수 있다. 또한, 그래프에서 할부를 고려했을 때와 안 했을 때 판매 금액의 차이가 나는 것으로 보아, 10월에 결제된 항목 중 할부로 계산된 항목이 많았음을 알 수 있다.

결과적으로 월, 연도별 판매금액의 차이가 있음을 확인 할 수 있다.

**3. 결제 방법에 따른 분석**

결제방법에 따라 판매금액간의 연관성을 알아보기 위해 '결제방법'을 기준으로 데이터를 이용하여 두 변수간의 연관성에 대해 알아보았다.

이 데이터의 특징은 다음과 같다.

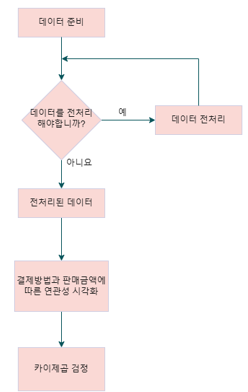
1) 결제방법은 맨 앞의 한 가지만 사용한 것으로 간주하므로, 맨 앞의 한 가지만 사용한 방법으로 결제 방법 변수를 전처리하였다.

2) 결제 방법 중 '무통장입금' , '가상계좌' , '현금간편결제'는 모두 '현금'이라는 새로운 변수에 저장하여 사용하였다.

3) 결제 방법은 '신용카드' , '현금' ,'후불, '포인트결제방법' ,'적립금' , '웰컴마일', '정기결제'이다.

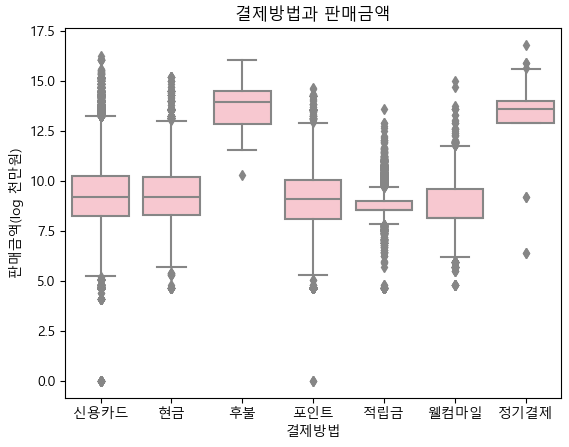
4) 판매금액이 크기 때문에 판매 금액을 로그변환 하였다.

**3-a. 워크 플로우**

****

**3-b. boxplot으로 결제 방법과 판매금액 간의 연관성**

결제 방법에 따라 판매 금액의 차이가 있는지 알아보기 위해 boxplot을 활용하였다. 결제 방법은 '신용카드' , '현금' ,'후불, '포인트결제방법' ,'적립금' , '웰컴마일', '정기결제'이고, 그에 따른 판매금액으로 그래프를 나타내었다.



<3.1> 결제 방법에 따른 판매금액 boxplot

먼저 적립금, 웰컴마일의 박스플롯을 보면, 중앙값이 없음을 확인할 수 있다. 신용카드, 현금, 포인트의 결제 방법들은 금액의 분포 차이가 나지 않는 것을 확인 할 수 있다. 후불과 정기 결제는 다른 결제 방법들 보다 높게 박스 플롯이 형성되어있음을 확인할 수 있다.

**3-c. 카이제곱 검정**

카이제곱 검정을 통해 결제 방법과 판매 금액의 연관성이 있는지 알아보았다. 먼저 다음과 같이 귀무가설(과 대립가설(을 설정하였다.

 결제방법과 판매금액은 연관성이 없다.

 결제방법과 판매금액은 연관성이 있다.

카이제곱 검정의 결과, 카이제곱 통계량은 350460.37(소수점 3번째 자리에서 반올림)로 나왔으며 P-value는 0.0이다. 이때, P-value <0.05이므로 , 즉 귀무가설을 기각하고, 대립가설을 채택하였다. 따라서 결제 방법과 판매금액은 연관성이 있다.

**결과**

본 프로젝트는 '온라인 쇼핑몰 판매금액의 영향을 미치는 요소들'을 주제로 진행되었다. 이를 위해 사전에 세운 가설은 '1) RFM등급이 판매금액과 연관성이 있다, 2)월, 연도별 판매금액의 차이가 있다, 3) 결제 방법과 판매금액이 연관성이 있다.'로 각 가설의 결과에 대해 논의하고자 한다.

1) RFM등급이 판매금액과 연관성이 있다.

위 가설을 검증하기 위해 RFM을 각각 3가지 그룹으로 나누었다.

R은 '1년, 2년, 그 외'로, F는 '10,000 이상, 10,000 미만 & 1,000 이상, 1,000 미만'을 기준으로, M은 '100,000,000 이상을 billion\_group, 10,000,000 이상 100,000,000 미만을 ten\_million\_group, 그 밑을 below\_group'으로 그룹을 나누어 진행하였다.

그림 <1.3> below\_group 그룹에 속한 업체 3개의 Recency, Frequency, Monetary의 점수를 통해 알게 된 사실은

모두 Recency 점수는 높았지만, Frequency, Monetary의 점수가 1점인 것을 확인할 수 있다. 이러한 사실을 통해 Recency, Frequency, Monetary의 총합이 높아야 판매금액이 잘 나온다는 것을 알 수 있다.

추가로, 카이제곱 검정을 이용하여 각 그룹 간의 연관성을 분석하였고, 귀무가설은 '두 변수 간의 연관성이 없다' , 대립가설은 '두 변수 간의 연관성이 있다'로 설정하였다.

3개의 카이제곱 검정 결과 모두 p-value 값이 0.0으로 모두 p-value의 값이 0.05보다 작아 귀무가설을 기각하였다. 따라서 RFM변수간의 연관성이 있다고 결론지을 수 있다.

즉, RFM변수간의 연관성이 있으므로 RFM의 총합의 크기가 판매금액과 관련이 있음을 알 수 있다.

2) 월, 연도별 판매금액의 차이가 있다.

<2.1> 월별 판매 금액 추이, <2.2> 연도별 판매 금액 추이의 그래프를 통해서 연도가 지남에 따라 매년 판매 금액이 증가하는 추세임을 확인할 수 있었고, 월별 판매 금액의 차이를 확인할 수 있었다.

3) 결제 방법과 판매금액이 연관성이 있다.

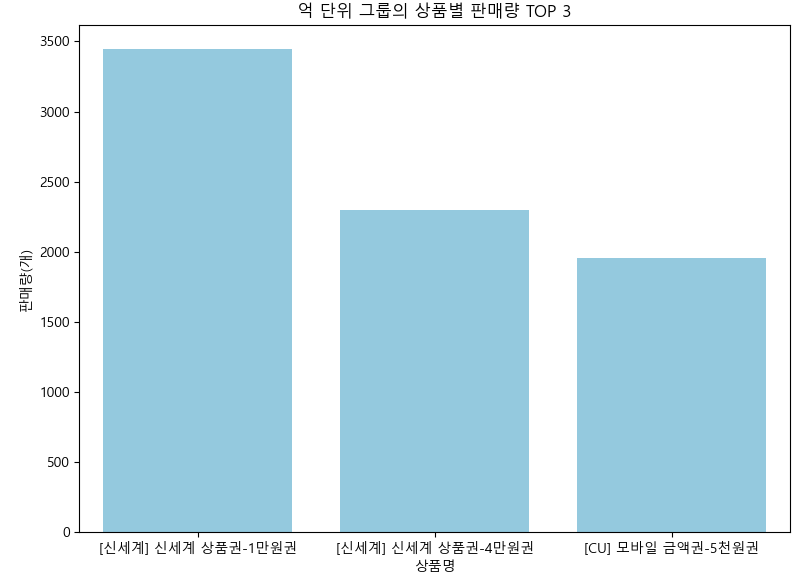
위 가설을 검증하기 위해 결제 방법과 판매금액 간의 연관성에 대해 분석하였다. 두 변수 간의 연관성 분석을 위해 카이제곱 검정을 사용하였다. 이때 귀무가설은 '결제 방법과 판매금액은 연관성이 없다'로, 대립가설은 '결제 방법과 판매금액은 연관성이 있다'로 설정하였다. 카이제곱 검정 결과 카이제곱 통계량은 '350460.37'이었으며, p-value 값은 0.0으로 p-value 값이 0.05보다 작아 귀무가설을 기각하였다.

따라서 결제 방법과 판매금액은 연관성이 있다.

**4. 토론**

RFM의 총합의 크기가 판매금액과 관련이 있다는 사실을 통해, RFM 분석을 이용하여 그룹을 나누고, 이에 따른 추가적인 분석이 유의미함을

RFM 점수가 판매금액의 영향을 준다는 것을 바탕으로 M 그룹에서 점수 3인 '억 단위' 그룹에서 가장 많이 팔린 상품별 판매량 1~3순위를 그래프로 나타내보았다.



위 그래프를 보면 1순위인 [신세계] 신세계 상품권-1만 원 권이 3,445개로 가장 많았으며. 2순위인 [신세계] 신세계 상품권 - 4만 원권이 2,298개였다. 마지막 3순위는 [CU] 모바일 금액권 - 5천원권은 1,951개였다. 1,2,3순위의 공통점을 분석해 본 결과 모두 상품권으로 드러났으며, 이에 따라 추후 상품권을 이용한 온라인 판매가 유의미한 판매금액 증대 효과를 가지고 올 것이라 기대한다.

RFM분석은 기업이 고객 기여도를 가지는 유의미함을 기존과 달라지지 않았으며, 프로젝트 결과에도 드러나듯, 추후에도 RFM 분석을 토대로 고객 기여도에 대한 기업의 구체적인 분석을 토대로 판매금액 증대의 구체적인 방안에 대한 논의가 이루어져야 한다.

▶ 사용한 라이브러리

1. pandas: The pandas development team. (2021). pandas-dev/pandas: Pandas (Version 2.1.4) [Software]. Zenodo.<https://doi.org/10.5281/zenodo.3509134>

2. sys: Python Software Foundation. Python Language Reference, version 3.11.7. Available at[http://www.python.org](http://www.python.org/)

3. NumPy: Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. In our experiments. NumPy version 1.26.4

4. Seaborn: Waskom, M. L. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. Seaborn version 0.12.2. (Seaborn: Statistical Data Visualization,[https://seaborn.pydata.org](https://seaborn.pydata.org/))

5. Matplotlib 및 matplotlib.pyplot: Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95. Matplotlib version 3.8.0

6. Plotly Express: Plotly Technologies Inc. (2015). Plotly: Collaborative data science. Montréal, QC. Available at:[https://plotly.com](https://plotly.com/). Plotly Express version 5.9.08. statsmodels.api: Seabold, Skipper, and Josef Perktold. "Statsmodels: Econometric and statistical modeling with python." Proceedings of the 9th Python in Science Conference. Vol. 57. 2010. statsmodels version 0.14.09. scipy: Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & van der Walt, S. J. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272. scipy version 1.11.4

1. A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa, RFM ranking – An effective approach to customer segmentation,Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 10,2021, Pages 1251-1257 <https://doi.org/10.1016/j.jksuci.2018.09.004> [↑](#footnote-ref-1)