

[통신사 이탈고객 특성 분석과 충성고객 확보를 위한 전략 수립]

3팀 2차 1조

- 데이터셋 선정

kaggle의 [Telco Customer Churn](#)이라는 통신사 이탈고객에 대한 데이터셋을 분석 대상으로 선정하였다. 약 7000개의 고객데이터 수가 있어 분석하는데 작지 않은 데이터 크기로 판단하였다. 그리고 고객 나이와 가족관계, 서비스 이용, 요금 납부, 계약 기간, 이탈자여부 등을 알 수 있어 분석을 통해 인사이트를 얻기에 충분하기 때문에 위의 데이터셋을 선정하였다.

- 프로젝트 목표

이탈고객의 특성을 분석하여 고객 이탈률을 줄일 수 있는 마케팅 방향을 제시한다. 오랜 시간동안 통신사를 이용하는 고객을 분석하여 지속적인 수익 창출을 목표로 한다.

- 데이터 전처리

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7032 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   SeniorCitizen         7032 non-null  object 
 1   Partner               7032 non-null  object 
 2   Dependents            7032 non-null  object 
 3   tenure                7032 non-null  int64  
 4   PhoneService          7032 non-null  object 
 5   MultipleLines         7032 non-null  object 
 6   InternetService       7032 non-null  object 
 7   OnlineSecurity        7032 non-null  object 
 8   OnlineBackup          7032 non-null  object 
 9   DeviceProtection      7032 non-null  object 
10   TechSupport           7032 non-null  object 
11   StreamingService      7032 non-null  object 
12   StreamingTV           7032 non-null  object 
13   StreamingMovies       7032 non-null  object 
14   Contract              7032 non-null  object 
15   PaperlessBilling      7032 non-null  object 
16   PaymentMethod         7032 non-null  object 
17   MonthlyCharges        7032 non-null  float64 
18   TotalCharges          7032 non-null  float64 
19   Churn                 7032 non-null  object 
dtypes: float64(2), int64(1), object(17)
memory usage: 1.1+ MB
```

1. **customerID**와 **gender** 열을 삭제하였다. **customerID**는 식별 용도라서 중복이 없어 삭제했다. **gender**는 카이제곱검정 결과 이탈에 영향을 주지 못한다 판단하였다.

카이제곱 검정 통계량: 0.5134

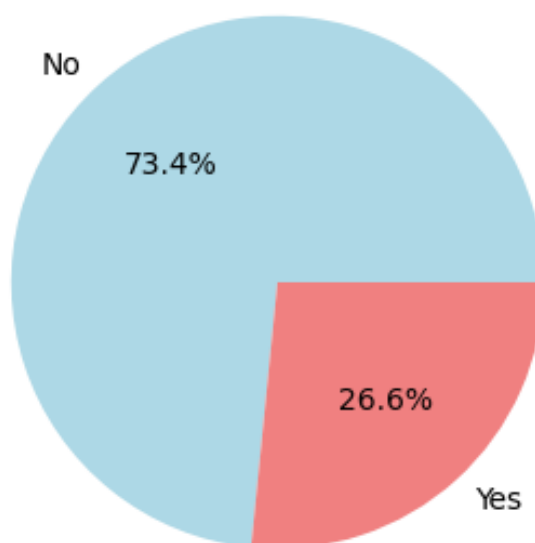
P-value: 0.9722

이탈과 성별은 통계적으로 유의미한 차이가 없습니다.

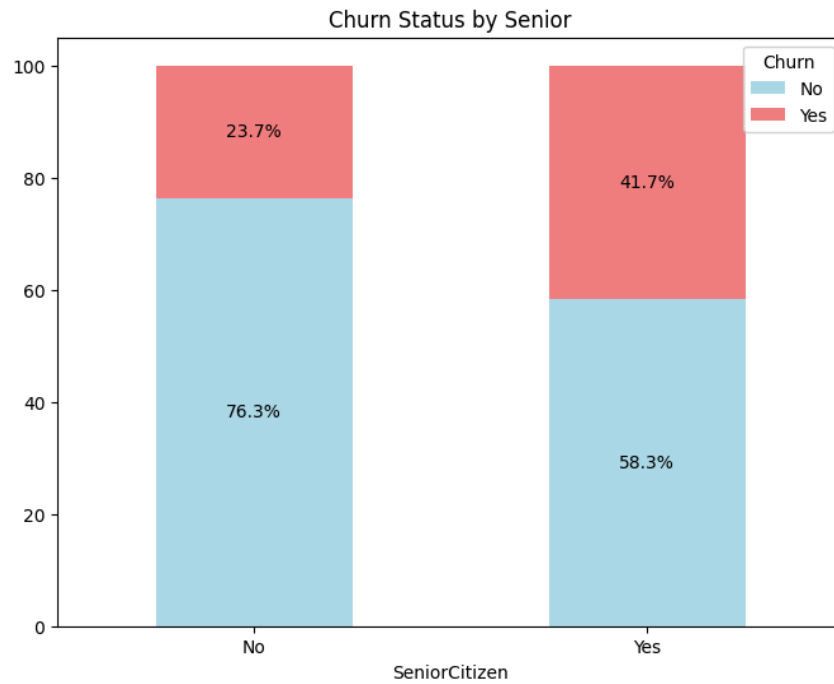
2. **TotalCharges** 데이터 타입을 숫자형으로 변환하였고, **SeniorCitizen**의 값을 0은 No, 1을 Yes로 변환했다.
3. 결측치는 **TotalCharges**에서 11개가 있었다. 모두 **Tenure**가 0인 행들이 있었고, 이용기간이 1달도 채 안된 고객은 이탈자 통계에 넣지 않는게 낫다고 판단하였다.
4. **StreamingTV**와 **StreamingMovies** 두 열의 큰 차이가 없다고 생각하여 **StreamingService**로 둘 중 하나의 서비스라도 이용하는 고객을 알 수 있는 열을 생성하였다.

- EDA

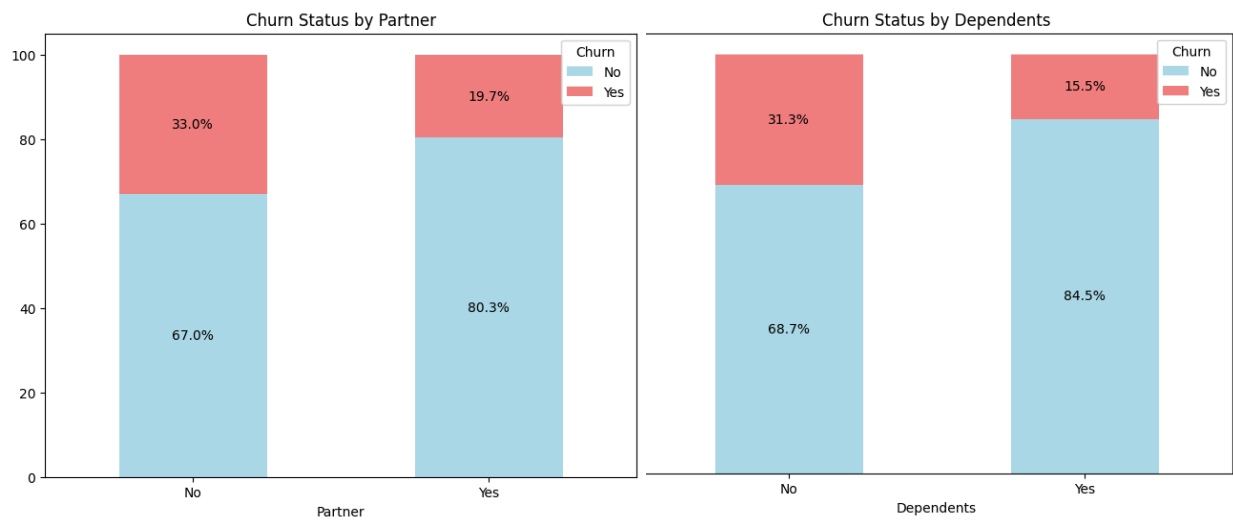
Distribution of Churn



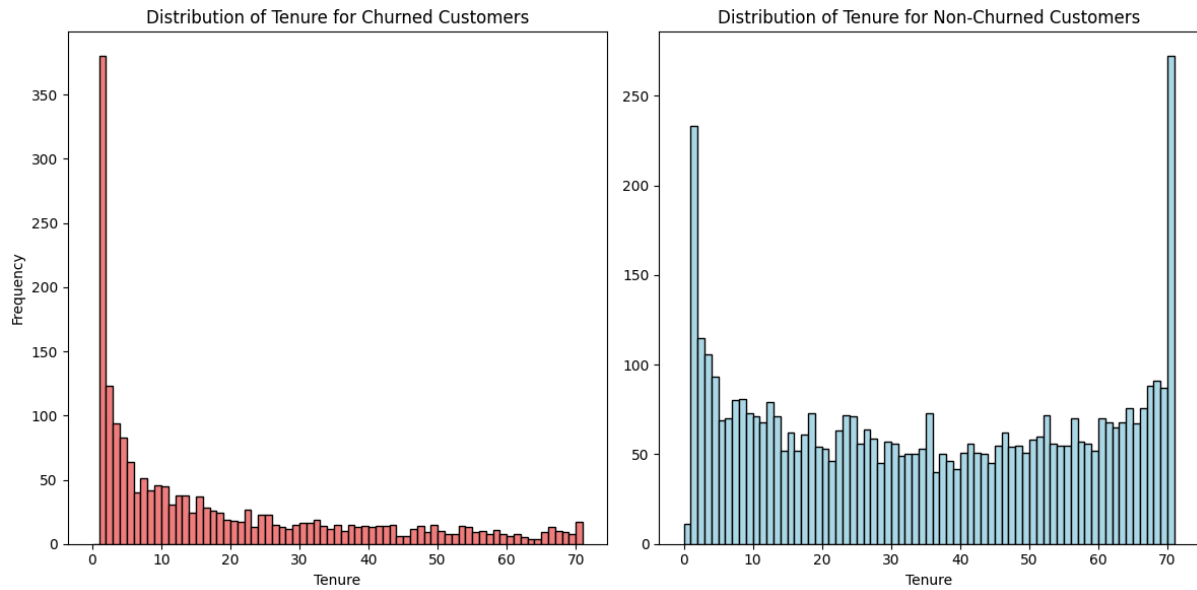
전체 이탈율은 26.6%이다.



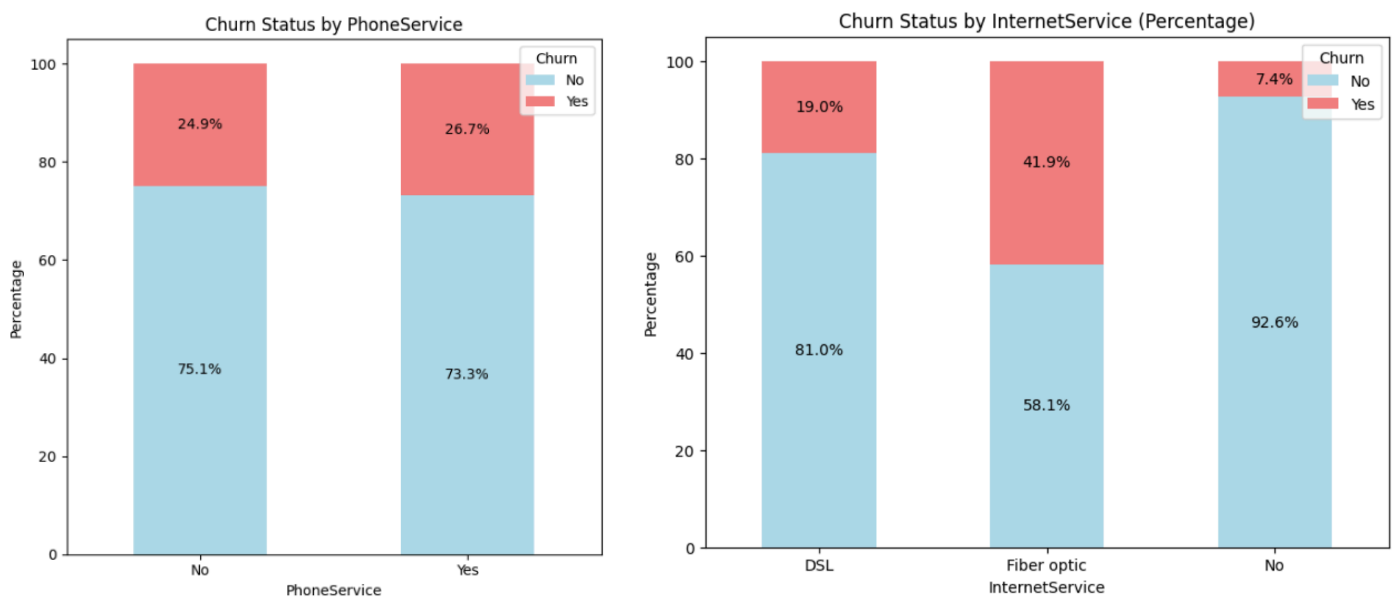
시니어 고객의 이탈률은 41.7%, 시니어가 아닌 고객의 이탈율은 23.7%이다.



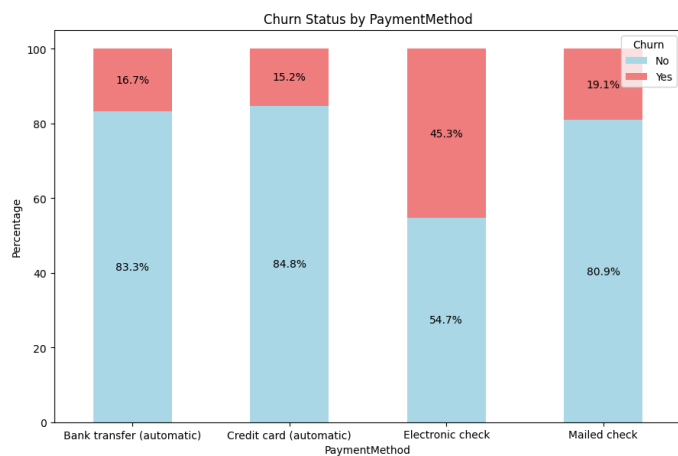
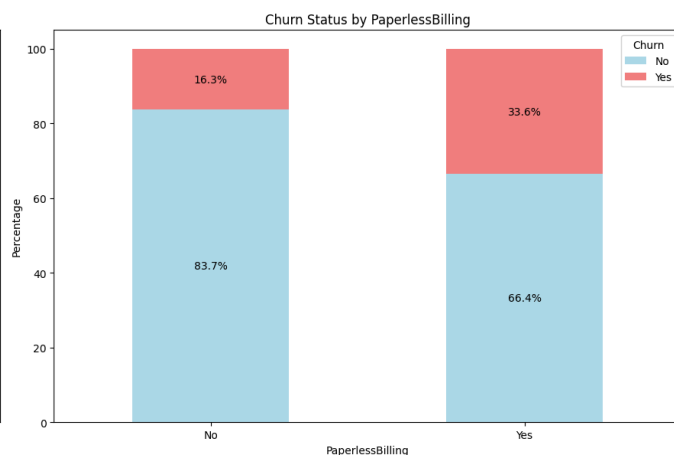
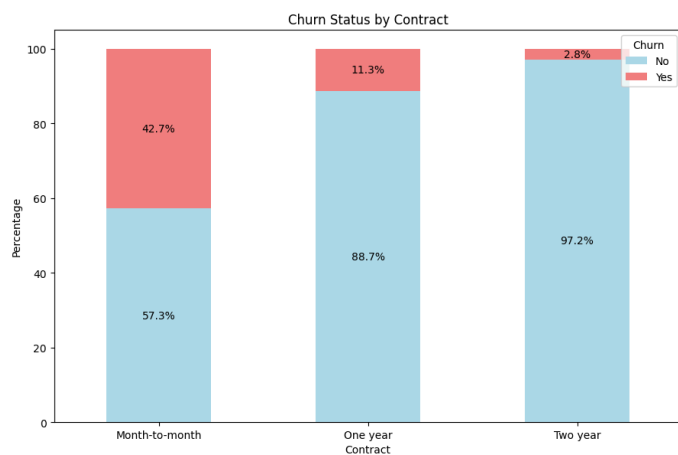
파트너가 있는 고객의 이탈율은 19.7%, 파트너가 없는 고객의 이탈율은 33.0%이다.
 부양가족이 있는 고객의 이탈율은 15.5%, 부양가족이 없는 고객의 이탈율은 31.3%이다.



왼쪽 그래프는 이탈자의 이용 개월수, 오른쪽 그래프는 비이탈자의 이용 개월 수이다.
이탈자는 첫달이 지나면 급격히 줄어들고 비이탈자는 장기간 이용하는 경우가 많다.



전화서비스를 이용하는 고객의 이탈율은 **26.7%**이고 이용하지 않는 고객의 이탈율은 **24.9%**이다. 인터넷 서비스 중 **DSL**(디지털 가입자 회선)고객의 이탈율은 **19.0%**, **Fiber optic InternetService**(광 섬유 인터넷)고객의 이탈율은 **41.9%**, 인터넷을 이용하지 않는 고객의이탈율은 **7.4%**이다.

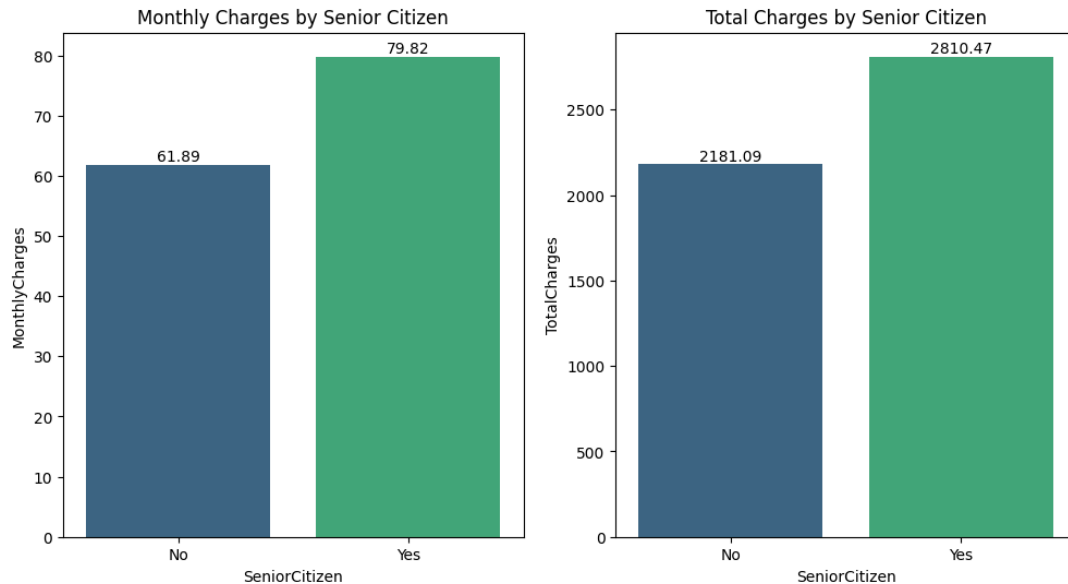


월마다 계약하는 고객의 이탈율은 **42.7%**이고 1년, 2년 단위로 계약하는 고객의 이탈율은 각각 **11.3%**, **2.8%**이다. 종이 없는 청구서를 이용하는 고객의 이탈율은 **33.6%**, 이용하지 않는 고객의 이탈율은 **16.3%**이다. 결제방법에서는 **Electronic check** 고객의 이탈율이 **45.3%**로 제일 높았으며 **Credit card(automatic)** 고객의 이탈율이 **15.2%**로 가장 낮았다.

- 데이터 분석 및 시각화

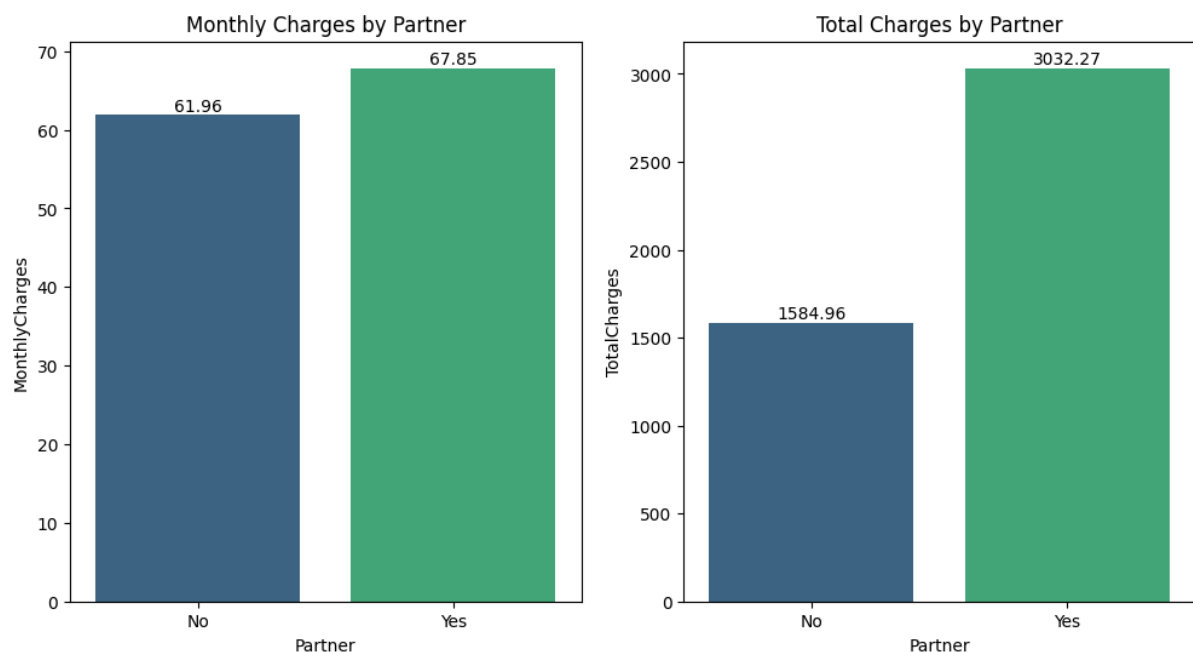
1)나이,가족관계 별 이탈자

(1)시니어의 월 요금과 총 요금



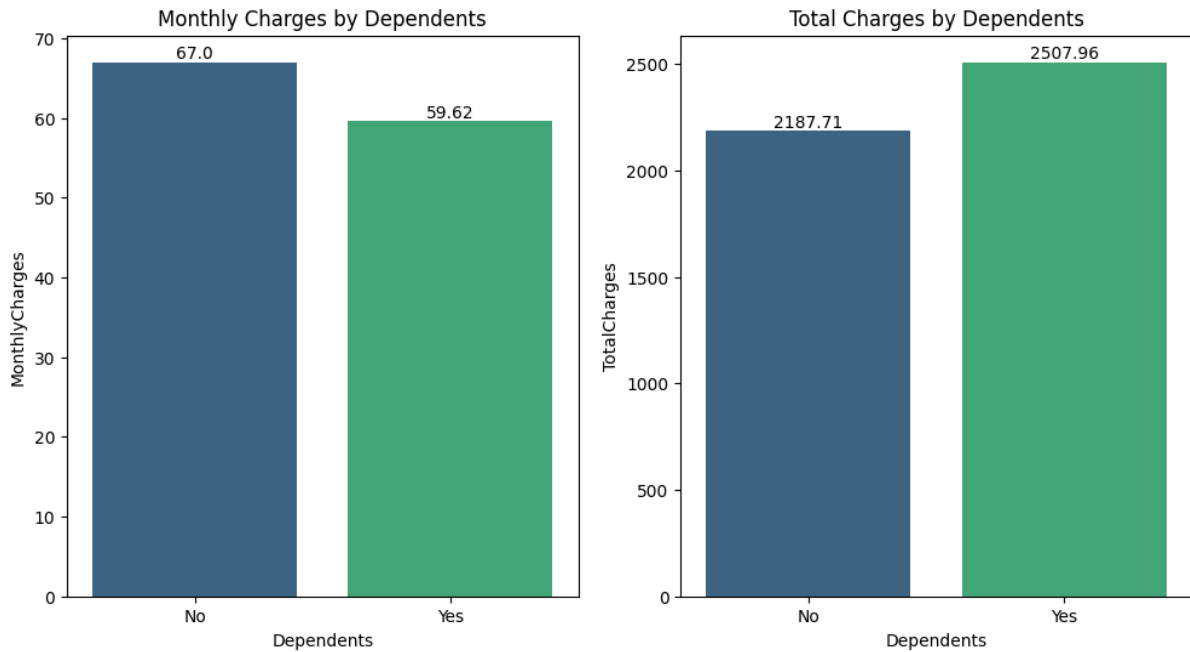
시니어 고객은 시니어가 아닌 고객보다 평균 월 요금과 총 요금이 더 크다. 하지만 시니어 고객 이탈율은 약 40%로 평균 이탈율보다 높기 때문에 시니어 고객을 유지하는 것이 필요하다. 따라서 시니어 고객은 수익을 증가하는데 중요한 타겟층이다.

(2)파트너 유무에 따른 월 이용요금과 총 이용요금



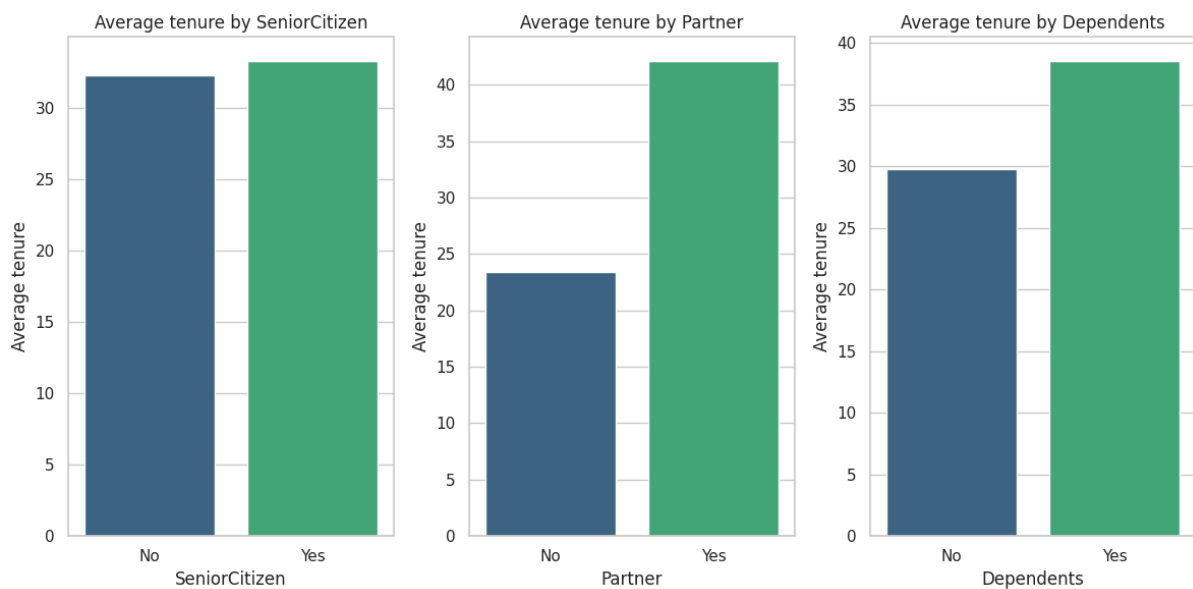
파트너가 있는 고객은 특히 총 이용요금이 크다. 파트너가 있는 고객이 상대적으로 충성도가 높다.

(3) 부양가족 유무에 따른 월 이용요금과 총 이용요금

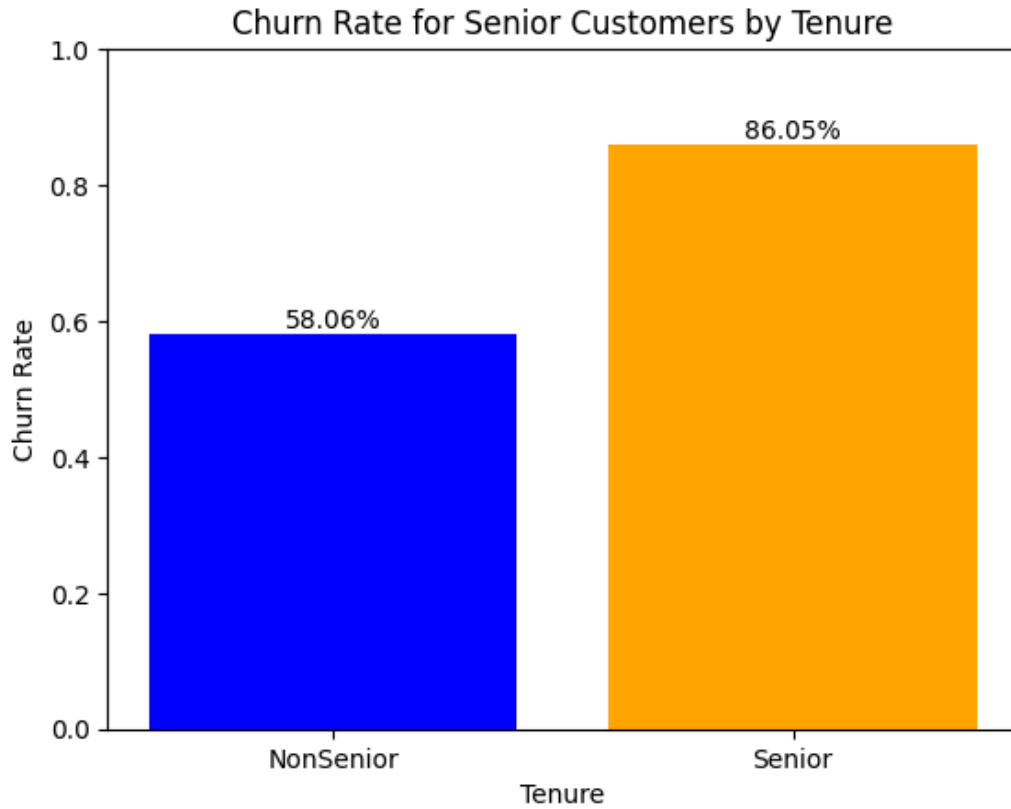


부양가족이 있는 고객은 월 이용요금은 작은 대신 총 이용요금은 더 크다. 부양가족이 있는 고객은 더 오래 서비스를 이용하고 있다.

(4) 나이와 가족형태에 따른 평균 이용기간

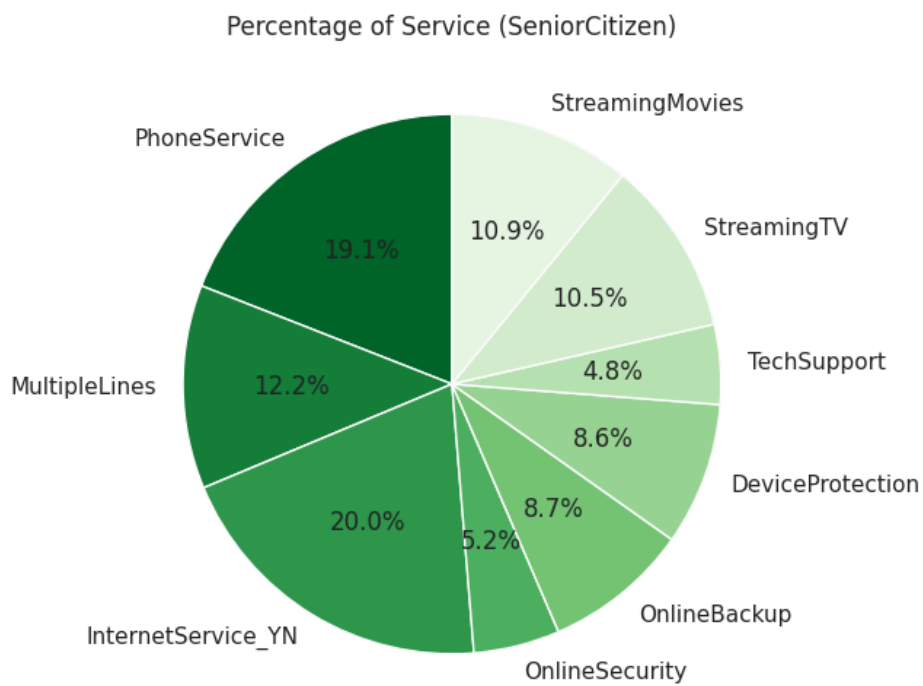


나이는 파트너와 부양가족 유무에 비해 이용기간이 차이가 나지 않는다. 시니어는 이용기간이 길다고 서비스를 계속 이용하지 않는다.

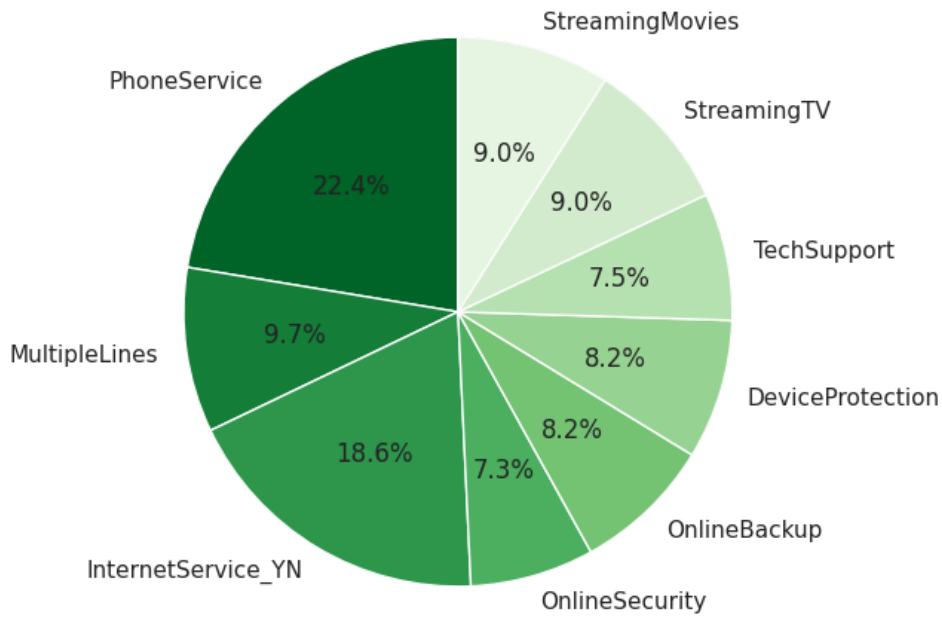


이용개월수가 1달에서 이탈한 고객이 시니어가 86.05%로 아닌 고객보다 높았다. 시니어 고객 초반 유지 전략을 해야 한다.

(5)시니어 고객과 시니어 아닌 고객의 서비스별 이용비율



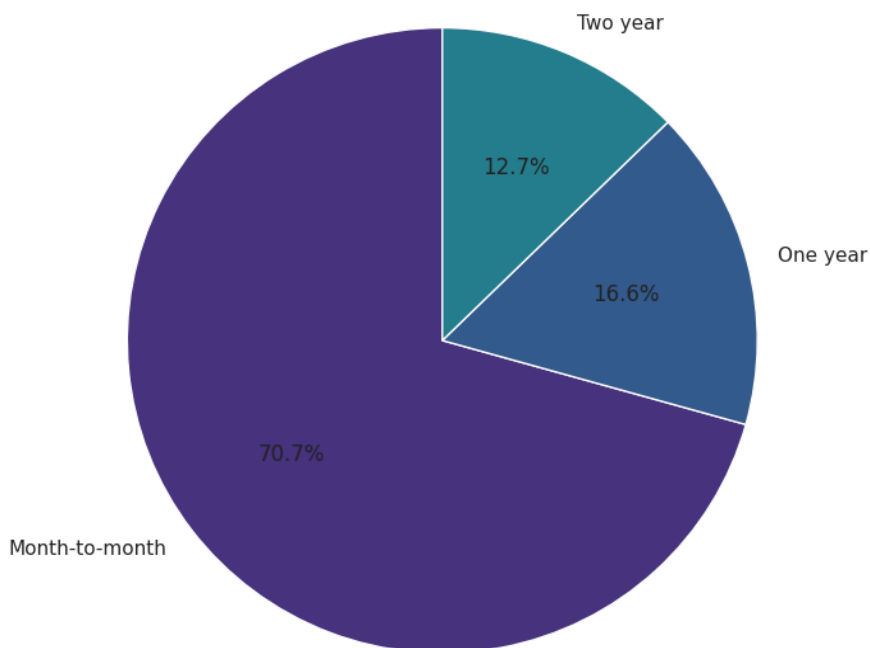
Percentage of Each Service (NonSeniorCitizen)



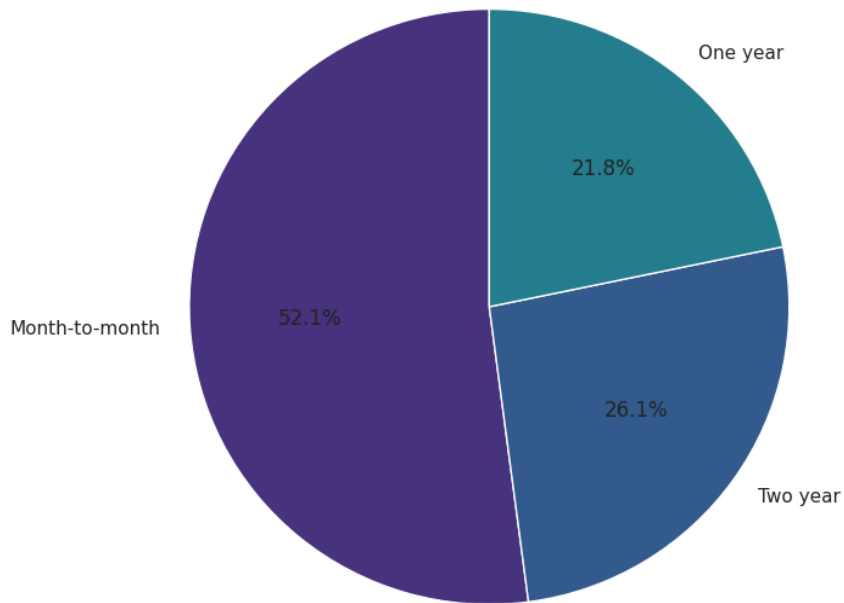
시니어 고객은 근소하게 인터넷 서비스와 스트리밍 서비스를 더 많이 이용하고 있다.

(6)시니어 고객과 시니어 아닌 고객의 계약기간별 비율

Senior Citizen Contract Distribution



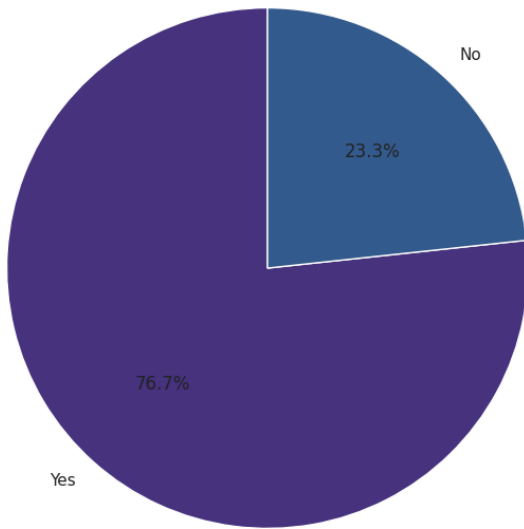
Non Senior Citizen Contract Distribution



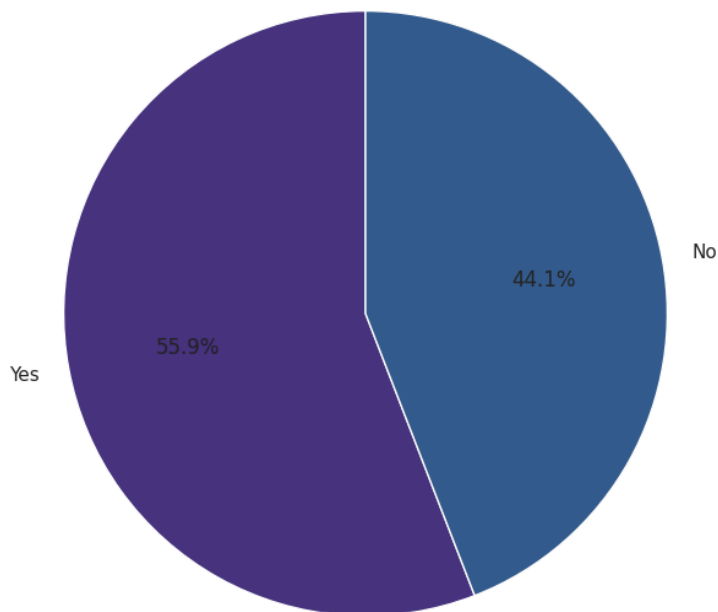
시니어는 월마다 계약을 많이 하고 있었다. 시니어 고객에게 계약기간을 월마다 하는 것보다 연단위로 하도록 유도해야 한다. 월마다 계약하는 것보다 연단위로 계약할수록 이탈할 확률은 줄어들이기 때문이다

(7)시니어 고객과 아닌 고객의 PaperlessBilling 비율

Senior Citizen PaperlessBilling Distribution



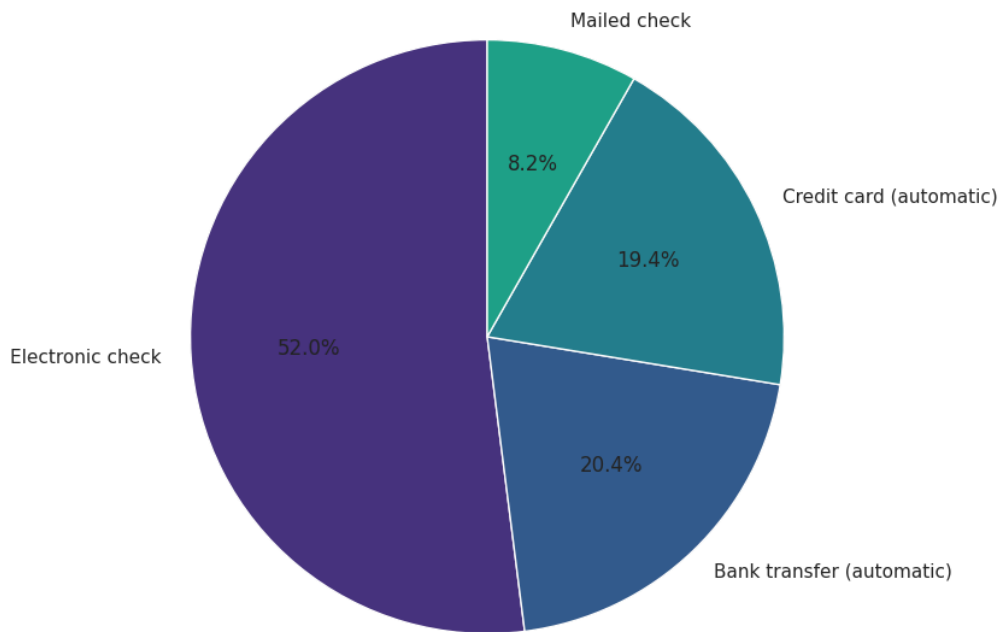
Non Senior Citizen PaperlessBilling Distribution



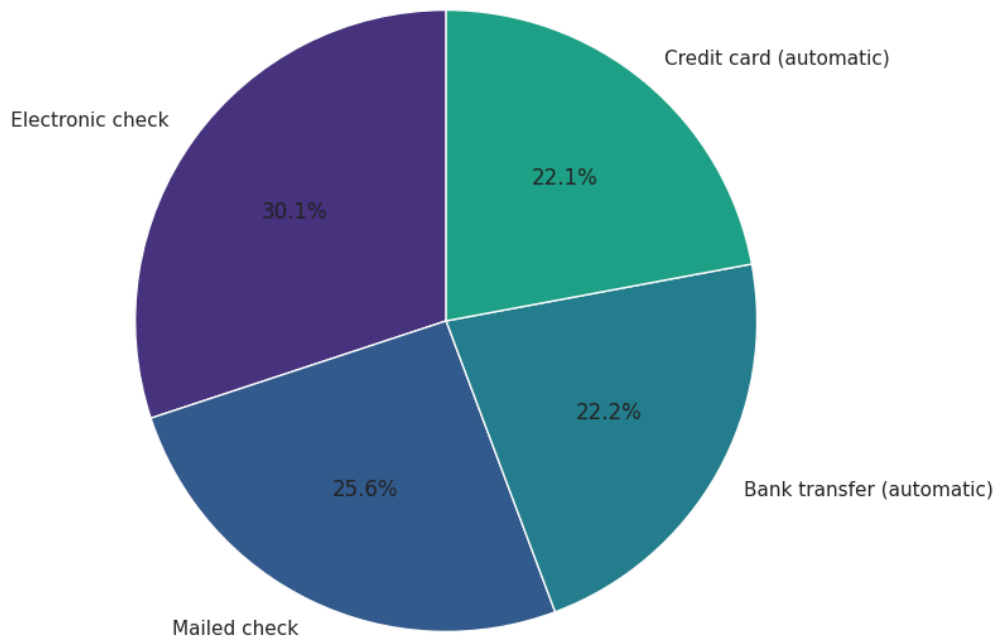
시니어 고객은 **Paperless Billing**을 더 이용하고 있었다. 보통 종이 없는 청구서는 고객과 기업에게 환경, 보안, 편의성 등에 대해 긍정적인 영향을 준다. 시니어가 종이 없는 청구서를 많이 사용한다고 이탈을 많이 한다 보기는 어려워 보인다.

(8)시니어 고객과 아닌 고객의 납부방법 비율

Senior Citizen PaymentMethod Distribution



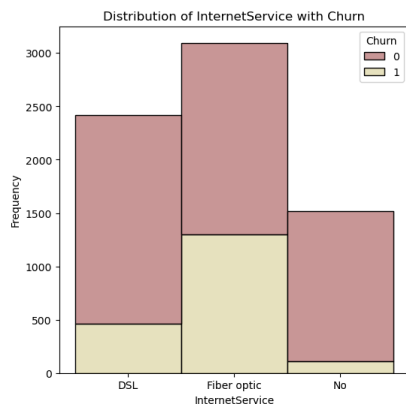
Senior Citizen PaymentMethod Distribution



시니어 고객은 상대적으로 **Electronic check**를 많이 사용하고 **Mailed check**는 적게 사용한다. 앞서 **Paperless Billing**과 비슷한 맥락 일 수 있다. 시니어가 오히려 디지털 결제를 잘 활용하는 것으로 보인다. 디지털결제가 불편한 것인지, 메리트가 별로 없는 것에 실망한 것인지 확실치는 않다. 분명한건 시니어는 디지털 결제를 선호한다.

2)인터넷 서비스와 핸드폰 서비스

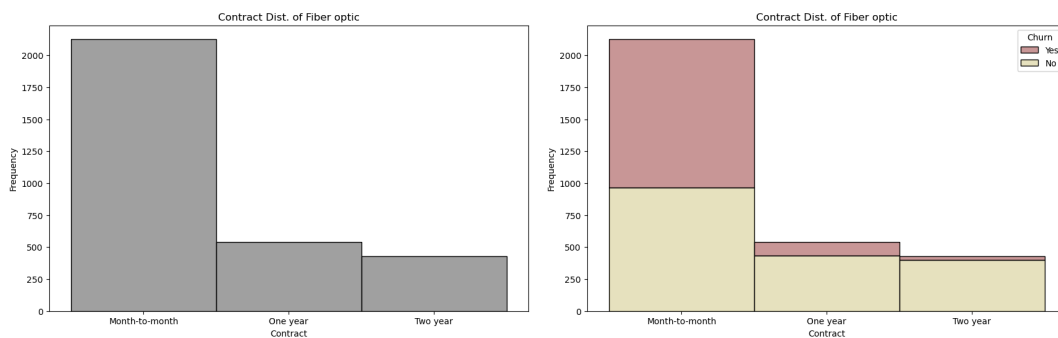
(1)인터넷 서비스



먼저 살펴보자면 이탈 고객 중 광통신(Fiber optic)을 이용하는 고객이 많았다.

각 인터넷 서비스를 계약 형태 (1) Fiber optic (2) DSL (3) No internet service 에 따라 살펴보았다.

1. Fiber optic

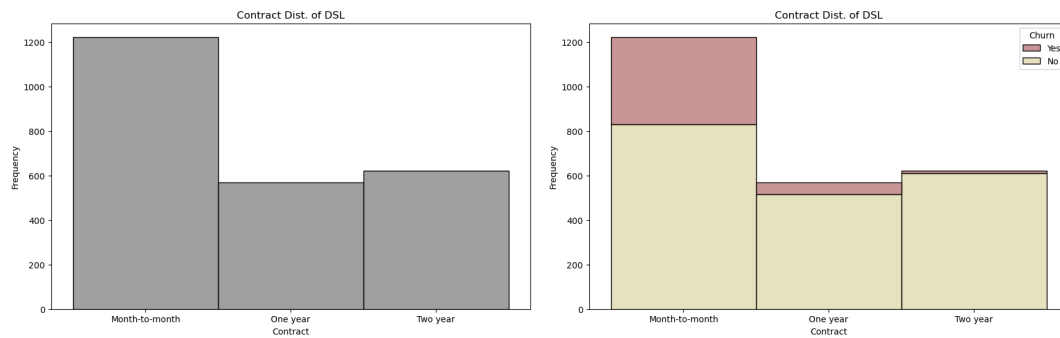


좌측 히스토그램은 광통신을 이용하는 고객의 계약 형태 분포이다.

우측 히스토그램은 이탈여부(Churn Yes/No) 정보도 포함한 분포이다.

Month-to-month 계약을 한 고객 중 절반 이상이 이탈하였음을 확인할 수 있다.

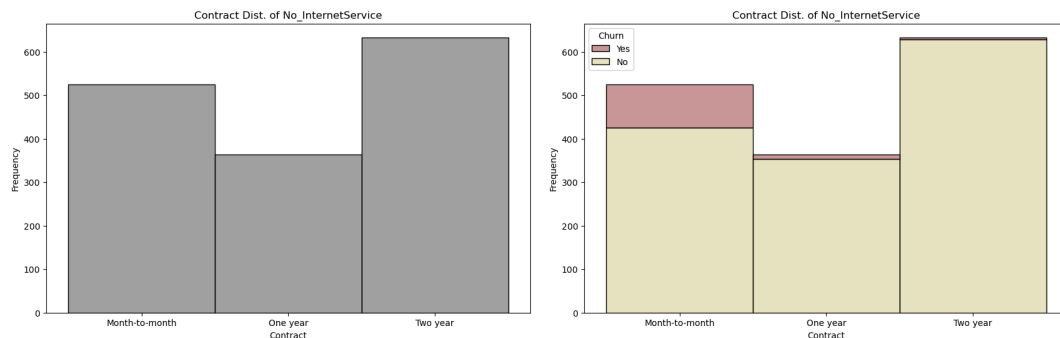
2. DSL



좌측 히스토그램은 DSL을 이용하는 고객의 계약 형태 분포이다.
우측 히스토그램은 이탈여부(Churn Yes/No) 정보도 포함한 분포이다.

Month-to-month 계약을 한 고객이 많이 이탈하였지만, Fiber optic을 이용한 고객의 이탈률보다는 이탈률이 낮았다.

3. No internet service



좌측 히스토그램은 인터넷서비스를 이용하지 않는 고객의 계약 형태 분포이다.
우측 히스토그램에 이탈여부(Churn Yes/No) 정보도 포함하였다.

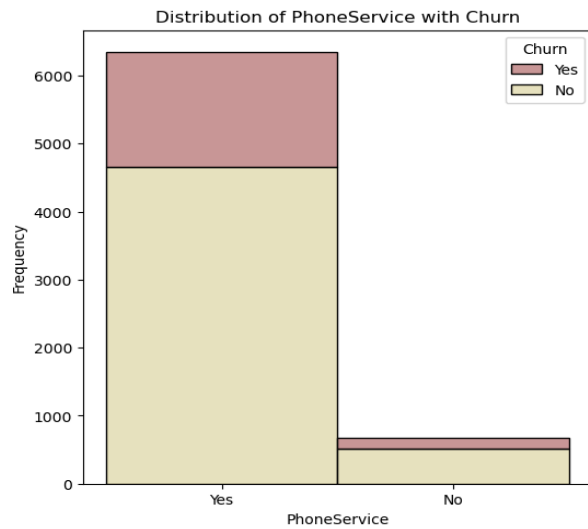
마찬가지로 Month-to-month 계약을 한 고객이 많이 이탈하였지만, Fiber optic이나 DSL의 이탈률과 비교하였을 때 더 낮은 이탈률이 나타났다.

종합하자면 이탈률은 Fiber optic > DSL > No Internet Service 순으로 높았으며, 이탈자는 대부분(수치가 필요하다면 수치도) Month-to-month 계약에서 발생하였다.

분석으로 얻은 결과는 Fiber optic > DSL > No Internet Service 순으로 이탈률이 높았다.
그리고 Month-to-month 계약 고객 자체로 이탈률이 높다. 일반적인 contract 비율 대비, Fiber optic은 Month-to-month 계약을 많이 한다. Fiber optic은 다른 인터넷 서비스에 비해 비싸고 개인가구용보다는 기관, 공용으로 쓰이고 때문이다. No internet service는 two-year 계약을 많이 하고 DSL은 일반적인 추세를 보인다.

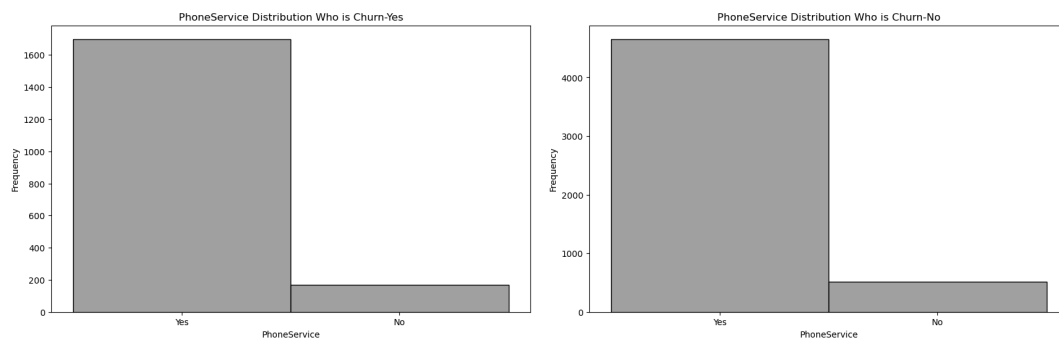
(2)핸드폰 서비스

1.PhoneService



핸드폰 서비스 사용 여부에 따른 고객의 분포를 이탈 여부와 함께 나타내었다.
핸드폰 서비스를 사용하는 고객이 사용하지 않는 고객보다 월등히 많음을 볼 수 있다.

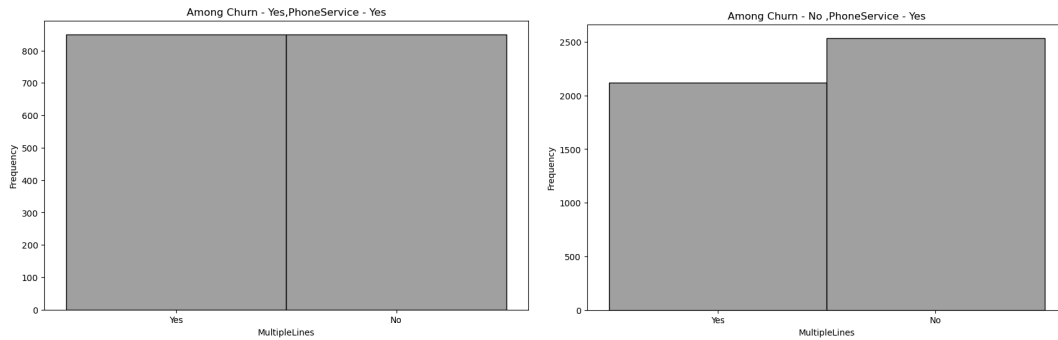
이탈률을 더 자세히 살펴보기 위해 이탈 고객과 비이탈 고객의 서비스 사용 여부를 따로 시각화하였다.



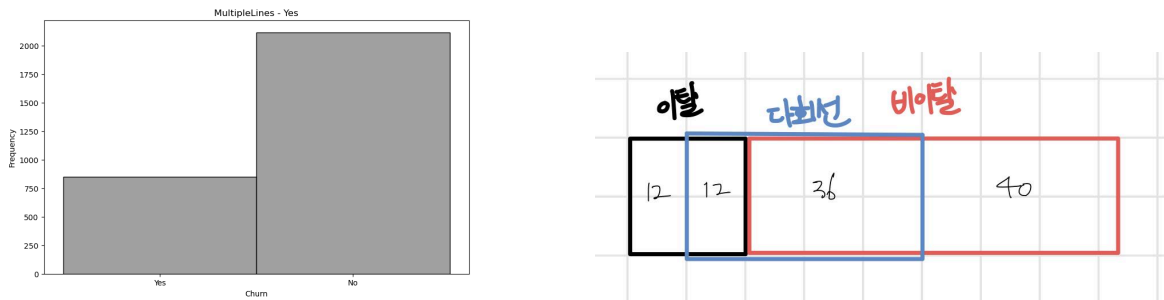
두 분포가 상당히 유사한 분포를 보인다.
핸드폰서비스의 사용 여부가 이탈을 결정짓지는 않은 것으로 판단된다.

2. MultipleLines(다회선)

핸드폰 서비스의 옵션으로 다회선을 이용할 수 있다.

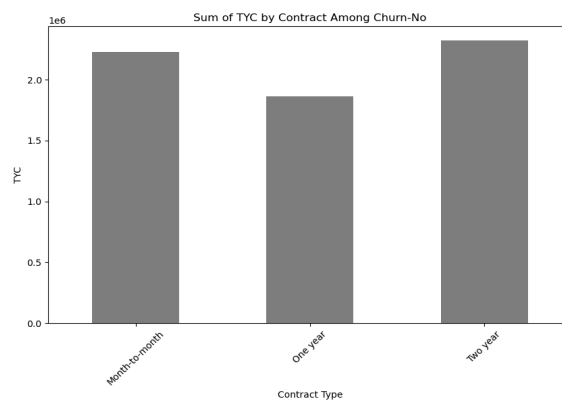


좌측 분포는 이탈 고객의 다회선 이용 비율을, 우측 분포는 비이탈 고객의 다회선 이용 비율을 나타낸다.

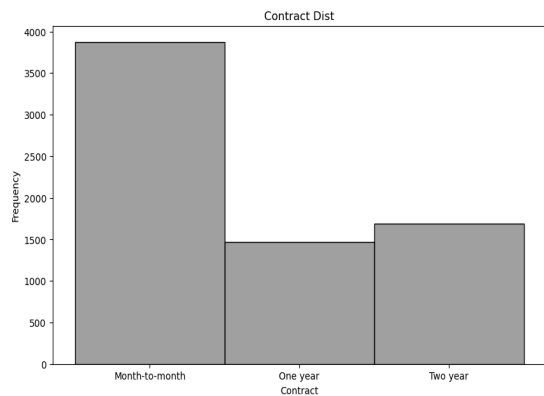


다음은 전체 비율에 대한 이해를 돕기 위한 그림이다. 비이탈자들이 다회선을 근소하게 덜 이용한 경향은 있다. 그렇지만 큰 차이가 없다고 볼 수 있는 추이이다. 다회선의 이용은 이탈자 분석에 큰 변인은 아니다.

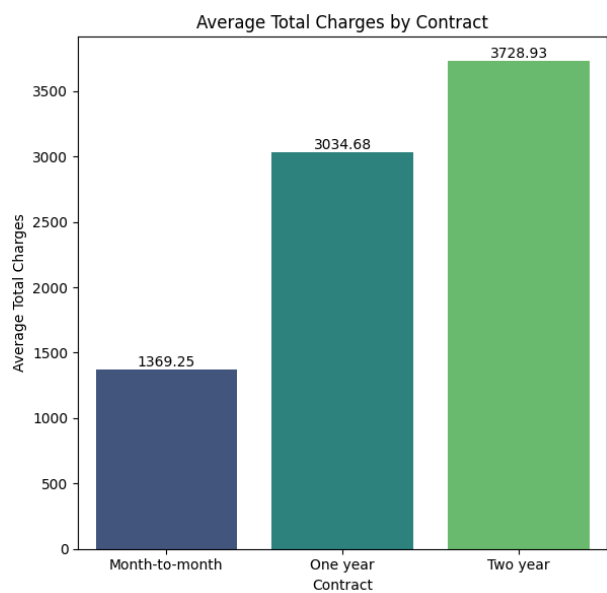
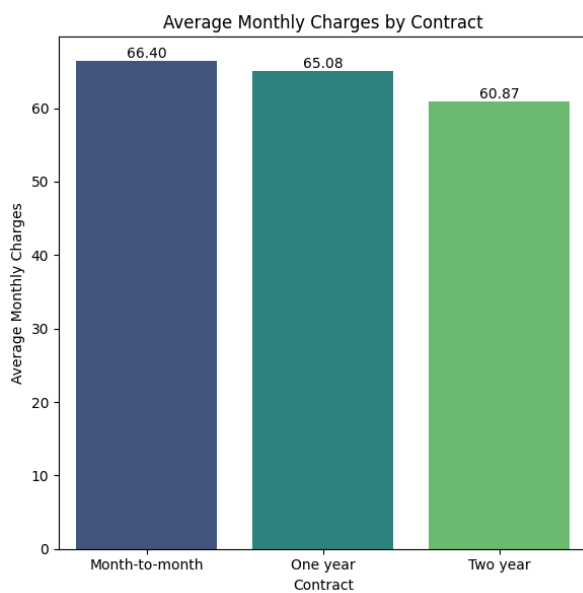
3) 서비스/납부방법/납부주기에 따른 금액



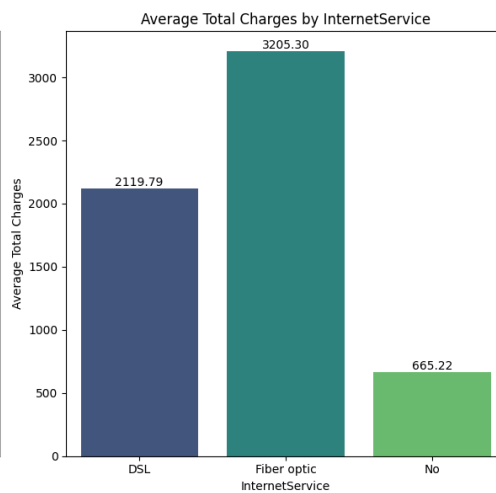
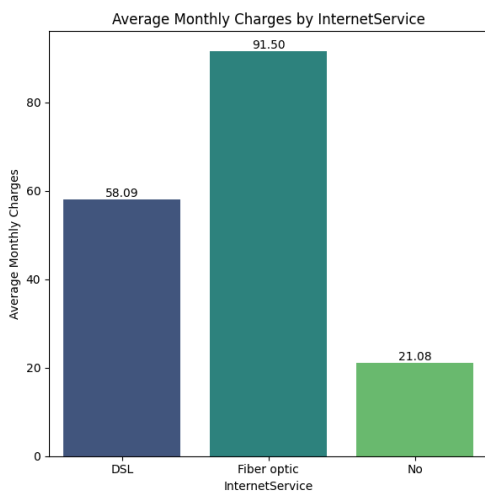
이탈고객의 유무만 나타날 뿐 고객이 이탈이 언제 이탈했는지 알 수 없다. 따라서, 일단 비이탈고객을 대상으로 현재 통신사에 지불하고 있는 금액만 계약형태 별로 알아보았다.



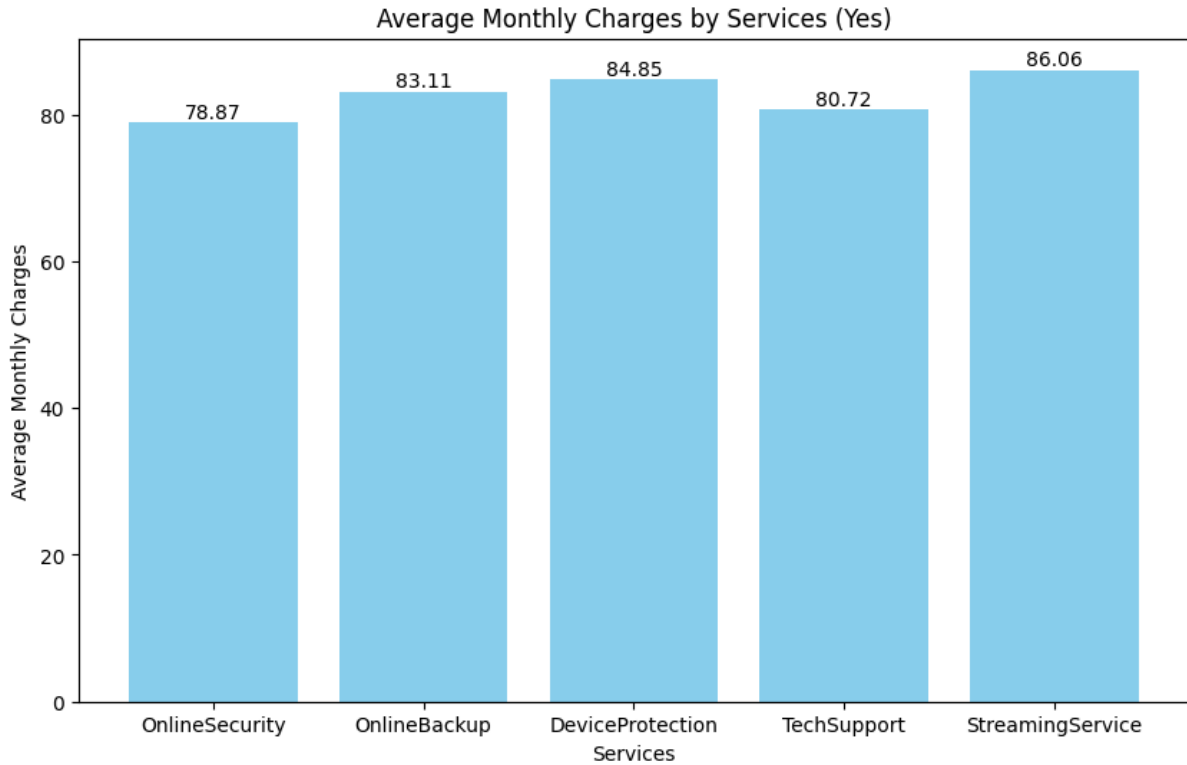
데이터 건수는 **Month-to-month**가 월등히 많지만, 계약기간이 짧고 이탈률이 높아서인지 **Month-to-month**의 매출이 적게 집계되었다. 기업의 입장에서 가장 매출에 도움이 되는 계약 형태는 장기 계약, 특히 **2년 계약**임을 알 수 있다.



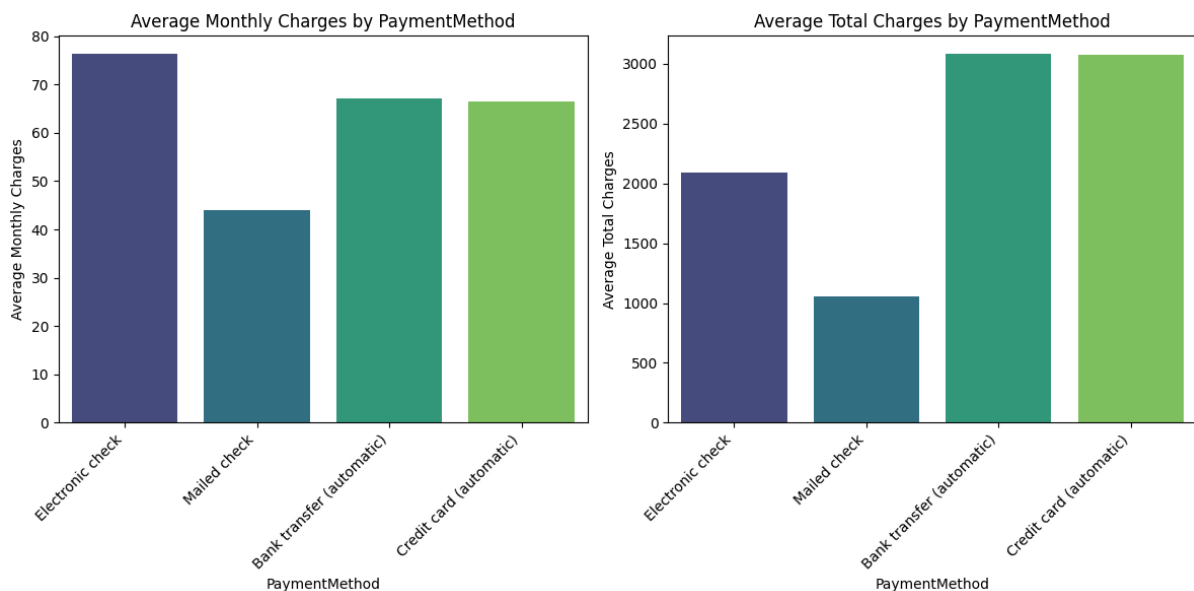
월마다 결제가 월 비용은 가장 높고 총 비용은 가장 낮다.



전화 서비스와 인터넷 서비스 중 인터넷 서비스를 이용하는 고객의 비용이 더 크다. (참고: 전화서비스와 인터넷서비스 중 한 서비스는 무조건 사용, 전화서비스만 사용=1520명, 인터넷서비스만 사용=680명, 두 서비스 사용=4832명)



서비스별 월 비용은 스트리밍서비스가 가장 높지만 모두 비슷하다.



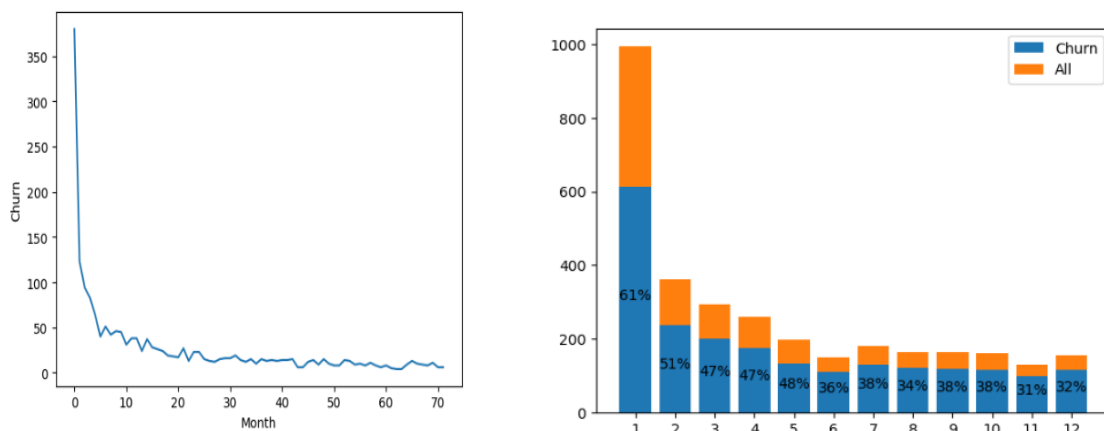
Electronic check가 월 비용이 가장 높았으며 Mailed check가 가장 낮았다. 시니어가 Electronic check를 많이 사용하지만 월 비용이 높은 것을 보아서 가격 측면에서 불만을 갖고 이탈했을수도 있다.

4)사용개월 수에 따른 이탈고객

통신사를 바꾸는 기준 중 대표적인 한 가지가 통신사 계약기간일 것이다. 보통 2~3년을 기준으로 계약을 하기 때문에 해당 기간을 기준으로 살펴보았다. 그리고 위에서 진행했던 분석대로 초반 1년 동안의 이탈율이 가장 눈에 띄게 많았고, 분석하려는 관점에 포함시켰다. 아래 3가지 관점에서 분석을 진행했다.

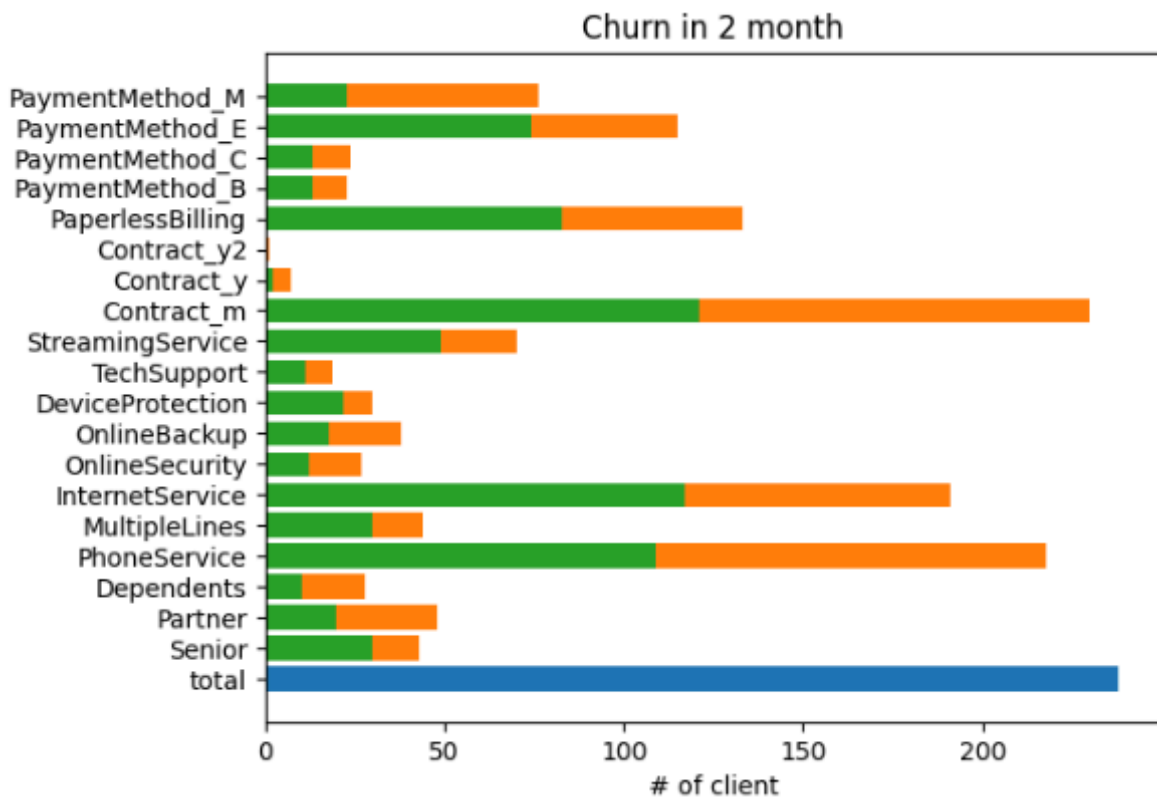
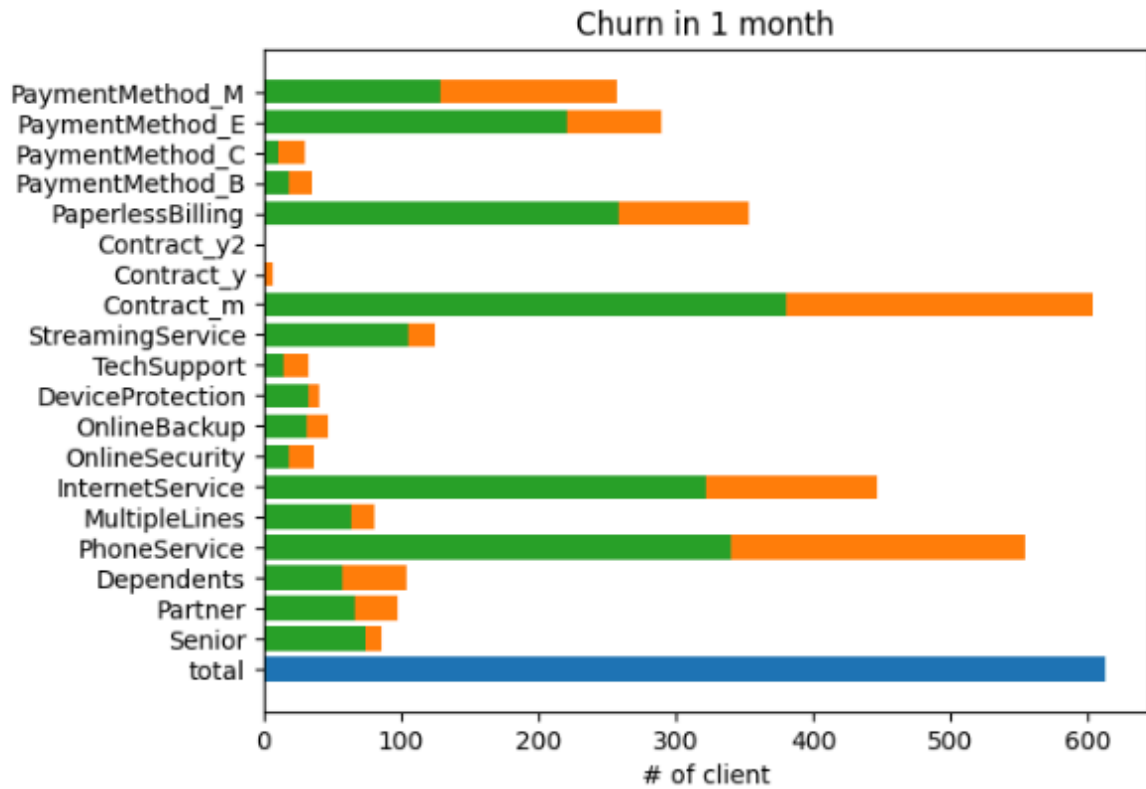
1. 초반 이탈고객
2. 계약기간 만료 이후의 이탈고객
3. 장기고객

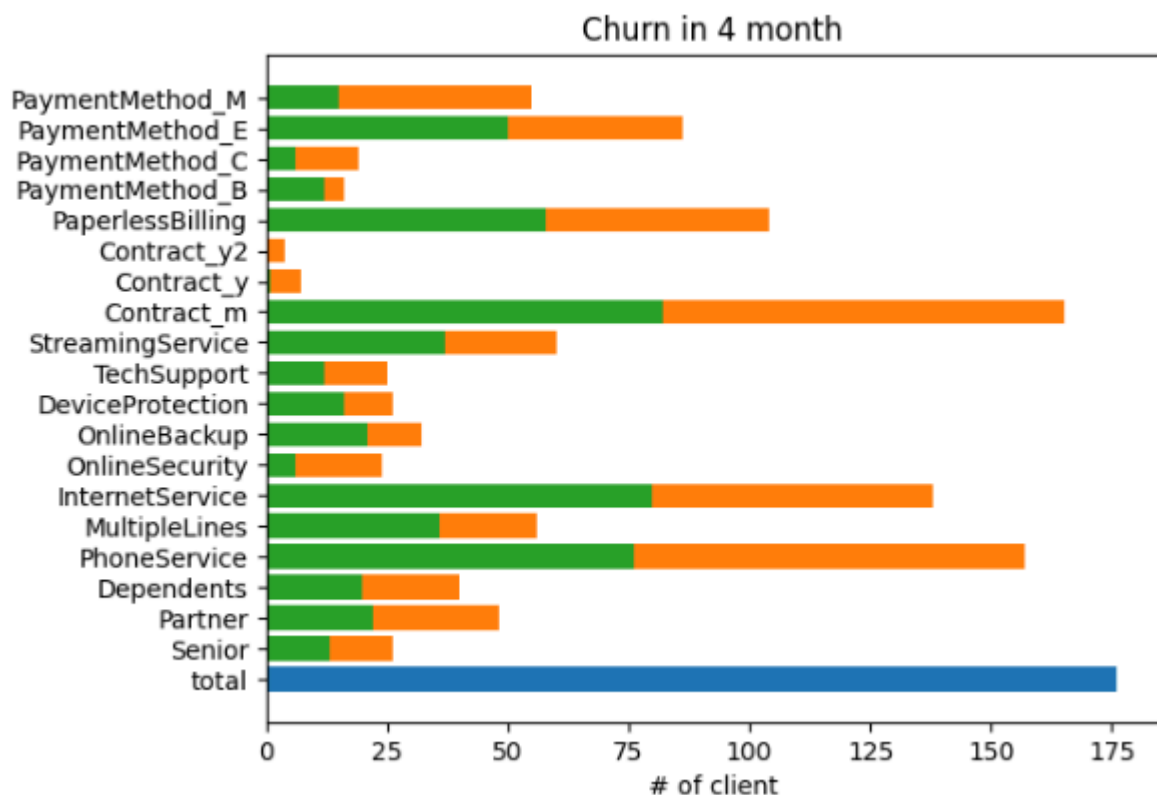
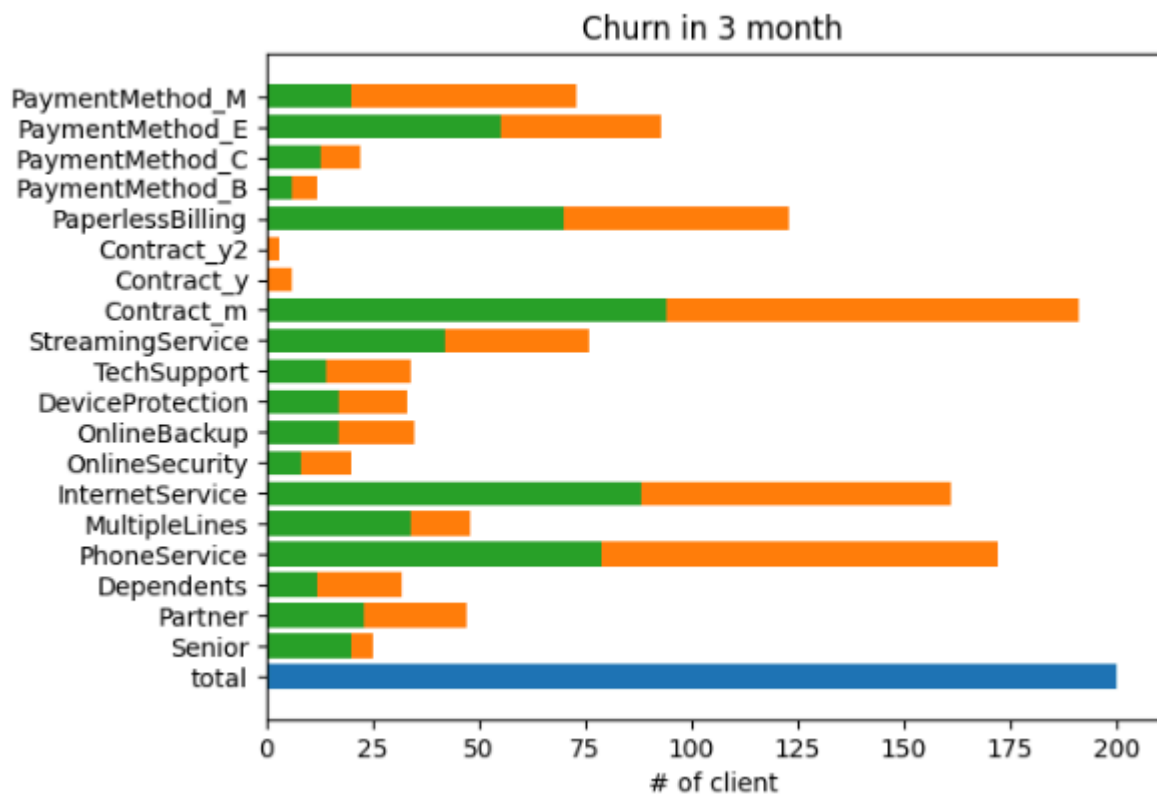
4-1. 첫 1년을 기준으로 봤을 때 1. 실제 이탈고객수도 많았고 2. 가입고객수 대비 이탈고객수도 많았다.

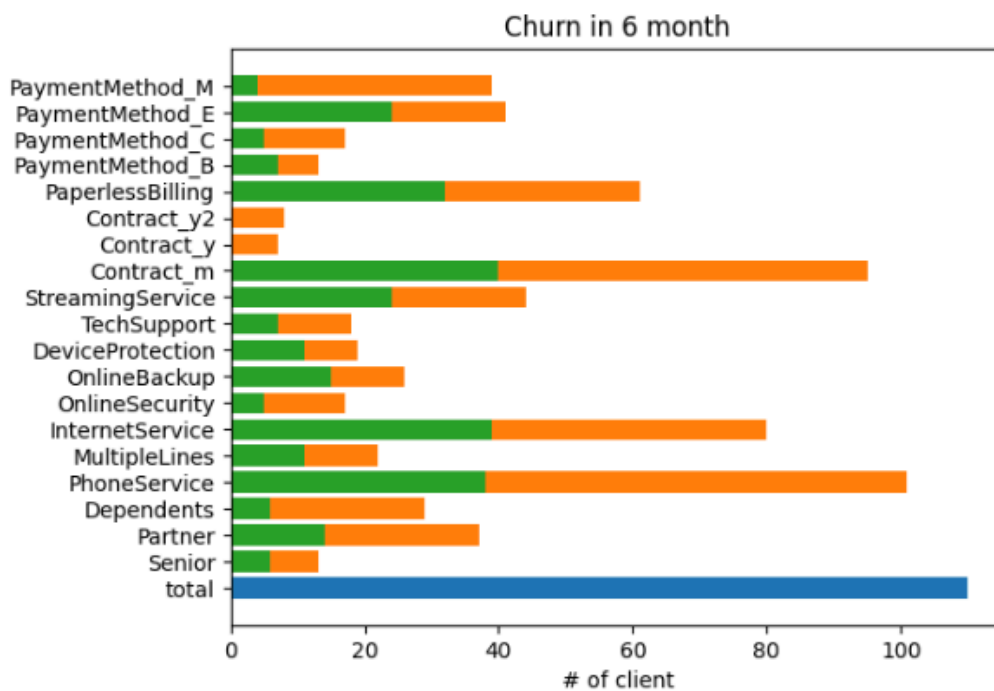
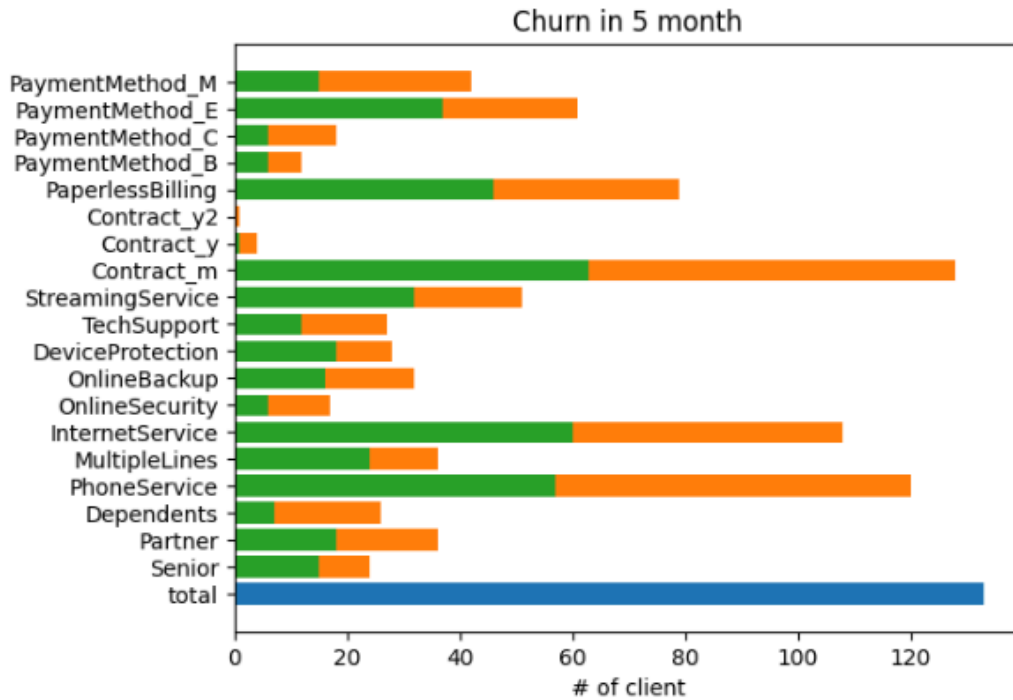


첫달에 이탈하는 고객은 380명으로 가장 많았고 그 이후 부터는 123명 94명 등 시간이 지남에 따라 점차 이탈자가 감소하는 추세가 보였다.

그렇다면 첫 1년동안의 이탈고객의 공통점을 분석하고, 이를 통해 더 이상 고객들이 이탈하지 않도록 해결방안을 모색해보도록 하자.







- * 파란색 (해당 개월 동안 이용하고 있는 전체 고객 수)
- * 주황색 (전체 고객들 중 label을 이용하고 있는 고객 수)
- * 초록색 (label을 이용하고 있는 고객들 중 이탈고객 수)

++6개월이후는 그래프 양상, 이탈율이 비슷하므로 유사데이터로 취급해 생략했다.

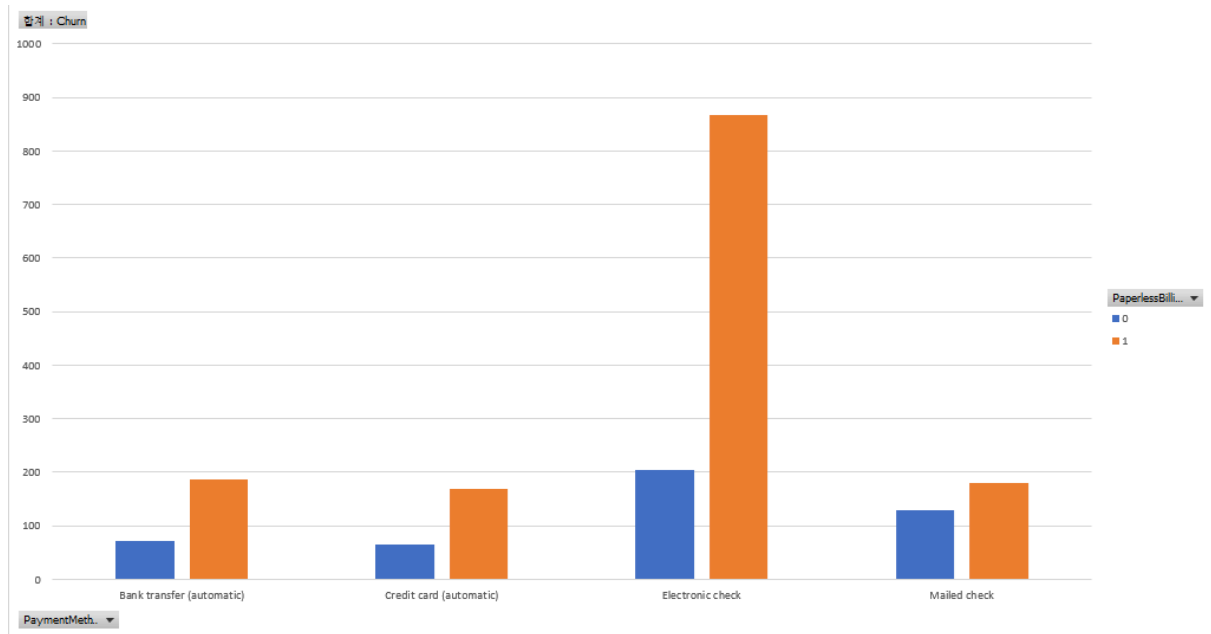
그래프를 봤을 때 높은 비율을 가지는 것은

1.PaperlessBilling, 2.Contract_month_by_month, 3.InternetService, 4.PhoneService 4가지이다.

(1) PaperlessBilling

사람들의 납부방식이 대부분PaymentMethod_E인 것을 감안하여 큰 양상을 보인 것이므로 추가적인 분석이 필요해보인다. (paperlessbilling - payment)

+ 납부방법에 따른 Paperbilling



납부 방법에 대한 PaperlessBilling 수.

위 그래프에서 알 수 있듯이 모든 납부방법은 대부분 PaperlessBilling 방법을 사용하고 있기 때문에 고객 분석할 때 높은 비율을 차지할 수 밖에 없다는 것을 알 수 있다. 한마디로 이탈율에는 큰 영향을 미치지 않는다는 것을 알 수 있다.

(2) Contract_month_by_month

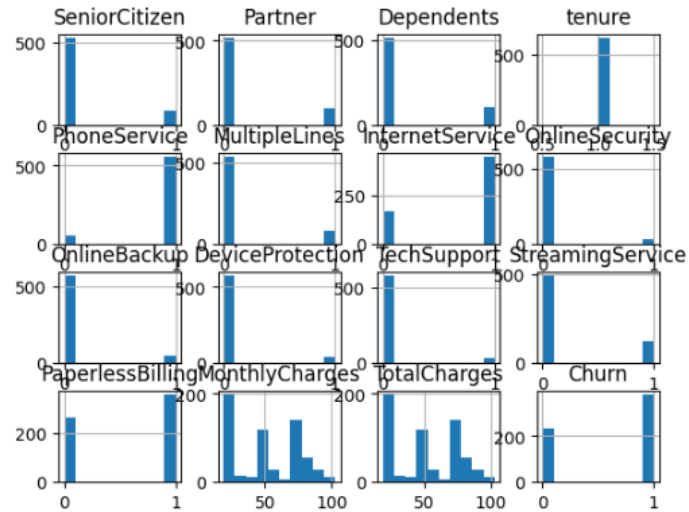
납부주기가 짧아 해당 통신사에 오래 있지 않아도 된다는 고객들의 생각이 짧은 기간 내에 이탈해버리는 가장 중요한 원인이라고 생각했다.

(3) InternetService / PhoneService

이 두 조건은 통신사에서 제시하는 필수조건에 해당하기 때문에 이탈율과는 무관할 것 같다. 물론 서비스 품질이나 조건이 마음에 안들 수도 있겠지만 해당 데이터로는 그러한 점에 대한 분석이 불가능할 거 같다.

따라서 초반 이탈고객들은 납부방법이 month-to-month이라는 공통점이 있기 때문에 처음 신규 고객과 계약할 때 장기적 계약을 하는 것이 고객이 초반에 이탈할 확률을 줄여준다고 볼 수 있다.

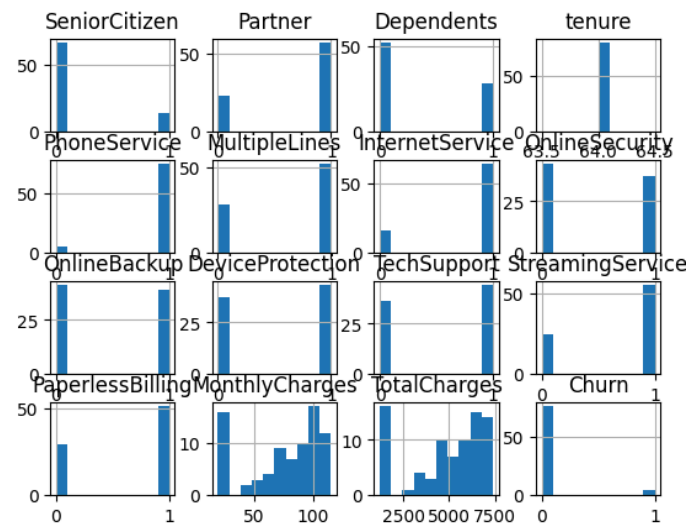
최대 이탈율을 보인 이용 개월 수의 이탈자들의 공통점을 봤으니, 이젠 최소 이탈율을 보인 이용 개월 수(64)의 이탈자와 비교하여 어떤 차이가 있는지 확인해보자



	Senior	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport
total	613	613	613	613	613	613	613	613	613	613
filter	86	98	104	555	81	447	37	47	41	33
churn	74	66	58	341	64	322	18	31	33	14

StreamingService	Contract_m	Contract_y	Contract_y2	PaperlessBilling	PaymentMethod_B	PaymentMethod_C	PaymentMethod_E	PaymentMethod_M
613	613	613	613	613	613	613	613	613
125	604	7	2	353	36	30	290	257
106	380	0	0	259	19	11	221	129

1개월 이용 고객들의 히스토그램



	Senior	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport
total	80	80	80	80	80	80	80	80	80	80
filter	14	57	28	75	52	64	37	39	43	44
churn	1	1	0	3	2	4	2	2	2	3

StreamingService	Contract_m	Contract_y	Contract_y2	PaperlessBilling	PaymentMethod_B	PaymentMethod_C	PaymentMethod_E	PaymentMethod_M
80	80	80	80	80	80	80	80	80
55	11	29	40	51	22	25	16	17
4	0	2	2	3	0	3	1	0

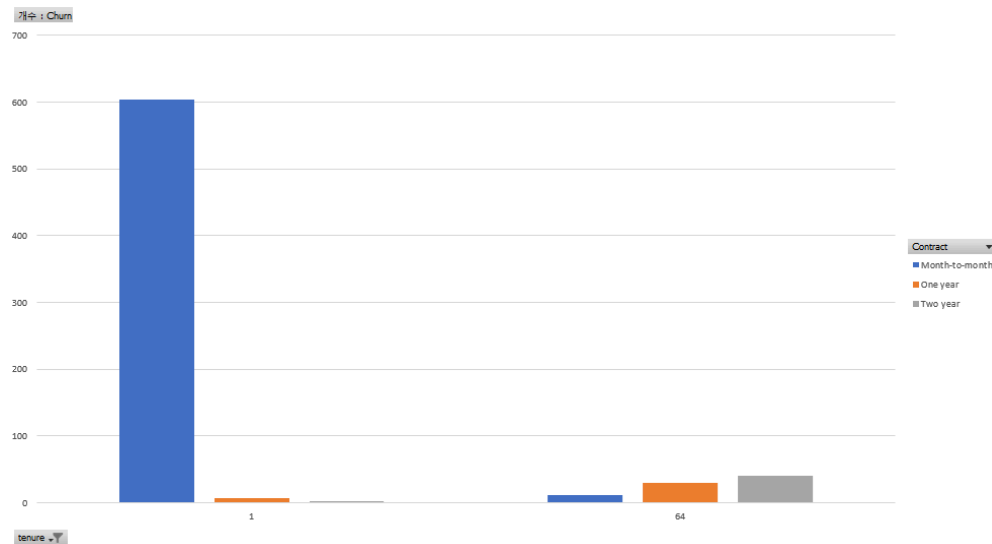
64개월 이용 고객의 히스토그램

1. total : 파란 막대기 총합
2. filter : 해당 label이 True값을 가지는 막대기
3. churn : True값을 가지는 막대기에서의 이탈 고객 (그래프에는 나타나지 않음 .. 비율확인)

가장 이탈자가 많은(380명) 1개월 이용 고객의 히스토그램과 이탈자가 가장 적은(4명) 64개월 이용 고객의 히스토그램과 비교했을 때 64개월동안 이용하고 있는 고객들이 다회선, 인터넷 서비스(OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingService) 등을 1개월 이용 고객보다 더 사용하고 있었다. 그로 인해 매달 청구 요금은 높을 수 밖에 없다.

한마디로 초반 고객들 중 통신사의 서비스를 사용하는 고객일 수록 이탈율이 적다고 볼 수 있다.

+ 이용 개월 수에 따른 계약기간

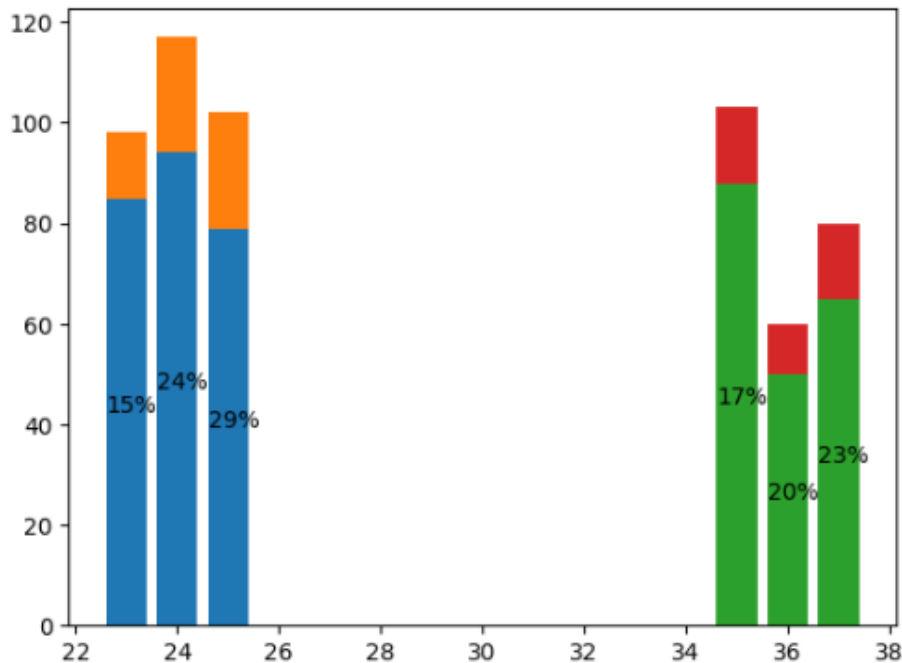


+추가적으로 계약기간 또한 month to month보단 year단위 계약 고객들이 상대적으로 많은 것을 보아 계약기간이 중요하다는 것을 다시 한 번 알 수 있다.

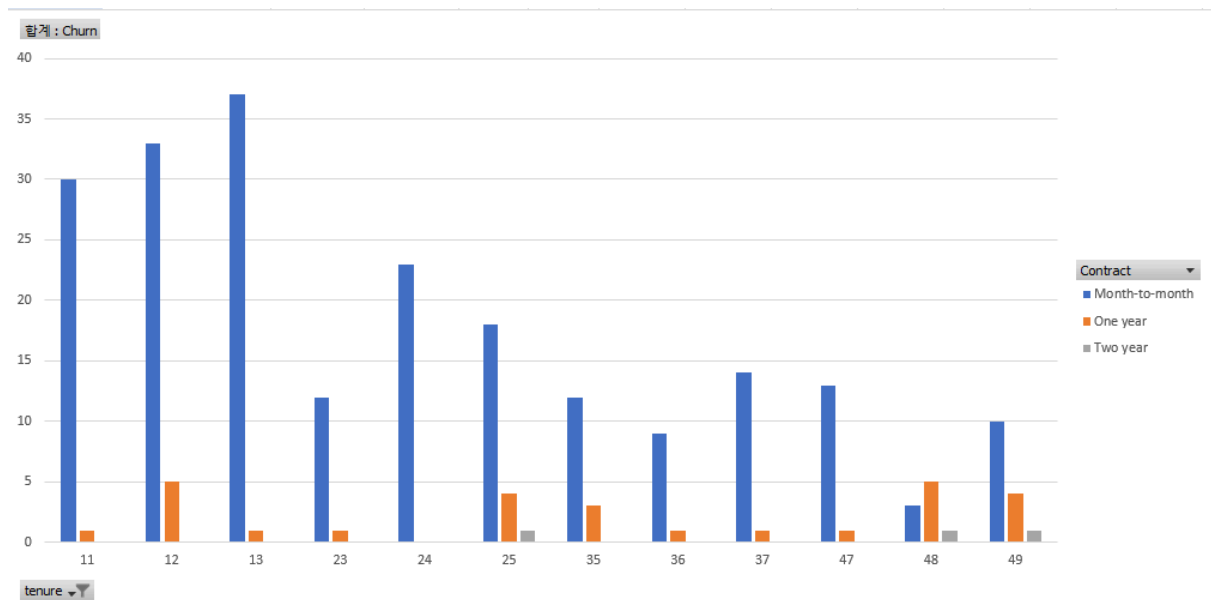
즉 초반 고객을 잡고 싶으면 1. 제공하는 서비스를 이용하도록 유도하고 2. 장기 계약을 따내는 방식을 하는 것이 좋아보인다.

4-2. 계약기간 만료 이후의 고객

계약기간이 평균 2~3년인 것을 바탕으로 정말 이탈율이 계약기간을 기준으로 바뀌는지 확인해 보았다.



위 그래프는 계약기간(24개월, 36개월)의 1개월 전후 이탈율을 나타냈다. 일반적 계약기간을 기준으로 이탈율이 점점 증가하고 있는 것을 알 수 있었다.



위 그래프는 계약기간(24개월 36개월 48개월)의 1개월 전후 계약종류 별 이탈 고객 수를 나타낸다.

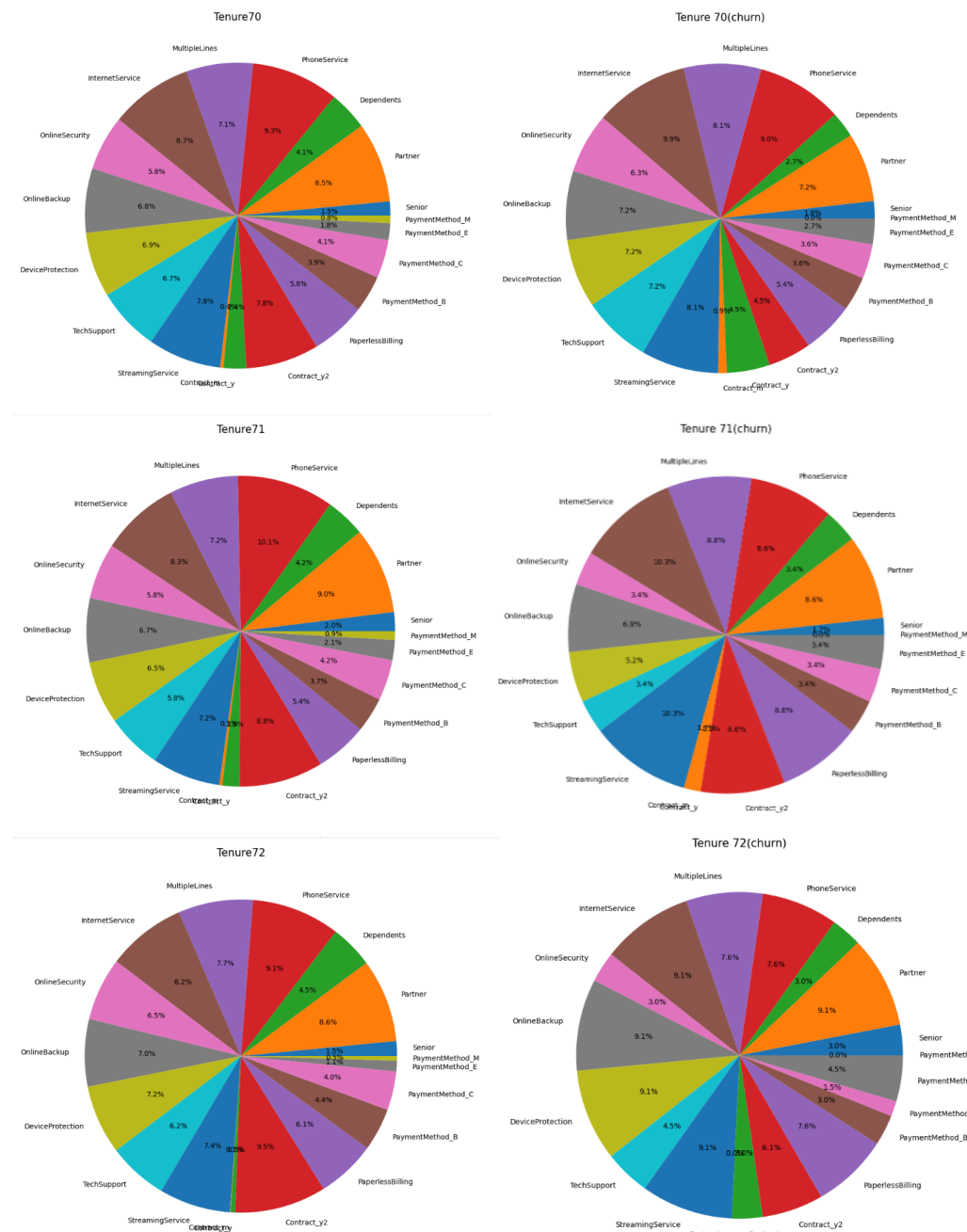
2년 계약 고객은 예상대로 계약기간(24개월 48개월) 혹은 그 다음달에 이탈할 것을 볼 수 있었고, 1년 계약고객은 첫 1년은 제외하곤 2년 계약고객과 동일하게 당월이나 다음 달에 대거 이탈했다.

따라서 계약기간이 다가올 수록 이탈율이 늘어나는 것을 알 수 있었고 계약기간을 연장할 수 있는 대책을 마련해야한다.

서비스 측면에선 이탈 고객과 비이탈 고객의 차이점이 없는 것으로 봐서 어느 정도 해당 통신사를 이용한 고객들에게는 서비스의 추가보단 요금 할인같은 방법을 사용하는 것이 좋아보인다.

4-3. 장기고객

장기고객의 확보는 매우 중요하다. 왜냐하면 신규고객 확보비용보다 훨씬 적은 비용이 소모되고, 해당 브랜드의 이미지 전파에도 적극적이고, 사회적인 증거로 사용되어 신규 고객의 전환을 늘리는데 도움이 될 수 있기 때문이다. 그럼 해당 통신사는 어떤 방법으로 장기고객을 유지하는지 알아봤다.



해당 개월 수에 따른 파이차트이다.

이탈자와 비이탈자들의 차트가 비슷한 것으로 보아 장기고객의 이탈이 통신사에서 제공하는 특정 서비스 때문이 아닌 다른 이유가 있을 것이라고 생각했다. 단 이탈자들이 적지않은 금액을 납부하고 있었다는 공통점은 있었다.

따라서 장기 고객은 마찬가지로 1. 요금할인을 해주던 2. 여론조사를 통해 불편함을 듣고 개선해주는 방법으로 잡을 수 있을 것이다.

=====

● 결론

1. 시니어의 CLTV(고객생애가치)는 다른 연령층에 비해 크지만 이탈율도 높다. 따라서 시니어 대상 마케팅이 중요하다.
2. 월마다 계약을 하는 것보다 1년이나 2년단위로 계약을 유도하는 것이 이탈을 방지하고 수익을 높일 수 있다.
3. 특히 통신사 이용 첫달에 이탈을 막는것이 향후 이탈고객을 줄일 수 있는 중요한 점이다.
4. 고객이 어느정도 안정권에 위치한 이후로는 서비스는 이탈에 큰 영향을 끼치지 않는다.

=====