

# 1. 分析背景与目的

---

## 1.1 分析背景

---

- 电信行业市场竞争异常激烈，据统计，每家电信运营商的年均客户流失率在15%至25%之间。客户流失率对企业经营而言是一项关键指标，因为客户留存成本远低于获客成本。关于客户留存有这样一个观点，如果将客户流失率降低5%，公司利润将提升25%-85%。
- 然而，大多数公司都有大量的客户，无法将客户留存精力放在每一个客户身上。如客户留存工作开展不精准，会使得留存成本过高，导致成效甚微甚至利润下降。随着市场饱和度的不断上升，电信运营商亟待解决增加客户黏性、延长客户生命周期的问题。
- 如果公司能够提前预测哪些客户会流失，就能更精准的开展留存工作，将重点放在这些“高风险”客户身上，有针对性的开展欲流失客户挽留工作。因此，电信客户流失分析与预测至关重要。

## 1.2 提出分析问题

---

**为提高客户留存率，探索该电信运营商的业务存在哪些可以改进的地方？**

针对分析的目的，提出以下三个方向：

1. 流失客户群体自身有哪些特征（性别/年龄/是否有家属等）？
2. 电信运营商提供的哪些服务不够完美，导致客户流失率较高？
3. 是否存在合同/付款方面的因素导致客户流失？

# 2. 分析维度

---

**根据以上设定的三个问题，提出来三个分析维度：**

将着重从电信运营商的**客户属性分析**、**服务属性分析**、**合同属性分析**三个方面进行分析，去探索和分析该电信运营商在产品和业务上有哪些可以改进的地方，并给出实际性的建议，以提升和改进客户留存率，提高电信运行商的营业收入和利润总额。

- 客户属性分析：客户的性别分布、年龄分布、是否有配偶以及是否有家属等对流失率的影响。

- 服务属性分析：客户订购电信服务时长、是否有电话服务、订购的互联网服务类型以及是否订购互联网附加服务等对流失率的影响。
- 合同属性分析：客户签署的合同期限、采取的支付方式、每月支付的金额以及总体支付的金额对流失率的影响。

## 3. 数据表理解与预处理

### 3.1 数据集描述

- 该数据集是 Kaggle 上的一个竞赛项目，名称为“Telco Customer Churn” (<http://www.kaggle.com/datasets/blastchar/telco-customer-churn>) 。
- 只有一张数据表：

WA\_Fn-UseC\_-Telco-Customer-Churn : 7043 \* 21

### 3.2 列名称理解

数据表名称：WA\_Fn-UseC\_-Telco-Customer-Churn

序号	字段名	数据类型	字段描述
1	customerID	String	顾客ID
2	gender	String	性别
3	SeniorCitizen	Integer	客户是否为老年人(1/0)
4	Partner	String	客户是否有配偶(是/否)
5	Dependents	String	客户是否有家属(是/否)
6	tenure	Integer	客户在公司停留的月数
7	PhoneService	String	客户是否有电话服务(是/否)
8	multipleLines	String	客户是否有多条线路(是/否/没有电话服务)
9	InternetService	String	客户的互联网服务提供商(DSL/光纤/否)
10	OnlineSecurity	String	客户是否具有在线安全性(是/否/没有互联网服务)

序号	字段名	数据类型	字段描述
11	OnlineBackup	String	客户是否有在线备份(是/否/没有互联网服务)
12	DeviceProtection	String	客户是否有设备保护(是/否/没有互联网服务)
13	TechSupport	String	客户是否有技术支持(是/否/没有互联网服务)
14	StreamingTV	String	客户是否有流媒体电视(是/否/没有互联网服务)
15	StreamingMovies	String	客户是否有流媒体电影(是/否/没有互联网服务)
16	Contract	String	客户的合同期限(每月/一年/两年)
17	PaperlessBilling	String	客户是否有无纸化账单(是/否)
18	PaymentMethod	String	客户的付款方式(电子支票/邮寄支票/银行转账(自动)/信用卡(自动))
19	MonthlyCharges	Integer	每月向客户收取的金额
20	TotalCharges	Integer	向客户收取的总金额
21	Churn	String	客户是否流失(是/否)

'Churn' 为目标值。同时，通过分析该数据集的各特征值，大致划分三个分析维度需用到的特征值：

- 客户属性分析：  
customerID, gender, SeniorCitizen , Partner, Dependents.
- 服务属性分析：  
PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies
- 合同属性分析：  
tenure, Contract, PaymentMethod, PaperlessBilling, MonthlyCharges, TotalCharges

### 3.3 数据预处理

采用 Python 语言对数据集进行数据预处理。预处理过程遵守“完全合一”规则：

1. 完整性：单条数据是否存在空值，统计的字段是否完善。
2. 全面性：观察某一列的全部数值，通过常识来判断该列是否有问题，比如：数据定义、单位标识、数据本身。
3. 合法性：数据的类型、内容、大小的合法性。比如数据中是否存在非ASCII字符，性别存在了未知，年龄超过了150等。
4. 唯一性：数据是否存在重复记录，因为数据通常来自不同渠道的汇总，重复的情况是常见的。行数据、列数据都需要是唯一的。

## 4. 客户属性分析

### 4.1 客户的性别、流失率特征分布

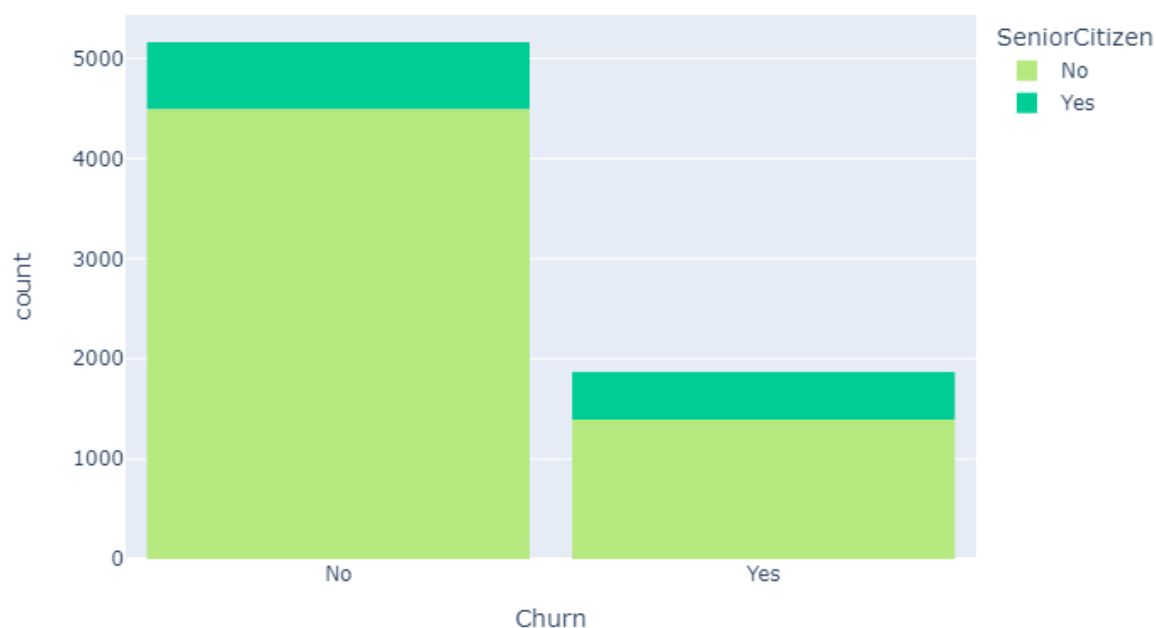
Gender and Churn Distributions



- 性别特征方面，有49.5%的客户为男性，有50.5%的客户为女性，男女占比差别很小。
- 流失率方面，有73.4%的客户未流失，26.6%的客户流失了。

### 4.2 客户年龄对流失率的影响

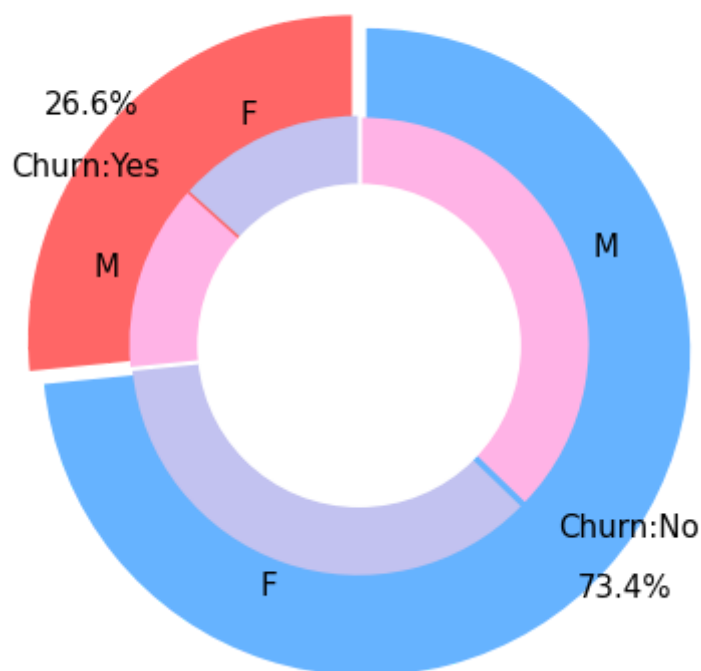
Chrun distribution w.r.t. Senior Citizen



- 老年客户数量占比很少。但老年用户的流失率高于非老年用户。

## 4.3 客户性别对流失率的影响

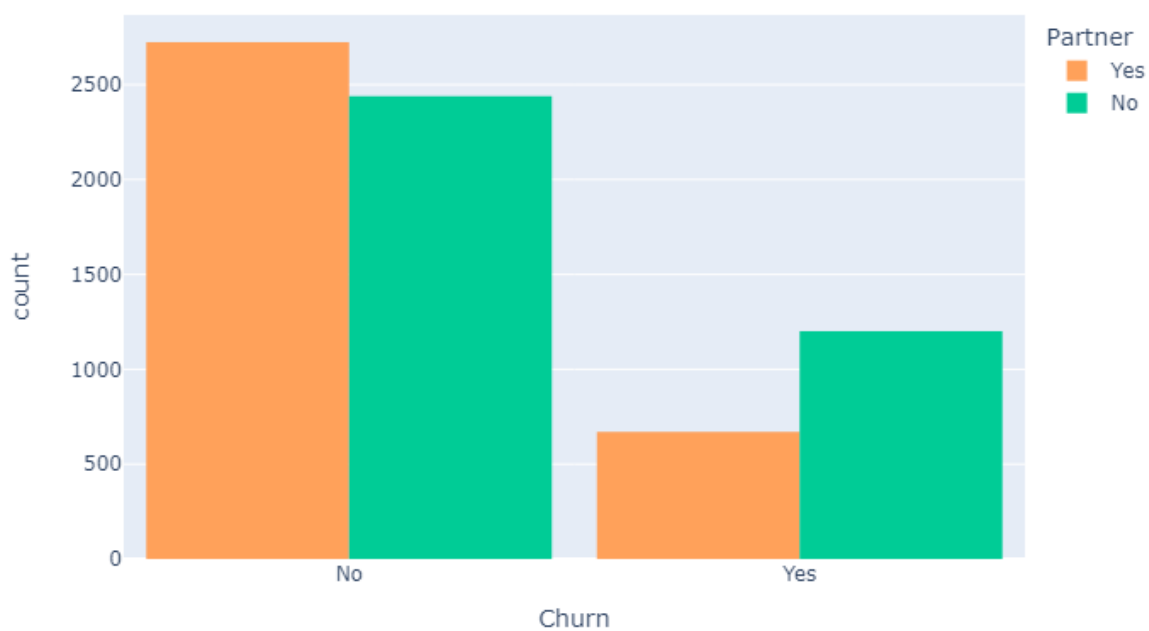
Churn Distribution w.r.t Gender: Male(M), Female(F)



- 客户流失与否与客户的性别基本无关。

## 4.4 客户有无配偶对流失率的影响

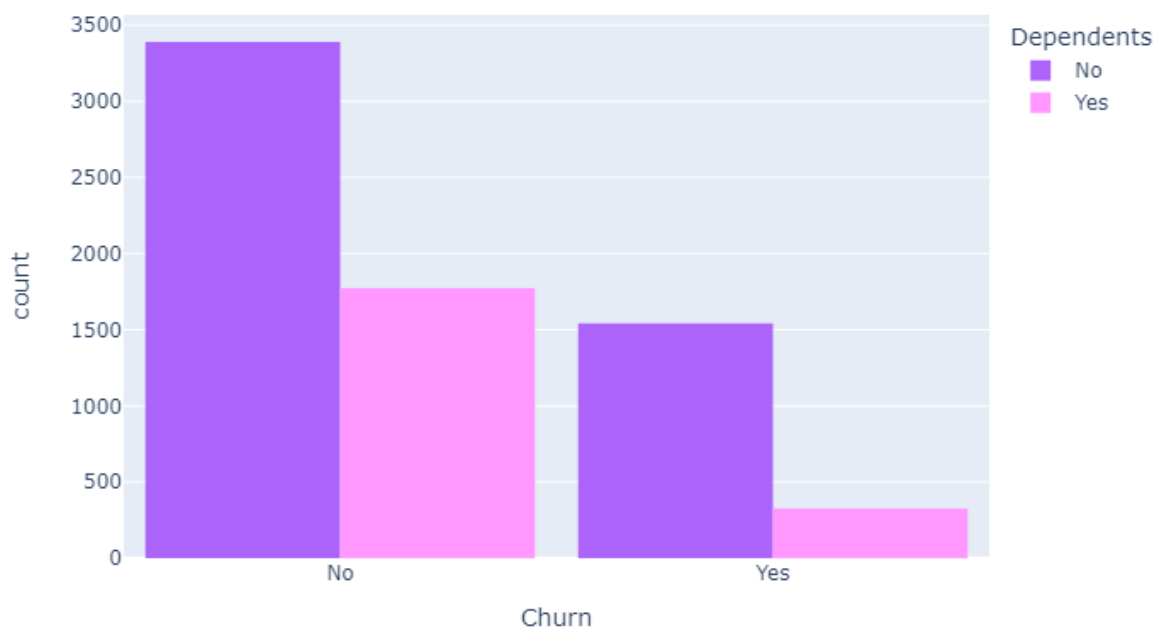
Churn distribution w.r.t. Partners



- 没有配偶的客户更容易流失。

## 4.5 客户有无家属对流失率的影响

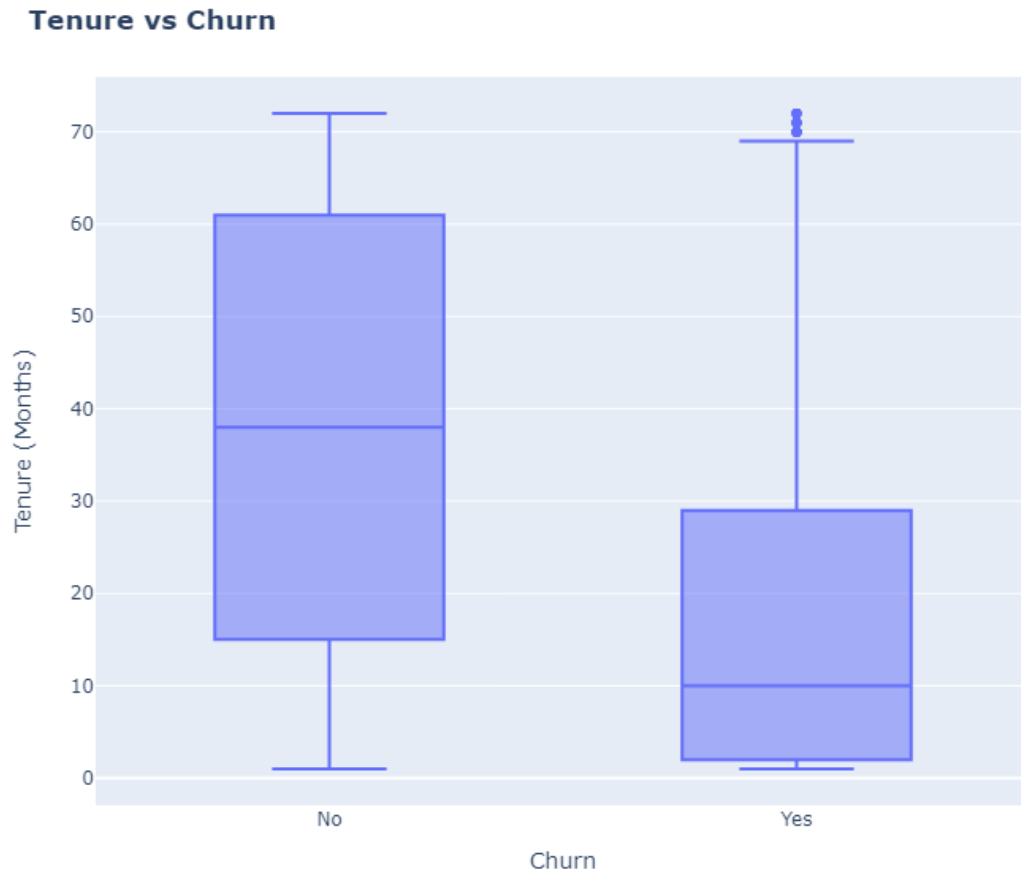
Dependents distribution



- 有家属的客户少。没有家属的客户更容易流失。

## 5. 服务属性分析

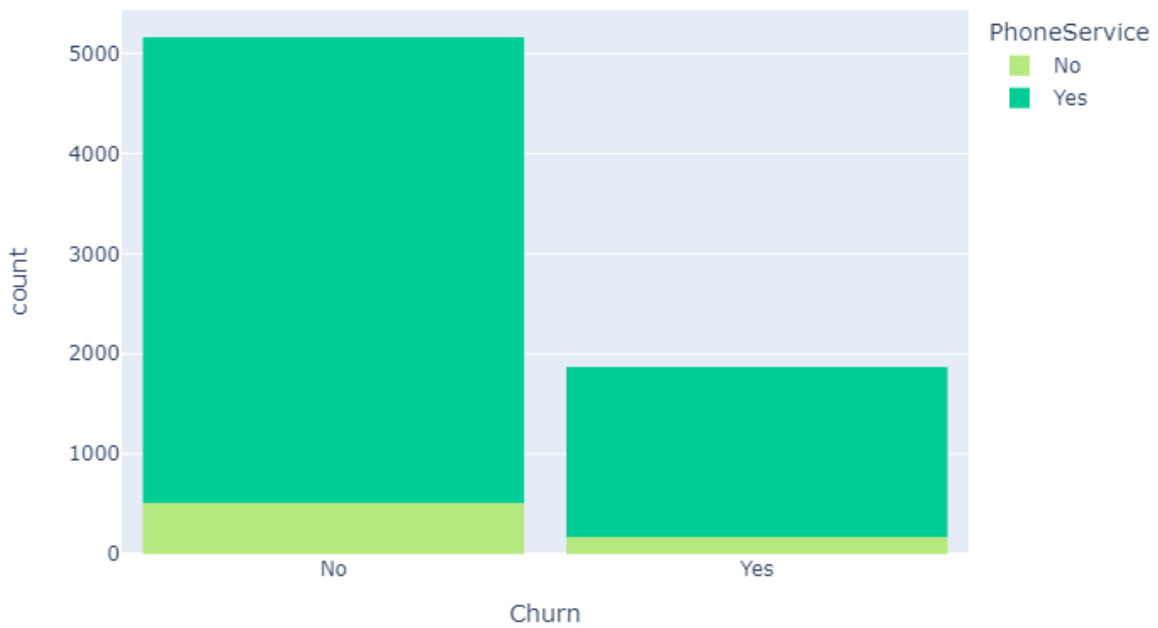
### 5.1 客户订购电信服务时长对流失率的影响



- 新客户更容易流失。

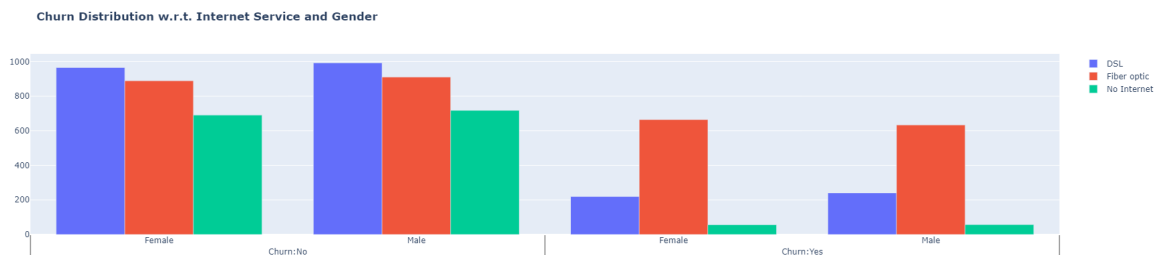
### 5.2 是否拥有电话服务对流失率的影响

### Churn distribution w.r.t. Phone Service



- 极少数客户没有电话服务。有电话服务的客户更有可能流失。

## 5.3 订购的互联网服务类型对流失率的影响

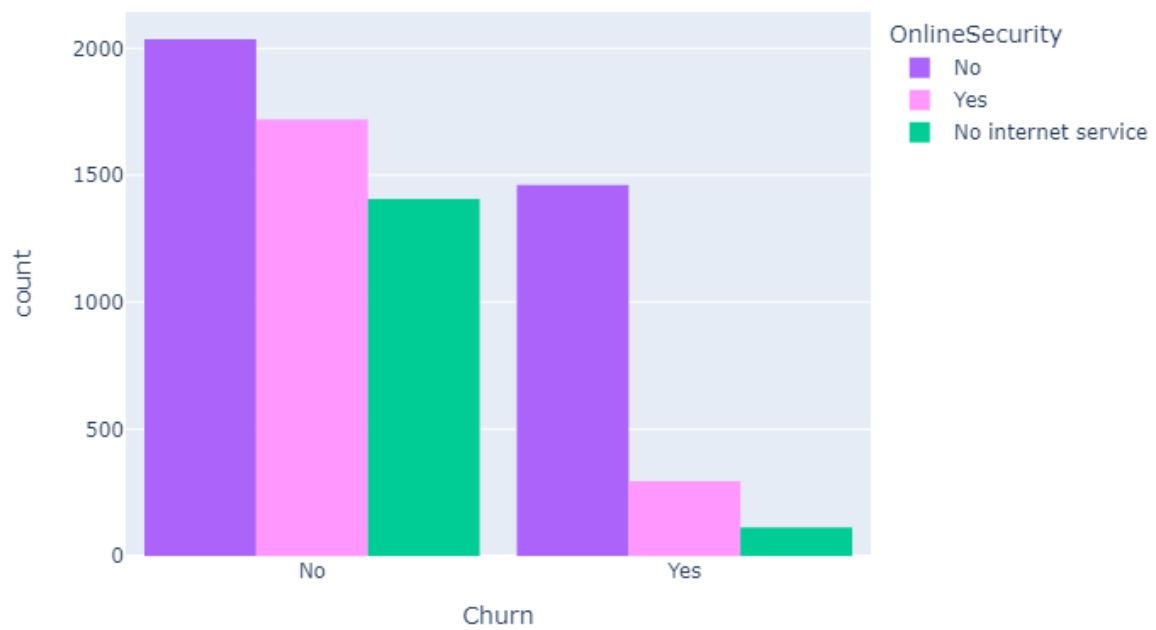


- 很多客户选择光纤服务，但使用光纤的客户流失率却很高，表明他们可能对该类互联网服务不满意。
- 使用 DSL 服务的客户也很多，与光纤服务相比，该群体的客户流失率更低。

## 5.4 是否订购网络安全服务对流失率的影响



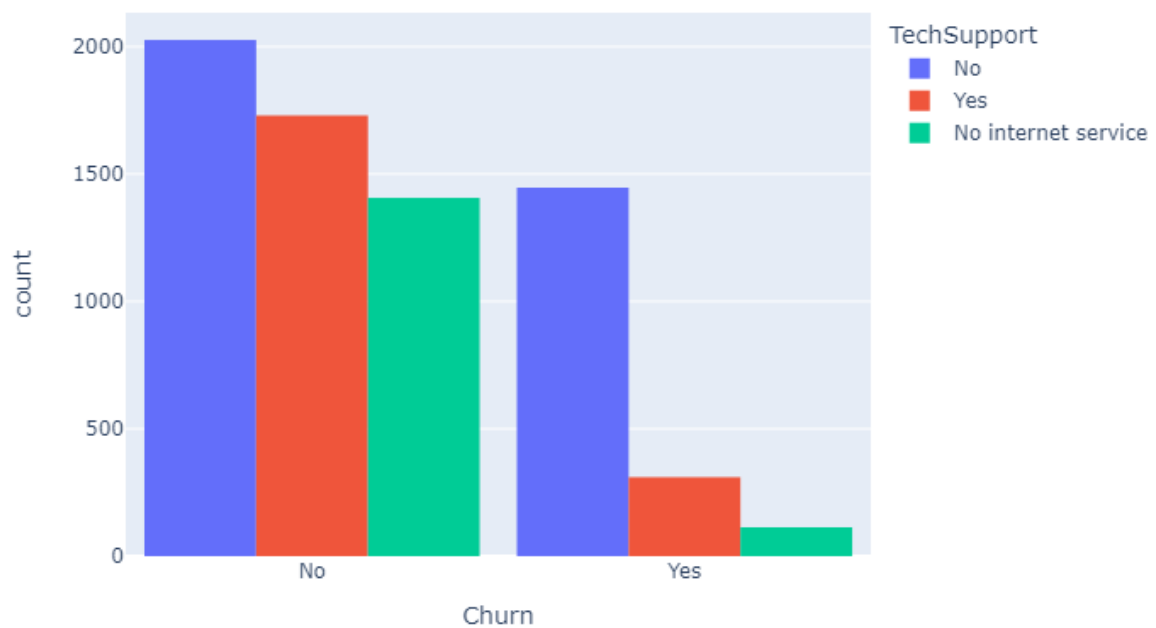
Churn w.r.t Online Security



- 未订购网络安全服务的用户群体流失率较高。

## 5.5 是否订购技术支持服务对流失率的影响

Churn distribution w.r.t. TechSupport

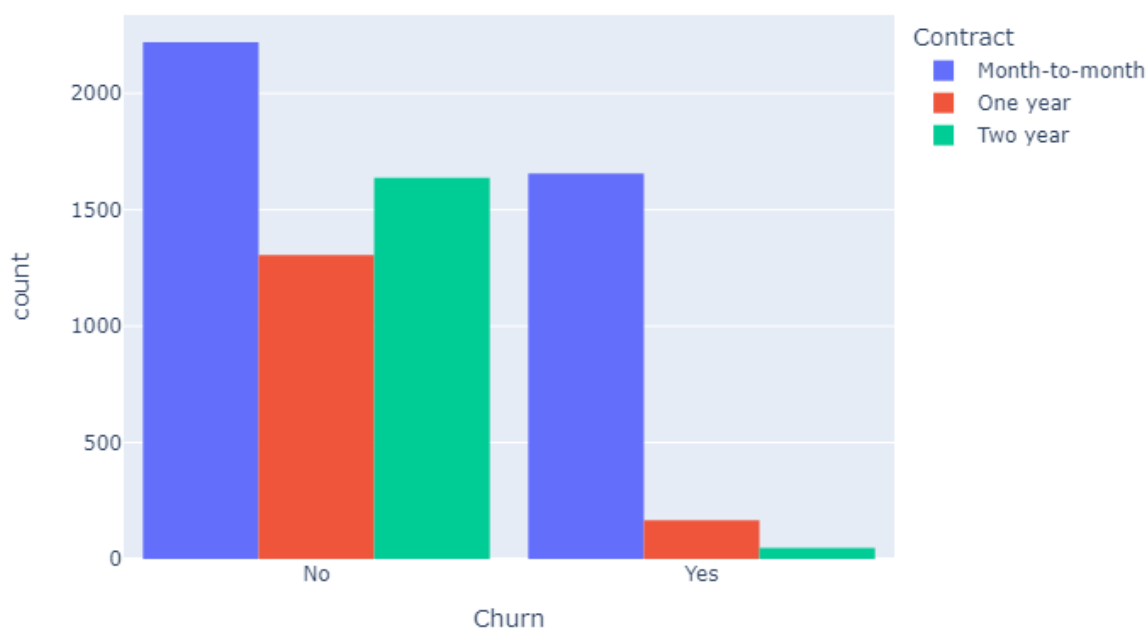


- 未订购技术支持服务的客户更容易流失。

## 6. 合同属性分析

### 6.1 合同签订期限对流失率的影响

Customer contract distribution



- 约55%用户按月签署合同，属于占比最大的类型。
- 按月签署合同的用户流失率也是最高的：
  - 按月签署合同的用户，流失率约为 43%。
  - 按年签署合同的用户，流失率约为11%。
  - 按两年为频率签署合同的用户，流失率约为3%。

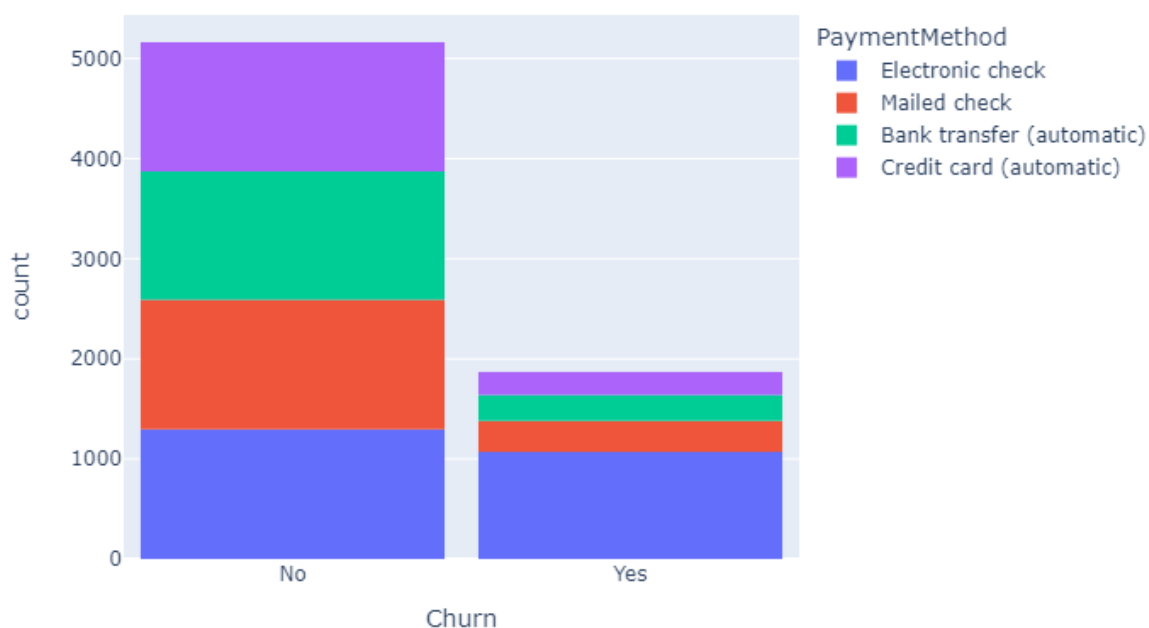
### 6.2 支付方式对流失率的影响

Payment Method Distribution



- 33.6%的用户使用电子账单支付费用；22.8%的用户采用邮寄支票的方式支付费用；21.9%的用户通过银行自动划款方式支付费用；21.6%的用户通过信用卡自动划款方式支付费用。

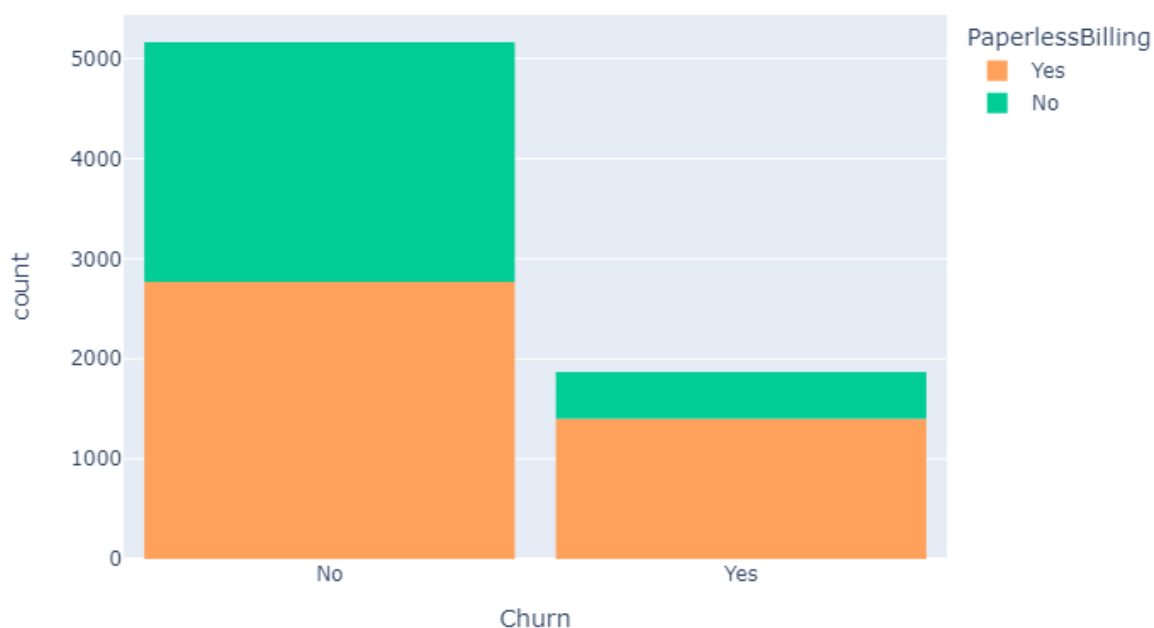
**Customer Payment Method distribution w.r.t. Churn**



- 使用电子支票作为支付方式的用户群体，流失率最高。可能该种支付方式体验一般。
- 选择信用卡自动扣款、银行自动扣款或邮寄支票作为支付方式的客户群体流失率较小。

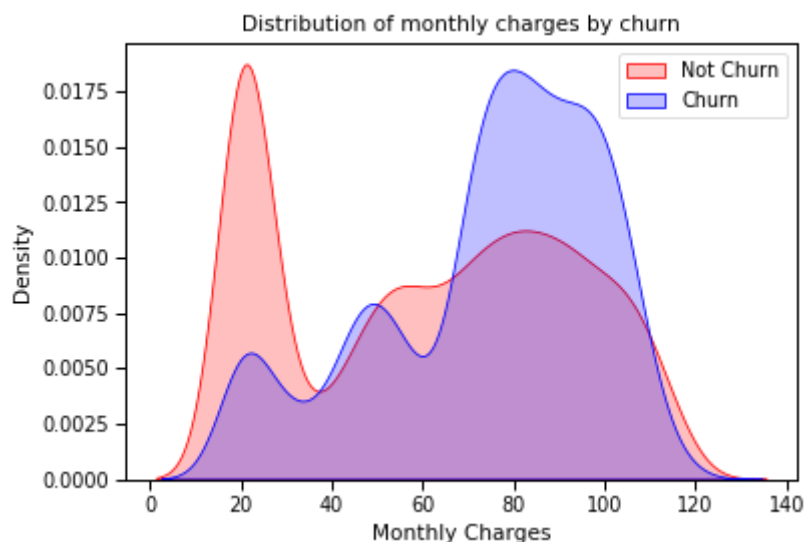
## 6.3 是否使用电子账单对流失率的影响

**Chrun distribution w.r.t. Paperless Billing**



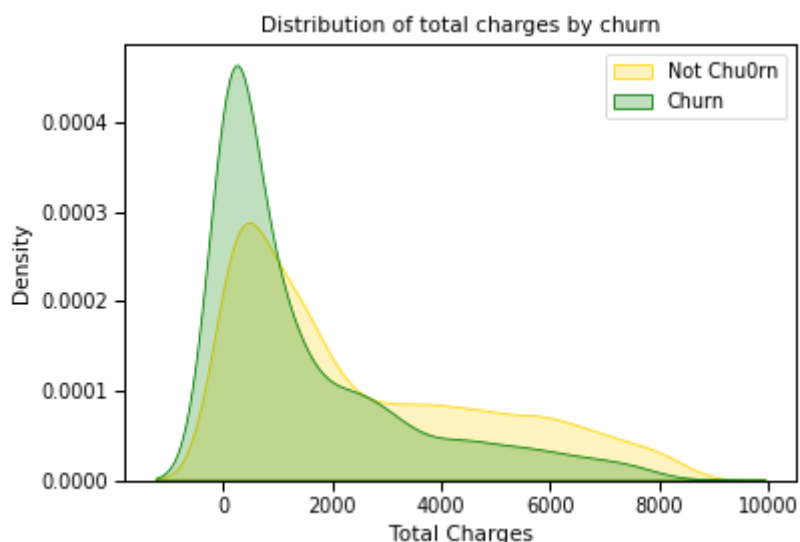
- 使用电子账单服务的客户更容易流失。

## 6.4 月消费金额对流失率的影响



- 月消费金额越高的客户，越容易流失。
- 月消费金额在 70至110 之间的客户流失的最多。

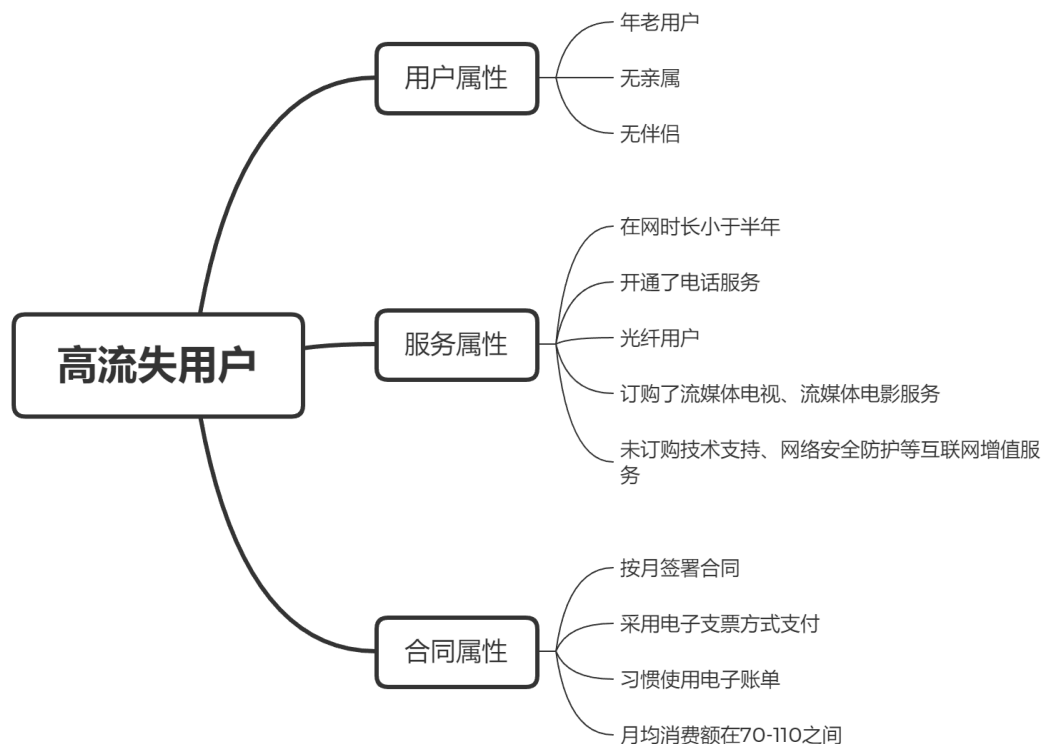
## 6.5 总消费金额对流失率的影响



- 长期来看，客户总消费越高，流失率越低，符合一般经验。

# 7. 分析结论及建议

## 7.1 分析结论



## 7.2 留存建议

针对上述结论，从业务角度给出相应建议：

一是根据预测模型，构建一个高流失率的客户列表。通过客户调研推出一个最小可行化产品功能，并邀请种子客户进行试用。

二是对电信运营商的业务、服务进行针对性改进。

### 1. 客户方面：

- 针对老年客户，无亲属、无伴侣客户群体推出亲属套餐、温暖套餐等定制服务，一方面加强与其他客户的关联度，另一方对特定客户提供个性化服务。

### 2. 服务方面：

- 针对新注册客户，推送半年优惠如赠送消费券，以渡过客户流失高峰期。
- 针对光纤客户和附加流媒体电视、电影服务客户，重点在于提升网络体验、增值服务体验。一方面推动技术部门提升网络指标，另一方面对客户承诺免费网络升级和赠送电视、电影等包月服务以提升客户黏性。
- 针对在线安全、在线备份、设备保护、技术支持等增值服务，应重点对客户进行推广介绍，如首月或半年可免费体验。

### 3. 合同方面：

- 针对单月合同客户，建议推出年合同付费折扣活动，将月合同客户转化为年合同客户，提高客户在网时长，以达到更高的客户留存。
- 针对采用电子支票支付客户，建议定向推送其它支付方式的优惠券，引导客户改变支付方式。