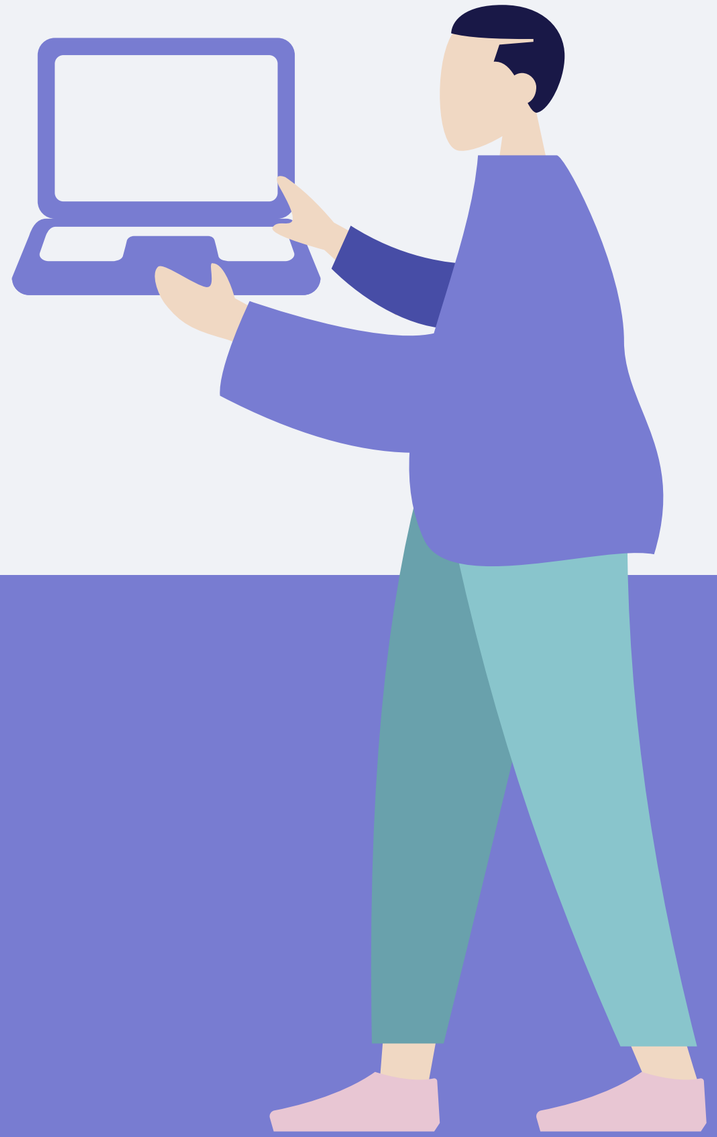
Internet Service Provider Customer Churn Analysis

FSDA • GROUP FINAL PROJECT • TEAM J

Project Objective



There is a big competition between Internet providers. If a providers want to increase its revenue they needs more subscriber but keep existing customer is more important than having new ones. So providers want to know which customer will likely cancel his service, **we call this as churn**. If the know who will go, maybe they can catch them with promotions.

Methodology

1 Preparation Phase

- Data Preparation
- Root Cause
- Problem Statement
- Data Cleaning

2 Analysis Phase

- Exploratory Data Analysis (EDA)
- Data Modeling
- Data Manipulation

3 Insight and Recommendation

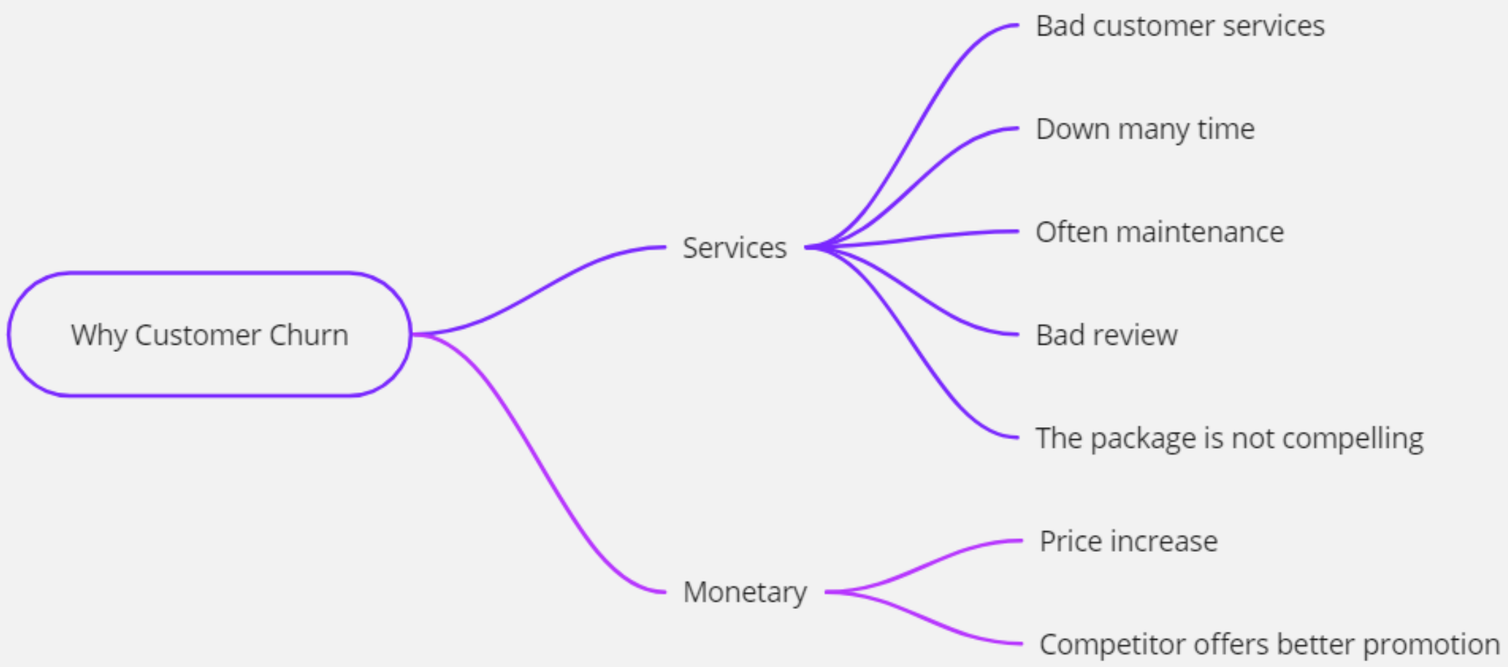
- Insight
- Recommendation

Dataset Overview

72K	Users
59K	Tv Subscriber
40K	Churn
24K	Movie Subscriber

The dataset is **Internet Service Provider Customer Churn** taken from
https://www.kaggle.com/mehmetsabrikunt/internet-service-churn/metadata
Consist of **72275 rows** and **11 columns**.

Root Cause


Why Customer Churn
Services
Bad customer services
Down many time
Often maintenance
Bad review
The package is not compelling
Monetary
Price increase
Competitor offers better promotion

Problem And Goals

Problem

This company **have a 58% churn rate** on this 3-month dataset which is a bad thing for internet service provider company that have services based on subscription

Goals

How to **reduce the churn rate by 5%** in the next quarter?

Data Cleaning

Before Outlier

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 71893 entries, 0 to 72273
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   id                            71893 non-null  int64
 1   is_tv_subscriber              71893 non-null  int64
 2   is_movie_package_subscriber   71893 non-null  int64
 3   bill_avg                      71893 non-null  int64
 4   service_failure_count         71893 non-null  int64
 5   download_avg                  71893 non-null  float64
 6   upload_avg                    71893 non-null  float64
 7   download_over_limit           71893 non-null  int64
 8   churn                         71893 non-null  int64
 9   subcription_days              71893 non-null  int64
 10  remaining_contract            71893 non-null  int64
dtypes: float64(2), int64(9)
memory usage: 6.6 MB
```

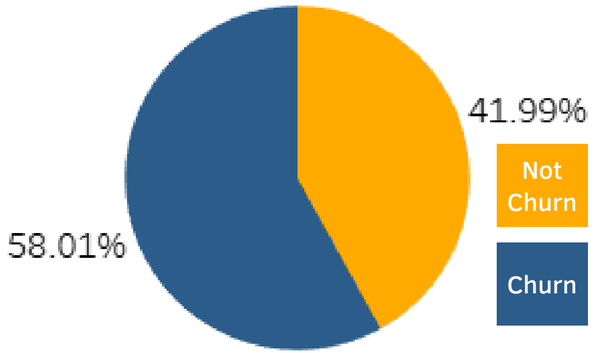
After Outlier

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 62496 entries, 0 to 72273
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   id                            62496 non-null  int64
 1   is_tv_subscriber              62496 non-null  int64
 2   is_movie_package_subscriber   62496 non-null  int64
 3   bill_avg                      62496 non-null  int64
 4   service_failure_count         62496 non-null  int64
 5   download_avg                  62496 non-null  float64
 6   upload_avg                    62496 non-null  float64
 7   download_over_limit           62496 non-null  int64
 8   churn                         62496 non-null  int64
 9   subcription_days              62496 non-null  int64
 10  remaining_contract            62496 non-null  int64
dtypes: float64(2), int64(9)
memory usage: 5.7 MB
```

In data cleaning, we handle missing values, remove duplicates, change data types and exclude the outliers on bill_avg, download_avg, and upload_avg

Churn Rate and Subscriber

Churn Rate



- After data cleaning, number of customer reduce to 62.496 user while churn rate is 58.01%
- Internet with TV Package have highest number of user while have highest number of churn
- Internet with Movie and TV Package proven to have lowest churn rate

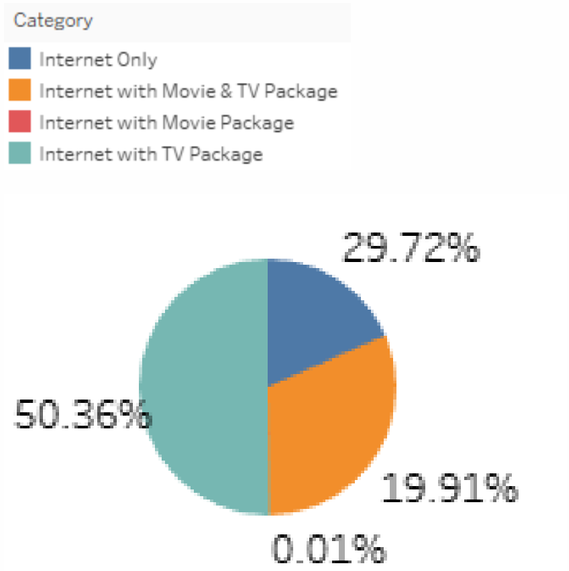
Number of Customer

62,496

Subscription	Number of Subscriber	Number of Churn	Churn/ Subscription
Internet Only	11.704	11.704	100%
Internet with Movie Package	2	2	100%
Internet with TV Package	30.932	18.260	59%
Internet with Movie and TV Package	19.857	7.220	36,35%

Average Bill With Churn

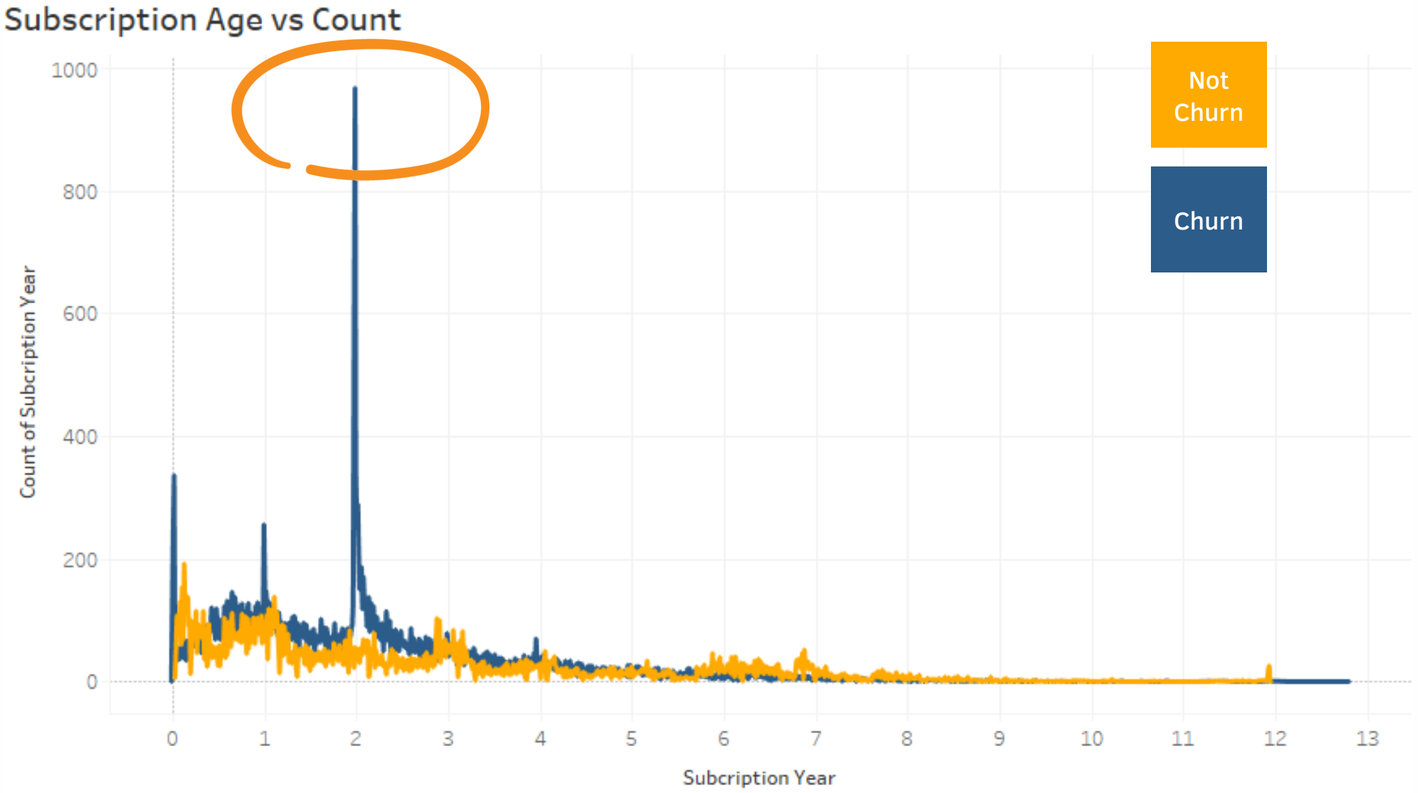
Subscription	Total Average Bills	Total Churn	Churn/ Subscription
Internet Only	195.239	10.774	100%
Internet with Movie Package	28	2	100%
Internet with TV Package	532.567	18.260	59%
Internet with Movie and TV Package	332.825	7.220	36,35%



Findings

The customers who only have tv subscription have a higher average bills with the highest churn rate. Internet with TV Package generates highest revenue

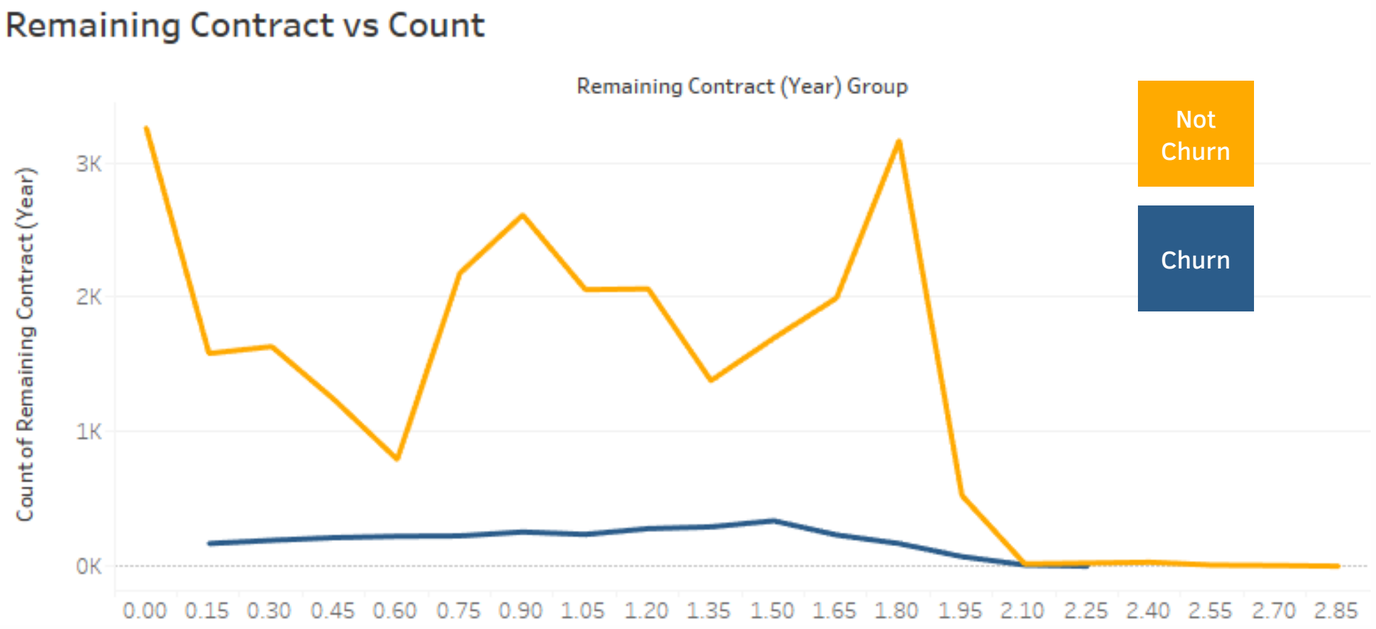
Subscription Age



Findings

users mostly have 0-4 years subscription age but they tend to churn mostly on the 2nd year

Remaining Contract with Churn



Findings

- Customers who have remaining contract less likely to churn.
- The longest remaining contract is 2 years.

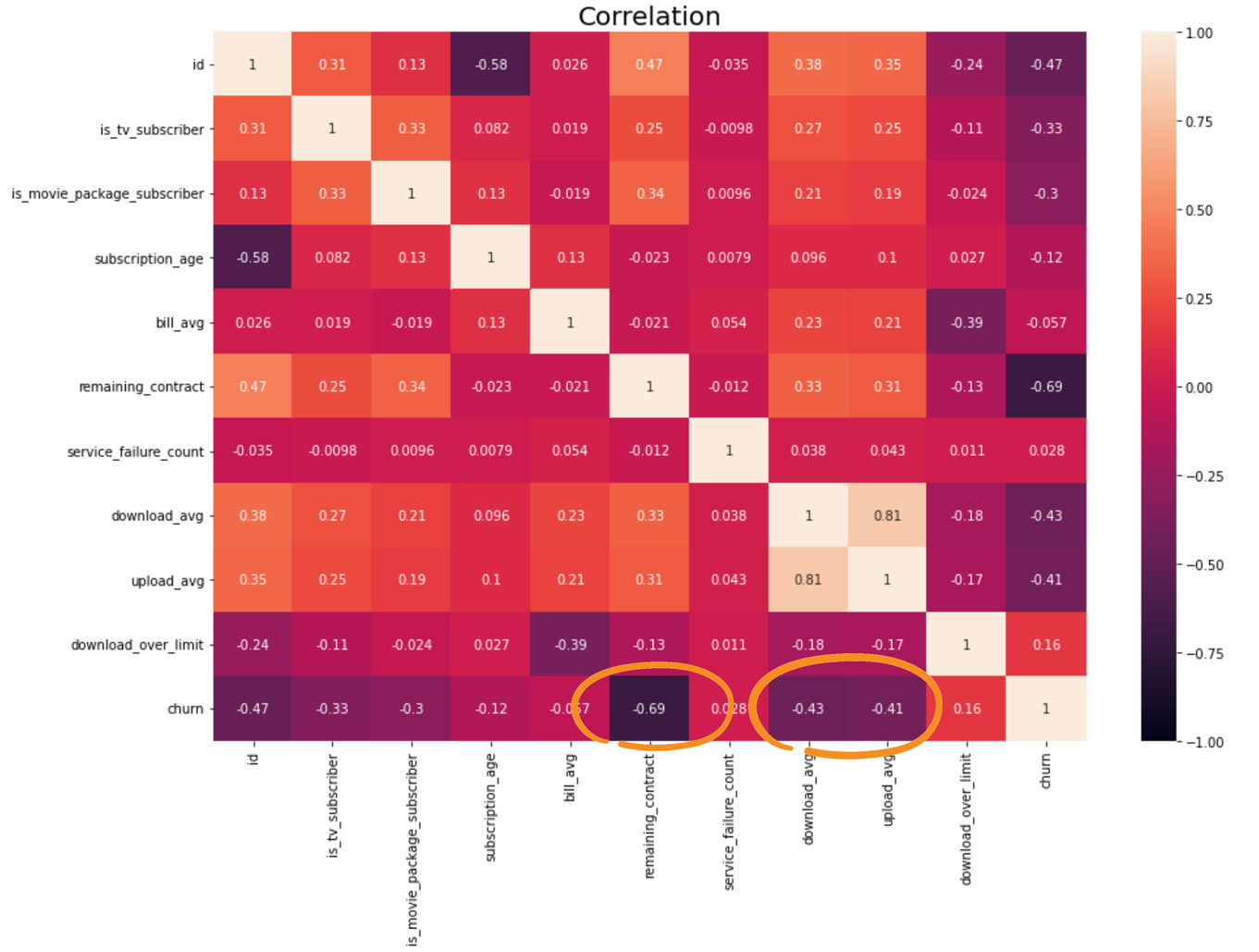
Download & Upload Average

Subscription	Download Average	Upload Average	Churn
Internet Only	163.890	13.406,4	10.774
Internet with Movie Package	83,2	5,7	2
Internet with TV Package	972.540	74.966,9	18.260
Internet with Movie and TV Package	804.493,4	60.080,2	7.220

Findings

The customers with Internet and TV Package has the highest download average

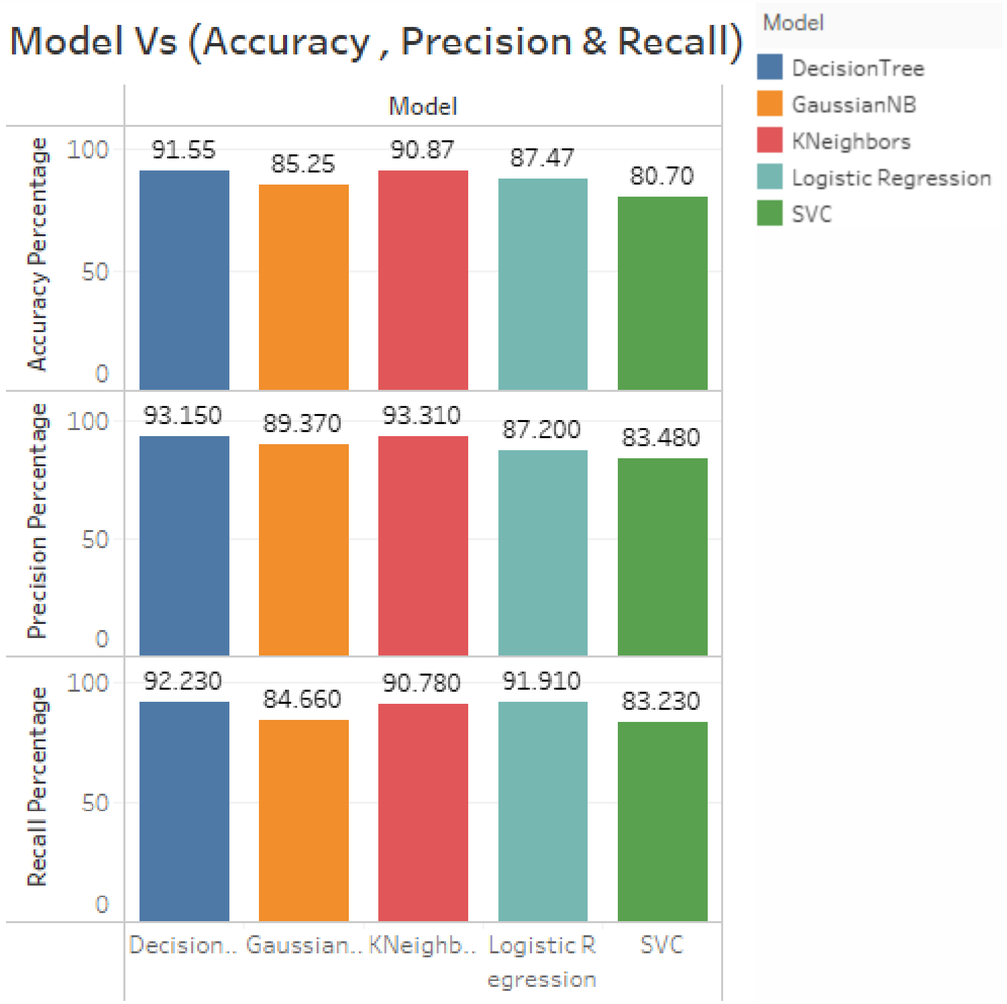
Correlation



Findings

- The remaining contract is highly affecting churn by -0,69 when download average and upload average is mildly affecting churn

Model Selection

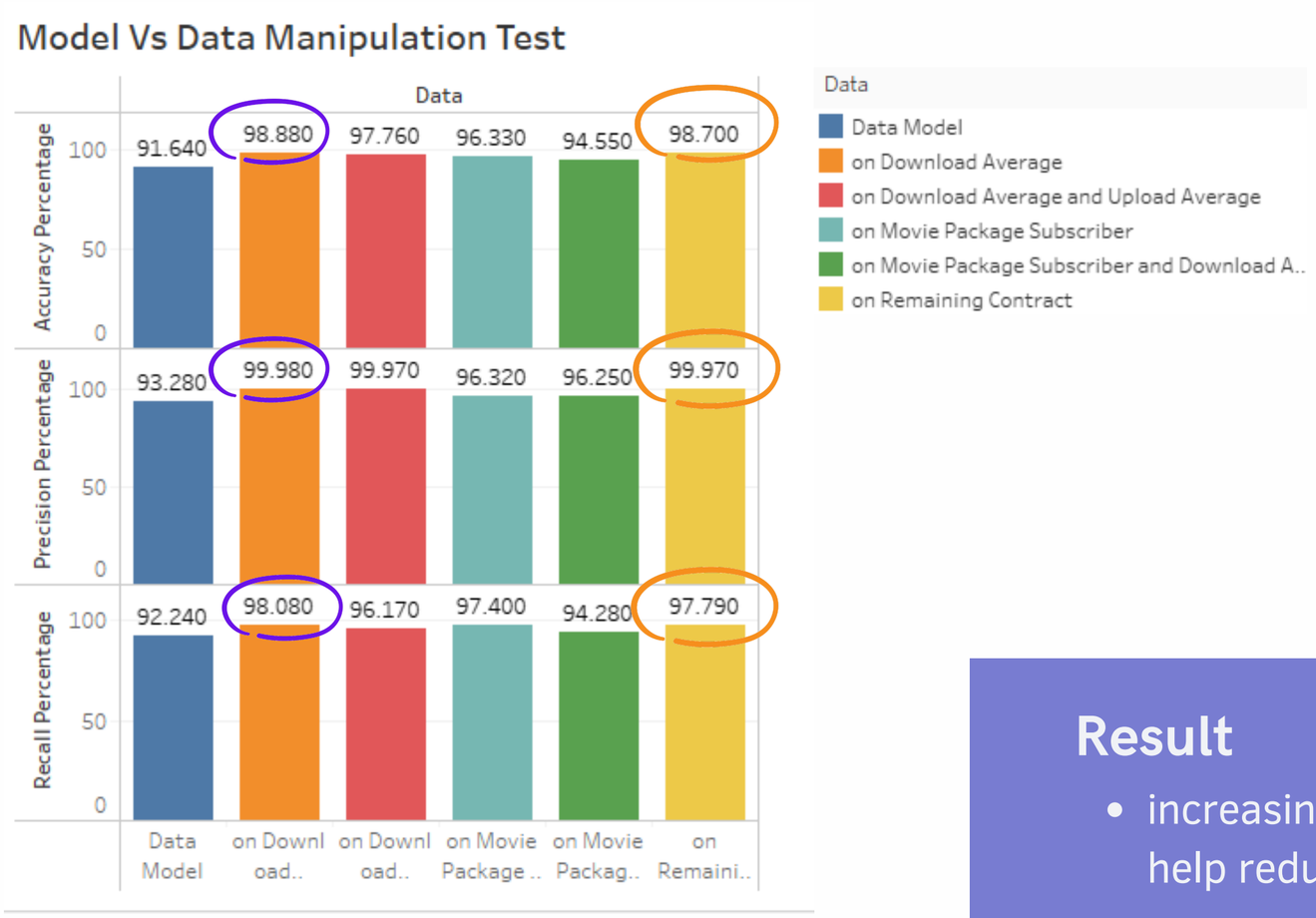


Findings

Decision Tree, Logistic Regression, and K-Nearest Neighbor shows highest number in term of accuracy, precision and recall.

Overall the best model based on accuracy, precision and recall percentage is decision tree modeling

Manipulation Data Test



Manipulation

- On Download Average: Replace zero values with "mean" value
- On Upload Average: Replace zero values with "mean" value
- On Movie Package Subscriber: Switch customer who didn't subscribe to subscribers
- On Remaining Contract: Replace zero values to "mean" value

Result

- increasing remaining contract and average download will help reducing churn rate more than 7%

Insight

- After data cleaning, number of customer reduce to 62.496 user while churn rate is 58.01%
- Internet with TV Package have highest number of user, generates highest revenue while have highest churn rate

- Remaining contract highly affecting churn while download average and upload average mildly affecting churn
- From the result of data manipulation, the remaining contract and average download have highest % of accuracy, precision, and recall
- With Decision Tree we found high possibility to reduce churn by increasing remaining contract and average download

Recommendation

- Create a loyalty-based campaign to increase remaining contract, download average, and upload average to reduce churn
- Create promotion for up-selling programs to increase the number of the Internet with Movie and TV Package that proven to have the lowest churn rate

- Create a promotion that target an audience segmentation "user with subscription age 1-2 years" because subscription age around 2 years more likely to churn
- Increase the number of remaining contracts by creating a campaign for converting "non-contract users" to "users with contract"

Appendix

- https://colab.research.google.com/drive/16QTRNYjD-gQLjnXE-iZ5l1RD4VRxBMCi?usp=sharing
- https://public.tableau.com/views/TeamJ-GroupFinalProject-RevoUBatchMayFixA/Dashboard2?:language=en-US&:display_count=n&:origin=viz_share_link
- https://www.canva.com/design/DAFJOlVjmxk/JiJd886RnMKXvvptv49CZA/edit?utm_content=DAFJOlVjmxk&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Meet Our Team

Praya

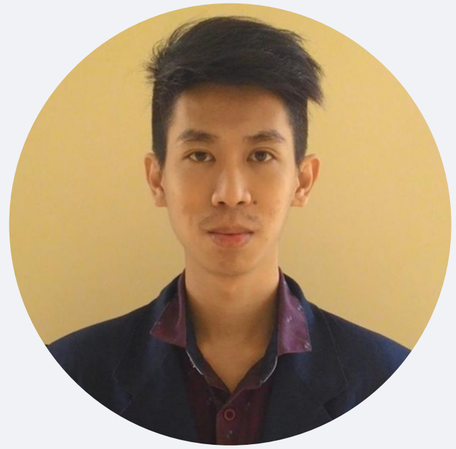
Project Manager
AKA Tim Hore

Bintang

Python Guru

Guntur

Tableu Shifu



Vito

All Rounder