Internet Service Provider Churn Prediction Model

James Oomens
DSB-122

Context

01

Intro

Problem Statement

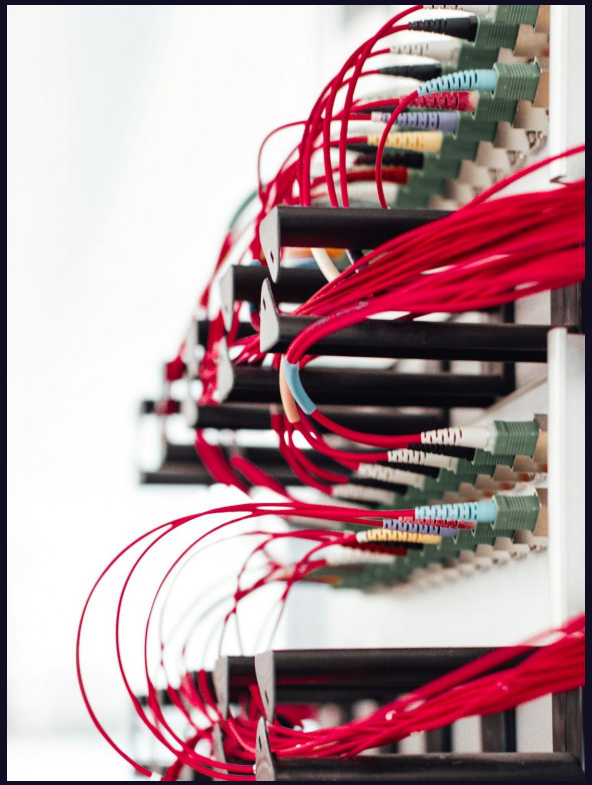
One of the biggest problems internet service providers face is retaining long-term customers as there is so much competition right now. Finding a way to retain these customers can be difficult. Some customers opt to pay month-by-month, some opt for contracts. The lying problem is finding the perpetrators as to why these customers give up their service after a short period of time. The goal is to find what these factors are. Though it may be easy to assume customers may churn due to internet speeds or lacking products, we need to know based on their actual accounts, the factors that cause this. Having a model to automate this will save much time in analyzing the data.

Objective

- Identify customer profiles that are likely to churn
- Find a way to reduce overall churn rate

Models to be tested:

- DecisionTree
- RandomForests
- Adaboost
- GradientBoost
- Logistic regression
- KNN



71,893 rows, 14 columns

ISP Dataset used for analysis:
https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn

Data Cleaning 02

Nulls

- **30% null**

"Remaining Contract" column

- **0.5% null**

"Download Avg" column

- **0.5% null**

"Upload Avg" column

Feature Engineering

All “NaN” in this column are customers who do not have a contract, therefore they have been replaced with 0s.

Then created an “is_contract” column to separate customer with or without contracts.

	reamining_contract
0	0.14
1	NaN
2	0.00
3	NaN
4	NaN
...	...
72269	1.25
72270	1.63
72271	2.19
72272	0.72
72273	0.82

72274 rows × 1 columns

	remaining_contract
0	0.14
1	0.00
2	0.00
3	0.00
4	0.00
...	...
72269	1.25
72270	1.63
72271	2.19
72272	0.72
72273	0.82

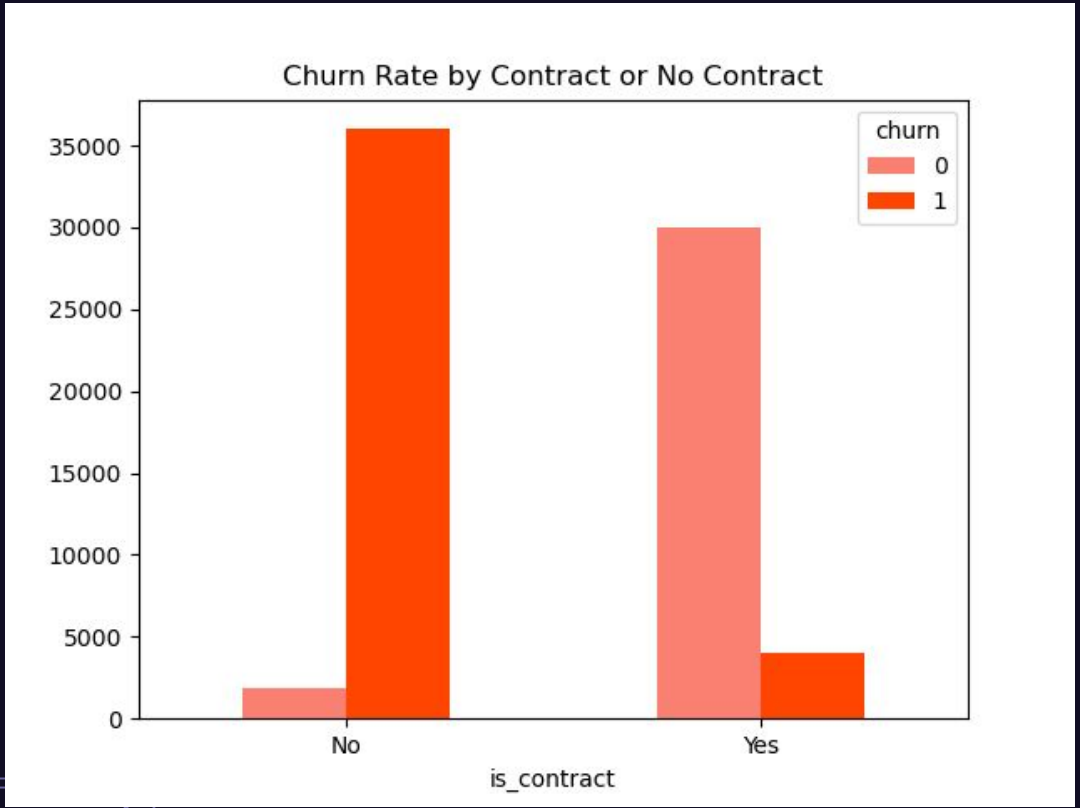
71893 rows × 1 columns

	is_contract
0	1
1	0
2	0
3	0
4	0
...	...
72269	1
72270	1
72271	1
72272	1
72273	1

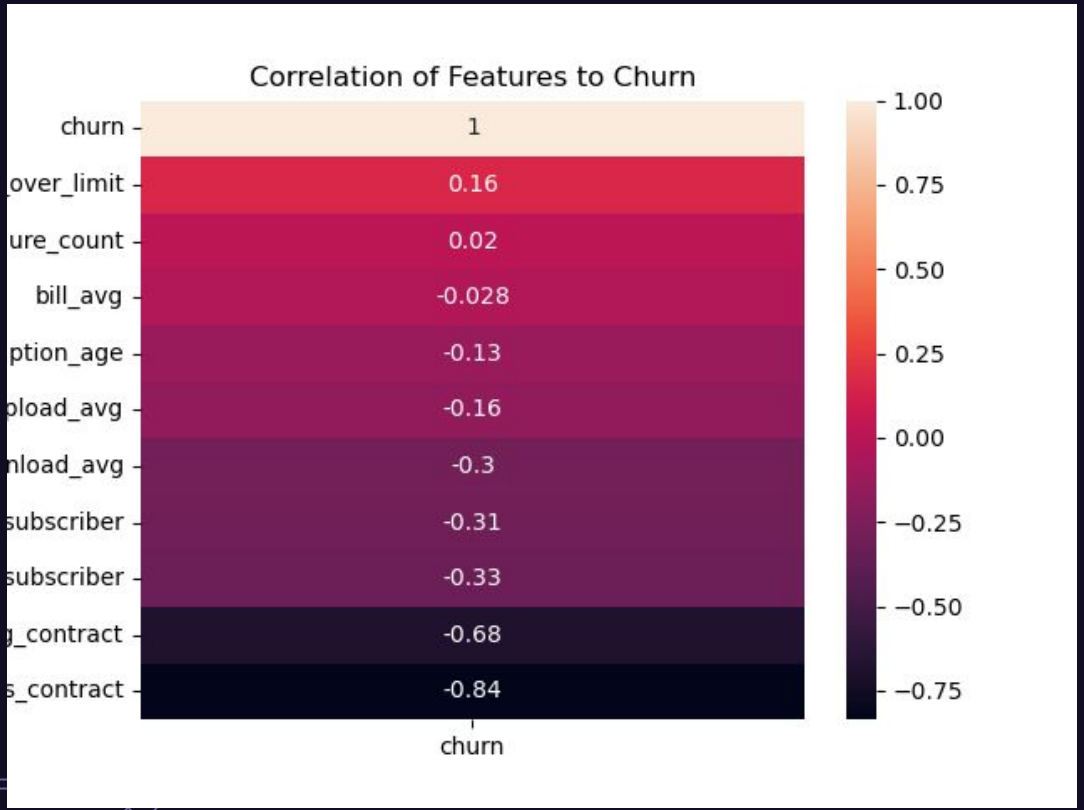
71893 rows × 1 columns

EDA 03

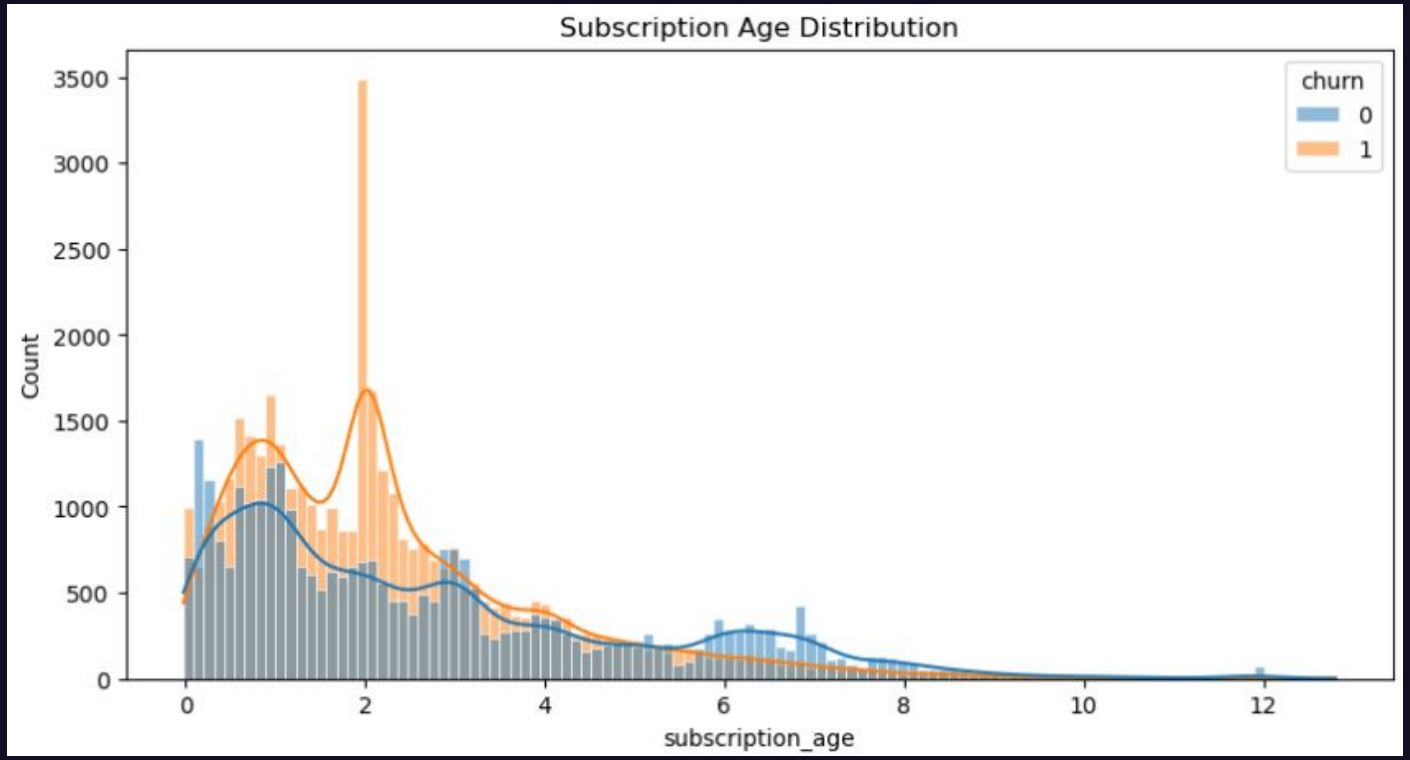
Customer Contracts



Mostly Negative Correlation of Features



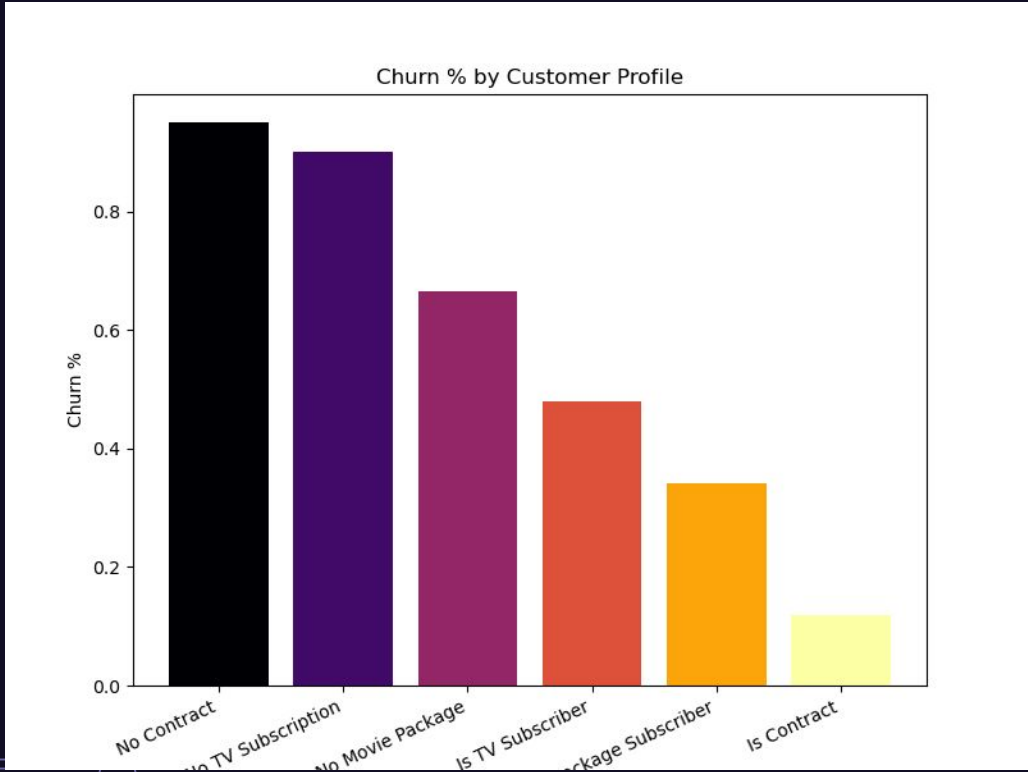
Subscription Age vs Churn



Churn % by Customer Account

CUSTOMER ACCOUNT	CHURN %
No Contract	95%
No Tv Subscription	90%
No Movie Package	66%
Is TV Subscriber	47%
Is Movie Package Subscriber	34%
Is Contract	11%

Another look at churn %

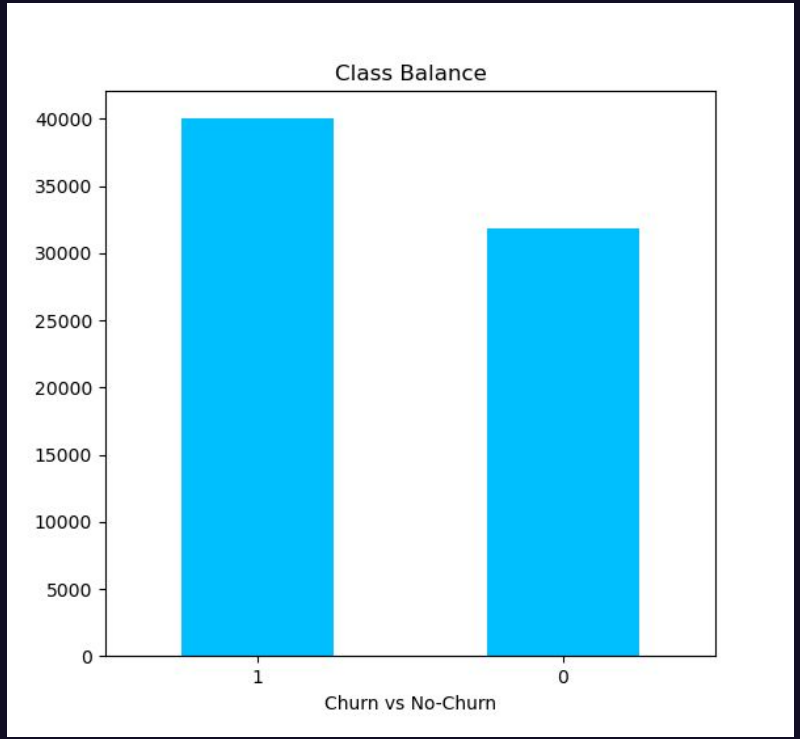


Modeling 04

Baseline Score & Metrics

Classes are balanced with **55%** churn rate. Oversampling was not needed.

To measure success, the goal is to create a model as accurate as possible, eliminating false negatives since expecting a customer to churn and not actually churning can help with budgeting and strategic business planning for the future.



Model of Choice: GradientBoostClassifier!

Initial Model Score

Train: 93.93%
Test: 93.77%

Score with grid searching

Train: 95.74%
Test: 94.13%

Best Parameters

learning_rate: 0.1
n_estimators: 150
max_depth: 7
max_features: log2

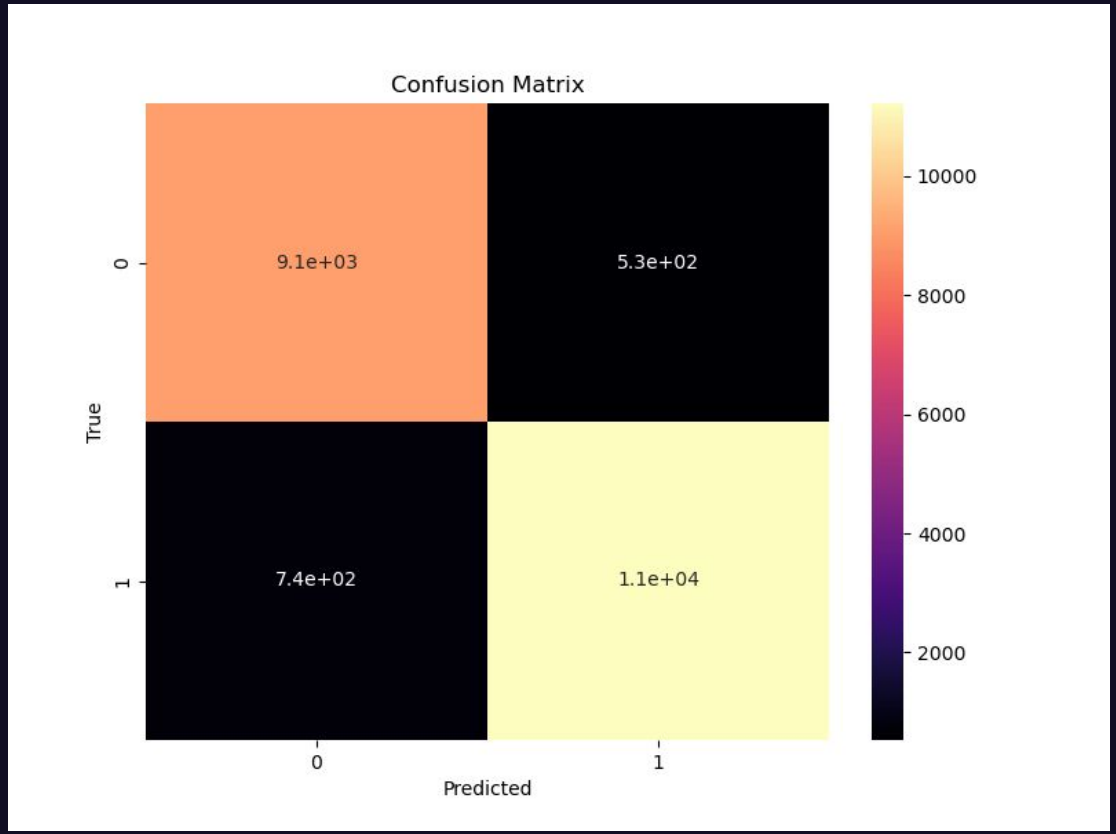
Feature Contributions

- Adding “is_contract” feature improved model by 1%
- Gridsearching improved score by about 1%

Inference 05

Inference

Model resulted in predicting 530 false negatives which is only 0.02% of all predictions.



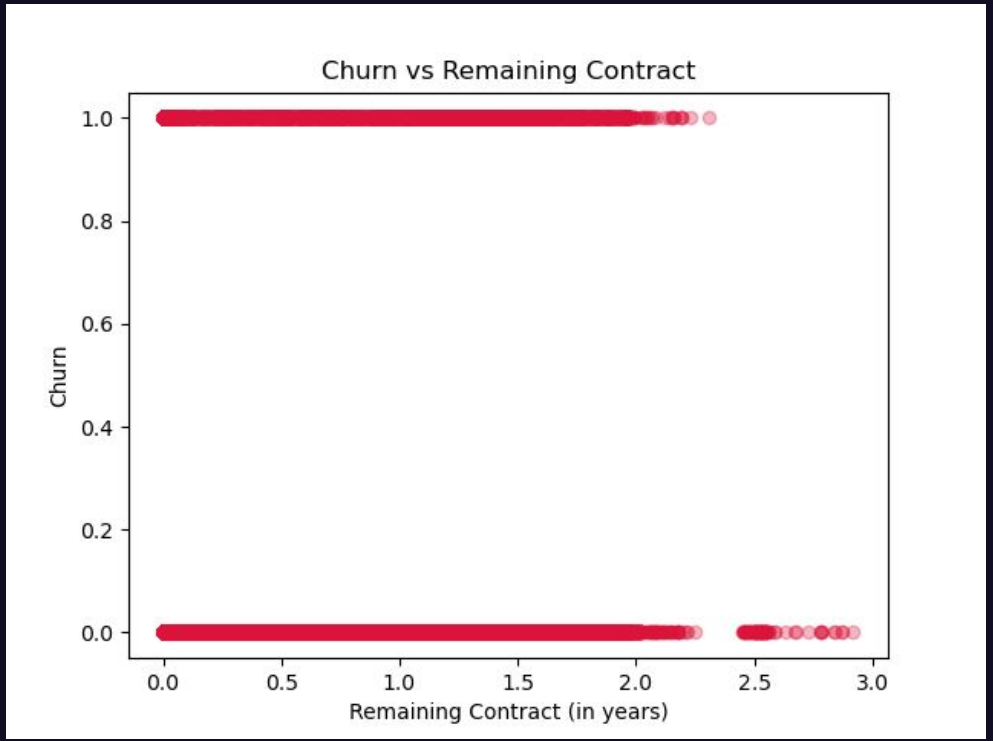
Conclusion & Next Steps

06

Business Recommendations

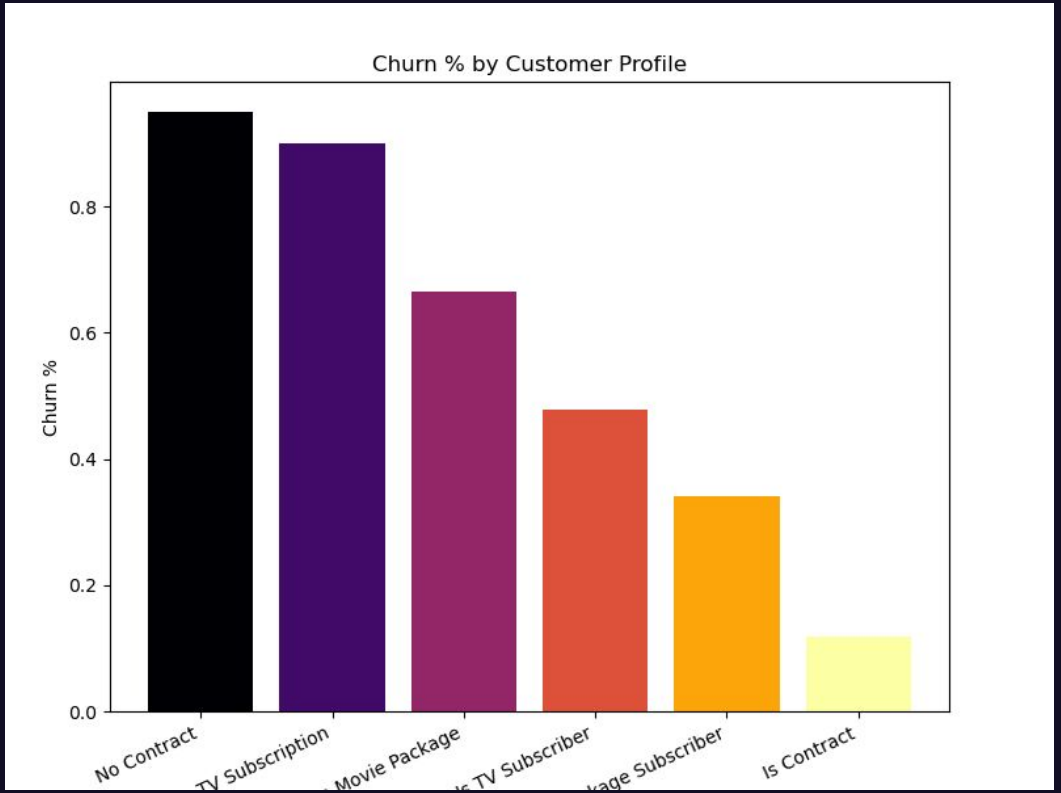
Customers with 2.5+ year contracts did not churn at all.

An ISP can retain long-term customers by delivering a product or business promotion that pushes customers to want to keep their plan for 2.5+ years.



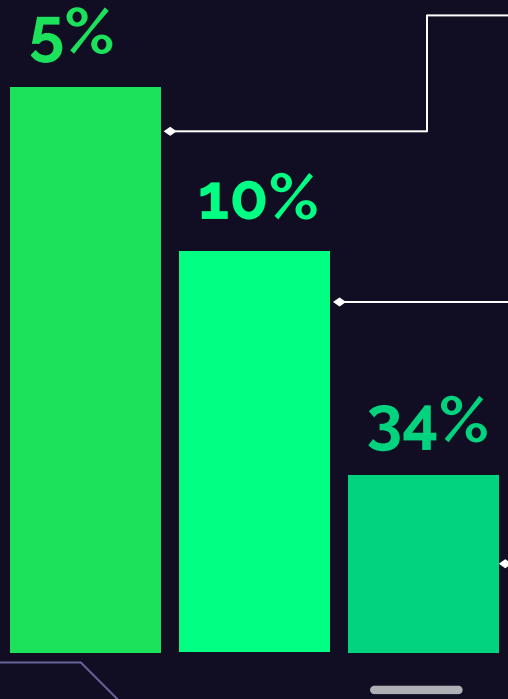
Business Recommendations

The first steps that should be taken to reduce customers from churning is for the business to focus on providing as much value in these 3 products to the customers as possible as the data shows, the long-term customers that can be gained can easily outweigh the cost of these product investments.



Business Recommendations

Customer profiles who churn the least



Internet Service Contracts

Customers with contracts

TV Subscriptions

Customers with TV subscriptions

Movie Package Subscriptions

Customers with movie package subscription

Future Steps

Here are some examples of data we should strive to attain going forward that will further establish a more effective business strategy.

Budget Data

With more financial info we can analyze where most of our profit is coming from and what products are not worth the cost.

Timeseries

With actual timestamps we can analyze what parts of the year people tend to churn and identify monthly or yearly trends.

Demographics

With information like gendar, household vs single adult, location, etc., we could further improve our model and identify more accurate customer profiles that we can target.

THANKS!

Any questions?

James Oomens
DSB-122
Capstone - ISP Churn Prediction