Predict Customers Churn

ADVANCED DATA SCIENCE PROJECT

UTSAV RAJ & THOMAS HASS

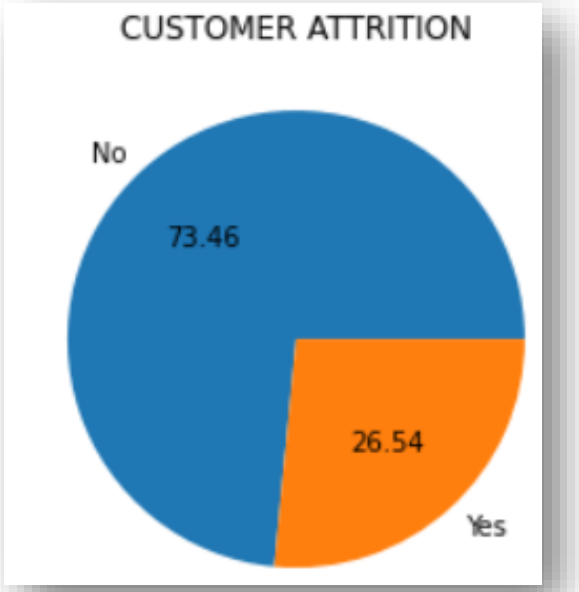
Introduction

Your phone service is bad?

You want to change your provider?

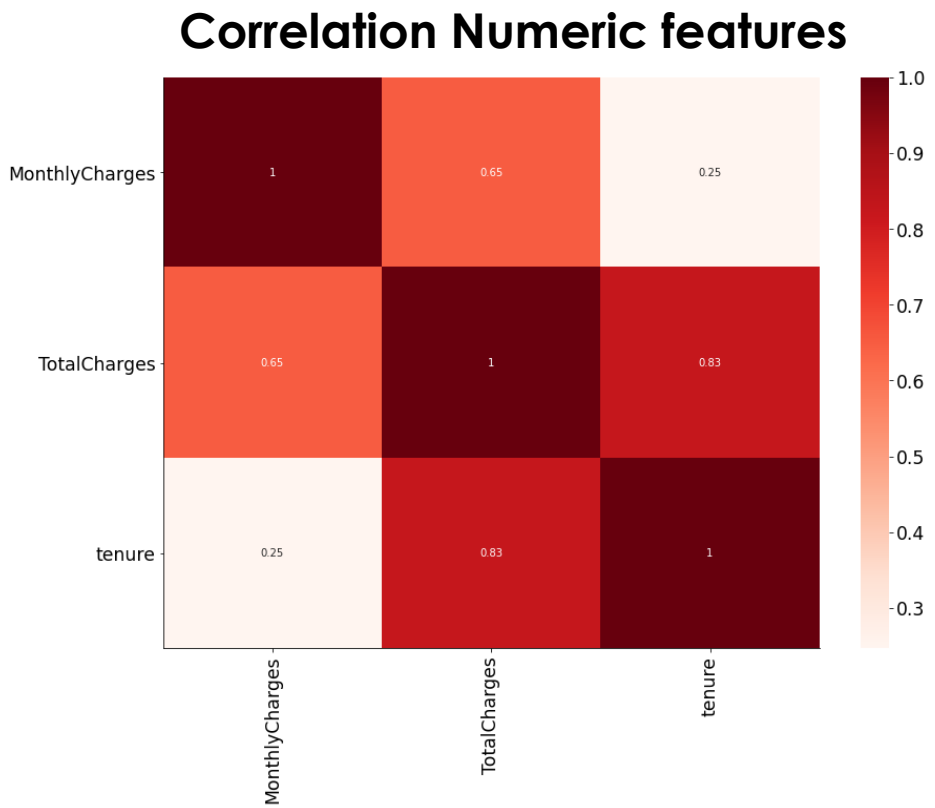
Well, your provider do not want you to do that!

Data Exploration

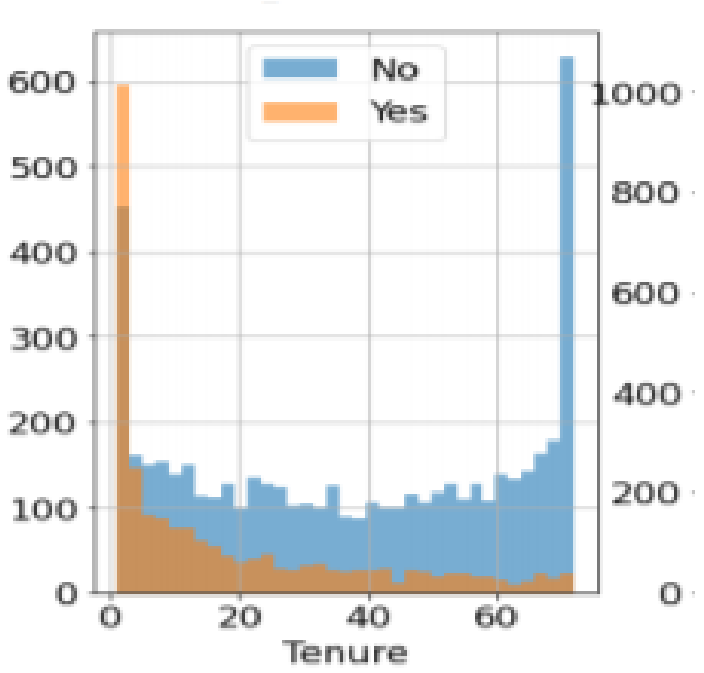


- Dataset from Kaggle
- Column „Churn“ → Yes/No → Left company or not
- Around 7000 columns
- 21 features: gender, tenure, monthly charges …

Data Exploration – numeric features

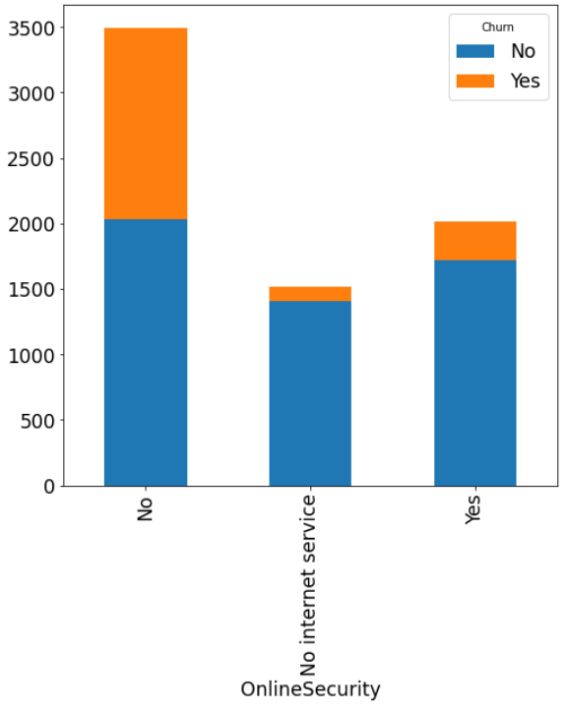


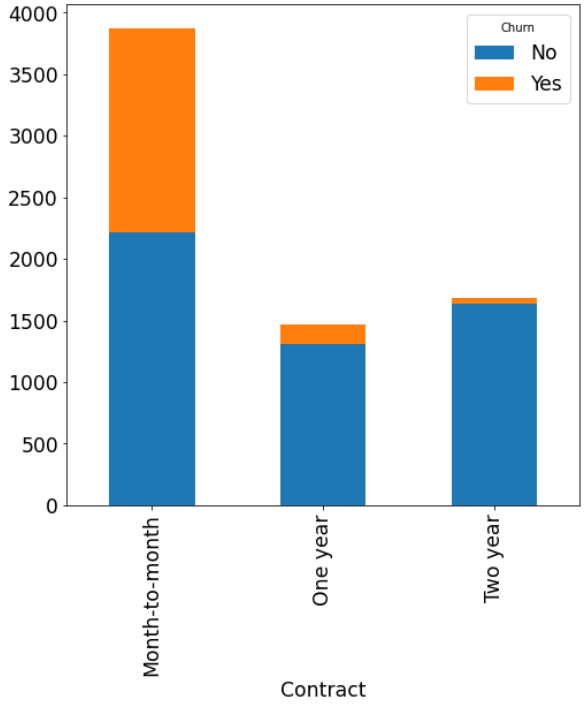
→ High correlation between tenure and Total Charges



→ Customers churn in first few months

Data Exploration – categorical features



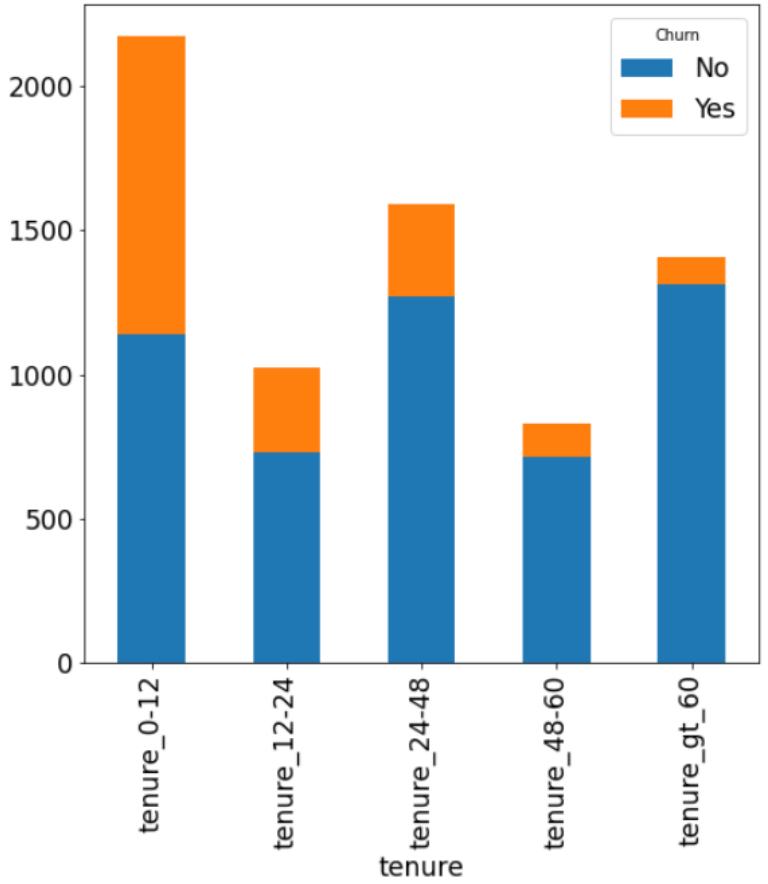


Obeservations:

- Customers with Online Security, Device Protection, Online Backup and Tech Support service are less likely to leave
- Increase in contact length → decrease in chrun rate
- Demographics effects churn rate
- Customers in relationship have lover churn rates than singles

Clean & Prepare Data

- Data is already almost cleaned and prepared
- Remove unusable columns (TotalCharges not given)
- Transform tenure to categories
- Remove CustomerID column
- Convert Churn values from Yes/No to 1/0



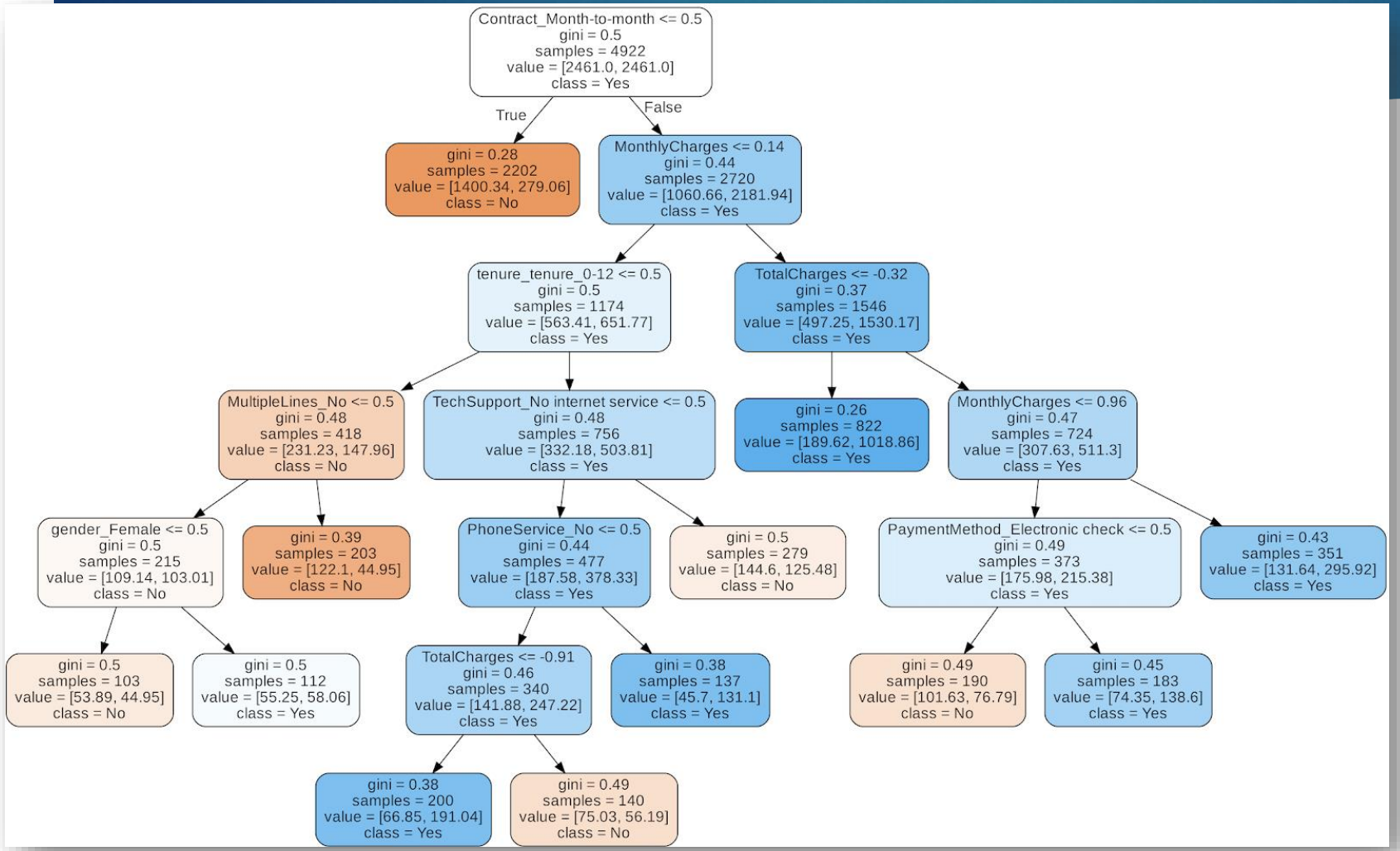
Modelling - Preparation

- Plan: Logistic Regression and Decision Tree
- Training set 70%, testing set 30%
- Class weight = balanced
- Standard Scaling on numerical features
- One Hot Coding on categorical features

Modelling – Decision Tree

- Concept: predict the churn by learning simple decision rules inferred from prior data
- GridSearch
- Gini Index

Modelling – Decision Tree



→ Most important features:

- Contract
- Monthly Charges
- Total Charges
- Tenure
- Tech Support

Performance:

	Accuracy	Macro F1-score
Train	0.7641	0.7276
Test	0.7550	0.7151

Modelling – Logistic Regression

- Uses the logistic function (with the features we are using) to predict churn
- >0.5 -> 1; <0.5 -> 0
- Grid Search
- Most important features: Monthly Charges, Tenure, Contract, Tech Support, Payment

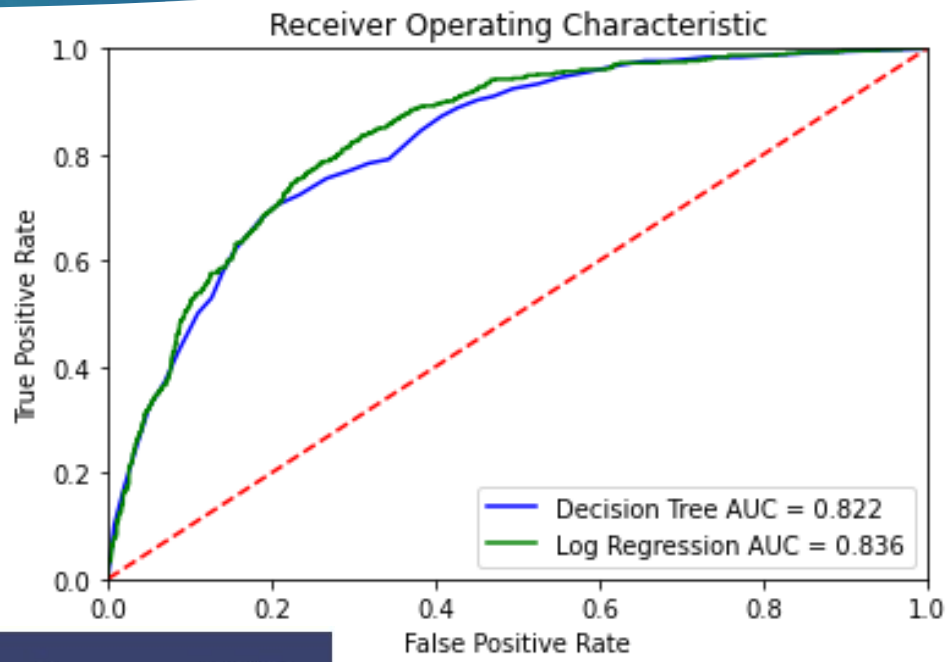
Performance:

	Accuracy	Macro F1-score	Macro Precision	Macro Recall
Train	0.7495	0.7350	0.7199	0.7640
Test	0.7517	0.7189	0.7114	0.7608

Compare Results

- Logistic Regression higher F1-Score
- Logistic Regression performs better in AUC

→ Logistic Regression outperforms Decision Tree



Model	Accuracy (Test)	Recall	Precision	F1-score	Area under curve (AUC)
Decision Tree	0.755	0.7451	0.7055	0.7151	0.8222
Logistic Regression	0.7517	0.7608	0.7114	0.7189	0.8357

Conclusion

Features Value	Churn	Reason
Monthly Charges	*Yes*	Price is the biggest reason to switch and higher prices pushes customers to find competitors with cheapest service.
Tenure (0-12 Months)		Customers are learning to use your product and deciding whether or not to stay with it
Contract (Month to Month)		12 purchasing decisions per year and hence, more time to reflect on the cost-effectiveness of Telco's services
Contract (Two Years)	*No*	Customers have only one purchasing decision per two year (albeit a larger one)
Tenure (More than 60M)		Customers will have enough time to implement the product and see the benefits of using it and hence, are more likely to commit.
Total Charges		The more customers have spent on the company, the more likely they are to not Churn.

Conclusion

→ **With the help of data science processes, new knowledge has been gained that can be used in the future to improve the service and to target and convince high-risk customers to remain loyal to the provider, thereby saving a considerable amount of money**