A data driven approach for Customer Churn prediction.

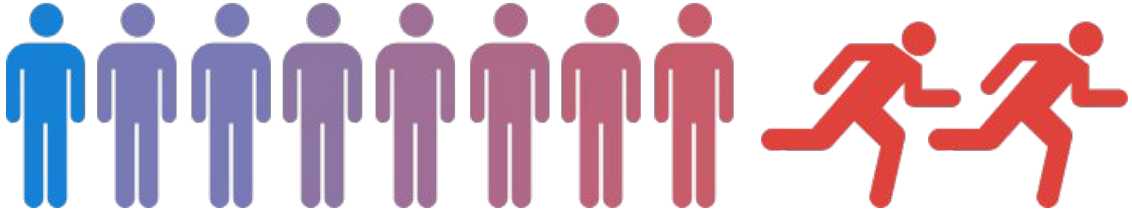
MSc. Noel Chaparro Portillo

TELCO

Outline

TELCO

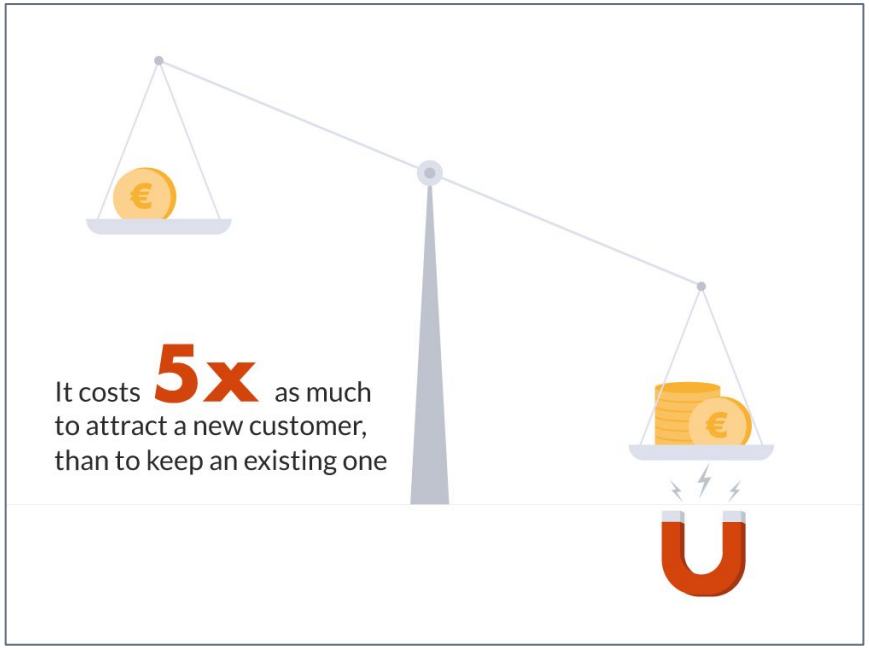
01. Introduction: What is customer churn?



Customer churn, also known as customer attrition, is when someone chooses to stop using your products or services

TELCO

01. Introduction: Why does churn matter?



According to the Harvard Business School… a **5%** increase in customer retention rates results in **25% – 95%** increase of profits.

Source: https://www.superoffice.com/blog/reduce-customer-churn/

01. Introduction: Problem statement.

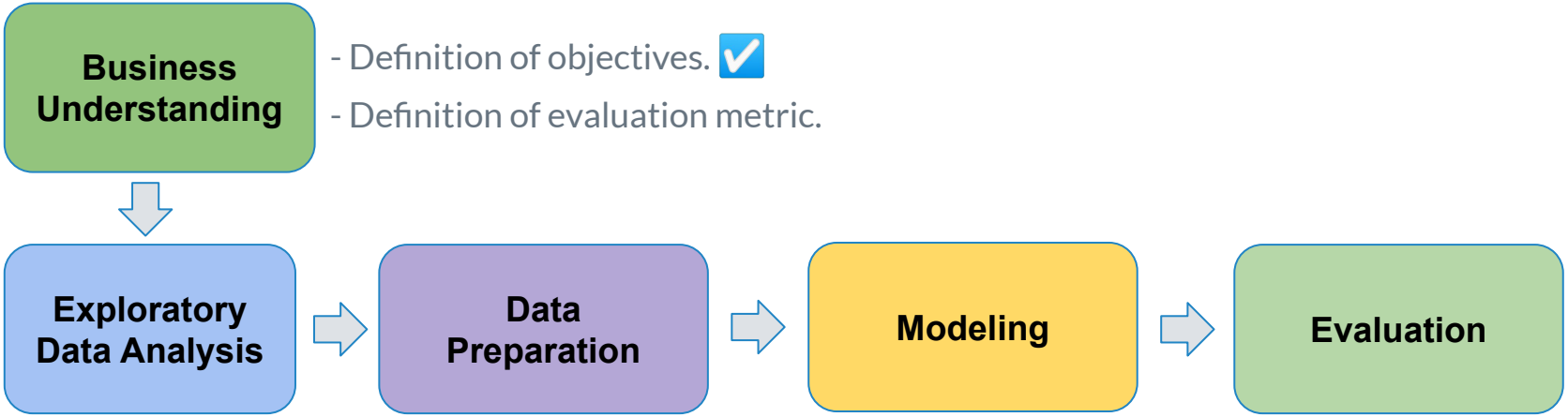
"The **TELCO** company is suffering
from a high customer churn rate
decreasing company revenue"

02. Objectives.

1. To create a model to predict customer churn.
2. To design a retention program to increase engagement.

TELCO

03. Methodology: overview.



- Univariate analysis.
- Bivariate analysis.
- Pearson's correlation.
- Cramer's V correlation,

- Data cleaning.
- Feature encoding.
- Dataset split.
- Feature scaling.

- Model selection.
- Hyperparameter tuning.
- K-fold cross validation.
- Generalization.

- Model evaluation.
- Statistical tests.
- Probability thresholds.

03. Methodology: Business Understanding

Model Evaluation Metric

Accuracy

F1-Score

F2-Score

TELCO

03. Methodology: EDA.

Demographic

- Gender.
- Senior Citizen.
- Partner.
- Dependents.

Customer

- Customer ID.
- Tenure.

Target

- Churn.

Services

- Phone Service.
- Multiple Lines.
- Internet Service.
- Online Security.
- Online Backup.
- Device.
- Protection.
- Tech Support.
- Streaming TV.
- Streaming Movies.

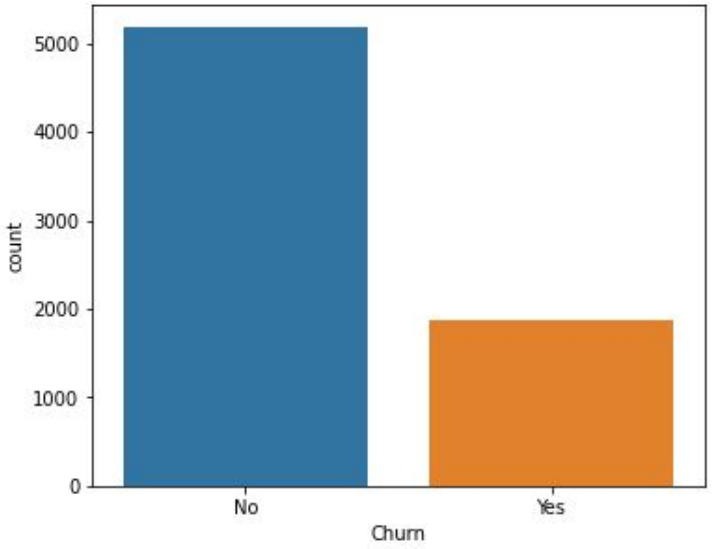
Billing

- Contract.
- Paperless Billing.
- Payment Method.
- Monthly Charges.
- Total Charges.

TELCO

03. Methodology: EDA.

Churn count

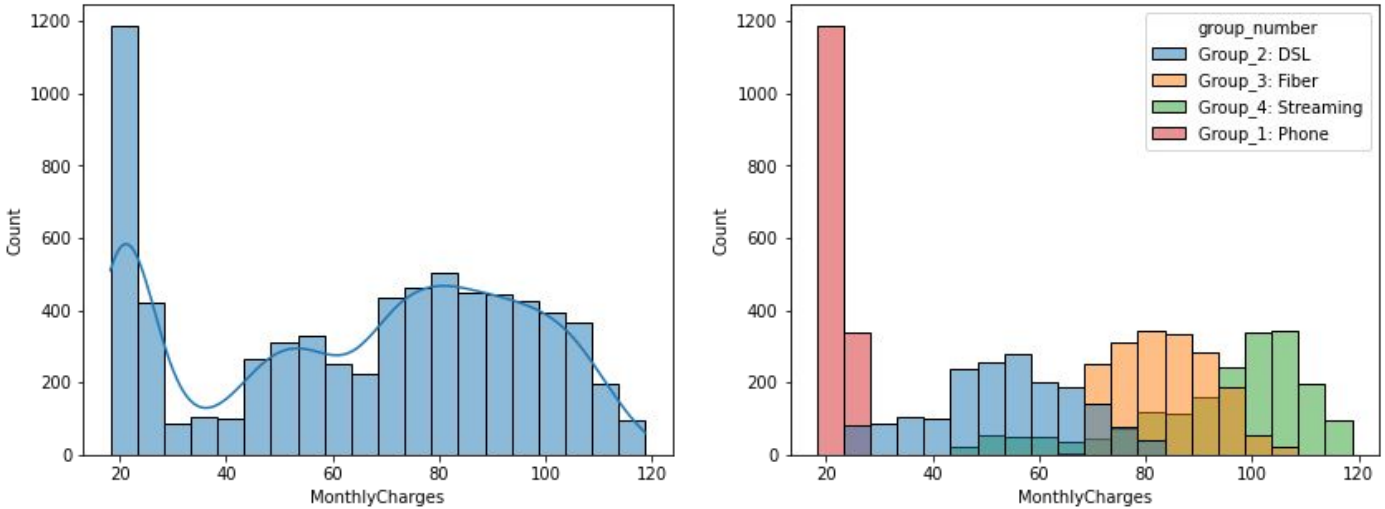


Observation: About **27%** of the total customers are churners.

TELCO

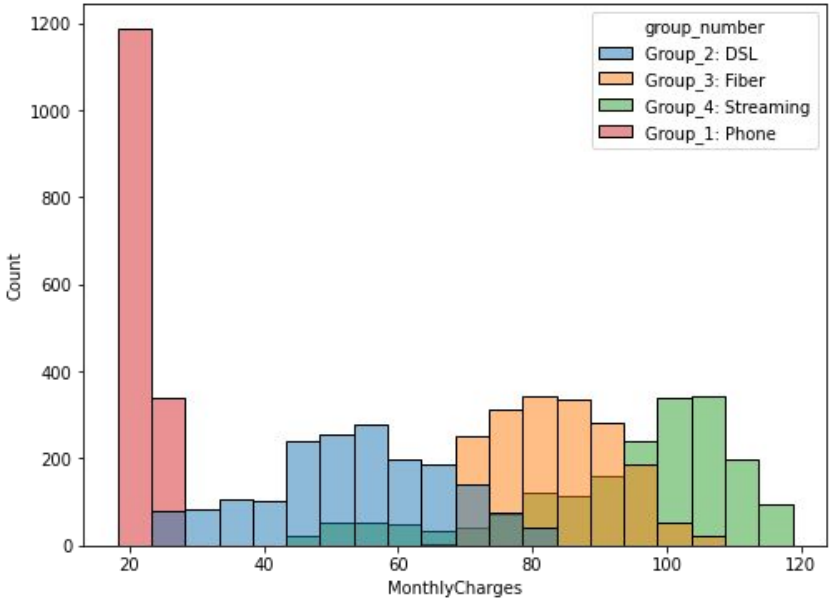
03. Methodology: EDA.

Monthly Charges distribution



Bundle type reduction: 13,122 —> 322 —> 4

03. Methodology: EDA segmentation.

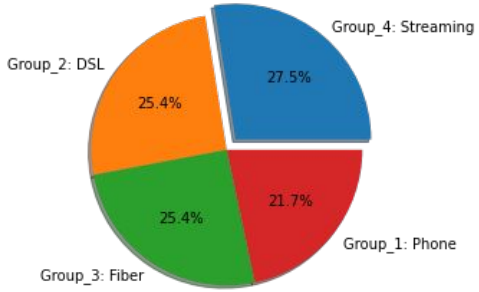


Group_1: Customers with PhoneService but no InternetService.

Group_2: Customers with InternetService (DSL).

Group_3: Customers with InternetService (Fiber optic).

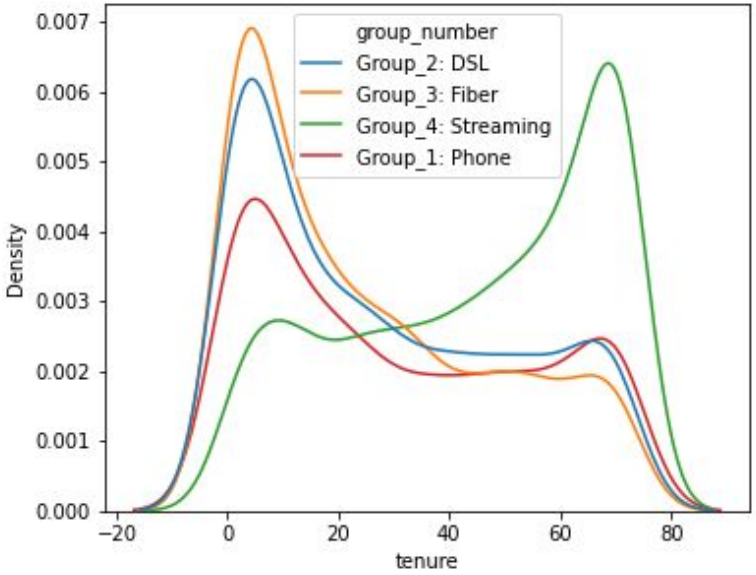
Group_4: Customers with StreamingTV or StreamingMovies.



TELCO

03. Methodology: EDA.

Tenure distribution vs. bundle type.

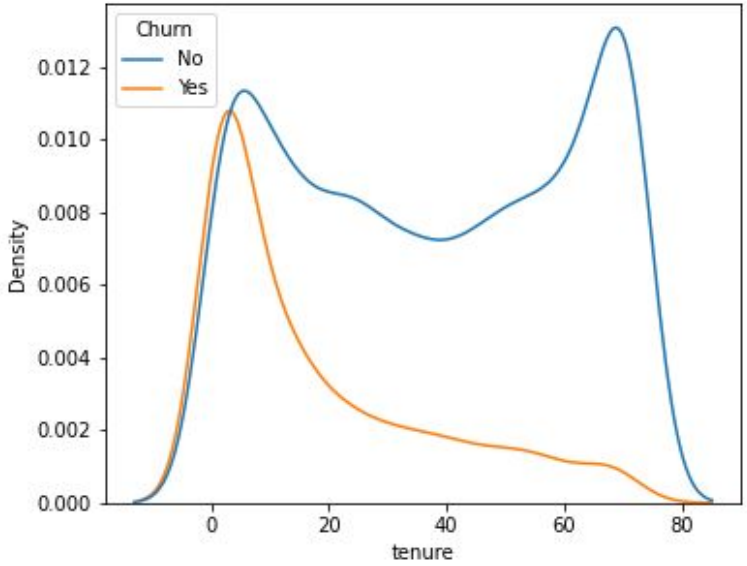


Observation: Customers consuming any streaming service are more likely to stay longer with the company.

TELCO

03. Methodology: EDA.

Tenure distribution vs. Churn.

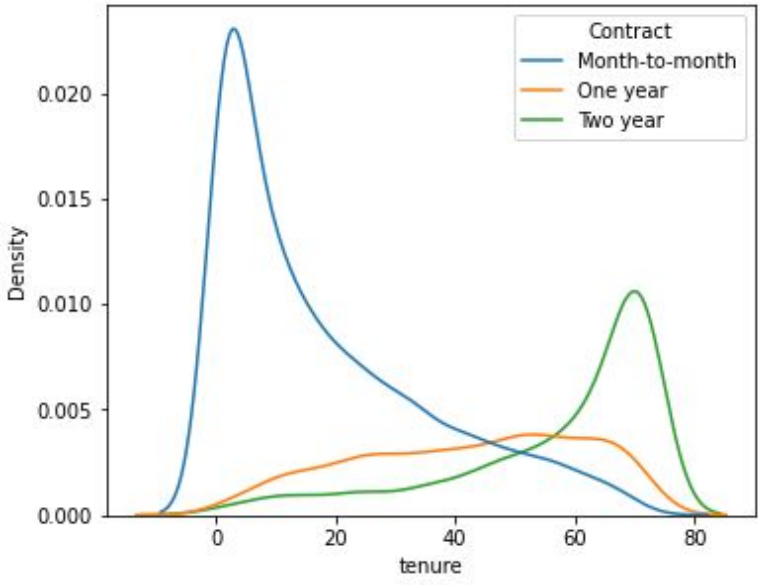


Observation: Most churners leave the company at very early stages.

TELCO

03. Methodology: EDA.

Tenure distribution vs. contract.

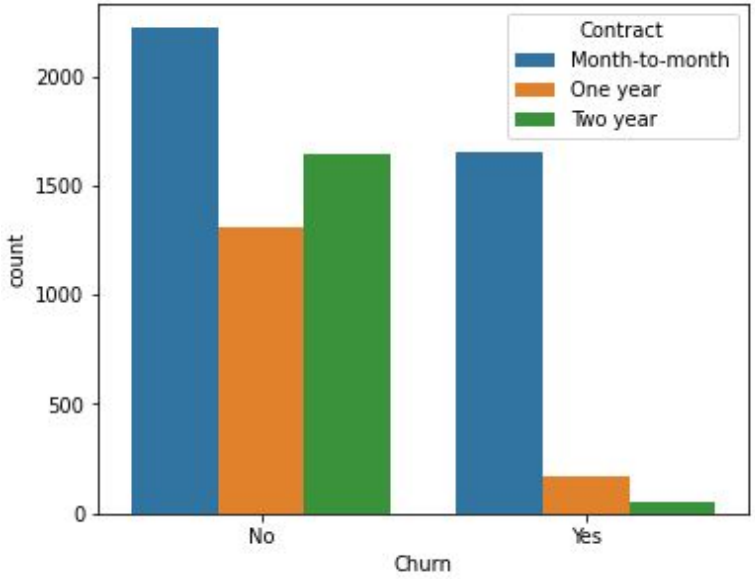


Observation: Month-to-month contract customers tend to leave the company much more easily.

TELCO

03. Methodology: EDA.

Churn vs. contract.

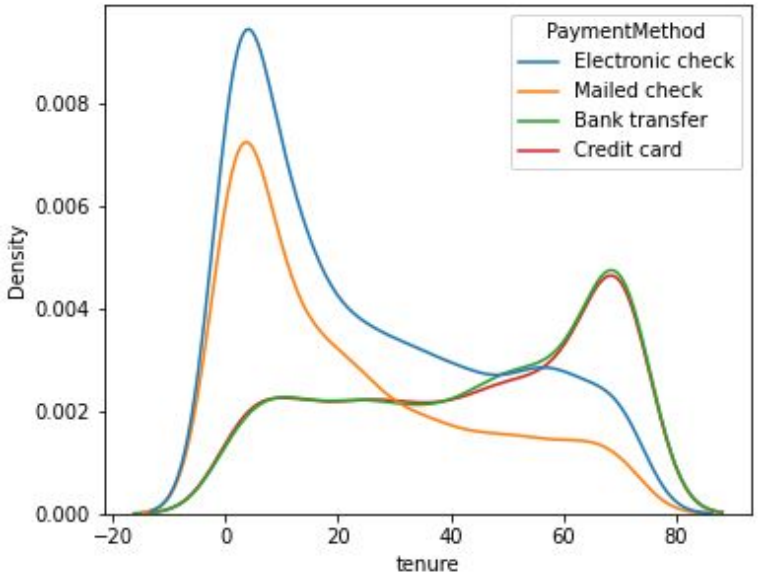


Observation: This just reaffirms our previous observation about Month-to-month contracts.

TELCO

03. Methodology: EDA.

Tenure distribution vs. payment method.

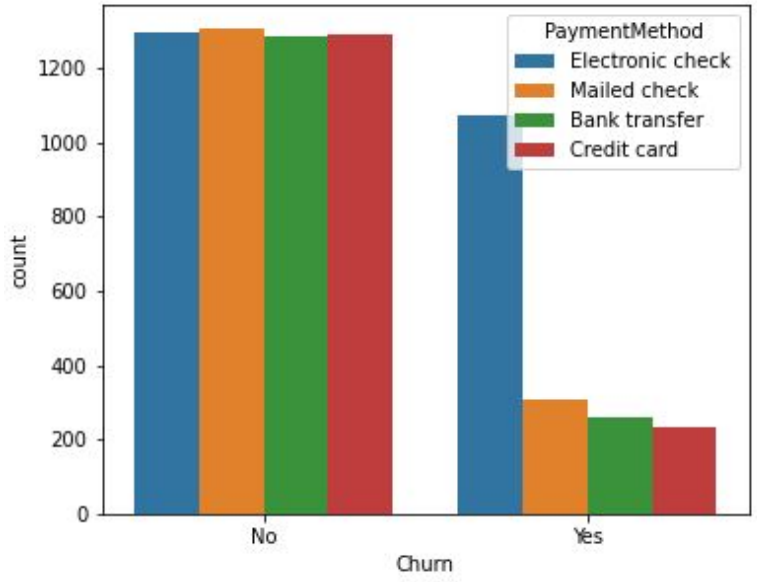


Observation: Automatic payment methods like bank transfer or credit card are well related with longer engagement.

TELCO

03. Methodology: EDA.

Churn vs. payment method.



Observation: Customers with Electronic Check as payment method are more likely to leave.

TELCO

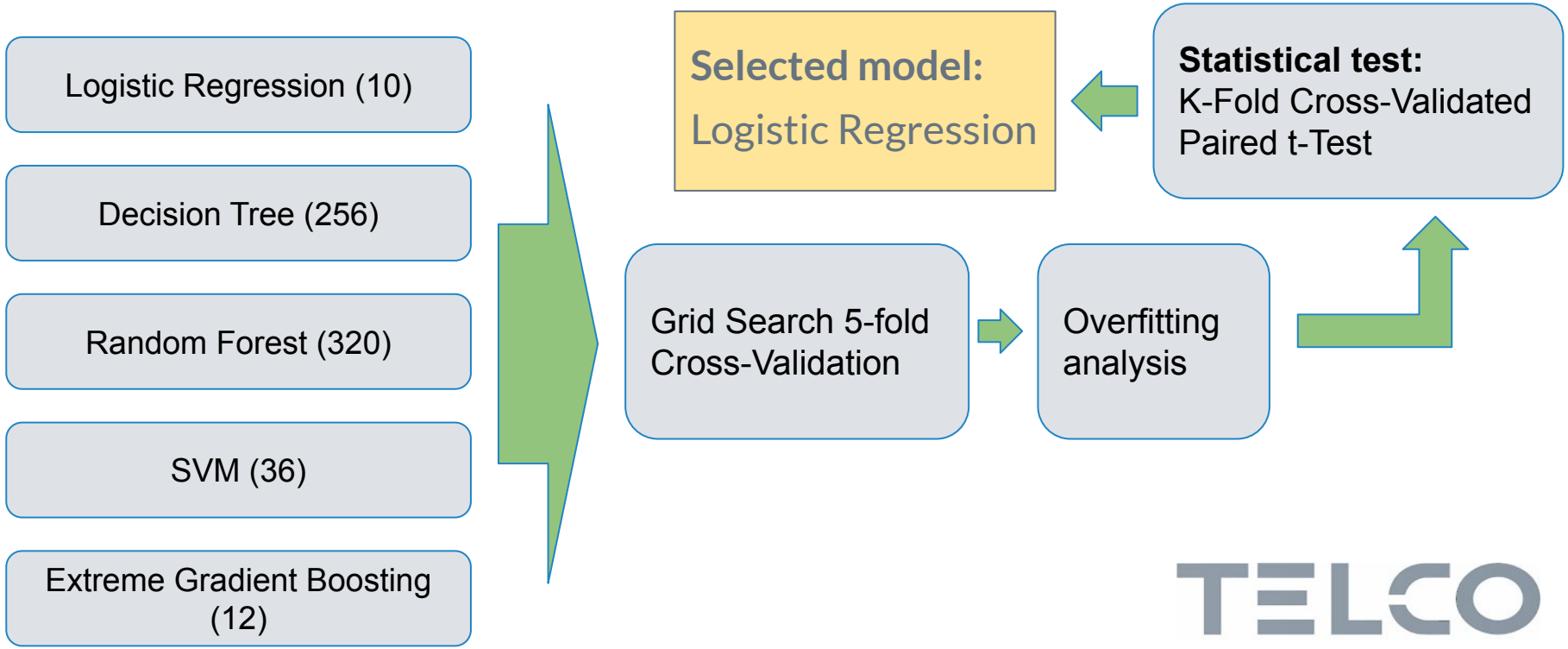
03. Methodology: Data preparation.

Actions taken:

- 11 blank values in “TotalCharges” were replaced by a zero value (0).
- No duplicate values found.
- Categorical features were integer encoded then OneHot encoded.
- Dataset was split in a 80% /20% ratio for training and test.
- For numerical features: a z-score normalization was applied whenever a model was trained or cross-validated.

TELCO

03. Methodology: Modeling.



03. Methodology: Modeling.

Overfitting analysis.

	mean_train_score	std_train_score	mean_test_score	std_test_score
Logistic Regression	0.72466	0.00502711	0.719093	0.0226338
Decision Tree	0.727275	0.0134499	0.714764	0.0154759
Random Forest	0.736085	0.00778334	0.725454	0.0158456
SVM	0.720131	0.00355627	0.720089	0.0142873
XG Boost	0.595087	0.00420724	0.550611	0.0269865

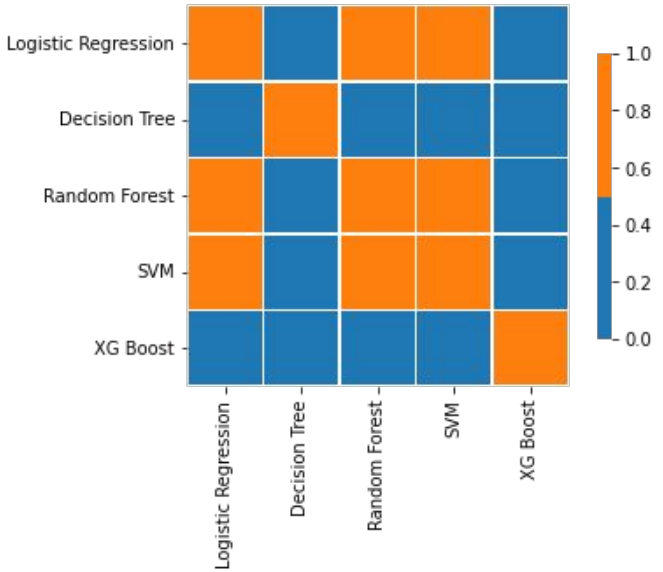
TELCO

04. Model performance: evaluation.

Models with the best hyperparameters.

	precision	recall	f2-score	accuracy
Baseline	0.27	1	0.64	0.27
Logistic Regression	0.52	0.81	**0.73**	0.75
Decision Tree	0.54	0.74	0.69	0.77
Random Forest	0.51	0.78	0.71	0.75
SVM	0.47	0.82	0.71	0.7
XG Boost	0.68	0.52	0.55	0.81

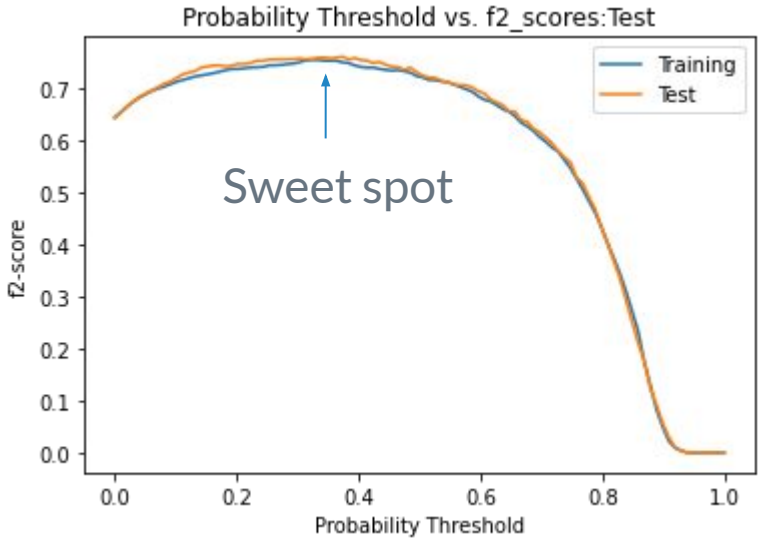
K-Fold Cross-Validated Paired t-Test. (20x5 scores)



Accepts null hypothesis: mean difference between paired observations is zero, α=0.05.

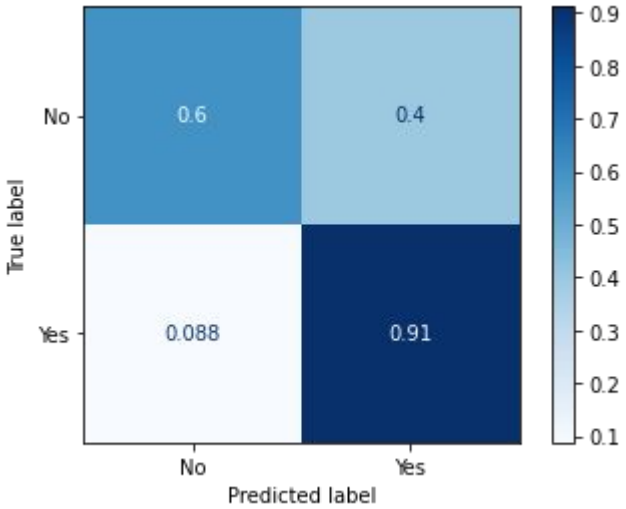
04. Model performance.

Moving probability thresholds for classifier predictions.



04. Model performance: final model.

Normalized confusion matrix.



Classification report.

	precision	recall	f1-score	support
0	0.95	0.60	0.74	1035
1	0.45	0.91	0.60	374
accuracy			0.68	1409
macro avg	0.70	0.76	0.67	1409
weighted avg	0.82	0.68	0.70	1409

F2-score: 0.76

TELCO

05. Retention program.

1

Retention Program 1: Encourage payment methods with autopay option by giving a 5% discount on the service.

2

Retention Program 2: Promote service bundle updates to include any streaming service.

3

Retention Program 3: Offer a substantial price discount on services to customers with longer contracts.

TELCO

05. Retention program cont.

4

Retention Program 4: Remove Month-to-month contracts.

5

Retention Program 5: Use the predictive model to gauge the level of risk for a person to churn and to create a proportional action.

Bonus: Reach anyone predicted to churn for a customer satisfaction / service evaluation survey. Act accordingly.

TELCO

Conclusions

- In this project, a data driven approach was implemented in order to reduce customer churn. A model with a very high recall (**91%**) of possible churners was generated. Even though the precision is around 50% the benefits overweights the related precision costs.

- Finally, a set of retention programas was proposed base on the information extracted from the data.