

---

---

# Um estudo sobre Churn

Descobrimos por qual motivo nossos clientes  
desistem

---

---



# Importância

Temos um alto custo de implementação do produto;

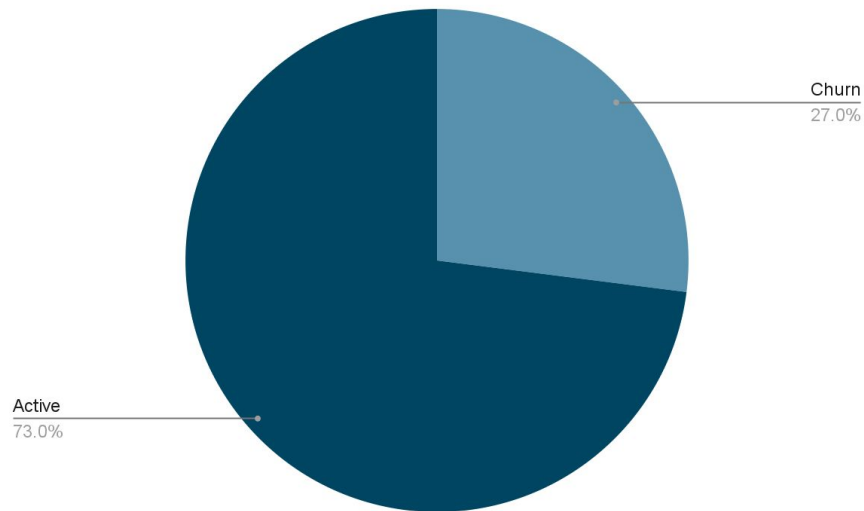
Quanto mais tempo um cliente fica ativo, maior o *revenue*;

A identificação de clientes prestes a *churnear* pode reduzir o custo das comunicações de retenção;

Características de Clientes Chuneados podem apontar áreas de melhoria no serviço.

# Data

- 7 mil clientes
- Informações em 21 *features*
  - Client personal information (ID, Gender, Senior, Partner or Dependents)
  - Hired product information (Multiple Lines, Public Service or Not, Streaming Services and Payment Methods)
- Total Payment 1400 USD
- Montly Payment 70 USD
- 29 months tenure



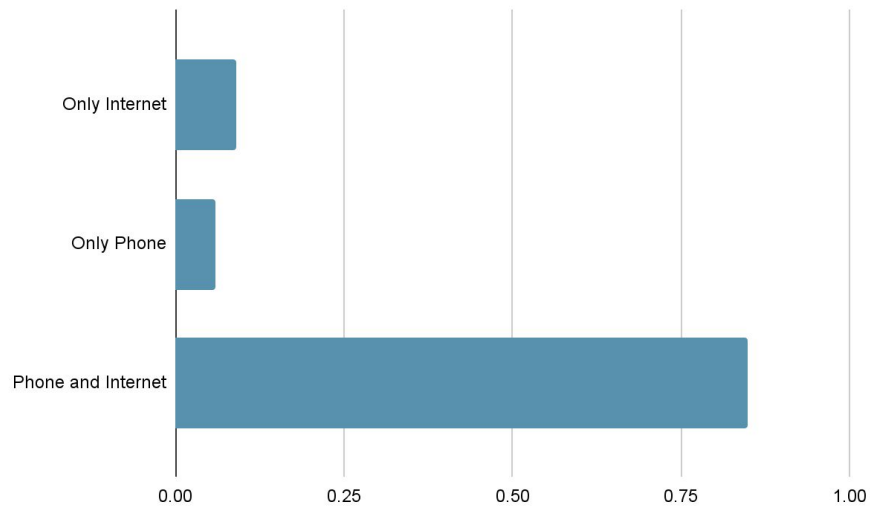
**1.6 mil**  
Mediana de  
Revenue Total de  
Ativos

**65 USD**  
Mediana de Revenue  
Mensal de Ativos

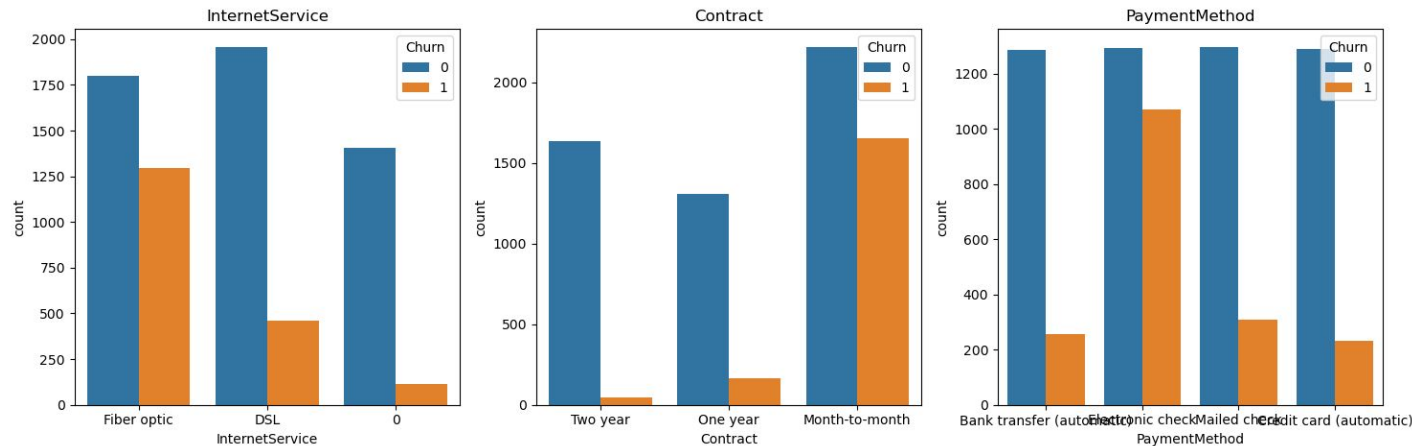
# Clients and Services

| <i>Feature</i>          | <i>No</i> | <i>Yes</i> |
|-------------------------|-----------|------------|
| Senior Citizen          | 83%       |            |
| Dependents              | 70%       |            |
| Have Phone Service      |           | 90%        |
| Have Internet Service   |           | 78%        |
| Have Internet and Phone |           | 68%        |

# Churn by Service



# Diferença entre Churn e Ativo



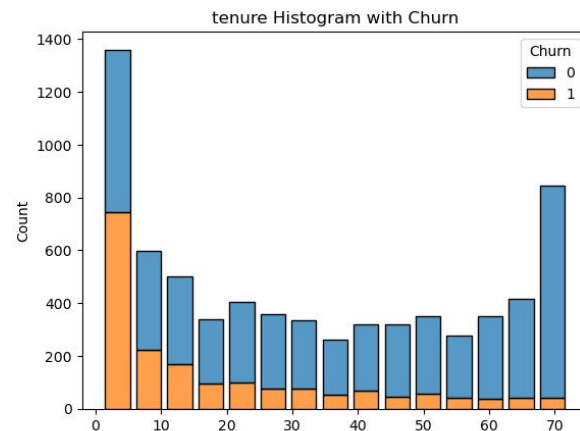
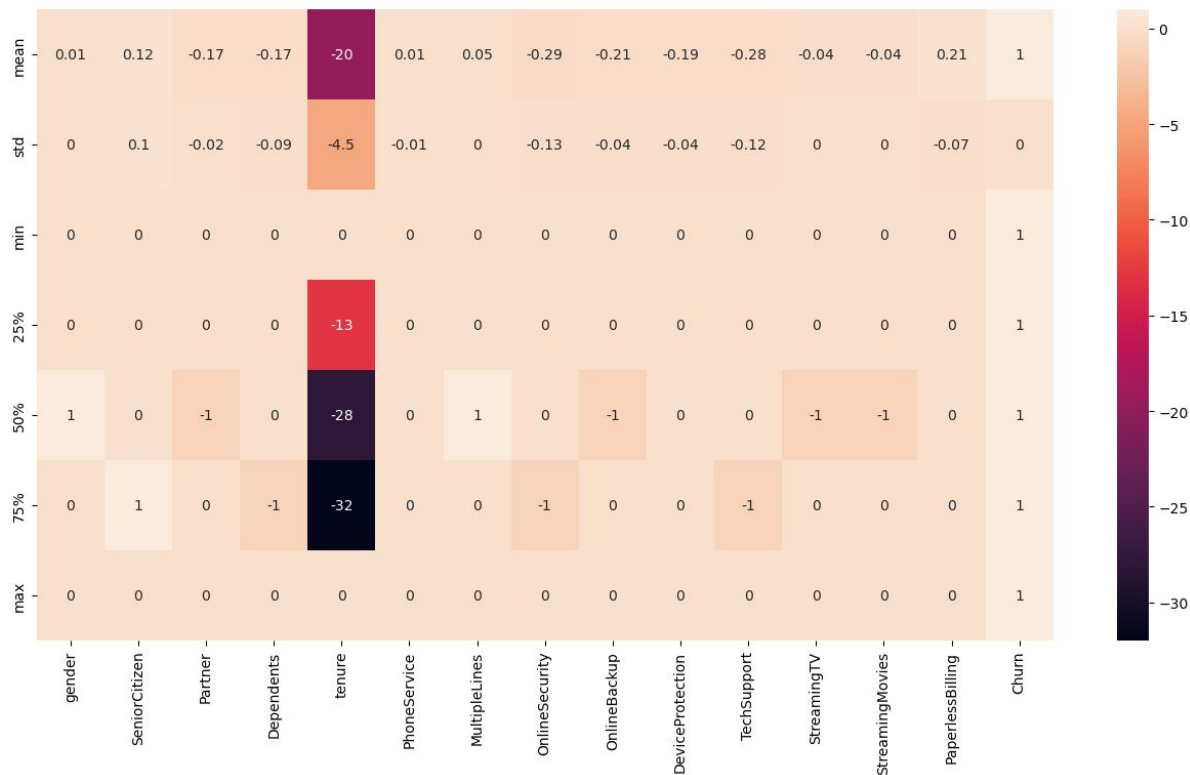
## Churned

|        | customerID | InternetService | Contract       | PaymentMethod    |
|--------|------------|-----------------|----------------|------------------|
| count  | 1869       | 1869            | 1869           | 1869             |
| unique | 1869       | 3               | 3              | 4                |
| top    | 8199-ZLLSA | Fiber optic     | Month-to-month | Electronic check |
| freq   | 1          | 1297            | 1655           | 1071             |

## Active

|        | customerID | InternetService | Contract       | PaymentMethod |
|--------|------------|-----------------|----------------|---------------|
| count  | 5163       | 5163            | 5163           | 5163          |
| unique | 5163       | 3               | 3              | 4             |
| top    | 7569-NMZYQ | DSL             | Month-to-month | Mailed check  |
| freq   | 1          | 1957            | 2220           | 1296          |

# Diferença entre Churn e Ativo



38  
Mediana de Tenue  
Total de Ativos

10  
Mediana de Tenue  
Total de  
Churneados

# Hipóteses

- Existe uma diferença estatística entre as features dos nossos dados e a condição de ter ou não ter *churnado*;
- Essas diferenças possibilitam a identificação de clientes *churnados* em algoritmos de aprendizado de máquina com uma precisão acima do aleatório.



# Teste: Correlação

All clients

| Continuous_Variable | Point-Biserial_Correlation | P_Value      | Interpretation  |
|---------------------|----------------------------|--------------|-----------------|
| gender              | 0.008545                   | 4.737357e-01 | Not Significant |
| SeniorCitizen       | 0.150541                   | 6.377295e-37 | Significant     |
| Partner             | -0.149982                  | 1.171290e-36 | Significant     |
| Dependents          | -0.163128                  | 3.896536e-43 | Significant     |
| PhoneService        | 0.011691                   | 3.269553e-01 | Not Significant |
| PaperlessBilling    | 0.191454                   | 4.879248e-59 | Significant     |
| Churn               | 1.000000                   | 0.000000e+00 | Significant     |

Hired internet

| Continuous_Variable | Point-Biserial_Correlation | P_Value       | Interpretation  |
|---------------------|----------------------------|---------------|-----------------|
| OnlineSecurity      | -0.280508                  | 3.365672e-100 | Significant     |
| TechSupport         | -0.274112                  | 1.364799e-95  | Significant     |
| OnlineBackup        | -0.195758                  | 9.673218e-49  | Significant     |
| DeviceProtection    | -0.176806                  | 6.052125e-40  | Significant     |
| Partner             | -0.161626                  | 1.408306e-33  | Significant     |
| Dependents          | -0.147970                  | 2.342520e-28  | Significant     |
| StreamingMovies     | -0.040525                  | 2.618977e-03  | Significant     |
| StreamingTV         | -0.036698                  | 6.432116e-03  | Significant     |
| gender              | 0.008608                   | 5.228634e-01  | Not Significant |
| PhoneService        | 0.055215                   | 4.104376e-05  | Significant     |
| SeniorCitizen       | 0.120978                   | 2.007380e-19  | Significant     |
| PaperlessBilling    | 0.142987                   | 1.433099e-26  | Significant     |
| Churn               | 1.000000                   | 0.000000e+00  | Significant     |

# Teste: ANOVA

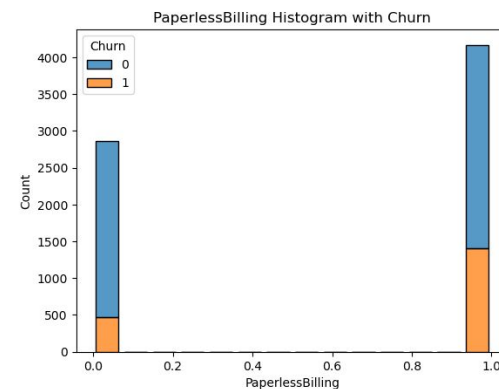
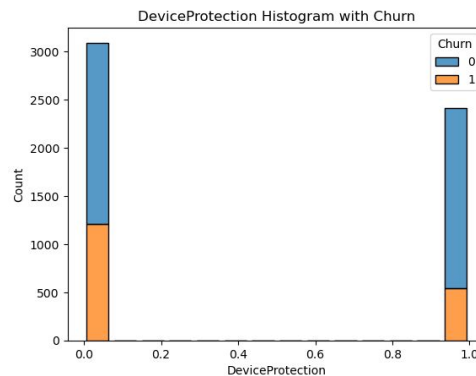
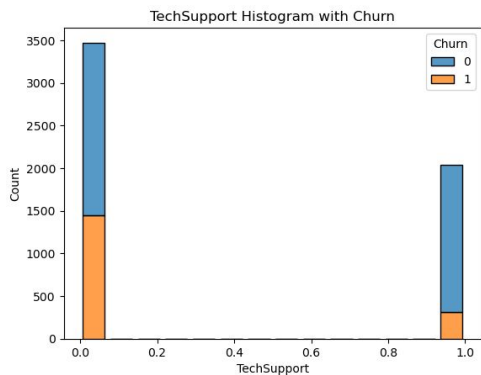
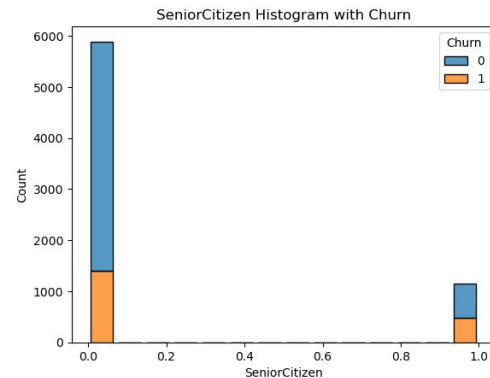
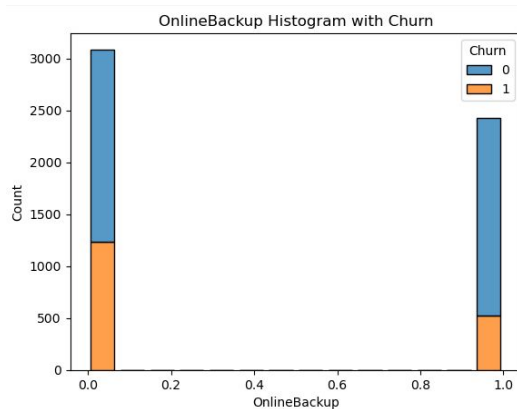
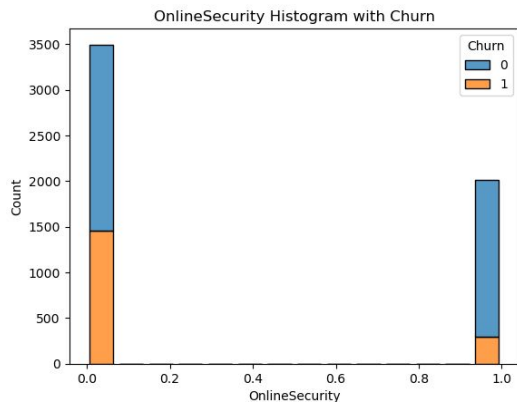
All clients

| Categorical_Variable | F_Statistic | P_Value       | Significant |
|----------------------|-------------|---------------|-------------|
| InternetService      | 406.294949  | 1.056959e-167 | True        |
| Contract             | 708.337681  | 5.633990e-281 | True        |
| PaymentMethod        | 236.751040  | 2.414992e-146 | True        |

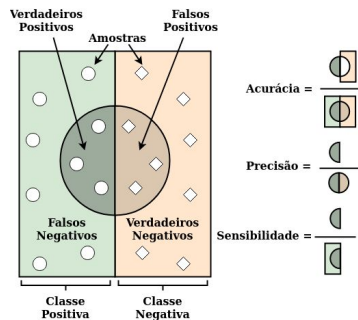
Hired internet

| Categorical_Variable | F_Statistic | P_Value       | Significant |
|----------------------|-------------|---------------|-------------|
| InternetService      | 348.236251  | 2.079449e-75  | True        |
| Contract             | 510.441250  | 4.247361e-204 | True        |
| PaymentMethod        | 152.504352  | 6.012633e-95  | True        |

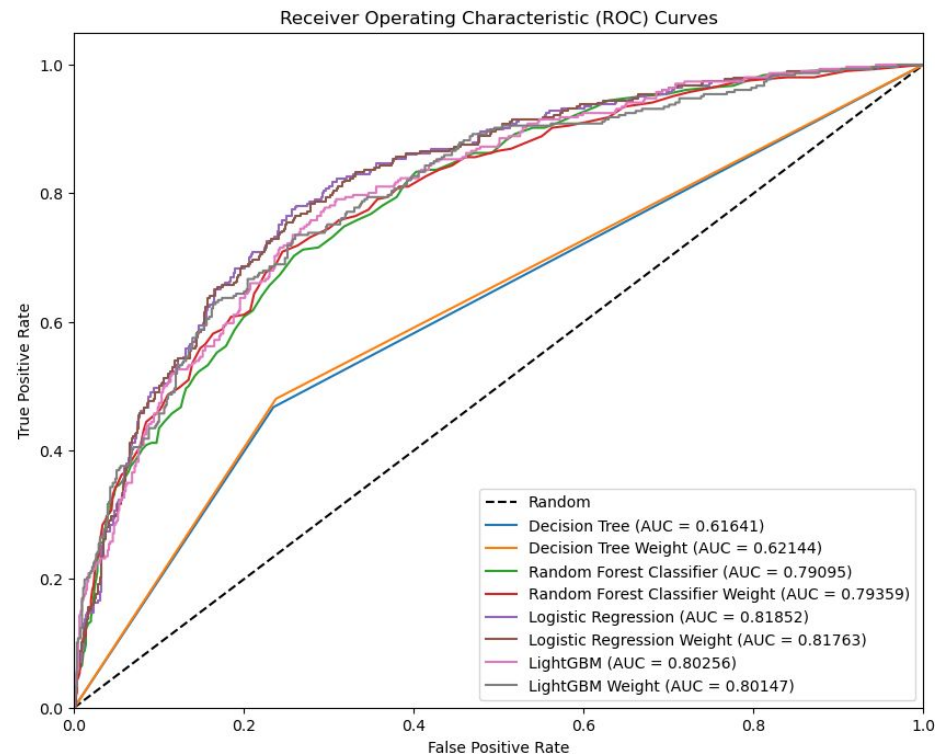
# Churn e Features



# Previendo Churn



| Model                           | Accuracy | Precision | Recall   | F1-Score |
|---------------------------------|----------|-----------|----------|----------|
| Random Forest Classifier Weight | 0.770434 | 0.597270  | 0.460526 | 0.520059 |
| LightGBM                        | 0.773987 | 0.592814  | 0.521053 | 0.554622 |
| Random Forest Classifier        | 0.766880 | 0.583871  | 0.476316 | 0.524638 |
| Logistic Regression             | 0.768301 | 0.577586  | 0.528947 | 0.552198 |
| LightGBM Weight                 | 0.740583 | 0.513915  | 0.728947 | 0.602829 |
| Logistic Regression Weight      | 0.733475 | 0.504146  | 0.800000 | 0.618515 |
| Decision Tree Weight            | 0.727790 | 0.496000  | 0.489474 | 0.492715 |
| Decision Tree                   | 0.720682 | 0.483204  | 0.492105 | 0.487614 |
| Dummy Classifier                | 0.729922 | 0.000000  | 0.000000 | 0.000000 |



# Estimativa de Impacto

Considerando que a mediana de revenue do cliente ativo no mes é de 65 USD, para cada cliente que estivessemos mantendo em nossa base, teríamos um valor médio de **65USD**.

Implementando o projeto até dezembro de 2023 e estimando uma precisão de 60% do modelo de churn; e também estimando uma taxa de aumento de permanência do cliente em mais um mês por iniciativas de comunicação como sendo  $\frac{1}{3}$ , teríamos em janeiro um aumento de **2145 USD** apenas nesse mes, ou de **25 mil USD** ao final de um ano.

# Conclusão

Fibra óptica teve a maior quantidade de churn

Contratos maiores tiveram menos churn

Transferência Bancária teve menos churn

Foi possível atingir uma ROC AUC de 0.81 com logistic regression

# Insights

Podemos investigar melhor por qual motivo o churn esta alto para Fibra Optica, fazendo uma análise de NLP em cima dos audios de contato entre a equipe e o cliente

O time de atendimento pode sugerir ao cliente para um tempo de contrato maior

Podemos pensar em promoções com um maior tempo de contrato

Adicionar mais features no modelo irá permitir termos uma capacidade maior de identificação de pessoas churneando.

# ROADMAP

