**INSPER**

**GRADUAÇÃO EM ADMINISTRAÇÃO**

**LUCAS GIMENES -** [lucasg4@al.insper.edu.br](mailto:lucasg4@al.insper.edu.br)

**APS1 DE ESTATISTICA 1**

Previsão de Churn de Clientes com base em análise de dados

Tatiana Terabayashi Melhado

**Resumo**

O objetivo dessa APS é, com base nos dados disponibilizados e a matéria ensinada em aula, desenvolver a análise estatística de Chun ocorrido em uma empresa. Churn é o indicador de quantos clientes cancelaram o contrato ou deixaram de consumir o produto ofertado, obtido por um número inteiro.

Os dados expostos apresentavam variáveis diversas e de vários tipos, como qual plano contratado pela empresa, valor pago ao mês e tempo que o cliente consome os produtos ofertados. Essas variáveis quando cruzadas poder nos mostrar correlações não vista antes, facilitando tomadas de decisões.

**Introdução**

Quando tentamos entender as ações tomadas por indivíduos, enfrentamos um desavio de entender que acontecimentos passados tornaram aquela ação possível. Desse modo, o acúmulo dessas tomadas de decisões pode indicar padrões de comportamento, e se seu incentivos forem parecidos, é possível tomar decisões para que o comportamento do publico mude.

Dentro de uma empresa real, várias métricas são importantíssimas para perceber quais melhores decisões a serem tomadas, reduzir despesas, aumentar investimento em propagandas ou contratar mais pessoas. Umas dessas métricas existentes é chamada de Churn, um indicador de quantos clientes deixaram de consumir o produto ofertado, essa variável é um número inteiro, quanto maior, mais clientes deixaram a empresa. Porém, apenas esse valor não nos permite ampliar nosso raio de visão sobre quais motivações fazem eles tomarem essa decisão. Desse modo, com outras variáveis disponibilizadas por um banco de dados, podemos cruzar informações e entender, quais delas tem correlação com o aumento de Churn ocorrido, para que possamos tomar uma decisão benéfica para a empresa e para os clientes

Os dados foram disponibilizados em uma dataframe chamado churn, no qual apresenta variáveis como ID do cliente, seu gênero, tipos de serviços contratados, tipo de renovação de contrato, tempo em contrato e claro, se o cliente realizou ou não o churn

**Proporção de Churn**

O primeiro passo, é saber qual é a proporção dos clientes que realizam churn ou não. Com poucas linhas de código, criamos uma tabela com a porcentagem de realizados.

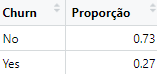


Tabela - Proporção de Churn realizados

O que podemos observar por enquanto é que, o valor apresentado em decimais, mostra que 27% deixam de consumir os produtos da empresa. Essa proporção é razoavelmente alta, mas considerável.

**Variáveis demográficas**

Nessa etapa, iniciaremos a análise dos churns comparados variáveis demográficas, como gênero e cliente Sênior. Com algumas linhas de código, plotamos um gráfico com a proporção de churns feitos por homens e mulheres. Gráfico apresentado abaixo.

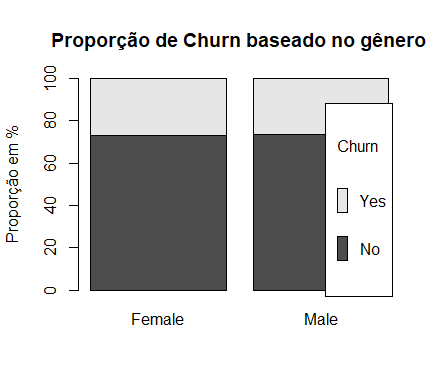


Gráfico - Proporção de Churn para cada gênero

O que podemos interpretar no gráfico é que, a proporção de churn para cada gênero não apresenta discrepâncias, ainda de ser também aproximado a proporção geral de churns feitos

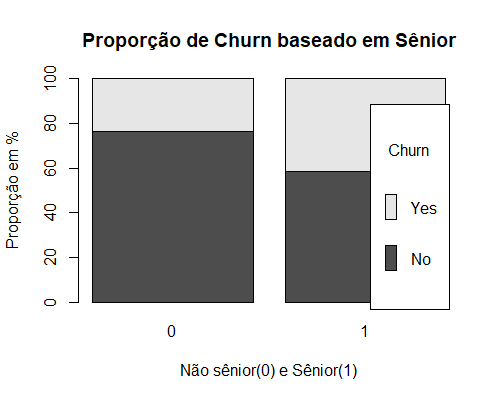


Gráfico - Proporção de churn para cliente Sênior ou não

Já a segunda variável demográfica apresentada, apresenta um diferença visivelmente considerável, podemos observar que Clientes Sênior tem maior propensão em deixa de consumir o produto

**Variáveis de Comportamento**

Enquanto as análises com variáveis demográficas podem nos indicar o padrão de cliente que realiza o churn, as de comportamento permitem observarmos que tipo de produtos o cliente consumia e quanto tempo ele consome os produtos da empresa. Essas variáveis podem mostrar motivações pelas quais, motivaram o churn.

A variável Tenure indica a quantidade de meses que, o cliente que consome o produto, está em vínculo com a empresa. Se compararmos medidas numéricas, os resultados de Tenure para clientes que realizaram churn ou não, podemos observar qual desses fica a mais tempo na empresa.

Tabela

Descrição gerada automaticamente

Tabela - Medidas numéricas de Tenure para Churn realizado ou não

Como podemos observar, clientes que continuam na empresa apresentam um Tenure maior na média, e próximo da mediana, indicando uma boa aproximação dos dados apesar de um desvio padrão maior. Os indicadores de máximo e mínimo continuam iguais para as duas condições, nos deixando poucas conclusões.

Podemos também observar o histograma de Tenure para churn realizado ou não.

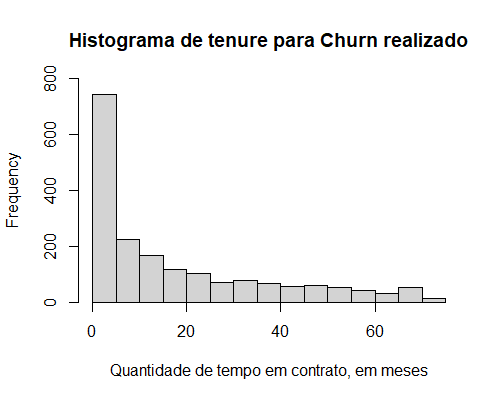


Gráfico - Histograma de Tenure para Churn realizado

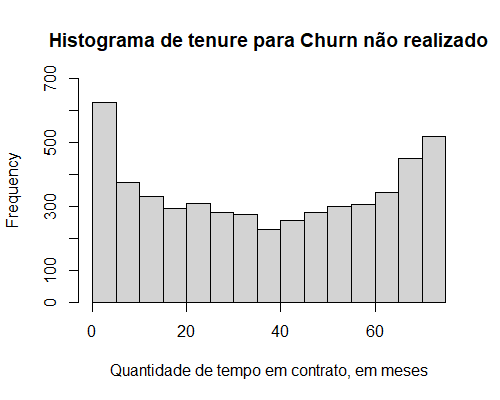


Gráfico - Histograma de Tenure para Churn não realizado

Esse gráfico já nos permite observar de outra forma os dados, o primeiro histograma uma grande assimétrica a direita, mostrando que a maioria dos casos de Churn ocorrem nos primeiros meses. Já no segundo gráfico, mostra-se um vale no meio, com duas grandes frequências nas extremidades. A percepção que temos é que existe uma entrada grande de clientes no começo e no fim das datas do dataframe, porém houve uma diminuição na metade desse tempo.

Outras variáveis que podemos observar também, é em relação a qual tipo de produto tem, o cliente que deixa de consumir o produto

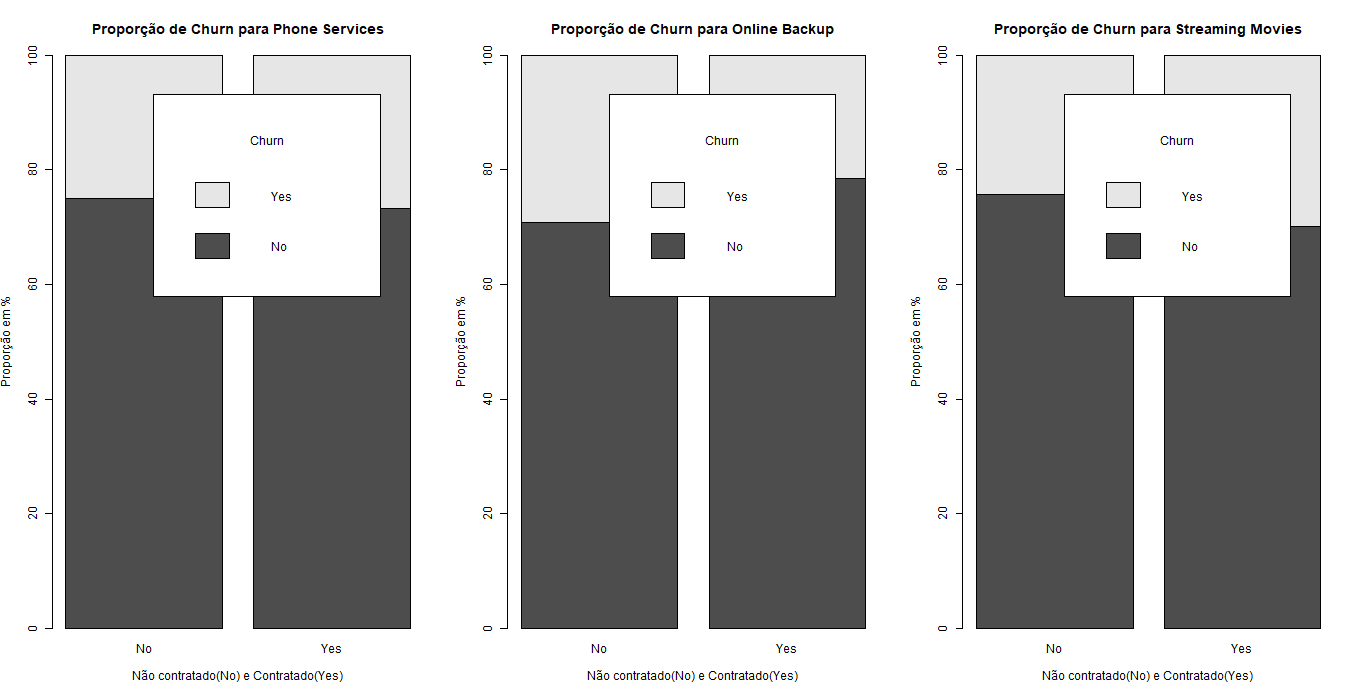


Gráfico , 6 e 7 - Proporção de Churn para tipos de produtos contratados(Phone Services, Online Backup, Streaming Movies)

Com a apresentação dessa proporção para os três tipos de produto, podemos observar que, a quantidade de pessoas que tem Online Backup e realizam churn é menor, podendo mostrar talvez, uma aceitação maior desse produto, em relação aos outros, pois eles apresentam churn maior para quem tem o produto contratado.

**Variáveis de Pagamento**

Alguma das variáveis que a base de dados já disponibiliza é o quanto cada cliente paga por mês total e a quantidade de serviços contratados, porém não apresenta quanto ele paga por serviço por mês, como se fosse um ticket médio. Por isso criaremos essa variável chamada *MCperQTY* a partir dos seguintes códigos:

Entrada:

“churn$MonthlyCharges = as.numeric(unlist(churn$MonthlyCharges))

churn$QtyServices = as.numeric((unlist(churn$QtyServices)))

churn$MCperQTY = cbind(churn$MonthlyCharges/churn$QtyServices)

colnames(churn$MCperQTY) = c(“MCperQTY”)”

Saída:

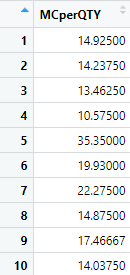


Tabela - Valores calculado de MCperQTY para os 10 primeiros índices

Essa nova variável é possível ser analisada em conjunto com outros valores. Através de medidas numéricas, vamos analisar seus valores para Churn realizado ou não.

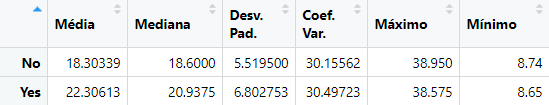


Tabela - Medidas numéricas de MCperQTY para Churn realizado ou não

Podemos perceber que, a diferença de MCperQTY para quem deixa de consumir o produto é na média do valor pago. Quem deixa de consumir é os que pagam mais por serviço por mês.

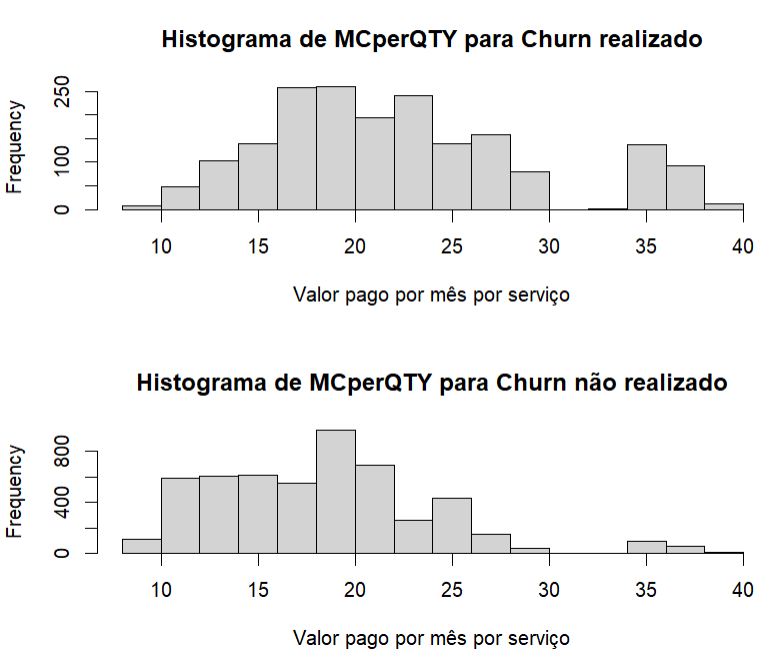


Gráfico 8 e 9 - Histograma de MCperQTY para Churn realizado ou não

Já na análise do histograma da variável criada, em relação ao Churn, podemos observar uma assimetria a direita maior para o segundo gráfico, mostrando a mesma conclusão anterior.

Outra variável que podemos analisar junto ao *MCperQTY* é o *contract*, que indica que tipo de pagamento o cliente faz, podendo ser mês a mês, a cada ano, ou a cada dois anos. Através do mesmo método, vamos analisar as medidas numéricas para MCperQTY para cada tipo de contrato.

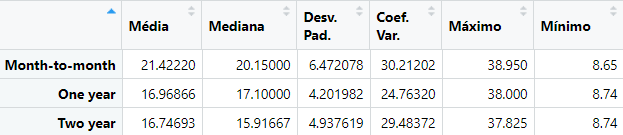


Tabela – Medidas numéricas do MCperQTY para cada tipo de contrato

Podemos observar que como mostrado pela média e mediana, quem paga mais são clientes com o tipo de contrato mês a mês, enquanto as outras medidas não apresentam muita diferença significante.

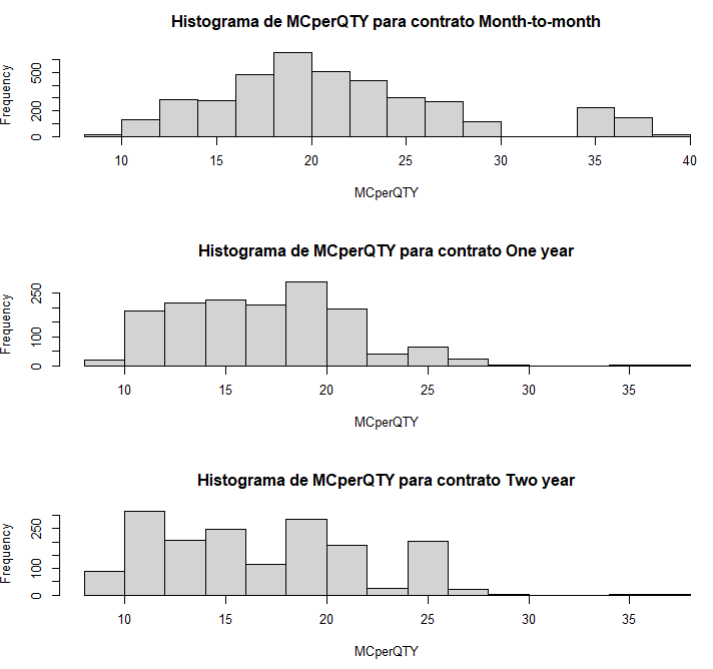


Tabela 10, 11 e 12 - Histograma de MCperQTY para cada tipo de contrato do cliente

Podemos observar com o Histograma o mesmo que com a tabela, com maior período de contrato, a assimetria a direita aumenta, deixando o valor pago por mês menor.

A primeira comparação permite observar qual valor pago por clientes que realizam o churn, enquanto a segunda comparação permite visualizar qual valor paga o cliente pelo tipo de contrato

Em conclusão, podemos traçar um perfil de cliente que realiza o churn, cliente que pagar um valor maior ao mês e considerado Sênior são mais propensos a deixarem de consumir os produtos, principalmente se estiver nos primeiros meses de contrato. Completando, outra possível forma de ver que tipo é mais propenso a realizar o churn, é observando que tipos de contratos realizam mais churn, mês a mês, ao ano ou a cada 2 anos.

