

# Análises de Dados e Big Data



**Sistemas de Informação / Ciências da Computação**

Aran Bey Tcholakian Morales, Dr. Eng.

Apostila 7: Análises descritiva e exploratória de dados

# Análises de Dados: Descritiva / Exploratória



# Análises Descritiva/Exploratória

## ANALYTICS

**BI**

What happened?

DESCRIPTIVE ANALYTICS



Why did it happen?

DIAGNOSTIC ANALYTICS



What will happen?

PREDICTIVE ANALYTICS



How can we make it happen?

PRESCRIPTIVE ANALYTICS



# Análises Descritiva/Exploratória



# Análises Descritiva/Exploratória

Análise de dados descritiva e análise de dados exploratória são duas abordagens diferentes para examinar e entender conjuntos de dados.

As duas técnicas possam ser usadas em conjunto, elas têm objetivos distintos.

# Análises Descritiva/Exploratória

A análise descritiva é usada para resumir e apresentar dados existentes por meio de tabelas, gráficos, medidas, enquanto a análise exploratória é uma abordagem mais abrangente que se concentra em descobrir padrões, tendências e relacionamentos que não são aparentes à primeira vista.

# Análises Descritiva/Exploratória

A análise exploratória inclui a utilização de gráficos, diagramas e outras ferramentas para visualizar os dados de uma forma que possa ser facilmente compreendida e ajudar a identificar hipóteses, insights e relacionamentos que possam exigir uma análise mais aprofundada.

# Análises Descritiva/Exploratória

**Análises descritivas** são as primeiras manipulações realizadas em um estudo quantitativo e consiste em **ORGANIZAR** e **RESUMIR** os dados coletados por meio de tabelas, gráficos ou medidas numéricas utilizando técnicas de **estatística descritiva e de visualização**.

A partir dos dados resumidos, identificar as **principais características dos dados**: medidas de tendência central, de ordenação, de dispersão, de correlação, para visualizar e encontrar padrões, relações, tendências e dependências entre os



# Análises Descritiva/Exploratória

O objetivo final é tornar mais clara a **descrição dos dados** a fim de ajudar o analista a desenvolver algumas hipóteses sobre o problema em questão e permitir a construção de modelos apropriados para tais dados, isto é auxiliar na **INTERPRETAÇÃO** dos dados.

As **Análises Descritivas** são um **pré-requisito** para uma análise de dados preditivas.

# Análises Descritiva/Exploratória

A **estatística descritiva** é um conjunto de técnicas que permite, de

forma sistemática, organizar, descrever, analisar e interpretar dados

por meio do uso de certas medidas-síntese que tornem possível a interpretação de resultados.

**Visualizar dados** é uma das técnicas mais importantes de descoberta e exploração de dados. A representação visual dos dados fornece fácil compreensão de dados complexos com vários

# Análises Descritiva/Exploratória

Os objetivos das **Análises Descritivas** são:

**A compreensão de dados:** fornece uma visão geral de cada atributo no conjunto de dados e interação entre os atributos;

Auxilia nas respostas a perguntas como qual é o valor típico de um atributo e se existem outliers no conjunto de dados.

Histogramas, por exemplo, permitem a compreensão da distribuição de um atributo.

# Análises Descritiva/Exploratória

Os objetivos das **Análises Descritivas** são:

**Preparação de dados:** Antes de aplicar outros algoritmos, precisamos preparar o conjunto de dados para a manipulação de: anomalias, valores atípicos, valores em falta, e remoção de instancias duplicadas ou atributos altamente correlacionados.

# Análises Descritiva/Exploratória

Os objetivos das **Análises Descritivas** são:

**Modelos analíticos:** podem auxiliar nos modelos de análises preditivas de dados. Por exemplo, gráficos de dispersão podem identificar os clusters, que podem auxiliar a desenvolver modelos de regressão ou de classificação.

**Interpretação dos resultados:** é usada para entender os resultados das técnicas analíticas como a previsão, como a classificação e o resultados de agrupamento.

# Análises Descritiva/Exploratória



Iris Setosa



Iris Versicolor



Iris Verginica

A base de dados Iris foi introduzida pelo estatístico e biólogo inglês Ronald Fisher, em 1936, como motivação para a técnica multivariada de análise discriminante linear;

Os dados consistem de 50 amostras de três espécies de Flores Iris e quatro medidas de cada flor: comprimento e largura da pétala e da sépala;

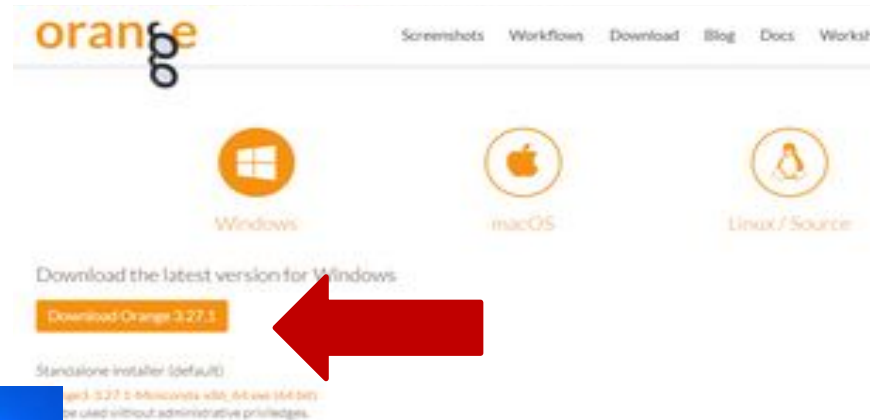
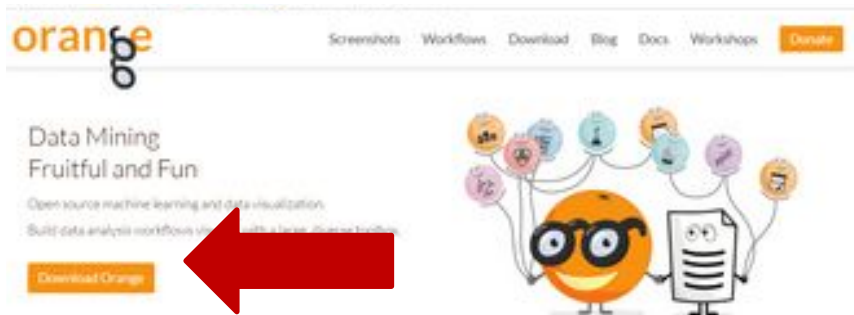
# Análises Descritiva/Exploratória

Temos a tarefa de criar um modelo de aprendizagem de máquina que com quatro medições das flores (comprimento sépala, largura sépala, comprimento pétala, e largura pétala), identifica (**classifica**) a qual das 3 espécies de Flores Iris pertence.

sepal.length	sepal.width	petal.length	petal.width	species
5.5	2.4	3.7	1.0	versicolor
5.5	2.5	4.0	1.3	versicolor
5.1	3.3	1.7	0.5	setosa
6.8	2.8	4.8	1.4	versicolor
6.7	3.3	5.7	2.5	virginica



# Análises Descritiva/Exploratória





# Análises Descritiva/Exploratória

Mover o “widgets” “File” para a área de trabalho e carregamos o arquivo “DadosIris.xls”. O atributo “TipoIris” deve ser o “target”, o atributo “ID”, posso “skip”

The screenshot shows the Orange3 software interface. On the left, the 'Data' widget is circled in red. On the right, the 'File' widget is open, showing the file 'DadosIris.xls' and a table of columns with their roles. The 'TipoIris' column is set as the 'target' and the 'ID' column is set as 'skip'.

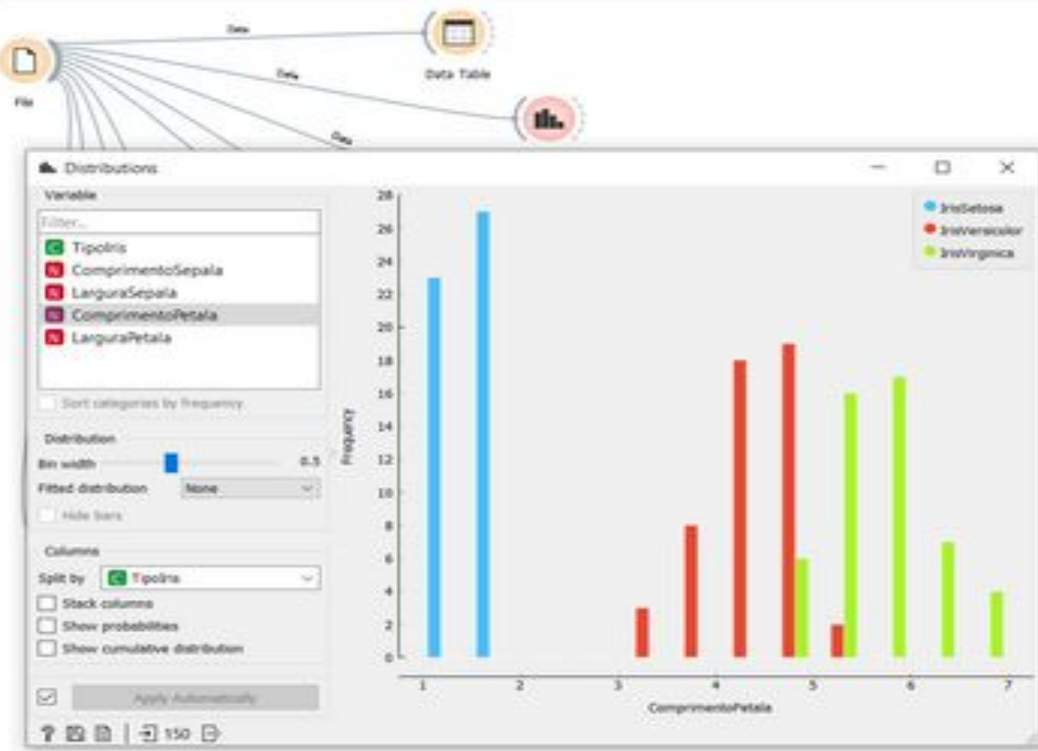
Name	Type	Role	Values
1 ComprimentoS...	N numeric	feature	
2 LarguraSepala	N numeric	feature	
3 ComprimentoPe...	N numeric	feature	
4 LarguraPetala	N numeric	feature	
5 <b>TipoIris</b>	C categorical	target	IrisSetosa,
6 ID	S text	feature	

# Análises Descritiva/Exploratória

Vamos visualizar os dados com "Data Table".



# Análises Descritiva/Exploratória



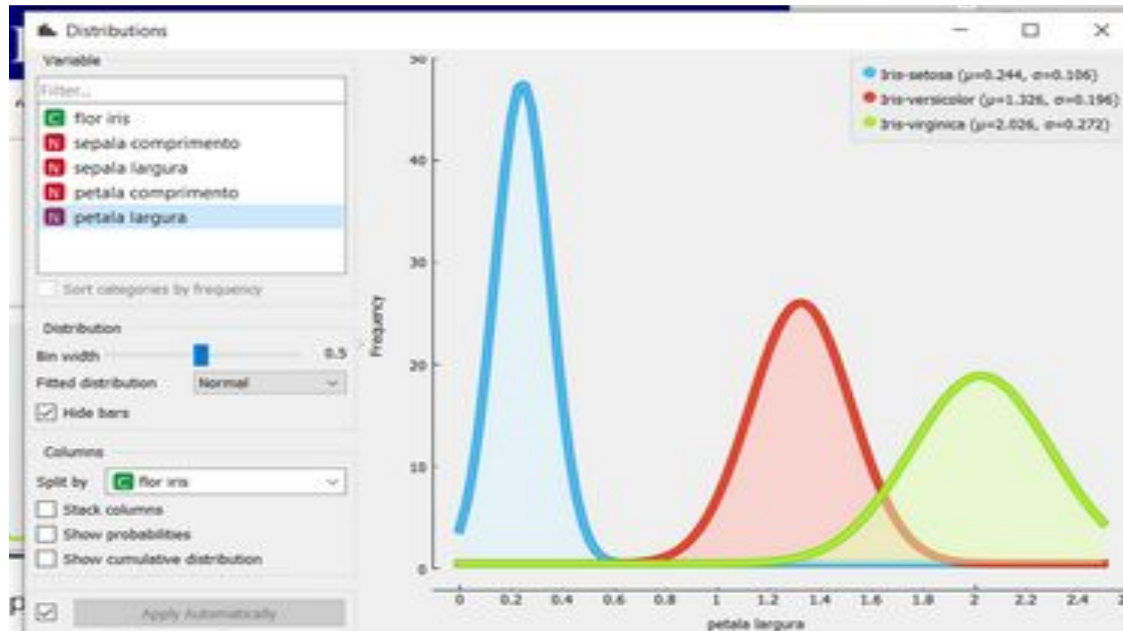
Vamos visualizar os dados com o “widget”, “Distributions”

Um histograma é uma das maneiras visuais para compreender a frequência de ocorrência de um intervalo de valores para uma variável.

Observa-se que na variável “comprimento da pétala”, existe uma clara distinção entre o tipo Setosa e os outros dois tipos de flores Iris.

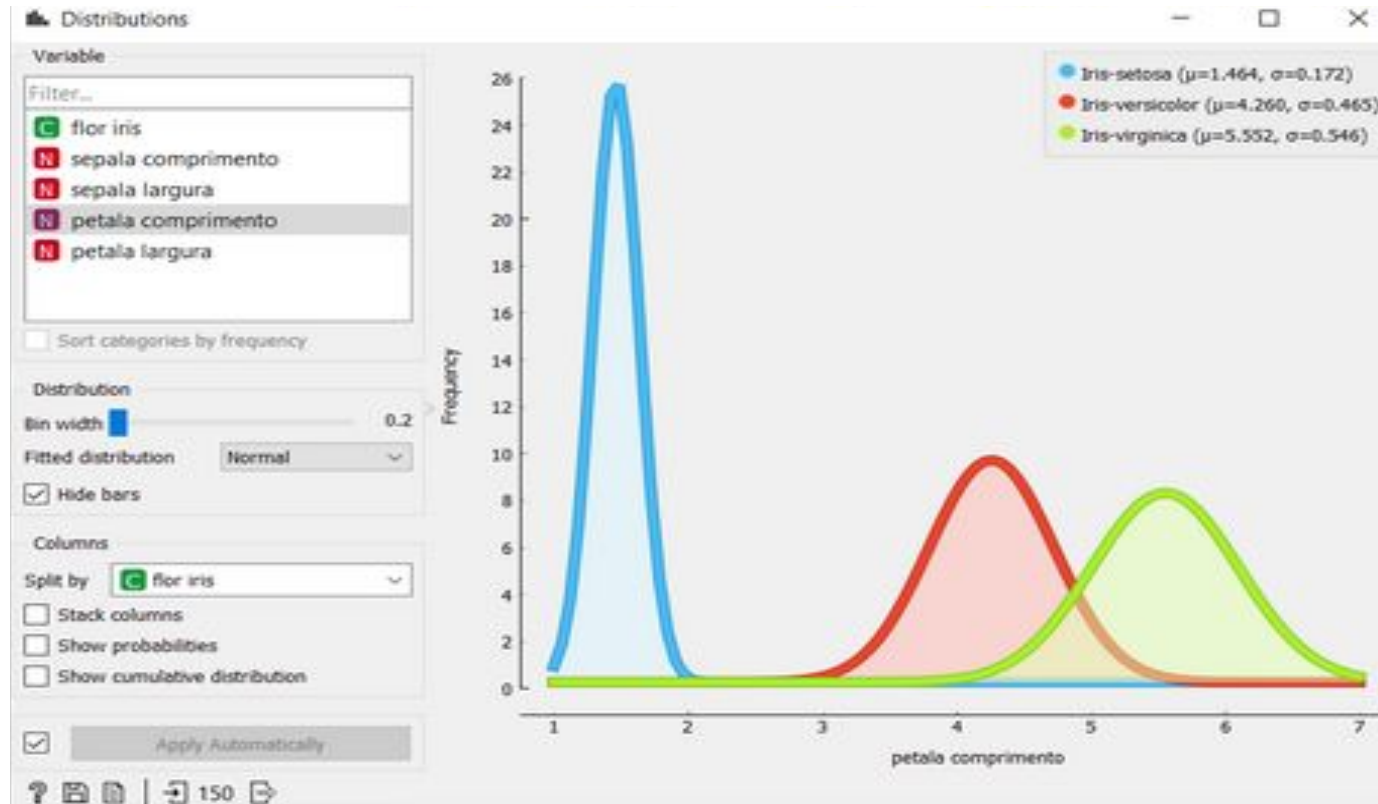
# Análises Descritiva/Exploratória

A partir do gráfico de distribuição podemos inferir que o comprimento pétala, por exemplo Iris Setosa é mais coeso e diferente do que Iris Versicolor e Iris Virginica.



A função de distribuição conta a probabilidade de ocorrência dos dados dentro de um intervalo. Se um conjunto de dados exibe distribuição normal, 68,2% de pontos de dados estão a um desvio padrão da média, 95,4% dos pontos caem dentro  $2\sigma$  e 99,7% dentro de  $3\sigma$  da média.

# Análises Descritiva/Exploratória



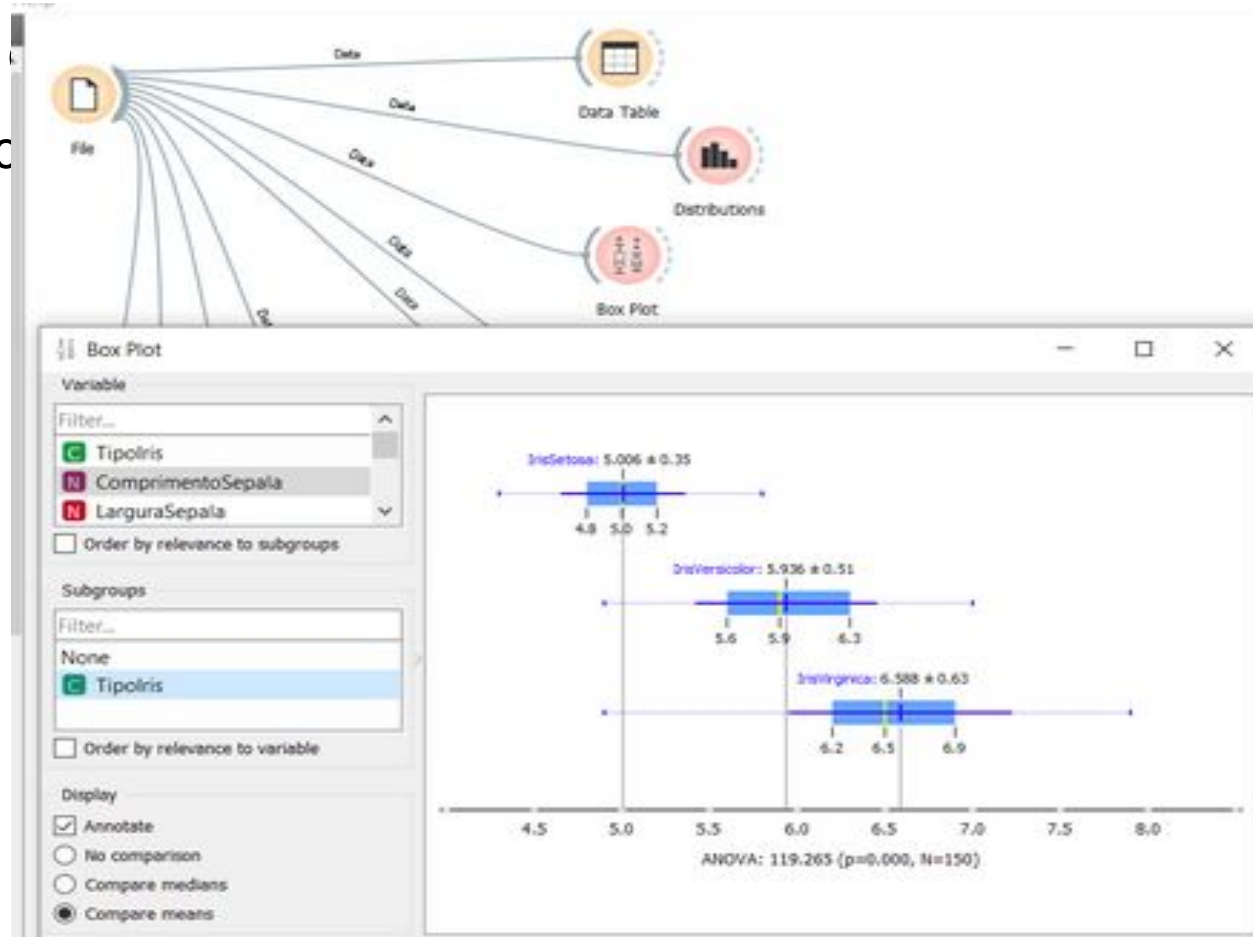
No gráfico de distribuição, pode-se inferir que o comprimento da pétala para a espécie Setosa, é mais distinta e coesa do que as outras espécies.

Para um comprimento de pétala de 1,5 cm, pode-se prever que a espécie seja Setosa, no entanto, para uma medida do comprimento da pétala de 5,0 cm, não há previsão clara, pois a espécie pode ser Versicolor ou Virgínica.

# Análises Descritiva/Exploratória

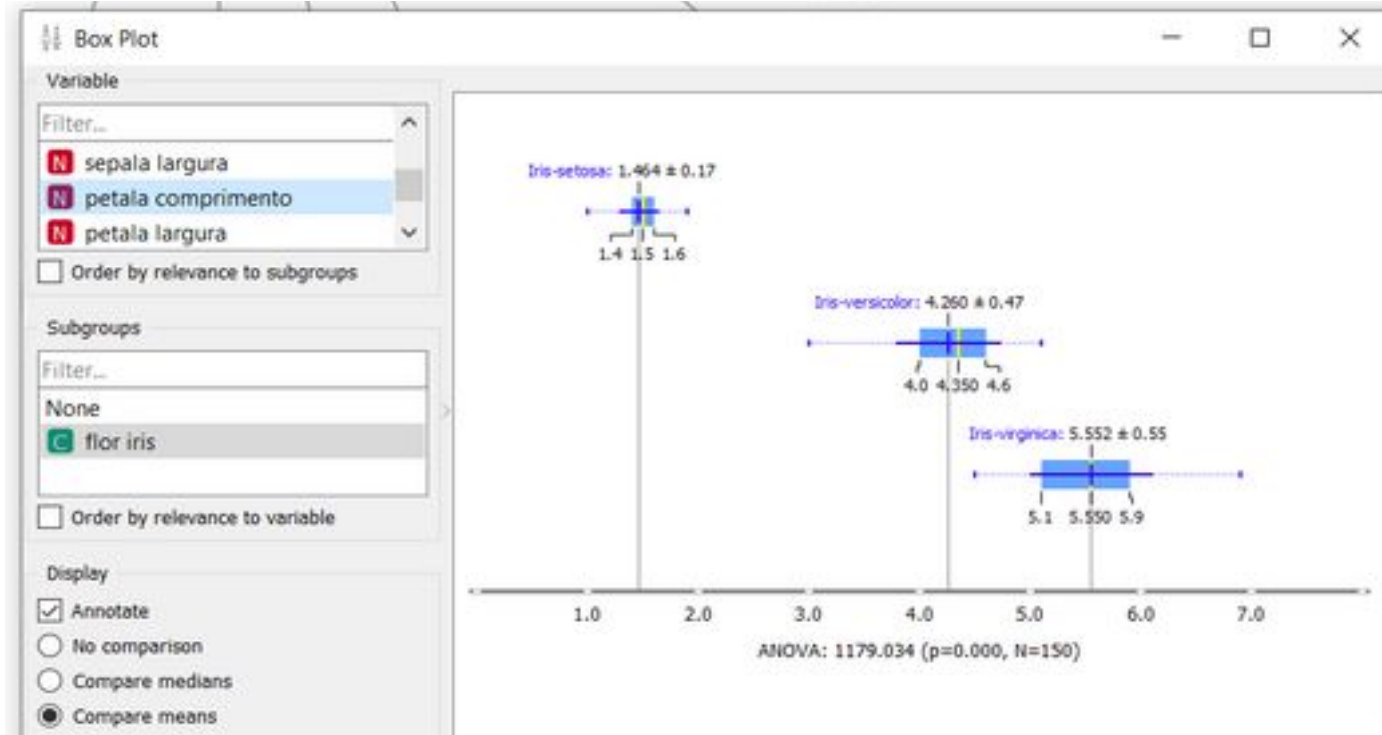
Vamos visualizar os dados com "Box Plot". A linha azul vertical é a "média" e a amarela a "mediana". Os valores da área azul, são os valores entre o 1 e 3 quartil.

Os extremos, são o menor e o maior valor, representa o desvio padrão



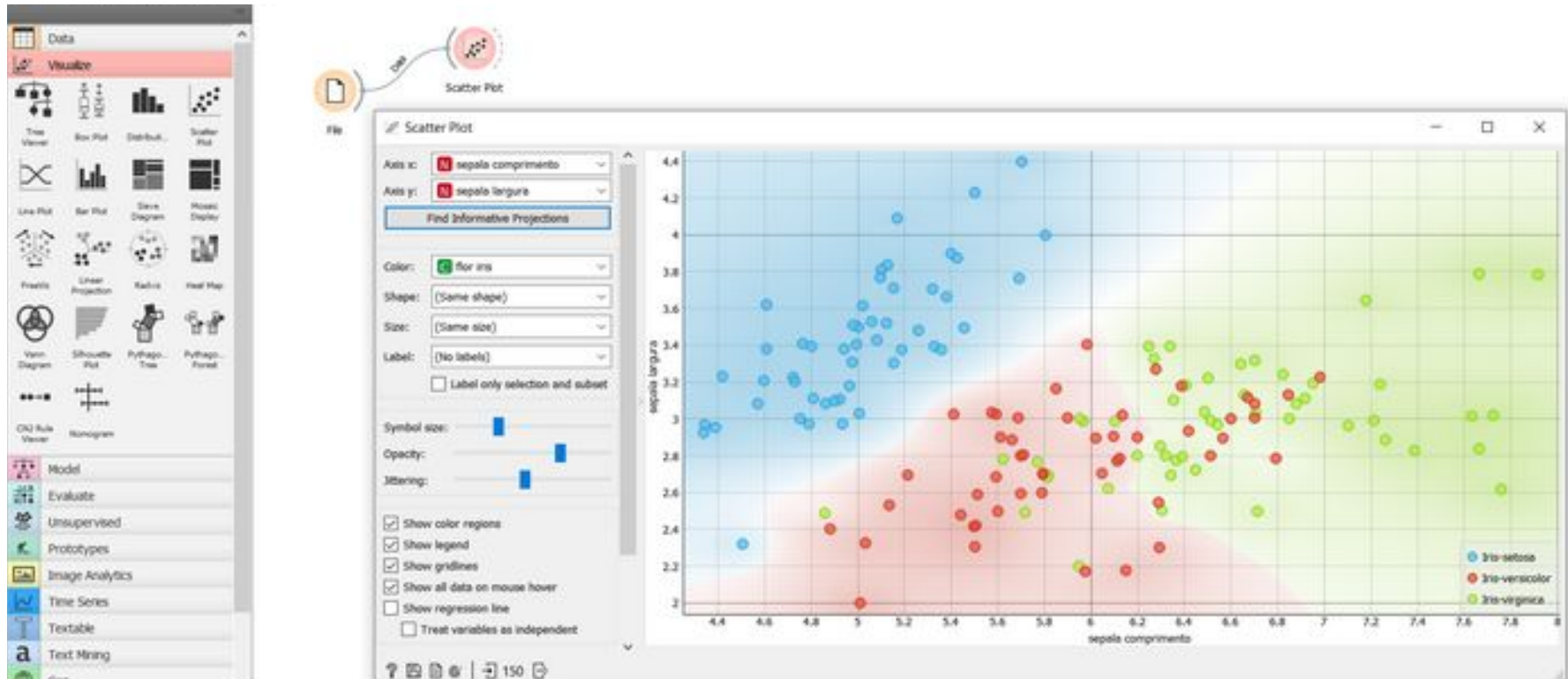
# Análises Descritiva/Exploratória

Observamos mais uma vez a diferença da variável “comprimento da pétala” da espécie Iris Setosa, quando comparadas as outras espécies.



# Análises Descritiva/Exploratória

Vamos visualizar os dados com “Scatter Plot” (gráfico de dispersão).



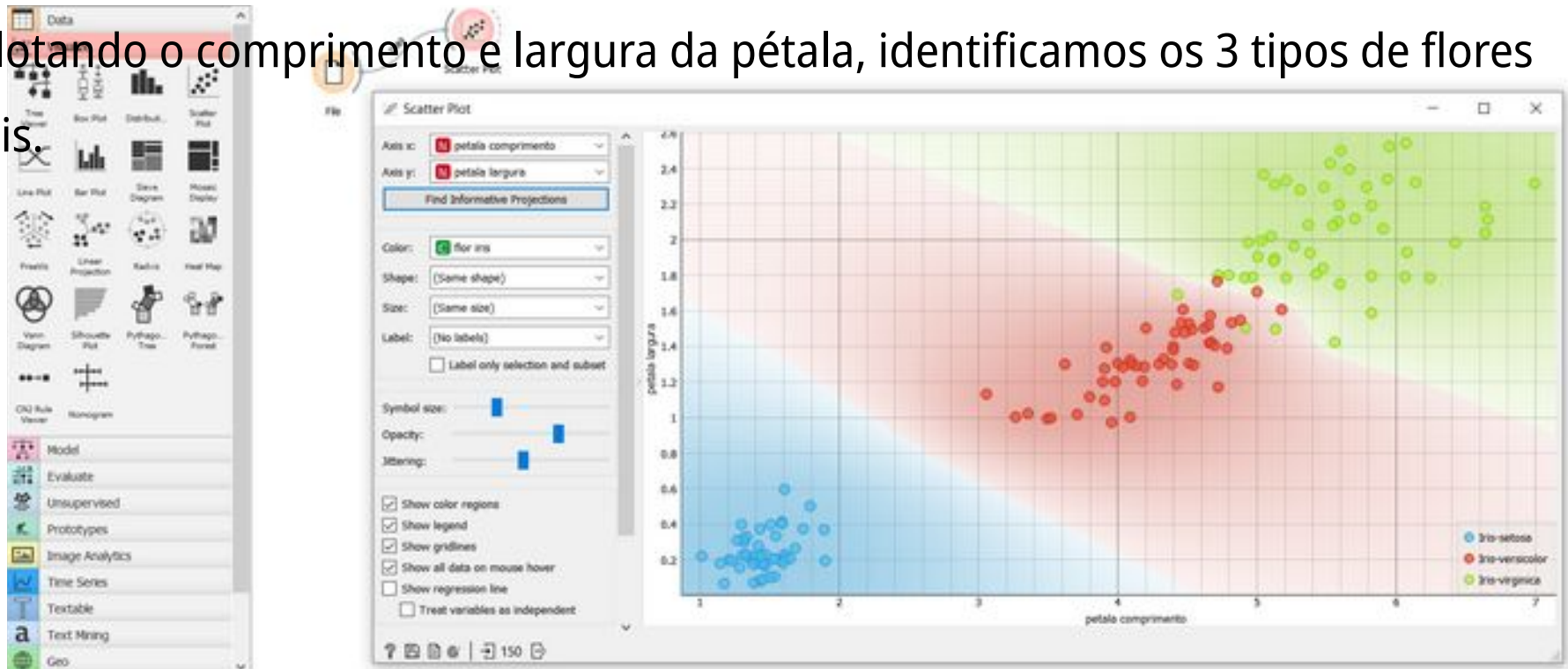
Nesta visualização, podemos observar a dispersão entre os tipos de flores Iris. Observa-se “claramente” a diferença entre o tipo Iris Setosa, e os outros tipos de flores Iris.



# Análises Descritiva/Exploratória

Vamos trocar as variáveis de visualização, para tentar identificar os 3 tipos de flores Iris.

Plotando o comprimento e largura da pétala, identificamos os 3 tipos de flores Iris.



# Análises Descritiva/Exploratória

O "widget", "Feature Statistics", apresenta a distribuição das variáveis e algumas medidas de cada variável.

The screenshot displays the Orange3 software interface. On the left is a 'Data' widget menu with various data handling options. In the center, a workflow diagram shows a 'File' widget connected to 'Data Table', 'Scatter Plot', 'Distributions', 'Box Plot', and 'Feature Statistics'. The 'Feature Statistics' widget is highlighted, showing its configuration and output.

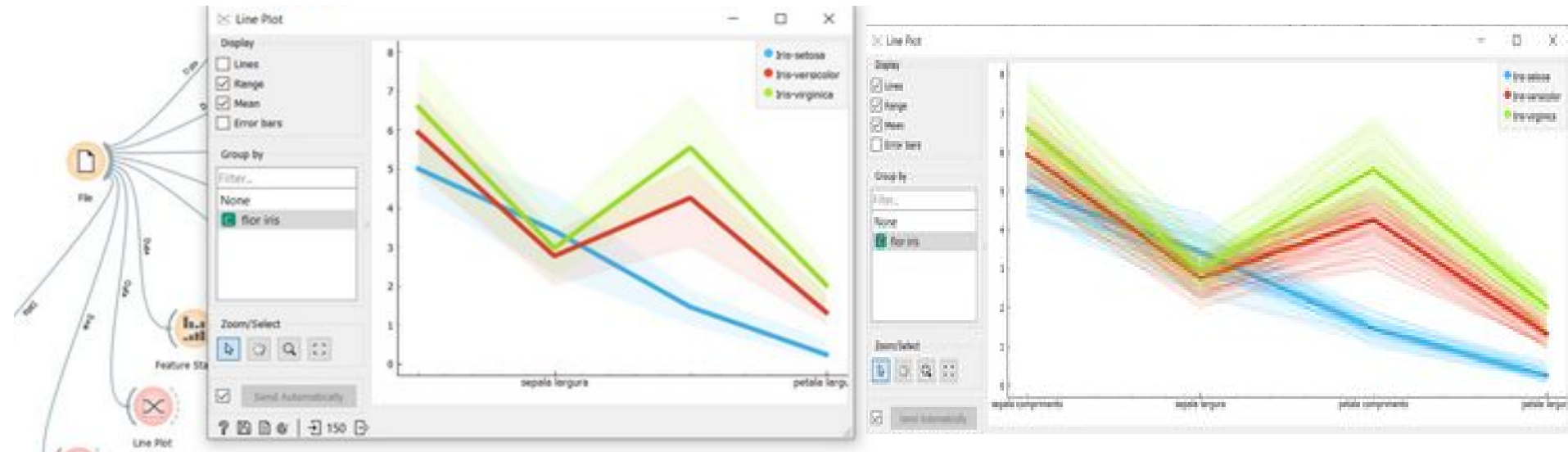
The 'Feature Statistics' widget window shows a histogram for the 'flor iris' variable. The output table is as follows:

Name	Distribution	Center	Dispersion
petala largura	[Histogram]	1.199	0.633
petala comprimento	[Histogram]	3.759	0.468
sepala largura	[Histogram]	3.054	0.142
sepala comprimento	[Histogram]	5.843	0.141
flor iris	[Histogram]	Iris-setosa	1.1

The 'Data Table (1)' widget window shows the following summary statistics:

Feature	Center	Dispersion	Min.	Max.	Missing
1 petala largura	1.19867	0.634549	0.1	2.5	0
2 petala comprim...	3.75867	0.46786	1	6.9	0
3 sepala largura	3.054	0.141502	2	4.4	0
4 sepala compri...	5.84333	0.141238	4.3	7.9	0
5 flor iris	0	1.09861	0	2	0

# Análises Descritiva/Exploratória



Notamos no gráfico de “Line Plot” que há uma sobreposição entre as três espécies do atributo de largura sépala, desta forma, esse atributo não pode ser a métrica utilizada para diferenciar as três espécies.

No entanto, existe uma clara separação das espécies para o comprimento pétala.

# Análises Descritiva/Exploratória

Colab, que é interpretador de python da Google, essa ferramenta em nuvem não requer nenhuma instalação no PC. Nesse sentido, o Colab usa a forma de “notebooks” que são interpretadores de arquivos em python com a extensão .ipynb, possibilitando também adicionar comentários e anotações a cada novo comando em python. <https://colab.research.google.com/>

Flores\_Iris\_Exploratoria\_Descritiva\_2023.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

**Análises visual, descritiva e exploratória de dados utilizando o conjunto de dados das Flores Iris**

O conjunto de dados inclui três espécies de iris com 50 amostras cada, bem como algumas propriedades sobre cada flor.

Neste exemplo, veremos como realizamos uma análises exploratória / descritiva antes da realização de uma análises preditiva (com a aplicação de algoritmos de machine learning). Vamos passar pelo seguinte, passo a passo:

- [Lendo e entendendo os dados](#)
- [Visualização de dados e análises exploratória / descritiva](#)

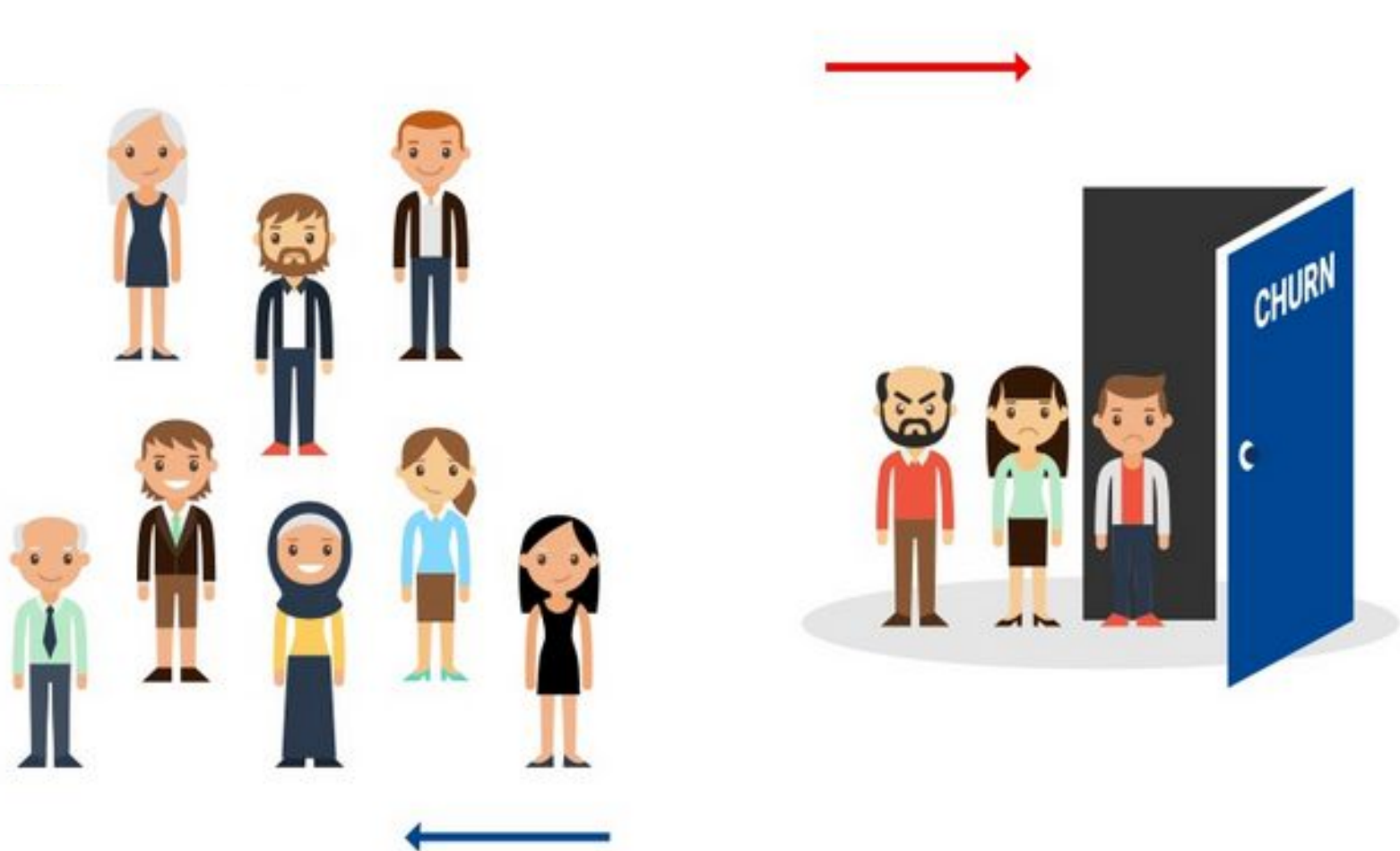
▶ **Lendo e entendendo os dados**

[ ] 49 células ocultas

▶ **Visualização de dados e análises exploratória / descritiva de dados**

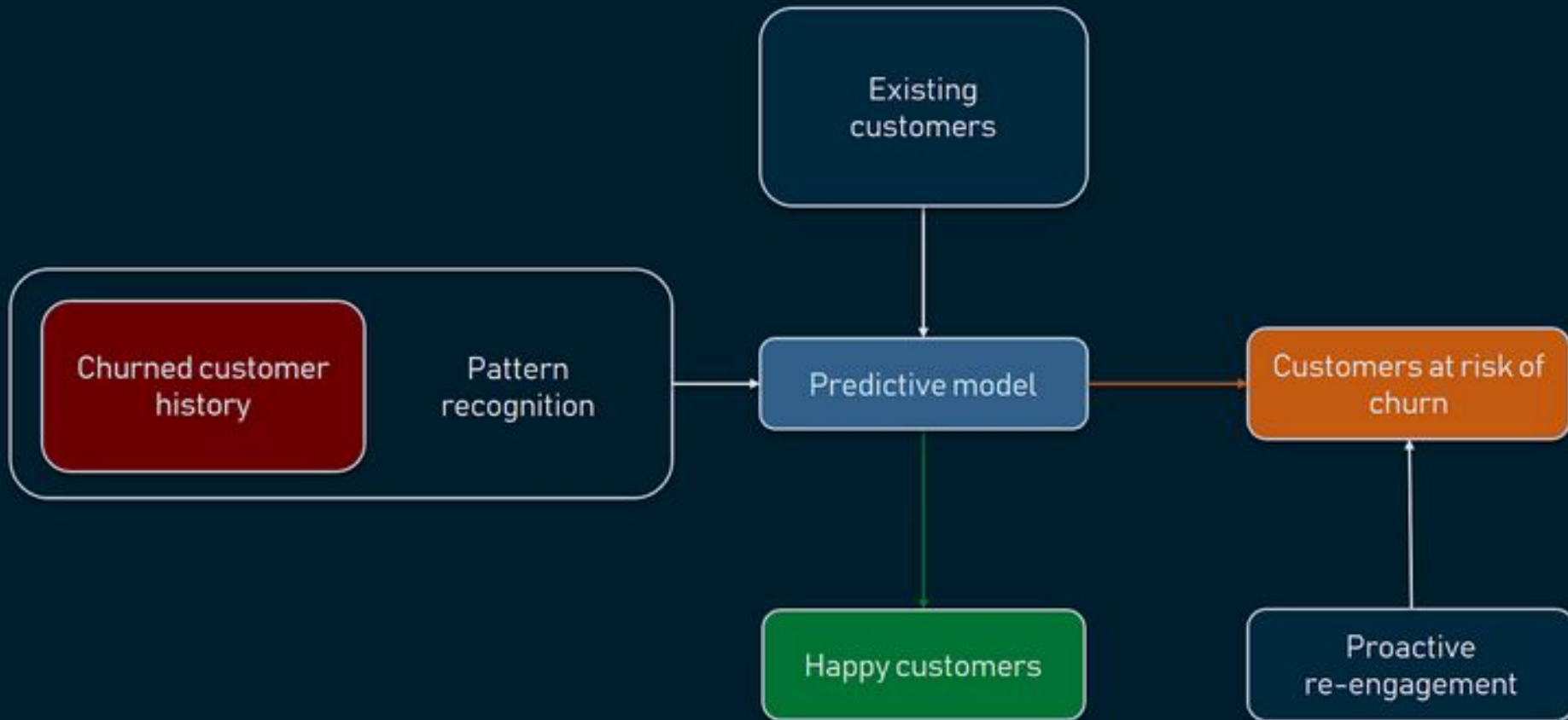
[ ] 49 células ocultas

# Rotatividade (churn) de Clientes



# Rotatividade (churn) de Clientes

## CHURN RATE PREDICTION WITH MACHINE LEARNING



# Rotatividade (churn) de Clientes

A **retenção de clientes** é um dos principais pilares de crescimento para produtos com um **modelo de negócios** baseado em assinatura.

Neste modelo de negócios, os clientes são livres para escolher entre muitos fornecedores, e se a quantidade de clientes insatisfeitos crescer, tanto as perdas materiais quanto os danos à reputação são enormes.

É extremamente útil para as empresas identificar e visualizar quais fatores contribuem para a perda de clientes.

Se as empresas puderem identificar os clientes que deixaram de utilizar seus serviços e os prováveis fatores que podem estar levando-os à desistência, eles podem então criar estratégias de marketing e retenção adequadas para reter os clientes.

Por exemplo, eles podem oferecer aos clientes ofertas em seus serviços ou criar planos de serviços mais condicentes com o perfil destes clientes.

# Rotatividade (churn) de Clientes

## Casos de uso para previsão de rotatividade de clientes

- **Os serviços de streaming de música e vídeo** são provavelmente os mais comumente associados ao modelo de negócios de assinatura (Netflix, Apple Music, YouTube, Spotify, Hulu, Amazon Video, Deezer, etc.).
- **Meios de comunicação.** Presença digital é uma necessidade para a imprensa, sendo assim as empresas de notícias oferecem aos leitores assinaturas digitais além das impressas (*Financial Times, The New York Times, Folha, O Globo, etc.*).
- **Empresas de telecomunicações:** empresas que podem fornecer vários produtos e serviços, como rede sem fio, internet, TV, telefone celular e telefone residencial (Vivo, Tim, Claro, AT & T, Verizon, etc.).
- **Software como prestadores de serviços.** A adoção de software hospedado na nuvem está crescendo. Segundo o [Gartner](#), o mercado de SaaS continua sendo o maior segmento do mercado de nuvem. A gama de produtos de provedores SaaS é extensa: edição gráfica e de vídeo (Adobe Creative Cloud), contabilidade (Sage 50cloud, FreshBooks), eCommerce (BigCommerce, Shopify), email marketing (MailChimp, Zoho Campaigns), e muitos outros.



# Rotatividade (churn) de Clientes

As empresas precisam responder a perguntas importantes sobre a saúde do relacionamento com o cliente. As respostas a essas perguntas as ajudam a entender o que precisam fazer para garantir que cada cliente receba valor das soluções e permaneça um cliente valioso e fiel.

## **1. Qual é a nossa taxa atual de rotatividade de clientes?**

O cálculo da taxa de rotatividade é um primeiro passo para entender a saúde do relacionamento com o cliente e o que você precisa fazer para melhorar (ou manter) essa saúde. Calcular sua taxa de rotatividade atual ajuda a entender se você tem um problema de rotatividade. Além disso, você pode observar as tendências históricas para avaliar se sua taxa de rotatividade atual melhorou ou piorou. Por fim, para saber se seus esforços de gerenciamento de rotatividade são efetivos amanhã, você precisa saber onde está sua rotatividade hoje.

# Rotatividade (churn) de Clientes

## 2. Quais são as principais causas de rotatividade de clientes?

Para corrigir um problema de rotatividade de clientes, primeiro você precisa identificar os motivos pelos quais seus clientes estão deixando você.

**A Ciência dos Dados, usando dados históricos, constroem modelos para prever a rotatividade de clientes.** Esses modelos analíticos ajudam a quantificar o grau em que diferentes fatores são responsáveis pela rotatividade de clientes.

Esses fatores podem incluir:

- 1) como os clientes usam o produto;
- 2) interações dos clientes com a empresa - via suporte por exemplo;
- 3) dados “demográficos” dos clientes;
- 4) opinião do cliente, entre outros.

# Rotatividade (churn) de Clientes

## **3. Quais clientes específicos estão em risco de deixar o serviço ou produto?**

Responder a essa pergunta ajuda as empresas a adotarem uma abordagem proativa para lidar com a rotatividade de clientes.

A Ciência dos Dados aplicam o modelo (pergunta 2) a novos clientes para classificar / identificar clientes com probabilidade de “rotatividade” no futuro.

O modelo usa as informações de cada cliente para classificar cada cliente em um de dois grupos: 1) com probabilidade de permanecer e 2) com probabilidade de deixar de usar os serviços.

**As empresas podem usar essas informações para “chegar” aos clientes em risco e salvar esses relacionamentos fazendo uma ação pro-ativa.**

# Rotatividade (churn) de Clientes

## 4. Como devemos segmentar nossos clientes?

Tradicionalmente, as empresas segmentam seus clientes com base em informações demográficas, incluindo localização geográfica, idade e sexo, para citar alguns.

A Ciência dos Dados podem aproveitar os dados existentes dos clientes para criar segmentos de clientes com base em seu valor para a empresa.

Ao identificar clientes de alto valor, **as empresas podem desenhar estratégias para atrair novos clientes que tenham as mesmas características dos seus clientes atuais mais valorizados e ainda elevando as taxas de retenção.**

# Rotatividade (churn) de Clientes

## 5. Qual campanha de marketing é mais eficaz na redução da rotatividade?

O marketing e o relacionamento com os clientes, podem ser usados para atrair novos clientes ou salvar / incentivar os relacionamentos existentes. Após a criação de segmentos de clientes ou após a identificação dos clientes em risco, o próximo passo é procurá-los proativamente para atrair tipos específicos de clientes ou salvar relacionamentos existentes, respectivamente. Diferentes campanhas de marketing podem ser testadas entre si para identificar qual delas resulta nas menores taxas de rotatividade.

# Rotatividade (churn) de Clientes

Um **modelo preditivo para prevenir a rotatividade de clientes** é uma ferramenta que define as etapas e fases de um cliente deixar o seu serviço ou produto.

Ter um **modelo preditivo** permite ter consciência e métricas quantificáveis para definir estratégias e esforços de retenção.

Sem uma ferramenta analítica, você estaria agindo em suposições amplas e de forma “ad-hoc”, um modelo baseado em dados dão indícios de como seus clientes agem. Isso lhe dá a capacidade verificar padrões de comportamento e hábitos de clientes que saem, e se antecipar aos clientes antes que ele deixe de usar seu produto ou serviço. Uma abordagem proativa à rotatividade de clientes depende necessariamente da análise de dados.

# Rotatividade (churn) de Clientes

A Telco Comunicações, empresa de telecomunicações deseja identificar os atributos que tem uma maior influencia na rotatividade de seus clientes.

Para isso, entregou a seus Cientistas de Dados uma amostra com mais de 7 mil clientes (arquivo Rotatividade\_Clientes\_Telco.csv), com informações referentes as informações demográficas do cliente, dos serviços usados pelo cliente e informações sobre o contrato e conta do cliente.

# Rotatividade (churn) de Clientes

## **Informações demográficas do cliente:**

ID\_Consumidor; Genero; Idoso (mais idade); Parceiro; Dependentes;

## **Informações sobre os serviços do cliente:**

Servico\_Telefone; Muitas\_Linhas\_Telefone; Servico\_Internet; Seguranca\_OnLine;  
Backup\_OnLine; Dispositivo\_Protecao; Suporte\_Tecnico; Streaming\_TV;  
Streaming\_Filmes;

## **Informações sobre a conta do cliente:**

Nro\_Meses\_Cliente\_na\_Empresa; Contrato; Faturamento\_Sem\_Papel;  
Metodo\_Pagamento; Valor\_Mensal; Valor\_Total Rotatividade.



# Rotatividade (churn) de Clientes

## **As seguintes hipóteses podem ser feitas:**

Quanto maior a duração do contrato, mais provável o cliente ficar na empresa;

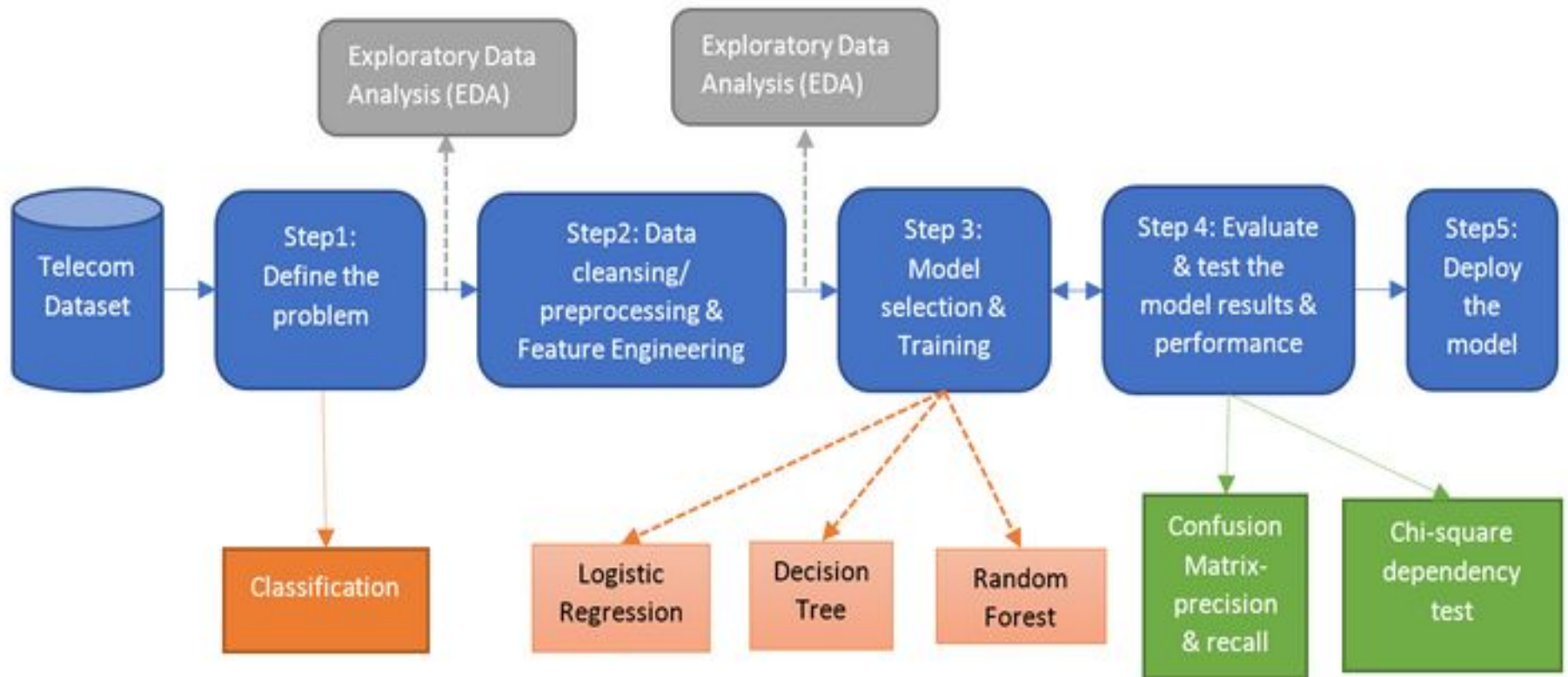
Clientes com mais serviços, são mais propensos a ficar na empresa;

Clientes com cônjuges e filhos podem migrar menos para manter os serviços funcionando para sua família.

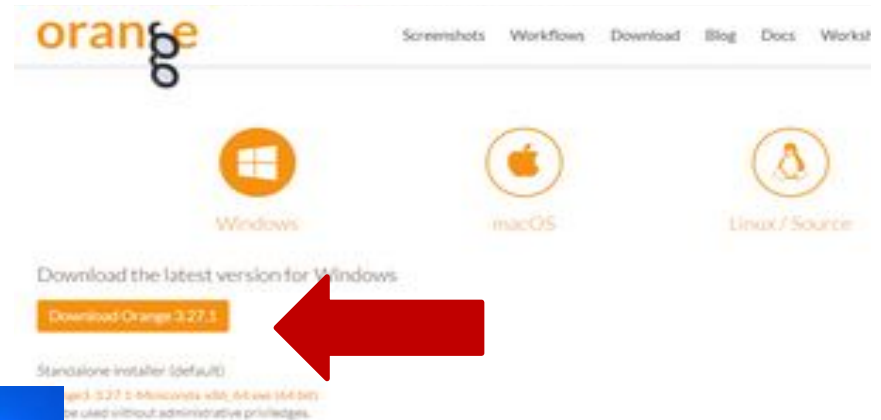
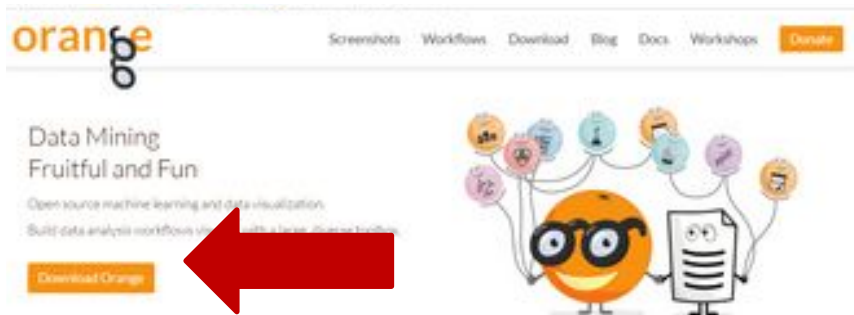
Contratos mais caros levam ao aumento da rotatividade, pois as chances de economizar dinheiro mudando de fornecedor podem ser maiores.

Os idosos tendem a diminuir a rotatividade "devido ao esforço" associado à rescisão de contratos.

# Rotatividade (churn) de Clientes

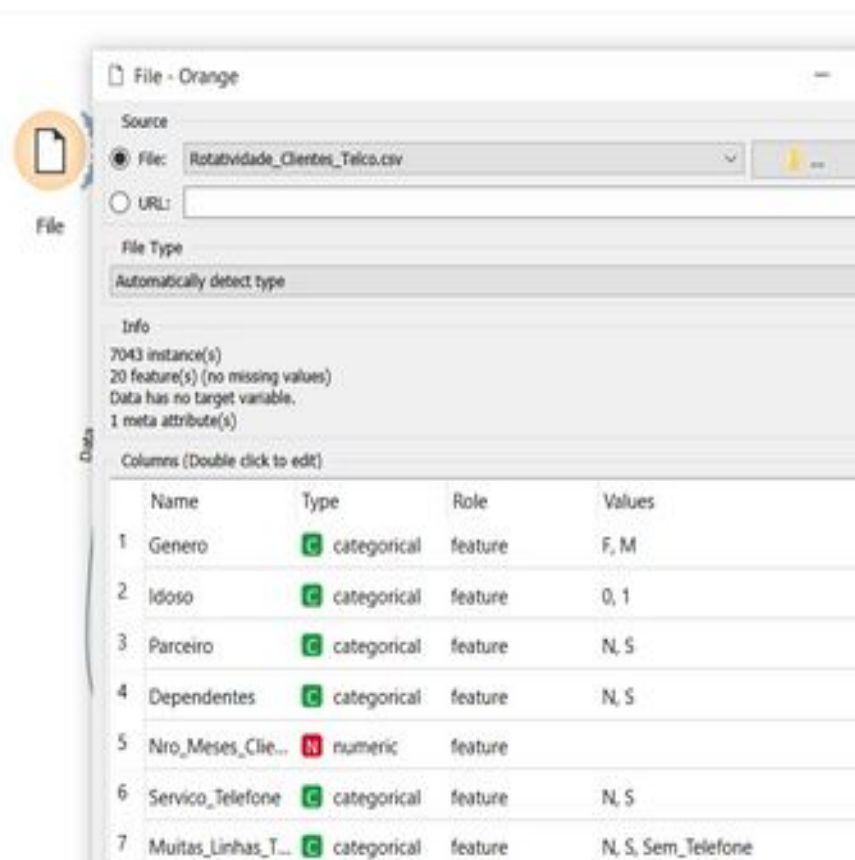
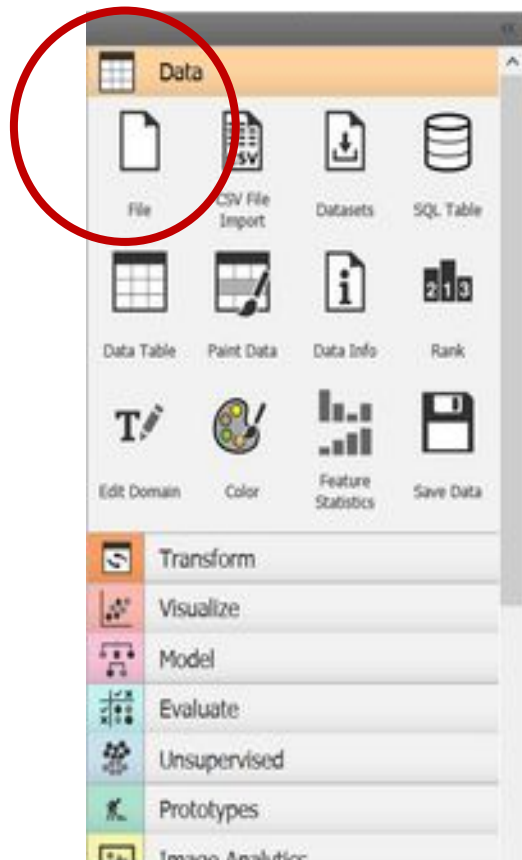


# Rotatividade (churn) de Clientes



# Rotatividade (churn) de Clientes

Mover o “widgets” “File” para a área de trabalho e carregamos o arquivo “Rotatividade\_Clientes\_Telco.csv”. O atributo “Rotatividade” deve ser o “target”, o atributo “ID\_Consumidor”, posso “skip”



# Rotatividade (churn) de Clientes

Vamos visualizar os dados com "Data Table".

The screenshot shows the Orange Data Mining interface. On the left, the 'Data' widget palette is visible, with the 'Data Table' widget icon circled in red. The main workspace displays a 'Data Table (1) - Orange' widget. The widget's configuration panel on the left shows the following settings:

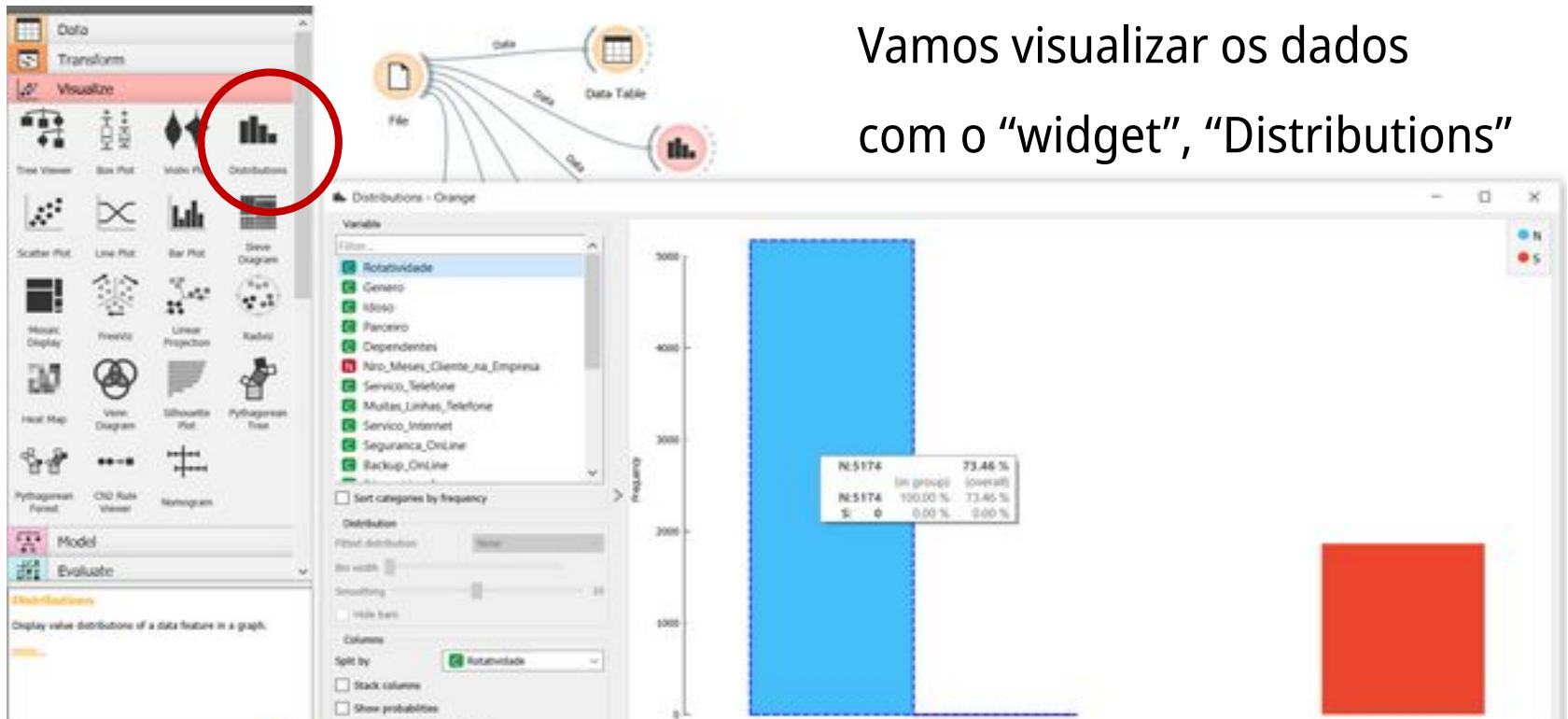
- Info: 7943 instances (no missing data), 19 features, Target with 2 values, 1 meta attribute.
- Variables:  Show variable labels (if present),  Visualize numeric values,  Color by instance classes.
- Selection:  Select full rows.

The main table displays the following data:

	Rotatividade	ID_Consumidor	Genero	Idoso	Pareceiro	Dependentes	meses_Cliente_na_Em	Servico_Telefone
1	N	7590-VHVEG	F	0	S	N	1	N
2	N	5575-GNVDE	M	0	N	N	34	S
3	S	3668-QPYBK	M	0	N	N	2	S
4	N	7795-CFOOW	M	0	N	N	45	N
5	S	9237-HQITU	F	0	N	N	2	S
6	S	9305-CDSKC	F	0	N	N	8	S
7	N	1452-KIOVK	M	0	N	S	22	S
8	N	6713-OKQMC	F	0	N	N	10	N
9	S	7892-POCKP	F	0	S	N	28	S
10	N	6388-TABGU	M	0	N	S	62	S
11	N	9763-GRSKD	M	0	S	S	13	S
12	N	7469-LKBCI	M	0	N	N	16	S
13	N	8091-TTVAX	M	0	S	N	58	S
14	S	0280-XJGEX	M	0	N	N	49	S
15	N	5129-ILPIS	M	0	N	N	25	S
16	N	3655-SNOYZ	F	0	S	S	69	S
17	N	8191-XWSZG	F	0	N	N	52	S
18	N	9959-WOFKT	M	0	N	S	71	S
19	S	4190-MFLUW	F	0	S	S	10	S
20	N	4181-MVFRB	F	0	N	N	21	C

# Rotatividade (churn) de Clientes

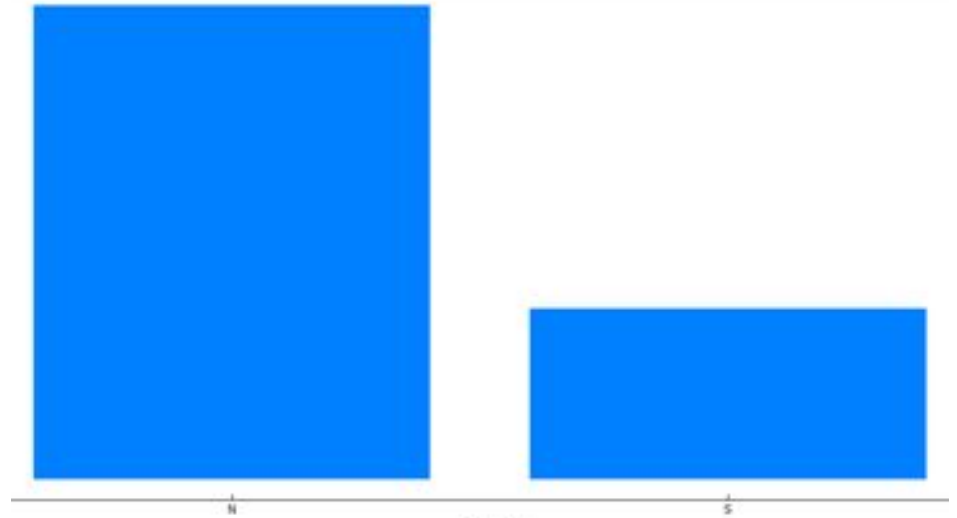
Vamos visualizar os dados com o “widget”, “Distributions”



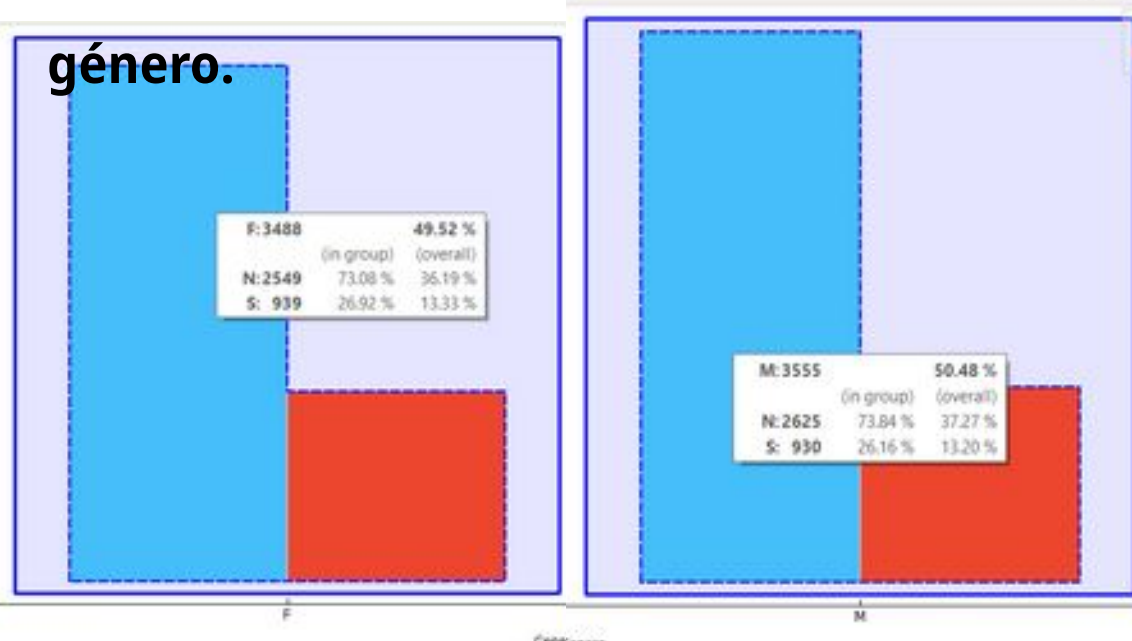
Um histograma é uma das maneiras visuais para compreender a frequência de ocorrência de um intervalo de valores para uma variável.

# Rotatividade (churn) de Clientes

Distribuição da Rotatividade  
(73,5% são da classe "N").



Distribuição da rotatividade por  
gênero.

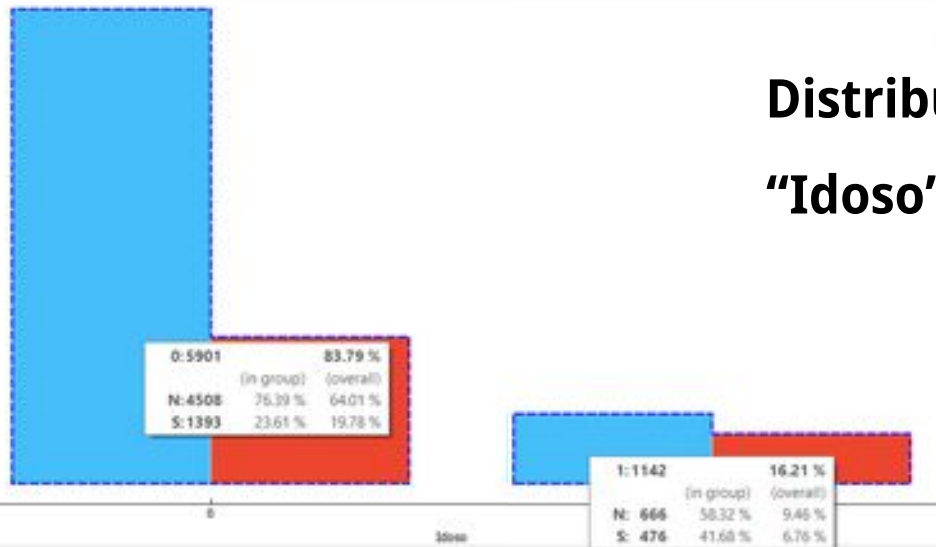


Observar que não existem predominância de sexo entre os clientes, e que pouco mais de 25% dos clientes deixaram a empresa.

Não existe tendencia em relação ao gênero.

# Rotatividade (churn) de Clientes

Distribuição da rotatividade por "Idoso".



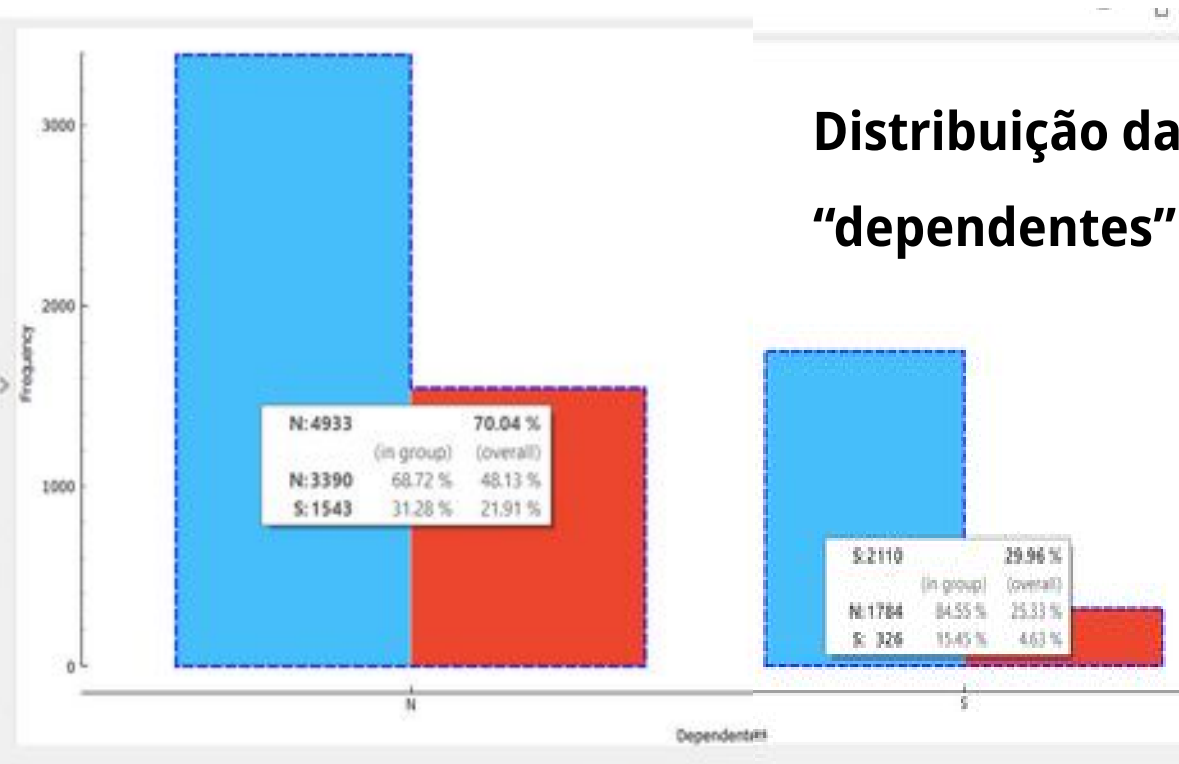
Dos 84% dos clientes com menos idade (não Idosos), 24% deixaram de ser clientes, dos clientes com mais idade (idosos) que são 16% do total, 42% deixaram de ser clientes;

**Existe uma tendencia de "Clientes idosos" de deixar de ser clientes muito maior que os "Clientes não idosos";**



# Rotatividade (churn) de Clientes

Distribuição da rotatividade por “dependentes”.



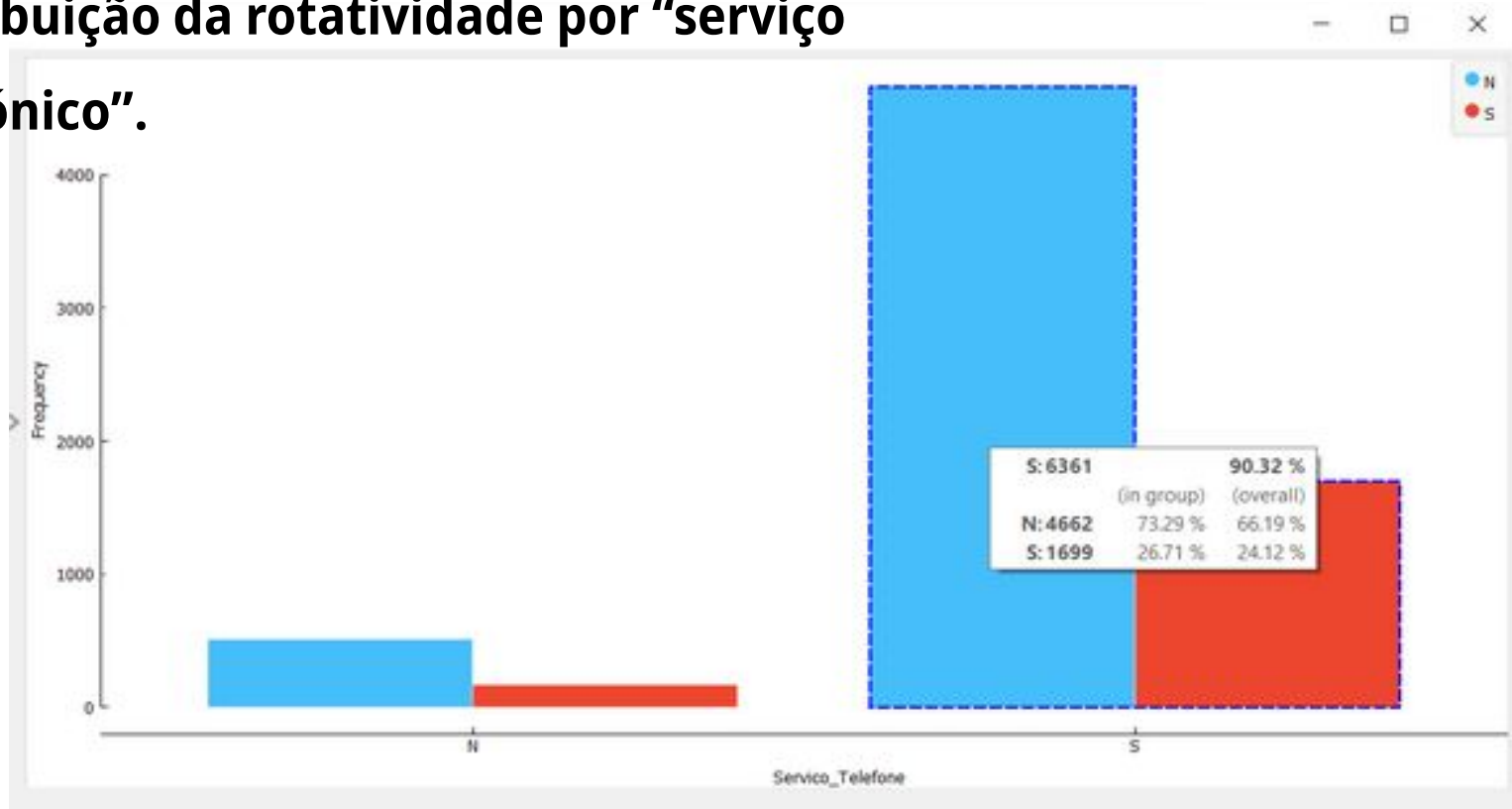
70% dos clientes tem dependentes, destes 31% deixaram de ser clientes, dos que não tem dependentes (30%), 15% deixaram de ser clientes;

**Existe uma tendencia de “Clientes que tem dependentes” de deixar de ser clientes maior dos “Clientes que não tem dependentes”;**

Analizando a distribuição por “parceiro”, temos informações similares.

# Rotatividade (churn) de Clientes

Distribuição da rotatividade por “serviço telefónico”.

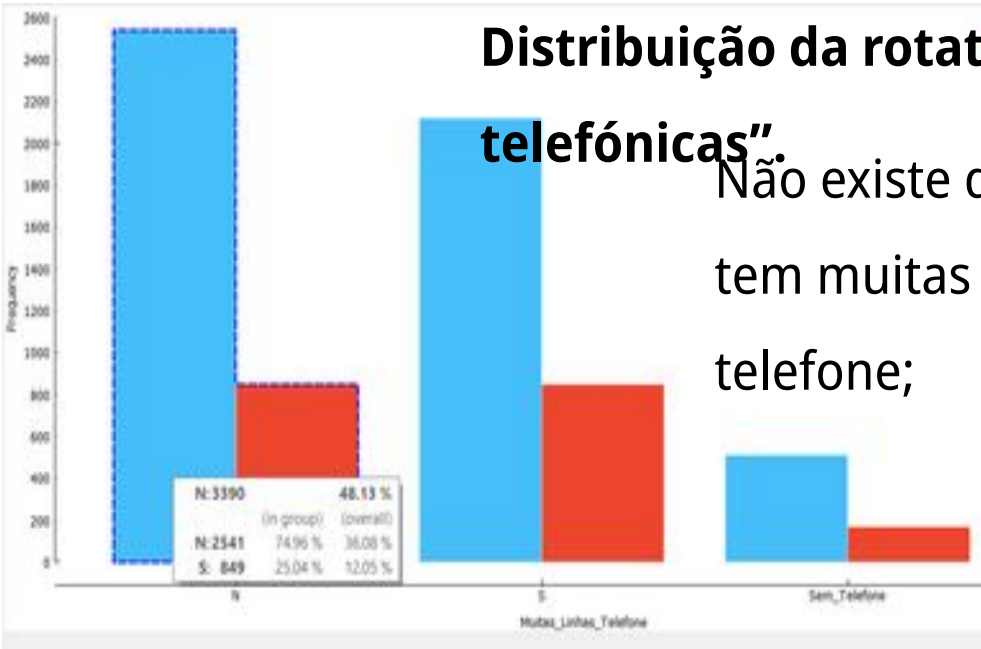


Não existe diferença de rotatividade entre os que tem e não tem serviço telefónico;

# Rotatividade (churn) de Clientes

## Distribuição da rotatividade por “linhas telefônicas”.

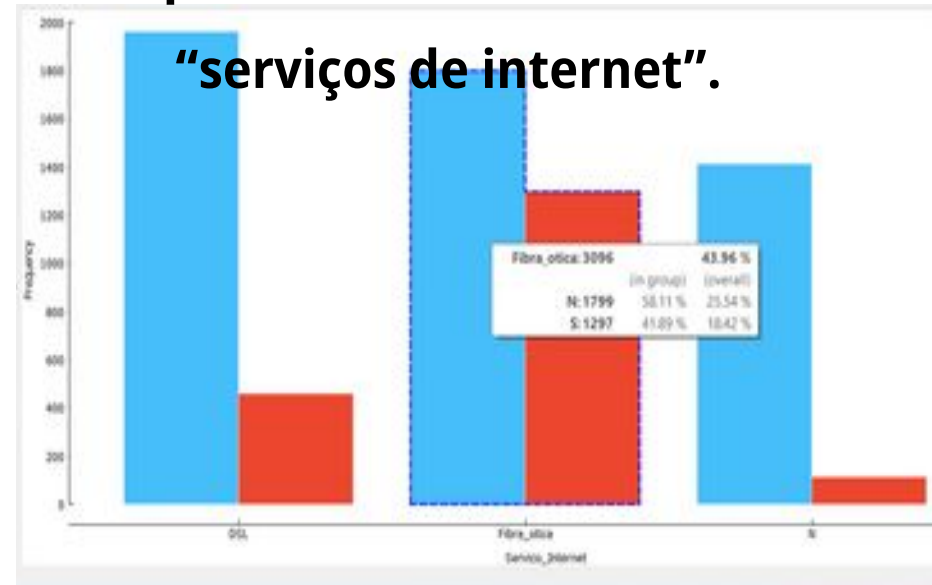
Não existe diferença de rotatividade entre os que tem muitas linhas telefônicas, poucas ou não tem telefone;



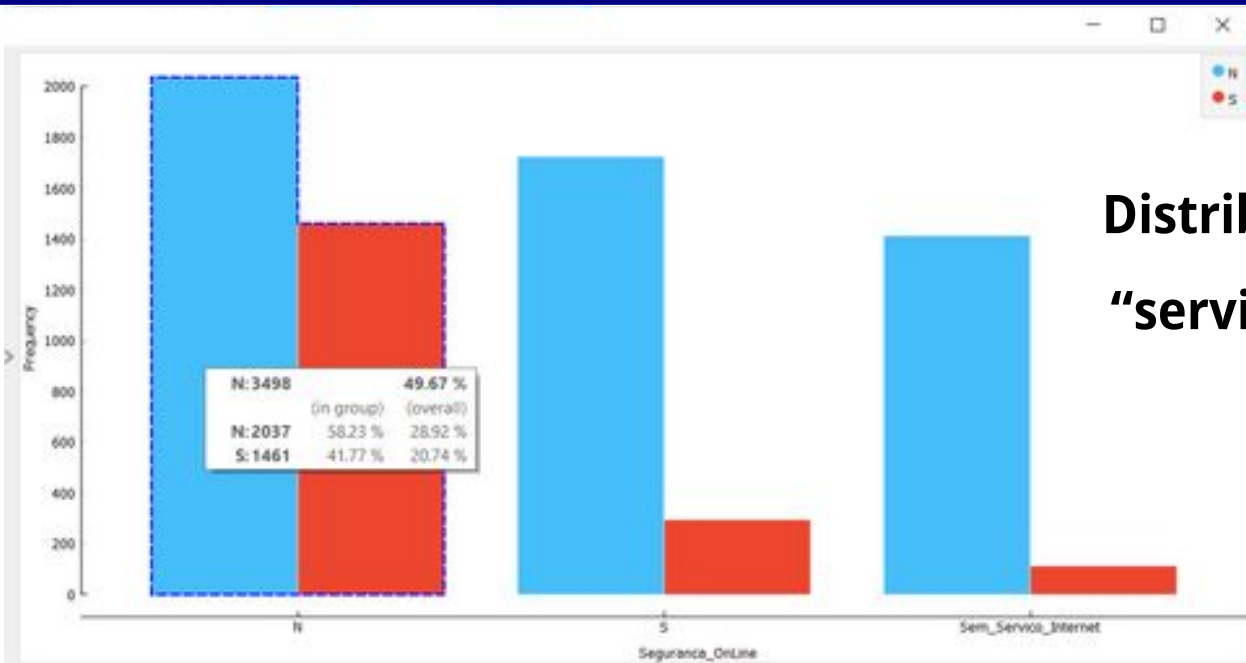
## Distribuição da rotatividade por

## “serviços de internet”.

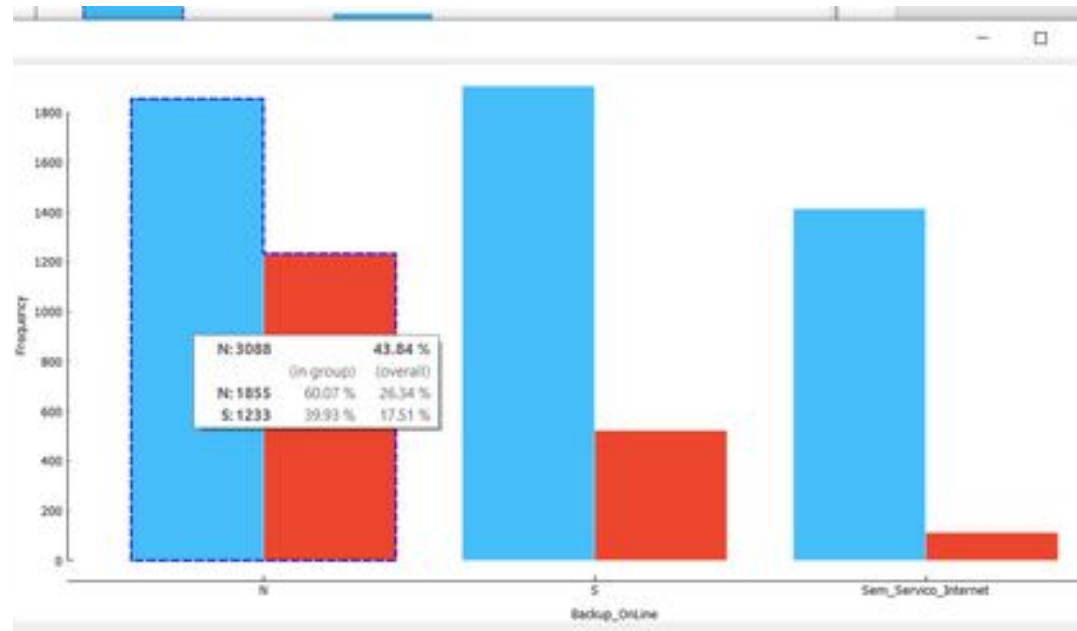
Quando é analisado o serviço de internet, que tem “fibra óptica” tem alta rotatividade, quando comparada com as outras opções;



# Rotatividade (churn) de Clientes



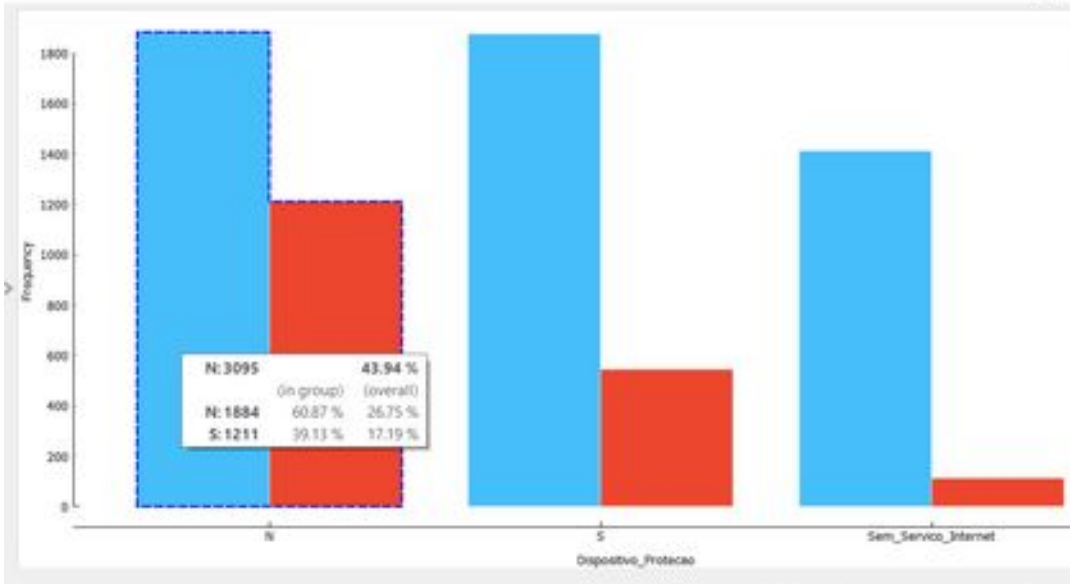
Distribuição da rotatividade por “serviços on-line” e “backup on-line”.



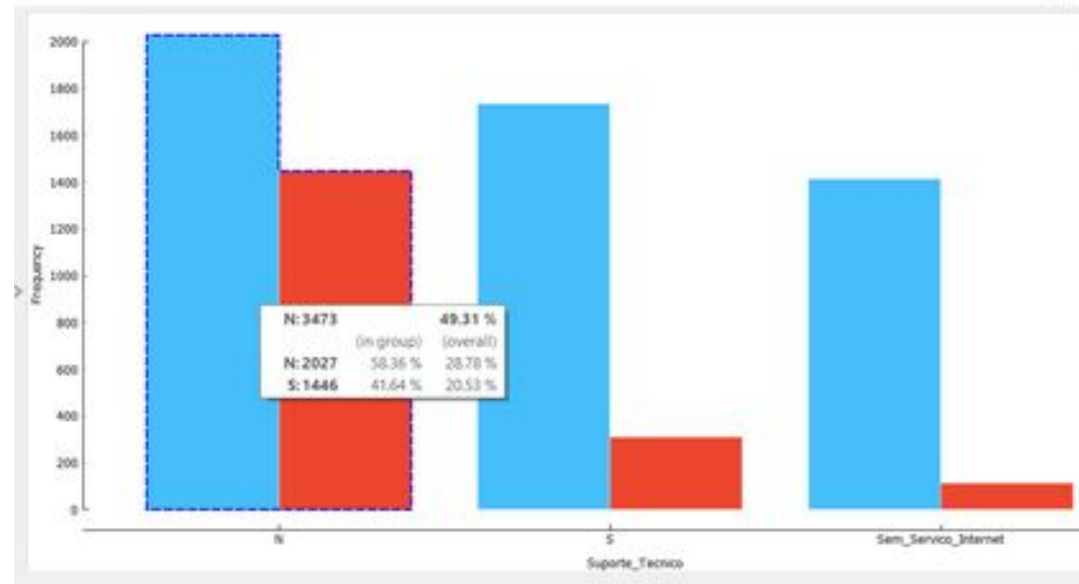
Quando é analisado o serviço on-line e backup on-line, quem não tem os serviços tem alta rotatividade, quando comparado com as outras opções;

# Rotatividade (churn) de Clientes

Distribuição da rotatividade por “dispositivo de proteção” e “suporte técnico”.

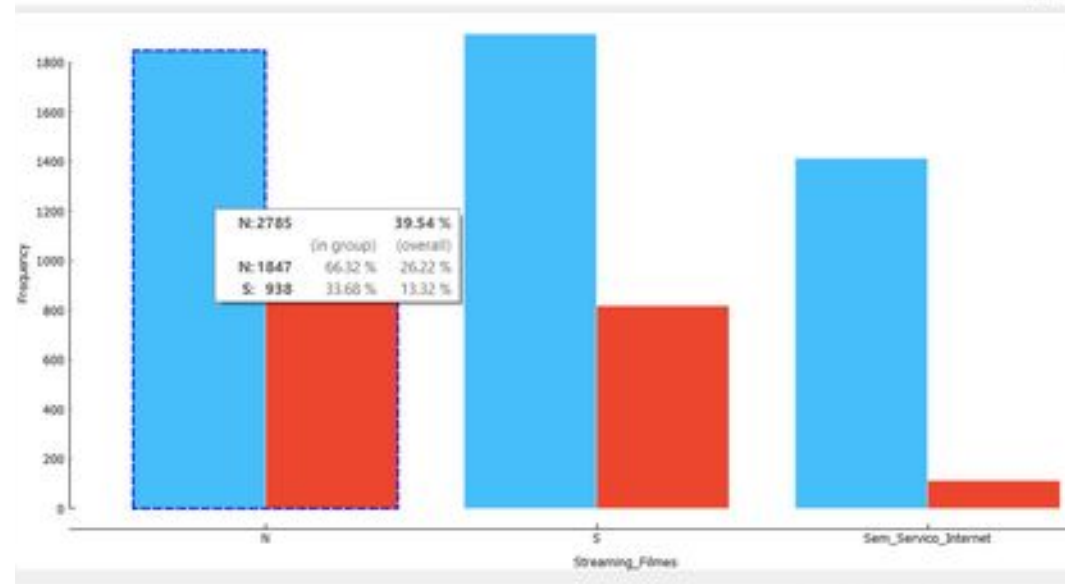
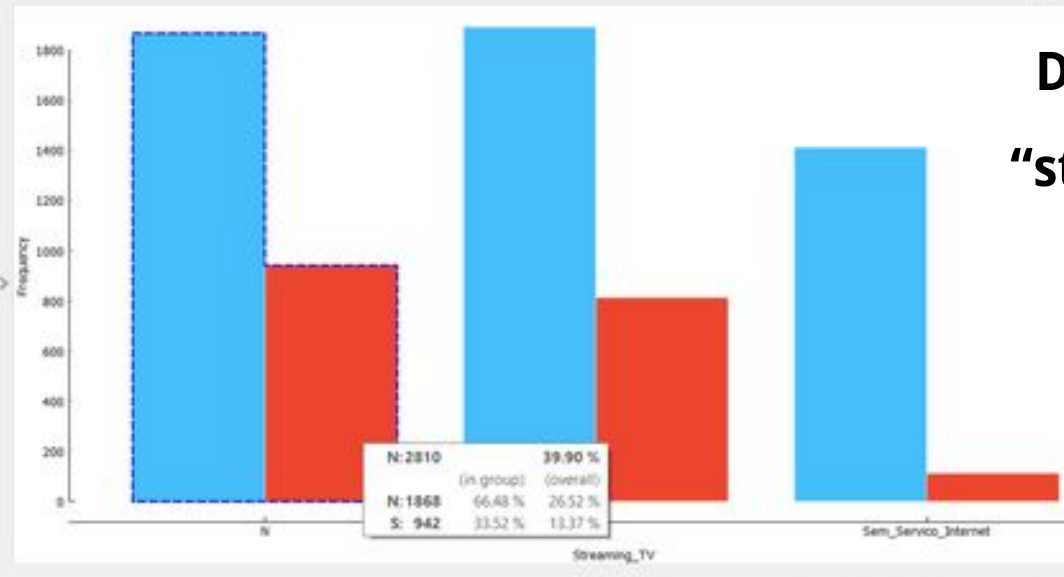


Quando é analisado o dispositivo de proteção e o suporte técnico, quem não tem os serviço tem alta rotatividade;



# Rotatividade (churn) de Clientes

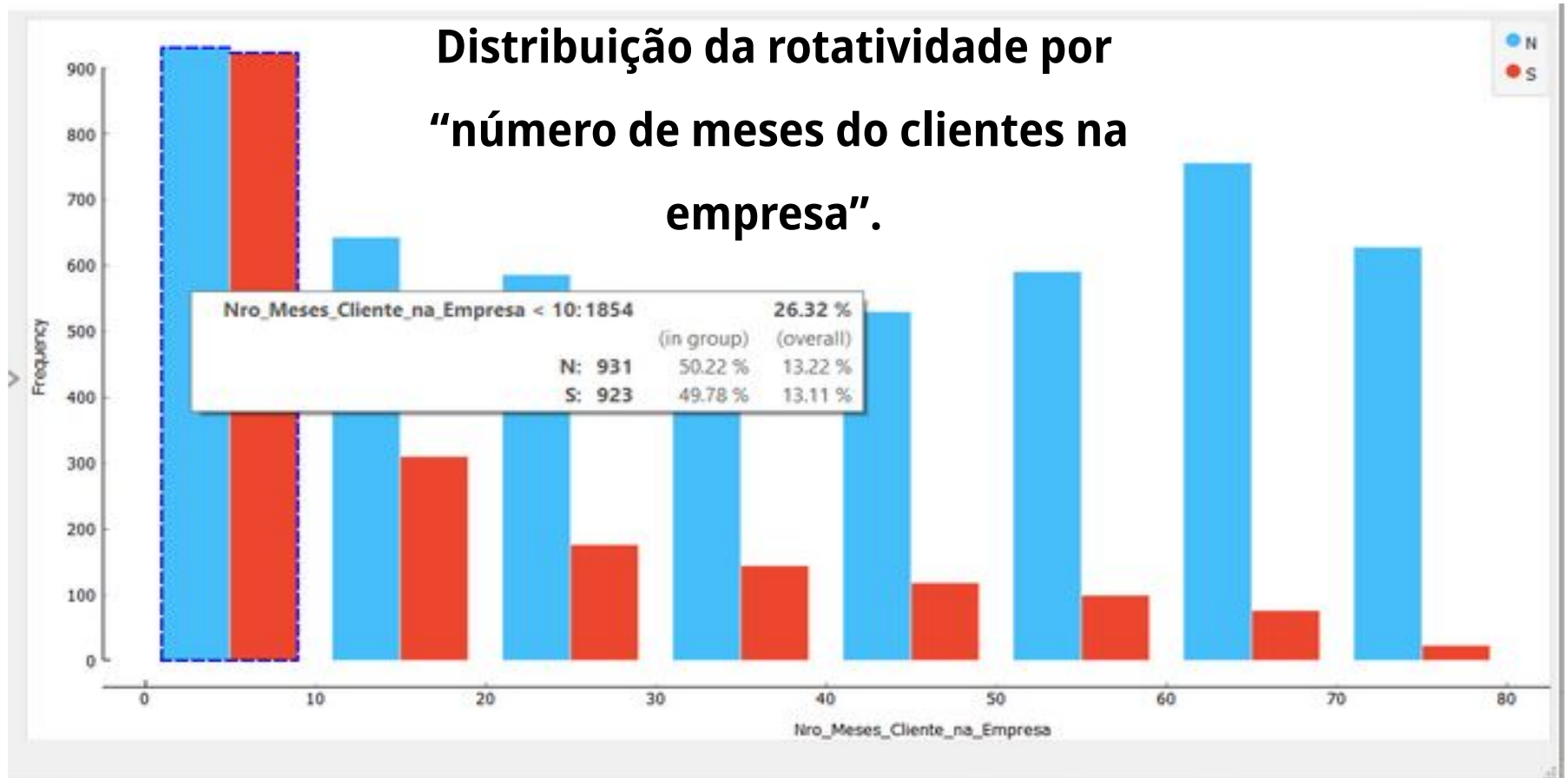
Distribuição da rotatividade por “streaming de TV” e “streaming de Filmes”.



Quando é analisado o Streaming de TV e de filmes, não existem diferenças entre quem tem o serviço e não tem o serviço quando observamos a rotatividade;

# Rotatividade (churn) de Clientes

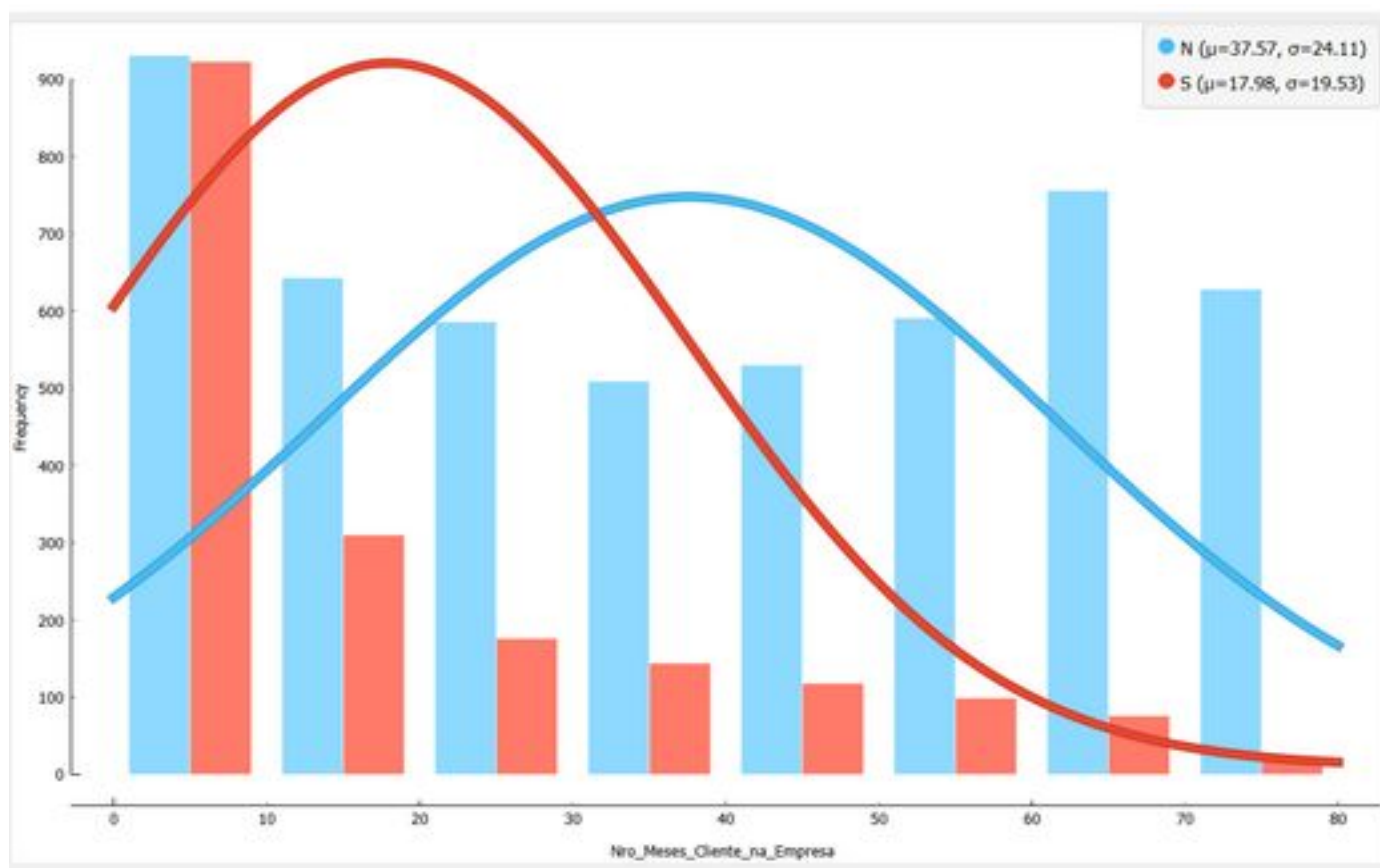
Distribuição da rotatividade por  
"número de meses do clientes na  
empresa".



Quem é cliente com menos tempo (até 10 meses), tem alta rotatividade (50%);

# Rotatividade (churn) de Clientes

Distribuição da rotatividade por “número de meses do clientes na

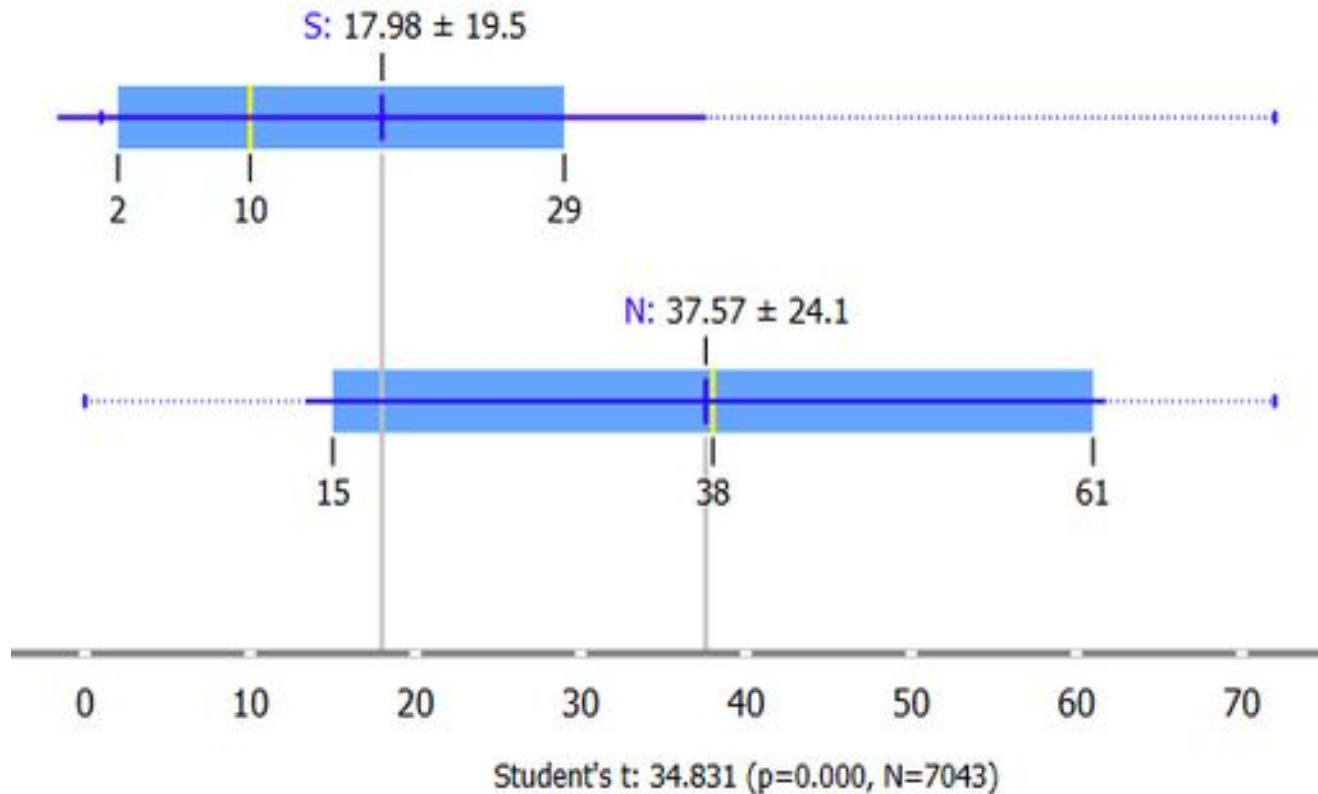


Outra visão do número de meses de clientes na empresa;



# Rotatividade (churn) de Clientes

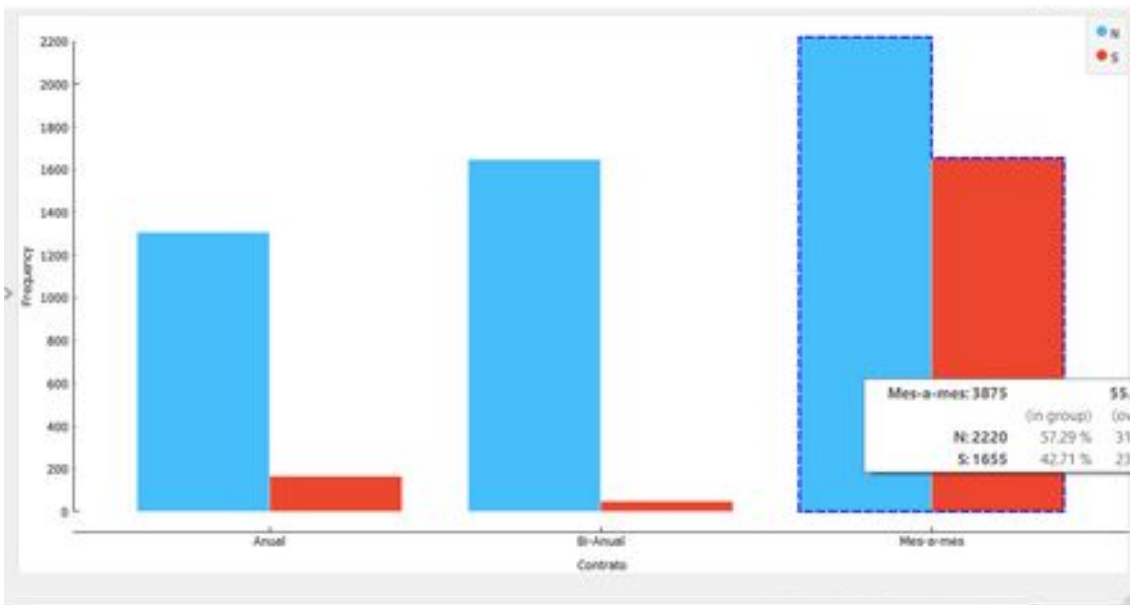
Distribuição da rotatividade por “número de meses do clientes na



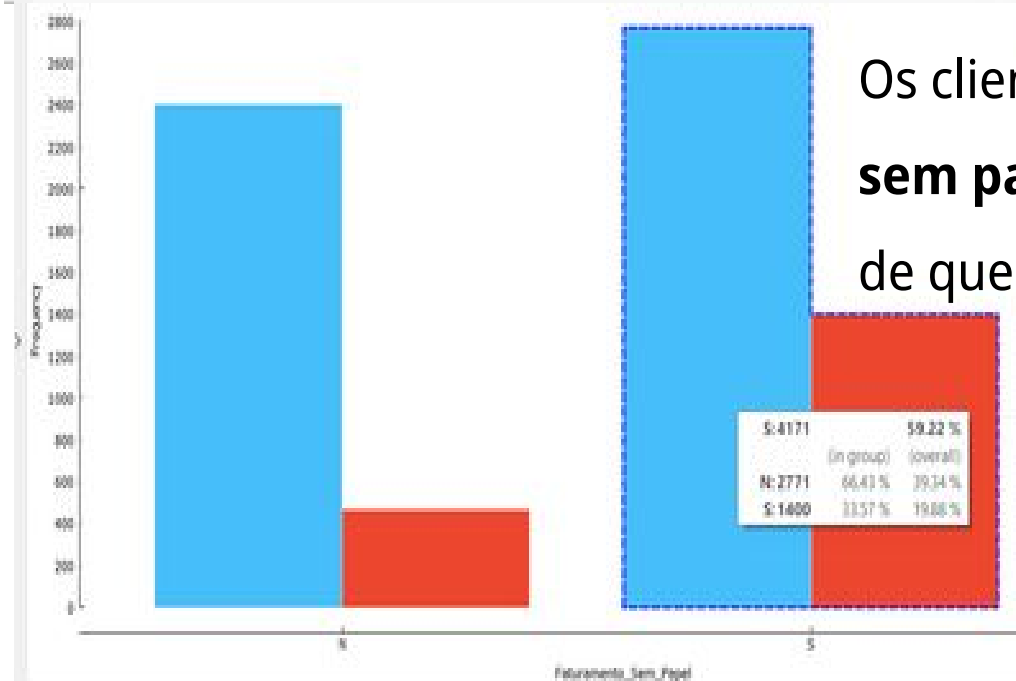
Outra visão do número de meses de clientes na empresa:

observamos que os clientes com rotatividade tem uma media de 10 meses de permanência da empresa contra 38 meses do não rotativos;

# Rotatividade (churn) de Clientes

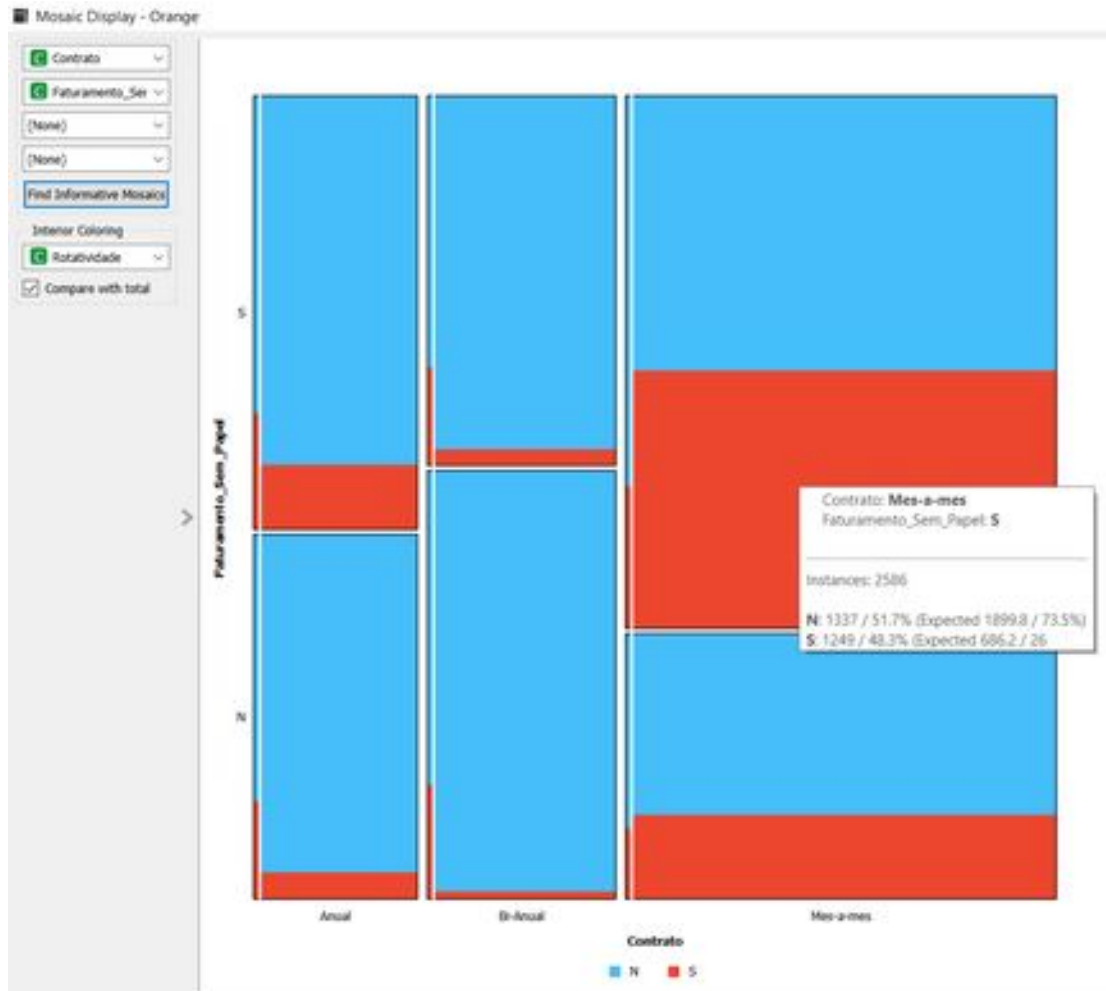


Quando é analisado o **tipo de contrato**, que tem tipo mês-a-mês, tem alta taxa de rotatividade;



Os cliente que optarão pelo **faturamento sem papel**, tem o dobro de rotatividade de quem não adoptou;

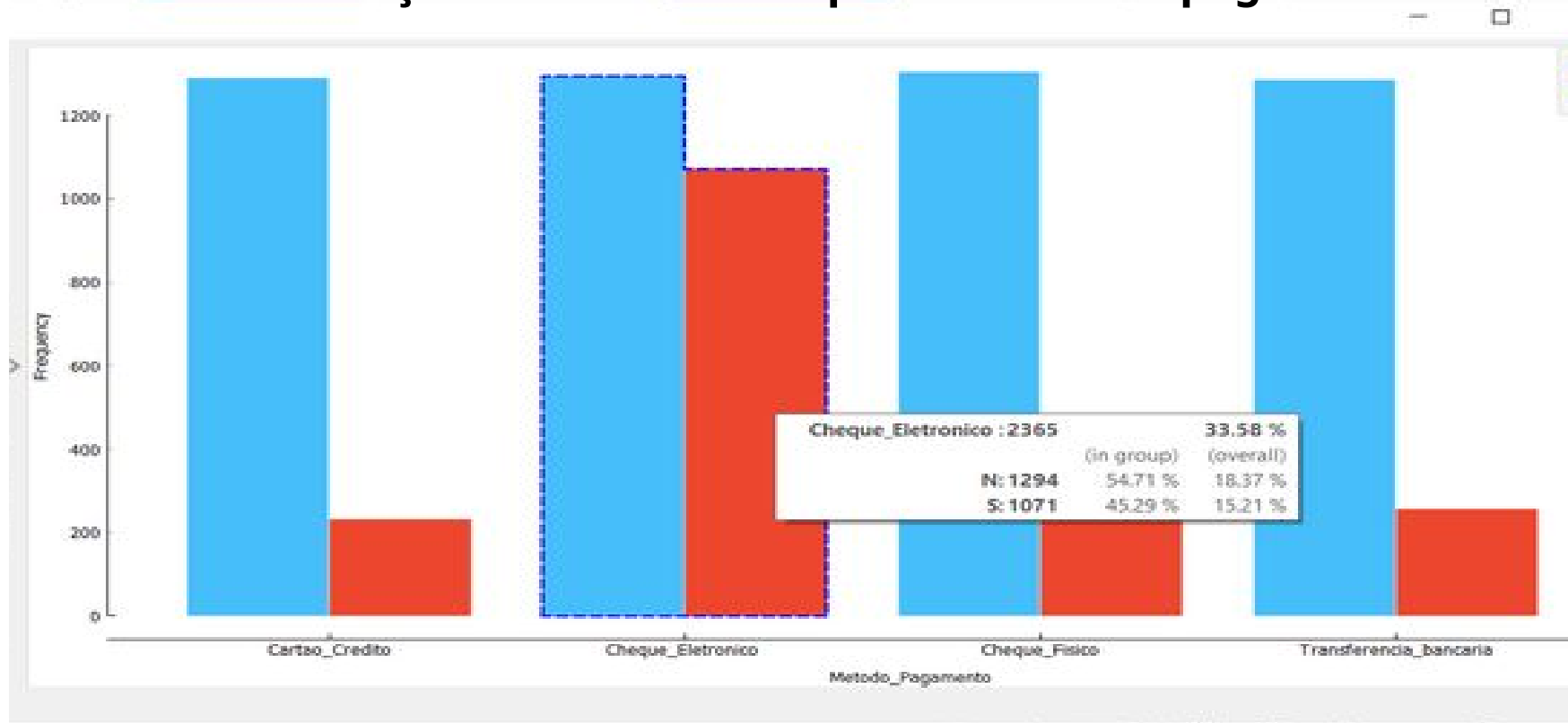
# Rotatividade (churn) de Clientes



Quase 50% dos clientes que optarão pelo **faturamento sem papel**, e **tipo de contrato mês-a-mês** deixaram de ser clientes;

# Rotatividade (churn) de Clientes

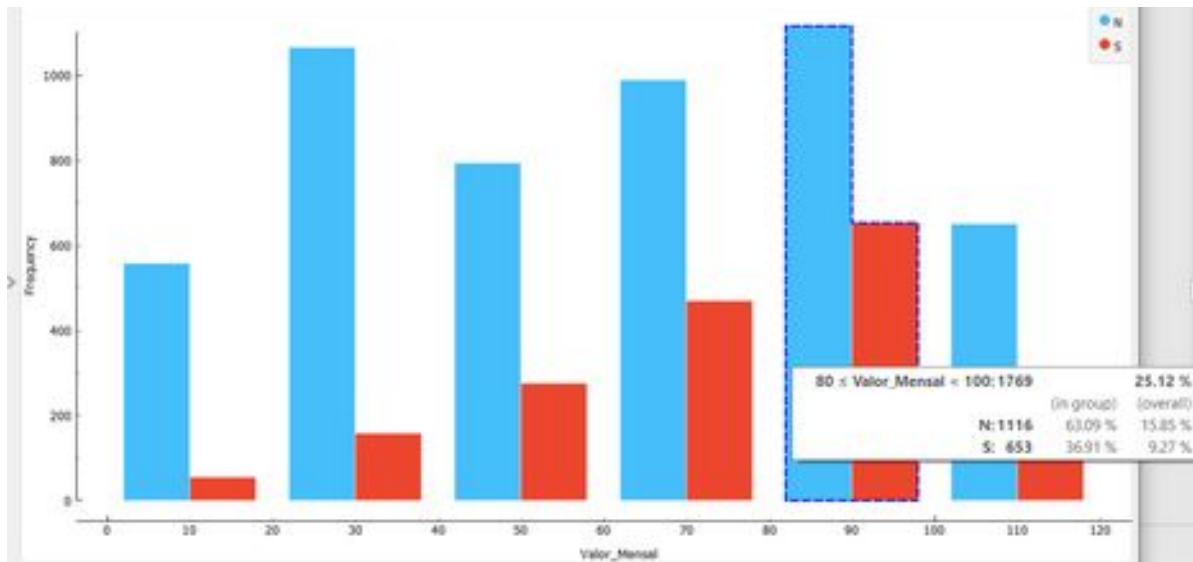
Distribuição da rotatividade por “método de pagamento”.



Existe uma alta rotatividade entre os cliente que optarão pelo método de faturamento “cheque eletrônico”;

# Rotatividade (churn) de Clientes

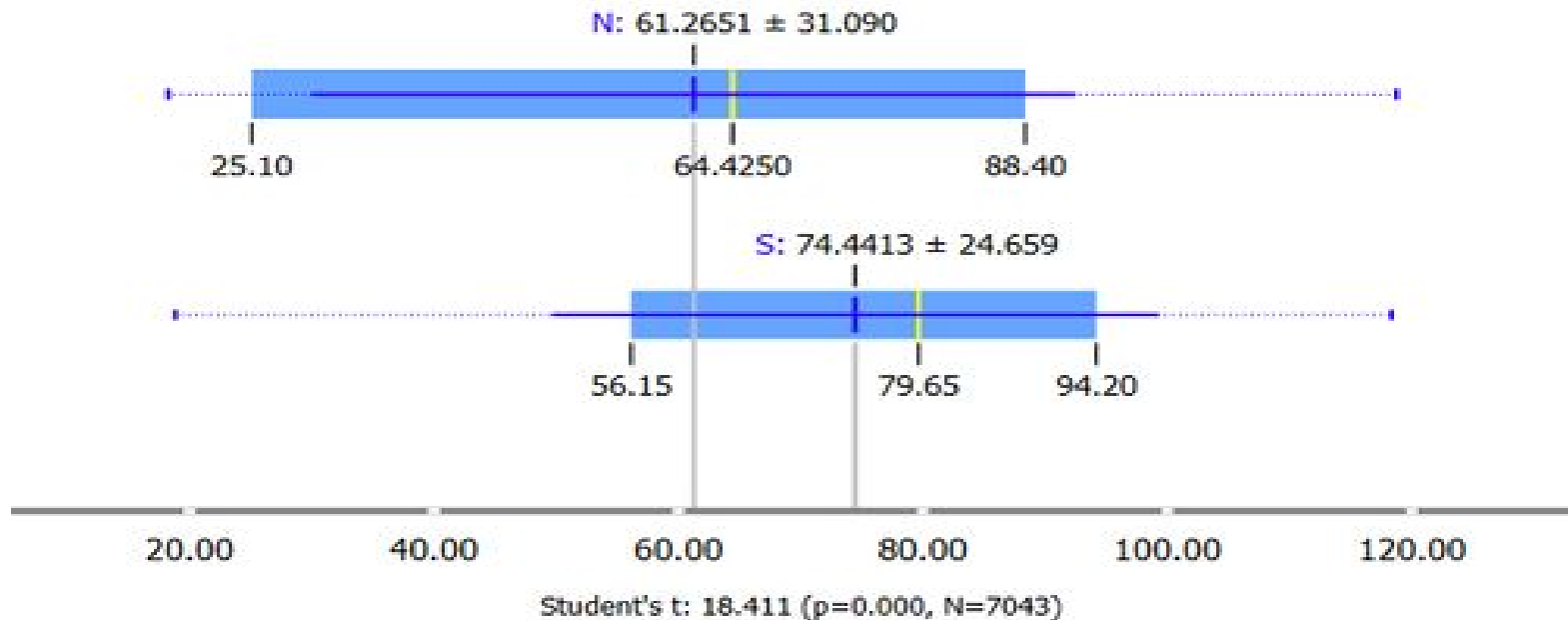
Distribuição da rotatividade por “valor mensal”.



Os clientes com um valor mensal mais “alto”, tem uma taxa de rotatividade maior;

# Rotatividade (churn) de Clientes

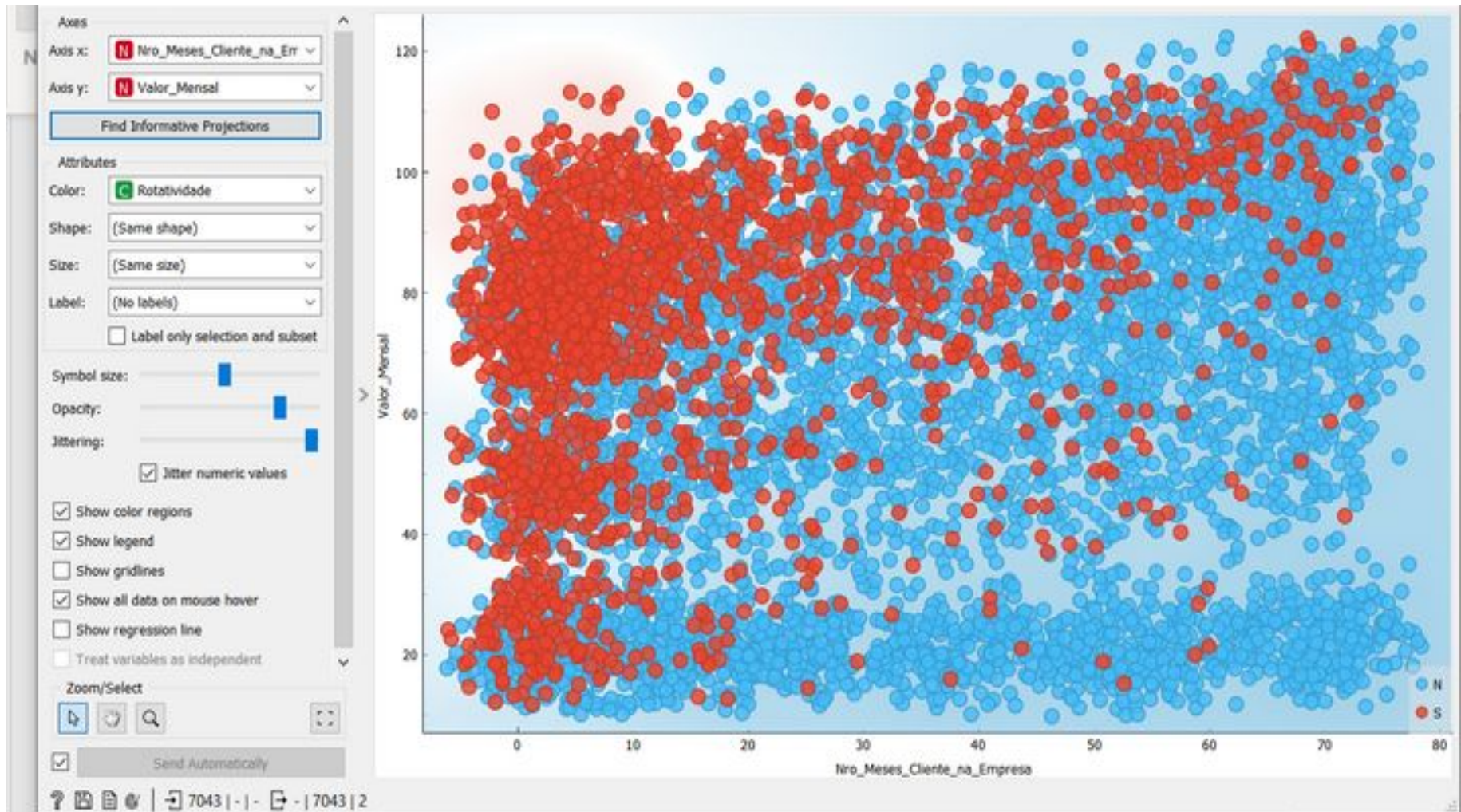
Distribuição da rotatividade por “valor mensal”.



Os clientes com um valor mensal mais “alto”, tem uma taxa de rotatividade maior,

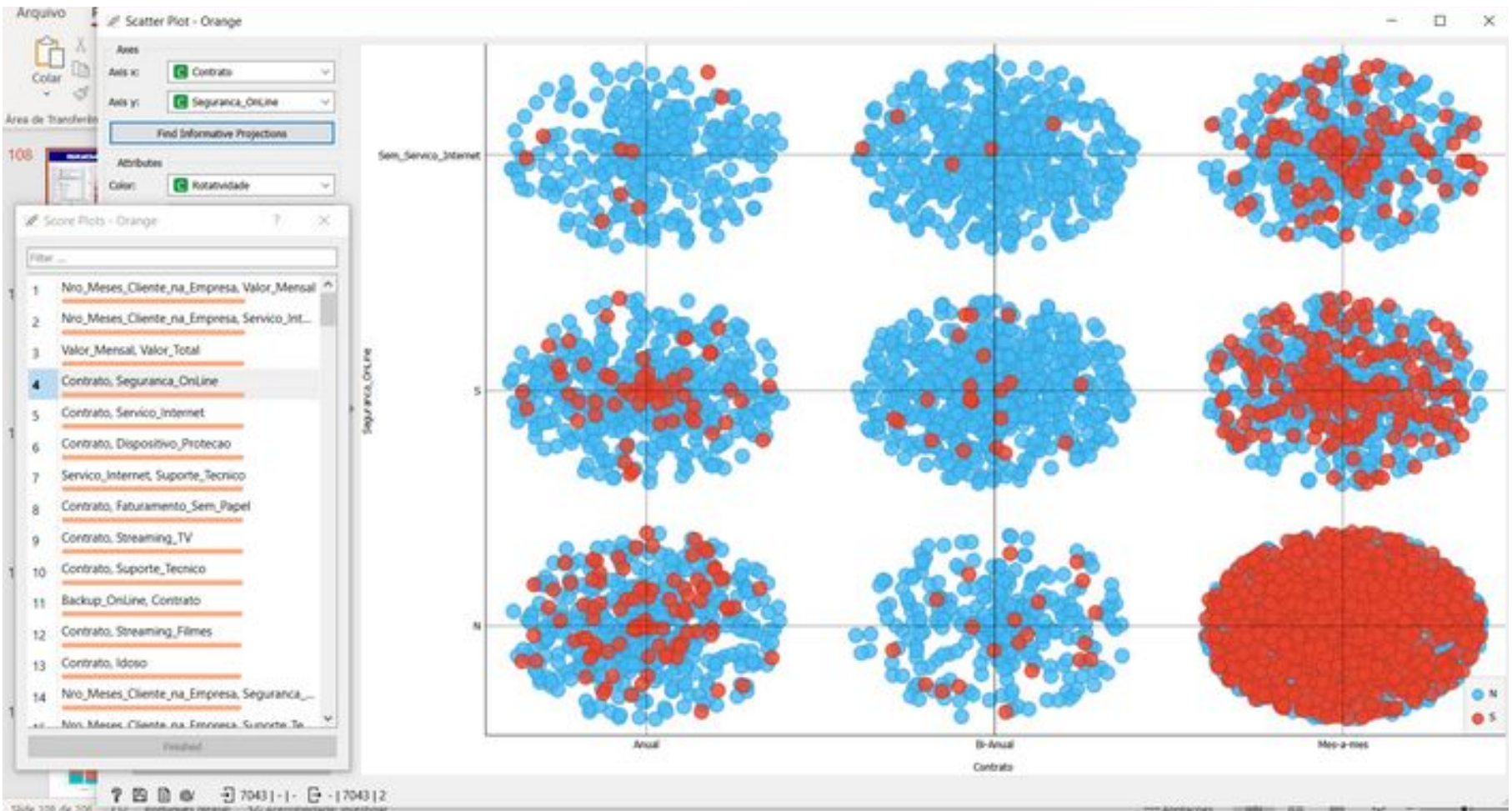
Os clientes rotativos tem uma mensalidade média de 74,4 dólares contra 61,3 dos não rotativos;

# Rotatividade (churn) de Clientes



Os clientes com poucos meses como clientes e valor mensal mais alto, tem uma alta rotatividade;

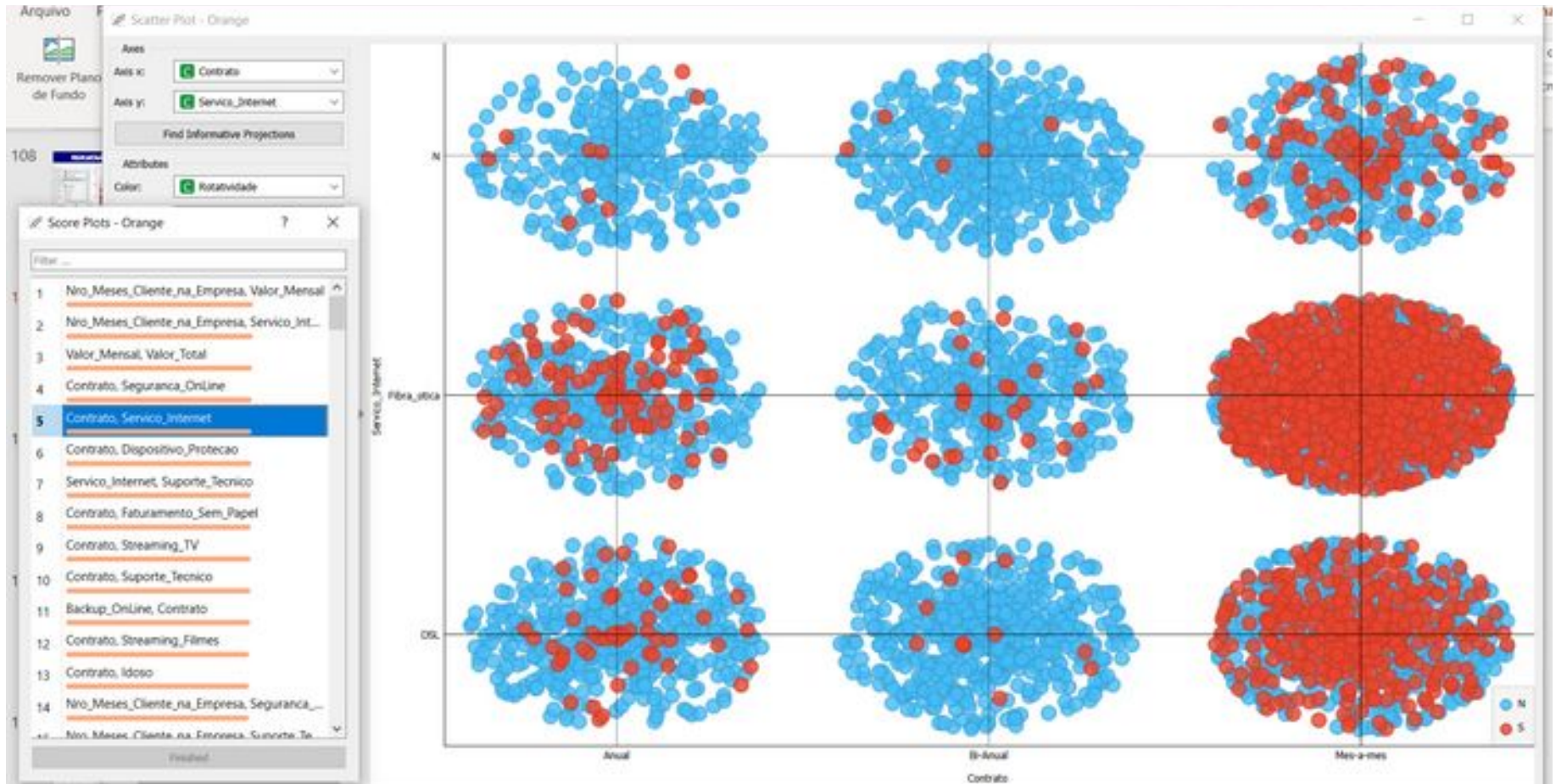
# Rotatividade (churn) de Clientes



Os clientes com tipo de contrato mês-a-mês e sem serviços online, tem uma alta rotatividade;

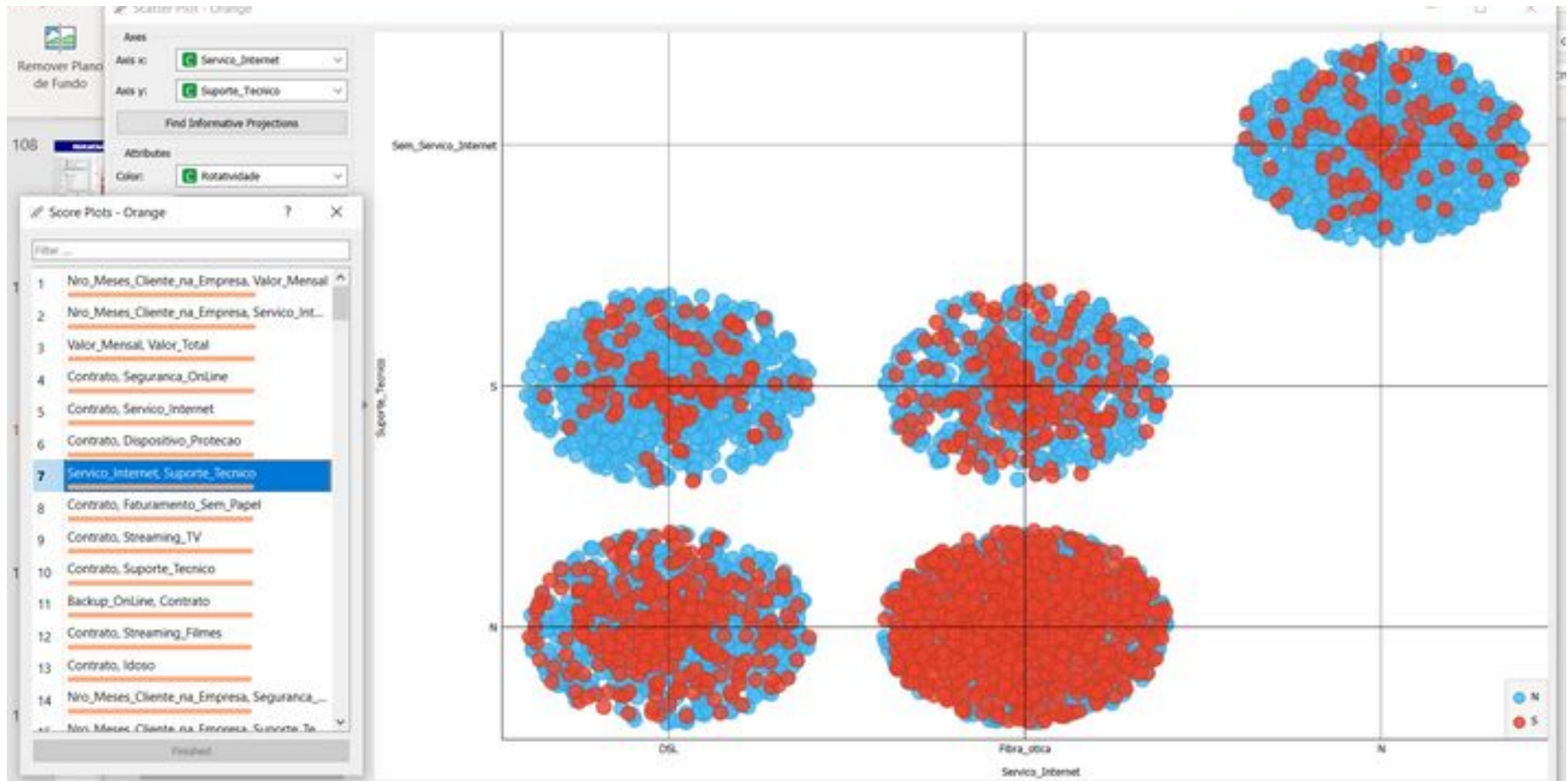


# Rotatividade (churn) de Clientes



Os clientes com poucos meses como clientes e com serviços de internet de “fibra ótica”, tem uma rotatividade muita alta;

# Rotatividade (churn) de Clientes



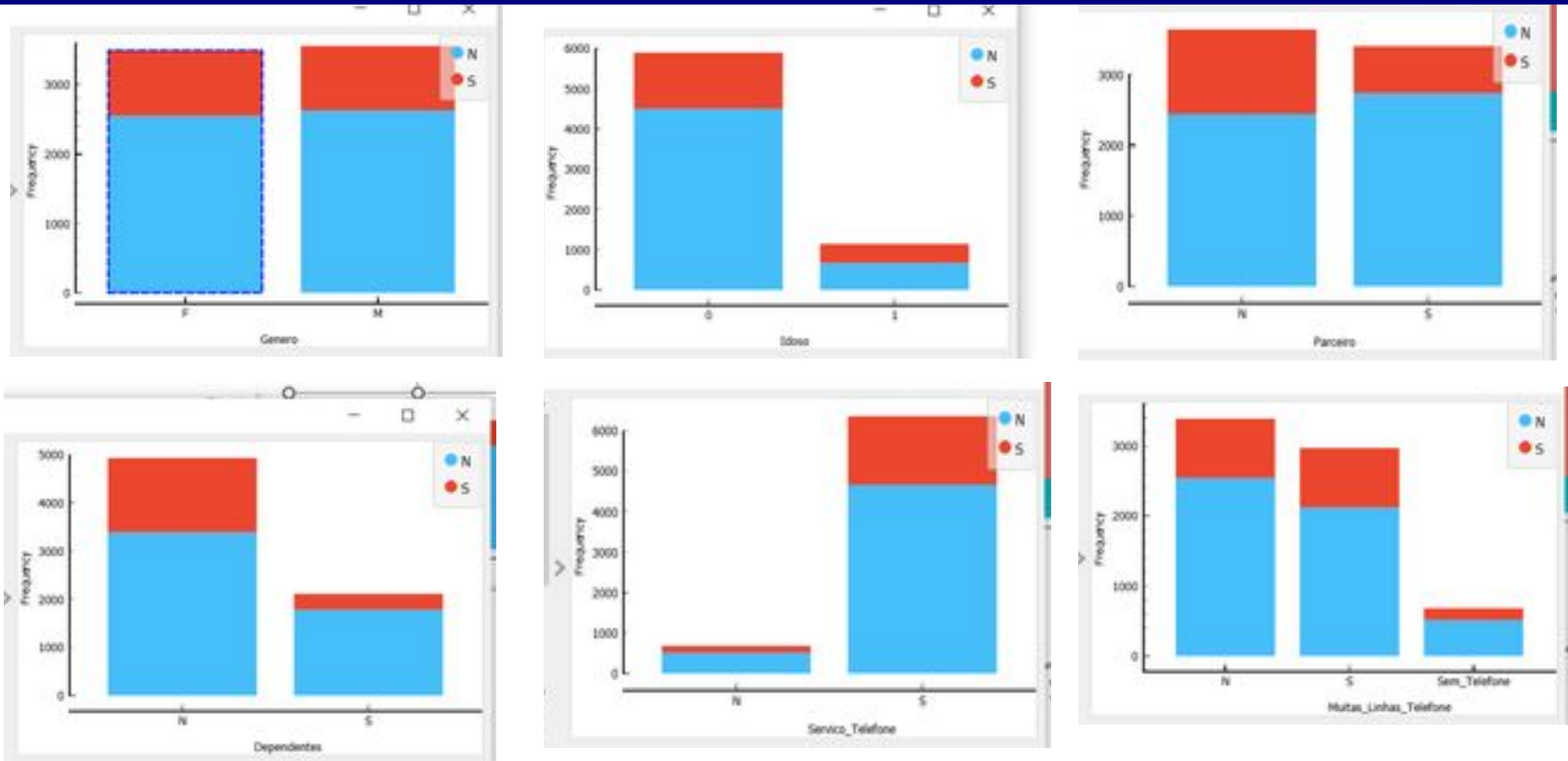
Os clientes sem suporte técnico tem rotatividade muito acima dos clientes com suporte técnico;

# Rotatividade (churn) de Clientes

**Resumo:** os clientes que tem perfil para **deixar de usar os serviços** da Telco são:

- Clientes com contrato mês-a-mês;
- Clientes que não tem serviços como Backup on-line; Segurança on-line; Dispositivo de proteção;
- Clientes que não tem parceiros ou dependentes;
- Clientes que tem internet de fibra ótica;
- Clientes com mais idade (idosos);
- Clientes que tem aderido ao Faturamento sem Papel;
- Clientes com valor mensal “alto” (faixa acima de \$60,00) ;

# Rotatividade (churn) de Clientes



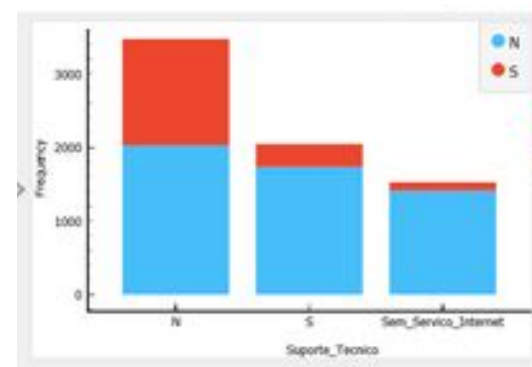
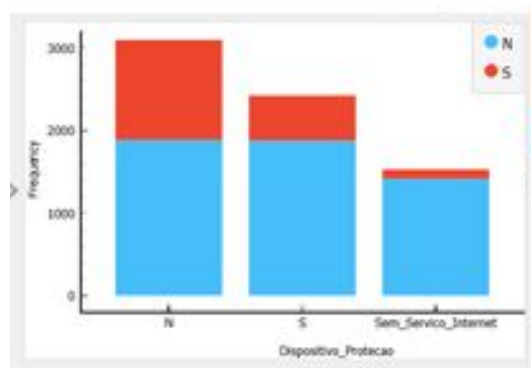
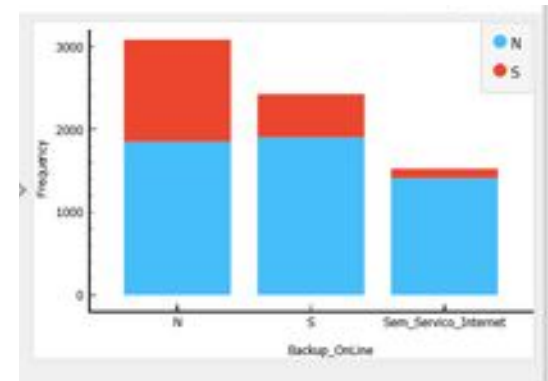
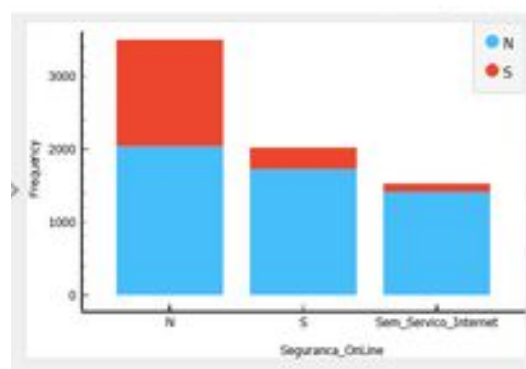
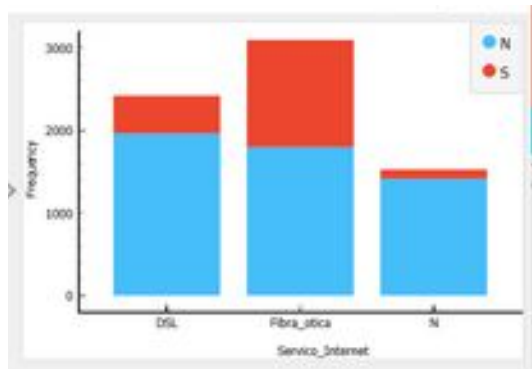
**Sexo:** percentual de rotatividade é quase igual no caso de homens e mulheres

A porcentagem de rotatividade é maior no caso de **idosos**

Clientes **com parceiros** e **dependentes** têm menor taxa de rotatividade em comparação com aqueles que não têm parceiros e dependentes.

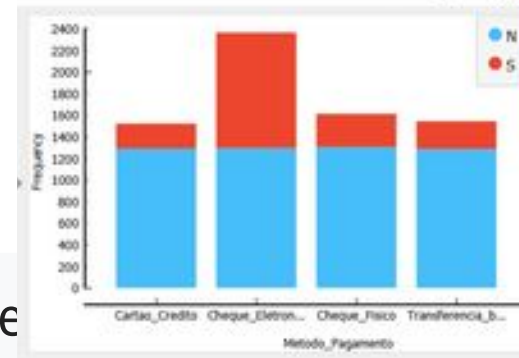
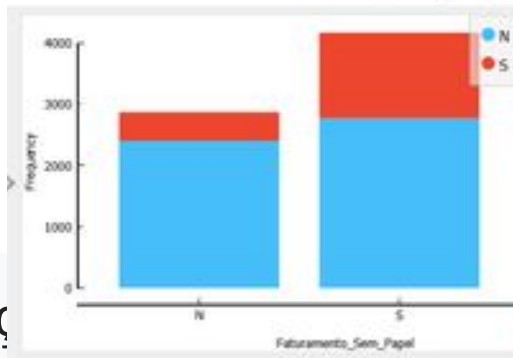
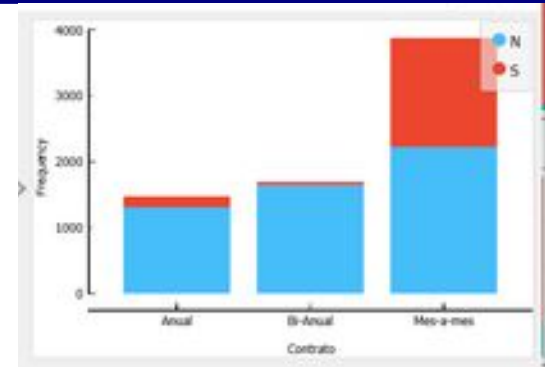
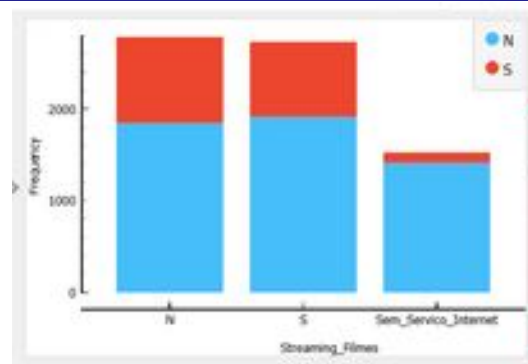
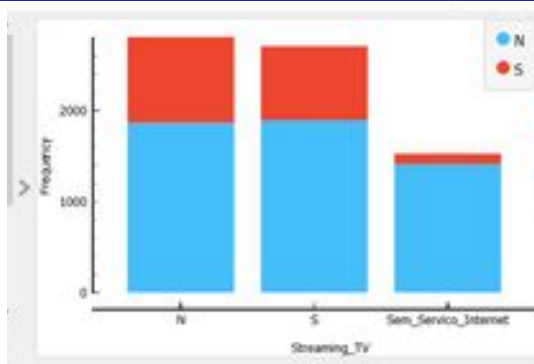
Não existe diferença para os atributos “serviços telefônicos” e “muitas linhas”

# Rotatividade (churn) de Clientes



A taxa de rotatividade é muito maior no caso dos serviços de Internet de fibra óptica. Clientes que não possuem serviços de “segurança on-line”, “backup on-line”, “dispositivo de proteção” e “suporte técnico”.

# Rotatividade (churn) de Clientes



Ter os serviços (streaming TV e filmes) não tem relação com a rotatividade.

Uma porcentagem maior de clientes com **assinatura mensal** tem alta rotatividade.

A porcentagem de rotatividade é maior no caso de clientes com **opção de cobrança sem papel**.

Os clientes que tem método de pagamento "**cheque eletrônico**" tendem a deixar a

# Rotatividade (churn) de Clientes

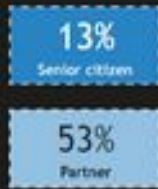
## CUSTOMER RISK ANALYSIS: WHO ARE MY CHURNERS?



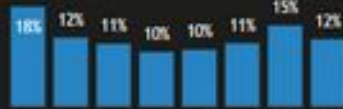
### CLIENT PROFILE

5174

#### DEMOGRAPHICS



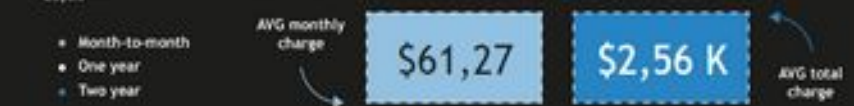
Tenure bins (10 month)



#### PHONE SERVICE



#### CONTRACT TERMS



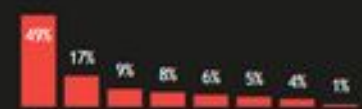
### CHURNER PROFILE

1869

#### DEMOGRAPHICS



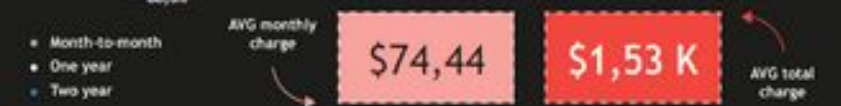
Tenure bins (10 month)



#### PHONE SERVICE



#### CONTRACT TERMS



# Rotatividade (churn) de Clientes



The screenshot shows a Jupyter Notebook interface. The browser address bar displays a Google Drive link. The notebook title is "Telecom\_churn\_Exploratoria\_Destritiva\_2023.ipynb". The menu bar includes "Arquivo", "Editar", "Ver", "Inserir", "Ambiente de execução", "Ferramentas", and "Ajuda". The left sidebar shows a file explorer with "Código" and "Texto" tabs. The main content area has a search icon and a "Conectar" button. A dropdown menu is open, showing "Análises exploratória / descritiva". Below this, a text cell contains the following text: "Churn é um dos maiores problemas da indústria de telecomunicações. A pesquisa mostrou que a taxa média mensal de rotatividade entre as 4 principais operadoras de telefonia móvel nos EUA é de 1,9% a 2%". Below the text cell is a code cell with the following Python code:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns # for creating plots
import matplotlib.ticker as mtick # For specifying the axes tick format
import matplotlib.pyplot as plt
```



# Análises de Dados e Big Data



**Sistemas de Informação / Ciências da Computação**

Aran Bey Tcholakian Morales, Dr. Eng.

Apostila 7: Análises exploratória / descritiva