

Desafio Técnico - Matheus R O Solon

Reclamações relacionadas a celulares

Metodologia

Toda a análise exploratória foi feita, na maior parte, com *Python*. O código foi feito inteiramente no [Colab](#) e o acesso está disponível caso queiram ver. O *Power BI* foi utilizado também para uma exploração dos dados, mas em menor parte. Seu foco foi a criação do dashboard a partir dos insights obtidos da análise com *Python*.

A análise exploratória começa pelo pré-processamento e depois pela visualização dos gráficos e relações entre os dados. Em algumas variáveis, utilizei testes estatísticos para garantir a robustez da análise. Por fim, criei um dashboard, a partir dos insights que se encontram no fim desse PDF, que mostra as principais relações entre as variáveis

Informações sobre a base de dados

Temos as seguintes colunas e suas respectivas informações

- **Tipo**: tipo de plano do cliente.
- **Sexo**: sexo do cliente.
- **Opinião**: opinião sobre o serviço.
- **Modelo**: modelo do celular.
- **Outra**: coluna sem informação.
- **Razão**: motivo da reclamação.
- **Renda**: renda do cliente.
- **Anos**: tempo de contrato.
- **Gasto**: gasto mensal.
- **Minutos**: uso de minutos.

Pré-processamento

Antes de começar a análise estatística em cima dos dados é imprescindível fazer o tratamento dos dados em relação aos valores espúrios, nulos *etc*. Fiz o tratamento desses dados tanto em *Python* quanto no *Power Query*, mas mostrarei aqui apenas a metodologia utilizada no *Power Query*.

Ao carregar a base de dados no *Power Query*, podemos encontrar valores incorretos e nulos em todas as colunas.

Os valores incorretos ocorrem de tal maneira que parecem que não foram preenchidos completamente antes de serem processados. Por exemplo, na coluna **Tipo** podemos inferir que ela assume dois valores: *Pós-pago* e *Pré-pago*. Entretanto, existem valores que estão digitados como "Pós-", "Pós-pa", "Pr" etc., esse padrão ocorre em todas as colunas. Portanto, podemos substituir esses valores pelos devidamente corretos. Segue o código no *Power Query* para fazer o tratamento desses dados:

```
#"Tratamento Coluna Tipo" = Table.ReplaceValue(
    #"Tipo Alterado",
    each [Tipo],
    each if Text.Start([Tipo], 2) = "Pr" then
        "Pré-pago"
    else if Text.Start([Tipo], 2) = "Pó" then
        "Pós-pago"
    else
        [Tipo], Replacer.ReplaceValue,{"Tipo"}
// e assim por diante para todas as colunas...
```

Esse código pode ser reaproveitado para as colunas restantes.

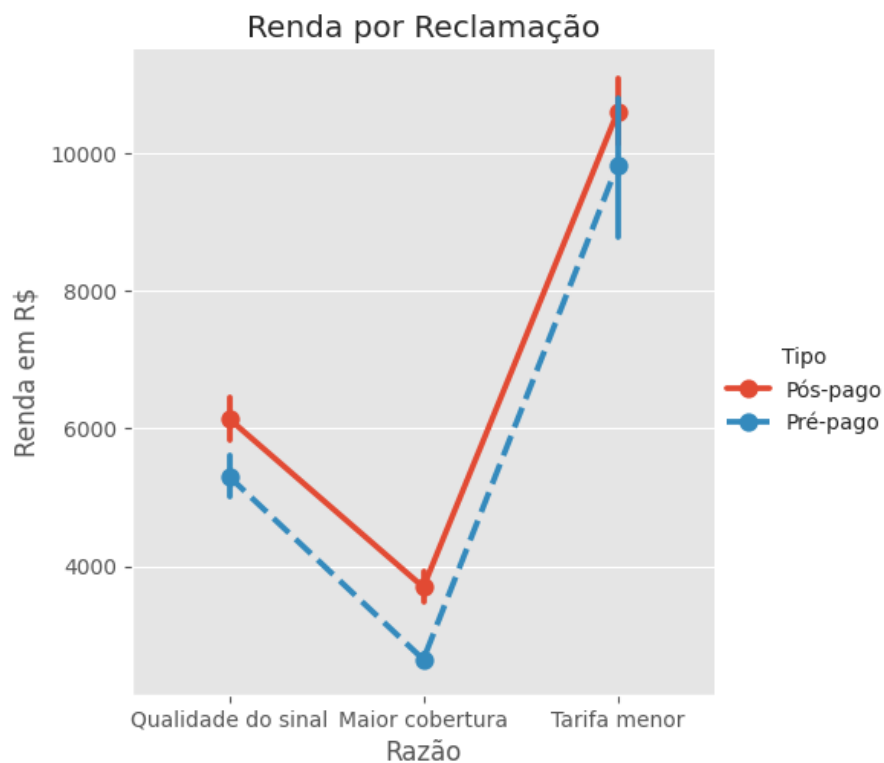
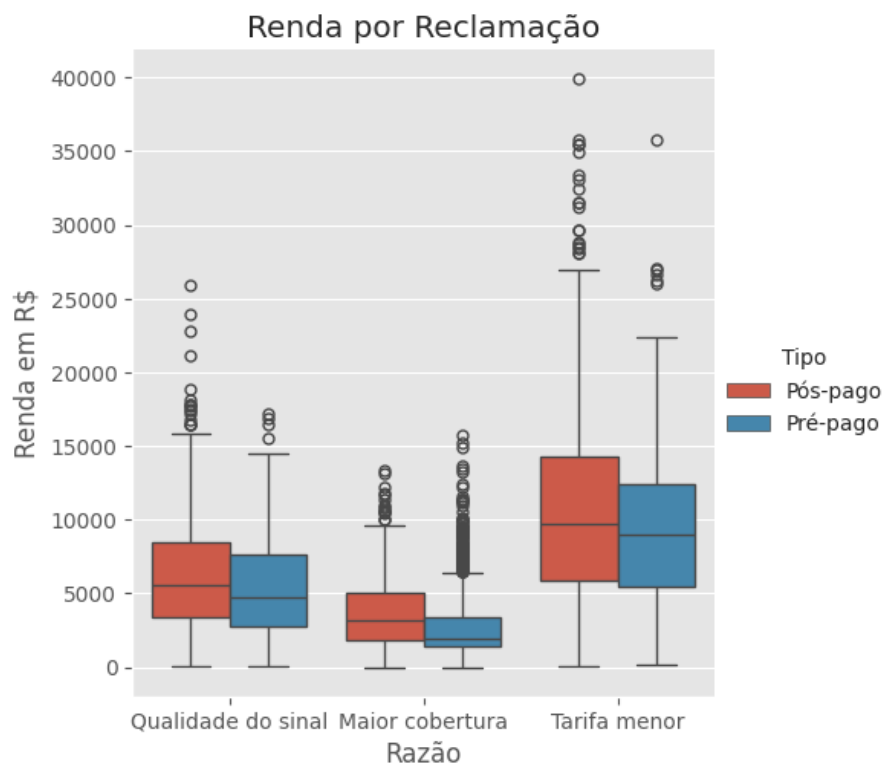
Com relação aos valores nulos, não há muito que possamos fazer. Idealmente, falaríamos com o time que coletou os dados para nos enviar os dados corrigidos ou nos indicar alguma regra para nós mesmos fazermos a correção. Visto que isso não é possível, podemos excluir as linhas contendo os dados ou isolá-las na análise. Como nosso objetivo é criar um dashboard, excluirei os registros para que os gráficos não sejam afetados.

Além disso, não recebi nenhuma informação sobre a coluna **Outra** e nem consegui inferir o que essa coluna poderia informar, portanto a considereei como sendo espúria. Por último, a coluna **Renda** possuía formatação errada, pois a vírgula estava sendo utilizada como divisor de milhares, necessitando correção.

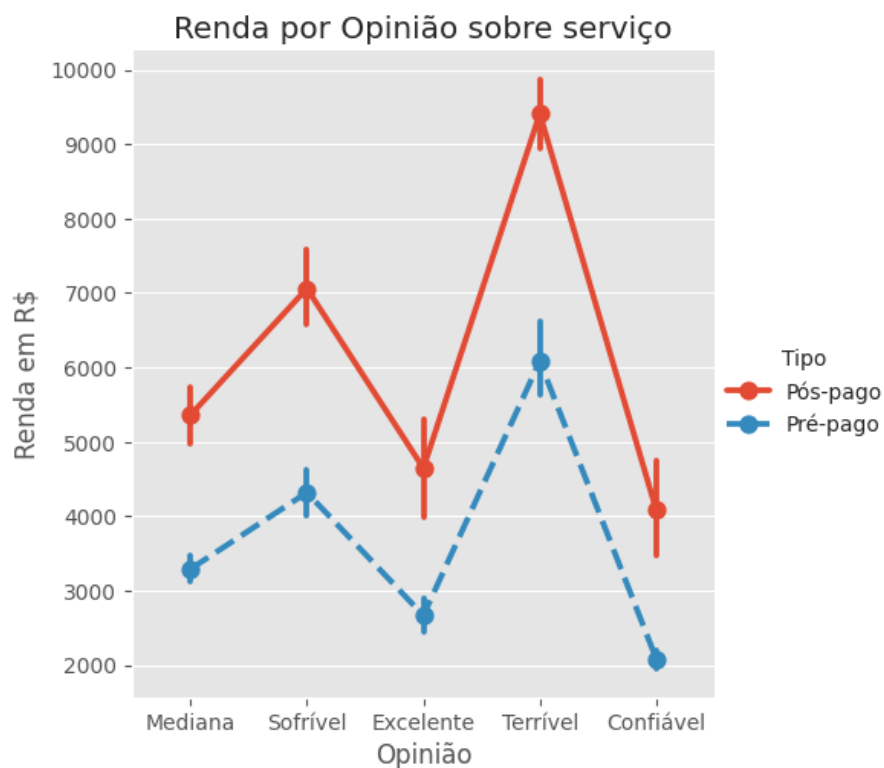
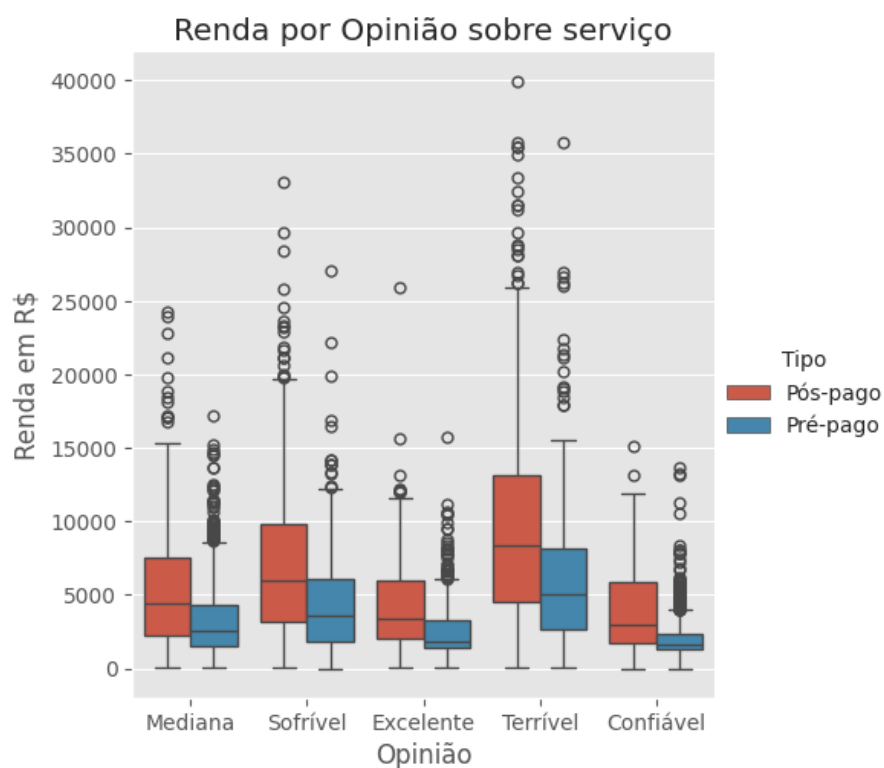
Análise Exploratória dos Dados

Renda

Há uma tendência entre clientes com mais renda reclamarem mais sobre a tarifa, é possível visualizar essa relação por meio dos gráficos de caixa e de ponto.



Outro ponto importante a destacar é que quanto maior a renda do cliente maior a insatisfação pelo serviço.

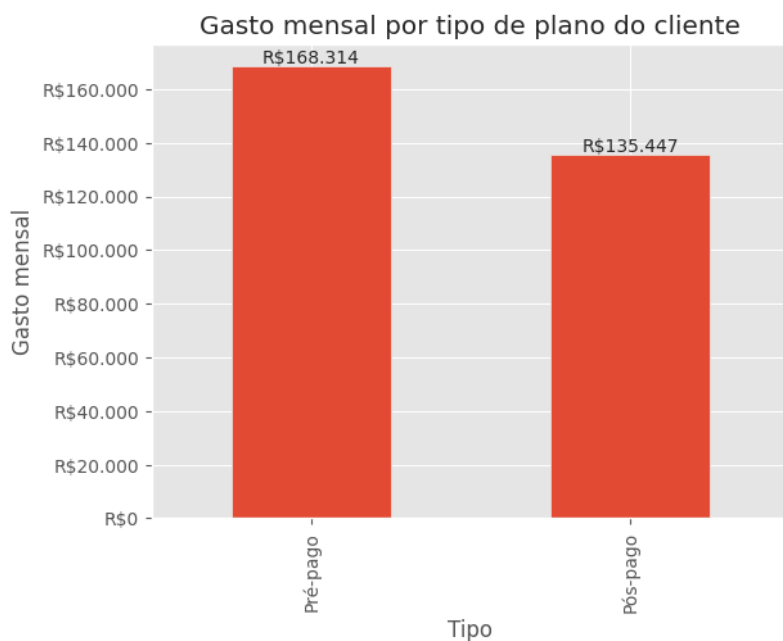
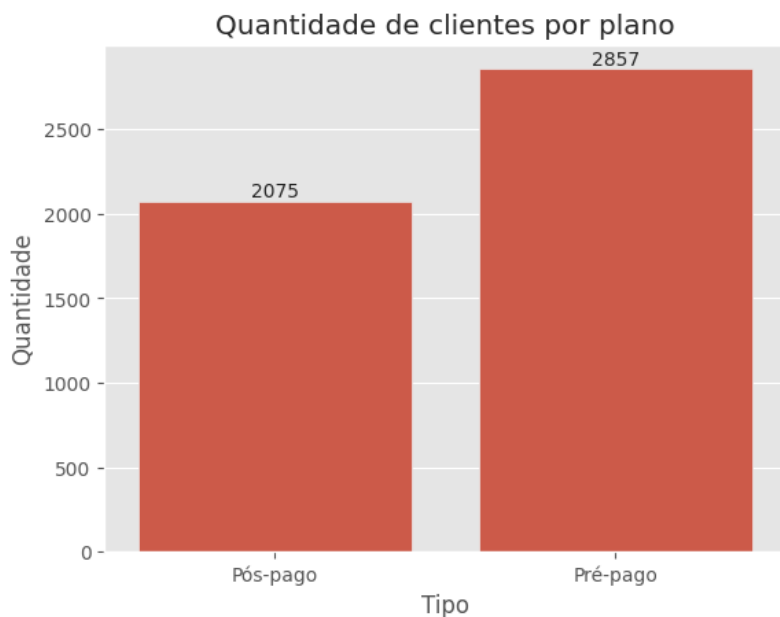


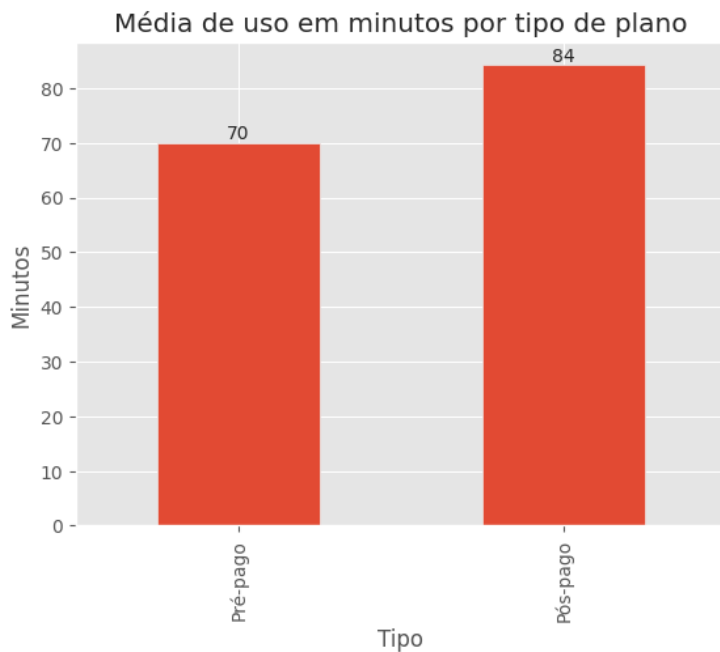
Entretanto, independentemente da renda, a opinião sobre o serviço, no geral, não é boa. A opinião está distribuída em:

Terrível	Sofrível	Mediana	Confiável	Excelente
25,1%	20%	29,2%	15,3%	10,1%

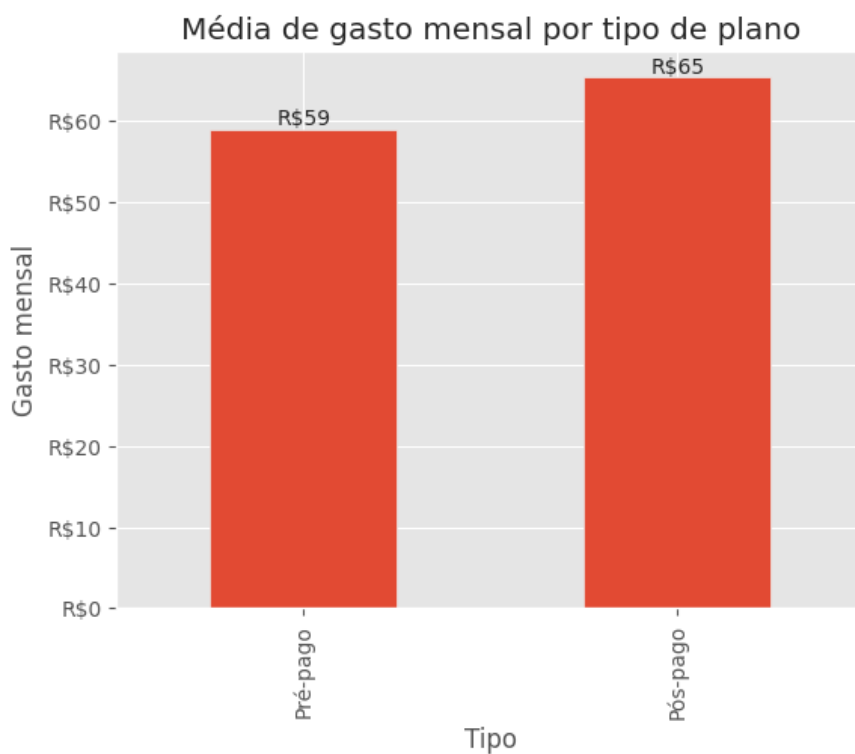
Plano do cliente

A quantidade de clientes com plano pré-pago é superior em, aproximadamente, 15% aos que possuem plano pós-pago. Clientes com plano pré-pago possuem um gasto mensal cerca de 10% maior. Além disso, os clientes com plano pré-pago utilizam em minutos, em média, 10% a menos do serviço do que clientes com plano pós-pago. Visto isso, podemos dizer que os clientes que, somados, gastam menos mensalmente estão utilizando mais do serviço do que aqueles que somam a maior parte dos gastos mensais. Fica mais intuitivo de entender essa inversão pelos gráficos abaixo.



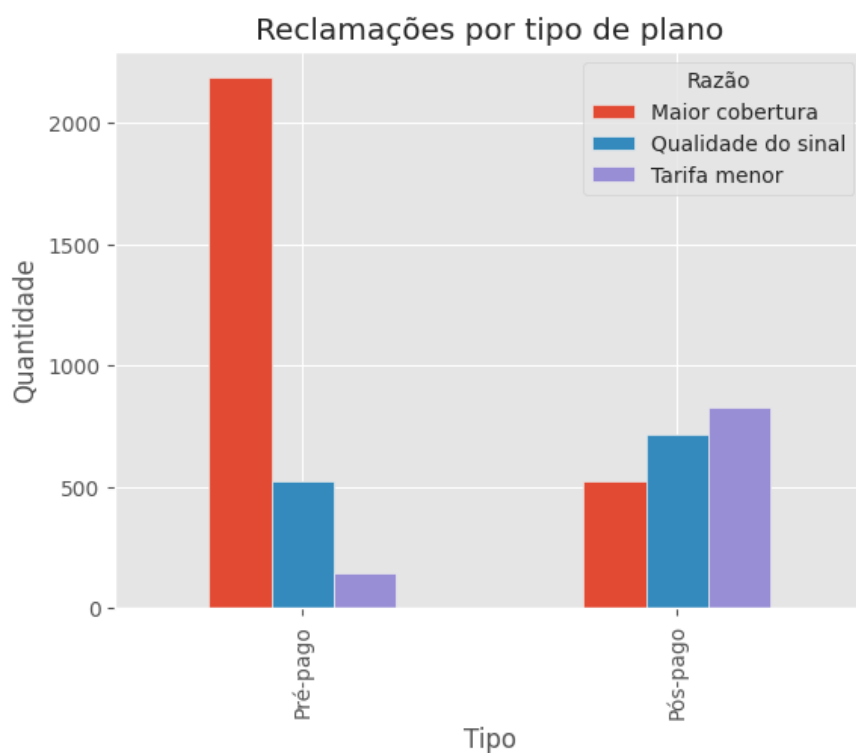


Convém mencionar que quando consideramos o gasto mensal em média, clientes pós-pago pagam, aproximadamente, 4% a mais que clientes pré-pago.

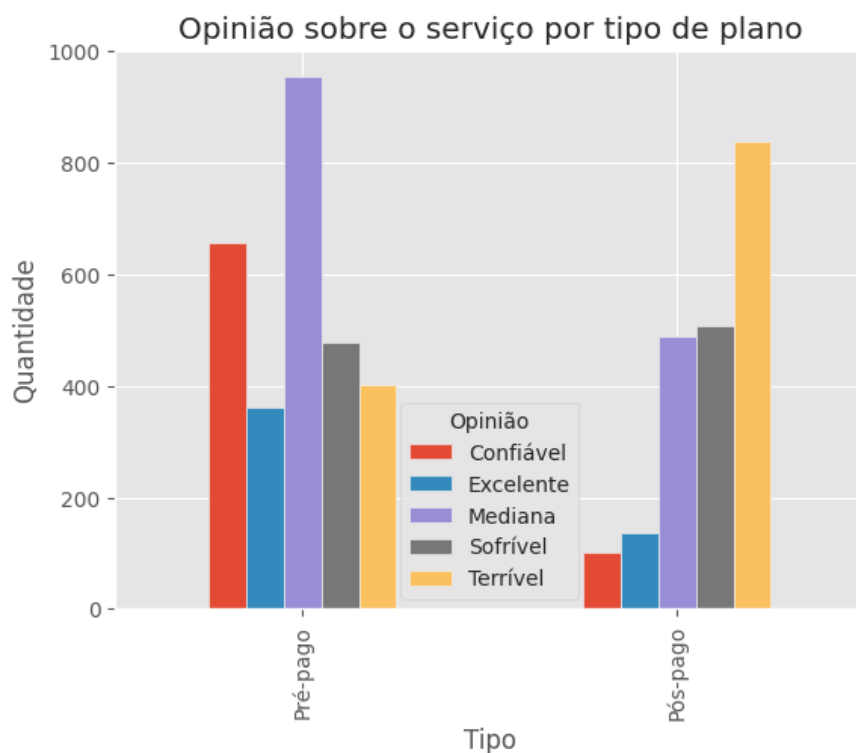


Para que essa análise com médias fosse robusta, realizei dois testes de hipótese com um *nível de significância* de **5%** para garantir que as médias, tanto de **Gasto mensal** como **Uso de minutos**, dos clientes pré-pago fossem realmente menores.

Ademais, é notável, a partir do gráfico abaixo, que os principais motivos de reclamação mudam com o plano.

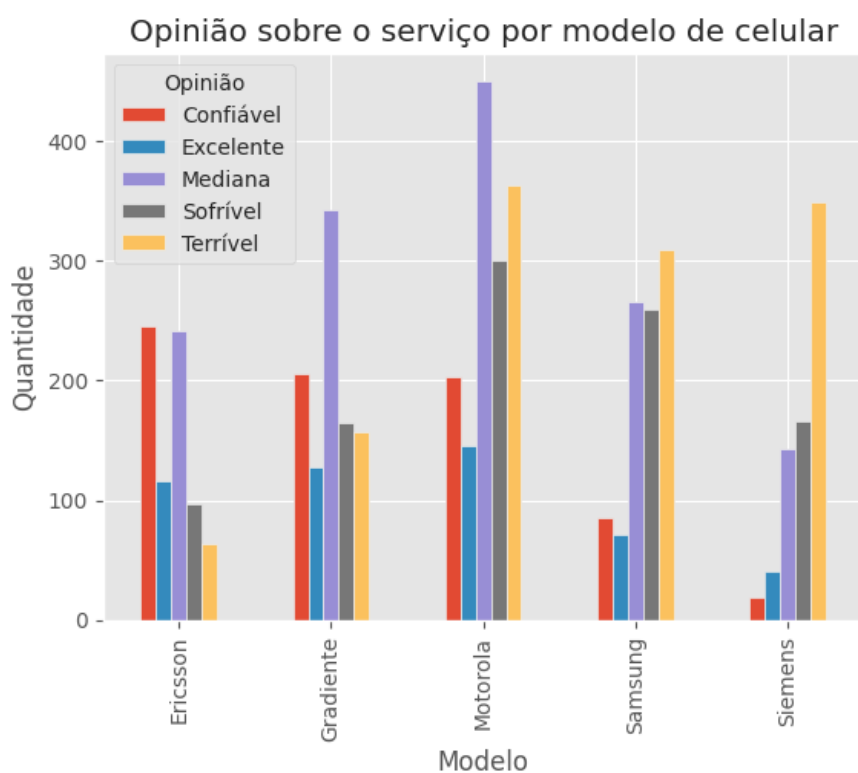
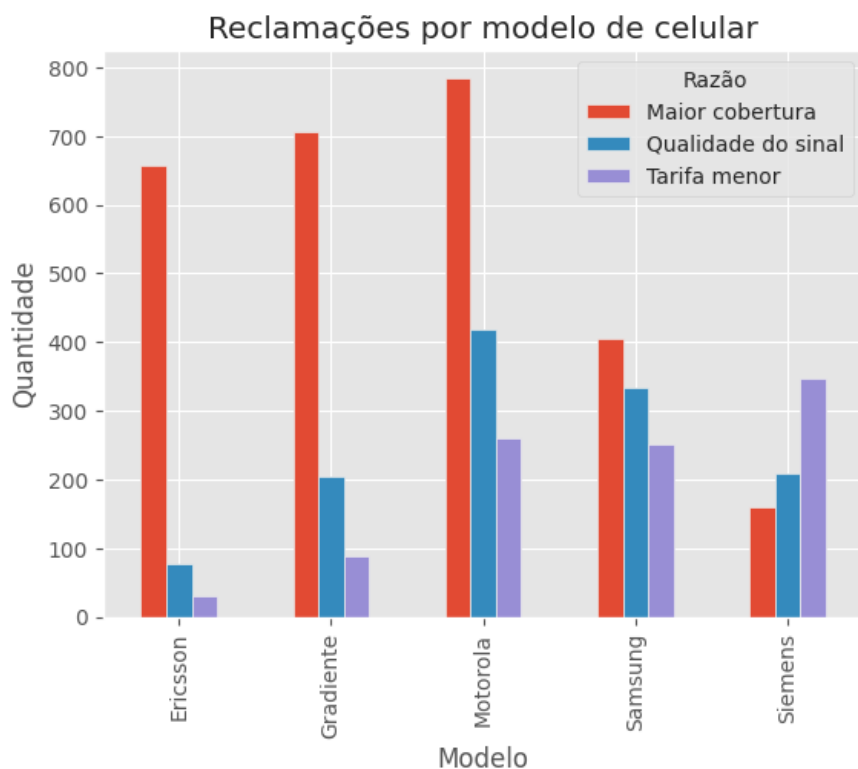


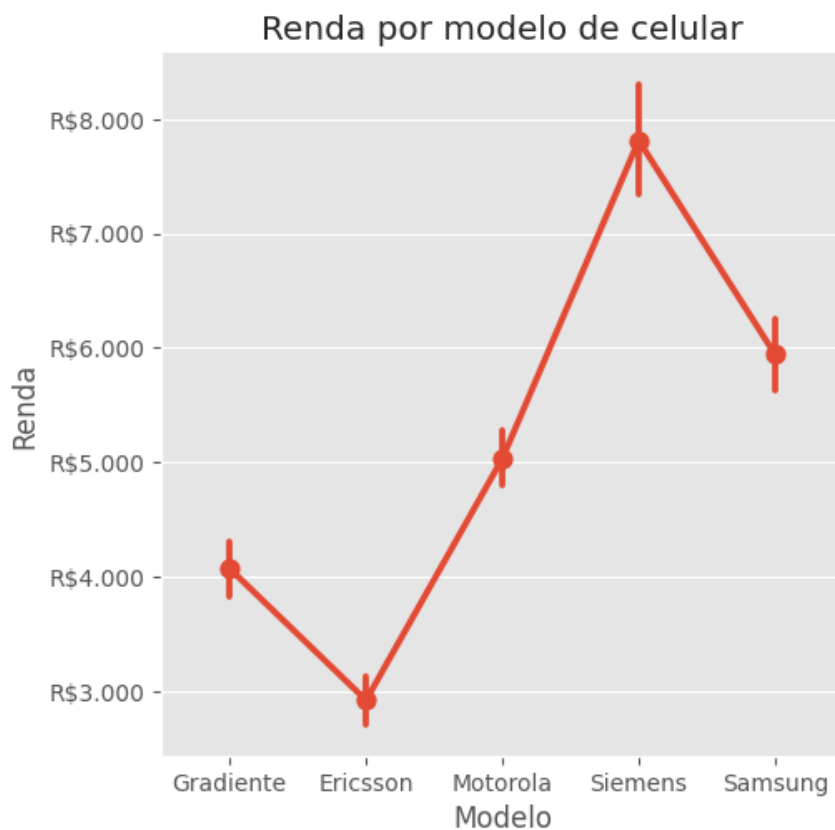
Por último, há um descontentamento muito grande entre a parcela dos clientes pós-pago com mais de 800 clientes avaliando o serviço como **Terrível**.



Modelos de celulares

Clientes com modelos de celulares *Samsung* e *Siemens* possuem as maiores rendas entre os clientes. Acho necessário fazer essa observação porque muito das relações que veremos já podem ser inferidas a partir da renda do cliente. O motivo de reclamação **Maior cobertura** prevalece entre clientes com *Motorola*, *Ericsson* e *Gradiente*. Entretanto, clientes *Samsung* possuem reclamações mais equilibradas e clientes *Siemens* reclamam mais da tarifa. Além disso, clientes *Samsung* e *Siemens* são os que mais possuem opiniões negativas sobre o serviço.



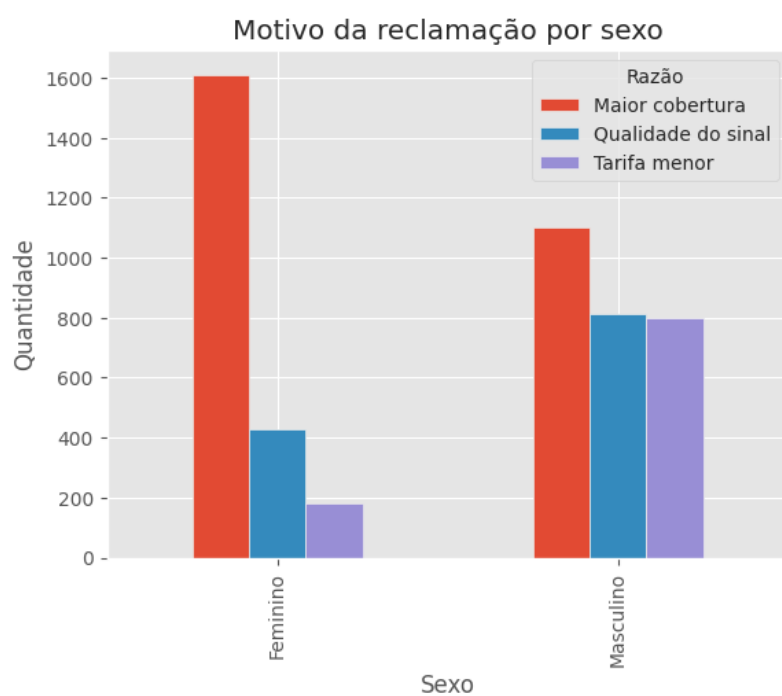
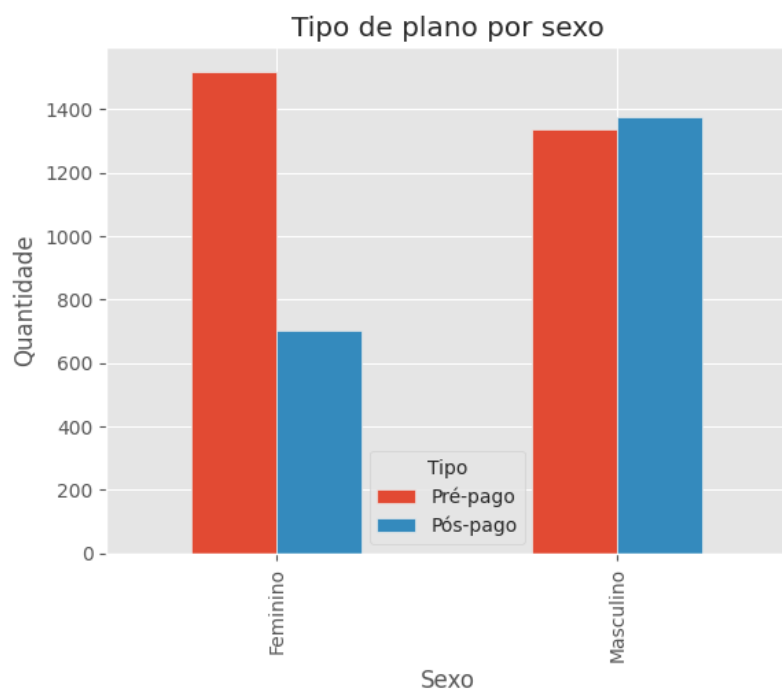


Nota

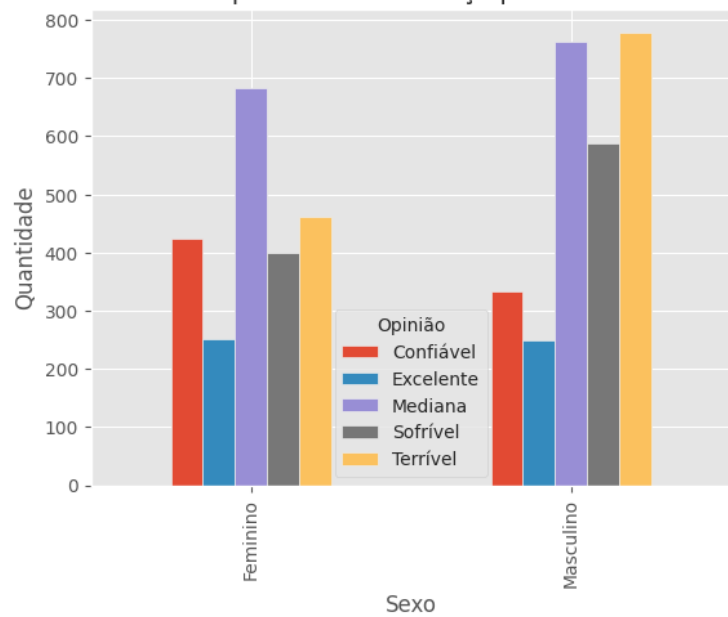
Note que clientes *Samsung* e *Siemens* são os que possuem a maior renda, portanto, eles tendem a serem os mais insatisfeitos com o serviço, reclamarem sobre a tarifa e possuírem plano pós-pago. É possível observar também que clientes *Motorola* são os mais "equilibrados" por possuírem clientes de ambos os planos e de uma renda em torno da média.

Sexo

Apesar de eu não acreditar que o sexo do cliente seja uma variável que cause alguma tendência nos dados, acho importante colocar na análise para mostrar como o **Tipo de plano** influencia as outras variáveis.

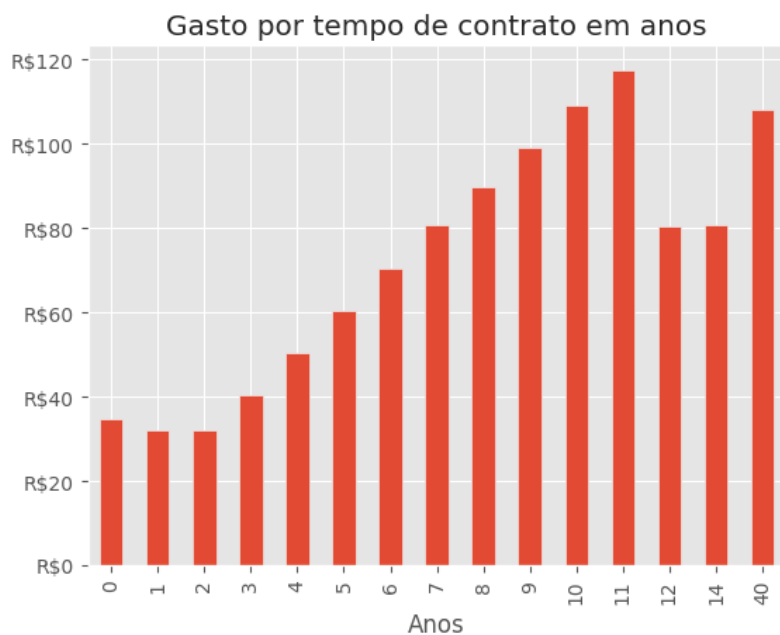
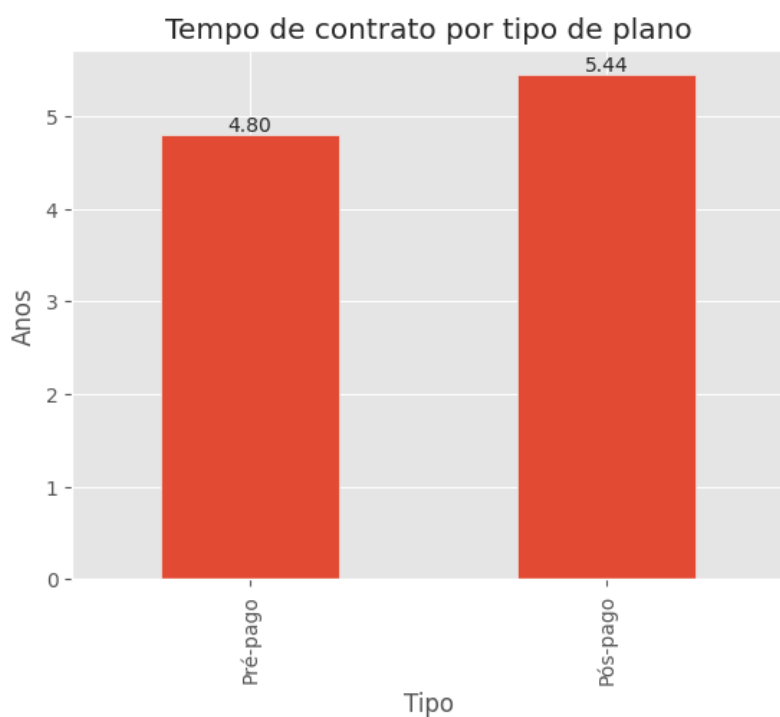


Opinião sobre serviço por sexo



Tempo de contrato

A variável **tempo de contrato** nos traz informações bem diretas. Clientes **pós-pago** possuem, em média, um tempo de contrato **7 meses** a mais que clientes pré-pago. Por fim, há uma tendência de quanto **mais tempo** um cliente tiver de contrato **maior** é o seu gasto mensal.



⚠ Nota

Os clientes com mais de 11 anos de contrato são muito poucos o que acaba influenciando na quebra da correlação acima.

Insights

Após análise dos gráficos acima é possível pontuar alguns insights e destacar algumas tendências:

- Podemos destacar a principal tendência dessa análise que é: clientes com **maior renda** tendem a optarem por **planos pós-pago**.
- Clientes **pós-pago** são os mais **insatisfeitos** com o serviço, possuindo opiniões majoritariamente negativas.
- Os motivos das reclamações são diferentes entre clientes de planos diferentes.
 - Clientes **pós-pago** reclamam mais da **tarifa**. O motivo talvez se dê pelo fato de, apesar de pagarem mensalmente **mais, em média**, o serviço não corresponde em qualidade ao valor cobrado dado as opiniões negativas.
 - Clientes **pré-pago** reclamam mais da **cobertura do sinal**. Não é possível obter da análise um motivo que esclareça bem o porquê dessa reclamação, mas podemos supor algumas coisas. Uma delas é que, por ser um plano **pré-pago**, o recurso alocado para esses clientes talvez seja menor, pois planos pré-pago geralmente não possuem tantos benefícios quanto planos pós-pago. Outro motivo pode ser a da má qualidade dos celulares, visto que clientes **pré-pago** possuem uma menor renda e tendem a ter celulares que possam atrapalhar na obtenção de um sinal. O primeiro motivo me parece mais coerente.
- Clientes **pré-pago** utilizam, em média, **menos o serviço** do que clientes **pós-pago**. Entretanto, ao considerar a **soma dos gastos mensais por tipo de plano**, vemos que clientes pré-pago são os que **mais contribuem**. Apesar do valor de gasto mensal de clientes pós-pago em média **ser maior**, a quantidade de clientes pré-pago **superam** a quantidade de clientes pós-pago, o que explica o fenômeno descrito.
- Há uma correlação positiva entre a variável **Gasto mensal** e **Tempo de contrato**. Quanto mais tempo de contrato um cliente possui maior o gasto mensal. Entretanto, faltam mais dados de clientes com mais de 10 anos de contrato para podermos confiar mais nessa correlação.