

Explicação

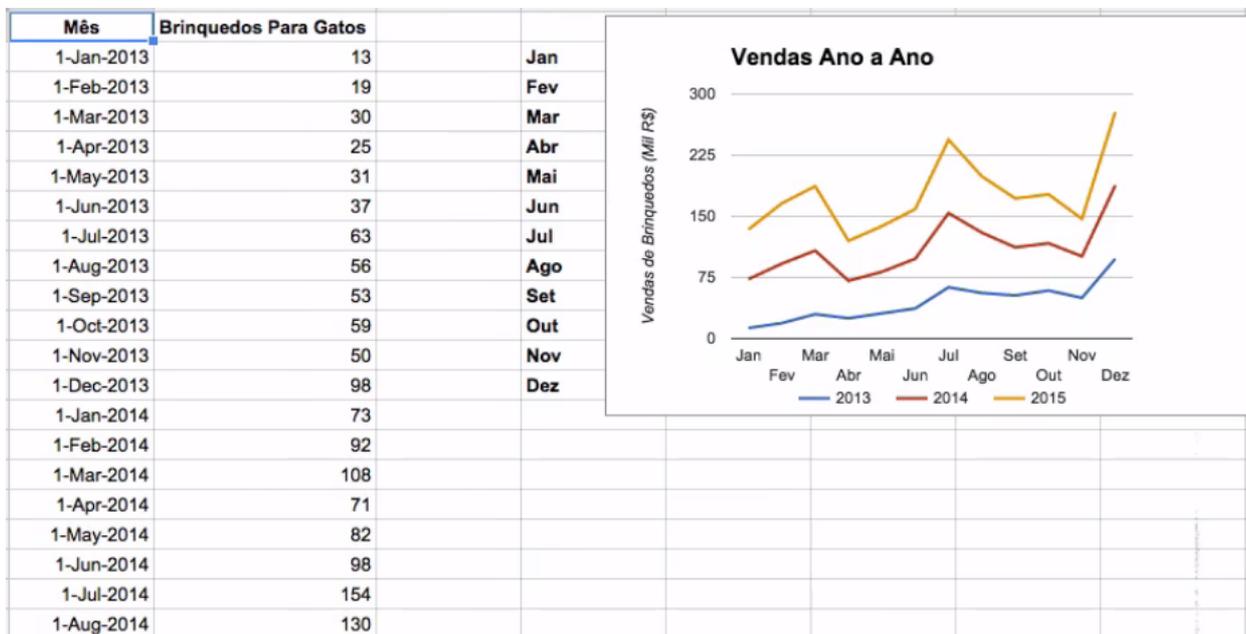
Análise de Vendas

Meu nome é Felipe Farias, eu sou instrutor do curso **Big Data: Uma introdução a data analysis na prática**. Montei o curso com base nas minhas experiências cotidianas e no meu trabalho diário.

Nós trabalharemos com a solução de problemas, de uma *start up* fictícia, chamada **Jumping Cats**. Trata-se de uma loja que oferece serviços e produtos voltados para **gatos**.



Conheceremos este *case*, mas você poderá adaptar o conteúdo para a sua realidade. Nós veremos um exemplo de análise de vendas, veremos como analisar os dados de cada mês, se o crescimento está sendo mais rápido. Veremos **séries temporais** e analisaremos se a empresa está crescendo ano a ano, se existe uma sazonalidade.



Aprenderemos também técnicas sobre **histogramas**, veremos o que é uma **medida de tendência central**, como ela é usada na prática. Iremos demonstrar como apresentar os dados para uma diretoria, como usaremos isso para estimar um **ROI** (Retorno de Investimento).

Também abordaremos o tema de médias, quais são os tipos e como elas se aplicam.

| A | Média Aritimética | B |
|----|-------------------|-------------|
| 50 | 75 | 100 |
| | 25 | 25 |
| | | |
| | Média Geométrica | |
| 50 | 70.71067812 | 100 |
| | 1.414213562 | 1.414213562 |
| | | |
| | Média Harmônica | |
| 50 | 66.66666667 | 100 |
| | 0.005 | 0.005 |

O curso não terá uma rigor matemático, porque o objetivo não é demonstrar fórmulas. Basicamente, iremos usar planilhas de cálculo. No nosso caso, utilizaremos o **Google Spreadsheets**, mas é possível usar o Excel, o LibreOffice-Calc ou outras ferramentas semelhantes. Veremos algumas técnicas bem simples, como por exemplo, multiplicação, divisão, com as quais poderemos pegar os dados brutos e extrair dos mesmos informações. Com isto poderemos dar um retorno para as empresas de forma mais direta, rápida e prática - um objetivo que buscamos atualmente para alcançar o sucesso.

Jumping Cats

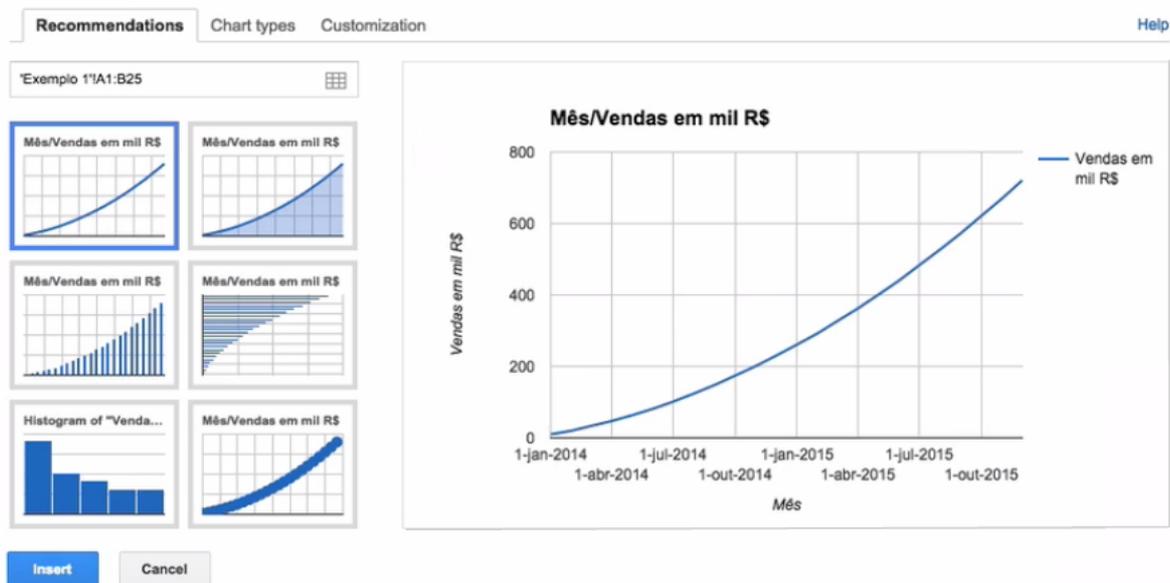
Imagine a situação, você começou a trabalhar na **Jumping Cats**, e o seu chefe pediu que alguns dados importantes fossem analisados. Ele deseja ter um panorama do que está acontecendo na empresa.

| Mês | Vendas em mil R\$ |
|------------|-------------------|
| 1-jan-2014 | 10 |
| 1-fev-2014 | 20 |
| 1-mar-2014 | 33 |
| 1-abr-2014 | 47 |
| 1-mai-2014 | 63 |
| 1-jun-2014 | 81 |
| 1-jul-2014 | 101 |
| 1-ago-2014 | 124 |
| 1-set-2014 | 148 |
| 1-out-2014 | 174 |
| 1-nov-2014 | 201 |
| 1-dez-2014 | 230 |
| 1-jan-2015 | 261 |
| 1-fev-2015 | 293 |
| 1-mar-2015 | 326 |
| 1-abr-2015 | 362 |
| 1-mai-2015 | 400 |
| 1-jun-2015 | 440 |
| 1-jul-2015 | 483 |
| 1-ago-2015 | 527 |
| 1-set-2015 | 573 |
| 1-out-2015 | 621 |
| 1-nov-2015 | 670 |
| 1-dez-2015 | 721 |

Vemos uma planilha com o total de venda em milhares de Reais, a cada mês. Qual é o procedimento que um analista de dados deve ter para analisar de forma simples esta tabela? Os títulos das colunas: "Mês" e "Vendas em mil R\$". Sabemos o que estamos medindo e a unidade. Devemos verificar se a tabela contém todos os dados e meses. Os meses estão em sequência? Sim. Não confie que todos os dados já estão presentes, sempre faça a checagem dos dados. Verifique se eles são coerentes.

A primeira coisa que faremos é criar o gráfico destes números. Clicaremos na opção **Insert**, do Google Spreadsheets, em seguida selecionaremos a opção **Charts**. Uma janela com opções de gráficos será aberta.

Chart Editor



Importante: quando forem construir gráficos, adicione títulos e especifique os eixos, para que você saiba no futuro do que se trata o gráfico. No nosso caso, o nome do nosso gráfico será "Vendas de produtos em mil R\$ por mês".



Existe a opção de mostrar no gráfico os valores alcançados por mês. Mas não é a minha recomendação. Se quisermos consultar as vendas de um mês específico, podemos verificar na nossa tabela. No gráfico, temos a oportunidade de vermos as **formas**. Conseguimos ver com clareza que chegamos de R\$ 0 até R\$ 800mil.

A partir do gráfico que criamos, conseguimos enxergar apenas uma linha relativamente linear. Não conseguimos identificar se durante algum mês houve uma queda nas vendas. Para ter esse tipo de informação, vamos criar uma nova coluna na tabela referente ao **aumento nas vendas**. Nela, iremos calcular o valor da venda de um mês subtraído pelo total do mês anterior. Por exemplo, em fevereiro de 2014 foram vendidos no total R\$ 20mil. Para saber o aumento, iremos subtrair deste o valor de R\$ 10mil, que foi o total de vendas de janeiro do mesmo ano. O resultado será R\$ 10mil.

| Mês | Vendas em mil R\$ | |
|------------|-------------------|----|
| 1-jan-2014 | 10 | 10 |
| 1-fev-2014 | 20 | |

Vamos repetir a operação para todos os meses.

| Mês | Vendas em mil R\$ | |
|------------|-------------------|----|
| 1-jan-2014 | 10 | 10 |
| 1-fev-2014 | 20 | 13 |
| 1-mar-2014 | 33 | 14 |
| 1-abr-2014 | 47 | 16 |
| 1-mai-2014 | 63 | 18 |
| 1-jun-2014 | 81 | 20 |
| 1-jul-2014 | 101 | 23 |
| 1-ago-2014 | 124 | 24 |
| 1-set-2014 | 148 | 26 |
| 1-out-2014 | 174 | 27 |
| 1-nov-2014 | 201 | 29 |
| 1-dez-2014 | 230 | 31 |
| 1-jan-2015 | 261 | 32 |
| 1-fev-2015 | 293 | 33 |
| 1-mar-2015 | 326 | 36 |
| 1-abr-2015 | 362 | 38 |
| 1-mai-2015 | 400 | 40 |
| 1-jun-2015 | 440 | 43 |
| 1-jul-2015 | 483 | 44 |
| 1-ago-2015 | 527 | 46 |
| 1-set-2015 | 573 | 48 |
| 1-out-2015 | 621 | 49 |
| 1-nov-2015 | 670 | 51 |
| 1-dez-2015 | 721 | |

Observe que a célula referente ao último mês ficará vazia, porque não temos um último mês para comparar.



Já temos um gráfico geral, vamos criar um gráfico para o aumento mês a mês das vendas.



Vamos aproximar os gráficos, para comparar o crescimento.



Podemos observar que assim como as vendas, o aumento também tem um crescimento. A cada mês a empresa vende mais e a velocidade também aumenta. Este é o efeito que chamamos de viral, uma pessoa consumiu o produto e recomendou para outra pessoa, que o recomendou novamente...

Como identificar a aceleração das vendas? O crescimento está sendo constante? Responderemos isto, criando uma nova coluna na tabela, que nomearemos como "Aceleração das Vendas". Para fazer este cálculo, faremos um cálculo simples: o aumento de um determinado mês subtraído pelo aumento do mês anterior.

| Mês | Vendas em mil R\$ | Aumento das Vendas | Aceleração das Vendas |
|------------|-------------------|--------------------|-----------------------|
| 1-Jan-2014 | 10 | 10 | 3 |
| 1-Feb-2014 | 20 | 13 | 1 |
| 1-Mar-2014 | 33 | 14 | 2 |
| 1-Apr-2014 | 47 | 16 | 2 |
| 1-May-2014 | 63 | 18 | 2 |
| 1-Jun-2014 | 81 | 20 | 3 |
| 1-Jul-2014 | 101 | 23 | 1 |
| 1-Aug-2014 | 124 | 24 | 2 |
| 1-Sep-2014 | 148 | 26 | 1 |
| 1-Oct-2014 | 174 | 27 | 2 |
| 1-Nov-2014 | 201 | 29 | 2 |
| 1-Dec-2014 | 230 | 31 | 1 |
| 1-Jan-2015 | 261 | 32 | 1 |
| 1-Feb-2015 | 293 | 33 | 3 |
| 1-Mar-2015 | 326 | 36 | 2 |
| 1-Apr-2015 | 362 | 38 | 2 |
| 1-May-2015 | 400 | 40 | 3 |
| 1-Jun-2015 | 440 | 43 | 1 |
| 1-Jul-2015 | 483 | 44 | 2 |
| 1-Aug-2015 | 527 | 46 | 2 |
| 1-Sep-2015 | 573 | 48 | 1 |
| 1-Oct-2015 | 621 | 49 | 2 |
| 1-Nov-2015 | 670 | 51 | |
| 1-Dec-2015 | 721 | | |

Também vamos criar o gráfico desta nova tabela:



É possível perceber uma variação na linha, porém, se analisarmos com cautela veremos que ela vai de 3 a 1. No fim, trata-se de uma reta. Ou seja, a aceleração do aumento pode ser considerada constante. Lembrando que a aceleração mostra o quanto varia o aumento de um mês para outro.

Observe que nós estamos nos aprofundando na análise dos dados. Estamos verificando se os valores têm uma aceleração ou não. Esta é uma técnica simples para apresentar análises, até aqui você já poderá informar que as vendas da empresa estão crescendo, o aumento das vendas também está crescendo - a tendência da curva do gráfico é crescer - e é constante. Se a aceleração fosse negativa, iria significar que estamos perdendo clientes, mas não é o caso do nosso exemplo.

Vale prestar a atenção na magnitude dos dados. O primeiro gráfico sobre as vendas, os valores foram de 0 a 800, e o desenho ficou linear. No gráfico da aceleração, os valores foram de 0 a 4, e a linha revelou uma variação, porém, comparando com a análise interior, podemos afirmar que a aceleração é constante.

Aprofundando nas análises

Agora, vamos conhecer outros dados da **Jumping Cat**. A empresa tem uma *newsletter*, em que informa novidades dos produtos e informações sobre como criar gatos da melhor forma. A newsletter é enviada para os clientes, como uma estratégia de aumentar a fidelidade a sua marca.

Abaixo temos número de assinaturas da newsletter feitos em cada mês:

| A | B |
|------------|--------------------------|
| Mês | Assinatura da Newsletter |
| 1-Jan-2014 | 0 |
| 1-Feb-2014 | 10 |
| 1-Mar-2014 | 21 |
| 1-Abr-2014 | 34 |
| 1-Mai-2014 | 49 |
| 1-Jun-2014 | 66 |
| 1-Jul-2014 | 85 |
| 1-Ago-2014 | 107 |
| 1-Set-2014 | 132 |
| 1-Out-2014 | 160 |
| 1-Nov-2014 | 190 |
| 1-Dez-2014 | 221 |
| 1-Jan-2015 | 252 |
| 1-Feb-2015 | 283 |
| 1-Mar-2015 | 313 |
| 1-Abr-2015 | 341 |
| 1-Mai-2015 | 367 |
| 1-Jun-2015 | 390 |
| 1-Jul-2015 | 410 |
| 1-Ago-2015 | 426 |
| 1-Set-2015 | 437 |
| 1-Out-2015 | 444 |
| 1-Nov-2015 | 449 |
| 1-Dez-2015 | 454 |

Já verificamos se a tabela inclui todos os meses e se a coluna de assinatura tem todos os valores. Em seguida, iremos criar um gráfico de linhas, mas você pode optar por barras.

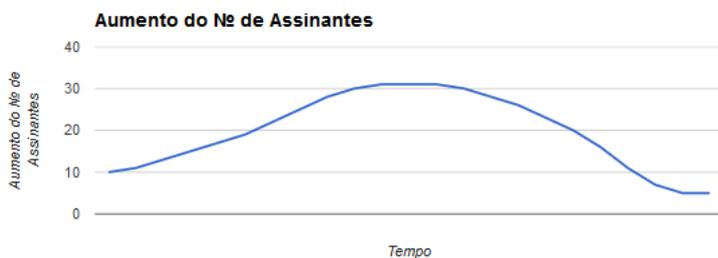


Observe que o gráfico apresenta as informações de forma bastante clara e qualquer pessoa pode compreendê-las. Quando um gráfico fica incompreensível fora do contexto, ele tem problemas.

A linha do gráfico é aparentemente linear. Também temos a impressão de que próximo da data de 1 de outubro de 2015, a linha tem um crescimento menor. Mas para termos certeza, usaremos novamente a técnica de especificar valores na tabela. Adicionaremos uma coluna na tabela, que receberá o nome "Aumento no Número de Assinantes".

| C | |
|---|-----------------------------|
| | Aumento no No de Assinantes |
| | 10 |
| | 11 |
| | 13 |
| | 15 |
| | 17 |
| | 19 |
| | 22 |
| | 25 |
| | 28 |
| | 30 |
| | 31 |
| | 31 |
| | 30 |
| | 28 |
| | 26 |
| | 23 |
| | 20 |
| | 16 |
| | 11 |
| | 7 |
| | 5 |
| | 5 |

Nós subtraímos o número de assinantes de um determinado pelo número do mês anterior. Também iremos criar o gráfico do aumento.



Agora, temos um dado interessante. O número de assinantes começou crescendo, chegou a um pico de 30 novas assinatura por mês, depois o número começou a diminuir. A diferença na linha do número de assinantes começa a ser mais sutil de um período a outro. Observando o gráfico do aumento, a tendência é que o número de assinantes não cresça tanto.

Vamos calcular a aceleração do número de assinantes.

| Aceleração do Nº de Assinantes |
|--------------------------------|
| 1 |
| 2 |
| 2 |
| 2 |
| 2 |
| 3 |
| 3 |
| 3 |
| 2 |
| 1 |
| 0 |
| 0 |
| -1 |
| -2 |
| -2 |
| -3 |
| -3 |
| -4 |
| -5 |
| -4 |
| -2 |
| 0 |

Em seguida, iremos criar o gráfico também.

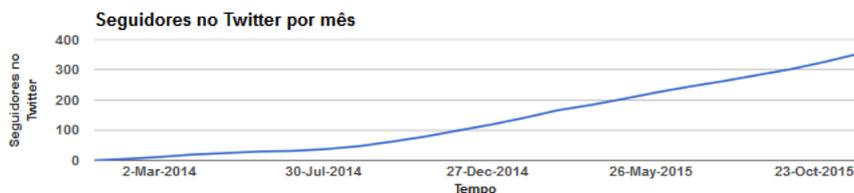


Como os valores do número de assinantes caiu, a aceleração ficou negativa. Os gráfico indicam que o número de novos assinantes deve estabilizar ou crescer pouco. Quando observamos os três gráfico, percebemos comportamentos diferentes.

Vamos usar mais um exemplo. Iremos analisar o número de **seguidores do Twitter**.

| Mês | Seguidores no Twitter |
|------------|-----------------------|
| 1-Jan-2014 | 0 |
| 1-Feb-2014 | 5 |
| 1-Mar-2014 | 11 |
| 1-Apr-2014 | 19 |
| 1-May-2014 | 24 |
| 1-Jun-2014 | 29 |
| 1-Jul-2014 | 31 |
| 1-Aug-2014 | 37 |
| 1-Sep-2014 | 47 |
| 1-Oct-2014 | 62 |
| 1-Nov-2014 | 79 |
| 1-Dec-2014 | 99 |
| 1-Jan-2015 | 119 |
| 1-Feb-2015 | 142 |
| 1-Mar-2015 | 165 |
| 1-Apr-2015 | 183 |
| 1-May-2015 | 203 |
| 1-Jun-2015 | 225 |
| 1-Jul-2015 | 244 |
| 1-Aug-2015 | 262 |
| 1-Sep-2015 | 282 |
| 1-Oct-2015 | 301 |
| 1-Nov-2015 | 325 |
| 1-Dec-2015 | 351 |

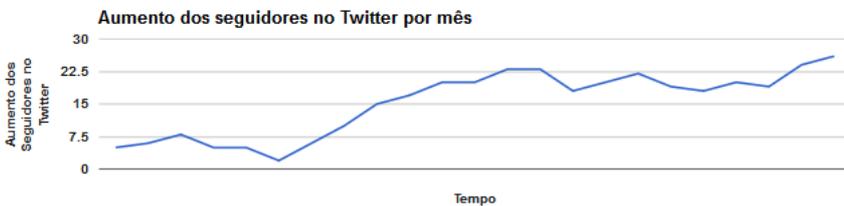
Após conferirmos na tabela que todos os dados da tabela foram preenchidos completamente, criaremos o gráfico.



Analisando o gráfico, podemos observar que até o mês de Agosto podemos observar um crescimento menor. Depois, vemos um crescimento rápido. Iremos calcular o aumento de cada mês.

| Aumento dos Seguidores | |
|------------------------|--|
| 5 | |
| 8 | |
| 8 | |
| 5 | |
| 5 | |
| 2 | |
| 6 | |
| 10 | |
| 15 | |
| 17 | |
| 20 | |
| 20 | |
| 23 | |
| 23 | |
| 18 | |
| 20 | |
| 22 | |
| 19 | |
| 18 | |
| 20 | |
| 19 | |
| 24 | |
| 28 | |

Vamos visualizar no gráfico o aumento.



No gráfico, vemos nitidamente que existe um ponto de em que inicio o aumento. É provável, que alguma ação tenha sido realizada, como uma campanha de Marketing ou a loja ganhou destaque. Porém, este momento não se repete nos gráficos anteriores, como os da Newsletter, por exemplo. Uma opção é conversar com o responsável pelo setor de marketing e investigar o que pode ter afetado positivamente os números.

Iremos desenhar o gráfico que mostra a **aceleração do número de seguidores**. Primeiramente, vamos detalhar na tabela os números em cada mês e depois, mostraremos visualmente no gráfico.

| Aceleração de Seguidores |
|--------------------------|
| 1 |
| 2 |
| -3 |
| 0 |
| -3 |
| 4 |
| 4 |
| 5 |
| 2 |
| 3 |
| 0 |
| 3 |
| 0 |
| -5 |
| 2 |
| 2 |
| -3 |
| -1 |
| 2 |
| -1 |
| 5 |
| 2 |



Como estamos trabalhando com séries temporais, geralmente, o eixo horizontal sempre será referente ao tempo.

Ao analisarmos o gráfico da aceleração, percebemos que houve uma variação. Existe um momento de aceleração no aumento e depois, ele ficou mais constante. Quando o número de assinantes fica estável, a aceleração se aproximará do 0. Vemos como separar as informações facilita a análise dos dados.

A técnica utilizada nos exemplos, se formos estudar matematicamente é uma **derivada numérica**. Temos um valor (Seguidores do Twitter), depois temos a **primeira derivada** (Aumento do Seguidores) e a **segunda derivada** (Aceleração de Seguidores).

Fique atento, que esta técnica pode ser utilizada para **séries temporais**, ou seja, valores que estão sendo analisados ao longo do tempo. Mais adiante, veremos outras análises.

Vimos alguns exemplos que nos dão exemplos de como podemos enriquecer a análise dos valores.

