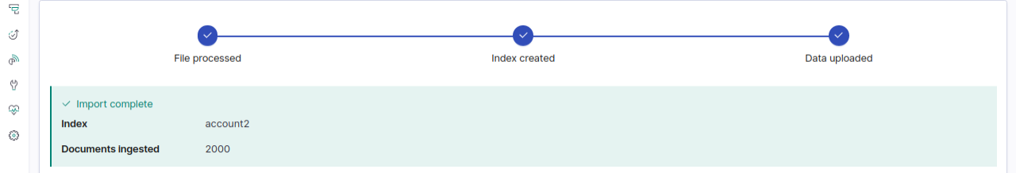
**2.1 - Importando novos conjuntos de dados**

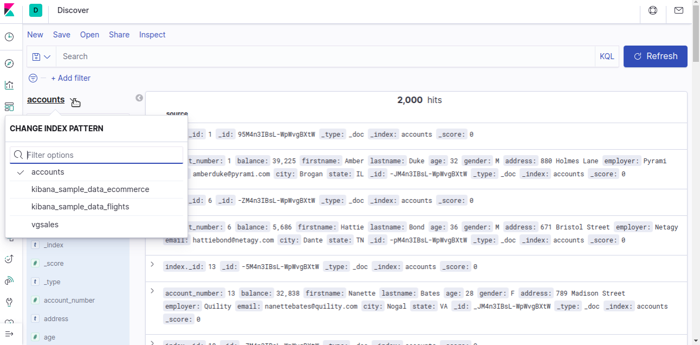
Primeiramente, vamos carregar um novo conjunto de dados no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana). Você pode baixar o conjunto de dados que iremos utilizar neste conteúdo através deste link: [download.elastic.co/demos/](https://leadfortaleza.com.br/ead2pcd/conteudo/tmp/myopenolat_1_103230062079325/aula/arquivos/accounts.zip)[kibana](https://leadfortaleza.com.br/ead/glossary/Kibana)/gettingstarted/8.x/accounts.zip. Salve o arquivo compactado na pasta de sua escolha. Descompacte e você irá extrair um arquivo chamado accounts.json. Esse arquivo contém dados de contas bancárias fictícias de um banco nos Estados Unidos, em formato JSON. Na seção 1.3 você carregou dados de exemplo no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) a partir da tela inicial. Para carregar os novos dados no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) siga os passos:

1. Acesse a página inicial do [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) (utilizando o endereço <http://localhost:5601/>" no navegador).
2. Acesse o link **Import** **a CSV, NDJSON, or** **log file**, disponível no centro da tela inicial do [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana).
3. Escolha a opção **Select or drag and drop a file**. Note que o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) permite adicionar arquivos de dados nos formatos JSON e CSV. Porém, internamente ele sempre trabalha com arquivos JSON, então caso seja carregado um arquivo CSV, o próprio [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) faz a conversão para JSON.
4. Selecione o arquivo accounts.json.
5. Será aberta uma tela mostrando um sumário dos dados, em sua forma bruta e estatísticas dos seus atributos. Acione o botão Import para importar os dados.
6. No campo **Index name**, escolha um nome para o conjunto de dados. Por exemplo, accounts. Esse será o nome que o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) irá se referir a esses dados. Lembre-se que no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) os conjuntos de dados são chamados index.
7. Acione o botão Import.

Deverá aparecer na tela uma imagem semelhante à figura a seguir. Isso significa que os dados foram importados corretamente!

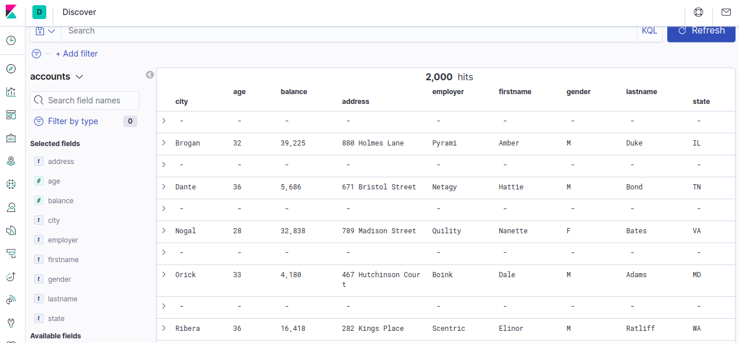
Figura 22 - Dados importados com sucesso

Antes de começar a criar o [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard), vamos explorar um pouco esses dados utilizando a ferramenta Discover. Acione o botão do Discover na barra de navegação e selecione o index que você criou, acionando-o. Deverá aparecer uma tela semelhante à figura a seguir, neste caso, com o index nomeado como accounts. Confira.

Figura 23 - Explorando o index accounts

Note que não é mostrada a distribuição dos dados ao longo do tempo porque esse conjunto de dados não possui informação temporal. Apesar disso, é possível conhecer os atributos dos dados, observando o painel de **Available fields** ou observando os dados diretamente em formato JSON. Observando os dados em JSON, você irá notar que o primeiro documento possui atributos diferentes do segundo e os tipos de documentos se alternam. Isso é apenas a forma como este JSON está estruturado e não irá interferir na visualização dos dados. Considere como relevantes apenas os documentos que possuem os atributos: account\_number, balance, firstname, lastname, age, gender, address, employer, email, city, state.

Podemos selecionar os atributos relevantes e observá-los em uma tabela, assim como já foi feito. Selecione os atributos citados anteriormente e deverá ser mostrada uma tabela semelhante à mostrada na figura a seguir.

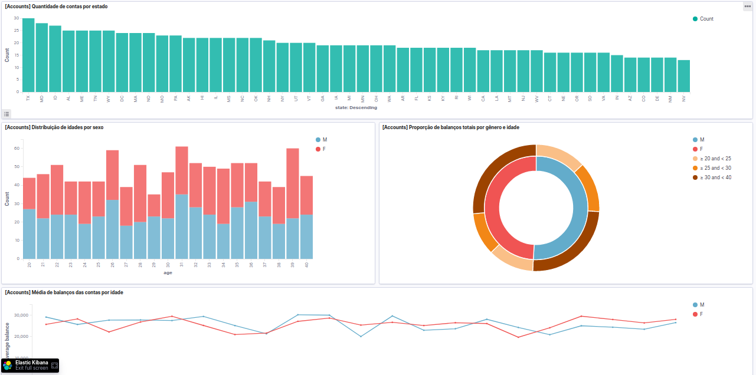
Figura 24 - Explorando atributos do index accounts

Pode-se notar que este conjunto de dados contém tanto atributos quantitativos como categóricos sobre os clientes, como nome, gênero, idade, cidade, estado e balanço da conta (saldo).

A partir dessas informações podemos começar a fazer algumas perguntas para conhecer melhor os clientes desse banco fictício. Esse banco possui mais clientes em quais estados? Qual a distribuição das idades dos clientes? Essa distribuição é a mesma para ambos os gêneros? Existe uma diferença nos balanços das contas de homens e de mulheres? Existe uma diferença por idade? Podemos responder a todas essas perguntas através de um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard), como será feito na próxima página.

## 2.2 - Criando visualizações

Para criar um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) é necessário criar os painéis de visualizações separadamente e adicioná-los ao [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard). Você já deve ter percebido que o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) é uma ferramenta muito ampla, que permite a criação de diversos tipos de visualizações, com muitas possibilidades de customização. Exemplificar todas as formas de criar visualizações seria inviável em uma aula. Porém, você perceberá que todas as visualizações seguem a mesma lógica de criação, utilizando Metrics e Buckets, e isto é o mais importante de se entender. Nesse conteúdo, vamos criar exemplos de visualizações para responder às perguntas feitas anteriormente, então vamos criar um gráfico de barras, um gráfico de barras empilhadas, um gráfico de rosca e um gráfico de linha. Para detalhes específicos sobre outros tipos de gráficos, consulte a documentação oficial do [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana), indicada nas referências. Ao final desse conteúdo, você terá construído um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) similar ao mostrado na figura a seguir.

Figura 25 - [Dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) com os dados das contas bancárias

Para criar visualizações, acione a opção Visualize (terceiro ícone da barra de navegação, no formato de um gráfico). Ao acionar essa opção, você abrirá uma página contendo todas as visualizações criadas no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana), semelhante ao mostrado na figura abaixo. Para criar uma nova visualização, acione o botão **Create visualization**. Irá surgir uma janela com diversos ícones para diferentes tipos de visualização, conforme ilustrado na figura a seguir. Vamos começar respondendo a pergunta de quantas contas existem em cada estado. Uma boa forma de visualizar essas quantidades é através de um gráfico de barras. Para isso, selecione a opção Vertical Bar (ícone de barras verticais, localizado na parte de baixo da tela de visualizações). Irá surgir as opções dos conjuntos de dados que você carregou no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) (index). Selecione o index referente ao conjunto de dados que você carregou das contas.

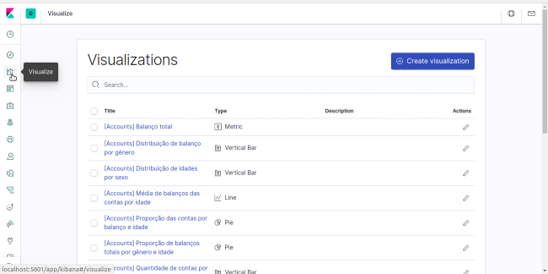
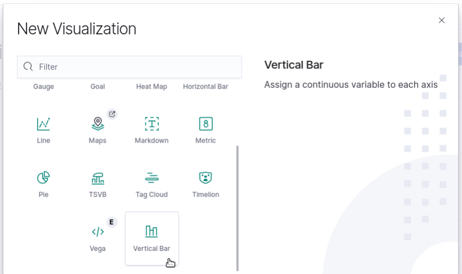
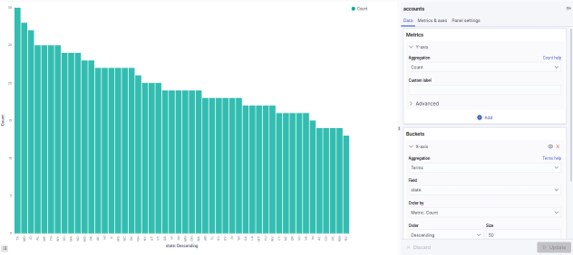
Figura 26 - página de visualizações

Figura 27 - criação de uma nova visualização

Você terá aberto a página de criação de visualizações. Lembra quando você editou uma visualização no tópico anterior? É a mesma página. Como você selecionou o tipo de visualização como gráfico de barras verticais, inicialmente haverá apenas uma única barra, contando o total de documentos. Isso porque não foi selecionada nenhuma métrica e nenhum bucket. Lembre-se que as métricas selecionam os valores que serão representados na visualização e os buckets indicam como os dados serão agrupados em torno de categorias. Nesse caso, queremos que as barras indiquem a quantidade de contas e cada barra represente um estado. Assim, vamos usar como métrica a contagem e como bucket os estados. Para isso, siga o passo a passo:

1. No painel **Metrics**, expanda **Y-axis**.
2. Configure **Aggregation** como **Count** (ou seja, contagem será representada no eixo y)
3. No painel **Buckets**, expanda **X-axis**.
4. Configure **Aggregation** como **Terms**.
5. Configure **Fields** como **state**. Ou seja, os dados serão agregados pelos estados e representados no eixo x.
6. Configure **Order by** como **Descending** e **Size** como 50 (os dados serão ordenados de maneira descendente e serão mostrados os primeiros 50 - que é a quantidade de estados nos Estados Unidos)
7. Acione o botão **Update**. O resultado deverá ser similar ao ilustrado na figura apresentada logo a seguir.
8. Por fim, salve a sua visualização acionando em **Save**, no canto superior esquerdo. De preferência, utilize um nome que remeta ao index dos dados que geraram a visualização, por exemplo: [Accounts] Quantidade de contas por estado.

Figura 28 - criando um gráfico de barras

Você já está começando a compreender a lógica de construção de visualizações no Kibana? Deve-se sempre pensar nos valores que se quer representar (Metrics) e em como esses valores serão divididos (Buckets). É possível utilizar em uma visualização mais de um Bucket, para que existam subdivisões na visualização. Vamos exemplificar isso criando mais uma visualização respondendo às seguintes perguntas: Qual a distribuição das idades dos clientes? Essa distribuição é a mesma dividindo por gênero? Como você deve saber, uma forma de representar uma distribuição é através de um histograma. Para criar a nova visualização, abra novamente a página **Visualize**, escolha a opção **Create Visualization**, selecione novamente **Vertical Bar** e escolha o index que representa as contas bancárias. Na página de criação de visualizações, siga o passo a passo:

1. No painel **Metrics**, expanda **Y-axis**.
2. Configure **Aggregation** como **Count** (ou seja, contagem será representada no eixo y)
3. No painel **Buckets**, expanda **X-axis**.
4. Configure **Aggregation** como **Histogram**.
5. Configure **Fields** como age. Ou seja, os dados serão agregados por intervalos de idade e representados no eixo x.
6. Em **Minimal interval**, você pode configurar o tamanho do bin, ou seja, o intervalo entre as idades. Nesse caso, você pode utilizar o valor 1.
7. Acione o botão **Update**. O resultado deverá ser similar ao ilustrado na figura a seguir.

Figura 29 - criando um histograma

O gráfico criado mostra a distribuição geral das idades dos clientes. Nota-se que é uma distribuição bem uniforme. E se quisermos saber a distribuição de idades por gênero? Vamos conferir.

Existem muitas formas de mostrar as duas distribuições na mesma visualização. Todas elas envolvem adicionar um novo bucket para sub agrupar os dados. Vamos utilizar uma abordagem bem compacta e que permite uma boa comparação chamado gráfico de barras empilhadas, onde divide-se cada barra pelas categorias que se deseja agrupar. Para criar um gráfico de barras empilhadas, siga este passo a passo:

1. No painel de **Bucket**, selecione **Add**.
2. Selecione a opção **Split series**. Essa opção diz que o novo bucket irá subdividir os elementos da visualização (nesse caso, as barras). A opção outra opção, **Split chart**, subdivide o painel em diversos gráficos.
3. Em **Sub aggregation**, selecione **Terms**. Isso significa que os dados serão subdivididos por um atributo categórico.
4. Em **Field**, selecione gender, indicando que os dados serão subdivididos pelo gênero.
5. Acione o botão **Update**.

O resultado deverá ser igual a este:

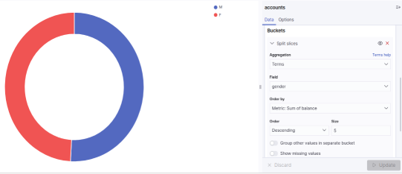
Figura 30 - criando um histograma

Salve essa visualização com o nome apropriado, por exemplo [Accounts] Distribuição de idades por sexo. Observe que cada barra ficou dividida mostrando a proporção de clientes masculinos e femininos em cada idade. Percebe-se que, de forma geral, as distribuições de idade por gênero são bem similares. Saiba que você pode checar os valores de cada elemento da visualização, selecionando-os. Um detalhe importante é que o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) escolhe as cores previamente, mas elas podem ser customizadas ao selecionar a legenda que indica a classe da cor e selecionando a cor apropriada. Quando você utilizar cores para representar categorias, é útil manter o mesmo padrão em todo o [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard).

Iremos agora criar um gráfico de rosca que mostre a proporção entre o total dos balanços para homens e mulheres para descobrir se existem grupos que concentram a maior parte do dinheiro. Para criar um gráfico de rosca, abra a página **Visualize**, selecione **Create Visualization**, selecione **Pie** e escolha o index que representa as contas bancárias. Na página de criação de visualizações, haverá um gráfico de rosca sem nenhuma divisão. Embora o gráfico de rosca seja diferente de um gráfico de barras, a lógica para construir os dois no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) é a mesma: em Metrics escolhe-se o valor que se deseja representar e em Buckets como esse valor será subdividido. Siga o passo a passo:

1. No painel **Metrics**, expanda **Slice size**.
2. Configure **Aggregation** como **Sum**.
3. Configure **Field** como balance. Ou seja, as fatias serão proporcionais ao somatório dos balanços.
4. No painel **Buckets**, selecione **Add**.
5. Selecione **Split slices**.
6. Configure **Aggregation** como **Terms**.
7. Configure **Field** como gender. Ou seja, os dados serão agregados pelo gênero e divididos no gráfico de rosca pela proporção do total de balanços em cada gênero.
8. Acione o botão **Update**.
9. Altere as cores para manter o padrão utilizado na visualização anterior.

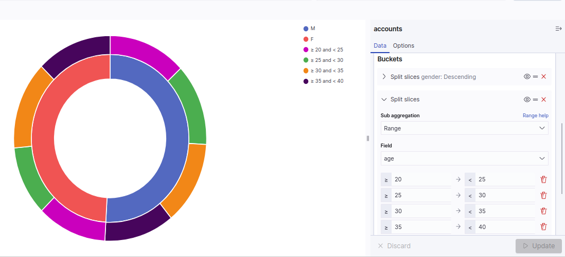
O resultado deverá ser similar a este:

Figura 31 - criando um gráfico de rosca

O gráfico anterior mostra que o total de balanços está dividido de forma muito próxima entre os gêneros. E se quisermos saber a proporção entre as idades? Uma maneira é criar um novo bucket que irá subdividir o gráfico criado em faixas etárias. Siga o passo a passo:

1. No painel de Bucket, selecione **Add**.
2. Selecione a opção Split slices. Essa opção diz que o novo bucket irá subdividir os elementos da visualização (nesse caso, as fatias).
3. Em **Sub aggregation**, selecione **Range**. Isso significa que os dados serão subdivididos por intervalos de um atributo numérico.
4. Em **Field**, selecione age, indicando que os dados serão subdivididos por faixas etárias.
5. Relembre do histograma onde vimos que a idade dos clientes vai de 20 a 40 anos. Podemos então definir os intervalos: 20 >= → < 25; 25 >= → < 30; 30 >= → < 35; 35 >= → < 40, Isso significa que os dados serão subdivididos nessas faixas etárias.
6. Acione o botão **Update**.

O resultado deverá ser similar a este:

Figura 32 - criando um gráfico de rosca com subdivisão

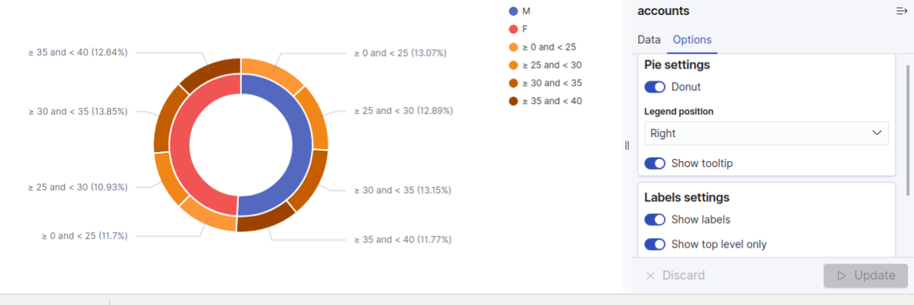
Salve a visualização com um nome adequado, como [Accounts] Proporção do balanço total por gênero e faixa etária.

Note que as fatias mais externas representam a proporção que cada faixa etária possui do somatório total. Note também que são proporções quase iguais para cada gênero e faixa etária. Essa coincidência pode ser explicada porque não são dados reais, e sim dados gerados para estudo.

Esse gráfico pode ser melhorado com alguns ajustes. O [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) escolherá cores automaticamente para cada faixa etária, mas como existe uma relação de ordem entre as idades e podemos escolher cores que reflitam isso, usando uma escala de intensidade. Selecione cores para as faixas etárias de uma mesma coluna das cores disponíveis, escolhendo a mais clara para a menor faixa etária e a mais escura para a maior faixa etária.

Como você já sabe, os valores exatos de cada fatia pode ser verificado passando o mouse em cima. Porém, você pode configurar a visualização para exibir os valores por padrão. Para isso selecione **Options** e ative a opção **Show Labels**. O painel Options possui muitas opções interessantes para configurar visualizações.

O resultado após as alterações deverá ser similar à figura a seguir. Não se esqueça de salvar a visualização.

Figura 33 - gráfico de rosca aprimorado

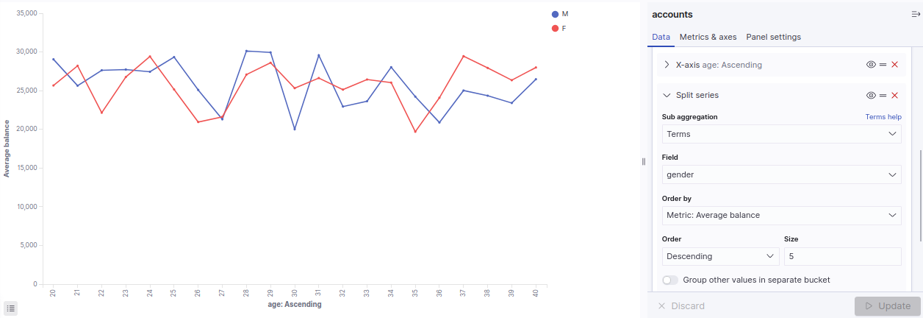
Por fim, vamos criar um gráfico de linha que mostre se há alguma relação entre a idade e o valor médio de balanço das contas. Crie uma nova visualização selecionando a opção Line e o index dos dados das contas bancárias. Siga então estes passos:

1. No painel **Metrics**, expanda **Y-axis**.
2. Configure **Aggregation** como **Average**.
3. Configure **Field** como balance. Ou seja, as no eixo y será mostrado a média dos balanços.
4. No painel **Buckets**, selecione Add.
5. Selecione **X-Axis**.
6. Configure **Aggregation** como **Terms**.
7. Configure **Field** como age. Ou seja, o eixo x mostrará a idade.
8. Configure **Order by** como **Alphabetical**, para que o eixo x mostre as idades em ordem.
9. Configurure **Order** como **Ascending**, para que a ordem seja crescente, de 20 a 40 anos.
10. Configure **Size** para qualquer valor acima de 20, para mostrar todas as idades.
11. Acione o botão **Update**. O resultado deverá ser similar ao ilustrado nesta figura.

Figura 34 - criando gráfico de linha

O gráfico resultante mostra que não existe uma tendência nítida da média do balanço em relação a idade. E se criarmos uma linha para cada gênero? Para fazer isso, basta adicionar um bucket. Siga os passos:

1. No painel de **Bucket**, acione o **Add**.
2. Selecione a opção **Split series**. Essa opção diz que o novo bucket irá subdividir os elementos da visualização (nesse caso, as linhas).
3. Em **Sub aggregation** selecione **Terms**. Isso significa que os dados serão subdivididos por um atributo categórico.
4. Em **Field**, selecione gender, indicando que os dados serão subdivididos pelo gênero.
5. Acione o botão **Update**.
6. Atualize as cores para o padrão escolhido para os gêneros masculino e feminino. O resultado deverá ser similar à figura abaixo.
7. Salve a visualização com um nome apropriado, como [Accounts] Média de balanços por idade e gênero.

Figura 35 - criando gráfico de linha subdividido

Pelo gráfico, notamos também que não há uma tendência nítida das médias de balanço em relação à idade e gênero. Muito bem, você criou as visualizações! Agora podemos exibi-las em um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard)!

**2.3 - Criando um dashboard**

Você já aprendeu que um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) no [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) é uma coleção de painéis de visualizações organizadas em uma página. Após criar as visualizações desejadas, adicioná-las a um [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) é muito simples. Siga o passo a passo a seguir:

1. Acione o ícone de [Dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) na barra de navegação.
2. Na página de dashboards, acione o botão **Create**[**dashboard**](https://leadfortaleza.com.br/ead/glossary/Dashboard). Você irá abrir a página de edição de dashboards.
3. Selecione **Add**. Irá aparecer uma lista com todas as visualizações pré-criadas.
4. Selecione as quatro visualizações que você criou.
5. O [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) irá distribuir as visualizações de maneira uniforme, deixando todas com o mesmo tamanho. Você pode reorganizar o [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard), modificando o tamanho e posição das visualizações de acordo com o que você achar mais adequado para cada visualização, assim como você fez na seção anteriormente.
6. Quando estiver satisfeito, acione o botão **Save** da página de edição de Dashboards, escolhendo um nome apropriado para o seu [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard).

Pronto, você fez o seu primeiro [dashboard](https://leadfortaleza.com.br/ead/glossary/Dashboard) utilizando o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana)! Como já foi dito, o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana) é uma ferramenta com muitas opções para visualizar e analisar dados e seria impossível cobrir todas as suas funcionalidades em uma única aula. Mas os conhecimentos que você desenvolveu até o momento já devem lhe dar uma noção bem fundamentada da lógica necessária para criar visualizações e dashboards com o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana). Procure explorar ainda mais os dados, experimentando diferentes opções de visualização. Quando surgirem dúvidas, você pode utilizar o botão Help, presente em qualquer tela do [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana), na barra superior, que te leva para a documentação do [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana), que é muito detalhada. Além disso, é muito útil explorar os dashboards que vêm pré-construídos nos dados de exemplo e observar como cada visualização foi desenvolvida através do modo de edição.

Caro estudante, você chegou ao fim desta aula. Neste conteúdo você aprendeu a instalar e executar o [Elasticsearch](https://leadfortaleza.com.br/ead/glossary/Elasticsearch) e o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana), a explorar dados com dashboards e a criar suas próprias visualizações e dashboards a partir de conjuntos de dados. Ufa! É bastante informação, não é mesmo? Porém são conhecimentos muito úteis que lhe darão grande poder de análise para explorar dados de forma dinâmica, permitindo descobrir padrões em vários níveis de detalhes. Procure sempre experimentar e aprender mais sobre essa ferramenta tão útil que é o [Kibana](https://leadfortaleza.com.br/ead/glossary/Kibana). Para isso, nada como exercitar e explorar conjuntos de dados.

Espero que você tenha aprendido bastante. Até a próxima!

### Referências

* WILKE, Claus O. Fundamentals of Data Visualization. O’Reilly, 2019.
* MUNZNER, Tamara. Visualization Analysis & Design. CRC Press, 2014.
* Kibana Guide, [https://www.elastic.co/guide/en/](https://www.elastic.co/guide/en/kibana/current/index.html)[kibana](https://leadfortaleza.com.br/ead/glossary/Kibana)/current/index.html - Acessado pela última vez em Junho de 2020.