



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

MODELAGEM DIMENSIONAL E INDICADORES DE DESEMPENHO/DASHBOARDS

Aula 3

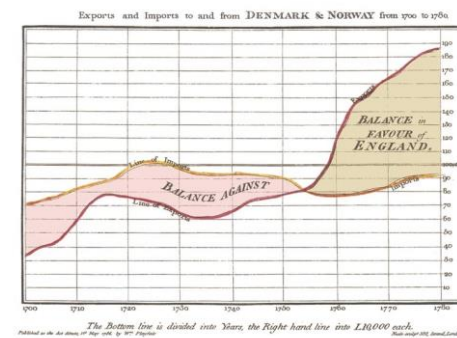
Prof. Marcondes Alexandre
marcondesalexandre@gmail.com



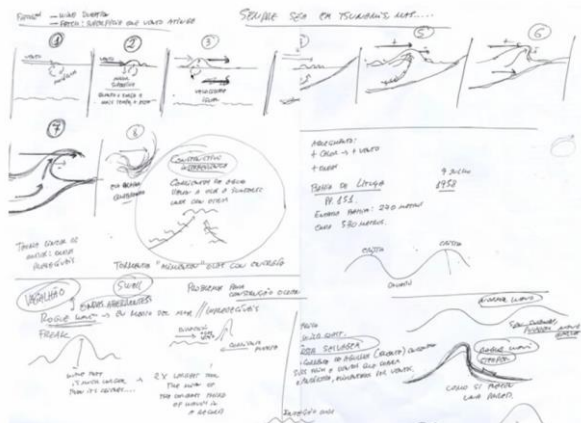
Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados

Willian Playfair (1759-1823)

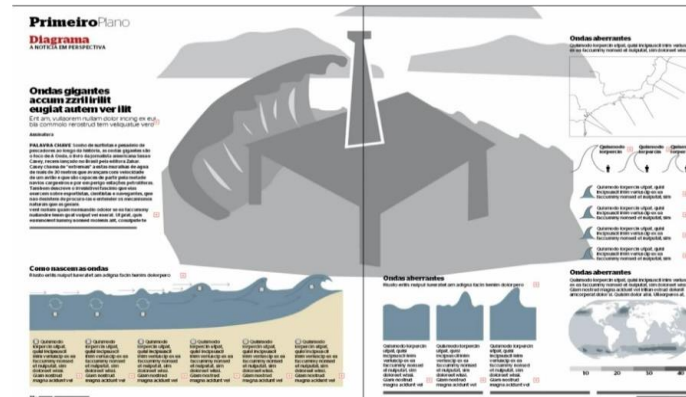


Visualização dos Dados

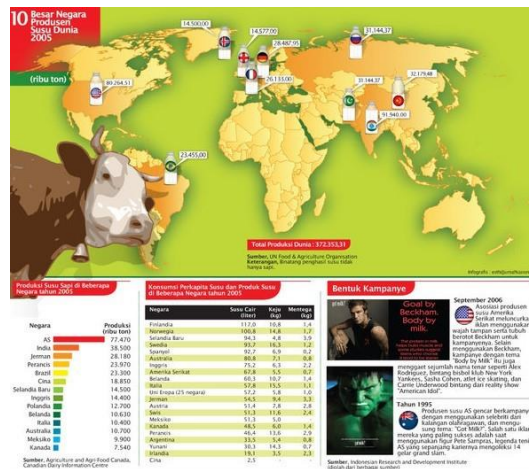


Visualização dos Dados

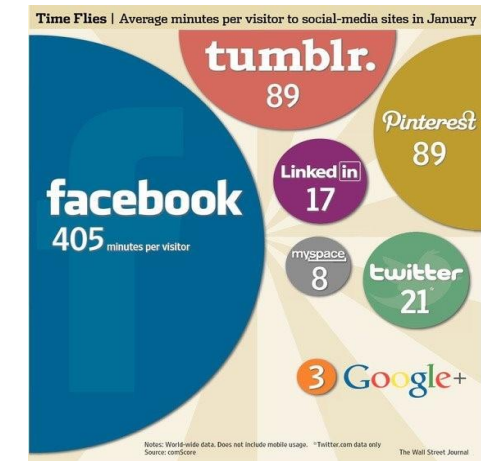
Esboço da infografia



Visualização dos Dados



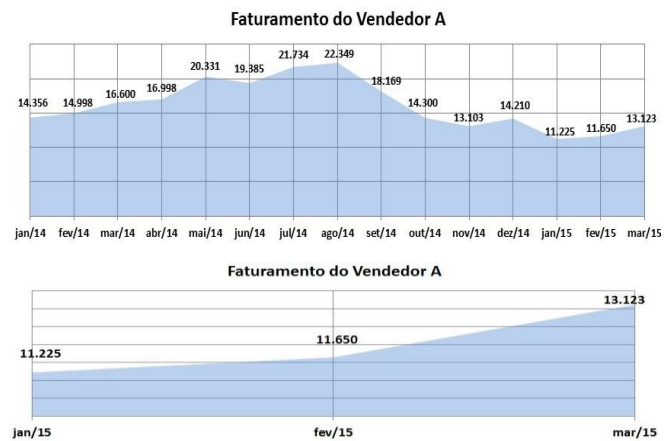
Visualização dos Dados





Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados



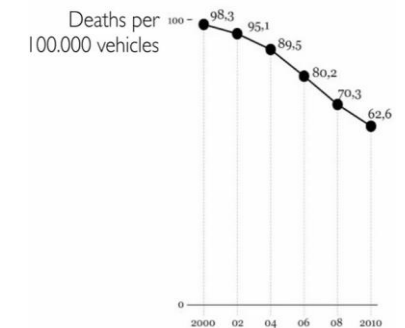
Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados

Revista Veja FICOU PIOR COM A LEI SECA



Veja o gráfico real



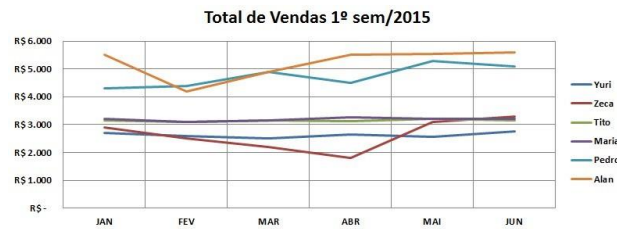


Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados

Gráfico de linha

compara a performance de 2 ou mais itens no tempo



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados

Gráfico de linha

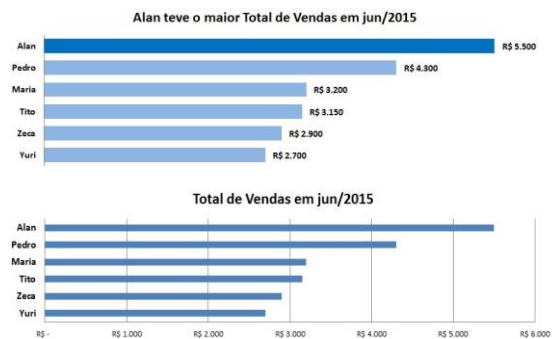
compara a performance de 2 ou mais itens no tempo





Visualização dos Dados

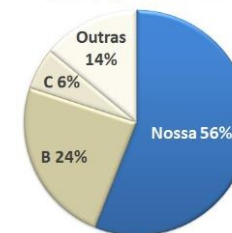
Gráfico de barras
usado para comparar itens



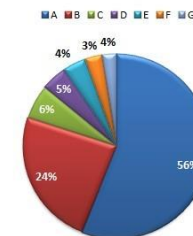
Visualização dos Dados

Gráfico de pizza
apresenta uma parte em relação ao todo

Nossa empresa está com 56% de participação no mercado



Participação no Mercado



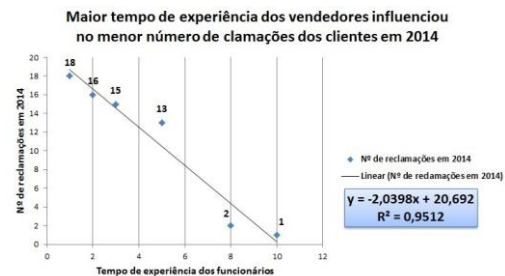


Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visualização dos Dados

Gráfico de correlação

Relaciona 2 variáveis (valores). E o eixo "x" não é tempo.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dashboards – Conceitos, dicas ..

- Muitas vezes, os termos "dashboard" e "relatórios" são utilizados como sinônimos, embora tenham significados diferentes.
- Um dashboard é um *contêiner* para um grupo relacionado de visualizações de tabelas de indicadores e relatórios organizados.
- Em outras palavras, um dashboard contém uma coleção de outros itens, tais como, tabelas de indicadores, relatórios e filtros.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dashboards – Conceitos, dicas ..



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dashboard e relatório são as mesmas coisas?

- Um relatório é a apresentação de dados transformados em informações formatadas e organizadas de acordo com requisitos de negócio específicos.
- Os relatórios podem ser imagens estáticas simples ou apresentações altamente interativas dos dados.

Podemos ordenar, filtrar e agregar, desagregar ou explorar dados em alguns tipos de relatório.

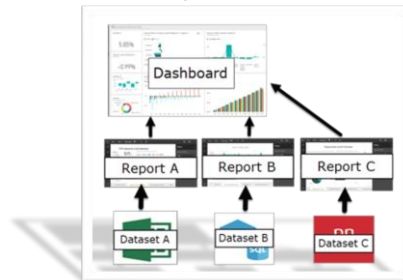




Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dashboard e relatório são as mesmas coisas?

- Um dashboard (em inglês painel de instrumentos ou painel de controle) típico pode conter uma tabela de indicadores, um relatório analítico e um gráfico analítico, sendo possível existirem diversas variações.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dashboard e relatório são as mesmas coisas?

- Alguns *dashboards* fornecem aos utilizadores um elevado nível de interatividade e outras apresentações estáticas. O grau e tipo de interatividade depende do programa utilizado na criação do dashboard.



Por que um dashboard?



Por que um dashboard?

- Um *dashboard* (ou "painel") é uma ótima forma de apresentar informações de maneira rápida e eficiente para diferentes pessoas, em diferentes áreas e níveis, de uma empresa.
- Eles comunicam claramente a informação e geram resultado.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Razões cognitivas para se fazer um dashboard

- Na maioria dos softwares empresariais ERP e até mesmo os BIs, que podem ser entendidos como um guarda-chuva conceitual que envolve a Inteligência Competitiva (CI), a Gerência de Conhecimento (KMS) e a IBI (Internet Business Intelligence), pesquisa e análise de mercado, relacionados a nova era da Economia da Informação, são dedicados a captura de dados, informações e conhecimentos que permitem as organizações competirem com maior eficiência no contexto atual.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Razões cognitivas para se fazer um dashboard

- Neste contexto de conjunto de ferramentas utilizadas para manipular uma massa de dados operacional em busca de informações essenciais para o negócio lidamos com um grande volume de informações, de diversas fontes e em diversos formatos, o que acaba criando uma massa de dados que gera apenas confusão.
- Emerge então a necessidade de sumarizar parte destes dados em um formato que forneça uma visão geral da performance, dê suporte à tomada de decisão e, claro, gere ação. O papel de um dashboard é justamente este: **sumarizar parte da informação disponível em uma forma clara, concisa e visual.**



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Menos é mais

- Não use sombreamento ou efeitos 3D, seu objetivo é comunicar eficientemente, e não fazer uma obra de arte computacional. O mesmo vale para cores: siga apenas uma cor e o tradicional “preto no branco” – facilita a leitura e não deixa espaço para dúvidas. Para tabelas, use apenas as bordas e grades necessárias para uma boa leitura dos dados.

• As regras básicas gerais são:

- 1 – Esqueça firulas, seja simples;
- 2 – Evite gráficos que ocupem espaço e não mostrem a informação completa (pizza);
- 3 – Economize espaços, tire borda, plano de fundo, eixos e títulos;
- 4 – Sobreponha gráficos se possível;
- 5 – Não coloque números muito pequenos.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Se você não sabe o objetivo do relatório, então você **NÃO TEM** relatório

- Especificar claramente qual é a “história” que se deseja contar em um dashboard é o primeiro passo.
- Não é aconselhável criar algo como “Dashboard da Empresa X”, o ideal seria mais próximo de “Dashboard de Tráfego no site da empresa X” – mais específico, mais conciso, menos megalomaniaco e com maior chance de sucesso.
- Embora um dashboard seja **generalista**, seu conteúdo deve ser o mais específico possível – é isso que o torna eficaz.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Se você não sabe o objetivo do relatório, então você **NÃO TEM** relatório

- É preciso ter claro também quem terá contato com este documento, um expert em internet ou um CEO mais preocupado com o lucro e o retorno obtido do que pelo número de visitantes únicos do site? Isso faz toda a diferença na forma como a tal "história" será contada.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Conheça um processo mais produtivo para se decidir o layout de um relatório

- O 5W2H, basicamente, é um checklist de determinadas atividades que precisam ser desenvolvidas com o máximo de clareza possível por parte dos colaboradores da empresa.
- Ele funciona como um mapeamento destas atividades, onde ficará estabelecido o que será feito, quem fará o quê, em qual período de tempo, em qual área da empresa e todos os motivos pelos quais esta atividade deve ser feita.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Conheça um processo mais produtivo para se decidir o layout de um relatório

- Para elaborar o dashboard não é diferente, para isso devemos usar o próprio nome da ferramenta que é composto pelas primeiras letras dos nomes em inglês das diretrizes usadas nesse processo.

1. What – O que será feito?
2. Why – Por que será feito?
3. Where – Onde será feito?
4. When – Quando será feito?
5. Who – Por quem será feito



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Conheça um processo mais produtivo para se decidir o layout de um relatório

1. How – Como será feito?
2. How much – Quanto custará fazer?





Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- A função de um dashboard não é vasculhar os detalhes e nuances dos dados, mas reduzi-los ao mínimo necessário para a compreensão.
- De fato, ao tratar das vendas de um e-commerce, por exemplo, você dispõe de uma série de métricas que parecem fantásticas e fundamentais.
- Mas em um dashboard você terá que cortar muitas delas. É um processo doloroso, mas necessário.
- **Primeiro corte:** apenas as métricas elementares para a compreensão da informação, aquelas sem as quais nada fará sentido.
- Ainda sobram algumas, então temos um **segundo corte:** apenas as métricas facilmente compreensíveis, que dispensem explicações adicionais.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- Um dashboard deve sumarizar a informação em uma única “tela”, sem a necessidade de virar a página, dar scroll na tela ou mudar de canal.
- Se, por alguma razão, necessitar de maior espaço, considere utilizar alguma forma de animação ou mudança automática, mas idealmente siga a regra acima: uma tela, nada mais.





Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- A maior parte dos exemplos de dashboards que conhecemos contém apenas gráficos e, no máximo, tabelas.
- Isso gera um certo preconceito ou a ideia de que o texto é “proibido” neste tipo de documento. Pelo contrário: muitas vezes a informação escrita é mais clara e objetiva do que uma enorme tabela.
- Além disso, um dashboard também pode ser o espaço para pequenas análises ou insights que direcionem a interpretação dos dados e, neste sentido, o bom e velho texto é a melhor saída.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- Aquela célula da tabela, em fonte grande e destacada, mostra: “1 milhão de visitas”. De fato, é um grande número... mas o que ele significa? Em um dashboard há pouco texto, normalmente não há a figura do “apresentador”, e é preciso que esteja claro se um número é bom, ruim ou indiferente.
- Esteja você usando Excel, softwares de design de dashboards ou outras ferramentas de business intelligence, o processo de criação deve ser o mais otimizado e automatizado possível.
- Devemos lembrar que um dashboard é entregue em intervalos relativamente curtos, geralmente há mais de uma versão ou mais de um dashboard e ele não elimina a necessidade de relatórios e análises. Por todos estes motivos, deve tomar o menor tempo possível.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- Um dashboard deve ser pensado e estruturado em torno de boas práticas de visualização de dados e voltada para a tomada de decisão.
- Seu objetivo não é “encantar” pessoas ou fornecer números detalhados ao extremo, mas comunicar eficientemente informações vitais para um negócio, para isso algumas boas práticas devem ser seguidas:



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- **Se apoie nas convenções:** Entre o criativo e o convencional, em um dashboard, opte pelo convencional. Use gráficos de barra, de colunas ou linhas. Muitas uma tabela também é uma boa opção.
- **Não use gráficos “pizza”:** Eles ocupam muito espaço, não oferecem uma visão de tendência ao longo do tempo e escondem sutilezas nos dados. Gráficos pizza podem até não ser tão malignos quanto se fala, mas há melhores opções em quase todos os casos.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Passos a serem seguidos até se chegar ao relatório

- **Use técnicas de otimização do espaço:** Pense em um dashboard como um apartamento pequeno e use soluções inteligentes. Qual porcentagem do seu espaço é ocupada por dados e informações relevantes? Procure sempre maximizar esta taxa, cortando fontes enormes ou legendas rebuscadas.
- **Não deixe espaço para dúvidas:** Gráficos devem ter legendas, nomes e unidades bem definidas e sem redundâncias ou inconsistências. Para cada métrica, ofereça o máximo de informação possível, antecipando as questões que podem ser criadas diante de cada número ("Qual a variação?", "O que causou o aumento?", "Qual era a meta?").



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

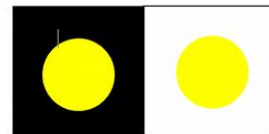
Passos a serem seguidos até se chegar ao relatório

- **Seja amigável:** Algumas pessoas são míopes, outras são daltônicas, outras possuem monitores desregulados. Para cada uma delas, fontes minúsculas ou cores estranhas podem ser um problema.
- **Procure usar cores comuns** (uma escala de cinzas geralmente resolve) e manter textos e gráficos em um tamanho razoável.
- **Pode parecer exagero**, mas em alguns casos é a diferença entre um documento ser visto como "útil" ou como "descartável".



As 9 regras para se usar cor

- Cor é absoluta, não relativa
- Ao contrário do que se pensa, a percepção de cor não é algo absoluto e sim relativo.
- A percepção da cor é totalmente influenciada pelo que existe na proximidade.
- Se duvida, responda: os círculos têm o mesmo tom de amarelo?



E agora?



As 9 regras para se usar cor

- Stephen Few listou 9 regras que deveríamos utilizar quando formos usar cor em nossos Dashboards (independente de onde sejam feitos, se em Excel, se em algum outro sistema):

Regra 1: Se você deseja que objetos com a mesma cor pareçam ter a mesma cor, então use uma cor de fundo que seja padronizada.

É como mostramos acima com os quadrados amarelos.

Não é porque a ferramenta (no caso o Excel) oferece este recurso que você deva usá-lo.

Refiro-me às opções de gradiente, à quantidade de gradientes diferentes em um gráfico. Se você for por este caminho, estará criando um [chart junk](#), não um Dashboard no Excel.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

- Regra 2: Se quer que a informação seja fácil de ser lida, use o **contraste**
- Exemplo típico é usar fonte preta combinada com preenchimento escuro. Isto pode acontecer quando você aplica uma formatação condicional ou quando fica mais preocupado em acertar uma cor para aquele gráfico/dashboard e se esquece do texto/valores, mantendo-os com a formatação padrão (automática) do Excel:

sem contraste

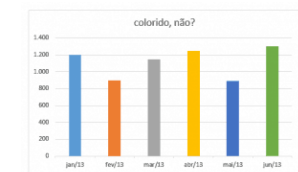
com contraste



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

- Regra 3: Cor é informação. Ela só muda se a informação muda.
- Exemplo típico: gráfico com colunas/barras em cores diferentes. Isto é mais um exemplo de que, apesar do Excel DEIXAR você configurar cores diferentes, você NÃO deve fazer isto:





Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

- **Regra 4:** Não use cor para decoração. Lembre-se de que o objetivo é o de informar, não entreter.
- Sim, tem gente que adora ficar colorindo seu Dashboard, seus gráficos. Afinal, você tem como objetivo mostrar a informação ou distrair o leitor?



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

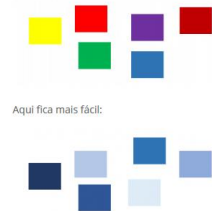
- **Regra 5:** Use cores suaves para o que não precisa de evidência e cores brilhantes e/ou escuras para o que precisa de atenção
- Se você usar o Excel, poderia criar Estilos ou Temas com conjuntos de cores brilhantes e com cores mais suaves, e usá-las de maneira padronizada.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

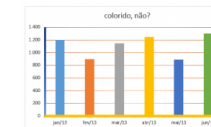
- **Regra 6:** Se for usar cores para identificar uma escala de valores, prefira mudança no tom à mudança na cor
- Se você precisar ordenar estas cores dificilmente acertará:



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

- **Regra 7:** As partes do gráfico/tabela que não forem a informação devem ser nada mais do que o razoável para serem visualizados.





Especialização em Ciência de Dados com Big Data, BI e Data Analytics

As 9 regras para se usar cor

- **Regra 8:** Será que você não tem algum leitor daltônico?
- Porque usar verde e vermelho se quem vai ler tem dificuldade em identificar estas cores é pedir para a pessoa não entender a informação...
- **Regra 9:** Nunca, em hipótese alguma invente de colocar efeitos visuais em seus gráficos



Especialização em Ciência de Dados
com Big Data, BI e Data Analytics

MODELAGEM DIMENSIONAL E INDICADORES DE DESEMPENHO/DASHBOARDS

Aula 4

Prof. Marcondes Alexandre
marcondesalexandre@gmail.com



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Agenda

- ✓ Introdução
- ✓ Gerenciando *DataSets*
- ✓ Pré-Processamento de Dados
- ✓ Usando Engenharia de Recursos e Seleção
- ✓ Prática Final

48



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Referências



49



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Introdução

Os principais componentes da aprendizagem da máquina para projeto no Microsoft Azure, são:

- *Machine Learning Studio*, um ambiente de desenvolvimento gráfico.
- Módulos de pré-processamento de dados.
- Algoritmos de aprendizagem de máquinas.
- APIs para integração a aplicativos.



50

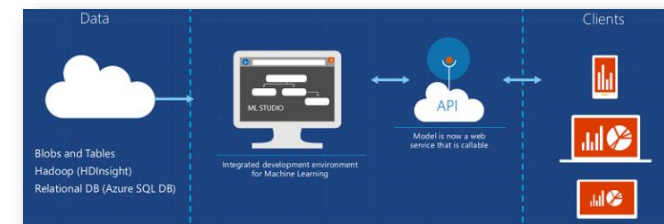


Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Introdução



Workflow Básico - Azure Machine Learning



51



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Introdução

Conceitos Iniciais sobre *Azure Machine Learning*

- *Machine Learning Workspace*
 - Container para projetos, modelos e experimentos
 - Vinculo com conta Microsoft para ser compartilhada
- Armazenamento
 - Free: Conta padrão de armazenamento possui limite de 10GB
 - Standard: Conta de armazenamento precificada conforme uso.
- Roles
 - User
 - Owner



52



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Introdução

Conceitos Iniciais sobre *Azure Machine Learning*

- *Experimentos*
 - Compreende pelo menos um conjunto de dados e um módulo
 - *Datasets* podem apenas ser conectados por módulos
- Projetos
 - Experimentos, datasets, anotações



53



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Gerenciando DataSets

Indagações Iniciais sobre a estrutura dos dados.

- *Crescimento do armazenamento da informação excede a lei de Moore*
 - as empresas raramente descartam dados
 - o crescimento de dados não estruturados é muito maior do que o crescimento de dados estruturados
- Essencial o entendimento sobre os tipos de dados
 - *Dados Estruturados*
 - *SGBD, SQL Server, Excel*
 - *Dados Não estruturados*
 - *Texto, Blogs, imagens, áudio e vídeo*



54



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Gerenciando DataSets

- *Big Data é desafiante por quê:*
 - *Pode contêm dados estruturados complexos*
 - *Dados não estruturados*
 - *A origem de dados pode não ser estática*
- *Os 3 Vs:*
 - *Volume, Variedade, Velocidade*
- *Tecnologias: Hadoop/HDInsight no Microsoft Azure*



55



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Gerenciando DataSets

Indagações Iniciais sobre o Propósito dos Dados.

- Como essa amostra foi feita?
- Que critérios foram usados para criar essa amostra?
- Essa amostra é de alguma forma especial?



56



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Gerenciando DataSets

Indagações Iniciais sobre o Treinamento dos Dados.

- 70% para treinamento, 30% para testes.
- 70% para treinamento, 20% para validação e 10% para testes.



Análise preditiva de dados não é a mesma coisa do que treinamento de dados

57



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Gerenciando DataSets

Importando dados para o *Azure Machine Learning*

- *Origens de Dados: Offline, Online e dados gerados*
- *Formato de Dados:*
 - *Txt, Csv, Tsv, Excel*
 - *Tabelas: SQL Database, Azure (Storage) e Hive*
 - *Odata, SvmLight, ARFF, zip, Rdata*
- *Tipos de Dados:*
 - *String, Integer, Double, Boolean, DateTime, TimeSpan*



58



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Pré-processamento de Dados no Azure Machine Learning

- Identificar questões que necessitam de uma exploração detalhada de dados
- Power BI pode ajudar durante esta fase de exploração
- As questões podem incluir:
 - *informações ausentes ou registros incompletos*
 - dados com ruído e valores atípicos
 - *Outras inconsistências e discrepâncias*
- *Ferramentas para Pré-processamento de Dados*
 - *Consultas SQL ou Hive*
 - *Scripts R ou Python*
 - *Módulos do Azure Machine Learning*



59



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Prática Final: Usando Classificação e Regressão Linear no Azure ML Studio

Laboratório 1

- Criando experimento simples com treinamento de dados

Laboratório 2

- Elaborando experimento utilizando algoritmo de classificação e regressão linear para aprendizagem supervisionada.

