



Modelagem, Extração e Manipulação de Dados

BLOCO: B.I. E ANÁLISE DE DADOS

PROF. RODRIGO EIRAS, M.SC.

[ETAPA 6] AULAS 1 E 2 – ANALISAR DADOS NO POWER BI



Na última aula...

- Dados agregados
- Escalas
- Análise Exploratória de Dados
- Estatística Descritiva e Inferencial



Agenda

- TP-3
- Dados contínuos e discretos
- Análise de dados
 - Descritiva
 - Prescritiva
 - Preditiva
 - Diagnóstica

O TP3 tem o objetivo de avaliar a seguinte competência:

Exibir Aus

2. Utilizar o Microsoft Power BI (MPBI).

Segundo a Microsoft, o Power BI (MPBI) é um conjunto de ferramentas de análise de negócios para analisar dados e compartilhar ideias. Se você é um analista de dados que entrega relatórios e análises à sua organização, o Power BI permitirá que você seja produtivo e criativo com sua criação. Power BI Desktop é um mashup de dados com recursos avançados e uma ferramenta de criação de relatórios. Combine os dados de bancos de dados diferentes, arquivos e serviços Web com ferramentas visuais que ajudam a compreender e corrigir automaticamente problemas na qualidade dos dados e de formatação.

Tendo em mente que vocês são futuros tomadores de decisão, vamos avaliar a habilidade de vocês na utilização de ferramentas analíticas, nesse caso, o Power BI.

Utilizando a base de dados de Categorização dos Municípios Turísticos - 2016 -, responda às questões.

<2016-categorizacao.csv>

TP-3

Segundo o Ministério do Turismo, a Categorização é um instrumento elaborado para identificar o desempenho da economia do setor nos municípios que constam no Mapa do Turismo Brasileiro. Esse instrumento, previsto como uma estratégia de implementação do Programa de Regionalização do Turismo, permite tomar decisões mais acertadas e implementar políticas que respeitem as peculiaridades dos municípios brasileiros.

fonte: URL: <http://www.turismo.gov.br/>

A coluna Valor Receita foi adicionada à Base de Categorização do Turismo Nacional para fins didáticos, não contém valor real.

TP-3

Questão 1)

Com a base de dados de Categorização do Turismo no Brasil carregada no Power BI, crie um relatório para análise do Turismo no Brasil. O objetivo é analisar os principais estados brasileiros com vocação para turismo. Esse é um relatório para tomada de decisão, o design é fundamental. Defina para o relatório: título, padrão de cores, e demais atributos da identidade visual. Crie uma aba de Resumo para apresentar os principais indicadores disponibilizados por região e estado do Brasil. Justifique a utilização dos gráficos escolhidos.

TP-3

Questão 2)

Faça uma análise exploratória da Base de Categorização do Turismo no Brasil apresentando para cada Estado do Brasil os valores totais, médias, mínimos e máximos. Faça uma análise breve dos resultados obtidos. Crie uma aba Exploração para apresentar os resultados e utilize o mesmo padrão visual definido para a questão 1. Justifique a utilização dos gráficos escolhidos.

Questão 3)

Crie um relatório para analisar as visitas internacionais e as visitas nacionais em comparação com a quantidade de empregos e estabelecimentos. Utilize os gráficos de combinação, recursos de segunda escala, se for o caso, para melhor visualização da informação. Faça uma análise breve dos resultados obtidos. Crie uma aba Visitação para apresentar os resultados e utilize o mesmo padrão visual definido para a questão 1. Justifique a utilização dos gráficos escolhidos.

TP-3

Questão 4)

Crie um relatório para analisar receitas obtidas em relação à quantidade de empregos gerados. Identifique qual o indicador discreto e o indicador contínuo que estão sendo analisados e defina os conceitos de dados discretos e contínuos. Escolha pelo menos duas representações gráficas para cada tipo de dado. Faça uma análise breve dos resultados obtidos. Crie uma aba Valor Gerado para apresentar os resultados e utilize o mesmo padrão visual definido para a questão 1.

Assim que terminar, salve todas as respostas em um arquivo PDF e poste na etapa correspondente no Moodle respeitando a seguinte nomenclatura: "nome_sobrenome_DR3_TP3.PDF".

Boa Sorte!

TP-3

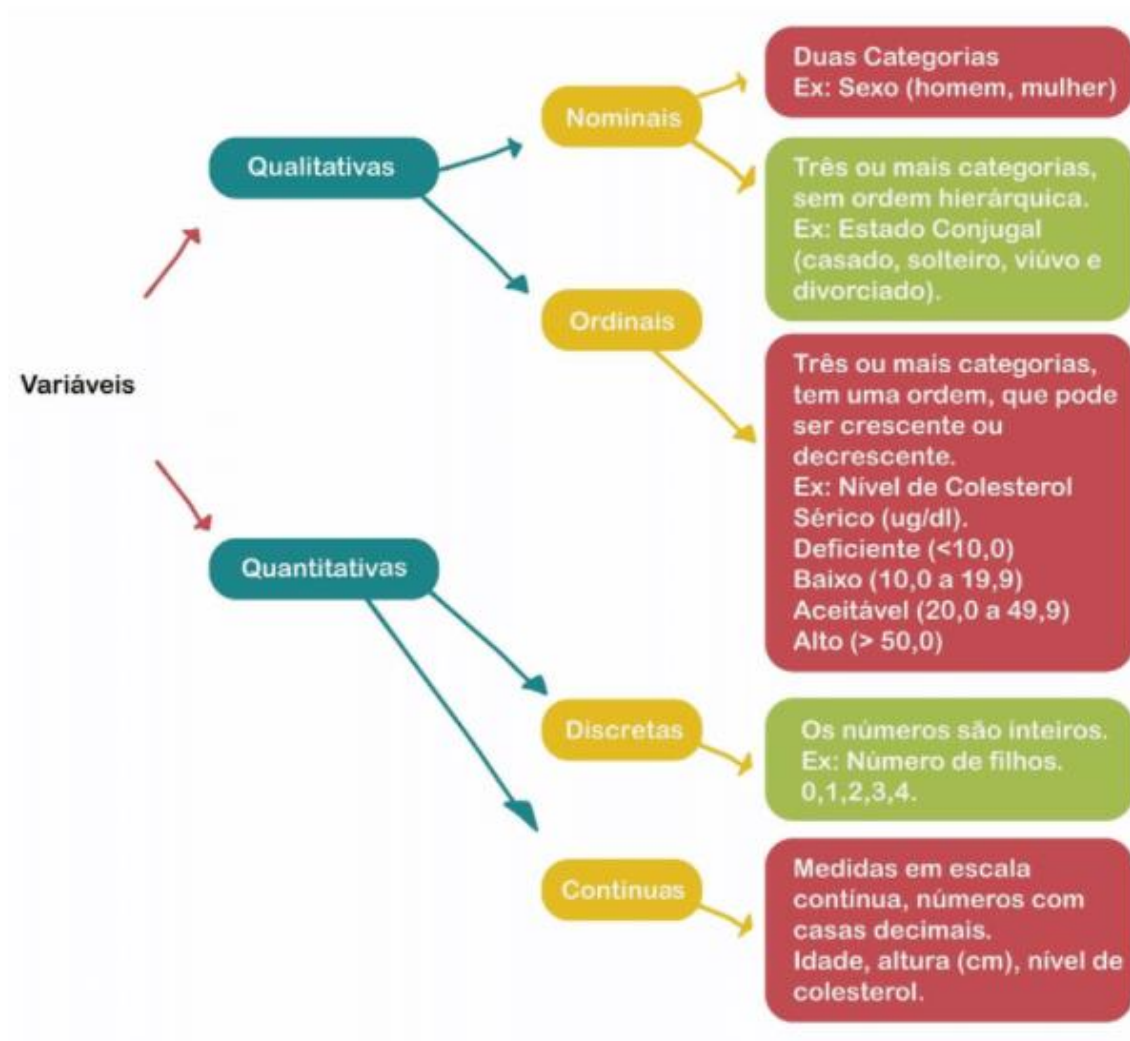
Dados contínuos e discretos

De forma simples, o dado discreto é avaliado em números que são resultados de contagens e, por isso, somente fazem sentido números inteiros.


- Exemplos: número de filhos, número de assaltos, números de gols.

Dado contínuo é avaliado em números que são resultados de medições e, por isso, podem assumir valores com casas decimais e devem ser medidos por meio de algum instrumento.

- Exemplos: massa (balança), altura (régua), tempo (relógio), pressão arterial, idade.



Dados contínuos e discretos



Visualização dos dados contínuos e discretos

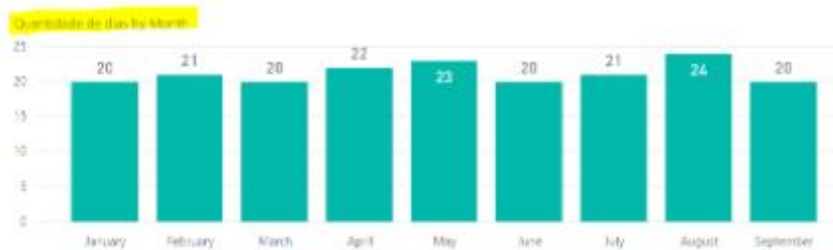
- O dado contínuo é refletido em um gráfico onde todos os pontos possuem valores significativos.
- O dado discreto, por sua vez, é representado por alguns pontos em um gráfico.
- Embora algumas vezes haja linhas conectando esses pontos, elas não representam valores naqueles pontos ao longo do domínio.

Visualização dos dados contínuos e discretos



Gastos Mensais no Intercambio EUA em R\$

Month	Gastos R\$	Quantidade de dias
January	68,40	20
February	68,00	21
March	67,80	20
April	67,20	22
May	68,00	23
June	65,40	20
July	64,40	21
August	85,00	24
September	95,00	20
Total	649,20	191



- Veja na imagem a representação da visualização de dados contínuos e discretos no Power BI
- Qual das visualizações mostra dados discretos? E contínuos?

Visualização dos dados contínuos e discretos



Gastos Mensais no Intercambio EUA em R\$

Month	Gastos R\$	Quantidade de dias
January	68,40	20
February	68,00	21
March	67,80	20
April	67,20	22
May	68,00	23
June	65,40	20
July	64,40	21
August	85,00	24
September	95,00	20
Total	649,20	191

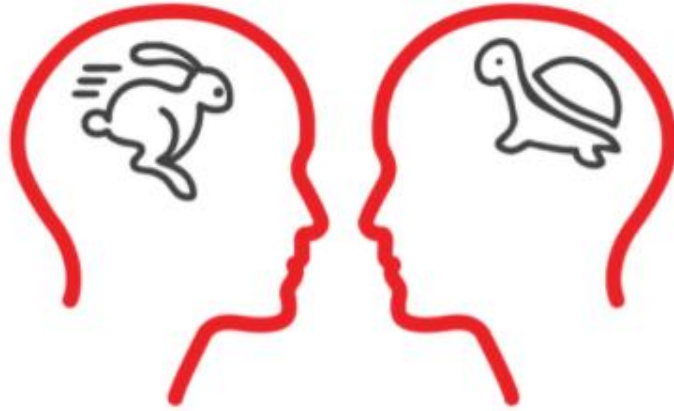


- A variável Gastos Mensais é mostrada na visualização gráfico de linhas, pois se trata de um dado contínuo e é importante mostrar que este pode assumir os valores de todos os pontos entre os intervalos dos meses.
- A variável Quantidade de Dias foi apresentada num gráfico de colunas pois os valores entre os intervalos de meses não se aplicam.



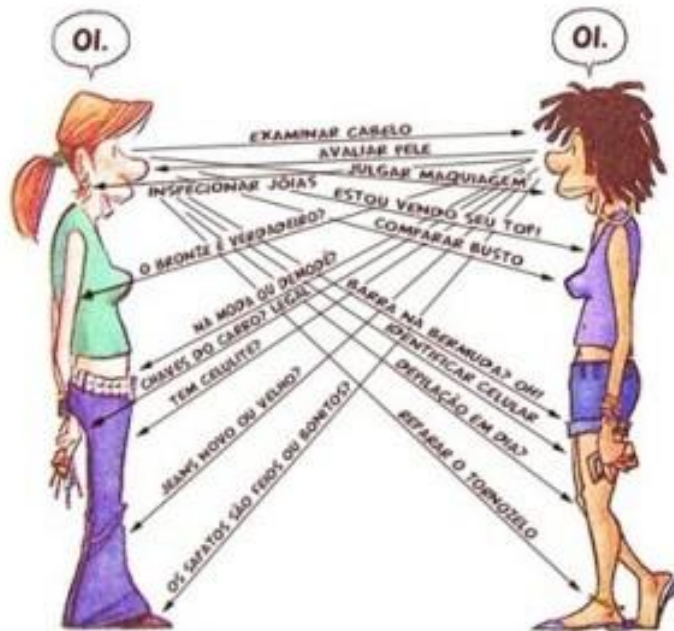
Exercício rápido

- Utilizando a base de dados OrderData.csv
 - <https://raw.githubusercontent.com/rodrigoeiras/Infnet-FundamentosBI/main/DataSets/OrderData.csv>
- Identificar variáveis numéricas contínuas e discretas

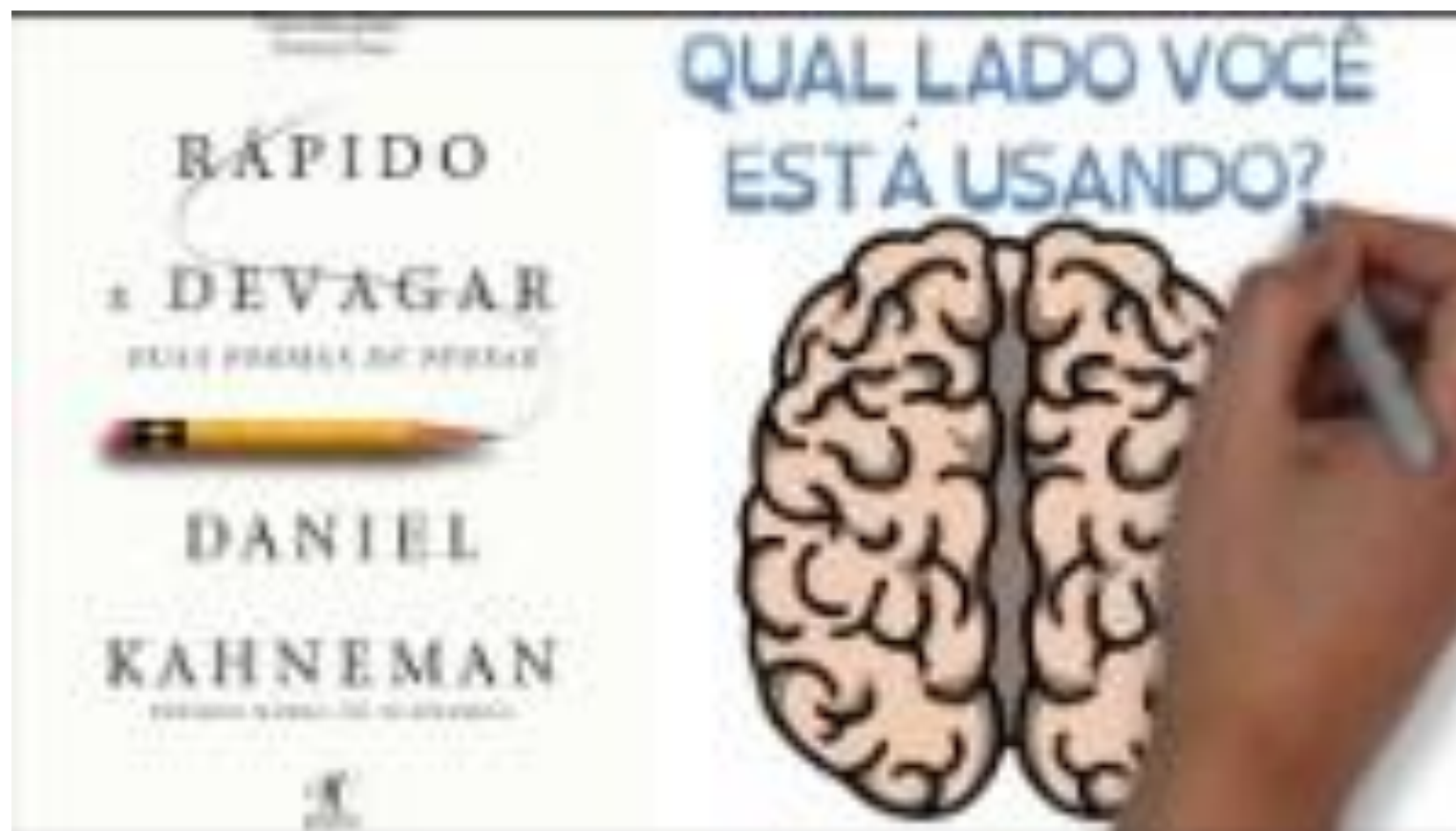


Analizando dados no Power BI

- Antes de começarmos a falar de análise de dados no Power BI é fundamental ver o recurso abaixo que fala dos dois modos de pensarmos: Rápido e Devagar, segundo Daniel Kahneman.
- EFEITO HALO: O motivo das pessoas defenderem políticos!
- RÁPIDO E DEVAGAR: DUAS FORMAS DE PENSAR



Rápido e Devagar

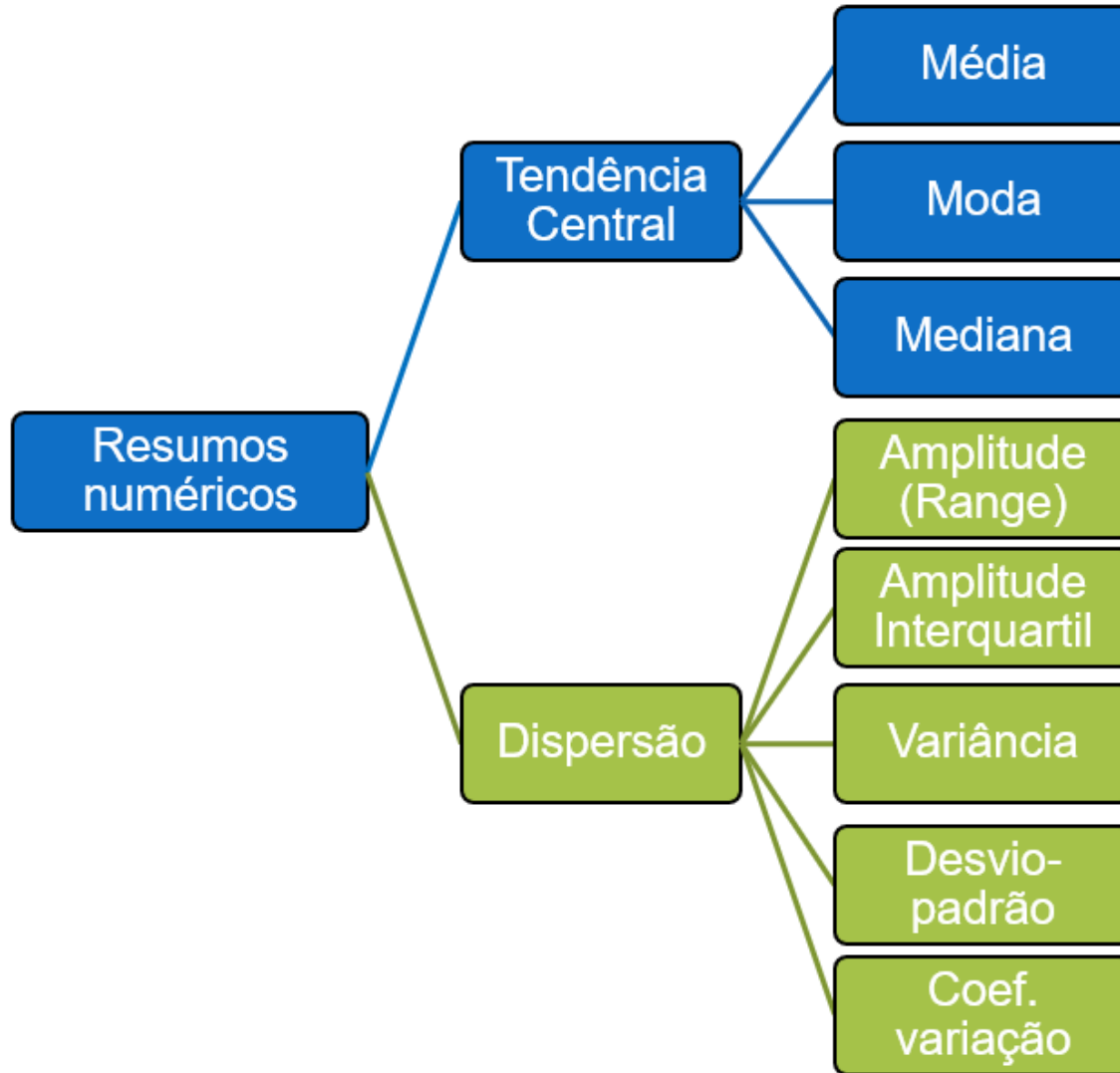




Análise de Dados

- A análise de dados é dividida em 4 categorias principais:
 - Descritiva
 - Diagnóstica
 - Preditiva
 - Prescritiva
- Vamos falar um pouco sobre cada uma delas.

Descritiva: o que está acontecendo?



- A análise descritiva fornece ao analista uma visão das principais medidas sobre aquilo que se está estudando.
- Exemplos de ações em uma análise descritiva são:
 - Organização e classificação das variáveis
 - Obtenção de medidas de posição, como a média, e de medidas de dispersão, como a variância
 - Categorização dos dados
 - Geração de gráficos e outras ferramentas de visualização

Power BI – Quick Insights

O Power BI Service possui um serviço de insights rápidos

Útil se você não tem conhecimento sobre os dados

Utiliza serviços de IA da Azure

A imagem abaixo mostra uma das visualização do insight rápido no Power BI.

Maioria (Principais fatores)

Encontra casos em que a maioria de um valor total pode ser atribuída a um único fator quando dividida por outra dimensão.



Power BI – Quick Insights

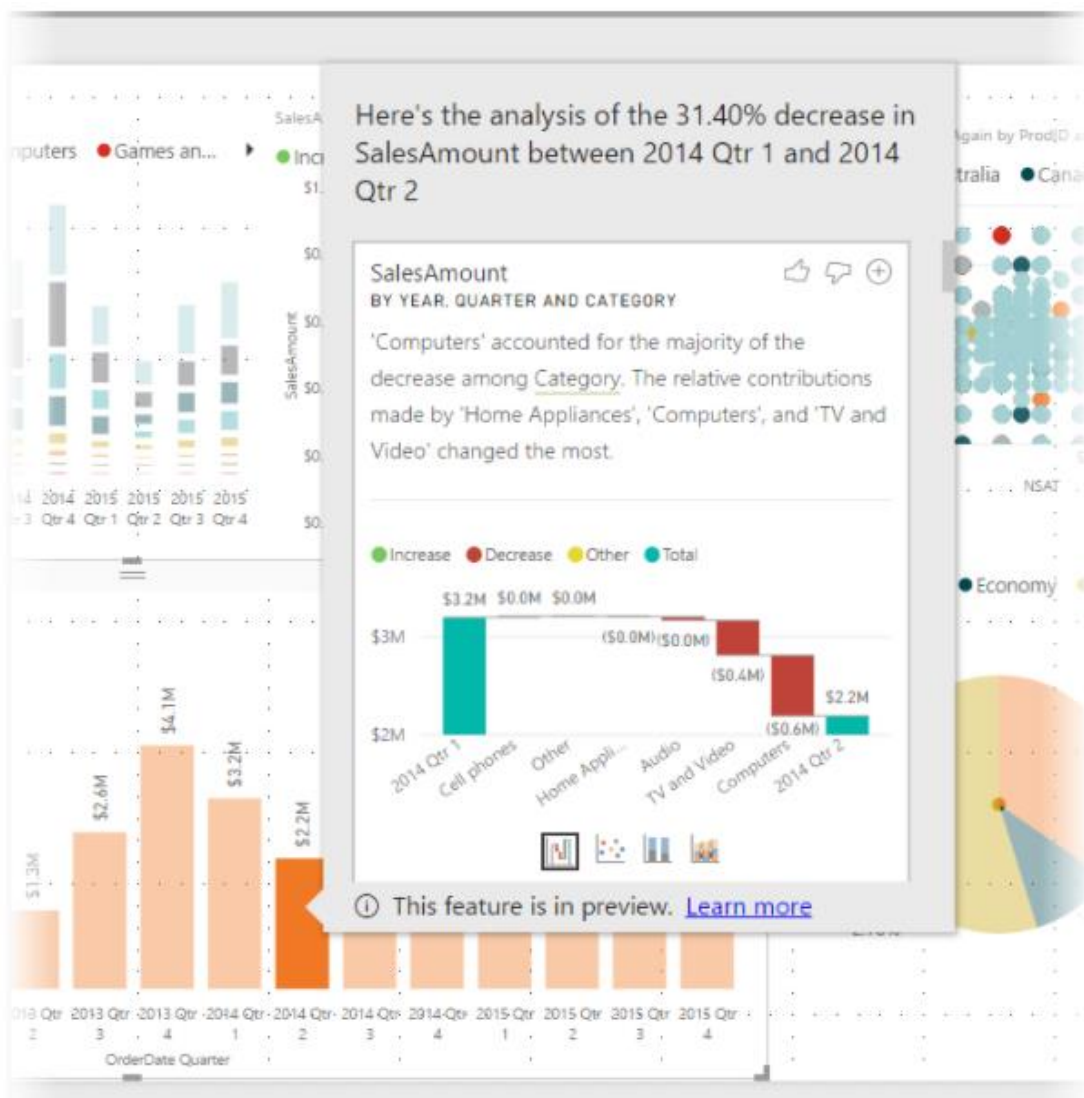
Diagnóstica: por que está acontecendo?



- A análise diagnóstica requer do analista conhecimento sobre o tema dos dados em análise.
- Por exemplo, para analisar os dados de uma partida de futebol, é necessário saber as regras básicas do esporte antes de realizar um diagnóstico.
- Painéis de informações bem planejados que incorporam a leitura de dados em séries temporais, com filtros e capacidade de detalhamento potencializam esse tipo de análise.
- Os Dashboards no Power BI são uma excelente forma de apresentação para análise diagnóstica.

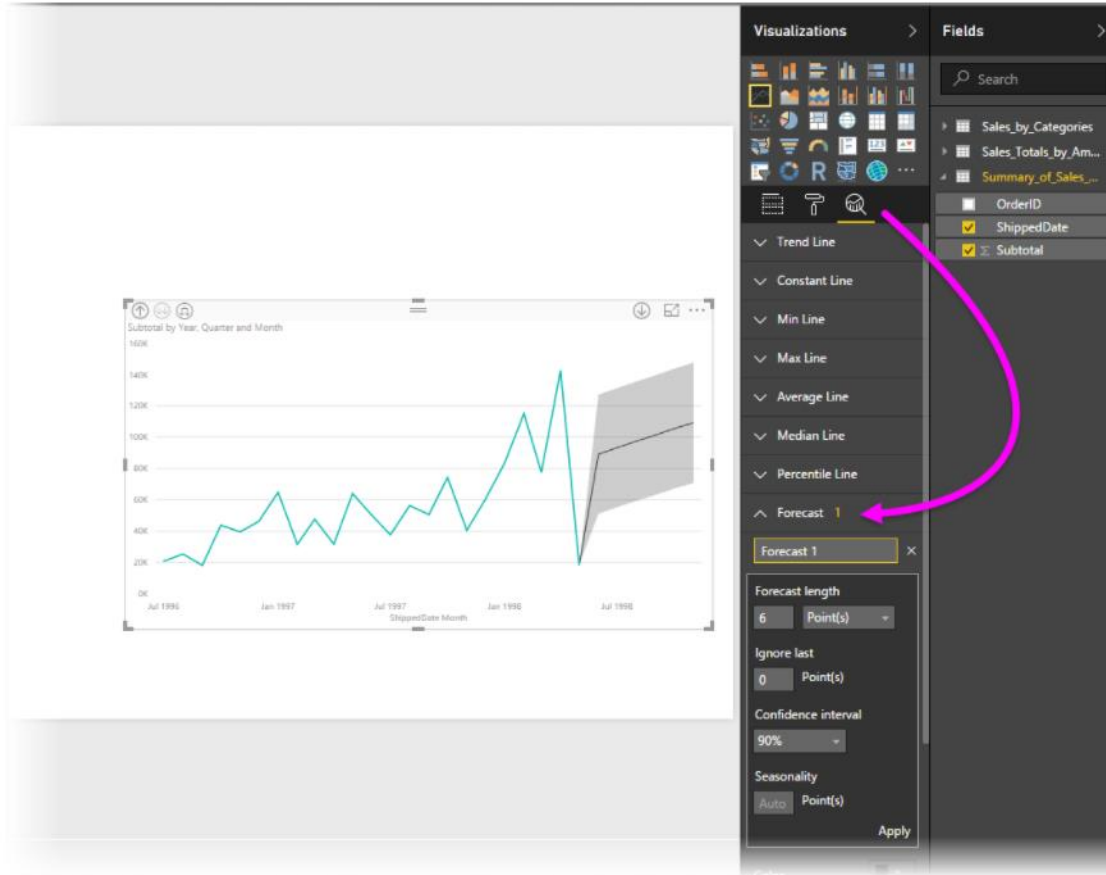
Power BI – Quick Insights

- A funcionalidade Insights/Percepções no Power BI é um diferencial para análise diagnóstica.
- O Power BI através de algoritmos sofisticados provê, através dessa funcionalidade, uma série de insights sobre os dados analisados.



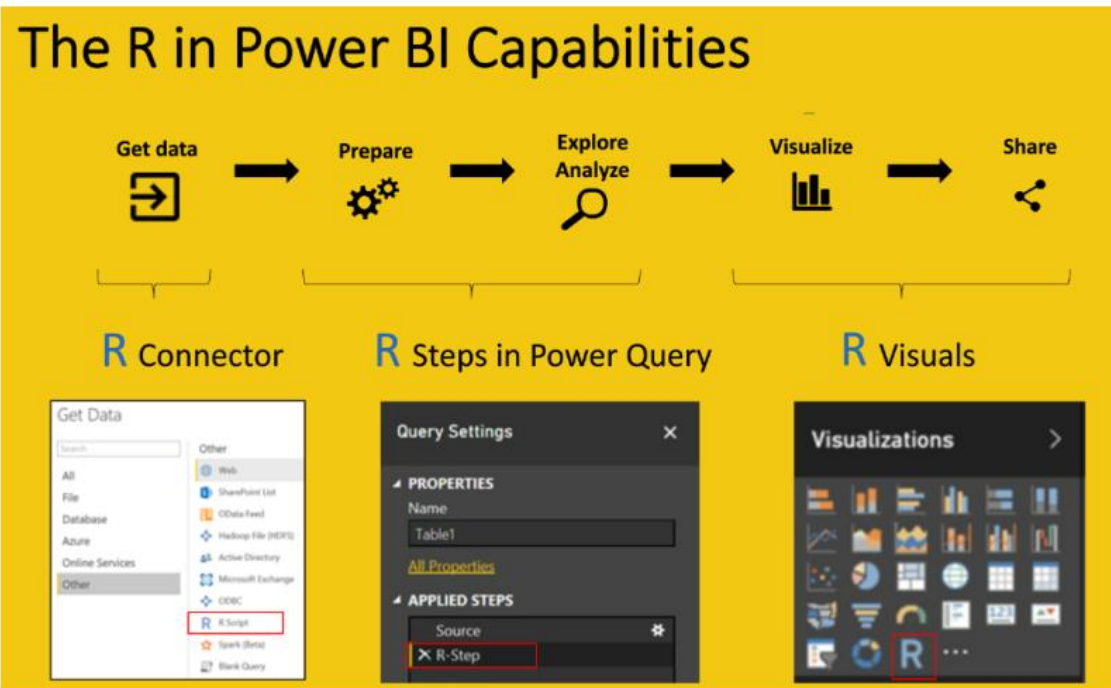
Preditiva: o que é provável que aconteça?

- A análise preditiva trata sobre prognósticos, inferências e estimativas.
- A probabilidade de um evento acontecer, a previsão sobre uma quantidade ou a estimação de um intervalo de tempo em que algo pode acontecer – tudo isso é feito através de modelos preditivos.
- Em um mundo de grande incerteza, poder prever permite tomar decisões melhores.

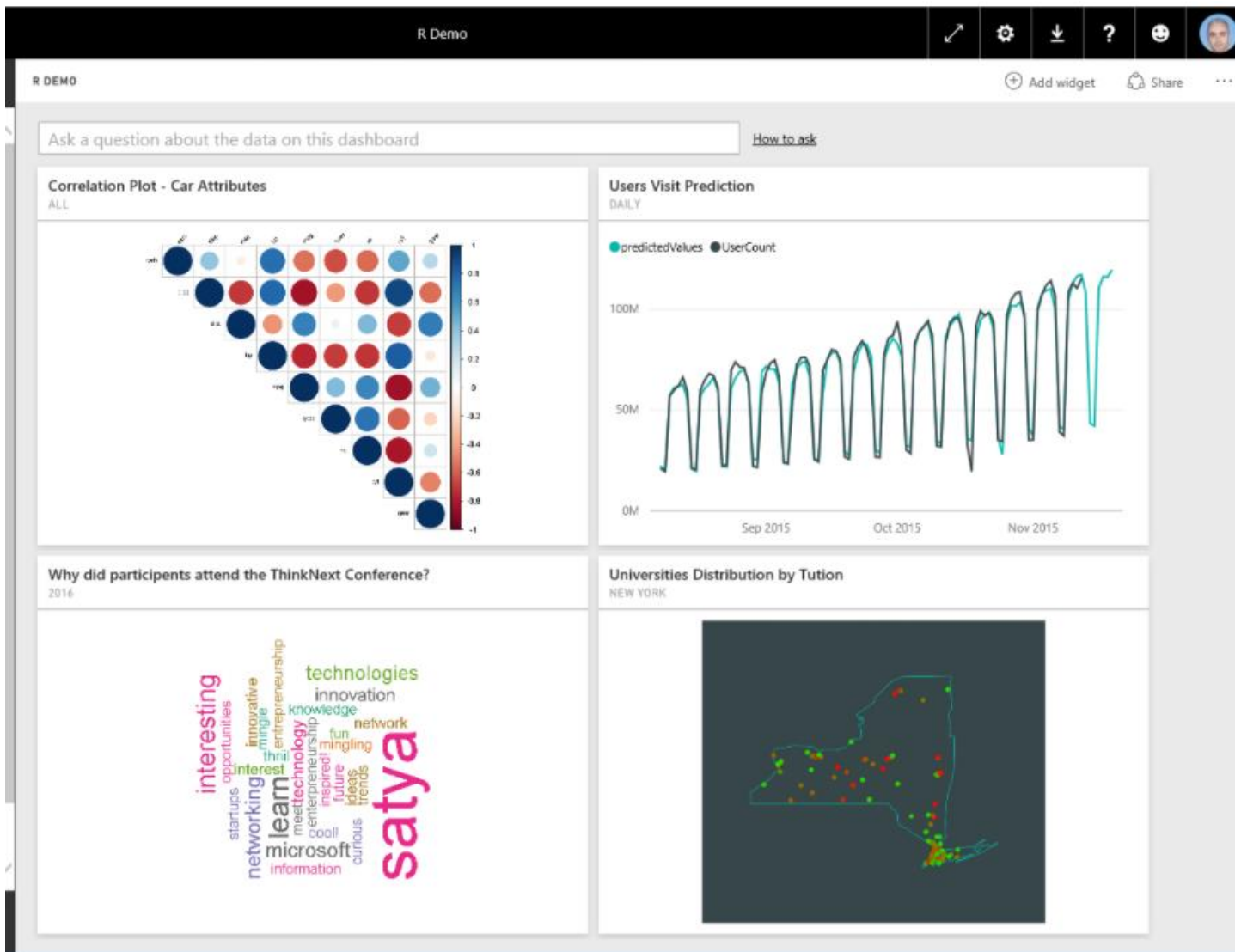


Preditiva: o que é provável que aconteça?

The R in Power BI Capabilities



- Para a análise preditiva, o Power BI disponibiliza a integração com o R, uma linguagem de programação amplamente utilizada pelos estatísticos.
- Um esquema de como o R pode ser utilizado no Power BI é apresentado na imagem.



Preditiva: o que é provável que aconteça?

- A imagem mostra um painel com informações geradas nos modelos do R e apresentadas no Power BI.

K-Means no Power BI

- Análise não supervisionada de grupos
- Busca criar um novo conceito dentro dos dados
- Exemplo com a base de dados IRIS

Prescritiva: O que eu preciso fazer?

- O modelo prescritivo utiliza uma compreensão do que aconteceu, o motivo de ter acontecido e mais uma variedade de análises do que “pode acontecer” para ajudar o usuário a determinar qual a melhor ação a ser realizada.
- Em geral, a análise prescritiva não consiste em uma ação individual, mas em uma série de ações conjuntas.
- Um exemplo disso é uma aplicação de tráfego que ajuda você a escolher a melhor rota para casa, levando em conta a distância de cada rota, a velocidade na qual se pode viajar em cada estrada e, fundamentalmente, as restrições de tráfego atuais.

Prescritiva: O que eu preciso fazer?

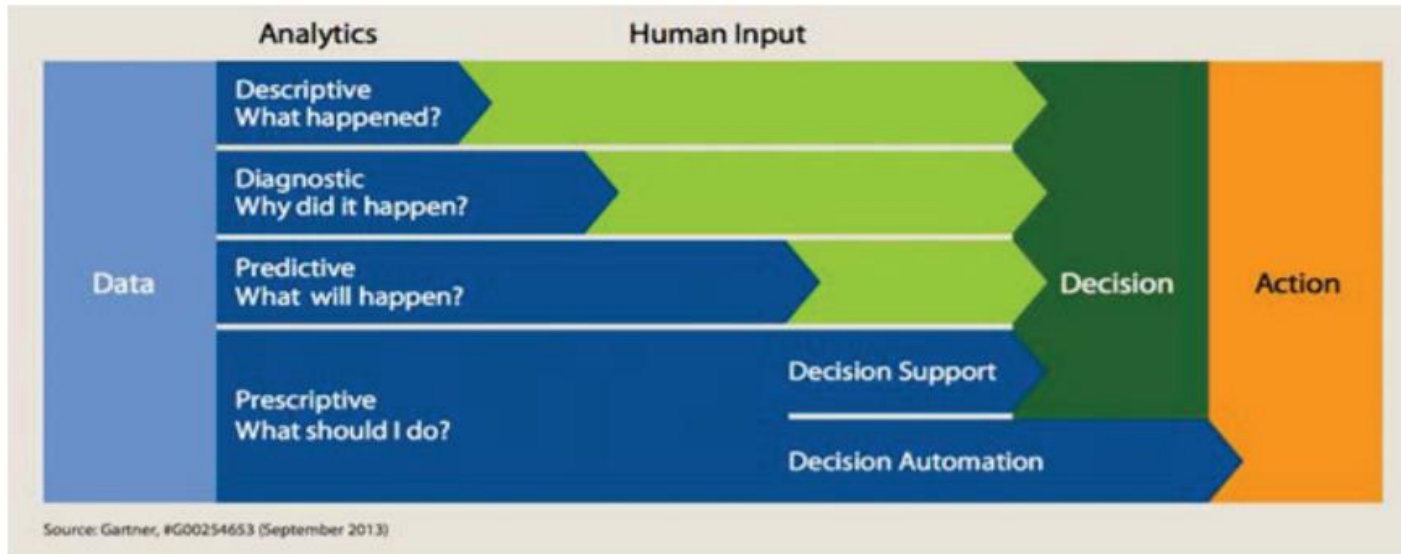
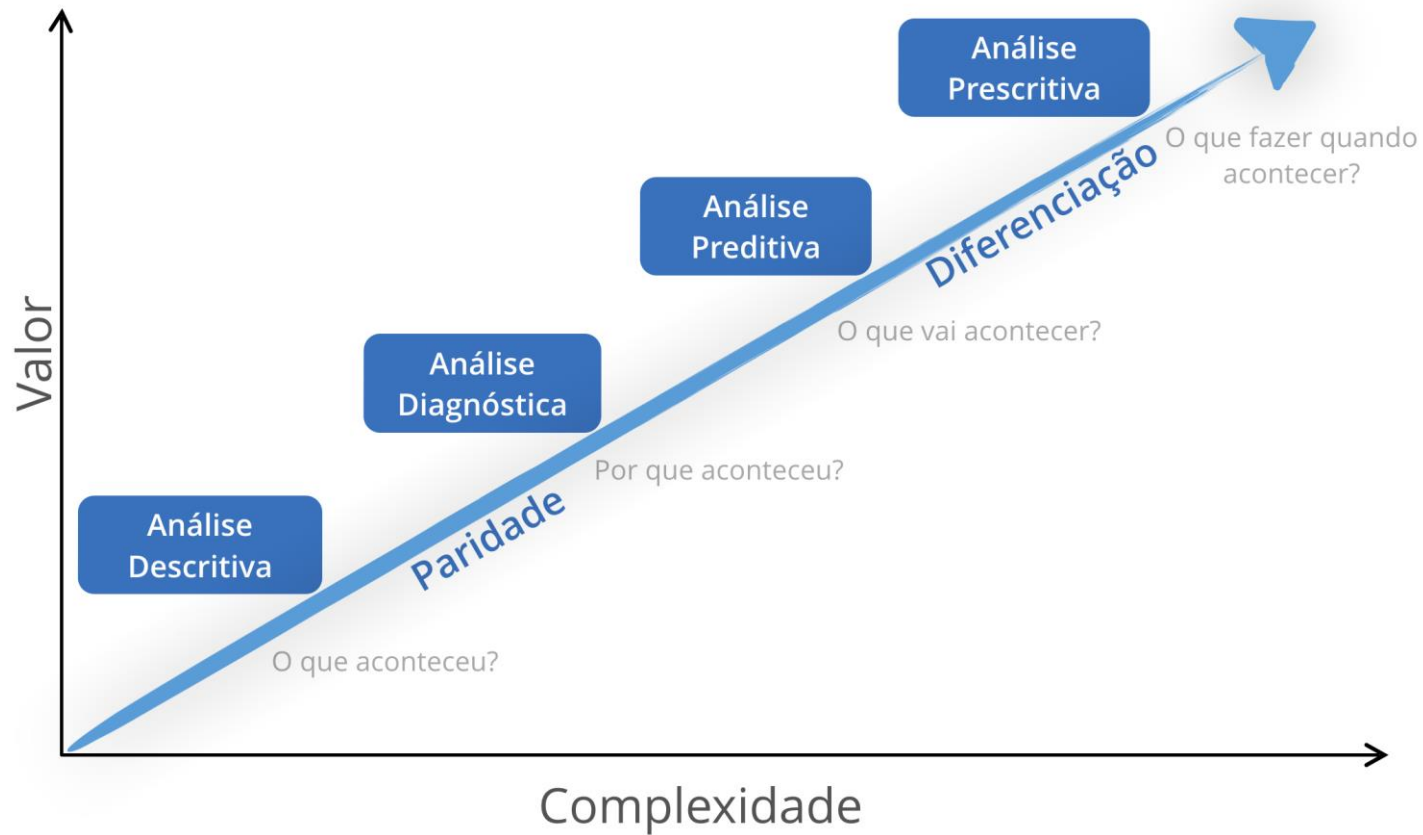


Figure 3. Gartner Analytics Maturity Model

- A análise prescritiva utiliza todo conjunto de funcionalidades do Power BI apresentadas para as demais análises.
- A prescrição não traz necessidades de novos objetos para análise, e sim novos objetivos.
- A imagem resume de forma simples a ligação de modelagem de dados, análise de dados, suporte a tomada de decisão e ação.

“Narratives” no Power BI





Resumindo!

Praticando!

Utilizando a base de dados (SSMM.csv) resolva as questões a seguir.

Faça eventuais ajustes nos dados para conseguir responder corretamente.

(https://raw.githubusercontent.com/rodrigoeiras/BI_AnaliseDados/main/Fundamentos_BI_AnaliseDados/Datasets/SSMM.csv)

1. Mostrar as medidas de posicionamento de todas as variáveis numéricas da tabela. Faça uma breve explicação dos resultados encontrados.
2. Mostrar as medidas de dispersão de todas as variáveis numéricas da tabela. Faça uma breve explicação dos resultados encontrados.
3. Mostrar as medidas de posicionamento e dispersão encontradas em um componente gráfico. Explique por que escolheu o gráfico utilizado.
4. Classifique as variáveis numéricas da tabela quanto ao seu tipo de dados.
5. Crie um gráfico de colunas configurado da seguinte forma: o eixo categórico terá a variável “Número da Loja” e o eixo de valor terá a variável “Número de Itens Comprados”. Qual loja vendeu mais itens? Ordene do maior valor para o menor. Certifique-se que o gráfico de colunas está com o eixo de categoria configurado de forma correta, de acordo com o tipo do dado encontrado na questão 4.

Praticando!

Utilizando a base de dados (SSMM.csv) resolva as questões a seguir.

Faça eventuais ajustes nos dados para conseguir responder corretamente.

(https://raw.githubusercontent.com/rodrigoeiras/BI_AnaliseDados/main/Fundamentos_BI_AnaliseDados/Datasets/SSMM.csv)

DESAFIOS PARA A PRÓXIMA SEMANA:

1. Crie um conector em R para carregar a tabela SSMM.csv diretamente do Github.
2. Crie uma visualização em R, um boxplot, para mostrar a distribuição dos dados por Região e Montante de Compras
3. É possível dizer que existe algum tipo de correlação entre as variáveis numéricas? Utilize o Power BI para tirar conclusões.

Na próxima aula...

Vamos continuar nosso estudo de visualização de dados no Power BI.

