

EDIT.

Módulo 3 – Sessão 1

EXPLORATORY
DATA ANALYSIS

TUTORA

Carla Cardoso

Freelancer AI Manager

29 de janeiro 2025




1


APRESENTAÇÕES

EDIT.


Matemáticas Aplicadas




Gestão de Informação




Gestão / Liderança




Filosofia




Consultant




Senior Consultant



Head of WAC



Head of AA & AI



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

2

2

E vocês, quem são?



3



Assiduidade e
Pontualidade
15%



Motivação e
Participação
20%



Domínio dos
Conteúdos das Aulas
15%



Aplicação
Técnica
20%



Trabalho Final
30%

4

ORGANIZAÇÃO DA SESSÃO



Alternar entre
conteúdos e
experimentação

Pausa para
jantar das
20:30 às 21:00



Quizz Final



AGENDA



INTRODUÇÃO




DADOS, DADOS E ...
MAIS DADOS



ANÁLISE E
REPRESENTAÇÃO

AGENDA


EDIT.



INTRODUÇÃO



DADOS, DADOS E ...
MAIS DADOS



ANÁLISE E
REPRESENTAÇÃO


DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

7

INTRODUÇÃO: CICLO DE VIDA PROJECTO DATA SCIENCE

EDIT.



Implementação
[Model Deployment]
Depois de seleccionado o modelo, este deverá ser implementado num ambiente de produção.

Problema
[Business understanding]
A primeira etapa é uma correcta definição do problema de negócio. Fazer as perguntas certas

Fontes de dados
[Data collection]
Identificar quais as fontes de dados disponíveis que permitam responder às perguntas levantadas na identificação do problema.

Preparação de dados
[Data Preparation]
Juntar fontes
Fazer transformações

Exploração
[EDA – Exploratory Data Analysis]
Análise Exploratória de Dados
Tipo e características das variáveis
Limpeza dos dados
Lidar com dados em falta (missing data)

Modelação
[Modelling]
Modelação estatística
Data Mining
Machine Learning
Redes Neuronais
Deep Learning

Avaliação
[Model Evaluation]
Fase de escolha do modelo a implementar

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

8

INTRODUÇÃO: ANÁLISE E EXPLORAÇÃO DE DADOS

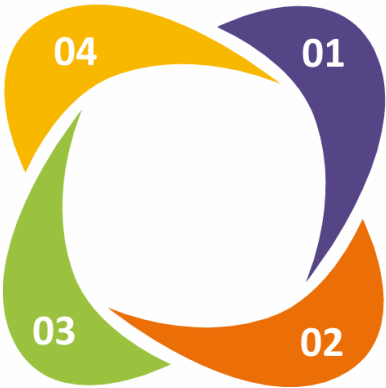


Storytelling

Depois de encontrada a solução para o problema, é necessário preparar a comunicação para a(s) audiência(s)

Análise

- Exploração com vista a:
- Conhecer melhor os dados
 - Responder às perguntas do problema



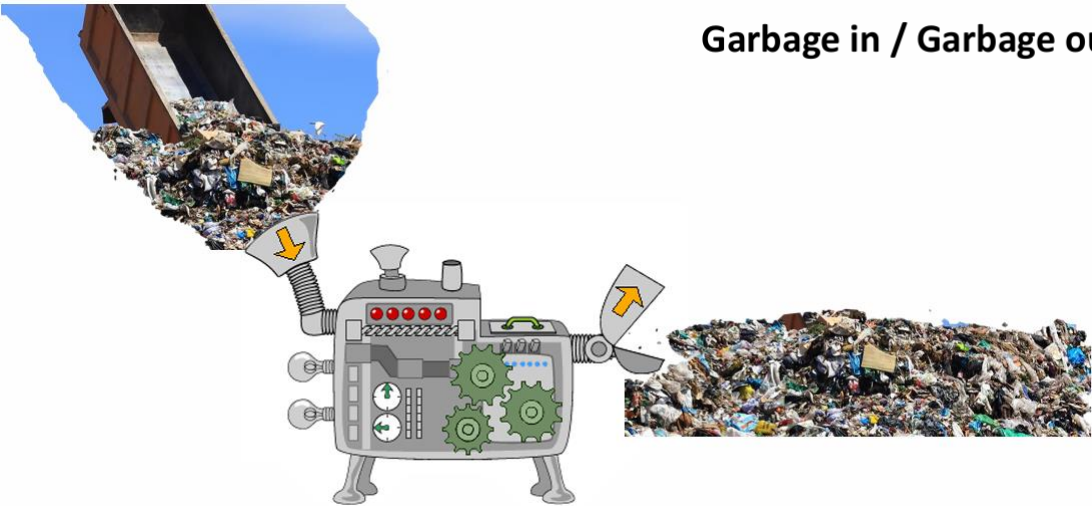
Pesquisa

Procura de fontes de dados cuja informação esteja relacionada com o problema a resolver

Preparação

Fusão de diversas fontes
Transformações necessárias para conseguir analisar a informação

INTRODUÇÃO: IMPORTÂNCIA

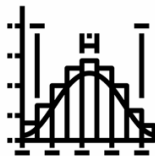


INTRODUÇÃO: OBJETIVO

E D I T.



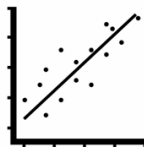
Responder a perguntas



Compreender as variáveis



Limpar os dados



Analisar a relação entre variáveis

11

AGENDA

E D I T.



INTRODUÇÃO



DADOS, DADOS E ...
MAIS DADOS

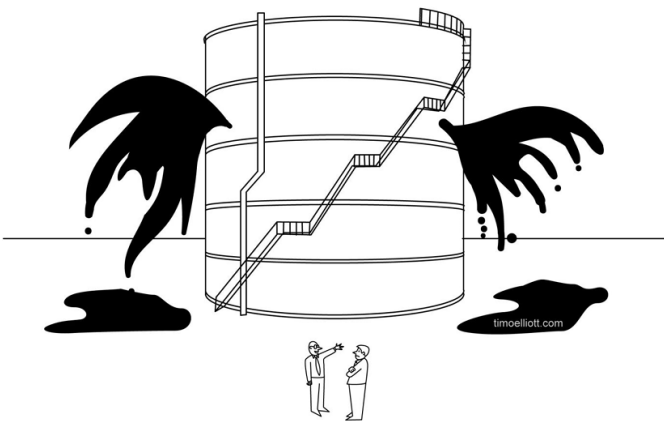


ANÁLISE E
REPRESENTAÇÃO

12

DADOS, DADOS E... MAIS DADOS: O NOVO PETRÓLEO?

EDIT.



“Data is the new oil? Absolutely—toxic if mishandled!...”

DADOS, DADOS E... MAIS DADOS: DIFERENTES MODOS DE APRESENTAÇÃO

EDIT.

DATA



DATA



SORTED



DATA



SORTED



ARRANGED



DADOS, DADOS E... MAIS DADOS: DIFERENTES MODOS DE APRESENTAÇÃO

EDIT.

DATA



SORTED



ARRANGED



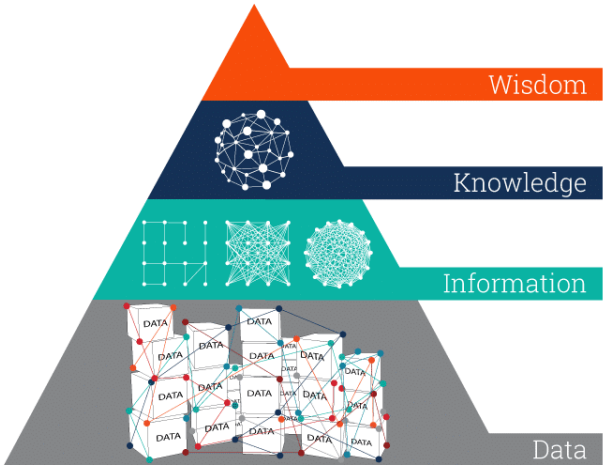
PRESENTED
VISUALLY



<https://medium.com/code-heroku/introduction-to-exploratory-data-analysis-eda-co257f888676>

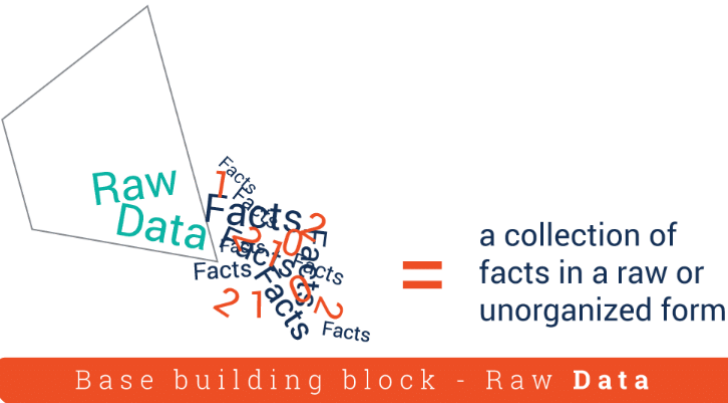
DADOS, DADOS E... MAIS DADOS: HIERARQUIA DIKW

EDIT.



Each step up
the pyramid
answers
questions
about and
adds value
to the initial data.

<https://www.ontotext.com/knowledgehub/fundamentals/dikw-pyramid/>



```
transaction_id,cust_id,tran_date,prod_subcat_code,prod_cat_code,Qty,Rate,Tax,tota_amt,Store_type
80712190438,270351,28-02-2014,1,1,-5,-772,405.3,-4265.3,e-Shop
29258453508,270384,27-02-2014,5,3,-5,-1497,785.925,-8270.925,e-Shop
51750724947,273420,24-02-2014,6,5,-2,-791,166.11,-1748.11,TeleShop
93274880719,271509,24-02-2014,11,6,-3,-1363,429.345,-4518.345,e-Shop
51750724947,273420,23-02-2014,6,5,-2,-791,166.11,-1748.11,TeleShop
97439039119,272357,23-02-2014,8,3,-2,-824,173.04,-1821.04,TeleShop
45649838090,273667,22-02-2014,11,6,-1,-1450,152.25,-1602.25,e-Shop
22643667930,271489,22-02-2014,12,6,-1,-1225,128.625,-1353.625,TeleShop
79792372943,275108,22-02-2014,3,1,-3,-908,286.02,-3010.02,MBR
50076728598,269014,21-02-2014,8,3,-4,-581,244.02,-2568.02,e-Shop
29258453508,270384,20-02-2014,5,3,5,1497,785.925,8270.925,e-Shop
25455265351,267750,20-02-2014,12,6,3,1360,428.4,4508.4,e-Shop
1571002198,275023,20-02-2014,6,5,4,587,246.54,2594.54,e-Shop
43134751727,268487,20-02-2014,3,2,-1,-611,64.155,-675.155,e-Shop
36554696014,269345,20-02-2014,3,5,3,1253,394.695,4153.695,e-Shop
56814940239,268799,20-02-2014,7,5,5,368,193.2,2033.2,e-Shop
54295803788,270787,20-02-2014,12,5,5,584,306.6,3226.6,e-Shop
25963520987,274829,20-02-2014,4,4,3,502,158.13,1664.13,Flagship store
17183929085,266863,20-02-2014,1,2,1,1359,142.695,1501.695,TeleShop
44783317894,269452,20-02-2014,3,1,3,825,259.875,2734.875,TeleShop
58957828601,269009,20-02-2014,11,5,4,846,355.32,3739.32,TeleShop
9647515879,268934,20-02-2014,7,5,3,1384,435.96,4587.96,TeleShop
20781077773,269223,20-02-2014,5,3,3,1203,378.945,3987.945,TeleShop
91116291703,268509,20-02-2014,1,2,4,1243,522.06,5494.06,MBR
```

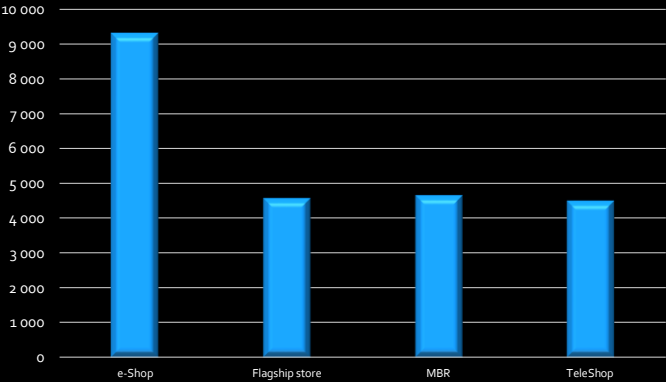
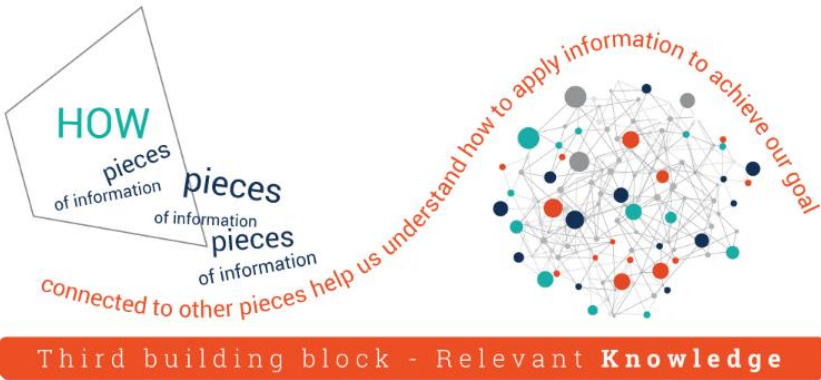
who
what
when
where

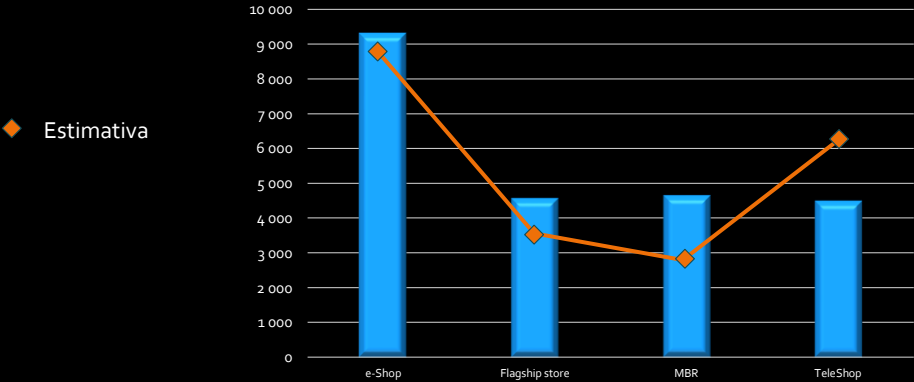
=

easier to measure,
visualize and analyze
data for a specific purpose

Second building block - Derived Information

transaction_id	cust_id	tran_date	prod_subcat_code	prod_cat_code	Qty	Rate	Tax	total_amt	Store_type
80712190438	270351	28/02/2014	1	1	-5	-772	405.3	-4265.3	e-Shop
29258453508	270384	27/02/2014	5	3	-5	-1497	785.925	-8270.925	e-Shop
51750724947	273420	24/02/2014	6	5	-2	-791	166.11	-1748.11	TeleShop
93274880719	271509	24/02/2014	11	6	-3	-1363	429.345	-4518.345	e-Shop
51750724947	273420	23/02/2014	6	5	-2	-791	166.11	-1748.11	TeleShop
97439039119	272357	23/02/2014	8	3	-2	-824	173.04	-1821.04	TeleShop
45649838090	273667	22/02/2014	11	6	-1	-1450	152.25	-1602.25	e-Shop
22643667930	271489	22/02/2014	12	6	-1	-1225	128.625	-1353.625	TeleShop
79792372943	275108	22/02/2014	3	1	-3	-908	286.02	-3010.02	MBR
50076728598	269014	21/02/2014	8	3	-4	-581	244.02	-2568.02	e-Shop
29258453508	270384	20/02/2014	5	3	5	1497	785.925	8270.925	e-Shop
25455265351	267750	20/02/2014	12	6	3	1360	428.4	4508.4	e-Shop
1571002198	275023	20/02/2014	6	5	4	587	246.54	2594.54	e-Shop
43134751727	268487	20/02/2014	3	2	-1	-611	64.155	-675.155	e-Shop
36554696014	269345	20/02/2014	3	5	3	1253	394.695	4153.695	e-Shop
56814940239	268799	20/02/2014	7	5	5	368	193.2	2033.2	e-Shop
54295803788	270787	20/02/2014	12	5	5	584	306.6	3226.6	e-Shop
25963520987	274829	20/02/2014	4	4	3	502	158.13	1664.13	Flagship store
17183929085	266863	20/02/2014	1	2	1	1359	142.695	1501.695	TeleShop
44783317894	269452	20/02/2014	3	1	3	825	259.875	2734.875	TeleShop
58957828601	269009	20/02/2014	11	5	4	846	355.32	3739.32	TeleShop
9647515879	268934	20/02/2014	7	5	3	1384	435.96	4587.96	TeleShop
20781077773	269223	20/02/2014	5	3	3	1203	378.945	3987.945	TeleShop
91116291703	268509	20/02/2014	1	2	4	1243	522.06	5494.06	MBR





DADOS, DADOS E... MAIS DADOS

EDIT.

BORA LÁ POR A MÃO NA MASSA



Inteligência Artificial: Gen AI & LLM


Módulo 1: Inteligência Artificial – Sessão 2

27


27

EXERCICIO EM EQUIPA #1


EDIT.




Equipa 1 (C)



Equipa 2



Equipa 3



Equipa 4 (P)

Slide 1	Andreia	Carolina M.	Ana	Alexandre
Slide 2	Carolina L.	Joana	José F.	Filipa
Slide 3	Stefane /Yohanna	José P M	José M.	João B.
Slide 4	Tamara	Sara	Susana	Rui

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

28

28

EXERCICIO EM EQUIPA #1

EDIT.



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

29

Ser-vos-á disponibilizado um conjunto de dados referentes às **transações diárias** de uma cadeia de **retalho**.

O objetivo deste desafio é transformar esse conjunto de **DADOS** em **SABEDORIA!**

Para tal vocês devem:

- **Importar** os dados para **excel**
- **Interpretar** o seu conteúdo
- Identificar possíveis **problemas nos dados** e corrigi-los
- Identificar e especificar uma **pergunta de negócio** que achem pertinente responder
- Dar **resposta à pergunta** de negócio com base no **análise** dos dados

29

EXERCICIO EM EQUIPA #1

EDIT.



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

30

O resultado do trabalho é um relatório, a ser entregue no final dos **30 minutos** que têm disponíveis para a realização deste exercício.

O relatório deve ter **EXATAMENTE 4 slides**, organizado da seguinte forma:

1. Apresentação dos **dados**
2. **Especificação do problema** que pretendem revolver
3. **Conclusões** alcançadas
4. **Próximos passos** que consideram relevantes para evolução do vosso trabalho

No final, escolhido **1 grupo para apresentar** o seu relatório (10 minutos) e **1 grupo para comentar** os resultados apresentados (10 minutos).

A escolha dos destes 2 grupos será **aleatória**.

30

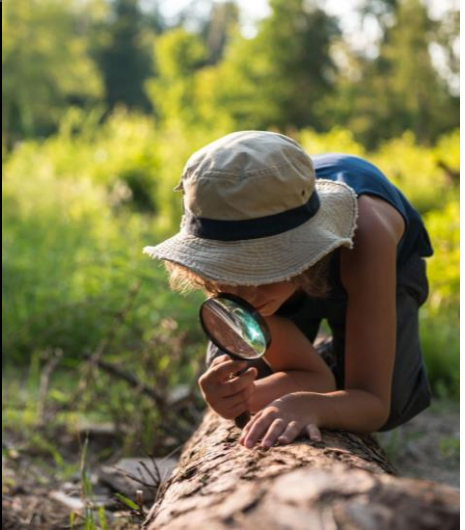
DADOS, DADOS E... MAIS DADOS: GARANTIR QUE TÊM UTILIDADE

EDIT.

Antes de iniciar qualquer análise, estudo, modelação, **OLHAR PARA OS DADOS!**

Verificar se:

- ✓ Os dados são todos **relevantes** para o trabalho que estamos a realizar
- ✓ Existem **duplicados**
- ✓ O **nível de detalhe** é o correto (ex.: por transação, por dia, etc.)
- ✓ Há **coerência** da informação (ex.: vendas aos fim de semana)
- ✓ É necessário alterar/corrigir **formatos** (datas, inteiro/decimal, ...)
- ✓ Todas as variáveis **têm informação** (missing data)
- ✓ Existem problemas de **escalas** ou valores **extremos** muito elevados (ex.: outliers)



DADOS, DADOS E... MAIS DADOS: GARANTIR QUE TÊM UTILIDADE

EDIT.


© NAEK ANDERSON, WWW.ANDERSTOONS.COM




"After analyzing all your data, I think we can safely say that none of it is useful."

AGENDA


EDIT.



INTRODUÇÃO



DADOS, DADOS E ...
MAIS DADOS



ANÁLISE E
REPRESENTAÇÃO

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

33

ANÁLISE E REPRESENTAÇÃO

EDIT.

TIPOS DE VARIÁVEIS

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

34

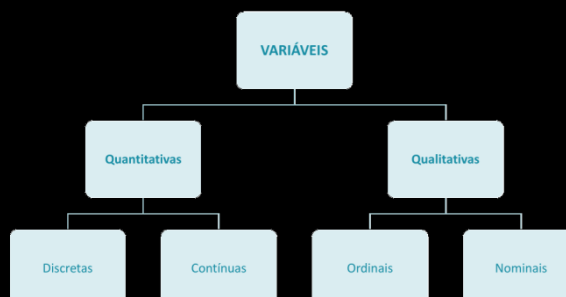
TIPOS DE VARIÁVEIS

EDIT.

Vamos chamar de **variável** a uma **característica comum**, que assume **valores diferentes** de indivíduo para indivíduo. Os indivíduos podem ser pessoas, empresas, transações, livros, etc.

Exemplo de variáveis:

- ❖ **Idade** de pessoas
- ❖ **Data de realização** de transações
- ❖ **Dimensão populacional** de países
- ❖ **Número de páginas** de livros
- ❖ **Realizador** de filmes
- ❖ ...



As variáveis podem ser **quantitativas** ou **qualitativas**, como vamos ver.

35

TIPOS DE VARIÁVEIS: QUANTITATIVAS

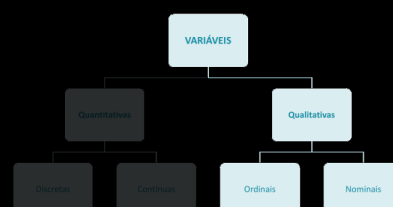
EDIT.

As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

Exemplo de variáveis:

- ❖ **Nível de escolaridade** de pessoas
- ❖ **Tipologia** de transações
- ❖ **Pertença ou não à EU** para os países
- ❖ **Score de Risco** de uma empresa
- ❖ **Marca** de um equipamento
- ❖ ...



36

TIPOS DE VARIÁVEIS: QUANTITATIVAS



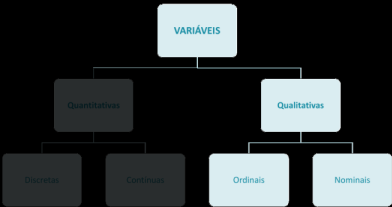
As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

Exemplo de variáveis:

ORDINAL

- ❖ **Nível de escolaridade** de pessoas
- ❖ **Tipologia** de transações
- ❖ **Pertença ou não à EU** para os países
- ❖ **Score de Risco** de uma empresa
- ❖ **Marca** de um equipamento
- ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

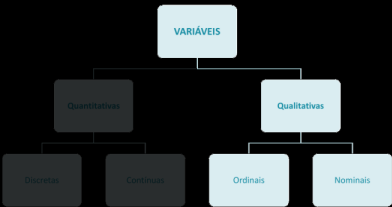
Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

Exemplo de variáveis:

ORDINAL

NOMINAL

- ❖ **Nível de escolaridade** de pessoas
- ❖ **Tipologia** de transações
- ❖ **Pertença ou não à EU** para os países
- ❖ **Score de Risco** de uma empresa
- ❖ **Marca** de um equipamento
- ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

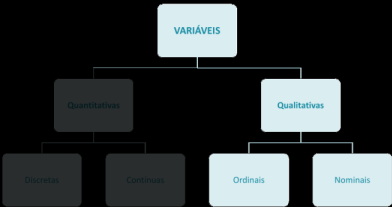
Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

Exemplo de variáveis:

- ORDINAL

NOMINAL

ORDINAL
- ❖ **Nível de escolaridade** de pessoas
 - ❖ **Tipologia** de transações
 - ❖ **Pertença ou não à EU** para os países
 - ❖ **Score de Risco** de uma empresa
 - ❖ **Marca** de um equipamento
 - ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

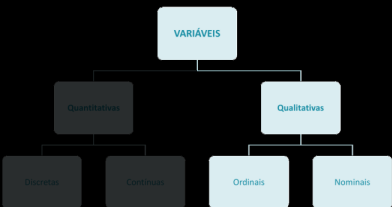
Exemplo de variáveis:

- ORDINAL

NOMINAL

ORDINAL

ORDINAL
- ❖ **Nível de escolaridade** de pessoas
 - ❖ **Tipologia** de transações
 - ❖ **Pertença ou não à EU** para os países
 - ❖ **Score de Risco** de uma empresa
 - ❖ **Marca** de um equipamento
 - ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS

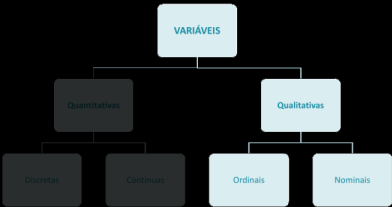


As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

Adicionalmente estas variáveis ainda se dividem entre **Ordinais**, quando a categoria permite uma ordenação, e **Nominais**, quando as categorias não permitem ordenação.

Exemplo de variáveis:

- ORDINAL** ❖ **Nível de escolaridade** de pessoas
- NOMINAL** ❖ **Tipologia** de transações
- ORDINAL** ❖ **Pertença ou não à EU** para os países
- ORDINAL** ❖ **Score de Risco** de uma empresa
- NOMINAL** ❖ **Marca** de um equipamento
- ❖ ...



41

TIPOS DE VARIÁVEIS: QUANTITATIVAS

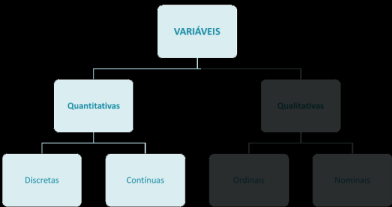


As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

Exemplo de variáveis:

- ❖ **Número de filhos** de pessoas
- ❖ **Valor** de transações
- ❖ **Duração** para filmes
- ❖ **Número de acidentes** em cruzamentos
- ❖ **Distancia** entre 2 cidades
- ❖ ...



42

TIPOS DE VARIÁVEIS: QUANTITATIVAS



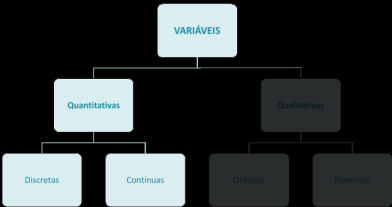
As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

Exemplo de variáveis:

DISCRETA

- ❖ **Número de filhos** de pessoas
- ❖ **Valor** de transações
- ❖ **Duração** para filmes
- ❖ **Número de acidentes** em cruzamentos
- ❖ **Distancia** entre 2 cidades
- ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

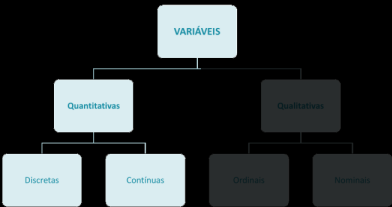
Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

Exemplo de variáveis:

DISCRETA

CONTÍNUA

- ❖ **Número de filhos** de pessoas
- ❖ **Valor** de transações
- ❖ **Duração** para filmes
- ❖ **Número de acidentes** em cruzamentos
- ❖ **Distancia** entre 2 cidades
- ❖ ...



TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

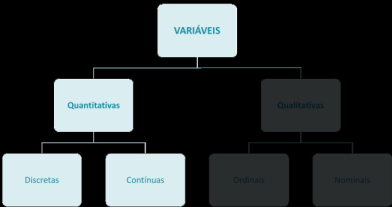
Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

Exemplo de variáveis:

- DISCRETA

CONTÍNUA

CONTÍNUA
- ❖ **Número de filhos** de pessoas
 - ❖ **Valor** de transações
 - ❖ **Duração** para filmes
 - ❖ **Número de acidentes** em cruzamentos
 - ❖ **Distancia** entre 2 cidades
 - ❖ ...



45

TIPOS DE VARIÁVEIS: QUANTITATIVAS



As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

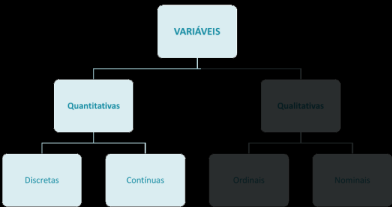
Exemplo de variáveis:

- DISCRETA

CONTÍNUA

CONTÍNUA

DISCRETA
- ❖ **Número de filhos** de pessoas
 - ❖ **Valor** de transações
 - ❖ **Duração** para filmes
 - ❖ **Número de acidentes** em cruzamentos
 - ❖ **Distancia** entre 2 cidades
 - ❖ ...



46

TIPOS DE VARIÁVEIS: QUANTITATIVAS

EDIT.

As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, e **Contínua**, quando representam medidas contínuas.

Exemplo de variáveis:

- | | |
|----------|---|
| DISCRETA | ❖ Número de filhos de pessoas |
| CONTÍNUA | ❖ Valor de transações |
| CONTÍNUA | ❖ Duração para filmes |
| DISCRETA | ❖ Número de acidentes em cruzamentos |
| CONTÍNUA | ❖ Distancia entre 2 cidades |
| | ❖ ... |



ANÁLISE E REPRESENTAÇÃO

EDIT.

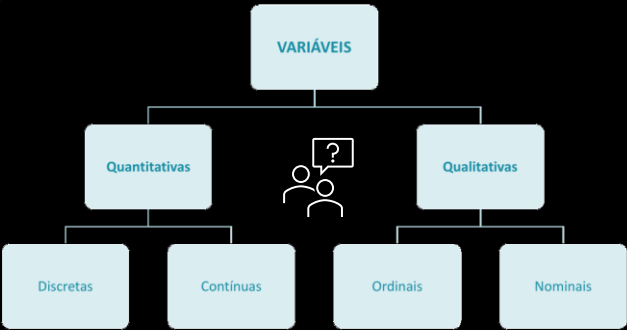
ORGANIZAÇÃO DOS DADOS

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**

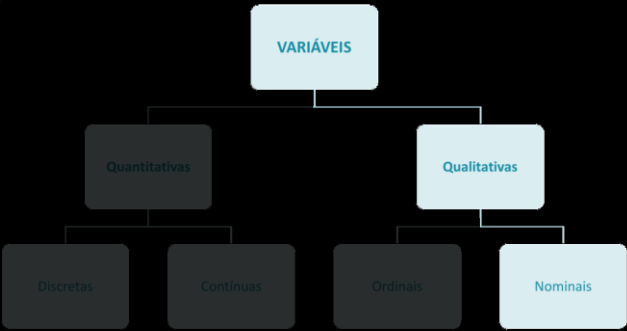


ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**



ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**

Tipo de Loja	Freq. Total	Freq. Relativa
e-Shop	9 311	40%
Flagship store	4 577	20%
MBR	4 661	20%
TeleShop	4 504	20%
Total	23053	100%

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**

Tipo de Loja	Freq. Total	Freq. Relativa
e-Shop	9 311	40%
Flagship store	4 577	20%
MBR	4 661	20%
TeleShop	4 504	20%
Total	23053	100%

$$\frac{\text{Frequência Total}}{\text{Total das Observações}}$$

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**

Tipo de Loja	Freq. Total	Freq. Relativa
e-Shop	9 311	40%
Flagship store	4 577	20%
MBR	4 661	20%
TeleShop	4 504	20%
Total	23053	100%

Frequência Total
Total das Observações

Total das Observações

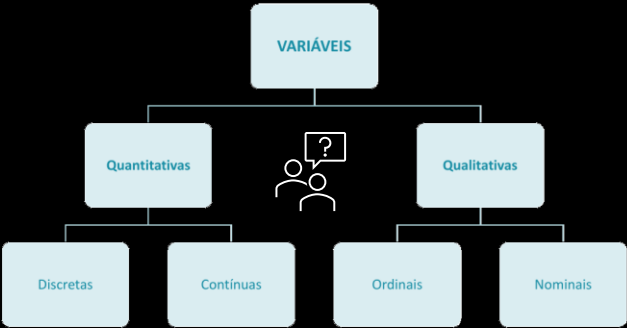
Sempre igual a 100%

53

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

Novo exemplo: **Rate**



54

Novo exemplo: **Rate**



Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

1. Calcular a amplitude total da variável → **Amplitude_Total** = (máximo – mínimo)

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

1. Calcular a amplitude total da variável → **Amplitude_Total** = (máximo – mínimo)

$$\begin{aligned} \text{Mínimo} &= -1.499 \\ \text{Máximo} &= 1.500 \\ \text{Amplitude_Total} &= 1.500 - (-1.499) = \mathbf{2.999} \end{aligned}$$

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

1. Calcular a amplitude total da variável → **Amplitude_Total** = 2.999
2. Calcular a amplitude dos intervalos como a divisão da amplitude total (passo 1) pelo número de classes **k** que queremos → **Amplitude_Intervalo** = $\text{Amplitude_Total} / k$

$$k = 5$$

$$\text{Amplitude_Intervalo} = 2.999 / 5 = 599.8 \sim 600$$

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

1. Calcular a amplitude total da variável → **Amplitude_Total** = 2.999
2. Calcular a amplitude dos intervalos → **Amplitude_Intervalo** = 600
3. Construir as classes da seguinte forma:
 - a) Classe 1: [mínimo : mínimo + Amplitude_Intervalo [

$$\text{Classe 1} = [-1.499 : -1.499 + 600 [= [-1.499 : -899[$$

Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

- 1. Calcular a amplitude total da variável → **Amplitude_Total** = 2.999
- 2. Calcular a amplitude dos intervalos → **Amplitude_Intervalo** = 600
- 3. Construir as classes da seguinte forma:
 - a) **Classe 1**: [-1.499 : -899[
 - b) **Classe 2**: [mínimo + Amplitude_Intervalo : mínimo + 2 * Amplitude_Intervalo [

Classe 2 = [-899 : -1.499 + 2 * 600 [= [-899 : -299[

Novo exemplo: **Rate**

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um **processo de discretização** para ser possível representa-las.

Etapas do processo de discretização:

- 1. Calcular a amplitude total da variável → **Amplitude_Total** = 2.999
- 2. Calcular a amplitude dos intervalos → **Amplitude_Intervalo** = 600
- 3. Construir as classes da seguinte forma:
 - a) **Classe 1**: [-1.499 : -899 [
 - b) **Classe 2**: [-899 : -299 [
 - c) ...
 - d) **Classe k**: [mínimo + (k-1) Amplitude_Intervalo : mínimo + k * Amplitude_Intervalo [

Classe k = [-1.499 + (5-1) * 600 : -1.499 + 5 * 600 [
= [901 : 1.501[

Novo exemplo: Rate

No caso das variáveis quantitativas, e especialmente no caso das quantitativas contínuas, temos de realizar um processo de discretização para ser possível representa-las.

Etapas do processo de discretização:

- 1. Calcular a amplitude total da variável → Amplitude_Total = 2.999
- 2. Calcular a amplitude dos intervalos → Amplitude_Intervalo = 600
- 3. Construir as classes da seguinte forma:
 - a) Classe 1: [-1.499 : -899 [
 - b) Classe 2: [-899 : -299 [
 - c) ...
 - d) Classe 5: [901 : 1.501 [

Novo exemplo: Rate

Rate Class	Freq. Total	Freq. Relativa
[-1499 : -899 [942	4%
[-899 : -299 [875	4%
[-299 : 301 [3 773	16%
[301 : 901 [8 671	38%
[901 : 1501 [8 792	38%
Total	23 053	100%

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



Novo exemplo: Rate

Rate Class	Freq. Total	Freq. Relativa
[-1499 : -899 [942	4%
[-899 : -299 [875	4%
[-299 : 301 [3 773	16%
[301 : 901 [8 671	38%
[901 : 1501 [8 792	38%
Total	23 053	100%

Mas como definir o k?
Um regra da estatística é considerar como k o menor inteiro tal que $2^k \geq \text{dimensão}$ da dados.

ORGANIZAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA



Novo exemplo: Rate

Rate Class	Freq. Total	Freq. Relativa
[-1499 : -899 [942	4%
[-899 : -299 [875	4%
[-299 : 301 [3 773	16%
[301 : 901 [8 671	38%
[901 : 1501 [8 792	38%
Total	23 053	100%

Mas como definir o k?
Um regra da estatística é considerar como k o menor inteiro tal que $2^k \geq \text{dimensão}$ da dados.

No nosso caso, temos que a dimensão dos dados é 23.053, logo $k = 15$, pois $2^{14} = 16.384$

ANÁLISE E REPRESENTAÇÃO

EDIT.

VISUALIZAÇÃO DOS DADOS

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

67

67

VISUALIZAÇÃO DOS DADOS: PIE CHART

EDIT.

A partir dos dados organizados é simples criar **representações gráficas** para melhor visualização.

A representação gráfica mais comum para dados **qualitativos** é o **Pie Chart**:

Distribuição por Tipo de Loja

Tipo de Loja	Valor	Porcentagem
e-Shop	9 311	40%
Flagship store	4 504	20%
MBR	4 661	20%
TeleShop	4 577	20%

- e-Shop
- Flagship store
- MBR
- TeleShop

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

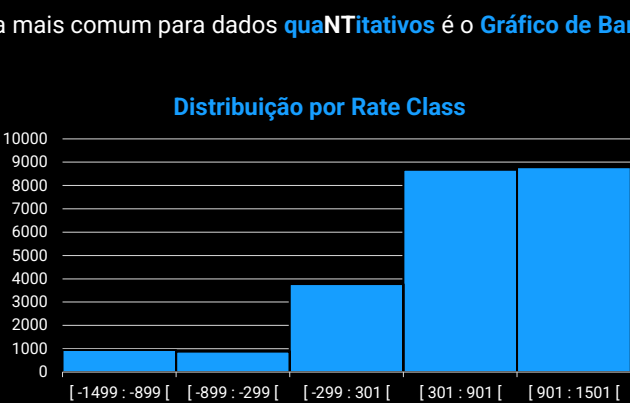
68

68

VISUALIZAÇÃO DOS DADOS: GRÁFICO DE BARRAS

E D I T.

A partir dos dados organizados é simples criar **representações gráficas** para melhor visualização.
A representação gráfica mais comum para dados **quanTitativos** é o **Gráfico de Barras**:

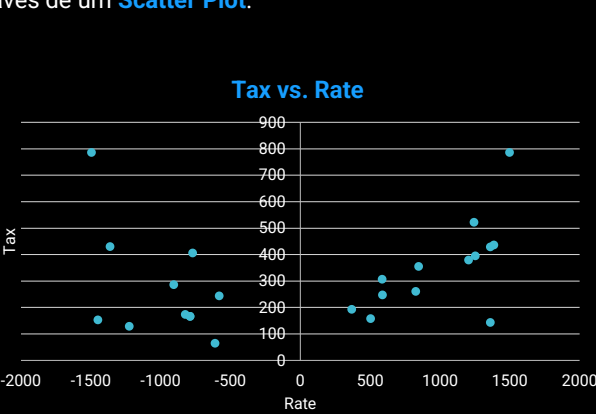


69

VISUALIZAÇÃO DOS DADOS: SCATTER PLOT

E D I T.

Continuando nas variáveis quantitativas, podemos querer analisar relações entre 2 variáveis, e a forma mais simples de o fazer é através de um **Scatter Plot**:



70

EXPLORATORY DATA ANALYSIS

EDIT.

Questionário

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 3 – Exploratory Data Analysis – Sessão 1

75