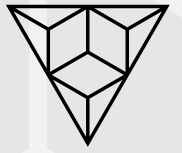


# Análise de Dados e Power BI

[ INTEGRIDADE DOS DADOS – AULA 3 ]



# **MAPA DO CURSO**

**AULA 1 – MUNDO DOS DADOS**

**AULA 2 – PENSAMENTO ANALÍTICO**

**AULA 3 - INTEGRIDADE DE DADOS**

**AULA 4 – FUNDAMENTOS DE BI**

**AULA 5 – JORNADA POWER BI - 1**

**AULA 6 – JORNADA POWER BI - 2**

**AULA 7 – LABORATÓRIO – Parte 1**

**AULA 8 – LABORATÓRIO – Parte 2**

**AULA 9 – VISUALIZAÇÃO DE DADOS**

**AULA 10 – O ANALISTA FORA DA CURVA**

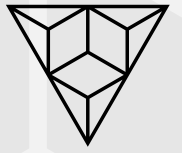
# 03

# INTEGRIDADE

## Integridade dos Dados

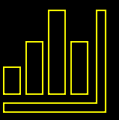
Confiança e assertividade na leitura dos dados.





# missão da aula

Aprender a importância de **dados inteligentes** e entender conceitos fundamentais para trabalhar com **bases padronizadas e confiáveis.**

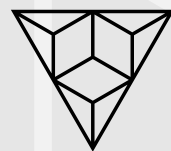




# DADOS INTELIGENTES

afinal, como garantir que os seus dados sejam inteligentes?





# COMO TER DADOS INTELIGENTES?

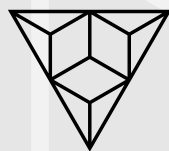
Para que isso aconteça é preciso dar atenção a 3 fatores:

**01** Dados Qualificados

**02** Modelos Estatísticos

**03** Digitalização + Inovação



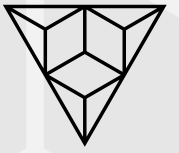


# DADOS QUALIFICADOS

Anteriormente já vimos como é importante termos dados qualificados, ter uma base de dados em que você tenha segurança de tomar ações baseada em análises que faz.

Para que isso seja possível é muito importante que você cuide da **coleta dos dados**, a **organização** deles e da **limpeza** de informações.



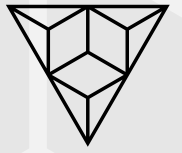


# MODELOS ESTATÍSTICOS

Enquanto os modelos estatísticos vão enriquecer sua atividade, trazendo **confiabilidade** para a análise e **segurança** na informação que está sendo extraída.







# DIGITALIZAÇÃO E INOVAÇÃO

Já a digitalização e inovação está relacionada à adaptação da empresa dentro do cenário de revolução digital, trazendo a **capacidade de gerir novas tecnologias** que suportem a manipulação do grande volume de dados.





# “Data Science”

## Ciência de Dados

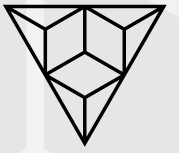
Quando o tema é inovação de tecnologia esse é um dos termos mais falados atualmente. E com ele outros diversos temas começam a fazer parte da rotina no universo dos dados.

*Inteligência Artificial...*

*Machine Learning....*

*Deep Learning...*





# Mas pra serve tudo isso?

**01 Tirar insights de uma grande gama de dados**  
(Big Data)

**02 Tomada de decisões inteligentes**

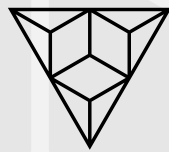
Hoje conseguimos fazer com que máquinas tomem decisões e façam previsões baseadas em análise de dados e como analistas precisamos estar antenados e podemos usar os benefícios da tecnologia ao nosso favor.





quem tem **certeza**  
dos dados que tem  
**confia** nas sugestões  
que faz.





# Estatística Essencial

É muito importante sempre manter frescos os conceitos básicos de estatística que podem te dar aquela mãozinha no dia-a-dia:

01 **Amostra mínima**

02 **Desvio Padrão**

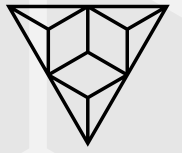
03 **Grau de Confiança**

04 **Margem de Erro**

05 **População**

06 **Amostra**





# Tipos de Amostra Mínima

## **por variável:**

utilizada quando se tem uma base que mostra o desvio padrão de uma informação específica a ser investigada

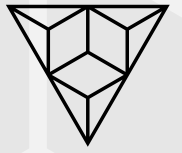
## **por população:**

utilizada quando se sabe o tamanho da população

## **por proporção:**

utilizada quando a proporção entre uma característica analisada e a população é conhecida





# Alguns conceitos importantes que precisam estar na ponta da língua:

## **população:**

conjunto de indivíduos que compartilham uma característica comum

## **amostra:**

subconjunto de indivíduos extraídos da população

## **margem de erro:**

diferença entre o resultado obtido com a amostra e o resultado real da população

## **grau de confiança:**

nível de certeza de que os dados medidos refletem a população





# MARGEM DE ERRO

# X

# GRAU DE CONFIANÇA

**exemplo:**

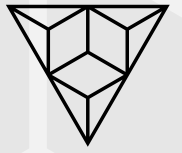
com 95% de confiança, posso afirmar que 52% dos dos clientes preferem o produto A, com uma margem de erro de 1% para cima ou para baixo.

**ou seja:**

tenho **95%** de certeza que entre **51 e 53%** dos clientes preferem o produto A.







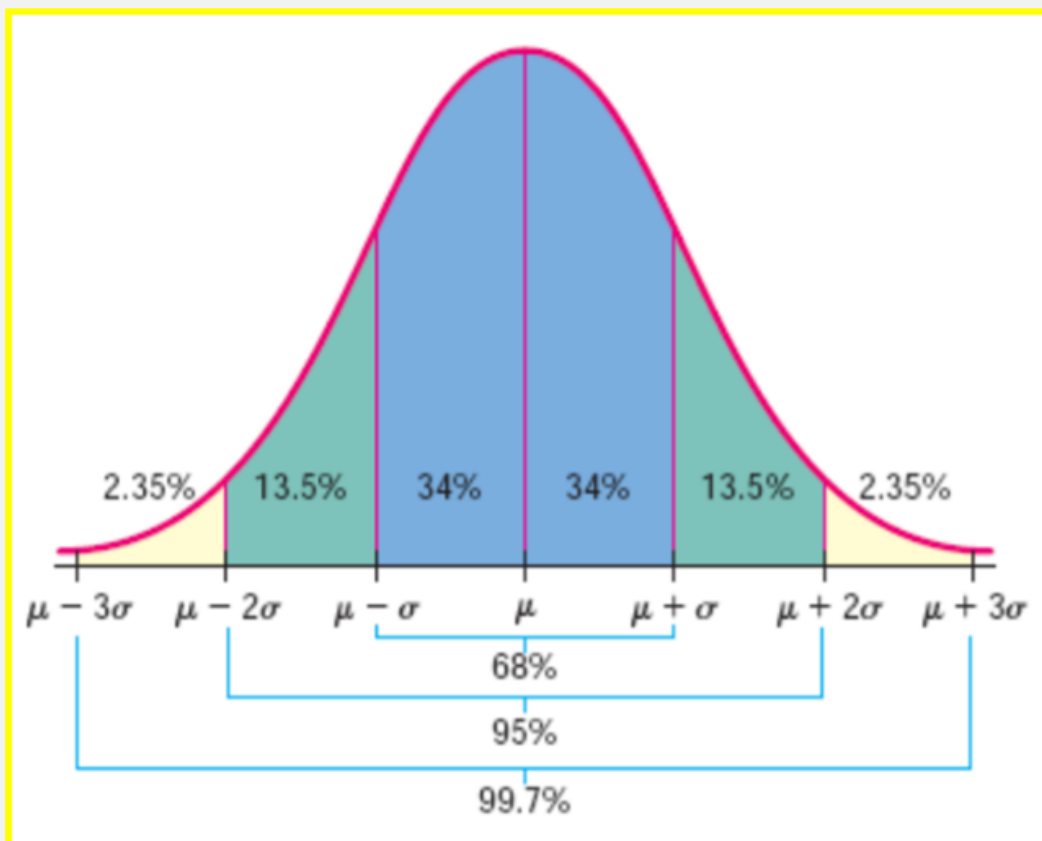
# Conceitos Estatísticos

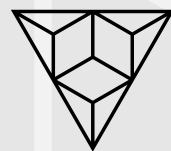
## curva normal:

tendência dos dados a se distribuírem em volta da média

## desvio padrão ( $\sigma$ ):

quantifica a dispersão dos valores de uma amostra





# Desvio Padrão no Excel

## **DESVPAD.A:**

calcula o desvio padrão a partir de uma amostra

## **DESVPAD.P:**

calcula o desvio padrão a partir de uma população

Aluno	Turma	Nota
1	A	9
2	A	6
3	A	4
4	A	10
5	A	4
6	A	6
7	A	7
8	A	7
9	B	6
10	B	5
11	B	5
12	B	10
13	B	10
14	B	8
15	B	8
16	B	8
17	B	9
18	C	8
19	C	9
20	C	10
21	C	6
22	C	7
23	C	4
24	C	5
25	C	6

Turma A
2,13

Turma B
1,94

Turma C
2,03

Geral
1,96





quem tem **certeza**  
dos dados que tem  
**confia** nas sugestões  
que faz.



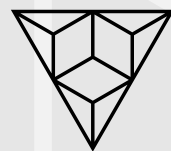


# LIMPEZA DE DADOS

Se te perguntarem agora, quais dados você tem à disposição, onde estão armazenados e quais são as possíveis conclusões que você pode chegar através deles, você saberia responder?

Grande parte das empresas não sabem responder essas perguntas com precisão. Mas num mundo onde dados são o novo petróleo, conhecer seus dados é fundamental.





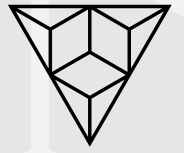
# pulo do gato

**Conheça seus  
dados como a  
palma da sua  
mão!**

Assegure-se que  
sua base está  
limpa e  
padronizada.

Se estiver seguro,  
 **siga em frente**





# Identificando Dados Ruins

Existem várias maneiras de identificar os dados ruins, ou seja, informações não confiáveis dentro da sua base.

Vamos focar nas 4 principais:

## **dados ilógicos:**

quando um dado não condiz com a coluna onde aparece ou é surreal comparado aos demais da mesma coluna

## **velocidade de respostas:**

quando uma pesquisa tem entradas com tempo de resposta muito abaixo da média

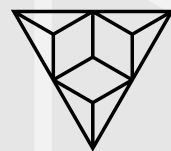
## **dados não padronizados:**

quando existe um padrão exato para os dados mas eles não são respeitados pela entrada

## **valores de comparação:**

quando um dado possui valores claros de referência que não atendidos por uma entrada

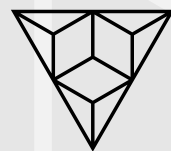




# pulo do gato

Crie **métodos automatizados** de identificação de erros e problemas com informações dentro da sua base.





# FÓRMULAS ÚTEIS

## **ESQUERDA / DIREITA**

Retorna o primeiro caractere ou caracteres em uma cadeia de texto baseado no número de caracteres especificado por você a partir da esquerda ou direita.

## **PROCURAR**

As funções PROCURAR e PROCURARB localizam uma cadeia de texto em uma segunda cadeia de texto e retornam o número da posição inicial da primeira cadeia de texto do primeiro caractere da segunda cadeia de texto.

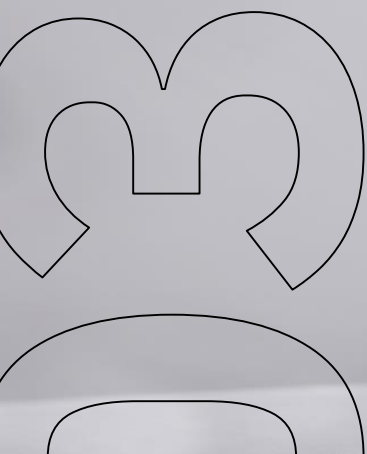
## **SE / E / OU**

As funções permitem fazer uma comparação lógica entre um valor e o que você espera, testando uma condição e retornando um resultado se for Verdadeiro ou Falso.

## **FORMATAÇÃO CONDICIONAL**







## desafio Conquer #03

**1.** Utilize os conhecimentos que você aprendeu ao longo da aula para **implementar** ou **reestruturar** uma análise no seu trabalho.

**2.** Identifique uma base de dados do seu trabalho ou da sua vida pessoal, e liste quais **tipos de dados não íntegros** você possui e possíveis meios de **automatizar** a identificação.

**3.** Busque mais exercícios e conteúdos para praticar seu conhecimento em **estatística**.



**sem dados** você é apenas mais  
uma pessoa **com uma opinião.**



*William Edwards*  
*escritor e matemático, séc XX*