

# Data Science : Apresentação da Disciplina

Prof. Ovídio Francisco



## Plano de aula

Data Science

Dados

Informação

Análise de Dados

# Introdução

# O contexto

Data Science é o **estudo** dos dados com propósito de **extrair conhecimentos** úteis sobre um assunto. Para isso, faz uso de:

- Data mining
- Big data
- Machine learning
- Estatística
- Algoritmos

“A ciência de dados combina vários campos, incluindo estatísticas, métodos científicos e análise de dados para extrair valor dos dados.”

<https://www.oracle.com/br/data-science/what-is-data-science/>

# Dados



Mas... o que são dados?



Acesse [menti.com](https://www.menti.com) e use o código **37 75 37 7**

# Dados

Um dado é o resultado de uma **observação**. É uma anotação que por si, **não tem valor**.

Exemplos:

Alguém mede um objeto e observa 1,25 metros

Alguém observa o termômetro que marca 34°C

Alguém vê no calendário a data 18/2

Após registrados, os valores

**1,25**

**34**

**18/2**

**são dados**

# Informação



# Informação

Informação é o **dado contextualizado**. Está associado a uma **necessidade**.

Se dissermos

A altura é **1,25**

A Temperatura do ambiente é **34°C**

O aniversário é **18/2**

**são informações**

**Saiba mais:** Na teoria matemática da informação de Shannon, informação é a quantidade de incerteza de um evento. Veja <http://www.bocc.ubi.pt/pag/fidalgo-antonio-quadros-incerteza.pdf>



# Análise de dados

Imagine uma desenvolvedora de jogos tem já tem no mercado dois games:

**Dark Blood**, um jogo de ação com elementos gore

e

**Falling Blocks**, um jogo casual no estilo puzzle.

# Análise de Dados

Imagine também que os jogadores avaliaram cada game com uma nota entre 0 e 100. Observe os dados na tabela tabela abaixo:

Jogador	Dark blood	Falling blocks
A	74	23
B	84	41
C	12	88
D	32	71
E	19	79
F	89	32
G	99	29
H	22	90
I	89	15
J	76	28
K	79	21
L	26	57
M	11	68
N	16	73
O	95	24
P	28	92
Q	82	62
R	71	45
S	18	95

# Análise de Dados

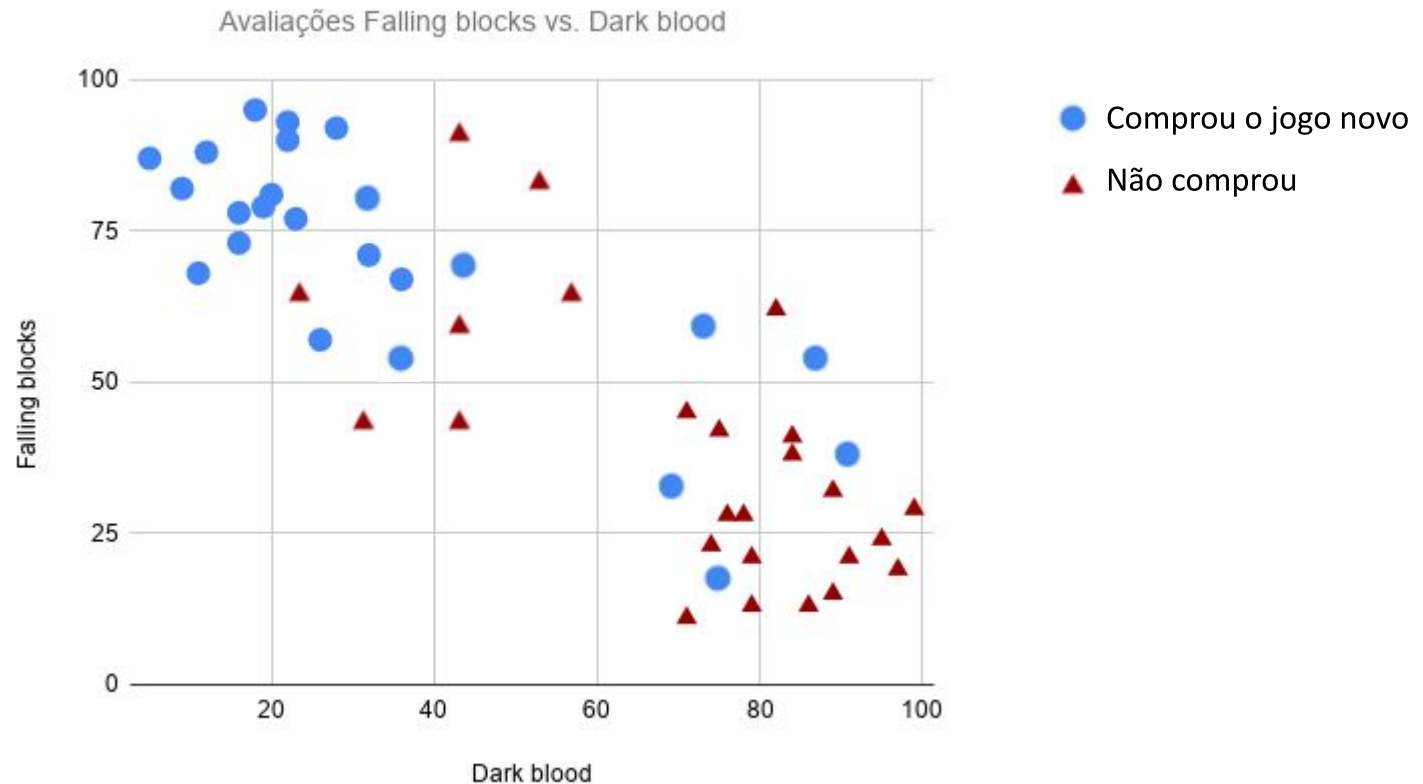
Agora acrescente à história que a desenvolvedora criou um novo jogo, o **Dancing Beans**, e o ofereceu aos usuários de **Dark Blood** e **Falling Blocks** por meio de anúncios.

Alguns desses adquiriram o novo game e outros não, conforme os dados abaixo:

Jogador	Dark blood	Falling blocks	Comprou Dancing Beans
A	74	23	N
B	84	41	N
C	12	88	S
D	32	71	S
E	19	79	S
F	89	32	N
G	99	29	N
H	22	90	S
I	89	15	N
J	76	28	N
K	79	21	N
L	26	57	S
M	11	68	S
N	16	73	S
O	95	24	N
P	28	92	S
Q	82	62	N
R	71	45	N
S	18	95	S

# Análise de Dados

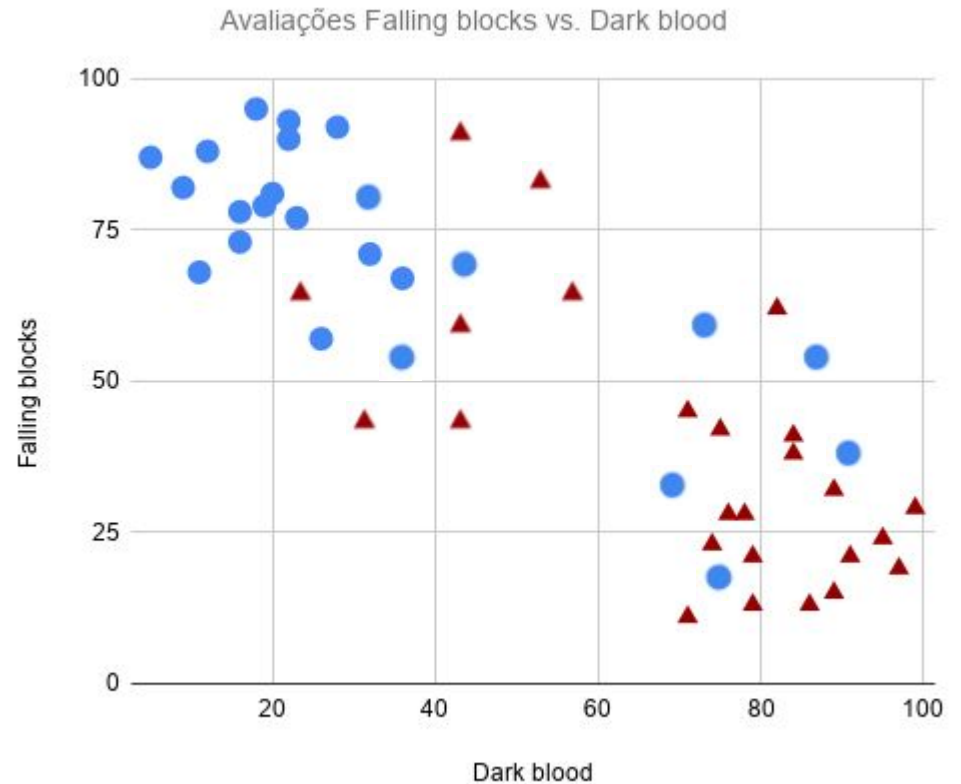
Veja os dados plotados em um gráfico de dispersão.



# Análise de Dados

Veja os dados plotados em um gráfico de dispersão.

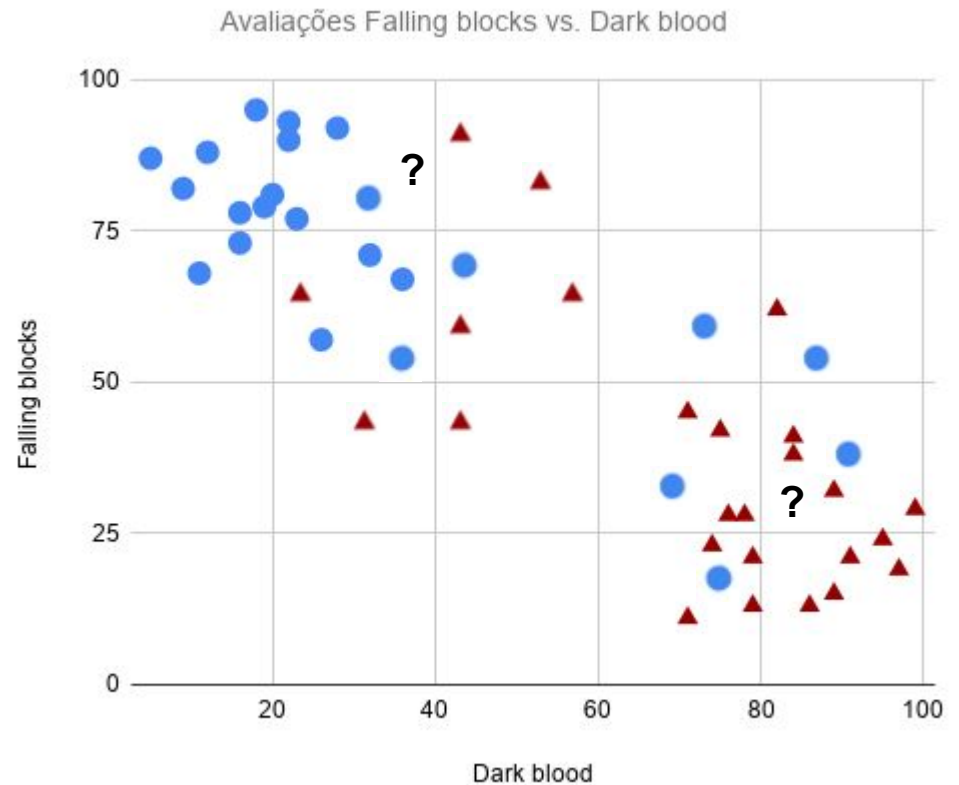
Jogador	Dark blood	Falling blocks	Comprou Dancing Beans
A	74	23	N
B	84	41	N
C	12	88	S
D	32	71	S
E	19	79	S
F	89	32	N
G	99	29	N
H	22	90	S
I	89	15	N
J	76	28	N
K	79	21	N
L	26	57	S
M	11	68	S
N	16	73	S
O	95	24	N
P	28	92	S
Q	82	62	N
R	71	45	N
S	18	95	S



# Análise de Dados

Considere agora que dois novos usuários avaliaram Dark blood e Falling blocks.

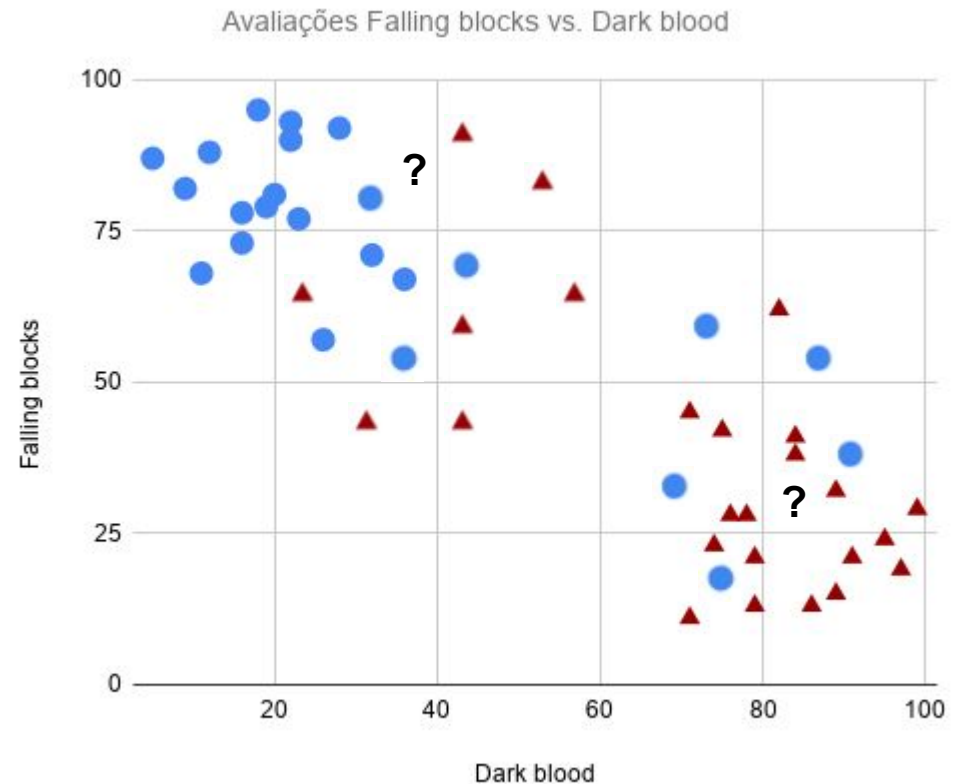
Jogador	Dark blood	Falling blocks	Comprou Dancing Beans
A	74	23	N
B	84	41	N
C	12	88	S
D	32	71	S
E	19	79	S
F	89	32	N
G	99	29	N
H	22	90	S
I	89	15	N
J	76	28	N
K	79	21	N
L	26	57	S
M	11	68	S
N	16	73	S
O	95	24	N
P	28	92	S
Q	82	62	N
R	71	45	N
S	18	95	S
T	38	83	?
U	84	29	?



# Análise de Dados

Considere agora que dois novos usuários avaliaram Dark blood e Falling blocks, mas ainda **não** receberam anúncios de Dancing Beans.

Jogador	Dark blood	Falling blocks	Comprou Dancing Beans
A	74	23	N
B	84	41	N
C	12	88	S
D	32	71	S
E	19	79	S
F	89	32	N
G	99	29	N
H	22	90	S
I	89	15	N
J	76	28	N
K	79	21	N
L	26	57	S
M	11	68	S
N	16	73	S
O	95	24	N
P	28	92	S
Q	82	62	N
R	71	45	N
S	18	95	S
T	38	83	?
U	84	29	?

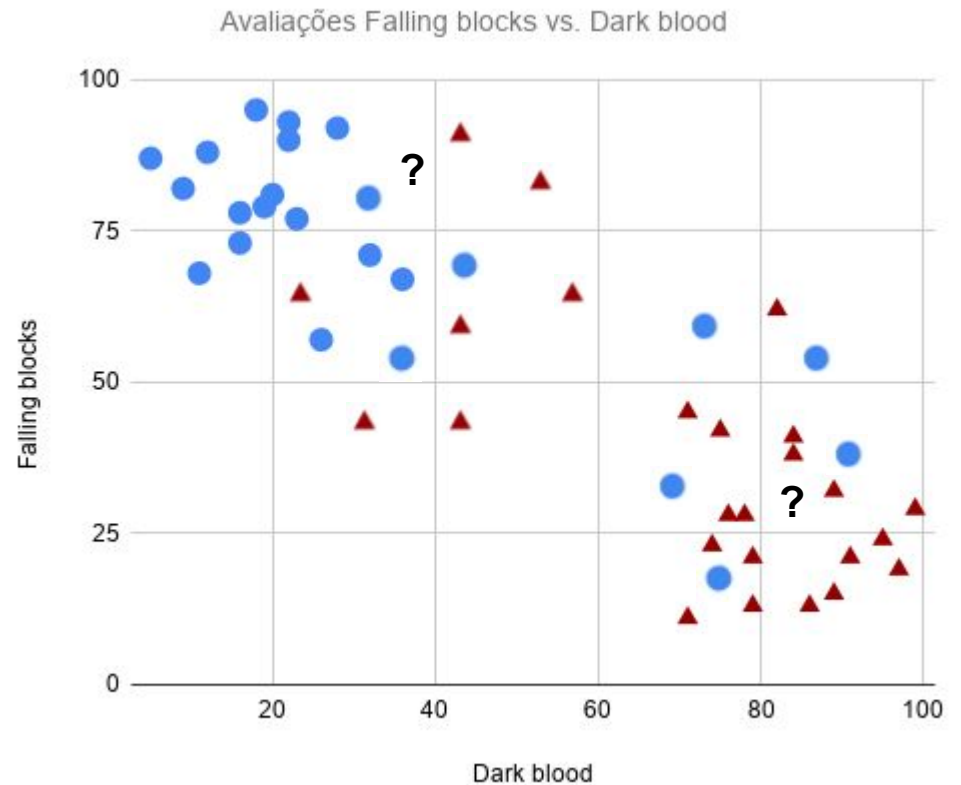


Qual a chance desses novos jogadores também jogarem Dancing Beans?



# Análise de Dados

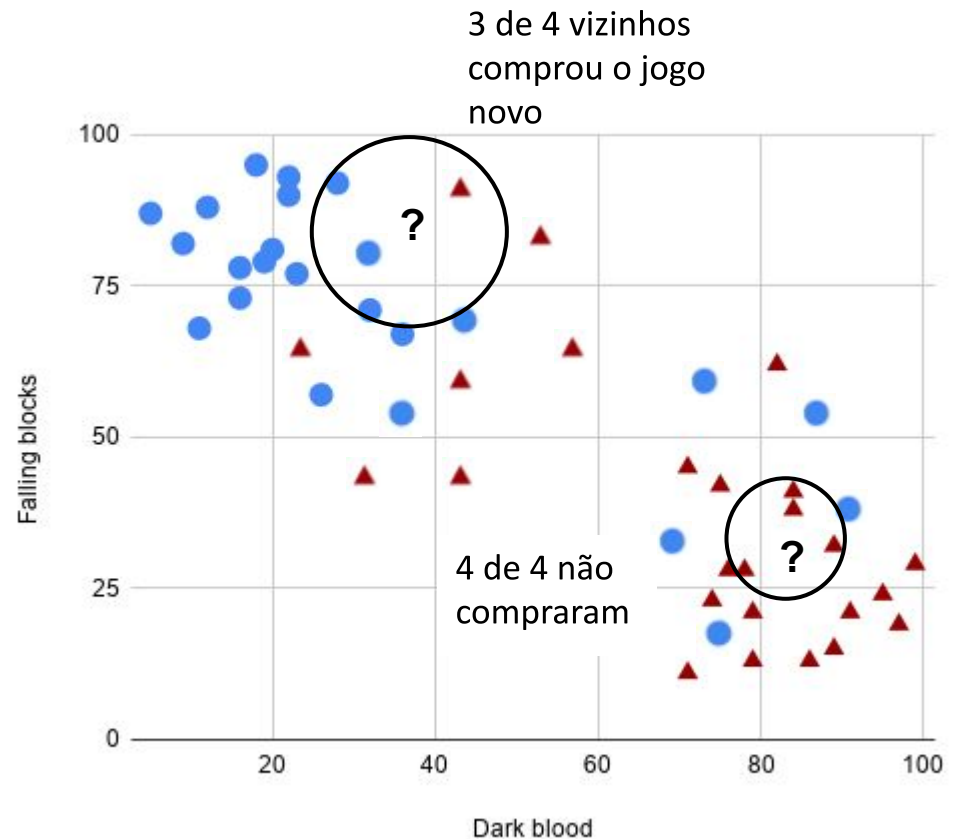
Como prever o comportamento de cada um dos novos jogadores?



# Análise de Dados

Podemos analisar qual o comportamento de seus vizinhos.

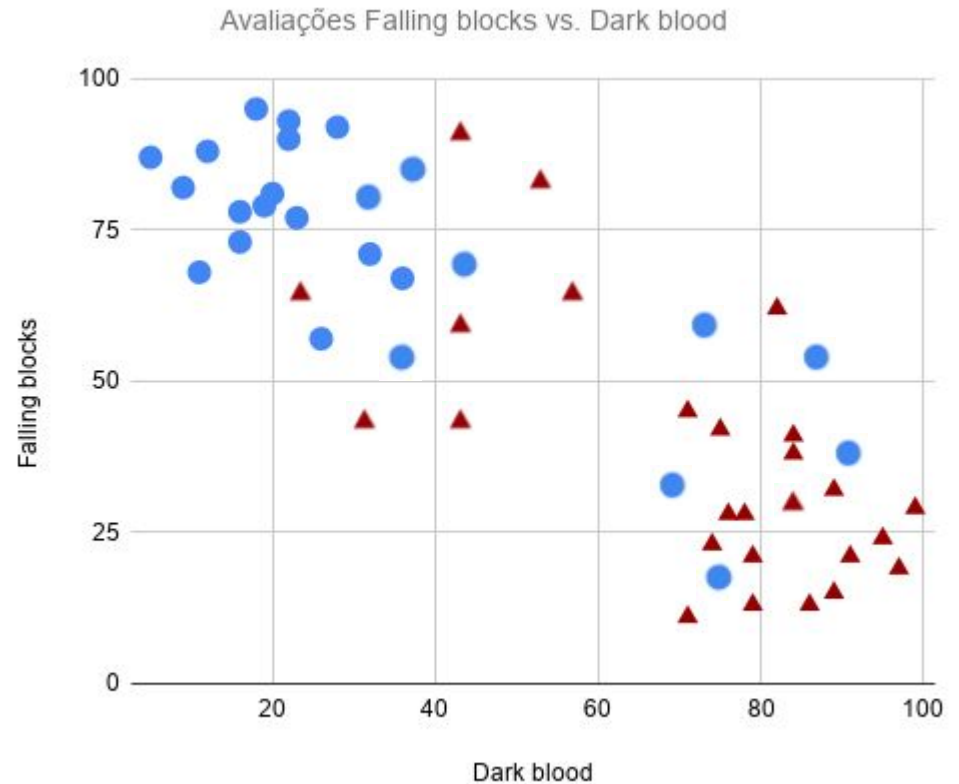
Nesse exemplo vamos ver os 4 vizinhos mais próximos.



# Análise de Dados

Usando a contagem da maioria, podemos inferir qual o comportamento mais provável para os novos jogadores.

Jogador	Dark blood	Falling blocks	Comprou Dancing Beans
A	74	23	N
B	84	41	N
C	12	88	S
D	32	71	S
E	19	79	S
F	89	32	N
G	99	29	N
H	22	90	S
I	89	15	N
J	76	28	N
K	79	21	N
L	26	57	S
M	11	68	S
N	16	73	S
O	95	24	N
P	28	92	S
Q	82	62	N
R	71	45	N
S	18	95	S
T	38	83	N
U	84	29	S



A técnica descrita nos últimos slides é o princípio de um dos mais conhecidos algoritmos de aprendizado de máquina, o KNN (*K Nearest Neighbours*) que significa K vizinhos mais próximos.

Saiba mais sobre o KNN em:

<https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>

# Para saber mais...

<https://sites.usp.br/datascience/spsas-learning-from-data/>

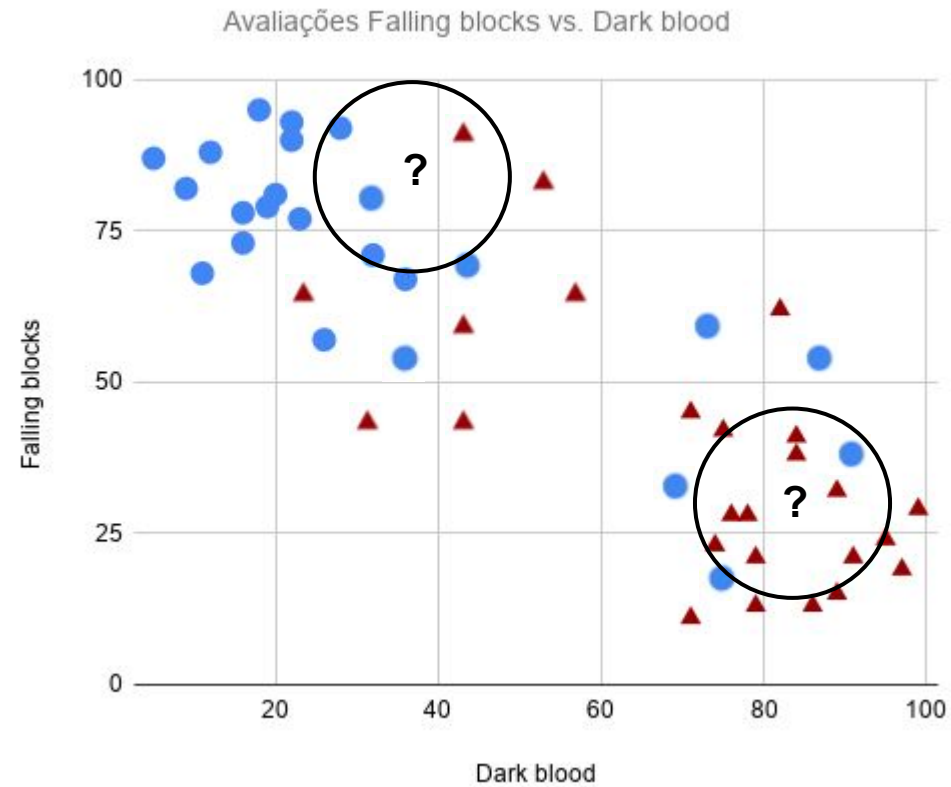
[https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)

<https://www.datageeks.com.br/o-que-e-data-science/>

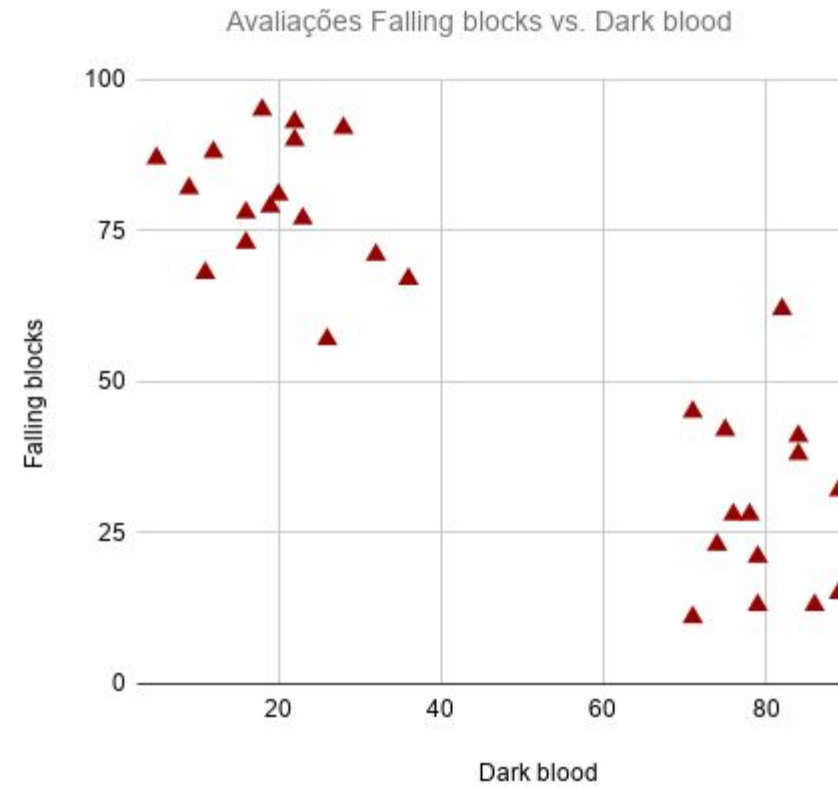
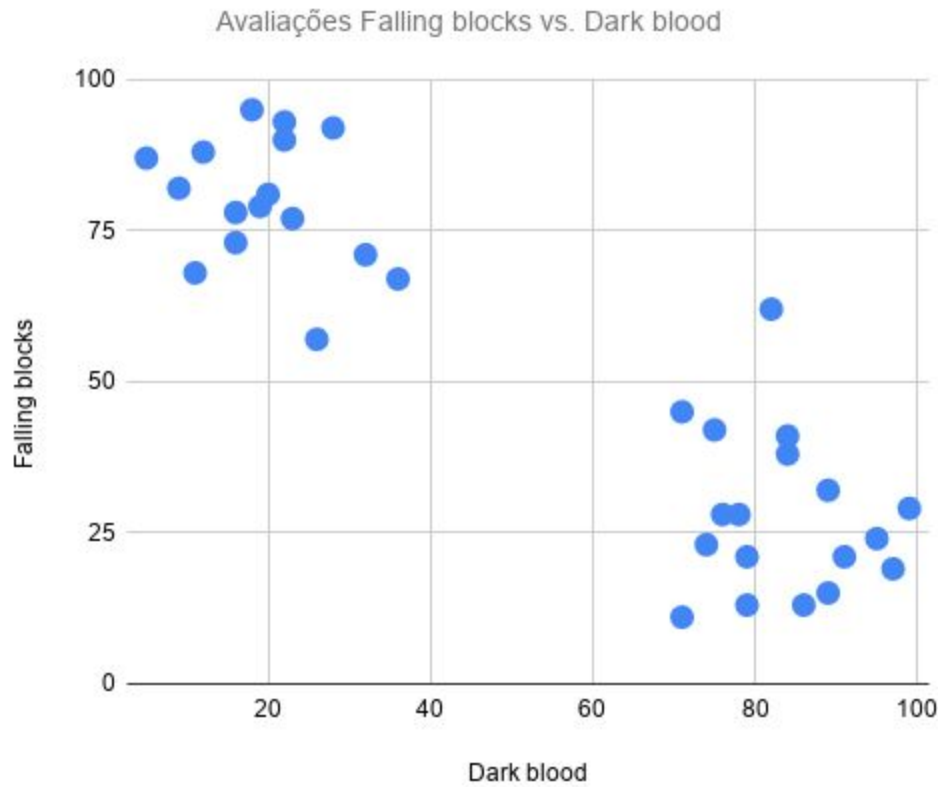
<https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>

# Análise de Dados



# O contexto



<https://dev.mysql.com/doc/refman/8.0/en/mysql-indexes.html>