

### A3 UC Inteligência Artificial

#### Integrantes:

Guilherme Moraes Siqueira - RA 8222240392

Heloisa Soares Ferreira - RA 824152581

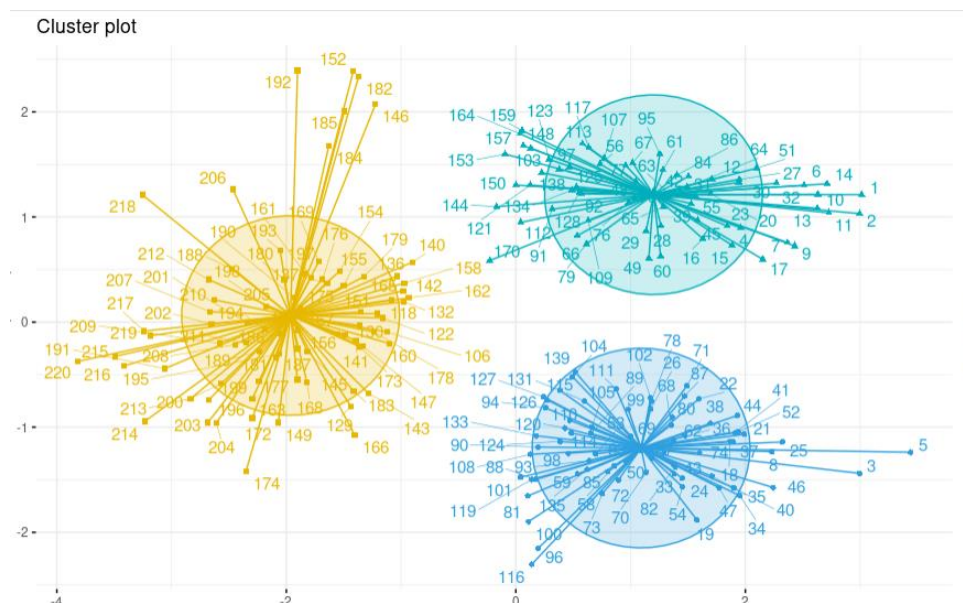
João Vitor Cordeiro Lopes - RA 82429891

Kauan Reis - 824124128

Thiago Amaral da Silva Barros – 822151695

#### a) Existem quantos clusters diferentes para esse exemplo?

Existem 3 clusters.



#### b) Qual a média salarial de cada cluster?

Cluster	Média Salarial
1	135.1667
2	154.0864
3	137.5821

Para descobrir as médias salariais de cada cluster, usamos a função "aggregate", que nos levou aos seguintes números, com base na coluna "salário".

**c) Existe alguma relação entre o sexo e salário desses indivíduos? Análise e explique.**

Primeiro, usamos novamente a função "aggregate", porém com a intenção de observar média salarial para cada valor da coluna "sexo" (0 e 1):

Sexo	Média salarial
0	140.4667
1	144.1103

Com isso, podemos observar que o grupo com o sexo codificado como 1 tem uma média salarial ligeiramente maior do que o grupo com o sexo codificado como 0.

Ademais, voltando as médias salariais por Cluster, parece haver sim uma relação entre o sexo e o salário, embora essa relação possa ser complexa e influenciada por outras variáveis, mesmo que, a análise direta mostre uma pequena vantagem salarial para o sexo codificado como 1 na base de dados geral, a análise por clusters revela nuances interessantes. O Cluster 2, com uma alta proporção do sexo codificado como 1, apresenta a maior média salarial de todos os clusters. O Cluster 3, predominantemente do sexo codificado como 0, tem uma média salarial ligeiramente superior à do Cluster 1, que é predominantemente do sexo codificado como 1.

```
salario  posicao  anosexperiencia  sexo
135.1667  4.083333      9.862500  1.0000000
137.5821  3.925373      6.708955  0.0000000
154.0864  6.888889     14.144444  0.9012346
```

**d) Se analisarmos os anos de experiência na função, é possível afirmar que quanto maior o tempo na função, maior o salário? Análise e explique.**

cluster	salário	anosexperiencia
1	135.1667	9.862500
2	154.0864	14.144444
3	137.5821	6.708955

Comparando a coluna da média salarial com a de anosexperiencia, conseguimos identificar a seguinte tendência:

O Cluster 2 possui a maior média de anos experiência (14.14) e também a maior média salarial (154.09). Isso sugere uma relação positiva entre anos experiência e salário.

O Cluster 3 possui a menor média de anos experiência (6.71) e uma média salarial intermediária (137.58).

O Cluster 1 possui a média de anos experiência intermediária (9.86) e a menor média salarial (135.17)

A análise sugere uma tendência geral de que mais experiência está associada a salários mais altos, especialmente quando comparamos o grupo com mais

experiência (Cluster 2). Entretanto, essa relação não é perfeitamente linear ao compararmos com o restante dos clusters. Dessa forma, decidimos rodar o comando de correlação para termos uma análise mais robusta dos dados.

O resultado foi 0.31, mais próximo de zero. Ou seja, existe uma relação, mas ela é fraca.

```
> # d) Existe correlação entre anos de experiência e salário?  
> cor(EXER_A3$anosexperiencia, EXER_A3$salario)  
[1] 0.3117473
```

**Segue o código que usamos no R e print dos resultados no console:**

```
# Carregando e padronizando os dados
```

```
data(EXER_A3)
```

```
df=scale(EXER_A3)
```

```
head(df, n=3)
```

```
# Instalando e carregando pacotes necessários
```

```
install.packages("factoextra")
```

```
library(factoextra)
```

```
library(ggplot2)
```

```
# Determinando o número ideal de clusters (WSS)
```

```
fviz_nbclust(df, kmeans, method = "wss") +
```

```
geom_vline(xintercept = 3, linetype = 2)
```

```
# a) Existem quantos Clusters diferentes para esse exemplo?
```

```
# Usando 3 clusters conforme o gráfico de WSS
```

```
set.seed(123)
```

```
km.res <- kmeans(df, centers = 3, nstart = 25)
```

```
print(km.res)
```

# b) Qual a média salarial de cada cluster?

```
EXER_A3_clustered <- cbind(EXER_A3, cluster = km.res$cluster)
```

```
aggregate(salario ~ cluster, data = EXER_A3_clustered, mean)
```

# c) Existe alguma relação entre o sexo e o salário desses indivíduos?

```
aggregate(salario ~ sexo, data = EXER_A3, mean)
```

# d) Existe correlação entre anos de experiência e salário?

```
cor(EXER_A3$anosexperiencia, EXER_A3$salario)
```

# Visualizando os clusters em gráfico

```
fviz_cluster(km.res, data = df,
```

```
  palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"),
```

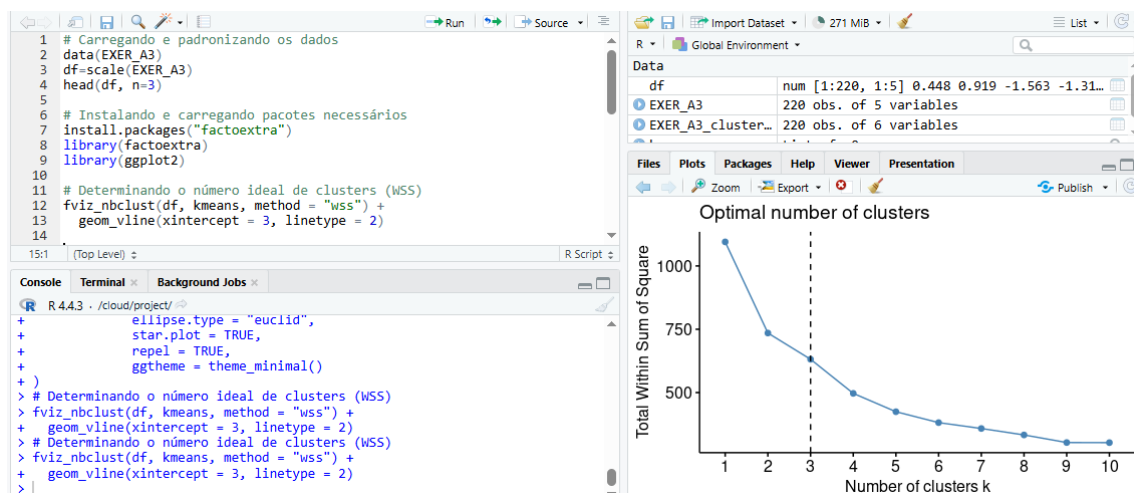
```
  ellipse.type = "euclid",
```

```
  star.plot = TRUE,
```

```
  repel = TRUE,
```

```
  ggtheme = theme_minimal()
```

```
)
```





```

25 # c) Existe alguma relação entre o sexo e o salário desses indivíduos?
26 aggregate(salario ~ sexo, data = EXER_A3, mean)
27
28 # d) Existe correlação entre anos de experiência e salário?
29 cor(EXER_A3$anosexperiencia, EXER_A3$salario)
30
31 # Visualizando os clusters em gráfico

```

28:1 (Top Level) ↕ R Script ↕

onsole Terminal × Background Jobs ×

R 4.4.3 · /cloud/project/ ↗

```

EXER_A3_clustered <- cbind(EXER_A3, cluster = km.res$cluster)
aggregate(salario ~ cluster, data = EXER_A3_clustered, mean)
cluster  salario
 1 135.1667
 2 154.0864
 3 137.5821

# c) Existe alguma relação entre o sexo e o salário desses indivíduos?
aggregate(salario ~ sexo, data = EXER_A3, mean)
sexo  salario
 0 140.4667
 1 144.1103

```

```

28 # d) Existe correlação entre anos de experiência e salário?
29 cor(EXER_A3$anosexperiencia, EXER_A3$salario)
30
31 # Visualizando os clusters em gráfico
32 fviz_cluster(km.res, data = df,
33               palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"),
34               ellipse.type = "euclid",
35               star.plot = TRUE,
36               repel = TRUE,
37               ggtheme = theme_minimal())
38 )

```

31:1 (Top Level) ↕ R Script ↕

Console Terminal × Background Jobs ×

R 4.4.3 · /cloud/project/ ↗

```

1      1 135.1667
2      2 154.0864
3      3 137.5821

> # c) Existe alguma relação entre o sexo e o salário desses indivíduos?
> aggregate(salario ~ sexo, data = EXER_A3, mean)
sexo  salario
 1      0 140.4667
 2      1 144.1103

> # d) Existe correlação entre anos de experiência e salário?
> cor(EXER_A3$anosexperiencia, EXER_A3$salario)
[1] 0.3117473
>

```

Cluster plot

