

# Fecomércio Sesc

Data Science – Princípios e Técnicas

Setembro

2024



Onde me encontrar:

https://www.linkedin.com/in/marco-mialaret-junior/

e

https://github.com/MatmJr









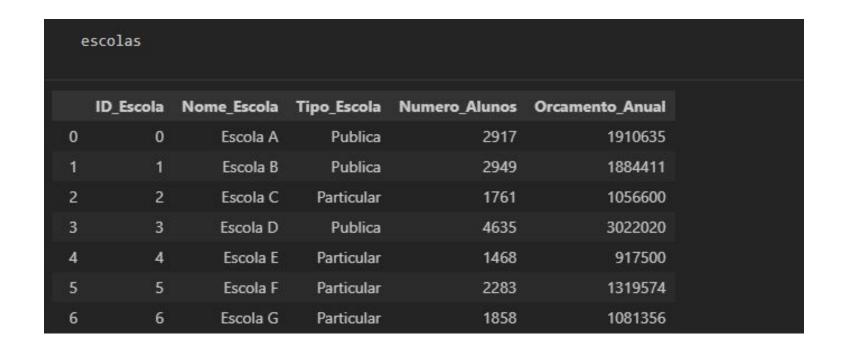
O engenheiro de dados da sua empresa forneceu acesso a dois conjuntos de dados pré-processados. Agora, cabe a você analisá-los cuidadosamente para identificar quais escolas utilizaram seus recursos de forma mais eficiente na preparação de estudantes que se destacaram nas Olimpíadas de Redação e de Matemática.



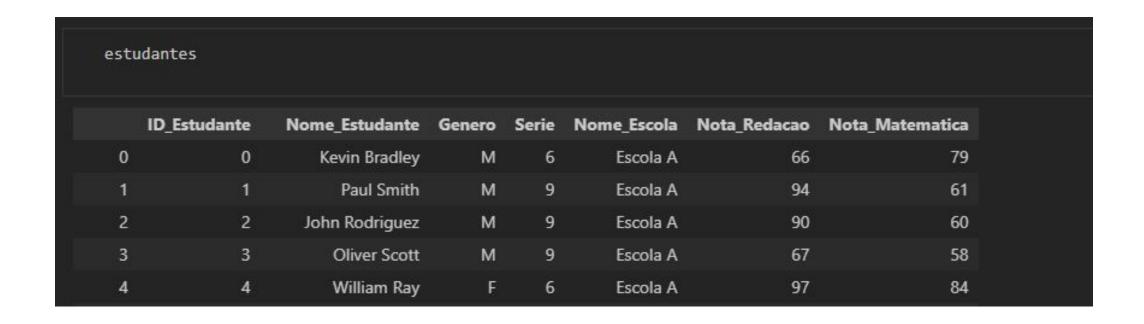
Link do dataset:

https://drive.google.com/uc?id=15aOJIGAyLMSY1gecjiCgu2ko
 rilcKQy











Perguntas:

Existem dados faltantes nas tabelas?

Existe algo em comum nas duas tabelas?



## Verificação de dados faltantes:

```
escolas.isnull().sum()
```

```
escolas.isna().sum()
```

Repetir o processo para escolas.



#### Combinando os datasets:



Entendendo as variáveis categóricas:

```
data["Genero"].unique()

data["Serie"].unique()

data["Tipo_Escola"].unique()
```



## Algumas perguntas:

- Qual o orçamento total das escolas?
- Qual a nota média dos alunos nas disciplinas analisadas?



#### Algumas perguntas:

- Qual o orçamento total das escolas?

```
escolas["Orcamento_Anual"].sum()
```

- Qual a nota média dos alunos nas disciplinas analisadas?

```
mediaRedacao = data["Nota_Redacao"].mean()
mediaMatematica = data["Nota_Matematica"].mean()
```



```
Quantos alunos ficaram com nota superior a 90 em redação?
   highRed = data[data["Nota Redacao"]>90]
   len(highRed)
E qual o percentual?
   len(highRed)/len(estudantes) * 100
```



Quantos alunos ficaram com nota superior a 90 em matemática?

```
highMat = data[data["Nota_Matematica"]>90]
len(highMat)
```

Qual o percentual?

len(highMat)/len(estudantes) \* 100





```
Quantos Alunos tiraram nota maior do que 90 nas duas disciplinas?
```

```
highBoth = data[(data["Nota_Redacao"]>90) & (data["Nota_Matematica"]>90)] len(highBoth)
```

Qual o percentual?

```
len(data[(data["Nota_Redacao"]>90) & (data["Nota_Matematica"]>90)])/len(estudantes) * 100
```





Quantos alunos que obtiveram alto desempenho em ambas as disciplinas são de escolas públicas?

```
highBoth["Tipo_Escola"].value_counts()

v 0.0s

Tipo_Escola
Publica 1321
Particular 1002
```



Name: count, dtype: int64

Quantas alunas obtiveram alto desempenho em ambas as disciplinas?

highBoth["Genero"].value\_counts()

Genero

M 1167

F 1156

Name: count, dtype: int64



```
Como ficou a distribuição dos alunos de alto desempenho pelas série?
   highBoth["Serie"].value counts()

√ 0.0s

Serie
6
      663
     604
     599
     457
Name: count, dtype: int64
```



```
Qual o total por escolas?

highBoth["Nome_Escola"].value_counts()

v 0.0s
```





Qual o total por escolas?

```
# Agrupando por 'Nome_Escola' e 'Tipo_Escola', e contando o número de registros em cada grupo
sumario = highBoth.groupby(['Nome_Escola', 'Tipo_Escola']).size().sort_values(ascending=False)
# Exibindo o sumário
print(sumario)
```

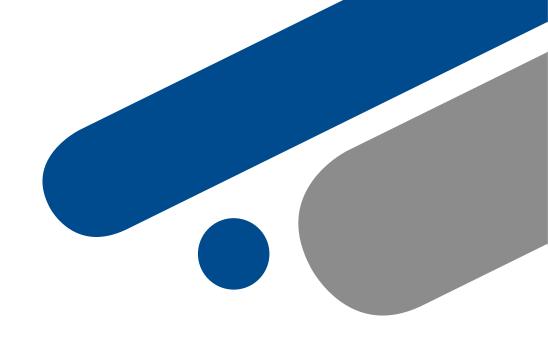


Qual o orçamento per capita de cada escola?

```
perCapita = escolas["Orcamento_Anual"]/escolas["Numero_Alunos"]
escolas["Per_Capita"] = perCapita
```

√ 0.0s









Quais escolas demonstraram maior eficiência?

#### Instruções:

1. Antes de analisar os dados, explique o que significa "eficiência" no contexto de escolas. Considere aspectos como o uso de recursos, desempenho acadêmico, custo-benefício, entre outros critérios relevantes.



- 2. Utilize os dados fornecidos neste documento para identificar quais escolas se destacam em termos de eficiência, de acordo com a definição estabelecida.
- 3. Quais diferentes formas de análise, novos resultados ou novas métricas podem ajudar a avaliar melhor a eficiência das escolas?



## **Dúvidas?**





#### **Marco Mialaret, MSc**

**Telefone:** 

81 98160 7018

E-mail:

marcomialaret@gmail.com

