**Análise de Dados**

**UFCD 10811 Projeto Final**

**Formador António Francisco da Costa Machado**

Projeto Final

Ana Fernandes

Anaïs Gilbert

Berlinda Caiongo

Faro, 8 de fevereiro de 2024

**Índice**

# Introdução

**Análise de dados cinematográficos**

**Objectivo:**

* Explorar o conjunto de dados cinematográficos e efetuar análise exploratória
* Não existe resultado correto.
* Pretende-se que sejam criativos e analíticos
* Entrega de um relatório simples que demonstre algum trabalho e capacidade de análise.
* Poderá ser feito em grupo de dois no máximo.
* As Tarefas podem ser possíveis ou não. Se não for possível, diga porquê.
* Uso do jupiter
* Tempo integral durante as aulas e ajuda

**Dados:**

* [**https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset**](https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset)

# Limpeza de dados

* + Tratar valores
  + Limpar duplicados
  + Certifique-se de que os tipos de dados sejam apropriados para cada coluna.

# Estatística Descritiva

* + Calcular média, mediana, desvio padrão por tipo de filme por exemplo.
  + Explore a distribuição de variáveis, especialmente as numéricas, como receita de bilheteria e avaliações.

# Visualização de dados

* + Crie visualizações para representar a distribuição de variáveis chave recorrendo a histogramas, gráficos de caixa ou outros gráficos relevantes.
  + Visualize a distribuição dos gêneros cinematográficos.

# Análise temporal

* + Evolução do mercado cinematográfico ao longo do tempo.

# Análise de receitas vs orçamento

* + Analise a relação entre orçamento e receita.

* + - Quais os filmes mais rentáveis.

# Notas Guia Temporário

1. Quantos filmes foram feitos em x ano, por exemplo por tipo/genero na analise temporal
2. Ate estatística descritiva e visualização de dados ja esta ok. Mais especifico analise temporal e analise de relacionamentos.
3. \*nota: ver as colunas no excel pois os dados vão ter que ser trabalhados.
4. \*Nota2: cada linha é um filme.
5. Nota3\*: ha varios que parece não terem budget. Confirmar
6. Exemplo harry potter- Filmes que pertencem a uma coleção (dados a tratar)\* budget = custou a fazer, ao se filtrar pelo harry potter mostra cada valor gasto em cada filme. Coluna revenue = quanto gerou de lucros. Para saber o total realmente ganho soma budget de cada coleção do filme - soma de cada receita=revenue de cada filme
7. Normalmente mostra-se uma linha para mostrar a respetiva variação no exemplo do harry Potter
8. Estat descritiva: quantos tem por tipo, qual a media,\* o codigo python é importante, \*nota mental: confirmar se na estat desc estaria por exemplo histogramas com as categorias de filmes feitos
9. Estat descr continuacao: qual e o orcamento medio de x genero; nota\*excel tabela dinâmica= pivot table vai ajudar a visualizar os dados e filtrar; \*nota mental: verificar quantos generos existem deve ser incluindo na estat descritiva. ;
10. Media, mediana, desvio padrao na estat desc
11. Tentar prever pode dar erros porque cada filme tem poucos dados, evitar

# Conclusões e previsões

* + Tente encontrar um modelo de ML que possa efetuar previsões.
    - Previsão de receitas de um filme Drama, Thriller....

# Conclusão

**Bibliografia**

**Webgrafia**