

Aluno(a)			RA:
Curso	Ciência da Computação	Ano:	
Disciplina	Ciência de Dados		
Professor	Prof. Eduardo Pena		
Nota			
Manipulação de Dados com Pandas			
<p>Orientações gerais:</p> <p>1 - Todos os materiais (códigos fontes, documentos, diagramas, etc) deverão ser entregues em um arquivo .zip identificado (Nome/RA/Ano).</p> <p>2 - A interpretação das questões é parte do processo de avaliação.</p>			

Nesta atividade, você será um cientista de dados iniciante e deverá analisar um conjunto de dados sobre filmes utilizando pandas. O objetivo é praticar manipulação de dados para responder a perguntas sobre desempenho, gêneros, diretores e características das produções.

Conjunto de Tarefas

1. Carregamento e Inspeção

- 1.1. Carregue o arquivo `imdb.csv` em um `DataFrame`.
- 1.2. Exiba as 5 primeiras linhas e use `.info()` e `.describe()` para explorar o dataset.
- 1.3. Identifique colunas com valores ausentes.

2. Limpeza e Transformação

- 2.1. Converta colunas como `year`, `duration`, `rating_imdb`, `vote` para tipos numéricos apropriados.
- 2.2. Crie uma nova coluna `profit = gross_world_wide - budget`.
- 2.3. Padronize nomes de colunas (sem espaços e com minúsculas).

3. Exploração e Agregação

- 3.1. Liste os 5 filmes com maior nota IMDb.
- 3.2. Calcule a média de duração por década.
- 3.3. Encontre o diretor com maior média de nota IMDb (mínimo 3 filmes).
- 3.4. Mostre o gênero mais recorrente no dataset.

4. Joins e Relações

- 4.1. Carregue o arquivo `oscars.csv` em um outro `DataFrame`.
- 4.2. Realize o merge entre o `DataFrame` principal e o `DataFrame` sobre `oscars`, utilizando as colunas apropriadas para identificar os filmes.
- 4.3. Considerando apenas os filmes vencedores da categoria *BEST PICTURE*, identifique o **diretor** com o maior número de prêmios nessa categoria.

5. Pivot e Reshaping

- 5.1. Crie uma `pivot_table` mostrando a média de `rating_imdb` por país e década.
- 5.2. Use `melt` para reorganizar colunas de bilheteria (`gross_us_canada`, `gross_world_wide`, etc.) em formato longo.

6. Exportação

- 6.1. Exporte a **tabela de pivot** (média de `rating_imdb` por país e década) em CSV com o nome: `rating_pivot_pais_decada.csv`.
 - Não incluir índice no arquivo (`index=False`).
 - Codificação: `utf-8` (ou `utf-8-sig` se for abrir no Excel).
 - Separador: vírgula.
 - Se houver valores ausentes, manter como vazio (sem escrever `'NaN'`).

Entrega e Instruções

Os alunos deverão entregar **dois notebooks (arquivos .ipynb)** desta atividade:

1. **Versão Manual:** feita exclusivamente com base na documentação oficial e consultas a fóruns, sem auxílio de ferramentas de IA.
 - Documentação Pandas: <https://pandas.pydata.org/docs/>
 - Documentação NumPy: <https://numpy.org/doc/>
 - Stack Overflow: <https://stackoverflow.com/>
2. **Versão com IA:** código gerado com o auxílio de uma ferramenta de Inteligência Artificial. Cronometrar o tempo de resposta da IA por etapas (Carregamento, Limpeza, etc).

Muito importante: Nesta etapa, não utilizem IA na primeira versão. A ideia é comparar seu próprio processo de raciocínio com a solução da IA. Em outras etapas da disciplina vamos poder utilizar IA.

Cada estudante deverá montar uma tabela comparativa com os seguintes campos:

- Tempo gasto na versão manual (cronometrem e anotem).
- Tempo gasto na versão IA.
- Para cada item da lista de tarefas: indicar se houve diferença significativa nas respostas.
- Indicar se a IA cometeu erro ou se o estudante cometeu erro.

Faça a cronometragem por etapa (Carregamento, Limpeza, etc).

Tabela de Comparação Manual x IA

Tarefa	Tempo Manual (min)	Tempo IA (seg)	Diferença?	Quem Errou	Observações
Carregamento e inspeção	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]
Limpeza e transformação	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]
Exploração e agregação	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]
Joins e relações	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]
Pivot e reshaping	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]
Exportação	[número]	[número]	(Sim / Não)	(Aluno / IA / Ninguém)	[texto curto]

Cada aluno deverá entregar:

- `atividade_manual_nomeAluno.ipynb` – notebook feito pelo aluno, sem uso de IA.
- `atividade_IA.ipynb` – notebook feito pela IA.
- `comparacao.csv` – tabela preenchida com os dados de comparação.