

Trabalho final

Data de entrega: 16/06

Objetivo

Aplicar os conhecimentos adquiridos ao longo da disciplina por meio da análise de dois conjuntos de dados distintos, utilizando ferramentas de programação, visualização e interpretação estatística.

Instruções Gerais

- A análise de **cada banco de dados deve ser realizada em notebooks distintos**.
- Utilize células em **Markdown** para descrever cada etapa da análise de forma clara.
- Comente o código e utilize **docstrings** sempre que necessário para explicitar o raciocínio adotado.
- Nomeie os arquivos no formato:
`TP_<NúmeroDoBanco>.ipynb`
- É permitido utilizar uma amostra do banco de dados para otimizar o tempo de execução, **desde que o critério de amostragem seja claramente documentado**.

1ª Análise: Séries Temporais de Vendas

Fonte de dados: `Sales_Transactions_Dataset_Weekly.csv`

Esse conjunto de dados contém o valor de transações de vendas semanais de diversos produtos, ao longo de 52 semanas.

Descrição das colunas:

- `Product_Code`: código identificador do produto.
- `W[0-51]`: valor da transação correspondente a cada semana

Etapas a serem realizadas:

1. **Transformação de formato:**
Converta o dataframe do formato *wide* (colunas W0 a W51) para o formato *long*.

2. Conversão de semana para dia da semana:

Utilize a função disponível no arquivo `week_day.py` para associar cada semana a um dia da semana.

Formato do arquivo:

| Código do produto | W0 | W1 | W2 |
|-------------------|----|----|----|
| PROD01 | 1 | | |
| PROD02 | | | |

Formato novo:

| Código do produto | Variable | Value |
|-------------------|----------|-------|
| PROD01 | W0 | 1 |
| PROD01 | W1 | |
| PROD01 | W2 | |
| PROD02 | W0 | |
| PROD02 | W1 | |
| PROD02 | W2 | |

3. Seleção de produtos:

Escolha três produtos distintos para análise.

4. Visualização temporal:

Construa um gráfico com a evolução do valor das transações ao longo do tempo para cada produto selecionado.

5. Análise de diferença entre semanas:

Para cada produto, calcule a diferença do valor de transações entre semanas consecutivas e gere um gráfico da série de diferenças.

6. Estatísticas descritivas:

Calcule para cada produto:

- Média
- Desvio-padrão
- Mínimo

- Mediana
- Máximo

7. **Média móvel [UM PRODUTO]:**

Calcule a média móvel de 3 semanas e apresente a série suavizada para ao menos um dos produtos.

2ª Análise: Receita por Visitas

Fontes de dados:

- **visitas.csv**: contém os registros de visitas ao site.
 - **id_visita**: identificador único da visita.
 - **id_cliente**: identificador do cliente.
- **receita.csv**: contém a receita gerada nas visitas.
 - **id_visita**: identificador da visita.
 - **receita**: valor gerado na visita (pode ser zero se não houve compra).

Etapas a serem realizadas:

1. **Junção de dados:**

Unifique os dois dataframes pela coluna **id_visita**. Preencha os valores ausentes de receita com zero.

2. **Estatísticas por cliente (**id_cliente**):**

| id_cliente | id_visita | receita |
|------------|-----------|---------|
| 1 | 0 | 100 |
| 1 | 1 | 50 |
| 1 | 2 | 0 |

- Média considerando todas as visitas (incluindo as que não geraram receita). Exemplo: 50
- Média considerando apenas as visitas com receita maior que zero. Exemplo: 75
- Percentual de visitas com receita (visitas com receita > 0 dividido pelo total de visitas). Exemplo: 0.67

3. Simulação de amostragem bootstrap:

Desenvolva um algoritmo que execute os seguintes passos:

Para **B** variando de 1 a 1000:

- Realize uma amostragem com reposição do dataframe original (use `df.sample(n=tamanho, replace=True)`).
- Calcule a **média da receita** da amostra (incluindo visitas com receita zero).
- Armazene o valor da média e o número da simulação em listas.

Ao final, construa um DataFrame com as colunas:

- `id_simulacao`: identificador da simulação (de 0 a 999)
- `media_amostrada`: média da receita em cada simulação

Exemplo:

| id_simulacao | media_amostrada |
|--------------|-----------------|
| 0 | 10 |
| 1 | 9.7 |

4. Visualização:

Gere um gráfico (ex: histograma ou densidade) com a **distribuição dos valores da coluna `media_amostrada`**.