

FIAP
ENGENHARIA DE SOFTWARE

LUCCA SABATINI TAMBELLINI - RM98169

DATA SCIENCE & STATISTICAL COMPUTING
GLOBAL SOLUTION
ECONOMIA AZUL

SÃO PAULO

2024

1. Descrição do Projeto e das Variáveis

O projeto “Water Quality - Every Drop Matters” utiliza um conjunto de dados que contém medições de seis parâmetros críticos amplamente utilizados na monitorização e análise da qualidade da água. Neste trabalho, exploraremos esses dados, realizaremos análises descritivas, criaremos gráficos e faremos inferências sobre a qualidade da água em diferentes contextos geográficos.

Neste conjunto de dados, temos as seguintes variáveis:

- City: Nome da cidade.
- Country: País onde a cidade está localizada.
- AirQuality: Qualidade do ar (valor numérico).
- WaterPollution: Poluição da água (valor numérico).

Para cada variável, vamos verificar:

- Tipo de Dados:

Se são categóricos (nominais) ou numéricos (contínuos ou discretos).

- Valores Nulos:

Quantidade de valores ausentes em cada coluna.

- Distribuição dos Atributos:

Como os valores estão distribuídos (média, desvio padrão, mínimo, máximo).

(Está no arquivo ipynb)

2. Estatísticas Descritivas

Vamos calcular métricas básicas para cada atributo:

- Média: Valor médio da variável.
- Mediana: Valor central da distribuição.
- Desvio Padrão: Medida de dispersão dos dados.
- Mínimo e Máximo: Valores extremos.
- Usado o .describe em países, cidades e regiões

(Está no arquivo ipynb)

3. Gráficos

- Comparação de AirQuality e WaterPollution em 5 Cidades Aleatórias.
- Comparação de AirQuality e WaterPollution por Países.
- Cidades com Maior e Menor Qualidade do Ar e Poluição da Água.
- Países com Maior e Menor Qualidade do Ar e Poluição da Água.
- Média da qualidade de Ar e Água entre todos os Países

(Está no arquivo ipynb)

4. Descrição das Análises Realizadas

1. Sorteio de 5 cidades aleatórias e comparação de AirQuality e WaterPollution:

- Primeiro, o código sorteia aleatoriamente 5 cidades a partir dos dados.
- Filtra os dados para incluir apenas as cidades selecionadas.
- O gráfico de barras é criado e compara os níveis de qualidade do ar AirQuality nessas cidades.
- Outro gráfico de barras é criado para comparar os níveis WaterPollution nessas mesmas cidades.

2. Países específicos e cálculo das médias de AirQuality e WaterPollution:

- O código filtra os dados para incluir apenas os países "Brazil", "United States of America", "Japan", "Russia".
- Em seguida, calcula as médias de AirQuality e WaterPollution para cada um desses países.
- O 1º gráfico de barras mostra as médias de qualidade do ar por país.
- O 2º gráfico de barras mostra as médias de poluição da água por país.

3. Comparação conjunta de AirQuality e WaterPollution para os quatro países:

- O código cria um gráfico de barras combinado para mostrar as médias de qualidade do ar e poluição da água para os quatro países selecionados.
- Os dados são representados lado a lado para facilitar a comparação entre os países.

4. Cidades com melhor qualidade do ar e água:

- O código seleciona as 5 cidades com a melhor qualidade do ar (com base no índice "AirQuality") e as 5 cidades com a melhor qualidade da água (com base no índice "WaterPollution").
- Ele imprime as listas dessas cidades.
- Também calcula a correlação entre a qualidade do ar e a poluição da água usando o coeficiente de correlação de Pearson.

(O coeficiente de correlação de Pearson é uma medida estatística que quantifica a força e a direção da relação linear entre duas variáveis numéricas. Seu valor varia de -1 a 1, onde 1 indica uma correlação positiva perfeita, -1 uma correlação negativa perfeita e 0 nenhuma correlação linear. Ele é amplamente utilizado em análises para identificar e quantificar relações entre conjuntos de dados.)

5. Países com melhor qualidade do ar e água:

- Similar ao primeiro trecho, mas agora considerando os países em vez das cidades.
- Ele seleciona os 5 países com a melhor qualidade do ar e os 5 países com a melhor qualidade da água.
- Imprime as listas desses países.
- Calcula a correlação entre a qualidade do ar e a poluição da água para esses países.

6. Média de qualidade do ar e da água para todos os países:

- Calcula a média de qualidade do ar e da água para todos os países.
- Exibe essas médias em formato de texto.

- Cria um gráfico de barras para comparar a média de qualidade do ar e poluição da água para todos os países.

5. Inferências e Testes Estatísticos

1. Inferências sobre as cidades com melhor qualidade do ar e água:

- As cidades com melhor qualidade do ar e da água foram selecionadas com base nos índices "AirQuality" e "WaterPollution".

- Essas cidades podem ser consideradas como referência para comparação com outras cidades.

- No entanto, é importante lembrar que essas seleções são baseadas em dados específicos e podem não representar todas as cidades.

2. Correlação entre qualidade do ar e poluição da água:

- A correlação entre a qualidade do ar e a poluição da água foi calculada usando o coeficiente de correlação de Pearson.

- O valor da correlação (correlation1) indica a direção e a força da relação entre essas duas variáveis.

- Se a correlação for positiva, significa que quando a qualidade do ar melhora, a poluição da água também tende a melhorar (ou vice-versa).

- Se a correlação for negativa, significa que essas duas variáveis têm uma relação inversa.

3. Inferências sobre os países com melhor qualidade do ar e água:

- Os países com melhor qualidade do ar e da água foram selecionados com base nos índices "AirQuality" e "WaterPollution".

- Esses países podem ser considerados como referência para comparação com outros países.

- Assim como nas cidades, essas seleções são específicas e podem não representar todos os países.

4. Correlação entre qualidade do ar e poluição da água para todos os países:

- A correlação entre a qualidade do ar e a poluição da água foi novamente calculada (correlation2) para todos os países.

- Esse valor global pode ajudar a entender se existe uma tendência geral entre essas duas variáveis em todos os países.

5. Média de qualidade do ar e da água para todos os países:

- As médias de qualidade do ar e poluição da água foram calculadas para todos os países.

- O gráfico de barras mostra essas médias lado a lado.

- A barra "Qualidade do Ar" representa a média de qualidade do ar para todos os países.

- A barra "Poluição da Água" representa a média de poluição da água para todos os países.

Conclusão

Essas análises nos ajudam a entender as relações entre qualidade do ar e poluição da água em diferentes contextos geográficos. No entanto, para tirar conclusões mais robustas, seria necessário realizar testes estatísticos específicos, como testes de hipóteses ou análises de variância.

Este projeto nos permite entender melhor a qualidade da água em diferentes regiões e contribui para a gestão ambiental. Os resultados obtidos podem ser aplicados em nosso projeto com parceria com a FIAP e ações para preservar nossos recursos hídricos.