

DATA SCIENCE

Kaique Maia Reis Silva

Descrição do Projeto e das Variáveis

O projeto foca em analisar a qualidade do ar e a poluição da água em diferentes cidades/países ao redor do mundo. Utilizamos um banco de dados contendo informações sobre cidades, regiões, países, qualidade do ar e poluição da água. Através de visualizações e análises estatísticas, buscamos identificar padrões e insights que possam ajudar em estudos ambientais.

Descrição das Variáveis:

City

- Tipo: String
- Descrição: Nome da cidade onde os dados foram coletados.
- Número de valores nulos: 0.

Region

- Tipo: String
- Descrição: Nome do país onde os dados foram coletados.
- Número de valores nulos: 425.

Country

- Tipo: String
- Descrição: Nome do país onde os dados foram coletados.
- Número de valores nulos: 0.

AirQuality

- Tipo: Numérico (float ou int)

- Descrição: Índice de qualidade do ar. Valores mais altos indicam melhor qualidade do ar.
- Número de valores nulos: 0.

WaterPollution

- Tipo: Numérico (float ou int)
- Descrição: Índice de poluição da água na cidade. Valores mais baixos indicam melhor qualidade da água.
- Número de valores nulos: 0.

Estatísticas Descritivas:

	<i>AirQuality</i>	<i>WaterPollution</i>
count	3963.000000	3963.000000
mean	62.253452	44.635372
std	30.944753	25.663910
min	0.000000	0.000000
25%	37.686567	25.000000
50%	69.444444	50.000000
75%	87.500000	57.719393
max	100.000000	100.000000

Média (mean):

- Qualidade do Ar: A média é 62.25, indicando que, em geral, a qualidade do ar é moderada
- Poluição da Água: A média é 44.64, indicando que a poluição da água é moderada.

Desvio Padrão (std):

- Qualidade do Ar: O desvio padrão é 30.94, mostrando uma alta variabilidade na qualidade do ar entre diferentes cidades.
- Poluição da Água: O desvio padrão é 25.66, indicando também uma alta variabilidade na poluição da água.

Quartis (25%, 50%, 75%):

- Qualidade do Ar: O 25% (primeiro quartil) é 37.69, a mediana (50%) é 69.44, e o 75% é 87.50. Isso mostra que metade das cidades tem uma qualidade do ar melhor que 69.44.
- Poluição da Água: O 25% é 25.00, a mediana é 50.00, e o 75% é 57.72. Isso indica que metade das cidades tem poluição da água menor que 50.

Valores Extremos (min e max):

- Mínimos: Ambas as variáveis têm valores mínimos de 0, indicando a presença de cidades com muito boa qualidade do ar e água.
- Máximos: Ambas as variáveis têm valores máximos de 100, indicando a presença de cidades com condições extremas de poluição.

Descrição das Análises Realizadas:

Gráfico de distribuição de Qualidade do Ar:

Em média, quanto maior a qualidade do ar, mais frequente é o mesmo.

Gráfico da distribuição da poluição da água:

Os níveis 0%, 25%, 50%, 75% e 100% de poluição de água têm picos na frequência, sendo 50% disparadamente o maior. É de se destacar que esses picos vêm exatamente a cada 25%.

Gráfico de Qualidade do Ar por Região (Top 15):

São Paulo é isoladamente a região com a pior qualidade do ar (em média). Excluindo SP, nenhuma região fica com média próxima ou abaixo de 60% na qualidade do ar. Nova York, Califórnia, Inglaterra, Texas e Michigan têm uma margem grande de diferencial (min. e max.) na qualidade do ar.

Gráfico de Poluição da Água por Região (Top 15):

Washington e Inglaterra têm a menor taxa de poluição da água. Illinois, Columbia, Washington, Inglaterra e São Paulo têm uma margem grande de diferencial (min. e max.) na poluição da água. Em média, tirando Inglaterra e Washington, as 13 regiões têm taxas parecidas, entre 20% e 55% de poluição na água.

Gráfico de comparação entre Qualidade do Ar e Poluição da Água por País (top 10):

Índia e Rússia são países preocupantes, se concentrando entre 50-80% na Poluição da Água e 15-55% na Qualidade do Ar.

Países/regiões como o Canadá, Reino Unido e Alemanha, estão concentrados entre 50-90% na Qualidade do Ar e 15-50% na Poluição da Água.

O país com as melhores taxas é a Espanha, concentrada entre 65-100% na Qualidade do Ar e 20-55% na Poluição da Água.

A China é disparadamente o país com as piores taxas, entre 35-90% de Poluição da Água e míseros 5-55% de Qualidade no Ar.

O Brasil se concentra entre 60-90% na Qualidade do Ar e 15-65% de Poluição na Água.

País mais presente no gráfico, o EUA, se concentram entre 40-100% de Qualidade no Ar e 0-85% de Poluição na Água.

A Polônia se concentra entre 20-50% de Qualidade no Ar e 25-60% na Poluição da Água.

As conclusões finais levando em consideração todos os insights discutidos, são:

A grande maioria dos conglomerados tem piores níveis na qualidade do ar e água, por isso países como Índia e China compartilham de dados similarmente ruins nos 2. Percebe-se que isso se agrava ainda mais em conglomerados pobres. Existe uma correlação negativa entre qualidade do ar e poluição da água de -0.45. Quando a qualidade do ar melhora, a poluição da água tende a piorar, e vice-versa.