Data Science Academy

Plano de Análise de Dados – Capítulo 9 – SQL Para Science

Etapa 1 - Definição do Objetivo

O objetivo deste trabalho é extrair informações sobre os índices de poluição do ar em localidades ao redor do mundo. As 5 perguntas abaixo deverão ser respondidas:

**# 1- Quais poluentes foram considerados na pesquisa?**

**# 2- Qual foi a média de poluição ao longo do tempo provocada pelo poluente ground-level ozone (o3)?**

**# 3- Qual foi a média de poluição causada pelo poluente ground-level ozone (o3) por país medida em µg/m³ (microgramas por metro cúbico)?**

**# 4- Considerando o resultado anterior, qual país teve maior índice de poluição geral por o3, Itália (IT) ou Espanha (ES)? Por quê?**

**# 5- Quais localidades e países tiveram média de poluição em 2020 superior a 100 µg/m³ para o poluente fine particulate matter (pm25)?**

Etapa 2 - Coleta e Extração de Dados

Os dados estão disponíveis em formato CSV e foram extraídos da fonte conforme indicado no manual em pdf no Capítulo 9. Os dados foram carregados via linha de comando no SGBD MySQL.

Etapa 3 - Escolha das Ferramentas

Para este trabalho usaremos Linguagem SQL uma vez que os dados estão estruturados e disponíveis em um banco de dados relacional.

Etapa 4 - Aplicação das Técnicas de Análise

Para responder às perguntas de negócio usaremos função window (quando for necessário cálculo ao longo do tempo) e calcularemos média e desvio padrão para explicar a variabilidade dos dados no item 4.

Etapa 5 - Entrega dos Resultados

Conclusão:

A Itália foi o país que apresentou o maior índice geral de poluição pelo poluente o3. Embora a Espanha tenha maior dispersão dos dados, a Itália tem uma média mais alta e a maioria dos pontos de dados em torno da média.

Entre as localidades com maior média de poluição pelo poluente pm25, duas ficam na Índia.