1.1. Tập dữ liệu về không khí

1.1.1. Tổng quan về tập dữ liệu

* Thời gian: 01/01/2015 đến 01/07/2020
* Địa điểm: Thành phố Delhi - Ấn Độ
* Phương pháp thu thập tập dữ liệu
* **Mạng lưới cố định của các trạm giám sát**: Các trạm giám sát cố định được phân bố rải rác trong toàn thành phố để đo lường chất lượng không khí. Các trạm này thường được trang bị các cảm biến đo lường khác nhau để đo các chất gây ô nhiễm như PM2.5, PM10, NO2, SO2, O3, và CO.
* **Sử dụng drones**: Đối với các vùng khó tiếp cận hoặc nơi có các vấn đề về an ninh, việc sử dụng drones có thể là một phương pháp hiệu quả để thu thập dữ liệu không khí. Drones có thể được trang bị cảm biến đo chất lượng không khí và có thể bay qua các khu vực khác nhau của thành phố để thu thập dữ liệu.
* **Thiết bị di động giám sát**: Ngoài các trạm cố định, các thiết bị di động như xe hơi, xe buýt hoặc thậm chí là đồng hồ thông minh được trang bị cảm biến đo chất lượng không khí để thu thập dữ liệu khi di chuyển qua các khu vực khác nhau trong thành phố. Điều này giúp thu thập dữ liệu từ nhiều vị trí và điều kiện khác nhau.

1.1.2. Mô tả chi tiết tập dữ liệu

* Kích thước của tập dữ liệu gồm 48193 hàng
* Mô tả các đặc trưng (feature hay các cột của tập dữ liệu)
* Tập dữ liệu cung cấp các cột sau
* Datetime: Đây là thời gian và ngày ghi nhận các thông số khí quyển.
* PM2.5: Nồng độ hạt nhỏ có đường kính nhỏ hơn hoặc bằng 2.5 micromet trong không khí (micrograms/m^3).
* PM10: Nồng độ hạt lớn hơn 10 micromet trong không khí (micrograms/m^3).
* NO: Nồng độ khí oxit nitơ (micrograms/m^3).
* NO2: Nồng độ khí dioxide nitơ (micrograms/m^3).
* NOx: Tổng nồng độ của khí oxit nitơ (NO) và dioxide nitơ (NO2) (micrograms/m^3).
* NH3: Nồng độ khí amoniac (micrograms/m^3).
* CO: Nồng độ khí carbon monoxit (micrograms/m^3).
* SO2: Nồng độ khí dioxide lưu huỳnh (micrograms/m^3).
* O3: Nồng độ khí ozone (micrograms/m^3).
* Benzene: Nồng độ của benzen trong không khí (micrograms/m^3).
* Toluene: Nồng độ của toluen trong không khí (micrograms/m^3).
* Xylene: Nồng độ của xylene trong không khí (micrograms/m^3).
* AQI: Chỉ số chất lượng không khí (Air Quality Index) được tính toán dựa trên nồng độ của các chất gây ô nhiễm khí quyển.
* AQI\_Bucket: Phân loại mức độ ô nhiễm không khí dựa trên AQI, có thể là “Severe”.
* Các hàng của tập dữ liệu có thể xem xét một số thông tin quan trọng như giá trị trung bình, giá trị tối đa, giá trị tối thiểu, độ lệch chuẩn, và cũng có thể vẽ biểu đồ để hiểu rõ hơn về xu hướng và phân phối của dữ liệu.
* Thống kê mô tả cơ bản: Tính mean (trung bình), median (trung vị), mode (mođô), min (giá trị nhỏ nhất), max (giá trị lớn nhất), và standard deviation (độ lệch chuẩn) cho từng cột dữ liệu.
* Phân phối dữ liệu: Vẽ histogram hoặc box plot cho từng cột dữ liệu để xem phân phối của chúng.
* Tìm kiếm các giá trị bất thường (outliers): Sử dụng box plot hoặc đơn giản là xem các giá trị rời rạc có giá trị lớn hoặc nhỏ hơn nhiều so với các giá trị khác.
* Phân tích mối tương quan: Tìm mối tương quan giữa các cột dữ liệu bằng cách tính ma trận tương quan hoặc vẽ biểu đồ scatter plot.
* Phân tích xu hướng thời gian (nếu thích hợp): Nếu cột "Datetime" có dạng thời gian, bạn có thể phân tích xu hướng thay đổi của các chỉ số qua thời gian.
* Mục đích sử dụng và ứng dụng của tập dữ liệu
* Đánh giá chất lượng không khí cung cấp thông tin về chất lượng không khí tại một số địa điểm cụ thể, giúp các nhà quản lý môi trường, chính phủ, và cộng đồng địa phương hiểu rõ hơn về mức độ ô nhiễm không khí.
* Dự đoán và phòng ngừa ô nhiễm không khí phân tích xu hướng và biến động của các chỉ số khí quyển có thể giúp dự đoán và phòng ngừa ô nhiễm không khí trong tương lai. Các biện pháp kiểm soát và giảm thiểu ô nhiễm có thể được áp dụng dựa trên thông tin này.
* Dữ liệu về chất lượng không khí có thể được sử dụng trong nghiên cứu y tế và môi trường để hiểu tác động của ô nhiễm không khí đối với sức khỏe con người và môi trường.
* Dữ liệu có thể được sử dụng để so sánh với các tiêu chuẩn và quy định về chất lượng không khí được đề xuất hoặc áp dụng bởi các tổ chức y tế và môi trường.