PROJECT

AIR QUALITY FORECAST IN USA

INTRODUCTION

- This project is about the exploratory data analysis of the air quality across states in USA using pyspark.

- From the year 2000 through 2021, this dataset contains daily statistics on four important gas pollutants: carbon monoxide, nitrogen dioxide, ground-level ozone, and sulfur dioxide.

Keywords in the Data set

▶ **Ozone molecule** (O3) is harmful to air quality outside of the ozone layer.

▶ **Carbon Monoxide** (CO) is a colorless, odorless gas that can be harmful when inhaled in large amounts.

▶ **Sulfur dioxide** (SO2) is a colorless, reactive air pollutant with a strong odor. This gas can be a threat to human health, animal health, and plant life.

▶ **Nitrogen dioxide** (NO2) is a gaseous air pollutant composed of nitrogen and oxygen and is one of a group of related gases called nitrogen oxides.

▶ **Air Quality Index** (AQI)

DATA SOURCE

➤ https://www.kaggle.com/alpacanonymous/us-pollution-20002021/download

➤ https://aqs.epa.gov/aqsweb/airdata/download_files.html

DATASET DETAILS:

▶ Number of rows: 608700

▶ Number of columns: 24

❑ Date, Year, Month, Day, Address, State, County, City, O3 Mean, O3 1st Max Value, O3 1st Max Hour, O3 AQI, CO Mean, CO 1st Max Value, CO 1st Max Hour, CO AQI, SO2 Mean, SO2 1st Max Value, SO2 1st Max Hour, SO2 AQI, NO2 Mean, NO2 1st Max Value, NO2 1st Max Hour, NO2 AQI

▶ Size of the dataset: 97.76 MB

Goals of the Project

- To perform the exploratory Data analysis using pyspark and predict which city has the highest AQI value.

- Identifying which pollutant gas is affecting the environment more in each city and state.

- Check which air pollutant has increased over a period across the states in USA.

THANK YOU