

Project Proposal: Air Index of world's All Cities Analysis

Jasraj_Singh_Johal_johalj11_400434346

Data

- My data set Air Index of world's All Cities 2017 to 2022
- My dataset format is csv
- Dimensions of the dataset are 7323 rows and 21 columns
- <https://www.kaggle.com/datasets/bilalwaseer/air-index-of-worlds-all-cities-2017-to-2022/>

Questions Why interesting?

- Trends in Air Quality: Analyzing the air quality index over time (2017-2022) to identify any trends or patterns. Understanding which cities consistently rank high or low in terms of air quality. Regional Disparities:
- Investigating if there are regional differences in air quality by comparing countries and cities. Exploring whether certain countries or regions consistently have worse air quality compared to others.

Relevant columns for analysis may include:

- Rank, Countries, City, Monthly air quality index data (JAN-DEC for each year from 2017 to 2022) Historical data (2017-2021)

Broad strokes of the Data Workflow Pipeline (DWP):

- Pivoting: Pivoting the dataset to transform the monthly air quality index data from wide to long format, which will make it easier to analyze trends over time.
- Separating: Separating the data into different subsets based on regions (e.g., countries) or specific time periods for more focused analysis.

- Character Data Cleaning: Cleaning any inconsistencies or inaccuracies in character data such as city names or country names to ensure consistency in analysis.
- Refomating the data: Reformatting the data to make it easier to work with and analyze. This may include creating new variables or aggregating data to a higher level of granularity. ## Plot types that will help answer the questions:
- Time Series Plots, Bar Plots, Scatter Plots