Statistical Analysis of U.S. Air Quality Index (AQI) Data

Key Findings Report By Yahya Bhara

Basic Info / Summary of findings

1. **Dataset**: This dataset contains air quality measurements (AQI) across various U.S. states. The data includes information such as state, county, AQI values, and pollutant type (Carbon Monoxide).
2. **Key Metrics**: Mean AQI values per state, with a focus on specific comparisons, including Los Angeles County vs. other California counties, New York vs. Ohio, and Michigan’s AQI against a threshold value of 10.
3. **Statistical Methods**: Hypothesis testing conducted using Python’s scipy.stats package, including two-sample t-tests for regional comparisons and a one-sample t-test for Michigan's policy implication analysis

Key Insights

1. The mean AQI for Los Angeles County is statistically higher than other counties in California, with a p-value just below the significance level (p = 0.0498). This difference suggests Los Angeles may require targeted interventions to improve air quality.
2. Analysis shows that New York has a statistically lower mean AQI than Ohio (p = 0.0304), making it a preferable location in terms of air quality. This could influence decisions on future regional office placement.
3. Michigan’s mean AQI does not significantly exceed the threshold of 10 (p = 0.9399), indicating that Michigan may not need immediate policy intervention but should maintain current monitoring practices.

Next Steps

1. High AQI readings in specific regions should be validated for data accuracy. This is to check that if the high readings correspond to genuine pollution concerns or is it just data inconsistencies.
2. Allocate resources to high-AQI areas like Los Angeles County for focused air quality initiatives to improve the AQI
3. Consult with the data provider to confirm that the dataset represents broader state-level trends accurately.