Causal Inference Term Paper

Parth Sandeep Rastogi 2022352

The “ EFFECT” of GRAP-4

CONTEXT and MOTIVATION

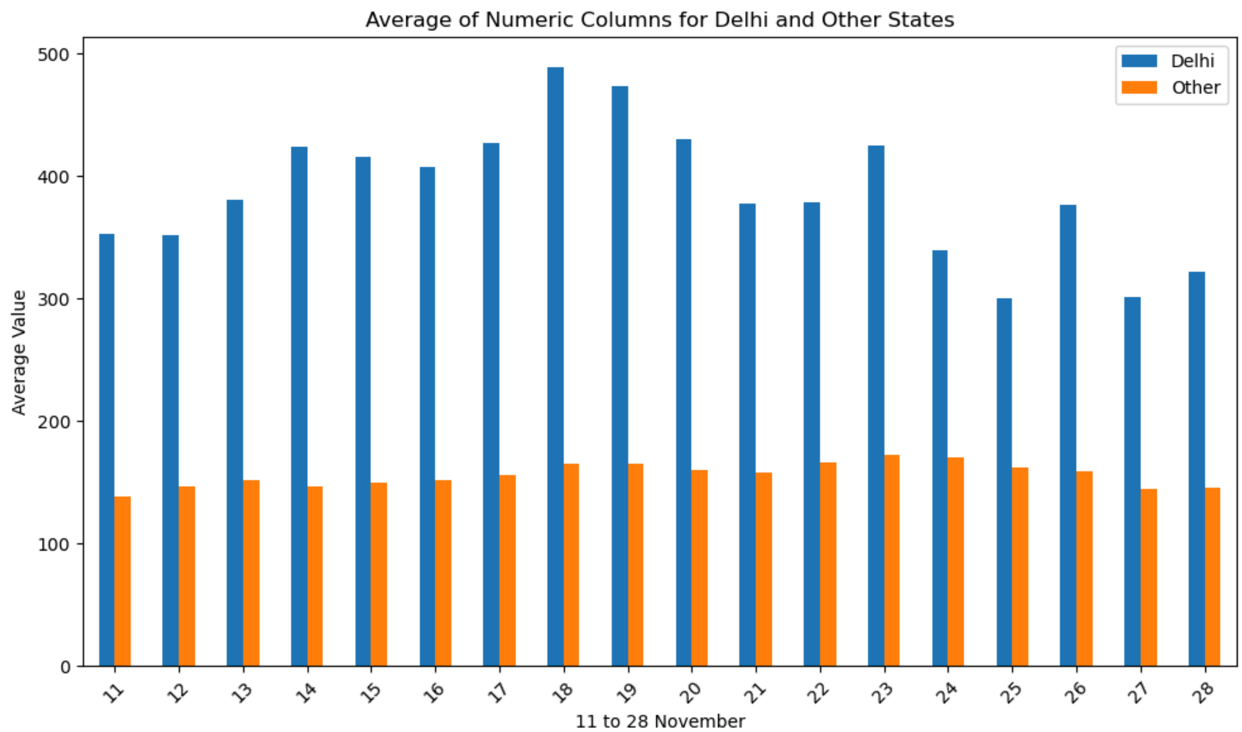
Air pollution in Delhi has reached an alarming level, mainly during the winter season. Several factors have to be taken into account when it reaches the alarming AQI of the city. Such factors include crop burning in neighboring states, frozen unmoving air of winter, fireworks and construction activities, overpopulation, and vehicular emission. Among these, crop burning by farmers in states like Punjab and Haryana is the big contributor, as they burn large areas of stubble following harvest, which releases substantial amounts of particulate matter in the atmosphere. Moreover, in Delhi, the winter season contributes to the problem because low temperatures and static air cause the inability of pollutants to disperse and thus develop smog and hazy conditions.

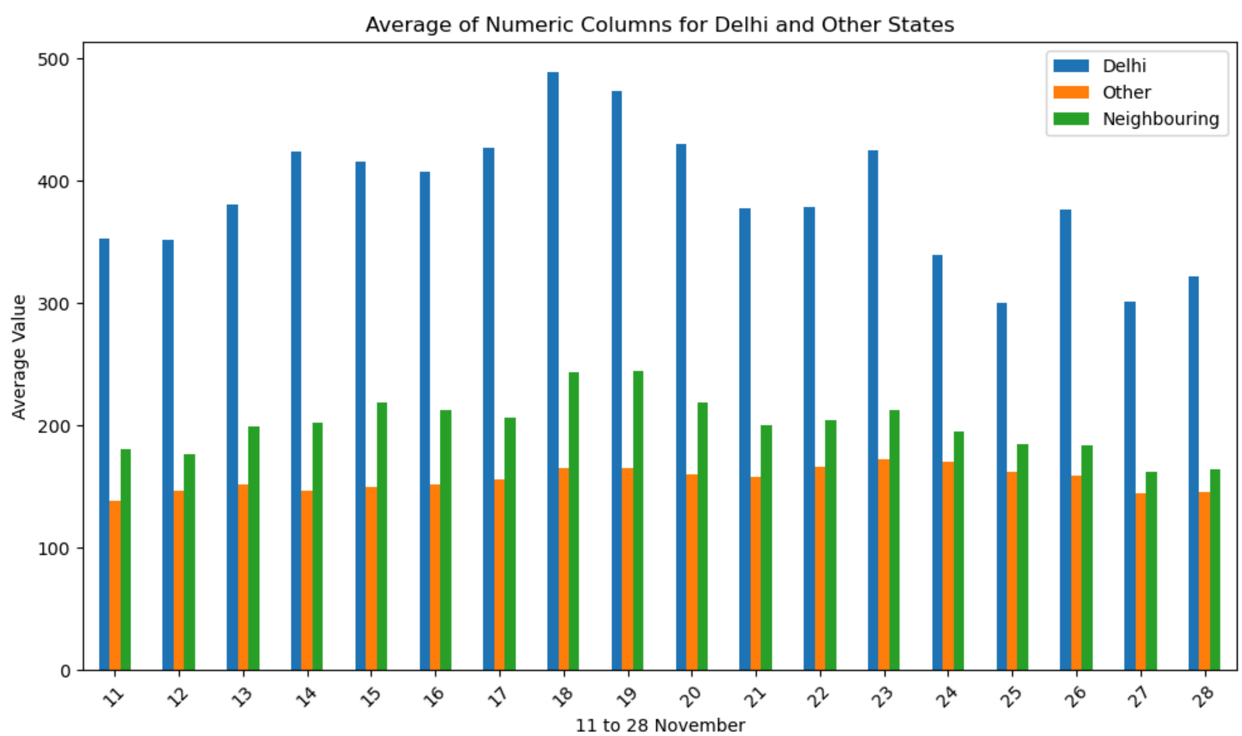
Due to this extreme environmental crisis of pollution, there is an implementation of a Graded Response Action Plan 4 to combat this pollution problem. This measure includes curbing various sectors like the entry of trucks into the city, construction of the building, and more on the online education with work-from-home that can reduce the emission produced by vehicles and industries also.

Primary focus of this study is on finding the impact of GRAP 4 on Delhi air quality. The variable which is to be put in the output is the AQI. The variable treatment is a comparison between measures taken by GRAP 4 in Delhi against the cities where GRAP 4 measures were not adopted. Comparing the AQI data of November 11th to November 28th, 2024, in the study, it can observe if GRAP 4 Improved air quality, is even significant or not.

Data Statistics and Description

Avg AQI of delhi and all other states





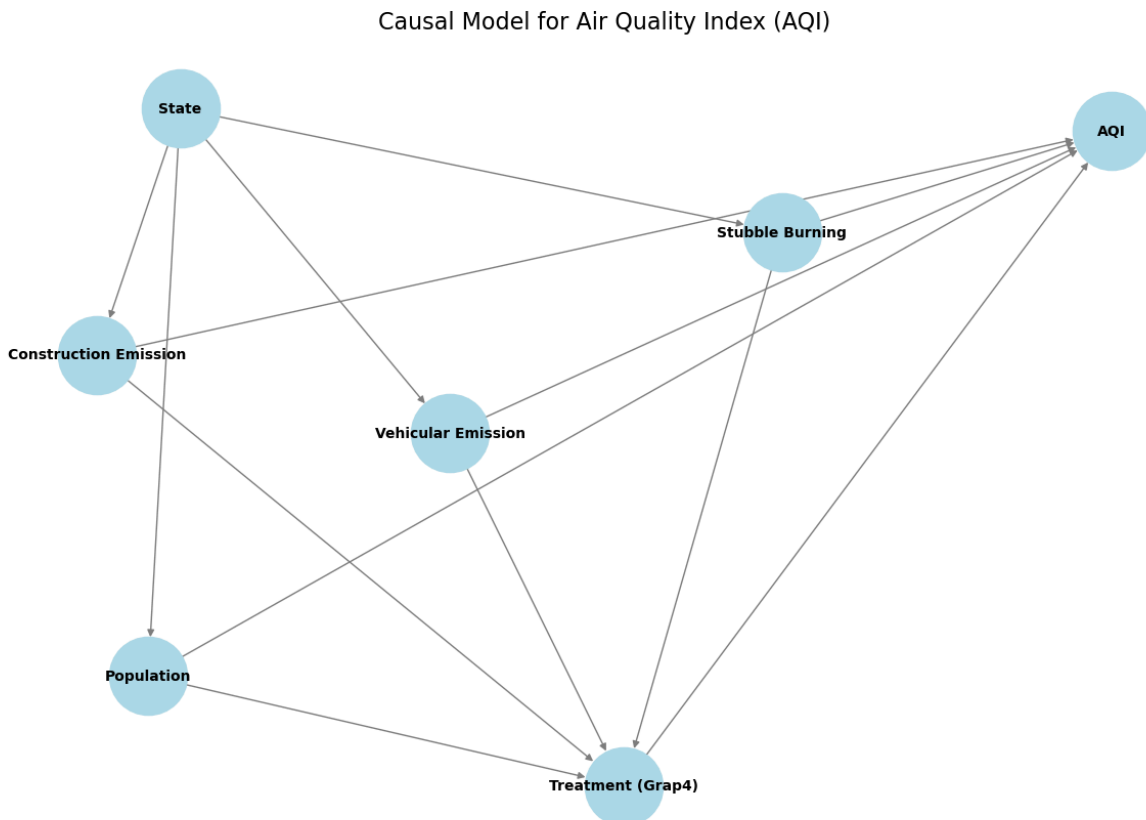
The dataset contains air quality index (AQI) values collected from 557 monitoring stations over the period from days 11 to 28, offering insights into air quality across three categories: Delhi, neighbouring states, and other states. Of the 557 stations, 39 are located in Delhi, while the remaining stations are distributed between neighbouring and other states.

Delhi exhibits a mean AQI of 387.39, reflecting severe air pollution. The standard deviation of 53.91 highlights significant variability, with AQI values ranging from 300 to 489 and a 75th percentile value of 424.88, indicating consistently poor air quality and occasional extreme pollution events. In contrast, other states show a mean AQI of 156.23 with a standard deviation of 9.81, indicating more stable air quality conditions. The AQI in these states ranges from 138.27 to 172.36, with a 75th percentile of 164.72. Neighbouring states, however, show a higher mean AQI of 200.69, with a standard deviation of 23.15, indicating moderate pollution levels with some variability. AQI values in these states range from 162.62 to 244.53, with a 75th percentile of 213.11, suggesting that pollution levels are higher than other states but still significantly better than Delhi.

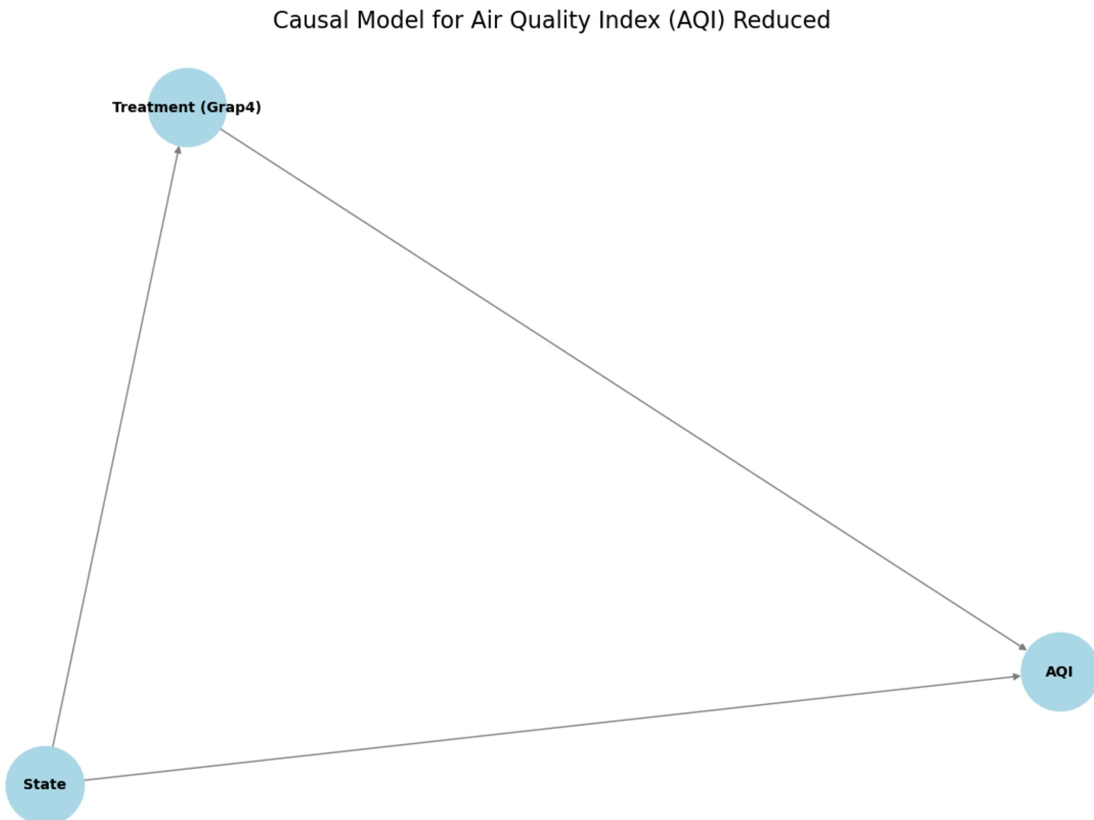
To address missing observations, the mean AQI for the respective state on the specific date was used, ensuring the dataset's completeness and consistency. This imputation approach maintains the representativeness of AQI trends across different regions.

The dataset reveals significant disparities in air quality, with Delhi experiencing the most severe pollution levels, far worse than neighbouring and other states. The station-level granularity provides a detailed perspective on regional air quality and highlights the urgent need for targeted interventions to address the critical pollution challenges faced by Delhi.

Causal Model



In the given DAG, latent variables such as construction emissions, stubble burning, population, and vehicular emissions indirectly influence AQI but are unobserved in the dataset. As we lack these confounders, we consider the state as a proxy confounder, introducing a backdoor effect on both the treatment and AQI. This approach helps account for state differences to isolate the causal impact of the treatment on AQI of delhi despite the unobserved latent factors.



This is the Reduced causal form.

Identification Assumptions and Methodology

My study employs Difference-in-Differences (DiD) to analyze the impact of policy interventions on Delhi's AQI compared to other states. Using I tested 3 assumptions: Canonical 2*2 ,parallel paths and parallel growth. The analysis uses three time periods: 11th (baseline, t=−1) , 18th (t=0), and 25th (t=1).

The canonical 2x2 DiD relies on the parallel paths assumption, which states that, in the absence of treatment, the change in average AQI would be the same for both the treated group (Delhi) and the control group (other states). While this assumption provides a straightforward baseline, it is overly simplistic given Delhi's inherently high AQI levels and large fluctuations. These characteristics make it unlikely that the treated and control groups naturally follow the same trajectory. However, this assumption serves as a useful starting point to establish initial comparisons before introducing additional pre-treatment periods for validation.

Regression used :- The regression model analyzes the impact of treatment on AQI, where the treatment variable (D) captures state-specific effects, and the time variable (T) distinguishes between the treated day (T=0) and the treated day plus 7 days (T=1), accounting for time-specific effects. The interaction term (D*T) measures the combined effect of treatment and time, with its coefficient representing the DiD.

$$Y = \mu_0 + \mu_1 D + \delta_0 T + \delta_1 (D \times T)$$

Extending to two pre-treatment periods strengthens the analysis by testing the parallel path assumption over a longer timeline. For instance, the AQI change between the 11th(Treated - 7 days) and 18th(Treated day) periods can be compared across treated and control groups to ensure their trends align before treatment. This approach addresses the limitations of the canonical assumption by providing more evidence that the groups behaved similarly prior to the intervention. If the trends are consistent, any divergence post-treatment can be attributed more confidently to the policy effect rather than random variability or unrelated factors.

Regression used :-   This regression model examines the effect of treatment on AQI across three time periods: T=0 (treated - 7 days), T=1 (the day treatment occurred), and T=2 (treated + 7 days). The treatment variable (D) captures state-specific effects, while P_1 and P_2 represent time-specific effects, with $P_1=1$ if T=1 and $P_2=1$ if T=2 (both 0 otherwise). The interaction terms ($D \times P_1$ and $D \times P_2$) account for the combined effect of treatment and time, with their coefficients measuring the differential impact on AQI in the respective periods. And difference in coefficient of interaction term is the DiD in our case $\gamma 1 - \gamma 0$

$$Y = \mu_0 + \mu_1 D + \delta_0 P_1 + \delta_1 P_2 + \gamma_0 (D \times P_1) + \gamma_1 (D \times P_2)$$

The parallel growth assumption incorporates two pre-treatment periods (11th and 18th), testing not just consistent levels but also consistent growth patterns. This ensures that both groups follow a similar trajectory in terms of both changes and rates of change before the treatment. It offers the strongest defense of the identification strategy by reducing the likelihood of biases caused by external shocks, or other unobserved factors. By demonstrating similar pre-treatment growth trends, the results post-treatment become far more credible, reinforcing that observed effects are due to the intervention.

Regression Used :-

This regression model examines the effect of treatment on AQI across three time periods: T=0 (treated - 7 days), T=1 (the day treatment occurred), and T=2 (treated + 7 days). The treatment variable (D) captures state-specific effects, while P_1 and P_2 represent time-specific effects, with $P_1=1$ if T=1 and $P_2=1$ if T=2 (both 0 otherwise). The interaction terms ($D \times P_1$ and $D \times P_2$) account for the combined effect of treatment and time, with their coefficients measuring the differential impact on AQI in the respective periods.

D2D will be $\gamma 1 - 2\gamma 0$

Results and Interpretation : -

Table of the first regression

==

	coef	std err	t	P>|t|
const	165.5029	3.442	48.088	0.000
D	323.4971	12.995	24.894	0.000
T	-19.7331	4.867	-4.054	0.000
D:T	-147.4208	18.378	-8.022	0.000

==

The regression results shed light on the effects of GRAP 4 over Delhi on AQI values. The constant term \(165.5) would represent the baseline AQI for the control group i.e., non-Delhi areas during the pre-intervention period. The coefficient associated with D ; (323.5) shows that in the pre-intervention periods, Delhi had much larger AQI values compared with the control group and depicts worse baseline pollution levels. The negative coefficient for T (−19.7) indicates that AQI in the control group decreased in the week following GRAP 4, indicating an overall trend of improving air quality that is unrelated to the treatment. The interaction term delta_0 = -147.4 , which is the treatment effect, shows that the AQI levels in Delhi have significantly decreased a week after GRAP 4 as compared to what would have happened without the intervention. This shows the effectiveness of the policy in improving air quality in Delhi. The high statistical significance of these coefficients underscores the robustness of the findings, demonstrating the significance of GRAP 4 in lessening the pollution while accounting for baseline differences and general trends in air quality.

For regression of the parallel path assumption

	coef	std err	t	P>|t|
const	138.2689	3.472	39.822	0.000
D	214.7568	13.110	16.381	0.000
P1	27.2340	4.910	5.546	0.000
P2	24.0851	4.910	4.905	0.000
P1:D	108.7403	18.541	5.865	0.000
P2:D	-77.1107	18.541	-4.159	0.000

The regression results analyze the impact of treatment (Delhi under GRAP 4) on AQI levels, with a baseline period of T−7 days, P1 representing the time of treatment (T), and P2 representing a week after treatment (T+7). The constant term (138.27) represents the baseline AQI for the control group during T−7. The coefficient for D (214.76) indicates that Delhi had significantly higher AQI than the control group during the baseline period. The coefficients for P1 (27.23) and P2 (24.08) reflect time effects for the control group, showing a slight decrease in AQI from T to T+7.

The interaction terms ($\gamma 0=108.74$ for P1:D and $\gamma 1=-77.11$ for P2:D) capture the treatment effects during T and T+7, respectively. The significant positive value for $\gamma 0$ suggests an initial increase in AQI in Delhi relative to the control group at the time of treatment, while the negative value for $\gamma 1$ indicates a substantial reduction in AQI a week later. Under the parallel path assumption, the difference ($\gamma 1-\gamma 0=-77.11-108.74=-185.85$) is the Difference-in-Differences (DiD) estimate, highlighting the net reduction in AQI due to the treatment over this period. This demonstrates the significant effectiveness of GRAP 4 in reducing pollution levels in Delhi.

Result for parallel growth

Under the assumption of parallel growth, there is no need to calculate a new regression. The Double Difference-in-Differences (D2D) estimate, calculated as $\gamma 1-2\gamma 0=-294.59$, highlights the substantial impact of GRAP 4. This result implies that, had pollution in Delhi followed the same growth rate, it would have been 294.59 points higher at T+7 (one week after treatment) than the observed value.

This demonstrates the effectiveness of GRAP 4 not only in halting the increase in pollution but also in achieving a significant reduction in AQI levels. The sharp decline underscores the intervention's success in mitigating the adverse effects of pollution in Delhi during the post-treatment period.

Robustness Check

As a robustness check, Bihar was used as a placebo group to assess the validity of the treatment effect. The coefficient for the interaction term was 33, with a p-value of 0.196, indicating it is statistically insignificant. This result suggests no measurable treatment effect in the placebo group, reinforcing the robustness of the findings for Delhi.

Conclusion and Limitations

This study demonstrates the stark effect of GRAP 4 in bringing down air pollution in Delhi, measured in terms of AQI. The robust canonical 2x2 Difference-in-Differences (DiD) with parallel paths and parallel growth demonstrate that GRAP 4 not only arrested the growth of pollution but also brought in meaningful reductions. The interaction terms and Double Difference-in-Differences (D2D) estimates validate that Delhi's AQI would have been substantially higher without GRAP 4. These results, validated by robustness checks using a placebo group, underscore the policy's efficacy in addressing Delhi's severe pollution crisis.

However, there are some limitations in this study. First, the GRAP 4 is still ongoing; thus, only short-term effects and not full effect are captured in this analysis. Second, possible spillover effects to neighboring cities where other pollution sources such as vehicular traffic or industrial activities must be taken into account to capture the full effect.

Third, the study uses state-level effects as proxies for unobserved confounders, but including other covariates such as stubble burning intensity, meteorological conditions, and localized pollution sources would strengthen causal inference. Finally, using only two pre-treatment periods limits validation of the parallel growth assumption. Expanding the analysis to include multiple pre-treatment periods (n-period analysis) would provide stronger evidence for these assumptions and ensure greater robustness of the findings. Addressing these limitations in future research would refine the conclusions.

References

What is GRAP 4? Delhi air pollution GRAP-4 restrictions; All You Need to Know - India Today :- For context

National Air Quality Index :- for data