Air Quality and Economic Progress in India

Nicholas Susemiehl
Georgia Institute of Technology

***Abstract*—Understanding the relationship between economic progress and the consequences of the methods used to achieve this progress is essential for future sustainable growth. In order to model the relationship between economic growth (proxied by unemployment rates) and concentrations of the pollutants SO_2, NO_2, CO, and particulate matter ($PM_{2.5}$), I first sought to synthesize a robust data set of pollutants and unemployment rates for 54 cities across India. Machine learning methods (e.g. XGBoost, KNN, ElasticNet) and complimentary data sets from other sources with high degrees of missingness were used to extrapolate a more reliable data set to cover an expanded date range of 2018 to 2022. I then used other statistical methods (including correlation analysis, linear regression, and random forests) to understand the relationship between the unemployment rate and these pollutants. This analysis was performed at both the city- and national-levels. Additionally, I visually examined the variations in pollutant levels within cities (across measuring locations) and examined the trajectory of emission and economic changes for each of the cities in the data set. At the city level, NO_2 is most highly correlated with unemployment rates (with a Pearson correlation coefficient of 0.22) and is the strongest predictor in a linear regression model of unemployment (coefficient of 0.29). Tree-based methods suggest that SO_2 is the most important predictor (relative importance of 0.3) at the city level. At the national level, NO_2 is the strongest predictor in a linear model (coefficient of 0.38) but CO is most highly correlated (-0.48) and the most important predictor in tree-based methods (0.8). I also found that most cities have a sufficient quantity and location of monitoring stations to measure within-city pollutant variations. While significant caveats regarding the treatment of missing data must be kept in mind, the results of this work suggest that SO_2 and CO are positively related to the unemployment rate while NO_2 is negatively related and $PM_{2.5}$ is unrelated.**

I. Introduction

In 2015, the United Nations released a list of 17 Sustainable Development Goals (SDGs) for tackling global health problems, poverty, and the climate crisis[1]. These SGDs provide a framework for the countries of the world to address the biggest problems facing our species and accomplishing them will pose significant challenges. In order to do so, humanity will need to rethink our approach to solving certain societal and economic problems because past approaches may no longer be effective. For instance, countries have attempted to combat poverty and hunger using industrialized methods which can be harmful to the environment. Approaching some SDGs with polluting processes would therefore hinder our ability to solve SDGs such as the climate crisis. At the forefront of this, it is critical to understand the how environmental factors affect economic factors and vise-versa. By concentrating on a particular nation, in this case India, more focused conclusions may be reached which could help this nation work towards accomplishing the SGDs.

One quantifiable aspect of the climate crisis is pollutant concentrations in the air. Pollutants such as SO_2, NO_2, CO, and $PM_{2.5}$ are environmentally harmful products of industrialized processes. Measuring concentrations of these pollutants can give scientists an understanding of the effects of the industrialized processes which created, but high quality data is difficult to come by. One such source is OpenWeather[2], an online service which curates and disseminates past, present and forecasted weather data. However, OpenWeather's pollutant data for India only goes back as far as November 2020. Since societal research typically spans longer time spans, an effort will be made to extend this data set backwards to 2018. Other sources, such as government and satellite data, span much longer periods of time but are of much lower quality. Limitations such as monitoring station locations and cloudy weather limit the completeness of these data sets. Additional factors, such as pandemic-era traffic levels, could also be useful in extrapolating the pollutant data. Once this more comprehensive data set of pollutant concentrations is synthesized, its relationship to economic growth, represented in this work by the unemployment rate, can be analyzed.

II. Methods

In order to investigate the relationship between air quality and economic progress in India, it was first necessary to collect data from multiple sources. The primary pollutant data set from OpenWeather provides high quality pollutant concentration data but only over a two-year time span of 2020 to 2022. Other data sets which span a longer period of time yet are of lower quality, including air quality data from government and satellite sources as well as traffic data, were also collected. Following an exploratory analysis of the pollutant and economic data sets, missing data were imputed. Once the pollutant and traffic data was prepared, they were then used as predictors to model the OpenWeather data set. These models were used to backfill the OpenWeather pollutant features so that they would extend to 2018. After the pollutant concentration data was extended, direct comparisons were made between it and the economic data using machine learning techniques. Once the main analysis was completed, further work was done to investigate the variations in pollutant data across monitoring stations in the government data set and classify cities according to their environmental and economic performance over time.

[1] https://sdgs.un.org/goals

[2] https://openweathermap.org/

A. Exploratory Data Analysis

The first step in this work was to gain an understanding of the data that would be used. This included checking for missing data and treating outliers. All of the data used in this work were provided by the practicum sponsors.

1) OpenWeather Data: The OpenWeather data set was considered to be the ground truth for among the pollutant data. It records daily atmospheric concentrations of the pollutants SO_2, NO_2, CO, and particulate matter ($PM_{2.5}$) for 54 Indian cities over a time period spanning from November 24, 2020 to July 25, 2022. These pollutants are each given in units of $\mu g/m^3$, so the pollutant features in each other data set will need to conform to this unit. Latitude and Longitude coordinates for each city are also included in this data set but these were not used. In total, the OpenWeather data set contains 32,832 rows and no missing values. I used three techniques to check for outliers: the IQR method (which flags observations as outliers if they are exceed the first or third quartile by 1.5 times the IQR), the Z-score method (which flags observations as outliers if they have a Z-score greater than three), and boxplots (which enable a visual inspection relative to the IQR). The IQR and Z-score methods assume the data to be normally distributed, which is not the case for any of the data sets used throughout this work. These two outlier detection methods identified outliers too readily, so I relied on boxplots to make decisions to exclude data points for this data set as well as all others. The removal of outliers was not necessary in the case of the OpenWeather data.

2) Government Data: The government data comes from an official Indian source for each of the same 54 cities found in the OpenWeather data set. Each city in this data set has data correspondong to one or more monitoring stations positioned at different locations within the cities. The government data set (once combined across cities and monitoring stations) contains 57,058 rows with columns corresponding to the same four pollutants present in the OpenWeather data set. This data contains daily pollution concentration measurements spanning from January 1, 2018 to December 31, 2021. The pollutant NH_3 is also included in the government data but was not used in this analysis due to its absence from the OpenWeather data. Before counting the number of missing values, the text ”None” as well as any pollutant values equal to or less than zero were replaced with ”NA”. Following this, 54% of the SO_2 and NO_2, 90% of the CO, and 58% of the $PM_{2.5}$ data were missing. CO is given in this data set in units of mg/m^3 whereas each other pollutant is in units of $\mu g/m^3$, so the government CO values were multiplied by 1000 to convert them to the proper units. Boxplots revealed several obvious outliers in the SO_2, CO, and $PM_{2.5}$ data which were removed due to their positions significantly above the 75th percentile. Lastly, some city names were changed to conform with the spellings present in the OpenWeather data.

3) Satellite: The satellite data contains 413,647 daily observations over a time period ranging from January 1, 1980 to August 11, 2022 and the same 54 cities present in the previous data sets. The features in this data set are also the same four pollutants present in the previous two. The first step taken was to replace the stand-in values of $-1.27 * 10^{30}$ and -9999 with NA markers. As before, non-positive pollutant values were also replaced with ”NA”. These steps led to missing value proportions of 77% for SO_2, 50% for NO_2, 79% for CO, and 93% for $PM_{2.5}$ for the full satellite data. The pollutants in this data set are given in units of molecule/cm^2 and so these data needed to be converted to $\mu g/m^3$. A research paper[3] provided by the practicum coordinators gives a formula for doing so:

$$V = A * \frac{avg.\ ground}{avg.\ satellite} \tag{1}$$

where V is the pollutant in volume density units ($\mu g/m^3$), A is the pollutant quantity in units of area density (molecule/cm^2), ”avg. ground” refers to the average pollutant quantity of ground-based measurements which are in units of $\mu g/m^3$ (taken as the mean of the respective OpenWeather pollutant features), and ”avg. satellite” refers to the average quantity of satellite measurements which are in units of molecule/cm^2. This conversion was performed independently for each of the four pollutants. The pollutant distributions of the satellite data following these conversions was visually compared to those of the other two data sets using histograms and this revealed that the satellite data are distributed similarly to the OpenWeather and satellite data. Therefore, this conversion was considered to be valid. Moving on, two outlier were found in the SO_2 satellite data and removed. Some cities were corrected to have consistent spellings with the OpenWeather data.

4) Combination of Pollutant Data: I produced several plots in order to verify the similarity of the OpenWeather, government, and satellite data. These data should be distributed somewhat similarly since they are measuring the same phenomena at the same locations, but some differences can be expected due to differences in collection methods. These comparisons were made after all unit conversions and outlier removals. First, I overplotted histograms of the SO_2, NO_2, CO, and $PM_{2.5}$ distributions in Figure 1.

These reveal the data overall are distributed similarly, but the NO_2 and CO features in the satellite data set are more concentrated at intermediate values than the same pollutants are in the other data sets. Next, I aggregated each of the pollutant variables over the cities and plotted the time series of pollutant concentration in Figure 2.

These plots show the periodic nature of pollutant concentrations. The three data sets show similar curves for NO_2, but the government data is typically valued lower than the OpenWeather and satellite data in the cases of the three other gases. These differences in the distributions of the three pollutant data sets suggest that the government and satellite data sets may not be perfect predictors of the OpenWeather data.

5) Traffic: The traffic data, which is provided by the Google Covid Mobility Reports, contains the percentage change in

[3]https://github.com/MacroXStudio/gatech/blob/main/Research_paper.pdf

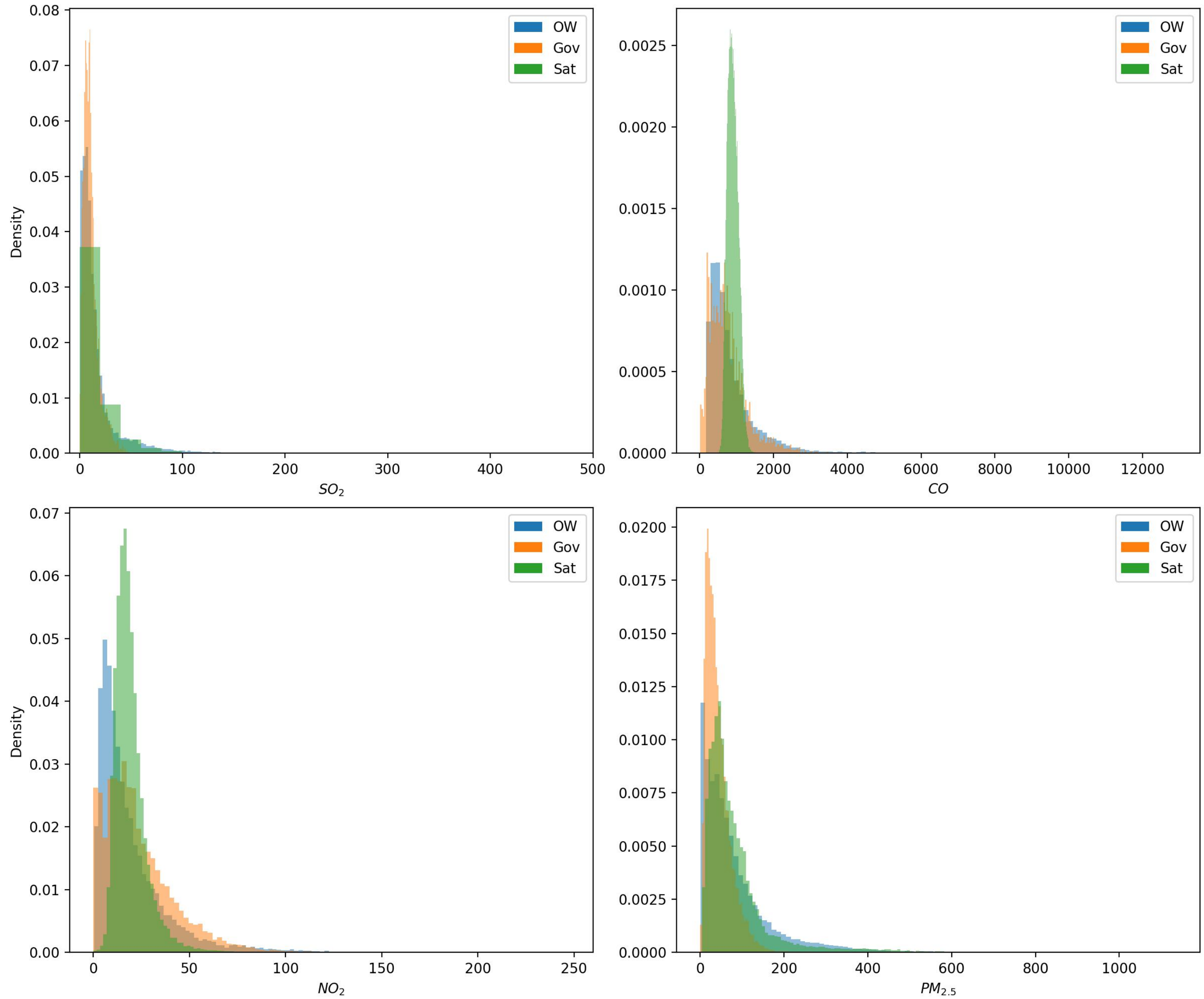

Fig. 1. Density histograms of the pollutant data across the three sources.

traffic quantities from a pre-Covid baseline in six sectors from March 1, 2020 to September 3rd, 2022 over 628 cities. In total, there are 587,101 rows in this data set. The sectors in which traffic changes were recorded as well as their repsective missing value proportions are as follows: retail and recreation, 7%; grocery and pharmacy, 14%; parks, 6%; transit stations, 5%; workplaces, 0.4%; residential, 6%. No outliers were found in this data set using boxplots.

6) Combination of Pollutant and Traffic Data: Following the exploration and initial processing of these data sets, the pollutant and traffic data sets were combined via an outer join on the ”Date” and ”City” columns. Any record corresponding to a date earlier than January 1, 2018 or a city not in the set of 54 present in the OpenWeather data was dropped. In total, there are 90,910 rows in this data set. The proportions of missing data for each government, satellite, and traffic feature in the full data set are shown in Table I.

7) Unemployment: The economic data provided by the practicum sponsors contains monthly records of unemployment rates in the same 54 Indian cities as the previous data sets over a time period from January 31, 2016 to July 31, 2022. There are no missing data in this data set and no definitive outliers were found via inspection of the boxplot of the unemployment rates. Two points with exceptionally high unemployment rates are potential outliers, but since they cover two months following the onset of the Covid-19 pandemic they were retained due to the turbulent nature of this time. No other pre-processing steps were performed on this data set. Since these data are monthly and the pollutant and traffic data are daily, steps were taken in the proceeding stages of this work to align the dates.

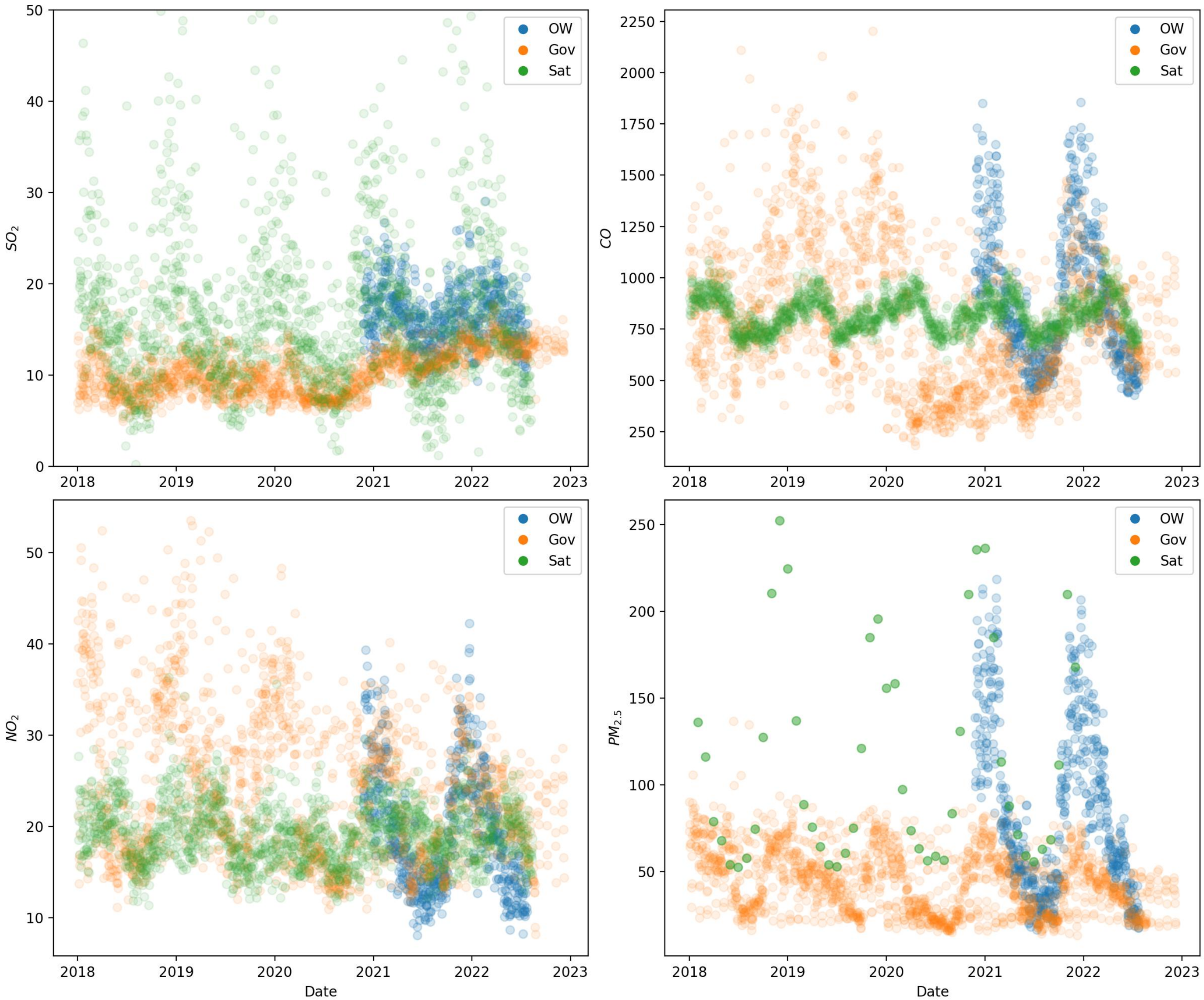

Fig. 2. Time-series plots of the pollutant data across the three sources.

TABLE I
MISSING VALUES BY FEATURE

Feature	Proportion of Missing Values
SO_2 (Gov.)	0.64
NO_2 (Gov.)	0.64
CO (Gov.)	0.91
$PM_{2.5}$ (Gov.)	0.67
SO_2 (Sat.)	0.76
NO_2 (Sat.)	0.49
CO (Sat.)	0.81
$PM_{2.5}$ (Sat.)	0.98
Traffic (All)	0.27

B. Handling Missing Data

Most features in the combined pollutant and traffic data contain are missing significant fractions of their observations. Figure 3 provides a visualization of how much data is missing and where these missing values exist for each column.

This missing data must be imputed in order for this data set to be useful in extending the OpenWeather data over a longer period of time. The K-Nearest Neighbors (KNN) algorithm was used to perform the imputation on each of the

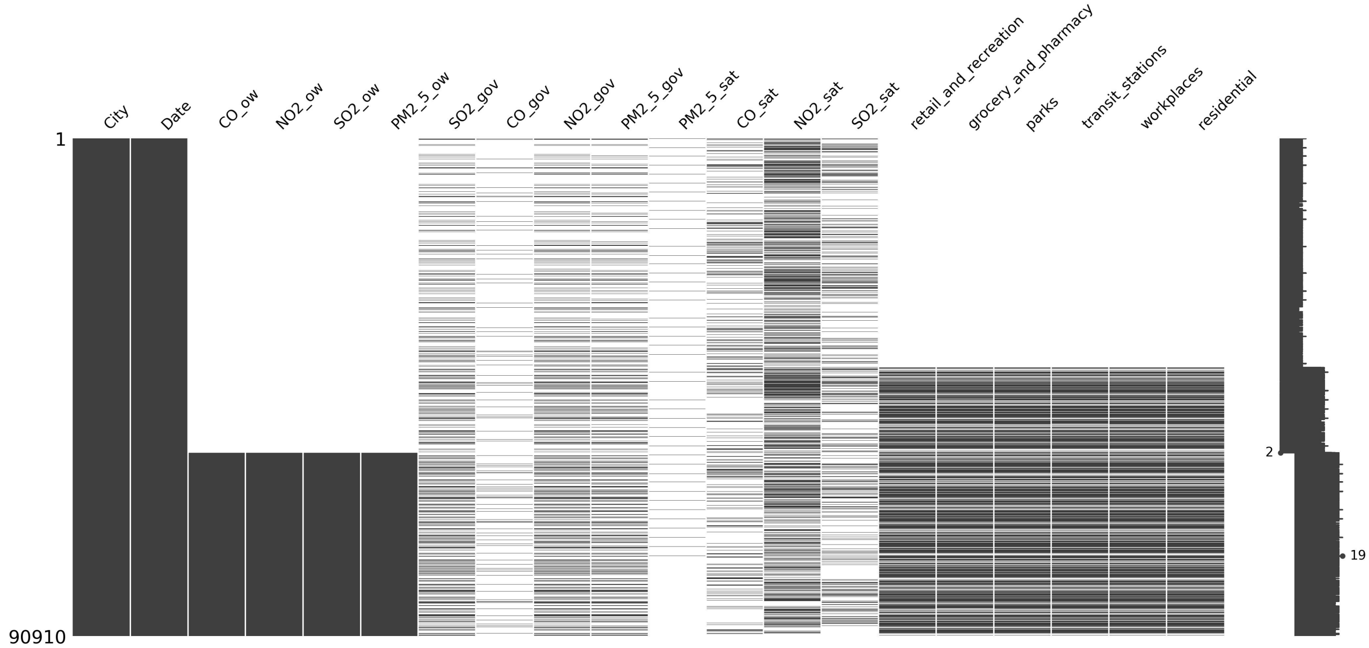

Fig. 3. Locations of missing values. Black regions are where values exist and blank regions are where values are missing. All values are sorted by date, with January 1, 2018 corresponding to the top and July 22, 2022 corresponding to the bottom.

individual columns in the government, satellite, and traffic data sets. First, the subset of the column of interest and the City and Date columns was taken from the respective data sets. Five-fold cross validation was used to finding the optimal nearest neighbor hyperparameter by iteratively testing the performance of KNN models using 1, 2, 4, or 8 neighbors trained and evaluated on the non-missing portions of the data. This procedure only tested a small number (relative to the size of the data) of nearest neighbors in order to preserve the cyclical variations in the data that could be lost if large number of neighbors are used (i.e. averaging over multiple periods). The mean absolute error was used as the error metric during training. Once the optimal number of nearest neighbors was found, the model was re-trained using this hyperparameter value on all of the non-missing data. The trained models were then used to predict values at the locations of the missing data. This process was performed twice on all of the data - once using normalized data and once using untransformed data. Transforming the data by scaling the mean to zero and standard deviation to unity resulted in lower mean absolute errors during training on the non-missing portions of the data, but the final predictions of the missing data looked more plausible when the untransformed data were used. Examples of this are provided in Figure 4.

The data imputed without transforming the data seems to more plausibly capture the variations present in the non-missing data, so these are what is used in the rest of this work. Because the traffic data measures changes in traffic patterns relative to a pre-Covid baseline, all traffic values prior to the starting date of this data, February 14, 2020, were filled with zeroes (indicating no change) rather than predicted values.

C. OpenWeather Modeling

With the missing data in the pollutant and traffic sets imputed, these data could now be used to extrapolate the OpenWeather data. This was performed on both the city-level (using all rows in the data set) and national-level (by averaging over the cities) using otherwise identical procedures. First, the data were split into three subsets based on date: one containing all observations between January 1, 2018 and the starting date of the OpenWeather data (November 24, 2020), another containing all records dated between the start of the OpenWeather data and December 31, 2021, and the final containing all rows with dates between the January 1, 2022 and the end of the OpenWeather data (July 25, 2022). The data prior to the start of the OpenWeather data would be used as the features to predict the OpenWeather during this preceding time period, the 2020-2021 data was used to train models to predict the OpenWeather data, and the 2022 data was used to evaluate these models. Three classes of algorithms were tested: XGBoost, ElasticNet, and KNN. The respective pollutant in the government and satellite data sets as well as all of the traffic variables were selected as features to predict each of the pollutants in the OpenWeather data. These features were then normalized by shifting the mean to zero and standard deviation to one. An optimal hyperparameter search was performed using the 2020-2021 data as the training set and the 2022 data as the testing set. The models which yielded the lowest mean-squared error (MSE) on the 2022 evaluation set were selected to make the final predictions. While the model selections were handled on a case-by-case basis for each variable, the ElasticNet model was

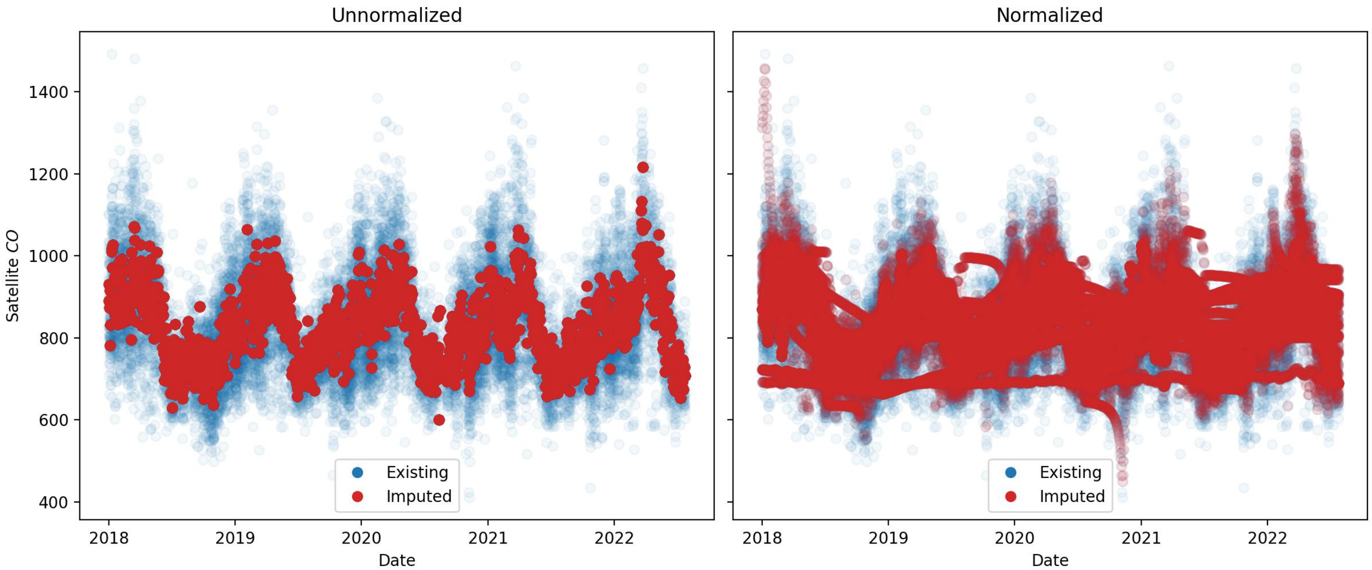

Fig. 4. Time-series plots of the satellite CO data imputed using the raw (unnormalized) and normalized data. Normalizing the data results in a imputations that are inconsistent with the cycles present in the data for this feature as well as others.

most commonly selected as the optimal model. The selected model (using the previously found hyperparameters) was then trained on the combined 2020-2021 and 2022 data and used to make backwards predictions of pollutant levels prior to the starting date of the OpenWeather data. These predictions were then combined with the OpenWeather data set to form a final data set with no missing values ranging from January 1, 2018 to July 25, 2022.

D. Economic/Environmental Analysis

Now that the pollutant data set is complete over an extended time period, it could be used to understand the relationship between air pollutants and unemployment rates. This was again performed at the city- and national-levels, but the procedures are identical other than the previous aggregation. As previously stated, the unemployment data is recorded at monthly intervals whereas the pollutant data is daily. These data sets were made aligned by averaging the pollutant data over each month. The monthly-aggregated city-level data set has 2,841 rows while the respective national-level data set has 54. The data were normalized to have a mean of zero and standard deviation of one in order to remove biases induced by the widely varying scales of different features. Once this was done, several statistical techniques were employed to investigate which pollutants are significant predictors for the unemployment rate. First, I examined the Pearson correlation between variables. Then, I fit a linear model with all of the pollutants as independent variables and the unemployment rate as the dependent variables and examined the model coefficients and p-values. I also employed tree-based methods, including single decision trees, random forests, and XGBoost models. These models provided the relative importance of difference variables when used as predictors.

E. Economic/Environmental Changes

Next, I investigate the changes in pollutant levels and unemployment rates for each city from quarter-quarter. To do this, I combined the pollutant features into a normalizaed air quality index (AQI) by first scaling all of the pollutants to be values between zero and one. Next, I took the averages of the scaled AQI's and unemployment rates for each of the quarters. I calculated the incremental changes in scaled AQI and unemployment rates by taking the lagging difference between quarters. I then classified each city into one of four quadrants of 2x2 matrices corresponding to positive and negative pollutant and economic change for each quarter. I also produced an interactive visualization to show the changes in scaled AQI and unemployment rate for each city over time.

F. Within City Variations

I also sought to investigate pollutant level variations within cities due to differences in the locations and quantities of monitoring stations. The government data set is the only one which stratified the pollutant data by monitoring station, so this was the only data used in this analysis. I again calculated a scaled AQI by averaging the pollutants detected by each monitoring station after scaling them to be between zero and one. I then produced an interactive visualization to show the average scaled AQI levels for each monitoring station of each city.

III. Results

The main results of this work include the extropolations of the OpenWeather data and the inferences regarding the relationship between economic growth and emissions drawn from these data. Also presented are analyses in the change in economic and emissions statuses and within-city variations for each city presented in the data set.

A. OpenWeather Modeling

Extrapolations of the OpenWeather data over the period ranging from January 1, 2018 to its start date of November 24, 2020 were produced at both the city- and national- (where pollutant and traffic values are averaged over the cities) levels. These processes resulted in daily pollutant (SO_2, NO_2, CO, and $PM_{2.5}$) data ranging from January 1, 2018 to July 25, 2022. Time series plots showing these extrapolations are presented in Figure 5 for the city-level data and Figure 6 for the national-level data.

The model with the lowest MSE on the 2022 subset from among XGBoost, ElesticNet, and KNN was selected to make the extrapolation. At the city level, the optimal model for SO_2 was ElasticNet with an MSE of 142.02; for NO_2 the optimal model was KNN with an MSE of 123.05; for CO the optimal model was ElasticNet with an MSE of 286608.20; for $PM_{2.5}$ the optimal model was ElasticNet with an MSE of 4010.21. At the national level, the optimal model for SO_2 was ElasticNet with an MSE of 8.42; for NO_2 the optimal model was KNN with an MSE of 70.08; for CO the optimal model was ElasticNet with an MSE of 82254.10; for $PM_{2.5}$ the optimal model was KNN with an MSE of 3216.32. In each case, the predictions on the nationally-aggregated data are associated with lower evaluation MSE's than the predictions on the city-level unaggregated data. Visually, the national-level predictions capture more cyclical variation in the pollution levels than the city-level predictions

B. Economic/Environmental Analysis

The metrics produced by the linear and tree-based models which used the pollutant concentrations to predict unemployment rates can be used to understand which, if any, pollutants are related to the unemployment rates. The Pearson pairwise correlation between the pollutant features and the unemployment rate, linear model (LM) coefficients and p-values, and tree-based relative importance statistics are presented in Table II for the city-level data and Table III for the national-level data.

At the city level, none of the pollutants are highly correlated with the unemployment rate. The pollutant most strongly correlated with the unemployment rate is NO_2, which also has the largest magnitude LM coefficient. All pollutants are positively correlated with the unemployment rate but the LM coefficient is negative in each case except for NO_2. At the 0.05 significance level, SO_2, NO_2, and CO are significant predictors of unemployment rate in a linear model. The tree-based methods do not suggest that one pollutant in particular is significantly more important than the others, but SO_2 may be slightly more important than NO_2 followed by $PM_{2.5}$ then CO.

The national-data suggests that different relationships between the pollutants and unemployment rate may exist. All of the pollutants are moderately and negatively correlated with the unemployment rate, with the CO correlation being the highest in magnitude. The LM coefficients for each pollutant are again all negative except for that of NO_2. Here, NO_2 and CO have large coefficient values. However, only CO is a significant predictor in the LM at the 0.05 level. The tree-based methods all indicate that CO is by far the most important predictor of the unemployment rate.

C. Economic/Environmental Changes over Time by City

Each city was classified into 2x2 matrices corresponding to positive and negative economic and emissions changes for each of the 18 quarters covering the date range spanned by this data set. One such matrix is provided in Figure 7 as an example and the other 17 can be found in the appendix.

An interactive visualization which shows the differences in the economic and emissions values for a given city over all quarters was also produced. An example plot is shown in Figure 8. Instructions for running visualization can be found in the appendix.

D. Within City Variations

The time-averaged scaled AQI aggregated by monitoring station for a given city is also depicted in an interactive visualization. An example is provided in Figure 9

The averaged scaled AQI tends to vary across monitoring stations for most of the cities. Agartala, in Figure 9, is an example of this.

IV. Summary

The goal of this work was to investigate the relationship between air pollutant concentrations (namely SO_2, NO_2, CO, and $PM_{2.5}$) and unemployment rates in India. Doing so requires several years of pollutant data, but the high-quality source OpenWeather only provides records going back to 2020. Supplementary pollutant data sets from government and satellite sources which span a longer period of time are available but are missing significant portions of data (ranging from 49-98%). Once these missing data points were imputed, the government and satellite sources as well as traffic data were used to train models capable of extending the OpenWeather data back to 2018. The relationships between the extended pollutant data and the unemployment rates were then investigated using linear and tree-based methods. This analysis was performed on the city- and national-levels and the mixed results suggest that some pollutants (mainly NO_2 and CO) may be adequate predictors of the unemployment rate. The change in emissions and unemployment levels over time was also investigated and each city was classified according to its performance. Lastly, an examination of within-city pollutant level variations suggest that monitoring stations are sufficiently located to measure these variations.

V. Conclusions

Any conclusions drawn from this work must be considered alongside the numerous ways the data was augmented. Several assumptions were made regarding the units of the units of the government and satellite pollution data which led to plausible yet unverified transformations. Following this, imputations were performed on features missing as much as 98% of

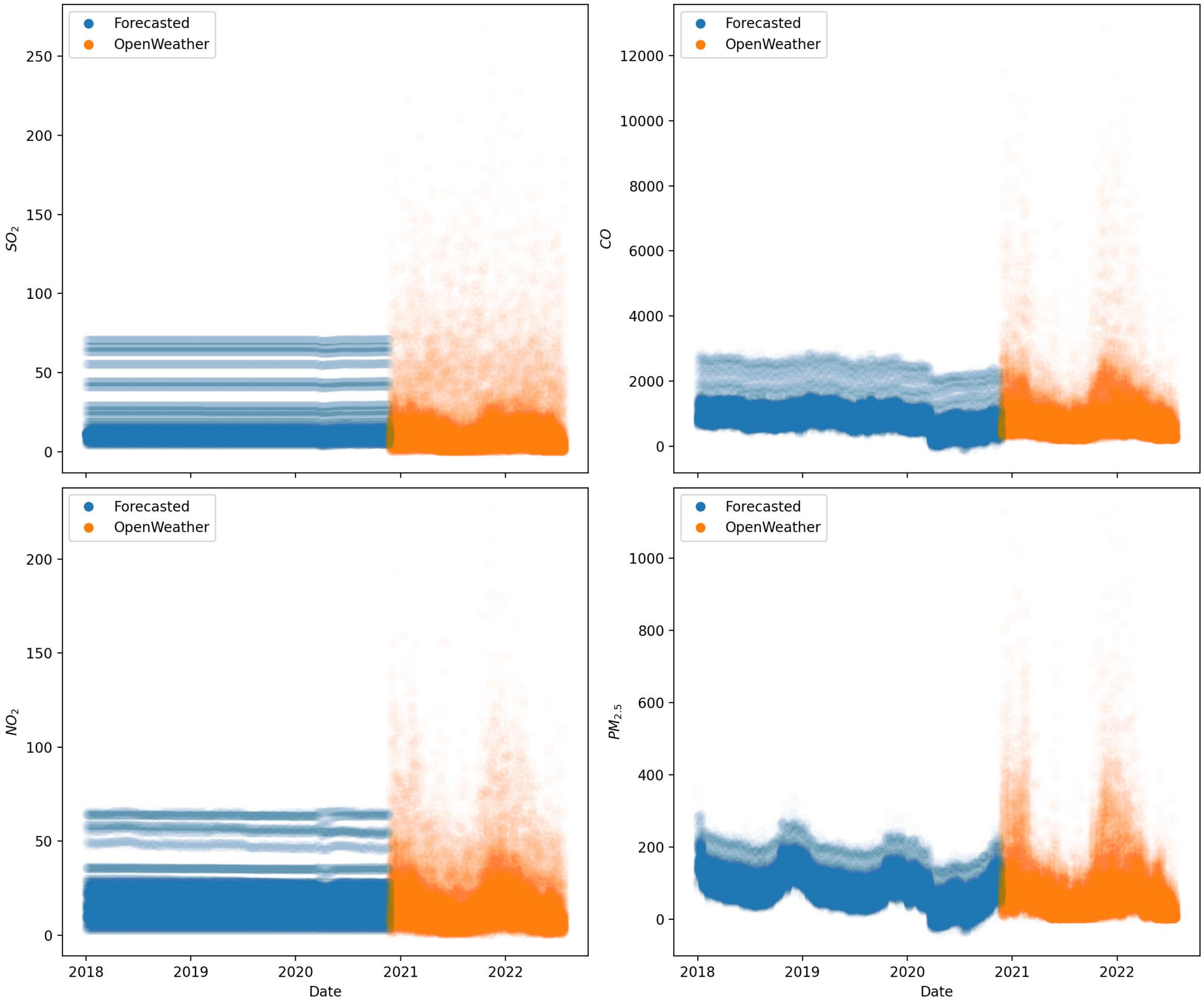

Fig. 5. Time-series plots of existing OpenWeather data extended by the forecasted values for each of the four pollutants (city-level).

TABLE II
CITY-LEVEL CORRELATIONS ("CORR."), LINEAR MODEL COEFFICIENT AND P-VALUES ("LM COEF.", "LM P-VALUE"), AND RELATIVE IMPORTANCES FROM DECISION TREE, RANDOM FOREST, AND XGBOOST MODELS ("DT IMPORT.", "RF IMPORT", AND "XGB IMPORT." RESPECTIVELY).

Pollutant	Corr.	LM Coef.	LM p-value	DT Import.	RF Import.	XGB Import.
SO_2	0.0877	-0.2499	0.0000	0.3321	0.3358	0.3239
NO_2	0.2215	0.6609	0.0000	0.3041	0.2938	0.2983
CO	0.1145	-0.2156	0.0012	0.1591	0.1798	0.1320
$PM_{2.5}$	0.0513	-0.0711	0.1136	0.2046	0.1906	0.2458

TABLE III
NATIONAL-LEVEL CORRELATIONS ("CORR."), LINEAR MODEL COEFFICIENT AND P-VALUES ("LM COEF.", "LM P-VALUE"), AND RELATIVE IMPORTANCES FROM DECISION TREE, RANDOM FOREST, AND XGBOOST MODELS ("DT IMPORT.", "RF IMPORT", AND "XGB IMPORT." RESPECTIVELY).

Pollutant	Corr.	LM Coef.	LM p-value	DT Import.	RF Import.	XGB Import.
SO_2	-0.3520	-0.0624	0.7396	0.0126	0.0563	0.0623
NO_2	-0.2657	0.5274	0.1524	0.1033	0.0919	0.0807
CO	-0.4843	-0.8203	0.0009	0.8262	0.7766	0.8117
$PM_{2.5}$	-0.2760	-0.0696	0.8358	0.0579	0.0751	0.0452

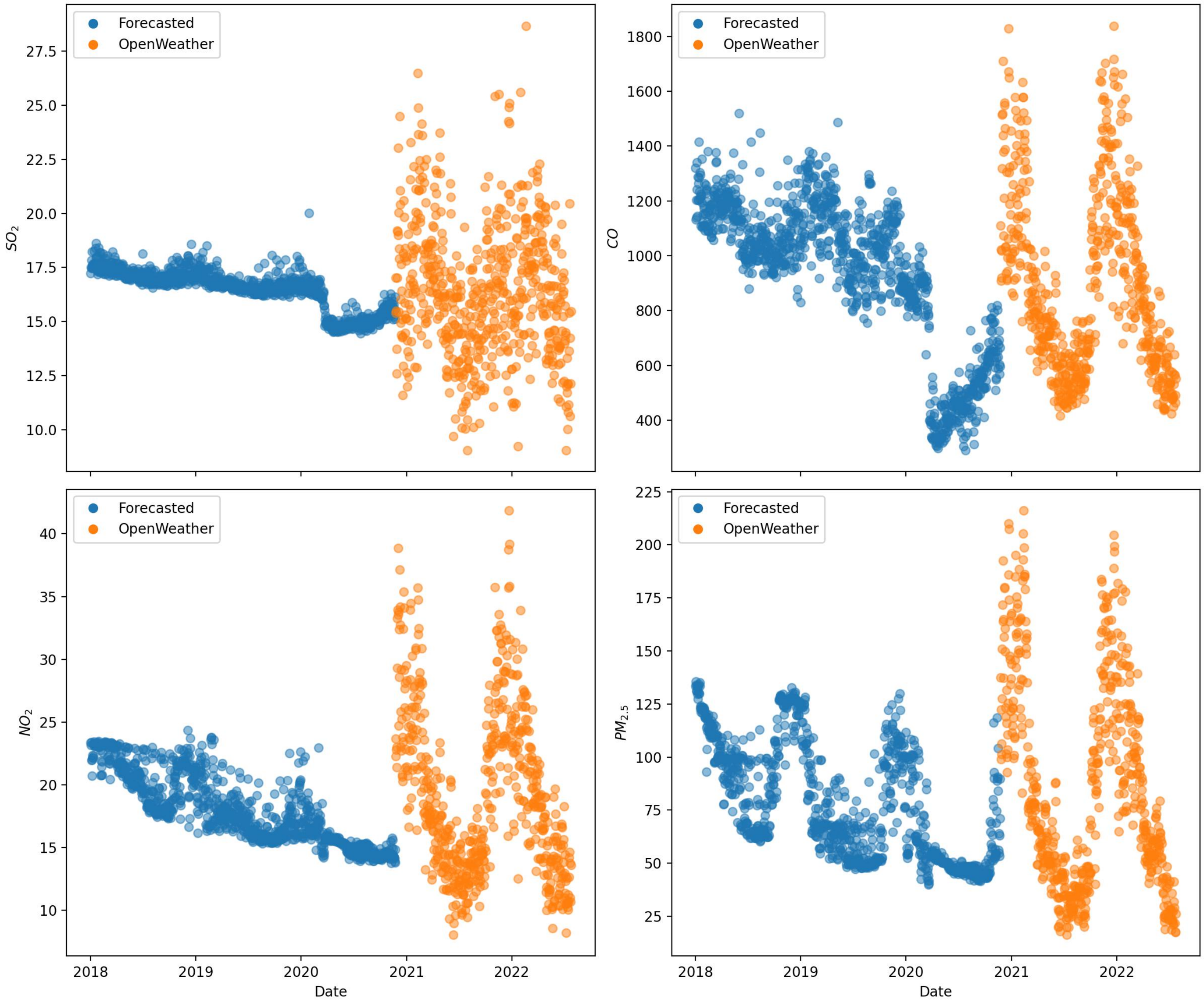

Fig. 6. Time-series plots of existing OpenWeather data extended by the forecasted values for each of the four pollutants (national-level).

their records. The data were also aggregated in several steps (to perform national-level analyses and to align the pollutant data frequency with that of the unemployment data). These extensive operations increase the abstraction of the data set and distance it from reality. However, the results of this work can still serve to construct a first-order understanding of the relationship between emissions and economic progress in India.

The various metrics used to assess the relationship between the pollutants and the unemployment rate suggest different conclusions at both the city and national levels. At the city level, the NO_2 coefficient in the linear model is both statistically significant and relatively large in magnitude. The tree-based methods also suggest that this pollutant is important relative to the others. Therefore, this view of the data suggests that NO_2 is the pollutant most strongly associated with the unemployment rate. Following similar logic, SO_2 and CO are also somewhat related to the unemployment rate (although perhaps to a lesser degree). $PM_{2.5}$ has the lowest correlation coefficient and is the only pollutant in this view with a non-significant p-value in the linear model, so it is less likely to be related to the unemployment rate in reality.

At the national level, CO is the only pollutant with a statistically significant p-value in the linear model. Its correlation coefficient is also the highest in magnitude and the tree-based methods all weigh this pollutant as the most important by far. Therefore, this analysis suggests that CO is a strong predictor of the unemployment rate when considering the national-level data.

The OpenWeather data models trained on the national-level data performed better during evaluation than those trained on the city-level data, so the national-level results should be con-

2018 Q2

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Delhi', 'Gorakhpur', 'Ghaziabad', 'Mumbai', 'Faridabad', 'Kanpur', 'Nashik', 'Moradabad', 'Lucknow', 'Prayagraj', 'Bareilly', 'Aurangabad', 'Solapur', 'Varanasi', 'Agra', 'Pune', 'Meerut'
Neg. Economic Change	None	'Kozhikode', 'Ludhiana', 'Mangalore', 'Kota', 'Agartala', 'Patna', 'Puducherry', 'Shillong', 'Srinagar', 'Thiruvananthapuram', 'Mysuru', 'Kolkata', 'Jalandhar', 'Kochi', 'Aizwal', 'Amritsar', 'Asansol', 'Belgaum', 'Bengaluru', 'Bhopal', 'Chandigarh', 'Chennai', 'Kohima', 'Coimbatore', 'Gandhinagar', 'Gwalior', 'Hyderabad', 'Imphal', 'Indore', 'Jabalpur', 'Jaipur', 'Jodhpur', 'Dehradun', 'Visakhapatnam'

Fig. 7. Emissions and economic changes for each city from the previous quarter.

sidered more highly than the city-level results. Synthesizing the city- and national-level conclusions suggests that CO is the pollutant most strongly associated with the unemployment rate. In both the city and national cases, the CO coefficient in the linear model is negative, suggesting that CO may be inversely proportional to the unemployment rate. SO_2 may also be inversely proportional and NO_2 may be directly proportional to the unemployment rate, but $PM_{2.5}$ seems to be unrelated.

More tangibly, these conclusions provide an understanding for the effects of industrialized economic growth. While industrialized processes can create jobs (lowering the unemployment rate), they may also increase emissions of some pollutants like SO_2 and CO. With these findings in mind, world leaders can better tackle the SDG's set by the UN. The classification of cities according to their economic and emissions progress and the assessment of the location and quantity of pollutant monitoring stations provided by the visualization presented in this work can also be applied to assist towards these ends.

VI. Appendix

All code used to perform this work can be found here: https://github.com/nsusemiehl/EconClimate

A. Economic/Emissions Change Matrices

Below are the economic and emission change matrices for each quarter other than Q2 of 2018, which can be found in Figure 7.

B. Instructions for Running Interactive Visualization

The interactive visualizations produced in this work were made using Ploty Dash[4]. First, ensure Dash is installed in the Python environment. Next, navigate to the directory ./EconClimate/Plotting and run either "change_dash.py" for the economic and emissions change visualization or "within_dash.py" for the within-city variation visualization using the terminal command "python within_dash.py" or "python change_dash.py" respectively. The output of this command will include the port that Dash is running on. Follow this link to load the visualization. A city can be selected using the dropdown menu at the bottom.

[4]https://dash.plotly.com/

Fig. 8. Emissions and economic changes for the selected city (here, Agartala) over all quarters.

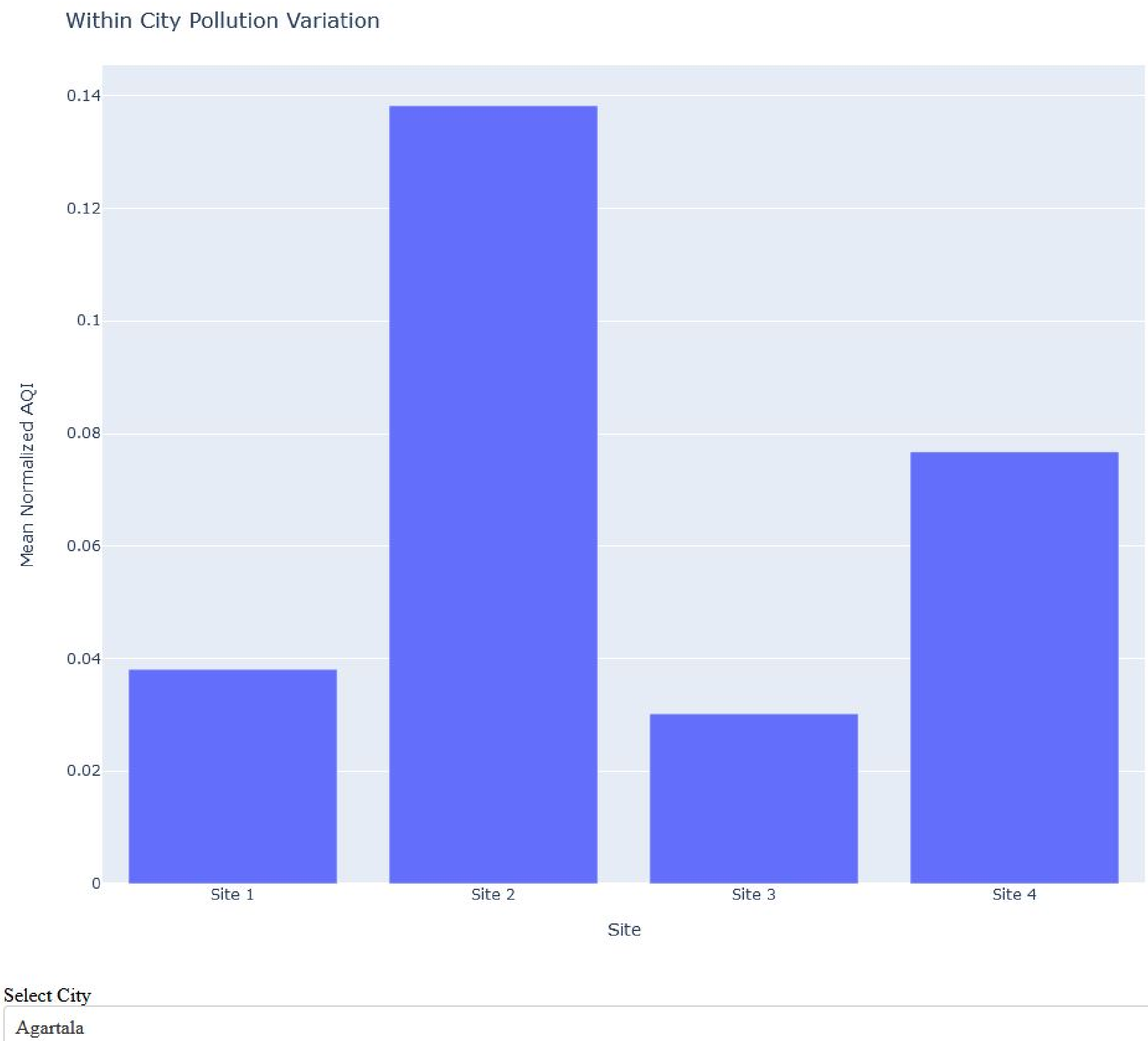

Fig. 9. Average scaled AQI for each monitoring station of the selected city (here, Agartala).

2018 Q3

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Agartala', 'Jabalpur', 'Jaipur', 'Varanasi', 'Jodhpur', 'Kanpur', 'Kochi', 'Kohima', 'Kota', 'Kozhikode', 'Lucknow', 'Ludhiana', 'Meerut', 'Moradabad', 'Patna', 'Prayagraj', 'Puducherry', 'Thiruvananthapuram', 'Indore', 'Imphal', 'Jalandhar', 'Gwalior', 'Agra', 'Aizwal', 'Amritsar', 'Bareilly', 'Bhopal', 'Chandigarh', 'Chennai', 'Visakhapatnam', 'Dehradun', 'Delhi', 'Gorakhpur', 'Faridabad', 'Ghaziabad', 'Coimbatore'
Neg. Economic Change	None	'Kolkata', 'Srinagar', 'Solapur', 'Shillong', 'Pune', 'Gandhinagar', 'Nashik', 'Mysuru', 'Mumbai', 'Aurangabad', 'Hyderabad', 'Belgaum', 'Bengaluru', 'Asansol', 'Mangalore'

Fig. 10. Emissions and economic changes for each city from the previous quarter.

2018 Q4

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Agartala', 'Indore', 'Jabalpur', 'Jaipur', 'Visakhapatnam', 'Jodhpur', 'Kanpur', 'Kohima', 'Imphal', 'Kota', 'Ludhiana', 'Meerut', 'Moradabad', 'Patna', 'Prayagraj', 'Srinagar', 'Varanasi', 'Lucknow', 'Hyderabad', 'Jalandhar', 'Gorakhpur', 'Bareilly', 'Gwalior', 'Bhopal', 'Aizwal', 'Amritsar', 'Chennai', 'Chandigarh', 'Dehradun', 'Agra', 'Faridabad', 'Gandhinagar', 'Ghaziabad', 'Coimbatore'	None
Neg. Economic Change	'Thiruvananthapuram', 'Solapur', 'Shillong', 'Pune', 'Puducherry', 'Asansol', 'Mangalore', 'Mysuru', 'Mumbai', 'Aurangabad', 'Belgaum', 'Bengaluru', 'Kozhikode', 'Kolkata', 'Kochi', 'Delhi', 'Nashik', 'Gangtok'	None

Fig. 11. Emissions and economic changes for each city from the previous quarter.

2019 Q1

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Jalandhar', 'Kanpur', 'Kohima', 'Kolkata', 'Lucknow', 'Ludhiana', 'Meerut', 'Moradabad', 'Mumbai', 'Nashik', 'Patna', 'Prayagraj', 'Puducherry', 'Pune', 'Shillong', 'Solapur', 'Srinagar', 'Varanasi', 'Visakhapatnam', 'Imphal', 'Gangtok', 'Aurangabad', 'Delhi', 'Bareilly', 'Hyderabad', 'Asansol', 'Ghaziabad', 'Chandigarh', 'Gorakhpur', 'Aizwal', 'Agra', 'Amritsar'
Neg. Economic Change	None	'Thiruvananthapuram', 'Belgaum', 'Bengaluru', 'Bhopal', 'Mysuru', 'Gwalior', 'Coimbatore', 'Mangalore', 'Dehradun', 'Kozhikode', 'Kota', 'Faridabad', 'Gandhinagar', 'Kochi', 'Jodhpur', 'Jaipur', 'Jabalpur', 'Indore', 'Chennai', 'Agartala'

Fig. 12. Emissions and economic changes for each city from the previous quarter.

2019 Q2

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Agartala', 'Hyderabad', 'Imphal', 'Nashik', 'Mysuru', 'Jaipur', 'Mumbai', 'Patna', 'Jodhpur', 'Kochi', 'Kohima', 'Moradabad', 'Kota', 'Kozhikode', 'Lucknow', 'Kanpur', 'Gorakhpur', 'Ghaziabad', 'Prayagraj', 'Agra', 'Aizwal', 'Varanasi', 'Thiruvananthapuram', 'Aurangabad', 'Bareilly', 'Belgaum', 'Bengaluru', 'Solapur', 'Shillong', 'Pune', 'Dehradun', 'Faridabad', 'Meerut', 'Mangalore'
Neg. Economic Change	None	'Srinagar', 'Puducherry', 'Jalandhar', 'Kolkata', 'Visakhapatnam', 'Jabalpur', 'Indore', 'Gwalior', 'Gandhinagar', 'Delhi', 'Coimbatore', 'Chennai', 'Chandigarh', 'Bhopal', 'Asansol', 'Amritsar', 'Ludhiana', 'Gangtok'

Fig. 13. Emissions and economic changes for each city from the previous quarter.

2019 Q3

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Faridabad', 'Hyderabad', 'Nashik', 'Patna', 'Gandhinagar', 'Kohima', 'Delhi', 'Dehradun', 'Coimbatore', 'Imphal', 'Chennai', 'Puducherry', 'Pune', 'Solapur', 'Aurangabad', 'Srinagar', 'Aizwal', 'Visakhapatnam', 'Mumbai'
Neg. Economic Change	None	'Mangalore', 'Ludhiana', 'Meerut', 'Lucknow', 'Agartala', 'Mysuru', 'Prayagraj', 'Shillong', 'Thiruvananthapuram', 'Varanasi', 'Moradabad', 'Kozhikode', 'Jalandhar', 'Kolkata', 'Agra', 'Amritsar', 'Asansol', 'Bareilly', 'Belgaum', 'Bengaluru', 'Bhopal', 'Chandigarh', 'Ghaziabad', 'Gorakhpur', 'Gwalior', 'Indore', 'Jabalpur', 'Jaipur', 'Jodhpur', 'Kanpur', 'Kochi', 'Kota', 'Gangtok'

Fig. 14. Emissions and economic changes for each city from the previous quarter.

2019 Q4

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Gangtok', 'Kota', 'Kolkata', 'Kohima', 'Kochi', 'Mangalore', 'Jodhpur', 'Visakhapatnam', 'Jaipur', 'Imphal', 'Hyderabad', 'Mumbai', 'Mysuru', 'Nashik', 'Kozhikode', 'Gandhinagar', 'Patna', 'Aizwal', 'Thiruvananthapuram', 'Asansol', 'Aurangabad', 'Belgaum', 'Faridabad', 'Bengaluru', 'Solapur', 'Shillong', 'Pune'	None
Neg. Economic Change	'Moradabad', 'Srinagar', 'Puducherry', 'Prayagraj', 'Varanasi', 'Meerut', 'Ludhiana', 'Agartala', 'Kanpur', 'Agra', 'Amritsar', 'Bareilly', 'Bhopal', 'Chandigarh', 'Chennai', 'Lucknow', 'Coimbatore', 'Delhi', 'Ghaziabad', 'Gorakhpur', 'Gwalior', 'Indore', 'Jabalpur', 'Dehradun', 'Jalandhar'	None

Fig. 15. Emissions and economic changes for each city from the previous quarter.

2020 Q1

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Agartala', 'Srinagar', 'Mysuru', 'Mangalore', 'Ludhiana', 'Kohima', 'Visakhapatnam', 'Imphal', 'Hyderabad', 'Gandhinagar', 'Faridabad', 'Delhi', 'Dehradun', 'Jalandhar', 'Chennai', 'Aizwal', 'Amritsar', 'Coimbatore', 'Belgaum', 'Bengaluru', 'Gangtok', 'Chandigarh'
Neg. Economic Change	None	'Gorakhpur', 'Varanasi', 'Thiruvananthapuram', 'Agra', 'Solapur', 'Shillong', 'Pune', 'Puducherry', 'Prayagraj', 'Patna', 'Nashik', 'Mumbai', 'Moradabad', 'Gwalior', 'Meerut', 'Asansol', 'Lucknow', 'Ghaziabad', 'Kota', 'Kolkata', 'Aurangabad', 'Kochi', 'Kanpur', 'Jodhpur', 'Bareilly', 'Jaipur', 'Jabalpur', 'Indore', 'Bhopal', 'Kozhikode'

Fig. 16. Emissions and economic changes for each city from the previous quarter.

2020 Q2

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Jalandhar', 'Jodhpur', 'Kanpur', 'Kochi', 'Kohima', 'Kolkata', 'Kota', 'Kozhikode', 'Lucknow', 'Ludhiana', 'Visakhapatnam', 'Mangalore', 'Moradabad', 'Mumbai', 'Mysuru', 'Nashik', 'Patna', 'Prayagraj', 'Puducherry', 'Pune', 'Shillong', 'Meerut', 'Jaipur', 'Jabalpur', 'Indore', 'Agra', 'Aizwal', 'Amritsar', 'Asansol', 'Aurangabad', 'Bareilly', 'Belgaum', 'Bengaluru', 'Bhopal', 'Chandigarh', 'Chennai', 'Coimbatore', 'Varanasi', 'Delhi', 'Faridabad', 'Gandhinagar', 'Ghaziabad', 'Gorakhpur', 'Gwalior', 'Hyderabad', 'Imphal', 'Solapur', 'Thiruvananthapuram'
Neg. Economic Change	None	'Agartala', 'Dehradun', 'Srinagar', 'Gangtok'

Fig. 17. Emissions and economic changes for each city from the previous quarter.

2020 Q3

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Jaipur', 'Kota', 'Dehradun', 'Jodhpur'	None
Neg. Economic Change	'Kanpur', 'Kochi', 'Kolkata', 'Kozhikode', 'Lucknow', 'Ludhiana', 'Mangalore', 'Meerut', 'Agartala', 'Mysuru', 'Nashik', 'Patna', 'Prayagraj', 'Puducherry', 'Pune', 'Shillong', 'Solapur', 'Srinagar', 'Varanasi', 'Mumbai', 'Moradabad', 'Jalandhar', 'Jabalpur', 'Agra', 'Aizwal', 'Amritsar', 'Asansol', 'Aurangabad', 'Bareilly', 'Belgaum', 'Bengaluru', 'Bhopal', 'Chandigarh', 'Visakhapatnam', 'Chennai', 'Delhi', 'Faridabad', 'Gandhinagar', 'Ghaziabad', 'Gorakhpur', 'Gwalior', 'Hyderabad', 'Imphal', 'Indore', 'Coimbatore', 'Gangtok'	'Kohima', 'Thiruvananthapuram'

Fig. 18. Emissions and economic changes for each city from the previous quarter.

2020 Q4

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Faridabad', 'Kohima', 'Lucknow', 'Kanpur', 'Jodhpur', 'Jaipur', 'Meerut', 'Imphal', 'Moradabad', 'Gorakhpur', 'Ghaziabad', 'Gandhinagar', 'Kota', 'Bareilly', 'Agra', 'Aizwal', 'Prayagraj', 'Varanasi', 'Shillong', 'Srinagar'	None
Neg. Economic Change	'Thiruvananthapuram', 'Ludhiana', 'Mangalore', 'Nashik', 'Solapur', 'Mumbai', 'Pune', 'Puducherry', 'Kozhikode', 'Mysuru', 'Patna', 'Agartala', 'Jalandhar', 'Kochi', 'Amritsar', 'Asansol', 'Aurangabad', 'Belgaum', 'Bengaluru', 'Bhopal', 'Chandigarh', 'Kolkata', 'Chennai', 'Dehradun', 'Delhi', 'Gwalior', 'Hyderabad', 'Indore', 'Jabalpur', 'Visakhapatnam', 'Coimbatore', 'Gangtok'	None

Fig. 19. Emissions and economic changes for each city from the previous quarter.

2021 Q1

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Agartala', 'Pune', 'Patna', 'Nashik', 'Mysuru', 'Mumbai', 'Ludhiana', 'Hyderabad', 'Delhi', 'Chennai', 'Jalandhar', 'Belgaum', 'Aurangabad', 'Amritsar', 'Chandigarh'	'Srinagar', 'Solapur', 'Puducherry', 'Coimbatore', 'Mangalore', 'Bengaluru', 'Gangtok'
Neg. Economic Change	'Thiruvananthapuram', 'Prayagraj', 'Kozhikode', 'Aizwal', 'Varanasi', 'Asansol', 'Meerut', 'Lucknow', 'Bhopal', 'Kohima', 'Dehradun', 'Faridabad', 'Gandhinagar', 'Ghaziabad', 'Gorakhpur', 'Kolkata', 'Indore', 'Imphal', 'Visakhapatnam', 'Jodhpur', 'Kochi'	'Jaipur', 'Moradabad', 'Gwalior', 'Agra', 'Shillong', 'Bareilly', 'Kanpur', 'Jabalpur', 'Kota'

Fig. 20. Emissions and economic changes for each city from the previous quarter.

2021 Q2

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Puducherry', 'Asansol', 'Bengaluru'	'Agartala', 'Jaipur', 'Visakhapatnam', 'Jodhpur', 'Kanpur', 'Kochi', 'Kohima', 'Kolkata', 'Kota', 'Kozhikode', 'Lucknow', 'Prayagraj', 'Mangalore', 'Meerut', 'Moradabad', 'Mumbai', 'Pune', 'Imphal', 'Hyderabad', 'Shillong', 'Agra', 'Aizwal', 'Varanasi', 'Aurangabad', 'Bareilly', 'Belgaum', 'Thiruvananthapuram', 'Mysuru', 'Nashik', 'Coimbatore', 'Dehradun', 'Delhi', 'Faridabad', 'Solapur', 'Ghaziabad', 'Gorakhpur', 'Chennai'
Neg. Economic Change	None	'Srinagar', 'Jalandhar', 'Ludhiana', 'Jabalpur', 'Indore', 'Gwalior', 'Gandhinagar', 'Chandigarh', 'Bhopal', 'Amritsar', 'Patna', 'Gangtok'

Fig. 21. Emissions and economic changes for each city from the previous quarter.

2021 Q3

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Indore', 'Patna', 'Bhopal', 'Jabalpur', 'Gwalior', 'Srinagar'
Neg. Economic Change	'Shillong', 'Puducherry', 'Gandhinagar', 'Visakhapatnam', 'Chandigarh', 'Chennai', 'Kota'	'Kozhikode', 'Ludhiana', 'Mangalore', 'Lucknow', 'Agartala', 'Mysuru', 'Moradabad', 'Mumbai', 'Nashik', 'Prayagraj', 'Pune', 'Solapur', 'Thiruvananthapuram', 'Varanasi', 'Meerut', 'Kolkata', 'Jalandhar', 'Kochi', 'Agra', 'Aizwal', 'Amritsar', 'Asansol', 'Aurangabad', 'Bareilly', 'Belgaum', 'Bengaluru', 'Coimbatore', 'Dehradun', 'Delhi', 'Faridabad', 'Ghaziabad', 'Gorakhpur', 'Hyderabad', 'Imphal', 'Jaipur', 'Jodhpur', 'Kanpur', 'Kohima', 'Gangtok'

Fig. 22. Emissions and economic changes for each city from the previous quarter.

2021 Q4

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Jalandhar', 'Srinagar', 'Shillong', 'Patna', 'Ludhiana', 'Kota', 'Kohima', 'Jodhpur', 'Jaipur', 'Imphal', 'Gandhinagar', 'Faridabad', 'Gangtok', 'Chandigarh', 'Aizwal', 'Amritsar'	None
Neg. Economic Change	'Dehradun', 'Mangalore', 'Meerut', 'Moradabad', 'Mumbai', 'Mysuru', 'Nashik', 'Prayagraj', 'Pune', 'Solapur', 'Agra', 'Thiruvananthapuram', 'Varanasi', 'Lucknow', 'Coimbatore', 'Kozhikode', 'Kolkata', 'Delhi', 'Chennai', 'Ghaziabad', 'Gorakhpur', 'Gwalior', 'Hyderabad', 'Bhopal', 'Aurangabad', 'Indore', 'Bengaluru', 'Belgaum', 'Kanpur', 'Kochi', 'Bareilly', 'Jabalpur', 'Agartala'	'Puducherry', 'Visakhapatnam', 'Asansol'

Fig. 23. Emissions and economic changes for each city from the previous quarter.

2022 Q1

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Agartala', 'Mysuru', 'Hyderabad', 'Bengaluru', 'Visakhapatnam'	'Jalandhar', 'Jabalpur', 'Indore', 'Mangalore', 'Delhi', 'Dehradun', 'Gwalior', 'Ludhiana', 'Chandigarh', 'Bhopal', 'Amritsar', 'Belgaum'
Neg. Economic Change	'Mumbai', 'Kohima', 'Asansol', 'Srinagar', 'Thiruvananthapuram', 'Aizwal', 'Pune', 'Gandhinagar', 'Shillong', 'Imphal'	'Puducherry', 'Prayagraj', 'Patna', 'Nashik', 'Agra', 'Moradabad', 'Meerut', 'Solapur', 'Aurangabad', 'Kozhikode', 'Kota', 'Kolkata', 'Kochi', 'Jodhpur', 'Varanasi', 'Jaipur', 'Gorakhpur', 'Ghaziabad', 'Faridabad', 'Coimbatore', 'Chennai', 'Bareilly', 'Lucknow', 'Kanpur'

Fig. 24. Emissions and economic changes for each city from the previous quarter.

2022 Q2

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	None	'Agartala', 'Jaipur', 'Jodhpur', 'Kochi', 'Gandhinagar', 'Faridabad', 'Delhi', 'Dehradun', 'Kohima', 'Kota', 'Kozhikode', 'Mangalore', 'Bengaluru', 'Belgaum', 'Mysuru', 'Patna', 'Shillong', 'Aizwal', 'Thiruvananthapuram', 'Hyderabad', 'Imphal'
Neg. Economic Change	'Puducherry', 'Gangtok'	'Jabalpur', 'Varanasi', 'Agra', 'Srinagar', 'Solapur', 'Amritsar', 'Pune', 'Asansol', 'Prayagraj', 'Aurangabad', 'Nashik', 'Bareilly', 'Mumbai', 'Indore', 'Moradabad', 'Bhopal', 'Ludhiana', 'Lucknow', 'Chandigarh', 'Chennai', 'Kolkata', 'Coimbatore', 'Ghaziabad', 'Kanpur', 'Gorakhpur', 'Visakhapatnam', 'Gwalior', 'Meerut', 'Jalandhar'

Fig. 25. Emissions and economic changes for each city from the previous quarter.

2022 Q3

	Pos. Emissions Change	Neg. Emissions Change
Pos. Economic Change	'Visakhapatnam', 'Gandhinagar'	'Ghaziabad', 'Moradabad', 'Kolkata', 'Kanpur', 'Jabalpur', 'Indore', 'Patna', 'Prayagraj', 'Gwalior', 'Gorakhpur', 'Lucknow', 'Meerut', 'Coimbatore', 'Chennai', 'Agra', 'Bhopal', 'Bareilly', 'Srinagar', 'Varanasi', 'Asansol'
Neg. Economic Change	'Jaipur', 'Shillong', 'Kota', 'Delhi'	'Mumbai', 'Puducherry', 'Thiruvananthapuram', 'Nashik', 'Solapur', 'Pune', 'Mysuru', 'Agartala', 'Jodhpur', 'Ludhiana', 'Kozhikode', 'Kohima', 'Kochi', 'Jalandhar', 'Imphal', 'Hyderabad', 'Faridabad', 'Chandigarh', 'Bengaluru', 'Belgaum', 'Aurangabad', 'Amritsar', 'Aizwal', 'Mangalore', 'Gangtok'

Fig. 26. Emissions and economic changes for each city from the previous quarter.