

# An Empirical Study of Existing and Prospective Air Pollutant level in UK Cities, and its Relationship with Weather

Shubham Rajesh

23/08/2022

## Abstract

**Context :** Over the past few years, air pollution in urban areas has become significantly worse. With expanding industrialisation and the release of more harmful compounds into the air, air pollution is increasing at an alarming rate. Forecasting time series data is challenging, as is identifying the factors that influence it. Understanding the air quality data may aid in forecasting the air quality, also allowing the air quality regulatory bodies to make crucial decisions promptly.

**Objectives :** The purpose of this study is to compare, visualise, and forecast air quality of Birmingham, Manchester, and Newcastle upon Tyne by using publicly available Air quality data. This study includes primarily suspended nitrogen dioxide  $NO_2$  and particulate matter  $PM_{10}$ , in compliance with the Essential Variables of the UK's Department of the Environment (EVs). Additionally, understand the correlation between atmospheric temperature and air quality.

**Method :** Data is mined, wrangled, and visually inspected with the help of Python, R, and PowerBI. In this research, Air quality data for all three cities will be compared from January to December 2021. PowerBI was employed to do the geographic visual analysis. The time series data were used to predict  $NO_2$  and  $PM_{10}$  concentrations using the FB Prophet model. R and PowerBI were utilised to understand the correlation of atmospheric temperature with  $NO_2$  and  $PM_{10}$  concentration.

**Results :** Visual and statistical analysis reveals that  $NO_2$  and  $PM_{10}$  concentrations are greater in Manchester City. While the average  $NO_2$  concentration in Birmingham is lower, and the average  $PM_{10}$  concentration in Newcastle is the lowest of the three, the presence of numerous outlier values indicates that significant fluctuations have occurred, perhaps owing to external sources. The  $NO_2$  and  $PM_{10}$  concentration hotspots of all three cities were identified using the geospatial visual assessment. FB Prophet model suggested the presence of trend and seasonality patterns, the model itself forecasted values with non-significant errors, ranging between 10 and 20% of MAPE on average. The Correlation analysis suggest positive correlation between  $NO_2$  and Atmospheric temperature.

**Novelty :** The work flow followed in this research, used combination of PowerBI, Time series FB Prophet and Correlation to identify the new insights which can be applied to uncover new evidence and discoveries for additional Air Pollutants variables and for different cities.

## Introduction

Air pollution is the major environmental concern to public health in the United Kingdom. The yearly mortality from air pollution caused to humans in the United Kingdom is roughly equivalent to 28,000 and 36,000 fatalities annually ('Air pollution', 2022) [1]. Air pollution can induce and exacerbate health impacts in all people, especially the most vulnerable members of the population. Long-term exposure to air pollution can result in chronic disease such as cardiovascular and respiratory ailments, as well as lung cancer, and a shortened lifespan. Short-term increases in air pollution levels are associated with a variety of adverse health outcomes, including impairment of lung function, asthma exacerbation, increases in respiratory and cardiovascular hospital admissions, and death ('Air pollution', 2022) [1]. The illness burden attributed to air pollution is now believed to be comparable to other significant worldwide health concerns, such as an unhealthy diet and cigarette usage, and air pollution is widely acknowledged as the widespread environmental danger to human health (World Health Organization, 2021) [2].

Common causes of air pollution include household combustion devices, motor vehicles, industrial operations, and forest fires. Particulate matter, carbon monoxide, ozone, nitrogen dioxide, and sulphur dioxide are pollutants that pose a significant threat to public health ('Air pollution WHO', 2022) [3]. In the research by (Curtis et al., 2006), the major focus is kept of two most dangerous pollutants Nitrogen dioxide and Particulate matter 10. Nitrogen dioxides are primarily created by industrial/vehicle combustion and n fertilizers oxidation. When reactive oxygen-containing molecules combine with nitrogen in the atmosphere, nitrogen dioxides can be produced (Curtis et al., 2006) [4].  $NO_2$  might result in a variety of negative effects on human health, the environment, and biological ecosystems, including acid rain, ozone depletion, and global warming. Epidemiological studies indicate that exposure to high levels of  $NO_2$  in the air as a pollutant increases the risk of lung cancer by around 5-7%. (ex-smokers and non-smokers) (Shams et al., 2021) [5].

According to recent research (World Health Organization, 2021) [2], air pollution (PM and O<sub>3</sub>) was implicated for 4.9 million fatalities in 2017 in both indoor and outdoor locations. Particulate matter (PM) air pollution is one of the leading causes of mortality globally, with both short- and long-term harmful effects on exposure having been evidenced (Stafoggia et al., 2019) [6]. Particulate Matter 10 ( $PM_{10}$ ) comprises any fine airborne solid or liquid particle having an aerodynamic diameter less than 10 micrometer. PM pollution may be caused by both natural and manmade causes, including the combustion of fossil fuels in the transportation sector (Garcia-Franco, 2020) [7].

Particulate matter 10 ( $PM_{10}$ ) has a considerable detrimental influence on human health; thus, predicting particulate matter has become one of the world's top priorities, as it plays a crucial role in the successful management of air pollution. Currently, the application of artificial intelligence technology is expanding rapidly, hence enhancing air quality (Rani Samal et al., 2020) [8].

Rapidly deteriorating air quality and extreme climate changes, such as heatwaves, are harming an increasing number of people throughout the world. Heatwaves have long been recognised as a major source of air pollution, resulting in a variety of adverse health, environmental, and economic effects (Cheval et al., 2009), (Garcia-Herrera et al., 2010) [9] [10]. Ozone ( $O_3$ ), particle matter (PM), and nitrogen dioxide ( $NO_2$ ) are air pollutants whose concentrations and effects are known to be impacted by heatwaves. When released into the atmosphere from a range of natural and manmade sources, these pollutants pose a significant hazard to human health (World Health Organization, 2010) [11]. As, the  $NO_2$  and  $PM_{10}$  had such a large impact on human life, so it became the air pollutants which were targeted in this research.

In past few years, forecasting models for air quality have been progressively enhanced and expanded. So that numerous methodologies and approaches for time series data may be utilised to assist focused control of air pollution and avert severe pollution occurrences (Liu et al., 2021) [12]. With the advancement of artificial intelligence and machine learning in past few years, forecasting models based on artificial intelligence are gaining popularity and garnering increasing attention. When dealing with nonlinearities and interacting linkages in air pollution modelling, intelligent predictors can reach higher levels of accuracy in their techniques (Zhang et al., 2012) [13]. The primary advantages of the models are that they can handle nonlinear aspects and solve issues using enormous amounts of data. The application does not need an in-depth comprehension

of the dynamic and chemical processes between air pollution levels and other atmospheric factors (Cabaneros et al., 2019) [14].

Many well-established businesses have also done extensive study and created their own prediction algorithms in response to the newly discovered possibilities in time-series forecasting. For instance, Facebook, the most popular social media platform, has just released a new time-series prediction algorithm called FBProphet. In the field of weather forecasting, the model's capacity to accommodate for seasonal shifts and the impact of holidays with unpredictable schedules is its most appealing characteristic (Verma et al., 2021) [15].

This research aims to comprehend the existing air pollution level and utilise the FB Prophet model to predict the prospective  $NO_2$  and  $PM_{10}$  levels based on the present data. The FBProphet model is selected because to its unique ability to categorize periodic changes and accommodate for the impact of holidays with irregular schedules, which is particularly valuable in the field of meteorology. This paper gives comprehensive guidance on data processing and model training for air quality time-series data for cities in the United Kingdom (Verma et al., 2021) [15].

The key contributions of this paper are to:

- Perform a comparative study of the existing air quality in Birmingham, Manchester and Newcastle, identify the fluctuations and finding.
- Conduct an empirical investigation with the objective of identifying seasonality trends more precisely and extending the analysis to forecast air quality.
- Understand the Correlation between atmospheric temperature and air quality.

## Literature Review

### Related work:

Several machine learning algorithms have been presented in recent years for handling air pollution prediction challenges. Important works in the subject will be presented and evaluated in this section. The classification of air quality index and its impact on health were investigated in the research by (Gore & Deshpande, 2017). The authors classified their data using the Decision tree and Naive Bayes J48 methods. The collected findings demonstrated that the accuracy of the decision tree algorithm is 91.9978 %. However, the drawback identified in this study, including the fact that the dataset utilised was inadequate. In addition, decision tree approaches are not expected to perform poorly with continuous variables and might have overfitting difficulties [16].

(Nagendra et al., 2007) review the next study, which evaluates the Air Quality around Bangalore, India's traffic junctions from 1997 to 2005. According to the study, the AQIs for the criteria pollutants  $SO_2$ ,  $NO_x$ ,  $RSPM$ , and SPM are calculated using the technique recommended by the EPA. The result reveals that air pollution levels at all three crossings (i.e., AQM stations) between 1997 and 2005 may be classed as "excellent" or "moderate" for  $SO_2$  and NOx, respectively. The AQI value for RSPM concentration was "good" or "moderate" 94% of the time between 1999 and 2005, "bad" 5% of the time, and "very poor" and "severe" 1% of the time. From 1997 to 2005, 91% of the AQI values for SPM concentrations were classed as "excellent" or "moderate," 7% as "poor" and the remaining 2% as "very poor" or "severe." The annual AQI scores for RSPM and SPM are declining, but the prevalence of "good" and "moderate" levels is increasing [17].

The subject of (Naseem et al., 2018) paper is to predict the future values of air quality parameters. The authors had employed the ARIMA model for the forecasts, as it is one of the most generic classes of models for time series analysis and forecasting. The air pollution data has been evaluated using time series analysis, and the anticipated and observed values have been compared. Using the ARIMA model yielded good and dependable results, as demonstrated by the findings. Mean square error, mean absolute error, and root mean square error have been calculated to evaluate the performance of the model [18].

(Athira.V et al., 2018) presented three deep learning models for predicting future  $PM_{10}$  concentration values. Gated Recurrent Unit (GRU), Recurrent Neural Networks (RNN), and Long Short-term Memory are the models examined (LSTM). After studying the variants in RNN model topologies, the most appropriate architecture was selected. The result revealed that the GRU network is more efficient than the other two networks. A trend was also established for  $PM_{10}$  concentration. Deep learning can provide accurate air quality forecasts by extracting air quality characteristics without any prior knowledge.[19]

(Kingsy et al., 2016) suggested research for identifying the air quality index. In their work, the authors implemented the K-means technique; however, the dataset utilised in this study was restricted. When attempting to predict future values, K-means algorithms have additional weaknesses [20]. In a separate study, B.K.Jha and S.Pande analysed the FB Prophet Additive and ARIMA sales models. Using the supermarket furniture information, they determined that the FB Prophet forecast was extremely close to actual data [21].

(Kumar Jha & Pande, 2021) have analysed and visualized urban pollution based on the geographical regions studied. They investigated data collected in Tehran between 2009 and 2013 using Apache Spark. They have discovered that the Naive Bayes approach is more accurate than other machine learning algorithms in classifying unknown air quality classes. However, the technique is not suitable for predicting actual time series [22].

(Zhu et al., 2018), discussed the forecasting of air pollutants including ozone, particle matter ( $PM_{2.5}$ ), and sulphur dioxide. They apply optimization and regularisation techniques to forecast the next day's air pollution levels. The values have been predicted using data from two stations. One station forecasts  $O_3$  and  $SO_2$  levels, while the other reports  $O_3$  and  $PM_{2.5}$  levels. They have modelled the data based on similarity and grouped them using liner regression. Root-mean-squared error (RMSE) was the criterion for assessment. The limits of this approach stem from the inability of Linear Regression models to predict or manage unexpected occurrences. In addition, just two stations' data are used in this study, limiting its generalizability [23].

Traditionally, time-series data were projected using typical regressive models like as AR, MA, ARMA, and ARIMA (Samira et al., 2014) [24]. These models are highly prevalent in economic and financial data forecasting. The models were not intended to study nonlinear behaviour between variables, nor did they account for the fact that conditional covariance implies variance change over time when the data display conditional covariance. However, this difficulty might be solved by combining it with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH), although optimising its parameters is extremely challenging (Mohamadi et al., 2019) [25].

## About Time series Data Mining

Time series is an interdisciplinary issue of significant practical importance. Because it permits discovering, with a certain margin of error, the future values of a series based on its historical values. Numerous successful applications in several domains, including economics, finance, and hydrology, have been documented in the pertinent literature (Tealab, 2018) [26]. Mining time series data originates from the requirement for individuals to visualise data models in accordance with their skills. People rely on intricate strategies to accomplish these jobs. The primary activities associated with time series include content-based querying, anomaly detection, pattern recognition, prediction, clustering, classification, and segmentation. A huge number of decision-making issues in the natural sciences and social sciences cannot be separated from forecasting. Forecasting is the foundation of decision-making (Zong et al. 2012) [27].

## About FB-Prophet model

Fbprophet is an open-source programme developed by the core data science team at Facebook for autonomous forecasting of univariate time series data using an additive model. Fbprophet works best with data that has substantial seasonal impacts and historical data that spans many seasons (minimum of 1 month). This model is resistant to missing data and is able to handle outliers and variations in trends with ease, allowing

it to produce accurate predictions. Additionally, the model might incorporate seasonality and the holiday season (Verma et al., 2021) [15]. In addition, this model is simple to apply, and anybody who lacks knowledge in forecasting may use it to produce relevant and reliable forecasts, as it only requires the adjustment of a few simple parameters. Three models comprise the Fbprophet: trend, seasonality, and holiday. Similar to additive models, Prophet attempts to fit several linear and nonlinear functions of time as components, with time as the regressor. The modelling equation is as follows:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

where,

- $g(t)$ : piece-wise linear or logistic growth curve for modelling non-periodic changes (i.e., growth over time)
  - $s(t)$ : periodic changes (e.g. weekly/yearly seasonality)
  - $h(t)$ : effects of holidays (user provided) with irregular schedules
  - $\epsilon_t$ : error term accounts for idiosyncratic changes that is not accommodated by the model
- 1) Trend. Over the trending or non-periodic portion of the time series, a piecewise linear curve or a saturation growth model is fitted to describe the trend. In order to make sure the model isn't too vulnerable to outliers or missing data; we do a linear fitting.
  - 2) Seasonality: Because of the seasonality factor  $s(t)$ , the model can account for variations that occur on a daily, weekly, monthly, and annual basis. Using the Fourier series, prophet is able to fit the impacts of seasonality and make predictions about such effects in the future. In order to get a rough estimate of seasonality, the formula:

$$s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P}))$$

where  $P$  is the duration of the time series (e.g.,  $P=365.25$  for annual data or  $P=7$  for weekly data when time is measured in days). The parameters  $a_1 \dots b_N$  must be calculated in light of the value of  $N$ .  $N$  specifies if high frequency variations may be modelled for fitting seasonal patterns that vary rapidly, or whether it is noise causing over-fitting, in which case this parameter should be set to a lower value.

- 3) Holiday: Holidays, such as Diwali, New Year's, and Christmas, generate unanticipated changes in predictions, which can lead to an increase in air pollution owing to the burning of firecrackers. The component  $h(t)$  contains these erratic schedules. Thus, prophet enables the inculcation of anomalies using a particular list of past and future occurrences; these days are examined independently, and extra parameters are added to the model to accommodate them.

## About correlation between the Air quality and temperature.

(Jayamurugan et al., 2013) have discovered a number of correlations between temperature and relative humidity and the concentration of pollutants throughout the summer, pre-monsoon, monsoon, and post-monsoon seasons of 2010-2011 [28]. The relevance of surface-based air temperature inversions in increasing air pollution concentrations, particularly in the winter, cannot be overstated. Similar to SBI, air pollution has a tight link with meteorological conditions, such as air temperature, relative humidity, wind speed, and sun radiation (Li et al., 2021) [29].

(Theoharatos et al., 2010) and (McMichael et al., 2003) discovered a substantial association between heat-waves and average hourly concentrations of O<sub>3</sub>, NO<sub>2</sub>, and SO<sub>2</sub>. Similarly, an increase in urban heat islands

(UHIs) and air pollution in London, England can be attributed to a combination of rising air pollution levels and high temperatures [30] [31].

Another research evaluated the correlation between temperature and air pollution in Birmingham during heatwaves. Overall, the results reveal a linear association and a positive correlation between temperature and air pollution during heatwaves, with the relationship between temperature and air pollution being more pronounced during heatwaves with high intensity and prolonged duration (Kalisa et al., 2018) [32].

## Accuracy Assessment Metrics

The model's performance is assessed by comparing the forecasted values with the test data. So, error is calculated by differencing the actual value ( $a_t$ ) with the forecasted value ( $f_t$ ). Then using these errors terms, we calculate the accuracy of the models (Verma et al., 2021) [15]. The metrics used are explained as follows:

- 1) Mean Absolute Error (MAE): It is the average of the absolute error values. In this, the absolute values are considered so that the positive and negative errors cancel out each other hence losing the error value.

$$MAE = \frac{1}{n} \sum |e_t|$$

where  $e_t$  is the error term.

- 2) Mean Absolute Percentage Error (MAPE): As its name implies, it represents the mean of absolute error values. In addition to MAE, the percentage is determined in order to eliminate ambiguity and standardise the data [9]. For instance, the inaccuracy of 1 is large for ten items but low for one thousand values.

$$MAPE = \frac{1}{n} \sum \frac{|e_t|}{a_t} * 100$$

where  $(e_t)$  is the error term and  $(a_t)$  is the actual value.

- 3) Root Mean Square Error (RMSE): Root mean square error may be determined using the formula.

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2}$$

where  $e_t$  is the error term.

## Methodology

### a. Data Collection

For this study, data about air pollutants in Newcastle and Birmingham were obtained from Urban observatory, the largest sensor deployment in the United Kingdom and the source of the most open environmental monitoring data. There are six urban observatories in the United Kingdom, including ones in Newcastle, Sheffield, Bristol, Birmingham, Manchester, and Cranfield [33]. The Newcastle Urban Observatory (UO) is

a multimillion-pound investment in ubiquitous urban sensing that delivers one of the most extensive collections of real-time environmental data. This research institution, manned mostly by research software engineers, provides open data to community organisations, local and national governments, and research projects ranging from cyber-security to flood predictions [34]. UK Air Information Resource was consulted for information on Manchester sites. Ricardo Energy & Environment hosts and maintains UK-AIR on behalf of the United Kingdom Department for Environment, Food & Rural Affairs and the Devolved Administrations. The United Kingdom has been separated into non-agglomeration zones and agglomeration zones for the purpose of monitoring and reporting air pollution via UK-AIR. There are fifteen non-agglomeration zones identified, which correspond to the Government Offices for the Regions in England and the Scottish Government, Welsh Government, and Department of the Environment in Northern Ireland.

**Study's location:** Newcastle urban observatory collects data from around 137 distinct places in the city, whereas Birmingham urban observatory collects data from just 25 distinct areas in the city. On the UK-AIR website, data was only accessible for four locations in Manchester. All locations are depicted in [figure-1], [figure-3], and [figure-5] below. All of these locations collected data on air pollutants including Nitrogen dioxide  $NO_2$ , Particulate matter 10 microns or smaller  $PM_{10}$ , sulfur dioxide  $SO_2$ , Nitrogen oxide  $NO$ , Particulate matter 1 microns or smaller  $PM_1$ , Particulate matter 2.5 microns or smaller  $PM_{2.5}$ , Particulate matter 4 microns or smaller  $PM_4$ , Carbon monoxide  $CO_3$ , and Ozone  $O_3$ . Since sensors record values every second, the amount of data for Newcastle was enormous. In contrast, sensors in Birmingham gathered data just once every half-hour, whilst sensors in Manchester collected data once every hour. This study focuses primarily on comparing and projecting  $NO_2$  and  $PM_{10}$  data. In order to maintain the objectivity of the analysis, hourly data from January 1, 2021 through December 31, 2021 was considered.

## Data Understanding and Data Preparation

The information for Birmingham was acquired manually from the Birmingham Urban Observatory website. The data was downloaded per location and consisted of the following variables: ‘index,’ ‘Timestamp,’ ‘Value,’ ‘has beginning,’ ‘has end,’ ‘duration,’ ‘longitude,’ ‘latitude,’ and ‘height (m)’. And while the data was downloaded according to location, the timestamp was used as the main key to combine the data.

Whereas the Manchester data was manually downloaded from the UK\_AIR webpage. The data was downloaded per location and each file had following variables: ‘Date’, ‘Time’, ‘Value’, ‘Status’. Here, the time and date had to concatenated to make single value names timestamp. And as the data was downloaded according to the location, the latitude and longitude values were assumed according to the Location name. E.g., Manchester Piccadilly Station- Latitude: 53.478691 and Longitude: -2.22895. Then all the location files were joined using the timestamp as primary key.

For Newcastle data was manually downloaded from the Newcastle Urban Observatory webpage. The data downloaded had following variables: ‘Sensor Name’, ‘Variable’, ‘Units’, ‘Timestamp’, ‘Value’, ‘Location (WKT)’, ‘Ground Height Above Sea Level’, ‘Sensor Height Above Ground’, ‘Broker Name’, ‘Third Party’, ‘Sensor Centroid Longitude’, ‘Sensor Centroid Latitude’, ‘Raw ID’. Data was standardised by keeping data per hour and removing all other columns.

After completing the data preparation stage, all data sets after this step contained five variables.

Table 1: Dataset description of Newcastle

Variable Name	Class	Information
Timestamp	Character	Time series of the data recorded
Variable	Numeric	Air pollutant name
Value	Numeric	Value of the sensor recorded
Latitude	Numeric	Latitude location of sensor
Longitude	Numeric	Longitude location of sensor

### **c. Data Cleaning and Exploratory data analysis**

The data was then cleansed in RStudio by handling missing values and finding and eliminating outliers. Box-plot and histogram were used to identify outliers, while the Interquartile Rule was utilised to eliminate them. To have a better understanding of data, RStudio was utilised. Using PowerBI, the relation between variables was visualised. When month-by-month comparisons of  $PM_{10}$  and  $NO_2$  measurements were undertaken for all three cities, crucial insights were discovered.

### **d. Geospatial Representation**

The data collected from Urban observatory and UK-AIR included sensor location information, therefore Longitude and Latitude values were utilised to depict the values on a map of the United Kingdom. The PowerBI map function was utilised for this purpose. The sites on the map were designated using latitude and longitude, and the average level of air pollution was determined and depicted using bubbles for each place. The bigger the bubble, the greater the average amount of air pollution. In addition, bubble colours were chosen based on the average air pollution level.

### **e. Predictive Data analytics: Time Series forecasting**

The Fbprophet Python library was installed in order to conduct time series forecasting. For training the FB-Prophet model, the data from January 1, 2021, to November 30, 2021, was used, and the data for the subsequent 30-day period was forecasted (December 1st, 2021 - December 31st, 2021). The estimated values were then cross validated using data from December 1, 2021 to December 31, 2021 that had been cleansed. Which provided RMSE and MAPE values for each day as output. In accordance with the predictions, three graphs are generated for each air pollutant ( $NO_2$  and  $PM_{10}$ ). In the first graph, the predicted concentration value is compared to the actual concentration value, and the performance metrics RMSE and MAPE are plotted against the number of days remaining on the Horizon (the range of days in forecasting).

### **f. Finding Correlation between Atmospheric Temperature and Air pollutant level:**

For determining the relationship between Atmospheric temperature and air pollutant level, the data from the Newcastle Urban observatory is utilised. Temperatures collected by sensors each hour are analysed alongside air pollutant levels recorded per hour. For this research, data spanning six months (1st July 2021 to 31st December 2021) are included, as well as data from a single site. First, the two datasets were combined using timestamp values, and then the correlation was computed and shown using the `corr()` method. In addition, PowerBI was utilised to comprehend the relationship between the variables in question and the time series. The X axis represents the Time series, while the two Y axes represent Temperature and Air pollutant level, respectively.

## **Results**

All the measures for  $NO_2$  and  $PM_{10}$  are in micrograms in each cubic metre units.

### **a. Geospatial Representation:**

Below are the PowerBI-generated Geospatial Representation of  $NO_2$  and  $PM_{10}$  concentrations by location in Birmingham, Manchester, and Newcastle for the year 2021. It can be seen that the areas marked by bubbles are the locations where the sensors are based, and red bubble are the locations where the Average levels of  $NO_2$  and  $PM_{10}$  concentration are high.



Figure 1: Location wise Concentration of *NO*<sub>2</sub> in Brimingham for year 2021



Figure 2: Location wise Concentration of *PM*10 in Brimingham for year 2021



Figure 3: Location wise Concentration of  $NO_2$  in Manchester for year 2021



Figure 4: Location wise Concentration of  $PM_{10}$  in Manchester for year 2021

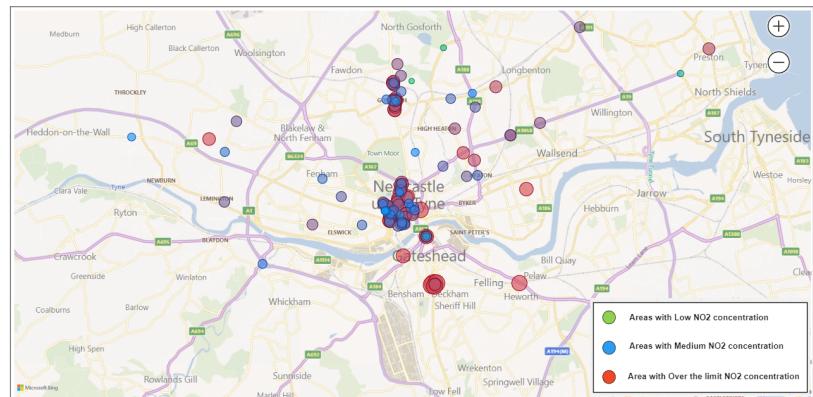


Figure 5: Location wise Concentration of  $NO_2$  in Newcastle for year 2021

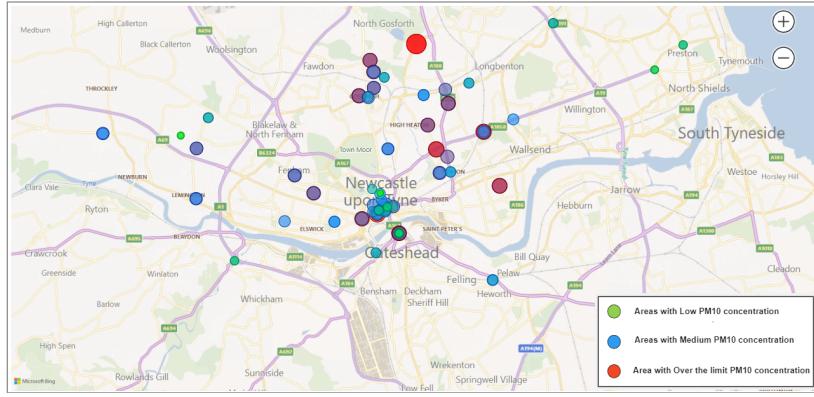


Figure 6: Location wise Concentration of  $PM_{10}$  in Newcastle for year 2021

#### b. Comparison of Month wise $NO_2$ levels in the UK cities in the year 2021

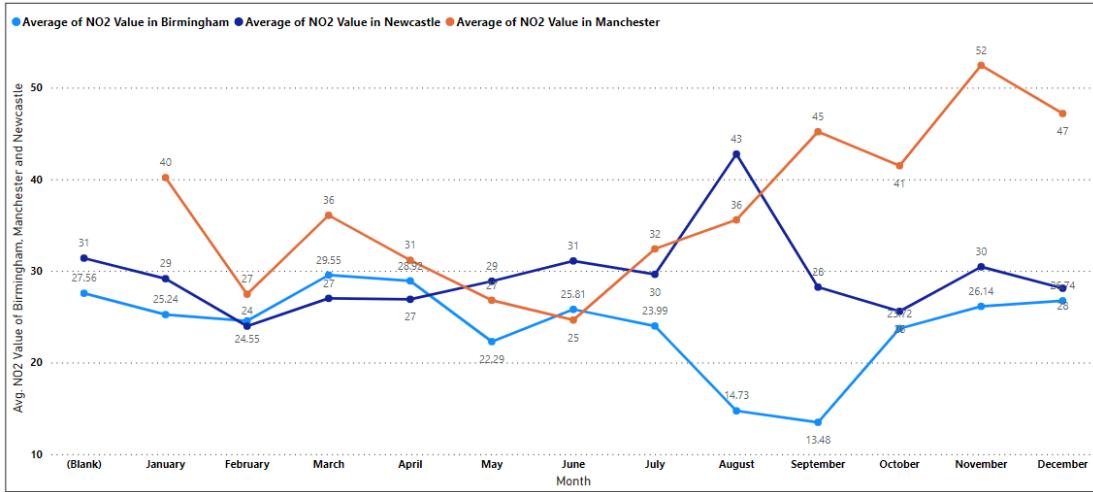


Figure 7: Month wise Comparison of Average  $NO_2$  level in the UK cities in the year 2021

For Manchester, The Highest average of  $NO_2$  concentration was recorded among all three cities. A pattern can be seen in the Manchester data, where the  $NO_2$  concentration is dropping by every month from January to June 2021 and after that from June to December 2021 gradually increase is visible. For Birmingham, the average  $NO_2$  concentration value throughout the year is in between 20 -30, except for months August and September 2021, where the  $NO_2$  levels have dropped significantly. Whereas in Newcastle, the average  $NO_2$  concentration value throughout the year is in between 25 -32, except for month August 2021, where the  $NO_2$  levels have hiked significantly and dropped again.

#### c. Comparison of Month wise $PM_{10}$ levels in the UK cites in the year 2021

For Birmingham, December 2021 had the highest Average of  $PM_{10}$  Value whereas the lowest was in February. The Average  $PM_{10}$  concentration value appears to be increasing throughout the year in Birmingham. Newcastle had the least average of  $PM_{10}$  concentration in year 2021 among all three cities. Whereas, Manchester recorded it highest average  $PM_{10}$  concentration value in month of March and April.

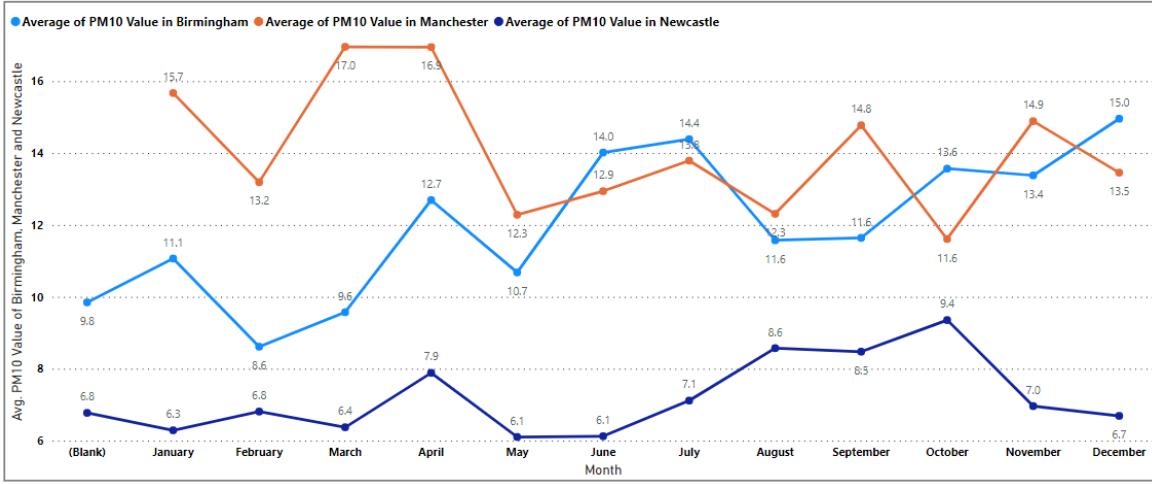


Figure 8: Month wise Comparison of Average  $PM_{10}$  level in the UK cities in the year 2021

#### d. Predictive Analysis- FB Prophet Model

The Fbprophet Python library was installed in order to conduct time series forecasting. For training the FB-Prophet model, the data from January 1, 2021, to November 30, 2021, was used, and the data for the subsequent 30-day period was forecasted (December 1st, 2021 - December 31st, 2021). The estimated values were then cross-validated using data from December 1, 2021 to December 31, 2021 that had been cleansed. Which provided RMSE and MAPE values for each day as output. In accordance with the predictions, three graphs are generated for each air pollutant ( $NO_2$  and  $PM_{10}$ ). In the first graph, the predicted concentration value is compared to the actual concentration value, and the performance metrics RMSE and MAPE are plotted against the number of days remaining on the Horizon (the range of days in forecasting).

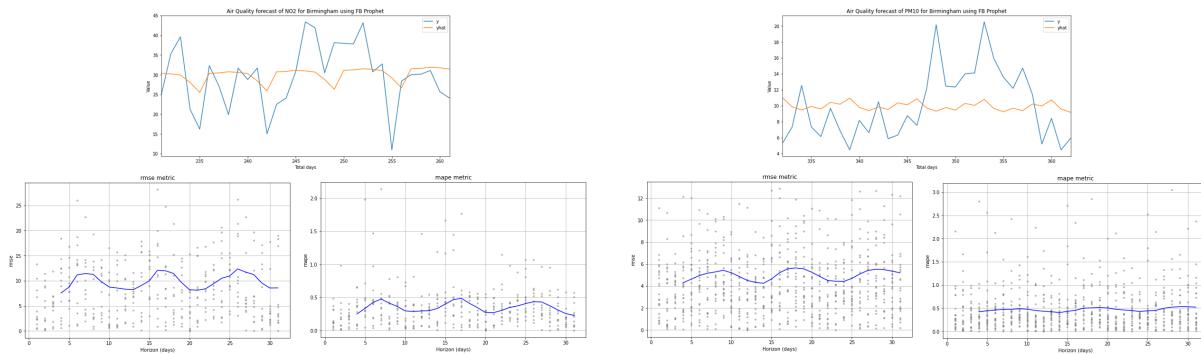


Figure 9: Prediction plot of  $NO_2$  and  $PM_{10}$  variables of Birmingham City

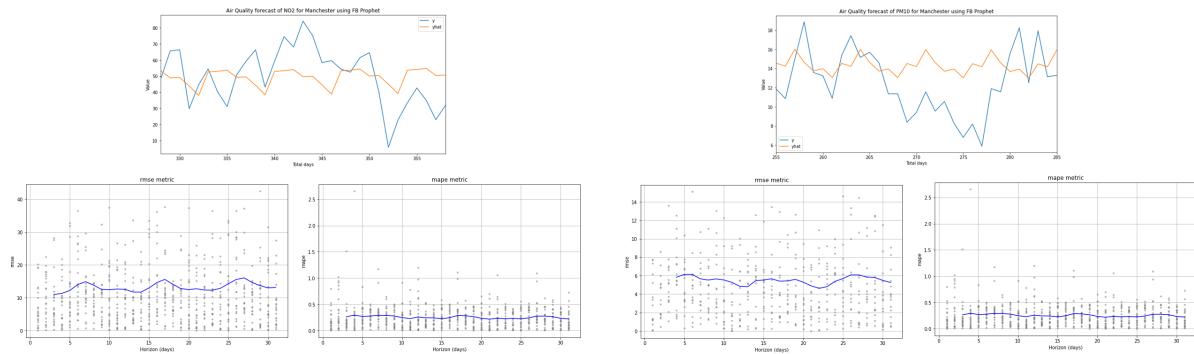


Figure 10: Prediction plot of  $NO_2$  and  $PM_{10}$  variables of Manchester City

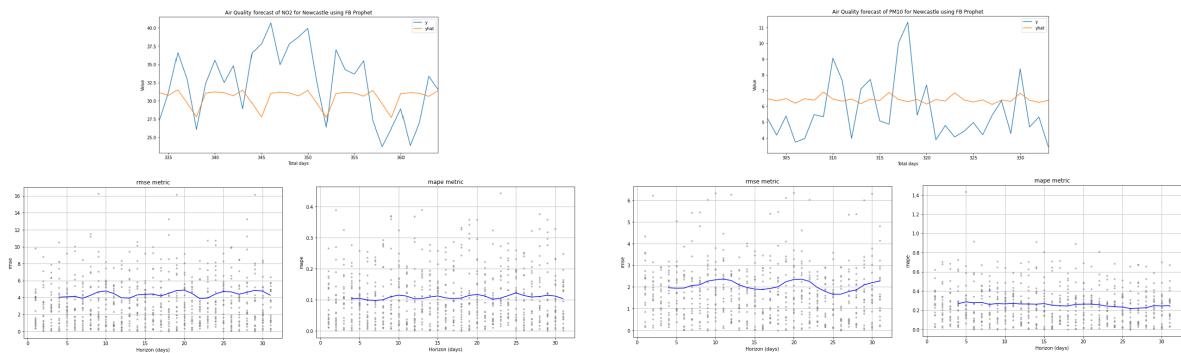


Figure 11: Prediction plot of  $NO_2$  and  $PM_{10}$  variables of Newcastle City

e. Correlation between temperature and Air pollutant levels

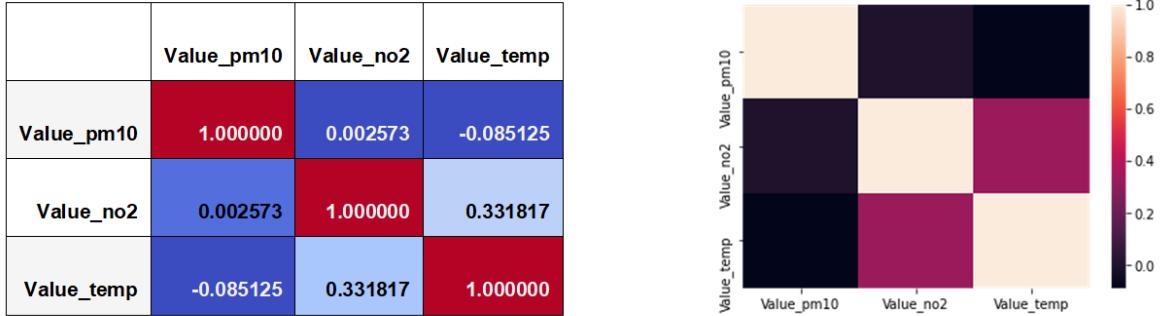


Figure 12: Correlation matrix and graph

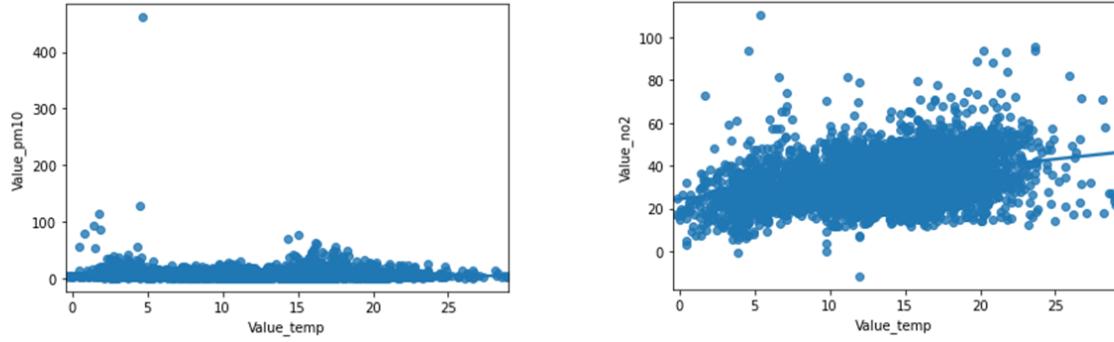


Figure 13: Scatter plot of correlation of Temperature VS  $PM_{10}$  and Temperature VS  $NO_2$

In [Fig – 12] and [Fig – 13], it can be observed from the table that Temperature value and  $NO_2$  value are significantly positively correlated. Whereas Temperature value and  $PM_{10}$  value has a trivial negative correlation between them.

f. Visualizing data in PowerBI, to find correlation between the Temperature VS  $NO_2$  &  $PM_{10}$

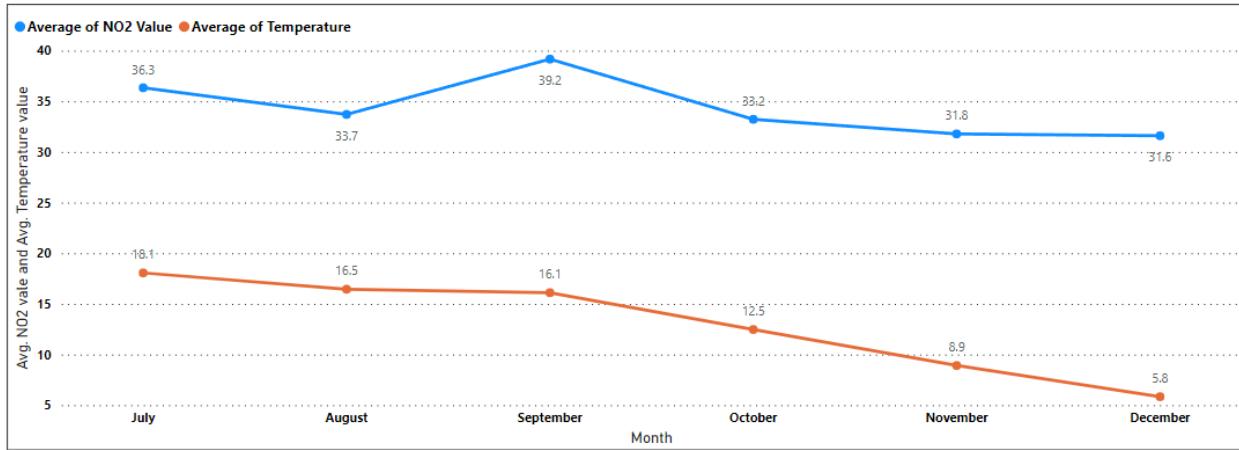


Figure 14: Month wise Comparison of Avg.  $NO_2$  value VS Avg. Temp value of Newcastle

According to [Fig – 14] Average of  $NO_2$  Value and total Average of Temperature are positively correlated with each other. Over the period of 6 months, both  $NO_2$  and temperature are dropping gradually.

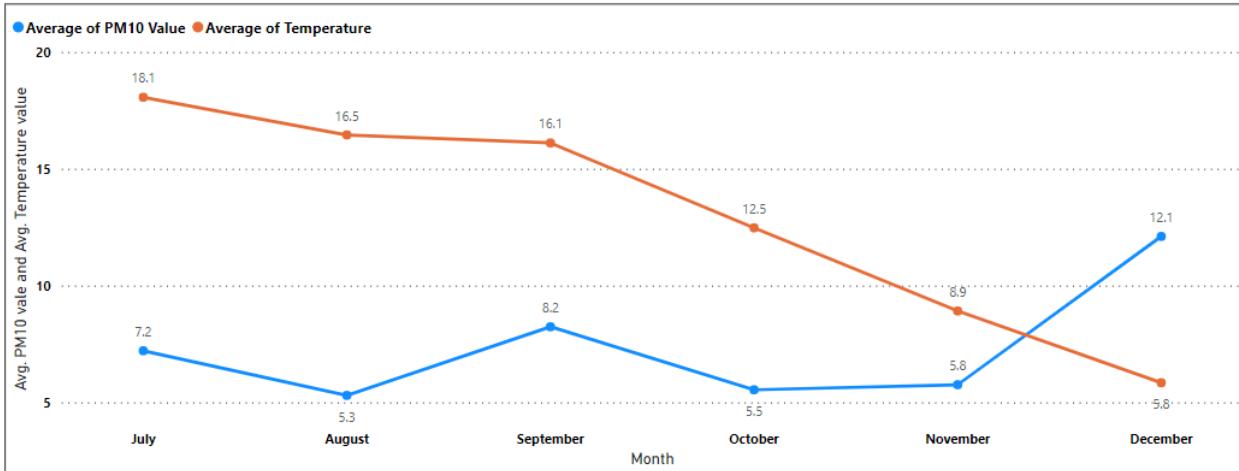


Figure 15: Month wise Comparison of Avg.  $PM_{10}$  value VS Avg. Temp value of Newcastle

According to [Fig – 15] Average Temperature value and Average of  $PM_{10}$  value has a no correlation until November 1st,2021 but after that negative correlation can be seen.

## Discussion

The data from the sensors gets recorded at the Birmingham Urban Observatory on a minute-by-minute basis, and these data may be downloaded for public use and research purposes by directly downloading it from the website. User need to select the theme and time frame, and the data will be downloaded. At Birmingham the sensors record the  $NO_2$  and  $PM_{10}$  value for 18 locations in the city. At the Newcastle Urban Observatory, the data from the sensors is captured minute-by-minute which can be obtained for public usage and research purposes via one of two methods: either manually or via an API. Using one of these approaches, one may access the data for a specific topic and time period. In Newcastle, the  $NO_2$  and  $PM_{10}$  values are recorded by 65 sensors. On the other hand, Manchester's Air quality data recorded by UK-AIR, for every 1 hour and allows users to retrieve the data only through a manual process. The  $NO_2$  and  $PM_{10}$  data is only recorded for 4 locations in the city, as a result the Manchester dataset for analysis is small. These locations were chosen as the pollutant and timeframe only matched for these cities. All other Air pollutants are not recorded at each station in the city of Birmingham and Manchester, causing this study to be limited to solely  $NO_2$  and  $PM_{10}$ .

At both Newcastle and Birmingham, the monthly and daily averages of  $NO_2$  and  $PM_{10}$  are uniformly distributed, with no noticeable trend. For Manchester, however, both daily and hourly comparisons reveal a trend. On the 14th to 16th of each month, the average  $NO_2$  value is significantly increased, and the hourly average  $NO_2$  value is greatest between 6:00 p.m. and 9:00 p.m. Additionally, the average daily  $PM_{10}$  concentration in Manchester is highest between the 15th and 17th of every month, and it is uniformly spread throughout the day. It is also possible to conclude that the  $NO_2$  and  $PM_{10}$  levels in Manchester in 2021 are greater than those in the other two cities. While the average  $NO_2$  concentration in Birmingham in 2021 is lower than that of the other cities, the average  $PM_{10}$  concentration in Newcastle in 2021 is the lowest of the three.

Geospatial visualisation is used by averaging the Air pollutant's concentration values and creating a colour hue segregation that ranges from low concentration to over the limit concentration of air pollutants to identify which location sensor in the city has recorded the highest and lowest average  $NO_2$  and  $PM_{10}$  concentration value over the course of 1 year. This information might be used to identify areas with the greatest  $NO_2$  and  $PM_{10}$  concentrations based on the colour hue representation.

The study involves in predicting the concentration of Air Pollutant's across the cities. In this study, the predictive analysis was carried out using Time Series Analysis in conjunction with the FB-Prophet model. It is clear that a trend can be observed when forecasting using historical data. The Performance Metrics RMSE and MAPE are shown versus the number of days remaining on the horizon. The [Fig-18] and [Fig-19] in the Appendix displays the performance metric values that were acquired across the horizon days based on the forecasting.

Further extension of study was to find if there is evidence of correlation between Atmospheric temperature and  $NO_2$  & Atmospheric temperature and  $PM_{10}$  value in Newcastle data. Correlation matrix, correlation graph, and scatter plot were shown for this purpose. Using PowerBI, even the correlation between variables was investigated. The conclusion of the experiment was determined that  $NO_2$  concentrations were positively correlated with atmospheric temperatures. And there is no significant evidence of a correlation between  $PM_{10}$  and the temperature of the atmosphere.

There are several external factors that affect the level of air pollutants emitted into the environment. Some of the parameters that may be examined are traffic data, population density in the vicinity, and other weather data. These factors will have influence on the levels of air pollutants. The future scope will involve addressing all of the factors that will help in the future examination of the Air Quality across cities or within cities.

## Future Implications

Future implications of this work would include a better understanding of the influence external elements such as humidity, wind speeds, and air pressure have on air quality. The data for those variables that are

available, and further analysis will be conducted to formulate a comparison between the cities. In addition, the predictive analysis would necessitate the application of Neural Network models, which have a reduced error rate when forecasting Air Quality data. The suggested technique performs well in terms of forecast accuracy for low-error time series data. However, performance can be enhanced with FB Prophet's fusion algorithms. Scalability might be an additional difficulty when evaluating a huge dataset. In combination with FB Prophet, transfer learning may be used to increase scalability and manage enormous datasets. High-precision real-time predictions are contingent on the model used for training and validation.

## Conclusion

This research focused on comparative study of existing air quality data for Birmingham, Manchester, and Newcastle. Visual and statistical comparisons of data between January and December 2021 represent that  $PM_{10}$  and  $NO_2$  concentrations are highest in Manchester City, whereas  $PM_{10}$  is lowest in Newcastle upon Tyne, and  $NO_2$  is lowest in Birmingham City. Even if the final results suggest that Newcastle upon Tyne has lower  $PM_{10}$  values, the appearance of many outlier numbers indicates that there have been considerable swings that may be attributable to external factors. The geographical visual analysis of the concentrations, based on the sensor locations in all three cities, indicates the  $NO_2$  and  $PM_{10}$  concentration levels are high in the Urban areas of the city. For the time series prediction of concentrations, FB Prophet model suggested the presence of trend and seasonality patterns, the model itself forecasted values with non-significant errors, ranging between 10 and 20% of MAPE on average. The Correlation analysis suggested positive correlation between  $NO_2$  and Temperature variable.

## References

- [1] Air pollution: applying All Our Health. [Online]. Available from: <https://www.gov.uk/government/publications/air-pollution-applying-all-our-health/air-pollution-applying-all-our-health> (Accessed 22 August 2022).
- [2] World Health Organization (2021) COP26 special report on climate change and health: the health argument for climate action. Geneva: World Health Organization. Available from: <https://apps.who.int/iris/bitstream/handle/10665/345334/9789240034433-eng.pdf>
- [3] Air pollution. [Online] [online]. Available from: <https://www.who.int/health-topics/air-pollution> (Accessed 15 May 2022).
- [4] Curtis, L., Rea, W., Smith-Willis, P., Fenyves, E. & Pan, Y. (2006) ‘Adverse health effects of outdoor air pollutants’, Environment International, 32(6), pp. 815-830. Available from: <https://doi.org/10.1016/j.envint.2006.03.012>
- [5] Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M. & Khorasani, N. (2021) ‘Artificial intelligence accuracy assessment in  $NO_2$  concentration forecasting of metropolises air’, Scientific Reports, 11(1), p. 1805. Available from: <https://www.nature.com/articles/s41598-021-81455-6>
- [6] Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scorticichini, M., Shtein, A., Viegi, G., Kloog, I. & Schwartz, J. (2019) ‘Estimation of daily PM<sub>10</sub> and PM<sub>2.5</sub> concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model’, Environment International, 124pp. 170-179. Available from: <https://doi.org/10.1016/j.envint.2019.01.016>
- [7] Air quality in Mexico City during the fuel shortage of January 2019 - ScienceDirect. [Online]. Available from: <https://doi.org/10.1016/j.atmosenv.2019.117131> (Accessed 1 June 2022).
- [8] Rani Samal, K.K., Sathya Babu, K., Panda, A.K. & Das, S.K. (2020) ‘Data Driven Multivariate Air Quality Forecasting using Dynamic Fine Tuning Autoencoder Layer’, in 2020 IEEE 17th India Council International Conference (INDICON). [Online]. December 2020 pp. 1-6
- [9] Cheval, S., Dumitrescu, A. & Bell, A. (2009) ‘The urban heat island of Bucharest during the extreme high temperatures of July 2007’, Theoretical and Applied Climatology, 97(3-4), pp. 391-401.
- [10] R. Garcia-Herrera, J. Diaz, R.M. Trigo, J. Luterbacher, E.M. Fischer ‘A review of the European summer heat wave of 2003’ Critical Reviews in Environmental Science and Technology, 40 (4) (2010), pp. 267-306
- [11] World Health Organization (ed.) (2010) Who guidelines for indoor air quality: selected pollutants. Copenhagen: WHO.
- [12] Liu, H., Yan, G., Duan, Z. & Chen, C. (2021) ‘Intelligent modeling strategies for forecasting air quality time series: A review’, Applied Soft Computing, 102p. 106957.
- [13] Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. & Baklanov, A. (2012) ‘Real-time air quality forecasting, part I: History, techniques, and current status’, Atmospheric Environment, 60pp. 632-655.
- [14] Cabaneros, S.M., Calautit, J.K. & Hughes, B.R. (2019) ‘A review of artificial neural network models for ambient air pollution prediction’, Environmental Modelling & Software, 119pp. 285-304.
- [15] Verma, P., Reddy, S.V., Ragha, L. & Datta, D. (2021) ‘Comparison of Time-Series Forecasting Models’, in 2021 International Conference on Intelligent Technologies (CONIT). [Online]. June 2021 pp. 1-7.
- [16] Gore, R.W. & Deshpande, D.S. (2017) ‘An approach for classification of health risks based on air quality levels’, in 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM). [Online]. October 2017 pp. 58-61.
- [17] S. M. S. Nagendra, K. Venugopal, and S. L. Jones, “Assessment of air quality near traffic intersections in Bangalore city using air quality indices,” Transportation Research Part D: Transport and Environment, vol. 12, no. 3, pp. 167-176, May 2007, doi: 10.1016/j.trd.2007.01.005.

- [18] Naseem, F., Rashid, A., Izhar, T., Khawar, M.I., Bano, S., Ashraf, A. & Adnan, M.N. (2018) ‘An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting’, Journal of Applied Agriculture and Biotechnology, 2(2), pp. 37-44.
- [19] V, A., P, G., R, V. & K p, S. (2018) ‘DeepAirNet: Applying Recurrent Networks for Air Quality Prediction’, Procedia Computer Science, 132pp. 1394-1403.
- [20] Kingsy, G.R., Manimegalai, R., Geetha, D.M.S., Rajathi, S., Usha, K. & Raabiathul, B.N. (2016) ‘Air pollution analysis using enhanced K-Means clustering algorithm for real time sensor data’, in 2016 IEEE Region 10 Conference (TENCON). [Online]. November 2016 pp. 1945-1949.
- [21] Kumar Jha, B. & Pande, S. (2021) ‘Time Series Forecasting Model for Supermarket Sales using FB-Prophet’, in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). [Online]. April 2021 pp. 547-554.
- [22] [M. Asgari, M. Farnaghi, and Z. Ghaemi, ‘Predictive mapping of urban air pollution using apache spark on a hadoop cluster,’ in Proc. Int. Conf. Cloud Big Data Comput., 2017, pp. 89-93.
- [23] D. Zhu, C. Cai, T. Yang, and X. Zhou, ‘A machine learning approach for air quality prediction: Model regularization and optimization,’ Big Data Cogn. Comput., vol. 2, no. 1, p. 5, 2018.
- [24] Samira, Muhammad & Salh, & Ahmed, Salah. (2014). Box -Jenkins Models For Forecasting The Daily Degrees Of Temperature In Sulaimani City. International Journal of Engineering Research and Applications (IJERA). 4. 2248-9622.
- [25] Salman Mohamadi, Farhang Yeganegi and Nasser M Nasrabadi. Detection and Statistical Modeling of Birth-Death Anomaly, 2019; arXiv:1906.11788.
- [26] Tealab, A. (2018) ‘Time series forecasting using artificial neural networks methodologies: A systematic review’, Future Computing and Informatics Journal, 3(2), pp. 334-340.
- [27] J. F. Li and Q. Zong, ‘The forecasting of the elevator traffic flow time series based on ARIMA and GP,’ Adv. Mater. Res., vols. 588-589, pp. 1466-1471, Nov. 2012.
- [28] Jayamurugan, R., Kumaravel, B., Palanivelraja, S. & Chockalingam, M.P. (2013) ‘Influence of Temperature, Relative Humidity and Seasonal Variability on Ambient Air Quality in a Coastal Urban Area’, International Journal of Atmospheric Sciences, 2013pp. 1-7.
- [29] Li, X.; Miao, Y.; Ma, Y.; Wang, Y.; Zhang, Y. Impacts of synoptic forcing and topography on aerosol pollution during winter in Shenyang, Northeast China. Atmos. Res. 2021, 262, 105764, DOI: <https://doi.org/10.1016/j.atmosres.2021.105764>
- [30] Theoharatos, G., Pantavou, K., Mavrakis, A., Spanou, A., Katavoutas, G., Efstatiou, P., et al. (2010). Heat waves observed in 2007 in Athens, Greece: Synoptic conditions, bioclimatological assessment, air quality levels and health effects. Environmental Research, 110(2), 152-161
- [31] McMichael, A. J., Campbell-Lendrum, D. H., Corvalan, C. F., Ebi, K. L., Githeko, A. K., Scheraga, J. D., et al. (2003). Climate change and human health: Risks and responses. Geneva: World Health Organization
- [32] Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L. & Habiyaremye, G. (2018) ‘Temperature and air pollution relationship during heatwaves in Birmingham, UK’, Sustainable Cities and Society, 43pp. 111-120.
- [33] Bham Urban Obs Data. [online]. Available from: <https://data.birminghamurbanobservatory.com/map/> platforms (Accessed 8 June 2022).
- [34] Urban Observatory. [Online]. Available from: <https://newcastle.urbanobservatory.ac.uk/> (Accessed 8 June 2022).
- [35] Siami-Namini, S., Tavakoli, N. & Siami Namin, A. (2018) ‘A Comparison of ARIMA and LSTM in Forecasting Time Series’, in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). [Online]. December 2018 pp. 1394-1401.

- [36] M. Krieger, “Time Series Analysis with Facebook Prophet: How it works and How to use it,” Medium, Mar. 12, 2021. <https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a> (accessed Aug. 29, 2021).
- [37] “Prophet: forecasting at scale,” Facebook Research, Feb. 23, 2017. <https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/> (accessed Jul. 23, 2022).
- [38] “Time Series Forecasts using Facebook’s Prophet,” Analytics Vidhya, May 10, 2018. <https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/> (accessed Jul. 29, 2022).
- [39] “Prophet,” Prophet. <http://facebook.github.io/prophet/> (accessed July. 25, 2022).
- [40] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “Evaluating algorithm performance metrics tailored for prognostics,” in 2009 IEEE Aerospace conference, Mar. 2009, pp. 1-13. doi: 10.1109/AERO.2009.4839666.

## Appendix

### Comparison of Day wise $NO_2$ and $PM_{10}$ levels in the UK cities in the year 2021

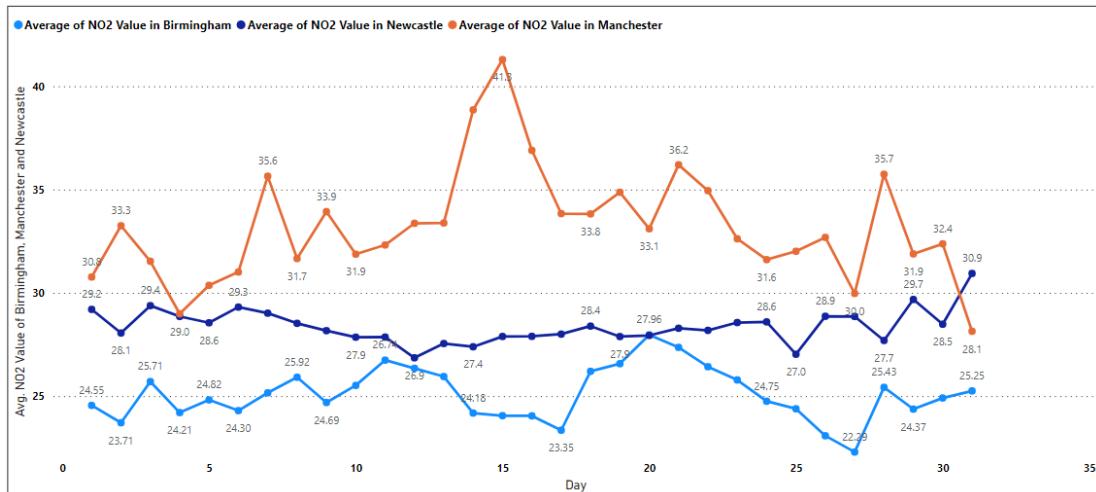


Figure 16: Day wise Comparison of Average  $PM_{10}$  level in the UK cities in the year 2021

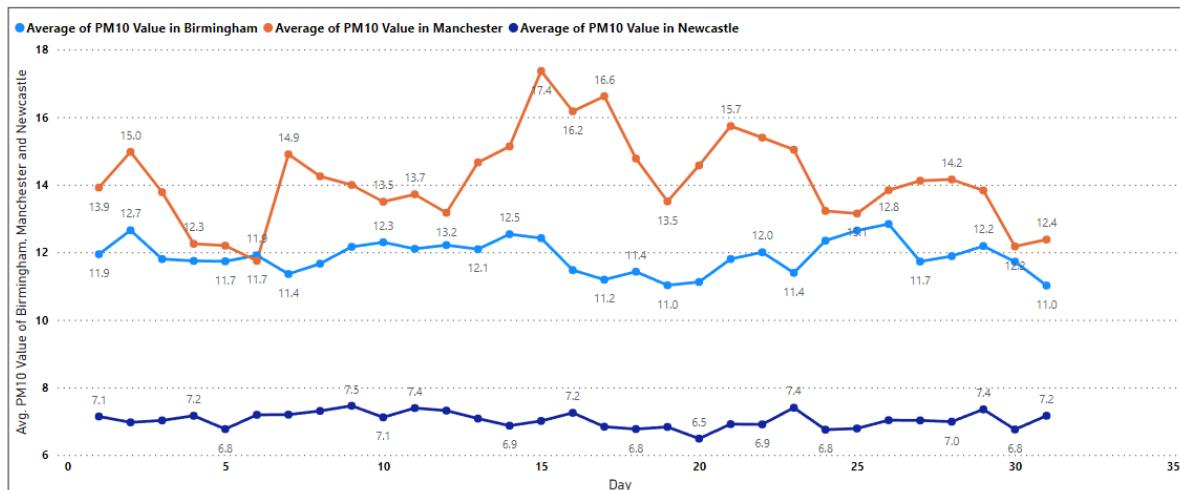


Figure 17: Day wise Comparison of Average  $PM_{10}$  level in the UK cities in the year 2021

Performance matrix produced by FB Prophet model:

## Birmingham

	horizon	mse	rmse	mae	mape	mdape	coverage
0	4 days	57.1443	7.55939	6.00525	0.24732	0.2033	0.80631
1	5 days	76.1276	8.72511	7.34594	0.3268	0.28464	0.72748
2	6 days	124.411	11.154	9.4483	0.42213	0.37709	0.53829
3	7 days	130.344	11.4168	9.73874	0.47345	0.34753	0.55631
4	8 days	126.93	11.2663	9.1837	0.41202	0.34677	0.53153
5	9 days	96.609	9.82899	8.12431	0.36252	0.33179	0.57658
6	10 days	76.4916	8.74595	6.95844	0.29495	0.25316	0.61261
7	11 days	72.9853	8.54314	6.83136	0.28459	0.25316	0.61036
8	12 days	69.17	8.31685	6.62873	0.29345	0.2124	0.68919
9	13 days	67.5085	8.21635	6.93278	0.29793	0.26784	0.75
10	14 days	81.7877	9.04366	7.6175	0.3312	0.29432	0.71847
11	15 days	99.5749	9.97872	8.67661	0.39721	0.36425	0.61486
12	16 days	145.167	12.0485	10.288	0.46351	0.40072	0.45721
13	17 days	143.68	11.9867	10.137	0.48522	0.39738	0.52703
14	18 days	131.021	11.4465	9.13147	0.40839	0.37152	0.63288
15	19 days	93.0196	9.64467	7.68591	0.34406	0.28004	0.71396
16	20 days	67.3211	8.20495	6.34631	0.27015	0.2390	0.77703
17	21 days	65.6744	8.10397	6.36394	0.26804	0.28004	0.77703
18	22 days	70.4805	8.39527	6.83389	0.30051	0.28524	0.74775
19	23 days	88.2553	9.39443	8.16839	0.34535	0.30775	0.67342
20	24 days	109.798	10.4785	9.08185	0.36948	0.30775	0.61486
21	25 days	119.938	10.9516	9.66775	0.40431	0.33283	0.55631
22	26 days	152.518	12.3498	10.7225	0.43297	0.42192	0.5045
23	27 days	137.008	11.7051	10.0249	0.42891	0.3808	0.55631
24	28 days	125.892	11.2202	9.02288	0.3725	0.3021	0.65991
25	29 days	91.0045	9.53963	7.39671	0.30908	0.26083	0.71622
26	30 days	73.096	8.54962	6.38286	0.254	0.19722	0.75
27	31 days	73.4159	8.56831	6.17468	0.22676	0.19351	0.75

Figure 18: Performance metrics values of Birmingham City

## Manchester

	horizon	mse	rmse	mae	mape	mdape	coverage
0	3 days	120.685	10.9857	8.61896	0.26322	0.16386	0.86957
1	4 days	128.374	11.3302	8.98612	0.29279	0.17663	0.89666
2	5 days	149.017	12.2073	9.37141	0.2690	0.1898	0.89477
3	6 days	196.132	14.0047	10.9309	0.27561	0.20793	0.8494
4	7 days	222.105	14.9032	12.1073	0.29109	0.27666	0.82411
5	8 days	192.426	13.8718	11.5888	0.28875	0.27043	0.86364
6	9 days	158.515	12.5903	10.5633	0.27608	0.25338	0.90119
7	10 days	155.36	12.4644	10.0475	0.24762	0.22663	0.88011
8	11 days	160.758	12.6791	9.66873	0.22679	0.17569	0.82609
9	12 days	157.601	12.5539	9.6264	0.2561	0.18937	0.82609
10	13 days	138.216	11.7565	8.99791	0.24016	0.17569	0.86957
11	14 days	137.521	11.7269	9.16855	0.24011	0.18469	0.88154
12	15 days	168.845	12.994	9.92222	0.24299	0.19155	0.85129
13	16 days	214.261	14.6377	11.4479	0.24416	0.24264	0.80781
14	17 days	243.509	15.6048	12.484	0.28627	0.25576	0.7917
15	18 days	199.747	14.1332	11.5609	0.28123	0.24722	0.85968
16	19 days	164.218	12.8148	10.5119	0.26508	0.23915	0.88867
17	20 days	156.149	12.496	9.99875	0.23438	0.21643	0.8643
18	21 days	162.408	12.7439	9.62925	0.21803	0.17426	0.82609
19	22 days	152.701	12.3572	9.93927	0.23173	0.19262	0.82609
20	23 days	151.852	12.3228	9.2303	0.22656	0.16886	0.85507
21	24 days	165.849	12.8782	9.88739	0.2307	0.19262	0.83869
22	25 days	202.94	14.2458	10.6874	0.22355	0.2098	0.82231
23	26 days	242.546	15.5739	12.0499	0.24126	0.21535	0.77505
24	27 days	258.56	16.0798	12.5768	0.27395	0.22505	0.7747
25	28 days	215.463	14.6787	11.6611	0.27461	0.21873	0.83267
26	29 days	188.269	13.7211	10.7542	0.26449	0.21873	0.87022
27	30 days	169.563	13.0216	10.1296	0.23242	0.19799	0.8643
28	31 days	172.909	13.1495	9.91705	0.22157	0.18365	0.82609

	horizon	mse	rmse	mae	mape	mdape	coverage
0	4 days	33.9229	5.82433	4.97371	0.48192	0.30958	0.80209
1	5 days	37.6294	6.13428	5.35609	0.50402	0.32342	0.77971
2	6 days	37.6491	6.13589	5.24711	0.43504	0.3269	0.77971
3	7 days	32.0401	5.6604	4.83354	0.1489	0.31204	0.84058
4	8 days	30.4833	5.52117	4.66106	0.3803	0.29264	0.86522
5	9 days	31.847	5.64332	4.7565	0.39486	0.28848	0.80145
6	10 days	30.5617	5.52827	4.45196	0.33975	0.28854	0.77971
7	11 days	27.7677	5.26951	4.02204	0.32654	0.23252	0.77971
8	12 days	23.4136	4.83876	3.62722	0.32456	0.16534	0.84058
9	13 days	23.0738	4.80352	3.81422	0.34783	0.22599	0.86522
10	14 days	24.8971	5.44816	4.49024	0.40169	0.32399	0.84493
11	15 days	30.57	5.52902	4.70064	0.41629	0.34598	0.86667
12	16 days	32.0564	5.66184	4.70719	0.39805	0.35997	0.86667
13	17 days	29.2104	5.40466	4.41497	0.4127	0.29877	0.88696
14	18 days	29.4548	5.42723	4.42342	0.39196	0.28097	0.84638
15	19 days	31.2759	5.59249	4.53434	0.40335	0.27503	0.80145
16	20 days	27.9704	5.28871	4.24577	0.4135	0.26519	0.80145
17	21 days	23.911	4.88988	3.86035	0.31543	0.22689	0.86522
18	22 days	21.3914	4.62508	3.64655	0.31996	0.20196	0.88551
19	23 days	23.0691	4.80302	3.9312	0.35048	0.20713	0.88841
20	24 days	28.8655	5.37267	4.50108	0.39845	0.24089	0.82609
21	25 days	32.2517	5.67904	4.78009	0.40611	0.30377	0.82174
22	26 days	36.9576	6.07927	4.99302	0.39273	0.31009	0.80145
23	27 days	36.8754	6.07251	4.72242	0.40588	0.29116	0.82174
24	28 days	34.1926	5.84744	4.56317	0.38839	0.28198	0.84348
25	29 days	33.7416	5.80876	4.52882	0.40863	0.28338	0.84348
26	30 days	30.1539	5.49126	4.30768	0.36996	0.28338	0.84493
27	31 days	27.6863	5.26178	4.12422	0.35912	0.22828	0.84493

Figure 19: Performance metrics values of Manchester City

## Newcastle upon Tyne

	horizon	mse	rmse	mae	mape	mdape	coverage
0	4 days	16.3069	4.03818	3.17386	0.10392	0.09143	0.77342
1	5 days	16.6024	4.07461	3.32233	0.10518	0.09793	0.79669
2	6 days	17.0397	4.12791	3.25359	0.09944	0.09143	0.79669
3	7 days	15.2513	3.9053	3.13833	0.09788	0.08847	0.80955
4	8 days	17.9499	4.23673	3.21607	0.10035	0.08532	0.78506
5	9 days	21.6856	4.65678	3.45846	0.11019	0.08088	0.76852
6	10 days	22.7672	4.7715	3.57971	0.11515	0.08088	0.71219
7	11 days	19.9191	4.46308	3.43106	0.11218	0.09606	0.72627
8	12 days	15.6873	3.96072	3.09921	0.10372	0.09337	0.72505
9	13 days	15.4266	3.92767	3.09974	0.10444	0.08625	0.76363
10	14 days	18.5193	4.30341	3.34081	0.10901	0.09021	0.74158
11	15 days	19.1058	4.37102	3.55815	0.11271	0.1024	0.7673
12	16 days	19.4253	4.40742	3.47946	0.10626	0.09079	0.78138
13	17 days	17.3927	4.17046	3.33657	0.10409	0.09079	0.78138
14	18 days	20.074	4.4804	3.38862	0.10596	0.08367	0.76975
15	19 days	23.1569	4.81217	3.57488	0.11412	0.08794	0.75444
16	20 days	23.7603	4.87446	3.65204	0.11754	0.08755	0.75321
17	21 days	20.024	4.47527	3.44306	0.11241	0.09034	0.76852
18	22 days	14.9474	3.86619	3.08706	0.10226	0.08693	0.79547
19	23 days	15.5296	3.94076	3.1619	0.10562	0.09075	0.79547
20	24 days	19.4979	4.41564	3.51661	0.11361	0.09075	0.75566
21	25 days	22.1609	4.70754	3.86327	0.12251	0.10278	0.71341
22	26 days	21.7576	4.6645	3.71132	0.1139	0.10278	0.71096
23	27 days	19.2045	4.3823	3.48453	0.10921	0.10087	0.72382
24	28 days	21.2375	4.60842	3.51987	0.11075	0.09705	0.72382
25	29 days	23.471	4.84469	3.6039	0.11476	0.09705	0.72505
26	30 days	22.5946	4.75338	3.50801	0.11228	0.07743	0.73913
27	31 days	18.1644	4.26197	3.19927	0.10392	0.07743	0.78016

	horizon	mse	rmse	mae	mape	mdape	coverage
0	4 days	3.85298	1.9629	1.64856	0.26759	0.24119	0.8629
1	5 days	3.77131	1.94199	1.66395	0.28663	0.24486	0.85161
2	6 days	3.80106	1.94963	1.62404	0.27709	0.21332	0.86613
3	7 days	4.29305	2.07197	1.68525	0.28107	0.22997	0.85
4	8 days	4.40194	2.09808	1.66082	0.26036	0.19938	0.86613
5	9 days	5.15574	2.27063	1.80602	0.26974	0.25531	0.80323
6	10 days	5.43277	2.33083	1.87249	0.26511	0.25531	0.76774
7	11 days	5.60889	2.36831	1.95695	0.27189	0.26735	0.72258
8	12 days	5.2926	2.30056	1.88974	0.26441	0.26735	0.76452
9	13 days	4.40544	2.09891	1.73254	0.26438	0.25159	0.82903
10	14 days	3.92139	1.98025	1.60234	0.2629	0.22112	0.8629
11	15 days	3.62445	1.9038	1.58884	0.26866	0.23142	0.86613
12	16 days	3.55025	1.88421	1.52565	0.25187	0.20406	0.83548
13	17 days	3.72758	1.9307	1.54576	0.24612	0.19838	0.80161
14	18 days	4.04431	2.01105	1.57922	0.24339	0.19084	0.80161
15	19 days	5.02295	2.24119	1.76579	0.26099	0.2464	0.76774
16	20 days	5.48927	2.34292	1.85593	0.26321	0.26762	0.76613
17	21 days	5.57642	2.36144	1.90437	0.26349	0.27517	0.7371
18	22 days	5.10483	2.25939	1.81379	0.25601	0.26518	0.78065
19	23 days	3.92152	1.98028	1.60736	0.24298	0.26303	0.8629
20	24 days	3.22972	1.79714	1.45626	0.23592	0.22536	0.89516
21	25 days	2.76945	1.66417	1.40531	0.23152	0.23397	0.91452
22	26 days	2.7774	1.66667	1.35768	0.2174	0.21264	0.8871
23	27 days	3.19579	1.78768	1.41123	0.22103	0.19908	0.86613
24	28 days	3.47794	1.86492	1.4478	0.22502	0.18809	0.85161
25	29 days	4.39467	2.09635	1.6577	0.24635	0.22047	0.80323
26	30 days	4.83808	2.19956	1.73385	0.24427	0.23288	0.78387
27	31 days	5.175	2.27486	1.82085	0.24833	0.24214	0.76935

Figure 20: Performance metrics values of Newcastle upon Tyne City