Air Quality Data Analytics Training Program

ekonnect
knowledge foundation
A PRASAD MODAK INITIATIVE

Prepared by:
Vishwa Vijaysheel
Jaykumar Mehta
Pranav Bhardwaj

TABLE OF CONTENTS

ekonnect

ABOUT THE COURSE

BACKGROUND

Air quality has assumed a national importance given the rising air pollution in our cities and around industrial estates. Out of the top 15 air polluted cities in the world, India has 12 polluted cities. It is estimated that in 2018, nearly 1.2 million people in India lost their lives due to air pollution.

In response, the Ministry of Environment, Forests and Climate Change (MoEFCC) has formulated a National Clean Air Program (NCAP). For the non-attaining or non-complying cities, city specific air quality action plans are under preparation. Further, under the directions of the Central Pollution Control Board (CPCB), continuous air quality and stack emission monitoring instruments have been installed by industries and urban local bodies and the status on air quality is reported in the form of an Air Quality Index. While these efforts will help in combating the menace of air pollution, it is important to build capacities of air quality professionals, managers and planners.

The call for data-driven decision-making renders data analytics important and relevant. While air quality data informs the user about compliance of parameters, it is necessary to view the data in context of meteorology, terrain and sources to make informed decisions and chalk out action plans.

To address this need, Ekonnect Knowledge Foundation (Ekonnect), Mumbai organized a 2-day training programs on air quality data analytics, the fourth of its kind on 8th and 9th August 2019.

OBJECTIVES OF THE PROGRAM

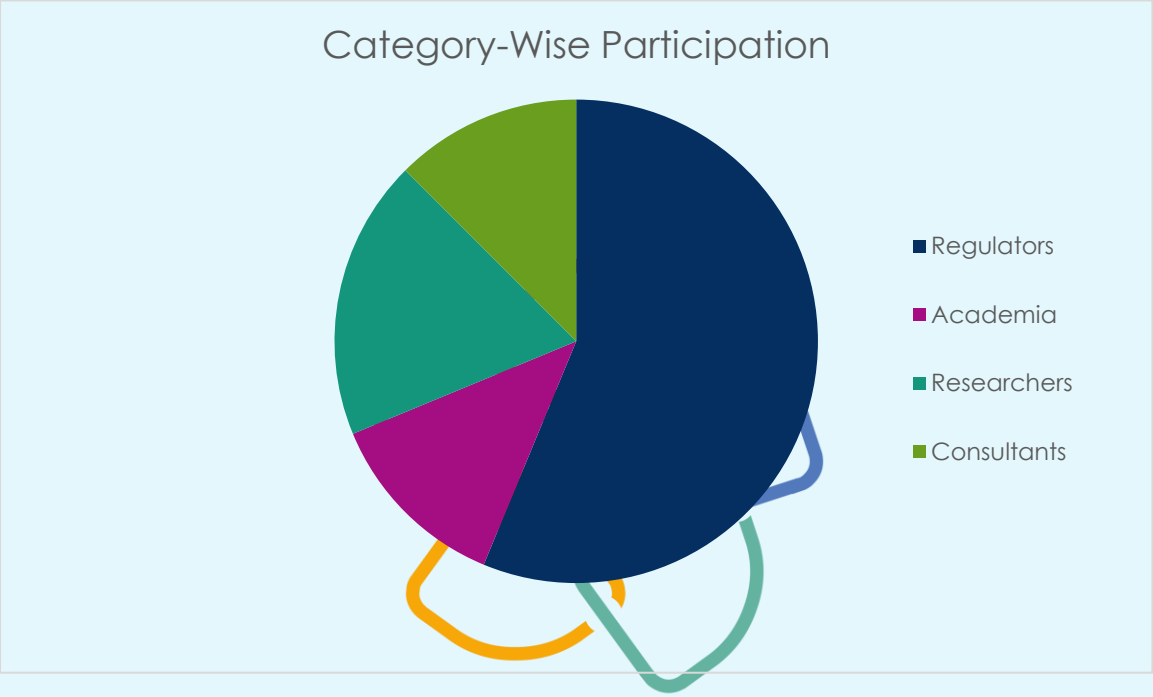
The overall objectives of the training programme can be summarized as below:

1. To enable air quality data users to check its authenticity and account for missing data and outliers.
2. To educate air quality data users towards application of statistical analysis methods for air quality data analysis.
3. To make users understand time-series plot, long term trends and short-term change computation, violation analysis and source apportionment of pollutants.
4. To demonstrate the capabilities of air quality dashboard as a tool for air quality data analytics
5. To enable users to understand decision making using data analysis.

PROGRAMME AGENDA

The training programme was scheduled over two days. The programme agenda covered all the aspects required to assess data on ambient air quality. The topics covered data cleaning, data quality checks, time-series visualization, detection of short-term and long-term change and meteorological data analysis. The following provides details of the training programme agenda.

The training was attended by regulators, researchers, academics and consultants, as shown by the pie chart below:



ekonnect

DAY-1: 8TH AUGUST 2019

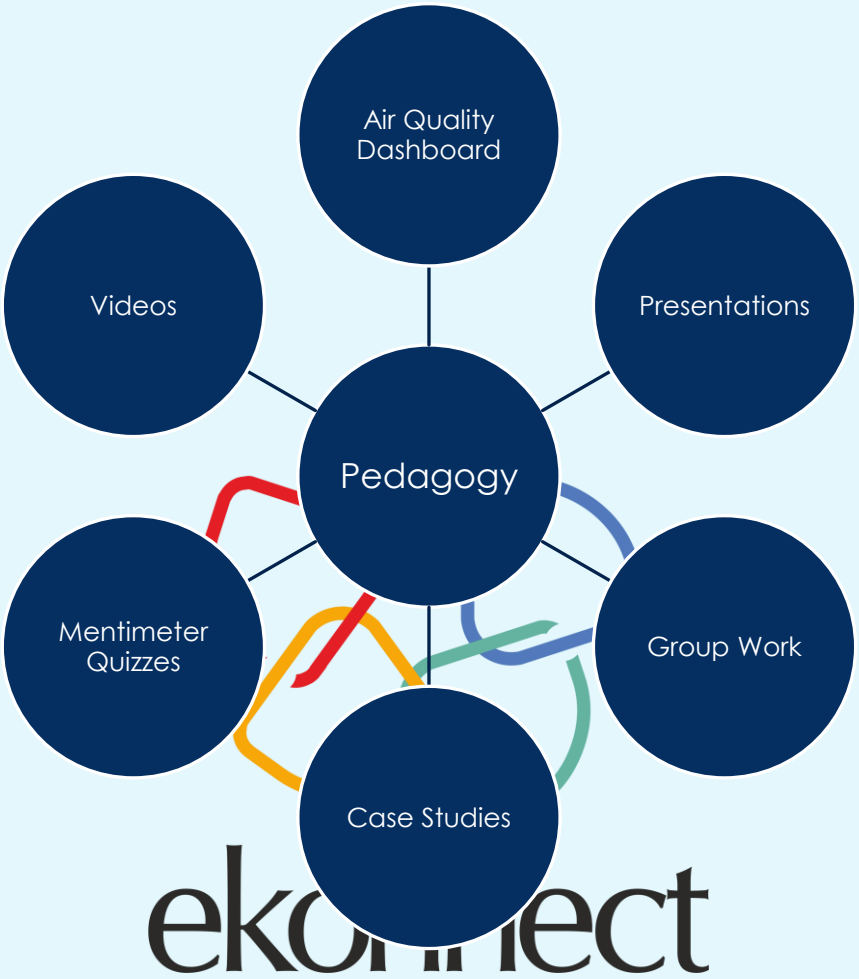
Task	Time Allotted
Session 1: Opening remarks and Data preparation	
Importance of Air Quality Data Analytics • What is the need for air quality monitoring? • Characteristics of air quality data and challenges in interpreting air quality data • Checking quality of concentrations and meteorological data	**1000 -1100 hrs**
Data Preparation • The process of data analysis and understanding air quality data types • Getting the data ready for the dashboard • Data cleaning using Microsoft Excel • Importance of removing unwanted and characters from a dataset and consistent representation of missing data • Why would data be missing in Manual and Automatic air quality data monitors? • Computing percent of missing data in excel	**1100 – 1200 hrs**
Session 2: Basic Processing of Air Quality Data	
Descriptive statistics, correlation and regression • Understanding basic statistical terminologies; Standard deviation, Coefficient of Variation, Quartiles, Skewness and Kurtosis • Correlation matrix, Correlation coefficient and coefficient of Determination • Data Visualization and Ratio plots	**1200 – 1300 hrs**
Dashboard demonstration • Uploading cleaned data on the dashboard • Preview of dataset, checking data quality, correlations, time series data visualization and ratio plots	**1345 – 1430 hrs**
Hands-On Work Session	**1515 – 1700 hrs**

DAY-2: 9TH AUGUST 2019

Task	Time Allotted
Session 1: Advanced Processing of air quality data	
Long-term Mann-Kendall and short-term change theory • Parametric and Non-parametric tests to detect air quality trends • Common tests for trend detection • Mann-Kendall test methodology and solved examples • Detecting short-term changes using Boxplots • Solved examples for short-term changes using Box-whisker diagrams	**0930 – 1030 hrs**
Violation Analysis theory and Violation Index • Violation of standards with examples; simultaneous violations • Theory of violation index and its computation	**1030 – 1100 hrs**
Dashboard demonstration • Box-whisker plots, Mann-Kendall tests and violation analysis using the dashboard and Detecting violations	**1100 – 1215 hrs**
Hotspot Analysis (Short-term change, long-term trends and violation index)	**1215 – 1300 hrs**
Session 2: Processing Meteorological data	
Theory of Wind rose, Pollution rose and Polar plots • Understanding conditional probability function and conditional bivariate probability function • Case Study of identifying sources using polar plots and Identification of sources of SO_2 and NO_X using conditional probability function	**1345- 1445 hrs**
Demonstration on the dashboard • Wind rose, Pollution rose and Polar plots • Theory of Persistent wind rose and demonstration on the dashboard	**1345- 1445 hrs**
Stack Influence theory and demonstration on the dashboard • Understanding Gaussian plume dispersion modelling • What is stack Influence and demonstration on the dashboard	**1515 – 1700 hrs**
Work session, discussions and closing remarks	**1515 – 1700 hrs**

PEDAGOGY

The pedagogies used consisted of classroom presentations and instructor-led training followed by practice sessions of hands-on training. Interactive tools to assess and engage the participants were used. Description of the different tools and topics covered under each type of pedagogy is given in the following sub-sections.



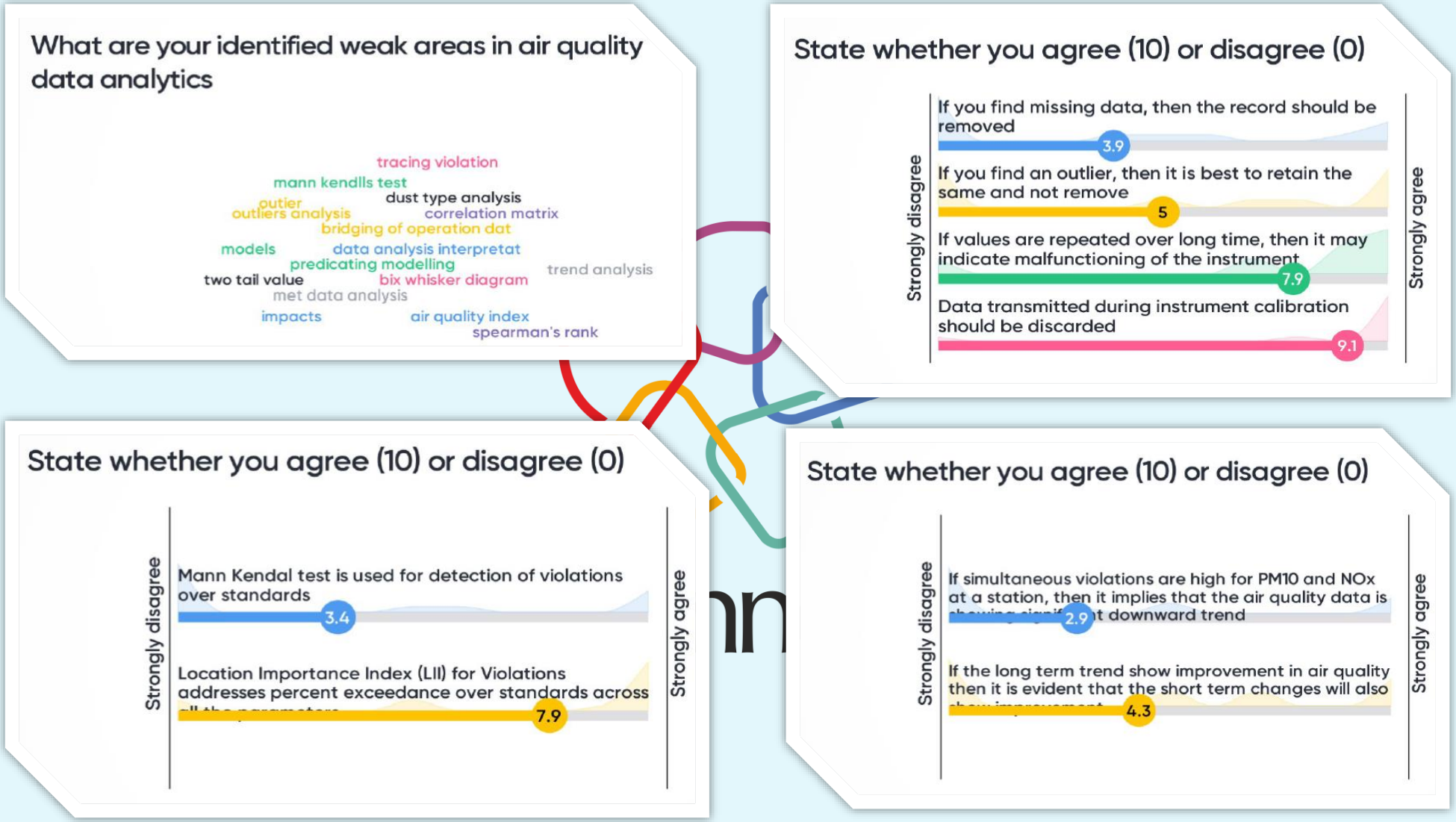


Figure 1: Interactive Mentimeter Questions for clarity of concepts

ABOUT THE AIR QUALITY DATA ANALYTICS DASHBOARD

The air quality analytics dashboard, conceived by Dr. Prasad Modk and developed under his guidance. Set of different functions from a simple time-series chart to Box-whisker plots and Mann-Kendal test, and for meteorological data analysis like Wind Rose plots, Pollution Rose plots Polar plots are as different modules in the dashboard that allow the user to investigate air quality data and deduce well informed decisions expeditiously, A list of the different modules in the dashboard is given in Appendix-3.

Figure 2 Discussion among participants during the group exercise

Following their exercise, the participants were asked to present their findings as depicted in Figure 2.

FEEDBACK

Feedback from the participants was solicited every day. Two methods were used to receive the feedback, namely feedback forms and Mentimeter.

FEEDBACK FORMS

The feedback form covered the following:

1. Course content and structure
2. Relevance of examples cited by the trainers for the better understanding of concepts
3. Trainers' ability to answers queries posed by the participants
4. Trainers' skills and competence in training and subject matter
5. Level of engagement of the sessions
6. Overall effectiveness of the training to improve participant's work

Rating used for evaluation was qualitative in nature, consisting of three options i.e. *Excellent, Satisfactory and Poor*. Scores of 5, 3 and 1 were assigned respectively to the aforementioned evaluation score. Table 1 summarizes the average score of all participants over four days for each of the six training evaluation criteria.

Table 1 Average score for training evaluation criteria[1]

Training Evaluation Criteria	Average Score (out of 5)
Course content and structure	4.65
Relevance of examples cited by the trainers for the better understanding of concepts	4.19
Trainers' ability to answers queries posed by the participants	4.79
Trainers' skills and competence in training and subject matter	4.79
Level of engagement of the sessions	4.53
Overall effectiveness of the training to improve participant's work	4.66
Overall average	**4.60**

Appendix 2 describes a typical sample survey form that was circulated to each participant at the end of each day.

[1] Scores: Excellent (5), Satisfactory (3) and Poor (1)

MENTIMETER

At the beginning of training, the participants were asked to self-assess their level of understanding of Microsoft Excel, data cleaning, basic statistics and in general the air quality data analytics. The average score of participants was 4.6 out of 10[2]. The participants rated themselves a little over 5 for MS Excel (5.6) and basic statistics (5.8), whereas, they identified Data cleaning (4.6) and Correlation and Regression (3.2) as relatively weaker areas. This initial assessment helped the trainers in the programme delivery and especially the practice sessions.

At the beginning of each day, a Mentimeter survey was conducted on level of understanding of topics discussed the previous day. Figure 3–Figure 5 illustrate the results of feedback surveys performed via Mentimeter.

At the end of training programme, participants were asked to re-evaluate their level of understanding of Microsoft Excel, data cleaning, basic statistics and various topics in air quality data analysis. The average score after two days rose to 7.2 out of 10[2] from the initial score of 4.6, implying a significant increase of understanding and skills from baseline assessment. Figure 6 And Figure 5 show such a change.

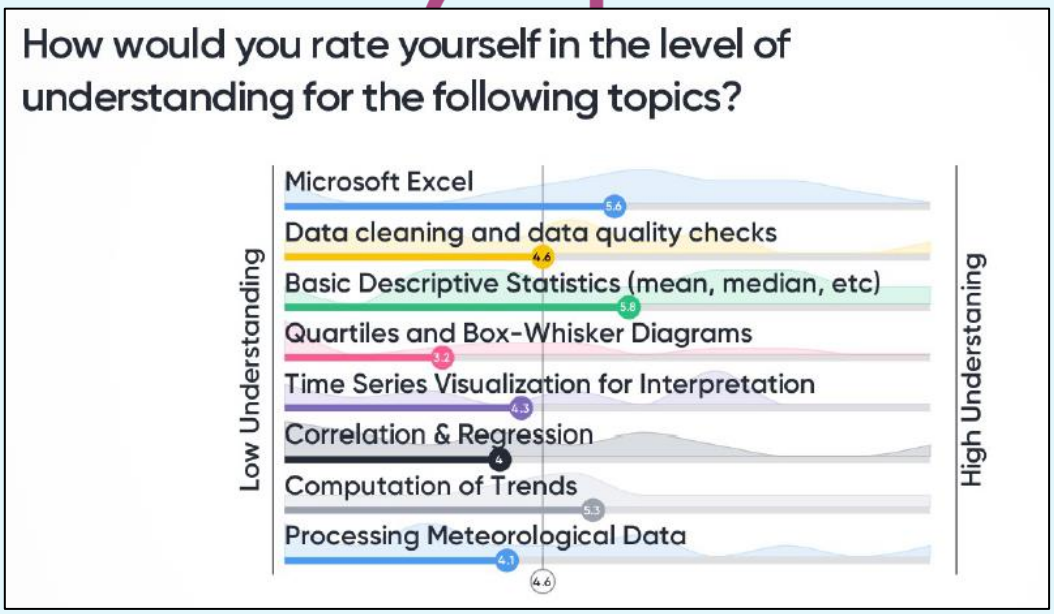


Figure 3 Self-evaluation of participants before the sessions on day 1

[2] The scale was from 0 to 10, with 0 signifying low level of understanding and 10 implying a high level

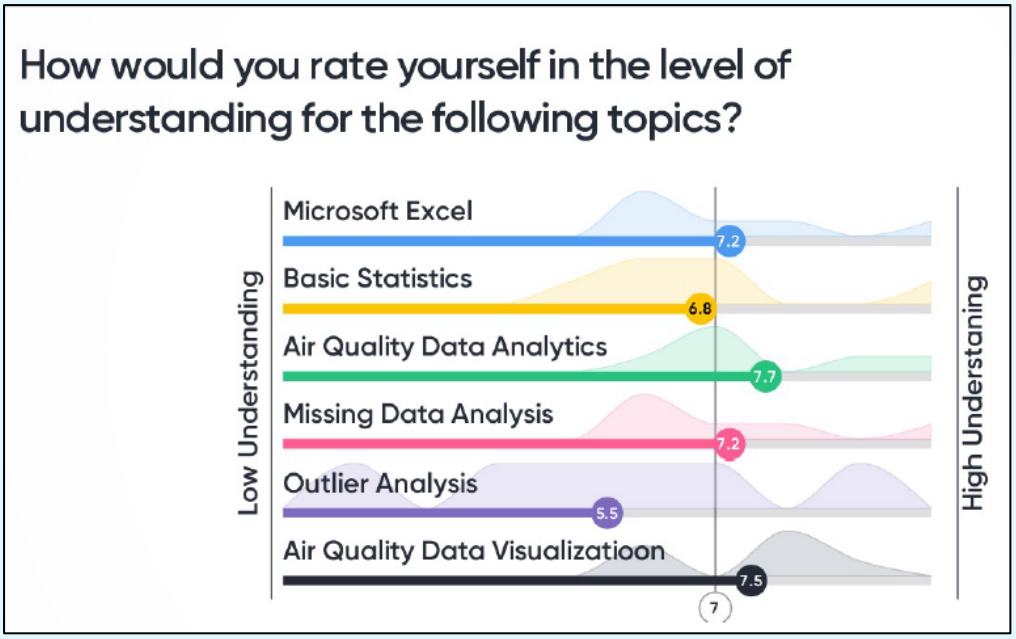


Figure 4 Self-evaluation of participants after the sessions on day 1

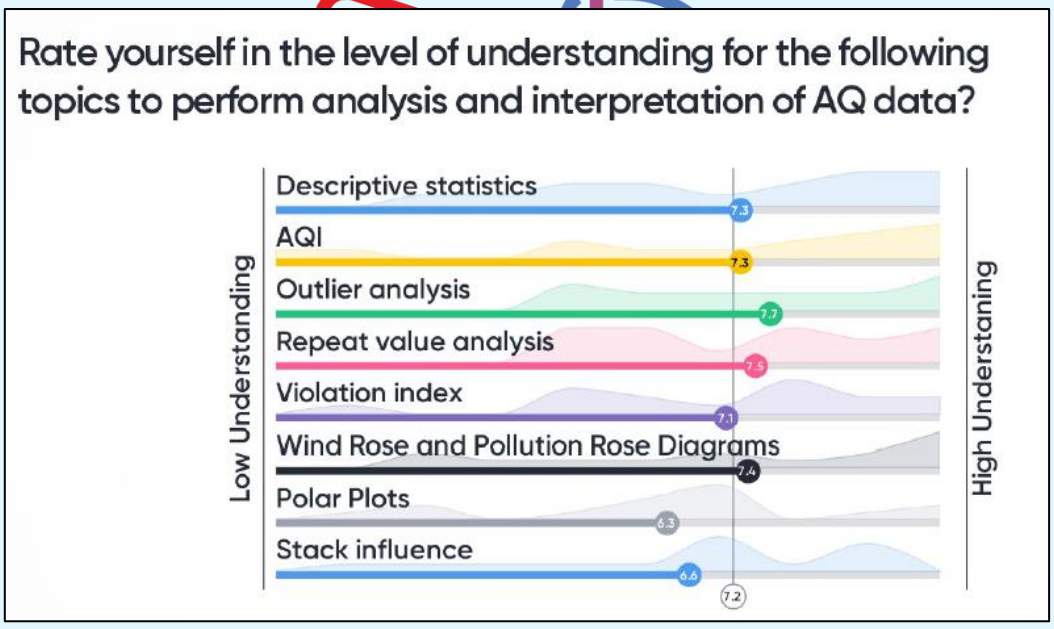


Figure 5 Self-evaluation of participants on day 2

The participants were also asked to rate the overall experience on various aspects of the program like the subject knowledge of trainers, ability to address doubts and the benefit of having a tool like the air quality analytics dashboard. The average rating was 8.9 out of 10. Figure 6 depict show the same.

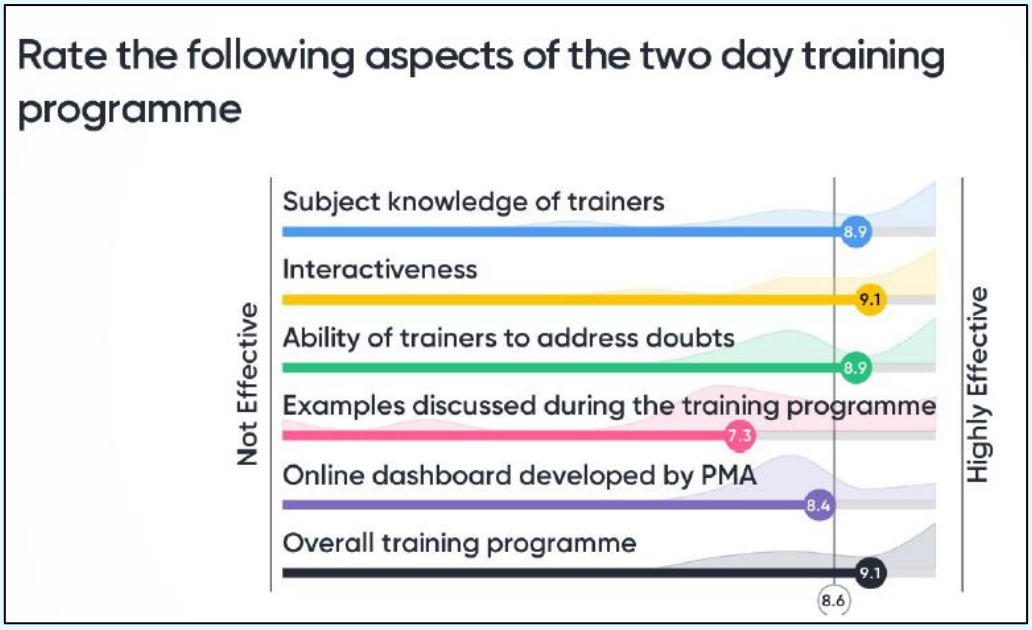


Figure 6 Feedback from participants on the various aspects of the training program

The participants were further asked to describe their overall experience of the two-day training in short sentences or keywords. These responses were collected as word cloud on a real-time basis. The most common words used for describing the programme were *"informative", "interactive" and "highly useful"*. The word cloud is represented in Figure 7.

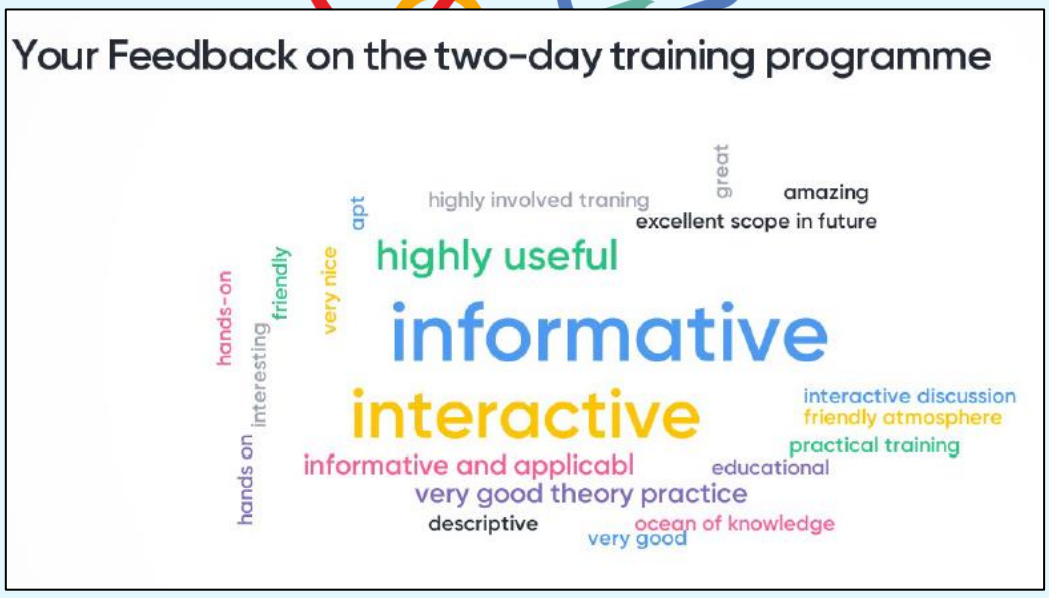


Figure 7 Word cloud of responses from participants on their overall experience of the two-day training programme

APPENDIX-1

Details of participants for the program.

Name	Organization	Category	Email ID
Unnat Shah	Gujarat Pollution Control Board	Regulator	unnatshah22@gmail.com
Dwijen Vyas	Gujarat Pollution Control Board	Regulator	envdwijen@gmail.com
Dr Vimal Poswal	Rajasthan Pollution Control Board	Regulator	vimalposwal007@gmail.com
Prasputita Nanda	Rajasthan Pollution Control Board	Regulator	nprasputita@yahoo.com
Dipsha Shah	Centre for Environment Planning & Technology	Academia	dipsha.shah@cept.ac.in
Archit Uprit	Central Pollution Control Board	Regulator	archituprit.cpcb@nic.in
Gautam Kumar Sharma	Central Pollution Control Board	Regulator	gksharma253@gmail.com
Anurag Sharma	Central Pollution Control Board	Regulator	anurag.cpcb@gov.in
Dr Neel Kamal	National Environmental Engineering Research Institute	Researcher	n_kamal@neeri.res.in
Ms Mayuri Shrirang	National Environmental Engineering Research Institute	Researcher	shrirangmayuri4@gmail.com
Bharat Ji Mehrotra	National Environmental Engineering Research Institute	Researcher	bharatjimehrotra@gmail.com
Nikhil More	Maharashtra Pollution Control Board	Regulator	srohq12@mpcb.gov.in
Mrs Sneha Kamble	Maharashtra Pollution Control Board	Regulator	srohq10@mpcb.gov.in
Hooma Ansari	Enviro Analysts & Engineers Pvt. Ltd.	Consultant	vivek.singh@eaepl.com
Nikhil Pawar	Enviro Analysts & Engineers Pvt. Ltd.	Consultant	nikhil.pawar@eaepl.com
Saurabh Desai	Enviro Analysts & Engineers Pvt. Ltd.	Academia	s.desai@eaepl.com

APPENDIX-2

A sample of answered feedback form.

ekonnect
knowledge foundation

Training Program on Air Quality Data Analytics
Organized by
Ekonnect Knowledge Foundation

Venue: Bandra-Kurla Complex, Mumbai
8th August -9th August 2019

TRAINING FEEDBACK FORM

Name: Sneha D. Kamble Date: 08.08.2019

Day: Day: 01

Training Evaluation Criteria	*Excellent*	*Satisfactory*	*Poor*
1. Course content and structure	✓		
2. Examples cited by the trainers during the training were relevant and enabled better understanding of the concepts		✓	
3. The trainers answered queries posed by the participants satisfactorily	✓		
4. Trainer's skill and competence in training and the subject matter	✓		
5. The presentations were clear and easy to understand	✓		
6. The workshop sessions were engaging and interesting	✓		
7. Overall effectiveness of training to improve my work	Highly effective	Will add value to my work ✓	Not effective
8. Any other comments or suggestions	I would like to understand & know more about how to handle huge quantities of data & analyse them properly, so that we can draw proper conclusions.		

Ekonnect Knowledge Foundation

APPENDIX-3

List of modules in the dashboard.

Module	Description
Descriptive Statistics	Mean, Max/Min, Standard Deviation, Coefficient of Variation.
Box-Whisker plots	Visualization in the form of Box-Whisker Diagram.
Outlier detection	A time series graph that shows outliers or suspect values based on the specified period and parameters.
Missing data Analysis	Summary of percentage of missing data over specified period and parameters, Visualization of missing data over time series.
Correlations	Correlation matrix based on "Pearson" or "Spearman" method.
Air Quality Index	Air Quality Index, or AQI is a simple representation of the quality of air, in terms of a number and colour. AQI will be computed as per the CPCB guidelines.
Time-Series analysis	Diurnal, Daily over a week, Weekends only, monthly, seasonal, annual for single/multiple pollutants and ratios over specified period.
Repeating Values	Spot repeating values in air quality data for specified "n" instances.
Violation Index	An index formulated on the basis of what magnitude (concentration) are the pollutants violating the standards, how many times did the pollutant violate the standards and for how long was the pollutant concentration contiguously violated?
Mann-Kendal trend analysis	Show direction and significance of trend following Mann-Kendall's test for the specified period and parameter.
Meteorological data analysis	Visualization of Wind Rose, Pollution Rose, Persistent wind rose and Polar plots.