

# Contents



Team Information	01
Project Overview	02
Data Cleaning	03
Data Preparation for Modeling	04
Model Training (Linear Regression)	05
Model Evaluation	06
Visualizing Regional Pollution	07
Conclusions and Future Work	08
Q&A	09

## GROUP-03

### MEMBERS

Aditya Singh - 202401100300020

Abhishekh Kumar Maurya- 202401100300008

Ansh Tandon - 202401100300053

Aryan Raj Pandey- 202401100300070

Anurag- 202401100300057

**Submitted to-** Mr Bikki Kumar



# 02

## Project Overview



# Summary of Findings

## Achievements

Data was successfully loaded, cleaned, and prepared, enabling model training. Initial model performance metrics showed meaningful results.

## Variations in AQI

Analysis identified variations in AQI across different cities.

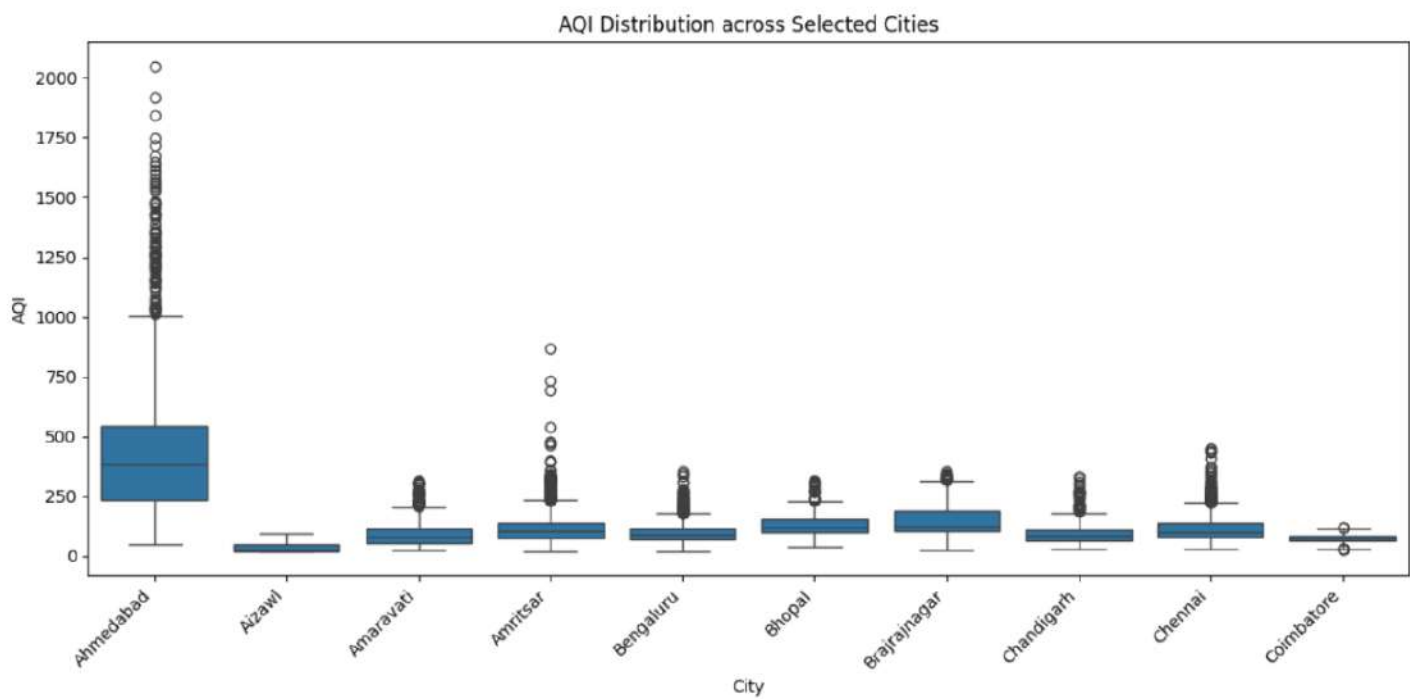
Customer Satisfaction	Recommendation rate
8.1	70%
8.2	71%
8.6	76%
7.9	69%
8.0	70%



# 08 Conclusions and Future Work



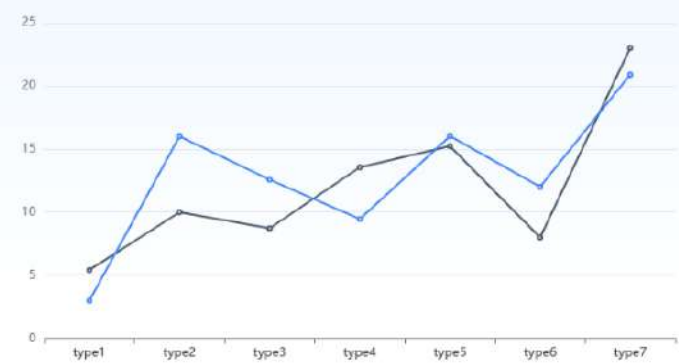
# Visualisation of AQI variation in Cities.



# 07 Visualizing Regional Pollution



# Performance Measurement



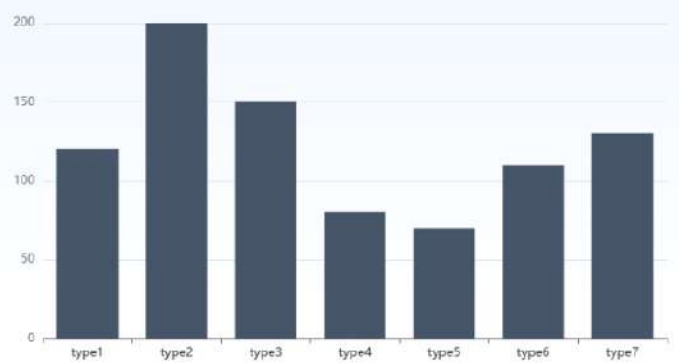
## Evaluation Metrics

Mean Squared Error (MSE) measures the average squared difference between predicted and actual values, with lower being better.



## R-squared (R2)

R-squared measures the proportion of variance in the target variable that can be predicted from features. A value closer to 1 is preferable.



## Results

Present the actual MSE and R2 values obtained during the evaluation.



# 06 Model Evaluation



# Chosen Model



## Model Selection

Linear Regression was selected as a starting point due to its simplicity and interpretability.



## Training Process

The model was trained by fitting it to the training data for prediction learning.

# 05 Model Training (Linear Regression)



# Model Inputs and Output

## Target Variable

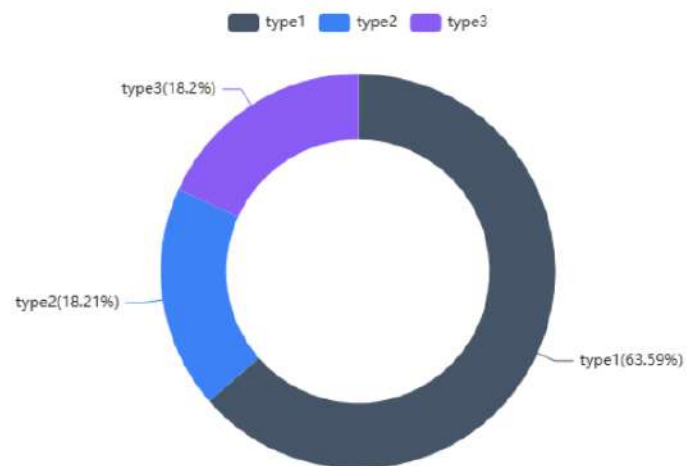
The target variable ( $y$ ) is AQI, which we aim to predict.

## Features

Features ( $X$ ) include numerical air pollutant measurements like PM2.5, PM10, and NO2.

## Data Splitting

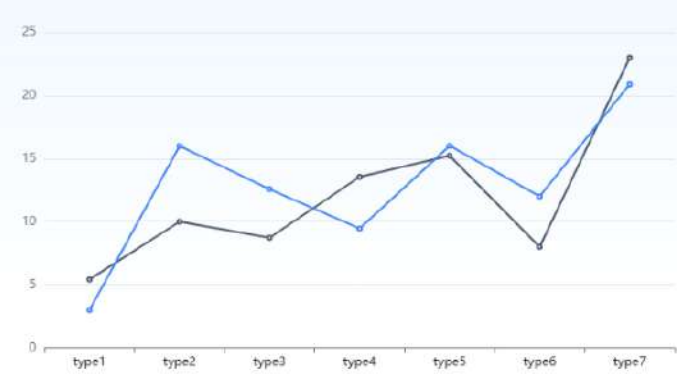
The dataset was split into training (80%) and testing sets (20%) for model evaluation.



# 04 Data Preparation for Modeling



# Importance of Data Cleaning



## Necessity

Handling missing values is crucial for accurate predictions and reliable models.



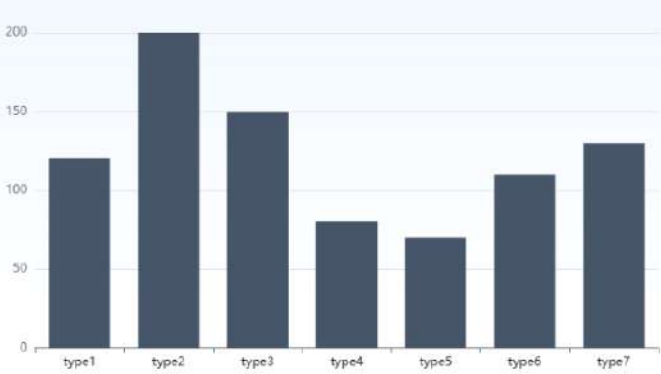
## Method Used

Numerical missing values were imputed with the median, while categorical values were imputed with the mode.



## Results

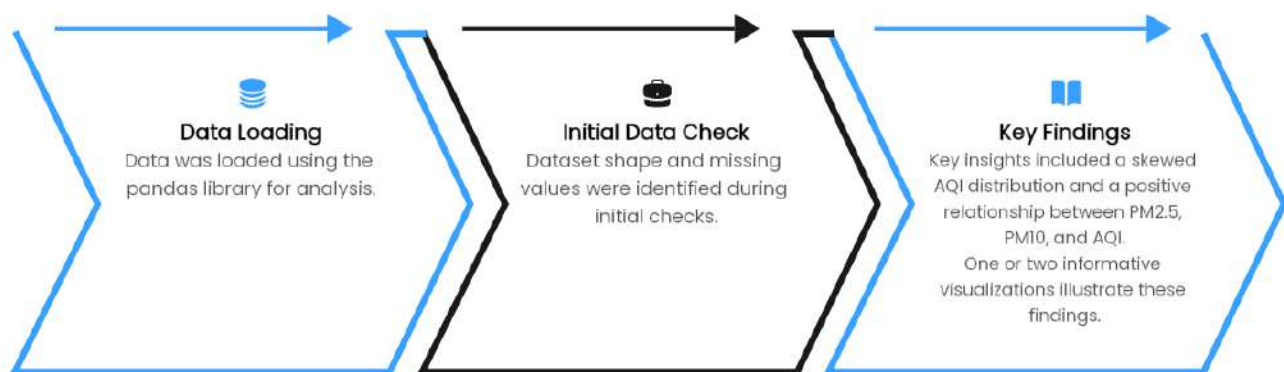
Missing values were successfully handled, ensuring data completeness for modeling.



# 03 Data Cleaning



## Data Loading and Initial Exploration





# Problem Statement

