Analysis and Predictions of Air Quality Index in India

Group 2
CS685 Group Project

Name	Roll No.	Email
Nikhil Agarwal	180475	nikhilag@iitk.ac.in
Supreeth Baliga	180801	supbal@iitk.ac.in
Sarthak Kapoor	180675	sartkap@iitk.ac.in
Jayesh Shaw	180330	jayeshaw@iitk.ac.in
Chinmay Goyal	180206	chinmayg@iitk.ac.in

About the Data

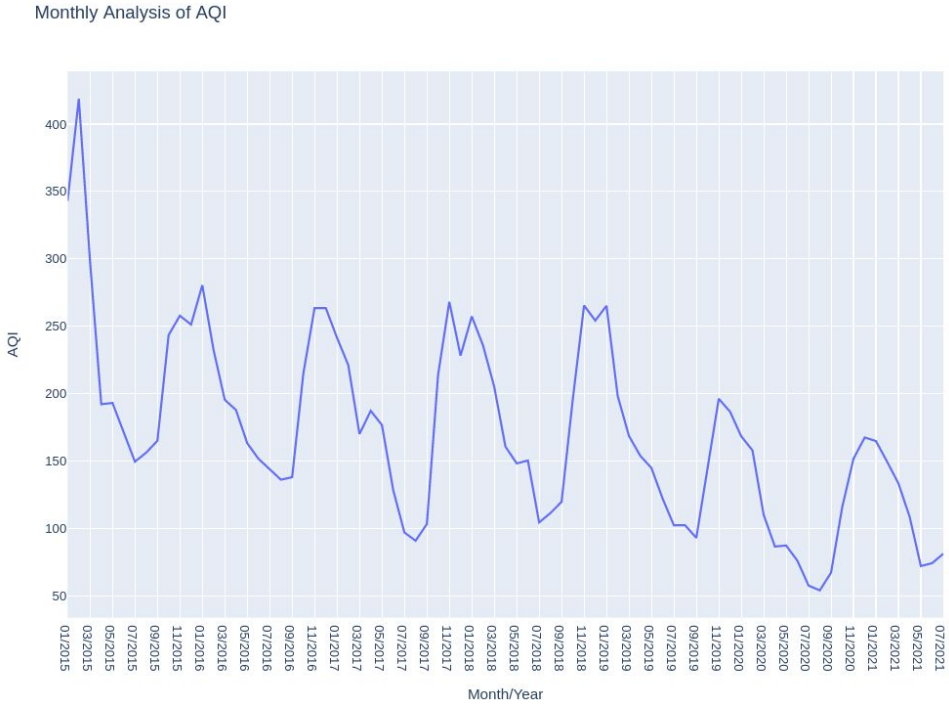
- Extracted from CPCB website by using their application (link here)
- Contains recordings of the concentrations of various pollutants at daily and hourly temporal frequencies
- Analysing data from 1st January 2015 to 1st July 2021
- Chose 25 cities spread over India for analysis

Other Links

- Link to Project Report
- Codebase: https://github.com/nikhilag2711/Project_CS685

Insights & Trends

Does AQI show a seasonality trend?

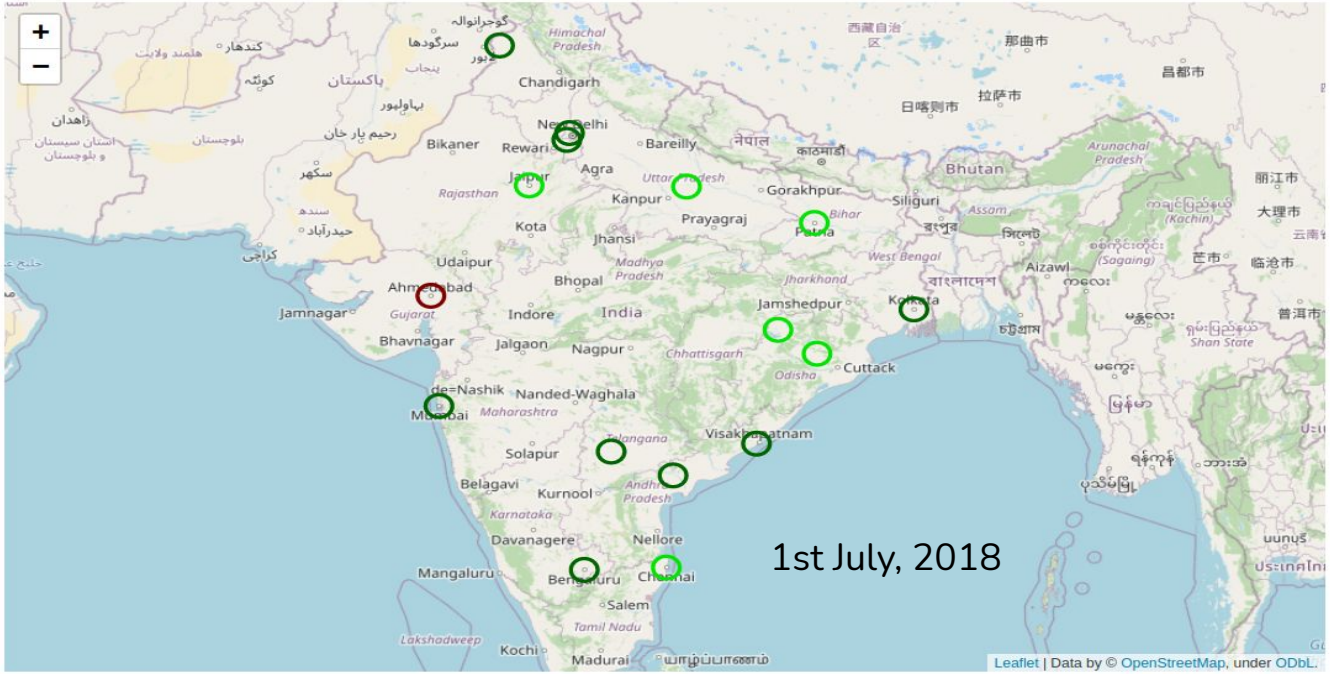


- Air quality is much better during summer as compared to winder
- Factors: Density, Humidity, Diwali
- Many of the individual cities show similar curves
- Lower peaks in 2020 & 2021 due to pandemic

Implications:

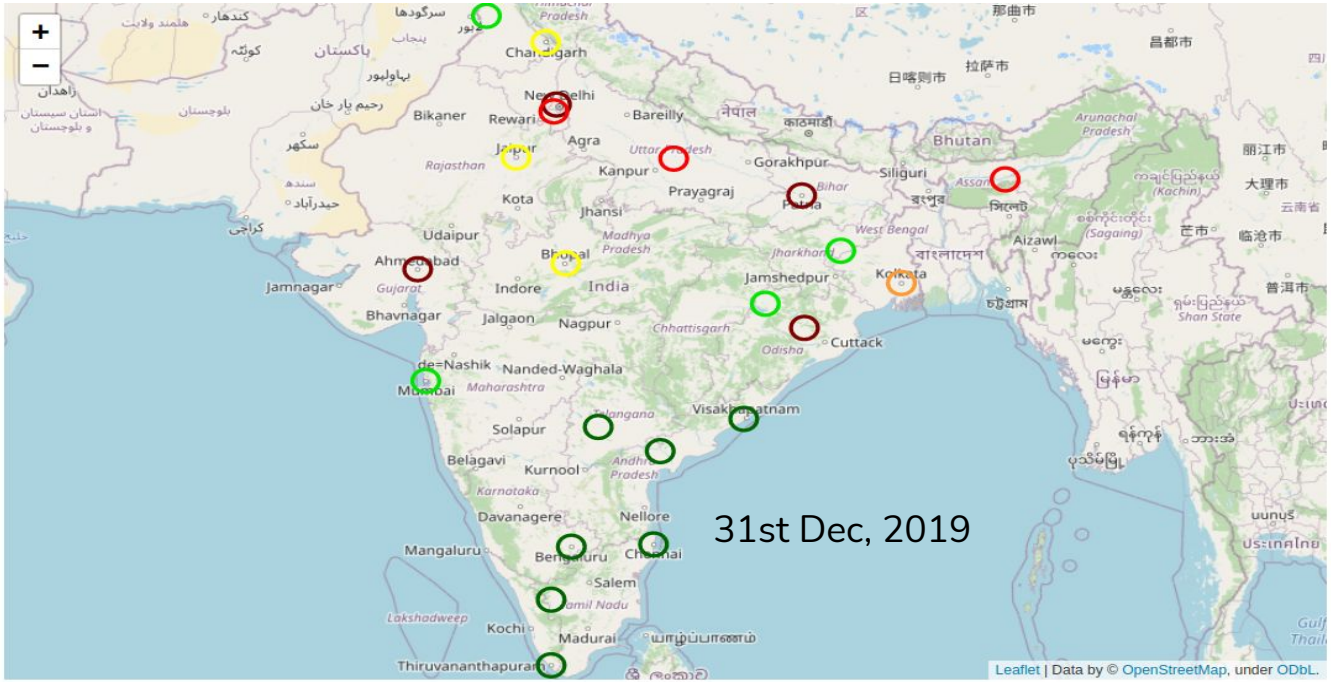
- Policy-makers can strategize to deal with higher air pollution during winter
- Mandate the pollution throughput of industries based on the season

Spatial and temporal comparison of AQI across cities



- High AQI for Ahmedabad since PM10 measured starting from 2019
- Rest of the cities have relatively better quality of air

Spatial and temporal comparison of AQI across cities



- Can see the drastic difference just in a span of 1.5 years
- Poor quality of air in northern cities like Ahmedabad, Delhi, Patna, Lucknow etc.
- Southern India has maintained a better AQI due to close vicinity to ocean
- In general, poorer air quality at locations with high density of industries, as expected

Dates on which cities show maximum & minimum AQI

City	2015	2016	2017	2018	2019	2020	2021
Ahmedabad	23-02-2015	14-07-2016	14-11-2017	19-02-2018	03-01-2019	19-02-2020	20-03-2021
Amaravati	N/A	N/A	08-12-2017	10-12-2018	04-01-2019	27-12-2020	01-01-2021
Amritsar	N/A	N/A	11-05-2017	15-06-2018	28-10-2019	29-11-2020	31-01-2021
Bengaluru	06-07-2015	13-12-2016	06-02-2017	01-05-2018	07-11-2019	02-11-2020	31-03-2021
Bhopal	N/A	N/A	N/A	N/A	28-11-2019	11-11-2020	04-01-2021
Brajrajnagar	N/A	N/A	13-12-2017	12-03-2018	14-01-2019	07-12-2020	03-03-2021
Chandigarh	N/A	N/A	N/A	N/A	28-10-2019	21-12-2020	01-01-2021
Chennai	28-03-2015	21-09-2016	19-10-2017	13-01-2018	27-10-2019	14-01-2020	28-01-2021
Coimbatore	N/A	N/A	N/A	N/A	15-10-2019	25-12-2020	30-01-2021
Delhi	07-11-2015	07-11-2016	09-11-2017	15-06-2018	03-11-2019	09-11-2020	15-01-2021
Gurugram	N/A	03-11-2016	30-08-2017	14-06-2018	03-11-2019	09-11-2020	01-01-2021
Guwahati	N/A	N/A	N/A	N/A	13-08-2019	15-01-2020	05-03-2021
Hyderabad	02-06-2015	16-06-2016	13-01-2017	04-01-2018	07-12-2019	26-12-2020	01-01-2021
Jaipur	N/A	N/A	02-12-2017	21-04-2018	08-04-2019	10-11-2020	30-03-2021
Jorapokhar	N/A	N/A	27-04-2017	11-12-2018	05-01-2019	21-02-2020	17-01-2021
Kolkata	N/A	N/A	N/A	08-11-2018	23-01-2019	29-12-2020	03-01-2021
Lucknow	22-03-2015	08-11-2016	14-11-2017	08-11-2018	02-01-2019	08-07-2020	03-01-2021
Mumbai	N/A	N/A	N/A	10-11-2018	25-12-2019	02-01-2020	06-01-2021
Patna	31-12-2015	01-01-2016	02-12-2017	21-12-2018	02-01-2019	27-12-2020	12-03-2021
Shillong	N/A	N/A	N/A	N/A	05-09-2019	19-02-2020	28-03-2021
Talcher	N/A	N/A	N/A	27-11-2018	14-02-2019	27-01-2020	19-01-2021
Thiruvananthapuram	N/A	N/A	26-11-2017	12-12-2018	01-01-2019	10-02-2020	29-01-2021
Visakhapatnam	N/A	12-09-2016	26-12-2017	08-11-2018	14-01-2019	24-10-2020	13-01-2021

Dates on which cities show maximum AQI

City	2015	2016	2017	2018	2019	2020	2021
Ahmedabad	24-07-2015	15-07-2016	23-10-2017	02-03-2018	23-12-2019	06-07-2020	17-05-2021
Amaravati	N/A	N/A	14-12-2017	16-07-2018	03-08-2019	04-09-2020	05-05-2021
Amritsar	N/A	N/A	29-05-2017	25-09-2018	18-08-2019	25-07-2020	04-02-2021
Bengaluru	18-09-2015	08-06-2016	02-12-2017	17-08-2018	20-07-2019	16-07-2020	30-05-2021
Bhopal	N/A	N/A	N/A	N/A	29-09-2019	23-08-2020	17-05-2021
Brajrajnagar	N/A	N/A	18-12-2017	16-08-2018	28-07-2019	21-08-2020	27-06-2021
Chandigarh	N/A	N/A	N/A	N/A	14-12-2019	08-07-2020	23-04-2021
Chennai	17-09-2015	24-09-2016	14-03-2017	16-08-2018	25-08-2019	03-12-2020	06-05-2021
Coimbatore	N/A	N/A	N/A	N/A	02-12-2019	16-07-2020	02-02-2021
Delhi	23-09-2015	22-08-2016	31-07-2017	24-09-2018	18-08-2019	31-08-2020	20-06-2021
Gurugram	N/A	02-08-2016	02-09-2017	28-07-2018	11-08-2019	31-08-2020	20-06-2021
Guwahati	N/A	N/A	N/A	N/A	03-07-2019	15-07-2020	11-05-2021
Hyderabad	25-07-2015	06-08-2016	17-09-2017	18-07-2018	05-09-2019	05-08-2020	15-06-2021
Jaipur	N/A	N/A	30-06-2017	23-09-2018	10-08-2019	31-08-2020	19-05-2021
Jorapokhar	N/A	N/A	24-07-2017	05-07-2018	10-09-2019	25-08-2020	14-02-2021
Kolkata	N/A	N/A	N/A	16-07-2018	24-09-2019	04-07-2020	17-05-2021
Lucknow	12-07-2015	31-07-2016	29-08-2017	01-08-2018	29-09-2019	07-07-2020	19-06-2021
Mumbai	N/A	N/A	N/A	31-07-2018	28-09-2019	06-08-2020	16-06-2021
Patna	30-10-2015	08-08-2016	03-10-2017	03-09-2018	29-09-2019	22-09-2020	28-05-2021
Shillong	N/A	N/A	N/A	N/A	23-11-2019	05-07-2020	11-05-2021
Talcher	N/A	N/A	N/A	13-10-2018	19-09-2019	23-06-2020	03-04-2021
Thiruvananthapuram	N/A	N/A	01-10-2017	15-07-2018	08-07-2019	17-07-2020	20-05-2021
Visakhapatnam	N/A	21-09-2016	19-11-2017	27-05-2018	23-09-2019	12-10-2020	16-05-2021

Dates on which cities show minimum AQI

- No specific relation between the dates for two different cities
- Reasoning would require extensive research of multiple factors for each city

Correlation among particle concentrations

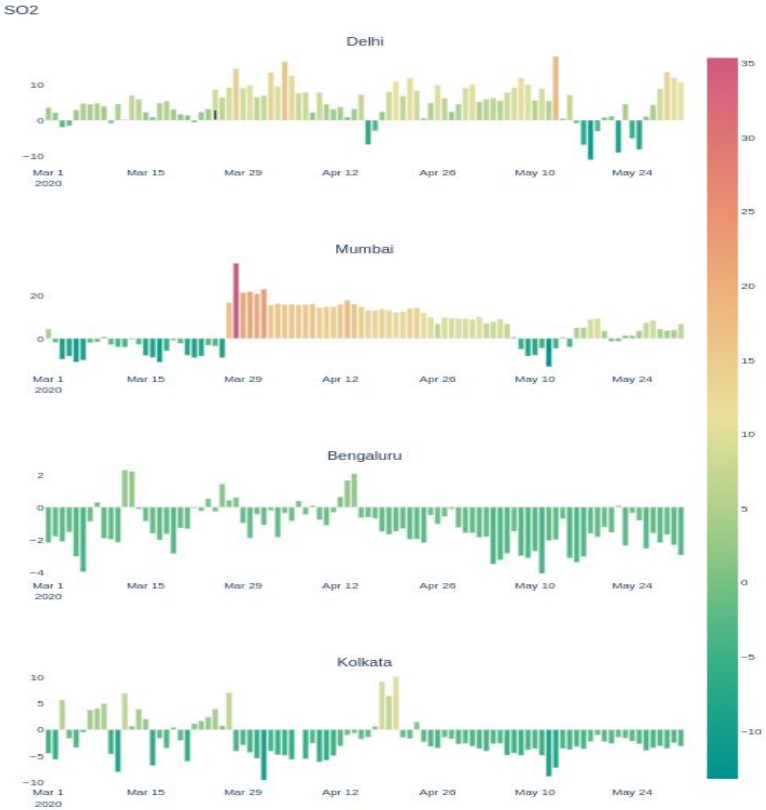
	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	AQI
PM2.5	1.000000	0.871494	0.469176	0.406199	0.475995	0.316530	0.108662	0.147238	0.181193	0.027294	0.140691	0.702553
PM10	0.871494	1.000000	0.519079	0.523979	0.556444	0.394602	0.132690	0.253499	0.258167	0.026689	0.197992	0.844010
NO	0.469176	0.519079	1.000000	0.500006	0.819155	0.220422	0.209814	0.183167	0.005071	0.041333	0.168642	0.462666
NO2	0.406199	0.523979	0.500006	1.000000	0.664503	0.299580	0.347848	0.399479	0.296289	0.027447	0.261889	0.548254
NOx	0.475995	0.556444	0.819155	0.664503	1.000000	0.228318	0.238680	0.255367	0.097633	0.043805	0.208113	0.503932
NH3	0.316530	0.394602	0.220422	0.299580	0.228318	1.000000	0.121293	-0.024005	0.109037	-0.002610	0.034683	0.294658
CO	0.108662	0.132690	0.209814	0.347848	0.238680	0.121293	1.000000	0.495671	0.051548	0.043718	0.237389	0.654849
SO2	0.147238	0.253499	0.183167	0.399479	0.255367	-0.024005	0.495671	1.000000	0.182535	0.031671	0.271102	0.465364
O3	0.181193	0.258167	0.005071	0.296289	0.097633	0.109037	0.051548	0.182535	1.000000	0.013854	0.108078	0.217084
Benzene	0.027294	0.026689	0.041333	0.027447	0.043805	-0.002610	0.043718	0.031671	0.013854	1.000000	0.774409	0.043167
Toluene	0.140691	0.197992	0.168642	0.261889	0.208113	0.034683	0.237389	0.271102	0.108078	0.774409	1.000000	0.257159
AQI	0.702553	0.844010	0.462666	0.548254	0.503932	0.294658	0.654849	0.465364	0.217084	0.043167	0.257159	1.000000

- NOx has a very high correlation with NO and NO2, something which is expected, since NOx is a mixture of these two gases.
- PM2.5 and PM10 have a high correlation. PM2. 5 is the subset of PM10 particles that have aerodynamic diameters less than or equal to 2.5 μm, thus such a result is very plausible.
- Benzene and toluene also have a high correlation. This is due to the fact that both these chemicals are very similar physically and chemically and thus are expected to exist under similar conditions.
- PM2.5, PM10, CO, SO_2, NO_2 have the most significant correlation with AQI while benzene, toluene, ozone and NH_3 have very little correlation with AQI.

Effect of lockdown on AQI

- Due to seasonality of AQI, we have taken difference of concentration of particle in month ‘x’ of 2020 from month ‘x’ in 2019.
- Air quality was found to be much better during the lockdown months compared to same months of 2019.
- Since vehicles are the major source of particles like PM10, PM2.5, NO_2 their concentration is significantly reduced.
- For SO_2 the concentration remains almost same or slightly increases for all cities since its major source is coal plants.

Effect of lockdown on AQI

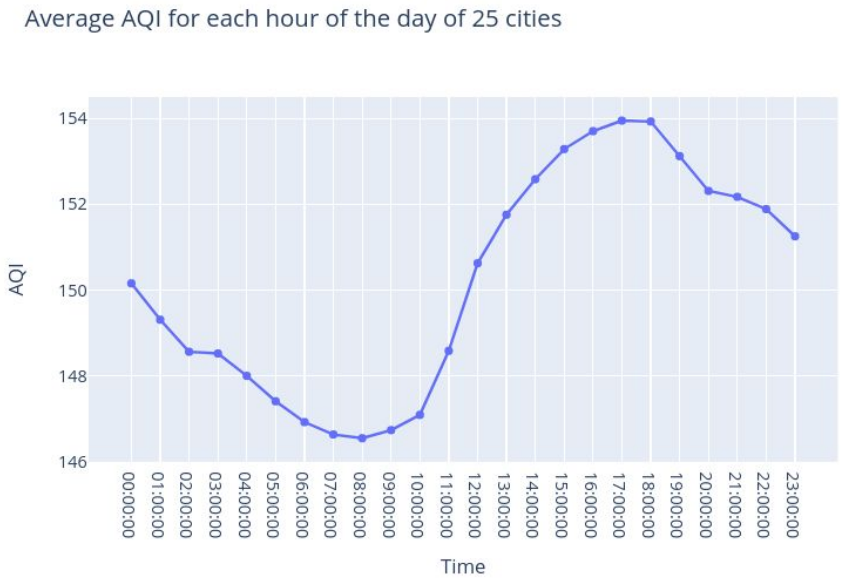


- The analysis has been done for both first wave (end March 2020 - May 2020) and second wave (April 2021 - June 2021).
- Both the time periods followed very similar trends as mentioned above.
- The same steps were followed for 4 major cities of India to ensure no outliers are present.

Implications:

- The results show that policy makers should emphasize on curbing vehicle usage since its reduction for only a couple of months had miraculous effect on AQI.

Is AQI high during daytime or night?



- AQI of Ahmedabad is very high so need to remove it from the dataset for this insight
- AQI is higher at daytime and evening
- Many of the individual cities show similar curves
- Lower AQI in night due to less vehicle count and industries operating at that time

Implications:

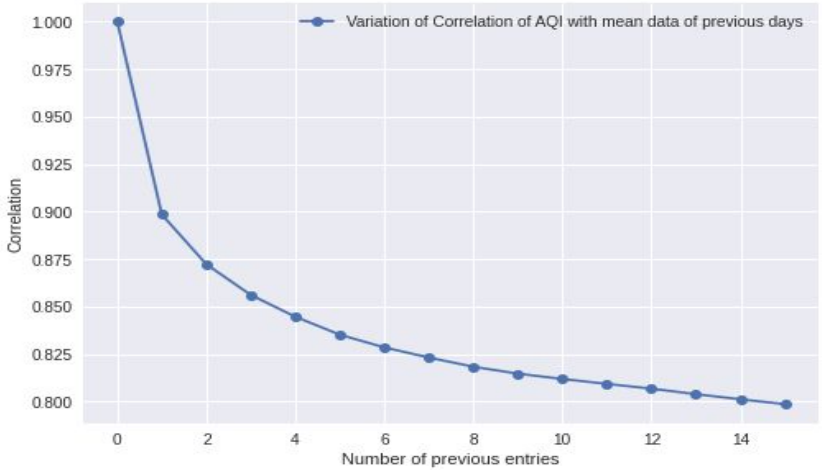
- Policy-makers can control the number of vehicles during the daytime
- Some of the industries can be operated in night rather than daytime to balance the air pollution

Particle with highest concentration in each city

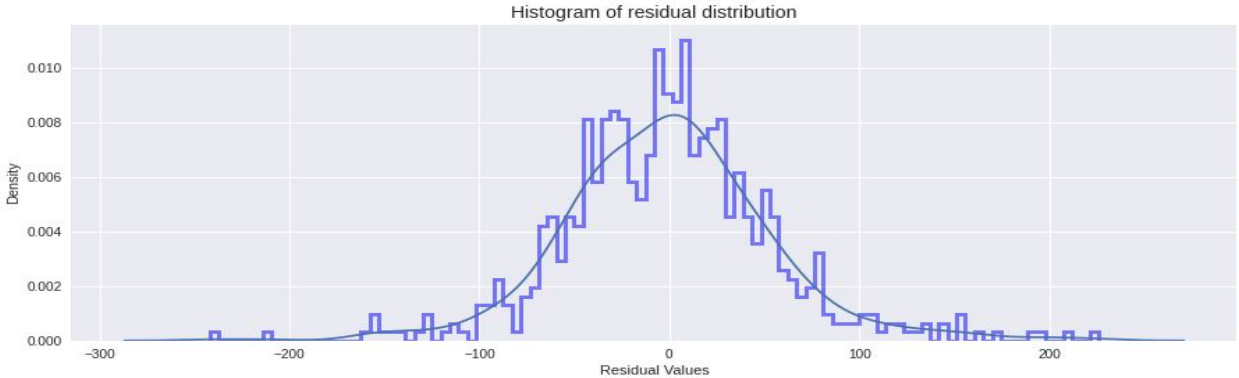
- Different particle's concentrations cannot be compared directly due to different units and different level of variations for high pollution for each of them.
- For most of the cities either PM2.5 or PM10 has the highest concentration.
- Ahmedabad has highest CO concentration due to highest number of vehicles and power plants.

Predictive Model

Linear Prediction Model



- Features based on average of previous day data have a good correlation with the next day AQI which suggests the set of features



- Plotting residuals shows that Linear Model can appropriately model the data

Comparison of Final Prediction