Statistical Analysis of the Factors attributing AQI Values of Major Cities in India

Team Members:
Hrishita Bapuram
Maryam Amir Ahmad
Shraddha Kodavade

Mentored by :
Dr. Hemant Kulkarni

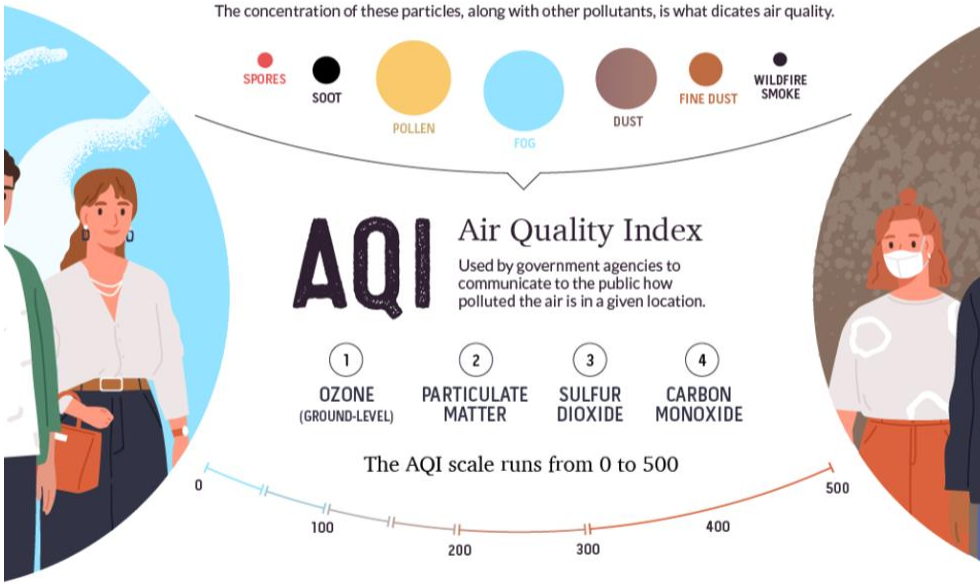
INTRODUCTION

- Air pollution is a burgeoning environmental problem with potential for climate change.

- Introspection into the various causes and mechanisms responsible for this sorry state of affairs, its impact on human health, and possible solutions to the problem is being studied in this project.

What is AQI (Air Quality Index)?

- Air Quality Index (AQI) is a tool to showcase air quality status.
- It transforms complex air quality data of various pollutants into a single number and color.
- The Environmental Protection Agency (EPA) calculates the AQI for five major air pollutants, for which national air quality standards have been established to safeguard public health.
    - Ground-level ozone
    - Particle pollution/particulate matter (PM2.5/pm 10)
    - Carbon Monoxide
    - Sulfur dioxide
    - Nitrogen dioxide

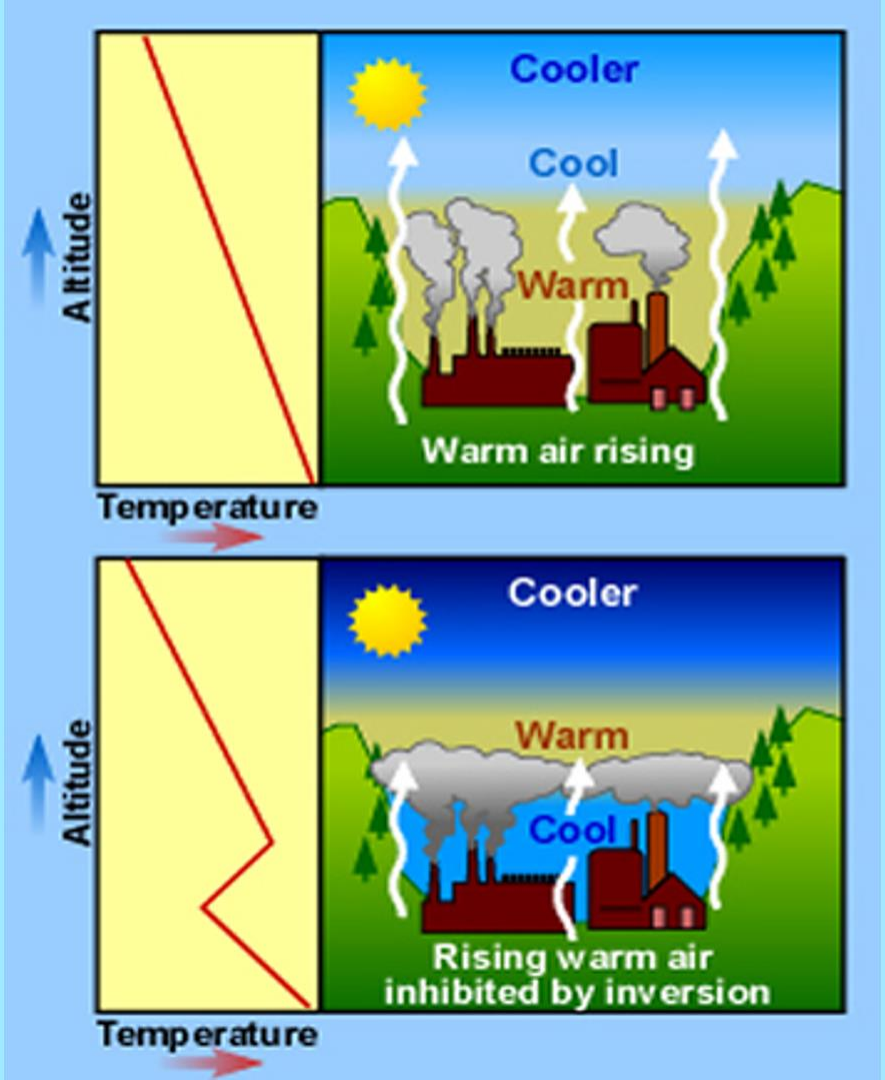


OBJECTIVES

1. To test the effect of **rainfall** and **temperature** on the AQI values of the cities studied
2. To test the effect of **Forest Area** and **Region (N/S)** on the AQI values of the cities studied
3. To test the effect of **Car density** and **Tier wise distribution of cities** on the AQI values of the cities studied
4. To study the **choice of cities** for individuals **before** and **after** informing them about the respective city's AQI levels

1.

To test the effect of **rainfall** and **temperature** on the AQI values of the cities studied



HYPOTHESES

NULL HYPOTHESES:

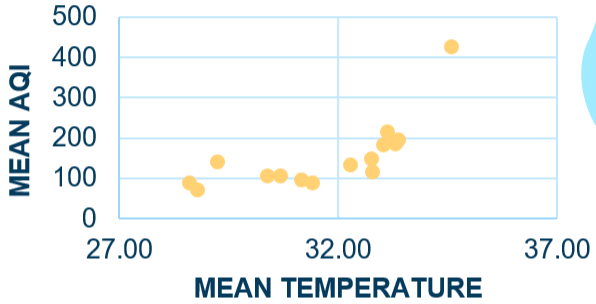
H_{01} : Air Quality is not significantly affected by changes in temperature conditions
H_{02} : Air Quality is not significantly affected by changes in rainfall

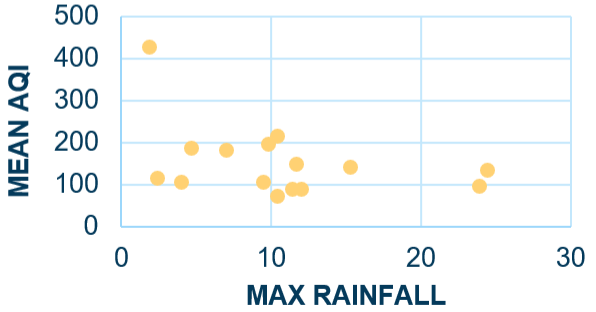
ALTERNATE HYPOTHESES:

H_{11} : Air Quality differs significantly due to temperature
H_{12} : Air Quality differs significantly due to rainfall

TEST USED:

Two Way ANOVA (Two Way Analysis of Variance)

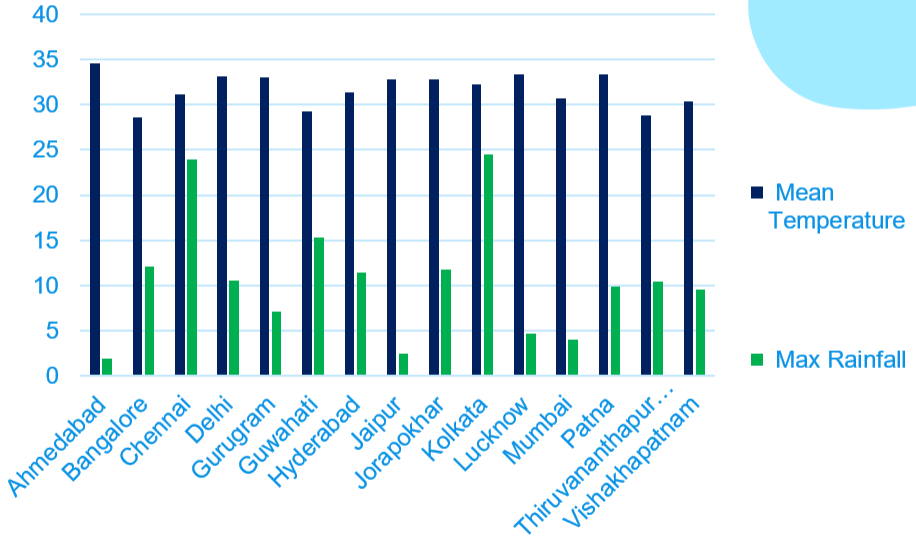




DATA

MEAN TEMPERATURE	
28-32	A
32-35	B

MAX RAINFALL	
1-10	Low
10-15	Moderate
15-25	High



DATA DISTRIBUTION

`	Low	Moderate	High	$T_{i.}$	$n_{i.}$
A	104.8398	82.4177	117.7744	305.0319	3
B	219.7202	180.5513	131.2134	531.4849	3
$T_{.j}$	324.5600	262.9690	248.9878		
$n_{.j}$	2	2	2		

G^2/N	116626.732
$\sum T_{i.}^2 / m_{i.}$	125173.5542
$\sum T_{.j}^2 / m_{.j}$	118243.4145

TSS	13120.84094
SST	8546.822189
SSB	1616.682518
SSE	2957.336237

N	6
$\sum y_{ijk}^2$	129747.5729
G	836.5168
h	3
k	2

ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DF	MS	F VALUE	F TABULATED ($\alpha = 0.1$)
Treatment	8546.822189	1	8546.822189	$F_1 = 5.780081469$	8.53
Block	1616.682518	2	808.3412591	$F_2 = 0.546668484$	9
Error	2957.336237	2	1478.668118		
Total	13120.84094	5			

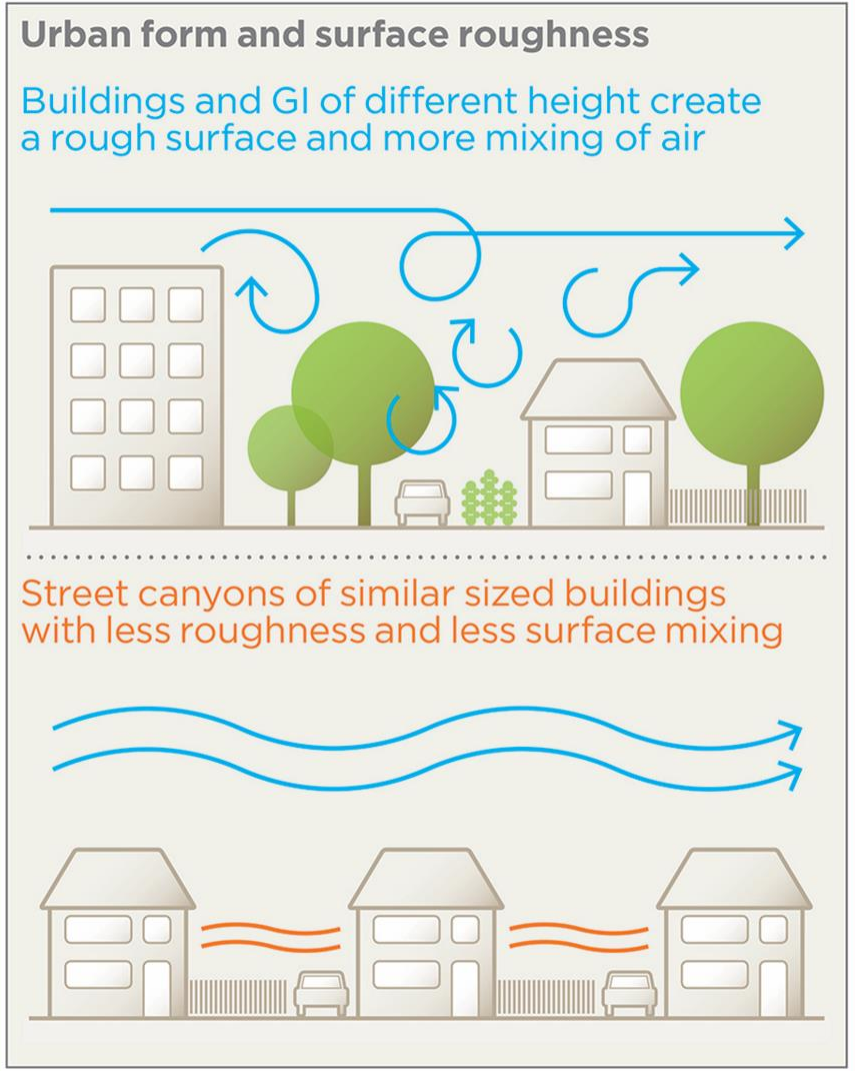
CONCLUSION:

Since $F_1 < F_{0.1}(1,2)$ and $F_2 < F_{0.1}(2,2)$, we **fail to reject** H_{01} and H_{02}

Therefore, the data indicates that Temperature and Rainfall **do not affect** Air Quality Index at 10% level of significance.

2.

To test the effect of **Forest Area** and **Region (N/S)** on the AQI values of the cities studied



HYPOTHESES

NULL HYPOTHESES:

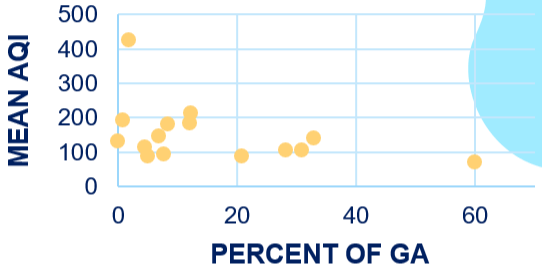
H_{01} : Air Quality is not significantly affected by changes in Green Area
H_{02} : Air Quality is not significantly affected by different regional differences

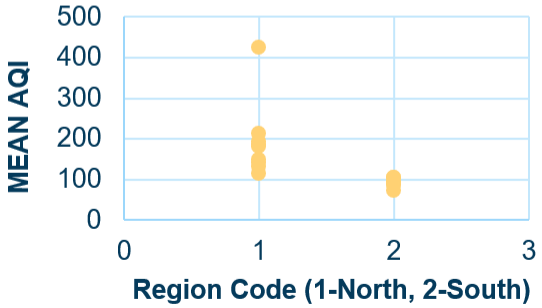
ALTERNATE HYPOTHESES:

H_{11} : Air Quality differs significantly due to proportion of green area in the city
H_{12} : Air Quality differs significantly due to the presence of city in either North or South of India

TEST USED:

Two Way ANOVA (Two Way Analysis of Variance)

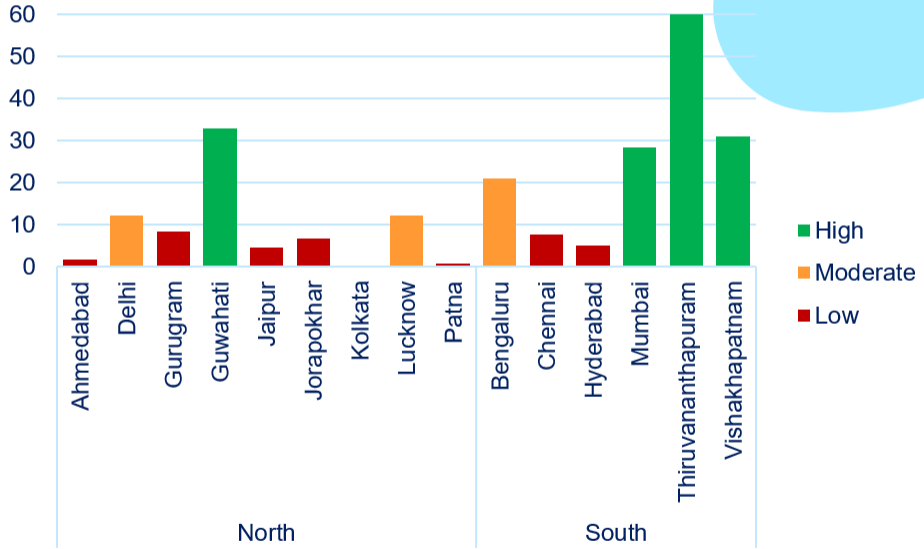




DATA

REGION
North
South

% OF GREEN AREA	
0-10	Low
10-25	Moderate
>25	High



DATA DISTRIBUTION

	North	South	$T_{i.}$	$m_{i.}$
A	198.788	91.581	290.369	2
B	199.092	87.559	286.651	2
C	140.512	93.749	234.261	2
$T_{.j}$	538.392	272.889		
$m_{.i}$	3	3		

G^2/N	109696.1435
$\sum T_{i.}^2 / m_{i.}$	123740.4486
$\sum T_{.j}^2 / m_{.j}$	811.281

TSS	14044.3051
SST	984.440548
SSB	11748.6405
SSE	1311.224052

N	6
$\sum y_{ijk}^2$	123740.4486
G	811.281
h	2
k	3

ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DF	MS	F VALUE	F TABULATED (α = 0.1)
Treatment	984.440548	2	492.220274	0.750779813	9
Block	11748.6405	1	11748.6405	17.92011134	8.53
Error	1311.224052	2	655.612026		
Total	14044.3051	5			

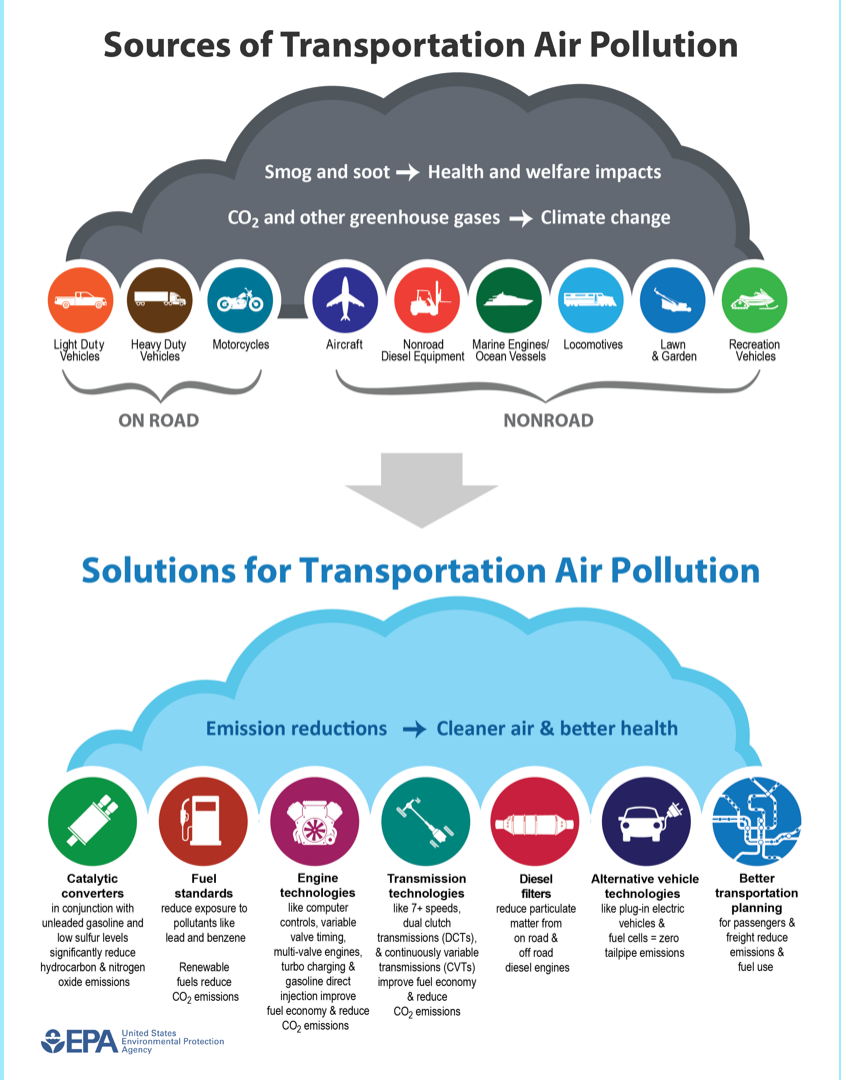
CONCLUSION:

Since $F_1 < F_{0.1}(2,2)$ and $F_2 > F_{0.1}(1,2)$, we **fail to reject** H_{01} and we **reject** H_{02}

Therefore, the data indicates that Percent of Green Area **does not affect** Air Quality Index and AQI **differs significantly** in the North and South of India at 10% level of significance.

3.

To test the effect of **Car density** and **Tier-wise distribution of cities** on the AQI values of the cities studied



HYPOTHESES

NULL HYPOTHESES:

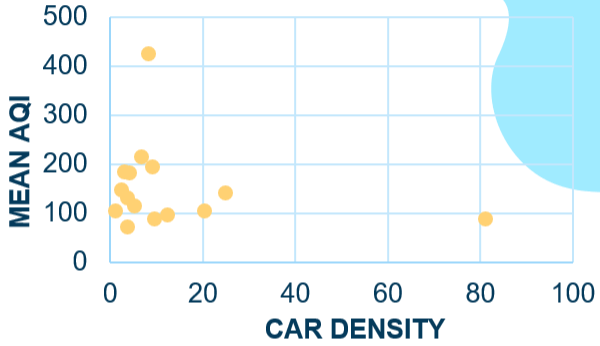
H_{01} : Air Quality is not significantly affected by changes in Car Density
H_{02} : Air Quality is not significantly affected by changes in population of the city

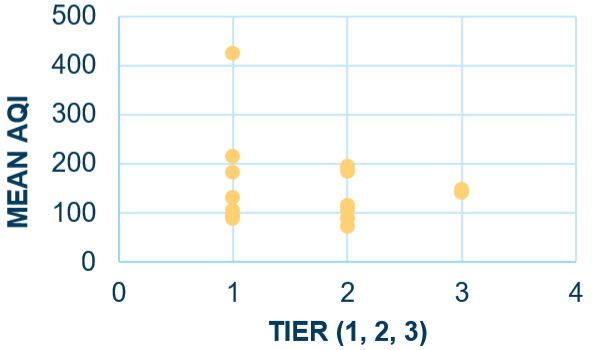
ALTERNATE HYPOTHESES:

H_{11} : Air Quality differs significantly due to car density in the cities
H_{12} : Air Quality differs significantly due to the population of the cities divided into tiers.

TEST USED:

Two Way ANOVA (Two Way Analysis of Variance)

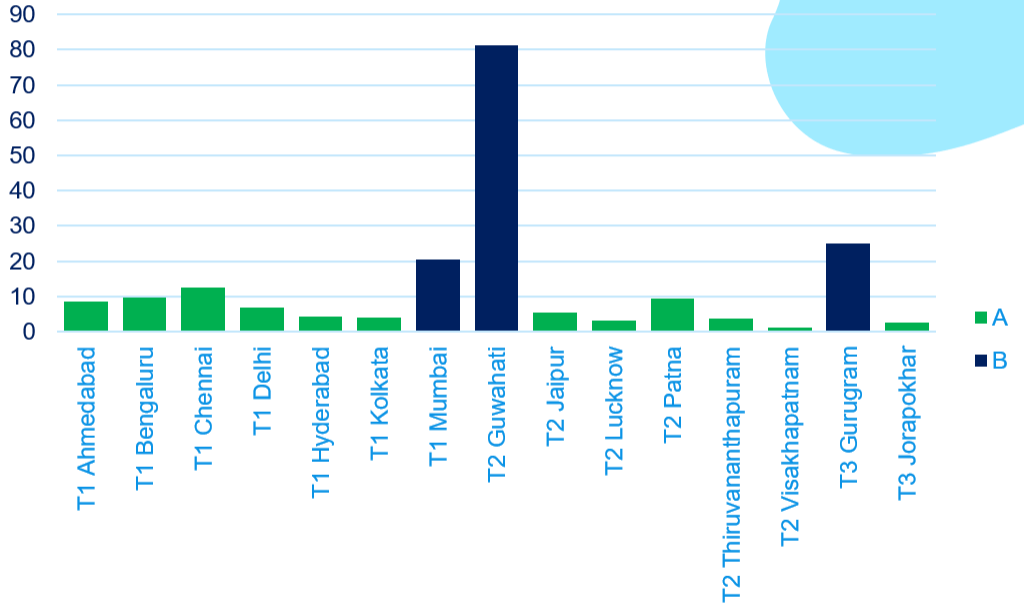




DATA

CAR DENSITY	
0-20	A
>20	B

TIER CATEGORY	
TIER 1	T1
TIER 2	T2
TIER 3	T3



DATA DISTRIBUTION

`	T1	T2	T3	$T_{i.}$	$n_{i.}$
A	188.92634	133.7808	147.43107	470.1382	3
B	104.86778	88.12546	88.125461	281.1187	3
$T_{.j}$	293.79413	221.9063	235.55653		
$n_{.j}$	2	2	2		

G^2/N	94064.49131
$\sum T_{i.}^2 / m_{i.}$	100019.2198
$\sum T_{.j}^2 / m_{.j}$	95522.12702

TSS	7791.338672
SST	5954.728503
SSB	1457.635708
SSE	378.9744613

N	6
$\sum y_{ijk}^2$	101855.83
G	751.256912
h	3
k	2

ANOVA TABLE

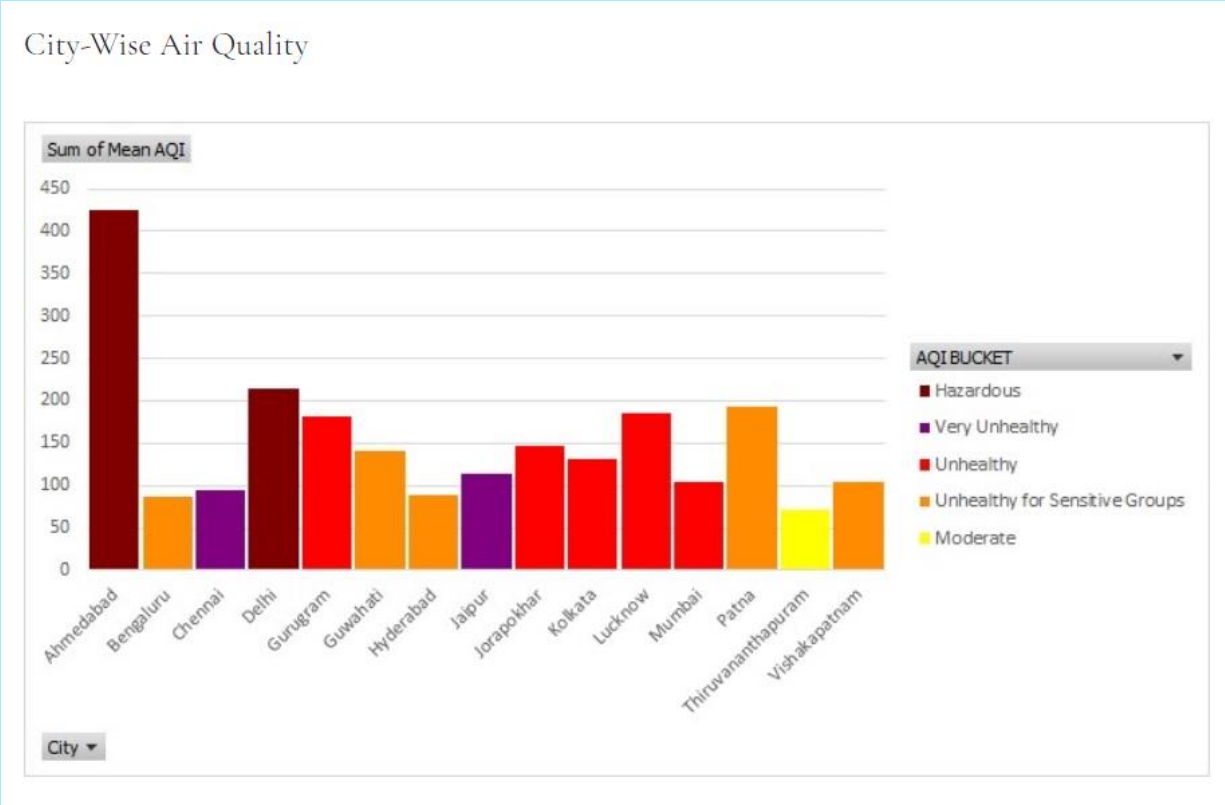
SOURCE OF VARIATION	SUM OF SQUARES	DF	MS	F VALUE	F TABULATED ($\alpha = 0.1$)
Treatment	5954.728503	1	5954.728503	31.425487	8.53
Block	1457.635708	2	728.8178538	3.846263684	9
Error	378.9744613	2	189.4872306		
Total	7791.338672	5			

CONCLUSION:

Since $F_1 > F_{0.1}(1,2)$ and $F_2 < F_{0.1}(2,2)$, we **reject** H_{01} and we **fail to reject** H_{02}

Therefore, the data indicates that Car Density **affects** Air Quality Index and Tier-wise differences **do not affect AQI** at 10% level of significance.

4. To study the **choice of cities** for individuals **before** and **after** informing them about the respective city's AQI levels



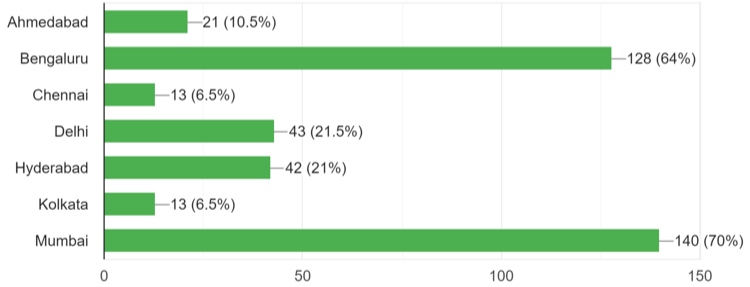
Questionnaire

TIER 1 CITIES CHOICE

WITHOUT AQI KNOWLEDGE

Pick any TWO of the following TIER 1 cities in which you'd want live in

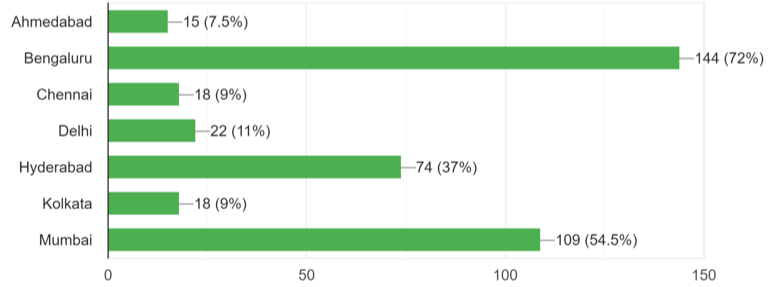
200 responses



WITH AQI KNOWLEDGE

With the above knowledge, pick any TWO of the following TIER 1 cities in which you'd want live in

200 responses



TIER 1 CITIES CHOICE

Rank Tier 1	City	City	Mean AQI	Category
1	Bengaluru	Bengaluru	87.55941	G
2	Hyderabad	Hyderabad	88.12546	G
3	Chennai	Chennai	95.03663	G
4	Mumbai	Mumbai	104.8678	G
5	Kolkata	Kolkata	131.2134	B
6	Delhi	Delhi	213.6716	B
7	Ahmedabad	Ahmedabad	424.9242	B

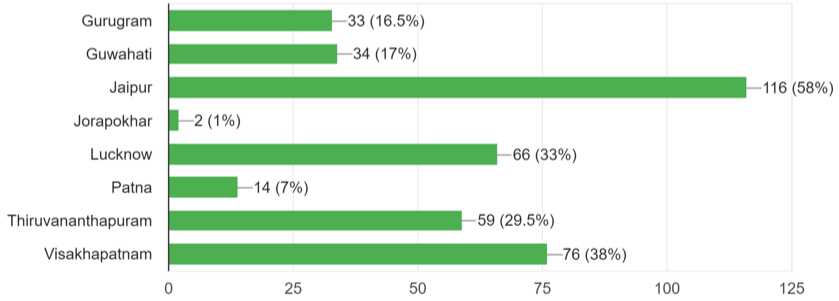
		AFTER		
		GOOD	BAD	TOTAL
BEFORE	GOOD	25	2	27
	BAD	152	21	173
	TOTAL	177	23	200

TIER 2 & 3 CITIES CHOICE

WITHOUT AQI KNOWLEDGE

Pick any TWO of the following TIER 2 and TIER 3 cities in which you'd want live in

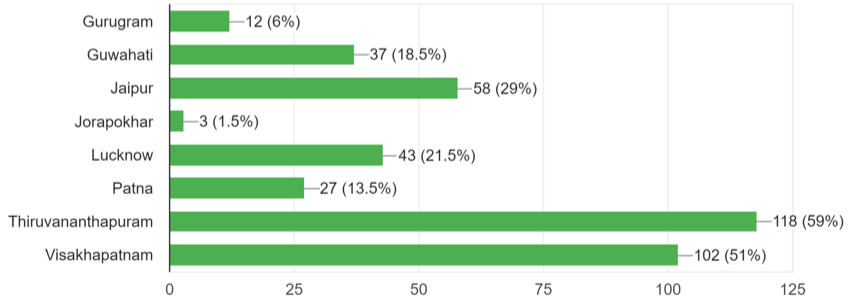
200 responses



WITH AQI KNOWLEDGE

Similarly, with the above knowledge, pick any TWO of the following TIER 2 and TIER 3 cities in which you'd want live in

200 responses



TIER 2 & 3 CITIES CHOICE

Rank Tier 2/3	City	City	Mean AQI	Category
1	Thiruvananthapuram	Thiruvananthapuram	71.56814	G
2	Vishakapatnam	Vishakapatnam	104.8119	G
3	Jaipur	Jaipur	114.2921	G
4	Guwahati	Guwahati	140.5122	B
5	Jorapokhar	Jorapokhar	147.4311	B
6	Gurugram	Gurugram	181.1529	B
7	Lucknow	Lucknow	184.513	B
8	Patna	Patna	193.7189	B

		AFTER		
		GOOD	BAD	TOTAL
BEFORE	GOOD	71	3	74
	BAD	108	18	126
	TOTAL	179	21	200

LIMITATIONS

- The study was not inclusive of all of the states in India. (We chose just 15 cities because of unavailability of data for other cities and hence the data was not representative of India
- Data was just restricted till Jun 2020
- All the factors were assumed to be independent of each other
- Car registration was considered constant due to the lack of unavailability of data
- Only a limited number of factors were considered.

FUTURE SCOPE

- Studying the **relationship** between **other contributors** to the AQI , i.e. Wind Speed Direction, Precipitation rate, Sun Radiations, Water Body distribution etc.

- Use **dashboard visualizations** like scatter plots to intuitively study of dependency of two or more variables and hence establish the interdependence among them. (Can hence be mathematically verified).

- To further contribute in the **research dimension** wherein results derived can be used to **study economic indicators** like QOL (Quality of Life) – HDI, Pollution Levels and others that contributes to further navigate the living index of a city and the healthcare of its territorial citizens.

- **Statistically deduce** if the formulations and policies have worked, and hence reason the deviations, if found.

- Many other factors like **odd-even rule**, other **festivals and occasions** (Holi, Diwali, New Year, etc), seasonal comparison (Summer vs Winter), Global Warming, Trading Peaks, proximity to airports and railway stations etc. can be examined.

REFERENCES

* data.gov.in – Open Government Data (OGD) Platform, India
* moef.gov.in – Ministry of Environment, Forest and Climate Change
* cpcb.nic.in – Central Pollution Control Board of India
* airnow.gov – Home of U.S. Air Quality Index
* scied.ucar.edu – UCAR Center for Science Education
* nrdc.org – National Resources Defence Council
* the-ies.org – The Institution of Environmental Sciences

ACKNOWLEDGMENTS

- *We would like to acknowledge and express our gratitude to our Mentor for this project, Dr. Hemant Kulkarni for helping us structure, format and understand the concepts and techniques employed in the project.*
- *We would like to thank NMIMS deemed to be university, Navi Mumbai for introducing us to the opportunity to work together and for providing us with all the necessary resources that were essential for carrying out this research.*
- *We would also like to express our gratitude to our college Library, for having provided us with the data sources that aided in the completion of our project.*
- *We acknowledge the contributions of our amazing respondents who took out the time to fill our survey so that we could get the primary data and helped us get the responses that were generated.*
- *Finally, we would like to express our gratitude to the various sources from which we extracted the data for this project, without which none of this would truly be possible.*

THIS IS OUR TEAM

HRISHITA BAPURAM

MARYAM AMIR AHMAD

SHRADDHA KODAVADE

“Just breathing can be such a luxury sometimes”

~ Walter Kirn

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Thanks!