DATA DRIVEN GOVERNANCE PROJECT REPORT

ON

ANALYZING THE RELATIONSHIP BETWEEN VEHICLES AND POLLUTION

Done By

BHAVIK DESAI (PRN: 20070243009)

KANIKA MEHTA (PRN: 20070243011)

ANAGHA PATIL (PRN: 20070243018)

SARAN S P. (PRN: 20070243028)

SHAH MANAV (PRN: 20070243030)

॥वसुधैव कुटुम्बकम्॥

DEPARTMENT OF DATA SCIENCE

SYMBIOSIS INSTITUTE OF GEO INFORMATICS

(Affiliated to Symbiosis University)

DECLARATION

We hereby declare that the project work entitled "**ANALYZING THE RELATIONSHIP BETWEEN VEHICLE REGISTRATION AND POLLUTION**" submitted to Symbiosis University in the partial fulfilment of the requirements for the award of MSc. Data Science is a record of original work submitted under the guidance of Mrs. APARNA JOSHI, Professor, Department of Data Science, Symbiosis Institute of Geoinformatics during our period of study in Symbiosis Institute of Geoinformatics, Pune.

We also declare that this project has not been submitted to any other university or institution for the award of any degree or diploma.

Place: PUNE
Date: 22 May 2021

ANAGHA PATIL

BHAVIK DESAI

KANIKA MEHTA

SARAN S P

SHAH MANAV

ABSTRACT

The project is "analyzing the relationship between pollution and vehicle registration" . In this project we focus on analyzing the relationship between pollution and vehicle registration in India. For the analysis we have taken 4 years of data that is from (2011-2014) of pollution reported and vehicle registration taken place across the 33 states of India. We will be analyzing the correlation between the components of pollutants and analyze the trends of rate of change of pollution taking place in each state and compare the pollution rates after classifying the states into 4 sectors North, South, East and West after comparing the classified sector using 4 years of data we will predict the next 6 years of data that's from (2015-2020) and analyze the difference and changes that may take place during the upcoming six years in the country and visualize using maps. And to study the second phase of project we have taken data set of carbon emission of 16 major states in India for 6 years (2009-2014) and data of vehicle registration taken place from the year 2009-2015 and basis of that data we are going to predict the expected carbon emission which may take place during the year 2015. For conducting our study, we are using Linear Regression, Label Encoder algorithms and Tableau for visualization.

Table of Contents

Chapters	Particulars		Page no.
1	**Introduction**		5
2	**Flowchart**		14
3	**Methodology**		16
	3.1	Datasets used	
	3.2	Relevant relationships between datasets	
4	**Analysis**		21
	4.1	Pollutant's analysis	
	4.2	Relationship between vehicle registered and carbon pollutants	
5	**Conclusion**		49
6	**Insights for data driven governance**		50
7	**Lessons learned**		51
8	**Bibliography**		52

INTRODUCTION

The atmosphere is an invisible substance around the Earth and gives us all the oxygen we breathe and plays a significant role in supporting life on Earth. But over time, fresh and clean air gradually becomes increasingly polluted because of increased air pollution. Air pollution is the presence of one or more substances within the concentration area above their natural levels, which has the potential to supply a negative effect. Air pollution is one of the most serious environmental problems in urban areas where most people are exposed to poor air quality. The rapid urbanization in India has resulted in a dramatic increase in the number of vehicles. As the number of vehicles continues to grow and congestion increases, cars are now becoming a major source of air pollution in Indian cities.

Rapid urbanization and growth of automobiles impose a significant effect on human life and its environment in recent years. Most of the cities of India are being suffered by extremely high levels of urban air pollution particularly in the form of CO2, SO2, NO2 and PM (Particulate Matter). Transport sectors contribute a serious share to environmental pollution (around 70%). Among these pollutants CO2 is the major pollutant coming from the transport sector, contributing 90% of total emission. Hydrocarbons are next to CO2. It is indeed interesting to study that the contribution of the transport sector to the particulate pollution is as less as 3-5%, most of the SPM (Suspended Particulate Matter) are generated thanks to re-suspension of dust out of which PM10 is that the most prominent air pollutant. NOx is another important air quality indicator. All these situations indicate that air pollution is becoming a major problem in Indian context and there is an essential need to build up a healthy environment and increase the level of research around the world. Most Indian cities are experiencing rapid urbanization and most of the country’s population is predicted to measure in cities within a span of next 20 years. The rapid development in urban India has also resulted in a tremendous increase within the number of automobiles and in some cities, this has doubled within the last decade. This is the main source of air pollution and poor ambient air quality impacting millions of dwellers.

Air pollution is one among the intense environmental concerns of the urban Asian cities including India, where the majority of the population is exposed to poor air quality. It has also resulted in a tremendous increase within the number of automobiles. The vehicle fleets have even doubled in some cities within the last one decade. This increased mobility, however, accompanies a high price. According to Wikipedia, out of total pollution, 27% pollution is caused by Vehicles in India. The growth rate of vehicles is the backbone of economic development and therefore the Indian automotive industry (the second fastest growing within the world). About 7-8 million vehicles are produced annually within the country today. In 2011 the country reported 141.8 million registered automobiles. The motorization rate in India is 260 vehicles per 1000 population, and this is lower than many developed countries like the

United Kingdom, China etc. China is the most populated country in the world, with a population of around 1,439,323,774. In 2020, its average air quality index (AQI) is 34.70% which will decrease as compared to 2019. The United Kingdom's population is around 67,886,004 & its average air quality index (AQI) is 8.30% which is very less, compared to other countries throughout the world (Brazil-222/1000 population in 2012, South Africa 153/1000 population.). Automobiles are the first source of pollution in India's major cities. In India the transport sector emits an estimated 261 tons of CO_2, of which 94.5% is contributed by road transport. According to data out of a total 3,000 metric tonnes of pollutants belched out a day, on the brink of two-third (66%) is from vehicles in Delhi. Similarly, the contribution of vehicles to urban air pollution is 52% in Mumbai and close to one-third (33%) in Kolkata. The transport sector in India consumes about 17% of total energy and is responsible for 60% production of the GHG from various activities. The pollution from vehicles is thanks to discharges like CO_2, unburnt HC, Pb.NO_2 and SO_2 and SPM from tail pipes.

Currently, in India, there is a high influx of population to urban areas, which has led to growing cities, sharp increase in traffic, rapid economic development and industrialization, and better levels of energy consumption in cities/UT like Delhi NCR, Lucknow, Ghaziabad, Gurugram etc. This unplanned urban and industrial development has led to the matter of pollution. The main contributor to this widespread pollution in urban areas is vehicular emission which is of great concern, as these are ground level sources and have maximum impact on the overall population.

Correlation between vehicle and pollution: -

• The automobile exhaust is that the part explanation for pollution, pollution and therefore the number of small cars have an excellent correlation.

• Exhaust gas emitted by vehicles could also be a quite scattered mobile source of pollution. Scientific analysis shows that the car exhaust contains quite 100 sorts of different compounds, among which the most pollutants are CO_2, HC, SO_2, NO_x, PM10 and PM2. 5 then on.

• Small cars emit most air pollutants on weekdays, followed by trucks, medium-sized vehicles emit the smallest amount of pollutants. For CO_2, the emission intensity of peak period is the highest of all time quantum, mainly caused by the vehicle flow.

• The pollutants contributed by automobiles mainly for the small car discharge. The contribution rate of small cars was the greatest (85%), which is thanks to the very fact that tiny cars occupied the most important percentage in vehicle flow, the sum of middle-sized vehicles and trunks is merely 15%. The emission of pollutants by small cars is 68% of the entire amount of CO_2, 74% of HC, 77% of NO_x and 77% of PM10.

Vehicular Pollution:

In India, the vehicle population is growing at the speed of over 5% once a year and as of 2020, the vehicle population is approximately 40 million. The vehicle mix is additionally unique to India therein there is a really high proportion of two-wheelers are 76% (SIAM, 2011). The expansion rate of vehicles is the backbone of economic development and therefore the Indian automotive industry (the second fastest growing within the world). Today, within the country about 7-8 million vehicles are produced annually. In 2011, the country reported 141.8 million registered automobiles.

Vehicular pollution is the introduction of harmful material into the environment by automobiles. These materials, referred to as pollutants, are released from vehicles and includes fugitive emissions of the fuel and therefore the source and level of those emissions depending upon the vehicle type, its maintenance, etc. the main pollutants released as vehicle/fuel emissions are, carbon dioxide gas (CO_2), nitrogen oxides (NO_x), photochemical oxidants, air toxics, namely benzene (C_6H_6), aldehydes, 1,3 butadiene (C_4H_6), lead (Pb), particulate (PM), hydrocarbon (HC), oxides of Sulphur (SO_2) and polycyclic aromatic hydrocarbons (PAHs). While the predominant pollutants in petrol/gasoline driven vehicles are hydrocarbons and carbon dioxide gas, the predominant pollutants from the diesel-based vehicles are Oxides of nitrogen and particulates. Transportation could also be a serious source of pollution in many countries around the world because of the high number of vehicles that are available on the roads today. A rise in purchasing power means more people can now afford cars and this is often bad for the environment. Vehicular pollution has grown at an alarming rate because of growing urbanization in India. The pollution from vehicles in urban areas, particularly in big cities, has become a severe problem.

Effects of Vehicular Pollutants on Human Health:

A conventional vehicle, equipped with an inside combustion engine, besides being a means of transport, is additionally a source of chemical and noise pollution. In addition to heat, CO_2 and water, exhaust gases contain several chemical compounds which may severely affect human health in high concentration.

The vehicular emissions have damaging effects on both human health and ecology. there's an honest range of adverse health/environmental effects of the pollutants released from vehicles. The consequences could also be direct also as in-direct covering right from reduced visibility to cancers and death in some cases of acute exposure to pollutants, especially carbon dioxide gas. These pollutants are believed to directly affect the respiratory and cardiovascular systems. Elevated levels of Sulphur dioxide and Suspended Particulate Matters are related to increased mortality, morbidity, and impaired pulmonary function.

Road vehicles, powered by combustion engines, use fuel in liquid form as an energy source. There are two kinds of pollutant emissions coming from vehicles with combustion engines:

- Primary: produced thanks to incomplete combustion and evaporation of fuel
- Secondary: produced within the atmosphere together with other chemicals

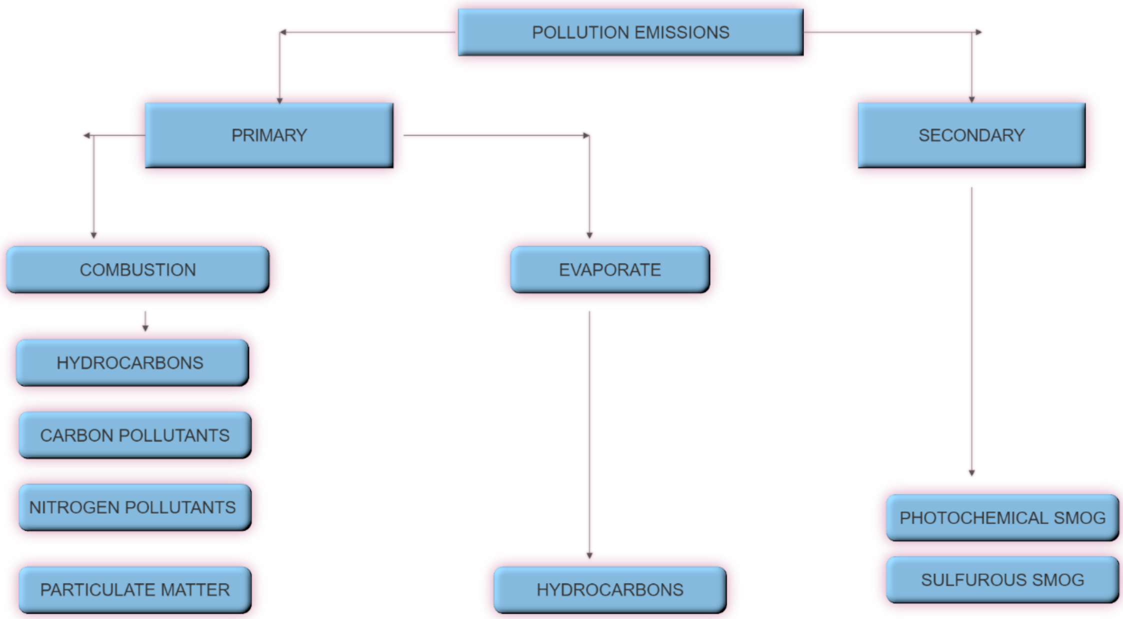


IMAGE: TYPES OF POLLUTANT EMISSIONS FROM VEHICLES WITH INTERNAL COMBUSTION ENGINES(ICE)

Depending on the sort of engine, compression ignition (diesel) or spark ignition (gasoline/petrol), the exhaust gases contain different levels of chemical substances.

The type of pollutant emissions, function of engine type, alongside the most reason of production, are summarized within the table below.

The major concentration of pollutants in the air are:

• Particulate Matter (PM2.5 and PM10): Particulate matter (PM) is a complex pollutant as it consists of a variety of components in different concentrations. The principal source of particulate matter is road traffic emissions, particularly from diesel vehicles. It is also emitted from industrial combustion plants and power generation, commercial and residential combustion, and a few non-combustion processes. Particulate matter is further categorized based on its size in micrometers. The particles under 10 micrometers, refers to PM10 sometimes called the 'coarse fraction'. The particles under 2.5 micrometers, refers to PM 2.5 sometimes called the 'fine fraction'. PM2.5 is more damaging to human health than PM10. The prominent health effects caused thanks to this are premature death, aggravation of respiratory and disorder. These particles include dust, dirt, soot, smoke, and liquid droplets. Some particulate is large enough to ascertain, but other particulate is often seen only with a micro

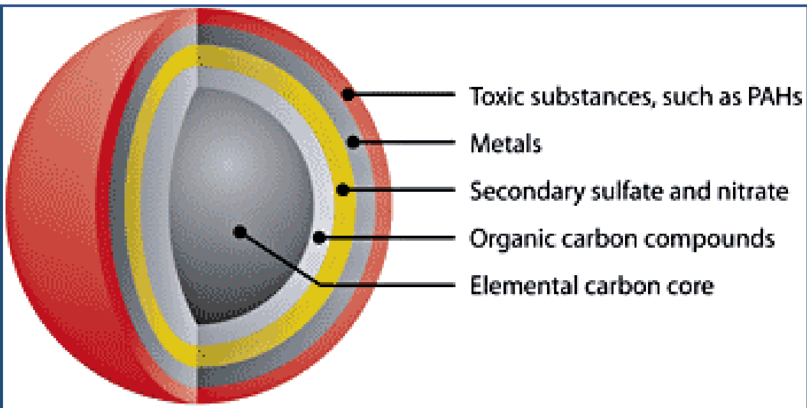
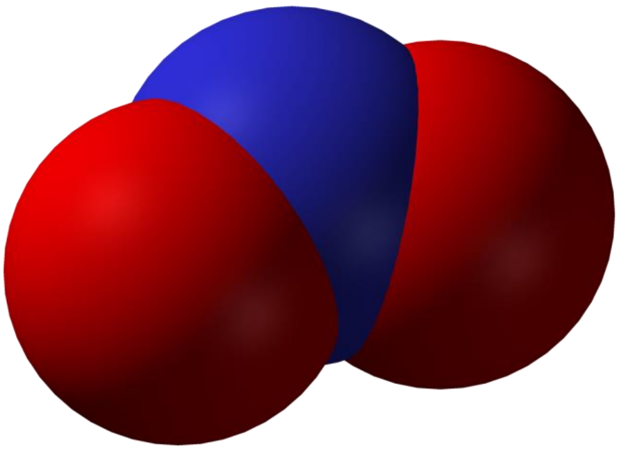


Image: Structure of a diesel particulate matter (PM)

• Nitrogen oxide (NOx): Oxides of Nitrogen are produced during industrial combustion processes and primarily as vehicular exhaust. NOx levels are highest in urban areas because of traffic. It is a crucial ingredient in the generation of photochemical smog which envelops the urban air with a haze like blanket. It has harmful effects like wide-range of respiratory problems in adults and youngsters. The main oxide s found within the exhaust gas are nitrogen oxide (NO) and dioxide (NO2). From the physical properties point of view, NO may be a sweet-smelling, colorless gas and NO2 may be a reddish-brown gas with irritating odors.



NO2

• Sulphur dioxide- automobiles create this pollutant by burning Sulphur-containing fuels, especially diesel. It can react within the atmosphere to make fine particles and may pose a health risk to young children and asthmatics. Short-term exposure to high levels of Sulphur dioxide is often life-threatening. Generally, exposures can cause a burning sensation within the nose and throat. Also, exposure can cause difficulty breathing, including changes within the body's ability to require a breath or breathe deeply, or absorb the maximum amount of air per breath. Long-term exposure to Sulphur dioxide can cause changes in lung function and aggravate existing heart conditions.

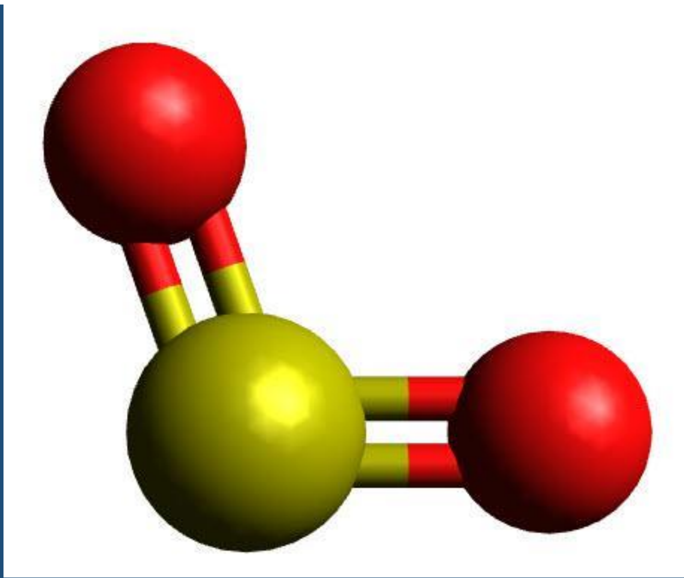


Image: Sulfur dioxide molecule structure

It is colorless gas with rotten egg odors. It is highly soluble in water. It liquefies easily.SO2 dissolves in water to form sulfurous acid due to which it possesses an acidic character.

$H2O + SO2 \rightarrow H2SO3$

It doesn't support combustion neither is it combustible.SO2 may be a strong oxidant. Acid rain occurs when released into the atmosphere by Sulphur dioxide (SO2) and nitrogen oxides (NOX) and carried by wind and air currents. The SO2 and NOX react to make sulfuric and nitric acids with water, oxygen, and other chemicals. They then combine before dropping to the table, with water and other materials.

Carbon Dioxide may be a toxic air pollutant which is produced by incomplete combustion of carbon- containing fuels. Vehicle deceleration and idling vehicle engines are one among its main causes. Higher amounts of carbon dioxide gas are found in urban zones, big cities, where there's heavy traffic and also industrial activities. Transportation features a major impact on the entire amount of carbon dioxide gas within urban areas. For these reasons, there are many government regulations in situ which try to scale back the pollutant emissions of road vehicles. Some cities went a step further and introduced "emission-free zones" which may be accessed by electric vehicles only.

Carbon Dioxide may be a colorless, odorless, and tasteless flammable gas that's slightly less dense than air and features a toxic effect on the physical body. Carbon dioxide inhibits the power of hemoglobin to move oxygen. Hemoglobin may be a metalloprotein found within the blood which is transporting oxygen from the lungs to the remainder of the body. By combining carbon dioxide gas with hemoglobin, we get carboxyhemoglobin, which may not deliver oxygen to the body tissues.

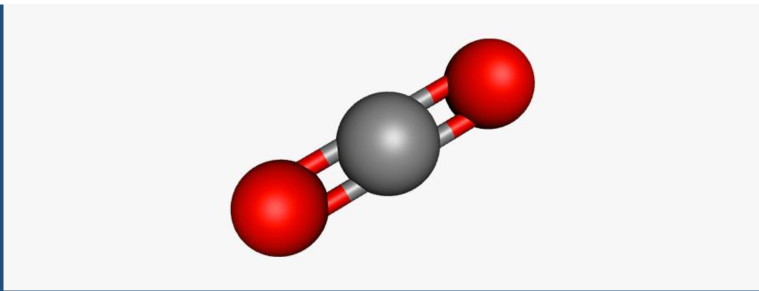


Image: Carbon Dioxide molecule structure

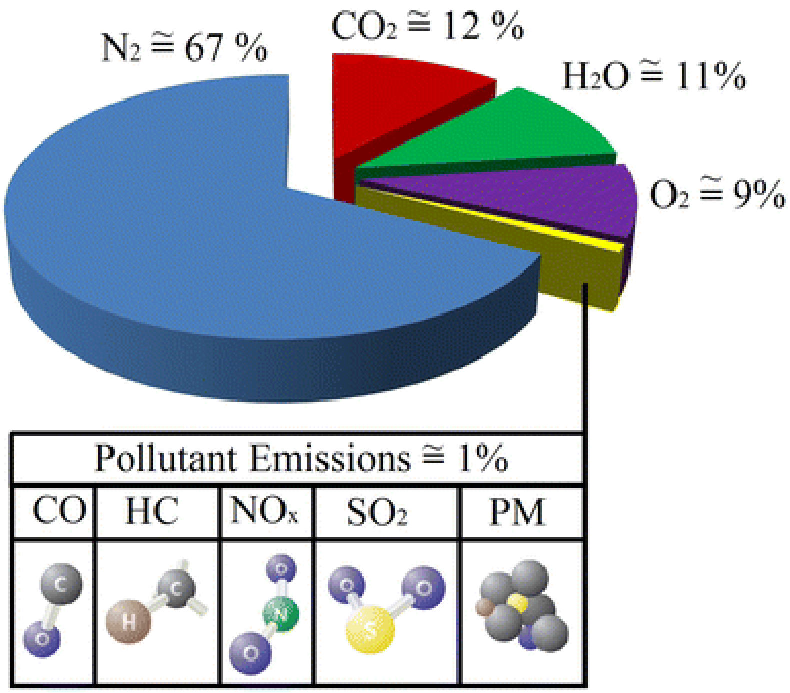
Causes of Vehicular Pollution:

The fundamental driver of vehicular pollution is the quickly developing number of vehicles. Different components of vehicular pollution in the metropolitan regions are 2-cycle motors, helpless fuel quality, old vehicles, deficient upkeep, blocked traffic, helpless street condition and old car advances and traffic the board framework.

Vehicular Pollution Problems in India:

Engine vehicles have been firmly related to expanding air pollution levels in metropolitan communities of the world. Other than considerable Carbon Dioxide (CO2) outflows, huge amounts of Carbon Dioxide (CO2), Hydrocarbon (HC), Nitrogen Oxide (NOx), Suspended Particulate Matter (SPM) and other air poisons are produced from these engine vehicles in the climate, causing genuine natural and wellbeing impacts.

In the same way as other different parts of the world, air pollution from engine vehicles is perhaps the most genuine and quickly developing issue in metropolitan habitats of India. The issue of air pollution has accepted genuine extent in a portion of the significant metropolitan urban areas of India and vehicular outflows have been recognized as one of the significant benefactors in the breaking down of air quality in these metropolitan places. The issue has additionally been compounded by the centralization of a tremendous number of vehicles and similarly high engine vehicles to populace proportions in these urban areas.



In the above pie chart, the four main pollutant emissions (CO2, HC, PM, and NO_x) from diesel engines are explained. Each type of emission is investigated individually and the impacts of each emission on environmental and health problems are also revealed.

Explanations behind expanding vehicular pollution issues in metropolitan India are as underneath:

- High vehicle thickness in Indian metropolitan communities.
- Older vehicles prevalent in vehicle vintage
- Predominance of private vehicles particularly vehicles and bikes, attributable to unsuitable public vehicle framework, accordingly causing higher sitting outflows and gridlock.
- Absence of sufficient land use arranging being developed of metropolitan zones, consequently causing more vehicle travel and fuel utilization
- Inadequate examination and upkeep offices.
- Adulteration of fuel and fuel items
- Improper traffic the board framework and street conditions
- Elevated levels of pollution at intersections
- Absence of successful mass quick vehicle framework and intra-city rail route organizations
- High populace departure to the metropolitan habitats.
- Increasing number soaring structures in the metropolitan zones makes the stagnation of the vehicular emanations the ground level and forestalling its appropriate scattering.

AQI-Air Quality Index:

The AQI is utilized by government organizations all throughout the planet to check how dirtied the air is in their urban communities. It's at present the best pointer of how vehicle outflows are answerable for dirtying our urban communities. The file utilizes five rules to check this:

- Particulate Matter (PM10 & PM2.5)

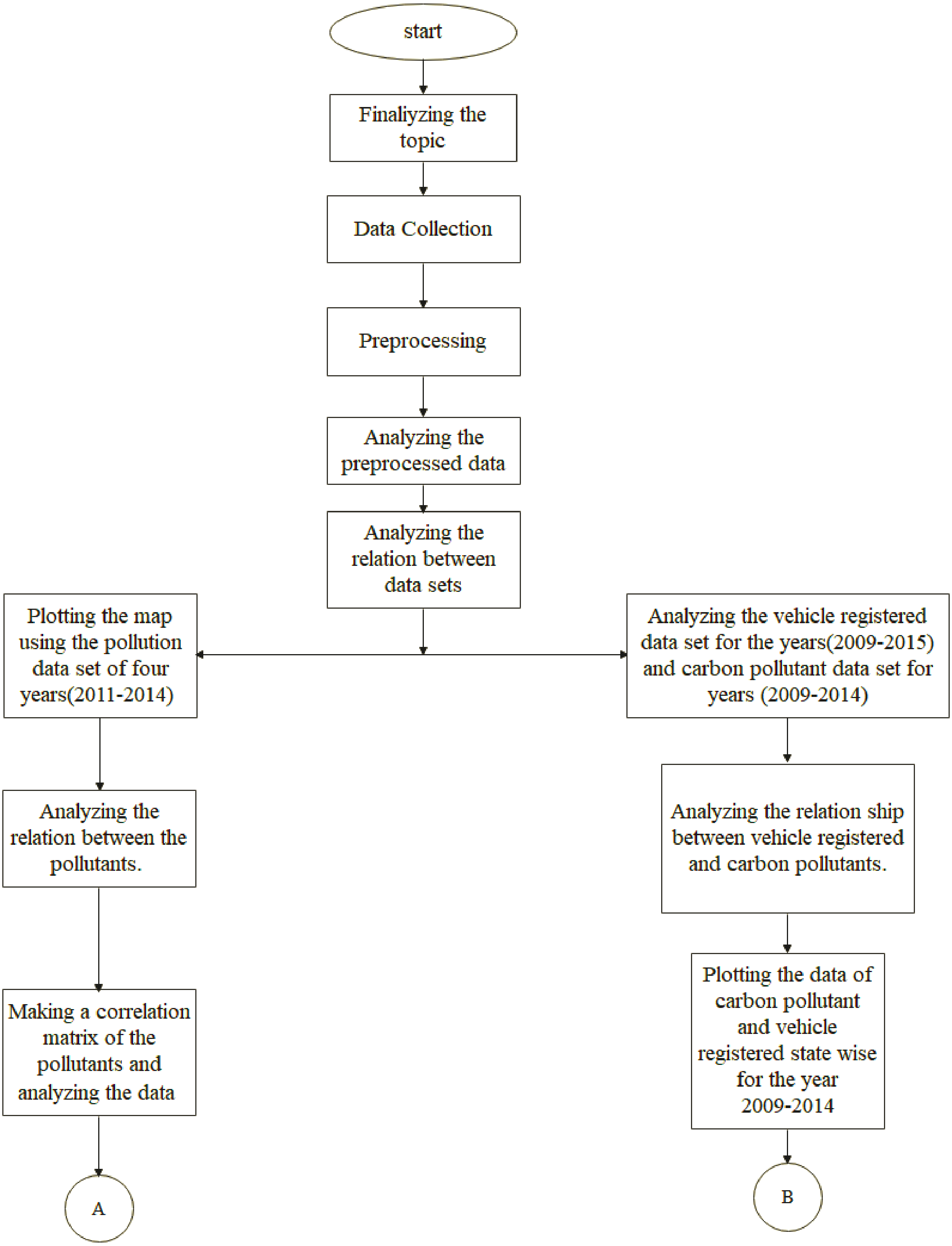
- Sulphur Dioxide (SO2)

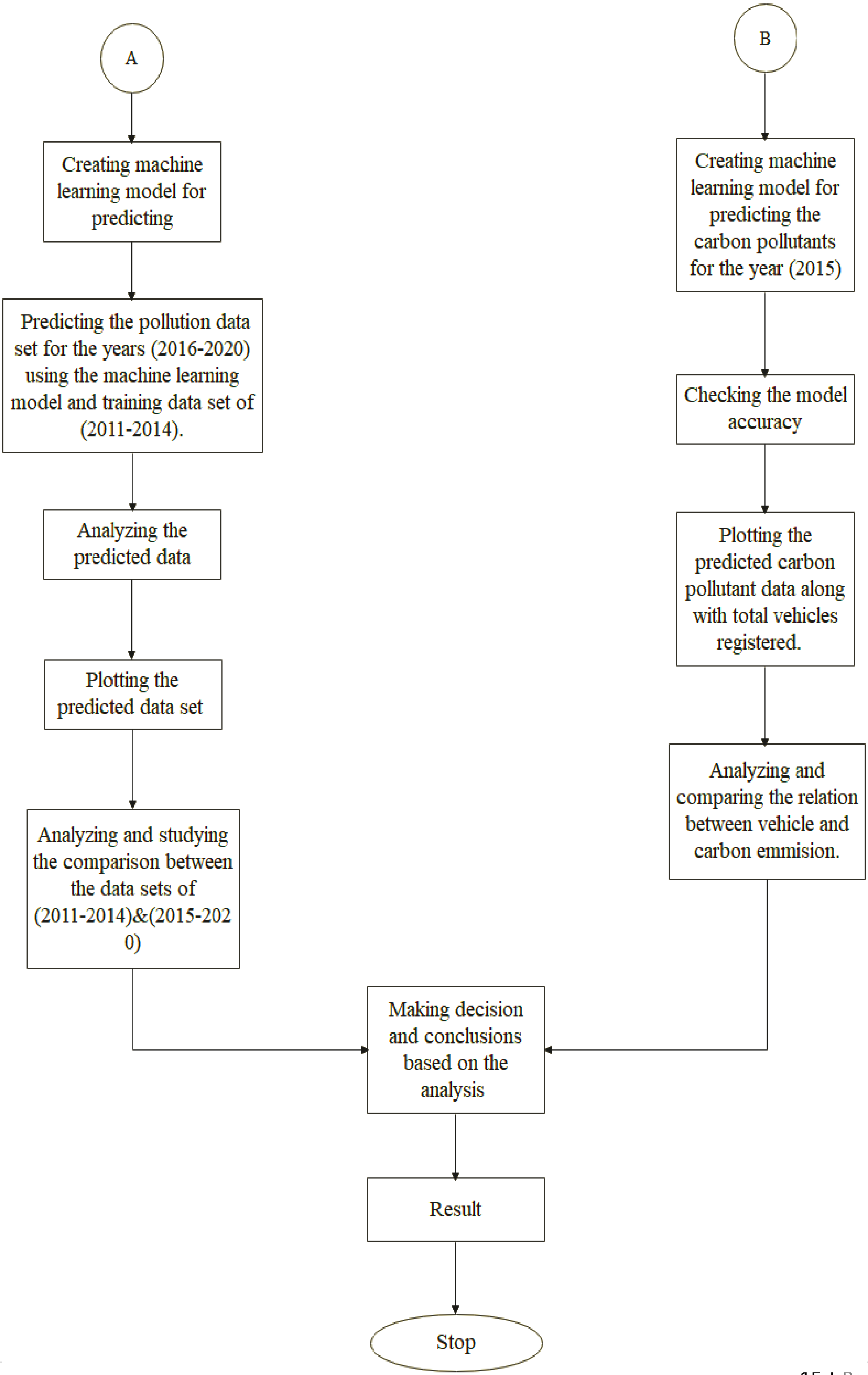
- Carbon Dioxide (CO2)

- Nitrogen Dioxide (NO2)

- Ozone (OZ)

FLOWCHART


start
Finalyzing the topic
Data Collection
Preprocessing
Analyzing the preprocessed data
Analyzing the relation between data sets
Plotting the map using the pollution data set of four years(2011-2014)
Analyzing the vehicle registered data set for the years(2009-2015) and carbon pollutant data set for years (2009-2014)
Analyzing the relation between the pollutants.
Analyzing the relation ship between vehicle registered and carbon pollutants.
Making a correlation matrix of the pollutants and analyzing the data
Plotting the data of carbon pollutant and vehicle registered state wise for the year 2009-2014
A
B


A
Creating machine learning model for predicting
Predicting the pollution data set for the years (2016-2020) using the machine learning model and training data set of (2011-2014).
Analyzing the predicted data
Plotting the predicted data set
Analyzing and studying the comparison between the data sets of (2011-2014)&(2015-2020)
B
Creating machine learning model for predicting the carbon pollutants for the year (2015)
Checking the model accuracy
Plotting the predicted carbon pollutant data along with total vehicles registered.
Analyzing and comparing the relation between vehicle and carbon emmision.
Making decision and conclusions based on the analysis
Result
Stop

METHODOLOGY

- **Label encoding:**
  It refers to the method of converting labels into numeric form and to convert it into machine-readable form. Machine learning algorithms decide in a better way on how these labels are operated. In supervised learning this is an important preprocessing step for the dataset to be structured.

- **Linear Regression:**
  It is a particular type of multiple time-series analysis for statically related data. It's a statistical technique for quantifying the relationship between variables.

- **DATASETS USED: -**

1) Dataset for number of vehicles registered during the period of years 2009-2015: -

- This dataset has been collected from www.data.gov.in .
- The dataset contains two attributes: STATE/UNION TERRITORY and TOTAL VEHICLE REGISTERED.
- These attributes describe the state/union territory and total vehicle registered (including transport and non-transport) in each year.
- The data has been collected from 2009 to 2015 for critical analysis and accurate prediction for the future.
- A snapshot of the dataset used is shown below.

MOTOR VEHICLES

Table 20.1-NUMBER OF MOTOR VEHICLES REGISTERED IN INDIA (TRANSPORT AND NON TRANSPORT as on 31st March)
(in numbers)

State /Union Territory	Grand Total (Transport+Non Transport)						
	2009	2010	2011	2012	2013	2014	2015
1	58	59	60	61	62	63	64
State:							
Andhra Pradesh	8058948	8923139	10189347	12424328	12687703	7002143	7882262
Arunachal Prades	22101	22101	144534	151279	151279	151279	151279
Assam	1234751	1383721	1582128	1807120	1877522	2217007	2509737
Bihar	1959553	2356986	2673209	3112880	3617227	4163396	4777596
Chhatisgarh	2114589	2435773	2766037	3104038	3437242	3871357	4313927
Goa	673909	727042	790075	865609	938372	1009362	1083668
Gujarat	10998651	11872573	12993135	14413717	15772453	17091599	18720567
Haryana	4425221	4791825	5377003	5978110	6599594	7238748	7927550
Himachal Prades	494398	538341	621714	736604	876061	974270	1077404
Jammu & Kashmi	668445	738905	926961	916898	1020786	1133077	1244271
Jharkhand	2038020	2767408	3113182	3157986	3416777	1718760	2065606
Karnataka	6952551	9043976	9930483	10909601	12063686	13335106	14784961
Kerala	4859918	5397652	6072019	6893314	7857822	8775041	9648320
Madhya Pradesh	6010691	6590576	7355702	8144159	8760039	9721625	11141127
Maharashtra	144590908	15768421	17434099	19432361	21488152	23393776	25562175
Manipur	147394	194452	206502	214523	295927	394544	307143
Meghalaya	141877	158113	175737	197838	220021	248265	267104
Mizoram	69882	80456	92648	101830	121336	137231	151486
Nagaland	239873	254483	272653	291438	295875	317917 &	333761&
Orissa	2607361	2931832	3338038	3759530	4215540	4701806	5218595
Punjab	4831531	5274254	5274254	6262939	6262939	6262939	6262939
Rajasthan	6489564	7165662	7986265	8985478	10071945	11133420	12378929
Sikkim	28551	33626	38783	43334	36277	39608	43210
Tamil Nadu	12890997	14061533	15638245	17412248	19232061	20863662	22518669
Telangana					6376477	7073109	7844716
Tripura	143823	160119	187673	204218	244922	256170	282134
Uttara Khand	787352	831372	997161	1244042	1460136	1639928	1827037
Uttar Pradesh	10778812	11988349	13287232	15445274	17048184	19114692	21635531
West Bengal	3043619	2747138	3260624	3860741	6110734	6745004	7403241
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

2) Dataset for emission of SO_2, NO_2 and PM for 33 states/union territory during 2011-2014: -

- This dataset has been collected from www.data.gov.in .
- The dataset contains five attributes:
  - STATE/UNION TERRITORY: -total amount of harmful gases such as SO_2, NO_2 and PM emitted through the heavy vehicles were recorded in 33 states/territories.
  - SO_2(Sulphur dioxide): automobiles create this pollutant by burning Sulphur-containing fuels, especially diesel. It can react within the atmosphere to make fine particles and may pose a health risk to young children and asthmatics.
  - NO_2: - Nitrogen Dioxide is produced during industrial combustion processes and primarily as vehicular exhaust. NO_x levels are highest in urban areas because of traffic. It is a crucial ingredient in the generation of photochemical smog which envelops the urban air with a haze like blanket.
  - PM: - Particulate matter (PM) is a complex pollutant as it consists of a variety of components in different concentrations. The principal source of particulate matter is road traffic emissions, particularly from diesel vehicles.
  - Year: -it describes the duration of the data collected.

- These attributes describe the emission of the harmful gases from vehicles such as SO_2, NO_2 and PM in different state/union territory in each year during the period of 2011-2014.
- This dataset is used for critical analysis and accurate prediction for the future.
- A snapshot of the dataset used is shown below.

	A	B	C	D	E
1	State	SO2	NO2	PM	Year
2	Andhra Pra	7	20	86	2011
3	Assam	9	24	113	2011
26	Rajasthan	6	27	129	2011
27	Tamil Nadı	11	20	72	2011
32	Lakshadw	2	6	113	2011
33	Tripura	2	6	113	2011
34	Andaman ;	2	6	113	2011
35	Andhra Pra	5	17	73	2012
56	Orissa	5	18	86	2012
57	Punjab	11	27	187	2012
58	Puduchery	6	13	38	2012
59	Rajasthan	7	29	168	2012
60	Tamil Nadı	9	20	70	2012
66	Tripura	2	6	113	2012
67	Andaman ;	2	6	113	2012
68	Andhra Pra	7	19	77	2013
69	Assam	10	23	114	2013
70	Bihar	10	23	114	2013
91	Puduchery	7	16	42	2013
92	Rajasthan	7	31	165	2013
99	Tripura	2	5	114	2013
100	Andaman ;	2	5	114	2013
101	Andhra Pra	7	18	72	2014
102	Assam	6	13	72	2014
103	Bihar	6	36	166	2014
104	Chandigarl	2	19	110	2014
105	Chhattisga	11	26	151	2014
106	Dadra & N	8	20	115	2014

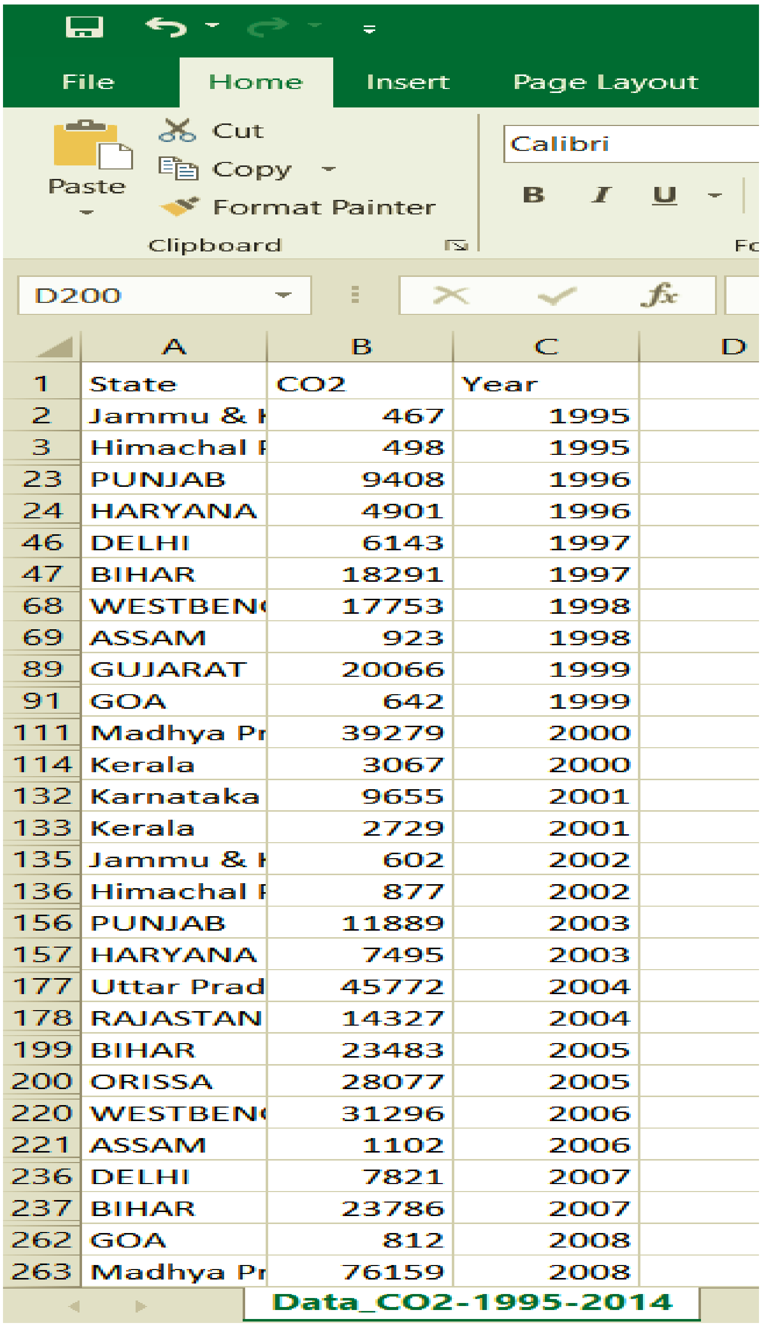
Data_PMSO2NO2_2011-2020

3) Dataset for emission of CO_2 for 16 states/unions territory during 1995-2008:

- This dataset has been collected from www.data.gov.in .
- The dataset contains three attributes:
  - STATE/UNION TERRITORY: - total amount of harmful gas (CO_2) emitted through the heavy vehicles were recorded in 16 states/territories.
  - CO_2: - CO_2 is a toxic air pollutant which is produced by incomplete combustion of carbon- containing fuels. Vehicle deceleration and idling vehicle engines are one among its main causes.
  - YEARS: - it describes the duration of the data collected.

- These attributes describe the emission of CO_2 in different state/union territory in each year during the period of 1995-2008.
- This dataset is used for critical analysis and accurate prediction for the future.

- A snapshot of the dataset used is shown below.

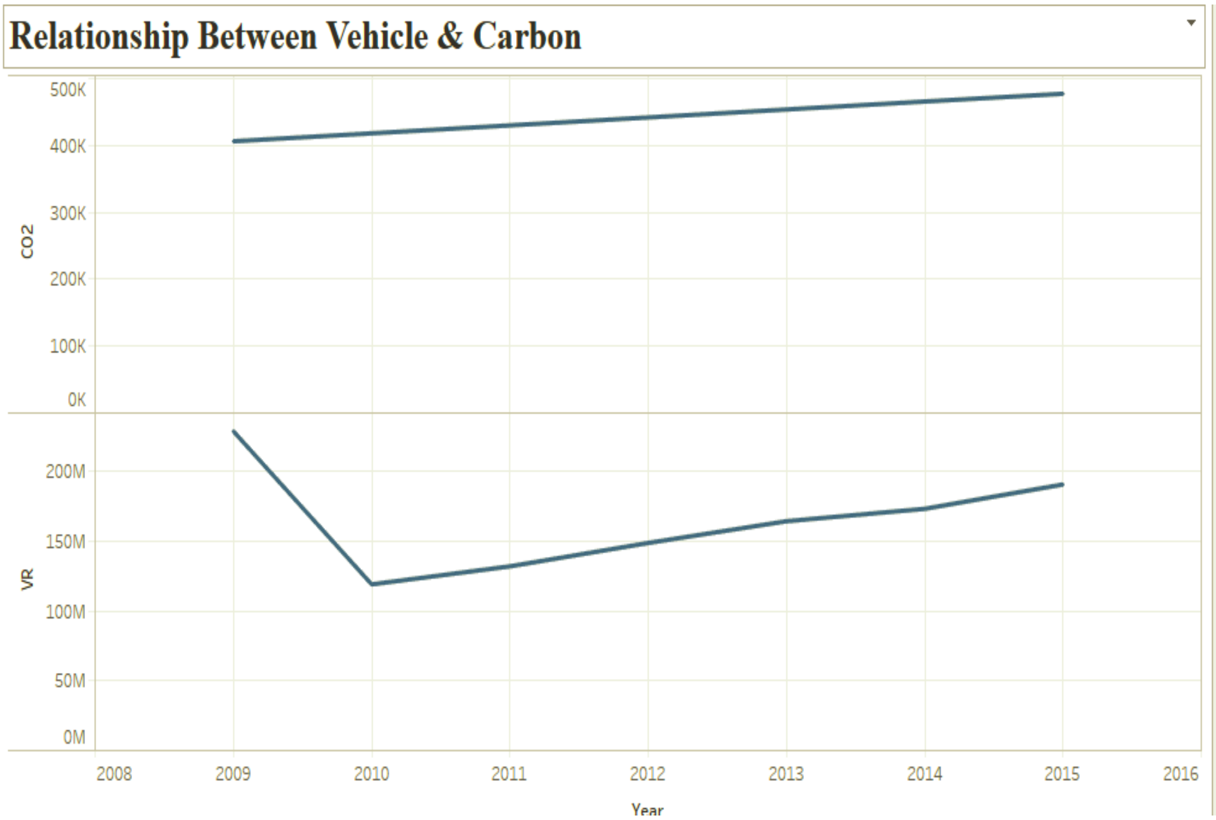


	A	B	C	D
1	State	CO2	Year	
2	Jammu & I	467	1995	
3	Himachal I	498	1995	
23	PUNJAB	9408	1996	
24	HARYANA	4901	1996	
46	DELHI	6143	1997	
47	BIHAR	18291	1997	
68	WESTBEN(	17753	1998	
69	ASSAM	923	1998	
89	GUJARAT	20066	1999	
91	GOA	642	1999	
111	Madhya Pr	39279	2000	
114	Kerala	3067	2000	
132	Karnataka	9655	2001	
133	Kerala	2729	2001	
135	Jammu & I	602	2002	
136	Himachal I	877	2002	
156	PUNJAB	11889	2003	
157	HARYANA	7495	2003	
177	Uttar Prad	45772	2004	
178	RAJASTAN	14327	2004	
199	BIHAR	23483	2005	
200	ORISSA	28077	2005	
220	WESTBEN(	31296	2006	
221	ASSAM	1102	2006	
236	DELHI	7821	2007	
237	BIHAR	23786	2007	
262	GOA	812	2008	
263	Madhya Pr	76159	2008	

Relevant relationships between datasets: -

Relationship between vehicle registration & Carbon: -

- In the Vehicle registration dataset, we have year wise vehicle registration in all the states in India Such as in 2009 Goa has 7644, Bihar has 19654 etc.



- In the above line chart, we can see the relationship between two major attributes in vehicle & Carbon dioxide dataset. With the help of both attributes, we can analyze Vehicle & Carbon dioxide year wise.
- In Pollution state wise dataset, we have Sulphur, Carbon & Particulate matter chemicals in air pollution. Also, we are CO2 relating the relationship between the pollutants.

ANALYSIS

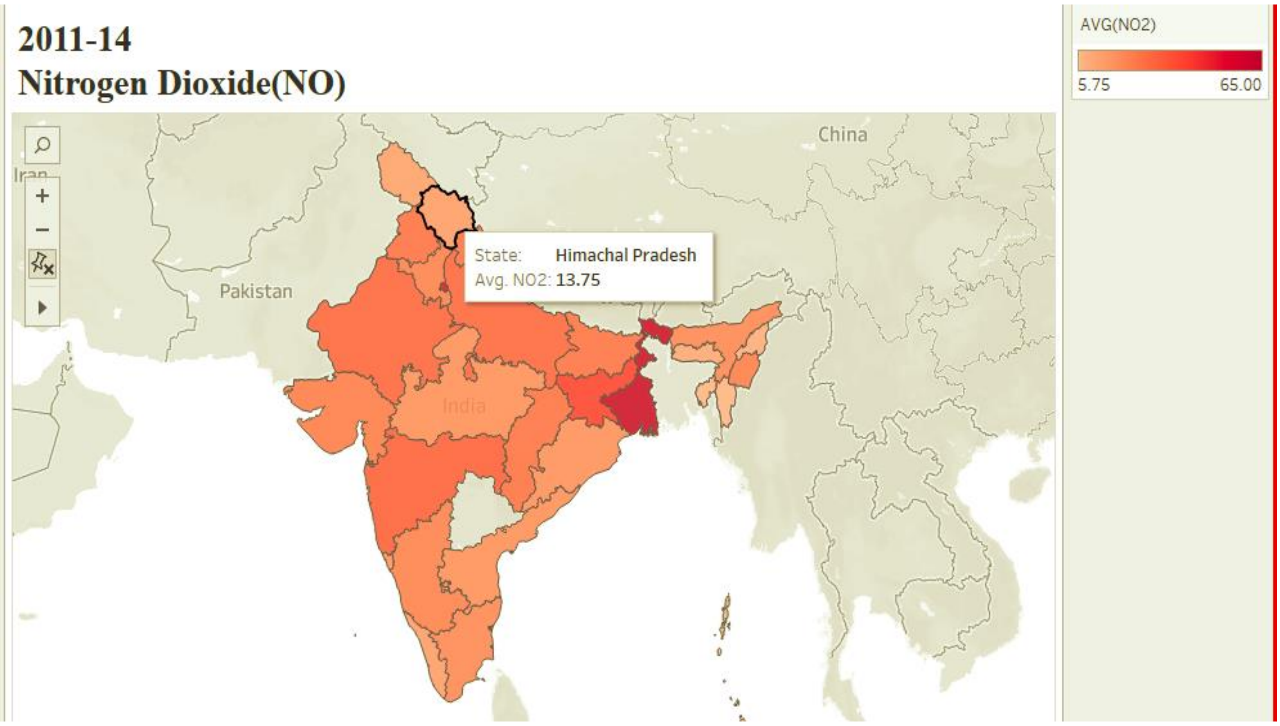
We are trying to conduct a study for analyzing the pollution of India and predicting the changes that may occur in upcoming years and analyzing the relationship between vehicle registered and carbon pollutants emitted across the state. We are conducting this study in two phases.

The first phase will be for analyzing the pollution of India across 33 states/ union territory during the years (2011-2014) and correlating the pollutants and will be predicting the expected pollution range for next six years (2015-2020) using the machine learning model. Once the data are predicted we study by comparing the map plotted using the data set of (2011-2014) and (2015-2020) and will analyze the major changes which may take place and according to that data we will try to provide data driven decisions.

Second phase will be analyzing the relationship between how the increase in vehicle registration across the states is having an impact on carbon emission. For conducting the study, we will be using carbon emission dataset of across 16 major states during the years (2009-2014) and vehicle registered dataset during the years (2009-2015). Once we study the impact of carbon pollutants due to the increase in vehicle registered, we will create a machine learning model for predicting the expected carbon pollutants which will be emitted during the year 2015 using the previous datasets of carbon emission used and along with the data set of vehicles registered during the year 2015. We will analyze the changes taking place in the predicted outcome and provide data driven decisions.

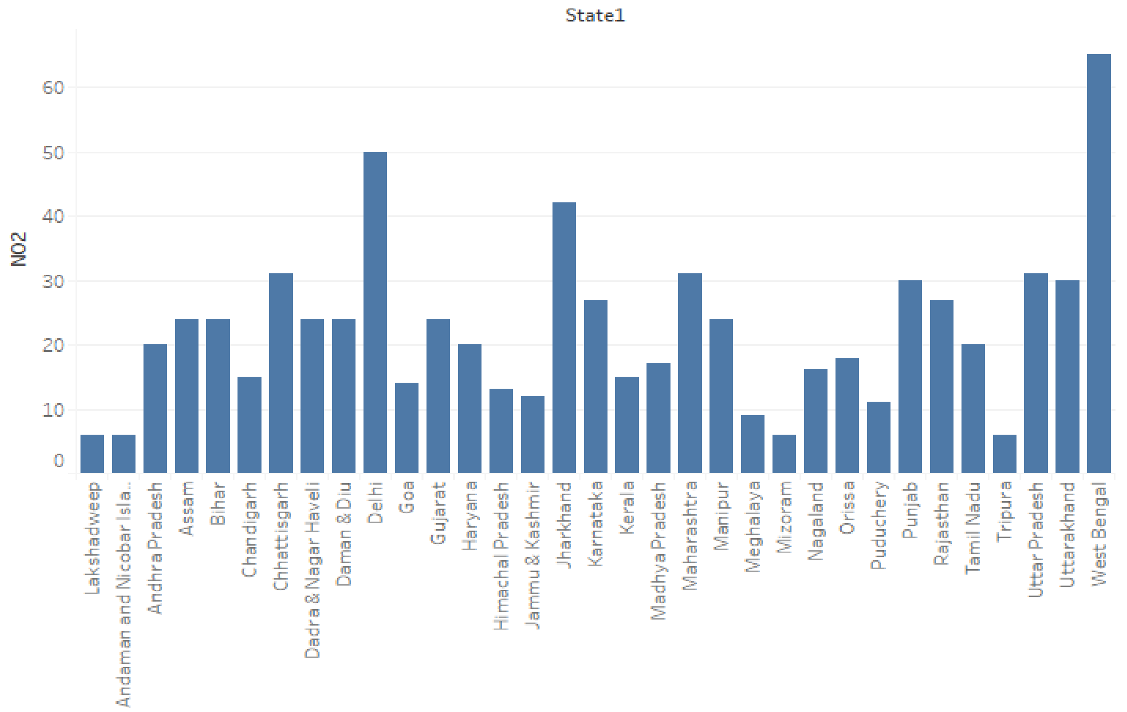
First Phase: Pollutant analysis

As for conducting our study in first phase:

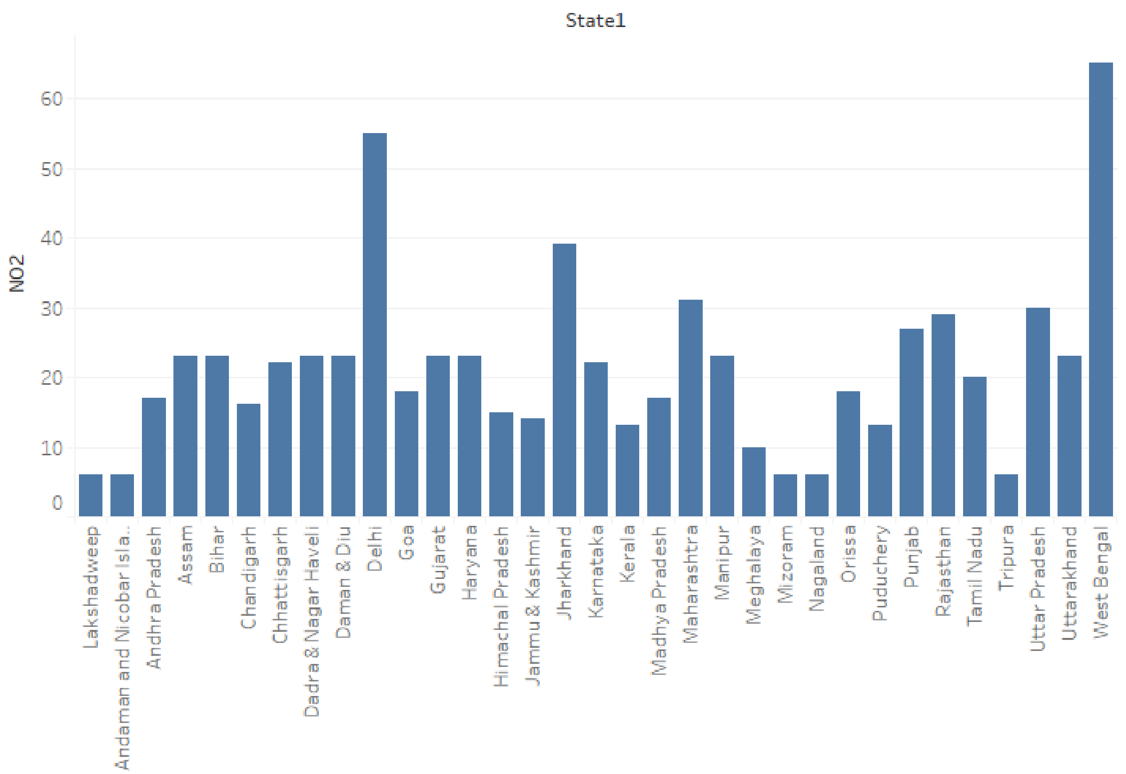


We have collected the data of NO emission for 33 states/union territories from 2011-2014. The above graph shows a sample of the map plotted. As the graph depicts the average NO2 value from the year 2011-2014. We can see Himachal Pradesh has an average NO2 value of 13.75.

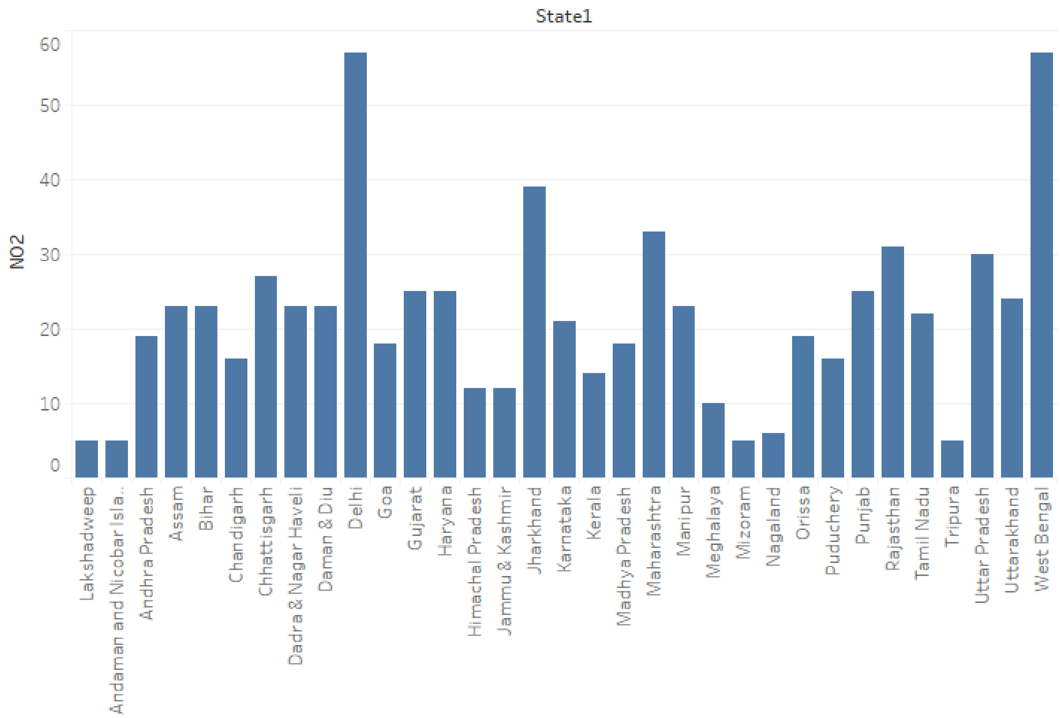
NO2 2011



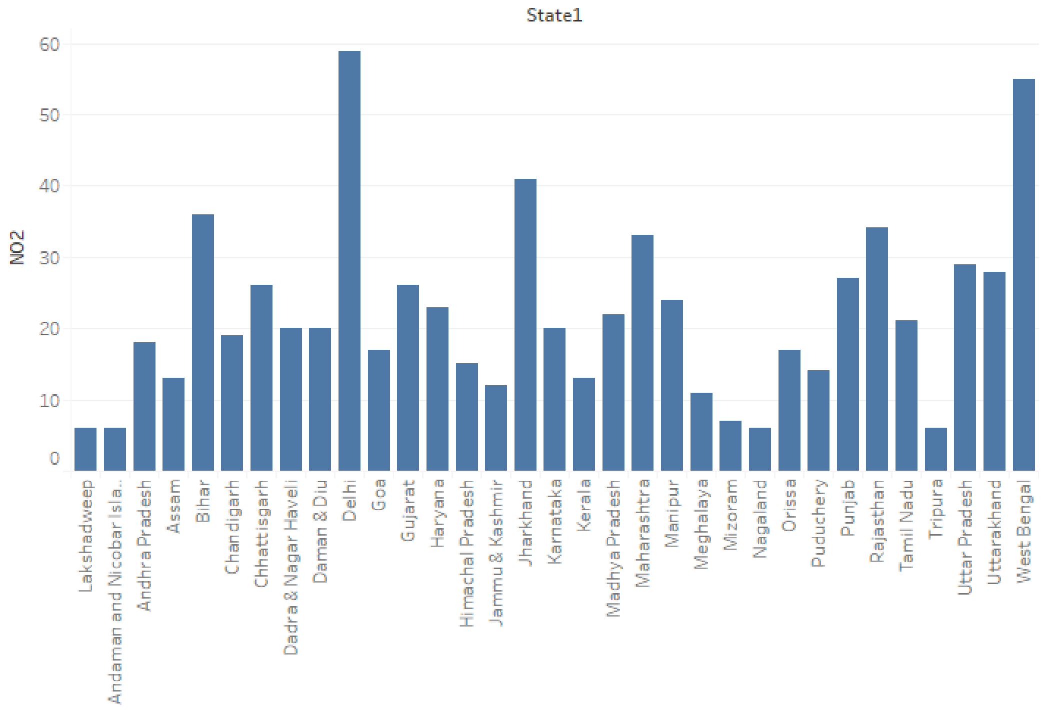
NO2 2012



NO2 2013



NO2 2014



For conducting further study, we predicted NO2 value for all 33 states for the year (2015&2016) using the data of (2011-2014) as a training data set. The image below shows the sample of predicted data.

ERROR PERCENTAGE OF PREDICTED AND ACTUAL FOR NO2.

(Top 6 and Bottom 4 states for NO2)

STATE	2014(ACTUAL)	2014(PREDICTED)	ERROR PERCENTAGE
DELHI	59	54	8.4%
JHARKHAND	41	40	2.4%
MAHARASHTRA	33	32	3.0%
UTTAR PRADESH	29	31	6.8%
WEST BENGAL	55	62	12.7%
MEGHALAYA	11	10	9.0%
ANDAMAN	6	6	0%
TRIPURA	6	6	0%
LAKSHADEEP	6	6	0%
MIZORAM	7	6	14.2%
		Overall error percentage	5.65%

```
In [55]:
test_2015_2016["NO2"] = pred
test_2015_2016
```

	Year	State	NO2
132	2015	Andhra Pradesh	7.423964
133	2015	Assam	7.588486
134	2015	Bihar	7.753008
135	2015	Chandigarh	7.917530
136	2015	Chhattisgarh	8.082052
...	...	...	...
193	2016	West Bengal	12.699020
194	2016	Jammu & Kashmir	9.737623
195	2016	Lakshadweep	7.434314
196	2016	Tripura	12.205453
197	2016	Andaman and Nicobar Islands	7.598836

66 rows × 3 columns

We have predicted NO2 value for the years (2017&18) using the training data set merging the original data set of (2011-2014) and predicted data set (2015-2016) to increase the accuracy of the model created. The image below shows the sample of the predicted data set.

```
In [31]:
test_2017_2018["NO2"] = pred
test_2017_2018
```

	Year	NO2	state_encoded
198	2017	18.246628	2
199	2017	18.479805	3
200	2017	18.712983	4
201	2017	18.946160	5
202	2017	19.179337	6
...	...	...	...
259	2018	25.187403	32
260	2018	20.990211	14
261	2018	17.725728	0
262	2018	24.487871	29
263	2018	17.958905	1

66 rows × 3 columns

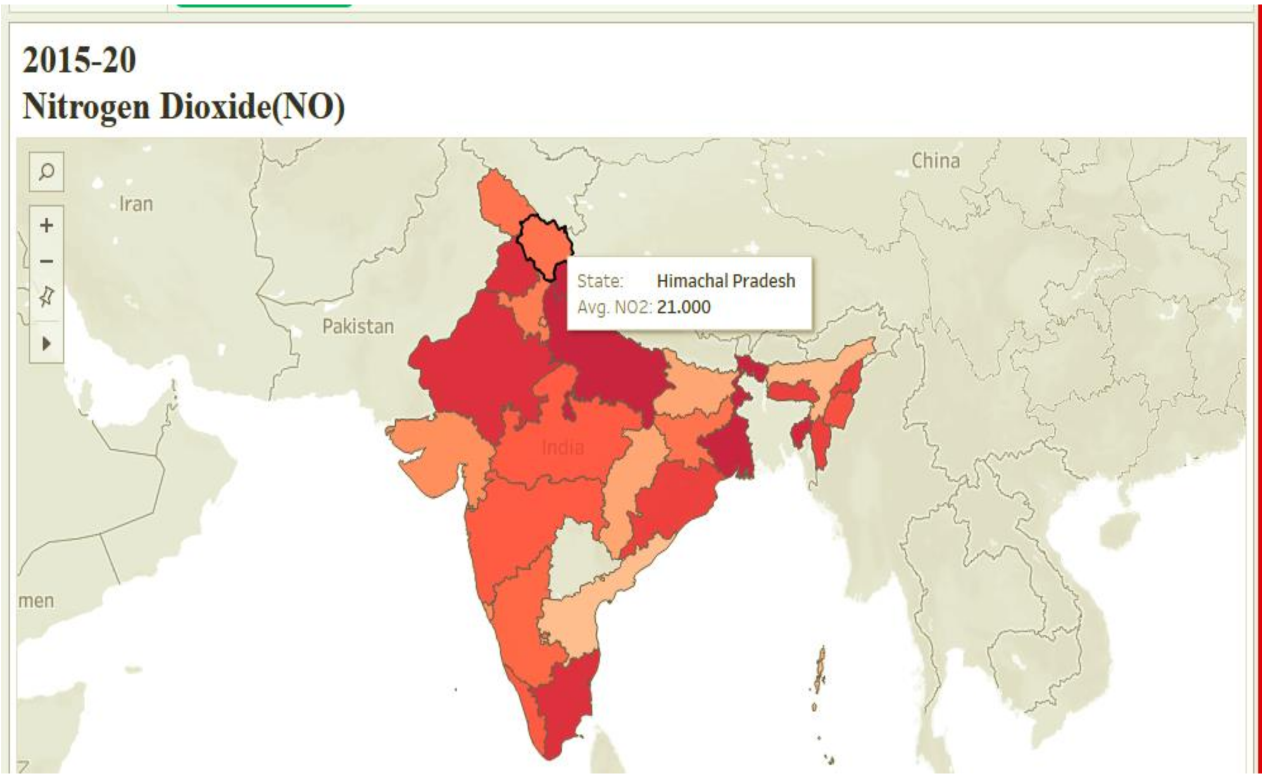
We have predicted NO2 value for the years (2019&2020) using the training data set merging the original data set of (2011-2014), predicted data set of years (2015-2016) and predicted data set of years (2017-2018) to increase the accuracy of the model created. The image below shows the sample of predicted data set of (2019-2020).

```
In [22]: test_2019_2020["NO2"] = pred
         test_2019_2020
```

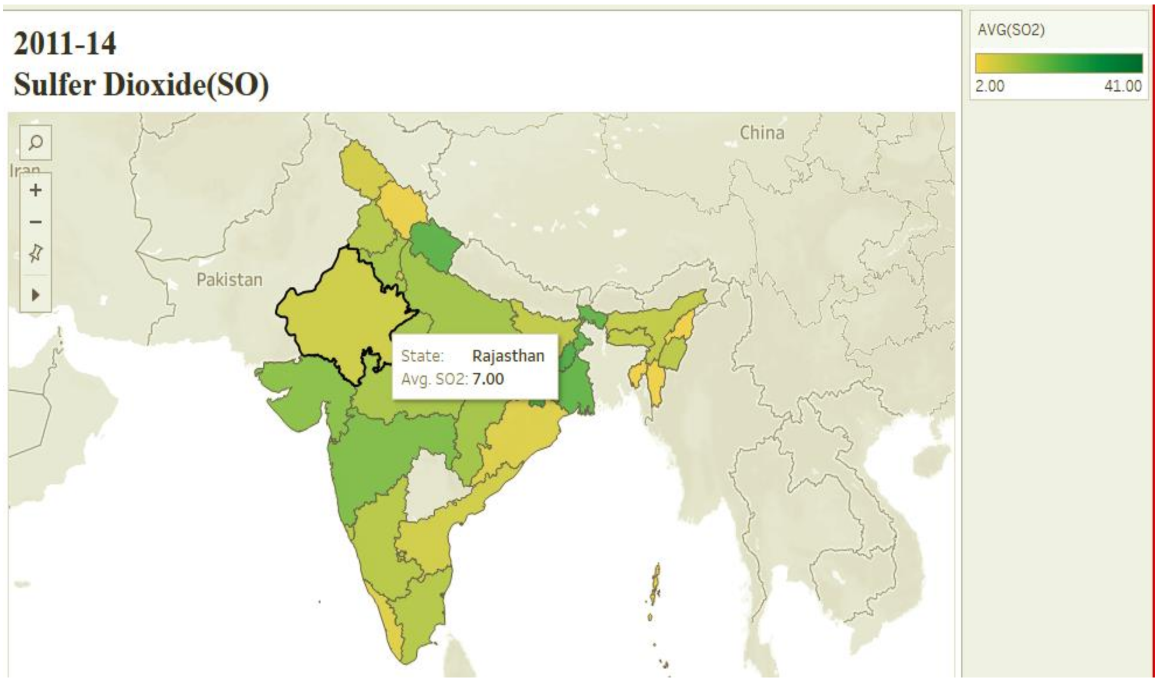
	Year	NO2	state_encoded
264	2019	18.138309	2
265	2019	18.371473	3
266	2019	18.604636	4
267	2019	18.837800	5
268	2019	19.070963	6
...	...	...	...
325	2020	25.078739	32
326	2020	20.881797	14
327	2020	17.617509	0
328	2020	24.379249	29
329	2020	17.850673	1

66 rows × 3 columns

Once our data of NO for the years (2015-2020) are predicted we plotted the data in a graph to compare the changes in NO taken place during the years (2015-2020) from (2011-2014). The image below shows the sample graph of Nitrogen dioxide.

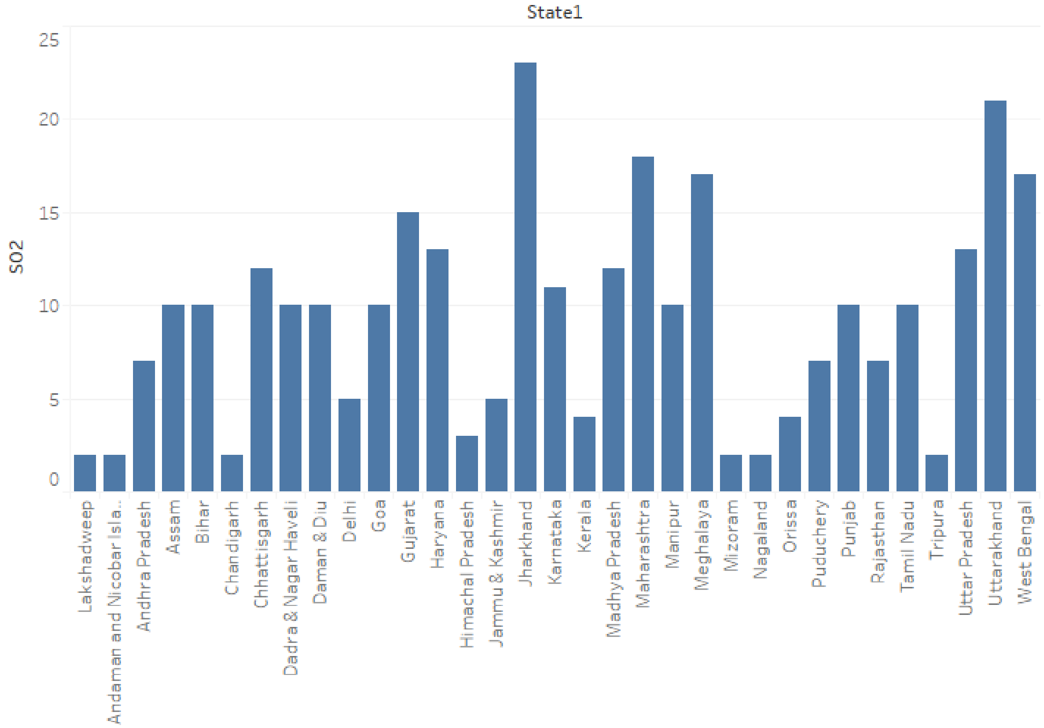


As the above graph depicts that the expected average value of NO2 of Himachal Pradesh for the year is 2021. Comparing the data with the data plotted from 2011-2014 shows a clear increase in NO from the year (2011-2014) to (2015-2020) with an increase from 13.7 to 21 in average value.

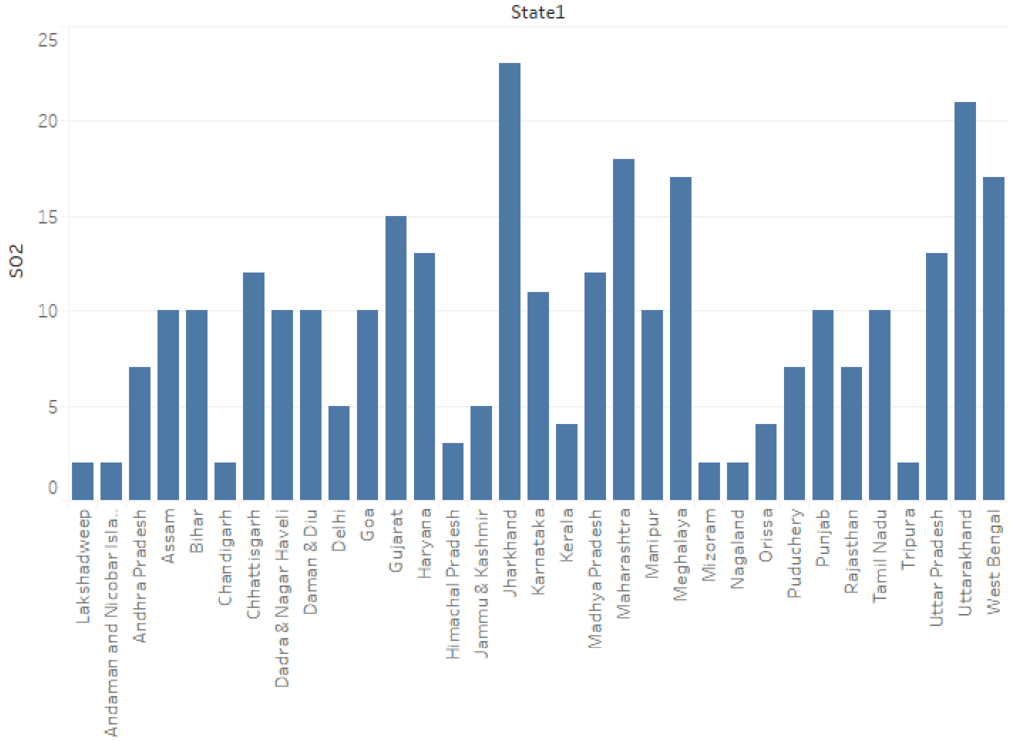


We have collected the data of SO emission for 33 states/union territories from 2011-2014. The graph above shows a sample of the map plotted. As the graph depicts the average SO2 value from the year 2011-2014. We can see Rajasthan has an average SO2 value of 7.

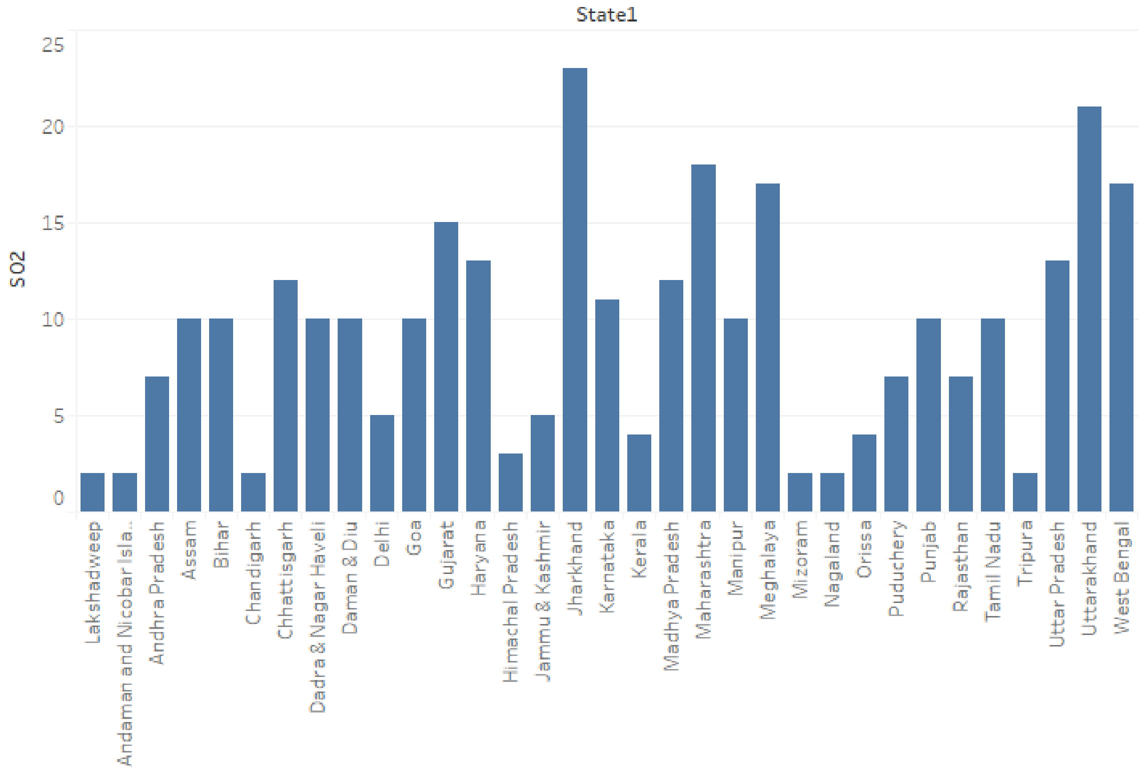
SO2 2011



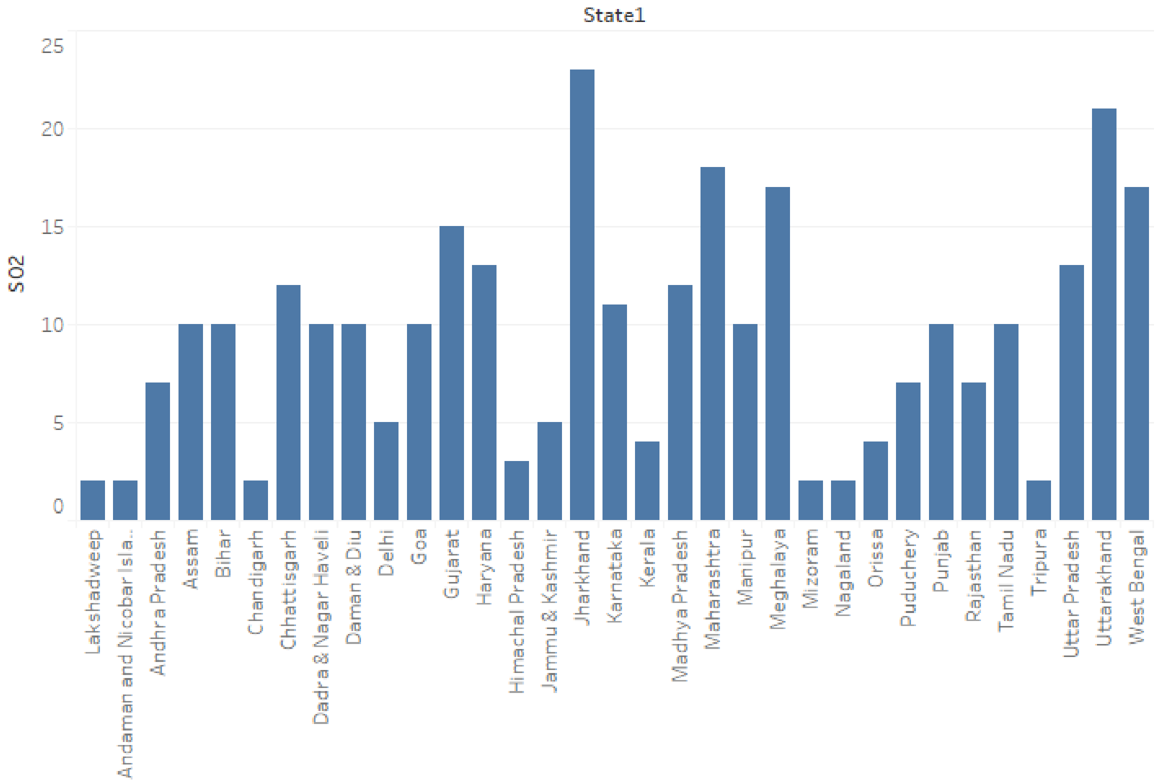
SO2 2012



SO2 2013



SO2 2014



ERROR PERCENTAGE OF PREDICTED AND ACTUAL FOR SO2.

(Top 7 and Bottom 5 states for SO2)

STATE	2014(ACTUAL)	2014(PREDICTED)	ERROR PERCENTAGE
GOA	11	9	18.1%
GUJARAT	15	15	0%
HARYANA	9	13	44.4%
JHARKHAND	23	23	0%
MAHARASHTRA	17	18	5.8%
MEGHALAYA	9	12	33.3%
UTTARAKHAND	26	17	34.6%
CHANDIGARH	2	2	0%
LAKSHADWEEP	2	2	0%
MIZORAM	2	2	0%
NAGALAND	2	2	0%
TRIPURA	2	2	0%
		Overall error %	11.35%

To conduct further study, we predicted the SO2 value for all 33 states for the year (2015&2016) using the data of (2011-2014) as a training data set. The image below shows the sample of predicted data.

In [42]:
```
test_2015_2016["SO2"] = pred
test_2015_2016
```

	Year	SO2	state_encoded
132	2015	7.423964	2
133	2015	7.588486	3
134	2015	7.753008	4
135	2015	7.917530	5
136	2015	8.082052	6
...	...	...	...
193	2016	12.699020	32
194	2016	9.737623	14
195	2016	7.434314	0
196	2016	12.205453	29
197	2016	7.598836	1

66 rows × 3 columns

We have predicted SO2 value for the years (2017&18) using the training data set merging the original data set of (2011-2014) and predicted data set (2015-2016) to increase the accuracy of the model created. The image below shows the sample of the predicted data set.

In [14]:
```
test_2017_2018["SO2"] = pred
test_2017_2018
```

	Year	SO2	state_encoded
198	2017	8.109945	2
199	2017	8.275108	3
200	2017	8.440270	4
201	2017	8.605433	5
202	2017	8.770596	6
...	...	...	...
259	2018	13.407682	32
260	2018	10.434754	14
261	2018	8.122477	0
262	2018	12.912194	29
263	2018	8.287640	1

66 rows × 3 columns

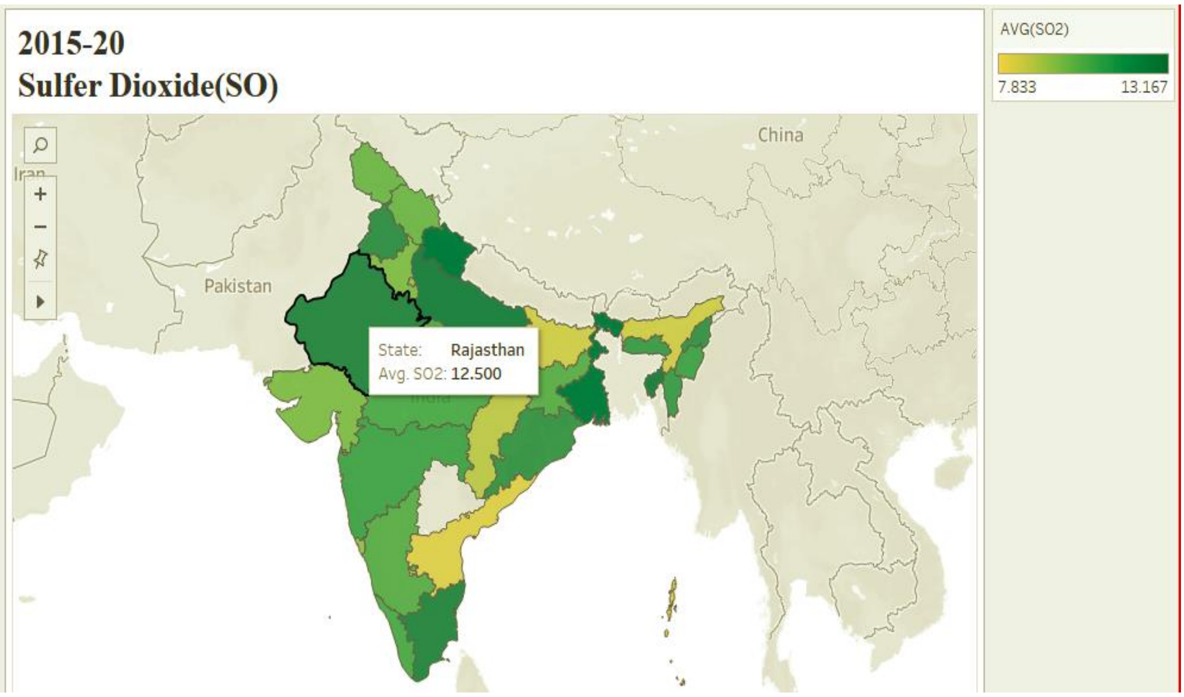
We have predicted SO value for the years (2019&2020) using the training data set merging the original data set of (2011-2014), predicted data set of years (2015-2016) and predicted data set of years (2017-2018) to increase the accuracy of the model created. The image below shows the sample of predicted data set of (2019-2020).

```
In [12]: test_2019_2020["SO2"] = pred
         test_2019_2020
```

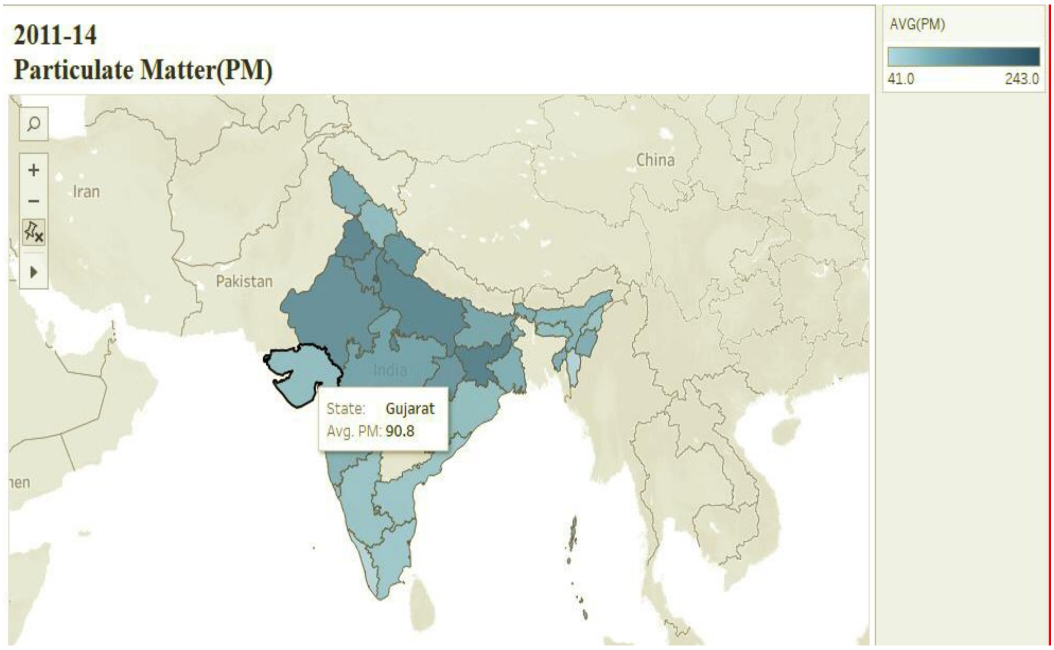
	Year	SO2	state_encoded
264	2019	8.773650	2
265	2019	8.939092	3
266	2019	9.104533	4
267	2019	9.269974	5
268	2019	9.435415	6
...	...	...	...
325	2020	14.076713	32
326	2020	11.098771	14
327	2020	8.782595	0
328	2020	13.580389	29
329	2020	8.948036	1

66 rows × 3 columns

Once our data of SO2 for the years (2015-2020) are predicted we plotted the data in a graph to compare the changes in SO2 taken place during the years (2015-2020) from (2011-2014). The image below is the sample graph of Sulphur dioxide.

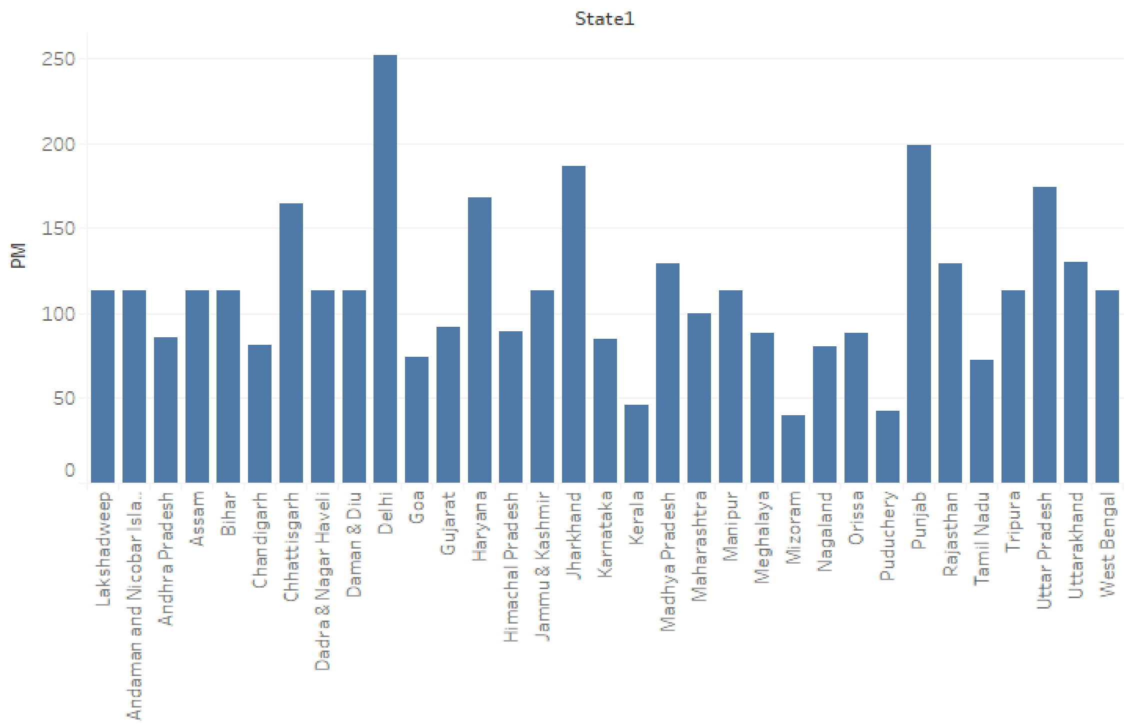


As the above graph depicts that the expected average value of SO2 of Himachal Pradesh for the year is 2021. Comparing the data with the data plotted from 2011-2014 shows a clear increase in SO from the year (2011-2014) to (2015-2020) with an increase from 7 to 12 in average value.

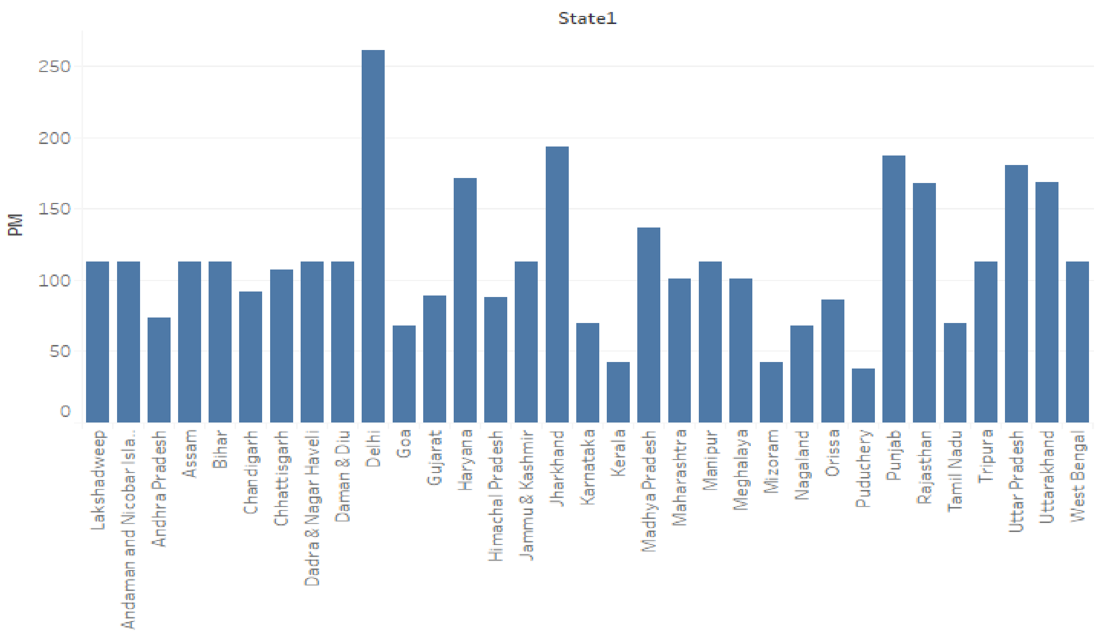


We have collected the data of PM emission for 33 states/union territories from 2011-2014. The image above shows a sample of the map plotted. As the graph depicts the average PM value from the year 2011-2014. We can see Gujarat has an average PM value of 90.8.

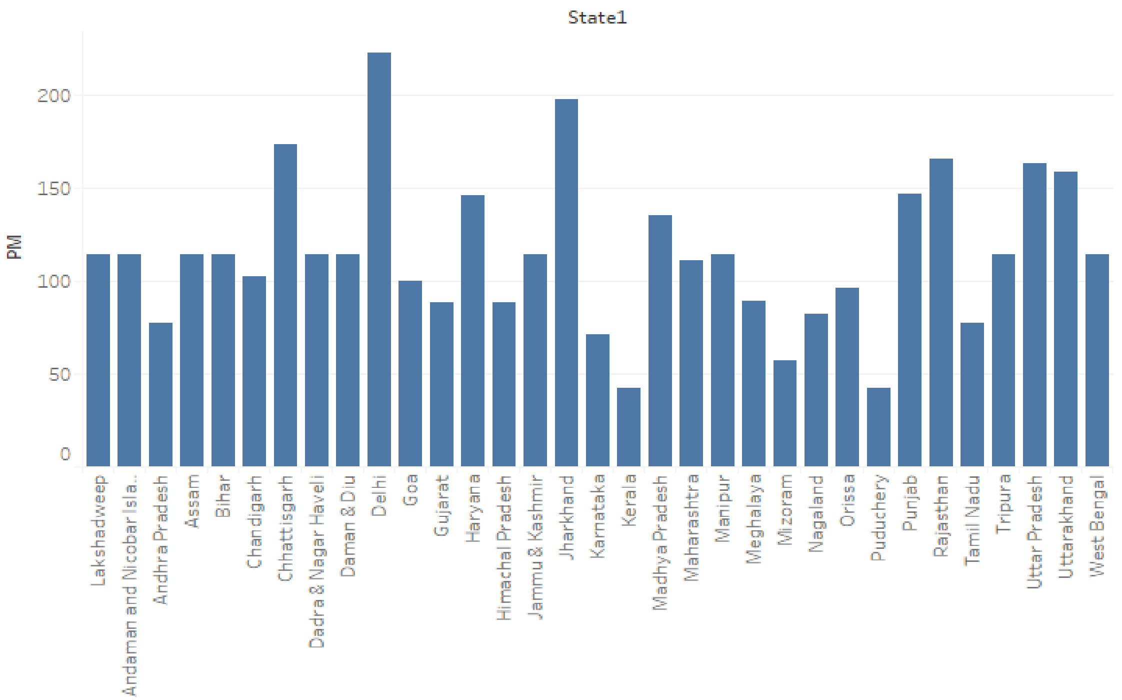
PM 2011



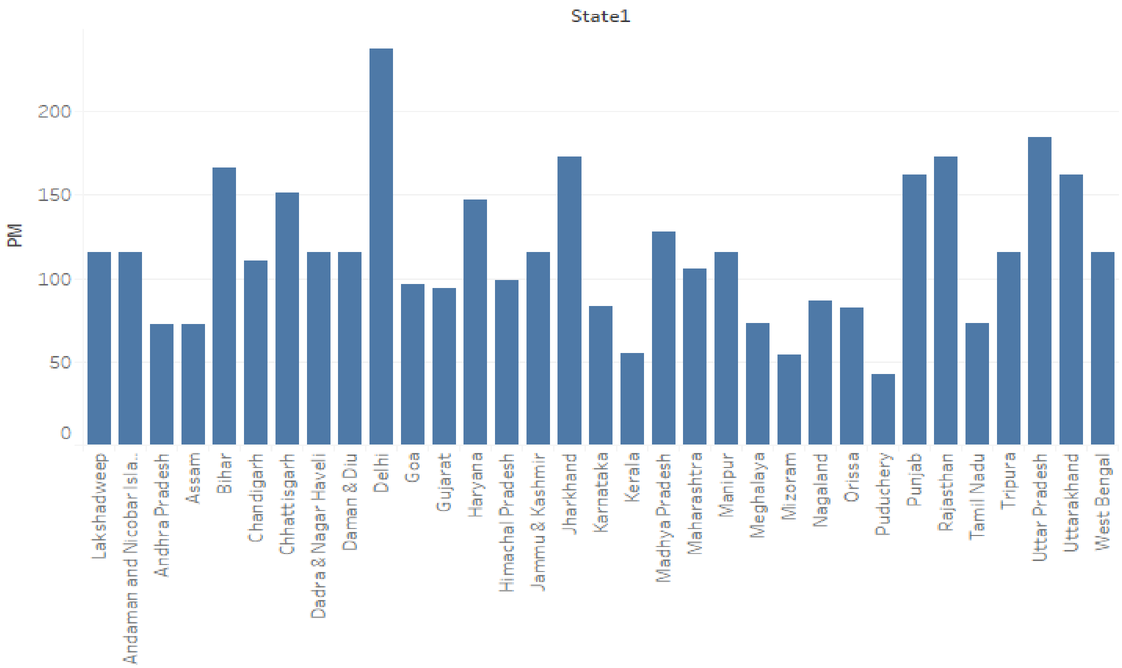
PM 2012



PM 2013



PM 2014



ERROR PERCENTAGE OF PREDICTED AND ACTUAL FOR PM.

(Top 7 and Bottom 5 states for PM)

STATE	2014(ACTUAL)	2014(PREDICTED)	ERROR PERCENTAGE
DELHI	237	245	3.3%
UTTAR PRADESH	184	173	4.8%
JHARKHAND	173	192	10.9%
RAJASTHAN	173	154	10.9%
BIHAR	166	114	13.25
GOA	96	81	15.6%
TAMIL NADU	73	73	0%
KERALA	55	44	20%
PUDUCHERRY	42	41	2.3%
MIZORAM	54	46	14.8%
		Overall error %	9.58%

For conducting further study, we predicted PM value for all 33 states for the year (2015&2016) using the data of (2011-2014) as a training data set. The image below shows the sample of predicted data.

In [88]:
```
test_2015_2016["PM"] = pred
test_2015_2016
```

	Year	PM	state_encoded
132	2015	116.037043	2
133	2015	116.019246	3
134	2015	116.001448	4
135	2015	115.983651	5
136	2015	115.965853	6
...	...	...	...
193	2016	116.294029	32
194	2016	116.614383	14
195	2016	116.863547	0
196	2016	116.347421	29
197	2016	116.845750	1

66 rows × 3 columns

We have predicted PM value for the years (2017&18) using the training data set merging the original data set of (2011-2014) and predicted data set (2015-2016) to increase the accuracy of the model created. The image below shows the sample of the predicted data set.

In [49]:
```
test_2017_2018["PM"] = pred
test_2017_2018
```

	Year	PM	state_encoded
198	2017	117.771442	2
199	2017	117.752559	3
200	2017	117.733675	4
201	2017	117.714791	5
202	2017	117.695908	6
...	...	...	...
259	2018	118.022248	32
260	2018	118.362154	14
261	2018	118.626526	0
262	2018	118.078899	29
263	2018	118.607642	1

66 rows × 3 columns

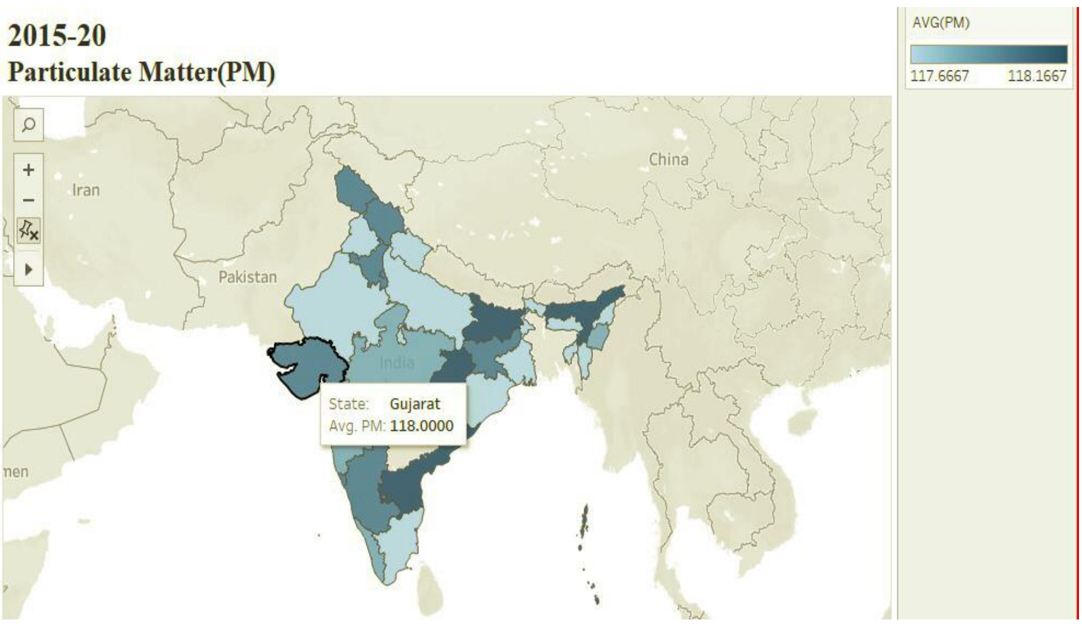
We have predicted PM value for the years (2019&2020) using the training data set merging the original data set of (2011-2014), predicted data set of years (2015-2016) and predicted data set of years (2017-2018) to increase the accuracy of the model created. The image below shows the sample of predicted data set of (2019-2020).

```
In [31]: test_2019_2020["PM"] = pred
         test_2019_2020
```

	Year	PM	state_encoded
264	2019	119.419794	2
265	2019	119.396148	3
266	2019	119.372501	4
267	2019	119.348855	5
268	2019	119.325208	6
...	...	...	...
325	2020	119.518844	32
326	2020	119.944479	14
327	2020	120.275528	0
328	2020	119.589783	29
329	2020	120.251882	1

66 rows × 3 columns

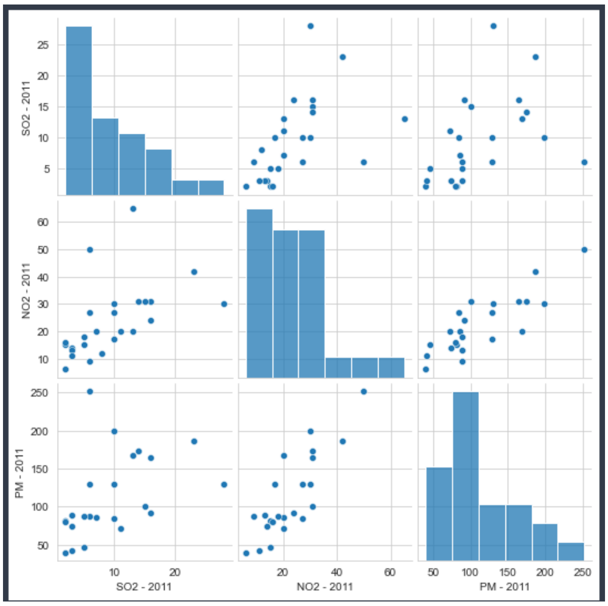
Once our data of PM for the years (2015-2020) are predicted we plotted the data in a graph to compare the changes in PM taken place during the years (2015-2020) from (2011-2014). The image below is the sample graph of Particulate Matter.



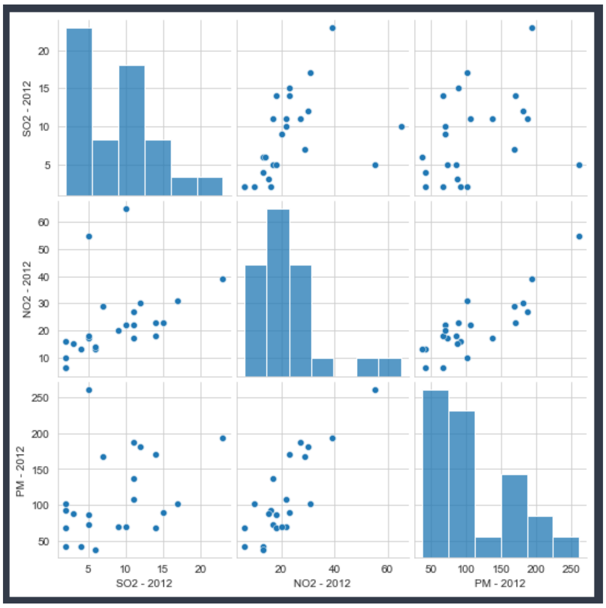
As the above graph depicts that the expected average value of PM of Gujarat for the year is 2021. Comparing the data with the data plotted from 2011-2014 shows a clear increase in PM from the year (2011-2014) to (2015-2020) with an increase from 90.8 to 118 in average value.

Correlation: It is the process of establishing a relationship or connections between two or more things.

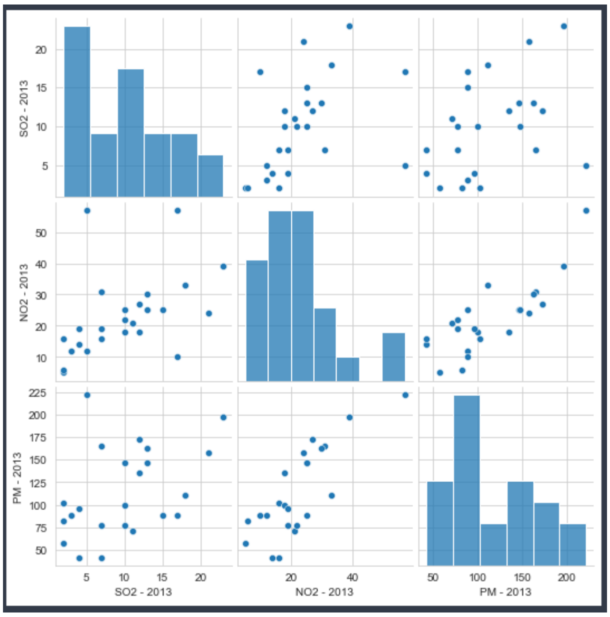
We have taken the correlation of the pollutants SO2, NO2 and Particulate Matter per year to identify the relationship between the pollutants and to analyze the changes taken place during each year.



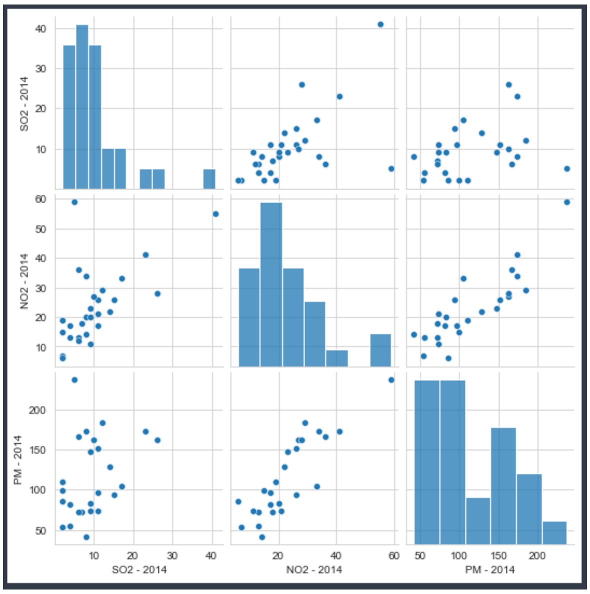
The above graph depicts the correlation of pollutants during the year 2011 and analyzing the above graph shows the best correlation between the pollutants of NO2 and PM. And the mostly scattered graph is shown between SO2 and PM.



The above graph depicts the correlation of pollutants during the year 2012 and analyzing the above graph shows the best correlation between the pollutants of NO2 and PM. And the mostly scattered graph is shown between SO2 and PM.



The above graph depicts the correlation of pollutants during the year 2013 and analyzing the above graph shows the best correlation between the pollutants of NO2 and PM. And the mostly scattered graph is shown between SO2 and PM.



The above graph depicts the correlation of pollutants during the year 2014 and analyzing the above graph shows the best correlation between the pollutants of NO2 and PM. And the mostly scattered graph is shown between SO2 and PM.

After analyzing all the four years CO2 relation between the pollutants it shows a major pattern that the NO2 and PM shows the highest CO2 relation.

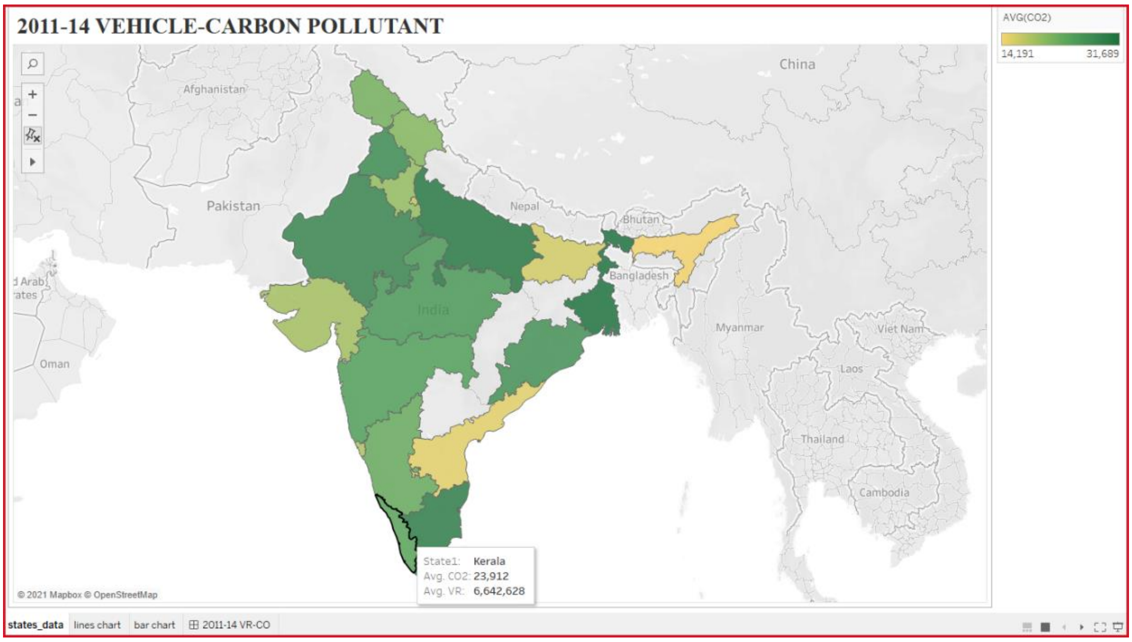
State	CO2	VR	Year
TOP 5 State/Union Territory for CO2			
West Bengal	33253	6745004	2014
Uttar Pradesh	32280	19114692	2014
Tamil Nadu	31308	20863662	2014
Rajasthan	30336	11133420	2014
Punjab	29364	6262939	2014
Bottom 5 State/Union Territory for CO2			
Gujarat	20615	17091599	2014
Goa	19643	1009362	2014
Delhi	18671	8292757	2014
Bihar	17698	4163396	2014
Andhra Pradesh	16726	7002143	2014

State	CO2	VR	Year
TOP 5 State/Union Territory for VR			
Andhra Pradesh	16726	7002143	2014
West Bengal	33253	6745004	2014
Punjab	29364	6262939	2014
Bihar	17698	4163396	2014
Assam	15754	2217007	2014
Bottom 5 State/Union Territory for VR			
Tamil Nadu	31308	20863662	2014
Uttar Pradesh	32280	19114692	2014
Gujarat	20615	17091599	2014
Rajasthan	30336	11133420	2014
Delhi	18671	8292757	2014

Phase 2: Analyzing the relationship between vehicle registered and carbon pollutants

Analyzing the relationship between vehicle registered and carbon pollutants emitted across the state. For conducting this study, we have used a data set of vehicles registered across the 16 states from the years 2009-2015 and carbon emitted across the 16 states for the years 2009-2014.

The image above shows a sample of the map plotted using the data set of vehicles



registered and carbon emitted during the years (2011-2014). We can see Kerala has an average CO2 value of 23,912 and 66,42,628 vehicles registered.

For conducting further study and estimating the CO2 emitted for the year 2015. We have created a machine learning model to predict the CO2 emitted for the year 2015 using the data set of CO2s emitted and vehicle registered during the year 2011-2014 as training data set and considering the vehicle registered during the year 2015 as testing data set.

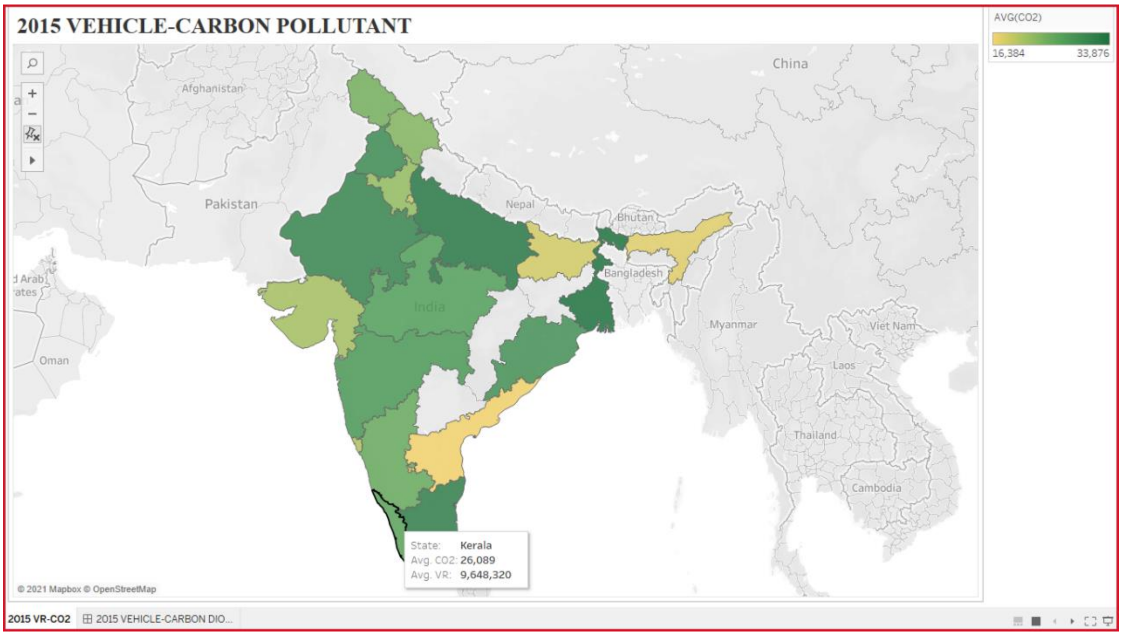
```
test_2015["CO2"] = pred
test_2015
```

	Year	VR	CO2	state_encoded
114	2015	7882262	16383.879386	0
115	2015	2509737	17389.828871	1
116	2015	4777596	18346.951333	2
117	2015	8850720	19292.536977	3
118	2015	1083668	20313.789047	4
119	2015	18720567	21172.693347	5
120	2015	7927550	22213.283301	6
121	2015	1077404	23228.675745	7
122	2015	14784961	25084.306651	9
123	2015	9648320	26088.748684	10
124	2015	25562175	27930.279911	12
125	2015	11141127	27050.824232	11
126	2015	5218595	29031.904175	13
127	2015	6262939	29996.845693	14
128	2015	12378929	30929.376095	15
129	2015	22518669	31836.192117	16
130	2015	21635531	32813.451499	17
131	2015	7403241	33876.020645	18
132	2015	1244271	24199.224912	8

The image above shows the sample of predicted data set of CO2 emitted during the year 2015 based on the vehicle registered during the year 2015.

We have checked the accuracy of the model created for predicting and the accuracy was 99.55%. The below graph shows the accuracy checked.

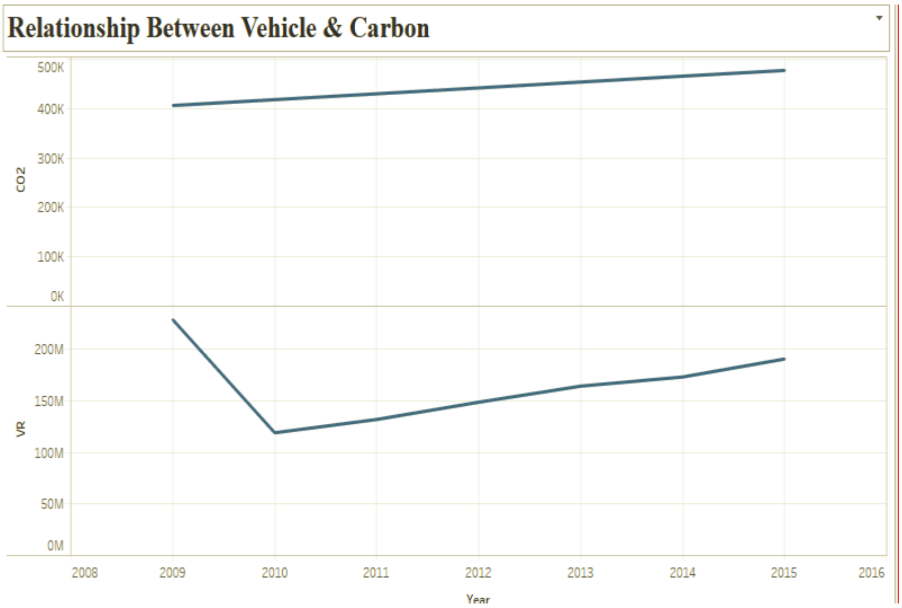
```
In [34]: lr = LinearRegression()
         lr_fir = lr.fit(x_train,y_train)
         pred = lr.predict(x_test)
         pred

array([24867.04390284, 32176.04947639, 19373.33167099, 21395.02559469,
       26886.20166313, 30588.63301895, 18722.65144136, 29979.9374367 ,
       28741.59199204, 26102.60217508, 27506.82521989, 25481.65554253,
       33223.15027069, 19985.04938502, 18762.01932691, 20597.02659192,
       23260.3938386 , 13733.74818765, 16212.02936858, 18496.10921597,
       30958.85358373, 22014.73507977, 31198.56223781])

In [36]: lr_fir.score(x_test,y_test)*100

99.55579715315086
```



The image above shows a sample of the map plotted using the predicted data set of vehicles registered and carbon emitted during the years (2015). We can see Kerala has an average CO2 value of 26,089 and 96,48,320 vehicles registered.

Comparing the above maps plotted using the data set of vehicles registered and carbon emitted during the year (2011-2014) and 2015. The maps show an overall increase in carbon emitted as per the vehicle registration increases. Example average CO2 emitted in Kerala with 66,42,628 vehicles registered was 23,912 and when the vehicle registration increases to 96,48,320 along with-it carbon emitted also increases to 26,089.

- **To analyze the impact of increase in vehicle registered on CO2 emitted we have plotted the line graph.**



The above graph clearly depicts an increase in CO2 emitted as there is an increase in vehicle registration.

Analyzing the above graph, it represents the relationship between the vehicle and carbon emission for the states. And the graph shows and dip in vehicle registered during the year (2009-2010) and after that there is a gradual growth of vehicle registration during the upcoming years from 2010 to 2015 and CO2 emission graph also shows a directly proportional relation with the vehicle registration as the vehicle registration increasing during the years carbon emission also shows a rate in increase from 2009 to 2015.

Tableau public for original & predicted pollutants:

https://public.tableau.com/profile/manav.shah2917#!/

Conclusion

Phase 1:

The agenda of this study was to use technology for creating awareness to decrease pollution by adopting proper measures. India is already ranked in the top 3 countries among the world's most polluted countries, so we considered it in the study. Few states with a high population load were selected. Python and tableau have been used for predicting future pollution levels in Indian states.

As the result indicates, the increasing trend of NO2, SO2 and PM in coming years can be attributed to increasing air pollution. Increase in these harmful gases can be attributed to dust, serious health issues mainly in infants, children and elderly people. This may result in the development of smog pollution. And if we look at it in terms of long-term prediction, then the atmosphere which includes air, water, humidity does not remain as it is now and this can result in worse air quality.

On the basis of previous years data, we predicted the emission of NO2, SO2 and PM for the year 2020. The predicted data would have been useful if there was no lockdown. Due to the pandemic, there was a lockdown in all the states and because of that transportation had come to a halt. This had a major fall in pollution and the air index increased positively. The states which had maximum pollution have also shown a decrease in the emission of pollutants due to lockdown. The year 2020 was seen to have the least emission of pollutants in a few decades.

Once the pollution data of 2020 would have been available then we could have compared the original data with the predicted one and would have analyzed the change in air pollution due to transportation and industry.

Phase2:

After analyzing the relationship between vehicle registration and carbon pollutants emitted across the states in India. It clearly shows that as vehicle registration is increasing carbon emission is also increasing and if the trend is going to be followed there will be a huge emission of carbon emission in the upcoming years leading to climate change and health issues. So the government should make clear policy in vehicle registration for decreasing the pollution.

- We mentioned a GitHub link of all the codes and datasets to get a deeper insight of how we analyzed the data and implemented it on tableau and python.
- all code link (https://github.com/manav3475/DDG-Model-Codes)
- all datasets link (https://github.com/manav3475/DDG-)

Insights for data driven governance

- As our analysis clearly predicted a rapid increase in pollution in upcoming years. The government can make data driven decisions from our analysis. We are trying to suggest some data driven decisions that can be taken by the government as follows.
- From our analysis and data govt. can easily identify the states which are performing good and bad in pollution control. Which clearly displays which of their states are benefited from the policies which are already made and which states need more strict policies to be made.
- The states with high pollution rate the govt should make policies to banning commercial vehicles which are more that 15 to 20 years old.
- Replacement all taxis and autos which are below BS3 stage and promote vehicles above stage BS4.
- Pollution testing of all vehicles should be made compulsory under the state RTO department only.
- Availability of improved and efficient public transports like local trains, metro, electric buses etc. which can reduce the use of private vehicles to an extent.
- Adoption and popularization of carpooling are an effective way to reduce pollution so that the government should take measures to ensure travelers safety.
- Make suitable steps in road development where the congestion is more during the rush hours which leads to more pollution.
- More use of electric vehicles should be entrained by the government making policy by providing charging stations across the country.
- Awareness campaigns can be done among people to help build a public awareness right from the Public level. Major health effects and environmental damages they can be aware of so that they can implement from their own extent and try to avoid things which result in extensive emission of pollutants.

Lessons learned

- While working on our project we found some difficulties but as we started proceeding, we got to know data better. Few problems we faced during data collection as the data was not up to date.
- In that case we predicted the data for 6 years (2015-2020) on the basis of 4 years (2011-2014) of SO_2, NO_2 and PM. And for CO_2 we used the dataset of vehicles registered state wise and carbon pollutants during (2009-2014).
- Then we started with data pre-processing. By cleaning the data sets. Then analysis on pre-processed data was done. We learned the analysis of the relationship between these datasets.
- While working with analysis and pre-processing of datasets we learned various algorithms and also the correlation matrix of the pollutants and analyzed data. Creating a machine learning model played the most important role in our project. On this we learned the machine learning model's accuracy. This model accuracy helped us in plotting the predicted data.
- We learned the decision-making techniques and the conclusions based on the analysis. This helped in getting the results.
- From this data analysis and prediction, we can conclude that it is a crucial aspect and predicted data can help in betterment of future pollution emission levels. We had an approach truly based on data-driven. We learned how data analysis can help us in day to day instances and also it helps us to deal with significant problems.
- This is the analysis of Indian states/territories which are slowly facing major pollution levels and are turning into a gas chamber. Not only India, many other countries are also facing major pollution levels resulting in harmful diseases and environmental damages.
- We learned that we should use measures to control this serious issue. Various methods of controlling this situation have already been mentioned above. We must find a cure for this problem as it is killing our nation slowly.

Bibliography

1) A. Sadatshojaie and M. R. Rahimpour, “Chapter 1 - CO2 emission and air pollution (volatile organic compounds, etc.)–related problems causing climate change,” in *Current Trends and Future Developments on (Bio-) Membranes*, A. Figoli, Y. Li, and A. Basile, Eds. Elsevier, 2020, pp. 1–30. doi: 10.1016/B978-0-12-816778-6.00001-1.

2) A. G. Sikirulahi and K. A. Salami, “contributions of CO2, NO_2 and SO_2 from automobile emission to environmental problems in Niger state, Nigeria,” 2013, Accessed: May 22, 2021. [Online]. Available: https://pubag.nal.usda.gov/catalog/7102381

3) C. Srivastava, S. Singh, and A. P. Singh, “Estimation of Air Pollution in Delhi Using Machine Learning Techniques,” in *2018 International conference on computing, Power and communication Technologies (GUCON)*, Sep. 2018, pp. 304–309. doi: 10.1109/GUCO2N.2018.8675022.

4) N. Sharma, S. Taneja, V. Sagar, and A. Bhatt, “Forecasting air pollution load in Delhi using data analysis tools,” *Procedia computer Science*, vol. 132, pp. 1077–1085, Jan. 2018, doi: 10.1016/j.procs.2018.05.023.

5) A. Aggarwal and D. Toshniwal, “Frequent Pattern Mining on Time and Location Aware Air Quality Data,” *IEEE Access*, vol. 7, pp. 98921–98933, 2019, doi: 10.1109/ACCESS.2019.2930004.

6) D. H. Stedman, G. A. Biship, and A. Peddle, “On-Road Motor Vehicle Emissions including NH3, SO2 and NO2,” Art. no. CARBRA-07-319, Oct. 2009, Accessed: May 22, 2021. [Online]. Available: https://trid.trb.org/view/928590

7) S. Taneja, N. Sharma, K. Oberoi, and Y. Navoria, “Predicting trends in air pollution in Delhi using data mining,” in *2016 1st India International conference on Information Processing (IICIP)*, Aug. 2016, pp. 1–6. doi: 10.1109/IICIP.2016.7975379.

8) M. Yadav, S. Jain, and K. R. Seeja, “Prediction of Air Quality Using Time Series Data Mining,” in *International conference on Innovative computing and communications*, Singapore, 2019, pp. 13–20. doi: 10.1007/978-981-13-2354-6_2.

9) M. Kandlikar and G. Ramachandran, “THE CAUSES AND CONSEQUENCES OF PARTICULATE AIR POLLUTION IN URBAN INDIA: A Synthesis of the Science,” *Annual Review of Energy and the Environment*, vol. 25, no. 1, pp. 629–684, 2000, doi: 10.1146/annurev.energy.25.1.629.

10) T. Goshala and R. Bhattacharyya, “Carbon Dioxide Emissions of Indian States: An Update,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2166900, Oct. 2012. doi: 10.2139/ssrn.2166900.

Machine Learning packages documentation:

1) https://matplotlib.org/
2) https://numpy.org/
3) https://scikit-learn.org/stable/
4) https://seaborn.pydata.org/#:~:text=Seaborn%20is%20a%20Python%20data,can%20read%20the%20introductory%20notes