3RI Technologies

3RI Technologies Pvt. Ltd.
Office No. 405, 4th Floor, Rainbow Plaza,
Pimple Saudagar, Pune - 411017.
An ISO 9001:2015 Certified Institute
Phone: +91 20 6500 3366, +91 830 810 3366

U72900PN2013PTC149923

Date: 14-Jun-2017

This is to certify that **Mr. Abhishek Hirve** Student of B.B.A (Computer Application) from **"Modern College ,Wing of Computer and Business Studies, Pune "** has worked as an Intern on our project "**WebSite Development using PHP Technology**" and developed a user friendly website as per our requirements and expectations.

This internship is viewed by us as being an educational opportunity for him. It has included training/orientation and primarily focused on learning and developing new skills and gaining a deeper understanding of concepts through hands-on application knowledge.

This is undertaken as a part of the Internship Training for B.B.A(Computer Application) students under the guidance of **Ms. Nirali Vaghela.** We appreciate Abhishek's efforts he has taken for project completion.

Thanks and Regards,
Rupali Sharma

Authorised Signature
3RI Technologies Pvt. Ltd.

Date : 31/8/2018

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. **Abhishek Digamber Hirve** was working with **Source Tech Systems** from 1st of March 2018 to 31stof August 2018.
The designation held by him at the time of leaving was **Intern Software Developer.**
His brief experience summary during the stay with the organization is given below :

- He has experience in Software Development with .Net framework and C++.
- Also Has experience in Web Application Development as contract and clients requirements assisting senior developer. HTML5,CSS3.0 , PHP/MY SQL , AJAX+(JSON) , XML with Xampp.
- Also excellent in PL/SQL, JAVA (advance) + J2EE on (TOMCAT) and JAVA SCRIPTING based work.
- He also performs well with his software engineering skills and helps senior engineers for documenting and peer code reviews.

For **Six Months** of working for us, he demonstrated as a diligent and truthful person. His interpersonal skills are outstanding. We wish him the very best in his future endeavors.
This letter is issued on the request of the employee.

We had paid stipend Rs.10,000/- Per Month.

Authorized Signatory

For Source Tech Systems

iON
TATA CONSULTANCY SERVICES

Digital Hub
Learn, Share, Collaborate

CERTIFIED

Cert. ID.: 72450-2689422-1016
Dated: 28 May 2017

This is to certify that

abhishek hirve

has successfully completed

Database Design

online course offered by NPTEL

Course End Date: 28 May 2017

Topics:

- Introductory Concepts
- Semantic Database Design
- Relational Model, Languages and Systems
- Database design using the relational model
- Storage and Indexing Structures
- Query Processing and Optimization
- Transaction Processing and Concurrency Control
- Distributed databases and issues
- Database recovery techniques
- Enhanced Data Models
- Case Study - Database Design

(Faculty)

TCS iON - Empire Plaza, Lal Bahadur Shastri Marg, Chandan Nagar, Vikroli West, Mumbai, Maharashtra 400083

NVIDIA. DEEP LEARNING INSTITUTE

IN ASSOCIATION WITH

IISER PUNE TIMESPRO

CERTIFICATE OF ATTENDANCE

This certificate is awarded to

Abhishek Hirve

for completion of

- Image classification using DIGITS™
- Image Segmentation using DIGITS™
- Neural Network Deployment™ using TensorRT™

Mark Ebersole
Director, NVIDIA Deep Learning Institute

Anish Srikrishna
President, Times Professional Learning

Date issued: **14/12/2017**
Venue: **Indian Institute of Science Education & Research, Pune**

Record of Achievement

openSAP is SAP's platform for open online courses. It supports you in acquiring knowledge on key topics for success in the SAP ecosystem.

abhishek hirve

Date of birth: November 21, 1997
E-mail: abhishek.hirve@yahoo.com

has successfully completed the following openSAP course:

Developing Web Apps with SAPUI5

Instructors: Janina Bläsius, Christiane Goebels, Michael Graf, Andreas Kunz, Thomas Marz, Thilo Seidel, and Martin Wittmann

This four-week online course was held from May 24 through June 28, 2016. It comprised 4-6 hours of learning effort per week, optional coding exercises, 4 weekly assignments, and 1 final exam.

The course covered the following topics:

- Learn the Essentials
- Become a Data Binding Expert
- Create Responsive Apps
- Master SAPUI5

Maximum score possible for this course: 240 points.

Dr. Bernd Welz
Executive Vice President
SAP Scale, Enablement & Transformation

Stefan Beck
Chief Product Owner
SAPUI5

Michael Graf
Course Instructor

openSAP

CERTIFICATE of ACHIEVEMENT

This is to certify that

Abhishek Hirve

has completed the course

SAS Programming 1: Essentials

May 4, 2020

Software AG
University Relations

Certificate

This certificate is presented to

Abhishek Hirve

for gaining expertise in the **ARIS Business Process Analysis Platform**

in corporation with **Software AG's University Relations**.

The participant has shown to be knowledgeable in:

- Fundamentals of Business Process management (BPM) and the principles of BPM methodology
- Design, analysis and evaluation of business processes with ARIS platform
- ARIS Enterprise Architecture
- Modelling business processes with Event-driven Process Chains (EPC) and Business Process Modeling Notation (BPMN)

November 30, 2019

Awarded on

Jürgen Powik
Senior Director University Relations
Software AG

Abhishek Hirve

has successfully completed the AWS Certification requirements and has achieved their:

AWS Certified Cloud Practitioner

Issue Date
Apr 16, 2019

Expiration Date
Apr 16, 2022

Maureen Lonergan
Director, Training and Certification

Validation Number DR40CZZ1DFFE1EG1
Validate at: http://aws.amazon.com/verification

Certificate

This is to certify that:

Abhishek Hirve

Has successfully passed the certification exam for
SCRUM FOUNDATION PROFESSIONAL CERTIFICATE (SFPC)

Date of Certification

Wed 24th Jun 2020

ID CP:

JLSRFLSHSP-VVNZCHCH-YBJQDBBWTB

MANAGING DIRECTOR

SCRUM FOUNDATION PROFESSIONAL CERTIFICATE (SFPC)

INTERNATIONAL MODEL UNITED NATIONS

Australian Embassy
Vietnam & Thailand

Certificate of Appreciation

This Certificate Recognizes with Appreciation

ABHISHEK HIRVE

For Successfully Participating as a Delegate in

IMUN ONLINE CONFERENCE 31.0

from November 14th - 15th, 2020

Mohneesh Bhardwaj
Executive Chairman

Galib Akhand
Academic Advisor

Progressive Education Society's

Modern College of Arts,Science & Commerce

Business Administration Campus

Shivajinagar, Pune-411005.

CERTIFICATE

This is to certify that

Abhishek Digambar Hirve Roll No:32021

Ketan Muralidhar Kulawade Roll No:32040

Have satisfactorily completed the project, entitled

Wireless security management System.

For the **Project Management** course towards partial fulfillment of **B.C.A. (Sem-V)** course of **Savitribai Phule Pune University** for the academic year **2017-2018** as a part of curriculum.

Date: 4/10/2017

College Stamp

Project Guide 4/10/17

Internal Examiner 16/10/17

External Examiner

PROJECT PERFORMANE CERTIFICATE

It is certify that the ***Wireless security management System.***

Project undertaken by Abhishek Digambar hirve and Ketan Murlidhar Kualawade is working satisfactorily .
The next version of system will betterly satisfy extra requirments and needs.
I would thankfull to project guide and teem for providing Experties for said project.

Date:- 05/11/2017

(Maharashtra police wireless)

Pune city police.

ACKNOWLEDGEMENTS

We have a great pleasure to present this project on-

Wireless security management System.

We have thankful to **P. E. Society's Modern College of Arts, Science & Commerce, Shivajinagar, Pune-41005.** For giving us the opportunity to carry out **B.B.A.-C.A (Sem-V)** project at their concern.

Many people have contribution in this project report, we are grateful to **Prof .M.D.Waghmare**, the Vice-Principal for his co-operation. Our special thanks to ***MRS.SHRUTI PAWAR*** for her constructive suggestions and comments and for lending her expertise and talent when needed. And above all, we thank the Almighty for his grace and blessings.

Finally, we would like thank each and everyone who was directly or indirectly involved with this project report.

Progressive Education Society's
Modern College of Arts,Science & Commerce
Business Administration Campus
Shivajinagar, Pune-411005.

CERTIFICATE

This is to certify that

Abhishek Digambar Hirve Roll No:32021 DIV:-A

Vineet Nair Roll No:32182 DIV:-B

Have satisfactorily completed the project, entitled

Aircraft Maintenance management System.

For the **Project Management** course towards partial fulfillment of **B.C.A. (Sem-VI)** course of **Savitribai Phule Pune University** for the academic year **2017-2018** as a part of curriculum.

Date:-

College Stamp

29/5/18

Project Guide

Internal Examiner

External Examiner

ACKNOWLEDGEMENTS

We have a great pleasure to present this project on-

Aircraft Maintenance management System..

We have thankful to **P. E. Society's Modern College of Arts, Science & Commerce, Shivajinagar, Pune-41005**. For giving us the opportunity to carry out **B.B.A.-C.A (Sem-VI)** project at their concern.

Many people have contribution in this project report, we are grateful to **Prof .M.D.Waghmare**, the Vice-Principal for his co-operation. Our special thanks to ***MRS.SHRUTI PAWAR*** for her constructive suggestions and comments and for lending her expertise and talent when needed. And above all, we thank the Almighty for his grace and blessings.

Finally, we would like thank each and everyone who was directly or indirectly involved with this project report.

WAC|2017
WHAT AFTER COLLEGE

COE PUNE 14TH & 15TH JANUARY (SATURDAY & SUNDAY)

Powered by

KYRION Technologies

CERTIFICATE OF PARTICIPATION

This certificate is awarded to

Abhishek Digambar Hirve

in recognition for participation in

ETHICAL HACKING & CYBER FORENSICS WORKSHOP

conducted by

Kyrion Technologies Pvt. Ltd.

on **14th & 15th January 2017**

at **College of Engineering, Pune**

Rishabh Mehta
Chairman & Editor-in-chief
What After College

Sujata Mehta
Chief Executive Officer
Kyrion Technologies Pvt. Ltd.

Skill Development Partner

N·S·D·C
National
Skill Development
Corporation

Transforming the skill landscape

IT CHAMP

RUNNER UP
WINNER

This is to certify that

Mr. Abhishek Digambar Hirve

Won the RUNNER UP Position & 5% Scholarship in

1st round of Jetking IT Champ - Inter College Quiz Competition in

Modern College Pune

held on 24th January ' 2017.

Jetking
Better Life

SURESH PATEL
Center Director
Jetking Hadapsar Learning Center

HARIPRASAD MARLA
Center Director
Jetking Swargate Learning Center

www.jetking.com | www.nsdcindia.org

Savitribai Phule Pune University

CERTIFICATE OF PARTICIPATION

seed
beyond the obvious

SEED IT Idol

This certificate is hereby awarded to Abhishek Hirve

for active participation in 'SEED IT Idol Competition' jointly organized

by SEED Infotech Ltd. and Savitribai Phule Pune University.

Date : 24/12/16

Executive Director
SEED Infotech Ltd.

MD & CEO
SEED Infotech Ltd.

www.seedinfotech.com

ANALYZING THE PATTERN OF AIR QUALITY IN INDIA

Abhishek D. Hirve (SID=347736209)

Department of Information Systems

The University of Auckland
(city campus)
Auckland, New Zealand

Email: ahir997@aucklanduni.ac.nz

INFOSYS 722: Data Mining and Big Data

Dr Ami Peiris

Jun19,2020

ABSTRACT

This paper Includes study on the pattern of Air pollution across India. The topic is based on the United Nations sustainable development goals. This paper is focusing on the methodology for Data mining and analysis. This paper is based on the practical application on the project carried out on the study of the pattern of Pollutant particles present in the air. The data used for the experiment carried out in this project is released by the Ministry of Environment and Forests and Central Pollution Control Board of India under the National Data Sharing and Accessibility Policy (NDSAP). The paper reflects insights from the experimental data mining carried out with the use of multiple software platforms. The technology stack used in the project is IBM SPSS, Python Pandas library, Pyspark, Tableau. This paper is also useful for decision making in order to generate a business enterprise based on innovative tech products. The methodology used in the proposed project and paper can be universally applicable for cross industry-based data mining and analysis.

Keywords: so2(Sulphur dioxide),no2(Nitrogen dioxide),rspm(respirable suspended perticulate matter) ,spm(suspended perticulate matter),pm2_5(atmospheric particulate matter of 2.5 microns),O3(ozon),NH3(Amonia),AQI(Air Quality Index).

Introduction

As per the count taken by the World Health organization in 2016, approx four million people die due to the effect of breathing decreases caused by Air pollution. Out of 10, every nine people inheld polluted air. Across the world, 7 million people die due to air pollution. Around 90% of deaths occur in mid-income countries or under development countries from Asia, Africa. Around 40% of the world's population do not have access to clean, breathable air. This research focuses on countries Which fall under High population creditors by United nations and mid-level income country in Asia like India. The research will follow Data Analysis collected by official authorities in India and offer solution-oriented insights based on quantitive and qualitative analysis. .("CLEAN AIR FOR HEALTH: Geneva Action Agenda", 2020) (*First WHO Global Conference on Air Pollution and Health*, 2018)

The Practical Problem

In the era of industrialization, India is the world 's largest developing country. As a United Nations report, India is one of the most polluted countries and has low air quality. Poor air quality has a health effect and a dangerous environmental impact. ("Air Pollution on the Move in India", 2020).The highest contaminated air levels in recent years has been reported in New Delhi, which is India's capital city. One of the NASA estimates claims 37% of the sun's rays reach the planet in India's New Delhi city instantly. ("AIR QUALITY FROM SPACE | Air Quality", 2020)

The Research Problem

With the analysis of reports and day to day life within India air quality is the biggest Major causing effect on human health. Also, if we look at global parameters, every highly populated country have issues with Air quality. Overall human health is effecting because of air pollution. Which generate need to study the p[attern of air pollution and make a sustainable solution. Therefore, providing clean and breath free air for people on the road has become a major societal issue. The key factors are to determine the pattern of pollutant matter present in the air.

To resolve the global issue and in context with the Indian population, there is a need to seek a solution to tackle a drop of **AQI** across the world and India. The solution generated with the study of Indian territory and with the aim of sustainable development will be a useful study case or a plan for all High populated countries. Therefore, analyzing the quality of air and providing a solution for the betterment of human health ad earth atmospheric health is necessary.

Research Objective

Project is focusing on the analysis of foreign elements present in the atmosphere in India like **so2** Sulphur dioxide, **no2** Nitrogen dioxide, **rspm** Respirable Suspended Particulate Matter, **spm** Suspended Particulate Matter. Based on local data state-wise from India.

1. To monitor the key indicators associated with foreign elements present in the air in India.

2. detect local trends with data available from different state and provisions.

3. relate the air quality changes to changes in Environmental Policy in India and try to predict solutions.

The report presents some results and findings on real-world simulation on data available. These findings will be useful to develop a sustainable solution to reduce pollutants effect in the air and human health.

Literature Review

Much work has been carried out and implemented on **AQI**(air Quality Indexing) and effect of foreign particles in the air. Also, there is a large amount of literature available for analysis regarding human respiration effect due to air pollution across the world. In this section, we have outlined and examined the work of a few researchers. Which helps to make our theoretical view up on air quality and its effect in context with India and world as it is a global issue raised by WHO and one of the sustainable development goals declared by United Nations.which is base of a combination of SDG 3,7,11,13(Good health and wellbeing, Affordable clean energy, sustainable cities and communities, climate action) ("SDGs .:. Sustainable Development Knowledge Platform", 2020)

The public safety effect of air emissions can be categorized into two groups: acute health result and lasting result from ozone intake and air contaminants from gas and can be measured on mortality levels, hospital attendance and Other clinically relevant health measures. (Deshpande & Ghude, 2020)

With the analysis of reports and the publications by the Indian ministry of earth science, NNAQS, and EPA the significant contributors in air pollution are **No2**(Nitrogen Dioxide), **So2**(Sulpher Dioxide), **NH3**(Ammonia), **O3**(Ozon), **PB**(Lead),**spm**, respond **pm** the **pm** stands for a particulate matter which is counted by per micron rpm and rpm are suspended particles present in the air which are respirable or non-respirable as category states. ("NAAQS Table | US EPA", 2020) **O3** Ozone is also known as a defender of an atmosphere of earth, which is constructed by Photochemical reactions of the light of sun and oxides catalyst pollutants, like the nitrogen oxides (**NOX**). (Deshpande & Ghude, 2020)

Also, with the analysis, multiple reports can found that air pollution is dependent on socio-economic and socio-cultural differences in India. As countries which are under development have different status and different lifestyles across the multiple regions of country.which show the impact on the pollutant immersion rate.

Air pollution ambient is 5th in mortality, as well as 7th in the overall health burden, among significant health risk factors in India. World Health Organization (WHO) reports that. (Public health foundation India, centre for environmental health, 2020)

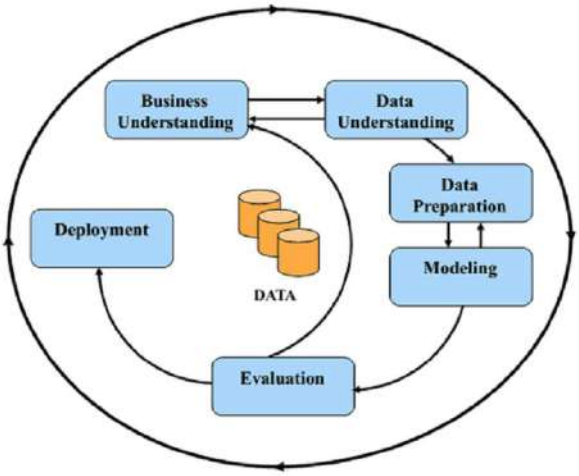
The primary impact is seen by the analysis conducted by WHO on the deceased accused by airpollution.there is a major impact on the health causing respirable and chronic diseases across the world and which is also the highest risk factor in India. (Appendix: A) ("WHO methods and data sources for the global burden of disease estimates", 2020)

There are standards placed on analysing an air quality worldwide called as EPA standards AQI calculation. Calculate the AQI for Clean Air with five primary air pollutants at the ground level of ozone, emission of particulate matter Air quality has a direct effect on (also called particulate matter) quality of life, carbon monoxide, sulphur dioxide and dioxide of Nitrogen. EPA has regional air quality guidelines for the safety of public health regarding each of these contaminants. By which we can categorise our result in healthy or unhealthy threshold. The Global and Indian Standards are documented in Appendix: B for reference, which is universally available and accessible. (APPENDIX: B) ("NATIONAL AIR QUALITY INDEX", 2014)

The Research Methodology

To conduct data mining and to model in this project, we followed steps described by KDD(knowledge discovery in Databases) model for data mining. It is a way to drive through for a journey to convert Initial data into knowledge and insights. The phases followed in the theoretical context of the process in KDD are a selection of data, preparation, transformation, mining, Interpretation and Knowledge. The reason behind the use of KDD is a traditional way to Size the ability of human knowledge to manage several bytes. (Appendix:C) (Fayyad & Uthurusamy, 1996)

With keeping the KDD process in mind, we used CRISP-DM(Cross Industry Reference Process for Data Mining) model methodology in our project to conduct data mining.



CRISP-DM Process (SPSS, 2007)

With the reference model journey for data mining begins with Business understanding. As it is a cross-industry model, the understanding, a business logic behind what we are going to perform in the mining and modelling process is primary stape. This initial step helps us to understand what resources we have and what strategy to turn raw data into information.

Furthermore, understanding the Business logic will help us to determine the objective that needs to be derived from the study of model and patterns. We get clear views and set the goals of data mining objective with this step. Data understanding is the next step in the model in which the collection of data set is done from various authorised platforms. With the proposed research objective dataset variables are studied in this phase. This gives a brief understanding of the information the data holds. When the data is known in the context and intent of the proposed analysis, in the Business understanding process, the collected data are generated by mathematical transmission by means of deletion of mistakes, missed values, incorrect entries, outliers, extremes and other columns to achieve 100% quality of the data. The Appendix :D refferes step followed to dealing with extremes and outlier values. (Appendix:D) This is a preparation phase of data. Once we prepare Data, there is a need to check Null and missing value present in the dataset. This is also known as data cleaning.

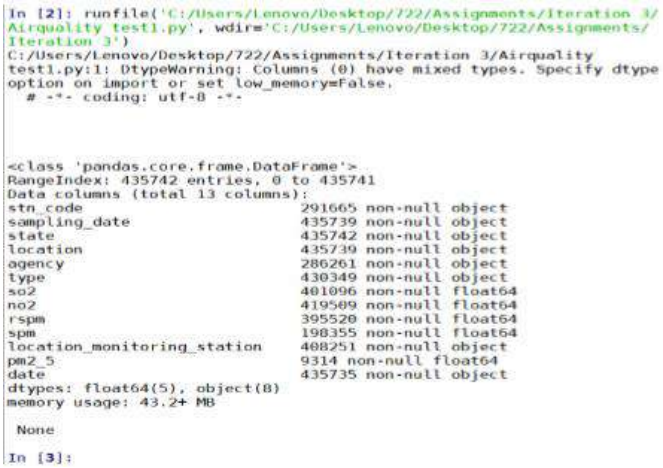
After cleaning, we move to the modelling step forward. In the modelling step, appropriate algorithm analysis and selection is done to achieve mining objective and business goal. In this step, studied data is matched with a bunch of algorithms. When the configuration of selected data matches with selected algorithms working behaviour data is loaded to an algorithm to analyze and make predictions. Furthermore, the applied model is evaluated based on the behaviour of prediction in the model with analysing some error matrix-like root mean squared error, variance etc. these all give the account of workability of the proposed model on our data and rate of accuracy of prediction. With all the above analysis decision is made whether the proposed algorithm and model is the best fit for our data or not, and do we achieve our accuracy goal.

With the accuracy value, multiple models are compared, and the final accurate model is interpreted and deployed into real-world scenarios which are used to give insights. The results from the loaded models are analysed with mind patterns and visualization, which we got while doing Business understanding. If we do not get an accurate rate of accuracy and if do not achieve the success rate, which we got from our business understanding multiple iterations with the model is done. Also, some changes in data and models are done. After the meeting, all the desired results project moves towards the deployment phase.

The Design Process

This section of the report includes real-time analysis and real-time insights and opinion upon designing the project to conduct analysis. This section light-ups some experiences and the issues which are come up with the analysis. This study will help to do triple loop learning which shows characteristics about learning about learning. (Peko, Danenberg & Sundaram, 2013)

The initial data collection and information collection is done multiple resources available online and 'Data.gov.in' online open government data portal India for getting access to related data. Also, the relevant data is found at open source repository. The analytics methods and the software used to load a data are IBM SPSS, pandas library with python, Pyspark. The visualization is done with both analytics platform as well as with Tableau. The visualization obtained from all software applications verifies that dataset contains 435742 entries and 13 columns.



```
In [2]: runfile('C:/Users/Lenovo/Desktop/722/Assignments/Iteration 3/
Airquality test1.py', wdir='C:/Users/Lenovo/Desktop/722/Assignments/
Iteration 3')
C:/Users/Lenovo/Desktop/722/Assignments/Iteration 3/Airquality
test1.py:1: DtypeWarning: Columns (0) have mixed types. Specify dtype
option on import or set low_memory=False.
  # -*- coding: utf-8 -*-



<class 'pandas.core.frame.DataFrame'>
RangeIndex: 435742 entries, 0 to 435741
Data columns (total 13 columns):
stn_code                    291665 non-null object
sampling_date               435739 non-null object
state                       435742 non-null object
location                    435739 non-null object
agency                      286261 non-null object
type                        430349 non-null object
so2                         401096 non-null float64
no2                         419509 non-null float64
rspm                        395520 non-null float64
spm                         198355 non-null float64
location_monitoring_station 408251 non-null object
pm2_5                       9314 non-null float64
date                        435735 non-null object
dtypes: float64(5), object(8)
memory usage: 43.2+ MB

 None

In [3]:
```

The data set contains attributes like stn_code: The unique identifier for each row. sampling_date: Date of the collection of readings. State:States in India. Location: Cities/location from where data is belonging to. Agency:Agency name who collect readings. type: Types of sources or types of locations. **so2**: Amount of Sulphur dioxide present in the air. **no2**: Amount of Nitrogen dioxide present in the air. **rspm**: Amount of Respirable Suspended Particulate Matter present in the air. **spm**: Amount of Suspended Particulate Matter present in the air. location_monitoring_station: location of the monitoring area. **pm2**_5: PSI is the Pollutant Quality Measure is based on the 24-hour maximum PM2 estimate. The other contaminants, five quantities. PM applies to PM2 and particulates. 5 during haze incidents is the primary pollutant. Date: Date of recording in a data set.

While analyzing with the attributes and values we need to take account of data type detection by each attribute for every kind of software, this is the issue raised mostly in SPSS modeller when it detects a wrong type of the attributes. When we deal with data type correction, we have to deal with null and missing parameters detected by the software.

While studying attributes and the data values when we obtain all errors, null and missing values, use integrated functions with several software's to remove null values and errors. The whole data set checked with multiple coding techniques used with inbuilt functions like SPSS modeller as it is an essential task in order to get successful results. When we began confirms details and account of cleaning specifications which are needed to perform we can move further to the removal, substitutes or nullify technique likewise by taking account of our research goal understanding and the characteristics of the attributes in whole research. We use removal for all the numeric columns the extremes and coerce the outliers. (Appendix: E) Appendix: E shows steps which can be considered for data cleaning in cross-industry mining project implementation.

To implement the standardized data inappropriate model, we conduct an exploratory analysis. Before that, we need to finalised our data set, which we are going to loading to algorithms for further analysis.

The Implementation Process

After cleaning the data in this project, we did versioning of the data by conduct mathematical calculation for Index for each pollutant and overall Quality Index. By the universal standardized formula to calculate AQI. The universal standard formula for Air Quality index is:

EQUATION:**{AQI=AQI(min)+PMObs−PM(Min)AQ/Max−AQ(IMin)*(PMMax−PMMin)}**

For calculating the quality index, there are certain range is given per element by EPA for each polutent.which we have used randomly in between a range of threshold value in our calculations. **No2**:0-400 , So2: 0-1600 , **rspm**:0-400 , **spm**:0-430 , **Pm2**_5:0-430 . (Appendix: F) ("Air Quality Index (AQI) Basics", 2020)

After finalising the Data which we are using to conduct analysis Implementation process carried out with multiple software stacks like SPSS modeller (Appendix: G), Pandats Python and Pyspark for visualising and conduct data mining, we can use multiple visualization platforms available in the market like MS Power BI, Tableau. To analyze the data and to equate specific essential characteristics with other important characteristics, we can use Tableau.

Concerning the above discussion, we have run models and different algorithms in different software. As the objective of this study is to monitor the behavioural pattern of Air pollution, some algorithms are adopted in order to conduct analysis. After doing a study of multiple algorithms, we can found that each one has its own constraint with respect to target values. As we have set of continuous type and nominal type of variables, we choose to use the classification technique for doing analysis, including the continuous type variables in SPSS and alternatively clustering technique. Secondly, the records of pollutants show nominal characters we used regression stack for checking the behaviour of those nominal values.

The initial stage of data and analysis:

In the first study which is conducted with IBM SPSS, we break our analysis on initial data which is the format of only collected readings of pollutant and we are analysing pattern with those readings without calculating indexes with SPSS. We decided analysis **Target 1:** (Target contain the amount of SO2 and NO2 which shows the proportion of polluted air gives prediction on gases versus State and Type of resource of pollution) and **Target 2:** (Analysing pattern of RSPM and SPM in the air By mean and distribution state-wise and Type and source of generation).

To analyze the proposed two targets with the Spss, we used supervised models which are in self-learning category. For target 1 we used Kohenen we are analysing state with the proportion of so2 and no2, which represents high dimensional data.kohonen is based on self-organized competitive type learning methods it is also used for experimentally acquired information. For target 2, we used Two stapes To help uncover trends, fragments data into distinct classes. This method is mainly used to compact the data and then slowly combine sub-clusters into mostly controlled sub-clusters. This model handles numerous field styles and broad datasets. (Appendix: H)

Versinasied Data and Analysis:

As discussed previously with the business objective and the universal standards analysis of air pollution, we used AQI calculation and Analysis of pattern using python technology stack.by which we have analysed nominal types of variables such as pollutant readings. While analyzing with the python stack, it gives us the more unobstructed view that before analysis of data which need to construct further the python stack is more useful than SPSS.where we can flexibly adapt the data preparation as per projects requirements. Which is not possible with the SPSS it gives a more visual view of the homogeneous and stationary type data.

To conduct an analysis with AQI and analysing dependencies, correlations in pollutants, we used regression techniques with the python stack with pandas and pyspark. (Appendix: I)

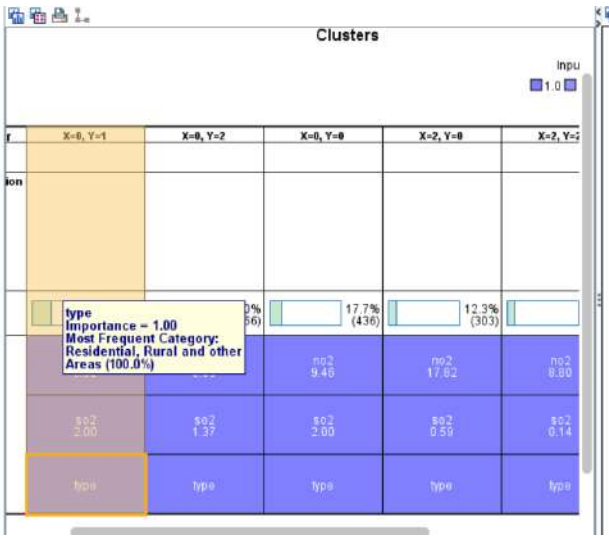
The Interpretation

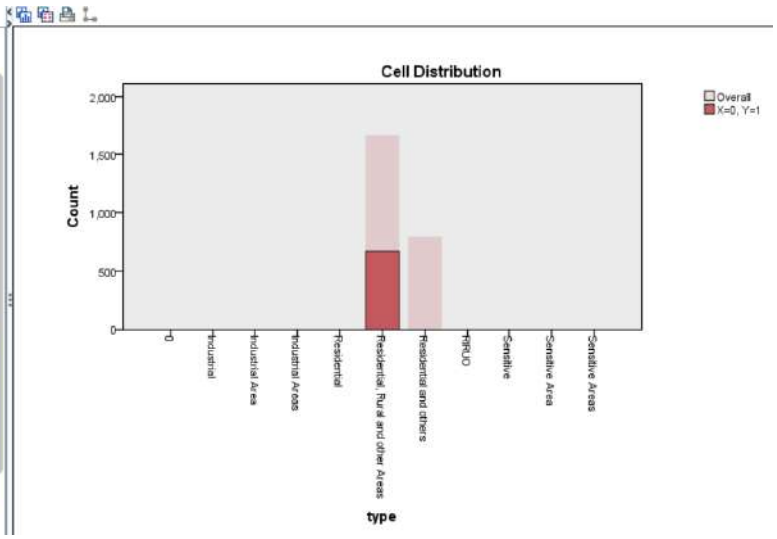
As we are testing the perfect fit of the model in this study splitting of data to get accuracy and correctness of study, and with this, we can make sure the data is deployed in the real world. We used Train-Test split technique for each Algorithms. For getting inaccurate test results and this Method helps to split a dataset and adopt an optimal pattern quickly. We have a split test for 30% and train the data 70%.

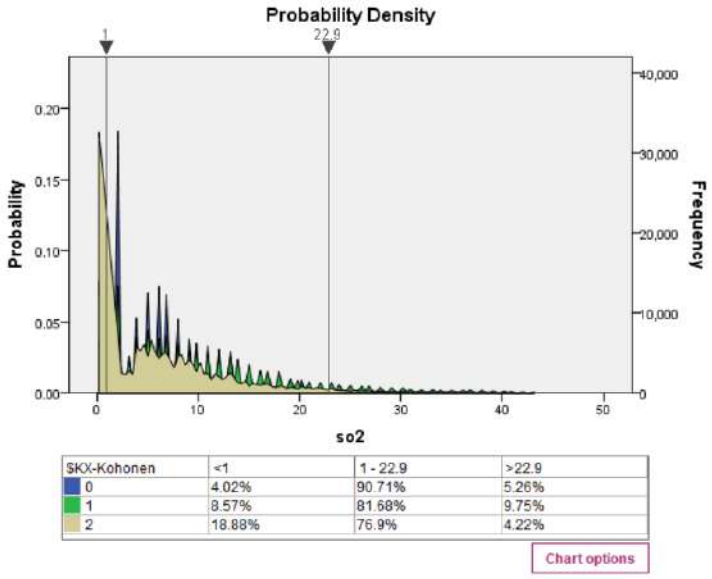
Based on the study of mind patterns and the visualisation of the tests of Algorithms, we got a predictors importance in the relationship with desired resultant target values. After exploratory analysis and conducting a comparison with other models, we use Kohonen and two steps algorithms in the initial stape of analysing the essential pollutant readings with the SPSS modeller, which gives a view dor the generation of pollutant particles and their immersion in the air. For distribution results (Appendix: J)

Interpretation for initial data:

Concerning previously mentioned targets in Kohonen cluster gives High importance is Rural area and Residential area, which shows a Proportion of **So2** over **No2** is up to 5500%. The trade associated with the frequency of collected data related presence of so2and no2 in the air and also the type of affected area most.



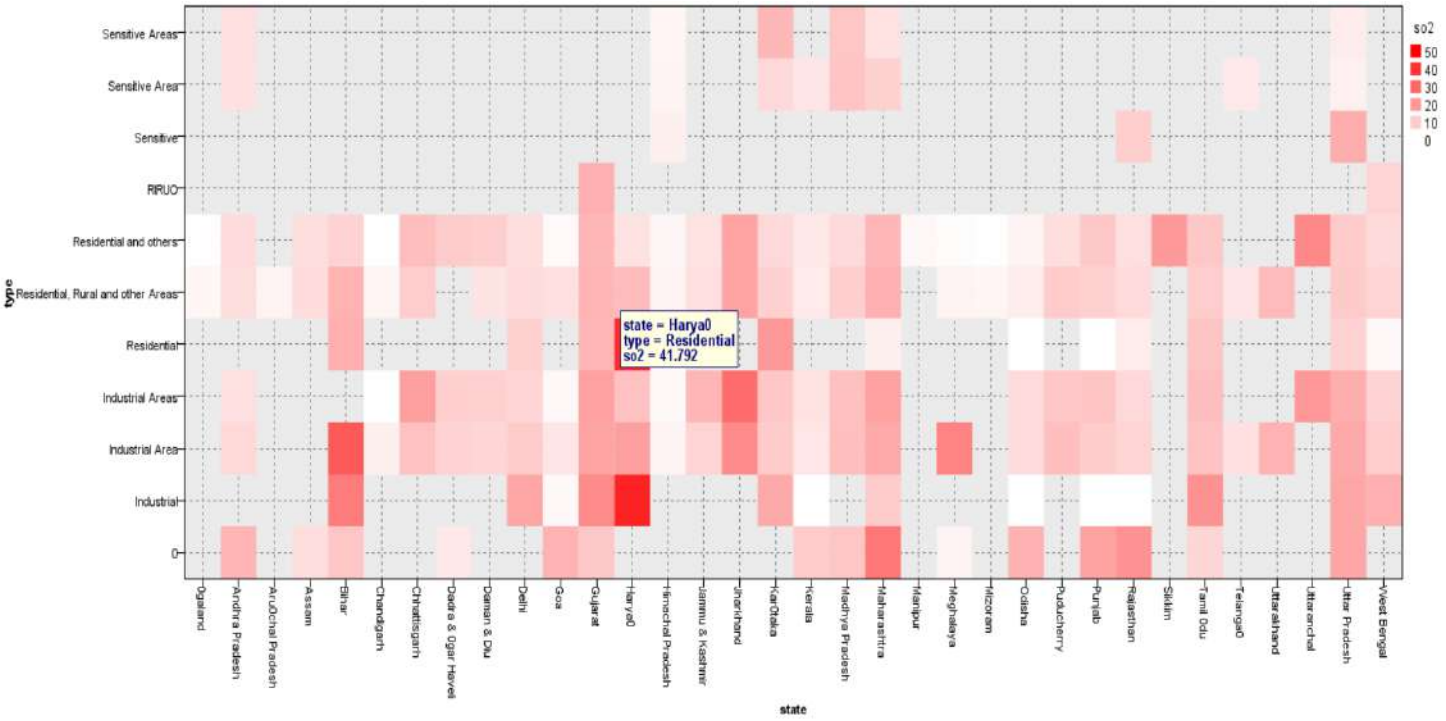




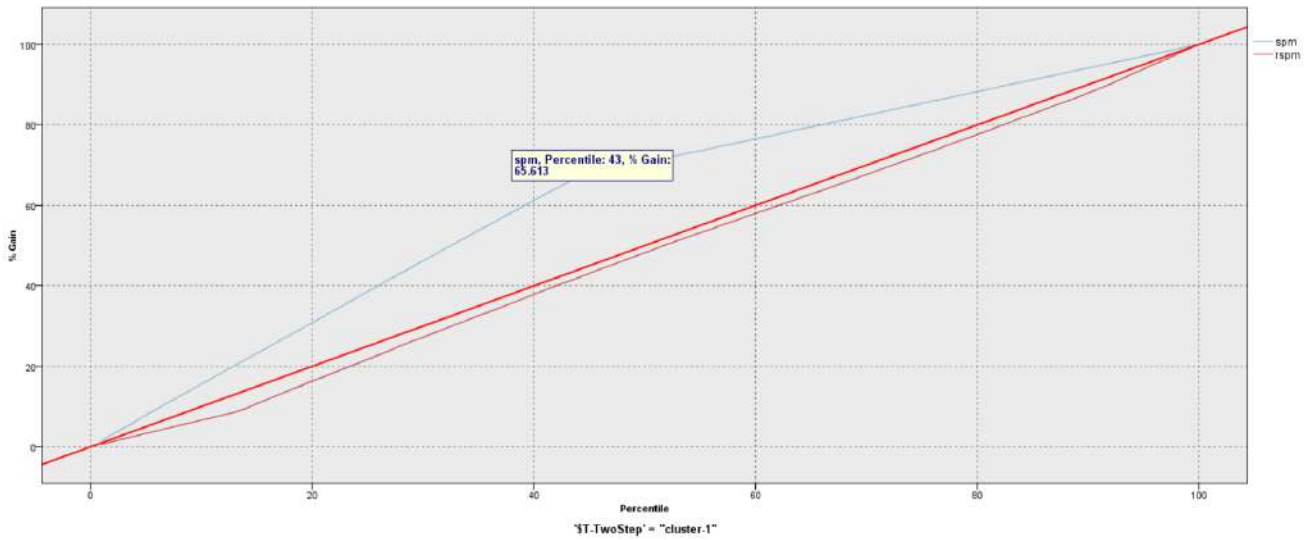
SKX-Kohonen	<1	1 - 22.9	>22.9
0	4.02%	90.71%	5.26%
1	8.57%	81.68%	9.75%
2	18.88%	76.9%	4.22%

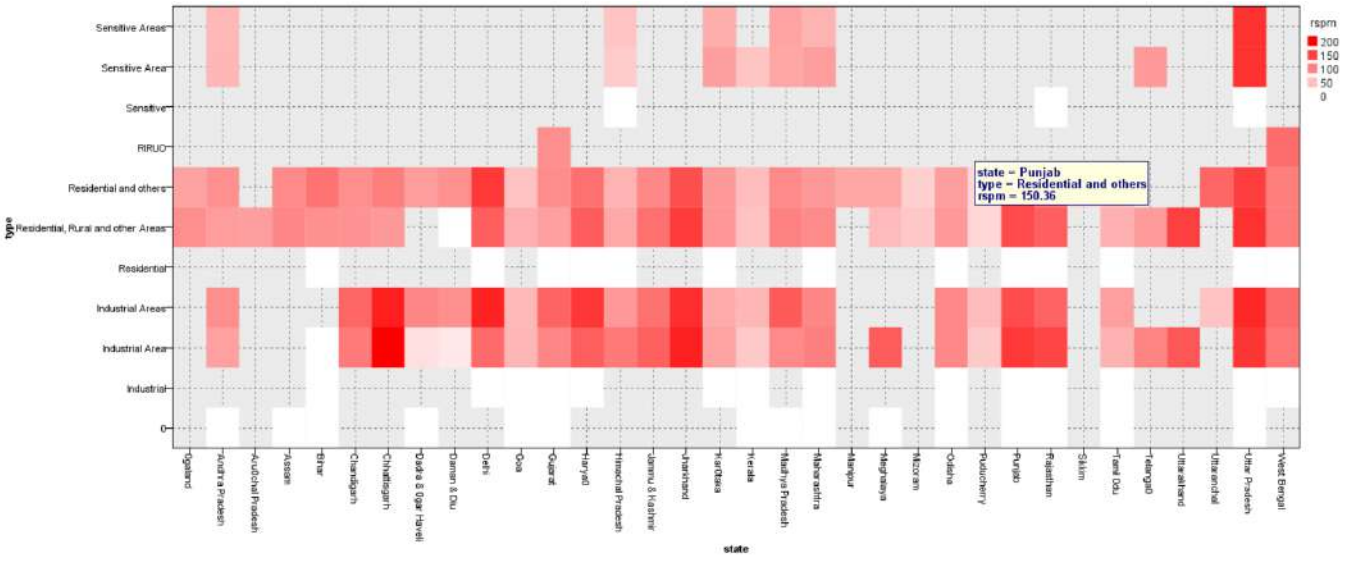
Chart options

As we can see, we have a more dangerous stage of **so2** in states: Hydrabad emission by Industrial areas, and there is more % of **So2** effecting on Reciandecial areas whereas states like Goa, Chattisgarh, Odisha, Punjab are near to 0-10% of **so2** presence which falls into the safe cluster.



The evaluation of result by two stapes algorithm for our target 2 where spm over **rspm** is plotted there is skewness with the distribution of **So2** with point gain of 65,613 over rspm. **rspm** is a liner with the distribution with very small skewness in initial point.





The % of rspm present in air states high-level fo polluted air is present. States like: Delhi, Punjab, Uttar Pradesh, Jharkhand and Chattisgarh are at high risk of pollution caused by rspm op to 200%. Moreover, areas effected are Industrial areas, residential and other areas and sensitive areas.

Interpretation with Index calculation:

With the versioned data with the AQI and the individual Index calculation we analysed regression family we perform Linear regression, random forest, gradient boosting . with the analysis of the model performances the best fit for the study which align with the targeted success rate in our study we choose linear regression. Linear regression is a good fit because It is used to predict the dependent variable from another independent variable.

With targeting AQI count, all individual Indexes are compared in the linear regression also we have calculated multiple error matrices which give us a more unobstructed view in our study. These study of the error matrix will help to analyse workability of the models in context with Data available in all kind of cross-industry projects.

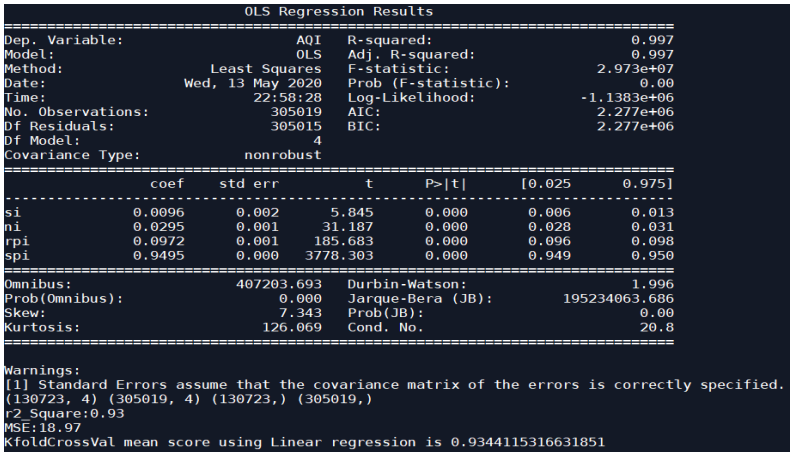
The use of all matrix evaluations listed below:

- **RMSE:** Average y meaning for the x value is expected in the regression section. Please notice that the y values distributed across the sum must also be calculated. We use the root-medium-square error to do this.

- **MSE:** The MSE is a measure of how near the line to the data points. The MSE error is a This is the sum of the difference between the predicted and the actual destination variables over all the data points divided by the number of data points. RMSE is MSE's square root. ("RMS Error", 2020)
- **K-Fold Cross Validation:** the Kfold cross-validation states an accuracy of the model. Cross-Validation is a technique which involves reserving a particular sample of a dataset on which we do not train the model. It estimates the accuracy of the model. ("3.1. Cross-validation: evaluating estimator performance — skit-learn 0.23.0 documentation", 2020)

- **AIC:** The proportional amount of details lost by a given model is determined by AIC: the less knowledge model losses, the better the efficiency of that model. AIC is concerned with the balance between the model's functionality and the model's simplicity.

- **BIC:** IT is estimated the likelihood of the model. A model has the lowest BIC is a good and accurate fit. (2020)



OLS Regression Results

Dep. Variable:	AQI	R-squared:	0.997
Model:	OLS	Adj. R-squared:	0.997
Method:	Least Squares	F-statistic:	2.973e+07
Date:	Wed, 13 May 2020	Prob (F-statistic):	0.00
Time:	22:58:28	Log-Likelihood:	-1.1383e+06
No. Observations:	305019	AIC:	2.277e+06
Df Residuals:	305015	BIC:	2.277e+06
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
si	0.0096	0.002	5.845	0.000	0.006	0.013
ni	0.0295	0.001	31.187	0.000	0.028	0.031
rpi	0.0972	0.001	185.683	0.000	0.096	0.098
spi	0.9495	0.000	3778.303	0.000	0.949	0.950

Omnibus:	407203.693	Durbin-Watson:	1.996
Prob(Omnibus):	0.000	Jarque-Bera (JB):	195234063.686
Skew:	7.343	Prob(JB):	0.00
Kurtosis:	126.069	Cond. No.	20.8

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
(130723, 4) (305019, 4) (130723,) (305019,)
r2_Square:0.93
MSE:18.97
KfoldCrossVal mean score using Linear regression is 0.9344115316631851

- **K-fold cross value mean score is 0.98, which shows 98% accuracy in model**.
- **Mean squared error MSE**: It is +ve 18.97 the ideal **MSE** is not 0 since then model would have a model that correctly predicts the data.
- **R-squared value** is 0.9, so it is indicated that around 100% attribute values indicate that the model explains all the variability of the response data around its mean.
- **AIC value (Akaike's Information Criteria):** is recorded2.2 which is constant and implies a high likelihood approximation of a constant plus the relative variance between the uncertain true probability function of the data and the model's corresponding probability function.
- **BIC value**:2.2 which is lower which does not affect likelihood value which indicate complexity in the model as BIC increases likelihood decreases and if BIC is low than likelihood will increase which is right fit Indication for the model.
- **Log_Likelihood:** -1.1 It is the sample size feature also the match of different coefficients may be contrasted. We want to maximize the likelihood. So higher likelihood is better Ex. -2 is better than -3.
- **Durbin-Watson**: It is a statistical test value for autocorrelation in the residuals in regression analysis. The value between 0 to 2 indicates +ve autocorrelation and value in between 2 to 4 is an -ve autocorrelation. Furthermore, we have scored here 1.99, which is +ve correlation.
- **The recrdurial stderr:** for all elements/attributes used in this model have 0 stderr, which indicate that the model fits the data perfectly likely due to overfitting.

The Conclusion

With the analysis and the overall study, the distribution of **So2** and **No2** is High in residential areas, whereas **spm** and **rspm** are in sensitive areas and industrial areas. Also, the correlation of AQI and the other particulate matters is effective is proved with the regression.

If we analyze the trend of the pattern of AQI detection with the Timestamp, the AQI is hitting to the unhealthy threshold in overall calculation across India in Past few years. We can see that only in the year 2000 year quality is moderate. In the year 1992,1997,2002 and 20003 air is unhealthy with low-risk factors. From 20010-15 air quality is hit continuously very unhealthy threshold. All risks with the % are mentioned in the given table associated with AQI.

The rate of RSPM, which are respirable hazardous particles is higher in the middle east and Northern side of India. The states like Maharashtra, Uttar Pradesh are profoundly affected. The cities count of AQI Delhi is at the highest risk stage. (Appendix: K)

There are some visualizations are present to support results in Appendix: K.

The above research concluded that it is evident that the following steps must be taken to deter reduce pollutant levels and improve quality index while predicting the trend of our results.

- Reduce the count of rspm in the air where the sensitive and industrial areas have a greater count of **rspm** as per our prediction. There is a need to generate a solution for reducing **rspm** in the air.
- There is a need to take immediate action in the areas which have recorded High threshold if bad AQI from the last few years.
- There is an Innovation gap generated, and it gives scope to generate new business ideas and ventures in the field of technology. With the conclusion of this study, there is scope for developing a technology-based solution to reduce pollutant count in the air to secure High-risk zones in the future.

Based on the software used in these projects, the best fit is the use of the Python stack because it allows to manipulate and restructure our data as needed by using user-defined functions or NumPy functions with lambda.

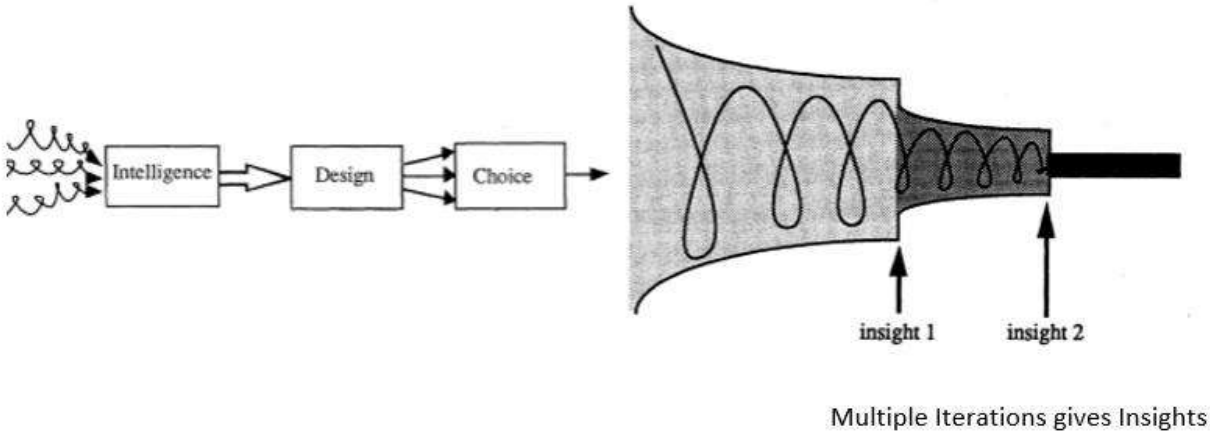
Discussion

From this kind of study which is carried out and practised in this project gives a scientific evaluation which should be observed an essential role to make strategies and guidelines and give exposure to policymakers and also for entrepreneurs and researchers to establish strategies based on studies like this.

Although the analysis of the main variables with pollutants behaviours is found through descriptive data mining, the results are generally not added to the experts as all of the discovery does not add new pieces of knowledge and factors to the previously knowledgeable study.

The study in this paper supports the practical application of Iterative decision-making model as the study is carried out with multiple iterative processes, which gives intelligence density from three-loop learning.

If we see the business perspective view with the current study, it follows the model for "Iterative Organizational Decision Making." (Langley, Mintzberg, Pitcher, Posada & Saint-Macary, 1995)



(Langley, Mintzberg, Pitcher, Posada & Saint-Macary, 1995)

Limitations

The current study drives to feature important variables which were responsible for Air Pollution. The scope of that dataset is for particular one country. So the study is limited for Inda.

The absence of value readings of pollutants generates ambiguity in the results. This can be seen as both ways as a limitation or the opportunity. By which we can fill up those gaps with mathematically using equations or the random values which can be the future-oriented dataset and which will give a future based predictable results and insights.

Also, within this project, we face some issues when we use different technology stack for analysis. All software have thair own workings and functionality. Some of them use stationary data only for analysis which can not be manipulated or restructured by using the software. Ex. While using SPSS, which mainly uses stationary and homogeneous data, it is not possible to do code for lambda functions to calculate Index values from regular readings.

Future Research

The analysis and study with this research paper can be further carried out with the availability of real-time data. With the business perspective view and innovation-based analysis, which needed to create an enterprise based on air quality improvement based products, the research can be done with the methodology mentioned in this paper.

This analysis and study are also further useful for entrepreneurs who are looking for sustainable and futuristic product development in the field of Environment Science and climate changes.

Ex: The Innovation based product can be launched with real-time data analysis in future with this methodology. The product which can improve air quality and which follows SMACT framework based enterprise or venture can be developed.

References

3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.23.0 documentation. (2020). Retrieved 15 May 2020, from https://scikit-learn.org/stable/modules/cross_validation.html

Air Pollution on the Move in India. (2020). Retrieved 11 March 2020, from https://earthobservatory.nasa.gov/images/92347/air-pollution-on-the-move-in-india

AIR QUALITY FROM SPACE | Air Quality. (2020). Retrieved 11 March 2020, from https://airquality.gsfc.nasa.gov/

Air Quality Index (AQI) Basics. (2020). Retrieved 15 May 2020, from https://cfpub.epa.gov/airnow/index.cfm?action=aqibasics.aqi

CLEAN AIR FOR HEALTH: Geneva Action Agenda. (2020). Retrieved 15 June 2020, from https://www.who.int/phe/news/clean-air-for-health/en/

CLEAN AIR FOR HEALTH: Geneva Action Agenda. (2020). Retrieved 15 June 2020, from https://www.who.int/phe/news/clean-air-for-health/en/

CRISP-DM Process (SPSS, 2007)"ibm spss modeler crisp-dm guide pdf - Google Search", *Google.com*, 2020. [Online]. Available: https://www.google.com/search?safe=active&sxsrf=ALeKk003urnQgoa5Vtl1Udho14Jnu7Z9qA%3A1592372474951&ei=-qzpXsu8OdGA9QOwyanoAw&q=ibm+spss+modeler+crisp-dm+guide+pdf&oq=IBM+SPSS+CIRSPDM&gs_lcp=CgZwc3ktYWIQARgCMgYIABANEB4yCAgAEAgQDRAeMggIABAIEA0QHjIICAAQCBANEB4yCAgAEAgQDRAeMggIABAIEA0QHjoHCAAQRxCwAzoECCMQJ1CakwJY2KYCYOXOAmgCcAB4AIABe4gB1AKSAQMxLjKYAQCgAQGqAQdnd3Mtd2l6&sclient=psy-ab. [Accessed: 17- Jun- 2020].

Deshpande, A., & Ghude, S. (2020). Scientific Evaluation of Air Quality Standards and Defining Air Quality Index for India. *Ndian Institute Of Tropical Meteorology, Pune, India Ministry Of Earth Sciences, Govt. Of India*, *RR-127*(IISN 0252-1075). doi: August 2010

Fayyad, U., & Uthurusamy, R. (1996). Data mining and knowledge discovery in databases. *Communications Of The ACM*, *39*(11), 24-26. doi: 10.1145/240455.240463.

Langley, A., Mintzberg, H., Pitcher, P., Posada, E., & Saint-Macary, J. (1995). Opening up Decision Making: The View from the Black Stool. *Organization Science*, *6*(3), 260-279. doi: 10.1287/orsc.6.3.260

NAAQS Table | US EPA. (2020). Retrieved 16 June 2020, from https://www.epa.gov/criteria-air-pollutants/naaqs-table

NATIONAL AIR QUALITY INDEX. (2014). Retrieved 16 June 2020, from http://www.indiaenvironmentportal.org.in/files/file/Air%20Quality%20Index.pdf.

Peko, G., Danenberg, J., & Sundaram, D. (2013). Adaptive Sustainable Enterprises: Interweaving the Deliberate and Emergent. *The International Journal Of Sustainability Policy And Practice*, *8*(4), 15-27. doi: 10.18848/2325-1166/cgp/v08i04/55401

Public health foundation India , center for environmental health. (2020). *Air Pollution and Health in India* (pp. 14-16). Retrieved from https://www.ceh.org.in/wp-content/uploads/2017/10/Air-Pollution-and-Health-in-Indi

RMS Error. (2020). Retrieved 15 May 2020, from http://statweb.stanford.edu/~susan/courses/s60/split/node60.html

SDGs .:. Sustainable Development Knowledge Platform. (2020). Retrieved 16 June 2020, from https://sustainabledevelopment.un.org/sdgs

WHO methods and data sources for global burden of disease estimates. (2020). *Global Health Estimates Technical Paper WHO/HIS/IER/GHE/2017.1*, *GHE/2017.1*, 15,31. doi: January 2017.

WHO. (2018). *First WHO Global Conference on Air Pollution and Health* [Video]. https://www.who.int/airpollution/events/conference/en/.

APPENDIX: A

	INDIA	
	% of total DALYs	Risk factor attribution (%)
Ischemic heart disease	7.28	38.61
Lower respiratory infections	4.48	62
Chronic obstructive pulmonary disease	4.63	55.74
Cerebrovascular disease	3.62	34.16
Tracheal, bronchus and lung cancer	0.4	46.29

APPENDIX: B

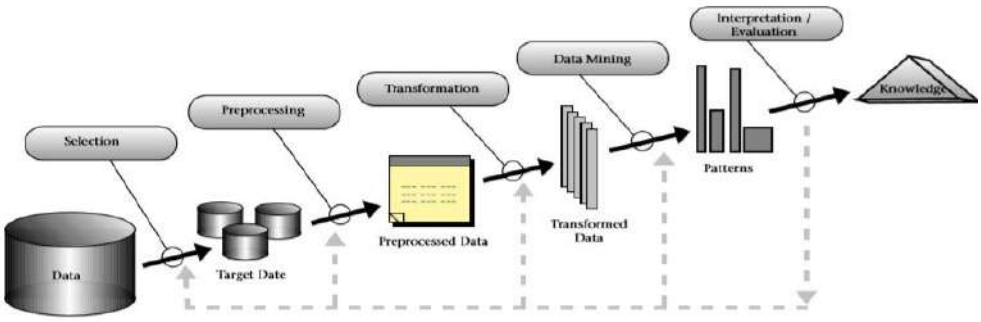
An Hourly calculation for pollutant contribution to AQI given by Central Pollution Control Board of India.

AQI Category (Range)	PM_{10} 24-hr	$PM_{2.5}$ 24-hr	NO_2 24-hr	O_3 8-hr	CO 8-hr (mg/m^3)	SO_2 24-hr	NH_3 24-hr	Pb 24-hr
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5 –1.0
Moderately polluted (101-200)	101-250	61-90	81-180	101-168	2.1- 10	81-380	401-800	1.1-2.0
Poor (201-300)	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748*	17-34	801-1600	1200-1800	3.1-3.5
Severe (401-500)	430 +	250+	400+	748+*	34+	1600+	1800+	3.5+

("NATIONAL AIR QUALITY INDEX", 2014)

Numerical Value	Color	Air Quality Index Levels of Health Concern	Meaning
0 - 50	Green	Good	Air quality is considered satisfactory, and air pollution poses little or no risk.
51 - 100	Yellow	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
101 - 150	Orange	Unhealty for sensitive groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
151 - 200	Red	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
201 -300	Purple	Very unhealthy	Health alert: everyone may experience more serious health effects.
301 - 500	Maroon	Hazardous	Health warnings of emergency conditions. The entire population is more likely to be affected.

APPENDIX:C



(Fayyad & Uthurusamy, 1996)

APPENDIX:D

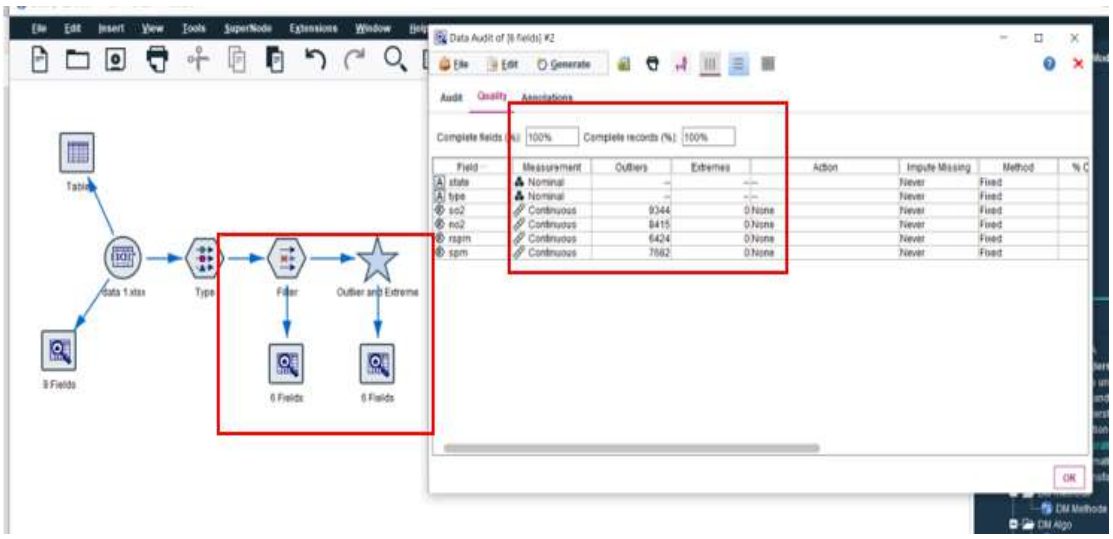
- Removing Extreme and outlier values with pandas Python.

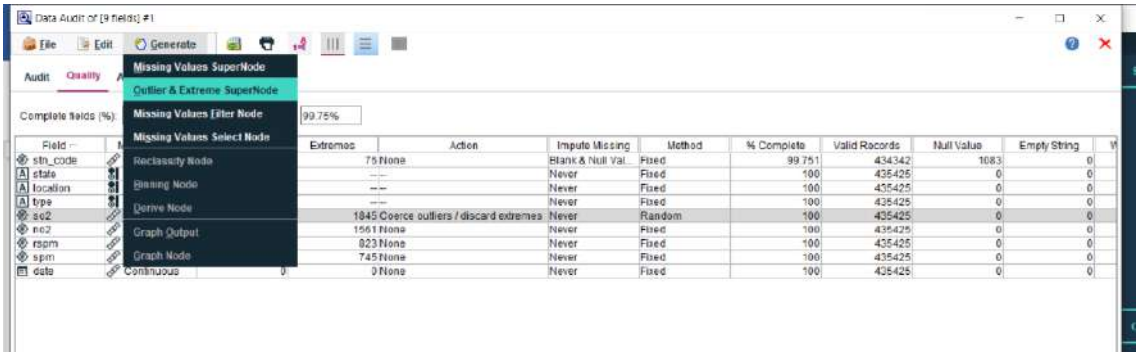
```
def remove_outlier(df_in, col_name):
    a1 = df_in[col_name].quantile(0.25)
    a2 = df_in[col_name].quantile(0.75)
    iqr = a2-a1 #Interquartile range
    fence_low  = a1-1.5*iqr
    fence_high = a2+1.5*iqr
    df_out = df_in.loc[(df_in[col_name] > fence_low) & (df_in[col_name] < fence_high)]
    return df_out
remove_outlier(air_data,'so2')
remove_outlier(air_data,'no2')
remove_outlier(air_data,'rspm')
remove_outlier(air_data,'spm')
print(remove_outlier)
print("after removel of Outlaird mean of attributes group")
print(air_data.groupby('state')[['spm','pm2_5','rspm','so2','no2']].mean())
print("\n")
```

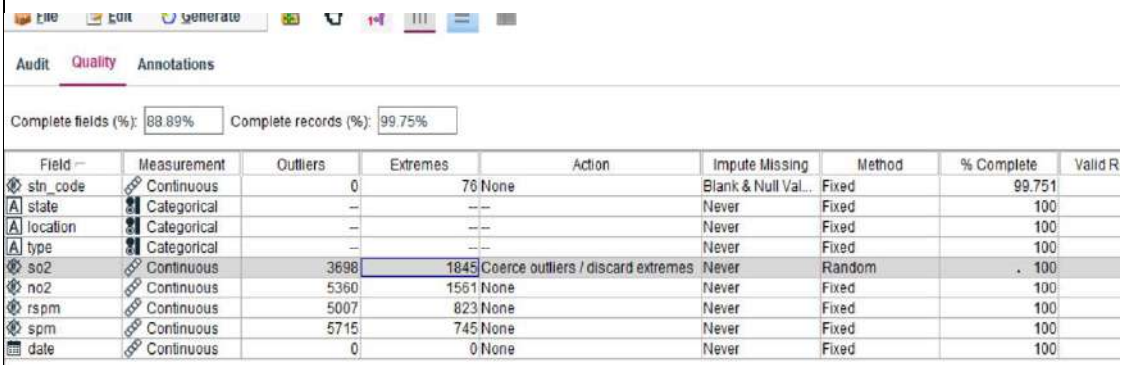
```
#Removing outlaired values
<function remove_outlier at 0x0000017A767E3840>
after removel of Outlaird mean of attributes group
                                spm      pm2_5        rspm       so2  \
state
Andhra Pradesh           200.260378        NaN   78.182824  7.284845
Arunachal Pradesh               NaN        NaN   76.629213  3.179104
Assam                    153.355386        NaN   93.724912  6.723263
Bihar                    276.917416        NaN  123.705176 10.381476
Chandigarh               206.056150        NaN   96.587079  2.676986
Chhattisgarh             231.290969        NaN  126.472390 12.846660
Dadra & Nagar Haveli     170.545024  30.511628   76.536530  8.939587
Daman & Diu              145.681416  27.886364   73.749431  8.192958
Delhi                    399.402088  95.113208  196.639771  8.737273
Goa                       67.254193  18.855612   61.212766  6.827913
Gujarat                  191.567930  30.729696   98.244510 16.656343
Haryana                  268.264804        NaN  149.868537 14.864957
Himachal Pradesh         208.575630        NaN   91.870202  2.667013
Jammu & Kashmir          196.221053        NaN  117.449483  7.180521
Jharkhand                277.948746        NaN  168.517763 23.485794
Karnataka                168.001743        NaN   79.371801 10.223099
Kerala                    84.419791        NaN   50.636064  5.322350
Lakshadweep                     NaN        NaN         NaN       NaN
Madhya Pradesh           210.067545  65.064565  114.717967 11.587410
Maharashtra              205.255823        NaN  101.479608 17.366863
```

```
                            no2
state
Andhra Pradesh        21.704451
Arunachal Pradesh      5.469697
Assam                 14.793691
Bihar                 36.575525
Chandigarh            18.619404
Chhattisgarh          24.815961
Dadra & Nagar Haveli  18.293959
Daman & Diu           16.168926
Delhi                 53.489147
Goa                   12.506337
Gujarat               24.065631
Haryana               23.428311
Himachal Pradesh      13.658688
Jammu & Kashmir       12.213181
Jharkhand             43.366341
Karnataka             22.702837
Kerala                14.421889
Lakshadweep                 NaN
```

- Dealing with Extreme and outlier values with IBM SPSS.


File Edit Insert View Tools SuperNode Extensions Window Help
Data Audit of [8 fields] #2
File Edit Generate
Audit Quality Annotations
Complete fields (%): 100%
Complete records (%): 100%
Field Measurement Outliers Extremes Action Impute Missing Method
state Nominal -- -- Never Fixed
type Nominal -- -- Never Fixed
so2 Continuous 9344 0 None Never Fixed
no2 Continuous 8415 0 None Never Fixed
rspm Continuous 6424 0 None Never Fixed
spm Continuous 7082 0 None Never Fixed
OK
Table
data 1.xlsx
Type
Filter
Outlier and Extreme
8 Fields
6 Fields
6 Fields
DM Methods
DM Algo


Data Audit of [8 fields] #1
File Edit Generate
Audit Quality
Missing Values SuperNode
Outlier & Extreme SuperNode
Missing Values Filter Node
Missing Values Select Node
Reclassify Node
Binning Node
Derive Node
Graph Output
Graph Node
Complete fields (%): 99.75%
Field Extremes Action Impute Missing Method % Complete Valid Records Null Value Empty String
stn_code 76 None Blank & Null Val... Fixed 99.751 434342 1083 0
state -- -- Never Fixed 100 435425 0 0
location -- -- Never Fixed 100 435425 0 0
type -- -- Never Fixed 100 435425 0 0
so2 1845 Coerce outliers / discard extremes Never Random 100 435425 0 0
no2 1561 None Never Fixed 100 435425 0 0
rspm 823 None Never Fixed 100 435425 0 0
spm 745 None Never Fixed 100 435425 0 0
date Continuous 0 0 None Never Fixed 100 435425 0 0


File Edit Generate
Audit Quality Annotations
Complete fields (%): 88.89%
Complete records (%): 99.75%
Field Measurement Outliers Extremes Action Impute Missing Method % Complete Valid R
stn_code Continuous 0 76 None Blank & Null Val... Fixed 99.751
state Categorical -- -- Never Fixed 100
location Categorical -- -- Never Fixed 100
type Categorical -- -- Never Fixed 100
so2 Continuous 3698 1845 Coerce outliers / discard extremes Never Random 100
no2 Continuous 5360 1561 None Never Fixed 100
rspm Continuous 5007 823 None Never Fixed 100
spm Continuous 5715 745 None Never Fixed 100
date Continuous 0 0 None Never Fixed 100

Appendix: E

Cleaning and Identifying Null values by using inbuilt functions and mathematical coding.

```
# null values present in data
print("# null values present in data")
print(air_data.isnull().sum())
print("\n")
```

```
# null values present in data
stn_code                       144077
sampling_date                       3
state                               0
location                            3
agency                         149481
type                             5393
so2                             34646
no2                             16233
rspm                            40222
spm                            237387
location_monitoring_station     27491
pm2_5                          426428
date                                7
dtype: int64
```

```
#% of nullvalues in data
print("#% of nullvalues in data")
print(air_data.isnull().sum()/len(air_data)*100)
print("\n")
```

```
#% of nullvalues in data
stn_code                       33.064749
sampling_date                   0.000688
state                           0.000000
location                        0.000688
agency                         34.304933
type                            1.237659
so2                             7.951035
no2                             3.725370
rspm                            9.230692
spm                            54.478797
location_monitoring_station     6.309009
pm2_5                          97.862497
date                            0.001606
dtype: float64
```

Dealing with Null values by filling or dropping functions.

```
print("\n filling missing values")
by_State=air_data.groupby('state')

def impute_mean(series):
    return series.fillna(series.mean())

air_data['rspm']=by_State['rspm'].transform(impute_mean)
air_data['so2']=by_State['so2'].transform(impute_mean)
air_data['no2']=by_State['no2'].transform(impute_mean)
air_data['spm']=by_State['spm'].transform(impute_mean)
air_data['pm2_5']=by_State['pm2_5'].transform(impute_mean)
#imputer = Imputer(missing_values = 'NaN', strategy = 'median', axis = 0)
#imputer = imputer.fit(air_data.groupby('state').iloc[:, 3:8].values)
#air_data.iloc[:,3:8] = imputer.transform(air_data.iloc[:, 3:8].values)
air_data.info()

for col in air_data.columns.values:
    if air_data[col].isnull().sum() == 0:
        continue
    if col == 'date':
        guess_values = air_data.groupby('state')['date'].apply(lambda x: x.mode().max())
    elif col=='type':
        guess_values = air_data.groupby('state')['type'].apply(lambda x: x.mode().max())
    elif col=='location':
        guess_values = air_data.groupby('state')['location'].apply(lambda x: x.mode().max())
    else:
        guess_values = air_data.groupby('state')['pm2_5'].apply(lambda x: x.mode().max())

print(air_data.head())
print("\n new % of null values")
print(air_data.isnull().sum()/len(air_data)*100)
#print(air_data.isnull())
#Fixing the missing values firstly for all the pollutants.
#We will consider taking mean for all the pollutants columns and make use of the Imputer class
from sklearn.preprocessing import Imputer
imputer = Imputer(missing_values = 'NaN', strategy = 'mean', axis = 0)
imputer = imputer.fit(air_data.iloc[:, 3:8].values)
air_data.iloc[:,3:8] = imputer.transform(air_data.iloc[:, 3:8].values)
#air_data.info()
#air_data.head()
#Fixing the missing values in the column 'type'
```

```
#We will consider taking mean for all the pollutants columns and make use of the Imputer class
from sklearn.preprocessing import Imputer
imputer = Imputer(missing_values = 'NaN', strategy = 'mean', axis = 0)
imputer = imputer.fit(air_data.iloc[:, 3:8].values)
air_data.iloc[:,3:8] = imputer.transform(air_data.iloc[:, 3:8].values)
#air_data.info()
#air_data.head()
#Fixing the missing values in the column 'type'
air_data['pm2_5'].describe()
#With 10 Unique labels, we will fill the null values by the most common type, which is 'Residential, Ru
common_value='Residential,Rural and other Areas'
air_data['pm2_5']=air_data['pm2_5'].fillna(common_value)
#air_data.info()
print(air_data.isnull().sum()/len(air_data)*100)
#Fixing the missing values in the column 'type'
air_data['type'].describe()
#With 10 Unique labels, we will fill the null values by the most common type, which is 'Residential, Ru
common_value='Residential,Rural and other Areas'
air_data['type']=air_data['type'].fillna(common_value)
air_data.info()
print("\n new % of null values")
print(air_data.isnull().sum()/len(air_data)*100)
```

No Null values present in the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 435742 entries, 0 to 435741
Data columns (total 9 columns):
state       435742 non-null object
location    435739 non-null object
type        435742 non-null object
so2         435742 non-null float64
no2         435742 non-null float64
rspm        435742 non-null float64
spm         435742 non-null float64
pm2_5       435742 non-null float64
date        435735 non-null object
dtypes: float64(5), object(4)
memory usage: 29.9+ MB

 new % of null values
state       0.000000
location    0.000688
type        0.000000
so2         0.000000
no2         0.000000
rspm        0.000000
spm         0.000000
pm2_5       0.000000
date        0.001606
dtype: float64

In [3]:
```

APPENDIX: F

Calculation of AQI with universal standards and lambda mathematical equations.

```
# Individual QI calculation
def calculate_si(so2):
    si=0
    if (so2<=40):
     si= so2*(50/40)
    elif (so2>40 and so2<=80):
     si= 50+(so2-40)*(50/40)
    elif (so2>80 and so2<=380):
     si= 100+(so2-80)*(100/300)
    elif (so2>380 and so2<=800):
     si= 200+(so2-380)*(100/420)
    elif (so2>800 and so2<=1600):
     si= 300+(so2-800)*(100/800)
    elif (so2>1600):
     si= 400+(so2-1600)*(100/800)
    return si
air_data['si']=air_data['so2'].apply(calculate_si)
df= air_data[['so2','si']]
print(df.head())
```

```
 # Individual QI calculation So2
   so2     si
0  4.8  6.000
1  3.1  3.875
2  6.2  7.750
3  6.3  7.875
4  4.7  5.875
```

```
# rspm individual pollutant index rpi
def calculate_(rspm):
    rpi=0
    if(rpi<=30):
     rpi=rpi*50/30
    elif(rpi>30 and rpi<=60):
     rpi=.+(rpi-30)*50/30
    elif(rpi>60 and rpi<=90):
     rpi=100+(rpi-60)*100/30
    elif(rpi>90 and rpi<=120):
     rpi=200+(rpi-90)*100/30
    elif(rpi>120 and rpi<=250):
     rpi=300+(rpi-120)*(100/130)
    else:
     rpi=400+(rpi-250)*(100/130)
    return rpi
air_data['rpi']=air_data['rspm'].apply(calculate_si)
df= air_data[['rspm','rpi']]
print(df.head())
```

```
 # Individual QI calculation no2
    no2      ni
0  17.4  21.750
1   7.0   8.750
2  28.5  35.625
3  14.7  18.375
4   7.5   9.375
```

```python
# no2 individual pollutant index ni
def calculate_ni(no2):
    ni=0
    if(no2<=40):
     ni= no2*50/40
    elif(no2>40 and no2<=80):
     ni= 50+(no2-40)*(50/40)
    elif(no2>80 and no2<=180):
     ni= 100+(no2-80)*(100/100)
    elif(no2>180 and no2<=280):
     ni= 200+(no2-180)*(100/100)
    elif(no2>280 and no2<=400):
     ni= 300+(no2-280)*(100/120)
    else:
     ni= 400+(no2-400)*(100/120)
    return ni
air_data['ni']=air_data['no2'].apply(calculate_ni)
df= air_data[['no2','ni']]
print(df.head())
```

```
 # Individual QI calculation rspm
       rspm       rpi
0  78.182824  97.72853
1  78.182824  97.72853
2  78.182824  97.72853
3  78.182824  97.72853
4  78.182824  97.72853
```

```python
# spm individual pollutant index spi
def calculate_spi(spm):
    spi=0
    if(spm<=50):
     spi=spm*50/50
    elif(spm>50 and spm<=100):
     spi=50+(spm-50)*(50/50)
    elif(spm>100 and spm<=250):
     spi= 100+(spm-100)*(100/150)
    elif(spm>250 and spm<=350):
     spi=200+(spm-250)*(100/100)
    elif(spm>350 and spm<=430):
     spi=300+(spm-350)*(100/80)
    else:
     spi=400+(spm-430)*(100/430)
    return spi
air_data['spi']=air_data['spm'].apply(calculate_spi)
df= air_data[['spm','spi']]
print(df.head())
```

```
 # Individual QI calculation Spm
          spm         spi
0  200.260378  166.840252
1  200.260378  166.840252
2  200.260378  166.840252
3  200.260378  166.840252
4  200.260378  166.840252
```

```python
# pm2_5 individual pollutant index pmi
def calculate_pmi(pm2_5):
    pmi=0
    if(pm2_5<=50):
     pmi=pm2_5*(50/50)
    elif(pm2_5>50 and pm2_5<=100):
     pmi=50+(pm2_5-50)*(50/50)
    elif(pm2_5>100 and pm2_5<=250):
     pmi= 100+(pm2_5-100)*(100/150)
    elif(pm2_5>250 and pm2_5<=350):
     pmi=200+(pm2_5-250)*(100/100)
    elif(pm2_5>350 and pm2_5<=450):
     pmi=300+(pm2_5-350)*(100/100)
    else:
     pmi=400+(pm2_5-430)*(100/80)
    return pmi
air_data['pmi']=air_data['pm2_5'].apply(calculate_pmi)
df= air_data[['pm2_5','pmi']]
print(df.tail())
```

```
 # Individual QI calculation pm
           pm2_5        pmi
435737  64.890625  64.890625
435738  64.890625  64.890625
435739  48.317274  48.317274
435740  48.317274  48.317274
435741  48.317274  48.317274
```

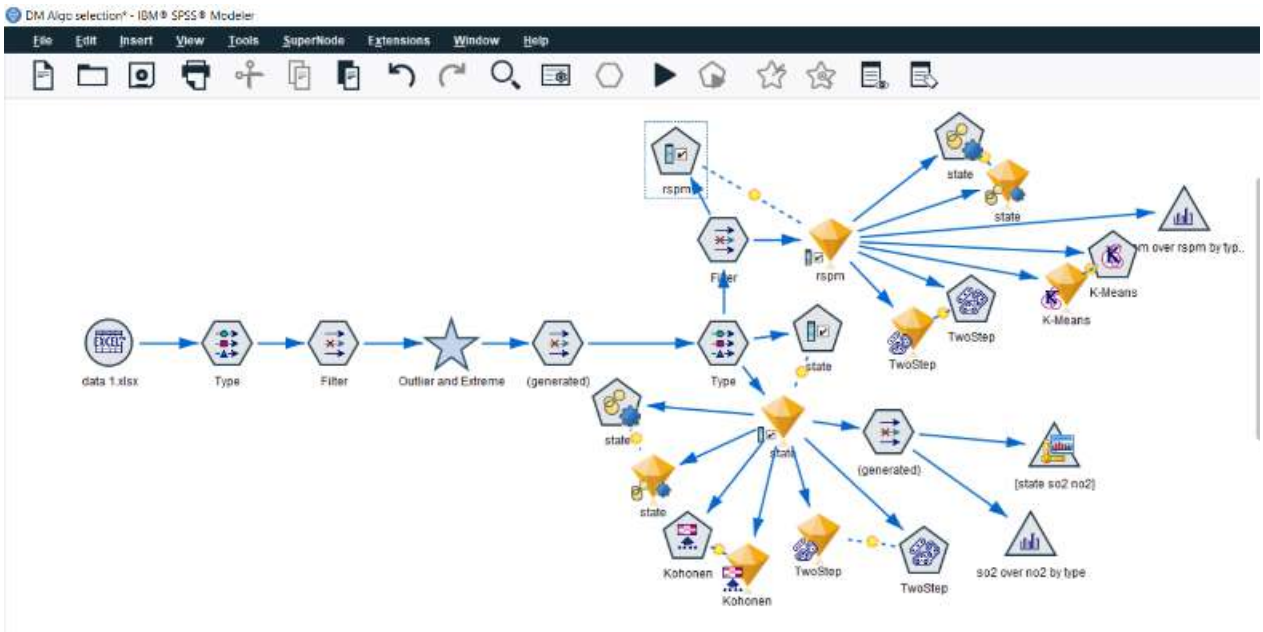
```
# the air quality index AQI of every data value
#its is calculated as per indian govt standards

def calculate_aqi(si,ni,spi,rpi):
    aqi=0
    if(si>ni and si>spi and si>rpi):
     aqi=si
    if(spi>si and spi>ni and spi>rpi):
     aqi=spi
    if(ni>si and ni>spi and ni>rpi):
     aqi=ni
    if(rpi>si and rpi>ni and rpi>spi):
     aqi=rpi
    return aqi
air_data['AQI']=air_data.apply(lambda x:calculate_aqi(x['si'],x['ni'],x['spi'],x['rpi']),axis=1)
df= air_data[['state','si','ni','rpi','spi','AQI']]
print(df.head())
```

```
 # AQI calculation
            state     si      ni       rpi        spi        AQI
0  Andhra Pradesh  6.000  21.750  97.72853  166.840252  166.840252
1  Andhra Pradesh  3.875   8.750  97.72853  166.840252  166.840252
2  Andhra Pradesh  7.750  35.625  97.72853  166.840252  166.840252
3  Andhra Pradesh  7.875  18.375  97.72853  166.840252  166.840252
4  Andhra Pradesh  5.875   9.375  97.72853  166.840252  166.840252
```

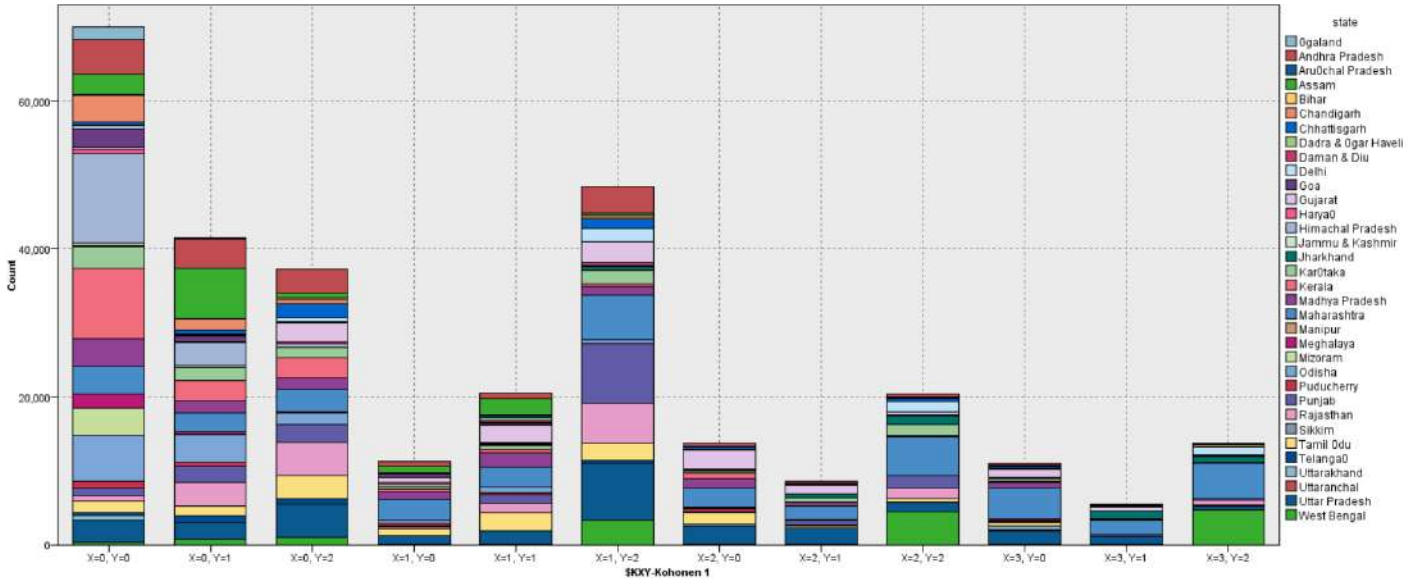
Appendix: G

Data mining with IBM SPSS modeller.

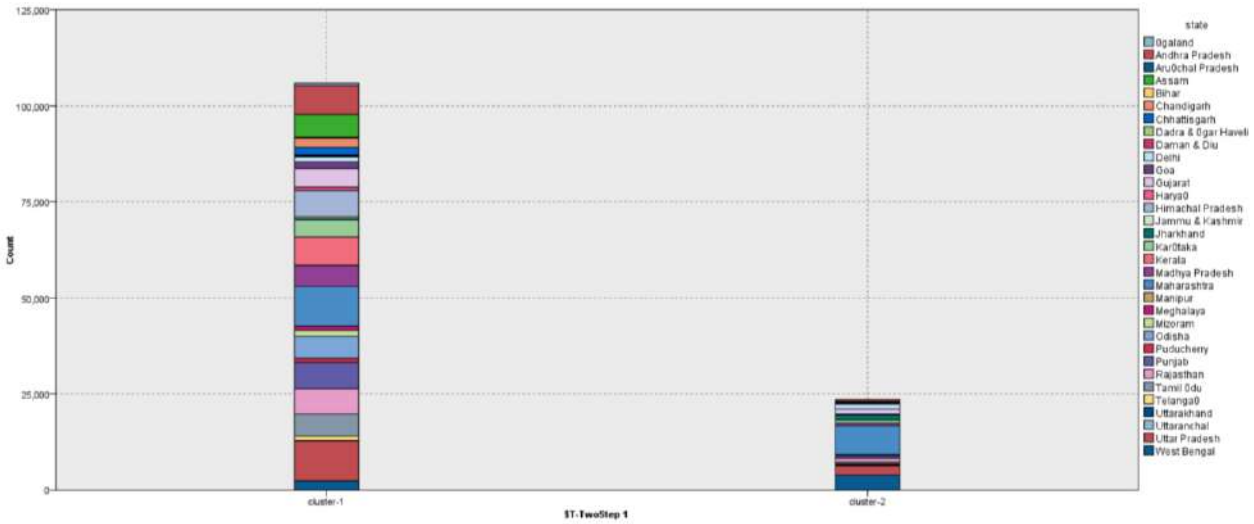


Appendix: H

The distribution of data in Kohonen model with 2 clusters and counts of distribution.

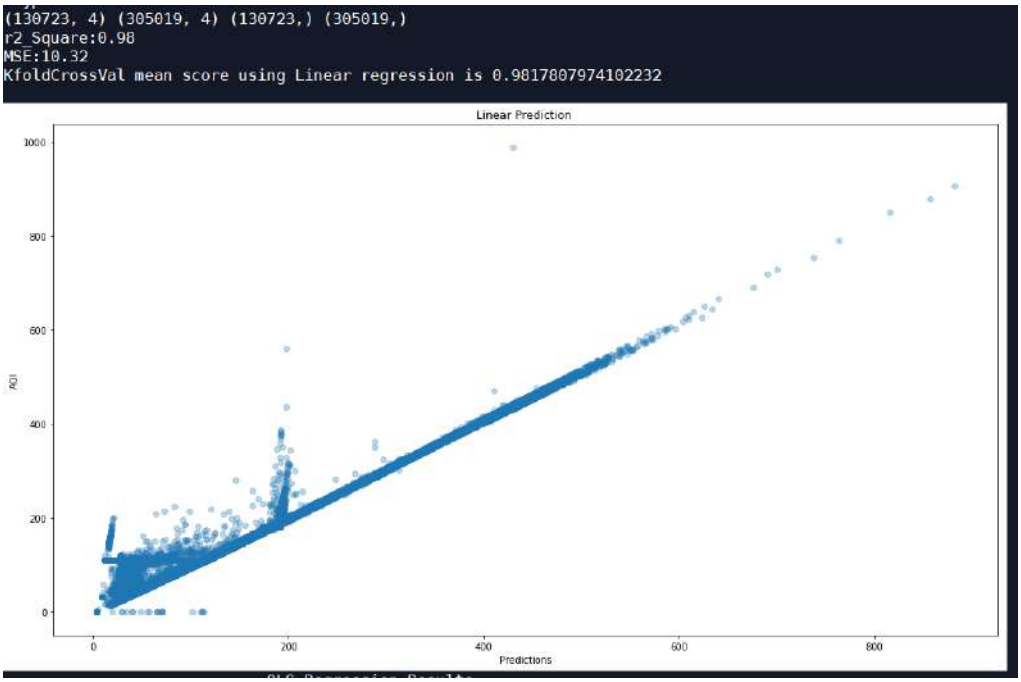


Distribution of States with a count of SO2 and No2 data in Two steps model with 2 clusters and counts of distribution.

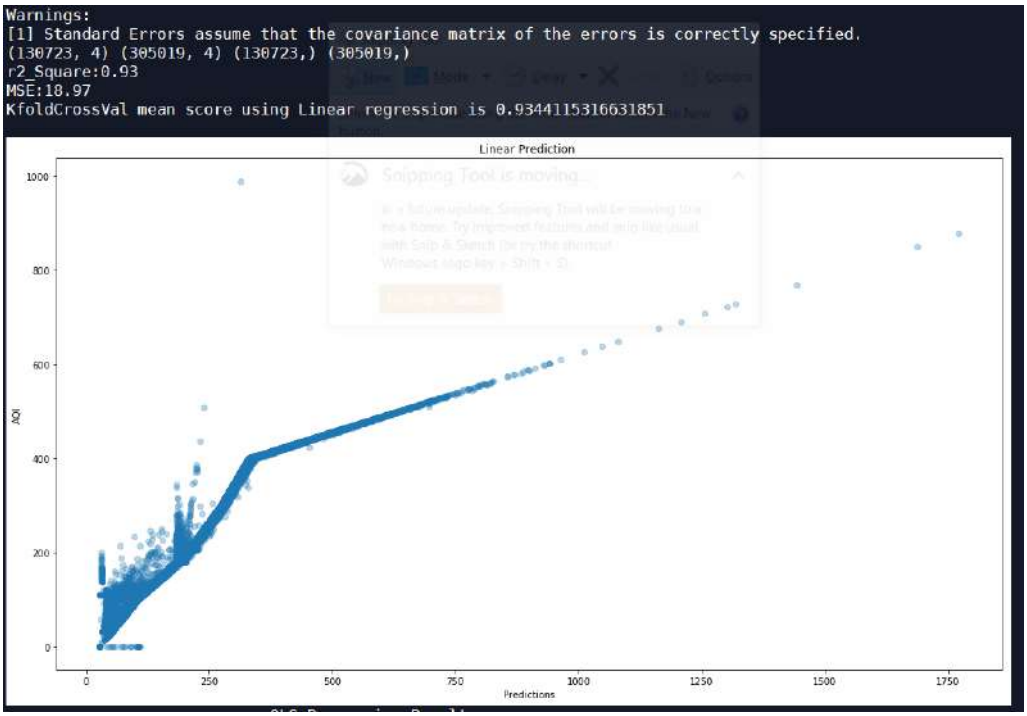


Appendix: I

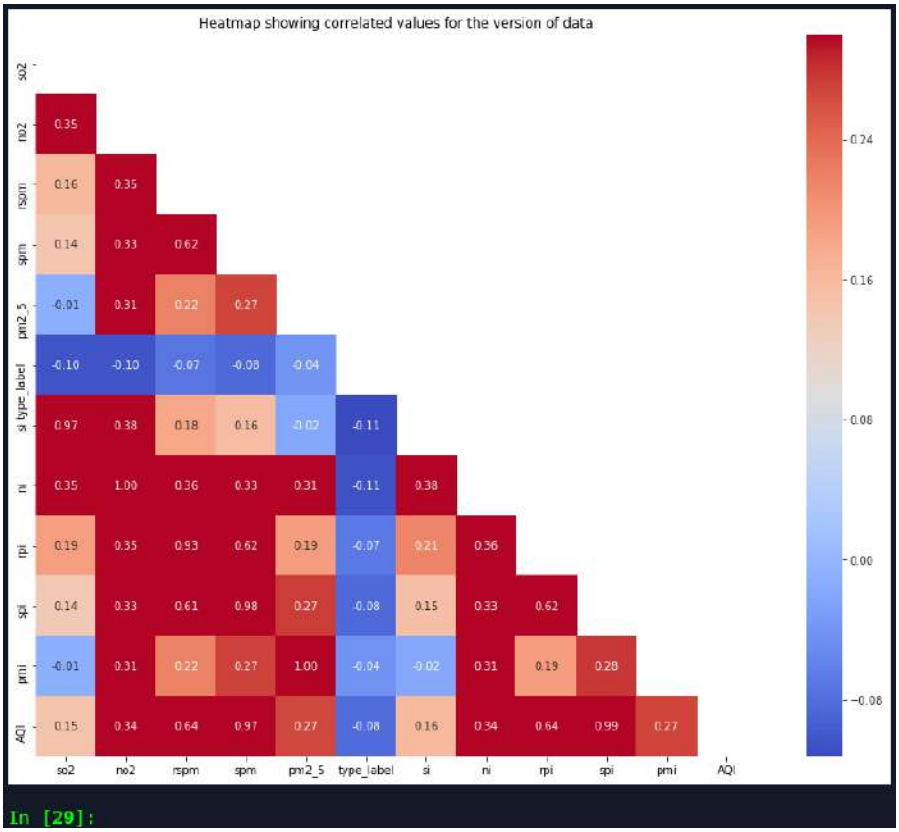
Linear Regression with Index calculation for so2,no2,spm,rspm,pm2_5.



Linear Regression for so2,no2,spm,rspm,pm2_5.



Correction in the dataset with numerical values.

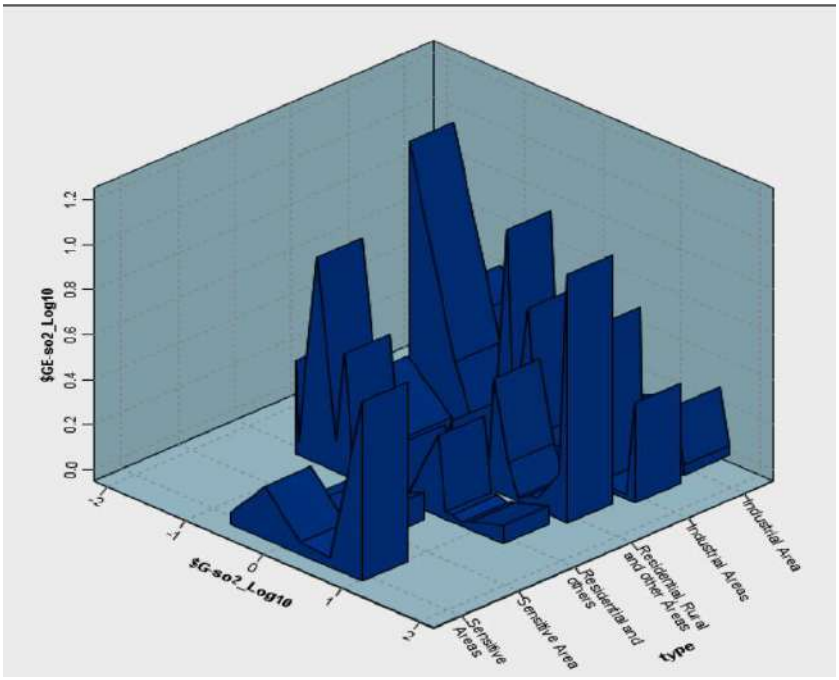

Heatmap showing correlated values for the version of data
so2
no2
rspm
spm
pm2_5
type_label
si
ni
rpi
spi
pmi
AQI
0.35
0.16 0.35
0.14 0.33 0.62
-0.01 0.31 0.22 0.27
-0.10 -0.10 -0.07 -0.08 -0.04
0.97 0.38 0.18 0.16 -0.02 -0.11
0.35 1.00 0.36 0.33 0.31 -0.11 0.38
0.19 0.35 0.93 0.62 0.19 -0.07 0.21 0.36
0.14 0.33 0.61 0.98 0.27 -0.08 0.15 0.33 0.62
-0.01 0.31 0.22 0.27 1.00 -0.04 -0.02 0.31 0.19 0.28
0.15 0.34 0.64 0.97 0.27 -0.08 0.16 0.34 0.64 0.99 0.27
0.24
0.16
0.08
0.00
−0.08

In [29]:

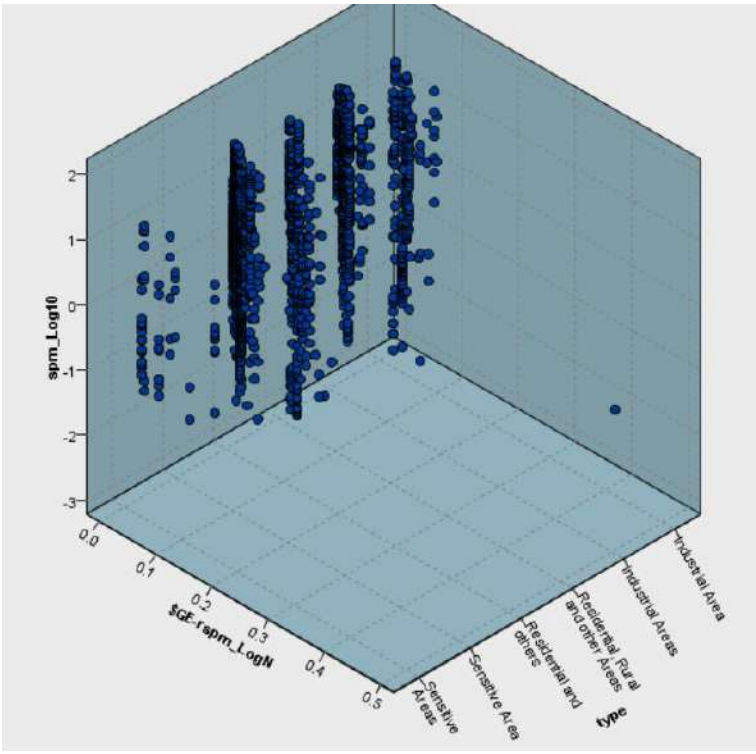
Correlation and dependency in attributes

Appendix: J

Algorithm result, which shows that the distribution of So2 is higher in Residencial rural and other areas followed by industrial and sensitive Areas.

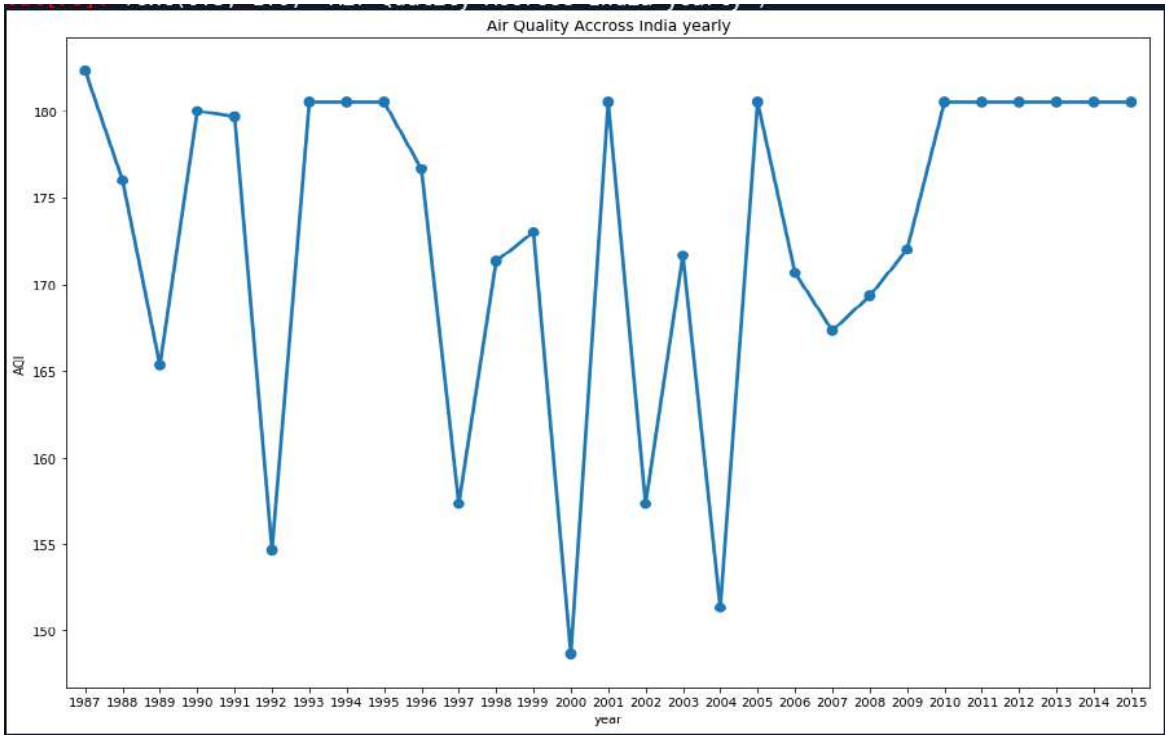


The distribution of rspm over spm based on emission type.

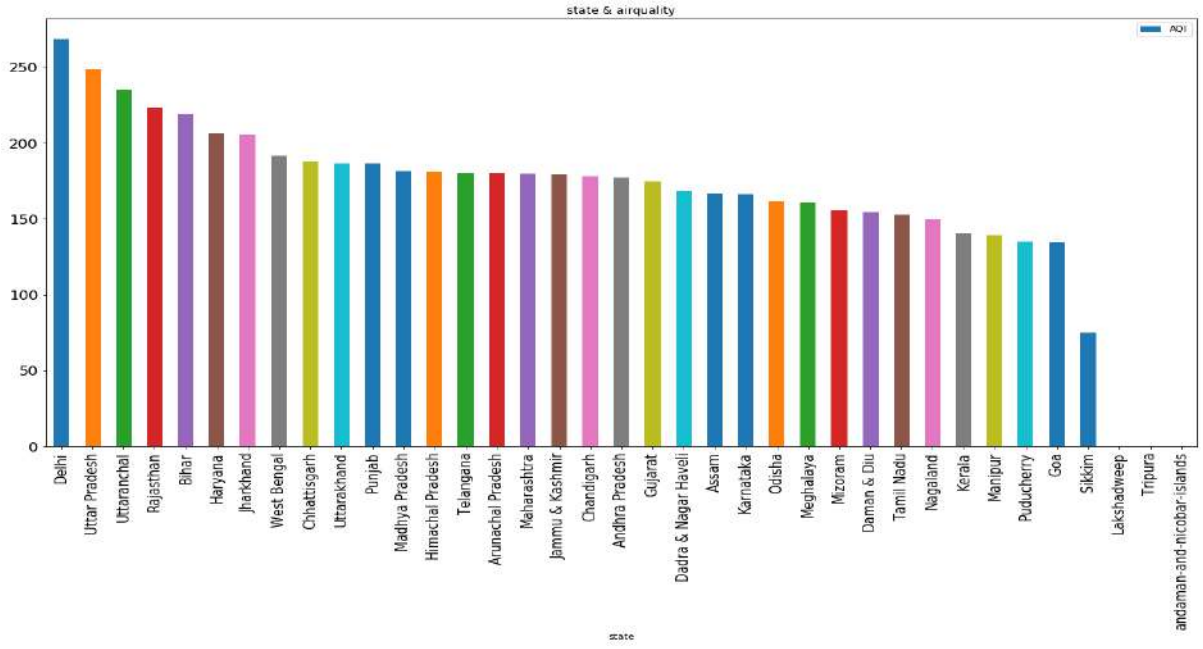


APPENDIX: K

AQI trande with a timestamp in last few years in INDIA

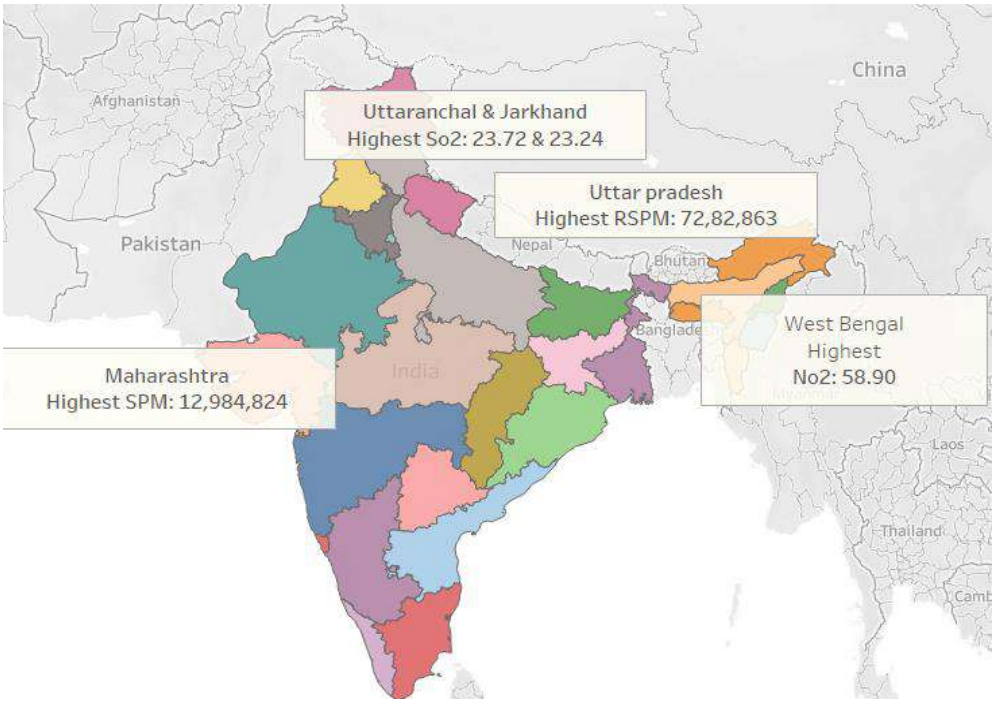


The Categorisation of risk zone based on AQI



Universal AQI analysis Table

Numerical Value	Color	Air Quality Index Levels of Health Concern	Meaning
0 - 50	Green	Good	Air quality is considered satisfactory, and air pollution poses little or no risk.
51 - 100	Yellow	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
101 - 150	Orange	Unhealty for sensitive groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
151 - 200	Red	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
201 -300	Purple	Very unhealthy	Health alert: everyone may experience more serious health effects.
301 - 500	Maroon	Hazardous	Health warnings of emergency conditions. The entire population is more likely to be affected.



"I acknowledge that the submitted work is my own original work in accordance with the University of Auckland guidelines and policies on academic integrity and copyright. (See: https://www.auckland.ac.nz/en/students/forms-policies-and-guidelines/studentpolicies-and-guidelines/academic-integrity-copyright.htmlLinks to an external site.).

I also acknowledge that I have appropriate permission to use the data that I have utilised in this project. (For example, if the data belongs to an organisation and the data has not been published in the public domain then the data must be approved by the rights holder.) This includes permission to upload the data file to Canvas. The University of Auckland bears no responsibility for the student's misuse of data."

How do Indian Business startups migrate their organizational services from on-premise to the cloud as a better solution choice with their limited resources in both technical skills and financial assets?

Abhishek D. Hirve
Department of science
(Master of Information Technology)
The University of Auckland
(city campus)
Auckland, New Zealand
email:-abhishek.girve@yahoo.com

ABSTRACT: **This report contains Information about Indian businesses which migrate to the cloud and with high performance and the organizational data. Migrating to cloud how it goes profitable for large and small scale organization's as effective audience. This research is conducted by a comparative study with different cloud service providers. It also includes price comparison with MS Azure platform vs. Amazon web services, which gives more better solutions to organizations for manage migration, and the growth of SaaS in India, Why ?- This research will light ups some Business and Technical Aspects for multi-scale Businesses In India for cloud migration. This paper can be referred to as solution processing for startups and give abstract to move on-premise to virtual services in the future. This paper is based on migration case studies. This paper includes a comparative view for startups to study different cost-effective and more relevant study while migrating services on the cloud. This study cases can be future example cases to make a relevant choice to choose more efficient cloud solution.**

Keywords:- on-premmises , IaaS, PaaS, SaaS, services, costings , LAN , CRM ,OLTP ,OLAP,framework.

I . Introduction

A. History of cloud computing

For understanding cloud concept it's needed to understand first 'what is computing ?'-computing is a process of utilizing computer technology to complete a task it involves hardware and software as a formation of any system. It can be defined as 'any objective oriented activity requiring benefiting from creating algorithmic processes through the computer."

There are multiple types involved in the term of computing. Cloud computing, ***Grid computing:*** multiple independent computing clusters which act like a 'Grid' because they are composed of resource nodes not located within a single administrative domain. Offering online computation known as utility computing like making 'virtual supercomputing,' utility computing, [22]

Distributed computing, cluster computing: it is a group of linked computers, working together with fast LAN usually deployed for improving performance. What is the cloud?: We can say that cloud is something, which is present at a remote location and provide public and private networks on WAN, LAN, or VPN application such as web apps, CRM's. [22]Cloud has multiple deployments models categorized as *Public cloud, private cloud, community cloud, Hybrid cloud,* which gives guessing for accessibility their names are self -explainers. [22] (Appendix: A)**Public categorizes cloud**: The public cloud allows system and services to be easily accessible to the public; it is less secure. **Private:** The private cloud allows system and services to be accessible within a more secure organization.**Hybrid:** It is a mixture of public and private services, in which critical transformation performed using private cloud and non-critical with the public cloud. Community cloud: community cloud allows systems and services to be accessible by a group of organizations.[22]

Research question: *what is a comparatively good solution to migrate like a new startup in India and Indian markets? Moreover, how it is cost-effective in which platforms AWS vs. Azure?*

II. Literature Review

A.migration views

Indian companies are continually looking for a cloud template that meets their requirements in a data-driven era. The cloud fracture has taken over the traditional software and platform as a service because of fresh IT expenditure from startups and medium-sized enterprises.[1] 'Gartner Reports that India

Public Cloud services industry will expand to $1.81 billion by 38% in 2017, with the greatest development powered by IaaS.

Cloud IaaS is going to exceed its traditional outsourcing information center (DCO) *'Dynamic creative optimization'* spending in India this year to average $677 million. By 2021, Cloud IaaS will have a total user expenditure of over $2.1 billion.' [1] "Companies are using IaaS for good effect, for getting the impactful effect with innovation and speed to their market values which gives them IaaS as best practice platform for availability, performance, and security. Most companies say that a progressive cloud strategy should include enterprise-wide IaaS and that businesses not investing in IaaS will increasingly find themselves struggling to keep pace with businesses." According to Mitesh Aggarwal, Vice President and CTO, Oracle India.[1]This Statements may present a view of Indian organizational culture mainly focused on IaaS, which will be beneficial for growth. Placing and maintaining their infrastructure for startups in India is quite a lengthy and costly process. Now if data management and computing services available virtually with cheap cost small and middle scale companies are started to make a choice to move their IT resources virtually, which can make them global in minus. Most of the companies are focusing on ERP solution and looking for 'EaaS'(Everything as services) from cloud vendors. Which is because of the limitation of resources and startup economy contributions in India. In startups and small scale businesses, it is easy to place infrastructure on manageable services rather than focus on hiring talent and setting up their infrastructure.

B.Why cloud?

- ***With the view of General purpose***

Cloud works in a distributed computing environment, which is agility. High scalability with 'OnDemand ' provisioning of resources large scale. Multi sharing which works as more efficient cost reduction by sharing common Unfractured. High

availability and reliability with no maintenance. Scalability and Elasticity via dynamic 'on demand' provisioning of resources on a fine-grained, self-service basis in near real-time(The VM startup time varies by VM type, location, OS, and cloud providers). The main reason for SME's(small medium scale enterprise's) migrating towards cloud productivity maintenance with quality in less time with less upfront and load .most of the services will work with Pay-per-use model.

cases. As per comparison, it concluded that service selection and based on service needs and funding organizAtions could compare cloud providers according to their needs.

C.Future Work

There is needed to study a perfect SOA for recently started business, and SME is a path to implement and things to consider while design SOA (Service-oriented Architecture) for active formation for the organizational database in small scale industries. The Indian cloud platform is increasing Indian Government is also taking part to provide effective solutions services are launched by NIC (National Informatics center) launched as '*Meghraj*' platform which is for the transformation of government on-premises services to the cloud. Most of the most prominent organizations are providing services in the private sector and GI-cloud services as well, like '*TATA COMMUNICATIONS*.'

References

[1]N. Arora, "Indian businesses open up to embracing cloud infrastructure", YourStory.com, 2019. [Online]. Available: https://yourstory.com/2017/07/indian-businesses-cloud-infrastructure. [Accessed: 21- May- 2019].

[2]P. Shukla, "How 'Cloud' Is Changing Grounds Of Indian Businesses", BW Businessworld, 2019. [Online].Available: http://www.businessworld.in/article/How-Cloud-Is-Changing-Grounds-Of-Indian-Businesses/04-08-2017-123511/.[Accessed:22- May- 2019].

[3]"Azure Cloud Solution Provider price list", *Docs.microsoft.com*, 2019. [Online]. Available: https://docs.microsoft.com/en-us/azure/cloud-solution-provider/billing/azure-csp-pricelist. [Accessed: 22- May- 2019].

[4]"Pricing – Azure SQL Database Single Database | Microsoft Azure", *Azure.microsoft.com*, 2019. [Online]. Available: https://azure.microsoft.com/en-us/pricing/details/sql-database/single/. [Accessed: 22- May- 2019].

[5]"Amazon RDS Pricing - Amazon Web Services", *Amazon Web Services, Inc.*, 2019. [Online]. Available: https://aws.amazon.com/rds/pricing/. [Accessed: 22- May- 2019].

[6]"Cloud Storage Pricing | S3 Pricing by Region | Amazon Simple Storage Service", *Amazon Web Services, Inc.*, 2019.[Online].Available: https://aws.amazon.com/s3/pricing/. [Accessed: 23- May- 2019].

[7]"Azure Page Blobs Storage Pricing | Microsoft Azure", *Azure.microsoft.com*,2019.[Online]. Available:https://azure.microsoft.com/en-us/pricing/details/storage/page-blobs/. [Accessed: 23- May- 2019].

[8]"CDN Pricing | Free Tier Eligible, Pay-as-you-go | Amazon CloudFront", *Amazon Web Services, Inc.*, 2019.[Online].Available: https://aws.amazon.com/cloudfront/pricing/?nc2=type_a. [Accessed: 23- May- 2019].

[9]"Pricing - Content Delivery Network (CDN) | Microsoft Azure", *Azure.microsoft.com*, 2019. [Online]. Available: https://azure.microsoft.com/en-us/pricing/details/cdn/. [Accessed: 23- May- 2019].

[10]"Pricing-FunctionsMicrosoft Azure", *Azure.microsoft.com*,2019.[Online]. Available:https://azure.microsoft.com/en-us/pricing/details/functions/. [Accessed: 23- May- 2019].

[11]"AWS Lambda – Pricing", *Amazon Web Services, Inc.*,2019.[Online].Available: https://aws.amazon.com/lambda/pricing/. [Accessed: 23- May- 2019].

[12]"Agriculture in India Goes Digital with AWS", *Amazon Web Services, Inc.*, 2019. [Online]. Available: https://aws.amazon.com/solutions/case-studies/iffco/. [Accessed: 23- May- 2019].

[13]D. Ghosh, "NASSCOM Cloud: Next Wave Of Growth In India 2019 - NASSCOM Community |The Official Community of Indian IT Industry", *NASSCOM Community |The Official Community of Indian IT Industry*, 2019. [Online]. Available: https://community.nasscom.in/communities/digital-transformation/nasscom-cloud-next-wave-of-growth-in-india-2019.html. [Accessed: 24- May- 2019].

.[14]"ABP News Case Study - Amazon Web Services (AWS)," *Amazon Web Services, Inc.*, 2019. [Online]. Available: https://aws.amazon.com/solutions/case-studies/abp_news/. [Accessed: 25- May- 2019].

[15]"Amazon Kinesis Video Streams Pricing," *Amazon Web Services, Inc.*, 2019. [Online]. Available: https://aws.amazon.com/kinesis/video-streams/pricing/. [Accessed: 24- May- 2019].

[16]"Pricing - Event Hubs | Microsoft Azure," *Azure.microsoft.com*,2019.[Online]. Available:https://azure.microsoft.com/en-us/pricing/details/event-hubs/. [Accessed: 24- May- 2019].

[17]"AWS | Amazon EMR | Pricing," *Amazon Web Services,Inc.*,2019.[Online].Available:https://aws.amazon.com/emr/pricing/?nc=sn&loc=4. [Accessed: 24- May- 2019].

[18]"Pricing - HDInsight (Hadoop) | Microsoft Azure," *Azure.microsoft.com*,2019.[Online]. Available:https://azure.microsoft.com/en-us/pricing/details/hdinsight/. [Accessed: 24- May- 2019].

[19]"AWS Partner Network," *Amazon Web Services, Inc.*,2019.[Online].Available:https://aws.amazon.com/partners/.[Accessed:25- May- 2019].

[20] J. G. Thoomkuzhy, "Cloud Adoption by Startups in India - Adoption, Preferred Cloud Models, Challenges andBusiness Transformation:ASurveyBased Investigation and Recommendation," vol. 7, no. 1, p. 7, 2016.

[21]N. Schäferhoff, "Popular CMS & Market Share(2018)websitesetup.org," *websitesetup.org*, 24-Apr-2019. [Online]. Available: https://websitesetup.org/popular-cms/.

[22]P. Kulkarni and J. Pawar, *RECENT TRENDS IN IT*, 1st ed. Pune,India: VISION PUBLICATIONS, 2010, pp. 5.1-5.25.

"I acknowledge that the submitted work is my own original work in accordance with the University of Auckland guidelines and policies on academic integrity and copyright. (See: https://www.auckland.ac.nz/en/students/forms-policies-and-guidelines/studentpolicies-andguidelines/academic-integrity-copyright.htmlLinks to an external site.). I also acknowledge that I have appropriate permission to use the data that I have utilised in this project. (For example, if the data belongs to an organisation and the data has not been published in the public domain then the data must be approved by the rights holder.) This includes permission to upload the data file to Canvas. The University of Auckland bears no responsibility for the student's misuse of data."

Are system design documents make a confused reflection of the system?

A study of various design specifications on understandability factors.

Abhishek D. Hirve
Department of science
(Master of Information Technology)
The University of Auckland
(city campus)
Auckland, New Zealand
email:-abhishek.hirve@yahoo.com

Abstract: **This paper gives a reflective view on document analysis .this research strategy gives a view of ambiguity generation and uncertainty generated in design documents. Qualitatively analyzed data which contains related themes and codes. This paper is based on case studies of organizational software design documents, user stories, a nonprofit organizational project which contains user story, system workflow, some design decisions, and some user manual with a verity of technical artifacts included in the documentation. Quantitative analysis help to understand the clarity of documents for better understanding.study includes ' Exploratory sequential mixed method' which is qualitative and a quantitative mixture of analysis.**

Key Words- *uncertainty, dependency, Relations, clarity*, Epicenes, use case, diagrams, synchronous, object-oriented software engineering.

I. Introduction

A layout document's primary objective is to create us more efficient by pushing us to reflect and collect input from others. People often believe that it is a matter of teaching others about a scheme or subsequently being used as paperwork. While these can be beneficial side impacts. [1]While writing a software design document, three questions have to take into consideration why? What how? Moreover, the process?. Design documents are a reflection of the system which gives usability and operational view for the world. Why:-design document will include valuable asset to overview system with multiple components with the technical design of the implemented advanced system. What:- design document reflects the virtual image of the system. It should contain an ample amount of system workflow information. How:- like multiple forms of implementing an SDS(Software Design specification) it should focus on clarity of the concept of a system which can be used for making a user manual of the proposed system in the future. Process:- while processing system documentation it must be needed to introduce Use perspective or actionable script in a more straightforward and less technical format, which is a more understandable document for the client if they do not have a technical background. If it is not precise documentation, it is difficult to understand and maintain processes whereas poor and missing parts in documents make the fault the full and low quality, which is converted further in cost raising problem in maintenance and updating. [2]

Research question: *Is uncertainty in design document give an ambiguous reflection of the built system? Moreover, how to reduce? What are the symptoms behind that?*

In this paper elements, those are needed to be included in understandable software documents are focused. The research question will focus on a document with Any components which are reflecting an ambiguity in understanding a process and workflow of the system performance, quality, understandability, and adverse impacts of unclear documentations .which will be helpful to reduce a fine line between uncertainty and ambiguity in the specification process.

References:

[1]A. Zhang, 2019. [Online]. Available: https://medium.freecodecamp.org/how-to-write-a-good-software-design-document-66fcf019569c. [Accessed: 27- May- 2019].

[2]L. Ilana, and Y. Aharon, "Raising Awareness of the Constituents of Software Design – The Case of Documentation", *Journal of Software Engineering and Applications*, vol. 03, no. 05, pp. 495-502, 2010. Available: 10.4236/jsea.2010.35056.

[3]N. Ikeya, R. Luck and D. Randall, "Recovering the emergent logic in a software design exercise", *Design Studies*, vol. 33, no. 6, pp. 611-629, 2012. Available: 10.1016/j.destud.2012.06.004.

[4]"Software development is a design activity - Extreme Uncertainty", *Extreme Uncertainty*. 2019.[Online].Available: https://www.extremeuncertainty.com/software-development-design-activity/. [Accessed: 27- May- 2019].

[5]Á.Garcíaand N. Medinilla, *Se.cs.toronto.edu*, 2019.[Online].Available:http://se.cs.toronto.edu/IWLU/papers/Ambiguity_Garcia.pdf. [Accessed: 28- May- 2019].

[6]"Shibboleth Authentication Request", *Go.galegroup.com.ezproxy.auckland.ac.nz*, 2019. [Online]. Available: http://go.galegroup.com.ezproxy.auckland.ac.nz/ps/i.do?id=GALE%7CA147240161&v=2.1&u=learn&it=r&p=ITOF&sw=w. [Accessed: 28- May- 2019].

[7]M. Rahmani, "Software modeling & design", *ACM SIGSOFT Software Engineering Notes*, vol. 36, no. 4, p. 35, 2011. Available: 10.1145/1988997.1989008.

[8]"Use Of Personas In Software Development", *Use Of Personas In Software Development*, 2019.

[9]R.Nikam, *object-oriented software engineering*, 1st ed. pune: vision publications, 2015, pp. 2.2 -2.3 ISBN-978-93-5016-334-4.

[10]T. Dybå, R. Prikladnicki, K. Rönkkö, C. Seaman and J. Sillito, "Qualitative research in software engineering", *Empirical Software Engineering*, vol. 16, no. 4, pp. 425-429, 2011. Available: 10.1007/s10664-011-9163-y.

[11]D. Bardon, D. Berry, C. Bjerke and D. Roberts, "Crafting the Compelling User Experience", *International Journal of Human-Computer Interaction*, vol. 14, no. 3, pp. 535-558,2002.Available:10.1207/s15327590ijhc143&4_14.

[12]*Analyzing qualitative data*, 1st ed. Routledge, 2002, pp. 106-119.

[13]"Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars.", *Ojs.aishe.org*, 2019. [Online]. Available: http://ojs.aishe.org/index.php/aishe-j/article/viewFile/335/553. [Accessed: 30- May- 2019].

[14] "John W. Creswell - Research Design_ Qualitative, Quantitative, and Mixed Methods Approaches (2013, SAGE Publications, Inc).epub." .

[15] H. Ziv, D. J. Richardson, and R. Klosch, "The Uncertainty Principle in Software Engineering," p. 13.

[16]*Itk.ilstu.edu*,2019.[Online].Available: http://www.itk.ilstu.edu/faculty/bllim/itk353/NotesSlidesDocs/MKS_Use_Cases.pdf. [Accessed: 30- M2019].

"I acknowledge that the submitted work is my own original work in accordance with the University of Auckland guidelines and policies on academic integrity and copyright. (See: https://www.auckland.ac.nz/en/students/forms-policies-and-guidelines/studentpolicies-andguidelines/academic-integrity-copyright.htmlLinks to an external site.). I also acknowledge that I have appropriate permission to use the data that I have utilised in this project. (For example, if the data belongs to an organisation and the data has not been published in the public domain then the data must be approved by the rights holder.) This includes permission to upload the data file to Canvas. The University of Auckland bears no responsibility for the student's misuse of data."

Name: Abhishek D. Hive.

Student Id : 347736209

Email: Abhishek.hirve @yahoo.com.

INFOSYS 737

ADAPTIVE ENTERPRISE SYSTEM

EXECUTIVE SUMMARY

This report contains an idea of creating a unique business solution A&A Power solutions. As creating a startup plan for a solution as a global issue of energy generation using renewable resources such as solar power. Project is about creating a solution for implanting solar panels for power generation, which is launching a new product in the market as Solar Power Tree(SPT). This report will describe that launching and implementing the project in high populated countries like India and as future scope for the global market. As taking account of global environmental issues and creating global as well as local level employability. The report also showcases multiple elements on which business plan is based on a key element, and that is essential to make the successful implementation of business logic and idea. With implementations of different business models and strategic analysis models for enterprise planning. The report also includes a business plan with value-added elements. The report also focuses on value chain activities and long term vision planning. The report also focuses on disruptive event management or risk management elements in enterprise planning. The business focuses on the Nish market in India and as getting exposure in future will concentrate on the global market. Which follows the principle of the stage of enterprise growth crucial transition and core management Mode. Strat and experiment with a small market and trivial towards globalisation.

REFERENCES

1. Barry J. Nalebuff, A. (1995). The right game: Use game theory to shape strategy. *Long Range Planning*, *28*(5), 128. doi: 10.1016/0024-6301(95)90326-7

2. Energy - United Nations Sustainable Development. (2019). Retrieved 25 August 2019, from https://www.un.org/sustainabledevelopment/energy/

3. Energy - United Nations Sustainable Development. (2019). Retrieved 14 August 2019, from https://www.un.org/sustainabledevelopment/energy/ .

4. Energy, Environment and Resilience - UN India. (2019). Retrieved 25 August 2019, from https://in.one.un.org/un-priority-areas-in-india/energy-environment-and-resilience/

5. Energy, Environment and Resilience - UN India. (2019). Retrieved 14 August 2019, from https://in.one.un.org/un-priority-areas-in-india/energy-environment-and-resilience/ .

6. G.S Laxmi (2019),ENERGY STATISTICS, 2019. (2019). Retrieved 13 August 2019, from http://www.mospi.gov.in/sites/default/files/publication_reports/Energy%20Statistics%202019-finall.pdf .

7. Gorner, M., & Scheffer, S. (2019). TCEP: Electric vehicles. Retrieved 15 August 2019, from https://www.iea.org/tcep/transport/electricvehicles/ .

8. https://singularityhub.com/2016/11/22/the-6-ds-of-tech-disruption-a-guide-to-the-digital-economy/

9. Jalagat, R. (2016). The impact of change and change management in achieving corporate goals and objectives: Organizational perspective. *International Journal of Science and Research (IJSR), 5*, 1233-1239.

10. *land acquisition issue in india*. (2019). [Image]. Retrieved from http://indpaedia.com/ind/index.php/Land_acquisition:_India

11. Murthy. (2014). India's solar energy future. Retrieved 14 August 2019, from http://www.indiaenvironmentportal.org.in/files/file/India's%20solar%20energy%20future.pdf

12. Steven Kotler. (2006). Steven Kotler's Six D's of Exponential Technology.
13. Sundaram, D., & Ahmed, M. (2012). Sustainability modelling and reporting: From roadmap to implementation. *Decision Support Systems*, *53*(3), 611-624. doi: 10.1016/j.dss.2012.02.004

14. Sustainable Development Goals. (2019). Retrieved 25 August 2019, from https://in.one.un.org/page/sustainable-development-goals/

15. Tomar, A., & Gautam, K. (2015). A Review of Solar Energy - Challenges, Economics & Policies in India. *International Journal Of Science And Research (IJSR)*, *6*(2319-7064), 2080 to 2082.