PROJECT REPORT
On

ANALYSIS OF HOUSING DATA

IN PARTIAL FULFILLMENT
OF
REQUIREMENT FOR THE DEGREE OF
B.Tech in Computer Science

Submitted by:

AKHILESH LAMBA (11CSU012)
ANMOL KHANNA (11CSU22)

Under the Guidance of

Dr. Latika Singh
Associate Professor

Department of CSE & IT
ITM University, Gurgaon

Acknowledgement

We express our heartiest gratitude & sincere regards to Dr Latika Singh for her continuous guidance, monitoring, & informal discussions which were a beam of light for us in the entire duration of this project.

We are also thankful to the Lab Assistants for their regular help in completing the project at the time.

Last, but not the least we would like to thank our friends Atul Bhatt and Devinder for helping us.

ANMOL KHANNA
11CSU022

AKHILESH LAMBA
11CSU012

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources.

ANMOL KHANNA
11CSU022

AKHILESH LAMBA
11CSU012

Certificate

This to certify that the **Final Report** of project titled **EPathshala** , which is being submitted is an authentic work carried out by **Anmol Khanna , Akhilesh Lamba** under the guidance of **Dr. Latika Singh .** They have successfully completed the above mentioned project.

Dr Latika Singh
Dept. of CSE & IT
ITM University

CONTENTS

ABSTRACT

In Our project titled the Analysis of housing data, we took Housing data from 2011 census. We took the data of Haryana Region covering all the districts and contained information like sources of electricity , source of cooking fuels , total number of households , households in good , liveable and bad condition etc. In this project we used a rapid miner tool to do our analysis work. For doing our analysis we used kmeans clustering algorithm , select operator etc. The main objective behind this project is to develop an information package helping in making the right distribution of funds for districts and ensuring that underdeveloped area are not left behind.

Introduction

As new technologies are coming up at a very fast rate and so is amount of data being generated day in day out is also increasing exponentially making it necessary for organizations to have tools and technology in place to make future company policies and understanding and analyzing huge data to make decisions to cater to customer demands. Therefore Business intelligence is becoming an integral part of many organization be it Google or Amazon or Flipkart.

Business intelligence (**BI**) is the set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis purposes. BI technologies are capable of handling large amounts of unstructured data to help identify, develop and otherwise create new strategic business opportunities. The goal of BI is to allow for the easy interpretation of these large volumes of data. Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability.

BI technologies provide historical, current and predictive views of business operations. Common functions of business intelligence technologies are reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics andprescriptive analytics.

BI can be used to support a wide range of business decisions ranging from operational to strategic. Basic operating decisions include product positioning or pricing. Strategic business decisions include priorities, goals and directions at the broadest level. In all cases, BI is most effective when it combines data derived from the market in which a company operates (external data) with data from company sources internal to the business such as financial and operations data (internal data). When combined, external and internal data can provide a more complete picture which, in effect, creates an "intelligence" that cannot be derived by any singular set of data.

STEPS DONE FOR THE COMPLETION OF THE PROJECT:

1.DATA PREPROCESSING: Our data set was having many parameters according to which the information was compiled in an excel sheet. The parameters varied from personal information to information about basic amenities available in the locality. So we thought of only dealing with

basic amenities that are necessary for a person and hence we chose the following listed parameters

- Total number of households in liveable condition
- Closed Drainage

- Water from treated Source
- Cooking Fuel used (LPG/PNG)
- There was also dense information regarding districts so we only chose District name attribute having the ward number acting as a primary key for our table.

2.DATA INTEGRATION AND DATA QUALITY: Using the Talend data integration and Talend data quality tools we checked the quality of our data set and this step helped us to know the outliers present in our data set.

3.RAPID MINER: This step was the most interesting and challenging one as it had many operators available to be applied on our data set Eg: Clustering Algo., Decision Tree, Verification and Validation. So choosing the correct operator which can be applicable on our data was really a challenge but it was interesting and an enjoyable task. So after going through many tutorials and guidance from Latika Mam, we finally used K-Means clustering Algorithm to get the outcome we intended to achieve before taking on this project.

Objective Of the Project

Our project titled **Analysis of Housing Data** is basically aimed at identifying districts in Haryana state which are under developed and which are developed. Through our project we have identified areas which are fine and in liveable condition and which are in penury condition (See the appendix).

This analysis of data will help the agencies to be in better position to know which districts are needing immediate need of funds and which are in ok condition and can be dealt with marginal distribution of funds. Through this analysis there can be an even distribution of funds and ensuring that no district is left underdeveloped or is left in poverty condition.

This project also identifies the resource distribution across areas i.e. identifies areas with abundance of one resource and some areas which don't even have pure drinking water and we have also identified areas which don't even have basic sanitation facilities.

The analysis of housing data will be an important step to ensure development takes place in the manner it should happen and there is no biasness towards distribution of funds based on political influence.

TOOL USED

MySQL: the world's second most widely used relational database management system (RDBMS) and most widely used open-source RDBMS. It is named after co-founder Michael Widenius's daughter, My. The SQL acronym stands for Structured Query Language.
The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedishcompany MySQL AB**,** now owned by Oracle Corporation.

Talend data integration: Talend Open Studio for Data Integration is an open source data integration product developed by Talend and designed to combine, convert and update data in various locations across a business.

Rapid Miner: is a software platform developed by the company of the same name that provides an integrated environment formachine learning, data mining, text mining, predictive analytics and business analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization. RapidMiner is developed on a business source model which means the core and earlier versions of the software are available under an OSI-certified open source license on Sourceforge. A Starter Edition is available for free download, a Personal Edition is offered for US$999, a Professional Edition is $2,999 and pricing for the Enterprise Edition is available from the developer.

MsExcel: is a spreadsheet application developed by Microsoft for Microsoft Windows, Mac OS X, and iOS. It features calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications. It has been a very widely applied spreadsheet for these platforms, especially since version 5 in 1993, and it has replaced Lotus 1-2-3 as the industry standard for spreadsheets. Excel forms part of Microsoft Office.

OUTCOMES

We have divided the total number of districts into different clusters using K-means algorithm, the clustering is based on the values of different parameters like Households in good condition, closed drainage, water from treated sources , LPG/PNG (See the appendix attached).

CLUSTER 0:

Parameters	Avg Value
Household in good condition	67.95
Closed drainage	13.43
Water from treated Resource	4.75
Cooking Fuel(LPG/PNG)	20.87
Total Districts	53

CLUSTER 1:

Parameters	Avg Value
Household in good condition	54.51
Closed drainage	75.70
Water from treated Resource	5.16
Cooking Fuel(LPG/PNG)	27.24
Total Districts	70

CLUSTER 2:

Parameters	Avg Value
Household in good condition	72.38
Closed drainage	69.41
Water from treated Resource	79.20
Cooking Fuel(LPG/PNG)	86.30
Total Districts	44

CLUSTER 3:

Parameters	Avg Value
Household in good condition	67.94
Closed drainage	28.61
Water from treated Resource	12.47
Cooking Fuel(LPG/PNG)	65.00
Total Districts	63

CLUSTER 4:

Parameters	Avg Value
Household in good condition	30.30
Closed drainage	30.37
Water from treated Resource	4.18
Cooking Fuel(LPG/PNG)	22.06
Total Districts	48

CLUSTER 5:

Parameters	Avg Value
Household in good condition	68.82
Closed drainage	82.58
Water from treated Resource	16.11
Cooking Fuel(LPG/PNG)	67.94
Total Districts	71

Total Districts in Gurgaon:**349**

Based on the above results, each cluster has some shortcomings and some strengths.

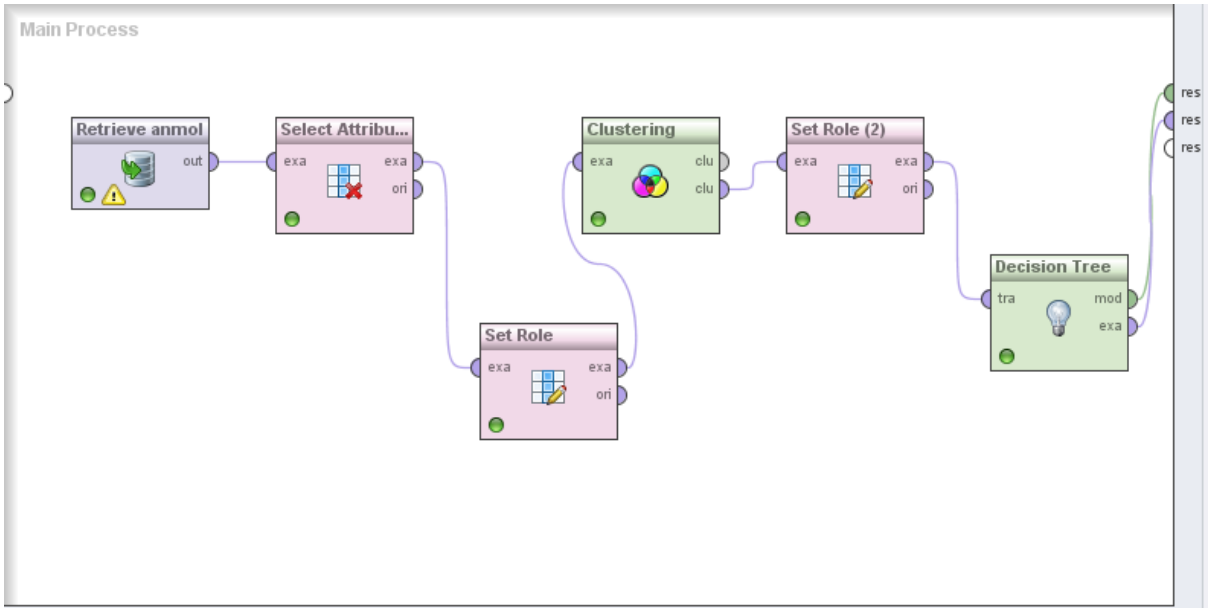

Fig showing the diagram made on the rapid miner tool

CONCLUSION

Based upon outcomes of our analysis,The following points are worth noticing:

- **Cluster 0:** It consists of regions having a very high percentage of houses in good condition but lacking in basic amenities very badly. So with allocation of funds towards this region for improving the basic amenities, these areas can provide excellent housing opportunities for the state.
- **Cluster 1:** Areas in this cluster are having adequate houses but the problem of availability of Pure drinking water and cooking fuel persists. So if the agencies pay more attention to these problems in these areas then it can serve them well for the people living there or for the future generations.
- **Cluster 2:** This cluster contains the best areas in terms of Houses available and also the availability of basic amenities. Hence in immediate future these areas can be dealt with marginal distribution of funds and thus preventing from wastage of funds.
- **Cluster 3:** This cluster consists of areas with ample good houses and cooking fuel is also available but problem is with closed drainage and availability of Pure drinking water. So agencies for these regions must focus on increasing availability of drinking water and also ensuring that measures are taken to improve drainage facilities.
- **Cluster 4:** This cluster has many backward areas in terms of development and hence these areas are the ones which need immediate action to be taken and if taken care of in the near future, then these areas can be boom for the state in terms of revenue generated and will also help in attracting more investment.
- **Cluster 5:** Areas within this cluster are only lacking in availability of water from the treated sources, hence steps must be taken to ensure that the problem is eradicated from the area as it can be seen from the data that there is ample infrastructure available.

APPENDIX

Areas wise distribution for

- **Cluster 0:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG
7	Nurgarh(12)	cluster_0	66.1	0	0	9.9
12	Husainka(15)	cluster_0	63.2	34.3	0	2.9
16	Chhilarki(14)	cluster_0	63.4	0	0.6	19.3
18	Brijpura(18)	cluster_0	47.8	3.5	2.7	1.8
23	Inchhapuri(264)	cluster_0	39.6	2	50.7	19.9
32	Bapas(266)	cluster_0	92.6	31.1	0	16.2
34	Khalilpur(272)	cluster_0	80.7	0.5	0.5	26.6
48	Lokra(289)	cluster_0	61.5	16.1	0.5	22.6
51	Khanpur(34)	cluster_0	76.4	1.7	0	36.8
52	Hera Heri(36)	cluster_0	85.8	0	87.5	0
55	Bas Padamka (41)	cluster_0	65.4	31.4	11.3	19.3
121	Medawas(85)	cluster_0	64.5	24.5	0	35.5
174	Ghausgarh(26)	cluster_0	51.5	0.1	1.2	31.5
178	Jatola(22)	cluster_0	53.6	27.8	1.2	40.7
179	Khandevila(21)	cluster_0	61.6	0	1.1	22.2
180	Mahchana(16)	cluster_0	89.7	0	3.2	10.8
181	Faredpur(15)	cluster_0	44.5	1.3	0	6.1
182	Karola(14)	cluster_0	47.1	0.8	0.7	20.1
195	Jhund Sarai Viran(123)	cluster_0	60	2.2	4.4	13.3
200	Farrukhnagar (Rural)(35)	cluster_0	66.4	28	0.5	35.2
212	Sultanpur(39)	cluster_0	52.6	22.6	2.9	31.1
233	Bhun Karka(137)	cluster_0	94.7	18	3.1	5.3
234	Prasoli(138)	cluster_0	76.2	1.1	6.4	21.6
236	Rathiwas(140)	cluster_0	70.8	18.6	4.9	38.7
237	Danokri(141)	cluster_0	97.2	0	0	48.1

238	Bhudka(142)	cluster_0	94.1	16.5	0.4	15.8
239	Pathrari(143)	cluster_0	75.2	32.5	1.1	35.7
255	Sehrawan(152)	cluster_0	86.9	34.8	0	9.6
257	Bar Gujar(156)	cluster_0	87.8	2.3	0.5	1.9
259	Sakatpur(159)	cluster_0	63.1	41.4	1	27.5
260	Hassanpur(163)	cluster_0	59.7	8.4	7.5	31.4
261	Darbaripur(162)	cluster_0	99.2	13.7	7.1	26.1
263	Kherki(3)	cluster_0	65.1	42.8	2.8	23.9
264	Baghanki(4)	cluster_0	68.2	42	3.6	24.5
272	Bandhwari(79)	cluster_0	63.7	20.5	1.6	12.3
273	Ulhawas(83) Part	cluster_0	79.9	34.1	4.7	26.2
274	Kadarpur(84)	cluster_0	68.7	33.7	0.8	39
275	Rethoj(169)	cluster_0	43.2	0.4	11.4	48.2
276	Sahjawas(170)	cluster_0	59	3.1	0.3	18.6
279	Mahendwara(178)	cluster_0	58.8	0	0.7	46.3
280	Kherla(177)	cluster_0	76.5	11.7	2.4	32.4
283	Abhepur(173)	cluster_0	82.2	5.4	1	28.7
287	Kuliaka(219)	cluster_0	56.2	0	0	0
288	Bai Khera (224)	cluster_0	100	31	3	1
291	Saramathla(225)	cluster_0	51.6	8.7	0.6	26.4
294	Bedhwaka(218)	cluster_0	50	0	0	0
299	Tolni(202)	cluster_0	94.4	27.8	0	16.7
300	Ranika Singhola (215)	cluster_0	48.1	15.6	0	1.3
304	Khuntpuri(209)	cluster_0	49.1	0	1.8	1.8
309	Chuharpur (203)	cluster_0	61.8	2.9	0	23.5
317	Lohatki(194)	cluster_0	61.9	14.8	16.5	25
318	Siraska(193)	cluster_0	72.4	1.7	0	0
326	Berka(181)	cluster_0	52	0.5	0	27

- **Cluster 1:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG
4	Sub-Dist - Pataudi	cluster_1	60.3	59.3	9.7	43
5	Sub-Dist - Pataudi	cluster_1	58.4	50.3	5.7	34.2
11	Ghudana(9)	cluster_1	76.4	76.7	0	32.1
13	Rajpura(8)	cluster_1	37.9	72.8	0	30.4
14	Muzaphara Alias Mandpura(262)	cluster_1	29.1	66.6	0	31.9
15	Sherpur(16)	cluster_1	27.9	65.3	1.2	22.6
17	Jasat(17)	cluster_1	94.3	83.5	0.5	3.1
21	Haqdarpur(19)	cluster_1	50.4	50.8	4.5	21.5
24	Milakpur(3)	cluster_1	33.9	75.1	0	13
25	Mirzapur(4)	cluster_1	57.6	80.7	0	31.8
26	Dewlawas(5)	cluster_1	53	86.7	2.4	39.8
27	Chhawan(2)	cluster_1	67.1	93.8	0.6	6.8
28	Ransika(31)	cluster_1	65.6	93.9	4.3	36.2
35	Ghilnawas(273)	cluster_1	22.2	97.6	0.8	3.2
38	Daultabad(28)	cluster_1	73.9	62	0.9	20.1
39	Gadaipur(30)	cluster_1	49.5	77.9	2.9	39.9
44	Lohchap(26)	cluster_1	26.6	74.1	3.1	2.4
45	Lokri(287)	cluster_1	61.9	60.1	2.2	30.4
47	Darapur(288)	cluster_1	80.9	86.4	3	35.3
50	Goriawas(33)	cluster_1	79.8	92.9	2	36.4
57	Uncha Majra(43)	cluster_1	57	73.9	7.8	40.9
58	Rampur(45)	cluster_1	70.7	81.6	1.7	41
59	Narhera(44)	cluster_1	66.3	68.2	4.4	42.3
68	Hailey Mandi (MC) - Ward No.8	cluster_1	40.7	95.3	19	44.7
71	Hailey Mandi (MC) - Ward No.11	cluster_1	43.4	59.7	0	28.9
72	Hailey Mandi (MC) - Ward No.12	cluster_1	43.3	96.4	11.8	34.8
73	Hailey Mandi (MC) - Ward No.13	cluster_1	37.7	80.9	7.1	44.6
84	Pataudi (MC) - Ward No.10	cluster_1	45.1	93.3	16.2	41.9
95	Budhera(43)	cluster_1	38.5	64.9	9.1	53.1
96	Mankrola(42)	cluster_1	91.1	87.9	2.9	15.9

99	Babupur(60)	cluster_1	56.9	100	0.8	35
103	Kharki Majra Dhankot(52)	cluster_1	32.7	75	11.4	54.5
112	Meoka(121)	cluster_1	12.5	97.2	11.1	1.4
116	Tikli(167)	cluster_1	58.1	52.6	6.1	35.4
117	Aklimpur(166)	cluster_1	79.4	95.3	3.2	39.6
164	Sub-Dist - Farrukhnagar	cluster_1	57.9	51.2	7.1	38.8
165	Sub-Dist - Farrukhnagar	cluster_1	57.7	47.1	7.3	37
183	Gugana(4)	cluster_1	53.4	69.6	0.7	31.1
184	Rajpur(13)	cluster_1	57.3	96	0	1.3
189	Fazalpur Badli(31)	cluster_1	54.3	59.9	5.4	21.9
190	Tajnagar(30)	cluster_1	60.1	49.7	0.2	24.1
191	Khawaspur(29)	cluster_1	38.1	65.5	5.5	49
192	Jarola(32)	cluster_1	74.6	73.2	0.2	40.4
199	Patli Hajipur(34)	cluster_1	46.2	57.6	9.2	29
201	Khurmpur(8)	cluster_1	56.1	59.7	3.3	4
202	Daboda(10)	cluster_1	54.2	55.3	0.8	1.9
205	Bir Hera(5)	cluster_1	79.2	98.9	2.7	26.8
206	Shekhupur Majri(3)	cluster_1	57.9	48	0.8	40.5
207	Siwari(1)	cluster_1	39.8	57.3	1	17.7
208	Jaraun(2)	cluster_1	61.4	96.7	0	30
210	Hari Nagar (7)	cluster_1	37.3	88.1	5	25.7
215	Kaliawas(41)	cluster_1	59.5	98.6	17.9	21.5
217	Farrukhnagar (MC) - Ward No.1	cluster_1	21.2	83.3	0.9	14
223	Farrukhnagar (MC) - Ward No.7	cluster_1	82.5	83.6	0.4	42
226	Farrukhnagar (MC) - Ward No.10	cluster_1	23.3	90	28.7	32.5
227	Farrukhnagar (MC) - Ward No.11	cluster_1	35.3	97	1	23.9
231	Nurpur Bahora(135)	cluster_1	88.3	70.4	1.9	12.2
232	Bahora Khurd(136)	cluster_1	48	74	0.6	18
245	Gwaliar(150)	cluster_1	49.6	82.3	8.6	26.3
249	Badha(113)	cluster_1	38.7	62.4	3.8	27.5
253	Kasan (129)	cluster_1	62.7	68.4	4.6	40.4
258	Navrangpur(157)	cluster_1	59	58.3	6	22.5
281	Damdma(174)	cluster_1	82.6	85.1	1.7	14.6
292	Ghangola(212)	cluster_1	87.2	67.6	15.4	8.9
298	Isaka (201)	cluster_1	48	78	2	22

307	Silani(207)	cluster_1	67.8	86.7	0.9	14.7
311	Kharoda (198)	cluster_1	82.9	69.6	0.6	23.4
322	Raipur (188)	cluster_1	20.9	85.9	1.7	7
327	Alipur(180)	cluster_1	69.6	60.7	1.2	25.3
342	Sohna (MC) - Ward No.9	cluster_1	23.3	94.8	58.1	22.9

- **Cluster 2:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG
1	District - Gurgaon	cluster_2	65.6	60.3	54.1	74.3
3	District - Gurgaon	cluster_2	68.7	66.6	71.5	86.6
63	Hailey Mandi (MC) - Ward No.3	cluster_2	85.5	74.4	57.6	64.4
65	Hailey Mandi (MC) - Ward No.5	cluster_2	77.5	99.4	56.9	98.1
88	Sub-Dist - Gurgaon	cluster_2	68.7	65.5	73	86.8
90	Sub-Dist - Gurgaon	cluster_2	69.2	66.2	76.5	88.5
98	Mohmadheri(58)	cluster_2	50	62.5	60.8	85
127	Gurgaon (M Corp. + OG)	cluster_2	69.4	66.1	77.9	88.8
128	Gurgaon (M Corp. + OG) - Ward No.1	cluster_2	61.7	44.1	77.5	87.4
129	Gurgaon (M Corp. + OG) - Ward No.2	cluster_2	86	76.1	96.4	95.8
130	Gurgaon (M Corp. + OG) - Ward No.3	cluster_2	65.4	82.7	95.3	94.2
131	Gurgaon (M Corp. + OG) - Ward No.4	cluster_2	71.2	97	95.3	94.9
132	Gurgaon (M Corp. + OG) - Ward No.5	cluster_2	63.8	95.3	93.7	95
133	Gurgaon (M Corp. + OG) - Ward No.6	cluster_2	64.2	43	93.4	91.4

134	Gurgaon (M Corp. + OG) - Ward No.7	cluster_2	83.2	87.7	96.6	95.8
135	Gurgaon (M Corp. + OG) - Ward No.8	cluster_2	73.1	92.5	98	98.5
136	Gurgaon (M Corp. + OG) - Ward No.9	cluster_2	71.9	76.5	89.6	89.5
137	Gurgaon (M Corp. + OG) - Ward No.10	cluster_2	45.8	37.4	50.4	92.4
139	Gurgaon (M Corp. + OG) - Ward No.12	cluster_2	59.8	33.6	69.3	86.5
140	Gurgaon (M Corp. + OG) - Ward No.13	cluster_2	78.3	74	84.4	86.8
141	Gurgaon (M Corp. + OG) - Ward No.14	cluster_2	78.8	67.4	79.9	81.4
142	Gurgaon (M Corp. + OG) - Ward No.15	cluster_2	79.1	75.8	78.7	85.7
143	Gurgaon (M Corp. + OG) - Ward No.16	cluster_2	95.4	97.5	96.8	96.1
144	Gurgaon (M Corp. + OG) - Ward No.17	cluster_2	76.9	68.1	82.3	84.9
146	Gurgaon (M Corp. + OG) - Ward No.19	cluster_2	75.1	48.1	52.3	87.6
147	Gurgaon (M Corp. + OG) - Ward No.20	cluster_2	77.3	94.9	62.6	87.5
148	Gurgaon (M Corp. + OG) - Ward No.21	cluster_2	89.8	85.7	92.4	93.1
149	Gurgaon (M Corp. + OG) - Ward No.22	cluster_2	80.9	82	84.9	87
150	Gurgaon (M Corp. + OG) - Ward No.23	cluster_2	66.7	67.6	71.9	80.4
151	Gurgaon (M Corp. + OG) - Ward No.24	cluster_2	57.2	42.4	74.1	74.4
152	Gurgaon (M Corp. + OG) - Ward No.25	cluster_2	72.9	81.1	89.8	92.4
154	Gurgaon (M Corp. + OG) - Ward No.27	cluster_2	86.3	87.1	91.4	95.5

155	Gurgaon (M Corp. + OG) - Ward No.28	cluster_2	77.9	80.9	87.9	92.8
156	Gurgaon (M Corp. + OG) - Ward No.29	cluster_2	46.9	51.3	51.5	90
157	Gurgaon (M Corp. + OG) - Ward No.30	cluster_2	70	67.2	76.2	94
158	Gurgaon (M Corp. + OG) - Ward No.31	cluster_2	78.6	36.3	83.2	90.6
159	Gurgaon (M Corp. + OG) - Ward No.32	cluster_2	84.2	65	88.3	92
160	Gurgaon (M Corp. + OG) - Ward No.33	cluster_2	58.5	31.1	57.2	92.4
162	Gurgaon (M Corp. + OG) - Ward No.35	cluster_2	64.1	66.8	83.8	64.7
214	Iqbalpur(40)	cluster_2	61.8	98.8	76.5	36.5
293	Khatrika(213)	cluster_2	100	0	100	38.5
339	Sohna (MC) - Ward No.6	cluster_2	74.2	90.2	91.7	95.4
345	Sohna (MC) - Ward No.12	cluster_2	80.9	98.6	84.4	97.4

- **Cluster 3:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG
2	District - Gurgaon	cluster_3	57.5	43.8	8.3	42
8	Haliaki(13)	cluster_3	89.2	1.3	0	58
9	Mangwaki(11)	cluster_3	91.4	0	0	64.1
20	Dadawas(20)	cluster_3	79.5	0	12.4	52.8
36	Mehaniawas(22)	cluster_3	60.6	7.1	2.4	65.4
40	Telpuri(23)	cluster_3	68.2	0	0	66.9
41	Nanu Khurd(29)	cluster_3	91.6	20.5	0	49.5
46	Mao(297)	cluster_3	85.3	28.1	0.3	51.1
56	Sapedar Nagar(37)	cluster_3	67.6	5.9	0	55.9
67	Hailey Mandi (MC) - Ward No.7	cluster_3	64	44.5	39	49.3
70	Hailey Mandi (MC) - Ward No.10	cluster_3	66.2	47.7	11.5	48.1
75	Pataudi (MC) - Ward No.1	cluster_3	55.2	40.9	2.6	49.8

83	Pataudi (MC) - Ward No.9	cluster_3	60.1	35.6	50.6	68.7
89	Sub-Dist - Gurgaon	cluster_3	60.2	54.1	13.9	57.6
92	Basharia(125)	cluster_3	53.5	46	0.3	73.2
97	Dharampur(59)	cluster_3	47.4	7.7	0.4	66.8
100	Bajghera(61)	cluster_3	64.5	3.3	37.5	72.4
101	Ghasula(88) Part	cluster_3	42.2	44.4	4.4	97.8
102	Tikampur(54)	cluster_3	100	0	0	100
106	Harsaru(107) Part	cluster_3	64	23.5	13.8	70.1
109	Hamirpur(116)	cluster_3	82.8	0	6.9	89.7
110	Wazirpur(115)	cluster_3	29.1	45.7	8.4	73.8
113	Kankrola(128)	cluster_3	70.5	16	12.1	74
118	Nurpur Jharsa(165)	cluster_3	69.9	39.8	0	57.5
120	Dhumaspur(86)	cluster_3	96.1	11.6	9.7	59.4
122	Nangli Umarpur(82) Part	cluster_3	100	0	0	68.8
138	Gurgaon (M Corp. + OG) - Ward No.11	cluster_3	50.4	19.1	8	80.8
153	Gurgaon (M Corp. + OG) - Ward No.26	cluster_3	41.8	23.3	52.5	71.1
163	Gurgaon (M Corp. + OG) - Ward No.36	cluster_3	96.2	24.4	4.3	62
167	Bahora Kalan(134)	cluster_3	49.6	45.8	6.1	51.2
168	Fakharpur(133)	cluster_3	85.8	41.4	2.4	63.7
170	Tatarpur(27)	cluster_3	91.7	32.9	32.9	57.1
173	Jamalpur(28)	cluster_3	59.1	46.5	22.3	58.2
177	Sanpka(23)	cluster_3	71.9	50.3	0.3	80.6
186	Basonda(18)	cluster_3	88.6	46.6	0.4	41.3
188	Joniawas(19)	cluster_3	63.4	39.6	0	53.9
197	Khetawas(118)	cluster_3	89.9	51.7	18.4	78.3
228	Sub-Dist - Manesar	cluster_3	58	37	13.3	58.8
229	Sub-Dist - Manesar	cluster_3	57.6	37.3	11	55.1
230	Sub-Dist - Manesar	cluster_3	60.4	35.7	26.3	79.7
235	Sidhrawali(139)	cluster_3	49.9	40.5	7.5	55.1
240	Udepuri(144)	cluster_3	83.3	0	0	100
243	Binola(147)	cluster_3	96	21.8	45.5	57.4
244	Chandla Dungerwas(148)	cluster_3	68	37.9	6.4	51.2
248	Nawada Fatehpur(112)	cluster_3	66.9	12.7	5.5	95.1

250	Sikanderpur Badha(109)	cluster_3	48.5	26.6	12.7	78.5
251	Lakhnola(110)	cluster_3	88.5	29.2	39.3	68.2
252	Naharpur Kasan (111)	cluster_3	28.8	30	18.5	90.8
254	Khoh(153)	cluster_3	38.5	32.1	27.3	94.6
262	Shikohpur(160)	cluster_3	63.1	27.6	7.3	50.2
265	Manesar (154) (CT)	cluster_3	60.4	35.7	26.3	79.7
266	Manesar (154) (CT) - Ward No.1	cluster_3	60.4	35.7	26.3	79.7
267	Sub-Dist - Sohna	cluster_3	58.5	46.9	14.6	44.3
270	Gual Pahari(77) Part	cluster_3	66.6	44.1	29.8	53.4
271	Balola(78)	cluster_3	88.7	22.6	25.6	46.4
282	Garhi Bazidpur(175)	cluster_3	72.6	43.8	3.9	61.1
296	Rahaka(216)	cluster_3	66.7	0	0	65
314	Lakhuwas(190)	cluster_3	69.8	50.5	2.1	51.6
316	Daula(197)	cluster_3	77.4	41.8	5.3	64.4
319	Khaika(192)	cluster_3	74.4	41.5	0	40.2
334	Sohna (MC) - Ward No.1	cluster_3	40.8	52.6	3.4	59.7
336	Sohna (MC) - Ward No.3	cluster_3	63.8	27.1	44.7	75.9
341	Sohna (MC) - Ward No.8	cluster_3	68.2	3	1	58.3

- **Cluster 4:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG
10	Bahmanwas(10)	cluster_4	32.4	2.8	0	22.5
19	Barheri Rehnwa(261)	cluster_4	43.2	47.9	0	32.6
29	Khor(32)	cluster_4	42.1	51.9	6.3	31.1
37	Balewa(271)	cluster_4	37.3	0	0.9	35.1
43	Bastpur(25)	cluster_4	15.8	58.2	1.9	38
49	Nanukalan(27)	cluster_4	30.3	27.7	7.1	42.3
77	Pataudi (MC) - Ward No.3	cluster_4	32.1	46	3.6	46
85	Pataudi (MC) - Ward No.11	cluster_4	21.5	39.3	7.6	27
108	Sadhrana(45)	cluster_4	39.7	35.2	32.8	41.7
115	Gairatpurbas(158)	cluster_4	14.4	56.7	2	6.2
119	Palra(164)	cluster_4	41.1	46.2	0	52.8
145	Gurgaon (M Corp. + OG) - Ward No.18	cluster_4	36.4	42.7	38.1	42.9

175	Janaula(25)	cluster_4	57.7	44.3	28.7	10.2
185	Alimudinpur(11)	cluster_4	44.2	47.1	10.8	9.9
187	Tripari(20)	cluster_4	43	45.3	0.8	15.6
194	Jhund Sarai Abad(124)	cluster_4	50.2	45.5	0.4	26.7
198	Sayyad Mohamadpur (117)	cluster_4	49.6	35.6	2.6	23.5
203	Garhi Nathekhan(9)	cluster_4	42.2	22.1	0.6	37.7
204	Palri(12)	cluster_4	26.7	56.4	5.2	29.7
211	Mubarikpur(37)	cluster_4	34.3	7.8	1.9	8
241	Langra(145)	cluster_4	41.5	41.9	0.4	9.3
242	Bilaspur(146)	cluster_4	52.5	46.8	3	27.6
256	Nainwal(155)	cluster_4	19.2	42.4	0	30.5
268	Sub-Dist - Sohna	cluster_4	54.2	33.9	4.8	29.1
277	Rojka Gujar(172)	cluster_4	0	0	0	0
278	Behlpa(171)	cluster_4	17.3	18.3	0	35.3
284	Mandawar(221)	cluster_4	20	11.1	0	18.4
285	Tethar(223)	cluster_4	13.3	52	5.3	2.7
286	Kherli Lala(220)	cluster_4	39.3	17.7	2.8	6.6
289	Badshahpur Tethar(222)	cluster_4	19.1	9.6	0.9	14.8
290	Loh Singhani(226)	cluster_4	45.2	36.8	0	14.8
295	Satlaka(217)	cluster_4	7.8	0	0	1.3
297	Nimot (200)	cluster_4	14.3	51.2	4.2	7.7
301	Jolahaka(214)	cluster_4	28.6	46.8	0	10.4
302	Bilaka(211)	cluster_4	17.6	0	0	3.7
303	Bhogpur(208)	cluster_4	9.4	0.7	1.3	12.1
305	Ratika Noabad(210)	cluster_4	0	0	0	48.9
306	Hajipur(15)	cluster_4	49.1	25.6	2.9	25.3
308	Karnki (206)	cluster_4	19.5	59	0	1
310	Harchandpur (199)	cluster_4	17.7	3.3	0.6	9.7
312	Nunera (204)	cluster_4	37.4	15.8	0	16.2
313	Sancholi (205)	cluster_4	25.5	3.6	0.3	8.3
321	Zakupur (189)	cluster_4	29.3	50.7	4.8	52.8
323	Sampki Nagli (186)	cluster_4	42.9	33.1	0	3.4
324	Mohammadpur Gujar (185)	cluster_4	32.2	44.7	0	10.5
325	Dhunela(182)	cluster_4	33.5	36.4	3.3	18
328	Hariahera(183)	cluster_4	8.7	0	2.6	26
329	Raisena(184)	cluster_4	25.1	17.8	12.4	35

- **Cluster 5:**

	District	cluster	Households in good condition	Closed drainage	Water from treated source	LPG/PNG

6	Sub-Dist - Pataudi	cluster_5	63.9	77.8	17.9	61
22	Shahpur Jat(263)	cluster_5	85.5	98.7	1.9	67.9
30	Pahari(265)	cluster_5	71.6	93.4	0	71.3
31	Mozamabad (21)	cluster_5	69.6	89	0	52.7
33	Khetiawas(268)	cluster_5	55.4	86.1	0.9	60.6
42	Said Shahpur(24)	cluster_5	42.9	82.5	0	58.7
53	Mumtajpur(38)	cluster_5	62	98.6	32.2	68.1
54	Turkapur(40)	cluster_5	24.3	91.9	27.2	78
60	Hailey Mandi (MC)	cluster_5	65.8	81.7	23.9	60.7
61	Hailey Mandi (MC) - Ward No.1	cluster_5	76.6	97.9	44	70.8
62	Hailey Mandi (MC) - Ward No.2	cluster_5	73.8	94.2	29.6	62.2
64	Hailey Mandi (MC) - Ward No.4	cluster_5	91.1	100	6.9	97
66	Hailey Mandi (MC) - Ward No.6	cluster_5	82.2	76.5	18.5	90.4
69	Hailey Mandi (MC) - Ward No.9	cluster_5	76.2	98.8	0	66.7
74	Pataudi (MC)	cluster_5	61.9	73.5	11.1	61.3
76	Pataudi (MC) - Ward No.2	cluster_5	95.2	91.7	12.6	87.8
78	Pataudi (MC) - Ward No.4	cluster_5	86.2	100	17.8	68.9
79	Pataudi (MC) - Ward No.5	cluster_5	64.3	99.6	17.1	58.7
80	Pataudi (MC) - Ward No.6	cluster_5	71.1	100	1.1	88.9
81	Pataudi (MC) - Ward No.7	cluster_5	95.5	82.9	1	83.4
82	Pataudi (MC) - Ward No.8	cluster_5	73.3	96.1	14.4	68.8
86	Pataudi (MC) - Ward No.12	cluster_5	77.3	98.9	0	77.8
87	Pataudi (MC) - Ward No.13	cluster_5	76.5	90.8	4.8	79.7
91	Dhana(126)	cluster_5	86.9	77	3.3	53.3
93	Baskushla(127)	cluster_5	89.4	57.3	24.8	80.3
94	Bhang Rola(122)	cluster_5	70	67.3	13.1	64.1
104	Dhankot(49) Part	cluster_5	50.6	91.5	26.3	68
105	Gopalpur(47) Part	cluster_5	90	76.7	6.7	66.7
107	Chandu(44)	cluster_5	64.4	93.8	8.1	67.8
111	Dhorka(120)	cluster_5	93.7	55.8	32	48.7
114	Hayatpur(114)	cluster_5	65.1	62	17.2	69.2
123	Garhi Harsaru (46) (CT)	cluster_5	47.9	67.5	9.6	66.7
124	Garhi Harsaru (46) (CT) - Ward No.1	cluster_5	47.9	67.5	9.6	66.7
125	Badshahpur (87) (CT)	cluster_5	62.7	73.6	18.2	76.8

126	Badshahpur (87) (CT) - Ward No.1	cluster_5	62.7	73.6	18.2	76.8
161	Gurgaon (M Corp. + OG) - Ward No.34	cluster_5	60	63.5	23.4	77.5
166	Sub-Dist - Farrukhnagar	cluster_5	60	89.2	5.5	55.1
169	Mokalwas(132)	cluster_5	74.1	67.9	42.6	67.2
171	Kharkhari(131)	cluster_5	46.6	66.6	20.8	57.5
172	Baslambi(130)	cluster_5	61	65.2	0.4	75.4
176	Jori(24)	cluster_5	80.6	67.3	11.8	48.9
193	Babra Bakipur(33)	cluster_5	73.5	64.2	3.1	60.6
196	Dhanawas(119)	cluster_5	72.8	79.4	1.2	50.6
209	Mushedpur(6)	cluster_5	62.5	95.9	1	51.4
213	Jhanjrola(38)	cluster_5	73.7	83.7	18.6	49.6
216	Farrukhnagar (MC)	cluster_5	60	89.2	5.5	55.1
218	Farrukhnagar (MC) - Ward No.2	cluster_5	49.5	84.5	2.7	68.2
219	Farrukhnagar (MC) - Ward No.3	cluster_5	77.7	99.3	0	69.8
220	Farrukhnagar (MC) - Ward No.4	cluster_5	73.1	99.4	25.7	96.4
221	Farrukhnagar (MC) - Ward No.5	cluster_5	88.4	99	0.5	63.3
222	Farrukhnagar (MC) - Ward No.6	cluster_5	90	63.8	0	75.4
224	Farrukhnagar (MC) - Ward No.8	cluster_5	65.6	99.3	0.4	60.4
225	Farrukhnagar (MC) - Ward No.9	cluster_5	46.2	100	4.3	71.5
246	Fazalwas(149)	cluster_5	68.1	59.6	4.5	57.1
247	Kukrola(151)	cluster_5	72.9	92	9.2	56.7
269	Sub-Dist - Sohna	cluster_5	66.7	71.6	33.5	73.4
315	Baluda(191)	cluster_5	83.3	93.7	53.2	61.1
320	Sohna (Rural)(Part)(187)	cluster_5	35.5	87.3	0.9	60.9
330	Ghamroj(179)	cluster_5	70.6	66.7	11.3	62.5
331	Bhondsi (168) (CT)	cluster_5	72.3	71.9	30.6	81.1
332	Bhondsi (168) (CT) - Ward No.1	cluster_5	72.3	71.9	30.6	81.1
333	Sohna (MC)	cluster_5	64.2	71.4	34.7	69.9
335	Sohna (MC) - Ward No.2	cluster_5	85.2	74.7	44.8	80.2
337	Sohna (MC) -	cluster_5	81.8	60.4	38.7	40.1

	Ward No.4					
338	Sohna (MC) - Ward No.5	cluster_5	94	84.8	32.1	86.3
340	Sohna (MC) - Ward No.7	cluster_5	59.4	74.8	42.8	77
343	Sohna (MC) - Ward No.10	cluster_5	66.8	88.5	20.7	56.4
344	Sohna (MC) - Ward No.11	cluster_5	44.2	96.3	39.5	84.1
346	Sohna (MC) - Ward No.13	cluster_5	49	99.7	46.1	55.5
347	Sohna (MC) - Ward No.14	cluster_5	45.5	62.4	22.2	69
348	Sohna (MC) - Ward No.15	cluster_5	65.9	93.9	11.3	71.5

REFERENCES

- **Our data source :**
  http://www.censusindia.gov.in/2011census/HLO/HL_PCA/Houselisting-housing-HLPCA.html

  We chose **Haryana State** and **Gurgaon District**
  http://www.censusindia.gov.in/2011census/HLO/HL_PCA/Houselisting-housing-HARYANA.html
- We referred tutorial manual available on rapid miner website (https://**rapidminer**.com/wp-content/.../**RapidMiner**-v6-user-manual.pd)
- Video Tutorials available on Youtube (https://www.youtube.com/watch?v=lZho66YQEIM)