

```
In [1]: #= miniproject by:-
```

NAME - KULDEEP GHORPADE

TOPIC - *data preprocessing operations on the dataset
Using Logistic regression, decision tree and Random Forest
algorithm and comparing the accuracy of the model*

Dataset - Lung cancer prediction =#

```
In [1]: using Pkg  
Pkg.add("CSV")
```

Updating registry at `C:\Users\kulde\.julia\registries\General.toml`
Resolving package versions...
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`

```
In [2]: Pkg.add("DataFrames")
```

Resolving package versions...
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`

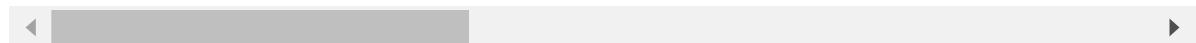
```
In [3]: using CSV
```

```
In [4]: using DataFrames
```

```
In [5]: iris=CSV.read("lungcancer.csv",DataFrame)
```

Out[5]: 1000×26 DataFrame 975 rows omitted

Row	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease
		Int64	String7	Int64	Int64	Int64	Int64	Int64	Int64	Int64
1	0	P1	33	1	2	4	5		4	3
2	1	P10	17	1	3	1	5		3	4
3	2	P100	35	1	4	5	6		5	5
4	3	P1000	37	1	7	7	7		7	6
5	4	P101	46	1	6	8	7		7	7
6	5	P102	35	1	4	5	6		5	5
7	6	P103	52	2	2	4	5		4	3
8	7	P104	28	2	3	1	4		3	2
9	8	P105	35	2	4	5	6		5	6
10	9	P106	46	1	2	3	4		2	4
11	10	P107	44	1	6	7	7		7	7
12	11	P108	64	2	6	8	7		7	7
13	12	P109	39	2	4	5	6		6	5
:	:	:	:	:	:	:	:		:	:
989	988	P989	39	2	6	8	7		7	7
990	989	P99	37	1	7	7	7		7	6
991	990	P990	49	1	6	5	6		5	5
992	991	P991	37	1	8	8	7		7	7
993	992	P992	26	2	7	7	7		7	7
994	993	P993	37	2	7	7	7		7	6
995	994	P994	33	1	6	7	7		7	7
996	995	P995	44	1	6	7	7		7	7
997	996	P996	37	2	6	8	7		7	7
998	997	P997	25	2	4	5	6		5	5
999	998	P998	18	2	6	8	7		7	7
1000	999	P999	47	1	6	5	6		5	5



In [9]: `lungcancer=CSV.read("lungcancer.csv",DataFrame,normalizenames=true)`

Out[9]: 1000×26 DataFrame 975 rows omitted

Row	index	Patient_Id	Age	Gender	Air_Pollution	Alcohol_use	Dust_Allergy	Occupational_I
			Int64	String7	Int64	Int64	Int64	Int64
1	0	P1	33	1	2	4	5	
2	1	P10	17	1	3	1	5	
3	2	P100	35	1	4	5	6	
4	3	P1000	37	1	7	7	7	
5	4	P101	46	1	6	8	7	
6	5	P102	35	1	4	5	6	
7	6	P103	52	2	2	4	5	
8	7	P104	28	2	3	1	4	
9	8	P105	35	2	4	5	6	
10	9	P106	46	1	2	3	4	
11	10	P107	44	1	6	7	7	
12	11	P108	64	2	6	8	7	
13	12	P109	39	2	4	5	6	
:	:	:	:	:	:	:	:	
989	988	P989	39	2	6	8	7	
990	989	P99	37	1	7	7	7	
991	990	P990	49	1	6	5	6	
992	991	P991	37	1	8	8	7	
993	992	P992	26	2	7	7	7	
994	993	P993	37	2	7	7	7	
995	994	P994	33	1	6	7	7	
996	995	P995	44	1	6	7	7	
997	996	P996	37	2	6	8	7	
998	997	P997	25	2	4	5	6	
999	998	P998	18	2	6	8	7	
1000	999	P999	47	1	6	5	6	



In [12]: `typeof(lungcancer)`

Out[12]: DataFrame

```
In [13]: names(lungcancer)
```

```
Out[13]: 26-element Vector[String]:  
  "index"  
  "Patient_Id"  
  "Age"  
  "Gender"  
  "Air_Pollution"  
  "Alcohol_use"  
  "Dust_Allergy"  
  "OccuPational_Hazards"  
  "Genetic_Risk"  
  "chronic_Lung_Disease"  
  "Balanced_Diet"  
  "Obesity"  
  "Smoking"  
  "Passive_Smoker"  
  "Chest_Pain"  
  "Coughing_of_Blood"  
  "Fatigue"  
  "Weight_Loss"  
  "Shortness_of_Breath"  
  "Wheezing"  
  "Swallowing_Difficulty"  
  "Clubbing_of_Finger_Nails"  
  "Frequent_Cold"  
  "Dry_Cough"  
  "Snoring"  
  "Level"
```

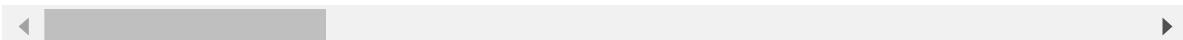
```
In [14]: size(lungcancer)
```

```
Out[14]: (1000, 26)
```

```
In [15]: first(lungcancer,5)
```

```
Out[15]: 5×26 DataFrame
```

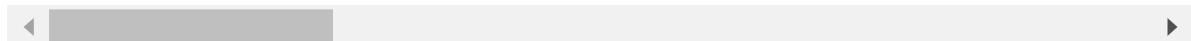
Row	index	Patient_Id	Age	Gender	Air_Pollution	Alcohol_use	Dust_Allergy	OccuPational_Hazards
	Int64	String7	Int64	Int64	Int64	Int64	Int64	Int64
1	0	P1	33	1	2	4	5	
2	1	P10	17	1	3	1	5	
3	2	P100	35	1	4	5	6	
4	3	P1000	37	1	7	7	7	
5	4	P101	46	1	6	8	7	



In [16]: `last(lungcancer,5)`

Out[16]: 5×26 DataFrame

Row	index	Patient_Id	Age	Gender	Air_Pollution	Alcohol_use	Dust_Allergy	Occupational_I
	Int64	String7	Int64	Int64	Int64	Int64	Int64	Int64
1	995	P995	44	1	6	7	7	
2	996	P996	37	2	6	8	7	
3	997	P997	25	2	4	5	6	
4	998	P998	18	2	6	8	7	
5	999	P999	47	1	6	5	6	



In [17]: `describe(lungcancer)`

Out[17]: 26×7 DataFrame

Row	variable	mean	min	median	max	nmissing	eltypes
	Symbol	Union... Any	Union... Any	Union... Any	Int64		DataType
1	index	499.5	0	499.5	999	0	Int64
2	Patient_Id		P1		P999	0	String7
3	Age	37.174	14	36.0	73	0	Int64
4	Gender	1.402	1	1.0	2	0	Int64
5	Air_Pollution	3.84	1	3.0	8	0	Int64
6	Alcohol_use	4.563	1	5.0	8	0	Int64
7	Dust_Allergy	5.165	1	6.0	8	0	Int64
8	OccuPational_Hazards	4.84	1	5.0	8	0	Int64
9	Genetic_Risk	4.58	1	5.0	7	0	Int64
10	chronic_Lung_Disease	4.38	1	4.0	7	0	Int64
11	Balanced_Diet	4.491	1	4.0	7	0	Int64
12	Obesity	4.465	1	4.0	7	0	Int64
13	Smoking	3.948	1	3.0	8	0	Int64
14	Passive_Smoker	4.195	1	4.0	8	0	Int64
15	Chest_Pain	4.438	1	4.0	9	0	Int64
16	Coughing_of_Blood	4.859	1	4.0	9	0	Int64
17	Fatigue	3.856	1	3.0	9	0	Int64
18	Weight_Loss	3.855	1	3.0	8	0	Int64
19	Shortness_of_Breath	4.24	1	4.0	9	0	Int64
20	Wheezing	3.777	1	4.0	8	0	Int64
21	Swallowing_Difficulty	3.746	1	4.0	8	0	Int64
22	Clubbing_of_Finger_Nails	3.923	1	4.0	9	0	Int64
23	Frequent_Cold	3.536	1	3.0	7	0	Int64
24	Dry_Cough	3.853	1	4.0	7	0	Int64
25	Snoring	2.926	1	3.0	7	0	Int64
26	Level		High		Medium	0	String7



In [18]: `lungcancer.Air_Pollution`

Out[18]: 1000-element Vector{Int64}:

```
2
3
4
7
6
4
2
3
4
2
6
6
4
⋮
6
7
6
8
7
6
6
6
4
6
6
```

In [19]: `lungcancer[:,3]`

Out[19]: 1000-element Vector{Int64}:

```
33
17
35
37
46
35
52
28
35
46
44
64
39
⋮
39
37
49
37
26
37
33
44
37
25
18
47
```

In [20]: `lungcancer[:,[1,2,3]]`

Out[20]: 1000×3 DataFrame

975 rows omitted

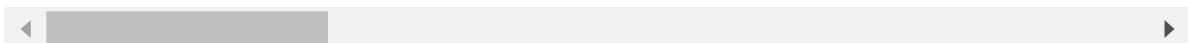
Row	index	Patient_Id	Age
	Int64	String7	Int64
1	0	P1	33
2	1	P10	17
3	2	P100	35
4	3	P1000	37
5	4	P101	46
6	5	P102	35
7	6	P103	52
8	7	P104	28
9	8	P105	35
10	9	P106	46
11	10	P107	44
12	11	P108	64
13	12	P109	39
:	:	:	:
989	988	P989	39
990	989	P99	37
991	990	P990	49
992	991	P991	37
993	992	P992	26
994	993	P993	37
995	994	P994	33
996	995	P995	44
997	996	P996	37
998	997	P997	25
999	998	P998	18
1000	999	P999	47



In [21]: `lungcancer[1:5,:]`

Out[21]: 5×26 DataFrame

Row	index	Patient_Id	Age	Gender	Air_Pollution	Alcohol_use	Dust_Allergy	Occupational_I
		Int64	String7	Int64	Int64	Int64	Int64	Int64
1	0	P1	33	1	2	4	5	
2	1	P10	17	1	3	1	5	
3	2	P100	35	1	4	5	6	
4	3	P1000	37	1	7	7	7	
5	4	P101	46	1	6	8	7	



In [22]: `lungcancer[1:5,1:2]`

Out[22]: 5×2 DataFrame

Row	index	Patient_Id	
		Int64	String7
1	0	P1	
2	1	P10	
3	2	P100	
4	3	P1000	
5	4	P101	



In [23]: `# exp 10`

In [24]: `using Pkg`

In [25]: `Pkg.add("Pandas")`

```
Resolving package versions...
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`
```

In [26]: `using Pandas`

```
WARNING: using Pandas.DataFrame in module Main conflicts with an existing identifier.
WARNING: using Pandas.describe in module Main conflicts with an existing identifier.
```

In [27]: `using Pandas`

```
In [28]: df_train=read_csv("lungcancer.csv")
```

Out[28]:

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
0	0	P1	33	1	2	4	5	4	3	2	
1	1	P10	17	1	3	1	5	3	4	2	
2	2	P100	35	1	4	5	6	5	5	4	
3	3	P1000	37	1	7	7	7	7	6	7	
4	4	P101	46	1	6	8	7	7	7	6	
5	5	P102	35	1	4	5	6	5	5	4	
6	6	P103	52	2	2	4	5	4	3	2	
7	7	P104	28	2	3	1	4	3	2	3	
8	8	P105	35	2	4	5	6	5	6	5	
9	9	P106	46	1	2	3	4	2	4	3	
10	10	P107	44	1	6	7	7	7	7	6	
11	11	P108	64	2	6	8	7	7	7	6	
12	12	P109	39	2	4	5	6	6	5	4	
13	13	P11	34	1	6	7	7	7	6	7	
14	14	P110	27	2	3	1	4	2	3	2	
15	15	P111	73	1	5	6	6	5	6	5	
16	16	P112	17	1	3	1	5	3	4	2	
17	17	P113	34	1	6	7	7	7	6	7	
18	18	P114	36	1	6	7	7	7	7	7	
19	19	P115	14	1	2	4	5	6	5	5	
20	20	P116	24	1	6	8	7	7	6	7	
21	21	P117	53	2	4	5	6	5	5	4	
22	22	P118	62	1	6	8	7	7	7	6	
23	23	P119	29	2	6	7	7	7	7	6	
24	24	P12	36	1	6	7	7	7	7	7	
25	25	P120	65	1	6	8	7	7	7	6	
26	26	P121	38	2	2	1	5	3	2	3	
27	27	P122	19	1	3	2	4	2	3	2	
28	28	P123	33	1	6	7	7	7	7	6	
29	29	P124	28	2	1	6	7	5	3	2	
30	30	P125	35	2	2	6	2	3	6	6	
31	31	P126	42	1	2	4	5	6	5	5	
32	32	P127	32	2	1	6	7	8	7	6	
33	33	P128	33	1	2	4	5	4	3	2	
34	34	P129	25	2	3	1	4	3	2	3	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
35	35	P13	14	1	2	4	5	6	5	5	
36	36	P130	27	2	3	1	4	2	3	2	
37	37	P131	28	1	6	7	8	7	6	7	
38	38	P132	32	1	2	3	6	7	7	7	
39	39	P133	45	2	1	2	4	5	6	5	
40	40	P134	27	2	3	1	4	3	2	3	
41	41	P135	26	2	2	3	1	4	3	2	
42	42	P136	48	1	4	2	3	2	1	2	
43	43	P137	17	2	1	2	3	4	4	3	
44	44	P138	22	1	2	1	3	4	3	5	
45	45	P139	42	1	2	1	2	3	4	3	
46	46	P14	24	1	6	8	7	7	6	7	
47	47	P140	35	1	1	3	2	4	2	6	
48	48	P141	24	2	1	2	2	3	2	4	
49	49	P142	38	2	3	2	3	2	3	2	
50	50	P143	18	2	3	2	1	3	2	1	
51	51	P144	23	2	4	2	3	4	2	3	
52	52	P145	24	2	3	2	2	1	1	1	
53	53	P146	35	2	2	1	2	1	2	1	
54	54	P147	38	2	5	2	3	1	2	3	
55	55	P148	47	2	2	3	1	3	2	5	
56	56	P149	52	2	3	2	1	2	3	5	
57	57	P15	53	2	4	5	6	5	5	4	
58	58	P150	44	2	2	3	2	1	3	2	
59	59	P151	38	2	2	3	5	2	1	1	
60	60	P152	62	2	3	2	1	3	2	4	
61	61	P153	61	2	2	3	4	2	1	1	
62	62	P154	55	1	3	1	1	1	2	3	
63	63	P155	45	2	1	2	3	4	2	4	
64	64	P156	38	1	6	8	7	7	7	6	
65	65	P157	44	1	2	1	5	3	2	3	
66	66	P158	45	2	3	2	4	2	3	2	
67	67	P159	33	2	1	6	7	8	7	6	
68	68	P16	62	1	6	8	7	7	7	6	
69	69	P160	32	2	1	6	7	5	3	2	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
70	70	P161	44	1	1	2	3	4	2	4	
71	71	P162	62	1	6	8	7	7	7	6	
72	72	P163	38	2	2	1	5	3	2	3	
73	73	P164	33	1	3	2	4	2	3	2	
74	74	P165	22	1	6	8	7	7	7	6	
75	75	P166	35	1	2	1	5	3	2	3	
76	76	P167	23	1	3	2	4	2	3	2	
77	77	P168	48	1	1	6	7	8	7	6	
78	78	P169	46	2	1	6	7	5	3	2	
79	79	P17	29	2	6	7	7	7	7	6	
80	80	P170	52	2	1	2	3	4	2	4	
81	81	P171	52	1	6	8	7	7	7	6	
82	82	P172	48	2	6	8	7	7	7	6	
83	83	P173	36	2	2	1	5	3	2	3	
84	84	P174	31	2	3	2	4	2	3	2	
85	85	P175	38	2	1	2	3	4	2	4	
86	86	P176	35	1	6	8	7	7	7	6	
87	87	P177	44	1	6	7	7	7	7	6	
88	88	P178	33	1	6	8	7	7	7	6	
89	89	P179	45	1	6	7	7	7	7	6	
90	90	P18	65	1	6	8	7	7	7	6	
91	91	P180	53	1	6	8	7	7	6	7	
92	92	P181	35	2	4	5	6	5	5	4	
93	93	P182	46	2	6	8	7	7	7	6	
94	94	P183	27	1	6	7	7	7	7	6	
95	95	P184	26	1	6	7	7	7	6	7	
96	96	P185	37	1	6	7	7	7	7	7	
97	97	P186	28	1	6	7	7	7	7	6	
98	98	P187	19	1	6	8	7	7	7	6	
99	99	P188	29	2	4	5	6	5	5	4	
100	100	P189	39	2	6	8	7	7	7	6	
101	101	P19	38	2	2	1	5	3	2	3	
102	102	P190	49	1	6	5	6	5	5	4	
103	103	P191	37	1	8	8	7	7	7	6	
104	104	P192	26	2	7	7	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
105	105	P193	37	2	7	7	7	7	6	7	
106	106	P194	33	1	6	7	7	7	7	7	7
107	107	P195	44	1	6	7	7	7	7	7	6
108	108	P196	37	2	6	8	7	7	7	7	6
109	109	P197	25	2	4	5	6	5	5	5	4
110	110	P198	18	2	6	8	7	7	7	7	6
111	111	P199	47	1	6	5	6	5	5	5	4
112	112	P2	25	2	3	1	4	3	2	3	
113	113	P20	19	1	3	2	4	2	3	2	
114	114	P200	26	2	8	8	7	7	7	7	6
115	115	P201	37	1	7	7	7	7	6	7	
116	116	P202	35	2	4	5	6	5	5	5	4
117	117	P203	33	1	2	4	5	4	3	3	2
118	118	P204	25	2	3	1	4	3	2	3	
119	119	P205	35	2	4	5	6	5	6	5	
120	120	P206	27	2	2	3	4	2	4	3	
121	121	P207	48	1	6	7	7	7	7	7	6
122	122	P208	64	1	6	8	7	7	7	7	6
123	123	P209	39	1	4	5	6	6	5	4	
124	124	P21	33	1	6	7	7	7	7	7	6
125	125	P210	27	2	3	1	4	2	3	2	
126	126	P211	73	1	5	6	6	5	6	5	
127	127	P212	17	1	3	1	5	3	4	2	
128	128	P213	34	1	6	7	7	7	6	7	
129	129	P214	36	1	6	7	7	7	7	7	
130	130	P215	14	1	2	4	5	6	5	5	
131	131	P216	24	1	6	8	7	7	6	7	
132	132	P217	53	2	4	5	6	5	5	4	
133	133	P218	62	1	6	8	7	7	7	7	
134	134	P219	29	2	6	7	7	7	7	7	6
135	135	P22	28	2	1	6	7	5	3	2	
136	136	P220	65	1	6	8	7	7	7	7	6
137	137	P221	38	2	2	1	5	3	2	3	
138	138	P222	19	1	3	2	4	2	3	2	
139	139	P223	33	1	6	7	7	7	7	7	6

index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
140	140	P224	28	2	1	6	7	5	3	2
141	141	P225	35	2	2	6	2	3	6	6
142	142	P226	42	1	2	4	5	6	5	5
143	143	P227	32	2	1	6	7	8	7	6
144	144	P228	33	1	2	4	5	4	3	2
145	145	P229	25	2	3	1	4	3	2	3
146	146	P23	35	2	2	6	2	3	6	6
147	147	P230	27	2	3	1	4	2	3	2
148	148	P231	28	1	6	7	8	7	6	7
149	149	P232	32	1	2	3	6	7	7	7
150	150	P233	45	2	1	2	4	5	6	5
151	151	P234	27	1	3	1	4	3	2	3
152	152	P235	26	1	2	3	1	4	3	2
153	153	P236	48	1	4	2	3	2	1	2
154	154	P237	17	2	1	2	3	4	4	3
155	155	P238	22	2	2	1	3	4	3	5
156	156	P239	42	1	2	1	2	3	4	3
157	157	P24	42	1	2	4	5	6	5	5
158	158	P240	35	1	1	3	2	4	2	6
159	159	P241	24	2	1	2	2	3	2	4
160	160	P242	38	2	3	2	3	2	3	2
161	161	P243	18	2	3	2	1	3	2	1
162	162	P244	23	2	4	2	3	4	2	3
163	163	P245	24	2	3	2	2	1	1	1
164	164	P246	35	2	2	1	2	1	2	1
165	165	P247	38	2	5	2	3	1	2	3
166	166	P248	47	2	2	3	1	3	2	5
167	167	P249	52	1	3	2	1	2	3	5
168	168	P25	32	2	1	6	7	8	7	6
169	169	P250	44	1	2	3	2	1	3	2
170	170	P251	38	1	2	3	5	2	1	1
171	171	P252	62	1	3	2	1	3	2	4
172	172	P253	61	1	2	3	4	2	1	1
173	173	P254	55	1	3	1	1	1	2	3
174	174	P255	45	2	1	2	3	4	2	4

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
175	175	P256	38	2	6	8	7	7	7	6	
176	176	P257	44	1	2	1	5	3	2	3	
177	177	P258	45	1	3	2	4	2	3	2	
178	178	P259	33	1	1	6	7	8	7	6	
179	179	P26	33	1	2	4	5	4	3	2	
180	180	P260	32	1	1	6	7	5	3	2	
181	181	P261	44	1	1	2	3	4	2	4	
182	182	P262	62	1	6	8	7	7	7	6	
183	183	P263	38	2	2	1	5	3	2	3	
184	184	P264	33	1	3	2	4	2	3	2	
185	185	P265	22	1	6	8	7	7	7	6	
186	186	P266	35	1	2	1	5	3	2	3	
187	187	P267	23	1	3	2	4	2	3	2	
188	188	P268	48	2	1	6	7	8	7	6	
189	189	P269	46	2	1	6	7	5	3	2	
190	190	P27	25	2	3	1	4	3	2	3	
191	191	P270	52	2	1	2	3	4	2	4	
192	192	P271	52	1	6	8	7	7	7	6	
193	193	P272	48	2	6	8	7	7	7	6	
194	194	P273	36	2	2	1	5	3	2	3	
195	195	P274	31	2	3	2	4	2	3	2	
196	196	P275	38	2	1	2	3	4	2	4	
197	197	P276	35	1	6	8	7	7	7	6	
198	198	P277	44	1	6	7	7	7	7	6	
199	199	P278	33	1	6	8	7	7	7	6	
200	200	P279	45	1	6	7	7	7	7	6	
201	201	P28	27	2	3	1	4	2	3	2	
202	202	P280	53	1	6	8	7	7	6	7	
203	203	P281	35	2	4	5	6	5	5	4	
204	204	P282	46	1	6	8	7	7	7	6	
205	205	P283	27	1	6	7	7	7	7	6	
206	206	P284	26	1	6	7	7	7	6	7	
207	207	P285	37	1	6	7	7	7	7	7	
208	208	P286	28	1	6	7	7	7	7	6	
209	209	P287	19	1	6	8	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
210	210	P288	29	2	4	5	6	5	5	4	
211	211	P289	39	2	6	8	7	7	7	6	
212	212	P29	28	1	6	7	8	7	6	7	
213	213	P290	49	1	6	5	6	5	5	4	
214	214	P291	37	1	8	8	7	7	7	6	
215	215	P292	26	2	7	7	7	7	7	6	
216	216	P293	37	2	7	7	7	7	6	7	
217	217	P294	33	1	6	7	7	7	7	7	
218	218	P295	44	1	6	7	7	7	7	6	
219	219	P296	37	2	6	8	7	7	7	6	
220	220	P297	25	2	4	5	6	5	5	4	
221	221	P298	18	2	6	8	7	7	7	6	
222	222	P299	47	1	6	5	6	5	5	4	
223	223	P3	35	1	4	5	6	5	6	5	
224	224	P30	32	1	2	3	6	7	7	7	
225	225	P300	26	2	8	8	7	7	7	6	
226	226	P301	37	1	7	7	7	7	6	7	
227	227	P302	35	1	4	5	6	5	5	4	
228	228	P303	33	1	2	4	5	4	3	2	
229	229	P304	25	2	3	1	4	3	2	3	
230	230	P305	35	1	4	5	6	5	6	5	
231	231	P306	27	1	2	3	4	2	4	3	
232	232	P307	48	1	6	7	7	7	7	6	
233	233	P308	64	1	6	8	7	7	7	6	
234	234	P309	39	1	4	5	6	6	5	4	
235	235	P31	45	2	1	2	4	5	6	5	
236	236	P310	27	2	3	1	4	2	3	2	
237	237	P311	73	1	5	6	6	5	6	5	
238	238	P312	17	1	3	1	5	3	4	2	
239	239	P313	34	1	6	7	7	7	6	7	
240	240	P314	36	1	6	7	7	7	7	7	
241	241	P315	14	1	2	4	5	6	5	5	
242	242	P316	24	1	6	8	7	7	6	7	
243	243	P317	53	2	4	5	6	5	5	4	
244	244	P318	62	1	6	8	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
245	245	P319	29	2	6	7	7	7	7	6	
246	246	P32	27	1	3	1	4	3	2	3	
247	247	P320	65	1	6	8	7	7	7	6	
248	248	P321	38	2	2	1	5	3	2	3	
249	249	P322	19	1	3	2	4	2	3	2	
250	250	P323	33	1	6	7	7	7	7	6	
251	251	P324	28	2	1	6	7	5	3	2	
252	252	P325	35	2	2	6	2	3	6	6	
253	253	P326	42	1	2	4	5	6	5	5	
254	254	P327	32	2	1	6	7	8	7	6	
255	255	P328	33	1	2	4	5	4	3	2	
256	256	P329	25	2	3	1	4	3	2	3	
257	257	P33	26	1	2	3	1	4	3	2	
258	258	P330	27	2	3	1	4	2	3	2	
259	259	P331	28	1	6	7	8	7	6	7	
260	260	P332	32	1	2	3	6	7	7	7	
261	261	P333	45	2	1	2	4	5	6	5	
262	262	P334	27	1	3	1	4	3	2	3	
263	263	P335	26	1	2	3	1	4	3	2	
264	264	P336	48	1	4	2	3	2	1	2	
265	265	P337	17	2	1	2	3	4	4	3	
266	266	P338	22	2	2	1	3	4	3	5	
267	267	P339	42	1	2	1	2	3	4	3	
268	268	P34	48	1	4	2	3	2	1	2	
269	269	P340	35	1	1	3	2	4	2	6	
270	270	P341	24	2	1	2	2	3	2	4	
271	271	P342	38	2	3	2	3	2	3	2	
272	272	P343	18	2	3	2	1	3	2	1	
273	273	P344	23	2	4	2	3	4	2	3	
274	274	P345	24	2	3	2	2	1	1	1	
275	275	P346	35	2	2	1	2	1	2	1	
276	276	P347	38	2	5	2	3	1	2	3	
277	277	P348	47	2	2	3	1	3	2	5	
278	278	P349	52	1	3	2	1	2	3	5	
279	279	P35	17	2	1	2	3	4	4	3	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
280	280	P350	44	1	2	3	2	1	3	2	
281	281	P351	38	1	2	3	5	2	1	1	
282	282	P352	62	1	3	2	1	3	2	4	
283	283	P353	61	1	2	3	4	2	1	1	
284	284	P354	55	1	3	1	1	1	2	3	
285	285	P355	45	2	1	2	3	4	2	4	
286	286	P356	38	2	6	8	7	7	7	6	
287	287	P357	44	1	2	1	5	3	2	3	
288	288	P358	45	1	3	2	4	2	3	2	
289	289	P359	33	1	1	6	7	8	7	6	
290	290	P36	22	2	2	1	3	4	3	5	
291	291	P360	32	1	1	6	7	5	3	2	
292	292	P361	44	1	1	2	3	4	2	4	
293	293	P362	62	1	6	8	7	7	7	6	
294	294	P363	38	2	2	1	5	3	2	3	
295	295	P364	33	1	3	2	4	2	3	2	
296	296	P365	22	1	6	8	7	7	7	6	
297	297	P366	35	1	2	1	5	3	2	3	
298	298	P367	23	1	3	2	4	2	3	2	
299	299	P368	48	2	1	6	7	8	7	6	
300	300	P369	46	2	1	6	7	5	3	2	
301	301	P37	42	1	2	1	2	3	4	3	
302	302	P370	52	2	1	2	3	4	2	4	
303	303	P371	52	1	6	8	7	7	7	6	
304	304	P372	48	2	6	8	7	7	7	6	
305	305	P373	36	2	2	1	5	3	2	3	
306	306	P374	31	2	3	2	4	2	3	2	
307	307	P375	38	2	1	2	3	4	2	4	
308	308	P376	35	1	6	8	7	7	7	6	
309	309	P377	44	1	6	7	7	7	7	6	
310	310	P378	33	1	6	8	7	7	7	6	
311	311	P379	45	1	6	7	7	7	7	6	
312	312	P38	35	1	1	3	2	4	2	6	
313	313	P380	53	1	6	8	7	7	6	7	
314	314	P381	35	2	4	5	6	5	5	4	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
315	315	P382	46	1	6	8	7	7	7	6	
316	316	P383	27	1	6	7	7	7	7	6	
317	317	P384	26	1	6	7	7	7	6	7	
318	318	P385	37	1	6	7	7	7	7	7	
319	319	P386	28	1	6	7	7	7	7	6	
320	320	P387	19	1	6	8	7	7	7	6	
321	321	P388	29	2	4	5	6	5	5	4	
322	322	P389	39	2	6	8	7	7	7	6	
323	323	P39	24	2	1	2	2	3	2	4	
324	324	P390	49	1	6	5	6	5	5	4	
325	325	P391	37	1	8	8	7	7	7	6	
326	326	P392	26	2	7	7	7	7	7	6	
327	327	P393	37	2	7	7	7	7	6	7	
328	328	P394	33	1	6	7	7	7	7	7	
329	329	P395	44	1	6	7	7	7	7	6	
330	330	P396	37	2	6	8	7	7	7	6	
331	331	P397	25	2	4	5	6	5	5	4	
332	332	P398	18	2	6	8	7	7	7	6	
333	333	P399	47	1	6	5	6	5	5	4	
334	334	P4	27	1	2	3	4	2	4	3	
335	335	P40	38	2	3	2	3	2	3	2	
336	336	P400	26	2	8	8	7	7	7	6	
337	337	P401	37	1	7	7	7	7	6	7	
338	338	P402	35	1	4	5	6	5	5	4	
339	339	P403	33	1	2	4	5	4	3	2	
340	340	P404	25	2	3	1	4	3	2	3	
341	341	P405	35	1	4	5	6	5	6	5	
342	342	P406	27	1	2	3	4	2	4	3	
343	343	P407	48	1	6	7	7	7	7	6	
344	344	P408	64	1	6	8	7	7	7	6	
345	345	P409	39	1	4	5	6	6	5	4	
346	346	P41	18	2	3	2	1	3	2	1	
347	347	P410	27	2	3	1	4	2	3	2	
348	348	P411	73	1	5	6	6	5	6	5	
349	349	P412	17	1	3	1	5	3	4	2	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
350	350	P413	34	1	6	7	7	7	6	7	
351	351	P414	36	1	6	7	7	7	7	7	7
352	352	P415	14	1	2	4	5	6	5	5	5
353	353	P416	24	1	6	8	7	7	6	7	
354	354	P417	53	2	4	5	6	5	5	5	4
355	355	P418	62	1	6	8	7	7	7	7	6
356	356	P419	29	2	6	7	7	7	7	7	6
357	357	P42	23	2	4	2	3	4	2	3	
358	358	P420	65	1	6	8	7	7	7	7	6
359	359	P421	38	2	2	1	5	3	2	3	
360	360	P422	19	1	3	2	4	2	3	2	
361	361	P423	33	1	6	7	7	7	7	7	6
362	362	P424	28	2	1	6	7	5	3	2	
363	363	P425	35	2	2	6	2	3	6	6	
364	364	P426	42	1	2	4	5	6	5	5	
365	365	P427	32	2	1	6	7	8	7	7	6
366	366	P428	33	1	2	4	5	4	3	2	
367	367	P429	25	2	3	1	4	3	2	3	
368	368	P43	24	2	3	2	2	1	1	1	
369	369	P430	27	2	3	1	4	2	3	2	
370	370	P431	28	1	6	7	8	7	6	7	
371	371	P432	32	1	2	3	6	7	7	7	
372	372	P433	45	1	1	2	4	5	6	5	
373	373	P434	27	1	3	1	4	3	2	3	
374	374	P435	26	1	2	3	1	4	3	2	
375	375	P436	48	1	4	2	3	2	1	2	
376	376	P437	17	2	1	2	3	4	4	4	3
377	377	P438	22	2	2	1	3	4	3	5	
378	378	P439	42	1	2	1	2	3	4	4	3
379	379	P44	35	2	2	1	2	1	2	1	
380	380	P440	35	1	1	3	2	4	2	6	
381	381	P441	24	2	1	2	2	3	2	4	
382	382	P442	25	2	3	2	3	2	3	2	
383	383	P443	38	2	3	2	1	3	2	1	
384	384	P444	43	2	4	2	3	4	2	3	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
385	385	P445	44	2	3	2	2	1	1	1	
386	386	P446	55	2	2	1	2	1	2	1	
387	387	P447	38	2	5	2	3	1	2	3	
388	388	P448	47	2	2	3	1	3	2	5	
389	389	P449	52	1	3	2	1	2	3	5	
390	390	P45	38	2	5	2	3	1	2	3	
391	391	P450	44	1	2	3	2	1	3	2	
392	392	P451	38	1	2	3	5	2	1	1	
393	393	P452	62	1	3	2	1	3	2	4	
394	394	P453	61	1	2	3	4	2	1	1	
395	395	P454	55	1	3	1	1	1	2	3	
396	396	P455	45	2	1	2	3	4	2	4	
397	397	P456	38	2	6	8	7	7	7	6	
398	398	P457	44	1	2	1	5	3	2	3	
399	399	P458	45	1	3	2	4	2	3	2	
400	400	P459	33	1	1	6	7	8	7	6	
401	401	P46	47	2	2	3	1	3	2	5	
402	402	P460	32	2	1	6	7	5	3	2	
403	403	P461	44	1	1	2	3	4	2	4	
404	404	P462	62	1	6	8	7	7	7	6	
405	405	P463	48	2	2	1	5	3	2	3	
406	406	P464	33	1	3	2	4	2	3	2	
407	407	P465	32	1	6	8	7	7	7	6	
408	408	P466	35	1	2	1	5	3	2	3	
409	409	P467	23	1	3	2	4	2	3	2	
410	410	P468	48	2	1	6	7	8	7	6	
411	411	P469	46	2	1	6	7	5	3	2	
412	412	P47	52	1	3	2	1	2	3	5	
413	413	P470	52	2	1	2	3	4	2	4	
414	414	P471	52	1	6	8	7	7	7	6	
415	415	P472	48	2	6	8	7	7	7	6	
416	416	P473	46	2	2	1	5	3	2	3	
417	417	P474	51	2	3	2	4	2	3	2	
418	418	P475	38	1	1	2	3	4	2	4	
419	419	P476	45	1	6	8	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
420	420	P477	54	1	6	7	7	7	7	6	
421	421	P478	33	1	6	8	7	7	7	6	
422	422	P479	45	1	6	7	7	7	7	6	
423	423	P48	44	1	2	3	2	1	3	2	
424	424	P480	63	1	6	8	7	7	6	7	
425	425	P481	45	1	4	5	6	5	5	4	
426	426	P482	46	1	6	8	7	7	7	6	
427	427	P483	37	1	6	7	7	7	7	6	
428	428	P484	36	1	6	7	7	7	6	7	
429	429	P485	37	1	6	7	7	7	7	7	
430	430	P486	28	1	6	7	7	7	7	6	
431	431	P487	19	1	6	8	7	7	7	6	
432	432	P488	39	1	4	5	6	5	5	4	
433	433	P489	39	1	6	8	7	7	7	6	
434	434	P49	38	1	2	3	5	2	1	1	
435	435	P490	49	1	6	5	6	5	5	4	
436	436	P491	37	1	8	8	7	7	7	6	
437	437	P492	36	1	7	7	7	7	7	6	
438	438	P493	37	1	7	7	7	7	6	7	
439	439	P494	33	1	6	7	7	7	7	7	
440	440	P495	44	1	6	7	7	7	7	6	
441	441	P496	37	1	6	8	7	7	7	6	
442	442	P497	25	1	4	5	6	5	5	4	
443	443	P498	28	1	6	8	7	7	7	6	
444	444	P499	47	1	6	5	6	5	5	4	
445	445	P5	48	1	6	7	7	7	7	6	
446	446	P50	62	1	3	2	1	3	2	4	
447	447	P500	26	2	8	8	7	7	7	6	
448	448	P501	37	1	7	7	7	7	6	7	
449	449	P502	35	1	4	5	6	5	5	4	
450	450	P503	33	1	2	4	5	4	3	2	
451	451	P504	25	2	3	1	4	3	2	3	
452	452	P505	35	2	4	5	6	5	6	5	
453	453	P506	27	2	2	3	4	2	4	3	
454	454	P507	48	2	6	7	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	455	455	P508	64	1	6	8	7	7	7	6
	456	456	P509	39	1	4	5	6	6	5	4
	457	457	P51	61	1	2	3	4	2	1	1
	458	458	P510	27	2	3	1	4	2	3	2
	459	459	P511	73	1	5	6	6	5	6	5
	460	460	P512	17	1	3	1	5	3	4	2
	461	461	P513	34	1	6	7	7	7	6	7
	462	462	P514	36	1	6	7	7	7	7	7
	463	463	P515	24	1	2	4	5	6	5	5
	464	464	P516	24	1	6	8	7	7	6	7
	465	465	P517	53	1	4	5	6	5	5	4
	466	466	P518	62	1	6	8	7	7	7	6
	467	467	P519	29	1	6	7	7	7	7	6
	468	468	P52	55	1	3	1	1	1	2	3
	469	469	P520	65	1	6	8	7	7	7	6
	470	470	P521	38	1	2	1	5	3	2	3
	471	471	P522	19	1	3	2	4	2	3	2
	472	472	P523	33	1	6	7	7	7	7	6
	473	473	P524	28	1	1	6	7	5	3	2
	474	474	P525	35	1	2	6	2	3	6	6
	475	475	P526	42	1	2	4	5	6	5	5
	476	476	P527	32	2	1	6	7	8	7	6
	477	477	P528	33	1	2	4	5	4	3	2
	478	478	P529	25	2	3	1	4	3	2	3
	479	479	P53	45	2	1	2	3	4	2	4
	480	480	P530	27	2	3	1	4	2	3	2
	481	481	P531	28	1	6	7	8	7	6	7
	482	482	P532	32	1	2	3	6	7	7	7
	483	483	P533	45	2	1	2	4	5	6	5
	484	484	P534	27	1	3	1	4	3	2	3
	485	485	P535	26	1	2	3	1	4	3	2
	486	486	P536	48	1	4	2	3	2	1	2
	487	487	P537	17	2	1	2	3	4	4	3
	488	488	P538	22	2	2	1	3	4	3	5
	489	489	P539	42	1	2	1	2	3	4	3

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	490	490	P54	38	2	6	8	7	7	7	6
	491	491	P540	35	1	1	3	2	4	2	6
	492	492	P541	24	2	1	2	2	3	2	4
	493	493	P542	38	2	3	2	3	2	3	2
	494	494	P543	18	2	3	2	1	3	2	1
	495	495	P544	23	2	4	2	3	4	2	3
	496	496	P545	24	1	3	2	2	1	1	1
	497	497	P546	35	1	2	1	2	1	2	1
	498	498	P547	38	2	5	2	3	1	2	3
	499	499	P548	47	2	2	3	1	3	2	5
	500	500	P549	52	1	3	2	1	2	3	5
	501	501	P55	44	1	2	1	5	3	2	3
	502	502	P550	44	1	2	3	2	1	3	2
	503	503	P551	38	1	2	3	5	2	1	1
	504	504	P552	62	1	3	2	1	3	2	4
	505	505	P553	61	1	2	3	4	2	1	1
	506	506	P554	55	1	3	1	1	1	2	3
	507	507	P555	45	2	1	2	3	4	2	4
	508	508	P556	38	2	6	8	7	7	7	6
	509	509	P557	44	1	2	1	5	3	2	3
	510	510	P558	45	1	3	2	4	2	3	2
	511	511	P559	33	1	1	6	7	8	7	6
	512	512	P56	45	1	3	2	4	2	3	2
	513	513	P560	32	1	1	6	7	5	3	2
	514	514	P561	44	1	1	2	3	4	2	4
	515	515	P562	62	1	6	8	7	7	7	6
	516	516	P563	38	2	2	1	5	3	2	3
	517	517	P564	33	1	3	2	4	2	3	2
	518	518	P565	22	1	6	8	7	7	7	6
	519	519	P566	35	1	2	1	5	3	2	3
	520	520	P567	23	1	3	2	4	2	3	2
	521	521	P568	48	2	1	6	7	8	7	6
	522	522	P569	46	2	1	6	7	5	3	2
	523	523	P57	33	1	1	6	7	8	7	6
	524	524	P570	52	2	1	2	3	4	2	4

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
525	525	P571	52	1	6	8	7	7	7	6	
526	526	P572	48	2	6	8	7	7	7	6	
527	527	P573	36	2	2	1	5	3	2	3	
528	528	P574	31	2	3	2	4	2	3	2	
529	529	P575	38	2	1	2	3	4	2	4	
530	530	P576	35	1	6	8	7	7	7	6	
531	531	P577	44	1	6	7	7	7	7	6	
532	532	P578	33	1	6	8	7	7	7	6	
533	533	P579	45	1	6	7	7	7	7	6	
534	534	P58	32	1	1	6	7	5	3	2	
535	535	P580	53	1	6	8	7	7	6	7	
536	536	P581	35	2	4	5	6	5	5	4	
537	537	P582	46	1	6	8	7	7	7	6	
538	538	P583	27	1	6	7	7	7	7	6	
539	539	P584	26	1	6	7	7	7	6	7	
540	540	P585	37	1	6	7	7	7	7	7	
541	541	P586	28	1	6	7	7	7	7	6	
542	542	P587	19	1	6	8	7	7	7	6	
543	543	P588	29	2	4	5	6	5	5	4	
544	544	P589	39	2	6	8	7	7	7	6	
545	545	P59	44	1	1	2	3	4	2	4	
546	546	P590	49	1	6	5	6	5	5	4	
547	547	P591	37	1	8	8	7	7	7	6	
548	548	P592	26	2	7	7	7	7	7	6	
549	549	P593	37	2	7	7	7	7	6	7	
550	550	P594	33	1	6	7	7	7	7	7	
551	551	P595	44	1	6	7	7	7	7	6	
552	552	P596	37	2	6	8	7	7	7	6	
553	553	P597	25	2	4	5	6	5	5	4	
554	554	P598	18	2	6	8	7	7	7	6	
555	555	P599	47	1	6	5	6	5	5	4	
556	556	P6	64	1	6	8	7	7	7	6	
557	557	P60	62	1	6	8	7	7	7	6	
558	558	P600	26	2	8	8	7	7	7	6	
559	559	P601	37	1	7	7	7	7	6	7	

index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
560	560	P602	35	1	4	5	6	5	5	4
561	561	P603	33	1	2	4	5	4	3	2
562	562	P604	25	2	3	1	4	3	2	3
563	563	P605	35	1	4	5	6	5	6	5
564	564	P606	27	1	2	3	4	2	4	3
565	565	P607	48	1	6	7	7	7	7	6
566	566	P608	64	1	6	8	7	7	7	6
567	567	P609	39	1	4	5	6	6	5	4
568	568	P61	38	2	2	1	5	3	2	3
569	569	P610	27	2	3	1	4	2	3	2
570	570	P611	73	1	5	6	6	5	6	5
571	571	P612	17	1	3	1	5	3	4	2
572	572	P613	34	1	6	7	7	7	6	7
573	573	P614	36	1	6	7	7	7	7	7
574	574	P615	14	1	2	4	5	6	5	5
575	575	P616	24	1	6	8	7	7	6	7
576	576	P617	53	2	4	5	6	5	5	4
577	577	P618	62	1	6	8	7	7	7	6
578	578	P619	29	2	6	7	7	7	7	6
579	579	P62	33	1	3	2	4	2	3	2
580	580	P620	65	1	6	8	7	7	7	6
581	581	P621	38	2	2	1	5	3	2	3
582	582	P622	19	1	3	2	4	2	3	2
583	583	P623	33	1	6	7	7	7	7	6
584	584	P624	28	2	1	6	7	5	3	2
585	585	P625	35	2	2	6	2	3	6	6
586	586	P626	42	1	2	4	5	6	5	5
587	587	P627	32	2	1	6	7	8	7	6
588	588	P628	33	1	2	4	5	4	3	2
589	589	P629	25	2	3	1	4	3	2	3
590	590	P63	22	1	6	8	7	7	7	6
591	591	P630	27	2	3	1	4	2	3	2
592	592	P631	28	1	6	7	8	7	6	7
593	593	P632	32	1	2	3	6	7	7	7
594	594	P633	45	2	1	2	4	5	6	5

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	595	595	P634	27	1	3	1	4	3	2	3
	596	596	P635	26	1	2	3	1	4	3	2
	597	597	P636	48	1	4	2	3	2	1	2
	598	598	P637	17	2	1	2	3	4	4	3
	599	599	P638	22	2	2	1	3	4	3	5
	600	600	P639	42	1	2	1	2	3	4	3
	601	601	P64	35	1	2	1	5	3	2	3
	602	602	P640	35	1	1	3	2	4	2	6
	603	603	P641	24	2	1	2	2	3	2	4
	604	604	P642	38	2	3	2	3	2	3	2
	605	605	P643	18	2	3	2	1	3	2	1
	606	606	P644	23	2	4	2	3	4	2	3
	607	607	P645	24	2	3	2	2	1	1	1
	608	608	P646	35	2	2	1	2	1	2	1
	609	609	P647	38	2	5	2	3	1	2	3
	610	610	P648	47	2	2	3	1	3	2	5
	611	611	P649	52	1	3	2	1	2	3	5
	612	612	P65	23	1	3	2	4	2	3	2
	613	613	P650	44	1	2	3	2	1	3	2
	614	614	P651	38	1	2	3	5	2	1	1
	615	615	P652	62	1	3	2	1	3	2	4
	616	616	P653	61	1	2	3	4	2	1	1
	617	617	P654	55	1	3	1	1	1	2	3
	618	618	P655	45	2	1	2	3	4	2	4
	619	619	P656	38	2	6	8	7	7	7	6
	620	620	P657	44	1	2	1	5	3	2	3
	621	621	P658	45	1	3	2	4	2	3	2
	622	622	P659	33	1	1	6	7	8	7	6
	623	623	P66	48	2	1	6	7	8	7	6
	624	624	P660	32	1	1	6	7	5	3	2
	625	625	P661	44	1	1	2	3	4	2	4
	626	626	P662	62	1	6	8	7	7	7	6
	627	627	P663	38	2	2	1	5	3	2	3
	628	628	P664	33	1	3	2	4	2	3	2
	629	629	P665	22	1	6	8	7	7	7	6

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	630	630	P666	35	1	2	1	5	3	2	3
	631	631	P667	23	1	3	2	4	2	3	2
	632	632	P668	48	2	1	6	7	8	7	6
	633	633	P669	46	2	1	6	7	5	3	2
	634	634	P67	46	2	1	6	7	5	3	2
	635	635	P670	52	2	1	2	3	4	2	4
	636	636	P671	52	1	6	8	7	7	7	6
	637	637	P672	48	2	6	8	7	7	7	6
	638	638	P673	36	2	2	1	5	3	2	3
	639	639	P674	31	2	3	2	4	2	3	2
	640	640	P675	38	2	1	2	3	4	2	4
	641	641	P676	35	1	6	8	7	7	7	6
	642	642	P677	44	1	6	7	7	7	7	6
	643	643	P678	33	1	6	8	7	7	7	6
	644	644	P679	45	1	6	7	7	7	7	6
	645	645	P68	52	2	1	2	3	4	2	4
	646	646	P680	53	1	6	8	7	7	6	7
	647	647	P681	35	2	4	5	6	5	5	4
	648	648	P682	46	1	6	8	7	7	7	6
	649	649	P683	27	1	6	7	7	7	7	6
	650	650	P684	26	1	6	7	7	7	6	7
	651	651	P685	37	1	6	7	7	7	7	7
	652	652	P686	28	1	6	7	7	7	7	6
	653	653	P687	19	1	6	8	7	7	7	6
	654	654	P688	29	2	4	5	6	5	5	4
	655	655	P689	39	2	6	8	7	7	7	6
	656	656	P69	52	1	6	8	7	7	7	6
	657	657	P690	49	1	6	5	6	5	5	4
	658	658	P691	37	1	8	8	7	7	7	6
	659	659	P692	26	2	7	7	7	7	7	6
	660	660	P693	37	2	7	7	7	7	6	7
	661	661	P694	33	1	6	7	7	7	7	7
	662	662	P695	44	1	6	7	7	7	7	6
	663	663	P696	37	2	6	8	7	7	7	6
	664	664	P697	25	2	4	5	6	5	5	4

index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
665	665	P698	18	2	6	8	7	7	7	6
666	666	P699	47	1	6	5	6	5	5	4
667	667	P7	39	1	4	5	6	6	5	4
668	668	P70	48	2	6	8	7	7	7	6
669	669	P700	26	2	8	8	7	7	7	6
670	670	P701	37	1	7	7	7	7	6	7
671	671	P702	35	1	4	5	6	5	5	4
672	672	P703	33	1	2	4	5	4	3	2
673	673	P704	25	2	3	1	4	3	2	3
674	674	P705	35	1	4	5	6	5	6	5
675	675	P706	27	1	2	3	4	2	4	3
676	676	P707	48	1	6	7	7	7	7	6
677	677	P708	64	1	6	8	7	7	7	6
678	678	P709	39	1	4	5	6	6	5	4
679	679	P71	36	2	2	1	5	3	2	3
680	680	P710	27	2	3	1	4	2	3	2
681	681	P711	73	1	5	6	6	5	6	5
682	682	P712	17	1	3	1	5	3	4	2
683	683	P713	34	1	6	7	7	7	6	7
684	684	P714	36	1	6	7	7	7	7	7
685	685	P715	14	1	2	4	5	6	5	5
686	686	P716	24	1	6	8	7	7	6	7
687	687	P717	53	2	4	5	6	5	5	4
688	688	P718	62	1	6	8	7	7	7	6
689	689	P719	29	2	6	7	7	7	7	6
690	690	P72	31	2	3	2	4	2	3	2
691	691	P720	65	1	6	8	7	7	7	6
692	692	P721	38	2	2	1	5	3	2	3
693	693	P722	19	1	3	2	4	2	3	2
694	694	P723	33	1	6	7	7	7	7	6
695	695	P724	28	2	1	6	7	5	3	2
696	696	P725	35	2	2	6	2	3	6	6
697	697	P726	42	1	2	4	5	6	5	5
698	698	P727	32	2	1	6	7	8	7	6
699	699	P728	33	1	2	4	5	4	3	2

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
700	700	P729	25	2	3	1	4	3	2	3	
701	701	P73	38	2	1	2	3	4	2	4	
702	702	P730	27	2	3	1	4	2	3	2	
703	703	P731	28	1	6	7	8	7	6	7	
704	704	P732	32	1	2	3	6	7	7	7	
705	705	P733	45	2	1	2	4	5	6	5	
706	706	P734	27	1	3	1	4	3	2	3	
707	707	P735	26	1	2	3	1	4	3	2	
708	708	P736	48	1	4	2	3	2	1	2	
709	709	P737	17	2	1	2	3	4	4	3	
710	710	P738	22	2	2	1	3	4	3	5	
711	711	P739	42	1	2	1	2	3	4	3	
712	712	P74	35	1	6	8	7	7	7	6	
713	713	P740	35	1	1	3	2	4	2	6	
714	714	P741	24	2	1	2	2	3	2	4	
715	715	P742	38	2	3	2	3	2	3	2	
716	716	P743	18	2	3	2	1	3	2	1	
717	717	P744	23	2	4	2	3	4	2	3	
718	718	P745	24	2	3	2	2	1	1	1	
719	719	P746	35	2	2	1	2	1	2	1	
720	720	P747	38	2	5	2	3	1	2	3	
721	721	P748	47	2	2	3	1	3	2	5	
722	722	P749	52	1	3	2	1	2	3	5	
723	723	P75	44	1	6	7	7	7	7	6	
724	724	P750	44	1	2	3	2	1	3	2	
725	725	P751	38	1	2	3	5	2	1	1	
726	726	P752	62	1	3	2	1	3	2	4	
727	727	P753	61	1	2	3	4	2	1	1	
728	728	P754	55	1	3	1	1	1	2	3	
729	729	P755	45	2	1	2	3	4	2	4	
730	730	P756	38	2	6	8	7	7	7	6	
731	731	P757	44	1	2	1	5	3	2	3	
732	732	P758	45	1	3	2	4	2	3	2	
733	733	P759	33	1	1	6	7	8	7	6	
734	734	P76	33	1	6	8	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
735	735	P760	32	1	1	6	7	5	3	2	
736	736	P761	44	1	1	2	3	4	2	4	
737	737	P762	62	1	6	8	7	7	7	6	
738	738	P763	38	2	2	1	5	3	2	3	
739	739	P764	33	1	3	2	4	2	3	2	
740	740	P765	22	1	6	8	7	7	7	6	
741	741	P766	35	1	2	1	5	3	2	3	
742	742	P767	23	1	3	2	4	2	3	2	
743	743	P768	48	2	1	6	7	8	7	6	
744	744	P769	46	2	1	6	7	5	3	2	
745	745	P77	45	1	6	7	7	7	7	6	
746	746	P770	52	2	1	2	3	4	2	4	
747	747	P771	52	1	6	8	7	7	7	6	
748	748	P772	48	2	6	8	7	7	7	6	
749	749	P773	36	2	2	1	5	3	2	3	
750	750	P774	31	2	3	2	4	2	3	2	
751	751	P775	38	2	1	2	3	4	2	4	
752	752	P776	35	1	6	8	7	7	7	6	
753	753	P777	44	1	6	7	7	7	7	6	
754	754	P778	33	1	6	8	7	7	7	6	
755	755	P779	45	1	6	7	7	7	7	6	
756	756	P78	53	1	6	8	7	7	6	7	
757	757	P780	53	1	6	8	7	7	6	7	
758	758	P781	35	2	4	5	6	5	5	4	
759	759	P782	46	1	6	8	7	7	7	6	
760	760	P783	27	1	6	7	7	7	7	6	
761	761	P784	26	1	6	7	7	7	6	7	
762	762	P785	37	1	6	7	7	7	7	7	
763	763	P786	28	1	6	7	7	7	7	6	
764	764	P787	19	1	6	8	7	7	7	6	
765	765	P788	29	2	4	5	6	5	5	4	
766	766	P789	39	2	6	8	7	7	7	6	
767	767	P79	35	2	4	5	6	5	5	4	
768	768	P790	49	1	6	5	6	5	5	4	
769	769	P791	37	1	8	8	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
770	770	P792	26	2	7	7	7	7	7	6	
771	771	P793	37	2	7	7	7	7	6	7	
772	772	P794	33	1	6	7	7	7	7	7	
773	773	P795	44	1	6	7	7	7	7	6	
774	774	P796	37	2	6	8	7	7	7	6	
775	775	P797	25	2	4	5	6	5	5	4	
776	776	P798	18	2	6	8	7	7	7	6	
777	777	P799	47	1	6	5	6	5	5	4	
778	778	P8	27	2	3	1	4	2	3	2	
779	779	P80	46	1	6	8	7	7	7	6	
780	780	P800	26	2	8	8	7	7	7	6	
781	781	P801	37	1	7	7	7	7	6	7	
782	782	P802	35	1	4	5	6	5	5	4	
783	783	P803	33	1	2	4	5	4	3	2	
784	784	P804	25	2	3	1	4	3	2	3	
785	785	P805	35	1	4	5	6	5	6	5	
786	786	P806	27	1	2	3	4	2	4	3	
787	787	P807	48	1	6	7	7	7	7	6	
788	788	P808	64	1	6	8	7	7	7	6	
789	789	P809	39	1	4	5	6	6	5	4	
790	790	P810	27	1	6	7	7	7	7	6	
791	791	P810	27	2	3	1	4	2	3	2	
792	792	P811	73	1	5	6	6	5	6	5	
793	793	P812	17	1	3	1	5	3	4	2	
794	794	P813	34	1	6	7	7	7	6	7	
795	795	P814	36	1	6	7	7	7	7	7	
796	796	P815	14	1	2	4	5	6	5	5	
797	797	P816	24	1	6	8	7	7	6	7	
798	798	P817	53	2	4	5	6	5	5	4	
799	799	P818	62	1	6	8	7	7	7	6	
800	800	P819	29	2	6	7	7	7	7	6	
801	801	P820	26	1	6	7	7	7	6	7	
802	802	P820	65	1	6	8	7	7	7	6	
803	803	P821	38	2	2	1	5	3	2	3	
804	804	P822	19	1	3	2	4	2	3	2	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	805	805	P823	33	1	6	7	7	7	7	6
	806	806	P824	28	2	1	6	7	5	3	2
	807	807	P825	35	2	2	6	2	3	6	6
	808	808	P826	42	1	2	4	5	6	5	5
	809	809	P827	32	2	1	6	7	8	7	6
	810	810	P828	33	1	2	4	5	4	3	2
	811	811	P829	25	2	3	1	4	3	2	3
	812	812	P83	37	1	6	7	7	7	7	7
	813	813	P830	27	2	3	1	4	2	3	2
	814	814	P831	28	1	6	7	8	7	6	7
	815	815	P832	32	1	2	3	6	7	7	7
	816	816	P833	45	2	1	2	4	5	6	5
	817	817	P834	27	1	3	1	4	3	2	3
	818	818	P835	26	1	2	3	1	4	3	2
	819	819	P836	48	1	4	2	3	2	1	2
	820	820	P837	17	2	1	2	3	4	4	3
	821	821	P838	22	2	2	1	3	4	3	5
	822	822	P839	42	1	2	1	2	3	4	3
	823	823	P84	28	1	6	7	7	7	7	6
	824	824	P840	35	1	1	3	2	4	2	6
	825	825	P841	24	2	1	2	2	3	2	4
	826	826	P842	38	2	3	2	3	2	3	2
	827	827	P843	18	2	3	2	1	3	2	1
	828	828	P844	23	2	4	2	3	4	2	3
	829	829	P845	24	2	3	2	2	1	1	1
	830	830	P846	35	2	2	1	2	1	2	1
	831	831	P847	38	2	5	2	3	1	2	3
	832	832	P848	47	2	2	3	1	3	2	5
	833	833	P849	52	1	3	2	1	2	3	5
	834	834	P85	19	1	6	8	7	7	7	6
	835	835	P850	44	1	2	3	2	1	3	2
	836	836	P851	38	1	2	3	5	2	1	1
	837	837	P852	62	1	3	2	1	3	2	4
	838	838	P853	61	1	2	3	4	2	1	1
	839	839	P854	55	1	3	1	1	1	2	3

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
840	840	P855	45	2	1	2	3	4	2	4	
841	841	P856	38	2	6	8	7	7	7	6	
842	842	P857	44	1	2	1	5	3	2	3	
843	843	P858	45	1	3	2	4	2	3	2	
844	844	P859	33	1	1	6	7	8	7	6	
845	845	P86	29	2	4	5	6	5	5	4	
846	846	P860	32	1	1	6	7	5	3	2	
847	847	P861	44	1	1	2	3	4	2	4	
848	848	P862	62	1	6	8	7	7	7	6	
849	849	P863	38	2	2	1	5	3	2	3	
850	850	P864	33	1	3	2	4	2	3	2	
851	851	P865	22	1	6	8	7	7	7	6	
852	852	P866	35	1	2	1	5	3	2	3	
853	853	P867	23	1	3	2	4	2	3	2	
854	854	P868	48	2	1	6	7	8	7	6	
855	855	P869	46	2	1	6	7	5	3	2	
856	856	P87	39	2	6	8	7	7	7	6	
857	857	P870	52	2	1	2	3	4	2	4	
858	858	P871	52	1	6	8	7	7	7	6	
859	859	P872	48	2	6	8	7	7	7	6	
860	860	P873	36	2	2	1	5	3	2	3	
861	861	P874	31	2	3	2	4	2	3	2	
862	862	P875	38	2	1	2	3	4	2	4	
863	863	P876	35	1	6	8	7	7	7	6	
864	864	P877	44	1	6	7	7	7	7	6	
865	865	P878	33	1	6	8	7	7	7	6	
866	866	P879	45	1	6	7	7	7	7	6	
867	867	P88	49	1	6	5	6	5	5	4	
868	868	P880	53	1	6	8	7	7	6	7	
869	869	P881	35	2	4	5	6	5	5	4	
870	870	P882	46	1	6	8	7	7	7	6	
871	871	P883	27	1	6	7	7	7	7	6	
872	872	P884	26	1	6	7	7	7	6	7	
873	873	P885	37	1	6	7	7	7	7	7	
874	874	P886	28	1	6	7	7	7	7	6	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
875	875	P887	19	1	6	8	7	7	7	6	
876	876	P888	29	2	4	5	6	5	5	4	
877	877	P889	39	2	6	8	7	7	7	6	
878	878	P89	37	1	8	8	7	7	7	6	
879	879	P890	49	1	6	5	6	5	5	4	
880	880	P891	37	1	8	8	7	7	7	6	
881	881	P892	26	2	7	7	7	7	7	6	
882	882	P893	37	2	7	7	7	7	6	7	
883	883	P894	33	1	6	7	7	7	7	7	
884	884	P895	44	1	6	7	7	7	7	6	
885	885	P896	37	2	6	8	7	7	7	6	
886	886	P897	25	2	4	5	6	5	5	4	
887	887	P898	18	2	6	8	7	7	7	6	
888	888	P899	47	1	6	5	6	5	5	4	
889	889	P9	73	1	5	6	6	5	6	5	
890	890	P90	26	2	7	7	7	7	7	6	
891	891	P900	26	2	8	8	7	7	7	6	
892	892	P901	37	1	7	7	7	7	6	7	
893	893	P902	35	2	4	5	6	5	5	4	
894	894	P903	33	1	2	4	5	4	3	2	
895	895	P904	25	2	3	1	4	3	2	3	
896	896	P905	35	1	4	5	6	5	6	5	
897	897	P906	27	1	2	3	4	2	4	3	
898	898	P907	48	1	6	7	7	7	7	6	
899	899	P908	64	1	6	8	7	7	7	6	
900	900	P909	39	1	4	5	6	6	5	4	
901	901	P91	37	2	7	7	7	7	6	7	
902	902	P910	27	2	3	1	4	2	3	2	
903	903	P911	73	1	5	6	6	5	6	5	
904	904	P912	17	1	3	1	5	3	4	2	
905	905	P913	34	1	6	7	7	7	6	7	
906	906	P914	36	1	6	7	7	7	7	7	
907	907	P915	14	1	2	4	5	6	5	5	
908	908	P916	24	1	6	8	7	7	6	7	
909	909	P917	53	2	4	5	6	5	5	4	

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	910	910	P918	62	1	6	8	7	7	7	6
	911	911	P919	29	2	6	7	7	7	7	6
	912	912	P92	33	1	6	7	7	7	7	7
	913	913	P920	65	1	6	8	7	7	7	6
	914	914	P921	38	2	2	1	5	3	2	3
	915	915	P922	19	1	3	2	4	2	3	2
	916	916	P923	33	1	6	7	7	7	7	6
	917	917	P924	28	2	1	6	7	5	3	2
	918	918	P925	35	2	2	6	2	3	6	6
	919	919	P926	42	1	2	4	5	6	5	5
	920	920	P927	32	2	1	6	7	8	7	6
	921	921	P928	33	1	2	4	5	4	3	2
	922	922	P929	25	2	3	1	4	3	2	3
	923	923	P93	44	1	6	7	7	7	7	6
	924	924	P930	27	2	3	1	4	2	3	2
	925	925	P931	28	1	6	7	8	7	6	7
	926	926	P932	32	1	2	3	6	7	7	7
	927	927	P933	45	2	1	2	4	5	6	5
	928	928	P934	27	1	3	1	4	3	2	3
	929	929	P935	26	1	2	3	1	4	3	2
	930	930	P936	48	1	4	2	3	2	1	2
	931	931	P937	17	2	1	2	3	4	4	3
	932	932	P938	22	2	2	1	3	4	3	5
	933	933	P939	42	1	2	1	2	3	4	3
	934	934	P94	37	2	6	8	7	7	7	6
	935	935	P940	35	1	1	3	2	4	2	6
	936	936	P941	24	2	1	2	2	3	2	4
	937	937	P942	38	2	3	2	3	2	3	2
	938	938	P943	18	2	3	2	1	3	2	1
	939	939	P944	23	2	4	2	3	4	2	3
	940	940	P945	24	2	3	2	2	1	1	1
	941	941	P946	35	2	2	1	2	1	2	1
	942	942	P947	38	2	5	2	3	1	2	3
	943	943	P948	47	2	2	3	1	3	2	5
	944	944	P949	52	1	3	2	1	2	3	5

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
	945	945	P95	25	2	4	5	6	5	5	4
	946	946	P950	44	1	2	3	2	1	3	2
	947	947	P951	38	1	2	3	5	2	1	1
	948	948	P952	62	1	3	2	1	3	2	4
	949	949	P953	61	1	2	3	4	2	1	1
	950	950	P954	55	1	3	1	1	1	2	3
	951	951	P955	45	2	1	2	3	4	2	4
	952	952	P956	38	2	6	8	7	7	7	6
	953	953	P957	44	1	2	1	5	3	2	3
	954	954	P958	45	1	3	2	4	2	3	2
	955	955	P959	33	1	1	6	7	8	7	6
	956	956	P96	18	2	6	8	7	7	7	6
	957	957	P960	32	1	1	6	7	5	3	2
	958	958	P961	44	1	1	2	3	4	2	4
	959	959	P962	62	1	6	8	7	7	7	6
	960	960	P963	38	2	2	1	5	3	2	3
	961	961	P964	33	1	3	2	4	2	3	2
	962	962	P965	22	1	6	8	7	7	7	6
	963	963	P966	35	1	2	1	5	3	2	3
	964	964	P967	23	1	3	2	4	2	3	2
	965	965	P968	48	2	1	6	7	8	7	6
	966	966	P969	46	2	1	6	7	5	3	2
	967	967	P97	47	1	6	5	6	5	5	4
	968	968	P970	52	2	1	2	3	4	2	4
	969	969	P971	52	1	6	8	7	7	7	6
	970	970	P972	48	2	6	8	7	7	7	6
	971	971	P973	36	2	2	1	5	3	2	3
	972	972	P974	31	2	3	2	4	2	3	2
	973	973	P975	38	2	1	2	3	4	2	4
	974	974	P976	35	1	6	8	7	7	7	6
	975	975	P977	44	1	6	7	7	7	7	6
	976	976	P978	33	1	2	4	5	4	3	2
	977	977	P979	45	1	3	1	4	3	2	3
	978	978	P98	26	2	8	8	7	7	7	6
	979	979	P980	53	1	3	1	4	2	3	2

index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lung Disease	I
980	980	P981	35	2	4	5	6	5	5	4
981	981	P982	46	1	6	8	7	7	7	6
982	982	P983	27	1	6	7	7	7	7	6
983	983	P984	26	1	3	2	4	2	3	2
984	984	P985	37	1	1	2	3	4	2	4
985	985	P986	28	1	6	7	7	7	7	6
986	986	P987	19	1	6	8	7	7	7	6
987	987	P988	29	2	4	5	6	5	5	4
988	988	P989	39	2	6	8	7	7	7	6
989	989	P99	37	1	7	7	7	7	6	7
990	990	P990	49	1	6	5	6	5	5	4
991	991	P991	37	1	8	8	7	7	7	6
992	992	P992	26	2	7	7	7	7	7	6
993	993	P993	37	2	7	7	7	7	6	7
994	994	P994	33	1	6	7	7	7	7	7
995	995	P995	44	1	6	7	7	7	7	6
996	996	P996	37	2	6	8	7	7	7	6
997	997	P997	25	2	4	5	6	5	5	4
998	998	P998	18	2	6	8	7	7	7	6
999	999	P999	47	1	6	5	6	5	5	4

```
In [29]: x=df_train[["Air Pollution","Alcohol use","Dust Allergy","Genetic Risk"]]
y = df_train["Level"]
```

```
Out[29]: 0 Low 1 Medium 2 High 3 High 4 High ... 995 High 996 High 997 High 998 High 999 High
Name: Level, Length: 1000, dtype: object
```

```
In [30]: Pkg.add("ScikitLearn")
Pkg.update()
```

```
Resolving package versions...
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`
  Updating registry at `C:\Users\kulde\.julia\registries\General.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
No Changes to `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`
```

In [31]: `using Pkg`

```
Pkg.add("sklearn")
```

The following package names could not be resolved:

- * sklearn (not found in project, manifest or registry)
- Suggestions: ScikitLearn ScikitLearnBase MLJScikitLearn

Stacktrace:

```
[1] pkgerror(msg::String)
    @ Pkg.Types C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\Types.jl:67
[2] ensure_resolved(ctx::Pkg.Types.Context, manifest::Pkg.Types.Manifest, pkgs::Vector{Pkg.Types.PackageSpec}; registry::Bool)
    @ Pkg.Types C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\Types.jl:952
[3] add(ctx::Pkg.Types.Context, pkgs::Vector{Pkg.Types.PackageSpec}; preserve::Pkg.Types.PreserveLevel, platform::Base.BinaryPlatforms.Platform, kwargs::Base.Pairs{Symbol, IJulia.IJuliaStdio{Base.PipeEndpoint}, Tuple{Symbol}}, NamedTuple{(:io,), Tuple{IJulia.IJuliaStdio{Base.PipeEndpoint}}})
    @ Pkg.API C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:264
[4] add(pkgs::Vector{Pkg.Types.PackageSpec}; io::IJulia.IJuliaStdio{Base.PipeEndpoint}, kwargs::Base.Pairs{Symbol, Union{}, Tuple{}}, NamedTuple{(), Tuple{}})
    @ Pkg.API C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:156
[5] add(pkgs::Vector{Pkg.Types.PackageSpec})
    @ Pkg.API C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:145
[6] #add#27
    @ C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:144 [inlined]
[7] add
    @ C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:144 [inlined]
[8] #add#26
    @ C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:143 [inlined]
[9] add(pkg::String)
    @ Pkg.API C:\Users\kulde\AppData\Local\Programs\Julia-1.8.3\share\julia\stdlib\v1.8\Pkg\src\API.jl:143
[10] top-level scope
    @ In[31]:2
[11] eval
    @ .\boot.jl:368 [inlined]
[12] include_string(mapexpr::typeof(REPL.softscope), mod::Module, code::String, filename::String)
    @ Base .\loading.jl:1428
```

In [32]: `using ScikitLearn: fit!, predict, @sk_import, fit_transform!`

```
In [33]: @sk_import model_selection: train_test_split
```

```
↳ Warning: Module model_selection has been ported to Julia - try `import ScikitLearn: CrossValidation` instead
└ @ ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\Skcore.jl:193
```

```
Out[33]: PyObject <function train_test_split at 0x000002398CEE6B90>
```

In [34]: `x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)`

Out[34]: 4-element Vector{PyCall.PyObject}:

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
541	6	7	7	7
440	6	7	7	7
482	2	3	6	7
422	6	7	7	7
778	3	1	4	3
..
106	6	7	7	7
270	1	2	2	2
860	2	1	5	2
435	6	5	6	5
102	6	5	6	5

[700 rows x 4 columns]

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
521	1	6	7	7
737	6	8	7	7
740	6	8	7	7
660	7	7	7	6
411	1	6	7	3
..
468	3	1	1	2
935	1	3	2	2
428	6	7	7	6
7	3	1	4	2
155	2	1	3	3

[300 rows x 4 columns]

PyObject	541	High
440	High	
482	Low	
422	High	
778	Low	
..	...	
106	High	
270	Low	
860	Medium	
435	High	
102	High	

Name: Level, Length: 700, dtype: object

PyObject	521	Medium
737	Medium	
740	Medium	
660	High	
411	Medium	
..	...	
468	Low	
935	Low	
428	High	
7	Low	
155	Low	

Name: Level, Length: 300, dtype: object

```
In [35]: @sk_import linear_model: LogisticRegression  
model=LogisticRegression()  
model.fit(x_train,y_train)
```

Out[35]: LogisticRegression()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [36]: println("Predicted Accuracy of the Model : ",model.score(x_test,y_test))  
println("Actual Accuracy of the Model : ",model.score(x_train,y_train))
```

Predicted Accuracy of the Model : 0.7
Actual Accuracy of the Model : 0.6957142857142857

```
In [37]: @sk_import model_selection: cross_val_score
```

↳ Warning: Module model_selection has been ported to Julia - try `import ScikitLearn: CrossValidation` instead
↳ @ ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\S
kcore.jl:193

Out[37]: PyObject <function cross_val_score at 0x000002398CFC6EF0>

```
In [38]: #5 fold cross validation  
cross_score = cross_val_score(model, x, y, cv=5)  
#print cross_val_score  
println("cross_validation_score: ", mean(cross_score))
```

cross_validation_score: 0.69

```
In [39]: # exp 11
```

```
In [40]: using Pandas
```

In [41]: `df_train=read_csv("lungcancer.csv")`

Out[41]:

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lun Diseases
0	0	P1	33	1	2	4	5	4	3	
1	1	P10	17	1	3	1	5	3	4	
2	2	P100	35	1	4	5	6	5	5	
3	3	P1000	37	1	7	7	7	7	6	
4	4	P101	46	1	6	8	7	7	7	
5	5	P102	35	1	4	5	6	5	5	
6	6	P103	52	2	2	4	5	4	3	
7	7	P104	28	2	3	1	4	3	2	
8	8	P105	35	2	4	5	6	5	6	
9	9	P106	46	1	2	3	4	2	4	

In [42]: `x=df_train[["Air Pollution","Alcohol use","Dust Allergy","Genetic Risk"]]
y = df_train["Level"]`

Out[42]: 0 Low 1 Medium 2 High 3 High 4 High ... 995 High 996 High 997 High 998 High 999 High
Name: Level, Length: 1000, dtype: object

In [43]: `using ScikitLearn: fit!, predict, @sk_import, fit_transform!`

In [44]: `@sk_import model_selection: train_test_split`

```
r Warning: Module model_selection has been ported to Julia - try `import ScikitLearn: CrossValidation` instead
└ @ ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\Skcore.jl:193
WARNING: redefinition of constant train_test_split. This may fail, cause incorrect answers, or produce other errors.
```

Out[44]: PyObject <function train_test_split at 0x000002398CEE6B90>

In [45]: `x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)`

Out[45]: 4-element Vector{PyCall.PyObject}:

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
541	6	7	7	7
440	6	7	7	7
482	2	3	6	7
422	6	7	7	7
778	3	1	4	3
..
106	6	7	7	7
270	1	2	2	2
860	2	1	5	2
435	6	5	6	5
102	6	5	6	5

[700 rows x 4 columns]

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
521	1	6	7	7
737	6	8	7	7
740	6	8	7	7
660	7	7	7	6
411	1	6	7	3
..
468	3	1	1	2
935	1	3	2	2
428	6	7	7	6
7	3	1	4	2
155	2	1	3	3

[300 rows x 4 columns]

PyObject	541	High
440	High	
482	Low	
422	High	
778	Low	
..	...	
106	High	
270	Low	
860	Medium	
435	High	
102	High	

Name: Level, Length: 700, dtype: object

PyObject	521	Medium
737	Medium	
740	Medium	
660	High	
411	Medium	
..	...	
468	Low	
935	Low	
428	High	
7	Low	
155	Low	

Name: Level, Length: 300, dtype: object

```
In [46]: @sk_import tree: DecisionTreeClassifier
model=DecisionTreeClassifier()
model.fit(x_train,y_train)
println("Predicted Accuracy of the Model : ",model.score(x_test,y_test))
println("Actual Accuracy of the Model : ",model.score(x_train,y_train))
```

Predicted Accuracy of the Model : 0.9066666666666666
 Actual Accuracy of the Model : 0.9114285714285715

```
In [47]: @sk_import model_selection: cross_val_score
```

↳ Warning: Module `model_selection` has been ported to Julia - try ``import ScikitLearn: CrossValidation`` instead
 ↳ @ `ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\Skcore.jl:193`
 WARNING: redefinition of constant `cross_val_score`. This may fail, cause incorrect answers, or produce other errors.

```
Out[47]: PyObject <function cross_val_score at 0x000002398CFC6EF0>
```

```
In [48]: #5 fold cross validation
cross_score = cross_val_score(model, x, y, cv=5)
#print cross_val_score
println("cross_validation_score: ", mean(cross_score))
```

cross_validation_score: 0.9100000000000001

```
In [49]: # exp 12
```

```
In [50]: using Pandas
```

```
In [51]: df_train=read_csv("lungcancer.csv")
```

```
Out[51]:
```

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	chronic Lun Diseases
0	0	P1	33	1	2	4	5	4	3	
1	1	P10	17	1	3	1	5	3	4	
2	2	P100	35	1	4	5	6	5	5	
3	3	P1000	37	1	7	7	7	7	6	
4	4	P101	46	1	6	8	7	7	7	
5	5	P102	35	1	4	5	6	5	5	
6	6	P103	52	2	2	4	5	4	3	
7	7	P104	28	2	3	1	4	3	2	
8	8	P105	35	2	4	5	6	5	6	
9	9	P106	46	1	2	3	4	2	4	

```
In [52]: x=df_train[["Air Pollution","Alcohol use","Dust Allergy","Genetic Risk"]]
y = df_train["Level"]
```

```
Out[52]: 0 Low 1 Medium 2 High 3 High 4 High ... 995 High 996 High 997 High 998 High 999 High
Name: Level, Length: 1000, dtype: object
```

```
In [53]: using ScikitLearn: fit!, predict, @sk_import, fit_transform!
```

```
In [54]: @sk_import model_selection: train_test_split
```

```
↳ Warning: Module model_selection has been ported to Julia - try `import ScikitLearn: CrossValidation` instead
↳ @ ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\S
kcore.jl:193
WARNING: redefinition of constant train_test_split. This may fail, cause inco
rrect answers, or produce other errors.
```

```
Out[54]: PyObject <function train_test_split at 0x000002398CEE6B90>
```

In [56]: `x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)`

Out[56]: 4-element Vector{PyCall.PyObject}:

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
541	6	7	7	7
440	6	7	7	7
482	2	3	6	7
422	6	7	7	7
778	3	1	4	3
..
106	6	7	7	7
270	1	2	2	2
860	2	1	5	2
435	6	5	6	5
102	6	5	6	5

[700 rows x 4 columns]

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
521	1	6	7	7
737	6	8	7	7
740	6	8	7	7
660	7	7	7	6
411	1	6	7	3
..
468	3	1	1	2
935	1	3	2	2
428	6	7	7	6
7	3	1	4	2
155	2	1	3	3

[300 rows x 4 columns]

PyObject	541	High
440	High	
482	Low	
422	High	
778	Low	
..	...	
106	High	
270	Low	
860	Medium	
435	High	
102	High	

Name: Level, Length: 700, dtype: object

PyObject	521	Medium
737	Medium	
740	Medium	
660	High	
411	Medium	
..	...	
468	Low	
935	Low	
428	High	
7	Low	
155	Low	

Name: Level, Length: 300, dtype: object

In [57]: `x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)`

Out[57]: 4-element Vector{PyCall.PyObject}:

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
541	6	7	7	7
440	6	7	7	7
482	2	3	6	7
422	6	7	7	7
778	3	1	4	3
..
106	6	7	7	7
270	1	2	2	2
860	2	1	5	2
435	6	5	6	5
102	6	5	6	5

[700 rows x 4 columns]

PyObject	Air Pollution	Alcohol use	Dust Allergy	Genetic Risk
521	1	6	7	7
737	6	8	7	7
740	6	8	7	7
660	7	7	7	6
411	1	6	7	3
..
468	3	1	1	2
935	1	3	2	2
428	6	7	7	6
7	3	1	4	2
155	2	1	3	3

[300 rows x 4 columns]

PyObject	541	High
440	High	
482	Low	
422	High	
778	Low	
..	...	
106	High	
270	Low	
860	Medium	
435	High	
102	High	

Name: Level, Length: 700, dtype: object

PyObject	521	Medium
737	Medium	
740	Medium	
660	High	
411	Medium	
..	...	
468	Low	
935	Low	
428	High	
7	Low	
155	Low	

Name: Level, Length: 300, dtype: object

```
In [58]: @sk_import model_selection: cross_val_score
```

```
↳ Warning: Module model_selection has been ported to Julia - try `import ScikitLearn: CrossValidation` instead
└ @ ScikitLearn.Skcore C:\Users\kulde\.julia\packages\ScikitLearn\7Csem\src\Skcore.jl:193
WARNING: redefinition of constant cross_val_score. This may fail, cause incorrect answers, or produce other errors.
```

```
Out[58]: PyObject <function cross_val_score at 0x000002398CFC6EF0>
```

```
In [59]: #5 fold cross validation
cross_score = cross_val_score(model, x, y, cv=5)
#print cross_val_score
println("cross_validation_score: ", mean(cross_score))
```

```
cross_validation_score: 0.9100000000000001
```

```
In [ ]:
```