Analysis on Air Quality Detecting IOT product

By

Christopher Immanuel Prince P J

Short note on Data Analysis :

No food waste is a company which serves food to the people who needs from the place where surplus food is available. They had designed an Air quality testing IOT product and they placed it in and around coimbatore and they were tracking the live data .

Here, I was provided with the dataset and asked to provide insights.

Attributes: Temperature, Humidity, Heat index, Air pollution, carbon mono oxide and noise pollution.

Overall Insights:

- While cleaning the data, it is found that there exist some places where the product is not working. Further information are provided under “Certain Places out of game”.
- On performing correlation matrix we found there exist singularities between heat index sensor and temperature sensor. Hence, suggest the client to have any one sensor in their product and thereby our suggestion could **reduce around 16% of their product cost.**
- Area based analysis and ranking is provided.

Contents:

- Removing the singularities.
- Certain places are out of game.
- Area based analysis and Ranking.
- Correlation and Linear Regression.
- Multiple Linear Regression.

Removing the Singularities

Insights:

While running the linear model, it became evident that there exist a singularity.

```
Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error t value Pr(>|t|)

(Intercept)             13.202107   0.213371  61.874   <2e-16
***
data$Temperature        -0.003891   0.004463  -0.872   0.3834

data$Relative.Humidity   0.003028   0.001794   1.688   0.0915
 .
data$Heat.Index                NA         NA      NA       NA

data$Carbon.Monoxide     0.374373   0.005751  65.098   <2e-16
***
data$Noise.Pollution     0.262325   0.004015  65.334   <2e-16
***
---
```

Steps taken to find the singularity:

Find the correlation matrix among the independent variables that will showcase the variable that has more correlation on other.

```
                [,1]          [,2]          [,3]          [,4]          [,5]
[1,]  1.0000000000  0.537889412  0.996426740 -0.002864487 -0.000697518
[2,]  0.5378894123  1.000000000  0.607169774 -0.002928457 -0.003740822
[3,]  0.9964267399  0.607169774  1.000000000 -0.002993283 -0.001032225
[4,] -0.0028644871 -0.002928457 -0.002993283  1.000000000  0.773613603
[5,] -0.0006975168 -0.003740820 -0.001032227  0.773613605  1.000000000
```

From the above we could see that there is a strong correlation among the variable one(Temperature) and variable three (Heat Index).

Steps taken to Remove the Singularity:

Now find the linear model for the air pollution with respect to heat index separately and between air pollution and Temperature separately. Take the variable which determines the Air pollution the most by considering the multiple r squared value.

The multiple R squared value of **Temperature**

Multiple R-squared: 1.528e-06

The multiple R squared value of **Heat index**

Multiple R-squared: 7.647e-07

Removing the singularity:

With the results above, it is proved that temperature determine the air pollution more than heat index. Hence, Heat Index is removed.

Result:

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            13.202107   0.213371  61.874   <2e-16 ***
data$Temperature       -0.003891   0.004463  -0.872   0.3834
data$Carbon.Monoxide    0.374373   0.005751  65.098   <2e-16 ***
data$Relative.Humidity  0.003028   0.001794   1.688   0.0915 .
data$Noise.Pollution    0.262325   0.004015  65.334   <2e-16 ***
```

Certain Places are out of game

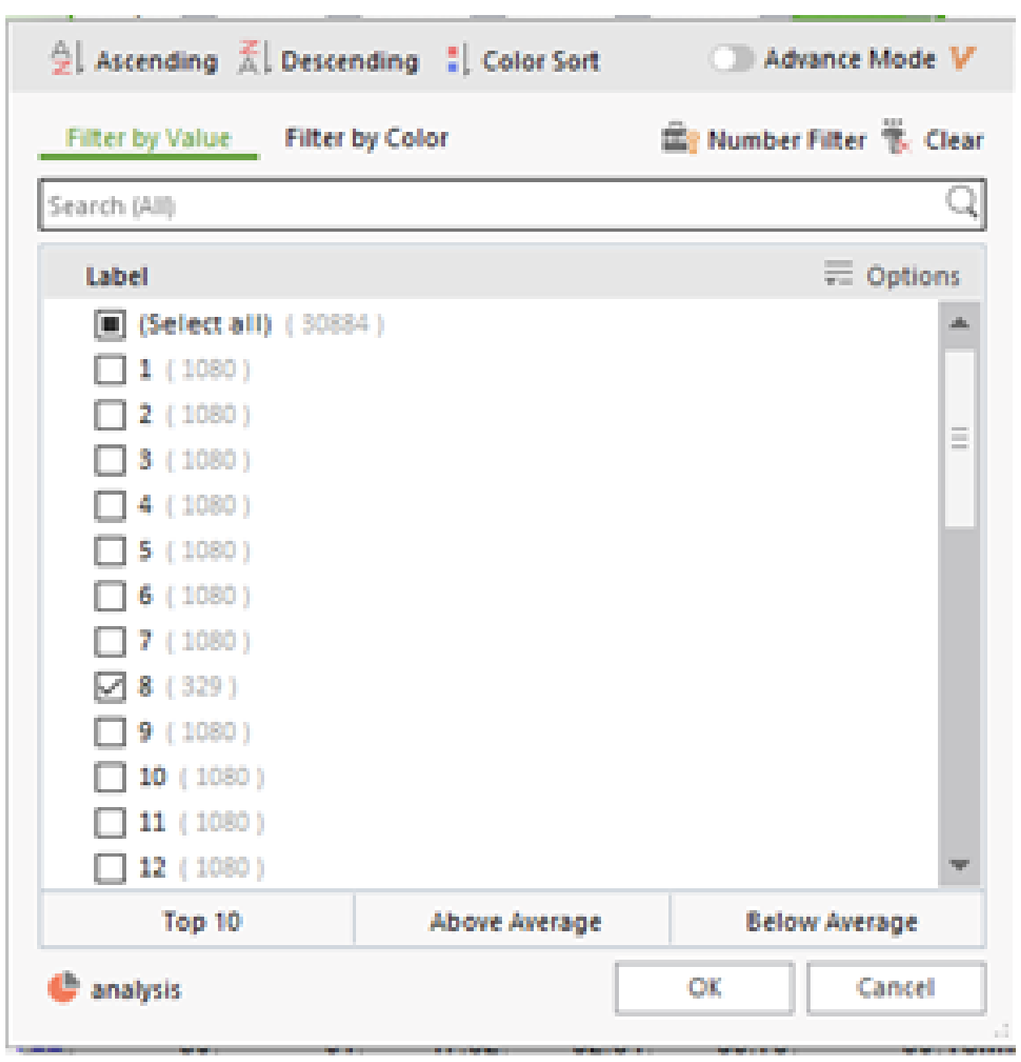
Insights:

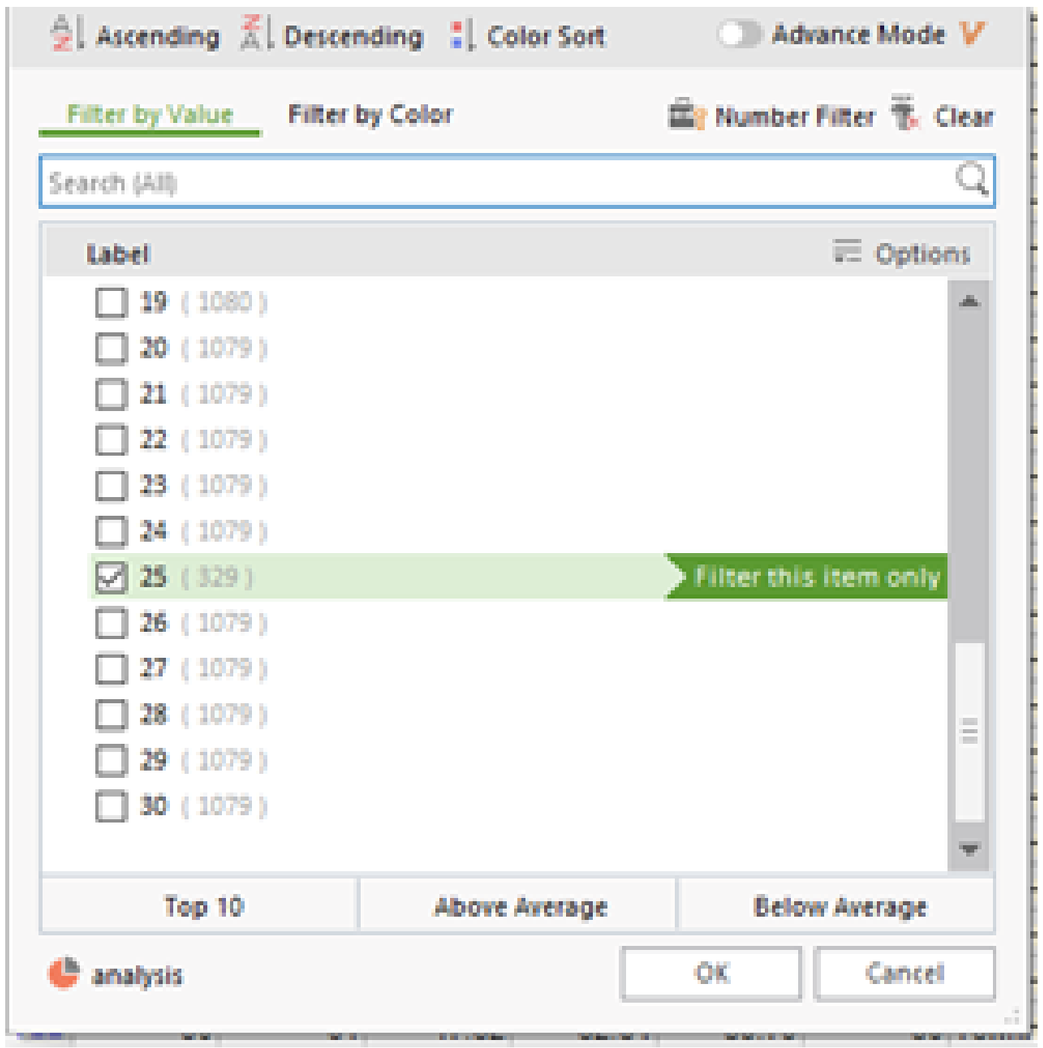
It is understood that the device 8 and 25 are malfunctioned. Hence, the places such as Ramanathapuram and Chinthamani cannot be considered for the prediction.

Reason:

While all devices have recorded 1080 data, the devices such as 8 and 25 doesn't even have recorded half of it (329).

Proof:





Area based analysis

1) Air Pollution:

The minimum of air pollution is recorded at Idyapalayam-palakad road(30) on 14/3/2018.
The maximum of air pollution is recorded at **Sundrapuram(81.36)** on **21/10/2017**.

2) Carbon Mono oxide:

The minimum of carbon mono oxide is recorded at **Selvapuram(30.03)** on **5/10/17**.
The maximum of carbon mono oxide is recorded at **Kavundampalayam(76.95)** on **19/10/2017**.

3) Noise Pollution:

The minimum of Noise pollution is recorded at **Selvapuram(30.04)** on **31/8/2017**.
The maximum of Noise pollution is recorded at **Saravanampatti(90.53)** on **1/1/2018**.

Ranks

1) Air Pollution:

Rank Location Average of Air Pollution

1 Selvapuram 35.09032
2 Podanur 35.11427
3 Idyapalayam-PalakadRoad 35.23735
4 Singanallur 38.58264
5 Puliyakulam 38.64071
6 Ondipudhur 38.68282
7 Vadavalli 38.78005
8 Idayarpalyam-Jn 38.79216
9 RaceCourse 47.53369
10 Sungam 47.55593
11 NewBusStand 47.55851
12 Upilipalaym 47.57296
13 GandhipuramBusstand 47.57356
14 WomensPolytechnic 47.58351
15 TownHall 47.58504
16 Gandhipuram-Tatabad 47.61369
17 PooMarket 47.63644
18 Gandhipuram 47.65366
19 SinganallurBusStand 47.68551
20 SaiBabaColony 47.69422
21 RSPuram 47.70571
22 Ukkadam 47.75604
23 TownHall-CCMC 47.76443
24 Ganapathi 49.64999
25 Thudiyalur 49.6508
26 Sundrapuram 49.77908
27 Kavundampalayam 49.81606
28 Saravanampatti 49.98696

2) Noise Pollution:

Rank	Location	Average of Noise Pollution
1	Selvapuram	37.63648
2	Idyapalayam-PalakadRoad	37.89736
3	Podanur	38.00832
4	Idayarpalyam-Jn	41.71547
5	Singanallur	41.72136
6	Vadavalli	41.75407
7	Puliyakulam	41.80591
8	Ondipudhur	41.95264
9	Ukkadam	57.67444
10	NewBusStand	57.68134
11	SinganallurBusStand	57.68424
12	TownHall-CCMC	57.69065
13	Gandhipuram	57.73976
14	Upilipalaym	57.75607
15	WomensPolytechnic	57.76033
16	TownHall	57.77813
17	RaceCourse	57.79719
18	Gandhipuram-Tatabad	57.79789
19	Sungam	57.847
20	RSPuram	57.85596
21	GandhipuramBusstand	57.89146
22	SaiBabaColony	57.97299
23	PooMarket	57.98084
24	Kavundampalayam	62.78642
25	Ganapathi	62.81537
26	Sundrapuram	62.8418
27	Thudiyalur	63.02881
28	Saravanampatti	63.05146

Correlation & Regression

Result:

Relationship between temperature and humidity is 48%.

Relationship between Air pollution and Noise pollution is 54.6%

Other Insights:

1) Relationship between temperature and other parameters:

- The relation between humidity and temperature is 48%.
- The relation between temperature and carbon mono oxide is .1%
- The relation between temperature and air pollution is 1.3%
- The relation between temperature and the noise pollution is 1.5%

	Temperature	Humidity	Carbon.Monoxide	Air.Pollution	Noise.Pollution
Temperature	1.000000000	0.48341323	0.001605298	0.01346443	0.01547675

2) Relationship between Humidity and other parameters:

- The relation between humidity and temperature is 48%.
- The relation between humidity and carbon mono oxide is 3.6%
- The relation between humidity and air pollution is 4.5%
- The relation between humidity and the noise pollution is 1.6%

	Temperature	Humidity	Carbon.Monoxide	Air.Pollution	Noise.Pollution
Humidity	0.483413227	1.00000000	0.036872460	0.04523131	0.01605351

3) Relationship between Carbon mono oxide and other parameters:

- The relation between carbon mono oxide and temperature is .1%.
- The relation between humidity and carbon mono oxide is 3.6%
- The relation between carbon mono oxide and air pollution is 90.7%
- The relation between carbon mono oxide and the noise pollution is 54.7%

	Temperature	Humidity	Carbon.Monoxide	Air.Pollution	Noise.Pollution
Carbon.Monoxide	0.001605298	0.03687246	1.000000000	0.90707296	0.54755120

4) Relationship between Noise pollution and other parameters:

- The relation between noise pollution and temperature is 1.54%.
- The relation between humidity and noise pollution is 1.6%
- The relation between noise pollution and air pollution is 54.6%
- The relation between carbon mono oxide and the noise pollution is 54.7%

	Temperature	Humidity	Carbon.Monoxide	Air.Pollution	Noise.Pollution
Noise.Pollution	0.015476748	0.01605351	0.547551200	0.54623214	1.00000000

5) Relationship between Air Pollution and other parameters:

- The relation between air pollution and temperature is 1.34%.
- The relation between humidity and air pollution is 4.5%
- The relation between noise pollution and air pollution is 54.6%
- The relation between carbon mono oxide and the air pollution is 90.7%

	Temperature	Humidity	Carbon.Monoxide	Air.Pollution	Noise.Pollution
Air.Pollution	0.013464427	0.04523131	0.907072958	1.00000000	0.54623214

Regression

1) Regression between temperature and humidity

Linear regression equation for temperature and humidity is
Temperature = 30.219866+0.007826Humidity

The above equation shows that for every unit increase in humidity there is a 0.007826 increase in temperature.
The **multiple R squared value** is **0.07** which shows that temperature does not explain the humidity well.
This shows that there exist **poor relation between temperature and humidity**.

```
lm(formula = data$Average.of.Temperature ~ data$Average.of.Relative.Humidity)

Coefficients:
                        (Intercept)  data$Average.of.Relative.Humidity
                          30.219866                           0.007826
```

2) Regression between air pollution and noise pollution

Linear regression equation for air-pollution and noise-pollution is
Air pollution = 14.303+0.573 Noise pollution

For every, unit increase in noise pollution there is an increase of .573 increase in air pollution.
The **multiple r squared value** is **99.19**, which shows that the Air pollution explains Noise pollution.
There exist a **good relation between air pollution and noise pollution.**

```
> lm2

Call:
lm(formula = data$Average.of.Air.Pollution ~ data$Average.of.Noise.Pollution)

Coefficients:
                        (Intercept)  data$Average.of.Noise.Pollution
                             14.303                            0.573
```

Multiple Linear Regression

Insights:

Multiple Linear regression equation is

A.P = −18.738937 +0.002069 T+0.001042H + 1.281859C + 0.057168 N.P

```
> fit
Call:
lm(formula = a$Air.Pollution ~ a$Temperature + a$Humidity + a$Carbon.Monoxide +
    a$Air.Pollution + a$Noise.Pollution)

Coefficients:
      (Intercept)      a$Temperature         a$Humidity  a$Carbon.Monoxide
        -18.738937           0.002069           0.001042           1.281859
a$Noise.Pollution
         0.057168
```

Here,

A.P -> Air Pollution

T -> Temperature

C-> Carbon Mono-oxide

N.P->Noise Pollution

The above equation explains us,

1. An unit increase in temperature increase the Air pollution by 0.002069.
2. An unit increase in Humidity increases the Air pollution by 0.001042.
3. An unit increase in Carbon Mono-Oxide increases the Air pollution by 1.281859.
4. An unit increase in Noise Pollution increases the Air pollution by 0.057168.
5. It is clear that Heat Index is eliminated because it correlate with the Temperature.

Further info,

1. Among the four factors, Carbon mono-oxide determines affects the most and then comes the noise pollution. Theses two are the major factors that determines the Air
pollution.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
a$Carbon.Monoxide  1.281859   0.028863  44.411  < 2e-16 ***
a$Noise.Pollution  0.057168   0.015794   3.620 0.000318 ***
```

2. The above prediction is good as the Multiple R-Squared value is **MultipleR-squared: 0.8265**.