Analysis of the impact of weather factors on air quality in Ho Chi Minh City

Pham Tien Duong [1,2,3], Tran Hoang Anh [1,2,3], Ngo Huynh Truong[1,2,3], Nguyen Huu Minh Tam [1,2,3], and Do Trong Hop [1,2,4]

[1] **University of Information Technology, Ho Chi Minh City, Vietnam**
[2] **Vietnam National University, Ho Chi Minh City, Vietnam**
[3] **{20521222, 20521079, 20522085, 20521871}@uit.edu.vn**
[4] **hopdt@uit.edu.vn**

Abstract: Weather is the state of the atmosphere at a time; the specific area is determined by meteorological phenomena, such as rain, heat and cold, monsoon, low humidity or dryness,... In a particular locality in a short period. Temperature is a vital weather factor, the or, the physical properties of matter, in other words, the scale of "hot" and elements such as rain, heat and cold, monsoon, humidity or dryness,...;...day based on the weather data of the previous three days and the experimental parameters means, in determining the life of humans and l creatures on earth. In this topic, our team will analyze the influence of weather factors on air quality in Ho Chi Minh City, hypothesizing the relationship between air quality and other weather attributes. After testing the above houses,, the team will conduct methods to model the problem of predicting the next day's air quality, the team will show processes their ways the problem of predicting the next day's air quality; the team will work procedures s airways air model the difficulty of predicting the next day's air lit grade lit quality air quality of the next day based on the weather data of the previous three day airways air quality of the next day based on the weather data of the last previous, the team will conduct methods to model the problem of predicting the next day's air quality; the team will show processes their ways the problem of predicting the next day's air quality; the team will work procedures stairways air model the difficulty of predicting the next day's air lit quality lit quality air quality of the next day based on the weather data of the previous three day airways air quality of the next day based on the weather data of the last previous three days and days and days the RI mental parameters measured.

Keywords: pm2.5, air quality, weather, experiment, hypothesis.

1 Introduction

1.1 Problem

We humans can survive for 30 days without eating, 3 days without drinking, but only 3 minutes without breathing. The need for human air is so great, we always want to live in clean and fresh air. However, in Ho Chi Minh City today, with more and more waste and pollutants being released into the environment, the air is also gradually becoming more polluted.

The dust or compound present in the dust is known as Particulate Matter, with the symbol is pm. There are many types of dust: pm10, pm2.5, pm1.0… Our group will be interested in pm2.5.

We will analyze, find out the influence of weather factors on the main factor that the group is interested in: pm2.5, making hypotheses and testing. After testing the hypotheses, we will conduct machine learning methods to model the problem of predicting air quality (the value of pm2.5 dust) of the next day based on the weather data of the previous three days and the measured experimental parameters.

1.2 Dataset

This dataset consists of two parts: weather data and air quality data. With the weather data, we collected from the NASA, the specific path will be in the References section. For air quality data, we only take the dust pm2.5, then proceed to combine the above two parts, preprocessing data. Finally, we get the final dataset in the form of a statistical table (file .csv), with information about the weather, air quality of Ho Chi Minh City in five years (from January 1st 2017, to December 31st 2021). The dataset after being processed consists of 1725 records, with 11 attributes as follows:

- **YEAR** (int64).
- **MO** (int64): Month of the year.
- **DY** (int64): The day of the month.
- **Relative_Humidity** (float64): Relative humidity.
- **Specific_Humidity** (float64): Specific humidity, the volume of steam (in grams) in 1kg of moist air mass.
- **Precipitation** (float64): Rainfall by day (mm).
- **Pressure** (float64) (kPa).
- **Wind_Speed** (float64): Wind speed (m/s).
- **Wind_Direction** (float64): Wind direction (degree).
- **PM2.5** (int64): The value of dust with a diameter of less than 2.5 microns.

	YEAR	MO	DY	Temperature	Relative_Humidity	Specific_Humidity	Precipitation	Pressure	Wind_Speed	Wind_Direction	pm25
0	2017	1	2	26.67	79.38	17.03	0.22	100.49	2.81	90.94	84
1	2017	1	3	26.91	77.19	16.91	0.06	100.46	1.99	106.56	109
2	2017	1	4	26.06	80.62	16.85	0.00	100.39	2.52	115.00	128
3	2017	1	5	26.54	81.00	17.27	0.00	100.34	3.47	83.19	118
4	2017	1	6	26.63	79.31	17.09	0.01	100.23	3.06	103.19	87
...	...	...	...	...	...	...	...	...	...	...	...
1720	2021	12	27	26.42	80.69	17.09	0.92	100.77	1.61	94.12	103
1721	2021	12	28	25.58	84.62	17.15	0.48	100.69	2.87	128.19	84
1722	2021	12	29	25.67	83.81	17.09	0.12	100.68	2.98	48.38	94
1723	2021	12	30	25.49	81.38	16.36	0.00	100.77	2.98	79.81	53
1724	2021	12	31	24.83	79.62	15.38	0.00	100.80	2.70	40.94	65

1725 rows × 11 columns

Figure 1. Dataset

2 Data analysis and visualization

In this section, we will implement a variety of visualization methods across each data field and on relationships between these data fields before analyzing them. Thereby, we will have a deeper insight into this weather dataset.

Value of pm2.5	Air Quality
0 – 50	Good
51 – 100	Moderate
101 – 150	Lightly Polluted (Sensitive group)
151 – 200	Medially Polluted (Unhealthy)
201 – 300	Heavily Polluted (Very Unhealthy)
301 – 500	Severely Polluted (Dangerous)

Table 1. Air quality classification in Vietnam by value of pm2.5

2.1 Classification of air quality in Vietnam by the value of pm2.5

From table 1, we can see that the higher the pm2.5 is, the more harmful to the air quality.

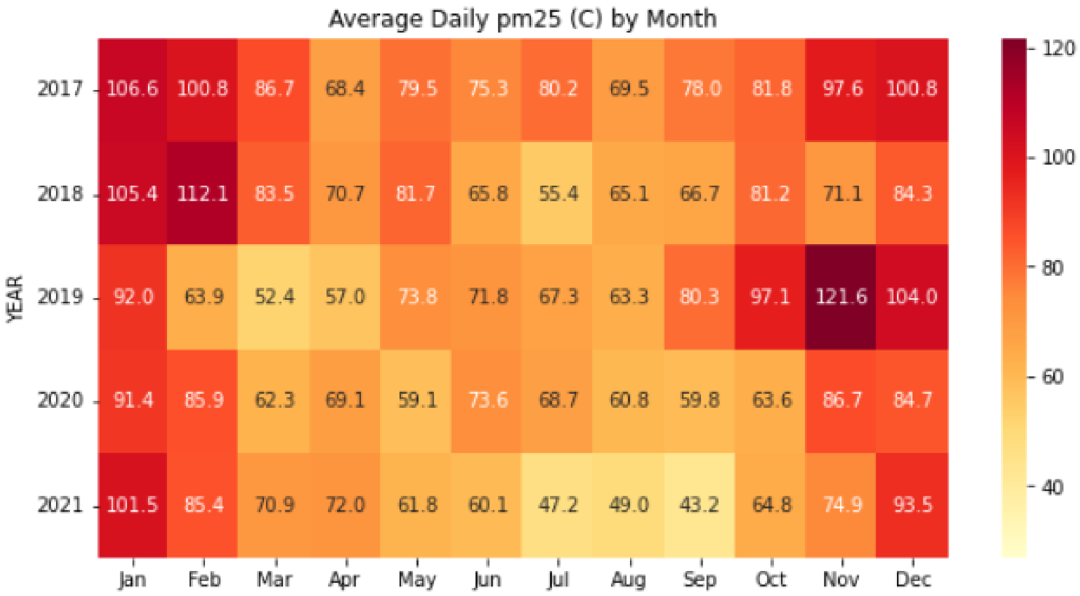


Figure 2. Average value of daily PM2.5 per month

Comment: Looking at Figure 2, we see the air quality in Ho Chi Minh City in 5 years, from 2017 to 2021. We see that months with bad air quality are usually concentrated from November to February of the following year (winter). Meanwhile, from March to October every year, air quality will usually be good or average. There is a difference in air quality in the Ho Chi Minh City because air quality is strongly affected by weather factors including factors such as wind, wind direction, temperature, humidity, precipitation and sunshine. Especially in summer with high temperature, heavy rain, a lot of light and high wind, pollutants rise quickly. As a result, the concentration of dust in the air will be lower, and pm2.5 will also be lower, therefore, the air quality will be better.

By contrast, in winter with low temperature, little light, less rain, the northeast monsoon overflows, bringing pollutants from the north to make the air more polluted. In particular, November 2019 had the highest average PM2.5, with the value of 121.6, while September 2021 had the lowest average PM2.5, with the value of 43.2.

2.2 The description of the dataset

	Temperature	Relative_Humidity	Specific_Humidity	Precipitation	Pressure	Wind_Speed	Wind_Direction	pm25
count	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000
mean	27.602006	77.684365	17.401304	4.186475	100.395954	2.965762	172.570910	76.351304
std	1.971128	12.761897	2.403383	8.556880	0.187268	1.117604	63.031659	24.749360
min	19.900000	43.000000	9.280000	0.000000	99.870000	0.680000	22.310000	5.000000
25%	26.380000	66.440000	15.560000	0.010000	100.270000	2.090000	122.000000	58.000000
50%	27.100000	82.500000	18.370000	0.870000	100.380000	2.780000	162.310000	73.000000
75%	28.830000	88.250000	19.350000	4.770000	100.520000	3.740000	234.120000	92.000000
max	33.700000	95.310000	20.940000	143.310000	101.060000	7.050000	343.560000	222.000000

Figure 3. The description of the dataset

Comment: The statistics above tell us about the amount of data (count), the mean, the variance (std), the smallest value (min), the lower quartile (25%), the median (50%), the upper quartile (75%) and the highest value (max) of each features included in the dataset, after the removal of the YEAR, MO and DY.

We see that the value range of each feature is neither too small nor too large, so we do not need to scale data.

2.3 Charts of frequency of properties in dataset

- **Temperature:**

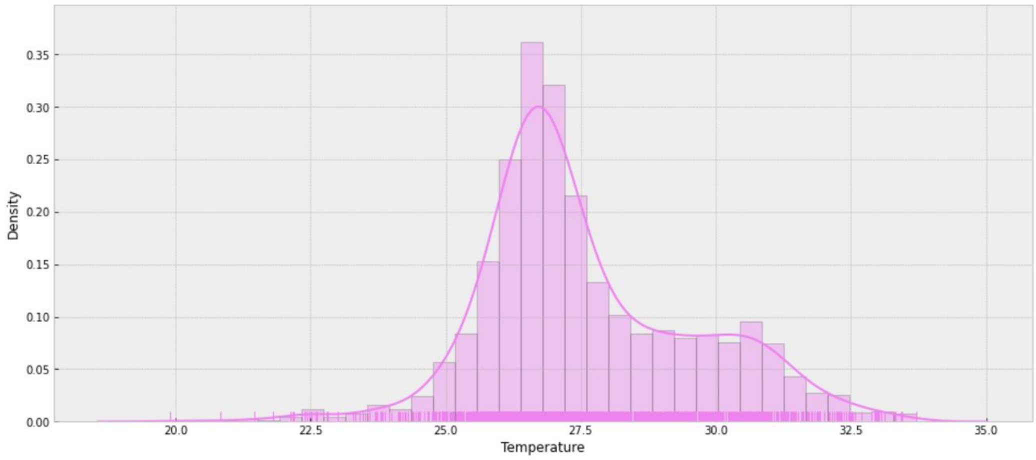


Figure 4. Temperature

Comment: Temperature in Ho Chi Minh City receives a value between 19.9 and 33.7 degrees Celsius, high temperature in summer and lower in winter. In particular, the temperature value between 26 and 27 degrees Celsius appears with the most frequency (the chart rises in the middle at the value between 26 and 27 degrees Celsius). The distribution chart of the temperature has a standard distribution, which protrudes in the middle and descends to the sides.

- **Relative Humidity:**

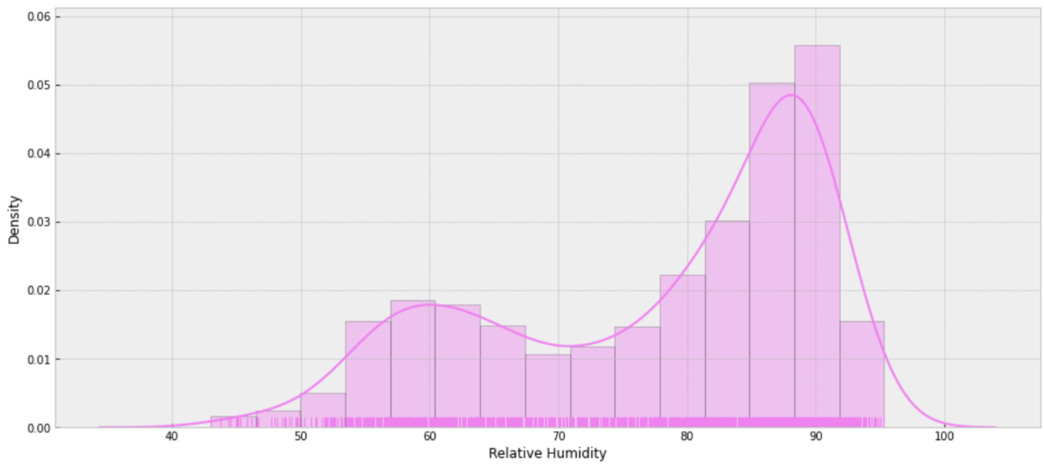


Figure 5. Relative Humidity

Comment: Relative Humidity, which represents the ratio between air vapor pressure and saturated vapor pressure its, receiving values between 43.00 and 95.31. Through the image above, we can see that this ratio is concentrated in about 85-90 percent (water evaporation occurs slowly, humid air).

- **Pressure:**

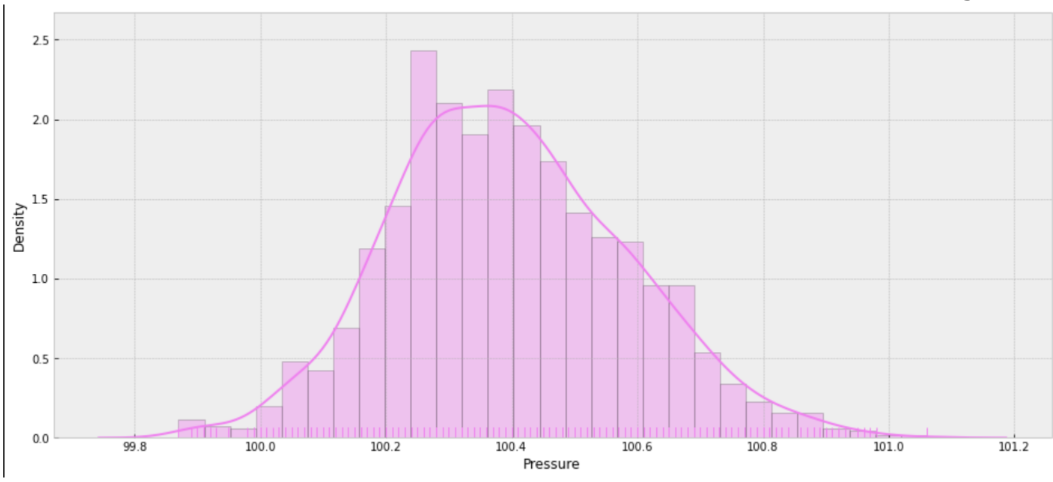


Figure 6. Pressure

Comment: Pressure has a value domain from 99.87 to 101.06 kPa, this value domain is quite small, we see the value of pressure distributed centrally between 100.2 and 100.4. The chart above is almost as a standard distribution.

- **Specific Humidity:**

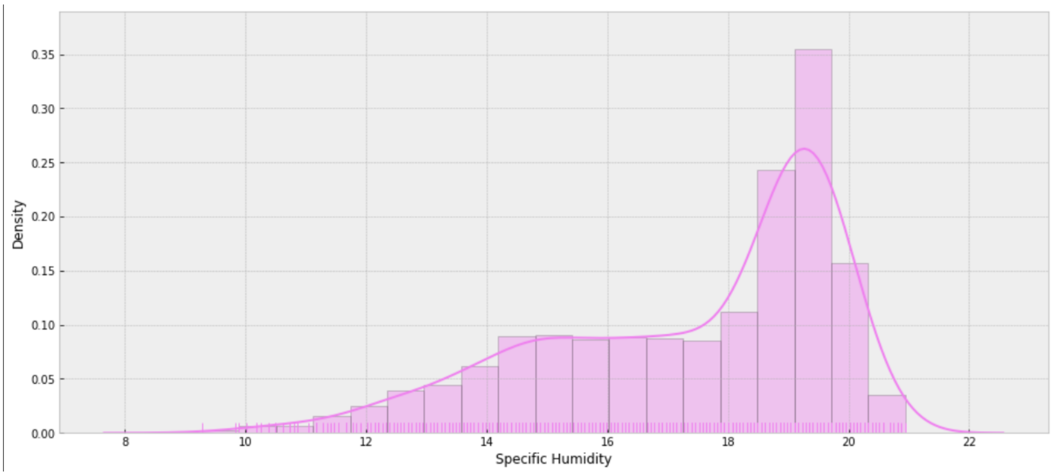


Figure 7. Specific Humidity

Comment: Specific Humidity has a value ranging from 9.28 to 20.94, where the value range from 14 to 18 has an almost equal frequency of occurrence, and the value of the features concentrates mainly between 18 and 20.

- **Precipitation:**

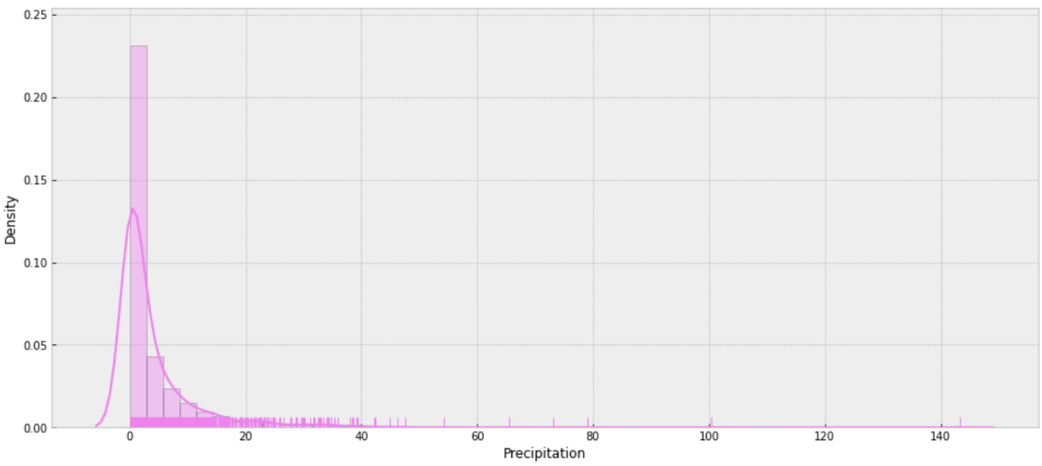


Figure 8. Precipitation

Comment: Precipitation, which has a value range of 0 to 143.31 mm. Looking at the chart, we see most days of the year, it doesn't rain, rainfall is usually low and heavy rain in this place is very rarely. The rainfall distribution chart does not follow a distribution at all.

- **Wind speed:**

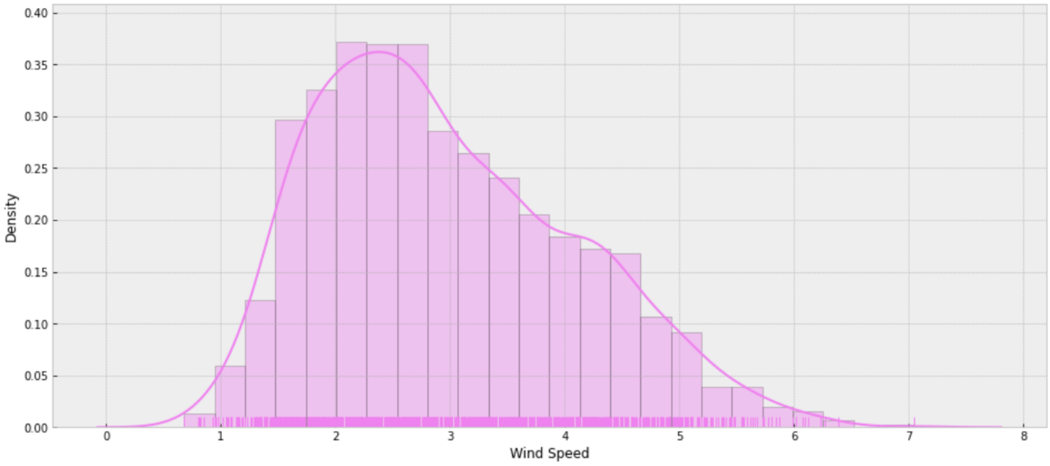


Figure 9. Wind Speed

Comment: Figure 9 show that Wind Speed has a domain value from 0.68 to 0.68 7.05 m/s, wind speed is usually concentrated in the range of 2-3 m/s; this is a light and moderate wind speed.

- **Wind Direction:**

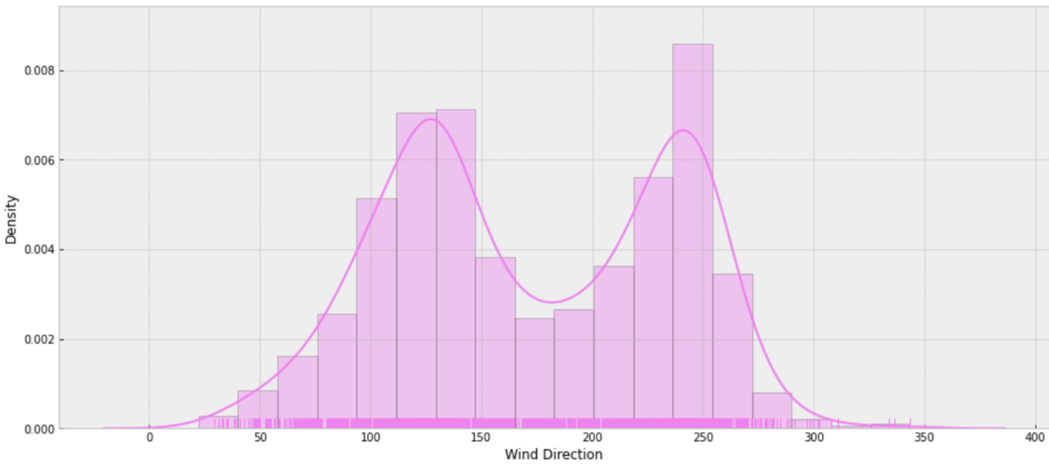


Figure 10. Wind Direction

Comment: Figure 10 show that Wind Direction receives values between 22.31 and 343.56 degrees. In which the wind direction of about 250 degrees accounts for the highest frequency, and the wind with a direction of about 145 degrees also accounts for a fairly high frequency. Ho Chi Minh City has the main wind direction is west - southwest, this is the reason for the above wind direction distribution.

- **PM2.5:**

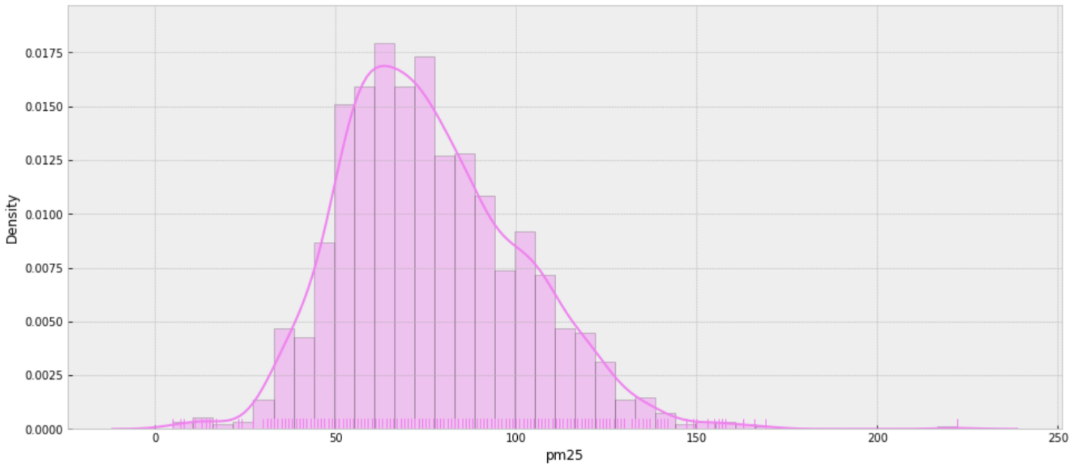


Figure 11. PM2.5

Comment: pm2.5 index with a value range of 5.00 to 222.00, of which the focus is mainly from 50 to 100. Such pm2.5 shows that air quality in Ho Chi Minh City is average, the frequency of good air quality is quite small, while bad air quality appears quite often. This is quite frightening in the 5 years from 2017 to 2021.

3 Analysis of factors

3.1 Time factor – Month

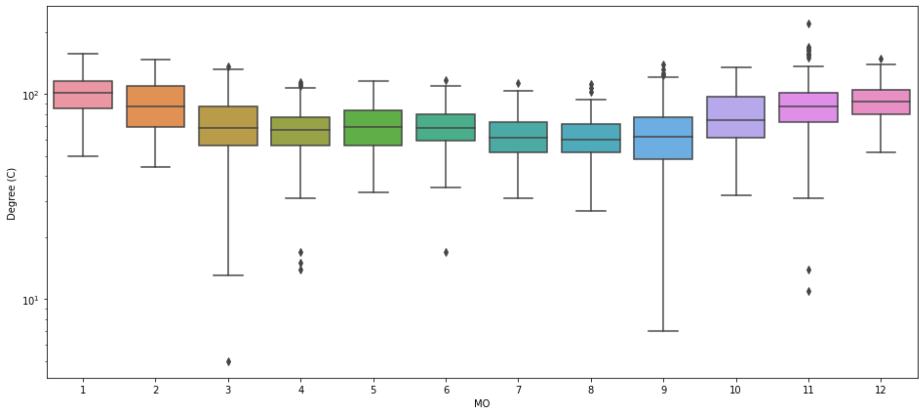


Figure 12. Distribution by pm2.5 of the Month factor

- H0: Months in the same group.
- H1: Months in different groups.

```
                Df Sum Sq Mean Sq F value Pr(>F)    
factor(df$MO)   11 272457   24769   54.15 <2e-16 ***
Residuals     1713 783547     457                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 13. One-way ANOVA on Month factor

The result show us that p_value < 2e-16 < 0.05. Therefore, we reject H0, i.e. months are in different groups.

```
                     Df Sum Sq Mean Sq F value Pr(>F)
factor(e345_data$MO)   3   1526   508.6   1.346  0.259
Residuals            570 215392   377.9               
```

Figure 14. One-way ANOVA on March, April, May and June

```
                     Df Sum Sq Mean Sq F value Pr(>F)
factor(e345_data$MO)   2    882   440.8   1.207    0.3
Residuals            441 161082   365.3
```

Figure 15.One-way ANOVA on July, August and September

```
                     Df Sum Sq Mean Sq F value Pr(>F)
factor(e345_data$MO)   1      7     6.9   0.009  0.924
Residuals            275 207477   754.5
```

Figure 16. One-way ANOVA on February and November

```
                     Df Sum Sq Mean Sq F value   Pr(>F)
factor(e345_data$MO)   2  33138   16569   35.45 5.73e-15 ***
Residuals            427 199596     467
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 17. One-way ANOVA on January, October and December

Figure 18

Comment: In January, October, and December, the results of the One-way ANOVA once again showed that the months January, October and December belong to different groups. We use Tukey HSD for analysis.

Pairs of months with confidence intervals do not contain 0 so the difference between these pairs of months is statistically significant. Therefore, it is possible to conclude that the Month factor can be divided into 6 groups: the March-April-May-June group, the July-August-September group, the February-November group, and the January group, the October group, the December group.

3.2 Time factor – Day

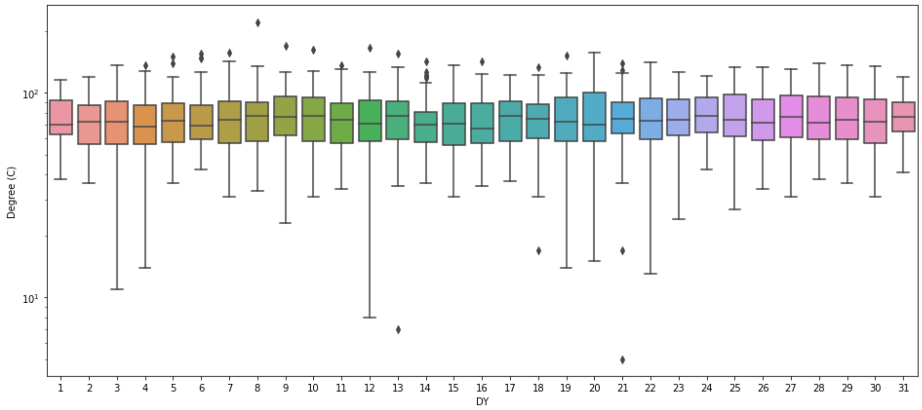


Figure 19. Distribution by pm2.5 of the Day factor

```
                 Df  Sum Sq Mean Sq F value Pr(>F)
factor(df$DY)    30    4624   154.1   0.248      1
Residuals      1694 1051379   620.6
```

Figure 20. One-way ANOVA on the Day factor

3.3 Analysis of natural conditions factors

– Two-way ANOVA

H0a: The months factor do not affect air quality.
H0b: Natural conditions factors do not affect air quality.
H0c: Natural conditions factors and months factors do not interact with each other.

3.3.1 Temperature

```
ANOVA Table (type II tests)

           Effect DFn  DFd      F        p p<.05      ges
1              MO  11 1701 51.891 7.93e-99     * 0.251000
2     Temperature   1 1701  1.152 2.83e-01       0.000677
3 MO:Temperature   11 1701 11.078 4.57e-20     * 0.067000
```

Figure 21. Two-way ANOVA on Month factor and Temperature

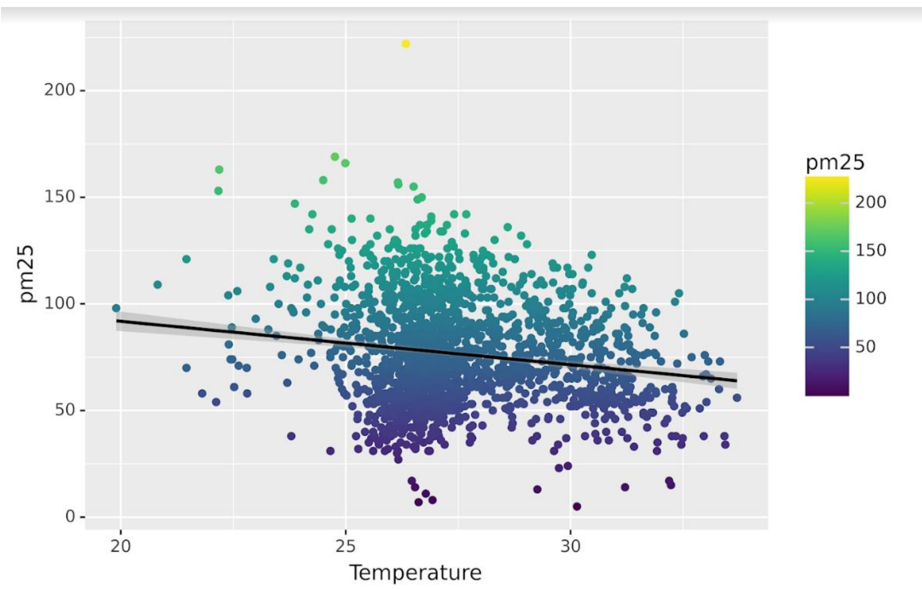


Figure 22

3.3.2 Relative Humidity

```
ANOVA Table (type II tests)

               Effect DFn  DFd      F          p p<.05   ges
1                  MO  11 1701 58.260 7.04e-110     * 0.274
2   Relative_Humidity   1 1701 40.369  2.69e-10     * 0.023
3 MO:Relative_Humidity  11 1701 16.358  5.82e-31     * 0.096
```

Figure 23. Two-way ANOVA on Month factor and Relative Humidity

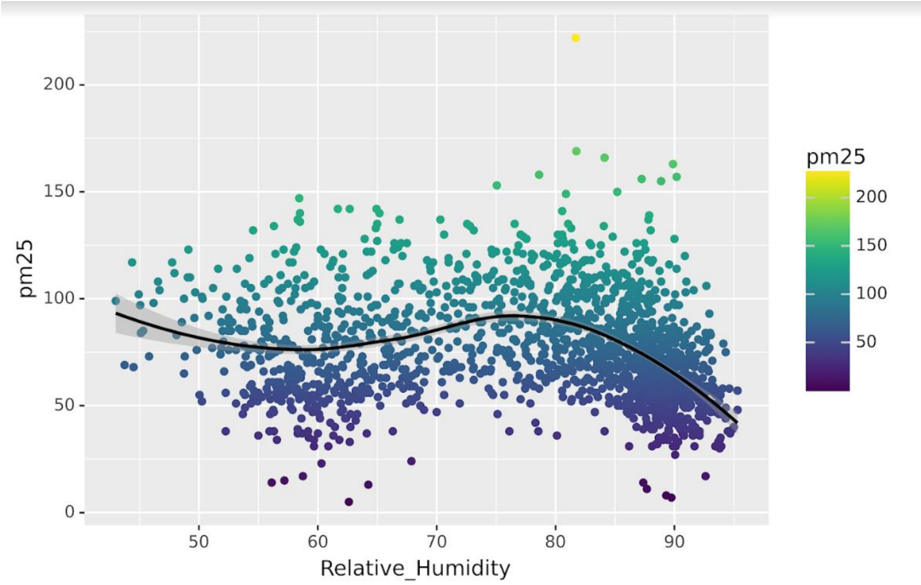


Figure 24

3.3.3 Specific Humidity

```
ANOVA Table (type II tests)

               Effect DFn  DFd      F         p p<.05   ges
1                  MO  11 1701 39.142 1.02e-75     * 0.202
2   Specific_Humidity   1 1701 50.068 2.16e-12     * 0.029
3 MO:Specific_Humidity  11 1701  7.475 1.17e-12     * 0.046
```

Figure 25. Two-way ANOVA on Month factor and Specific Humidity

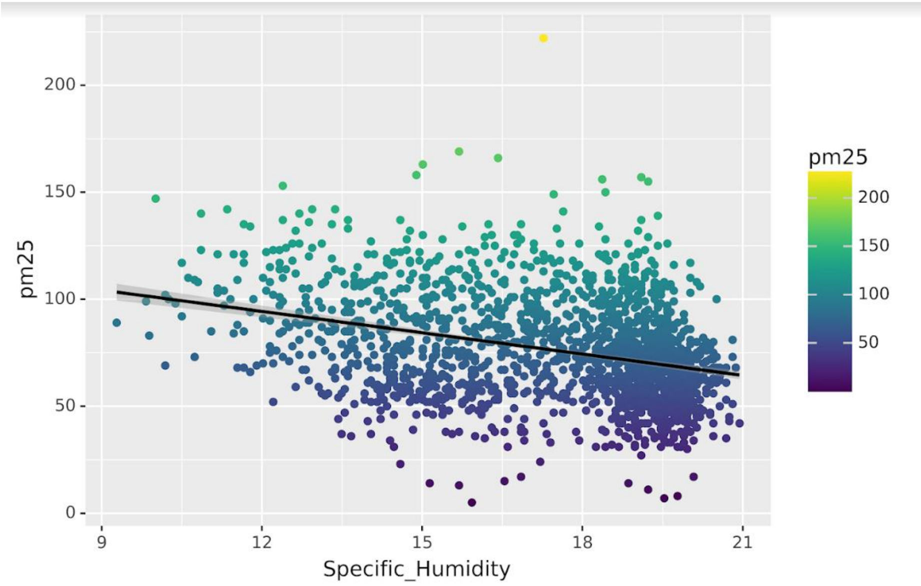


Figure 26

3.3.4 Pressure

ANOVA Table (type II tests)

	Effect	DFn	DFd	F	p	p<.05	ges
1	MO	11	1701	27.075	1.47e-52	*	0.149
2	Pressure	1	1701	55.286	1.64e-13	*	0.031
3	MO:Pressure	11	1701	4.373	1.71e-06	*	0.027

Figure 27. Two-way ANOVA on Month factor and Pressure

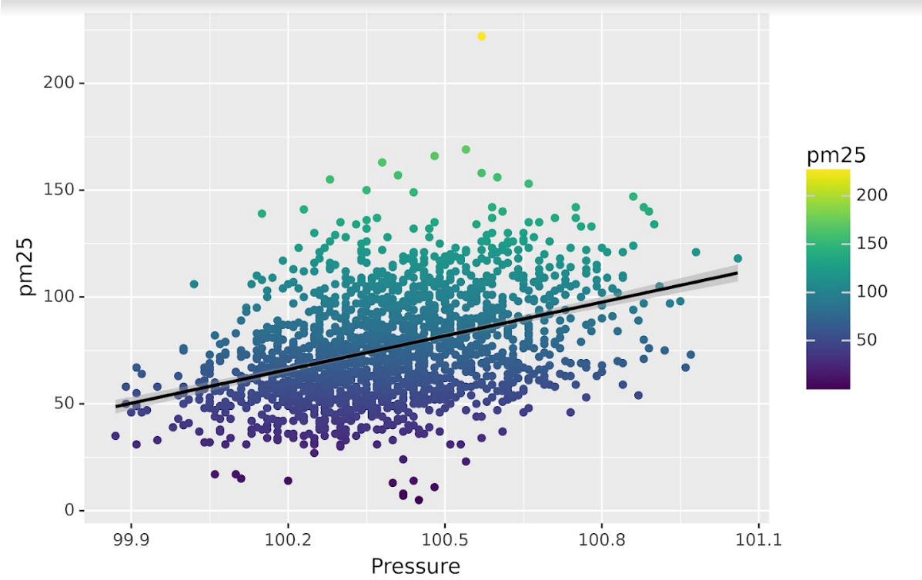


Figure 28

3.3.5 Precipitation

ANOVA Table (type II tests)

	Effect	DFn	DFd	F	p	p<.05	ges
1	MO	11	1701	47.403	7.62e-91	*	0.235
2	Precipitation	1	1701	22.787	1.97e-06	*	0.013
3	MO:Precipitation	11	1701	3.923	1.21e-05	*	0.025

Figure 29. Two-way ANOVA on Month factor and Precipitation

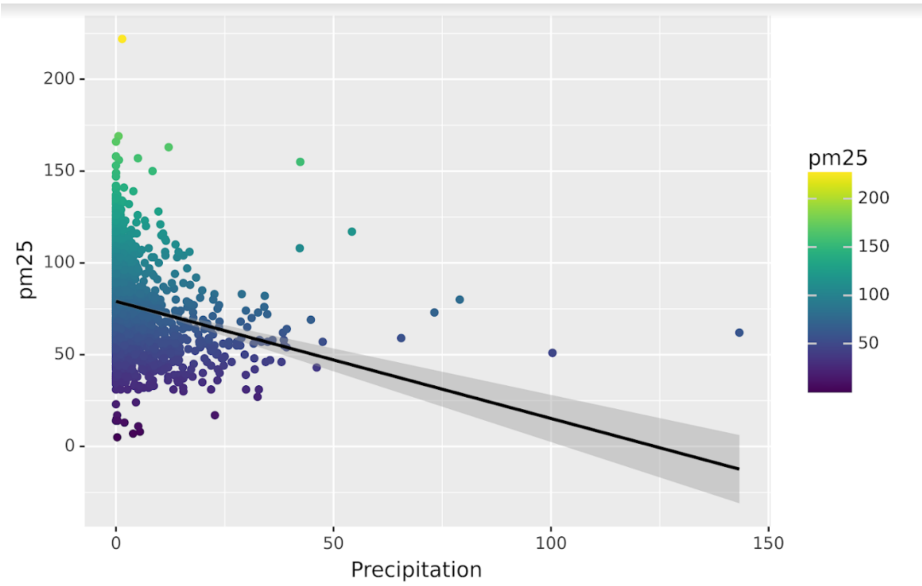


Figure 30

Comment: We see that in ANOVA table, precipitation has interacted with pm2.5. However, in the scatterplot, we can see the peculiarity of the features in an area, not following any pattern. Therefore, we can conclude that Precipitation and PM2.5 has no interaction with each other (or has but very little).

3.3.6 Wind Speed

```
ANOVA Table (type II tests)

         Effect DFn  DFd       F        p p<.05   ges
1            MO  11 1701  36.340 1.88e-70     * 0.190
2    Wind_Speed   1 1701 426.402 1.03e-84     * 0.200
3 MO:Wind_Speed  11 1701   6.763 3.27e-11     * 0.042
```

Figure 31. Two-way ANOVA on Month factor and Wind Speed

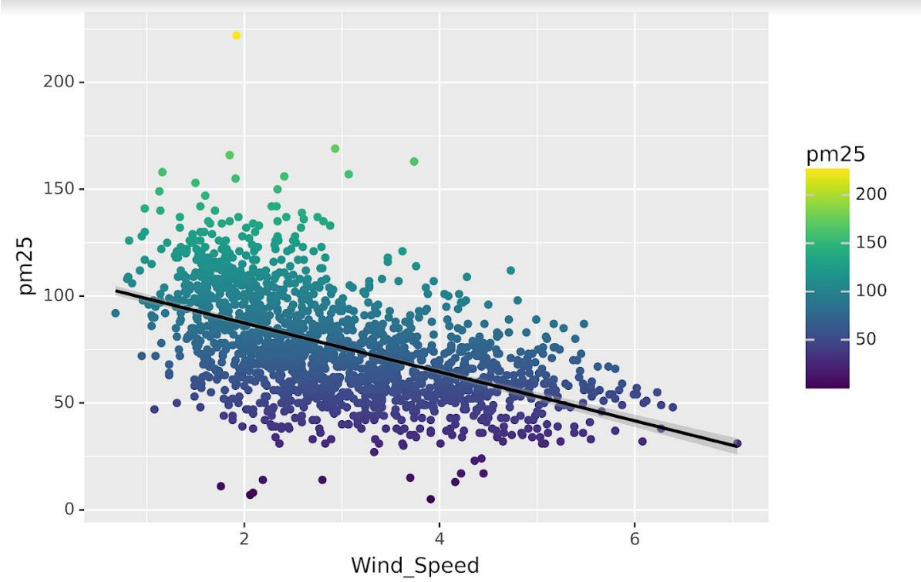


Figure 32

3.3.7 Wind Direction

```
ANOVA Table (type II tests)

             Effect DFn  DFd      F        p p<.05      ges
1                MO  11 1701 38.262 4.53e-74     * 1.98e-01
2    Wind_Direction   1 1701  0.048 8.26e-01       2.85e-05
3 MO:Wind_Direction  11 1701  4.393 1.57e-06     * 2.80e-02
```

Figure 33. Two-way ANOVA on Month factor and Wind Direction

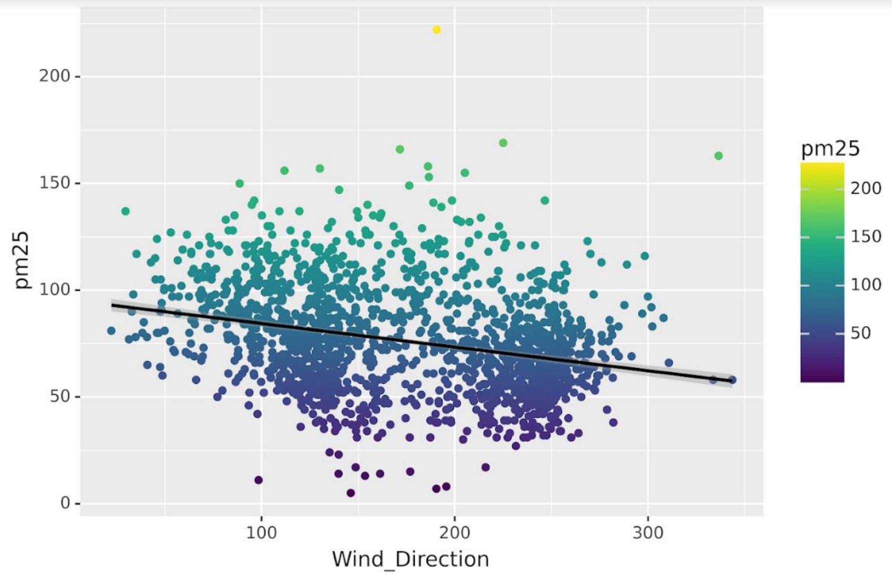


Figure 34

Conclusion:

- Relative Humidity, Specific Humidity, Pressure and Wind Speed factors have interactions with Month and PM25.
- Temperature, Wind Direction and Precipitation factors interact with months but do not interact with pm2.5.

4 Regression Model

After performing One-way ANOVA and Two-way ANOVA tests on weather, air quality as well as natural factors, we proceeded to build linear regression models on factors that have been identified as interacting with each other or having an effect on air quality factors.

4.1 Experiment 1

At the first experiment we built a linear regression model based on all of the factors in the dataset.

```
regressor = lm(formula = pm25 ~
                              YEAR_1 + MO_1 + DY_1 + pm25_1 + Temperature_1 +
Relative_Humidity_1+

Specific_Humidity_1+Precipitation_1+Pressure_1+Wind_Speed_1+Wind_Direction_1+

                              YEAR_2 + MO_2 + DY_2 + pm25_2 + Temperature_2 +
Relative_Humidity_2+

Specific_Humidity_2+Precipitation_2+Pressure_2+Wind_Speed_2+Wind_Direction_2+

                              YEAR_3 + MO_3 + DY_3 + pm25_3 + Temperature_3 +
Relative_Humidity_3+

Specific_Humidity_3+Precipitation_3+Pressure_3+Wind_Speed_3+Wind_Direction_3,
          data = training_set)
```

Result:
RMSE = 13.62056
MAE = 10.53653
MAD = 12.72094

4.2 Experiment 2

At the second experiment we built a linear regression model based on the hypothesis that Month and pm25 interacted with Relative Humidity.

```
regressor = lm(formula = pm25 ~ (MO_1+pm25_1)*Relative_Humidity_1
            + (MO_2+pm25_2)*Relative_Humidity_2
            + (MO_3+pm25_3)*Relative_Humidity_3 ,
            data = training_set)
```

Result:
RMSE = 13.91894
MAE = 10.4454
MAD = 11.83975

4.3 Experiment 3

At experiment 3,we built a linear regression model based on the hypothesis that Month and pm25 interacted with Specific Humidity.

```
regressor = lm(formula = pm25 ~ (MO_1 + pm25_1)*Specific_Humidity_1
            + (MO_2+pm25_2)*Specific_Humidity_2
            + (MO_3+pm25_3)*Specific_Humidity_3,
            data = training_set)
```

Result:
RMSE = 13.72975
MAE = 10.43576
MAD = 12.12984

4.4 Experiment 4

At the 4th experiment, we built a linear regression model based on the hypothesis that Month and pm25 have interaction with Pressure.

```
regressor = lm(formula = pm25 ~ (MO_1 + pm25_1)*Pressure_1
            + (MO_2+pm25_2)*Pressure_2
            + (MO_3+pm25_3)*Pressure_3,
            data = training_set)
```

Result:
RMSE = 13.94949
MAE = 10.67646
MAD = 12.44928

4.5 Experiment 5

At the 5th experiment, we built a linear regression model based on the hypothesis that Month and pm25 have interaction with Wind Speed.

```
regressor = lm(formula = pm25 ~ MO_1*Wind_Speed_1
            + MO_2*Wind_Speed_2
            + MO_3*Wind_Speed_3,
            data = training_set)
```

Result:
RMSE = 19.52017
MAE = 15.5655
MAD = 19.57665

4.6 Experiment 6

At the 6th experiment we built a linear regression model that was built on the hypothesis of there is an interaction between Month, PM25 with Relative Humidity, Specific Humidity, Pressure, Wind Speed factors.

```
regressor = lm(formula = pm25 ~
                    (MO_1 + pm25_2)*(Relative_Humidity_1 + Specific_Humidity_1 +
Pressure_1 + Wind_Speed_1) +

                    (MO_2 + pm25_2)*(Relative_Humidity_2 + Specific_Humidity_2 +
Pressure_2 + Wind_Speed_2) +

                    (MO_3 + pm25_3)*(Relative_Humidity_3 + Specific_Humidity_3 +
Pressure_3 + Wind_Speed_3),
            data = training_set)
```

Result:
RMSE = 15.51062
MAE = 12.04725
MAD = 14.48284

4.7 Experiment 7

At the 7th experiment we built a linear regression model based on the hypothesis that there is a interaction between Months, PM25 with Relative Humidity, Specific Humidity, Pressure, Wind Speed, and interaction between Month and PM25.

```
regressor = lm(formula = pm25 ~
                    (MO_1* pm25_1) + (MO_1 + pm25_2)*(Relative_Humidity_1 +
Specific_Humidity_1 + Pressure_1 + Wind_Speed_1) +

                    (MO_2* pm25_2) + (MO_2 + pm25_2)*(Relative_Humidity_2 +
Specific_Humidity_2 + Pressure_2 + Wind_Speed_2) +

                    (MO_3* pm25_3) + (MO_3 + pm25_3)*(Relative_Humidity_3 +
Specific_Humidity_3 + Pressure_3 + Wind_Speed_3),
            data = training_set)
```

Result:
RMSE = 13.57332
MAE = 10.48986
MAD = 12.7695

4.8 Experiment 8

At the 8th experiment we built a linear regression model based on all of the factors in the dataset. (Except for these factors: Day, Year, Temperature, Wind Speed and Precipitation).

```
regressor = lm(formula = pm25 ~

              MO_1 + pm25_1 + Relative_Humidity_1+
              Specific_Humidity_1+Pressure_1+Wind_Speed_1+


              MO_2 + pm25_2 + Relative_Humidity_2+
              Specific_Humidity_2+Pressure_2+Wind_Speed_2+


              MO_3 + pm25_3 + Relative_Humidity_3+
              Specific_Humidity_3+Pressure_3+Wind_Speed_3,
            data = training_set)
```

Result:
RMSE = 13.65421
MAE = 10.5766
MAD = 12.40426

4.9 Experiment 9

At the 9th experiment we built a linear regression model based on all of the features in the dataset. (Except for these factors: Day, Year, Temperature, Wind Speed, Precipitation) and the interaction between Month and pm25.

```
regressor = lm(formula = pm25 ~
         MO_1 + pm25_1 + Relative_Humidity_1+
         Specific_Humidity_1+Pressure_1+Wind_Speed_1+ (pm25_1*MO_1) +

         MO_2 + pm25_2 + Relative_Humidity_2+
         Specific_Humidity_2+Pressure_2+Wind_Speed_2+ (pm25_2*MO_2) +

         MO_3 + pm25_3 + Relative_Humidity_3+
         Specific_Humidity_3+Pressure_3+Wind_Speed_3 + (pm25_3*MO_3),
        data = training_set)
```

Result:

RMSE = 13.62631
MAE = 10.56236
MAD = 12.50514

5 Experimental results

5.1 Experiment 1

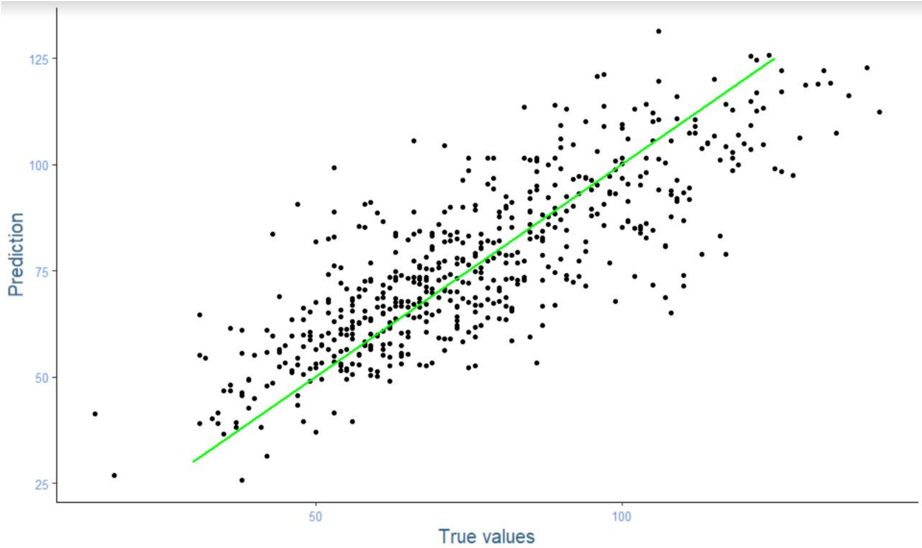


Figure 35. Experiment 1 result

```
Residuals:
    Min      1Q  Median      3Q     Max
-83.517  -9.074  -0.171   8.811  85.165

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          3.700e+03  8.350e+02   4.432 1.03e-05 ***
YEAR_1              -3.502e+01  2.425e+02  -0.144  0.88520
MO_1                -3.720e+00  2.015e+01  -0.185  0.85355
DY_1                 6.549e-02  6.754e-01   0.097  0.92277
pm25_1               6.008e-01  3.218e-02  18.673  < 2e-16 ***
Temperature_1       -1.445e+00  2.290e+00  -0.631  0.52804
Relative_Humidity_1  5.322e-01  6.386e-01   0.833  0.40481
Specific_Humidity_1 -3.518e+00  2.400e+00  -1.466  0.14306
Precipitation_1     -1.689e-02  7.049e-02  -0.240  0.81069
Pressure_1           1.417e+01  6.249e+00   2.267  0.02356 *
Wind_Speed_1        -4.929e+00  7.513e-01  -6.561 8.16e-11 ***
Wind_Direction_1    -3.694e-02  1.298e-02  -2.845  0.00452 **
YEAR_2               3.848e+02  3.347e+02   1.150  0.25056
MO_2                 3.293e+01  2.783e+01   1.183  0.23695
DY_2                 8.357e-01  9.336e-01   0.895  0.37091
pm25_2               5.251e-03  3.712e-02   0.141  0.88751
Temperature_2        1.294e+00  2.628e+00   0.493  0.62245
Relative_Humidity_2 -2.800e-01  7.130e-01  -0.393  0.69456
Specific_Humidity_2  1.465e+00  2.650e+00   0.553  0.58052
Precipitation_2      1.380e-01  7.453e-02   1.851  0.06440 .
Pressure_2          -8.776e+00  8.882e+00  -0.988  0.32331
Wind_Speed_2         1.357e+00  9.090e-01   1.493  0.13574
Wind_Direction_2     5.897e-03  1.516e-02   0.389  0.69740
YEAR_3              -3.514e+02  2.372e+02  -1.481  0.13879
MO_3                -2.937e+01  1.973e+01  -1.489  0.13675
DY_3                -9.367e-01  6.596e-01  -1.420  0.15584
pm25_3               5.361e-02  3.150e-02   1.702  0.08905 .
Temperature_3       -1.126e+00  2.315e+00  -0.486  0.62686
Relative_Humidity_3 -2.514e-01  6.234e-01  -0.403  0.68676
Specific_Humidity_3  1.312e+00  2.359e+00   0.556  0.57817
Precipitation_3     -1.147e-01  6.329e-02  -1.812  0.07031 .
Pressure_3          -8.958e+00  6.470e+00  -1.385  0.16647
Wind_Speed_3         1.136e+00  7.873e-01   1.443  0.14944
Wind_Direction_3    -9.905e-03  1.362e-02  -0.727  0.46713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.36 on 1125 degrees of freedom
Multiple R-squared:  0.6454,	Adjusted R-squared:  0.635
F-statistic: 62.05 on 33 and 1125 DF,  p-value: < 2.2e-16
```

Figure 36

5.2 Experiment 2

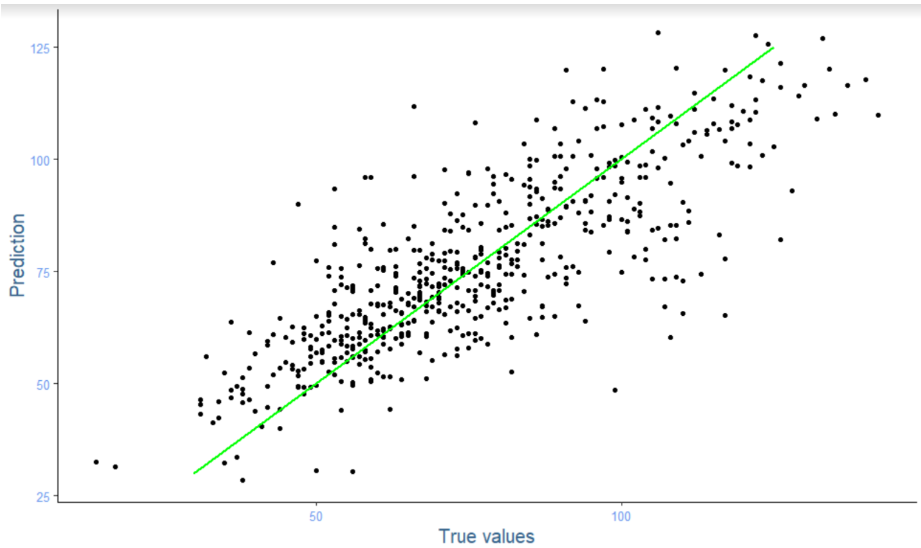


Figure 37. Experiment 2 result

```
Residuals:
    Min      1Q  Median      3Q     Max 
-90.160  -9.558  -0.789   9.439  79.074 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                23.865413  14.606407   1.634   0.1026    
MO_1                       -7.317615   3.205787  -2.283   0.0226 *  
pm25_1                      1.080561   0.186511   5.794  8.9e-09 ***
Relative_Humidity_1        -0.190156   0.336097  -0.566   0.5717    
MO_2                       -1.130098   4.245968  -0.266   0.7902    
pm25_2                     -0.266942   0.247356  -1.079   0.2807    
Relative_Humidity_2        -0.397291   0.438180  -0.907   0.3648    
MO_3                        7.069855   3.491267   2.025   0.0431 *  
pm25_3                     -0.027448   0.198910  -0.138   0.8903    
Relative_Humidity_3         0.454806   0.347811   1.308   0.1913    
MO_1:Relative_Humidity_1    0.076132   0.039347   1.935   0.0533 .  
pm25_1:Relative_Humidity_1 -0.004532   0.002349  -1.929   0.0540 .  
MO_2:Relative_Humidity_2    0.023882   0.051914   0.460   0.6456    
pm25_2:Relative_Humidity_2  0.003390   0.003067   1.105   0.2693    
MO_3:Relative_Humidity_3   -0.082091   0.041267  -1.989   0.0469 *  
pm25_3:Relative_Humidity_3  0.001511   0.002489   0.607   0.5440    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.16 on 1143 degrees of freedom
Multiple R-squared:  0.6012,	Adjusted R-squared:  0.596 
F-statistic: 114.9 on 15 and 1143 DF,  p-value: < 2.2e-16
```

Figure 38

5.3 Experiment 3

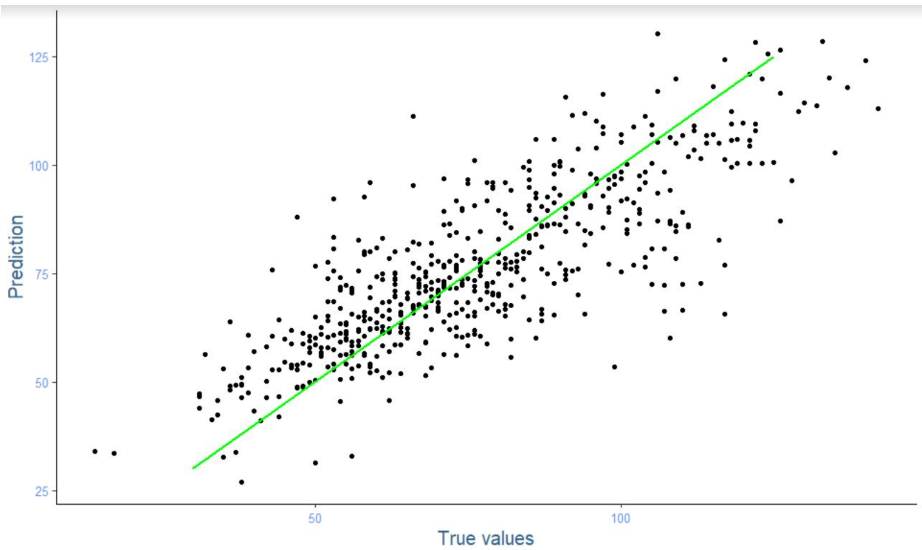


Figure 39. Experiment 3 result

```
Residuals:
    Min      1Q  Median      3Q     Max 
-90.168  -9.176  -1.002   9.188  82.575 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                21.410551  16.495600   1.298  0.19456    
MO_1                       -4.068011   2.406348  -1.691  0.09120 .  
pm25_1                      1.274401   0.200868   6.344 3.21e-10 ***
Specific_Humidity_1        -0.659660   1.423161  -0.464  0.64308    
MO_2                       -0.103812   3.312062  -0.031  0.97500    
pm25_2                     -0.500459   0.254308  -1.968  0.04932 *  
Specific_Humidity_2        -1.670243   1.816237  -0.920  0.35797    
MO_3                        3.199904   2.490017   1.285  0.19902    
pm25_3                      0.157405   0.210859   0.746  0.45552    
Specific_Humidity_3         1.895736   1.494800   1.268  0.20498    
MO_1:Specific_Humidity_1    0.201066   0.143939   1.397  0.16272    
pm25_1:Specific_Humidity_1 -0.032857   0.011369  -2.890  0.00392 ** 
MO_2:Specific_Humidity_2    0.070232   0.203912   0.344  0.73059    
pm25_2:Specific_Humidity_2  0.028584   0.014207   2.012  0.04445 *  
MO_3:Specific_Humidity_3   -0.197009   0.149331  -1.319  0.18734    
pm25_3:Specific_Humidity_3 -0.004147   0.011932  -0.348  0.72822    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.01 on 1143 degrees of freedom
Multiple R-squared:  0.6087,	Adjusted R-squared:  0.6035 
F-statistic: 118.5 on 15 and 1143 DF,  p-value: < 2.2e-16
```

Figure 40

5.4 Experiment 4

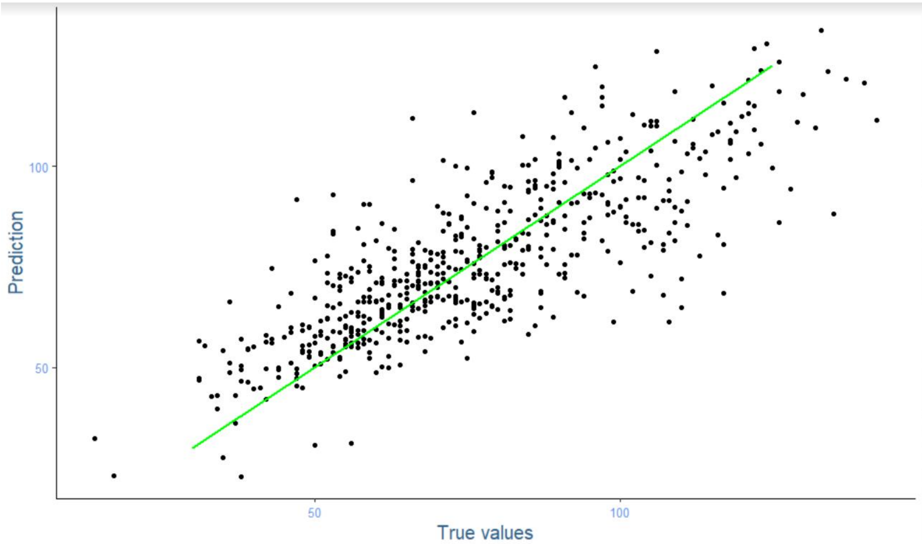


Figure 41. Experiment 4 result

```
Residuals:
    Min      1Q  Median      3Q     Max 
-85.484  -9.247  -0.616   9.232  78.074 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)  
(Intercept)       -965.1969  1267.7966  -0.761   0.4466  
MO_1                82.3126   176.8147   0.466   0.6416  
pm25_1              -3.4799    14.0226  -0.248   0.8041  
Pressure_1          33.5780    17.1109   1.962   0.0500 *
MO_2               356.9940   263.8394   1.353   0.1763  
pm25_2              16.0375    17.3399   0.925   0.3552  
Pressure_2          21.6581    23.7885   0.910   0.3628  
MO_3              -336.5472   181.9351  -1.850   0.0646 .
pm25_3             -20.1563    15.0104  -1.343   0.1796  
Pressure_3         -45.4714    17.9287  -2.536   0.0113 *
MO_1:Pressure_1     -0.8326     1.7583  -0.474   0.6359  
pm25_1:Pressure_1    0.0416     0.1396   0.298   0.7658  
MO_2:Pressure_2     -3.5339     2.6219  -1.348   0.1780  
pm25_2:Pressure_2   -0.1596     0.1727  -0.924   0.3555  
MO_3:Pressure_3      3.3448     1.8078   1.850   0.0645 .
pm25_3:Pressure_3    0.2015     0.1495   1.348   0.1779  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.95 on 1143 degrees of freedom
Multiple R-squared:  0.6118,	Adjusted R-squared:  0.6067 
F-statistic: 120.1 on 15 and 1143 DF,  p-value: < 2.2e-16
```

Figure 42

5.5 Experiment 5

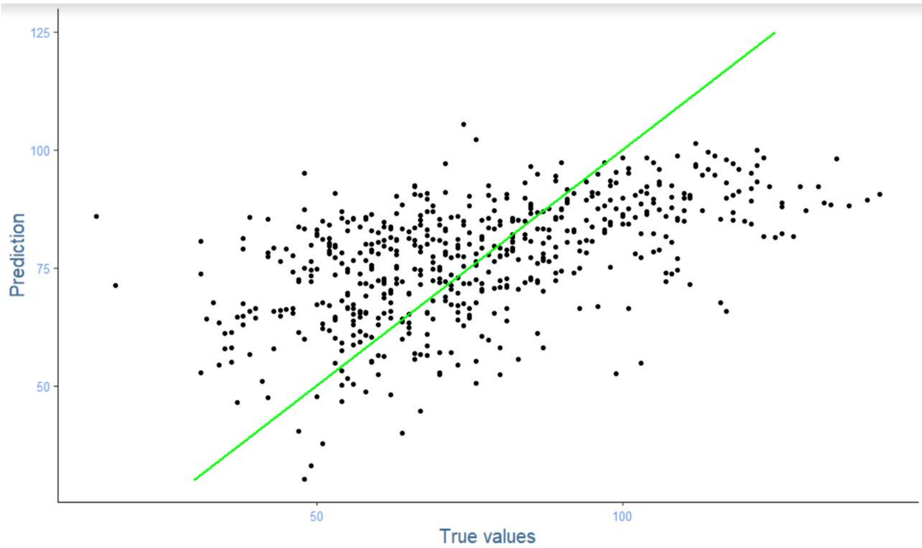


Figure 43. Experiment 5 result

```
Residuals:
    Min      1Q  Median      3Q     Max
-77.522 -14.640  -0.859  13.719 148.906

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       113.0160     4.3024  26.268  < 2e-16 ***
MO_1               -3.3207     1.5627  -2.125   0.0338 *
Wind_Speed_1      -11.1628     2.0359  -5.483 5.14e-08 ***
MO_2                1.8454     2.2304   0.827   0.4082
Wind_Speed_2       -0.1901     2.6698  -0.071   0.9433
MO_3                1.3172     1.6172   0.814   0.4155
Wind_Speed_3        1.3808     2.1369   0.646   0.5183
MO_1:Wind_Speed_1   0.1301     0.2713   0.480   0.6316
MO_2:Wind_Speed_2  -0.0755     0.3589  -0.210   0.8334
MO_3:Wind_Speed_3  -0.3725     0.2782  -1.339   0.1808
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.19 on 1149 degrees of freedom
Multiple R-squared:  0.2445,	Adjusted R-squared:  0.2386
F-statistic: 41.31 on 9 and 1149 DF,  p-value: < 2.2e-16
```

Figure 44

5.6 Experiment 6

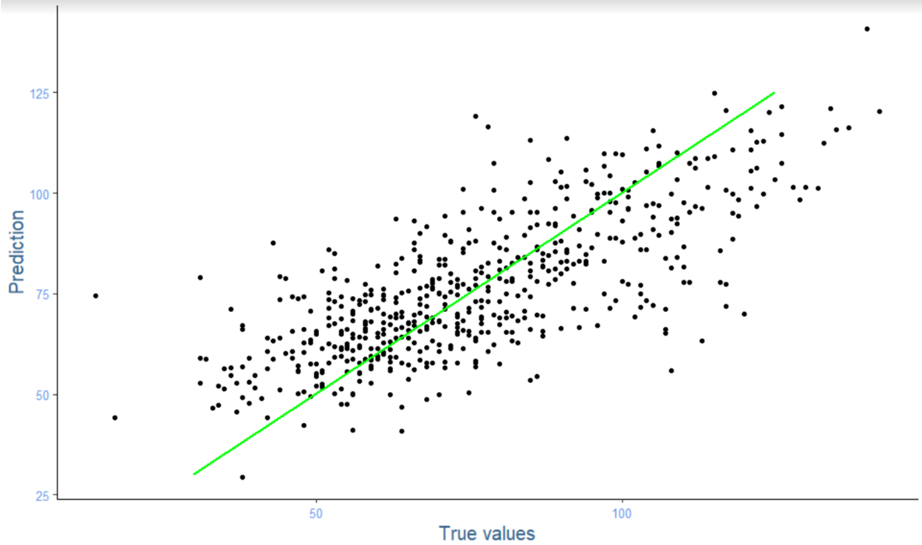


Figure 45. Experiment 6 result

```
Residuals:
    Min      1Q  Median      3Q     Max
-74.035 -11.004  -0.503  10.596 108.664

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -1.581e+03  2.246e+03  -0.704  0.48156
MO_1                       5.806e+01  2.146e+02   0.271  0.78678
pm25_2                     2.891e+01  2.317e+01   1.247  0.21252
Relative_Humidity_1       -6.836e-01  8.485e-01  -0.806  0.42058
Specific_Humidity_1        9.763e-01  3.584e+00   0.272  0.78539
Pressure_1                 4.298e+01  2.616e+01   1.643  0.10067
Wind_Speed_1              -2.103e+00  3.193e+00  -0.659  0.51012
MO_2                       2.507e+02  3.100e+02   0.809  0.41874
Relative_Humidity_2        9.528e-01  9.783e-01   0.974  0.33030
Specific_Humidity_2       -6.087e+00  4.050e+00  -1.503  0.13311
Pressure_2                 1.808e+01  3.277e+01   0.552  0.58121
Wind_Speed_2              -2.668e+00  3.600e+00  -0.741  0.45879
MO_3                      -3.763e+02  2.241e+02  -1.679  0.09341 .
pm25_3                    -6.912e+00  2.248e+01  -0.308  0.75850
Relative_Humidity_3       -3.755e-01  5.939e-01  -0.632  0.52738
Specific_Humidity_3        3.738e+00  2.818e+00   1.326  0.18497
Pressure_3                -4.458e+01  2.394e+01  -1.862  0.06291 .
Wind_Speed_3               2.368e+00  2.923e+00   0.810  0.41804
MO_1:Relative_Humidity_1   3.949e-02  7.155e-02   0.552  0.58115
MO_1:Specific_Humidity_1   1.384e-01  2.678e-01   0.517  0.60531
MO_1:Pressure_1           -6.406e-01  2.129e+00  -0.301  0.76360
MO_1:Wind_Speed_1          1.864e-02  2.483e-01   0.075  0.94017
pm25_2:Relative_Humidity_1 1.282e-02  8.206e-03   1.562  0.11854
pm25_2:Specific_Humidity_1 -8.324e-02 3.444e-02  -2.417  0.01580 *
pm25_2:Pressure_1         -1.805e-01  2.716e-01  -0.665  0.50635
pm25_2:Wind_Speed_1       -9.323e-02  3.487e-02  -2.674  0.00761 **
MO_2:Relative_Humidity_2   3.045e-02  8.808e-02   0.346  0.72966
MO_2:Specific_Humidity_2   1.197e-01  3.375e-01   0.355  0.72296
MO_2:Pressure_2           -2.534e+00  3.080e+00  -0.823  0.41074
MO_2:Wind_Speed_2          7.040e-02  3.101e-01   0.227  0.82043
pm25_2:Relative_Humidity_2 -1.187e-02 8.839e-03  -1.342  0.17978
pm25_2:Specific_Humidity_2 6.338e-02  3.636e-02   1.743  0.08164 .
pm25_2:Pressure_2         -9.940e-02  3.118e-01  -0.319  0.74994
```


```
MO_2:Wind_Speed_2          7.040e-02  3.101e-01   0.227  0.82043
pm25_2:Relative_Humidity_2 -1.187e-02 8.839e-03  -1.342  0.17978
pm25_2:Specific_Humidity_2 6.338e-02  3.636e-02   1.743  0.08164 .
pm25_2:Pressure_2         -9.940e-02  3.118e-01  -0.319  0.74994
```


```
pm25_2:Wind_Speed_2        4.546e-02  3.767e-02   1.207  0.22782
MO_3:Relative_Humidity_3  -1.345e-01  6.809e-02  -1.975  0.04847 *
MO_3:Specific_Humidity_3   8.617e-02  2.479e-01   0.348  0.72823
MO_3:Pressure_3            3.835e+00  2.221e+00   1.727  0.08445 .
MO_3:Wind_Speed_3          2.243e-01  2.576e-01   0.871  0.38402
pm25_3:Relative_Humidity_3 7.083e-03  4.422e-03   1.602  0.10949
pm25_3:Specific_Humidity_3 -2.043e-02 2.456e-02  -0.832  0.40559
pm25_3:Pressure_3          6.793e-02  2.219e-01   0.306  0.75955
pm25_3:Wind_Speed_3        1.986e-03  2.918e-02   0.068  0.94574
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.83 on 1117 degrees of freedom
Multiple R-squared:  0.5256,	Adjusted R-squared:  0.5082
F-statistic: 30.18 on 41 and 1117 DF,  p-value: < 2.2e-16
```

Figure 46

5.7 Experiment 7

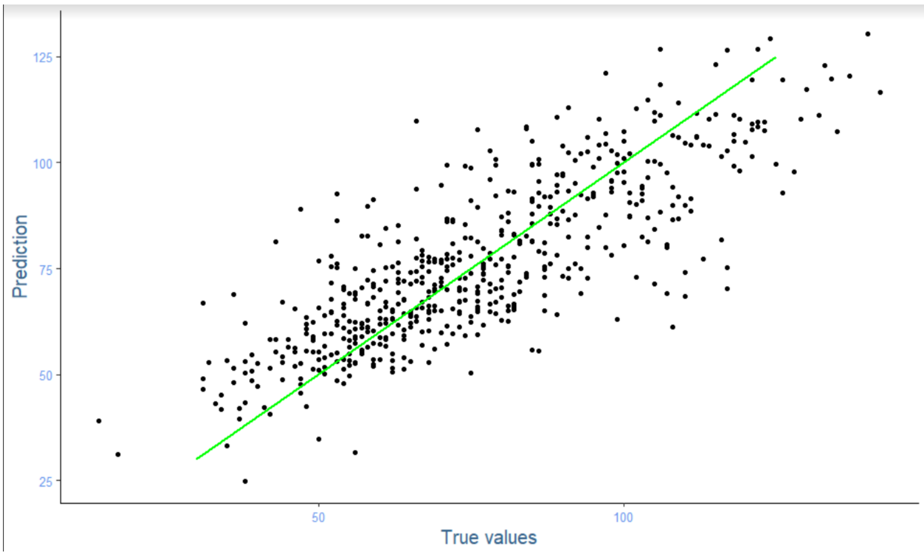


Figure 47. Experiment 7 result

```
Residuals:
    Min      1Q  Median      3Q     Max 
-81.396  -9.427  -0.725   9.018  86.336 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 -1.043e+03  1.961e+03  -0.532  0.59482    
MO_1                         4.980e+01  1.871e+02   0.266  0.79016    
pm25_1                       6.270e-01  7.186e-02   8.726  < 2e-16 ***
pm25_2                       1.427e+01  2.042e+01   0.699  0.48493    
Relative_Humidity_1         -3.928e-01  7.468e-01  -0.526  0.59897    
Specific_Humidity_1          1.522e+00  3.150e+00   0.483  0.62900    
Pressure_1                   4.621e+01  2.285e+01   2.022  0.04338 *  
Wind_Speed_1                -1.433e+00  2.825e+00  -0.507  0.61196    
MO_2                         2.568e+02  2.702e+02   0.950  0.34216    
Relative_Humidity_2          2.861e-01  8.583e-01   0.333  0.73893    
Specific_Humidity_2         -4.633e+00  3.563e+00  -1.300  0.19378    
Pressure_2                  -3.184e+00  2.865e+01  -0.111  0.91153    
Wind_Speed_2                -7.206e-01  3.174e+00  -0.227  0.82046    
MO_3                        -3.140e+02  1.956e+02  -1.605  0.10868    
pm25_3                      -8.277e-01  1.970e+01  -0.042  0.96649    
Relative_Humidity_3         -5.503e-04  5.317e-01  -0.001  0.99917    
Specific_Humidity_3          2.126e+00  2.477e+00   0.858  0.39094    
Pressure_3                  -3.214e+01  2.094e+01  -1.534  0.12520    
Wind_Speed_3                 8.265e-01  2.632e+00   0.314  0.75353    
MO_1:pm25_1                 -2.173e-03  9.244e-03  -0.235  0.81417    
MO_1:Relative_Humidity_1     5.904e-02  6.248e-02   0.945  0.34492    
MO_1:Specific_Humidity_1     8.509e-02  2.341e-01   0.364  0.71625    
MO_1:Pressure_1             -5.686e-01  1.857e+00  -0.306  0.75951    
MO_1:Wind_Speed_1            8.455e-02  2.305e-01   0.367  0.71379    
pm25_2:Relative_Humidity_1   9.232e-03  7.243e-03   1.275  0.20273    
pm25_2:Specific_Humidity_1  -7.407e-02  3.033e-02  -2.442  0.01476 *  
pm25_2:Pressure_1           -2.585e-01  2.373e-01  -1.089  0.27637    
pm25_2:Wind_Speed_1         -5.165e-02  3.060e-02  -1.688  0.09173 .  
pm25_2:MO_2                  5.716e-03  1.215e-02   0.470  0.63813    
MO_2:Relative_Humidity_2     4.165e-02  7.737e-02   0.538  0.59042    
MO_2:Specific_Humidity_2    -6.680e-02  2.976e-01  -0.224  0.82244    
MO_2:Pressure_2             -2.569e+00  2.685e+00  -0.957  0.33879    
MO_2:Wind_Speed_2            1.307e-02  2.778e-01   0.047  0.96248    
pm25_2:Relative_Humidity_2  -1.123e-02  7.724e-03  -1.454  0.14612    
pm25_2:Specific_Humidity_2   8.465e-02  3.204e-02   2.642  0.00835 ** 
pm25_2:Pressure_2            1.168e-01  2.729e-01   0.428  0.66874    
```

```
pm25_2:Wind_Speed_2           2.098e-02  3.288e-02   0.638  0.52358
MO_3:pm25_3                  -2.991e-03  1.066e-02  -0.281  0.77908
MO_3:Relative_Humidity_3     -1.278e-01  5.985e-02  -2.136  0.03293 *
MO_3:Specific_Humidity_3      1.624e-01  2.220e-01   0.732  0.46459
MO_3:Pressure_3               3.198e+00  1.938e+00   1.650  0.09913 .
MO_3:Wind_Speed_3             1.006e-01  2.394e-01   0.420  0.67441
pm25_3:Relative_Humidity_3    7.140e-03  4.128e-03   1.730  0.08396 .
pm25_3:Specific_Humidity_3   -3.248e-02  2.148e-02  -1.512  0.13083
pm25_3:Pressure_3             9.091e-03  1.944e-01   0.047  0.96272
pm25_3:Wind_Speed_3           5.449e-03  2.568e-02   0.212  0.83199
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.54 on 1113 degrees of freedom
Multiple R-squared:  0.641,	Adjusted R-squared:  0.6265
F-statistic: 44.17 on 45 and 1113 DF,  p-value: < 2.2e-16
```

Figure 48

5.8 Experiment 8

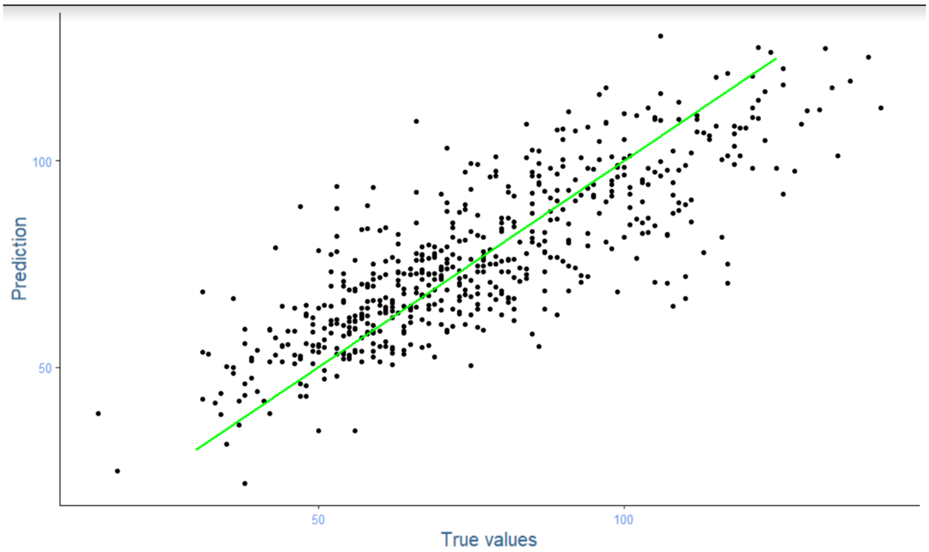


Figure 49. Experiment 8 result

```
Residuals:
    Min      1Q  Median      3Q     Max
-82.849  -9.197  -0.686   9.107  88.208

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           6.61332  434.57106   0.015 0.987861
MO_1                 -1.12790    0.95146  -1.185 0.236088
pm25_1                0.61900    0.03203  19.328  < 2e-16 ***
Relative_Humidity_1   0.77149    0.20139   3.831 0.000135 ***
Specific_Humidity_1  -4.77452    0.88397  -5.401 8.05e-08 ***
Pressure_1           20.79284    6.00689   3.461 0.000557 ***
Wind_Speed_1         -4.79859    0.74950  -6.402 2.23e-10 ***
MO_2                  1.14902    1.33924   0.858 0.391092
pm25_2                0.02743    0.03714   0.739 0.460332
Relative_Humidity_2  -0.39392    0.25360  -1.553 0.120623
Specific_Humidity_2   2.25359    1.17575   1.917 0.055523 .
Pressure_2          -14.07034    8.76518  -1.605 0.108715
Wind_Speed_2          1.40517    0.91326   1.539 0.124172
MO_3                 -0.04318    0.96287  -0.045 0.964241
pm25_3                0.07206    0.03148   2.289 0.022259 *
Relative_Humidity_3  -0.17827    0.19201  -0.928 0.353378
Specific_Humidity_3   0.76850    0.86753   0.886 0.375886
Pressure_3           -6.36405    6.41417  -0.992 0.321317
Wind_Speed_3          1.42453    0.77929   1.828 0.067814 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.58 on 1140 degrees of freedom
Multiple R-squared:  0.6302,	Adjusted R-squared:  0.6244
F-statistic: 107.9 on 18 and 1140 DF,  p-value: < 2.2e-16
```

Figure 50

5.9 Experiment 9

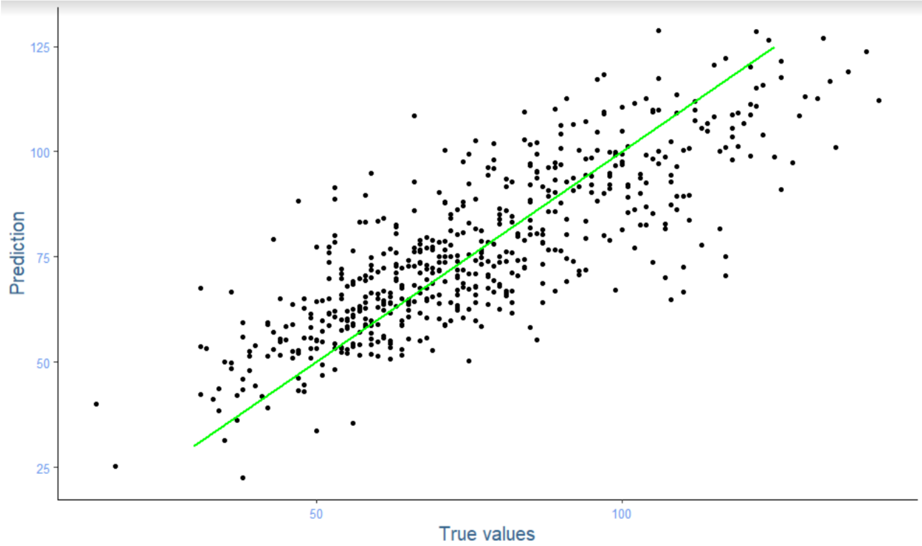
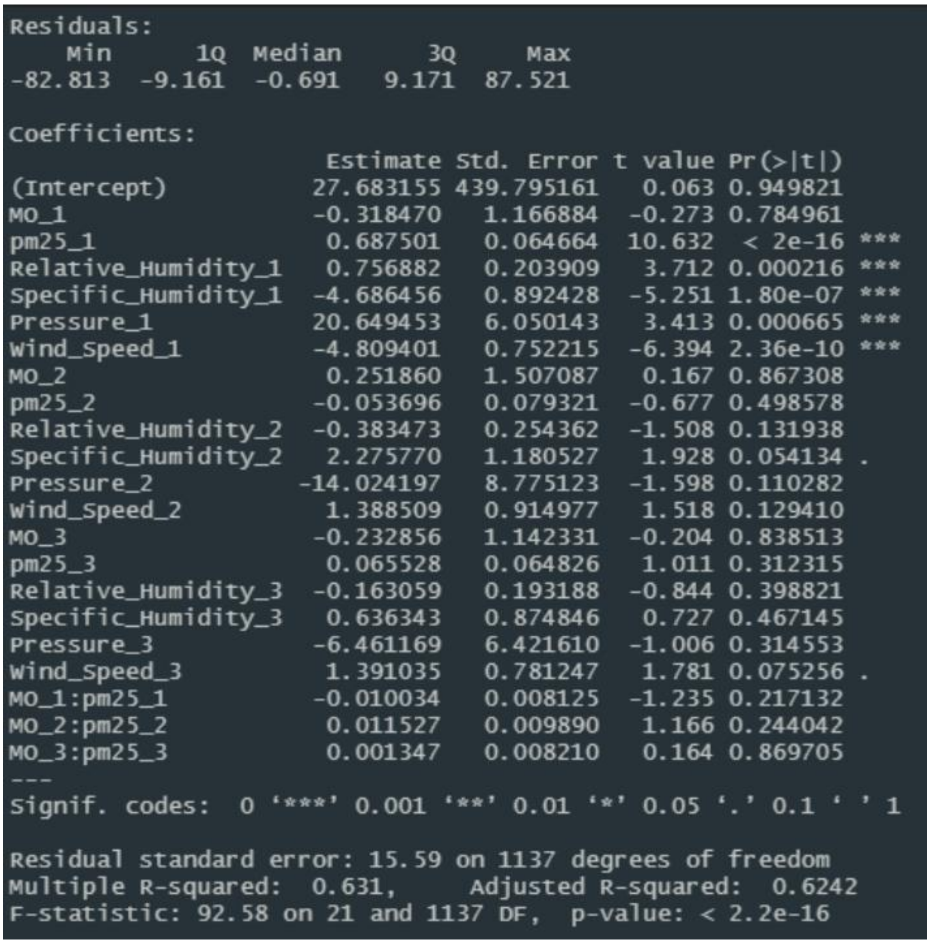


Figure 51. Experiment 9 result



```
Residuals:
    Min      1Q  Median      3Q     Max
-82.813  -9.161  -0.691   9.171  87.521

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          27.683155 439.795161   0.063 0.949821
MO_1                 -0.318470   1.166884  -0.273 0.784961
pm25_1                0.687501   0.064664  10.632  < 2e-16 ***
Relative_Humidity_1   0.756882   0.203909   3.712 0.000216 ***
Specific_Humidity_1  -4.686456   0.892428  -5.251 1.80e-07 ***
Pressure_1           20.649453   6.050143   3.413 0.000665 ***
Wind_Speed_1         -4.809401   0.752215  -6.394 2.36e-10 ***
MO_2                  0.251860   1.507087   0.167 0.867308
pm25_2               -0.053696   0.079321  -0.677 0.498578
Relative_Humidity_2  -0.383473   0.254362  -1.508 0.131938
Specific_Humidity_2   2.275770   1.180527   1.928 0.054134 .
Pressure_2          -14.024197   8.775123  -1.598 0.110282
Wind_Speed_2          1.388509   0.914977   1.518 0.129410
MO_3                 -0.232856   1.142331  -0.204 0.838513
pm25_3                0.065528   0.064826   1.011 0.312315
Relative_Humidity_3  -0.163059   0.193188  -0.844 0.398821
Specific_Humidity_3   0.636343   0.874846   0.727 0.467145
Pressure_3           -6.461169   6.421610  -1.006 0.314553
Wind_Speed_3          1.391035   0.781247   1.781 0.075256 .
MO_1:pm25_1          -0.010034   0.008125  -1.235 0.217132
MO_2:pm25_2           0.011527   0.009890   1.166 0.244042
MO_3:pm25_3           0.001347   0.008210   0.164 0.869705
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.59 on 1137 degrees of freedom
Multiple R-squared:  0.631,     Adjusted R-squared:  0.6242
F-statistic: 92.58 on 21 and 1137 DF,  p-value: < 2.2e-16
```

Figure 52

Result:

Experiment	RMSE	MAE	MAD	p-value
1	13.62056	10.53653	12.72094	< 2.2e-16
2	13.91894	10.4454	11.83975	< 2.2e-16
3	13.72975	10.43576	12.12984	< 2.2e-16
4	13.94949	10.67646	12.44928	< 2.2e-16
5	19.52017	15.5655	19.57665	< 2.2e-16

6	15.51062	12.04725	14.48284	< 2.2e-16
7	13.57332	10.48986	12.7695	< 2.2e-16
8	13.65421	10.5766	12.40426	< 2.2e-16
9	13.62631	10.56236	12.50514	< 2.2e-16

6 Conclusion

We can see that many factors are removed and added through various analyses and experiments, even in the ANOVA inspections. It shows whether or not these factors have implications for air quality, but when building a linear regression model, the experiments showed that the model's difference is not big, but the reliability is relatively high (p-value < 0.5).

Despite the addition of the regression model of interactions of natural factors and indicators and different interactive combinations, the results have improved. However, the RMSE, MAE, AND MAD measurements are still relatively high, not too other, and the predicted results still have some practical significance.

7 Development

In the future, we plan to use a variety of machine learning and deep learning methods for this problem to improve predictability. We also apply more data visualization and analysis techniques to gain more insight into this dataset.

8 References

1. https://vn.usembassy.gov/embassy-consulate/ho-chi-minh-city/air-quality-monitor/
2. https://www.airnow.gov/international/us-embassies-and-consulates/#Vietnam/Ho_Chi_Minh_City
3. https://www.scribbr.com/statistics/one-way-anova/
4. https://www.scribbr.com/statistics/two-way-anova/
5. https://www.datanovia.com/en/lessons/ancova-in-r/#:~:text=The%20Analysis%20of%20Covariance%20
6. https://power.larc.nasa.gov/data-access-viewer/
7. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html?highlight=linear%20regression#sklearn.linear_model.LinearRegression
8. https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=taskUp&view=table
9. https://scikit-learn.org/stable/