|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Data analytics for Health and Safety in mining sector  Final deliverable |
| |  |  |  | | --- | --- | --- | | MEHARBAN SINGH & Alisha Panday | 11/13/17 | Applied IT Project | |

Contents

[Descriptive Analysis 2](#_Toc498307718)

[Basic features of Data 2](#_Toc498307719)

[Graphical Visualization of data 3](#_Toc498307720)

[Correlation matrix 6](#_Toc498307721)

[Analyzing relation between data variables by drawing graphs 7](#_Toc498307722)

[Prescriptive Analysis 13](#_Toc498307723)

[Linear regression models 22](#_Toc498307724)

[Predict Particulates 22](#_Toc498307725)

[Predict Sulphur-Dioxide 22](#_Toc498307726)

[Predict Carbon-monoxide 23](#_Toc498307727)

[Multiple-Linear regression models 24](#_Toc498307728)

[Predict Particulates 24](#_Toc498307729)

[Predict Sulphur-Dioxide 25](#_Toc498307730)

[Predict Carbon-monoxide 26](#_Toc498307731)

[An Example of multiple linear regression model in python 27](#_Toc498307732)

[Predicted values of Carbon-monoxide with above model 28](#_Toc498307733)

[Residual graph and test values of linear regression models 29](#_Toc498307734)

[Time Series Forecasting 34](#_Toc498307735)

[Time Series graph of Temperature, Humidity, Particulates, SO2, CO 34](#_Toc498307736)

[Time Series graph of data After Normalization 35](#_Toc498307737)

# Descriptive Analysis

## **Basic features of Data**

A screenshot of a cell phone

Description generated with very high confidence

The above data represent some basic features of descriptive analysis. The count represents the number of rows in data (73673) and number of columns in data with their mean, minimum, maximum and their standard Deviation value of each column.

## **Graphical Visualization of data**

This is Histogram graph of temperature.

**A screenshot of a cell phone

Description generated with very high confidence**

This is Histogram graph of Humidity.

**A screenshot of a cell phone

Description generated with very high confidence**

This is Histogram graph of A screenshot of a cell phone

Description generated with very high confidenceSulphur-dioxide.

A screenshot of a cell phone

Description generated with very high confidence

A screenshot of a cell phone

Description generated with very high confidence

The above histogram graph helps us to visualize the data and its trend and now we will analyze this data and will find how this data relate to each other by making correlation matrix and drawing graphs

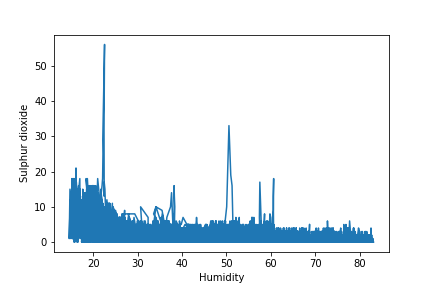
## **Correlation matrix**

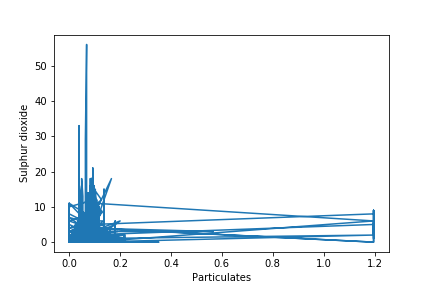
A screenshot of a cell phone

Description generated with high confidence

## A screenshot of a cell phone Description generated with very high confidence **Analyzing relation between data variables by drawing graphs**

A screenshot of a cell phone

Description generated with very high confidenceA screenshot of a cell phone

Description generated with very high confidenceA close up of a logo

Description generated with high confidenceA screenshot of a cell phone

Description generated with very high confidenceA close up of a map

Description generated with high confidenceA screenshot of a cell phone

Description generated with very high confidenceA screenshot of a cell phone

Description generated with very high confidence

After Analyzing these graphs, we can see visually how strongly these variable related to each other.

Now we need to draw line of regression to find what kind of relation these variables have (positive, negative or non-linear relationship)

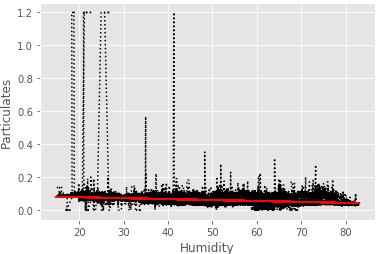
# **Prescriptive Analysis**

# 

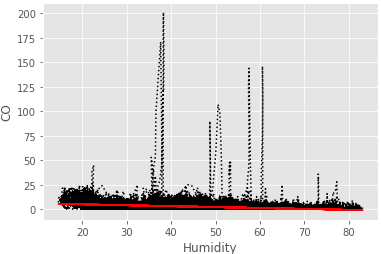
A close up of a map

Description generated with high confidence

This graph and line of regression tell us there is Strong negative relationship between the temperature and humidity. So hence, this conclusion support that if temperature decrease humidity will increase and vice-versa.



This graph and line of regression tell us there is small negative relationship between the Particulates and humidity. So hence, this conclusion support that if Humidity Increase Particulates Will Decrease and vice-versa.



This graph and line of regression tell us there is negative relationship between the Carbon-dioxide and humidity. So hence, this conclusion support that if Humidity Increase Carbon-dioxide value will decrease.

A screenshot of a cell phone

Description generated with high confidence

This graph and line of regression tell us there is Positive relationship between the Carbon-dioxide and Particulates. So hence, this conclusion support that if Particulates Increase Carbon-dioxide value will also increase.

A screenshot of a cell phone

Description generated with high confidence

This graph and line of regression tell us there is Positive relationship between the Particulates and Sulphur-dioxide. So hence, this conclusion support that if Particulates Increase Sulphur-dioxide value will also increase.

A screenshot of a cell phone

Description generated with high confidence

This graph and line of regression tell us there is Positive relationship between the Carbon-monoxide and Sulphur-dioxide. So hence, this conclusion support that if Carbon-monoxide Increase Sulphur-dioxide value will also increase.

A screenshot of a cell phone

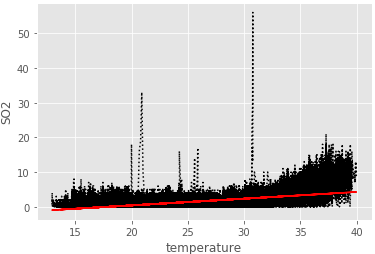
Description generated with high confidence

This graph and line of regression tell us there is Positive relationship between the Carbon-monoxide and Temperature. So hence, this conclusion support that Temperature could be another factor in increase of carbon-monoxide.

A close up of a white wall

Description generated with high confidence

This graph and line of regression tell us there is Positive relationship between the Particulates and Temperature. So hence, this conclusion support that Temperature could be another factor in increase of Particulates.



This graph and line of regression tell us there is Positive relationship between the Sulphur-dioxide and Temperature. So hence, this conclusion support that Temperature could be another factor in increase of Sulphur-dioxide.

## Linear regression models

### **Predict Particulates**

|  |  |
| --- | --- |
| Independent variables | Models |
| Temperature | y = [-0.00075357] + [ 0.00263042]x |
| Humidity | y = [ 0.08939778] + [-0.00056874]x |
| Sulphur-dioxide | y = [ 0.05566477] + [ 0.00405585]x |
| Carbon-monoxide | y = [ 0.05715204] + [ 0.00121544]x |

### **Predict Sulphur-Dioxide**

|  |  |
| --- | --- |
| Independent variables | Models |
| Temperature | y = [-3.47607525] + [ 0.19320779]x |
| Humidity | y = [ 4.0523311] + [-0.05903551]x |
| Particulates | y = [-0.08735171] + [ 17.44996502]x |
| Carbon-monoxide | y = [ 0.52878725] + [ 0.21659056]x |

### **Predict Carbon-monoxide**

|  |  |
| --- | --- |
| Independent variables | Models |
| Temperature | y = [-5.06072711] + [ 0.30597952]x |
| Humidity | y = [ 6.85837899] + [-0.09342701]x |
| Particulates | y = [ 0.58554517] + [ 22.9376002]x |
| Sulphur-dioxide | y = [ 1.0470696] + [ 0.95003701]x |

## Multiple-Linear regression models

### **Predict Particulates**

|  |  |  |
| --- | --- | --- |
| No | Independent variables | Models |
| 1 | Temperature, Humidity | y = [-0.0559] + [0.0039]x + [0.0005]x |
| 2 | Temperature, Sulphur-dioxide | y = [-0.0023] + [0.0027]x + [-0.0005]x |
| 3 | Temperature, Carbon-monoxide | y = [-0.0030] + [0.0028]x + [-0.0004]x |
| 4 | Humidity, Sulphur-dioxide | y = [0.0815] + [-0.0005]x + [0.0020]x |
| 5 | Humidity, Carbon-monoxide | y = [0.0873] + [-0.0005]x + [0.0003]x |
| 6 | Sulphur-dioxide, Carbon-monoxide | y = [0.0552] + [0.0037]x + [0.0004]x |
| 7 | Temperature, Humidity, Sulphur-dioxide | y = [-0.0560] + [0.0039]x + [0.0005]x + [-0.0001]x |
| 8 | Temperature, Humidity, Carbon-monoxide | y = [-0.0565] + [0.0040]x + [0.0040]x + [-0.0003]x |
| 9 | Temperature, Sulphur-dioxide, Carbon-monoxide | y = [ -0.0035] + [ 0.0028]x + [ -0.0002]x + [ -0.0004]x |
| 10 | Humidity, Sulphur-dioxide, Carbon-monoxide | y = [ 0.0813] + [ -0.0005]x + [ 0.0019]x + [ 4.436e-05]x |
| 11 | Temperature, Humidity, Sulphur-dioxide, Carbon-monoxide | y = [-0.0563] + [0.0040]x + [0.0005]x + [0.0001]x + [-0.0004]x |

### **Predict Sulphur-Dioxide**

|  |  |  |
| --- | --- | --- |
| No | Independent variables | Models |
| 1 | Temperature, Humidity | y = [-1.3677] + [0.1451]x + [-0.0191]x |
| 2 | Temperature, Particulates | y = [-3.4774] + [0.1979]x + [-1.7989]x |
| 3 | Temperature, Carbon-monoxide | y = [-2.8515] + [0.1554]x + [0.1234]x |
| 4 | Humidity, Particulates | y = [3.4400] + [-0.0551]x + [6.8500]x |
| 5 | Humidity, Carbon-monoxide | y = [3.0947] + [-0.0460]x + [0.1396]x |
| 6 | Particulates, Carbon-monoxide | y = [ -0.2050] + [ 12.8398]x + [ 0.2010]x |
| 7 | Temperature, Humidity, Particulates | y = [-1.3941] + [0.1469]x + [-0.0189]x + [-0.4720]x |
| 8 | Temperature, Humidity, Carbon-monoxide | y = [ -1.1578] + [ 0.1174]x + [ -0.0155]x + [ 0.1202]x |
| 9 | Temperature, Particulates, Carbon-dioxide | y = [ -2.8535] + [ 0.1573]x + [ -0.6711]x + [ 0.1231]x |
| 10 | Humidity, Particulates, Carbon-dioxide | y = [ 2.5642] + [ -0.0427]x + [ 6.0803]x + [ 0.1377]x |
| 11 | Temperature, Humidity, Particulates, Carbon-monoxide | y = [-1.1348] + [0.1158]x + [-0.0157]x + [0.4075]x + [0.1204]x |

### **Predict Carbon-monoxide**

|  |  |  |
| --- | --- | --- |
| No | Independent variables | Models |
| 1 | Temperature, Humidity | y = [-1.7454] + [0.2303]x + [-0.0301]x |
| 2 | Temperature, Particulates | y = [ -5.0676] + [ 0.3301]x + [ -9.1611]x |
| 3 | Temperature, Sulphur-dioxide | y = [ -2.7985] + [ 0.1802]x + [ 0.6508]x |
| 4 | Humidity, Particulates | y = [6.3588] + [-0.0902]x + [5.5883]x |
| 5 | Humidity, Sulphur-dioxide | y = [ 3.9749] + [ -0.0514]x + [ 0.7116]x |
| 6 | Particulates, Sulphur-Dioxide | y = [0.6661] + [6.8439]x + [0.9223]x |
| 7 | Temperature, Humidity, Particulates | y = [ -2.1536] + [ 0.2587]x + [ -0.0264]x + [ -7.3050]x |
| 8 | Temperature, Humidity, Sulphur-dioxide | y = [ -0.8720] + [ 0.1377]x + [ -0.0179]x + [ 0.6386]x |
| 9 | Temperature, Humidity, Particulates | y = [-2.1536] + [0.2587]x + [-0.0264]x + [ -7.3050]x |
| 10 | Humidity, Particulates, Sulphur-dioxide | y = [ 3.9160] + [ -0.0511]x + [ 0.7239]x + [ 0.7101]x |
| 11 | Temperature, Humidity, Particulates, Sulphur-dioxide | y = [-1.2645] + [0.1650]x + [-0.0144]x + [-7.0040]x + [0.6378]x |

### **An Example of multiple linear regression model in python**

A screenshot of a social media post

Description generated with very high confidence

This is an example of multiple regression model with four independent variables. we can see that Adjusted R-Squared value is very low 0.253 and condition number is very high 2.52e+0.3, so we can conclude that this model is not ideal to predict the value of carbon-monoxide.

### **Predicted values of Carbon-monoxide with above model**

A screenshot of a social media post

Description generated with very high confidence

### **Residual values of Carbon-monoxide**

### A screenshot of a computer Description generated with very high confidence

### A screenshot of a map Description generated with very high confidence Residual graph and test values of linear regression models

This residual graph is between temperature vs predicted value of humidity

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between temperature vs predicted value of Particulates

A screenshot of a cell phone

Description generated with very high confidence

This residual graph is between temperature vs predicted value of Sulphur-Dioxide.

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between temperature vs predicted value of Carbon-monoxide.

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between humidity vs predicted value of Particulates.

A screenshot of a social media post

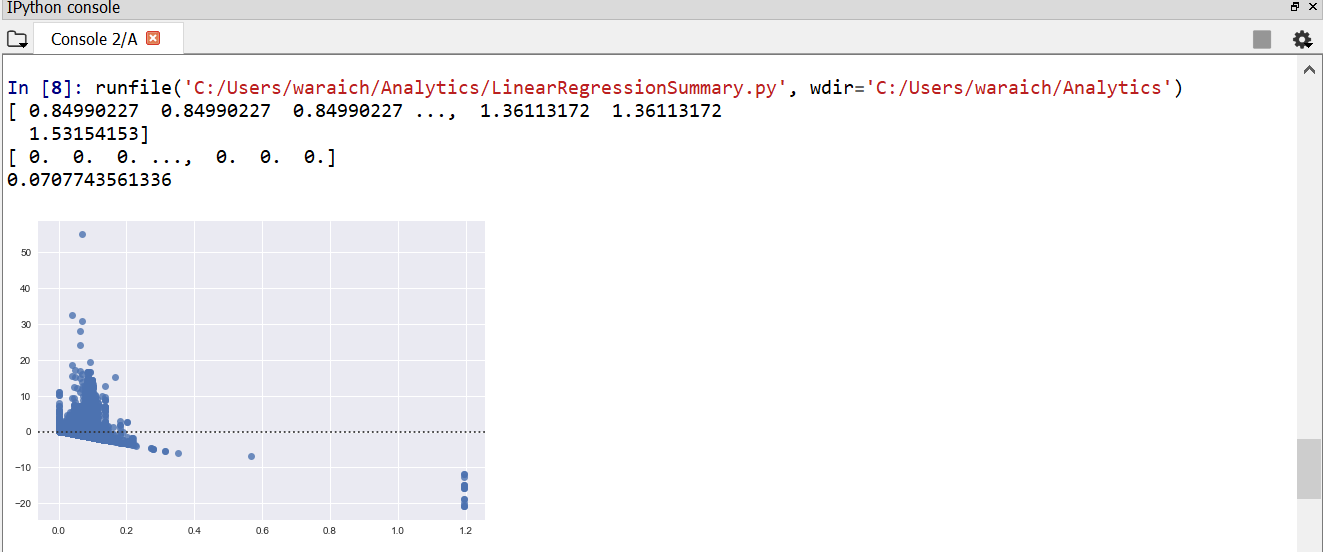
Description generated with very high confidence

This residual graph is between humidity vs predicted value of Sulphur-dioxide.

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between humidity vs predicted value of carbon-monoxide.



This residual graph is between particulates vs predicted value of Sulphur-Dioxide.

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between Particulates vs predicted value of Carbon-monoxide.

A screenshot of a social media post

Description generated with very high confidence

This residual graph is between Sulphur-dioxide vs predicted value of Carbon-monoxide.

# **Time Series Forecasting**

## **Time Series graph of Temperature, Humidity, Particulates, SO2, CO**

A close up of a device

Description generated with high confidence

## **Time Series graph of data After Normalization**

A screenshot of a cell phone

Description generated with very high confidence