Taniya Ahmed

21BDS0059

1. **Z – Score method**

   CODE:

   z_score = scale(df, center = TRUE, scale = TRUE)

   outlier = which(abs(z_score) > 3, arr.ind = TRUE)

   outlier

   OUTPUT:

   ```
   > outlier
         row col
   [1,]  62    1
   [2,] 117    1
   [3,]  48    3
   > print("Taniya Ahmed 21BDS0059")
   [1] "Taniya Ahmed 21BDS0059"
   >
   ```

2. **Inter quartile method**

   CODE:

   outlier_iqr = apply(df, 2, function(col){

    if(is.numeric(col)){

      q1 = quantile(col, 0.25)

      q3 = quantile(col, 0.75)

      iqr = q3 - q1

      lower_bound = q1 - 1.5 * iqr

      upper_bound = q3 + 1.5 * iqr

      return(which(col < lower_bound | col > upper_bound))

    }

   })

outlier_iqr

print("Taniya Ahmed 21BDS0059")

OUTPUT:
```
> outlier_iqr
$Ozone
 [1]  30  62  69  70  71  86  99 100 101 117 121 122 123 124 127

$Solar.R
integer(0)

$Wind
[1]  9 18 48

$Temp
integer(0)

$Month
integer(0)

$Day
integer(0)

> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
```
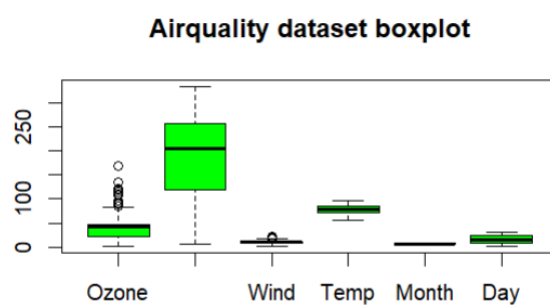
3. Boxplot method

CODE:

boxplot(df, main = "Airquality dataset boxplot", col = "green")

OUTPUT:

4. Mahalanobis distance

CODE:

mahalanobis_dist = mahalanobis(df, colMeans(df), cov(df))

threshold = qchisq(0.99, df = ncol(df)) # setting a threshold based on chi squared

outlier_mahalanobis = which(mahalanobis_dist > threshold)

outlier_mahalanobis

print("Taniya Ahmed 21BDS0059")

OUTPUT:

```
> outlier_mahalanobis
[1]  62 117
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
>
```

5. **Median Absolute Deviation method**

CODE:

med_abs_dev_value = mad(df$Ozone)

median_value = median(airquality$Ozone)

lower_bound_mad = median_value - 3 * med_abs_dev_value

upper_bound_mad = median_value + 3 * med_abs_dev_value

outlier_mad = df$Ozone[df$Ozone < lower_bound_mad | df$Ozone > upper_bound_mad]

outlier_mad

print("Taniya Ahmed 21BDS0059")

OUTPUT:

```
> outlier_mad
[1] 135 168
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
>
```

6. DBSCAN (Density – Based Spatial Clustering)

CODE:

```
db = dbscan(df[, c("Ozone", "Temp")], eps = 10, minPts = 5)

outlier_dbscan = airquality[db$cluster == 0, ]

outlier_dbscan

print("Taniya Ahmed 21BDS0059")
```

OUTPUT:

```
> outlier_dbscan
    Ozone Solar.R Wind Temp Month Day
30    115     223  5.7   79     5  30
62    135     269  4.1   84     7   1
86    108     223  8.0   85     7  25
99    122     255  4.0   89     8   7
101   110     207  8.0   90     8   9
117   168     238  3.4   81     8  25
121   118     225  2.3   94     8  29
> print("Taniya Ahmed 21BDS0059")
[1] "Taniya Ahmed 21BDS0059"
> |
```