Final Report (Output)

**-------------IQR OUTLIER DETECTION-------------**

**DATA BEFORE FILTERING**

OUTLIERS :

Indexes of rows with Outliers = [483, 711, 983, 1752, 1864, 2089, 2482, 2498, 2540, 3283, 4006, 4171, 4172, 4173, 4174, 4202, 4593, 4801, 4998, 5120, 5279, 5395, 5751, 7125, 7126, 7127, 7341, 7421, 7452, 7454, 7455, 7458, 7470, 7471, 7527, 7967, 7968, 7969, 7970, 7971, 7972, 8042, 8064, 8084, 8186, 8253, 9055, 9176, 9353, 9376, 9647, 9726, 9919]

Values of Outliers = [0, 6, 0, 0, 8, 0, 0, 0, 0, 0, 0, 6, 6, 6, 6, 0, 0, 0, 0, 0, 0, 0, 0, 8, 8, 8, 0, 0, 0, 0, 0, 0, 0, 0, 0, 8, 8, 8, 8, 8, 8, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

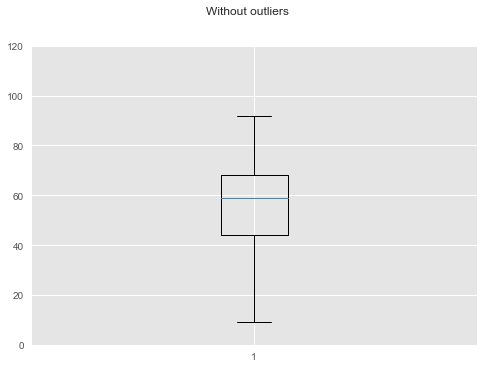
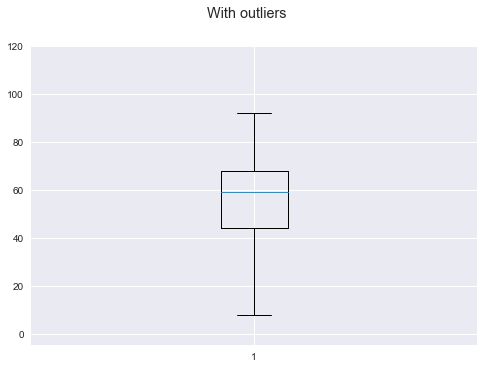
**DATA AFTER FILTERING**

OUTLIERS :

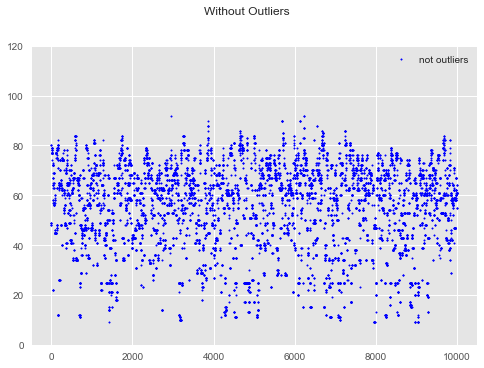
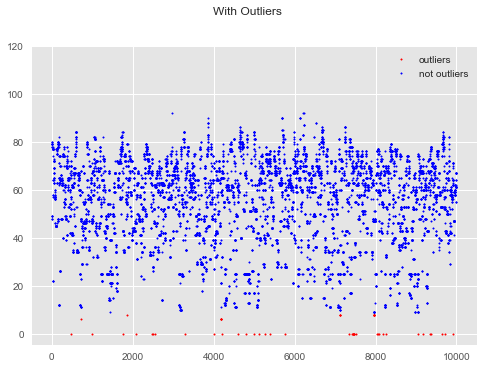
Indexes of rows with Outliers = [] (NO OUTLIER)

Values of Outliers = []

**-------------------BOX PLOTS-------------------**



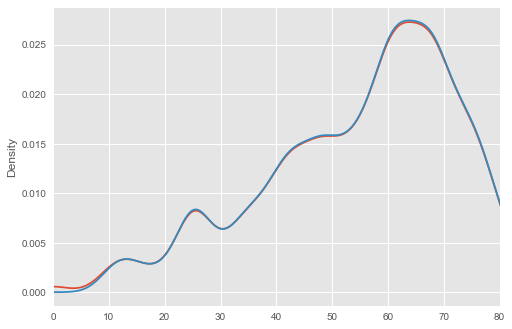
**-------------------SCATTER PLOTS-------------------**



**----------------KDE plot of data with and without outliers----------------**

RED - Unclean Data

BLUE - Cleaned Data



**------------ZSCORE OUTLIER DETECTION-----------**

**DATA BEFORE FILTERING**

OUTLIERS :

Indexes of rows with Outliers = [483, 711, 983, 1426, 1752, 1864, 2089, 2482, 2498, 2540, 3283, 4006, 4171, 4172, 4173, 4174, 4202, 4593, 4801, 4998, 5120, 5279, 5395, 5751, 7125, 7126, 7127, 7341, 7421, 7452, 7454, 7455, 7458, 7470, 7471, 7527, 7960, 7961, 7962, 7963, 7964, 7965, 7966, 7967, 7968, 7969, 7970, 7971, 7972, 8042, 8064, 8084, 8186, 8253, 8968, 8969, 8970, 9042, 9043, 9044, 9045, 9046, 9055, 9176, 9353, 9376, 9647, 9726, 9919]

Values of Outliers = [0, 6, 0, 9, 0, 8, 0, 0, 0, 0, 0, 0, 6, 6, 6, 6, 0, 0, 0, 0, 0, 0, 0, 0, 8, 8, 8, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9, 9, 9, 9, 9, 9, 9, 8, 8, 8, 8, 8, 8, 0, 0, 0, 0, 0, 9, 9, 9, 9, 9, 9, 9, 9, 0, 0, 0, 0, 0, 0, 0]

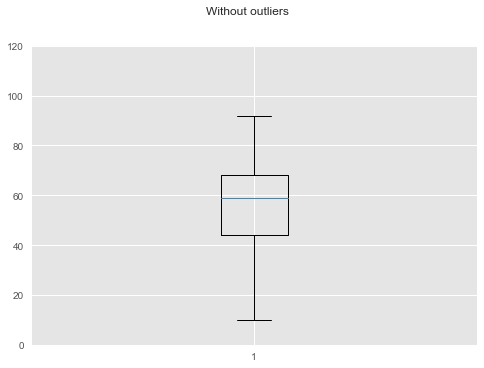
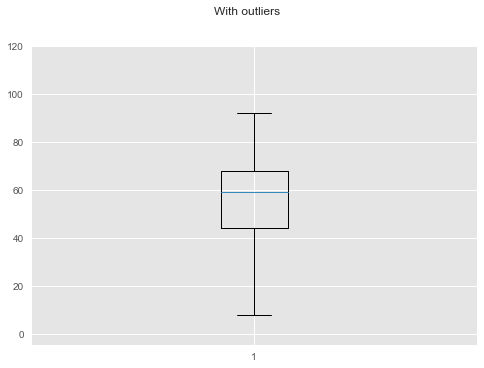
**DATA AFTER FILTERING**

OUTLIERS :

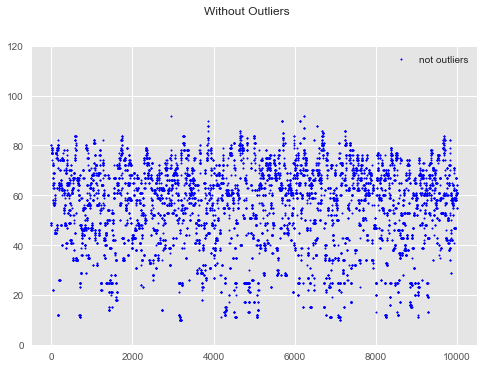
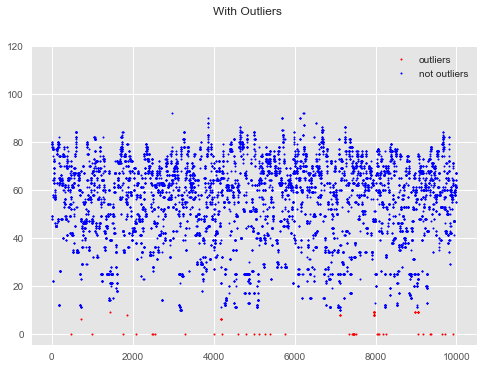
Indexes of rows with Outliers = [706, 707, 708, 709, 3150, 3151, 3152, 3153, 3154, 3155, 3156, 3167, 3168, 3169, 3170, 3171, 3172, 3173, 3174, 3175, 3176, 3177, 3178, 4163, 4514, 4515, 4521, 4522, 4523, 4524, 4525, 4526, 4527, 5046, 5047, 5048, 5049, 5050, 5051, 5052, 6848, 6849, 6850, 6851, 6852, 6853, 6854, 6855, 6856, 6857, 6858, 6859, 6860, 7052, 7053, 7054, 7055, 7056, 7057, 7058, 7059, 7060, 7061, 7067, 7086, 7087, 7088, 7098, 7099, 7100, 8183, 8184, 8185, 8186, 8187, 8977, 8978]

Values of Outliers = [11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11]

**-------------------BOX PLOTS-------------------**



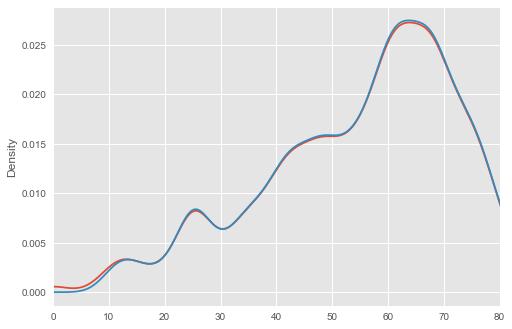
**-------------------SCATTER PLOTS-------------------**



**----------------KDE plot of data with and without outliers----------------**

RED - Unclean Data

BLUE - Cleaned Data



----------------------------------------------------------------------

**-------MODIFIED ZSCORE OUTLIER DETECTION-------**

**DATA BEFORE FILTERING**

OUTLIERS :

Indexes of rows with Outliers = [483, 707, 708, 709, 710, 711, 983, 1426, 1752, 1864, 2089, 2482, 2498, 2540, 3160, 3161, 3162, 3163, 3164, 3165, 3166, 3177, 3178, 3179, 3180, 3181, 3182, 3183, 3184, 3185, 3186, 3187, 3188, 3283, 4006, 4171, 4172, 4173, 4174, 4179, 4202, 4531, 4532, 4538, 4539, 4540, 4541, 4542, 4543, 4544, 4593, 4801, 4998, 5066, 5067, 5068, 5069, 5070, 5071, 5072, 5120, 5279, 5395, 5751, 6872, 6873, 6874, 6875, 6876, 6877, 6878, 6879, 6880, 6881, 6882, 6883, 6884, 7076, 7077, 7078, 7079, 7080, 7081, 7082, 7083, 7084, 7085, 7091, 7110, 7111, 7112, 7122, 7123, 7124, 7125, 7126, 7127, 7341, 7421, 7452, 7454, 7455, 7458, 7470, 7471, 7527, 7960, 7961, 7962, 7963, 7964, 7965, 7966, 7967, 7968, 7969, 7970, 7971, 7972, 8042, 8064, 8084, 8186, 8236, 8237, 8238, 8239, 8240, 8253, 8968, 8969, 8970, 9034, 9035, 9042, 9043, 9044, 9045, 9046, 9055, 9176, 9353, 9376, 9647, 9726, 9919]

Values of Outliers = [0, 11, 11, 11, 11, 6, 0, 9, 0, 8, 0, 0, 0, 0, 11, 11, 11, 11, 11, 11, 11, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 0, 0, 6, 6, 6, 6, 11, 0, 11, 11, 11, 11, 11, 11, 11, 11, 11, 0, 0, 0, 11, 11, 11, 11, 11, 11, 11, 0, 0, 0, 0, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 10, 10, 10, 8, 8, 8, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9, 9, 9, 9, 9, 9, 9, 8, 8, 8, 8, 8, 8, 0, 0, 0, 0, 11, 11, 11, 11, 11, 0, 9, 9, 9, 11, 11, 9, 9, 9, 9, 9, 0, 0, 0, 0, 0, 0, 0]

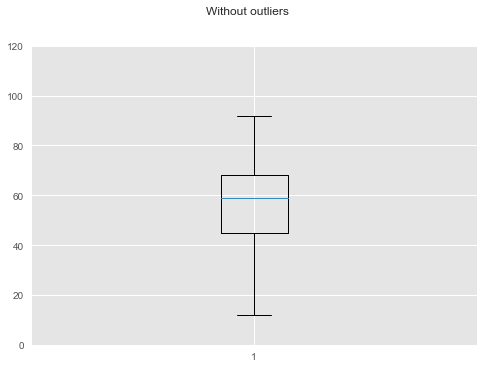
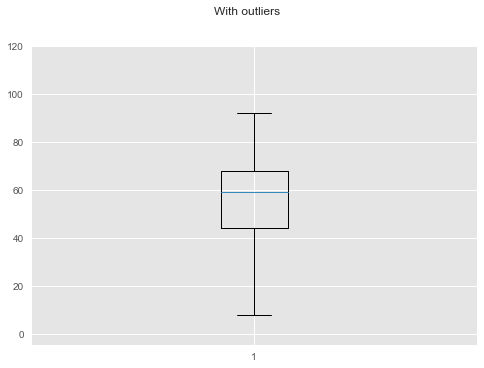
**DATA AFTER FILTERING**

OUTLIERS :

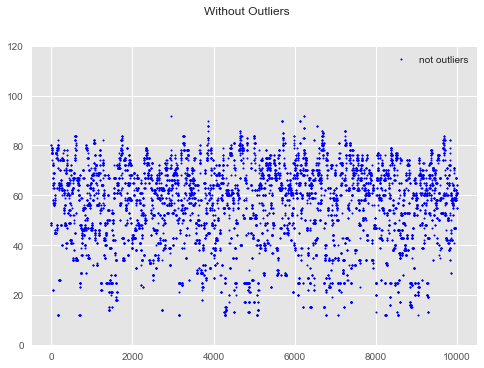
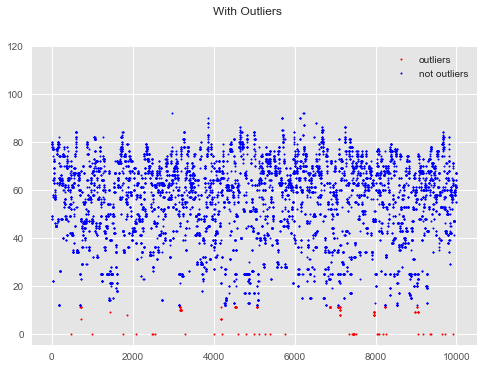
Indexes of rows with Outliers = [] (NO OUTLIER)

Values of Outliers = []

**-------------------BOX PLOTS-------------------**



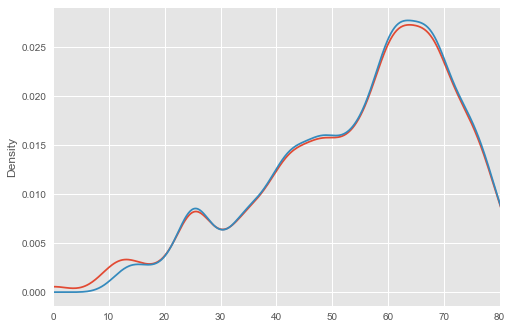
**-------------------SCATTER PLOTS-------------------**



**----------------KDE plot of data with and without outliers----------------**

RED - Unclean Data

BLUE - Cleaned Data



------------------------------------------------------------------------