Report for Q4

# OVERVIEW

Pandas was used to clean the data and fill in the empty values. Numpy was used to calculate the euclidean distance and matplotlib was used for graphing.

# OBJECTIVES

1. Estimate the values of PM10 and PM25 for the first day of 2017 using KNN
2. See how the results vary with normalisation.

# RESULTS (First 50)

1. For PM10, the original values were:

[ 16.0,

19.944863214296003,

21.0,

19.944863214296003,

19.944863214296003,

19.944863214296003,

28.0,

24.0,

19.944863214296003,

19.944863214296003,

19.944863214296003,

28.0,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

33.0,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

44.0,

55.0,

19.944863214296003,

19.944863214296003,

24.0,

19.944863214296003,

37.0,

19.944863214296003,

47.0,

19.944863214296003,

19.944863214296003,

42.0,

19.944863214296003,

19.944863214296003,

51.0,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

19.944863214296003,

73.0,

38.0,

19.944863214296003,

40.0,

35.0 ...]

1. For PM10, the estimated values(recursive) were:

[ 16.2,

19.944863214296003,

18.166917928577604,

19.944863214296003,

23.977945285718402,

19.944863214296003,

24.6,

23.8,

27.977945285718402,

18.988972642859203,

20.155890571436803,

27.0,

19.944863214296003,

24.566917928577602,

23.566917928577602,

31.566917928577606,

26.2,

19.566917928577602,

21.5558905714368,

19.944863214296003,

19.944863214296003,

32.7889726428592,

19.944863214296003,

19.944863214296003,

38.0,

27.566917928577602,

19.944863214296003,

19.944863214296003,

21.0,

19.944863214296003,

27.1779452857184,

19.944863214296003,

45.4,

19.944863214296003,

19.944863214296003,

39.7889726428592,

26.7669179285776,

23.7558905714368,

37.2,

22.755890571436804,

21.555890571436805,

24.755890571436804,

22.5558905714368,

21.355890571436802,

22.5558905714368,

35.177945285718394,

23.555890571436805,

19.944863214296003,

43.4,

24.577945285718403 …]

For PM10 , the values with rolling are:

16.0889726428592

19.944863214296

17.8614042500072

19.944863214296

23.372431607148

19.944863214296

23.7

20.8

27.072431607148

21.583458964288802

18.950376892866398

25.2944863214296

22.7614042500072

27.8669179285776

23.272431607148

27.9669179285776

22.588972642859197

20.8669179285776

20.7503768928664

19.944863214296

19.944863214296

35.0944863214296

20.750376892866402

22.750376892866402

34.0

23.755890571436797

19.944863214296

20.050376892866403

19.6

19.944863214296

24.7779452857184

22.8614042500072

34.394486321429596

20.566917928577602

28.077945285718403

35.394486321429596

25.8669179285776

23.1558905714368

31.588972642859197

22.6614042500072

20.750376892866402

25.1558905714368

27.8669179285776

23.1614042500072

29.172431607148003

29.677945285718398

26.1614042500072

19.944863214296

32.394486321429596

24.0669179285776

1. For PM10, the values after normalisation were:

11.0

19.944863214296003

14.188972642859202

19.944863214296003

29.7889726428592

19.944863214296003

28.2

24.0

34.377945285718404

14.188972642859202

19.944863214296003

27.4

19.944863214296003

22.555890571436805

31.9669179285776

30.555890571436798

27.4

19.944863214296003

22.355890571436802

19.944863214296003

19.7558905714368

27.7889726428592

19.944863214296003

19.944863214296003

38.3889726428592

27.566917928577602

19.944863214296003

19.944863214296003

19.0

19.944863214296003

27.577945285718403

19.944863214296003

37.177945285718394

21.5558905714368

25.355890571436802

42.8

25.355890571436806

18.3889726428592

37.3889726428592

19.944863214296003

23.766917928577605

24.755890571436804

30.555890571436798

19.944863214296003

25.166917928577604

30.555890571436798

31.388972642859198

23.755890571436804

42.8

25.366917928577603

1. For PM25, the original values were:

[ 9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

25.0,

21.0,

44.0,

9.947407877285046,

9.947407877285046,

23.0,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

29.0,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

30.0,

9.947407877285046,

9.947407877285046,

37.0,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

43.0,

9.947407877285046,

89.0,

38.0,

9.947407877285046,

9.947407877285046,

45.0,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

9.947407877285046,

33.0,

9.947407877285046,

34.0,

9.947407877285046,

1. For PM25, the expected values(recursive) were :

[ 11.494765776675708,

10.449160948042218,

11.087753493647046,

9.947407877285054,

14.640370605309196,

11.289506564404212,

19.4,

17.833333333333332,

15.32809911000904,

12.896518847432871,

10.852667089556547,

20.65264210060935,

12.291259635161376,

15.32809911000904,

14.235111393037702,

16.3351113930377,

14.936864463794867,

17.329827191819003,

11.289506564404212,

9.947407877285054,

11.419333756223214,

13.826346039251877,

11.354420160313712,

10.715827614708884,

23.633333333333333,

26.89649385848567,

9.947407877285054,

11.552667089556547,

17.759654383638008,

12.686000422889881,

22.557901312880844,

12.291259635161376,

24.2,

11.92459296849471,

15.963185514099537,

22.47721008015686,

12.991259635161377,

16.30703727197586,

22.314049555004523,

14.603531130461532,

11.087753493647048,

10.715827614708884,

15.194765776675705,

13.826346039251876,

17.401778059704366,

26.615802625761685,

17.026346039251873,

12.686000422889881,

23.8,

16.801778059704368,

Expected values using rolling:

13.073703938642524

9.947407877285046

11.557926301828036

9.947407877285046

18.189481575457012

12.763185514099531

20.8

18.4

16.76318551409953

12.457926301828035

11.557926301828036

24.6

14.168444726371026

14.96318551409953

14.86844472637103

16.973703938642526

20.694740787728506

18.1

12.763185514099531

9.947407877285046

9.947407877285046

15.363185514099536

9.947407877285046

9.947407877285046

29.4

35.2

9.947407877285046

9.947407877285046

20.4

9.947407877285046

26.6

11.457926301828035

28.3

13.368444726371028

14.963185514099532

32.5

11.457926301828035

16.473703938642522

32.2

17.184222363185516

12.763185514099531

9.947407877285046

14.96318551409953

10.552667089556541

13.56318551409953

33.7

14.75792630182804

9.947407877285046

26.1

20.27896315091402

1. For PM25, the expected values after normalisation were:

20.031495440476533

19.944863214295996

19.057728464293596

20.948539000009596

21.683458964288796

19.3485390000096

25.8

23.866666666666667

32.5944863214296

18.951963523812264

19.552214785723198

25.731495440476532

21.020719345246665

22.48371022619973

20.62990880953066

24.689223904770127

28.933333333333334

20.88922390477013

22.191061797626926

19.944863214295996

19.139098273814668

29.5981621071432

20.517043559533064

20.248539000009597

32.394486321429596

20.68738601191333

19.944863214295996

20.248539000009597

20.157477202382665

21.350376892866397

22.161152988096266

19.8522147857232

32.961152988096266

20.52439513096026

33.65380141666906

31.9926484285728

21.787386011913327

19.451963523812264

32.522305976192534

21.08738601191333

22.72439513096026

22.08554811905653

25.459566357150393

21.250376892866395

24.92255723810346

25.59473758334053

29.261152988096267

24.559566357150395

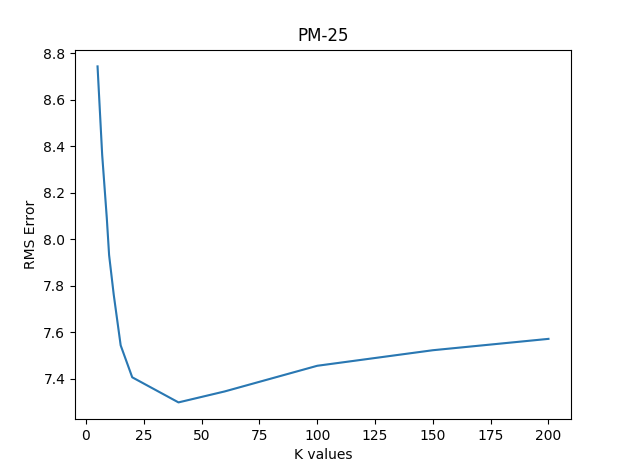
28.82230597619253

21.391061797626932

# Choosing Best k Value

We iterated over a set of possible k values and calculated the RMSE error for various k.

The graph obtained from this process looked like:



This graph shows that an ideal k value would be between 25 to 50, approximately 35. However, we can see that the RMSE does not decrease very vastly between values after 20 and to achieve a better performance, k values such as 15,20 can also be used.