Analitica Hackathon

Title: Market segmentation based on local customer activity

Team Members:

- Satish Palaniappan (tpsatish95@gmail.com)
- Siddharth G (gxrockinstyle@gmail.com)
- Vijayalakshimi MLS (vijayalakshimisethuraman@gmail.com)

BE CSE, SSN College of Engineering,2015

Human Activity Recognition

- What is it?

  Using smart phones (or smart devices) to monitor physical activity of human beings.

- Why is it necessary?

  - Health care researchers would give a fortune to understand physical activity of groups of people
  - To understand the sequence of common activities one performs throughout the day

The idea

- We use Human Activity Recognition and pattern mining to present a mirage of implementations where this would make a tremendous impact. The sectors we focus on:
  - Market Segmentation for targeted and focused product sales
  - Remote patient activity monitoring, for psychiatrists, doctors, healthcare professionals
  - Predicting the profession of an individual
  - Fitness app which gives feedback based on user activity patterns
  - A fun use case we present is, suggesting songs based on the mood of the person, which is reflected by the physical activity sequence they perform.

Our System

- To predict a sequence of humanly activities (from a basic set of 6) that any individual with a gyroscope, accelerometer & GPS (or to be more specific, having a smartphone) on them, is performing at any instant of time and there-by mining the relationships and associations between activity patterns of people.
- As our first step, we classify signals from the smartphone to indicate that the subject is in one of the following six activities:
  - SITTING
  - STANDING
  - WALKING
  - WALKING UP
  - WALKING DOWN
  - LAYING

Our System (cont.)

- With a classifier trained for physical activity recognition, our system collects similar but new data from GPS devices over time to form activity sequences (eg. STANDING – STANDING – SITTING – WALKING – WALKING – LAYING – LAYING – LAYING – LAYING – SITTING) and separate it on a user basis.

Data source

- University of California Irvine has a huge repository of datasets
- We used the dataset – Human Activity Recognition using Smartphones to train our classifier to identify physical human activity
- http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

DATASET

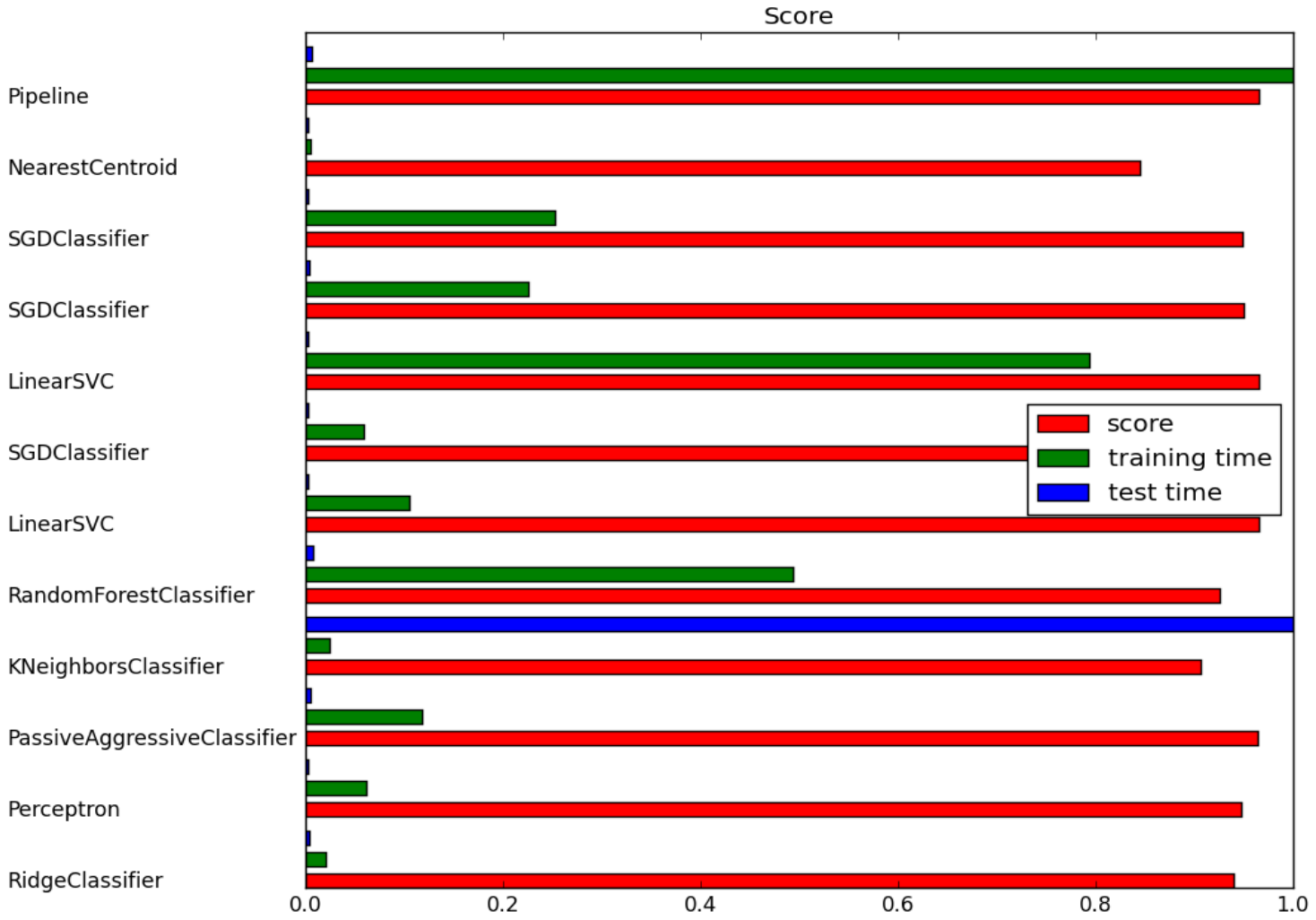
- The initial dataset was arrived at by analysing the gyroscope and accelerometer data from 30 people to label 6 activities.
- Features such as time domain acceleration, magnitude and a mirage of other features were extracted and Fast Fourier transformed to the frequency domain also.
- After some filters such as low pass filters were used to remove noise, the cleaned up dataset looks like this:
  - A matrix of filtered and normalized inertial signals (float numbers), of size 7352 * 561 where each row represents a training example of 561 features, so in essence, we have 7352 training examples
  - A class label matrix of size 7352 * 1, where each of the elements are from 1 – 6 representing one of the six classes

Dataset looks like this. This is a partial set of features of the first training example

CLASSIFIERS

- We have a report of the accuracy of various classifiers we trained over this dataset in the file named ‘report_classifiers.txt’.
- We have chosen a linear SVM (with L2 penalty) since it gave us a 97% accuracy at a fast training time.
- Comparison report is available on the next slide

Classifier comparison



The report for linear-SVC

==========

L2 penalty

Training:

LinearSVC(C=1.0, class_weight=None, dual=False, fit_intercept=True,intercept_scaling=1, loss='l2', max_iter=1000, multi_class='ovr',penalty='l2', random_state=None, tol=0.001, verbose=0)

train time: 3.031s

test time: 0.047s

accuracy: 0.965

dimensionality: 561

density: 1.000000

classification report:

	precision	recall	f1-score	support
1	0.96	1.00	0.98	496
2	0.98	0.96	0.97	471
3	1.00	0.98	0.99	420
4	0.97	0.87	0.92	491
5	0.90	0.98	0.94	532
6	1.00	1.00	1.00	537
avg / total	0.97	0.97	0.96	2947

confusion matrix:

```
[[495   0   1   0   0   0]
 [ 20 450   1   0   0   0]
 [  2   5 413   0   0   0]
 [  0   3   0 429  57   2]
 [  1   0   0  11 520   0]
 [  0   0   0   0   0 537]]
```

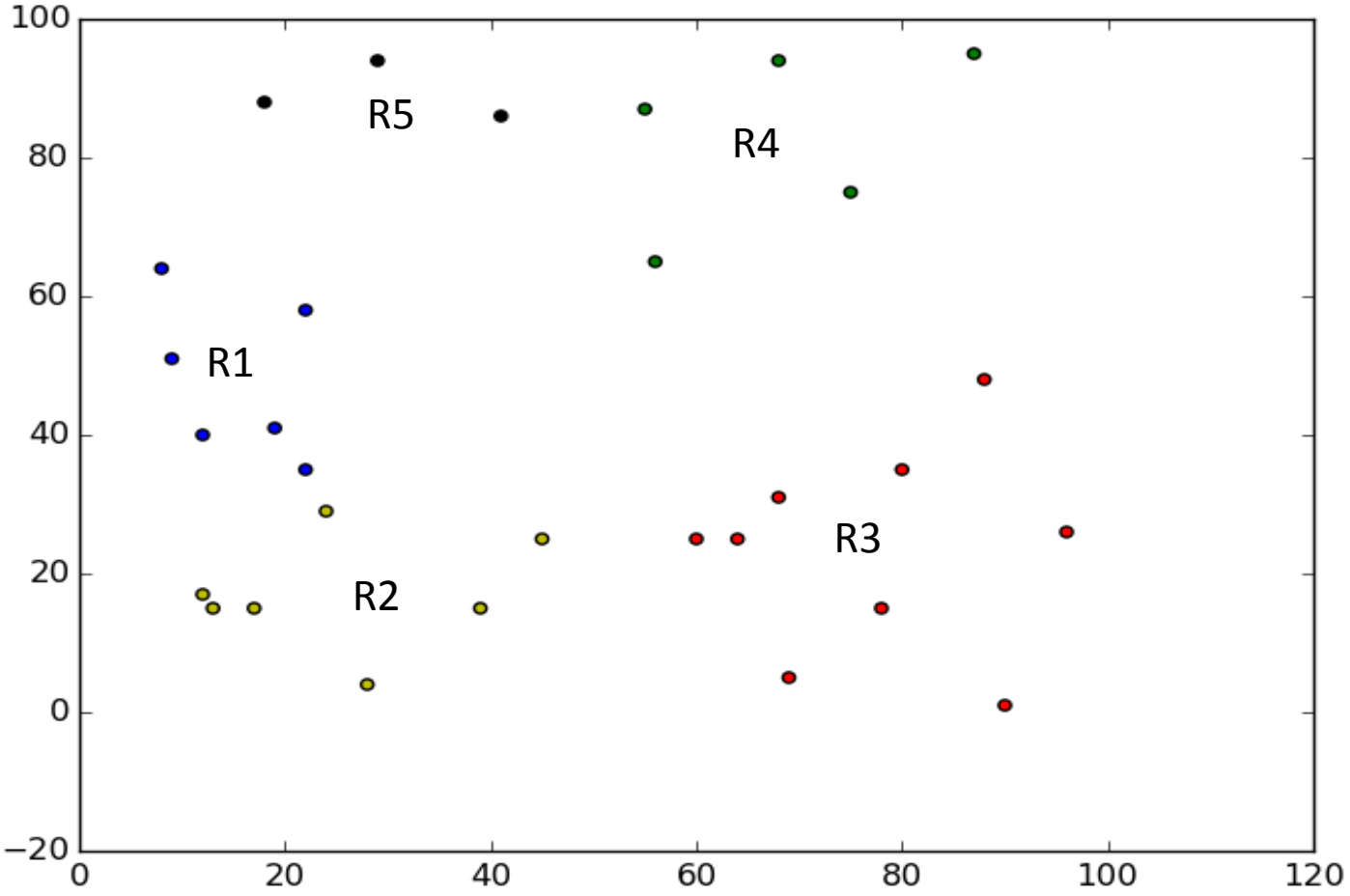
MARKET SEGMENTATION

- Then we input GPS locations along with training data of inertial signals for separate users where the signals are time ordered
  - Input of the form of the following tuple for ever user:

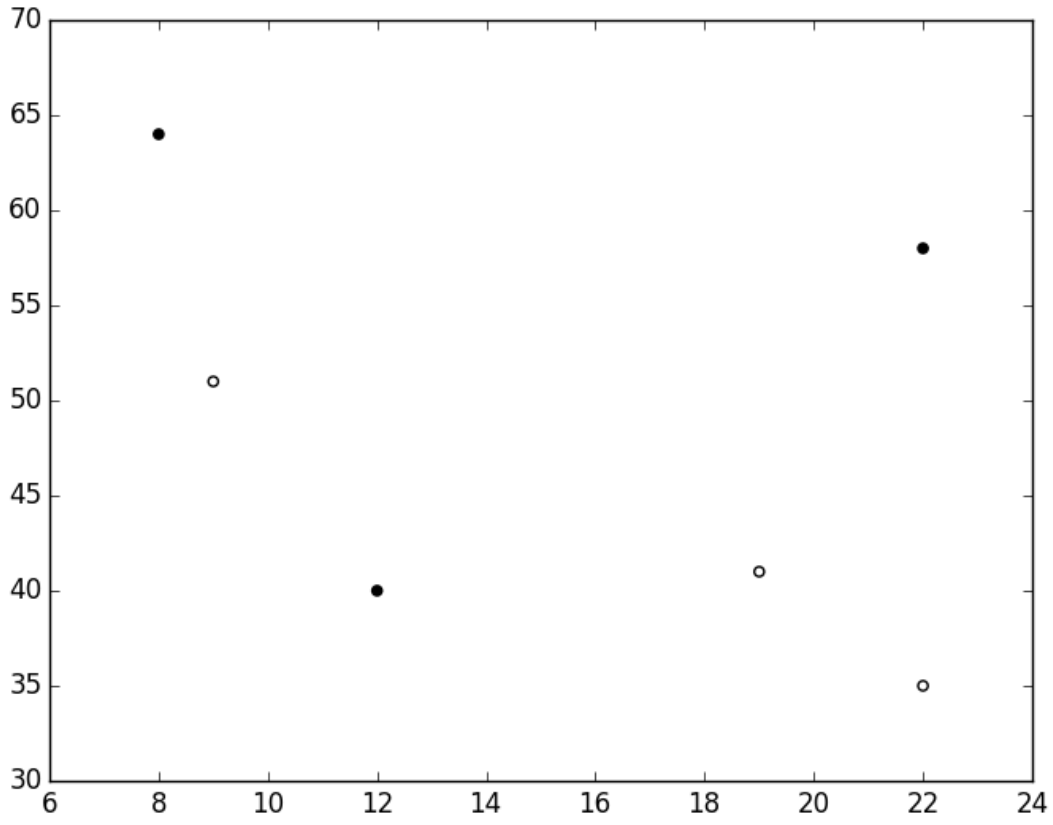
    (GPS coordinates, (Training data in time sequence))

- Then we cluster the GPS coordinated based on geographical nearness.
- Each cluster is then analysed to produce activity report of populace

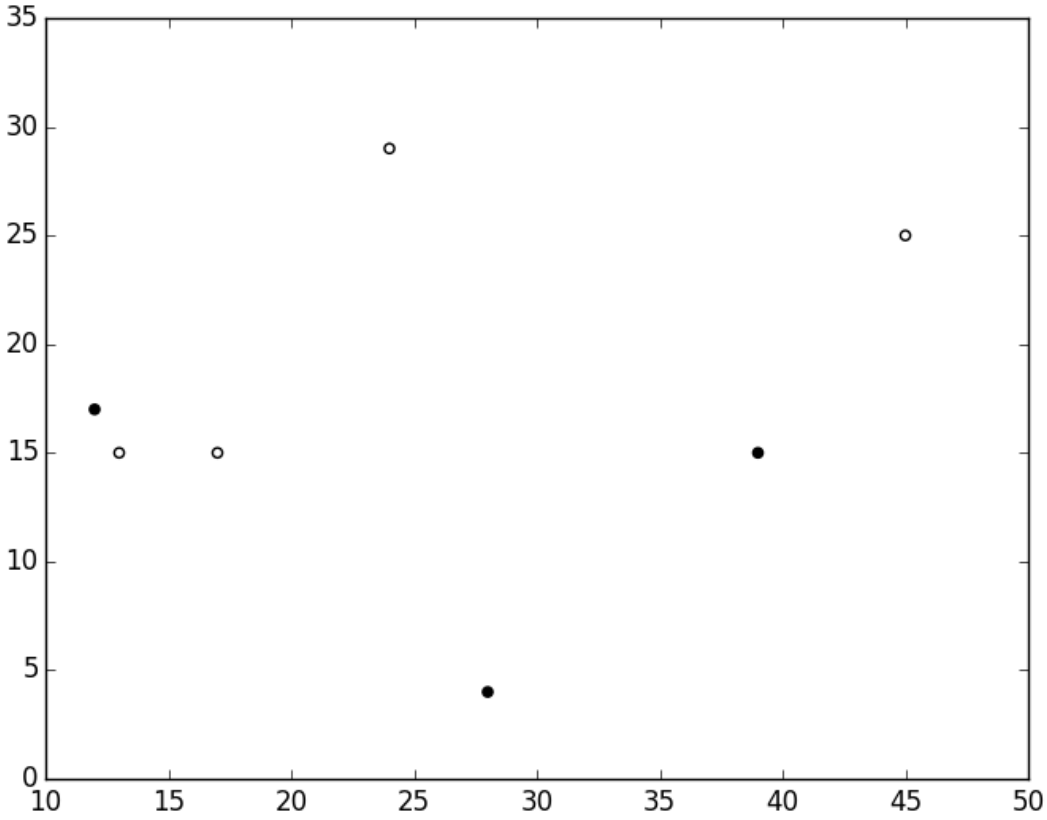
Geographic clusters



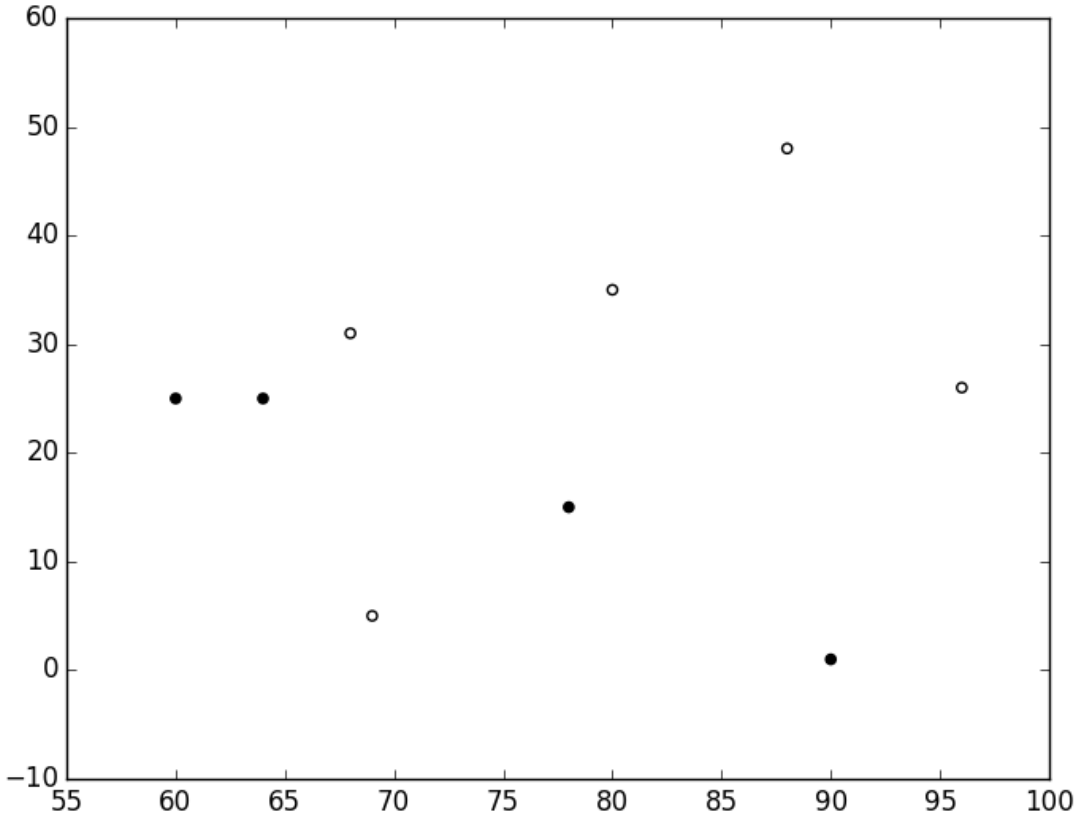
Activity analysis – the black and white labels here denote two different types of user activity patterns. The following image is of Region 1



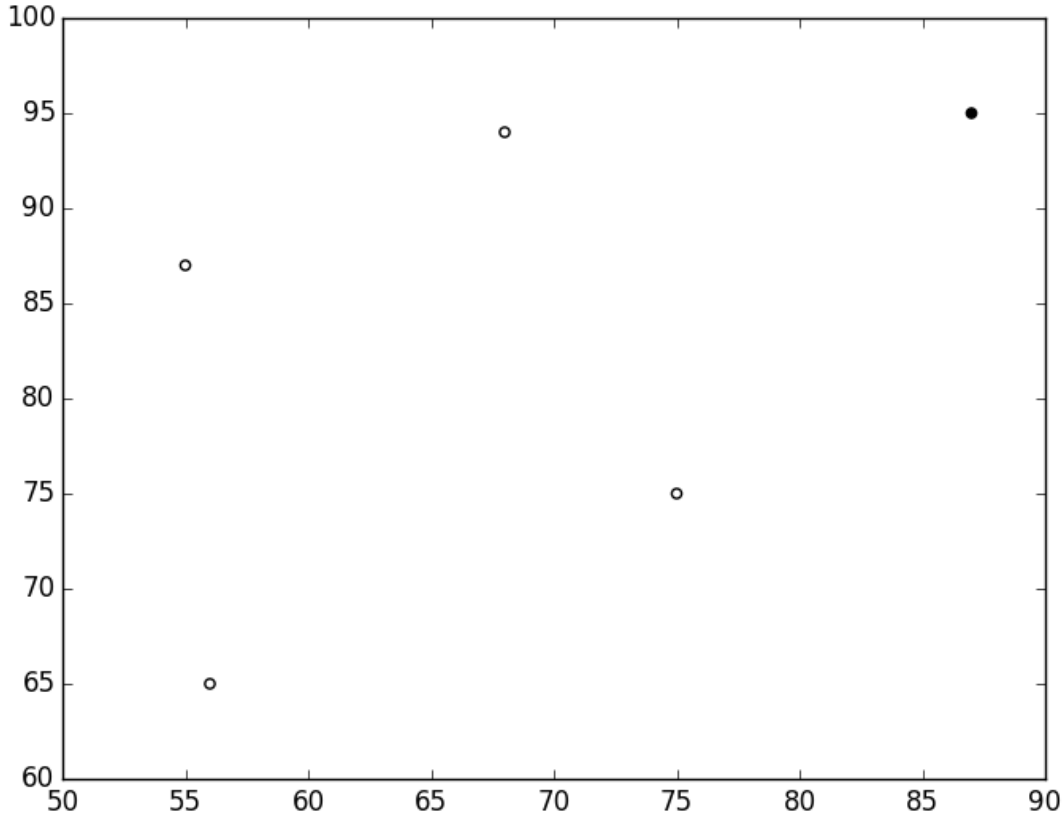
Region 2



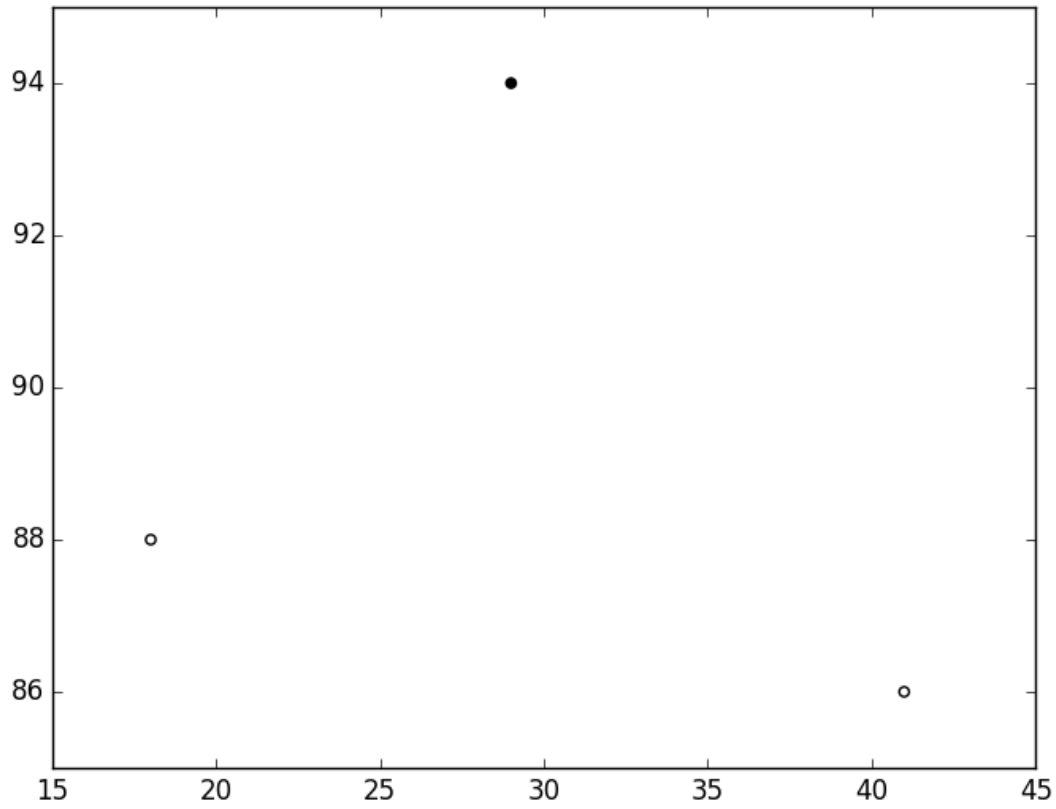
Region 3



Region 4



Region 5



OUTPUT

- We have made frequency analysis of the activities performed by individuals in a particular geographical cluster.
- This gives scope to target products to audience based on their activity patterns.
  - For eg, a location with high physical activity can be targeted to sell sports drinks
  - A location with low physical activity can be aimed at to sell fitness products, etc.
- The frequency analysis report for clusters in the images can be viewed at the file ‘geo_frequency_output.txt’

What difference does this make?

- We can go a huge way to destroying obesity which is a plague infesting nations today.
- We can predict people's professions from their activity patterns and sell them products they need to enhance the overall health of the community
- A great but quirky use case would be monitoring mentally challenged patients in an asylum. There will be a single place to monitor activity of all patients and anomaly detection can be used to seek out erratic activity and render help to them really quickly.

Thank you for watching! ☺