K-Means - Aarhus city Road traffic Data.

The k-means clustering method is utilized for the clustering of segments in Aarhus city traffic dataset. Dataset contain two types of files. Several combinations of traffic features are investigated in order to achieve the best results.

a. Traffic Data: Columns in Road data

i. |-- status: string (nullable = true)
ii. |-- avgMeasuredTime: string (nullable = true)
iii. |-- avgSpeed: string (nullable = true)
iv. |-- extID: string (nullable = true)
v. |-- medianMeasuredTime: string (nullable = true)
vi. |-- TIMESTAMP: string (nullable = true)
vii. |-- vehicleCount: string (nullable = true)
viii. |-- _id: string (nullable = true)
ix. |-- REPORT_ID: string (nullable = true)

b. Meta Data : Column in MetaData

i. |-- status: string (nullable = true)
ii. |-- avgMeasuredTime: string (nullable = true)
iii. |-- avgSpeed: string (nullable = true)
iv. |-- extID: string (nullable = true)
v. |-- medianMeasuredTime: string (nullable = true)
vi. |-- TIMESTAMP: string (nullable = true)
vii. |-- vehicleCount: string (nullable = true)
viii. |-- _id: string (nullable = true)
ix. |-- REPORT_ID: string (nullable = true)
x. |-- POINT_1_STREET: string (nullable = true)
xi. |-- DURATION_IN_SEC: string (nullable = true)
xii. |-- POINT_1_NAME: string (nullable = true)
xiii. |-- POINT_1_CITY: string (nullable = true)
xiv. |-- POINT_2_NAME: string (nullable = true)
xv. |-- POINT_2_LNG: string (nullable = true)
xvi. |-- POINT_2_STREET: string (nullable = true)
xvii. |-- NDT_IN_KMH: string (nullable = true)
xviii. |-- POINT_2_POSTAL_CODE: string (nullable = true)
xix. |-- POINT_2_COUNTRY: string (nullable = true)
xx. |-- POINT_1_STREET_NUMBER: string (nullable = true)
xxi. |-- ORGANISATION: string (nullable = true)
xxii. |-- POINT_1_LAT: string (nullable = true)
xxiii. |-- POINT_2_LAT: string (nullable = true)
xxiv. |-- POINT_1_POSTAL_CODE: string (nullable = true)
xxv. |-- POINT_2_STREET_NUMBER: string (nullable = true)

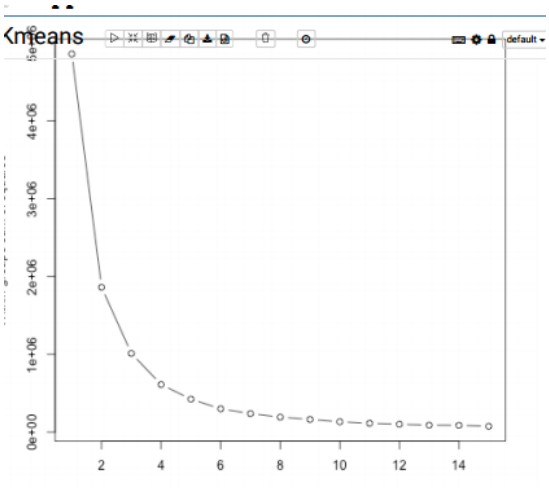
xxvi. |-- POINT_2_CITY: string (nullable = true)
xxvii. |-- ROAD_TYPE: string (nullable = true)
xxviii. |-- POINT_1_LNG: string (nullable = true)
xxix. |-- POINT_1_COUNTRY: string (nullable = true)
xxx. |-- DISTANCE_IN_METERS: string (nullable = true)
xxxi. |-- REPORT_NAME: string (nullable = true)
xxxii. |-- RBA_ID: string (nullable = true)

2. Merged Meta and Road datasets. There are 18213793 rows in the dataset. Each Row is Five minutes of traffic data with time variant and spatial data.
3. Data aggregated hourly. Now data set contain 10632 rows. Each row includes average speed, Latitude, Longitude and hour in 24-hour format.
4. Data grouped into five time zones. Each zone approximately to five hours of data.
5. Divided data into five time zones
   a. 22- 6 (10PM - 6 AM) - Zone 1 & 5
   b. (7 AM – 11 AM) - Zone 2
   c. (12 AM – 4 PM) - Zone 3
   d. (5 PM - 9 PM) – Zone 4
6. Filtered rows with mean speed less than 30 Km/hr(These is a chance less speed means chance of traffic congestions)
7. Traffic congestion Recognition is presented based on the clustering of traffic segments. We implemented K-means on formatted dataset. Before implementing model we try to find the Best K- Value

```
%r
mydata <- data1
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata,centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
```

8.

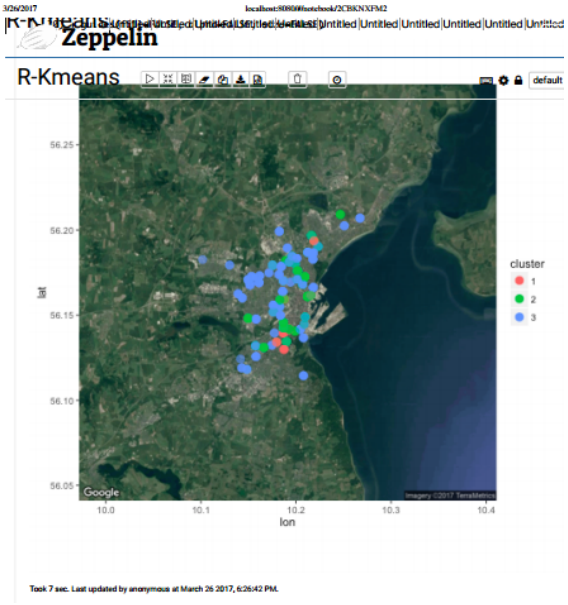
Plotted K values

9. 

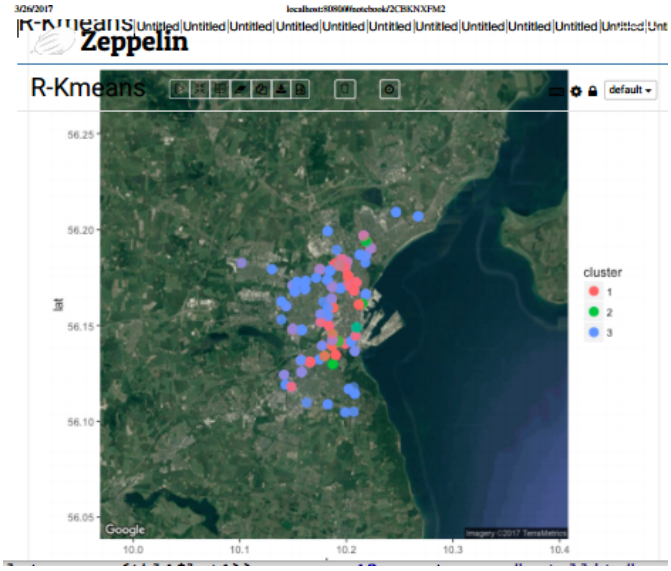
10. We have taken three as best K value. With this K value plotted maps for five time zones.

Time Zone -1

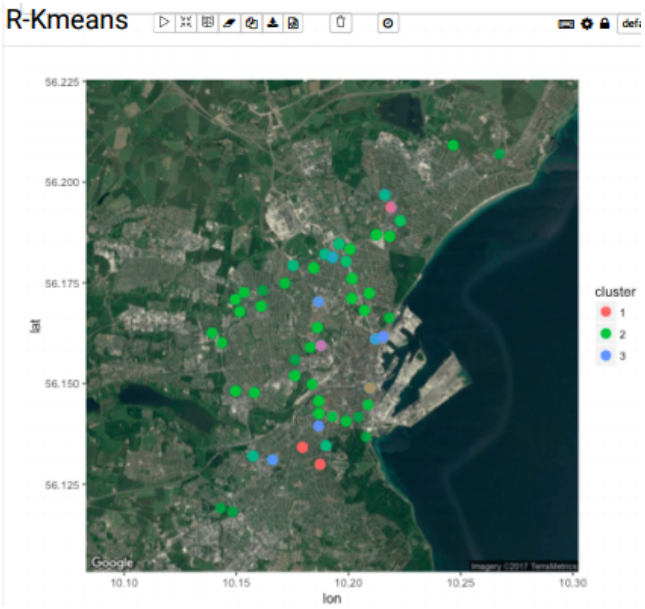
TimeZone -2



Time Zone -3



Time Zone -4



Time Zone - 5

- ok 4 sec. Last updated by anonymous at March 26 2017, 6:28:10 PM.

- By observing different time zones maps, we can say that Zone -1 and Zone - 5 (22 PM to 6 AM) traffic with average speed less than 30 km/hour, traffic scattered all around the city.
- But Zone -1, Zone-2, Zone -3 (7 AM - 21 PM) less speed (implies more traffic) areas are closer. All less traffic areas mostly in the nearer neighbourhood.
- We clearly identify road congestions are more in day time. And also which areas in city are prone to traffic congestions.
- By investigating more thoroughly, we may be able to identify even roads that are causing more traffic congestions.