MIDTERM
Part 3

Team 5

Team members:

Dipti Pamnani
Jayesh Samyani
Suraj Sharma

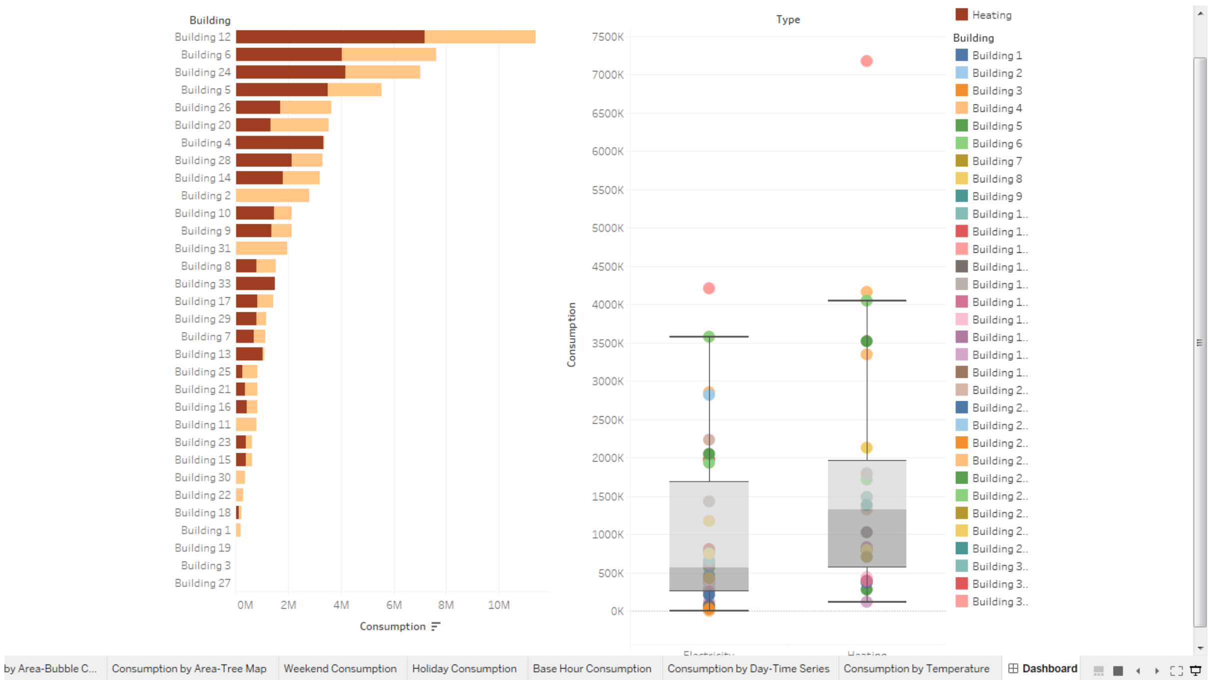
Under guidance of :

Prof. Sri Krishnamurthy
Ashwin Dinoriya

Part 3

Dashboard - 1

Heat and Electricity consumption for each building:



The above two graphs show the consumption for each building segregated by heat and electricity consumption.

The bar chart shows the segregation of heat and electricity consumption by colour.

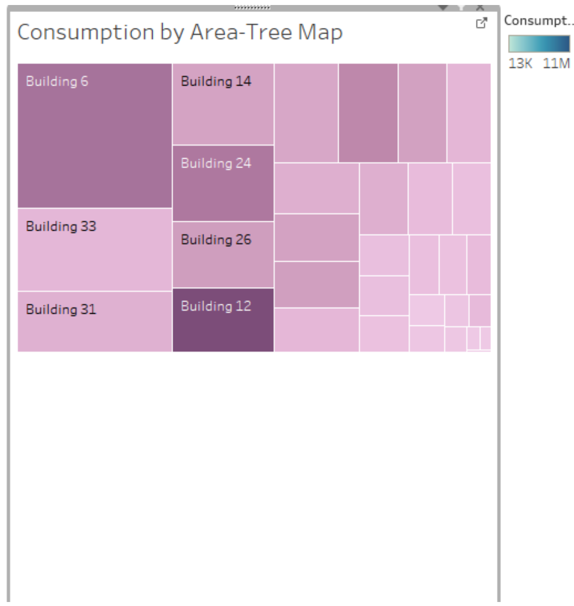
The box plot shows the distribution for electricity consumption for all buildings. The box pot gives a range of consumption over the buildings.

Building 12 is seen as a major outlier, hence it is inferred that the consumption for building 12 is huge.

Dashboard 2

Chart 1: Consumption by Area (Bubble Chart)

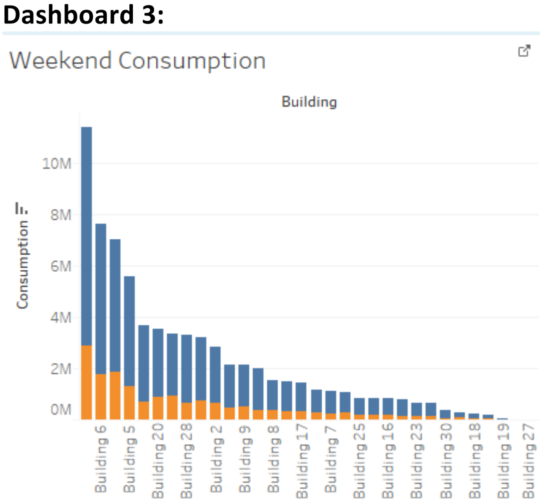
Chart 2: Consumption by Area (Tree Map)

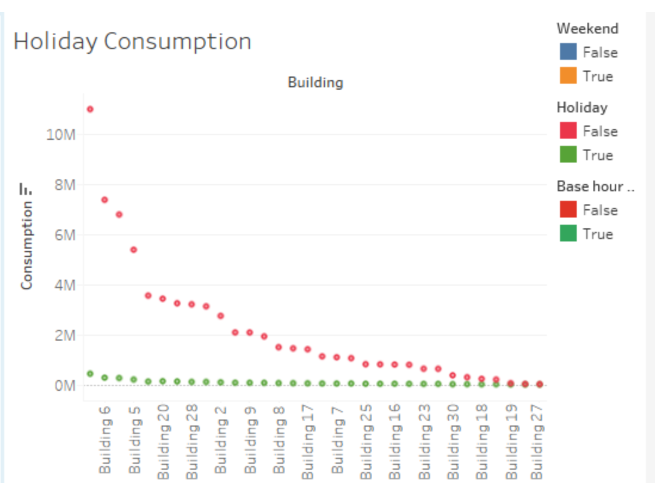


The above two charts show the consumption for each building.
From the bubble chart, we have considered parameters like area and consumption.
The size of the bubble infers the area of the building.
The colour of the bubbles infer the consumption for the respective building.
Thus, looking at the graph we conclude that Building 12 with the darkest shade of blue has the greatest consumption (though it is not the one with greatest area)

There is also another way of representing the same data, which is shown in chart 2 with the help of tree map.
The size of the Box implies the area of that building, and the colour specifies the consumption (Dark to light for High to Low)
Thus here we conclude that inspite of Building 6 having a large area does not consume heat/electricity as compared to that of Building 12.

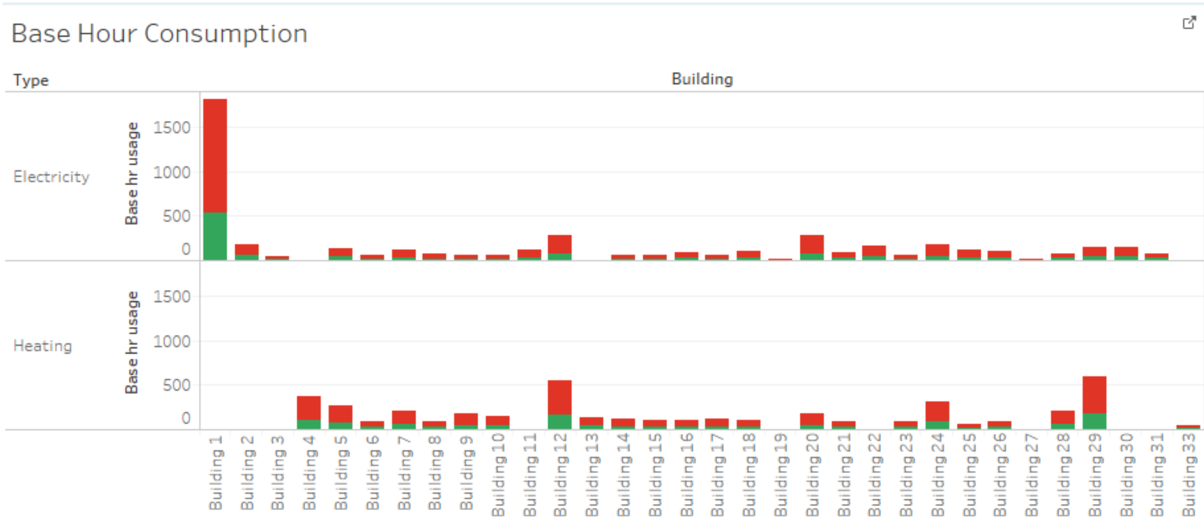
Dashboard 3:





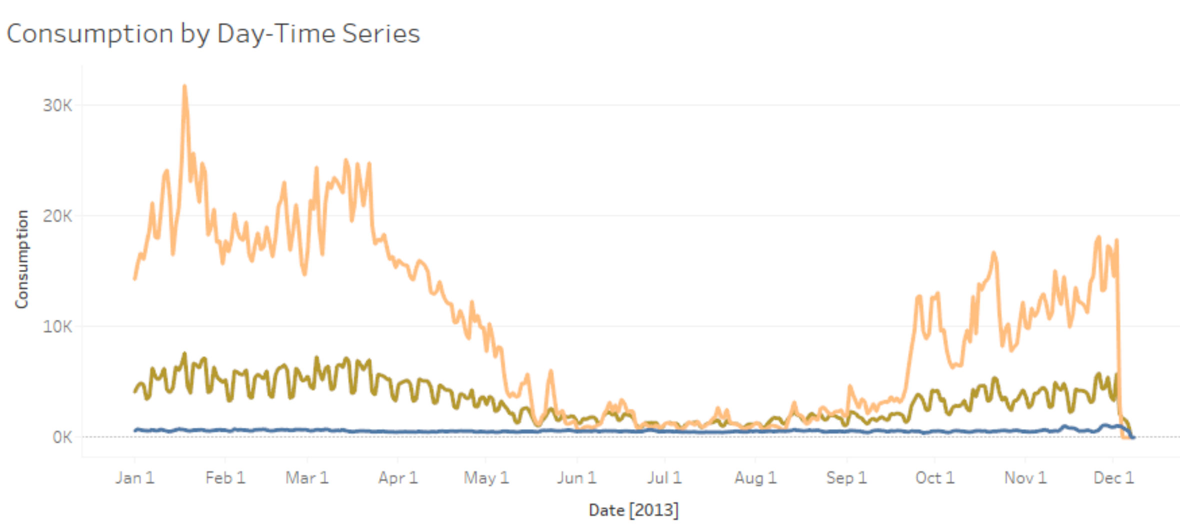
These charts give the weekend and holiday consumption for each building.

Conclusion: From the above charts it is clear that Building 6 consumes more heat/electricity on weekends/holidays that any other building.

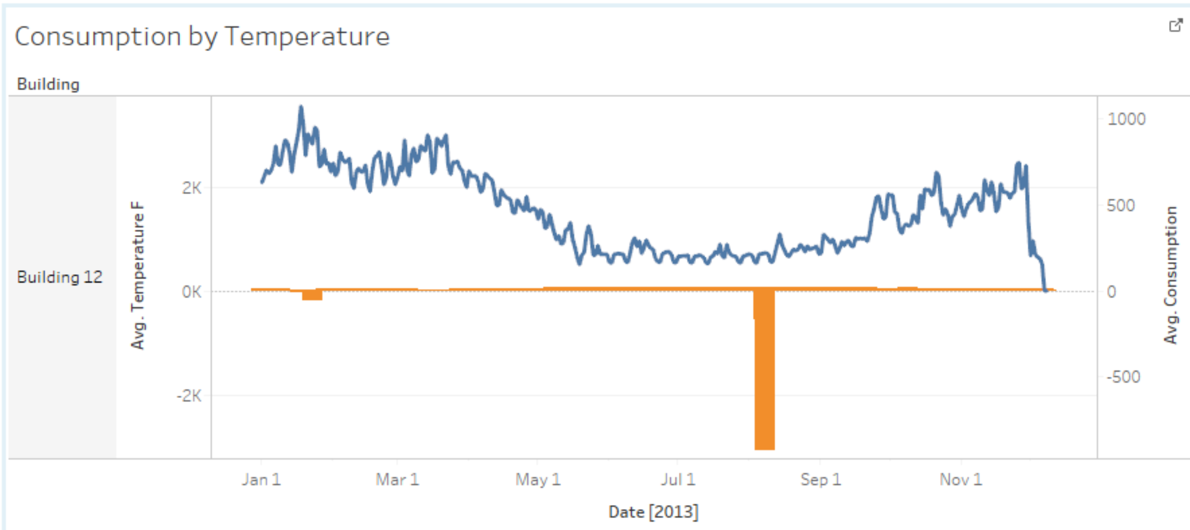


The above graph shows the base hour consumption for each building where ideally the consumption should be minimal. Building 1 seems to have more electricity consumption as compared to other building during base hour time frame.

Dashboard 4:



This graph shows the per day consumption of heat and electricity in a time series format. Building 4 shows a very fluctuating usage with very minimal usage during June-Sept.



This graph shows the dependency between temperature change and consumption for each building which implies that there is not much difference for this building with respect to temperature.

Dashboard 5

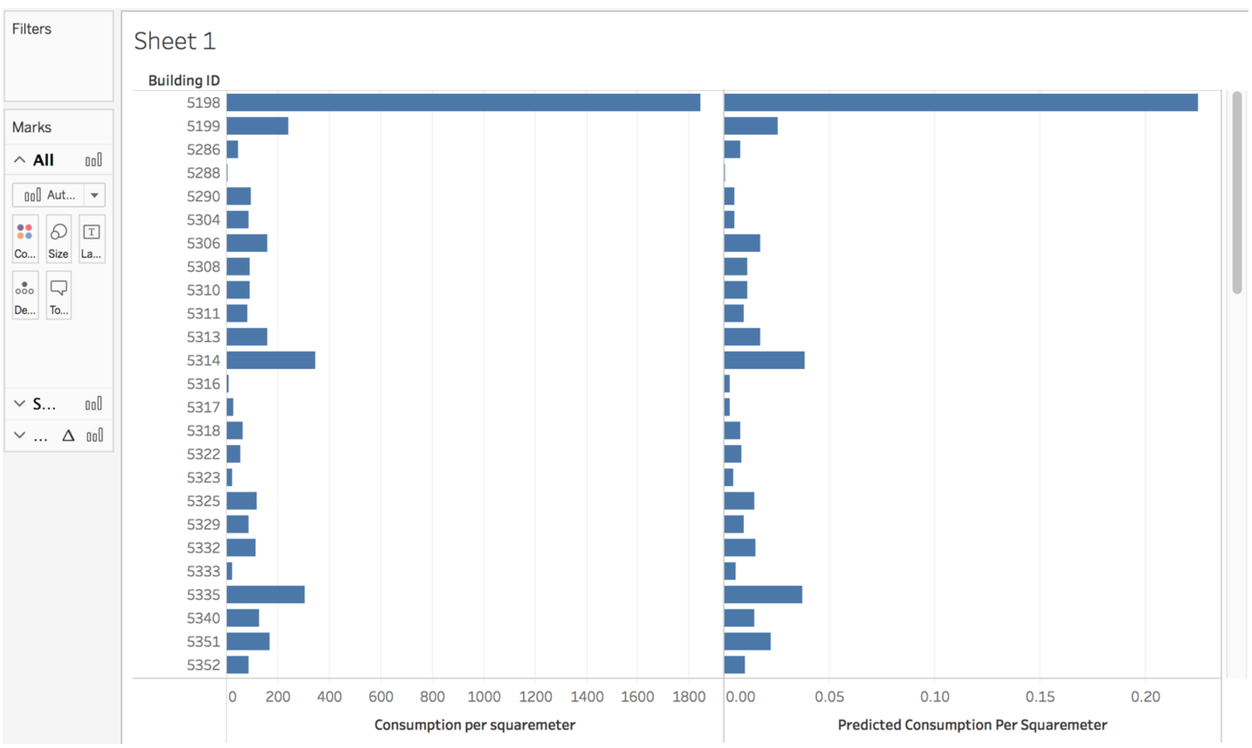
Consumption of heat and electricity for each moth for a particular building.
Building 12 shows a constant use of electricity for each month except December, whereas the heat consumption for heat varies.

After ingesting data into data source of tableau.
We need to connect tableau to R using Rserve service.

Steps to connect R to tableau using Rserve

Step 1. Install package Rserve in R (install.packages("Rserve"))
Stpe 2. Load Rserve (library(Rserve))
Step 3. Start Rserve (run.Rserve() for linux and Rserve() for windows)
Step 4. Open Tableau. Go to Help -> Go to Systems and Performance -> Select Manage external service connection .
Step 5. Type Localhost in Server and 6311 in port number.

After successful connection.



<u>Prediction using Linear regression</u>

Steps to write R script and run in tableau for best models of Prediction, Classification and clustering

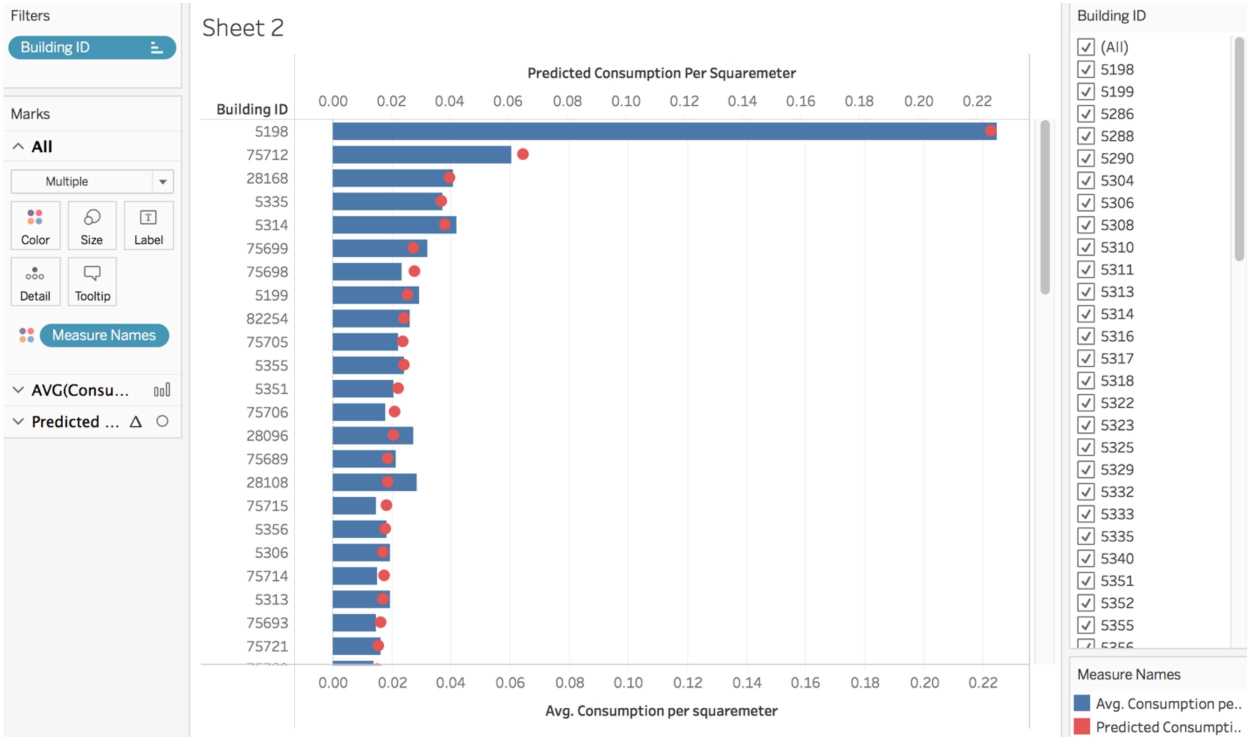
1. Create a new Calculated Field in measures
2. Write R script for linear regression for predicting Consumption per squaremeter

R SCRIPT:

SCRIPT_REAL('mydata <- data.frame(Area=.arg1,Base_hr_usage=.arg2, Consumption_per_squaremeter=.arg3,Consumption=.arg4, Dew_pointF=.arg5,Sea_Level_PressureIn=.arg6, Wind_SpeedMPH=.arg7, Wind_Dir_Degrees=.arg8);

lrmodel <- lm(Consumption_per_squaremeter~ Area + Base_hr_usage+Consumption_per_squaremeter + Consumption + Dew_pointF + Sea_Level_PressureIn + Wind_SpeedMPH + Wind_Dir_Degrees, data = mydata);
prob <- predict(lrmodel, newdata = mydata, type = "response")',
AVG([Area Floor. M.Sqr.X]),AVG([Base hr usage]),AVG([Consumption per squaremeter]),AVG([Consumption]),AVG([Dew PointF]),AVG([Sea Level PressureIn]),AVG([Wind SpeedMPH]),AVG([Wind Dir Degrees]))

3. Based on Consumption per square meter, seprate building ID with their actual and predicted Consumption per square meter

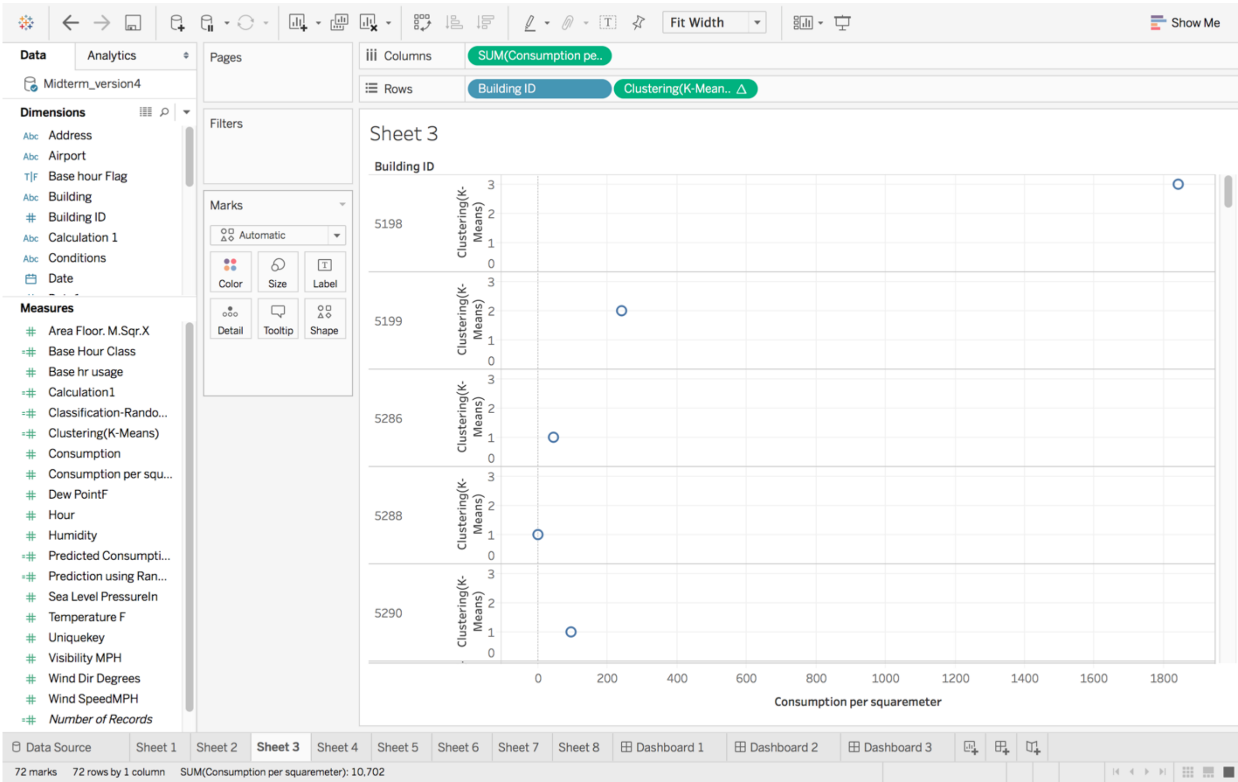


<u>Prediction using Linear regression</u>

Above sheet is replicated.

Additional Steps:

4. Right Click on prediction based on consumption per squaremeter, select dual axis.
5. Use bar graph to denote actual consumption and and solid dots to denote prediction Consumption



Clustering using K-means

1. Create a new Calculated Field in measures

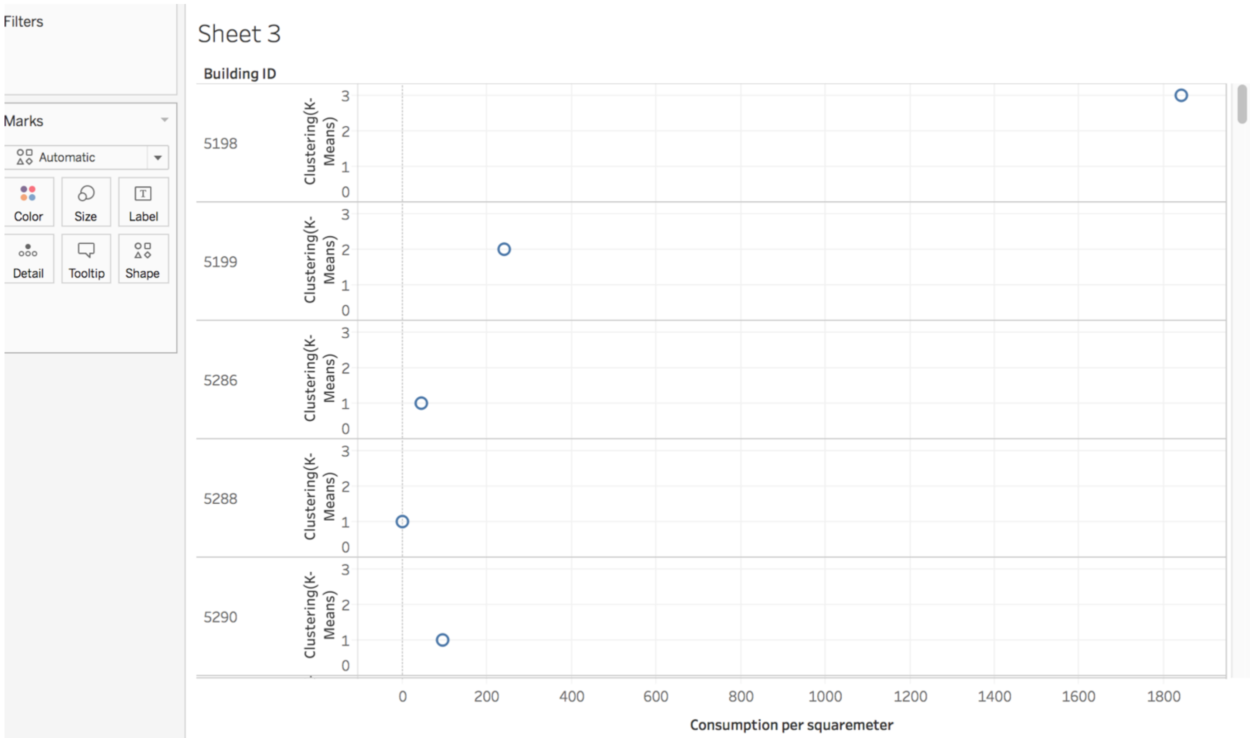
2. Write R script for linear regression for predicting Consumption per squaremeter

R Script:

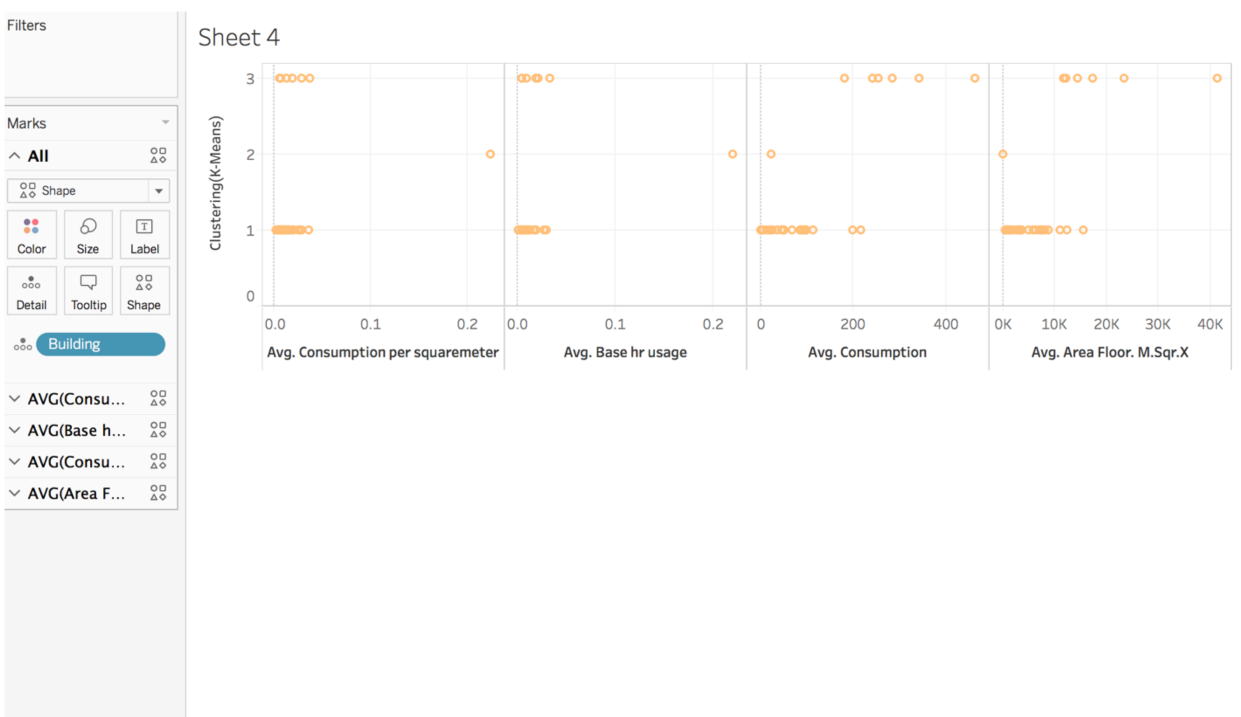
```
SCRIPT_REAL("consumption <- ( .arg1 - mean(.arg1) ) / sd(.arg1)
   normalized.consumption <- ( .arg2 - mean(.arg2) ) / sd(.arg2)
   consumption.base <- ( .arg3 - mean(.arg3) ) / sd(.arg3)
   area.sqrm <- ( .arg4 - mean(.arg4) ) / sd(.arg4)
   dat <- cbind(consumption, normalized.consumption, consumption.base, area.sqrm)
   num <- 3
   kmeans(dat, num)$cluster",

AVG( [Consumption] ), AVG( [Consumption per squaremeter] ), AVG( [Base hr usage] ),
AVG( [Area Floor. M.Sqr.X] ))
```

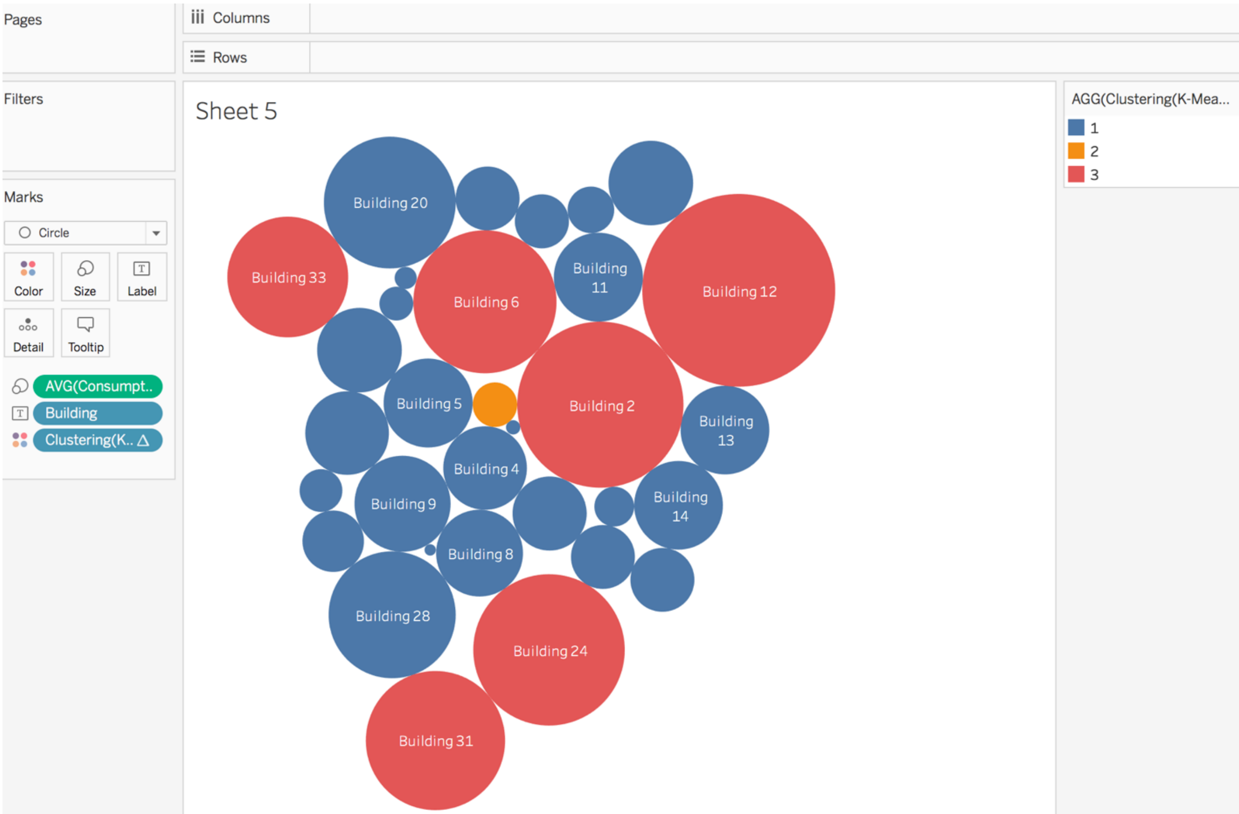
3. Based on Consumption per square meter, cluster building ID’s



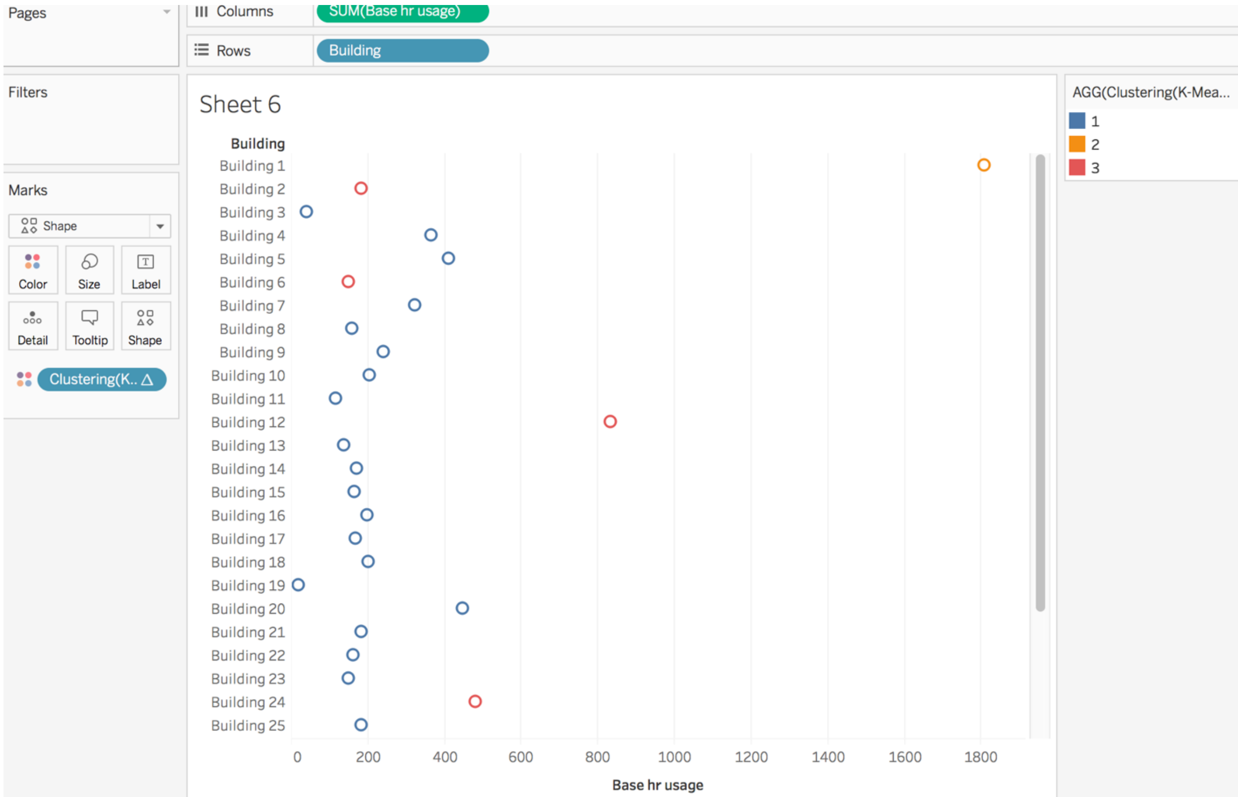
Different way of representing clustering



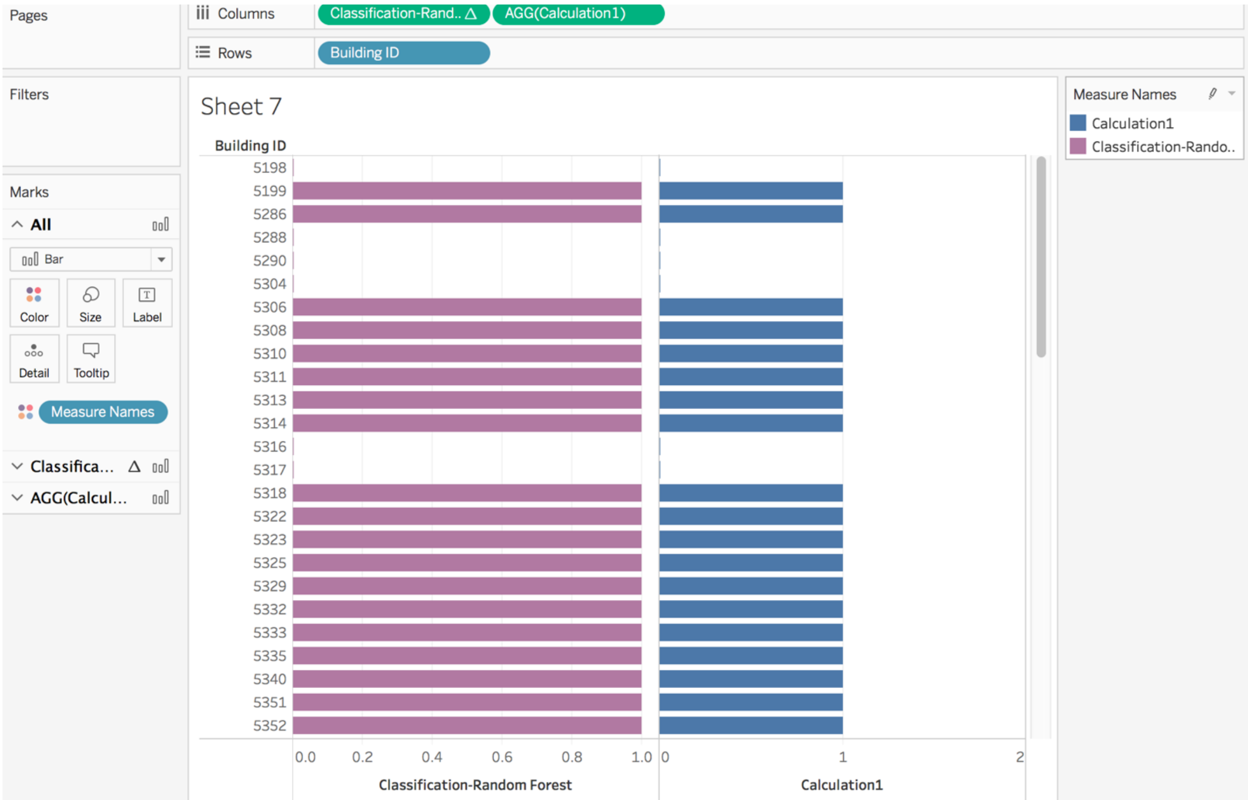
Different way of representing clustering



<u>Different way of representing clustering</u>



<u>Different way of representing clustering</u>



Classification using Random Forest

1. Create a new Calculated Field in measures

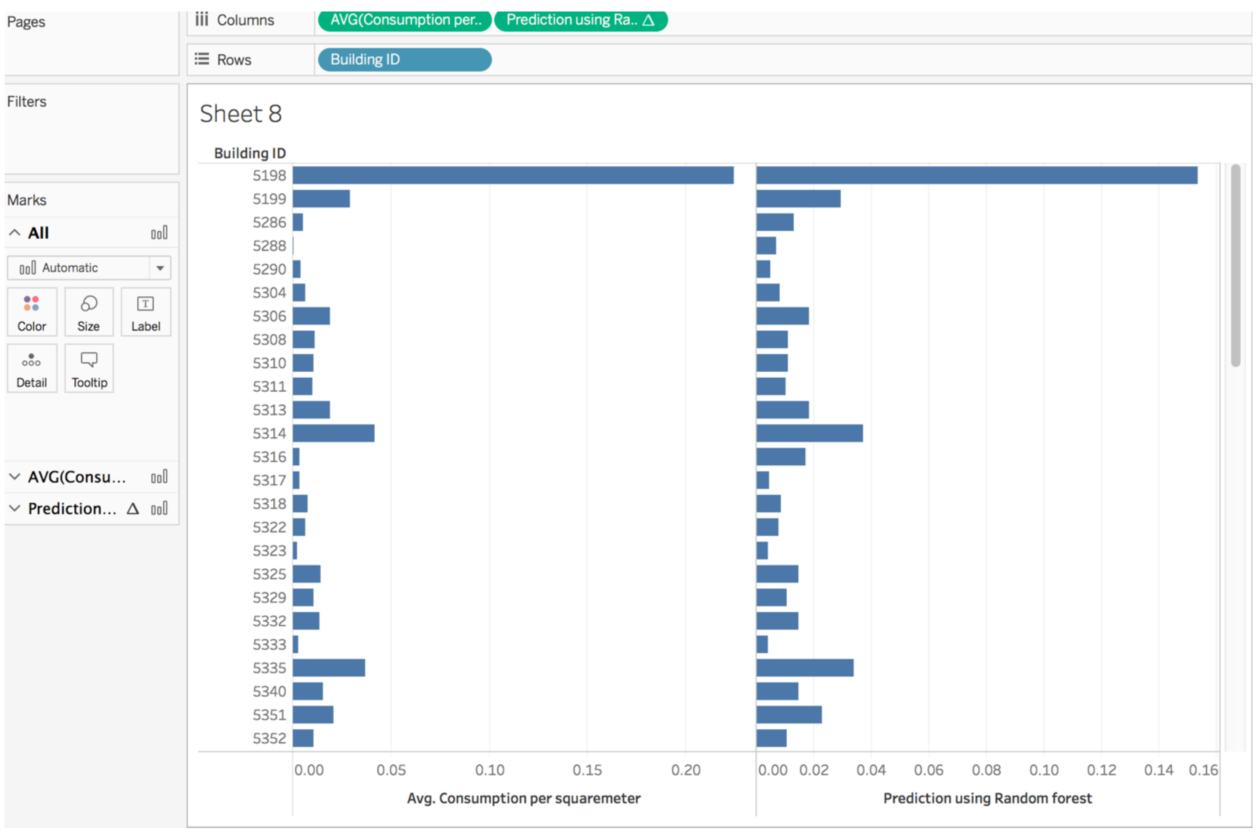
2. Write R script for random forest for classification Base hour class

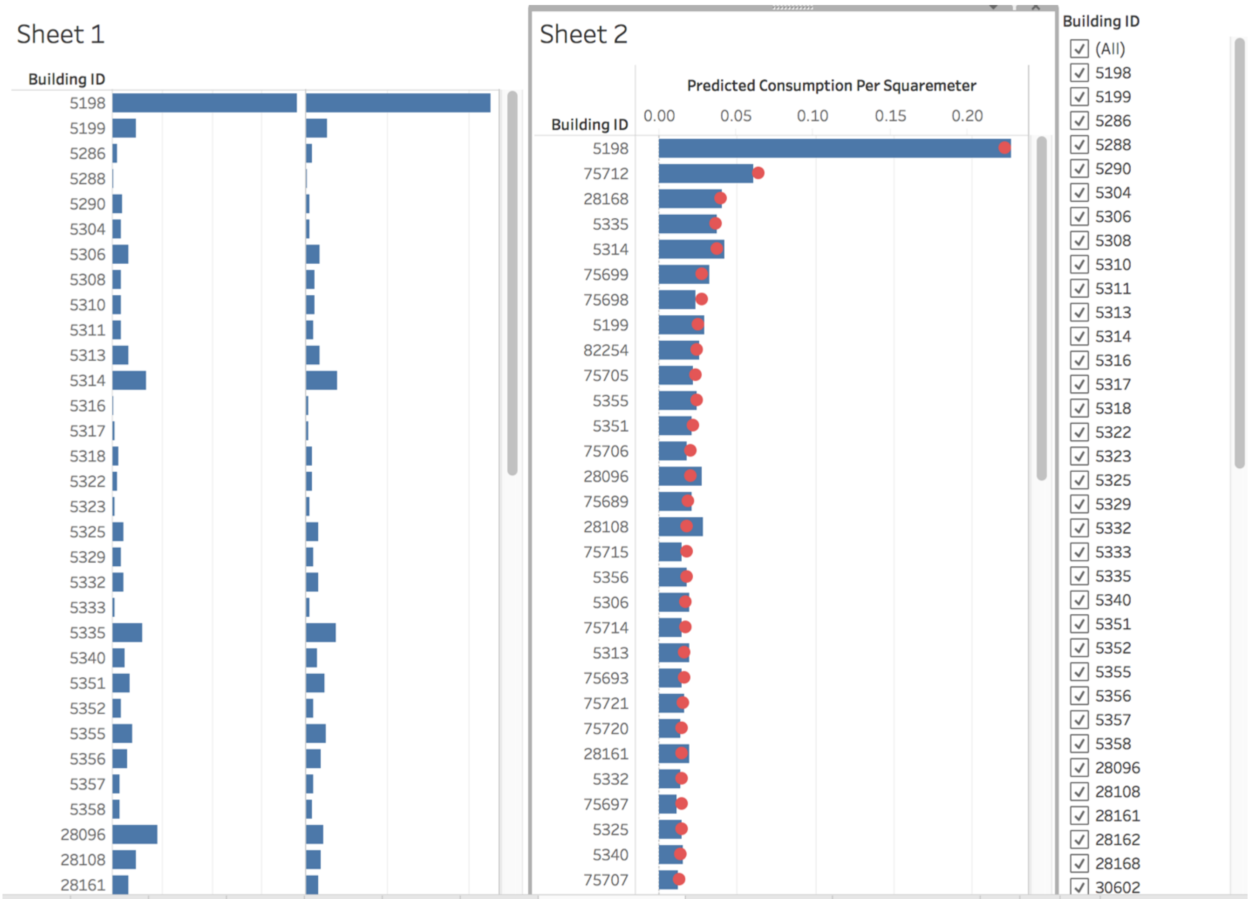
R SCRIPT

```
SCRIPT_REAL('install.packages("randomForest");
library(randomForest);
mydata <- data.frame(Area=.arg1,Base_hr_usage=.arg2,
Consumption_per_squaremeter=.arg3,Consumption=.arg4,
Dew_pointF=.arg5,Sea_Level_PressureIn=.arg6, Wind_SpeedMPH=.arg7,
Wind_Dir_Degrees=.arg8,Base_hour_class=.arg9);
fit<-randomForest( Base_hour_class ~ .,
data=mydata,mtry=floor(sqrt(ncol(mydata))),ntree=500);
prob <- predict(fit, newdata = mydata, type = "response");prob<-round(prob)',


AVG([Area Floor. M.Sqr.X]),AVG([Base hr usage]),AVG([Consumption per
squaremeter]),AVG([Consumption]),AVG([Dew PointF]),AVG([Sea Level PressureIn]),AVG([Wind
SpeedMPH]),AVG([Wind Dir Degrees]),AVG([Base Hour Class]),AVG([Base Hour Class]))
```

3. Based on Base hour class , classify the data using Random Forest





<u>**Prediction using Random Forest**</u>

1. Create a new Calculated Field in measures

2. Write R script for linear regression for predicting Consumption per squaremeter

R SCRIPT

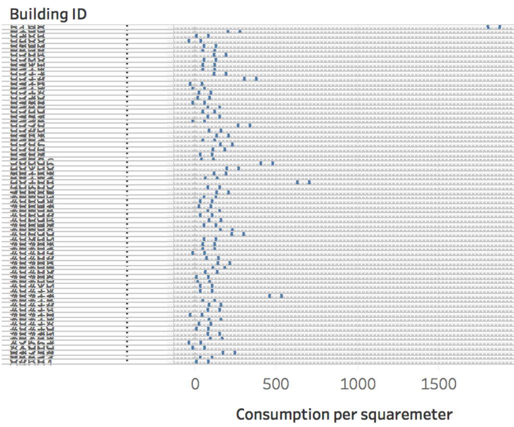
```
SCRIPT_REAL('install.packages("randomForest");
library(randomForest);
mydata <- data.frame(Area=.arg1,Base_hr_usage=.arg2,
Consumption_per_squaremeter=.arg3,Consumption=.arg4,
Dew_pointF=.arg5,Sea_Level_PressureIn=.arg6, Wind_SpeedMPH=.arg7,
Wind_Dir_Degrees=.arg8,Base_hour_class=.arg9);
fit<-randomForest( Consumption per squaremeter ~ .,
data=mydata,mtry=floor(sqrt(ncol(mydata))),ntree=500);
prob <- predict(fit, newdata = mydata, type = "response");prob<-round(prob)',


AVG([Area Floor. M.Sqr.X]),AVG([Base hr usage]),AVG([Consumption per
squaremeter]),AVG([Consumption]),AVG([Dew PointF]),AVG([Sea Level PressureIn]),AVG([Wind
SpeedMPH]),AVG([Wind Dir Degrees]),AVG([Base Hour Class]),AVG([Base Hour Class]))
```

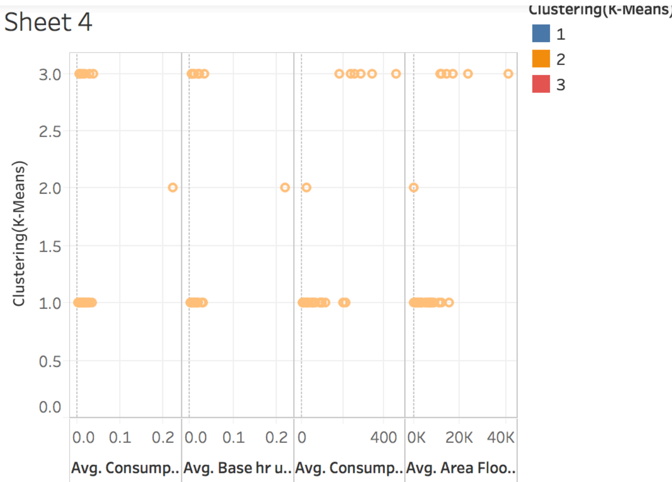
3. Based on Base hour class , classify the data using Random Forest

Team 5 Midterm

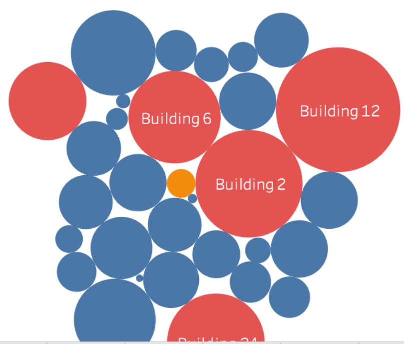
Sheet 3



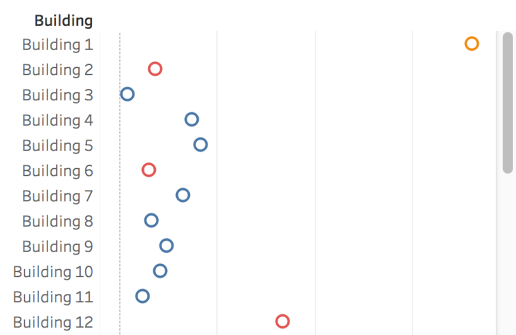
Sheet 4



Sheet 5



Sheet 6



Dipti Pamnani Jayesh Samyani Suraj Sharma

Team 5 Midterm

Dipti Pamnani Jayesh Samyani Suraj Sharma