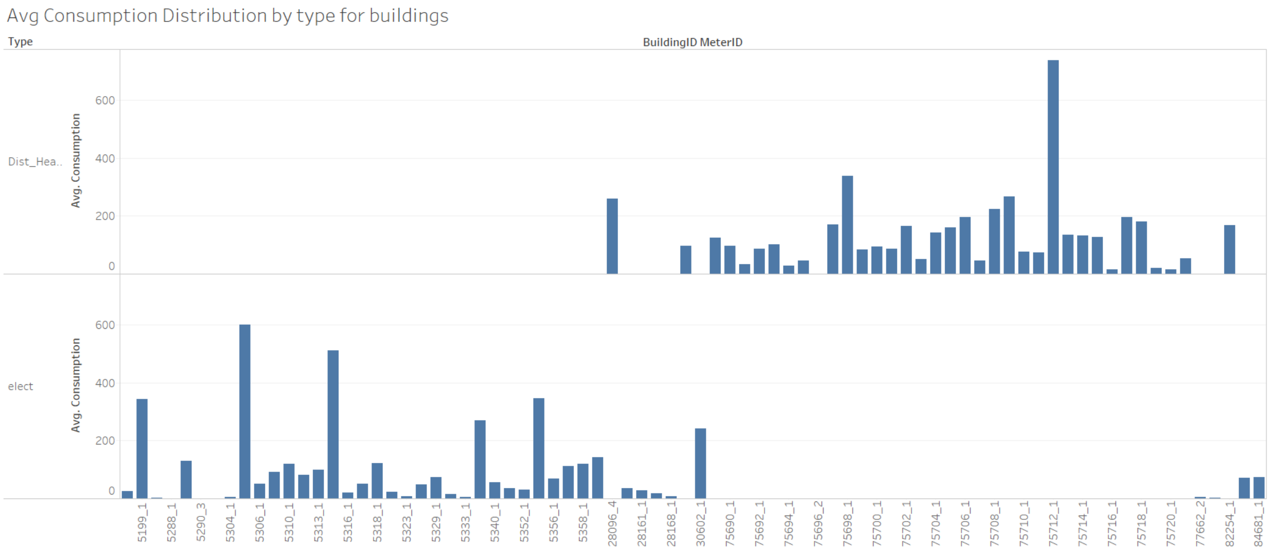
Part 3: Visualization and Dashboard

1. Consumption Analysis Dashboard

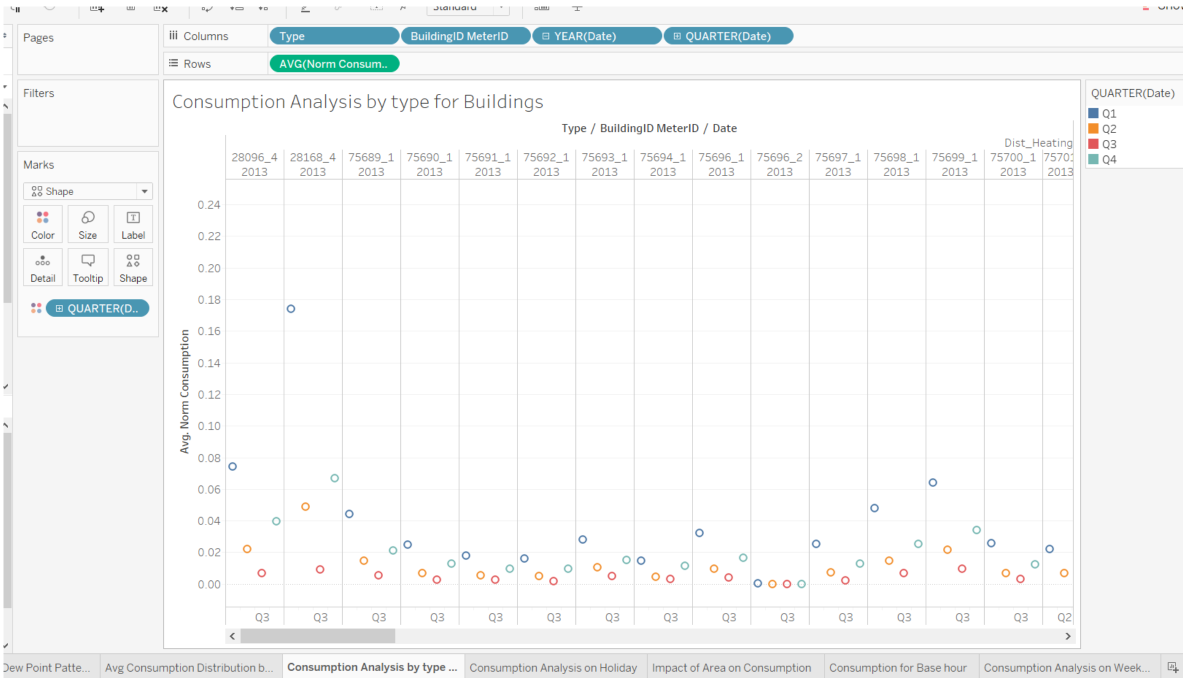
a) Avg Consumption comparison by type for Buildings



Average of Consumption for each BuildingID MeterID broken down by Type.

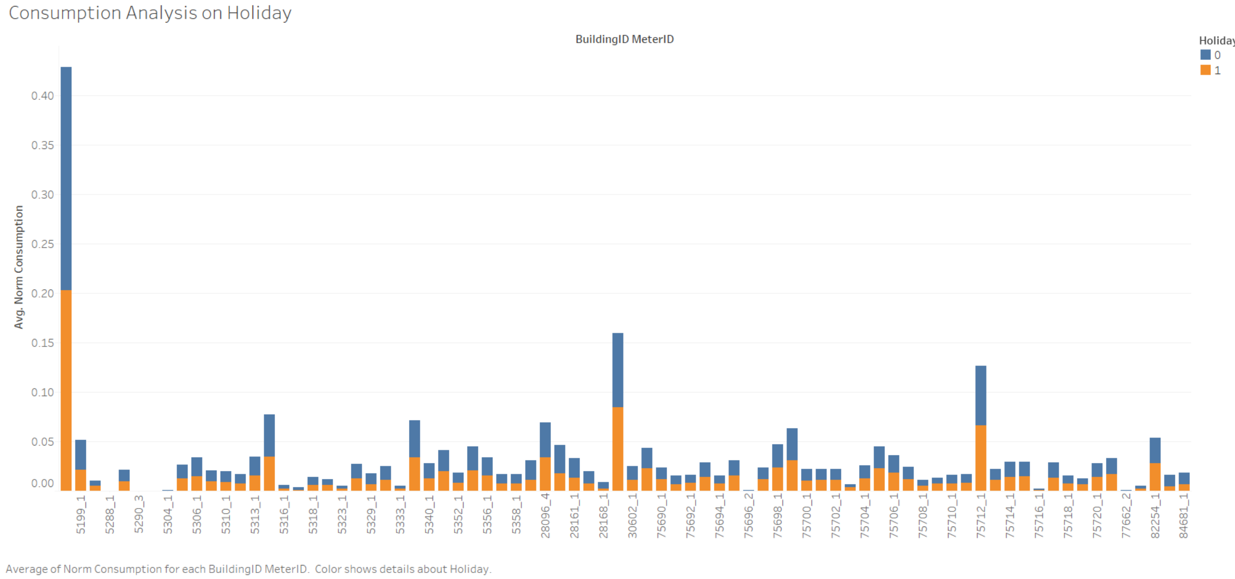
- Plot displays the Average normalised consumption of a building. Here the building refers to Building_MeterId which is unique and is of either Electricity of heating type.
- Each histogram bar gives value per building for the entire year.
- Based on the same it can be said that contribution of Heating to energy consumption is more as compared to Electricity for the year.

b) Consumption Trend Analysis by type for Building based on year hierarchy



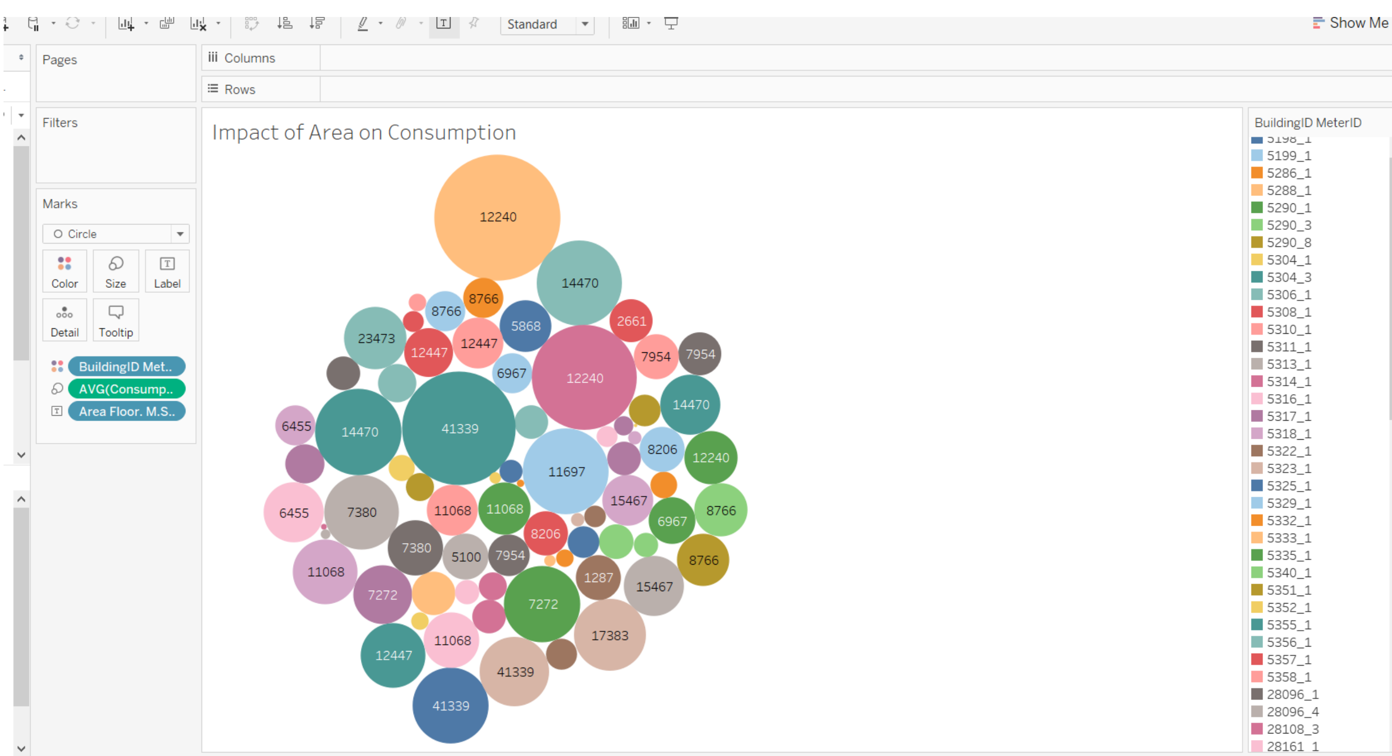
- Plot gives a detail picture of Avg Normalized consumption for each building based on year hierarchy.
- Thus, you can see the trend for consumption for each building. There was no generalised trend for each of the building with consumption dropping, increasing or staying constant through the year.

c) Consumption Analysis on Holiday



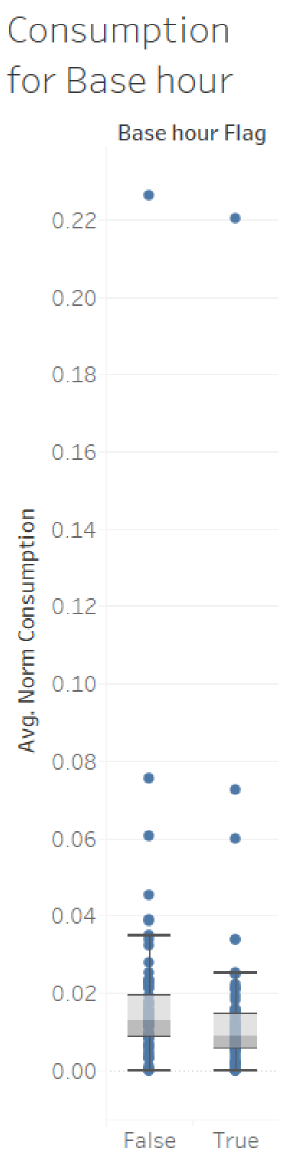
- The Average Consumption of building on holidays revealed that for most of them the consumption was higher on working days and less on holidays with a few exceptions

d) Impact of Area on consumption



- As expected this plot reveals the greater the area of the building their total consumption was found higher with respect to others.
- However, the Normalized consumption (Consumption/Area) did not necessarily reveal the same story. Meaning the normalized consumption for a small building was higher as opposed to the large area building.

e) Consumption for base Hour

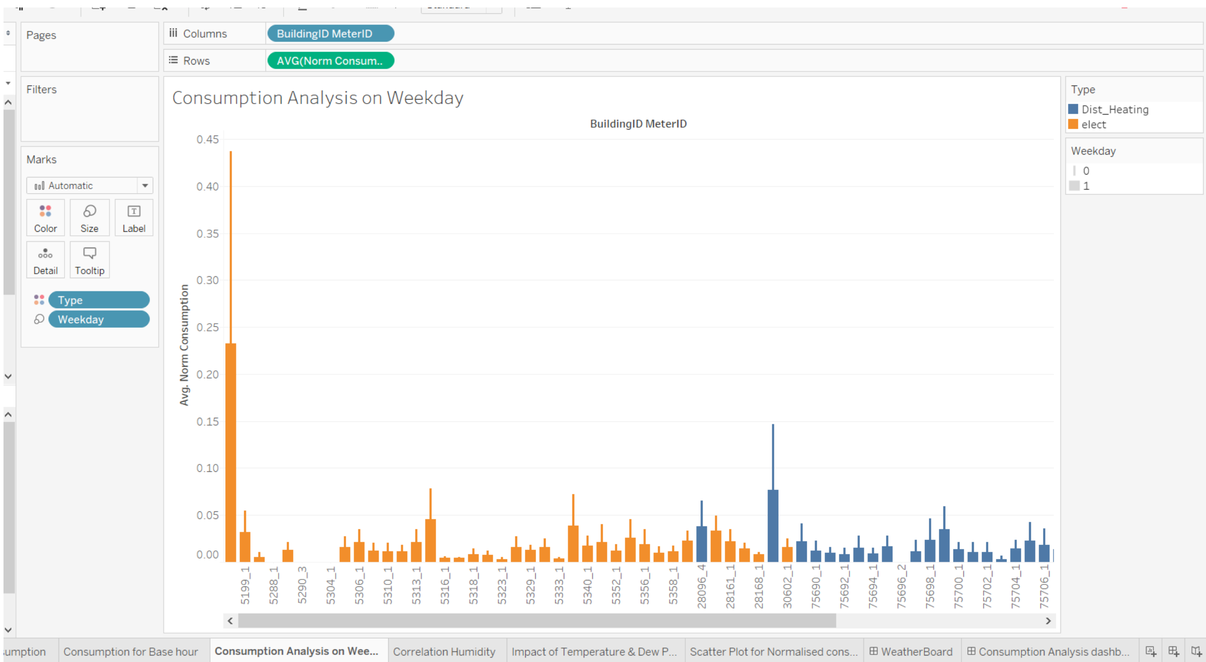


Average of Norm Consumption for each Base hour Flag. Details are shown for Type and BuildingID MeterID.

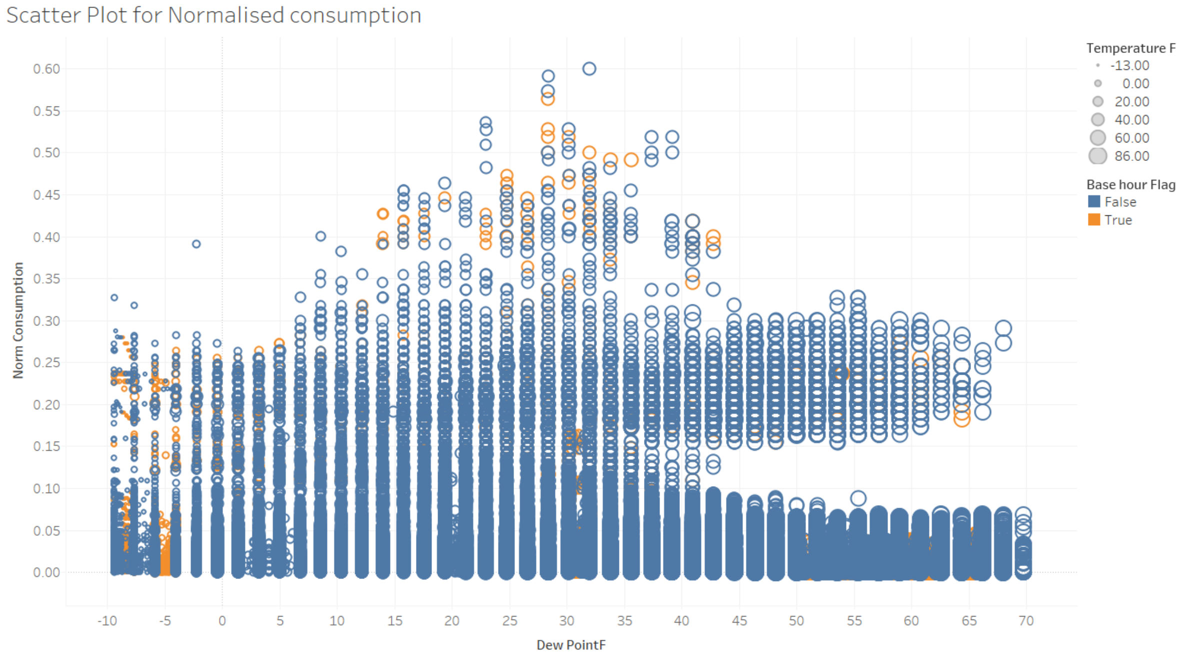
- Shows the distribution of avg Normalised consumption for base hour for all buildings.

f) Consumption Analysis on Weekday

- Shows the Avg normalised consumption for each building by type based on Weekday.
- It was found that in general the consumption was higher on weekdays than on weekend for most of the buildings.



g) Scatter plot for normalised Consumption



Dew PointF vs. Norm Consumption. Color shows details about Base hour Flag. Size shows details about Temperature F. Details are shown for Weekday and Holiday.

- Shows the correlation of normalised consumption (on Y axis) with the variables as Temperature, Dewpoint, Weekday, Holiday,BaseHourFlag.
- Thus Base on the same we would need to model our data so as to predict the normalised consumption value.

Best model:

After getting all the results, we compared the evaluation metrics of Prediction, Classification and Clustering and showed the best results.

<u>Prediction: -</u>

For prediction, we merged the evaluation output of all the models in a table. From the merged output, we grouped the value by BuildingID_MeterID and selected the min of RMSE from the grouped data. This way we got the minimum of all the 78 meters. The output clearly shows that **Random Forest** with a ntree value of 100 has the lease RMSE value for all the 78 meters and hence the best value of Prediction.

RMSE	MAE	MAPE	model	Buildingl	algorithm
0.006175	0.004325	Inf	train$Bas	28096_1	Random Forest
0.010341	0.006881	Inf	train$Bas	28108_3	Random Forest
0.005108	0.003618	Inf	train$Bas	28161_1	Random Forest
0.003827	0.002589	Inf	train$Bas	28162_1	Random Forest
0.007082	0.004174	Inf	train$Bas	28168_1	Random Forest
0.002887	0.001933	Inf	train$Bas	30602_1	Random Forest
0.032596	0.020316	Inf	train$Bas	5198_1	Random Forest
0.005785	0.003929	Inf	train$Bas	5199_1	Random Forest
0.002952	0.001619	Inf	train$Bas	5286_1	Random Forest
0	0	NA	train$Bas	5288_1	Random Forest
0.002643	0.001844	Inf	train$Bas	5290_1	Random Forest
0	0	NA	train$Bas	5290_3	Random Forest
0	0	NA	train$Bas	5290_8	Random Forest
0.000211	4.70E-05	Inf	train$Bas	5304_1	Random Forest
0.002675	0.001899	Inf	train$Bas	5304_3	Random Forest
0.004813	0.00342	Inf	train$Bas	5306_1	Random Forest
0.002024	0.00136	Inf	train$Bas	5308_1	Random Forest
0.001997	0.001375	Inf	train$Bas	5310_1	Random Forest
0.002975	0.001953	Inf	train$Bas	5311_1	Random Forest
0.005614	0.003751	Inf	train$Bas	5313_1	Random Forest
0.006539	0.00452	Inf	train$Bas	5314_1	Random Forest

<u>Classification: -</u>

For Classification, we merged the evaluation output of all the models in a table. From the merged output, we grouped the value by BuildingID_MeterID and selected the maximum of accuracy from the grouped data. The output clearly shows that **Random Forest** with a ntree value of 100 has the highest number of True positives and True Negatives for all the 78 meters and hence the best value of Classification.

	High	Low	Model	BuildingI	algorithm
High	4203	603	train$Bas	28096_1	Random Forest
Low	285	1065	train$Bas	28096_1	Random Forest
High1	4181	382	train$Bas	28108_3	Random Forest
Low1	627	966	train$Bas	28108_3	Random Forest
High2	3680	753	train$Bas	28161_1	Random Forest
Low2	359	1364	train$Bas	28161_1	Random Forest
High3	3796	571	train$Bas	28162_1	Random Forest
Low3	466	1323	train$Bas	28162_1	Random Forest
High4	3317	479	train$Bas	28168_1	Random Forest
Low4	565	1795	train$Bas	28168_1	Random Forest
High5	4297	276	train$Bas	30602_1	Random Forest
Low5	815	768	train$Bas	30602_1	Random Forest
High6	2131	955	train$Bas	5198_1	Random Forest
Low6	862	2208	train$Bas	5198_1	Random Forest
High7	4723	350	train$Bas	5199_1	Random Forest
Low7	360	723	train$Bas	5199_1	Random Forest
High8	2378	932	train$Bas	5286_1	Random Forest
Low8	1695	1151	train$Bas	5286_1	Random Forest
High9	4022	442	train$Bas	5290_1	Random Forest
Low9	691	1001	train$Bas	5290_1	Random Forest
High10	127	43	train$Bas	5304_1	Random Forest

Clustering

We find the value of between_ss/total ss and a higher ratio indicates better results.

For K means clustering with k=10: -

```
Within cluster sum of squares by cluster:
 [1]       0.0       0.0  609421.4 1389757.9  111070.2  305624.8  350525.7 4450483.7  553842.9
[10]  823611.0
 (between_SS / total_SS =  99.6 %)
```

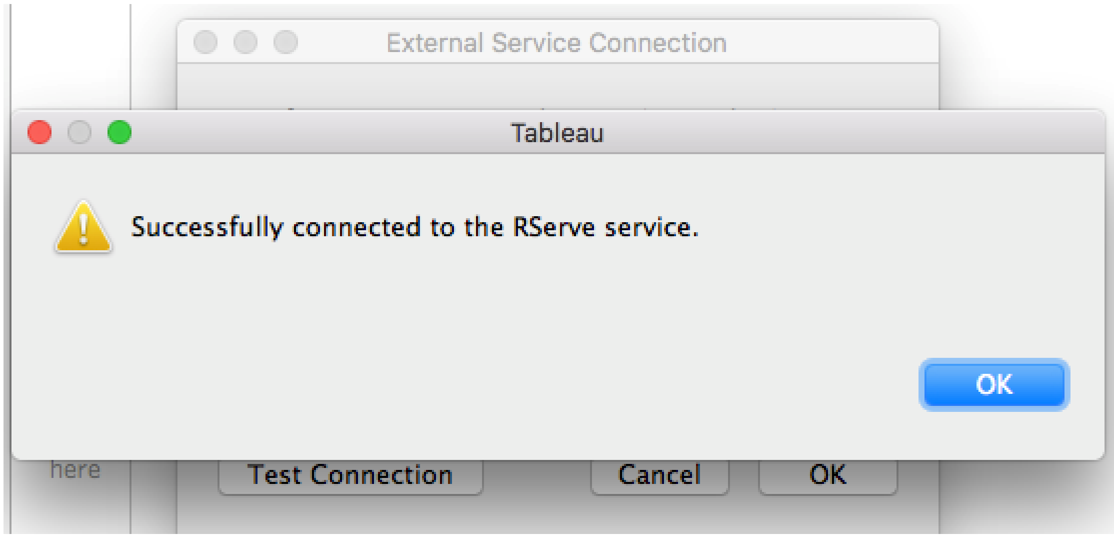
For Hierarchal clustering of k=12: -

```
> print(elbow.obj)
$k
[1] 12

$ev
[1] 0.9545588

$inc.thres
[1] 0.01

$ev.thres
[1] 0.95
```

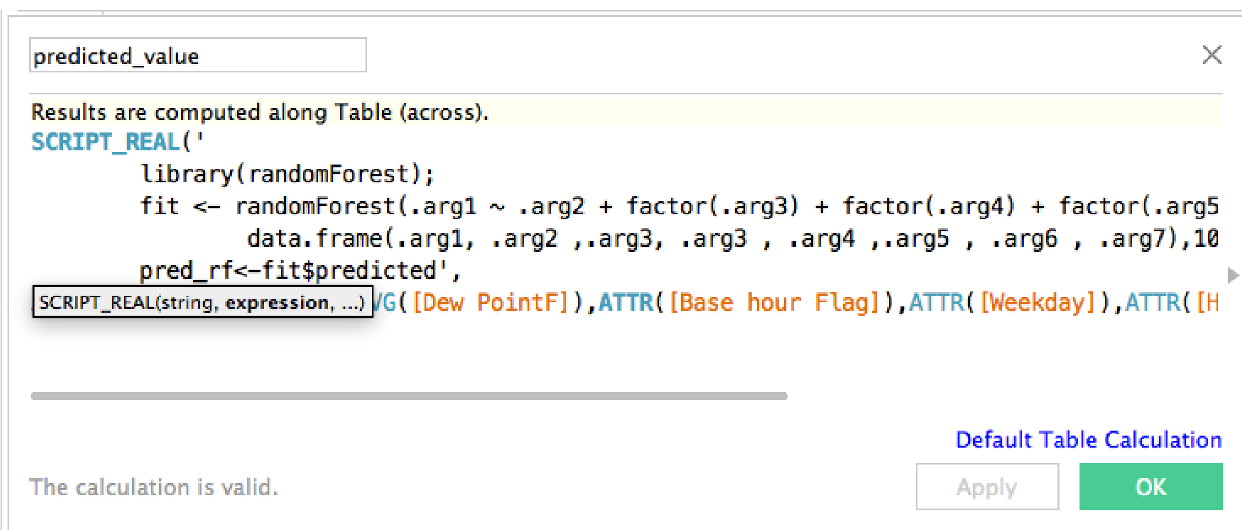
We can’t compare k means clustering directly with Hierarchal clustering because they have different evaluation metrics. However, time complexity of K-means clustering O(n) is better than time complexity of Hierarchal clustering, hence we chose k means as the better of the two.

Part 3: RServe

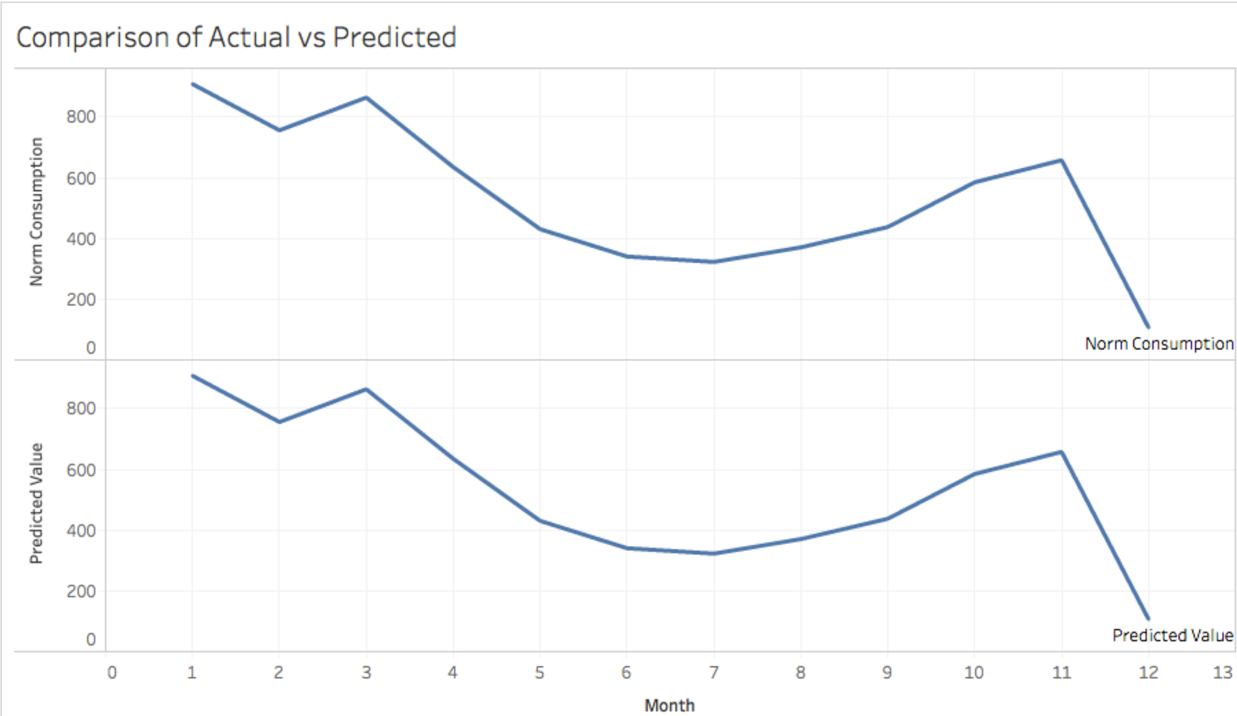
Step 1 : Invoking R from Tableau



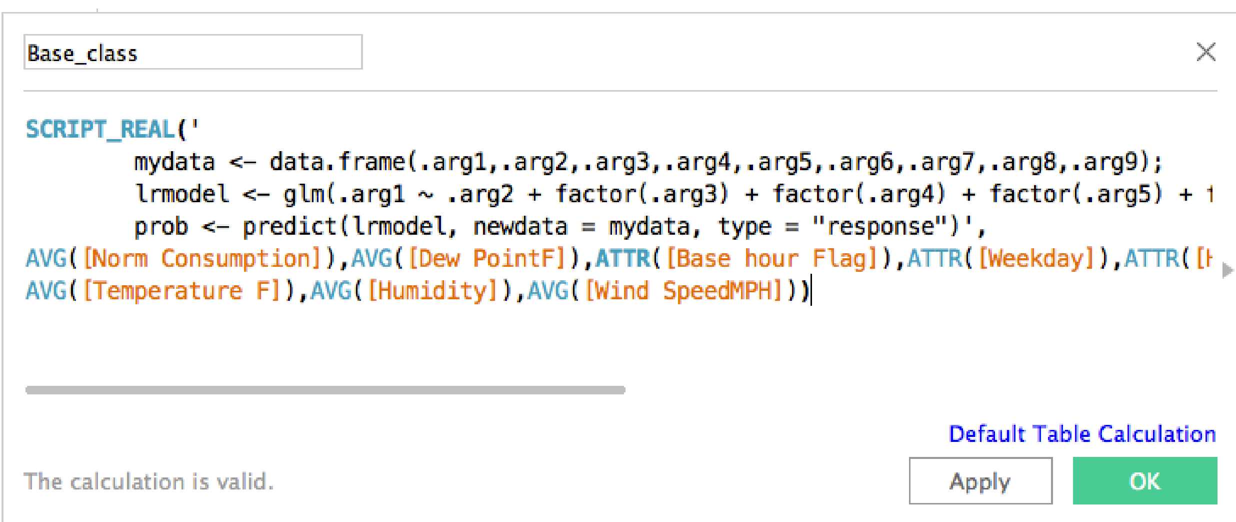
Step 2: Invoking best prediction model from Tableau using RServe



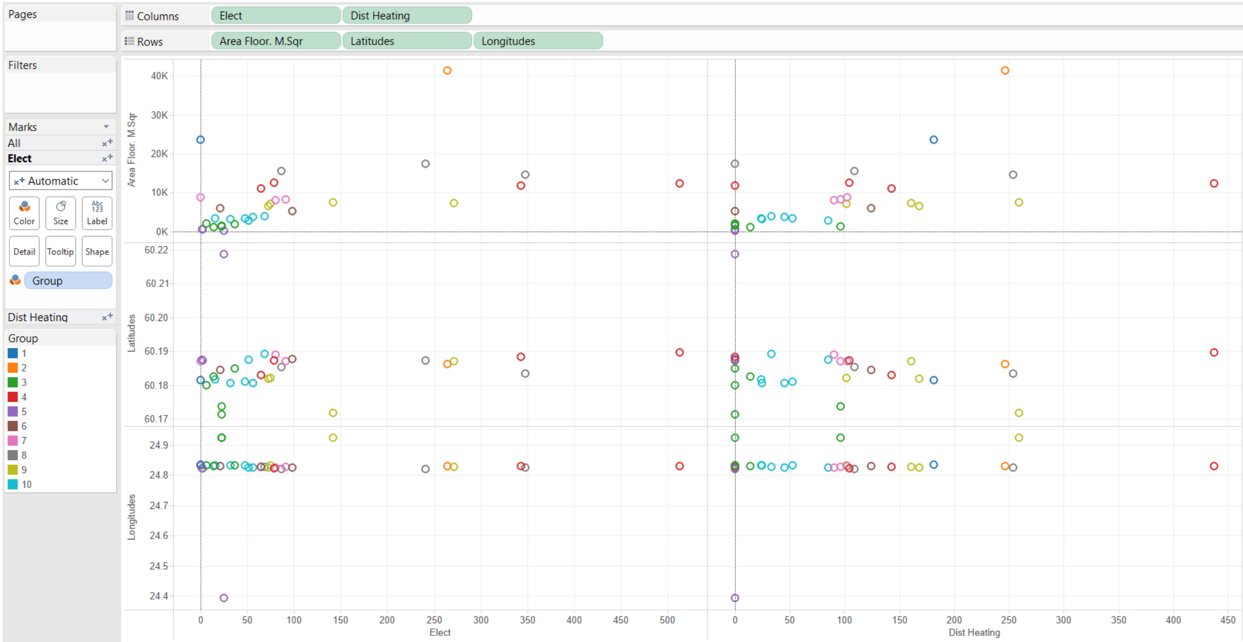
Since, we donot have future data, by simply comparing actua value to the predicted value, we can see that they are really close and hence our model has a good accuracy



Step 3: Invoking best classification model from Tableu using RServe



Clustering:



Amulya Aankul Saksham Agrawal Nand Govind Modani