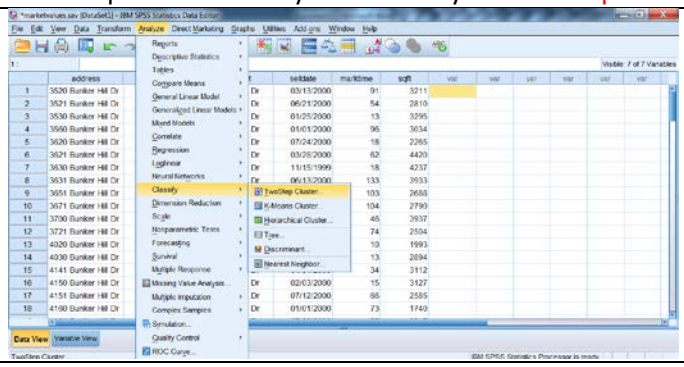
Dr. Gaurav Bansal
Bus Adm 216: Cluster Analysis Using SPSS
Start with an existing data file.

marketvalues.sav

The SPSS file is provided as an attachment to this document

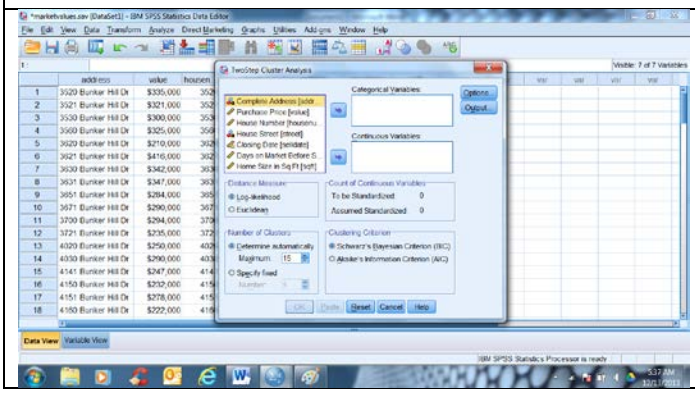
Dataset screen shot	Next step: Click on analyze >> Classify >> Two Step Cluster

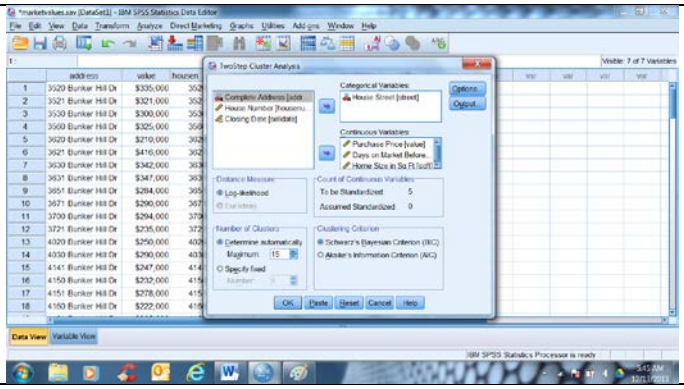
	address	value	housen...	street	selldate	marktime	sqft
1	3520 Bunker Hill Dr	$335,000	3520	Bunker Hill Dr	03/13/2000	91	3211
2	3521 Bunker Hill Dr	$321,000	3521	Bunker Hill Dr	06/21/2000	54	2810
3	3530 Bunker Hill Dr	$300,000	3530	Bunker Hill Dr	01/25/2000	13	3295
4	3560 Bunker Hill Dr	$325,000	3560	Bunker Hill Dr	01/01/2000	96	3034
5	3620 Bunker Hill Dr	$210,000	3620	Bunker Hill Dr	07/24/2000	18	2265
6	3621 Bunker Hill Dr	$416,000	3621	Bunker Hill Dr	03/28/2000	62	4420
7	3630 Bunker Hill Dr	$342,000	3630	Bunker Hill Dr	11/15/1999	18	4237
8	3631 Bunker Hill Dr	$347,000	3631	Bunker Hill Dr	06/13/2000	133	3933
9	3651 Bunker Hill Dr	$284,000	3651	Bunker Hill Dr	03/29/2000	103	2688
10	3671 Bunker Hill Dr	$290,000	3671	Bunker Hill Dr	01/01/2000	104	2790
11	3700 Bunker Hill Dr	$294,000	3700	Bunker Hill Dr	07/06/2000	46	3937
12	3721 Bunker Hill Dr	$235,000	3721	Bunker Hill Dr	09/08/1999	74	2504
13	4020 Bunker Hill Dr	$250,000	4020	Bunker Hill Dr	01/01/2000	10	1993
14	4030 Bunker Hill Dr	$290,000	4030	Bunker Hill Dr	07/31/2000	13	2894
15	4141 Bunker Hill Dr	$247,000	4141	Bunker Hill Dr	01/01/2000	34	3112



This will then open the following window.

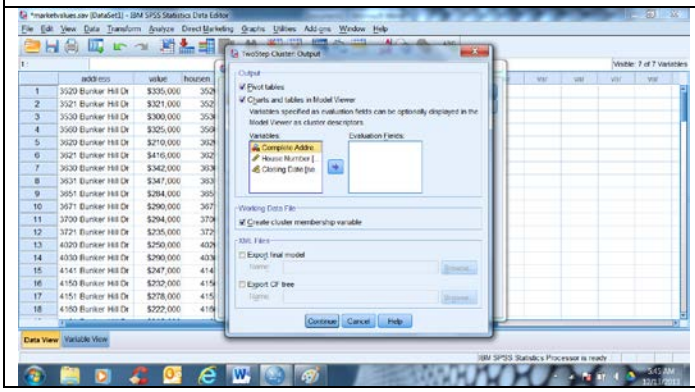
You can select both categorical and continuous variables for this analysis.	We selected House street as the categorical variable. We selected house price, square feet and days in market as continuous variables. After selecting the variables, click on the OUTPUT button.

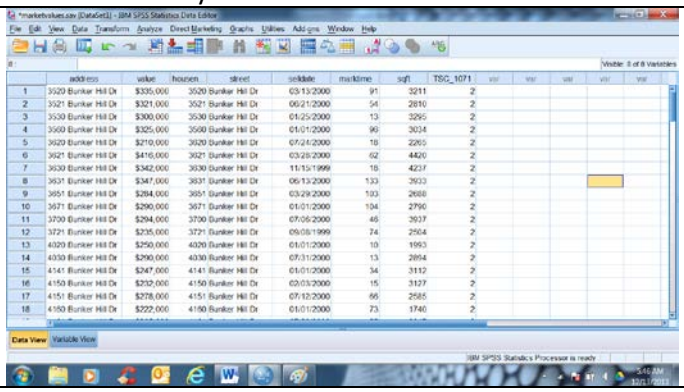




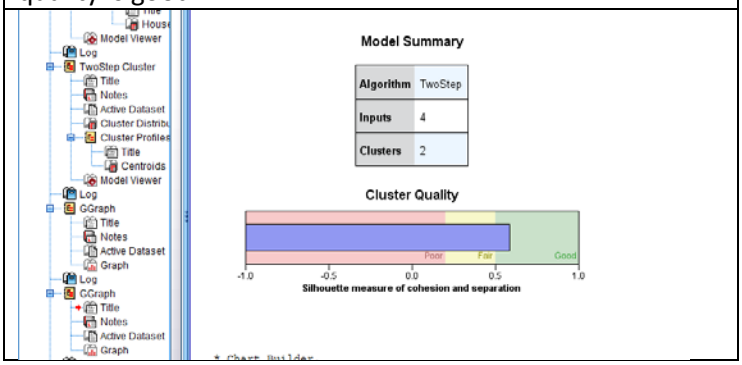
Next step:

In the following window click on the CREATE CLUSTER MEMBERSHIP variable	You can go to the data view there you will find a new variable added to your list of variables towards the end. This variable identifies the cluster membership of all the observations in your dataset.

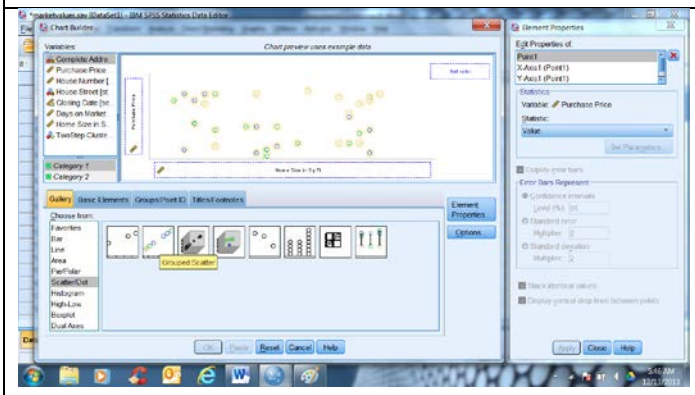
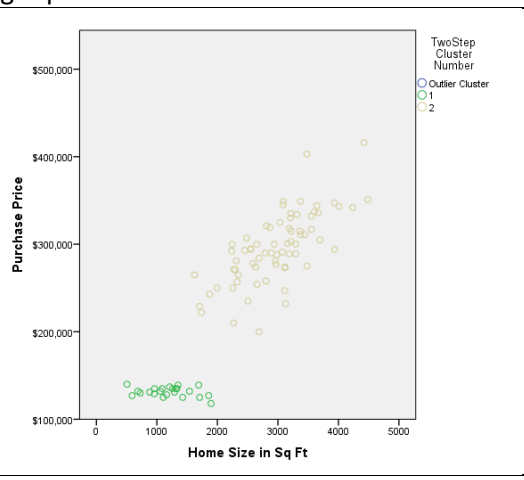




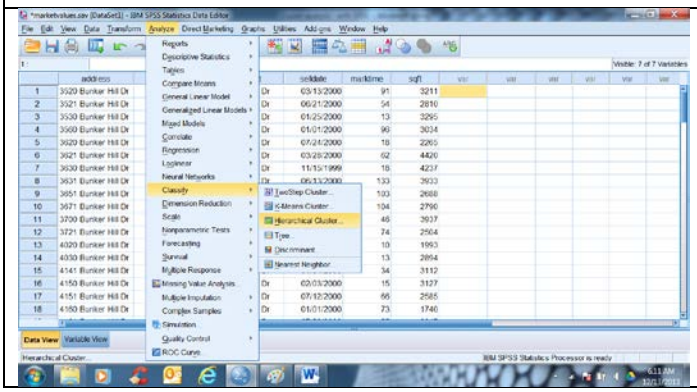
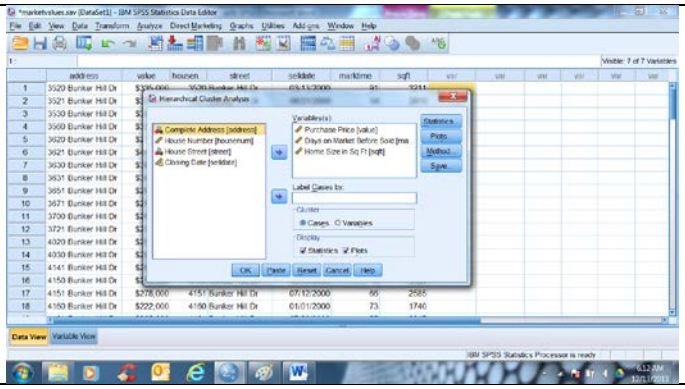
Next:

We can identify from the SPSS output that the cluster quality is good.


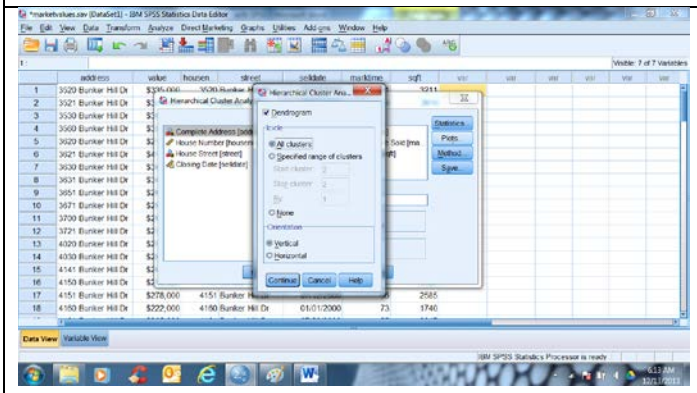
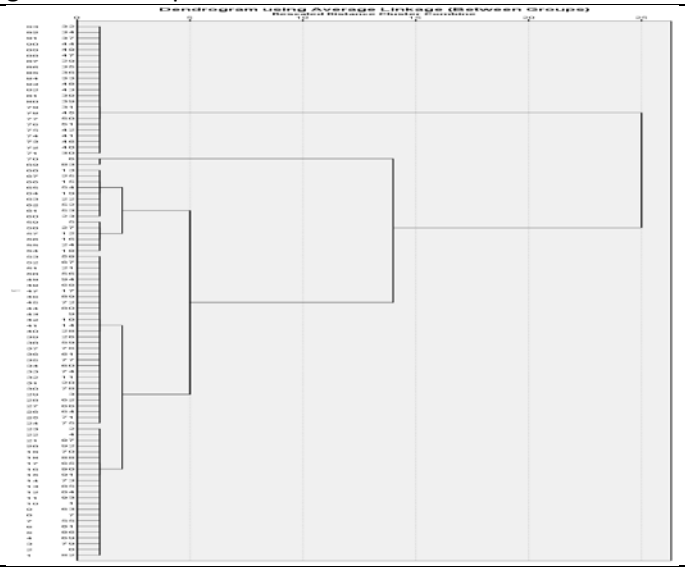
Next:

Then click on Graphs and then select Chart Builder. Select Scatter / Dot plots. Then select GROUPED SCATTER plot.	Then select Sq Ft as x variable and Purchase price as y variable. Use Two Step Cluster Membership variable to SET COLOR (shown in the upper right corner of the graph window). You need to drag and drop the variables at the right place. Then hit ok.
	

In the above TWO STEP analysis we could choose both categorical and continuous variables and the algorithm automatically identifies the suitable number of clusters possible with the variables and the data set provided. However there is another approach as well. That is to first use the Hierarchical cluster modeling and examine the dendogram. The visual examination of the dendogram would then be helpful in identifying the correct number of clusters. You can then plot them using the scatter plot (as shown above). You need to be creative in selecting the right mix of x and y variables to demonstrate the clusters on your scatter plot. The hierarchical cluster modeling only allows use of continuous variables.

Analyze >> Classify >> Hierarchical Cluster	Select the variables. It only lets you select continuous variables. Click on PLOT button.
	

>>

Click on plot and select Dendogram.	Look at the Dendogram. The Dendogram suggests three good clusters possible with the variables selected.
	

Now we do the TWO STEP cluster analysis again and we specify 3 cluster solution. And then we plot the grouped scatter plot again. This time I have modified the X and Y variables to clearly see the three clusters. You may want to refer to TWO STEP cluster analysis shown earlier in this document.	I select the same variables as I selected for Hierarchical cluster analysis. And do the cluster analysis again with Two Step algorithm. This time I specify three cluster solution. The SPSS output suggests that 3 clusters happen to be a good solution with the variables I selected.