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



The SPSS file is provided as an attachment to this document

Data	set screen	shot				Next step:	Click on	analyze	e >> (	Classi	fy >	> Tv	/o St	ep	Clust	ter		
	OOLO DOIMEI II							🔁 *marketvalues sav (DotaSet1) - IBM	SPSS Statistics Data Editor	and the second se	and the second second	2	3		1	100	X	
	address	value 1	housen	street	selldate	marktime	sqft	Eile Edt Vere Data Transform	Analyze Direct Marketing Gra	phs Willies Add-gra	Window Help						-	
1	3520 Bunker Hill Dr	\$335,000	3520 Bu	nker Hill Dr	03/13/2000	91	3211	🥮 🔚 🛞 📖 🖛 🤈	Regota   Descriptive Statedics	* * =	2 🖬 🚅	<u>ک</u> ۹	6					
2	3521 Bunker Hill Dr	\$321,000	3521 Bu	nker Hill Dr	06/21/2000	54	2810	1:	Tables +							Visible:	7 of 7 Vanables	4
3	3530 Bunker Hill Dr	\$300,000	3530 Bu	nker Hill Dr	01/25/2000	13	3295	1 3520 Bunker Hill Dr	Coyyoare Means  Qeneral Linear Model Generalget Linear Models +	Dr 03/13/200	marktime 0 91	3211	Vir	Ver Ve	- We	- 267	484	
4	3560 Bunker Hill Dr	\$325,000	3560 Bu	nker Hill Dr	01/01/2000	96	3034	2 3521 Bunker Hill Dr 3 3530 Banker Hill Dr		Dr 06/21/200 Dr 01/25/200	0 54 0 13	2810						
5	3620 Bunker Hill Dr	\$210,000	3620 Bu	nker Hill Dr	07/24/2000	18	2265	4 3550 Bunker Hill Dr	Correlate +	Dr 01/01/200	0 96	3034						
6	3621 Bunker Hill Dr	\$416,000	3621 Bu	nker Hill Dr	03/28/2000	62	4420	6 3521 Bunker Hill Dr	Begression	Dr 03/28/200	0 62	4420						
7	3630 Bunker Hill Dr	\$342,000	3630 Bu	nker Hill Dr	11/15/1999	18	4237	7 3530 Bunker Hill Dr 8 3531 Bunker Hill Dr	Neural Networks	Dr 11/15/199 Dr 05/13/200	9 18	4237						
8	3631 Bunker Hill Dr	\$347,000	3631 Bu	nker Hill Dr	06/13/2000	133	3933	9 3651 Bunker Hill Dr	Classify +	TwoStep Cluster	103	2688						
9	3651 Bunker Hill Dr	\$284,000	3651 Bu	nker Hill Dr	03/29/2000	103	2688	10 3671 Bunker Hill Dr 11 3700 Bunker Hill Dr	Scyle +	Berachical Cluster	45	3937						
10	3671 Bunker Hill Dr	\$290,000	3671 Bu	nker Hill Dr	01/01/2000	104	2790	12 3721 Bunker Hill Dr	Nonparametric Tests	El Tyre	74	2504						
11	3700 Bunker Hill Dr	\$294,000	3700 Bu	nker Hill Dr	07/06/2000	46	3937	13 4020 Bunker Hill Dr 14 4030 Bunker Hill Dr	Sunival +	B Discriminant	13	2894						
12	3721 Bunker Hill Dr	\$235,000	3721 Bu	nker Hill Dr	09/08/1999	74	2504	15 4141 Bunker Hill Dr 16 4150 Bunker Hill Dr	Multiple Response +	Dr 02/03/201	34	3112						
13	4020 Bunker Hill Dr	\$250,000	4020 Bu	nker Hill Dr	01/01/2000	10	1993	17 4151 Bunker Hill Dr	Muljple Imputation	Dr 07/12/200	0 65	2585						
14	4030 Bunker Hill Dr	\$290,000	4030 Bu	nker Hill Dr	07/31/2000	13	2894	18 4160 Bunker Hel Dr	Complex Samples	Dr 01/01/200	0 73	1740	_		_		-	
15	4141 Bunker Hill Dr	\$247,000	4141 Bu	nker Hill Dr	01/01/2000	34	3112	Carta View Versitive View	Guality Control ,		1.20		-	DES Statistics	Processor is a	-		

## 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									
월 "mateholen.ar Fondell - 99/3/25 Seines Dira Gales [Per [de ]we geta ]newlow ganges Diret@retty gants [blies: Adoges Window (blie 	Image: The state of the sta									
Image: Second	1:         Under Hote         Description           1 2020 Burner HI10         305000         20									

Next step:

In the following window click on the CREAT	You can go to the data view there you will find a new														
MEMBERSHIP variable	variable added to your list of variables towards the end.														
		This variable identifies the cluster membership of all the													
	observations in your dataset.														
🕼 *marketvalkes.sav (DataSet1) - BM SPSS Statistics Data Editor		🕼 *marketnaker.com (DataSett) - BM SFSS Statistics Data Editor													
Ele Edi Yew Data Transform Analyze Direct Harlweing Graphic Utilities Addigns Window Help		Ein Edi	t View Data Transform	n <u>A</u> nalyze D	rect Marketing Graphs U	illies Addigens <u>W</u> a	ndow <u>Help</u>								
		(2) h	1 @ III m	-				39	16						
Carl and the Carl and the Carl and Carl	Master 2 of 7 Valation			110 84		Henry 14-					hours	an 2 of 8 Variables			
address whe borned Odput	war unit wir war	<u> </u>	and the serves	without 1	ninen street	Indiate	marktime	and 1	TSC 1071	Mar uni	T wher				
1 3520 Bunker Hill Dr \$335.000 3521 # Protitives	E	1	3520 Bunker Hill Dr	\$335.000	3520 Bunker Hill Dr	03/13/2000	91	3211	2			E			
2 3521 Bunker Hill Dr \$321,000 352 Cparts and tables in Model Viewer	and the second se	2	3521 Bunker Hill Dr	\$321,000	3521 Banker Hill Dr	06/21/2000	54	2810	2						
3 3530 Bunker Hill Dr \$300,000 353 Variables specified as evaluation fields can be optionally displayed in the		3	3530 Bunker Hill Dr	\$300,000	3530 Banker Hill Dr	01/25/2000	13	3295	2						
4 3560 Bunkar Hill Dr \$325,000 359		4	3560 Bunker Hill Dr	\$325,000	3560 Bunker Hill Dr	01/01/2000	96	3034	2						
5 3920 Bunker Hill Dr \$210,000 392		5	3620 Bunker Hill Dr	\$210,000	3620 Sunker Hill Dr	07/24/2000	18	2265	2						
6 3621 Bunker Hill Dr \$416,000 362		6	3621 Bunker Hill Dr	\$416,000	3021 Bunker Hill Dr	03/25/2000	62	4420	2						
7 3630 Bunker Hill Dr \$342,000 363 🖉 Closing Date (tria		7	3630 Bunker Hill Dr	\$342,000	3630 Banker Hill Dr	11/15/1999	18	4237	2						
8 3531 Dunker Hill Dr \$547,000 353		8	3631 Dunker Hill Dr	\$347,000	3631 Dunker Hill Dr	06/13/2000	133	3933	2			4			
9 3051 Direct Hill (P \$204,000 305)		9	3651 Burker Hill Dr	\$254,000	3051 Dunker Hel Dr	03/29/2000	103	2000							
10 2011 Bartise Hill Dr. \$204,000 307 Honora Data File		10	1200 Burker Hit Dr	\$294,000	3700 Dunker Mit Dr	01/01/2000	46	2090							
12 1/21 Duraw Hill Dr. \$235.000 1/2		12	3721 Bucker Mil Dr	\$215,000	3721 Dunker MI Dr	09/08/1999	74	240.4	2						
13 4020 Durater Hill Dr 5250 000 403 1000 1415		13	4020 Durker Hill Dr	\$250.000	4020 Busker Hill Dr	01/01/2000	10	1993	2						
14 4030 Burker Hill Dr \$290.000 403 DEport Inal model		14	4030 Bucker Hill Dr	\$290,000	4030 Bunker Hill Dr.	07/31/2000	13	2094	2						
15 4145 Burker Hill Dr \$247,000 414		15	4545 Bunker Hill Dr	\$247,000	4141 Bunker Hil Dr	01/01/2000	34	3112	2						
16 4150 Burker Hill Dr \$232,000 415 Export OF bree		16	4150 Bunker Hill Dr	\$232,000	4150 Banker Hill Dr.	02/03/2000	15	3127	2						
17 4151 Burker Hil Dr 5278,000 415 James		17	4151 Bunker Hill Dr	\$278,000	4151 Bunker Hill Dr	07/12/2000	66	2585	2						
18 4150 Burker Hill Dr \$222,000 410		18	4160 Bunker Hill Dr	\$222,000	4160 Bunker Hill Dr	01/01/2000	73	1740	2						
Contrast Cancel Help			1				**					2			
	dics Processor is ready	Data Vie	Valizzie View	0	é 🗰 🛯		188	3.57		estes Piscesser in	ready for all	SIGAM			



Next:



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.

nury20 >>	Classify /	~	inera	I CHI		JSLEI					Sei	• • •		es. II		ou sele		ILIII	JUUS	2		
											vai	lables.	Click of	n PLC	JI button.							
navketvolues.sav (DataSet1) - IBM S	PSS Statistics Data Editor		AND ROOM AND		270 C			-	-0	X	E *mark	tvslues.sav (DataSet1) - B	M SPSS Statistics Data Edit	lor	Monad (ref.			-		<b>3</b> 3		
Edit View Data Transform	Analyze Deect Marketing Gra	phs Us	ites Add.ges We	ndow Help		1			-		Ele Ed	Vew Data Transform	n Analyze Deect Harle	ting Quarter (	Utilities Addigns Window Help							
HAMERA	Reports +	*	1 = 5		A A ANS									1 H ¥	🗑 💟 📖 🚓 🛄 📢 G	A16						
La (my mag	Descriptive Statistics +	- later	and here we	1111 (J.)4																		
	Tables •	-							procese: 7 of 7 v	arresers.	10			1007				-	proster: 7 or	A / VANNES		
BOD ESS	Compare Moans *	10	SERVICE	maname	SQT V	5	- WH	491	Var s	101		1002 CS5	value nousen	Steel	seikare mankime	SQE VE	- VII - VII	4 1 WH	Als.			
3520 Bonker Hill Dr	General Linear Model 🔰	Dr.	0513/2000	91	3211						2	3520 Bunker Hill Dr	of the Herenchical Clu	rden Acalvas	accenter an	- 2 -						
VS10 Bunker Hill Dr	Generalized Linear Models +	Dr	01/25/2000	13	3295							3530 Bunker Hill Dr	-	1000 X 100		1.						
3500 Bunker Hill Dr	Miged Models +	Dr	01/01/2000	66	301.8							3500 Bunker Hill Dr			Variables(s)	Etablistics :						
3620 Burker Hit Dr	Gorelate P Begression P Logiment P	Dr	07/24/2000	18	2265						5	3020 Burker Hit Dr	Complete Ad	vess (sources)	<ul> <li>Portnase Price (value)</li> <li>On-a on Market Defice Sale</li> </ul>	Piets						
3621 Bunker Hit Dr		Dr	03/25/2000	10	4420						6	3621 Bunker Hit Dr	Se A House Street	[street]	Home Size in So Pt Ingti	Method						
3630 Bunker Hill Dr		Dr	11/15/1999	18	4237						7	3630 Bunker Hill Dr	a Closing Date	(seikdar)		Sam						
3631 Bunker Hill Dr	Neural Networks	Dr	06/13/2000	133	3903							3631 Bunker Hill Dr	57									
3651 Bunker Hill Dr	Classify +	BIL-	oStep Cluster	103	2688						9	3851 Bunker Hill Dr	5		Label Gaues by:							
3671 Dunker Hill Dr	Dimension Reduction +	E 64	Means Cluster	104	2790						10	3671 Dunker Hill Dr	53									
1 3700 Bunker Hill Dr	Scale +	10144	rathical Gluber	46	3937						11	3700 Bunker Hill Dr	52		Christer							
2 3721 Bunker Hill Dr	Nonparametric Tests	CIT.		74	2504						12	3721 Bunker Hill Dr	52		Cases O Vasables							
4020 Euriker Hill Dr	Forecasting +	10.00		10	1993						13	4020 Bunker Hill Dr	52		Display							
4 4030 Bunker Hill Dr	Surved >	-		13	2894						14	4030 Bunker Hill Dr	\$2		Statistics & Flats							
5 4545 Bunker Hill Dr	Myltple Response +	100	ares negros	34	3112						15	4545 Bunker Hill Dr	si	OK	Patte Reset Cancel Hep							
6 4150 Bunker Hill Dr	Missing Value Analysis	Dr	62/03/2000	15	3127						16	4150 Bunker Hill Dr	5	Contraction of								
7 4151 Bunker Hill Dr	Nuljple Imputation +	Dr	07/12/2000	66	2585						17	4151 Bunker Hil Dr	\$278,000 4151	Sunker Hill Dr	07/12/2000 66	2585						
4150 Banker Hill Dr	Complex Samples +	Dr	01/01/2000	73	1740						18	4160 Banker Hill Dr	\$222,000 4160	Bunker Hill Dr	01/01/2000 73	1740						
	5 Simulation	-		**	1000													- 11 mar	_			
View Vaciable View	Quality Control +			10.00							Data Vie	w Variable View			1000							
Concession of the local division of the loca	ROC Curve											Contraction of the local division of the loc										





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	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							
see the three clusters. You may want to refer to TWO STEP	specify three cluster solution. The SPSS output							
cluster analysis shown earlier in this document.	suggests that 3 clusters happen to be a good							
	solution with the variables I selected.							
Stop.000- Stop.000-	Model Summary Algorithm TwoStop Inputs 3 Clusters 3 Cluster Quality -1.0 0.5 00 0.5 1.0 Silhouette measure of cohesion and separation							