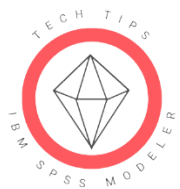




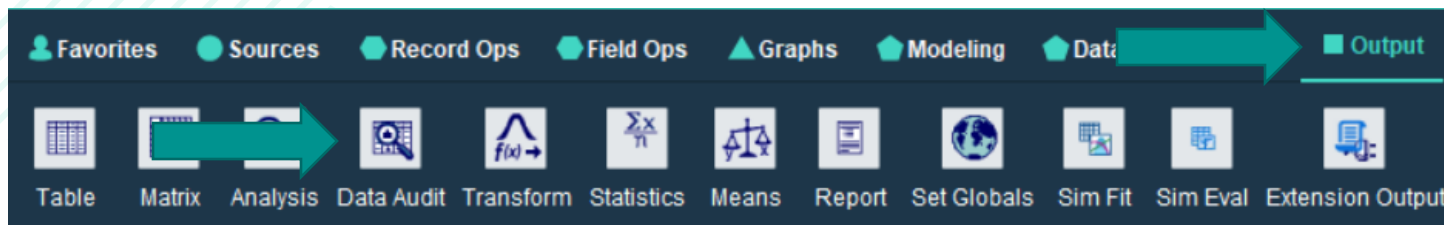
Tech Tips - Quickly Audit Data in IBM SPSS Modeler

Tech Tips - Quickly Audit Data in IBM SPSS Modeler



Tech Tips - Quickly Audit Data

- Here's a quick tip to audit your data file in IBM SPSS Modeler.
- Users can obtain a quick audit and overview of a data file by using the **Data Audit** node. The **Data Audit** node is located on the **Output** palette and provides visualizations of fields within the data file. The **Data Audit** also identifies outliers, extreme cases and missing data.



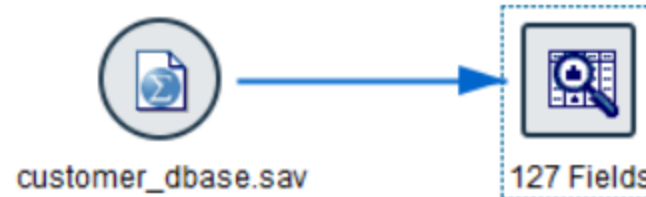
The Data Audit window displays a table of field statistics and sample graphs. The table includes columns for Field, Sample Graph, Measurement, Min, Max, Mean, Std Dev, Skewness, Median, Mode, Unique, and Valid. The fields listed are region, townsize, age, agecat, birthmonth, ed, educ, and jobcat.

Field	Sample Graph	Measurement	Min	Max	Mean	Std Dev	Skewness	Median	Mode	Unique	Valid
region	[Sample Graph]	Nominal	1,000	5,000	—	—	—	—	5,000	5	5,000
townsize	[Sample Graph]	Ordinal	1,000	5,000	—	—	—	—	1,000	5	4,998
age	[Sample Graph]	Continuous	18,000	79,000	47,026	17,770	0.091	47,000	18,000	—	5,000
agecat	[Sample Graph]	Ordinal	2,000	6,000	—	—	—	—	4,000	5	5,000
birthmonth	[Sample Graph]	Nominal	—	—	—	—	—	—	September	12	5,000
ed	[Sample Graph]	Continuous	6,000	23,000	14,543	3,281	0.004	14,000	14,000	—	5,000
educ	[Sample Graph]	Ordinal	1,000	5,000	—	—	—	—	2,000	5	5,000
jobcat	[Sample Graph]	Nominal	1,000	6,000	—	—	—	—	2,000	6	5,000

* Indicates a multimode result * Indicates a sampled result

Tech Tips - Quickly Audit Data

- To audit data go to the **Output** palette. Select the **Data Audit** node and drag it onto the stream canvas. You can also double click the node to drop it onto the stream canvas. Once it is on the canvas you can connect it to your stream.
- When you connect the node to your stream (data) you will see the number of fields in your data file. In this case, there are 127 fields in our data file. We want to get an overview of all fields.



Tech Tips - Quickly Audit Data

- Double click to open the node. If we wanted to audit a subset we could click the **Use custom fields** button. In this case, we will select **Default** to audit all fields. We have also asked for **Graphs**, and **Basic Statistics**. There is also the option to **Calculate the median and mode**.



Data Audit of [127 fields] #3

Audit Quality Annotations

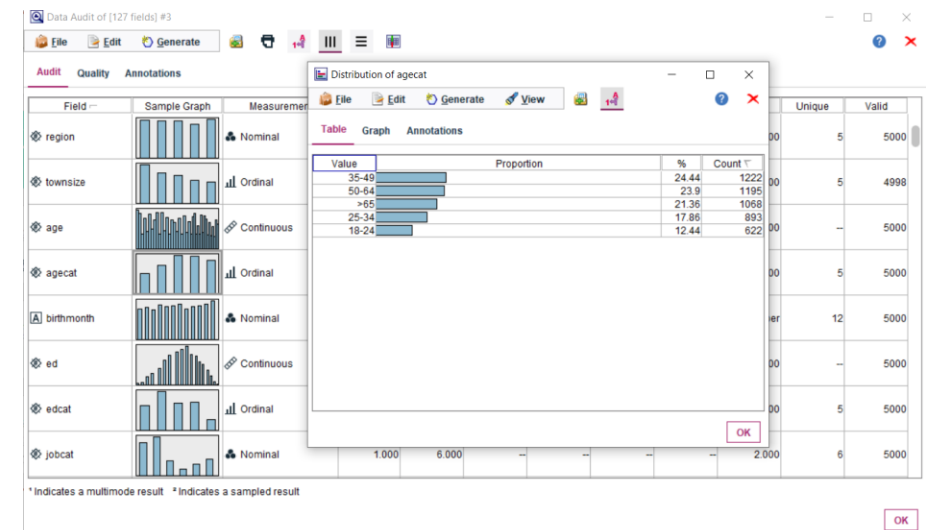
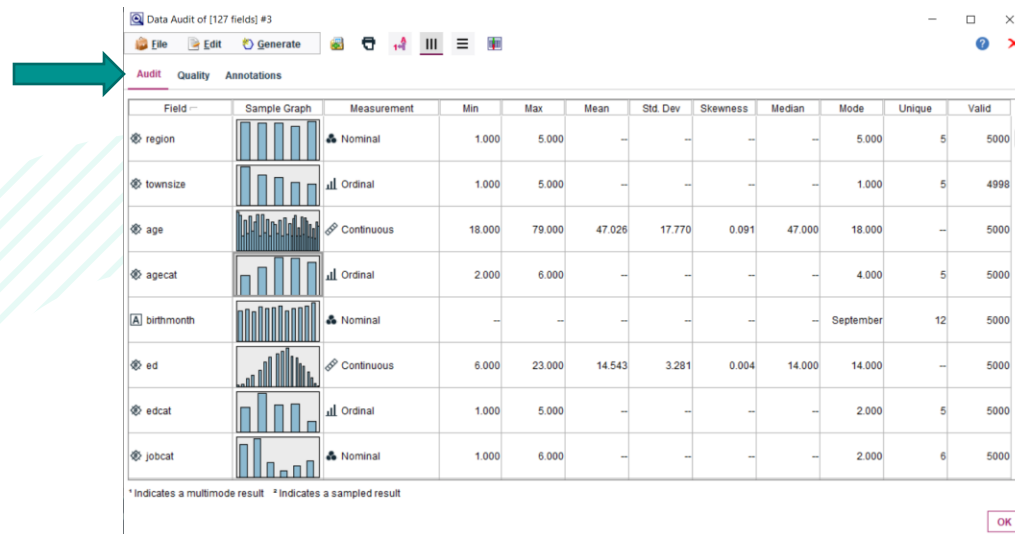
Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Median	Mode	Unique	Valid
region		Nominal	1.000	5.000	--	--	--	--	5.000	5	5000
townsize		Ordinal	1.000	5.000	--	--	--	--	1.000	5	4998
age		Continuous	18.000	79.000	47.026	17.770	0.091	47.000	18.000	--	5000
agecat		Ordinal	2.000	6.000	--	--	--	--	4.000	5	5000
birthmonth		Nominal	--	--	--	--	--	--	September	12	5000
ed		Continuous	6.000	23.000	14.543	3.281	0.004	14.000	14.000	--	5000
edcat		Ordinal	1.000	5.000	--	--	--	--	2.000	5	5000
jobcat		Nominal	1.000	6.000	--	--	--	--	2.000	6	5000

* Indicates a multimode result * Indicates a sampled result

OK

Tech Tips - Quickly Audit Data

- On the **Audit** tab each field is listed with a thumbnail graph and statistics. Users can click on a thumbnail graph to see distributions or histograms.



Tech Tips - Quickly Audit Data

- On the **Quality** tab, **Field**, **Measurement**, **Outliers**, **Extremes** and missing data are shown. Users can sort on **% Complete** to see which fields have missing data. Users can click on the column top and use the arrow up or arrow down option to sort the fields on missing.
- The **Data Audit** is a fast overview of your data available in just a few clicks.

Complete fields (%) 88.19% Complete records (%) 11.55%

Field	Measurement	Outliers	Extremes	Action	Imp	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
region	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
townsize	Ordinal	---	---	Never	Fixed	99.98	4999	2	0	0	0
age	Continuous	0	0	Never	Fixed	100	5000	0	0	0	0
agecat	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
bathsmth	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
ed	Continuous	0	0	Never	Fixed	100	5000	0	0	0	0
educat	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
jobcat	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
union	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
employ	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
empcat	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
retire	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
income	Continuous	63	23	Never	Fixed	100	5000	0	0	0	0
ininc	Continuous	15	0	Never	Fixed	100	5000	0	0	0	0
inccat	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
debtinc	Continuous	52	1	Never	Fixed	100	5000	0	0	0	0
creddebt	Continuous	50	23	Never	Fixed	100	5000	0	0	0	0
incdebtcat	Continuous	27	1	Never	Fixed	99.98	4999	1	0	0	0
obtdcat	Continuous	65	28	Never	Fixed	100	5000	0	0	0	0
hothdebt	Continuous	24	0	Never	Fixed	99.98	4999	1	0	0	0
default	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
jobcat	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
marital	Ordinal	---	---	Never	Fixed	100	5000	0	0	0	0
married	Continuous	---	---	Never	Fixed	100	5000	0	0	0	0
spoused	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
spousedcat	Continuous	0	0	Never	Fixed	100	5000	0	0	0	0
reside	Continuous	38	0	Never	Fixed	100	5000	0	0	0	0
reside	Continuous	38	0	Never	Fixed	100	5000	0	0	0	0
pets	Continuous	36	1	Never	Fixed	100	5000	0	0	0	0
pets_cats	Continuous	24	7	Never	Fixed	100	5000	0	0	0	0
pets_dogs	Continuous	153	2	Never	Fixed	100	5000	0	0	0	0
pets_birds	Continuous	88	70	Never	Fixed	100	5000	0	0	0	0
pets_reptiles	Continuous	0	0	Never	Fixed	100	5000	0	0	0	0
pets_small	Continuous	83	90	Never	Fixed	100	5000	0	0	0	0
pets_smallfish	Continuous	11	44	Never	Fixed	100	5000	0	0	0	0
pets_freshfish	Continuous	32	0	Never	Fixed	100	5000	0	0	0	0
homeown	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0
homeown	Nominal	---	---	Never	Fixed	100	5000	0	0	0	0

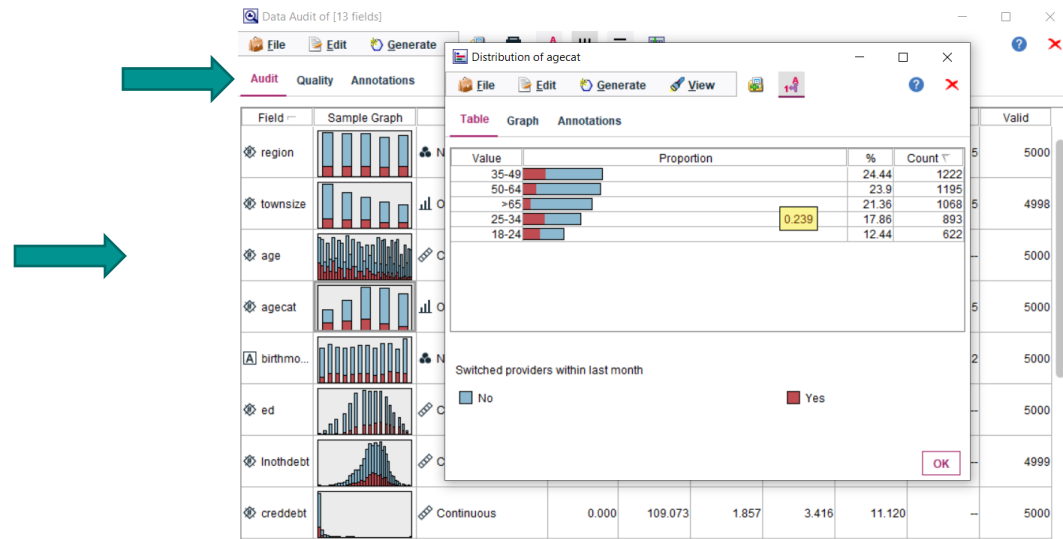
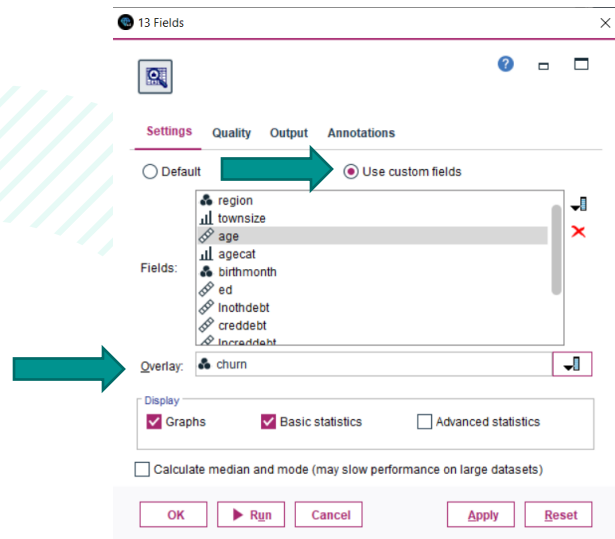


Complete fields (%) 88.19% Complete records (%) 11.55%

Field	Measurement	Outliers	Extremes	Action	Impulse Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
inwten	Continuous	5	0	None	Never	Fixed	26.88	1344	3656	0	0	0
inwten	Continuous	10	0	None	Never	Fixed	26.88	1344	3656	0	0	0
leaquipm	Continuous	8	0	None	Never	Fixed	34.08	1704	3296	0	0	0
leaquipm	Continuous	15	0	None	Never	Fixed	34.08	1704	3296	0	0	0
idmten	Continuous	3	0	None	Never	Fixed	47.56	2378	2622	0	0	0
idmten	Continuous	35	0	None	Never	Fixed	47.56	2378	2622	0	0	0
icardten	Continuous	58	0	None	Never	Fixed	71.56	3578	1422	0	0	0
icardmten	Continuous	7	0	None	Never	Fixed	71.62	3581	1419	0	0	0
longten	Continuous	79	28	None	Never	Fixed	93.54	4697	3	0	0	0
longten	Continuous	27	0	None	Never	Fixed	93.54	4697	3	0	0	0
lowincat	Ordinal	---	---	Never	Fixed	---	99.96	4998	2	0	0	0
commuttime	Continuous	25	0	None	Never	Fixed	99.96	4998	2	0	0	0
cardten	Continuous	62	19	None	Never	Fixed	99.96	4998	2	0	0	0
increddebt	Continuous	27	1	None	Never	Fixed	99.98	4999	1	0	0	0
increddebt	Continuous	24	0	None	Never	Fixed	99.98	4999	1	0	0	0
region	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
age	Continuous	---	---	Never	Fixed	---	100	5000	0	0	0	0
apcat	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
bathsmth	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
ed	Continuous	0	0	None	Never	Fixed	100	5000	0	0	0	0
educat	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
jobcat	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
union	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
employ	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
empcat	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
retire	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
income	Continuous	63	23	None	Never	Fixed	100	5000	0	0	0	0
ininc	Continuous	15	0	None	Never	Fixed	100	5000	0	0	0	0
inccat	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
debtinc	Continuous	52	1	None	Never	Fixed	100	5000	0	0	0	0
creddebt	Continuous	50	23	None	Never	Fixed	100	5000	0	0	0	0
incdebtcat	Continuous	27	1	None	Never	Fixed	100	5000	0	0	0	0
obtdcat	Continuous	65	28	None	Never	Fixed	100	5000	0	0	0	0
hothdebt	Continuous	24	0	None	Never	Fixed	100	5000	0	0	0	0
default	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
jobcat	Ordinal	---	---	Never	Fixed	---	100	5000	0	0	0	0
married	Continuous	---	---	Never	Fixed	---	100	5000	0	0	0	0
spoused	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
spousedcat	Continuous	0	0	None	Never	Fixed	100	5000	0	0	0	0
reside	Continuous	38	0	None	Never	Fixed	100	5000	0	0	0	0
reside	Continuous	38	0	None	Never	Fixed	100	5000	0	0	0	0
pets	Continuous	36	1	None	Never	Fixed	100	5000	0	0	0	0
pets_cats	Continuous	24	7	None	Never	Fixed	100	5000	0	0	0	0
pets_dogs	Continuous	153	2	None	Never	Fixed	100	5000	0	0	0	0
pets_birds	Continuous	88	70	None	Never	Fixed	100	5000	0	0	0	0
pets_reptiles	Continuous	0	0	None	Never	Fixed	100	5000	0	0	0	0
pets_small	Continuous	83	90	None	Never	Fixed	100	5000	0	0	0	0
pets_smallfish	Continuous	11	44	None	Never	Fixed	100	5000	0	0	0	0
pets_freshfish	Continuous	32	0	None	Never	Fixed	100	5000	0	0	0	0
homeown	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0
homeown	Nominal	---	---	Never	Fixed	---	100	5000	0	0	0	0

One more tip...

- You can also look at fields by a target field you are interested in. For example, we want to build a model to predict churn. We can use the Overlay option and chose churn.
- Select Use custom fields and use Overlay to select churn.
- On the Audit tab we can see 'agecat' by "churn."



SPSS Tech Tips

For more **Tech Tips**, go to the Tech Tips section within the Learning Hub.

Contact us



SPSSAdmin@version1.com



+ 44 (0) 203 859 4790
+ 353 (0) 1 865 7800



www.spssanalyticspartner.com



Thank you.