



Quickly Audit Data

ech Tips – IBM SPSS Modeler

Classification: Controlled. Copyright ©2024 Version 1. All rights reserved.

Classification: Controlled

- 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.







- 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.







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.







• 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.





- 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.



🖨 Eile 📄 Ed	Annotations	a 🕈 🚽										0
Complete fields (9	6): 88.19% Cor	mplete records (%):	11.56%									
Field	Measurement	Outliers	Extremes	Action Impute Missin	g Method	% Complete /	Valid Records	Null Value	Empty String	White Space	Blank Value	
Inwmon	& Continuous	5	ONone	Never	Fixed	26.88	1344	3656	0	0	0	
Inwten	& Continuous	10	0 None	Never	Fixed	26.88	1344	3656	0	0	0	
Feaulpmon	& Continuous	5	0 None	Never	Fixed	34.08	1704	3296	0	0	0	
leguipten	& Continuous	15	0 None	Never	Fixed	34.08	1704	3296	0	0	0	
Intmon	& Continuous	3	0 None	Never	Fixed	47.56	2378	2622	0		0	
Itoliten	& Continuous	35	ONone	Never	Fixed	47.56	2378	2622	0	0	0	
lcardten	& Continuous	58	0 None	Never	Fixed	71.56	3578	1422	0	0	0	
lcardmon	& Continuous	7	0 None	Never	Fixed	71.62	3581	1419	0	0	0	
longten	& Continuous	79	28 None	Never	Fixed	99.94	4997	3	0	0	0	
Inlongten	& Continuous	27	0 None	Never	Fixed	99.94	4997	3	0	0	0	
townsize	I Ordinal	-		Never	Fixed	99.96	4998	2	0	0	0	
commutatime	S Continuous	25	ONone	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	
Inothdebt	Continuous	24	ONone	Never	Fixed	99.98	4999	1	0	0	0	
region	A Nominal	-		Never	Fixed	100	5000	0	0	0	0	
age	Continuous	0	0 None	Never	Fixed	100	5000	0	0	0	0	
agecat	d Ordinal			Never	Fixed	100	5000	0	0	0	0	
birthmonth	A Nominal	-		Never	Fixed	100	5000	0	0	0	0	
ed	Continuous	0	0 None	Never	Fixed	100	5000	0	0	0	0	
edcat	d Ordinal			Never	Fixed	100	5000	0	0	0	0	
jobcat	A Nominal	-		Never	Fixed	100	5000	0	0	0	0	
union	A Nominal			Never	Fixed	100	5000	0	0	0	0	
employ	d Ordinal			Never	Fixed	100	5000	0	0	0	0	
empcat	I 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	I 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	
othdebt	& Continuous	65	28 None	Never	Fixed	100	5000	0	0	0	0	
default	Nominal			Never	Fixed	100	5000	0	0	0	0	
jobsat	I Ordinal	-		Never	Fixed	100	5000	0	0	0	0	
marital	Nominal	-		Never	Fixed	100	5000	0	0	0	0	
spoused	S Continuous	0	0 None	Never	Fixed	100	5000	0	0	0	0	
spousedcat	II Ordinal			Never	Fixed	100	5000	0	0	0	0	
A secolds	Al Continuous		di bita na s	Manufact	Without all	100	E 0.0.0	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.











Thank You

For more information

please visit spssanalyticspartner.com

Classification: Controlled. Copyright ©2024 Version 1. All rights reserved.

Classification: Controllec