

Exploring data one variable at time

- a) Nie wszystkie dane się zgadzają. Różnice są widoczne w nazwach oraz ilościami kolumn.
- b) Nie zauważono znaczących różnic.

Exploring data two variables at a time - tabulate

a)

	BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ
1	Bad Risk	1100	25860	39025	Homelmp	Other	10,5	0	
2	Bad Risk	1300	70053	68400	Homelmp	Other	7	0	
3	Bad Risk	1500	13500	16700	Homelmp	Other	4	0	
4	Bad Risk	1500		*			*	*	
5	Good Risk	1700	97800	112000	Homelmp	Office	3	0	
6	Bad Risk	1700	30548	40320	Homelmp	Other	9	0	
7	Bad Risk	1800	48649	57037	Homelmp	Other	5	3	
8	Bad Risk	1800	28502	43034	Homelmp	Other	11	0	
9	Bad Risk	2000	32700	46740	Homelmp	Other	3	0	
10	Bad Risk	2000		62250	Homelmp	Sales	16	0	
11	Bad Risk	2000	22608	*			18	*	
12	Bad Risk	2000	20627	29800	Homelmp	Office	11	0	
13	Bad Risk	2000	45000	55000	Homelmp	Other	3	0	
14	Good Risk	2000	64536	87400		Mgr	2,5	0	
15	Bad Risk	2100	71000	83850	Homelmp	Other	8	0	
16	Bad Risk	2200	24280	34687	Homelmp	Other	*	0	
17	Bad Risk	2200	90957	102600	Homelmp	Mgr	7	2	
18	Bad Risk	2200	23030	*			19	*	

b)

Columns	N	N Missing	N Categories
JOB	5681	279	6

Columns	N	N Missing	Min	Max	Mean	Std Dev
YOJ	5445	515	0	41	8,922268	7,573982

Equity - Tabulate - JMP Pro

Tabulate

To add to the table, drag and drop columns or statistics into the column header or row label area of the table.

14 Columns

- BAD
- LOAN
- MORTDUE
- VALUE
- REASON
- JOB
- Y0J
- DEROG
- DELINQ
- CLAGE
- NINQ
- CLNO
- DEBTINC
- Validation

Freq: Weight: Page Column:

Include missing for grouping columns
 Order by count of grouping columns
 Add Aggregate Statistics

Default Statistics

Format: Use the same decimal format

You can change the numeric format for displaying specific statistics. Each format consists of two integers: the field width and the number of decimal places. For the 'Rest Format' use the keyword 'Rest' in place of

REASON	BAD			
	Good Risk		Bad Risk	
	Row %	N	Row %	N
DebtCon	81,03%	3183	18,97%	745
Homelmp	77,75%	1384	22,25%	396
JOB				
Mgr	76,66%	588	23,34%	179
Office	86,81%	823	13,19%	125
Other	76,80%	1834	23,20%	554
ProfExe	83,39%	1064	16,61%	212
Sales	65,14%	71	34,86%	38
Self	69,95%	135	30,05%	58

Equity - Tabulate - JMP Pro

Tabulate

To add to the table, drag and drop columns or statistics into the column header or row label area of the table.

14 Columns

- BAD
- LOAN
- MORTDUE
- VALUE**
- REASON
- JOB
- Y0J
- DEROG
- DELINQ
- CLAGE
- NINQ
- CLNO
- DEBTINC
- Validation

Freq: Weight: Page Column:

Include missing for grouping columns
 Order by count of grouping columns
 Add Aggregate Statistics

Default Statistics

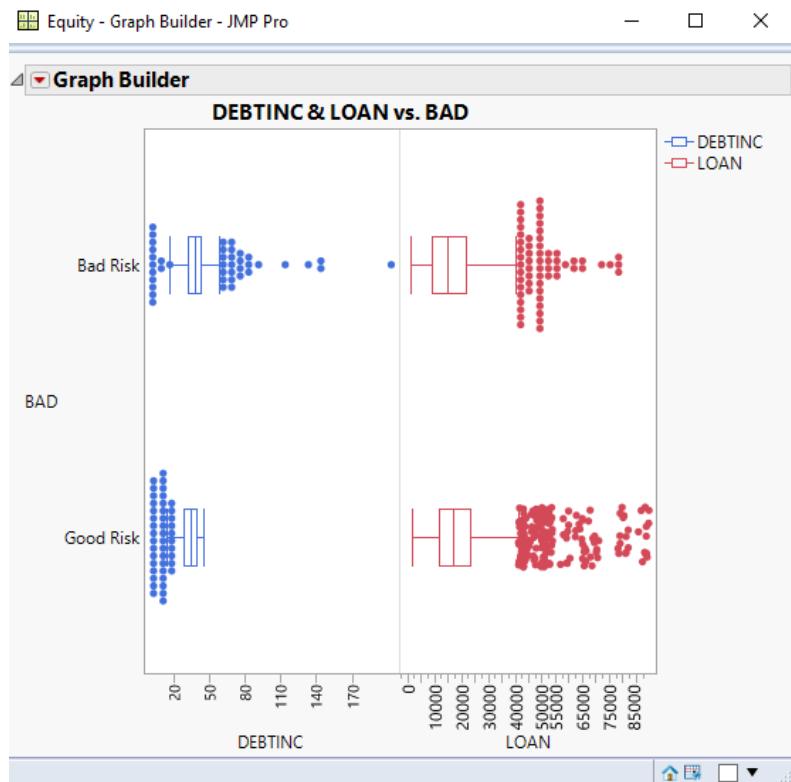
Format: Use the same decimal format

You can change the numeric format for displaying specific statistics. Each format consists of two integers: the field width and the number of decimal places. For the 'Rest Format' use the keyword 'Rest' in place of.

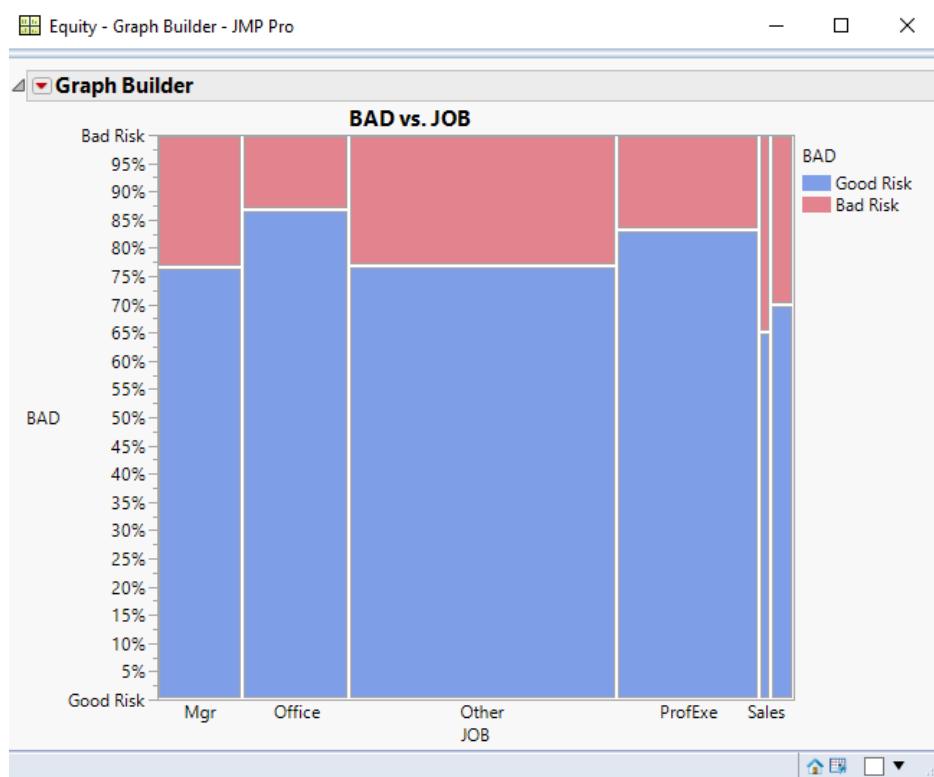
	BAD					
	Good Risk		Bad Risk			
	LOAN	MORTDUE	VALUE	LOAN	MORTDUE	VALUE
Mean	19028,107315	74829,25	102595,9	16922,119428	69460,45	98172,85

Exploring data two variables at a time – graph

a)

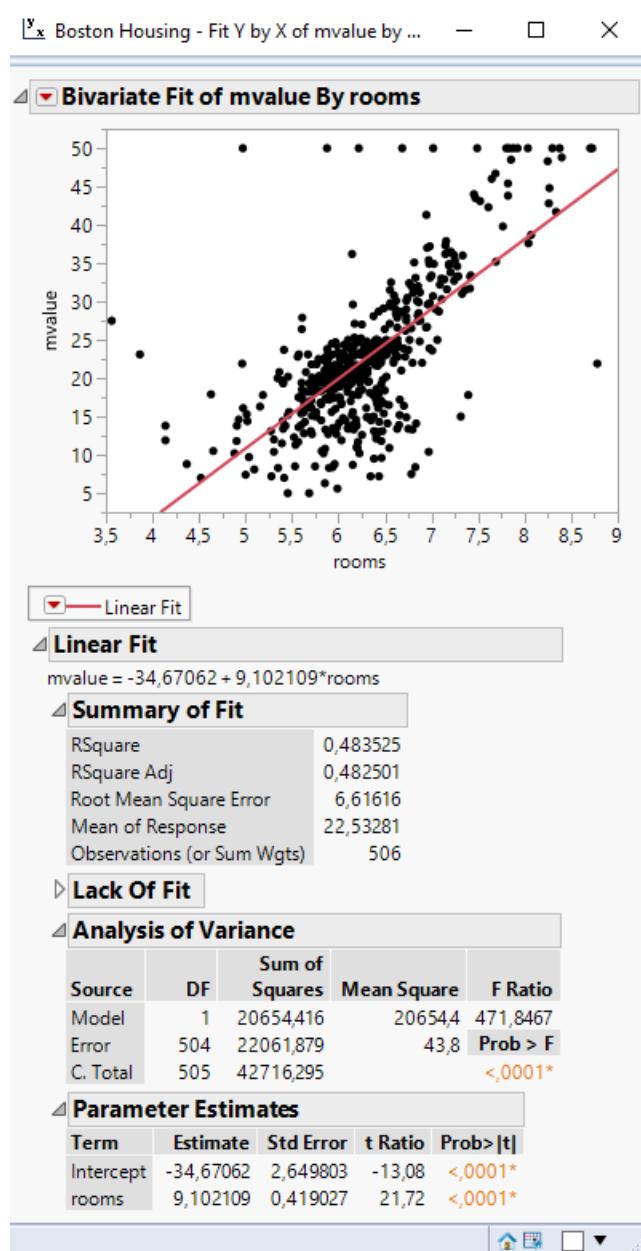


b)

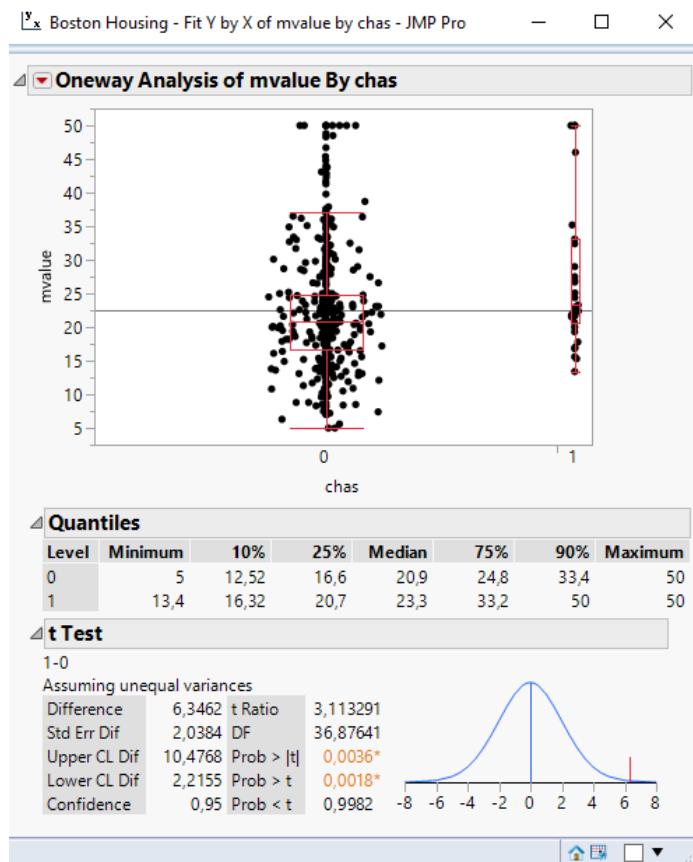


Exercise 2

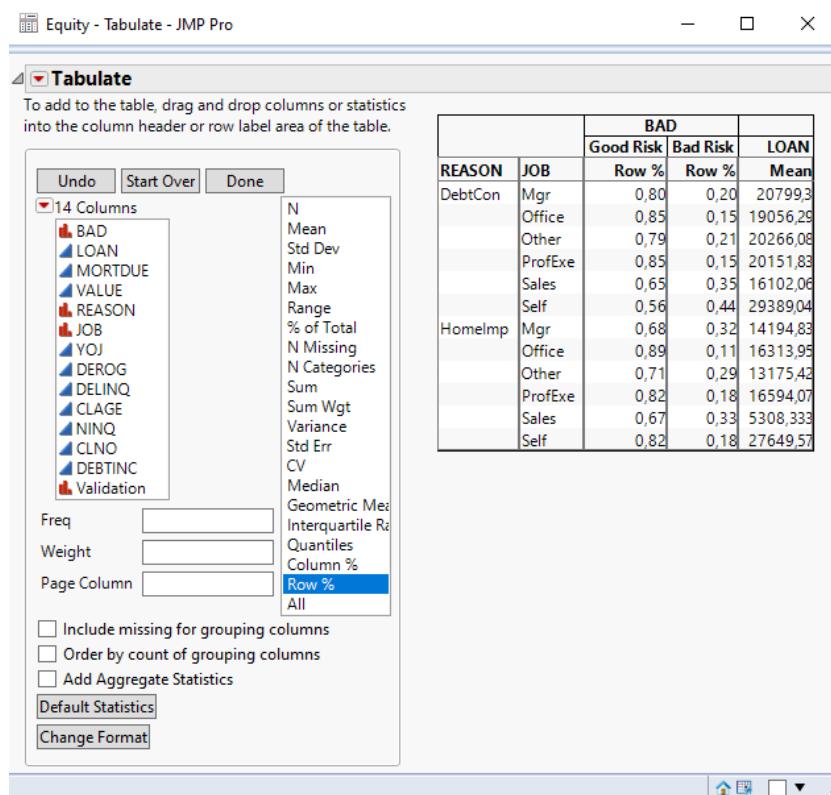
a)

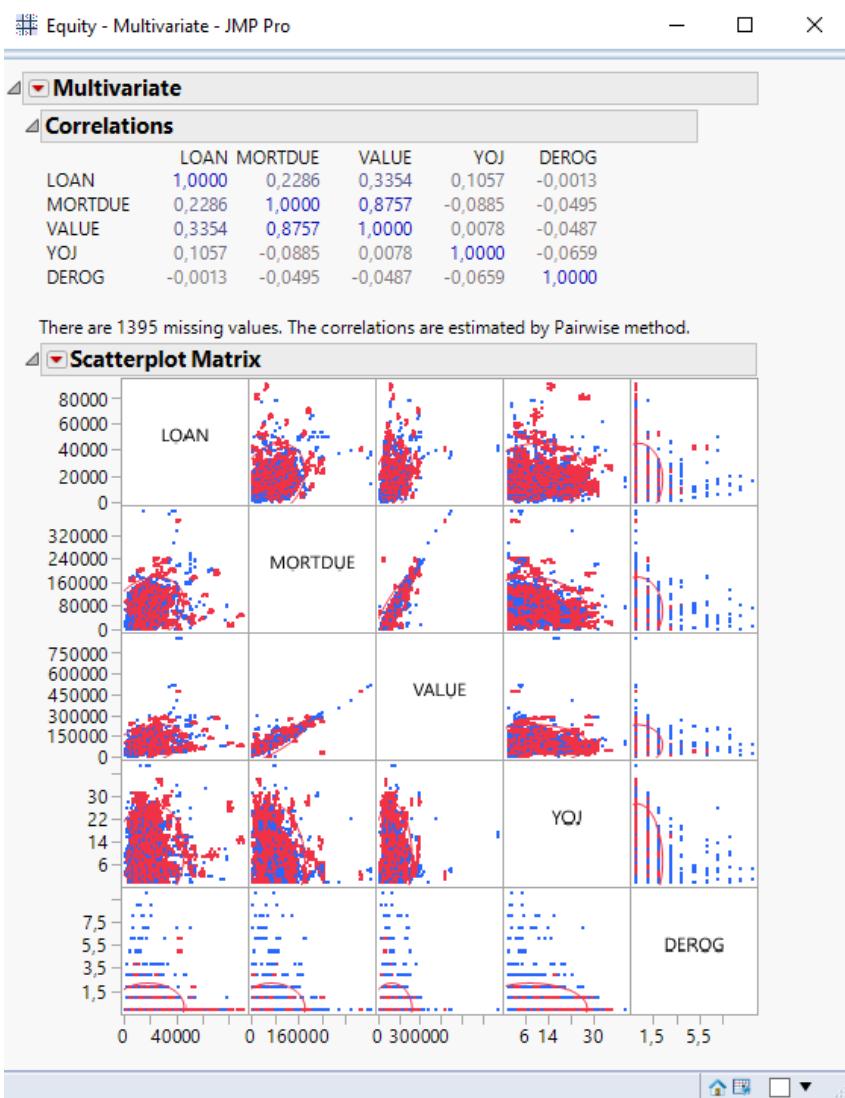


b)

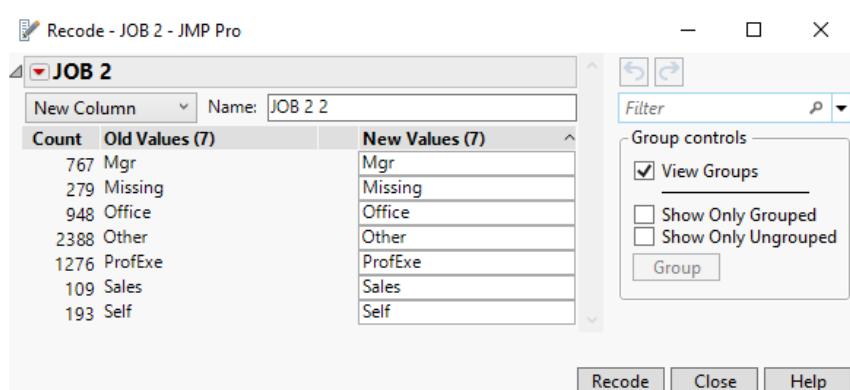


Exercise 3 – Exploring data three or more variables at a time



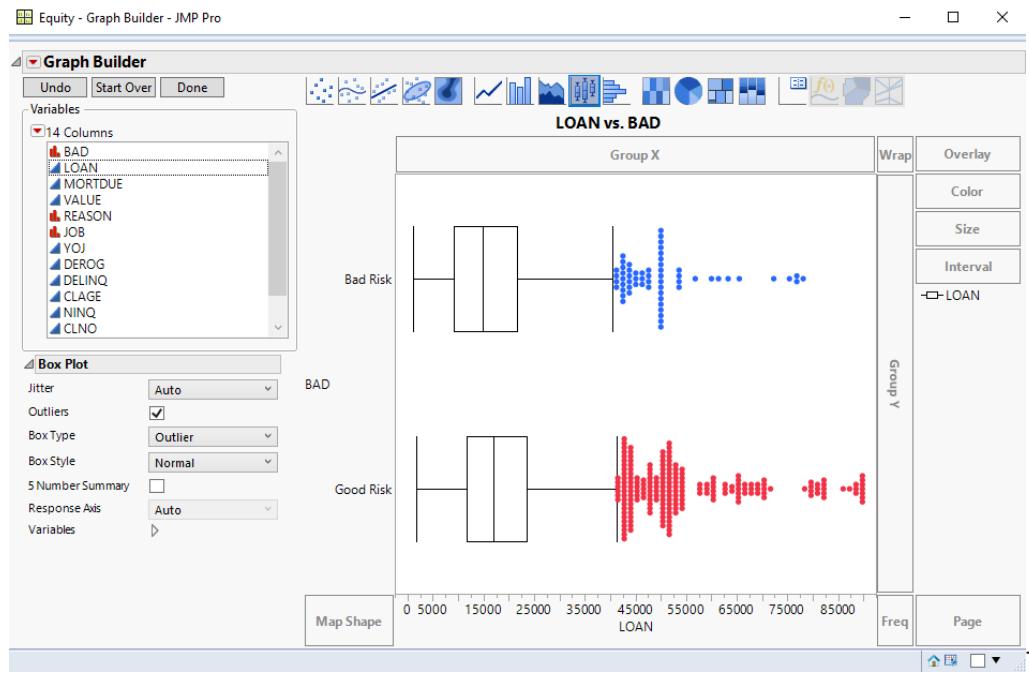


Exercise 4 – Preparing data for modeling



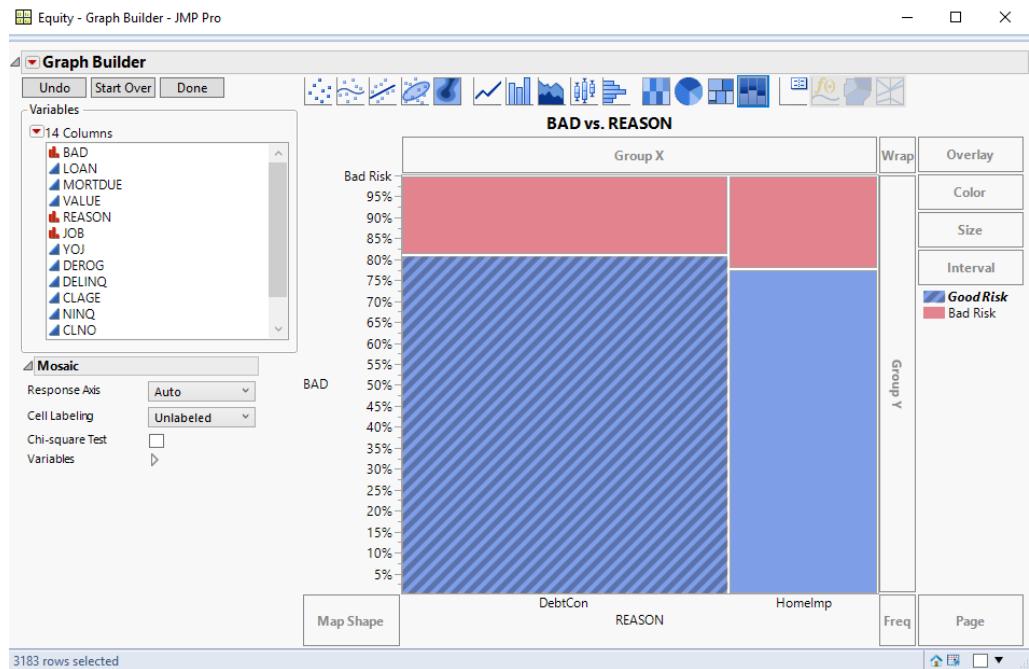
Exercise 5

a)



Osobny podział dla dwóch grup - brak powiązania.

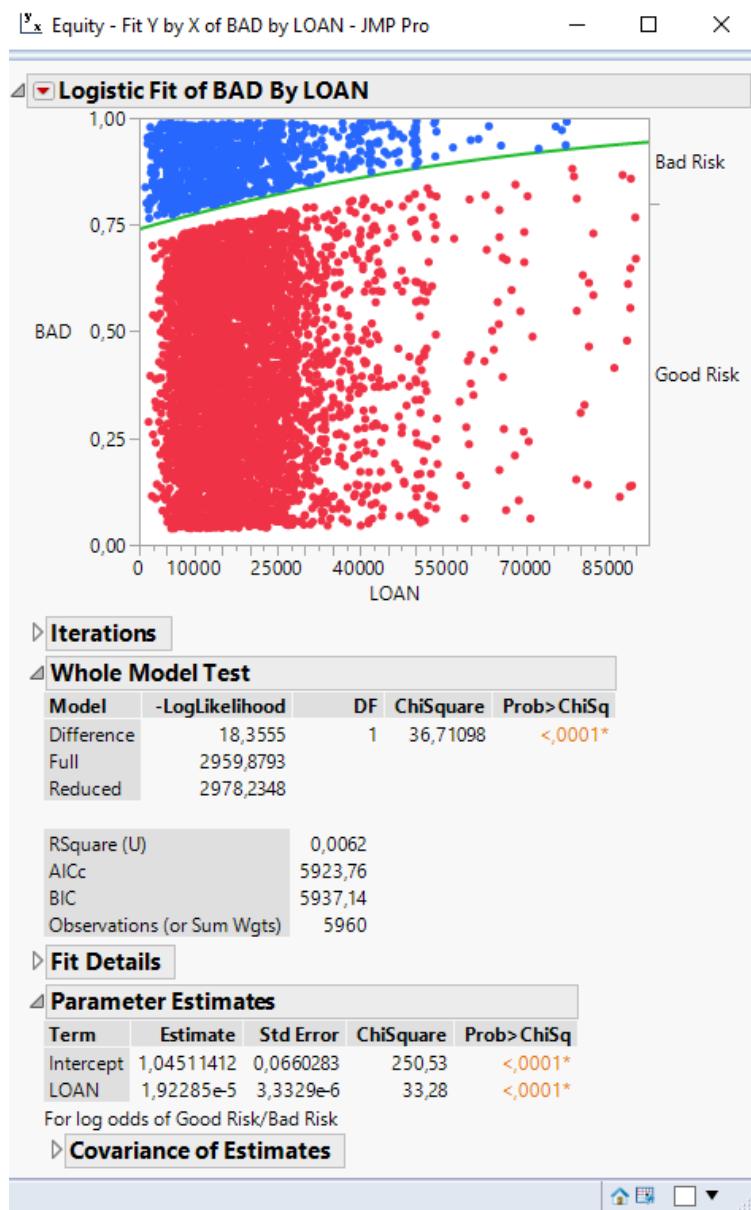
b)



81% of debtcon to good risk, 19% debtcon of bad risk. 77.8% home improvement good risk and 22.2% bad risk.

Powiązanie pomiędzy „badem” i „reasonem” jest takie, że w zależności od „reasona” wyszczególnione są osoby które są w grupie „good risk” i „bad risk”.

c) 1)

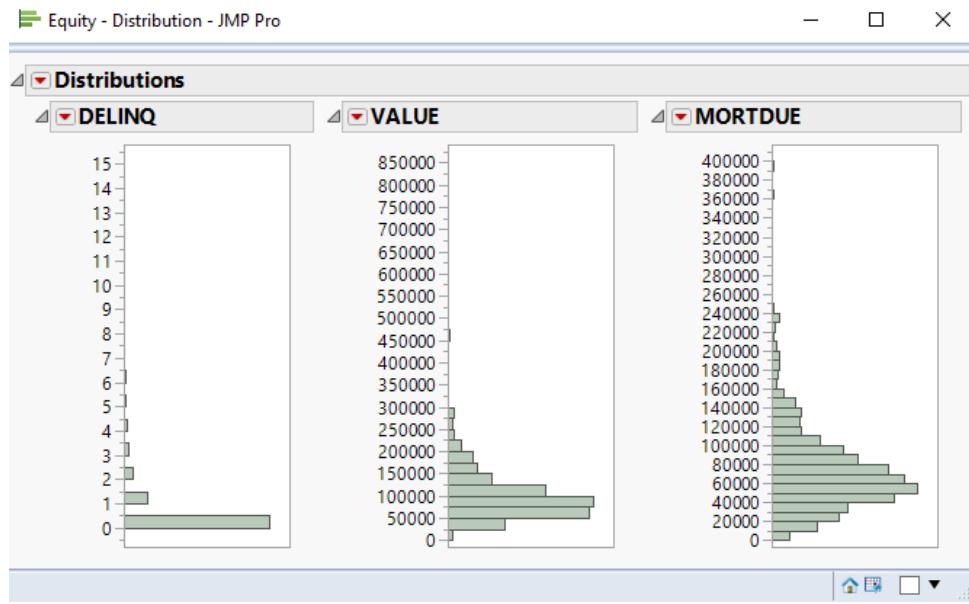


Im większa kwota pożyczki, tym mniej chętnych na tę kwotę.

c) 2)

Powiązanie pomiędzy „badem” i „reasonem” jest takie, że w zależności od „reasona” wyszczególnione są osoby które są w grupie „good risk” i „bad risk”.

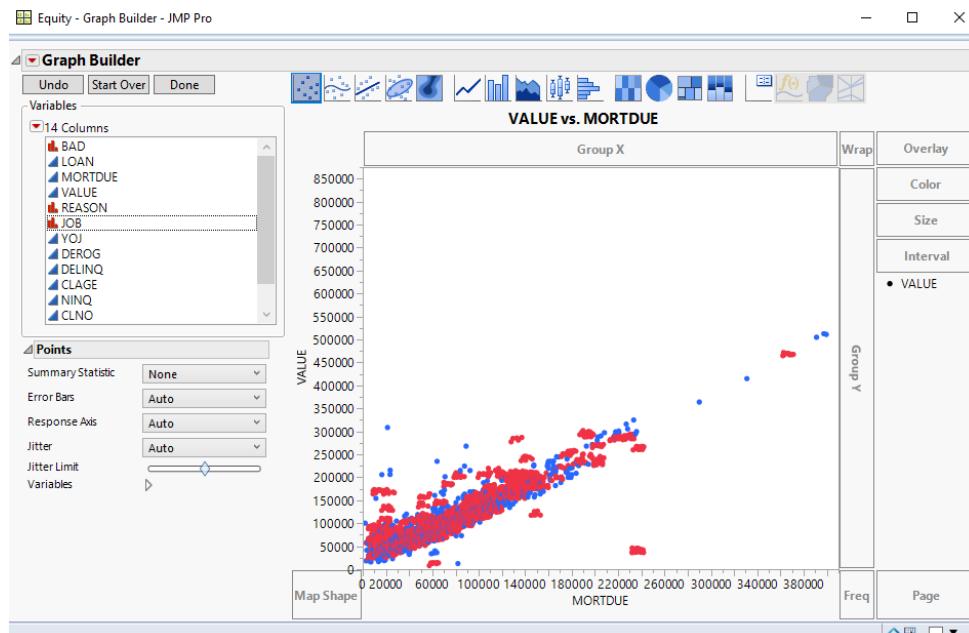
d)



Delinq ukazuje wykres słupkowy, natomiast value i mortdue to histogramy, ponieważ słupki są połączone – jest to asymetria lewostronna.

e) do dokończenia

f) 1)



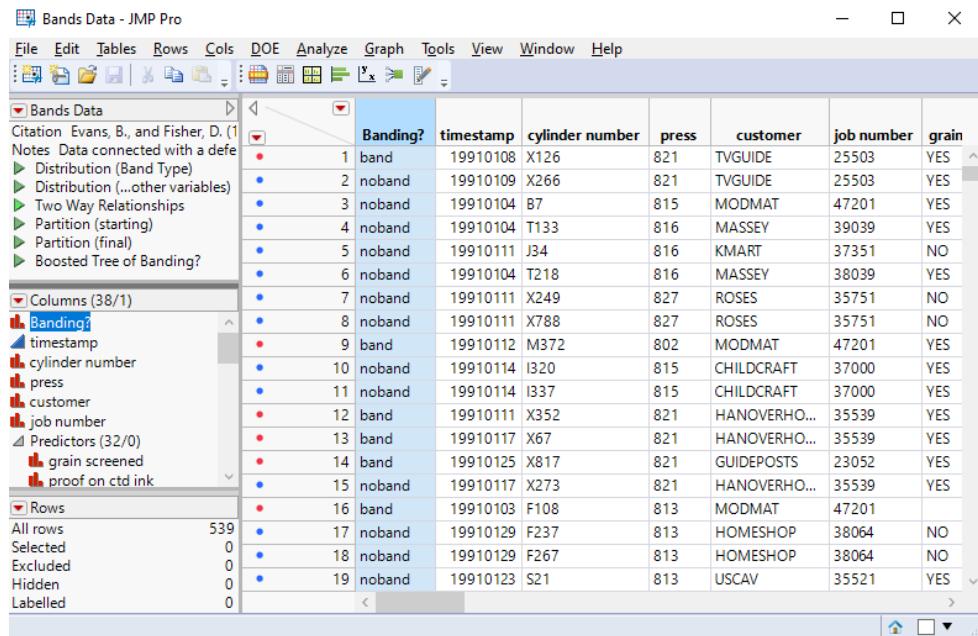
Wykres punktowy przedstawia dane w najdokładniejszy sposób.

f) 2)

Ma sens, ponieważ dane są wtedy najbardziej czytelne na wykresie.

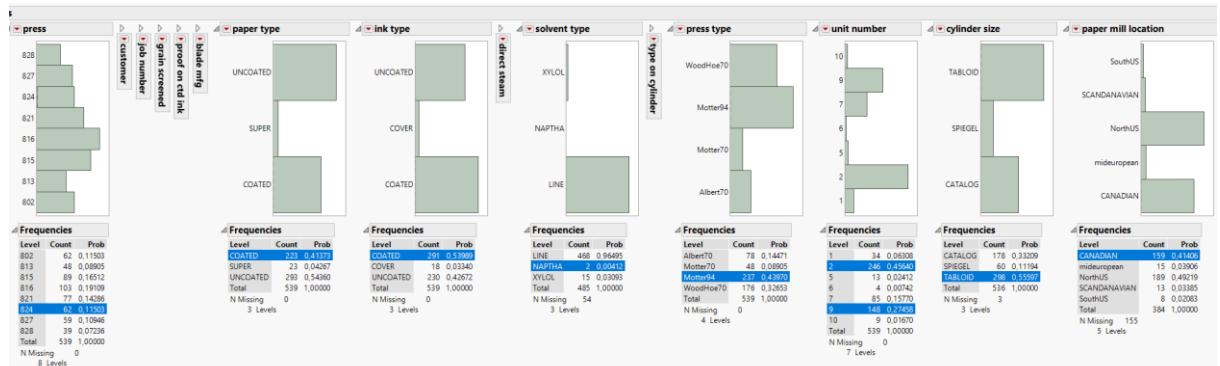
Excercise 6

a)



37 zmiennych kategorycznych oraz 20 ciągłych.

b) 1)



Miedzy 2-10 kategorii - press, paper type, ink type, press type, solvent type, unit number, cylinder size, paper mill location.

b) 2)

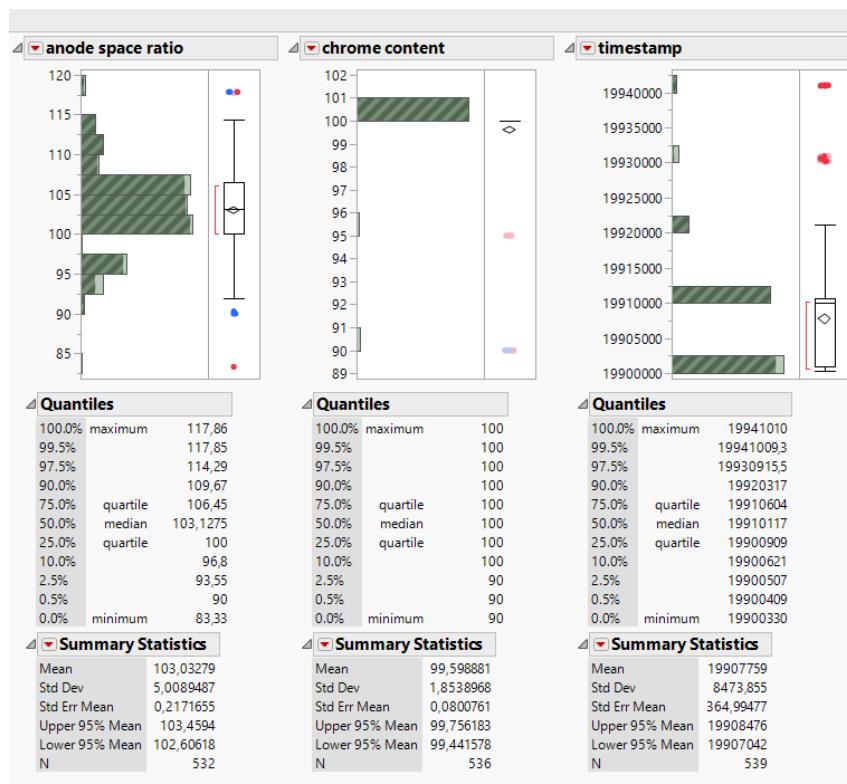


Proof on ctd ink, blade mfg, direct steam, solvent type, type on cylinder

c) 1)

Esa Voltage – asymetria lewostronna, Wax – asymetria prawostronna, Amperage – wykres słupkowy

c) 2)



Dla chrome content – 100, liczba elementów wynosi 510

Dla 95 – liczba elementów wynosi 9

Dla 90 – liczba elementów wynosi 17

Można wybrać pojedyncze elementy z kategorii co powoduje ukazanie z którymi elementami z (innych kategorii) jest powiązanie.

d)

W miejscu gdzie jest ciągła, jest to relacja 1 do wielu, a tam gdzie jest kategoryczny jest 1:1

e)

job number	grain screened	proof on ctd ink	blade mfg	paper type	ink type
522	47403			COATED	COATED
523	85741			UNCOATED	UNCOATED
524	85750			SUPER	UNCOATED
525	47405			COATED	COATED
526	37191			COATED	COATED
527	35069			COATED	COATED
528	37191			COATED	COATED
529	35425			SUPER	UNCOATED
530	71331			SUPER	UNCOATED
531	85813			SUPER	COATED
532	38240			UNCOATED	UNCOATED
533	38240			UNCOATED	UNCOATED
534	85813			SUPER	UNCOATED
535	85813			SUPER	UNCOATED
536	38064			SUPER	COATED
537	85814			SUPER	COATED
538	85814			SUPER	UNCOATED
539	38064			SUPER	UNCOATED

Nie są informacyjne ponieważ dla niektórych „bandingów” nie pokazują żadnych informacji

f)

Banding?	timestamp	cylinder number	press	customer	job number	grain
band	19910108	X126	821	TVGUIDE	25503	YES
noband	19910109	X266	821	TVGUIDE	25503	YES
noband	19910104	B7	815	MODMAT	47201	YES
noband	19910104	T133	816	MASSEY	39039	YES
noband	19910111	J34	816	KMART	37351	NO
noband	19910104	T218	816	MASSEY	38039	YES
noband	19910111	X249	827	ROSES	35751	NO
noband	19910111	X788	827	ROSES	35751	NO
band	19910112	M372	802	MODMAT	47201	YES
noband	19910114	I320	815	CHILDCRAFT	37000	YES
noband	19910114	I337	815	CHILDCRAFT	37000	YES
band	19910111	X352	821	HANOVERHO...	35539	YES
band	19910117	X67	821	HANOVERHO...	35539	YES
band	19910125	X817	821	GUIDEPOSTS	23052	YES
noband	19910117	X273	821	HANOVERHO...	35539	YES
band	19910103	F108	813	MODMAT	47201	
noband	19910129	F237	813	HOMESHOP	38064	NO
noband	19910129	F267	813	HOMESHOP	38064	NO
noband	19910123	S21	813	USCAV	35521	YES

Są powiązane

g)

The screenshot shows a JMP Data Table titled "Bands Data - JMP Pro". The table has columns: blade mfg, paper type, ink type, direct steam, solvent type, and type on cylinder. The data shows various combinations of these variables, with some entries having empty or null values.

	blade mfg	paper type	ink type	direct steam	solvent type	type on cylinder
•	505	COATED	COATED	NO		YES
•	506	SUPER	UNCOATED	NO		YES
•	507	SUPER	UNCOATED	NO		YES
•	508	SUPER	UNCOATED	NO		YES
•	509	SUPER	UNCOATED	NO		YES
•	510	SUPER	UNCOATED	NO		YES
•	511	SUPER	UNCOATED	NO		YES
•	512	SUPER	UNCOATED	NO		YES
•	513	SUPER	COVER	NO		YES
•	514	SUPER	COVER	NO		YES
•	515	COATED	COATED	NO		YES
•	516	SUPER	UNCOATED			YES
•	517	COATED	COATED			YES
•	518	COATED	COATED			YES
•	519	SUPER	COVER			YES
•	520	SUPER	UNCOATED			YES
•	521	UNCOATED	UNCOATED			YES
•	522	COATED	COATED			YES
•	523	UNCOATED	UNCOATED			

Problem istnieje – występują brakujące wartości które są widoczne w data table. Data table ukazuje, że dla niektórych kategorii wartości mają puste pola. Problematyczne pola są puste przez co nie możemy odczytać niektórych wartości. Powinna być jakakolwiek informacja, czy puste pole jest uzasadnione, np. wartość null, czy pole rzeczywiście jest puste i nie ma żadnej wartości.

h)

The screenshot shows a JMP Data Table titled "Bands Data - JMP Pro". The table has columns: press, blade pressure, varnish pct, press speed, ink pct, solvent pct, ESA Voltage, and Et. The data shows various process parameters with some entries having empty or null values.

	press	blade pressure	varnish pct	press speed	ink pct	solvent pct	ESA Voltage	Et
•	396	0,75	30	0	1600	62,5	27,5	5
•	397	1	35	0	1650	55,5	44,5	0
•	398	625	20	3,6	1800	60,2	36,2	0
•	399	625	33	•	1200	•	•	0
•	400	125	40	0	1500	58,8	41,2	3
•	401	0,5	32	0	1750	62,5	37,5	0
•	402	625	20	10,5	1550	52,6	36,9	0
•	403	625	20	0	1550	62,5	37,5	0
•	404	625	46	0	1350	60,2	39,8	0
•	405	625	48	0	1350	63,9	36,1	0
•	406	875	20	18,1	1800	47,6	34,3	0
•	407	0,75	33	0	1500	56,8	43,2	0
•	408	1	28	0	1800	61,7	38,3	0
•	409	625	22	0	1300	60	40	1
•	410	0,75	26	0	1450	58,8	41,2	1
•	411	•	29	0	1000	58,8	41,2	0
•	412	625	30	0	1950	58,8	41,2	0
•	413	0,75	50	0	1700	60,2	39,8	0
•	414	0,75	50	0	1700	63,9	36,1	0

Kropki powinny przedstawiać jakąkolwiek wartość, aby łatwiej było je zidentyfikować.

