

Erstellung eines Klassen-Histogramms zur Überprüfung der Häufigkeitsverteilung

Diese Histogramme sollten Sie nach der Aufbereitung der Rohdaten, d.h. nach Neuordnung und Zuordnung der Lagemerkmale, der Stadtbezirke und der Ortsteile, nach Ausfiltern von Doubletten, und nach Ausfiltern der Datensätze, die aus anderen Teilmärkten stammen, als erste Untersuchung auf die Güte und Eignung der Daten der Mietpreissammlung erstellen.

Hierbei geht es um den Teilmarkt „marktübliche Mietwerte für Wohnungen in einem konkreten Jahr“.

- Zuerst werden nun offensichtliche Doubletten ausgeschieden
- Dazu gehören keine Neubauten und Erstvermietungen im Baujahr oder dem Folgejahr des Baujahrs.
- Dazu gehören keine Wohnungen des öffentlich geförderten Wohnungsbaus (erkennbar an dem Merkmal WBS)
- Dazu gehören keine möblierten oder teilmöblierten Wohnungen
- Dazu gehören keine Apartments. Apartments sind Wohnungen mit einer Wohnfläche bis 30 m²
- Dazu gehören keinen Wohnungen, die keinem Stadtbezirk zugeordnet werden können, denn diese sind wahrscheinlich im Rahmen der Vermarktung bereits zusätzlich mit Adresse oder Stadtbezirksangabe erfasst.
- Schließlich werden noch die Datensätze ausgeschieden, die keinem Stadtbezirk zugeordnet werden können.

Im Beispiel bei Wohnungen, die in Krefeld im Jahr 2013 angeboten wurden, waren in der Rohdatenbank bei insgesamt 6.543 Datensätzen 266 ohne Zuordnung enthalten. Nach der Ausfilterung von Doubletten sind noch 117 Datensätze ohne Zuordnung verblieben. Damit wurden immerhin 56 % der nicht zugeordneten Fälle im Rahmen der Doublettenprüfung in den vollständigen Datensätzen wiedergefunden.

Die beispielhafte Statistik der Auswertung ergibt sich wie folgt:

	Anzahl	Prozentual
Ursprünglich vorhandene Rohdatensätze	6.543	100%
Nach Ausscheiden von		
Neubauten	133	2,0%
öffentlich gefördertem Wohnungsbau	59	0,9%
Möbliert und Teilmöbliert	18	0,3%
Apartments	171	2,6%
Doubletten	1.159	17,7%
Datensätze ohne Zuordnung nach Doublettenprüfung	117	1,8%
Unplausiblen Datensätze (MW \pm 4,5 Sigma)	13	0,2%
Unvollständige gelöschte Datensätze	60	0,9%
Insgesamt ausgefilterte Datensätze	1.730	26,4%
verbleiben valide Datensätze	4.813	73,6%

Sämtliche nicht zum zu untersuchenden Teilmarkt gehörenden Datensätze werden ausgefiltert, aber nicht gelöscht, sondern in jeweils eigene Tabellen verschoben. Denn je nach Fragestellung könnte man sie wieder brauchen.

Danach erzeugt man eine **statistische Auswertung** über die validierten Datensätze laufen. Als Ergebnis liefert diese eine Statistik mit folgenden Aussagen:

© Rolf Schubert & Kollegen, 50769 Köln, www.ImmoWert-RS.com											± 1,5 Interq.	± 2 Sigma	
WHG-Mieten Krefeld-Gesamt Jahr 2013	BJ	REN	MBJ	WFL	EnKW	Angeb.€	Faktor	NKM €/Mo	NKM €/m²	NK €/Mo	WARM €/m²	Interquartil	Konfident
Anzahl	2.623	1.540	4.190	4.761	452	4.761	4.761	4.761	4.761	4.644	4.761	4.455	4.278
Minimum	1588	1965	1960	31 m²	22	105,00 €	0,994	104,37 €	2,01 €/m²	18,00 €	2,30 €/m²	4,14 €/m²	4,32 €/m²
25% Quartil	1938	2008	1993	53 m²	114	300,00 €	0,994	298,20 €	5,31 €/m²	100,00 €	7,20 €/m²	5,31 €/m²	5,31 €/m²
Mittelwert	1956	2009	1996	69 m²	145	414,71 €	0,994	412,22 €	5,99 €/m²	145,30 €	8,09 €/m²	5,89 €/m²	5,84 €/m²
Median	1960	2011	2000	66 m²	140	380,00 €	0,994	377,72 €	5,85 €/m²	140,00 €	8,02 €/m²	5,80 €/m²	5,78 €/m²
75% Quartil	1976	2013	2010	80 m²	173	490,00 €	0,994	487,06 €	6,55 €/m²	180,00 €	8,87 €/m²	6,46 €/m²	6,39 €/m²
Maximum	2010	2014	2013	180 m²	282	1.176,00 €	0,994	1.168,94 €	10,82 €/m²	420,00 €	13,82 €/m²	7,83 €/m²	7,46 €/m²
Streuungsmaße													
Median / Arithm.Mittel	0,2%	0,1%	0,2%	-4,4%	-3,3%	-8,4%	0,0%	-8,4%	-2,2%	-3,7%	-0,9%	-1,5%	-1,1%
Interquartilabstand	38	5	17	27 m²	59	190,00 €	0,000	188,86 €	1,23 €/m²	80,00 €	1,67 €/m²	1,15 €/m²	1,08 €/m²
Standardabweichung(n)	32	5	16	23 m²	43	156,44 €	0,000	155,50 €	1,00 €/m²	60,14 €	1,32 €/m²	0,79 €/m²	0,72 €/m²
Variationskoeffizient	1,6%	0,3%	0,8%	32,7%	29,8%	37,7%	0,0%	37,7%	16,6%	41,4%	16,3%	13,4%	12,3%
Zusammenhangsmaße													
	DATUM	MBJ	WFL	Schwellenwerte unten				4,14 €/m² Interquartil				4,31 €/m² Stabw.	
Bestimmtheitsmaß (Spalte : rel.NKM)	0,0026	0,0342	0,0012	Schwellenwerte oben				7,83 €/m² Interquartil				7,46 €/m² Stabw.	
Korrelationskoeffizient (SP : rel.NKM)	0,0505	0,1848	-0,0352										
Kovarianz(s) (SP zu rel.NKM)	5,1422	3,0512	-0,7892										
Ausreißerprüfungen bei 5 Standardabweichungen													
unterer Grenzwert				-44 m²	-71	-367,50 €		-365,30 €	1,01 €/m²	-155,38 €	1,48 €/m²		
oberer Grenzwert				182 m²	361	1.196,92 €		1.189,74 €	10,96 €/m²	445,99 €	14,70 €/m²		

Dabei werden für die angegebenen Merkmale von Baujahr bis Warmmiete in €/m² jeweils die Anzahl und die Lagemaße vom Minimum bis zum Maximum sowie die Streuungsmaße und die Zusammenhangsmaße für die Beziehungen zwischen Datum und relativer NKM, Mietbaujahr und relativer NKM sowie Wohnfläche und relativer NKM berechnet. Gleichartige Auswertungen kann man auch für jeden einzelnen Stadtbezirk erstellen. Unter dieser Auswertung werden für eine erste Ausreißerprüfung die Schwellenwerte bei ± 5 Standardabweichungen berechnet. Die davon betroffenen Datensätze werden in die Tabelle UNPL verschoben.

Die Grenzen aus dem Interquartilabstand und der 2-fachen Standardabweichung werden für 2 hintereinandergeschaltete Ausreißerprüfungen genutzt und in den beiden rechten Spalten wird die relative Nettokaltmiete je m² nach den beiden Ausreißerprüfungen ausgeworfen.

Aus den Ergebnissen der rechten Spalte werden dann gleichzeitig die (vorläufigen) **Eckwerte** der bereinigten statistischen Auswertung in einer zusammengefasst tabelliert, damit man übersichtlich stabile Erkenntnisse über den mittleren Bereich des Marktes – ohne die ggf. auch validen Extremwerte erhält.

Eckwerte	WHG-Mieten Krefeld nach STBZ						Abw.
© Rolf Schubert & Kollegen, Köln	2013						vom Ø
STBZ	Anzahl	25 % Q.	Median	75 % Q.	Mittelwert		in %
Stadtmitte	1.121	5,25 €/m²	5,68 €/m²	6,11 €/m²	5,73 €/m²		-1,88%
Kempener Feld/Baackeshof	303	5,37 €/m²	5,89 €/m²	6,36 €/m²	5,88 €/m²		0,62%
Inrath/Kliedbruch	295	5,45 €/m²	5,88 €/m²	6,45 €/m²	5,91 €/m²		1,18%
Cracau	755	5,27 €/m²	5,78 €/m²	6,45 €/m²	5,84 €/m²		0,03%
Dießem/Lehmheide	331	5,17 €/m²	5,53 €/m²	6,09 €/m²	5,67 €/m²		-2,95%
Benrad-Süd	53	4,97 €/m²	5,36 €/m²	5,80 €/m²	5,43 €/m²		-7,06%
Forstwald	21	5,96 €/m²	6,50 €/m²	6,63 €/m²	6,30 €/m²		7,86%
Benrad-Nord	83	4,75 €/m²	5,16 €/m²	5,82 €/m²	5,32 €/m²		-8,92%
Hülser Berg	3	5,78 €/m²	6,36 €/m²	6,41 €/m²	6,01 €/m²		2,88%
Traar	28	6,05 €/m²	6,50 €/m²	6,97 €/m²	6,45 €/m²		10,42%
Verberg	22	6,46 €/m²	6,90 €/m²	7,13 €/m²	6,72 €/m²		14,96%
Gartenstadt	34	5,20 €/m²	5,43 €/m²	5,93 €/m²	5,63 €/m²		-3,61%
Bockum	290	5,74 €/m²	6,33 €/m²	6,86 €/m²	6,26 €/m²		7,17%
Linn	78	5,30 €/m²	5,49 €/m²	5,99 €/m²	5,68 €/m²		-2,68%
Gellep-Stratum	9	6,39 €/m²	6,43 €/m²	6,82 €/m²	6,56 €/m²		12,35%
Oppum	164	5,26 €/m²	5,47 €/m²	6,06 €/m²	5,66 €/m²		-3,06%
Fischeln	263	5,44 €/m²	5,86 €/m²	6,43 €/m²	5,90 €/m²		1,00%
Uerdingen	266	5,47 €/m²	5,94 €/m²	6,41 €/m²	5,96 €/m²		1,97%
Hüls	159	5,50 €/m²	6,12 €/m²	6,61 €/m²	6,10 €/m²		4,38%
Gesamt	4.278	5,31 €/m²	5,78 €/m²	6,39 €/m²	5,84 €/m²	100,00%	

In der rechten Spalte wird ein vorläufiges Ranking der einzelnen Stadtteile ausgeworfen.

Gleichzeitig werden die statistischen Kontrollwerte stadtbezirksbezogen (ggf. auch ortsteilbezogen) berechnet:

Statist.Kontrollwerte Krefeld STBZ	Anzahl	2013		Kennzahlen nach 2 Außerprüfungen	
		Δ Median Arithm \bar{O}	STABW(n) 1 Sigma	Variationskoeffizient STABW/Arithm.M	
Stadtmitte	1.121	-0,9%	0,67 €/m ²		11,61%
Kempener Feld/Baackeshof	303	0,3%	0,69 €/m ²		11,69%
Inrath/Kliedbruch	295	-0,5%	0,71 €/m ²		12,07%
Cracau	755	-1,0%	0,75 €/m ²		12,79%
Dießem/Lehmheide	331	-2,4%	0,69 €/m ²		12,21%
Benrad-Süd	53	-1,2%	0,65 €/m ²		12,06%
Forstwald	21	3,2%	0,59 €/m ²		9,35%
Benrad-Nord	83	-2,9%	0,65 €/m ²		12,15%
Hülser Berg	3	5,9%	0,57 €/m ²		9,47%
Traar	28	0,7%	0,71 €/m ²		11,06%
Verberg	22	2,8%	0,61 €/m ²		9,06%
Gartenstadt	34	-3,5%	0,59 €/m ²		10,41%
Bockum	290	1,1%	0,69 €/m ²		10,95%
Linn	78	-3,5%	0,68 €/m ²		12,04%
Gellep-Stratum	9	-2,0%	0,43 €/m ²		6,58%
Oppum	164	-3,5%	0,70 €/m ²		12,35%
Fischeln	263	-0,7%	0,68 €/m ²		11,57%
Uerdingen	266	-0,2%	0,67 €/m ²		11,30%
Hüls	159	0,4%	0,73 €/m ²		11,96%
Σ Krefeld	4.278	-1,1%	0,72 €/m ²		12,33%
zul. Toleranzen		< \pm 5,0 %			< 30%
Die Grundgesamtheit ist <normal-verteilt>.					

Hier wird die Grob-Prüfung auf „Normal Verteilung“ mittels der Division von Median und MW gezeigt. Wir sehen, dass die 3 Fälle aus Hülser Berg diese Prüfung $\pm < 5 \%$ nicht bestehen. Das ist aber bei 3 Fällen auch nicht zu erwarten. Es wird eine Standard-Abweichung berechnet und der Variationskoeffizient, ein relatives Streuungsmaß. Hier erkennt man, dass trotz der fehlenden „Normal Verteilung“ der Variationskoeffizient für Hülser Berg mit 9,47 % sehr gut ist. Es kommt bei der explorativen Statistik nicht auf die üblichen aus der induktiven Statistik bekannten Testverfahren an.

Nun müssen diese Daten weiteren Tests unterzogen werden:

Dazu gehört als erstes der klassifizierte Häufigkeitstest. Damit testet man die absolute Nettokaltmiete je Monat. Da es hierbei im Endeffekt darum geht, tatsächliche Ausreißer und unplausible Daten zu entdecken, insbesondere solche, die nicht in der Masse der großen Zahl untergehen und deshalb einen Einfluss auf die Aussagen der Grundgesamtheit haben, darf man an dieser Stelle mit größeren Spannen arbeiten. Wilfried Mann hat in seiner Grundlagenforschung zum integrativen Auswertemodell hier die Grenzen bei $\pm 4,5$ bis 5 Standardabweichungen für immobilienwirtschaftliche Statistiken festgestellt.

Da der absolute Mietwert pro Monat ja auch stark von der Wohnfläche abhängig ist, hat diese Auswertung das Ziel tatsächliche Unplausibilitäten aufzudecken. Zur Vorbereitung bedient man sich deshalb dabei zunächst der relativen Mietwerte in €/m² und berechnet die Standardabweichung aus allen validen Datensätzen, dann die 5-fache Standardabweichung und bestimmt den unteren und den oberen Grenzwert.

Der untere Grenzwert liegt ausgehend vom arithmetischen Mittel von 5,99 €/m² - 5 × eine Standardabweichung von 1,00 €/m² = 1,01 €/m². Der obere Schwellenwert liegt bei 5,99 €/m² + 5 × 1,00 €/m² = 10,96 €/m². Siehe Screenshot der statistischen Auswertung auf Seite 2.

Wir setzen dann den Cursor in der letzten Datenzeile in die Spalte „V“, die den NKM €/m² enthält und drücken F8 (Erweiterung der Auswahl) und setzen den Cursor in Zeile 1 der Datenbank. Nachdem alle Zeilen in dieser Spalte markiert sind, wählt man im Menüpunkt <Daten> den Filter aus.

R	S	T	U	V	W
EnKW	ANG €/M	Faktor	NKM €/Mo	NKM €/r	NK €/Mo
	440,00 €	0,994	437,36 €	5,33 €/m ²	137,00 €
	350,00 €	0,994	347,90 €	6,44 €/m ²	100,00 €
	430,00 €	0,994	427,42 €	5,62 €/m ²	140,00 €
	400,00 €	0,994	397,60 €	6,12 €/m ²	150,00 €
	320,00 €	0,994	318,08 €	6,36 €/m ²	140,00 €

Durch Klick auf das kleine Dreieck in der Zelle „V1“ öffnet sich der Zahlenfilter.

Suchen	NK €/Mo
1.186	1.186
<input type="checkbox"/> (Alle auswählen)	50,00 €
<input checked="" type="checkbox"/> 2,01 €/m ²	93,25 €
<input checked="" type="checkbox"/> 2,30 €/m ²	139,28 €
<input checked="" type="checkbox"/> 2,48 €/m ²	120,00 €
<input checked="" type="checkbox"/> 2,49 €/m ²	160,00 €
<input checked="" type="checkbox"/> 2,61 €/m ²	291,13 €
<input checked="" type="checkbox"/> 3,28 €/m ²	
<input checked="" type="checkbox"/> 3,40 €/m ²	
<input checked="" type="checkbox"/> 3,43 €/m ²	66,75 €
<input checked="" type="checkbox"/> 3,51 €/m ²	39,16 €
<input checked="" type="checkbox"/> 3,60 €/m ²	239,41 €
<input checked="" type="checkbox"/> 3,61 €/m ²	
<input checked="" type="checkbox"/> 3,64 €/m ²	249,07 €
<input checked="" type="checkbox"/> 3,68 €/m ²	178,8%
<input checked="" type="checkbox"/> 3,69 €/m ²	-358,86 €
<input type="checkbox"/> 3,70 €/m ²	637,42 €

Suchen	NK €/Mo
1.186	1.186
<input checked="" type="checkbox"/> (Alle auswählen)	50,00 €
<input checked="" type="checkbox"/> 2,01 €/m ²	93,25 €
<input checked="" type="checkbox"/> 2,30 €/m ²	139,28 €
<input checked="" type="checkbox"/> 2,48 €/m ²	120,00 €
<input checked="" type="checkbox"/> 2,49 €/m ²	160,00 €
<input checked="" type="checkbox"/> 2,61 €/m ²	291,13 €
<input checked="" type="checkbox"/> 3,28 €/m ²	
<input checked="" type="checkbox"/> 3,40 €/m ²	
<input checked="" type="checkbox"/> 3,43 €/m ²	66,75 €
<input checked="" type="checkbox"/> 3,51 €/m ²	39,16 €
<input checked="" type="checkbox"/> 3,60 €/m ²	239,41 €
<input checked="" type="checkbox"/> 3,61 €/m ²	
<input checked="" type="checkbox"/> 3,64 €/m ²	249,07 €
<input checked="" type="checkbox"/> 3,68 €/m ²	178,8%
<input checked="" type="checkbox"/> 3,69 €/m ²	-358,86 €
<input type="checkbox"/> 3,70 €/m ²	637,42 €

Man schaltet nun die Auswahl <Alles auswählen> aus und klickt die Werte einzeln an, die außerhalb der beiden berechneten Grenzwerte liegen. Nun werden nur noch die Datensätze angezeigt, die außerhalb der 5-Sigma-Grenze liegen. Dabei hat man die Möglichkeit sich auch die anderen Merkmale dieser Datensätze anzusehen und kann entscheiden, ob alle so identifizierten oder nur einige dieser Datensätze in die Tabelle „UNPL“ verschoben werden. Dadurch bleiben diese als unplausibel qualifizierten Datensätze erhalten, man kann sie später z.B. wenn es um die Bewertung von Mieten im oberen Segment geht erneut in die valide Datenbank zurück verschieben.

Hier sieht man z.B. einige, der als unplausibel erkannten Datensätze, weit weniger als 1 % der validen Datensätze:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	S	T	U	V	W	X						
1	Datum	LAG	STB	OT#	LKR	Stac	PLZ	OT	Str	Zimr	Obj	Typ	BJ	REN	MBJ	WFL	ANG	€/M	Faktor	NKM	€/Mo	NKM	€/m²	NK	€/Mo	Warm		
2	11.09.2013	1	02	21	KR	Kref	47803	Kemr	Prin.	1	WHG				2000	445	m²	247,00	€	0,994	245,52	€	0,55	€/m²	104,00	€	0,79	
3	11.03.2013	4	17	170	KR	Kref	47807	Fisch		3	WHG				2000	70	m²	65,00	€	0,994	64,61	€	0,92	€/m²	199,00	€	3,77	
4	19.08.2013	3	18	181	KR	Kref	47802	Uerdi	Krie	7	MAIS		1993		2011	250	m²	1.845,00	€	0,994	1.833,93	€	7,34	€/m²	500,00	€	9,34	
5	14.10.2013	3	18	180	KR	Kref	47829	Uerdi		8	WHG		1756	2001	2007	275	m²	2.395,00	€	0,994	2.380,63	€	8,66	€/m²	600,00	€	10,84	
6	04.03.2013	1	13	130	KR	Kref	47809	Bock		4	WHG		2012	2012	186	m²	1.950,00	€	0,994	1.938,30	€	10,42	€/m²	370,00	€	12,41		
7	20.05.2013	1	03	31	KR	Kref	47803	Inrath	Wilr	2	WHG				2013	2013	45	m²	500,00	€	0,994	497,00	€	11,04	€/m²	136,00	€	14,07
8	21.01.2013	1	03	31	KR	Kref	47803	Inrath	Wilr	2	WHG				2013	2013	50	m²	580,00	€	0,994	576,52	€	11,53	€/m²	130,00	€	14,13
9	06.05.2013	1	13	130	KR	Kref	47809	Bock		1	WHG		1925		2013	53	m²	650,00	€	0,994	646,10	€	12,19	€/m²		12,19		
10	18.03.2013	1	13	130	KR	Kref	47809	Bock		2	WHG		1887	2012	2012	87	m²	1.130,00	€	0,994	1.123,22	€	12,91	€/m²	170,00	€	14,86	
11	30.07.2013	1	13	130	KR	Kref	47809	Bock		7	WHG					220	m²	3.000,00	€	0,994	2.982,00	€	13,55	€/m²	488,00	€	15,77	
12	29.04.2013	1	13	130	KR	Kref	47809	Bock		5	WHG		1887	2011	2012	198	m²	2.970,00	€	0,994	2.952,18	€	14,91	€/m²	396,00	€	16,91	
13	11.11.2013	1	13	130	KR	Kref	47809	Bock		3	WHG		1887	2011	1960	131	m²	1.965,00	€	0,994	1.953,21	€	14,91	€/m²	390,00	€	17,89	
14	10.12.2013	3	18	183	KR	Kref	47829	Uerdi	Lösc	3	WHG		2001		2010	81	m²	4.475,00	€	0,994	4.448,15	€	54,92	€/m²	95,00	€	56,09	

Nun bereitet man die Daten vor, die man zur grafischen Darstellung der Häufigkeitsverteilung benötigt:

	A	B	C	D	E	F	G	H
1	Datum	BRW	LAG	STB	OT#	LKR	Stadt/Kre	PLZ
4867	Minimum		104,37 €/m²	Schiefe	1,437	rechtsschiefe	Verteilung	
4868	Maximum		1.168,94 €/m²	Wölbung	2,886	spitzer	gewölbt	
4869	Spannweite		1.064,57 €/m²	Varianz	36	σ²		
4870								
4871	Vorbereitung eines Histogramms zur Darstellung der Häufigkeitsverteilung							
4872					HÄUFIGKEIT(U2:U4762;D4874:D4885)			
4873			Klasse	Werte bis	Anzahl			
4874	1/10 d. Spannweite		1	110,00 €/m²	1	Führen Sie nun das Array		
4875			2	220,00 €/m²	193	in Zelle E4874 aus.		
4876			3	330,00 €/m²	1.426	Maximum Sie den		

Das sind Minimum und Maximum (in Excel MIN() bzw. MAX()). Diese berechnen wir nun aus der Spalte „U“, denn hier stehen die um die Verhandlungsspanne 0,994 bereinigten Abschlussmieten.

Die Rohdatenbank enthält nur Angebotsmieten. Diese werden fast zu 100 % nach Aussagen der Makler und Verwalter auch vertraglich vereinbart. Um den Mietpreis wird bei Mietvertragsabschlüssen nur sehr selten gehandelt. Gerade in stark nachgefragten Gebieten von Deutschland wird fast überhaupt nicht gehandelt, fast jede Mietforderung findet einen Mieter, das

ist der offizielle Grund für die Einführung der Mietpreisbremse. Der IVD hat im März 2015 in einer bundesweiten Umfrage bei einem Rücklauf von etwa 65 % die statistische Grundlage für den Faktor 0,994 geliefert. – Siehe IVD Minutenumfrage – eine intersubjektive Umfrage unter Fachleuten.

Aus Minimum und Maximum wird die Spannweite R (von engl. RANGE) ausgerechnet. Diese ist hier 1.065 €/Monat. Bei im Beispiel 4761 Datensätzen kann man mit der Excel-Formel =SCHIEFE(U2:U4762) die Schiefe der Verteilung messen. Diese Daten sind definitiv nicht normal verteilt, sondern, da die Schiefe größer 0 ist, deutlich rechtsschief. Das ist übrigens nach den Untersuchungen von Mann typisch für den deutschen Immobilienmarkt. Hier würde jeder halbwegs versierte Statistiker schnell sein Veto einlegen, sofern er um die Besonderheiten des Immobilienmarktes nicht Bescheid weiß und sich auf die induktive Statistik stützt. Aber Sachverständige der Immobilienbewertung wirken in einer anderen Welt. In der explorativen Statistik ist jede Miete und jeder Kaufpreis ein einmaliges, nicht wiederholbares Experiment.

Mit der Excel-Formel =KURT(U2:U4762) berechnet man die Wölbung und erhält den Wert 2,886. Da dieser positive Wert deutlich über 0 liegt ist die Verteilung der Daten spitzer gewölbt als bei einer Normalverteilung. Das bedeutet, die mittleren Werte liegen enger beieinander als bei einer normal verteilten Streuung.

Nun bildet man Klassen, um die fast 5.000 Datensätze übersichtlich und ohne „rauschen“ beurteilen zu können. Es hat sich bewährt ca. 6 bis 12 Klassen zu bilden. Da die Spannweite der Daten von ca. 1.065 €/Monat bekannt ist, bildet man gleich weit auseinander liegende Klassen mit einem Abstand von 110 €/m². D.h. man verwendet in diesem Fall ein Intervall von 110 €/Monat.

In die Spalte C, in diesem Beispiel ab Zeile 4874 schreibt man die Klassen von 1 bis 12.

In Spalte D ab Zeile 4874 die Werte bis von 110 € bis 1.320 €. In die Spalte E4873 schreibt man „Anzahl“ und man müsste nun in die nächste Zeile, in die Zelle E4874 die Anzahl der Datensätze schreiben, für die die Bedingung Nettokaltmiete < 110 € = WAHR zutrifft, schreiben. In die Zelle E4875 müsste man die Anzahl der Datensätze schreiben, für die die Bedingung Nettokaltmiete < 220 € = wahr zutrifft. Das geht natürlich einfacher.

	A	B	C	D	E	F	
1	Datum	BRW	LAG	STB	OT#	LKR	Sta
4872				HÄUFIGKEIT(U2:U4762;D4874:D4885)			
4873			Klasse	Werte bis	Anzahl		
4874	1/10 d. Spannweite		1	110,00 €/m ²	1	Führen Sie nun	
4875			2	220,00 €/m ²	193	in Zelle E4874 ;	
4876			3	330,00 €/m ²	1.426	Markieren Sie d	
4877			4	440,00 €/m ²	1.561	Zellbereich von	

Wir schreiben in die Zelle E4874 die Formel =HÄUFIGKEIT(U2:U4762;D4874:D4885). Es wird 1 angezeigt. Dies ist eine Matrix-Formel, die zur gültigen Eingabe eine besondere Methode benötigt. Zunächst setzen wir den Cursor in die Zelle E4874 und markieren den Bereich der Matrix unter der Zelle E4874 bis Zelle E4885 und drücken dann die Taste <F2>. Dadurch wird der Bereich der Matrix unter der Zelle E4874 hellblau eingefärbt und der Matrix-Bereich von D4874 bis E 4885 markiert. Siehe folgender Screenshot:

1	Datum	BRW	LAG	STB	OT#	LKR	Stadt/Kre PLZ
4872				HÄUFIGKEIT(U2:U4762;D4874:D4885)			
4873			Klasse	Werte bis	Anzahl		
4874	1/10 d. Spannweite		1	=HÄUFIGKEIT(U2:U4762;D4874:D4885) das Array			
4875			2	220,00 €/m ²	193	in Zelle E4874 aus.	
4876			3	330,00 €/m ²	1.426	Markieren Sie den	
4877			4	440,00 €/m ²	1.561	Zellbereich von Zelle	
4878			5	550,00 €/m ²	856	E4874 bis E4885,	
4879			6	660,00 €/m ²	396	drücken Sie die Taste <F2>	
4880			7	770,00 €/m ²	154	und anschließend	
4881			8	880,00 €/m ²	93	gleichzeitig die Tasten	
4882			9	990,00 €/m ²	45	<STRG> + <UMSCHAL> + <E	
4883			10	1.100,00 €/m ²	24	(= Pfeil nach oben)	
4884			11	1.210,00 €/m ²	12		
4885			12	1.320,00 €/m ²	0		
4886							

Nun drücken wir gleichzeitig die Tasten <STRG>, <UMSCHALTTASTE> und <ENTER>. (Mit der Umschalttaste wechseln Sie üblicherweise zwischen Groß- und Kleinschreibung.)

Sofort wird der Matrixbereich ausgerechnet:

	B	C	D	E	F	G	H
1	BRW	LAG	STB	OT#	LKR	Stadt/Kre	PLZ
4869		1.064,57 €/m ²	Varianz	36	σ^2		
4870							
4871	hms zur Darstellung der Häufigkeitsverteilung						
4872			HÄUFIGKEIT(U2:U4762;D4874:D4885)				
4873		Klasse	Werte bis	Anzahl			
4874		1	110,00 €/m ²	1	Führen Sie nun das Array		
4875		2	220,00 €/m ²	193	in Zelle E4874 aus.		
4876		3	330,00 €/m ²	1.426	Markieren Sie den		
4877		4	440,00 €/m ²	1.561	Zellbereich von Zelle		
4878		5	550,00 €/m ²	856	E4874 bis E4885,		
4879		6	660,00 €/m ²	396	drücken Sie die Taste <F2>		
4880		7	770,00 €/m ²	154	und anschließend		
4881		8	880,00 €/m ²	93	gleichzeitig die Tasten		
4882		9	990,00 €/m ²	45	<STRG> + <UMSCHAL> + <E		
4883		10	1.100,00 €/m ²	24	(= Pfeil nach oben)		
4884		11	1.210,00 €/m ²	12			
4885		12	1.320,00 €/m ²	0			
4886							

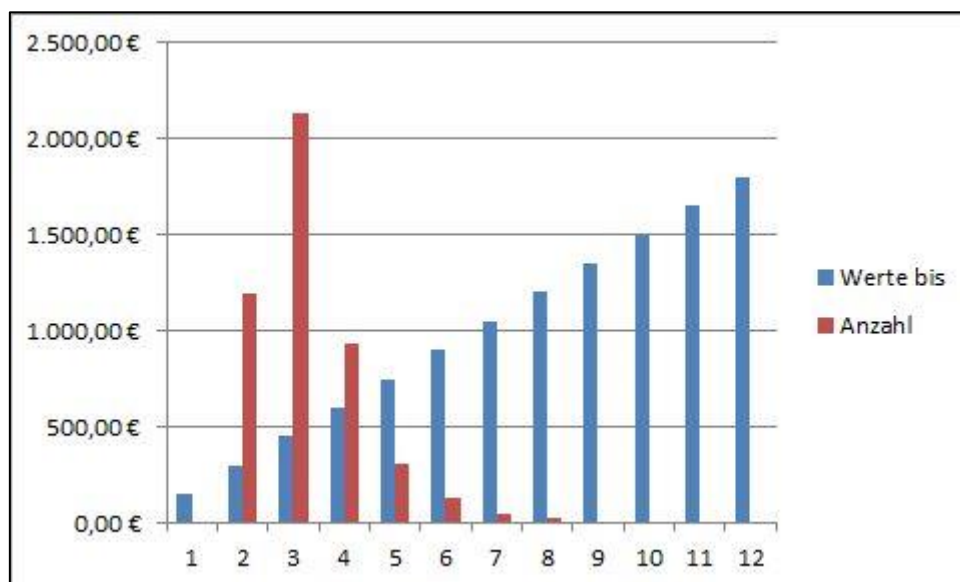
In der Formelzeile fx sieht man, dass die Matrixformel geschweifte Klammern bekommen hat. ({})

Nun sind für die einzelnen Klassen die Anzahlen ausgerechnet worden.

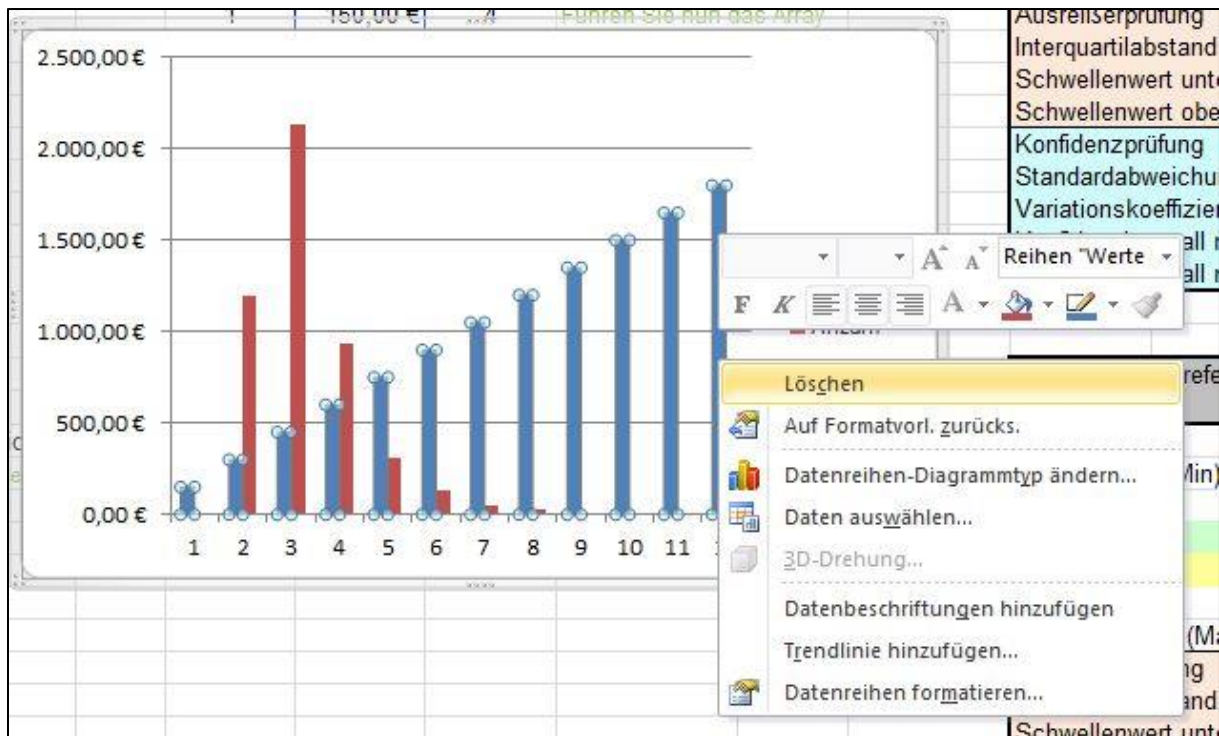
Man markiert nun die Matrix einschließlich der Überschrift von Zelle D4873 bis Zelle E4885 und drückt zunächst die Schaltfläche <kopieren> um dann im Menü <Einfügen> <Diagramme> das erste Säulendiagramm per Mausclick auszuwählen.

	B	C	D	E	F
1	BRW	LAG	STB	OT#	LKR
4870					
4871	ms zur Darstellung der Häufigkeitsverteilung				
4872			HÄUFIGKEIT(U2:U4762:D4874:D4886)		
4873		Klasse	Werte bis	Anzahl	
4874		1	110,00 €/m²	1	Führen Sie
4875		2	220,00 €/m²	193	in Zelle E4
4876		3	330,00 €/m²	1.426	Markieren S
4877		4	440,00 €/m²	1.561	Zellbereich
4878		5	550,00 €/m²	856	E4874 bis
4879		6	660,00 €/m²	396	drücken Si
4880		7	770,00 €/m²	154	und anschl
4881		8	880,00 €/m²	93	gleichzeitig
4882		9	990,00 €/m²	45	<STRG> +
4883		10	1.100,00 €/m²	24	(= Pfeil nac
4884		11	1.210,00 €/m²	12	
4885		12	1.320,00 €/m²	0	
4886					

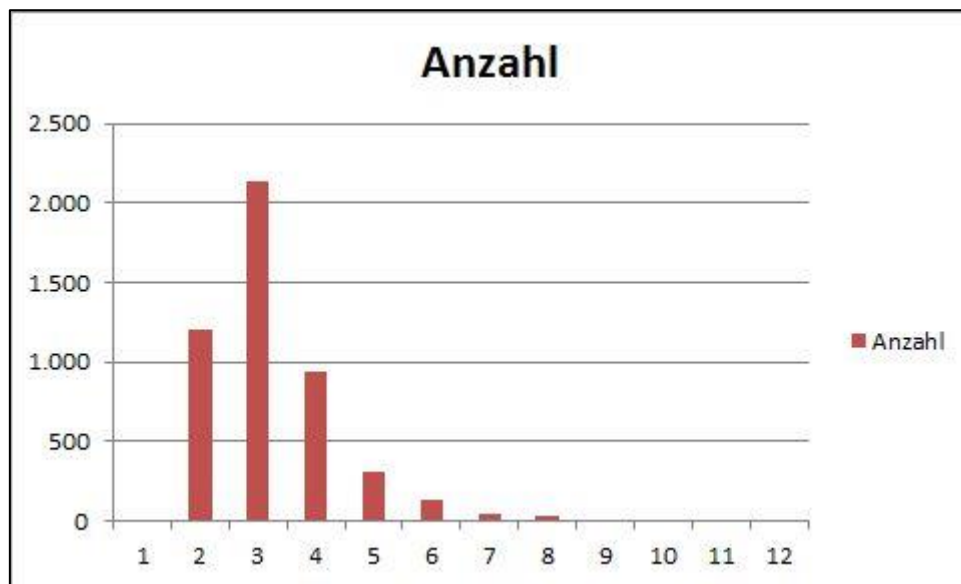
Das Säulendiagramm wird zunächst wie folgt erzeugt:



Man muss nun den rechten blauen Balken der Zahlenreihe „Werte bis“ mit der rechten Maustaste anklicken und die Option löschen anklicken:

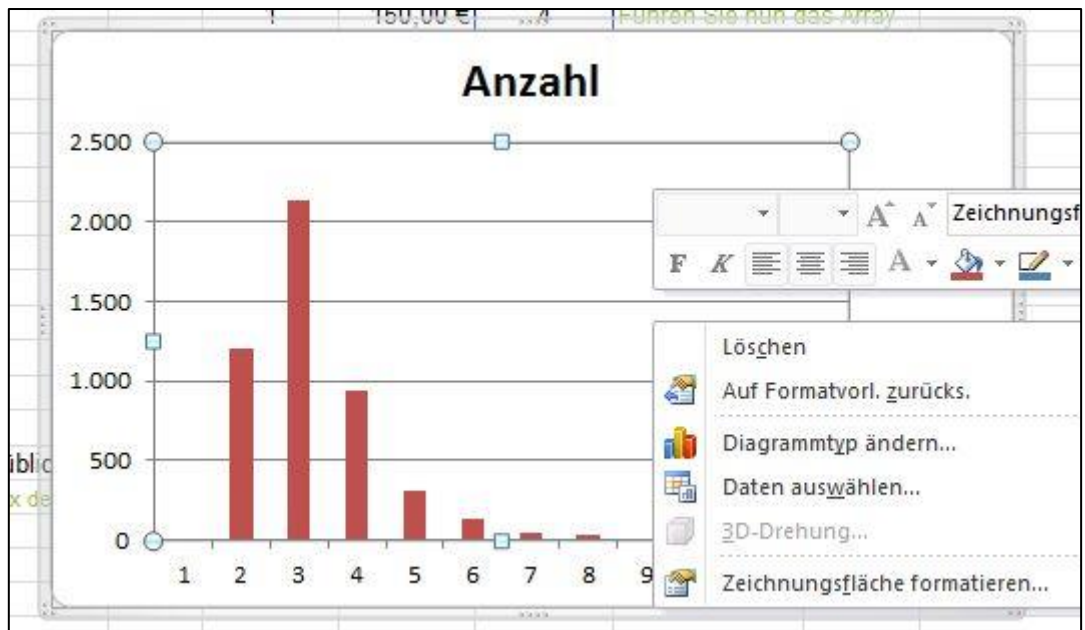


Danach ändert sich das Histogramm wie folgt:

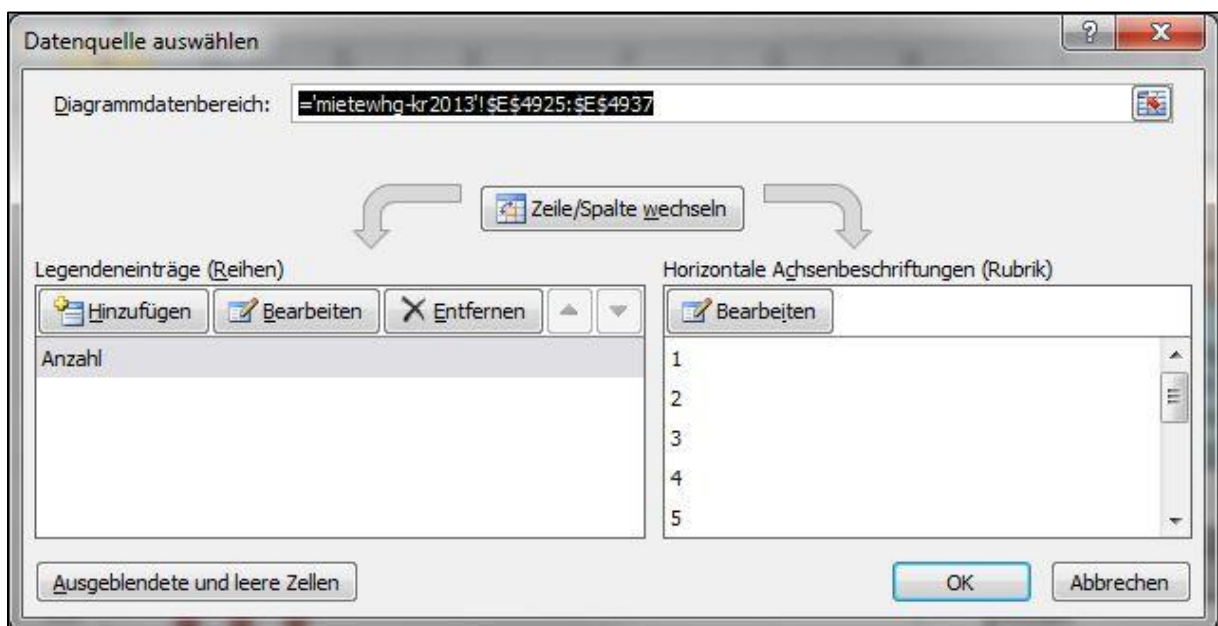


Nun klickt man ggf. zweimal mit der rechten Maustaste in den mittleren Bereich des Diagramms (in die Zeichenfläche)

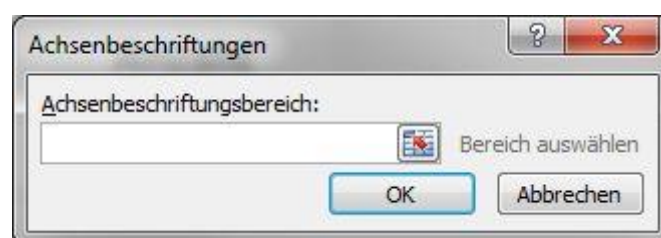
Ein Untermenü wird angezeigt, in dem man <Daten auswählen ...> anklicken kann.



Hier klickt man <Daten auswählen> an und mit dem Auswahldialog <Datenquelle auswählen> behilft. Es macht Sinn, wenn man das Säulendiagramm seitlich neben die Wertematrix schiebt, damit man diese später gut erreichen kann.

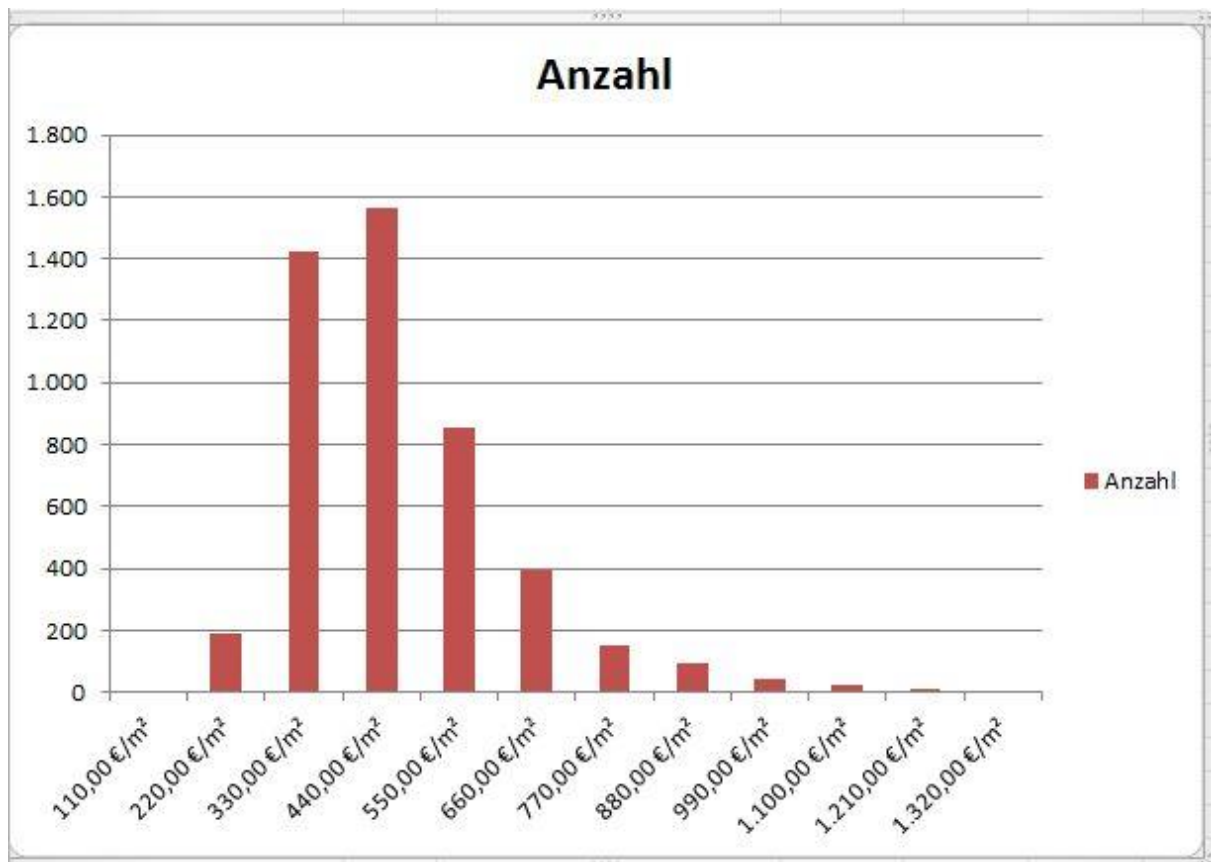
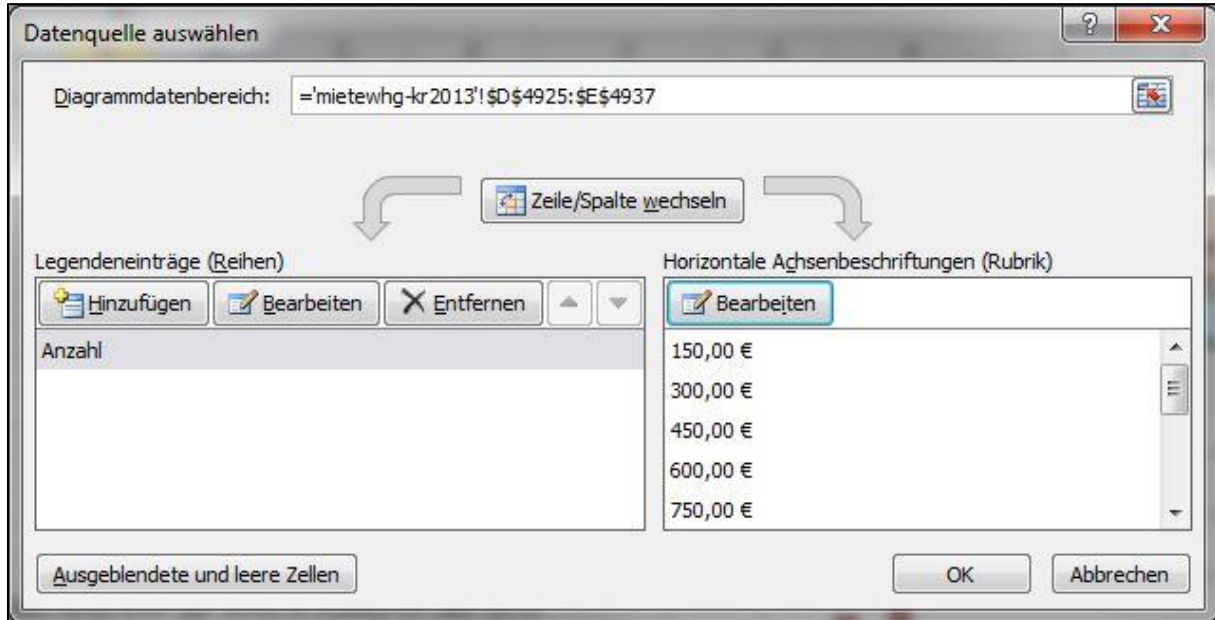


Hier klickt man auf die Schaltfläche <Bearbeiten> und es öffnet sich der Eingabe Dialog Achsenbeschriftungen.



Mit Klick in das Eingabefeld und der Markierung der Zellen D4873 bis D4885 in der Excel Tabelle werden diese in den Erfassungsdialog geschrieben.

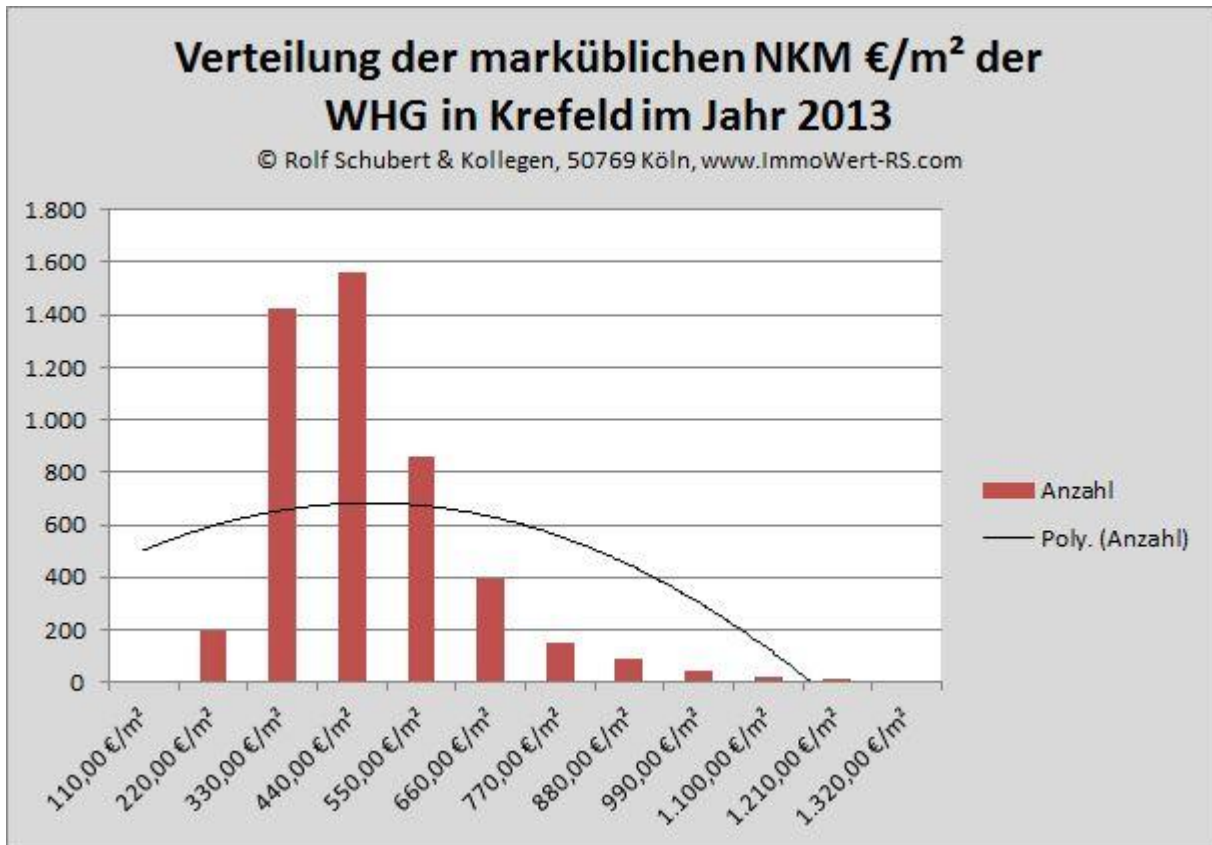
Drücken Sie <OK>. Nun stehen die Werte der Klassen im Dialog <Datenquelle auswählen> in der Spalte <Horizontale Achsenbeschriftungen (Rubrik)> und unter dem Säulendiagramm im Diagramm.



Man markiert nun die Überschrift <ANZAHL> und ersetzen Sie diese durch die im Beispiel in Zelle A4887 vorbereitete Diagramm-Überschrift. Sollte das nicht auf Anhieb klappen kann man nach Klick

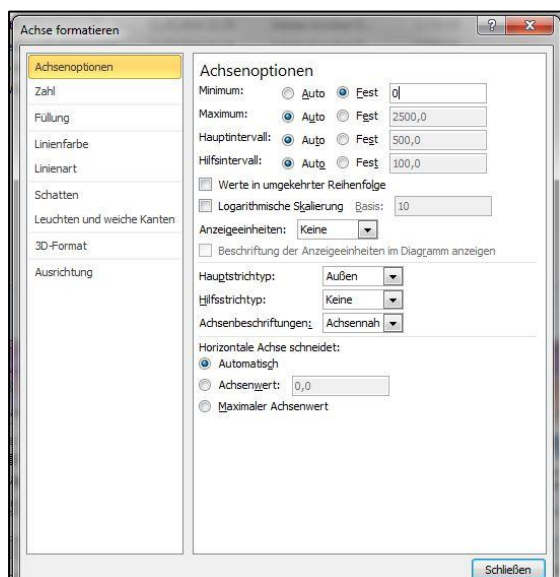
auf das Diagramm über den Menüpunkt <Diagrammtools> <Diagrammtitel> die Option <Über Diagramm> auswählen.

Man klickt nun noch in die Freifläche neben dem Titel an und fügt eine Hintergrundfarbe über das Start-Menü ein. Z.B. grau 25 %. Schließlich fügt man noch eine polynomische Trendlinie hinzu. Hier benötigt man eine polynomische Trendlinie 2. oder 3. Grades.



Ab dem 3. Grad kann sich die linke Skala bis in den negativen Skalenbereich erweitern.

Deshalb muss man dann nochmal auf die linke Skala klicken und im Dialog <Achse formatieren> und beim Minimum den Fest-Wert 0 eintragen.



Nun ist das Histogramm der klassierten Verteilung fertig.

Sollten sich bei einem so erstellten Histogramm Lücken innerhalb der mittleren Säulen ergeben, **dann wären solche Daten statistisch nicht zu gebrauchen**. Man erkennt das an einer nicht nach oben gewölbten polynomischen Trendkurve. Sie steht dann auf dem Kopf.