



31250 Introduction to Data Analytics

Assignment #2

Data Exploration and Preparation

Florian Lubitz

FEIT

13688799

florian.lubitz@student.uts.edu.au

Contents

1	Initial data exploration	1
1.1	Attribute features	1
1.2	Field_Info1	1
1.3	Field_Info2	2
1.4	Field_Info3	2
1.5	Field_Info4	3
1.6	Coverage_Info1	3
1.7	Coverage_Info2	3
1.8	Coverage_Info3	3
1.9	Sales_Info1	4
1.10	Sales_Info2	4
1.11	Sales_Info3	5
1.12	Sales_Info4	5
1.13	Sales_Info5	5
1.14	Personal_Info1	6
1.15	Personal_Info2	6
1.16	Personal_Info3	6
1.17	Personal_Info4	9
1.18	Personal_Info5	9
1.19	Property_Info1	9
1.20	Property_Info2	10
1.21	Property_Info3	10
1.22	Property_Info4	11
1.23	Property_Info5	11
1.24	Geographic_Info1	11
1.25	Geographic_Info2	11
1.26	Geographic_Info3	12
1.27	Geographic_Info4	12
1.28	Geographic_Info5	12
2	Data Preprocessing	13
3	Summary	13
4	Projekthalt	13
4.1	Aktivitäten	13
	List of Figures	15
	List of Tables	15

1 Initial data exploration

1.1 Attribute features

1.1.1 Quote_Id

Attribute type: nominal The "Quote_Id" is a customer number. It is neither orderable nor is the number a quantitative value.

Statistics: As the attribute is nominal you can not perform any mathematical calculations. The Id ranges between 638 to 434297.

1.1.2 Quote_Date

Attribute type: scale "Quote_Date" seems to be at date. These have a specified point of zero. Therefore the type of the attribute is scale.

Statistics: **Missing Values:** None

Range: 02/01/2013 – 18/05/2015

1.1.3 Quote_Flag

Attribute type: nominal (dichotomous) The description of the attributes declares "Quote_Flag" as information about whether an insurance was purchased. This can be only answered with yes or no.

Statistics: **Missing Values:** None

Value	Absolute Frequency	Relative Frequency
0	1605	0.8025
1	395	0.1975

1.2 Field_Info1

Attribute type: nominal

Statistics: **Missing Values:** None

Value	Absolute Frequency	Relative Frequency
J	390	0.365
F	540	0.27
B	730	0.195
E	200	0.1
C	39	0.0505
K	101	0.0195

1.3 Field_Info2

Attribute type: scale All values of "Field_Info2" are numbers between 0 and 2 and have a precision of 4 decimals. This suggests that the values are on a zero-based scale.

Statistic	Value
Missing Values	None
Minimum	0.8746
Maximum	1.0101
Mean	0.9391
Std. Dev.	0.0373
90th Perc.	1.00051
10th Perc.	0.8922

Statistics:

Histogramm :

1.4 Field_Info3

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.5 Field_Info4

Attribute type: nominal (dichotomous) In this Attribute only the values "Y" and "N" appear. Those values are often used as a short version on "Yes" and "No". This field therefore seems to be a yes-no answer.

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
N	1860	0.93
Y	140	0.07

1.6 Coverage_Info1

Attribute type: interval

Statistic	Value
Missing Values	None

1.7 Coverage_Info2

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.8 Coverage_Info3

Attribute type: nominal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
E	673	0.3365
D	388	0.194
G	245	0.1225
K	229	0.1145
F	221	0.1105
A	116	0.058
J	110	0.055
I	7	0.0035
B	4	0.002
C	3	0.0015
H	2	0.001
L	2	0.001

1.9 Sales_Info1

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
1	1490	0.745
0	510	0.255

1.10 Sales_Info2

Attribute type: ordinal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
5	1093	0.5465
3	519	0.2595
4	295	0.1475
2	93	0.0465

1.11 Sales_Info3

Attribute type: nominal

Statistic	Value
Missing Values	None

1.12 Sales_Info4

Attribute type: nominal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
K	410	0.205
P	376	0.188
T	332	0.166
Q	312	0.156
V	298	0.149
R	156	0.078
M	116	0.058

Statistics:

1.13 Sales_Info5

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.14 Personal_Info1

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
N	1993	0.9965
Y	7	0.0035

Statistics:

1.15 Personal_Info2

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.16 Personal_Info3

Attribute type: nominal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
ZA	937	0.4685
XR	104	0.052
XM	81	0.0405
XJ	73	0.0365
XD	63	0.0315
XX	62	0.031
XB	59	0.0295
YH	58	0.029
XH	54	0.027
ZT	51	0.0255
XO	50	0.025
ZF	46	0.023
ZR	40	0.02
ZN	39	0.0195
ZH	37	0.0185
XS	35	0.0175
YF	29	0.0145
XW	23	0.0115
ZG	20	0.01
XE	20	0.01
ZW	18	0.009
YE	17	0.0085
ZC	16	0.008
XC	14	0.007
XQ	14	0.007
XL	7	0.0035
ZE	7	0.0035
ZJ	6	0.003
XI	5	0.0025
ZD	5	0.0025
ZK	4	0.002
XZ	4	0.002
ZU	2	0.001

Statistics:

1.17 Personal_Info4

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
0	1999	0.9995
1	1	0.001

Statistics:

1.18 Personal_Info5

Attribute type: nominal (dichotomous)

Statistics:

Missing Values : 966

1.19 Property_Info1

Attribute type: nominal (dichotomous)

Statistics:

Missing Values : 1

Value	Absolute Frequency	Relative Frequency
N	1738	0.869
Y	261	0.1305

1.20 Property_Info2

Attribute type: nominal

Statistic	Value
Missing Values	None

Statistics:

1.21 Property_Info3

Attribute type: nominal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
O	573	0.2865
R	533	0.2665
J	285	0.1425
D	198	0.099
S	146	0.073
N	92	0.046
I	77	0.0385
A	31	0.0155
Q	30	0.015
E	10	0.005
H	9	0.0045
K	6	0.003
F	5	0.0025
L	3	0.0015
G	2	0.001

Statistics:

1.22 Property_Info4

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
1	1364	0.682
0	636	0.318

Statistics:

1.23 Property_Info5

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.24 Geographic_Info1

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.25 Geographic_Info2

Attribute type: interval

Statistic	Value
Missing Values	None

Statistics:

1.26 Geographic_Info3

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
-1	1942	0.971
hline 25	58	0.029

Statistics:

1.27 Geographic_Info4

Attribute type: nominal (dichotomous)

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
N	1961	0.9805
Y	39	0.0195

Statistics:

1.28 Geographic_Info5

Attribute type: nominal

Statistic	Value
Missing Values	None

Value	Absolute Frequency	Relative Frequency
CA	730	0.365
NJ	540	0.27
TX	491	0.2455
IL	239	0.1195

Statistics:

2 Data Preprocessing

3 Summary

4 Projektinhalt

4.1 Aktivitäten

ID	Aktivität
A	Projektmanager

Table 1: Aktivitäten im Projekt



Figure 1: Organisation-Breakdown-Struktur

List of Figures

1	Organisation-Breakdown-Strukture	14
---	--	----

List of Tables

1	Aktivitäten im Projekt	13
---	----------------------------------	----