



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	Quote_Id	1
1.2	Quote_Date	1
1.3	Quote_Flag	1
1.4	Field_Info1	2
1.5	Field_Info2	2
1.6	Field_Info3	3
1.7	Field_Info4	3
1.8	Coverage_Info1	3
1.9	Coverage_Info2	4
1.10	Coverage_Info3	4
1.11	Sales_Info1	4
1.12	Sales_Info2	4
1.13	Sales_Info3	4
1.14	Sales_Info4	5
1.15	Sales_Info5	5
1.16	Personal_Info1	5
1.17	Personal_Info2	5
1.18	Personal_Info3	6
1.19	Personal_Info4	6
1.20	Personal_Info5	6
1.21	Property_Info1	6
1.22	Property_Info2	7
1.23	Property_Info3	7
1.24	Property_Info4	7
1.25	Property_Info5	7
1.26	Geographic_Info1	8
1.27	Geographic_Info2	8
1.28	Geographic_Info3	8
1.29	Geographic_Info4	8
1.30	Geographic_Info5	8
2	Data Preprocessing	9
3	Summary	9
4	Projekthalt	9
4.1	Aktivitäten	9
	List of Figures	10
	List of Tables	10

1 Initial data exploration

1.1 Quote_Id

1.1.1 Attribute type: nominal

The "Quote_Id" is a customer number. It is neither orderable nor is the number a quantitative value.

1.1.2 Statistics:

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

1.2 Quote_Date

1.2.1 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.

1.2.2 Statistics:

Missing Values: None

Range: 02/01/2013 – 18/05/2015

1.3 Quote_Flag

1.3.1 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.

1.3.2 Statistics:

Missing Values: None

Value	Absolute Frequency	Relative Frequency
0	1605	0.8025
1	395	0.1975

1.4 Field_Info1

1.4.1 Attribute type: nominal

1.4.2 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%

1.5 Field_Info2

1.5.1 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.

1.5.2 Statistics:

Statistic	Value
Missing Values	None
Minimum	0.8746
Maximum	1.0101
Mean	0.9391
Std. Dev.	0.0373

Histogramm :

1.6 Field_Info3

1.6.1 Attribute type: interval

1.6.2 Statistics:

Missing Values : None

1.7 Field_Info4

1.7.1 Attribute type: nominal (dichotomous)

1.7.2 Statistics:

Missing Values : None

1.8 Coverage_Info1

1.8.1 Attribute type: interval

1.8.2 Statistics:

Missing Values : None

1.9 Coverage_Info2

1.9.1 Attribute type: interval

1.9.2 Statistics:

Missing Values : None

1.10 Coverage_Info3

1.10.1 Attribute type: nominal

1.10.2 Statistics:

Missing Values : None 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

1.11 Sales_Info1

1.11.1 Attribute type: nominal (dichotomous)

1.11.2 Statistics:

Missing Values : None 1 1490 0.745 0 510 0.255

1.12 Sales_Info2

1.12.1 Attribute type: ordinal

1.12.2 Statistics:

Missing Values : None 5 1093 0.5465 3 519 0.2595 4 295 0.1475 2 93 0.0465

1.13 Sales_Info3

1.13.1 Attribute type: nominal

1.13.2 Statistics:

Missing Values : None

1.14 Sales_Info4

1.14.1 Attribute type: nominal

1.14.2 Statistics:

Missing Values : None 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

1.15 Sales_Info5

1.15.1 Attribute type: interval

1.15.2 Statistics:

Missing Values : None

1.16 Personal_Info1

1.16.1 Attribute type: nominal (dichotomous)

1.16.2 Statistics:

Missing Values : None N 1993 0.9965 Y 7 0.0035

1.17 Personal_Info2

1.17.1 Attribute type: interval

1.17.2 Statistics:

Missing Values : None

1.18 Personal_Info3

1.18.1 Attribute type: nominal

1.18.2 Statistics:

Missing Values : None 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

1.19 Personal_Info4

1.19.1 Attribute type: nominal (dichotomous)

1.19.2 Statistics:

Missing Values : None 0 1999 0.9995 1 1 0.001

1.20 Personal_Info5

1.20.1 Attribute type: nominal (dichotomous)

1.20.2 Statistics:

Missing Values : 966

1.21 Property_Info1

1.21.1 Attribute type: nominal (dichotomous)

1.21.2 Statistics:

Missing Values : 1 N 1738 0.869 Y 261 0.1305

1.22 Property_Info2

1.22.1 Attribute type: nominal

1.22.2 Statistics:

Missing Values : None

1.23 Property_Info3

1.23.1 Attribute type: nominal

1.23.2 Statistics:

Missing Values : None 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

1.24 Property_Info4

1.24.1 Attribute type: nominal (dichotomous)

1.24.2 Statistics:

Missing Values : None 1 1364 0.682 0 636 0.318

1.25 Property_Info5

1.25.1 Attribute type: interval

1.25.2 Statistics:

Missing Values : None

1.26 Geographic_Info1

1.26.1 Attribute type: interval

1.26.2 Statistics:

Missing Values : None

1.27 Geographic_Info2

1.27.1 Attribute type: interval

1.27.2 Statistics:

Missing Values : None

1.28 Geographic_Info3

1.28.1 Attribute type: nominal (dichotomous)

1.28.2 Statistics:

Missing Values : None -1 1942 0.971 25 58 0.029

1.29 Geographic_Info4

1.29.1 Attribute type: nominal (dichotomous)

1.29.2 Statistics:

Missing Values : None N 1961 0.9805 Y 39 0.0195

1.30 Geographic_Info5

1.30.1 Attribute type: nominal

1.30.2 Statistics:

Missing Values : None CA 730 0.365 NJ 540 0.27 TX 491 0.2455 IL 239 0.1195

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	9
---	--	---

List of Tables

1	Aktivitäten im Projekt	9
---	----------------------------------	---