SPSS Case Study Data Understanding

WEEK 3

Data Science Academy

CASE STUDY

Disclaimer: While the case and data is real, case study modified for educational purposes.
Open "German Credit.csv" dataset. Answer following question.

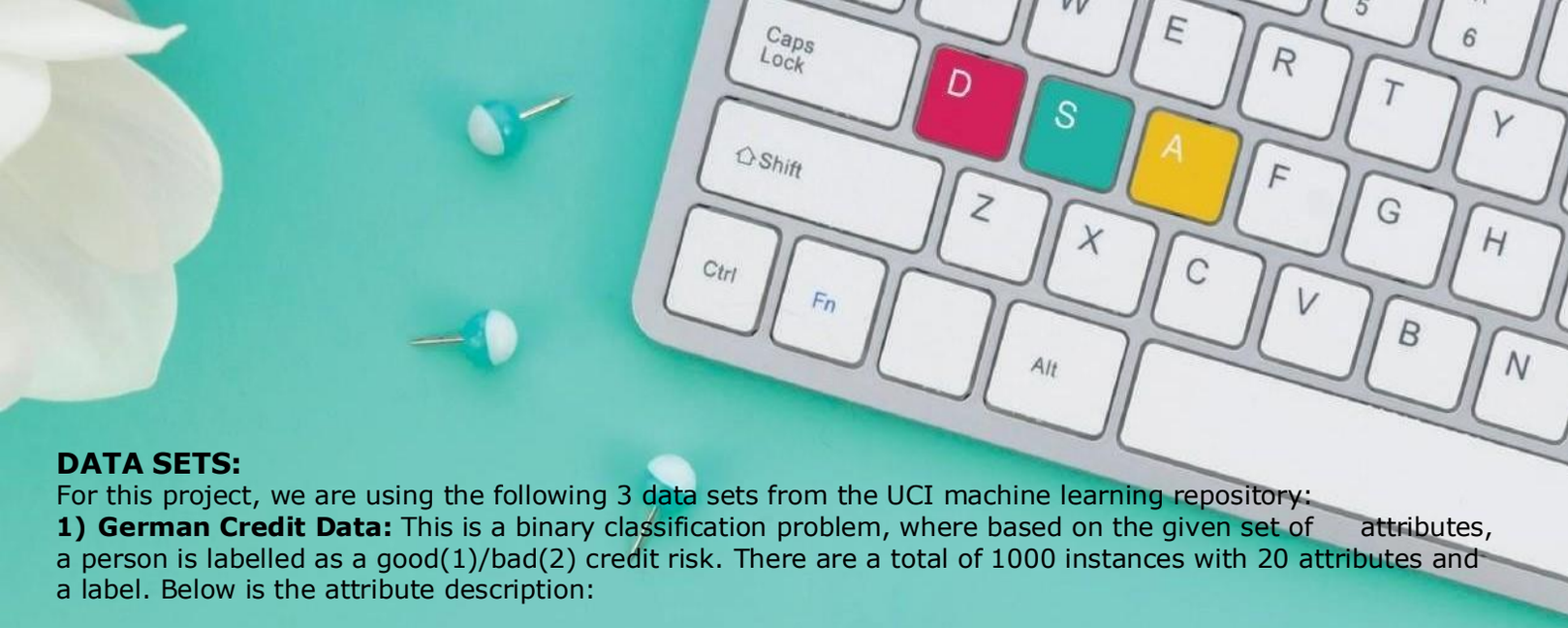
1. Use all diagnostic analysis tests (e.g. Pearson, Chi-Square etc.) on SPSS Statistics and find relationships between variables.

   Note: Assume all variables are normally distributed.

Open "Freight transportation visualization case study.xlsx" dataset. Answer following question.

2. In Driver & Vehicle sheet, use Driver and Vehicle field to create a bar chart that demonstrates top vehicles and drivers.
3. In Marketing vs Sales sheet, find the relationship between marketing expenses and number of orders using scatter plot.
4. In Trucking Price Data sheet, compare distance and price of transportations according to road conditions using box plot.
5. In Time Series sheet, conduct time series analysis for each transportation type using time plot or multiplot in Graph Palette.
6. Explore all data using data audit node and include meaningful insights in your report.
7. Audit Time Series sheet in freight transportation dataset. Change outlier and extreme settings: outlier - 5 standard deviations, extreme – 7 standard deviations. Compare results.

Data Science Academy



DATA SETS:

For this project, we are using the following 3 data sets from the UCI machine learning repository:

1) German Credit Data: This is a binary classification problem, where based on the given set of attributes, a person is labelled as a good(1)/bad(2) credit risk. There are a total of 1000 instances with 20 attributes and a label. Below is the attribute description:

Attribute	Type	Description	Example
Attribute 1	Qualitative	Status of existing checking account	A11 : ... < 0 DM A12 : 0 <= ... < 200 DM A13 : ... >= 200 DM /salary assignments for at least 1year A14 : no checking account
Attribute 2	Qualitative	Credit history	A30 : no credits taken/ all credits paid back duly A31 : all credits at this bank paid back duly A32 : existing credits paid back duly till now A33 : delay in paying off in the past A34 : critical account/ other credits existing (not at this bank)
Attribute 3	Qualitative	Purpose	A40 : car (new) A41 : car (used) A42 : furniture/equipment A43 : radio/television A44 : domestic appliances A45 : repairs A46 : education A47 : (vacation - does not exist?) A48 : retraining A49 : business A410 : others
Attribute 4	Qualitative	Savings status	A61 : ... < 100DM A62 : 100 <= ... < 500 DM A63 : 500 <= ... < 1000 DM A64 : .. >= 1000 DM A65 : unknown/ no savings account
Attribute 5	Qualitative	Present employment since	A71 : unemployed A72 : ... < 1 year A73 : 1 <= ... < 4 years A74 : 4 <= ... < 7 years A75 : .. >= 7 years
Attribute 6	Qualitative	Personal status and gender	A91 : male : divorced/separated A92 : female : divorced/separated/married A93 : male : single A94 : male : married/widowed A95 : female : single

Attribute7	Qualitative	Other debtors/ guarantors	A101 : none A102 : co-applicant A103 : guarantor
Attribute8	Qualitative	Property	A121 : real estate A122 : if not A121 : building society savings agreement/ life insurance A123 : if not A121/A122 : car or other, not in attribute 6 A124 : unknown / no property
Attribute9	Qualitative	Other installment plans	A141 : bank A142 :stores A143 : none
Attribute 10	Qualitative	Housing	A151 : rent A152 : own A153 : for free
Attribute 11	Qualitative	Job	A171 : unemployed/ unskilled - non-resident A172 : unskilled - resident A173 : skilled employee / official A174 : management/ self-employed/highly qualified employee/ officer
Attribute 12	Qualitative	Telephone	A191 : none A192 : yes, registered under the customer's name
Attribute 13	Qualitative	Foreign Worker	A201 : yes A202 : no
Attribute 14	Numerical	Installment rate in percentage of disposable income	1,2,3,4 etc.
Attribute 15	Numerical	Present residence since	1,2,3,4 etc.
Attribute 16	Numerical	Age in years	22, 49 etc.
Attribute 17	Numerical	Number of existing credits at this bank	1,2,3 etc.
Attribute 18	Numerical	Duration in month	6, 48 etc
Attribute 19	Numerical	Credit amount	1169, 5951 etc.
Attribute 20	Numerical	Number of people being liable to provide maintenance for	1,2
Label	Binary	Indicates good/bad risk	1 = Good, 2 = Bad