Data Analysis Skills — German Credit Dataset

Small quiz based on the German Credit dataset

German Credit - EDA

This is a short quiz to help you develop your pandas skills. The grade is returned is formative only (i.e., does not contribute towards CA).

When entering a numerical answer to questions be mindful of:

Percentages should be entered as decimals, for example to enter 34.2% use 0.342

Floating point numerical answers have a required accuracy. This is typically stated as the number of correct — after rounding — digits after the decimal point (abbreviated to DP).
For example, if the answer to a question was 265.15765054700429743 and you were asked to be accurate to 2DP then as long as you enter at last two digits after the decimal point then you will be fine.

(I tend to just cut and paste numerical answers (from python/excel/etc.), so I don't worry about the number of places as I will have more than enough.)

Question 1:

What is the total of Loan_Amount?

Question 2:

What is the mean Loan_Duration to 2DP?

Question 3:

What is the standard deviation of Loan_Amount (to 2DP)?

Question 4:

What is the mean Loan_Amount for applications that have Year_of_Residence equal to 1 (answer to 2DP)?

Question 5:

What is the range of Loan_Amount?

Question 6:

Classify the Age distribution as:

- Left-skewed
- Symmetric

- Right-skewed

Question 7:

What is the interquartile range (IQR) of Loan_Amount (to 2DP)?

Question 8:

Using the 1.5*IQR from the quartiles rule, how many outlier observations are in Loan_Amount?

Question 9:

What is the third most popular reason cited for loan?

- (A410) others
- (A49) business
- (A40) car (new)
- (A45) repairs
- (A46) education
- (A44) domestic appliances
- (A41) car (used)
- (A48) retraining
- (A42) furniture/equipment
- (A43) radio/television

Question 10:

How many distinct reasons are cited for asking for a loan?

Question 11:

Which reason cited for loan has the highest rate of bad claims?

- (A40) car (new)
- (A48) retraining
- (A44) domestic appliances
- (A41) car (used)
- (A46) education
- (A49) business
- (A42) furniture/equipment
- (A410) others
- (A43) radio/television
- (A45) repairs

Question 12:

Which Credit_History has the lowest rate of bad claims?

- (A32) existing credits paid back duly till now
- (A33) delay in paying off in the past
- (A30) no credits taken/
- (A34) critical account/
- (A31) all credits at this bank paid back duly

Question 13:

What is the average age of applicants that have Credit_Rating of Bad?

Question 14:

How many reasons cited for loans have a mean Loan_Duration greater than the mean of Loan_Duration overall?

Question 15:

What is the correlation between Loan_Duration and Loan_Amount (to 2DP)?

Question 16:

What is the probability of an applicant is female? (enter answer as a decimal to 2DP)?

Question 17:

What is the probability of an applicant that owns their own property, does not have a telephone? (enter answer as a decimal to 2DP)?

Question 18:

Which categorical feature has the most impact on the target (using chi2 as the metric)?

- Housing
- Installment_Plans
- Savings
- Loan_Purpose
- Other_Debtors_Guarantors
- Telephone
- Foreign
- Gender_and_Status
- Gender
- Years_Employed
- Account_Balance
- Property
- Job

- Credit_History