**Practical 6 - Data from an online microlending platform has been collected. This data contains details of the purpose for which the loans**

**would be used and how the loan is funded. Additional information on the country of loan recipient and the poverty levels of**

**the country are also given.**

**It is to be seen whether a loan would be funded or not based on the available data.**

**Variable Description:**

**Parameter Description**

**activity Activity for which loan was requested**

**borrower\_genders Gender of the borrowers**

**country Country in which loan was disbursed**

**country\_code ISO country code**

**currency\_policy The currency policy in which loan was disbursed**

**distribution\_model Loan disbursed through field partner or not**

**lender\_count the total number of lenders that contributed to this loan**

**original\_language language of the original loan application**

**loan\_amount The amount disbursed by the field agent to the borrower(USD)**

**repayment\_interval intervel between payments**

**sector High level category**

**status The status of a loan : whether funded,not funded**

**term\_in\_months The duration for which the loan was disbursed in months**

**rMPI Multiple Poverty Index**

**1. How many columns are of ‘object’ data type? Read the given data “lendingdata.csv” and save it**

**as a dataframe called data, and answer the questions below:**

**2. Find the total number of missing values in the data set?**

**3. Identify which of the columns contain redundant information and can be dropped from the**

**dataframe.**

**4. What is the third quartile value of the variable “loan\_amount”?**

**5. What is the percentage split of the different categories in the column “repayment\_interval” after**

**dropping the missing values?**

**6. What is the minimum loan amount disbursed in the Agriculture sector?**

**Solution :**

**# Import necessary libraries**

import pandas as pd

**# Load the data into a dataframe**

data = pd.read\_csv('lendingdata.csv')

**# Step 1: Determine the number of columns with 'object' data type**

object\_columns = data.select\_dtypes(include=['object']).columns

num\_object\_columns = len(object\_columns)

print(f"Number of columns with 'object' data type: {num\_object\_columns}")

print(f"Columns with 'object' data type: {object\_columns.tolist()}")

**# Step 2: Find the total number of missing values in the dataset**

total\_missing\_values = data.isnull().sum().sum()

print(f"Total number of missing values in the data set: {total\_missing\_values}")

**# Step 3: Identify and drop redundant columns**

**# Assuming 'country\_code' is redundant because 'country' is present**

columns\_to\_drop = ['country\_code']

data = data.drop(columns=columns\_to\_drop)

print(f"Columns after dropping redundant information: {data.columns.tolist()}")

**# Step 4: Calculate the third quartile value (Q3) of the variable 'loan\_amount'**

third\_quartile\_value = data['loan\_amount'].quantile(0.75)

print(f"Third quartile value of the variable 'loan\_amount': {third\_quartile\_value}")

**# Step 5: Calculate the percentage split of different categories in 'repayment\_interval'**

**# Drop rows with missing values in 'repayment\_interval'**

repayment\_data = data['repayment\_interval'].dropna()

**# Calculate the percentage split of different categories**

percentage\_split = repayment\_data.value\_counts(normalize=True) \* 100

print("Percentage split of different categories in 'repayment\_interval':")

print(percentage\_split)

**# Step 6: Find the minimum loan amount disbursed in the Agriculture sector**

min\_loan\_amount\_agriculture = data[data['sector'] == 'Agriculture']['loan\_amount'].min()

print(f"Minimum loan amount disbursed in the Agriculture sector: {min\_loan\_amount\_agriculture}")