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}")
```