PANDAS ASSIGNMENT

DESCRIPTION

This is a two part assignment. In the first part you will write a python program to generate a Pandas data frame of random test data. This is an extremely common task when getting larger systems to work. This data will be stored as an Excel workbook. In the second part, you will write a python program to read the excel workbook and perform a variety of summary and aggregation tasks.

PART I

In order to generate random data, you will need to write functions like the following

```
1  import pandas as pd
2  import numpy as np
3
4
5  def create_random_state(n):
6      states = ['MO', 'IN', 'TN', 'TX', 'NY', 'OH']
7      rs = np.random.choice(states, n, replace=True)
8      return list(rs)
```

You need to generate 10,000 records containing the following fields and written to an Excel Workbook using Pandas:

- First Name (Given Name) – Have at least 20 options. You may use names appropriate to your country and culture.
- Last Name (Last Name) – Have at least 20 options. You may use names appropriate to your country and culture.
- Gender – Male/Female. Should match name (Sue is Female)
- Region – Have at least 10. These can be US States or states/regions from any country.
- Age – Reported in decades: 1 for teens, 2 for twenties, and so forth
- Customer Id – Random string of the form: A123-e7. First character is a capital alphabetic, then 3 digits a hyphen and a digit.
- Loan – A large random number in whatever currency
- Principal – Random Amount paid, less than loan
- Owed – Difference between Loan and Principal
- Interest Due – Keep it simple. Just 20% of Owed

PART 2

Using the Excel Workbook created in the first part, create panda's data frames addressing the following questions. You may answer more than one question per data frame. You decide the best way to deliver the data. Write each data frame to a sheet in a workbook.

- What is the mean loan size per region?
- What is the mean loan size per gender?
- What is the mean load size per age group?
- What is the maximum amount owed by region for each age group?
- What is the maximum amount owed by gender for each age group?
- How much interest can the company expect to make from each region?
- Which man in each region owes the most?