Mini Project - Python
Asset Management Company Dataset

Introduction

The Objective is to identify the best countries and a suitable investment type for making the investment. The overall strategy is to invest where others are investing, implying that the best countries are the ones ‘where most investors are investing’

Data Description

This is real investment data taken from crunchbase.com

Data Dictionary

Companies Details	
Attributes	Description
Permalink	Unique ID of Companies
name	name of Companies
homeurl_page	Website URL
Category_list	Categories to which company belong
Status	Operational Status
country_code	country_code
state_code	state
rounds2	
Attributes	Description
company_permalink	Unique ID of company
funding_round_permalink	Unique ID of funding round

funding_round_type	Type of funding - Venture,angel, private equity
funding_round_code	Round of venture funding (round A,B..)
Funding_at	Date of funding
raised_amount_usd	Money raised in funding (USD)

Answer the below questions -

1. Reading the Rounds2, companies.txt files.

2. How many unique companies are present in rounds2?

3. How many unique companies are present in the companies file?

4.Write the code to convert companies[‘permalink’] and Round2[‘company_permalink’] columns to uppercase.

5. Are there any companies in the rounds2 file which are not present in companies.txt ?

6. Merge the two data frames so that all variables (columns) in the company's frame are added to the rounds2 data frame. Name the merged frame master_dataframe. How many observations are present in master_frame ?

7. Write the code to drop the redundant column ‘'company_permalink' from the master_dataframe.

8. Write the code to find the percentage missing value (column-wise) in master_dataframe. Drop unnecessary columns.

9. Create a pivot table to compare the mean and median values for ‘raised_amount_usd’ and ‘funding_round_type’ across different funding round categories..

10. Filter the master_dataframe on ‘raised_amount_usd’ such as ‘raised_amount_usd’ lie between 4 Millions to 15 Millions.

11.Identify the top 9 countries in terms of highest investment in ‘venture’ fund type.

12. Create the ‘main_category’ column by extracting the main sector using the column ‘category_list’.
Hint - Use the Lambda function or string function.

13. Using the user defined function convert the ‘rasied_amount_used’ column into a categorical column as follow-

a) If the amount is less than 5 Millions then recode as TypeA.
b) If the amount is greater than 5 Millions and less than 7 Millions then recode as TypeB.
c) If the amount is greater than 7 Millions then recode as TypeC.