PGPDSE FT Mini-Project – Subject
Asset Management Company Dataset

Participant List: Mayur, Pratiksha, Ashish, Piyush, Pratik **(GROUP - 6)**

Instructions for report preparation

- Create a separate page for each of the following sections
- Do not copy-paste the contents, please provide references to the contents of your literature
- Keep the header-footer as given.
- Submit the PDF format

Industry Review

- Industry Review – Current practices, Background Research
- Literature Survey - Publications, Application, past and undergoing research

Dataset and Domain

- Data Dictionary
- Dataset description
- Business Importance

Business Questions Discussion (Shared with the project document)

- Write the question
- Provide understanding
- Code snippet
- Solution
- Inference – if required

Business Questions Discussion (Identified by the students)

- Write the question
- Provide understanding
- Code snippet
- Solution
- Inference – if available

Conclusion

- Write the lessons learned
- Skills used
- Domain understanding developed

Industry Review

What is assets management?

The process of planning and controlling the acquisition, maintenance, operations, renewal and disposal of organizational assets is known as asset management. Assets management improves the deliver potential of assets and minimizes the risks and costs involved.

What is an asset management company?

An asset management company (AMC) is a firm that invests a pooled fund of capital on behalf of its clients. The capital is used to fund different investments in various asset classes. Asset management companies are commonly referred to as money managers or money management firms as well.

Assets are the core part of any business; therefore, it is necessary and challenging to manage fixed assets as it grows exponentially with the growth of any company.

Individual investors usually lack the expertise and resources to consistently produce strong investment returns over time. Therefore, many investors rely on asset management companies to invest capital on their behalf.

Asset management companies are usually a group of investment professionals with broad market expertise. With a large amount of pooled capital, they are able to utilize diversification and complex investment strategies to generate returns for investors.

Asset management companies come in many different forms and structures, such as:

- Hedge funds
- Mutual funds
- Index funds
- Exchange-traded funds
- Private equity funds
- Other funds

In addition, they invest on behalf of various types of clients, such as:

- Retail investors
- Institutional investors
- Public sector (government organizations)
- Private sector
- High-net-worth clients

Current Practices followed:

There are 3 methods of managing assets

Manual organization with spreadsheets and process agreements.

DAM (Digital Data Management) software purpose-built for managing your assets.

Asset management tools provided with centralized storage systems.

Analyse Asset Performance: Keep track of performance of the assets with the best tool possible.

Ensure Accurate Depreciation Tracking: Choose appropriate software management tool to assess the depreciation of the asset.

Maintain Asset: The regular maintenance is mandatory.

Determine Asset Value: Set levels of services while removing ghost and zombie assets.

Regulatory Compliance: Finding accurate reports.

Dataset and Domain

Data Dictionary

Company 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)

Data Description

This is real investment data taken from crunchbase.com.

Company details: The data for different companies has been provided for investing in different sectors by taking reference from past investment strategies on several sectors in various countries. The company details include the unique company id's their names, URL's for their websites the category to which these said companies belong operational statuses countries in which they operate and their codes.

Round2: gives us the funding rounds, its types which are coded and include types such as seeds venture, private equity and many more. The date of fundings and money raised as per the investments.

Business Importance

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

The given data set is to be filtered and manipulated by into python workbench

Thus, the first step involves importing functions to solve the provided data files and worked upon them by using:

import numpy as np

import pandas as pd

1) Reading the Rounds2, companies.txt files.

data = pd.read_csv('rounds2.csv',encoding = 'ANSI')

data.head()

solution: The csv is uploaded in the data base to perform operation we import the csv file using the above command

2) How many unique companies are present in rounds2?

len(rounds2['company_permalink'].unique())

solution: to find unique number of companies from the imported rounds2 csv file we use length(len)

inference: we get 90247 unique companies

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

len(companies['permalink'].unique())

solution: to find unique number of companies from the imported companies csv file we use length(len)

inference: we get 66368 unique companies

4) Write the code to convert companies['permalink'] and Round2['company_permalink'] columnsto uppercase.

companies['permalink'] = companies['permalink'].str.upper()

companies.head()

rounds2['company_permalink'] = rounds2['company_permalink'].str.upper()

rounds2.head()

solution: we find that all permalinks and company_permalinks have been converted in uppercase

inference: This is done to make the permalinks and company_permalinks uniques to discard any duplication while performing further tasks while merging the data.

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

df1 = pd.DataFrame(rounds2.company_permalink.unique())

df2 = pd.DataFrame(companies.permalink.unique())

df2.equals(df1)

solution: False

inference: There are no duplicated attributes in the data of permalinks

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 observationsare present in master_frame ?

master_dataframe = pd.merge(rounds2 , companies, left_on = 'company_permalink' , right_on= 'permalink', how = 'left')

master_dataframe.head().len(master_dataframe)

solution: we get a common data frame in the form of master_frame which includes all the columns and rows from both data frames of companies and rounds. They are both combined on the columns of company_permalinks and permalinks respectively.

Inference: We do this to make it easy to perform operations on a single data frame and avoid complications and mistakes while working with two data frames.

7) Write the code to drop the redundant column ''company_permalink' from the master_dataframe?

Master_dataframe = master_dataframe.drop('company_permalink', axis = 1)

master_dataframe.head()

solution: The data frame is now removed of unnecessary column which contains similar values as permalink

inference: removing clutter and duplicate column

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

(master_dataframe.isnull().sum()/ len(master_dataframe))*100

master_dataframe.drop('funding_round_code', axis = 1)

master_dataframe.drop('founded_at', axis = 1)

master_dataframe.head()

solution/inference: The funding_round_code does not serve us any purpose thus has been removed

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.

df_pivot=pd.pivot_table(master_dataframe,index=['funding_round_type'], values=['raised_amount_usd'], aggfunc=[np.median,np.mean])

df_pivot

solution: A table comparing median and mean according to the raised_amount_usd is shown as output.

Inference: To find the average values of different funding types for investment.

10) Filter the master_dataframe on 'raised_amount_usd' such as 'raised_amount_usd' liebetween 4 Millions to 15 Millions.

master_dataframe.loc[(master_dataframe['raised_amount_usd']>400000) &

(master_dataframe['raised_amount_usd']>1500000)]

master_dataframe.head()

solution: We come across a table showing different columns of companies their names and countries which have raised amounts between 4 million and 15 million

inference: The objective here is to find out companies and countries which have invested money, check competitors and realise investing opportunities.

11) Identify the top 9 countries in terms of highest investment in 'venture' fund type.

master_dataframe = master_dataframe.loc[master_dataframe['funding_round_type']=='venture']

master_dataframe ()

top=pd.DataFrame(master_dataframe.groupby('country_code')['raised_amount_usd'].sum().sort_values(ascending= False).head(9))

top

solution: The output is given as a table that shows us investments in venture type across the top 9 countries.

Inference: From the above operation we get a clear picture of countries intrested in the venture type funding oportunities.

12) Create the 'main_category' column by extracting the first value using the column 'category_list'

Hint - Use the Lambda function or string function.

master_dataframe ['main_category'] = master_dataframe ['category_list'].str.split('|').str[0].copy()

master_dataframe

solution: To order the category for easy access and understanding as too many subtypes are present

13) Using the user defined function convert the 'rasied_amount_used' column into a categoricalcolumn as follow-

If the amount is less than 5 Millions then recode as TypeA.

If the amount is greater than 5 Millions and less than 7 Millions then recode as TypeB.

If the amount is greater than 7 Millions then recode as TypeC.

```
def categorical_column(value):

   if value < 500000:

      return 'TypeA'

   elif value > 500000 and value< 700000:

      return 'TypeB'

   elif value > 700000:

      return 'Type C'
```

master_dataframe['categorical_column'] = master_dataframe['raised_amount_usd'].map(categorical_column)

display(master_dataframe.head(15))

solution: the master_dataframe is returned ordered by categorising into only 3 types.

Inference : This makes it easy to find companies and which deal with small, medium and large scale investment

Additional Questions:

Q1) How many company have venture or seed funding round type ?

Solution:In round 2 csv file, we have to first find which company have venture and seed funding_round_type and total how many company have venture and seed funding_round_type

filter1= rounds2['funding_round_type']=='venture'

filter2= rounds2['funding_round_type']=='seed'

len(rounds2[filter1|filter2])

Inference: Applying filter on the column funding_round_type.filter Filter the venture seed funding_round_type and just by applying len function we get number of company which having venture or seed funding round type.

Q2) In compnaies.text file find the name of company which having folowing url 'http://inovatiff.com'

Solution: In this question we have to find the name of company having url 'http://inovatiff.com'.

companies[companies['homepage_url']=='http://www.zznode.com']['name']

Inference: We need filtering condition on column ['homepage_url'] .If the following condition true it return true or false.

[companies['homepage_url']=='http://www.zznode.com']

But we interested in name of company so just putting ['name'] column after giving condition we get name of company which having above url.

Q3) Create a pivot table to show the amount raised by each country in each category.

Solution: In this question we have to find, how the distribution of 'raised_amount_usd' by each country and each category.

df_pivot1 = pd.pivot_table(master_dataframe,index=['country_code'] , columns=['categorical_column'], values=['raised_amount_usd'])

Inference: The solution may be written using pivot table, The pivot table takes simple column-wise data as input, and groups the entries into a two-dimensional table that provides a multidimensional summarization of the data.

In above syntax country code columns taken as index and the numerical column taken as value using this pivot table we get distribution according to each country and each category.

Q4) Companies want to invest in top 2 English speaking countries like USA and IND.Create three separate data frames D1 and D2 for each of the three countries containing the observations of funding type FT falling within the 10-15 million USD range. The two data frames should contain:

All the columns of the master_frame along with the primary sector and the main sector.

The total number (or count) of investments for each main sector in a separate column

The total amount invested in each main sector in a separate column

Solution: We have to create the two data frame one is for IND and other for USA,but the condition is the funding type FT falling within the 10-15 million USD range. In this two dataframe we want the column of the master frame along with primary and main sector And the total amount of investment of main sector should have separate column.

master_dataframe = master_dataframe[(master_dataframe['country_code'] == 'USA') | (master_dataframe['country_code'] == 'IND')]

master_dataframe.head()

D1 = master_dataframe[(master_dataframe['country_code'] == 'USA') & (master_dataframe['raised_amount_usd'] >= 10000000) &

(master_dataframe['raised_amount_usd'] <= 15000000)]

USA=D1[['raised_amount_usd','main_category']].groupby('main_category').agg(['sum', 'count']).rename(columns={'sum':'Total_amount','count' : 'Total_count'})

D1 = D1.merge(USA, how='left', on ='main_category')

D1.head()

D2 = master_dataframe[(master_dataframe['country_code'] == 'IND') & (master_dataframe['raised_amount_usd'] >= 10000000) &(master_dataframe['raised_amount_usd'] <= 15000000)]

IND=D2[['raised_amount_usd','main_category']].groupby('main_category').agg(['sum', 'count']).rename(columns={'sum':'Total_amount','count' : 'Total_count'})

D2 = D2.merge(IND, how='left', on ='main_category')

D2.head()

Inference: First part is to create the two separate dataframe for USA and IND having funding type FT falling within the 10-15 million USD range.for that we need filter condition on ['raised_amount_usd']

Next applying groupby on categorical column ['main_category'] and aggregate function on numerical column ['raised_amount_usd'] i.e sum and count and just giving merge left module function on categorical column we get accepted result.

Q5) Find the top 5 categories with most amount raised usd for USA and IND respectively.

Solution: We intrested on top 5 company so we need sorting in descending manner for USA and IND country.

USA.sort_values(by=('raised_amount_usd','Total_amount'),ascending=False).head()

IND.sort_values(by=('raised_amount_usd','Total_amount'),ascending=False).head()

Inference: This problem solved using sorting method i.e sort by values.

by: Single/List of column names to sort Data Frame by. We need 'raised_amount_usd','Total_amount' and asending =false since we need the data in desending order.head gives top element by giving upto which data we need.

Conclusion:

- The NumPy allows us to perform mathematical and scientific computations, which contains multidimensional arrays and matrices.
- Huge amount of data can be effectively filtered by the performed operations.
- The operations require analytical, logical skills to deliver the desired output.

Referring from Additional Question, Q4 and Q5

There are 14 different types of investments.

USA amounts as the most invested country.

• China follows USA, United Kingdom, India Canada are among the Top 5 countries to attract investments.

• France, Germany, Israel and Japan are other among the top 9 invested countries.

• USA receives the highest investment amount (approx. 420k Million USD)

The 5 main sectors to invest in USA are:

Biotechnology

Software

Enterprise Software

Advertising

Healthcare

The 5 main sectors to invest in India are

E-Commerce

Mobile

Software

Analytics

Real Estate