

Mini-Project

TITLE : Asset Management Company Dataset

Submitted by :

Sneha Sanjeevkumar (project no :2)

Problem Definition

- Dataset:

Companies, rounds2

- Problem Statement :

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 Set Description

• 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

Data Set Description

- **DATA DICTIONARY:**

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 Set Description

- This is real investment data taken from crunchbase.com.
- 1. 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 include seeds venture, private equity and many more and funding round codes, dates of fundings and money raised in USD per investment.

Business Importance of Problem

- Investment Decision Making
- Client servicing
- Risk management

To identify the best sectors, countries, and a suitable investment type for making investments.

Goals of data analysis:

- Investment type analysis
- Country Analysis
- Sector Analysis

Project Flow – Question 1

- Reading the Rounds2, companies.txt files.

Code –

```
rounds=pd.read_csv(r"C:\Users\akshaya_gv\Desktop\NPV\Project 2\Dataset\rounds2 (1).csv",encoding='ANSI')
rounds|
```

Output -

	company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd
0	/organization/-fame	/funding-round/9a01d05418af9f794eeb7f7ace91f638	venture	B	05/01/2015	10000000.0
1	/ORGANIZATION/-QOUNTER	/funding-round/22dacff496eb7acb2b901dec1dfe5633	venture	A	14/10/2014	NaN
2	/organization/-qounter	/funding-round/b44fbb94153f6cdef13083530bb48030	seed	NaN	01/03/2014	700000.0
3	/ORGANIZATION/-THE-ONE-OF-THEM-INC-	/funding-round/650b8f704416801069bb178a1418776b	venture	B	30/01/2014	3406878.0
4	/organization/0-6-com	/funding-round/5727accaaaa57461bd22a9bdd945382d	venture	A	19/03/2008	2000000.0
...
114944	/organization/zzzzapp-com	/funding-round/8f6d25b8ee4199e586484d817bcda05	convertible_note	NaN	01/03/2014	41313.0
114945	/ORGANIZATION/ZZZZAPP-COM	/funding-round/ff1aa06ed5da186c84f101549035d4ae	seed	NaN	01/05/2013	32842.0
114946	/organization/ãeron	/funding-round/59f4dce44723b794f21ded3daed6e4fe	venture	A	01/08/2014	NaN
114947	/ORGANIZATION/Ã"ASYS-2	/funding-round/35f09d0794651719b02bbfd859ba9ff5	seed	NaN	01/01/2015	18192.0

Project Flow – Question 1

- Reading the Rounds2, companies.txt files.

Code –

```
company=pd.read_csv(r"C:\Users\akshaya_gv\Desktop\NPV\Project 2\Dataset\companies (1).txt",sep='\t')
company
```

Output -

	permalink	name	homepage_url	category_list	status	country_code	state_code	region	city	founded_
0	/Organization/-Fame	#fame	http://livfame.com	Media	operating	IND	16	Mumbai	Mumbai	Na
1	/Organization/-Qounter	:Qounter	http://www.qounter.com	Application Platforms Real Time Social Network...	operating	USA	DE	DE - Other	Delaware City	04-09-201
2	/Organization/-The- One-Of-Them-Inc-	(THE) ONE of THEM,Inc.	http://oneofthem.jp	Apps Games Mobile	operating	NaN	NaN	NaN	NaN	Na
3	/Organization/0-6-Com	0-6.com	http://www.0-6.com	Curated Web	operating	CHN	22	Beijing	Beijing	01-01-200
4	/Organization/004- Technologies	004 Technologies	http://004gmbh.de/en/004- interact	Software	operating	USA	IL	Springfield, Illinois	Champaign	01-01-201
...
66363	/Organization/Zznode- Science-And- Technology-Co...	ZZNode Science and Technology	http://www.zznode.com	Enterprise Software	operating	CHN	22	Beijing	Beijing	Na
66364	/Organization/Zzzzapp- Com	Zzzzapp Wireless Ltd.	http://www.zzzzapp.com	Advertising Mobile Web Development Wireless	operating	HRV	15	Split	Split	13-05-201
66365	/Organization/ÄEron	ÄEERON	http://www.aeron.hu/	NaN	operating	NaN	NaN	NaN	NaN	01-01-201
66366	/Organization/ÄAsys-2	Äasys	http://www.oasys.io/	Consumer Electronics Internet of Things Teleco...	operating	USA	CA	SF Bay Area	San Francisco	01-01-201
66367	/Organization/ ÄNovatiff-Reklam-Ve- TanÄTÄM-H...	Änovatiff Reklam ve TanÄTÄm Hizmetleri Tic	http://inovatiff.com	Consumer Goods E- Commerce Internet	operating	NaN	NaN	NaN	NaN	Na

66368 rows × 10 columns

Project Flow – Question 2

- How many unique companies are present in rounds2?

.nunique() gives the count of unique of the companies in rounds2 dataset

Code – `rounds['company_permalink'].nunique()`

`Out[457]: 90247`

Output -

Project Flow – Question 3

- How many unique companies are present in the companies file?

.nunique() gives the count of unique of the companies in companies dataset

Code – `company['name'].nunique()`

Output - `66102`

Project Flow – Question 4

- Write the code to convert `companies['permalink']` and `Round2['company_permalink']` columns to uppercase

`.str.upper()` converts the strings in the series in uppercase.

Code – `rounds['company_permalink'].str.upper()`

Output -

```

0          /ORGANIZATION/-FAME
1      /ORGANIZATION/-QOUNTER
2      /ORGANIZATION/-QOUNTER
3      /ORGANIZATION/-THE-ONE-OF-THEM-INC-
4      /ORGANIZATION/0-6-COM
...
114944      /ORGANIZATION/ZZZZAPP-COM
114945      /ORGANIZATION/ZZZZAPP-COM
114946      /ORGANIZATION/ÄERON
114947      /ORGANIZATION/Ä"ASYS-2
114948      /ORGANIZATION/Ä°NOVATIFF-REKLAM-VE-TANÄ±TÄ±M-H...
Name: company_permalink, Length: 114949, dtype: object

```

Project Flow – Question 4

- Write the code to convert `companies['permalink']` and `Round2['company_permalink']` columns to uppercase

Code – `company['permalink'].str.upper()`

Output -

```

0          /ORGANIZATION/-FAME
1          /ORGANIZATION/-QOUNTER
2          /ORGANIZATION/-THE-ONE-OF-THEM-INC-
3          /ORGANIZATION/0-6-COM
4          /ORGANIZATION/004-TECHNOLOGIES
...
66363      /ORGANIZATION/ZZNODE-SCIENCE-AND-TECHNOLOGY-CO...
66364          /ORGANIZATION/ZZZZAPP-COM
66365          /ORGANIZATION/ÄERON
66366          /ORGANIZATION/Ä”ASYS-2
66367      /ORGANIZATION/Ä°NOVATIFF-REKLAM-VE-TANÄ±TÄ±M-H...
Name: permalink, Length: 66368, dtype: object

```

Project Flow – Question 5

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

First we got the unique companies in each dataset and noticed that some of the strings in rounds datasets are in uppercase. So we converted both the series in uppercase and checked for companies in rounds2 dataset that not present in companies dataset by adding '~' not operator.

Code –

```
2 company['permalink'].unique()

array(['/Organization/-Fame', '/Organization/-Qounter',
      '/Organization/-The-One-Of-Them-Inc-', ...,
      '/Organization/Ã\x81Eron', '/Organization/Ã”Asys-2',
      '/Organization/Ã°Novatiff-Reklam-Ve-Tanä±Tä±M-Hizmetleri-Tic'],
      dtype=object)
```

```
1 rounds['company_permalink'].unique()

array(['/organization/-fame', '/ORGANIZATION/-QOUNTER',
      '/organization/-qounter', ..., '/organization/ã\x81eron',
      '/ORGANIZATION/Ã”ASYS-2',
      '/organization/ã°novatiff-reklam-ve-tanä±tä±m-hizmetleri-tic'],
      dtype=object)
```

Project Flow – Question 5

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

Code –

```
rounds[~(rounds['company_permalink'].str.upper()).isin(company['permalink'].str.upper())]
# companies of rounds2 dataset that are not in companies dataset
```

Output -

	company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd
729	/ORGANIZATION/51WOFANG-Æ — Å¿\$Æ~Æ*¿	/funding- round/346b9180d276a74e0fbb2825e66c6f5b	venture	A	06/07/2015	5000000.0
45176	/organization/huizuche-com- æf ç\$ÿè½!	/funding- round/8f8a32dbeeb0f831a78702f83af78a36	seed	NaN	18/09/2014	NaN
54751	/ORGANIZATION/LAWPĂ DĂ	/funding- round/56a8ba75784c0028e2eacd0c0e205119	grant	NaN	23/03/2015	5000.0

Project Flow – Question 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_dataframe ?

master_dataframe is created by merging rounds and company data frame by the permalinks.

Code –

```
rounds['company_permalink']=rounds['company_permalink'].str.upper()
company['permalink']=company['permalink'].str.upper()
master_dataframe=pd.merge(left=rounds,right=company,left_on='company_permalink',right_on='permalink',how='inner')
master_dataframe
```

Output -

	company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	
0	/ORGANIZATION/-FAME	/funding-round/9a01d05418af9f794eebf7ace91f638	venture	B	05/01/2015	10000000.0	/ORG
1	/ORGANIZATION/-QOUNTER	/funding-round/22dacff496eb7acb2b901dec1dfe5633	venture	A	14/10/2014	NaN	
2	/ORGANIZATION/-QOUNTER	/funding-round/b44fbb94153f6cdef13083530bb48030	seed	NaN	01/03/2014	700000.0	
3	/ORGANIZATION/-THE-ONE-OF-THEM-INC-	/funding-round/650b8f704416801069bb178a1418776b	venture	B	30/01/2014	3406878.0	/ORC O
4	/ORGANIZATION/0-6-COM	/funding-round/5727acc8aa57461bd22a9bdd945382d	venture	A	19/03/2008	2000000.0	/ORGAN
...
114941	/ORGANIZATION/ZZZAPP-COM	/funding-round/8f6d25b8ee4199e586484d817bcda05	convertible_note	NaN	01/03/2014	41313.0	/ORGANI
114942	/ORGANIZATION/ZZZAPP-COM	/funding-round/ff1aa06ed5da186c84f101549035d4ae	seed	NaN	01/05/2013	32842.0	/ORGANI
114943	/ORGANIZATION/ÄÆERON	/funding-round/59fdce44723b794f21ded3daed6e4fe	venture	A	01/08/2014	NaN	/ORGAN
114944	/ORGANIZATION/ÄÆASYS-2	/funding-round/35f09d0794651719b02bbfd859ba9ff5	seed	NaN	01/01/2015	18192.0	/ORGANI

Project Flow – Question 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_dataframe ?

len() gives the number of observations in the data frame.

Code –

```
print('No. of observations:', len(master_dataframe))
```

Output -

```
No. of observations: 114946
```


Project Flow – Question 7


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

.drop() allows to drop the company_permalink column in master_dataframe and inplace=True makes the change as permanent in the data frame.

Code –

```
master_dataframe.drop(columns='company_permalink',inplace=True)
```

Output – company permalink is dropped from its position.



	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	permalink	nan
0	/funding-round/9a01d05418af9f794eebff7ace91f638	venture	B	05/01/2015	10000000.0	/ORGANIZATION/-FAME	#fan
1	/funding-round/22dacff496eb7acb2b901dec1dfe5633	venture	A	14/10/2014	NaN	/ORGANIZATION/-QOUNTER	:Qount
2	/funding-round/b44fbb94153f6cdef13083530bb48030	seed	NaN	01/03/2014	700000.0	/ORGANIZATION/-QOUNTER	:Qount

Project Flow – Question 8

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

Total of null values for each column divided by total observations and multiplying the whole with 100 gives the null value percentages of each column.

Code –

```
(master_dataframe.isna().sum()/len(master_dataframe))*100
```

Output -

```
funding_round_permalink    0.000000
funding_round_type         0.000000
funding_round_code        72.909888
funded_at                  0.000000
raised_amount_usd         17.389905
permalink                  0.000000
name                       0.000870
homepage_url               5.336419
category_list              2.964870
status                     0.000000
country_code               7.547892
state_code                 9.520992
region                     8.843283
city                       8.840673
founded_at                17.852731
dtype: float64
```

Project Flow – Question 8

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

funding_round_code has null values more than 50% (72.90%) and it is dropped.

Code –

```
master_dataframe.drop(columns=['funding_round_code']) # deleting column with more than 50% null values
```

Output -

	funding_round_permalink	funding_round_type	funded_at	raised_amount_usd	permalink	name	homepage.
0	/funding-round/9a01d05418af9f794eebf7ace91f638	venture	05/01/2015	10000000.0	/ORGANIZATION/-FAME	#fame	http://livfame.c
1	/funding-round/22dacff496eb7acb2b901dec1dfe5633	venture	14/10/2014	NaN	/ORGANIZATION/-QOUNTER	.Qounter	http://www.qounter.c
2	/funding-round/b44fbb94153f6cdef13083530bb48030	seed	01/03/2014	700000.0	/ORGANIZATION/-QOUNTER	.Qounter	http://www.qounter.c
3	/funding-round/650b8f704416801069bb178a1418776b	venture	30/01/2014	3406878.0	/ORGANIZATION/-THE-ONE-OF-THEM-INC-	(THE) ONE of THEM, Inc.	http://oneofther
4	/funding-round/5727acc8aa57461bd22a9bdd945382d	venture	19/03/2008	2000000.0	/ORGANIZATION/0-6-COM	0-6.com	http://www.0-6.c
...
114941	/funding-round/8f6d25b8ee4199e586484d817bcda05	convertible_note	01/03/2014	41313.0	/ORGANIZATION/ZZZZAPP-COM	Zzzzapp Wireless Ltd.	http://www.zzzzapp.c
114942	/funding-round/f1aa06ed5da186c84f101549035d4ae	seed	01/05/2013	32842.0	/ORGANIZATION/ZZZZAPP-COM	Zzzzapp Wireless Ltd.	http://www.zzzzapp.c
114943	/funding-round/59f4dce44723b794f21ded3daed6e4fe	venture	01/08/2014	NaN	/ORGANIZATION/ÄÄERON	ÄÄERON	http://www.aeron
114944	/funding-round/35f09d0794651719b02bbfd859ba9ff5	seed	01/01/2015	18192.0	/ORGANIZATION/ÄÄASYS-2	ÄÄasys	http://www.oasys
114945	/funding-round/af942669878d2cd788ef5189b435ebc4	grant	01/10/2013	14851.0	/ORGANIZATION/ÄÄNOVATIFF-REKLAM-VE-TANÄÄTÄM-H...	ÄÄnovatiff Reklam ve TanÄÄtÄsm Hizmetleri Tic	http://inovatiff.c

Project Flow – Question 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.

Using pivot table, mean and median of each funding_round_type can be displayed

Code –

```
pd.pivot_table(rounds, index=['funding_round_type'], values=['raised_amount_usd'], aggfunc=['mean', 'median'])
```

Output -

funding_round_type	mean	median
	raised_amount_usd	raised_amount_usd
angel	9.586945e+05	400000.0
convertible_note	1.453439e+06	272000.0
debt_financing	1.704353e+07	1100000.0
equity_crowdfunding	5.383682e+05	100000.0
grant	4.300576e+06	201684.0
non_equity_assistance	4.112031e+05	60000.0
post_ipo_debt	1.687046e+08	19950000.0
post_ipo_equity	8.218249e+07	12262852.5
private_equity	7.330859e+07	20000000.0
product_crowdfunding	1.363131e+06	183915.0
secondary_market	7.964963e+07	32600000.0
seed	7.198180e+05	275000.0
undisclosed	1.924237e+07	1018680.0
venture	1.174895e+07	5000000.0

Project Flow – Question 10

- Filter the master_dataframe on 'raised_amount_usd' such as 'raised_amount_usd' lie between 4 Millions to 15 Millions.

The master_dataframe is filtered based on both the conditions by '&' operator

Code –

```
master_dataframe[(master_dataframe['raised_amount_usd']>=4000000) & (master_dataframe['raised_amount_usd']<=15000000)]
```

Output -

	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	permalink	n
0	/funding-round/9a01d05418af9f794eebf7ace91f638	venture	B	05/01/2015	10000000.0	/ORGANIZATION/-FAME	#
12	/funding-round/e1cfcbe1bdf4c70277c5f29a3482f24e	venture	A	19/07/2014	8900000.0	/ORGANIZATION/0XDATA	Hz
21	/funding-round/11c228f58831bc7ed337ef69ecc560c2	private_equity	NaN	01/02/2015	6000000.0	/ORGANIZATION/1-800-PUBLICRELATIONS-INC-	1-PublicRelat
22	/funding-round/b952cbaf401f310927430c97b68162ea	venture	NaN	17/03/2015	5000000.0	/ORGANIZATION/1-MAINSTREAM	1 Mainsti
28	/funding-round/0facbbcc5818dc5326469f13f5a8ac8	venture	A	09/10/2014	4000000.0	/ORGANIZATION/10-MINUTES-WITH	10 Minutes
...
114918	/funding-round/85443f5438bfd50104b9591eabc95c94	grant	NaN	26/03/2014	7500000.0	/ORGANIZATION/ZYOMYX-INC	ZYO
114919	/funding-round/b4fce7f8a13267353d57db09099800e8	venture	NaN	11/07/2013	14219999.0	/ORGANIZATION/ZYOMYX-INC	ZYO
114927	/funding-round/c05aab2155e390066d51c6852ff61464	grant	NaN	11/09/2009	5400000.0	/ORGANIZATION/ZYRAZ-TECHNOLOGY	2 Techno
114928	/funding-round/e949f077c943db197f20e43c2eb8d0e6	venture	A	09/10/2009	7991547.0	/ORGANIZATION/ZYRAZ-TECHNOLOGY	2 Techno
114933	/funding-round/7ea37c0ff4132b084832039a9380d34e	venture	A	16/12/2004	8500000.0	/ORGANIZATION/ZYSTOR	Z

22213 rows × 15 columns

Project Flow – Question 11

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

Code – New data frame was created by merging rounds and company data frames.

```
rounds['company_permalink']=rounds['company_permalink'].str.upper()
company['permalink']=company['permalink'].str.upper()
df=pd.merge(left=rounds,right=company,left_on='company_permalink',right_on='permalink',how='inner')
df
```

Output -

	company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	
0	/ORGANIZATION/-FAME	round/9a01d05418af9f794eebf7ace91f638	venture	B	05/01/2015	10000000.0	/ORG
1	/ORGANIZATION/-QOUNTER	round/22dacff496eb7acb2b901dec1dfe5633	venture	A	14/10/2014	NaN	
2	/ORGANIZATION/-QOUNTER	round/b44fbb94153f6cdef13083530bb48030	seed	NaN	01/03/2014	700000.0	
3	/ORGANIZATION/-THE-ONE-OF-THEM-INC-	round/650b8f704416801069bb178a1418776b	venture	B	30/01/2014	3406878.0	/ORC O
4	/ORGANIZATION/0-6-COM	round/5727accacaa57461bd22a9bdd945382d	venture	A	19/03/2008	2000000.0	/ORGAN
...	
114941	/ORGANIZATION/ZZZZAPP-COM	round/8f6d25b8ee4199e586484d817bcdad05	convertible_note	NaN	01/03/2014	41313.0	/ORGANI
114942	/ORGANIZATION/ZZZZAPP-COM	round/ff1aa06ed5da186c84f101549035d4ae	seed	NaN	01/05/2013	32842.0	/ORGANI
114943	/ORGANIZATION/ÄÄERON	round/59f4dce44723b794f21ded3daed6e4fe	venture	A	01/08/2014	NaN	/ORGAN
114944	/ORGANIZATION/ÄÄSYS-2	round/35f09d0794651719b02bbfd859ba9ff5	seed	NaN	01/01/2015	18192.0	/ORGANI
114945	/ORGANIZATION/ÄÄNOVATIFF-REKLAM-VE-TANÄÄTÄÄM-H...	round/af942869878d2cd788ef5189b435ebc4	grant	NaN	01/10/2013	14851.0	ÄÄNOVA

Project Flow – Question 11

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

Code – Pivot table is used to the maximum amount based on each funding round type and each country. .sort_values() are used to obtain on venture’s top 9 countries.

```
pd.pivot_table(df, index=['funding_round_type', 'country_code'], values=['raised_amount_usd'], aggfunc='max').sort_values(
    by=['funding_round_type', 'raised_amount_usd'], ascending=[False, False]).head(9)
```

Output -

		raised_amount_usd
funding_round_type	country_code	
venture	USA	1.760000e+10
	CHN	1.100000e+09
	IND	1.000000e+09
	GBR	9.150000e+08
	NLD	7.500000e+08
	SWE	5.260000e+08
	FRA	5.000000e+08
	SGP	3.500000e+08
	NOR	3.324638e+08

Project Flow – Question 12

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

Code – Main category should contain the first word before |. Lambda is used to apply split('|') in all the observations of category_list.

```
df['main_category']=df['category_list'].astype(str).apply(lambda x:x.split('|')[0] if '|' in x else x)
df
```

Output

permalink	name	homepage_url	category_list	status	country_code	state_code	region	city	founded_at	main_category
GANIZATION/-FAME	#fame	http://livfame.com	Media	operating	IND	16	Mumbai	Mumbai	NaN	Media
/ORGANIZATION/-QOUNTER	:Qounter	http://www.qounter.com	Application Platforms Real Time Social Network...	operating	USA	DE	DE - Other	Delaware City	04-09-2014	Application Platforms
/ORGANIZATION/-QOUNTER	:Qounter	http://www.qounter.com	Application Platforms Real Time Social Network...	operating	USA	DE	DE - Other	Delaware City	04-09-2014	Application Platforms
ORGANIZATION/-THE-ONE-OF-THEM-INC-	(THE) ONE of THEM,Inc.	http://oneofthem.jp	Apps Games Mobile	operating	NaN	NaN	NaN	NaN	NaN	Apps

Project Flow – Question 13

- Using the user defined function convert the 'raised_amount_usd' 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.

Code – A user defined function was created including all the conditions. Rows with null values in raised_amount_usd are dropped to avoid the condition being applied to the null values. The user defined function is then applied to get new column 'amount_category'.

```
#13. Using the user defined function convert the 'raised_amount_usd' 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.  
  
def convert_num_to_cat(n):  
    if n<float(5000000):  
        return 'TypeA'  
    elif n>=float(5000000) and n<float(7000000):  
        return 'TypeB'  
    else:  
        return 'TypeC'
```

```
3]: rounds.dropna(axis=0,inplace=True)
```

```
1]: rounds['amount_category']=rounds['raised_amount_usd'].apply(convert_num_to_cat)
```

```
2]: rounds
```

Project Flow – Question 13

- Using the user defined function convert the 'raised_amount_usd' 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.

Output -

company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	amount_category
/organization/-fame	round/9a01d05418af9f794eebf7ace91f638	venture	B	05-01-2015	10000000.0	TypeC
/ORGANIZATION/-THE-ONE-OF-THEM-INC-	round/650b8f704416801069bb178a1418776b	venture	B	30-01-2014	3406878.0	TypeA
/organization/0-6-com	round/5727accaaaa57461bd22a9bdd945382d	venture	A	19-03-2008	2000000.0	TypeA
/organization/0xdata	round/3bb2ee4a2d89251a10aaa735b1180e44	venture	B	09-11-2015	20000000.0	TypeC
/organization/0xdata	round/e1cfcbe1bdf4c70277c5f29a3482f24e	venture	A	19-07-2014	8900000.0	TypeC
...
/organization/zyomyx-inc	round/4191cb5258d3863b0212d1083881da64	venture	B	30-06-2013	12000000.0	TypeC
/ORGANIZATION/ZYRAZ-TECHNOLOGY	round/e949f077c943db197f20e43c2eb8d0e6	venture	A	09-10-2009	7991547.0	TypeC
/organization/zystor	round/7ea37c0ff4132b084832039a9380d34e	venture	A	16-12-2004	8500000.0	TypeC
ORGANIZATION/ZYTOPROTEC	round/0c43e717298296d2fb334fb236300f6e	venture	A	29-01-2013	2686600.0	TypeA
/ORGANIZATION/ZZNODE-SCIENCE-AND-TECHNOLOGY-CO...	round/171693e1e07e0230defd8fddb6de4515	venture	A	01-04-2012	1587301.0	TypeA

× 7 columns

Major Challenge

Q1: Determine what percentage of the total funding was raised through each round type

Total of values for each column divided by total observations in raised_amount_usd and multiplying the whole with 100 gives the total funding percentages of each column.

Code :

```
In [479]: #Determine what percentage of the total funding was raised through each round type
total_amount=rounds['raised_amount_usd'].sum()
total_amount
```

```
(rounds.groupby('funding_round_type')['raised_amount_usd'].sum()/total_amount)*100
```

Output:

```
funding_round_type
angel                0.007100
convertible_note     0.002802
debt_financing       0.013881
post_ipo_equity      0.001409
private_equity       0.034034
seed                 0.070964
undisclosed          0.013944
venture              99.855866
Name: raised_amount_usd, dtype: float64
```

Major Challenge

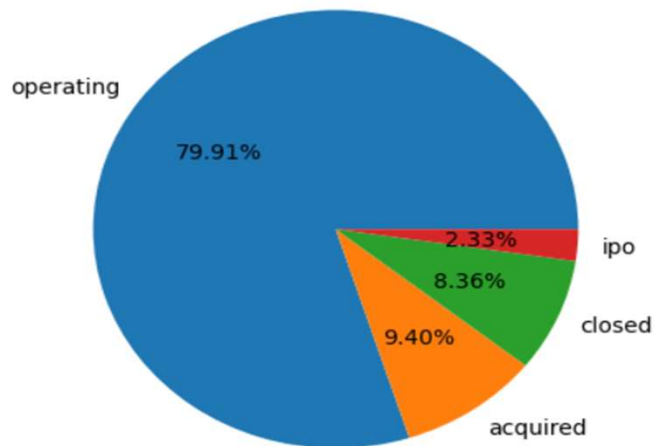
Q2: Visualize the distribution of company statuses (e.g., operating, closed) using a pie chart or bar plot.

Code :

```
#Visualize the distribution of company statuses (e.g., operating, closed) using a pie chart or bar plot.  
plt.pie(x=company['status'].value_counts(), labels=company['status'].unique(), autopct='%.2f%')  
plt.title('Distribution of company statuses')  
plt.show()
```

Distribution of company statuses

Output:



Operating companies are higher than companies with other statuses. Ipo status is the lowest in the distribution

Major Challenge

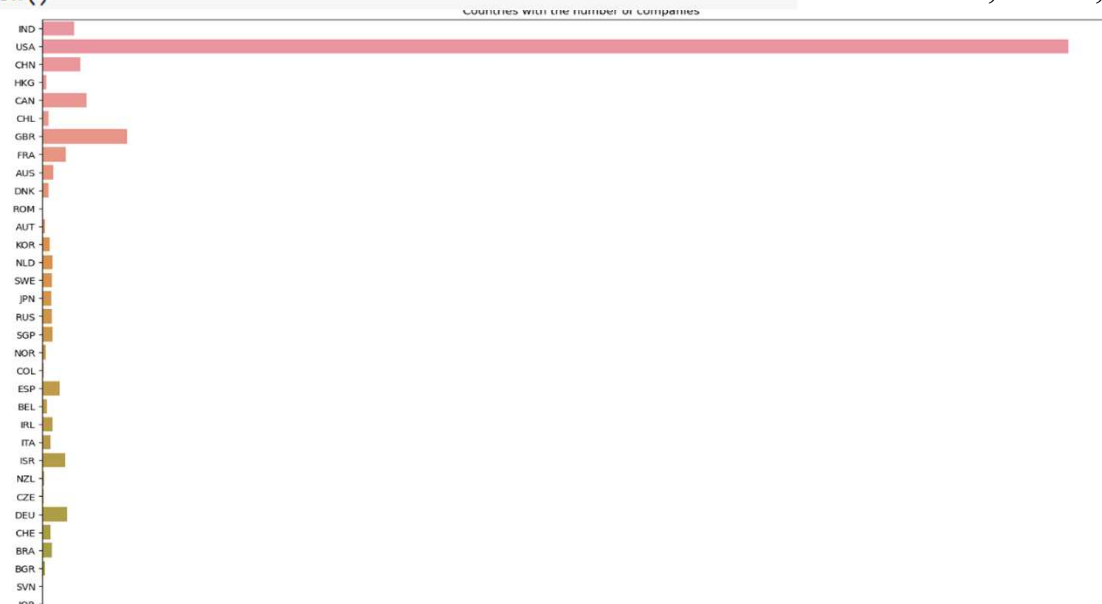
Q3: Visualize the countries with more number of companies in the master_dataframe

Code:

```
plt.figure(figsize=(20,50))
sns.countplot(data=master_dataframe, y='country_code')
plt.xlabel('Number of companies')
plt.ylabel('Country codes')
plt.title('Countries with the number of companies')
plt.show()
```

USA has the highest number of companies followed by GBR, CAN, CHN and IND. Countries with lowest number (1) of companies are KNA, BRN, DZA, TGO, and GGY.

Output:



Major Challenge

Q4: Which country has the highest growth rate for the last decade

Data of year 2005 and 2015 were retrieved. Raised_amount_usd were summed for each country for year 2005 and 2015. Growth rate is then calculated for all countries based on the formula:

Growth rate = $((\text{amount raised in 2015} / \text{amount raised in 2005})^{(1/10)}) \times 100$

Code:

```
master_dataframe['date'] = pd.to_datetime(master_dataframe['funded_at'], format='%d/%m/%Y')
```

```
master_dataframe.sort_values(by='date', ascending=False) # Taking years from 2005-2015
```

```
master_dataframe['year'] = master_dataframe['date'].dt.year
```

```
df1 = master_dataframe[(master_dataframe['year'] == 2015) & (master_dataframe['raised_amount_usd'] > 1)]  
df1
```

```
df2 = master_dataframe[(master_dataframe['year'] == 2005) & (master_dataframe['raised_amount_usd'] > 1)]
```

```
pivot_table_2015 = pd.pivot_table(df1, index='country_code', values='raised_amount_usd', aggfunc='sum')
```

```
pivot_table_2015.rename(columns={'raised_amount_usd': 'raised_amount_2015'}, inplace=True)
```


Major Challenge

Q4: Which country has the highest growth rate for the last decade

Code:

```
pivot_table_2005=pd.pivot_table(df2,index='country_code',values='raised_amount_usd',aggfunc='sum')
pivot_table_2005.rename(columns={'raised_amount_usd':'raised_amount_2005'},inplace=True)
```

```
growth_rate_data=pd.merge(left=pivot_table_2005,right=pivot_table_2015,how='inner',on='country_code')
```

```
growth_rate_data['growth_rate']=(((growth_rate_data['raised_amount_2015']/growth_rate_data['raised_amount_2005'])**(1/10))-1)*100
```

```
growth_rate_data.sort_values(by='growth_rate',ascending=False)
```

```
growth_rate_data.sort_values(by='growth_rate',ascending=False).iloc[0]
```

Output:

	raised_amount_2005	raised_amount_2015	growth_rate
country_code			
BRA	3.110000e+06	1.499989e+09	85.495238
HKG	5.000000e+05	2.330874e+08	84.865077
RUS	5.000000e+06	1.609636e+09	78.145878
IDN	2.000000e+06	6.426125e+08	78.111482
NZL	3.000000e+06	4.969075e+08	66.692259
ZAF	3.307000e+06	3.925292e+08	61.229231
IND	1.118600e+08	1.073934e+10	57.844798

```
raised_amount_2005    3.110000e+06
raised_amount_2015    1.499989e+09
growth_rate           8.549524e-01
Name: BRA, dtype: float64
```

Country BRA has the highest growth rate in the past decade

Major Challenge

Q5: Find the companies that have been to more funding rounds and arrange them based on the companies, dates, and round code.

Companies with more than 1 value counts are filtered. Funded_at is converted to a proper date format. Then, the companies were sorted based on the company permalink, date, and funding round code.

Code:

```
in [26]: rounds['company_permalink'].value_counts()
value_counts=rounds['company_permalink'].value_counts()
more_than1=value_counts[value_counts>1].index.tolist()
more_than1

: df=rounds[rounds['company_permalink'].isin(more_than1)]

df['date']=pd.to_datetime(df['funded_at'],format='%d/%m/%Y')

df.sort_values(by=['company_permalink','date','funding_round_code'], ascending=[True,True,True])
```


Major Challenge

Q5: Find the companies that have been to more funding rounds and arrange them based on the companies, dates, and round code.

Output:

	company_permalink	funding_round_permalink	funding_round_type	funding_round_code	funded_at	raised_amount_usd	amount_usd
12	/ORGANIZATION/OXDATA	round/e1cfcbe1bdf4c70277c5f29a3482f24e	venture	A	19/07/2014	8900000.0	
10	/ORGANIZATION/OXDATA	round/3bb2ee4a2d89251a10aaa735b1180e44	venture	B	09/11/2015	20000000.0	
49	/ORGANIZATION/1001-MENUS	round/b6c28ea4ebe32db7083052cf87e7c368	venture	A	25/10/2013	1800000.0	
47	/ORGANIZATION/1001-MENUS	round/8025123b90c1ab687f9857ffe18bfd3b	venture	A	13/11/2013	1736910.0	
46	/ORGANIZATION/1001-MENUS	round/645b4cd895450b8e0268027cd7813047	venture	B	04/11/2015	6602694.0	
...
114910	/ORGANIZATION/ZYNGA	round/8411431b5a0389a874e3fae93016c089	venture	B	24/04/2010	15000000.0	
114912	/ORGANIZATION/ZYNGA	round/cf8299f657f7fd4cf01054ee1934109c	venture	B	14/06/2010	300000000.0	
114911	/ORGANIZATION/ZYNGA	round/a6ebc53cf3c0c97d1cdd075f6bbf31fa	venture	C	18/02/2011	490000000.0	
114917	/ORGANIZATION/ZYNSTRA	round/129127802dd48be42d6cd45f65895245	venture	A	12/09/2013	3800000.0	

Conclusions

- We learned and practiced visualization techniques such as barplot and pie chart. We were able to examine the performances and get the data from the asset management datasets by applying npv codes such as crosstab, sortvalue, and pivot tables. analyze company trends and learn about investment strategies
- More techniques, such as null value analysis and treatments can be used in asset management data sets.
- We got to apply pivot tables to perform country-wise and funding round type-wise analyses and use visualization techniques on the data.
- We also learned about investing in companies based on the countries that received most of the investments and how growth rate also plays an important role in investing when we focus on the funded years.

Thank You !!!....