

# Dropping not available values in the columns "InvestorsName" and "StartupName" using dropna() method

Jupyter 3rd Last Checkpoint: 19 hours ago (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)

Run Code

[2372 rows x 10 columns]

```
In [23]: df.dropna(subset=["InvestorsName", "StartupName"], inplace=True)
df.reset_index(drop = True, inplace = True)
print(df)
```

	SNo	Date	StartupName	IndustryVertical	\
0	0	01/08/2017	TouchKin	Technology	
1	1	02/08/2017	Ethinos	Technology	
2	2	02/08/2017	Leverage Edu	Consumer Internet	
3	3	02/08/2017	Zepo	Consumer Internet	
4	4	02/08/2017	Click2Clinic	Consumer Internet	
...	...	...	...	...	...
2359	2367	29/01/2015	Printvenue	NaN	
2360	2368	29/01/2015	Graphene	NaN	
2361	2369	30/01/2015	Mad Street Den	NaN	
2362	2370	30/01/2015	Simplotel	NaN	
2363	2371	31/01/2015	couponmachine.in	NaN	

		SubVertical	CityLocation	\
0		Predictive Care Platform	Bangalore	
1		Digital Marketing Agency	Mumbai	
2	Online platform for Higher Education Services		New Delhi	
3	DIY Ecommerce platform		Mumbai	
4	healthcare service aggregator		Hyderabad	
...	...	...	...	...
2359		NaN	NaN	
2360		NaN	NaN	
2361		NaN	NaN	
2362		NaN	NaN	
2363		NaN	NaN	

		InvestorsName	InvestmentType	\
0		Kae Capital	Private Equity	
1		Triton Investment Advisors	Private Equity	
2	Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...		Seed Funding	
3	Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...		Seed Funding	
4	Narottam Thudi, Shireesh Palle		Seed Funding	
...	...	...	...	...

# Correcting the spelling mistake of startup names and ignoring "undisclosed investors" in column "InvestorsName"

jupyter 3rd Last Checkpoint: 19 hours ago (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel) O

Run

```
In [29]: df["StartupName"].replace("Flipkart.com", "Flipkart", inplace=True)
df["StartupName"].replace("Ola Cabs", "Ola", inplace=True)
df["StartupName"].replace("OlaCabs", "Ola", inplace=True)
df["StartupName"].replace("OlaCabs", "Ola", inplace=True)
df["StartupName"].replace("Oyo Rooms", "Oyo", inplace=True)
df["StartupName"].replace("OyoRooms", "Oyo", inplace=True)
df["StartupName"].replace("Oyorooms", "Oyo", inplace=True)
df["StartupName"].replace("OYO Rooms", "Oyo", inplace=True)
df["StartupName"].replace("Paytm Marketplace", "Paytm", inplace=True)
df = df[df.InvestorsName != 'Undisclosed Investors']
df = df[df.InvestorsName != 'Undisclosed investors']
df = df[df.InvestorsName != 'undisclosed investors']
df = df[df.InvestorsName != 'undisclosed investor']
print(df.head())
```

	SNo	Date	StartupName	IndustryVertical \
0	0	01/08/2017	TouchKin	Technology
1	1	02/08/2017	Ethinos	Technology
2	2	02/08/2017	Leverage Edu	Consumer Internet
3	3	02/08/2017	Zepo	Consumer Internet
4	4	02/08/2017	Click2Clinic	Consumer Internet

		SubVertical	CityLocation \
0		Predictive Care Platform	Bangalore
1		Digital Marketing Agency	Mumbai
2	Online platform for Higher Education Services		New Delhi
3		DIY Ecommerce platform	Mumbai
4		healthcare service aggregator	Hyderabad

		InvestorsName	InvestmentType \
0		Kae Capital	Private Equity
1		Triton Investment Advisors	Private Equity
2	Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...		Seed Funding
3	Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...		Seed Funding
4		Narottam Thudi, Shireesh Palle	Seed Funding

	AmountInUSD	Remarks
0	1,300,000	NaN
1	NaN	NaN

Used iterrow() method to iterate through the each rows of the dataframe and then splitting the investors names and append Them to investor\_names list and also corresponding startup name to startup\_name list and then created a new Data frame having "InvestorsName" and "StartupName" columns and then using nunique() method to get the count of Unique values in column "StartupName" for each investorsName and then sorted the dataframe by values in descending order

jupyter 3rd Last Checkpoint: 19 hours ago (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)

Run Code

```
0 1,300,000 NaN
1      NaN NaN
2      NaN NaN
3   500,000 NaN
4   850,000 NaN
```

```
In [64]: investor_name=[]
startup_name=[]

for index,row in df.iterrows():
    i=row["InvestorsName"]
    s=row["StartupName"]

    i=str(i)
    temp=i.split(',')
    for j in temp:
        if(j!=''):
            investor_name.append(j)
            startup_name.append(s)

newDf=pd.DataFrame({"InvestorsName":investor_name,"StartupName":startup_name})
newDf=newDf.groupby('InvestorsName')['StartupName'].nunique()
newDf=newDf.sort_values(ascending=False).head(5)
print(newDf)
```

```
InvestorsName
Sequoia Capital      34
Accel Partners       33
Indian Angel Network 33
Kalaari Capital      30
Ratan Tata           28
Name: StartupName, dtype: int64
```

## Top 5 Investors who have invested in different startups

jupyter 3rd Last Checkpoint: 19 hours ago (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)

Run Code

```
Kalaari Capital      30
Ratan Tata           28
Name: StartupName, dtype: int64
```

```
In [65]: plt.pie(newDf.values, labels=newDf.index, autopct="%.2f%%")
plt.axis("equal")
plt.show()
```

