Journey of Startups in India

Author- Sanjay Chauhan

Date- May 25, 2022

Introduction

The project is regarding the recent wave of startups in the country, as we are witnessing a continuous rise in the startups and many of them are competing in the world stages. In India, several cities are becoming startup hubs and giving healthy competition to each other. But this also raises many questions such as "When exactly does this startup wave starts to take off?" "Which are the top-rated cities with the startup-friendly environment?" or "Which are the most common or rare sectors to start a startup among all?" and many others properly described and answered as briefly as possible through visualization and analysis boards. This report contains two phases first is **Data cleaning & manipulation** to maintain the integrity of the data and it has been made visible to ensure transparency in the analysis. The second phase is the **Analysis phase**. The timeline of data is from the year 1963 till 2021.

In [1]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
```

Now, let's see the data

In [2]:
```
df=pd.read_csv("Indian Startup.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	Company/Brand	Founded	Headquarters	Sector	What it does	Founder/s	Investor/s	
0	CollegeDekho	2015.0	Gurgaon	E-learning	Collegedekho.com is Student's Partner, Friend ...	Ruchir Arora	Disrupt ADQ, QIC	$35
1	BOX8	2012.0	Mumbai	Food & Beverages	India's Largest Desi Meals Brand	Anshul Gupta, Amit Raj	Tiger Global	$40
2	Simpl	2015.0	Bangalore	Consumer Services	Simpl empowers merchants to build trusted rela...	Nitya, Chaitra Chidanand	Valar Ventures, IA Ventures	$40
3	8i Ventures	2018.0	Mumbai	Venture Capital & Private Equity	8i is a Mumbai & Bangalore based early stage f...	Vikram Chachra	NaN	$50
4	PayGlocal	2021.0	Bangalore	Financial Services	PayGlocal is a FinTech solving for global paym...	Prachi Dharani, Rohit Sukhija, Yogesh Lokhande	Sequoia Capital India	$4

An overview of data and datatypes of its attributes

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1209 entries, 0 to 1208
Data columns (total 10 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Company/Brand  1209 non-null   object 
 1   Founded        1208 non-null   float64
 2   Headquarters   1208 non-null   object 
 3   Sector         1209 non-null   object 
 4   What it does   1209 non-null   object 
 5   Founder/s      1205 non-null   object 
 6   Investor/s     1146 non-null   object 
 7   Amount         1202 non-null   object 
 8   Stage          781 non-null    object 
 9   Date           1209 non-null   object 
dtypes: float64(1), object(9)
memory usage: 94.6+ KB
```

Data cleaning process

Now, if we look at the datatypes of the "Amount column" it is showing it as a string. So we first need to rectify it as numeric and column names need to be consistent as well as should be in lcase format to make it convenient while writing codes and functions.

Step-1 Consistency in column names

In [5]:
```
df.columns= [x.lower() for x in df.columns]
```

In [6]:
```
df.columns=df.columns.str.replace(" ","_")
```

In [7]:
```
df.rename(columns={"investor/s":"investor","founder/s":"founder","company/brand":"comp
```

In [8]:
```
df.head()
```

Out[8]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amou
0	CollegeDekho	2015.0	Gurgaon	E-learning	Collegedekho.com is Student's Partner, Friend ...	Ruchir Arora	Disrupt ADQ, QIC	$35,000,0
1	BOX8	2012.0	Mumbai	Food & Beverages	India's Largest Desi Meals Brand	Anshul Gupta, Amit Raj	Tiger Global	$40,000,0
2	Simpl	2015.0	Bangalore	Consumer Services	Simpl empowers merchants to build trusted rela...	Nitya, Chaitra Chidanand	Valar Ventures, IA Ventures	$40,000,0
3	8i Ventures	2018.0	Mumbai	Venture Capital & Private Equity	8i is a Mumbai & Bangalore based early stage f...	Vikram Chachra	NaN	$50,000,0
4	PayGlocal	2021.0	Bangalore	Financial Services	PayGlocal is a FinTech solving for global paym...	Prachi Dharani, Rohit Sukhija, Yogesh Lokhande	Sequoia Capital India	$4,900,0

Step-2 Dealing with the inconsistencies in columns' elements

Removal of "," & "$" and replacing "Undisclosed" with "0" from the "amount" column to turn them into numeric datatypes from string for calculation purposes in the analysis phase.

In [9]:
```
df.amount=df.amount.str.replace(",","")
df.amount=df.amount.str.replace("$","")
df.amount=df.amount.str.replace("Undisclosed","0")
df.amount=df.amount.str.replace("undisclosed","0")
```

```
C:\Users\chauhan\AppData\Local\Temp\ipykernel_8824\3462054536.py:2: FutureWarning: Th
e default value of regex will change from True to False in a future version. In addit
ion, single character regular expressions will *not* be treated as literal strings wh
en regex=True.
  df.amount=df.amount.str.replace("$","")
```

Changing the month name from string pattern to the month number pattern to convert it into 'Datetime' datatypes.

In [10]:
```
df.funding_date=df.funding_date.str.replace("Dec","12")
df.funding_date=df.funding_date.str.replace("Nov","11")
df.funding_date=df.funding_date.str.replace("Oct","10")
df.funding_date=df.funding_date.str.replace("Sep","09")
df.funding_date=df.funding_date.str.replace("Aug","08")
df.funding_date=df.funding_date.str.replace("Jul","07")
df.funding_date=df.funding_date.str.replace("Jun","06")
df.funding_date=df.funding_date.str.replace("May","05")
df.funding_date=df.funding_date.str.replace("Apr","04")
df.funding_date=df.funding_date.str.replace("Mar","03")
df.funding_date=df.funding_date.str.replace("Feb","02")
df.funding_date=df.funding_date.str.replace("Jan","01")
```

In [11]:
```
df.head()
```

Out[11]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amount
0	CollegeDekho	2015.0	Gurgaon	E-learning	Collegedekho.com is Student's Partner, Friend ...	Ruchir Arora	Disrupt ADQ, QIC	3500000(
1	BOX8	2012.0	Mumbai	Food & Beverages	India's Largest Desi Meals Brand	Anshul Gupta, Amit Raj	Tiger Global	4000000(
2	Simpl	2015.0	Bangalore	Consumer Services	Simpl empowers merchants to build trusted rela...	Nitya, Chaitra Chidanand	Valar Ventures, IA Ventures	4000000(
3	8i Ventures	2018.0	Mumbai	Venture Capital & Private Equity	8i is a Mumbai & Bangalore based early stage f...	Vikram Chachra	NaN	5000000(
4	PayGlocal	2021.0	Bangalore	Financial Services	PayGlocal is a FinTech solving for global paym...	Prachi Dharani, Rohit Sukhija, Yogesh Lokhande	Sequoia Capital India	490000(

Step-3 Dealing with NAs

Filling NAs with " " to check the NA values or rows and to deal with them according to the need of the hour.

In [12]:
```
df.fillna(" ",inplace=True)
```

But we cannot replace NAs of the "amount" column with " " otherwise it will throw an error while changing datatypes as we cannot mix strings and integers in the same column. For that purpose, we are replacing them with "0".

In [13]:
```
df.amount=df.amount.str.replace(" ","0")
```

Step-4 Rectifying datatypes

In [14]:
```
df.amount=pd.to_numeric(df.amount)
```

In [15]:
```
df.funding_date=pd.to_datetime(df.funding_date,format="%m-%Y")
```

In [16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1209 entries, 0 to 1208
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype         
---  ------        --------------  -----         
 0   company       1209 non-null   object        
 1   founded       1209 non-null   object        
 2   headquarters  1209 non-null   object        
 3   sector        1209 non-null   object        
 4   what_it_does  1209 non-null   object        
 5   founder       1209 non-null   object        
 6   investor      1209 non-null   object        
 7   amount        1209 non-null   int64         
 8   stage         1209 non-null   object        
 9   funding_date  1209 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(1), object(8)
memory usage: 94.6+ KB
```

Note- We can see in the above chunk that "founded" column still has a string datatype and it needs to be rectified as int.But we cannot do it as of now because some cells in the column have blanks. We can rectify datatype only after filling those blank cells.So, we will do that after researching and filling or dropping those blank cells.

Step-5 Checking and handling duplicate rows

In [17]:
```
df.shape
```

Out[17]:
```
(1209, 10)
```

In [18]:
```
df.duplicated().sum()
```

Out[18]:
```
19
```

In [19]:
```
df.loc[df.duplicated(),:]
```

Out[19]:

	company	founded	headquarters	sector	what_it_does	founder	investor
31	Trinkerr	2021.0	Bangalore	Capital Markets	Trinkerr is India's first social trading platf...	Manvendra Singh, Gaurav Agarwal	Accel India
32	Zorro	2021.0	Gurugram	Social network	Pseudonymous social network platform	Jasveer Singh, Abhishek Asthana, Deepak Kumar	Vijay Shekhar Sharma, Ritesh Agarwal, Ankiti Bose
33	Ultraviolette	2021.0	Bangalore	Automotive	Create and Inspire the future of sustainable u...	Subramaniam Narayan, Niraj Rajmohan	TVS Motor, Zoho
34	NephroPlus	2009.0	Hyderabad	Hospital & Health Care	A vision and passion of redefining healthcare ...	Vikram Vuppala	IIFL Asset Management
35	Unremot	2020.0	Bangalore	Information Technology & Services	Unremot is a personal office for consultants!	Shiju Radhakrishnan	Inflection Point Ventures
36	FanAnywhere	2021.0	Bangalore	Financial Services	Celebrity NFT platform	Varun Chaudhary, Amit Kumar	Oasis Capital, Scorpio VC, DeltaHub Capital
37	PingoLearn	2021.0	Pune	E-learning	PingoLearn offers language learning courses wi...	Mohit Menghani, Shubham Maheshwari	Titan Capital, Haresh Chawla, AngelList Syndicate
38	Spry	2021.0	Mumbai	Music	THE BEST ROYALTY-FREE MUSIC TO CREATE TRENDING...	Gaurav Dagaonkar, Meghna Mittal	9Unicorns, Ashneer Grover
39	Enmovil	2015.0	Hyderabad	Information Technology & Services	Enmovil delivers a Business Intelligence Platf...	Ravi Bulusu, Nanda Kishore, Venkat Moganty	Anicut Angel Fund
40	ASQI Advisors	2019.0	Mumbai	Financial Services	Bringing Blockchain technology intro mainstrea...	Swapnil Pawar	Founders Room Capital
41	Insurance Samadhan	2018.0	New Delhi	Insurance	The insurance industry will help in representa...	Deepak Bhuvenshwari Uniyal	9Unicorns, ZNL

	company	founded	headquarters	sector	what_it_does	founder	investor
42	Evenflow Brands	2020.0	Mumbai	Consumer Goods	Identify and partner with the upcoming eCommer...	Utsav Agarwal, Pulkit Chhabra	
43	MasterChow	2020.0	Food & Beverages	Hauz Khas	A ready-to-cook Asian cuisine brand	Vidur Kataria, Sidhanth Madan	WEH Ventures
44	Fullife Healthcare	2009.0	Pharmaceuticals #REF!	Primary Business is Development and Manufactur...	Varun Khanna	Morgan Stanley Private Equity Asia	$22,000,000
464	Curefoods	2020.0	Bangalore	Food & Beverages	Healthy & nutritious foods and cold pressed ju...	Ankit Nagori	Iron Pillar, Nordstar, Binny Bansal
466	Bewakoof	2012.0	Mumbai	Apparel & Fashion	Bewakoof is a lifestyle fashion brand that mak...	Prabhkiran Singh	InvestCorp
468	FanPlay	2020.0	Computer Games	Computer Games	A real money game app specializing in trivia g...	YC W21	Pritesh Kumar, Bharat Gupta, Upsparks
474	Advantage Club	2014.0	Mumbai	HRTech	Advantage Club is India's largest employee eng...	Sourabh Deorah, Smiti Bhatt Deorah	Y Combinator, Broom Ventures, Kunal Shah
476	Ruptok	2020.0	New Delhi	FinTech	Ruptok fintech Pvt. Ltd. is an online gold loa...	Ankur Gupta	Eclear Leasing

In [20]:
```
df.drop_duplicates(keep="first",inplace=True)
```

In [21]:
```
df.shape
```

Out[21]:
```
(1190, 10)
```

Step-6 Dealing with the missing data or NAs whom we replaced with the blank cells in step-3

In [22]:
```
df[df.founded==" "]
```

Out[22]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amount	stage
494	Smart Express		Mumbai	Logistics	India's Most Innovative and Awarded Express Lo...	Yogesh Dhingra	IIFL India Private Equity Fund, Smiti Holding ...	10000000	Seed

In [23]:
```
df.at[494,"founded"]=2018
```

Now, after filling the gap in the "founded" column its ready to get rectified as numeric from string datatypes.

In [24]:
```
df.founded=pd.to_numeric(df.founded)
```

In [25]:
```
df.loc[494]
```

Out[25]:
```
company                                              Smart Express
founded                                                     2018.0
headquarters                                                Mumbai
sector                                                   Logistics
what_it_does    India's Most Innovative and Awarded Express Lo...
founder                                                Yogesh Dhingra
investor        IIFL India Private Equity Fund, Smiti Holding ...
amount                                                   10000000
stage                                                        Seed
funding_date                                  2021-08-01 00:00:00
Name: 494, dtype: object
```

Now, let's have an altogether estimate of blank cells. For that purpose we will create a new dataframe with blank rows of original data.

In [26]:
```
empty_cells=df[(df.company==" ")|(df.headquarters==" ")|(df.sector==" ")|(df.what_it_c
```

In [27]:
```
empty_cells.head()
```

Out[27]:

	company	founded	headquarters	sector	what_it_does	founder	investor	am
1	BOX8	2012.0	Mumbai	Food & Beverages	India's Largest Desi Meals Brand	Anshul Gupta, Amit Raj	Tiger Global	4000
3	8i Ventures	2018.0	Mumbai	Venture Capital & Private Equity	8i is a Mumbai & Bangalore based early stage f...	Vikram Chachra		5000
5	Curefit	2016.0	Bangalore	Health, Wellness & Fitness	Tata Digital-backed Curefit	Mukesh Bansal, Ankit	Zomato	14500
7	CHARGE+ZONE	2018.0	Vadodara	Automotive	CHARGE+ZONE is a tech-driven EV Charging infra...	Kartikey Hariyani	Venture Catalysts	1000
13	Probus Insurance	2002.0	Mumbai	Insurance	Probus Insurance is a leading InsurTech platfo...	Rakesh goyal	BlueOrchard Impact Investment Managers	670

Let's check the blank cells of particular columns to have a better understanding.

In [28]:
```
df.stage.value_counts()
```

Out[28]:
```
                421
Seed            243
Pre-series A    146
Series A        125
Series B         49
Pre-seed         47
Series C         45
Debt             27
Series D         21
Series E         17
Pre-series B     10
Pre-Series A      9
Series F          6
Pre-series A1     4
Pre-series        2
Series G          2
Bridge            2
Series H          2
Series D1         1
PE                1
Series B3         1
Seies A           1
$6,000,000        1
Series A2         1
Series A+         1
Series F2         1
Seed+             1
Pre-Seed          1
Series I          1
Series F1         1
Name: stage, dtype: int64
```

If we look at the "stage" column element count we can see that there is a stage as "$6,000,000" which means the values of the "amount" column got interchanged with the "stage" column and we have to clean all those sort of errors. We can also see that it has 428 blank cells.

Rectification of interchanged values is below-

In [29]:
```
df[df.stage=="$6,000,000"]
```

Out[29]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amount	
784	MYRE Capital	2020.0	Mumbai	Commercial Real Estate	Democratising Real Estate Ownership	Own rent yielding commercial properties	Aryaman Vir	0	$6

In [30]:
```
df.at[784,"amount"]=6000000.0
df.at[784,'stage']=" "
```

In [31]:
```
df[df.amount==0].head()
```

Out[31]:

	company	founded	headquarters	sector	what_it_does	founder	investor
22	FanAnywhere	2021.0	Bangalore	Financial Services	Celebrity NFT platform	Varun Chaudhary, Amit Kumar	Oasis Capital, Scorpio VC, DeltaHub Capital
30	Fullife Healthcare	2009.0	Pharmaceuticals #REF!	Primary Business is Development and Manufactur...	Varun Khanna	Morgan Stanley Private Equity Asia	$22,000,000
45	MoEVing	2021.0	Gurugram #REF!	MoEVing is India's only Electric Mobility focu...	Vikash Mishra, Mragank Jain	Anshuman Maheshwary, Dr Srihari Raju Kalidindi	$5,000,000
88	Karkinos Healthcare	2020.0	Mumbai	Hospital & Health Care	Medical center for treatment of complex cancer...	R Venkataramanan, Ravi Kant	Reliance Digital Health
92	Zorgers	2013.0	Mohali	Hospital & Health Care	Zorgers Home Healthcare is India's most truste...	Varun Gupta, Anil Kumar, Abhinav Gupta	Ritu Marya Family Office.

In [32]:
```
df.at[30,"amount"]=22000000
df.at[45,"amount"]=5000000
df.at[335,"amount"]=0
df.at[383,"amount"]=1000000
df.at[30,"sector"]="Pharmaceuticals"
df.at[335,"sector"]="Online Media"
df.at[30,"what_it_does"]="Primary Business is Development and Manufacturing of Novel H
df.at[45,"what_it_does"]="MoEVing is India's only Electric Mobility focused Technology
df.at[335,"what_it_does"]="Sochcast is an Audio experiences company that give the list
df.at[30,"founder"]="Varun Khanna"
df.at[45,"founder"]="Vikash Mishra, Mragank Jain"
df.at[335,"founder"]="CA Harvinderjit Singh Bhatia, Garima Surana, Anil Srivatsa"
df.at[30,"investor"]="Morgan Stanley Private Equity Asia"
df.at[45,"investor"]="Anshuman Maheshwary, Dr Srihari Raju Kalidindi"
df.at[335,"investor"]="Vinners, Raj Nayak, Amritaanshu Agrawal"
df.at[383,"investor"]="NA"
```

In [33]:
```
df.at[30,"headquarters"]="Mumbai"
df.at[45,"headquarters"]="Gurugram"
df.at[45,"sector"]="Automotive"
df.at[335,"headquarters"]="Bangalore"
```

In [34]:
```
df.loc[[30,45,335,383,972]]
```

Out[34]:

	company	founded	headquarters	sector	what_it_does	founder	investor	a
30	Fullife Healthcare	2009.0	Mumbai	Pharmaceuticals	Primary Business is Development and Manufactur...	Varun Khanna	Morgan Stanley Private Equity Asia	22(
45	MoEVing	2021.0	Gurugram	Automotive	MoEVing is India's only Electric Mobility focu...	Vikash Mishra, Mragank Jain	Anshuman Maheshwary, Dr Srihari Raju Kalidindi	5(
335	Sochcast	2020.0	Bangalore	Online Media	Sochcast is an Audio experiences company that ...	CA Harvinderjit Singh Bhatia, Garima Surana, A...	Vinners, Raj Nayak, Amritaanshu Agrawal	
383	Godamwale	2016.0	Mumbai	Logistics & Supply Chain	Godamwale is tech enabled integrated logistics...	Basant Kumar, Vivek Tiwari, Ranbir Nandan	NA	1(
972	Onsurity	2020.0	Bangalore	HealthCare	Onsurity is an employee healthcare platform pr...	Kulin Shah, Yogesh Agarwal	Jitendra Gupta, Harsh Shah	

In [35]:
```
df[df.headquarters==" "]
```

Out[35]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amount	stage	fur
915	Vidyakul	2018.0		EdTech	Vidyakul is an vernacular e-learning platform ...	Raman Garg, Tarun Saini	JITO Angel Network, SOSV	500000	Seed	2

In [36]:
```
df.at[915,"headquarters"]="Gurugram"
```

In [37]:
```
df.loc[915]
```

Out[37]:
```
company                                                 Vidyakul
founded                                                   2018.0
headquarters                                             Gurugram
sector                                                     EdTech
what_it_does    Vidyakul is an vernacular e-learning platform ...
founder                                       Raman Garg, Tarun Saini
investor                                    JITO Angel Network, SOSV
amount                                                    500000
stage                                                       Seed
funding_date                                 2021-04-01 00:00:00
Name: 915, dtype: object
```

In [38]:
```
df[df.company=="Smart Express"]
```

Out[38]:

	company	founded	headquarters	sector	what_it_does	founder	investor	amount	stage
494	Smart Express	2018.0	Mumbai	Logistics	India's Most Innovative and Awarded Express Lo...	Yogesh Dhingra	IIFL India Private Equity Fund, Smiti Holding ...	10000000	Seed

Summaries and visualization

1. Startup trend in the country

In [39]:
```
yr=df.founded.value_counts()
```

In [40]:
```
yr=yr.to_frame(name="number_of_stps")
yr.index.name="year"
yr.reset_index(inplace=True)
```

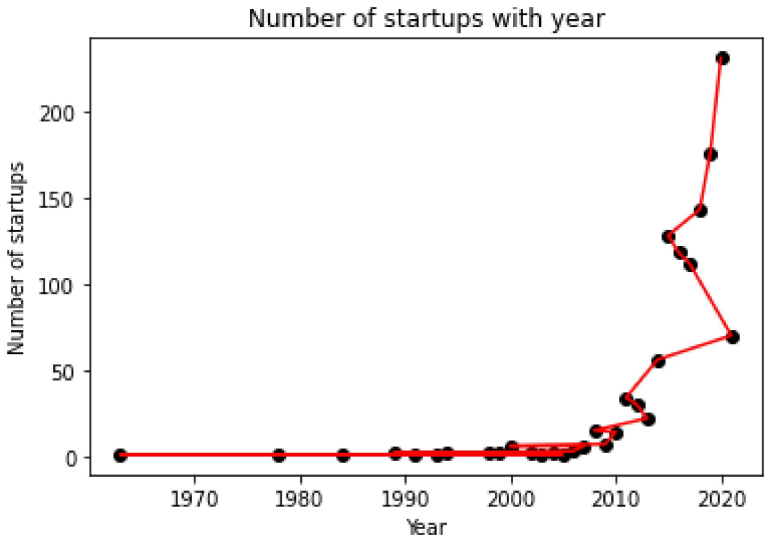
In [41]:
```
yr
```

Out[41]:

	year	number_of_stps
0	2020.0	231
1	2019.0	176
2	2018.0	143
3	2015.0	128
4	2016.0	118
5	2017.0	112
6	2021.0	70
7	2014.0	56
8	2011.0	34
9	2012.0	30
10	2013.0	22
11	2008.0	15
12	2010.0	14
13	2009.0	7
14	2000.0	6
15	2007.0	6
16	2006.0	3
17	1994.0	2
18	1989.0	2
19	2004.0	2
20	1998.0	2
21	2002.0	2
22	1999.0	2
23	2005.0	1
24	1963.0	1
25	2003.0	1
26	1993.0	1
27	1991.0	1
28	1984.0	1
29	1978.0	1

In [42]:
```
plt.plot(yr.year,yr.number_of_stps,"r")
plt.scatter(yr.year,yr.number_of_stps,color="black")
plt.title("Number of startups with year")
plt.xlabel("Year")
plt.ylabel("Number of startups")
```

Out[42]:
```
Text(0, 0.5, 'Number of startups')
```



2. Cities with most startups

In [43]:
```
hq=df.headquarters.value_counts().head(10)
```

In [44]:
```
hq=hq.to_frame(name="Number of startups")
hq.index.name="City"
```

In [45]:
```
hq=hq.sort_values(by="Number of startups",ascending=True)
hq.reset_index(inplace=True)
```

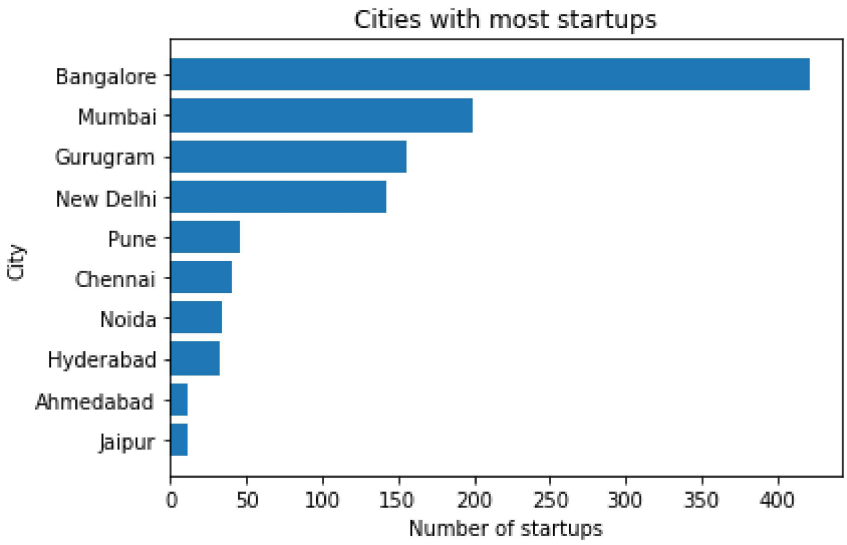
In [46]:
```
hq
```

Out[46]:

	City	Number of startups
0	Jaipur	11
1	Ahmedabad	11
2	Hyderabad	32
3	Noida	34
4	Chennai	40
5	Pune	46
6	New Delhi	142
7	Gurugram	156
8	Mumbai	199
9	Bangalore	422

In [47]:
```
plt.barh(hq["City"],hq["Number of startups"])
plt.xlabel("Number of startups")
plt.ylabel("City")
plt.title("Cities with most startups")
```

Out[47]:
```
Text(0.5, 1.0, 'Cities with most startups')
```



3. Most common sectors among all

In [48]:
```
sect=df.sector.value_counts().head(20).to_frame(name="Number of startups").sort_values
```

In [49]:
```
sect.index.name="Sector"
sect.reset_index(inplace=True)
```

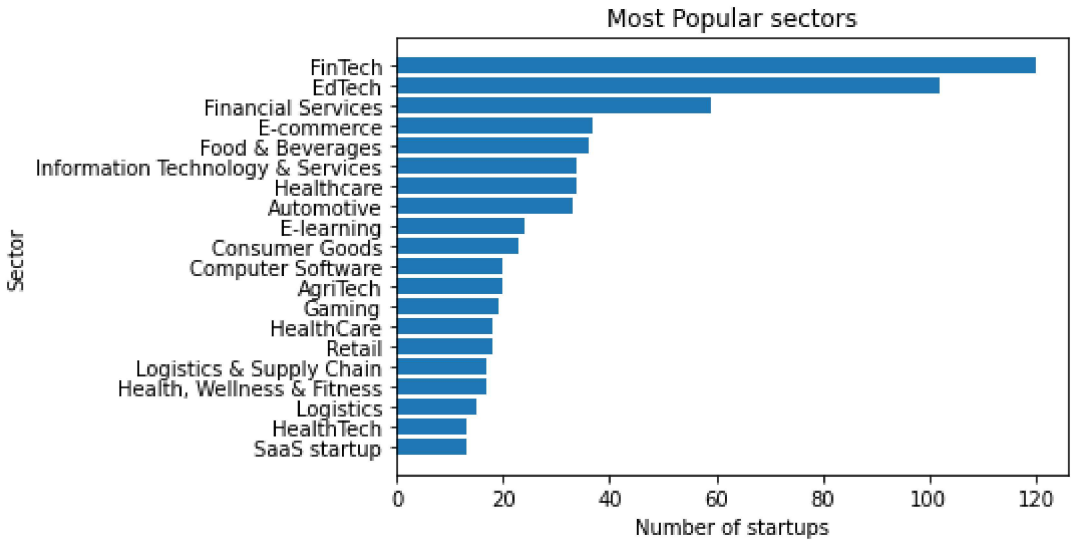
In [50]:
```
sect.head()
```

Out[50]:

	Sector	Number of startups
0	SaaS startup	13
1	HealthTech	13
2	Logistics	15
3	Health, Wellness & Fitness	17
4	Logistics & Supply Chain	17

In [51]:
```
plt.barh(sect["Sector"],sect["Number of startups"])
plt.xlabel("Number of startups")
plt.ylabel("Sector")
plt.title("Most Popular sectors")
```

Out[51]:
```
Text(0.5, 1.0, 'Most Popular sectors')
```



4. Ranking by Average investment received by startups

In [52]:
```
psect=df[df.amount!=0][["sector","amount"]]
```

In [53]:
```
psect=psect.groupby("sector")
```

In [54]:
```
psect=psect.mean().sort_values(by="amount",ascending=False).head(10)
psect.reset_index(inplace=True)
psect.sort_values(by="amount",ascending=True,inplace=True)
```

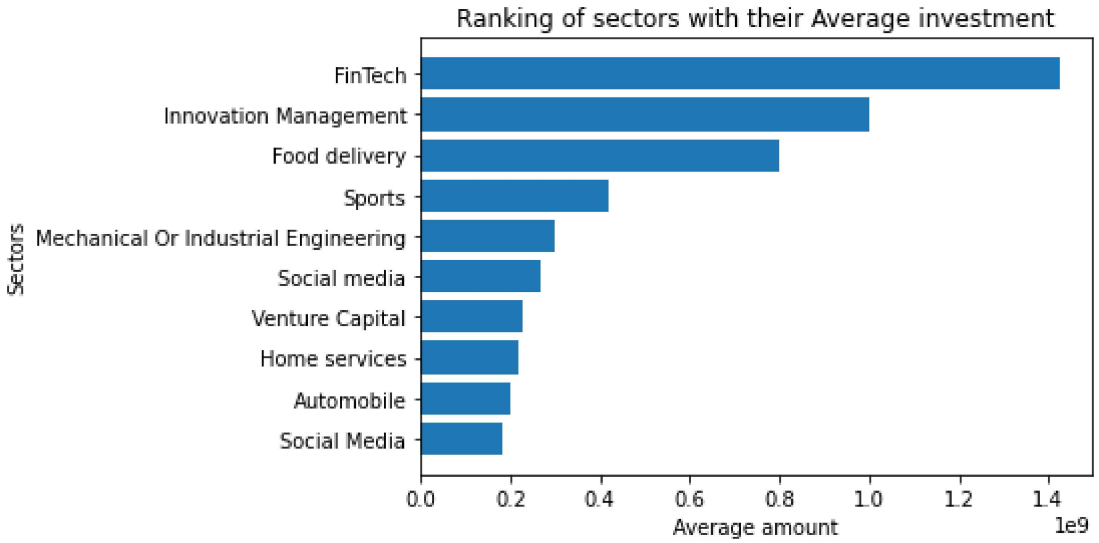
In [55]:
```
psect
```

Out[55]:

	sector	amount
9	Social Media	1.810000e+08
8	Automobile	2.000000e+08
7	Home services	2.190000e+08
6	Venture Capital	2.250000e+08
5	Social media	2.660000e+08
4	Mechanical Or Industrial Engineering	3.002000e+08
3	Sports	4.210000e+08
2	Food delivery	8.000000e+08
1	Innovation Management	1.000000e+09
0	FinTech	1.426267e+09

In [56]:
```
plt.barh(psect.sector,psect.amount)
plt.title("Ranking of sectors with their Average investment")
plt.xlabel("Average amount")
plt.ylabel("Sectors")
```

Out[56]:
```
Text(0, 0.5, 'Sectors')
```



5. Overview of months with total amount of investment in startups

In [57]:
```
fnds_dt= df.funding_date.value_counts().to_frame(name='Number of investors')
fnds_dt.index.name="month"
fnds_dt.reset_index(inplace=True)
```

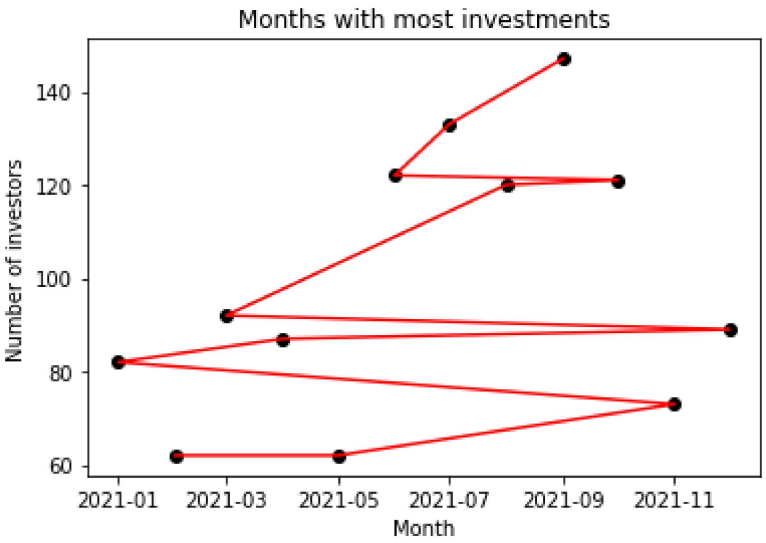
In [58]:
```
fnds_dt
```

Out[58]:

	month	Number of investors
0	2021-09-01	147
1	2021-07-01	133
2	2021-06-01	122
3	2021-10-01	121
4	2021-08-01	120
5	2021-03-01	92
6	2021-12-01	89
7	2021-04-01	87
8	2021-01-01	82
9	2021-11-01	73
10	2021-05-01	62
11	2021-02-01	62

In [59]:
```
plt.plot(fnds_dt["month"],fnds_dt["Number of investors"],color="red")
plt.scatter(fnds_dt["month"],fnds_dt["Number of investors"],color="black")
plt.xlabel("Month")
plt.ylabel("Number of investors")
plt.title("Months with most investments")
```

Out[59]:
```
Text(0.5, 1.0, 'Months with most investments')
```



6. Stages of startups

In [60]:
```
stg=df[df.stage!=" "].stage.value_counts().to_frame(name="Number of startups").sort_va
stg.index.name="Stage"
stg.reset_index(inplace=True)
stg.head()
```

Out[60]:

	Stage	Number of startups
0	Series F1	1
1	Pre-Seed	1
2	Seed+	1
3	Series F2	1
4	Series A+	1

a. Most common stages

In [61]:
```
stg_comn=stg.tail(20)
```

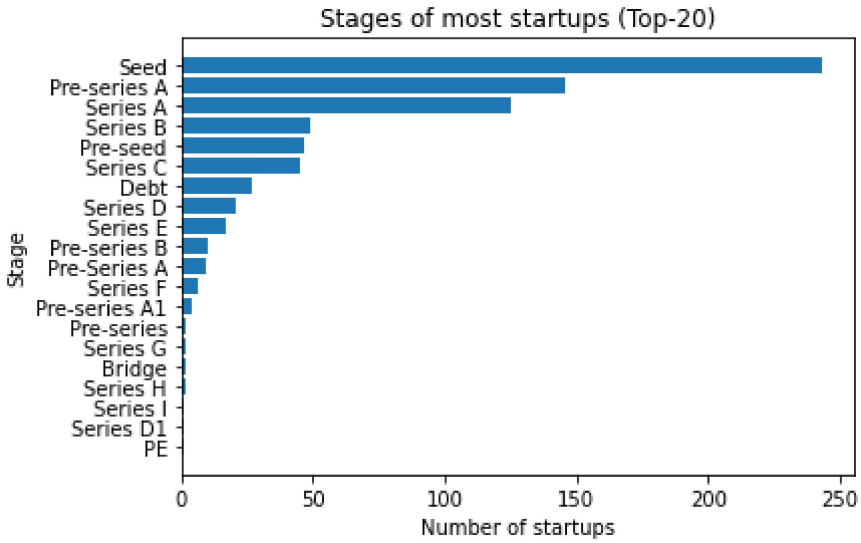
In [62]:
```
stg_comn
```

Out[62]:

	Stage	Number of startups
8	PE	1
9	Series D1	1
10	Series I	1
11	Series H	2
12	Bridge	2
13	Series G	2
14	Pre-series	2
15	Pre-series A1	4
16	Series F	6
17	Pre-Series A	9
18	Pre-series B	10
19	Series E	17
20	Series D	21
21	Debt	27
22	Series C	45
23	Pre-seed	47
24	Series B	49
25	Series A	125
26	Pre-series A	146
27	Seed	243

In [63]:
```
plt.barh(stg_comn["Stage"],stg_comn["Number of startups"])
plt.title("Stages of most startups (Top-20)")
plt.ylabel("Stage")
plt.xlabel("Number of startups")
```

Out[63]:
```
Text(0.5, 0, 'Number of startups')
```



b. Rare stages

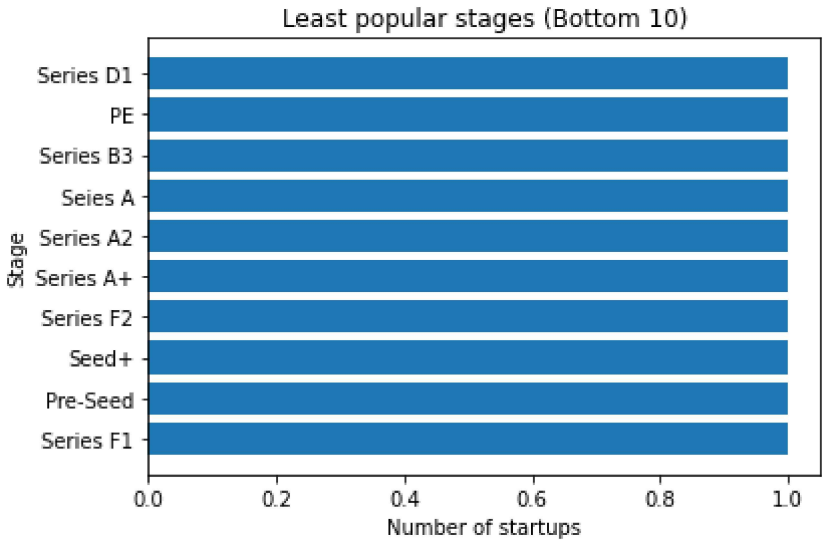
In [64]:
```
stg_uncmn=stg.head(10)
stg_uncmn
```

Out[64]:

	Stage	Number of startups
0	Series F1	1
1	Pre-Seed	1
2	Seed+	1
3	Series F2	1
4	Series A+	1
5	Series A2	1
6	Seies A	1
7	Series B3	1
8	PE	1
9	Series D1	1

In [65]:
```
plt.barh(stg_uncmn["Stage"],stg_uncmn["Number of startups"])
plt.title("Least popular stages (Bottom 10)")
plt.xlabel("Number of startups")
plt.ylabel("Stage")
```

Out[65]:
```
Text(0, 0.5, 'Stage')
```



8. Startups which raised most funds

In [66]:
```
funds_rankings=df.loc[:, ~df.columns.isin(["what_it_does","founder"])].sort_values(asc
```

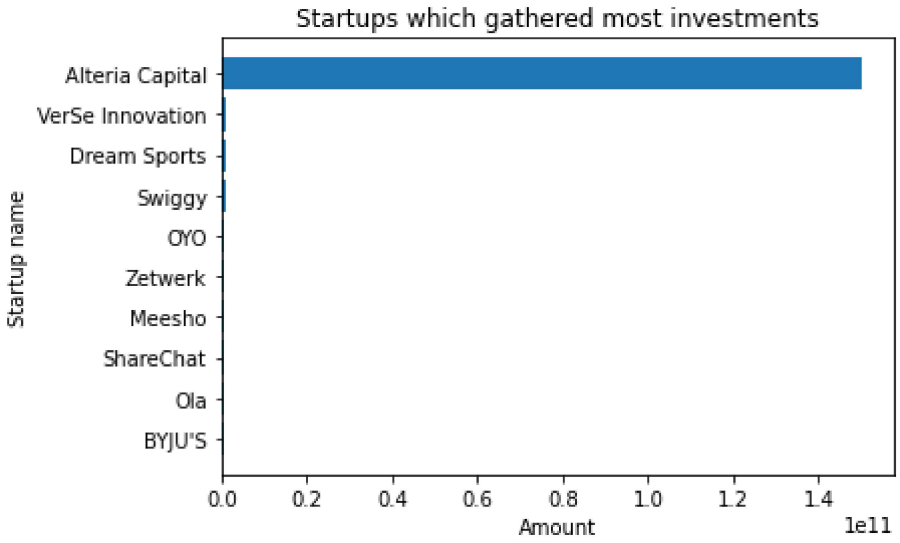
In [67]:
```
top_fnds=funds_rankings.head(10).sort_values(by="amount",ascending=True)
top_fnds
```

Out[67]:

	company	founded	headquarters	sector	investor	amount	stage	funding_da
995	BYJU'S	2011.0	Bangalore	EdTech	BlackRock, Owl Ventures	460000000	Series F	2021-03
55	Ola	2010.0	Bangalore	Mobility	Marquee international institutional investors	500000000		2021-12
956	ShareChat	2015.0	Bangalore	Social Media	Twitter Ventures, Pawan Munjal	500000000		2021-04
305	Meesho	2015.0	Bangalore	Social commerce	Fidelity, B Capital	570000000		2021-09
218	Zetwerk	2018.0	Bangalore	Mechanical Or Industrial Engineering	TradeCred	600000000		2021-10
642	OYO	2013.0	Gurugram	Hospitality	SoftBank Vision Fund, Hindustan Media Venture	660000000		2021-07
939	Swiggy	2014.0	Bangalore	Food delivery	Carmignac, Falcon Edge Capital	800000000		2021-04
170	Dream Sports	2008.0	Mumbai	Sports	Falcon Edge, DST Global, D1 Capital, Redbird C...	840000000		2021-11
1096	VerSe Innovation	2007.0	Bangalore	Innovation Management	Canaan Valley Capital, Glade Brook Capital Par...	1000000000	Series H	2021-02
959	Alteria Capital	2018.0	Mumbai	FinTech		150000000000	Debt	2021-04

In [68]:
```
plt.barh(top_fnds.company,top_fnds.amount)
plt.title("Startups which gathered most investments")
plt.ylabel("Startup name")
plt.xlabel("Amount")
```

Out[68]:
```
Text(0.5, 0, 'Amount')
```



9. Startups with lowest ranking in fund raising

In [69]:
```
bttm_rankings=funds_rankings[funds_rankings.amount!=0].tail(10).sort_values(by="amount
bttm_rankings
```

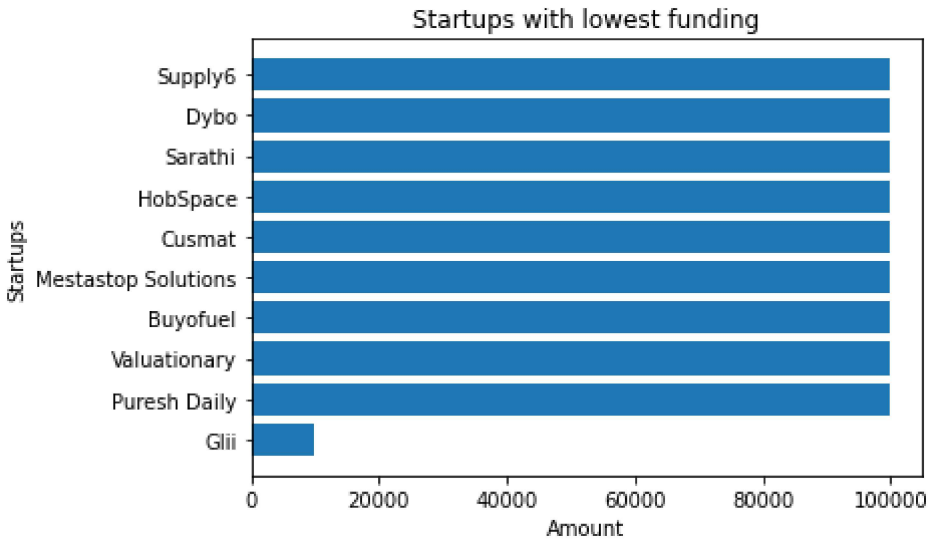
Out[69]:

	company	founded	headquarters	sector	investor	amount	stage	funding_dat
96	Glii	2020.0	Noida	Dating	We Founder Circle	10000	Seed	2021-12-0
927	Puresh Daily	2019.0	Ranchi	Milk startup	Alfa Ventures, Agility Venture Partners	100000	Seed	2021-04-0
132	Valuationary	2020.0	Surat	E-learning	Ashish Jain, Ajay Surana	100000	Pre-seed	2021-11-0
129	Buyofuel	2020.0	Coimbatore	Oil & Energy	Inflection Point Ventures	100000	Seed	2021-11-0
239	Mestastop Solutions	2018.0	Bangalore	Biotechnology	CIIE.CO	100000		2021-10-0
307	Cusmat	2016.0	Hyderabad	Computer Software	We Founder Circle	100000	Pre-series A	2021-09-0
1204	HobSpace	2019.0	Mumbai	EdTech	Siddharth Bhaskar Shah, Upsparks	100000	Pre-series A1	2021-01-0
877	Sarathi	2020.0	Jodhpur	HealthCare		100000	Seed	2021-05-0
930	Dybo	2019.0	Bangalore	IT	Spanache	100000	Seed	2021-04-0
1087	Supply6	2018.0	Bangalore	Helathcare	Rohit Goutamchand, India CXO Fund	100000	Seed	2021-02-0

In [70]:
```
plt.barh(bttm_rankings.company,bttm_rankings.amount)
```

```
plt.title("Startups with lowest funding")
plt.ylabel("Startups")
plt.xlabel("Amount")
```

Out[70]: Text(0.5, 0, 'Amount')



Conclusion

Major conclusions of analysis-

1. **Startup Trend in India from early ages to 2021**- If we monitor the trend line of startups in India, we will find that it looks constant from early age till the year 2010 and after that, it started moving upward and especially after 2016 it went up exponentially till 2020 by fell in 2020 which is not a good sign and reason for that could be the upside-down changes in the environment due to pandemic. However, the overall trend is indicating a positive sign. Startups in India are becoming common these days and looking at the trendline we can predict that it will keep rising continuously in upcoming years as well.

1. **Cities with startup-friendly environments**- Bengaluru remains the city with the most number of startups. An overall analysis of top-ranked cities gives indications that many tier-2 cities are also contributing along with the metropolitan cities.

1. **Most and least common sectors to start a startup**- In India 10% of all startups are in the finTech sector while 8.5% are in the EdTech and then Financial services with almost 5.00% followed by E-commerce and Food & Beverages. On the other hand, if we look at the least popular sectors then we will see that Fashion & Lifestyle, Automobile, wholesale, etc are getting out of fashion. we can conclude that the conventional business sectors are becoming unpopular while the Tech startups are taking off.

1. **Stages of startups**- While concluding the stages of startups, we cannot blindly rely on this analysis due to the significant amount of unavailability of data in this particular area.

However, if we have to take an overall estimate then it is safe to say that most of the startups are in the Seed phase and there are very few who belong to Series F2, F1. What we can conclude is that most of the startups are in the early stages which shows that it's just the beginning of the startup wave in India. The negative factor is that they are vulnerable, there thriving is not ensured. Hence the market is still booming and new to most of them.

1. **Which sectors do investors focus on most?**- The analysis suggests that FinTech remains the most preferred sector and it is because 30% of startups are from the FinTech sector. However, the second spot is taken by Innovation Management which is nowhere on the list of popular sectors. It shows that this sector has the potential to thrive as investors rate it very high but it is not very common in India if we compare it with the sectors like FinTech, EdTech, Food delivery, etc. Food delivery and sports are also the centers of investors' attention.