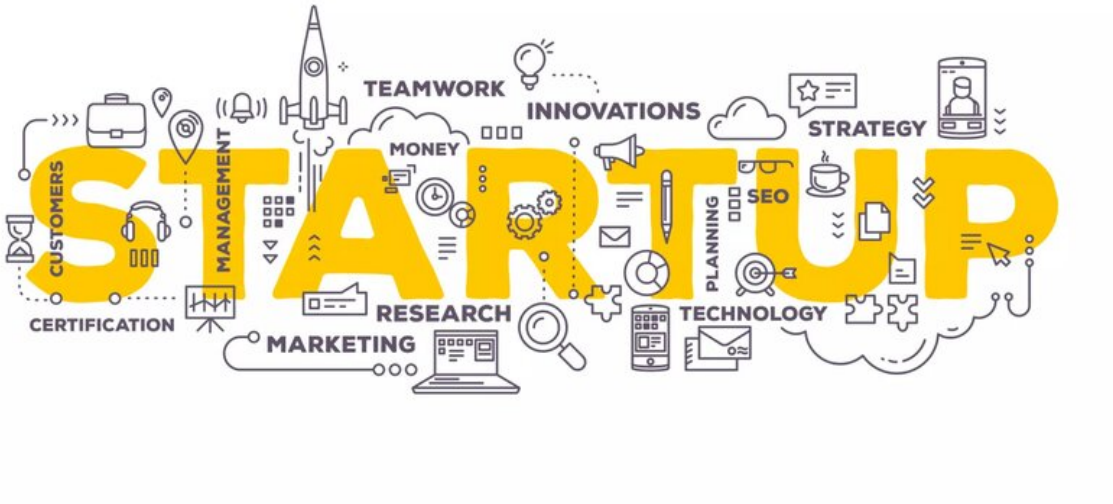
Indian startups



About Dataset

Content The following dataset has data about the Top 300 startups in India. Details about the columns are as follows:

- Company - Name of the Startup.
- City - The City in which the startup is started.
- Starting Year - The Year in which the startup was started.
- Founders - Name of the founders of the startup.
- Industries - Industrial domain in which the startup falls.
- No. of Employees - Number of employees in the startup.
- Funding Amount in USD - Total funding amount funded to the startup.
- Funding Rounds - Funding rounds are the number of times a startup goes back to the market to raise more capital. The goal of every round is for founders to trade equity in their business for the capital they can utilize to advance their companies to the next level.
- No. of Investors - Number of investors in the startup.

Introduction

This dataset gives us information about new companies, known as startups. Startups are different from regular new businesses because they want to grow really big, not just stay small. At the beginning, startups often face a lot of problems and many of them don't succeed. But some of them do become successful and have a big impact on the business world. This dataset helps us understand which industries these startups are in and how they're doing.

Workflow

- Understanding Data
- Data Cleaning and Preprocessing
- EDA (Exploratory Data Analysis)
- Insights
- Conclusion

In [1]:

```
#Importing Libraries

import warnings
warnings.filterwarnings("ignore")

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [49]:

```
df = pd.read_csv("Startups1.csv")
df
```

Out[49]:

	Unnamed: 0	Company	City	Starting Year	Founders	Industries	Description	Em
0	0	Urban Company	Gurgaon	2014	Abhiraj Singh Bhal, Raghav Chandra, Varun Khaitan	Apps, Home Services, Marketplace, Service Indu...	Urban is a marketplace for independent contrac...	10
1	1	Classplus	Noida	2018	Bhaswat Agarwal, Bikash Dash, Mukul Rustagi, N...	B2B, E-Learning, EdTech, Education, Mobile App...	Classplus is a mobile-first SaaS platform that...	
2	2	Paytm	Noida	2010	Akshay Khanna, Vijay Shekhar Sharma	E-Commerce, Finance, Financial Services, Inter...	Paytm is a payment gateway that allows users a...	5
3	3	Apna	Mumbai	2019	Nirmit Parikh	Employment, Human Resources, Recruiting, Staff...	Apna is a professional networking and job-sear...	
4	4	Razorpay	Bengaluru	2014	Harshil Mathur, Shashank Kumar	Financial Services, FinTech, Payments, Software	Razorpay is a payment acceptance, processing, ...	10
...	...	...	...	...	...	...	...	
295	295	SafexPay	Thane	2017	Ravi Gupta	FinTech, Payments	Safexpay is a B2B2B finance company that speci...	
296	296	Pariksha	Pune	2015	Deepak Choudhary, Karanvir Singh Shekhawat, Ut...	E-Learning, EdTech, Education, Skill Assessment	Pariksha - The Success App is India's leading ...	
297	297	Fyllo	Bengaluru	2019	Sachin Gautam, Sudhanshu Rai, Sumit Sheoran	Agriculture, AgTech, Artificial Intelligence, ...	Fyllo is a precision agriculture service provi...	
298	298	CredFlow	New Delhi	2019	Kunal Aggarwal	Credit, Financial Services	CredFlow offers financial services that help y...	
299	299	HalaPlay Technologies	Bengaluru	2016	Swapnil Saurav	Digital Entertainment, Fantasy Sports, Sports	HalaPlay Technologies is a platform for daily ...	

300 rows × 11 columns

Understanding Data

Let's see the shape of the data

In [5]:
```
1 df.shape
```

Out[5]:
```
(300, 11)
```

Let's see if there any null value in the dataset

In [8]:
```
1 df.isnull().sum()
```

Out[8]:
```
Unnamed: 0             0
Company                0
City                   0
Starting Year          0
Founders               0
Industries             0
Description            0
No. of Employees       0
Funding Amount in $    0
Funding Round          0
No. of Investors       0
dtype: int64
```

Let's see the information of dataset

In [9]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Unnamed: 0           300 non-null    int64
 1   Company              300 non-null    object
 2   City                 300 non-null    object
 3   Starting Year        300 non-null    int64
 4   Founders             300 non-null    object
 5   Industries           300 non-null    object
 6   Description          300 non-null    object
 7   No. of Employees     300 non-null    object
 8   Funding Amount in $  300 non-null    int64
 9   Funding Round        300 non-null    int64
 10  No. of Investors     300 non-null    int64
dtypes: int64(5), object(6)
memory usage: 25.9+ KB
```

In [10]:
```
1 df.head() #Head - returns a specified number of rows, string from the t
```

Out[10]:

	Unnamed: 0	Company	City	Starting Year	Founders	Industries	Description	No. of Employees
0	0	Urban Company	Gurgaon	2014	Abhiraj Singh Bhal, Raghav Chandra, Varun Khaitan	Apps, Home Services, Marketplace, Service Indu...	Urban is a marketplace for independent contrac...	1001-5000
1	1	Classplus	Noida	2018	Bhaswat Agarwal, Bikash Dash, Mukul Rustagi, N...	B2B, E-Learning, EdTech, Education, Mobile App...	Classplus is a mobile-first SaaS platform that...	101-250
2	2	Paytm	Noida	2010	Akshay Khanna, Vijay Shekhar Sharma	E-Commerce, Finance, Financial Services, Inter...	Paytm is a payment gateway that allows users a...	501-1000
3	3	Apna	Mumbai	2019	Nirmit Parikh	Employment, Human Resources, Recruiting, Staff...	Apna is a professional networking and job-sear...	101-250
4	4	Razorpay	Bengaluru	2014	Harshil Mathur, Shashank Kumar	Financial Services, FinTech, Payments, Software	Razorpay is a payment acceptance, processing, ...	1001-5000

In [11]:
```
1 df.tail() #Tail - returns a specified number of last rows
```

Out[11]:

	Unnamed: 0	Company	City	Starting Year	Founders	Industries	Description	Em
295	295	SafexPay	Thane	2017	Ravi Gupta	FinTech, Payments	Safexpay is a B2B2B finance company that speci...	
296	296	Pariksha	Pune	2015	Deepak Choudhary, Karanvir Singh Shekhawat, Ut...	E-Learning, EdTech, Education, Skill Assessment	Pariksha - The Success App is India's leading ...	
297	297	Fyllo	Bengaluru	2019	Sachin Gautam, Sudhanshu Rai, Sumit Sheoran	Agriculture, AgTech, Artificial Intelligence, ...	Fyllo is a precision agriculture service provi...	
298	298	CredFlow	New Delhi	2019	Kunal Aggarwal	Credit, Financial Services	CredFlow offers financial services that help y...	
299	299	HalaPlay Technologies	Bengaluru	2016	Swapnil Saurav	Digital Entertainment, Fantasy Sports, Sports	HalaPlay Technologies is a platform for daily ...	

Data Cleaning & Pre-processing

Let's Drop Unamed: 0 column as it appears to be an index or identifier with no relevance to the data analysis.

In [16]:
```
1 df.drop("Unnamed: 0", axis=1, inplace=True)
```

EDA - Exploratory Data Analysis

Let's see the starting year of the compaines or compaines founded in year

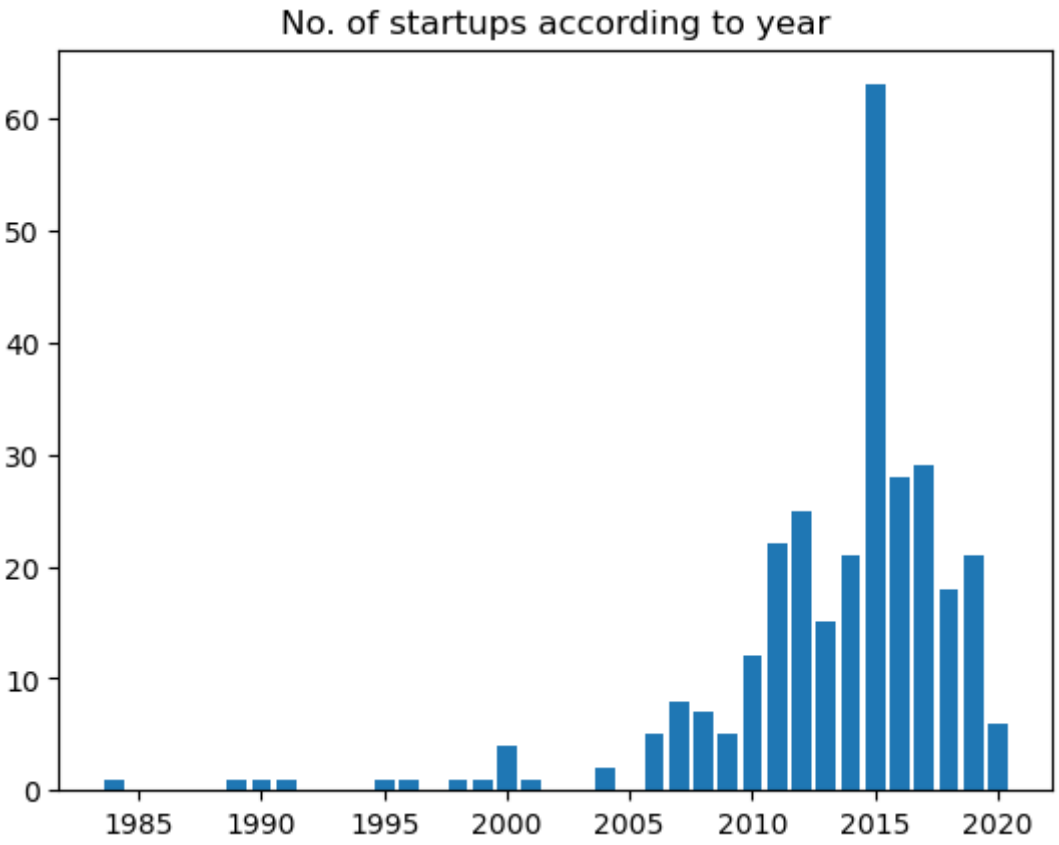
In [21]:
```
1 df["Starting Year"].value_counts()
```

Out[21]:
```
2015    63
2017    29
2016    28
2012    25
2011    22
2014    21
2019    21
2018    18
2013    15
2010    12
2007     8
2008     7
2020     6
2006     5
2009     5
2000     4
2004     2
1990     1
2001     1
1984     1
1996     1
1999     1
1989     1
1995     1
1998     1
1991     1
Name: Starting Year, dtype: int64
```

Let's plot it

In [26]:
```
1 plt.bar(df['Starting Year'].value_counts().index,df['Starting Year'].va
2 plt.title("No. of startups according to year")
```

Out[26]:
```
Text(0.5, 1.0, 'No. of startups according to year')
```



Insights

- the most number of startup founded in the year of 2015

Let's the see cities where companies founded most

In [27]:
```
1 df["City"].value_counts()
```

Out[27]:
```
Bengaluru       123
Mumbai           55
Gurgaon          47
New Delhi        22
Noida            12
Chennai           6
Hyderabad         5
Pune              4
Ahmedabad         4
Delhi             3
Haryana           3
Thane             2
Kormangala        2
Faridabad         2
Whitefield        1
Nungambakkam      1
Azadpur           1
Kilpauk           1
Patna City        1
Sohna             1
Kolkata           1
Surat             1
Andheri           1
Jaipur            1
Name: City, dtype: int64
```

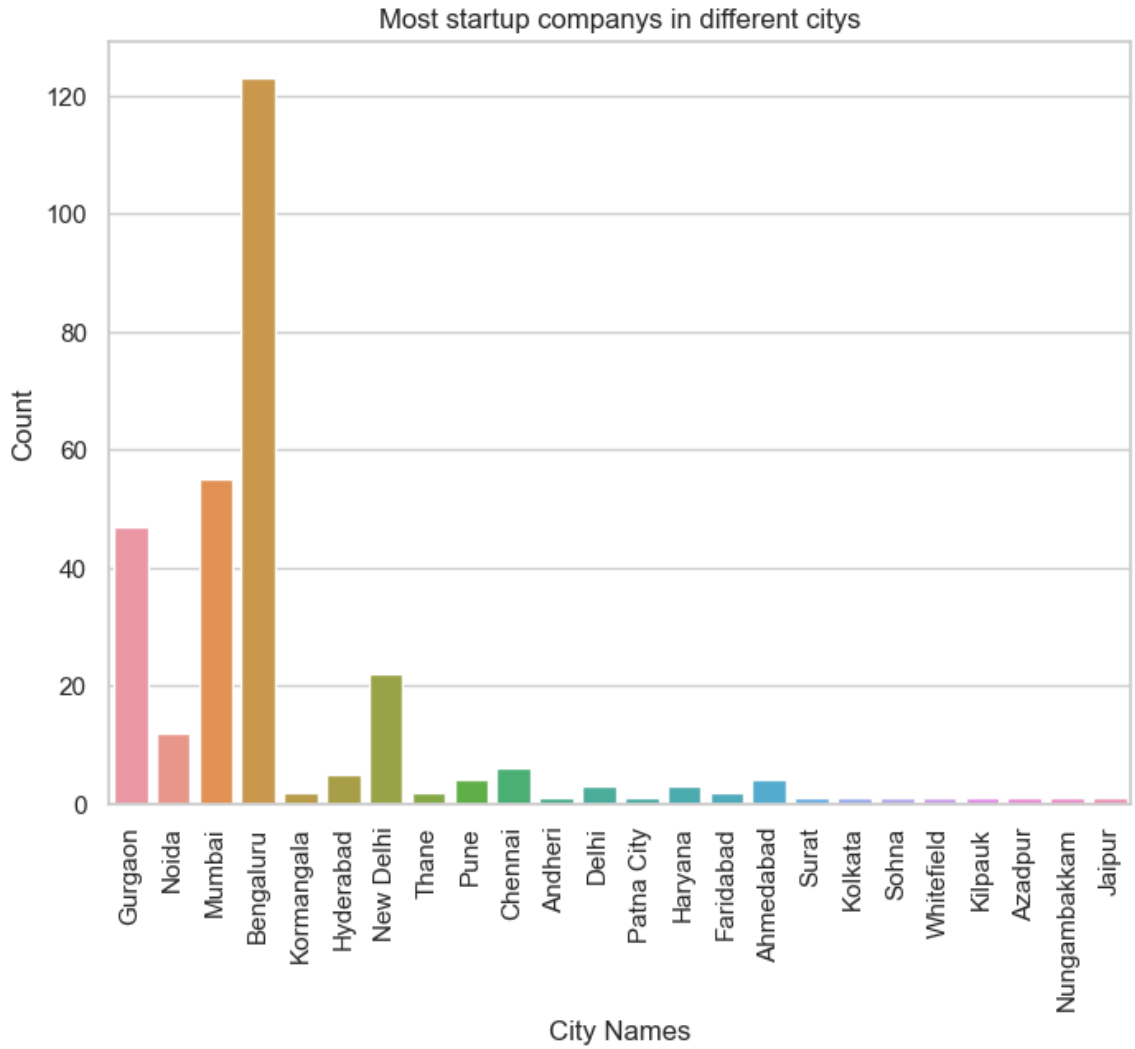
Let's Plot it

In [33]:

```
plt.figure(figsize=(8, 6))
sns.set(style="whitegrid")
sns.countplot(data=df,x="City")
plt.title("Most startup companys in different citys")
plt.xlabel('City Names')
plt.ylabel('Count')
plt.xticks(rotation=90)
```

Out[33]:

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 1
6,
        17, 18, 19, 20, 21, 22, 23]),
 [Text(0, 0, 'Gurgaon'),
  Text(1, 0, 'Noida'),
  Text(2, 0, 'Mumbai'),
  Text(3, 0, 'Bengaluru'),
  Text(4, 0, 'Kormangala'),
  Text(5, 0, 'Hyderabad'),
  Text(6, 0, 'New Delhi'),
  Text(7, 0, 'Thane'),
  Text(8, 0, 'Pune'),
  Text(9, 0, 'Chennai'),
  Text(10, 0, 'Andheri'),
  Text(11, 0, 'Delhi'),
  Text(12, 0, 'Patna City'),
  Text(13, 0, 'Haryana'),
  Text(14, 0, 'Faridabad'),
  Text(15, 0, 'Ahmedabad'),
  Text(16, 0, 'Surat'),
  Text(17, 0, 'Kolkata'),
  Text(18, 0, 'Sohna'),
  Text(19, 0, 'Whitefield'),
  Text(20, 0, 'Kilpauk'),
  Text(21, 0, 'Azadpur'),
  Text(22, 0, 'Nungambakkam'),
  Text(23, 0, 'Jaipur')])
```



Insight

- Most number of most startup companys located in Bengaluru,After that Mumbai,Gurganon, New Delhi
- And jaipur has lowest number of startup

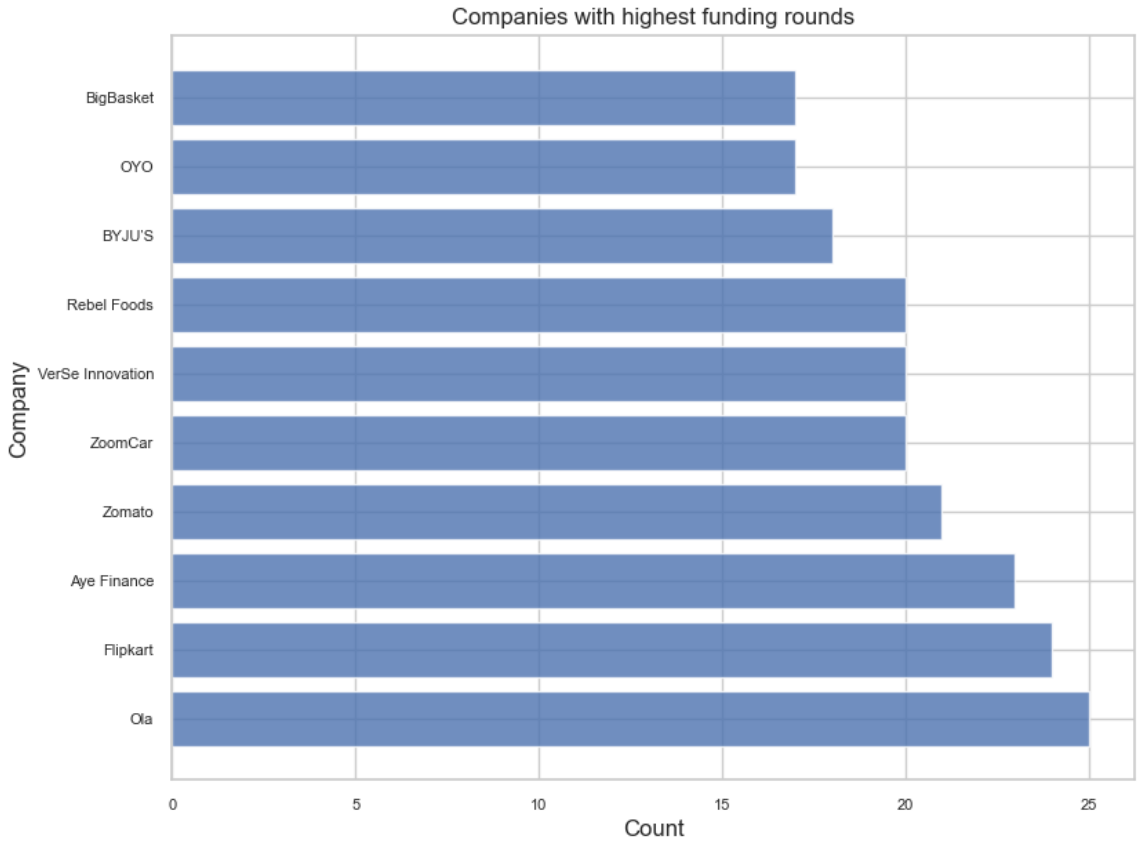
Let's see the which company has most funding round

In [52]:
```
sorting_funding_round = df["Funding Round"].sort_values(ascending=False
sorting_funding_round.index
```

Out[52]:
```
Int64Index([11, 41, 113, 107, 50, 181, 80, 23, 239, 8], dtype='int64')
```

In [53]:
```
sorting_funding_round1 = df.loc[[11, 41, 113, 107, 50, 181, 80, 23, 239
```

In [55]:
```
plt.figure(figsize=(9,7))
plt.barh(sorting_funding_round1,sorting_funding_round,alpha=0.8)
plt.title('Companies with highest funding rounds')
plt.ylabel('Company')
plt.xlabel('Count')
plt.tick_params(labelsize=8)
plt.show()
```



Insight

- Ola has highest funding round followed by flipkar and aye finance

Let's see the which company has most investor

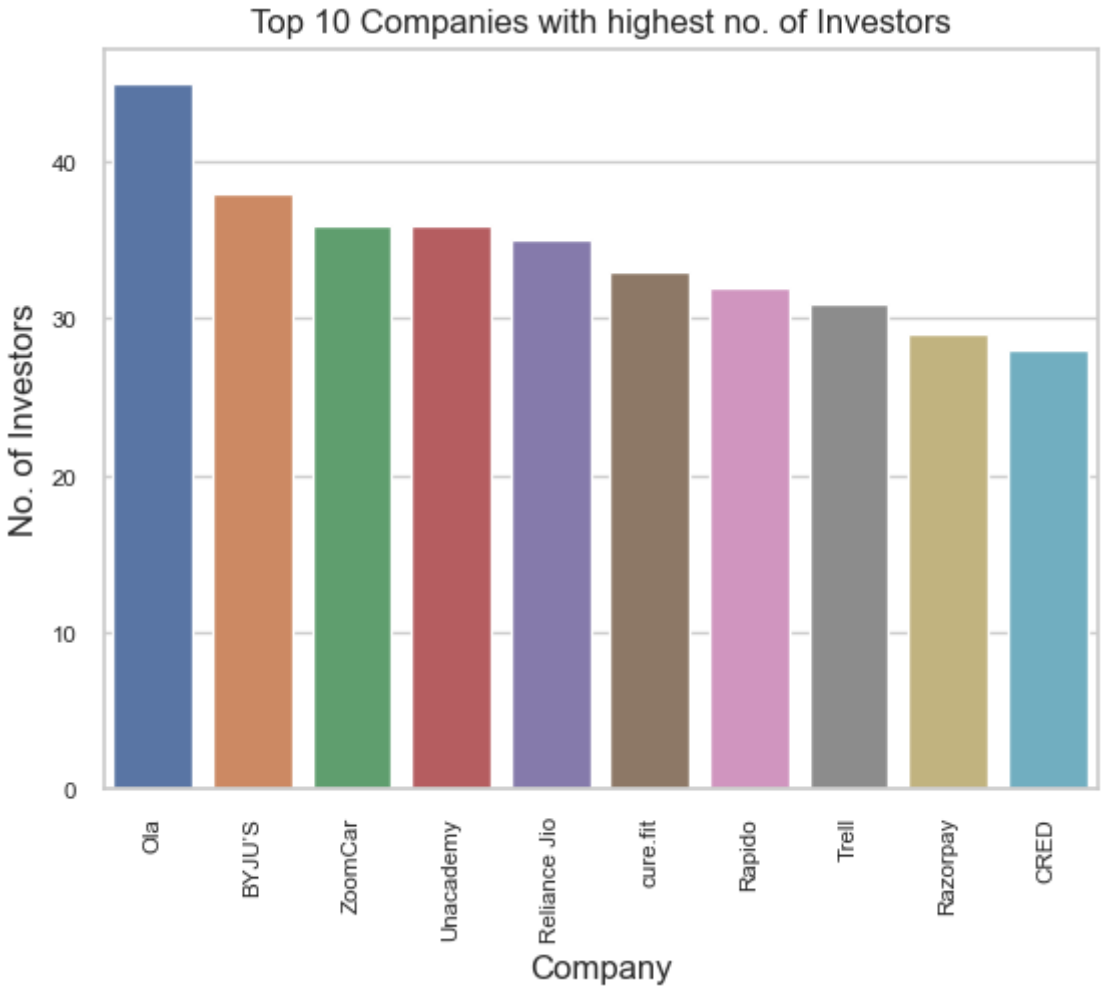
In [56]:
```
investor_sorting = df["No. of Investors"].sort_values(ascending=False).
investor_sorting
```

Out[56]:
```
11    45
23    38
50    36
10    36
32    35
69    33
31    32
88    31
4     29
13    28
Name: No. of Investors, dtype: int64
```

In [57]:
```
investor_sorting1 = df.loc[[11, 23, 50, 10, 32, 69, 31, 88, 4, 13],"Com
investor_sorting1
```

Out[57]:
```
11            Ola
23         BYJU'S
50        ZoomCar
10      Unacademy
32   Reliance Jio
69       cure.fit
31         Rapido
88          Trell
4        Razorpay
13           CRED
Name: Company, dtype: object
```

In [58]:
```
sns.barplot(x=investor_sorting1,y=investor_sorting,data=df)
plt.xticks(rotation='vertical')
plt.title('Top 10 Companies with highest no. of Investors')
plt.xlabel('Company',size=12)
plt.ylabel('No. of Investors',size=12)
plt.tick_params(labelsize='x-small')
```



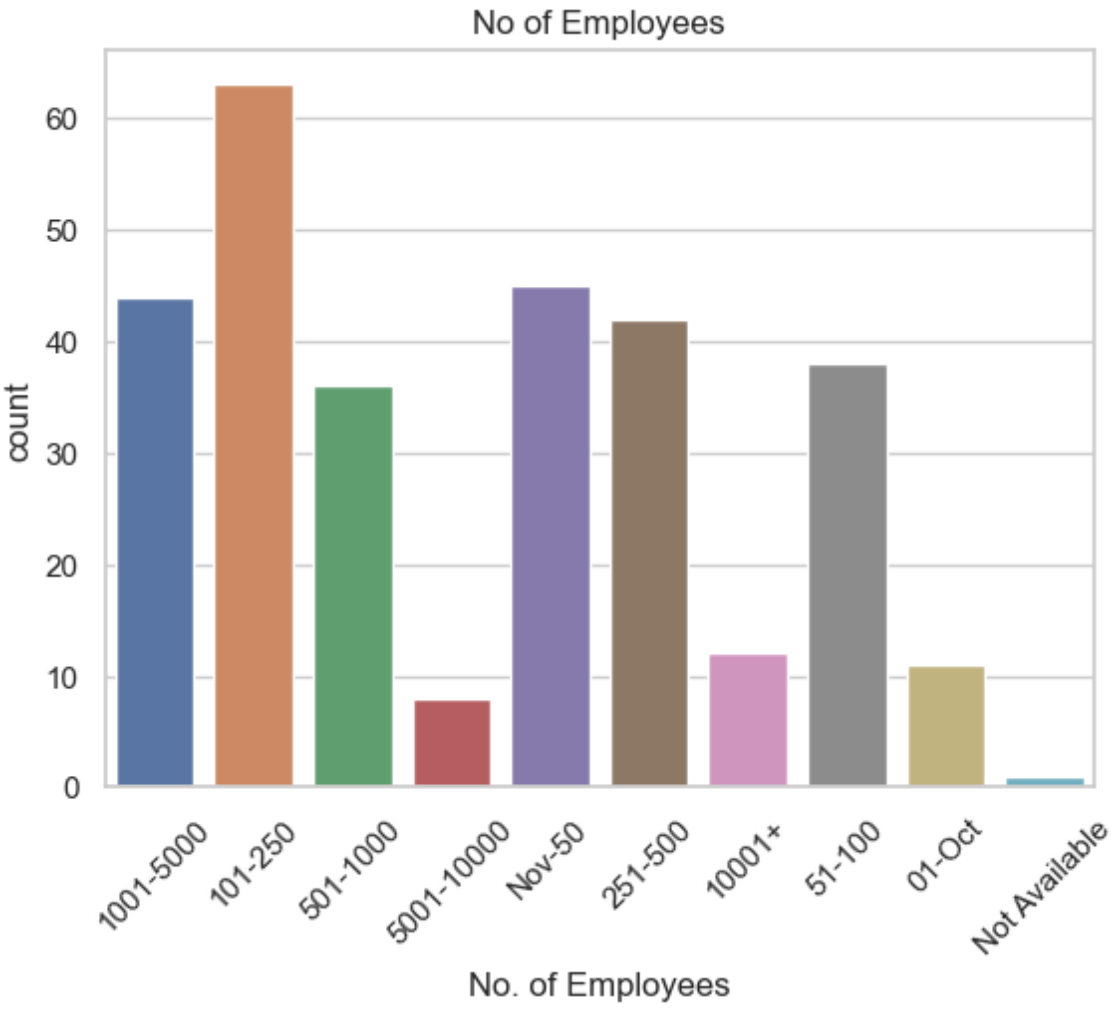
Insight

- Ola has the highest number of investor follwed by byju's and zoomcar

Let's see which company has most no. of employees

In [64]:
```
1 sns.countplot(data=df,x="No. of Employees")
2 plt.xticks(rotation=45)
3 plt.title('No of Employees')
```

Out[64]:
```
Text(0.5, 1.0, 'No of Employees')
```



Insight

- Most companies has employees count of 101 - 250

Conclusion

In conclusion, this dataset offers a window into the world of startups, providing insights into the industries they operate in and their potential for growth and success. While startups face significant challenges and uncertainties at the outset, a subset of them demonstrates resilience and the capacity to make a significant impact on their respective fields. By studying these startups and their journeys, we gain a better understanding of the ever-evolving landscape of entrepreneurship and the potential for innovation and transformation in the business world. This dataset serves as a valuable resource for exploring the dynamics of emerging companies and their contributions to the broader economy.