

STARTUP CASE STUDY AND ANALYSIS

In [1]:



```
#importing required Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams['figure.figsize']=(16,7)
```

In [2]:

```
#importing the dataset for the analysis
data = pd.read_csv("startup_funding.csv")

#creating a copy of original data
df=data.copy(deep=True)

#displaying the first 10 records in the dataset
df.head(10)
```

Out[2]:

	Sr No	Date dd/mm/yyyy	Startup Name	Industry Vertical	SubVertical	City Location	Investor
0	1	09/01/2020	BYJU'S	E-Tech	E-learning	Bengaluru	Tiger Global Management
1	2	13/01/2020	Shuttl	Transportation	App based shuttle service	Gurgaon	Susquehanna Growth Equity
2	3	09/01/2020	Mamaearth	E-commerce	Retailer of baby and toddler products	Bengaluru	Sequoia Capital
3	4	02/01/2020	https://www.wealthbucket.in/	FinTech	Online Investment	New Delhi	\ Khatri
4	5	02/01/2020	Fashor	Fashion and Apparel	Embroided Clothes For Women	Mumbai	Silicon Valley Partners
5	6	13/01/2020	Pando	Logistics	Open-market, freight management platform	Chennai	Chirba Venkatesh
6	7	10/01/2020	Zomato	Hospitality	Online Food Delivery Platform	Gurgaon	Ant Financial
7	8	12/12/2019	Ecozen	Technology	Agritech	Pune	Satish Catech Adv
8	9	06/12/2019	CarDekho	E-Commerce	Automobile	Gurgaon	Pir G Voyager
9	10	03/12/2019	Dhruva Space	Aerospace	Satellite Communication	Bengaluru	Mu Ar Ravil R

In [3]:

#changing column names of the given data

```
df.columns=["Sno","Date","StartupName","IndustryVertical","SubVertical","City","InvestorName","InvestmentType","AmountinUSD","Remarks"]
df.head()
```

Out[3]:

	Sno	Date	StartupName	IndustryVertical	SubVertical	City	InvestorName
0	1	09/01/2020	BYJU'S	E-Tech	E-learning	Bengaluru	Tiger G Manage
1	2	13/01/2020	Shuttl	Transportation	App based shuttle service	Gurgaon	Susqueh Growth E
2	3	09/01/2020	Mamaearth	E-commerce	Retailer of baby and toddler products	Bengaluru	Sec Capital
3	4	02/01/2020	https://www.wealthbucket.in/	FinTech	Online Investment	New Delhi	\ Khat
4	5	02/01/2020	Fashor	Fashion and Apparel	Embroided Clothes For Women	Mumbai	S Ve Par

In [4]:

#data cleaning

```
def clean_string(x):
    return str(x).replace("\\xc2\\xa0","").replace("\\\\xc2\\\\xa0","")

for col in ["Sno","Date","StartupName","IndustryVertical","SubVertical","City","InvestorName","InvestmentType","AmountinUSD","Remarks"]:
    df[col]=df[col].apply(lambda x: clean_string(x))
```

In [5]:

```
print("Rows : ",df.shape[0])
print("Columns : ",df.shape[1])
```

Rows : 3044
Columns : 10

In [6]:

```
print("Columns Names ")  
print(df.columns)
```

Columns Names

```
Index(['Sno', 'Date', 'StartupName', 'IndustryVertical', 'SubVertical', 'City',  
      'InvestorName', 'InvestmentType', 'AmountinUSD', 'Remarks'],  
      dtype='object')
```

DATA CLEANING

In [7]:

```
df.isnull().sum()
```

Out[7]:

```
Sno                0  
Date              0  
StartupName       0  
IndustryVertical  0  
SubVertical       0  
City              0  
InvestorName      0  
InvestmentType    0  
AmountinUSD       0  
Remarks          0  
dtype: int64
```

In [8]:

```
df.isna().sum()
```

Out[8]:

```
Sno                0  
Date              0  
StartupName       0  
IndustryVertical  0  
SubVertical       0  
City              0  
InvestorName      0  
InvestmentType    0  
AmountinUSD       0  
Remarks          0  
dtype: int64
```

In [9]:

```
#Since remarks has many NaN values we will remove the remark columns  
df=df.drop(['Remarks'],axis=1)
```

In [10]:

```
print("Data Set after removing remarks ")
df.head()
```

Data Set after removing remarks

Out[10]:

	Sno	Date	StartupName	IndustryVertical	SubVertical	City	InvestorN
0	1	09/01/2020	BYJU'S	E-Tech	E-learning	Bengaluru	Tiger G Manage
1	2	13/01/2020	Shuttl	Transportation	App based shuttle service	Gurgaon	Susqueh Growth E
2	3	09/01/2020	Mamaearth	E-commerce	Retailer of baby and toddler products	Bengaluru	Sec Capital
3	4	02/01/2020	https://www.wealthbucket.in/	FinTech	Online Investment	New Delhi	\ Khat
4	5	02/01/2020	Fashor	Fashion and Apparel	Embroided Clothes For Women	Mumbai	S Ve Par

In [11]:

```
# Lets convert the amount column into numerical, so that we can analyze the values inside i

# function to clean the AmounInUsd Column
def clean_amount(x):
    x = ''.join([c for c in str(x) if c in ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']])
    x = str(x).replace(",", "").replace("+", "")
    x = str(x).lower().replace("undisclosed", "")
    x = str(x).lower().replace("n/a", "")
    if x == '':
        x = '-999'
    return x

# Lets apply the function on the column
df["AmountinUSD"] = df["AmountinUSD"].apply(lambda x: float(clean_amount(x)))
```

In [12]:

```

## Cleaning the dates

# doesn't show warnings
import warnings
warnings.filterwarnings('ignore')

df['Date'][df['Date']=='12/05.2015'] = '12/05/2015'
df['Date'][df['Date']=='13/04.2015'] = '13/04/2015'
df['Date'][df['Date']=='15/01.2015'] = '15/01/2015'
df['Date'][df['Date']=='22/01//2015'] = '22/01/2015'
df['Date'][df['Date']=='05/072018'] = '05/07/2018'
df['Date'][df['Date']=='01/07/015'] = '01/07/2015'
df['Date'][df['Date']=='\\\\\\\\xc2\\\\\\\\xa010/7/2015'] = '10/07/2015'

```

DATA ANALYSIS

1.How does funding of ecosystem changes w.r.t time?

In [13]:

```

#converting to specific format eg. 201702 represent 2nd feb in 2017
df["yearmonth"] = (pd.to_datetime(df['Date'],
                                format='%d/%m/%Y').dt.year*100)+(pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.m

temp = df['yearmonth'].value_counts().sort_values(ascending = False).head(10)
print("Number of funding per month in decreasing order(Top 10)\n",temp)
year_month = df['yearmonth'].value_counts()

```

Number of funding per month in decreasing order(Top 10)

201601	104
201602	100
201508	98
201507	98
201606	98
201510	93
201511	92
201512	89
201509	88
201608	87

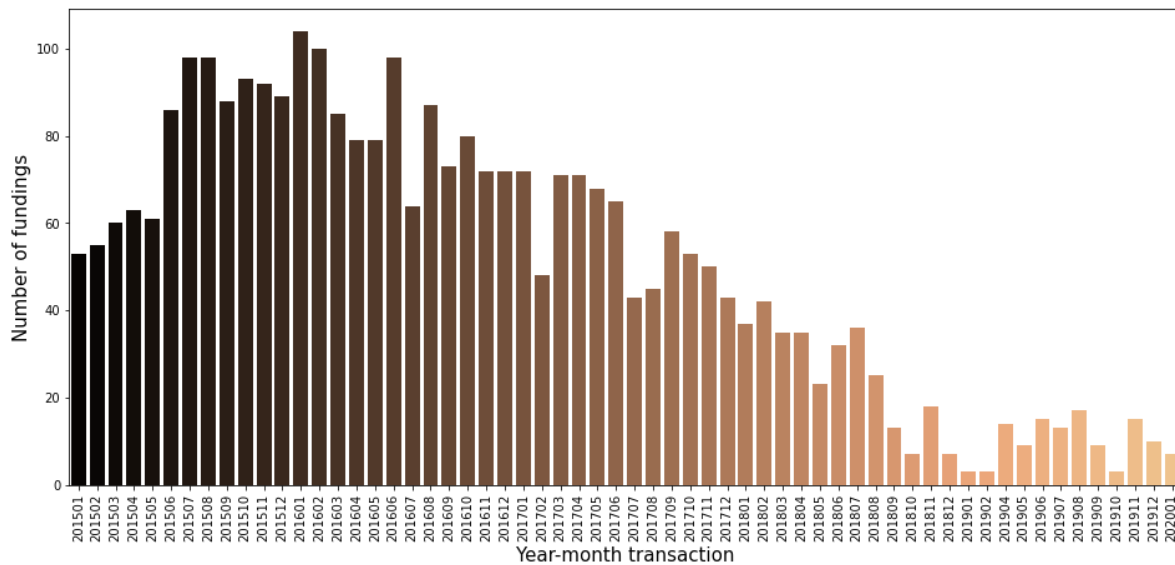
Name: yearmonth, dtype: int64

In [14]:

```
#barplot
sns.barplot(year_month.index, year_month.values,palette="copper")
plt.xticks(rotation=90)
plt.xlabel("Year-month transaction",fontsize=15)
plt.ylabel("Number of fundings",fontsize=15)
```

Out[14]:

```
Text(0, 0.5, 'Number of fundings')
```



2.WHAT IS THE GENRAL AMOUNT THAT STARTUP GET IN INDIA

In [15]:

```
#Maximum, Minimum and average funding for a startup
print("Maximum funding for a startup : ",df['AmountinUSD'].dropna().max())
print("Minimum funding for a startup : ",df['AmountinUSD'].dropna().min())
print("Average funding for a startup : ",df['AmountinUSD'].dropna().mean())
```

```
Maximum funding for a startup : 3900000000.0
Minimum funding for a startup : -999.0
Average funding for a startup : 13270058.261169514
```

In [16]:



#10 startups with Least funding

df[['AmountinUSD', 'StartupName']].sort_values(by='AmountinUSD', ascending=True).head(10)

Out[16]:

	AmountinUSD	StartupName
2125	-999.0	Splitkart
1464	-999.0	TAC Security
708	-999.0	FableStreet
709	-999.0	Monsoon Fintech
710	-999.0	MonkeyBox
1463	-999.0	Fitternity
2166	-999.0	STAGE 3
713	-999.0	Creator\\€\™s Gurukul
2465	-999.0	LoanCircle
1461	-999.0	Zzungry

In [17]:



10 startups with most funding

df[['AmountinUSD', 'StartupName']].sort_values(by='AmountinUSD', ascending=True).tail(10)

Out[17]:

	AmountinUSD	StartupName
33	5.850000e+08	Udaan
188	6.000000e+08	True North
2459	6.800000e+08	Paytm
2648	7.000000e+08	Flipkart.com
31	1.000000e+09	Paytm
966	1.400000e+09	Flipkart
830	1.400000e+09	Paytm
83	1.510950e+09	Pratilipi
651	2.500000e+09	Flipkart
60	3.900000e+09	Rapido Bike Taxi

In [18]:



#Total Startups Funding

print("Total Startups funding : ",len(df['StartupName'].unique()))

Total Startups funding : 2457

In [19]:

```
startupname=df['StartupName'].value_counts().head(10)
startupname
```

Out[19]:

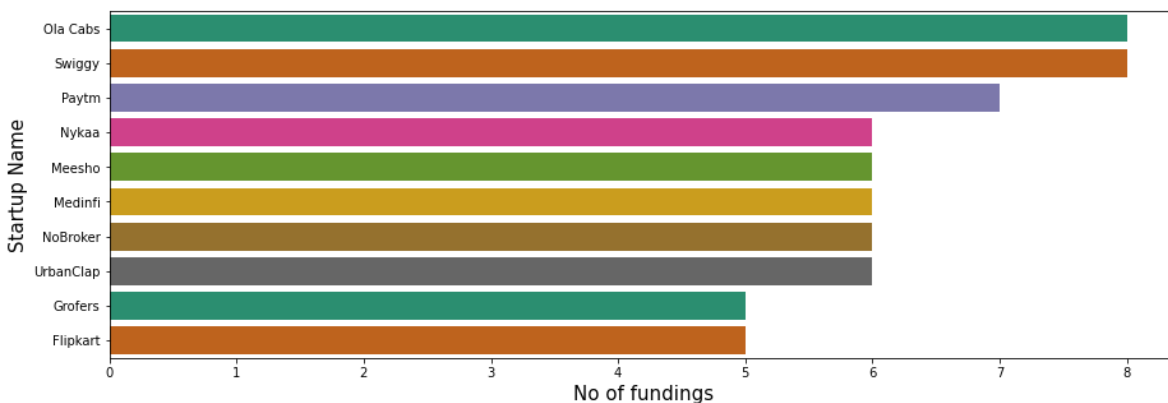
```
Ola Cabs      8
Swiggy        8
Paytm         7
Nykaa         6
Meesho        6
Medinfi       6
NoBroker      6
UrbanClap     6
Grofers       5
Flipkart      5
Name: StartupName, dtype: int64
```

In [30]:

```
#barplot
plt.rcParams['figure.figsize'] = (15, 5)
sns.barplot(y=startupname.index,x=startupname.values,palette='Dark2')
plt.ylabel("Startup Name",fontsize=15)
plt.xlabel("No of fundings",fontsize=15)
```

Out[30]:

```
Text(0.5, 0, 'No of fundings')
```



3. WHAT KIND OF INDUSTRIES ARE PREFERRED MORE FOR STARTUPS

In [31]:

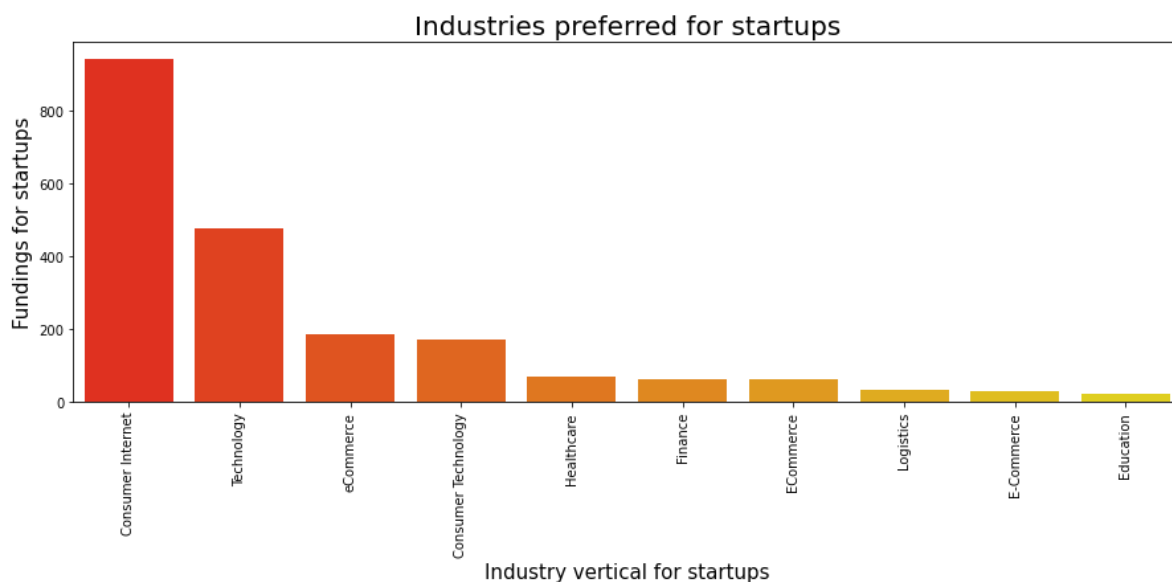
```
df['IndustryVertical']=df['IndustryVertical'].replace('nan','Consumer Technology')

Industry=df['IndustryVertical'].value_counts().head(10)
print(Industry)
```

```
Consumer Internet      941
Technology             478
eCommerce             186
Consumer Technology    172
Healthcare             70
Finance                62
ECommerce             61
Logistics             32
E-Commerce            29
Education             24
Name: IndustryVertical, dtype: int64
```

In [33]:

```
#plotting for industry vertical
plt.rcParams['figure.figsize'] = (15, 5)
sns.barplot(x=Industry.index,y=Industry.values,palette='autumn')
plt.xticks(rotation='90')
plt.title("Industries preferred for startups",fontsize=20)
plt.xlabel("Industry vertical for startups",fontsize=15)
plt.ylabel("Fundings for startups",fontsize=15)
plt.show()
```



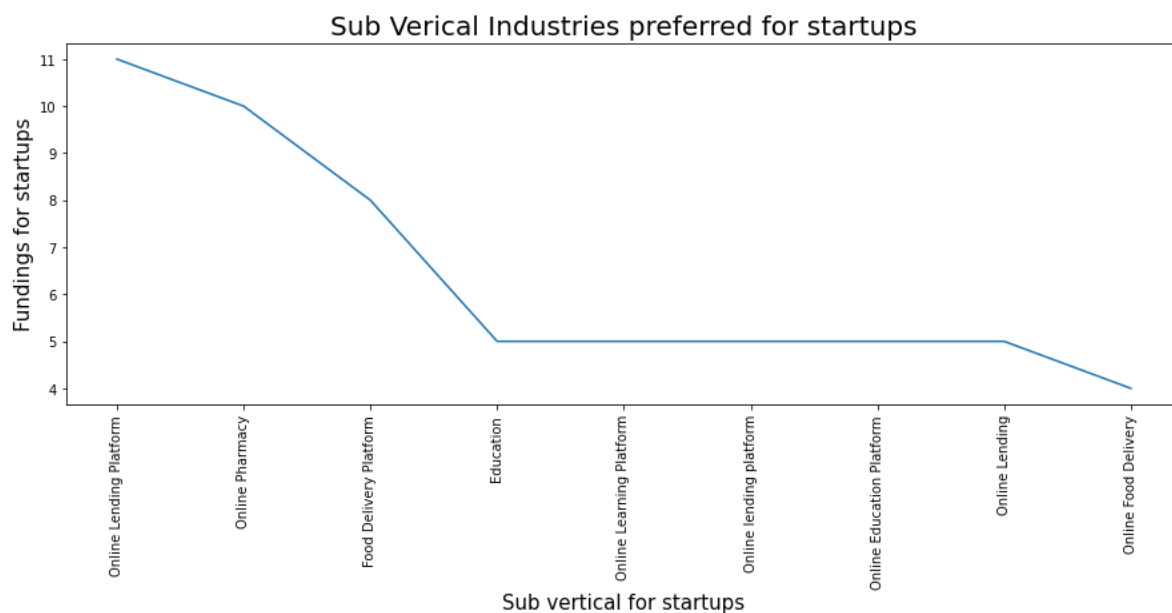
In [39]:

```
#plotting for Subvertical
subind=df['SubVertical'].value_counts().head(10)
subvertical=subind[1:]
print(subvertical)
```

```
Online Lending Platform      11
Online Pharmacy             10
Food Delivery Platform      8
Education                   5
Online Learning Platform    5
Online lending platform     5
Online Education Platform   5
Online Lending              5
Online Food Delivery        4
Name: SubVertical, dtype: int64
```

In [44]:

```
plt.rcParams['figure.figsize'] = (15, 5)
sns.lineplot(x=subvertical.index,y=subvertical.values,palette='winter')
plt.xticks(rotation='90')
plt.title("Sub Verical Industries preferred for startups",fontsize=20)
plt.xlabel("Sub vertical for startups",fontsize=15)
plt.ylabel("Fundings for startups",fontsize=15)
plt.show()
```



4.DOES LOCATION PLAY IMPORTNT ROLE ON INDIAN STARTUPS

In [46]:

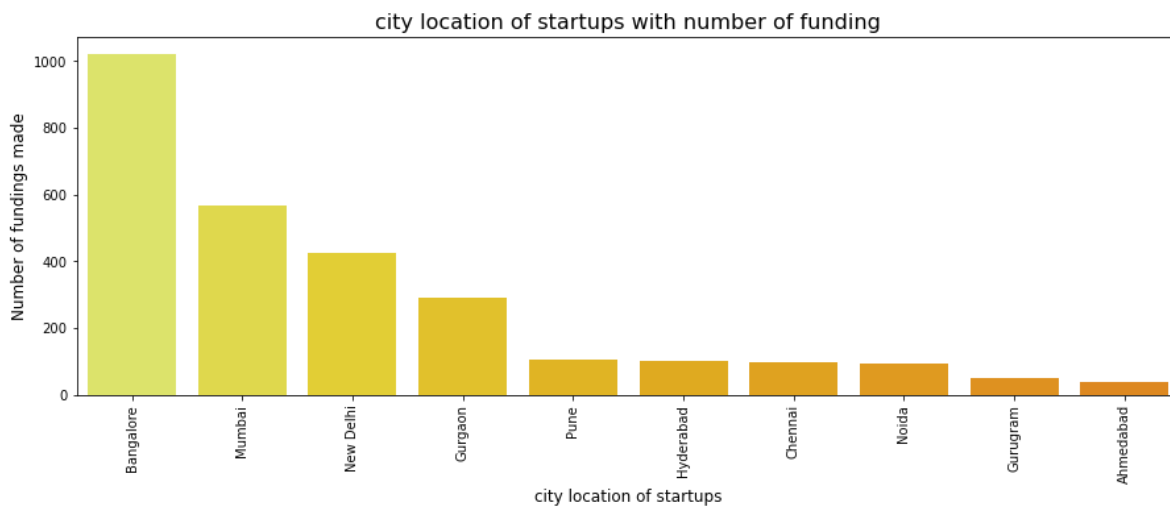
```
df['City'] = df['City'].replace(('Bengaluru', 'nan'), ('Bangalore', 'Bangalore'))

city = df['City'].value_counts().head(10)
print(city)
```

```
Bangalore    1022
Mumbai        568
New Delhi    424
Gurgaon       291
Pune          105
Hyderabad      99
Chennai        97
Noida          93
Gurugram       50
Ahmedabad      38
Name: City, dtype: int64
```

In [47]:

```
# plot
sns.barplot(city.index, city.values, palette = 'Wistia')
plt.xticks(rotation='vertical')
plt.xlabel('city location of startups', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("city location of startups with number of funding", fontsize=16)
plt.show()
```



5.WHO PLAYS THE MAIN ROLE IN INDIAN STARTUPS

In [52]:

```
df['InvestorName'][df['InvestorName'] == 'Undisclosed investors'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'undisclosed Investors'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'undisclosed investors'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'Undisclosed investor'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'Undisclosed Investor'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'Undisclosed'] = 'Undisclosed Investors'
df['InvestorName'][df['InvestorName'] == 'nan'] = 'Undisclosed Investors'
```

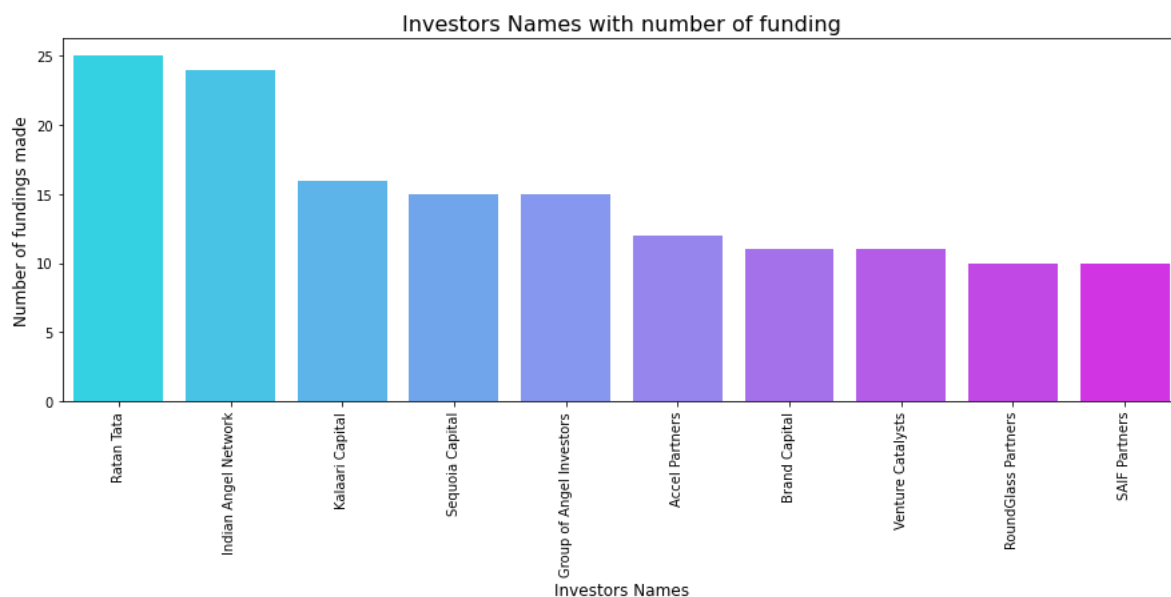
In [54]:

```
# value counts
investors = df['InvestorName'].value_counts()[1:].head(10)
print(investors)

# plot the data
sns.barplot(investors.index, investors.values, palette = 'cool')
plt.xticks(rotation='vertical')
plt.xlabel('Investors Names', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("Investors Names with number of funding", fontsize=16)
plt.show()
```

Ratan Tata	25
Indian Angel Network	24
Kalaari Capital	16
Sequoia Capital	15
Group of Angel Investors	15
Accel Partners	12
Brand Capital	11
Venture Catalysts	11
RoundGlass Partners	10
SAIF Partners	10

Name: InvestorName, dtype: int64



6. WHAT ARE THE DIFFERENT TYPES PF FUNDING FOR

STARTUPS

In [56]:



```
investment = df['InvestmentType'].value_counts().head(10)
print(investment)
```

```
Private Equity      1356
Seed Funding        1355
Seed/ Angel Funding    60
Seed / Angel Funding  47
Seed\\nFunding       30
Debt Funding         25
Series A             24
Seed/Angel Funding   23
Series B             20
Series C             14
Name: InvestmentType, dtype: int64
```

In [62]:



```
# Lets clean the dataset
df['InvestmentType'][df['InvestmentType'] == 'SeedFunding'] = 'Seed Funding'
df['InvestmentType'][df['InvestmentType'] == 'Crowd funding'] = 'Crowd Funding'
df['InvestmentType'][df['InvestmentType'] == 'PrivateEquity'] = 'Private Equity'

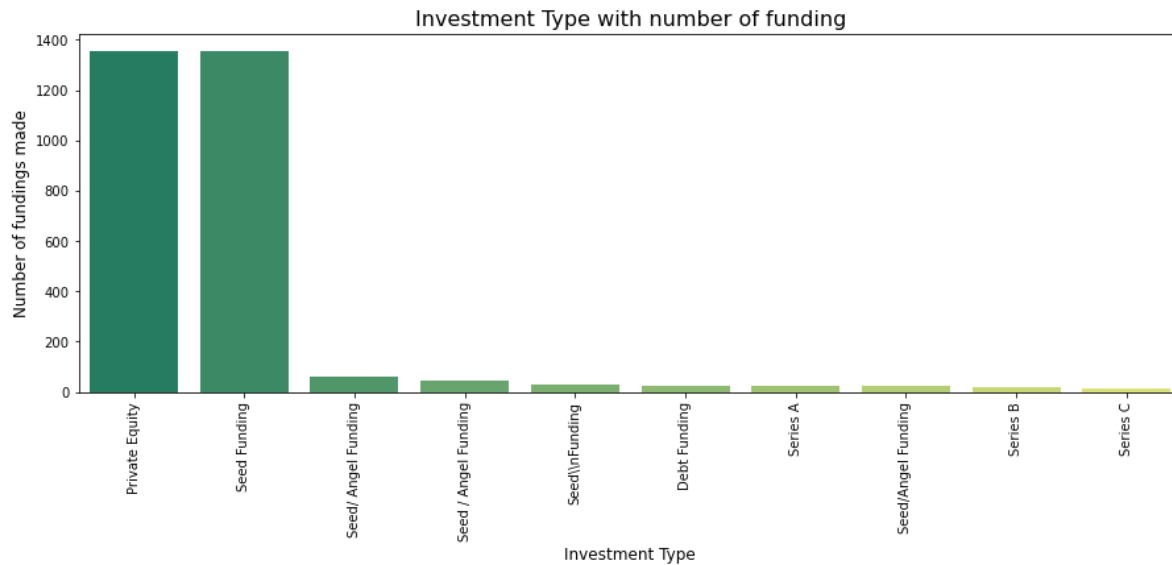
investment = df['InvestmentType'].value_counts().head(10)
print(investment)
```

```
Private Equity      1357
Seed Funding        1355
Seed/ Angel Funding    60
Seed / Angel Funding  47
Seed\\nFunding       30
Debt Funding         25
Series A             24
Seed/Angel Funding   23
Series B             20
Series C             14
Name: InvestmentType, dtype: int64
```

In [63]:



```
sns.barplot(investment.index, investment.values, palette = 'summer')
plt.xticks(rotation='vertical')
plt.xlabel('Investment Type', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("Investment Type with number of funding", fontsize=16)
plt.show()
```



In []:

