# Startup Funding Ecosystem

ANALYSE THE DATA ABOUT START-UP VENTURES, INVESTORS
INVESTMENT TRENDS AND VISUALIZE THE DATA

Nithin Gollanapally (SJSU Id: 013820599) | CMPE256 | 08/09/2019

# Introduction

Nowadays, all high growth technology startups have emerged from no more than 3-4 startup ecosystems, namely Silicon Valley and Boston, but now this trend seems to have reached its end. And moreover, with a global expansion of entrepreneurship has been an explosion in the process of rising new startup ecosystems around the globe.

For the new information economy high growth startups look to be the primary growth engine, the recent developments that are taking place in this ecosystem has big consequences for the future global economy.

Lately, India has been very active in creating a healthy startup ecosystem, and yearly there are increased number of startups growing. In fact, it is one of the top five startup communities around the world. Being so startup friendly, India has attracted numerous investors from both national and international which resulted in large amounts of money being poured into the startup ecosystem. This report basically tries to analyse the data about Indian startup ventures, investment trends and funding related ecosystem.

# Problem Statement

The objective of this project was to give possible answers for the questions below:

What is the Indian funding ecosystem trend with time?

How much funding does Indian startups generally get?

What type of industries does generally investors favour?

Does location play a major role in funding?

Who are all the important investors in the ecosystem?

What are different types of funding for startups?

# Implementation

## Dataset :

The data was scraped from the website trak.in. The data consists of over 500 startups across India which includes funding information of the Indian startups from January 2015 to August 2017. The data is in csv format.

<u>Input data format:</u>

The csv data contains columns with the StartupName, date founded, IndustryVertical, SubVertical, the city the startup is based out of, the names of the funders and the amount invested in USD.

## Data collection and Data Preprocessing :

The data obtained was inconsistent and data cleaning was necessary. The missing data has been dropped and cleaned data with the removal of the unnecessary punctuations and characters from the attributes. Also, as a part of the pre-processing process, the stop words has been removed. After this, it is converted to dataframe (pandas) with required features.

```python
# missing data
total = funding_data.isnull().sum().sort_values(ascending = False)
percent = ((funding_data.isnull().sum()/funding_data.isnull().count())*100).sort_values(ascending = False)
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent %'])
missing_data.head()
```

|  | Total | Percent % |
|---|---|---|
| Remarks | 1953 | 82.335582 |
| SubVertical | 936 | 39.460371 |
| AmountInUSD | 847 | 35.708263 |
| CityLocation | 179 | 7.546374 |
| IndustryVertical | 171 | 7.209106 |

```python
funding_data["AmountInUSD"] = funding_data["AmountInUSD"].apply(lambda x: float(str(x).replace(",","")))
funding_data["AmountInUSD"] = pd.to_numeric(funding_data["AmountInUSD"])
funding_data.head()
```

## Implementation :

**Question 1:** What is the funding ecosystem trend with time?

```python
funding_data['Date'][funding_data['Date']=='12/05.2015'] = '12/05/2015'
funding_data['Date'][funding_data['Date']=='13/04.2015'] = '13/04/2015'
funding_data['Date'][funding_data['Date']=='15/01.2015'] = '15/01/2015'
funding_data['Date'][funding_data['Date']=='22/01//2015'] = '22/01/2015'
funding_data["yearmonth"] = (pd.to_datetime(funding_data['Date'],format='%d/%m/%Y').dt.year*100)+(pd.to_datetime(funding_dat
temp = funding_data['yearmonth'].value_counts().sort_values(ascending = False).head(10)
print("Number of funding per month in decreasing order(Top 10)\n",temp)
year_month = funding_data['yearmonth'].value_counts()
plt.figure(figsize=(15,8))
sns.barplot(year_month.index, year_month.values, alpha=0.9, color=color[0])
plt.xticks(rotation='vertical')
plt.xlabel('Year-Month of transaction', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("Year-Month Distribution", fontsize=16)
plt.show()
```

Year-Month Distribution
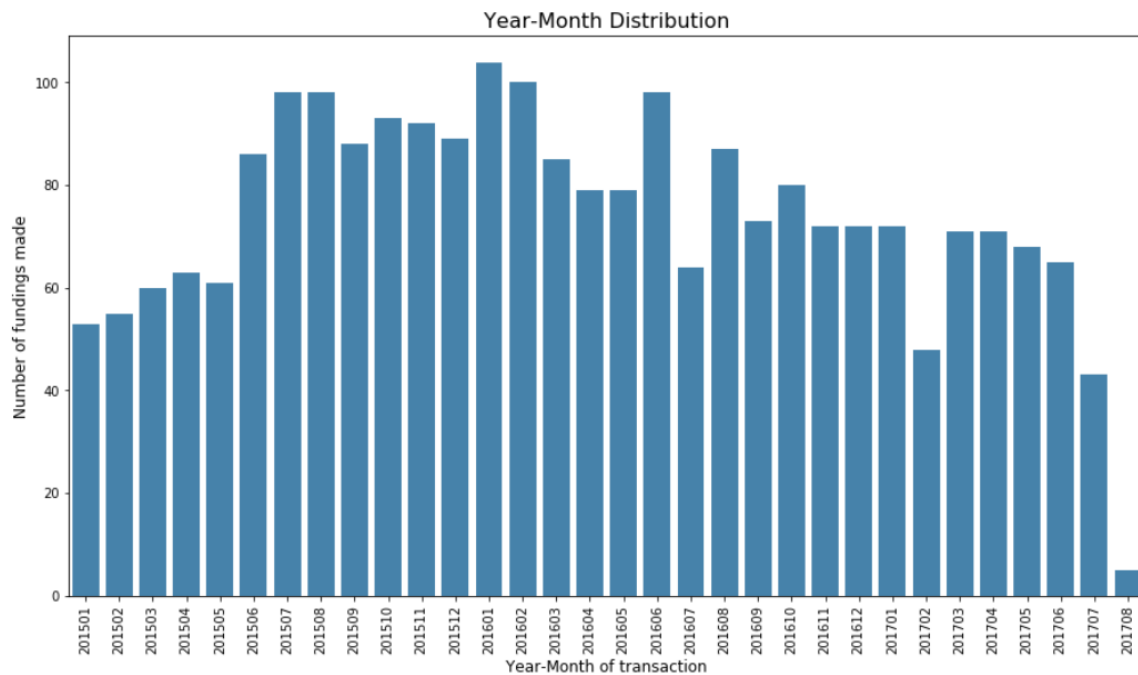
So, from the above result it is evident that startups got more funding in the month of January 2016. The above visualization shows funding variation from one month to another.

**Question 2:** How much funding does Indian startups generally get?

So, let's try to answer this question in the form of average funding, minimum funding, maximum funding and number of fundings.

Maximum Funding :

```
print("Maximum funding to a Startups is : ",funding_data["AmountInUSD"].dropna().sort_values().max())
```

Maximum funding to a Startups is :  1400000000.0

```
funding_data[funding_data.AmountInUSD == 1400000000.0]
```

| | SNo | Date | StartupName | IndustryVertical | SubVertical | CityLocation | InvestorsName | InvestmentType | AmountInUSD | yearmonth |
|---|---|---|---|---|---|---|---|---|---|---|
| 158 | 158 | 18/05/2017 | Paytm | ECommerce | Mobile Wallet & ECommerce platform | Bangalore | SoftBank Group | Private Equity | 1.400000e+09 | 201705 |
| 294 | 294 | 21/03/2017 | Flipkart | eCommerce | ECommerce Marketplace | Bangalore | Microsoft, eBay, Tencent Holdings | Private Equity | 1.400000e+09 | 201703 |

```
funding_data[funding_data.StartupName == 'Paytm']
```

| | SNo | Date | StartupName | IndustryVertical | SubVertical | CityLocation | InvestorsName | InvestmentType | AmountInUSD | yearmonth |
|---|---|---|---|---|---|---|---|---|---|---|
| 158 | 158 | 18/05/2017 | Paytm | ECommerce | Mobile Wallet & ECommerce platform | Bangalore | SoftBank Group | Private Equity | 1.400000e+09 | 201705 |
| 821 | 821 | 30/8/2016 | Paytm | eCommerce | Mobile Wallet & ECommerce platform | Bangalore | MediaTek Inc. | Private Equity | 6.000000e+07 | 201608 |
| 1787 | 1787 | 29/09/2015 | Paytm | E-Commerce & M-Commerce platform | NaN | New Delhi | Alibaba Group, Ant Financial | Private Equity | 6.800000e+08 | 201509 |
| 2218 | 2218 | 13/03/2015 | Paytm | NaN | NaN | NaN | Ratan Tata | Private Equity | NaN | 201503 |
| 2276 | 2276 | 05/02/2015 | Paytm | NaN | NaN | NaN | Ant Financial Services (Alipay) | Private Equity | NaN | 201502 |

## Minimum Funding:

```
print("Minimum funding to a Startups is : ",funding_data["AmountInUSD"].dropna().sort_values().min())
```
```
Minimum funding to a Startups is :  16000.0
```

```
funding_data[funding_data.AmountInUSD == 16000.0]
```

| | SNo | Date | StartupName | IndustryVertical | SubVertical | CityLocation | InvestorsName | InvestmentType | AmountInUSD | yearmonth |
|---|---|---|---|---|---|---|---|---|---|---|
| 2345 | 2345 | 19/01/2015 | Hostel Dunia | NaN | NaN | NaN | Hyderabad Angels (at Startup Heroes event) | Seed Funding | 16000.0 | 201501 |
| 2346 | 2346 | 19/01/2015 | Play your sport | NaN | NaN | NaN | Hyderabad Angels (at Startup Heroes event) | Seed Funding | 16000.0 | 201501 |
| 2347 | 2347 | 19/01/2015 | Yo Grad | NaN | NaN | NaN | Hyderabad Angels (at Startup Heroes event) | Seed Funding | 16000.0 | 201501 |
| 2348 | 2348 | 19/01/2015 | Enabli | NaN | NaN | NaN | Hyderabad Angels (at Startup Heroes event) | Seed Funding | 16000.0 | 201501 |
| 2349 | 2349 | 19/01/2015 | CBS | NaN | NaN | NaN | Hyderabad Angels (at Startup Heroes event) | Seed Funding | 16000.0 | 201501 |

## Average Funding:

```
print("On Average indian startups got funding of : ",funding_data["AmountInUSD"].dropna().sort_values().mean())
```
```
On Average indian startups got funding of :  12031073.099016393
```

```
print("Total startups funded : ", len(funding_data["StartupName"].unique()))
print(funding_data["StartupName"].value_counts().head(10))
startupname = funding_data['StartupName'].value_counts().head(20)
plt.figure(figsize=(15,8))
sns.barplot(startupname.index, startupname.values, alpha=0.9, color=color[0])
plt.xticks(rotation='vertical')
plt.xlabel('Startup Name', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("Number of funding a startup got", fontsize=16)
plt.show()
```
```
Total startups funded :  2001
Swiggy          7
UrbanClap       6
Medinfi         5
NoBroker        5
Jugnoo          5
Paytm           5
Flipkart        4
Capital Float   4
TinyStep        4
HolaChef        4
Name: StartupName, dtype: int64
```
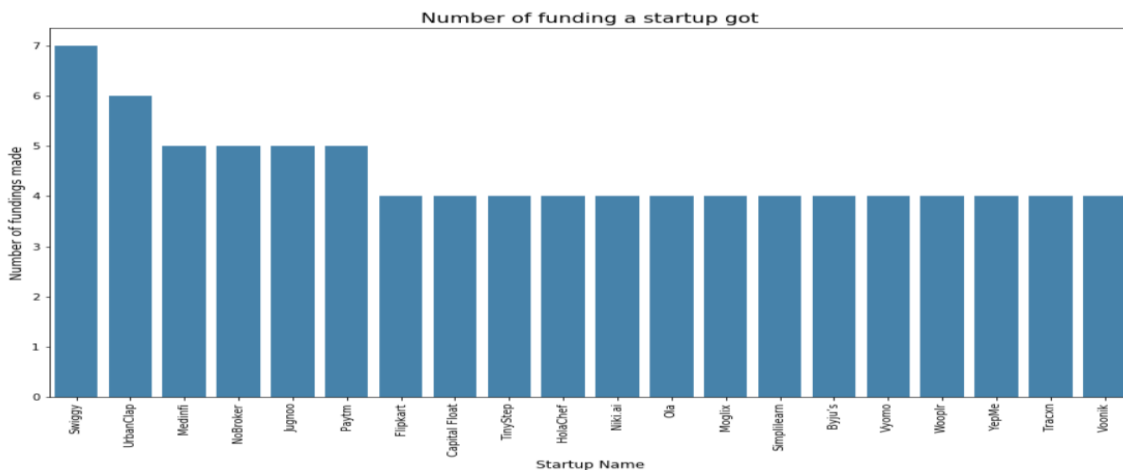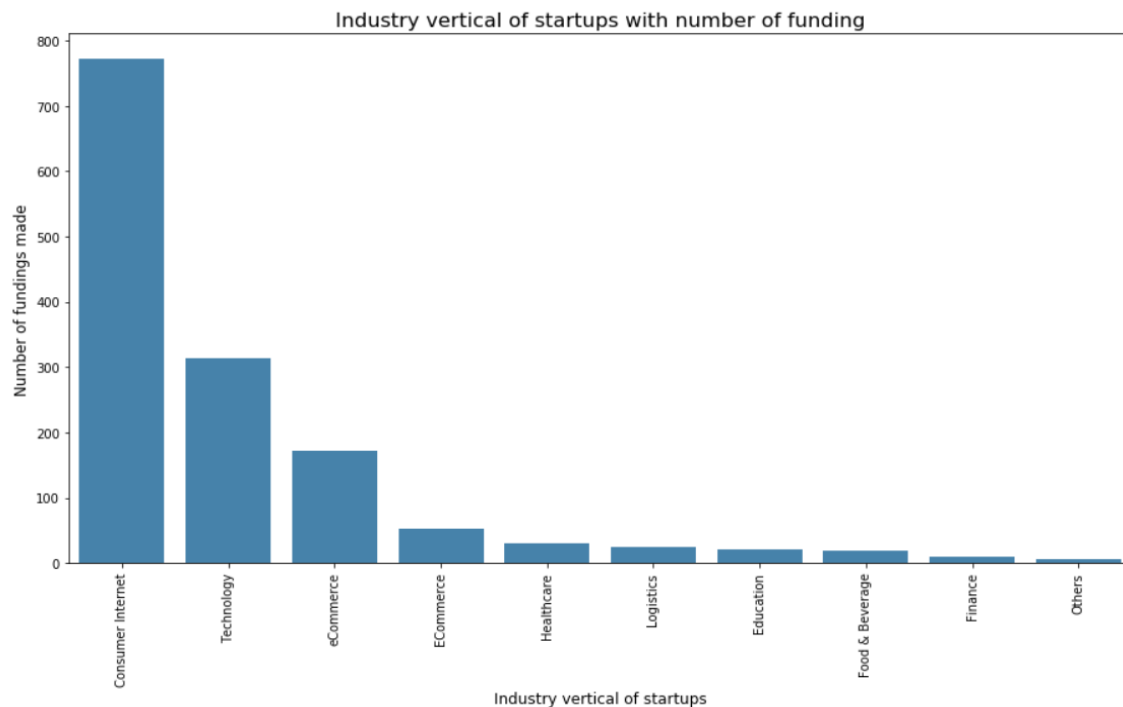


From the above result, Swiggy got maximum number of fundings (7) and there are 2001 Indian startups funded in the given timeframe.

**Question 3:** What type of industries does generally investors favour?

```
industry = funding_data['IndustryVertical'].value_counts().head(10)
print(industry)
plt.figure(figsize=(15,8))
sns.barplot(industry.index, industry.values, alpha=0.9, color=color[0])
plt.xticks(rotation='vertical')
plt.xlabel('Industry vertical of startups', fontsize=12)
plt.ylabel('Number of fundings made', fontsize=12)
plt.title("Industry vertical of startups with number of funding", fontsize=16)
plt.show()
```



Industry vertical of startups with number of funding

From the above result, Consumer Internet got maximum number of fundings (772) followed by technology and E-Commerce.

**Question 4:** Does location play a major role in funding?

```
city = funding_data['CityLocation'].value_counts().head(10)
print(city)
plt.figure(figsize=(15,8))
sns.barplot(city.index, city.values, alpha=0.9, color=color[0])
plt.figure(figsize=(15,8))
count = funding_data['CityLocation'].value_counts()
squarify.plot(sizes=count.values,label=count.index, value=count.values)
plt.title('Distribution of Startups across Top cities')
```

Distribution of Startups across Top cities
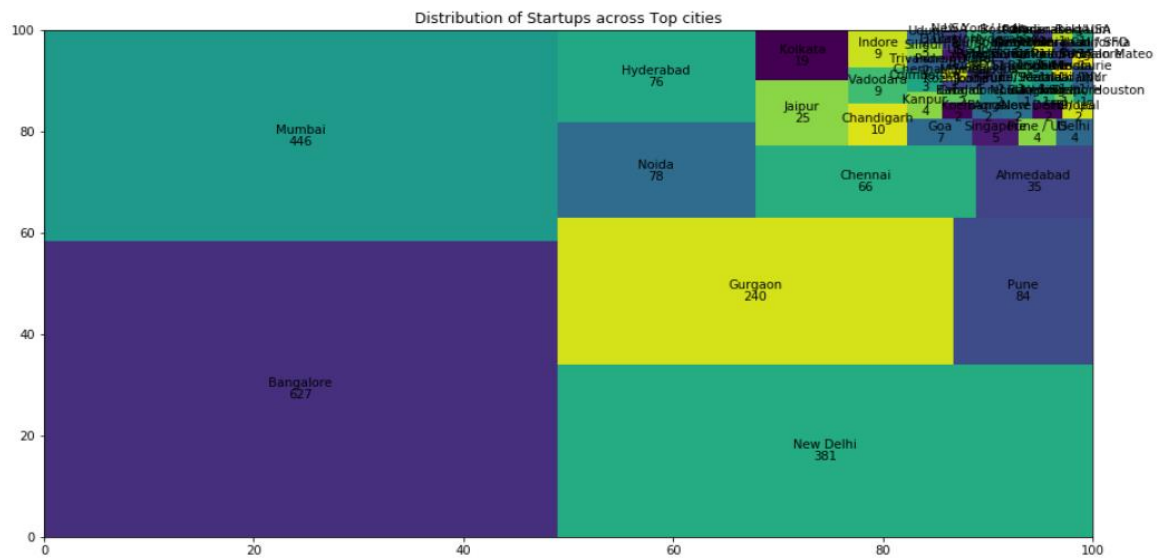
From the above results, we can say that Bangalore attracts a greater number of investors followed by Mumbai and New Delhi.

**Question 5:** Who are all the important investors in the ecosystem?

```python
from wordcloud import WordCloud

names = funding_data["InvestorsName"][~pd.isnull(funding_data["InvestorsName"])]
#print(names)
wordcloud = WordCloud(max_font_size=50, width=600, height=300).generate(' '.join(names))
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'Undisclosed investors'] = 'Undisclosed Investors'
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'undisclosed Investors'] = 'Undisclosed Investors'
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'undisclosed investors'] = 'Undisclosed Investors'
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'Undisclosed investor'] = 'Undisclosed Investors'
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'Undisclosed Investor'] = 'Undisclosed Investors'
funding_data['InvestorsName'][funding_data['InvestorsName'] == 'Undisclosed'] = 'Undisclosed Investors'
investors = funding_data['InvestorsName'].value_counts().head(10)
print(investors)
plt.figure(figsize=(15,8))
sns.barplot(investors.index, investors.values, alpha=0.9, color=color[0])
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()
```



Investors Names with number of funding
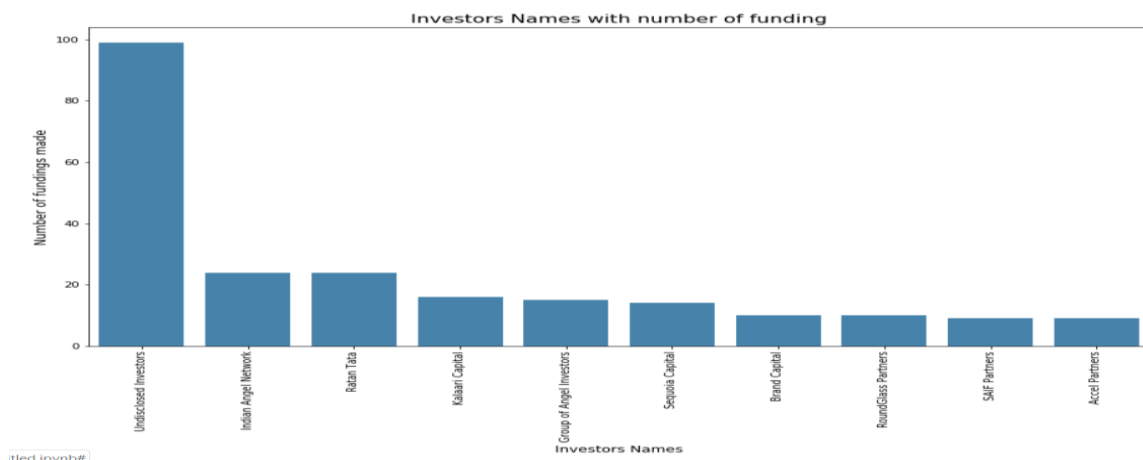
From the above result, we can say that **Indian Angel network** and **Ratan tata** have funded maximum number of startups followed by **Kalaari Caitals**.

**Question 6:** What are the different types of funding for startups ?

```python
investment = funding_data['InvestmentType'].value_counts()
funding_data['InvestmentType'][funding_data['InvestmentType'] == 'SeedFunding'] = 'Seed Funding'
funding_data['InvestmentType'][funding_data['InvestmentType'] == 'Crowd funding'] = 'Crowd Funding'
funding_data['InvestmentType'][funding_data['InvestmentType'] == 'PrivateEquity'] = 'Private Equity'
investment = funding_data['InvestmentType'].value_counts()
print(investment)
plt.figure(figsize=(15,8))
sns.barplot(investment.index, investment.values, alpha=0.9, color=color[0])
temp = funding_data["InvestmentType"].value_counts()
labels = temp.index
sizes = (temp / temp.sum())*100
trace = go.Pie(labels=labels, values=sizes, hoverinfo='label+percent')
layout = go.Layout(title='Types of investment funding with %')
data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="BorrowerGender")
```

Types of investment funding with %
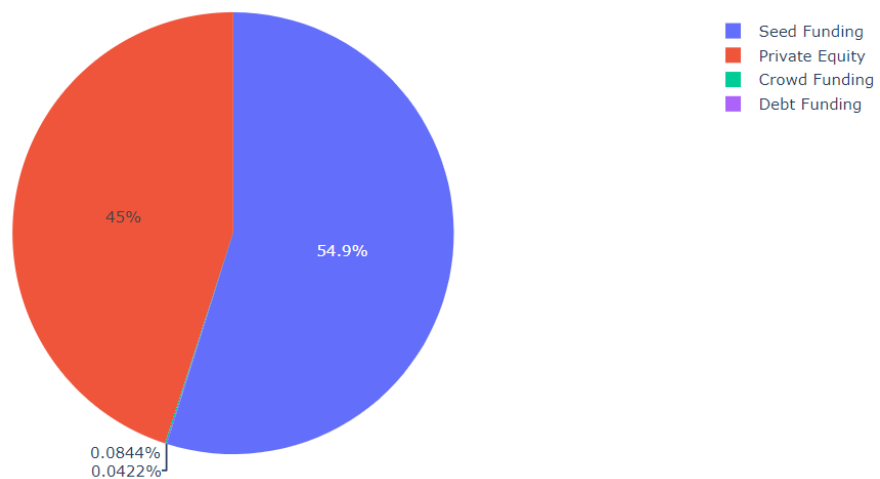


From the above result, we can say that **Seed Funding** is in **Top** followed by Private Equity.

## Conclusion:

For startup investments, the year 2015 seems to be a boom period but this has seen a downward trend in the year 2016.

In this project, I have tried to focus on analysing the dataset from different angles and focused on finding the top players in the startup space as far as the investments are concerned.