INDIAN STARTUP ECOSYSTEM EDA ANALYTICS PROJECT

If you want to be the first to be informed about new projects, please do not forget to follow us - by Fatma Nur AZMAN

Fatmanurazman.com | Linkedin | Github | Kaggle | Tableau

Table of Contents

- Understanding The Data
  - Project Description
  - About the Datasets
  - Import The Libraries
  - Performing Essential Statistical Analysis on the Dataset
  - Data Cleaning and Preparation
  - Feature Engineering
- How Does the Funding Ecosystem changes with respect to Time?
- What is the General Amount that Startups get in India?
- Which Kind of Industries are more preferred for Startups?
- Does Location also play a role, In determining the Growth of a Startup?
- Who plays the main role in Indian Startups Ecosystem?
- What are the different Types of Funding for Startups?
- Recommendations for New Investors

Understanding The Data

Project Description:

- In the evolving landscape of startups, understanding the dynamics of funding is crucial. This dataset provides comprehensive insights into the startup funding ecosystem in India, capturing various aspects of funding activities over the years. From the types of investments to the key players and preferred industries, this dataset offers a rich tapestry of information that can help stakeholders make informed decisions and identify emerging trends.

- **Purpose:**

- **Project Importance:**

Examining this project helps new investors make more informed and strategic decisions in the Indian startup ecosystem. Data-driven insights and visualizations enable investors to minimize risks and capitalize on high-potential opportunities.

About the Datasets

Dataset Descriptions: 'stocks_daily_prices.csv' / 'stocks_daily_returns.csv'

- **Content:** Daily stock prices for various companies.
- **Rows:** 3044
- **Columns:** 10
  - **Sr No:** Serial number.A unique identifier for each record.
  - **Date dd/mm/yyyy:** The date when the funding event took place.
  - **Startup Name:** The name of the startup receiving the funding.
  - **Industry Vertical:** The primary industry to which the startup belongs.
  - **SubVertical:** A more specific category within the primary industry.
  - **City Location:** The city where the startup is headquartered.
  - **Investors Name:** The names of the investors or investment firms involved in the funding.
  - **InvestmentnType:** The type of investment (e.g., Seed, Series A, Series B).
  - **Amount in USD:** The amount of funding received in US dollars.
  - **Remarks:** Additional comments or details about the funding event.

Import The Libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import seaborn as sns
import folium
from folium.plugins import MarkerCluster
from wordcloud import WordCloud

# Enable inline plotting
%matplotlib inline

import warnings

warnings.filterwarnings("ignore")
warnings.warn("this will not show")
```

```
df = df0.copy()
```

Performing Essential Statistical Analysis on the Dataset

In [3]:
```
# Dimensions of the Data Set - (rows, columns)
df.shape
```

Out[3]: (3044, 10)

In [4]:
```
# Preview of Data Set
df.head()
```

Out[4]:

	Sr No	Date dd/mm/yyyy	Startup Name	Industry Vertical	SubVertical	City Location	Inve N
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 Gr E
2	3	09/01/2020	Mamaearth	E-commerce	Retailer of baby and toddler products	Bengaluru	Se 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	Embroiled Clothes For Women	Mumbai	S Ve Par

In [5]:
```
# Data Type Properties
df.info()
```

```
 0   Sr No               3044 non-null   int64
 1   Date dd/mm/yyyy     3044 non-null   object
 2   Startup Name        3044 non-null   object
 3   Industry Vertical   2873 non-null   object
 4   SubVertical         2108 non-null   object
 5   City  Location      2864 non-null   object
 6   Investors Name      3020 non-null   object
 7   InvestmentnType     3040 non-null   object
 8   Amount in USD       2084 non-null   object
 9   Remarks             419 non-null    object
dtypes: int64(1), object(9)
memory usage: 237.9+ KB
```

In [6]:
```
df.describe(include='object').T
```

Out[6]:

	count	unique	top	freq
Date dd/mm/yyyy	3044	1035	02/02/2015	11
Startup Name	3044	2459	Ola Cabs	8
Industry Vertical	2873	821	Consumer Internet	941
SubVertical	2108	1942	Online Lending Platform	11
City Location	2864	112	Bangalore	700
Investors Name	3020	2412	Undisclosed Investors	39
InvestmentnType	3040	55	Private Equity	1356
Amount in USD	2084	471	10,00,000	165
Remarks	419	72	Series A	175

In [7]:
```
# Checking Null Values
(df.isnull().sum() / df.shape[0] *
 100).sort_values(ascending=False).round(2).astype(str) + ' %'
```

Out[7]:
```
Remarks              86.24 %
Amount in USD        31.54 %
SubVertical          30.75 %
City  Location         5.91 %
Industry Vertical     5.62 %
Investors Name        0.79 %
InvestmentnType       0.13 %
Sr No                  0.0 %
Date dd/mm/yyyy        0.0 %
Startup Name           0.0 %
dtype: object
```

Data Cleaning and Preparation

Output: to do

for creating charts.

- The 'Remarks' property contains about 86.23% empty values, so we can remove it.

In [42]:
```
#Changing commas in the 'Amount in USD' column
df['Amount in USD'] = df['Amount in USD'].apply(
    lambda x: str(x).replace(',', ''))
```

In [43]:
```
# Correction of incorrect values from 'Amount in USD' column
replace_map = {
    "undisclosed": "0",
    "Undisclosed": "0",
    "unknown": "0",
    "14342000+": "0",
    "\\\\xc2\\\\xa010000000": "0",
    "\\\\xc2\\\\xa05000000": "0",
    "\\\\xc2\\\\xa019350000": "0",
    "\\\\xc2\\\\xa0600000": "0",
    "\\\\xc2\\\\xa020000000": "0",
    "\\\\xc2\\\\xa0N/A": "0",
    "\\\\xc2\\\\xa016200000": "0",
    "\\\\xc2\\\\xa0685000": "0",
    "nan": "0"
}

df['Amount in USD'] = df['Amount in USD'].apply(lambda x: replace_map.get(str(x), x))
```

In [44]:
```
# Conversion to digital data
df['Amount in USD'] = pd.to_numeric(df['Amount in USD'])
```

In [11]:
```
# Replacing 'Amount in USD' 0 values with empty values
df['Amount in USD'] = df['Amount in USD'].replace(0, np.nan)
```

In [45]:
```
# Replace empty values with average
df['Amount in USD'].fillna(df['Amount in USD'].mean(), inplace=True)
```

In [46]:
```
# Correcting incorrect date values
date_replace_map = {
    '12/05.2015': '12/05/2015',
    '13/04.2015': '13/04/2015',
    '15/01.2015': '15/01/2015',
    '22/01//2015': '22/01/2015',
    '05/072018': '05/07/2018',
    '01/07/015': '01/07/2015',
    '\\xc2\\xa010/7/2015': '10/07/2015',
    '\\\\xc2\\\\xa010/7/2015': '10/07/2015'
}

df['Date dd/mm/yyyy'] = df['Date dd/mm/yyyy'].apply(lambda x: date_replace_map.get(x, x
```

In [47]:
```
# Convert to datetime type by specifying the date format
df['Date dd/mm/yyyy'] = pd.to_datetime(df['Date dd/mm/yyyy'],
```

```
df.drop('Remarks', axis=1, inplace=True)
```

In [48]:
```
# Replacing 'Bengaluru' used in the data set with the more common name 'Bangalore'
df['City  Location'][df['City  Location'] ==
                         'Bengaluru'] = 'Bangalore'
```

In [49]:
```
# Change the name in the 'Undisclosed investors' column to 'Undisclosed Investors'
investor_replace_map = {
    'Undisclosed investors': 'Undisclosed Investors',
    'Undisclosed Investor': 'Undisclosed Investors',
    'undisclosed investors': 'Undisclosed Investors',
    'Undisclosed': 'Undisclosed Investors'
}

df['Investors Name'] = df['Investors Name'].apply(lambda x: investor_replace_map.get(x,
```

In [50]:
```
# Removal of the gap in 'Ola Cabs'.
df['Startup Name'][df['Startup Name'] == 'Ola Cabs'] = 'OlaCabs'
```

In [51]:
```
# Replace with a more commonly used word
investment_type_replace_map = {
    'Seed/ Angel Funding': 'Seed / Angel Funding',
    'Seed\\nFunding': 'Seed Funding',
    'Seed/Angel Funding': 'Seed / Angel Funding',
    'Angel / Seed Funding': 'Seed / Angel Funding'
}

df['InvestmentnType'] = df['InvestmentnType'].apply(lambda x: investment_type_replace_m
```

In [52]:
```
# Standardizing common industry terms using regex and string replacement
replacements = {
    r'\be[ -]?commerce\b': 'e-commerce',
    r'\bfintech\b': 'fintech',
    r'\bhealth[ -]?tech\b': 'healthtech',
    r'\bedu[ -]?tech\b': 'edtech',
    r'\bfood[ -]?(tech|delivery)\b': 'food & beverage',
    r'\btransportation|logistics\b': 'transportation & logistics',
    r'\bconsumer internet\b': 'consumer internet',
    r'\btechnology\b': 'technology',
    r'\bagri[ -]?tech\b': 'agritech',
    r'\bauto[ -]?tech\b': 'autotech',
    r'\bmedia\b': 'media',
    r'\bfinance\b': 'finance',
    r'\bunknown\b': 'other'
}

# Applying replacements
for pattern, replacement in replacements.items():
    df['Industry Vertical'] = df['Industry Vertical'].str.replace(pattern, replacement,
```

In [53]:
```
#Fill in missing values in the 'Industry Vertical' column
df['Industry Vertical'] = df['Industry Vertical'].fillna('unknown')

#Convert all inputs to lowercase
```

```
    cleaned_parts = []
    for part in parts:
        if part not in cleaned_parts:
            cleaned_parts.append(part)
        if len(cleaned_parts) == 2:
            break
    return ' & '.join(cleaned_parts)

df['Industry Vertical'] = df['Industry Vertical'].apply(clean_industry)
```

Feature Engineering

In [54]:
```
# Create the 'Year Month' column
df['Year Month'] = (df['Date dd/mm/yyyy'].dt.year *
                          100) + df['Date dd/mm/yyyy'].dt.month

# Let's check that the conversion was successful
df[['Date dd/mm/yyyy', 'Year Month']].head()
```

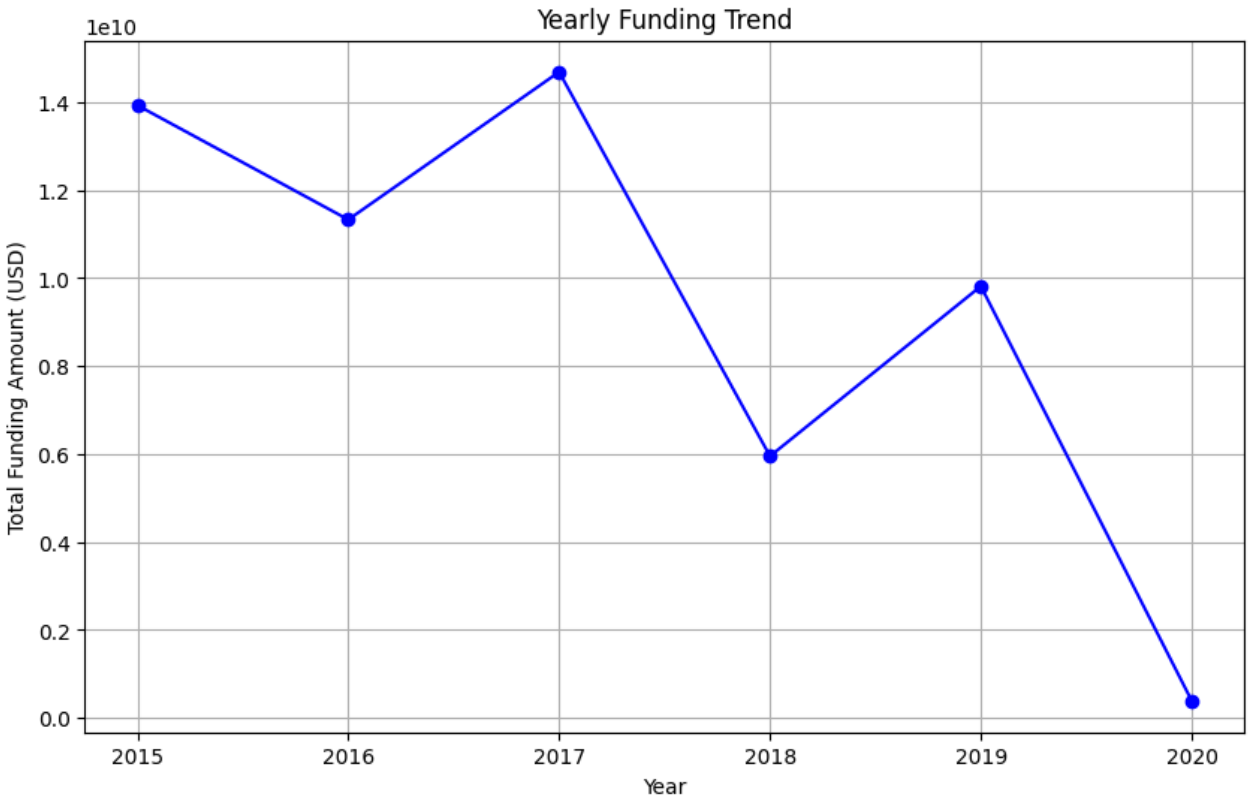
Out[54]:

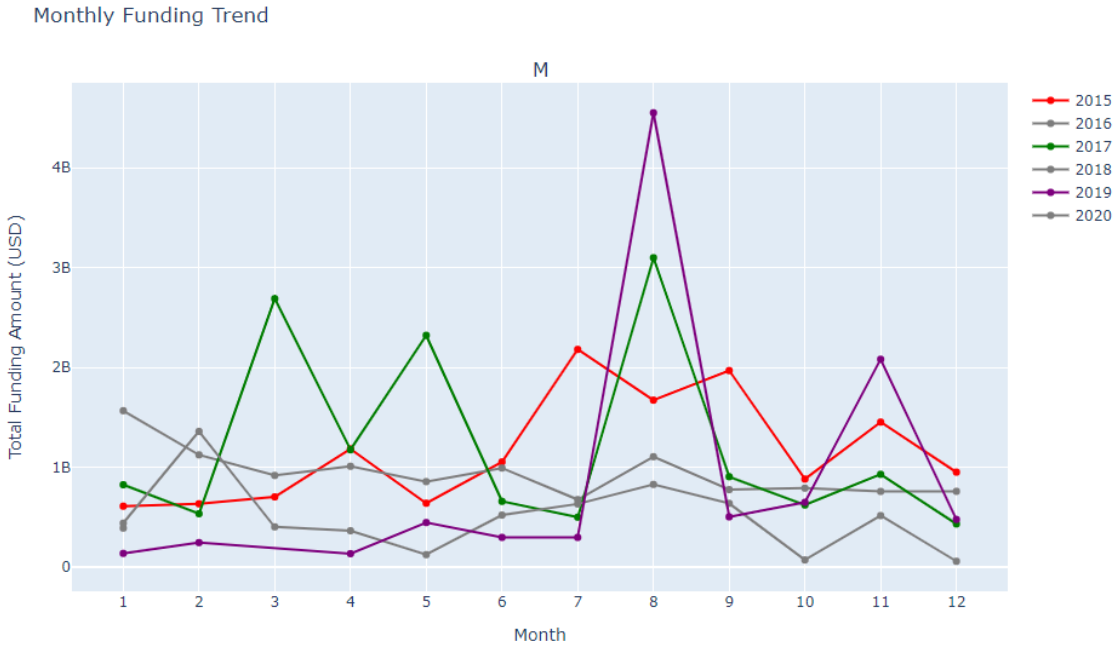
	Date dd/mm/yyyy	Year Month
0	2020-01-09	202001
1	2020-01-13	202001
2	2020-01-09	202001
3	2020-01-02	202001
4	2020-01-02	202001

How Does the Funding Ecosystem changes with respect to Time?

In [55]:
```
import pandas as pd
import matplotlib.pyplot as plt
import plotly.graph_objects as go
from plotly.subplots import make_subplots

# Convert 'Date' column to datetime format
df['Date'] = pd.to_datetime(df['Date dd/mm/yyyy'], format='%d/%m/%Y')

# Add Year and Month columns
df['Year'] = df['Date'].dt.year
df['Month'] = df['Date'].dt.month

# Yearly funding trend
funding_trend_yearly = df.groupby('Year')['Amount in USD'].sum().reset_index()
```

```
# Matplotlib ile Yıllık Fonlama Trendi Grafiği
plt.figure(figsize=(10, 6))
plt.plot(funding_trend_yearly['Year'], funding_trend_yearly['Amount in USD'], marker='o
plt.title('Yearly Funding Trend')
plt.xlabel('Year')
plt.ylabel('Total Funding Amount (USD)')
plt.grid(True)
plt.show()

# Plotly ile Aylık Fonlama Trendi Grafiği
fig = make_subplots(rows=1, cols=1, subplot_titles=("Monthly Funding Trend"))

# Monthly funding trend with selected years highlighted
colors = {2015: 'red', 2017: 'green', 2019: 'purple'}
for year in funding_trend_monthly['Year'].unique():
    monthly_data = funding_trend_monthly[funding_trend_monthly['Year'] == year]
    color = colors.get(year, 'gray')
    fig.add_trace(
        go.Scatter(x=monthly_data['Month'], y=monthly_data['Amount in USD'], mode='line
                   name=str(year), line=dict(color=color)),
        row=1, col=1
    )

# Update layout
fig.update_layout(title_text="Monthly Funding Trend", height=600)
fig.update_xaxes(title_text="Month", row=1, col=1)
fig.update_yaxes(title_text="Total Funding Amount (USD)", row=1, col=1)

# Show the plot
fig.show()
```





Output:Observation:

- A noticeable decline followed in 2018, and another rise was observed in 2019.
- The data for 2020 shows a steep decline, but it's likely incomplete due to the year being partial or ongoing.

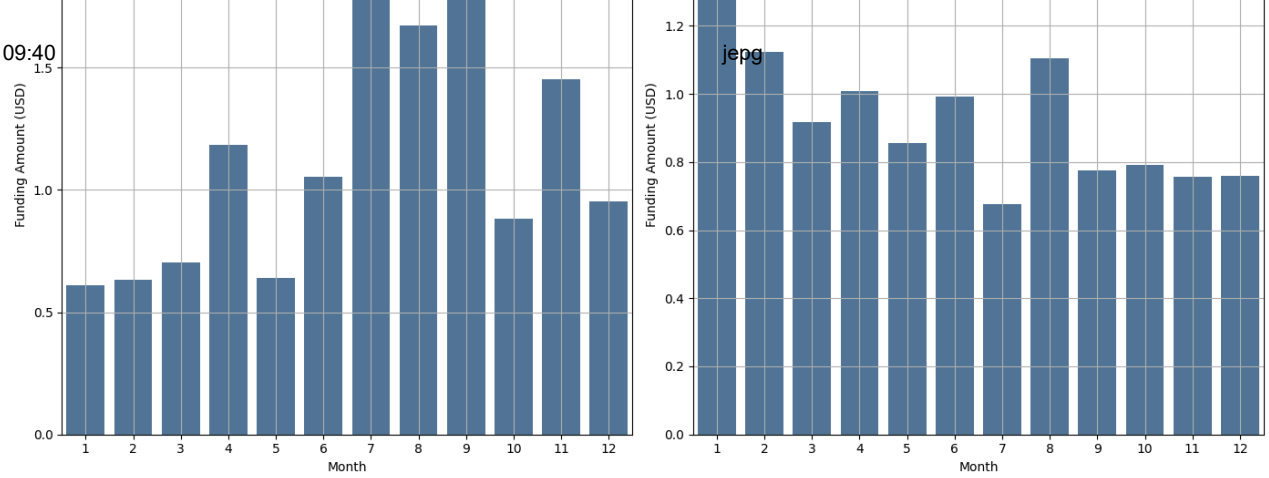
Possible Explanations:

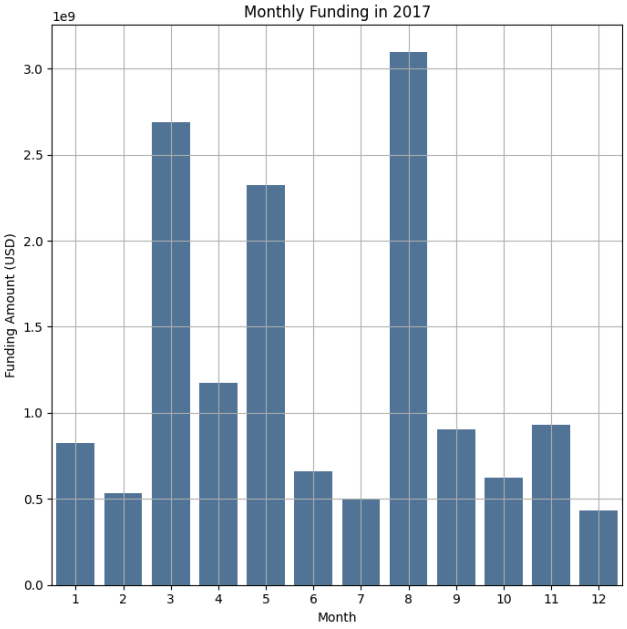
- Economic cycles, investor confidence, and macroeconomic factors could be influencing these fluctuations.
- Major funding rounds or significant investments in particular years can cause spikes.

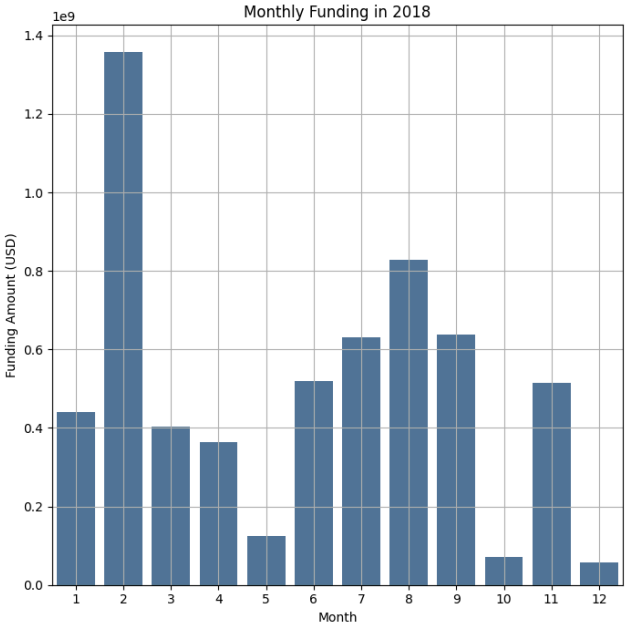
Monthly Funding Trend

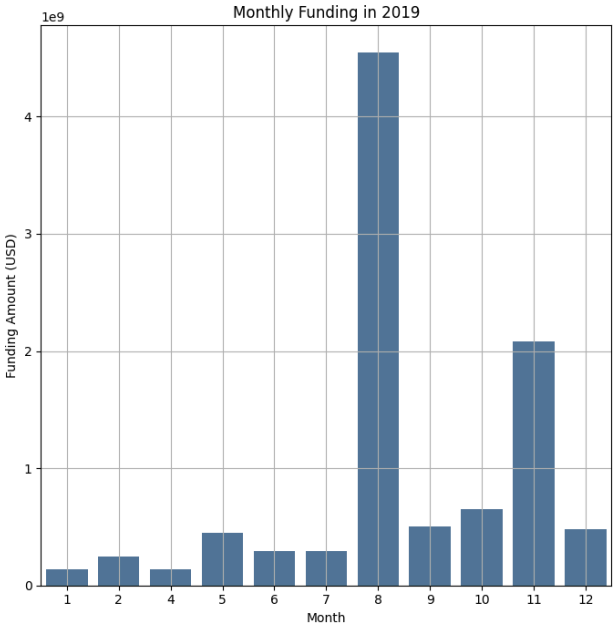
- The monthly funding trend reveals more granular details, with multiple peaks and troughs.
- There are significant spikes in certain months, notably in mid-2017 and mid-2019.
- There seems to be some seasonality, with certain periods consistently showing higher funding activities.

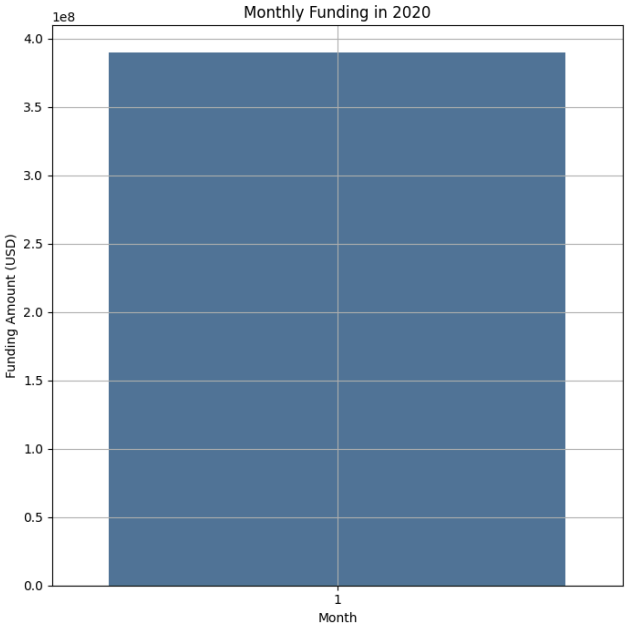
In [56]:
```
# Extract year and month for monthly funding analysis
df['Year'] = df['Date dd/mm/yyyy'].dt.year
df['Month'] =df['Date dd/mm/yyyy'].dt.month

# Group by Year and Month to get monthly funding amounts
monthly_funding = df.groupby(['Year', 'Month'])['Amount in USD'].sum().reset_index()

# Unique years in the dataset
years = monthly_funding['Year'].unique()

# Plotting monthly funding amounts for each year using subplots
fig, axs = plt.subplots(len(years) // 2, 2, figsize=(14, 7 * (len(years) // 2)))
color = '#4878A2'
for i, year in enumerate(years):
    row = i // 2
    col = i % 2
    sns.barplot(x='Month',
                y='Amount in USD',
                data=monthly_funding[monthly_funding['Year'] == year],
                color=color,
                ax=axs[row, col])
    axs[row, col].set_title(f'Monthly Funding in {year}')
    axs[row, col].set_xlabel('Month')
    axs[row, col].set_ylabel('Funding Amount (USD)')
    axs[row, col].grid(True)

plt.tight_layout()
plt.show()
```











Output:Observation:

Seasonal Trends:

- Each year has different months with peaks, indicating that specific events or investments might drive funding activities.

Funding Peaks:

- Significant spikes in funding amounts can be attributed to large investment rounds or high-profile startups securing funding.

This analysis provides a clear view of how funding is distributed across months each year, highlighting key periods of investment activity.

What is the General Amount that Startups get in India?

In [25]:
```
# Preview of the Top 10 Most Funded Initiatives
df['Amount in USD'].sort_values(ascending=False).head(10)
```

Out[25]:
```
60      3.900000e+09
651     2.500000e+09
966     1.400000e+09
830     1.400000e+09
31      1.000000e+09
2648    7.000000e+08
2459    6.800000e+08
188     6.000000e+08
33      5.850000e+08
2244    5.000000e+08
Name: Amount in USD, dtype: float64
```

In [26]:
```
# Preview of the details of the 10 most funded Initiatives
df.sort_values(by='Amount in USD', ascending=False).head(5)
```

60	61	2019-08-27	Bike Taxi	transportation	Bike Taxi	Bangalore	Capital	Se
651	652	2017-08-11	Flipkart	ecommerce	Online Marketplace	Bangalore	Softbank	Private I
966	967	2017-03-21	Flipkart	ecommerce	ECommerce Marketplace	Bangalore	Microsoft, eBay, Tencent Holdings	Private I
830	831	2017-05-18	Paytm	ecommerce	Mobile Wallet & ECommerce platform	Bangalore	SoftBank Group	Private I
31	32	2019-11-25	Paytm	fintech	Mobile Wallet	Noida	Vijay Shekhar Sharma	Funding F

In [27]:
```
# Calculating the average funding received by a startup
df['Amount in USD'].mean()
```

Out[27]:
```
18429897.27080872
```

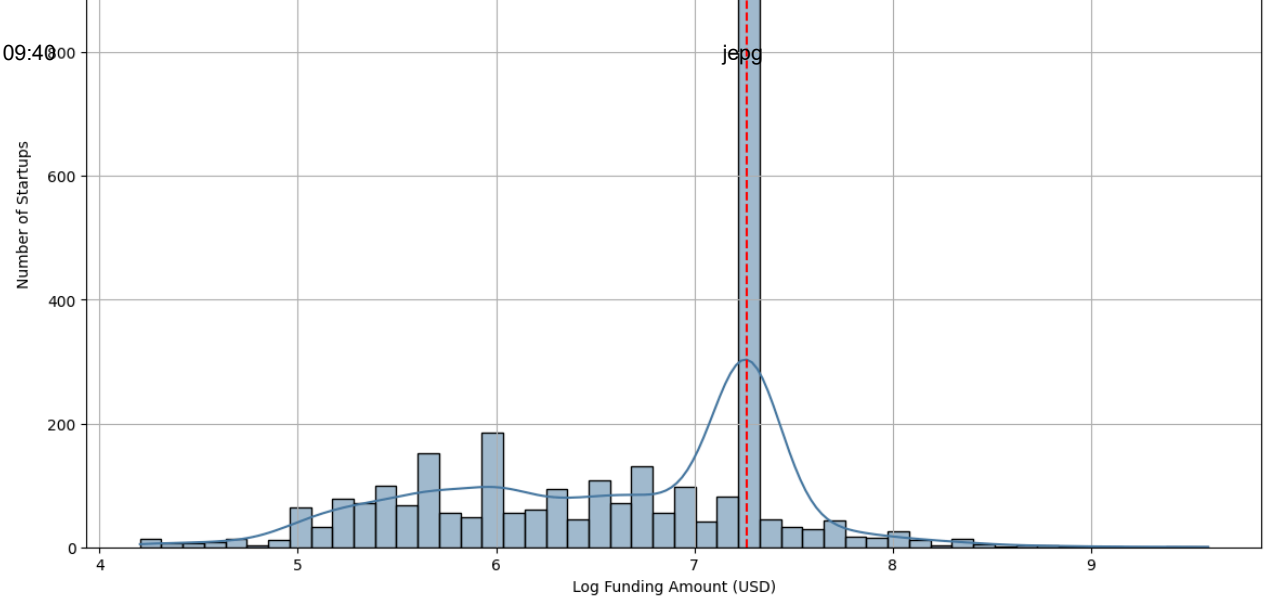
In [28]:
```
# Preview of the least funded initiatives
df['Amount in USD'].sort_values().head(10)
```

Out[28]:
```
3020    16000.0
3021    16000.0
3019    16000.0
3018    16000.0
3017    16000.0
2933    16600.0
2934    16600.0
2935    16600.0
2936    16600.0
2937    16600.0
Name: Amount in USD, dtype: float64
```

In [29]:
```
# Preview of the details of the least funded initiatives
df.sort_values(by='Amount in USD').head(5)
```

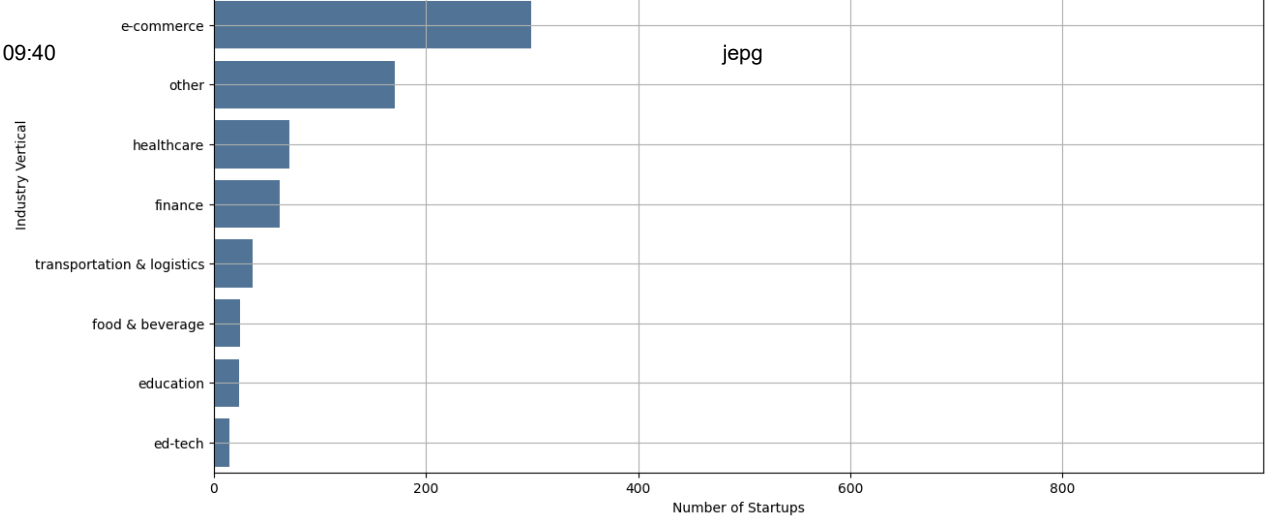
3020	3021	2015-01-19	Enabli	unknown	NaN	NaN	Angels (at Startup Heroes event)	Seed Fundin
3021	3022	2015-01-19	CBS	unknown	NaN	NaN	Hyderabad Angels (at Startup Heroes event)	Seed Fundin
3019	3020	2015-01-19	Yo Grad	unknown	NaN	NaN	Hyderabad Angels (at Startup Heroes event)	Seed Fundin
3018	3019	2015-01-19	Play your sport	unknown	NaN	NaN	Hyderabad Angels (at Startup Heroes event)	Seed Fundin
3017	3018	2015-01-19	Hostel Dunia	unknown	NaN	NaN	Hyderabad Angels (at Startup Heroes event)	Seed Fundin

In [57]:
```
# Calculate the average funding amount
average_funding = df['Amount in USD'].mean()

# Log-transform the funding amounts for better visualization
df['Log Amount in USD'] = np.log10(df['Amount in USD'] + 1)

# Plot the log-transformed funding amount distribution
plt.figure(figsize=(14, 8))
sns.histplot(df['Log Amount in USD'], bins=50, kde=True, color='#4878A2')
plt.axvline(np.log10(average_funding + 1), color='r', linestyle='--', label=f'Log Avera
plt.title('Log-Scaled Distribution of Funding Amounts for Startups in India')
plt.xlabel('Log Funding Amount (USD)')
plt.ylabel('Number of Startups')
plt.legend()
plt.grid(True)
plt.show()
```



Which Kind of Industries are more preferred for Startups?

In [58]:
```
# Identify the top 10 industries
top_industries = df[(df['Industry Vertical'] != 'unknown')]['Industry Vertical'].value_
top_industries.columns = ['Industry Vertical', 'Count']

# Create the bar plot
plt.figure(figsize=(14, 8))
sns.barplot(x='Count', y='Industry Vertical', data=top_industries, color='#4878A2')
plt.title('Top 10 Preferred Industries for Startups')
plt.xlabel('Number of Startups')
plt.ylabel('Industry Vertical')
plt.grid(True)
plt.show()
```



Output:Industry Preferences Analysis

Number of Funding Rounds per Industry

Top Industries:

- Consumer Internet: Leading with the highest number of funding rounds (589 rounds).
- Technology: Second most active with 310 funding rounds.
- E-commerce: Significant presence with 170 funding rounds.

Total Funding Amount per Industry

Top Funded Industries:

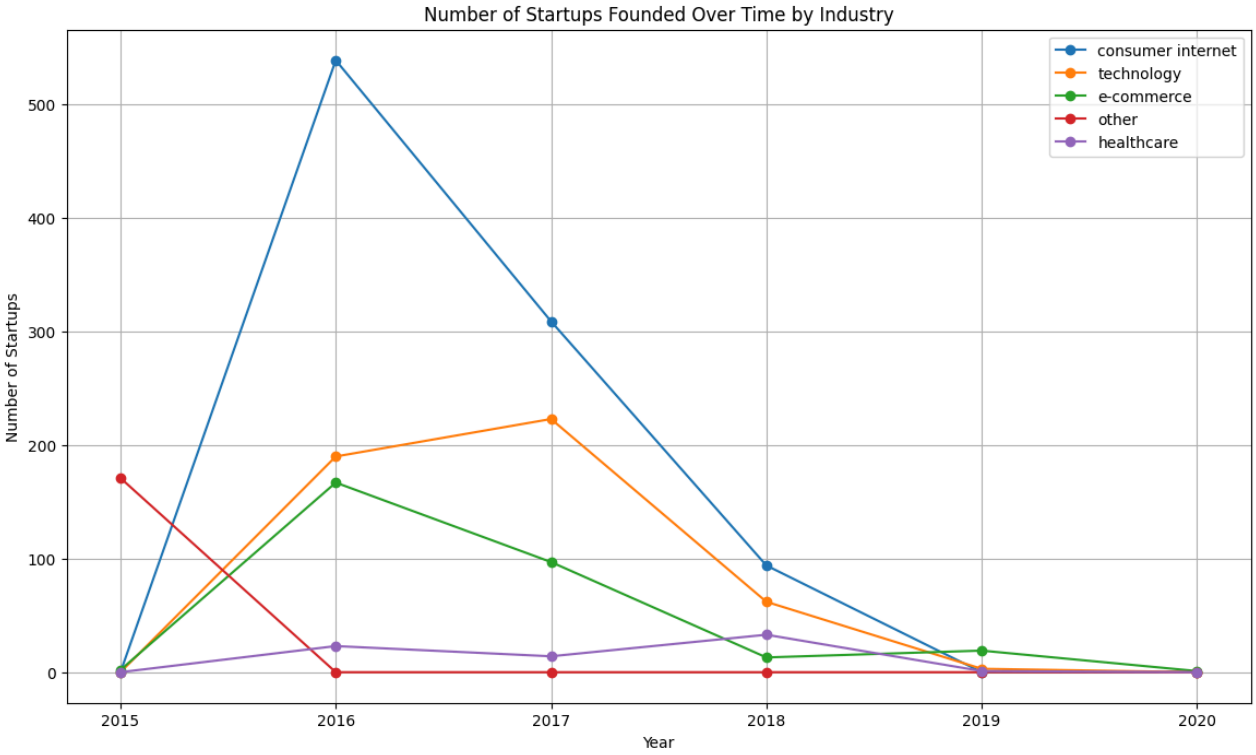
- E-commerce: Secured the highest total funding amount, indicating large investments in this sector ($7.16 billion).
- Consumer Internet: Close behind with substantial funding ($6.25 billion).
- Technology: Also received considerable funding ($2.23 billion).

Insights

- Active Sectors: Consumer Internet and Technology sectors are highly active in terms of funding rounds.
- High Investment Sectors: E-commerce and Consumer Internet attract the highest total funding, reflecting investor confidence and market potential in these sectors.
- Industry Dynamics: The analysis highlights which industries are more preferred by investors and which sectors secure larger investments.

In [59]:
```
# Ensure 'Date' column is in datetime format
df['Date'] = pd.to_datetime(df['Date dd/mm/yyyy'], format='%d/%m/%Y')

# Extract year and group by year and industry
df['Year'] = df['Date'].dt.year
yearly_industry_count = df.groupby(['Year', 'Industry Vertical']).size().unstack().fill
```

```
    plt.plot(yearly_industry_count.index, yearly_industry_count[industry], marker='o',
plt.title('Number of Startups Founded Over Time by Industry')
plt.xlabel('Year')
plt.ylabel('Number of Startups')
plt.legend()
plt.grid(True)
plt.show()
```



Output:Top 5 Industry Choice Analysis

Consumer Internet:

- Peak in 2016: The number of consumer internet startups saw a significant peak in 2016 with over 500 startups founded.
- Sharp Decline: After 2016, there is a sharp decline, indicating a reduction in new consumer internet startups over the subsequent years.

Technology:

- Steady Growth and Decline: Technology startups grew steadily, peaking in 2016 with around 200 startups, followed by a decline similar to the consumer internet trend.
- Consistency: Despite the decline, the number of technology startups remains relatively consistent compared to other industries.

E-Commerce:

- Initial Growth: E-commerce startups showed initial growth, peaking in 2016 with about 150 startups.

- Stability: The healthcare industry shows relative stability with slight fluctuations, peaking modestly in 2016 and maintaining a lower but consistent presence.

Insights

- 2016 as a Pivotal Year:

Most industries, especially consumer internet, technology, and e-commerce, peaked in 2016. This indicates a significant year for startup formations across these sectors. Post-2016, there is a noticeable decline in new startup formations, which could be due to market saturation, changing investment climates, or shifts in entrepreneurial focus.

- Consumer Internet and Technology Leading:

These two sectors have the highest peaks, indicating high interest and investment in these areas during their peak years. The sharp decline post-2016 suggests potential over-saturation or a shift in investor interest.

- Steady but Low Growth in Healthcare:

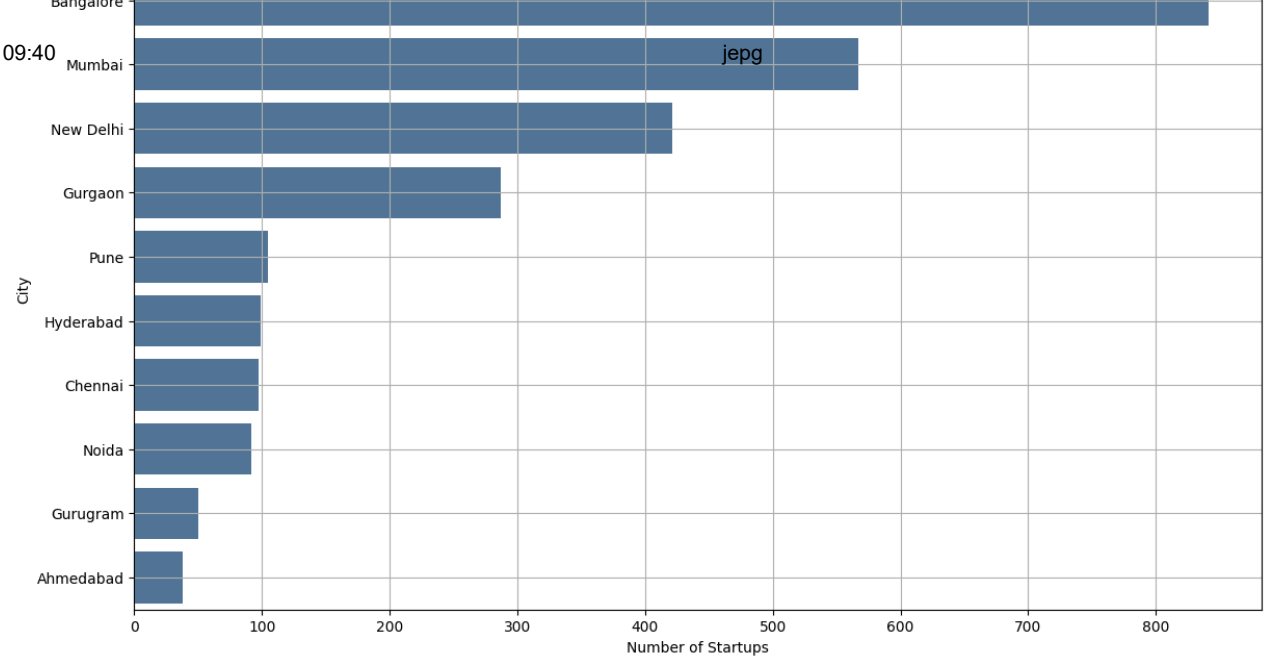
Healthcare startups show steady but lower growth compared to other sectors, suggesting a more stable but less explosive industry.

Potential Reasons for Trends:

- Economic Factors: Changes in the economic environment, funding availability, and investor sentiment could explain the peak and subsequent decline.
- Market Saturation: High initial growth could lead to market saturation, causing a drop in new startup formations in subsequent years.
- Shifts in Focus: Emerging technologies and changing market demands might shift entrepreneurial focus to other areas over time.

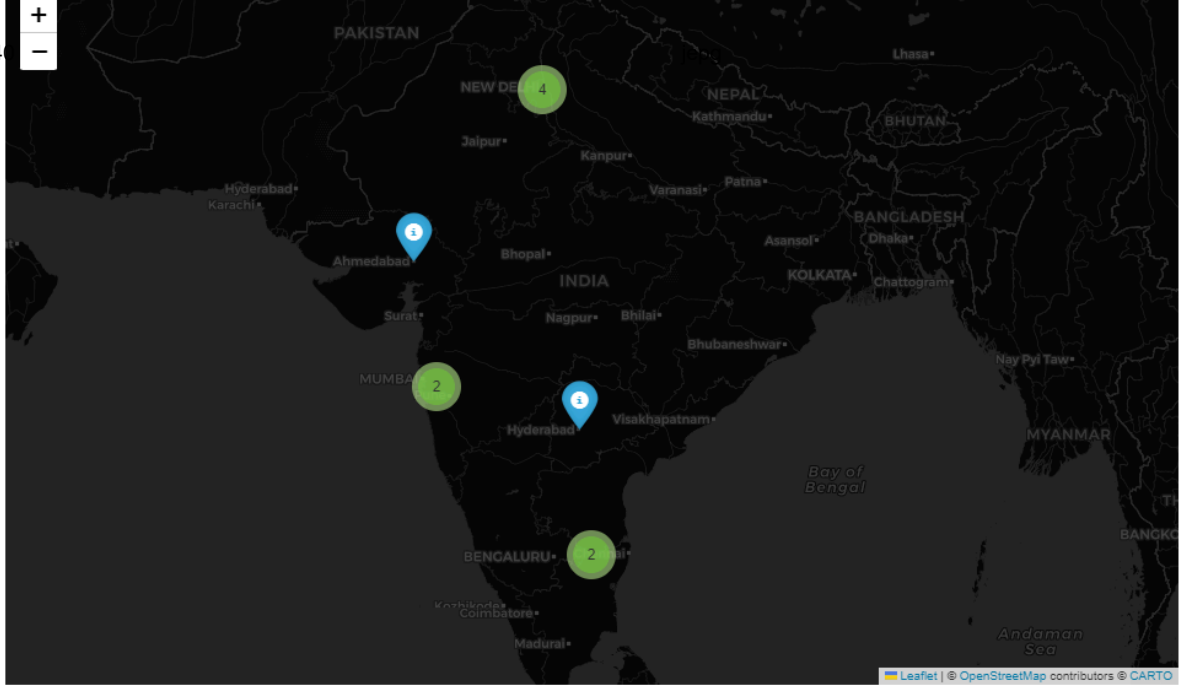
Does Location also play a role, In determining the Growth of a Startup?

In [60]:
```
# Count the number of startups in each city
top_cities_count = df['City  Location'].value_counts().head(10).reset_index()
top_cities_count.columns = ['City', 'Count']

# Plot the number of startups by city
plt.figure(figsize=(14, 8))
sns.barplot(x='Count', y='City', data=top_cities_count, color='#4878A2')
plt.title('Top 10 Cities by Number of Startups')
plt.xlabel('Number of Startups')
plt.ylabel('City')
```



In [ ]:

```
# Sample data: Coordinates for the cities
city_coords = {
    'City': ['Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon', 'Pune', 'Hyderabad', 'Chenn
    'Latitude': [12.9716, 19.0760, 28.6139, 28.4595, 18.5204, 17.3850, 13.0827, 28.5355
    'Longitude': [77.5946, 72.8777, 77.2090, 77.0266, 73.8567, 78.4867, 80.2707, 77.391
}

# Convert to DataFrame
city_coords_df = pd.DataFrame(city_coords)

# Calculate the number of startups
city_counts = df['City  Location'].value_counts().reset_index()
city_counts.columns = ['City', 'Count']

# Merge with coordinates
city_counts = city_counts.merge(city_coords_df, on='City', how='left')

# Filter out NaN values
city_counts = city_counts.dropna(subset=['Latitude', 'Longitude'])

# Select all cities in the first graph
selected_cities = city_counts[city_counts['City'].isin(city_coords['City'])]

# Create a Folium map with the "dark" map style
m = folium.Map(location=[20.5937, 78.9629], zoom_start=5, tiles='CartoDB dark_matter')

# Add markers to the map for each city with the number of startups
marker_cluster = MarkerCluster().add_to(m)

for idx, row in selected_cities.iterrows():
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=f"{row['City']}: {row['Count']} Startups",
        icon=folium.Icon(color='blue')
    ).add_to(marker_cluster)
```



Output:Analysis of the Top 10 Cities by Number of Startups

Bangalore as the Primary Hub:

- Bangalore's significant lead in the number of startups highlights its role as the primary tech and innovation hub in India. The city's infrastructure, talent pool, and supportive ecosystem attract a large number of startups.

Mumbai and New Delhi's Strong Presence:

- Mumbai and New Delhi's high ranks underscore their importance in the Indian startup ecosystem. Mumbai's financial prowess and New Delhi's political and incubator support contribute to their strong startup cultures.

Emergence of Other Cities:

- Cities like Gurgaon, Pune, and Hyderabad show significant numbers of startups, indicating the diversification of the startup ecosystem beyond the primary hubs. These cities offer favorable conditions such as talent availability, infrastructure, and government support.

Regional Clusters:

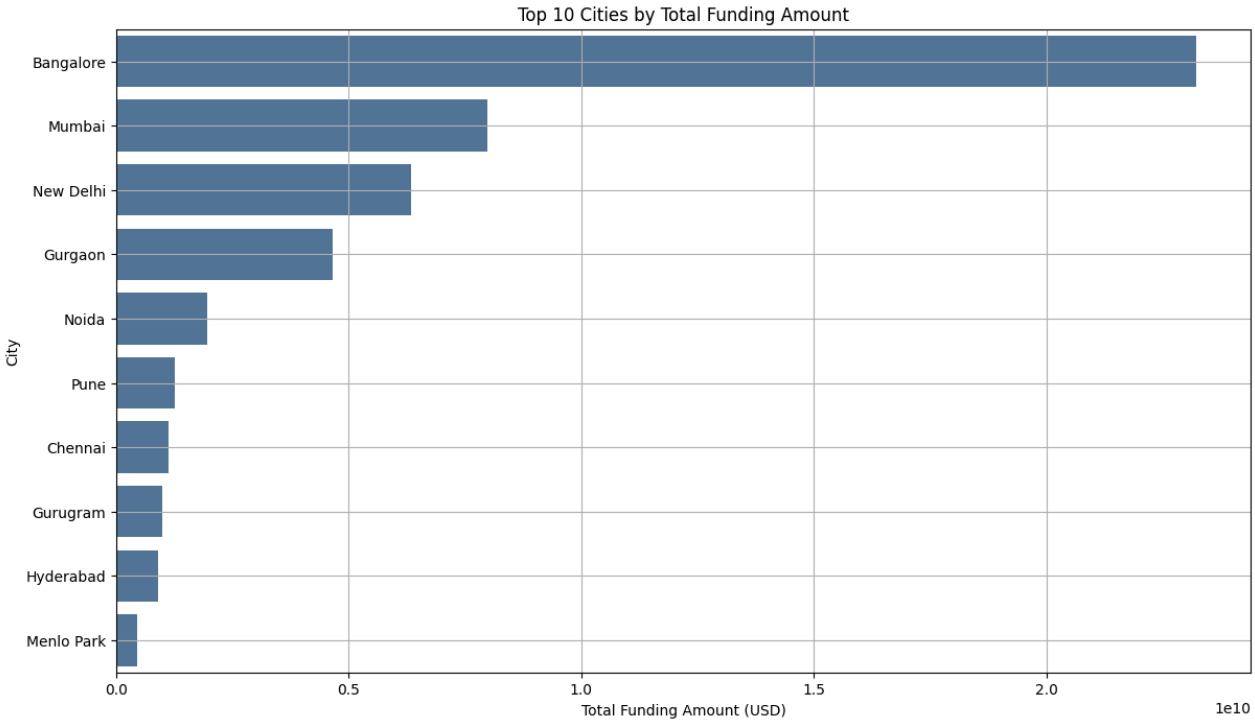
- The presence of multiple cities from the National Capital Region (NCR) like New Delhi, Gurgaon, and Noida highlights the region's attractiveness for startups. Proximity to the capital and good connectivity are key factors.

Supporting Infrastructure and Ecosystems:

Conclusion

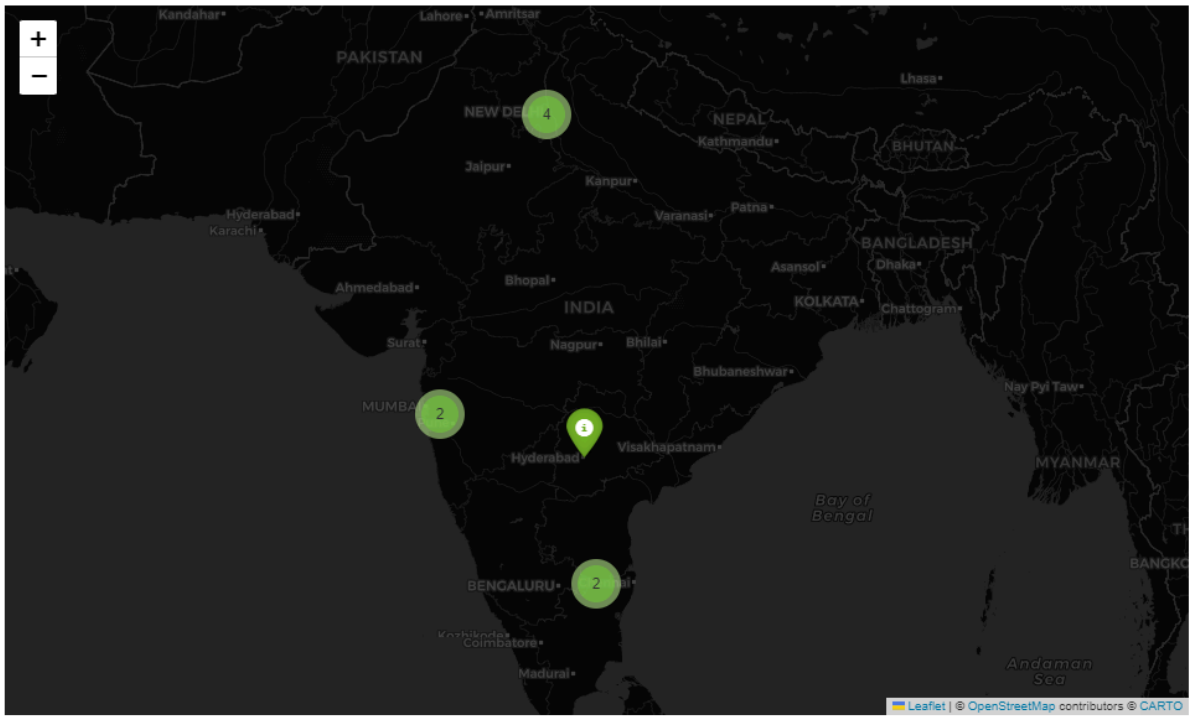
- The chart indicates that location significantly influences startup growth. Cities with strong ecosystems, infrastructure, and support systems tend to have higher concentrations of startups. Understanding these dynamics can help stakeholders, including investors, entrepreneurs, and policymakers, make informed decisions about where to focus their efforts and resources.

In [62]:
```
# Calculate the total funding amount received by startups in each city
top_cities_funding = df.groupby('City  Location')['Amount in USD'].sum().reset_index()
top_cities_funding = top_cities_funding.sort_values(by='Amount in USD', ascending=False
top_cities_funding.columns = ['City', 'Total Funding Amount']

# Plot the total funding amount by city
plt.figure(figsize=(14, 8))
sns.barplot(x='Total Funding Amount', y='City', data=top_cities_funding, color='#4878A2
plt.title('Top 10 Cities by Total Funding Amount')
plt.xlabel('Total Funding Amount (USD)')
plt.ylabel('City')
plt.grid(True)
plt.show()
```



In [ ]:
```
# Sample data: Coordinates for the cities
city_coords = {
    'City': ['Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon', 'Noida', 'Pune', 'Chennai',
    'Latitude': [12.9716, 19.0760, 28.6139, 28.4595, 28.5355, 18.5204, 13.0827, 28.4595
    'Longitude': [77.5946, 72.8777, 77.2090, 77.0266, 77.3910, 73.8567, 80.2707, 77.026
}

# Convert to DataFrame
city_coords_df = pd.DataFrame(city_coords)
```

```
# Merge with coordinates
city_funding = city_funding.merge(city_coords_df, on='City', how='left')

# Filter out NaN values
city_funding = city_funding.dropna(subset=['Latitude', 'Longitude'])

# Select all cities in the first graph
selected_cities_funding = city_funding[city_funding['City'].isin(city_coords['City'])]

# Create a Folium map with the "dark" map style
m_funding = folium.Map(location=[20.5937, 78.9629], zoom_start=5, tiles='CartoDB dark_m

# Add markers to the map for each city with the total funding amount
marker_cluster_funding = MarkerCluster().add_to(m_funding)

for idx, row in selected_cities_funding.iterrows():
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=f"{row['City']}: ${row['Total Funding Amount']:,.2f}",
        icon=folium.Icon(color='green')
    ).add_to(marker_cluster_funding)

# Display the map
m_funding
```



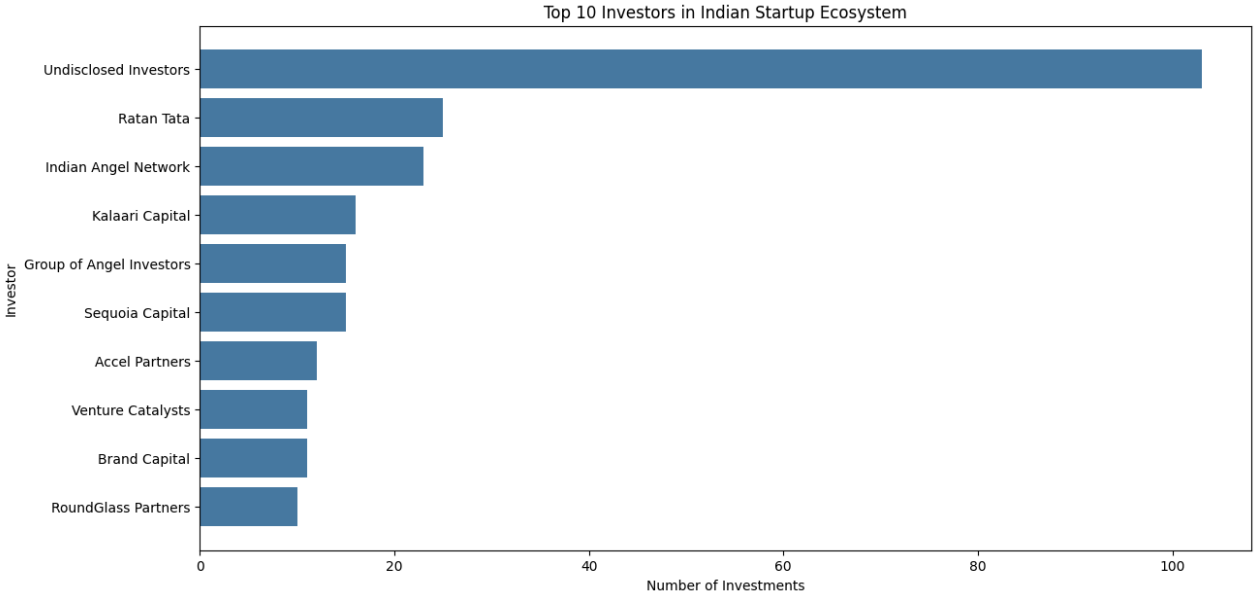
Output:Analysis of the Top 10 Cities by Number of Startups

Conclusion

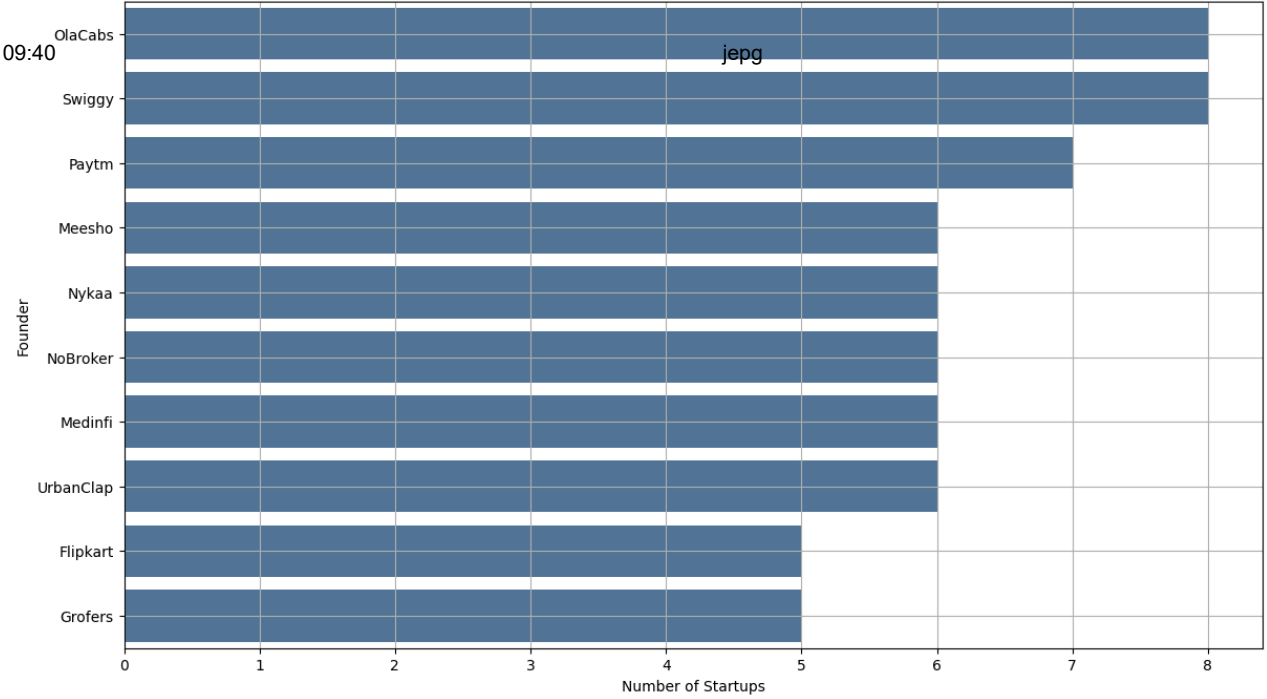
- The chart indicates that location plays a crucial role in determining the growth and success of startups in terms of funding. Major tech hubs and cities with strong ecosystems attract the most investment. Understanding these dynamics can help stakeholders, including investors,

Who plays the main role in Indian Startups Ecosystem?

In [64]:
```
# Investor analysis
investor_funding = df['Investors Name'].value_counts().reset_index()
investor_funding.columns = ['Investor', 'Number of Investments']

# Top 10 most invested investors
top_investors = investor_funding.head(10)

plt.figure(figsize=(14, 7))
plt.barh(top_investors['Investor'],
         top_investors['Number of Investments'],
          color='#4878A2')
plt.xlabel('Number of Investments')
plt.ylabel('Investor')
plt.title('Top 10 Investors in Indian Startup Ecosystem')
plt.gca().invert_yaxis()
plt.show()
```



In [65]:
```
# Extract the most active startup founders (this is a simplification)
# Split multiple founders in a single row (assuming founders are listed in 'Startup Nam
df['Founders'] = df['Startup Name'].fillna('Unknown').str.split(',')

# Explode the list of founders into separate rows
founders_exploded = df.explode('Founders')
top_founders = founders_exploded['Founders'].value_counts().head(10).reset_index()
top_founders.columns = ['Founder', 'Number of Startups']

# Plot the top founders
plt.figure(figsize=(14, 8))
sns.barplot(x='Number of Startups', y='Founder', data=top_founders, color='#4878A2')
plt.title('Top 10 Founders by Number of Startups')
plt.xlabel('Number of Startups')
```



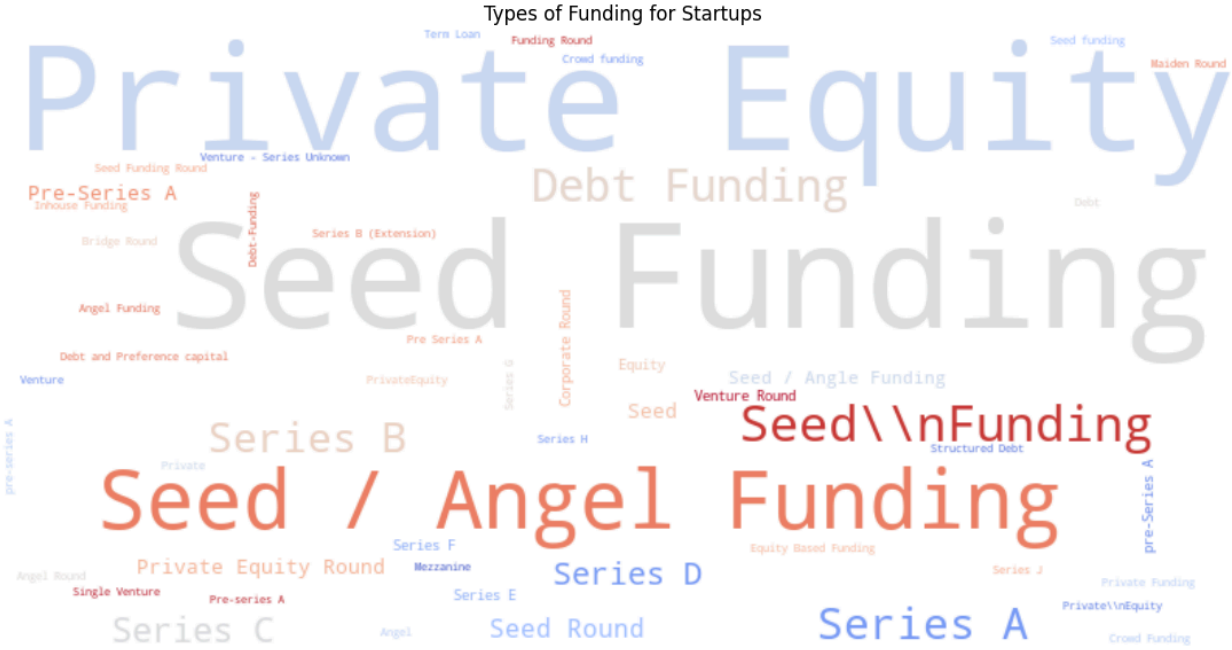
What are the different Types of Funding for Startups?

In [39]:
```
df.head()
```

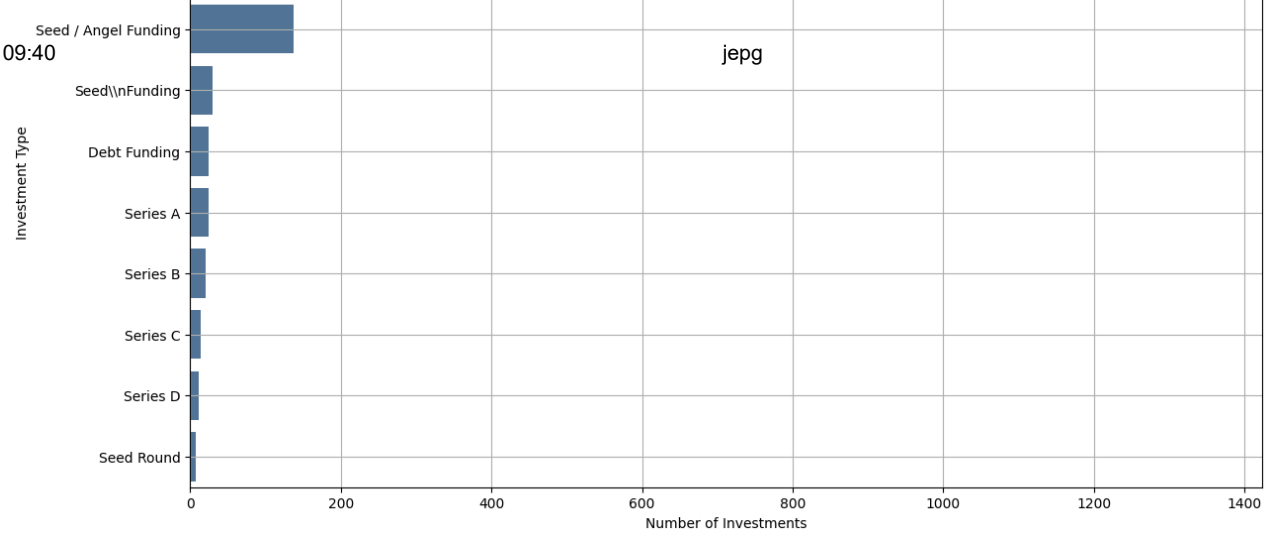
Out[39]:

	Sr No	Date dd/mm/yyyy	Startup Name	Industry Vertical	SubVertical	City Location	Inve N
0	1	2020-01-09	BYJU'S	e-tech	E-learning	Bangalore	Tiger G Manage
1	2	2020-01-13	Shuttl	transportation	App based shuttle service	Gurgaon	Susqueh Gr E
2	3	2020-01-09	Mamaearth	e-commerce	Retailer of baby and toddler products	Bangalore	Sec Capital
3	4	2020-01-02	https://www.wealthbucket.in/	fintech	Online Investment	New Delhi	V Khat
4	5	2020-01-02	Fashor	fashion and apparel	Embroiled Clothes For Women	Mumbai	Sı Ve Par

```
investment_types = df['InvestmentnType'].value_counts().reset_index()
investment_types.columns = ['Investment Type', 'Number of Investments']

# Convert the investment types to a dictionary
investment_dict = dict(zip(investment_types['Investment Type'], investment_types['Numbe

# Generate a word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white', colormap='coolwa

# Plot the word cloud
plt.figure(figsize=(14, 8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Types of Funding for Startups')
plt.show()
```



In [66]:
```
# Assuming df is the cleaned dataframe with startup data
# Count the number of each investment type
investment_types = df['InvestmentnType'].value_counts().head(10).reset_index()
investment_types.columns = ['Investment Type', 'Number of Investments']

# Plot the investment types
plt.figure(figsize=(14, 8))
sns.barplot(x='Number of Investments', y='Investment Type', data=investment_types, colo
plt.title('Types of Funding for Startups')
plt.xlabel('Number of Investments')
plt.ylabel('Investment Type')
plt.grid(True)
plt.show()
```



Output:Analysis of the Types of Funding for Startups

Private Equity: Most Common Funding Type: Private Equity is the most common funding type, with nearly 1,400 instances. This indicates that many startups in the dataset have reached a level of maturity where they can attract significant private equity investments.

Seed Funding: Second Most Common: Seed Funding is close behind Private Equity, with a similar number of instances. This suggests that many startups are in the early stages of their lifecycle, seeking initial capital to develop their ideas and products.

Seed / Angel Funding: Early-Stage Investments: Seed / Angel Funding is also prominent, with a significant number of instances. This type of funding is crucial for startups to get off the ground and demonstrates the active role of angel investors in the ecosystem.

Diverse Funding Landscape: The chart demonstrates a diverse landscape of funding types, from early-stage seed funding to later-stage private equity. This diversity is crucial for catering to the varying needs of startups at different stages of their growth.

Importance of Early-Stage Funding: The high frequency of Seed Funding and Seed / Angel Funding underscores the importance of early-stage investments in nurturing new startups. These funding types are critical for startups to develop their initial ideas and products.

Private Equity's Dominance: The dominance of Private Equity highlights the significant role of large-scale investments in the startup ecosystem. It suggests that many startups in the dataset have achieved substantial growth and maturity, making them attractive targets for private equity investors.

Growth Funding Rounds: The presence of Series A, B, C, and D funding rounds, although less frequent, indicates a structured path for startups to secure additional capital as they grow. Each subsequent round typically involves larger amounts of funding and is aimed at scaling the business.

Recommendations for New Investors

Focus on Early-Stage Investments:

- Seed Funding and Seed/Angel Funding:

Observation: Seed Funding and Seed/Angel Funding are among the most common types of investments, indicating a high level of activity at the early stages of startup development.

Recommendation: New investors should consider participating in early-stage investments. This allows them to support innovative ideas and potentially benefit from high returns if the startups succeed.

Diversify Across Funding Stages:

- Private Equity and Series Funding:

Observation: While Private Equity dominates, there are also significant numbers of Series A, B, C, and D funding rounds.

Recommendation: Diversifying investments across different stages of a startup's lifecycle can mitigate risk. Early-stage investments offer high growth potential, whereas later-stage investments in Series rounds or Private Equity can provide more stability and lower risk.

Leverage the Power of Syndicates and Co-Investments:

Observation: High-frequency investors often participate in investment syndicates or co-investments. Recommendation: Collaborate with experienced investors and join syndicates to pool resources and share risks. This approach also provides access to a broader network and better deal flow. Identify and Support High-Potential Sectors:

Prominent Sectors: Observation: Key sectors like Consumer Internet, Technology, and E-Commerce show high levels of startup activity and funding. Recommendation: Focus on high-potential sectors that align with your expertise and interests. Investing in sectors with robust growth prospects can enhance the likelihood of successful exits.

Consider Regional Investment Strategies:

- Top Cities for Startups:

Observation: Cities like Bangalore, Mumbai, and New Delhi dominate the startup ecosystem, attracting the most startups and funding.

Recommendation: Consider focusing investments in these key startup hubs. These cities offer vibrant ecosystems, better infrastructure, and access to a large pool of talent and resources.

based on market opportunity, team quality, product-market fit, and growth metrics. Use a data-driven approach to make informed investment decisions.

Support Startups with Strategic Value: Observation: Investors not only provide capital but also strategic support, mentoring, and networking opportunities.

Recommendation: Offer more than just financial investment. Provide strategic guidance, mentorship, and leverage your network to help startups grow. This adds value to your investment and increases the chances of startup success.

Stay Updated with Market Trends and Innovations: Observation: The startup ecosystem is dynamic, with continuous innovations and evolving market trends.

Recommendation: Stay informed about the latest market trends, technological advancements, and emerging sectors. Continuous learning and adaptability are crucial for identifying new investment opportunities and staying ahead in the competitive landscape.

Conclusion By following these recommendations, new investors can strategically navigate the startup ecosystem, identify promising opportunities, and build a diversified and high-potential investment portfolio. Balancing early-stage and later-stage investments, focusing on key sectors and regions, and providing strategic support to startups will enhance the likelihood of successful investments and sustainable growth.

THANK YOU

If you want to be the first to be informed about new projects, please do not forget to follow us - by Fatma Nur AZMAN

Fatmanurazman.com | Linkedin | Github | Kaggle | Tableau