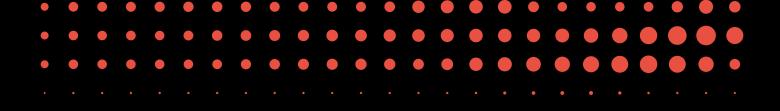




- 1. Introduction: This project analyse startup funding in India over the last 10 years.
- 2. Objective: Understand funding trends, top sectors, investors, city-wise distribution, and startup success rates.
- 3. Dataset: 560 entries, 11 columns including Startup Name, Sector, City, Funding Round, Funding Amount, Lead Investor, Date, Outcome, etc.







Import Libraries

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

• All libraries needed for data cleaning, analysis, and plotting.





Load Dataset

```
df = pd.read_excel(r"C:\Users\Om\Downloads\startup_funding_india_2014_2024_messy.csv.xlsx",
   engine='openpyxl')
   df.head()
   df.info()
```

- Initial check to understand dataset columns and types.
- Look at first few rows to get an idea of the data.



Check Missing Values and Shape

```
V V
```

```
df.isnull().sum()
 ✓ 0.0s
Startup Name
Sector
                         35
Location City
                         28
Funding Round
                          0
Funding Amount Raw
Lead Investor
                         24
Valuation Raw
                        155
Funding Date Raw
Outcome
funding amt clean me
Lead Investor 2
                        537
dtype: int64
```

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

- Identify columns with missing values.
- Check total number of rows and columns.

Observation:

- Valuation_Raw and Lead_Investor_2 have many missing values.
- Some columns like Funding_Amount are stored as strings.



Data Cleaning

```
# Clean funding amount
df['funding_amt_clean_me'] = df['funding_amt_clean_me'].astype(str).str.replace(r'[\$,]', '',
    regex=True)
df['funding_amt_clean_me'] = pd.to_numeric(df['funding_amt_clean_me'], errors='coerce')

# Clean valuation
df['Valuation_Raw'] = df['Valuation_Raw'].astype(str).str.replace(r'[\$,]', '', regex=True)
df['Valuation_Raw'] = pd.to_numeric(df['Valuation_Raw'], errors='coerce')
df.isnull().sum()
```

- Drop irrelevant columns: Lead_Investor_2 (too many nulls)
- Convert funding amounts to numeric: ercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

V V

Convert funding date to datetime and extract Year

```
df['Funding_Date_Raw'] = pd.to_datetime(df['Funding_Date_Raw'], errors='coerce')
df.isnull().sum()
df.info()
```

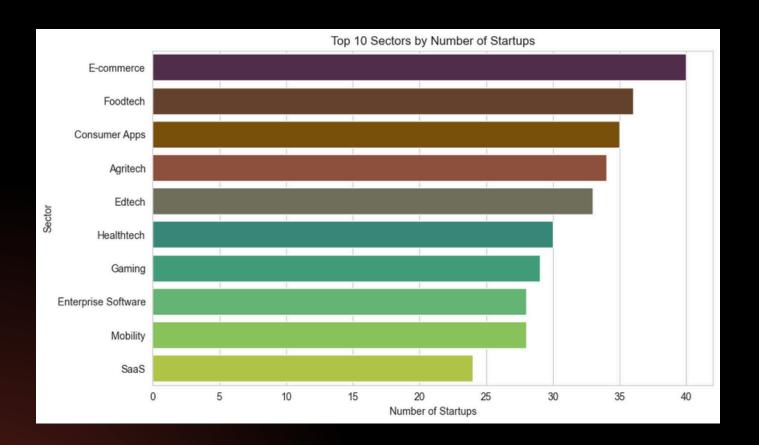


Top Sectors by Number of Startups

```
# Count number of startups per sector
top_sectors = df['Sector'].value_counts().head(10)

# Plot
plt.figure(figsize=(10,6))
sns.barplot(x=top_sectors.values, y=top_sectors.index, palette='viridis')
plt.title("Top 10 Sectors by Number of Startups")
plt.xlabel("Number of Startups")
plt.ylabel("Sector")
plt.show()
```

This code identifies the top 10 sectors with the highest number of startups by counting how many startups belong to each sector. It then visualizes the result using a bar chart, making it easy to see which sectors are the most active in the startup ecosystem. This helps investors and stakeholders quickly identify which industries are attracting the most entrepreneurial activity.



- E-commerce, FoodTech, and Consumer Apps are the most crowded startup sectors, indicating very high market interest but intense competition.
- Agritech, Edtech & Healthtech also show strong participation, suggesting increasing innovation in impact-driven sectors.
- SaaS and Enterprise Software are slightly lower in count, but these sectors generally focus on B2B and high-revenue models rather than volume.



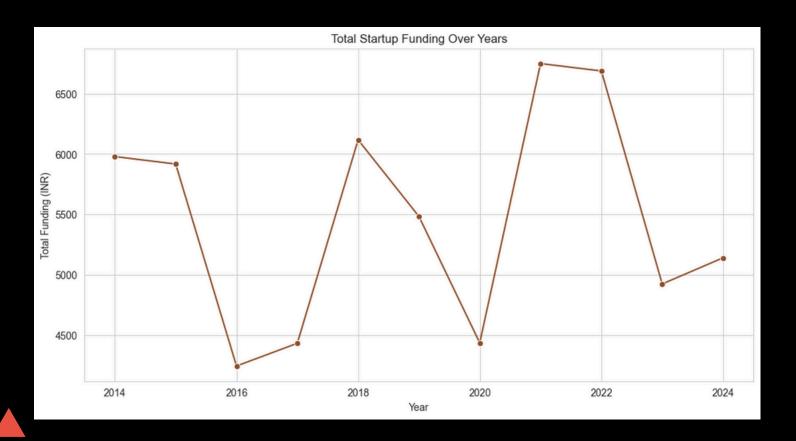
Funding Trend Over Years

```
V V V
```

```
# Total funding per year
funding_by_year = df.groupby('Year')['Funding_Amount_Raw'].sum()

# Plot
plt.figure(figsize=(12,6))
sns.lineplot(x=funding_by_year.index, y=funding_by_year.values, marker='o')
plt.title("Total Startup Funding Over Years")
plt.xlabel("Year")
plt.ylabel("Total Funding (INR)")
plt.show()
```

This code calculates the total amount of funding raised by startups in each year. It then displays the trend using a line chart to show how startup funding has grown or fluctuated over time. This helps understand which years saw the highest investor activity and the overall growth pattern of the startup ecosystem.



- Peak and Trough: Funding reached its peak in 2021 (and remained high in 2022) with over 6500 Cr, while the lowest point (the trough) was in 2016 and 2020, with funding around INR 4200 4500 Cr.
- Major Swings: There were sharp drops, most notably from 2015 to 2016, and a dramatic decrease from the 2021-2022 high to the 2023 level around INR 4900 Cr.
- Recent Recovery: There appears to be a slight recovery in 2024 compared to 2023, though it remains significantly below the 2021-2022 peak.

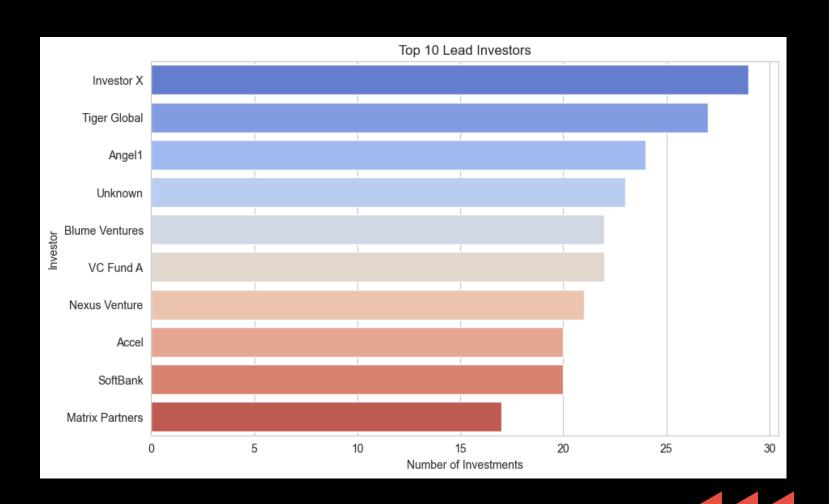
>>> Top Lead Investors

This code identifies the top 10 investors who have funded the highest number of startups. It counts how many investment deals each investor has made and displays it using a bar chart. This helps us understand which investors are most active and playing a major role in driving startup funding

```
# Count number of investments per lead investor
top_investors = df['Lead_Investor'].value_counts().head(10)

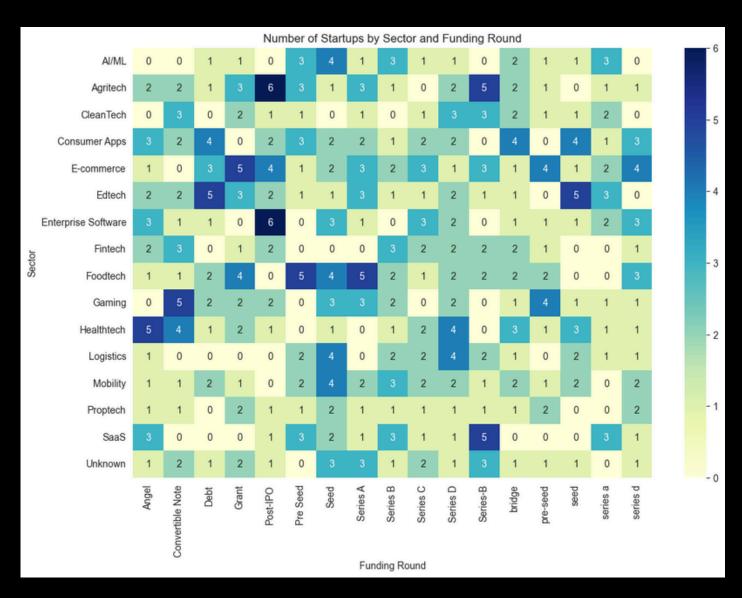
# Plot
plt.figure(figsize=(10,6))
sns.barplot(x=top_investors.values, y=top_investors.index, palette='coolwarm')
plt.title("Top 10 Lead Investors")
plt.xlabel("Number of Investments")
plt.ylabel("Investor")
plt.show()
```

- Top Investors: Investor X and Tiger Global are the clear leaders, both having made close to 30 lead investments.
- Mid-Tier Activity: Angel1 and an Unknown investor are also highly active, with 23 and 22 lead investments, respectively.
- VC Fund Presence: Well-known VC funds like Blume Ventures, VC Fund A, Nexus Venture, Accel, and SoftBank feature prominently among the top 10, typically with investment counts ranging from 17 to 22.
- Matrix Partners rounds out the top 10 with the lowest number of investments on this list, around 17.



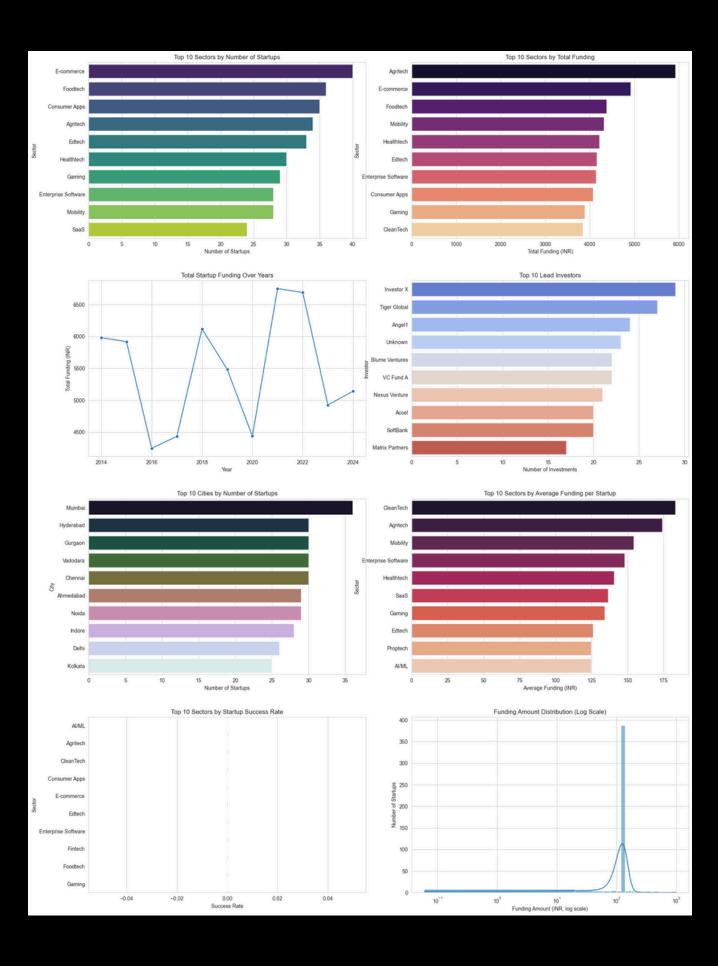
```
# Create pivot table: sectors vs funding rounds, count of startups
sector_round_pivot = df.pivot_table(index='Sector', columns='Funding_Round',
values='Startup_Name', aggfunc='count', fill_value=0)

# Plot heatmap
plt.figure(figsize=(12,8))
sns.heatmap(sector_round_pivot, annot=True, fmt='d', cmap='YlGnBu')
plt.title("Number of Startups by Sector and Funding Round")
plt.xlabel("Funding Round")
plt.ylabel("Sector")
plt.show()
```



This code creates a pivot table to show how many startups are present in each sector at each funding round (like Seed, Series A, Series B, etc.). It then visualizes this as a heatmap, where darker shades indicate higher concentration of startups. This helps identify which sectors are raising more early-stage or latestage funding, giving a strong business and investor trend insight.

- Most Active Sectors (Overall): Edtech, Healthtech, Gaming, and Agritech appear to have a high overall number of startups that have successfully raised funding across various rounds.
- Early-Stage Funding Dominance: For almost all sectors, funding is concentrated in the Angel, Seed, and Pre-Seed rounds, indicating a large pipeline of early-stage startups.
- Healthtech shows the highest number of Angel investments (5).
- Gaming and Foodtech have the most significant activity in Seed rounds (5 and 4, respectively).
- Late-Stage Success: Fewer startups make it to the later stages (Series C, D, etc.).
- Agritech shows a high count of 5 in Series B funding, which is notable.
- E-commerce and AI/ML also show some activity in the later-stage Series D and Series E rounds.
- Specific Round Strength: Enterprise Software and AI/ML have a high count (6) in the Post-IPO and Pre-Seed rounds, respectively, suggesting success in reaching maturity or strong initial interest.



Dashboard of All Plots

Combine all key charts in one dashboard page to visualize trends, top sectors, investors, funding distribution, and success rates.

Advanced Business Insights

- Smaller sectors sometimes attract higher average funding per startup.
- Success does not always correlate with total funding.
- Certain cities dominate specific sectors, showing regional specialization.
- Top investors focus on select sectors, showing concentrated investment patterns.

Conclusion

- Indian startup funding is growing steadily and concentrated in top sectors and cities.
- Investors are strategically focused, but smaller sectors also attract high-value funding.
- Funding amount alone does not determine success; sector-specific trends are key.
- Dataset and insights can guide investors, entrepreneurs, and business strategy.

