Project-2

Startup Investment Analysis

1. Introduction

Shark Tank India is a popular business reality show where aspiring entrepreneurs pitch their innovative business ideas to a panel of investors ("Sharks") seeking funding and mentorship. This dataset provides valuable insights into the startups, founders, investments, sectors, and deals made during the show. Analyzing this dataset helps us understand investment patterns, popular industries, gender representation among founders, and which Sharks invest the most.

2. Abstract

The main objective of this project is to perform exploratory data analysis (EDA) on the Shark Tank India dataset to uncover trends and patterns in startup funding. By analyzing data related to pitches, investments, valuations, and sectors, we aim to identify key insights such as:

- Which industry sectors attract the most investment
- Average equity and deal amounts
- Which Shark invests most frequently
- Gender diversity among founders
- Relationship between startup valuation and deal amount

The analysis will support data-driven insights about entrepreneurship trends and investor behavior in India's startup ecosystem.

3. Tools Used

- Python – For data analysis and visualization
- Pandas – To clean and manipulate the dataset
- NumPy – For numerical computations
- Matplotlib / Seaborn – For creating visual insights and trend graphs
- Jupyter Notebook / Google Colab – For executing and documenting the analysis
- Excel – For preliminary data exploration and validation

4. Steps Involved in Building the Project

Step 1: Data Collection

- Imported the Shark Tank India.csv dataset which includes details like startup name, industry, founders, amount invested, valuation, and Sharks involved.

Step 2: Data Cleaning & Preprocessing

- Handled missing values, formatted numeric columns, and standardized data types.
- Removed irrelevant columns and corrected inconsistent entries (like missing values for investments or equity).

Step 3: Exploratory Data Analysis (EDA)

- Analyzed distributions of investments across industries.
- Examined which Shark invested the most and how frequently.
- Studied average equity given and total deal amounts.
- Visualized sector-wise deal count, average investment, and valuation using bar and pie charts.
- Compared solo vs group investments and gender representation among founders.

Step 4: Data Visualization

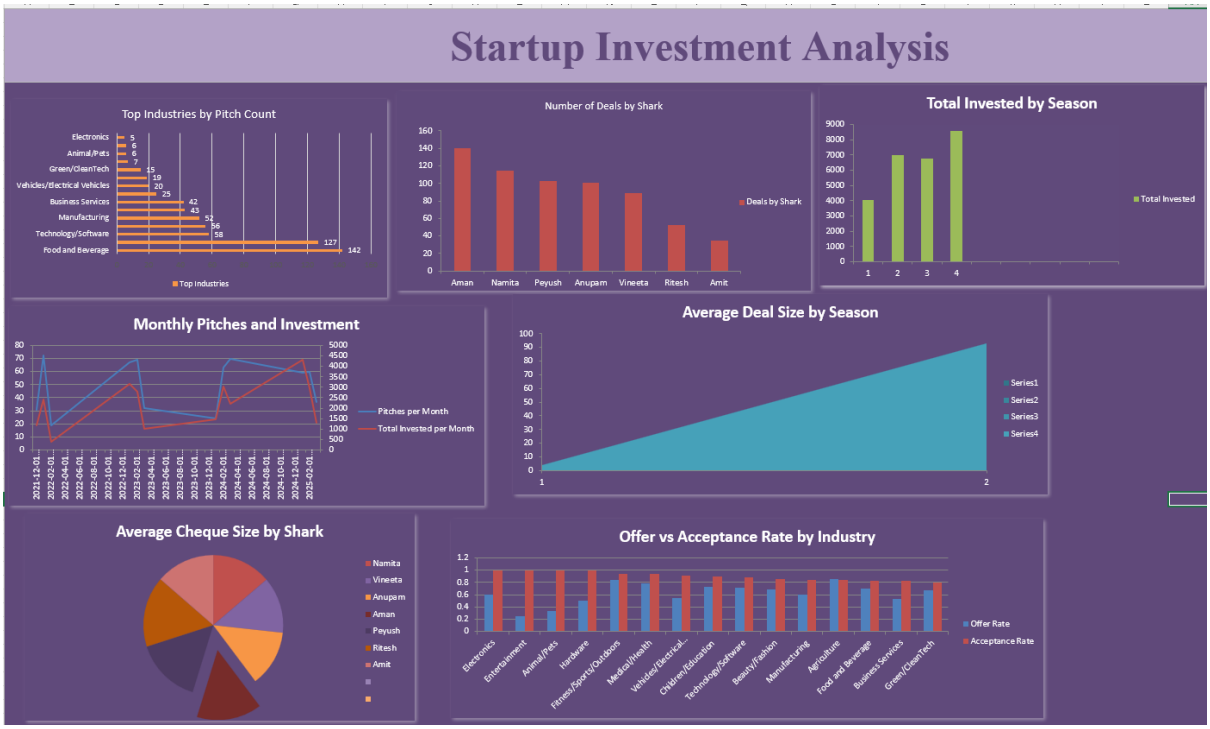
- Created meaningful visualizations to present findings:
    - Top 5 Industries with Most Deals
    - Sharks with Highest Total Investment
    - Investment Amount vs Equity Percentage
    - Male vs Female Founders

Step 5: Insights & Findings

- Most investments occurred in sectors like Food & Beverage, Technology, and Lifestyle.
- Aman Gupta and Anupam Mittal were among the most active investors.
- Startups with multiple founders tended to get higher investments.
- Female-led startups received fewer deals but showed strong business valuations.

5. Conclusion

This project provides a comprehensive overview of the investment landscape in Shark Tank India. The analysis highlights which industries attract the most investor interest, how equity and valuations are related, and the participation levels of each Shark.
Such insights are useful not only for future entrepreneurs but also for understanding emerging business trends in India's growing startup ecosystem. The dataset demonstrates how data analytics can uncover real-world business intelligence from television-based entrepreneurship content.



In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [8]:
```
# Read CSV
df_shark = pd.read_csv('Shark Tank India.csv', encoding='utf-8')
df_shark.head()
```

Out[8]:

	Season Number	Startup Name	Episode Number	Pitch Number	Season Start	Season End	Original Air Date	Episode Title	
0	1	BluePineFoods	1	1	20-Dec-21	4-Feb-22	20-Dec-21	Badlegi Business Ki Tasveer	
1	1	BoozScooters	1	2	20-Dec-21	4-Feb-22	20-Dec-21	Badlegi Business Ki Tasveer	
2	1	HeartUpMySleeves	1	3	20-Dec-21	4-Feb-22	20-Dec-21	Badlegi Business Ki Tasveer	
3	1	TagzFoods	2	4	20-Dec-21	4-Feb-22	21-Dec-21	Insaan, Ideas Aur Sapne	
4	1	HeadAndHeart	2	5	20-Dec-21	4-Feb-22	21-Dec-21	Insaan, Ideas Aur Sapne	

5 rows × 80 columns

In [9]:
```
# Basic shape and head
print(df_shark.shape)
print(df_shark.head())
```

```
(634, 80)
   Season Number      Startup Name  Episode Number  Pitch Number Season Start  \
0              1     BluePineFoods               1             1    20-Dec-21
1              1      BoozScooters               1             2    20-Dec-21
2              1  HeartUpMySleeves               1             3    20-Dec-21
3              1         TagzFoods               2             4    20-Dec-21
4              1      HeadAndHeart               2             5    20-Dec-21

  Season End Original Air Date                Episode Title           Anchor  \
0   4-Feb-22         20-Dec-21  Badlegi Business Ki Tasveer  Rannvijay Singh
1   4-Feb-22         20-Dec-21  Badlegi Business Ki Tasveer  Rannvijay Singh
2   4-Feb-22         20-Dec-21  Badlegi Business Ki Tasveer  Rannvijay Singh
3   4-Feb-22         21-Dec-21      Insaan, Ideas Aur Sapne  Rannvijay Singh
4   4-Feb-22         21-Dec-21      Insaan, Ideas Aur Sapne  Rannvijay Singh

                      Industry  ... Invested Guest Name All Guest Names  \
0            Food and Beverage  ...      Ashneer Grover  Ashneer Grover
1  Vehicles/Electrical Vehicles  ...      Ashneer Grover  Ashneer Grover
2               Beauty/Fashion  ...                 NaN  Ashneer Grover
3            Food and Beverage  ...      Ashneer Grover  Ashneer Grover
4           Children/Education  ...                 NaN  Ashneer Grover

   Namita Present  Vineeta Present  Anupam Present  Aman Present  \
0             1.0              1.0             1.0           1.0
1             1.0              1.0             1.0           1.0
2             1.0              1.0             1.0           1.0
3             1.0              1.0             1.0           1.0
4             1.0              1.0             1.0           1.0

   Peyush Present  Ritesh Present Amit Present Guest Present
0             NaN             NaN          NaN           1.0
1             NaN             NaN          NaN           1.0
2             NaN             NaN          NaN           1.0
3             NaN             NaN          NaN           1.0
4             NaN             NaN          NaN           1.0

[5 rows x 80 columns]
```

In [10]:

```python
# Normalize some key columns to numeric where relevant
num_cols = [
    'Original Ask Amount','Original Offered Equity','Valuation Requested','Total
    'Total Deal Debt','Debt Interest','Deal Valuation','Number of Sharks in Deal
    'Yearly Revenue','Monthly Sales','Gross Margin','Net Margin','EBITDA','Cash
    'Namita Investment Amount','Vineeta Investment Amount','Anupam Investment Am
]
for c in num_cols:
    if c in df_shark.columns:
        df_shark[c] = pd.to_numeric(df_shark[c], errors='coerce')

# High-level metrics
n_rows = len(df_shark)
seasons = sorted(df_shark['Season Number'].dropna().unique().tolist()) if 'Seaso
unique_episodes = df_shark['Episode Number'].nunique() if 'Episode Number' in df
pitches = df_shark['Pitch Number'].nunique() if 'Pitch Number' in df_shark.colum
received_offer_rate = None
if 'Received Offer' in df_shark.columns:
    received_offer_rate = df_shark['Received Offer'].dropna().mean()
accepted_deal_rate = None
if 'Accepted Offer' in df_shark.columns:
    accepted_deal_rate = df_shark['Accepted Offer'].dropna().mean()
```

```
total_invested = df_shark['Total Deal Amount'].sum(skipna=True) if 'Total Deal A
median_valuation = df_shark['Deal Valuation'].median(skipna=True) if 'Deal Valua

print(n_rows)
print(seasons)
print(unique_episodes)
print(pitches)
print(received_offer_rate)
print(accepted_deal_rate)
print(total_invested)
print(median_valuation)
```

```
634
[1, 2, 3, 4]
53
634
0.667192429022082
0.851063829787234
26284.91356
1250.0
```

In [11]:
```
# Top industries by count
top_industries = pd.Series(dtype='int')
if 'Industry' in df_shark.columns:
    top_industries = df_shark['Industry'].value_counts().head(10)
print(top_industries)
```

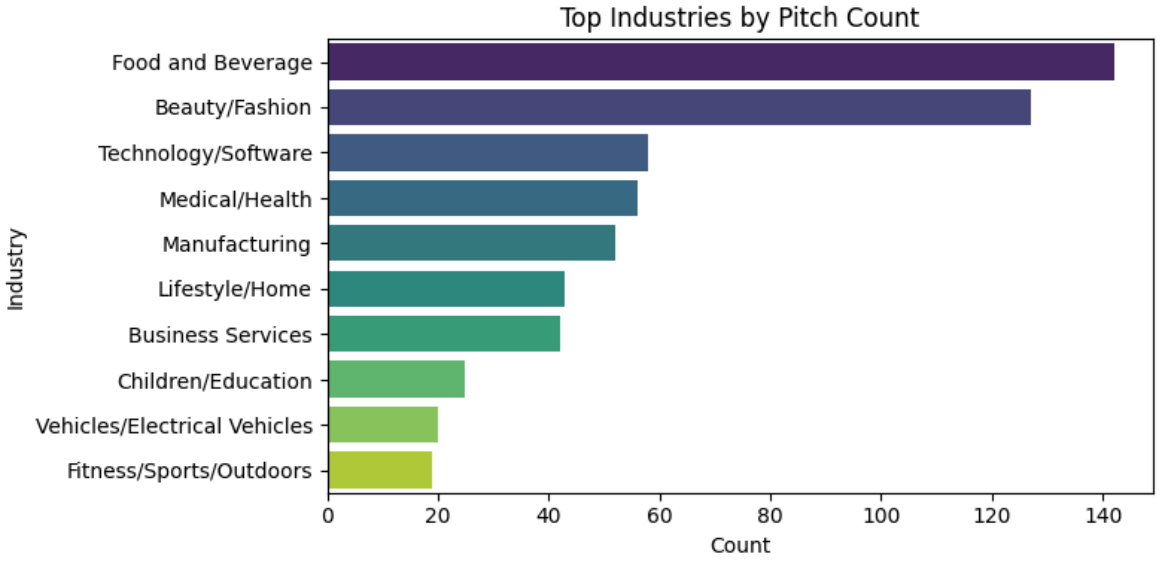
```
Industry
Food and Beverage              142
Beauty/Fashion                 127
Technology/Software             58
Medical/Health                  56
Manufacturing                   52
Lifestyle/Home                  43
Business Services               42
Children/Education              25
Vehicles/Electrical Vehicles    20
Fitness/Sports/Outdoors         19
Name: count, dtype: int64
```

In [12]:
```
# Shark participation counts: count of investments per shark
shark_amount_cols = [
    ('Namita Investment Amount','Namita'),
    ('Vineeta Investment Amount','Vineeta'),
    ('Anupam Investment Amount','Anupam'),
    ('Aman Investment Amount','Aman'),
    ('Peyush Investment Amount','Peyush'),
    ('Ritesh Investment Amount','Ritesh'),
    ('Amit Investment Amount','Amit'),
    ('Guest Investment Amount','Guest')
]
participation = {}
for col, name in shark_amount_cols:
    if col in df_shark.columns:
        participation[name] = (df_shark[col].fillna(0) > 0).sum()

participation_series = pd.Series(participation).sort_values(ascending=False)
print(participation_series)
```

```
Aman       140
Namita     114
Peyush     103
Anupam     101
Vineeta     89
Guest       63
Ritesh      52
Amit        35
dtype: int64
```

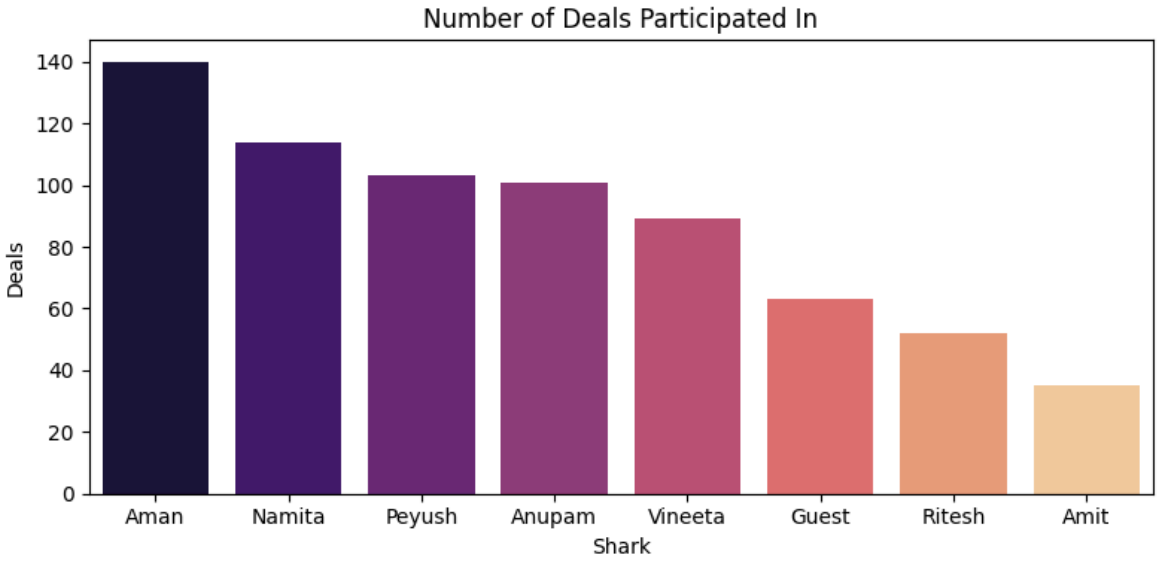
In [13]:
```
# Plot 1: Top industries
if not top_industries.empty:
    plt.figure(figsize=(8,4))
    sns.barplot(x=top_industries.values, y=top_industries.index, orient='h', pal
    plt.title('Top Industries by Pitch Count')
    plt.xlabel('Count')
    plt.ylabel('Industry')
    plt.tight_layout()
    plt.show()
```

```
C:\Users\kkjeg\AppData\Local\Temp\ipykernel_6764\1492733257.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.barplot(x=top_industries.values, y=top_industries.index, orient='h', palett
e='viridis')
```

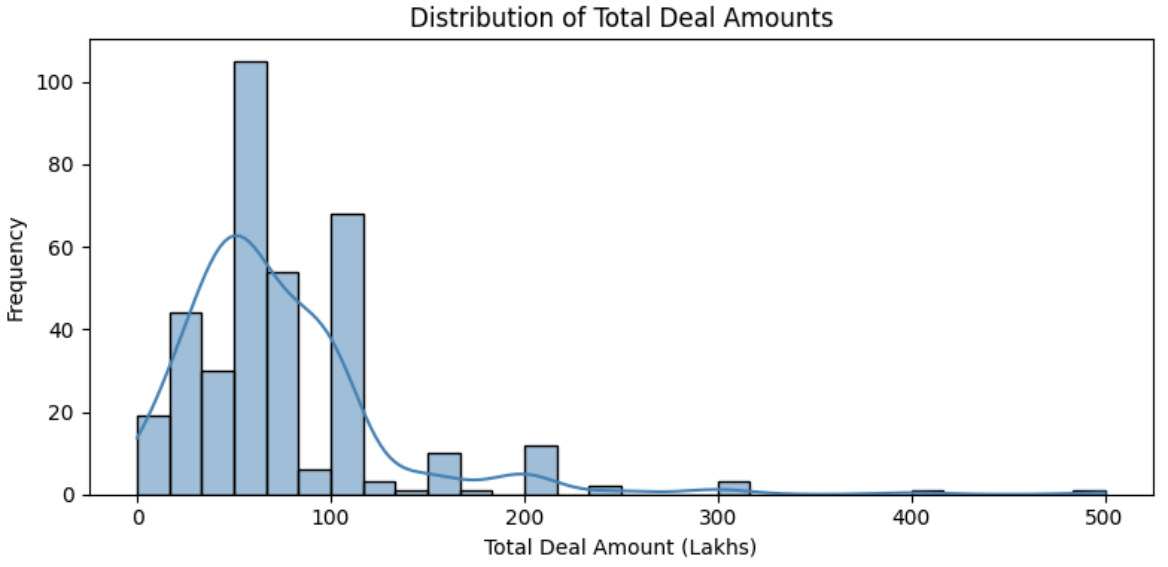


In [14]:
```
# Plot 2: Investment participation by shark
if len(participation_series) > 0:
    plt.figure(figsize=(8,4))
    sns.barplot(x=participation_series.index, y=participation_series.values, pal
    plt.title('Number of Deals Participated In')
    plt.xlabel('Shark')
    plt.ylabel('Deals')
    plt.tight_layout()
    plt.show()
```

```
C:\Users\kkjeg\AppData\Local\Temp\ipykernel_6764\1645072389.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.barplot(x=participation_series.index, y=participation_series.values, palett
e='magma')
```



In [15]:
```
# Plot 3: Deal amounts distribution
if 'Total Deal Amount' in df_shark.columns and df_shark['Total Deal Amount'].not
    plt.figure(figsize=(8,4))
    sns.histplot(df_shark['Total Deal Amount'].dropna(), bins=30, kde=True, colo
    plt.title('Distribution of Total Deal Amounts')
    plt.xlabel('Total Deal Amount (Lakhs)')
    plt.ylabel('Frequency')
    plt.tight_layout()
    plt.show()

print('Loaded data, computed summaries, and plotted key charts')
```




```
Loaded data, computed summaries, and plotted key charts
```