Analyzing Startup Ecosystem and Funding Trends in India

Objective

Explore the Indian startup ecosystem using real-world funding data. Identify industry trends, city-wise growth patterns, top investors, and build predictive models to estimate funding amounts based on startup characteristics.

Dataset

- **Source:** [Startup Funding in India – Kaggle (Free)]
- **Link:** https://www.kaggle.com/datasets/sudalairajkumar/indian-startup-funding

Columns include:

- Startup Name, Industry Vertical, SubVertical, City Location, Investors Name
- Investment Type (e.g., Seed, Series A), Amount in USD, Date, Remarks

Use Cases

- Analyze city- and industry-wise funding patterns
- Identify most active investors and funding rounds
- Detect seasonal or annual funding booms/busts
- Predict funding amount based on sector, city, investor type
- Help aspiring founders understand investor trends

SECTION A: Python & Data Cleaning

- Load the dataset, display structure, and inspect datatypes
- Convert Date to datetime and sort by year
- Clean Amount in USD column (remove strings, convert to float)
- Standardize inconsistent entries in City Location, Industry Vertical, and Investment Type
- Add derived columns:
    - Year from date
    - Funding Category based on amount: Low, Medium, High
    - Number of investors (split & count from Investors Name)

SECTION B: SQL Operations

- Import cleaned data into SQL
- Write queries to:
    - List top 10 most funded startups by total amount
    - Identify cities with highest total and average funding
    - Count startups funded each year by investment type
    - Find most common combinations of investor and city
    - Identify industries with highest median funding

SECTION C: EDA & Statistics

Exploratory Data Analysis:

- Plot bar charts of startup count per city and industry
- Line chart of total funding amount per year (trendline)
- Box plots comparing funding amounts across cities
- Pie chart of investment types distribution
- Word cloud of most frequent investor names
- Count plot of startups per funding round (Seed, Series A, etc.)
- Correlation analysis between funding amount and number of investors

Descriptive Statistics:

- Mean, median, standard deviation of funding amount
- Create summary tables: Avg. funding per industry per year
- Calculate coefficient of variation by city and industry
- Identify cities/industries with most stable vs. volatile funding
- Calculate IQR and detect funding outliers
- Determine top 3 sectors that received highest median investment post-2018

SECTION D: Tableau Dashboard

- Create an interactive dashboard with:
    - Filters for industry, city, year, funding type
    - Time-series line chart of yearly total funding
    - Horizontal bar chart of average funding per industry
    - Top 10 investors and their preferred sectors
    - Scatter plot: funding vs number of investors
    - Map view: city-wise funding distribution
    - Dynamic investor tracker: click an investor → see startups & total funding

SECTION E: Machine Learning

Regression Task:

- Predict Amount in USD using:
  - Industry Vertical, City, Investment Type, No. of Investors, Year
- Models:
  - Linear Regression
  - Random Forest Regressor
  - XGBoost (optional)

Classification Task:

- Classify startups as Low, Medium, High funded
- Label encode categorical variables
- Use Decision Tree, Logistic Regression, and Random Forest
- Evaluate with accuracy and feature importance