Project Report: Startup Investment Analysis (Shark Tank India Data)

Abstract

This report documents the analysis of the Shark Tank India dataset, spanning multiple seasons and over 600 startup pitches. The objective was to uncover key investment trends, evaluate the success drivers among founding teams, and identify the industry preferences of individual investors ("Sharks"). The analysis utilized Python for data processing and aggregation, generating structured data insights on deal success rates across various industries and founder demographics (age, gender, location). Key findings revealed that All-Female teams demonstrated the highest deal success rate, while the **Beauty/Fashion** and **Food and Beverage** sectors received the largest cumulative funding amounts. The data prepared is intended for use in an interactive **Power BI** dashboard to visualize these trends.

1. Introduction

Step	Detail	Output
1. Data Loading & Cleaning	The Shark_tank_india.csv dataset was loaded. Columns related to investment (Total Deal Amount, Accepted Offer) were cleaned and converted to numeric types.	Cleaned Pandas DataFrame.
2. Industry Trend Analysis	Data was grouped by **Industry** to calculate the **total pitches**, **successful deals**, **deal success rate**, and **total funding**.	industry_analysis.csv
3. Founder Profile Analysis	Startups were categorized by **gender composition** (All-Female, All-Male, Mixed Team) and **age group** (Young, Middle, Old) to calculate success rates for each demographic.	founder_gender_analysis.csv and founder_age_analysis.csv
4. Investor Trend Analysis	The dataset was **unpivoted** using Python (simulating complex SQL UNION operations) to aggregate the total investment amount and number of deals made by each individual Shark within every industry.	shark_industry_investment_analysis.csv
5. Visualization Preparation	All generated data tables were exported as structured CSV files, ready for direct connection to **Power BI** for dashboard creation.	**5 Analytical CSV files**

The global startup ecosystem is heavily influenced by venture capital trends and investor behavior. This project uses publicly available transactional data from the Indian reality show *Shark Tank India* as a proxy for early-

stage investment analysis. The core goal was to transform raw pitch data into actionable insights, providing a clear view of which industries and founder profiles are most successful at securing funding on the show.

2. Tools Used

Tool	Purpose
Python	Primary tool for data manipulation, aggregation, and statistical analysis.
Pandas (Library)	Used for cleaning the raw CSV data, structuring the 80+ columns, and running group-by calculations (e.g., success rates, total funding).
SQL (MySQL/Queries)	Developed queries to structure data by domain, calculate funding metrics, and unpivot investor-specific data, preparing the foundation for Python execution.
Power BI (Target)	Intended for final visualization and creation of an interactive dashboard based on the exported CSV results.

3. Conclusion

The analysis provided a robust view of the Shark Tank India investment landscape.

- **Success Rate by Gender:** All-Female founding teams recorded the highest deal success rate (**64.8%**), underscoring a strong correlation between female-led pitches and successful funding outcomes.
- **Funding Volume:** The highest investment concentrations were found in the consumer-facing sectors of **Beauty/Fashion** and **Food and Beverage**, confirming these as the most popular and capital-intensive domains on the show.
- **Investor Preferences:** Detailed analysis showed distinct preferences, such as **Namita Thapar's** focus on **Medical/Health** and **Aman Gupta's** significant commitment to the **Beauty/Fashion** sector.

These structured data files provide all the necessary inputs for creating a professional and interactive **Power BI** dashboard, fulfilling the core deliverables of this project.