

# Shark Tank India Data Analysis Report

## Introduction

- **Dataset:** Shark-Tank-India.csv
- **Source:** Shark Tank India Episodes Dataset
- **Total Records:** 634 startups across 4 seasons

This notebook presents an in-depth analysis of startup pitches, investments, and investor behavior on Shark Tank India using Apache PySpark. It explores industry trends, geographic distribution, financial metrics, investor participation, and startup performance.

## 1. Dataset Description

**Columns:** 80 attributes including Season Number, Startup Name, Episode Number, Industry, Yearly Revenue, Deal Valuation, and detailed shark investment fields.

**Data Quality:** Clean dataset with minimal missing values.

## 2. Data Exploration & Cleaning

- Checked for null values and data types
- Performed summary statistics on numerical columns
- Verified categorical distributions

## 3. Descriptive Analytics

- **Industry Distribution:** Top sectors: Food & Beverage (22.4%), Beauty/Fashion (20.0%), Technology/Software (9.1%)
- **Geographic Distribution:** Maharashtra leads (24.8%), followed by Delhi (13.9%) and Karnataka (10.9%)
- **Revenue & Valuation Distributions:** Average revenue ₹665.9L, median ₹230L, success rate 38.2%

## 4. Advanced Analysis

- **Shark Participation:** Multi-shark deals correlate with higher investments
- **High-Valuation Startups:** 145 startups with valuations  $\geq$  ₹2000L
- **Gender Distribution:** Presenters proxy shows male vs female distribution
- **Investment Over Time:** Seasonal trends in top industries

## 5. Key Insights

- Food & Beverage and Beauty/Fashion dominate pitches and investments
- Technology startups command premium valuations
- Multi-investor deals yield larger funding amounts
- Maharashtra and Delhi are primary startup hubs

## 6. Visualizations

- Industry distribution pie chart
- Top 10 states bar chart
- Revenue distribution histogram
- Deal valuation vs total deal amount scatter
- Gender distribution pie chart (presenters): male vs female
- Seasonal bar chart: top industries by investment

## 7. Conclusion

The Shark Tank India dataset reveals a dynamic entrepreneurial ecosystem with strong sectoral and regional patterns. Key success factors include proven revenue, team composition, and industry relevance. This analysis provides actionable insights for entrepreneurs, investors, and policymakers to drive growth and innovation in the Indian startup landscape.

Future work could include predictive modeling for deal success and deeper demographic analysis of founders.