

A black background with a black square

Description automatically generated with medium confidence

**RV COLLEGE OF ENGINEERING®**

**(Autonomous Institution Affiliated to Visvesvaraya Technological University, Belagavi)**

**Exploratory Data Analysis on Capital Funding invested on Startups**

**COMPUTING & ANALYTICS LAB**

**22MCE14L**

**PROJECT REPORT**

Submitted by

**Name (USN)**

Under the guidance of

**Prof. Name**

Assistant Professor,

Department of Computer Science and Engineering

RV College of Engineering

Bengaluru - 560059

***Submitted in partial fulfillment of the requirements for the award of degree of***

**MASTER OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**

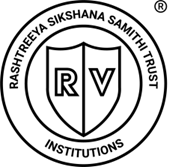
**DEPARTMENT OF COMPUTER SCIENCE AND**

**ENGINEERING**

2023 - 24

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

[**Bengaluru**](http://www.bengaluruairport.com/)**– 560059**



**CERTIFICATE**

Certified that the project work titled **“Exploratory Data Analysis on Capital Funding invested on Startups”** carried out by **Name (USN)** bona fide student of **RV College of Engineering, Bengaluru** submitted in partial fulfillment for the award of **Master of Technology** in **Computer Science Engineering** of **RV College of Engineering, Bengaluru affiliated to Visvesvaraya Technological University, Belagavi** during the year **2023-24**. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirement in respect of the project work prescribed for the said degree.

**Dr. Ramakanth Kumar P Prof. Name**

Head of Department, Assistant Professor,

Department of Computer Department of Computer

Science and Engineering Science and Engineering,

RVCE, Bengaluru - 560059 RVCE, Bengaluru - 560059

**ABSTRACT**

The Indian startup ecosystem has witnessed remarkable growth in recent years, attracting significant attention from investors and analysts alike. This project delves into an exploratory analysis of top funding startups in India, aiming to uncover key trends, insights, and opportunities within this dynamic landscape.

The dataset comprises information on funding rounds, sectors, locations, investors, invested amount into the startups, meticulously curated and preprocessed for analysis. Through comprehensive exploratory data analysis (EDA), we present a detailed overview of the dataset, highlighting key summary statistics and visualizations that shed light on the distribution and characteristics of the data.

Our analysis focuses on identifying the top funding startups in India, examining their funding amounts, sectors, funding rounds, and prominent investors. Additionally, we delve into regional variations, sector-wise trends, and investor preferences to glean deeper insights into the Indian startup ecosystem.

The findings of this study not only provide valuable insights for entrepreneurs, investors, and policymakers but also offer a foundation for further research and strategic decision-making. By understanding the dynamics of funding in the Indian startup landscape, stakeholders can better navigate challenges, capitalize on opportunities, and contribute to the continued growth and innovation of the ecosystem.

Overall, this project contributes to the existing body of knowledge on Indian startups, offering a comprehensive analysis that illuminates key trends and opportunities within the vibrant and rapidly evolving startup ecosystem.

i

Table of Contents

ABSTRACT i

[CHAPTER 1 1](#_TOC_250040)

[INTRODUCTION 5](#_TOC_250039)

* 1. [OVERVIEW 5](#_TOC_250038)
  2. [OBJECTIVES 6](#_TOC_250037)
  3. [SCOPE 7](#_TOC_250036)
  4. [PROBLEM STATEMENT 8](#_TOC_250035)

[CHAPTER 2 7](#_TOC_250034)

[METHODOLOGY OF THE EXPLORATORY DATA ANALYSIS ON CAPITAL FUNDING RECEIVED BY STARTUPS 9](#_TOC_250033)

* 1. [BASIC METHODOLOGY 9](#_TOC_250032)

[CHAPTER 3 11](#_TOC_250020)

[DATASET AND PREPROCESSING 11](#_TOC_250019)

* 1. [ABOUT THE DATASET 11](#_TOC_250018)
  2. [DATA PREPROCESSING 13](#_TOC_250017)

[CHAPTER 4 17](#_TOC_250016)

[HYPOTHESIS TESTING 17](#_TOC_250015)

* 1. [HYPOTHESIS TESTING 17](#_TOC_250014)
  2. [SHAPIRO WILK TEST AND LEVENE’S TEST 18](#_TOC_250013)
  3. [KRUSKAL WALLIS H TEST 19](#_TOC_250012)

[CHAPTER 5 20](#_TOC_250011)

[EXPLORATORY DATA ANALYSIS 20](#_TOC_250010)

1. [.1 EXPLORATORY DATA ANALYSIS SULT FOR THE PREDICTIONS 20](#_TOC_250009)

[CHAPTER 6 42](#_TOC_250006)

[REQUIREMENT SPECIFICATION 42](#_TOC_250005)

* 1. [SOFTWARE REQUIREMENT 42](#_TOC_250004)
  2. [HARDWARE REQUIREMENT 42](#_TOC_250003)
  3. [LIBRARIES AND TOOLS REQUIREMENT 43](#_TOC_250002)

[CHAPTER 7 46](#_TOC_250006)

[CONCLUSION 46](#_TOC_250005)

* 1. [CONCLUSION 46](#_TOC_250003)

List Of Figures

|  |  |  |
| --- | --- | --- |
| Fig 3.1 | Dataset Columns Screenshot | 13 |
| Fig 3.2 | Missing Data | 15 |
| Fig 3.3 | Type Conversion | 16 |
| Fig 3.4 | Data Reformatting | 16 |
| Fig 3.5 | Duplicate Values | 17 |
| Fig 4.1 | Hypothesis Testing | 18 |
| Fig 5.1 | Funding Statistics with Time | 23 |
| Fig 5.2 | Top 20 months that received maximum number of fundings | 23 |
| Fig 5.3 | Bottom 15 months that received the least number of fundings | 24 |
| Fig 5.4 | Funding Distribution Over the Years | 24 |
| Fig 5.5a | Maximum Funding | 25 |
| Fig 5.5b | Investigating the particulars of startups to gain further insights | 25 |
| Fig 5.5c | Breakdown the investors and their investment | 26 |
| Fig 5.6 | Minimum Funding | 27 |
| Fig 5.7 | Count of Funding Rounds | 28 |
| Fig 5.8 | Number of Fundings received | 29 |
| Fig 5.9 | Pie Chart of Funding Industry Distribution | 30 |
| Fig 5.10 | Bar Graph of Funding Industry Distribution | 31 |
| Fig 5.11 | Distribution of Sub Industries Funding | 32 |
| Fig 5.12 | Distribution of Funding across Cities | 34 |
| Fig 5.13 | Bar Graph of Investors Investment Count | 36 |
| Fig 5.14 | Word Cloud of Popular Investors | 38 |
| Fig 5.15 | Bar Graph of Investment Type | 39 |
| Fig 5.16 | Pie Chart of Investment Type | 41 |

**Chapter 1: INTRODUCTION**

The Overview of the Project, the Problem Statement, Objectives, and the Basic Methodology or the Flow of the model is defined and discussed in this chapter. This is helpful to further discuss the concepts behind the model and project in depth.

### OVERVIEW

The success of a start-up is heavily reliant on ideas, innovation, and execution. A startup is a company or project begun by an entrepreneur to seek, develop, and validate a scalable business model. While entrepreneurship refers to all new businesses, including self-employment and businesses that never intend to become registered, startups refer to new businesses that intend to grow large beyond the solo founder. Startups face high uncertainty and have high rates of failure, but a minority of them do go on to be successful and influential. Some startups become unicorns, privately held startup companies valued at over US $1 billion.

One of the startup's first tasks is raising a substantial amount of money to further develop the product. To do that, they have to make a strong argument, if not a prototype, that supports their claim that their idea is truly new or a great improvement to something on the market. However, these alone may not be sufficient. It is the financial backing or “funding” from investors that enables or allows start-ups and other entrepreneurial projects with great aspirations to dream big, achieve wealth, and make a significant impact. In this project, I will be investigating the funding that start-ups in India received from various investors. The dataset provided includes separate csv files for each year’s funding data. These files contain details about the start-ups, the funding they received, and the investors’ information.

Exploratory Data Analysis is an important step in any Data Analysis or Data Science project. It is the process of investigating the dataset to discover patterns, and anomalies or outliers, and form hypotheses based on our understanding of the dataset. Exploratory Data Analysis involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand Exploratory Data Analysis with the help of an example dataset. In this dataset, we used Pandas, Numpy, Matplotlib, Seaborn libraries.

### 

### OBJECTIVES

The objectives collectively provide a roadmap for the analysis and exploration of top funding startups in India, facilitating a comprehensive understanding of the dynamics and trends within the Indian startup ecosystem

1. **To analyze the distribution of funding amounts among top startups in India:** This objective focuses on understanding the range and distribution of funding received by startups, identifying outliers, and exploring factors influencing funding amounts.
2. **To identify key sectors attracting the most investment:** This objective aims to determine the sectors or industries that have been most successful in attracting investment, uncovering trends and patterns in sector-wise funding distribution.
3. **To examine regional variations in startup funding:** This objective involves analyzing the geographic distribution of startup funding across different regions of India, exploring variations in funding amounts, sector preferences, and investor activity.
4. **To profile top funding startups and their characteristics:** This objective focuses on profiling the startups that have received the highest amounts of funding, examining their sector, business model, founding team, and other relevant characteristics.
5. **To understand investor behavior and preferences:** This objective seeks to analyze the behavior and preferences of investors participating in the Indian startup ecosystem, identifying key investors, investment patterns, and factors influencing investment decisions.
6. **To derive insights for stakeholders:** This objective aims to derive actionable insights from the analysis that can inform strategic decision-making for entrepreneurs, investors, policymakers, and other stakeholders involved in the startup ecosystem.

### SCOPE

By defining this scope, the project aims to provide a structured and focused analysis of top funding startups in India while considering various dimensions such as sectors, regions, and investor behavior

* + 1. ***Exploratory Data Analysis (EDA):***

The project will primarily focus on conducting exploratory data analysis (EDA) on a dataset containing information about top funding startups in India. EDA will involve descriptive statistics, data visualization, and summary analysis to gain insights into various aspects of the dataset.

* + 1. ***Top Funding Startups:*** The analysis will prioritize examining startups that have received significant amounts of funding, with a focus on understanding their characteristics, funding sources, and sectoral distribution.
    2. ***Sector Analysis and Regional Analysis:***

The project will include an in-depth analysis of sectors or industries that have attracted substantial investment, identifying trends, and patterns in sector-wise funding distribution. Regional variations in startup funding across different states or cities in India will be explored to understand geographical trends and preferences.

* + 1. ***Investor Analysis***:

The project will analyze the behavior and preferences of investors participating in the Indian startup ecosystem, identifying key investors, investment patterns, and factors influencing investment decisions.

* + 1. ***Challenges and Opportunities***

While not the primary focus, the project may briefly discuss challenges faced by startups in India and opportunities for growth and innovation within the ecosystem based on the analysis.

* + 1. ***Recommendations:***

The project may provide recommendations for stakeholders such as entrepreneurs, investors, and policymakers based on the insights derived from the analysis.

### PROBLEM STATEMENT

The problem statement for our exploratory data analysis is a group of business questions that can be answered and inferred from the analysis. We define these questions in this section. In addition, we also identify the problem statement to discover new patterns and trends that may not be intuitive or cannot be defined beforehand. Thus the following questions form the foundation for our exploration, as well as we set out to find particular patterns and trends.

1 . How has the funding trend for Indian start-ups changed?

2. Which sectors or industries received the most funding?

3. Who are the major investors in the Indian start-up ecosystem?

4. Are there any specific cities in India that are attracting more start-ups than the others?

5. Which cities in India have attracted the most funding?

6. Which start-ups have shown the most growth in terms of funding received over the years?

## Chapter 2: METHODOLOGY OF THE EXPLORATORY DATA ANALYSIS ON CAPITAL FUNDING RECEIVED BY STARTUPS

1. ***Business Understanding:*** This is the initial phase where business understanding is to be defined by identifying the problems, objectives, and outcomes. We identified the key stakeholders, investors, and other pertinent attributes of the startup ecosystem.
2. ***Loading and extracting the dataset***: We must import the dataset into Jupyter notebook for further processing. In this case, I am choosing the Top Funding Startups dataset which contains the necessary information to carry out the analysis and load it into Jupyter notebook using pandas library.
3. ***Installing and importing all the necessary dependencies***: Numpy, pandas, matplotlib, plotly, squarify, wordcloud, seaborn, and other libraries are to be installed. Since we have Anaconda Navigator, we can install these libraries by running Anaconda Prompt and running the command pip install followed by the name of the library of interest.
4. ***Data cleaning and preprocessing***: This phase involves preparing the data for analysis. This could include data cleaning, data transformation, and data integration. We load the data from the dataset. Data Cleaning is performed by identifying and correcting errors in the data, deleting all null values, removing duplicate and corrupted data, separating the data, removing punctuations and stop words, and formatting the attributes to follow a consistent structure.
5. ***Hypothesis Testing:***

Hypothesis testing is a fundamental concept in data science used to make inferences about a population based on sample data. Hypothesis testing typically involves these steps. We start by defining the null hypothesis (H0) and alternative hypothesis (H1). The null hypothesis represents the default assumption or no effect, while the alternative hypothesis represents what you're trying to prove or disprove. We apply three statistical test suitable for our dataset that include Shapiro-Wilke test, Levene’s test, and Kruskal-Wallis test. Based on the comparison of p-value and alpha value, we will conclude the variance among data.

1. ***Exploratory Data Analysis***: This stage involves exploring the data to understand its characteristics and patterns. This involves visualizing the data, calculating descriptive statistics, and identifying correlations. We will utilize various techniques to gain insights into the dataset.

Visualizations such as histograms, box plots, scatter plots, bar graphs, pie charts, word clouds, square plot will be employed to visualize the distribution of funding amounts, explore the relationships between different variables, and identify any outliers or anomalies. Descriptive statistics including mean, median, standard deviation, and percentiles will be calculated to summarize the central tendency, spread, and variability of the data. Additionally, correlation analysis will be conducted to assess the strength and direction of relationships between pairs of variables, providing further insight into potential dependencies or associations within the dataset.

Through these exploratory techniques, we aim to uncover underlying patterns, trends, and relationships that will inform subsequent stages of analysis and interpretation.

1. ***Analysis of Business Questions***: This phase involves using the insights gained from the exploratory data analysis and hypothesis testing to answer the business questions identified earlier. This could involve building predictive trends, segmenting the data, or conducting further statistical tests. With the analysis and close observation with the help of visualization, we will be equipped to answer the identified questions.

### 

### Chapter 3: DATASET and PREPROCESSING

### ABOUT THE DATASET

More than 1870 startups are included in the dataset. The funding received by these startups in various rounds and time exceeds 2300. The dataset has the Creative Commons 1.0 Public License which means the information is open to the public and can be used for a variety of analytical tasks of interest. The dataset it built by collecting the information from trak.in website.

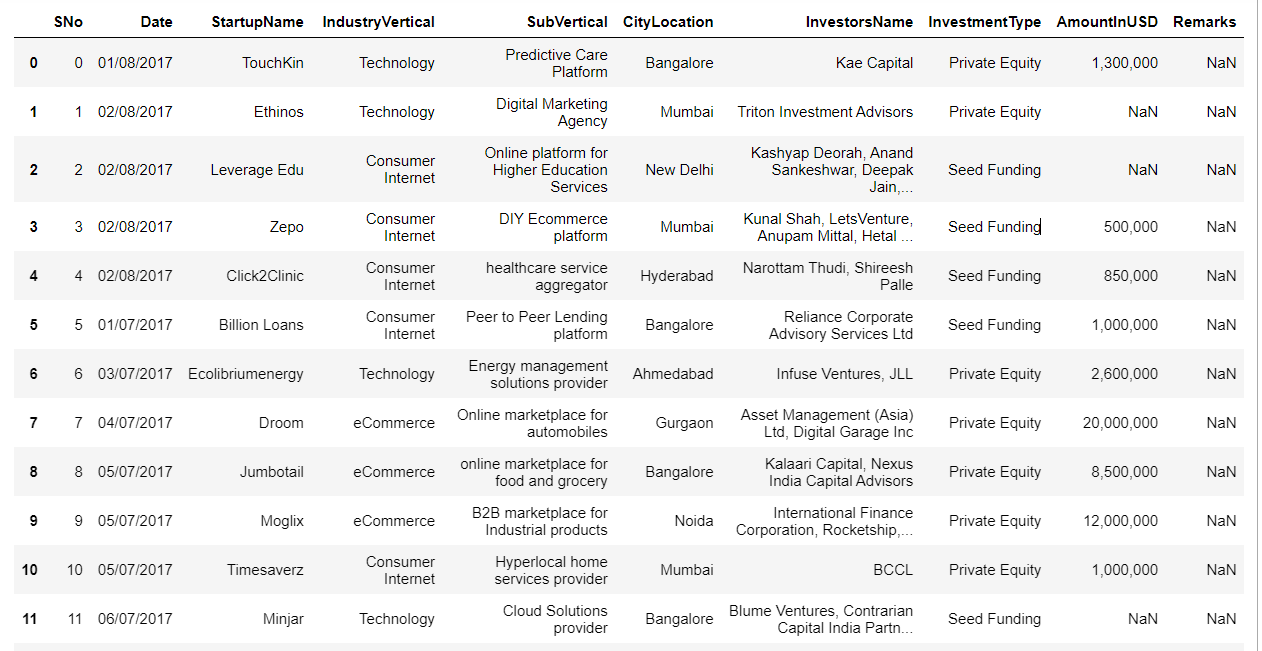
Each tuple consists of Serial Number, Date, Startup Name, Industry Vertical, Sub Vertical, City and Location, Investors Name, Investment Type, Amount in USD, and Remarks about the funding. It comprises of 10 columns and 2372 rows in the dataset.

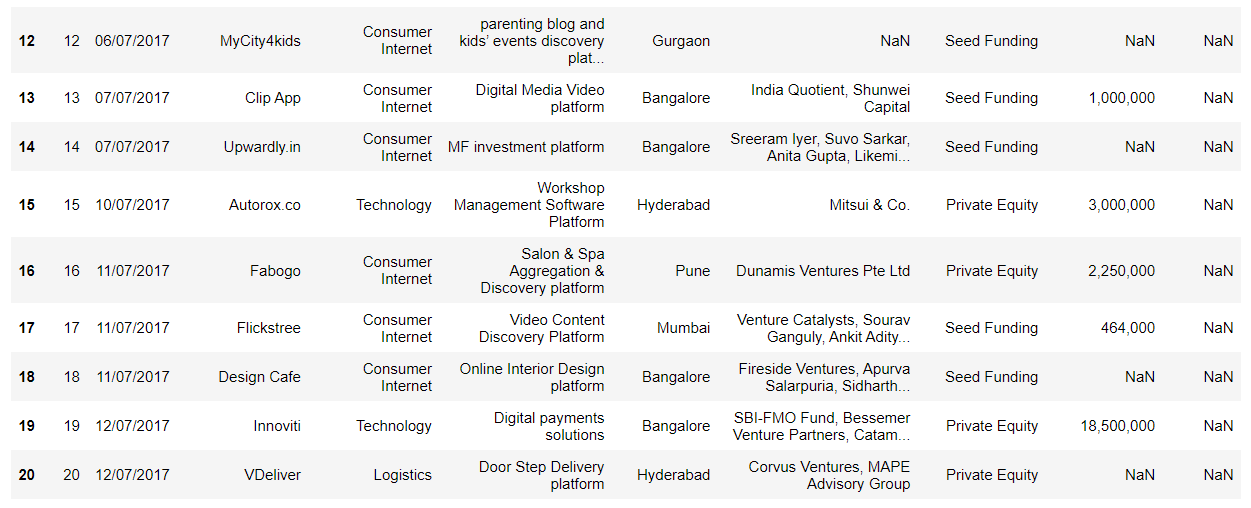
The dataset is almost clean and well-structured. The columns have been clearly identified and labelled. However, we will need to reformat some of the rows of the data present to make it suitable for analysis.

Below are the columns present in the Dataset:

1. **SNo**: Serial number that indexes the entries.
2. **Date**: Date in the format dd/mm/yyyy on which the funding was finalized.
3. **StartupName**: The startup's name.
4. **IndustryVertical**: Category of the industry in which the startup specializes in.
5. **SubVertical**: Sub-category of the industry. It’s a further specialization of industry vertical.
6. **CityLocation**: The location where the startup is based.
7. **InverstorsName**: The name of the investor’s funding the startups.
8. **InvestmentType**: This specifies the type of investment made.
9. **AmountInUSD**: The amount of money received by the startup from funding in US Dollars.

10. **Remarks**: This specifies any additional information about the funding





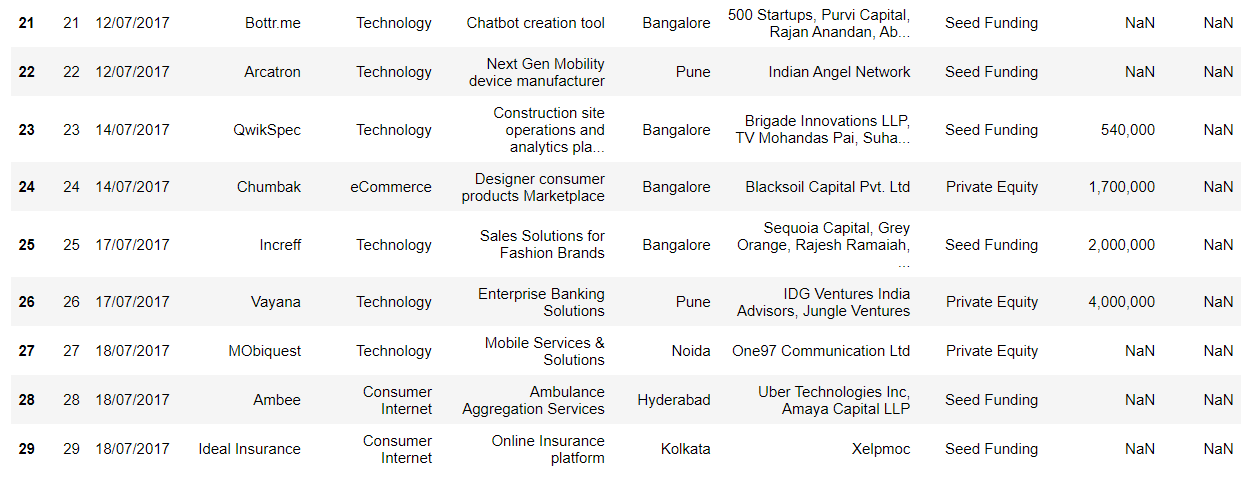


Fig 3.1 : Dataset Columns Screenshot

### DATA PRE-PROCESSING

In preparing the startup funding dataset for analysis, several crucial pre-processing and cleaning steps are essential to ensure the quality and integrity of the data. Firstly, missing values within the dataset need to be addressed. This involves identifying the presence of missing data in each column and determining the most appropriate strategy for handling them, whether through imputation techniques or deletion of incomplete records.

Following this, attention must be given to the data types present in the dataset. Checking the data types of each column ensures that they align with the analysis objectives and computational requirements. Conversion of data types may be necessary, particularly for fields representing numerical values or date/time information, to facilitate subsequent analysis.

Duplicate records pose a risk to the accuracy and reliability of the dataset. Therefore, identifying and removing duplicate entries is an essential step in the cleaning process. This helps to streamline the dataset and prevent redundancy in subsequent analyses.

Text data within the dataset may require standardization to ensure consistency and comparability across entries. This involves tasks such as converting text to a consistent case (e.g., lowercasing), removing unnecessary punctuation or special characters, and normalizing text representations to ensure uniformity.

Outliers in numerical data can skew statistical analyses and modeling outcomes. Thus, it is important to detect and address outliers using appropriate methods such as z-score or interquartile range analysis. Strategies for handling outliers may include removal, transformation, depending on the nature of the data and analysis goals.

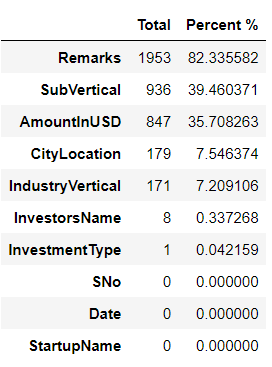
Feature engineering plays a crucial role in enhancing the predictive power of the dataset. This involves creating new features from existing ones or transforming features to better capture underlying patterns or relationships. Additionally, categorical variables must be encoded using suitable methods such as one-hot encoding or label encoding to enable their inclusion in analytical models.

***Missing Values:***

The following script will help us recognize the missing data in the dataset.

A white board with colorful objects

Description automatically generated with medium confidence

We obtain the results as follows.

A yellow square on a purple background

Description automatically generatedA better graphical visualization of the missing data in the dataset is summarized below.

Fig 3.2: Missing Data

We notice that the "Remarks" column is missing 82% of data. Since it doesn't describe the data accurately, we can clear it. As such high missing data would be redundant for our analysis.

***Type Conversion:***

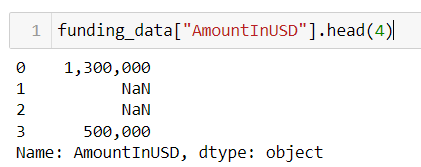
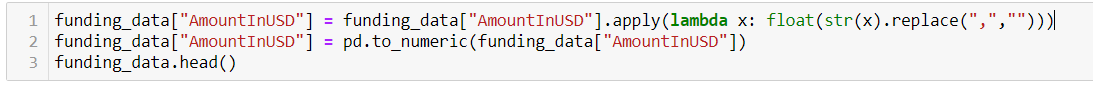


Fig 3.3 Type Conversion

The amount funded to the startup, represented by AmountInUSD column, is actually present as a literal string. For our analysis, we will need it to be a numeric in order to carry out certain mathematical plots. So we will change the type of AmountInUSD from string to integer.

The following script will achieve that:

T

A screenshot of a computer program

Description automatically generated

Fig 3.3 Type Conversion

The conversion is complete. As you can see, the dtype or the datatype of AmountInUSD has been converted from object to float64.

***Data Reformatting:***

The date column has some inconsistent date syntax in countable numbers. We reformat it to a suitable consistent syntax for easier understanding and to maintain integrity.

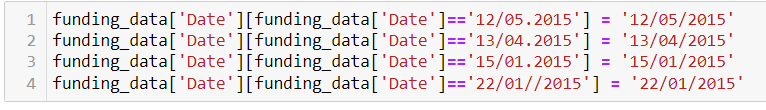


Fig 3.4: Data Reformatting

***Duplicate Values:***

We check to see if the dataset contains any replicated rows. If the data is duplicated, it is redundant and we can remove it make a succinct dataset.

A screenshot of a computer

Description automatically generated

Fig 3.5a: Duplicate Values

We observe that the same startup names have been repeated. Let’s delve into one of these to analyze if the rows are identical by displaying it.

A screenshot of a computer

Description automatically generated

Fig 3.5b: Duplicate Values

We notice that even though it’s the same startup name, the row is not repeated as they represent a separate funding from different investors and at a different point of time. So our dataset doesn’t contain duplicates.

### Chapter 4: HYPOTHESIS TESTING

### HYPOTHESIS TESTING

The Hypothesis testing is a statistical method that allows us to make inferences or draw conclusions about a population based on a sample of data. This phase was particularly important as it enabled me to test assumptions, compare groups, and determine the statistical significance of the patterns and relationships I had observed in the dataset. This step was crucial in providing reliable and meaningful insights from the data.

Upon completing the data cleaning process, which involved handling missing values, correcting inconsistencies, and transforming data types as necessary, we transition into testing the hypothesis. The necessary hypotheses is defined which are the Null and Alternative hypotheses.

Null Hypothesis: There is no difference in the amount of funding received by start-ups across different industry verticals.

Alternative Hypothesis: There is a difference in the amount of funding received by start-ups across different industry verticals.

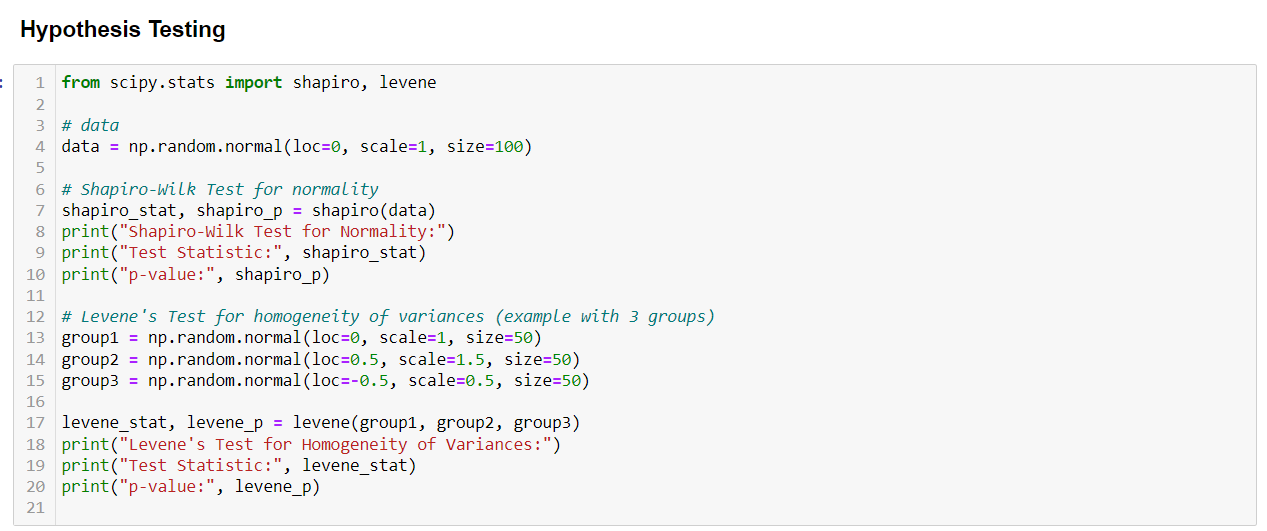


Fig 4.1: Hypothesis Testing

**The Shapiro-Wilk Test for normality:** It is a statistical test used to determine whether a given dataset follows a normal distribution or not. It assesses whether the observed data points are likely to have been drawn from a population with a Gaussian distribution, also known as a normal distribution.

The test produces a test statistic and associated p-value. The null hypothesis of the Shapiro-Wilk test is that the data are normally distributed. A low p-value indicates evidence against the null hypothesis, suggesting that the data significantly deviate from a normal distribution. Conversely, a high p-value suggests that there is no significant departure from normality, supporting the assumption that the data are normally distributed.

A black text with black letters

Description automatically generated with medium confidence

Fig 4.1a: Shapiro Wilk Test

**Levene's Test for homogeneity of variances**: It is a statistical test used to assess whether the variances of two or more groups are equal or homogeneous. It evaluates whether the spread or dispersion of data points within each group is similar across all groups.

The test works by comparing the absolute deviations of individual data points from their group means, rather than comparing the raw variances directly. This can make the test more robust against violations of normality assumptions compared to other tests for homogeneity of variances.

Levene's Test produces a test statistic and associated p-value. A low p-value suggests evidence against the null hypothesis of equal variances, indicating that the assumption of homogeneity of variances may be violated. Conversely, a high p-value suggests that there is no significant difference in variances among groups, supporting the assumption of homogeneity.

A number and a number

Description automatically generated with medium confidence

Fig 4.1b: Levene's Test

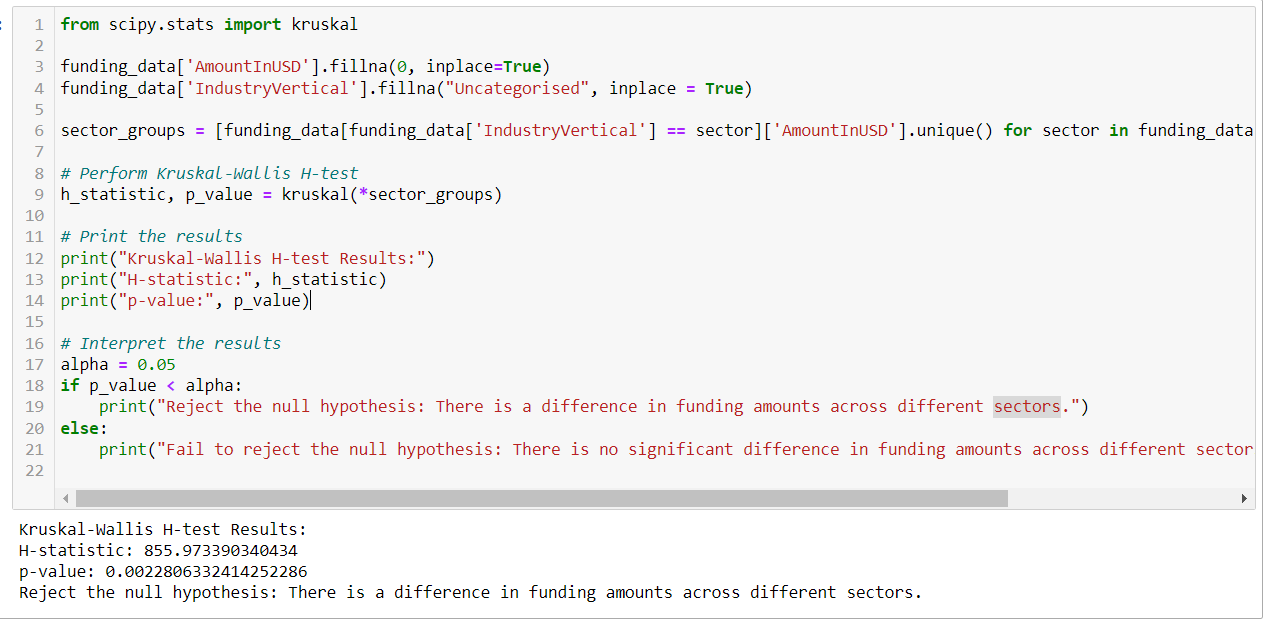
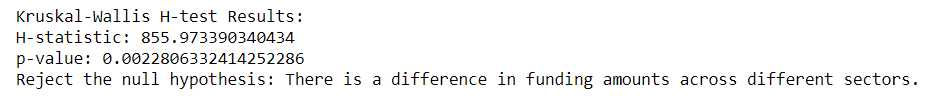


Fig 4.1c: Kruskal Wallis H Test

**Kruskal-Wallis H-test:** It is a non-parametric statistical test used to determine whether two or more independent samples originate from populations with the same distribution or not. The test works by ranking all data points from all groups together, then calculating a test statistic based on the ranked data. The test statistic, denoted as H, measures the degree of difference among the group means ranks. If the samples are drawn from populations with the same distribution, the ranks of observations across groups should be similar.

Thus, a low H-value suggests that the distributions are similar, while a high H-value indicates significant differences in distributions among groups. The test produces a p-value, where a small p-value (typically below a predetermined significance level, like 0.05) suggests rejecting the null hypothesis, indicating that at least one group significantly differs from the others in terms of central tendency.



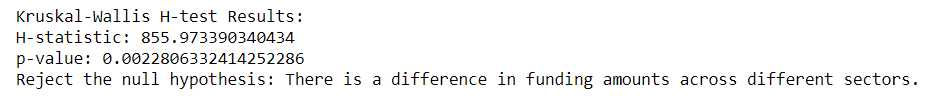


Fig 4.1d: Kruskal Wallis H Test

This indicates that our initial null hypothesis said that there is no difference in the amount of funding received by start-ups across different industry verticals is false. We notice lower p value and higher H-statistic indicating that the taken columns of Industry Vertical and Amount they received as funding has a large variance.

**Chapter 5: EXPLORATORY DATA ANALYSIS**

Exploratory Data Analysis is carried out subsequently. It is a crucial initial step in data analysis that focuses on summarizing the main characteristics of a dataset, often employing graphical and statistical techniques. Its primary goal is to gain insights into the data, understand its structure, identify patterns, detect anomalies, and formulate hypotheses for further investigation.

Exploratory Data Analysis is a foundational approach serving as a preliminary step to understand the structure, patterns, and relationships within a dataset. By employing a combination of data visualization, descriptive statistics, and exploratory techniques, it aims to uncover insights, detect anomalies, and formulate hypotheses that guide further analysis. Through this, analysts gain a deeper understanding of the dataset's characteristics, identifying trends, outliers, and potential correlations between variables. This process not only aids in data cleaning and preparation but also informs subsequent modeling, hypothesis testing, and decision-making.

It is an iterative process, allowing analysts to refine their understanding of the data and generate new insights as they explore different aspects of the dataset. It plays a crucial role in extracting actionable insights from data, enabling informed decision-making, and driving business value. Here’s the insights we can derive from the startup ecosystem:

*Historical Perspective*: The analysis begins by collecting historical data on funding received by startups, businesses, or projects on a monthly basis over a certain period, spanning several years. This data typically includes information such as the amount of funding, the number of deals, the types of investors involved, and the industries or sectors receiving funding.

*Temporal Trends*: The analysis examines temporal trends in funding, looking for patterns or fluctuations in funding levels, deal activity, and investor behavior over time. This involves visualizing the data using time series plots, line charts, or heatmaps to identify trends, seasonality, or cyclical patterns.

*Key Metrics*: Key metrics such as total funding amount, average deal size, funding rounds, and the number of unique investors are analyzed to understand the overall health and dynamics of the funding ecosystem. Changes in these metrics over time can provide insights into shifts in investor sentiment, market conditions, and the maturity of the ecosystem.

*Sectoral Analysis*: The analysis explores how funding trends vary across different sectors or industries. By categorizing funding data based on industry verticals (e.g., technology, healthcare, fintech), the analysis can identify sectors experiencing rapid growth, emerging trends, or declining investment activity.

*Geographical Analysis*: Geographical variations in funding trends are examined to understand regional dynamics and investment hotspots. This involves comparing funding levels, deal flow, and investor preferences across different regions, countries, or cities to identify areas of growth and opportunities for investment.

*Impact of Events*: The analysis considers the impact of significant events or external factors such as economic downturns, regulatory changes, technological advancements, or global events (e.g., pandemics) on the funding ecosystem. By correlating funding trends with external events, the analysis can assess how these factors influence investor behavior and funding patterns.

*Predictive Insights*: Finally, based on historical trends and patterns, the analysis provides predictive insights into how the funding ecosystem may evolve in the coming years. This involves forecasting future funding trends, identifying emerging sectors or technologies, and anticipating changes in investor preferences or market dynamics.

**How has the funding ecosystem changed with time and will evolve in the coming years?**

We will analyze the characteristics of funding received monthly over time to derive the historical trends in the investment made. Bar graph can be plotted to visualize the data.

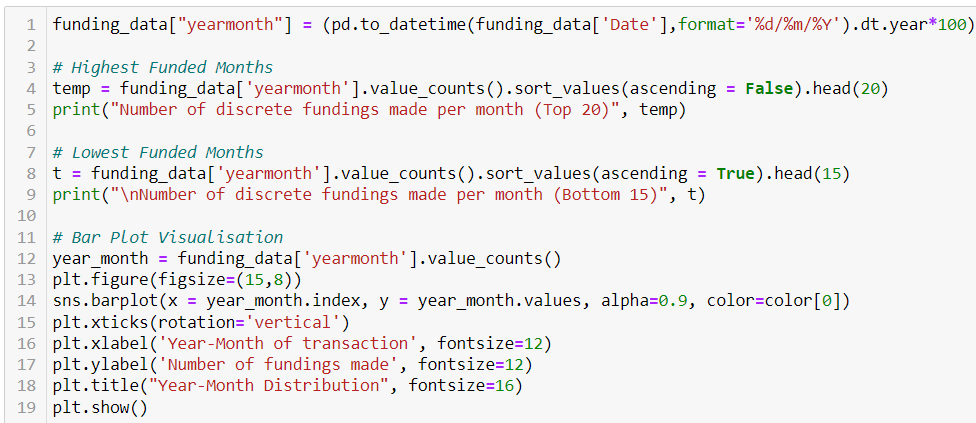
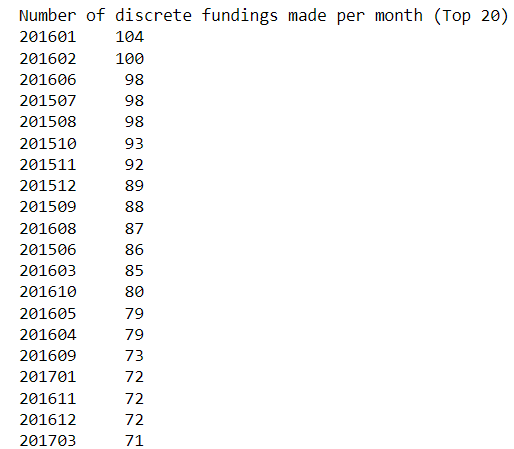


Fig 5.1: Funding Statistics with Time

**Fig 5.2: Top 20 months that received maximum number of fundings**



**Fig 5.3: Bottom 15 months that received the least number of fundings**

A screenshot of a computer screen

Description automatically generated

A graph of a number of columns

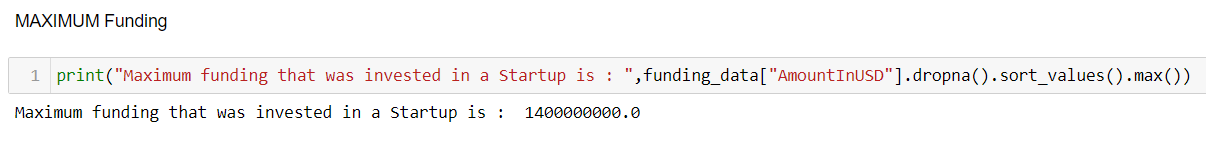
Description automatically generated with medium confidence

Fig 5.4: Funding Distribution Over the Years

We derive the information that startups received peak funding at the start of 2016 as the number exceeded over 100 with January having the highest offers of 104. We also notice that the numbers gradually increased from 2015 to 2016, then we see a gradual decrease. Overall, the fundings made is consistent with an average of 70 different investments.

**Let's analyze the amount of funding received by the startups. We will uncover the maximum, minimum, average value of the money invested.**

*Fig 5.5a: Maximum Funding*

**

The highly valued startup received a funding of $1.4 billion. For a startup, this is a lot of valuation. These startups are respected and labelled as a unicorn for its rarity. If a startup can receive over a billion dollars of funding, it must be pioneering in a particular domain and has shown significant growth.

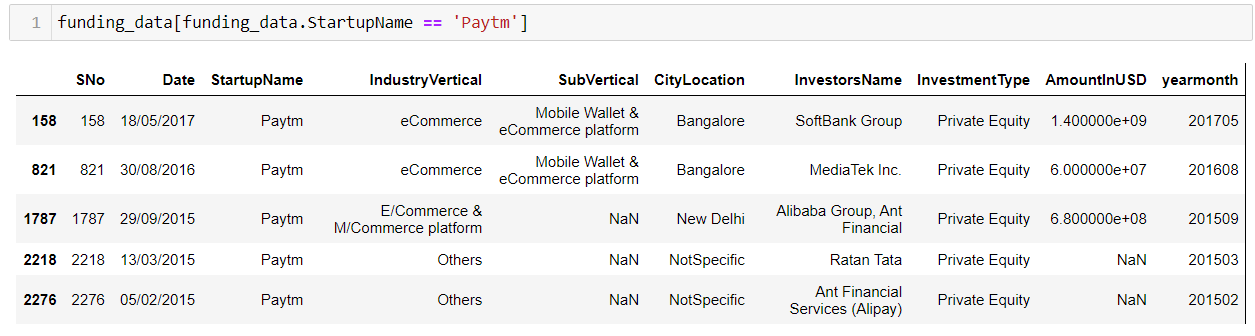
*Fig 5.5b: Investigating the particulars of these startups to gain further insights*

A screenshot of a computer

Description automatically generated

Paytm and Flipkart are the unicorn startups that have a valuation of over a billion of USD dollars. Paytm is a payment system and financial technology company. Initially launched as a platform for mobile recharging and utility bill payments, it has since diversified into various segments such as online shopping, ticket booking, digital wallets, and financial services. It has emerged as one of leading fintech companies, playing crucial role in driving the country's digital economy and financial inclusion.

Flipkart is one of the largest e-commerce platforms, offering a wide range of products and services to customers across the country. It operates as an online marketplace, connecting buyers and sellers across various product categories. It provides a platform for purchasing a diverse range of products, including electronics, fashion apparel, home appliances, books, beauty products, groceries, and more. Through continuous technological advancements, strategic partnerships, and a customer-centric approach, Flipkart continues to redefine the way India shops, making it a trusted destination for consumers and sellers alike.

*Fig 5.5c: Breakdown the investors, their investment, as well as where and when did they fund:*

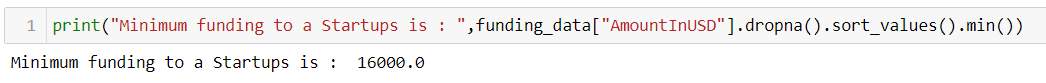
Observe that majority of the investment or funding was received after the startup established a strong foothold in its domain. We see that in 2015, the startup started getting support with 680 million being offered at the end of that fiscal year. It was sufficient to scale up their operations. Then at the end of 2016, a year later, it received a 60 million dollars. It continued growing, and in the mid of the year 2017, it received the highest unicorn value funding. Also notice it switched it’s industry as it grew and even had subvertical domains.

A screenshot of a computer

Description automatically generated

Observe the same type of trend with Flipkart. As the online retail market started to blow in 2015, it received 50 million in funding. They were one of the few online shopping markets in India and this was seen as an opportunity. They received a large 700 million USD Dollar a month later. This was sufficient to grow their operations as we don’t see any fundings in the next year. In the year 2017, they had grown as a big player of eCommerce market and that was when they received the billion dollar funding. Also, we can notice that all of these fundings received by these startups are of private equity. It implies that these startups had a tremendous idea that everyone wanted a share of.

*Fig 5.6a: Minimum Funding*:



The least amount of funding that was made was 16,000 dollars. For a small startup at its earliest stage, this is a sufficient amount of money to keep the operations afloat. If the idea or the product sold by these startups are exceptional and they perform well in the industry, they will grow up. Many of the investors test the waters by giving them with minimum sufficient funds. If they are able to utilize that and succeed, they will receive more funding.

*Fig 5.6b: Investigating the particulars of these startups to gain further insights*

A screenshot of a computer

Description automatically generated

We notice another interesting fact. All the low funded startups doesn’t have a specific industry they excel at. So we can say if a startup has a clear mission and field they work on, they get more funding. Also interesting to notice that the same investor has multiple fundings, all are seed funding which is the investment at the beginning stage of a startup. This investor is making a small amount of funding to numerous startups showcasing the trust in startup ecosystem and encouraging the development of the startups.

*Fig 5.7: Count of Funding Rounds:*

A screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

Top Funded Startups: The table lists the startups that have received the highest number of fundings. Swiggy and UrbanClap lead the list with 7 and 6 fundings respectively, indicating that they have been successful in attracting multiple rounds of investment.

Distribution of Fundings: The majority of startups listed have received between 4 to 7 fundings, with a total of 1981 fundings distributed among them. This suggests that the startup ecosystem is dynamic, with various companies attracting different levels of investor interest.

Diverse Industries: Startups from various industries are represented in the list, including food delivery (Swiggy), home services (UrbanClap), healthcare (Medinfi), transportation (Jugnoo), e-commerce (Flipkart), and fintech (Paytm). This diversity indicates that funding opportunities exist across different sectors of the economy.

Investor Confidence: The presence of well-known startups such as Flipkart, Paytm, and Byju's among the top funded companies reflects investor confidence in these established players. Additionally, startups like Tracxn and NoBroker with relatively fewer fundings but strong market presence suggest that investor interest may not solely depend on the number of fundings but also on factors like growth potential, market traction, and business model innovation.

Potential Growth Areas: Startups like Wooplr, Niki.ai, and TinyStep with moderate fundings have the potential for growth and may attract further investments in the future. Understanding the factors driving their success and market positioning could provide valuable insights for investors and entrepreneurs alike.

A graph of a number of small blue bars

Description automatically generated with medium confidence

Fig 5.8: Number of Fundings received

**What type of industries are favored by investors for funding (OR) the domain of the startups that receive funding**

*Let's analyze the SECTOR or INDUSTRY of the startups that receiving funding*

A screen shot of a computer code

Description automatically generated

A screen shot of a computer

Description automatically generated

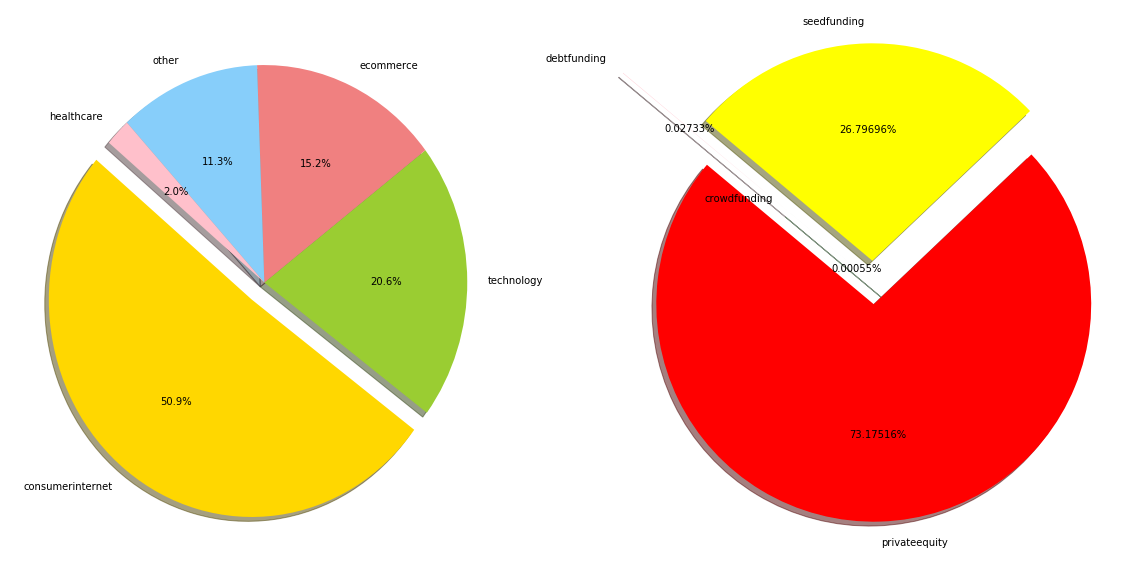


Fig 5.9: Pie Chart of Funding Industry Distribution

A graph of a number of green squares

Description automatically generated with medium confidence

Fig 5.10: Bar Graph of Funding Industry Distribution

Dominance of Consumer Internet and Technology: The data indicates that the Consumer Internet and Technology sectors attract the highest number of fundings, with 772 and 313 fundings respectively. This suggests a strong investor interest in startups operating in these sectors, which encompass a wide range of online services, platforms, and technological innovations.

Rise of e-commerce: The e-commerce sector follows closely behind, with 224 fundings. This reflects the continued growth and potential of online retail businesses in capturing market share and meeting consumer demand for convenient shopping experiences.

Non-technological and Specialized segments: While Consumer Internet, Technology, and e-commerce dominate the funding landscape, the presence of industries such as Healthcare, Logistics, Education, and Food & Beverage indicates a diverse range of opportunities for startups. These industries may offer niches for innovation and disruption.

*Analysing even specific industries or sub industries for available data:*

A screen shot of a computer code

Description automatically generated

A white background with black text

Description automatically generated

A graph of scatters with numbers

Description automatically generated with medium confidence

Fig 5.11: Distribution of Sub Industries Funding

Overall, the funding distribution among specific sub-industries reflects the diverse opportunities and evolving trends within the startup ecosystem, with sectors such as healthcare, food delivery, fintech, education, and e-commerce witnessing significant investor interest and innovation. Understanding the dynamics and growth potential of these sub-industries can inform investment decisions, strategic partnerships, and entrepreneurial endeavors within the startup ecosystem.

Online Pharmacy: With 9 fundings, the online pharmacy sector demonstrates strong investor interest, likely driven by the increasing adoption of e-commerce in healthcare and the growing demand for convenient and accessible pharmaceutical services. Startups in this sub-industry aim to digitize the pharmacy experience, offering online ordering, prescription fulfillment, and doorstep delivery of medications.

Food Delivery Platform: Following closely with 8 fundings, food delivery platforms continue to attract significant investment, reflecting the popularity of online food ordering and delivery services among consumers. These platforms leverage technology to connect users with a wide range of restaurants, cafes, and eateries, offering convenience, choice, and efficiency in meal delivery.

Online Lending Platform: With 5 fundings, online lending platforms address the evolving needs of consumers and businesses for accessible and efficient financial services. These platforms leverage technology and data analytics to streamline the lending process, offering digital loan applications, quick approvals, and flexible repayment options to borrowers.

Online Learning Platform: With 4 fundings, online learning platforms cater to the growing demand for remote education and skill development opportunities. These platforms offer a variety of courses, tutorials, and training programs across diverse subjects and industries, enabling individuals to enhance their knowledge and skills conveniently from anywhere.

eCommerce Marketplace: With 4 fundings, eCommerce marketplaces remain a fundamental part of the digital commerce landscape, providing a platform for buyers and sellers to connect and transact online. These marketplaces offer a wide range of products across various categories, empowering businesses to reach a broader audience and consumers to access a diverse selection of goods and services.

**Do cities where the startups are based in play a major role in funding? Analyzing the investment received based on location**

*Analyzing the geography of where the startups are located*

A screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

A colorful squares with black text

Description automatically generated

Fig 5.12: Distribution of Funding across Cities

Startup Hub in Bangalore: Bangalore emerges as the leading startup hub with 634 startups located in the city. Known as the Silicon Valley of India, Bangalore's vibrant ecosystem boasts a conducive environment for entrepreneurship, driven by factors such as access to talent, robust infrastructure, supportive government policies, and a thriving network of investors and mentors.

Significant Presence in Mumbai and New Delhi: Mumbai and New Delhi follow closely behind Bangalore, with 449 and 385 startups respectively. These cities, being major economic and commercial centers, offer diverse opportunities for startups across various sectors, including finance, media, entertainment, technology, and healthcare. The presence of a large number of startups in Mumbai and New Delhi reflects their status as key hubs for business and innovation in India.

Rise of Gurgaon as a Startup Destination: Gurgaon emerges as a notable startup destination with 241 startups located in the city. As a satellite city of Delhi and a prominent part of the National Capital Region (NCR), Gurgaon offers infrastructure, connectivity, and business-friendly policies that attract startups, particularly in sectors such as technology, e-commerce, and fintech.

Pune and Noida as Emerging Startup Ecosystems: Pune and Noida represent emerging startup ecosystems with 91 and 79 startups respectively. These cities benefit from their proximity to major metropolitan areas like Mumbai and Delhi, as well as their own strengths in areas such as education, research, and manufacturing. With growing support from local governments, educational institutions, and industry associations, Pune and Noida are attracting increasing attention from startups and investors alike.

Diverse Presence Across Other Cities: Hyderabad, Chennai, and Ahmedabad also have a notable presence of startups, with 77, 67, and 35 startups respectively. These cities offer unique advantages such as strong technology infrastructure, skilled workforce, and strategic location, making them attractive destinations for startups seeking growth opportunities in sectors like IT, biotechnology, manufacturing, and logistics.

**Who are the important INVESTORS in the startup ecosystem?**

A computer code with text

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph with colorful squares

Description automatically generated

Fig 5.13: Bar Graph of Investors Investment Count

The data highlights the diverse investor landscape in the Indian startup ecosystem, with a mix of individual investors, angel investor networks, venture capital firms, and strategic investors playing crucial roles in fueling innovation, fostering entrepreneurship, and driving economic growth. Understanding the preferences, motivations, and investment strategies of these investors is essential for startups seeking funding and navigating the fundraising process.

Significance of Undisclosed Investors: Undisclosed investors represent the largest category with 99 investments. While the identity of these investors is not specified, their significant presence underscores the prevalence of private investments made by individuals, family offices, or investment firms that prefer to remain anonymous. Undisclosed investments are common in early-stage funding rounds and may indicate a willingness to support promising startups without seeking public recognition.

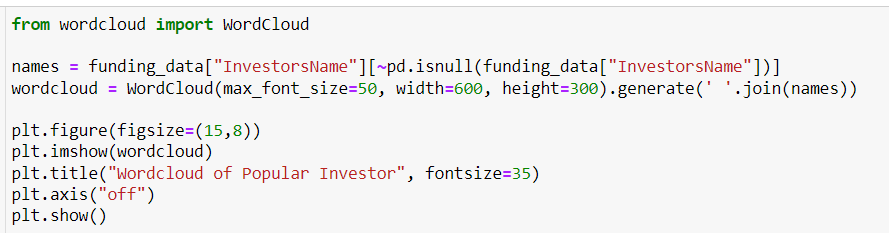
Impact of Individual Investors: Individual investors like Ratan Tata play a prominent role in the Indian startup ecosystem, leveraging their experience, networks, and resources to support entrepreneurial ventures. Ratan Tata's involvement in 24 investments highlights his active participation as an angel investor and mentor to startups, contributing not only capital but also strategic guidance and industry insights.

Angel Investor Networks: Angel investor networks such as the Indian Angel Network and Group of Angel Investors collectively account for 39 investments. These networks provide a platform for individual investors to pool their resources, share expertise, and co-invest in early-stage startups, thereby reducing risk and increasing the likelihood of success. Their involvement signals a supportive ecosystem for budding entrepreneurs and a willingness to fund innovative ideas with growth potential.

Venture Capital Firms: Venture capital firms like Kalaari Capital, Sequoia Capital, Accel Partners, and SAIF Partners are well-known players in the Indian startup ecosystem, specializing in providing growth capital and strategic support to high-potential startups. Their involvement in funding rounds, with a combined total of 48 investments, reflects their confidence in the scalability, market potential, and leadership of the startups they back.

Strategic Investments: Brand Capital, with 10 investments, focuses on strategic investments in startups that offer synergies with its parent company's business interests. This strategic approach enables Brand Capital to leverage its brand equity, market reach, and industry expertise to drive mutual value creation and market differentiation for both the startup and the parent company.   
  
*Analysis through wordcloud of the Investors data:*

The wordcloud library is a package specifically designed for creating word clouds from text data. Word clouds are visual representations of text data in which the size of each word indicates its frequency or importance within the text. They are often used to visualize word frequency distributions in a visually appealing and intuitive manner.



A black background with words

Description automatically generated

Fig 5.14: Word Cloud of Popular Investors

**Let's take a look into the TYPES of Funding and analyze the characteristics of the Fundings such as at what level and stage is this funding made**

A computer screen shot of a code

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A graph showing a number of furring

Description automatically generated

Fig 5.15: Bar Graph of Investment Type

Seed funding refers to the initial capital provided to startups or early-stage companies to support their business idea, product development, and initial operations. It is typically the first round of investment raised by a startup, often obtained from angel investors, venture capital firms, or accelerators. Seed funding is used to validate the market potential of the startup's product or service, conduct market research, build prototypes, and hire key team members. This early-stage financing helps startups progress from the concept stage to the point where they can demonstrate proof of concept or achieve key milestones that make them more attractive to larger investors in subsequent funding rounds.

Private equity refers to investments made in privately-held companies or assets by institutional investors, high-net-worth individuals, or private equity firms. Private equity investments typically involve acquiring ownership stakes in companies, either through equity or debt instruments, with the aim of generating attractive returns over the long term. Private equity investors provide capital to support various corporate activities such as growth initiatives, acquisitions, restructuring, or operational improvements. Unlike publicly traded companies, which raise capital through stock markets, private equity investments are not traded on public exchanges and are characterized by longer investment horizons, greater control over portfolio companies, and potentially higher returns but also higher risk.

Crowdfunding is a method of raising capital for a project, business, or cause by soliciting small contributions or investments from a large number of individuals or entities, typically through online platforms or social media channels. Crowdfunding platforms connect project creators with potential backers, allowing them to pitch their ideas, set fundraising goals, and offer rewards or incentives to supporters in exchange for their contributions.

Debt funding involves raising capital for a business through borrowing funds from lenders or financial institutions, typically in the form of loans or debt securities. Unlike equity financing, where ownership in the company is exchanged for capital, debt financing requires the borrower to repay the borrowed amount along with interest over a specified period. Debt funding can take various forms, including bank loans, lines of credit, bonds, or convertible debt instruments.

A pie chart with numbers and a red circle

Description automatically generated

Fig 5.16: Pie Chart of Investment Type

Prevalence of Seed Funding: Seed funding emerges as the most common investment type, with 1301 fundings made in this category. Seed funding typically represents the initial capital provided to startups or early-stage companies to support product development, market validation, and initial operations. The high number of seed fundings suggests a vibrant ecosystem of early-stage startups and investors willing to support innovative ideas and entrepreneurial ventures.

Significant Presence of Private Equity: Private equity (PE) investments follow closely behind seed funding, with 1066 fundings recorded in this category. Private equity investments typically involve larger sums of capital provided to more mature companies in exchange for ownership stakes. The presence of a significant number of private equity fundings indicates investor interest in established businesses with proven track records, stable revenue streams, and potential for growth and profitability.

Limited Participation in Crowd Funding, Debt Funding, and Private Funding: Crowd funding, debt funding, and private funding represent a smaller proportion of the total fundings, with only 2, 1, and 1 fundings respectively. Crowd funding involves raising capital from a large number of individuals or investors through online platforms, while debt funding involves raising capital through loans or debt instruments. Private funding typically involves investments from individuals or private entities outside of traditional venture capital or private equity channels. The limited participation in these categories suggests that they may be less common or less preferred avenues for fundraising compared to seed funding and private equity for startups.

# CHAPTER 6: REQUIREMENT SPECIFICATION

The requirements that are necessary in order to carry out the project is described in this chapter. The software and hardware requirements are mentioned and explained.



### SOFTWARE REQUIREMENTS

1. **Jupyter Notebook**

Jupyter notebook is an interactive computing environment that enables users to create and share documents containing live code, equations, visualizations, and explanatory text. It combines the features of a traditional code editor with the flexibility of a narrative-driven document, allowing users to run Python scripts, perform data analysis, and perform machine learning tasks.

1. **Kaggle**

Kaggle is an online network for data science amateurs and experts that provides access to datasets, tournaments, and machine learning and data analysis courses. It promotes collaboration, knowledge exchange, and skill development by creating a community-driven environment. Kaggle is owned by Google and offers a variety of tasks ranging from beginner to intermediate levels, allowing individuals to refine their skills and compete for prizes. To assist users in their learning journey, the site also provides resources such as notebooks, tutorials, and forums.

### HARDWARE REQUIREMENTS

* + - * 1. **Processor**: For data processing, a multi-core CPU with a clock speed of at least 2 GHz is advised.
        2. **RAM**: At least 8 GB of RAM is recommended for running the machine learning algorithms and natural language processing tasks.
        3. **Storage**: Sufficient storage capacity is required to store the Top Funding Startup dataset which includes the data
        4. **Graphics Processing Unit (GPU)**: A GPU is not strictly required, but it can significantly speed up the training of deep learning models, especially for larger datasets.
      1. **LIBRARIES and TOOLS REQUIREMENTS**
         1. **Pandas:**

Python's Pandas module is a widely used open-source tool for handling and analysing data. For interacting with structured data, including tabular, time series, and matrix data, it offers a number of data types and operations. Among the essential characteristics of pandas are:

Data Frame: A two-dimensional data structure that resembles a table and can hold different types of data. It enables for simple data management and analysis and is comparable to a spreadsheet or SQL table.

Series: An unlabeled, one-dimensional array that can hold any kind of data. It is frequently used to represent a data column in a Data Frame.

Data Cleaning: Pandas has a variety of functions for cleaning data, including completing blank fields, eliminating duplicate records, and converting data.

Data exploration is simple with Pandas because of its groupby, pivot tables, and descriptive statistics capabilities.

Data Visualisation: Pandas interacts with well-liked libraries for data visualization, such as Matplotlib and Seaborn, making it simple to produce graphs and charts for data visualization.

Because of its simplicity, adaptability, and capacity for handling big datasets, Pandas is frequently used ML projects. In numerous other fields, including economics, social sciences, and finance, it is also employed.

* + - * 1. **NUMPY:**

In data science, machine learning, scientific computing, and other fields of study that call for quick and effective mathematical operations on massive arrays and matrices, NumPy is a Python library for numerical computations that is widely utilised.

Additionally, a wide range of mathematical operations are supported by NumPy, including linear algebra, Fourier transformations, random number generation, and more. Along with providing a whole ecosystem for scientific computing in Python, NumPy works nicely with other scientific computing libraries like SciPy, Matplotlib, Pandas, and scikit-learn.

Among NumPy's primary characteristics are the following:

1. Effective manipulation and storage of massive arrays and matrices.

2. Mathematical operations on arrays that are quick and vectorized.

3. A wide range of mathematical operations for performing calculations.

4. Python integration with more scientific computing libraries.

5. Support for matrices and arrays with multiple dimensions.

6. Strong array indexing and slicing operations.

Providing a strong and effective array-based computing environment for a variety of numerical applications, NumPy is an essential library for scientific computing in Python.

* + - * 1. **MATPLOTLIB**

It is a popular library used to produce data visualisations including graphs, charts, and plots. Line plots, scatter plots, bar plots, histograms, and other visualisation techniques are all available through Matplotlib. Additionally, a large variety of customization options are supported, giving users control over elements like colours, typefaces, labels, and axis restrictions. It has a big and vibrant community of users and developers who contribute to its ongoing development, making it very extendable.

In conclusion, Matplotlib is a robust and adaptable Python library for building high-quality data visualisations, and it is frequently used for data analysis, scientific research, and other purposes.

* + - * 1. **SEABORN**

Seaborn is a Python data visualization library based on Matplotlib that provides a high-level interface for creating attractive and informative statistical graphics. It is built on top of Matplotlib and tightly integrated with Pandas data structures, making it particularly well-suited for visualizing data from DataFrames. Its key features include:

Statistical Visualization: Seaborn provides a wide range of functions for creating statistical plots such as scatter plots, bar plots, box plots, violin plots, and histograms. These plots are designed to reveal patterns, trends, and relationships in the data, making it easier to understand complex datasets.

Categorical Data Support: Seaborn offers robust support for visualizing categorical data, including functions for creating categorical scatter plots, box plots, and bar plots. It also provides options for grouping data by categorical variables and applying statistical aggregations.

Themes and Contexts: Seaborn allows users to easily customize the overall look and feel of their plots by choosing from different built-in themes and contexts. Themes control the overall aesthetic style of the plots, while contexts control the size and scale of plot elements.

* + - * 1. **SCIPY**

It is a Python library that builds on top of NumPy and provides additional functionality for scientific computing. It offers a wide range of modules for numerical integration, optimization, interpolation, signal processing, linear algebra, statistics, and more.

We are interested in scipy.stats module that includes functions for conducting hypothesis tests, such as t-tests, chi-square tests, Kolmogorov-Smirnov tests, and many others. These functions allow users to test hypotheses about population parameters, compare sample distributions, and assess the significance of observed differences.

# CHAPTER 7: CONCLUSION

Based on the exploratory data analysis (EDA) conducted across various aspects of the startup ecosystem, including funding, industries, locations, and investors, several key conclusions can be drawn:

* There were more than 2000, specifically 2372, new startups funded according to the data
* The Funding startups are highly dispersed as investors are specific about choosing startups
* Paytm and Flipkart were funded most
* Top 10 investments are made through private equity
* Paytm was mostly funded on different days by single investor
* Swiggy was invested by most number of investors
* As per startups 2 and 3 were most frequent combination of investors
* There were 50% relation between funded amount and number of investors per startup
* There were more than 1900 unique investors
* Steadview capital and existing investors invested highest amount.
* Ola was Funded most frequent number of times.
* Consumer internet was the top most choice for all investors.
* Top investors funded ecommerce and consumer internet most in terms of amount.
* Most funding came through private equity
* Crowd funding and dept funding were less preferred by the investors
* Seed funding was less dispersed but private equity accounted for large investments per startup
* Consumer internet sector got most amount of funding and attracted the most investors
* Bangalore had the most average funding
* Starting months of 2015, mid of 2016 and 3rd Quarter of 2017 were funded most
* 2016 had highest amount of investments
* In 2015 there was slight fund generated through crowd funding
* In 2016 there was slight variation in seed funding by amount
* In 2017 there dept funding witnessed for first time but in less fraction

Investment Trends: Seed funding emerges as the most common investment type, followed by private equity investments. This suggests a strong interest from investors in supporting early-stage startups and providing growth capital to more mature companies. Additionally, individual investors like Ratan Tata and institutional investors such as venture capital firms play significant roles in funding startups and driving innovation.

Industry Landscape: The data highlights the dominance of sectors such as consumer internet, technology, and e-commerce in attracting investments, reflecting the growing digital economy and consumer-centric business models. However, there are also emerging sectors like healthcare, logistics, education, and food & beverage that present opportunities for innovation and investment.

Geographical Distribution: Bangalore emerges as the leading startup hub, followed by Mumbai, New Delhi, and Gurgaon. These cities offer conducive environments for startups, including access to talent, infrastructure, funding, and networking opportunities. Emerging startup ecosystems in cities like Pune, Noida, Hyderabad, Chennai, and Ahmedabad indicate the geographical diversification of entrepreneurship across the country.

Investor Diversity: The presence of diverse investors, including angel investors, venture capital firms, private equity investors, and strategic investors, underscores the depth and breadth of the investment landscape. Each investor category brings unique perspectives, expertise, and resources to the table, contributing to the growth and success of startups in different stages of development.

Future Opportunities: Despite the challenges and uncertainties in the startup ecosystem, the data points towards promising opportunities for innovation, growth, and investment in India. By leveraging emerging technologies, addressing market gaps, and fostering collaboration between startups, investors, and ecosystem enablers, India can continue to strengthen its position as a leading destination for startups and entrepreneurship.

In conclusion, the exploratory data analysis provides valuable insights into the current state and future potential of the startup ecosystem in India, highlighting areas of strength, opportunities for growth. By leveraging these insights we can foster a supportive ecosystem.