**A Major Qualifying Report submitted to the faculty of**

**BML MUNJAL UNIVERSITY**

****

**“Partially meets the requirements of the Degree Bachelor of Technology in Computer Science and Engineering”**

**I**

**By:**

**Gagan Singhal 220461**

**Ayushi Singh 220462**

**Ishika Kochhar 220435**

**Saksham Pandla 220382**

**Academic Year – (2022 – 23)**

**New Start-up trends among Indian youths, their profitability and sustainability**

**COURSE –** MATHEMATICS FOR ENGINEERING- II

**COURSE FACULTY –** DR. RANJIB BANERJEE

**Acknowledgment**

We would like to express our sincere gratitude to our teacher, Dr. Ranjib Banerjee, for his invaluable guidance throughout the project. His unwavering support and advice were instrumental in resolving our doubts and challenges, and his motivation was a driving force behind our successful completion of the project.

We would also like to extend our heartfelt appreciation to all the members of our group who worked tirelessly and gave their best to ensure that our project was of the highest quality. Their dedication and hard work were key to our success.

Finally, we would like to express our gratitude to all those who provided their support and encouragement, both during the project and in our personal lives. Your unwavering support was a great source of inspiration, and we could not have completed this project without you.

**Abstract**

In recent years, the start-up culture has been rapidly growing in India, with a significant number of young entrepreneurs entering the market. The report analyzes the current trends in the Indian start-up ecosystem. The study uses R programming and its statistical models to evaluate the financial performance and growth of start-ups in India. It includes comparisons of different start-up performances and investment trends in various cities and industries all over India, highlighting the cities with the most promising start-up ecosystems. It also evaluates the impact of various external factors, such as investment amount, cities, and yearly growth. Finally, the report provides the top 15 start-ups for aspiring entrepreneurs and investors looking to enter the Indian start-up market. Overall, the report finds that analyzing the start-up ecosystem in different cities of India and their investments can provide valuable insights into the profitability and sustainability of start-ups. By using data-driven approaches and understanding the unique characteristics of each city's start-up ecosystem, entrepreneurs can make informed decisions and improve their chances of success.

**Introduction**

India's startup ecosystem is growing rapidly and is primarily fueled by young, motivated entrepreneurs. In recent years, India has seen a surge in the number of startups. These startups have shown great potential for growth, profitability, and sustainability. The government has launched several initiatives such as 'Startup India' , 'Make in India', and 'Digital India' to promote entrepreneurship and create a conducive environment for startups to grow. The government's focus on entrepreneurship, improved access to funding, and a thriving technology sector. As a result, India has become the third-largest startup ecosystem in the world after the United States and China, with over 50,000 startups operating in the country. The Startup ecosystem in India has evolved beyond the traditional sectors such as IT and software services, with startups emerging in a diverse range of industries such as e-commerce, FinTech, healthcare, education, food and agriculture, logistics, and more1.

The report will focus on the start-up growth of India from 2015-2021 and compare the trends based on investment, funding, cities, type of industries, and the number of investments made every year. The report will also examine the factors that have contributed to the growth of these startups and how e-commerce is one of the most significant success stories of the Indian startup ecosystem, with companies like Flipkart and Snapdeal revolutionizing the retail industry in India. Fintech is another sector that has seen significant growth in recent years, with startups like Paytm, PhonePe, and Razorpay transforming the way payments and financial services are offered in the country. In the healthcare industry, startups like Practo, Netmeds, and 1mg are providing innovative solutions to improve healthcare access and affordability. The education sector has also seen the emergence of startups like Byju's, Unacademy, and Vedantu, which are changing the way education is delivered and consumed in India2.

Moreover, the Indian startup ecosystem has seen a rise in the number of startups emerging from tier-2 and tier-3 cities, reflecting the growing entrepreneurial spirit across the country. The government's focus on creating a conducive ecosystem for startups has led to the emergence of startup incubators, accelerators, and co-working spaces across the country, providing the necessary infrastructure and support for startups to grow. These facilities have played a vital role in creating a culture of entrepreneurship in these regions, by providing aspiring entrepreneurs with the necessary resources and support to start and grow their businesses. They have also helped in bridging the gap between smaller cities and metropolitan cities, by providing startups with access to resources and opportunities that were previously available only in big cities. The report will be really beneficial to entrepreneurs, investors, and anyone interested in understanding the dynamics of the Indian startup ecosystem.

<https://blog.mygov.in/editorial/startup-india-what-it-means-for-the-youth/#:~:text=This%20is%20what%20Start%2Dup,grabbing%20a%20job%20after%20college>.

<http://www.isec.ac.in/WP%20514%20-%20Fakih%20Amrin%20Kamaluddin%20-%20Final.pdf>

**Methodology**

The methodology adopted for this report involves a rigorous and comprehensive analysis of the Indian startup ecosystem from 2015 to 2021. To ensure the reliability and representativeness of the data used in the analysis, we collected data from various sources, including secondary research reports, news articles, and public databases such as Kaggle. We carefully pre-processed the data obtained from Kaggle to ensure its accuracy and completeness. We removed any missing values and manually added information that was not provided in the database. This helped us to obtain reliable and accurate data for our analysis.

The report focuses on four key parameters, namely, investment, funding, cities, and type of industries, to analyze the growth of startups in India. For each of these parameters, we collected data on the number of investments made in each year, the total amount of funding received, the cities where startups are located, and the industries in which they operate. To analyze the data and draw meaningful insights, we used R programming language, which is a powerful tool for statistical analysis and data visualization. We conducted in-depth analysis and graphical comparisons on the basis of all the above parameters, using various data visualization techniques, such as bar charts and line graphs.

The report is generated using R Markdown, which is an open-source tool for creating dynamic documents that combine text, code, and graphical output. This ensures that our analysis is adequately represented in the report, and readers can easily reproduce our findings. Overall, the methodology adopted for this report is robust, and reliable, and provides a comprehensive understanding of the Indian startup ecosystem and its growth from 2015 to 2021.

suppressWarnings({  
data <- read\_excel("D:\\Sem 2\\Mathematics for Engineers - II\\Project\\1. Startup Datasets\\Final Project.xlsx")  
is.na(data)  
summary(data)  
})

## Date Startup\_name Vertical   
## Min. :2015-01-02 00:00:00.00 Length:5195 Length:5195   
## 1st Qu.:2016-04-27 00:00:00.00 Class :character Class :character   
## Median :2017-11-28 00:00:00.00 Mode :character Mode :character   
## Mean :2018-07-03 22:48:12.47   
## 3rd Qu.:2020-11-27 00:00:00.00   
## Max. :2021-12-31 00:00:00.00   
##   
## City Investors\_Name Investment\_Type Amount   
## Length:5195 Length:5195 Length:5195 Min. :1.270e+04   
## Class :character Class :character Class :character 1st Qu.:6.000e+05   
## Mode :character Mode :character Mode :character Median :2.300e+06   
## Mean :7.965e+07   
## 3rd Qu.:1.000e+07   
## Max. :1.500e+11   
## NA's :1342

This Summary shows we have 1342 NA values in dataset.

#### Droping NA values from the dataset(data) using drop ‘na’ function

startup<-data %>% drop\_na()  
summary(startup)

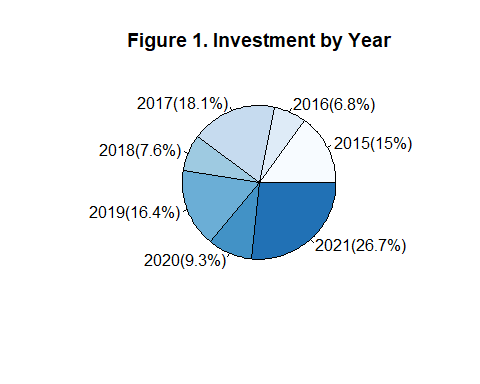
## Date Startup\_name Vertical   
## Min. :2015-01-02 00:00:00.00 Length:3103 Length:3103   
## 1st Qu.:2016-03-07 12:00:00.00 Class :character Class :character   
## Median :2017-08-21 00:00:00.00 Mode :character Mode :character   
## Mean :2018-04-03 22:37:51.60   
## 3rd Qu.:2020-09-18 00:00:00.00   
## Max. :2021-12-31 00:00:00.00   
## City Investors\_Name Investment\_Type Amount   
## Length:3103 Length:3103 Length:3103 Min. :1.270e+04   
## Class :character Class :character Class :character 1st Qu.:5.000e+05   
## Mode :character Mode :character Mode :character Median :2.000e+06   
## Mean :1.856e+07   
## 3rd Qu.:1.000e+07   
## Max. :3.900e+09

# Distribution of amount of investment made in several years

startup$Year <- as.integer(format(as.Date(startup$Date), "%Y"))  
investment\_by\_year <- aggregate(Amount ~ Year, data = startup, FUN = sum)  
max\_investment\_year <- investment\_by\_year$Year[which.max(investment\_by\_year$Amount)]  
investment\_by\_year$Percentage <- investment\_by\_year$Amount / sum(investment\_by\_year$Amount) \* 100  
labels <- paste(investment\_by\_year$Year, "(", round(investment\_by\_year$Percentage, 1), "%)", sep = "");labels

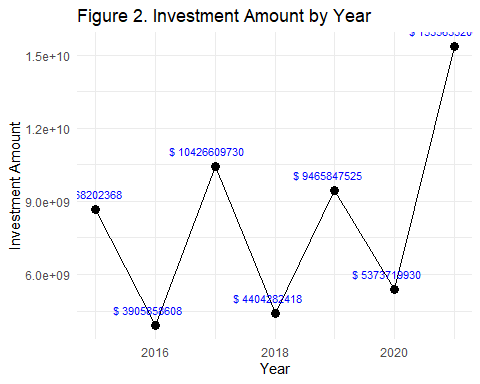
## [1] "2015(15%)" "2016(6.8%)" "2017(18.1%)" "2018(7.6%)" "2019(16.4%)"  
## [6] "2020(9.3%)" "2021(26.7%)"

pie(investment\_by\_year$Amount, labels = labels, main = "Figure 1. Investment by Year",col = brewer.pal(9, "Blues"))



A pie chart showing the investment amounts for different years, and each year will be labeled with its percentage of the total investment. The year with the highest investment amount will be highlighted in the chart.

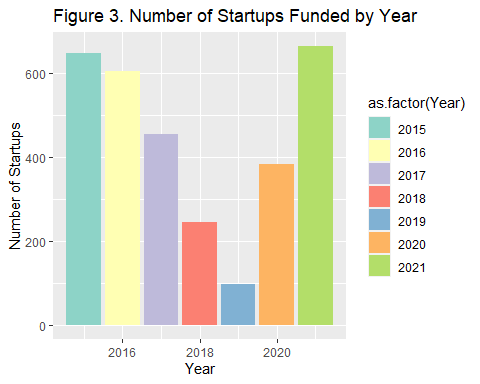
ggplot(investment\_by\_year, aes(x = Year, y = Amount)) +  
 geom\_line() +  
 labs(x = "Year", y = "Investment Amount", title = "Figure 2. Investment Amount by Year") +  
 geom\_text(aes(label = paste("$", Amount)), vjust = -1.2, hjust = +0.6, size = 3, color = "blue") +  
 geom\_point(size = 3, color = "black") +  
 theme\_minimal()



A line graph that shows the investment amounts for different years. The year with the highest investment amount will be highlighted in the graph, and the investment amounts will be displayed as labels on the graph.

# Distribution of number of startups over several years

startup$Year <- as.integer(substr(startup$Date, 1, 4))  
startups\_by\_year <- aggregate(Startup\_name ~ Year, data = startup, FUN = length)  
max\_funded\_year <- startups\_by\_year$Year[which.max(startups\_by\_year$Startup\_name)]  
ggplot(startups\_by\_year, aes(x = Year, y = Startup\_name, fill = as.factor(Year))) +  
 geom\_bar(stat = "identity") +  
 labs(x = "Year", y = "Number of Startups", title = "Figure 3. Number of Startups Funded by Year") +  
 scale\_fill\_brewer(palette = "Set3")

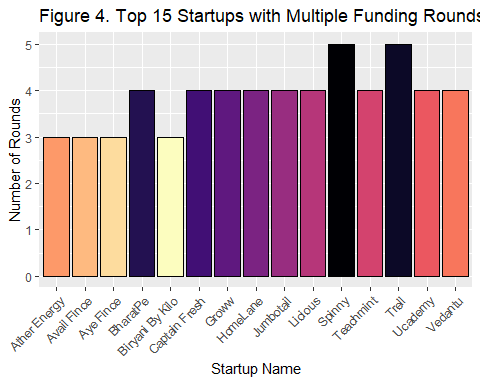


# Top 15 Startups that got multiple rounds of funding

startup\_rounds <- aggregate(Investment\_Type ~ Startup\_name, data = startup, FUN = function(x) length(unique(x)))  
startup\_multiple\_rounds <- subset(startup\_rounds, Investment\_Type > 1);  
top\_15\_startups <- head(startup\_multiple\_rounds[order(startup\_multiple\_rounds$Investment\_Type, decreasing = TRUE), ], 15);top\_15\_startups

## Startup\_name Investment\_Type  
## 1979 Spinny 5  
## 2171 Trell 5  
## 211 BharatPe 4  
## 324 Captain Fresh 4  
## 901 Groww 4  
## 980 HomeLane 4  
## 1140 Jumbotail 4  
## 1254 Licious 4  
## 2089 Teachmint 4  
## 2212 Ucademy 4  
## 2262 Vedantu 4  
## 143 Ather Energy 3  
## 156 Avail Fince 3  
## 173 Aye Fince 3  
## 237 Biryani By Kilo 3

ggplot(top\_15\_startups, aes(x = Startup\_name, y = Investment\_Type)) +  
 geom\_bar(stat = "identity", fill = magma(15), color = "black") +  
 labs(x = "Startup Name", y = "Number of Rounds", title = "Figure 4. Top 15 Startups with Multiple Funding Rounds") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



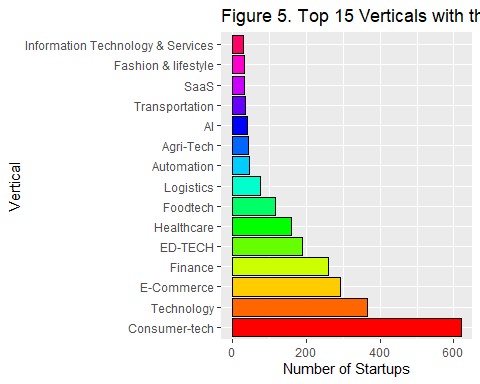
A bar chart that displays the number of funding rounds for each startup. Only the startups with multiple rounds will be shown in the chart. The x-axis represents the startup names, and the y-axis represents the number of rounds. The chart will help identify the startups that have received multiple rounds of funding.

# Top 15 vertical with highest number of startups

top\_verticals <- head(sort(table(startup$Vertical), decreasing = TRUE), 15);top\_verticals

##   
## Consumer-tech Technology   
## 620 365   
## E-Commerce Finance   
## 293 259   
## ED-TECH Healthcare   
## 190 158   
## Foodtech Logistics   
## 115 76   
## Automation Agri-Tech   
## 44 41   
## AI Transportation   
## 39 35   
## SaaS Fashion & lifestyle   
## 32 31   
## Information Technology & Services   
## 30

ggplot(data.frame(vertical = names(top\_verticals), count = as.numeric(top\_verticals)),   
 aes(x = count, y = reorder(vertical, -count))) +  
 geom\_bar(stat = "identity", fill = rainbow(15), color = "black") +  
 labs(x = "Number of Startups", y = "Vertical", title = "Figure 5. Top 15 Verticals with the Highest Number of Startups") +  
 theme(axis.text.y = element\_text(hjust = +1.0))



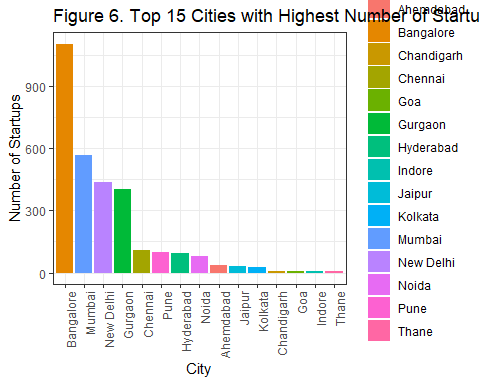
The x-axis represents the number of startups, and the y-axis represents the verticals. The verticals are reordered based on the count of startups in descending order.

# Cities with highest number of startups

top\_cities <- head(sort(table(startup$City), decreasing = TRUE), 15)  
top\_startup <- subset(startup, startup$City %in% names(top\_cities));top\_cities

##   
## Bangalore Mumbai New Delhi Gurgaon Chennai Pune Hyderabad   
## 1103 566 434 401 110 97 94   
## Noida Ahemdabad Jaipur Kolkata Chandigarh Goa Indore   
## 81 38 33 26 9 9 8   
## Thane   
## 8

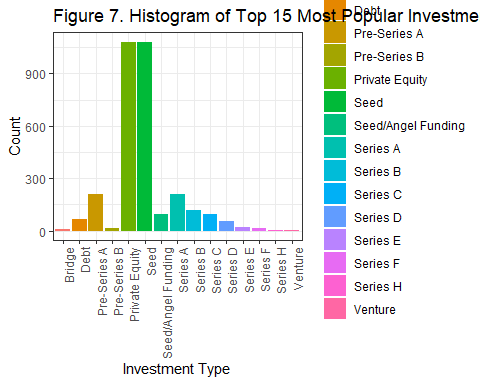
ggplot(top\_startup, aes(x = reorder(City, -table(City)[City]), fill = City)) +  
 geom\_bar() +  
 labs(x = "City", y = "Number of Startups", title = "Figure 6. Top 15 Cities with Highest Number of Startups") +  
 scale\_fill\_discrete(name = "City") +  
 theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



The x-axis represents the cities, and the y-axis represents the number of startups. The bars are filled with colors based on the city variable to differentiate the bars. The bar plot displays the number of startups in each of the top 15 cities, allowing for easy comparison and identification of the city with the highest number of startups. The x-axis labels are angled for better readability when there are many cities.

# Most Popular Investment type

top\_investment\_types <- head(sort(table(startup$Investment\_Type), decreasing = TRUE), 15)  
top\_startup <- subset(startup, startup$Investment\_Type %in% names(top\_investment\_types));  
ggplot(top\_startup, aes(x = Investment\_Type, fill = Investment\_Type)) +  
 geom\_bar() +  
 labs(x = "Investment Type", y = "Count", title = "Figure 7. Histogram of Top 15 Most Popular Investment Types") +  
 scale\_fill\_discrete(name = "Investment Type") +  
 theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



The x-axis represents the investment types, and the y-axis represents the count of startups. The bars are filled with colors based on the investment type variable to differentiate the bars. The histogram displays the distribution of startups across the top 15 most popular investment types, allowing for easy comparison and identification of the investment type that is most commonly used. The x-axis labels are angled for better readability when there are many investment types.

## Seed type investment in different cities

data\_seed <- startup[startup$Investment\_Type == "Seed", ]  
a<-aggregate(Startup\_name ~ City, data\_seed, length);a

## City Startup\_name  
## 1 Ahemdabad 14  
## 2 Amritsar 1  
## 3 Andhra Pradesh 1  
## 4 Bangalore 360  
## 5 Bhilwara 1  
## 6 Bhopal 1  
## 7 Bihar 1  
## 8 Chandigarh 5  
## 9 Chennai 35  
## 10 Faridabad 1  
## 11 Goa 1  
## 12 Gujarat 1  
## 13 Gurgaon 117  
## 14 Guwahati 1  
## 15 Gwalior 1  
## 16 Haryana 1  
## 17 Hyderabad 52  
## 18 Indore 6  
## 19 Jaipur 8  
## 20 Jodhpur 3  
## 21 Kanpur 2  
## 22 Kerala 3  
## 23 Kochi 1  
## 24 Kolkata 1  
## 25 Lucknow 1  
## 26 Maharashtra 3  
## 27 Mangalore 1  
## 28 Mumbai 189  
## 29 Mussoorie 1  
## 30 New Delhi 191  
## 31 Noida 28  
## 32 Orissa 1  
## 33 Pune 34  
## 34 Punjab 1  
## 35 Seoul 1  
## 36 Silvassa 1  
## 37 Tamil Nadu 3  
## 38 Telangana 1  
## 39 Thane 1  
## 40 Thiruvananthapuram 2  
## 41 Uttar Pradesh 1  
## 42 Vadodara 3  
## 43 Varanasi 1  
## 44 West Bengal 1

summary(a)

## City Startup\_name   
## Length:44 Min. : 1.00   
## Class :character 1st Qu.: 1.00   
## Mode :character Median : 1.00   
## Mean : 24.61   
## 3rd Qu.: 5.25   
## Max. :360.00

## Private Equity type investment in different cities

data\_seed <- startup[startup$Investment\_Type == "Private Equity", ]  
a<-aggregate(Startup\_name ~ City, data\_seed, length);a

## City Startup\_name  
## 1 Ahemdabad 12  
## 2 Bangalore 391  
## 3 Bhopal 1  
## 4 Bhubaneswar 1  
## 5 Bihar 1  
## 6 Chandigarh 3  
## 7 Chennai 36  
## 8 Faridabad 1  
## 9 Goa 2  
## 10 Gurgaon 156  
## 11 Hyderabad 41  
## 12 India 1  
## 13 Jaipur 22  
## 14 Karnataka 3  
## 15 Kochi 1  
## 16 Kolkata 7  
## 17 Lucknow 1  
## 18 Maharashtra 3  
## 19 Mumbai 221  
## 20 New Delhi 139  
## 21 Noida 31  
## 22 Singapore 1  
## 23 Surat 1  
## 24 Tamil Nadu 4  
## 25 Uttar Pradesh 1  
## 26 Vadodara 1

summary(a)

## City Startup\_name   
## Length:26 Min. : 1.00   
## Class :character 1st Qu.: 1.00   
## Mode :character Median : 3.00   
## Mean : 41.62   
## 3rd Qu.: 28.75   
## Max. :391.00

## Pre-Series A type of investment in different cities

data\_seed <- startup[startup$Investment\_Type == "Pre-Series A", ]  
a<-aggregate(Startup\_name ~ City, data\_seed, length);a

## City Startup\_name  
## 1 Ahemdabad 4  
## 2 Bangalore 71  
## 3 Chennai 6  
## 4 Goa 2  
## 5 Gujarat 1  
## 6 Gurgaon 24  
## 7 Haryana 1  
## 8 Indore 1  
## 9 Kanpur 1  
## 10 Kochi 1  
## 11 Kolkata 2  
## 12 Lucknow 1  
## 13 Mumbai 35  
## 14 New Delhi 32  
## 15 Noida 7  
## 16 Pune 18  
## 17 Punjab 1  
## 18 Thane 1

summary(a)

## City Startup\_name   
## Length:18 Min. : 1.00   
## Class :character 1st Qu.: 1.00   
## Mode :character Median : 2.00   
## Mean :11.61   
## 3rd Qu.:15.25   
## Max. :71.00

## Series A type of investment in different cities

data\_seed <- startup[startup$Investment\_Type == "Series A", ]  
a<-aggregate(Startup\_name ~ City, data\_seed, length);a

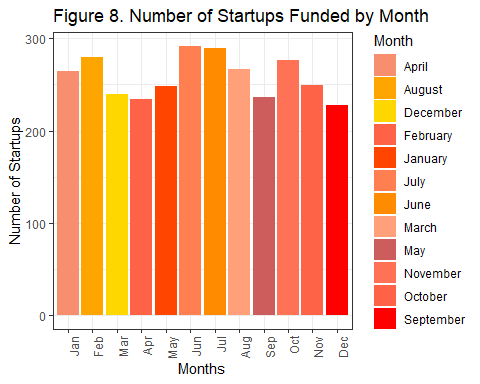
## City Startup\_name  
## 1 Bangalore 88  
## 2 Chandigarh 1  
## 3 Chennai 12  
## 4 Faridabad 1  
## 5 Goa 2  
## 6 Gurgaon 27  
## 7 Haryana 1  
## 8 Karnataka 1  
## 9 Mumbai 31  
## 10 New Delhi 18  
## 11 Noida 3  
## 12 Pune 20  
## 13 Singapore 1  
## 14 Surat 1  
## 15 Thane 4  
## 16 Uttar Pradesh 1

summary(a)

## City Startup\_name   
## Length:16 Min. : 1.00   
## Class :character 1st Qu.: 1.00   
## Mode :character Median : 2.50   
## Mean :13.25   
## 3rd Qu.:18.50   
## Max. :88.00

# Distribution of number of startups in different months

startup$Date <- as.Date(startup$Date)  
startup$Month <- format(startup$Date, "%B")  
startup\_count <- table(startup$Month)  
highest\_month <- names(startup\_count)[which.max(startup\_count)]  
ggplot(startup, aes(x = Month, fill = Month)) +  
 geom\_bar() +  
 labs(x = "Months", y = "Number of Startups", title = "Figure 8. Number of Startups Funded by Month") +  
 scale\_x\_discrete(labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")) +  
 scale\_fill\_manual(values = c("#F78F6F", "#FFA500", "#FFD700", "#FF6347", "#FF4500", "#FF7F50",  
 "#FF8C00", "#FFA07A", "#CD5C5C", "#FF7256", "#FF6347", "#FF0000")) +  
 theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



The x-axis represents the months, and the y-axis represents the number of startups funded. The bars correspond to the count of startups for each month. The bar chart displays the number of startups funded by month, allowing you to identify the month with the highest number of startups funded. The x-axis labels show the corresponding month names.

# Calculate summary statistics

num\_startups <- nrow(startup)  
num\_investors <- startup %>% distinct(Investors\_Name) %>% nrow()  
total\_investment <- sum(startup$Amount)  
num\_cities <- startup %>% distinct(City) %>% nrow()

# Print summary information

cat("Summary of the Dataset:\n")

## Summary of the Dataset:

cat("Number of Startups:", num\_startups, "\n")

## Number of Startups: 3103

cat("Number of Investors:", num\_investors, "\n")

## Number of Investors: 2466

cat("Total Investment:", total\_investment, "\n")

## Total Investment: 57600852579

cat("Number of Cities:", num\_cities, "\n")

## Number of Cities: 53

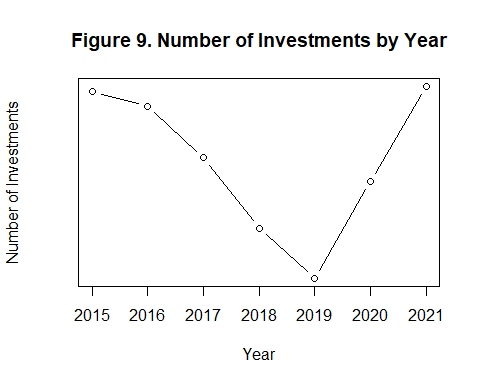
# Calculate average investment amount by city

avg\_investment <- aggregate(Amount ~ City, data = startup, FUN = mean)  
print(avg\_investment)

## City Amount  
## 1 Ahemdabad 16888421.1  
## 2 Amritsar 300000.0  
## 3 Andhra Pradesh 500000.0  
## 4 Bangalore 25532063.6  
## 5 Bangkok 9000000.0  
## 6 Bhilwara 1500000.0  
## 7 Bhopal 265332.5  
## 8 Bhubaneswar 23000000.0  
## 9 Bihar 3164500.0  
## 10 Chandigarh 5522222.2  
## 11 Chennai 8730281.8  
## 12 Faridabad 65750000.0  
## 13 Gandhinagar 2000000.0  
## 14 Goa 23427555.6  
## 15 Gujarat 880000.0  
## 16 Gurgaon 18542608.2  
## 17 Guwahati 400000.0  
## 18 Gwalior 500000.0  
## 19 Haryana 12200000.0  
## 20 Hyderabad 6508904.3  
## 21 India 42000000.0  
## 22 Indore 834000.0  
## 23 Jaipur 10420310.0  
## 24 Jodhpur 1320000.0  
## 25 Kanpur 1700000.0  
## 26 Karnataka 31800000.0  
## 27 Kerala 68000.0  
## 28 Kochi 4700000.0  
## 29 Kolkata 30265715.4  
## 30 Lucknow 1033333.3  
## 31 Maharashtra 4702800.7  
## 32 Mangalore 900000.0  
## 33 MP 18000000.0  
## 34 Mumbai 13840927.4  
## 35 Mussoorie 350000.0  
## 36 New Delhi 9389904.3  
## 37 Noida 28823814.8  
## 38 Orissa 1300000.0  
## 39 Pune 32216773.2  
## 40 Punjab 3200000.0  
## 41 Seoul 1400000.0  
## 42 Silvassa 300000.0  
## 43 Singapore 4600000.0  
## 44 Surat 20343333.3  
## 45 Tamil Nadu 7182142.9  
## 46 Telangana 300000.0  
## 47 Thane 10062500.0  
## 48 Thiruvananthapuram 1050000.0  
## 49 Udaipur 500000.0  
## 50 Uttar Pradesh 31775000.0  
## 51 Vadodara 4924800.0  
## 52 Varanasi 52000.0  
## 53 West Bengal 8000000.0

# How does the number of investments vary by year?

startup$Date <- as.Date(startup$Date)  
startup$Year <- format(startup$Date, "%Y")  
investment\_count <- table(startup$Year)  
plot(as.numeric(names(investment\_count)), investment\_count, type = "b",   
 xlab = "Year", ylab = "Number of Investments",  
 main = "Figure 9. Number of Investments by Year")



A line plot with points that shows number of investments by year . The x-axis represents the year, and the y-axis represents the number of investments.

# Are there any outliers or anomalies in the investment amounts?

q1 <- quantile(startup$Amount, 0.25)  
q3 <- quantile(startup$Amount, 0.75)  
iqr <- q3 - q1  
upper <- q3 + 1.5 \* iqr  
lower <- q1 - 1.5 \* iqr  
outliers <- startup$Amount[startup$Amount > upper | startup$Amount < lower]  
  
cat("Number of outliers in investment amounts:", length(outliers), "\n")

## Number of outliers in investment amounts: 420

cat("Outliers in investment amounts:\n")

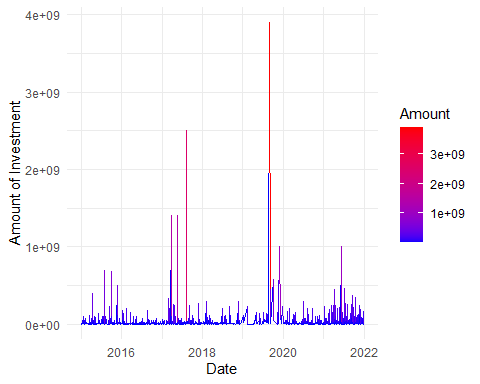
## Outliers in investment amounts:

print(outliers)

## [1] 50000000 100000000 56000000 31000000 50000000 26000000  
## [7] 60000000 80000000 40000000 30000000 110000000 50000000  
## [13] 25000000 30000000 75000000 150000000 25000000 400000000  
## [19] 50000000 25000000 35000000 40000000 50000000 28000000  
## [25] 100000000 85000000 35000000 137000000 50000000 40000000  
## [31] 50000000 60000000 30000000 100000000 30000000 100000000  
## [37] 30000000 25000000 60000000 100000000 700000000 50000000  
## [43] 500000000 100000000 30000000 90000000 25000000 30000000  
## [49] 63000000 50000000 25000000 40000000 45000000 40000000  
## [55] 60000000 225000000 75000000 37500000 30000000 680000000  
## [61] 100000000 36000000 25000000 50000000 34000000 500000000  
## [67] 25000000 52000000 27000000 120000000 30000000 25000000  
## [73] 70000000 30000000 180000000 100000000 145000000 35000000  
## [79] 35000000 30000000 200000000 75000000 150000000 25000000  
## [85] 100000000 30000000 50000000 60000000 100000000 25000000  
## [91] 32000000 25000000 82000000 30000000 25000000 62000000  
## [97] 175000000 37000000 40000000 60000000 50000000 31300000  
## [103] 51000000 50000000 25000000 30000000 27000000 55000000  
## [109] 75000000 35000000 57000000 35000000 55000000 55000000  
## [115] 38700000 70000000 330000000 200000000 1400000000 56000000  
## [121] 100000000 25000000 40000000 260000000 250000000 73700000  
## [127] 104500000 1400000000 30000000 80000000 50000000 50000000  
## [133] 31000000 71000000 35000000 25000000 2500000000 36000000  
## [139] 80000000 25000000 34000000 30000000 26000000 250000000  
## [145] 30000000 44000000 50000000 110000000 38700000 77000000  
## [151] 32000000 25000000 26000000 200000000 270000000 60000000  
## [157] 27000000 27000000 36230000 25000000 47000000 50000000  
## [163] 300000000 200000000 100000000 100000000 50000000 87000000  
## [169] 50000000 115000000 38400000 82000000 62000000 55000000  
## [175] 200000000 27700000 29000000 49700000 100000000 45000000  
## [181] 120000000 100000000 35000000 66000000 35000000 225000000  
## [187] 157200000 100000000 32000000 30000000 27799000 50000000  
## [193] 300000000 31000000 29500000 42000000 50000000 110000000  
## [199] 226000000 45000000 26000000 150000000 50000000 51000000  
## [205] 140000000 51000000 75000000 26000000 26000000 52000000  
## [211] 38080000 150000000 60000000 140000000 38080000 125000000  
## [217] 51000000 37000000 110000000 70000000 3900000000 450000000  
## [223] 45000000 585000000 150000000 26000000 135000000 220000000  
## [229] 283000000 1000000000 70000000 50000000 30000000 50000000  
## [235] 231000000 35000000 200000000 30000000 210000000 60000000  
## [241] 105000000 30000000 97500000 90000000 125000000 113000000  
## [247] 110000000 25000000 40000000 296000000 32000000 30000000  
## [253] 50000000 30000000 43700000 43000000 55000000 27000000  
## [259] 60000000 44000000 26500000 46761000 100000000 56000000  
## [265] 28000000 170000000 90000000 30000000 40000000 54000000  
## [271] 35000000 42500000 28000000 28000000 50000000 30000000  
## [277] 25000000 121000000 150000000 100000000 55000000 25000000  
## [283] 35000000 25000000 200000000 25000000 100000000 35000000  
## [289] 75000000 45000000 40000000 80000000 40000000 30000000  
## [295] 81000000 100000000 30000000 100000000 75000000 100000000  
## [301] 30000000 1000000000 40000000 120000000 27000000 460000000  
## [307] 195000000 75000000 45000000 50000000 30000000 51000000  
## [313] 25000000 160000000 53000000 215000000 350000000 83000000  
## [319] 26000000 65000000 35000000 140000000 30000000 30000000  
## [325] 50000000 43000000 75000000 270000000 250000000 50000000  
## [331] 38000000 70000000 200000000 45000000 26000000 42000000  
## [337] 65000000 36000000 67000000 150000000 108000000 75000000  
## [343] 100000000 45000000 35000000 192000000 65000000 41000000  
## [349] 100000000 40000000 225000000 25000000 37000000 370000000  
## [355] 44000000 125000000 48000000 90000000 100000000 35000000  
## [361] 75000000 30000000 30000000 450000000 31000000 50000000  
## [367] 100000000 50000000 75000000 76000000 55000000 26000000  
## [373] 30000000 40000000 28000000 260000000 140000000 175000000  
## [379] 30000000 125000000 35000000 90000000 50000000 250000000  
## [385] 30000000 75000000 250000000 78000000 100000000 115000000  
## [391] 40000000 255000000 50000000 150000000 30000000 135000000  
## [397] 25000000 248000000 220000000 35000000 40000000 70000000  
## [403] 100000000 30000000 32000000 150000000 32000000 150000000  
## [409] 96000000 70000000 32000000 60000000 266000000 100000000  
## [415] 85000000 40000000 30000000 53000000 86000000 111000000

# A colorful line graph for the amount of investment made over time

ggplot(data = startup, aes(x = Date, y = Amount, color = Amount)) +  
 geom\_line() +  
 scale\_color\_gradient(low = "blue", high = "red") +  
 labs(x = "Date", y = "Amount of Investment") +  
 theme\_minimal()

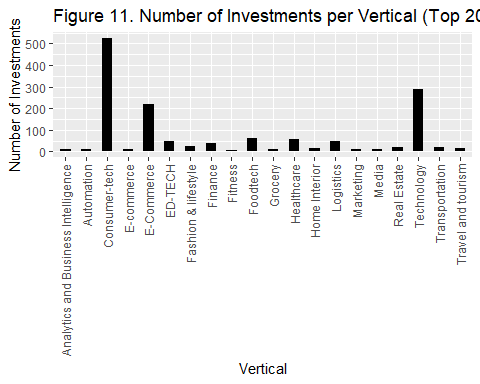
 Figure 10.

# How many investments were made in each vertical in a specific time period?

startup$Date <- as.Date(startup$Date)

### From year 2015-2017:

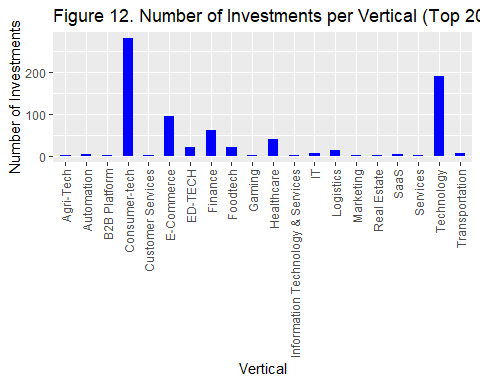
start\_date <- as.Date("2015-01-01")  
end\_date <- as.Date("2017-12-31")  
investments\_period <- subset(startup, Date >= start\_date & Date <= end\_date)  
investments\_count <- table(investments\_period$Vertical)  
sorted\_investments <- sort(investments\_count, decreasing = TRUE)  
top\_20\_verticals <- names(sorted\_investments)[1:20]  
investments\_top\_20 <- subset(investments\_period, Vertical %in% top\_20\_verticals)  
investments\_count\_top\_20 <- table(investments\_top\_20$Vertical)  
ggplot(data = data.frame(Vertical = names(investments\_count\_top\_20), Count = as.vector(investments\_count\_top\_20)), aes(x = Vertical, y = Count)) +  
 geom\_bar(stat = "identity", fill = "black", width = 0.5) +  
 labs(x = "Vertical", y = "Number of Investments", title = "Figure 11. Number of Investments per Vertical (Top 20)") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))



A bar chart where the x-axis represents the verticals and the y-axis represents the number of investments.

### From year 2017-2019:

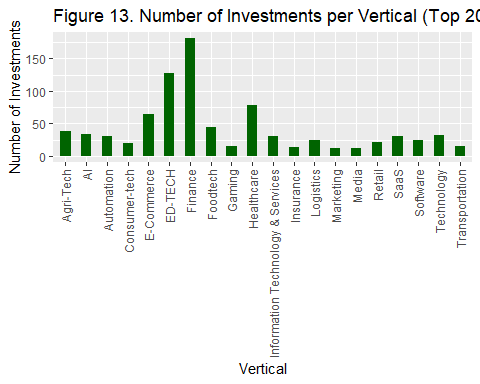
start\_date <- as.Date("2017-01-01")  
end\_date <- as.Date("2019-12-31")  
investments\_period <- subset(startup, Date >= start\_date & Date <= end\_date)  
investments\_count <- table(investments\_period$Vertical)  
sorted\_investments <- sort(investments\_count, decreasing = TRUE)  
top\_20\_verticals <- names(sorted\_investments)[1:20]  
investments\_top\_20 <- subset(investments\_period, Vertical %in% top\_20\_verticals)  
investments\_count\_top\_20 <- table(investments\_top\_20$Vertical)  
ggplot(data = data.frame(Vertical = names(investments\_count\_top\_20), Count = as.vector(investments\_count\_top\_20)), aes(x = Vertical, y = Count)) +  
 geom\_bar(stat = "identity", fill = "blue", width = 0.5) +  
 labs(x = "Vertical", y = "Number of Investments", title = "Figure 12. Number of Investments per Vertical (Top 20)") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))



A bar chart where the x-axis represents the verticals and the y-axis represents the number of investments.

### From year 2019-2021:

start\_date <- as.Date("2019-01-01")  
end\_date <- as.Date("2021-12-31")  
investments\_period <- subset(startup, Date >= start\_date & Date <= end\_date)  
investments\_count <- table(investments\_period$Vertical)  
sorted\_investments <- sort(investments\_count, decreasing = TRUE)  
top\_20\_verticals <- names(sorted\_investments)[1:20]  
investments\_top\_20 <- subset(investments\_period, Vertical %in% top\_20\_verticals)  
investments\_count\_top\_20 <- table(investments\_top\_20$Vertical)  
ggplot(data = data.frame(Vertical = names(investments\_count\_top\_20), Count = as.vector(investments\_count\_top\_20)), aes(x = Vertical, y = Count)) +  
 geom\_bar(stat = "identity", fill = "darkgreen", width = 0.5) +  
 labs(x = "Vertical", y = "Number of Investments", title = "Figure 13. Number of Investments per Vertical (Top 20)") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))



A bar chart where the x-axis represents the verticals and the y-axis represents the number of investments.

**Conclusion**

The Indian startup ecosystem has expanded significantly since 2015, yet there are still issues that need to be resolved. A competitive environment must be negotiated by new businesses, and they must be ready to come up with creative fixes to fill in gaps in the market.

The outlook for new companies in India, however, is positive given continuing government support and improved capital access.

As per our analysis, we concluded that the year 2021 has the highest amount of investment made and we have represented it using a pie chart (figure 1). Also, we’ve represented the highest investment made through a line graph (figure 2). Most startups are funded in the year 2021 followed by years 2015,2016,2017,2020,2018 and 2019 in the queue (figure 3). But, Spinny and Trell startups got the highest multiple rounds of funding but there are the following 15 top startups that got multiple rounds of funding – Ather Energy, Avail Finance, Aye Finance, Bharatpe, Biryani by Kilo, Captain Fresh, Groww, HomeLane, Jumpotail, Licious, Spinny, Teachmint, Trell, Unacademy, Vedantu. (figure 4)

Verticals having the highest number of startups are Consumer tech, Technology, and E-commerce (figure 5) followed by other startups in our data. Bangalore tops the list when it comes to having the highest number of startups followed by Mumbai, New Delhi, Gurgaon, and so on (figure 6). By comparing the investment types we understand that Private Equity and Seed are the most popular investment types in India and venture is the least popular investment type ( figure 7).

Bangalore has the highest number of startups in the Seed investment type and there are a total of 44 cities that have startups in the Seed investment type Bangalore also has the highest number of startups in the Private Equity investment type but there is a total of 26 cities that have startups in Private Equity type. Most of the startups are funded in February, June, and July (figure 8).

The summary of the dataset that we used to prepare our report depicts that there are **3103 startups** established in India from 2015 to 2021. There are a total of **2466 Investors** and **Rs 57600852579 Investment** has been done in **53 cities**.

The line graph used to represent the Number of Investments by Year depicts that year 2019 has the least number of Investments (figure 9). The line graph for the amount of investment made over time depicts the biggest investments were achieved at the end of 2019 followed by some smaller ones in 2017 (figure 10).

There were 420 outliers found in investment amounts in the dataset. Consumer tech has the highest number of Investments made from year 2015 to 2017 which is almost **500+** as per the top 20 verticals in India (figure 11). Consumer-tech has the highest Investments as per the Verticals which is almost **300+** from the year 2017 to 2019 (figure 12), Finance has the highest and Marketing has the least number of Investments as per the Verticals taken from the year 2019 to 2021 (figure 13).

The startup environment is continuously changing, and new trends emerge all the time. As an investor, it's critical to stay current on these trends in order to make informed judgments about where to put your money.

**REFERENCES:-**

1. <http://www.isec.ac.in/WP%20514%20-%20Fakih%20Amrin%20Kamaluddin%20-%20Final.pdf>
2. <https://blog.mygov.in/editorial/startup-india-what-it-means-for-the-youth/#:~:text=This%20is%20what%20Start%2Dup,grabbing%20a%20job%20after%20college>.
3. <https://startuptalky.com/top-indian-startups-revenue/>
4. <https://trak.in/india-startup-funding-investment-2015/amp/>
5. <https://www.statista.com/statistics/1155602/india-start-up-recognized-businesses/>
6. <https://inc42.com/reports/the-state-of-indian-startup-ecosystem-report-2022/>
7. <https://www.kaggle.com/datasets>

Our Dataset can be accessed through <https://docs.google.com/spreadsheets/d/1hTpopTd5q7YxlraO8NBs05ORvVJyuqUR/edit#gid=959743127>

