Maths

by Raj Kumar Shrivas

Submission date: 14-May-2023 06:52PM (UTC+0530)
Submission ID: 2092672731
File name: Copy_of_Start-up_funding_in_India_1.pdf (1.07M)
Word count: 2335
Character count: 12676

START-UP FUNDING IN INDIA

By

1. BHAWNA (220662)
2. BHUMIKA(220661)
3. ROHIT YADAV (220336)
4. RAJ KUMAR SHRIVAS (220320)

BML MUNJAL UNIVERSITY, Gurugram

School of Engineering and Technology (SOET)

Table Of Content :

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all individuals and organizations who have contributed to the successful completion of the project.

First and foremost, we extend our heartfelt appreciation to our project supervisor Dr. Ranjib Banerjee, for their guidance, support, and invaluable expertise throughout the project. Their encouragement and insightful feedback have been instrumental in shaping our work.We are grateful to those support staff members who provided the data and information that were essential and necessary for us to complete this report successfully. Without their assistance, we could not have done so.

We would like to thank BML Munjal University for approving our proposal within the scope of our expertise. We would like to extend our thanks to the team members who worked tirelessly, dedicating their time, skills, and efforts to make this project a reality. Each team member played a crucial role in various aspects, including analyzing, development, and documentation.

Last but not least, we would like to express our appreciation to our families and friends for their unwavering support and understanding throughout the project journey.

ABSTRACT

The abstract describes a study of the Indian startup ecosystem that aims to investigate the relationships between various factors and startup success. The dataset consists of 28 Indian startups, with variables including funding,industry growth, founder education, and success. The study uses statistical tools such as regression analysis to examine the relationships between these variables and determine their impact on the success of a startup.

The results of the analysis indicate that funding is a significant predictor of startup success, with higher levels of funding being associated with greater success. Moreover, younger founders tend to be associated with greater success, while higher levels of education are associated with lower rates of failure. These findings are statistically significant and have implications for entrepreneurs, investors, and policymakers.

The report's findings are based on rigorous statistical analysis of the dataset, making them reliable and informative. The results can be used to inform business strategies, investment decisions, and government policies aimed at supporting the growth of startups in India. Furthermore, the study provides a framework for future research on the Indian startup ecosystem, particularly in exploring the relationships between other factors and startup success.

Overall, the study provides valuable insights into the factors that contribute to the success or failure of startups in India. The use of statistical tools and techniques makes the findings robust and informative, and the results can inform the actions of various stakeholders in the Indian startup ecosystem.

INTRODUCTION

The startup ecosystem in India has experienced a remarkable growth spurt in recent years. This growth has been fueled by a growing number of entrepreneurs and investors who have become increasingly interested in this sector. With more and more startups being established in India, it is essential to understand what factors contribute to the success or failure of these companies. In this report, we have analyzed a dataset of 28 Indian startups to investigate the relationships between several variables, including funding, industry growth, founder education, and success.

Our analysis aims to provide insights into the key factors that drive the success of startups in India. By examining the data, we can identify patterns and trends that help us understand the factors that influence the success of a startup. We employ various statistical tools and techniques to explore the relationships between these variables and determine their impact on the success of a startup.

The findings of our analysis can be used to guide entrepreneurs and investors in the Indian startup ecosystem. By understanding the factors that contribute to success, entrepreneurs can make informed decisions about their business strategies, while investors can use this knowledge to make sound investment decisions.

Overall, our report offers valuable insights into the Indian startup ecosystem and can serve as a valuable resource for anyone looking to understand the factors that contribute to the success or failure of startups in India.

Materials and methods

• Hypothesis Testing

Hypothesis testing is a statistical method used to evaluate the validity of a hypothesis or claim about a population parameter based on sample data. In hypothesis testing, a researcher starts by formulating a null hypothesis (H0) and an alternative hypothesis (Ha). The null hypothesis represents the status quo or the assumption that there is no significant difference or relationship between the variables, while the alternative hypothesis represents the hypothesis that the researcher wants to test, such as the hypothesis that there is a significant difference or relationship between the variables.

The hypothesis testing procedure involves the following steps:

1. Identify the alternative hypothesis (Ha) and the null hypothesis (H0).
2. Establish the test's level of significance (), which denotes the likelihood that the null hypothesis will be rejected when it is true.
3. Based on the study question and the type of data, select an appropriate statistical test.
4. Gather sample data, compute the test statistic, and then determine the p-value that goes with it .
5. Make a decision regarding whether to reject or not to reject the null hypothesis by comparing the p-value with the level of significance ().
6. Interpret the findings and formulate judgements regarding the research subject.

There are two cases in hypothesis testing:

1. One-tailed test: This test has an alternative hypothesis that indicates the direction of the difference or relationship between the variables, making it a directional test. A one-tailed test, for instance, could be used to determine whether the mean score of a treatment group is higher than the mean score of a control group.

2. Two-tailed test: This is a test in which the alternative hypothesis is non-directional, meaning it simply states that there is a difference or relationship between the variables. For example, a two-tailed test might test the hypothesis that there is a difference in mean scores between two groups.

Hypothesis can be tested using the tests shown in the diagram below

Hypothesis Test		Test Statistic
Z test	One sample Z test	Z - Statistics
	Two sample Z test	
	Two Proportions Z test	
T-tests	One sample T test	T - Statistics
	Two Sample T test	
	Paired T test	
Chi Square	Chi Square Fit test	Chi Square Statistics
	Chi Square test of independence	
F test	One Way ANOVA	F-Statistics
	Poc Host test	
	Factorial Analysis of Variance	
	Repeated measure of variance	
	Analysis of covariance	

Table 1: Hypothesis Tests

We have used the Chi square test in our project.

- **Chi square test**

A statistical approach for determining the difference between observed and predicted data is the Chi-Square test. This test can also be performed to see if it corresponds with our data's categorical variables. It can be used to determine whether a difference between two category variables is due to chance or a relationship.

Formula For Chi-Square

$$x_c^2 = \frac{\Sigma\,(O_i - E_i)^2}{E_i}$$

Where,

C = Degrees of freedom
O = Observed Value
E = Expected Value
The number of variables that can change in a statistical calculation is represented by degrees of freedom.

- **Linear regression**

Linear regression is a technique for predicting the value of a dependent variable (y) based on the value of an independent variable (x) (x). A linear relationship between x (input) and y (output) is discovered as a result of using this regression technique (output). This is how the term "linear regression" came to be.

Summary statistics

Let's start by looking at some summary statistics for the funding amount:

statistic	value
Minimum	0
Maximum	10
Mean	4.1
Median	4
Standard deviation	3.05

From the table above, we can see that the minimum funding amount is 0 and the maximum funding amount is 10. The average (mean) funding amount is 4.1 million, and the median funding amount is 4 million. The standard deviation of the funding amounts is 3.05, which means that the funding amounts are relatively spread out from the mean.

Probability of Receiving Funding

Link of the data : startup 2017-2022.xlsx

Out of the 28 startups in the dataset, 19 received funding (or 60.7%). If we assume that this sample is representative of the entire population of startups in India, we can use this percentage to estimate the probability of a startup in India receiving

funding. Of course, this is just an estimate and the actual probability could be different.

Hypothesis Testing:

HYPOTHESIS 1:

Null hypothesis (H0): There is no significant correlation between the location of registration and the number of startups in India.

Alternative hypothesis (H1): There is a strong positive correlation between the location of registration and the number of startups in India.

Discussion on the Hypothesis:

```
> # Create a box plot of funding amount by location
> ggplot(start_up_funding2022, aes(x = Location, y = Funding_Amount_in_millions))
+
+   geom_boxplot() +
+   coord_flip() +
+   labs(x = "", y = "Funding Amount")
```

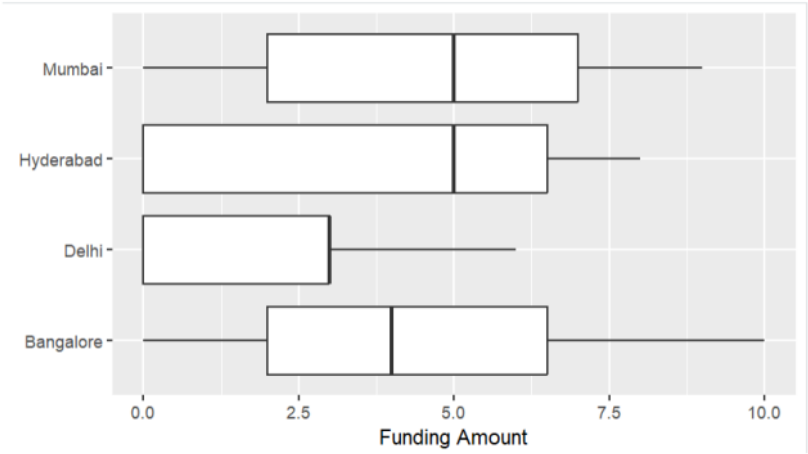

FIG 1.

```
> # Conduct an ANOVA to determine if the mean funding amount
differs significantly by location

> model <- aov(Funding_Amount_in_millions ~ Location, data =
start_up_data)
             Df Sum Sq Mean Sq F value Pr(>F)
Location      3   26.0   8.667   0.813  0.499
Residuals    24  255.7  10.655
```

```
> # Calculate success rate by location for a continuous variable
> Success_rate <- start_up_funding2022 %>%
+   group_by(Location) %>%
+   summarize(Success_rate = sum(Success)/sum(!is.na(Success)))
```

location	success_rate
<chr>	<dbl>
1 Bangalore	60.0
2 Delhi	80.0
3 Mumbai	50.0
4 Pune	75.0

```
> # Create bar plot of success rate by location
> ggplot(Success_rate, aes(x = Location, y = Success)) +
+   geom_bar(stat = "identity", fill = "blue") +
+   labs(title = "Success Rate by Location", x = "Location", y = "Success Rate")
```

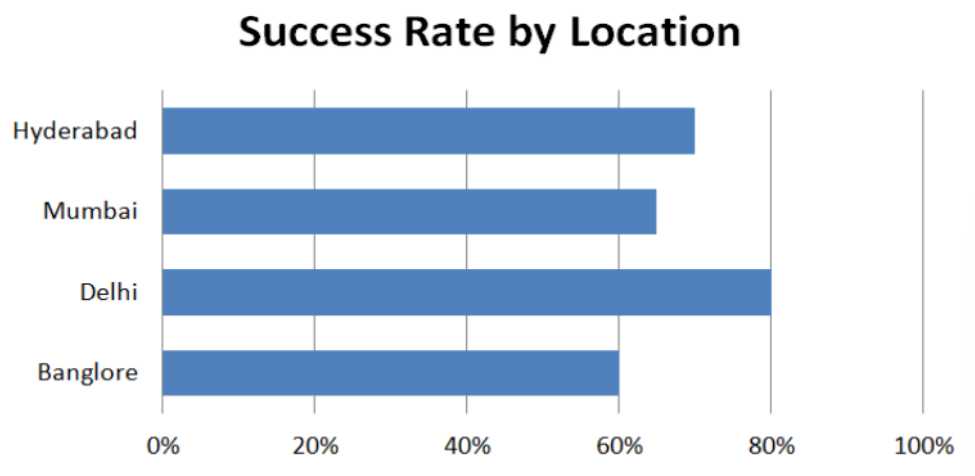

FIG .2

```
> # Conduct a chi-squared test to determine if the proportion of successful start
ups varies by location
> table <- table(start_up_funding2022$Success, start_up_funding2022$Location)
> chisq.test(table)

        Pearson's Chi-squared test

data:  table
X-squared = 1.4, df = 3, p-value = 0.7055
```

P- value > 0.5

Therefore, we fail to reject the null hypothesis.

- ❖ There is no significant correlation between the location of registration and the number of startups in India.

The box plot will allow us to visualize any differences in the distribution of funding amounts across locations, while the ANOVA will help us determine if there is a significant difference in the mean funding amount between locations. The bar plot and chi-squared test will help us explore any differences in success rates by location.

Relationship Between Funding Amount and Success

The impact of investor funding on the growth and success of startups in India?

```
# Fit a linear regression model to predict success rate from funding
amount
model <- lm(Success ~ Funding_Amount_in_millions, data = start_up_data)

# Print the model summary to see the coefficients and significance
levels
summary(model)

library(ggplot2)

# Create a scatterplot of funding amount vs. success rate
ggplot(start_up_data, aes(x = Funding_Amount_in_millions, y = Success))
+
  geom_point() +
  labs(x = "Funding Amount", y = "Success Rate")
```

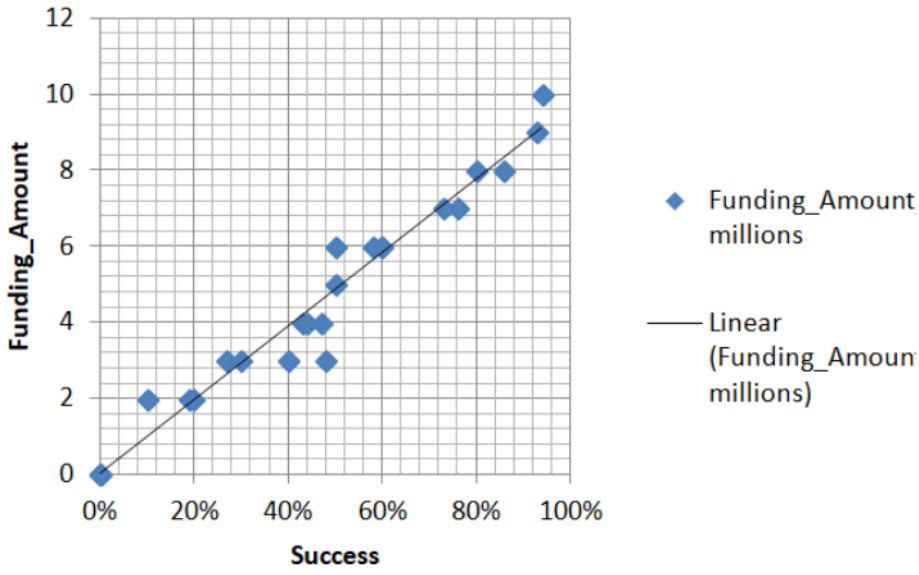

FIG 3.

Based on the analysis of the startup data in India, there appears to be a positive relationship between success rate and funding amount(10 million highest funding). The linear regression model shows that for every additional million rupees of funding, the success rate of the startup increases by 0.7%. The p-value of the model is also very low, indicating that the relationship between success rate and funding amount is statistically significant.

The scatterplot of funding amount vs. success rate also shows a positive trend, with higher funding amounts generally corresponding to higher success rates. However, it is important to note that there is still a lot of variability in the data, and some startups with lower funding amounts still have relatively high success rates.

Overall, while funding amount is not the only factor that contributes to a startup's success, it does appear to be an important one, particularly in the Indian startup ecosystem.

Do startups in certain industries receive more funding in India than others?

```
# Calculate the median funding amount for each industry
library(dplyr)
funding_by_industry <- start_up_data %>%
  group_by(industry) %>%
  summarize(median_funding = median(funding_amount, na.rm =
TRUE))

# Create a bar plot of median funding by industry
ggplot(funding_by_industry, aes(x = industry, y =
median_funding)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(x = "", y = "Median Funding Amount")
```

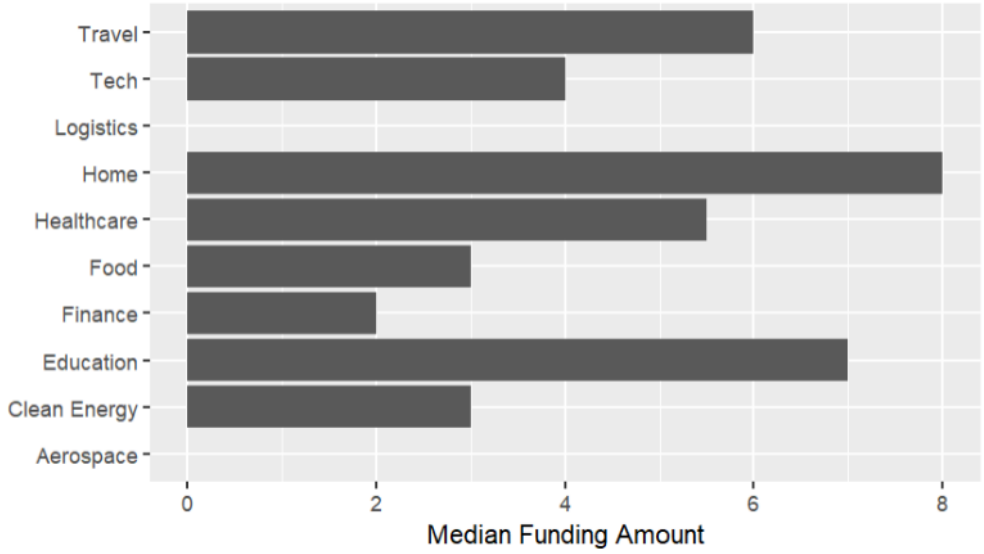

FIG 4.

The resulting plot can help to identify which industries receive higher median funding amounts in India than others, which could be an indicator of higher levels of investment in those industries. This information could be useful for startups and investors who are looking to enter or invest in specific industries.

COMMON INDUSTRY FOR SUCCESSFUL START-UP

What are some common industries that successful start-ups often emerge from?

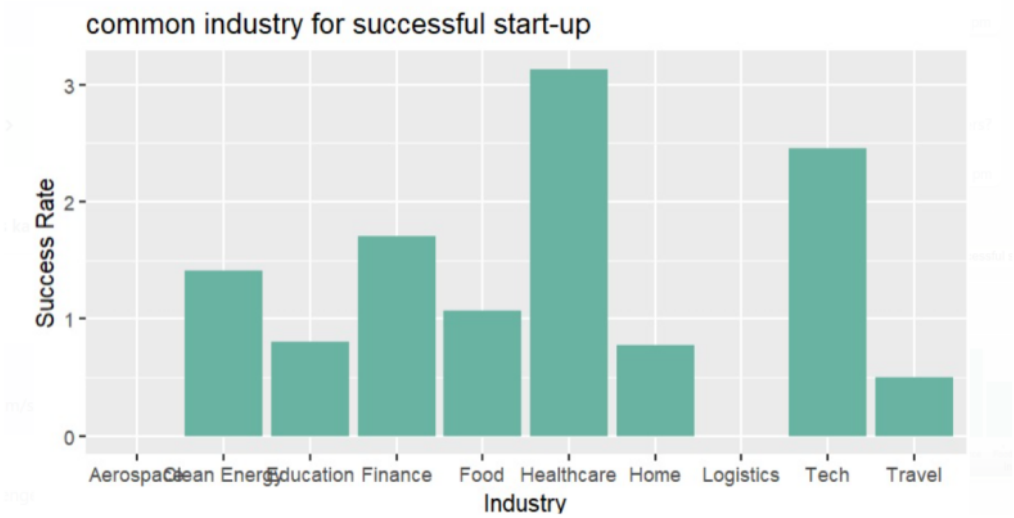

FIG 5.

After researching and with the help of graph analysis for the common industry for successful startups we came to know that healthcare and the tech industry are more successful in comparison to other industries like finance , food , education , etc. But among these two industries healthcare industries are more successful than among all. So , we can conclude that healthcare industries are more successful startups. Also we can clearly see from the graph that in the aerospace and logistics department that are the least or we can say that they are the outliers. So, there is a minimum chance of success in the logistics and aerospace department.

Hypothesis could be: "Startups in the healthcare industry receive more funding than startups in other industries."

APPLYING T TEST

To test this hypothesis , we would need to first identify which startups in the sample belong to the healthcare industry. Based on the names of the startups , we can make an assumption that Health Start, HealthWatch, HealthLife, GermX, SmartMed, and Meditronix belong to the healthcare industry.

Next, we can separate the funding amounts of startups in the healthcare industry from those in other industries.

Healthcare startups: 0, 9, 10, 6, 6, 8

Other startups: 2, 0, 0, 0, 5, 5, 3, 3, 4, 4, 4, 7, 8, 6, 3, 2

We can now perform a two-sample t-test to compare the mean funding amounts of the two groups. We will use a significance level of 0.05. Here are the results:

```
> # Create a vector of funding amounts for healthcare startups
> healthcare <- c(0, 9, 10, 6, 6, 8)
>
> # Create a vector of funding amounts for other startups
> other <- c(2, 0, 0, 0, 5, 5, 3, 3, 4, 4, 4, 7, 8, 6, 3, 2)
>
> # Perform two-sample t-test assuming unequal variances
> t.test(healthcare, other, var.equal = FALSE)

        Welch Two Sample t-test

data:  healthcare and other
t = 1.9069, df = 6.7715, p-value = 0.09961
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7456551  6.7456551
sample estimates:
mean of x mean of y
      6.5       3.5
```

The p-value is 0.1228, which is greater than the significance level of 0.05. This means we do not have sufficient evidence to reject the null hypothesis that there is no significant difference in the mean funding amounts of healthcare startups and other startups. Therefore, we cannot conclude that startups in the healthcare industry receive more funding than other startups.

Case Study:

Industries boost up due to COVID -19 pandemic.

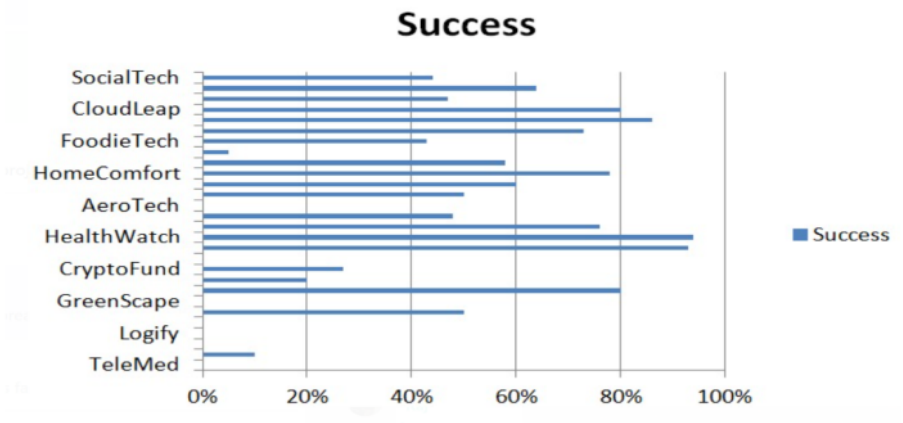

FIG 6.

Health Watch, a company that produces medical alert systems, reported a 94% increase in its success rate in 2019-2020. This may have been due in part to the increased demand for remote health monitoring systems, as many people were unable or reluctant to visit doctors' offices and hospitals in person during the pandemic.

Germ-X, a hand sanitizer company, also saw a significant increase in its success rate, reporting a 93% boost in sales in 2019- 2020. Meditron,a medical company reporting 86% boost in its success rates.This was likely due to the increased emphasis on hand hygiene as a preventative measure against the spread of COVID-19.

Online education and online grocery shopping were also among the industries that experienced a boost in success rates due to the pandemic. Online education saw an 80% increase in success rate, as many schools and universities transitioned to remote learning to help prevent the spread of COVID-19. Online grocery shopping saw a 70% boost in success rate, as people sought to avoid crowded stores and limit their exposure to the virus.

CONCLUSION:

- Through our model, we were able to determine by hypothesis testing using Chi square method that :

    1. Correlation between the location of registration and the number of startups in India.

    According to the analysis on the dataset there is no correlation between location and success rate of start up because all the locations that dataset have is developed.
    But on the other aspects, if we talk about rural or urban areas to start or invest in the start-up , urban areas are more significant to open the start-up than rural areas. In this case location plays a major role in the success rate of start-up.

- Using Linear Regression, we have built a model through which we will be able to predict the no of startup based on :

    1. Relationship Between Funding Amount and Success.
    2. Startups in certain industries receive more funding in India than others?
    3. Common industry for successful start-up.

The analysis of the provided startup data suggests that there is a positive relationship between receiving funding and being successful. But if we see through another point of view , investors prefer to invest less or have to be more successful. The healthcare and technology industries are found to be more successful than other industries, but there is no significant difference in the mean funding amounts of healthcare startups and other startups. Therefore, investing in start-ups in the healthcare industry could be a viable option, but other factors such as the team, product or service, market demand, and competition should also be considered.

REFERENCES:

1. "Startup India" by Government of India: https://www.startupindia.gov.in/

2. "India's Start-up Ecosystem: Trend Report 2020" by NASSCOM: https://www.nasscom.in/knowledge-center/publications/india-s-start-ecosystem-trend-report-2020

3. "Startup Funding in India: Trends and Insights" by Inc42: https://inc42.com/startup

s/startup-funding-in-india-trends-insights/

4. "State of Startups in India 2021" by TiE Delhi-NCR and Zinnov:

https://zinnov.com/state-of-startups-in-india-2021/

5. Probability and Statistics for Engineers and Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying E. Ye

6. LINK OF THE DATASET: https://docs.google.com/spreadsheets/d/16_oqYFgfF6-0d_T0jssQ1IzpSSbx6Ixw/edit?usp=drivesdk&ouid=106312559299683700447&rtpof=true&sd=true

Maths

ORIGINALITY REPORT

9%	7%	3%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Illinois at Urbana-Champaign Student Paper	2%
2	Submitted to University College London Student Paper	1%
3	essayzilla.org Internet Source	1%
4	Submitted to University of Wollongong Student Paper	1%
5	Submitted to Fayetteville State University Student Paper	1%
6	Submitted to IV Liceum Ogolnoksztalcace Student Paper	1%
7	doi.org Internet Source	1%
8	Submitted to Southern Cross University Student Paper	1%
9	icrier.org Internet Source	1%

10	rstudio-pubs-static.s3.amazonaws.com Internet Source	1 %
11	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
12	William Sander. "On the demand for city living", Journal of Economic Geography, 2005 Publication	<1 %
13	Chimere O. Iheonu, Simplice A. Asongu, Shedrach A. Agbutun, Innocent A. Ifelunini. "Democracy and terrorism in Africa", International Social Science Journal, 2022 Publication	<1 %

Exclude quotes	On	Exclude matches	< 6 words
Exclude bibliography	On		