



**Analysis On Startups**

**By Team 7**

1. DYUTI DASMAHAPATRA (210C2030142)

2. HIMANSHI AGGARWAL (210C2030205)

3. ADARSH WADAL (210C2030145)

4. SAMAKSH GUPTA (210C2030126)

5. RITIK DATA (210C2030174)



SCHOOL OF ENGINEERING AND TECHNOLOGY BML MUNJAL UNIVERSITY GURGAON Year 2022





**Abstract**

India is a global hotspot with several unicorn startups competing against the rest of the globe on multiple platforms and sectors. In addition to privately owned, LLP and foreign companies, India registers more than 10,000 private limited companies each month with the Department of Corporate Affairs. Analyzing data on registration places and business types can reveal the direction the industry is moving. This gives us a glimpse of the future.

**Introduction**

A startup is a business in its development.

Founders often fund their own enterprises and may seek outside money before they launch. Family and friends, venture capitalists, crowdfunding, and loans are all potential sources of finance. Startups must also assess their legal structure and the location of their operations. Startups have a high chance of failure, but they may also be very unique places to work, with great rewards, an emphasis on innovation, and good learning opportunities.

**GROWTH IN INDIA**

According to Sanjeev Malhotra, CEO of NASSCOM Center of Excellence for Internet of Things (IoT) and Artificial Intelligence, India has the third-largest startup ecosystem in the world, with the number of such companies growing by 10% every year.

**FACTORS ON WHICH STARTUP DEPENDS UPON:**

**1-Start-up capital**

Capital is one of the most crucial parts of financial management for new businesses. The amount of money you'll need to establish and run your business is referred to as capital. It could be in the form of non-cash or cash. To figure out how much money is needed, a list is created of all the potential investments that will be needed to start and manage the business, such as real estate construction, purchase, operational expenses and leasing payments.

**2-Importance of hiring the right workforce**

Starting a new business venture necessitates in taking on various tasks in various situations. However, as the company grows, it will require a strong staff to succeed. Recruiting proper employees for the new startup is just as critical as having a fantastic concept or product. People are what make a business successful.

**3-Government compliances and company registrations**

Before you begin operations, make sure you have all of the appropriate registrations in order to legalize your business. To comply with government rules, you'll need an address proof, a business license, a PAN card, and (if required) a GST registration certificate. To minimize complications, remember to file your GST and other tax returns on time.

**4-Marketing for start-ups**

Remember that having a great product, a talented crew, and enough cash to build your business and make it a success isn't enough. To grow it, you'll need to attract more customers, sales, and money, which you can do with a complete start-up marketing strategy. Unlike traditional tactics, start-up marketing is important. The start-up can be promoted via a variety of tactics and platforms. However, it would be good if you investigated your target audience, their activities, and your company's existence. You must choose the best marketing platforms, combine them, and create an unmatched strategy.

**5-Accounting Analysis**

Finally, before starting a business, ensure that you have a solid financial reporting system in place. You can keep track of what's going on in your firm with a trustworthy accounting and financial reporting system. Consider how Indian start-ups are using the digital platform to revolutionize how they work today. Not only that, but there was a beneficial change in traditional cash flow and all other financial services when startups like OkCredit entered the market.

**WHAT WE HAVE DONE IN THIS PROJECT?**

'In this project, we are going to analyze two factors 'location of startup' and 'type of business’, we are going to use hypothesis testing and regression analysis to test whether the number of startups depend on these factors or not.

**Aims and Objective of the study:**

**(i) To know if the startups in India are dependent on location of registration**

**(ii) To know if the startups in India are dependent on the type of business.**

**(iii) Linear relationship between number of startups and location of registration.**

**(iv) Linear relationship between number of startups and type of business.**

**Research Hypotheses:**

**HYPOTHESIS 1**

H0: There is no relation between the number of startups and where they are registered in India.

H1: The number of startups and the place where they are registered in India have a strong correlation.

**HYPOTHESIS 2**

H0: In India, there is no discernible link between the number of startups and the type of firm.

H1: In India, the number of startups and the type of business have a substantial link.

**Materials and methods**

* **Hypothesis testing**

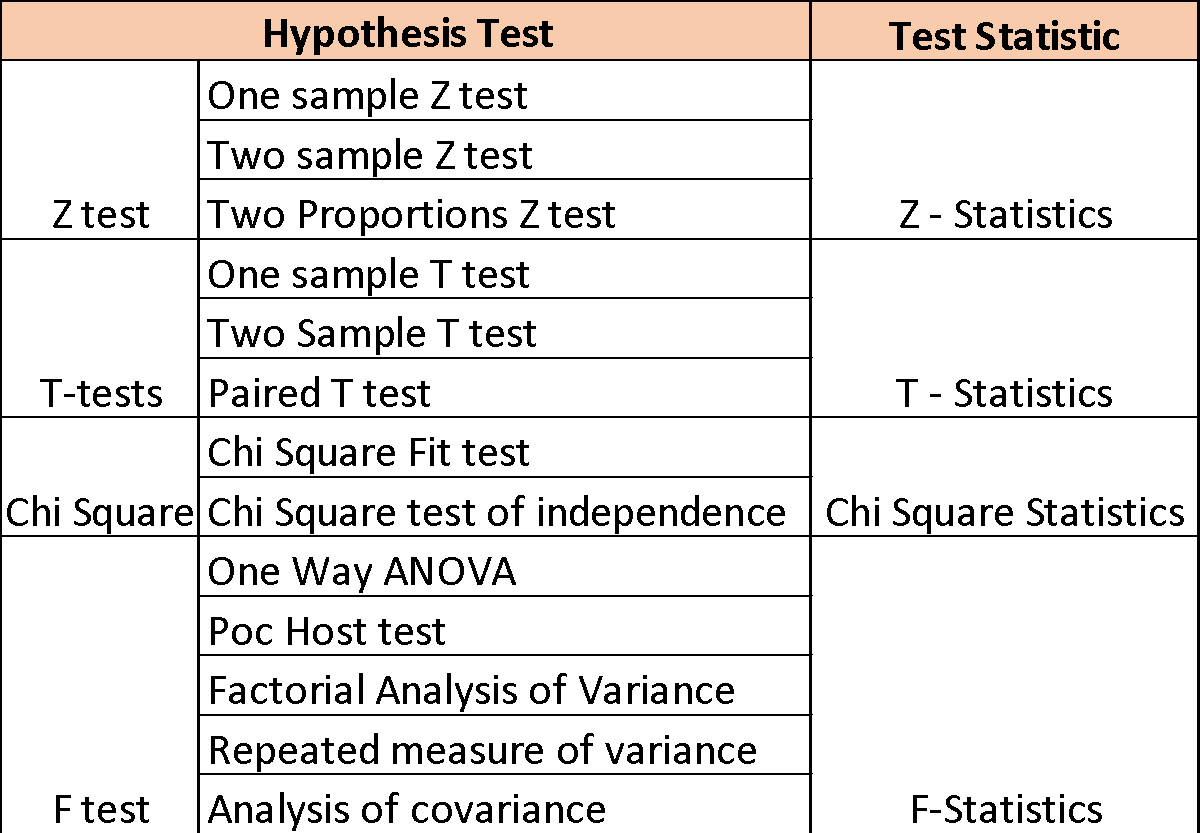
A hypothesis is a possible true or incorrect assertion concerning population distribution. In other words, a hypothesis is a question with a "yes" or "no" response; the "yes" response is considered accepting, or not rejecting, the null hypothesis. H0

Hypothesis testing is a type of statistical analysis in which your assumptions about a population parameter are put to the test. It is used to determine the relationship between two statistical variables.

There are two types of hypotheses in statistics: null and alternative hypotheses.

According to the Null Hypothesis, the event will not occur. The null hypothesis has no bearing on the study's outcome unless it is disproved. H0 is the symbol for it.

The Alternate Hypothesis' logical polar opposite is the Null Hypothesis. When the null hypothesis is rejected, the alternative hypothesis is accepted. H1 is the symbol for it.

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

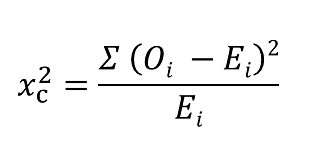
**Table 1: Hypothesis Tests**

We have used the Chi square test in our project.

* **Chi square test**

The Chi-Square test is a statistical method for detecting the difference between observed and expected data. 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 Test:

Where,

c = Degrees of freedom

O = Observed Value

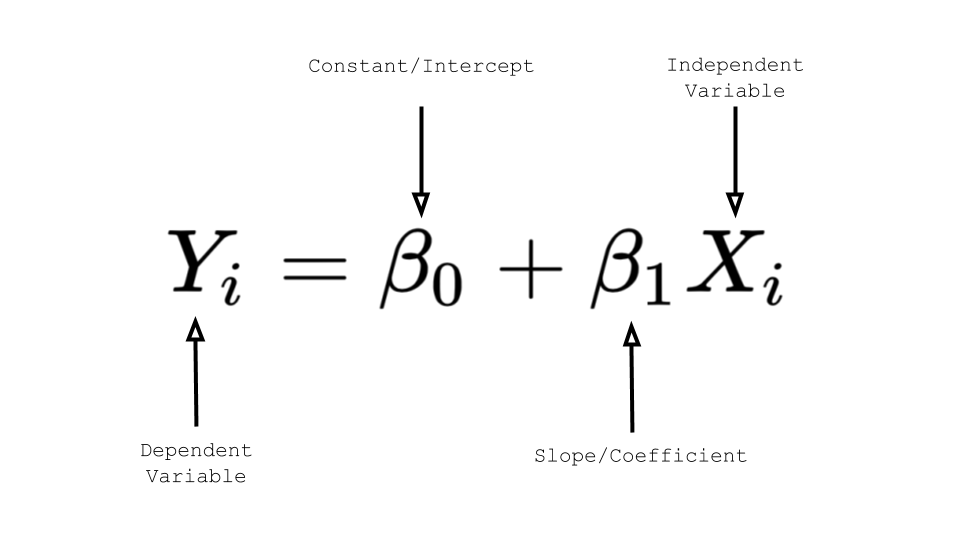
E = Expected Value

The number of variables that can change in a statistical calculation is represented by degrees of freedom.

* **Linear regression**

The machine learning method linear regression is a supervised learning technique. It performs a regression procedure. A target prediction value is calculated using independent factors. It's largely used in forecasting and figuring out how variables are related. The sort of link between dependent and independent variables that different regression models examine, as well as the number of independent variables that they use, varied.

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.

Formula for linear regression:

* **Python and Jupyter notebook**

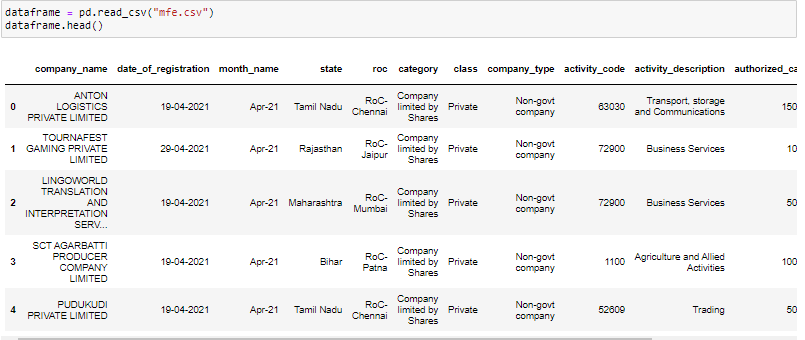
Python is a widely used, high-level programming language. In web development, machine learning applications, and all cutting-edge software technology, the Python programming language (latest Python 3) is used.

Jupyter Notebook is a free and open-source web application that allows you to create and share documents with live code, equations, visualizations, and text.

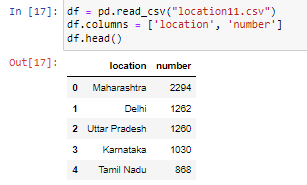
**Results and Discussion:**

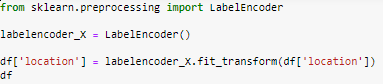
**Hypothesis Testing using Chi-Square test:**

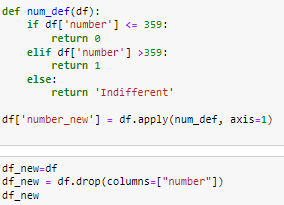
1. Testing dependency on location of startup companies in India.

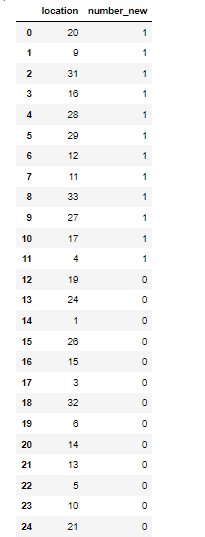
Uploading csv file:

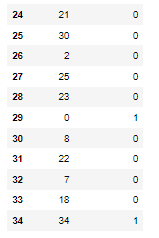
1. Uploading data set:



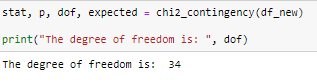
1. Encoding Categorical data



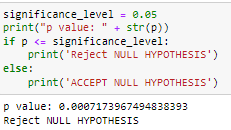




1. Calculating degrees of freedom:



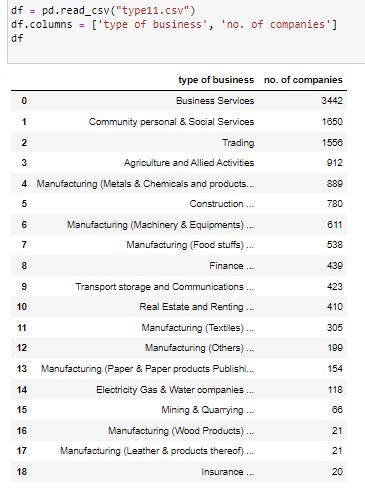
1. Testing Hypothesis:

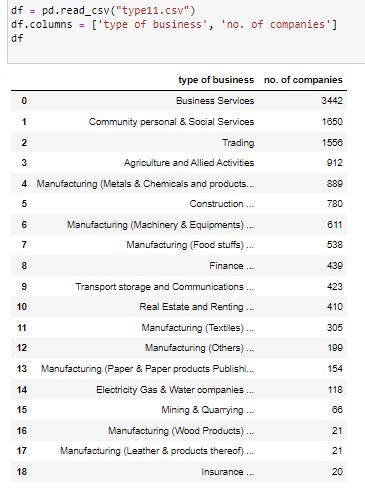


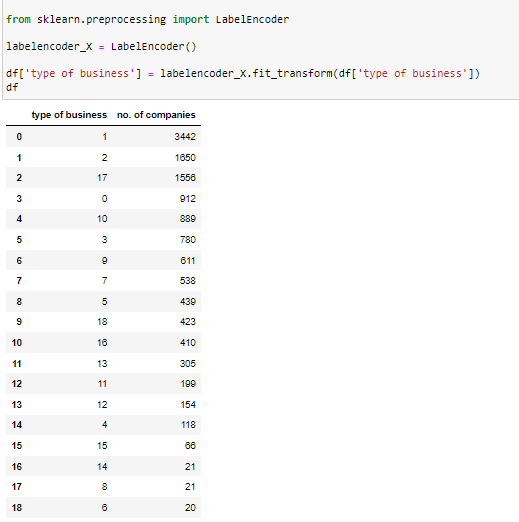
Where **ho-> No. of Startups and their location are independent of each other.**

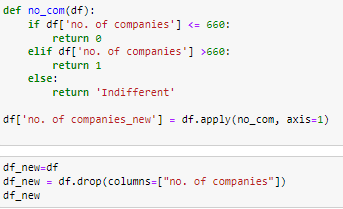
**h1-> No. of Startups and their location are dependent on each other.**

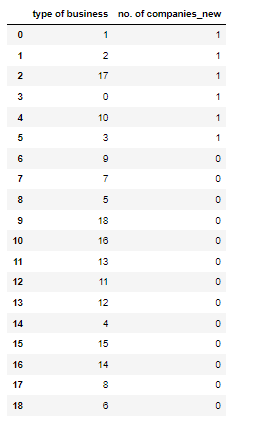
b. Testing dependency on type of business of startup companies in India.

1. Uploading data set:

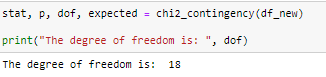


1. Encoding categorical data:

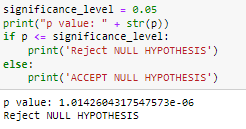




1. Calculating degree of freedom:

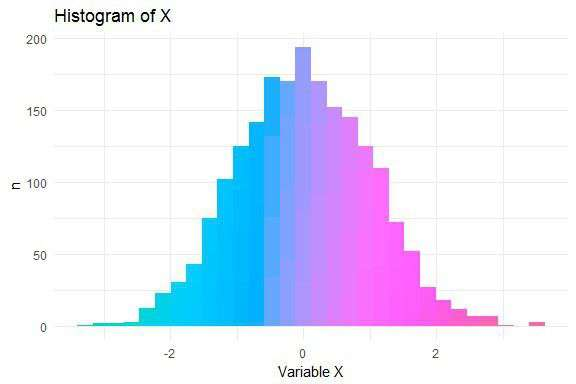


1. Testing hypothesis:

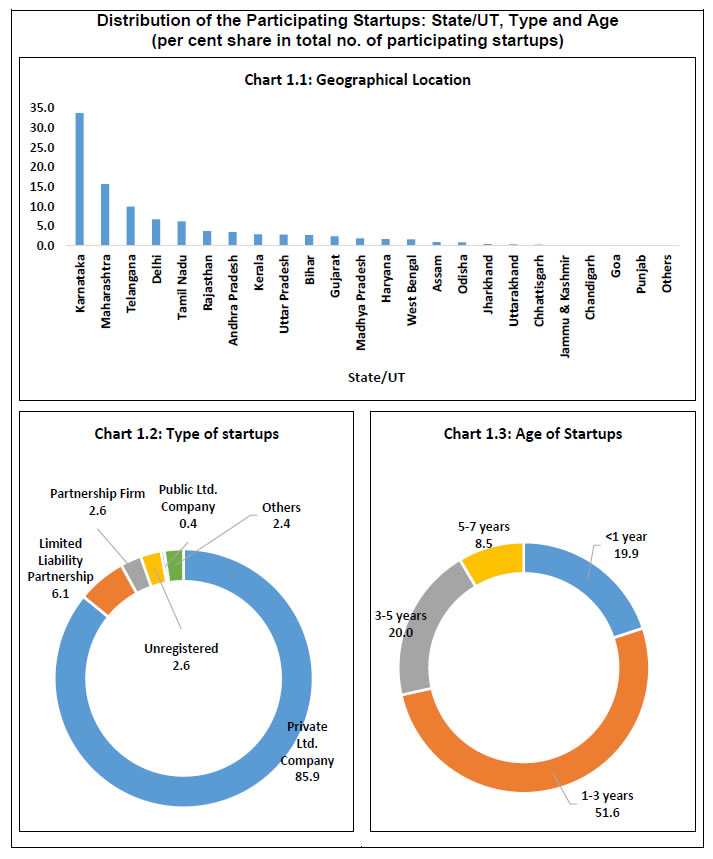


Where **ho-> No of Startups and its type are independent of each other.**

**h1->= No of Startups and its type are dependent on each other.**



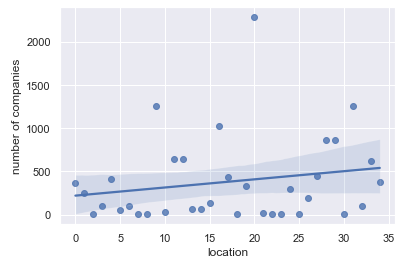
**Graph  for normal distribution location and number of startup**



**Regression Analysis:**

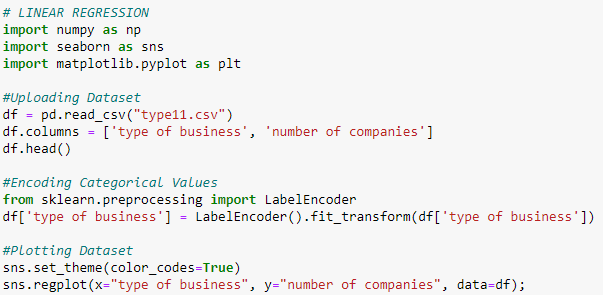
1. **Relationship between no of companies and its location.**

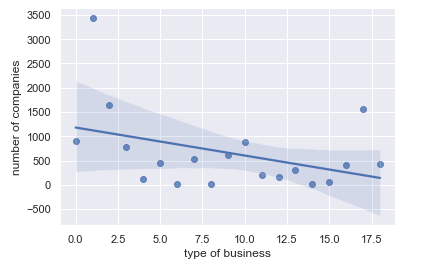
****

****

**Graph 1: Relationship between no of companies and its location**

1. **Relationship between no of companies and type of business.**

****

****

**Graph 2: Relationship between no of companies and type of business**

**Further Work:**

* We can use the above hypothesis testing model to determine the relationship between the number of startups and other factors such as capital, class, company etc.
* We can also use the linear regression model to predict the number of startups based on other factors.

**Conclusions:**

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

1. Number of startups and their location are dependent on each other.
2. Number of startups and their type are dependent on each other.

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

1. Location of the startup
2. Type of business of the startup

Thus, we came to a conclusion that the number of startups is dependent on both the location and the type of business.

**References**

* <https://towardsdatascience.com/chi-square-test-with-python-d8ba98117626> : Chi square test using Python
* <https://realpython.com/linear-regression-in-python/>: Linear regression in python
* <https://www.kaggle.com/>: Datasets
* <https://www.analyticsvidhya.com/blog/2021/11/startups-profit-prediction-using-multiple-linear-regression/>: Multiple linear regression
* [How to Perform Simple Linear Regression in Python (Step-by-Step) (statology.org)](https://www.statology.org/simple-linear-regression-in-python/#:~:text=How%20to%20Perform%20Simple%20Linear%20Regression%20in%20Python,following%20form%3A%20%C5%B7%20%3D%20b0%20%2B%20b1x%20): For Simple Linear Regression
* [seaborn.pairplot — seaborn 0.11.2 documentation (pydata.org)](https://seaborn.pydata.org/generated/seaborn.pairplot.html)
* Probability and Statistics for Engineers and Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying E. Ye

**Acknowledgments**

With immense satisfaction, I would like to express my heartfelt gratitude and appreciation to Ziya Uddain Sir for the sincere support he has given us. We are grateful for being given the opportunity to complete this project and for providing us with the chance to expand our and everyone else's understanding of hypothesis testing and regression analysis.

I'd also like to thank my classmates for being so cooperative and helpful during the entire endeavor.