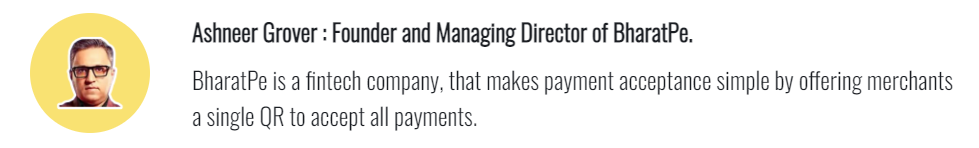
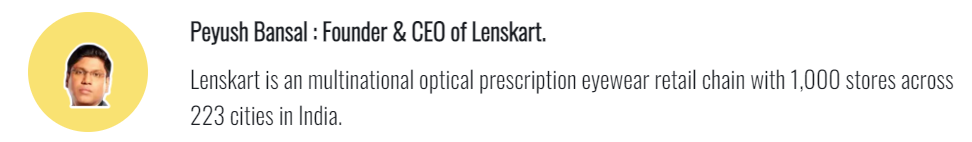
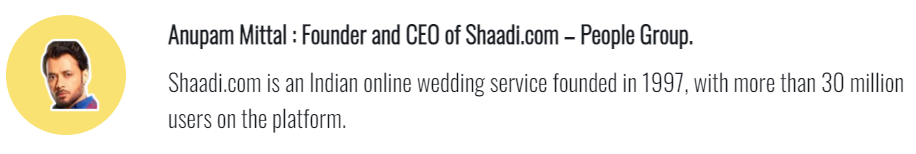
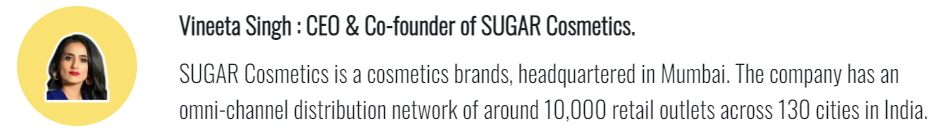
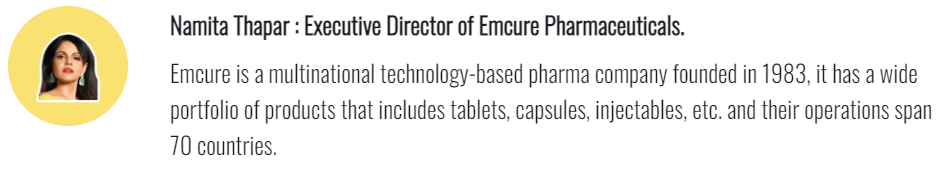
 ­­­

 Shark Tank India is an Indian Hindi-language business reality television series that airs on Sony Entertainment Television. It shows [entrepreneurs](https://en.wikipedia.org/wiki/Entrepreneurship) making business presentations to a panel of investor or sharks, who decide whether to invest in their company. Shark Tank India (season 1) aired on Sony LIV and Sony Entertainment Television from 20 December 2021 to 4 February 2022. In season 1 there are seven Sharks, Here’s a brief introduction of each shark in:



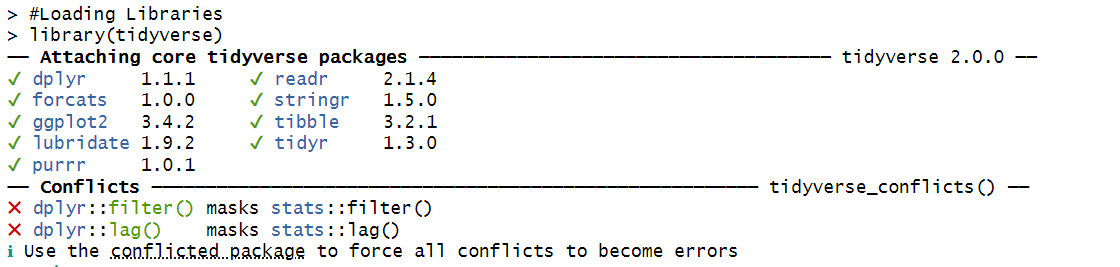


**DATA COLLECTION**

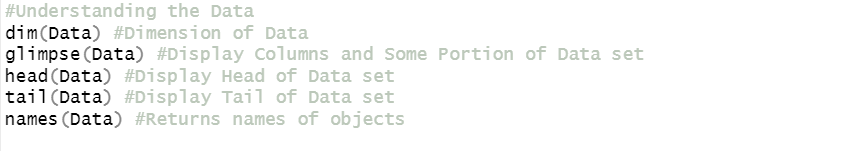
This study is based on secondary data. The secondary data for this project was obtained from publicly available sources such as Wikipedia, data.world, github and kaggle.

You can easily download the data from the link below:

**** **IMPORTING THE LIBRERIES AND LOADING THE FILE**

* **Output:-**

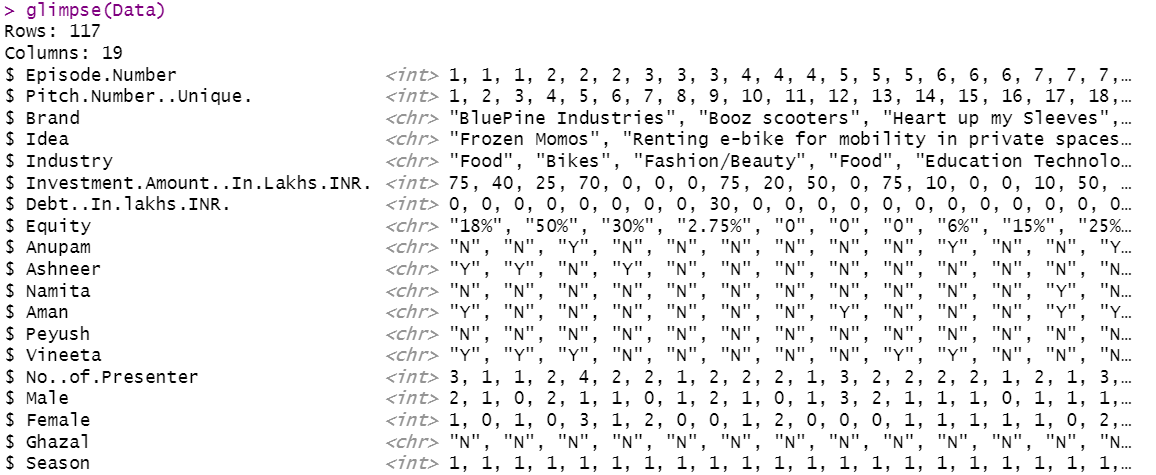
**UNDERSTANDING THE DATASET**

****

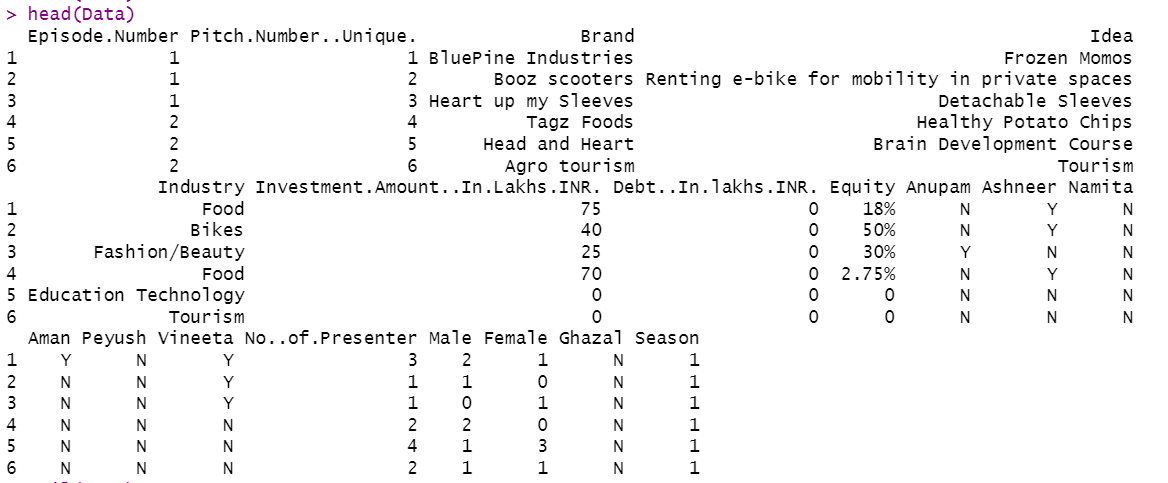
**Dimension of Data-set**

****

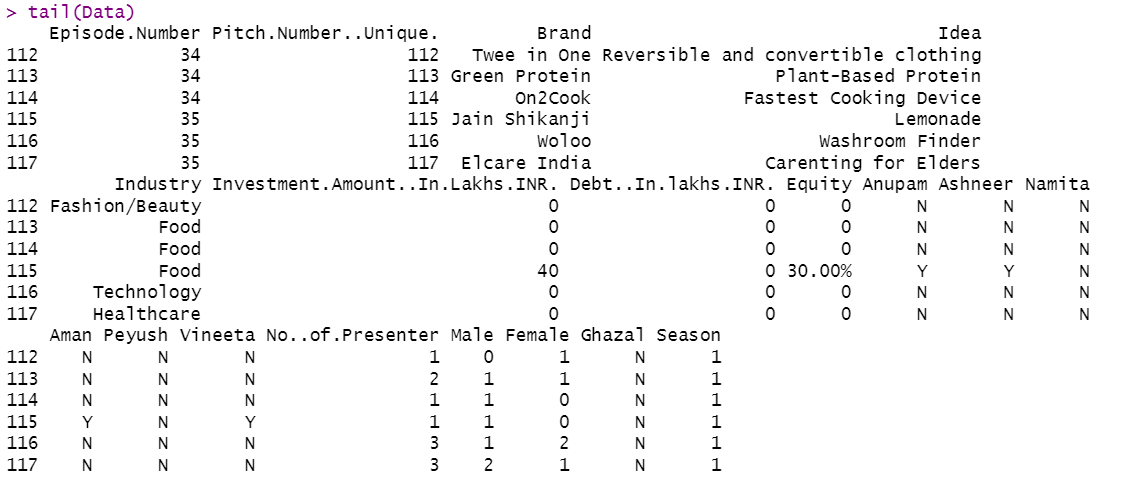
**Glimpse of Data-set**

****

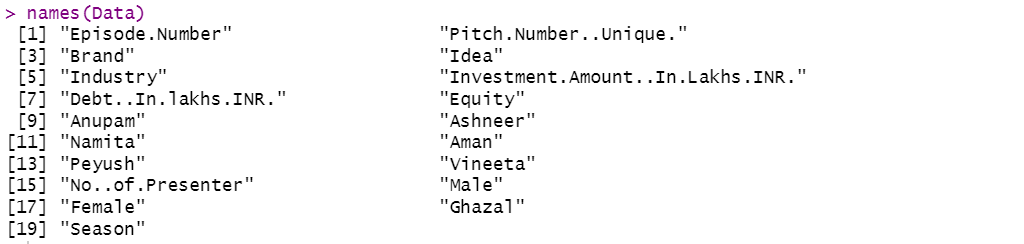
**Head of Data-set**

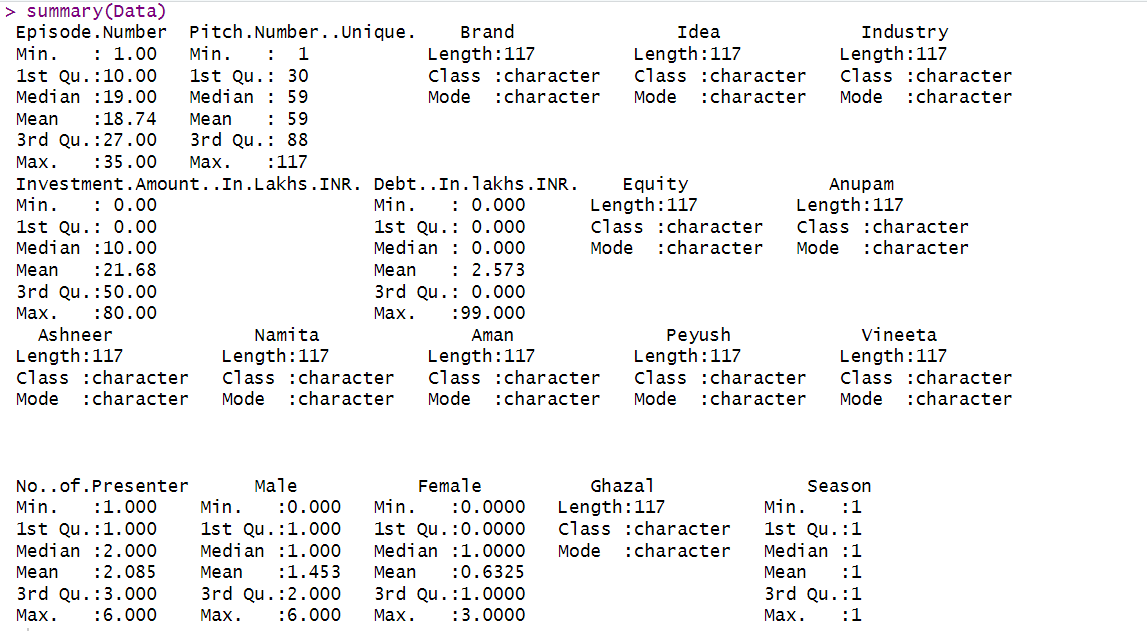
****

**Tail of Data-set**

****

**Names/Columns of Data-set**

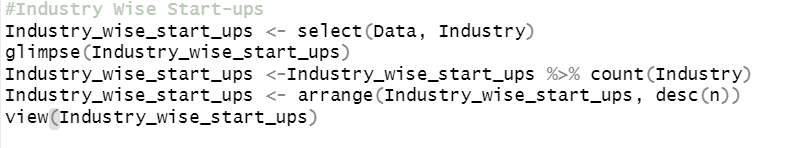
****

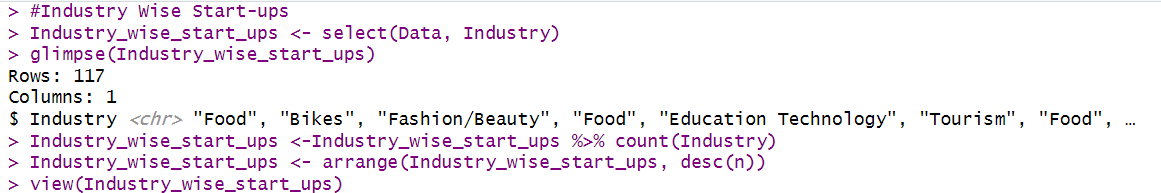
** Summary of Data-set**

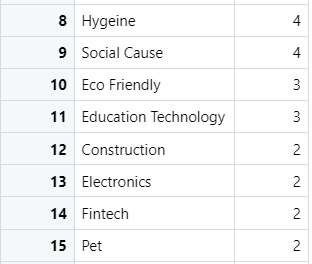
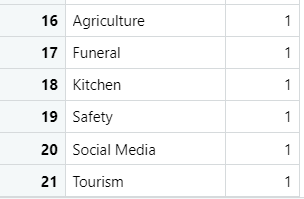
**VISUALIAING DATA USING GGPLOT2**

Visualization plays an important role in the decision-making process after analysing relevant data. For data visualizing we are going to use ggplot2 , ggplot2 package in R Programming Language also termed as Grammar of Graphics is a free, open-source, and easy-to-use visualization package widely used in R. It is the most powerful visualization package written by Hadley Wickham.

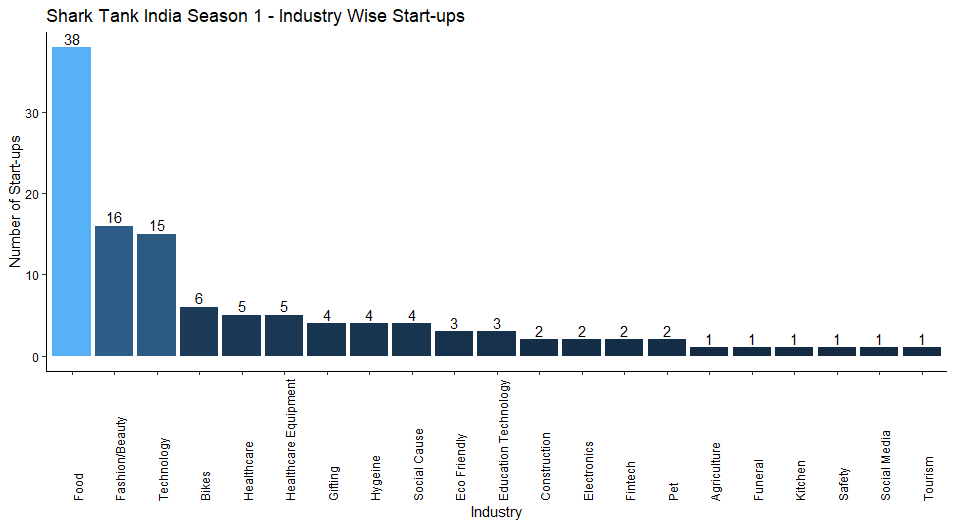
**Industry Wise Start-ups**

****

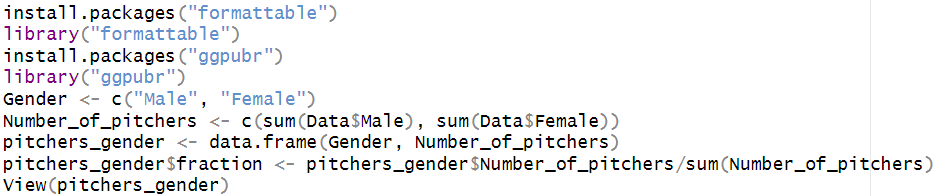
* **Output**

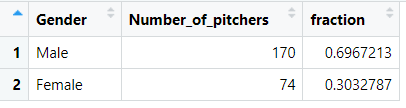
****

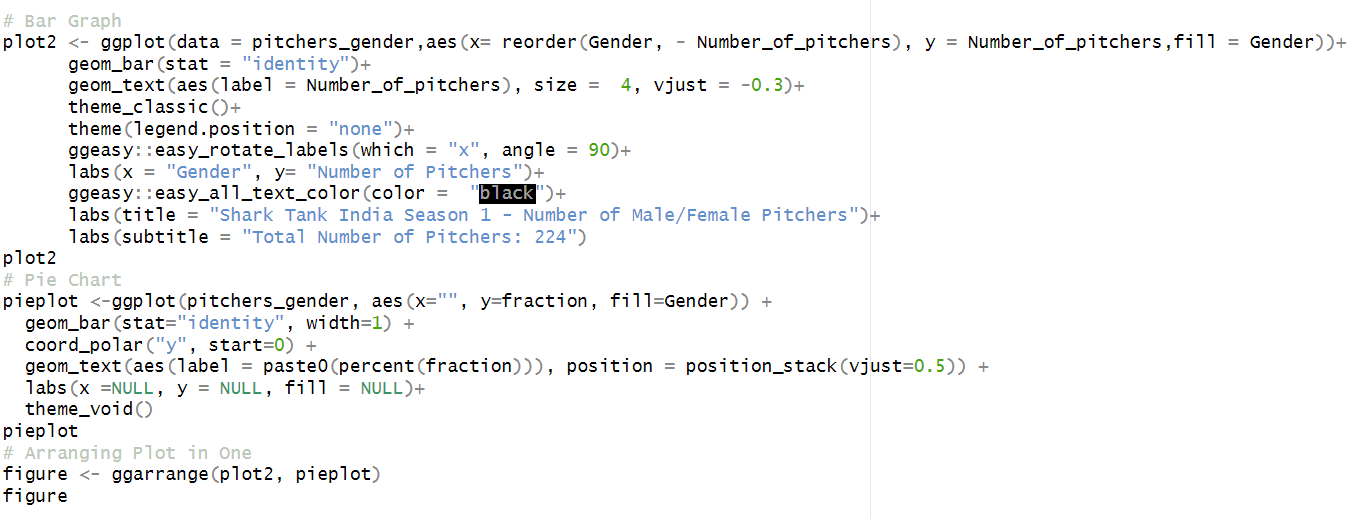
****

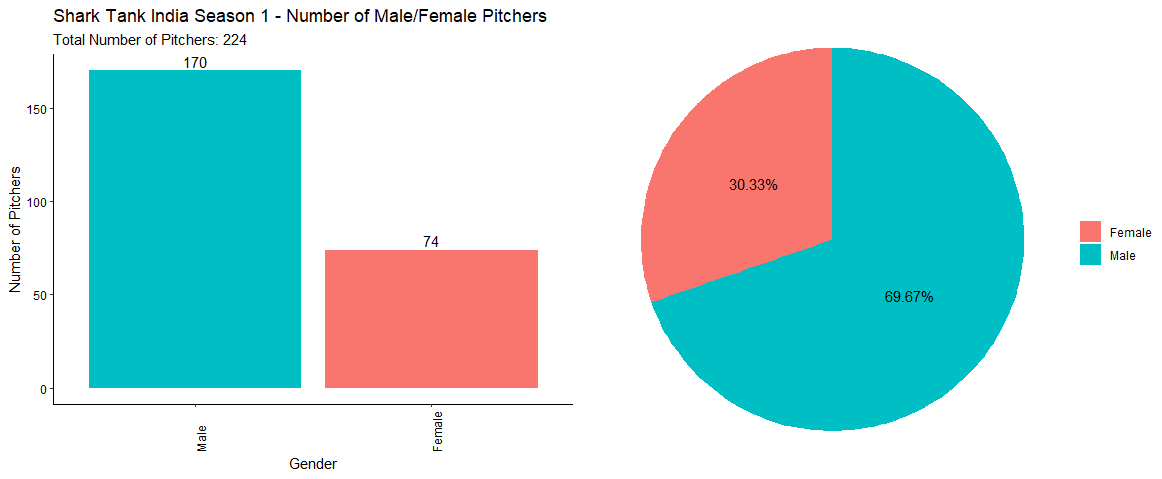
****

**Number of Pitchers Gender Wise**

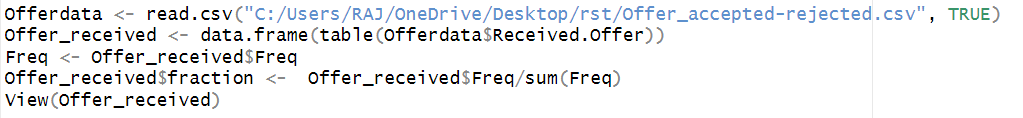
****

****

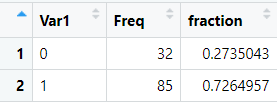
****

****

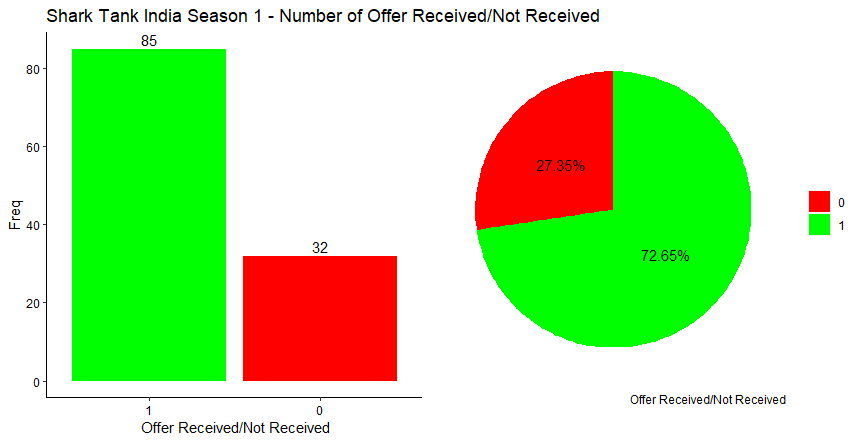
**Offer Received/Not Received**

****

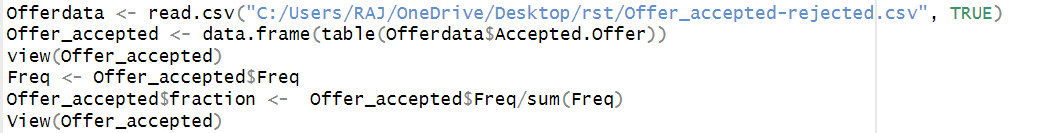
****



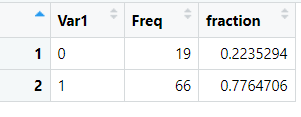
* **­­73% of start-ups received offers from sharks and 27% didn’t received offers from sharks**



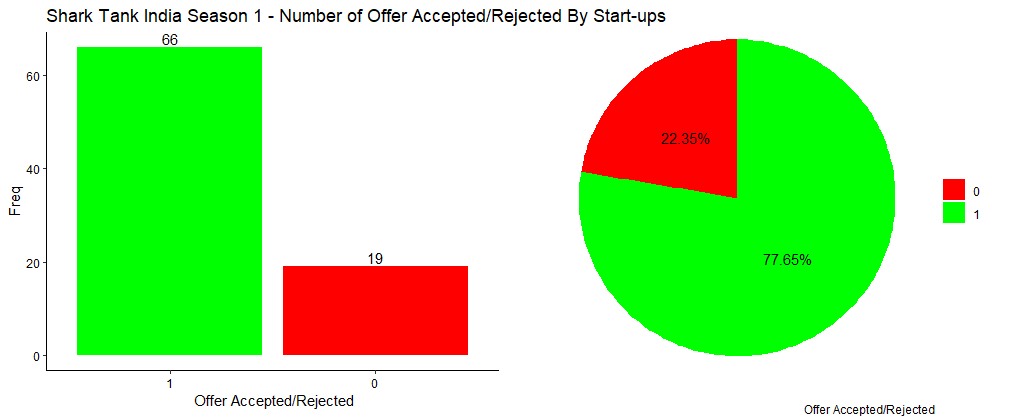
**Offer Accepted/Rejected**

****

****



* **23% of** **start-ups accepted offers from sharks and 78% of start-ups rejected offers from sharks.**



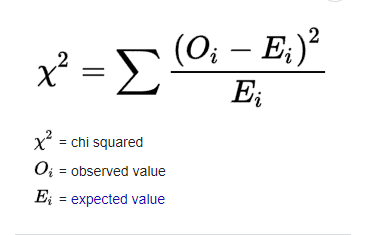
# 

# CHI-SQUARE STATISTICS

Pearson’s chi-square (Χ2) tests, often referred to simply as chi-square tests, are among the most common nonparametric tests. Nonparametric tests are used for data that don’t follow the assumptions of parametric tests, especially the assumption of a normal distribution. If you want to test a hypothesis about the distribution of a categorical variable you’ll need to use a chi-square test or another nonparametric test. Categorical variables can be nominal or ordinal and represent groupings such as species or nationalities. Because they can only have a few specific values, they can’t have a normal distribution.

**The chi-square formula**

Both of Pearson’s chi-square tests use the same formula to calculate the test statistic, chi-square:



The larger the difference between the observations and the expectations (O − E in the equation), the bigger the chi-square will be. To decide whether the difference is big enough to be statistically significant, you compare the chi-square value to a critical value.

There are two types of chi-square tests. Both use the chi-square statistic and distribution for different purposes:

* A chi-square goodness of fit test determines if sample data matches a population. For more details on this type, see: Goodness of Fit Test.
* A chi-square test for independence compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each another.

**When to use a chi-square test**

A Pearson’s chi-square test may be an appropriate option for your data if all of the following are true:

1. You want to test a hypothesis about one or more categorical variables. If one or more of your variables is quantitative, you should use a different statistical test. Alternatively, you could convert the quantitative variable into a categorical variable by separating the observations into intervals.
2. The sample was randomly selected from the population.
3. There are a minimum of five observations expected in each group or combination of groups.

**How to perform the chi-square test of independence**

Follow these steps to perform a chi-square test of independence:

**Step 1: Calculate the expected frequencies**

Use the contingency table to calculate the expected frequencies following the formula:

\begin{equation*}\dfrac{(\textup{Row}\,\, r \,\,\textup{total}\,\, \times \textup{Column}\,\, c \,\,\textup{total})}{\textup{Grand}\,\, \textup{total}}\end{equation*}

**Step 2: Calculate chi-square**

Use the Pearson’s chi-square formula to calculate the test statistic:

\begin{equation*}X^2 = \sum {\frac {(O - E)^2}{E}}\end{equation*}

**Step 3: Find the critical chi-square value**

You can find the critical value in a chi-square critical value table or using statistical software. You need to know two numbers to find the critical value:

* **The degrees of freedom (df):** For a chi-square test of independence, the df is (number of variable 1 groups − 1) \* (number of variable 2 groups − 1).
* **Significance level (α):** By convention, the significance level is usually .05.

**Step 4: Compare the chi-square value to the critical value**

Is the test statistic big enough to reject the null hypothesis? Compare it to the critical value to find out.

**Step 5: Decide whether to reject the null hypothesis**

* If the Χ2 value is greater than the critical value, then the difference between the observed and expected distributions is statistically significant (p < α).
  + The data allows you to reject the null hypothesis that the variables are unrelated and provides support for the alternative hypothesis that the variables are related.
* If the Χ2 value is less than the critical value, then the difference between the observed and expected distributions is not statistically significant (p > α).
  + The data doesn’t allow you to reject the null hypothesis that the variables are unrelated and doesn’t provide support for the alternative hypothesis that the variables are related.

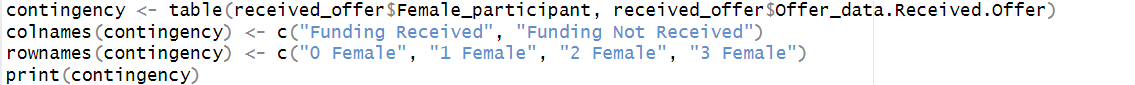
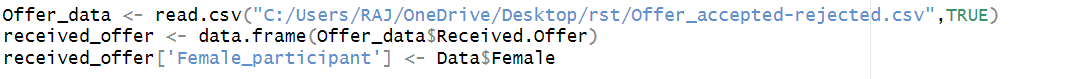
**THE CHI-SQUARE TEST OF INDEPENDENCE OF ATTRIBUTES**

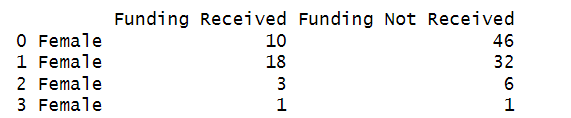
In order to examine whether gender and offer received are associated, we will perform chi-square test of independence of attributes.

Let us take the following hypothesis:

***Null Hypothesis*** **𝐻0*:*** There is no significant association between gender and offer received.

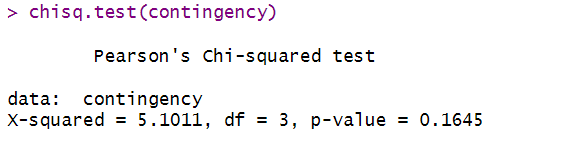
***Alternative Hypothesis* 𝐻𝑎*:*** Significant Association between gender and received.

**** **Contingency Table:**

** Output:**

**Pearson’s Chi-Square Test**

****

** Output:**

* **P-value is greater than 0.05 (Level of Significance “α”) : We fail to reject Null Hypothesis**

**i.e., No significant association between gender and offer received.**