

DSC 682-001 - Summer 2021

Data Visualization for Business Analytics

Xuan Liu

Weijie Zhao

6/29/2021

## Process Book

### Link of Dashboards:

[https://public.tableau.com/views/Final\\_Visualization\\_16248379321770/EntrepreneurialCompetency?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Final_Visualization_16248379321770/EntrepreneurialCompetency?:language=en-US&:display_count=n&:origin=viz_share_link)

### • Motivation and background:

We are both interested in entrepreneurship. And our major is about business which is well related to entrepreneurship. To have a better understanding of how to become an entrepreneur helps us choose a better career in the near future.

Furthermore, in recent years, many new entrepreneurs have succeeded in their different fields. Entrepreneurial competency has been an important nation-wide scale in university students. The features or traits of successful entrepreneurs makes us feel so curious. They must have more unique and brilliant qualities or traits than other people. Hence, during this project, we tried to know more about students' features or traits of what make great entrepreneurs. It would be very helpful and practical for university students to consider whether being an entrepreneur or not. If they want to be entrepreneurs, because of our project, they could understand what kinds of features or traits they need to own or build during their daily life.

• **Research questions:**

1. Which features are the most important for making a great entrepreneur?
2. Why are male entrepreneurs more than female entrepreneurs?
3. How does staying in a city affect entrepreneurship success?

• **Data:**

Source of the data(link):

<https://www.kaggle.com/namanmanchanda/entrepreneurial-competency-in-university-students>. We used Kaggle (<https://www.kaggle.com/datasets>) to find this dataset.

The dataset was collected in 2019 by Utkarsh Sharma and Naman Manchanda who are from India. And the dataset comprises 16 features collected from university students in India. These 16 features consists of education sector, if the student builds personal project, age of student, sex of student, if the student stays in a city, if the student is influenced by someone, rating of a student based upon perseverance, rating of a student based upon desire to take initiative, competitive rating, self reliance rating, strong need to achieve a goal rating, self confidence rating, good physical health rating, if there is any mental disorder, key traits of the student and reason for lack of entrepreneurship culture. The features are highly correlated with the target variable which is whether the student is likely to become an entrepreneur or not. Hence, there are 219 observations and 17 variables in this dataset.

• **ETL:**

1. First of all, we set the working directory where we put the target dataset and used the read function to get the dataset. Here, we label null values of the “ReasonsForLack” column into “NA” because we want to label it all into specific and reasonable phrases later.

Besides, the next steps are checking the number of rows, the number of columns, and structure of the dataset.

2. Moreover, in order to do transformation of data, we called the needed library “stringr”.

By looking through the dataset, we found that the target variable named “y” is not clear enough for audiences. Therefore, we rename it to “Entrepreneur”, which means whether the student seems to become an entrepreneur or not. Then, show 91 missing values by applying `is.na()` function and detect missing values in columns and observations. we fill out empty values(“NA” we mentioned before) with “Null(be an entrepreneur)” that would be more accessible for audiences. Besides, for the column “ReasonsForLack”, the many reasons showing there which seem a little disorganized and random, we make all phrases to lower-case characters and trim the unnecessary spaces that we may not notice. And by applying string matching, fix the different random reasons into just a few groups. Then, in the “EducationSector” column, there is “Teaching Degree (e.g., B.Ed)”, the parentheses are unnecessary. Hence, we still use string matching to justify it as a “Teaching Degree”, which is more concise. Furthermore, the target variable(the last column) values are “0” or “1” which would be confusing for audiences, so we change it into categorical factors “yes” or “no”.

3. Finally, using `write.csv()` function to load or save the new dataset. We can find it on our computer file.

#### • Design of visualization:

Firstly, in the data visualization, we almost used bar charts, stacked bar charts, pie charts, packed bubble graphs and treemaps to show different information we want to present. The reason why we chose bar and stacked bar charts was that it can clearly show different categories like

gender, yes or no choices, and the amount of entrepreneurs. It would also show the comparison of dissimilar categories and make conclusions easier. Besides, using pie charts is better to learn the proportion of gender and yes or no category. And for packed bubble graphs, they were clear enough to show the different categories' proportion such as different key traits. Besides, the reason why applying tree maps was that it allows us to show information about “key traits” in a very efficient way in a limited space. The main purpose of it was to make non-accurate comparisons between different levels of the hierarchy according to the hue of colors and the proportions.

Secondly, we also applied diverse colors to the graphs, which is important to help separate or label different groups appropriately. It would be easier to extract information from the charts.

Then, we also added subtitles and highlighted the important words to every chart that help the audience read the message directly and clearly. Furthermore, we label the information of numbers or groups in each graph, which makes data messages integrally and the comparison easier.

### • **Implementation:**

For the functionality of the main interactive visualization, we justified the tooltips of most graphs since it would help extract messages easier, clearer and more interactive. For instance, we deleted some unnecessary or repetitive messages and added more related charts into the tooltips which make information connected to each other and more interactive. Then, filters were also applied by us. According to our first research question, studying people who can become entrepreneurs is more important. So we filter the people who could not be entrepreneurs. And for the second question, it is about different sex. Hence, there is a gender filter. For the third

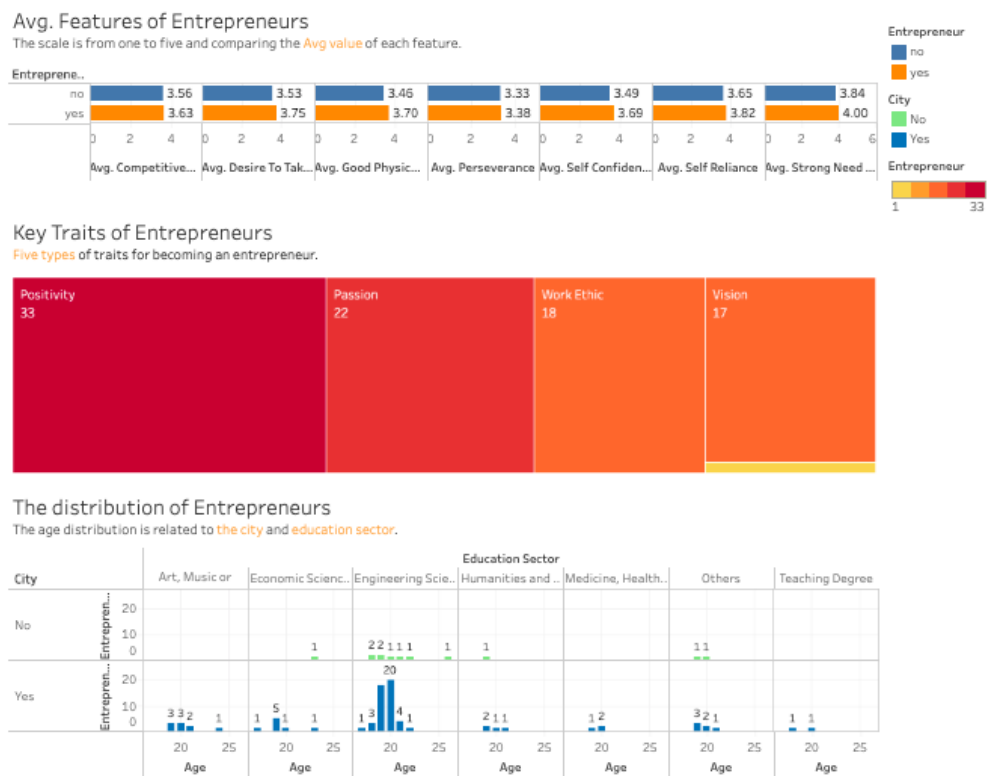
question, we talk about how staying in a city affects entrepreneurship success. so there is another filter of whether staying in a city or not.

• **Results and conclusions:**

Results: First of all, to answer the first question, after we did our project visualization, we found that lots of different vital features affect whether a student becomes an entrepreneur or not. Students can not become entrepreneurs with only one or two qualities, which means entrepreneurs have to own or build various essential traits. Besides, to answer the second question, there still are a few different elements to influence male entrepreneurs more than female entrepreneurs. Because male students who study engineering science, stay in a city, and be influenced are more than females. Their key traits and other unique characteristics are stronger than females. In the end, for the third question, staying in a city is beneficial for students to become entrepreneurs. It is because they are much more easily influenced by other entrepreneurs and build unique key traits or other qualities in the city.

Conclusion: from this final project, actually, we learned a lot of information from it. In terms of this specific dataset, we understand that most people who are entrepreneurs are majoring in engineering science and have various traits or characteristics like self-confidence, strong need to achieve, positivity or passion etc. Male entrepreneurs are generally more than females and staying in a city is advantageous for being entrepreneurial. Moreover, knowing about how to apply the knowledge that we learned from the class in a real world scenario is so helpful and practical for us. From the analytic view, we realized that data visualization is a very powerful and efficient approach to analyze multiple data. And we acquired lots of different messages which we did not expect from it, which made the dataset become more valuable and important. Then, in order to answer one research question, using multiple information instead of

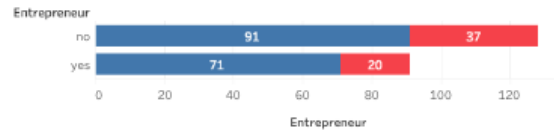
a few of it is necessary. We learned that we have to try different combinations of dissimilar variables many times to do those reasonable and appropriate graphs. Above all, for this project, we knew that data visualization is time-consuming but powerful. And it is vital that we concentrate on and check many details such as color choice, the tooltips or the axis. To be patient and careful is essential for doing it. Of course, we would like to apply data visualization to our future work. It would be so valuable and powerful to analyze diverse datasets.



Gender  
Female  
Male

### The Number of students

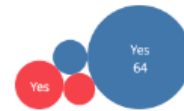
Illustrating the structure of the student sample which is 219 students.



The Number of Entrepreneurs  
Gender of ninety-one entrepreneurs

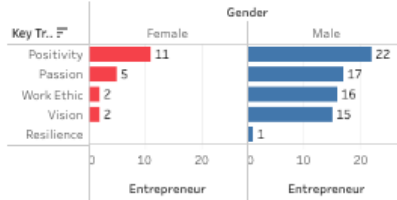


Gender and Influenced



### Gender and Key Traits

Five key traits' distribution from different gender entrepreneurs



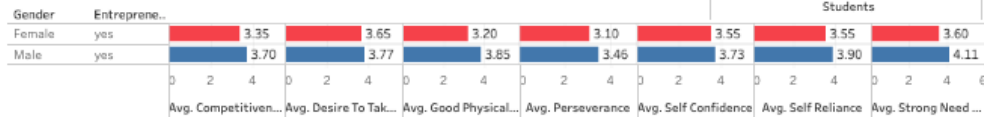
### Reasons for lack

A few reasons why students abandon becoming an entrepreneur.



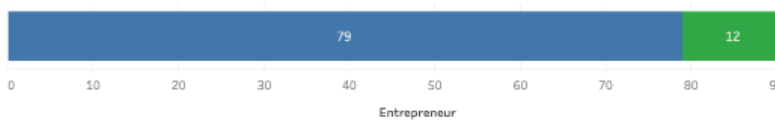
### Avg. Features

Comparing the Avg value of seven features.



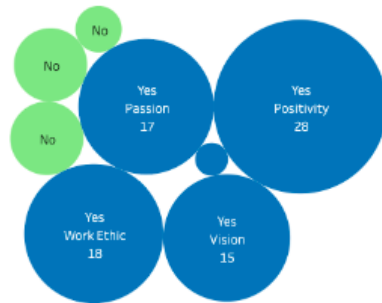
### Influenced entrepreneurs and City

In the city, influenced entrepreneurs are 87.34%. Not in the city, influenced entrepreneurs are 75%.



### City and Key Traits

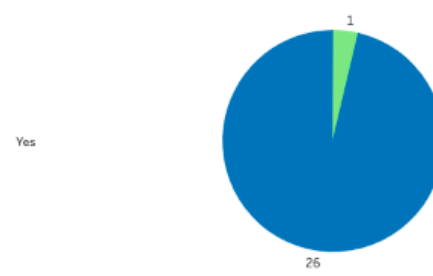
Five key traits from entrepreneurs in the city or not, and correlating gender and avg age.



### Mental Disorder and City

The condition of mental disorder and city, and correlating education sector.

Mental Disorder



### City and Features

Comparing the Avg value of seven features.

