INFO 5709 Data Visualization Final Project

Spring 2020

7th August

Garikipati Vaishnavi

11409444

Table of Contents

* Introduction
* Exploratory Data Analysis
* Hypothesis
* Visualizations
* Conclusion
* References

**Introduction:**

**Tool wise:** Data visualization is used for representing data. The data is described in graphs, charts, and maps in the visualization tool. These visualization tools help in understanding trends and patterns of the data processed. Processing big data like millions of data is not understandable; thus, the visualization tool quickly helps understand the direction of the data. Here for visualization purpose tableau software tool is used (Godfrey et al., 2016)

**Dataset wise:** [Campus recruitment](https://yello.co/blog/campus-recruitment/) is a strategy for sourcing, engaging, and hiring young talent for internship and entry-level positions. College recruiting is typically a tactic for medium- to large-sized companies with high-volume recruiting needs, but can range from small efforts (like working with university career centers to source potential candidates) to large-scale operations (like visiting a wide array of colleges and attending recruiting events throughout the spring and fall semester). Campus recruitment often involves working with university career services centers and attending career fairs to meet in-person with college students and recent graduates. Some industries participate in campus recruiting more than others; finance, technology, business consulting, manufacturing, and engineering are a few of the most popular.

**Exploratory Data Analysis:**

Source for input data: <https://mksaad.wordpress.com/2020/06/30/datasets-for-visualization/>

**Campus Recruitment- Academic and Employability Factors influencing placement**

The data set consists of Placement data of students from anonymous campus in India. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offered to the placed students. The dataset collected has 215 rows and 15 columns. Columns include both quantitative data and qualitative data.

After the data is collected, the cleaning process is to be done. The cleaning process varies from dataset to dataset. So here common steps are being used to structure the data and filtering the data (Su, S. et al., 2016)

The first step includes removing unwanted data like Null values. Duplicate and irreverent data are removed in this step. Then the structure of the data is processed as per requirement.

The model performance depends upon the outliers of the data. Outliers could contain massive data but removing the data can affect the information of the visualization. Handling the data is dependency.

Then, in this .xlsx dataset used for processing, has no missing dataset. No machine learning algorithms have been used for this project. The data set is processed after removing some unwanted data and filtering the data. To perform the EDA here, Tableau software has been used. Outliers (highlighted in red circle) must be identified and removed before performing necessary analysis. The Univariate and Bi-variate-analysis is done using the Tableau software (Setlur et al., 2016)

A screenshot of a computer

Description automatically generated

Univariate analysis:

It is about looking at one variable at a time only. The distribution of the data visualization is done in Histogram and Boxplot. It is a natural process in the software tool by selecting the Analysis pane and after that Aggregate measure.

Bi Variate Analysis:

Here in this analysis involves only two variables, and these are donated as X and Y. the process of the processing of data is the following:

The x variable is to be dragged in the column pane and the Y variable to pull in the column pane (Setlur et al., 2016).

By default, the values would be aggregated into sum or count. Then the aggregate measures from the analysis option to be done, which is in the left pane.

**Hypothesis:**

1. Which factor influenced a candidate in getting placed?
2. Does percentage matter for one to get placed?
3. Which degree specialization is much demanded by corporate?

**Visualizations:**

**1.Highlight tableA picture containing screenshot

Description automatically generated**

**Hypothesis 1:** Which factor influenced a candidate in getting placed?

**Explanation:** From the above highlight tables, it can be inferred that Men followed by women from Central board during Secondary and higher secondary education, graduated in Commerce and Management and having MBA in Marketing and finance without any prior experience are getting placed. Next to them, students from other boards with same degree and MBA are getting placed. Gender did have influence i.e., less than half of men were placed among women.

Students with prior experience also are getting placed but only next to students without having any experience. Sadly, Science and technology students are observed to have scored least placements from that campus.

Hence, it can be said that the most influencing factors for placements would be

1. Male students
2. Central board during Secondary and Higher Secondary Education
3. Commerce & Management field in degree
4. Marketing and Finance in MBA
5. No experience needed/ Freshers

**2.Side-by-Side Bar Graph**

A picture containing fence

Description automatically generated

**Hypothesis2:** Does percentage matter for one to get placed?

**Explanation:** Percentage clearly has high influence on students getting recruited. The above bar chart shows details of average percentage in MBA, Degree, Higher Secondary & Secondary Education of students who got placed and who did not get placed. It can be observed that **Placed students have Higher Percentage** compared to those who did not get placed. Stats give better idea about influence of percentages. A screenshot of a cell phone

Description automatically generated

**3. Line Graph**

A picture containing map, text, boat, person

Description automatically generated

**Hypothesis3: Which degree specialization is much demanded by corporate?**

**Explanation:** From the line graph below it can comprehended that corporate bodies are much interested in recruiting students **graduated with Commerce and Marketing and specialized in Marketing and Finance**. Highest number of students recruited from that combination is 68 and least is 2 from other fields of graduation. Science and Technology seems to have managed an average placement number of 25.

**Conclusion:**

Placements are quiet an important aspect for students whether post-graduates or graduates. Recruits concentrate on different aspects while hiring students and by this project I have analyzed some important factors that interest recruiters by studying parameters that effect the placement rate. Visualizations are presented to strength my argument. The dataset under consideration is dedicated only to a specific college in India and hence we cannot comprehend the results as applicable for any region but only to that college/campus. From the analysis I came to conclusion that male students with MBA in Marketing and Finance with average percentage about 70% and Central board during higher and secondary education have secured more campus placements. Relatively girls from same sector of education have secured placements next to boys. Experience did not help much as freshers are recruited more than experienced students. Science and Technology students did not do well compared to management and other fields of specialization. Therefore, students need not worry whether they have prior experience or not as long as they maintain good percentage.

**References:**

Butcher, J.C., Hill, A.T. and Norton, T.J.T., 2016. Symmetric general linear methods. *BIT Numerical Mathematics*, *56*(4), pp.1189-1212.

Costa, D. and Martins, M.A., 2016, June. A tableau system for quasi-hybrid logic. In *International Joint Conference on Automated Reasoning* (pp. 435-451). Springer, Cham.

Godfrey, P., Gryz, J. and Lasek, P., 2016. Interactive visualization of large data sets. *IEEE Transactions on Knowledge and Data Engineering*, *28*(8), pp.2142-2157.

Hamersky, S., 2016. Tableau desktop. *Mathematics and Computer Education*, *50*(2), p.148.

Setlur, V., Battersby, S.E., Tory, M., Gossweiler, R. and Chang, A.X., 2016, October. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 365-377).

Setlur, V., Battersby, S.E., Tory, M., Gossweiler, R. and Chang, A.X., 2016, October. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 365-377).

Su, S., Perry, V., Cantner, N., Kobayashi, D. and Leigh, J., 2016, October. High-resolution interactive and collaborative data visualization framework for large-scale data analysis. In *2016 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 275-280). IEEE.