Group Coursework Submission Form

Specialist Masters Programme

Please list all names of group members: (Surname, first name) 1. Badhe, Tanaya 2. Bhalla, Kartikey 3. Chawla, Sarthak	4. Gola, Shashank 5. Joshi, Rutva **GROUP NUMBER: 11**
MSc in: Business Analytics	
Module Code: SMM635	
Module Title: Data Visualization	
Lecturer: Prof. Simone Santoni	**Submission Date:** 14/11/2022

Declaration:

By submitting this work, we declare that this work is entirely our own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work we acknowledge that we have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. We also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission:

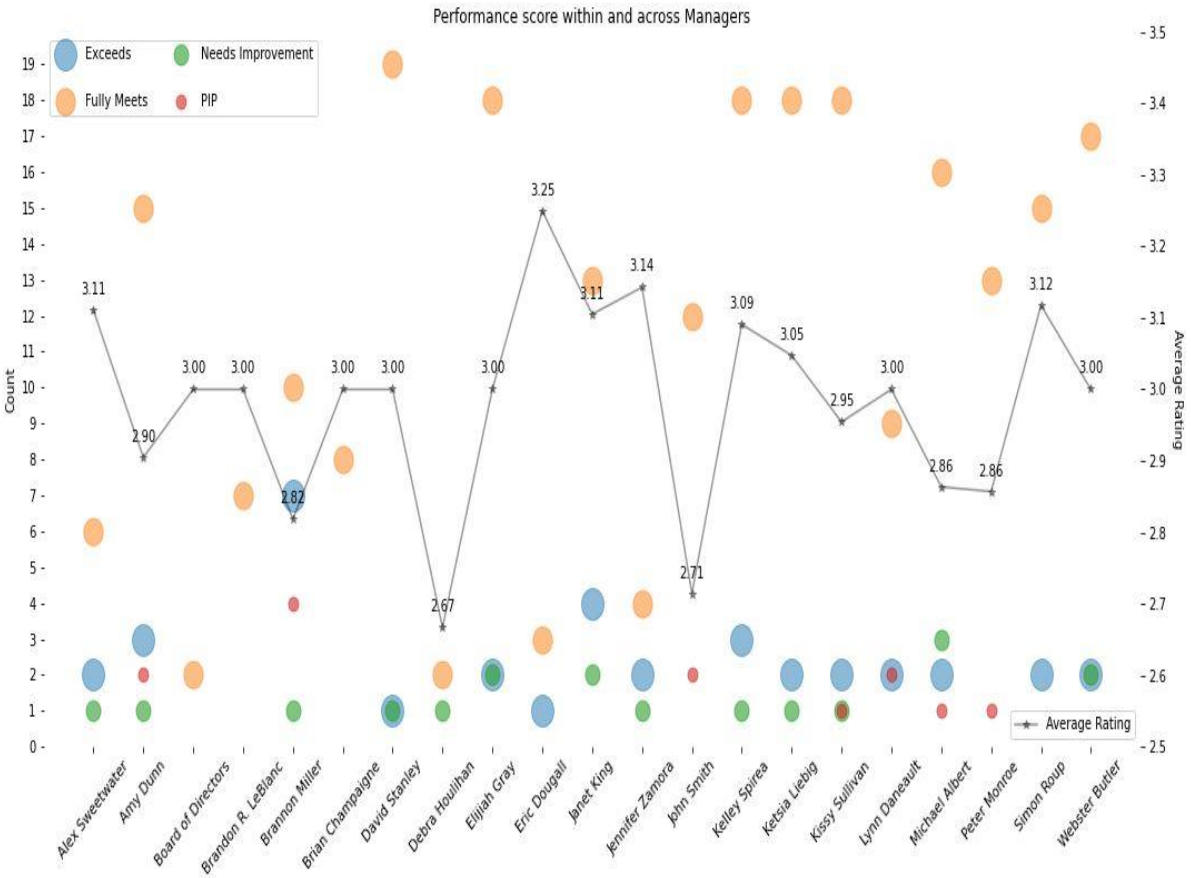
Final Mark: %

SMM635
Data
Visualization

Mid-Term Project

Visualization #1

How do employees' performance scores change within and across managers?



What are the key design features of your visualization?

The plot is a scatter-line plot where X-axis = "Manager Name"; the twin Y-axis has been equipped where the y (left) axis = count, y (right) axis = average ratings. Representation of performance metrics is divided into 4 categories: named exceeds, fully meet, needs improvement, and pip. Assorted colours have been assigned to the designated performance metric. The trend line signifies the change of average rating across the managers whereas the data points of the scatter plot represent the count of the performance scores for employees within the manager. The Star symbol has been used on the trend line.

Why did you choose the above-mentioned design features?

Scatter plot has been used as it reports the value of individual data points as well as the pattern among the complete data - line plot has been used for depicting the change in trend over the average performance across the managers. The scatter-line plot gave the flexibility of representing the performance within and across the managers in a single plot making the process of analysing and user interaction simplified. The twin axis approach was used for efficient illustration of the information in the limited space for discovering and comparing dual trends for improved client interactions. A Star marker was used in the line plot for tranquil identification and to circumvent the overlapping of data points with the scatter plot. Alpha attributes were used for the depiction of overlapping data points within the scatter plot for better user readability.

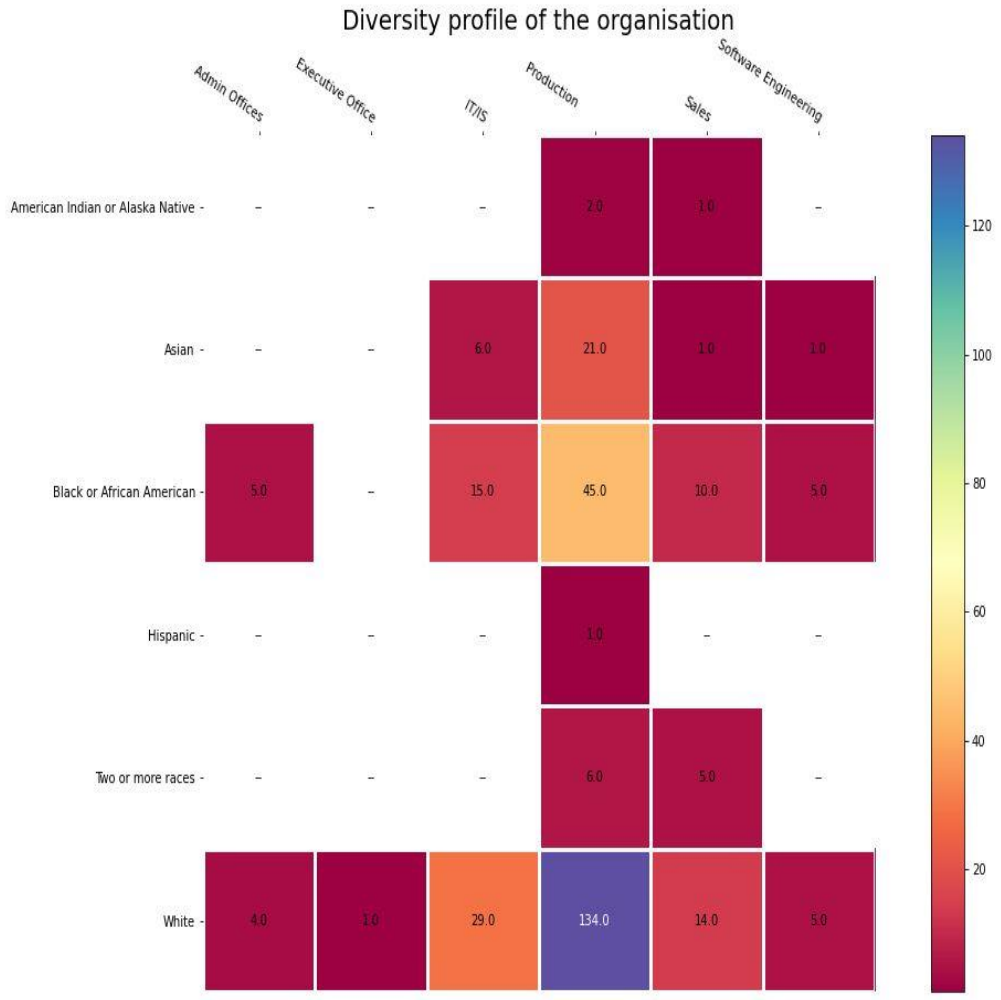
What is the main insight of the chart?

Insights are as follows:

1. Highest average performance is under manager Eric Dougall - 3.25. However, number of employees working under Eric Dougall are limited to four.
2. Lowest average performance is under the manager Debra Houlihan - 2.66.
3. Maximum variability in performance is under the manager Brannon Miller. Maximum number of exceptional performing employees (exceeds) has been recorded under Brannon Miller. In addition, employees under performance improvement plan extends to four.

Visualization #2

What is the overall diversity profile of the organization?



What are the key design features of your visualization?

Heat-map has been plotted to represent the overall diversity profile of the organization across the two axis variables 'Department' and 'Race description'. The legend has been used to depict the spectrum of counts and their relation to colour. Each cell depicts the count of employees of a particular race in the respective department. Spectral colour coding has been implemented.

Why did you choose the above-mentioned design features?

Heat-map was one the most suitable visualization for mapping the diversity distribution across the departments in the organization instead of being soiled into separate bars making it difficult to understand for the users/clients. The density of the individual race across the departments takes the form of colour shapes. Higher the count of a particular race in a department denser the colour block. Spectral colour has been equipped keeping in mind the readability as shades of red is one the easiest to identify and correlate. Moreover, the count has been assigned to the cells for a better understanding of the statistical inference of the data.

What is the main insight of the chart?

Insights are as follows:

1. The production department is the most diverse part of the organization.
2. The Admin office is the least diverse among all departments.
3. White employees have traces in all the departments of the organization. Followed by the black or African Americans.
4. Hispanic / American Indian or Alaskan Native / two or more races follow a similar trend of being distributed over a single or at most 2 departments.
5. Asians are nominally distributed within the organization.

Visualization #3

What are our best recruiting sources if we want to ensure a diverse organization?

What are the key design features of your visualization?

The Donut plot has been plotted to scrutinise the distribution of the recruitment process to visualise the proportions of the diversity factor among the various recruitment sources of the organization. Individual donuts are allocated to each of the respective recruitment sources taking the number to nine. Every race has been allocated a colour code. Legend has been described as colour coding for an individual race. The arc segment inside a donut chart signifies the racial background of the employee. The count of each arc segment has been annotated.

Why did you choose the above-mentioned design features?

The Donut plot is effective in visualising the relative proportions of the data and is a powerful tool for the comparison of categories among the data. Auditing the share of races among each recruitment source is important to signify the recruitment sources for improving the diversity of the organization. Colour codes for each race (arc segment) have been used for better readability and understanding of the data. The count has been annotated for the effective identification of statistical distribution in each arch segment. Nine donuts have been plotted for getting extensive and exploratory information in a single plot for pinpointing the hiring trend.

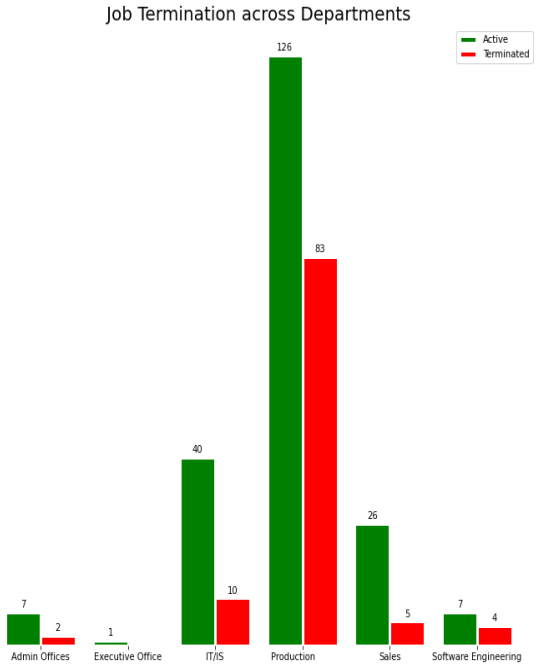
What is the main insight of the chart?

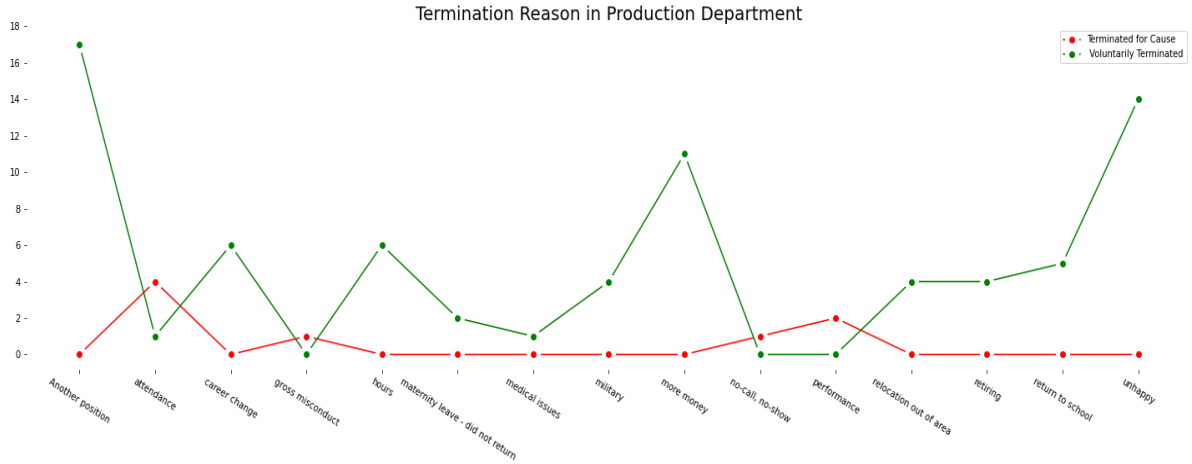
Insights are as follows:

1. Indeed and LinkedIn, as a recruiting source, has the most racial diversity.
2. In distinction, Black or African Americans dominate the Diversity job fair accounting for 100%.
3. 4 of every 10 candidates hired from Indeed and LinkedIn are non-white.
4. LinkedIn, Indeed, and Google search account for 61% of the total non-whites recruited in the organization.

Visualization #4

What are the most frequent factors for job termination?





What are the key design features of your visualization?

Grouped Bar chart (Job Termination across Departments) and multiple line plot (Term reasons for Job Termination in Production Department) have been used and aligned equipping the GridSpec function. Grouped bar plot has been plotted where the x-axis = Department and y-axis = count. Legend has been described to categorise active and terminated employees using colour scheme. The count of employees has been annotated on the bar.
Multiple line plot has been used where the x-axis = termination reason and y-axis = count (Terminated employees). Legend has been described as diversifying 'Terminated for cause' and 'Voluntarily Terminated' taking in the use of a colour scheme. Point markers were used with a dash-dot line style.

Why did you choose the above-mentioned design features?

Grouped bar and multiple line plots have been plotted for drawing the comparison between the two data series and information depiction utilising the space and flexibility to represent the trend and statistical inference. GridSpec has been utilised due to its control and pliability to place the subplots. Colour coding in the bar and the line chart is used for differentiating, classifying, and efficiently analysing the trend of the 2 data series along the axis. Dash-dot line style is used in the line plot for improvised visualization and better visibility of the data points. Annotations in the bar chart illustrate the statistical inference for optimised recognition of trends and efficient readability.

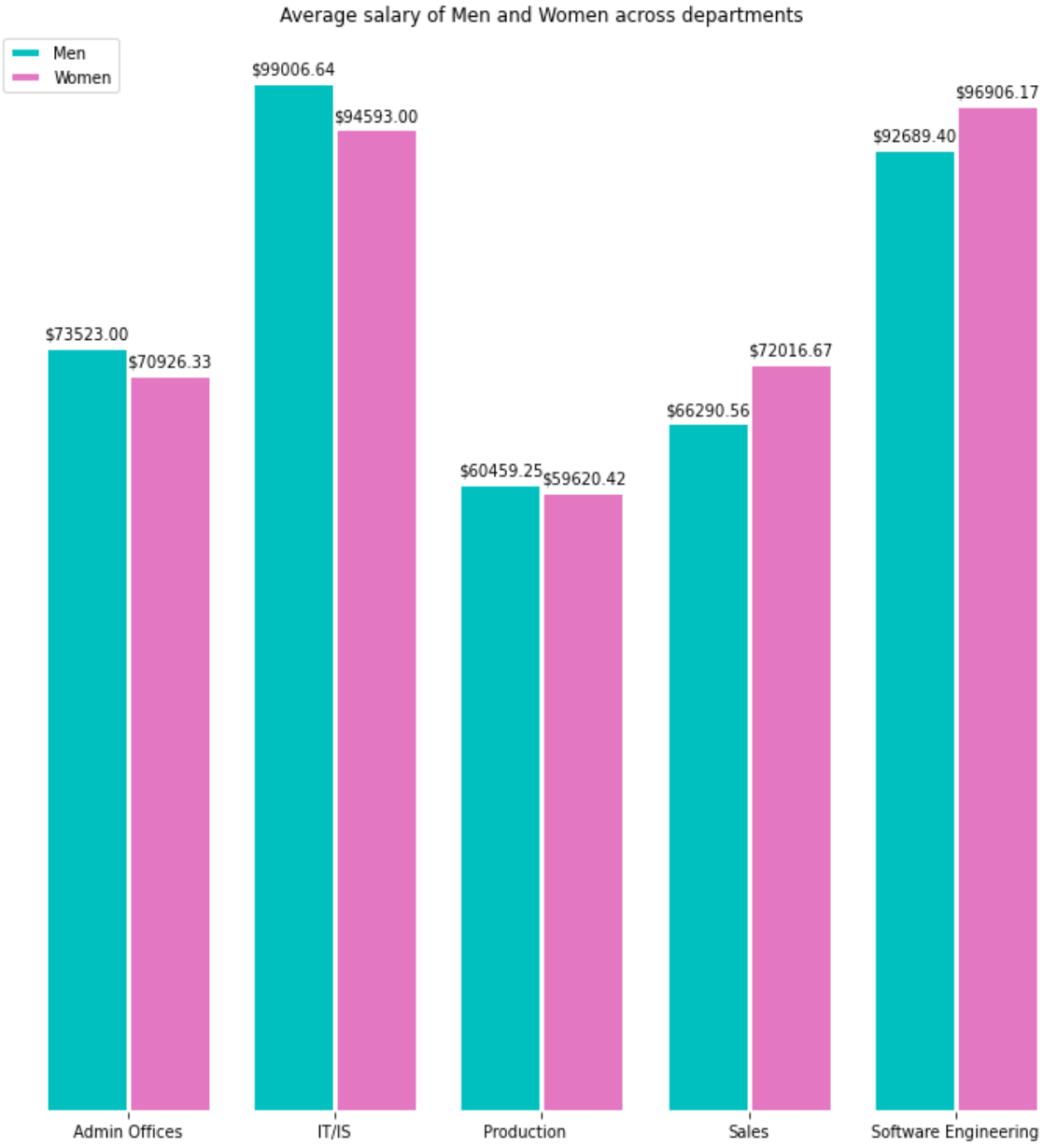
What is the main insight of the chart?

Insights are as follows:

1. Across the departments, production experienced maximum termination accounting for 79% of the total termination in the organization.
2. 83 employees were terminated out of the total 209 in the production department.
3. Another position is one of the most frequent factors for employees leaving the company voluntarily. Followed by 'more money and 'unhappy'.
4. Attendance and Performance are the most frequent factor for 'termination for cause.

Visualization #5

Are there areas of the company where pay is not equitable?



What are the key design features of your visualization?

Multiple bar chart has been plotted where X-axis = Department, and Y-axis = Salary in $. Distributed among 2 series of Male and Female. Legend has been defined to categorise males and females with different colours. The average salary has been annotated on each bar. The height of the plot has been adjusted to maximise the visual accuracy of the bar plot.

Why did you choose the above-mentioned design features?

Multiple bar chart is extremely helpful to represent abundant information in a small space making it easy to understand for the users/clients. Moreover, it is efficient in visualizing multiple data series (categorical) on the same axis and grouped into parent categories. Different colours for males and females are allocated for easy identification and interaction. The average salary over the top of the bar would help in recording the statistical readings for optimised analysis. Longer height in comparison to the width would help users correlate with the salary in terms of the varying height of the bars.

What is the main insight of the chart?

Insights are as follows:

1. Across the departments' the salary in the production department is way lesser.
2. Females in production on average are paid 29% lesser than the other departments whereas males in production on average get 27% lesser.
3. The average salary of a male in the organisation is $78393.
4. The average salary of females in the organisation is $78812.
5. Sales department have maximum variability in the average salaries among male and female .i.e $5726.11 .