Data Analysis Report

PAY RATE ANALYSIS

TONY WONG, MATTHEW RODRIGUEZ, DILPREET SINGH, ZACHARY NABAVIAN
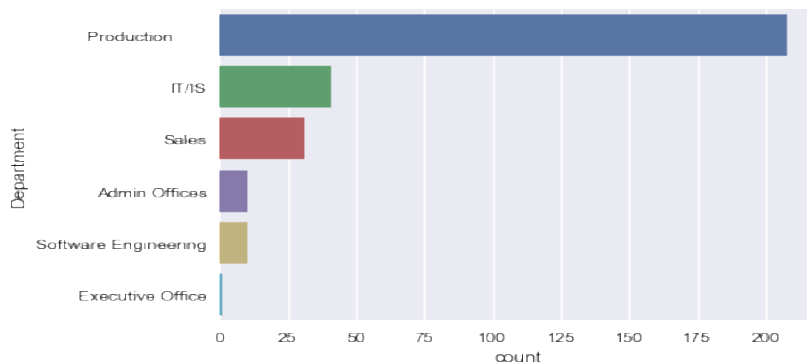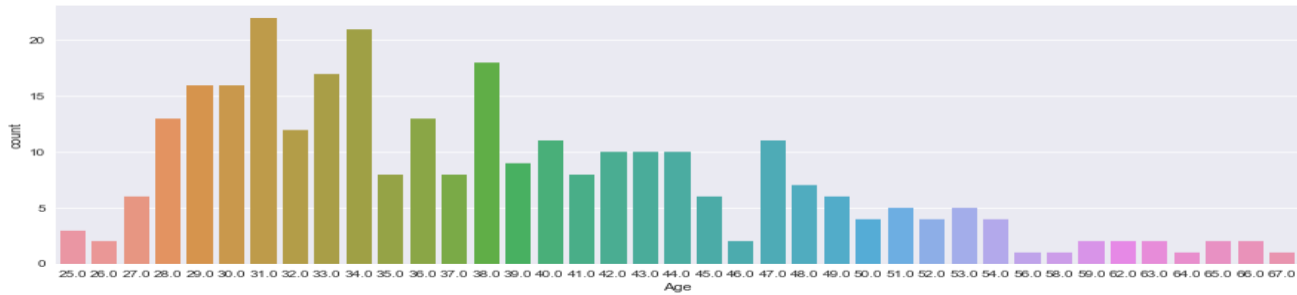
INTRODUCTION

The goal of the analysis was to describe various aspects of the employment staff at Dental Magic. We gathered data on the employees from Kaggle as well as information on all employees' genders, age, race, occupation, employment status (employed currently versus unemployed) and then compared the pay rates of different groups of employees. We find that the vast majority of employees, over 200, are in the production occupation. Conversely, less than 20 are in software engineering. Most employees are in their 30's with an extreme minority in their 20's and an even smaller amount in their 50's and older. The majority of employees are female. We displayed all of our findings in the form of charts and graphs that can be found in our Jupyter Notebook.

Some of the most interesting results include box plots showing pay rates for different occupations, box plots comparing wage rates of different faces, box plots comparing the wage rates of different genders, and box plots comparing citizenship status to pay rate. In addition, we have a regression plot comparing age to pay rate. We find that at Dental Magic, occupation is the most reliable predictor of pay rate. Predictably, the President and CEO have the highest pay rate. On average, the Production Technician 1 has the lowest pay rate. The median for the IT staff lies in the middle but interestingly has an extensive range. Among races, Hispanics appear to have the highest median pay rate while the other races seem to have roughly equal medians. We find no discernable wage gap between genders. There is an extremely weak positive correlation between age and pay rate. Citizenship appears to have no significant effect on pay rate, but US citizens appear to have a slightly lower average pay rate than Non-Citizens. This difference seems statistically in-significant.

METHODOLOGY

We wanted to do a data analysis on the employment conditions in a certain company. The core questions we wanted to tackle included descriptive statistics about the demographics of the employees at the fictional company Dental Magic such as race, age distribution, and gender distribution. We then sought to determine which of the above categories served as the best predictor of employees' pay rates. We downloaded our data source from Kaggle. The source took the form of four CSV files, and we ended up using the "core-dataset" for our analysis. Since all of the data was already in a CSV file, web-scraping was not necessary to clean up the data. In our code, we have a section where we display the first five rows of the dataset to obtain an understanding of what our data looks like and what information we have. Our analysis consisted of displaying various tables and graphs generated from the data using pandas and seaborn. At one point in our coding, we needed to convert categorical variables into numerical variables so that we could plot them on the x-axis of a graph.
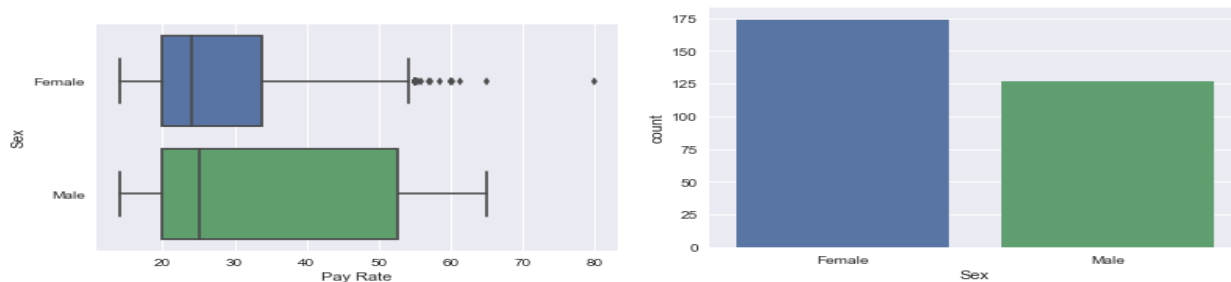
DESCRIPTIVE STATISTICS

The top graph illustrates the age of each employee. The bottom graph shows number of employees within each department.

The vast majority of employees have careers in production. Predictably, only a few employees are in the Executive Office.

GENDER VS PAY RATE

Next, we plotted a few graphs that illustrate the age distribution and gender distribution. A majority of employees are female. Most of our data analysis consists of a construction of box plots show pay rates compared with different demographics. The spread of females across various pay rates and job roles are lesser as compared to males, but there are individual outliers in the higher pay rate region which shows that there's some but not an adequate representation of females in the more senior management roles. We find no wage gap between genders regarding median wage gap. The company CEO is a female, which is why the highest pay rate rests on the female side of the graph.
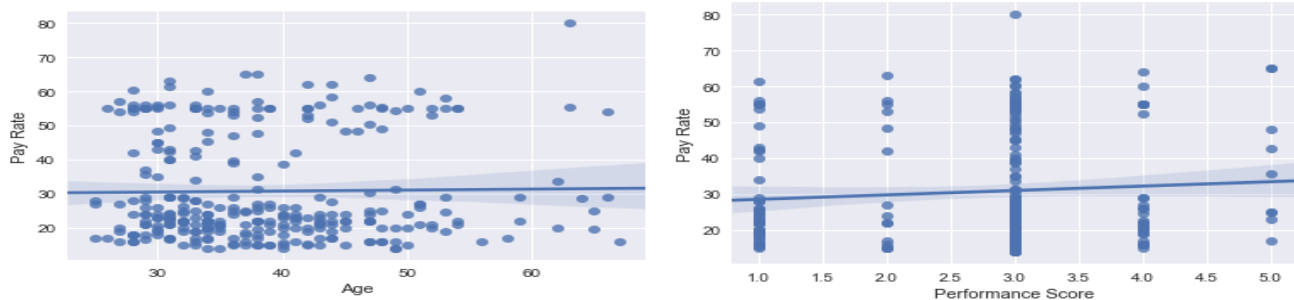
We have two regression plots that compare age and "performance score" to pay rates. Both show a very weak positive correlation with pay rates. In the case of performance score, we mapped five qualitative categories from the core dataset to numerical values. Neither of these metrics, age and performance score, serves as a good predictor of pay rate.

Additionally, the company mostly accommodates middle-aged people, there are some outliers of age between 60 and 70, but this doesn't affect their pay rate. The company doesn't discriminate by giving lower wages to elderly people which is a good HR policy sign.

POSITION VS PAY RATE
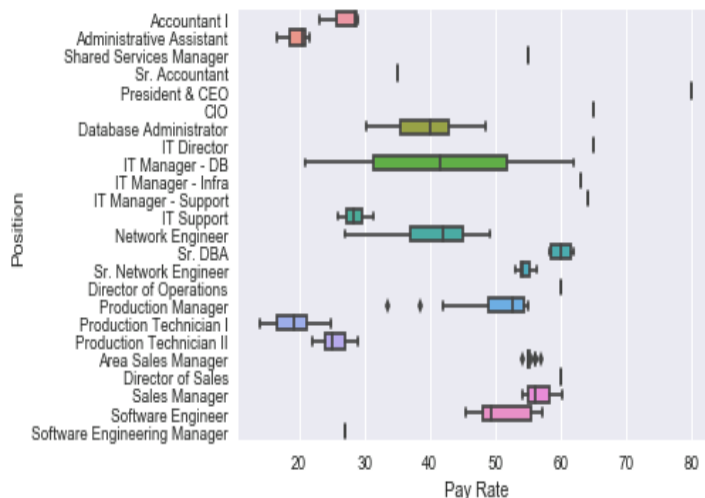
The data suggests that an employee's occupation serves as the best predictor of pay rates. From the box plot, the difference in median pay rate is more discernable across all of the different occupations than in any box plot previously.

The position of IT Manager in Database has the most significant spread of pay rate which spans from $20 to over $60 and holds a uniform distribution. There is either consistent pay hike policy for this role or the work in this role is required at different levels of uniformly increasing pay rates. Only single openings are available for some senior positions like Shared services manager, Senior Accountant, President and CEO, CIO, IT Director, IT Managers, Director of sales and software engineering manager. Area sales manager of different areas get almost the same pay; the slight difference might be due to the performance bonus that is given on meeting targets. The boxplot for accountants is left-skewed to a great extent which shows that lower 50% employees are spread across a more extensive pay rate range than the top 50% employees who are getting more or less the same pay rate at the higher end.

RACE VS PAY RATE
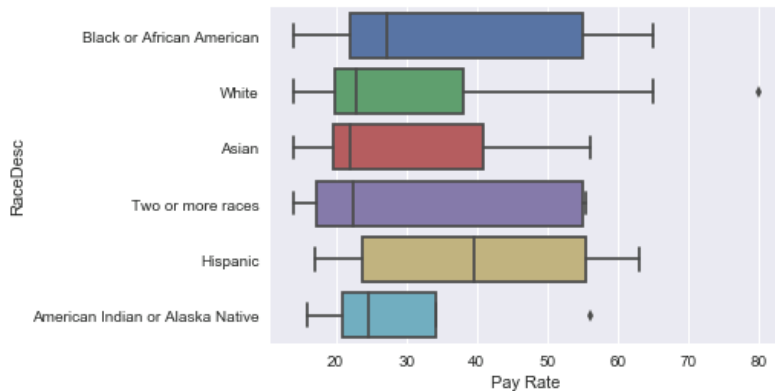
American Indian or Alaska Native has the least diversity in pay-rate which is also skewed towards the left. They all are a getting similar pay-rate. They are hired on the roles of product technicians, accountants and administrative assistants only. White, Asian, two or more race & American Indian have a very similar median (which are near to mean in the absence of large outliers), but their distributions are quite different. The top paid person who is the President and CEO is a White. Hispanic has highest median pay role and almost evenly distributed among all pay rate; they are more or less equally distributed among the different roles and pay-rates.

CITIZEN VS PAY RATE
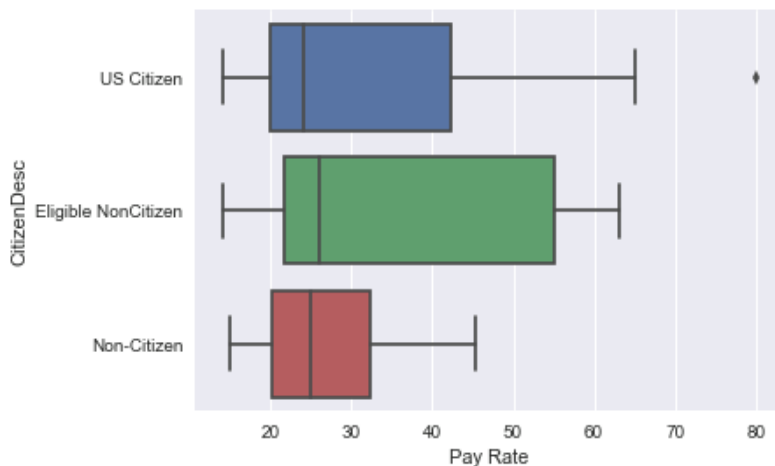
The median salaries of all the citizenship categories are almost the same. A small number of eligible American citizens are spread across the different paying positions. They are hired for multiple varied tasks. Non-US citizens are never hired for some senior positions like Sr. DBA, Sr. Network Engineer, Sales Manager, and software engineers as well. CIO, IT Director, President, and CEO always come from US citizens. The boxplot for US citizens is skewed towards the right which shows that 50% of the top paid US citizens are spread across a big range of pay rates.

EMPLOYMENT STATUS VS PAYRATE
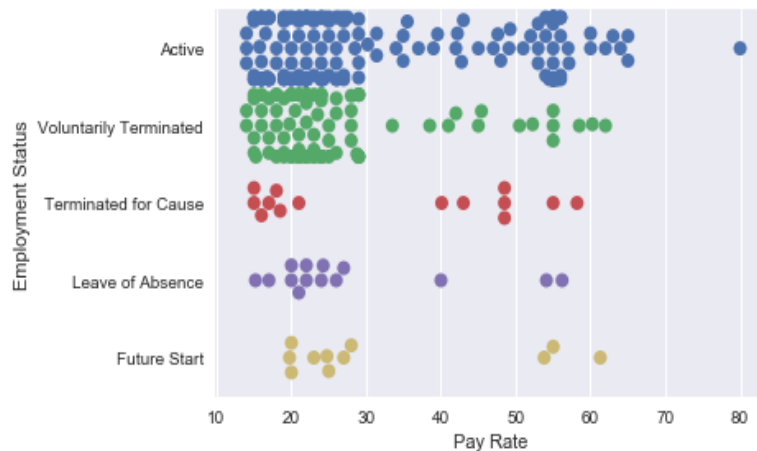
Both active and voluntarily terminated employees have the same pay rate distribution, high concentration in the low pay rate region as more employees are required for positions like product technicians and account-ants. Senior management people next to President and CEO are paid high and usually voluntarily terminated, there are much lesser layoffs in that category. The employees that are terminated for a cause are either from very junior positions or positions like software engineers and IT managers. The future start employees are concentrated either on less paying jobs or high paying ones. There's a gap in the average pay rate range which shows that hiring takes place either at the lower level of positions or the higher ones. Employees come to the middle region through promotions or hikes.

SOURCE

Human Resources Data set was obtained from Kaggle.
Link: **https://www.kaggle.com/rhuebner/human-resources-data-set/data**