Assignment #2 US Patent Office Analysis

Group 4:

Will Stephenson	261082056
Hugo Garcia	260791363
Liliana Tretyakova	261086215
Emery Dittmer	260658030

1.1. US Patent Office

The United States Patent Office (USPTO) USPTO advises the president of the United States, the secretary of commerce, and U.S. government agencies on intellectual property (IP) policy, protection, and enforcement; and promotes the stronger and more effective IP protection around the world according to their website.

The USPTO is a large employer of patent examiners. In this basic study we are examining the human metrics of these patent employees. This includes turnover rate, mobility within the company and more. To accomplish this we are using a dataset provided by the USPTO (simplified by our instructor). Additional details can be found here and the datasets here.

1.2. Data Exploration

To explore the data we first looked into the NAs for the first name. As we knew that both the race and the gender packages were based on name. We wanted to also make sure that the data made sense for first names but as the dataset was too large to validate manually we looked at first names at somewhat random intervals to ensure that they were entered in a consistent format.

Further exploration was required for the patent process. To determine our measure for marking a patent for the clerk per month we needed to determine which date to use. As the filing system didn't represent the actual work of the clerks, we looked to the filing date. We also took into consideration dates that the patent was abandoned. Unfortunately, when exploring these dates we did find some nonsense dates, such as the year 2050, 2900, or 3000. We believe that this could likely be attributed to human error and we discussed how we could replace this data with our best guesses informed by the surrounding data (ie for the data that was in the year 3000, would the date make sense based on the application or status change date and if the 3 was a 2?) then could we change it to our guess?

We didn't land on consensus for this and opted to remove them instead, but this line of thought did go into our exploration discussion. Once we had gone through the dates and names, we felt more comfortable working with this dataset and importantly, felt that we understood what information would be most valuable for the task.

1.3. Data aggregation

Once done with the data exploration we proceeded with data aggregation. Focusing the attention on biometric data by examiner and quarter we aggregated all data based on examiner quarterly data. This data included the number of applications of each type per quarter (i.e. issued, pending, abandoned). The current art unit for each examiner, the number of people, women and examiners by ethnicity.

2.1. Analysis

We evaluated basic metrics for the data set. We found that the examiners had an average mobility of 2.4 art units throughout their careers. This would indicate that on average, each examiner goes through 2.4 art units and is highly mobile throughout their careers. This number has limitations as the movement back to a previous art unit was not counted as 'mobility' and the annual average might give more insights and a comparable metric for benchmarking.

Furthermore, we found that the turnover was on average 24% from 2000-2016. This would indicate that 24% of their staff left the company between these years. There were no direct observations or indicators of whether or not an examiner had left. Therefore, These numbers were gathered from the transactional data, where we investigated the last quarter where activity was logged. We assumed that the ratio of pre-2017 and 2017 activity was a good stand-in metric for turnover. A lack of activity in 2017 indicates turnover whereas activity in 2017 indicates employment.

2.2. Insights – Graphs

Further to these key metrics we investigated basic biometric data for all the patent examiners.

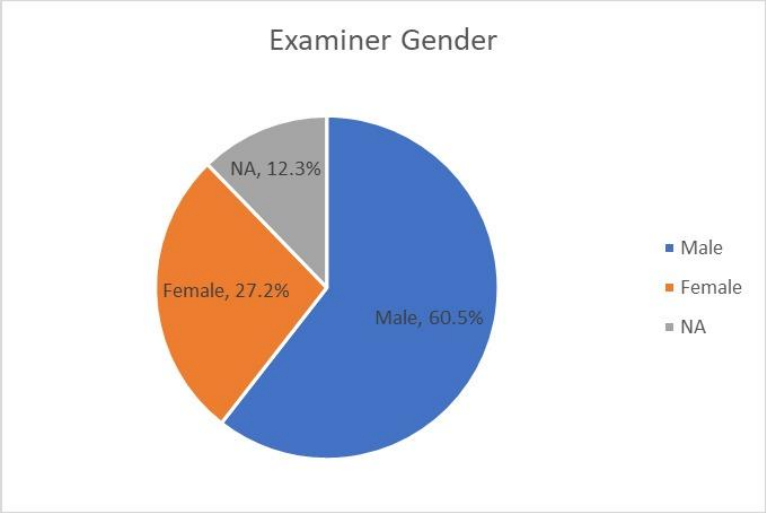

Figure 1: Examiner Gender distribution across all examiners

Most examiners are male, indicating an under-representation of women in patent examination.

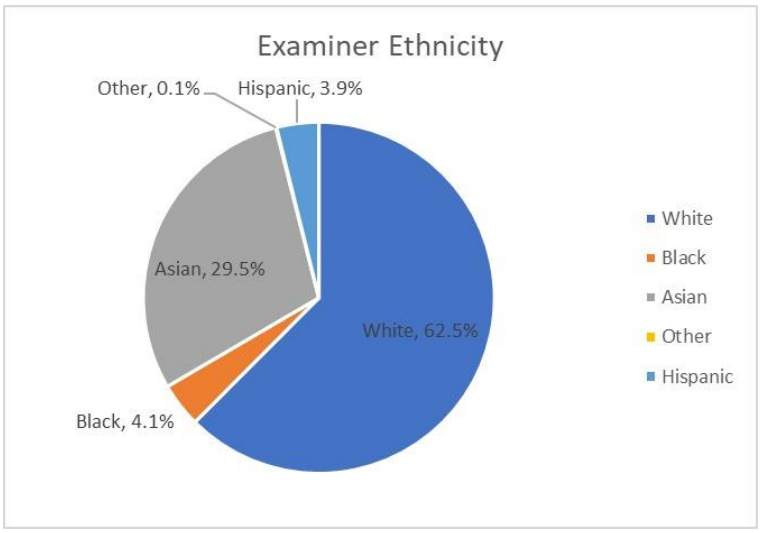

Figure 2: Pie Chart of Examiner ethnicity for all examiners in dataset

Most examiners tend to be white or Asian. Black and Hispanic people make up only a small proportion of all examiners.

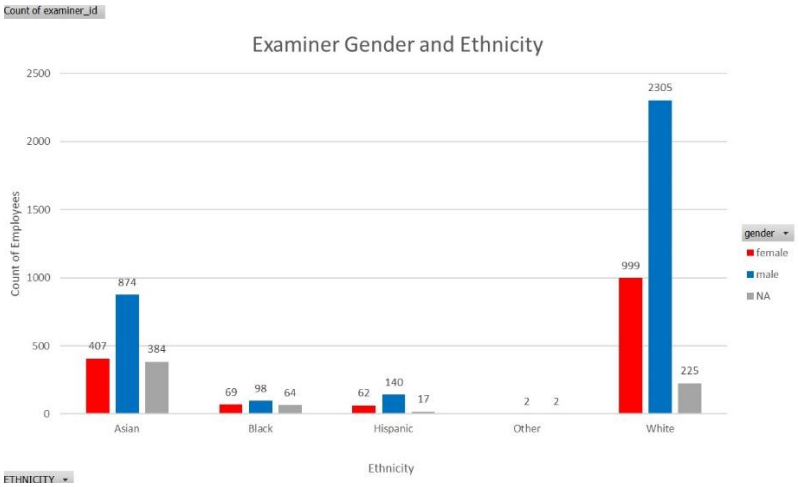

Figure 3: Patent Examiner by Gender and Ethnicity.

All ethnicity metrics indicate a predominant male distribution. The largest groups are white men and women while the groups with the lowest populations are Black and Hispanic women.

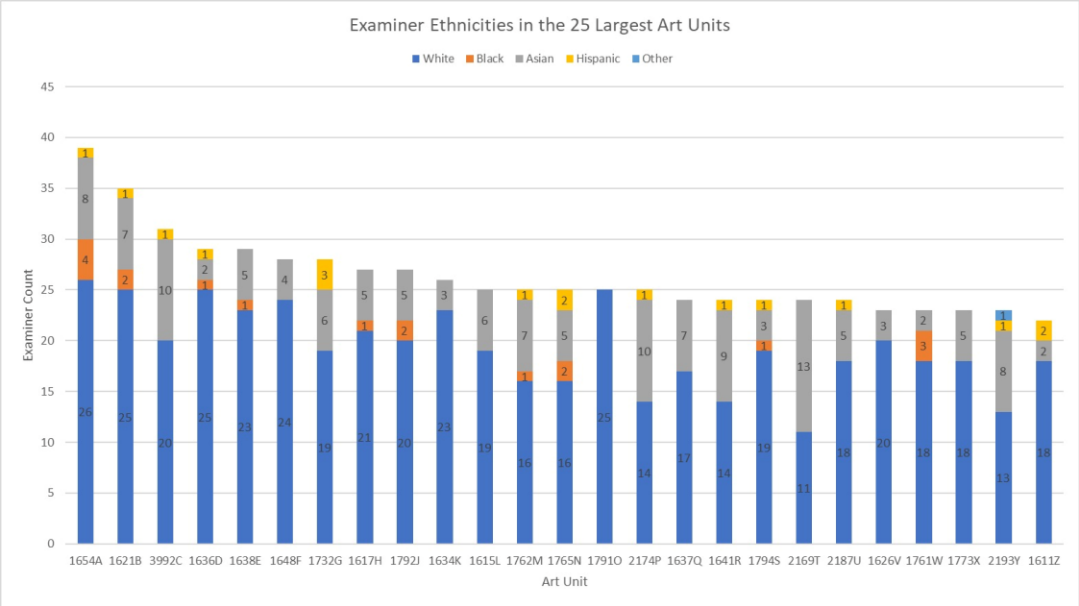

Figure 4: Number of Examiners per art unit and per ethnicity