BURNING GLASS DATA EXERCISE

Katherine Yu

APPROACHING “ANALYTICAL”

After looking through the data, abilities, technology skills, and tasks seemed to be the most relevant in answering the research question

- Abilities (occupation level) vs Tasks (task level)
  - Decided to approach using data related to Tasks as opposed to Abilities
  - Abilities was broken down into broader categories than Tasks
  - Tasks are more detailed and relates to the duties of the actual job itself the most, giving a more accurate indicator of how analytical the job duties are
  - We do have to assume that the majority of are captured by the different Tasks listed for each occupation; this is not a big of a problem for Abilities since the ability categories are the same for all occupations and are rather broad
- Technology Skills vs Tasks
  - The Technology Skills dataset doesn’t have data values and technologies can have many different uses (i.e a data engineer creating an Access database vs an assistant entering data)

After comparing, chose to study Tasks for approaching “analytical”.

SORTING AND CALCULATING "ANALYTICAL"

SORTING/CATEGORIZING

Decide which tasks can be categorized as analytical

- First, establish what analytical means
  - Critical thinking, being able to extract information out of data
- Second, create a list of keywords that encapsulate what analytical means and mark if they appear in either the task descriptions or the tasks' DWA
  - ```
    'analyze',
    'research',
    'evaluate',
    'recommend',
    'investigate',
    'advise',
    'consult',
    'forecast',
    'statistic'
    'analyses'
    ```

CALCULATING

Apply the following formula (after scaling Importance to the same scale as Relevance) to get a Task Score only **if** the analytical flag is greater than 0:

Task Score = √(Importance*Relevance)

Task Score falls between 0 and 100. The higher the score the more analytical the task is.

Next, we aggregate at the occupation level with the following formula:

Occupation Score = $\Sigma TS_i / (i \times 100)$
Where i = the number of tasks listed for each occupation.

In order to merge with the gender data, we do a simple average for each 6 digit code.
    ```

DISCUSSING WEAKNESSES, STRENGTHS, ASSUMPTIONS

	Weaknesses	**Strengths**	**Assumptions**
Approach to "analytical"/categorizing	Depends heavily on word choice of keywords.. It's also important to choose words that were not too broad; i.e. words that can be used in an un-analytical context as well, but also captured the analytical task. Engineering occupations ended up scoring low for this reason although they are STEM.	Using task ratings addresses specific instances people are performing at their job.	The keywords chosen cover the majority of analytical duties and don't cause bias.
Calculating analytical score for occupations	Without numbers for the sub-occupations, I could only do a simple average for each occupation.	The formula used takes into account the number of tasks. For example, a CEO has as many analytical tasks than an analyst but has more overall tasks, thus the analyst will be more analytical.	Assume that importance and relevance are a good indicator for being "analytical".

LOOKING AT THE DATA

occupationScore	SOCName	% Female (ACS)	TOT_EMP	# Female
67.98001735876030	Sociologists	51.948917564253	2710.0	1407.8156659912600
65.03532699914980	Animal Scientists	36.4767948619925	2530.0	922.8629100084100
64.04118894353530	Biochemists and Biophysicists	48.4413500234551	28500.0	13805.784756684700
62.491828162360800	Survey Researchers	51.948917564253	11690.0	6072.828463261180
57.88739198990670	Economists	33.357528957529	18650.0	6221.179150579160
55.40218789009270	Mathematicians	46.7496218012912	2580.0	1206.1402424733100
54.013219454123000	Hydrologists	31.0543514372561	6290.0	1953.3187054034100
53.72577734943020	Statisticians	46.7496218012912	39920.0	18662.449023075400
53.443358550657500	Credit Analysts	56.039029535865	74820.0	41928.401898734200
53.081453017979900	Political Scientists	51.948917564253	5660.0	2940.30873413672
52.82852862988680	Anthropologists and Archeologists	51.948917564253	6020.0	3127.324837368030
52.37335651709970	Budget Analysts	61.0012842858453	52810.0	32214.778231354900
50.43880600353320	Epidemiologists	52.9731796411208	7060.0	3739.906482663130
49.36697166425070	Anthropology and Archeology Teachers, Postsecondary	51.8247774989852	5890.0	3052.479394690230
49.293843726660500	Chiropractors	29.0215568040806	34740.0	10082.0888337376
46.744327192056500	Market Research Analysts and Marketing Specialists	57.9085305525011	638200.0	369572.241986062
46.186107979046900	Atmospheric and Space Scientists	15.891412811057	9310.0	1479.4905327094100
44.78905147122380	Compensation, Benefits, and Job Analysis Specialists	80.5747668985709	83550.0	67320.21774375600

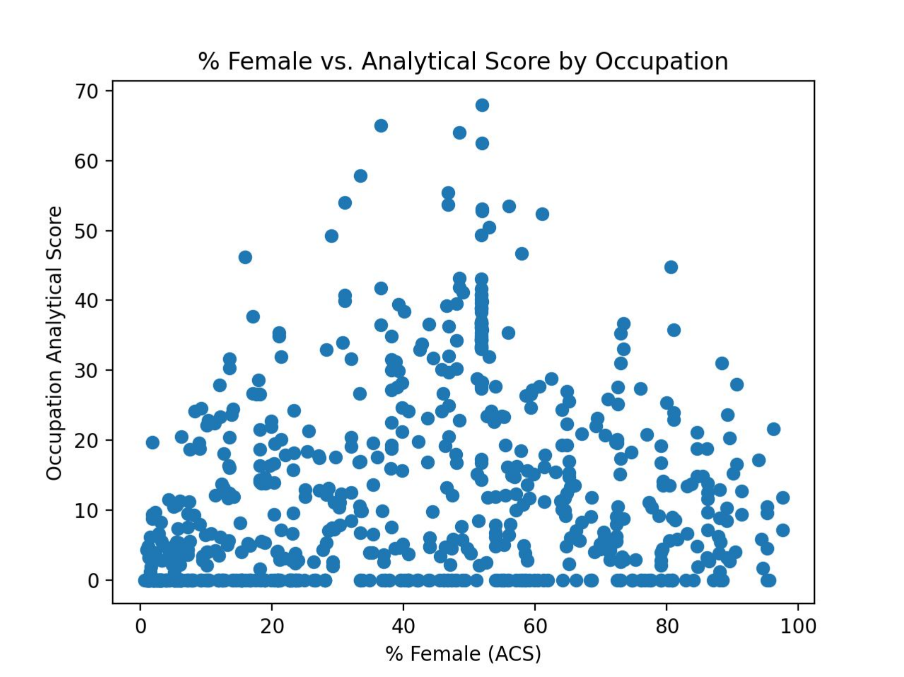
The top 18 highest analytical scores by occupation are shown to the left.

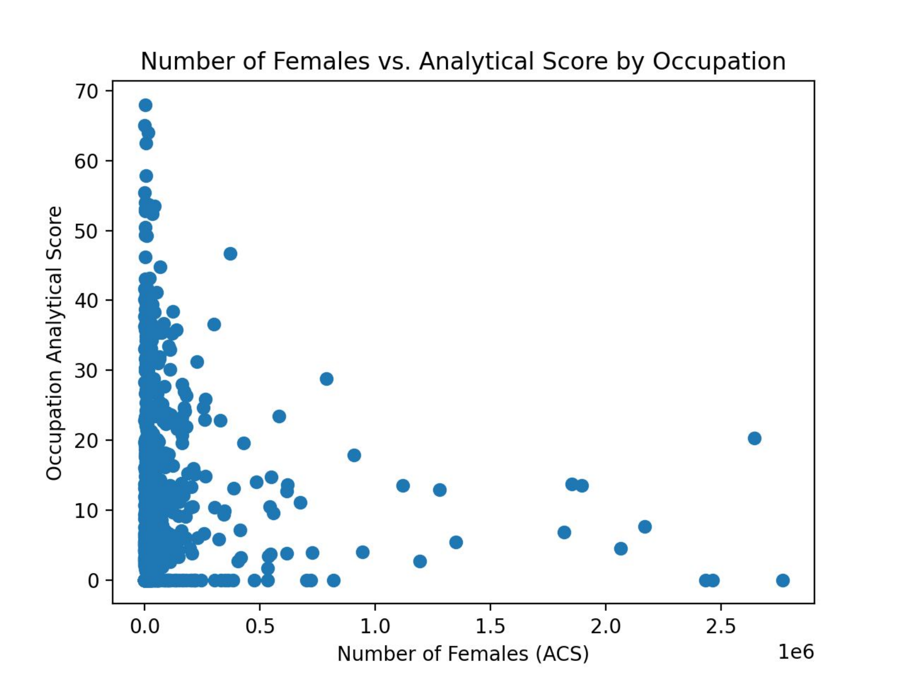
Unsurprisingly, many occupations in the natural sciences and math (STEM) found their way to the top, along with professors.

Chiropractors was rather surprising, and scored high because many of the tasks include language like advising, consulting

Distribution of occupationScore is skewed right.

EVALUATING THE GENDER GAP





EVALUATING THE GENDER GAP

- Overall, the data seems to look pretty normally distributed when comparing % female to analytical score. We can't see an obvious trend that the higher the analytical score, the lower % female.
- The average analytical score (for occupations with at least one task that is analytical) is 17%
- Below the average, we see that there is a pretty even spread between % female and analytical score on the previous graph.
- Above the average, the data is more sparse and holds a bell curve shape.

	% female mean
Occupations with at least 1 analytical task, above mean	46.48
Occupations with at least 1 analytical task, below mean	44.49
All occupations below mean	35.26

- Since the client is looking specifically for women in STEM, this would mean that they are interested in the occupations with higher analytical scores.
  - Among the higher analytical scores, women usually hold a larger percentage of post secondary teaching jobs.
  - Men seem to be the majority in natural science research and engineering jobs.

FINDINGS

1

On an overall level, the amount of females in analytical occupations seems to be normally distributed.

2

However, from the analysis we see that there are indeed certain sectors that are influenced by gender.

Engineering and other natural science positions in research have the biggest gender gap, being overwhelmingly male occupations. On the other hand, teaching jobs tend to be equal or majority women.

3

Occupations that scored above 30% on the analytical score and are below 50% female are going to be of most interest. A 30% analytical score is high enough to avoid any non-STEM occupations that have higher scores. **The positions that stand out the most in this "quadrant" that are highly analytical and have a large gender gap include animal scientists, hydrologists, economists, atmospheric scientists, and various engineering fields.**

When presenting to the client, I would emphasize the bolded sentences. Concentrating efforts to create equality in the mentioned occupations would probably create the most impact.

FURTHER RESEARCH

For defining “analytical”:

- In relation to improving the current method, if there was more time I would have liked to experiment with machine learning (NPL for keywords and clustering for gender analysis) to see if that would give new insights.
- In the O*NET data, the data in the original Task Ratings dataset had a wide range of dates (early 2000’s to present). There may be a discrepancy in the time the Task Rating data vs the time the % Female data was collected, which could result a discrepancy in the relevance of the tasks. Another way to look at the “analytical” would be to directly search through job postings and formulate a way to categorize analytical jobs based on the language of the job posting.

For studying the “analytical” gender gap:

- Wage is certainly an important part of the gender gap in the workplace; it’s usually the case that women have lower salaries. This could be addressed by including wage data from the Bureau of Labor Statistics.
- If more relevant data was collected, it would also be possible to construct a regression analysis of some sort to make the analysis more robust.