# Case Study: Compensation Research Analyst

*Note: This exercise is theoretical and does not reflect an actual business initiative of D. E. Shaw & Co., L.P. This document and its contents are confidential and may not be redistributed or shared.*

## Directions

* You’ll have 72 hours to complete the case study, though it is only intended to take 2-3 hours. When you have completed the assignment, please submit your response to Sarah Jameson ([Sarah.Jameson@deshaw.com](mailto:Sarah.Jameson@deshaw.com)) and Mary Claire Ecclesine ([MaryClaire.Ecclesine@deshaw.com](mailto:MaryClaire.Ecclesine@deshaw.com)).
* R or Python should be used to complete the first three prompts.
* Written responses and any accompanying tables or visualizations may be sent in Word or PDF format.
* Organize files and code with the expectation that teammates may reference them in future analyses.
* Save all outputs in a zip file and begin the file name with the ID listed in the email from your recruiter (e.g., “65873\_case\_study.docx”). This will help us to review cases anonymously.
* Note that you may be asked to discuss the case study with an interviewer.

## Prompts

1. Join candidate\_profile and recruiting\_activity datasets, keeping each row in the recruiting\_activity data.
2. Create a "Recruiting Funnel" view (as shown in the Legend tab) for Highest Degree. You will need to calculate the highest degree for each candidate using the education columns (the table in the Legend tab shows the order of degrees to consider from highest to lowest). The funnels should show the number of unique candidates who moved through each recruiting stage including offer response. Count each candidate only once per Recruiting Funnel; if a candidate applied to multiple positions within a given department, count them once for the application that reached the furthest recruiting stage.

A screenshot of a computer

Description automatically generated with medium confidence

1. Our Head of Recruiting, Cameron, has asked some questions about compensation expectations, as outlined below. When answering these questions, please focus only on the data available in our simulated dataset and write the responses as if you were writing to Cameron, who runs a large department and does not have a quantitative background.
   1. Do applicants with a PhD expect to make more than candidates who don’t have a PhD?
      1. At a high level, it appears that PhD candidates expect to make more than those who don't have a PhD. However, that is largely a result of PhD candidates applying to higher level roles in the first place, skewing their averages. As we do more of an apples to apples comparison (comparing across department and position), we observe no evidence that PhD candidates expect to make more than non-PhD candidates.
   2. Do the compensation expectations for PhD/non-PhD candidates differ across departments?
      1. See above
   3. In addition to the PhD and department, please check to see if there is a relationship between compensation expectations and years of experience and/or current industry.
      1. We observe that on average, candidates with more years of experience expect a higher compensation. More specifically, for every year of experience, we estimate the expected compensation to be higher by ~3%. This means that an employee with 10 years of experience will have an expected salary 30% higher than someone with no experience. Note that the exact rate varies by department - Engineering candidates value their experience at the highest rate (4%), and IT at the lowest (2%.)
      2. Another factor that impacts the expected salary is the role the candidate is applying for. Some of the roles appear to be much higher level, and hence command a premium in salary expectations. Note that there is no evidence of industry having any impact on expected salary.
   4. A colleague on the analytics team asked about your *deduplication* and *analysis* methods for a, b, and c above and why you chose that approach. Please draft a few bullets to explain.
      1. **Deduplication**: There are two types of duplicates present in the data. The first type is a candidate who applied for more than one role within the same department, and the second one is a candidate who applied for more than 1 roles across distinct departments
      2. **Deduplication**: The second type of duplication is not a problem, as one candidate can be equally qualified for two different roles across different departments as they can be quite different from one another
      3. **Deduplication**: However, the first type could be problematic. Roles within the same department are quite different - the seniority appears to differ a lot. It would not make sense for one employee to qualify for two different levels of seniority
      4. **Deduplication**: In that case, their salary expectations might not be the most accurate and may be muddied by the perception of the other role. So for that reason, we have decided to only retain the role for which the candidate progressed the farthest into, as the best indication of qualification for that role, and hence proper ability to estimate compensation
      5. **Analysis**: The analysis that I used throughout was a multiple linear regression model, ran across departments
      6. **Analysis**: Most models had very good performance suggested by their R^2, especially after we applied a log transformation
      7. **Analysis**: We then inferred the impact of various factors by looking at their coefficients and the significance associated with them
      8. **Analysis**: We chose this model due to its simplicity, high interpretability, and ability to identify the isolated impact of various factors
2. The COO suspects that engineering and sales roles may not be priced properly and has asked for thoughts from the analytics team. Please draft an email to your analytics team manager that outlines the analysis you propose to answer this question. The analysis plan should include:
   1. How you would determine whether the compensation we’d offer a candidate is appropriate for these roles.
   2. What internal/external data you might collect or analyze to answer this question.
   3. The most important points for the COO to know.
   4. Any other information you think could be relevant to the COO’s question.

*[Note: No analysis is required here; we just want to see how you would approach the question.]*

*Hi COO-Name,*

*I am writing regarding the idea that you expressed about engineering and sales roles being inappropriately priced. My team and I have been analyzing this issue and have two possible solutions we would like to discuss.*

*The first idea revolves around using a 3rd party benchmark to properly identify how we are pricing these roles, relative to market expectations. In my experience, these benchmarks can be informative and bring to the table otherwise inaccessible data. This would require the least amount of work from our end and would be quicker but would be more costly.*

*The second idea revolves around building a predictive model to help us predict pay for a given candidate and role. This would require a dataset that comprises of fair compensation expectations, as this is the data our model would learn from. As such, building this model would require more time as we need to carefully examine our data. We could also feed in additional data, such as geographical information, internal compensation benchmarks for the role and more descriptive role information such as expected span of control and such.*

*This model, if built properly, could help us implement a sustainable and practical process for making pay decisions for new hires, while also ensuring fairness and consistency, as it would not be biased against external factors like gender and ethnicity. It would be more cost-effective than the other approach but would not be able to provide market benchmarks.*

*I will schedule a call for us to go over both options and present a set of possible timelines and associated costs. We can then evaluate them both properly and see which direction you would rather go in.*

*Thanks,*

*Steven*