Your CEO has decided that the company needs a full-time data scientist, and possibly a team of them in the future. She thinks she needs someone who can help drive data science within then entire organization and could potentially lead a team in the future. She understands that data scientist salaries vary widely across the world and is unsure what to pay them. To complicate matters, salaries are going up due to the great recession and the market is highly competitive. Your CEO has asked you to prepare an analysis on data science salaries and provide them with a range to be competitive and get top talent. The position can work offshore, but the CEO would like to know what the difference is for a person working in the United States. Your company is currently a small company but is expanding rapidly.

Prepare your analysis in an R file. Your final product should be a power point presentation giving your recommendation to the CEO. CEOs do not care about your code and don’t want to see it. They want to see visuals and a well thought out analysis. You will need to turn in the power point and the code as a flat R file.

The metadata are as follows:

| Column | Description |
| --- | --- |
| work\_year | The year the salary was paid. |
| experience\_level | The experience level in the job during the year with the following possible values: EN Entry-level / Junior MI Mid-level / Intermediate SE Senior-level / Expert EX Executive-level / Director |
| employment\_type | The type of employement for the role: PT Part-time FT Full-time CT Contract FL Freelance |
| job\_title | The role worked in during the year. |
| salary | The total gross salary amount paid. |
| salary\_currency | The currency of the salary paid as an ISO 4217 currency code. |
| salary*in*usd | The salary in USD (FX rate divided by avg. USD rate for the respective year via fxdata.foorilla.com). |
| employee\_residence | Employee's primary country of residence in during the work year as an ISO 3166 country code. |
| remote\_ratio | The overall amount of work done remotely, possible values are as follows: 0 No remote work (less than 20%) 50 Partially remote 100 Fully remote (more than 80%) |
| company\_location | The country of the employer's main office or contracting branch as an ISO 3166 country code. |
| company\_size | The average number of people that worked for the company during the year: S less than 50 employees (small) M 50 to 250 employees (medium) L more than 250 employees (large) |

**R Project Rubric**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Possible**  **Points** | **Earned**  **Points** |
| 1. Presentation – Clear, concise analysis  * Introduction to analysis * Clear & concise analysis * Recommendation | 20 |  |
| 1. Coding Style  * Code blocks properly commented * Use of functions for common data operations * Descriptive naming conventions * Proper use of built in functions (Ex: dplyr pipe) * Clean, organized code with logical code blocks for each part of the analysis (cleaning, analysis, plots, etc) | 50 |  |
| 3. Analysis   * Original thought – go beyond counts and presenting rows. This is a data science course. * Aggregation – you should be using aggregation multiple times in your analysis to present data. * Compelling plots, do they tell the story & are they properly labeled with clear axes. | 30 |  |

**Points 100**