**Report for summarizing all steps you have already done**

This report summarizes the analytical steps undertaken on the our dataset, primarily focused on exploring statistical characteristics and building a linear regression model.

Initially, I conducted basic statistical calculations on the 'salary\_in\_usd' column. This included determining the mean, median, mode, standard deviation, minimum and maximum values, quartiles, and range. These metrics provided a comprehensive overview of the salary distribution, revealing central tendencies and variability within the data.

Next, I examined the unique values of various categorical variables, such as experience level, employment type, job titles, salary currency, employee residence, remote ratio, company location, and company size. This analysis highlighted the dataset's diversity, covering a wide range of job titles, geographical locations, and company sizes, thus reflecting the dataset's global and multifaceted nature.

Finally, a linear regression model was built to predict salaries (in USD). The model used numerical variables from the dataset, including 'work\_year', 'salary', and 'remote\_ratio'. The regression analysis revealed a relatively low R-squared value, indicating limited explanatory power of the model in accounting for salary variability. Each predictor's coefficient was analyzed for its impact on the salary. Notably, the model exhibited potential issues such as multicollinearity, as indicated by a high condition number. The inclusion of 'salary' as both an independent and dependent variable was identified as a probable anomaly, potentially skewing the model's effectiveness and interpretation.

Overall, these analyses provided valuable insights into the dataset's characteristics and raised important considerations for further refining the predictive model