

DS311

Group Project Documentation

Salary Data Set Analysis Plan

One of the ultimate goals in this project is to identify the specific job-titles, locations, and companies of data-related jobs, which a worker wants, based on the salaries the worker would make in those positions. This is a common research questions on the supply side of the labor market for data analysts to solve in Human Resources platforms like LinkedIn, Glassdoor, Indeed, Dice, etc ...

Key Performance Indicator (KPI): Paid Wage Per Year (in salary_data)

There are three high-level questions with more details questions you should be able to answer with the dataset.

- I. Do specific sub-types of data-related jobs have higher or lower salaries than others?
 - Which companies have the highest salaries for those sub-types?
 - Changes with location of the job?
 - Will the answer change if I take standard of living into account?
- II. What states (of those I am willing to move to) have the highest paying data-related salaries?
 - Differences between job sub-categories?
 - Which companies have the highest salaries for those sub-types?
 - Will the answer change if I take standard of living into account?
- III. How do offered salaries compare to the prevailing wage?
 - Are there job sub-categories that tend to get over-paid or under-paid?
 - Are there companies that tend to over-pay or under-pay?
 - Will the answer change if I take standard of living into account?

By answering the three high-level questions, you should have a better understanding about the data-related job market. The second part of the project is to come up with three other high-level questions and answer them with some specific questions to the dataset.

For each question you are answering in this project, the answers MUST be supported by the data, which can be tables or graphs. However, you should be aware that you will never have a chance in the real-world presentation to flash of 15 plots with 20 tables because no stakeholders will wait for your presentation of all these plots and tables. Be wise on selecting the right presenting material and make sure telling a story from each table and plot.