2023

Salary Breakdown in the Data Science Field

[DOCUMENT SUBTITLE]

STUDENT'S NAME

A1. Research Question

What are the characteristic of the salary of those in the data science field? Are there salary benefits to being a contractor as opposed to a freelancer? Do executives consistently make more money than senior level employees? And how much more do mid level employees make compared to entry level?

A2. Context and background

The technology job market is ever increasing and evolving. For the many people who are interested in getting into a career in technology, a basic overview of salary expectations can provide context of what to expect with certain job categories as well as give those interviewing for jobs a basic expectation for salary negotiations.

A3. Three summarized sources

In an article titled “What is a Data Scientist”, it says “The Bureau of Labor Statistics projects 35.8% employment growth for data scientists between 2021 and 2031.” This article also goes on to describe these jobs as coming from a wide variety of employers from tech start ups to fortune 500 companies. With that much variety in job market and that much growth, having clear expectations of payment and salary is very important. This is a good article for my research because with this much growth in the job market for data scientists, there will be many people who will have questions about salary expectations for the field of work they are interested in.

Aside from the regular data science jobs, there are also data science executive positions. Someone who may be aspiring to join the data science field may not have the deep technical ability to be a data scientist, but possess the skills necessary to lead a team. This could be an opportunity for some. This analysis will include them in salary expectations. A quick article from Ladder.com shows the expectations for a “Chief Data Officer” across the United States. This article is good evidence toward my analysis because many may believe that you must be a programmer or a statistician in order to be in this job field, but that is not the case. There are many opportunities including “Chief Data Officer”, “Chief Diversity Officer”, and “Chief Technology Officer”. None of these jobs require a deep technical knowledge, but offer someone a chance to be in this line of work. This article helps to understand the salaries for those areas of the data science field.

The third piece of evidence that I will use to bring efficacy to this analysis is an article from Fox News quoting statistics from the Bureau of Labor Statistics showing the 100 highest paying jobs in America. The technology field claims 13 out of 100, with “Data Scientist” being number 64. With so many jobs in America relying on Technology, it is an amazing opportunity to make a decent amount of money. This also provides evidence to the fact that the data science job market will be highly sought after. With such a large implication of job availability and growth, both employees and employers should have a good grasp on expected salaries.

A4. Summary of Data Analytics Solution

This analysis will utilize a data set from Kaggle.com listing salaries across job categories and employee type categories. I will be using Microsoft Excel for the data analysis. Specifically the data analysis

toolpack t-test models. I will compare the mean salaries and standard deviations of contractors vs freelancers. I will do the same comparison for entry level employees vs mid level employees and mid level vs senior level employees. Lastly I will do the comparison for executives vs senior level empolyees. What will be delivered will be scatter plots showing salary ranges for the employee types listed. Mean salaries and their standard deviations for each employee type listed, and some T-test results comparing two of the employee types listed.

A5. Benefit of Solution

I hope to be able to present some deeper information pertaining to salaries for employees in the Data Science industry. It may provide those interested in working in that field with some insights on what they can expect to earn. Or if they have an earning expectation, which job field they should aspire to.

B1. Goals, Objectives, and Deliverables

Goal: Produce an Excel workbook that accurately breaks down salary statistics for jobs in data science.

Objective A: Data Collection

Deliverable 1: Download a Kaggle dataset

Deliverable 2: A cleaned dataset

Objective B: Provide statistical analysis of jobs in the dataset

Deliverable 1: mean, standard deviation, T-test, and scatter plots of contactors vs freelancers salaries

Deliverable 2: mean, standard deviation, T-test, and scatter plots of entry level vs mid level employees and mid level vs senior level employees salaries

Deliverable 3: mean, standard deviation, T-test, and scatter plots of senior level vs executive employee salaries.

B2. Scope

The scope of this work will include an initial .csv file from Kaggle that will be downloaded into an Excel workbook. It will be broken down into individual spreadsheets within that workbook for the comparisons listed above. The scope will only be for recent salaries, from 2020-2023, and will not provide a deeper analysis in the sense of historical salaries.

B3. Planning Methodology

I plan to use the ADDIE method to work through this project. ADDIE stands for Analyze, Design, Develop, Implement, and Evaluate. Here is my plan.

Analyze – After downloading the .csv from Kaggle I will need to spend some time making sure the dataset is clean and complete. Clean in the sense that I have uniform data types for each column, and complete in that I’m not missing data points in each row.

Design – The design phase will be breaking up each portion of the .csv into individual spreadsheets so the specific analysis can be done with the groups listed in my goals. I will also use this phase to decided how charts and results will be formatted and presented.

Develop – Once the dataset is broken into specified spreadsheets, I will use the built in Excel functions and the data analysis toolkit to provide the results of my analysis specified above.

Evaluate – From the analysis I will draw some conclusions on the salary characteristics of jobs in the data science field.

B4. Milestone Timeline

Milestone	Start Date	End Date	Duration
Analyze	6/19/2023	6/20/2023	2 day
Design	6/21/2023	6/23/2023	2 day
Develop	6/24/2023	6/26/2023	2 days
Implement	6/27/2023	6/29/2023	2 day
Evaluate	6/30/2023	7/1/2023	2 day

B5. Resources and Costs

Resources	Cost
Kaggle dataset, Microsoft Excel, Excel Data Analysis Toolkit	Free

B6. Measurable Criteria for Success

The measure for success in this analysis will be to uncover insights on salary expectations with a clean set of data. The first measure is to have no missing data points in my dataset. The second measure of success is to clearly show and explain the statistics that I have mentioned in my goals in B1.

C1. Project Hypothesis

> Contractors make more money than freelancers. Staying in a job from entry level until mid level will result in a significant increase in salary, while mid level to senior level will provide a minimal increase in salary. My final hypothesis is that to make the most money in the data science field, someone should aspire to be an executive rather than a senior level employee.

C2. Analytical method (descriptive, diagnostic, predictive, or prescriptive)

What I am analyzing would fall under descriptive analytics. I'm taking what has already happened and will describe important characteristics about it.

C3. Describe tools and environments you'll use

I am planning on using the Microsoft Excel environment along with the data analytics toolpack addition.

C4. Metrics for statistical significance

The metric for statistical significance will be the mean and standard deviations of the salaries themselves as well as P-values less than .05 of the T-tests run.

C4a.

Being able to see the mean and standard deviation of the salaries will provide a good metric on the average salary and how much fluctuation the salary can have. With the T-test a P-value of .05 is statistically significant and has widely accepted standard for determining statistical significance. With a P-value of less than . 05, you should reject the null hypothesis.

C5. Assess practical significance

The practical significance of this analysis will be to allow incoming employees to the data science field get an idea of what they can expect from a salary in different job categories and allow them to set their own career goals accordingly.

C6. Describe tools and graphics to visually show findings

The Microsoft Excel analytics toolpack will allow me to do the T-test and will return a visual chart listing means, P-values, and degrees of freedom. I can also use some chart functionality to plot some simple data points.

D1. ID data source

I got my dataset from Kaggle. It is the 4th listed source in my sources section.

D2. Why is the data set appropriate

The dataset is appropriate because it is current data, so there is no need to adjust for inflation, and it contains all the data characteristics I need to do my analysis. It also contains a large enough data sample to gain the insights I'm looking for.

D3. Data collection method

I downloaded the .csv file from Kaggle.

D4. Summarize observations on quality and completeness of the data

The data contains strings and integers. There are also no row of missing data, so no cleaning is necessary.

D5. data governance, privacy and security, ethical, legal, and regulatory compliance considerations that relate to the dataset and the proposed project.

Because this data comes from Kaggle.com it is free to use and I do not feel it breaks any laws or ethical standards for use.

E. Sources

1. *Data Scientist Ranks Among Best Jobs of 2023*. (n.d.). https://money.usnews.com/careers/best-jobs/data-scientist

2. *Chief Data Officer (CDO) Salary - Find & Compare 100K Job Salaries | Ladders*. (n.d.). Ladders. https://www.theladders.com/salary/chief-data-officer-cdo-salary

3. Berkman, S., & Givens, A. (2023, June 2). *100-highest-paying-jobs-in-america*. https://fox2now.com. https://fox2now.com/news/national/100-highest-paying-jobs-in-america/

4. *Data Science Salaries 2023* 💸. (2023, April 13). Kaggle. https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023