

# Data Science: Where in the World am I Going to Work?

Arthur Richardson

2022-10-12

## Contents

<b>Dedication:</b>	<b>2</b>
<b>Introduction:</b>	<b>3</b>
<b>Objective:</b> . . . . .	3
<b>Data Installation:</b> . . . . .	4
<b>Data Analysis:</b> . . . . .	5
<b>Dimensions and Summary</b> . . . . .	5
<b>Cost of Living and Pay</b>	<b>6</b>
<b>Career Exploration</b>	<b>8</b>
<b>Career Location</b>	<b>13</b>
<b>Conclusion</b>	<b>17</b>
<b>Reference Section:</b>	<b>18</b>

## Dedication:

This project is dedicated to my two daughters.

*Reflect on the important things to gain a better introspective and  
always move forward.*

*Love, Dad.*

## **Introduction:**

I will analyze Data Science Job Salaries from Ruchi Bhatia. This data set contains salaries of jobs in the Data Science domain. The dataset variable terminology is located in the Data Science Job Salaries Data Set Glossary and Terminology Section of this document.

## **Objective:**

Data Science is a diverse and rewarding career. In this project, we will explore the data associated with the career field of Data Science to give you an insight into the different job positions that are available to you based on Salary, Cost of Living, Position, Country, and more!

## Data Installation:

Upload following packages and libraries for data exploration.

```
library(tidyverse)
library(caret)
library(data.table)
library(RColorBrewer)
library(rmarkdown)
library(dslabs)
library(gtable)
library(hexbin)
library(gt)
library(dplyr)
library(ggpmisc)
library(gridExtra)
library(janitor)
library(lubridate)
library(highcharter)
library(viridisLite)
library(broom)
library(scales)
library(xfun)
library(htmltools)
library(mime)
library(ggfortify)
library(gtsummary)
library(tinytex)
library(vroom)
library(curl)
library(gtools)
library(hrbrthemes)
library(viridis)
library(latexpdf)
library(kableExtra)
library(showtext)
library(remotes)
library(extrafont)
```

## Data Analysis:

Upload the data set. The file can be downloaded from <https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

## Dimensions and Summary

## [1] 607 12

The dataset has 607 observations for 12 variables.

## Data Science Job Salaries Data Set Glossary and Terminology:

1. Analyst ID: Provides an unique ID number to each Data Analyst Job.
2. Work Year: The year the salary was paid.
3. 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.
4. Employment Type: The type of employment for the role: PT Part-time, FT Full-time, CT Contract, FL Freelance.
5. Job Title: The role worked in during the year.
6. Salary: The total gross salary amount paid.
7. Salary Currency: The currency of the salary paid as an ISO 4217 currency code.
8. Salary in USD: The salary in USD (FX rate divided by avg. USD rate for the respective year via [fxdata.foorilla.com](https://fxdata.foorilla.com)).
9. Employee Residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.
10. 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%).
11. Company Location: The country of the employer's main office or contracting branch as an ISO 3166 country code.
12. Company\_Size: The average number of people that worked for the company during the year: S means less than 50 employees (small). M means 50 to 250 employees (medium). L means more than 250 employees (large).

## Cost of Living and Pay

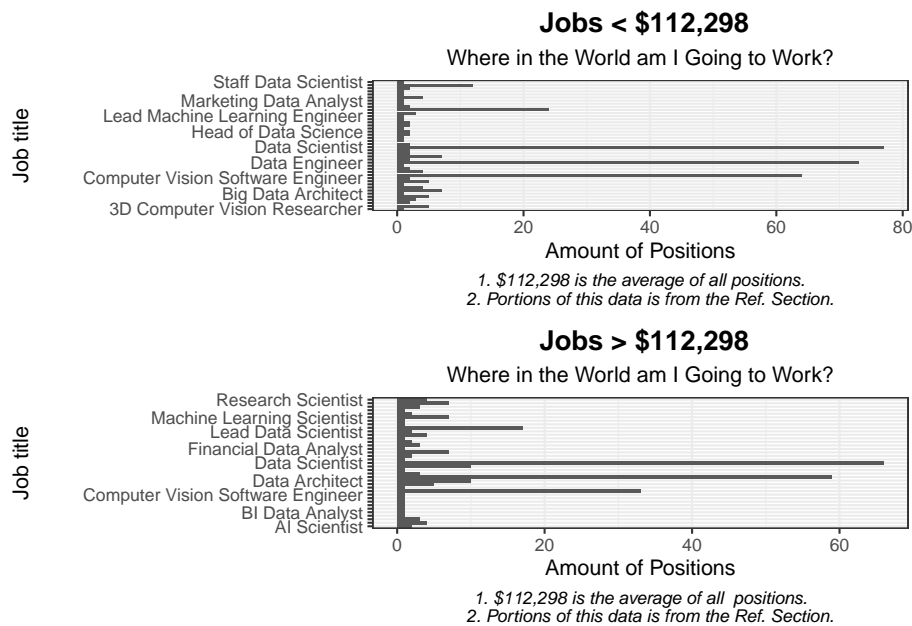
As we start our journey in the Data Science career field, we have to analyze the cost of living for the country we would like to work in. What is the Cost of Living? Cambridge Dictionary (“Cambridge”, n.d.) states, “the amount of money a person needs to live.” To expound upon the definition even further,

Per Investopedia (“Banton, C.”, 2022), “The cost of living is the amount needed to cover basic expenses such as housing, food, taxes, and healthcare in a particular place and time. The cost of living is used to compare how expensive it is to live in one city versus another. The cost of living is tied to wages. If expenses are higher in a city, such as New York, salary levels must be higher so that people can afford to live in that city.”

Now that we understand the cost of living, let us explore the average salary for the Data Science Field, Highest Paying Job, and Lowest Paying Job.

Based on all available positions in this dataset, let's get the average salary in USD. After, we will take a look at all jobs making less than and more than the average salary.

```
## [1] "$112,298"
```



Now that we have an idea of the highest and lowest earners within the Data Science field, let's take a look at the highest and lowest pay jobs.

### Highest Paying Job

Data Analyst Jobs 2020-2022

Job Title	Max. Salary	Exp. Level	Company Loc.	Company Size
Principal Data Engineer	\$600,000	EX	US	L

Salary is in USD

EX = Experience level is Executive-level

US = United States

L = Large Company, more than 250 employees

Portions of this data is from the Reference Section.

### Least Paying Job

Data Analyst Jobs 2020-2022

Job Title	Min. Salary	Exp. Level	Company Loc.	Company Size
Data Scientist	\$2,859	MI	MX	S

Salary is in USD.

MI = Experience level is Mid-Level

MX = Mexico

S = Small Company, less than 50 employees

Portions of this data is from the Reference Section.

After reviewing the highest and lowest-paying jobs, we have to factor in the cost of living. Per World Population Review (2022), "Mexico's general minimum wage (Salario Minimo) was raised on January 1, 2022, to 172.87 pesos per work day, a 23% rise on the 2021 level of 141.70 pesos. The rate for 2022 along the 'Northern Border Zone' was raised to 260.34 pesos per work day, a rise of 23% on the 2021 rate of 213.39 pesos. Mexico's Minimum Wage Increased by 23% for 2022." In USD, that's about 8 dollars a day, which is \$2920 a year. They are making below minimum wage and the cost of living is 491.91(USD) per month without rent factored into the equation!

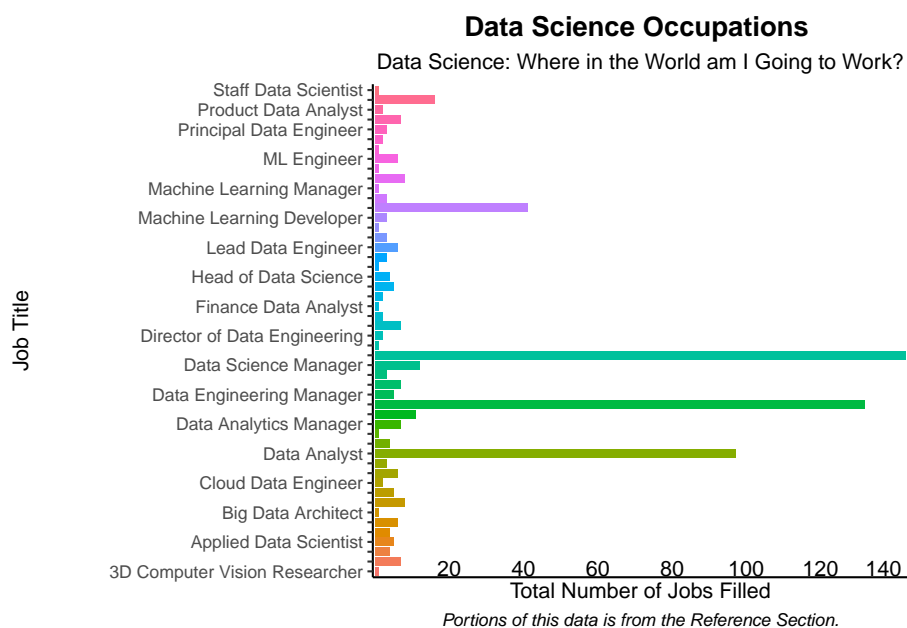
Per Numbeo ("Cost of Living", n.d.), "The average household in the United States spends 61,334 (USD) a year on expenses. On average, 34.9% of spending, or roughly 1,784 (USD) a month, is dedicated to housing and housing-related costs." With an average salary of \$112,298 in the data science field, you are looking at 55% surplus in funds after the annual cost of living is subtracted.

As we can see, the cost of living can really affect if the salary is enough to support your lifestyle in a particular country.

Let us narrow down our specific job within the Data Science career field.

## Career Exploration

As we explore the different jobs within the Data Science field, we need to know which positions are the most populated, which type of company would be more advantageous for us in regards to pay, and which country has the best career opportunities pay to fit your needs.



Take a look at the jobs with the most personnel.

```
## [1] "Data Scientist"
```

```
## [1] 143
```

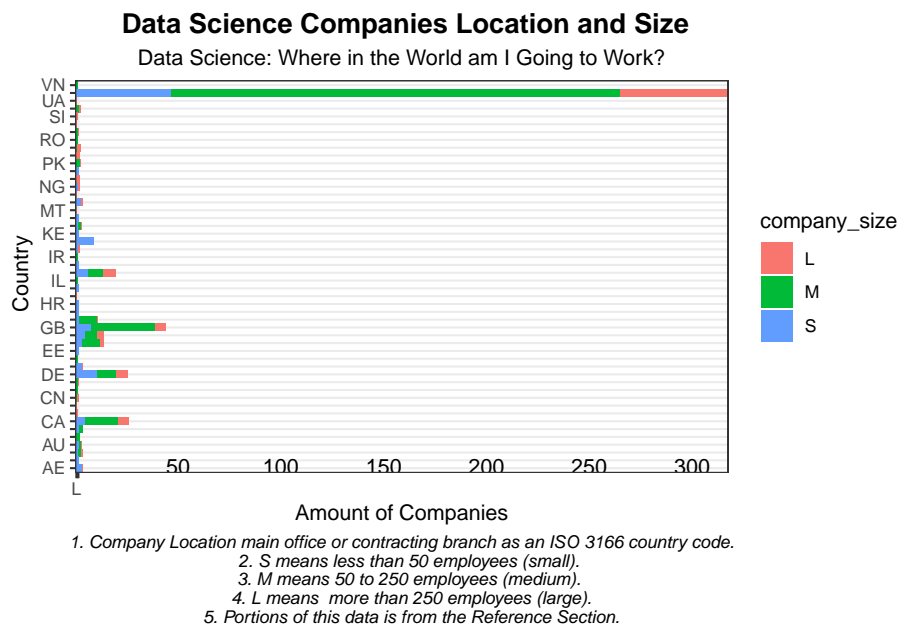
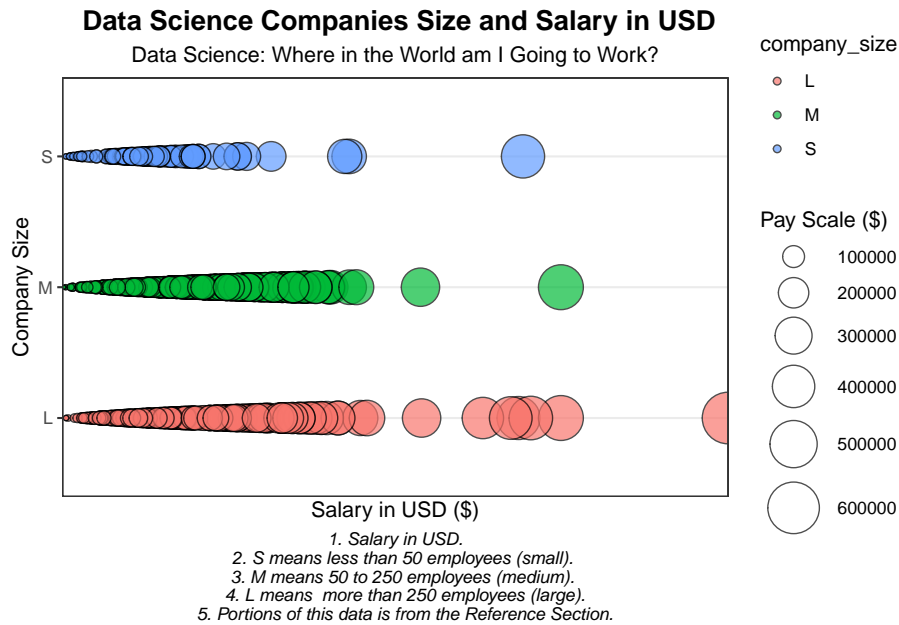
Now lets evaluate the jobs with the least amount of personnel.

```
## [1] "3D Computer Vision Researcher" "Big Data Architect"
## [3] "Data Analytics Lead"          "Data Specialist"
## [5] "Finance Data Analyst"         "Head of Machine Learning"
## [7] "Lead Machine Learning Engineer" "Machine Learning Manager"
## [9] "Marketing Data Analyst"       "NLP Engineer"
## [11] "Staff Data Scientist"
```

```
## [1] 1
```

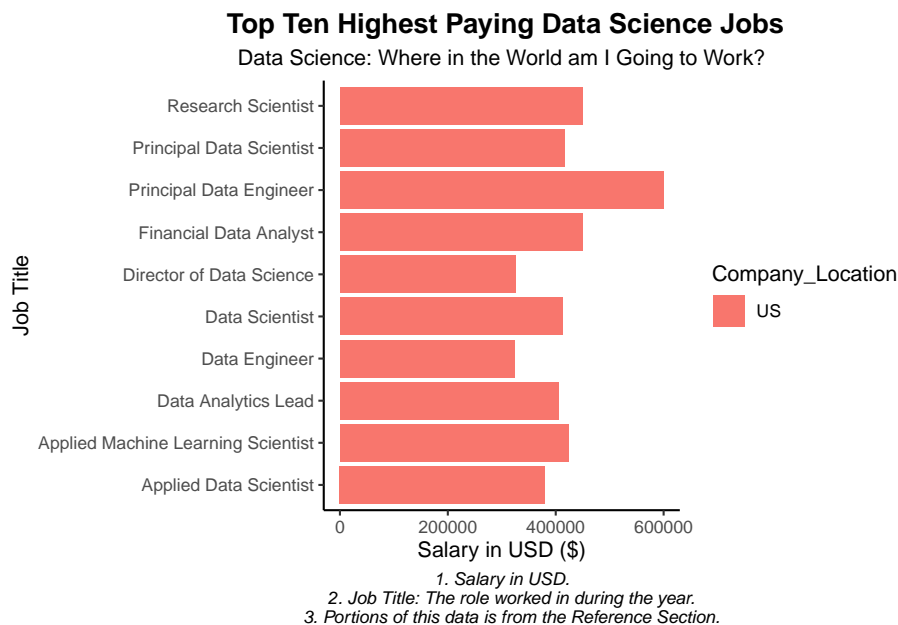


As you can see, the Data Science field has various positions. Data Scientist, Data Analyst, and Data Engineer are some of the most populated positions based on our data set. Next, we will visualize the countries with the largest, medium sized and smallest companies.

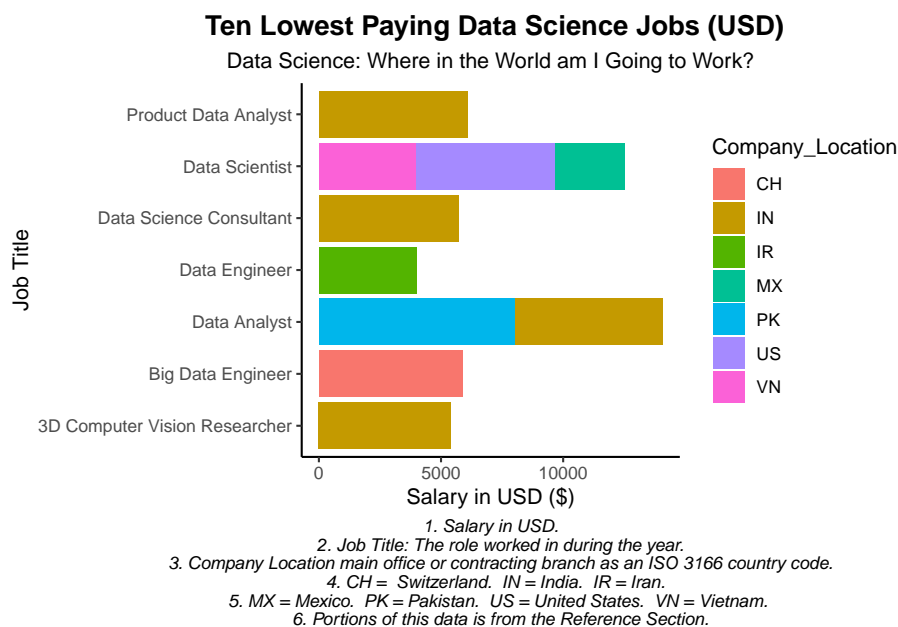


Wow, all three company types pay handsomely over the Data Science career field average and the largest companies pay over 5x more than the average.

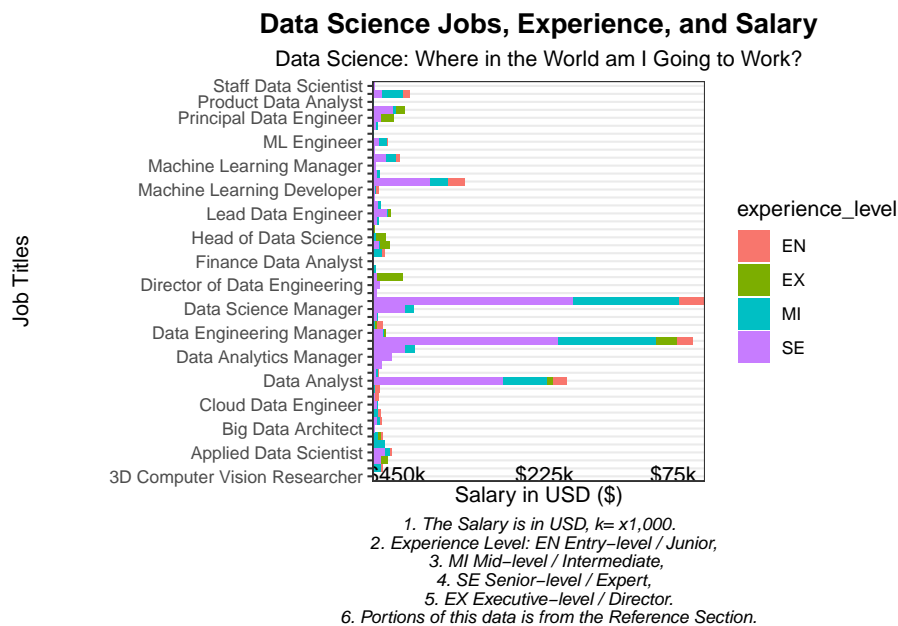
Since we have the Salary data, let's explore which country has the ten highest-paying and lowest-paying jobs.



Now let's view the Bottom 10 positions by Salary.



As we can see, the United States is the leader in the highest-paid positions. We know which jobs are the highest and lowest paying. Additionally, we know where they are located, but we all have different experience levels. Our experience level matters. You may be a novice analyst or sought-after expert within a specific field, but which job will pay you the most for your experience level? Let's explore it.



Based on Experience, let us see which job pays the most and what country.

```
## # A tibble: 1 x 4
##   Title                      SalaryUSD Company_Location Exp
##   <chr>                      <dbl> <chr>             <chr>
## 1 Machine Learning Engineer  250000 US              EN

## # A tibble: 1 x 4
##   Title                      SalaryUSD Company_Location Exp
##   <chr>                      <dbl> <chr>             <chr>
## 1 Financial Data Analyst    450000 US              MI

## # A tibble: 1 x 4
##   Title                      SalaryUSD Company_Location Exp
##   <chr>                      <dbl> <chr>             <chr>
## 1 Data Scientist           412000 US              SE

## # A tibble: 1 x 4
```

```
## Title SalaryUSD Company_Location Exp
## <chr> <dbl> <chr> <chr>
## 1 Principal Data Engineer 600000 US EX
```

Now lets see if the benifits from Experience, Salary, and Location.



For the Expert experience level, we can ascertain that it has fewer opportunities but higher pay based on the visualization. Additionally, we can conclude that most countries have more Mid-level and Senior Level positions than any other position. Now that we know our opportunities based on our experience level, let us explore remote job opportunities.

## Career Location

Since February 11 2020, COVID-19 has changed the landscape of in person working experience. Most companies have transitioned to remote working. According to (J. Wise, 2022 ), “statistics indicate to us that around the world, 18% of people are now working remotely full-time. In America, over 4.3 million people work remotely, which is 3.2% of the workforce. Remote work is a growing trend, but 44% of companies currently do not let their employees work remotely.”

Per ApolloTechnical (“Apollo Technical LLC”, 2022), “Upwork estimates that 1 in 4 Americans, over 26% of the American workforce, will work remotely through 2021. They also estimate that 22% of the workforce (36.2 Million Americans) will work remotely by 2025.”

The remote worker for this dataset will be defined by remote ratio. The remote ratio for this dataset is 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%). Let us explore the remote working opportunities within the data science field.

### Ten Best Remote Paying Jobs

```
## # A tibble: 10 x 6
##   Title                SalaryUSD Company_Location Exp RemoteWork Year
##   <chr>                <dbl> <chr>          <chr>    <int> <int>
## 1 Data Architect      266400 US           SE        100  2022
## 2 ML Engineer        270000 US           MI        100  2021
## 3 Data Engineer      324000 US           EX        100  2022
## 4 Director of Data Science 325000 US           EX        100  2020
## 5 Applied Data Scientist 380000 US           SE        100  2022
## 6 Data Analytics Lead  405000 US           SE        100  2022
## 7 Data Scientist     412000 US           SE        100  2020
## 8 Principal Data Scientist 416000 US           EX        100  2021
## 9 Financial Data Analyst 450000 US           MI        100  2021
## 10 Principal Data Engineer 600000 US           EX        100  2021
```

**Note: 1. Salary is in USD.**

**2. 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.**

**3. Employee Residence: Employee’s primary country of residence in during the work year as an ISO 3166 country code.**

**4. 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%).**

## Bottom Ten Remote Paying Jobs

```
## # A tibble: 10 x 6
##   Title                               SalaryUSD Company_Locat~1 Exp   Remot~2 Emplo~3
##   <chr>                               <dbl> <chr>           <chr>   <int> <chr>
## 1 3D Computer Vision Researcher      5409 IN             MI       50 IN
## 2 Data Science Consultant            5707 IN             EN       50 IN
## 3 Data Analyst                       8000 PK             MI       50 PK
## 4 Data Analyst                     10354 ES             EN       50 ES
## 5 Machine Learning Scientist        12000 PK            MI       50 PK
## 6 Data Scientist                    20171 TR            SE       50 TR
## 7 Data Engineer                     21637 IN            EN       50 IN
## 8 Data Scientist                    21669 IT            EN       50 IT
## 9 Machine Learning Engineer         21844 CO            EN       50 CO
## 10 Machine Learning Engineer        24823 SI            MI       50 SI
## # ... with abbreviated variable names 1: Company_Location, 2: RemoteWork,
## # 3: 'Employee Residence'
```

Note: 1. *IN = India. PK = Pakistan. ES = Spain. TR = Turkey. IT = Italy. SI = Slovenia. CO = Columbia.*

2. *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.*

3. *Employee Residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.*

4. *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%).*

5. *Portions of this data is from the Reference Section*

## Ten Best Non Remote Paying Jobs

```
## # A tibble: 10 x 6
##   Title                SalaryUSD Company_Location Exp RemoteW~1 Emplo~2
##   <chr>                <dbl> <chr>          <chr>    <int> <chr>
## 1 Research Scientist    450000 US             MI         0 US
## 2 Lead Data Engineer    276000 US             SE         0 US
## 3 Machine Learning Scientist 260000 JP             SE         0 JP
## 4 Director of Data Science 250000 US             EX         0 US
## 5 Data Science Manager   240000 US             SE         0 US
## 6 Data Engineer         220110 US             SE         0 US
## 7 Principal Data Scientist 220000 US             SE         0 US
## 8 Data Scientist        215300 US             SE         0 US
## 9 Data Engineer         206699 US             MI         0 US
## 10 Data Scientist       205300 US             SE         0 US
## # ... with abbreviated variable names 1: RemoteWork, 2: 'Employee Residence'
```

Note: 1. *JP = Japan. US = United States.*

2. *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.*

3. *Employee Residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.*

4. *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%).*

5. *Portions of this data is from the Reference Section*

## Bottom Ten Non Remote Paying Jobs

```
## # A tibble: 10 x 6
##   Title                SalaryUSD Company_Location Exp RemoteWork Employ~1
##   <chr>                <dbl> <chr>          <chr>    <int> <chr>
## 1 Data Scientist      2859 MX          MI        0 MX
## 2 Data Scientist      4000 VN          EN        0 VN
## 3 Big Data Engineer    5882 CH          EN        0 IN
## 4 Data Analyst         6072 IN          EN        0 IN
## 5 Data Scientist      9466 IN          MI        0 IN
## 6 Data Engineer       12103 TR          MI        0 TR
## 7 Data Scientist     12901 BR          MI        0 BR
## 8 Big Data Engineer   18000 MD          MI        0 MD
## 9 Computer Vision Engineer 18907 BR          SE        0 BR
## 10 Product Data Analyst 20000 HN          MI        0 HN
## # ... with abbreviated variable name 1: 'Employee Residence'
```

Note: 1. *MX = Mexico. VN = Vietnam. CH = Switzerland. IN = India. TR = Turkey. BR = Brazil MD = Moldova HN = Honduras.*

2. *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.*

3. *Employee Residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.*

4. *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%).*

5. *Portions of this data is from the Reference Section*



## Conclusion

After exploring the different positions within the Data Science field, one may be able to make a sound decision on which position they would like to endeavor in, the country that provides them the most opportunities based on pay and experience level.

## Reference Section:

1. Irizarry, R. A. (2022, July 7). Introduction to Data Science. HARVARD Data Science. Retrieved August 8, 2022, from <https://rafalab.github.io/dsbook/> This project utilized “Introduction to Data Science Data Analysis and Prediction Algorithms with R” by our course instructor Rafael A. Irizarry published 2022-07-07.
2. Cost of living. Cost of Living. (n.d.). Retrieved October 11, 2022, from <https://www.numbeo.com/cost-of-living/> Numbeo is the world’s largest cost of living database. Numbeo is also a crowd-sourced global database of quality of life data: housing indicators, perceived crime rates, healthcare quality, transport quality, and other statistics.
3. Banton, C. (2022, February 8). Cost of living. Investopedia. Retrieved October 11, 2022, from <https://www.investopedia.com/terms/c/cost-of-living.asp>
4. Cost of living. Cambridge Dictionary. (n.d.). Retrieved October 11, 2022, from <https://dictionary.cambridge.org/dictionary/english/cost-of-living>
5. Cost of living index by state 2022. (n.d.). Retrieved October 11, 2022, from <https://worldpopulationreview.com/state-rankings/cost-of-living-index-by-state>
6. Wikimedia Foundation. (2022, October 5). List of ISO 3166 country codes. Wikipedia. Retrieved October 11, 2022, from [https://en.wikipedia.org/wiki/List\\_of\\_ISO\\_3166\\_country\\_codes](https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes)
7. Statistics on remote workers that will surprise you (2022). Apollo Technical LLC. (2022, May 31). Retrieved October 11, 2022, from <https://www.apollotechnical.com/statistics-on-remote-workers>
8. Wise, J. (2022, July 28). Remote work statistics 2022: How many people work from home during Covid? EarthWeb. Retrieved October 11, 2022, from <https://earthweb.com/remote-work-statistics>