Data Jobs Salaries in Mexico in January 2022

Author: Daniel Eduardo López

January 2022

Executive Summary

With the emergence of the big data, new jobs have appeared demanding new sets of skills and expertise for extracting value from often raw and unstructured information. To cope with this challenge, several data profiles have evolved: Business Analysts, Data Analysts, Data Architects, Data Engineers and Data Scientists.

For job seekers and recruiters, it is relevant to know what the salaries for each of those categories are. So, the question guiding the present analysis was: Which data job category has the highest salary in the Mexican labor market in January 2022 according to the OCC website? Thus, in the present study, data about job titles, salaries, employers, and locations was collected from the OCC website (Mexico) through web scrapping. Then, the data was cleaned and prepared in Power Query and explored and analyzed in Excel.

It was found that the Data Engineers and Data Scientists enjoyed the highest salaries (figure 1).
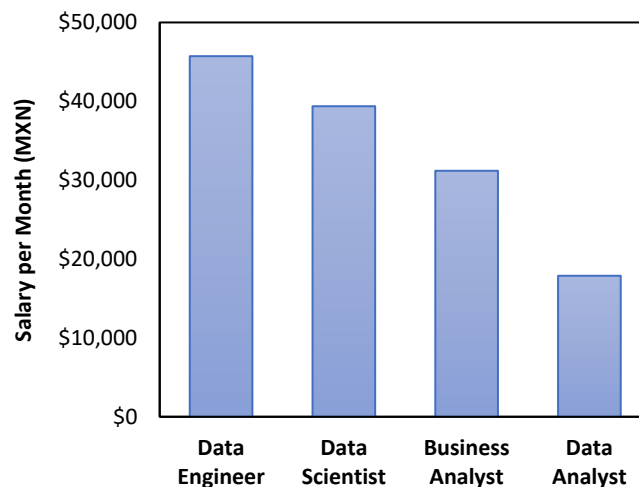

Figure 1. Average salary for each data job category in Mexico in January 2022 (own elaboration).

Even though the average salary for Data Engineer was higher than that for Data Scientists, it was found that the difference was not statistically significant. Moreover, it was observed that Ciudad de México and Estado de México were the states where the greatest number of data jobs were concentrated along with the highest salaries.

Additionally, the companies with the greatest demand of data positions (Grupo Salinas, Softtek and Banamex) and the ones with the highest salaries (Mobiles Apps Enterprises, Compartamos Servicios and Cuarto Origen) were also identified.

Table of Contents

1. Introduction

With the massification of the digital devices and technologies, it has been possible to collect large amounts of data as it had not been done in the past. Thus, novel technologies, innovative companies and new jobs have appeared to deal with such amount of information in order to extract value and, hopefully, financial return. In this sense, with the emergence of the big data, a new data jobs market has appeared to provide the skilled labor needed to accomplish this task.

According to Axis Talent (2020), the data jobs market can be analyzed and classified into several sometimes-overlapping job profiles: Business analysts, data analysts, data architects, data engineers and data scientists.

Business Analysts (BA) optimize organizational resources by using data-driven analysis (Axis Talent, 2020). Because of the nature of their responsibilities, business analysts need substantial knowledge and expertise on their particular industry and organization. Conversely, Data Analysts (DA) extract information to support and enhance the decision-making process (Kudyba, 2014). They usually use spreadsheets applications, business intelligence software as well as data visualization tools to perform their responsibilities (Axis Talent, 2020; Kudyba, 2014).

On the other hand, Data Architects (DR) design the company data management framework. Therefore, deep knowledge and skills on data bases, ETL process and big data technologies are mandatory to fulfill this position (Axis Talent, 2020). Whereas Data Engineers (DE) are responsible for building, implementing maintaining and optimizing the framework and processes that will extract, transform, load, and deliver data to relevant users; as well as taking steps to ensure the quality and reliability of the same (Axis Talent, 2020).

Finally, Data Scientists (DS) extract insights from both structured and unstructured data, by acquiring, exploring, cleaning, and transforming information to create machine learning models that help them to make predictions on some outcome of interest (Grus, 2019).

Even though the different data jobs overlap at certain extent, they undoubtedly require a different set of knowledge and skills to coupe successfully with their functions and responsibilities. Thus, for jobseekers and recruiters it is relevant to know what set of knowledge and skills is more valued in the Mexican labor market currently.

In this context, the purpose of the present study is to identify which data job category has the highest salary in the Mexican labor market in January 2022 according to the OCC website.

2. Methodology

The methodology of the present study is based on Rollin's Foundational Methodology for Data Science (Rollins, 2015).

2.1 Objective

General objective

To identify which data job category has the highest salary in the Mexican labor market in January 2022 according to the OCC website.

Specific objectives

- To explore in which Mexican state the demand of data jobs is the greatest.
- To identify what companies demand the greatest number of data jobs.
- To identify which Mexican state the highest salaries are offered.
- To identify what companies offer the highest salaries.

2.2 Research question

Which data job category has the highest salary in the Mexican labor market in January 2022 according to the OCC website?

2.3 Hypothesis

In accordance with Davenport & Patil's statement that Data Scientist is the sexiest job of the 21st century (Davenport & Patil, 2012), it was hypothesized that Data Scientist is the data job category with the highest salary in the Mexican labor market in January 2022.

2.4 Analytical approach

Descriptive and inferential statistics were used in order answer the research question.

2.5 Data requirements

Data about job positions such as job title, salary, employer, and location were set as data requirements. For the purposes of the present study, no distinction was made among entry level, middle or senior positions.

2.6 Data collection

Data was collected from the OCC website (Mexico) on 06 January 2022, through web scrapping with original Python code. The OCC website was selected due to their more lenient policy for automated process of information retrieval.

2.7 Data understanding and preparation

Data was prepared and cleaned with Power Query. Then, a duplicate data set was created. In the first data set, only the job positions with full salary information were kept for answering the research question. Whereas in the second data set, the whole vacancies were maintained to explore the data jobs demand geographically and by company according to the specific objectives of the present study. After the two data sets were created, data was explored and visualized in Excel and Power BI.

2.8 Statistical analysis

Data was analyzed through descriptive and inferential statistics in Excel. Average salaries were obtained for each data job category. Furthermore, a histogram and a box plot were constructed to visualize the distribution of the salaries, overall and across the data job categories.

On the other hand, statistical tests were performed to assess the statistical significance of the results obtained on a data set cleaned from outliers. Firstly, a one-way analysis of variance (ANOVA) procedure was carried to detect whether the average salary differences for the distinct data jobs categories were statistically significant. Then, T-tests for two independent samples assuming unequal variances were used to assess whether the average salary for different category pairs was significantly different.

3. Results and Discussion

3.1 Data collected

Data collection and cleaning allowed to identify 322 data jobs in Mexico in January 2022 from the OCC Website, of which 58 had salary information. The whole data set was used in the data jobs demand analysis below and the job positions with salary information were used in the salary analysis more below.

3.2 Data jobs demand analysis

From the whole universe of data jobs identified, almost half of the vacancies correspond to Data Analyst's positions (figure 2).
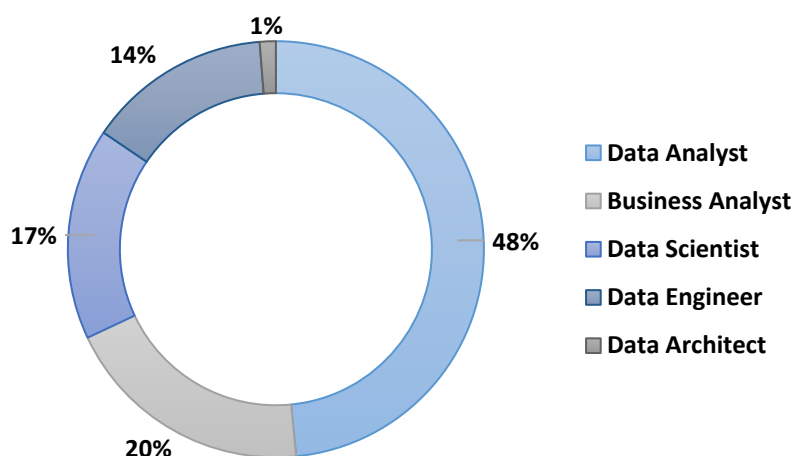

Figure 2. Proportion of vacancies for each data job category in Mexico in January 2022 (own elaboration).

In addition, from figure 2, it is also possible to observe that Business Analysts, Data Scientists and Data Engineers positions account for the other half of the data jobs demand in Mexico at that moment. On the contrary, Data Architects account for just 1% of the demand, which suggest that said data job profile is still not well developed in the Mexican labor market and that their functions are fulfilled by Data Engineers.

On the other hand, as expectable, most of the data jobs demand is located in Ciudad de México, Estado de México, Nuevo León and Jalisco, the most important economic poles in the country (figure 3).
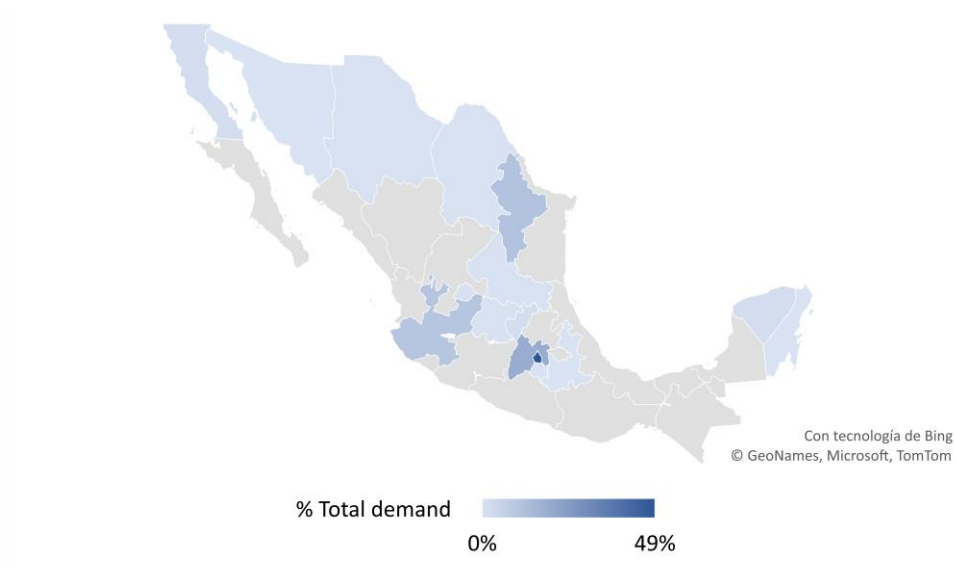
Figure 3. Data jobs demand per state in Mexico in January 2022 (own elaboration).

In view of figure 3, it is observed that the jobs demand is greater in Ciudad de México and Estado de México with the 49% and 19%, respectively, of the data jobs demand in the country. Therefore, the combined demand in Ciudad de México and Estado de México accounted for 68% of the vacancies, and the combined demand in the four most important states accounted for 90%.

Furthermore, the data jobs demand per category in Ciudad de México, Estado de México, Nuevo León and Jalisco is shown in the figure 4.
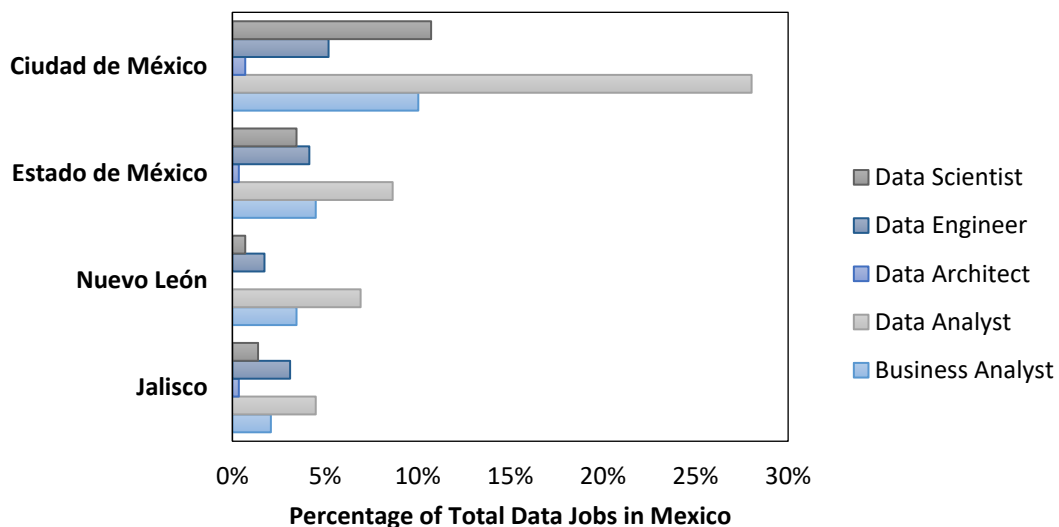

Figure 4. Data jobs demand per category in Ciudad de México, Estado de México, Nuevo León and Jalisco in January 2022 (own elaboration).

From figure 4, Data Analyst's, Data Scientist's and Business Analyst's jobs demand in Ciudad de México is almost the triple than the one in Estado de México, Nuevo León and Jalisco. In this sense, Data Analyst's, and Data Scientist's demand in Ciudad de Mexico accounts for 28% and 11%, respectively, of all the data jobs in the country. Thus, undoubtedly, Ciudad de México is the location where the most part of the data jobs are concentrated.

Regarding the companies with the greatest demand of data jobs, the top 5 are shown in the table 1.

Table 1. Top 5 companies with the greater demand of data jobs in Mexico in January 2022.

Company	Data Job Vacancies
Grupo Salinas	6
Manpower	5
Softtek	4
Banamex	4
Axa	4

So, Grupo Salinas is nowadays the biggest seeker of data skills in the Mexican labor market, along with Softtek, Banamex and Axa. In the case of Manpower, it is reasonable to think that their data job vacancies are published on behalf of other companies.

3.3 Salary analysis

Overall, the average salary of the data jobs in Mexico in January 2022 was **$ 30,980 MXN** per month. The distribution of the salary is shown in the figure 5.
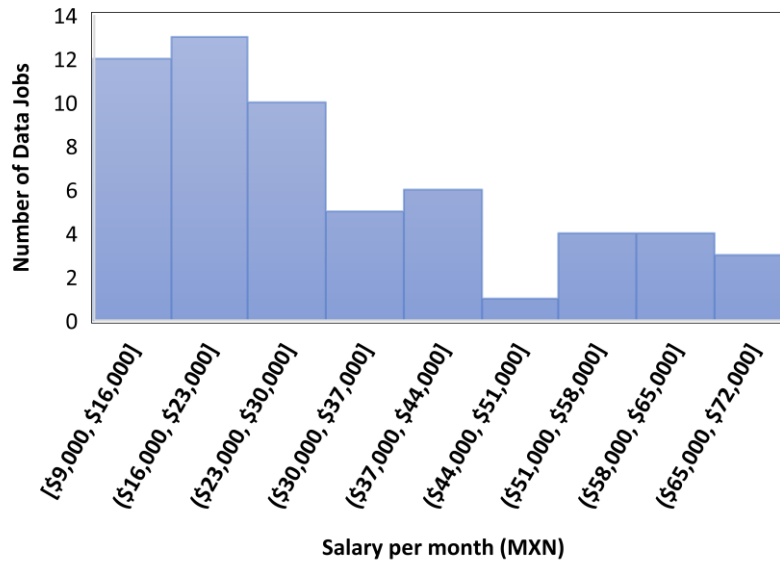
Figure 5. Histogram of salary distribution of the data jobs in Mexico in January 2022 (own elaboration).

From figure 5, it is possible to see that the salary distribution is not normal. Conversely, most of the salaries are found in the range from $9,000 to $30,000.

Regarding the salaries for each data job category, they are shown in table 2.

Table 2. Average salary for each data job category in Mexico in January 2022.

Data job category	Average salary
Data Engineer	$45,722.22
Data Scientist	$39,375.04
Business Analyst	$31,192.31
Data Analyst	$17,879.57
Data Architect	N/A
Total general	**$30,980**

From table 2 it can be seen that Data Engineers enjoy the highest average salary with $45,722 MXN, followed by Data Scientists with $39,375 MXN and Business Analysts with $31,192.31 MXN. On the other hand, on the bottom, Data Analysts have an average salary of $17,879.57 MXN. Unfortunately, no salary information could be retrieved for Data Architects.

Thus, even though Data Analysts are the most demanded position nowadays, it is the data job with the lowest salary.

The distribution of salaries for each data job category is shown in the box plot in figure 6.
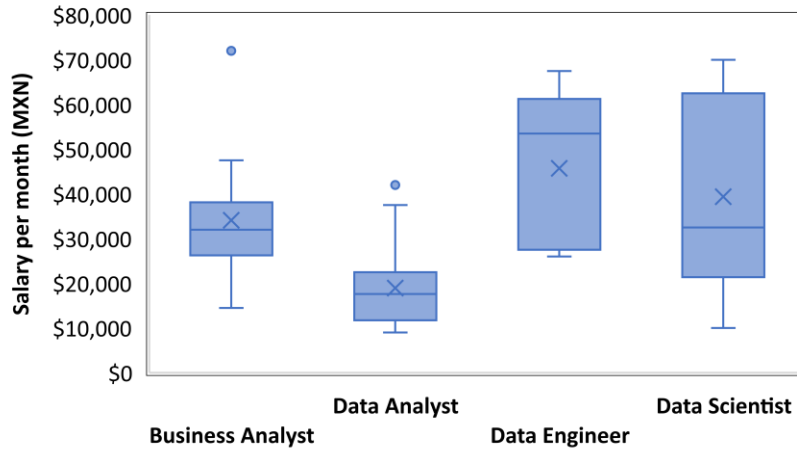
Figure 6. Box plot of salary distributions for each data job category in Mexico in January 2022 (own elaboration).

From figure 6, it is possible to see that no salary distribution for each data job category is normal. Furthermore, it is possible to see that salary distributions for Data engineers and Data scientists overlap, which raises the question whether the salaries for Data Engineers are actually higher than those for Data Scientists.

Thus, a one-way ANOVA was carried out to determine whether the average salary differences for the distinct data jobs categories were statistically significant. The hypotheses are as follows:

$$H_0: \quad \mu_{BA} = \mu_{DA} = \mu_{DE} = \mu_{DS}$$

$$H_1: \quad Not\ H_0$$

$$\alpha = 0.05$$

Where μ_{BA} is the average salary for Business Analysts; μ_{DA}, average salary for Data Analysts; μ_{DE}, average salary for Data Engineers; and μ_{DS}, average salary for Data Scientists.

The results of the ANOVA are shown in the table 3.

Table 3. ANOVA for the average salary differences from the distinct data jobs categories.

Origin of variations	Sum of squares	Degrees of freedom	Mean squares	F	Probability	Critical value for F
Between groups	6529634026	3	2176544675	11.93	4.69E-06	2.78
Within groups	9491105928	52	182521268			
Total	16020739954	55				

Thus, from table 3, it is observed that the average salary differences for the distinct data jobs categories are statistically significant, as the mean square between groups is greater than the estimated population variance within groups. This is reflected in the magnitude of the F statistic (11.93) over its critical value (2.78), which indicates that the statistical test falls under the rejection region; and that the probability of obtaining the observed outcome would be very unlikely under the null hypothesis (p-value equals 4.69E-06).

Then, a series to T-test were performed to determine whether the average salary for different category pairs was significantly different. In first place, the average salary for Data Engineers and Data Scientists was compared under the following hypothesis:

$$H_0: \quad \mu_{DE} - \mu_{DS} \leq 0$$

$$H_1: \quad \mu_{DE} - \mu_{DS} > 0$$

$$\alpha = 0.05$$

The results of the T-test are shown in the table 4.

Table 4. T-test of the average salary differences for data engineers and data scientists.

	Data Engineer	Data Scientist
Means	45722.2	39375.0
Variance	309381944.4	454321193.2
Observations	9	12
Hypothetical difference of means	0	
Degrees of freedom	19	
T statistic	0.747	
P(T<=t) one tail	0.232	
Critical value of t (one-tailed)	1.729	

Therefore, from table 4 it is observed that the average salary differences for the Data Engineers and Data Scientists were not statistically significant, as the T statistic (0.747) were smaller than the critical value of t (1.729). This is reflected in the p-value of 0.232, which indicates that the probability of obtaining the observed outcome is not unlikely under the null hypothesis at the selected level of significance. In other words, Data Engineer's salaries are not significantly higher than those for Data Scientists.

Following the above logic, the results of the other T-tests are shown in the table 5, using the same signification level of $\alpha = 0.05$.

Table 5. T-test results of the average salary differences for other data job categories.

Data jobs		T statistic	Critical value of t (one-tailed)	P(T<=t) one tail	Null hypothesis rejected?
Data Scientist	Data Analyst	3.392	1.782	0.005	Yes
Data Scientist	Business Analyst	1.233	1.753	0.118	No
Business Analyst	Data Analyst	4.579	1.720	8.141E-05	Yes
Data Engineer	Data Analyst	4.598	1.833	6.468E-04	Yes
Data Engineer	Business Analyst	2.282	1.796	0.021	Yes

In view of the table 5, Data Scientist's salaries are significantly higher than those for Data Analysts but not significantly higher than those for Business Analysts. Similarly, Business Analyst's salaries are significantly higher than those for Data Analysts. Furthermore, Data Engineer's salaries are significantly higher than those for both Business Analysts and Data Analysts.

All in all, average salaries for Data Engineers and Data Scientists are the highest ones in the current Mexican Labor market according to the data collected.

Regarding the assessment of locations where the highest salaries for data jobs could be found, due to the concentration of the job demand in a few Mexican States, it was only possible to estimate the average salaries for some data job categories in Ciudad de México, Estado de México, and Nuevo León (figure 7).
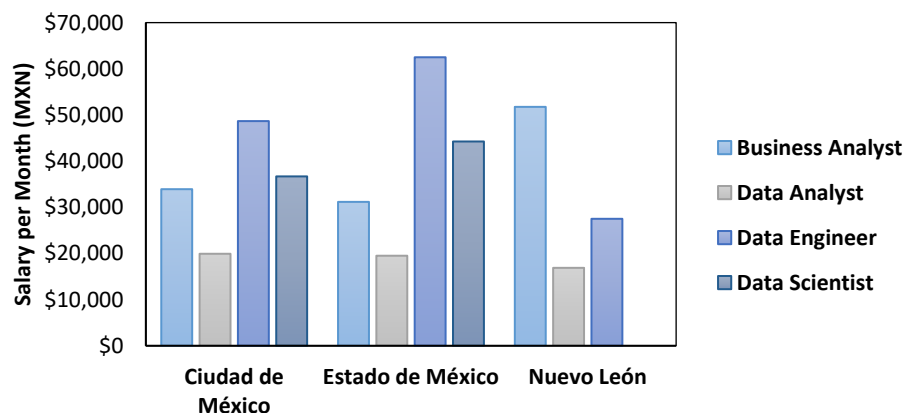

Figure 7. Average salaries for each data job category in Ciudad de México, Estado de México, and Nuevo León in January 2022 (own elaboration).

From figure 7, salaries for Business Analysts are higher in Nuevo León ($51,750) and lower in Estado de México ($31,167); salaries for Data Analysts are higher in Ciudad de México ($19,925) and lower in Nuevo León ($16,875); salaries for Data Engineers are higher in Estado de México ($62,500) and lower in Nuevo León ($27,500); and salaries for Data Scientists are slightly higher in Estado de México ($44,250) and lower in Ciudad de México ($36,688). So, overall, data jobs salaries are higher in Ciudad de México ($32,128) and Estado de México ($33,667).

Nonetheless, the above results must be taken with caution as not enough data was possible to collect to calculate a meaningful average salary for each data job category in each state. However, is seems to be reasonable that the highest salaries could be found in the capital and its surroundings.

Finally, in regard to the companies offering the highest salaries, the figure 8 shows the top ten companies with the highest salaries for all data jobs in Mexico.
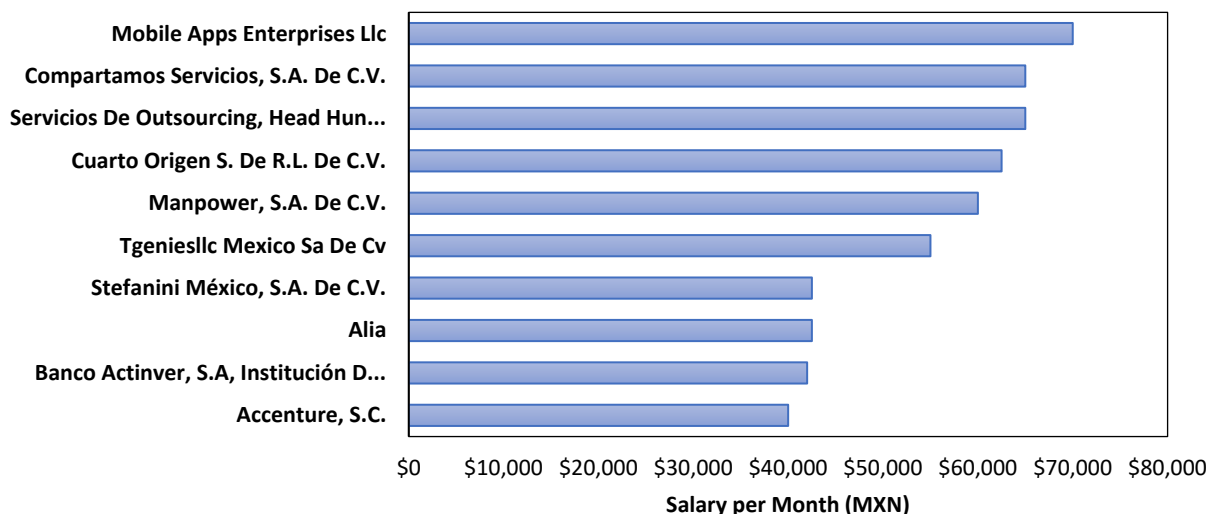

Figure 8. Companies with the highest average salaries for data jobs in Mexico (own elaboration).

Similarly, in the case of Servicios de Outsourcing and Manpower, it is reasonable to consider that their data job vacancies are published on behalf of other companies.

Again, the results from figure 8 must be taken with caution, as most likely they refer to senior data positions. Nevertheless, they provide an insight about what companies offer the highest salaries for data jobs in the Mexican labor market.

4. Conclusions

The data job categories with the highest salaries in the Mexican labor market in January 2022 according to the OCC website were Data Engineer and Data Scientist. Even though the average salary for Data Engineer was higher than that for Data Scientists, it was found that the difference was not statistically significant.

Moreover, Ciudad de México and Estado de México were the states where it is possible to find the greatest jobs demand and the highest salaries.

Furthermore, the companies with the greatest demand of data positions were Grupo Salinas, Softtek and Banamex; nevertheless, the organizations that offered the highest salaries were Mobiles Apps Enterprises, Compartamos Servicios and Cuarto Origen.

Finally, regarding the limitations of the present study, it is important to bear in mind that the data was collected solely from the OCC website and only for a very short period of time, collected data mostly correspond to Ciudad de México, Estado de México, Nuevo León and Jalisco, no salary data was retrieved for Data Architects and no distinction was made among entry level, middle and senior positions.

Thus, as future perspectives, it would be advisable to gather data from more job websites, retrieve information for a longer time span, collect more salary data for other Mexican states, identify tentative salaries for Data Architects in Mexico and make a distinction among entry level, middle and senior positions.

5. References

Axis Talent. (2020). *The Ecosystem of Data Jobs - Making sense of the Data Job Market.* Retrieved from Axis Talent: https://www.axistalent.io/blog/the-ecosystem-of-data-jobs-making-sense-of-the-data-job-market

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review.*

Grus, J. (2019). *Data Science from Scratch.* Sebastopol: O'Reilly Media.

Kudyba, S. (2014). Information Creation through Analytics. In S. Kudyba, *Big Data, Mining, and Analytics* (pp. 19-47). Boca Raton: Auerbach Publications.

Rollins, J. B. (2015). *Metodología Fundamental para la Ciencia de Datos.* Somers: IBM Corporation. Retrieved from https://www.ibm.com/downloads/cas/WKK9DX51