

# DATA JOBS SALARIES IN MEXICO IN FEBRUARY 2023



**Daniel Eduardo  
López**

© 2023

Daniel Eduardo López

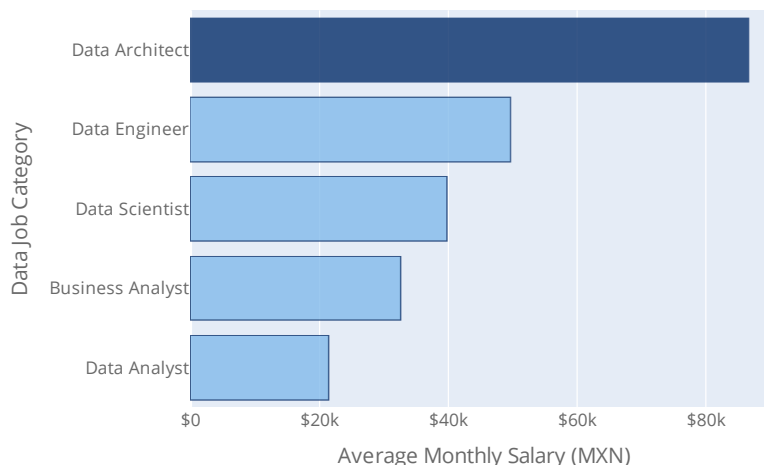
ALL RIGHTS RESERVED

Cover Picture Credit: © Lukas: <https://www.pexels.com/es-es/foto/foto-de-primer-plano-del-portatil-gris-577210/>

## Executive Summary

With the emergence of the big data, novel 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 February 2023 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 scraping. Then, the data was explored, prepared, analyzed and visualized using Python 3 and its libraries. It was found that Data Architects enjoyed the highest earnings with an average monthly salary of about 86,562 MXN, respectively, whereas the Data Analyst positions had the lowest ones with an average monthly salary of about 21,240 MXN (figure 1).



**Figure 1.** Average monthly salary for each data job category in Mexico in February 2023 (own elaboration).

Indeed, the average salary for Data Architects was significantly higher than that for Data Scientists, being the mean salary difference between those two data job categories of 41,875 MXN. However, it is important to bear in mind that the observations of salaries for Data Architect positions were scarce and might not be fully representative.

Moreover, it was observed that Ciudad de México was the location where the greatest number of data jobs were concentrated along with the highest salaries. And, finally, it was identified that the companies with the greatest demand of data positions were *Manpower*, *Atento Servicios*, *Praxis* and *Softek*; and the ones offering the highest salaries were *Especialistas en Talento*, *Zemsania México*, *Resources Connection México*, *Zegovia RH*, and *Reclutamiento en Tecnología*.

# Table of Contents

<b>Executive Summary.....</b>	<b>1</b>
<b>Table of Contents .....</b>	<b>2</b>
<b>1. Introduction.....</b>	<b>3</b>
<b>2. Methodology .....</b>	<b>4</b>
2.1 Objective.....	4
2.2 Research Question .....	4
2.3 Hypothesis .....	4
2.4 Analytical Approach .....	5
2.5 Data Requirements .....	5
2.6 Data Collection.....	5
2.7 Data Understanding and Preparation.....	5
2.8 Data Visualization .....	5
2.9 Statistical Analysis .....	6
<b>3. Results and Discussion .....</b>	<b>7</b>
3.1 Data Collected .....	7
3.2 Data Jobs Demand Analysis .....	7
Most demanded data job.....	7
Location where most data jobs are demanded .....	8
Data jobs demand per state .....	9
Companies demanding more data jobs .....	11
Data jobs demand per company.....	12
Location of the top companies demanding data jobs.....	13
3.3 Salary Analysis .....	14
Salaries distribution.....	14
Data job with the highest salary.....	18
Salaries per location.....	24
Companies offering the highest salaries.....	26
3.4 Dashboard .....	28
<b>4. Conclusions .....</b>	<b>29</b>
<b>5. References .....</b>	<b>30</b>

# 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 the next 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 databases, 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 cope 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 the present moment (February 2023).

## 2. Methodology

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

All data and calculations can be found in the correspondent [GitHub repository](#).

### 2.1 Objective

#### General objective

To identify which data job category has the highest salary in the Mexican labor market in February 2023 according to the OCC website.

#### Specific objectives

- To find what is the most demanded data job.
- To identify in which Mexican state the demand of data jobs is the greatest.
- To explore the data jobs demand per Mexican state.
- To identify what companies demand the greatest number of data jobs.
- To explore the data jobs demand per company.
- To explore the location of the companies demanding data jobs.
- To assess the statistical significance of the salary differences for each data job category.
- To identify in 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 February 2023 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 February 2023.

## 2.4 Analytical Approach

Descriptive and inferential statistics were used in order answer the research question (see 2.9 *Statistical analysis* below).

## 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 07 February 2023, through web scraping with Python 3 and its libraries Selenium and BeautifulSoup. The OCC website was selected due to their more lenient policy for automated process of information retrieval and its importance as a platform for both recruiters and job seekers in Mexico.

## 2.7 Data Understanding and Preparation

Data was explored and prepared with Python 3 and its libraries Numpy, Pandas, Matplotlib and Seaborn.

## 2.8 Data Visualization

Data was visualized with Python 3 and its libraries Matplotlib, Seaborn, Folium, and Plotly. In this sense, a variety of figures such as pie charts, choropleth maps, bar charts, heatmaps, treemaps, histograms, lollipop charts and boxplots were used to extract insights from the data.

Furthermore, an [interactive Dashboard](#) was built using Plotly and Dash, and it was deployed on Render.

## 2.9 Statistical Analysis

Data was analyzed through descriptive and inferential statistics in Python 3 and its libraries Pandas, Scipy, Statsmodels and Scikit-learn. Average salaries were obtained for each data job category. Furthermore, a box plot and histograms were constructed to visualize the distribution of the salaries, overall and across the data job categories.

Both parametric (ANOVA and t-test with unequal variance) and non-parametric (Kruskal-Wallis H and Mann-Whitney U) tests were carried out to assess the significance of the obtained results.

Firstly, a D'Agostino-Pearson normality test (omnibus) was carried out to assess the normality of the data jobs salary distribution.

Then, a one-way analysis of variance (ANOVA) procedure and a Kruskal-Wallis H test were carried out to detect whether the average salary differences for the distinct data jobs categories were statistically significant.

After that, pairwise t-tests for two independent samples assuming unequal variances and Mann-Whitney U tests were used to assess whether the average salary for different data jobs categories was significantly different.

Finally, an effect size analysis was carried out by computing the absolute mean salary differences, the Cohen's d, and the bootstrap confidence intervals for the mean for each data job category. This, in order to assess whether the salary differences are significant from a practical point of view.



### 3. Results and Discussion

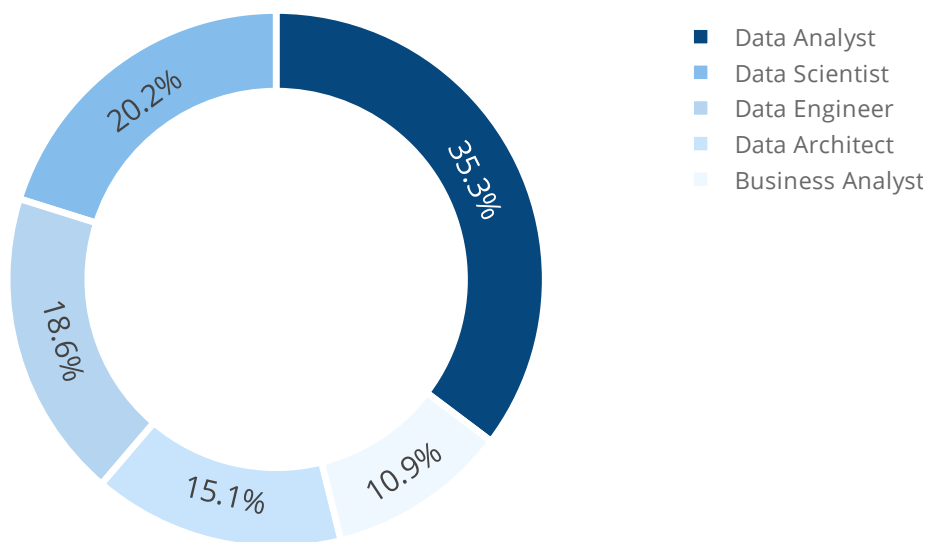
#### 3.1 Data Collected

Data collection and cleaning allowed to identify 258 data jobs in Mexico in February 2023 from the OCC Website, of which 105 had salary information (40.7%). 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

##### Most demanded data job

From the whole universe of data jobs identified, more than a third of the data jobs (35.3%) correspond to positions of **Data Analyst**, rendering them as **the most demanded** ones in the Mexican labor market at the time of this study (figure 2).



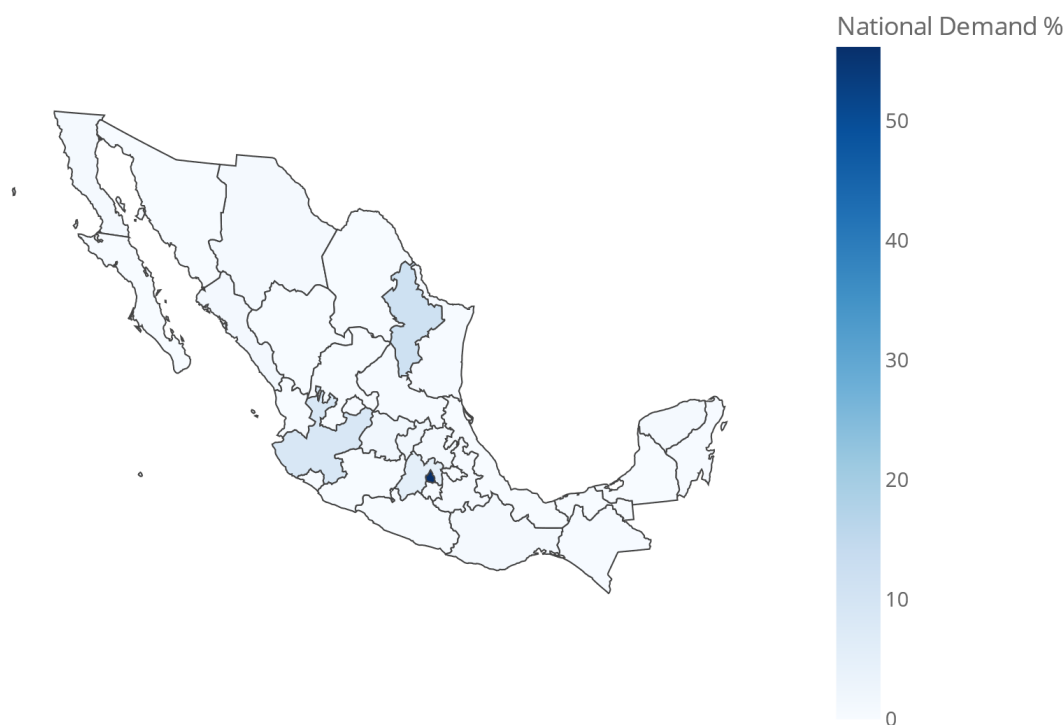
**Figure 2.** Proportion of vacancies for each data job category in Mexico in February 2023 (own elaboration).

In addition, from figure 2, it is also possible to observe that Data Scientists (20.2%) and Data Engineers (18.6%) positions account for the other half of the data jobs demand in Mexico.

On the contrary, Business Analyst positions are the less demanded, with only a 10.9% out of the total.

### Location where most data jobs are demanded

On the other hand, as expectable, most of the data jobs demand is located in **Ciudad de México, Nuevo León, Jalisco, and Estado de México**, the most important economic poles in the country (figure 3).



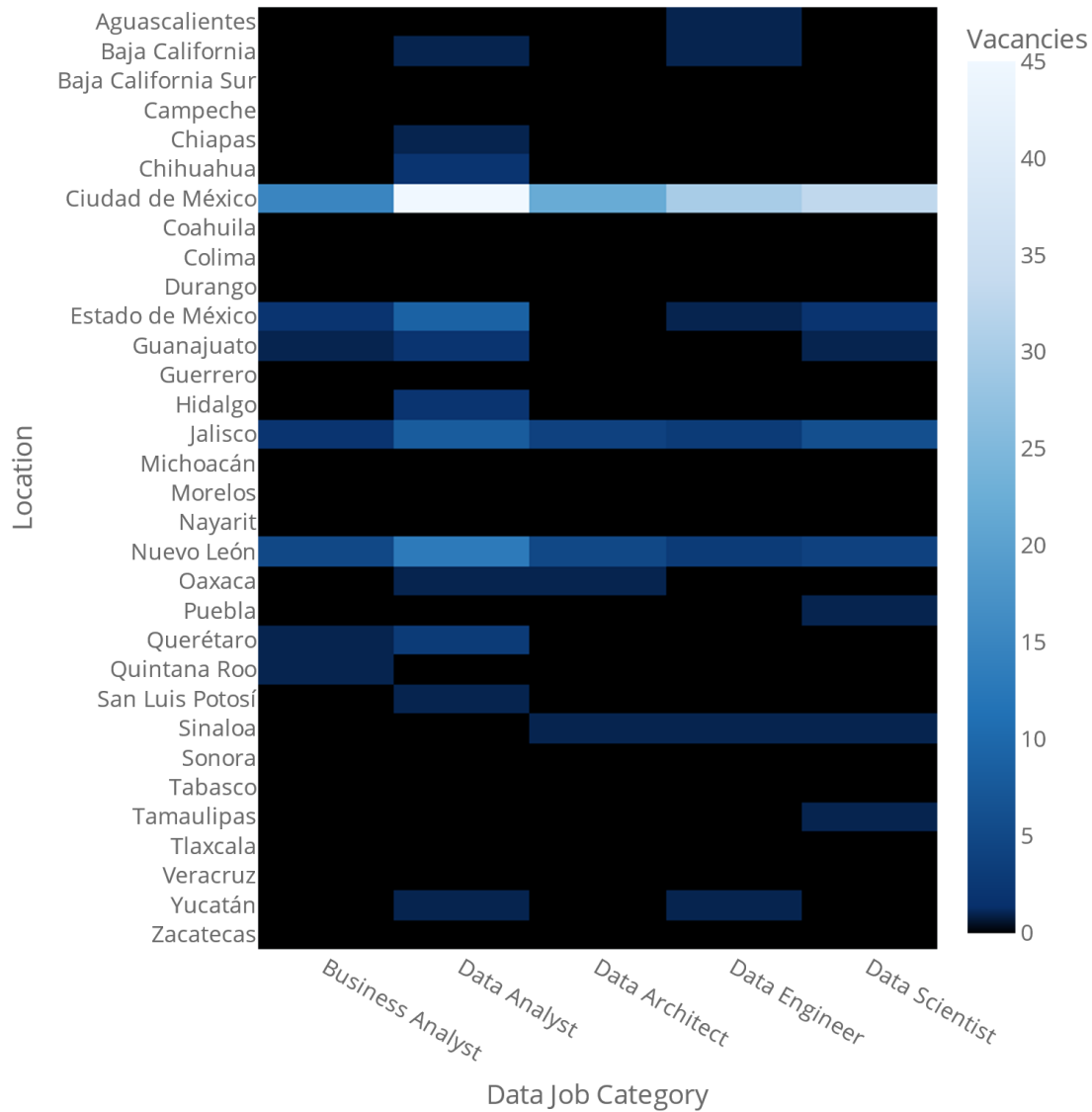
**Figure 3.** Data jobs demand per state in Mexico in February 2023 (own elaboration).

In view of figure 3, it is observed that the data jobs demand is **highly concentrated** in **Mexico City** (“Ciudad de México”, in Spanish) with about the **56% of the total national demand**. On the other hand, Nuevo León represented a distant second place with about the 11.6% of the demand. Finally, Estado de México and Jalisco accounted for about the 8.9% and 5.4% of the demand, respectively.

Therefore, the combined demand in Ciudad de México, Nuevo León, Jalisco, and Estado de México accounted for 82% of the vacancies, which suggest the extremely high level of concentration of the data jobs demand in the country and the lag in terms of technology in the rest of the country.

## Data jobs demand per state

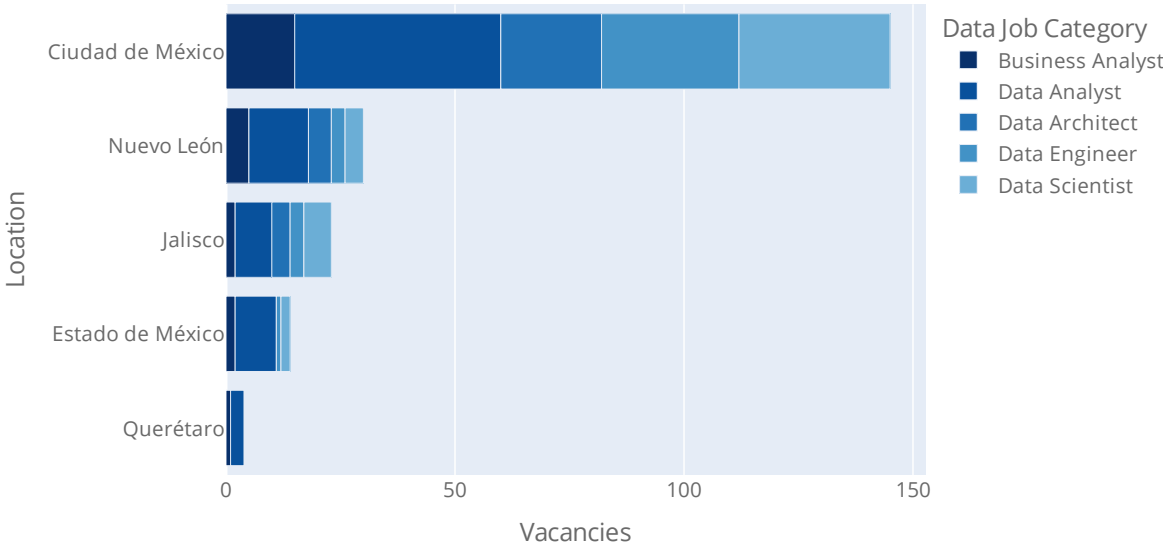
The data jobs demand per state is shown in the figure 4.



**Figure 4.** Data jobs demand per category and location in February 2023 (own elaboration).

From the plot above, it is possible to observe that, undoubtedly, **Ciudad de México**, **Nuevo León**, **Jalisco**, and **Estado de México** are the locations where most of the data jobs are demanded, while the rest of the country is lagging in terms of data jobs creation. Moreover, it is noteworthy that the **Data Analyst** position is the one most demanded across the Mexican states along with **Data Engineer** and **Data Scientist** positions, whereas **Data Architect** positions are highly concentrated in Ciudad de México, Nuevo León and Jalisco.

As shown above, the data jobs demand is largely concentrated in **Ciudad de México**, whereas **Nuevo León**, **Jalisco**, and **Estado de México** are distant second, third and fourth places, respectively. In this sense, the data jobs demand per category in those locations, the top 5 states with the highest demand, was further explored (figure 5).



**Figure 5.** Data jobs demand per category in Ciudad de México, Estado de México, Nuevo León, Jalisco and Querétaro in February 2023 (own elaboration).

From figure 5, data jobs demand in Ciudad de México exceeds by far the demand in any other state. In this sense, Data Analyst’s, Data Scientist’s, and Data Engineers’ demand in Ciudad de México accounts for 17.4%, 12.7%, and 11.6%, respectively, of all the data jobs in the country.

On the other hand, considering the national demand of data jobs per category, the demand in Ciudad de México and its weight is shown in the table 1.

**Table 1.** Data jobs demand in Ciudad de México and its weight in the national demand.

Data Job Category	Demand in Ciudad de México (Vacancies)	National Demand (Vacancies)	Demand in Ciudad de México (% of National Demand)
Business Analyst	15	28	53.6%
Data Analyst	45	91	49.5%
Data Architect	22	39	56.4%
Data Engineer	30	48	62.5%
Data Scientist	33	52	63.5%
All	145	258	56.2%

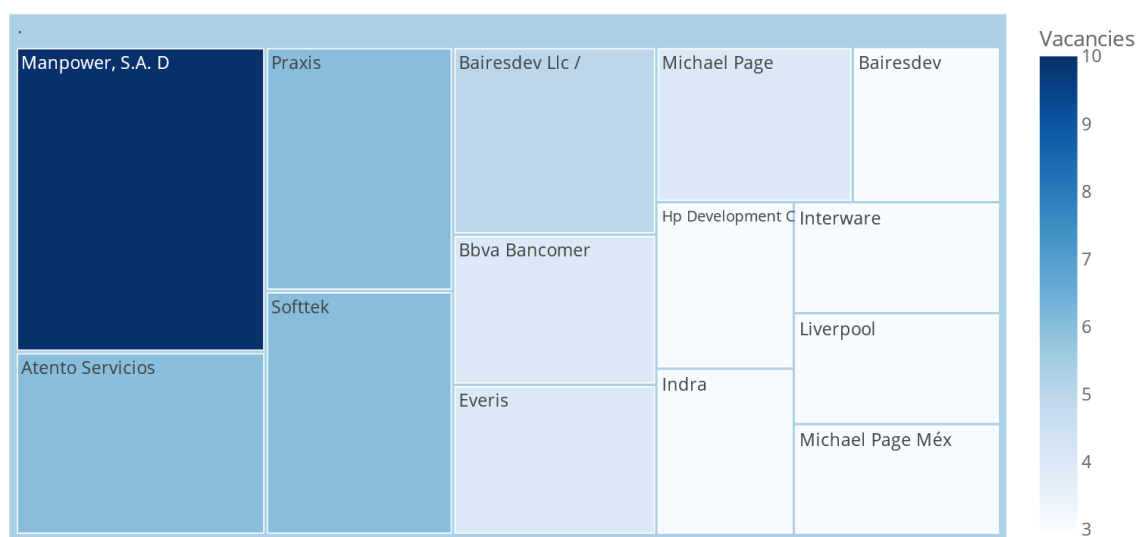
Thus, in view of the table 1, some positions such as **Data Scientist**, **Data Engineer**, and **Data Architect** are **highly concentrated in Ciudad de México**, with 63.5%, 62.5%, and 56.4% of the national demand for said categories, respectively. On the other

hand, **Data Analyst** positions are the ones more demanded outside Ciudad de México with just a 49.5% of the national demand.

On the other hand, it is noteworthy that much of the demand is comprised by Data Analyst positions in most of the states. However, there is also a significant demand of Business Analyst positions across the country. On the contrary, demand for Data Architect positions could only be found mostly in Ciudad de México, Nuevo León and Jalisco.

### Companies demanding more data jobs

Regarding the companies with the greatest demand of data jobs, the top 15 are shown in the figure 6.



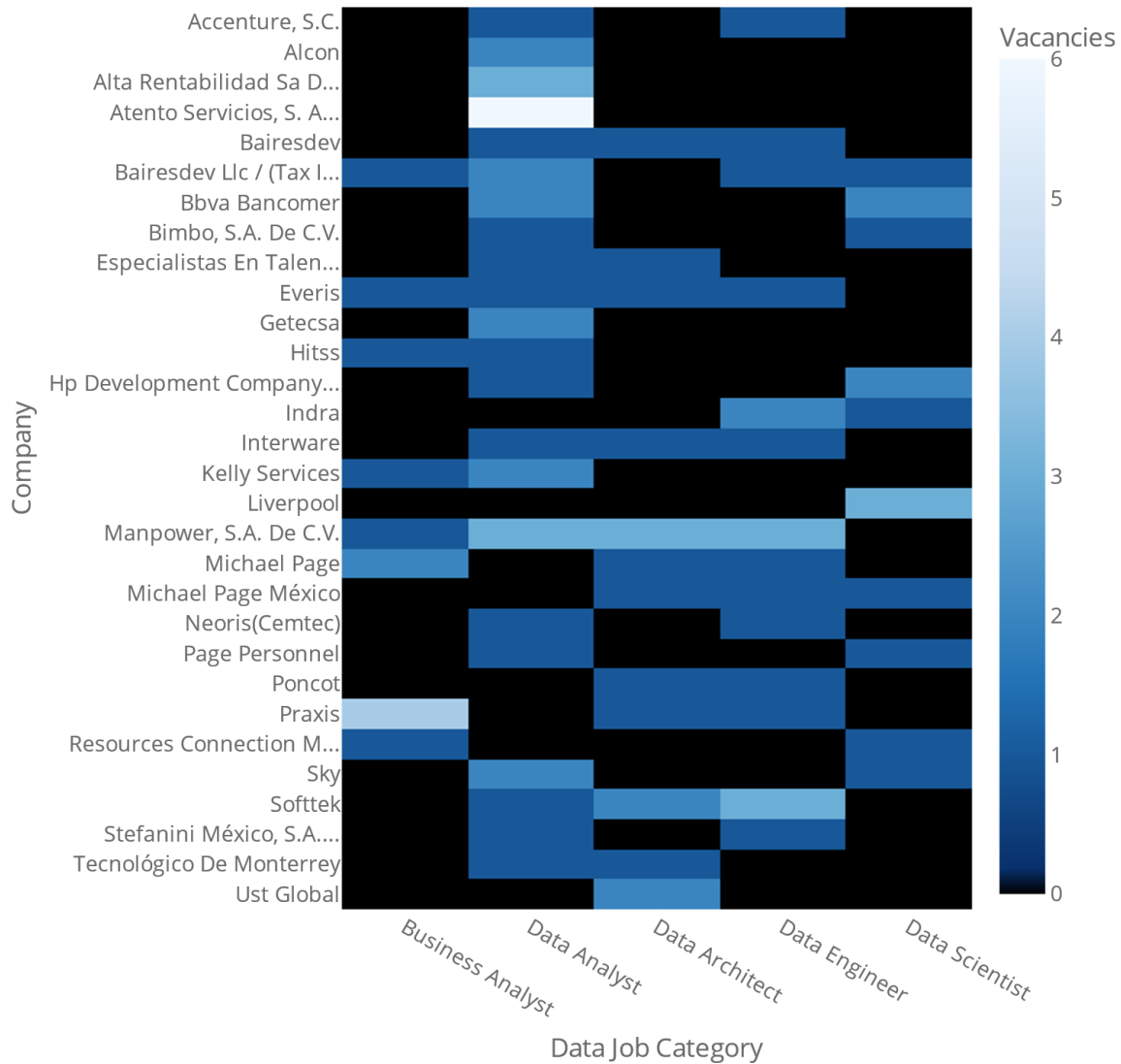
**Figure 6.** Top 15 companies with the highest demand of data jobs in Mexico in February 2023 (own elaboration).

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

On the other hand, **Praxis** and **Softtek** are tech consultancy firms. Other well-known organizations in the top 15 of companies demanding data jobs in Mexico are **BBVA**, **HP** and **Liverpool**. So, even though, the data jobs demand may vary over time, the current interest of such companies in data science and analytics might convert them in interesting prospects for job seekers.

## Data jobs demand per company

The data jobs demand per company in the top 30 companies with the highest data jobs demand is shown in the figure 7.

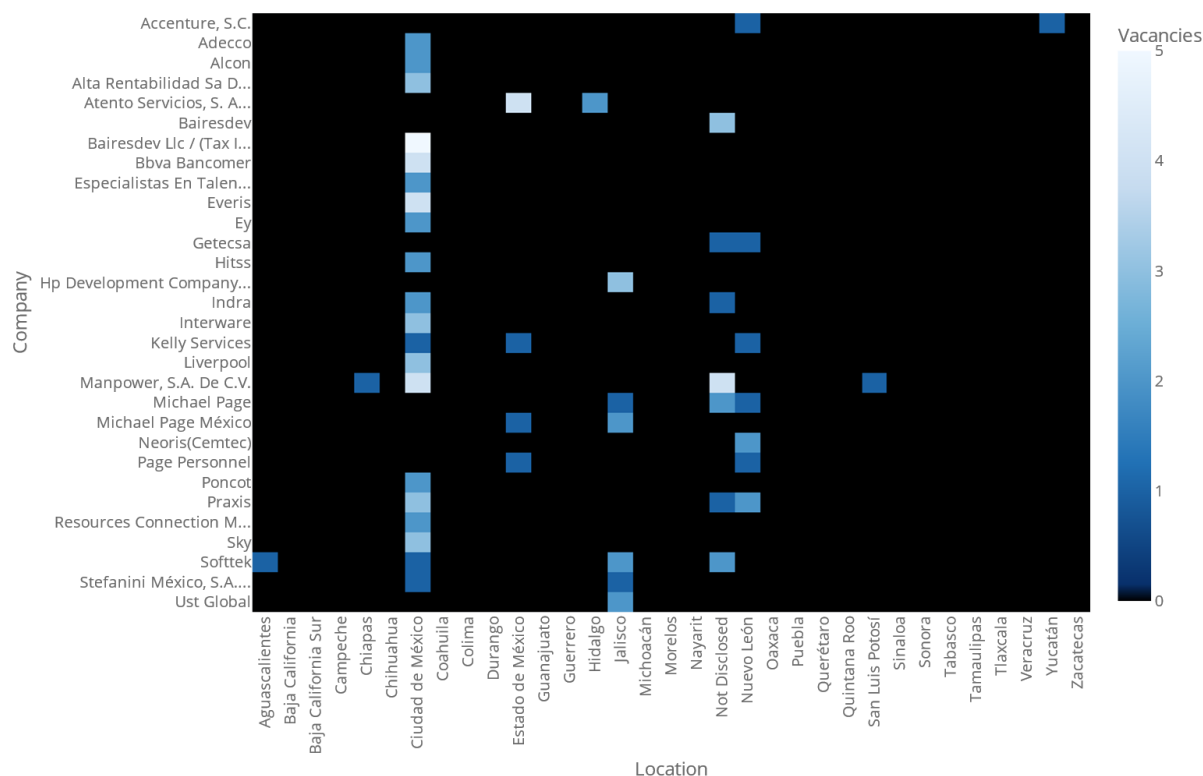


**Figure 7.** Demand per data job category in the top 30 companies with the highest data jobs demand in February 2023 (own elaboration).

The above heatmap suggests that **Data Analyst**, **Data Engineer** and **Data Scientist** positions are more demanded across different organizations. On the contrary, **Data Architect** vacancies are more likely to be found in more specific organizations like tech consultancy firms (*BairesDev*, *Everyis*, *Praxis*, *Softek*, etc.).

## Location of the top companies demanding data jobs

The location of the top 30 companies demanding data jobs is shown in the figure 8.

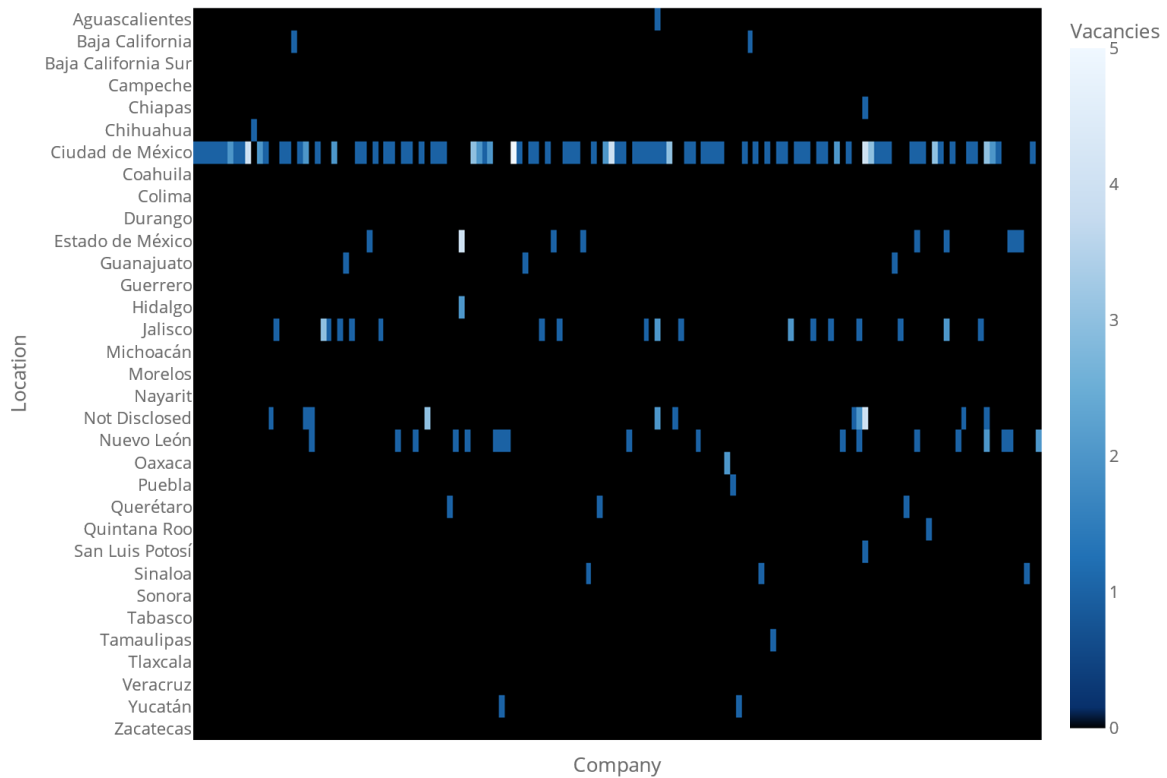


**Figure 8.** Top 30 companies demanding data jobs per location in February 2023 (own elaboration).

As expectable, most of the companies are located in **Ciudad de México** as the large majority of the vacancies are offered there. However, the heatmap shows that there are some organizations that spread across several Mexican states such as *Accenture*, *Praxis* or *Softek*.

Furthermore, there are few well-known companies whose data jobs demand is not located in the capital region, such as *HP* (Jalisco).

On the other hand, a heatmap showing all the companies demanding data jobs by location is shown in the figure 9.



**Figure 9.** Companies demanding data jobs per location in February 2023 (own elaboration).

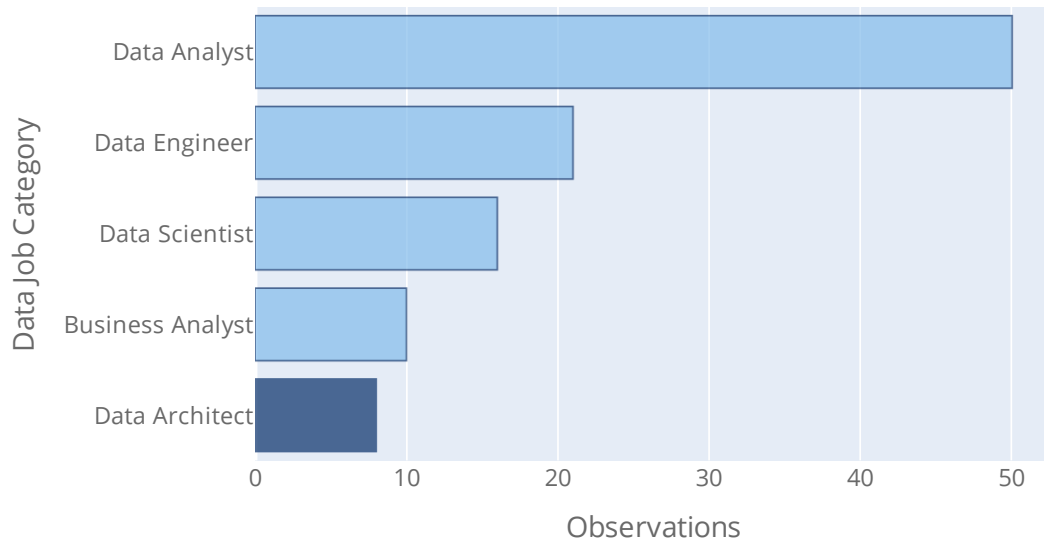
Likewise, the above heatmap beautifully shows that most of the data jobs are concentrated in **Ciudad de México** and, into a lesser extent, in **Nuevo León, Jalisco** and **Estado de México**.

### 3.3 Salary Analysis

#### Salaries distribution

For the salary analysis, the number of salary observations per data job category is shown in the figure 10.

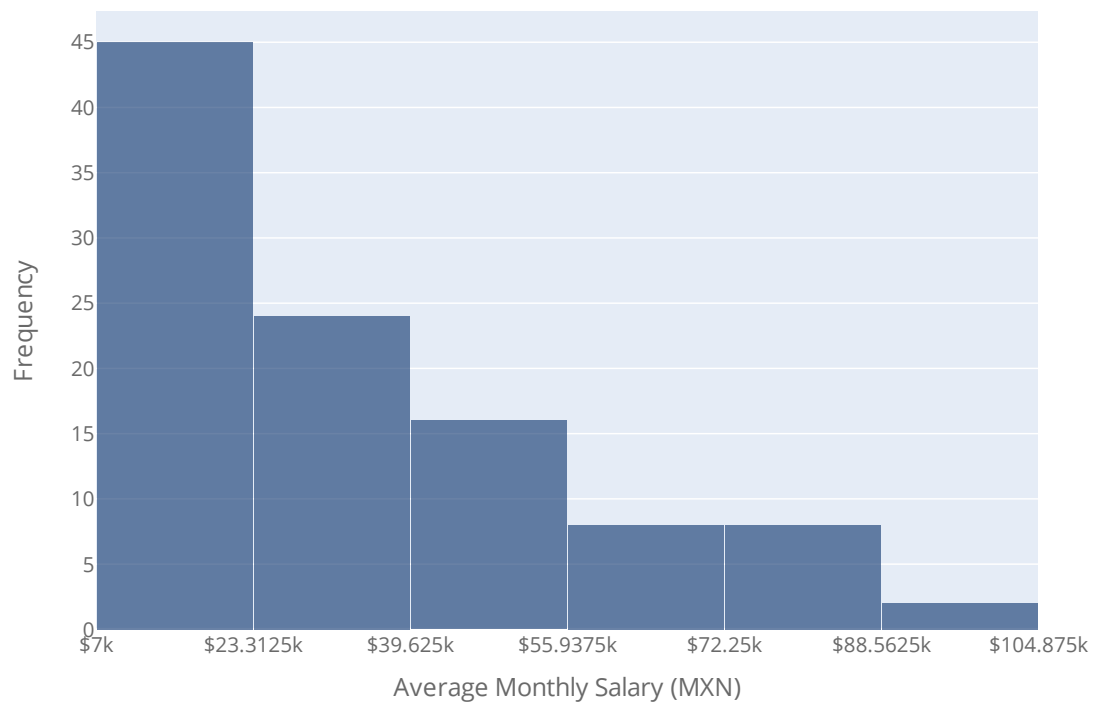




**Figure 10.** Number of salary observations per data job category (own elaboration).

In view of the figure 10, it is necessary to be cautious by interpreting the salary results for the Data Architect and Business Analyst positions, as very few observations were collected.

Overall, the average salary of the data jobs in Mexico in February 2023 was **35,757.79 MXN ( $SD = 26,219.92$ )** per month. The distribution of the salary is shown in the figure 11.



**Figure 11.** Histogram of salary distribution of the data jobs in Mexico in February 2023 (own elaboration).

From figure 11, it is noteworthy that the salary distribution is skewed to the right. So, it might not comply with the normality assumption. In this sense, **most of the salaries are found in the range from 7,000 to 23,312 MXN.**

In this context, a D'Agostino-Pearson normality test (omnibus) was carried out to assess whether the data jobs salary distribution complied with the assumption of normality. The hypotheses are as follows:

$$H_0: \text{Normal Distribution}$$

$$H_1: \text{Not } H_0$$

$$\alpha = 0.05$$

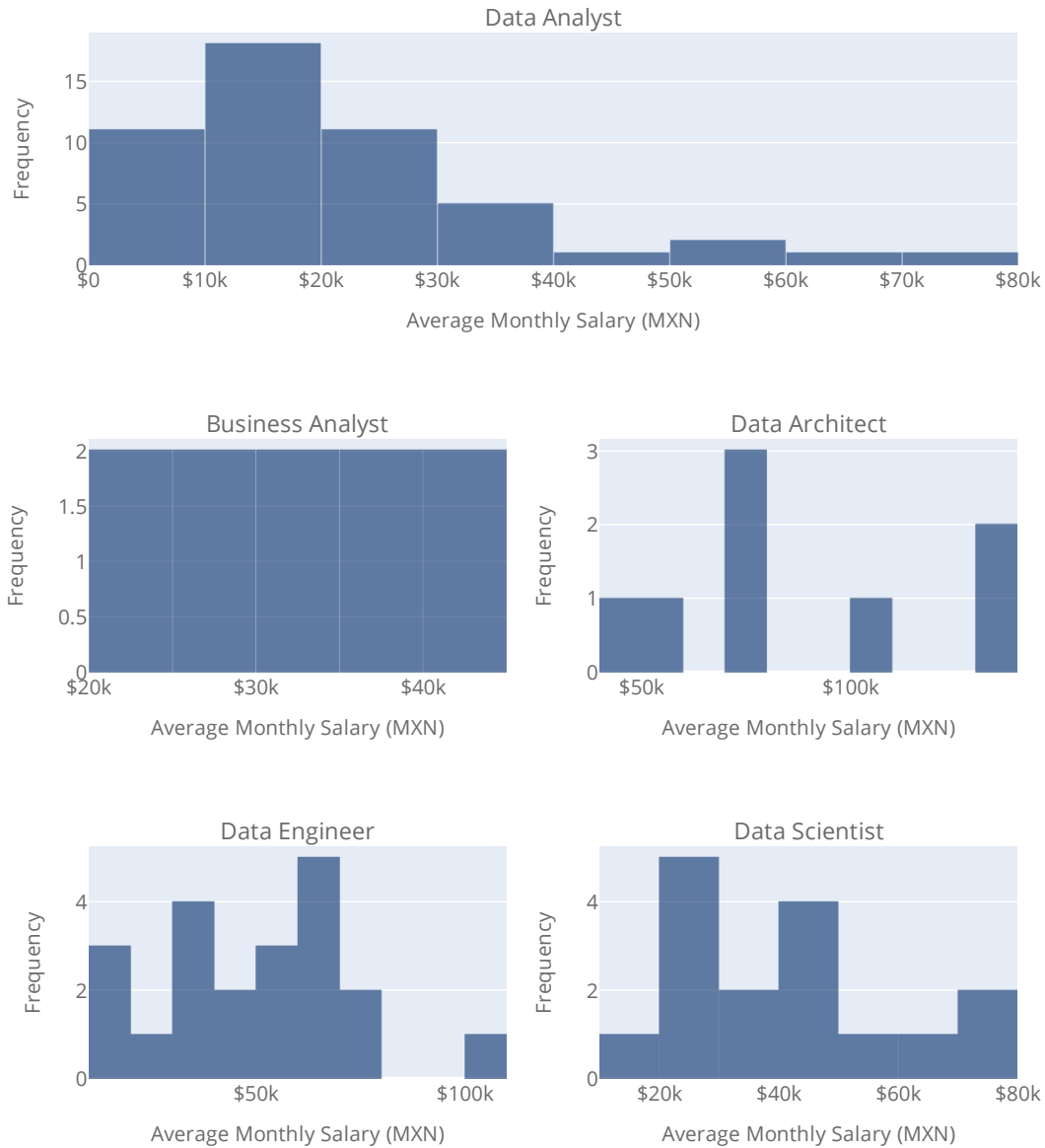
The results of the D'Agostino-Pearson normality test are shown in the table 2.

**Table 2.** D'Agostino-Pearson normality test results of the data jobs salary normality distribution at a signification level of  $\alpha = 0.05$ .

<i>K<sup>2</sup> statistic</i>	<i>p-value</i>
39.029	<0.001

As the  $p$ -value is significant ( $p < 0.05$ ), the null hypothesis that the sample comes from a normal distribution is **rejected**.

Indeed, only the salary observations for the **Data Scientist** and **Data Engineer** positions somewhat resemble a normal distribution, whereas it is not possible to state the same with the observations for the other data job categories (figure 12).

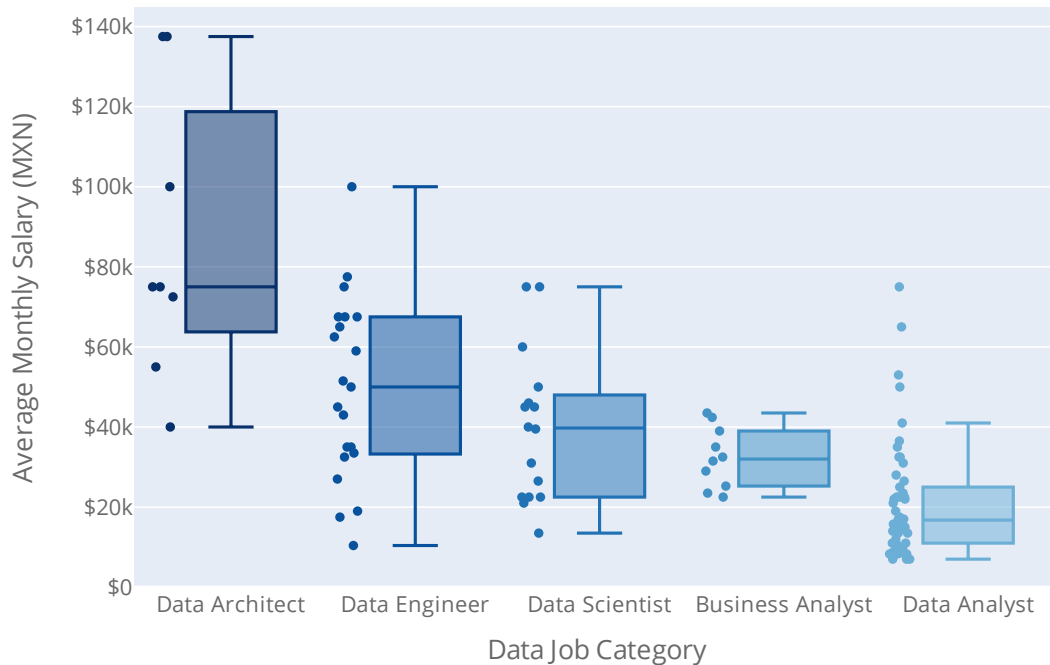


**Figure 12.** Histogram of salary distributions per data job category in Mexico in February 2023 (own elaboration).

Notwithstanding with the above, for the purposes of the present study, both parametric (ANOVA and t-test with unequal variance) and non-parametric (Mann-Whitney U and Kruskal-Wallis H) tests were carried out to assess the significance of the obtained results.

### Data job with the highest salary

The salaries distributions for each data job category are shown in the box plot (figure 13).



**Figure 13.** Box plot of salary distributions for each data job category in Mexico in February 2023 (own elaboration).

Figure 13 suggests that, after removing outliers from the calculation of the average salaries, **Data Architect** positions are still the ones with the highest salaries in the current Mexican labor market with a monthly median salary of about **75,000 MXN**.

As the salaries population likely do not follow a normal distribution as more low-income jobs are being offered than high-income positions, the median monthly salary figures could represent in a more accurate manner the central tendency for each data job category; however, the mean figures are also shown for sake of completeness and comparison (table 3).

**Table 3.** Mean and median salaries for each data job category in February 2023.

Data Job Category	Mean Monthly Salary (MXN)	Median Monthly Salary (MXN)
Data Architect	86,562	75,000
Data Engineer	49,567	50,000
Data Scientist	39,687	39,750
Business Analyst	32,415	32,000
Data Analyst	21,240	16,750

From table 3, it is possible to see that the mean and median figures are consistent for each data job category and, in this sense, Data Architect positions are the ones with the highest salaries.

Moreover, in order to verify whether the salary differences among the different data jobs categories were statistically significant, a one-way analysis of variance (ANOVA) procedure and a Kruskal-Wallis H test were performed. The hypotheses are as follows:

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

$$H_1: \text{Not } H_0$$

$$\alpha = 0.05$$

Where  $\mu_{BA}$  is the average salary for Business Analysts;  $\mu_{DA}$ , average salary for Data Analysts;  $\mu_{DR}$ , average salary for Data Architects;  $\mu_{DE}$ , average salary for Data Engineers; and  $\mu_{DS}$ , average salary for Data Scientists.

The results of the one-way ANOVA and the Kruskal-Wallis H test are shown in the table 4.

**Table 4.** ANOVA and Kruskal-Wallis H test results for the average salary differences for the different data jobs categories at a signification level of  $\alpha = 0.05$ .

Test	Statistic	p-value
One-way ANOVA	$F = 13.963$	$<0.001$
Kruskal-Wallis	$H = 27.973$	$<0.001$

Thus, from table 3, it is observed that the average salary differences for the distinct data jobs categories are statistically significant, as the probability of obtaining the observed outcome would be very unlikely under the null hypothesis ( $p$ -value is less than 0.001 for both tests). Thus, the null hypothesis that the averages salaries are equal is **rejected**.

Then, a series of pairwise t-tests with unequal variance (Welch's tests) and Mann-Whitney U tests were performed to determine whether the average salary for different data jobs categories was significantly different. In first place, the average salaries for Data Architects (DR) and Data Scientists (DS) were compared under the following hypothesis:

$$H_0: \mu_{DR} \leq \mu_{DS}$$

$$H_1: \mu_{DR} > \mu_{DS}$$

$$\alpha = 0.05$$

The results of the t-test and the Mann-Whitney U test are shown in the table 5.

**Table 5.** T-test and Mann-Whitney U test results of the average salary differences for Data Architects and Data Scientists at a signification level of  $\alpha = 0.05$ .

Test	Statistic	p-value
t-test	$t = 2.864$	0.013
Mann-Whitney U test	$U = 74.500$	0.004

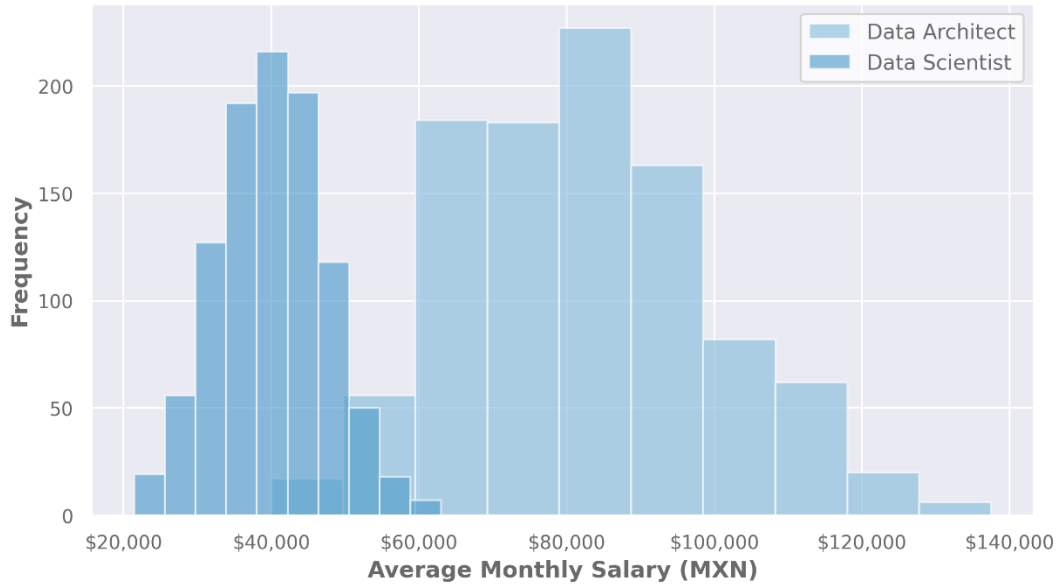
Therefore, from table 4 it is observed that the average salary differences for the Data Architects and Data Scientists were statistically significant, as the  $p$ -value of both tests indicates that the probability of obtaining the observed outcome is very unlikely under the null hypothesis at the selected level of significance. Therefore, the null hypothesis is rejected. In other words, **Data Architect's salaries are significantly higher than those for Data Scientists.**

Then, the difference between the mean salaries for each data job category were calculated in order to quantify the size of the effect as well as the confidence intervals for the mean salaries. However, as the salary observations do not follow a Gaussian distribution, the confidence intervals for the mean were estimated using bootstrapping and the effect size was primarily calculated in absolute terms. Nonetheless, for sake of completeness, Cohen's  $d$  was also estimated. The results are shown in the table 6.

**Table 6.** Absolute mean difference, Cohen's  $d$  and bootstrap confidence intervals for the mean for Data Architect and Data Scientist salaries at a signification level of  $\alpha = 0.05$ .

Effect size	Value
Absolute mean difference	41,875 MXN
Cohen's $d$	$d = 1.73$ (Large effect size)
95% Confidence interval for the mean for DR salaries	(51 667, 116 667) MXN
95% Confidence interval for the mean for DS salaries	(26 925, 54 002) MXN

Furthermore, the bootstrap distributions for the Data Architect and Data Scientist mean salaries are shown in the figure 14.



**Figure 14.** Bootstrap distributions for the Data Architect and Data Scientist mean salaries (own elaboration).

So, from the effect size analysis, it is possible to conclude that the mean salary difference between **Data Architect** and **Data Scientist** positions is, not only statistically significant, but also **practically significant**. Certainly, a **difference of \$41,875 MXN per month** is huge in the Mexican labor market. And, in this sense, it is noteworthy that the bootstrap confidence intervals for both positions are very different from each other, and the overlap shown between the bootstrap distributions for the Data Architect and Data Scientist mean salaries in the figure 14 is minimal.

Following the above logic, the results of the other t-tests and Mann-Whitney U tests are shown in the table 7, using the same signification level of  $\alpha = 0.05$ .

**Table 7.** T-test and Mann-Whitney U tests results of the average salary differences for other data job categories at a signification level of  $\alpha = 0.05$ .

Null hypothesis ( $H_0$ )	t-test		Mann-Whitney U test		Reject null hypothesis?
	t statistic	p-value	U statistic	p-value	
$\mu_{DE} \leq \mu_{DS}$	0.457	0.326	104.500	0.391	No
$\mu_{DS} \leq \mu_{DA}$	2.829	0.005	361.000	0.001	Yes
$\mu_{DS} \leq \mu_{BA}$	1.229	0.117	72.000	0.296	No
$\mu_{DR} \leq \mu_{DE}$	2.610	0.018	73.000	0.006	Yes
$\mu_{DE} \leq \mu_{BA}$	1.749	0.049	83.500	0.104	No
$\mu_{DE} \leq \mu_{DA}$	3.280	0.002	365.000	0.001	Yes
$\mu_{BA} \leq \mu_{DA}$	2.538	0.009	239.000	0.003	Yes

In view of the table 6, regarding the comparison of the salary averages for Data Engineers and Data Scientists, the obtained  $p$ -value from both the t-test and the Mann-Whitney U test was not significant ( $p > 0.05$ ). Therefore, the **Data Engineers** salaries are **not significantly higher** than those for **Data Scientists**, so the null hypothesis is not rejected.

Moreover, **Data Scientist's** salaries are significantly higher than those for **Data Analysts** but not for those of **Business Analysts**.

On the other hand, regarding the comparison of the salary averages for **Data Architect** and **Data Engineers**, the obtained  $p$ -value from the t-test was not significant ( $p > 0.05$ ), but the obtained  $p$ -value was significant from Mann-Whitney U test ( $p < 0.05$ ). In this sense, taking into account that the salary observations do not follow a normal distribution, the results from the Mann-Whitney U test could be deemed as more accurate. Thus, the Data Architect salaries are significantly higher than those for Data Engineers, so the null hypothesis is rejected.

On the contrary, regarding the comparison of the salary averages for **Data Engineers** and **Business Analysts**, the obtained  $p$ -value from the t-test was significant ( $p < 0.05$ ), whereas the obtained  $p$ -value was not significant from Mann-Whitney U test ( $p > 0.05$ ). Likewise, for the reasons stated before, the Mann-Whitney U test could be deemed as more accurate. Thus, the Data Engineer salaries are not significantly higher than those for Business Analysts, so the null hypothesis is not rejected.

In this context, **Data Engineer's** salaries are significantly higher than those for **Data Analysts**, and **Business Analyst's** salaries are significantly higher than those for **Data Analysts**.

On the other hand, the effect size analysis for the same pair-wise comparisons in table 6 is shown in the table 8.



**Table 8.** Absolute mean difference, Cohen's d and bootstrap confidence intervals for the mean for other data job categories at a signification level of  $\alpha = 0.05$ .

Data jobs	Absolute mean difference (MXN)	Cohen's d	95% Confidence interval for the first job (MXN)	95% Confidence interval for the second job (MXN)	Significant Practical Difference?
$\mu_{DE} - \mu_{DS}$	3,458	0.17 (Small)	(28,122, 58,216)	(27,357, 54,073)	No
$\mu_{DS} - \mu_{DA}$	16,562	0.99 (Large)	(26,641, 53,718)	(17,419, 31,315)	Yes
$\mu_{DS} - \mu_{BA}$	7,136	0.44 (Medium)	(26,000, 55,502)	(25,844, 39,602)	Yes
$\mu_{DR} - \mu_{DE}$	38,417	1.55 (Large)	(51,667, 116,667)	(29,712, 58,861)	Yes
$\mu_{DE} - \mu_{BA}$	10,594	0.63 (Large)	(29,096, 57,813)	(25,750, 40,200)	Yes
$\mu_{DE} - \mu_{DA}$	20,020	1.18 (Large)	(29,498, 58,191)	(17,125, 31,691)	Yes
$\mu_{BA} - \mu_{DA}$	9,426	0.66 (Large)	(25,875, 39,627)	(17,085, 31,774)	Yes

From the effect size analysis, it is possible to conclude that the mean salary difference between **Data Engineer** and **Data Scientist** positions is also **practically not significant**. As a **difference of \$3,458 MXN per month** is neglectable in the Mexican labor market for technology. And, in this sense, it is noteworthy that the bootstrap confidence intervals for both positions overlap in a large extent.

On the other hand, the mean salary differences between **Data Scientist** and **Data Analyst** positions are not only statistically significant, but also **practically significant**. As a **difference of \$16,562 MXN per month** is important in the Mexican labor market. And, in this sense, it is noteworthy that the bootstrap confidence intervals for both positions are different from each other.

However, the mean salary difference between **Data Scientist** and **Business Analyst** positions is more contentious, as it is not statistically significant but a **difference of \$7,136 MXN per month** is not neglectable in the Mexican labor market as this quantity represents about the national average salary in 2022 (Infobae, 2022). Therefore, notwithstanding with their lack of statistical significance, based on the effect size analysis, it is possible to state that there is a **significant practical difference** between the mean salary difference of the **Data Scientist** and the **Business Analyst** positions.

Similarly, even though the mean salary difference between **Data Engineers** and **Business Analyst** positions is not statistically significant, a **difference of \$10,594 MXN per month** is practically significant. Therefore, based on the effect size analysis, it is possible to conclude that there is a **significant practical difference** between the mean salary difference of the **Data Engineers** and the **Business Analyst** positions.

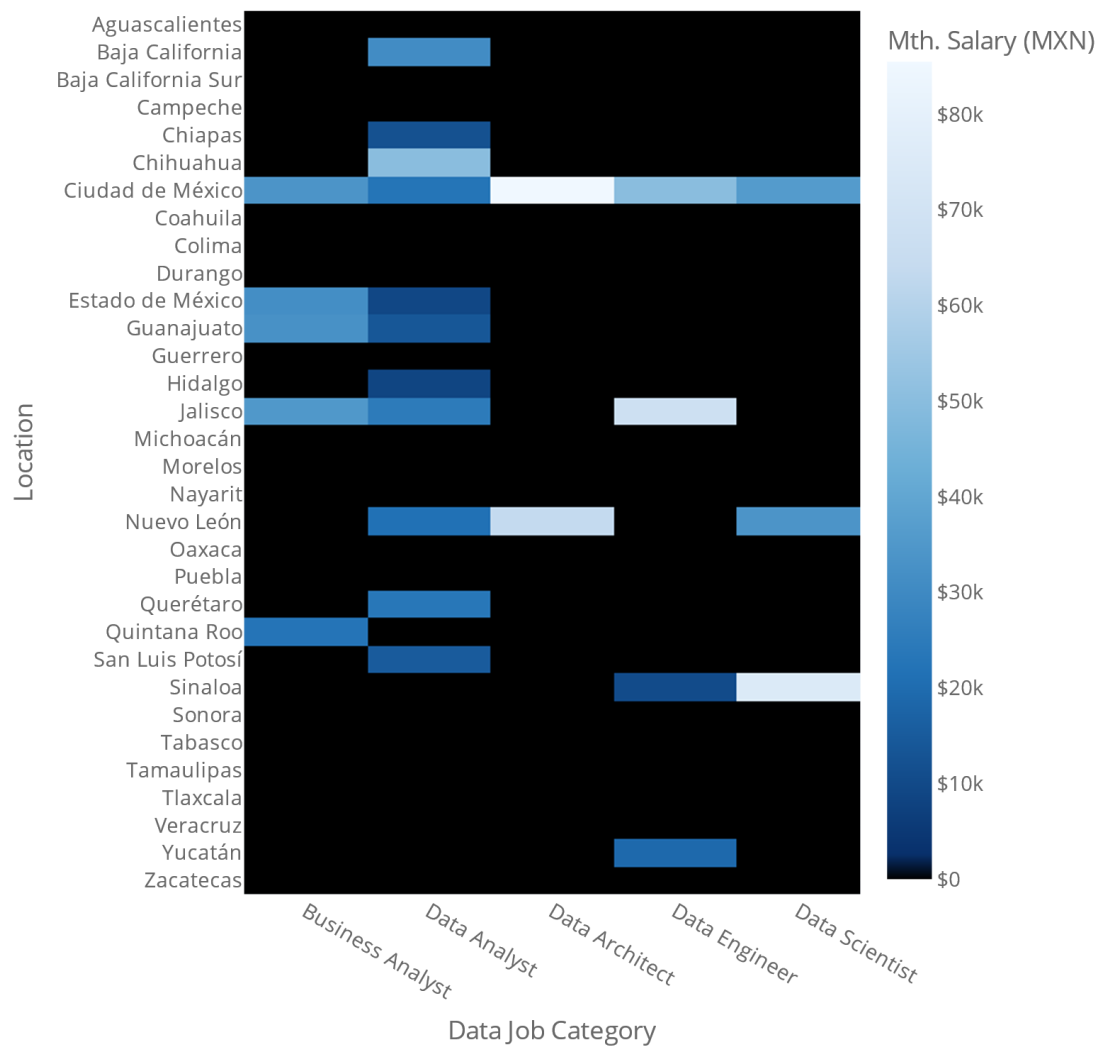
Moreover, it is possible to conclude that the mean salary difference between **Data Architect** and **Data Engineer** positions is, not only statistically significant, but also practically significant as a **difference of \$38,417 MXN per month** is important in the Mexican labor market. And, in this sense, it is noteworthy that the bootstrap confidence intervals for both positions are different from each other.

Again, from the effect size analysis, it is possible to conclude that the mean salary difference between **Data Engineer** and **Data Analyst** positions is, not only statistically significant, but also **practically significant**. As a **difference of \$20,020 MXN per month** is important.

In this context, according to the results from the present statistical analysis, average salaries for **Data Architects** are the **highest ones** in the current Mexican labor market.

### **Salaries per location**

Regarding the assessment of locations where the highest salaries for data jobs could be found, the following heatmap shows the relationship between locations, data job categories and average monthly salaries (figure 15).



**Figure 15.** Average salaries per location and data job category in February 2023 (own elaboration). In general, figure 14 suggests that the highest salaries can be found in Ciudad de México, Nuevo León, Jalisco and Sinaloa. However, the observation for the latter state is atypical and should be interpreted with caution.

More specifically, the highest salaries for the different data jobs categories can be found in the following states (table 9):

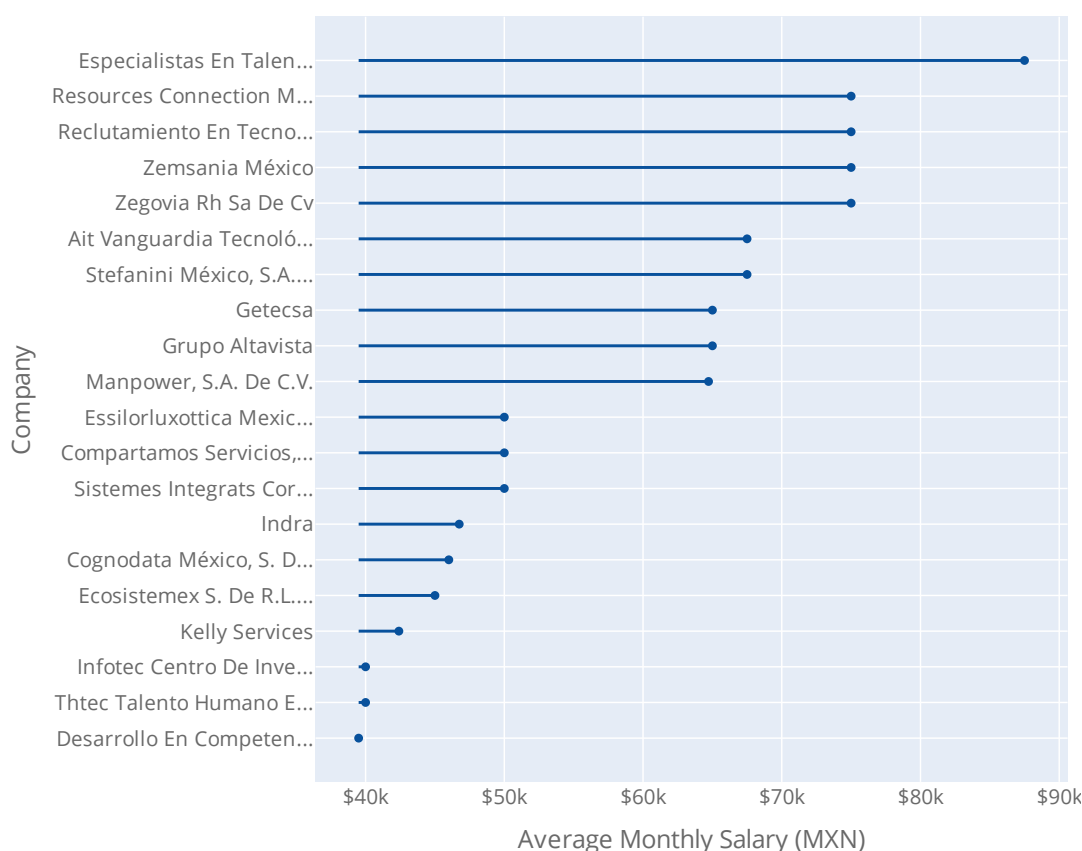
**Table 9.** Locations with the highest average monthly salaries by data job category.

Data Job Category	Locations with the highest salaries
Business Analyst	Ciudad de México
Data Analyst	Chihuahua
Data Architect	Ciudad de México
Data Engineer	Jalisco
Data Scientist	Sinaloa

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, it seems to be reasonable that the highest salaries could be found in the capital region.

### Companies offering the highest salaries

Regarding to the companies offering the highest salaries, the figure 16 shows the top 20 companies offering highest average salaries for all data jobs in Mexico.



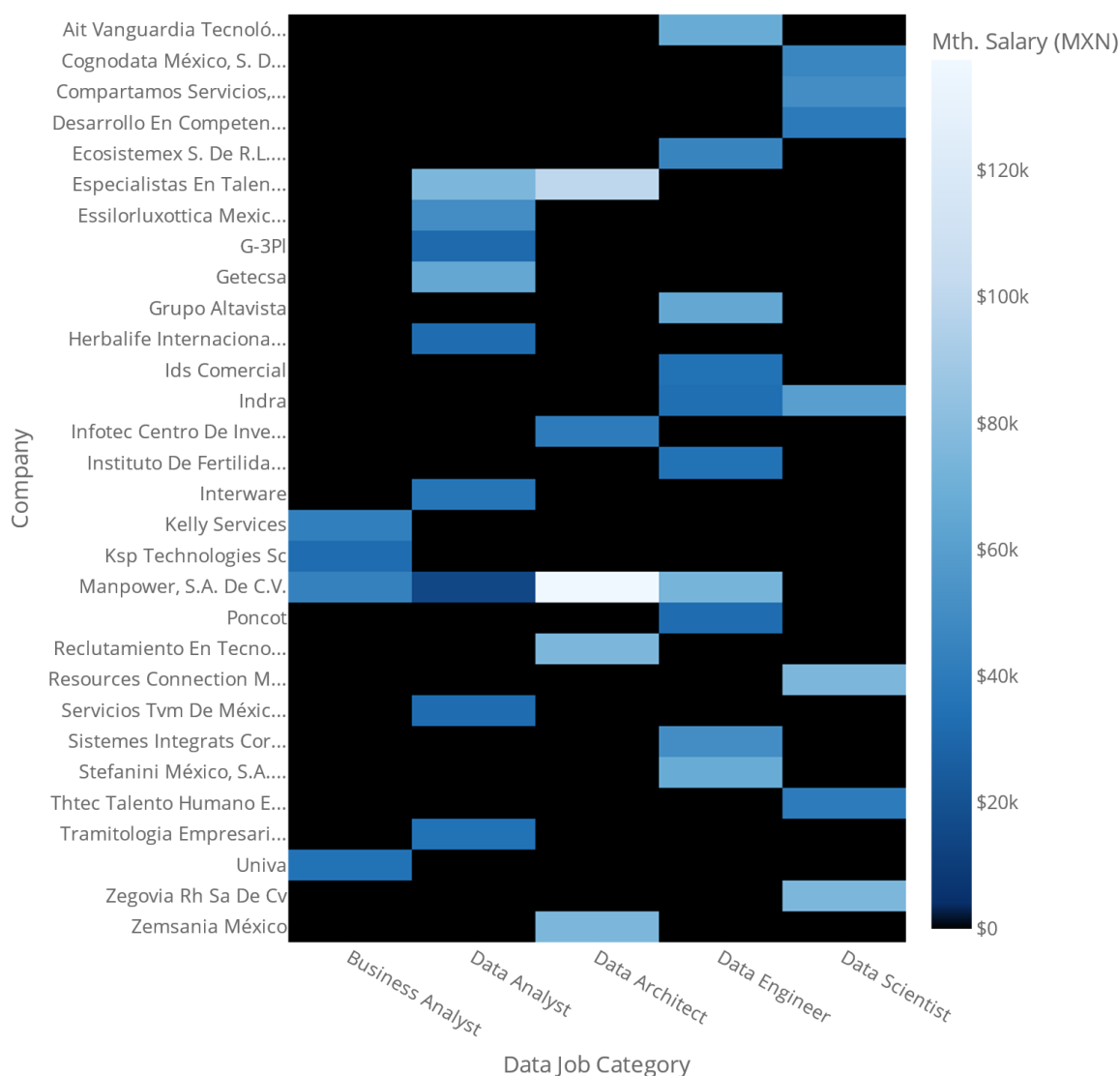
**Figure 16.** Top 20 companies with the highest average salaries for data jobs in Mexico in February 2023 (own elaboration).

From the plot above, the companies offering the highest salaries are **Especialistas en Talento**, **Zemsania México**, **Resources Connection México**, **Zegovia RH**, and **Reclutamiento en Tecnología**, which correspond to recruiting agencies and tech consulting firms.

However, in the case of *Especialistas en Talento*, *Zegovia RH* and *Reclutamiento en Tecnología* it is reasonable to think that their highly paid data job positions are demanded on behalf of other companies.

Again, the results from figure 16 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 current Mexican labor market.

Furthermore, the figure 17 shows a closer look of the top 30 companies offering the highest salaries by data job category.



**Figure 17.** Companies with the highest average salaries by data job category in Mexico in February 2023 (own elaboration).

In view of figure 17, for **Business Analyst** positions, the organizations offering higher salaries are *Kelly Services* and *Manpower*. Moreover, for **Data Analyst** positions, the

organizations offering higher salaries are *Especialistas en Talento* and *Getecsa*. Furthermore, for **Data Architect** positions, the organizations offering higher salaries are *Manpower* and *Especialistas en Talento*. In addition, for **Data Engineer** positions, the organizations offering higher salaries are *Manpower*, *Ait Vanguardia Tecnológica*, and *Stefanini México*. And, finally, for **Data Scientist** positions, the organizations offering higher salaries are *Resources Connection* and *Zegovia Rh*.

### 3.4 Dashboard

A dashboard was built and deployed to explore the data in a more interactive way. Please visit this [link](#) to play with the tool. A screenshot of the dashboard is shown below (figure 18).

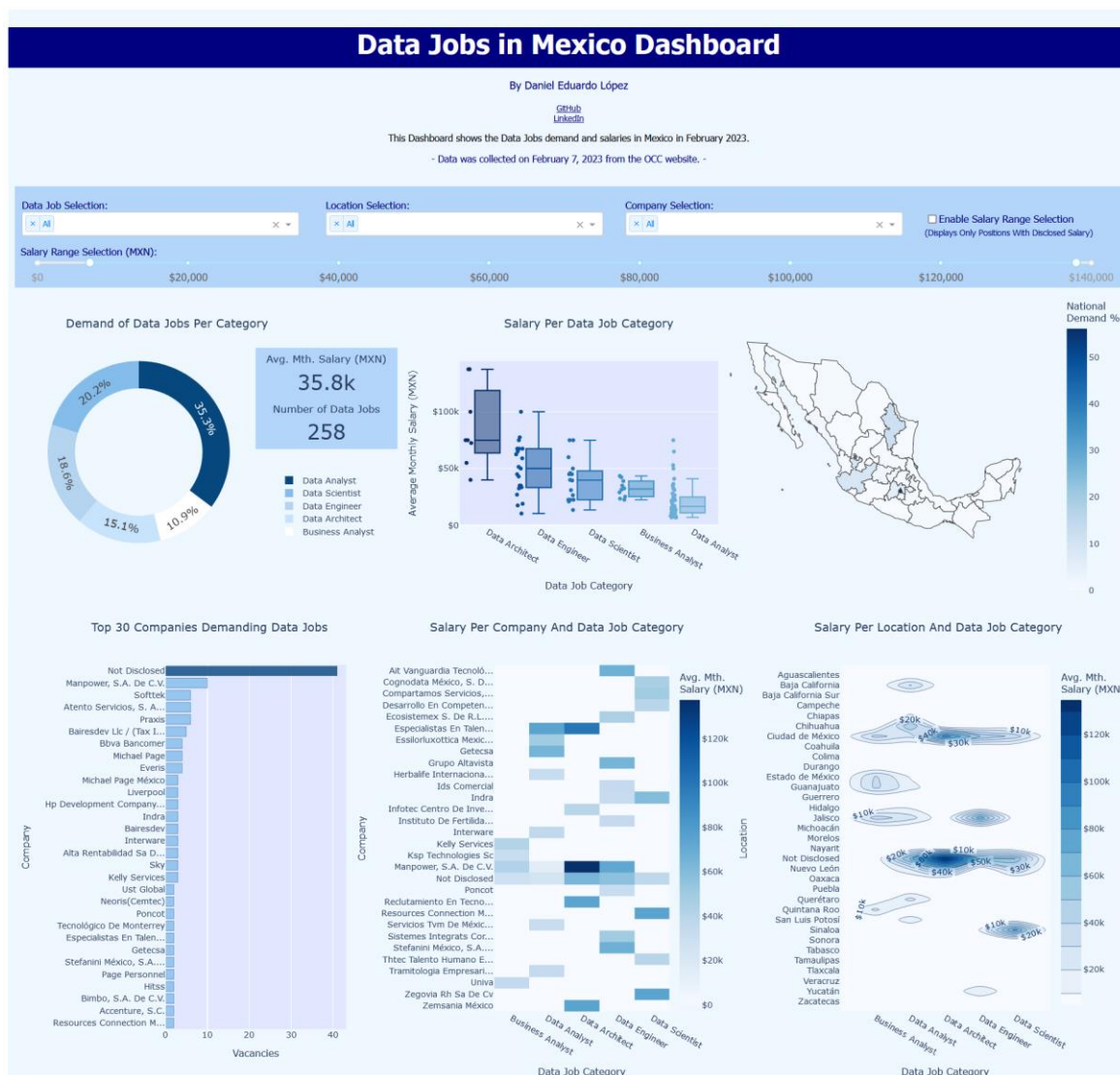


Figure 18. Dashboard of the data jobs in Mexico in February 2023 (own elaboration).

## 4. Conclusions

The data job category with **the highest salary** in the Mexican labor market in February 2023 according to the OCC website was **Data Architect**. Indeed, the average salary for Data Architect positions was significantly higher than for both **Data Engineers** and **Data Scientists**. Thus, the present study's hypothesis is rejected.

On the other hand, the data job category **most demanded** in the Mexican labor market was **Data Analyst**; even though it was also the one with **the lowest salary**. Also, this data job category is the most demanded across the different Mexican states, whereas **Data Architect** and **Data Scientist** positions were the most concentrated in certain locations, namely, Ciudad de México, Nuevo León, and Jalisco.

Moreover, **Ciudad de México** was the location where it is possible to find **the highest jobs demand and the highest salaries**. However, **Nuevo León, Estado de México, and Jalisco** were locations where the demand of data jobs and the salaries offered are the highest after the capital.

Furthermore, the companies with the greatest demand of data positions were **Manpower, Atento Servicios, Praxis** and **Softek**; nevertheless, the organizations that offered the highest salaries were **Especialistas en Talento, Zemsania México, Resources Connection México, Zegovia RH**, and **Reclutamiento en Tecnología**, which correspond to recruiting agencies and tech consulting firms.

Additionally, it was also found that the data jobs demand from some companies spread across several Mexican states such as *Accenture, Praxis* or *Softek*, and that there are some well-known organizations whose data jobs demand is not located in the capital region, such as *HP* (Jalisco).

The results of the present study suggest that Data Analyst, Business Analyst, and Data Engineer positions were more demanded across different organizations. On the contrary, Data Scientist and, certainly, Data Architect vacancies could only be found in more specific organizations like tech companies or tech consultancy firms.

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. Thus, very few observations were obtained for the least demanded data jobs categories: Data Architect and Business Analyst. Also, the collected data mostly corresponded to Ciudad de México, Nuevo León, Jalisco, and Estado de México, 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, make a distinction among entry level, middle and senior positions, and collect more salary data for Data Architect and Business Analyst positions as well as for other Mexican states.

## 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.
- Infobae. (2022, Enero 28).** Cuánto gana en promedio un mexicano y por qué 30 mil pesos no son un “sueldito”. *Infobae*. Retrieved from <https://www.infobae.com/america/mexico/2022/01/29/cuanto-gana-en-promedio-un-mexicano-y-por-que-30-mil-pesos-no-son-un-sueldito/>
- 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>