DATA JOBS SALARIES IN MEXICO IN NOVEMBER 2023

Daniel Eduardo López

14 January 2024

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 Intelligence Analysts, Business Analysts, Data Analysts, Data Architects, Data Engineers, Data Scientists, and Machine Learning Engineers. 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 November 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 ML Engineers and Data Architects enjoyed the highest earnings with an average monthly salary of about 65,000 MXN and 54,222 MXN, respectively, whereas the Data Analyst positions had the lowest ones with an average monthly salary of about 24,026 MXN (figure 1).

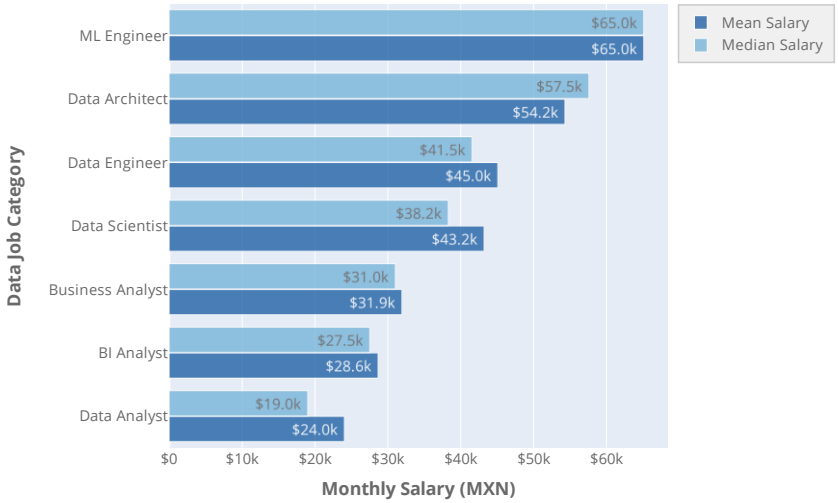

Figure 1. Mean and median monthly salary for each data job category (own elaboration).

Indeed, the average salary for ML Engineers was not significantly lower than that for Data Architects, being the mean salary difference between those two data job categories of 10,375 MXN. However, it is important to bear in mind that the observations of salaries for ML Engineers and 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; however, the highest salaries could be found in remote positions. And, finally, it was identified that the companies with the greatest demand of data positions were *Bairesdev, Banamex, Pepsico* and *Softek*; and the ones offering the highest salaries were *Ecosistemex, Caspex Corp, Addon Technologies, Enterprise Solutions,* and *Softek*.

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.

The data jobs market can be classified into the next sometimes-overlapping job profiles: Business Intelligence Analysts, Business Analysts, Data Analysts, Data Architects, Data Engineers, Data Scientists, and Machine Learning Engineers (Axis Talent, 2020; ai-jobs.net, 2023).

Business Intelligence Analysts (BI) are professional working business and market data to help organizations to make informed decisions (ai-jobs.net, 2023). On the other hand, 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); whereas Machine Learning Engineers (ML) are devoted to the designing, building, testing and deploying of machine learning models to automate processes or improve decision-making (ai-jobs.net, 2023).

Even though the different data jobs overlap at certain extent, they 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.

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 November 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 identify in which Mexican state the highest salaries are offered.
- To identify what companies offer the highest salaries.
- To assess the statistical significance of the salary differences for each data job category.
- To assess the practical differences between the salary differences for each data job category.

2.2 Research Question

Which data job category has the highest salary in the Mexican labor market in November 2023 according to the OCC website?

2.3 Hypothesis

From previous studies (López, 2023), it was hypothesized that Data Architects is the data job category with the highest salary in the Mexican labor market in November 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 26 November 2023, through web scraping with Python 3 and its libraries Selenium and Beautiful Soup. The OCC website was selected due to its importance as a platform for both recruiters and job seekers in Mexico and its more lenient policy for automated process of information retrieval.

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, Tukey-Kramer, one-sample and two-sample t-tests) and non-parametric (Kruskal-Wallis H, Dunn, Wilcoxon signed-rank 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, post hoc tests were performed to identify where the differences lie among the different categories. To do so, both a Tukey-Kramer and Dunn's tests were carried out.

Later, pairwise tests were used to further test the statistical significance for the mean salary difference among data job categories.

When only one salary observation was retrieved for a given category, one-sample t-test and Wilcoxon signed-rank tests were carried out, using the category with just one salary observation as the population mean in the case of the one-sample t-test, and subtracting from the array of salaries of the other category the value of the salary for the category with just one observation.

On the other hand, when an array of salaries could be collected for both data job categories under comparison, two-sample t-tests assuming unequal variances and Mann-Whitney U tests were performed.

Finally, an effect size analysis was carried out by computing the absolute mean salary difference**,** the percentage difference of the two mean samples**,** 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 were significant from a practical point of view.

3. Results and Discussion

3.1 Data Collected

Data collection and cleaning allowed to identify **563** data jobs in Mexico on 26 November 2023 from the OCC Website, of which **197** had salary information (35%). The whole data set was used in the data jobs demand analysis below, whereas 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 (**33%**) 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).

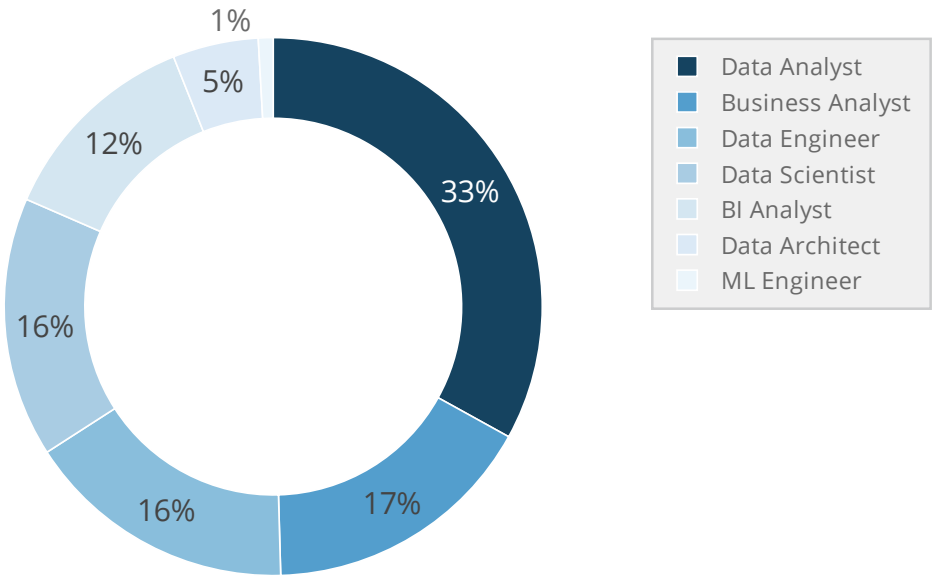

Figure 2. Proportion of vacancies for each data job category in Mexico in November 2023 (own elaboration).

In addition, from figure 2, it is also possible to observe that Business Analyst (17%) and Data Engineers (16%) positions account for the other third of the data jobs demand in Mexico.

On the contrary, ML Engineer positions are the less demanded, with only a 1% 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 November 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 **57% of the total national demand**. On the other hand, Jalisco and Nuevo Léon represented a distant second place with about the 9% of the demand. Finally, Estado de México accounted for about the 4% of the demand, respectively.

Therefore, the combined demand in Ciudad de México, Nuevo León, Jalisco, and Estado de México accounted for about 80% 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.

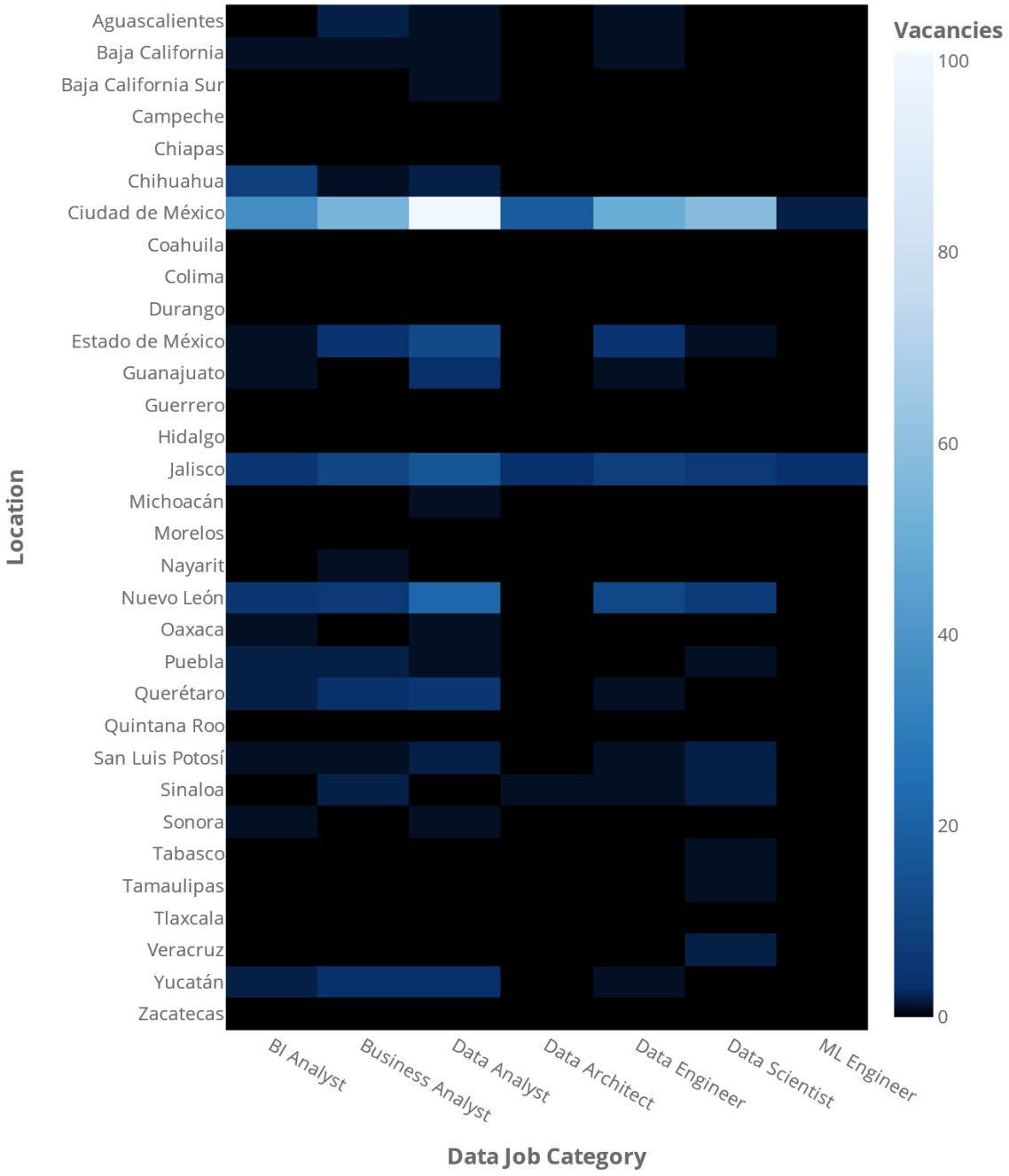

Figure 4. Data jobs demand per category and location in November 2023 (own elaboration).

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 **Business Analyst** and **BI Analyst** positions, whereas **Data Architect** and **ML Engineer** positions are highly concentrated in Ciudad de México and Jalisco.

As shown above, the data jobs demand is largely concentrated in **Ciudad de México**, whereas **Jalisco and Nuevo León**, and **Estado de México** are distant second and third 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).

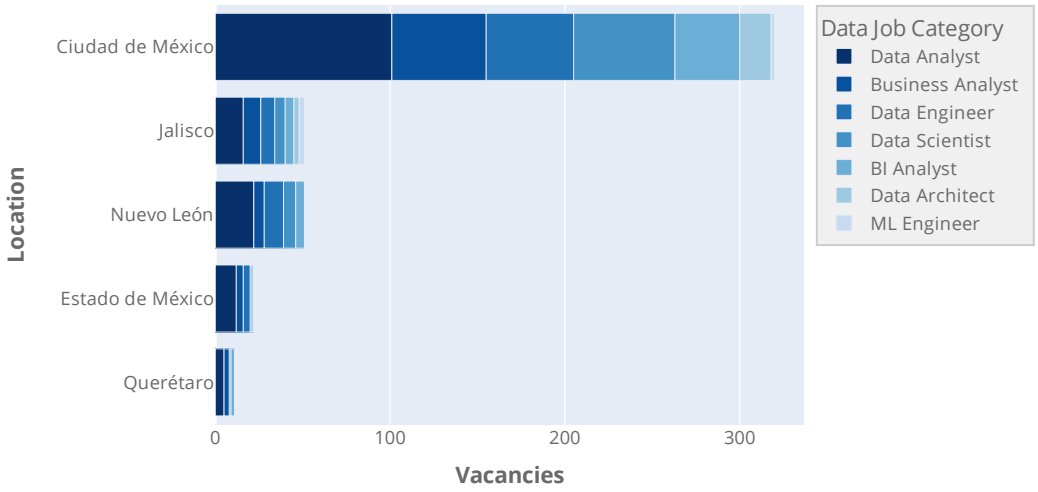

Figure 5. Data jobs demand per category in Ciudad de México, Jalisco, Nuevo León, Estado de México, and Querétaro in November 2023 (own elaboration).

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)
BI Analyst	37	70	52.9%
Business Analyst	54	93	58.1%
Data Analyst	101	186	54.3%
Data Architect	18	29	62.1%
Data Engineer	50	92	54.3%
Data Scientist	58	88	65.9%
ML Engineer	2	5	40.0%
All	320	563	56.8%

Thus, in view of the table 1, some positions such as **Data Scientist and Data Architect,** are **highly concentrated in Ciudad de México**, with 65.9% and 62.1% of the national demand for said categories, respectively. On the other hand, **ML Engineer** positions are the ones more demanded outside Ciudad de México with just a 40.0% of the national demand.

Companies demanding more data jobs

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

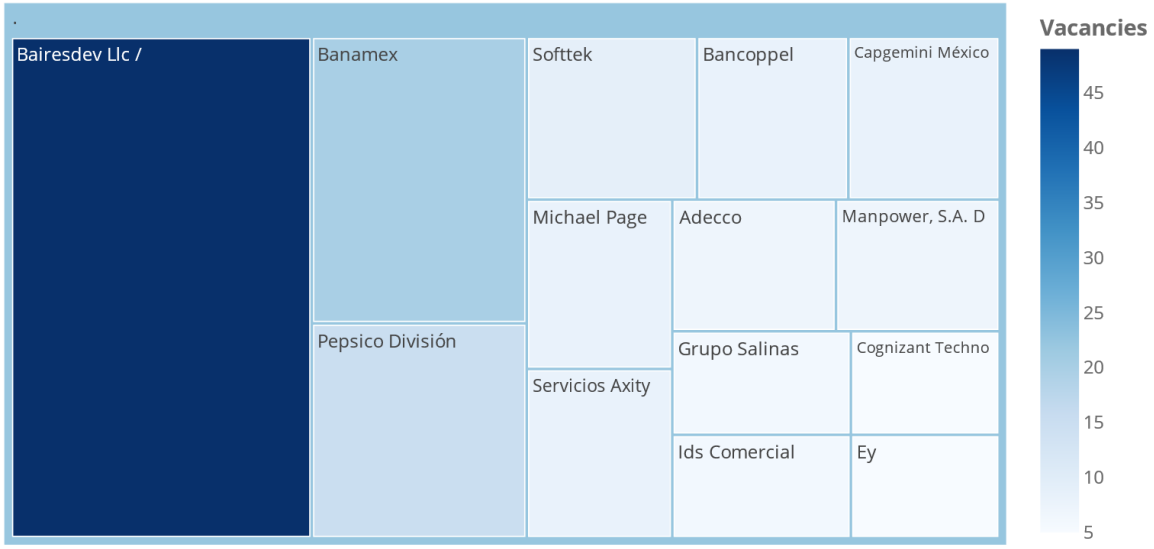

Figure 6. Top 15 companies with the highest demand of data jobs in Mexico in November 2023 (own elaboration).

So, ***Bairesdev*** is nowadays the biggest seeker of data skills in the Mexican labor market, along with ***Banamex***, ***Pepsico*** and ***Softek***. In the case of *Michael Page, Adecco, and Manpower* and, it is reasonable to think that their data job vacancies are published on behalf of other companies.

On the other hand, *Bairesdev* and *Softek* are tech consultancy firms. Other well-known organizations in the top 15 of companies demanding data jobs in Mexico are *Banamex*, *Pepsico, Bancoppel, Grupo Salinas and EY*. 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.

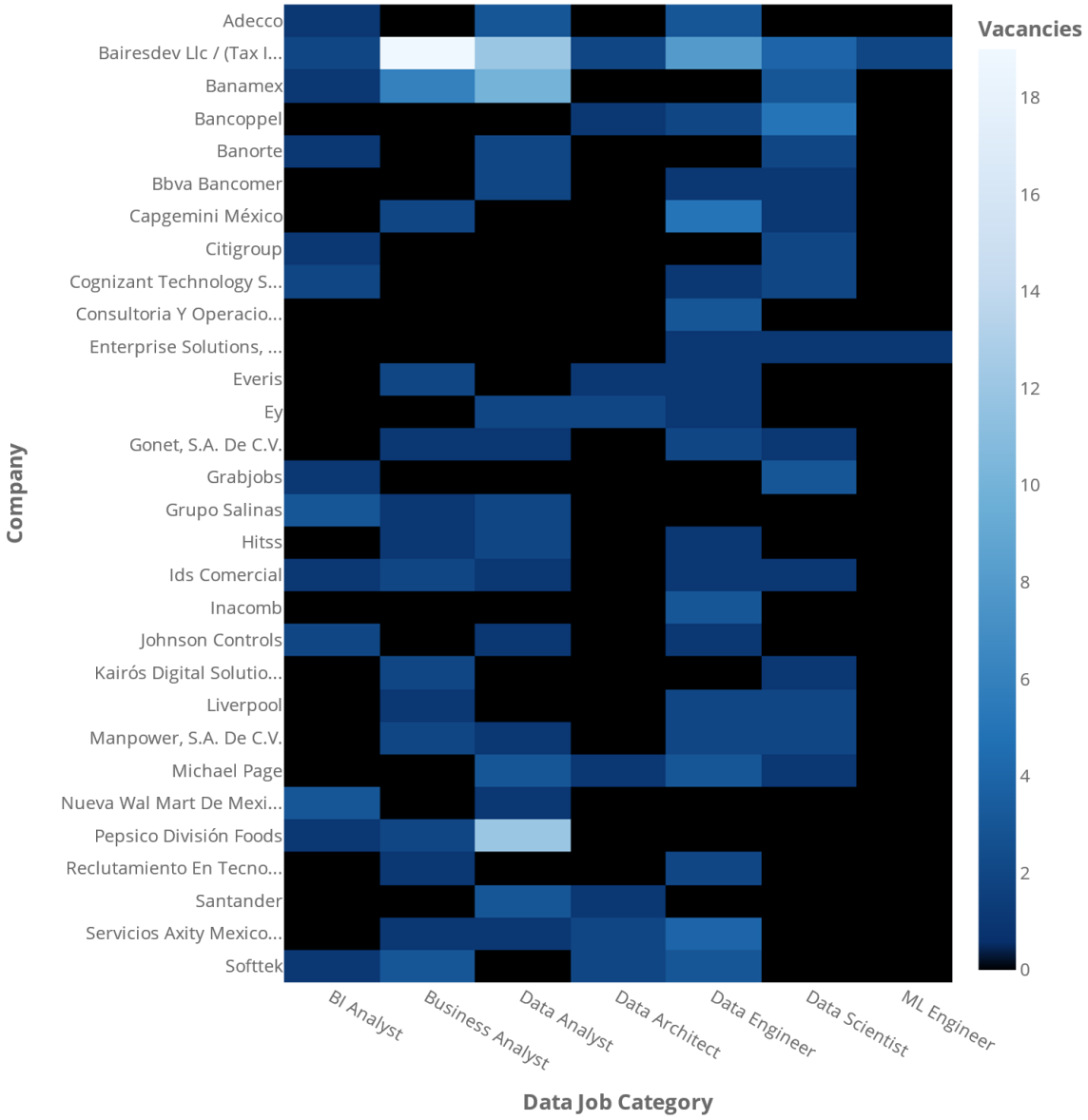

Figure 7. Demand per data job category in the top 30 companies with the highest data jobs demand in November 2023 (own elaboration).

The above heatmap suggests that **Data Engineer** positions are more demanded across different organizations. On the contrary, **ML Engineer** and **Data Architect** vacancies are more likely to be found in more specific organizations like tech consultancy firms and banks (*Bairesdev*, *Everys*, Bancoppel, *EY*, *Santander*, *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.

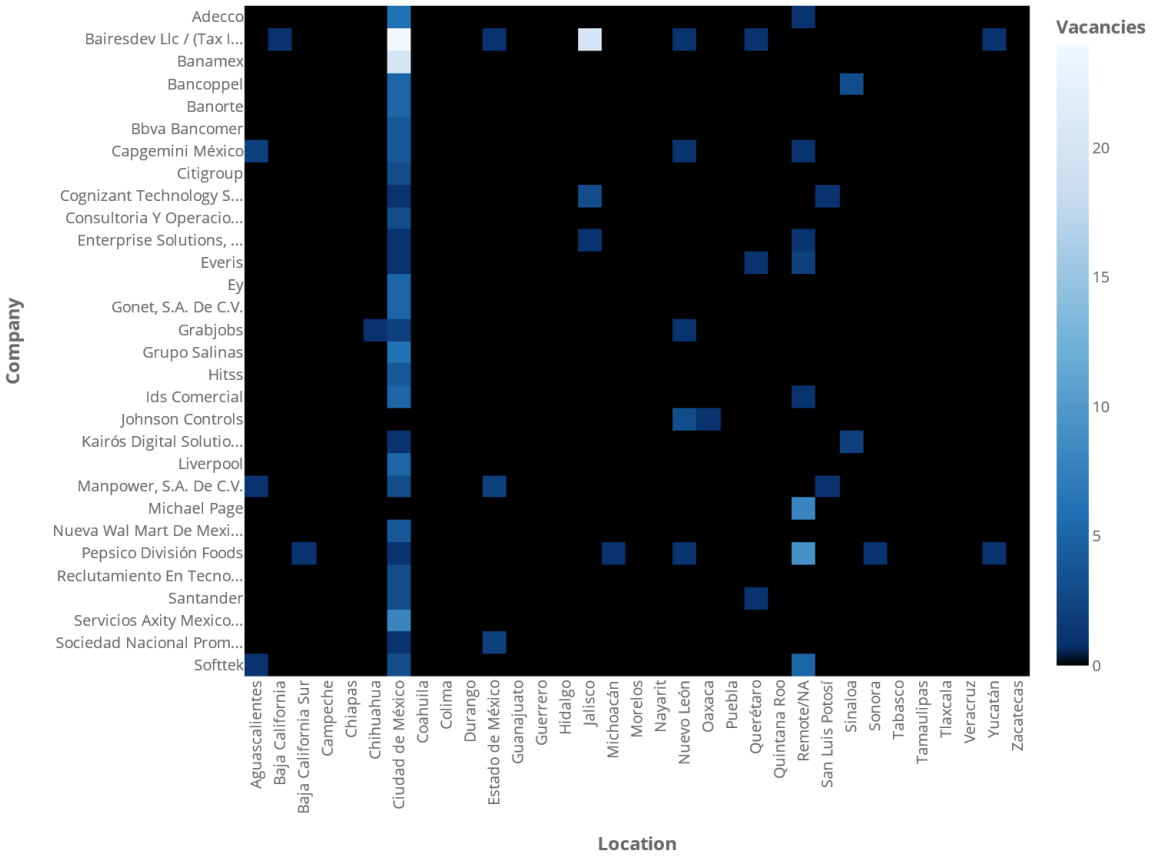

Figure 8. Top 30 companies demanding data jobs per location in November 2023 (own elaboration).

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 such as *Bairesdev* or *Pepsico*.

Furthermore, there are few well-known companies whose data jobs demand is not located in the capital region, such as such as *Jonhson Controls*, which is located in Nuevo León.

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

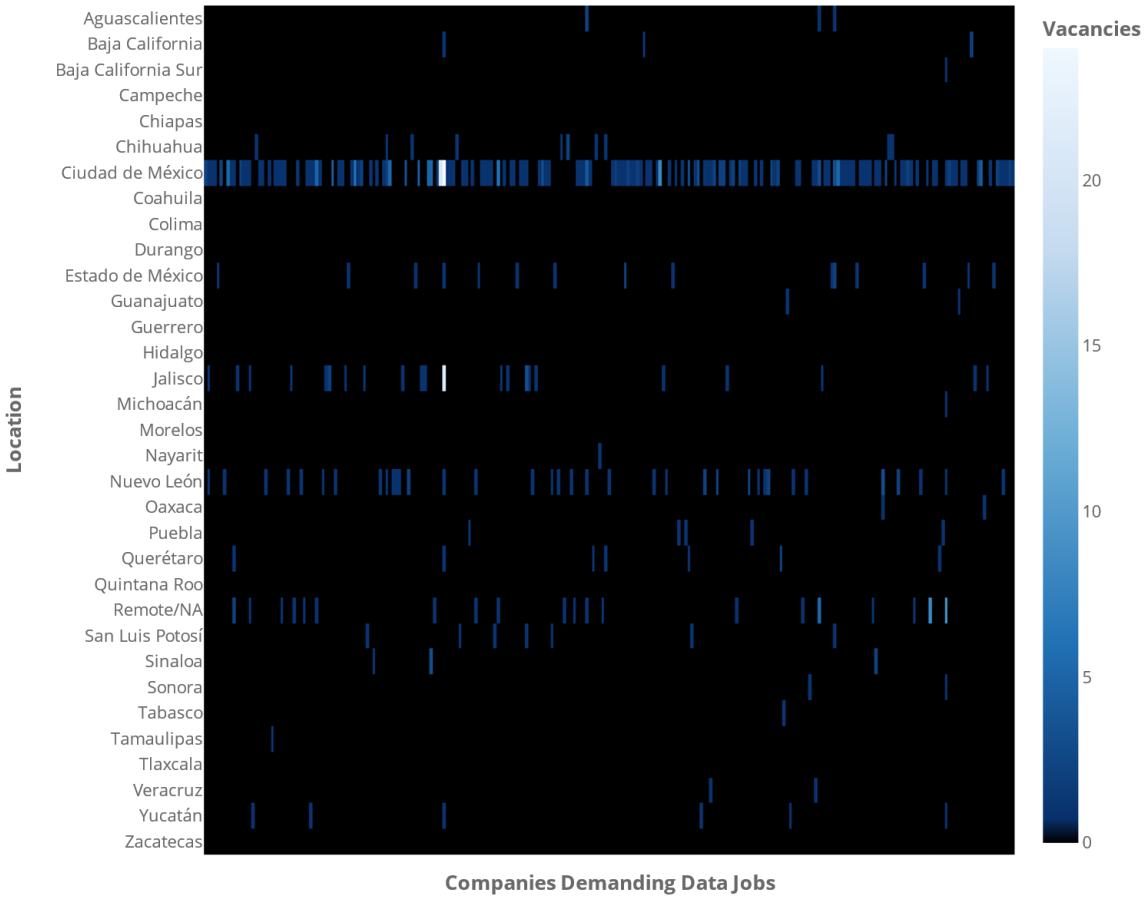

Figure 9. Companies demanding data jobs per location in November 2023 (own elaboration).

Likewise, the above heatmap beautifully shows that most of the data jobs are concentrated in **Ciudad de México** and, into a much 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.

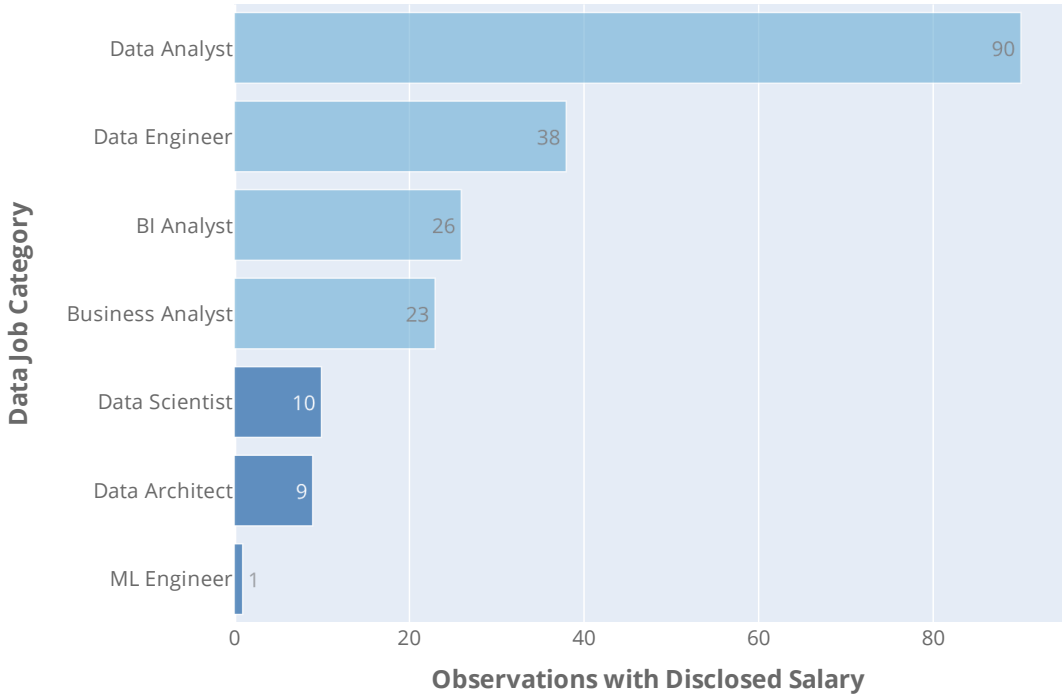

Figure 10. Number of salary observations per data job category (own elaboration).

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

Overall, the average salary of the data jobs in Mexico in November 2023 was **32,162.97 MXN (*SD* = 19,417.17)** per month. The distribution of the salary is shown in the figure 11.

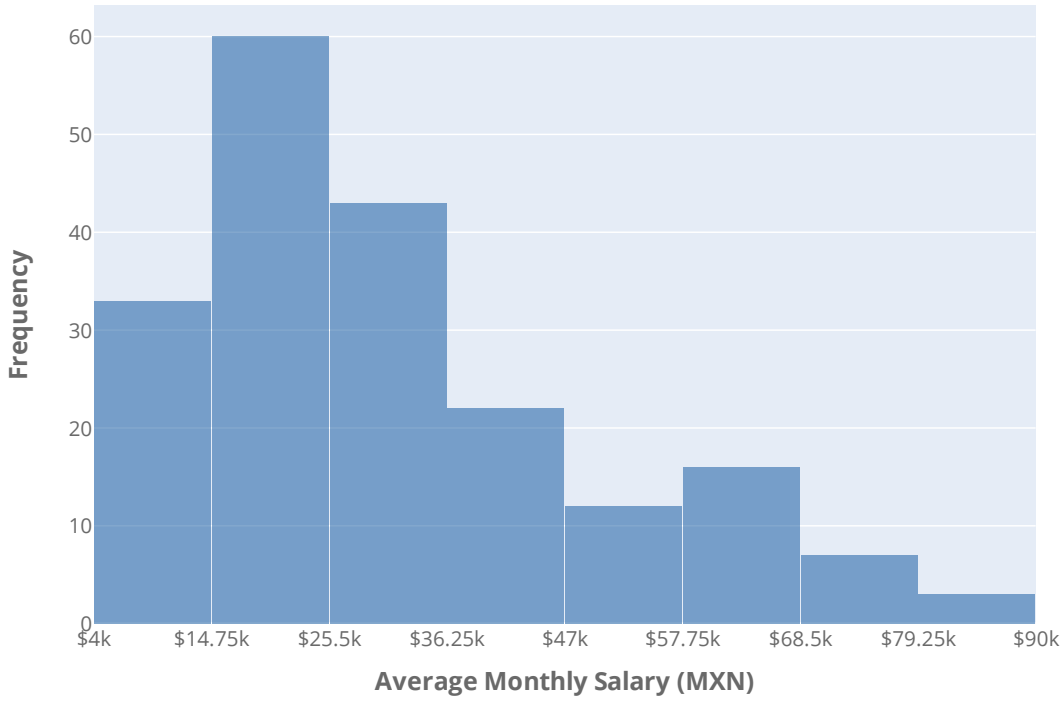

Figure 11. Histogram of salary distribution of the data jobs in Mexico in November 2023 (own elaboration).

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 14,750 to 36,250 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:

$$H_0: \quad Normal\ Distribution$$

$$H_1: \quad Not\ H_0$$

$$\alpha = 0.05$$

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 $\boldsymbol{\alpha = 0.05}$.

K^2 **statistic**	***p*-value**
24.808	<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 **BI Analyst** and **Business Analyst** 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).

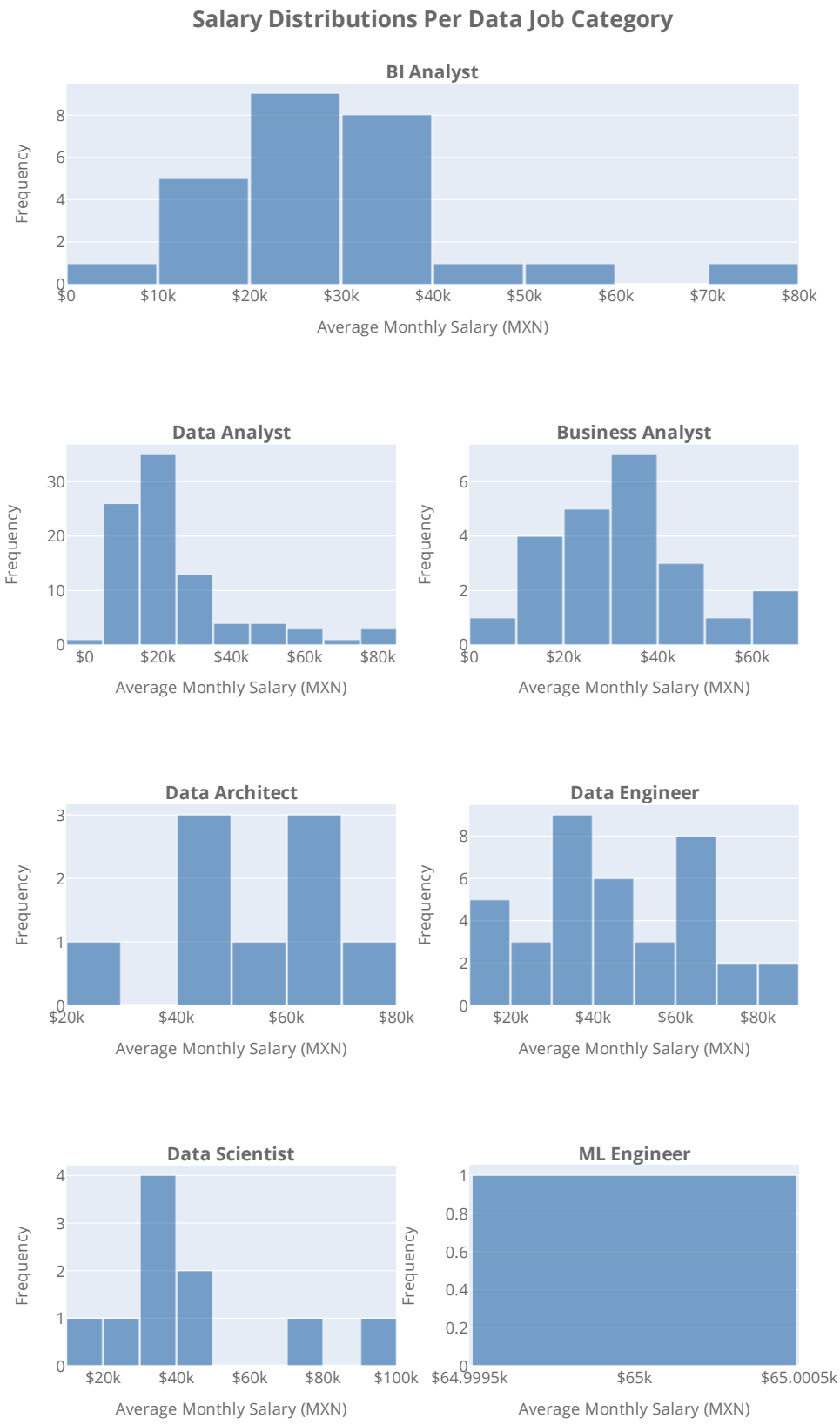

Figure 12. Histogram of salary distributions per data job category in Mexico in November 2023 (own elaboration).

Notwithstanding with the above, for the purposes of the present study, both parametric (ANOVA, Tukey-Kramer, one-sample and two-sample t-tests) and non-parametric (Kruskal-Wallis H, Dunn, Wilcoxon signed-rank and Mann-Whitney U) 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).

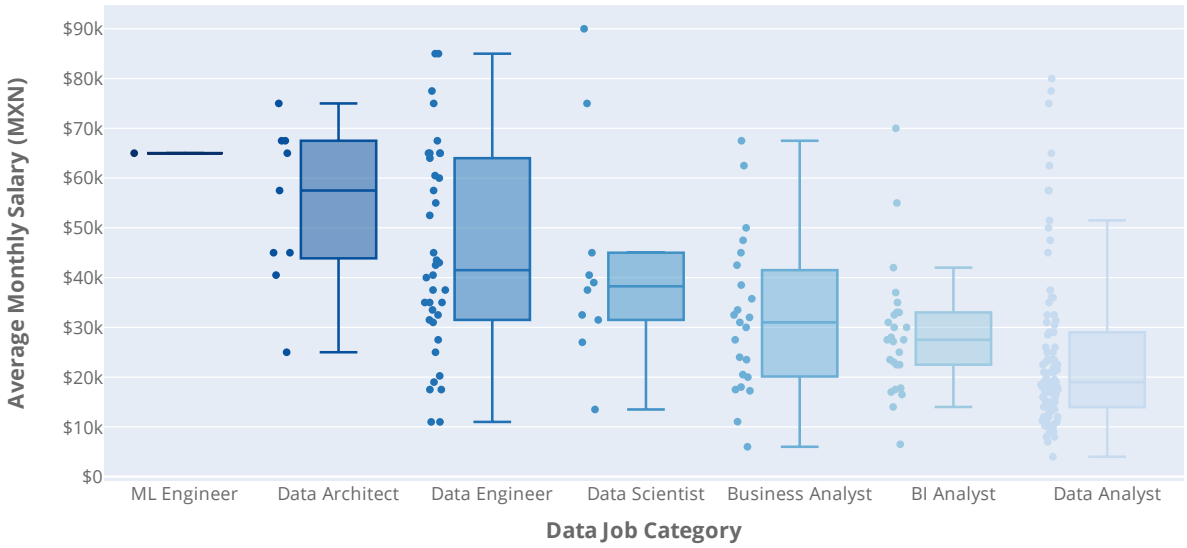

Figure 13. Box plot of salary distributions for each data job category in Mexico in November 2023 (own elaboration).

Figure 13 suggests that, after removing outliers from the calculation of the average salaries, **ML Engineer and Data Architect** positions are still the ones with the highest salaries in the current Mexican labor market with a monthly median salary of about **65,000 MXN and 57,500 MXN, respectively.**

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 November 2023.

Data Job Category	Mean Monthly Salary (MXN)	Median Monthly Salary (MXN)
ML Engineer	65,000	65,000
Data Architect	54,222	57,500
Data Engineer	45,033	41,500
Data Scientist	43,150	38,250
Business Analyst	31,893	31,000
BI Analyst	28,633	27,500
Data Analyst	24,026	19,000

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, ML Engineer 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:

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

$$H_1: \quad Not\ H_0$$

$$\alpha = 0.05$$

Whereμ_{BI} is the average salary for BI Analysts; μ_{BA}, average salary for Business Analysts; μ_{DA}, average salary for Data Analysts; μ_{DR}, average salary for Data Architects; μ_{DE}, average salary for Data Engineers; μ_{DS}, average salary for Data Scientists; and, μ_{ML}, average salary for ML Engineers.

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 $\boldsymbol{\alpha = 0.05}$.

Test	Statistic	*p*-value
One-way ANOVA	$F = 8.657$	<0.001
Kruskal-Wallis	$H = 42.662$	<0.001

Thus, from table 3, it is observed that the average salary differences for the 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, post hoc tests were applied to identify where the significant average salary differences for the data jobs lie. To do so, a Tukey-Kramer and a Dunn's test were performed. The results for the parametric test are shown in the table 5; whereas the results for the non-parametric test are shown in the table 6.

Table 5. Tukey-Kramer post hoc test results for the average salary differences for the data jobs categories at a signification level of $\boldsymbol{\alpha = 0.05}$.

	BI Analyst	**Business Analyst**	**Data Analyst**	**Data Architect**	**Data Engineer**	**Data Scientist**	**ML Engineer**
BI Analyst	False	False	False	True	True	False	False
Business Analyst	False	False	False	False	False	False	False
Data Analyst	False	False	False	True	True	True	False
Data Architect	True	False	True	False	False	False	False
Data Engineer	True	False	True	False	False	False	False
Data Scientist	False	False	True	False	False	False	False
ML Engineer	False	False	False	False	False	False	False

Table 6. Dunn's post hoc test results for the average salary differences for the data jobs categories at a signification level of $\boldsymbol{\alpha = 0.05}$.

	BI Analyst	**Business Analyst**	**Data Analyst**	**Data Architect**	**Data Engineer**	**Data Scientist**	**ML Engineer**
BI Analyst	False	False	False	True	True	False	False
Business Analyst	False	False	False	False	False	False	False
Data Analyst	False	False	False	True	True	True	False
Data Architect	True	False	True	False	False	False	False
Data Engineer	True	False	True	False	False	False	False
Data Scientist	False	False	True	False	False	False	False
ML Engineer	False	False	False	False	False	False	False

Thus, according to the results above from tables 5 and 6:

- The salary mean for **BI Analyst** positions is significantly different from those for Data Architect, and Data Engineer positions.
- The salary mean for **Business Analyst** positions is significantly different from those for Data Analyst, Data Architect, Data Engineer, and Data Scientist positions.
- The salary mean for **Data Analyst** positions is significantly different from those for Business Analyst, Data Architect, Data Engineer, and Data Scientist positions.
- The salary mean for **Data Architect** positions is significantly different from those for BI Analyst, Business Analyst, and Data Analyst positions.
- The salary mean for **Data Engineer** positions is significantly different from those for BI Analyst, Business Analyst, and Data Analyst positions.
- The salary mean for **Data Scientist** positions is significantly different from those for Data Analyst positions.
- Finally, as only one salary observation was retrieved for **ML Engineer** positions, no appropriate comparisons could be carried out by the post hoc tests.

Later, specific data jobs categories were later compared using the one-sample t-test, the two-sample t-test with unequal variance (Welch's test), the Wilcoxon signed-rank test and the Mann-Whitney U test to further confirm the significance of the above results.

The one-sample t-test and the Wilcoxon signed-rank test were used when only one salary observation was collected for a data job category, i.e., ML Engineer positions; whereas the two-sample t-test with unequal variance (Welch's test) and the Mann-Whitney U test were performed whenever the sample size of salaries for both data job categories under comparison was larger than 1, i.e., the rest of data job categories.

In first place, the average salaries for ML Engineers (ML) and Data Architects (DA) were compared under the following hypothesis:

$$H_0: \quad \mu_{DR} \geq \mu_{ML}$$

$$H_1: \quad \mu_{DR} < \mu_{ML}$$

$$\alpha = 0.05$$

The results of the one-sample t-test and the Wilcoxon signed-rank test are shown in the table 7.

Table 7. One-sample t-test and Wilcoxon signed-rank test results of the average salary differences for ML Engineers and Data Architects at a signification level of $\boldsymbol{\alpha = 0.05}$.

Test	Statistic	*p*-value
One-sample t-test	t = -1.360	0.116
Wilcoxon signed-rank test	T = 6.000	0.219

Therefore, from table 7 it is observed that the average salary differences for ML Engineers and Data Architects were not statistically significant, as the *p*-value of both tests indicated that the probability of obtaining the observed outcome is not unlikely under the null hypothesis at the selected level of significance. Therefore, the null hypothesis is failed to be rejected. In other words, **Data Architect's salaries are not significantly lower than those for ML Engineer**.

Again, caution must be taken as only one salary observation for ML Engineer was collected.

Then, the difference between the mean salaries for each data job category were calculated in order to quantify the size of the effect, by computing the absolute difference and the percentage difference. The results are shown in the table 8.

Table 8. 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 $\boldsymbol{\alpha = 0.05}$.

Effect size	Value
Absolute mean difference	10,375 MXN
Percentage difference	17.35%

So, from table 8, it is possible to conclude that the mean salary difference between ML Engineer and Data Architect positions is neither statistically nor practically significant. As a difference of 10,375 MXN per month, or a percentage difference of 17.35%, is not that important when the salary observation for a ML Engineer is 65,000 MXN and the mean salary for a Data Architect position is about 54,625 MXN.

Indeed, according to the *Reporte del Mercado Laboral de TI México 2023* (Spanish for *Mexico IT Labor Market Report 2023*) by Olvera (2023), only about 20% of the surveyed IT professionals in Mexico would be willing to a job change for a salary increase of 20% or less; whereas the 79% would be willing to a job change for a salary increase of 20% or more. So, as a heuristic rule, a percentage difference of **20% was selected as a cut-off value for practical significance**.

Later, the average salaries for Data Architects (DR) and Data Engineers (DE) were compared under the following hypothesis:

$$H_0: \quad \mu_{DR} \leq \mu_{DE}$$

$$H_1: \quad \mu_{DR} > \mu_{DE}$$

$$\alpha = 0.05$$

The results of the two-sample t-test with unequal variance and the Mann-Whitney U test are shown in the table 9.

Table 9. Two-sample t-test with unequal variance and Mann-Whitney U test results of the average salary differences for Data Architects and Data Scientists at a signification level of $\boldsymbol{\alpha = 0.05}$.

Test	Statistic	*p*-value
Two-sample t-test with unequal variance	$t = 1.101$	0.153
Mann-Whitney U test	$U = 119.000$	0.113

Therefore, from table 9 it is observed that the average salary differences for the Data Architects and Data Engineer were not statistically significant, as the *p*-value of both tests indicates that the probability of obtaining the observed outcome is not unlikely under the null hypothesis at the selected level of significance. Therefore, the null hypothesis is failed to be rejected. In other words, **Data Architect's salaries are not significantly higher than those for Data Engineers**.

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. Nonetheless, for sake of completeness, Cohen's d was also estimated. The results are shown in the table 10.

Table 10. Absolute mean difference, percentage difference, Cohen's d and bootstrap confidence intervals for the mean for Data Architect and Data Engineer salaries at a signification level of $\boldsymbol{\alpha = 0.05}$.

Effect size	Value
Absolute mean difference	9,238 MXN
Percentage difference	18.47%
Cohen's d	$d = 0.48$ (Medium effect size)
95% Confidence interval for the mean for DR salaries	(33,750 MXN, 72,083 MXN)
95% Confidence interval for the mean for DE salaries	(36,472 MXN, 54,384 MXN)

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

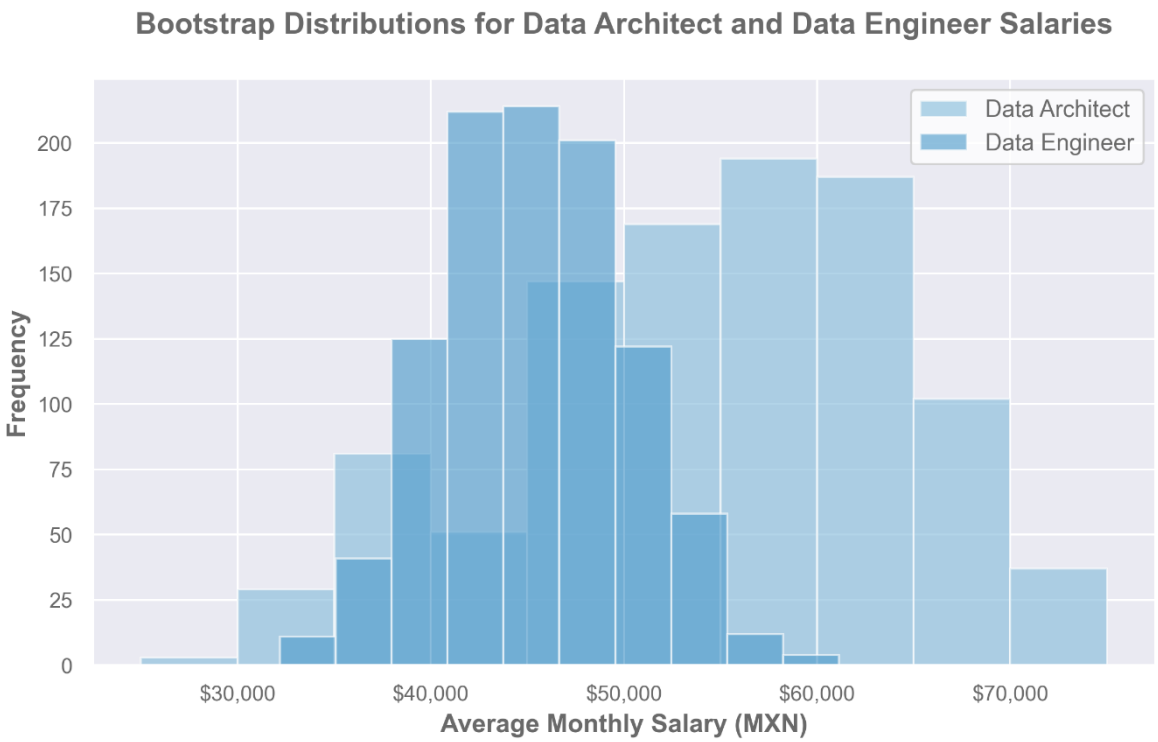

Figure 14. Bootstrap distributions for the Data Architect and Data Engineer mean salaries (own elaboration).

So, from the effect size analysis, it is possible to conclude that the mean salary difference between **Data Architect** and **Data Engineer** positions is neither statistically nor **practically significant**. Certainly, a **difference of 9,238 MXN per month** is not that huge in the Mexican labor market when the mean salaries are 54,222 MXN and 45,033 MXN, for Data Architects and Data Engineer positions, respectively. And, in this sense, it is noteworthy that the bootstrap confidence intervals for Data Engineer positions are encompassed by the bootstrap confidence intervals for Data Architects.

Following the above logic, the results from all the hypothesis testing are shown in the table 11, using the same signification level of $\alpha = 0.05$.

Table 11. Parametric and non-parametric testing results of the average salary differences for all data job categories at a signification level of $\boldsymbol{\alpha = 0.05}$.

	Parametric		Non-parametric		
Null hypothesis (H_0)	**Statistic**	***p*-value**	**Statistic**	***p*-value**	**Reject null hypothesis?**
$\mu_{DR} \geq \mu_{ML}$	-1.360	0.116	6.000	0.219	No
$\mu_{DE} \geq \mu_{ML}$	-5.620	<0.001	18.500	<0.001	Yes
$\mu_{DS} \geq \mu_{ML}$	-2.311	0.027	3.000	0.020	Yes
$\mu_{BA} \geq \mu_{ML}$	-10.259	<0.001	1.500	<0.001	Yes
$\mu_{BI} \geq \mu_{ML}$	-11.271	<0.001	1.000	<0.001	Yes
$\mu_{DA} \geq \mu_{ML}$	-21.190	<0.001	4.000	<0.001	Yes
$\mu_{DR} \leq \mu_{DE}$	1.101	0.153	119.000	0.113	No
$\mu_{DR} \leq \mu_{DS}$	0.867	0.201	31.500	0.183	No
$\mu_{DR} \leq \mu_{BA}$	2.744	0.015	115.500	0.007	Yes
$\mu_{DR} \leq \mu_{BI}$	2.995	0.010	101.000	0.007	Yes
$\mu_{DR} \leq \mu_{DA}$	3.854	0.005	345.000	0.001	Yes
$\mu_{DE} \leq \mu_{DS}$	0.096	0.463	127.500	0.401	No
$\mu_{DE} \leq \mu_{BA}$	2.839	0.003	487.500	0.005	Yes
$\mu_{DE} \leq \mu_{BI}$	3.303	0.001	453.500	0.001	Yes
$\mu_{DE} \leq \mu_{DA}$	5.293	<0.001	1593.500	<0.001	Yes
$\mu_{DS} \leq \mu_{BA}$	1.331	0.108	122.500	0.088	No
$\mu_{DS} \leq \mu_{BI}$	1.543	0.079	119.500	0.024	Yes
$\mu_{DS} \leq \mu_{DA}$	2.219	0.029	411.500	0.003	Yes
$\mu_{BA} \leq \mu_{BI}$	0.434	0.333	252.000	0.300	No
$\mu_{BA} \leq \mu_{DA}$	2.020	0.025	999.000	0.006	Yes
$\mu_{BI} \leq \mu_{DA}$	1.543	0.066	869.000	0.008	Yes

In view of the table 11, the salaries for all the data job categories were, indeed, significantly lower than that for ML Engineers, except for Data Architect positions ($p > 0.05$). Moreover, the salaries for Data Architect were not significantly higher than those for Data Engineer and Data Scientist positions ($p > 0.05$).

In a similar fashion, the salaries for Data Engineers were not significantly higher than those for Data Scientists ($p > 0.05$); the salaries for Data Scientists were not significantly

higher than those for Business Analysts ($p > 0.05$), and the salaries for Business Analysts were not significantly higher than those for BI Analysts ($p > 0.05$).

On the other hand, regarding the comparison of the salary averages for Data Scientists and BI Analysts, the obtained *p*-value from the two-sample t-test with unequal variance was not significant ($p > 0.05$), but the obtained *p*-value was significant from the 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 were deemed as more accurate. Thus, the Data Scientists salaries are significantly higher than those for BI Analysts, and the null hypothesis was rejected.

All in all, ML Engineer salary was significantly higher than the average salaries for all the data job categories (except Data Architect positions); Data Architect and Data Engineer average salaries were significantly higher than those for Business Analysts, BI Analysts, and Data Analysts; Data Scientists average salary was significantly higher than those for BI Analysts and Data Analysts; and, Business Analysts and BI Analysts average salaries were significantly higher than those for Data Analysts.

Moreover, the effect size analysis for the same pair-wise comparisons in table 11 is shown in the table 12.

Table 12. Absolute mean difference, mean percentage difference, Cohen's d and bootstrap confidence intervals for the mean for all data job categories at a signification level of $\boldsymbol{\alpha = 0.05}$.

Data jobs	Absolute mean difference (MXN)	Mean percentage difference (%)	Cohen's d	95% Confidence interval for the first job (MXN)	95% Confidence interval for the second job (MXN)	Significant practical difference?
$\mu_{DR} - \mu_{ML}$	10,375	17.35	NA	NA	NA	No
$\mu_{DE} - \mu_{ML}$	19,612	35.53	NA	NA	NA	Yes
$\mu_{DS} - \mu_{ML}$	20,531	37.51	NA	NA	NA	Yes
$\mu_{BA} - \mu_{ML}$	33,107	68.34	NA	NA	NA	Yes
$\mu_{BI} - \mu_{ML}$	35,053	73.84	NA	NA	NA	Yes
$\mu_{DA} - \mu_{ML}$	40,693	91.13	NA	NA	NA	Yes
$\mu_{DR} - \mu_{DE}$	9,238	18.47	0.48	(33,750, 72,083)	(36,472, 54,384)	No
$\mu_{DR} - \mu_{DS}$	10,156	20.50	0.45	(36,833, 72,083)	(23,875, 71,875)	No
$\mu_{DR} - \mu_{BA}$	22,732	52.55	1.41	(30,917, 70,000)	(22,776, 41,412)	Yes
$\mu_{DR} - \mu_{BI}$	24,678	58.36	1.64	(33,750, 72,500)	(22,749, 39,650)	Yes
$\mu_{DR} - \mu_{DA}$	30,318	76.82	1.94	(33,750, 72,083)	(19,596, 29,860)	Yes
$\mu_{DE} - \mu_{DS}$	919	2.04	0.04	(35,539, 54,317)	(24,562, 71,897)	No
$\mu_{DE} - \mu_{BA}$	13,494	34.92	0.77	(36,665, 55,675)	(23,784, 41,184)	Yes
$\mu_{DE} - \mu_{BI}$	15,441	40.99	0.90	(36,350, 55,168)	(22,488, 39,125)	Yes
$\mu_{DE} - \mu_{DA}$	21,081	60.50	1.27	(35,898, 54,400)	(19,518, 30,084)	Yes
$\mu_{DS} - \mu_{BA}$	12,575	32.94	0.69	(24,812, 70,894)	(22,657, 41,298)	Yes
$\mu_{DS} - \mu_{BI}$	14,522	39.03	0.82	(25,117, 69,961)	(22,643, 38,651)	Yes
$\mu_{DS} - \mu_{DA}$	20,162	58.63	1.21	(26,053, 69,375)	(19,430, 29,902)	Yes
$\mu_{BA} - \mu_{BI}$	1,947	6.30	0.13	(23,594, 41,506)	(22,560, 39,100)	No
$\mu_{BA} - \mu_{DA}$	7,587	27.00	0.49	(22,987, 40,867)	(19,582, 29,529)	Yes
$\mu_{BI} - \mu_{DA}$	5,640	20.79	0.38	(22,463, 38,381)	(19,475, 29,708)	No

From the effect size analysis (table 12), it is possible to conclude that the mean salary difference between ML Engineer and Data Architect positions is neither statistically nor practically significant. Certainly, a difference of 10,375 MXN per month, or a percentage difference of 17.35%, is not that important when the salary observation for a ML Engineer is 65,000 MXN and the mean salary for Data Architect positions is about 54,625 MXN.

On the other hand, the difference among the ML Engineer salary observation and the rest of the data job categories (except Data Architect positions) was practically significant, with a mean percentage difference ranging from 91% to 35%.

Moreover, it was noteworthy that the mean salary difference among Data Architects, Data Engineers, and Data Scientist is not practically significant (<10,156 MXN); whereas the mean salary difference among Data Architects and the rest of the data jobs was practically significant (>22,732 MXN).

The mean salary difference among Data Engineer and Data Scientist positions was also neither statistically nor practically significant (919 MXN). However, the salary difference among Data Engineer and Data Scientist positions in regard to Business Analyst, BI Analyst and Data Analyst positions was practically significant (>12,575 MXN).

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. That is the reason why an effect size analysis should be carried out along with the hypothesis testing.

The mean salary difference between Business Analyst and BI Analyst was not practically significant (1,947 MXN); however, the difference between Business Analyst and Data Analyst was practically significant (7,587 MXN). Even though the mean absolute difference seems like a low figure (7,587 MXN), it represents a mean percentage difference of 27%, which is non-neglectable in the Mexican labor market.

Finally, even though the mean salary difference between BI Analyst and Data Analyst positions was statistically significant, it was deemed as not practically significant. Indeed, their mean monthly salaries were 28,633 MXN and 24,026 MXN, respectively, and a difference of 5,640 MXN or a 20.79% percentage difference is not very important in the Mexican labor market for technology. So, again, here it lies the importance of performing an effect size analysis along with the hypothesis testing.

In this context, according to the results from the present statistical analysis, average salaries for **ML Engineers and Data Architects** were 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).

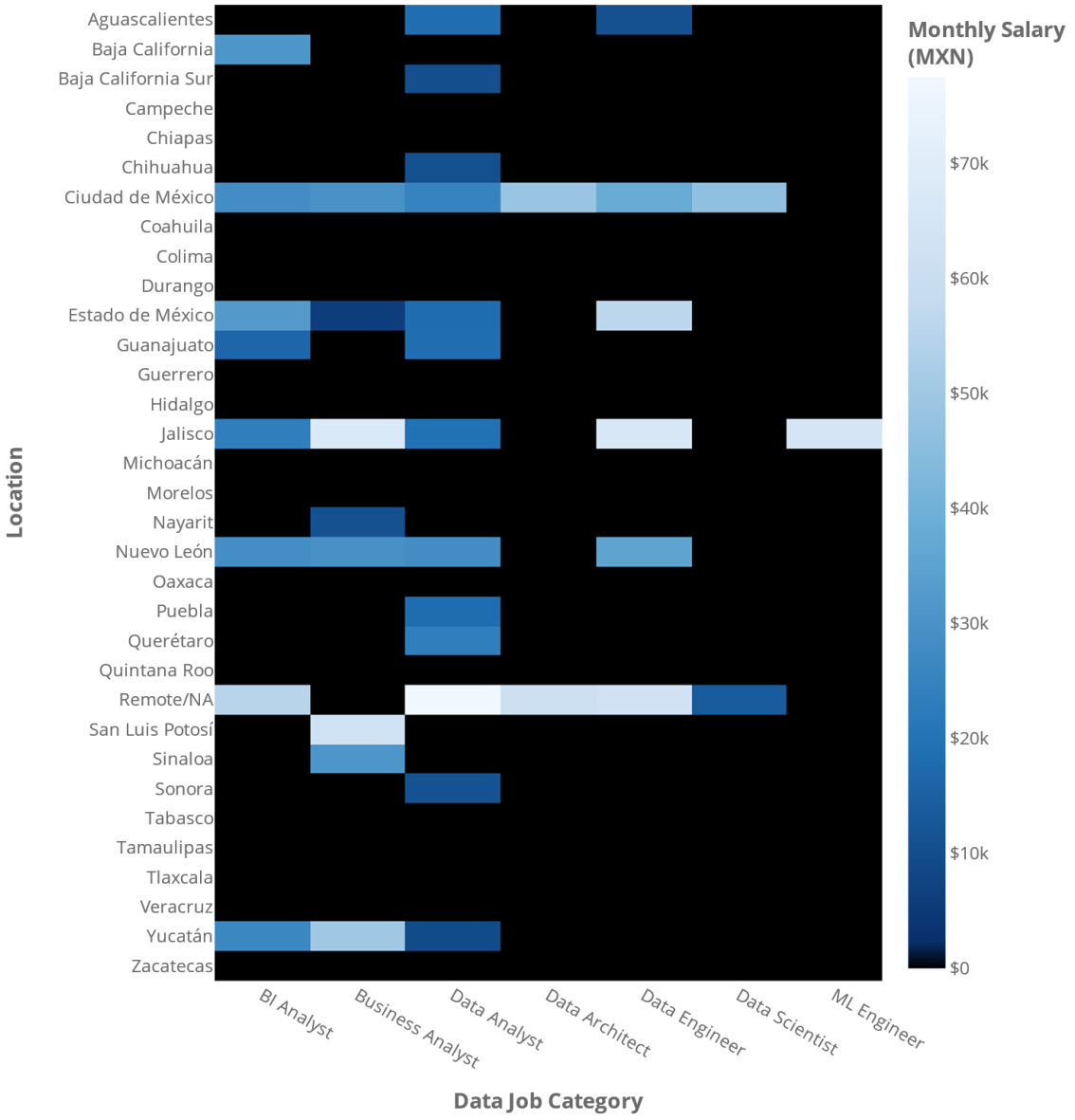

Figure 15. Average salaries per location and data job category in November 2023 (own elaboration).

In general, figure 15 suggests that the highest salaries can be found in Remote positions, Ciudad de México. This supports the growing importance of the remote jobs in the current Mexican labor market.

More specifically, the highest salaries for the different data jobs categories can be found in the following states (table 13):

Table 13. Locations with the highest average monthly salaries by data job category.

Data Job Category	Locations with the highest salaries
BI Analyst	Remoto
Business Analyst	Jalisco
Data Analyst	Remoto
Data Architect	Remoto
Data Engineer	Jalisco
Data Scientist	Ciudad de México
ML Engineer	Jalisco

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 is clear a tendency in which the highest salaries could be found in the remote positions, Ciudad de México, and Jalisco.

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.

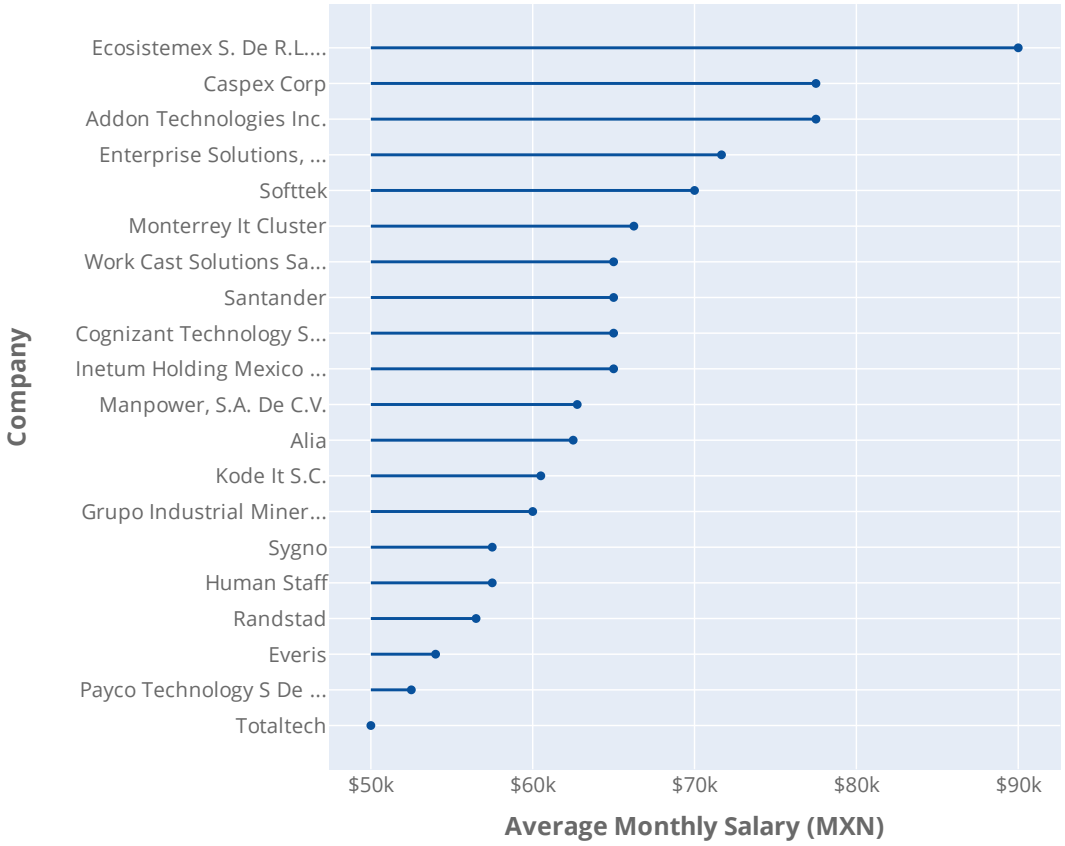

Figure 16. Top 20 companies with the highest average salaries for data jobs in Mexico in November 2023 (own elaboration).

From the plot above, the companies offering the highest salaries are *Ecosistemex*, *Caspex*, *Addon Technologies*, *Enterprises Solutions*, and *Softtek*. From those, 2 are recruiting agencies and 3 are tech consulting firms.

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.

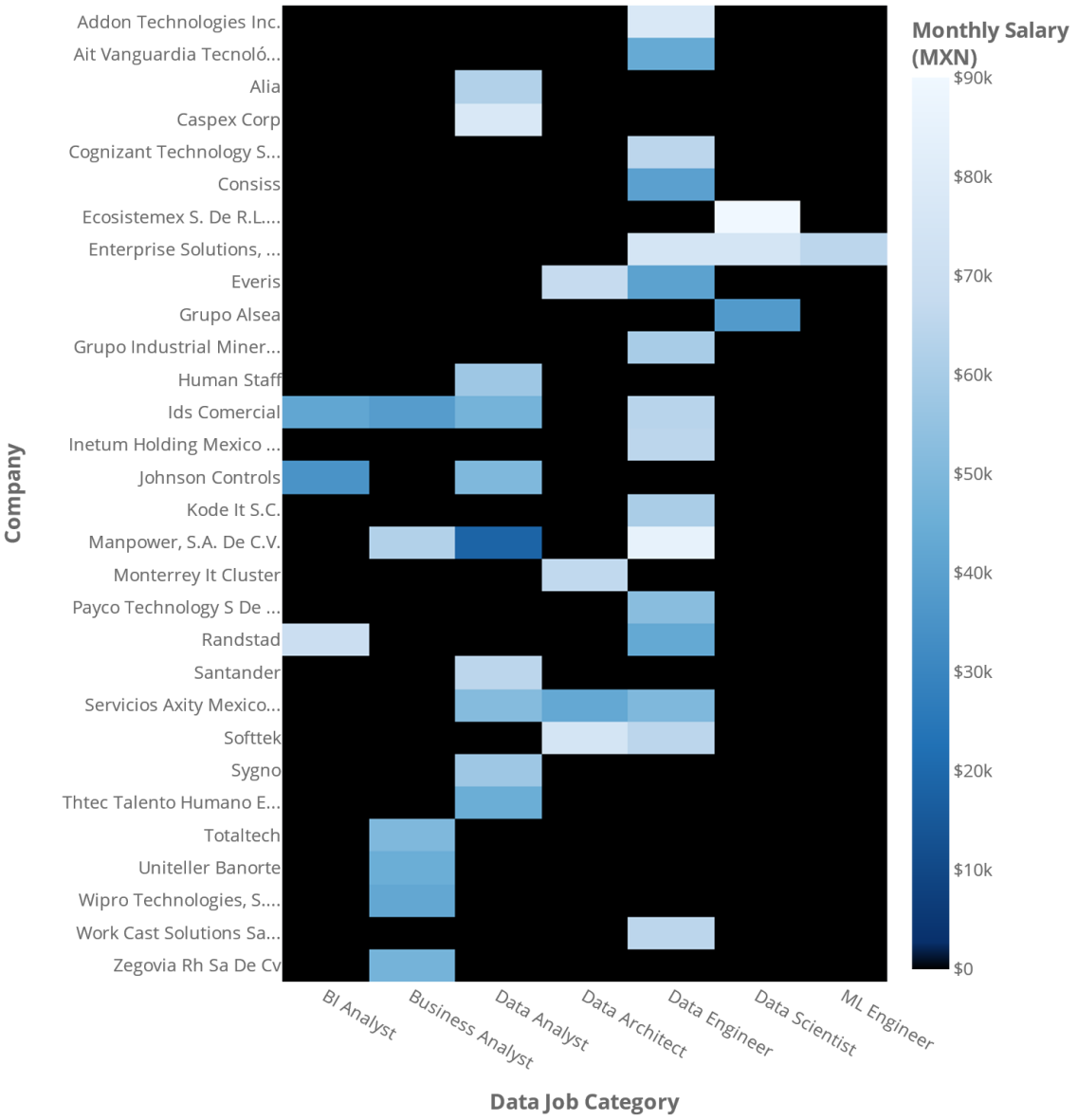

Figure 17. Companies with the highest average salaries by data job category in Mexico in November 2023 (own elaboration).

In view of figure 17, So, for **BI Analyst** positions, the company offering higher salaries were *Randstad* and *Ids Comercial*. For **Business Analyst** positions, the organizations offering higher salaries were *Manpower* and *Totaltech*. Moreover, for **Data Analyst** positions, the organizations offering higher salaries were *Caspex* and *Santander*. Furthermore, for **Data Architect** positions, the organizations offering higher salaries were *Softtek* and *Everis*. In addition, for **Data Engineer** positions, the organizations offering higher salaries were *Manpower*, and *Addon Technologies*. For **Data Scientist** positions, the organizations offering higher salaries were *Ecosistemex* and *Enterprises Solutions*. And, finally, for **ML Engineer** positions, the only company with a disclosed salary was *Enterprise Solutions*.

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 it. A screenshot of the dashboard is shown below (figure 18).

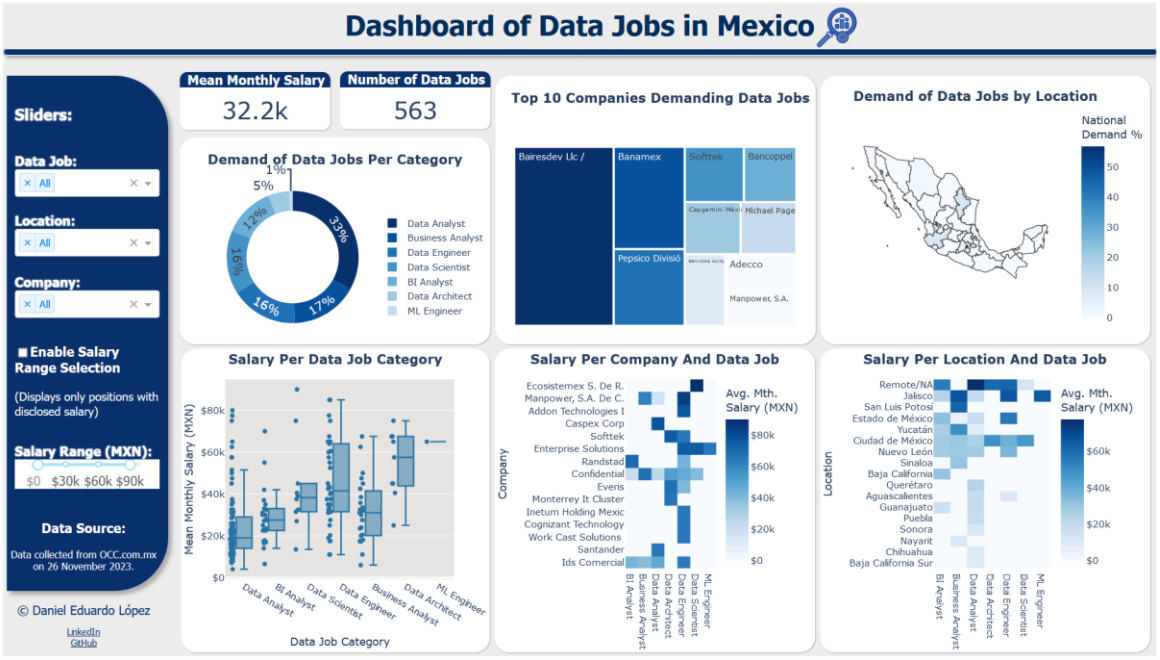

Figure 18. Dashboard of the data jobs in Mexico in November 2023 (own elaboration).

4. Conclusions

The data job category with the highest salaries in the Mexican labor market in November 2023 according to the OCC website were **Data Architect** and **ML Engineer**. Indeed, the average salary for Data Architect positions was not significantly lower than for ML Engineers. Thus, the present study's hypothesis is rejected. However, this result must be taken with caution as only one salary observation could be retrieved for the latter position.

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 **ML Engineer** and **Data Architect** 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. However, **Estado de México** and **Jalisco** were locations along with **remote positions** where the demand of data jobs was the highest after the capital city. Similarly, in those locations the highest salaries can be found.

Furthermore, the companies with the greatest demand of data positions were ***Bairesdev***, ***Banamex***, ***Pepsico*** and ***Softek***; nevertheless, the organizations that offered the highest salaries were ***Ecosistemex***, ***Caspex Corp***, ***Addon Technologies***, ***Enterprise Solutions***, and ***Softek***, 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 ***Bairesdev***, ***Pepsico*** or ***Softek***, and that there are some well-known organizations whose data jobs demand is not located in the capital region, such as ***Johnson Controls*** (Nuevo León).

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 Architect** and, certainly, **ML Engineer** vacancies could only be found in more specific organizations like tech consultancy firms or banks.

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 some of the least demanded data jobs categories: Data Scientist, Data Architect, and ML Engineer. Also, the collected data mostly corresponded to Ciudad de México, Nuevo León, Estado de México, and Jalisco, 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 Scientist, Data Architect, and ML Engineer positions as well as for other Mexican states.

5. References

ai-jobs.net. (2023). *Machine Learning Engineer vs. Business Intelligence Data Analyst*. Retrieved from https://ai-jobs.net/insights/machine-learning-engineer-vs-business-intelligence-data-analyst/

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

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.

López, D. E. (2023). *Data Jobs Salaries in Mexico in February 2023*. Retrieved from GitHub: https://github.com/DanielEduardoLopez/DataJobsMX2023

Olvera, E. (2023). *Reporte del Mercado Laboral de TI México 2023*. Retrieved from Hireline: https://hireline.io/mx/estudio-mercado-laboral-y-empleos-de-ti-mexico

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