



---

Nombre: **Roel Adrián De la Rosa Castillo**

Entrega: **Reporte final de "Los salarios"**

Clase: **Inteligencia artificial avanzada para la ciencia de datos Grupo 102**

Profesora: **Blanca Rosa Ruiz Hernandez**

Fecha: **18 de Septiembre de 2022**

---

## **1 Resumen**

Durante este reporte se analizará una base de datos que proviene de Kaggle. En este archivo nos encontramos con la información de los ingresos de diferentes personas en el área relacionada al análisis de datos. El objetivo planteado es que seamos capaces de reconocer algunos patrones dentro de la información y responder algunas preguntas que se encontrarán a continuación. A grandes rasgos, los resultados obtenidos son que en los últimos años ha habido un aumento en los salarios, cuales contratos son los que reciben mejores salarios y que la distribución de los salarios en dólares no pertenecen a una distribución normal.

## **2 Introducción**

Es importante que conozcamos el área e industria en el que vamos a trabajar durante nuestra vida laboral. Una buena parte de los alumnos de esta concentración somos de Ingeniería en Ciencias de Datos y Matemáticas, por lo que poder aprender lo máximo acerca de los potenciales ingresos es algo importante que tendrá bastante relevancia en nuestro futuro.

Es por eso que algunas de las preguntas que buscamos resolver son las siguientes:

1. ¿Se han incrementado los salarios a lo largo del tiempo?
2. ¿Qué tipo de contrato ofrece mejores salarios? ¿Qué tipo de contrato es el más conveniente?
3. ¿El salario en USD sigue una distribución normal? ¿Cómo lo podemos comprobar o rechazar?

### 3 Análisis de los resultados

Lo primero que se realizó fue una serie de gráficas y medidas estadísticas con las que podamos conocer los datos con los que estamos trabajando. A partir de aquí nos dimos cuenta que el promedio del salario en dolares es \$112297.87, la desviación estándar del salario en dolares es 70957.2594 y la mediana del salario en dolares es \$101570.

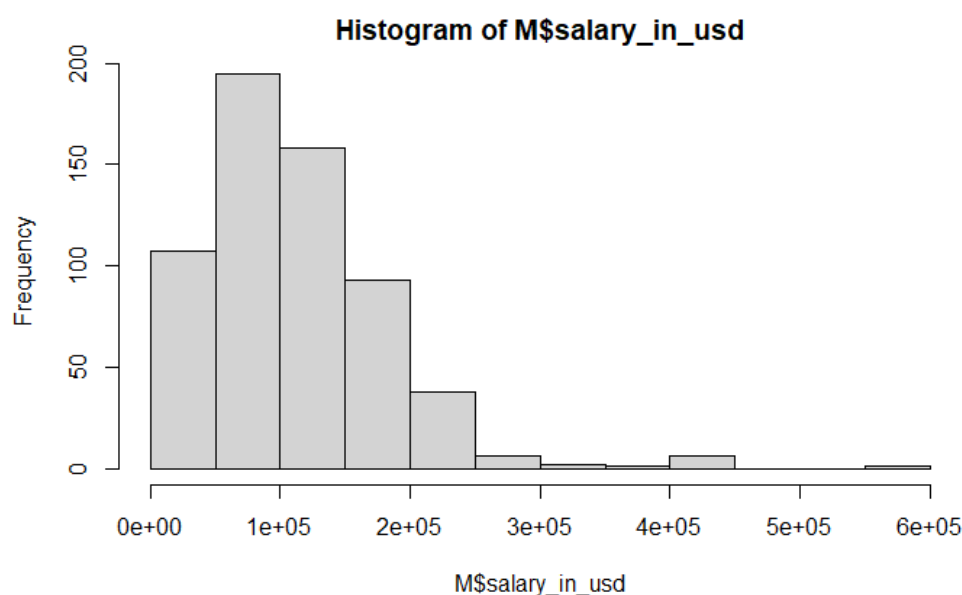


Figure 1: Histograma de los salarios en USD

A partir de estas dos primeras gráficas podemos observar la forma en la que los Salarios en USD se distribuyen. A su vez generamos un gráfico de pastel para observar las diferentes categorías que se tienen. Así vemos que la mayoría de las personas son de los Estados Unidos.

De la misma forma se hace un gráfico de pastel sobre el tamaño de las compañías.

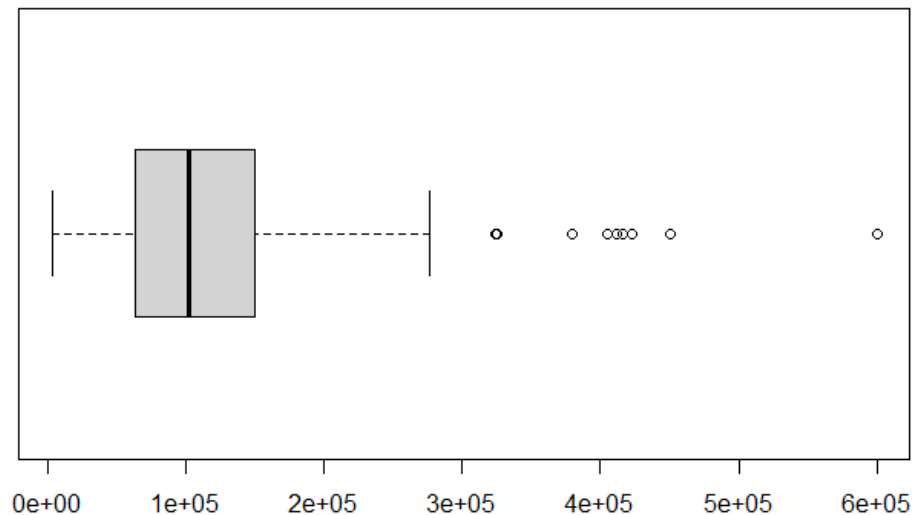


Figure 2: Boxplot de los salarios en USD

Finalmente se hace un gráfico de pastel sobre las diferentes posiciones analizadas. Se puede observar que en su mayoría son Data Scientists, Data Engineers, Data Analysts y Machine Learning Engineers, con muchas otras posiciones con menos cantidad de personas en ellas.

### 3.1 Pregunta 1

**¿Se han incrementado los salarios a lo largo del tiempo?**

A partir de la figura anterior podemos observar que tras cada año, el salario promedio ha aumentado. Sin embargo, solo se tiene información de 3 años, por lo que para poder asegurar esta conclusión sería necesario ya sea más años de información, u otra columna que pueda especificar el mes o trimestre del año.

### 3.2 Pregunta 2

**¿Qué tipo de contrato ofrece mejores salarios? ¿Qué tipo de contrato es el más conveniente?**

Tipos empleo con mayores salarios:

1. CT (Contract) - \$184575

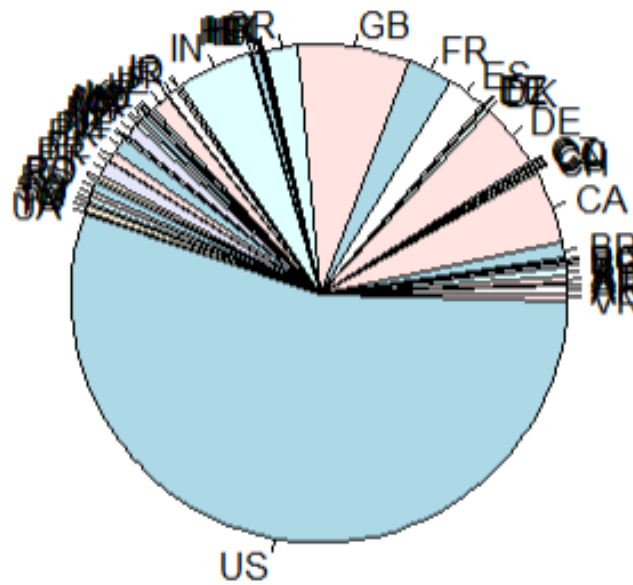


Figure 3: Pie Chart de los empleados

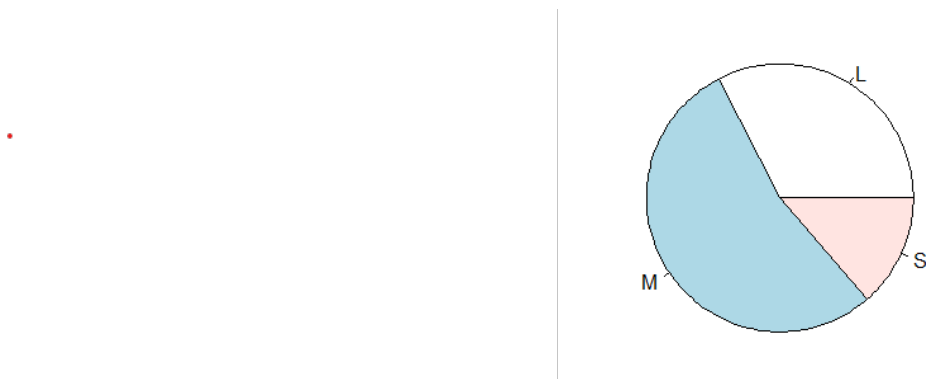


Figure 4: Pie Chart del tamaño de las empresas

2. FT (Full time) - \$113468

3. FL (Freelancer) - \$48000

4. PT (Part Time) - \$33070

### 3.3 Pregunta 3

¿El salario en USD sigue una distribución normal? ¿Cómo lo podemos comprobar o rechazar?

Lo primero que hicimos para saber si los salarios siguen una distribución normal es

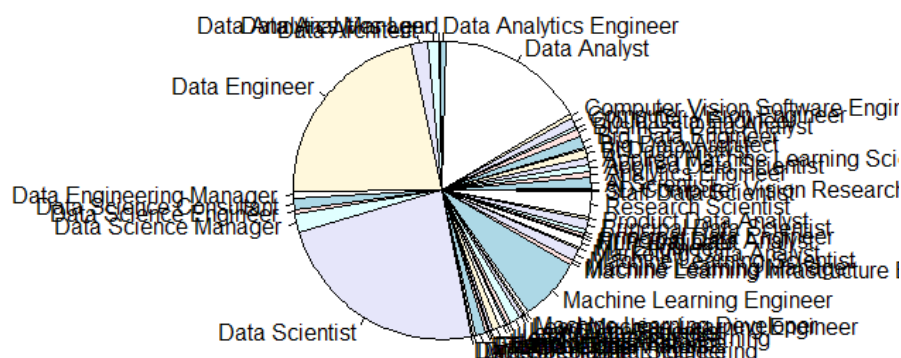


Figure 5: Pie Chart de las posiciones

Group.1 <int>	x <dbl>
2020	95813.00
2021	99853.79
2022	124522.01

Figure 6: Salario Promedio por año

observar el histograma.

A primera vista, no parece que siga una distribución normal, pero para poder asegurarlo se deben de tomar algunos tests estadísticos. En nuestro caso tomaremos el test de Shapiro:

$$\begin{aligned} H_0 : \rho &= 0 \text{ Se sigue una distribución normal} \\ H_1 : \rho &\neq 0 \text{ No se sigue una distribución normal} \end{aligned} \tag{1}$$

Regla de decisión: Rechazar  $H_0$  si el p-value  $< 0.05$  (2)

Dado que el p-value obtenido por la prueba de Shapiro es muy inferior a 0.05, se rechaza la hipótesis nula y se asume que no se sigue una distribución normal.

$$[1]$$

CT	184575.0
FL	48000.0
FT	113468.1
PT	33070.5

Figure 7: Salario Promedio por tipo de empleo

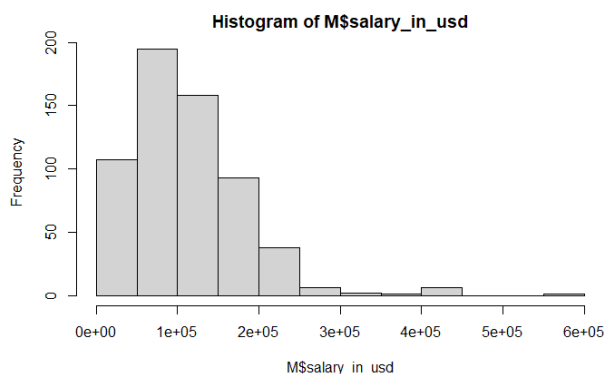


Figure 8: Histograma de los salarios en USD

```

shapiro-wilk normality test

data:  df$salary_in_usd
W = 0.89836, p-value < 2.2e-16

```

Figure 9: Shapiro Test de los salarios en USD

## 4 Conclusiones

Las conclusiones a las que hemos llegado son las siguientes:

- Por lo menos durante los últimos 3 años, el promedio de ingresos ha aumentado
- El tipo de empleo que es mejor pagado es CT (Contract)
- La distribución de los salarios no es una distribución normal.

## 5 Referencias

### References

- [1] Aleksander Dietrichson, P. (2022). 7.2 Prueba de Shapiro-Wilks — Métodos Cuantitativos. Retrieved 18 September 2022, from <https://bookdown.org/dietrichson/metodos->

cuantitativos/test-de-normalidad.html

## 6 Anexos

Liga con el archivo Markdown utilizado para el análisis:

<https://drive.google.com/drive/folders/1mxoJPfQE3mi-VQglX-bxEbdbivV37XTQ?usp=sharing>