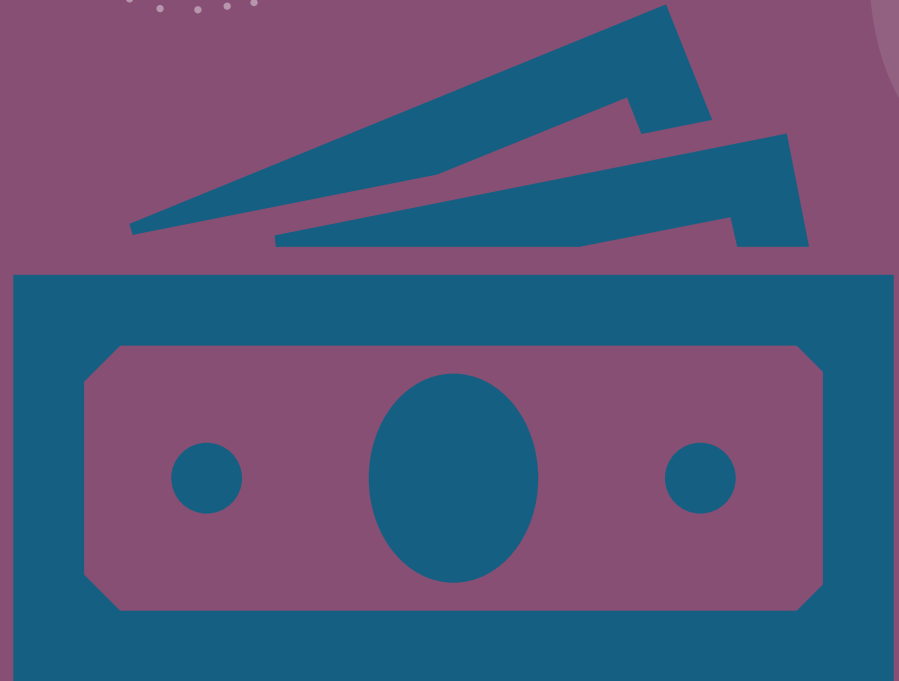


# Predicción de salarios IT

Scarramberg, Julián



***CODERHOUSE***

# Abstracto y audiencia

***CODERHOUSE***

Buscaremos encontrar relaciones entre los distintos rubros, lugares de trabajo y tecnologías que se utilizan con los salarios que perciben los trabajadores. Esto quiere decir por ejemplo que se podría identificar cuales rubros tienen mejores salarios, en que lugares se paga mejor o que tecnologías son más redituables.

Este estudio alcanzaría a todos aquellos trabajadores y empleadores relacionados con los puestos de trabajo relacionados al mundo de la tecnología. Estos se podrían beneficiar de este análisis en el caso por ejemplo en el que un trabajador quisiera conocer nuevas tecnologías, pudiendo identificar con mayor facilidad las más redituables.

# Objetivo

***CODERHOUSE***

¿Se pueden encontrar relaciones entre los rubros y los salarios para identificar oportunidades para aquellos que buscan empleo?

# Hipótesis y preguntas de interés

***CODERHOUSE***

Los trabajadores perciben diferentes salarios según donde trabajen.

Los trabajadores perciben diferentes salarios según que profesión ejercen.

¿A mayor experiencia los trabajadores perciben mejores sueldos?

¿Mientras más grande es la empresa los trabajadores perciben mejores salarios?

# Data Acquisition

***CODERHOUSE***

Los datos fueron adquiridos de manera externa (Second Party) provenientes de una encuesta realizada en Junio y Julio del año 2023 a través de la plataforma colaborativa Openqube y publicada por la comunidad Sysarmy que nuclea a profesionales del área de sistemas.

# Resumen de la información

## Dataset statistics

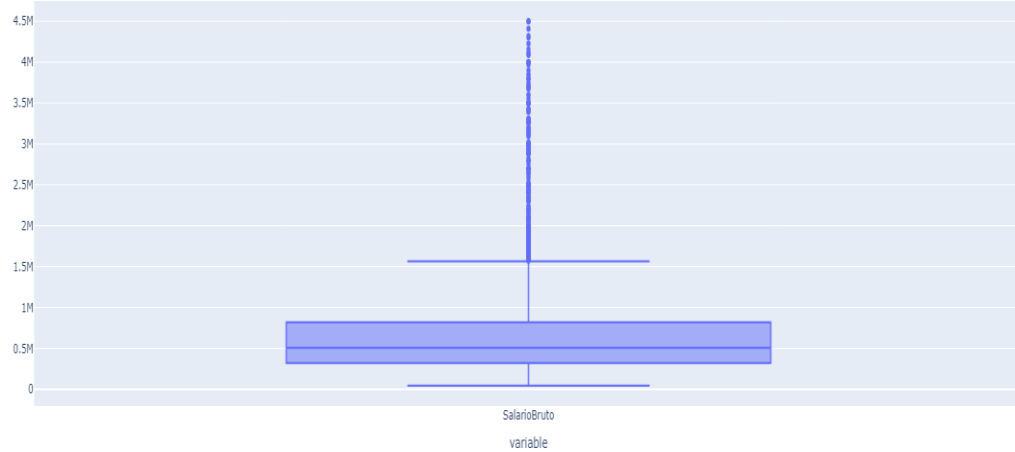
Number of variables	47
Number of observations	5422
Missing cells	38695
Missing cells (%)	15.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.9 MiB
Average record size in memory	369.0 B

## Variable types

Categorical	23
Numeric	13
Text	10
Boolean	1

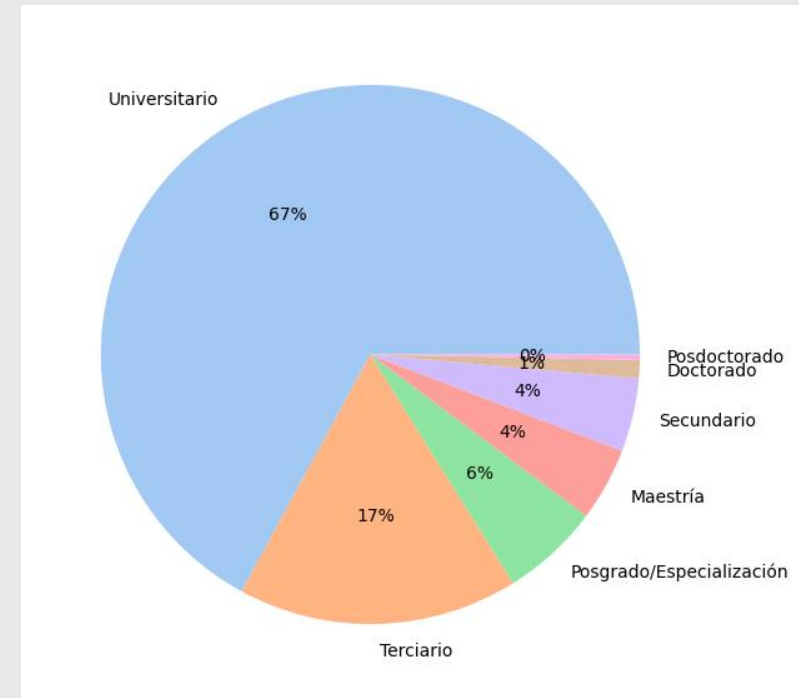
# Exploratory Data Analysis

***CODERHOUSE***



Boxplot de variable “SalarioBruto”: se observa que tenemos una gran cantidad de valores atípicos.

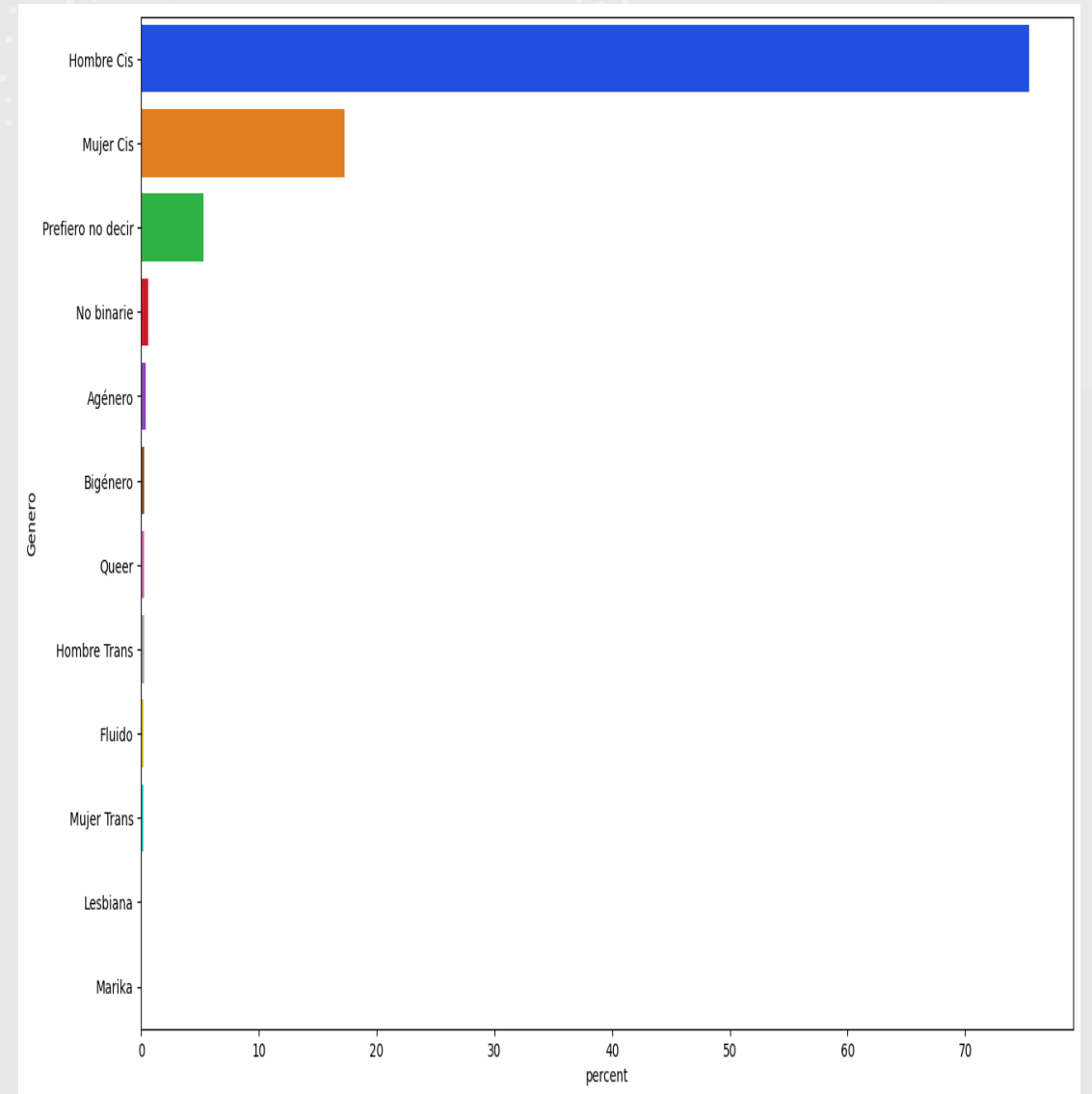
Piechart de variable “MaximoNivelEstudios”: se observa que la mayoría de los encuestados son graduados universitarios.

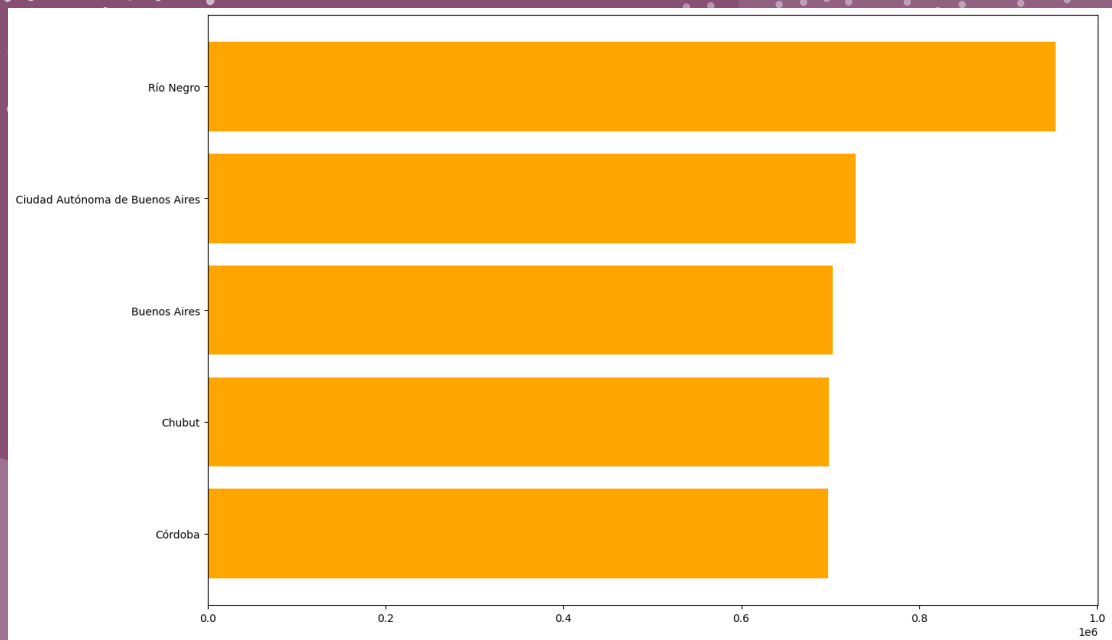




Barchart de variable “SalarioBruto”: se observa que nuestra información abarca principalmente a hombres.

**CODERHOUSE**





Barchart de variables “SalarioBruto” y “Provincia”: se observa en cuales provincias se paga mayor sueldo promedio.

Scatterplot entre variables “SalarioBruto” y “Experiencia”: se observa una moderada relación entre el salario y la experiencia.



**CODERHOUSE**

# Modelaje

***CODERHOUSE***

Como nuestro objetivo es predecir salarios nos encontramos en un problema de regresión.

Con esto en mente utilizaremos diferentes herramientas para realizar los análisis e intentaremos también optimizar estas para alcanzar los mejores resultados posibles.

# Conclusiones

***CODERHOUSE***

La métrica en que utilizaremos para evaluar el modelo será el Error Absoluto Medio o MAE.

El mejor resultado que conseguimos para el MAE fue de 257909.

Este lo conseguimos a través del modelo Gradient Boosting optimizado por RandomizedSearchCV.

Considerando este número, observamos que nuestras predicciones se alejan significativamente de lo que buscamos.