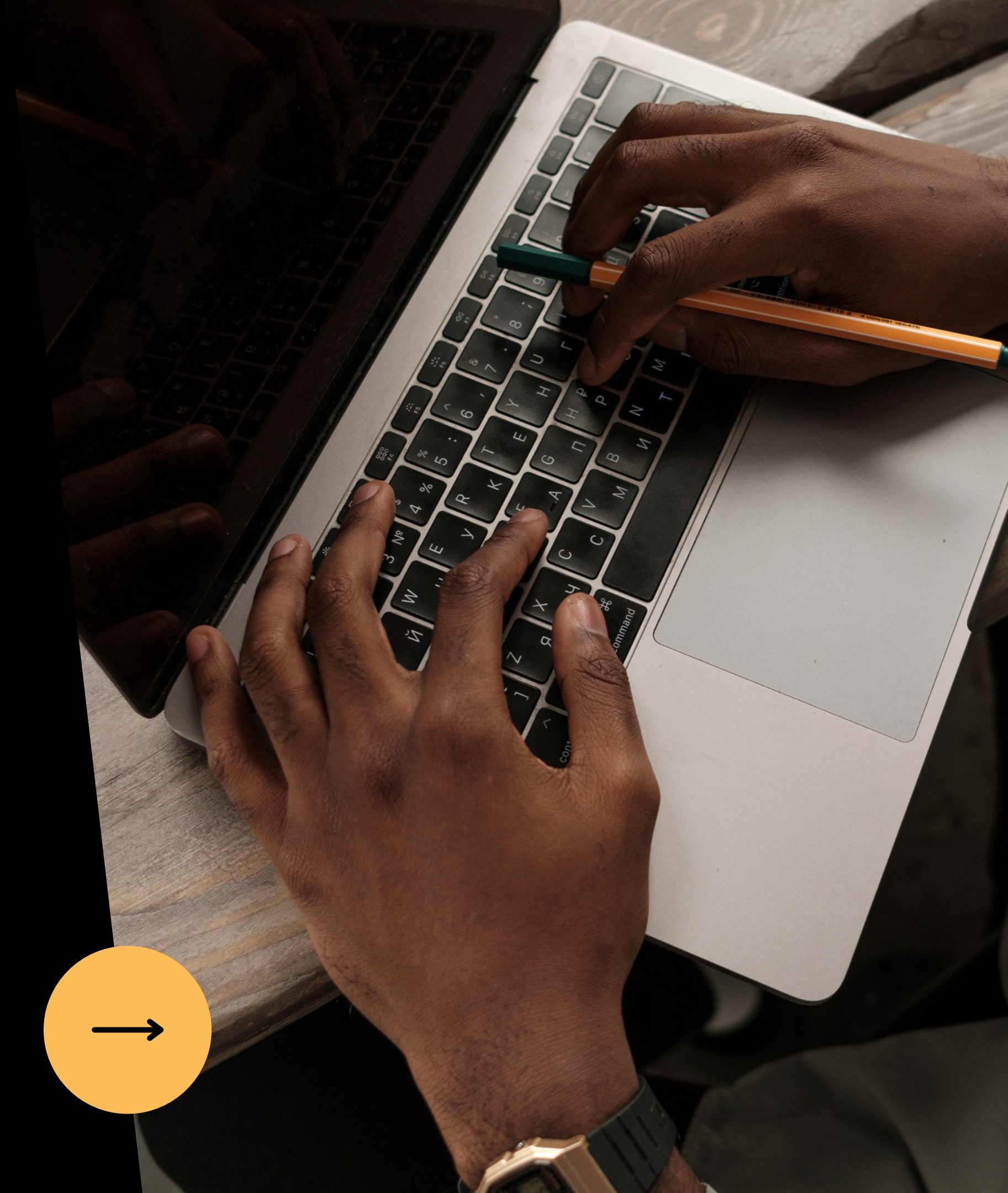


**Beatriz Zatarain**

Marzo 2023

# Predicción Salarial

Proyecto Kaggle



# Contenidos

<b>01</b>	<b>Exploración de datos</b>
<b>02</b>	<b>Limpieza de datos</b>
<b>03</b>	<b>Creación del modelo</b>
<b>04</b>	<b>Resultado</b>

# Exploración de los datos

01 Ausencia de nulos y filas repetidas

7 columnas categóricas vs. 4 numéricas.

02

03 Baja colinealidad entre columnas numéricas.



# Limpieza de datos

## **Get Dummies**

Aplicación directa para las  
columnas con pocas  
categorías

**Etiqueta en función  
de la frecuencia +  
get dummies**

**Eliminación de  
columnas  
redundantes**

**Unificación de  
columnas y  
transformación  
numérica**

# Resultado

Una vez dividida la información en muestras de entrenamiento y testeo se ha empleado una función para determinar el modelo de predicción óptimo.

## Modelo Elastic

Resultante de la sugerencia al aplicar la función.

## Hiperparámetros

$\alpha=0.5$  y  $l1\_ratio=0.5$

## Error (RMSE)

57238.80854422544

## Error con todos datos

52978.03633

Gracias