

Universidad de los Andes

Edgar Balaguera Ascencio

MIAD

Documentación del modelado

1. Variables en base de datos original:

La base de datos está compuesta por 17 variables que se describen a continuación:

No	Nombre	Tipo	Descripción
1	How old are you?	Texto	Rangos de edades
2	Industry	Texto	Sector laboral
3	Job title	Texto	Tipo de profesión
4	Additional context on job title	Texto	Sector donde se han desempeñado
5	Annual salary	Número	Salario anual recibido
6	Other monetary comp	Número	Compensaciones salariales
7	Currency	Texto	Moneda
8	Currency - other	Texto	Descripción de otro tipo de moneda
9	Additional context on income	Texto	Descripciones de otros ingresos adicionales.
10	Country	Texto	País donde se encuentra ubicado
11	State	Texto	Estado donde se encuentra
12	City	Texto	Ciudad donde se encuentra ubicado
13	Overall years of professional experience	Texto	Total de años de experiencia
14	Years of experience in field	Texto	Años de experiencia en el campo profesional
15	Highest level of education completed	Texto	Nivel de educación
16	Gender	Texto	Tipo de Género
17	Race	Texto	Tipo de raza

2. Variables luego de modeladas.

En esta parte primero se tradujeron las variables y se asignaron nombres en español, se transformaron algunas y se crearon otras.

Las variables transformadas según TMR son:

- Salario Anual COP
- Compensaciones COP

Variables creadas:

- Salario y compensaciones COP

A continuación se describen las nuevas variables:

No	Nombre	Tipo	Descripción
1	Edad	Texto	Rangos de edades
2	Industria	Texto	Sector laboral
3	Profesión	Texto	Tipo de profesión
4	Contexto profesión	Texto	Sector donde se han desempeñado
5	Salario COP	Fórmula	Salario anual recibido en COP según TMR
6	Otras compensaciones COP	Fórmula	Compensaciones salariales en COP según TMR
7	Moneda	Texto	Moneda
8	Otras Monedas	Texto	Descripción de otro tipo de moneda
9	Contexto Ingreso	Texto	Descripciones de otros ingresos adicionales.
10	País	Geográfica	País, variable transformada por ubicación geográfica
11	Estado	Geográfica	Estado, variable transformada por ubicación geográfica
12	Ciudad	Geográfica	Ciudad, variable transformada por ubicación geográfica
13	Años de experiencia Total	Texto	Total de años de experiencia
14	Años de experiencia en el campo	Texto	Años de experiencia en el campo profesional
15	Nivel de educación	Texto	Nivel de educación
16	Genero	Texto	Tipo de Género
17	Raza	Texto	Tipo de raza
18	Salario anual y compensaciones COP	Fórmula	Suma de salario anual y compensaciones

3. Descripción del modelado

Este proceso se realizó por etapas:

Extracción:

En esta etapa se extrae y visualiza la base de datos mediante la herramienta data studio de google.

Limpieza:

En esta etapa se realiza la limpieza de la base de datos, debe ser un proceso exhaustivo en el que se analiza cada variable teniendo en cuenta datos faltantes, datos repetidos, datos con asignación diferente pero que representan el mismo grupo. En esta base de datos se realizó limpieza de las variables ciudad y país para garantizar la homogeneidad de los datos que representaban la misma ubicación geográfica.

Transformación:

Sobre la base de datos se realizaron transformaciones en Salario COP y otras compensaciones COP, estas se realizaron por fórmula según la Tasa de cambio con fecha de 28 de febrero de 2022. En esta etapa también se creó una variable adicional denominada Salario y compensaciones COP en la que se creó una fórmula para sumar las dos variables salario COP y otras compensaciones COP.

Visualización y publicación:

En esta etapa se creó un dashboard mediante la herramienta de data studio, se crearon los gráficos pertinentes según las indicaciones dadas y direccionados sobre la historia que se buscaba narrar “¿Cuánto Quieres ganar al año ?” Finalmente se publicó a través de github.

URL: <https://github.com/edgarbalagueraa/MIAD-ML/blob/main/README.md>