

## Documentación ejercicio semana 3 Storytelling

Juan Felipe Uribe - 201329374

### Documento original:

Este trabajo se realizó con la base de datos “Ask a Manager”, pueden revisar el formulario en <https://www.askamanager.org/2021/04/how-much-money-do-you-make-4.html> y los datos quedan alojados en un Google Sheets público que se actualiza constantemente: <https://docs.google.com/spreadsheets/d/1IPS5dBSGtwYVbjsfbaMCYIWnOuRmJcbequohNxCyGVw/edit?resourcekey#gid=1625408792>.

En la base original encontrarán las siguientes variables:

- Timestamp: Fecha en la cuál fue tomado el registro, tipo date.
- How old are you?: Rango de edad de la persona encuestada, tipo string.
- Industry: Industria en la que trabaja, tipo string.
- Job title: Título de puesto laboral, tipo String.
- Additional context on job title: Comentarios adicionales sobre el tipo de puesto laboral, tipo string.
- Annual salary: Salario anual que percibe el encuestado, tipo string.
- Other monetary comp: Otros ingresos monetarios, tipo string.
- Currency: Moneda en la cual ingresa su salario, tipo string.
- Currency – other: Moneda en la cual ingresa su salario (si puso other en Currency), tipo string.
- Additional context on income: Información adicional sobre sus ingresos, tipo string.
- Country: País de trabajo, tipo string.
- State: Estado de trabajo (si vive en USA), tipo string.
- City: Ciudad de trabajo, tipo string.
- Overall years of professional experience: Rango de años de experiencia laboral, tipo string.
- Years of experience in field: Rango de años de experiencia laboral en el campo actual, tipo string.

- Highest level of education completed: Nivel más alto de educación terminada, tipo string.
- Gender: Género, tipo string.
- Race: Raza o grupo étnico, tipo string.

Para realizar el reporte se agregaron las siguientes variables:

- salario\_anual: Es el salario anual convertido a pesos colombianos (COP) según la TRM actual, tipo int.
- compensaciones: Son las compensaciones adicionales recibidas convertidas a pesos colombianos (COP) según la TRM actual, tipo int.
- Total\_ingresos: Suma de salario\_anual y compensaciones. Son los ingresos anuales totales, tipo int.
- Lugar: Es la concatenación de ciudad, estado y país unidos con “,” , tipo string.
- Pais: Es el país en el que trabaja (homogenizado), tipo string.
- Ciudad, es la ciudad en la que trabaja (homogenizado), tipo string.

Para replicar el proceso, después de cargar los nuevos datos, deben seguir los siguientes pasos:

1. Crear la variable salario\_anual: Esta variable se crea transformando la variable Annual Salary. primero hay que quitarle los puntos o comas que las personas suelen poner para separar mil, millón, etc. Posteriormente, teniendo ya una variable numérica, se multiplica por la TRM según la variable Currency. En este reporte se eliminaron las monedas diferentes a las usuales y se utilizó la siguiente conversión:

```
{"USD":3917.52, "GBP": 5343.23, "CAD":3094.32, "EUR": 4471.69, "AUD/NZD": 2811.54,"CHF":4253.90, "ZAR":258.89, "SEK":422.01, "HKD":505.09, "JPY":34.13}
```

Para el futuro se puede extender las monedas permitidas y cambiar la estos valores por la TRM se ese momento. Después se multiplicó por 1000 todos los valores inferiores a 12.000.000 (pues mucha gente pone números como 10 refiriéndose a 10.000).

2. Se crea la variable compensaciones: Se sigue el mismo proceso que en salario\_anual pero con la variable Other monetary comp. No se realiza el último paso (de multiplicar por 1000).
3. Se crea la variable Total\_ingresos: Se suma salario\_anual y compensaciones.
4. Se crea la variable Lugar concatenando las variables de city, state y country unidos con “,”.
5. Se eliminan los datos que tienen Lugar vacío.
6. Se crea una aplicación en Nominatim con ayuda de geopy con el siguiente código (importante importar nominatim de geopy.geocoders):

```
geolocator = Nominatim(user_agent="Mi_aplicacion_intento1")
```

```
geocode = partial(geolocator.geocode, language="en")
```

El user\_agent puede tener el nombre que se desee.

7. Se busca la información de localización según el Lugar del dato. Esto retorna un json. Se hace lo mismo solo con la City, después solo con State y después solo con Country.
8. Se extrae el país que hay en el json (es como un diccionario entonces solo es llamar “country”).
9. Se extrae la ciudad, para esto se hace primero con “city”, después con “town”, “place”, “suburb”, “village”, “locality”, “county”. Esto es porque la denominación real del lugar que pusieron como ciudad, dependiendo del tamaño, puede variar.
10. Por último, si no se encontró en algún sitio la ciudad, se mira por latitud y longitud el centro del estado (posteriormente del país si no funciona) y se vuela a ver el paso 9.

Para finalizar, ya se tienen los datos en el mismo formato que requiere el dashboard.

Todo el código se encuentra documentado en el notebook.