**Documentación del modelado y actualización de datos de la base de datos Ask A Manager Salary Survey 2021**

1. **Tipo de dato y descripción de cada una de las variables de la base de datos original.**
   * **Timestamp:** variable de tipo fecha e indica la fecha (en formato m/d/yyyy h:mm) en que se diligenció el formulario.
   * **How old are you?:** lista de variables categóricas que muestra el rango de edad en el que se encuentra el encuestado.
   * **What industry do you work in?:** lista de variables categóricas que muestran la industria en la que trabaja el encuestado.
   * **Job title:** variable categórica que indica el cargo del encuestado.
   * **If your job title needs additional context, please clarify here:** variable categórica que indica si el cargo del encuestado necesita tener un contexto adicional.
   * **What is your annual salary? (You'll indicate the currency in a later question. If you are part-time or hourly, please enter an annualized equivalent -- what you would earn if you worked the job 40 hours a week, 52 weeks a year.):** variable numérica que indica el salario anual del encuestado.
   * **How much additional monetary compensation do you get, if any (for example, bonuses or overtime in an average year)? Please only include monetary compensation here, not the value of benefits.:** variable numérica que indica la compensación monetaria que recibe el encuestado en caso de tener.
   * **Please indicate the currency:** lista con variables categóricas que muestran la moneda en que el encuestado recibe el salario.
   * **If "Other," please indicate the currency here:** variable categórica que indica la moneda en que el encuestado recibe el salario en caso de la opción anterior haya sido *Other*.
   * **If your income needs additional context, please provide it here:** variable categórica que indica si los ingresos del encuestado necesitan tener un contexto adicional.
   * **What country do you work in?:** variable categórica que indica el país en el que trabaja el encuestado.
   * **If you're in the U.S., what state do you work in?:** lista con variables categóricas que muestran el estado donde trabaja el encuestado en caso tal de que resida en Estados Unidos.
   * **What city do you work in?:** variable categórica que indica la ciudad en donde trabaja el encuestado.
   * **How many years of professional work experience do you have overall?:** lista de variables categóricas que muestran el rango de años de experiencias en general del encuestado.
   * **How many years of professional work experience do you have in your field?:** lista de variables categóricas que muestran el rango de años de experiencias en el campo en el cual se desempeña el encuestado.
   * **What is your highest level of education completed?:** lista de variables categóricas que muestran el nivel educativo más alto completado por el encuestado.
   * **What is your gender?:** lista de variables categóricas que muestran el género del encuestado.
   * **What is your race? (Choose all that apply.):** lista de variables categóricas que muestran la raza del encuestado.
2. **Variables de la base de datos modelada.** En esta etapa, se renombraron los nombres de todas las columnas, en donde se cambió el idioma a español y se recortaron sus nombres.
   * **FECHA:** variable de tipo fecha e indica la fecha (en formato m/d/yyyy h:mm) en que se diligenció el formulario.
   * **EDAD:** variable categórica que indica el rango de edad en el que se encuentra el encuestado.
   * **INDUSTRIA:** variable categórica que indica la industria en la que trabaja el encuestado.
   * **CARGO:** variable categórica que indica el cargo del encuestado.
   * **CONTEXTO CARGO:** variable categórica que indica si el cargo del encuestado necesita tener un contexto adicional.
   * **SALARIO ANUAL:** variable numérica que indica el salario anual del encuestado.
   * **COMPENSACION:** variable numérica que indica la compensación monetaria que recibe el encuestado en caso de tener.
   * **MONEDA:** variable categórica que indica la moneda en que el encuestado recibe el salario.
   * **OTRA MONEDA:** variable categórica que indica la moneda en que el encuestado recibe el salario en caso de la opción anterior haya sido *Other*.
   * **CONTEXTO INGRESOS:** variable categórica que indica si los ingresos del encuestado necesitan tener un contexto adicional.
   * **PAIS:** variable categórica que indica la ciudad en donde trabaja el encuestado.
   * **ESTADO USA:** lista con variables categóricas que muestran el estado donde trabaja el encuestado en caso tal de que resida en Estados Unidos.
   * **CIUDAD:** variable categórica que indica la ciudad en donde trabaja el encuestado.
   * **EXPERIENCIA:** variable categórica que indica el rango de años de experiencias en general del encuestado.
   * **EXPERIENCIA CAMPO:** variable categórica que indica el rango de años de experiencias en el campo en el cual se desempeña el encuestado.
   * **EDUCACION:** variable categórica que indica el nivel educativo más alto completado por el encuestado.
   * **GENERO:** variable categórica que indica el género del encuestado.
   * **RAZA:** variable categórica que indica la raza del encuestado.
3. **Procedimiento para modelado y actualización de datos.**

Gran parte del proceso de modelado de datos se realizó en Python, en donde se inició leyendo la base de datos y renombrando las columnas. Luego de esto, se removieron las comas de los valores de la columna SALARIO ANUAL. Seguidamente, se inspeccionaron los registros de los encuestados que tuvieran un salario anual superior a un millón de dólares, y se encontró que hay una persona que devenga 102 millones de unidades monetarias. A pesar de que la moneda que aparece es **USD**, en el campo OTRA MONEDA y CONTEXTO INGRESOS se observa que la moneda en que recibe el salario es en pesos colombianos, por lo que se cambió a **Other** en el campo **MONEDA**. Una vez hecho esto, se cambiaron los valores **nan** de la columna **COMPENSACION** por 0. Después, se limpiaron las variables **PAIS** y **CIUDAD**, en donde se removieron los espacios y se convierten en mayúsculas todos los registros. Para esta parte, debido a que se decidió visualizar el salario anual promedio por país, se realizó una limpieza más profunda a dicha variable, en donde se identificaron todos los países que tuvieran distintos nombres para poderlos unificar en uno solo. Una vez realizado esto, se redujo el número de valores distintos de las columnas **INDUSTRIA** y **GENERO** con el fin de poder visualizar mejor la información de ambas variables en función de los ingresos anuales promedio. Finalmente, se exporta la base de datos resultante como un libro de Excel el cual se utilizó para la construcción del dashboard en Power BI.

Para la visualización de la información en Power BI, se utilizaron las siguientes tres fuentes de datos:

* El libro de Excel que se exportó desde Python.
* <https://tradingeconomics.com/currencies?base=cop>
* <https://www.sport-histoire.fr/en/Geography/Currencies_countries_of_the_world.php>

Los dos últimos fueron muy útiles ya que permitieron identificar la moneda del país en donde trabaja cada encuestado, la tasa de cambio de cada una de las monedas frente a la colombiana y la fecha de corte de la tasa de cambio.

A continuación, se explicará la aplicación del modelado que se ha diseñado.

Antes de cargar las tres fuentes de datos, lo primero que se debe hacer es abrir el código Python que realiza el modelado de datos y luego ejecutarlo. Sin embargo, en este caso, se deben tener en cuenta algunos aspectos. El primero de ellos es saber el nombre de la base de datos original que se va a leer y el formato que tiene. Se sugiere que dicho archivo esté en **formato CSV** y el nombre sea **salary\_survey**, ya que esto no obligaría a cambiar el código de Python. Otro aspecto que se debe tener en cuenta es identificar el lugar donde debe estar almacenada la base de datos que se va a importar desde Python, la cual debe estar en el directorio predeterminado. El último aspecto es tener claro el sitio dónde se guardará el archivo de Excel que resulta de realizar la exportación desde Python puesto que es una de las fuentes de datos que se importará a Power BI.

Luego de ejecutar el código Python que permite realizar gran parte del modelado de datos, se debe abrir Power BI y se cargan las tres fuentes anteriormente mencionadas, en donde se debe tener en cuenta que una de ellas es un archivo de Excel, mientras que las otras dos son sitios web. Para el libro de Excel, en la pestaña **Inicio** hay una opción llamada **Obtener datos** y clicar en **Libro de Excel**. Luego, se debe buscar y seleccionar el archivo que se ha creado desde Python, se selecciona la hoja y clicar en **Cargar**. Ahora, se procede a importar los sitios web. Para esto, nuevamente se dirige a la opción **Obtener datos** de la pestaña **Inicio** y, esta vez, se clica en **Web**. En el espacio debajo de **Dirección URL**, se pega uno de los dos sitios web y se clica en **Aceptar**. Después, se selecciona la casilla de la primera tabla que aparece en la venta **Navegador** y se clica en **Cargar**. El mismo paso anterior se realiza con el otro sitio web. Luego de haber importado las tres fuentes de datos a Power BI, se realizarán algunas transformaciones. Para eso, se clica en la opción **Transformar datos** de la pestaña **Inicio** la cual abrirá una ventana en donde se podrán observar las tres fuentes previamente cargadas al lado izquierdo. Las tres tablas se renombrarán clicando dos veces en cada una de ellas. La primera fuente importada a Power BI fue el libro de Excel y se llamará **Ingresos**; la fuente que contiene las tasas de cambio se renombrará a **Tasa de cambio**; y, la faltante se le pondrá **Monedas**. Una vez hecho esto, se clica en la tabla **Tasa de cambio**, se seleccionan las columnas **Day**, **%**, **Weekly**, **Monthly** y **YoY**, y se eliminan clicando en la opción **Quitar columnas** de la pestaña **Inicio**. Luego, se selecciona la primera columna que tiene por nombre **Crosses** y, en la pestaña **Transformar**, se escoge la opción **Extraer**. Allí, se cicla en **Últimos caracteres** y, en el espacio debajo de **Recuento** se anota **3** y se clica en **Aceptar**. Despuesto de esto, se cicla dos veces en cada uno de los nombres de tres las columnas y se anota **Codigo moneda**, **Tasa de cambio** y **Fecha**, respectivamente. Seguidamente, se selecciona la tabla **Monedas** y se cicla en la opción **Usar la primera fila como encabezado** de la pestaña **Inicio**. Una vez terminado eso, se elimina la columna **Currency** de la misma manera como se eliminaron las columnas de la tabla **Tasa de cambio**. Luego de esto, en la esquina superior izquierda se cicla en la opción **Cerrar y aplicar** de la pestaña **Inicio**. Después, en la parte lateral izquierda se podrán observar tres íconos y se selecciona el que está en la mitad en forma de tabla el cual dice **Datos** al pasar el curso sobre él. Allí, en la parte lateral derecha se podrán observar las tres tablas que se han cargado. Se selecciona la tabla Moneda y se cicla en la opción **Nueva columna** de la pestaña **Herramientas de tablas**. En el panel de funciones, se anota la siguiente función:

Luego, se vuelve a agregar una nueva columna y se escribe la siguiente función en el panel de funciones:

Las anteriores funciones permitirán retornar la tasa de cambio y en la **Fecha** en la tabla **Moneda**.

Una vez realizado esto, se selecciona la tabla **Ingresos** y se agregan cuatro columnas de la misma forma como se detalló anteriormente, y se anotan las siguientes funciones en el panel de funciones:

El anterior procedimiento permita realizar todo el proceso de modelado de las tres fuentes importadas para poder realizar la visualización de los datos.

En seguida, se presentará el procedimiento de actualización de los datos:

En caso tal de que se siga alimentado la base de datos original, se debe descargar dicho conjunto de datos para ejecutar el código Python que se encargará de importarlo, depurarlo y exportarlo a un nuevo libro de Excel. Luego de esto, en la opción **Transformar datos** de la pestaña **Inicio** en Power BI, se selecciona **Configuración de origen de datos**. Allí, se cicla en la fuente de Excel y se luego en **Cambiar origen…**, en donde se abrirá una ventana donde se examina la ruta de acceso al nuevo archivo de Excel previamente exportado desde Python. Una vez hecho esto, se cicla en aceptar y después en cerrar. Finalmente, se selecciona la opción Actualizar de la pestaña Inicio para que se puedan refrescar las visualizaciones con los nuevos datos.