

Dashboard 2: visualización y documentación para narrar a un equipo de trabajo

Pregunta 1

Vamos a poner en práctica lo estudiado a lo largo de la semana con una base de datos real. Es una fuente que nos habla de los sueldos que tienen profesionales en diferentes disciplinas (principalmente en Estados Unidos) y estos datos son recolectados por “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>

Algunos campos del formulario son campos de texto donde las personas pueden ingresar libremente la información y esto naturalmente produce diferencias en las formas de escribir datos que deberían ser homogéneos para poderlos analizar y visualizar. Por ejemplo: “City” o “Country”. Al mismo tiempo, el campo de “Annual Salary” es propenso a recibir valores que no corresponden con la medida esperada, por ejemplo, alguien puede escribir 80 (refiriéndose a coloquialismos como 80K) para decir que su salario anual es 80,000. El resultado que encuentran en la base de datos es una buena oportunidad para poner en práctica sus habilidades de limpiar, filtrar y en general modelar los datos para ser analizados y visualizados.

Mientras realizan el ejercicio, imaginen que están trabajando en un equipo (empresa, colectivo, proyecto en grupo) y lo que están produciendo son los insumos para que como equipo puedan investigar y diseñar unos objetivos propios a partir de la base de datos. El **dashboard** y **documentación** que van a crear, son materiales de trabajo que comunican **internamente** las potencias de un proyecto. En ese sentido, no necesariamente está pensado para un público general y por ende se busca más la utilidad y claridad para un equipo que esté trabajando con esos datos. Esto nos sirve de simulacro y preparación para las dinámicas del proyecto final y ponemos en práctica las formas de narrar a nosotros mismos y al público especializado.

Modelar los datos

Para poder visualizar y documentar tienen que realizar los siguientes pasos de modelado primero:

1. Extraer la base de datos “Ask A Manager Salary Survey 2021” de AskAManager.org (<https://docs.google.com/spreadsheets/d/1IPS5dBSGtwYVbjsfbaMCYIWnOuRmJcbequohNxCyGVw/edit?resourcekey#gid=1625408792>)
2. Limpiar los campos de “Country” y “City” para homogenizar los nombres de los lugares.
3. Crear 2 campos nuevos: “salario_anual” y “compensaciones” convirtiendo sueldos y compensaciones a **Pesos Colombianos** basados en la tasa de cambio del día que hacen el ejercicio.
4. Crear un campo adicional sumando salario anual y compensaciones en pesos colombianos.

Presentar un dashboard básico:

Después de realizar el modelado de los datos elaboren un dashboard que contenga los siguientes módulos:

- Módulo 1: Contador básico con total de respuestas.
- Módulo 2: Mapa (decidir si lo muestran por ciudades o sólo países).
- Módulo 3: Relación industria y promedio de salario. Ustedes deciden el tipo de visualización que les permita describir esta relación efectivamente, usar los campos nuevos en pesos colombianos que crearon anteriormente.
- Módulo 4: Libre – Ustedes deciden qué quieren visualizar y eligen la estrategia para hacerlo.
- Créditos a la base de datos, incluir la fecha en la que extraen la TRM.

Documentar el modelado

Ahora, elaboren un documento escrito en español con las siguientes secciones:

- Variables en base de datos original. Tengan en cuenta que, al describir esta base de datos, los nombres de las variables deben estar escritas en inglés, tal cual las encuentran en la base de datos original, mientras que las descripciones sí deben estar escritas en español. Esto permite a los otros miembros del grupo volver a la base original y reconocer las variables que están describiendo:
 - Tipo de variable (número, texto, etc.).
 - Descripción de variable.
- Variables luego de modeladas. Durante el modelado, ustedes pueden cambiar el nombre de las variables, las pueden dejar en inglés o traducirlas al español. En cualquier caso, por buena práctica, decidan qué idioma van a usar y sean consistentes con esa decisión en todas las variables:
 - Tipo de variable.
 - Descripción corta, máximo 1 párrafo.
- Describir paso a paso lo que una persona debería hacer para actualizar los datos y aplicar el modelado que han diseñado. Piensen esta sección como una ayuda a alguien que los va a reemplazar en el trabajo. Esta persona debe entender con precisión cómo replicar lo que diseñaron. Este paso a paso asume que la estructura de la base de datos original no cambia en nuevas versiones. Puede ser tan detallada como ustedes quieran, pero debe ser muy claro cómo crear los nuevos campos y procedimientos para limpiar.

¡Puntos extra!

Ustedes pueden alcanzar el 100% de la nota cumpliendo con los requisitos anteriormente mencionados. Pero, esta semana pueden acumular 10% extra que se suma a la entrega con la nota más baja que tengan al final del curso. Para sumar estos puntos extras, simplemente deben presentar su trabajo de esta semana completamente unificado en un sitio web. Es decir, presentar el dashboard y la documentación en una página web como si fuera el sitio oficial de documentación para un proyecto de Analytics. Pueden hacerlo tan sencillo como una entrada de blog con el tablero embebido o crear un micro-sitio propio (también pueden crear un wiki o repositorio en Github, las formas de presentar documentación en sitios web es diversa, busquen una que les parezca cómoda). La idea es que exploren autónomamente las maneras de hacer publicaciones web y voy a premiar aquellos que decidan invertir tiempo y esfuerzo en hacer una presentación de algo que normalmente es de uso interno a un equipo, en un espacio web. En las empresas con equipos grandes, es una muy buena práctica hacer micro-sitios o blogs de documentación para los recursos internos de la empresa ya que son fáciles de diseminar y mantener actualizados en un solo lugar.

Rúbrica

Dimensión	Peso	Deficiente	Bien	Muy bien	Excelente
Dashboard: módulos 1-3	30%	No entrega	Falta algún módulo, o, la estrategia de visualización no es muy clara para cumplir con el objetivo.	Módulos completos, cumple con cada uno de los objetivos. La estrategia de visualización podría mejorar.	Cumple claramente con los objetivos y las estrategias de visualización sintetizan con claridad los objetivos.
Dashboard: módulo 4	10%	No entrega	No es del todo claro lo que busca mostrar con esta visualización.	Se entiende el objetivo de la visualización, la estrategia de visualización o presentación podría mejorar.	La propuesta es clara y revela información implícita en los datos. La estrategia de visualización es clara y efectiva.
Documentación: Descripción de variables tabla original	10%	No entrega	Hay errores en la tipología de las variables o las descripciones no son claras.	Bien tipificadas las variables, las descripciones podrían ser más claras.	Bien tipificadas las variables, las descripciones son efectivas y claras.
Documentación: Modelado	25%	No entrega	No están todos los campos nuevos o hay errores en la conversión de variables. Hay errores en la tipología de las variables o las descripciones no son claras.	Bien tipificadas las variables, las descripciones podrían ser más claras.	Bien tipificadas las variables, las descripciones son efectivas y claras.
Documentación: paso a paso	25%	No entrega	No es muy claro cómo replicar el modelado sobre datos actualizados.	La descripción es clara, pero faltan pasos o la redacción es confusa.	Es perfectamente claro cómo repetir el modelado sobre datos actualizados.

Entregables

Si realiza su entrega a través de enlaces, escríbalos en el campo a continuación. De lo contrario, utilice el espacio que está más adelante para adjuntar los archivos que dan cuenta de la visualización y la documentación. La plataforma recibirá un único archivo, por lo que es necesario que los comprima.

Dashboard 2: visualización y documentación para narrar a un equipo de trabajo

Desarrollado por: Juan Pablo Mogollón Avaunza

Variables en base de datos original

1. timestamp: Marca de tiempo en la que se realizó la encuesta (fecha y hora).
2. How old are you?: Rango de edad del encuestado (texto categórico).
3. What industry do you work in?: Industria en la que trabaja el encuestado (texto categórico).
4. Job title: Título del cargo del encuestado (texto categórico).
5. If your job title needs additional context, please clarify here:: Descripción breve de las responsabilidades del cargo (texto).
6. 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.): Salario anual del encuestado en la moneda declarada (número).
7. 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.): Compensaciones extra anuales (bonos acciones etc.) (número).
8. Please indicate the currency: Moneda en la que se reporta el salario (texto categórico).
9. If "Other," please indicate the currency here:: Moneda adicional en caso de tener ingresos en otra divisa (texto categórico).
10. If your income needs additional context, please provide it here: Información adicional proporcionada por el encuestado (texto).
11. What country do you work in?: País de residencia del encuestado (texto categórico).
12. If you're in the U.S., what state do you work in?: Estado en EE.UU. si aplica (texto categórico).
13. What city do you work in?: Ciudad en la que trabaja el encuestado (texto categórico).
14. How many years of professional work experience do you have overall?: Años de experiencia laboral total (número).
15. How many years of professional work experience do you have in your field?: Años de experiencia en la posición actual o similar (número).
16. What is your highest level of education completed?: Nivel educativo alcanzado (texto categórico).
17. What is your gender?: Identificación de género del encuestado (texto categórico).
18. What is your race?: (Choose all that apply.): Autoidentificación étnica del encuestado (texto categórico).

Variables luego de transformar y modelar datos

1. momento de captura: Marca de tiempo en la que se realizó la encuesta (fecha y hora).
2. rango de edad: Grupo etario del encuestado (texto categórico).
3. Industria_es: Industria en la que trabaja el encuestado en español (texto categórico).
4. experiencia total: Rango de años de experiencia laboral total (texto categórico).
5. experiencia específica: Rango de años de experiencia en el área específica del trabajo (texto categórico).
6. nivel educativo: Nivel de educación alcanzado por el encuestado en español (texto categórico).
7. género: Género con el que se identifica el encuestado (texto categórico).
8. país_corregido: Nombre del país corregido después de la limpieza de datos (texto categórico).
9. ciudad_corregido: Nombre de la ciudad corregido después de la limpieza de datos (texto categórico).
10. salario_anual COP: Salario anual del encuestado convertido a pesos colombianos (número decimal).
11. compensaciones COP: Ingresos adicionales al salario base, convertidos a pesos colombianos (número decimal).
12. ingreso anual total COP: Suma del salario anual y las compensaciones adicionales en pesos colombianos (número decimal).

Transformación de datos

Renombrar columnas en español

Se cambio el nombre de las columnas a un nombre más corto y comprensible en español, según el mapeo siguiente:

titulo original	nuevo título
Timestamp	momento de captura
How old are you?	rango de edad
What industry do you work in?	industria
Job title	posicion
If your job title needs additional context, please clarify here:	descripcion de la posición
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.)	salario anual
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.	compensaciones extra anuales
Please indicate the currency	moneda
If "Other," please indicate the currency here:	otra moneda
If your income needs additional context, please provide it here:	contexto adicional
What country do you work in?	pais
If you're in the U.S., what state do you work in?	estado us
What city do you work in?	ciudad
How many years of professional work experience do you have overall?	experiencia total
How many years of professional work experience do you have in your field?	experiencia especifica
What is your highest level of education completed?	nivel educativo
What is your gender?	genero
What is your race? (Choose all that apply.)	etnia

Verificar formato de fecha

Verificar que la columna “momento de captura” contenga datos en formato dd/mm/aaaa hh:mm:ss y tratar los datos que no correspondan con este formato.

Mapear géneros a español y agruparlos

Mapear los géneros así:

- "Woman": "Mujeres",
- "Man": "Hombres",
- "Non-binary": "Otros",
- "Other or prefer not to answer": "No informado",
- "Prefer not to answer": "No informado"

Mapear los niveles educativos a español

Realizar el mapeo de los niveles educativos así:

- "Master's degree": "Maestría",
- "College degree": "Título universitario",
- "PhD": "Doctorado",
- "Some college": "Estudios universitarios incompletos",
- "High School": "Bachillerato",
- "Professional degree (MD, JD, etc.)": "Título profesional"

Borrar espacios

Eliminar los espacios al inicio y final de las cadenas de texto de las columnas: 'ciudad', 'pais', 'otra moneda', 'industria'.

Mapear los códigos de las monedas

Crear la columna 'moneda corregida' y utilizando la información del archivo `codigosMoneda.csv` mapear las monedas, primero las monedas en la columna `moneda`, y luego las incluidas en la columna `otra moneda`. El valor mapeado se debe cargar en la columna 'moneda corregida'.

Mapear las tasas de cambio

Crear la columna `COP` y rellenar con el valor correspondiente a la tasa de cambio de cada moneda en pesos colombianos. El archivo `TasasDeCambio.csv` contiene la equivalencia de cada moneda en pesos colombianos, el valor del cambio se obtuvo utilizando la función `monedas` de Excel el día 9 de febrero de 2025.

Crear las columnas 'salario_anual COP', 'compensaciones COP', 'ingreso anual total COP'

- Crear la columna 'salario_anual COP', igual al producto de la columna 'COP' y la columna 'salario anual',
- Crear la columna 'compensaciones COP', igual al producto de la columna 'COP' y la columna 'compensaciones extra anuales',
- Crear la columna 'ingreso anual total COP', igual a la suma de la columna "salario_anual COP" y la columna 'compensaciones COP'.

Mapear los nombres de las industrias a español

Utilizando el archivo `mapearIndustrias.csv` convertir los nombres de las industrias a español y almacenarlos en la columna `industrias_es`.

Mapear los países y las ciudades

Utilizando el archivo `mapearPaisCiudad.csv` homologar los nombres de países y ciudades. La clave del archivo `mapearPaisCiudad.csv` corresponde al resultado de concatenar las columnas 'pais', 'estado us' y 'ciudad' separadas por '|'.

Seleccionar las columnas de interés

Seleccionar las columnas 'momento de captura', 'rango de edad', 'industria_es', 'experiencia total', 'experiencia especifica', 'nivel educativo', 'genero', 'pais_corregido', 'ciudad_corregido', 'salario_anual COP', 'compensaciones COP', 'ingreso anual total COP' y crear un archivo csv para usarlo como fuente del tablero.

Tablero

El tablero se encuentra disponible en: <https://lookerstudio.google.com/reporting/063a1df1-c242-465c-b9f7-55dfe7fe2508>

Documentación

La documentación está disponible en <https://github.com/ZabakJL/ManageSalarySurvey2021.git>