**Cálculos de Tablas, Dashboards Avanzados, Storytelling**

## **Mapping: Cómo Establecer Roles Geográficos. Intro a Segmentación demográfica**

**Algo importante.** Considere la siguiente definición de **Segmentación demográfica**: La segmentación demográfica consiste en dividir el mercado en grupos más pequeños, este tipo de segmentación toma en cuenta variables como la edad, ingresos, grado de estudios, nacionalidad, raza, religión, ocupación, etc. (QuestionPro). A partir del contenido que hay en el *dataset* ‘P1-UK-Bank-Customers’ situado en este directorio, que contiene datos simulados de clientes de un Banco del Reino Unido, nos interesa proceder con nuestra segmentación inicial: localizar las cuatro regiones en las que se sitúan los clientes en el Reino Unido.

### **Rastrear mis puntos geográficos marginados**

Nos interesa saber en principio cómo darles un reconocimiento a los valores *desconocidos* parcialmente por Tableau al ser arrastrados a una Worksheet; para la ocasión, el ejercicio tomará como valores desconocidos valores de tipo *geográfico*. Manos a la obra. Observe que cuando arrastra el campo ***Region*** de tipo *Geográfico* con formato ‘*CC. AA./Estado/Provincia/Dpto.*’ a su *Worksheet* de Tableau, no se reconoce a las regiones propiamente de **United Kingdom** dentro del mapa(no se ven, no se *marcan*). Estas regiones son: **England, Wales, Scotland & Northern Ireland.** Las regiones se supone que debieran verse, pues, justamente arrastramos a la *Worksheet* el campo que las contiene: ***Region.* ¿Qué ha pasado?**

Generalmente cuando sucede esto se debe a que realmente Tableau no está reconociendo a los valores, no dentro de la visualización propiamente ya configurada: la razón más común para este *bug* podría deberse a que se haya generado algún tipo de inconsistencia en los datos, o valores, cuando fueron arrastrados a la Worksheet.

En Tableau, los valores *desconocidos* se ven así, por ejemplo: 

En esta ocasión como le interesa trabajar con *regiones*, y no las ve dentro de la Worksheet, es probable que sus 4 regiones no estén sido reconocidas y, de hecho, son 4 los valores desconocidos por la visualización: **England, Wales, Scotland & Northern Ireland.**

#### **Tratamiento de valores desconocidos**

Veamos. Hagamos clic en . Se nos abrirá una ventana emergente, observe.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

La opción que seleccionaremos es ***Editar ubicaciones…*** ; pues, en efecto, nos interesa corregir las ubicaciones desconocidas (estamos trabajando con datos geográficos que están siendo desconocidos actualmente). Observe la siguiente ventana emergente.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Note que las *regiones* no son coincidentes con la ubicación (Colombia), lo cual tiene sentido porque estas regiones no se ubican dentro del país Colombiano (como por defecto lo intuye Tableau), sino dentro del Reino Unido (UK). **Resolución**: Cambiemos entonces a ‘Colombia’ por ‘Reino Unido’.

Verá lo siguiente:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Hace clic en ‘Aceptar’ y listo!, sus datos geográficos ya han sido reconocidos dentro del mapa: las 4 divisiones de UK. Observe.

Primeramente, debe seleccionar el siguiente diseño de mapa: 

Mapa

Descripción generada automáticamente

Vamos a distinguir visualmente a cada región por medio de colores, quedando así:

Mapa

Descripción generada automáticamente

Y listo!, así es como hemos establecido roles geográficos, por medio del mapeo (Mapping) de datos.

### **Diferencias entre Recuento (Distintos) & Recuento**

Ahora, la idea es responder a la consulta que requiere saber el total de número de clientes registrados en el Banco del Reino Unido para cada región; es decir, hay que *agregar* datos, específicamente correr una función de agregación de tipo **Conteo**.

Aclaremos algo importante. En el *dataset* ‘P1-UK-Bank-Customers’ cada registro pertenece a un cliente diferente, es decir, no hay cliente duplicado en ningún registro de esta fuente de datos. Entonces, dicho eso, no sería explícitamente necesario hacer uso de una función de reconteo que cuente sólo valores *distintos* según el parámetro o campo pasado (en nuestro caso sería el campo **Customer ID**); pues, no hay clientes duplicados para ningún registro; es decir, no es necesaria aplicar la medida: Recuento (Distintos)de Tableau. En esta ocasión particular, no habría ninguna diferencia entre usar **Recuento (Distintos) & Recuento**.

Arrastre su columna **Customer ID** a su marca **Etiqueta** (usamos esta marca porque nos interesa ver explícitamente los números agregados correspondientes sobre cada región de UK) y aplique una medida de tipo **Recuento o Recuento (Distintos)** sobre ella. Como la visualización ya se encuentra categorizada o granulada por Regiones, por las 4 regiones en las que opera el Banco del Reino Unido: **England, Wales, Scotland & Northern Ireland**, usted entonces verá el número de clientes que habría para este banco por cada Región.

## **Creando Cálculos de Tablas por Género**

### **Campo calculado: Conteo de nº de registros**

A veces es necesario tener un campo (columna) dentro de su tabla de estudio que contabilice el numero total de registros de la tabla en cuestión, ya sea que desee un conteo de los registros duplicados o no duplicados (esto depende de lo que pretenda responder la empresa). En Tableau, este campo tiene que ser calculado, pues, por defecto no se genera.

Por ejemplo, ya sabemos que los registros del *dataset* ‘P1-UK-Bank-Customers’ no se repiten; sin embargo, nos interesa contar la totalidad de los ID, de los clientes, para saber así el número total de ellos que estén vinculados al Banco de UK. El interés final consiste en representar un diagrama de sectores, popularmente conocido como ‘Gráfico de Torta’ o ‘Gráfico Circular’, que me diga en proporciones equivalentes qué parte de la torta pertenece a los clientes de género masculino y a los clientes de género femenino; es decir, nos interesa representar en el gráfico de torta el número total (conteo) de clientes y el número total (conteo) de clientes mujeres.

Entonces, lo primero, crear nuestro campo calculado que corresponda al número total de registros del dataset ‘P1-UK-Bank-Customers’: *Total Records*. Sin embargo, como nos interesa categorizar los datos por *Género,* para el diagrama circular, debemos previamente arrastrar el campo ‘Gender’ (incluso antes de crear el campo calculado). Veamos:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente con confianza media

Ahora sí, contar todos los registros de ‘P1-UK-Bank-Customers’ por medio de un campo calculado (recuerde, esta vez no es necesario hacer uso de ‘Distintos’). Observe.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Ahora que ya tenemos el conteo de todos los registros del *dataset* ‘P1-UK-Bank-Customers’, nos interesa saber qué tanto de ese conteo corresponde a clientes femeninos y a clientes masculinos. Para esto, debe arrastrar su campo calculado ‘*Total Records*’ al campo previamente arrastrado *‘Gender’* (Género),¿esto para qué? Para que, en efecto, se cuente el total numérico de clientes mujeres y el total numérico de clientes hombres. Veamos cómo queda.

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

### **Diagrama de sectores o gráfico de Torta**

Ahora, ¿ya lo notaste? No estamos ante un diagrama circular, que es lo que necesitamos, sino ante una tabla. La tabla fue creada automáticamente al arrastrar nuestro campo *Gender* a la Worksheet; sin embargo, podemos modificar manualmente esta representación visual de los datos.

Seleccione ‘Circular’.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Verá lo siguiente:

Gráfico, Gráfico de burbujas

Descripción generada automáticamente

Los valores para la categoría *Gender* son: **Female & Male**. Entonces, se verá un gráfico circular por cada valor o nivel de su categoría *Género*, pues, anteriormente habíamos definido el campo *Gender* en función de Filas: . Cada valor o nivel de su categoría asumiría una fila. En consecuencia, cada gráfico circular correspondería a un valor o nivel de la categoría que esté evaluando, *Gender*, y cada valor o nivel se ubicaría sobre una fila.

Para suprimir el total de gráficos circulares que podría haber para cada uno de sus géneros, debe quitar *Gender* de Filas y colocar dicho campo, *Gender*,en la marca *Colores;* así pues, usted vería un solo gráfico circular o de *torta*, pero distinguiría visualmente cada uno de los valores de la categoría evaluada por medio de colores particulares asignados para cada valor o nivel de *Gender* (justamente nuestra categoría evaluada). Veamos:

Gráfico, Gráfico circular

Descripción generada automáticamente

#### **Proporcionalidad & Marca Ángulo**

Los valores visualizados en la *torta* sabe bien que corresponden al campo *Total Records* y que ya se encuentran categorizados por el campo *Gender*; es decir, el total de *Total Records* se divide entre **Female & Male.** Sin embargo, no le hemos dicho a Tableau que queremos que se respeten también, de manera visual, las proporciones correspondidas; es decir, que se respete la proporción en la *torta* de lo que le corresponde a cada grupo, sea el grupo de clientes masculinos o el grupo de clientes femeninos.

Para decirle a Tableau que queremos que *Total Records* guarde proporcionalidad correspondida en el gráfico circular, debemos arrastrar de nuevo el campo *Total Records* ahora a la marca “Angulo”. Con esto se suprime el *sesgo* o manipulación de los datos sobre el gráfico circular evaluado. Su diagrama de sectores se vería así:

Interfaz de usuario gráfica, Texto, Aplicación, Teams

Descripción generada automáticamenteGráfico, Gráfico circular

Descripción generada automáticamente

Esta es la forma manual de generar nuestro gráfico circular; sin embargo, hay una forma automática, … sólo sería necesario hacer clic en:  para generar nuestro tipo de gráfico ideal ya diseñado.

#### **Valores expresados en porcentajes**

Si bien tenemos representado el número total de empleados masculinos (o femeninos), en nuestro gráfico circular, ahora nos interesaría saber qué porcentaje (%) de la porción de empleados masculinos (o femeninos) se asume o se corresponde sobre la totalidad de los empleados en sí, sin distinguir género sexual (100%); es decir, cuál porcentaje asume el valor *absoluto* 2.165 (2.165) sobre el total de los empleados (2.165 + 1.849). De hecho, lo que se busca generalmente en un gráfico circular es representar proporciones con respecto a un total, es decir, representar **porcentajes**.

#### **Cálculo de Tablas Rápido**

Pasar de valores absolutos, visualizables textualmente a causa de la marca *Etiqueta*, a valores porcentuales (*relativos*), en un diagrama de sectores, se logra de la siguiente manera: Debe hacer clic derecho sobre el objeto que constituye los valores absolutos que desea ver, ahora, en términos porcentuales; para nuestro caso puntual, este objeto sería: , pues es quien se encarga propiamente de las *etiquetas* que hay alrededor de nuestro gráfico circular (**2.165 + 1.849**). Luego, seleccionamos la opción ‘Cálculo de tablas rápido’ y, posteriormente, la opción ‘Porcentaje del total’. Veamos.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Una vez hecho clic sobre ‘Porcentaje del total’, verá las proporciones de cada uno de sus valores o niveles (Male, Female) sobre el total de la categoría evaluada (Gender) en términos porcentuales (%). Observe.

Gráfico, Gráfico circular

Descripción generada automáticamente

**Dato de interés:** No se recomienda hacer uso de Diagramas de sectores cuando se evalúan 3 o más de 3 grupos, divisiones (porciones) en el gráfico.

**Conclusiones claves:** Hasta ahora, por medio de la segmentación demográfica, hemos categorizado la totalidad de los datos tanto por región como por género sexual. En el siguiente apartado continuaremos con el análisis de la *segmentación demográfica* de los clientes registrados en el *dataset* ‘P1-UK-Bank-Customers’.

## **Agrupaciones (Bins) & Distribuciones por Edad**

#### **Uso de Gráfico de Barras**

Nos interesa continuar con nuestra segmentación demográfica ahora categorizando a nuestros clientes por *Edades*; es decir, nos interesa saber cómo se distribuyen los clientes en los diferentes intervalos o franjas de edad. **Como se espera que sean más de 2 franjas de edad las evaluadas, incluso más que 5, lo ideal es representar estos datos por medio de un gráfico de barras** (especialmente se usa gráficos de barras en los casos en los que pretenda conocer la cantidad numérica de una variable A, una variable tipo *medida*, sobre cada grupo o franja de una variable categórica, una variable tipo *dimensión*).

Tenemos que la variable *medida* correspondería al campo calculado que cuenta el número total de registros (*Total Records*); el cual tiene que *desagregarse* para poder conocer la cantidad exacta de ellos, de los *registros* (que es lo mismo que decir *personas*), sobre cada franja de edades evaluadas (*Age*): nuestra variable *dimensión*.

Es probable que usted encuentre que Tableau automáticamente *agregue* el campo *Total Records* al hacer un conteo total de la cantidad de registros que hay sobre el *dataset* ‘P1-UK-Bank-Customers’, de hecho es eso lo que justamente hace dicho campo calculado, y no es precisamente eso lo que usted desea. Cada registro corresponde a una persona; por lo cual, su interés sería entonces saber la cantidad de registros (personas) que hay para cada franja de edades evaluada (toca *desagregar* a *Total Records*). Vamos allá.

Como queremos hacer un gráfico de barras, nos toca pasar nuestro campo *medida* (*Total Records*) a una columna (o fila) y lo propio con nuestro campo *dimensión* (*Age*). **Definamos *Total Records* en Filas & *Age* en Columnas** (todo esto en una nueva Worksheet)**.** Tal que así:

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

#### **El paso de Medida a Dimensión**

Si se percata, Tableau está interpretando la variable *Age* como una variable de tipo *medida* (esto por conservar *números* como tipo de dato). Sin embargo, son números correspondientes a las edades, no es que uno necesite hacer operaciones matemáticas con ellos; es decir, *Age* realmente es una variable de tipo *dimensión* puesto que describe una característica (edad) de alguien. Vamos a convertir a *Age* en una variable *dimensión*.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Listo! Vamos a ver nuestra visualización ahora.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Lo tenemos! Sin embargo, nos interesa que la visualización sea un gráfico de barras (no de líneas). Veamos.

Gráfico, Histograma

Descripción generada automáticamente

Algunos ***insighsts***: En la edad de 34 están 185, es decir, 185 personas tienen la edad de 34 años.

Ahora, si te percatas, el tener tantas *barras* en un diagrama de barras como éste, puede ser agotador. Este diagrama de barras puede dificultar la tarea de poder leer e interpretar lo que la visualización pretende enseñar, y tampoco deseamos ese desenlace… He aquí donde surge la necesidad de crear *Agrupaciones (Bins)*; es decir, para nuestro caso puntual, permitir dividir nuestras edades en diferentes intervalos. Crear *Bins* es crear Intervalos sobre datos ya existentes, es justamente eso lo que haremos con nuestras edades.

#### **Creando Agrupaciones o Bins**

Nos interesa crear intervalos en las edades de 5 en 5; es decir, que se evalúe el numero de personas con edades que van desde los 5 años hasta los 9, luego desde los 10 años hasta los 14, luego desde los 15 hasta los 19, y así sucesivamente. Si se da cuenta, **los intervalos siempre son cerrados a izquierda y abiertos a derecha.** Veamos cómo podemos hacerlo.

**Importante:** Los *Bins* se crean por medio de variables tipo *medidas* o, al menos, que así las interprete inicialmente Tableau (no importa si luego las convierte en *dimensiones* como es el caso de *Age*). Ahora, a las *Bins* propiamente sí se les considera categorías o *dimensiones* (luego de creadas). Es decir, de lo que se trata es de cómo se convierte una variable numérica ya existente a una nueva variable categórica, sin suprimir ninguna de las dos, a través de Agrupaciones o Bins. **Sí, al crear un *Bin* se crea un nuevo campo de tipo categórico o *dimensión* que represente esa agrupación que acaba de configurar.** Veamos.

Ya habíamos planteado anteriormente que queríamos crear los *intervalos* sobre las edades. Vaya a su variable *Age,* haga clic derecho sobre ella, luego hace clic en la opción ‘Crear’ y, posteriormente, hace clic en la opción ‘Agrupaciones… ’. Observe el procedimiento.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Verá la siguiente ventana emergente:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Tableau por defecto le sugiere un tamaño de agrupación *ideal*, sin embargo usted puede reconfigurar esto. En nuestro caso, por ejemplo, deseamos que los intervalos en las edades vayan de 5 en 5. Dicho lo anterior, definamos entonces el ***tamaño de agrupaciones*** como: **5**. Observe.

, hacemos clic en ***Aceptar*** y se creará un nuevo campo de tipo *dimensión* que sería el equivalente a la Agrupación del campo *Age.* Observe. 

El siguiente paso es sencillo: Suprima su actual campo *Age* de las columnas y arrastre ahí el nuevo campo generado *Age (agrupación)* para poder ver una visualización donde las edades se agrupen en intervalos de 5 en 5.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Listo! Vamos a ver nuestra visualización ahora.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ahora, queremos saber los valores porcentuales (*relativos*) que hay sobre cada barra; es decir, qué porcentaje ocupa cada intervalo de edades de los clientes sobre el total de los clientes. Esto ya lo sabe hacer y se hace por medio de **cálculo de Tablas Rápido.** La escala de *Total Records* también cambiará: de valores absolutos a valores relativos (%). Veamos.

Gráfico, Gráfico de barras

Descripción generada automáticamente