



RETO GRUPAL

ESTADOS DE ESTADOS UNIDOS

ESCUELA DE TALENTO | FUNDACIÓN NTTDATA



ÁREA IV: VISUALIZACIÓN DE DATOS



METODOLOGÍA

En este documento presentamos nuestro trabajo correspondiente a la resolución de los ejercicios del Área IV - Reto Grupal “Estados de Estados Unidos”.

Para la realización de las tareas descritas en el reto hemos trabajado bajo la modalidad de pair programming, como ya ocurrió en una parte del Área II. Jesé ha compartido compartiendo su pantalla y ha aplicado las acciones que se deciden en conjunto. Cabe destacar que este reto lo hemos trabajado en una sesión

El proceso ha sido documentado por Luisa y Ana, quien se ha encargado de la edición de texto e imagen, y ha sido revisado y aprobado por el resto de los componentes del equipo.

Gracias a nuestro profesor Miguel Palomares por las clases y aclaraciones en los primeros momentos del reto.

PARTICIPANTES



- JÉSSICA RÍOS
- LUISA ROMERO
- ANA DELGADO
- JAIME SHIMOHIRA
- JESÉ MUÑOZ
- SERGIO BULBARELA POPO



ÍNDICE

EJERCICIO 1: Power Query.

Carga de csv, análisis exploratorio y preparación de datos.

EJERCICIO 2: Modelo de datos.

Uso de SWITCH.

EJERCICIO 3: Diseño del cuadro de mandos.

Segmentación de variables, creación de tablas, tarjetas KPI's y gráficos.

EJERCICIO 4: Mostrar visualizaciones.

Creación de mapa.

EJERCICIO 5: Creación de una página índice con botones.

EJERCICIO 1

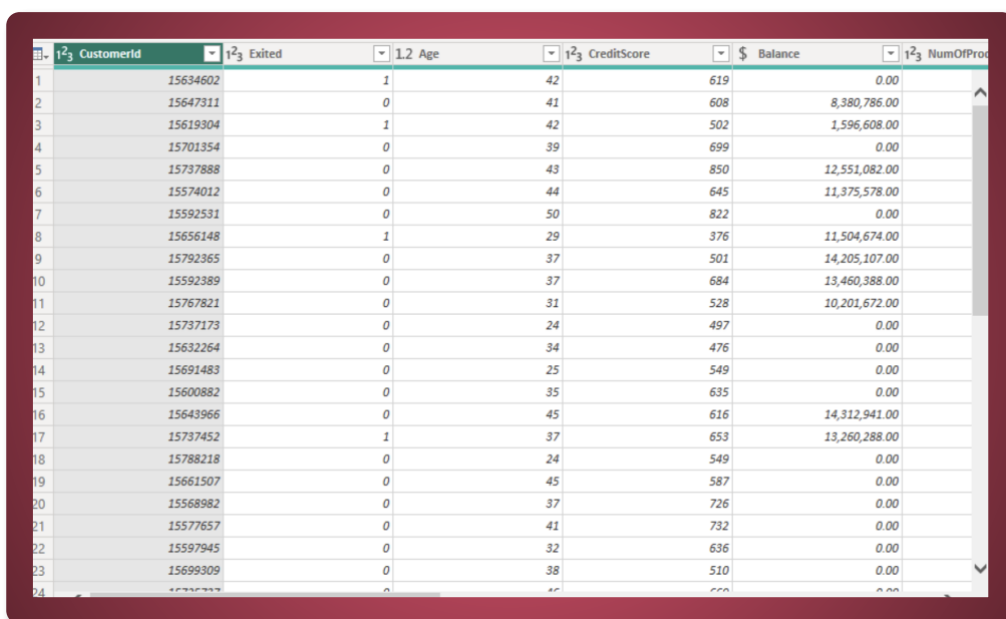
De acuerdo el enunciado del ejercicio, se nos pide primero cargar el csv, por lo que lo descargamos y Jesé lo carga en Power Bi, compartiendo pantalla como indicamos en la metodología. El ejercicio sugiere hacer primero un análisis exploratorio, por lo que analizamos el csv, que a simple vista es muy similar al que teníamos para el reto anterior, pero con algunas pequeñas modificaciones de importancia.

Ana abre Jupyter con el notebook del reto anterior, para comparar ambos csvs y poder seguir una línea conductora en este último trabajo. Las columnas son casi idénticas, por lo que decidimos que el análisis anterior nos puede ayudar más adelante, cuando tengamos que preparar los datos.

Dado que se nos pide eliminar las columnas que no serán vinculantes, y revisar que los tipos de datos de las columnas sean correctos, ya que Ana tiene abierto Jupyter nos va indicando que columnas podrían ser eliminadas y valorar si podrían aplicarse en este caso. De acuerdo a esto, propone eliminar las columnas "Rownumber", una columna sin nombre con los índices y la columna "Unnamed", las tres representan el índice y todas tienen el mismo dato, que no nos sirve para este análisis.

También eliminamos, la columna "Surname", que corresponde al apellido y que no influye en el objetivo de estos retos que es valorar qué influye en la permanencia o abandono de la entidad bancaria, por parte del cliente.

Por unanimidad eliminamos las columnas mencionadas anteriormente y procedemos con la revisión del tipo de dato y los datos contenidos en el resto de columnas. Verificamos que los tipos de datos están bien, exceptuando la columna "Age", dónde Luisa es la primera en darse cuenta de que, si bien el tipo de dato es correcto, las edades están multiplicadas por 10, por lo que hay que reajustarlas. La solución que Ana y Jesé proponen es la de dividir los datos de "Age" por 10, aunque desconocían como aplicar esos cálculos en Power Bi, Jaime nos da indicaciones de cómo hacer este cálculo a través de la pestaña "Transformar", con la columna seleccionada y a través botón "Estándar". Seleccionamos "Dividir", se introducen los valores que queremos y aplicamos.



	i ₂ CustomerId	i ₂ Exited	1.2 Age	i ₂ CreditScore	\$ Balance	i ₂ NumOfProducts
1	15634602	1	42	619	0.00	
2	15647311	0	41	608	8,380,786.00	
3	15619304	1	42	502	1,596,608.00	
4	15701354	0	39	699	0.00	
5	15737888	0	43	850	12,551,082.00	
6	15574012	0	44	645	11,375,578.00	
7	15592531	0	50	822	0.00	
8	15656148	1	29	376	11,504,674.00	
9	15792385	0	37	501	14,205,107.00	
10	15592389	0	37	684	13,460,388.00	
11	15767821	0	31	528	10,201,672.00	
12	15737173	0	24	497	0.00	
13	15632264	0	34	476	0.00	
14	15691483	0	25	549	0.00	
15	15600882	0	35	635	0.00	
16	15643966	0	45	616	14,312,941.00	
17	15737452	1	37	653	13,260,288.00	
18	15788218	0	24	549	0.00	
19	15661507	0	45	587	0.00	
20	15568982	0	37	726	0.00	
21	15577657	0	41	732	0.00	
22	15597945	0	32	636	0.00	
23	15699309	0	38	510	0.00	
24	15699309	0	38	510	0.00	

EJERCICIO 2

Siguiendo con la línea del primer ejercicio se nos pide mostrar los datos con separadores de miles cuando sean numéricos. En este punto Jesé ha propuesto que las columnas “Balance” y “EstimatedSalary”, se les aplique este separador de miles, pero no en “Age”, que es la edad de la persona, ya que no tendría sentido.

Según las indicaciones de Ana vamos a la función “Tipo de datos” y luego a “Decimal fijo”, aunque hubo que realizar varios intentos e investigar ya que no sabíamos el sitio exacto donde realizar los cambios y tuvimos que ir probando.

Al aplicar esto, comprobamos que nos aparecen la separación, pero también se añaden dos decimales extra, así que aplicamos redondeo. Las cifras de las columnas seleccionadas. Pero con dos decimales extras, que bien pueden confundir, así es que redondeamos los números.

Exited	Age	CreditScore	Balance	NumOfProducts	EstimatedSalary	Tenure	Exited_predict	Gender	HasCrCard	IsActiveMember	Geography
0	41	506	\$ 0	2	\$ 100,608,435	1	0	Female	1	0	Texas
0	41	828	\$ 0	2	\$ 100,608,435	7	0	Female	1	0	Texas
0	40	512	\$ 0	2	\$ 100,608,435	5	0	Male	1	1	Texas
0	30	694	\$ 0	2	\$ 100,608,435	9	0	Female	1	1	Texas
0	80	652	\$ 0	2	\$ 100,608,435	4	0	Female	1	1	Texas
0	38	711	\$ 0	2	\$ 100,608,435	3	0	Male	1	0	Texas
0	35	753	\$ 0	2	\$ 100,608,435	3	0	Male	1	1	Texas
0	27	650	\$ 0	2	\$ 100,608,435	6	0	Female	1	0	Texas
0	57	664	\$ 0	2	\$ 100,608,435	6	0	Male	1	1	Texas
0	28	754	\$ 0	2	\$ 100,608,435	8	0	Male	1	1	Texas
0	42	468	\$ 0	2	\$ 100,608,435	5	0	Female	1	0	Texas
0	28	561	\$ 0	2	\$ 100,608,435	7	0	Male	1	0	Texas
0	25	521	\$ 0	2	\$ 100,608,435	7	0	Male	1	1	Texas
0	41	794	\$ 0	2	\$ 100,608,435	7	0	Female	1	1	Texas
0	33	735	\$ 0	2	\$ 100,608,435	4	0	Female	1	0	Texas
0	39	621	\$ 0	2	\$ 100,608,435	6	0	Male	1	1	Texas
0	45	703	\$ 0	2	\$ 100,608,435	7	0	Female	1	1	Texas
0	36	703	\$ 0	2	\$ 100,608,435	2	0	Male	1	0	Texas
0	35	656	\$ 0	2	\$ 100,608,435	6	0	Male	1	0	Texas
0	33	731	\$ 0	2	\$ 100,608,435	4	0	Male	1	1	Texas
0	29	786	\$ 0	2	\$ 100,608,435	4	0	Female	1	0	Texas
0	44	583	\$ 0	2	\$ 100,608,435	8	0	Female	1	1	Texas
0	47	707	\$ 0	2	\$ 100,608,435	3	0	Male	1	0	Texas
0	35	605	\$ 0	2	\$ 100,608,435	6	0	Male	1	1	Texas
0	40	652	\$ 0	2	\$ 100,608,435	1	0	Female	1	0	Texas

En el punto dos y tres se nos pide hacer una recategorización, primero creando una nueva columna y luego tomando los valores de “CreditScore” y luego recategorizar según rangos de numéricos. También se nos pide hacer lo mismo, pero con la variable “Edades”.

El enunciado pide hacerlo con la función SWITCH por lo que Jaime nos enseña cómo podemos realizar el proceso, localiza la formula necesaria y tanto él como Jesé lo adecuan para conseguir nuestro objetivo.

```
1 Grupos CreditScore = SWITCH(TRUE(), [CreditScore] >= 350 && [CreditScore] <= 450, "Grupo 1", [CreditScore] > 450 && [CreditScore] <= 550, "Grupo 2", [CreditScore] > 550 && [CreditScore] <= 650, "Grupo 3", [CreditScore] > 650 && [CreditScore] <= 800, "Grupo 4", [CreditScore] > 800, "Grupo 5")
```

Al ejecutar el código nos dimos cuenta de que hay valores que en el enunciado están en dos categorías, como, por ejemplo: Grupo 1: 350 a 450 y Grupo 2: 450 a 550 y no nos habíamos percatado hasta ese momento. Como se ve el número 450 está en dos categorías, por lo que realizamos modificaciones de los grupos, decidimos que de 450 pertenezca al Grupo 1 y desde 451 en adelante se considere Grupo 2, esto lo hicimos con todos los grupos que nos pidieron.

El mismo proceso sucedió con la variable edades, dónde Jéssica ha modificado el código que nos proporcionó Jaime y que adecuamos en el anuncio anterior y lo adecuó a la variable "Age". Ana ha propuesto que los grupos de las edades se llamen: "18 a 24" para el primer rango y así sucesivamente para el resto ya que era más visual y fácil de entender.

Finalmente comprobamos que las dos nuevas columnas tengan rangos en sus filas, es decir que todos los datos fueron recategorizados correctamente.

EJERCICIO 3

En este ejercicio comenzamos con la creación de la visualización de los datos. En primer lugar, debemos crear segmentadores para algunas variables, usando el método SWITCH. Para la creación de dichos segmentadores, usamos el método "Slicer".

Posteriormente pasamos a crear una tabla que muestre el grupo de crédito score de un lado, y del otro, el promedio de "Creditscore", Jesús lo hace con el método "Table". Una vez creada la tabla, pasamos a crear tarjetas de KPIs, para "Cantidad de customers", "Credit score promedio", "Tenure promedio" y "NumOfProducts promedio".

Aquí tratamos de llegar a un consenso, ya que consideramos que el enunciado no era suficientemente claro para saber con qué variable tendríamos que comparar las variables pedidas, Jaime propone que es el campo "CustomerId", que es el único identificador de las personas, fundamentando que cada vez que se filtra, lo único que puede diferenciar entre una y otra persona dicho dato. Luisa concuerda, ya que, en momento de generarse la duda con la interpretación del enunciado, estábamos en clase síncrona, solicitamos aclaración del enunciado a Miguel, quien nos confirma que, tal y como habíamos debatido, efectivamente debemos tener en cuenta ese campo.

Procedemos a aplicar "DISTINCTCOUNT" para hacer que "CustomerId" no repita y poder contar.

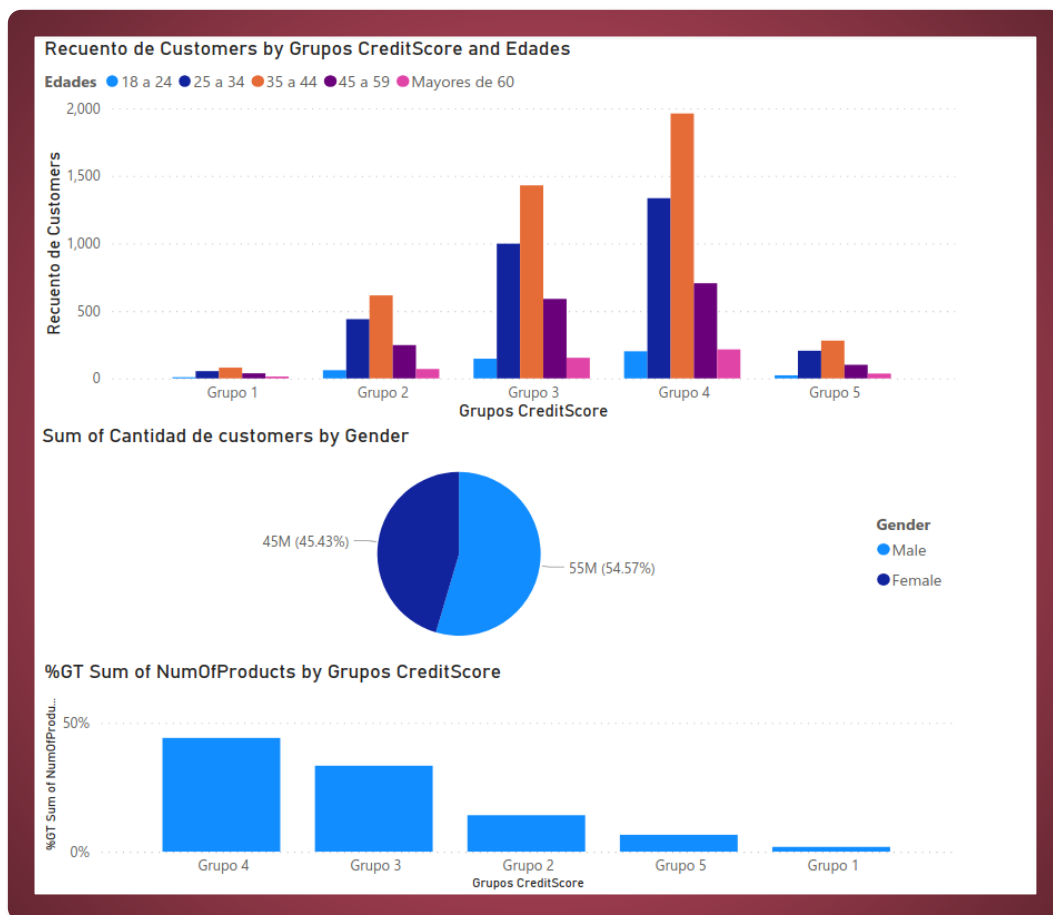
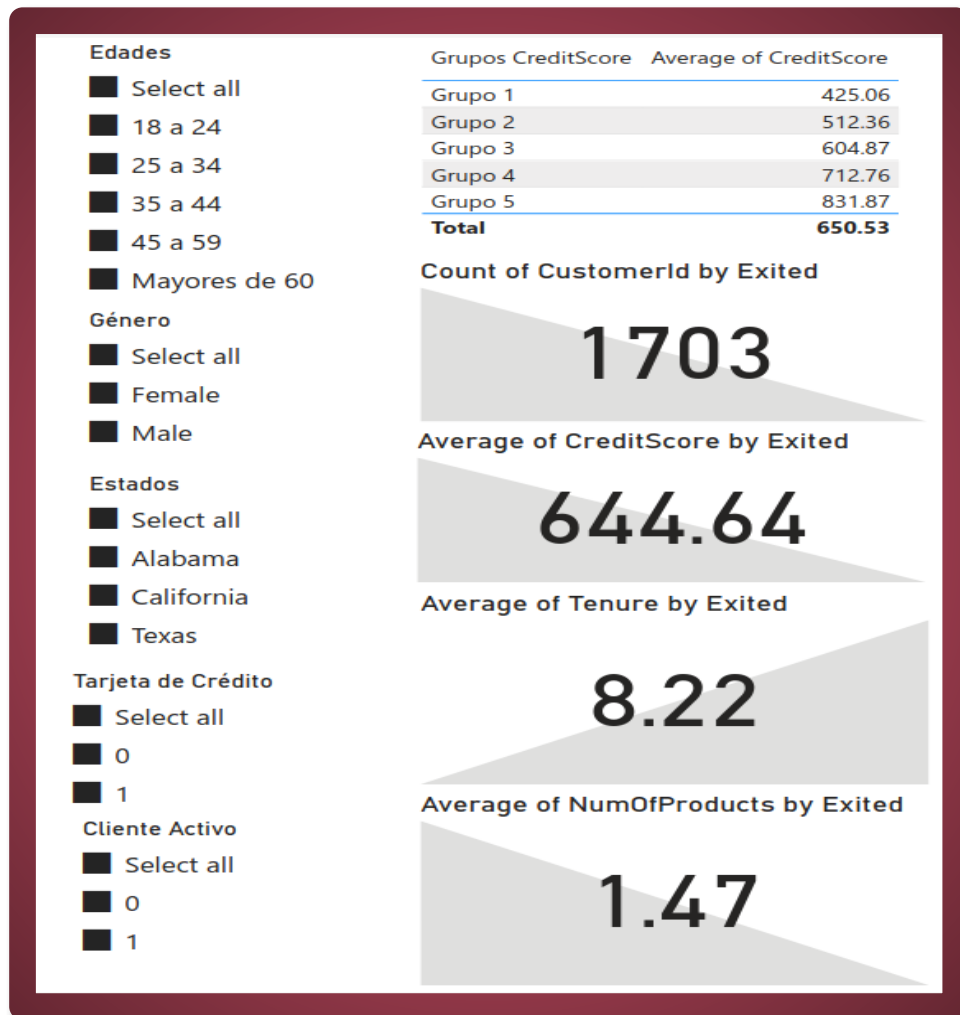
```
1 Cantidad de customers = DISTINCTCOUNT(df_banco_f[CustomerId])
```

Intentamos con la función crear una nueva columna contando las filas únicas con DAX, cómo se pide en el primer punto, pero en Power query nos dio error así que lo probamos en el fichero de Power Bi, creando una nueva columna usando la función y con el parámetro a calcular, en nuestro caso "CustomerId". Pudimos crear esta nueva columna y nos dio un total de 10000 clientes, y pudimos crear las siguientes Kpis con las distintas variables.

Pasamos al siguiente punto, que consiste en crear un gráfico de columnas agrupadas, que muestre la cantidad de clientes por grupo de "Credit Score" y "Age". Jesús lo pudo graficar con el método "clustered column chart".

Asimismo, creamos un gráfico de torta para mostrar la cantidad de clientes por género, que Jesús creó con el método "donut chart".

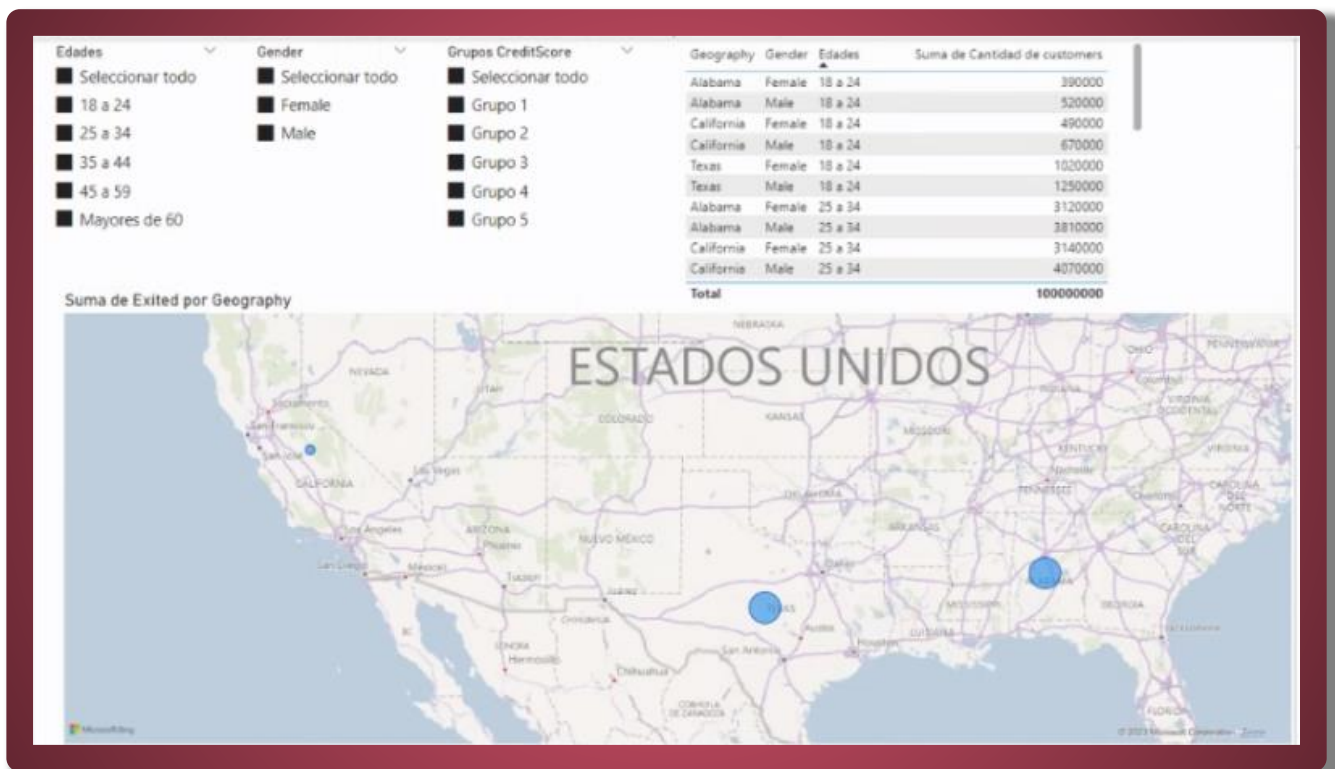
Por último, creamos un gráfico de columnas agrupadas que muestre el porcentaje de productos, que tiene cada grupo de credit score. El porcentaje debe ser sobre el total, dando entre los 5 grupos un 100%. Hicimos aquí un "clustered column chart" con las variables "NumberOfProducts" y los grupos "CreditScore".



EJERCICIO 4

En el ejercicio procedemos con la creación de una segunda página donde incluyamos un mapa con la ubicación de la variable “Geography” y que realizamos a través de “Mapa”.

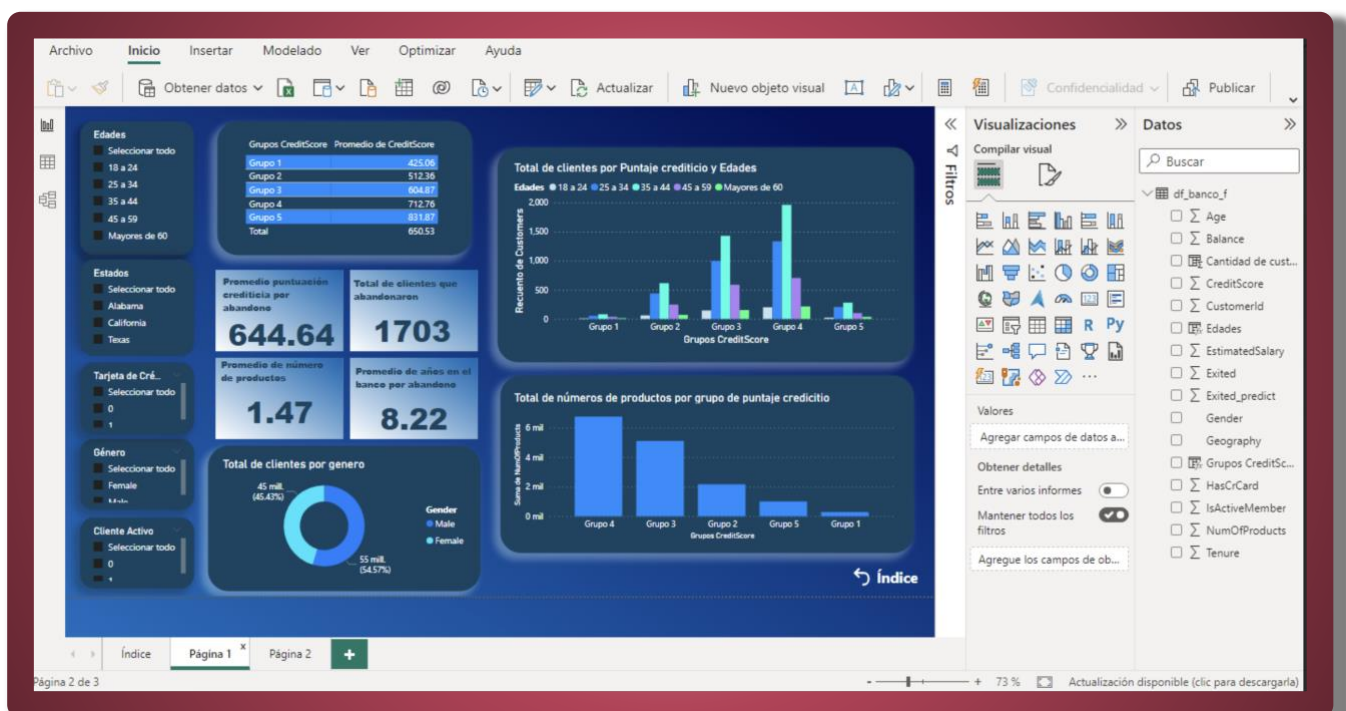
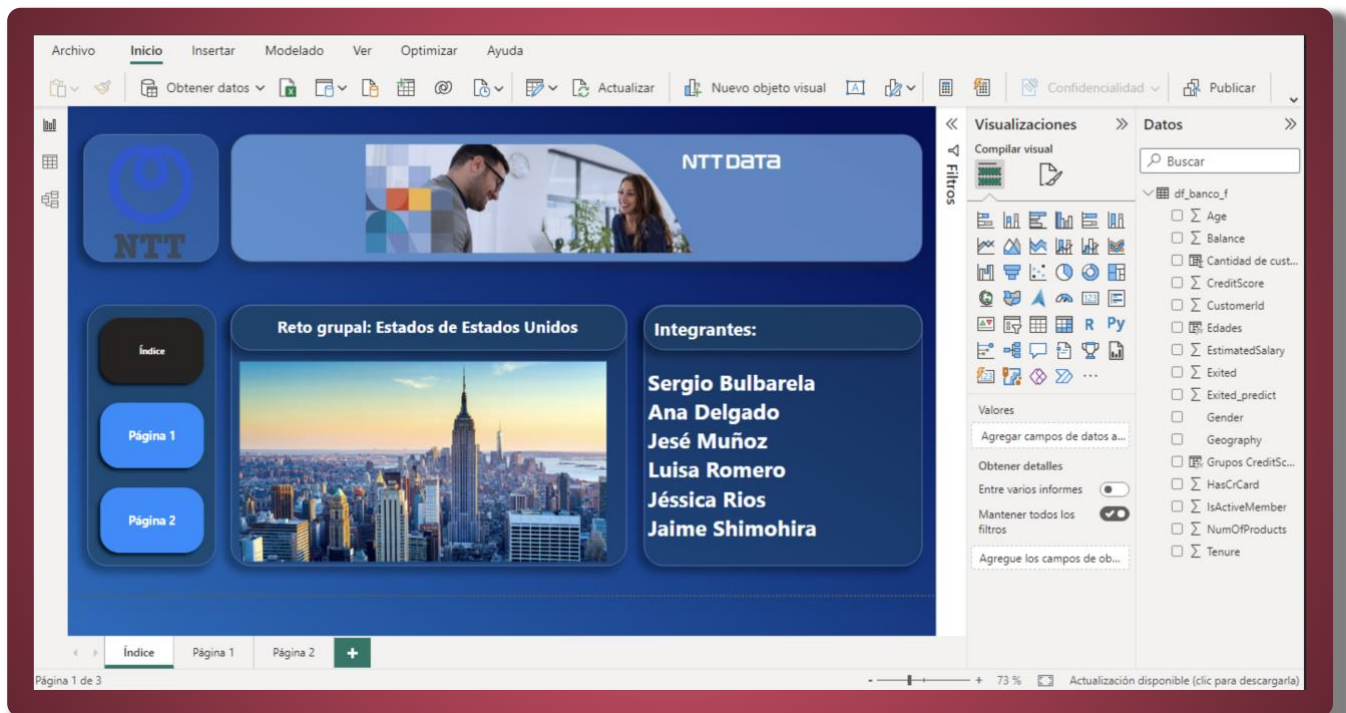
Asimismo, colocamos segmentadores correspondientes a la edad, “Age”, “Gender”, y al grupo “CreditScore”, aplicando “Filtro” por cada una de las variables, esta parte nos mostró Jesé como llevarla a cabo. También usamos de nuevo “Tabla” para la creación de una que contuviera “Geography”, “Gender”, “Age” y Cantidad de customers.

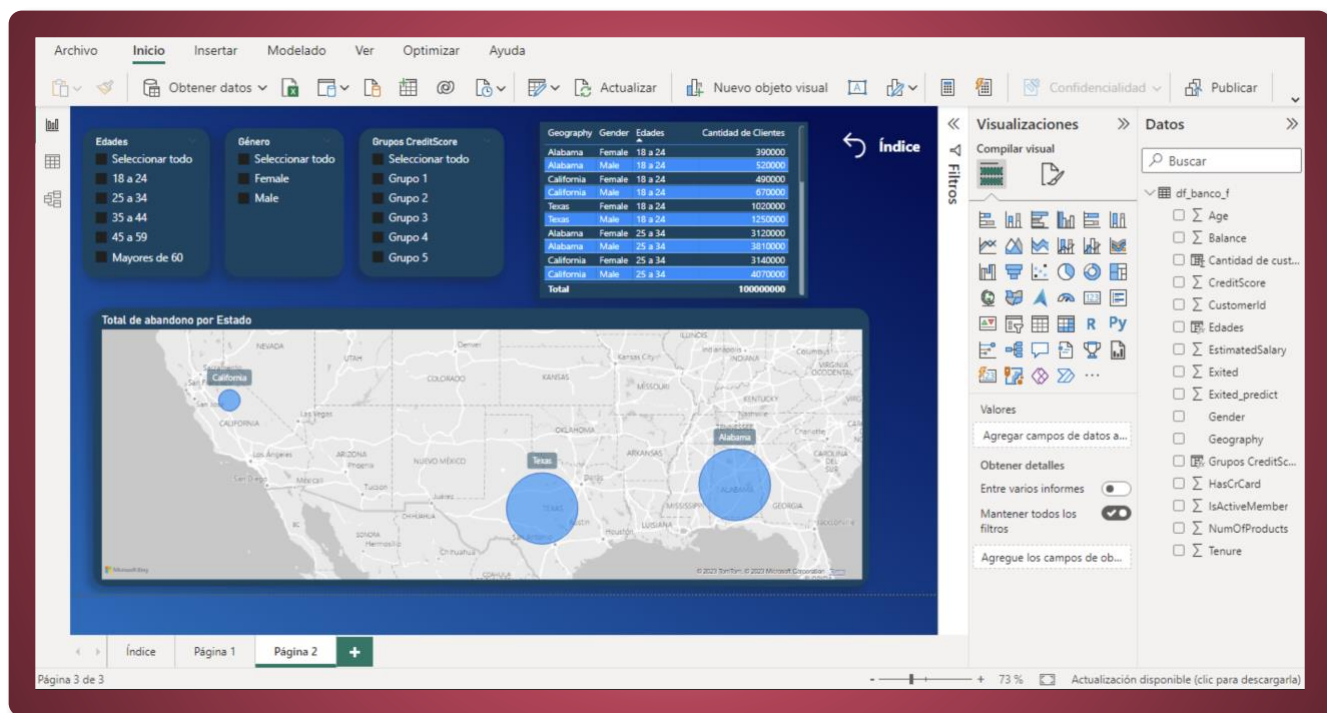


EJERCICIO 5

Aunque era opcional, decidimos completar este ejercicio en el que Jaime además de crear la página inicial, se encargó de la elección de los colores, ayudado por Jéssica y del diseño. Dándole una imagen más homogénea y profesional, con un punto corporativo.

La presentación queda de esta manera más fácil y atractiva de cara a presentarla a un cliente.







ANEXO

Adjuntamos en este enlace el archivo correspondiente a nuestro trabajo con POWER BI y un GIF realizado para que se vea el uso de botones.

[POWER BI](#)

[GIF](#)