

Tabla de contenido

Integrantes	1
Análisis Crediticio - préstamos personales	3
Introducción	3
Hipótesis	3
Target del Informe	3
Descripción del dataset	3
Diagrama Entidad - Relación	5
Listado de tablas	6
Columnas por tablas y definiciones de tipo de dato	6
Extracción de información	11
Carga de archivos planos en Microsoft SQL Server Manage	ment Studio
	11
Carga de tablas en Power Bl Desktop	11
Modelo relacional en Power BI Desktop	12
Modificación de la tabla ScoreExterno	14
Transformación de datos	15
Columnas y medidas calculadas	19
Análisis funcional del tablero	22

Análisis Crediticio - préstamos personales

Introducción

En este documento, nuestra propuesta es realizar un análisis crediticio sobre un dataset de préstamos personales otorgados por una entidad financiera. Uno de los principales objetivos como analistas de riesgos será el de poder identificar a los clientes con mayor probabilidad de no pagar sus obligaciones.

Hipótesis

A priori nuestra intuición nos dice que los hombres, con menor grado de calificación y en situaciones laborales poco calificadas tendrán una mayor probabilidad de ser morosos respecto al resto.

Target del Informe

El presente informe esta dirigido a analistas de riesgo crediticio, bancarios y asesores de préstamos. Encargados de identificar a los clientes con mayores probabilidades de generar deudas y no cumplir con sus obligaciones bancarias.

Descripción del dataset

El dataset elegido está conformado por un total de 122 columnas (variables) y 307.511 registros. Debido a la extensión del dataset, se decidió por separarlo en 5 archivos planos diferentes (hay un sexto), cada uno con una cierta lógica de agrupación de información:

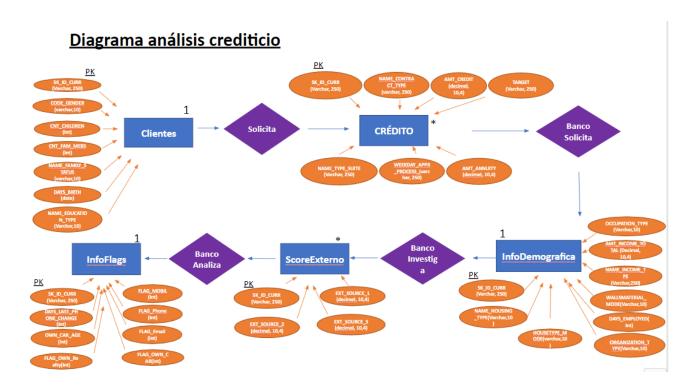
- InfoCliente: contiene información de la persona solicitante.
- InfoCredito: contiene información sobre el préstamo otorgado.
- ScoreExterno: contiene información de 3 pedidos de scores a entidades financieras externas.
- InfoFlags: contiene información en formato de "flags" o marcas (0/1).
- InfoDemografica: contiene información socioeconómica del cliente.

- InfoFechas: contiene la fecha de otorgamiento del préstamo por cliente.
- InfoExtra: contiene información adicional por si llegamos a necesitarla.
 (aquí se presentan las variables remanentes no incluidas en los archivos mencionados anteriormente).

Como podrá observarse en las siguientes secciones, la información incluida en los datasets propuestos anteriormente, se encuentra en el idioma inglés ya que su dataset original está confeccionado de esa manera. A medida que vayamos avanzando en el análisis, iremos traduciendo los campos que sean necesarios.

Diagrama Entidad - Relación

A continuación exponemos el diagrama de Entidad - Relación (E-R) desarrollado en base a la información seleccionadas:



Listado de tablas

A continuación presentamos el listado de tablas/datasets que serán utilizadas en este proyecto final.

# Tabla	Nombre	# Columnas	Primary Key (PK)
1	InfoCliente	7	SK_ID_CURR
2	InfoCredito	7	SK_ID_CURR
3	ScoreExterno	4	SK_ID_CURR
4	InfoFlags	8	SK_ID_CURR
5	InfoDemografica	9	SK_ID_CURR
6	InfoExtra	92	SK_ID_CURR

Columnas por tablas y definiciones de tipo de dato

A continuación presentamos por cada dataset, su composición en cuanto a columnas y tipo de datos asociados.

InfoCliente

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	CODE_GENDER	307.511	varchar
2	CNT_CHILDREN	307.511	integer
3	CNT_FAM_MEMBERS	307.509	float
4	NAME_FAMILY_STATUS	307.511	varchar
5	NAME_EDUCATION_TYPE	307.511	varchar
6	DAYS_BIRTH	307.511	integer

InfoCredito

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	NAME_CONTRACT_TYPE	307.511	varchar
2	AMT_CREDIT	307.511	float
3	AMT_ANNUITY	307.499	float
4	WEEKDAY_APPR_PROCESS	307.511	varchar
5	NAME_TYPE_SUITE	306.219	varchar
6	TARGET	307.511	integer

InfoDemog

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	AMT_INCOME_TOTAL	307.511	float
2	NAME_INCOME_TYPE	307.511	varchar
3	DAYS_EMPLOYED	307.511	integer
4	HOUSETYPE_MODE	153.214	varchar
5	WALLSMATERIAL_MODE	151.170	varchar
6	OCCUPATION_TYPE	211.120	varchar
7	NAME_HOUSING_TYPE	307.511	varchar
8	ORGANIZATION_TYPE	307.511	varchar

InfoFlags

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	FLAG_MOBIL	307.511	integer
2	FLAG_PHONE	307.511	integer
3	FLAG_EMAIL	307.511	integer
4	FLAG_OWN_CAR	307.511	varchar
5	FLAG_OWN_REALTY	307.511	varchar
6	DAYS_LAST_PHONE_CHAN	307.510	float
7	OWN_CAR_AGE	104.582	float

ScoreExterno

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	EXT_SOURCE_1	134.133	float
2	EXT_SOURCE_2	306.851	float
3	EXT_SOURCE_3	246.546	float

InfoFechas

#	Columna	#Datos no nulos Tipo de dato
	0 IdCliente	307.512 Integer
	1 Fecha	307.512 Date

InfoExtra (i)

#	Columna	# Datos no nulos	Tipo de dato
0	SK_ID_CURR	307.511	integer
1	AMT_GOODS_PRICE	307.233	float
2	REGION_POPULATION_REI	307.511	float
3	DAYS_REGISTRATION	307.511	float
4	DAYS_ID_PUBLISH	307.511	integer
5	FLAG_EMP_PHONE	307.511	integer
6	FLAG_WORK_PHONE	307.511	integer
7	FLAG_CONT_MOBILE	307.511	integer
8	REGION_RATING_CLIENT	307.511	integer
9	REGION_RATING_CLIENT_'	307.511	integer
10	HOUR_APPR_PROCESS_ST.	307.511	integer
11	REG_REGION_NOT_LIVE_R	307.511	integer
12	REG_REGION_NOT_WORK	307.511	integer
13	LIVE_REGION_NOT_WORK	307.511	integer
14	REG_CITY_NOT_LIVE_CITY	307.511	integer
15	REG_CITY_NOT_WORK_CI1	307.511	integer
16	LIVE_CITY_NOT_WORK_CI	307.511	integer
17	APARTMENTS_AVG	151.450	float
18	BASEMENTAREA_AVG	127.568	float
19	YEARS_BEGINEXPLUATATION	157.504	float
20	YEARS BUILD AVG	103.023	float

InfoExtra (ii)

#	Columna	# Datos no nulos	Tipo de dato
	COMMONAREA_AVG	92.646	float
22	ELEVATORS_AVG	143.620	float
23	ENTRANCES_AVG	152.683	float
24	FLOORSMAX_AVG	154.491	float
25	FLOORSMIN_AVG	98.869	float
26	LANDAREA_AVG	124.921	float
27	LIVINGAPARTMENTS_AVG	97.312	float
28	LIVINGAREA_AVG	153.161	float
29	NONLIVINGAPARTMENTS_	93.997	float
30	NONLIVINGAREA_AVG	137.829	float
31	APARTMENTS_MODE	151.450	float
32	BASEMENTAREA_MODE	127.568	float
33	YEARS_BEGINEXPLUATATION	157.504	float
34	YEARS_BUILD_MODE	103.023	float
35	COMMONAREA_MODE	92.646	float
36	ELEVATORS_MODE	143.620	float
37	ENTRANCES_MODE	152.683	float
38	FLOORSMAX_MODE	154.491	float
39	FLOORSMIN_MODE	98.869	float
40	LANDAREA_MODE	124.921	float

InfoExtra (iii)

#	Columna	# Datos no nulos	Tipo de dato
41	LIVINGAPARTMENTS_MOD	97.312	float
42	LIVINGAREA_MODE	153.161	float
43	NONLIVINGAPARTMENTS_	93.997	float
44	NONLIVINGAREA_MODE	137.829	float
45	APARTMENTS_MEDI	151.450	float
46	BASEMENTAREA_MEDI	127.568	float
47	YEARS_BEGINEXPLUATATION	157.504	float
48	YEARS_BUILD_MEDI	103.023	float
49	COMMONAREA_MEDI	92.646	float
50	ELEVATORS_MEDI	143.620	float
51	ENTRANCES_MEDI	152.683	float
52	FLOORSMAX_MEDI	154.491	float
53	FLOORSMIN_MEDI	98.869	float
54	LANDAREA_MEDI	124.921	float
55	LIVINGAPARTMENTS_MED	97.312	float
56	LIVINGAREA_MEDI	153.161	float
57	NONLIVINGAPARTMENTS_	93.997	float
58	NONLIVINGAREA_MEDI	137.829	float
59	FONDKAPREMONT_MODE	97.216	varchar
60	TOTALAREA_MODE	159.080	float

InfoExtra (iv)

#	Columna	# Datos no nulos	Tipo de dato
61	EMERGENCYSTATE_MODE	161.756	varchar
62	OBS_30_CNT_SOCIAL_CIR(306.490	float
63	DEF_30_CNT_SOCIAL_CIRC	306.490	float
64	OBS_60_CNT_SOCIAL_CIR(306.490	float
65	DEF_60_CNT_SOCIAL_CIRC	306.490	float
66	FLAG_DOCUMENT_2	307.511	integer
67	FLAG_DOCUMENT_3	307.511	integer
68	FLAG_DOCUMENT_4	307.511	integer
69	FLAG_DOCUMENT_5	307.511	integer
70	FLAG_DOCUMENT_6	307.511	integer
71	FLAG_DOCUMENT_7	307.511	integer
72	FLAG_DOCUMENT_8	307.511	integer
73	FLAG_DOCUMENT_9	307.511	integer
74	FLAG_DOCUMENT_10	307.511	integer
75	FLAG_DOCUMENT_11	307.511	integer
76	FLAG_DOCUMENT_12	307.511	integer
77	FLAG_DOCUMENT_13	307.511	integer
78	FLAG_DOCUMENT_14	307.511	integer
79	FLAG_DOCUMENT_15	307.511	integer
80	FLAG_DOCUMENT_16	307.511	integer

InfoExtra (v)

#	Columna	# Datos no nulos	Tipo de dato		
81	FLAG_DOCUMENT_17	307.511	integer		
82	FLAG_DOCUMENT_18	307.511	integer		
83	FLAG_DOCUMENT_19	307.511	integer		
84	FLAG_DOCUMENT_20	307.511	integer		
85	FLAG_DOCUMENT_21	307.511	integer		
86	AMT_REQ_CREDIT_BUREAU_HOUR	265.992	float		
87	AMT_REQ_CREDIT_BUREAU_DAY	265.992	float		
88	AMT_REQ_CREDIT_BUREAU_WEEK	265.992	float		
89	AMT_REQ_CREDIT_BUREAU_MON	265.992	float		
90	AMT_REQ_CREDIT_BUREAU_QRT	265.992	float		
91	AMT REQ CREDIT BUREAU YEAR	265.992	float		

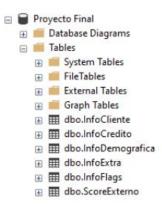
Extracción de información

La siguiente sección intenta detallar los procesos seguidos para la utilización de la herramienta Power BI Desktop.

Carga de archivos planos en Microsoft SQL Server Management Studio

Una vez dentro del programa, haciendo click derecho sobre la carpeta "databases" elegimos la opción "New Database..." y creamos un nuevo database llamado "Proyecto Final".

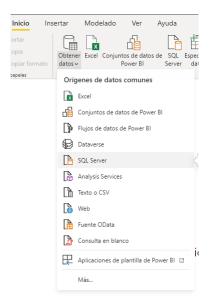
En el database recientemente creado cargamos los archivos planos presentamos como fuentes de información finales para la entrega del proyecto final. Por consiguiente, subimos los 6 archivos planos (.csv) mencionados en la sección anterior.



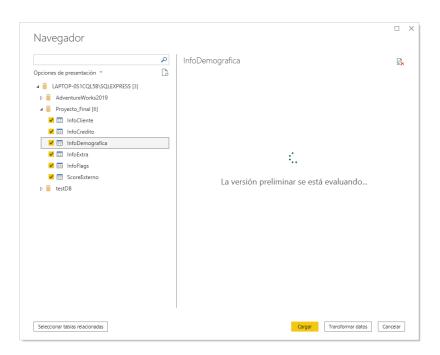
Una vez terminada esta etapa, procedimos a abrir y cargar dichas tablas en el Microsoft Power BI Desktop.

Carga de tablas en Power BI Desktop

Al cargar el programa, el primer paso fue hacer la conexión con nuestro servidor de SQL dentro de Power BI.

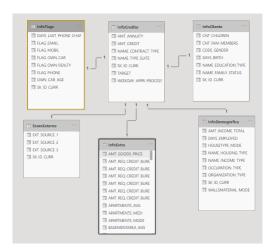


En el siguiente paso, se realizó la conexión al servidor local y acto seguido se especificó el database y las tablas a ser importadas.

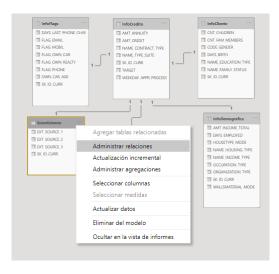


Modelo relacional en Power BI Desktop

La IA (Inteligencia Artificial) del programa había seleccionado correctamente el Primary Key pero había seleccionado erróneamente como tabla "madre" la denominada InfoCliente. Lo que hicimos fue modificar las relaciones del modelo Entidad-Relación para que todas las tablas tomen como referencia la tabla InfoCredito ya que es esa la tabla que utilizaremos como tabla principal.



Los cambios realizados mencionados anteriormente fueron hechos mediante la opción "Administrar relaciones". Allí se fueron modificando manualmente las conexiones.



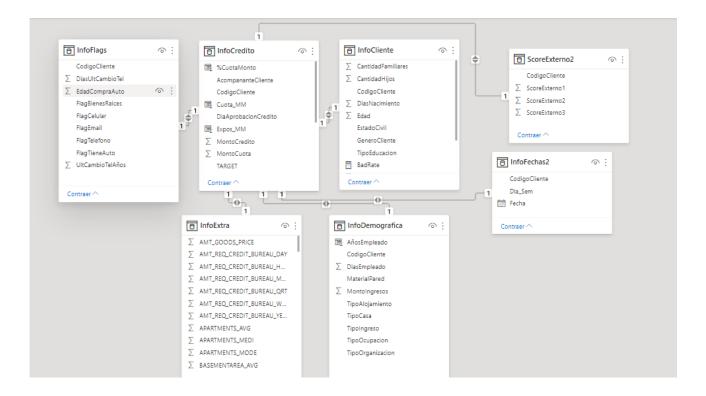
Modificación de la tabla ScoreExterno

Al importar la tabla ScoreExterno nos dimos cuenta que la misma estaba trayendo valores sumamente equivocados para las columnas Ext_Source_1, Ext_Source_2 y Ext_Source_3. Esto se debía al delimitador del archivo plano (.csv) el cual usaba la coma (,) y confundía los números pasándolos a números enteros demasiado grandes. Decidimos cambiar en el archivo plano el delimitador (,) por (|).

Sabiendo que los campos Ext_Source_[i] representaban probabilidades (cuyo dominio son los valores entre 0 y 1) decidimos crear mediante un query una nueva tabla llamada ScoreExterno2 donde transformamos y renombramos dichos campos. A continuación dejamos expresado el query mencionado anteriormente.

```
select SK_ID_CURR as CodigoCliente,
cast(EXT_SOURCE_1 as decimal(4,4)) as ScoreExterno1,
cast(EXT_SOURCE_2 as decimal(4,4)) as ScoreExterno2,
cast(EXT_SOURCE_3 as decimal(4,4)) as ScoreExterno3
into Proyecto_Final.dbo.ScoreExterno2
from Proyecto Final.dbo.ScoreExterno
```

El nuevo diagrama de Entidad-Relación quedó de la siguiente manera



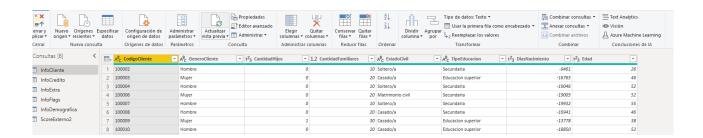
Transformación de datos

Debido a que la base de datos utilizada se encontraba en inglés, procedimos a realizar la traducción tanto de las columnas de cada tabla como también la información en ellas contenida. Para ello utilizamos la herramienta de manipulación de datos "Transformar datos" para realizar las modificaciones necesarias.

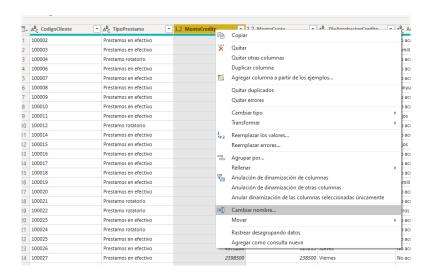


Al utilizar la herramienta, se habilitó una ventana aparte que nos permitió navegar por las distintas tablas y sus contenidos.

Nuestro primer paso fue renombrar todas las columnas a su traducción al español, siguiendo el consejo de escribir los nombres con el estilo Camel Case.



Renombramos los atributos utilizando la opción Cambiar nombre haciendo clic derecho sobre el nombre del atributo correspondiente.



Lo mencionado anteriormente fue repetido para todas las tablas y atributos salvo por la tabla de datos llamada "InfoExtra". Debido a la cantidad de

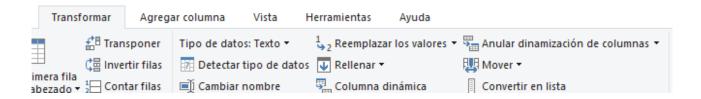
atributos allí presentes, decidimos dejarla como tabla back up para extraer más información en caso de ser necesario. Llegado el momento se modificarán los atributos que se consideren pertinentes y necesarios.

Una vez renombrados los atributos, pasamos a transformar los datos internos de cada tabla.

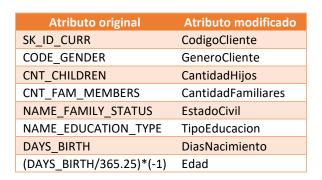
A continuación presentamos el detalle de las modificaciones realizadas para cada tabla importada dentro de nuestro proyecto final.

INFOCLIENTE

Se tradujeron al español los atributos y la información contenida en ellos también. El proceso de modificación de información fue realizado mediante la opción de Transformar/Reemplazar los valores la cual nos permitió ir modificando los datos en bloques.



Por otro lado, se creó un nuevo atributo Edad, el cual venía a resolver el problema del campo DAYS_BIRTH el cual cuenta con la cantidad de días pero con signo negativo. Utilizamos esa información para crear la Edad del cliente. En la siguiente tabla podrán encontrar la fórmula utilizada para dicho cálculo.



INFOCREDITO

Para este set de datos procedimos a traducir los atributos y datos.

Atributo original	Atributo modificado
SK_ID_CURR	CodigoCliente
NAME_CONTRACT_TYPE	TipoPrestamo
AMT_CREDIT	Montocredito
AMT_ANNUITY	MontoCuota
WEEKDAY_APPR_PROCESS_START	DiasAprobacionCredito
NAME_TYPE_SUITE	AcompananteCliente
TARGET	Target

INFOFLAGS

Para este set de datos procedimos a traducir los atributos y datos respectivamente. Al igual que en la tabla InfoCliente, para solucionar el dato de DAYS_LAST_PHONE_CHANGE el cual venía en valores negativos, procedimos a crear un nuevo atributo llamado "UltCambioTelAño" el cual transforma la información de días en años.

Atributo original	Atributo modificado
SK_ID_CURR	CodigoCliente
FLAG_MOBIL	FlagCelular
FLAG_PHONE	FlagTelefono
FLAG_EMAIL	FlagEmail
FLAG_OWN_CAR	FlagTieneAuto
FLAG_OWN_REALTY	FlagBienesRaices
DAYS_LAST_PHONE_CHANGE	DiasUltCambioTel
OWN_CAR_AGE	EdadCompraAuto
(DAYS_LAST_PHONE_CHANGE/365.25)*(-1)	UltCambioTelAños

SCOREEXTERNO2

Como fue mencionado anteriormente, esta tabla fue creada en base a ScoreExterno. Dicha transformación fue realizada directamente en SQL.

INFODEMOGRAFICA

Como en las tablas anteriores, tradujimos tantos los atributos como los datos en ellos.

Atributo original	Atributo modificado
SK_ID_CURR	CodigoCliente
AMT_INCOME_TOTAL	MontoIngresos
NAME_INCOME_TYPE	TipoIngreso
DAYS_EMPLOYED	Dias Empleado
HOUSETYPE_MODE	TipoCasa
WALLSMATERIAL_MODE	MaterialPared
OCCUPATION_TYPE	TipoOcupacion
NAME_HOUSING_TYPE	TipoAlojamiento
ORGANIZATION_TYPE	TipoOrganizacion

Columnas y medidas calculadas

Como primer paso decidimos crear 2 columnas calculadas en base a la información disponible en las tablas InfoCredito e InfoDemografica.

%CUOTAMONTO

Creamos la columna "%CuotaMonto" la cual indica la relación porcentual entre lo que el cliente deberá pagar como cuota del préstamo y el monto total del préstamo otorgado.

Para crear esta columna utilizamos la opción "Nueva columna" en Herramientas de tablas. Allí realizamos el cálculo de la siguiente manera:

%CuotaMonto = InfoCredito[MontoCuota]/InfoCredito[MontoCredito]

AÑOSEMPLEADO

Para crear dicha columna utilizamos la información provista por el atributo DiasEmpleado de la tabla InfoDemografica. Al crear esta nueva columna nos dimos cuenta de que el atributo DiasEmpleado presentaba muchas veces el valor 365243 y después de analizar dicha columna como los atributos, llegamos a la conclusión que ese número se utiliza como default para aquellos registros que no parecen tener un empleo activo; por ejemplo, el número 365243 aparece sobre todo en aquellos clientes cuyo estado de TipoIngreso es "Desempleado" o "Pensionado".

Por consiguiente decidimos calcular el nuevo atributo con un condicional: si la persona presenta valores en DiasEmpleado que sean menores a cero, entonces se transforman los días en años, caso contrario utilizamos el valor menos uno (-1) como indicador de valor "nulo". A continuación mostramos la fórmula de cálculo:

AñosEmpleado = IF(InfoDemografica[DiasEmpleado]<0,(1)*(InfoDemografica[DiasEmpleado]/365.25),-1)</pre>

Una vez creadas estas dos columnas, proseguimos a crear las 4 medidas calculadas, que se verán reflejadas en la sección Informe. El detalle de las medidas calculadas se encuentran a continuación:

- Sum_Monto_Credito: es la sumatoria del monto de la totalidad de prestamos otorgados. Como puede observarse, dicho monto equivale a \$1,84 billones.

```
Sum_Monto_Credito = CALCULATE(SUM(InfoCredito[MontoCredito]))
```

- Prom_Monto_Cuota: es el promedio de las cuotas que pagan la totalidad de los clientes. La misma equivale a una cuota promedio de \$271,09 mil.

```
Prom_Monto_Cuota = CALCULATE(AVERAGE(InfoCredito[MontoCuota]))
```

 Prom_Edad: es el promedio de edad de los clientes. Dicho valor es de 44 años (promedio).

```
Prom_Edad = CALCULATE(ROUND(AVERAGE(InfoCliente[Edad]),0))
```

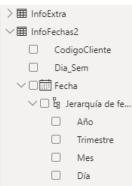
 Max_Cant_Hijos: es el valor máximo de cantidad de hijos de cada cliente que presenta la base. Como se puede ver, hay clientes que tienen hasta un máximo de 19 hijos.

```
Max Cant Hijos = CALCULATE(MAX(InfoCliente[CantidadHijos]))
```

Debido a que la base de datos no contaba con un atributo de tipo fecha, decidimos generarlos en una nueva tabla. Para ello, aplicando una fórmula en Excel, tomamos el atributo idCliente y le asignamos aleatoriamente fechas de días hábiles correspondientes al año 2020 entre los meses de enero y febrero. Una vez realizado esto, guardamos dicho archivo como un archivo .csv y lo importamos al SQL Server. Allí realizamos una transformación del campo Fecha para que quede en formato date y lo guardamos en una nueva tabla a la cual llamamos InfoFechas2. Lo mencionado anteriormente fue realizado con el siguiente query:

```
SELECT IdCliente, cast(fecha as date) as Fecha, Dia_Sem
INTO [Proyecto_Final].[dbo].[InfoFechas2]
FROM [Proyecto_Final].[dbo].[InfoFechas]
```

Acto seguido, importamos la tabla en Power BI, nos aseguramos de que la relación a través del Primary Key (PK) sea la correcta y el programa a través de su IA creó automáticamente un campo Fecha con la jerarquía correcta de tipo Calendar como puede verse a contuanción.



Análisis funcional del tablero

Finalmente, para el proyecto final presentar la siguiente información:

- 1. En la primera solapa un **Resumen** de los préstamos otorgados, abiertos por Tipo de Préstamo. En la misma, mostramos:
 - Monto total préstamos otorgados, expresados en millones de \$.
 - Q total de préstamos.
 - Monto total préstamos morosos, expresados en millones de \$.
 - Q total de préstamos morosos.
- 2. Página 2: Conozca a su cliente. En la misma, mostramos información referida las variables blandas y duras de los clientes.
 - Apertura por estado civil y acompañante.
 - Cantidad de clientes por género.
 - Máximo nivel educativo alcanzado.
- Página 3: <u>Performance.</u> En esta solapa, se pueden visualizar las principales características de los clientes, abiertos por morosos/no morosos.
- 4. Página 4: <u>Historia.</u> Muestra la cuota promedio y la exposición abierta por día y morosidad.

FORMATO DEL DASHBOARD

Para nuestro Dashboard elegimos la tipografía Segoe y sus variantes Bold y light:

Segoe UI

Seleccionamos esta tipografía ya que la misma expresa claridad y elegancia, dos cualidades valoradas en el ámbito bancario.

Respecto de la paleta de colores, utilizamos una variación del verde que, junto con el negro de la tipografía, mantienen el tono serio e institucional. Además, el verde es un color que socialmente se encuentra asociado a las transacciones bancarias.

El mismo lo obtuvimos de una de las páginas proporcionadas por el curso.

