UNIVERSIDAD DON BOSCO FACULTAD DE INGENIERÍA ESCUELA DE COMPUTACIÓN



"Desafío No. 1"

DATAWAREHOUSE Y MINERIA DE DATOS

CATEDRATICA:

Inga. Karens Medrano

PRESENTADO POR:

Cristian Alberto Montoya Reyes	MR211303
Victor Amilcar Elías Peña	EP171613
Kevin Alexander Hernández Cerón	HC162209
Katherine Paola Pineda Rodriguez	PR232427

sábado 16 de mar. de 24 SOYAPANGO, EL SALVADOR

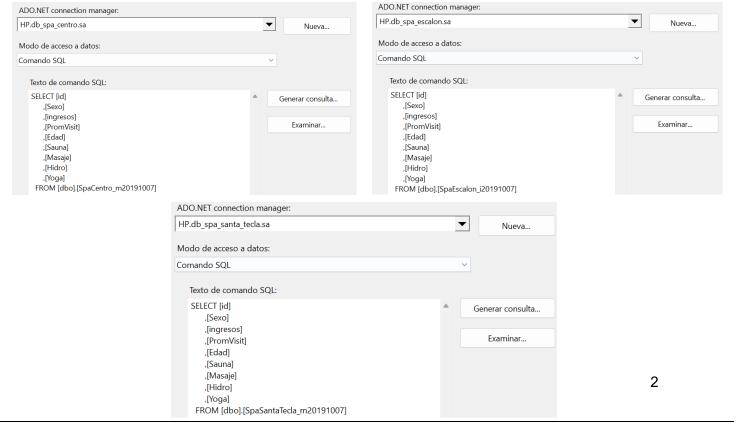
EJERCICIO 1

El Spa, "Diego", necesita segmentar sus clientes, para realizar una campaña de fidelización, y le pide a usted que efectué un análisis de sus tres sucursales, que defina cuantos grupos y que características tienen

Para este ejercicio se solvento así.



Cada una de las DATAs fueron importadas a la base de datos SQL Server. Una bd por cada una y luego en los orígenes se configuraron con las siguientes consultas SQL



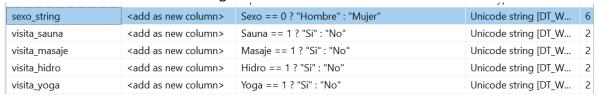
Con un Union unimos todos los datos de los tres orígenes

Output Column Name	Union All Input 1	Union All Input 2	Un
id	id	id	id
Sexo	Sexo	Sexo	Se
ingresos	ingresos	ingresos	ing
PromVisit	PromVisit	PromVisit	Pro
Edad	Edad	Edad	Ed
Sauna	Sauna	Sauna	Sa
Masaje	Masaje	Masaje	Ma
Hidro	Hidro	Hidro	Hic
Yoga	Yoga	Yoga	Yo

Agrupamos por cliente. Sumamos los montos y promediamos en general el promedio que ya nos da

Input Column	Output Alias	Operation	Com
id	id	Group by	
Sexo	Sexo	Group by	
ingresos	ingresos	Sum	
PromVisit	PromVisit	Average	
Edad	Edad	Group by	
Sauna	Sauna	Group by	
Massis	Massis	Craus bu	

Las columnas derivadas son las siguientes

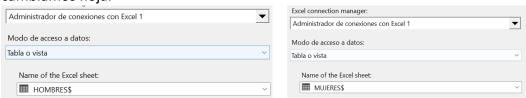


Con esto logramos dividir el genero masculino del femenino y además poner en con palabra "Si" o "No" si consume un servicio

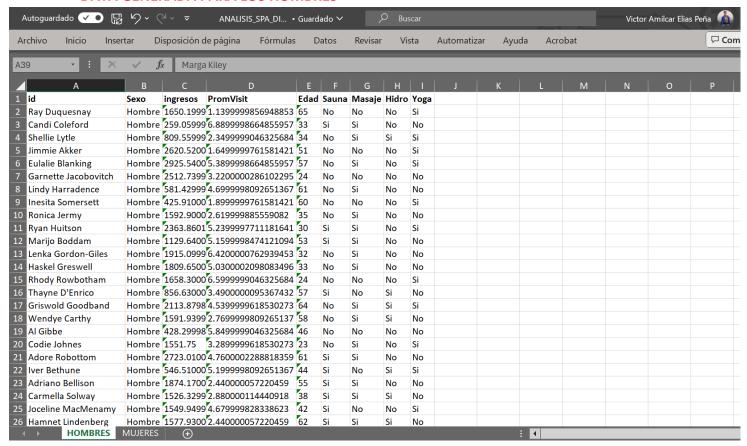
En la condicional separamos los géneros

Order	Output Name	Condition
1	MUJER	Sexo == 1
2	HOMBRE	Sexo == 0

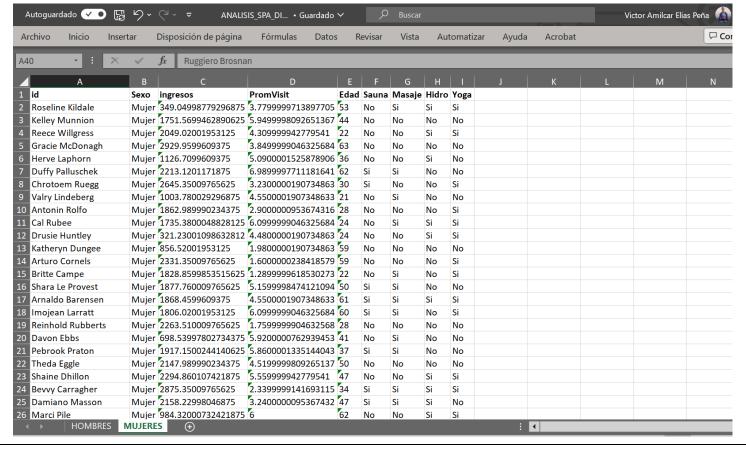
Esto nos genera dos datas diferentes, pero lo estamos en el mismo documento, solo cambiamos hoja.



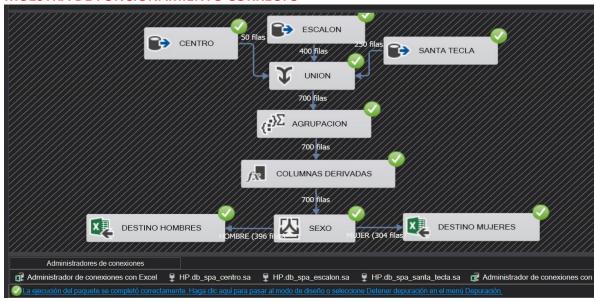
DATA GENERADA PARA LOS HOMBRES



DATA GENERADA PARA LAS MUJERES



MUESTRA DE FUNCIONAMIENTO CORRECTO



PORCENTAJE ALCANZADO: 100%

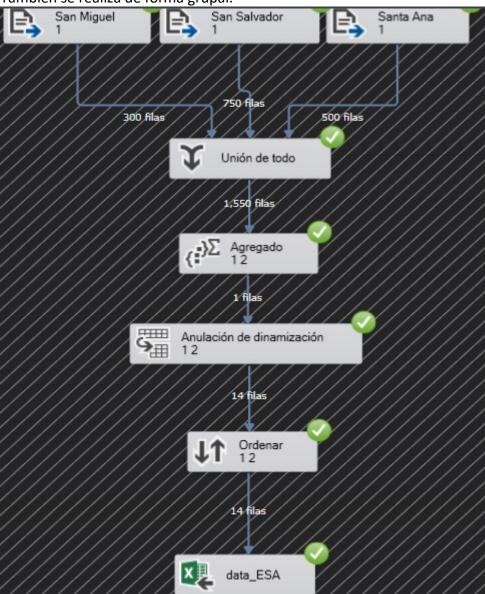
EJERCICIO 2

La Floristería "Fiorella" quiere saber cómo se compran sus productos, y tiene la data de tres departamentos del país, por lo cual les pide su opinión sobre qué productos sobresalen, que combinaciones son mejores y quieren este estudio por departamento y también por país.

El ejercicio 2 se realizo de la siguiente manera. Datas divididas por sucursal para conocer la información individual de cada una.



También se realiza de forma grupal.



Con esto obtenemos la información de cada una de los departamentos, pero también la información agrupada o sumada de todo el salvador.

DATA DE SAN SALVADOR

1	producto	cantidad
2	Listón	690
3	Rosas	612
4	Globos	587
5	Macetas	392
6	Aurora	384
7	Tarjetas	384
8	fOrquÃ-dia	380
9	Hortensia	374
10	Girasoles	371
11	Tierra	368
12	Lirios	365
13	Tulipanes	357
14	CarmesÃ-	353
15	Claveles	350

DATA DE SAN SAN MIGUEL

producto	cantidad
Aurora	160
Lirios	160
CarmesÃ-	158
fOrquÃ-dia	158
Hortensia	157
Rosas	157
Globos	151
Girasoles	150
Listón	149
Tulipanes	149
Tarjetas	143
Macetas	141
Tierra	141
Claveles	137
	Aurora Lirios CarmesÃ- fOrquÃ-dia Hortensia Rosas Globos Girasoles Listón Tulipanes Tarjetas Macetas Tierra

DATA DE SAN SANTA ANA

1	producto	cantidad
2	Lirios	270
3	Girasoles	266
4	Aurora	260
5	fOrquÃ-dia	259
6	Tarjetas	252
7	Tulipanes	247
8	Claveles	246
9	Macetas	245
10	Hortensia	243
11	CarmesÃ-	236
12	Tierra	236
13	Rosas	176
14	Globos	154
15	Listón	136

DATA DE EL SALVADOR

DATAD	DATA DE EL SALVADOR		
1	producto	cantidad	
2	Lirios	270	
3	Girasoles	266	
4	Aurora	260	
5	fOrquÃ-dia	259	
6	Tarjetas	252	
7	Tulipanes	247	
8	Claveles	246	
9	Macetas	245	
10	Hortensia	243	
11	CarmesÃ-	236	
12	Tierra	236	
13	Rosas	176	
14	Globos	154	
15	Listón	136	

DE AQUÍ PODEMOS VER CUAL ES EL PRODUCTO QUE MAS SE VENDE, POR DEPARTAMENTO Y TOTALIZADO EN EL PAIS

FUNCIONAMIENTO AL 100%

Administradores de conexiones

gill Administrador de conexiones con Excel 1

gill Administrador de conexiones con Excel 3

gill Administrador de conexiones con Excel 4

gill Administrador de conexiones con Excel 5

gill Administrador de conexiones con Excel 4

gill Administrador de conexiones con Excel 5

gill Administrador de conexiones con Excel 3

gill Administrador de conexiones con Excel 4

gill Administrador de conexiones con Excel 3

gill Administrador de conexiones con Excel 4

gill Administrador de conexiones con

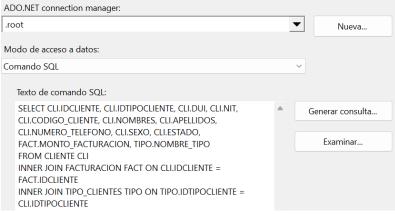
EJERCICIO 3

La telefonía "FioDio" solicita realizar un ETL que exporte una base de datos de Mysql y SQL Server, al final el destino serán dos archivos de Excel en donde en un archivo estarán los clientes preferenciales y ejecutivos y en el segundo los de gobierno y turista, adicional en los archivos de Excel se deberá crear un campo código de país, que se llenará sustraendo los dos primeros caracteres de código cliente.

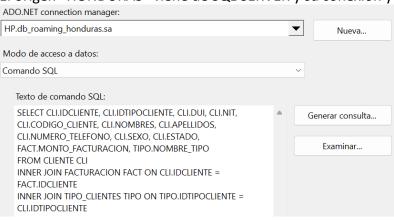
Se desarrolló de la siguiente manera:



El origen "COSTA RICA" viene de MySql y aquí se muestra la conexión y consulta SQL



El origen "HONDURAS" viene de SQL SERVER y su conexión y consulta SQL es la siguiente



Esta información tenemos que juntarla y para eso utilizamos un UNION ALL

Output Column Name	Union All Input 1	Union All Input 2
IDCLIENTE	IDCLIENTE	IDCLIENTE
IDTIPOCLIENTE	IDTIPOCLIENTE	IDTIPOCLIENTE
DUI	DUI	DUI
NIT	NIT	NIT
CODIGO_CLIENTE	CODIGO_CLIENTE	CODIGO_CLIENTE
NOMBRES	NOMBRES	NOMBRES
APELLIDOS	APELLIDOS	APELLIDOS
NUMERO_TELEFONO	NUMERO_TELEFONO	NUMERO_TELEFONO
SEXO	SEXO	SEXO
ESTADO	ESTADO	ESTADO
MONTO_FACTURACION	MONTO_FACTURACION	MONTO_FACTURACION
NOMBRE_TIPO	NOMBRE_TIPO	NOMBRE_TIPO

La agrupación nos sirve para sumar todos los montos de registros que sean idénticos (Mismo cliente)

Input Column	Output Alias	Operation	Corr
APELLIDOS	APELLIDOS	Group by	
NUMERO_TELEFONO	NUMERO_TELEFONO	Group by	
SEXO	SEXO	Group by	
ESTADO	ESTADO	Group by	
MONTO_FACTURACION	MONTO_FACTURACION	Sum	
NOMBRE_TIPO	NOMBRE_TIPO	Group by	

En columna derivada creamos las siguientes columnas

Derived Column Name	Derived Column	Expression	Data Type	Le
codigo_pais	<add as="" column="" new=""></add>	SUBSTRING(CODIGO_CLIENTE,1,2)	Unicode string [DT_W	2
nombre_completo	<add as="" column="" new=""></add>	NOMBRES + ", " + APELLIDOS	Unicode string [DT_W	50
sexo_nombre	<add as="" column="" new=""></add>	SEXO == "m" ? "MASCULINO" : "FEMENINO"	Unicode string [DT_W	9
codigo_cliente_upper	<add as="" column="" new=""></add>	UPPER(CODIGO_CLIENTE)	Unicode string [DT_W	15
dui_formato	<add as="" column="" new=""></add>	SUBSTRING(DUI,1,8) + "-" + SUBSTRING(DUI,9,1)	Unicode string [DT_W	10
nit_formato	<add as="" column="" new=""></add>	SUBSTRING(NIT,1,4) + "-" + SUBSTRING(NIT,5,6)	Unicode string [DT_W	17
telefono_formato	<add as="" column="" new=""></add>	SUBSTRING(NUMERO_TELEFONO,1,4) + "-" + SU	Unicode string [DT_W	9
estado_nombre	<add as="" column="" new=""></add>	ESTADO == "t" ? "ACTIVO" : "INACTIVO"	Unicode string [DT_W	8
			-	

El condicional nos sirve para que determinemos los tipos de cliente. Ya que el ejercicio nos pide separar los Preferenciales y Ejecutivos de los de Gobierno y Turistas

Order	Output Name	Condition
1	PREFERENCIAL_EJECUTIVO	(IDTIPOCLIENTE == 1 IDTIPOCLIENTE == 2)
2	GOBIERNO_TURISTA	(IDTIPOCLIENTE == 3 IDTIPOCLIENTE == 4)

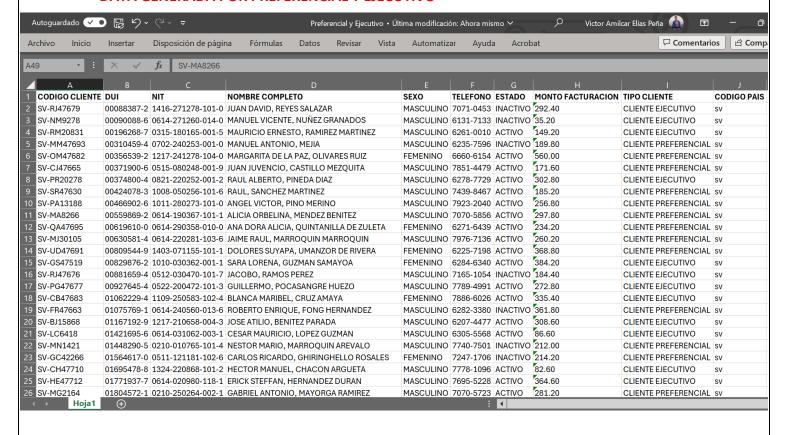
Ambos ordenamientos los hacemos por el DUI

Input Column	Output Alias	Sort Type	Sort Order	Com
DUI	DUI	ascending	1	
	201	uscending		

Por último, cada condición genera un Excel diferente



DATA GENERADA POR PREFERENCIAL Y EJECUTIVO



DATA GENERADA POR GOBIERNO Y TURISTA

Gobierno v T • Última modificación: sáb. a las 12:34 >

Archivo	Inicio	Insertar	Dispo	sición de págir	na Fórmulas Da	atos Revisar Vist	a Automat	tizar Ayı	uda Acro	bat	☐ Comentario	s
A57 · \vdots $ imes$ $ imes$ $ imes$ SV-DM47697												
4	Α			С		D						
1 COD	IGO CLIENTE	DUI	NIT		NOMBRE COMPLETO		SEXO	TELEFONO	ESTADO	MONTO FACTURACION	TIPO CLIENTE	CODIGO PAI
2 SV-P	W9858	00108338-6	0617-	210164-002-0	WALTER MAURICIO, P.	ALACIOS COLOCHO	MASCULINO	6227-2957	ACTIVO	164.60	CLIENTE GUBERNAMENTAL	sv
3 SV-N	S47705	00115399-5	0604-	090668-101-7	SIMEON, NAVARRETE	LEONOR	MASCULINO	6124-2645	ACTIVO	199.00	CLIENTE TURISTA	sv
4 SV-B	A47686	00119483-2	0614-	251259-006-0	ANA IRMA, BENITEZ AF	RGUETA	FEMENINO	6270-5458	ACTIVO	390.80	CLIENTE TURISTA	sv
5 SV-V	S5897	00141561-4	0610-	091159-101-0	SALVADOR, VASQUEZ	MELARA	MASCULINO	6333-5201	ACTIVO	334.40	CLIENTE GUBERNAMENTAL	SV
6 SV-R	A2158	00287394-6	0210-	250167-105-8	ANA LIZ, RODRIGUEZ [DE TOVAR	FEMENINO	6440-5701	INACTIVO	270.20	CLIENTE GUBERNAMENTAL	sv
7 SV-EI	M47658	00374984-9	0619-	030179-101-7	MARVIN VITELIO, ERAZ	O VASQUEZ	MASCULINO	7786-5956	INACTIVO	93.40	CLIENTE TURISTA	SV
SV-M	E47714	00449568-9	1009-	100279-101-7	ERIC ALEXANDER, ME.	IIA MOLINA	MASCULINO	7083-1594	ACTIVO	61.40	CLIENTE GUBERNAMENTAL	sv
9 SV-AI	E47669	00453608-7	0811-	301161-101-8	ESTANISLAO, ALVARA	DO ALVAREZ	MASCULINO	7993-9305	ACTIVO	88.20	CLIENTE GUBERNAMENTAL	SV
0 SV-N	J26082	00481311-2	0703-	250365-101-1	JOSE ANTONIO, NOLA	SCO LOPEZ	MASCULINO	7710-9443	ACTIVO	219.80	CLIENTE GUBERNAMENTAL	SV
1 SV-A	L47672	00492838-9	0210-	020382-105-2	LUIS ALONSO, ALVARI	Z HERNANDEZ	MASCULINO	7663-1438	INACTIVO	301.80	CLIENTE TURISTA	sv
2 SV-SI	R47678	00542374-0	0210-	150150-002-2	RICARDO ANTONIO, S	ILIEZAR SALINAS	MASCULINO	6293-0678	ACTIVO	61.20	CLIENTE GUBERNAMENTAL	SV
3 SV-JJ	47659	00557486-8	0203-	230383-101-7	JOSE RODOLFO, JIMEI	NEZ PINTIN	MASCULINO	6263-4944	ACTIVO	52.00	CLIENTE GUBERNAMENTAL	sv
4 SV-R	C47664	00569371-9	0614-	020383-132-1	CARLOS RODOLFO, R	OSALES MACHADO	MASCULINO	6566-6345	ACTIVO	367.60	CLIENTE TURISTA	sv
5 SV-D	M47697	00729179-1	0609-	150457-001-8	MANUEL HERIBERTO,	DURAN RODRIGUEZ	MASCULINO	7880-4336	INACTIVO	352.40	CLIENTE GUBERNAMENTAL	sv
6 SV-RI	M47671	00738151-2	0614-	310570-114-1	MANUEL ALFREDO, RI	VERA MEDRANO	MASCULINO	7189-5007	ACTIVO	302.80	CLIENTE TURISTA	sv
7 SV-D	R47657	00864458-1	0312-	050365-001-4	ROSSANA ELISABETH,	DIAZ DE CORVERA	FEMENINO	7457-5307	ACTIVO	283.80	CLIENTE TURISTA	sv
8 SV-M	F47700	00889131-6	0614-	091258-016-3	FRANCISCO REMBERT	O, MIXCO LOPEZ	MASCULINO	6276-3473	INACTIVO	155.20	CLIENTE TURISTA	SV
9 SV-P	C47670	01040243-2	0714-	230573-102-6	CRUZ MARINA, PALAC	IOS DE GUARDADO	FEMENINO	6203-9568	ACTIVO	194.60	CLIENTE TURISTA	sv
0 SV-FI	F47704	01054713-5	0511-	070282-101-9	FRANCISCO STEVE, FL	ORES SARMIENTO	MASCULINO	7921-1113	ACTIVO	70.80	CLIENTE TURISTA	SV
1 SV-G	M47660	01129927-2	0805-	120572-101-3	MANUEL ENRIQUE, GR	RANDE CASTELLANOS	MASCULINO	6284-5726	INACTIVO	186.20	CLIENTE GUBERNAMENTAL	sv

MUESTRA DE FUNCIONAMIENTO CORRECTO



PORCENTAJE ALCANZADO: 100%