

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

ADO.NET connection manager:
HP.db_spa_centro.sa Nueva...

Modo de acceso a datos:
Comando SQL

Texto de comando SQL:

```
SELECT [id]
,[Sexo]
,[Ingresos]
,[PromVisit]
,[Edad]
,[Sauna]
,[Masaje]
,[Hidro]
,[Yoga]
FROM [dbo].[SpaCentro_m20191007]
```

Generar consulta...
Examinar...

ADO.NET connection manager:
HP.db_spa_escalon.sa Nueva...

Modo de acceso a datos:
Comando SQL

Texto de comando SQL:

```
SELECT [id]
,[Sexo]
,[Ingresos]
,[PromVisit]
,[Edad]
,[Sauna]
,[Masaje]
,[Hidro]
,[Yoga]
FROM [dbo].[SpaEscalon_i20191007]
```

Generar consulta...
Examinar...

ADO.NET connection manager:
HP.db_spa_santa_tecla.sa Nueva...

Modo de acceso a datos:
Comando SQL

Texto de comando SQL:

```
SELECT [id]
,[Sexo]
,[Ingresos]
,[PromVisit]
,[Edad]
,[Sauna]
,[Masaje]
,[Hidro]
,[Yoga]
FROM [dbo].[SpaSantaTecla_m20191007]
```

Generar consulta...
Examinar...

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	inc
PromVisit	PromVisit	PromVisit	Pre
Edad	Edad	Edad	Ed
Sauna	Sauna	Sauna	Sa
Masaje	Masaje	Masaje	M
Hidro	Hidro	Hidro	Hi
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	Corr
id	id	Group by	
Sexo	Sexo	Group by	
ingresos	ingresos	Sum	
PromVisit	PromVisit	Average	
Edad	Edad	Group by	
Sauna	Sauna	Group by	
Masaje	Masaje	Group by	

Las columnas derivadas son las siguientes

sexo_string	<add as new column>	Sexo == 0 ? "Hombre" : "Mujer"	Unicode string [DT_W...	6
visita_sauna	<add as new column>	Sauna == 1 ? "Si" : "No"	Unicode string [DT_W...	2
visita_masaje	<add as new column>	Masaje == 1 ? "Si" : "No"	Unicode string [DT_W...	2
visita_hidro	<add as new column>	Hidro == 1 ? "Si" : "No"	Unicode string [DT_W...	2
visita_yoga	<add as new column>	Yoga == 1 ? "Si" : "No"	Unicode string [DT_W...	2

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.

Administrador de conexiones con Excel 1

Modo de acceso a datos:
Tabla o vista

Name of the Excel sheet:
HOMBRES\$

Excel connection manager:
Administrador de conexiones con Excel 1

Modo de acceso a datos:
Tabla o vista

Name of the Excel sheet:
MUJERES\$

DATA GENERADA PARA LOS HOMBRES

Autoguardado

ANÁLISIS SPA_DI... • Guardado

Buscar

Victor Amílcar Elias Peña

Archivo

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar







Ayuda


Acrobat


Com




A39

DATA GENERADA PARA LAS MUJERES


Autoguardado     ANALISIS_SPA_DI... • Guardado   Buscar

Victor Amilcar Elias Peña 

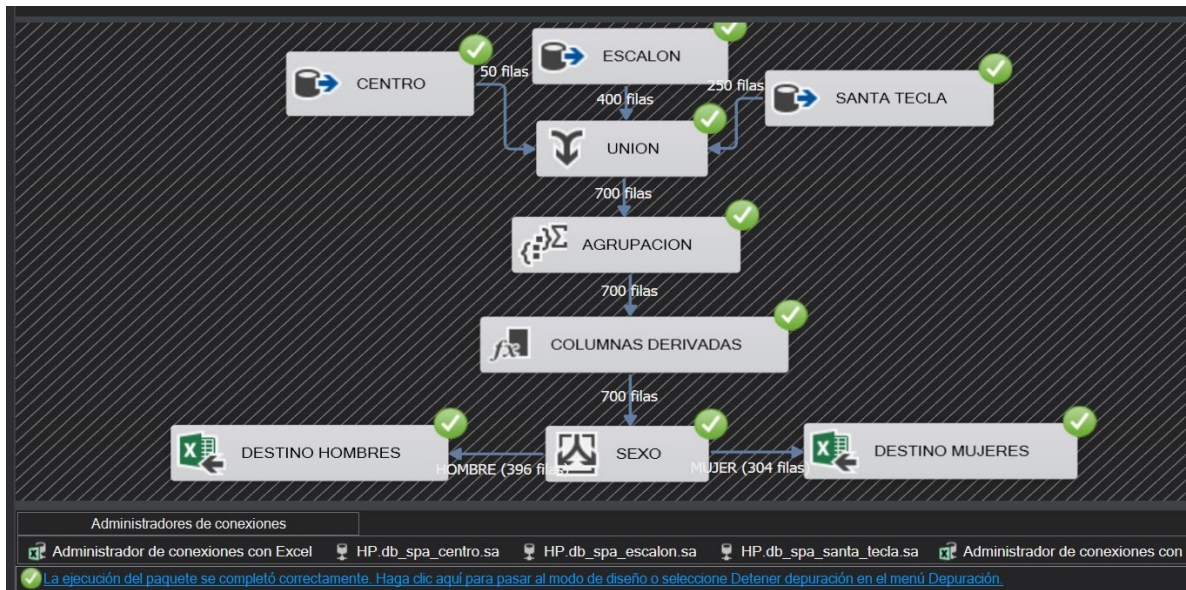
Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda Acrobat  Cor

A40    Ruggiero Brosnan

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	Sexo	Ingresos	PromVisit	Edad	Sauna	Masaje	Hidro	Yoga					
2	Roseline Kildale	Mujer	349.04998779296875	3.7799999713897705	53	No	Si	Si	Si					
3	Kelley Munnion	Mujer	1751.5699462890625	5.9499998092651367	44	No	No	No	No					
4	Reece Wungress	Mujer	2049.02001953125	4.309999942779541	22	No	No	Si	Si					
5	Gracie McDonagh	Mujer	2929.9599609375	3.8499999046325684	63	No	No	No	No					
6	Herve Laphorn	Mujer	1126.7099609375	5.0900001525878906	36	No	No	Si	No					
7	Duffy Palluschek	Mujer	2213.1201171875	6.9899997711181641	62	Si	Si	No	No					
8	Chrotoem Ruegg	Mujer	2645.35009765625	3.2300000190734863	30	Si	No	No	Si					
9	Valry Lindeberg	Mujer	1003.780029296875	4.5500001907348633	21	No	Si	No	No					
10	Antonin Rolfo	Mujer	1862.989990234375	2.9000000953674316	28	No	No	No	Si					
11	Cal Rubee	Mujer	1735.3800048828125	6.0999999046325684	24	No	Si	Si	Si					
12	Drusie Huntley	Mujer	321.23001098632812	4.4800000190734863	24	No	No	Si	Si					
13	Katheryn Dungee	Mujer	856.52001953125	1.9800000190734863	59	No	No	No	No					
14	Arturo Cornels	Mujer	2331.35009765625	1.6000000238418579	59	No	No	No	Si					
15	Britte Campe	Mujer	1828.8599853515625	1.2899999618530273	22	No	Si	No	Si					
16	Shara Le Provest	Mujer	1877.760009765625	5.1599998474121094	50	Si	Si	No	No					
17	Arnaldo Barenzen	Mujer	1868.4599609375	4.5500001907348633	61	Si	Si	Si	Si					
18	Imojean Larratt	Mujer	1806.02001953125	6.0999999046325684	60	Si	Si	No	Si					
19	Reinhold Rubberts	Mujer	2263.510009765625	1.7599999904632568	28	No	No	No	No					
20	Davon Ebbs	Mujer	698.53997802734375	5.9200000762939453	41	No	Si	No	No					
21	Pebrook Praton	Mujer	1917.1500244140625	5.8600001335144043	37	Si	Si	No	No					
22	Theda Eggle	Mujer	2147.98990234375	4.5199999809265137	50	No	No	No	No					
23	Shaine Dhillon	Mujer	2294.860107421875	5.559999942779541	47	No	No	Si	Si					
24	Bevvy Carragher	Mujer	2875.35009765625	2.3399999141693115	34	Si	Si	Si	Si					
25	Damiano Masson	Mujer	2158.22998046875	3.2400000095367432	47	Si	Si	Si	No					
26	Marci Pile	Mujer	984.32000732421875	6	62	No	No	Si	Si					

HOMBRES 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.

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 sustrayendo 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

ADO.NET connection manager:

.root Nueva...

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT CLI.IDCLIENTE, CLI.IDTIPOCLIENTE, CLI.DUI, CLI.NIT,
CLI.CODIGO_CLIENTE, CLI.NOMBRES, CLI.APELLIDOS,
CLI.NUMERO_TELEFONO, CLI.SEXO, CLI.ESTADO,
FACT.MONTO_FACTURACION, TIPO.NOMBRE_TIPO
FROM CLIENTE CLI
INNER JOIN FACTURACION FACT ON CLI.IDCLIENTE =
FACT.IDCLIENTE
INNER JOIN TIPO_CLIENTES TIPO ON TIPO.IDTIPOCLIENTE =
CLI.IDTIPOCLIENTE
```

Generar consulta...

Examinar...

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

ADO.NET connection manager:

HP.db_roaming_honduras.sa Nueva...

Modo de acceso a datos:

Comando SQL ▼

Texto de comando SQL:

```
SELECT CLI.IDCLIENTE, CLI.IDTIPOCLIENTE, CLI.DUI, CLI.NIT,
CLI.CODIGO_CLIENTE, CLI.NOMBRES, CLI.APELLIDOS,
CLI.NUMERO_TELEFONO, CLI.SEXO, CLI.ESTADO,
FACT.MONTO_FACTURACION, TIPO.NOMBRE_TIPO
FROM CLIENTE CLI
INNER JOIN FACTURACION FACT ON CLI.IDCLIENTE =
FACT.IDCLIENTE
INNER JOIN TIPO_CLIENTES TIPO ON TIPO.IDTIPOCLIENTE =
CLI.IDTIPOCLIENTE
```

Generar consulta...

Examinar...

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

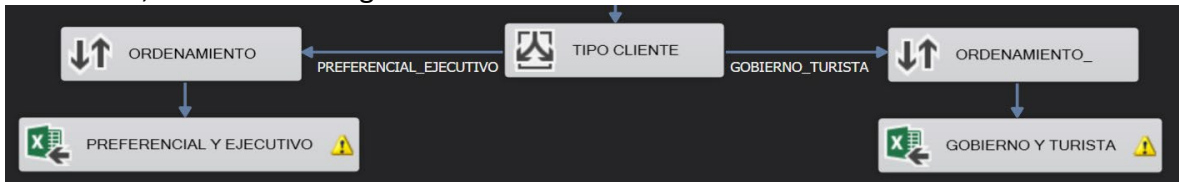
El condicional nos sirve para que determinemos los tipos de cliente. Ya que el ejercicio nos pide separar los Preferenciales y Ejecutivos de los de Gobierno y Turistas

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	

Por último, cada condición genera un Excel diferente



DATA GENERADA POR PREFERENCIAL Y EJECUTIVO

Autoguardado ✓ Preferencial y Ejecutivo • Última modificación: Ahora mismo Víctor Amílcar Elías Peña										
Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda Acrobat Comentarios Comp										
A49 SV-MA8266										
	A	B	C	D	E	F	G	H	I	J
1	CODIGO CLIENTE	DUI	NIT	NOMBRE COMPLETO	SEXO	TELEFONO	ESTADO	MONTO FACTURACION	TIPO CLIENTE	CODIGO PAIS
2	SV-RJ47679	00088387-2	1416-271278-101-0	JUAN DAVID, REYES SALAZAR	MASCULINO	7071-0453	INACTIVO	292.40	CLIENTE EJECUTIVO	sv
3	SV-NM9278	00090088-6	0614-271260-014-0	MANUEL VICENTE, NUÑEZ GRANADOS	MASCULINO	6131-7133	INACTIVO	35.20	CLIENTE EJECUTIVO	sv
4	SV-RM20831	00196268-7	0315-180165-001-5	MAURICIO ERNESTO, RAMIREZ MARTINEZ	MASCULINO	6261-0010	ACTIVO	149.20	CLIENTE EJECUTIVO	sv
5	SV-MM47693	00310459-4	0702-240253-001-0	MANUEL ANTONIO, MEJIA	MASCULINO	6235-7596	INACTIVO	189.80	CLIENTE PREFERENCIAL	sv
6	SV-OM47682	00356539-2	1217-241278-104-0	MARGARITA DE LA PAZ, OLIVARES RUIZ	FEMENINO	6660-6154	ACTIVO	560.00	CLIENTE PREFERENCIAL	sv
7	SV-CJ47665	00371900-6	0515-080248-001-9	JUAN JUVENCIO, CASTILLO MEZQUITA	MASCULINO	7851-4479	ACTIVO	171.60	CLIENTE PREFERENCIAL	sv
8	SV-PR20278	00374800-4	0821-220252-001-2	RAUL ALBERTO, PINEDA DIAZ	MASCULINO	6278-7729	ACTIVO	302.80	CLIENTE EJECUTIVO	sv
9	SV-SR47630	00424078-3	1008-050256-101-6	RAUL, SANCHEZ MARTINEZ	MASCULINO	7439-8467	ACTIVO	185.20	CLIENTE PREFERENCIAL	sv
10	SV-PA13188	00466902-6	1011-280273-101-0	ANGEL VICTOR, PINO MERINO	MASCULINO	7923-2040	ACTIVO	256.80	CLIENTE PREFERENCIAL	sv
11	SV-MA8266	00559869-2	0614-190367-101-1	ALICIA ORBELINA, MENDEZ BENITEZ	MASCULINO	7070-5856	ACTIVO	297.80	CLIENTE PREFERENCIAL	sv
12	SV-QA47695	00619610-0	0614-290358-010-0	ANA DORA ALICIA, QUINTANILLA DE ZULETA	FEMENINO	6271-6439	ACTIVO	234.20	CLIENTE PREFERENCIAL	sv
13	SV-MJ30105	00630581-4	0614-220281-103-6	JAIME RAUL, MARROQUIN MARROQUIN	MASCULINO	7976-7136	ACTIVO	260.20	CLIENTE PREFERENCIAL	sv
14	SV-UD47691	00809544-9	1403-071155-101-1	DOLORES SUYAPA, UMANZOR DE RIVERA	FEMENINO	6225-7198	ACTIVO	368.80	CLIENTE PREFERENCIAL	sv
15	SV-GS47519	00829876-2	1010-030362-001-1	SARA LORENA, GUZMAN SAMAYOA	FEMENINO	6284-6340	ACTIVO	384.20	CLIENTE EJECUTIVO	sv
16	SV-RJ47676	00881659-4	0512-030470-101-7	JACOBO, RAMOS PEREZ	MASCULINO	7165-1054	INACTIVO	184.40	CLIENTE EJECUTIVO	sv
17	SV-PG47677	00927645-4	0522-200472-101-3	GUILLERMO, POCASANGRE HUEZO	MASCULINO	7789-4991	ACTIVO	272.80	CLIENTE EJECUTIVO	sv
18	SV-CB47683	01062229-4	1109-250583-102-4	BLANCA MARIBEL, CRUZ AMAYA	FEMENINO	7886-6026	ACTIVO	335.40	CLIENTE EJECUTIVO	sv
19	SV-FR47663	01075769-1	0614-240560-013-6	ROBERTO ENRIQUE, FONG HERNANDEZ	MASCULINO	6282-3380	INACTIVO	361.80	CLIENTE PREFERENCIAL	sv
20	SV-BJ15868	01167192-9	1217-210658-004-3	JOSE ATILIO, BENITEZ PARADA	MASCULINO	6207-4477	ACTIVO	308.60	CLIENTE EJECUTIVO	sv
21	SV-LC6418	01421695-6	0614-031062-003-1	CESAR MAURICIO, LOPEZ GUZMAN	MASCULINO	6305-5568	ACTIVO	86.60	CLIENTE EJECUTIVO	sv
22	SV-MN1421	01448290-5	0210-010765-101-4	NESTOR MARIO, MARROQUIN AREVALO	MASCULINO	7740-7501	INACTIVO	212.00	CLIENTE PREFERENCIAL	sv
23	SV-GC42266	01564617-0	0511-121181-102-6	CARLOS RICARDO, GHIRINGHELLO ROSALES	FEMENINO	7247-1706	INACTIVO	214.20	CLIENTE PREFERENCIAL	sv
24	SV-CH47710	01695478-8	1324-220868-101-2	HECTOR MANUEL, CHACON ARGUETA	MASCULINO	7778-1096	ACTIVO	82.60	CLIENTE EJECUTIVO	sv
25	SV-HE47712	01771937-7	0614-020980-118-1	ERICK STEFFAN, HERNANDEZ DURAN	MASCULINO	7695-5228	ACTIVO	364.60	CLIENTE EJECUTIVO	sv
26	SV-MG2164	01804572-1	0210-250264-002-1	GABRIEL ANTONIO, MAYORGA RAMIREZ	MASCULINO	7070-5723	ACTIVO	281.20	CLIENTE PREFERENCIAL	sv

DATA GENERADA POR GOBIERNO Y TURISTA

Autoguardado ☒ Gobierno y T... Última modificación: sáb. a las 12:34 Victor Amílcar Elias Peña

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda Acrobat Comentarios Compa

A57

	A	B	C	D	E	F	G	H	I	J
	CODIGO CLIENTE	DUI	NIT	NOMBRE COMPLETO	SEXO	TELEFONO	ESTADO	MONTO FACTURACION	TIPO CLIENTE	CODIGO PAIS
1	SV-PW9858	00108338-6	0617-210164-002-0	WALTER MAURICIO, PALACIOS COLOCHO	MASCULINO	6227-2957	ACTIVO	164.60	CLIENTE GUBERNAMENTAL	sv
2	SV-NS47705	00115399-5	0604-090668-101-7	SIMEON, NAVARRETE LEONOR	MASCULINO	6124-2645	ACTIVO	199.00	CLIENTE TURISTA	sv
3	SV-BA47686	00119483-2	0614-251259-006-0	ANA IRMA, BENITEZ ARGUETA	FEMENINO	6270-5458	ACTIVO	390.80	CLIENTE TURISTA	sv
4	SV-VS5897	00141561-4	0610-091159-101-0	SALVADOR, VASQUEZ MELARA	MASCULINO	6333-5201	ACTIVO	334.40	CLIENTE GUBERNAMENTAL	sv
5	SV-RA2158	00287394-6	0210-250167-105-8	ANA LIZ, RODRIGUEZ DE TOVAR	FEMENINO	6440-5701	INACTIVO	270.20	CLIENTE GUBERNAMENTAL	sv
6	SV-EM47658	00374984-9	0619-030179-101-7	MARVIN VITELIO, ERAZO VASQUEZ	MASCULINO	7786-5956	INACTIVO	93.40	CLIENTE TURISTA	sv
7	SV-ME47714	00449568-9	1009-100279-101-7	ERIC ALEXANDER, MEJIA MOLINA	MASCULINO	7083-1594	ACTIVO	61.40	CLIENTE GUBERNAMENTAL	sv
8	SV-AE47669	00453608-7	0811-301161-101-8	ESTANISLAO, ALVARADO ALVAREZ	MASCULINO	7993-9305	ACTIVO	88.20	CLIENTE GUBERNAMENTAL	sv
9	SV-NJ26082	00481311-2	0703-250365-101-1	JOSE ANTONIO, NOLASCO LOPEZ	MASCULINO	7710-9443	ACTIVO	219.80	CLIENTE GUBERNAMENTAL	sv
10	SV-AL47672	00492838-9	0210-020382-105-2	LUIS ALONSO, ALVAREZ HERNANDEZ	MASCULINO	7663-1438	INACTIVO	301.80	CLIENTE TURISTA	sv
11	SV-SR47678	00542374-0	0210-150150-002-2	RICARDO ANTONIO, SILIEZAR SALINAS	MASCULINO	6293-0678	ACTIVO	61.20	CLIENTE GUBERNAMENTAL	sv
12	SV-JJ47659	00557486-8	0203-230383-101-7	JOSE RODOLFO, JIMENEZ PINTIN	MASCULINO	6263-4944	ACTIVO	52.00	CLIENTE GUBERNAMENTAL	sv
13	SV-RG47664	00569371-9	0614-020383-132-1	CARLOS RODOLFO, ROSALES MACHADO	MASCULINO	6566-6345	ACTIVO	267.60	CLIENTE TURISTA	sv

UNION

160 filas

AGRUPACION

80 filas

COLUMNAS DERIVADAS

80 filas

ORDENAMIENTO

38 filas

PREFERENCIAL Y EJECUTIVO

TIPO CLIENTE

PREFERENCIAL_EJECUTIVO (38 filas) GOBIERNO_TURISTA (42 filas)

ORDENAMIENTO_

42 filas

GOBIERNO Y TURISTA

Administradores de conexiones

.root Administrador de conexiones con Excel Administrador de conexiones con Excel 1 Administrador de conexiones con Excel 2 HP.db_roaming_honduras.sa

La ejecución del paquete se completó correctamente. Haga clic aquí para pasar al modo de diseño o seleccione Detener depuración en el menú Depuración.

PORCENTAJE ALCANZADO: 100%