

# PROYECTO EXPLORACION DE DATOS (EDA)

```
Select * From INFORMATION_SCHEMA.TABLES
```

00 %

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE
1	DataAnalytics	gold	customers	BASE TABLE
2	DataAnalytics	gold	products	BASE TABLE
3	DataAnalytics	gold	sales	BASE TABLE

```
Select * From INFORMATION_SCHEMA.COLUMNS
```

100 %

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION
1	DataAnalytics	gold	customers	customer_key	1
2	DataAnalytics	gold	customers	customer_id	2
3	DataAnalytics	gold	customers	customer_number	3
4	DataAnalytics	gold	customers	first_name	4
5	DataAnalytics	gold	customers	last_name	5
6	DataAnalytics	gold	customers	country	6
7	DataAnalytics	gold	customers	marital_status	7
8	DataAnalytics	gold	customers	gender	8
9	DataAnalytics	gold	customers	birthdate	9
10	DataAnalytics	gold	customers	create_date	10
11	DataAnalytics	gold	products	product_key	1
12	DataAnalytics	gold	products	product_id	2
13	DataAnalytics	gold	products	product_number	3
14	DataAnalytics	gold	products	product_name	4
15	DataAnalytics	gold	products	category_id	5
16	DataAnalytics	gold	products	category	6
17	DataAnalytics	gold	products	subcategory	7
18	DataAnalytics	gold	products	maintenance	8
19	DataAnalytics	gold	products	cost	9
20	DataAnalytics	gold	products	product_line	10

Select \* From INFORMATION\_SCHEMA.COLUMNS  
Where TABLE\_NAME = 'customers'

100 %

Results Messages

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE
1	DataAnalytics	gold	customers	customer_key	1	NULL	YES	int
2	DataAnalytics	gold	customers	customer_id	2	NULL	YES	int
3	DataAnalytics	gold	customers	customer_number	3	NULL	YES	nvarchar
4	DataAnalytics	gold	customers	first_name	4	NULL	YES	nvarchar
5	DataAnalytics	gold	customers	last_name	5	NULL	YES	nvarchar
6	DataAnalytics	gold	customers	country	6	NULL	YES	nvarchar
7	DataAnalytics	gold	customers	marital_status	7	NULL	YES	nvarchar
8	DataAnalytics	gold	customers	gender	8	NULL	YES	nvarchar
9	DataAnalytics	gold	customers	birthdate	9	NULL	YES	date
10	DataAnalytics	gold	customers	create_date	10	NULL	YES	date

Exploración de dimensión

Select Distinct country from gold.customers

00 %

Results Messages

	country
1	Germany
2	United States
3	Australia
4	United Kingdom
5	Brasil
6	Canada
7	France

-- Seleccionar todas las categorías de productos

Select Distinct category from gold.products

100 %

Results Messages

	category
1	Accessories
2	Bikes
3	Clothing
4	Components
5	Mountain
6	Road
7	Touring

```
Select Distinct category ,subcategory, product_name from gold.products  
Order By 1,2,3
```

100 %

Results Messages

	category	subcategory	product_name
1	Accessories	Bike Racks	Hitch Rack - 4-Bike
2	Accessories	Bike Stands	All-Purpose Bike Stand
3	Accessories	Bottles and Cages	Mountain Bottle Cage
4	Accessories	Bottles and Cages	Road Bottle Cage
5	Accessories	Bottles and Cages	Water Bottle - 30 oz.
6	Accessories	Cleaners	Bike Wash - Dissolver
7	Accessories	Fenders	Fender Set - Mountain
8	Accessories	Helmets	Sport-100 Helmet- Black
9	Accessories	Helmets	Sport-100 Helmet- Blue
10	Accessories	Helmets	Sport-100 Helmet- Red
11	Accessories	Hydration Packs	Hydration Pack - 70 oz.
12	Accessories	Lights	Headlights - Dual-Beam
13	Accessories	Lights	Headlights - Weatherproof
14	Accessories	Lights	Taillights - Battery-Powered
15	Accessories	Locks	Cable Lock

```
-- Explorar las columnas de fecha  
-- Encontrar las fechas del 1er y ultimo pedido
```

```
Select  
MIN(order_date) as primerPedido , MAX(order_date) as ultimoPedido  
From gold.sales
```

100 %

Results Messages

	primerPedido	ultimoPedido
1	2010-12-29	2014-01-28

```
-- Cuantos años de ventas hay disponibles
```

```
Select  
MIN(order_date) as primerPedido ,  
MAX(order_date) as ultimoPedido ,  
DATEDIFF(YEAR,MIN(order_date), MAX(order_date) ) as rangoPedidosenAnios  
From gold.sales
```

100 %

Results Messages

	primerPedido	ultimoPedido	rangoPedidosenAnios
1	2010-12-29	2014-01-28	4

```
└─ Select
    MIN(order_date) as primerPedido ,
    MAX(order_date) as ultimoPedido ,
    DATEDIFF(YEAR,MIN(order_date), MAX(order_date) ) as rangoPedidosenAnios,
    DATEDIFF(MONTH,MIN(order_date), MAX(order_date) ) as rangoPedidosenMeses
  From gold.sales
```

0 %

Results Messages

	primerPedido	ultimoPedido	rangoPedidosenAnios	rangoPedidosenMeses
1	2010-12-29	2014-01-28	4	37

```
-- Encontra al cliente mas joven y al mas viejo
```

```
└─ Select
    MIN(birthdate) as clienteViejo,
    MAX(birthdate) as clienteJoven
  From gold.customers
```

100 %

Results Messages

	clienteViejo	clienteJoven
1	1916-02-10	1986-06-25

```
-- Encontra al cliente mas joven y al mas viejo
```

```
└─ Select
    DATEDIFF(YEAR,MIN(birthdate),GETDATE() ) as EdadViejo,
    DATEDIFF(YEAR,MAX(birthdate),GETDATE() ) as EdadJoven
  From gold.customers
```

100 %

Results Messages

	EdadViejo	EdadJoven
1	109	39

```
-- Encontrar las ventas Totales  
Select  
    SUM(price) as ventasTotales  
From gold.sales
```

100 %

Results Messages

	ventasTotales
1	29355710

```
-- Encontrar cuantos articulos se venden  
Select  
    SUM(quantity) as cantidadTotal  
From gold.sales
```

100 %

Results Messages

	cantidadArticulosVendidos
1	60423

```
-- Encontrar el precio de venta promedio  
Select  
    AVG(price) as precioPromedio  
From gold.sales
```

100 %

Results Messages

	precioPromedio
1	486

```
-- Encontrar el total de pedidos  
Select  
    COUNT(Distinct order_number) as totalPedidos  
From gold.sales
```

100 %

Results Messages

	totalPedidos
1	27659

```
-- Encontrar el total de numeros de productos  
Select COUNT(product_id) as totalProductos from gold.products  
Select COUNT(Distinct product_id) as totalProductos from gold.products
```

10 %

Results Messages

	totalProductos
1	295

```
-- Encontrar el total de numeros de clientes  
Select COUNT(Distinct customer_id) as totalClientes from gold.customers
```

100 %

Results Messages

	totalClientes
1	18484

```
-- Encontrar numero total de clientes que han realizado un pedido  
Select COUNT(Distinct customer_key) as totalClientes from gold.customers  
Select COUNT(Distinct customer_key) as totalClientes from gold.sales
```

100 %

Results Messages

	totalClientes
1	18484

```
-- ***** Generar un Reporte que muestren todas las metricas de negocio  
Select 'ventas Totales' as nombreMedida , SUM(sales_amount) as valorMedida from gold.sales  
Union All  
Select 'cantidad Total' as nombreMedida , SUM(quantity) as valorMedida from gold.sales  
Union All  
Select ' precio Promedio' as nombreMedida , AVG(price) as valorMedida from gold.sales  
Union All  
Select ' Total de pedidos' as nombreMedida , COUNT(Distinct order_number) as valorMedida from gold.sales  
Union All  
Select ' Total de Productos' as nombreMedida , COUNT(Distinct product_key) as valorMedida from gold.products  
Union All  
Select ' Total de Clientes' as nombreMedida , COUNT(Distinct customer_key) as valorMedida from gold.customers
```

100 %

Results Messages

	nombreMedida	valorMedida
1	ventas Totales	29356250
2	cantidad Total	60423
3	precio Promedio	486
4	Total de pedidos	27659
5	Total de Productos	295
6	Total de Clientes	18484

```
-- Encontrar el total de clientes por pais
Select country,
       COUNT(customer_key) totalCliente
From gold.customers
Group By country
Order By totalCliente desc
```

100 %

Results Messages

	country	totalCliente
1	United States	7482
2	Australia	3591
3	United Kingdom	1913
4	France	1810
5	Germany	1780
6	Canada	1571
7	Brasil	337

```
-- Encontrar el total de cliente por genero
Select gender,
       COUNT(customer_key) totalCliente
From gold.customers
Group By gender
Order By totalCliente desc
```

100 %

Results Messages

	gender	totalCliente
1	Male	9341
2	Female	9128
3	n/a	15

```
-- Encontrar el total de productos por categoria
```

```
Select category ,
       COUNT(product_key) totalProductos
From gold.products
Group By category
Order By totalProductos desc
```

100 %

Results Messages

	category	totalProductos
1	Components	127
2	Bikes	97
3	Clothing	35
4	Accessories	29
5	Mountain	3
6	Road	3
7	Touring	1

```
-- cual es el costo medio en cada categoria  
  
Select category ,  
       AVG(cost) costoMedio  
From gold.products  
Group By category  
Order By costoMedio desc
```

100 % ▾

Results Messages

	category	costoMedio
1	Bikes	949
2	Components	264
3	Touring	36
4	Mountain	27
5	Road	27
6	Clothing	24
7	Accessories	13

```
-- cual es el total de ingreso generado en cada categoria
```

```
Select  
p.category,  
SUM(s.sales_amount) as totalIngreso  
From gold.sales as s  
Left Join gold.products as p ON s.product_key = p.product_key  
Group By p.category  
Order By totalIngreso desc
```

100 % ▾

Results Messages

	category	totalIngreso
1	Bikes	28316272
2	Accessories	700262
3	Clothing	339716

```
-- cual es el ingreso total generado en cada cliente

Select
c.customer_key ,
c.first_name,
c.last_name ,
SUM(sales_amount) as totalIngreso
From gold.sales as s
Left Join gold.customers as c ON s.customer_key = c.customer_key
Group By c.customer_key, c.first_name, c.last_name
Order By totalIngreso desc
```

100 %

Results Messages

	customer_key	first_name	last_name	totalIngreso
1	1302	Nichole	Nara	13294
2	1133	Kaitlyn	Henderson	13294
3	1309	Margaret	He	13268
4	1132	Randall	Dominguez	13265
5	1301	Adriana	Gonzalez	13242
6	1322	Rosa	Hu	13215
7	1125	Brandi	Gill	13195
8	1308	Brad	She	13172
9	1297	Francisco	Sara	13164
10	434	Maurice	Shan	12914
11	440	Janet	Munoz	12488
12	242	Lisa	Cai	11468
13	418	Lacey	Zheng	11248
14	421	Jordan	Tumer	11200
15	243	Larry	Munoz	11067

```
-- Cual es la distribucion de los articulos vendidos en los diferentes paises

Select c.country,
COUNT(s.quantity) as TotalVendidos
from gold.sales as s
Left Join gold.customers as c ON s.customer_key = c.customer_key
Group By c.country
Order By TotalVendidos desc
```

100 %

Results Messages

	country	TotalVendidos
1	United States	20473
2	Australia	13345
3	Canada	7620
4	United Kingdom	6906
5	Germany	5625
6	France	5558
7	Brasil	871

```
-- Cuales son los 5 productos que generan mayores ingresos

Select Top 5 p.product_name,
       SUM(s.sales_amount) as totalIngresos
  From gold.sales as s
 Left Join gold.products as p ON s.product_key = p.product_key
 Group By p.product_name
 Order By totalIngresos desc
```

100 %

Results Messages

	product_name	totalIngresos
1	Mountain-200 Black- 46	1373454
2	Mountain-200 Black- 42	1363128
3	Mountain-200 Silver- 38	1339394
4	Mountain-200 Silver- 46	1301029
5	Mountain-200 Black- 38	1294854

```
-- Cuales son los 5 productos que generan mayores ingresos
-- con Windows Functions
Select *
  From (
    Select p.product_name,
           SUM(s.sales_amount) as totalIngresos,
           ROW_NUMBER() Over(Order By SUM(s.sales_amount) desc) as Rank_productos
      From gold.sales as s
     Left Join gold.products as p ON s.product_key = p.product_key
    Group By p.product_name
  ) t
 Where Rank_productos <= 5
```

100 %

Results Messages

	product_name	totalIngresos	Rank_productos
1	Mountain-200 Black- 46	1373454	1
2	Mountain-200 Black- 42	1363128	2
3	Mountain-200 Silver- 38	1339394	3
4	Mountain-200 Silver- 46	1301029	4
5	Mountain-200 Black- 38	1294854	5

```
-- Cuales son los 5 productos que generan menor ingresos

Select Top 5 p.product_name,
       SUM(s.sales_amount) as totalIngresos
  From gold.sales as s
 Left Join gold.products as p ON s.product_key = p.product_key
 Group By p.product_name
 Order By totalIngresos
```

100 %

Results Messages

	product_name	totalIngresos
1	Racing Socks- L	2430
2	Racing Socks- M	2682
3	Patch Kit/8 Patches	6382
4	Bike Wash - Dissolver	7272
5	Touring Tire Tube	7440

```
-- Encontrar el top 10 de clientes que generaron mayor ingreso
```

```
Select Top 10
  c.customer_key ,
  c.first_name,
  c.last_name ,
  SUM(sales_amount) as totalIngreso
 From gold.sales as s
 Left Join gold.customers as c ON s.customer_key = c.customer_key
 Group By c.customer_key, c.first_name, c.last_name
 Order By totalIngreso desc
```

100 %

Results Messages

	customer_key	first_name	last_name	totalIngreso
1	1302	Nichole	Nara	13294
2	1133	Kaitlyn	Henderson	13294
3	1309	Margaret	He	13268
4	1132	Randall	Dominguez	13265
5	1301	Adriana	Gonzalez	13242
6	1322	Rosa	Hu	13215
7	1125	Brandi	Gill	13195
8	1308	Brad	She	13172
9	1297	Francisco	Sara	13164
10	434	Maurice	Shan	12914

```
-- Encontrar el top 3 de clientes que generaron menos pedidos realizados

Select Top 3
c.customer_key ,
c.first_name,
c.last_name ,
COUNT(Distinct s.order_number) as totalPedidos
From gold.sales as s
Left Join gold.customers as c ON s.customer_key = c.customer_key
Group By c.customer_key, c.first_name, c.last_name
Order By totalPedidos |
```

100 %

Results Messages

	customer_key	first_name	last_name	totalPedidos
1	21	Jordan	King	1
2	37	Jennifer	Russell	1
3	17	Wyatt	Hill	1