



**IPN**  
*Instituto Politécnico Nacional*



**UPIIZ**  
*Unidad Profesional Interdisciplinaria de Ingeniería  
Campus Zacatecas*

**Ingeniería en:**  
*Sistemas Computacionales*

**Materia:**  
*Data Mining*

**Docente:**  
*Héctor Alejandro Acuña Cid*

**Actividad:**  
*“Práctica 01. Operaciones de Cubos OLAP”*

**Alumno:**  
*Alejandro Tamayo Castro*

**Grupo:**  
*7CM1*

***Zacatecas, Zac. septiembre 20 de 2023***

# Código

Para crear el data set de 500 se implemento un llenado aleatorio de datos partiendo de diccionarios establecidos, en este caso se uso un generador aleatorio de fechas de pertenecientes al año 2021, 5 productos, 5 categorias, 5 municipios y 5 vendedores.

```
# Crear listas con valores posibles para cada columna
fechas = [f"2021-{str(month).zfill(2)}" for month in range(1, 13)] * 42
productos = ['Manzana', 'Leche', 'Pan', 'Platano', 'Jugo de Naranja']
categorias = ['Frutas', 'Lácteos', 'Panadería', 'Frutas', 'Bebidas']
municipios = ['Zacatecas', 'Guadalupe', 'Fresnillo', 'Tlaltenango', 'Jerez']
vendedores = ['Ana', 'Héctor', 'Francisco', 'Nora', 'Claudio']
```

Se genero el dataset de 500 filas y se guardo en un DataFrame para su uso en el código.

Adicionalmente se genero un segundo data set de 500 elementos pero en este caso para el inventario que se usará en el ejercicio 6. Se usaro nuevamente el generador de fechas y los mismos productos anteriores.

```
# Crear listas con valores posibles para cada columna en el inventario
fechas_inventario = [f"2021-{str(month).zfill(2)}" for month in range(1, 13)] * 42
productos_inventario = ['Manzana', 'Leche', 'Pan', 'Platano', 'Jugo de Naranja']
```

Para realizar la practica se realizó el siguiente código.

```

# 1. Drill-Down
dd = df.groupby(['Fecha', 'Categoría', 'Producto'])['Ventas'].sum().reset_index()
print("1. Drill-Down Result:")
print(dd)

# 2. Roll-Up
ventas_region = df.groupby('Municipio')['Ventas'].sum().reset_index()
ventas_totales = ventas_region['Ventas'].sum()
print("\n2. Roll-Up Result:")
print(ventas_region)
print(f"Ventas Totales: {ventas_totales}")

# 3. Slice and Dice
slice_result = df[(df['Fecha'].str.startswith('2021-01')) & (df['Municipio'] == 'Zacatecas')]
dice_result = slice_result[slice_result['Categoría'] == 'Frutas']
print("\n3. Slice and Dice Result:")
print(dice_result)

# 4. Pivot (o Rotate)
pivot_result = pd.pivot_table(df, values='Ventas', index='Vendedor', columns='Fecha', aggfunc=sum, fill_value=0)
print("\n4. Pivot Result:")
print(pivot_result)

# 5. Drill-Through
total_ventas_frutas_2021_01 = df[(df['Fecha'].str.startswith('2021-01')) & (df['Categoría'] == 'Frutas')]['Ventas'].sum()
drill_through_result = df[(df['Fecha'].str.startswith('2021-01')) & (df['Categoría'] == 'Frutas')]
print("\n5. Drill-Through Result:")
print(drill_through_result)

# 6. Drill-Across (Simulación ya que no tenemos un DataFrame real de inventario)
# El df_inventario es el otro conjunto de datos.
drill_across_result = pd.merge(df, df_inventario, on=['Producto', 'Fecha'])
print("\n6. Drill-Across:")
print(drill_across_result)

# 7. Consolidación
consolidation_result = df.groupby('Vendedor').agg({'Ventas': ['sum', 'mean', 'max'], 'UnidadesVendidas': ['sum', 'mean', 'max']}).reset_index()
print("\n7. Consolidación Result:")
print(consolidation_result)

```

Para el primer ejercicio se usa una agrupación de Fecha, Categoría y Producto donde se hace la suma de las ventas. El segundo ejercicio agrupa por municipio y suma las ventas para cada uno, finalmente se suman todas las ventas de todos los municipios para las ventas totales. El tercer código busca por fecha de Enero 2021 y en el municipio de Zacatecas para finalmente buscar la categoría Frutas. El ejercicio 4 que pide hacer un pivot, se hace mediante Vendedor y Fecha para sumar cada una de las ventas. El ejercicio 5 busca las ventas totales de fritas de Enero 2021 y realiza el drill para buscar la categoría de frutas. En el ejercicio 6 se hace uso del segundo dataframe de inventario, en el se juntan los productos y fechas para desplegar el resultado. Para finalizar la consolidación se agrupan los vendedores y se hace la suma, promedio y el máximo de las ventas y unidades vendidas. Obteniendo de resultados lo siguiente.

1. Drill-Down Result:									
	Fecha	Categoría	Producto	Ventas					
0	2021-01	Bebidas	Jugo de Naranja	1072					
1	2021-01	Frutas	Manzana	1108					
2	2021-01	Frutas	Platano	895					
3	2021-01	Lácteos	Leche	735					
4	2021-01	Panadería	Pan	690					
5	2021-02	Bebidas	Jugo de Naranja	1820					
6	2021-02	Frutas	Manzana	935					
7	2021-02	Frutas	Platano	1731					
8	2021-02	Lácteos	Leche	552					
9	2021-02	Panadería	Pan	1749					
10	2021-03	Bebidas	Jugo de Naranja	450					
11	2021-03	Frutas	Manzana	1115					
12	2021-03	Frutas	Platano	883					
13	2021-03	Lácteos	Leche	1016					
14	2021-03	Panadería	Pan	854					
15	2021-04	Bebidas	Jugo de Naranja	829					
16	2021-04	Frutas	Manzana	1249					
17	2021-04	Frutas	Platano	615					
18	2021-04	Lácteos	Leche	1262					
19	2021-04	Panadería	Pan	1381					
20	2021-05	Bebidas	Jugo de Naranja	1076					
21	2021-05	Frutas	Manzana	1454					
22	2021-05	Frutas	Platano	866					
23	2021-05	Lácteos	Leche	1134					
24	2021-05	Panadería	Pan	1064					
25	2021-06	Bebidas	Jugo de Naranja	925					
26	2021-06	Frutas	Manzana	1133					
27	2021-06	Frutas	Platano	604					
28	2021-06	Lácteos	Leche	711					
29	2021-06	Panadería	Pan	1293					
30	2021-07	Bebidas	Jugo de Naranja	669					
31	2021-07	Frutas	Manzana	567					
32	2021-07	Frutas	Platano	816					
33	2021-07	Lácteos	Leche	905					
34	2021-07	Panadería	Pan	485					
35	2021-08	Bebidas	Jugo de Naranja	718					
36	2021-08	Frutas	Manzana	652					
37	2021-08	Frutas	Platano	610					
37	2021-08	Frutas	Platano	610					
38	2021-08	Lácteos	Leche	1326					
39	2021-08	Panadería	Pan	1096					
40	2021-09	Bebidas	Jugo de Naranja	1024					
41	2021-09	Frutas	Manzana	662					
42	2021-09	Frutas	Platano	814					
43	2021-09	Lácteos	Leche	1016					
44	2021-09	Panadería	Pan	1283					
45	2021-10	Bebidas	Jugo de Naranja	799					
46	2021-10	Frutas	Manzana	1557					
47	2021-10	Frutas	Platano	895					
48	2021-10	Lácteos	Leche	1227					
49	2021-10	Panadería	Pan	885					
50	2021-11	Bebidas	Jugo de Naranja	232					
51	2021-11	Frutas	Manzana	920					
52	2021-11	Frutas	Platano	1739					
53	2021-11	Lácteos	Leche	1839					
54	2021-11	Panadería	Pan	1431					
55	2021-12	Bebidas	Jugo de Naranja	736					
56	2021-12	Frutas	Manzana	1821					
57	2021-12	Frutas	Platano	1116					
58	2021-12	Lácteos	Leche	1194					
59	2021-12	Panadería	Pan	1404					
2. Roll-Up Result:					Municipio	Ventas			
0	Fresnillo	14219							
1	Guadalupe	12168							
2	Jerez	12682							
3	Tlaltenango	10572							
4	Zacatecas	11998							
					Ventas Totales:	61639			

3. Slice and Dice Result:							
	Fecha	Producto	Categoría	Municipio	Vendedor	Ventas	UnidadesVendidas
134	2021-01	Platano	Frutas	Zacatecas	Héctor	170	26
188	2021-01	Platano	Frutas	Zacatecas	Ana	59	27
233	2021-01	Manzana	Frutas	Zacatecas	Héctor	121	91
253	2021-01	Platano	Frutas	Zacatecas	Ana	134	79

4. Pivot Result:												
Fecha	2021-01	2021-02	2021-03	2021-04	2021-05	2021-06	2021-07	2021-08	2021-09	2021-10	2021-11	2021-12
Vendedor												
Ana	938	2147	915	845	1162	909	925	762	678	1341	1185	1119
Claudio	510	1399	785	1114	1457	730	353	735	936	936	1193	885
Francisco	1166	1188	819	823	1456	1267	766	710	1150	556	1264	1981
Héctor	1085	1439	727	855	882	802	932	794	855	1169	1635	1516
Nora	801	614	1072	1699	637	958	466	1401	1180	1361	884	770
5. Drill-Through Result:												
Fecha	Producto	Categoría	Municipio	Vendedor	Ventas	UnidadesVendidas						
1	2021-01	Manzana	Frutas	Guadalupe	Francisco	198	29					
19	2021-01	Platano	Frutas	Guadalupe	Héctor	194	62					
76	2021-01	Platano	Frutas	Tlaltenango	Nora	87	74					
112	2021-01	Manzana	Frutas	Guadalupe	Nora	132	86					
134	2021-01	Platano	Frutas	Zacatecas	Héctor	170	26					
139	2021-01	Manzana	Frutas	Fresnillo	Claudio	114	55					
188	2021-01	Platano	Frutas	Zacatecas	Ana	59	27					
215	2021-01	Manzana	Frutas	Guadalupe	Héctor	120	68					
233	2021-01	Manzana	Frutas	Zacatecas	Héctor	121	91					
253	2021-01	Platano	Frutas	Zacatecas	Ana	134	79					
309	2021-01	Platano	Frutas	Guadalupe	Francisco	183	84					
319	2021-01	Manzana	Frutas	Jerez	Francisco	111	92					
338	2021-01	Manzana	Frutas	Guadalupe	Francisco	114	94					
415	2021-01	Manzana	Frutas	Jerez	Nora	198	68					
435	2021-01	Platano	Frutas	Fresnillo	Héctor	68	26					

6. Drill-Across:

	Fecha	Producto	Categoría	Municipio	Vendedor	Ventas	UnidadesVendidas	Stock
0	2021-05	Pan	Panadería	Jerez	Ana	65	47	472
1	2021-05	Pan	Panadería	Jerez	Ana	65	47	458
2	2021-05	Pan	Panadería	Jerez	Ana	65	47	270
3	2021-05	Pan	Panadería	Jerez	Ana	65	47	160
4	2021-05	Pan	Panadería	Jerez	Ana	65	47	392
...	...	...	...	...	...	...	...	...
4152	2021-10	Pan	Panadería	Zacatecas	Nora	81	64	292
4153	2021-10	Pan	Panadería	Zacatecas	Nora	81	64	438
4154	2021-10	Pan	Panadería	Zacatecas	Nora	81	64	466
4155	2021-10	Pan	Panadería	Zacatecas	Nora	81	64	428
4156	2021-10	Pan	Panadería	Zacatecas	Nora	81	64	265

[4157 rows x 8 columns]

7. Consolidación Result:

	Vendedor	Ventas				UnidadesVendidas		
		sum	mean	max		sum	mean	max
0	Ana	12926	127.980198	200		6019	59.594059	100
1	Claudio	11033	113.742268	200		5876	60.577320	100
2	Francisco	13146	128.882353	200		5599	54.892157	100
3	Héctor	12691	122.028846	198		6226	59.865385	99
4	Nora	11843	123.364583	199		5575	58.072917	100

- ¿Qué insights puedes obtener de cada operación?

Proporcionan operaciones importantes que incluyen desglose, resumen y filtrado de datos para comprender patrones, rendimiento y detalles específicos de ventas, ayudando en la toma de decisiones y la identificación de áreas clave de enfoque.

- ¿Cómo podrían estas insights beneficiar a la empresa?

Tomar decisiones para operar estratégicamente para modificar y administrar su inventario, estrategias de ventas e identificar oportunidades de crecimiento.

- ¿Qué otras métricas o dimensiones serían útiles para analizar?
  - Costos de adquisición de productos.
  - Margen de beneficio por producto.
  - Retorno de inversión publicitaria.
  - Análisis de la competencia en cada municipio.
  - Datos demográficos de los clientes.
  - Análisis de tendencias de precios en el mercado.
  - Segmentación de clientes y análisis de comportamiento de compra.