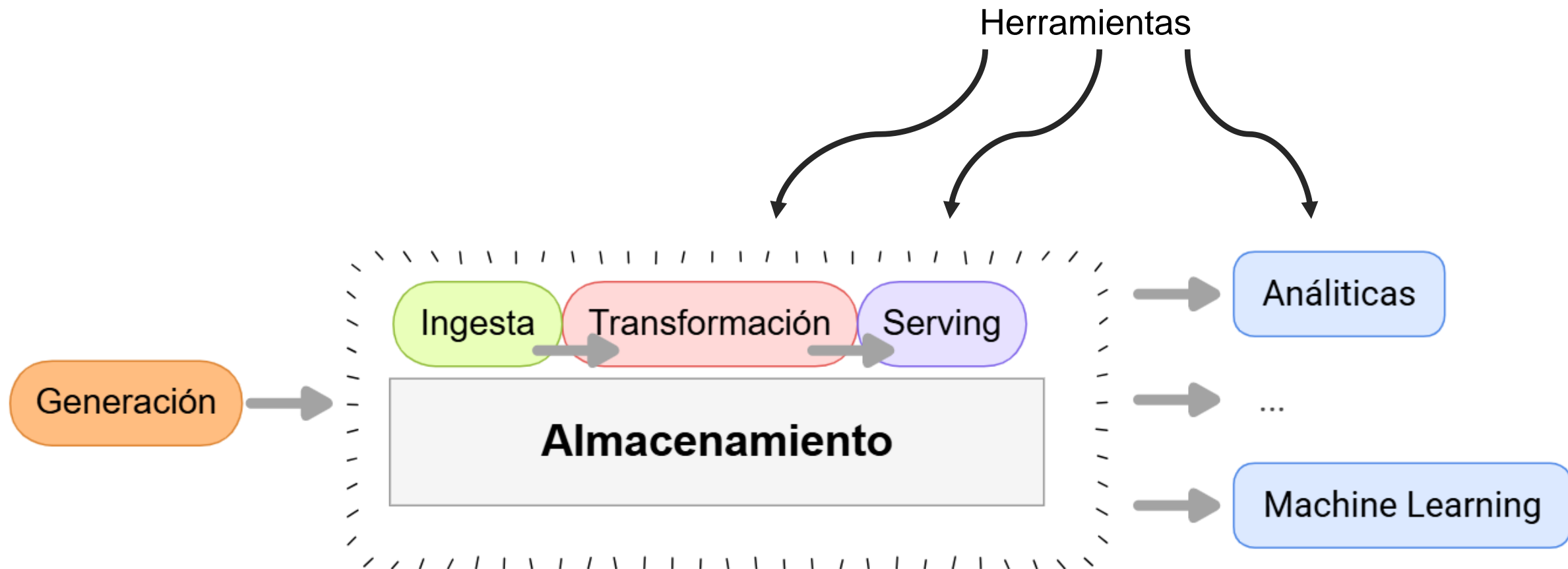


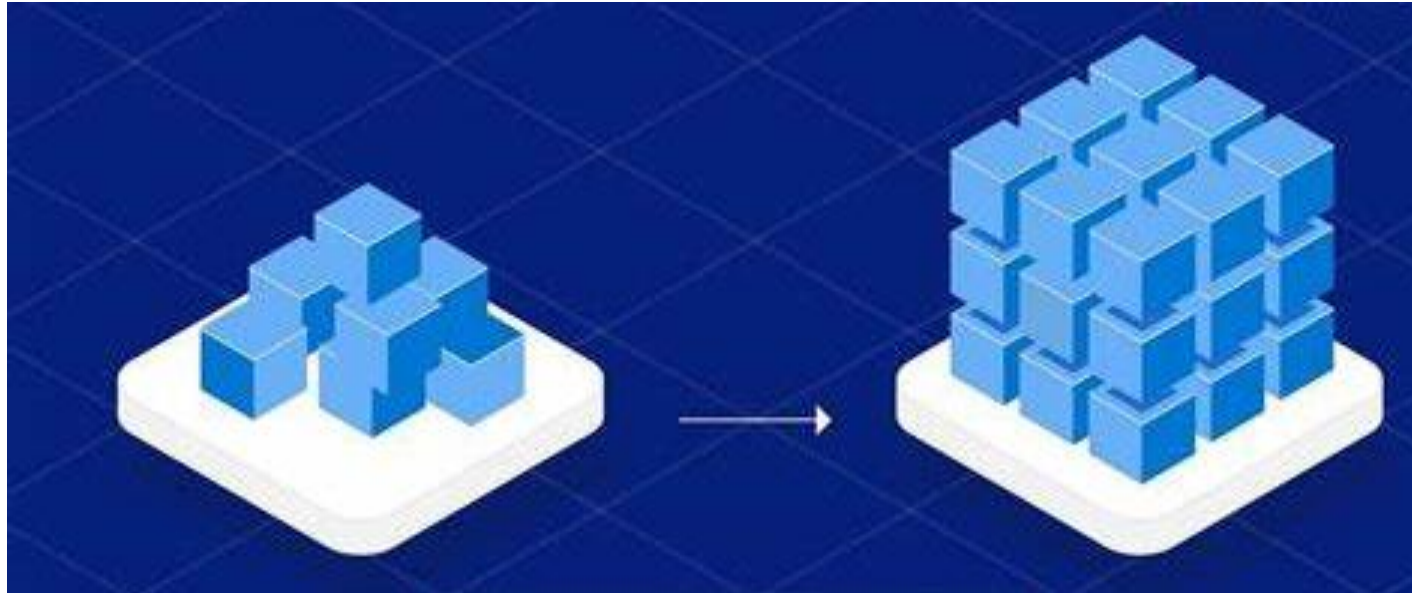
# Tecnologías de la Ingeniería de Datos Transformación y Servicio



---

Seguridad	Gestión de Datos	DataOps	Arquitectura de Datos	Orquestación	Ingeniería de Software
-----------	------------------	---------	-----------------------	--------------	------------------------

# Transformación



# Transformación



# Transformación: Limpieza

ANTES:				DESPUÉS:		
ID	Producto	Valor		ID	Producto	Valor
1	Laptop	\$1000		1	Laptop	\$1000
1	Laptop	\$1000	→	2	Mouse	\$20
2	Mouse	\$20				

# Transformación: Limpieza

ANTES:				DESPUÉS:		
ID	Producto	Valor		ID	Producto	Valor
1	Laptop	\$1000		1	Laptop	\$1000
1	Laptop	\$1000	→	2	Mouse	\$20
2	Mouse	\$20				

ANTES:			DESPUÉS:	
Paciente	Fecha_Nacimiento		Paciente	Fecha_Nacimiento
A	1990-05-15		A	1990-05-15
B	2025-13-02	→	B	NaN (fecha futura/mes inválido)
C	15/07/1985		C	1985-07-15 (formato apropiado)

# Transformación: Limpieza

ANTES:				DESPUÉS:		
ID	Producto	Valor		ID	Producto	Valor
1	Laptop	\$1000		1	Laptop	\$1000
1	Laptop	\$1000	→	2	Mouse	\$20
2	Mouse	\$20				

ANTES:			DESPUÉS:	
Paciente	Fecha_Nacimiento		Paciente	Fecha_Nacimiento
A	1990-05-15		A	1990-05-15
B	2025-13-02	→	B	NaN (fecha futura/mes inválido)
C	15/07/1985		C	1985-07-15 (formato apropiado)

Datos: [22, 23, 21, 25, 19, 28, 31, -273, 24]  
Outlier: -273 (posiblemente error de sensor)

# Transformación: Limpieza

ANTES:

ID	Producto	Valor
1	Laptop	\$1000
1	Laptop	\$1000
2	Mouse	\$20

→

DESPUÉS:

ID	Producto	Valor
1	Laptop	\$1000
2	Mouse	\$20

ANTES:

Paciente	Fecha_Nacimiento
A	1990-05-15
B	2025-13-02
C	15/07/1985

→

DESPUÉS:

Paciente	Fecha_Nacimiento
A	1990-05-15
B	NaN (fecha futura/mes inválido)
C	1985-07-15 (formato apropiado)

Datos: [22, 23, 21, 25, 19, 28, 31, -273, 24]

Outlier: -273 (posiblemente error de sensor)

ANTES:

Producto	Precio
X	\$120.50
Y	95,99€
Z	"1,200"

→

DESPUÉS:

Producto	Precio
X	120.50
Y	95.99
Z	1200.00



# Transformación: Normalización

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

# Transformación: Normalización

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

ANTES (Raw values):

Sensor\_A: [200, 400, 600]

Sensor\_B: [10, 20, 30] →

(Rango A: 200-600)

(Rango B: 10-30)

DESPUÉS (Normalizado):

Sensor\_A: [0.0, 0.5, 1.0]

Sensor\_B: [0.0, 0.5, 1.0]

(Ambos en [0,1])

# Transformación: Estandarización

$$Z = \frac{X - \mu}{\sigma}$$

# Transformación: Estandarización

$$Z = \frac{X - \mu}{\sigma}$$

ANTES (Puntajes crudos):

FICO: [650, 720, 580]

Vantage: [550, 600, 450] →

(mu=650, sigma=62.5)

(mu=550, sigma=44.7)

DESPUÉS (Z-scores):

FICO: [0.23, 1.12, -1.35]

Vantage: [0.23, 1.12, -1.35]

(Comparables directamente)

# Transformación: Encoding

Product Name	Price (USD)	Rating (out of 10)	In Stock	Quality Level	Brand
iPhone 14	799	9	Yes	Excellent	Apple
Samsung S23	899	8.5	No	Good	Samsung
Google Pixel	699	8.2	Yes	Fair	Google
OnePlus 10	599	8	Yes	Excellent	OnePlus
Nokia 2.3	149	6.5	No	Poor	Nokia

# Transformación: Encoding

Binario

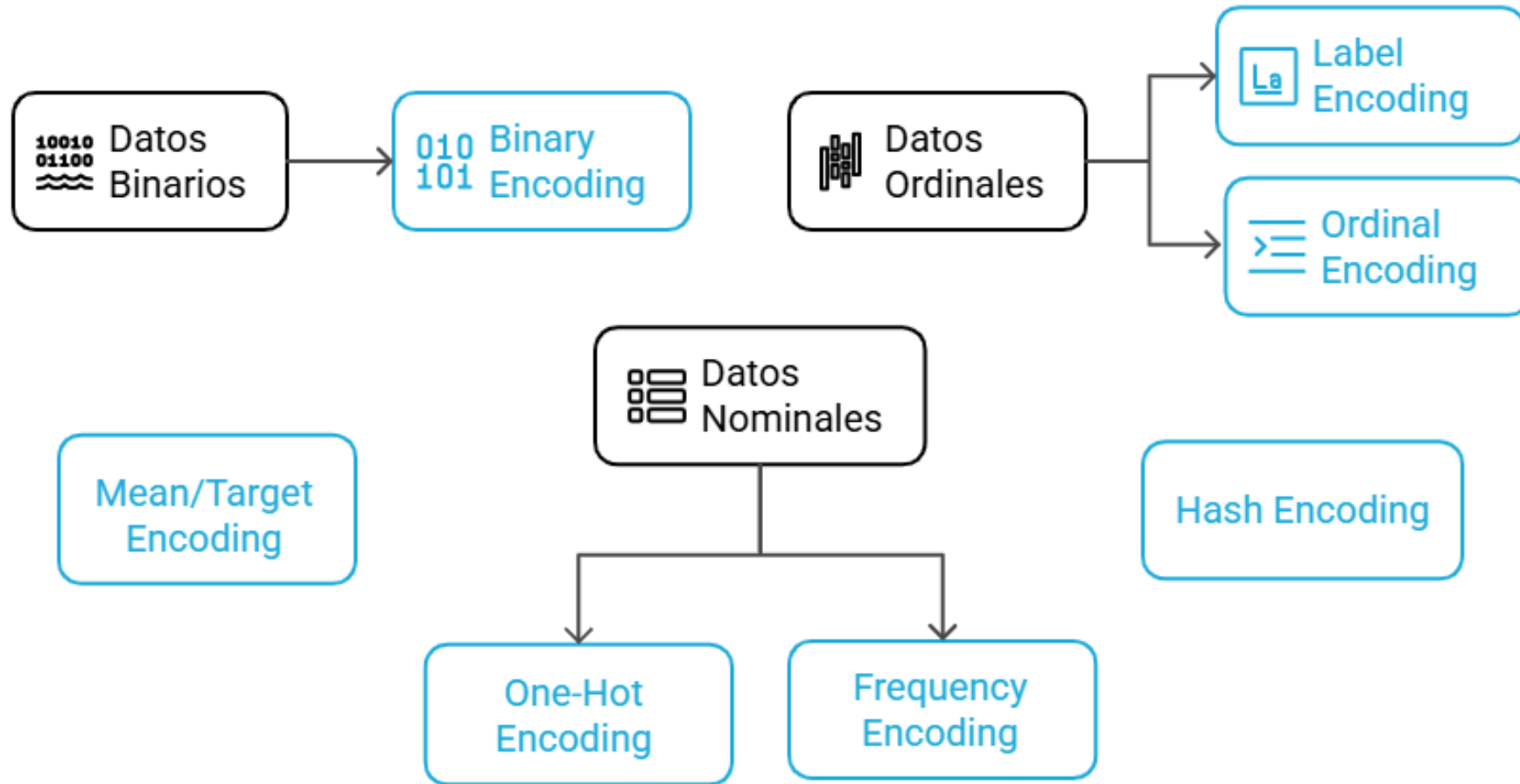
Ordinal

Nominal



Product Name	Price (USD)	Rating (out of 10)	In Stock	Quality Level	Brand
iPhone 14	799	9	Yes	Excellent	Apple
Samsung S23	899	8.5	No	Good	Samsung
Google Pixel	699	8.2	Yes	Fair	Google
OnePlus 10	599	8	Yes	Excellent	OnePlus
Nokia 2.3	149	6.5	No	Poor	Nokia

# Transformación: Encoding



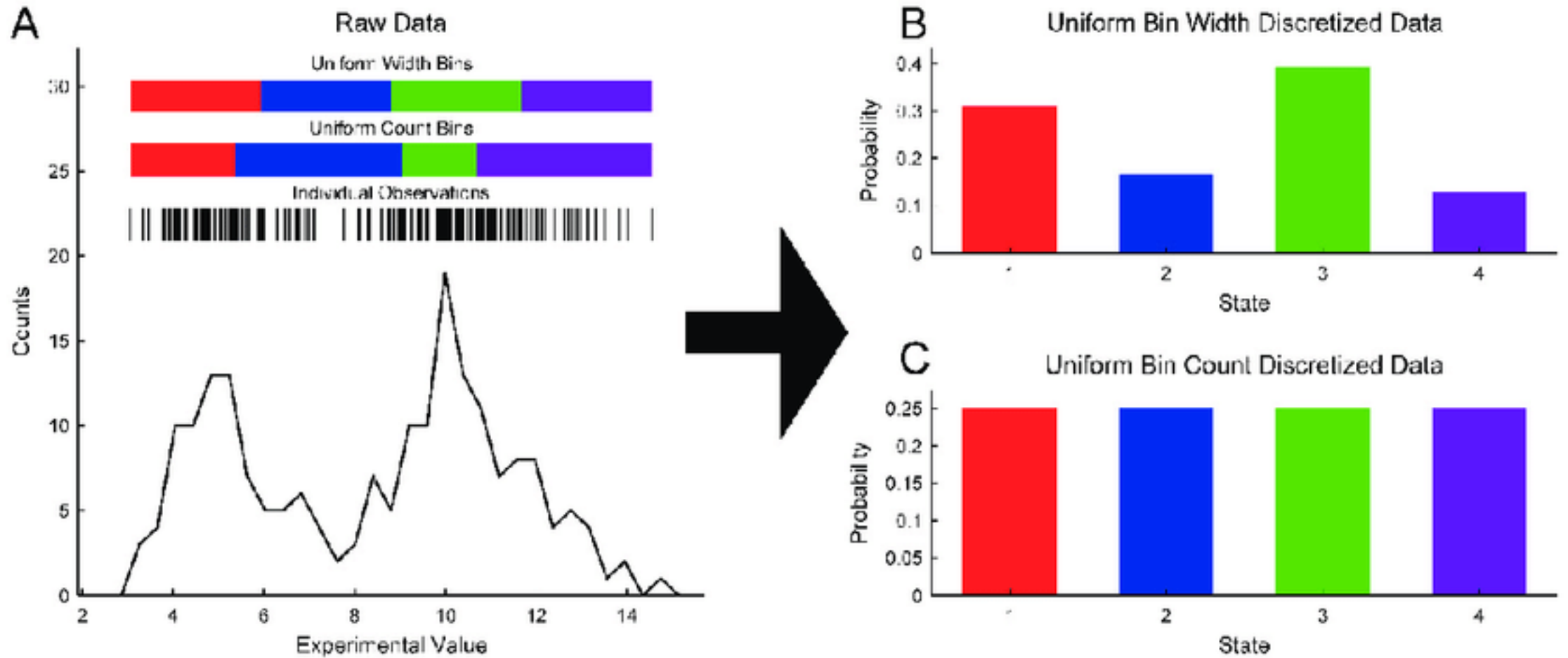
# Transformación: Encoding

ANTES:		DESPUÉS (One-Hot):		
Color		Rojo	Verde	Azul
Rojo	→	1	0	0
Verde		0	1	0
Azul		0	0	1

ANTES:		DESPUÉS:
Talla		Talla_Encoded
S	→	0
M		1
L		2



# Transformación: Discretizar



# Transformación: Discretizar

ANTES (Edad):		DESPUÉS (Bins):
[15,		["0-18",
25,	→	"19-35",
70,		"60+",
32]		"19-35"]

# Transformación

Imputación

Generación de nuevos atributos

Procesamiento de texto



# Transformación: ETL y ELT en el Procesamiento de Datos



# Transformación:

## ETL y ELT en el Procesamiento de Datos

Transformación  
en Tránsito



**ETL**

Validación estricta y  
transformaciones  
complejas

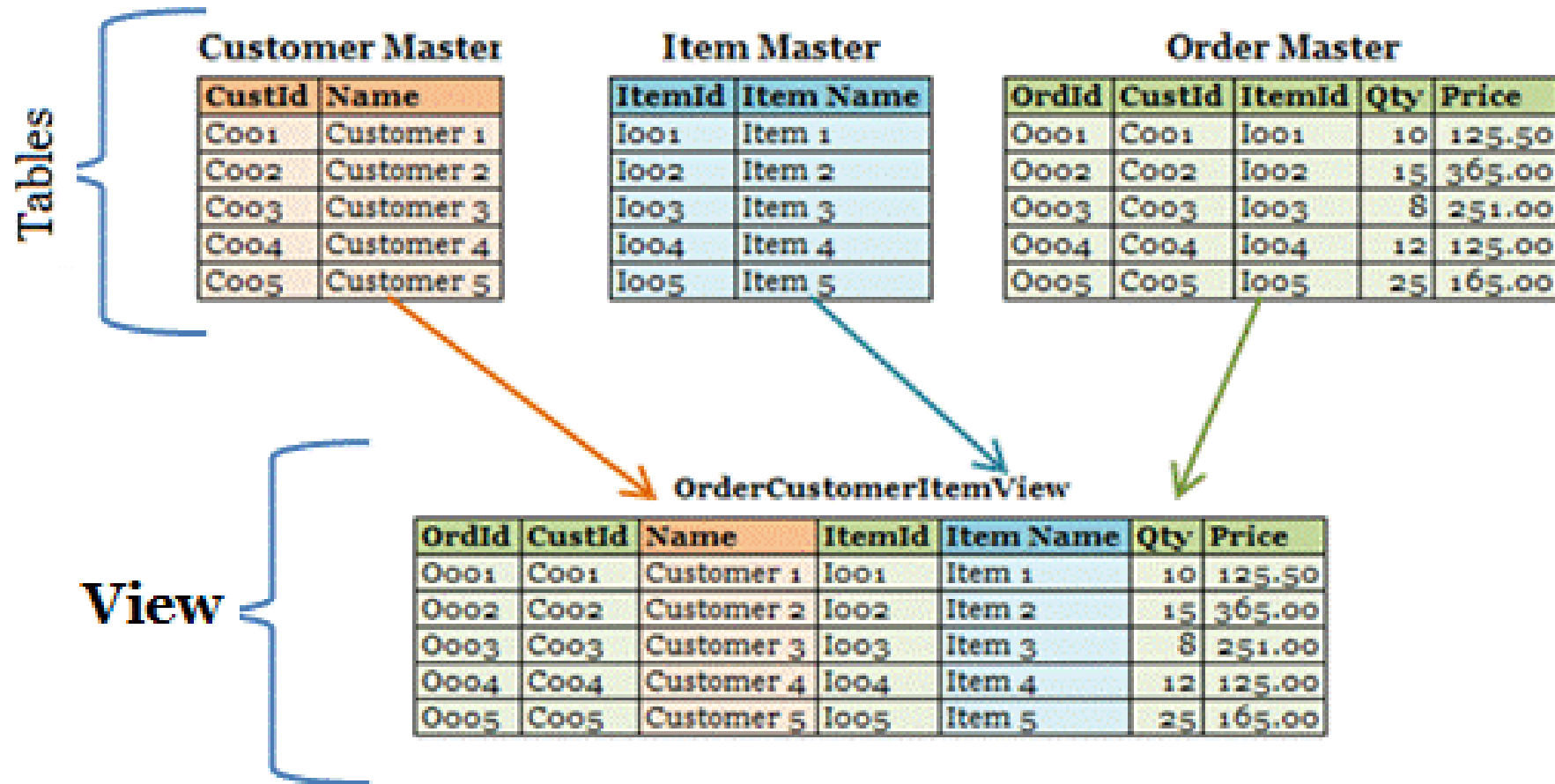
Transformación  
en Destino



**ELT**

Escalabilidad y análisis  
de datos en bruto.  
data warehouse/lake

# Transformación: Views



# Transformación: dbt (data build tool)



# Transformación: Spark

Realiza Análisis Exploratorio de  
Datos (EDA)

Escala horizontal

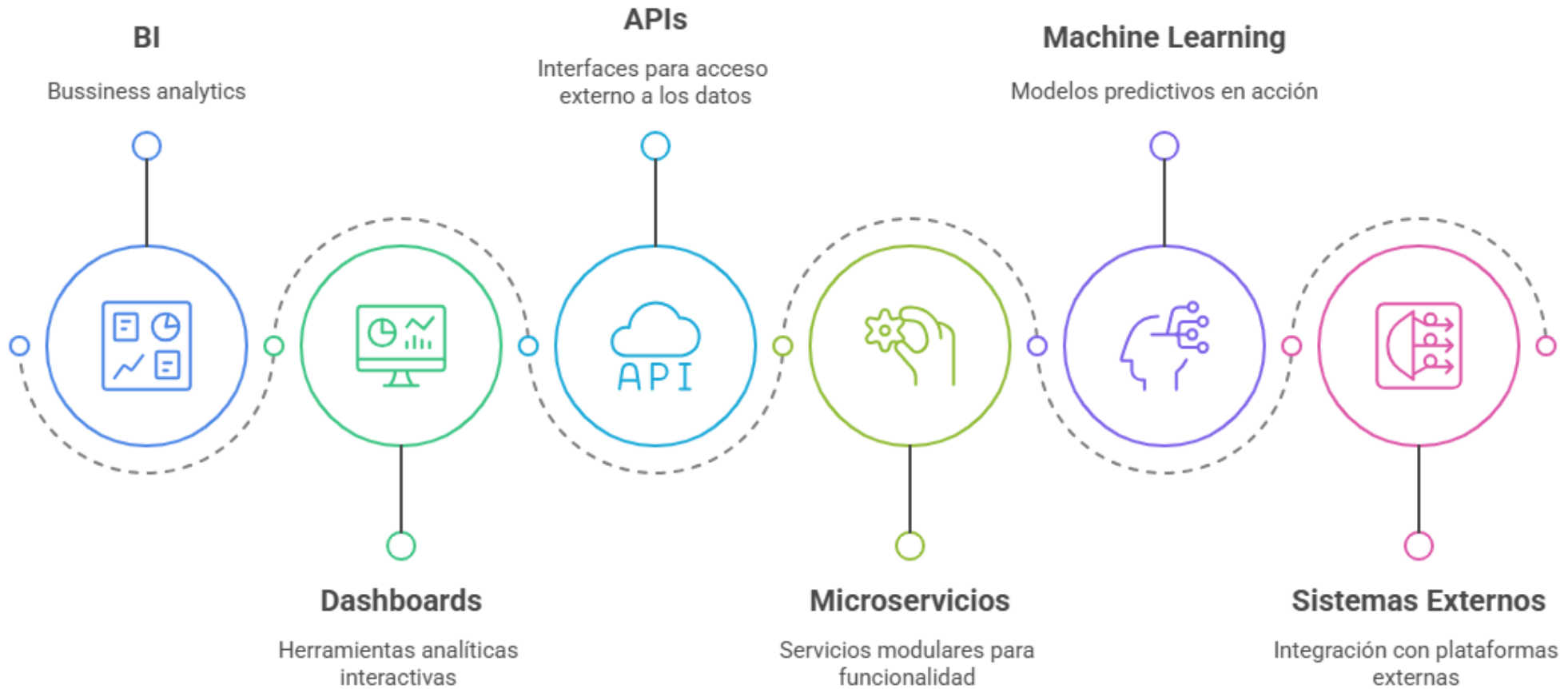
Rápidas ejecución

MLlib





# Serving



# Serving: APIs

```
@app.get("/sales/{region}")
async def get_sales(region: str,
                    start: date,
                    end: date):
    return db.query(Sales)
        .filter(Sales.region == region)
        .filter(Sales.date.between(start, end))
```

# Serving: APIs

## Paginación

```
{  
  "data": [...],  
  "pagination": {  
    "total": 1000,  
    "page": 2,  
    "per_page": 50,  
    "next": "/api/v1/sales?page=3"  
  }  
}
```

## Versionado

- /api/v1/sales
- /api/v2/sales

## Circuit Breaker

```
@retry(stop=stop_after_attempt(3),  
        wait=wait_exponential(multiplier=1))  
def fetch_external_data():  
    ...
```

# Serving: APIs

## Paginación

```
{  
  "data": [...],  
  "pagination": {  
    "total": 1000,  
    "page": 2,  
    "per_page": 50,  
    "next": "/api/v1/sales?page=3"  
  }  
}
```

## Versionado

- /api/v1/sales
- /api/v2/sales

## Circuit Breaker

```
@retry(stop=stop_after_attempt(3),  
        wait=wait_exponential(multiplier=1))  
def fetch_external_data():  
    ...
```

# Herramientas de Análisis y Visualización: Tableau

- drag-and-drop
- Soporte para big data
- Dashboards interactivos y compartibles
- Tableau Embedding API





## Campaign Analysis

20,790

PD

28,822

PA

404

FV

185

BPA

200

MPP



### Filters

Note: \*These are global filters, applicable on all the visualizations.



Select FY  
FY 16-17

Select Months  
All

### Location Filters

Retail/IB  
All

Zone  
All

Region  
All

State  
All

City  
All

Location Name  
All

### Article Filters

Segment  
All

Family  
All

Class  
All

MC  
All

Brand Name  
All

Manufacturer  
All

Salt  
All

Sku Name  
All

### Other Filters

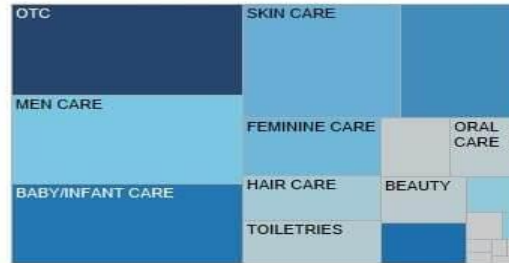
Customer Type  
All

Home Delivery  
All

Scheme Made  
All

Hour Group  
All

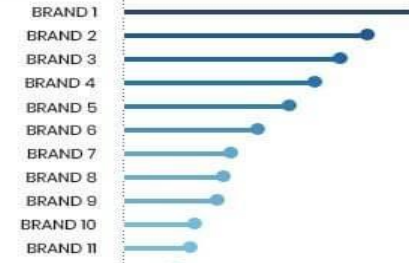
### Major Category



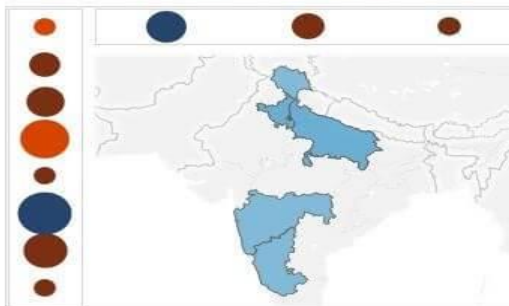
### Category



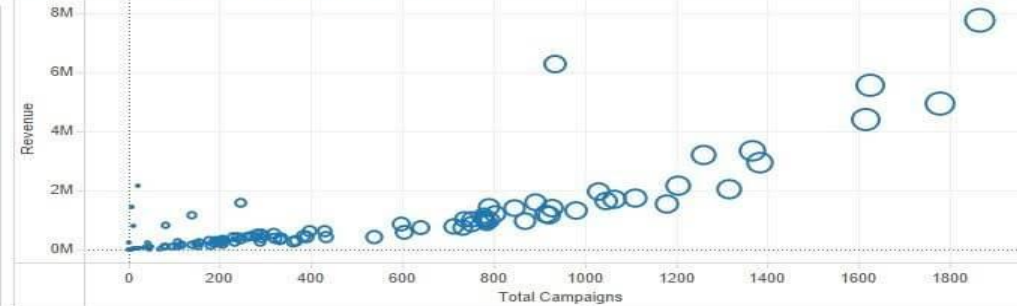
### Brand



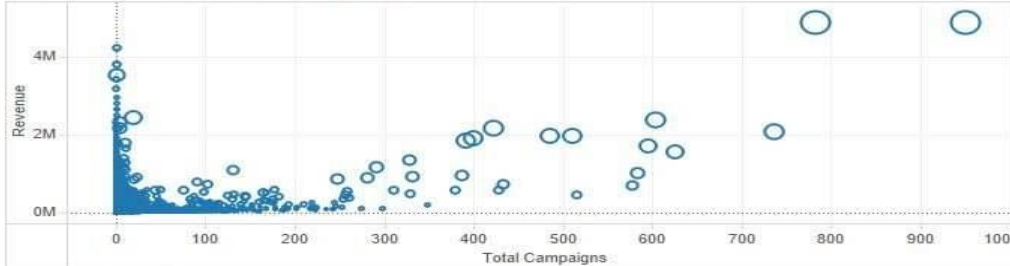
### Effect On Stores During Camps



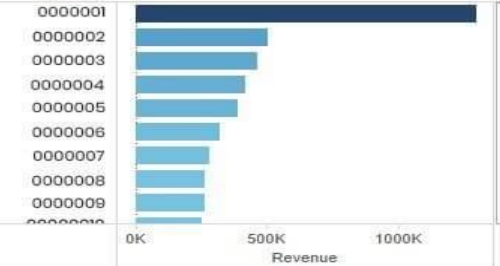
### Store wise Sale due to campaign



### Sale of Products due to campaign



### Camp Codes

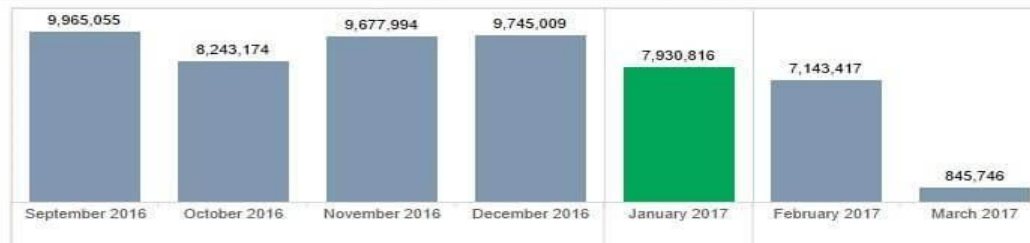


### Promotion Effectiveness

Please select a Month, Year to see promotion Effectiveness.

Jan

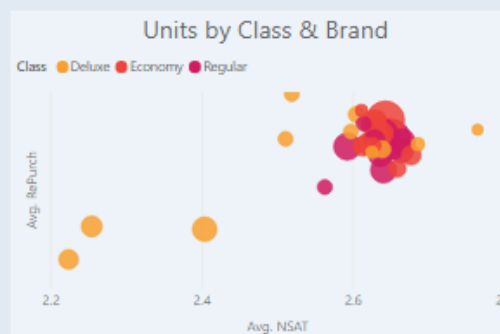
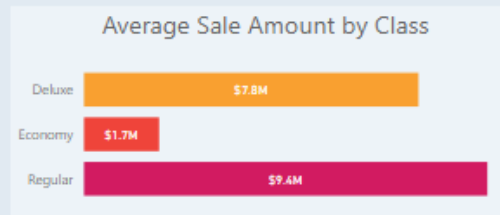
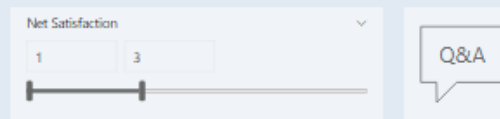
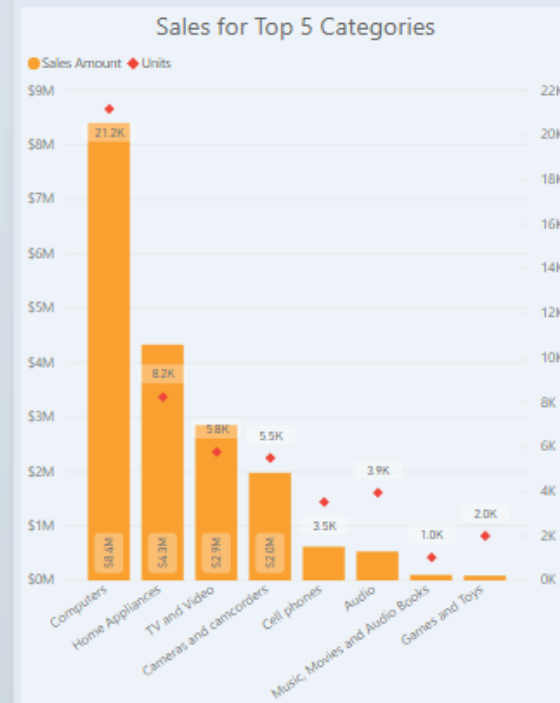
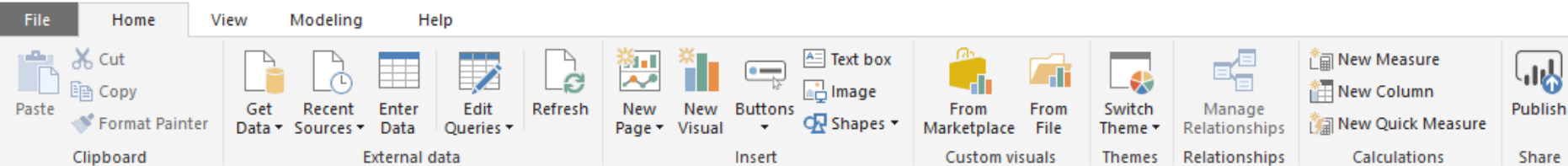
2017



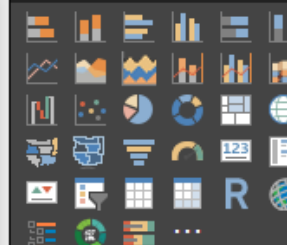
# Herramientas de Análisis y Visualización: Power BI

- Amplia gama de estilos (gráficos, diagramas, plots, etc)
- Gran soporte para visualizaciones en mapas
- Disponible en varias aplicaciones, gratis y de pago.
- Power BI Embedded





## VISUALIZATIONS



## Values

Drag data fields here

## FILTERS

Page level filters

Drag data fields here

Drillthrough filters

Drag drillthrough fields here

Report level filters

Drag data fields here

## FIELDS

Search

## Sales

- BrandName
- Category
- Class
- Color
- Country
- CustomerAcc...
- Manufacturer
- NSAT
- OrderDate
- ProdID
- ProductDescri...
- ProductLabel
- ProductName
- RePurch
- Sale Size
- SalesAmount
- SalesAmount ...
- SalesChannel...
- SalesOrderID
- StockType
- StoreKey
- StyleName
- SubCategory
- Units



# Herramientas de Análisis y Visualización



Looker Studio



**Metabase**