



06-11-2023

DocRed Test - Ingeniero de datos V.I

Juan Pablo Mantilla



# PARTE I: MODELO RELACIONAL PARA UN SOFTWARE DE VENTAS

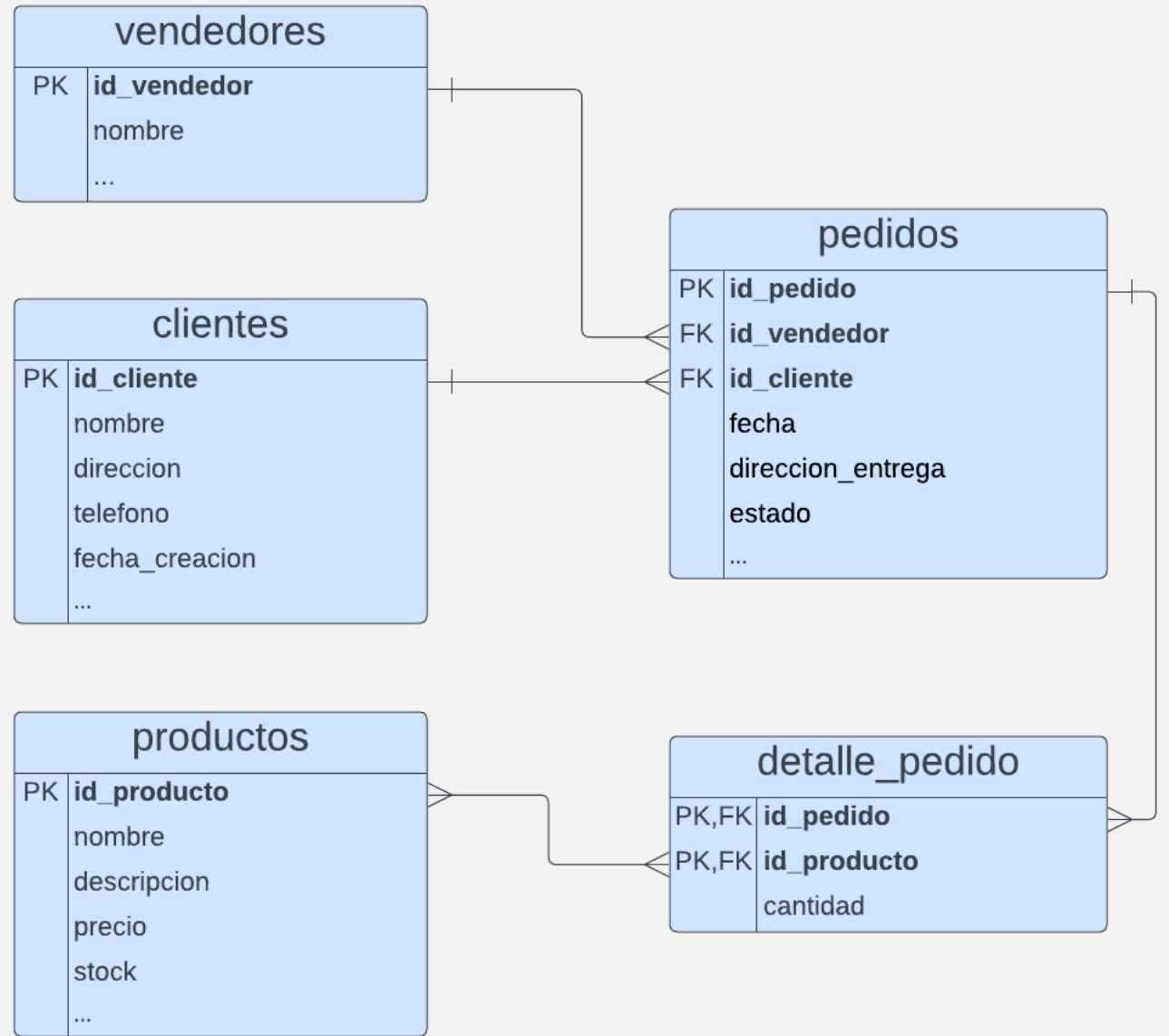
*El modelo debe almacenar la siguiente información:*

1. **Cientes:** nombre, dirección, teléfono, y cualquier otro dato relevante
2. **Productos:** nombre, descripción, precio, y cualquier otro dato relevante.
3. **Pedidos:** fecha, y cualquier otro dato relevante

# PARTE I: MODELO RELACIONAL PARA UN SOFTWARE DE VENTAS

```

CREATE TABLE Clientes(
    ID_Cliente SERIAL PRIMARY KEY,
    Nombre VARCHAR(100),
    Direccion VARCHAR(100),
    Telefono VARCHAR(15),
    Fecha_Creacion DATE
)
CREATE TABLE Vendedores (
    ID_Vendedor SERIAL PRIMARY KEY,
    Nombre VARCHAR(100),
    ...
)
CREATE TABLE Productos (
    ID_Producto SERIAL PRIMARY KEY,
    Nombre VARCHAR(100),
    Descripcion VARCHAR(255),
    Precio DECIMAL(10, 2),
    Stock INT
)
CREATE TABLE Pedidos (
    ID_Pedido SERIAL PRIMARY KEY,
    Direccion_Entrega VARCHAR(100),
    Fecha DATE,
    Estado INT,
    ID_Cliente INT,
    FOREIGN KEY (ID_Cliente) REFERENCES Clientes(ID_Cliente)
    FOREIGN KEY (ID_Vendedor) REFERENCES Vendedores(ID_Vendedor)
)
CREATE TABLE Detalle_Pedido (
    ID_Pedido INT,
    ID_Producto INT,
    Cantidad INT,
    PRIMARY KEY (ID_Pedido, ID_Producto),
    FOREIGN KEY (ID_Pedido) REFERENCES Pedidos(ID_Pedido),
    FOREIGN KEY (ID_Producto) REFERENCES Productos(ID_Producto)
)
    
```



# *¿Por qué SQL?*

## Datos estructurados

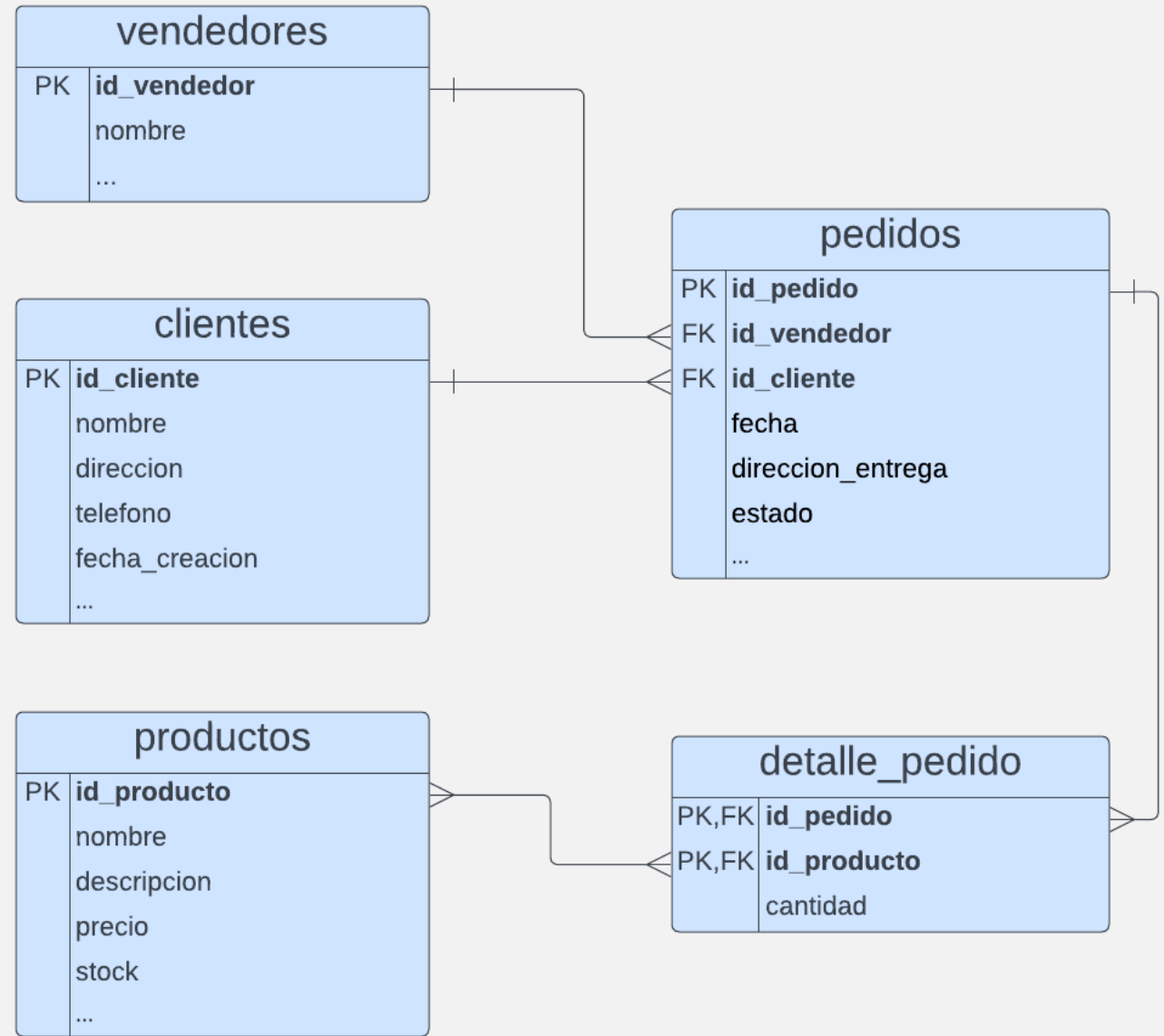
Los campos de cada tabla están definidos pre-implementación

## Operaciones entre varias tablas

Más eficiencia al realizar operaciones de cuenta, unión u cálculos con datos de varias tablas + Transacciones ACID (actualizar stock )

## Simplicidad

Fácil de implementar y manejar



## PARTE 2: MODELADO DE BASE DE DATOS PARA SENSORES DE EVENTOS DE USUARIO

*El modelo debe almacenar la siguiente información:*

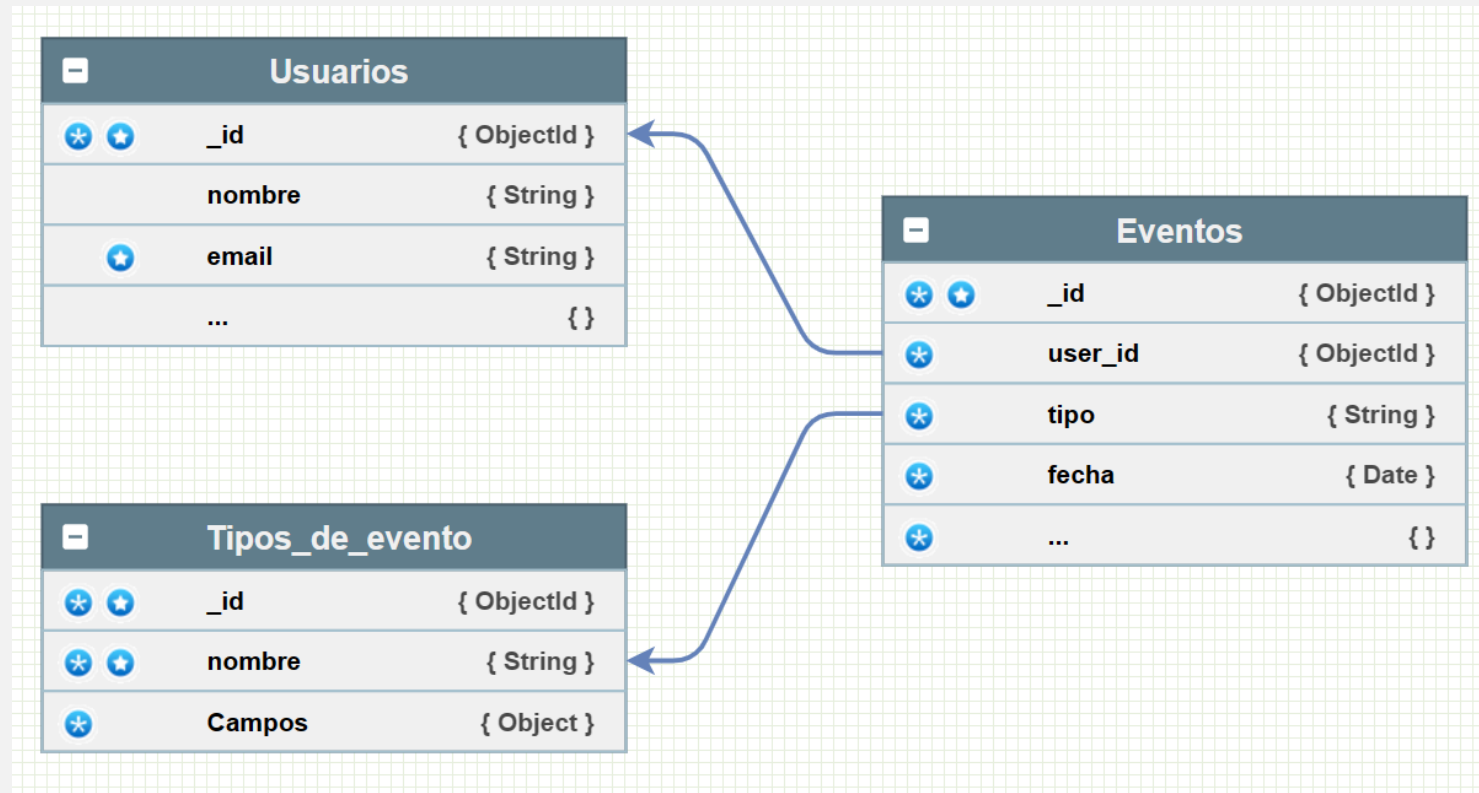
1. **Usuarios:** nombre, dirección de correo electrónico, y cualquier otro dato relevante.
2. **Eventos:** fecha y hora del evento, tipo de evento, y cualquier otro dato relevante

## Parte 2: Modelado de Base de Datos para Sensores de Eventos de Usuario.

Problema: Heterogeneidad de datos bajo sombrilla de “Evento”

- Clics
- Movimientos del ratón
- Eventos de teclado
- Eventos de formulario
- Eventos de navegación
- Eventos de sesión
- Eventos de interacción con el contenido
- Eventos de interacción con elementos de la interfaz de usuario
- Eventos de error
- Eventos de interacción con publicidad
- ....

## Parte 2: Modelado de Base de Datos para Sensores de Eventos de Usuario.



## Parte 2: Modelado de Base de Datos para Sensores de Eventos de Usuario.

```
{
  "schemaType": "Collection",
  "name": "Usuarios"
  "fields": [
    {
      "schemaType": "Field",
      "name": "_id",
      "type": "ObjectId",
      "required": true,
      "unique": true,
    },
    {
      "schemaType": "Field",
      "name": "nombre",
      "type": "String",
      "required": false,
      "unique": false,
    },
    {
      "schemaType": "Field",
      "name": "email",
      "type": "String",
      "required": false,
      "unique": true,
    },
    ...
  ]
}
```

Usuarios		
* *	_id	{ ObjectId }
	nombre	{ String }
*	email	{ String }
	...	{ }



## Parte 2: Modelado de Base de Datos para Sensores de Eventos de Usuario.

```
{
  "schemaType": "Collection",
  "name": "Tipos_de_evento"

  "fields": [
    {
      "schemaType": "Field",
      "name": "_id",
      "type": "ObjectId",
      "required": true,
      "unique": true,
    },
    {
      "schemaType": "Field",
      "name": "nombre",
      "type": "String",
      "required": true,
      "unique": true,
    },
    {
      "schemaType": "Field",
      "name": "Campos",
      "type": "Object",
      "required": true,
      "unique": false,
    }
  ]
}
```

Tipos_de_evento		
* *	<b>_id</b>	{ ObjectId }
* *	<b>nombre</b>	{ String }
*	<b>Campos</b>	{ Object }

## Parte 2: Modelado de Base de Datos para Sensores de Eventos de Usuario.

```
{
  "unique": false,
  "schemaType": "Collection",
  "name": "Eventos",
  "fields": [
    {
      "schemaType": "Field",
      "name": "_id",
      "type": "ObjectId",
      "required": true,
      "unique": true,
    },
    {
      "schemaType": "Field",
      "name": "user_id",
      "type": "ObjectId",
      "required": true,
      "unique": false,
    },
    {
      "schemaType": "Field",
      "name": "fecha",
      "type": "Date",
      "required": true,
      "unique": false,
    },
    {
      "schemaType": "Field",
      "name": "tipo",
      "type": "String",
      "required": true,
    }
  ]
}
```

**Campos propios de cada tipo de evento**

Eventos		
*	* _id	{ ObjectId }
*	user_id	{ ObjectId }
*	tipo	{ String }
*	fecha	{ Date }
*	...	{ }

# *¿Por qué NoSQL?*

## Flexibilidad de esquemas

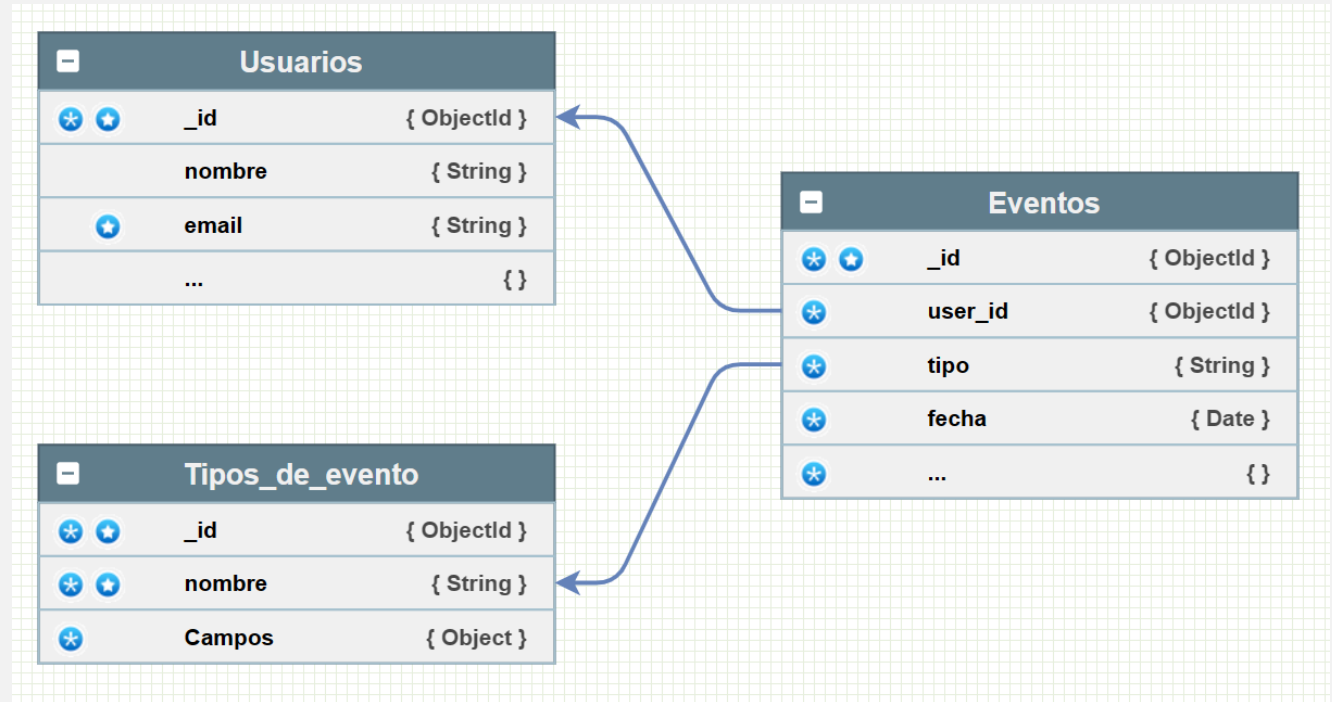
No requiere un esquema fijo, lo que significa que se pueden almacenar diferentes tipos de eventos con diferentes estructuras de datos

## Escalabilidad

Fácilmente escalables con el volumen de datos

## Alto rendimiento

Puede manejar un gran número de operaciones de lectura y escritura por segundo (registrar eventos en tiempo real)



## PARTE 3: ANÁLISIS DE DATOS

1. **Análisis de datos:** explorar la base de datos de ejemplo y realizar un análisis de los datos relevantes para el contexto
2. **Selección de criterios:** elegir los criterios de selección para filtrar los datos según considere necesario.
3. **Generación de informes:** generar informes personalizados que presenten los datos de manera clara y comprensible.
4. **Interpretación de resultados:** interpretar los resultados obtenidos a través del análisis y los informes generados.

# ANÁLISIS DE DATOS

- Distribución
- Tendencias
- Periodicidad

# DATASET

id	Id de Usuario
[nombre, apellido, email, numero de documento, tipo de documento, numero de telefono movil ]	Datos de identificación y contacto
<b>especialidad medica</b>	['Hematologia', 'Pediatría', 'Dermatología', 'Ortopedia', 'Cardiología', 'Gastroenterología', 'Medicina de familia', 'Neurología', 'Medicina general', 'Oncología', 'Psiquiatría', 'Endocrinología']
<b>pais</b>	['Uruguay', 'Ecuador', 'Guatemala', 'Honduras', 'Argentina', 'Panama', 'Peru', 'Colombia', 'Chile', 'Mexico', 'Bolivia', 'Venezuela', 'Cuba', 'Paraguay', 'Costa Rica', 'Puerto Rico']
<b>genero</b>	['Masculino', 'Femenino']

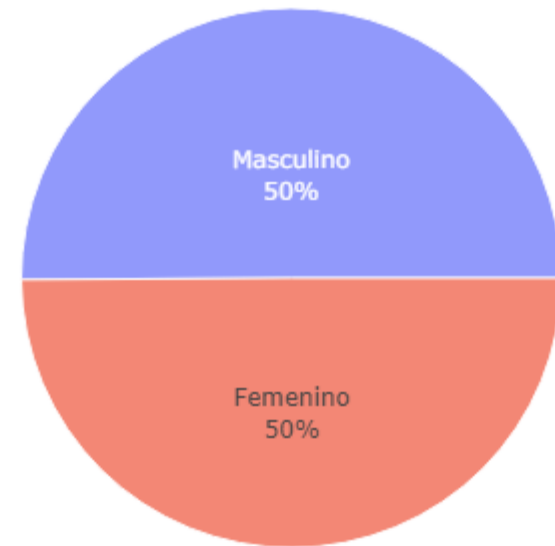
Distribución según países



Distribución de especialidades

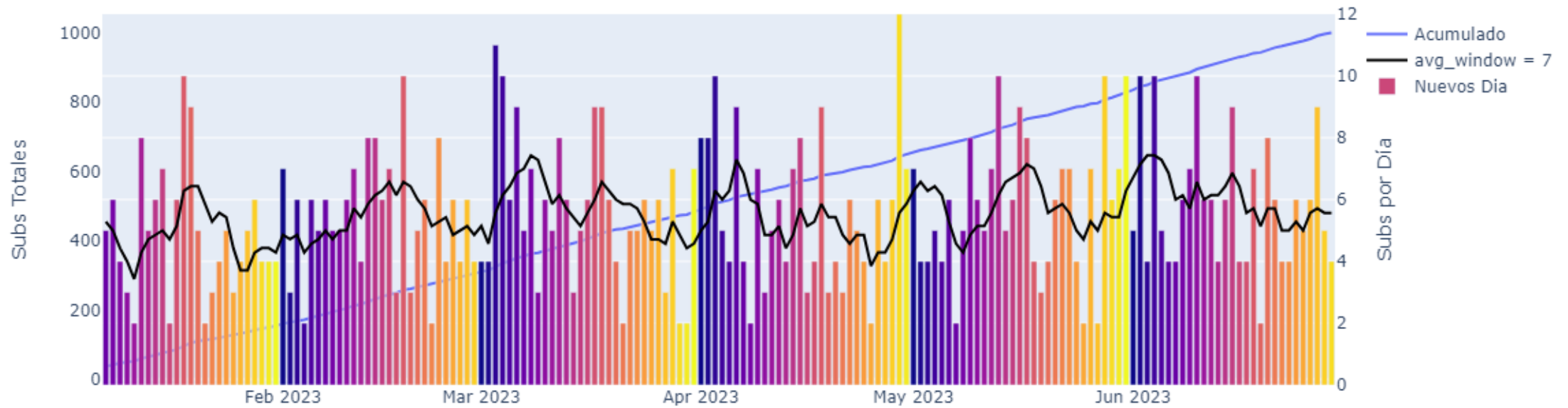


Distribución según género



\*detalles inspeccionables en dashboard

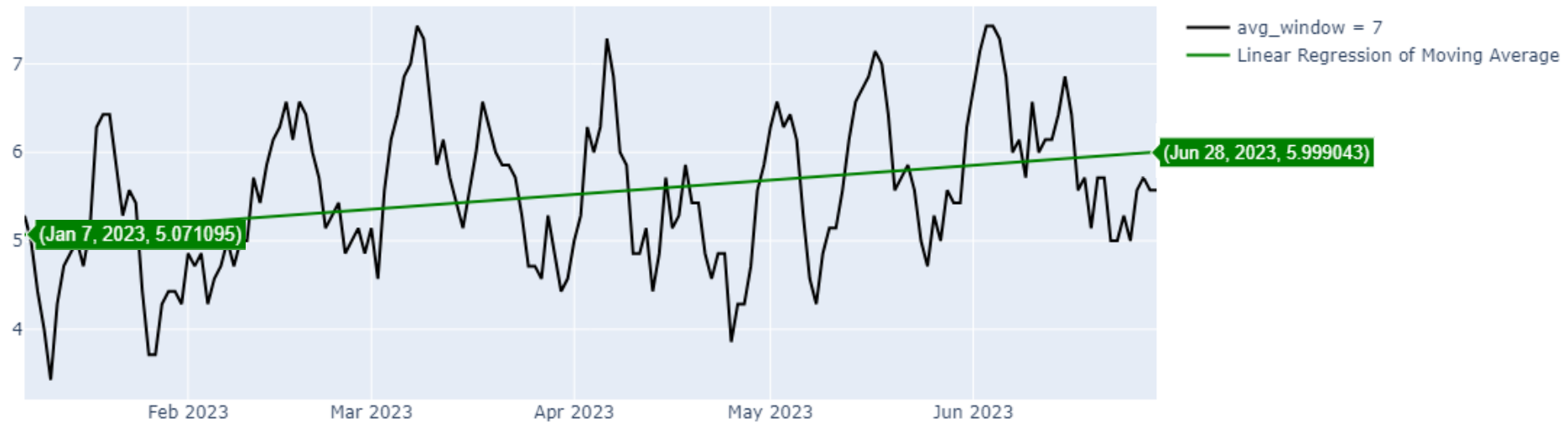
## Subscripciones por día



\*detalles inspeccionables en dashboard

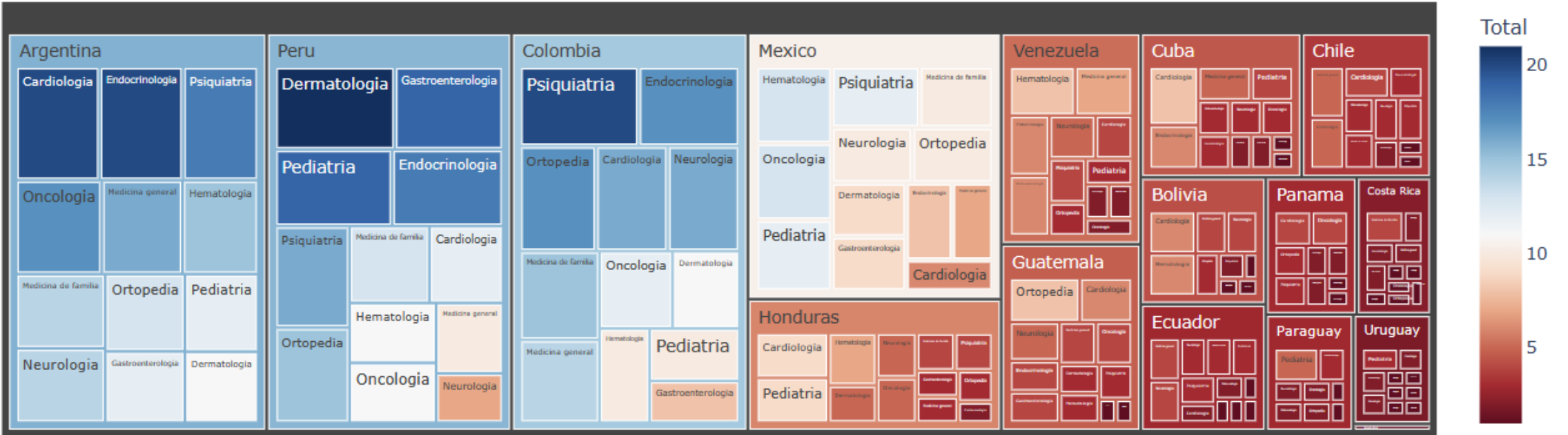


# Tendencia General



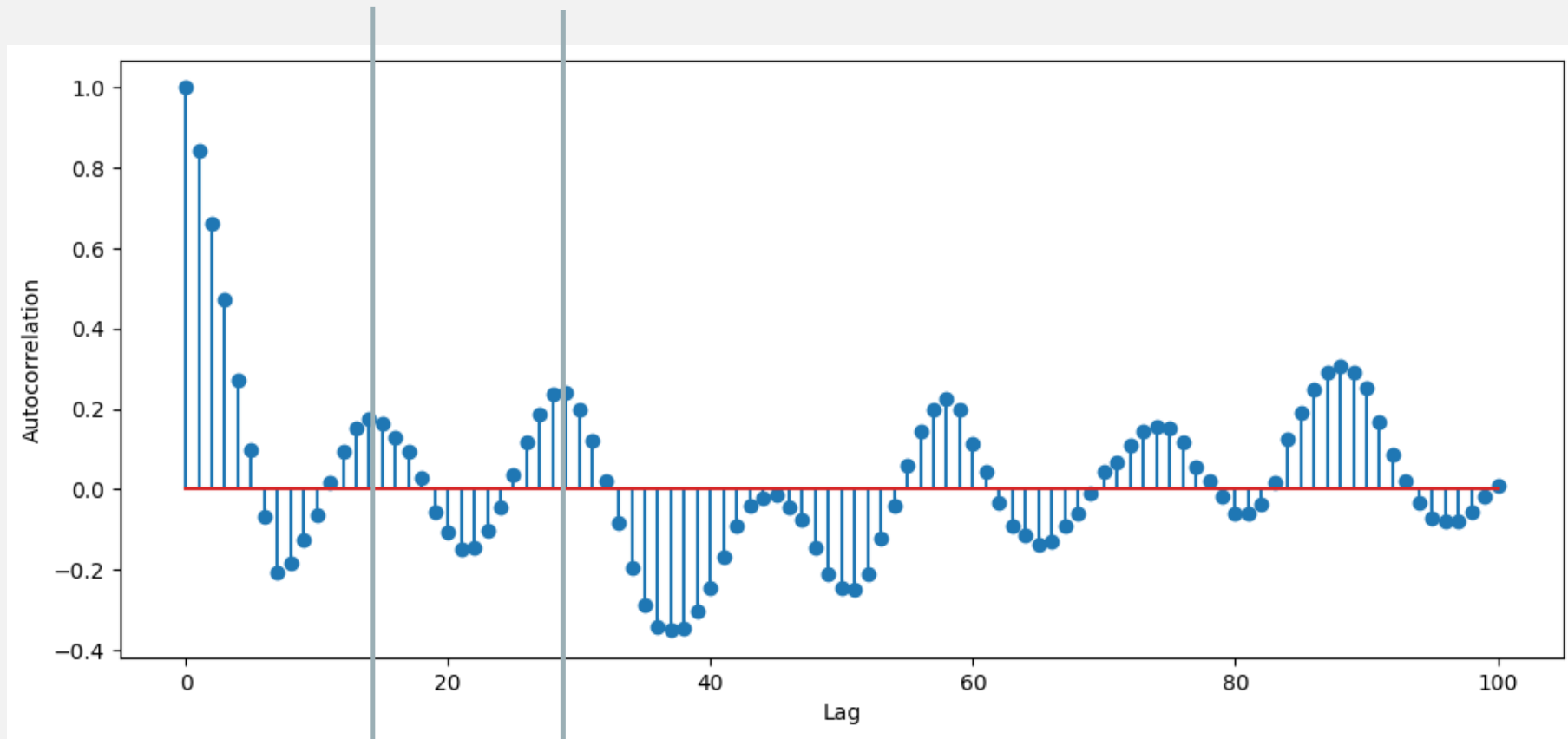
\*detalles inspeccionables en dashboard

# Distribucion



\*detalles inspeccionables en dashboard

# Periodicidad



## CONCLUSIONES

En primer lugar, la tendencia general es al alza, lo que indica un crecimiento positivo y sostenido.

Este crecimiento es impulsado en gran medida por la fuerte concentración de clientes en Argentina, Colombia, Perú y México, que juntos representan más del 60% de la base de nuevos clientes.



## CONCLUSIONES

Se observa un equilibrio tanto en términos de especialidades médicas como de género.

Esto sugiere una amplia aceptación de nuestro servicio en diversas áreas de la medicina y entre todos los géneros.



## CONCLUSIONES

Se identificó una autocorrelación de  $14 \pm 1$  días en los datos, lo que indica que hay picos de nuevos usuarios cada dos semanas.

Esta periodicidad puede estar relacionada con las campañas publicitarias que se realizan con esta frecuencia o alternatively, puede ser el resultado de las actividades propias del personal médico, que parece tener más tiempo para inscribirse a servicios como este a principios de quincena.



## CONCLUSIONES

En resumen, los resultados del proyecto son prometedores, con una tendencia de crecimiento al alza y una base de clientes diversa y equilibrada.

Las conclusiones extraídas de este análisis y posteriores análisis de subgrupos del Dataset nos permitirán tomar decisiones informadas para continuar mejorando y expandiendo nuestros servicios.

