

# TP Final Base de Datos Relacional

Responsable: Mauro Aguirregaray

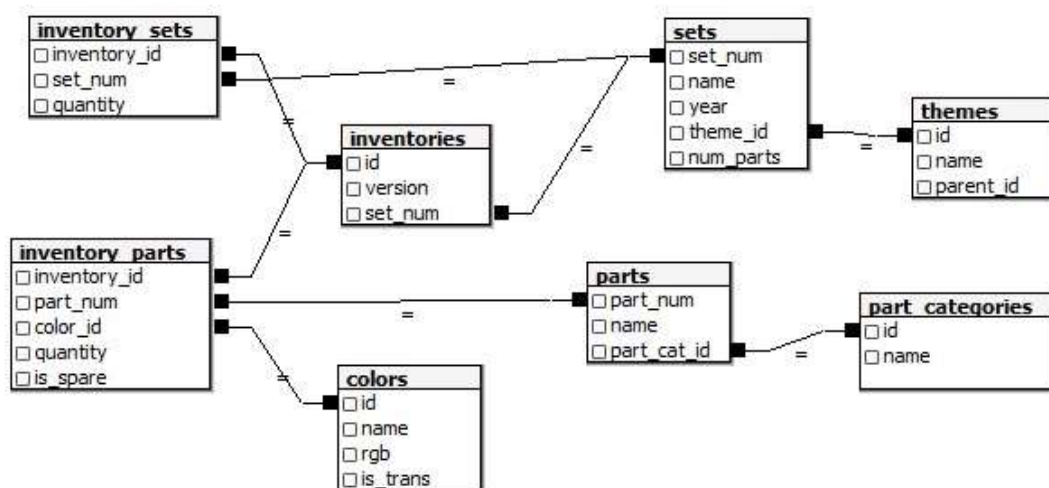
## Base de datos elegida: Legos

### Primera Parte: Formación de la base de datos

Se recibieron 8 archivos .csv que corresponden con las tablas de la base de datos de Legos:

- Colors.csv
- Inventories.csv
- Inventory\_parts.csv
- Inventory\_sets.csv
- Part\_categories.csv
- Parts.csv
- Sets.csv
- Themes.csv

Estas 8 tablas se relacionan de la siguiente manera:



Para este proyecto se utilizó el motor de PostgreSQL, y se cargó la base de datos en el entorno de pgAdmin4 para hacer las queries.

Este proceso requirió crear las tablas una a una e importar los archivos acordemente.

### Segunda Parte: Queries y análisis

En esta segunda parte se nos pide trabajar como si fuéramos parte del equipo de marketing de LEGO y presentar un informe especial con los siguientes puntos como eje:

1. Colores más utilizados en los 90
2. Colores únicos
3. Tendencia de piezas por sets a lo largo de los años
4. Temáticas más populares de los 90



## Análisis Histórico y Tendencias de LEGO: Evolución de Colores, Piezas y Temáticas

En este informe especial analizaremos datos históricos y tendencias para obtener una visión integral de cómo han evolucionado nuestros productos y preferencias de los fanáticos. Este análisis nos permitirá comprender mejor las decisiones pasadas y planificar estrategias futuras basadas en información concreta.

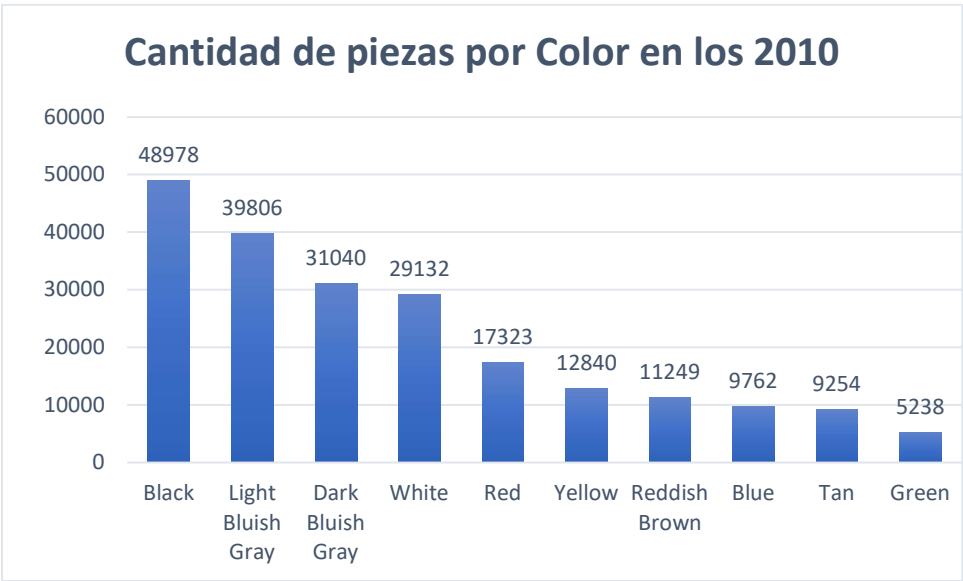
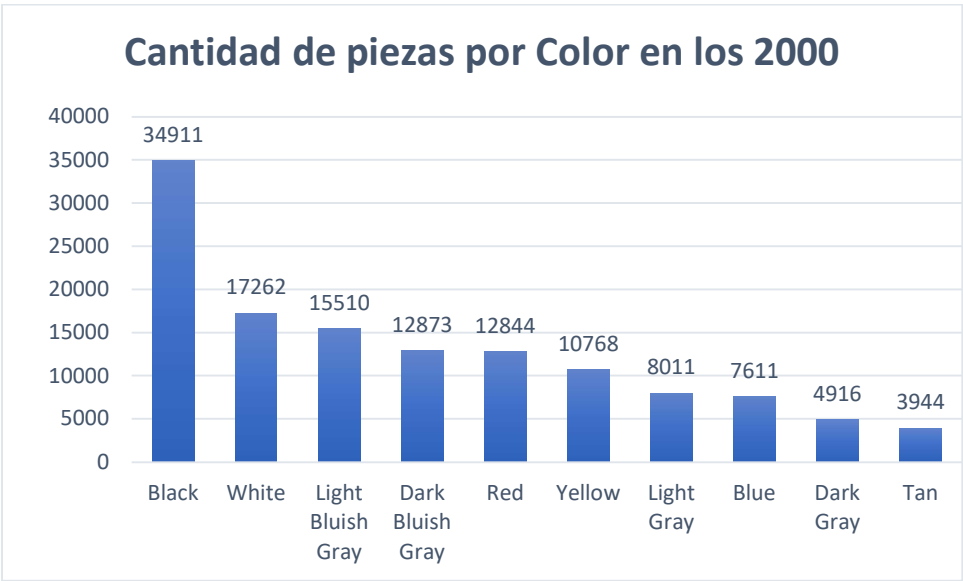
Uno de los primeros puntos que queremos enfocar en el análisis es la evolución de los colores en los sets de LEGO a lo largo de los años. Exploraremos cómo han cambiado las preferencias y la popularidad de ciertos colores en diferentes décadas.

Por ejemplo, en la década de los 90 de un total de 79253 piezas de sets lanzados en esos años, se presenta la siguiente distribución por color:



Color	cantidad_piezas_por_color	suma_total_cantidad_piezas
Black	19655	79253
Light Gray	11233	79253
White	11203	79253
Red	9616	79253
Yellow	8364	79253
Blue	5769	79253
Dark Gray	2460	79253
Green	2136	79253
Brown	1519	79253
Trans-Clear	748	79253

Se puede observar que más de la mitad de las piezas lanzadas en esos años están escala de grises “Black, Gray y White”. Es interesante ver si en los siguientes años varían los colores más utilizados:



Si bien en las siguientes décadas se dispararon al doble o triple de piezas totales el negro y el blanco siguen siendo los colores más predominantes, con una proyección en los últimos años del “light bluish y dark bluish grays” que prácticamente no tenían relevancia en los 90.

Esto podemos ver que es un comportamiento esperable, pues el blanco y negro son colores que pueden aparecer en cierta medida en cualquier set.

*La query utilizada en el entorno para obtener estos resultados fue la siguiente:*

```

SELECT c.name,
       COUNT(ip.inventory_id) AS Cantidad_Piezas_Por_Color,
       (SELECT COUNT(ip2.inventory_id)
        FROM sets s2
        JOIN inventories i2 ON i2.set_num = s2.set_num
        JOIN inventory_parts ip2 ON ip2.inventory_id = i2.id
        WHERE s2.year < 2000 AND s2.year > 1989) AS Suma_Total_Cantidad_Piezas
FROM sets s
JOIN inventories i ON i.set_num = s.set_num
JOIN inventory_parts ip ON ip.inventory_id = i.id
JOIN colors c ON ip.color_id = c.id
WHERE s.year < 2000 AND s.year > 1989
GROUP BY c.name
ORDER BY Cantidad_Piezas_Por_Color DESC
LIMIT 10;

```

En total y a lo largo de todos los años, las piezas de lego han tenido 129 colores únicos, sin contar transparente ni “unknown”.

La query utilizada en el entorno para obtener estos resultados fue la siguiente:

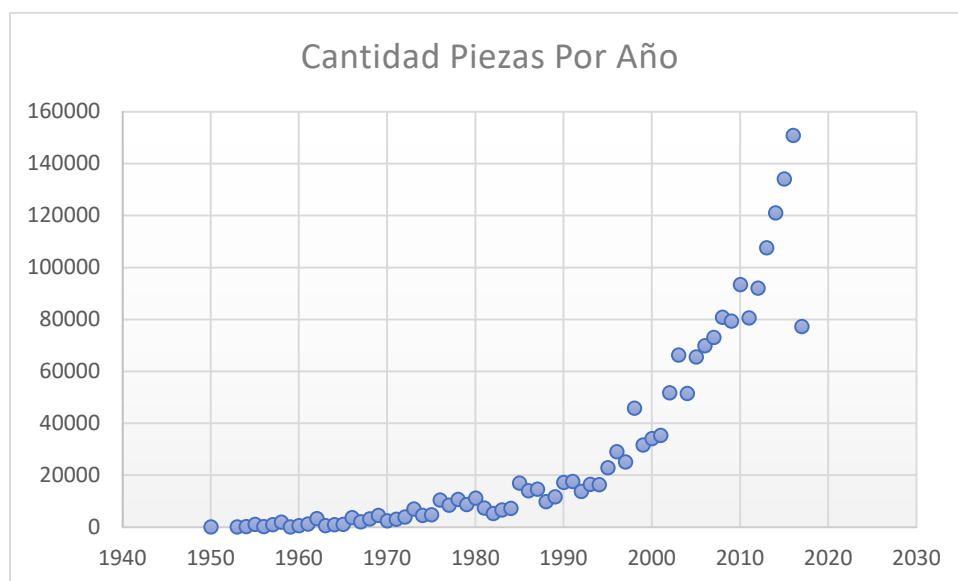
```

-- Cantidad de colores únicos
SELECT DISTINCT c.name, ip.color_id
FROM inventory_parts ip
JOIN colors c ON ip.color_id = c.id
ORDER BY ip.color_id;

```

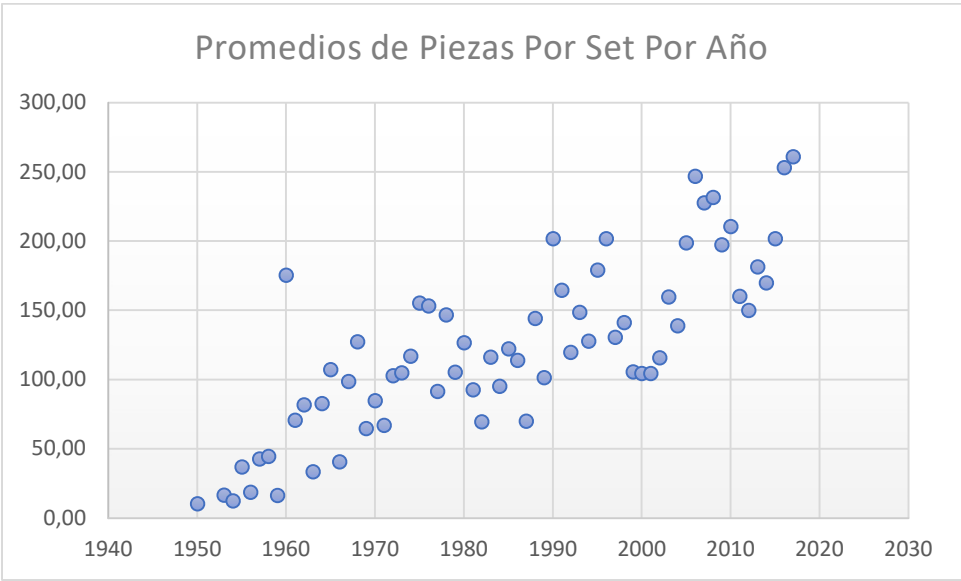
Ya vimos los colores y observamos como década a década hay más piezas totales, sin embargo no sabemos aún si son porque año a año salen más sets o si en cambio los sets tienen más piezas en promedio.

Así que vamos a estudiar eso, cómo han evolucionado la cantidad de piezas por set a lo largo de los años. Primero comprobemos lo que veníamos adelantando, ¿cada año salen más piezas?



La respuesta es que sí, todos los años se han sacado más piezas y se presenta un claro comportamiento exponencial, entendible pensando en la globalización de la marca LEGO.

En cuanto a lo que son piezas promedio por set en un año, vemos un aumento también. Esta vez parece ser un comportamiento lineal de que cada año, los sets en promedio tienen más piezas.



La query utilizada en el entorno para obtener estos resultados fue la siguiente:

```
-- Promedio de cantidad piezas por set por año
SELECT year, AVG(num_parts) AS Cantidad_Piezas_Promedio_Por_Año
FROM sets
GROUP BY year
ORDER BY year;
```

Por último, hagamos un chequeo similar con las temáticas, centrándonos en cuáles fueron las más populares este siglo.

name	cantidad_de_sets_por_tematica
City	287
Friends	269
Creator	249
Gear	240
Star Wars	210
Duplo	205
Ninjago	191
Basic Set	158
Technic	153
Star Wars Episode 4/5/6	143
Key Chain	131
Bulk Bricks	124
Police	122
Clikits	105
Construction	101
Batman	93
LEGO Brand Store	88

Supplemental	86
Legends of Chima	84

Mientras que si vemos las temáticas previas al 2000, tenemos:

name	cantidad_de_sets_por_tematica
Supplemental	410
Technic	282
Service Packs	201
Basic	101
Basic Set	99
Construction	93
Airport	90
Race	89
Classic Space	89
4.5V	88
Traffic	87
Fabuland	83
Train	78
Classic Basic	69
Basic Model	68
Freestyle	61
Fire	55
Police	52
12V	50
Building	43

Es interesante observar como en los últimos años las principales temáticas están dadas por grandes franquicias del mundo del entretenimiento (Star Wars, Batman, etc), mientras que en el siglo anterior era más popular los sets genéricos para construir como el de la siguiente imagen.

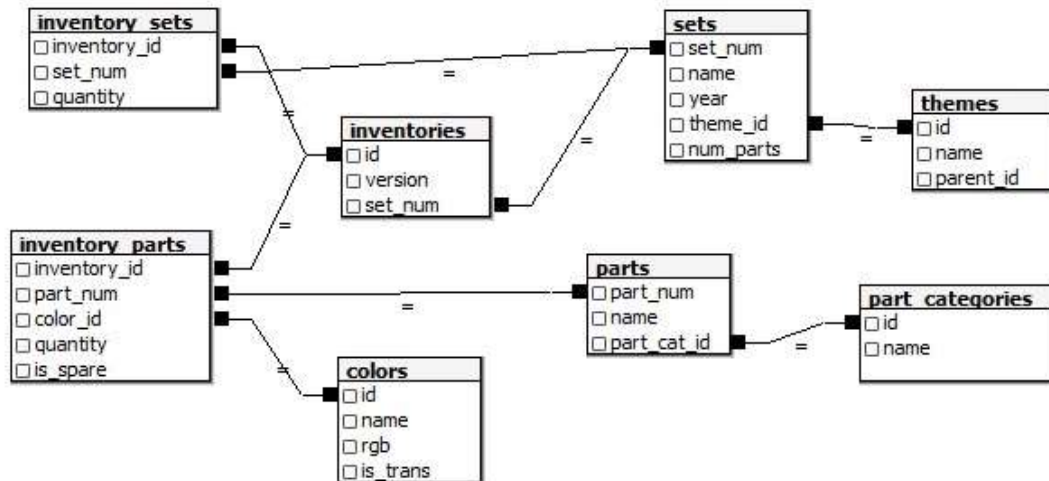


*La query utilizada en el entorno para obtener estos resultados fue la siguiente:*

```
-- Temáticas más populares luego de los 2000
SELECT t.name, COUNT(s.set_num) AS Cantidad_de_sets_por_tematica
FROM themes t
JOIN sets s ON s.theme_id = t.id
WHERE s.year >1999
GROUP BY t.name
ORDER BY Cantidad_de_sets_por_tematica DESC;
```

## Anexo: Comentarios sobre la base de datos

Al importar la base de datos al entorno de pgAdmin4 me encontré con algunos problemas, veamos de nuevo las relaciones:



Hay dos puntos importantes e interesantes a ver para estos datos:

- La tabla `inventory_sets` no parece ser tan necesaria, al menos yo no le encontré utilidad, toma dos FK que en `inventories` ya están (siendo `id` PK de `inventories`). Por lo que parecería que el único dato nuevo “quantity” podría pertenecer sin problema a la tabla `inventories` y ahorrarse una tabla.
- Por otro lado, al intentar ingresar las restricciones no me fue posible declarar como FK de `inventories` a “set\_num”, ya que hay datos de “set\_num” en la tabla de `inventories` que no están en la tabla `sets`, cuando en `sets` “set\_num” es una PK. De esta manera, no se estaría respetando la integridad referencial de la variable y de esa manera el motor no me deja avanzar. Para este informe y con las queries que fui presentando, no resulto un problema puesto que al hacer los JOIN, me descarto estos datos, sin embargo, si fuera en un entorno profesional habría que chequear este punto.

Después otro dato menor, pero que, creo que puede ayudar a la claridad, es mejorar la notación, varias columnas tienen el mismo nombre “id” y se refieren a datos en tablas diferentes. Se podría cambiar por ejemplo “color\_id” en la tabla `colors`, para que mantuviera el mismo nombre con la relación de FK en la tabla `inventory_parts`.



## Conclusiones finales

El trabajo estuvo interesante pues se pareció más a un trabajo real, ya que no todo estaba ya “armado” como para simplemente hacer las queries. Además, el tema era interesante y se podían sacar conclusiones y análisis de los datos resultantes de las queries.

Personalmente me hubiese gustado haber sido más claro en la presentación del problema con el tema de que el database puede tener problemas de integridad, en mi caso, creía que el error era mío.