



# LEGO Trend ANALYSIS

*Since 1970... and still building!*

# Index

- I. Company Profile
- II. Project Objectives & Dataset Description
- III. First Part: *Data Cleaning & Exploration (MySQL)*
- IV. Second Part: *Data Visualization & Analysis (Tableau Dashboard)*
- V. Conclusion: *Key Insights*





# I. Company Profile

Founded in 1932 by Ole Kirk Christiansen in Denmark, the LEGO® Group has grown into a global leader in the toy industry.

It is best known for its signature interlocking plastic bricks, which inspire creativity and hands-on learning through play.

LEGO's product portfolio spans both original themes, such as LEGO City and Technic, and popular licensed franchises including Star Wars, Harry Potter, and Marvel.

## II. Project Objectives & Dataset Description

- LEGO sets released from **1970** to **2022**, including details on set's theme, subtheme, pieces, retail price, image and recommended age.
- The **aim** of this project is to show the **LEGO sets trend analysis** over the past 50 years, with the purpose to identify some pattern or differences between the **price**, **theme** and **pieces** across these years.
- To identify the **principal theme** for **each decade** and give some insights regarding the price and the possible correlation with the LEGO minifigures numbers.
- To guide this analysis, I first established a clear roadmap outlining the **key objectives** and the appropriate **tools** needed to explore each one effectively.



# III. First Part: Data Cleaning & Exploration

## Data Cleaning with MySQL

To clean and prepare the dataset for initial exploration, I used SQL as the primary tool.

### Issues identified:

Column	Problem Type	Description
pieces	Data Type Error	Stored as text, expected integer
minifigs	Data Type Error	Stored as text, expected integer
agerange_min	Data Type Error	Stored as text, expected integer
US_retailPrice	Data Type Error	Stored as text, expected decimal
All Columns	Missing Values	Some fields were empty or non-numeric
General	Irregular Formats (n/a, unknown, etc.)	Needed to set NULL for consistency



The first step was to create a backup table (**lego\_sets\_clean**), allowing safe and structured preparation of the dataset.



This enabled a smooth transition into the data formatting phase.

```
SELECT * FROM lego_sets;
```

```
--- Create a backup table
```

```
CREATE TABLE lego_sets_clean as  
SELECT * FROM lego_sets;
```

# III. First Part: Data Cleaning & Exploration

## Data Cleaning with MySQL

- Detecting non-numeric Values and setting “null” for consistency.
- Correct data types and column names.
- Ensure data consistency.

### Main steps performed:

- ✓ Standardized Column types: Converted fields like pieces, minifigs, and US\_retailPrice to proper INT o DECIMAL types.
- ✓ Renamed Columns: Cleaned column names for clarity, such as: **themeGroup**→ **theme\_group**, **bricksetURL** → **brickset\_url**, etc.
- ✓ Handled Non-Numeric Values: Replaced invalid strings with NULL for numerical fields.
- ✓ Verified Themes: Ran queries to inspect and validate values.

### Tools Used:

- MySQL.
- Regular expressions for validation (REGEXP).
- ALTER TABLE, UPDATE, SELECT for formatting and exploration.

```
ALTER TABLE lego_sets_clean
MODIFY COLUMN pieces int,
MODIFY COLUMN minifigs int,
MODIFY COLUMN agerange_min int,
MODIFY COLUMN US_retailPrice decimal (10,2);
```

```
ALTER TABLE lego_sets_clean
CHANGE COLUMN themeGroup theme_group text,
...
CHANGE COLUMN imageURL image_url text;
```

```
UPDATE lego_sets_clean
SET
theme = trim(theme),
subtheme = trim(subtheme),
theme_group = trim(theme_group),
category = trim(category);
```

*Sample SQL queries used for formatting and cleaning*



# III. First Part: Data Cleaning & Exploration

## Data Exploration with MySQL & Excel

Through targeted queries, I addressed key business questions, laying the foundation for deeper analysis.



### 1. LEGO Sets Overview

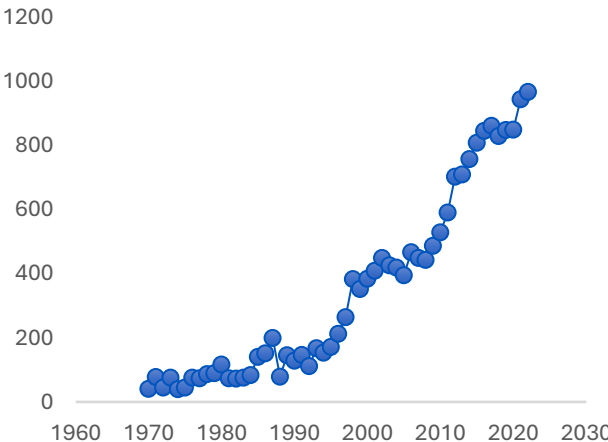
- Goal: Understand the evolution of LEGO Sets over time.
- How many sets were released each year?

```
SELECT year, COUNT(*) AS
total_sets_released
FROM lego_sets_clean
GROUP BY year;
```

```
-- Count Themes per decade
-- Rank using RANK() OVER by
decade
-- Select top 1 per decade
```

*Sample SQL queries used*

Total Set Released since 1970



### 2. Top Theme by Decade

- Which themes dominated each decade?

DECADE	TOP THEME	SET COUNT
1970s	LEGOLAND	150
1980s	TOWN	190
1990s	TOWN	346
2000s	GEAR	892
2010s	GEAR	1306
2020s	GEAR	585

# III. First Part: Data Cleaning & Exploration

## Data Exploration with MySQL & Excel

3. Price Distribution by Age Group – How do prices vary by recommended age?

recommended_age	total_sets	avg_price	min_price	max_price
18	114	\$ 195.65	\$ 34.99	\$ 849.99
16	74	\$ 212.15	\$ 19.99	\$ 849.99
14	36	\$ 188.60	\$ 79.99	\$ 499.99
12	72	\$ 108.81	\$ 24.99	\$ 449.99
11	22	\$ 186.35	\$ 99.99	\$ 289.99
10	218	\$ 55.44	\$ 9.99	\$ 359.99
9	246	\$ 80.39	\$ 9.99	\$ 169.99
8	670	\$ 50.43	\$ 2.99	\$ 199.99
7	925	\$ 29.16	\$ 3.99	\$ 199.99
6	1078	\$ 22.02	\$ 1.99	\$ 789.99
5	649	\$ 18.59	\$ 1.99	\$ 99.99
4	226	\$ 23.37	\$ 2.49	\$ 99.99
3	17	\$ 51.46	\$ 9.99	\$ 239.95
2	237	\$ 34.93	\$ 6.99	\$ 284.95
1	70	\$ 17.88	\$ 4.99	\$ 49.99

Higher recommended age = higher average price

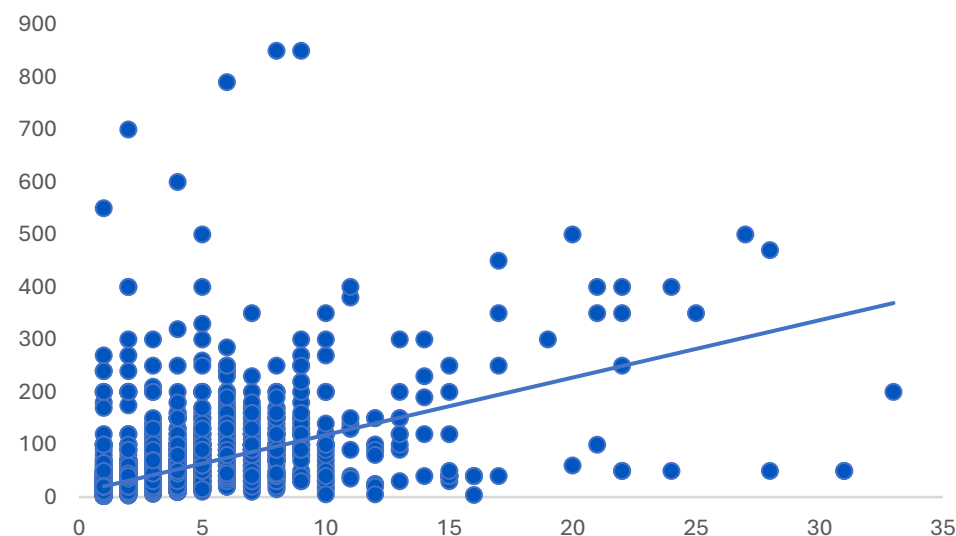
### SQL Queries Steps:

- Grouped data by **age\_range\_min**
- Calculated **average**, **min**, and **max** price
- Filtered null values

4. Minifigs vs. Price – Is there a correlation between the number of minifigures and set price?

**Correlation Coefficient: 0.54**

☐ Moderate positive correlation





# IV. Second Part: Data Visualization

## *Data Visualization in Tableau*

In the second phase, I used **Tableau** to develop an interactive dashboard that reveals key trends and changes in LEGO sets over the past 50+ years.

To guide the analysis, I adopted a stakeholder perspective, identifying relevant KPIs and framing business-oriented questions.



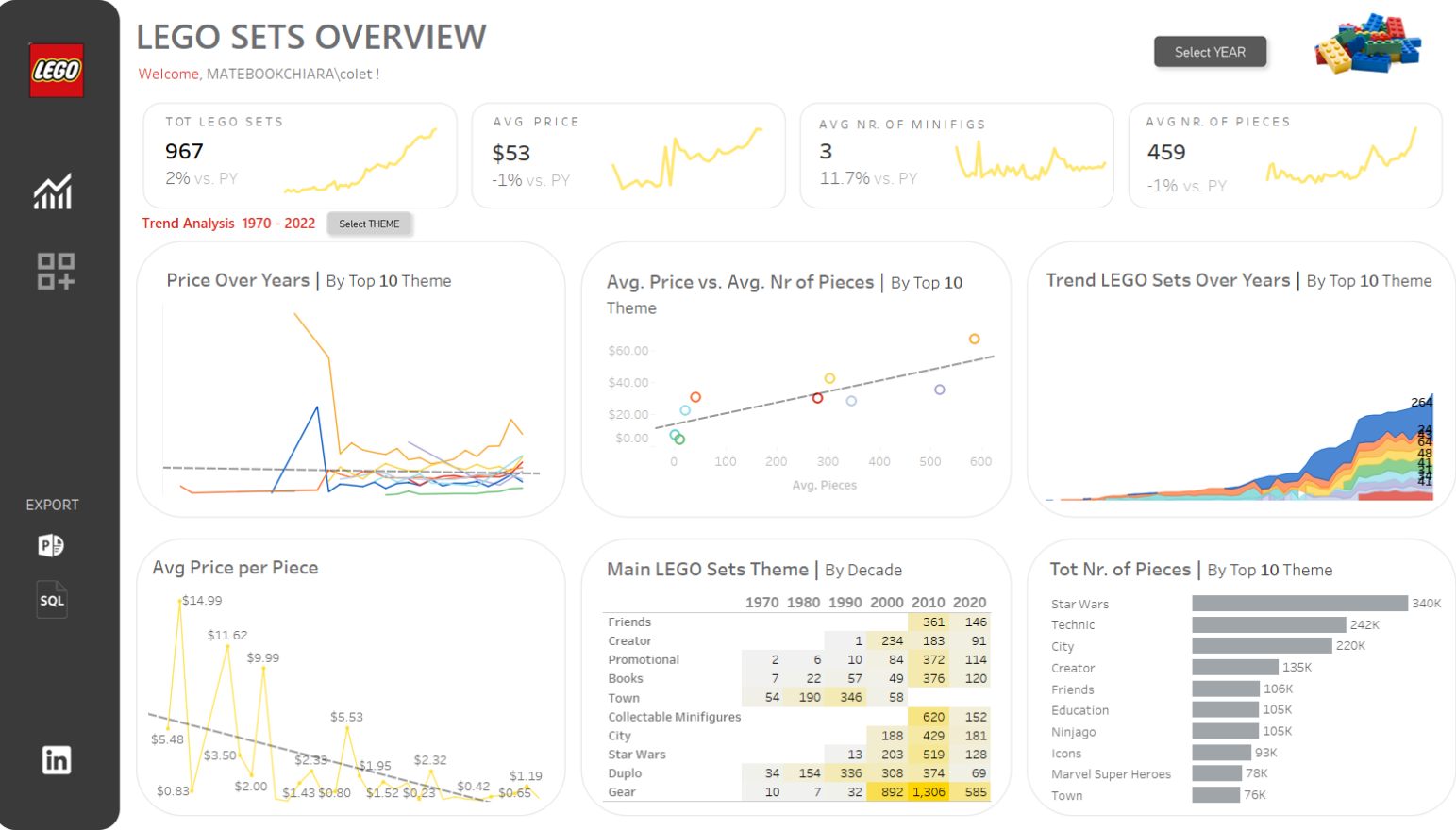
### **Analysis Goals:**

- ✓ LEGO sets evolution across 50 years
- ✓ Changes in the number of pieces per set
- ✓ Price vs. number of pieces: is there a correlation?
- ✓ Most popular themes by decade
- ✓ Are minifigures more common in licensed sets?
- ✓ Most expensive vs. cheapest LEGO sets
- ✓ How have average prices changed over time?



# IV. Second Part: Data Visualization

## Dashboard 1: LEGO Sets Overview



This dashboard provides a comprehensive overview of LEGO set trends from 1970 to 2022.

It highlights **key metrics** such as the number of sets released, average price, number of pieces, and minifigures along with **top-performing themes** and the evolution of LEGO sets over time.



The dashboard was built using core dataset measures (e.g., price, pieces, year) and custom-calculated fields to highlight trends, KPIs, and theme performance.

# IV. Second Part: Data Visualization

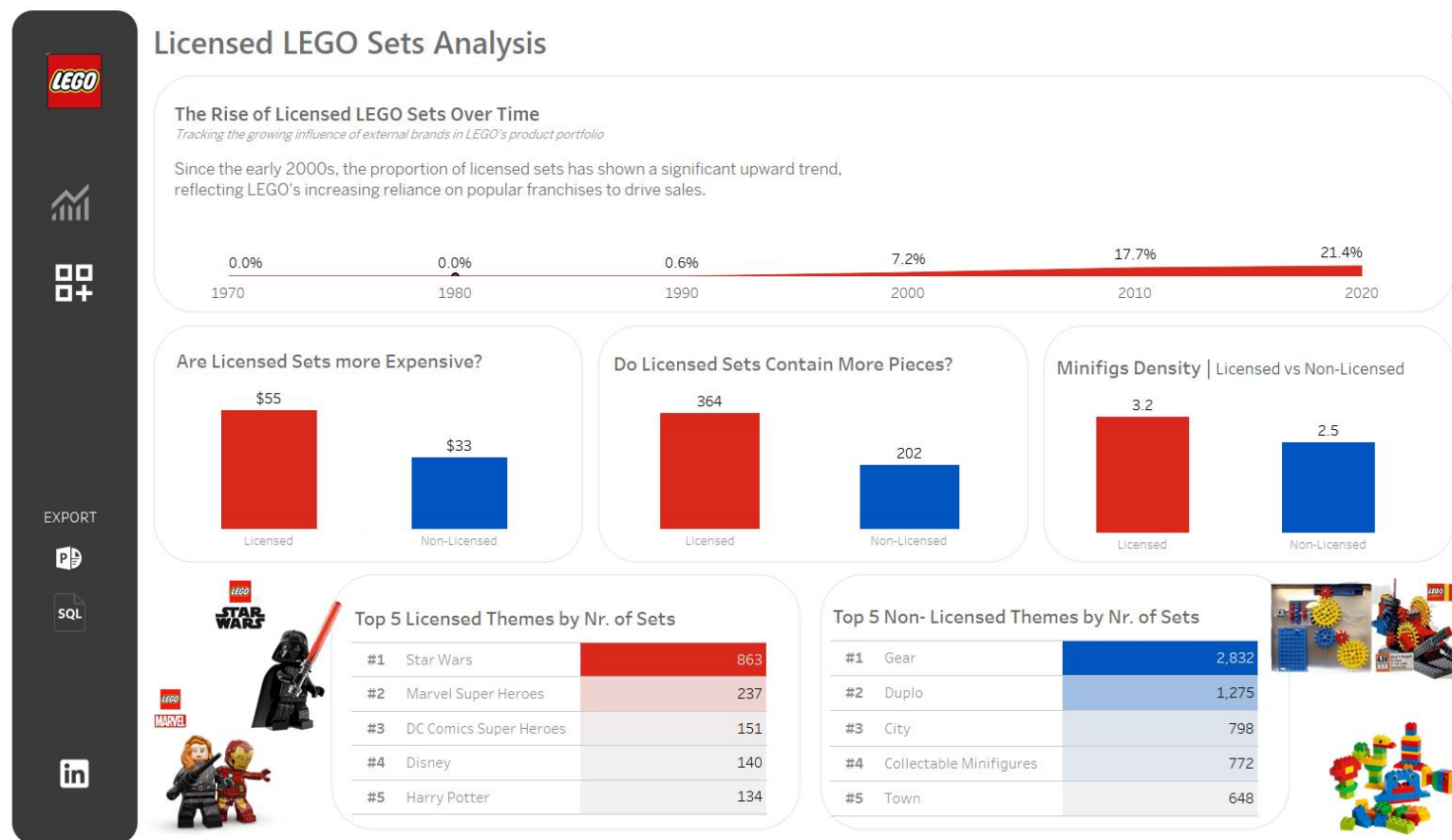
## Dashboard 2: Licensed vs Non-Licensed Sets Analysis

This dashboard focuses on comparing Licensed vs. Non-Licensed LEGO sets, highlighting how external brand partnerships have influenced LEGO's evolution.

It explores:

- The rise in the percentage of licensed sets over time
- Price and complexity differences (e.g., avg. price, pieces, minifigs)
- The top 5 licensed franchises vs. original LEGO themes

The analysis was supported by calculated fields to classify themes (e.g., "Licensed Category") and compute averages across both categories.



Comparison of Licensed vs. Non-Licensed sets, analysing trends in price, minifigs, pieces, and theme popularity.

# V. Conclusion

## Key Insights:

- There is a clear **positive correlation** between the average number of pieces and the average set price across top LEGO themes.
- The number of LEGO sets released has **grown significantly**, from 41 in 1970 to 967 in 2022.
- The **average set size** has increased steadily over the past decade, reaching **459 pieces per set** in 2022.
- The **average number of minifigures per set** has remained relatively stable at around **3 per set**.
- **Price per piece** has shown a **gradual decline**, indicating greater value offered to consumers over time.
- In **2003**, LEGO experienced a notable drop in average set prices, due to financial and strategic crisis. A leadership change in the following year led to a successful turnaround by refocusing on core products.
- From the early 2000s, there has been a **notable rise in licensed sets** (e.g., Star Wars, Marvel), which have influenced higher prices, more pieces, and increased minifigs density.
- While original LEGO themes (e.g., Gear, Duplo) remain noticeable in volume, **licensed themes have had greater impact on pricing and product complexity**.

