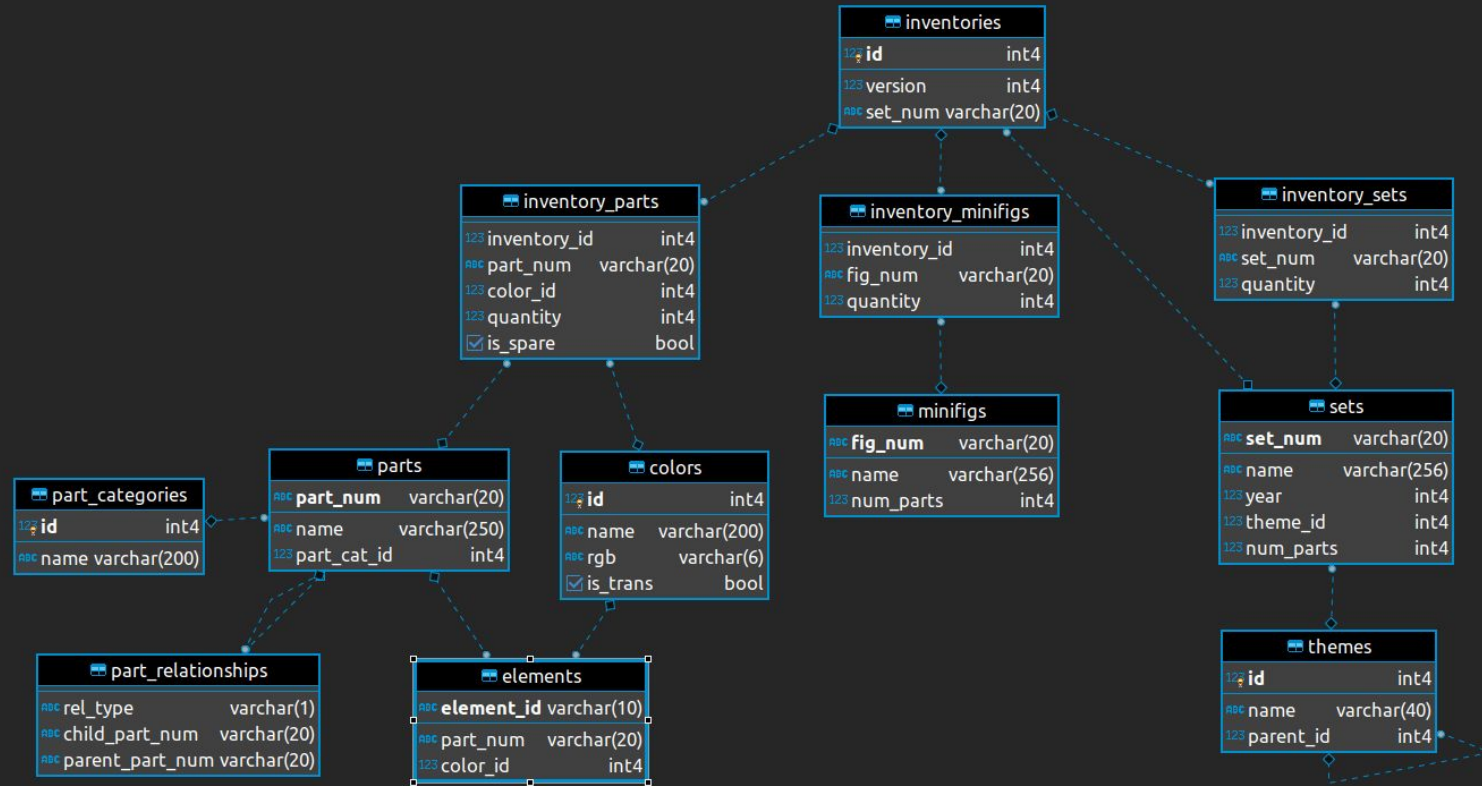




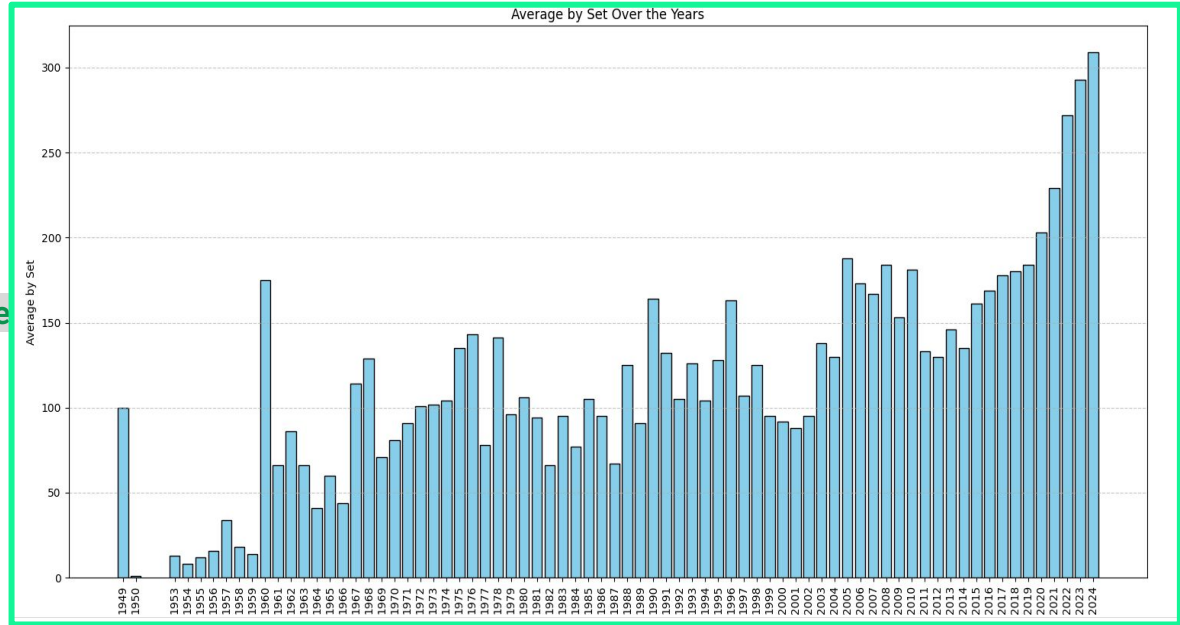
# An Analysis of Lego Dataset

# The structure of the dataset “Lego”



# Question: What is the overall trend in average set values from 1949 to 2024?

```
SELECT
  year,
  round(AVG(num_parts)) AS
averagebyset
FROM
  `lego-june2024.lego_dataset.sets`
GROUP BY
  year
```



# Which year had the lowest average set values and why?

```
Select
year, name, set_num,
num_parts
FROM
`lego-june2024.lego_dataset.
sets`
WHERE year = 1950
```

| A    | B   | C         | D         |
|------|---|-----------|-----------|
| year | name  | set_num   | num_parts |
| 1950 | Single 2 x 4 Brick (ABB)                    | 700.1.1-1 | 1         |
| 1950 | Single 2 x 2 Brick (ABB)                    | 700.1.2-1 | 1         |
| 1950 | Single 1 x 4 x 2 Window without Glass (ABB) | 700.B.1-1 | 1         |
| 1950 | Single 1 x 2 x 3 Window without Glass (ABB) | 700.B.2-1 | 1         |
| 1950 | Single 1 x 2 x 2 Window without Glass (ABB) | 700.B.3-1 | 1         |
| 1950 | Single 1 x 2 x 4 Door without Glass (ABB)   | 700.B.4-1 | 1         |
|      |   |           |           |

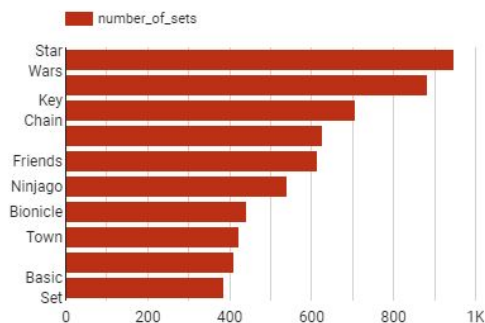
# Diversity of Themes Over Time

```
SELECT
    t.name AS theme_name,
    COUNT(s.set_num) AS
number_of_sets
FROM

`lego-june2024.lego_dataset.sets` s
JOIN

`lego-june2024.lego_dataset.themes` t
ON
    s.theme_id = t.id
GROUP BY
    theme_name
ORDER BY
    number_of_sets DESC
LIMIT 10;
```

|     | theme_name   | number_of_sets |
|-----|--------------|----------------|
| 1.  | Star Wars    | 949            |
| 2.  | Technic      | 883            |
| 3.  | Key Chain    | 706            |
| 4.  | Gear         | 628            |
| 5.  | Friends      | 614            |
| 6.  | Ninjago      | 540            |
| 7.  | Bionicle     | 442            |
| 8.  | Town         | 423            |
| 9.  | Classic Town | 411            |
| 10. | Basic Set    | 386            |



## We also can see the distribution of *rare LEGO parts* across different themes

```
WITH
  SELECT
    COUNT(DISTINCT
      ) AS
  FROM
    `lego-june2024.lego_dataset.inventory_parts`
  GROUP BY

  HAVING
    < 10
)
SELECT
  AS
  COUNT(DISTINCT
    ) AS
  FROM
    `lego-june2024.lego_dataset.sets`
  JOIN
    `lego-june2024.lego_dataset.inventory_sets`
  ON
    =
  JOIN
    `lego-june2024.lego_dataset.inventory_parts`
  ON
    =
  JOIN
    `lego-june2024.lego_dataset.themes`
  ON
    =
  GROUP BY
  ORDER BY
    DESC;
```

| theme_name         | number_of_rare_sets |
|--------------------|---------------------|
| Bionicle           | 20                  |
| Star Wars          | 14                  |
| Knights Kingdom II | 12                  |
| Soccer             | 9                   |
| X-Pod              | 5                   |
| Technic            | 5                   |
| Clikits            | 5                   |
| Orient Expedition  | 4                   |
| Blacktron II       | 4                   |
| Classic            | 4                   |
| Mindstorms         | 4                   |



# What is the most popular color of lego in terms of parts produced

```
SELECT color_name, SUM(quantity) AS total_quantity
FROM (
  SELECT
    ip.color_id,
    ip.inventory_id,
    ip.part_num,
    CAST(ip.quantity AS numeric) AS quantity,
    ip.is_spare,
    c.name AS color_name,
    c.rgb,
    p.name AS part_name,
    p.part_material,
    pc.name AS category_name
  FROM redi-demo.Rebrickable.inventory_parts ip
  INNER JOIN redi-demo.Rebrickable.colors c ON ip.color_id = c.id
  INNER JOIN redi-demo.Rebrickable.parts p ON ip.part_num = p.part_num
  INNER JOIN redi-demo.Rebrickable.part_categories pc ON p.part_cat_id = pc.id
) AS inventories
GROUP BY inventories.color_name
ORDER BY total_quantity DESC
LIMIT 10
```

| Row | color_name ▼      | total_quantity ▼ |
|-----|-------------------|------------------|
| 1   | Black             | 786635           |
| 2   | Light Bluish Gray | 483407           |
| 3   | White             | 471018           |
| 4   | Dark Bluish Gray  | 345073           |
| 5   | Red               | 305367           |
| 6   | Blue              | 203962           |
| 7   | Yellow            | 203892           |
| 8   | Tan               | 167446           |
| 9   | Reddish Brown     | 143650           |
| 10  | Light Gray        | 105858           |

## How many sets were created in each century?

```
SELECT
    Century,
    COUNT(set_num) AS total_set_num
FROM
    (
        SELECT
            s.set_num,
            CASE
                WHEN s.year BETWEEN 1901 AND 2000 THEN '20th_Century'
                WHEN s.year BETWEEN 2001 AND 2100 THEN '21st_Century'
            END AS Century
        FROM
            redi-demo.Rebrickable.sets s
    ) AS sets_by_century
GROUP BY
    Century;
```

| Row | Century ▼    | total_set_num ▼ |
|-----|--------------|-----------------|
| 1   | 20th_Century | 5079            |
| 2   | 21st_Century | 17703           |