

## BAIS: 3200:000A

### Lego Database Project Proposal

Ethan Backhaus, Lance Regehr, Sebastian Cochran, Lucas Jensen, Vivian Lin

#### Introduction

LEGO is a multi-billion-dollar toy company that is internationally recognizable, but how do they maintain their dominance in the industry? In this project, we will conduct a descriptive analysis of LEGO's inventory to understand how they have evolved over time. By examining the composition of LEGO sets including the variety of pieces, colors, and themes we aim to uncover trends in LEGO's design philosophy and company evolution.

#### Data

For our proposed project, we analyzed data from a LEGO database set posted on Kaggle<sup>1</sup>, last updated in July 2017. The original dataset had eight different files which included colors, inventories, inventory parts, inventory sets, part categories, parts, sets, and themes. We trimmed this down to five files, removing inventory sets, part categories, and parts for simplicity, as those three branch off to existing entities. The three files removed are also not significant when answering the proposed questions. With the inventories, inventory parts, and sets files all accumulating to over our given data limit, we plan to randomly reduce the rows covered to ensure a significant spread in data while staying under the appropriate data limit. Specifically, in our Inventory\_Parts table, for each InventoryID, we trimmed its number of parts to at most five. This enables us to include all our sets in our queries and maintain the table's original purpose while cutting down the tables' exorbitant number of records. Included below is a data dictionary, with respective data types and descriptions.

---

<sup>1</sup> <https://www.kaggle.com/datasets/ratatman/lego-database?select=colors.csv>

*Table 1 Data Dictionary*

Field	Type	Description
InventoryID	Numeric	Unique ID for inventory taken at Lego
SetNum	Numeric	Lego set number
PartNum	Numeric	Unique ID for specific Lego part
ColorID	Numeric	Unique ID for color
Quantity	Numeric	Quantity of specific pieces included in set
Spare	Binary	T if part is spare, F otherwise

Name	Text	Name of the color
RGB	Text	Color mix using red, green, and blue
IsTranslucent	Binary	T if the color is transparent/translucent, F otherwise
Name	Numeric	Unique ID for specific Lego set
Year	Numeric	Year the set was released
ThemeID	Numeric	Unique ID for Lego theme
SetPieces	Numeric	Number of pieces included in Lego set
Name	Text	Name of the Lego theme

This data is comprised of 5 entities (COLORS, INVENTORIES, PIECES, INVENTORY\_Parts, THEMES) with unique identifiers being ColorID, InventoryID, SetNumber, ThemeID, while Inventory\_Parts has a composite id. All attributes are required because we are talking about inventories, and all data is integral to defining items in inventory.

We normalized the data and created a relational schema with 5 tables. The COLORS table is a reference table, and the Inventory Parts table uses a composite primary key made up of InventoryID, PartNumber, and ColorID. The Inventories table contains foreign key SetNumber, which references the SETS table. The ThemeID in the Sets table acts as a foreign key referencing the Themes table. This is because there is a one-to-many relationship between Themes and Sets, so ThemeID is placed in Sets table (the “many” sides of the relationship). Lastly, the ParentID in the Themes table is a recursive foreign key that references the ThemeID within the same table, creating a parent-child relationship between themes.

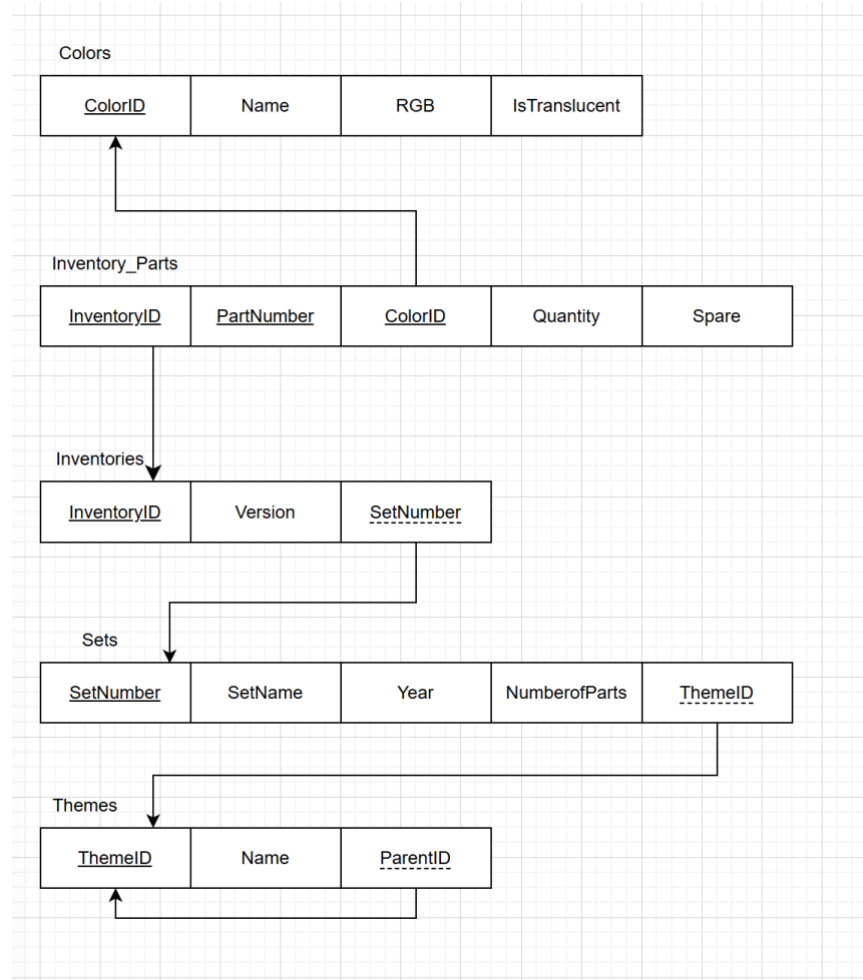


Figure 1 showcases the Relationship Schema between the entities

## Database Implementation

To create the database in Apex Oracle, we wrote CREATE TABLE commands for each table in the relational schema. As Themes has only a unary relationship, we created this table first.

### THEMES:

CREATE TABLE Themes (

ThemeID Number(\*,0) Not Null,

Name VARCHAR2(50) not null,

ParentID Number,

Constraint Themes\_PK Primary Key (ThemeID),

Constraint Themes\_FK Foreign Key (ParentID) References Themes(ThemeID)

);

Since Sets only references Themes, we created this table next.

SETS:

```
CREATE TABLE SETS(  
    SETNUM VARCHAR2(25) NOT NULL,  
    NAME VARCHAR2(50) NOT NULL,  
    YEAR NUMBER(4,0) NOT NULL,  
    THEMEID NUMBER(*,0) NOT NULL,  
    NUMBEROFPARTS NUMBER(4,0) NOT NULL,  
    CONSTRAINT SETS_PK PRIMARY KEY (SETNUM),  
    CONSTRAINT SETS_FK FOREIGN KEY (THEMEID) REFERENCES THEMES(THEMEID)  
);
```

We created Colors next.

COLORS:

```
CREATE TABLE Colors (  
    ColorID Number(*, 0) Not Null,  
    Name VARCHAR2(50) not null,  
    rgb VARCHAR2(8) not null,  
    Is_Translucent VARCHAR2(1) DEFAULT 'F' CHECK (UPPER(SPARE) IN ('T', 'F')) NOT NULL,  
    Constraint Color_PK Primary Key (ColorID)  
);
```

Next was Inventories.

INVENTORIES:

```
CREATE TABLE INVENTORIES(  
    INVENTORYID VARCHAR2(5) NOT NULL,  
    VERSION CHAR(1) NOT NULL,  
    SETNUM VARCHAR2(25) NOT NULL,  
    CONSTRAINT INVENTORY_PK PRIMARY KEY (INVENTORYID)
```

);

And finally we can create the Inventory\_Parts table as its referenced tables are created.

#### INVENTORY\_PARTS:

```
CREATE TABLE INVENTORY_PARTS (  
    INVENTORYID VARCHAR2(5) NOT NULL,  
    PARTNUM VARCHAR2(50) NOT NULL,  
    COLORID Number NOT NULL,  
    QUANTITY VARCHAR2(4) NOT NULL,  
    SPARE CHAR(1) DEFAULT 'F' CHECK (UPPER(SPARE) IN ('T', 'F')) NOT NULL,  
    CONSTRAINT INVENTORY_PARTS_PK PRIMARY KEY (PARTNUM, InventoryID, ColorID),  
    CONSTRAINT INVENTORY_PARTS_FK FOREIGN KEY (INVENTORYID) REFERENCES  
    INVENTORIES(INVENTORYID),  
    Constraint Inventory_Parts_FK2 Foreign Key (ColorID) REFERENCES Colors(ColorID)  
);
```

Across all tables, the appropriate data types and field sizes we selected create an environment of better data integrity. For example, the spare and Is\_Translucent fields in the Inventory\_Parts and Colors tables were limited to the 't' and 'f' characters to limit human input error. At times, our VARCHAR2 restriction in terms of characters was quite high. We made this decision as we found that many identifiers and names had high variations in length.

The cleaned and normalized data was inserted into our created tables using Apex's tools of uploading into these tables.

### **Analysis**

This analysis is intended to show Lego's evolution over time by summarizing the change over time of their most popular themes, their use of branded themes, the number of sets created, their sets' complexity, their use of different colors, and usage of their branded themes. This data can be useful to a variety of users. Lego itself may find this data interesting as they look to continue evolving in the future, potential investors may be interested in this data to see the future outlook of the toy company, and inherently, this analysis is very intriguing to anyone that's a Lego fan and curious about their favorite brand's evolution.

#### Question1: Most Popular Themes Over Time

*How have the most used themes changed over time?* We wanted to understand how Lego's themes have changed over time. We wrote two queries to compare Lego's most popular themes in the 1950s, their first decade in making sets, and the 2010s, our most recent decade of data.

We did a simple join query for both queries, first specifying the years less than 1960 and grouping by the sets name.

```
SELECT T.Name, COUNT(S.SetNum) AS NumberOfSets
```

```
FROM Themes T
```

```
JOIN Sets S ON S.ThemeID = T.ThemeID
```

```
WHERE S.Year < 1960
```

```
GROUP BY T.Name
```

```
ORDER BY NumberOfSets DESC
```

```
Fetch First 10 Rows Only;
```

The results of the query are shown below (Figure 2). Based on the Lego database, we found that in the 1950s there were only 6 themes used, with the most used themes being Supplemental and Town Plan. This is expected as Lego was just starting to release sets within this decade. From our dataset we found that Supplemental and Town Plan were foundational themes for Lego.

NAME	NUMBEROFSETS
Supplemental	88
Town Plan	35
Basic Set	5
Books	2
Mosaic	1
HO 1:87 Vehicles	1

*Figure 2 1950s Number of Sets Per Theme*

The next query is similar but not we are looking into the most recent decade of data, the 2010s.

```
SELECT T.Name, COUNT(S.SetNum) AS NumberOfSets
```

```
FROM Themes T
```

```
JOIN Sets S ON S.ThemeID = T.ThemeID
```

```
WHERE S.Year > 2010
```

```
GROUP BY T.Name
```

```
ORDER BY NumberOfSets DESC
```

Fetch First 10 Rows Only;

The results of the query are shown below in Figure 3. In our dataset, we found that the number of themes used and the number sets of per theme grew. Even though we only took the top 10 rows, we say the number of themes grew from 6 to 10 as well as our most used theme grew from 88 sets to 269. Overall, the composition of themes has changed entirely. No theme from the 1950s cracks the top 10 used themes of the most current decade. The most common themes in the 2010s are now full of collaborations with large media brands like Star Wars, Ninjago, and Batman. Comparing these two results delivers an insight into how Lego has evolved their themed sets.

NAME	NUMBEROFSETS
Friends	269
Star Wars	193
Gear	191
Ninjago	191
Duplo	157
Creator	125
City	118
Legends of Chima	84
Key Chain	80
Batman	79

*Figure 3 2010s Number of Sets Per Theme*

#### Question 2: Amount of Sets Created Over Time

*Has Lego increased their output of sets over their company's history?* Next, we wanted to show how Lego has grown over the years. We quantified this by looking for how many sets Lego has released per year till modern day.

In this query, we used the CASE function which enabled us to categorize the number of sets released over time into decades.

SELECT

CASE

WHEN year < 1960 THEN '1950s'

WHEN year < 1970 THEN '1960s'

WHEN year < 1980 THEN '1970s'

WHEN year < 1990 THEN '1980s'

WHEN year < 2000 THEN '1990s'

```

    WHEN year < 2010 THEN '2000s'

    ELSE '2010s'

END AS Decade,

COUNT(setnum) AS NumberOfSetsReleased

FROM Sets

GROUP BY

CASE

    WHEN year < 1960 THEN '1950s'

    WHEN year < 1970 THEN '1960s'

    WHEN year < 1980 THEN '1970s'

    WHEN year < 1990 THEN '1980s'

    WHEN year < 2000 THEN '1990s'

    WHEN year < 2010 THEN '2000s'

    ELSE '2010s'

END

ORDER BY COUNT(setnum) DESC;

```

The results of the query are shown in Figure 4. Looking at the query output, we can see that the number of sets released per decade has increased in every consecutive decade. This result gives us an insight into the overall growth of the Lego brand. In every decade, Lego invests in more creative resources to come up with more sets. This is a positive sign for the company, its shareholders, and potential investors.

DECADE	NUMBEROFSETSRELEASED
2010s	4425
2000s	3584
1990s	1635
1980s	1030
1970s	564
1960s	303
1950s	132

*Figure 4 Number of Sets Released Per Decade*

### Question 3: Set Complexity Over Time

*Have Lego sets become more complex throughout time?* We were also interested in the complexity of sets over time. To measure this, we used the metric of parts per sets.

We just used a simple query.

```
Select Year, Round(Sum(NumberOfParts) / Count(SetNum), 2) As PartsPerSet
```

```
From Sets
```

```
Group By Year
```

```
Order By Year;
```

A portion of the results of the query are shown below in Figure 5. In the results, we can see that the number of parts per set has increased relatively consistently over the history of Lego. Once again, we can see an interesting insight into the evolution of their set composition. This insight into the evolution of their parts per set metric seems to give evidence of their growth and their consumer base. It suggests that their more recent consumer base enjoys solving more complex sets that require more time and have more pieces. Like the increasing number of sets released, it's promising to see that they are investing resources into creating more pieces in every set and potentially catering to a wider consumer base.

YEAR	PARTSPERSET
1950	10.14
1953	16.5
1954	12.36
1955	36.86
1956	18.5
1957	42.62
1958	44.45
1959	16.25
1960	175.33
1961	70.59
1962	81.75

*Figure 5 Number of Parts Per Set Per Year*

#### Question 4: Set Colorfulness Over Time

*How has Lego's use of different colors evolved over time?* We were interested in how Lego has evolved in how many colors they use in their newly released sets per year.

We just used a simple join query.

```
Select Year, Count(Distinct(C.Name)) As ColorsUsed
```

From Sets S Join Inventories I On S.SetNum = I.SetNum Join Inventory\_parts IP ON I.InventoryID = IP.InventoryID Join Colors C On IP.ColorID = C.ColorID

Group By Year

Order By Year Desc;

A portion of the results of the query are shown below in Figure 6. In this query, we saw a slight increase over time in the number of colors used in newly released sets every year until quick growth in the late 1990s. After a peak in 2005, this trend tails off, leading to a more decayed oscillation as the data leads into modern day. This data once again gives us insight into the set creation and its trends. It appears there was a deliberate move by Lego to make their sets more colorful leading into the 2000s. The data makes you wonder why. Could it have been a social trend, the creation of multitudes of new sets which required more diverse pieces (this could be the case looking at Figure 5), or just a new shift in branding?

YEAR	COLORSUSED
2017	54
2016	61
2015	60
2014	56
2013	57
2012	62
2011	59
2010	51
2009	45
2008	42
2007	49

Figure 6 Number of Parts Per Set Per Year

#### Question 5: Number of Ninjago Sets Over Time

*How has Lego used their Ninjago brand over time in their sets?* Finally, we were interested in how Lego has used their original themes over time. We focused on Ninjago, a Lego original theme that has competed in popularity with the likes of Star Wars and Batman by being a popular kids show on Cartoon Network.

We just used simple join query.

Select T.Name, Year, Count(S.SetNum) As NumberOfSets

From Themes T Join Sets S On S.ThemeID = T.ThemeID

Group By T.Name, Year

Having T.Name = 'Ninjago'

Order by Year Asc;

The results are shown below in Figure 7. Using this query, we found that the use of this original theme seemed to fluctuate over the 8 years it was a theme. There is a peak in 2011-12 and 2015-16 with lulls in the other years. This unsteadiness seems to allude to the fact that Lego might be testing how their own original theme does as a Lego set. Another reason could be that they release sets coinciding new seasons of the popular show, but more research would better clarify that. Overall, this query gives an interesting insight into how Lego uses one of their own specific original theme.

NAME	YEAR	NUMBEROFSETS
Ninjago	2011	42
Ninjago	2012	47
Ninjago	2013	8
Ninjago	2014	8
Ninjago	2015	46
Ninjago	2016	32
Ninjago	2017	8

*Figure 7 Number of Parts Per Set Per Year*

## Web Application

[https://apex.oracle.com/pls/apex/r/database\\_final\\_project/lego-database272847/home?session=963576091235](https://apex.oracle.com/pls/apex/r/database_final_project/lego-database272847/home?session=963576091235)

### Home Page

The home page of the web application contains a brief description of what the APP contains (along with a hyperlink to the original data source), a main navigation menu that is organized with nest elements and custom icons, and an image of Lego figures to greet users we found on google. Throughout the entire application we kept a consistent accent theme that resembles Lego. Additionally, we customized the icons within the navigation menu to match with the visualization displayed on the page along with a dropdown menu to group the tables under one heading. Figure 8 shows a screenshot of the home page and navigation menu.

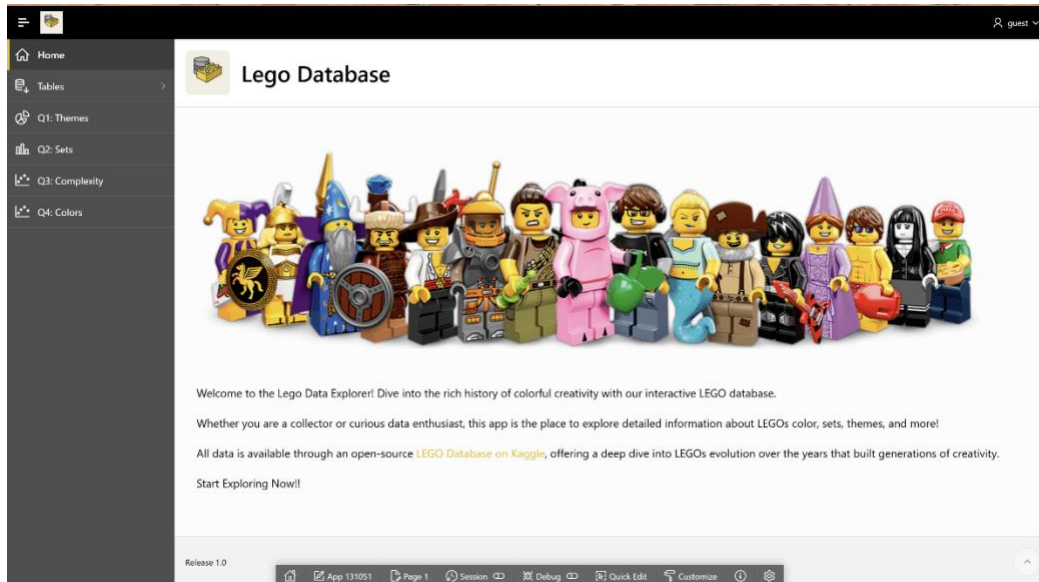


Figure 8 Home Page

## Tables

We created an interactive report for each database table what allows the users to filter, search and group data (Figures 9-13). The search bar can be used to look up all columns relating to the keyword. The filter in the actions button can be used to filter more specific data like which column to look for the keyword in. Column headings and number format have been adjusted for easy interpretation. Each page also includes a textbox above the table that describes the data.

Colors Table			
The Colors Table shows the ColorID, Color, RGB, and whether or not the LEGO is translucent. It includes all of the colors LEGO used from 1950 to 2017. This table has a total of 135 rows.			
<input type="text"/> <input type="button" value="Go"/> <input type="button" value="Actions"/>			
ColorID	Color	RGB	Is Translucent
-1	Unknown	003382	f
0	Black	05131D	f
1	Blue	0055BF	f
2	Green	237841	f
3	Dark Turquoise	008F98	f
4	Red	C91A09	f
5	Dark Pink	C870A0	f
6	Brown	583927	f
7	Light Gray	9BA19D	f
8	Dark Gray	6D6E5C	f
9	Light Blue	B4D2E3	f
10	Bright Green	4B9F4A	f
11	Light Turquoise	55A5AF	f
12	Salmon	F2705E	f
13	Pink	FC97AC	f
14	Yellow	F2CD37	f

Figure 9 Colors

Inventory Table		
The Inventories table below shows information regarding the InventoryID, Version, and the Set_Num. The table has a total of 11,675 rows with data from 1950 to 2017.		
<input type="text"/> <input type="button" value="Go"/> <input type="button" value="Actions"/>		
InventoryID	Version	Set_Num
1649	1	66436-1
1651	1	6201-1
1653	1	7574-9
1654	1	8831-12
1655	1	7687-5
1656	1	6510-1
1657	1	118-2
1658	1	3603-1
1660	1	79002-1
1661	1	4448-1
1662	1	3885-1
1664	1	8614-2
1665	1	3316-16
1666	1	8274-1
1668	1	7103-1

Figure 10 Inventory

Inventory_Parts					
The Inventory Parts tables includes InventoryID, Part_Num, ColorID, Quantity and Spare. This table has a total of 49,754 rows with data from 1950 to 2017					
<input type="text"/> <input type="button" value="Go"/> <input type="button" value="Actions"/>					
InventoryID	Part_Num	ColorID	Quantity	Spare	
452	3005	14	1	f	
454	23306	383	1	f	
454	2412b	71	6	f	
454	2413	71	1	f	
454	2420	15	8	f	
454	2431	4	4	f	
457	11477	0	2	f	
457	11477	191	2	f	
457	15392	72	1	f	
457	15392	72	1	t	
457	15403	0	1	f	
459	2301	1	2	f	
459	2302	10	2	f	
459	3011	10	3	f	
459	3011	1	3	f	
459	3011	14	3	f	

Figure 11 Inventory Parts

Sets				
The Sets table includes Set_Num, Name, Year, ThemeID, and Number_of_Parts. The table has 11,673 rows with data from 1950 to 2017.				
<div> <input type="text"/> <input type="button" value="Go"/> <input type="button" value="Actions"/> </div>				
Set_Num	Name	Year	ThemeID	Number_Of_Parts
1076-22	Advent Calendar 1999 (Day 21) Police Car	1999	217	11
1076-23	Advent Calendar 1999 (Day 22) Dog with Red Hat	1999	217	12
1076-24	Advent Calendar 1999 (Day 23) Police Helicopter	1999	217	10
1076-25	Advent Calendar 1999 (Day 24) Santa	1999	217	11
1076-3	Advent Calendar 1999 (Day 2) Snowman	1999	217	11
1076-4	Advent Calendar 1999 (Day 3) Speedboat	1999	217	7
1076-5	Advent Calendar 1999 (Day 4) Girl	1999	217	8
1076-6	Advent Calendar 1999 (Day 5) Sailboat	1999	217	9
1076-7	Advent Calendar 1999 (Day 6) Reindeer	1999	217	12
1076-8	Advent Calendar 1999 (Day 7) Plane	1999	217	10
1076-9	Advent Calendar 1999 (Day 8) Girl	1999	217	8
1077-1	Supplementary Set	1976	528	170
10801-1	Baby Animals	2016	504	13
10802-1	Savanna	2016	504	15
10803-1	Arctic	2016	504	32
10804-1	Jungle	2016	504	86

Figure 12 Sets

Themes		
The Themes table below shows the ThemeID, name and parentID. It records all of the themes LEGO has done from 1950 to 2017. The table has 614 rows.		
<div> <input type="text"/> <input type="button" value="Go"/> <input type="button" value="Actions"/> </div>		
ThemeID	Name	ParentID
1	Technic	
2	Arctic Technic	1
3	Competition	1
4	Expert Builder	1
5	Model	1
6	Airport	5
7	Construction	5
8	Farm	5
9	Fire	5
10	Harbor	5
11	Off-Road	5
12	Race	5
13	Riding Cycle	5
14	Robot	5
15	Traffic	5
16	RoboRiders	1

Figure 13 Themes

## Queries

We presented the results of all our queries based on several questions. The first question was “What was the most popular theme throughout the decades?” (Figure 14). This page includes two pie charts. These pie charts describe the top 10 most popular themes of sets that were released in the 1950s and 2010s along with the percentage each theme took up. We included a text box at the top to describe our findings on the results based on our research question.

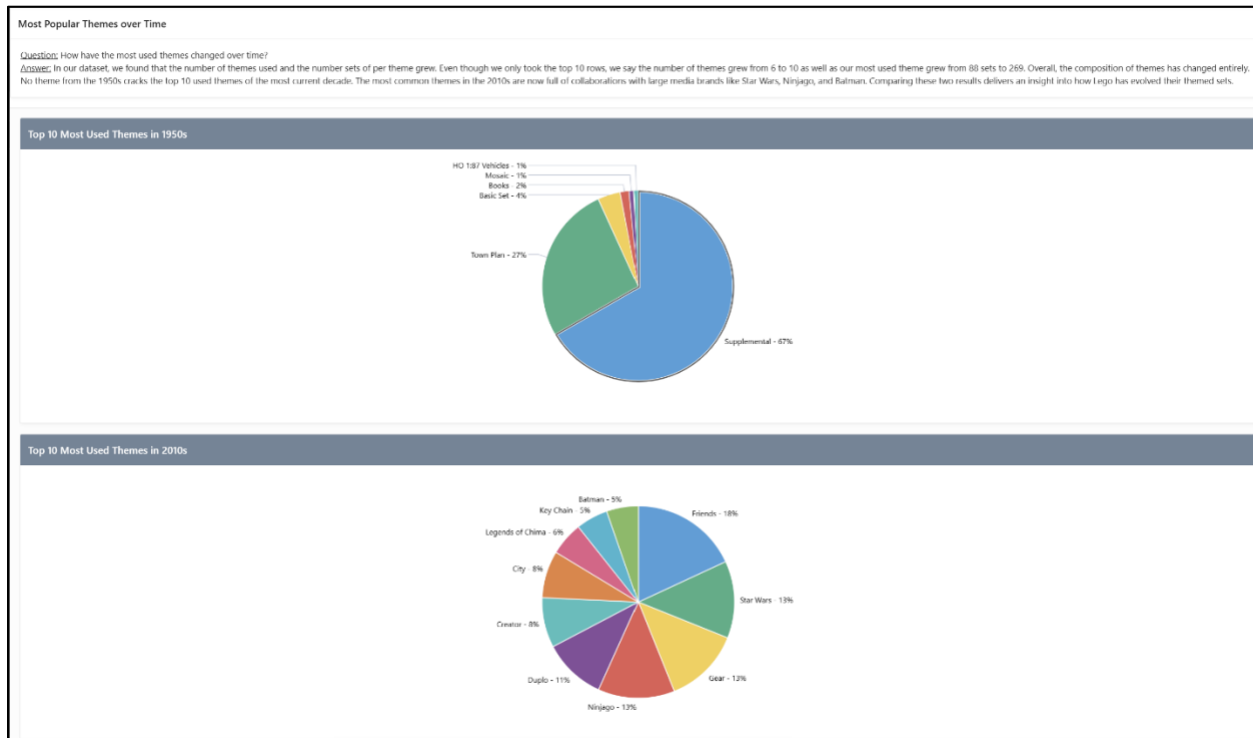


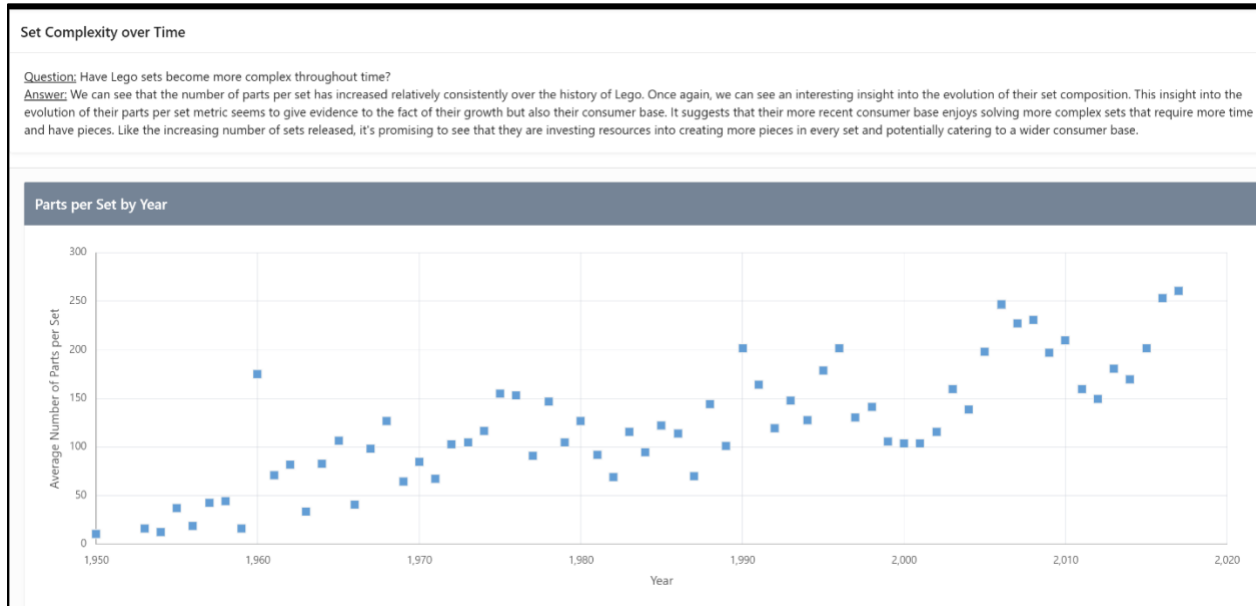
Figure 14 Popular Themes over time

Our second research question was about the number of sets throughout the years (Figure 15). The page includes a bar chart that shows the number of sets from 1950s to 2010s. The page also includes a text box that includes our insights on our findings.



Figure 15 Number of Sets Created Over Time

The third research question examines the set complexity over time (Figure 16). The page includes a scatterplot that showcases the number of parts in a set over time. There is also a textbox above the visualization that describes the plot along with insights given by the result.



*Figure 16 Set Complexity over Time*

Our fourth research question analyzes how the colorfulness in the set changes over time (Figure 17). The results were presented in a scatterplot with number of colors on the y axis and the years on the x axis. We also included a text box that provides our insights on our findings based on our results.

### Set Colorfulness over Time

**Question:** How has Lego's use of different colors evolved over time?

**Answer:** From the scatterplot below, we saw a slight increase over time in the number of colors used in newly released sets every year until quick growth in the late 1990s. After a peak in 2005, this trend tails off, leading to a more decayed oscillation as the data leads into modern day. This data once again gives us insight into the set creation and its trends. It appears there was a deliberate move by Lego to make their sets more colorful leading into the 2000s. The data makes you wonder why. Could it have been a social trend, the creation of multitudes of new sets which required more diverse pieces (this could be the case looking at Q2), or just a new shift in branding?

#### Most Colors Used per Year

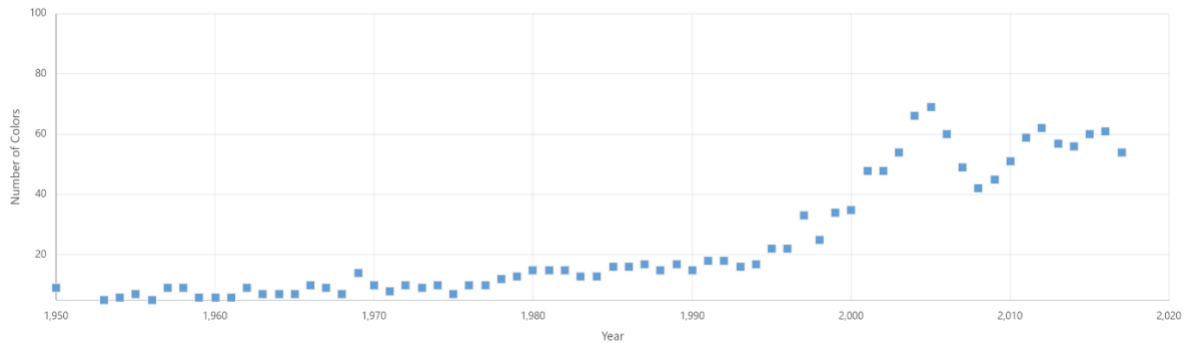


Figure 17 Set Colorfulness Over Time

Our final research question looks to see how Lego has used its Ninjago brand over time (Figure 18). The results were displayed in a bar chart with the year on the x axis and the number of sets released on the y axis. We also included a text box that provides our insights on our findings based on our results.

### Ninjago Brand Usage Over Time

**Question:** How has Lego used their Ninjago brand over time in their sets?

**Answer:** The results are shown below. Using this query, we found that the use of this original theme seemed to fluctuate over the 8 years it was a theme. There is a peak in 2011-12 and 2015-16 with lulls in the other years.

This unsteadiness seems to allude to the fact that Lego might be testing how their own original theme does as a Lego set. Another reason could be that they release sets coinciding new seasons of the popular show, but more research would better clarify that.

Overall, this query gives an interesting insight into how Lego uses one of their own specific original theme.

#### Ninjago Sets Over Time

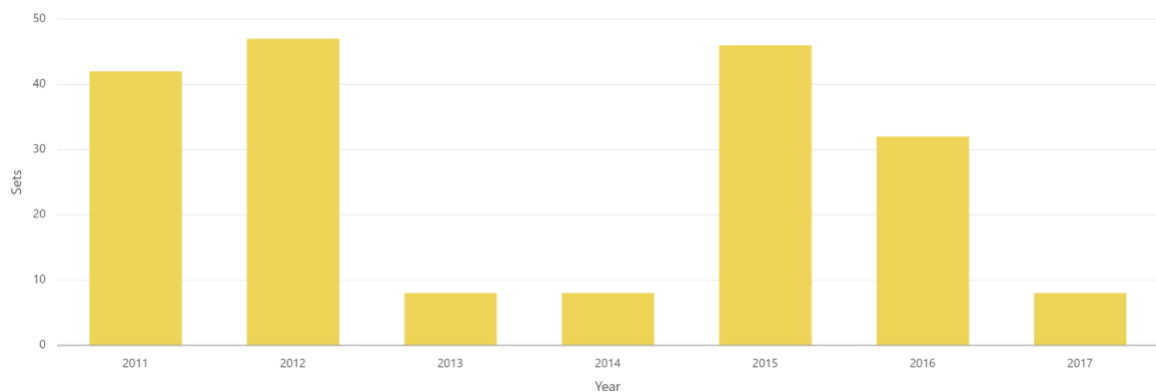


Figure 18 Ninjago Sets Over Time