BedaDSC550.Week2

March 20, 2024

This LEGO dataset contains information about various LEGO sets over the years, including their set number, name, year released, number of parts, image URL, and theme name. LEGO sets have been a popular toy for decades, offering creativity, building and fun for people of all ages. This dataset provides data insights of LEGO sets over time.

The questions to explore are:

How has the number of LEGO sets released evolved over the years?

Which themes have been the most prevalent in LEGO set releases?

Is there a relationship between the number of parts in a LEGO set and its year of release?

[6]:

```python
import pandas as pd
import urllib.request
import pandas as pd
import matplotlib.pyplot as plt
```

[13]:

```python
# Load the dataset into a DataFrame
url = "https://raw.githubusercontent.com/cheribeda/datamining/main/
 ↪LEGO_Sets%20_Themes%20.csv"
response = requests.get(url)

# Save the content to a file
with open('LEGO_Sets_Themes.csv', 'wb') as file:
    file.write(response.content)

# Read the dataset into a DataFrame
df = pd.read_csv('LEGO_Sets_Themes.csv')

# Display the first few rows of the DataFrame
print(df.head())
```

```
  set_number                         set_name  year_released  number_of_parts  \
0      001-1                            Gears           1965               43
1      002-1   4.5V Samsonite Gears Motor Set           1965                3
2     1030-1  TECHNIC I: Simple Machines Set           1985              210
3     1038-1              ERBIE the Robo-Car           1985              120
4     1039-1             Manual Control Set 1           1986               39

                                           image_url theme_name
```

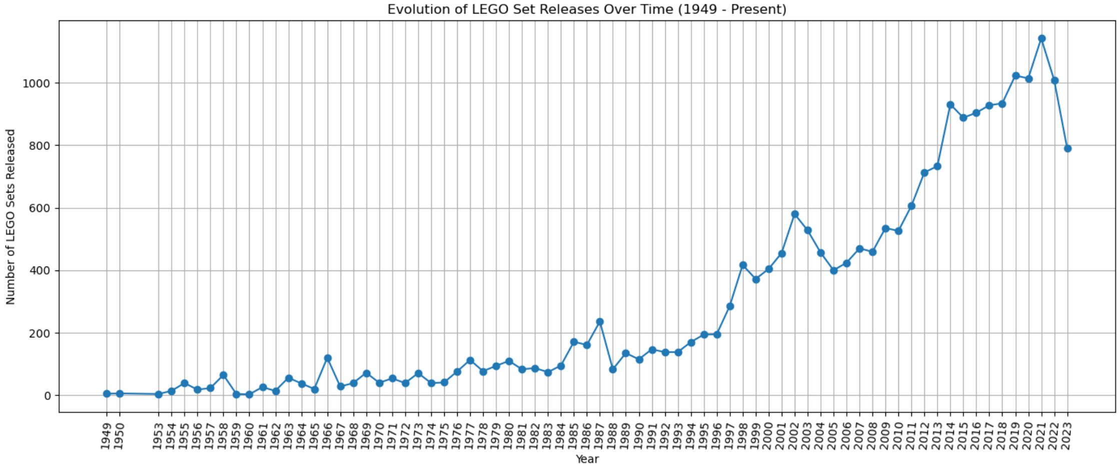
```
0   https://cdn.rebrickable.com/media/sets/001-1.jpg    Technic
1   https://cdn.rebrickable.com/media/sets/002-1.jpg    Technic
2  https://cdn.rebrickable.com/media/sets/1030-1.jpg    Technic
3  https://cdn.rebrickable.com/media/sets/1038-1.jpg    Technic
4  https://cdn.rebrickable.com/media/sets/1039-1.jpg    Technic
```

[24]:

```
# Filter data for years starting from 1949 to current year
df_filtered = df[df['year_released'] >= 1949]

# Count the number of sets released each year
set_count_per_year = df_filtered['year_released'].value_counts().sort_index()

# Create a line plot
plt.figure(figsize=(14, 6))
plt.plot(set_count_per_year.index, set_count_per_year.values, marker='o', linestyle='-')
plt.title('Evolution of LEGO Set Releases Over Time (1949 - Present)')
plt.xlabel('Year')
plt.ylabel('Number of LEGO Sets Released')
plt.grid(True)
plt.xticks(set_count_per_year.index, rotation=85)  # Rotate x-axis labels for better readability
plt.tight_layout()  # Adjust layout to prevent crowding of labels
plt.show()
```



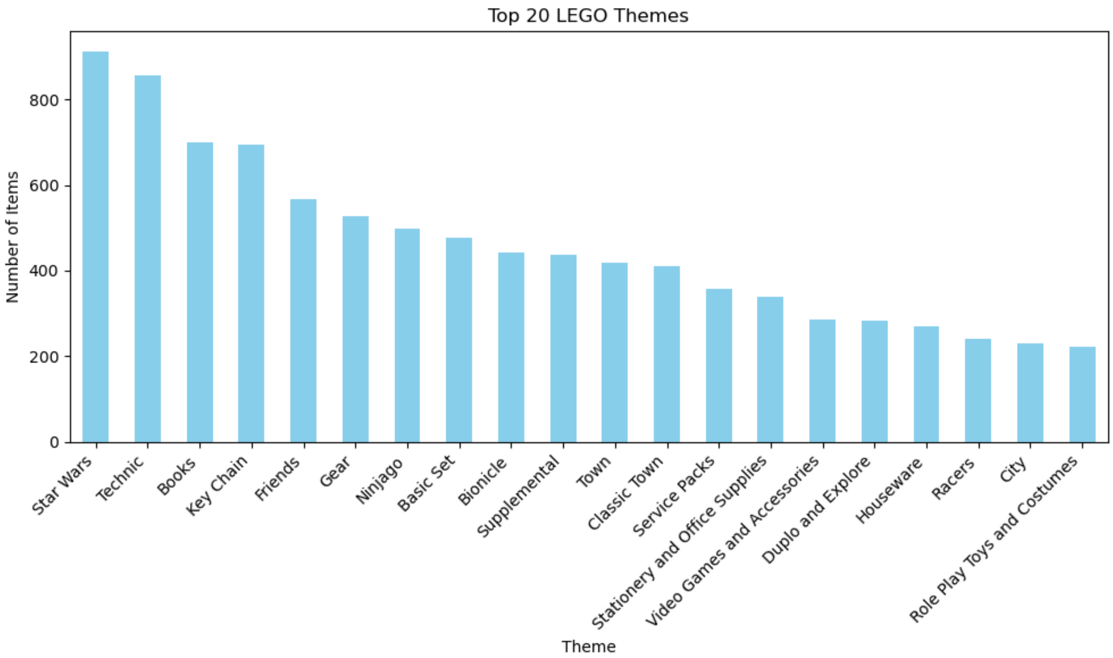
This graph show the evolution of LEGO as the toys popularity grew the number of sets released grew. It appears that the number of different sets started to steadily increase in 1996 with an all time high in 2021.

[47]:

```
# Count the number of sets for each theme
theme_counts = df['theme_name'].value_counts()
```

```
# Select the top 20 most prevalent themes
top_n_themes = 20
top_themes = theme_counts.head(top_n_themes)

# Plotting the bar chart
plt.figure(figsize=(10, 6))
top_themes.plot(kind='bar', color='skyblue')
plt.title('Top 20 LEGO Themes')
plt.xlabel('Theme')
plt.ylabel('Number of Items')
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for better␣
 ↪readability
plt.tight_layout()  # Adjust layout to prevent crowding of labels
plt.show()
```



This bar chart displays the top LEGO themes. Star Wars emerges as the clear winner, closely followed by Technic. Upon reviewing the data, I found categories described as books, which are not sets but books based on LEGO characters or instructional guides on building. Housewares include items like lunchboxes and drinkware. Costumes and role play are attire and accessories for imaginative play. This bar chart illustrates the popularity of sets and the appeal of additional merchandise, contributing to the brand's success.

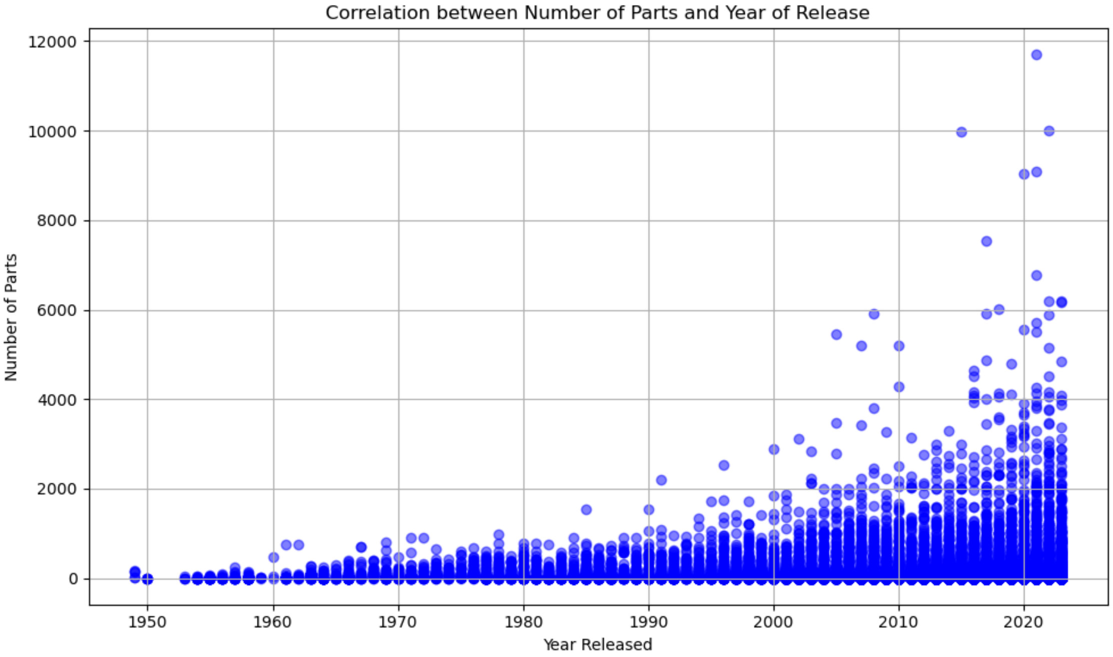
[50]:

```
# Filter out any rows with missing values in 'number_of_parts' or␣
 ↪'year_released' columns
```

```
df = df.dropna(subset=['number_of_parts', 'year_released'])

# Filter the DataFrame to include years from 1949 to current year
df_filtered = df[(df['year_released'] >= 1949) & (df['year_released'] <= pd.
 ↪Timestamp.now().year)]

# Create a scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df_filtered['year_released'], df_filtered['number_of_parts'],␣
 ↪alpha=0.5, color='blue')
plt.title('Correlation between Number of Parts and Year of Release')
plt.xlabel('Year Released')
plt.ylabel('Number of Parts')
plt.grid(True)
plt.tight_layout()
plt.show()
```



The scatterplot illustrates LEGO's ongoing expansion of set complexity over the years. The set with the highest number of pieces, the World Map released in 2021, and contains 11,695 pieces. However, such large sets appear as outliers, with the majority of sets containing under 2,000 pieces. Overall there appears to be a correlation between the year of release and the number of parts, indicating a trend toward increasingly intricate LEGO sets over time.

Conclusion: The analysis of LEGO sets and merchandise showcases the brand's expansion over time. The line plot illustrating the evolution of LEGO set releases indicates a steady growth trajectory notably since 1996, with a peak in 2021. This indicates LEGO's enduring popularity and market expansion. Star Wars appears as the top theme in the bar chart of top LEGO themes, closely followed by Technic, reflecting the brand's strategic partnerships and diversified product offerings, including books, housewares, and costumes, which contribute to LEGO's success. The scatterplot revealing a positive correlation between the year of release and the number of parts highlights LEGO's continuing innovation. The release of the World Map set in 2021, features 11,695 pieces illustrating the increasing complexity of the LEGO sets over time. These large sets stand as outliers, the majority of sets contain under 2,000 pieces, suggesting a balance between intricate designs and accessibility. Overall, these findings highlight LEGO's journey of adaptation, creativity, and enduring appeal and committment to imaginative play.

[ ]: