Exploring 67 years of LEGO with pandas in python

Mohsen Mirhashemi

August 24, 2020

Project definition

The Rebrickable database includes data on every LEGO set that ever been sold; the names of the sets, what bricks they contain, what color the bricks are, etc. It might be small bricks, but this is big data! In this project, you will get to explore the Rebrickable database. To do this you need to know your way around pandas dataframes and it's recommended that you have a medium knowledge of python and data manipulation with pandas library.

Project Tasks

In this project, we will analyze a fascinating dataset on every single **lego** block that has ever been built!

1. **Understanding data**: A comprehensive database of lego blocks is provided by Rebrickable. The data is available as csv files and the schema is shown below.

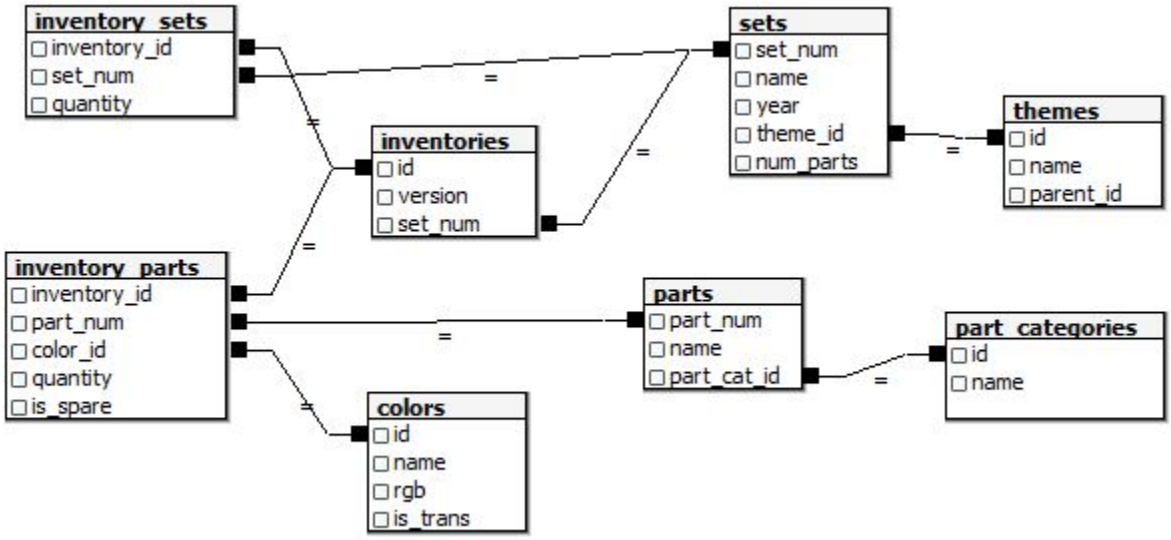

Figure 1: schema of logo database.

2. **Reading data**: Read the color and sets dataset and display the first 10 rows of each dataframe.

3. **Exploring Colors**: Now that we have read the colors data, we can start exploring it! Let us start by understanding the number of colors available. How many distinct colors are available? use a bar plot to visualize your finding. **tip**: you can use countplot in seaborn library.

4. **Transparent Colors in Lego Sets**: The colors data has a column named is_trans that indicates whether a color is transparent or not. It would be interesting to explore the distribution of transparent vs. non-transparent colors. how distributed colors base on transparency?

5. **Explore Lego Sets**: Another interesting dataset available in this database is the sets data. It contains a comprehensive list of sets over the years and the number of parts that each of these sets contained.

	set_num	name	year	theme_id	num_parts
0	00-1	Weetabix Castle	1970	414	471
1	0011-2	Town Mini-Figures	1978	84	12
2	0011-3	Castle 2 for 1 Bonus Offer	1987	199	2
3	0012-1	Space Mini-Figures	1979	143	12
4	0013-1	Space Mini-Figures	1979	143	12

Figure 2: Sets dataframe.

first create a summary of average number of parts by year and **second** plot trends in average number of parts by year.

6. **Lego Themes Over Years**: Lego blocks ship under multiple themes. Let us try to get a sense of how the number of themes shipped has varied over the years. display number of themes shipped by year.

Good luck