notebook

January 14, 2018

0.1 Introduction

Everyone loves Lego (unless you ever stepped on one). Did you know by the way that "Lego" was derived from the Danish phrase leg godt, which means "play well"? Unless you speak Danish, probably not.

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

```
In [4]: # Nothing to do here.
```

0.2 Reading Data

This comprehensive database of lego blocks is provided by Rebrickable. The data is available as csv files and the schema is shown below.

Let us start by reading in the colors data to get a sense of the diversity of lego sets!

```
In [6]: # Import modules
        import pandas as pd

        # Read colors data
        colors = pd.read_csv('datasets/colors.csv')

        # Print the first few rows
        print(colors.head())
```

```
   id            name     rgb is_trans
0  -1         Unknown  0033B2        f
1   0           Black  05131D        f
2   1            Blue  0055BF        f
3   2           Green  237841        f
4   3  Dark Turquoise  008F9B        f
```

0.3 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.

```
In [8]: # How many distinct colors are available?
        num_colors = colors.name.count()

        print(num_colors)
```

```
135
```

0.4 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.

```
In [10]: # colors_summary: Distribution of colors based on transparency
         colors_summary = colors.groupby(colors.is_trans).count()

         print(colors_summary)
```

```
           id  name  rgb
is_trans
f         107   107  107
t          28    28   28
```

```
In [11]: print(colors_summary.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2 entries, f to t
Data columns (total 3 columns):
id      2 non-null int64
name    2 non-null int64
rgb     2 non-null int64
dtypes: int64(3)
memory usage: 64.0+ bytes
None
```

0.5 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.

Let us use this data to explore how the average number of parts in lego sets has varied over the years.

```
In [13]: import matplotlib.pyplot as plt

         sets = pd.read_csv('datasets/sets.csv')

         parts_by_year = sets[["year","num_parts"]].\
```

```
groupby(sets.year,as_index = False).\
mean()

# Plot trends in average number of parts by year

parts_by_year.plot(x = 'year',y= 'num_parts')
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60718e59b0>
```

0.6 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.

```
In [15]: # themes_by_year: Number of themes shipped by year

         themes_by_year = sets[['year', 'theme_id']].\
           groupby('year',as_index = False).\
           agg({"theme_id": pd.Series.count})
         themes_by_year.mean()
         print(themes_by_year)
```

```
    year  theme_id
0   1950         7
1   1953         4
2   1954        14
3   1955        28
4   1956        12
5   1957        21
6   1958        42
7   1959         4
8   1960         3
9   1961        17
10  1962        40
11  1963        18
12  1964        11
13  1965        10
14  1966        89
15  1967        21
16  1968        25
17  1969        69
18  1970        29
19  1971        45
20  1972        38
21  1973        68
22  1974        39
23  1975        31
24  1976        68
25  1977        92
```

```
26  1978        73
27  1979        82
28  1980        88
29  1981        79
..   ...       ...
36  1988        68
37  1989       114
38  1990        85
39  1991       106
40  1992       115
41  1993       111
42  1994       128
43  1995       128
44  1996       144
45  1997       194
46  1998       325
47  1999       300
48  2000       327
49  2001       339
50  2002       447
51  2003       415
52  2004       371
53  2005       330
54  2006       283
55  2007       319
56  2008       349
57  2009       403
58  2010       444
59  2011       502
60  2012       615
61  2013       593
62  2014       715
63  2015       670
64  2016       609
65  2017       470

[66 rows x 2 columns]
```

0.7 Wrapping It All Up!

Lego blocks offer an unlimited amoung of fun across ages. We explored some interesting trends around colors, parts and themes.

```
In [17]: # Nothing to do here
```