Mark Kipper

Final Report:

How to Identify the Theme of a Lego set

Problem Statement

Lego is a plastic, self-locking brick system patented by the Lego Group in 1958. Beginning in 1978, the Lego Group began marketing a series of sets based on the present, past, and future. These were termed a “system within the system” and led to the birth of the idea of Lego themes. The first themes were Town, Castle, and Space (representing the present, past, and future, respectively.) All sets released prior to this were labeled as the “System” theme. Since then, new themes have been introduced as older, less popular themes discontinued. 1999 saw the introduction of licensed themes such as Star Wars. There are currently 444 themes in the Lego database. This data includes all sets made since 1949.

As a father of 5, and as a personal Lego fan, I am intimately familiar with Lego. The greatest thing about Lego is that anything constructed can be taken apart again, and the pieces reused to make new things. As a collection of Lego grows, collecting various sets from themes that interest or as gifts from holidays or birthdays, and as instructions and boxes get lost or thrown away, one is left with a myriad of pieces in many colors. What can you build with this? Can you identify which theme certain parts belong to? Can these parts be used to build other themed sets?

By using the Lego Rebrickable database, I have created a tool to identify the theme of a Lego set based the number of parts, the color of the parts, the year it was produced, etc. This can be used by the company for more targeted generic building sets based on the more popular themes. Also, I was recently informed of an app that scans your loose Lego parts and generates building instructions for new build ideas. Using my tool, a Theme element can be incorporated into this app and it can generate theme-specific building ideas based on the parts scanned.

Data Wrangling

Exploratory Data Analysis

In-Depth Analysis

Model Selection

Results

Future Work