Welcome to the Python project Exploring the Evolution of Lego!

You are a Data Analyst at Lego working with the Sales/Customer Success teams. The Account Executive responsible for the Star Wars partnership has asked for specific information in preparation for their meeting with the Star Wars team. Although Star Wars was critical to the survival of the brand, Lego has since introduced a wide variety of licensed sets over subsequent years.

Your two questions are as follows:

Q1. What percentage of all licensed sets ever released were Star Wars themed? Save your answer as a variable the_force in the form of an integer (e.g. 25).

Q2. In which year was Star Wars not the most popular licensed theme (in terms of number of sets released that year)? Save your answer as a variable new_era in the form of an integer (e.g. 2012). The method through which you approach this question is up to you, but one thing to keep in mind is that the dataset is not necessarily clean, and may require the removal rows where there are values missing from critical columns.

Q3. Breakdown the data_set into sets launched per year and store them in a new dataframe called : sets_per_year.

Information About the datasets:

Let's look at Lego sets!

Lego is a household name across the world, supported by a diverse toy line, hit movies, and a series of successful video games. In this project, we are going to explore a key development in the history of Lego: the introduction of licensed sets such as Star Wars, Super Heroes, and Harry Potter.

It may not be widely known, but Lego has had its share of ups and downs since its inception in the early 20th century. This includes a particularly rough period in the late 90s. As described in [this article](), Lego was only able to survive due to a successful internal brand (Bionicle) and the introduction of its first licensed series: Star Wars. In the instructions panel are the two questions you will need to answer to complete this project.

Before diving into our analysis though, let's become familiar with the two datasets that will help you with this project:

datasets/lego_sets.csv

- **set_num:** A code that is unique to each set in the dataset. ***This column is critical, and a missing value indicates the set is a duplicate or invalid!***
- **set_name:** A name for every set in the dataset (note that this can be the same for different sets).
- **year:** The date the set was released.
- **num_parts:** The number of parts contained in the set. ***This column is not central to our analyses, so missing values are acceptable.***
- **theme_name:** The name of the sub-theme of the set.
- **parent_theme:** The name of the parent theme the set belongs to. Matches the `name` column of the `parent_themes` csv file.

datasets/parent_themes.csv

- **id:** A code that is unique to every theme.
- **name:** The name of the parent theme.
- **is_licensed:** A Boolean column specifying whether the theme is a licensed theme.