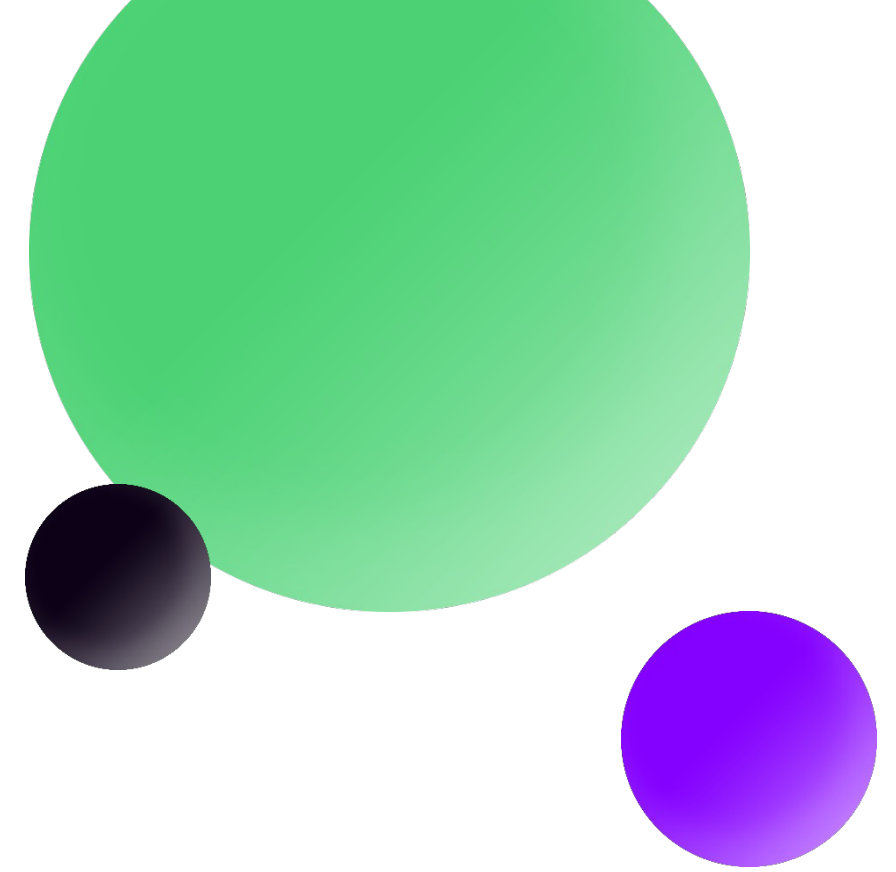


# Python en Pandas

## Case



**wortell**

# Introductie - LEGO

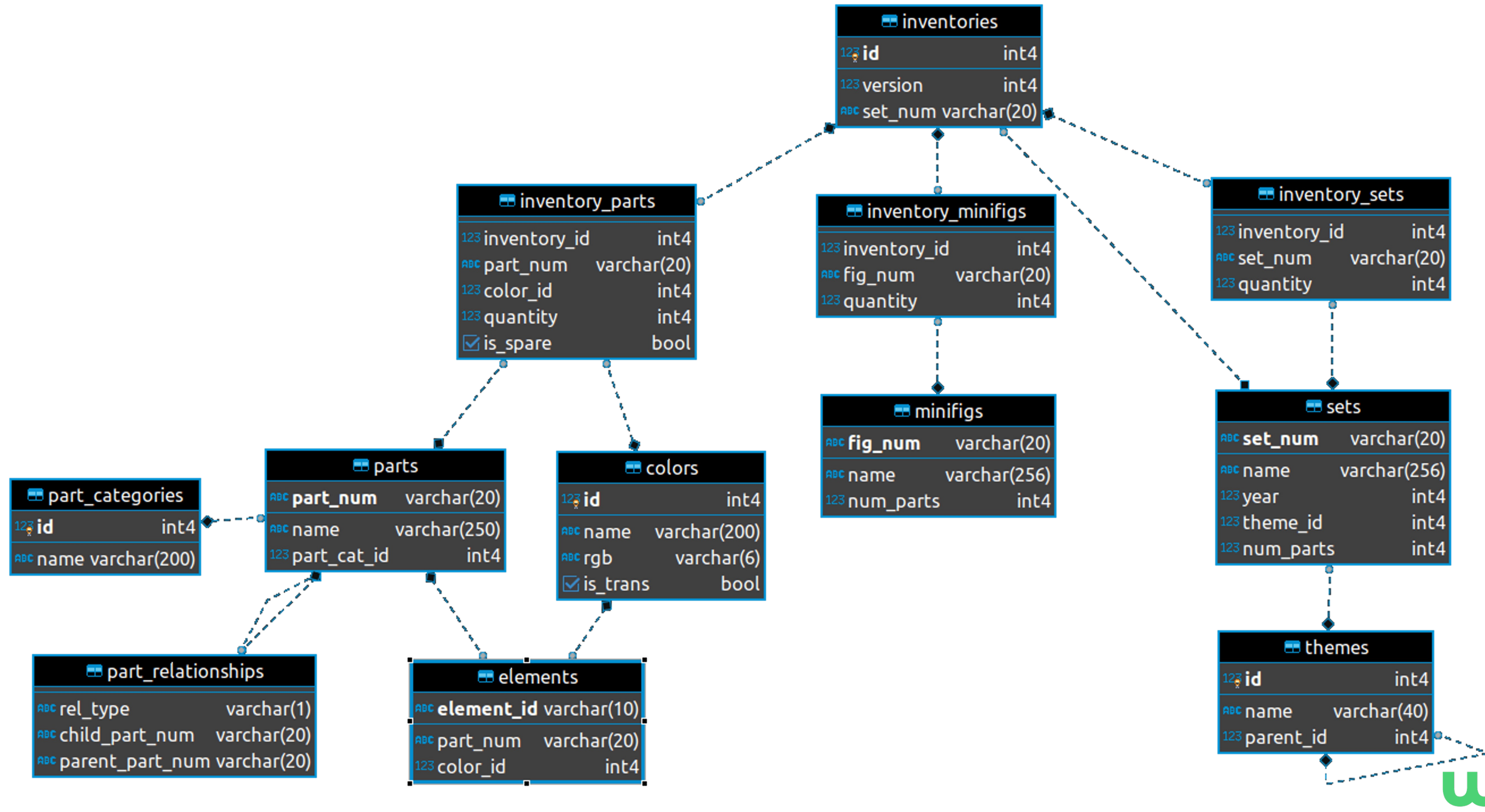
In Billund, Denmark, the world of LEGO started in 1916. At that time the blocks were made of wood, but since 1947 they are made of plastic. Over the years, the blocks have become even better and more durable (they last a very long time) and have already been named Toys of the Century twice. LEGO is the largest toy supplier in Europe and the second largest in the world. Many new themes, minifigs and technical features have been developed over the years. 20 billion stones are produced every year. The range consists of 2,200 different stones that are manufactured in 55 different colours. The blocks must meet high quality requirements so that all bricks produced since 1958 fit together.

# Do you have all the necessary Legos?

You are a big lego fan also called AFOL (Adult Fan of LEGO) and have now collected *all* Creator sets together. Now that you have it complete, you would like to build something else. You will end up on the 2007 Hobby Train (set 10183-1). However, it is no longer available. You want to know what you already own from this train and what you should buy separately, or whether there might be a set that contains the missing pieces.



# The (Rebrickable) data model



Voorbeeld



Set: 60218-1,  
Name: Desert Rally  
Racer,  
75 parts  
Theme: City

wortell

# The (Rebrickable) data model

- The database contains a catalog of LEGO Sets, Parts, Colors and various relationships. This LEGO catalog has been put together over many years with the help of an amazing volunteer admin team, with submissions from thousands of Rebrickable members.
- This database stores only official LEGO items - Sets, Parts and Minifigs (no B-Models, Sub-Sets, MOCs)
- Sets and Minifigs contain one or more Inventories (inventories.csv)
- Inventories can contain Sets (inventory\_sets.csv) and/or Parts (inventory\_parts) and/or Minifigs (inventory\_minifigs). The quantities are of members of the Rebrickable-club.
- Part Relationship rel\_types are: (P)rint, Pai(R), Su(B)-Part, (M)old, Pa(T)tern, (A)lternate
- A LEGO Set can consist of one or more Inventories (e.g. 8088-1) We might have multiple Inventories if the Parts in the Set were changed during it's production period, or perhaps to cater for different geographic regions.
- Each Inventory can contain Parts and/or Sets. Most Inventories contain just Parts, but some consist solely of Sets (e.g. 5004559-1) or a combination of Parts and Sets (e.g. 9509-1).



# Step-by-step plan; question 1

- Provide a data frame with the data of sets.
- Expand it with the following information:
  - The name of the theme (“Themes”) Note: There is a hierarchy in the themes, so make sure you know the highest level per set.
  - Create a new variable that indicates with 0/1 whether the set belongs to the Creator theme
- Question: What % of all sets do you own?

# Step-by-step plan; question 2

2. Filter the dataset (table/dataframe) based on your created variable 'Creator'.
3. Link the data sources together so that you know which parts it contains for all sets together. You need the following tables for this: Inventories, Inventory\_parts, Parts, Colors. This table is unique on part\_num and Color\_id.

**Question: How many stones do you have in the color Red?**

# Step-by-step plan; question 3

- You are also curious how many parts you now have per category.
- Link your just created table to the table 'Parts\_Categories' and count the number of parts and sort this table descending.
- **Questions:**
  - Which category do you have the most parts from?
  - What is the average number of tiles per creator set?



# Step-by-step plan; question 4

- Determine which parts are included in the Hobby Train set (set 10183-1).
- Match this table with the table of all your Creator Collection items.  
Create a new variable that counts how much you miss of each part.
- **Questions:**
  - How many % of all required stones are you still missing?
  - What is the top 5 of the parts that you already have and the bottom 5 that you are still missing?

# Bonus assignments

- You bought every set from the “Creator” theme in the year it came out. Make an overview per year of which parts you received.
- Make a list of unique parts that only appear in one set.
- Which recent set(s) (2017-now) contain the most parts you still need?
- Is this set still available? Answer this last question by scraping the LEGO website (tip: BeautifulSoup)