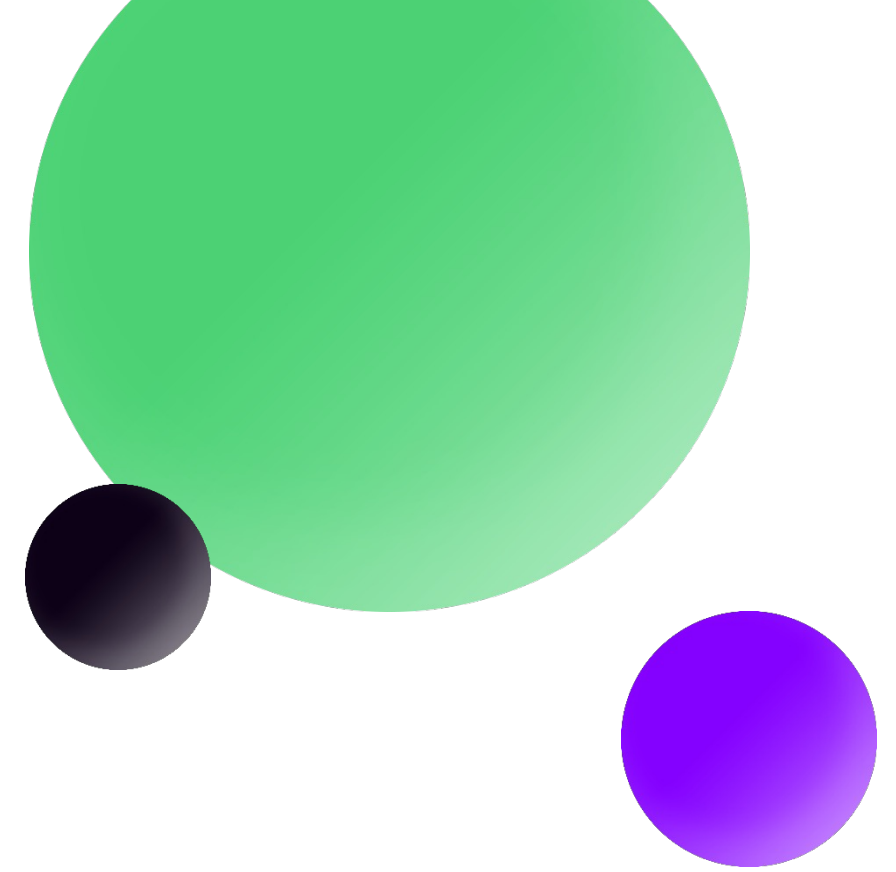


# Casus Python en Pandas



**wortell**

# Introductie - LEGO

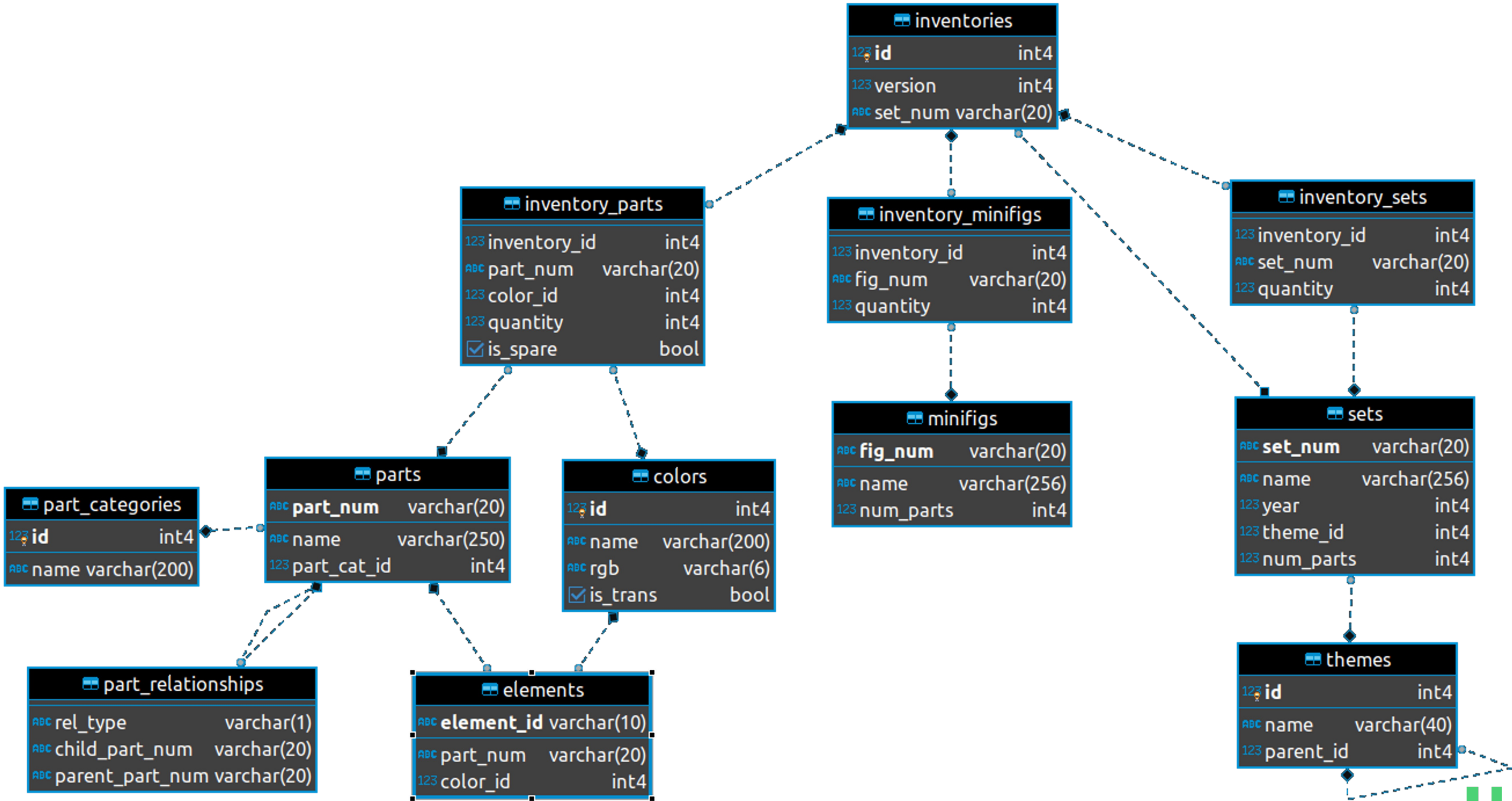
In Billund, Denemarken, is de wereld van LEGO begonnen in 1916. Destijds bestonden de blokjes uit hout, maar sinds 1947 zijn ze van plastic. Door de jaren heen zijn de blokjes nog beter en duurzamer geworden (ze gaan ontzettend lang mee) en zijn ze al twee keer uitgeroepen tot Speelgoed van de Eeuw. LEGO is de grootste speelgoedleverancier in Europa en de tweede grootste ter wereld. Door de jaren heen zijn er veel nieuwe thema's, minifigs en technische features ontwikkeld. Jaarlijks worden er 20 miljard steentjes geproduceerd. Het assortiment bestaat uit 2.200 verschillende stenen die vervaardigd worden in 55 verschillende kleuren. De blokjes moeten aan hoge kwaliteitseisen voldoen zodat alle stenen die sinds 1958 geproduceerd zijn op elkaar passen.

# Heb je alle benodigde blokjes?

Je bent een grote lego-fan ook wel AFOL (Adult Fan of LEGO) genoemd en hebt inmiddels *alle* Creator-sets bij elkaar gespaard. Nu je die compleet hebt wil je graag eens wat anders bouwen. Je komt uit op de Hobby Train uit 2007 (set 10183-1). Echter is deze niet meer verkrijgbaar. Je wilt weten wat je van deze trein al in je bezit hebt en wat je nog los aan zou moeten schaffen, of dat er wellicht een set is die de ontbrekende stukjes bevat.



# Het (Rebrickable) datamodel



Voorbeeld



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



# Het (Rebrickable) datamodel

- 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).

# Stappenplan en vraag 1

- Zorg voor een dataframe met de gegevens van sets.
- Breid deze uit met de volgende informatie:
  - De naam van het thema (“Themes”)
    - Let op: Er zit een hiërarchie in de thema’s, zorg dus dat je het hoogste niveau uiteindelijk weet per set.
  - Maak een nieuwe variabele aan die aangeeft met 0/1 of de set bij het Creator-thema hoort

**Vraag: Hoeveel % van alle sets heb je in je bezit?**

# Stappenplan en vraag 2

2. Filter de dataset (tabel/dataframe) op basis van je aangemaakte variabele 'Creator'.
3. Koppel de databronnen aan elkaar zodat je voor alle sets samen weet welke onderdelen het bevat. Je hebt hiervoor de volgende tabellen nodig: Inventories, Inventory\_parts, Parts, Colors. Deze tabel is uniek op part\_num en Color\_id.

**Vraag:** Hoeveel steentjes heb je in de kleur Red?

# Stappenplan en vraag 3

- Je bent ook benieuwd hoeveel onderdelen je nu per categorie hebt. Koppel je net gemaakte tabel aan de tabel 'Parts\_Categories' en tel het aantal onderdelen en sorteer deze tabel van veel naar weinig.
- **Vragen:**
  - Van welke categorie heb je de meeste onderdelen?
  - Wat is het gemiddelde aantal steentjes per creator-set?



# Stappenplan en vraag 4

- Bepaal welke onderdelen er in de set van de Hobby Train zitten (set 10183-1).
- Match deze tabel met de tabel van al je onderdelen van je Creator-verzameling. En maak een nieuwe variabele aan die telt hoeveel je mist van elk onderdeel.
- **Vragen:**
  - Hoeveel % van alle benodigde steentjes mis je nog?
  - Wat is de top-5 van de onderdelen die je allemaal al hebt en de bottom-5 welke je nog mist?

# Bonusopdrachten

- Je hebt elke set uit het “Creator” thema gekocht in het jaar dat deze uitkwam. Maak een overzicht per jaar welke onderdelen je erbij hebt gekregen.
- Maak een lijst van unieke onderdelen die maar in één set voorkomen.
- Welke recente set(s) (2017-nu) bevat de meeste onderdelen die je nog nodig hebt?
- Is deze set nog leverbaar? Beantwoord deze laatste vraag door het scrapen van de LEGO-website (tip: BeautifulSoup)