

A LEGO Data Analysis: The Playing Generation

A Data Dictionary by **Noortje Offreins**

INDEX

DATA SET 1 // LEGO SALES	3
DATA SET LEGO SALES: VARIABLES BEFORE CLEANING.....	4
DATA SET 2 // LEGO CATALOG	6
MAPPING THE RELATIONAL DATABASE (or: Entity Relationship Diagram).....	7
DATA SET LEGO CATALOG: VARIABLES BEFORE CLEANING.....	8
FACT TABLE.....	8
DIMENSION TABLES.....	8
LEGEND	12
VARIABLES DICTIONARY.....	12
TYPE OF COATING OVERVIEW.....	13
PINKS.....	14
COLOR NAMING DIVERSION OVERVIEW.....	14

DATA SET 1 // LEGO SALES

Citation: Boiński, T. M., Oczkoś, W., Podgórski, B., & Szczepańska, W.
(2023). Data on LEGO sets release dates and retail prices combined with
aftermarket transaction prices between June 2018 and June 2023. (1-)
[Dataset]. Gdańsk University of Technology. <https://doi.org/10.34808/s25h-sx91>

DATA SET LEGO SALES: VARIABLES BEFORE CLEANING

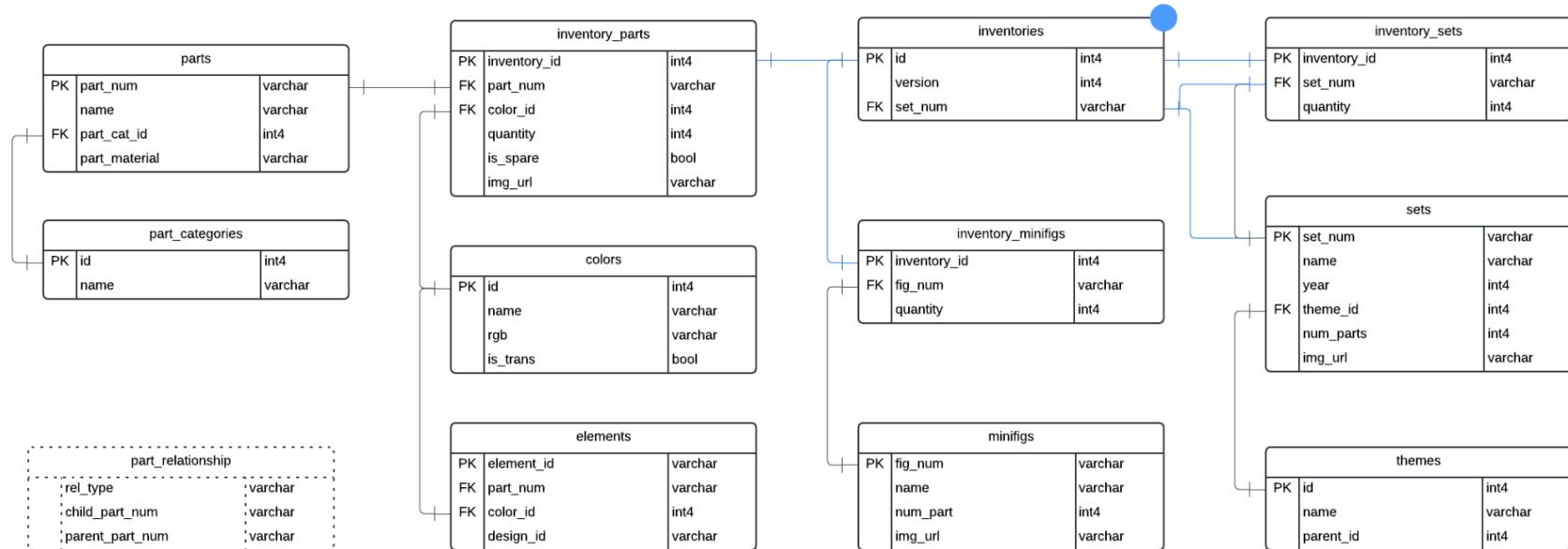
Name	Data type	Description
setID	int64	internal Brickset.com LEGO set identification number
number	int64	official LEGO set ID
numberVariant	int64	official LEGO set sub variant (e.g. different minifigure hidden in a random bag)
name	object	official LEGO set name
year	int64	the set release year
theme	object	official name of the set theme
themeGroup	object	official name of the set themes group (if available)
subtheme	object	official name of the set sub-theme (if available)
category	object	brickset.com internal set type (overview here)
released	int64	indicates whether the set was officially released (1) or not (0)
pieces	float64	number of parts in the set
minifigs	float64	number of minifigures in the set
ownedBy	float64	number of brickset.com users claiming that he or she owns the set
wantedBy	int64	number of brickset.com users claiming that he or she wants to buy the set
rating	float64	average set rating according to brickset.com users
reviewCount	int64	number of the set reviews written by brickset.com users
packagingType	object	type of packaging for the set (if specified)
availability	object	indicates whether the set was available in retail shops or only on official LEGO shop website
instructionsCount	int64	number of books with building instructions added to the set
minAge	float64	LEGO recommended minimal user age for the set
maxAge	float64	LEGO recommended maximal user age for the set (either not specified or 99)
tags	object	list of brickset.com assigned set tags
lastUpdated	object	the date and time of the last update of the data in brickset.com in ISO 8601 format
urlRetailPriceCheckPLN	object	URL where retail price in PLN was downloaded from
US_retailPrice	float64	retail price in United States in US dollars
US_dateFirstAvailable	object	date and time when the set became available in United States in ISO 8601 format
US_dateLastAvailable	object	date and time when the set stopped being officially available

		in United States in ISO 8601 format
UK_retailPrice	float64	retail price in United Kingdom in GBP
UK_dateFirstAvailable	object	date and time when the set became available in United Kingdom in ISO 8601 format
UK_dateLastAvailable	object	date and time when the set became available in United Kingdom in ISO 8601 format
CA_retailPrice	float64	retail price in Canada in Canadian dollars
CA_dateFirstAvailable	object	date and time when the set became available in Canada in ISO 8601 format
CA_dateLastAvailable	object	date and time when the set became available in Canada in ISO 8601 format
DE_retailPrice	float64	retail price in Germany in EUR
DE_dateFirstAvailable	object	date and time when the set became available in Germany in ISO 8601 format
DE_dateLastAvailable	object	date and time when the set became available in Germany in ISO 8601 format
PL_retailPrice	float64	retail price in Poland in PLN
Date	object	year and month for which the PriceMonthPLN is given
priceMonthPLN	float64	price in PLN read from promoklocki.pl for year and month specified in Date column
status	object	official status of the set (if available) in LEGO web shop
urlRetailPriceHistoryPLN	object	URL containing retail and aftermarket price changes from the day of the release of the set, in PLN

DATA SET 2 // LEGO CATALOG

Citation: [Rebrickable.com](https://rebrickable.com)

MAPPING THE RELATIONAL DATABASE (or: Entity Relationship Diagram)





DATA SET LEGO CATALOG: VARIABLES BEFORE CLEANING

FACT TABLE

Table: inventories

Stores information about the contents of a set.

Key	Name	Data Type	Description / Attributes
	id	int4	<ul style="list-style-type: none">- A unique identification number that refers to the contents of a set. A new inventory is created when LEGO changes the contents of a set over time.- Primary Key.
	version	int4	<ul style="list-style-type: none">- Unique identifier for the version of inventory.
	set_num	varchar (20)	<ul style="list-style-type: none">- A unique identification number for a set based on official LEGO set numbers but complemented by Rebrickable identifiers.- Foreign Key.

DIMENSION TABLES

Table: colors

Stores information about the color of the bricks.




Key	Name	Data Type	Description / Attributes
	id	int4	<ul style="list-style-type: none">- Unique identification number for the color assigned by Rebrickable.- Primary Key.
	name	varchar(200)	<ul style="list-style-type: none">- Name of the color assigned by Rebrickable¹.
	rgb	varchar(6)	<ul style="list-style-type: none">- 6 digital standard format for displaying colors on screens.
	is_trans	bool	<ul style="list-style-type: none">- Color is transparent: true / false.

Table: elements

Stores information about the LEGO parts.

Key	Name	Data Type	Description / Attributes
	element_id	varchar(10)	<ul style="list-style-type: none">- A 6/7 digit number assigned by Lego that codes for a combination of shape and color.- Primary Key.
	part_num	varchar(20)	<ul style="list-style-type: none">- A unique identification number for the shape or design of a piece assigned by Rebrickable.
	color_id	int4	<ul style="list-style-type: none">- Unique identification number for the color assigned by

¹The color name follows the Rebrickable assigned LEGO name. They provide [an overview](#) matching their names with official LEGO and other fan community names. See [the legend below](#) for more details.

			Rebrickable. - Foreign Key.
	design_id	varchar	- A 5/6 digit number assigned by LEGO that codes for the shape of the part.

Table: inventory_minifigs

Stores information about the minifigs in a specific inventory



Key	Name	Data Type	Description / Attributes
	inventory_id	int4	- A unique identification number that refers to the contents of a set. - Primary Key.
	fig_num	varchar (20)	- Unique identification number sequencing the minifig assigned by Rebrickable. - Foreign Key.
	quantity	int4	- Number of minifigs in the inventory

Table: inventory_parts

Stores information about the parts in an inventory






Key	Name	Data Type	Description / Attributes
	inventory_id	int4	- A unique identification number that refers to the contents of a set. - Primary Key.
	part_num	varchar(20)	- A unique identification number for the shape or design of a piece assigned by Rebrickable. - Foreign Key.
	color_id	int4	- Unique identification number for the color assigned by Rebrickable. - Foreign Key.
	quantity	int4	- Total number of parts in the inventory.
	is_spare	bool	- Part is a spare part: true / false.
	image_ulr	varchar	- Links to source image.

Table: inventory_sets

Stores information about which set corresponds to the inventory.

Key	Name	Data Type	Description / Attributes
	inventory_id	int4	- A unique identification number that refers to the contents of a set. A new inventory is created when LEGO changes the contents of a set over time.
	set_num	varchar	- A unique identification number for a set based on

			official LEGO set numbers but complemented by Rebrickable identifiers. - Foreign Key.
	quantity	int4	- Total number of parts in the set

Table: minifigs

Stores information about unique minifigs.


Key	Name	Data Type	Description / Attributes
	fig_num	varchar(20)	- Unique identification number sequencing the minifig assigned by Rebrickable. - Primary Key.
	name	varchar(256)	- Name of Minifigs as assigned by Rebrickable.
	num_part	int4	- Numeric value indicating the size of the inventory.
	image_ulr	varchar	- Links to source image.

Table: parts

Stores information about unique parts.



Key	Name	Data Type	Description / Attributes
	part_num	varchar(20)	- A unique identification number for the shape or design of a piece assigned by Rebrickable. - Primary Key.
	name	varchar(20)	- Name of the part as assigned by Rebrickable.
	part_cat_id	int4	- Unique identification number for the category of parts. - Foreign Key.
	part_material	varchar	- Name of the material used for the part.

Table: part_categories

Stores information about the categories the unique parts are assigned to.


Key	Name	Data Type	Description / Attributes
	Id	int4	- Unique identification number for the category of parts. - Primary Key.
	name	varchar(200)	- Name of the category of parts.

Table: part_relationships

Stores information about how single pieces such as hands, arms, stickers relate to parts.

Key	Name	Data Type	Description / Attributes
	rel_type	varchar	- A categorical identifier how a child part relates to the

			parent part. ²
	child_part_num	varchar	- Unique identifier for single part.
	parent_part_num	varchar	- Unique identifier for part that is made up of single parts.

Table: sets

Stores information about the sets.




Key	Name	Data Type	Description / Attributes
	set_num	varchar(20)	- A unique identification number for a set based on official LEGO set numbers but complemented by Rebrickable identifiers. - Primary Key.
	name	varchar(256)	- Name of the LEGO set.
	year	int4	- The release year in numbers.
	theme_id	int4	- A unique identification number for the set theme assigned by Rebrickable. - Foreign Key.
	num_parts	int4	- The number of parts in the set.
	image_ulr	varchar	- Links to source image.

Table: themes

Stores information about the themes the sets are assigned to.

Key	Name	Data Type	Description / Attributes
	id	int4	- A unique identification number for the set theme assigned by Rebrickable. - Primary Key.
	name	varchar(40)	- Name of the set theme assigned by Rebrickable.
	parent_id	int4	- Refers to the id (themes) if the theme is also appearing as a subtheme of another theme, i.e. City & Airport.

² Part Relationship rel_types are: (P)rint, Pai(R), Su(B)-Part, (M)old, Pa(T)tern, (A)lternate

LEGEND

VARIABLES DICTIONARY

Part	A shape or design of a piece, i.e. wheel or brick.
Element	A combination of part and color, i.e. black wheel, 2x4 red brick.
UPN	Unknown Part Number
Set numbers	<p>In the past, LEGO used the same set number for different sets, forcing fans to invent a suffix. Rebrickable complemented official LEGO set numbers, including references to set type:</p> <ul style="list-style-type: none">- For Normal and Supplemental sets we use the LEGO set number followed by "-1"- B-models have the full set number followed by "-b1", "-b2", etc.;- Subsets we use "-s1", for Subsets of B-Models "-b1-s1".- Super-Sets sometimes have an official LEGO set number;- If not, we use the set number of one of the included sets, preceded with a "K" (for kit) or a "C" (for collection).- Models seen in books, leaflets, magazines, or instruction manuals that do not step-by-step instructions are called LEGO Idea reBuilds (LIBs) . We use the prefix "-c1".
Theme	User themes are useful because they are a way to easily narrow down areas of interest. Rebrickable follows the lead of LEGO in terms of assigning the themes, however historically the theme setting was a fan initiative and therefore discrepancies could exist.
Inventory	Is the overview of unique parts in a set. The size of the inventory can differ from the quantity, when unique parts are used more often in a set. For example, a set can have an inventory of 123 parts with a quantity of 571 parts.

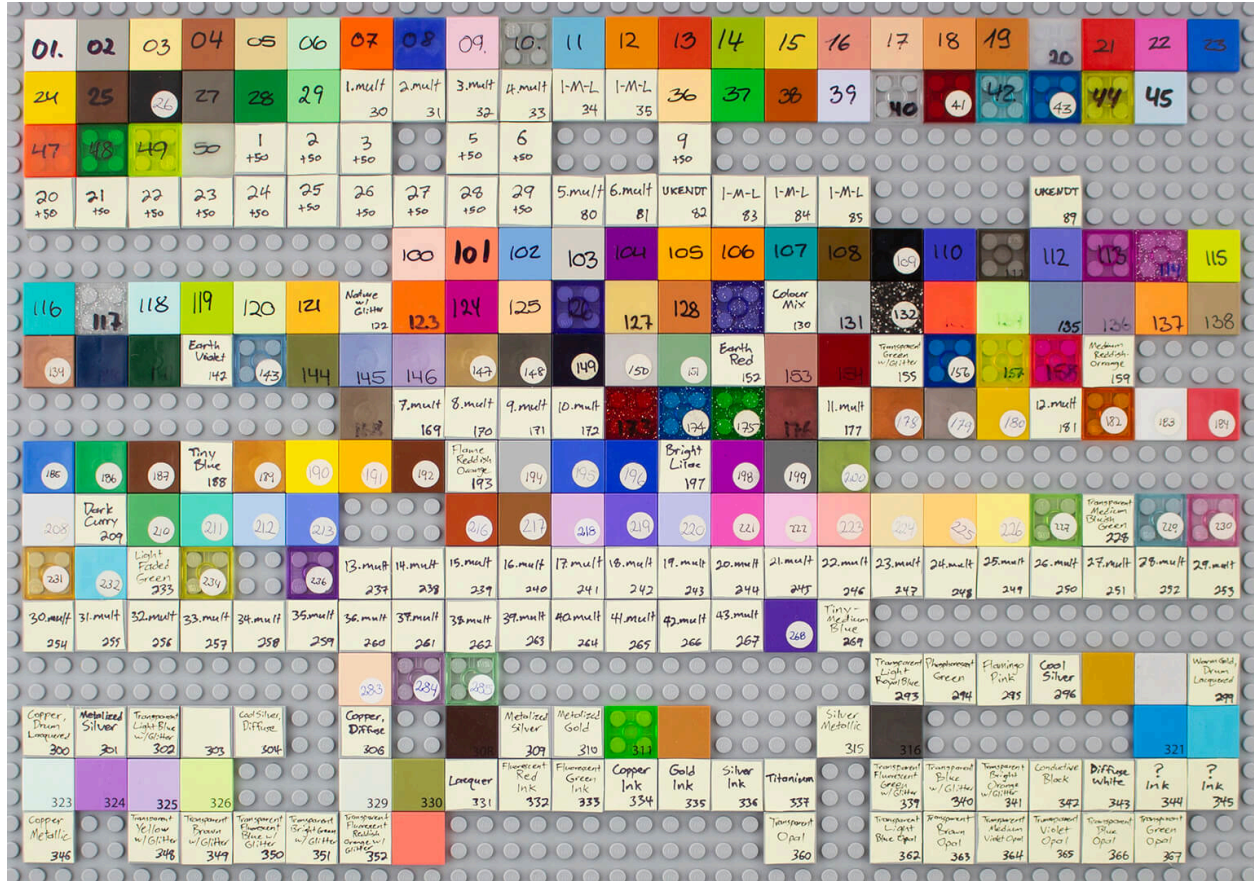


Image: [Rylie Howerter](#)

TYPE OF COATING OVERVIEW

Type	Description
Solid	The most common type. Pretty self explanatory.
Transparent	Another common type. All the clear colors that light happily shines through.
Metallic	A dull reflective coating on top of a molded color. Popularized after LEGO made the decision to stop making Chrome parts.
Chrome	A highly shiny reflective coating.
Pearl	Confusingly referred to as Metallic by LEGO, these are molded colors not an applied effect. They can look kind of swirled.
Glitter	Metallic particles are mixed with a color (usually transparent for obvious reasons) before molding to add the glittery effect.
Satin	The newest type added in 2020, we call them Opal (Opalescent) because they have that kind of iridescent look like an opal.

Speckled	A flecked/spotted coating of metallic. These colors get very confusing in terms of how LEGO defined them. Luckily they didn't last long so the confusion is limited.
Milky	A very limited use type, almost exclusive to Glow in the Dark parts. There's a color called Milky White which is actually just the un-dyed pellets LEGO melt to make parts. LEGO calls it 'Nature' but it hasn't been seen since 2008.