CONTEXT

The world of merchandise related to Anime characters and stories has a value of about 30 B$ (2023) with a CAGR of 8 - 9.5 % (2030).
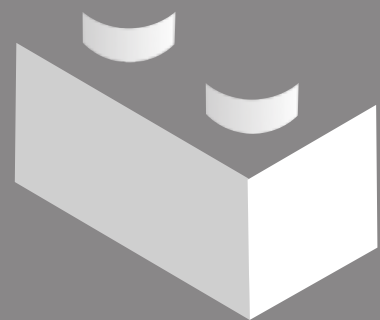
Over the last 20 years, LEGO has increased its revenue almost ten times, from B$0.98 (2003) to B$9.51 (2023).
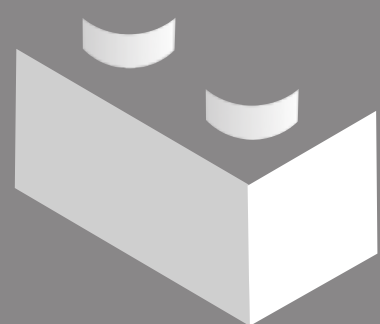
QUESTION

Are there opportunities for LEGO designers and creators to propose Anime-based sets in co-branding with Anime production companies and enter the merchandise market?

THE LEGO CREATIVE PIPELINE 🧱

LEGO DESIGNERS

- Have an idea
- Make a concept
- Submit to a design manager
- Submit to a reviewers panel

LEGO IDEAS

- Have a great idea (that follows the rules)
- Build it
- Submit to the LEGO IDEAS website
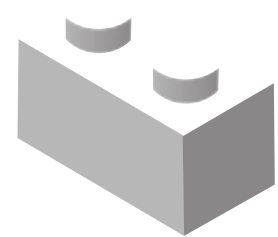- Spread to the world to gain support
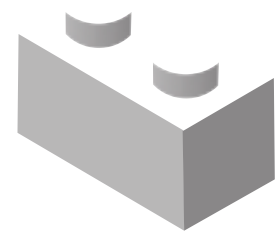- After 10K supporter enter in the Expert review status

POSITIVE REVIEW

LEGO includes the set in product line

TOOLS, TECHNOLOGIES AND ARCHITECTURE

DATA COLLECTION — python ⬇

ETL — python | pandas | Chroma
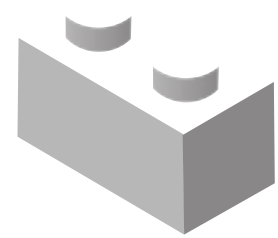
STORAGE AND PROVISIONING — CSV | MongoDB Atlas

DATA VISUALIZATION — tableau

ARCHITECTURE — On Premis | docker | Cloud
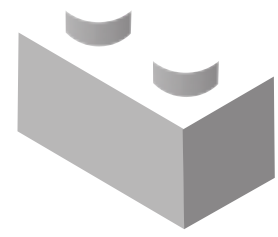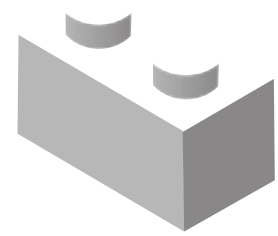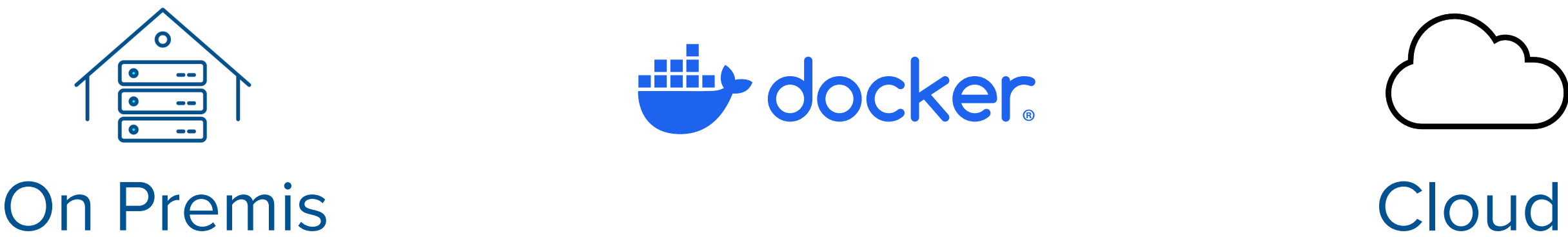
ORIGINAL DATASETS

	ANIME	LEGO
SOURCES	MyAnimeList myanimelist.net	BRICKSET YOUR LEGO® SET GUIDE brickset.com
COLLECTION METHOD	Scraping/API	Download
FORMAT	JSON	CSV
PROPERTIES PER RECORD	31	42
TOTAL RECORDS	14693	20140

DATA EXPLORATION AND CLEANING

ANIME

The Anime dataset shows some issues with missing data and overabundance of categories that required specific management.

LEGO

The LEGO dataset shows problems with missing data, incorrect field formatting and the presence of inconsistent records compared to the ongoing analyses.

The cleaning operations were applied differently to both datasets:

- first, keep only sets that are released and remove duplicates

- then delete the columns that do not serve our purposes

- set the correct format for the columns that required it

- managed the missing data

At the end of the process we kept 11714 records and 21 properties (columns)

At the end of the process we kept 12812 records and 20 properties (columns)

STAGE DATA - ANIME 🧱

Among all the properties that have been retained and treated from the original dataset, the most useful ones for the purposes of the current study are listed here, with the introduction of the calculated field final_score that takes into account the number of users and their behavior.

The number of genres present in the initial dataset has been reduced from more then 70 to 13 by merging the common ones.

title: name of anime (with season number)/movie

num_scoring_user: number of users that vote the anime

completed: number of users who completed the anime

watching: number of users who is currently watching the anime

plan_to_watch: number of users who wants to watch the anime

on_hold: number of users who paused watching the anime/movie

dropped: numbeR of users who stopped watching the anime/movie

studio: name of the production studio(s).

media_type: method of publishing the anime/movie

inspired_by: source of inspiration for the anime/movie

$$final\ score = 10 * \frac{log(score)}{max(log(score))} \in (0,10]$$

$$
\begin{aligned}
score = (vote &* scoring\ user) \\
&+ (6 * watching) + (8 * completed) + (5.5 * on\ hold) \\
&+ (1 * dropped) + (4 * plan\ to\ watch)
\end{aligned}
$$

final_score: calculated field based on MyAnimeList users' vote

anime_quality: value based on final_score

rating: the minimum recommended age rating for viewing

genres: the list of genres (merged)

STAGE DATA - LEGO

Among all the properties that have been retained and treated from the original dataset, the most useful ones for the purposes of the current study are listed here.

To address the lack of data on LEGO set prices, a new MixedPrice field was introduced that uses *USRetailPrice*, if present, or the *BrickLinkSoldPriceNew,* the price in the exchange market. At the end, this field was not taken into account for final analyses because it created too many outliners both in the set distribution per price and in the calculation of the price per piece.

SetID: the commercial ID of the LRGo set

SetName: name of the LEGO set

YearFrom: year of LEGO set release

ThemeGroup: macro-group of LEGO sets

USRetailPrice: price of the LEGO set in $

Pieces: number of pieces

Minifigs: number of minifigures

PackagingType (group): packaging of the LEGO set

MinAge: the minimum recommended age

MixedPrice: the USRetailPrice, if present, the price in the exchange market of the website.

DATA INTEGRATION

In order to integrate the two datasets of Anime and LEGO, two particular dimensions were chosen:

the minimum recommended age

Anime are classified by movie rating, while LEGO sets have a minimum recommended age. A connection minimum age field was then built, aligned as follows:

Movie rating	Minimum Age
G - General audience	0
PG - Parental Guidance	10
PG-13 - Parents Cautioned	13
R - Restricted	18

the genre

To classify the LEGO sets by the chosen subset Anime genres, the Chroma vector database was used to calculate the distance between vectors made by the chosen genres (as reference) and the ones composed by title, theme group, theme and subtheme of the LEGO sets.

Genres : ['Sci-Fi' , 'Arts' , 'Drama' , 'Suspence' , 'Action' , 'Detective' , 'Horror' , 'Adventure' , 'Comedy' , 'Romantic' , 'Fantasy' , 'Music' , 'Kids']

THE FACT TABLES

Once the ETL process was completed, the fact tables were generated in CSV format, with a relational structure that can be conveniently imported into any SQL DBMS.

Age

Age/LEGO

Age/Anime

Studios

LEGO

Anime

Genres/Anime

Genres/LEGO

Word embeddings

Genres

Main table

Related table

Join table

Integration table
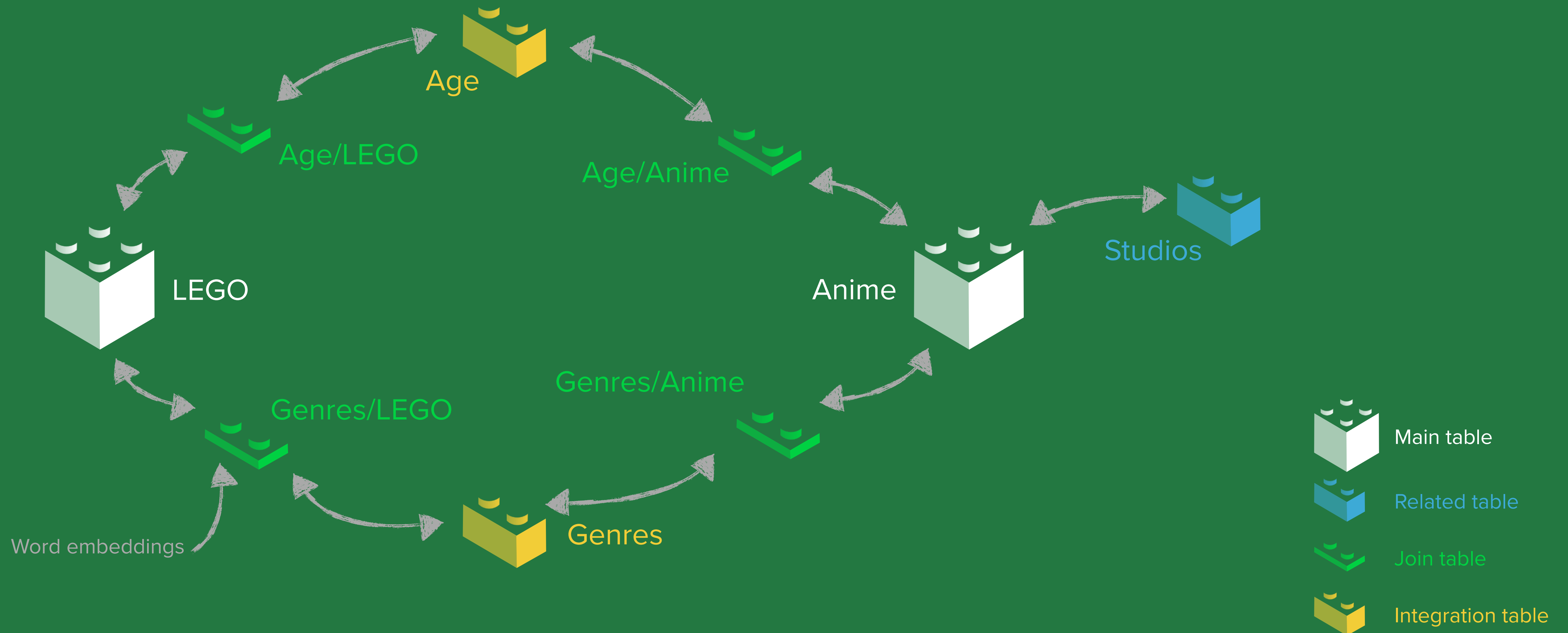
DATA VISUALIZATION

For data visualization, three different dashboards were created with Tableau, each with a specific purpose.

The ANIME Dashboard, to exploreto the ratings, quality and audience response to released anime

Here you can explore anime productions based on genre, year, media type and user response, looking not only at the rating but also at their behavior.

The LEGO Dashboard, to explore prices, piece count and types of lego sets produced
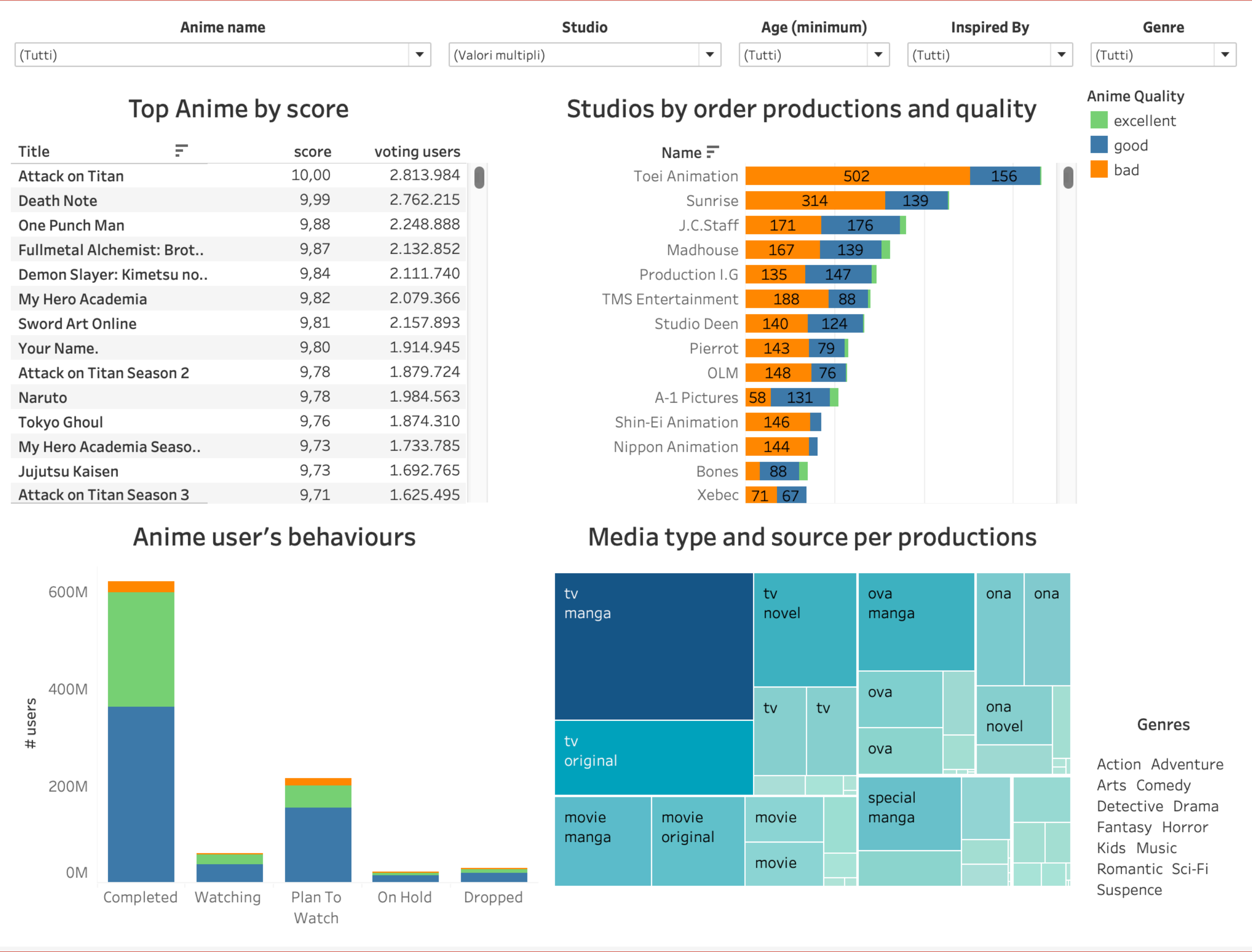
Here you can explore the marketing of Lego sets by year, genre, age, price and number of pieces

The SUMMARY OVERVIEW Dashboard, to support the choices of designers and creators

Here you can get an overview of the requirements that the prototype must have by comparing it with the sets already produced

DASHBOARD - ANIME

In the Anime Dashboard it can be possible to analyze the quality of the Anime productions, their score and the behavior of the users, applying some filters such as those related to genre, minimum age and source of inspiration.

Anime name	Studio	Age (minimum)	Inspired By	Genre
(Tutti)	(Valori multipli)	(Tutti)	(Tutti)	(Tutti)

Top Anime by score

Title	score	voting users
Attack on Titan	10,00	2.813.984
Death Note	9,99	2.762.215
One Punch Man	9,88	2.248.888
Fullmetal Alchemist: Brot..	9,87	2.132.852
Demon Slayer: Kimetsu no..	9,84	2.111.740
My Hero Academia	9,82	2.079.366
Sword Art Online	9,81	2.157.893
Your Name.	9,80	1.914.945
Attack on Titan Season 2	9,78	1.879.724
Naruto	9,78	1.984.563
Tokyo Ghoul	9,76	1.874.310
My Hero Academia Seaso..	9,73	1.733.785
Jujutsu Kaisen	9,73	1.692.765
Attack on Titan Season 3	9,71	1.625.495

Studios by order productions and quality

Anime Quality
- excellent
- good
- bad

Name		
Toei Animation	502	156
Sunrise	314	139
J.C.Staff	171	176
Madhouse	167	139
Production I.G	135	147
TMS Entertainment	188	88
Studio Deen	140	124
Pierrot	143	79
OLM	148	76
A-1 Pictures	58	131
Shin-Ei Animation	146	
Nippon Animation	144	
Bones	88	
Xebec	71	67

Anime user's behaviours

#users — 600M, 400M, 200M, 0M

Completed, Watching, Plan To Watch, On Hold, Dropped

Media type and source per productions

tv manga, tv novel, ova manga, ona, ona, tv original, tv, tv, ova, ova, ona novel, ova, movie manga, movie original, movie, special manga, movie

Genres
Action Adventure Arts Comedy Detective Drama Fantasy Horror Kids Music Romantic Sci-Fi Suspence
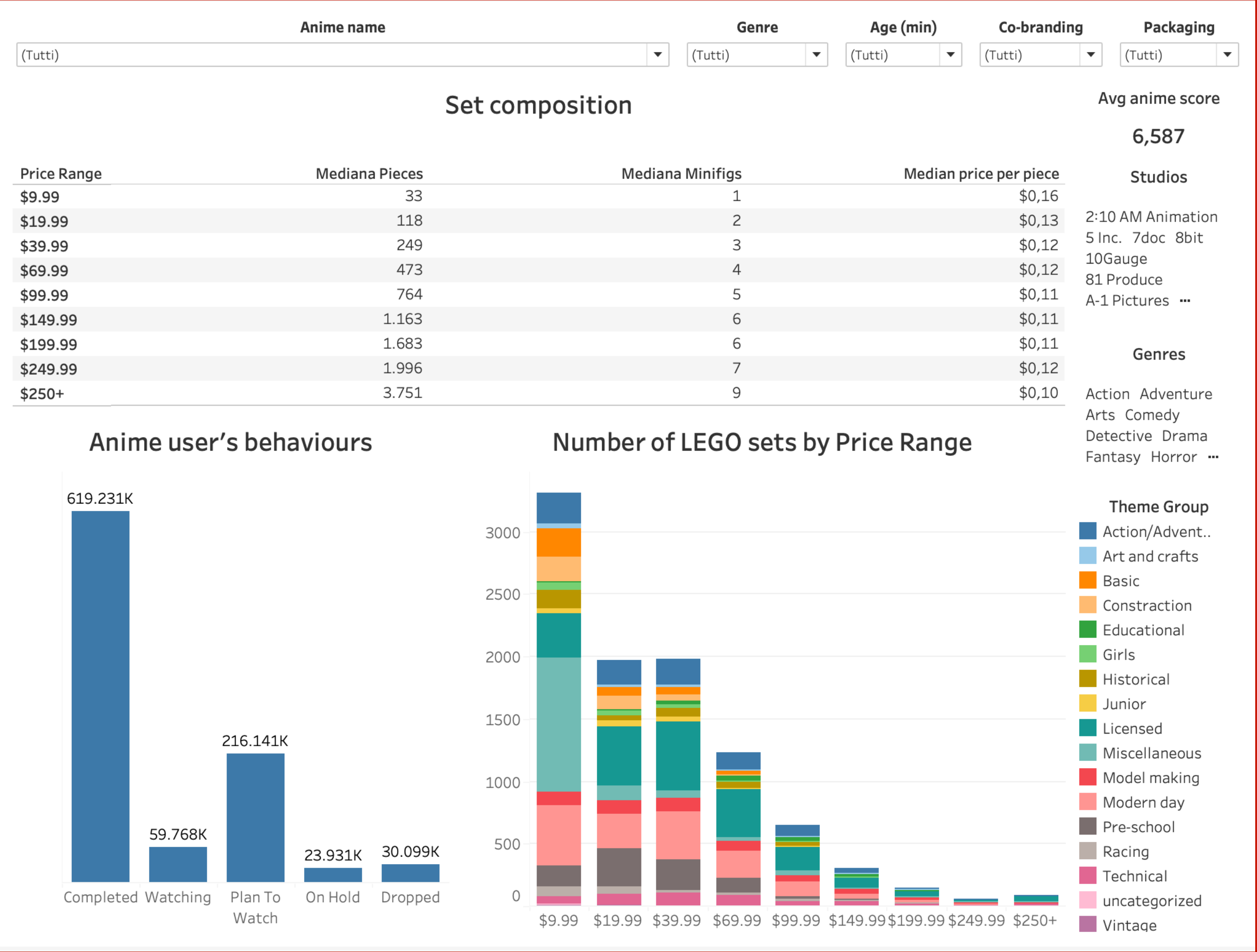
DASHBOARD - LEGO

In the LEGO Dashboard, it can be possible to analyze the history of the sets marketed, divided by Theme Group and displaying detailed values such as the distribution of price per piece, the number of sets per price range and the trend of the price per piece.

Filters can also be applied to all of this as for the Anime dashboard, by genre and minimum age, in addition to specific ones such as the type of packaging and the possibility to select only the sets produced under license.
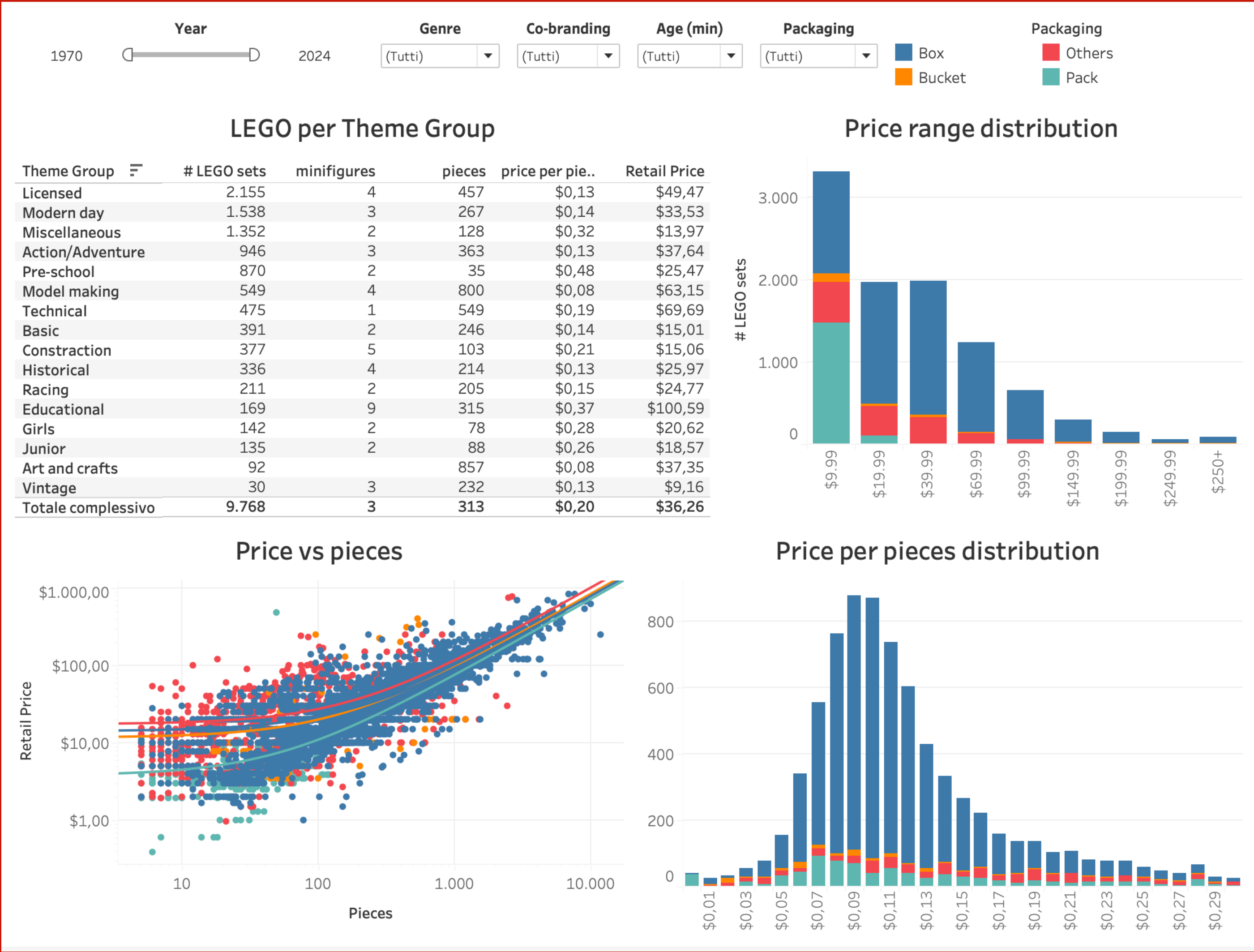
DASHBOARD - SUMMARY OVERVIEW

In the Summary Overview Dashboard, after selecting the Anime(s), it is possible to view the information useful for defining the prototype to submit to the approval chain. It will be possible to verify the requirements of the Lego sets based on the price ranges, the sets actually marketed by theme group, and the possible target audience. Furthermore, the information associated with the selection on genres (also filterable), production houses and the average score are reported
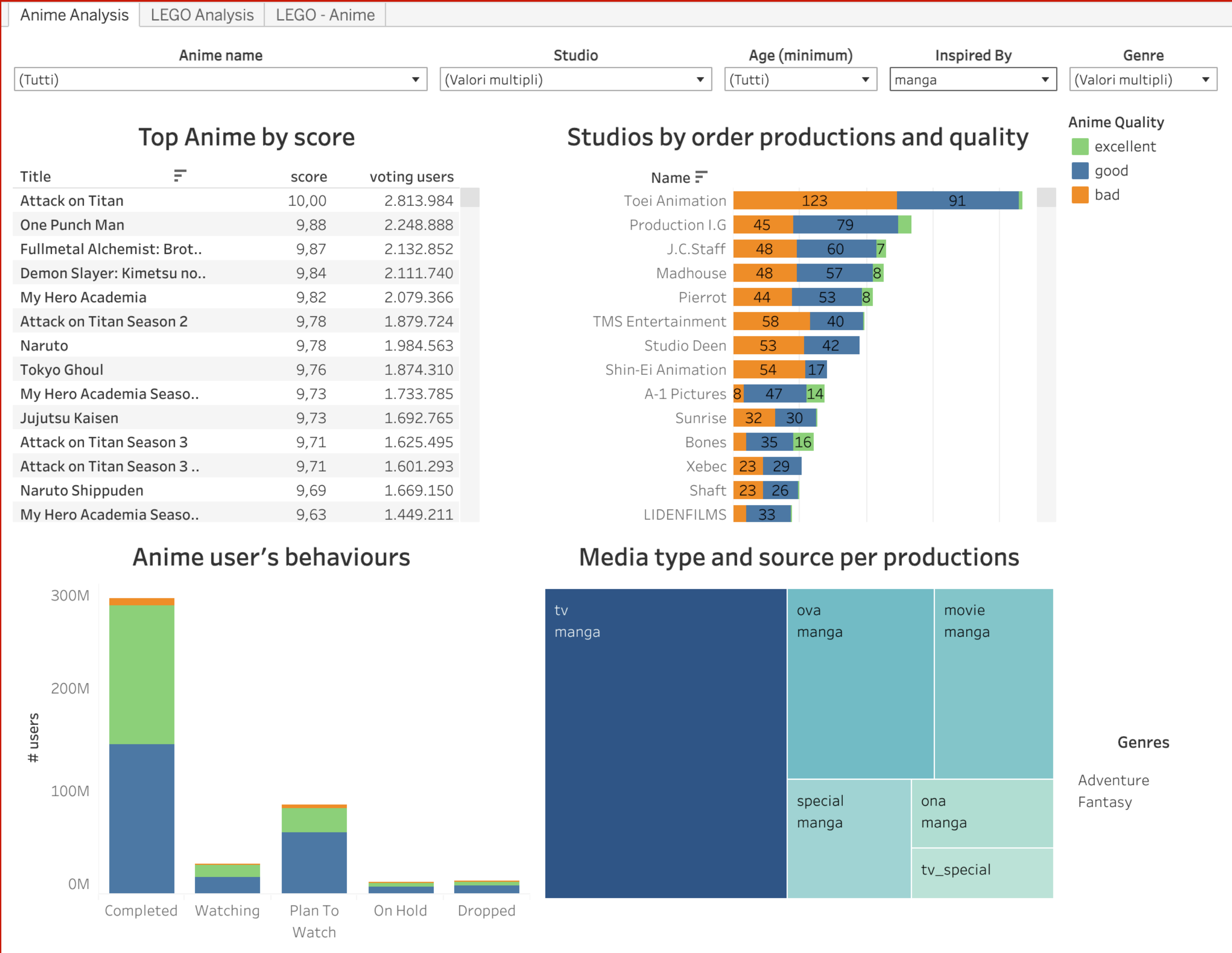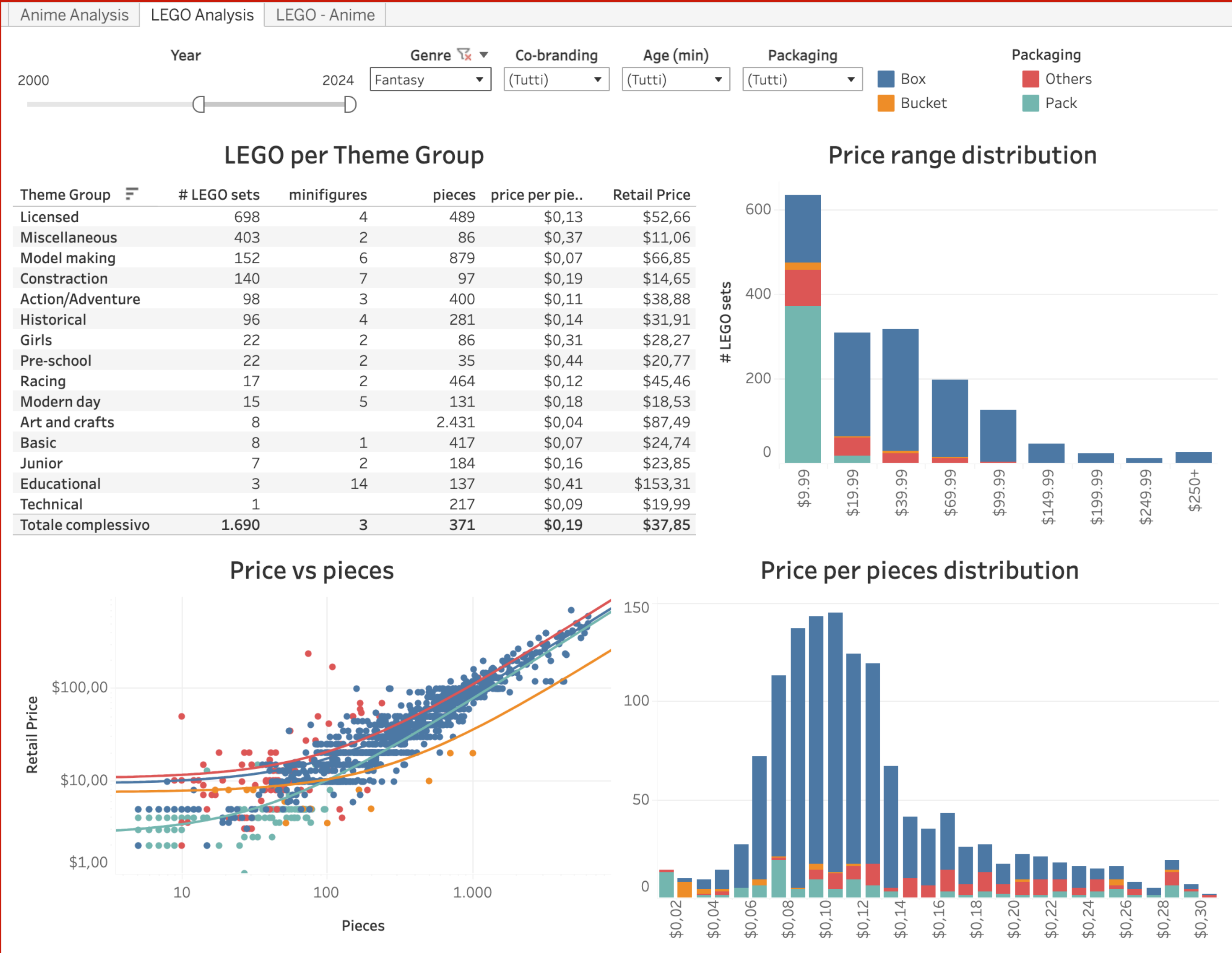
DASHBOARD - USE CASE

Starting from the study of Anime, selecting some genres [Adventure and Fantasy] and the source of inspiration [manga], you can visualize the list of different productions ordered by score, the associated studios and the behavior of the public with the Anime quality and the types of product.

NOTE: the different seasons of the same Anime have been voluntarily kept separate because the choice of representing a character, a specific scene or a setting through a LEGO set can depend on the data for the particular season (scores, user's behaviors and so on).

DASHBOARD - USE CASE

Selecting one of the previous genres [Fantasy] in the LEGO dashboard and choosing the sets produced after the year 2000, you can visualize the overview of the set created: the list of Theme Group with related average data, the distribution of sets per price range, divided by packaging anche the trend and distribution of price per piece.
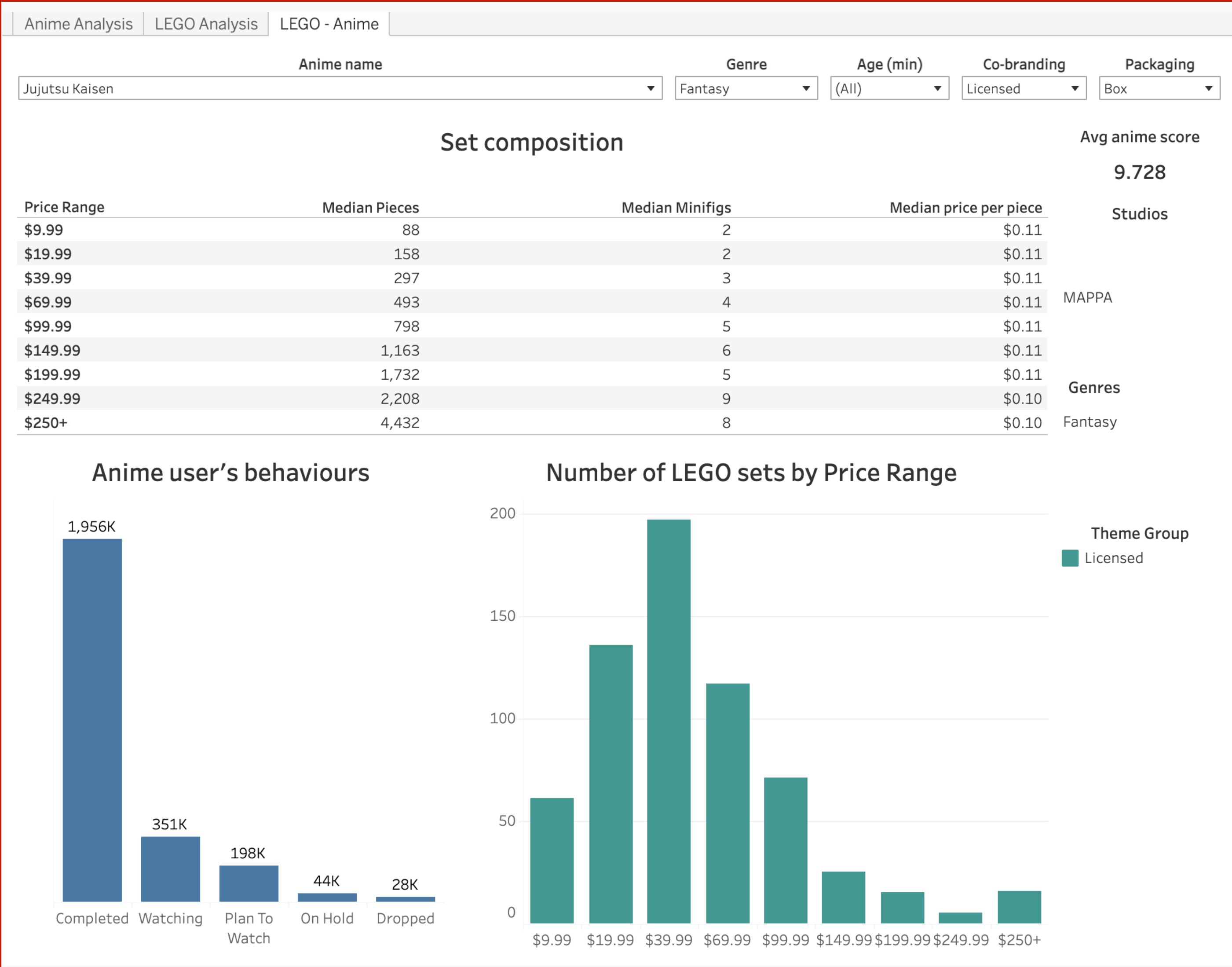
DASHBOARD - USE CASE

Finally, by selecting one of the Anime identified in the first view [Jujutsu Kaisen], the construction parameters of the second [Genre: Fantasy, Packaging: Box] and the filter for Co-branding productions [Licensed] you obtain the data useful for creating the prototype to be submitted to the approval board.

Creation of a new LEGO set with the battle scene between Itadori and Mahito (Licensed by studio MAPPA):

for the price rage of $ 39.99, in a box you have to consider a set with ~ 300 pieces, 3 minifigures, with the possibilities to reach more then 2M of customers.

Anime Analysis	LEGO Analysis	LEGO - Anime

Anime name	Genre	Age (min)	Co-branding	Packaging
Jujutsu Kaisen ▼	Fantasy ▼	(All) ▼	Licensed ▼	Box ▼

Set composition

Avg anime score
9.728

Studios

Price Range	Median Pieces	Median Minifigs	Median price per piece
$9.99	88	2	$0.11
$19.99	158	2	$0.11
$39.99	297	3	$0.11
$69.99	493	4	$0.11
$99.99	798	5	$0.11
$149.99	1,163	6	$0.11
$199.99	1,732	5	$0.11
$249.99	2,208	9	$0.10
$250+	4,432	8	$0.10

MAPPA

Genres

Fantasy

Anime user's behaviours

1,956K — Completed
351K — Watching
198K — Plan To Watch
44K — On Hold
28K — Dropped

Number of LEGO sets by Price Range

Theme Group
■ Licensed

Price ranges: $9.99, $19.99, $39.99, $69.99, $99.99, $149.99, $199.99, $249.99, $250+

GOALS

RESULTS ACHIEVED

Datasets unification via genre and minimum age.

Functional tool for decision making, starting from both the analysis of the released LEGO sets and the analysis of the published Anime.

A proof of concept for studying new groups of themes for LEGO sets.

ISSUE AND IMPROVEMENTS

ABOUT DATA QUALITY

- Anime genre association to LEGO sets

- Recovering missing retail prices for LEGO sets

- Managing different editions of the same Anime

- If inside LEGO company, availability of additional data (such as sales data)

ABOUT DATA TREATMENT

- Create an automated pipeline to update data from new Anime and LEGO set releases

FUTURE DEVELOPMENTS

- Integration with datasets from other topics then Anime

LEGO & ANIME