**ABOUT THE DATASET**

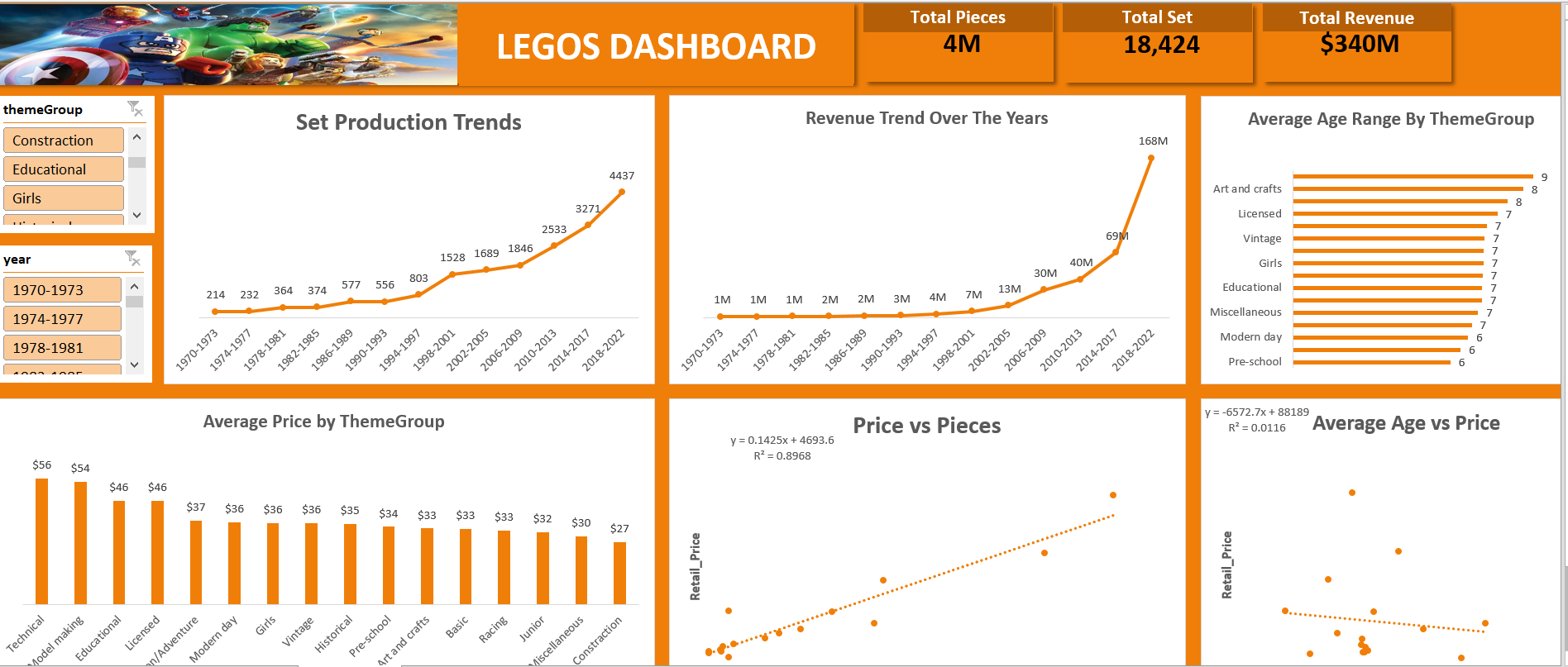
The LEGO dataset consists of data that contains sets released from 1970 to 2022, including details on each set's theme, pieces, recommended age, retail price, and image.

The dataset was gotten from Maven Analytics as a challenge to analysts to use our imagination and analytical prowess to piece together an interactive dashboard or visual that lets users explore the history and evolution of LEGO sets from the past 5 decades.

**Data Cleaning**

* Changed data type of columns
* Deleted 24 rows from set\_id as they weren’t in the same format as other set\_ids’
* 32 duplicate values were removed
* Removed last 3 columns containing URLs since they were not relevant for the analysis
* Blank spaces in numerical columns were replaced by the Mean of each respective columns.

**Dashboard**



**Key Findings**

* Over the past five decades, LEGO set production has shown a general increasing trend. The production rate has significantly risen, especially in recent years, indicating a growing diversity in LEGO offerings.
* The analysis of revenue trends reveals a positive correlation with set production. As set production increased, so did overall revenue, emphasizing the correlation between the number of sets produced and revenue generated.
* The average recommended age varies across different theme groups. Some themes are designed for younger audiences, while others cater to older age groups. Although analysis showed that most theme groups were designed for children aged 7years.
* Different theme groups show variations in average set prices. Could be that some themes such as; Technical, Model-Making, Educational and Licensed are positioned as premium, hence the reason why they have higher prices, while others are more budget-friendly.
* There is a positive correlation between the number of pieces in a set and its price. Larger sets with more pieces tend to have higher retail prices. An R-Square determines the strength of the model or how accurate it is and with an R-Square value of 0.8968,it means that the dependent variable(price) explains the variation of the independent variable(pieces) at an accuracy of 89.68%
* There is a positive correlation between the number of pieces in a set and its price. Larger sets with more pieces tend to have higher retail prices. An R-Square determines the strength of the model or how accurate it is and with an R-Square value of 0.8968,it means that the dependent variable(price) explains the variation of the independent variable(pieces) at an accuracy of 89.68%
* There is no correlation between Average Age and Price. With an R-Square value of 0.0116, it means that the independent variable (price) explains only a small proportion of the variation in the dependent variable (Average Age).

**Recommendations**

* To capitalize on the increasing demand, LEGO should continue diversifying its set themes and exploring collaborations to maintain the upward trajectory in set production.
* LEGO should strategically invest in themes and sets that historically contribute to higher revenue, ensuring a balanced mix of popular and innovative offerings.
* Continue offering a range of set pieces to cater to different customer preferences. Clearly communicate the value proposition of larger sets, emphasizing the complexity and play value they bring.

**Conclusion**

The LEGO dataset provides valuable insights into production, revenue, age, and pricing dynamics. Leveraging these insights can guide strategic decisions to enhance market competitiveness and meet diverse consumer preferences. The recommendations aim to optimize product offerings, marketing strategies, and pricing models to ensure sustained growth in the LEGO product line.