Continued Analyses of One-Sixth Scale Action Figure Collecting

Leonard Armstrong  
David Ladd  
IST-652 Scripting for Data Science  
March 18, 2021

# Introduction

## Birth of the Action Figure

The 1960s saw the advent of the action figure with *G.I. Joe*®, conceived by licensing agent Stan Weston and brought to life in 1964 by Hasbro executive Don Levine. To combat concerns about whether or not boys would play with dolls, Levine insisted that his team refer to G.I. Joe as an “*action soldier*” or “*action figure*.” Anyone who referred to G.I. Joe as a “doll” risked losing their job. Thus, to counter the cultural biases of the era, the “action figure” was born. The phrase “*Action Soldier*” can be seen on the original G.I. Joe box shown in Figure 1.

In 1964, G.I. Joe was an articulated[[1]](#footnote-1) figure slightly under 12-inches tall, making him one-sixth scale compared to the average male height. G.I. Joe’s early success led to the production of one-sixth scale figures by other companies during the mid-1960s. However, one-sixth scale action figures faded away from store shelves in the late 1970s to be replaced by smaller, less articulated figures such as Kenner’s Star Wars line of 3 ¾-inch toys. A key driver behind this change was the rising price of petroleum-based plastics. Even G.I. Joe reinvented himself in 1982 with the 3 ¾-inch *“A Real American Hero”* line of toys, a distinctly different and more expansive concept from the original everyman G.I. Joe. One-sixth scale action figures were very sparse from the late 1970s through the entire decade of the 1980s.

Figure - A 1964 G.I. Joe Action Soldier

However, one-sixth scale figures returned in 1992 when Hasbro introduced one-sixth scale versions of popular characters from the *G.I. Joe: A Real American Hero* line, initially available only as 3 ¾-inch figures. The new one-sixth scale G.I. Joe figures caught the attention of young fathers and mothers as well as nostalgic collectors.

## Action Figures as Big Business

In 2020, the toy industry market size in the United States alone was approximately $32.6 billion[[2]](#footnote-2). Action figures account for $1.66 billion of that market.[[3]](#footnote-3) Figure 2 shows action figure sales trends for the decade spanning 2011 to 2020. Additionally, Levine’s “*No ‘d’-word*” edict must have hit a broader nerve as The NPD group tracks action figures and dolls as separate and distinct “super categories.”

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated | Table  Description automatically generated |

*Figure 2- US Retail Sales for Action Figures and Accessories*

Action figures continue to have value on both the primary and secondary markets. The prototype for the original GI Joe action figure sold at auction in 2003 for $200,000.[[4]](#footnote-4)

# The Data Sets

The research documented here extends analyses from Armstrong, 2021[[5]](#footnote-5). In this research, the ***Joebase*** relational data set again serves as one structured data source. A second, semi-structured data set is also used. This second data set consists of Twitter postings, also known as *tweets*, from the five highest volume manufacturers in ***Joebase***.

## Data Set 1: Joebase

***Joebase*** is a multi-table, relational, MS SQL Server database that catalogs one of the researcher’s private one-sixth scale action figure collection. The entity-relationship diagram shown in Figure 3 describes the ***Joebase*** data model.

Diagram

Description automatically generated

Figure - Action Figure Database ER Diagram

A single query joining all tables in ***Joebase***’s relational model denormalized the data into a single, comprehensive, comma-separated values (CSV) file[[6]](#footnote-6) used in these analyses. The primary unit of the CSV structure is an action\_figure record. The denormalized model translates the one-to-many relationship between an action figure and its genres with one-hot encoding resulting in 30 new binary fields added to each action figure record. This research uses a snapshot of ***Joebase*** from February 9, 2021.

### Joebase Data Elements

The table in Figure 4 lists all the fields in the ***Joebase*** CSV.

| Column[[7]](#footnote-7) | Type | Description |
| --- | --- | --- |
| ProductId, product\_id | Integer | Unique identifier of a product. A product is an atomic unit of sales. A product typically contains one figure, but in some cases, it may include multiple figures. |
| FigureId, figure\_id | Integer | Unique identifier of a single action figure. An action figure is a single 1:6-scale plastic human form. |
| Manufacturer, manufacturer | String | Name of the company that manufactured the action figure. |
| Product, product\_name | String | Name of the product as sold. |
| Release Year, year | Integer | The year that the manufacturer released the action figure which is not necessarily the year of purchase of the action figure. However, almost all figures produced after 1995 were acquired within a year of their release. |
| Product Description, product\_descr | String | A longer description of the product. |
| Product Type | String | In all cases, this field will contain the value “1:6 figure”. This field is in the database for expansion to additional collectibles other than one-sixth scale action figures. |
| Purchased From, seller | String | Retailer or another source that sold or gifted the product. |
| purchase\_price, price | Float | The price paid for the product. |
| exclusive\_to\_retailer\_id | Integer | Unique identifier of a retailer that sold the product as an exclusive, if applicable. |
| af\_descr | String | A longer description of the action figure. This description often repeats information in product\_descr or is left null when the parent product contains only a single action figure. |
| likeness | String | The real person whose likeness is represented by the action figure. |
| storage\_location | String | Indicator of where the action is figure is either stored or displayed in the owner’s collection. |
| Genres (30 fields):  Adventure, Air Force, Armor, Army, Astronaut, Avengers, Celebrity, Civilian, Coast Guard, Comics, DC Comics, Fashion, Fire Fighter, Foreign, Horror, Knight, Marines, Martial Arts, Marvel Comics, Navy, Police, RAH/Cobra, Sci-Fi, Sports, Spy, TV/Film, Warrior, Western, World Leader, X-Men | Binary | Thirty, one-hot encoded genres that classify action figures. Multiple genres often categorize each action figure. For example, an action figure of Robert Downey Jr.’s *Iron Man* is simultaneously part of the “**TV/Film**,” “**Comics,**” and “**Marvel Comics**” genres. |

Figure - Fields in the Input Data File

### Joebase Data Cleansing and Shaping

After reading the CSV file, data cleansing commences. First, the cleansing process renames a subset of data fields. Renaming simplifies some attribute names and gives many names a consistent snake case form without whitespace. After the renaming step, null year values are set to 0 to represent an unknown purchase year. Then, the process alters the data type of the figure\_id and year columns from float64 to the more appropriate int64.

Finally, a new field, Half Decade, is derived from the year field and added to the data set. For example, the years 1995 through 1999, inclusive, are mapped to the half-decade starting in 1995. Likewise, 2000-2004 are mapped to the half-decade beginning in 2000, and so on.

### Basic Descriptive Statistics of Joebase[[8]](#footnote-8)

Immediately after intake, ***Joebase*** contains 497 records with 43 attributes per record. After cleansing, the data set consists of 491 records with 44 attributes per record.

#### Year of Release

***Joebase*** action figure years-of-release span from 1964 (vintage G.I. Joe figures) to 2020. Figure 5 shows a distribution of figure counts across these years. The figure shows a peak of over 60 action figures from 2001. Action figures released over the past decade (2011-2020) come in at more modest numbers of ten or less per year.

Chart, histogram

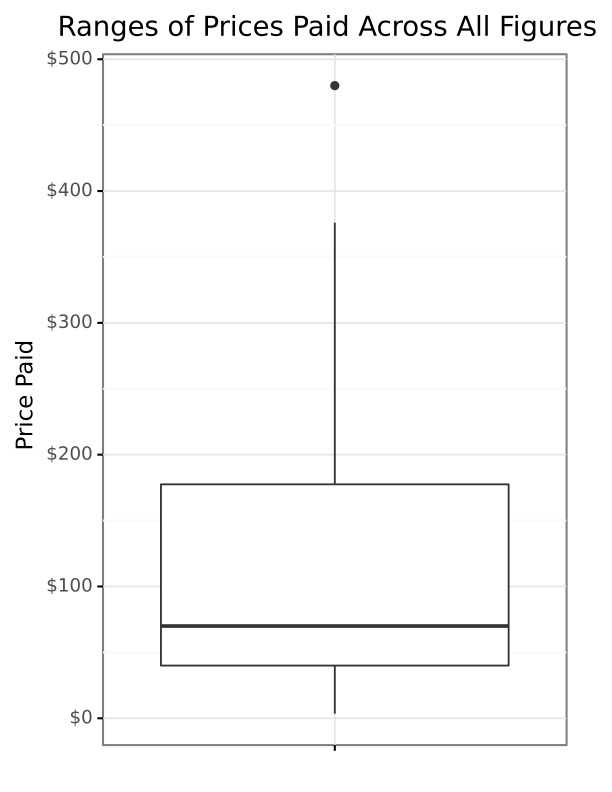
Description automatically generated

Figure - Distribution of Figures by Year of Release

The chart shows that the collection contains a small volume of figures from the “vintage” era of one-sixth scale figures, 1964-1976; however, most action figures in the collection were produced in 1996 and after. In fact, the collection started with the purchase of a single reproduction, G.I. Joe Action Soldier figure that was packaged and sold along with a book titled *G.I. Joe: The Story Behind the Legend[[9]](#footnote-9)* as a “Masterpiece Edition” boxed set.

#### Prices

There are 166 figures with non-zero prices recorded with a mean cost of $110.31 per product. Figure 6’s boxplot and associated quartile statistics show the range of prices across all action figures in ***Joebase***.

Table

Description automatically generated

Figure - Range of Action Figure Prices

## Data Set 2: Tweet Data

A series of action figure-related Twitter tweets serve as a second source of research data. Five separate tweet sets were pulled, one set for each of the top five highest-volume manufacturers in ***Joebase***, as shown in Figure 7.

|  |  |  |
| --- | --- | --- |
| Manufacturer | Twitter Screen Name | Figure Volume |
| Hasbro | Hasbro | 92 |
| Sideshow | collectsideshow | 56 |
| Hot Toys | hottoysofficial | 48 |
| DC Direct/DC Collectibles | DCCollectibles | 31 |
| G.I. Joe Collector's Club | The\_GIJoe\_Club | 29 |

Figure - Top Five Manufacturers in Joebase

### Tweet Selection Process

Tweet data pulls used the Tweepy[[10]](#footnote-10) Twitter API for Python. Tweet requests required English-language Tweets, in reverse chronological order, in batches of 100 tweets, until no more tweets were available (i.e., the Twitter feed was exhausted), or a 3000-tweet threshold was met or exceeded. Batch sizes are 100 because Tweepy will not return more than 100 tweets at a time, regardless of how many are requested. The tweet limit of 3,000 is a volume that is sufficient to show trends while not overloading Twitter to a point where access would be temporarily disabled or throttled due to excessive requests.[[11]](#footnote-11) The query string used for each tweet batch was the manufacturer’s Twitter screen name, shown in Figure 7. The “@” character is not part of the query string, so tweets *from* the manufacturer account as well as tweets *mentioning* the manufacturer account would all be pulled.

After pulling the tweets, the five batches were stored on disk in Python’s fast read/write pickle format. Extracting the data one time, pickling the results, and using only the pickle files for subsequent research ensured consistency of tweet data across time and team member usage.

### Tweet Data Elements

Tweepy returns Twitter status objects (tweets) as JSON documents[[12]](#footnote-12). For easier processing, the JSON documents are converted to rows in a pandas DataFrame whose columns align with the JSON attributes. The resulting DataFrame has the column structure shown in Figure 8.

| Column | Type | Description[[13]](#footnote-13) |
| --- | --- | --- |
| created\_at | String | Creation time of this tweet in UTC. |
| id | Integer | The unique identifier of the status. |
| id\_str | String | The unique identifier of the status, as a string. |
| full\_text | String | The text of the status from extended mode searches. |
| truncated | Boolean | Indicates whether the value of the text parameter was truncated, for example, as a result of a retweet exceeding the original Tweet text length limit of 140 characters. |
| display\_text\_range | List | For extended mode searches, an array of two Unicode code point indices, identifying the inclusive start and exclusive end of the displayable content of the Tweet. |
| entities | Dictionary | The parsed entities of the status such as hashtags, URLs, etc. |
| extended\_entities | Dictionary | When between one and four native photos or one video or one animated GIF are in Tweet, this field contains an array of media metadata. |
| metadata | Dictionary |  |
| source | String | The utility used to post the Tweet as an HTML-formatted string. |
| in\_reply\_to\_status\_id | Integer | The identifier of the status being replied to. |
| in\_reply\_to\_status\_id\_str | String | The identifier of the status being replied to as a string. |
| in\_reply\_to\_user\_id | Integer | The identifier of the user being replied to. |
| in\_reply\_to\_user\_id\_str | String | The identifier of the user being replied to as a string. |
| in\_reply\_to\_screen\_name | String | The screen name of the user being replied to. |
| user | Dictionary | The User object of the poster of the status. |
| geo | Dictionary | The geo object of the status. |
| coordinates | Dictionary | Represents the geographic location of this Tweet as reported by the user or client application. |
| place | Dictionary | When present, it indicates that the tweet is associated (but not necessarily originating from) a Place. |
| contributors |  | Deprecated |
| is\_quote\_status | Boolean | Indicates whether the status is a quoted status or not. |
| retweet\_count | Integer | The number of retweets of the status. |
| favorite\_count | Integer | The number of likes of the status. |
| favorited | Boolean | Indicates whether the authenticated user has favorited the status or not. |
| retweeted | Boolean | Indicates whether the authenticated user has retweeted the status or not. |
| possibly\_sensitive | Boolean | Indicates whether the status is sensitive or not. |
| lang | String | The language of the status. |
| retweeted\_status | Dictionary | Retweets can be distinguished from typical Tweets by the existence of a retweeted\_status attribute. This attribute contains a representation of the original retweeted Tweet. |
| quoted\_status\_id | Integer | For quoted tweets, this is the identifier of the quoted tweet. |
| quoted\_status\_id\_str | String | The identifier of the quoted tweet as a string. |
| quoted\_status | Dictionary | This attribute contains the original quoted tweet object. |

Figure - Original Tweet Data Attributes

### Tweet Data Cleansing and Derivations

One of the tweet data research questions requires time series analysis, highlighting the importance of the created\_at element. However, as seen in Figure 8, the created\_at field is provided as a string, owing mainly to the lack of date and time types in JSON. The created\_at field was parsed, and a series of related fields were added to the DataFrame, enabling time series analysis. Figure 9 shows the derived date and time-related attributes.

| Column[[14]](#footnote-14) | Type | Description |
| --- | --- | --- |
| day\_of\_week | String | Three-character day of the week (Sun-Sat) parsed from created\_at. |
| month | Integer | Month (1-12) parsed from created\_at. |
| day\_of\_month | Integer | Day/date within the month (1-31) parsed from created\_at. |
| hour | Integer | Hour of the day in 24-hour format (0-23) parsed from created\_at. |
| minutes | Integer | Minutes (0-59) parsed from created\_at. |
| Seconds | Integer | Seconds (0-59) parsed from created\_at. |
| tz\_offset | String | Timezone offset parsed from created\_at. All Tweets seem to be listed in UTC, making tz\_offset consistent at “-0:00”. |
| year | Integer | Four-digit year parsed from created\_at. |
| datetime | datetime.datetime | created\_at, parsed, and recreated in datetime format. |
| report\_timestamp | datetime.datetime | The datetime field, standardized to AM or PM. All AM hours have the timestamp of 0:00. All PM hours have the timestamp of 12:00. This field supports half-day reporting. |
| group | String | The search string used to find the tweet. This field supports reporting. |

Figure - Derived Tweet Fields

### Basic Descriptive Statistics for Tweets

Twitter provides approximately one-week of tweet history using the Tweepy API. This analysis uses tweets pulled on Saturday, March 13, 2021, at 3:03 PM EST. Figure 10 shows the resulting tweet volumes, and Figure 11 shows a pie chart of the percent of the total tweets for each manufacturer.

|  |  |  |
| --- | --- | --- |
| Manufacturer | Tweets Pulled[[15]](#footnote-15) | Oldest Tweet |
| Hasbro | 3,000 | March 12 |
| Sideshow | 3,097 | March 9 |
| Hot Toys | 1,112 | March 5 |
| DC Direct/DC Collectibles | 7 | March 6 |
| G.I. Joe Collector's Club | 1 | March 10 |

Figure - Manufacturer Tweet Batches

Chart, pie chart

Description automatically generated

Figure - Tweet Percentage Breakdown

Since the volumes of tweets for Hot Toys, DC Collectibles, and the G.I. Joe Collector’s Club are all less than 3000, the implication is that the Twitter feed for those queries was exhausted. However, Hasbro and Sideshow both hit the tweet limit. Hasbro’s tweets go back only one day prior, implying that Hasbro was the most popular query subject of the five during that week of March 2021. Sideshow’s tweet limit was reached after 4-days of tweet history, March 9 through March 13.

# Analysis

Analyses of the ***Joebase*** data set targeted the following series of questions:

* What function predicts the average figure price in a given year?
* What is the relationship between manufacturers and genres?
* Who are the “influencers” in the available period of data?
* How active are action figure conversations on social networks?

## Question 1 – What function predicts the average figure price in a given year?

Armstrong, 2021 asked the question, *“What are the average prices, per year-of-release, of figures?”* In answer to that question, Figure 12 showed the rise in the mean purchase price of action figures in the collection produced from 1995 to 2020, along with the statement *“One can see a trend of prices rising at a superlinear rate.”*

Chart, line chart

Description automatically generated

Figure - Mean Figure Price Per Year of Release

The question of finding a predictive function for prices is a logical follow-on. Given the “superlinear” comments from Armstrong, 2021, three different regressions were compared, using year as the sole predictor variable. The three regressions were: a linear regression, an order-2 polynomial regression using the square roots of the year variable, and an exponential regression using the natural logarithms of the year variable.

*Appendix A – Results Three Regressions on Average Annual Price* shows the three regressions' full results. However, Figure 13 summarizes the R2 and Prob (F) values of the regressions. Additionally, all three regressions have P-values much less than 0.05 on all coefficients.

|  |  |  |
| --- | --- | --- |
| Regression Form | R2 | Prob(F) |
| Linear | 0.814 | 2.98⨉10-10 |
| Polynomial | 0.816 | 2.61⨉10-10 |
| Exponential | 0.778 | 2.63⨉10-9 |

Figure - Regression Results

Both the linear and polynomial regressions show similar R2 results making a statistics-based case in favor of either function reasonable. However, predicting 2-years out, the average price for a figure grows by approximately $66 or around 27% for the polynomial model. Conversely, the average price increases by $20 or 8% over the same period in the linear model. The US expects a 2.2% inflation rate in 2021 and 1.6% inflation in 2022.[[16]](#footnote-16) The linear model price increases are more than double that inflation rate making the polynomial model increases seem excessive and not economically sustainable.

With this in mind, the linear model is selected as the final price predictor with the prediction function, . This function has an x-origin of 1995 and is not valid before that year.[[17]](#footnote-17) Figure 12 shows the prediction line along with individual product prices (the small grey x marks) and the annual mean price (the larger blue dots).

Chart, scatter chart

Description automatically generated

Figure - Linear Predictor for Average Product Prices Per Year

## Question 2 – What Is the Relationship Between Manufacturers and Genre?

Figure 15[[18]](#footnote-18) shows the volumes of action figures within each of the 30 tracked genres.

| Genre | Volume | Genre | Volume | Genre | Volume |
| --- | --- | --- | --- | --- | --- |
| TV/Film | 128 | World Leader | 18 | Warrior | 6 |
| Comics | 95 | Spy | 13 | Fire Fighter | 6 |
| Army | 64 | Civilian | 13 | Western | 6 |
| Foreign | 50 | Marvel Comics | 11 | DC Comics | 5 |
| Adventure | 43 | Astronaut | 10 | Fashion | 5 |
| Sci-Fi | 36 | Sports | 9 | Armor | 4 |
| Horror | 34 | Martial Arts | 9 | Knight | 2 |
| Navy | 31 | Celebrity | 8 | X-Men | 1 |
| Air Force | 27 | RAH/Cobra | 8 | Avengers | 1 |
| Marines | 24 | Police | 7 | Coast Guard | 1 |

Figure - All Genres, Sorted by Volume of Figures in Each Genre

***Joebase*** currently tracks 66 manufacturers. Figure 16 shows a histogram of figure volumes for the top 16 manufacturers. Hasbro, the manufacturer of the G.I. Joe brand of figures, tops the collection with over 90 figures. Conversely, the sixteenth largest manufacturer provides only five figures to the collection. Although not shown in Figure 16, one-half of the manufactures tracked by ***Joebase*** (33 of 66) represent only one figure each.

Chart, histogram

Description automatically generated

Figure - Top 16 Manufacturers by Action Figure Volume

The heatmap in Figure 17 visualizes the relationship between manufacturer and genre.[[19]](#footnote-19) The heatmap shows only the top 11 manufacturers, each contributing 15 or more action figures to the collection.

TV/Film for Sideshow and Hot Toys are the darkest points on the map due to the many move franchise licenses held by both companies such as Star Wars (both), Indiana Jones (both), Universal Monsters (Sideshow), James Bond (Sideshow), the Marvel Cinematic Universe (Hot Toys), and the DC Comics Extended Universe (Hot Toys). DC Collectibles and Hot Toys also show strong representation in the Comics genre.

Hasbro has representation in several genres, including being the top manufacturer of Air Force, Army, Marines, and Navy figures in the collection. A reasonable question here is, *“Should individual military branches be separate genres, or should ‘military’ be a single consolidated genre?”* When G.I. Joe first hit toy shelves in 1964, the line contained distinct Action Soldier, Action Sailor, Action Pilot, and Action Marine figures representing each of the military's four branches. Since 1:6 scale action figures started with a military line that distinguished individual branches of the military, it was logical to use separate military branch genres.

Chart, scatter chart

Description automatically generated

Figure - Heatmap of Genres vs. Top 16 Manufacturers

However, a heatmap with combined military services has a different look, as shown in Figure 18 with Air Force, Army, Coast Guard, Marines, and Navy merged into a single column named *Military*. Eight of the top 11 manufacturers contribute to the military genre, with the Hasbro/Military cell now the “hottest” section on the map.

Chart, scatter chart

Description automatically generated

Figure - Heatmap of Genres vs. Top 16 Manufacturers (Combining Military Genres)

## Question 3 – Who Are the “Influencers” in the Available Period of Tweet Data?

In addition to the targeted vendor screen names, the tweet data contains additional screen names of users interested in action figures. As users increase Twitter traffic for specific topics, trends emerge. Tweet volumes of the top 30 users across all captured tweets help determine the influencers. Figure 19 provides a Pareto chart[[20]](#footnote-20) showing these top users. The highest volume tweeter, collectsideshow (the Twitter screen name for Sideshow Collectibles), is one of the five manufacturers. All other screen names are from non-manufacturers.

**Chart

Description automatically generated**

Figure - Pareto Chart of Top 30 Screen Names

Even including the high tweet volume from Sideshow, it takes 12 of the top 30 accounts (40%) to get to 80% of the total tweet volume. This 80/40 ratio far exceeds the Pareto Principle, wherein 80% of the consequences come from 20% of the causes.

The question of influencers is still somewhat open. One can visually determine three groups of users from the Pareto based on significant drops in tweets between them. The first group contains only Sideshow. The second group includes the quartet of HobsonAnton, XenaGifts, BigTopApps, and jkwdiamond. After jkwdiamond the per-user tweet volume has another significant drop. The last group of users ranges from Champs through (from left-to-right) ActionMarvel. Sideshow’s tweet volume seems to place them in the category of an influencer, a reasonable classification since Sideshow is both a producer of 1:6 scale figures and the official distributor of Hot Toys 1:6 scale figures in North America and Europe.[[21]](#footnote-21)

## Question 4 – How Active Are Action Figure Conversations on Social Networks?

Using Twitter as a representative social network, the process searched for the screen names of each of the five highest volume manufacturers in ***Joebase***. To interpret the results, one must remember a few key constraints. First, Twitter does not provide an infinite stream of tweets. Search results will extend back approximately one week of tweet history. Second, organizations that provide APIs into their data and services are conscious about the activity volumes that hit their services. High rates of activity from a single endpoint can imply undesirable activities such as denial-of-service attacks or abuse of terms of service. These activities may lead to a user’s account being disabled.[[22]](#footnote-22) Thus, the developed process halted data collection for a manufacturer after approximately 3,000 tweets to avoid being disabled or rate limited.

Figure 20 shows a time series of activity volumes for each of the top five manufacturers in ***Joebase***. Volumes were summed over 12-hour increments starting on the afternoon of March 13, 2021, and extending back as far as March 5 for one manufacturer. The choice to limit tweet collection to 3,000 tweets per manufacturer made for odd timeline visuals for Hasbro and Sideshow, who reached their 3,000 tweet volumes within two and five calendar days of tweet history, respectively.

Chart, line chart

Description automatically generated

Figure - Twitter Timeseries

Figure 20 only shows four of the top five manufacturers. The query for “The\_GIJoe\_Club” returned only one tweet[[23]](#footnote-23), an insufficient volume to form a new line. Also, DC Collectibles has a shallow tweet volume during this period, with only seven tweets. Thus, the DC Collectibles line is very near zero as no one 12-hour period contained more than two tweets.

### The Hasbro Activity Spike

The remaining three manufacturers all show a spike in activity at some point during the period shown. Hasbro’s and Sideshow’s spikes are relatively large, between 1,000 and 1,500 tweets each, whereas Hot Toys’ spike is smaller, about 250 tweets during the AM hours of March 11, 2021. To better understand the events that drove those three activity spikes, word clouds were created for each manufacturer's spike period. The green word cloud in Figure 21 represents Hasbro’s spike from the PM hours of March 12, 2021. A combination of the tokens found in the cloud plus a spot check of the tweets during this time shows the spike resulted, in large part, from the announcement of a new line of *The Real Ghostbusters*[[24]](#footnote-24) toys from Hasbro. Words like ghostbusters, ecto, ghost, and real provide clear clues to this topic.

Text

Description automatically generated

Figure - Hasbro Word Cloud

### The Sideshow Activity Spike

Figure 22’s blue word cloud is for Sideshow’s March 10, 2021, AM spike. This spike is mainly due to Sideshow’s announcement of a giveaway promotion of the Rocket Racoon 1:6 scale figure. Tokens like rocket, raccoon, giveaway, chance, and win are clear indicators of the announcement. The prominent token “@brandondavisbd” is the Twitter screen name of the Sideshow employee who posted the original, oft retweeted, announcement.

Text

Description automatically generated

Figure - Sideshow Word Cloud

### The Hot Toys Activity Spike

The final red word cloud, shown in Figure 23 represents the smaller-in-volume Hot Toys spike from the AM hours of March 11, 2021. The tweets centered mainly on the announcement of Hot Toys of Vision and Scarlett Witch figures from the just-completed *WandaVision* series on Disney+. The word cloud contains tokens for wandavision, scarlettwitch, vision, thevision, wanda, scarlett, witch, and elizabetholsen. As an inside look into the 1:6 scale collecting hobby, one can also see the tokens hair and sculpted. The preference for sculpted vs. rooted hair for 1:6 scale female figures can ferociously bifurcate the 1:6 scale collecting community. The new Wanda/Scarlett Witch figure was released with sculpted hair, whereas prior releases of the character had rooted hair.[[25]](#footnote-25)

Text

Description automatically generated

Figure - Hot Toys Word Cloud

# Conclusion

The authors’ analysis of the 1:6 scale action figure domain used data from a single action figure collection and tweets related to the top five manufacturers in that collection.

First, a linear function was derived that predicts average action figure prices for a given year. 81% of the variability of average prices across years can be explained solely by a figure’s year of production.

Second, a heatmap helped to visualize the relationship between manufacturers and genres. The heatmap has a very different look after combining the five military branch genres[[26]](#footnote-26) into a single genre.

Third, tweet volumes helped to define potential action figure word-of-mouth influencers. Sideshow Collectibles, the producer of multiple lines of action figures and a distributor for Hot Toys action figures, led all other users in tweet volume.

Finally, a time series of action figure-related tweets showed periods where manufacturer-related tweets spiked in volume. For the short timespan evaluated, “high volume” activities translate to between 1,000 and 1,500 tweets in 12 hours.

# Appendix A – Results Three Regressions on Average Annual Price

## Linear Regression

The full results of the linear regression are:

**REGULAR LINEAR REGRESSION**

OLS Regression Results

==============================================================================

Dep. Variable: avg\_price R-squared: 0.814

Model: OLS Adj. R-squared: 0.807

Method: Least Squares F-statistic: 105.2

Date: Tue, 16 Mar 2021 Prob (F-statistic): 2.98e-10

Time: 19:21:23 Log-Likelihood: -129.24

No. Observations: 26 AIC: 262.5

Df Residuals: 24 BIC: 265.0

Df Model: 1

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const -1.944e+04 1905.451 -10.200 0.000 -2.34e+04 -1.55e+04

year 9.7371 0.949 10.259 0.000 7.778 11.696

==============================================================================

Omnibus: 1.637 Durbin-Watson: 0.994

Prob(Omnibus): 0.441 Jarque-Bera (JB): 1.461

Skew: 0.471 Prob(JB): 0.482

Kurtosis: 2.321 Cond. No. 5.37e+05

==============================================================================

### Polynomial Regression

The full results of the order-2 polynomial regression are:

**QUADRATIC REGRESSION**

OLS Regression Results

==============================================================================

Dep. Variable: sqrt\_price R-squared: 0.816

Model: OLS Adj. R-squared: 0.809

Method: Least Squares F-statistic: 106.7

Date: Tue, 16 Mar 2021 Prob (F-statistic): 2.61e-10

Time: 19:25:12 Log-Likelihood: -50.577

No. Observations: 26 AIC: 105.2

Df Residuals: 24 BIC: 107.7

Df Model: 1

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const -945.4005 92.486 -10.222 0.000 -1136.283 -754.518

year 0.4758 0.046 10.328 0.000 0.381 0.571

==============================================================================

Omnibus: 2.450 Durbin-Watson: 0.962

Prob(Omnibus): 0.294 Jarque-Bera (JB): 2.048

Skew: 0.575 Prob(JB): 0.359

Kurtosis: 2.246 Cond. No. 5.37e+05

==============================================================================

### Exponential Regression

The full results of the exponential regression are:

**EXPONENTIAL REGRESSION**

OLS Regression Results

==============================================================================

Dep. Variable: log\_price R-squared: 0.778

Model: OLS Adj. R-squared: 0.768

Method: Least Squares F-statistic: 83.98

Date: Tue, 16 Mar 2021 Prob (F-statistic): 2.63e-09

Time: 19:25:12 Log-Likelihood: -13.406

No. Observations: 26 AIC: 30.81

Df Residuals: 24 BIC: 33.33

Df Model: 1

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const -198.5069 22.140 -8.966 0.000 -244.202 -152.812

year 0.1011 0.011 9.164 0.000 0.078 0.124

==============================================================================

Omnibus: 0.990 Durbin-Watson: 1.187

Prob(Omnibus): 0.609 Jarque-Bera (JB): 0.765

Skew: 0.401 Prob(JB): 0.682

Kurtosis: 2.752 Cond. No. 5.37e+05

==============================================================================

1. Movable joints and body parts allowing the figure to be posed. [↑](#footnote-ref-1)
2. The Toy Association. *U.S. Sales Data.* The NPD Group. 2021. <http://www.toyassociation.org/ta/research/data/u-s-sales-data/toys/research-and-data/data/us-sales-data.aspx>. [↑](#footnote-ref-2)
3. Data and graphic are copyright © Statistica. *Toy action figure and accessory retail sales in the United States from 2011 to 2020*. 2021. <https://www.statista.com/statistics/247398/toy-sales-in-the-us--action-figures-accessories-and-role-play/>. [↑](#footnote-ref-3)
4. Coyle, Anthony. Gazette Review. *Top Ten Most Expensive and Valuable Action Figures*. 2017. <https://gazettereview.com/2016/08/top-ten-expensive-valuable-action-figures/>. [↑](#footnote-ref-4)
5. Armstrong, Leonard. *Analysis of a One-sixth Scale Action Figure Collection*. Homework #1. IST-652, Syracuse University. February 2021. [↑](#footnote-ref-5)
6. The denormalized structure maintained all significant information except for the series data which is not part of this analysis. [↑](#footnote-ref-6)
7. In the cases where the **Column** description contains two values, one represents the name that is present in the CSV file’s header row and the other representing a column renaming performed during data cleansing. [↑](#footnote-ref-7)
8. Section 2.1.3 and its subsections are taken from Armstrong, 2021. However, only key elements of ***Joebase*** descriptive statistics that are applicable to this new research are reproduced here. [↑](#footnote-ref-8)
9. Levine, Don with John Michlig, *G.I. Joe: The Story Behind the Legend, An Illustrated History of America’s Greatest Fighting Man*. Chronicle Books, 1996. [↑](#footnote-ref-9)
10. https://www.tweepy.org/ [↑](#footnote-ref-10)
11. Remembering that these requests were made repeatedly during the develop/test/debug cycle. [↑](#footnote-ref-11)
12. Using the \_json attribute of a returned Tweepy status object. [↑](#footnote-ref-12)
13. The descriptions provided come from either Tweepy or Twitter documentation. [↑](#footnote-ref-13)
14. In the cases where the **Column** description contains two values, one represents the name that is present in the CSV file’s header row and the other representing a column renaming performed during data cleansing. [↑](#footnote-ref-14)
15. Tweets are pulled in up-to 100 tweets batches until the total pull greater than or equal to 3000. Some requests for 100 tweets may come back with less resulting in the total tweets pulled ranging from 3000 to 3099. [↑](#footnote-ref-15)
16. Trading Economics. 2021, United States Inflation Rate. <https://tradingeconomics.com/united-states/inflation-cpi#:~:text=In%20the%20long%2Dterm%2C%20the,according%20to%20our%20econometric%20models>. [↑](#footnote-ref-16)
17. An argument can be made that the function is only valid from 1996 where it crosses the y=0 origin making prices for 1995 negative. [↑](#footnote-ref-17)
18. First shown in Armstrong, 2021. [↑](#footnote-ref-18)
19. Converse to the defaults on many heatmap generators, the heatmaps produced here use lighter colors to represent lesser values and darker colors to represent larger values. [↑](#footnote-ref-19)
20. The Pareto chart was created in Excel, not Python. [↑](#footnote-ref-20)
21. Hot Toys in a Hong Kong-based company. [↑](#footnote-ref-21)
22. Per <https://developer.twitter.com/en/developer-terms/policy>: Your license agreement with Twitter limits your use of the Twitter API and Twitter Content. Among other things, the Twitter API has rate limits which help to ensure fair data usage and to combat spam on the platform. You may not exceed or circumvent rate limits, or any other limitations or restrictions described in this Policy or your agreement with Twitter, listed on the Developer Site, or communicated to you by Twitter. [↑](#footnote-ref-22)
23. The G.I. Joe Collector’s club is no longer in existence. [↑](#footnote-ref-23)
24. *The Real Ghostbusters* was a 1980s cartoon series based off the *Ghostbusters* film franchise. [↑](#footnote-ref-24)
25. If rooted vs. sculpted hair on your collectibles is one of your *causes du jour*, you lead a pretty charmed life. [↑](#footnote-ref-25)
26. Army, Navy, Air Force, Marines, and Coast Guard. To the best of the authors’ knowledge no Space Force figures have yet been produced. 🚀 [↑](#footnote-ref-26)