

# **Marvel & DC Character Analysis**

Gabriel Valenzuela

Bellevue University

DSC 680: Applied Data Science

Professor Catie Williams

November 21, 2020

## Abstract

Marvel and DC comics are famous comic book creators that have focused on creating stories based on superheroes and villains for more than eighty years. Throughout this long time as top creators and rivals, they have created nearly over twenty thousand characters from famous superheroes such as Batman and Spiderman to villains like Thanos and the Joker. With each comic that is released throughout generations,



Marvel and DC Characters – Appendix A

the more diverse each of these characters become in various ways. It can either be from their gender, their sex, if they have a dual secret and public identity, to even their hair color. By analyzing the data generated by each record of the characters that are released to the public, this analysis attempts to predict whether a future character will be aligned as a good or bad character from both Marvel and DC. To complete this analysis of predicted characteristics, I utilized a decision tree model to predict the aspects of a character

and what the comic book world is trending in the creation of future heroes and villains in how they are appearing to better represent their audiences as a whole.

Across the number of characters that have appeared in all of the comics, it has predominately been a certain character type that has appeared for heroes with the villains staying consistent within certain decades. However, as the culture and the society of our nation changes, so do the characters that we share stories with and desire to become. Based on the results of this analysis, it will present information to those interested in past, current, and future characters of these worlds as they continue to grow in popularity in our culture.

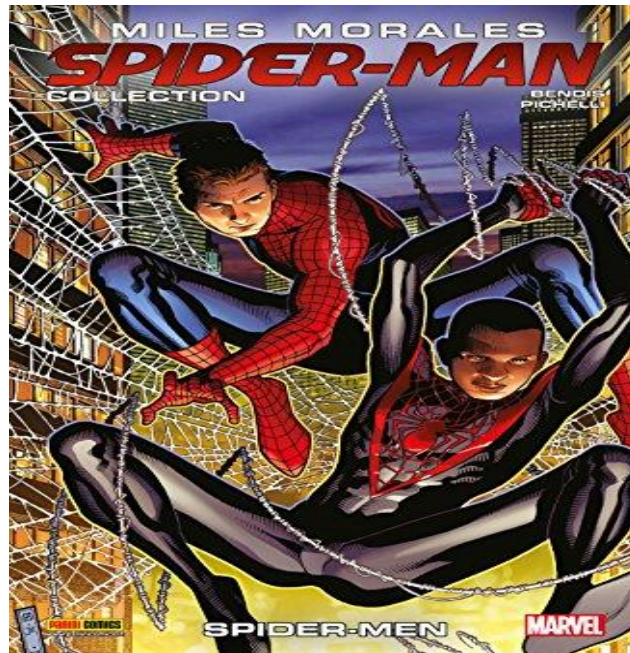
## **Marvel & DC Character Analysis**

### **Background**

Superhero comics have been around since the 1930s where they began to rise in popularity throughout the nation during the 1940s and remained there until the 1960s. Within these comics, they will describe various stories about superheroes and the universes these characters inhabit including all of the villains that accompany them. As these comics came about, two major companies became the main faces of the industry. One being Marvel including famous heroes such as the X-men, Spiderman, and the Avengers, and the other being DC which would have characters such as Wonder woman, Batman, and Superman.

When it comes to characteristics of each of these characters from both DC and Marvel, they can be broken down into a few factors such as their alignment of good versus bad, physical aspects including eye and hair color, as well as the their current status and number of appearances throughout every comic that has been released even if it does not include their own series. As generations have come and gone, these characters' attributes have been altered or built upon by adding more complexity to them. However, as the years have passed throughout different ages of comics for both Marvel and DC, they have typically have been a mirror of their creators, the audience, and the world around them. The goal was always to create stories that allowed for people to connect those in the story by seeing themselves and stopping fears that they may come across by gaining confidence from the heroes they idled.

Because creators and the companies understood the connection to the audience, they wanted to reflect them in some form within their comics by creating different variations of characters or new ones all together. For example, creators understood that more people of different races began to show more interest in various superheroes. Therefore, a new Spiderman was created that was able to speak Spanish and was of different backgrounds named Miles Morales.



#### Diverse Spiderman – Appendix B

With this new wave of heroes and villains appearing within the stories along with keeping the old character types of the old ones, it has created a diverse universe of characters that allows for many communities to relate to and continue reading their stories allowing for both of these creators to continue releasing more adventures with both the new and old characters. At the same time, with the immense popularity of superhero movies and shows being released on yearly basis, it allowing for more and more people to identify with these characters expanding the audience further which in turn expands the diversity of those that we watch on the screen.

## **Methods**

This analysis will be using the “Comic Characters” data set gathered by G\_R\_S from Kaggle. The data set consists of all of the characters that have made appearances in both DC and Marvel comics since the first issue published by each one of them which is more than twenty thousand characters. The data set contains information about each character including identity, alignment such as good or bad, their gender, hair and eye color, and the number of appearances that they have made across various comic series. With comics having different waves of popularity having different versions of similar characters and spawning multiple television shows and movies from them, the data set allows someone to focus on various characters and trends to determine future heroes and villains that may appear from each of the comic giants.

After importing the data into a Jupyter Notebook by utilizing Python, I first began by creating a separate data frame for DC, Marvel, and a combined set of them both to gain a general picture of the character. At the same time, in order to create the features for this set, I needed to convert all of the categorical variables into numerical variables that can be used within the model to be able to predict the newest character in each comic. For each of the records on a character, I also had to remove certain aspects since it did not involve descriptive sides of each character such as their source within a certain wiki.

Once the data has been prepared and edited for feature selection and modeling, the data was split by seventy and thirty for training and for testing. After the data has been split appropriately, I initialized a decision tree model to be able to predict a superhero comic book character’s alignment, good character or bad character. The goal of this model is to see how accurate it could predict these aspects.

## **Results**

The main goal of this project is to examine Marvel and DC comic characters to try and predict the future characters of each of the comic giants. Throughout various comics throughout the decades, both of these

groups had altered their characters to incorporate the current culture of the nation which included more characters of different races and sexes.

From data explorations thus far in the analysis, the distribution of the characteristics within the comics include the following:

	Sex	Identity	Align	Alive
<b>DC</b>	28% Female / .2% Genderless / 69% Male / .01% Transgender	35% Public / 34% Secret / 31% Unknown	41% Bad / 41% Good / 8% Neutral / 10% Reformed Criminals	24% Deceased / 76% Alive
<b>Marvel</b>	.2 Agender / 23% Female / .01 Genderfluid / 71% Male	.09% Known to Authorities / 10% No Dual Identity / 27% Public / 38% Secret	41% Bad / 28% Good / 13% Neutral	22% Deceased / 78% Alive

Table of Data Exploration Results – Appendix C

Looking at these basic results of the exploration, it is showing that there are certain characteristics of most characters in both comic atmospheres. Based on the model that was deployed, the Marvel decision tree model was able to predict future alignments with 64.80% accuracy, while the DC decision tree model was only able to predict with a 54.27% accuracy. A main reason for the observable difference in accuracies is due to the number of characters that have appeared in each set as well as the other alignments that existed within them. For Marvel, it had the ability to learn with about 7000 more records of characters. At the same time, DC had to be concerned with both neutral and reformed criminals within the data set which limited the amount of records that could be used to determine good and bad alignments within its characters.

## **Discussion/Conclusion**

Even though there is a high percentage towards certain aspects such as male, alive, and secret identity, there is a possibility of characters having different traits depending on the trends within the past decade or so. Because of the golden age of comics, there was a massive boom of characters that generally had the same characteristics both good and bad where characters were not killed and continued reappear on a repetitive basis. Therefore, from this analysis, I think it will provide an insight and future predictions of characters to not only understand DC and Marvel, but the attitude of the current society since comics like to mirror the culture and its audience. Based on this analysis of the characters and the history of superhero comics, we will begin to see characters that represent the newest generation of readers to better connect with the stories being told with the villains being re-imagined to include different genders, sexes, and other characteristics that we saw from this dataset of both comic creators.

## **Acknowledgements**

Marvel Comics, DC Comics, Kaggle, Bellevue University

## **References**



DeForest, T. (2020, October 02). Marvel Comics. Retrieved October 31, 2020, from <https://www.britannica.com/topic/Marvel-Comics>

DeForest, T. (2014, January 17). DC Comics. Retrieved October 31, 2020, from <https://www.britannica.com/topic/DC-Comics>

Marvel. (2020). Marvel Characters, Super Heroes, & Villains List: Marvel. Retrieved October 31, 2020, from <https://www.marvel.com/characters>

DC. (2020). Characters. Retrieved October 31, 2020, from <https://www.dccomics.com/characters>

Murphy, K. (2016, February 01). Analyzing Female Gender Roles in Marvel Comics from the Silver Age (1960) to the Present. Retrieved October 31, 2020, from <http://www.inquiriesjournal.com/articles/1449/analyzing-female-gender-roles-in-marvel-comics-from-the-silver-age-1960-to-the-present>

Shendruk, A. (2017, April 1). Analyzing the Gender Representation of 34,476 Comic Book Characters. Retrieved October 31, 2020, from <https://pudding.cool/2017/07/comics/>

Kirkpatrick, E., & Scott, S. (2015, September 15). IN FOCUS: Gender Identity and the Superhero Representation. Retrieved October 31, 2020, from [https://cdn.ymaws.com/www.cmstudies.org/resource/resmgr/In\\_Focus\\_Archive/InFocus\\_55-1.pdf](https://cdn.ymaws.com/www.cmstudies.org/resource/resmgr/In_Focus_Archive/InFocus_55-1.pdf)

Wilson, M. (2014, October 14). Superhero Demographics Still Overwhelmingly Male. Retrieved October 31, 2020, from <https://comicsalliance.com/fivethirtyeight-superhero-demographic/>

Wilson, M. (2014, October 14). Superhero Demographics Still Overwhelmingly Male. Retrieved October 31, 2020, from <https://comicsalliance.com/fivethirtyeight-superhero-demographic/>

McMillan, G. (2019, March 11). 10 Superheroes That Switched Up Demographics From Their Predecessors. Retrieved October 31, 2020, from <https://www.hollywoodreporter.com/heat-vision/superheroes-who-changed-race-gender-915103>

## Appendix

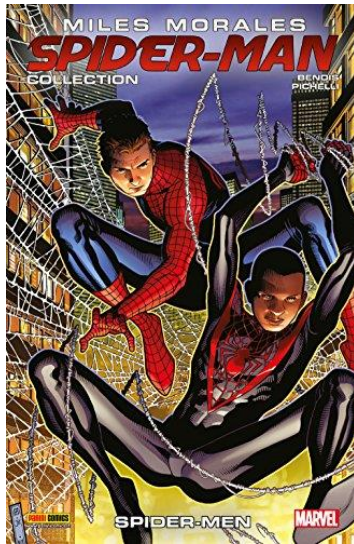
### Appendix A

#### Marvel and DC Characters



### Appendix B

#### Diverse Spiderman – Appendix B



## Appendix C

Table of Data Exploration Results

	Sex	Identity	Align	Alive
<b>DC</b>	28% Female / .2% Genderless / 69% Male / .01% Transgender	35% Public / 34% Secret / 31% Unknown	41% Bad / 41% Good / 8% Neutral / 10% Reformed Criminals	24% Deceased / 76% Alive
<b>Marvel</b>	.2 Agender / 23% Female /	.09% Known to Authorities /	41% Bad / /	22% Deceased / 78% Alive

	.01 Genderfluid / 71% Male	10% No Dual Identity / 27% Public / 38% Secret	28% Good / 13% Neutral	
--	-------------------------------	---	---------------------------------	--