

Are You the Main Character? Movie Character Personality Compatibility Matches

I. Dataset

Raw data: <https://www.kaggle.com/datasets/subinium/movie-character-mbti-dataset>

The raw dataset is a csv web scraped from personality-database.com, posted by a Kaggle user 2 years ago. Personality Database is a fan wiki/database in which **users vote and guess** their favorite fictional characters' personality types using famous tests like the [MBTI](#) or [enneagram](#). These types are not canonical by the creators of these fictional characters. This dataset only contains movie characters.

CSV files after data cleaning (for the cleaning process, see **preproc.ipynb**)

- **highlighted columns** indicate columns that we made

"mbti_clean.csv"

- NaN/unknown were dropped from "mbti" and enneagram". The original dataset was about 17,000 rows and the resulting is almost 12,000 rows.

Columns:

- "mbti": string, mbti type, in all caps e.g. "INFP"
 - **Insight:** we can make a scoring system based on these types
 - see "Scoring System" below
- "enneagram": string, enneagram type. e.g. "9w1". Enneagram types have a "wing" which indicates an adjacent personality. "9w1" means 9 is the primary type but has type 1 characteristics.
 - **Insight:** this can be used to diversify the scoring system.
 - see "Scoring System" below
- "role": string, movie character
- "movie": string, movie or movie franchise
- "img_url": string, url to image of the "role"
- **enn_1**: int, enneagram type
 - **Insight:** some people don't know their wings, and this would be easier to implement
- **enn_2**: int, enneagram wing type
 - **Insight:** in case we have time to implement this subscore
- **downthemall**: string, "img_url" appended with ``
 - we used <https://www.downthemall.net/> per Prof. Rzeszotarski's suggestion to avoid hotlinking. This column was used to generate links from the column to use this extension.
- **img_filepath**: string, local path to images downloaded from downthemall
 - **Insight:** Distributions are nice to see but we want people to see specific people they are compatible with.

Scoring System

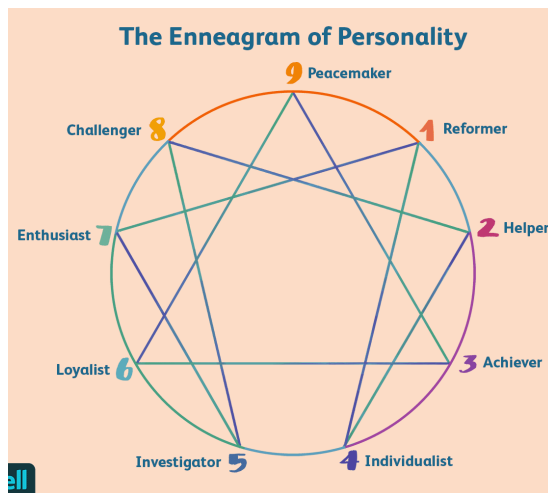
When the user inputs their personality scores, we want them to see who they are similar to and where they stand in the dataset (“main character” quality). To do this, we implemented a scoring system. Every MBTI type has 4 letters which are all binary. When the user inputs their MBTI, we can assign a “compatibility score” for each of the characters in the dataset. The enneagram is sorted like a network. Some types are related to other types and we can also score based on the same or similar types.

Example:

1. user selects “INFP”

Character A is also INFP. A gets a score of 4 because every letter matches. Character B is ESFP. B gets a score of 2 because 2 letters match. Character C is an ESTJ. C gets a score of 0 because 0 letters match.

2. Then, the user selects enneagram “9”



Character A is also a 9. A gets 4 more points (to make the MBTI and enneagram equal weight) for a total and max of 8 points.

$8/8 = 100\%$ compatibility

Character B is a 3, which is related to 9. Character B gets 2 points for a total of 4 points.

$4/8 = 50\%$ compatibility

Character C is a 7, which is not the 9 or the related 3 and 6 gets 0 more points.
 $0/8 = 0\%$ compatibility

“mbti_counts.csv”

- This csv is a simple group by and count by MBTI from the mbti_clean.csv.

Columns:

- **mbti**: string, MBTI type, caps only
- **count**: int, count of all characters with that MBTI

II. Original Storyboard

The interactions that we originally designed for this dataset were filters made through a selector for an MBTI character, a selector for an enneagram value, and a button for the user to select whether they wanted a character similar to or different from them. The MBTI and enneagram selectors would generate a histogram that held the different counts of characters that had 4 letters of the MBTI in common with the user, 3 letters in common with the user, 2 letters in common with the user, 1 letter in common with the user, and 0 letters in common with the user. It would also generate a few random character cards for characters in each level of commonality.

We decided to use filtering as an interaction because we thought it would be most interesting to a user to see which characters are similar/different from them. We chose to use button-type selectors because there are not that many values for MBTI and enneagrams. For the MBTI selection, the user would have to click through the characters in order to find the one they want, but because there aren't that many values, we thought this type of selection would be appropriate. Furthermore, for the enneagram selection, we wanted to mimic the common visualization of the enneagram in which the values lie in a circle with edges between the ones that are connected based on personality. We thought this interaction was appropriate because again, there are not that many values to choose from and they can be displayed at once.

We included the character cards because we also wanted to include a type of 'details-on-demand' interaction. The cards would provide more information on specific characters.

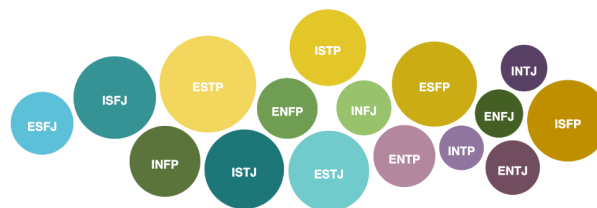
III. Final Interactive Visualization

Are You the Main Character?
Find Your Movie Character Personality Compatibility Matches

[Personality Database](#) is a popular wiki website in which fans vote and guess their favorite fictional characters' personality types. This visualization uses a [dataset](#) that contains almost 12,000 movie characters and their voted MBTI and Enneagram types, collected in 2021 by a Kaggle user. Selecting your MBTI and enneagram scores to discover how many movie characters are similar to you and who some of them are!

► *How are my compatibilites calculated?*

Overall Distribution of MBTI for Movie Characters

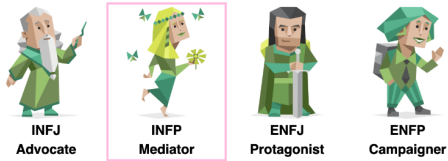


What's your MBTI?

If you don't know your Myers-Briggs Type Indicator, take the test [here](#)

image sources: <https://www.16personalities.com/>

DIPLOMATS

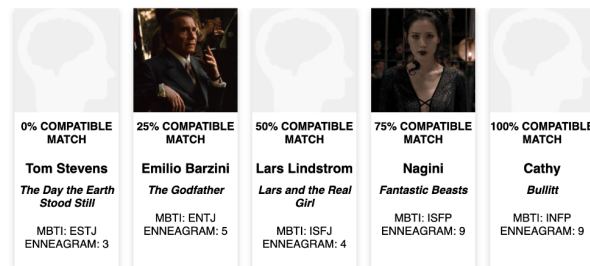
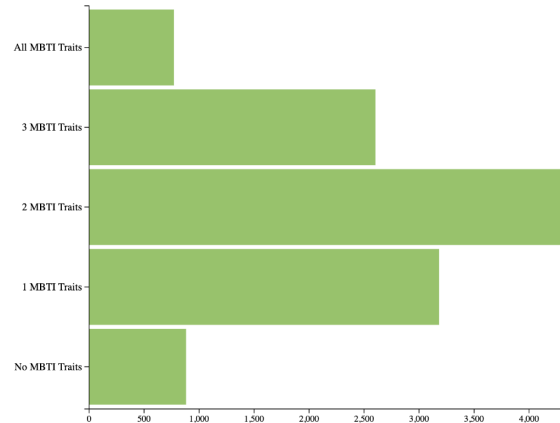


Optional: What's your Enneagram?

If you don't know your enneagram, take the test [here](#)



Your Movie Character Matches!



A. Bubble Chart

- **Marks:** Circles representing each MBTI type
- **Channels:** Size representing a count of each MBTI type, hue representing MBTI group, saturation indicating MBTI within the group
- This bubble chart is an overview of the distribution of the 16 MBTI types in the dataset.

B. MBTI Selector

- The 16 MBTI types are categorized into 4 groups: Diplomats, Consuls, Adventurers, and Analysts and are officially colored by these groups. We wanted a carousel/ slide show selector between these 4 groups to indicate their relationship and also save space. We also used the official pictures from <https://www.16personalities.com/personality-types> as button selectors.

C. Enneagram Selector

- The enneagram is often depicted in a circle with lines connecting their relationships. We recreated this network as the selector so that users unfamiliar

with the enneagram would be able to know the types they are related to because it is related to the scoring system. When the user clicks on a node, the relationship links are highlighted.

D. Histogram

- **Marks:** rectangles representing counts of compatibility scores
- **Channels:** hue representing MBTI group, vertical aligned position, horizontal aligned position
- A count of all characters belonging to each compatibility score bin. The histogram dynamically updates based on the user's personality selections.

E. Cards

- Random character examples from each of the bins generated by the user's personality selections and the histogram bins. Cards contain compatibility score, MBTI, picture, and the enneagram.

IV. Development Process

A. Changes from Original Storyboard

- **Bubble plot:** We added a bubble plot to indicate the overall distribution of MBTI types in the dataset. We were inspired by Ben Shneiderman's mantra of "overview first, zoom and filter, then details-on-demand." While we had filtering (selectors) and details-on-demand (cards), we realized that users will probably be unfamiliar with the distribution of the dataset.
- **"like attracts like" or "opposites attract" button:** We realized this button was redundant because the information presented would be the same, but ordered in a different way. We do not have that many bins, so we decided not implementing this was fine.
- **Enneagram wing:** We were already having a lot of trouble implementing the enneagram selector as is and the histogram bins. While this would have diversified our scores, we didn't have time to implement this.

B. Trade-offs and Issues

- **MBTI priority over enneagram:** In this graph, you cannot generate the histogram or cards with only an enneagram, but you can with only an MBTI. We were limited by time constraints, but justified this with the MBTI being more prevalent than the enneagram, and made the enneagram optional. We decided to emphasize this focus by only creating the overview bubble chart only for MBTI. In our scoring system, we weight these scores equally for those that do care about the enneagram (4 pts max each).
- **Histogram axis labels:** We were debating on how to be informative about the

histogram bins. We originally had 100%, 75%, 50%, 25%, and 0% as labels for the bins but realized that it was not informative, and switched to All of the same MBTI traits, 3 of the same MBTI traits, and so on. However, this restricted us in describing the bins including the enneagram scores. To be inclusive of both personality scores, we decided to use “100% personality match, 87.5% personality match, and so on for 8 bins. We include a paragraph on the site about the scoring system for those who are interested in learning more.

- **Character cards:** We included this as a way to display details from each character in the dataset; however, this way of displaying the data begs the question of whether the user should be able to see all of the characters in a particular bin upon demand. We decided to stick with the character cards instead of opening the door into another interaction to see all the characters because of time constraints, but we think this way of displaying provides interesting details from the dataset. Given more time, this could be alleviated by including another interaction, such as another ‘on click’ or ‘on hover’ that would display all the characters within a particular bin. The code for the cards is also a bit buggy when the user selects a certain combination of enneagram choices in succession; we tried to fix this in a variety of ways, but ultimately did not have time to fix all the bugs.

C. Development Process/Design Choices

- **Bubble plot:** This could have been done in a histogram, but we decided to make a bubble plot because we wanted to make it visually distinct from our other histogram.
- **MBTI colors:** We used the canonical colors of the MBTI and tried to implement wherever we could, including the bar plot, the bubble plot, and the MBTI selector.
- **Mouseover/Mouseout/Click:** Because our selectors do not traditionally look like buttons or drop downs, we added hover and event feedback to indicate they were clickable. We used a gray outline for both to be consistent.
- **Static networks:** We used d3 networks to implement the enneagram selector and the bubble plot, and made them static because they are not the main focus of project.

V. Division of Labor

Estelle

- data cleaning: 1 hr
- enneagram selector: 5 hr
- styling/polish: 4 hr
- write-up: 2 hr

Eva

- mbti selector: 5 hr

- Histogram: 5 hr
- Enneagram selector: 2 hr
- data cleaning: 1 hr

Gaby:

- scoring system function: 3
- random card generation: 5
- styling: 3

VI. Citations

INFO 3300 Lecture code by Prof. Rzeszotarski

- histogram: 9/26 d3 interactive bar chart
- enneagram selector, bubble chart: 10/26 d3 network layouts
- enneagram selector: 11/28 d3 chord diagram
- hover box: 10/5 d3 choropleth (adding data)

INFO 4310 Lecture code by Prof. Rzeszotarski

- .classed(): 2/13 SPLOM + Responsiveness

Images:

- downloading mass batch of images: downthemall.net
- character images: <https://www.personality-database.com/>
- MBTI images: <https://www.16personalities.com/personality-types>

Code:

- enneagram selector, transforming objects around a circle:
<https://spin.atomicobject.com/2015/06/12/objects-around-svg-circle-d3-js/>