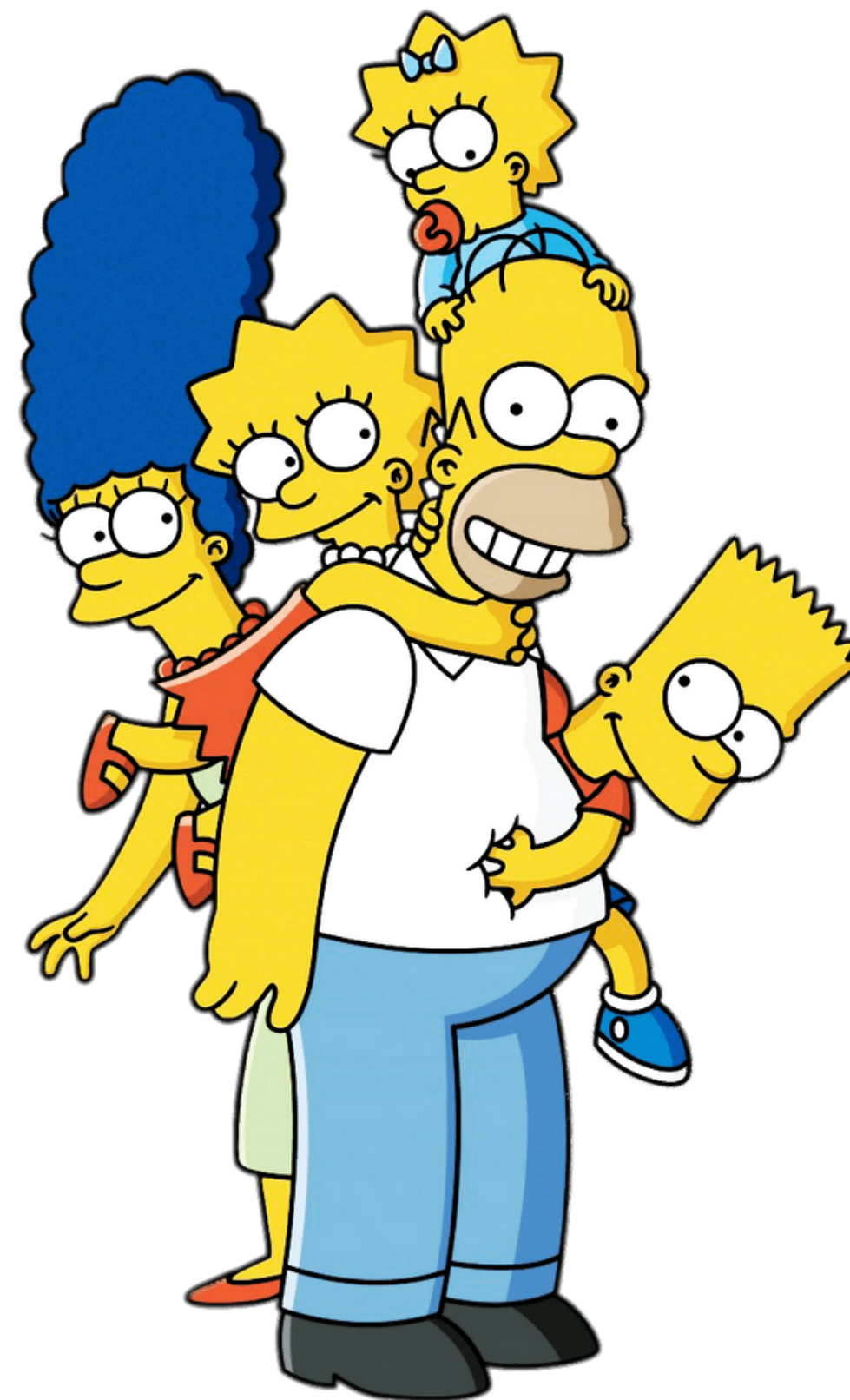


Data Management

Julia Tsymbal, Paola Cavana, Edoardo Fava





Introduction

The idea behind the project was to create a database of all the Simpsons' characters and episodes.

The goal is to answer the following questions:

- WHICH CHARACTERS MAKES THE EPISODES **MORE** ENJOYABLE?
- WHICH CHARACTERS MAKES THE EPISODES **LESS** ENJOYABLE?

Data Acquisition

Fandom Characters Scrapping

Characters

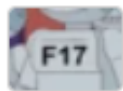
Category page

This is the main category page for all of the characters that have appeared in *The Simpsons* and its related stories (both current and former).

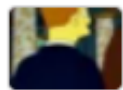
All items (5570)

• A • B • C • D • E • **F** • G • H • I • J • K • L • M • N • O • P • Q • R • S • T • U • V • W • X • Y • Z • OTHER

F



F17



Faberge Egg Seller



Faceless Man

Martin Prince, Jr.

3 | VIEW SOURCE |

"Pick me, teacher, I'm ever so smart!"

—Martin Prince^[src]

"Dickety! Highly dubious."

—Martin questioning Abe's story

Martin Prince is a recurring character in *The Simpsons*. He is a friend and classmate of **Bart Simpson**, and rivals **Lisa**'s intelligence. He is **Nelson Muntz**'s favorite target for bullying and is an academically brilliant teacher's pet and is portrayed as a stereotypical nerd. Martin is a fourth-grade student at **Springfield Elementary School**. He is usually shown on friendly terms with Bart due to the fact that he is occasionally seen with him, along with Nelson, Milhouse, and other kids.

Contents [hide]

1. Biography
2. Appearance
3. Relationship with bullies
 - 3.1. Nelson Muntz
 - 3.2. Jimbo Jones, Dolph Starbeam & Kearney Zyzwicz
4. Non-Canon Appearances
 - 4.1. Future
 - 4.2. The Simpsons Game
 - 4.3. Treehouse of Horror series

Martin Prince



Gender/Sex **Male**

Status **Alive**

AKA Martin Priss
Marcia Princess
Fartin Prince (by **Bart Simpson**)

Fandom Episodes Scrapping

List of Episodes

This is a **list of *The Simpsons* episodes** by broadcast order. The series has so far broadcasted 740 episodes.

For a list of episodes in production order, see *List of episodes by production order*.

☰

Contents

[hide]

1. Season 1 (1989/90)

2. Season 2 (1990/91)

3. Season 3 (1991/92)

4. Season 4 (1992/93)

5. Season 5 (1993/94)

6. Season 6 (1994/95)

7. Season 7 (1995/96)

8. Season 8 (1996/97)

9. Season 9 (1997/98)

10. Season 10 (1998/99)

11. Season 11 (1999/2000)

12. Season 12 (2000/01)



13. Season 13 (2001/02)

14. Season 14 (2002/03)



15. Season 15 (2003/04)

16. Season 16 (2004/05)


17. Season 17 (2005/06)


[VIEW SOURCE](#)



Life on the Fast Lane


[VIEW SOURCE](#)


[Episode](#)
[References](#)
[Gags](#)
[Appearances](#)
[Gallery](#)
[Quotes](#)
[Credits](#)


The Telltale Head

Life on the Fast Lane

Homer's Night Out


"I can teach you everything."
—Jacques

"Life on the Fast Lane" is the ninth episode of **Season 1**. It was planned to be the eleventh episode of the season. The episode marks the debuts of **Lenny Leonard**, **Jacques** and **Helen Lovejoy**. This also marks the first episode to win a **Primetime Emmy Award for Outstanding Animated Program**.

- ☰ **Contents** [hide]
- 1. Synopsis
- 2. Story
- 3. Transcript
- 4. Citations



IMDb Episodes Scrapping



I Simpson (1989–)
Episode List

Season: OR Year:

Season 34



S34, Ep1

Habeas Tortoise25 Sep. 2022

★ 6.5 (494) ☆ Rate

After humiliating himself at a town meeting, Homer looks for acceptance in an internet group on the hunt for a missing tortoise and finds something much more sinister.



S34, Ep2

One Angry Lisa2 Oct. 2022

★ 6.3 (395) ☆ Rate

Lisa gets called for jury duty, while Marge becomes obsessed with her exercise bike.

S21.E1 < Tutti gli episodi >

Il Cast e la Troupe · Recensioni degli utenti · Quiz · IMDbPro · Tutti gli argomenti

Homer the Whopper

L'episodio è andato in onda il 20 set 2010 · TV-14 · 30min

+ Animazione Commedia

+ Aggiungi una trama nella tua lingua

A Hollywood studio makes a movie based on Comic Book Guy's comic book character "Everyman," and Homer is cast in the lead. However, Homer's weight fluctuations cause problems for the production.

Regia Mike B. Anderson (supervising director) · Lance Kramer

Sceneggiatura Matt Groening · James L. Brooks · Sam Simon >

Star Dan Castellaneta (voce) · Julie Kavner (voce) · Nancy Cartwright (voce) >

IMDbPro Vedi produzione, biglietteria e informazioni sull'azienda

VALUTAZIONE IMDb 7,1/10 1521

LA TUA VALUTAZIONE ☆ Valuta

+ Aggiungi alla lista dei preferiti

3 Recensioni degli utenti

Data Cleaning, Data Enrichment & Data Integration

- We **normalized** some attributes, such as the characters name
- **Conversion** from text format to numeric format
- **Cancellation** of future episodes as these have missing values
- We guaranteed **consistency**



Fandom Episode
url
title
episode_number_absolute
airdate
main_characters
season
production_code
written_by
directed_by

+

IMDb Episode
url
title
episode_number_absolute
episode_number_relative
airdate
season
rating
reviews_amount

=

Episode
episode_number_absolute
episode_number_relative
season
title
rating
reviews_amount
fandom_url
imdb_url

Data Quality

Accuracy

In our case, it is currently impossible to establish a source of truth for the collected data. We do not have reference values or validated sources where we can get extra data for comparison.

Fandom and IMDb are not a hundred percent guarantor of the information presented on the site.

Completeness

- Fandom Characters: the percentage of completeness of this dataset is 56.2%
- Fandom Episodes: the percentage of completeness of this dataset is 93.7%
- IMDb Episodes: the percentage of completeness of this dataset is 100%

Currency

Currency measures how quickly data is updated, so, in our case the series "The Simpsons" is in continuous production and that's why the currency of our database will decrease every week.

- To try to ensure currency data should be collected at least weekly from the first scraping.

Consistency

We check the consistency of the data formats that we integrate into one dataset.

Fandom Columns	Fandom DType	IMDb DType	IMDb Columns
url	string	string	imdb_url
title	string	string	title
season	float64	int64	season
episode_number_absolute	int64	int64	episode_number_absolute
airedate	date	date	airdate
main_characters	string	-	-
image_url	string	-	-
production_code	string	-	-
-	-	int64	rating
-	-	int64	episode_number_relative
-	-	int64	reviews_amount
written_by	string	-	-
directed_by	string	-	-

Data Model

Why use a Relational Database?

- It is the best option for managing **many-to-many** relations
- It is easy to execute **complex queries** using SQL
- Helps in maintaining data **integrity** and data **accuracy**
- Reduces data **redundancy** to minimum or zero
- It is very **well known**



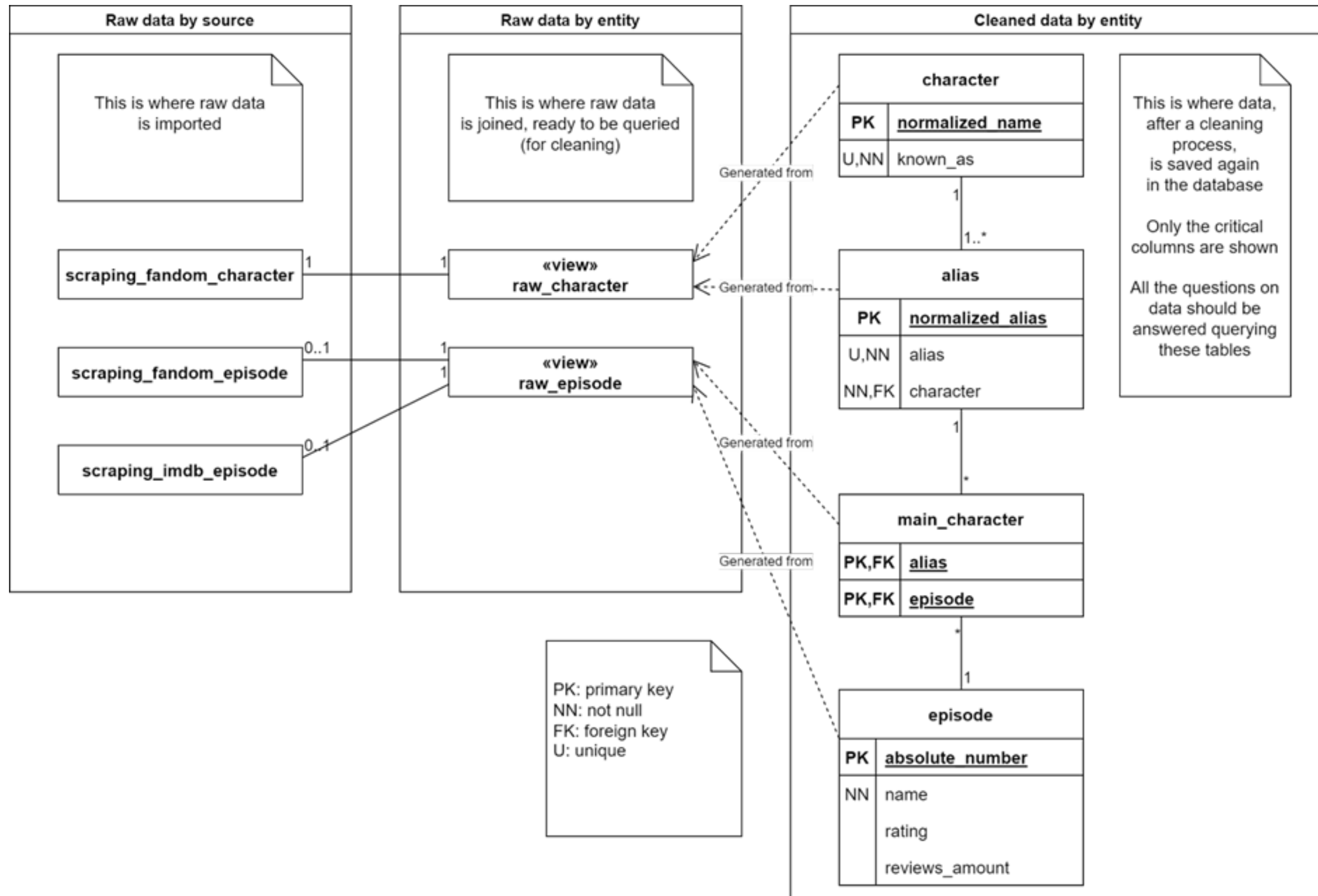
Why use a Relational Database?

- *Key-Value Stores*: not suitable for relations
- *Column Family Stores*: not suitable for relations
- *Graph databases*: we don't need rich relations
- *RDF databases*: we don't need such expressive power



We chose PostgreSQL because
it's a **powerful, open source** object-relational database system
with a strong reputation for **reliability**, feature **robustness**, and **performance**

Data Storage



- CharacterTable

	normalized_name [PK] character varying (128)	known_as character varying (128)	fandom_url character varying (256)
50	agnes skinner	Agnes Skinner	https://simpsons.fandom.com/wiki/Agnes_Skinner
51	a hole in the ground gang	A Hole in the Ground Gang	https://simpsons.fandom.com/wiki/A_Hole_in_the_Ground_Gang
52	ahura mazda	Ahura Mazda	https://simpsons.fandom.com/wiki/Ahura_Mazda
53	aiden freemont	Aiden Freemont	https://simpsons.fandom.com/wiki/Aiden_Freemont
54	aileen	Aileen	https://simpsons.fandom.com/wiki/Aileen
55	air force guard	Air Force Guard	https://simpsons.fandom.com/wiki/Air_Force_Guard

- Alias Table

	normalized_alias [PK] character varying (128)	alias character varying (128)	character character varying (128)
1	10 year old voiced man	10 Year Old Voiced Man	10 year old voiced man
2	3 little pigs	3 Little Pigs	3 little pigs
3	4h leader	4-H leader	4h leader
4	50 cent character	50 Cent (character)	50 cent character
5	98th airborne veteran	98th Airborne veteran	98th airborne veteran
6	a27	A27	a27

- Main Character Table

	episode [PK] smallint		alias [PK] character varying (128)
1		2	bart simpson
2		3	homer simpson
3		3	mr burns
4		4	homer simpson
5		5	bart simpson
6		5	grampa simpson

- Episode Table

	episode_number_absolute [PK] smallint	episode_number_relative smallint	season smallint	title character varying (128)	rating smallint	reviews_amount integer	fandom_url character varying (256)	imdb_url character varying (256)
127	127	24	6	Lemon of Troy	87	3300	https://simpso...	https://www.imdb.com/titl...
128	128	25	6	Who Shot Mr. Burns? (Part One)	92	4200	https://simpso...	https://www.imdb.com/titl...
129	129	1	7	Who Shot Mr. Burns? (Part Two)	90	3800	https://simpso...	https://www.imdb.com/titl...
130	130	2	7	Radioactive Man	83	2900	https://simpso...	https://www.imdb.com/titl...
131	131	3	7	Home Sweet Homediddly-Dum-Doo...	84	2800	https://simpso...	https://www.imdb.com/titl...
132	132	4	7	Bart Sells His Soul	88	3800	https://simpso...	https://www.imdb.com/titl...

Data Exploration

First Query

```
SELECT c.normalized_name, c.know_as, c.fandom_url
COUNT (e.episode_number_absolute) AS number_of_episodes,
AVG (e.rating) AS average_episode_rating,
COUNT (e.episode_number_absolute)/2 + AVG(e.rating) AS popularity_score
FROM "character" AS c
    INNER JOIN "alias" AS a ON c.normalized_name = a."character"
    INNER JOIN main_character AS m ON a.normalized_alias = m."alias"
    INNER JOIN episode AS e ON m.episode = e.episode_number_absolute
GROUP BY c.normalized_name
HAVING COUNT COUNT(e.episode_number_absolute)>1
ORDER BY COUNT COUNT(e.episode_number_absolute)/2 * AVG(e.rating) DESC
LIMIT 10;
```

Which character makes the episodes more enjoyable?

1



2



3



Second Query

```
SELECT c.normalized_name, c.know_as, c.fandom_url
COUNT (e.episode_number_absolute) AS number_of_episodes,
AVG (e.rating) AS average_episode_rating,
COUNT (e.episode_number_absolute)/2 * AVG(e.rating) AS popularity_score
FROM "character" AS c
    INNER JOIN "alias" AS a ON c.normalized_name = a."character"
    INNER JOIN main_character AS m ON a.normalized_alias = m."alias"
    INNER JOIN episode AS e ON m.episode = e.episode_number_absolute
GROUP BY c.normalized_name
HAVING COUNT COUNT(e.episode_number_absolute)>10
ORDER BY COUNT COUNT(e.episode_number_absolute)/2 * AVG(e.rating) ASC
LIMIT 5
```

Which character makes the
episodes less enjoyable?

1



2



3



Third Query

```
SELECT e.episode_number_absolute,e.season, e.title, e.rating  
FROM episode AS e  
ORDER BY e.rating DESC  
LIMIT 10
```


What are the best episodes?

1



"Homer's Enemy"
S8.E23

2

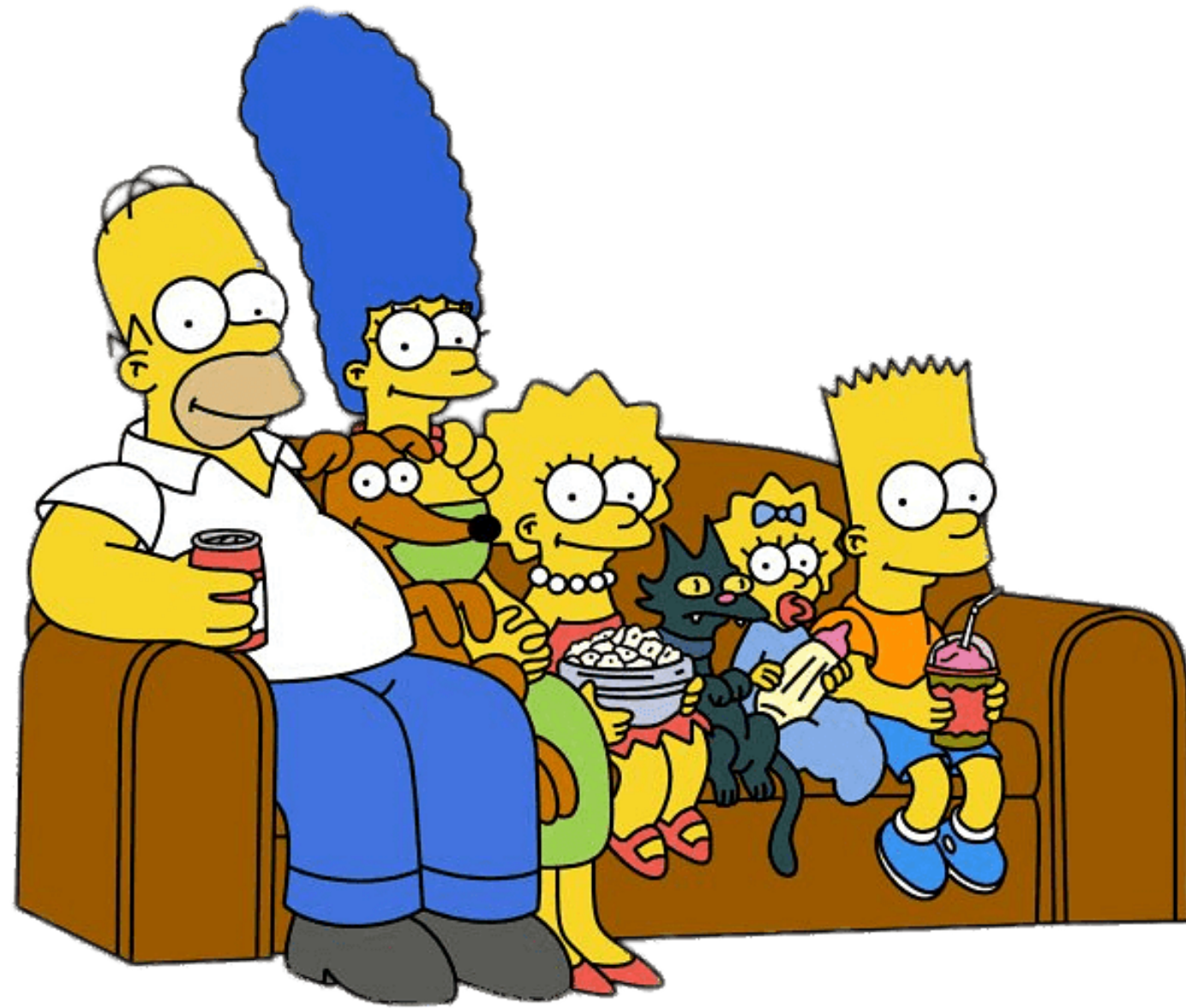


"Who Shot Mr. Burns?"
(Part One)
S6.E25

3



"Treehouse of Horror V"
S6.E6



The End.

**Thanks for your
attention!**