

# Analyzing *Jeopardy!* Data

M. Ball, M. Farrow, J. Harrison

**Abstract**—For this project, the team’s primary goal is to create a database using *Jeopardy!* data and visualize insights. For the team to proceed, the team used R to crawl the website, J! Archive [1] and extract the data. Using MySQL Workbench, a schema was created and the database was populated. Once complete, the team will be able to connect the database with an RShiny application.

**Index Terms**—database, relational database, text analysis, text mining

## I. INTRODUCTION

**J**EOPARDY! is a long running quiz-style television game show where contestants are presented clues in the form of answers from a series of categories and must phrase their responses in the form of a question. Each game is composed of three rounds: Jeopardy, Double Jeopardy, and Final Jeopardy. The first two rounds are made up of six categories with five answers of increasing difficulty and monetary value. If a contestant provides the correct question, they receive the amount that question is worth, otherwise they lose that amount.

Although the original version of *Jeopardy!* premiered in 1964, the show’s revival in 1984 with host Alex Trebek led to a run of more than 8,000 games over 37 years [2].

Using data from 6,775 of those games, we wanted to analyze using the data provided on J! Archive to examine what has made this show such a game show staple. We examined trends across game categories, player appearances, and Daily Doubles.

## II. EXISTING RESEARCH

Jeopardy! is a game show that lends itself to trends and statistical studies, and trivia’s broad appeal gives the game an accessibility to players and viewers alike. Sites such as The Jeopardy! Fan [3] and the Jeopardy! History Wiki [4] have provided facts and statistics from the game, with data often coming from J! Archive. Even sites like FiveThirtyEight have turned their analytical prowess on the game, analyzing two of the show’s most notable players, Ken Jennings and James Holzhauer [5].

The J!-Archive has enabled all of this work, but one of the site’s limitations is its ability to interact with the data. Our intention with this project was to take the data from J!-Archive and create a database that could be used by researchers to analyze the data for themselves.

This paper was submitted as the final term project for SMU’s MSDS 7330: File Organization and Database Management course under the supervision of Dr. Sohail Rafiqi.

Megan Ball is with \_\_\_\_\_Austin, TX USA (e-mail: ballm@smu.edu).

## III. DATA

Our approach to this project involved crawling the J!-Archive, balancing the wealth of available data with the goals of this project. Due to the connected nature of the data, we made the decision to build our project in a relational database using MySQL. Initial data collection and subsequent analysis took place using the R programming language. All source data files and analysis queries are located on GitHub [6].

### A. Data Collection

The `whatr` [7] package was used for the majority of the data collection for this project. Using the package’s included functions, we were able to crawl the J! Archive and extract the air date, board details, Daily Double information, final scores, player information, and game synopsis for almost 7,000 games spanning more than three decades. Once the data had been crawled, a function with a `for` loop was used to iterate over each game and bind the rows of each piece of information into a single data set.

### B. Data Models

Once we understood which variables were pertinent to the project’s overall goal, we were able to create a normalized schema which is represented by the enhanced entity-relationship diagram (Figure 1). There are three many-to-many relationships within the schema: one between players and episodes, another between Daily Doubles and their scores, and finally between the game synopsis and players. Therefore, the team had to create specialized tables that could represent these many-to-many relationships: *players\_has\_episode*, *doubles\_has\_scores*, and *synopsis\_has\_players*. The other relationships are one-to-many/many-to-one.

Based on the schema in the enhanced entity-relationship diagram, the team determined the database was normalized. The data was manipulated in R to satisfy the schema that had been design in MySQL Workbench, and was then uploaded into the database.

Matt Farrow is with Southern Methodist University, Dallas, TX 75025 USA. (e-mail: mfarrow@smu.edu).

Jake Harrison is with \_\_\_\_\_ (e-mail: harrisonjp@mail.smu.edu)



A word cloud of Daily Double answers seems to bear that hypothesis out, as seen in Figure 5.

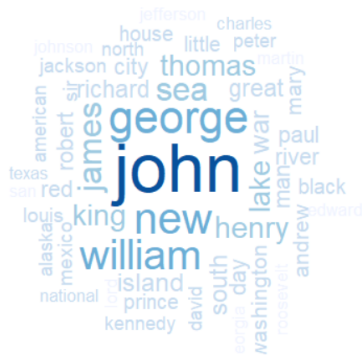


Figure 5: Daily Double answers

With 75 game appearances, Ken Jennings is not only one of the most well-known players, he also got the most Daily Double clues during his time on the show.

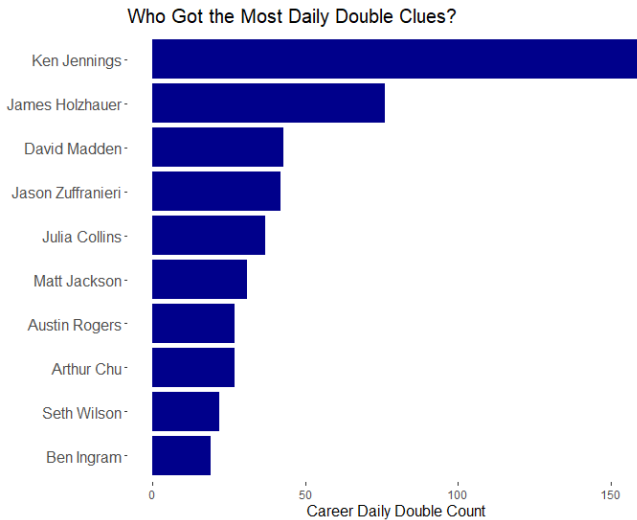


Figure 6: Plot of career Daily Double counts for the top ten players.

### C. Player Statistics

Following Ken Jennings' historic run

### D. Notable Players

TABLE I  
NOTABLE PLAYER STATISTICS\*

Player	Highest Score	Cumulative Correct Answers	Cumulative Incorrect Answers	Total Number of Games
James Holzhauer	\$131,127	1154	35	33
Ken Jennings	\$75,000	2643	240	75
Jason Zuffranieri	\$58,400	565	34	29
Julia Collins	\$35,000	504	42	21
David Madden	\$34,200	470	36	20
Matt Jackson	\$51,000	389	14	14
Austin Rogers	\$69,000	322	42	13
Arthur Chu	\$58,200	309	42	12
Seth Wilson	\$31,200	307	30	13
Jason Keller	\$36,900	246	24	10

\*all statistics exclude any tournament or special game series

### V. RSHINY APP

Should I insert a time trend of any notable info for specific players?

Table of worst games?

### VI. REFERENCES

- [1] "J! Archive," [Online]. Available: <https://j-archive.com>. [Accessed 22 03 2021].
- [2] K. Q. Seelye, "The New York Times," 08 11 2020. [Online]. Available: <https://www.nytimes.com/2020/11/08/arts/television/alex-trebek-dead.html>. [Accessed 22 03 2021].
- [3] "The Jeopardy! Fan," [Online]. Available: <https://thejeopardyfan.com>. [Accessed 22 03 2021].
- [4] "Jeopardy! History Wiki," [Online]. Available: [https://jeopardyhistory.fandom.com/wiki/Jeopardy!\\_Statistics](https://jeopardyhistory.fandom.com/wiki/Jeopardy!_Statistics). [Accessed 22 03 2021].
- [5] O. Roeder, "FiveThirtyEight," 30 04 2019. [Online]. Available: <https://fivethirtyeight.com/features/the-battle-for-jeopardy-supremacy/>. [Accessed 22 03 2021].
- [6] M. Farrow, "GitHub," [Online]. Available: <https://github.com/mattfarrow1/7330-term-project>.

- [7] K. Nicholls, "whatr," [Online]. Available: <https://kiernann.com/whatr/>. [Accessed 22 03 2021].
- [8] R. Kalland, "'Jeopardy!' Champ James Holzhauer's Daily Double Dominance Is The Evolution Of Recent Strategy," 25 04 2019. [Online]. Available: <https://uproxx.com/sports/jeopardy-james-holzhauer-april-25-daily-doubles-strategy-historic-wagers/>. [Accessed 22 03 2021].
- [9] C. Jacobs, "The Federalist," 15 08 2020. [Online]. Available: <https://thefederalist.com/2020/08/15/how-alex-trebeks-memoir-explains-the-enduring-success-of-jeopardy/>. [Accessed 22 03 2021].
- [10] C. McNear and K. Jennings, Answers in the Form of Questions, New York: Twelve, 2020.