



Analysis and Machine Learning

The Seinfeld

- Created by Larry David & Jerry Seinfeld
- 174 Episodes
- 9 Seasons
- 1989-1998
- Jerry, George, Elaine, Kramer
- **“The Elaine Dance”** ----->



Seinfeld is one of the most popular TV Shows of all time.

It consists of 174 episodes over 9 seasons spanning 1989-1998

It catalogues the life and exploits of comedian Jerry Seinfeld (played by himself) his best friend George, his friend and ex-girlfriend Elaine and his neighbor Kramer.

The Data Cleaning

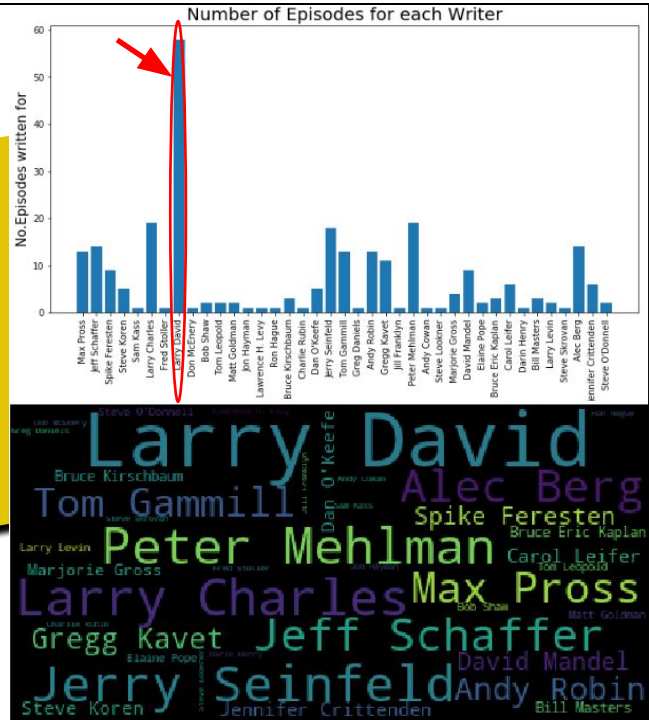
- Kessler -> Kramer
- S01E00 -> 'Pilot'
- The “missing” episodes



Both the original pilot and the first 'regular episode were designated S01E01. The original pilot for Seinfeld has been relabelled to episode zero to avoid confusion. 'Kessler' has been corrected to Kramer
Missing clip episodes

The Writers

- Larry David
 - 58 episodes
- Highest contributing writer

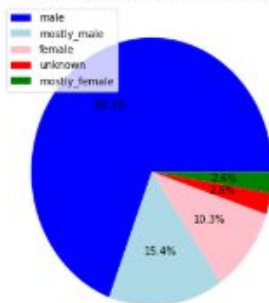


Numerous writers across the show but Larry David's contribution is singularly monumental
 He is by far the highest contributing writer with 58 writing credits

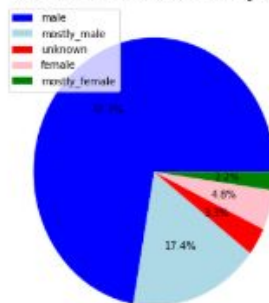
The Gender Disparity

- Heavily male writers room
- <10% of episodes have a female writer
- 3 of 9 seasons have no female writer

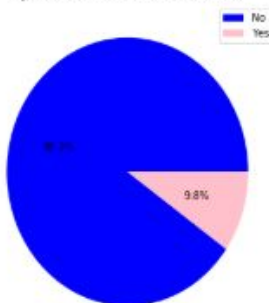
Gender estimates breakdown of writers



Gender estimates breakdown of writing credits



Episodes that have a female writer



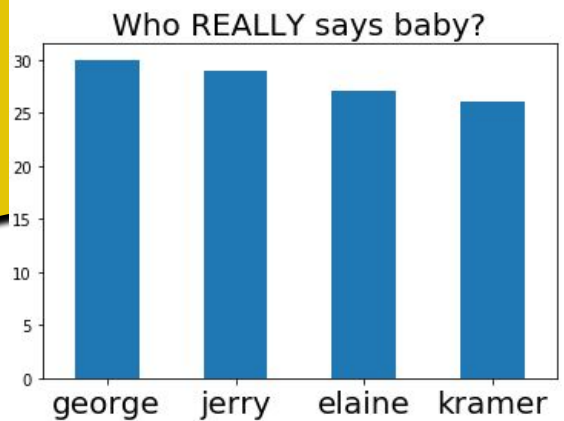
The writing credits on Seinfeld are heavily skewed towards male writers.

<10% of episodes have one or more female writer credited

Improves from the first 2 seasons where Larry David does most of the writing but no season has more than 4 episodes with a female writer

The Dialogue

- The Superman
 - 24 times, 17 episodes
- The “Babies”
 - George’s “babies”



Jerry Seinfeld is a well known Superman super fan and Seinfeld is known for its background Superman references. Superman is referenced in the dialogue 24 times across 17 episodes.

George likes to use the word baby

The Larry David Identifier

- Did Larry David write this?
- Natural Language Processing - Classifier
- 70% accuracy



Larry David contributed greatly to the Seinfeld canon but I wanted to see if he had a distinct writing style or signature that could be distinguished from episodes written by other writers

Natural Language processing is used on a Keras Deep Learning model and a model was created with 70% prediction accuracy

The Future Work

- More Data
- Cleaner Data
 - Remove simple greetings



The model performs significantly better than random chance but could be improved by the addition of more data

There is no more Seinfeld but Larry David also wrote for other TV Shows such as 'Curb Your Enthusiasm' and 'Saturday Night Live'

The removal of simple greeting from the dialogue set could improve the accuracy of the model by focusing more on distinctive language patterns

The Future Work (2)

- Other Classifiers
 - Character
 - Season



Other classification tasks could be developed such as a character identifier or Identifying by Season to see how writing and characters develop over time

The Business Recommendations

- Seinfeld: The billion dollar brand
 - Chatbot, interactive experience



Seinfeld is still broadcast in reruns and attracts over a million viewers on a nightly basis. As a brand it has earned over 3 billion dollars and counting. Seinfeld is still popular today, 30 years after its debut. Building products for that brand such as a chatbot or a Larry David's Seinfeld interactive experience

The Business Recommendations

- CEO Fraud Prevention
 - Flag imposter correspondence



More broadly this kind of writer detection can be used for fraud prevention. By training a model on business correspondence by the C-Level executives for a company imposter scams could be avoided by flagging suspicious correspondence that doesn't appear to be written by its supposed author

Seinfeld

THANK YOU!



Thank you