Stone Barrett

CSE 590: Python Data Analytics

Homework 3

Part 1

The three documents I chose are text conversions of a novel, a television series episode, and a song.

Novel: 1984 by George Orwell

TV Episode: Person of Interest - Season 3, Episode 23: "Deus Ex Machina" by Greg Plageman and David Slack

Song: Welcome to the Machine by Pink Floyd

I chose these three texts because they all either directly deal with or at least vaguely relate to themes involving artificial intelligence,

while each approaching the topic differently.

Part 2

a)

A picture containing text, orange, close

Description automatically generated

b)

Deus Ex Machina

Chart, histogram

Description automatically generated

1984

Chart, bar chart, histogram

Description automatically generated

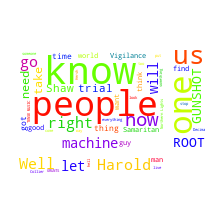
Welcome to the Machine

Chart, bar chart, histogram

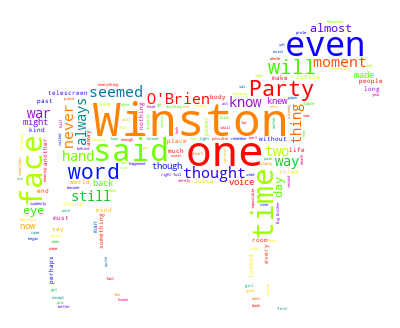
Description automatically generated

c)

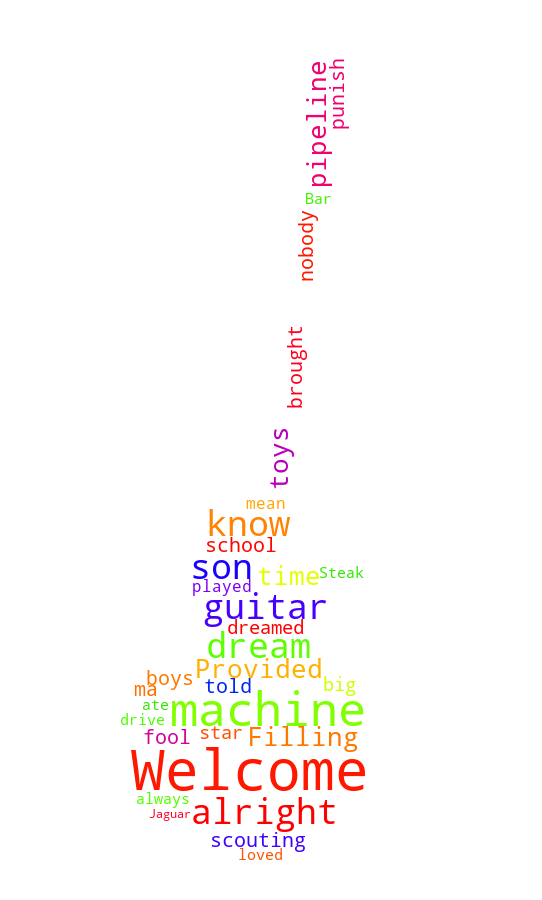
Deus Ex Machina



1984



Welcome to the Machine



d) For part b, there is a clear distinction to the counts and words picked up for Welcome to the Machine in comparison to the other two documents. This is because for some reason, I was unable to remove the special newline character from Welcome to the Machine. The other reason it is very different is because it is a song and very much shorter than that screenplay and the novel. The sheer word count being significantly lower than the other two documents coupled with the fact that songs often repeat themselves multiple times results in a very skewed word occurrence graph. As for part c, all clouds turned out well and I can tell which is which just by looking at them. The guitar is for Welcome to the Machine, the computer is for Deus Ex Machina, and the horse is for 1984.

Part 3

a)

Text

Description automatically generated

b)

I suspect Welcome to the Machine’s score is relatively higher because it is a song and uses fairly simple language to maintain a rhyme scheme and get a point across in a short amount of time. The three documents all score in the same ballpark (which is what I expected) because they are all documents written around the same time period and similar cultures.

Part 4

a)

A screenshot of a computer

Description automatically generated with medium confidence

b)

The scores are slightly higher than I expected but I did expect them to be fairly high given that all three documents were written in similar time periods and cultures, and dealing with similar subject matter.

Part 5

a)

Printout is far too large to paste here. Can be found by running Part5() in the script.

b)

Deus Ex Machina

Chart, bar chart, histogram

Description automatically generated

1984

Chart, histogram

Description automatically generated

Welcome to the Machine

Chart, bar chart

Description automatically generated

c)

Deus Ex Machina

Chart, histogram

Description automatically generated

1984

Chart, histogram

Description automatically generated

Welcome to the Machine

Chart, bar chart

Description automatically generated

d)

For the same reasons as in the previous Parts, Welcome to the Machine suffers in being analyzed in this way due to its length. I do find it interesting that the Pareto principle can start to be observed in the 1984 graph, though.