

Summary of established dataset: A collection of 20 TV scripts from the pilot episodes from popular shows.

Data Dictionary

Variable	Description	Type	Value or Characteristics
Show Name	The title of the show	String	N/A
Compound Sentiment Score	The aggregated sentiment score which summarizes sentiment of the text.	Float (numerical)	Ranges from -1 (extremely negative) to +1 (extremely positive) where Compound score > 0.05: Positive sentiment. Compound score < -0.05: Negative sentiment. Compound score between -0.05 and 0.05: Neutral sentiment
Positive Sentiment Score	The proportion of the text that expresses a positive sentiment.	Float (numerical)	Ranges from 0-1
Negative Sentiment Score	The proportion of the text that expresses a negative sentiment.	Float (numerical)	Ranges from 0-1
Neutral Sentiment Score	The proportion of the text that is neutral or lacks clear sentiment.	Float (numerical)	Ranges from 0-1
Runtime (in number of seasons)	Number of seasons produced for a show	Float (numerical)	Value of 1 or greater than 1

Dataframe of Final Data for Analysis (column names described in data dictionary above)

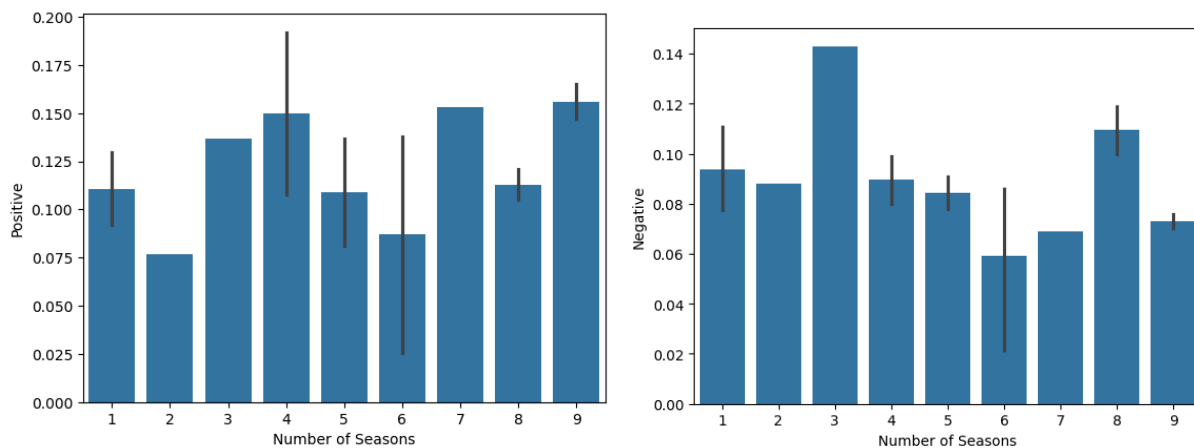
all figures below use the following dataframe

	Show_Name	Positive	Negative	Neutral	Compound	Runtime
0	Barry	0.108	0.099	0.794	0.9974	4
1	Lost	0.084	0.103	0.813	-1.0000	6
2	Mindhunter	0.077	0.088	0.835	-0.9996	2
3	Dead_to_Me	0.137	0.143	0.720	-0.9993	3
4	Breaking_Bad	0.081	0.091	0.828	-0.9995	5
5	Dexter	0.105	0.119	0.776	-0.9997	8
6	Game_of_Thrones	0.121	0.100	0.779	0.9999	8
7	Mad_Men	0.153	0.069	0.778	1.0000	7
8	Peaky_Blinders	0.102	0.086	0.812	0.9998	6
9	The_Office	0.147	0.076	0.777	0.9999	9
10	The_Sopranos	0.095	0.086	0.819	0.9986	6
11	Baby_Reindeer	0.141	0.076	0.783	1.0000	1
12	Cruel_Intentions	0.131	0.108	0.761	0.9997	1
13	Station_Eleven	0.085	0.074	0.842	0.9993	1
14	Ghosted	0.095	0.123	0.782	-0.9996	1
15	Girlboss	0.100	0.088	0.812	0.9977	1
16	Gossip_Girl	0.152	0.064	0.784	1.0000	6
17	How_I_Met_Your_Mother	0.165	0.070	0.764	1.0000	9
18	Jane_the_Virgin	0.137	0.078	0.785	1.0000	5
19	The_Good_Place	0.192	0.080	0.729	1.0000	4

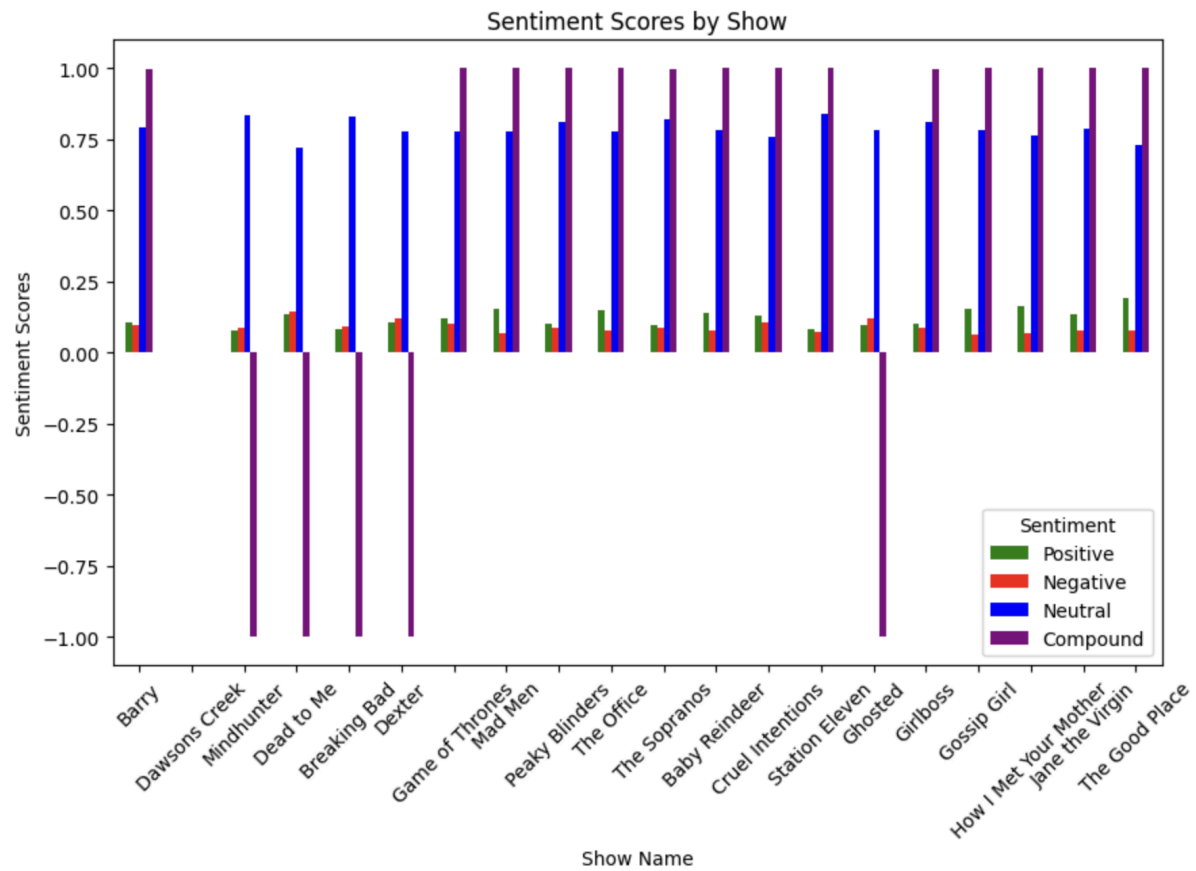
Exploratory + Analysis Plots

1. Bar plots of positive and negative sentiment versus show length (number of seasons)
2. Bar plot for individual sentiment scores by show
3. Grouped scatterplot of positive, negative and compound score
4. Scatterplot with actual vs. predicted runtime

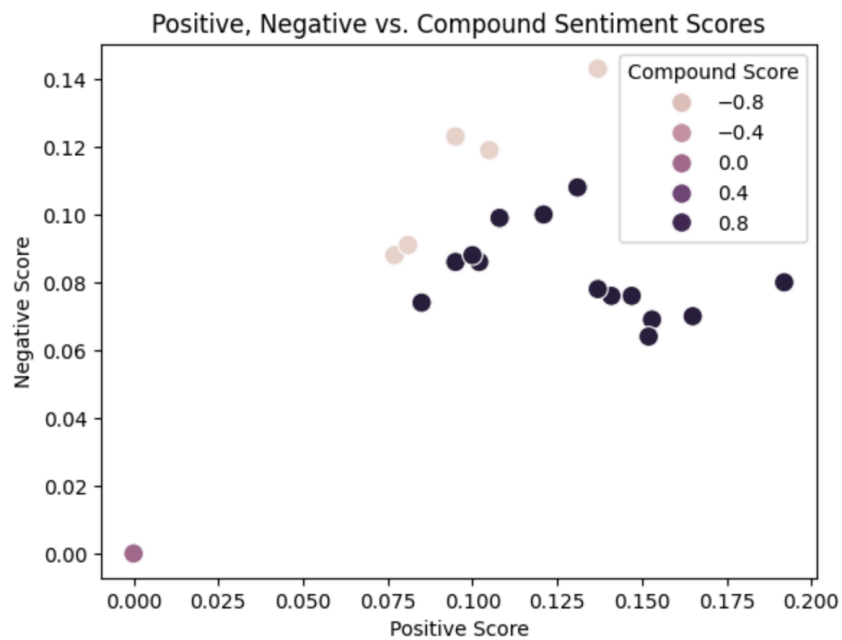
(1)



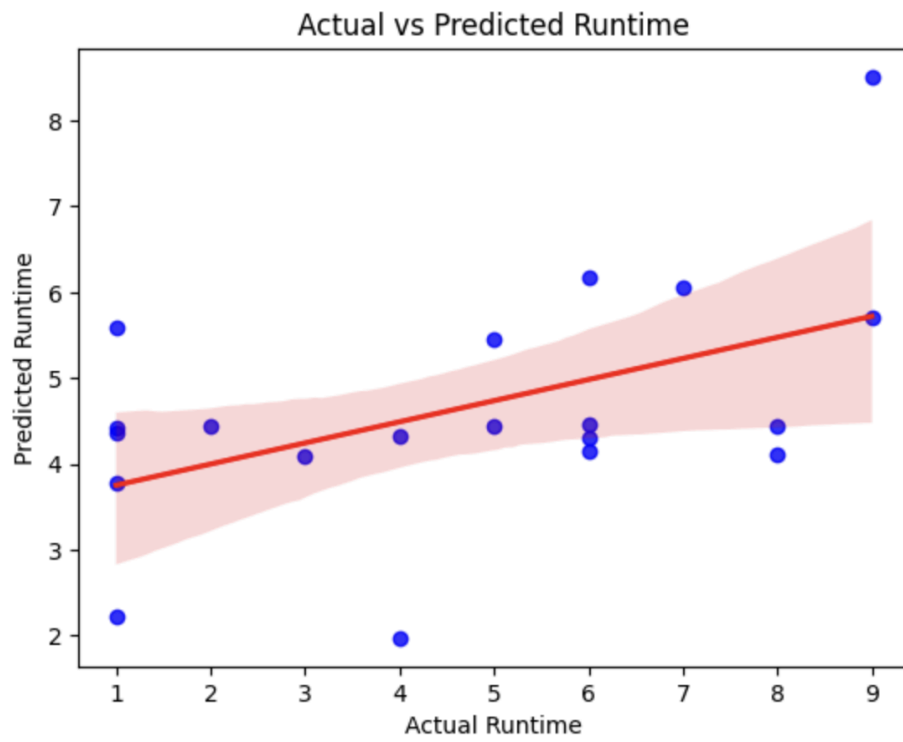
(2)



(3)



(4)



Analysis Results (Multiple Linear Regression)

OLS Regression Results

Dep. Variable:	Runtime	R-squared:	0.246			
Model:	OLS	Adj. R-squared:	0.045			
Method:	Least Squares	F-statistic:	1.224			
Date:	Thu, 13 Feb 2025	Prob (F-statistic):	0.342			
Time:	16:51:04	Log-Likelihood:	-45.739			
No. Observations:	20	AIC:	101.5			
Df Residuals:	15	BIC:	106.5			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	4.6500	0.615	7.560	0.000	3.339	5.961
Positive	-68.9601	45.842	-1.504	0.153	-166.669	28.749
Negative	-44.8240	28.927	-1.550	0.142	-106.480	16.832
Neutral	-69.6406	45.715	-1.523	0.148	-167.080	27.799
Compound	-0.2609	0.936	-0.279	0.784	-2.256	1.734
=====						
Omnibus:	0.940		Durbin-Watson:	0.940		
Prob(Omnibus):	0.625		Jarque-Bera (JB):	0.790		
Skew:	-0.188		Prob(JB):	0.674		
Kurtosis:	2.102		Cond. No.	168.		