

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Change in Themes of Billboard Top 100 Songs

Jayant Sharma, MS in Business Analytics, Oklahoma State University

John Harden, Business Management Professional, Sandia National Laboratories

USERS PROGRAM



Change in Themes of Billboard Top 100 Songs

Jayant Sharma, MS in Business Analytics, Oklahoma State University

John Harden, Business Management Professional, Sandia National Laboratories

ABSTRACT

Rapid advances in technology have empowered musicians all across the globe to share their music easily, resulting in intensified competition in the music industry. For this reason, musicians and record labels need to be aware of factors that can influence the popularity of their songs. The focus of our study is to determine how sentiments, topics and terms within song lyrics have changed over time and how these changes may have influenced the popularity of songs. Moreover, We also plan to run time series analysis on the numeric attributes of Billboard Top 100 songs in order to determine the appropriate combination of relevant attributes that influences a song's popularity. The findings of our study can potentially benefit musicians and record-labels in understanding the necessary lyrical construction, overall topics and terms, in an effort to reach the highest chart position on the Billboard Top 100. The Billboard Top 100 is an optimal source of data, as the list is an objective measure of popularity. Our data has been collected from open sources. Our dataset consists of all 334,784 Billboard Top 100 observations for the years 1955-2015, with metadata covering all 26,869 unique songs that have appeared on the chart for that period. Our lyric dataset currently contains 18,002 of those songs.

METHODOLOGY

Data Preparation and Analysis

- We collected billboard Top 100 data from the official American Top 40 Website, lyrical data was collected from open source lyric websites using web crawlers and further supplemental data was collected from the dataset of the Whitburn Project.
- Our consolidated dataset included following variables: Song Title, Artist, Week Date, This week position, Last week position, song lyrics, song lyrics with chorus, duration, beats per minutes, words per minute etc.
- We did our analysis in two steps. In the first step, we performed the text mining in order to identify period specific popular topics and terms used in the song lyrics. In the second step, we performed time series analysis on numeric properties of Billboard top 100 songs such as words per minute(WPM), beats per minute(BPM) etc.

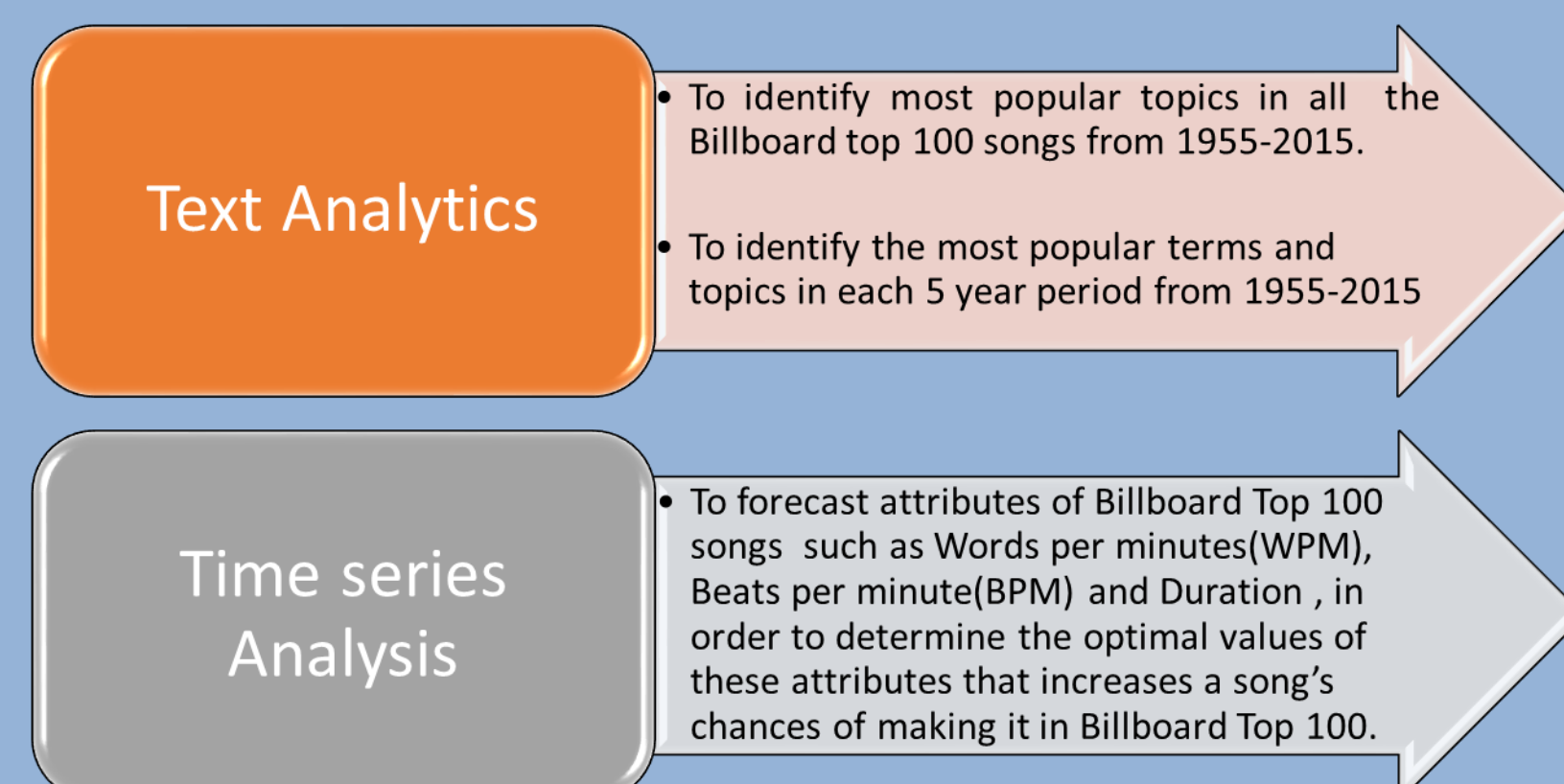


Fig. 1 Outline of Analysis method

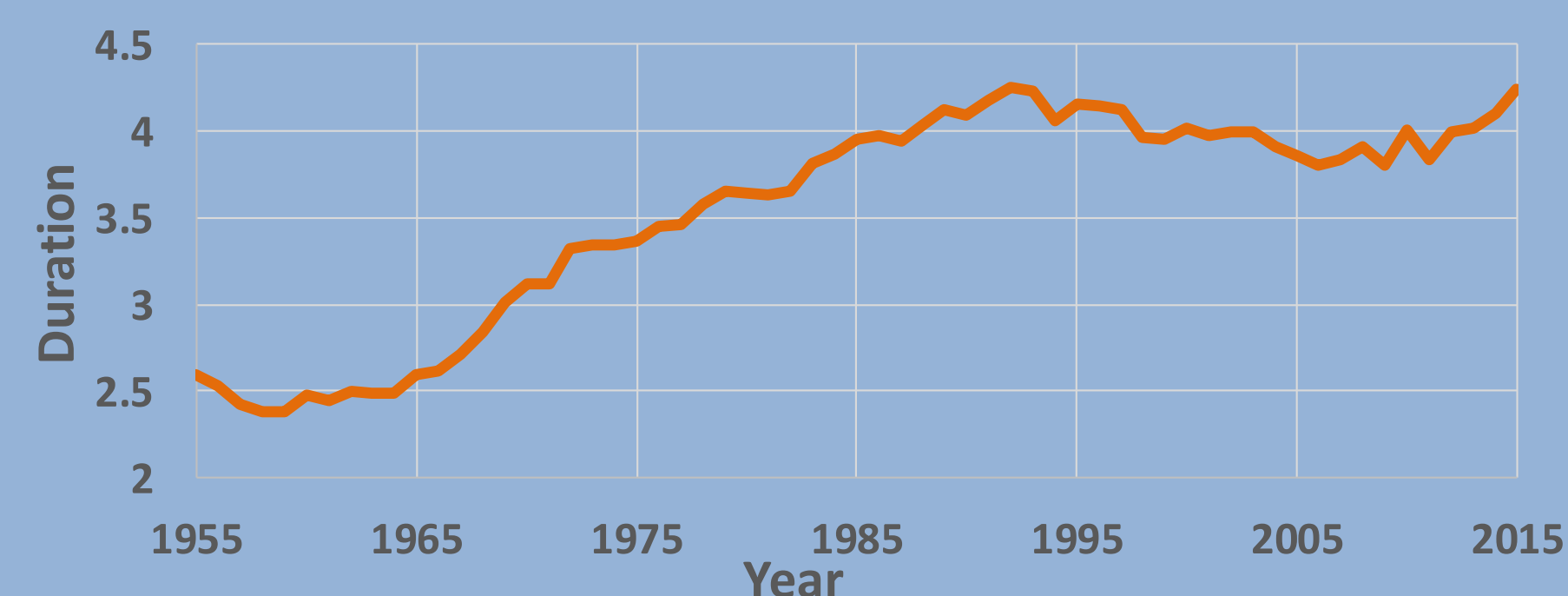


Fig. 2 Average Duration of songs through 1955-2015

METHODOLOGY

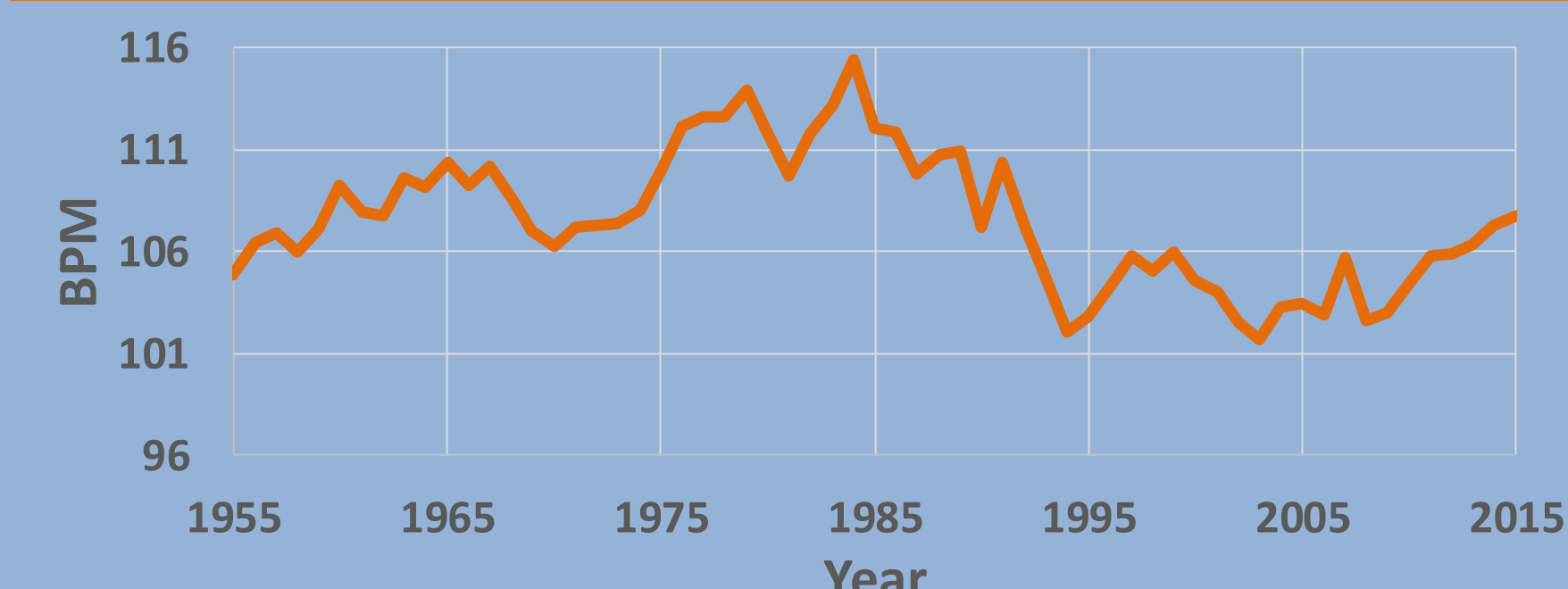


Fig. 3 Avg. BPM of songs through 1955-2015

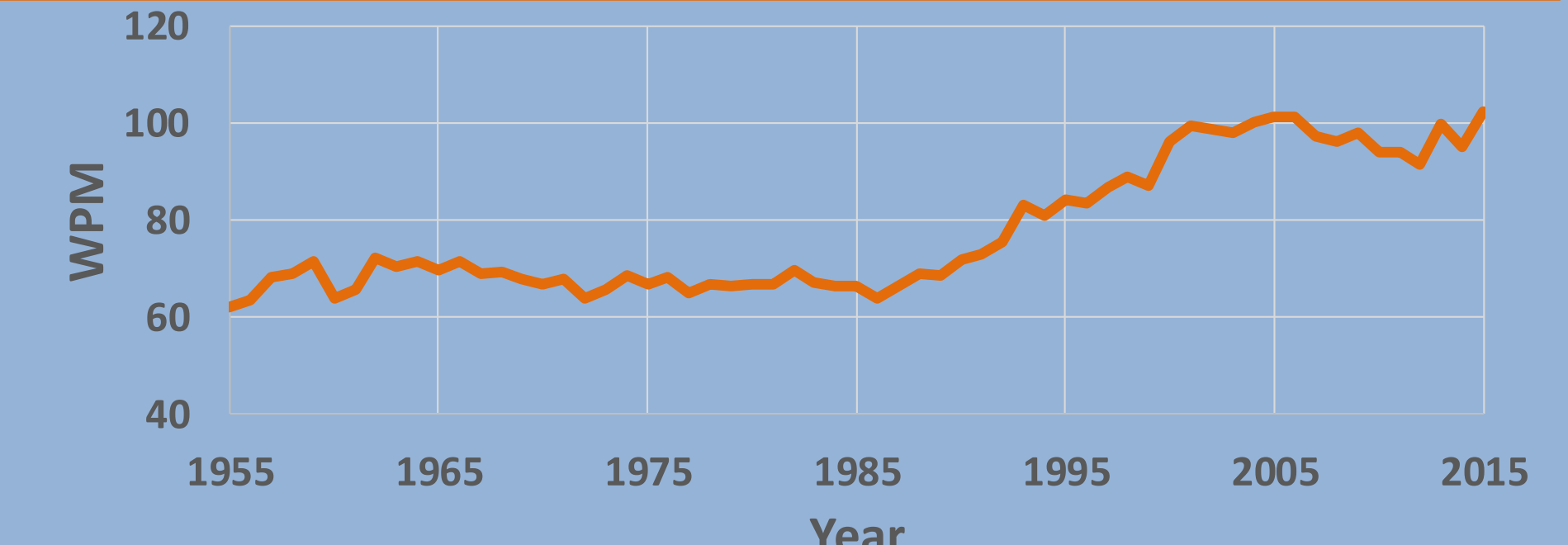


Fig. 4 Avg. Words per minute through 1955-2015

- Figure 2 gives us information on how duration of Billboard top 100 songs have varied through years. It can be seen that duration of these successful songs after 1995 has pretty much been between 3.5 minutes to 4.3 minutes. we can use this historical information to produce an optimal forecast.
- Figure 3 provide us insight about variation of beats per minute or tempo of billboard Top 100 songs with time. Clearly there are some period specific variations and we can capitalize it to produce forecasts.
- Figure 4 shows the words per minute of billboard Top 100 songs with time. It is evident that from year 1985 the number of words per minute of songs have been on a rise.
- With all the individual Univariate forecast of above properties of Billboard Top 100 songs, we can equip the upcoming artist, song writers and record labels with optimal information of duration, WPM and BPM, so that they can increase their chances of making into the billboard Top 100 by following the forecast.

Text Analytics

- We utilized SAS® Enterprise Miner 13.1 to execute the text mining process by utilizing the text parsing, text filter and text topic nodes to identify the most popular topics and terms of Billboard Top 100 songs, firstly for all songs from 1955-2015, and then for every 5 year period ranging from 1955 to 2015. We use the term cutoff values to decide the rankings of the topics and terms.

Term Cutoff ▼	Topic	Number of Terms	# Docs
0.061	+kiss,+money,+well,+home,+big	109	4637
0.061	+chorus,+da,+verse,+gotta,+repeat	69	2369
0.057	+right,+thing,+wrong,+know,+gotta	93	4859
0.056	+boy,+man,+young,+die,+god	80	4112
0.052	+ya,+keep,+stop,+gotta,+lovin	83	4034
0.052	+long,+night,+night long,+wrong,+wait	64	3012
0.051	+fire,+burn,+light,+high,+soul	70	2952
0.050	+eye,+look,+smile,+close,+face	82	4378
0.049	+down,+fall,+lie,+forever,+town	67	4049
0.049	+shake,+body,+move,+party,+come on	68	3088

Fig. 5 Top 10 Key Topics and included Terms used in Billboard Top 100 songs from 1955-2015

Change in Themes of Billboard Top 100 Songs

Jayant Sharma, MS in Business Analytics, Oklahoma State University

John Harden, Business Management Professional, Sandia National Laboratories

RESULTS

- From Figure 5, we can see the Top 10 Topics of all billboard Top 100 songs from 1955 to 2015. This means that Song Artists, record labels, writers etc. can all increase their likelihood of entering Billboard Top 100 if they use the terms and topics present in the Top 10 topics list, to construct the lyrics of songs.



Fig. 6 Top 5 Topics used in Billboard Top 100 songs for every 5 year period between 1955-2015.

- Figure 6 shows us how the Top 5 topics and terms of billboard top 100 songs have changed every 5 year starting from 1955-2015.

RESULTS

- From figure 6, we discover that there are many topics that are common in songs belonging to specific 5 year periods and these corresponding topics are also present in the overall Top 10 topics of all Billboard Top 100 songs from 1955-2015.
- Some of the common Topics and terms are:
 - 1) Money- Money, Pay, Buy, Price, Big, home and similar related terms and synonyms.
 - 2) Love- Lover, kiss, soul, dear, dream, darling, friend, babe and similar related terms and synonyms.
 - 3) Feelings- Heart, soul, treat, feel, stay, belong, wrong and similar related terms and synonyms.
 - 4) Body- Shake, body, booty, party, work, pump, jump and similar related terms and synonyms.
 - 5) Religious- Pray, heaven, help, god, angel, child, brother, soul and similar related terms and synonyms.
 - 6) Age/Gender – old, young, age, die, time, run, boy, man, girl and similar related terms and synonyms.
- We also discovered that songs having Chorus and verses are more likely to enter into billboard top 100. As chorus and verses are present in more than 50 percent of the billboard top 100 songs from 1955-2015. This can be validated by looking in the overall top 10 topic list (Fig. 5), where the terms Chorus and Verse are ranked 2nd
- It can be seen from Figure 6 that song lyrics starting from 1955-1990 focused more on topics such as Feelings, love, religion and age/gender described above. Whereas, song lyrics from 1990s- 2015 are more inclined towards topics like Money and Body.
- Also, starting 2000, a disposition of using slang and racial slur in song lyrics can be seen. We can see use that popular billboard Top 100 songs of this period heavily use slang terms like black, nigga, booty, ass etc.
- Also the year 2000 marks dominant use of slang contraction such as 'aint', 'gotta', 'da'. From year 2000 onwards a trend of consonant reduction in songs can be seen, for example, them is reduced to em, nothing is reduced to nothin, him is reduced to im etc.

Time Series Analysis

- Since we have Beats per minute, words per minutes and song duration as properties of songs which have already been in Billboard Top 100, we can make forecasts about these properties such that new songs adhering to the values of our forecast might have higher likelihood of making into the Billboard top 100.
- SAS® Enterprise Guide 6.1 was utilized to perform the times series analysis on attributes of Billboard Top 100 songs such as Words per minute(WPM), Beats per minute(BPM), duration of songs etc.
- We used PROC VARMAX functionality to make optimum forecasts of Beats per minute, Words per minute and duration of billboard top 100 songs.
- Moving Average of past 5 years was selected to forecast Beats per minute, whereas Autoregressive moving average model(ARMA(1,5)) was selected to forecast words per minute(WPM) and duration of songs.

Change in Themes of Billboard Top 100 Songs

Jayant Sharma, MS in Business Analytics, Oklahoma State University

John Harden, Business Management Professional, Sandia National Laboratories

RESULTS

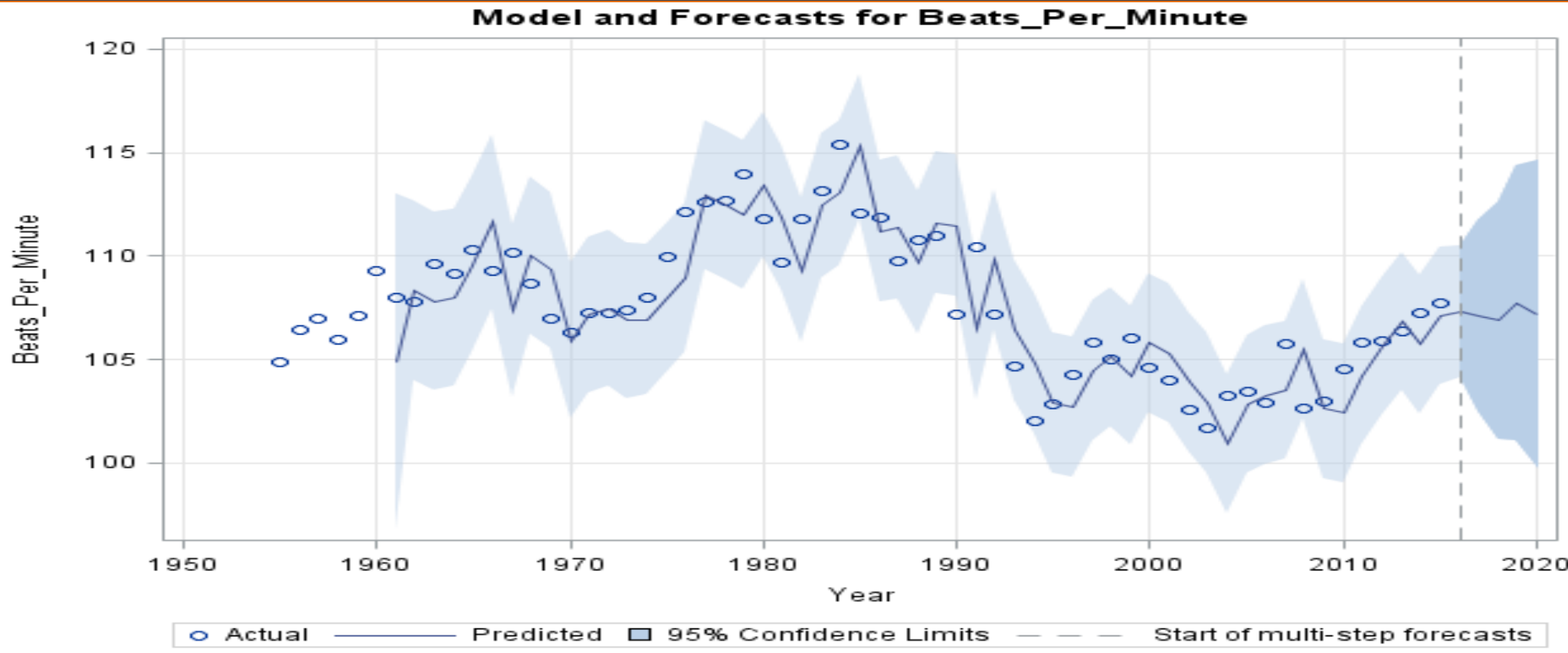


Fig.7 MA(5) Forecast Model For Beats per Minute(BPM)

Type of Model			MA(5)			
Estimation Method			Maximum Likelihood Estimation			
Forecasts						
Variable	Obs	Time	Forecast	Standard Error	95% Confidence Limits	
Beats_Per_Minute	62	2016	107.28062	1.62730	104.09116	110.47007
	63	2017	107.08631	2.35441	102.47175	111.70087
	64	2018	106.89824	2.92448	101.16637	112.63011
	65	2019	107.71355	3.40028	101.04913	114.37797
	66	2020	107.16422	3.80224	99.71197	114.61647

Fig. 8 Forecasts of BPM from year 2016 to 2020

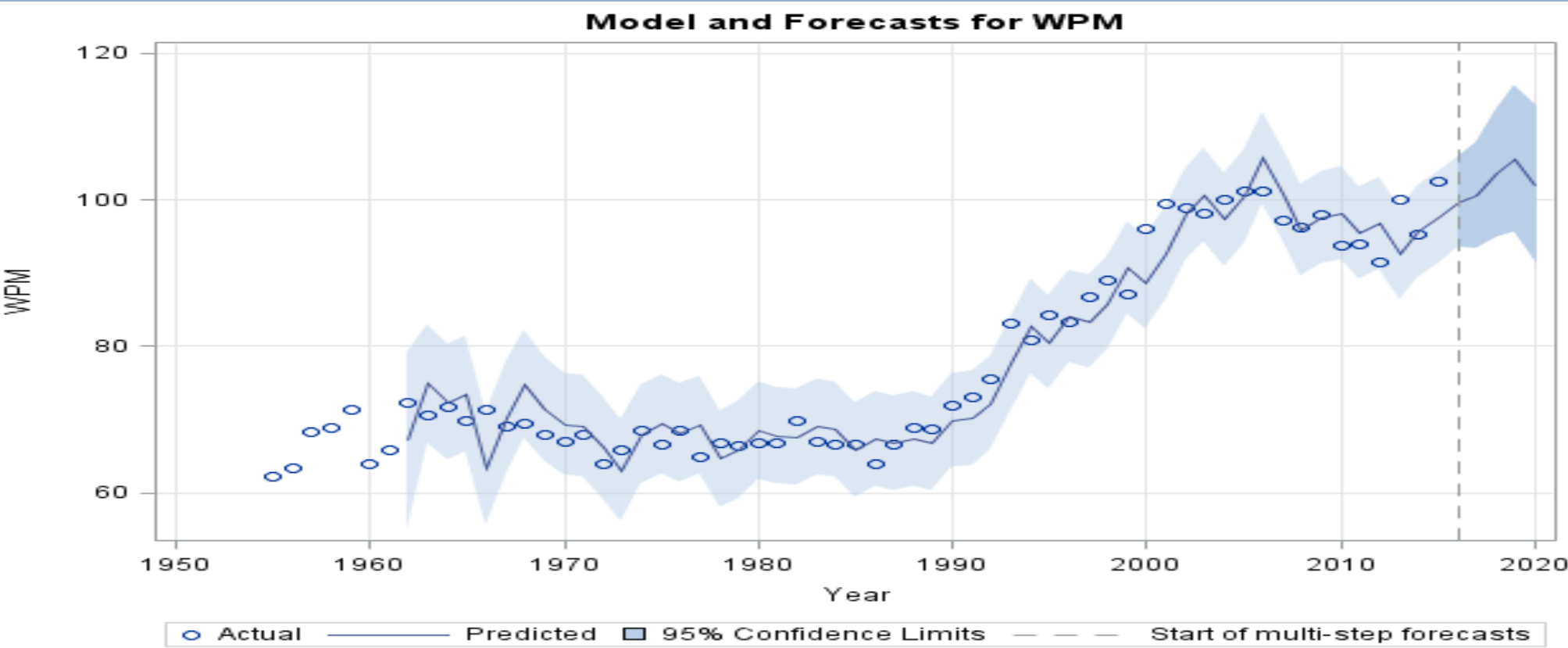


Fig. 9 ARMA(1,5) Forecast model for WPM

Type of Model			ARMA(1,5)			
Estimation Method			Maximum Likelihood Estimation			
Forecasts						
Variable	Obs	Time	Forecast	Standard Error	95% Confidence Limits	
WPM	62	2016	99.73835	3.14843	93.56753	105.90917
	63	2017	100.55645	3.68657	93.33090	107.78199
	64	2018	103.60306	4.45573	94.86998	112.33613
	65	2019	105.62262	5.05583	95.71336	115.53187
	66	2020	102.01435	5.52459	91.18636	112.84235

Fig. 10 Forecasts of WPM for year from year 2016-2020

- From figure 7 and 8, we observe that the forecast for Beats per Minute(BPM) from year 2016-2020 is almost 107. This means that song artists, record labels etc. can have more chances of making into billboard top 100 if they keep the BPM of their songs close to 107.

- Figure 9 and 10 imparts a better understanding of the forecasted values of Words per minute(WPM) from year 2016 to 2020 which ranges from 99.73 to 105.62. By following the forecast of WPM for respective year as shown above, song writers, record label, artist etc. can increase their chances of making into the billboard top 100 list.

- From Figure 11 and 12 we can witness that the forecast for song duration from 2016-2020 is on an increasing trend, increasing from 4.19 to 4.47 minutes. Thus, keeping the duration of songs between 4-4.5 minutes can assist artist, song writers and record labels in making into the billboard top 100 list.

RESULTS

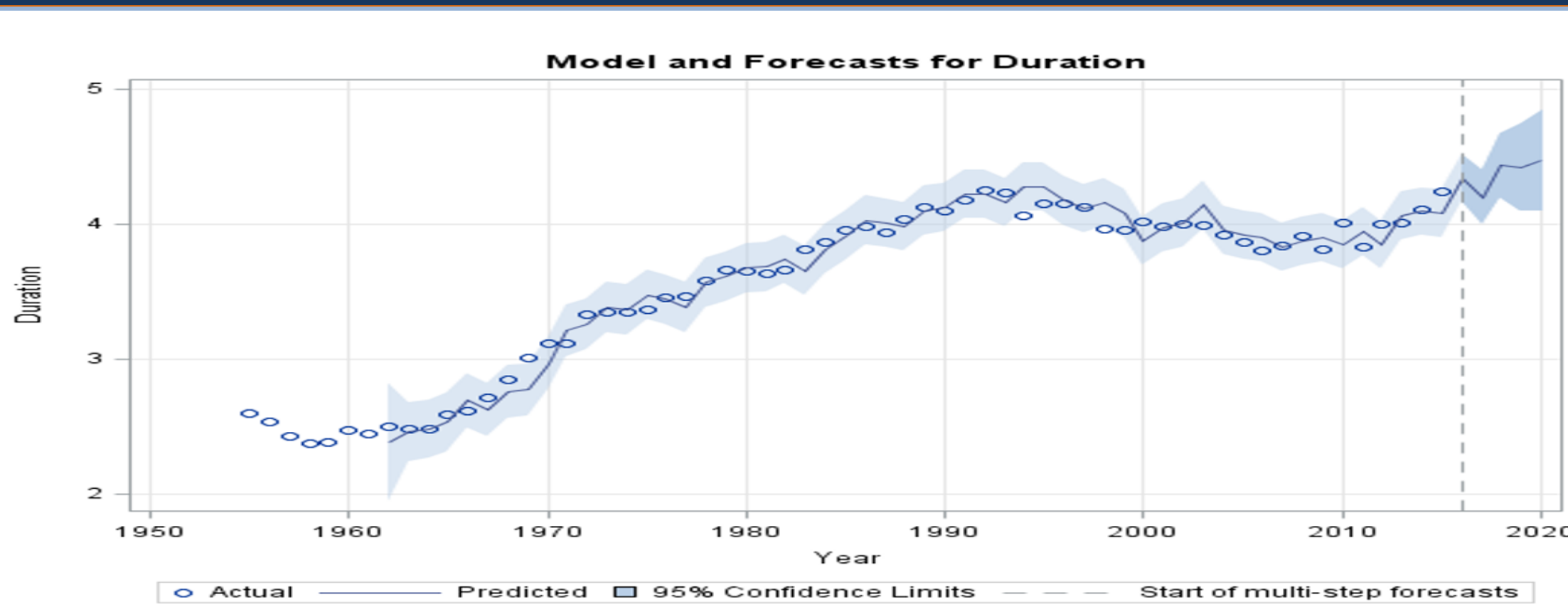


Fig. 11 ARMA(1,5) Forecast model for Song Duration

Type of Model				ARMA(1,5)		
Estimation Method				Maximum Likelihood Estimation		
Forecasts						
Variable	Obs	Time	Forecast	Standard Error	95% Confidence Limits	
Duration	62	2016	4.34066	0.08889	4.16645	4.51487
	63	2017	4.19892	0.10106	4.00085	4.39699
	64	2018	4.43497	0.12152	4.19679	4.67315
	65	2019	4.41779	0.16529	4.09383	4.74176
	66	2020	4.47393	0.19072	4.10012	4.84774

Fig. 12 Forecasts of Duration from 2016-2020

CONCLUSION

The following factors can increase the likelihood of songs making into the Billboard top 100 list:

- By using the following topics and terms in songs: Money-Money, Pay, Buy, Price, Big, home and similar related terms. Body- Shake, body, booty, party, work, pump, jump and similar related terms Feelings- Heart, soul, treat, feel, stay, belong, wrong and similar related terms.
- By making use of Chorus and verses in a song. Additionally, the presence of slangs, slang reduction, slurs and consonant reduction in song may increase the probability of entering Billboard top 100.
- By keeping the beats per minutes (BPM) of upcoming songs close to 107 and the duration between 4 to 4.5 minute.
- By keeping the words per minute of upcoming songs between 95 to 105 words per minute

REFERENCES

[Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS by Goutam Chakraborty, Murali Pagolu, Satish Garla.](#)
[Lyric Complexity and Song Popularity: Analysis of Lyric Composition and Relation among Billboard Top 100 Songs.](#)
[Sentiments of Billboard Top 100 Songs over Time.](#)

ACKNOWLEDGEMENT

We would like to thank Dr. Goutam Chakraborty, Professor, Marketing and Program Head, Business Analytics, Oklahoma State University and Dr. Miriam McGaugh, Clinical Assistant Professor, Marketing, Oklahoma State University, for their constant motivation, guidance and support.



SAS[®] GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL