

Predicting the Winner of American Idol

May 4, 2009

There are countless predictions of who will win American Idol, but most look at one factor or are based largely on opinion. I wanted to see if there was any objective way to predict the winner of American Idol? What would happen if I collected available data and created a predictive model?

The American Idol Prediction Model

Final Two = Adam Lambert and Danny Gokey
Winner = most likely Adam Lambert

	Raw Score	Expected Finish
Adam Lambert	1.923236	1
Danny Gokey	1.688147	2
Kris Allen	1.020240	4
Alison Iraheta	0.386957	3

Nearly tied with Danny
 Nearly tied with Adam
Gender adjustment, see below

Since Adam and Danny are nearly tied, the final result will be heavily based on how each is perceived in there final three weeks of performances and on-air presentation.

Kris and Alison will most likely not make the Final Two. Alison has a very low overall score. Although based on the Raw Score, Kris should be #3, the presence of two even stronger male performers and only one female performer should affect the results.

Adjustment for Gender-Based Voting Patterns

There is an odd, but consistent pattern for the Top 4 Elimination. With only 7 cases to examine, this factor has very little predictive reliability. In addition, in the years (1,2,5,6,7) when the gender split was 3 to 1, the 4th place finisher was one of the 3. The only exceptions are AI3 with an all-Female Top 3 and AI4 with an even split.

Gender and Finish Order

	American Idol Finish Order			
	1	2	3	4
AI1	F	M	F	F
AI2	M	M	F	M
AI3	F	F	F	F
AI4	F	M	F	M
AI5	M	F	M	M
AI6	F	M	F	F
AI7	M	M	F	M

The gender of the Top 4 Elimination has always been the same as the Winner.

Background Information on the American Idol Prediction Model

What are the accurate predictors?

Some of these factors make intuitive sense; others are a bit more mysterious.

Predictive Factors

Age
Bottom 3 thru Top 5 Results
Performance Quality
Post Semi-Final Consumer Awareness
Consumer Appeal
Consumer Research Rankings
Buzz Media/Online Share
Geographic Distribution of Supports

Non-Predictive Factors

Geographic Location
Hometown in “Live” Broadcast Area
Race/Ethnic Origin
Betting Odds¹

Factors with Insufficient Data

Musical Genre
Sexual Preference

¹ Betting odds does effectively identify the top 4 finishers.

Previous Bottom 3 Appearance

One of the best predictors of elimination in the last 3 Rounds, is appearances in the Bottom 3. Danny Gokey is the only AI8 contestant who has not been in the Bottom 3. Note: Kris and Adam appeared in the Bottom 3 in the first ever Top 5 Bottom 3. In previous years there is no information for who came in third on the Top 5 show.

AI8 Bottom 3 Appearances thru Top 5 Results

	Number	Weeks
Kris Allen	1	F5
Danny Gokey	0	
Alison Iraheta	3	F11,9,7ii
Adam Lambert	1	F5

AI1 – AI7 Bottom 3 Appearances thru Top 5 Results Show

Season	American Idol Finish Order							
	1		2		3		4	
AI1	Kelly Clarkson	0	Justin Guarini	1	Nikki McKibbin	4	Tamyra Gray	0
AI2	Ruben Studdard	1	Clay Aiken	0	Kimberley Locke	3	Josh Gracin	3
AI3	Fantasia Barrino	1	Diana DeGarmo	3	Jasmine Trias	3	LaToya London	2
AI4	Carrie Underwood	0	Bo Bice	1	Vonzell Solomon	2	Anthony Fedorov	4
AI5	Taylor Hicks	0	Katharine McPhee	1	Elliott Yamin	3	Chris Daughtry	1
AI6	Jordin Sparks	0	Blake Lewis	1	Melinda Doolittle	0	LaKisha Jones	1
AI7	David Cook	0	David Archuleta	0	Syeshia Mercado	6	Jason Castro	1

Consumer Appeal

There are a number of measures of consumer appeal that are used by advertisers to determine who has the potential to be the best spokesperson. These measures have also proven to reliably predict political races and past American Idol contest outcomes. These indices are based on how test audiences scored them on eight factors: Appeal, aspiration, awareness, endorsement, influence, breakthrough, trendsetter, and trust.²

Brand Marketing Index	Contestant
46.47	Adam Lambert
46.46	Anoop Desai
44.24	Danny Gokey
41.27	Lil Rounds
39.62	Allison Iraheta
39.26	Kris Allen
37.29	Matt Giraud

Consumer Attitudes

Consumer research surveys have been very accurate at predicting the strong and weak candidates on American Idol. HCD Research conducts a weekly consumer survey with a large (1,500+) sample of representative Americans. Every week, they ask survey participants who they feel are the top three performers of the week that should move on to the next round. Here are the predictions and actual outcome in the results show.³

The study, by HCD Research, was conducted among 1,500+ American Idol viewers each Tuesday night to track viewers' perceptions of the judges. Americans who were asked which American Idol contestant should be voted off.

After the semi final round, consumers were surveyed on who they thought should move on the Finals. Here are the results for the current Top 4:

% who thought contestant should be in the Finals	
Danny Gokey	50.3%
Alison Iraheta	49.4%
Adam Lambert	48.5%
Kris Allen	24.9%

Kris Allen is the only one of the Top 4 who wasn't widely selected. Danny, Alison and Adam got the highest ratings.

² Source: "Data: Adam Lambert has edge on Danny Gokey in American Idol finals", Newsday, May 1, 2009.

³ The study, by HCD Research, is conducted from a panel 1,500+ American Idol viewers each Tuesday night to track viewers' perceptions of the judges and contestants. Americans who were asked which American Idol contestant should be voted off.

The results as the contest progressed are even more interesting. The survey seems to be a good measure of view sentiment. March 17th, for example, was Adam's performance of Ring of Fire which resulted in 15% of viewers thinking he should be voted off. Anoop Desai survived longer than predicted but he also has a very high brand marketing index; his likability may have overcome the ratings of the performances. Alexis Grace and Jasmine Murray were not the lowest scorers these two early-round performances were poorly rated by the AI judges and on whatnottosign.com.

% who thought contestant should be voted off

Date	in alphabetical order				Elimination Order								
	Kris Allen	Danny Gokey	Alison Iraheta	Adam Lambert	5	6	7	8	9	10	11	13	13
4/14/2009	8.9%	7.2%	8.7%	6.0%	14.9%	17.3%	37.1%	<i>Matt was eliminated but got Judges' save</i>					
4/7/2009	6.4%	5.1%	7.5%	7.9%	5.5%	18.5%	22.5%	26.9%					
3/31/2009	2.8%	4.3%	4.7%	5.2%	4.7%	29.4%	7.4%	5.5%	36.0%				
3/25/2009	3.7%	3.2%	5.1%	7.5%	4.3%	11.8%	6.5%	10.1%	31.2%	16.6%			
3/17/2008	4.4%	3.9%	4.9%	15.0%	3.9%	10.1%	8.4%	8.1%	11.5%	22.4%	7.9%		
3/10/2009	3.5%	5.5%	4.8%	4.8%	2.8%	18.4%	7.3%	6.9%	10.3%	6.7%	4.1%	8.0%	17.0%



Lowest ranked performer(s)



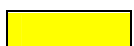
Bottom 3

Performance Quality Perception

It may be a statistically anomaly, but the whatnottosign.com ratings (used in the model as a measure of Performance Quality Perception) are negatively correlated to the Final 4 outcome. In other words, it takes more than great performances to win, and very high scorers often lose. I believe that this may mean that "Likeability" and/or musical genre will beat Performance Quality in the final rounds.

Performance Quality Perception - All Performances (whatnottosign.com)

Season	American Idol Finish Order							
	1		2		3		4	
A11	Kelly Clarkson	76.8	Justin Guarini	52.5	Nikki McKibbin	45.8	Tamyra Gray	72.9
A12	Ruben Studdard	63.6	Clay Aiken	68.3	Kimberley Locke	64.7	Josh Gracin	42.4
A13	Fantasia Barrino	65.4	Diana DeGarmo	58.9	Jasmine Trias	38.3	LaToya London	77.8
A14	Carrie Underwood	57.3	Bo Bice	72.0	Vonzell Solomon	56.1	Anthony Fedorov	40.1
A15	Taylor Hicks	61.3	Katharine McPhee	55.7	Elliott Yamin	65.5	Chris Daughtry	65.4
A16	Jordin Sparks	59.6	Blake Lewis	56.5	Melinda Doolittle	78.7	LaKisha Jones	55.9
A17	David Cook	64.7	David Archuleta	56.5	Syeshya Mercado	53.5	Jason Castro	44.4



Highest rank



Lowest rank of Top 4

Buzz Media

Buzz media measurement has been used to successfully predict political elections and has accurately predicted the AI6 and AI7 winners. Buzz media is a measure of social media coverage (blogs/micro blogs, message boards, forums, micro-blogs, online news media and social networks). It measures share of mentions for each of the contestants (volume of coverage) and the tonality of the coverage. In AI7, David Cook was the clear leader of consumer discussion online.⁴

Contestant	Buzz Volume
David Cook	14.3%
David Archuleta	12.5%
Jason Castro	10.5%

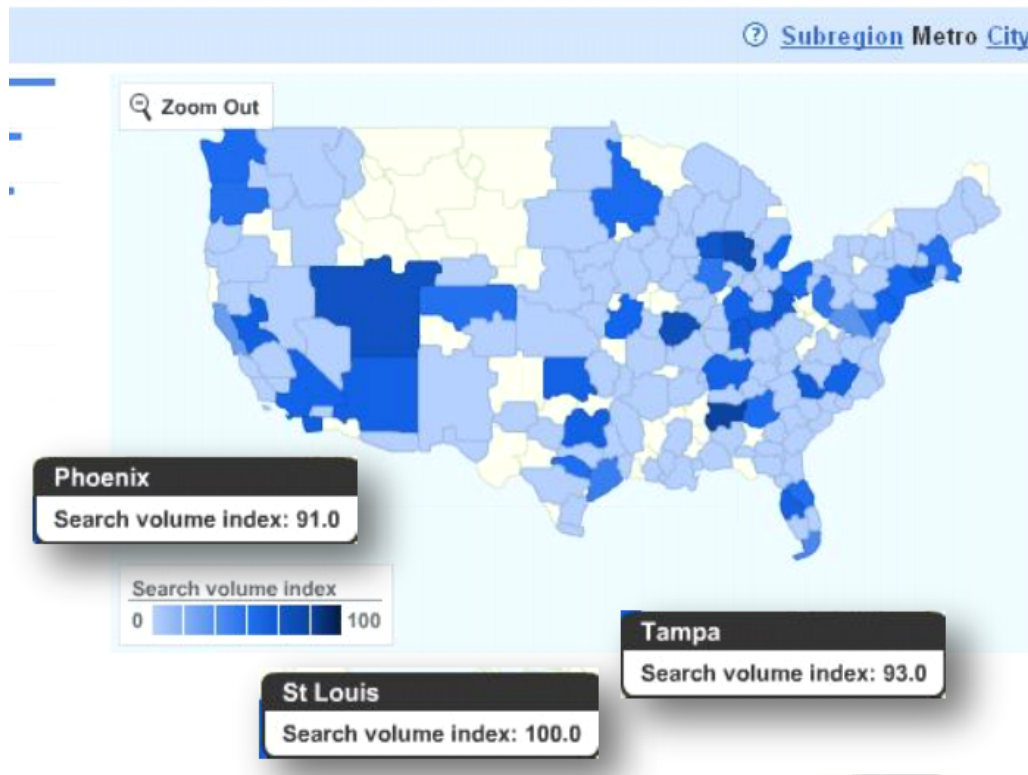
This year the clear leader in Buzz Media volume is Adam Lambert, and he is the clear leader in both positive and negative discussion. On Twitter, Adam is far in the lead and accelerating, Danny is second with Kris close behind and gaining rapidly.

On Facebook, Danny had an early lead but Adam has dominated since late February. Kris Allen's Facebook presence has grow rapidly and he moved into second place in Mid-March.

⁴ "The Nielsen Company Measures the American Idol Phenom", May 15, 2008, Nelson Media Research

Online Traffic

Overall online traffic is at least a measure of interest in a contestant and it has been and accurate predictor of past Idol winners. This is one case were Adam is in a class of his own. On Google, for example, Adam's search volume is 12 times Danny's and 6 times Kris'. Online interest in Adam Lambert is spread across the US and around the World. For comparison, Kris Allen's online popularity is concentrated in Little Rock, AR.



Source: ymarketing The Idol Prediction Project Post #4⁵

⁵ <http://blog.ymarketing.com/idol-global-appeal/Default.aspx?RewriteStatus=1>

From January 31, 2009 through May 1, 2009, the search term “Adam Lambert” was the #3 keyword driving traffic to americanidol.com, and the only one of the Top 5 keywords that wasn’t a form of “American Idol”.

Online interest in Adam and Kris is growing rapidly; interest in Danny is growing much more slowly. But even with Danny’s headstart, Kris has now passed him in average monthly search referrals over the entire AI8 Season.

Average Monthly Online Search Referrals - Top 5 Sites
January 31, 2009 thru May 1, 2009

Search Term (broad match)	youtube.com	myspace.com	americanidol.com	newsday.com	mtv.com
Adam Lambert	335,487	136,975	59,246	101,843	74,662
Danny Gokey	14,265	14,089	8,878		
Kris Allen	19,548	19,568			8,296

Search Term (broad match)	beliefnet.com	rickey.org	wikipedia.org	krisallenation.com
Adam Lambert				
Danny Gokey	42,088	11,327		
Kris Allen			10,267	9,408

Search Term (broad match)	Average Total Monthly Traffic to Top 5 Sites
Adam Lambert	708,213
Danny Gokey	93,156
Kris Allen	64,313

Total keyword search traffic for Alison Iraheta's top five destinations less than 10,000 per month.
Traffic measured from 1/31/09 through 5/1/09. Source complete.com

What really doesn’t matter?

First of all, there is a lot of American Idol “folklore” that just doesn’t hold up under statistical analysis. I identified a number of potential predictors that do not have a proven track record of proven winners.

Notably, I have not used betting odds. In the past bookmakers have not been very accurate on their weekly or final picks. For example, bookmakers (betters) overwhelmingly selected David Archuleta and Clay Aiken as winners.⁶

Other predictors, were effective for the first few seasons of American Idol, but no longer seem valid. The Southeastern dominance that was clear for the first 5 seasons of American Idol, completely vanished in AI6 and AI7. This may have been a shift in American musical taste or a

⁶ Note to Clay’s fans: There some evidence that Clay did not win because of lack of phone capacity vs. lack of supporters. There is an equally good argument that Clay’s support was not evenly distributed across the US.

results of the American Idol producers efforts to diversify the types of performers that started in Season AI4.

"Southeastern Bias" appears to have vanished

Season	American Idol Finish Order and Home State							
	1		2		3		4	
AI1	TX	SE	GA	SE	TX	SE	GA	SE
AI2	AL	SE	NC	SE	TN	SE	CA	
AI3	NC	SE	GA	SE	HI		CA	
AI4	OK	SE	AL	SE	FL	SE	PA	
AI5	AL	SE	CA		VA	SE	NC	SE
AI6	AZ		WA		TN	SE	MD	
AI7	MO		UT		FL	SE	TX	SE

This might have been related to the diminished importance of Live broadcast in the contestants' home states.

Home State and Live Broadcast do not seem to matter either -- anymore

Season	American Idol Finish Order							
	1		2		3		4	
AI1	TX	Live	GA	Live	TX	Live	GA	Live
AI2	AL	Live	NC	Live	TN	Live	CA	Tape
AI3	NC	Live	GA	Live	HI	Tape	CA	Tape
AI4	OK	Live	AL	Live	FL	Live	PA	Live
AI5	AL	Live	CA	Tape	VA	Live	NC	Live
AI6	AZ	Tape	WA	Tape	TN	Live	MD	Live
AI7	MO	Live	UT	Tape	FL	Live	TX	Live

The diminished importance of a contestant originating from a "Live" broadcast area may be a result of increased access to online video and broadcasts, or it may be simply an effect of the lack of success for Southeastern performers in AI6 and AI7.

Viewership Location Bias

There is a geographic bias with more viewership in the Eastern and Central time zones where the show is a live broadcast, but this did not appear to skew the results.

American Idol Season 7 Viewership by Region			
Region	% of Total US	% of American Idol Audience	% Over or Under National Average
Southeast	20%	23%	15%
Northeast	20%	21%	5%
Pacific	21%	19%	-10%
East Central	12%	14%	17%
West Central	15%	14%	-7%
Southwest	12%	9%	-25%

Source: The Nielsen Company May, 2008