

dancing with the stars

By Dan Graettinger

- Is it possible to predict the finalists on Dancing with the Stars before the contestants ever step on the floor?
- Is being young and beautiful a sure ticket into the Top 3?
- Do Cheryl Burke and Maksim Chmerkovskiy have the best track records for guiding their partners to the top?

The staff at Discovery Corps wanted to know! Discovery Corps Inc. is a data mining company, and we spend most of our time helping businesses make predictions -- to make them more profitable.¹ But statistics and predictions can also be used for fun, and trying to predict the outcome of Dancing with the Stars was too good to pass up!



So we set to work pulling together data from the twelve previous seasons of DWTS. Knowing the show and what has happened in the past, we started collecting the characteristics of the contestants that our intuition said might be important: age, professional partner, how well-known contestants were, etc. But we also wanted to check out aspects of the contestants that might not seem to be relevant at first, like gender and occupation and marital status. We decided to throw all the data into the mix -- not prejudicing our expectations based on anecdotes, but letting the data tell the story.

And we came up with some really interesting results! Here's what we found, summarized in a Q&A style format.

What attributes of the contestants were most closely associated to making it to the Finals?

1. Dance Background. Contestants who have had previous experience with dancing or choreography really do have a leg up. Top 3 performers like Jennifer Grey, Kristi Yamaguchi, and Joey Fatone come to mind. In fact, 10 of the 12 contestants who came to DWTS with some dance background made their way to Finals Night.
2. Professional Partner. With so many different professionals having appeared on the show -- and having appeared various numbers of times -- we decided to create a

percentage for each professional indicating how often their celebrity partners make it to the Finals. Here are some of those percentages:

<u>Professional Dancer</u>	<u>Top 3 Finishes</u>	<u>Seasons</u>	<u>Top 3 %</u>
Mark Ballas	4	8	50%
Kym Johnson, Maksim C. ²	4	9	44%
Derek Hough	3	7	43%
Cheryl Burke ²	4	10	40%
Lacey S., Julianne Hough	2	5	40%
Anna Trebunskaya	2	7	29%
Karina Smirnoff	1	9	11%

3. Youth. From Shawn Johnson at age 17 to Cloris Leachman at age 82, DWTS has brought in celebrities to appeal to all demographics. But placing in the top 3 is a different matter. The average age of all contestants through Season Twelve is about 40, but the average age of finalists is 34. That's not to say that youth makes someone a shoo-in. In Season Seven, 36- and 37-year-olds Warren Sapp and Brooke Burke made the Finals while 19-year old Cody Linley didn't. But on the whole, having some youthful spring in your step is a big help.



4. Occupation: Athlete. Professional athletes have done very well on DWTS. Whether it's their familiarity with an exhausting training schedule or their ability to play through the pain of the small injuries that the dancing competition brings, athletes have finished in the top 3 more frequently than the other celebrities on the show. Of the twenty-eight athletes on the show, twelve have reached the Finals.

What attributes didn't seem to matter?

In a bit of a surprise, beauty, fame level³, and ethnicity seem to matter very little. Apparently, thanks to the judges and the voters, it really is a *dance* competition.

What attributes have a negative correlation to making the Finals?

Age Group - Over 60. No contestant over age 60 has ever made it to the Finals.

Occupation: Model. Contestants who are models have a poorer track record of making it to the Finals than contestants with other occupations.

How well were we able to predict Final 3 contestants?

The Discovery Corps predictive model performed really well! We trained our model using a fraction of the results of the first twelve seasons, looking only at the

contestants' characteristics -- no dancing scores -- and their finishing place on the show. The model assigned each contestant a decimal score between 0 and 1, with higher numbers indicating a stronger likelihood of making it into the top 3. We then assigned each contestant to a 'confidence group' based on their score. For instance, those in Group 1 got the highest scores, and we had high confidence in them. Group 2 got the next highest scores but a little less confidence from us ... and so on down to Group 10, the lowest group. (Many of our business clients only pursue the top group.) In the twelve past seasons, a total of 135 contestants have competed⁴ and 36 have finished in the top 3. So a person randomly throwing darts at the names of the contestants would have about a 27% chance of selecting a finalist. But contestants placed into Group 1 by our predictive model landed in the Finals 66% of the time! That's two and a half times better -- and that's why data mining and predictive modeling are so useful.

But what happens if, for a given season, our model only puts one contestant into the high-confidence group (Group 1)? For example, in Season Twelve our model predicted only Chelsea Kane as a high-confidence pick. What if DCI had to choose a top 3, no matter how confident we might be about each of them? Turns out, we still fare pretty well. We would have successfully chosen 22 of the 36 contestants who made it to Finals Night over 12 seasons. That's 61%. So in a typical season, our model is very close to being able to pick two of the three contestants who would make it to the Finals. And for three of the seasons, we would have 'run the table', correctly picking all three finalists!



Does that mean that, before the dancers ever take to the floor, we can pretty well predict two of the three finalists?

At first, this thought discouraged me. Being able to make such good predictions might make the show less exciting. But then some things occurred to me:

- Our success rate of 61% is for the entire length of the show, not for any specific season. There have been three seasons where we predicted only one of the three finalists, and another where we struck out completely. The finalists were dark horses who came out of nowhere! We don't know until the season unfolds who is going to perform way better than expected, and who will take an early dive. Who could've predicted that 52-year-old, no-dance-background Donny Osmond would win the whole thing in Season 9???!

- When it comes to Dancing with the Stars, a big part of the fun is rooting for an underdog or for a contestant who has a great attitude. Yes, there may be the odds-on favorites out there, but our hearts go out to contestants for all kinds of personal reasons.

Who does the model predict for Season Thirteen?

- Hope Solo got the highest score and a high-confidence rating.
- Kristin Cavallari and Rob Kardashian received the next highest scores, but with diminishing confidence for each.



How about some other fun facts?

- Biggest Overperformers: **Kirstie Alley and Marie Osmond** - Our model put both into Group 9, almost as low as our scores go. But Marie made magic in Season Five, garnering third place. And Kirstie Alley last season? Forget about it!
- Biggest Underperformers (at least as far as our predictive model is concerned): **Sabrina Bryan and Natalie Coughlin** - Both were young. Bryan had a dance background and Coughlin was an Olympic athlete. Our model put both into Group 1. Nevertheless, they only placed seventh and tenth, respectively.
- Last-minute contestants have done just fine. Occasionally DWTS will have a planned contestant drop out just before the season starts and will find a last-minute replacement. Of the four such replacements, one of them (Melissa Rycroft) actually made it into the Top 3.
- The height difference between the celebrity and his/her professional partner does not seem to matter. Intuitively we thought it might; but the data says otherwise.

The Recap:

With the analytic and predictive tools of data mining, we can make some excellent predictions -- even about reality TV shows. But perfect predictions elude us. Human motivations and heart are hard to quantify, especially when there's a mirror ball trophy to be had!

Dan Graettinger is a data mining consultant for Discovery Corps, Inc., a Pittsburgh-area company specializing in data mining, visualization, and predictive analytics. You can visit Discovery Corps on the web at <http://www.discoverycorpsinc.com>. Your comments and questions about this article are welcome. Please contact Dan at dgraettinger@discoverycorpsinc.com

Text of this article is © 2011, Discovery Corps, Inc.

Photos are courtesy of ABC.com and Walt Disney Internet Group.

¹ For instance, we help pharmaceutical companies choose the compounds that have the greatest likelihood of becoming a life-saving medicine, or retail chains decide where stores should be located, or marketers know who will respond to direct marketing efforts.

² Both Maksim Chmerkovskiy and Cherly Burke had partners who pulled out of the competition with injuries, so we did not include those seasons in their total number of seasons participating.

³ Both 'beauty' and 'fame level' are subjective characteristics. We did the best we could with these.

⁴ Actually 138 contestants have begun the show; but three withdrew before being voted off.