Apply BBN model to identify consumer persona in high dimensional parameter space

Yong Zhang
(P&G – Corporate Function R&D, Data & Modeling Science, DS/AI)

Nicole Russell, Hollis Cuffie, Melissa Kreuzer
(SPC R&D Setlist team)

2021 October BayesiaLab Conference at the Laval Virtual World

DATA & MODELING SCIENCES
Unlocking Innovation

Background

Setlist Beauty fulfilled the consumer desire for a digital beauty app. Learning how to categorize consumers interested in a digital app concept and provides data for guiding the next steps needed for the brand.

How?

Study consumers to categorize into segments of interest.

Introducing…

Methodology

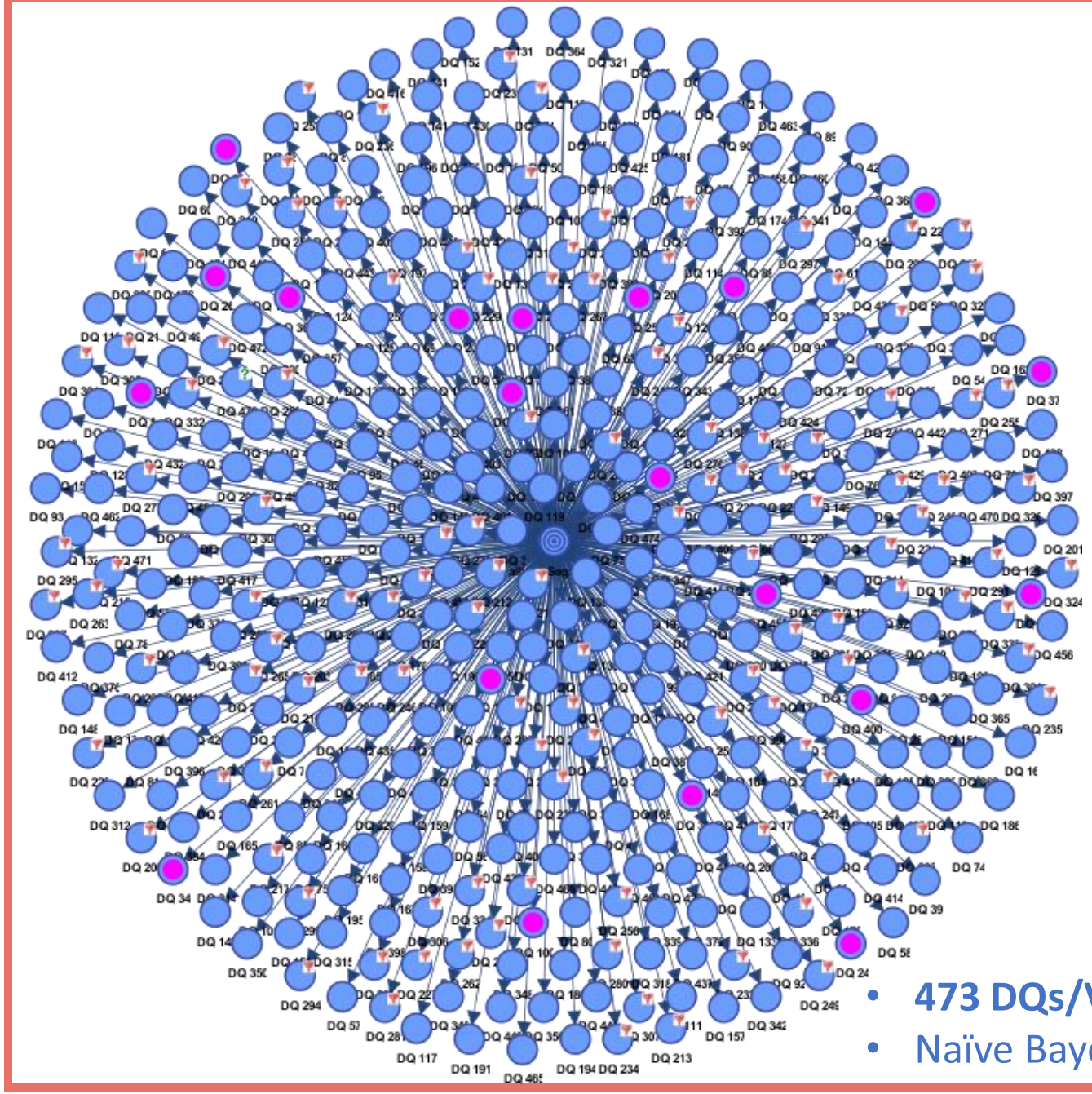
One External Agency

- K-Means Clustering

P&G Data Science

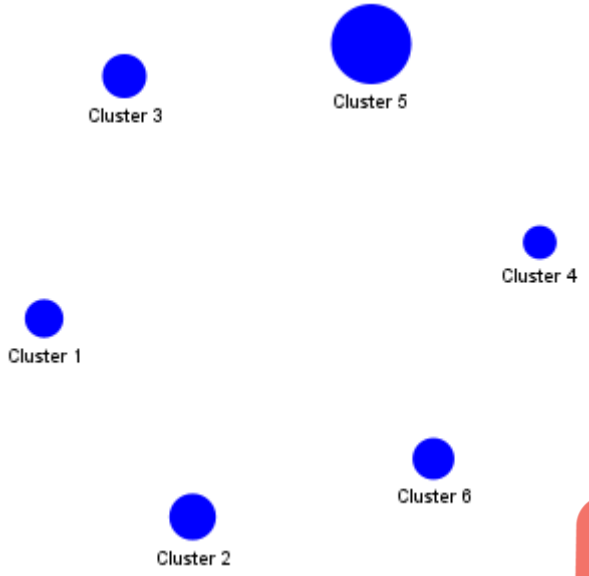
- HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)
- Bayesian Belief Network (BBN)
- Profile Regression
- Polytomous variable Latent Class Analysis (PoLCA)

Other methods have mostly generated 2 segments or few segments without clear distinction in **high dimensional space**.
Variables used are **anonymized**. Some numbers are made up for demonstration and to protect P&G business interest.



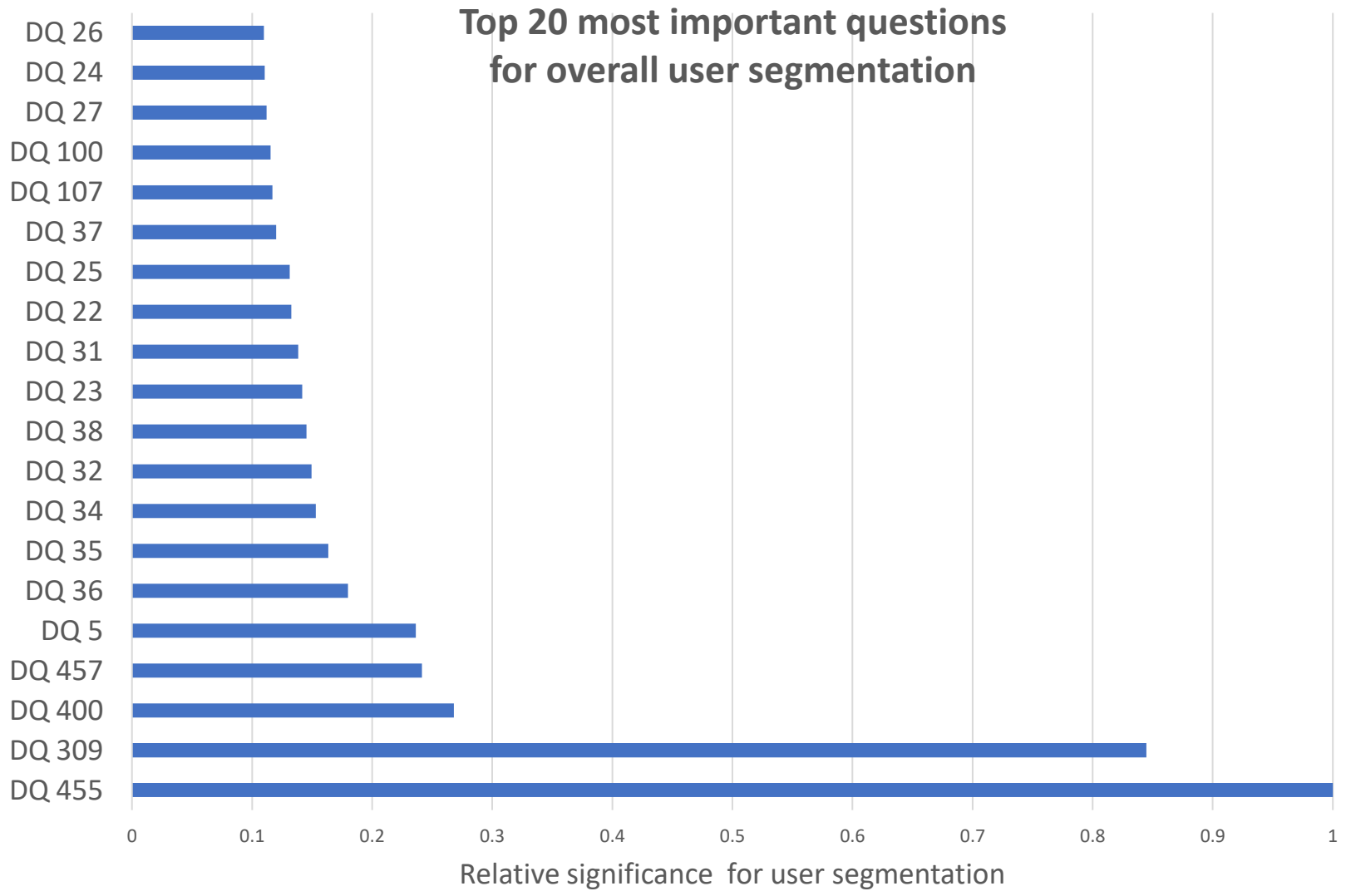
6 Segments Identified

Cluster	Size	Purity
Cluster 1	11.2%	100%
Cluster 2	16.2%	99.99%
Cluster 3	14.7%	99.81%
Cluster 4	8.3%	99.98%
Cluster 5	36.5%	99.96%
Cluster 6	13.2%	99.93%



- **473 DQs/Variables**
- Naïve Bayes Model + Meta Clustering

Top questions to differentiate user segments



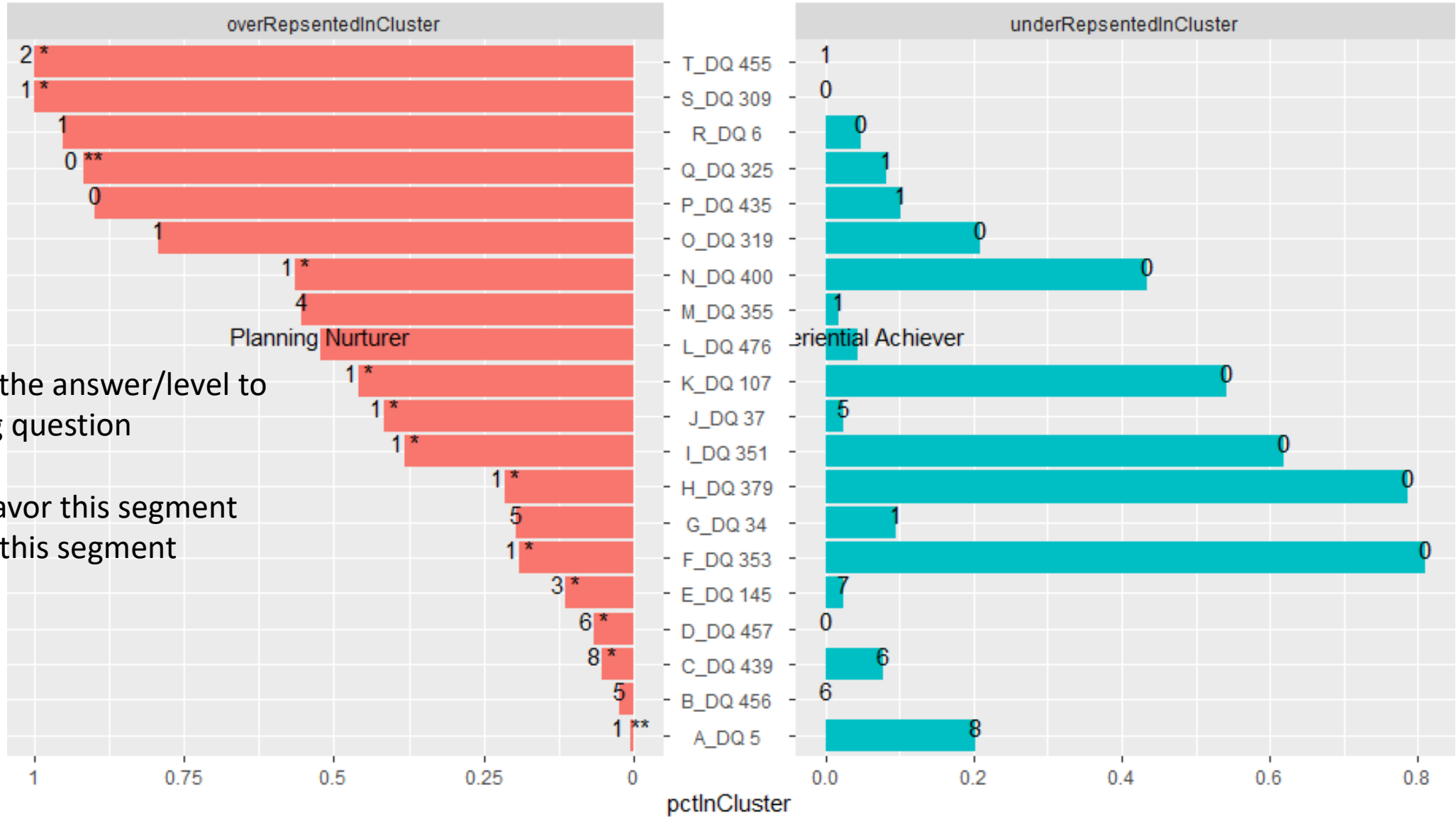
Top 20 Questions & Its Answers Differentiate Segment 1 (11.2%)

Node	Binary Mutual Information	Relative Binary Mutual Information	Binary Relative Significance	Posterior Mean Value	Max Bayes Factor			Min Bayes Factor		
DQ 309	0.2151	42.4556%	1.0000	1.0000	1	99.9950%	3.2470	0	0.0050%	0.0001
DQ 455	0.1681	33.1872%	0.7817	2.0000	2	99.9950%	2.5544	1	0.0050%	0.0001
DQ 400	0.0553	10.9158%	0.2571	0.5655	1	56.5470%	2.6451	0	43.4530%	0.5527
DQ 37	0.0304	6.0059%	0.1415	2.0537	1	41.6645%	1.8079	5	2.3827%	0.2566
DQ 457	0.0282	5.5752%	0.1313	3.0839	6	6.5475%	2.0419	0	0.0006%	0.0000
DQ 476	0.0234	4.6122%	0.1086	2.2607	Planning Nurturer	52.1689%	1.5108	Experiential Achiever	4.3516%	0.2409
DQ 351	0.0220	4.3333%	0.1021	0.3810	1	38.0964%	2.1600	0	61.9036%	0.7516
DQ 5	0.0193	3.8068%	0.0897	7.7260	1	0.5962%	4.4404	8	20.2371%	0.7631
DQ 107	0.0152	3.0082%	0.0709	0.4583	1	45.8337%	1.7067	0	54.1663%	0.7405
DQ 353	0.0141	2.7857%	0.0656	0.1905	1	19.0507%	2.6154	0	80.9493%	0.8731
DQ 34	0.0137	2.6993%	0.0636	3.5416	5	19.6429%	1.4776	1	9.5248%	0.4817
DQ 319	0.0135	2.6654%	0.0628	0.7916	1	79.1638%	1.2952	0	20.8362%	0.5359
DQ 456	0.0133	2.6345%	0.0621	2.8156	5	2.3827%	1.4257	6	0.0020%	0.0027
DQ 355	0.0123	2.4185%	0.0570	3.4404	4	55.3541%	1.3344	1	1.7880%	0.2788
DQ 439	0.0117	2.3088%	0.0544	4.1132	8	5.3577%	2.5865	6	7.7384%	0.5971
DQ 145	0.0114	2.2408%	0.0528	6.9763	3	11.3091%	1.7635	7	2.3814%	0.5021
DQ 325	0.0113	1.1915%	0.0524	0.0834	0	91.6625%	3.6984	1	8.3375%	1.3861
DQ 6	0.0108	2.1321%	0.0502	0.9523	1	95.2336%	1.1324	0	4.7664%	0.2998
DQ 435	0.0108	2.1237%	0.0500	0.1012	0	89.8770%	1.1761	1	10.1230%	0.4293
DQ 379	0.0107	2.1060%	0.0496	0.2143	1	21.4314%	2.1383	0	78.5686%	0.8732

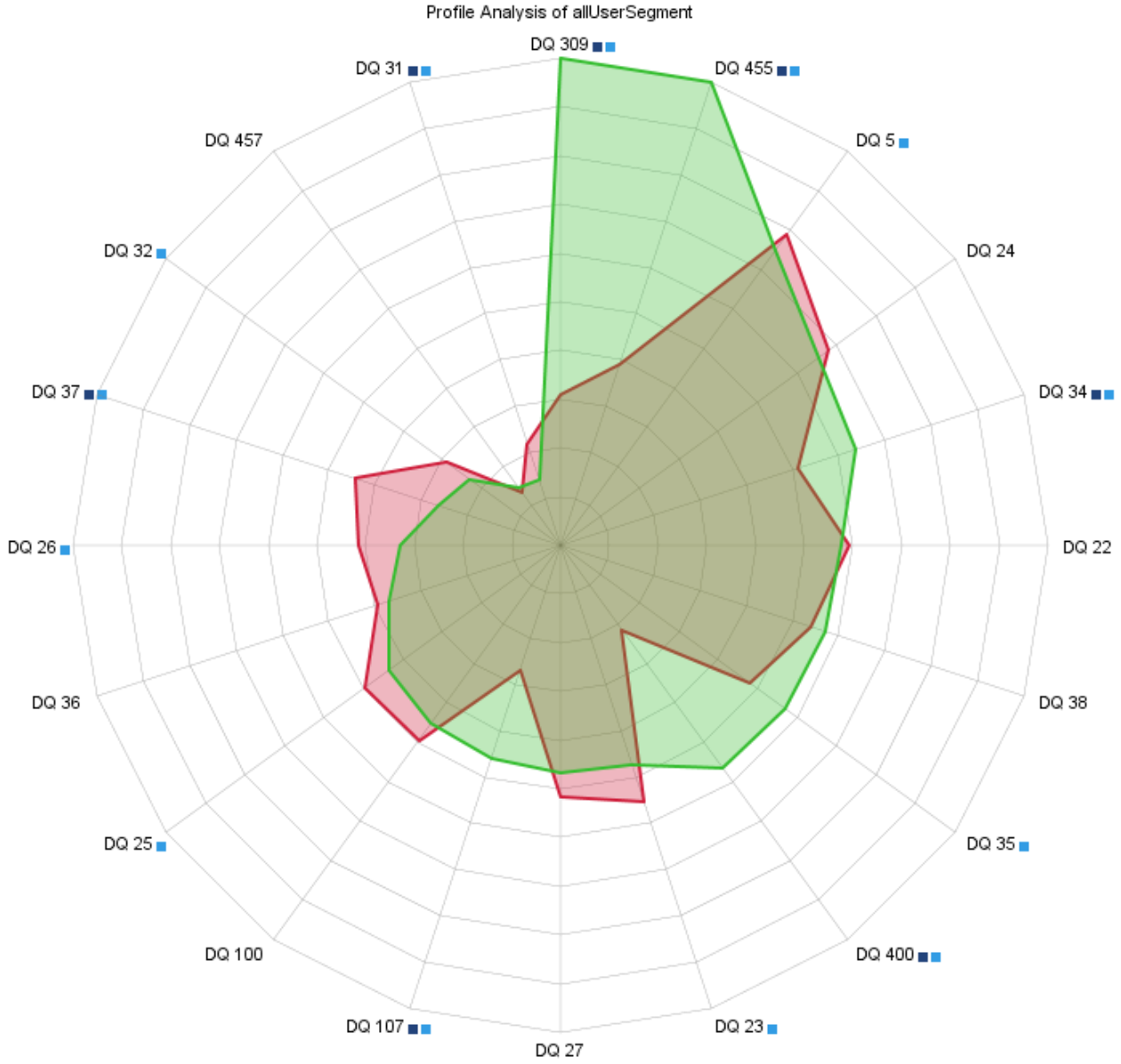
Top 20 Questions & Its Answers Differentiate Segment 1 (11.2%)

Number on bar is the answer/level to the corresponding question

* Substantially favor this segment
** Strongly favor this segment

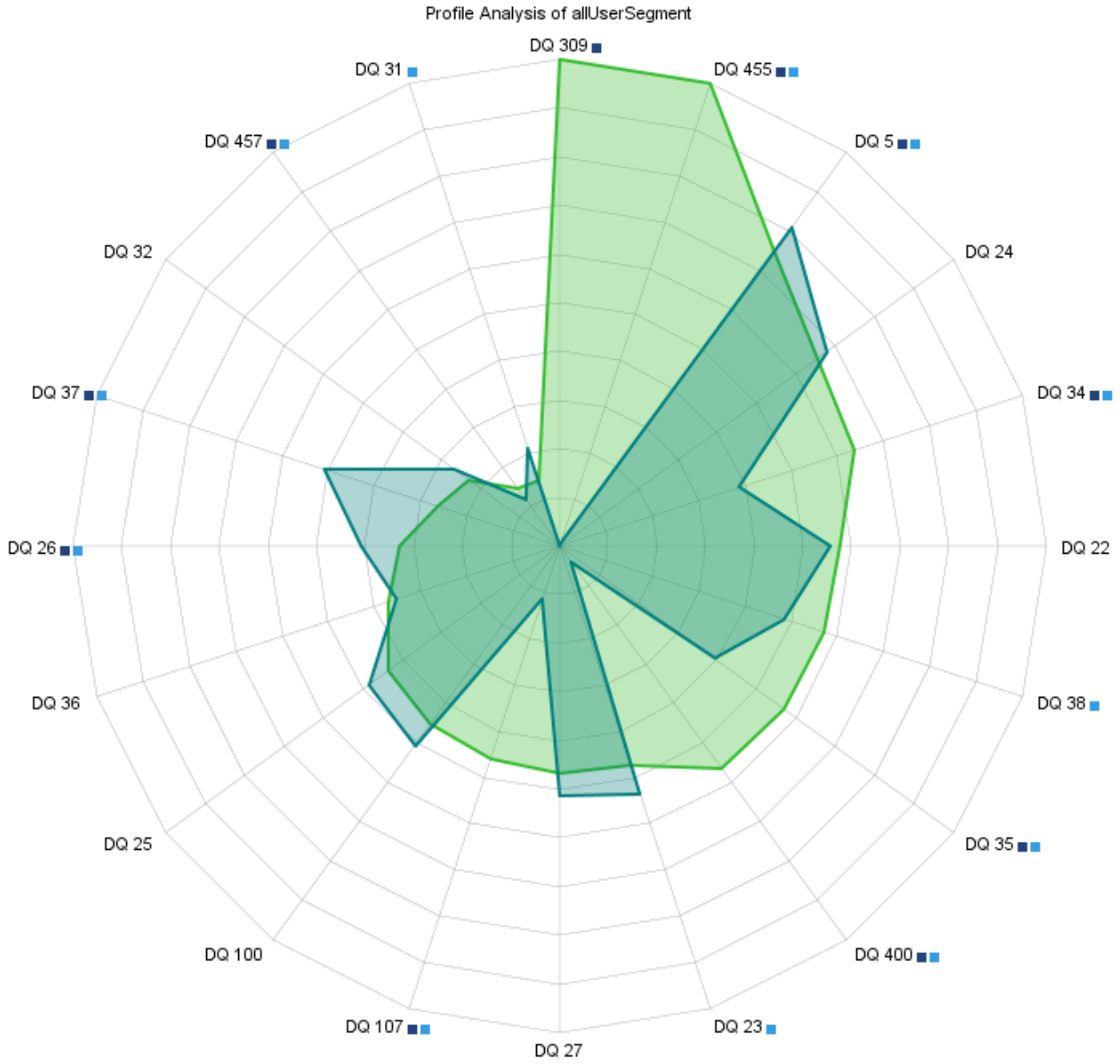


Profile of Segment 1 (Green) and Overall Population (Red)



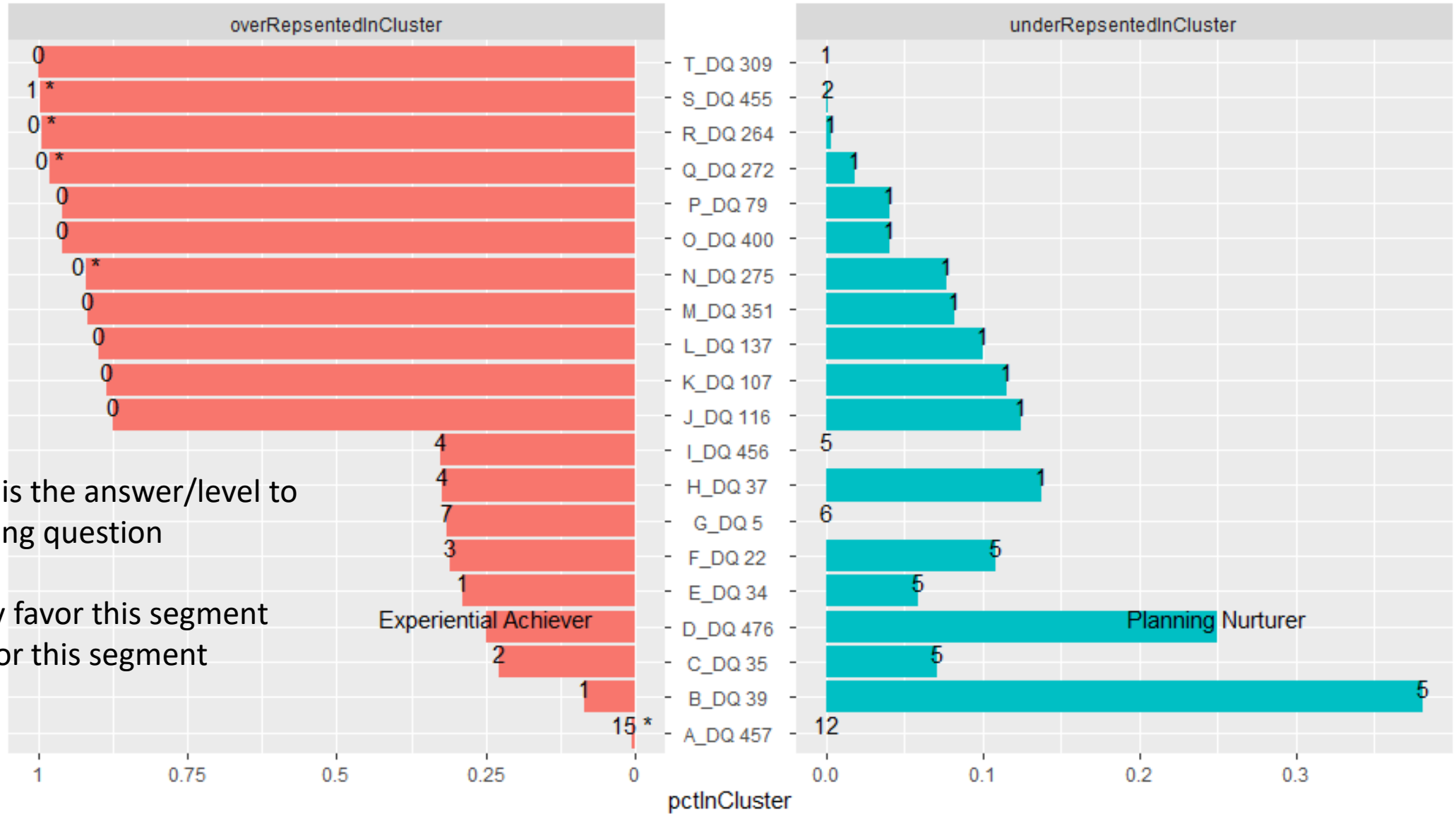
Light blue square indicates statistic significance
Dark blue square indicates practical meaningful

Profile of Segments 1 (Green) and 2(Blue)



Light blue square indicates statistic significance
Dark blue square indicates practical meaningful

Top 20 Questions & Its Answers Differentiate Segment 5 (36.5%)

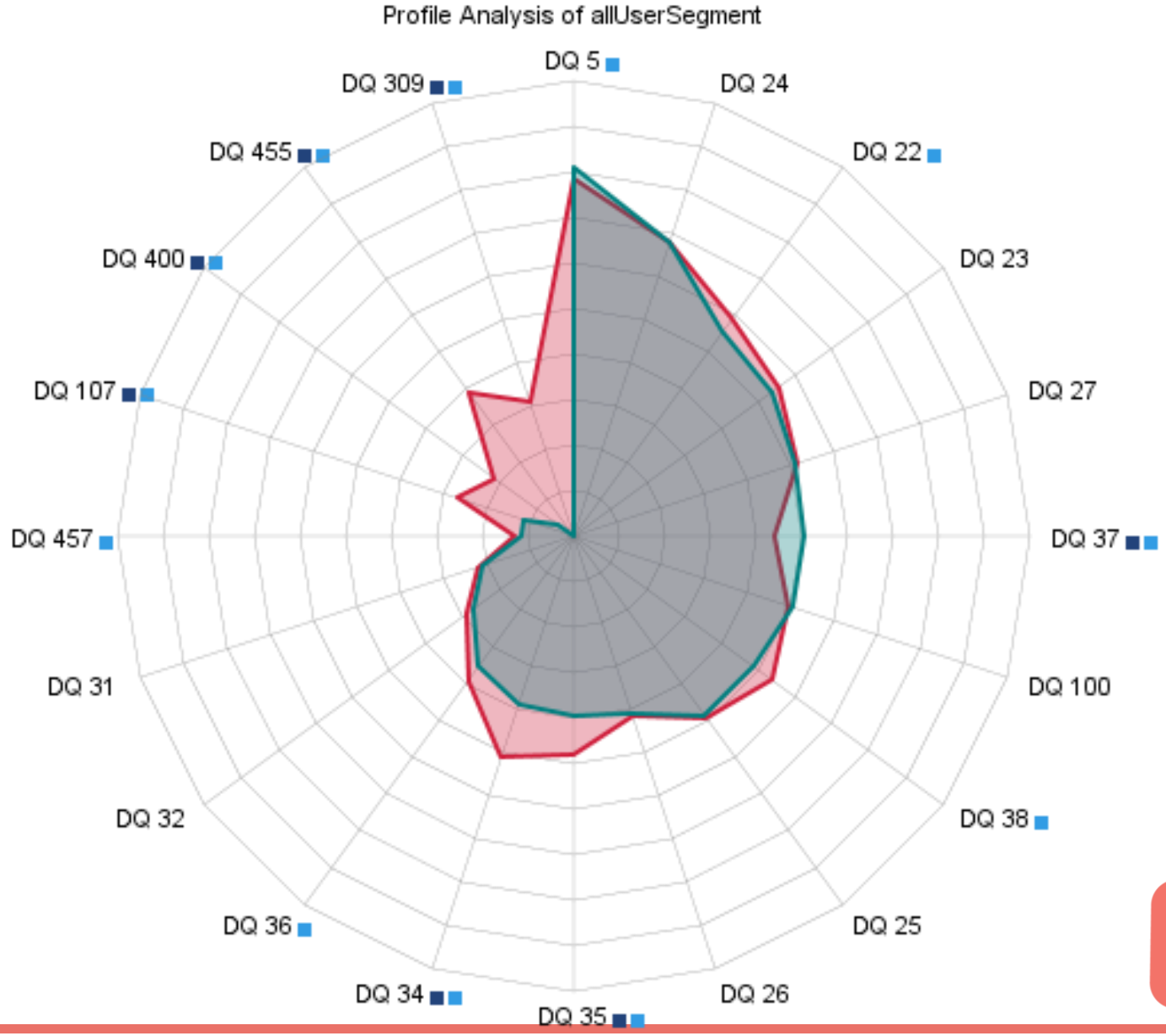


Number on bar is the answer/level to the corresponding question

* Substantially favor this segment

** Strongly favor this segment

Profile of Segment 5 (Dark Green) and Overall Population (Red)



Light blue square indicates statistic significance
Dark blue square indicates practical meaningful

Example - Segment 2

40-50 Y

XXXX

Employed full-time

Desire to look younger.

What they want: The magic bullet in skin care.

Pain Point: xxxxxxx.

Why they are with us: Willing to try something to get results.

Insight: xxxxxxx.

Statement: “I am willing to do something to look younger.”

Example – Understand User Segments & Drive Actions

N = 1498	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
% of Total	11%	16%	15%	8%	37%	13%
Concept Appeal **Scale = 1 -4**	★★★★	★★★	★★★	★★★	★	★
Setlist Appeal **Scale = 1 -4**	★★★★	★★★★	★★★	★	★★★	★★
JTBD	xxxxxxx	Find the best thing in anti-aging to slow down the look of aging skin.	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx
% Setlist Users	x%	20.0%	x%	x%	x%	x%
NPS Score	x	8	x	x	x	x
Likelihood to Download **Scale = 1(👎👎) – 4(👍👍)**	👍👍	👍👍	👍👍	👍	👎👎	👎