

- **ARTICLE**

[Analyzing Diversity & Inclusion with SQL - KDnuggets](#)

- **SUMMARY**

The article introduces how a forgotten aggregate function in Database Mgmt. using SQL, might help Data professionals who seek answers related to Diversity, Equity, and Inclusion within enterprises. After a few descriptions of data classes and operations for categorical & numeric variables -- generally, aggregations between these offer a plethora of operations compared to two categorical combined -- the author sees a simple statistical index called Entropy to compensate for such a shortage and uses it on an HR dataset, retrieved on Kaggle.com, to explore ways to analyze DEI. With facile steps, he gives a description of Entropy *as a numeric measure to describe how diverse something is* and shows how it works using a bag of marbles with only two colors as an example. The result is very intuitive.

- If the bag has 100 marbles of one color & 0 of the other one: The entropy value's 0.
- If the bag has 50 marbles of one color & 50 of the other one: The entropy value's 1.

Then, he applies the same logic to the dataset that would allow him to construct the query to answer his questions. Finally, he concludes his article by restating why using such an approach is pragmatic.

- **ANALYSIS**

Reading this article helped me discover a new aggregate function in SQL to retrieve precious information hidden in a dataset - especially when cross-categorical values are involved. My experience with relational databases is not vast, I only took an IBM professional where I got to play with SQL for four months. However, it has been intriguing to discover that there is always something to learn, even for Database Mgmt. Analysts with years of experience accrued – it just fortifies my idea that there's no finish line in science fields but an asymptotic learning curve. Furthermore, I happened to apply the Entropy index during my course in Multivariate analysis when I was a student at the University of Rome years ago. It is a pleasure to re-encounter concepts I've studied in the past & see those applied in different contexts. Hopefully, it will be useful & precious throughout my career as Analyst//Scientist.