

Adding Value to Market Research Surveys

Background and Business Problem

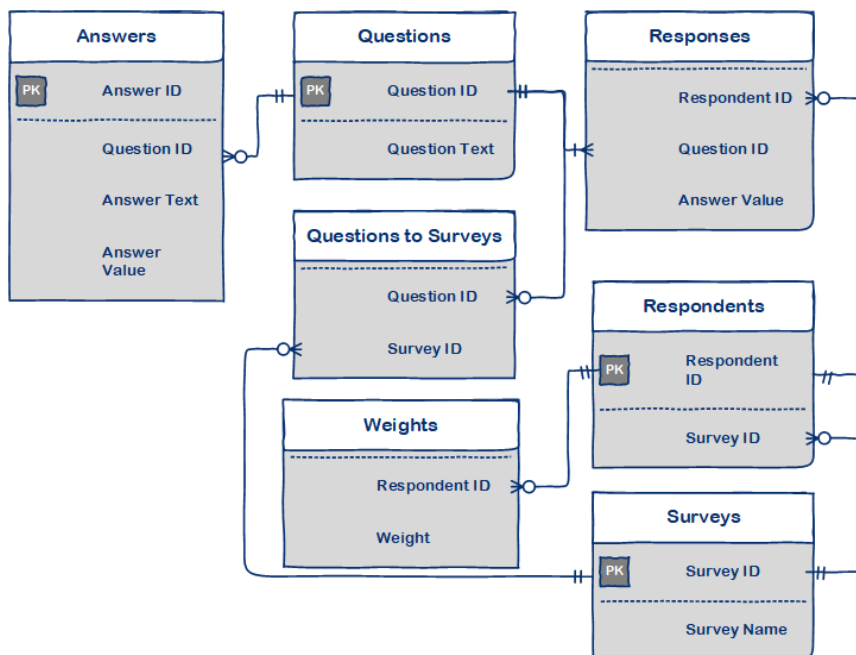
Our client had just completed a large Worldwide survey which had cost many millions to do. The agency which had coordinated the survey were using SPSS to calculate the results and they presented many hundreds of pages of cross tabulations to the client. Our task was to see if could take raw data from the survey and process it on the fly.

We would perform the tasks of cleaning and weighting the data and then computing the metrics. Filters could change the calculations. This would help us to automate survey reporting.

There were so many ways to cut the data and our client wished to perform data mining. For example, each survey respondent lived in a different country, worked in different sizes of company, worked in different industries, had different qualifications, sex, age etcetera. The survey agency had prepared summaries by each country and by industry, but it was not practical to delve deeper. It was hard to visualize what was going on.

The Solution

We started by designing a database schema which could hold all the results of the survey.



The Challenge

Huge investment in running a global survey. The client wanted an online dashboard report to show the results of key metrics.

Users would be able to filter by many audiences and the system would compute the results directly and chart them.

The calculations require working with sets. As each filter is selected the system must find the corresponding set of respondents.

As the sets from each filter are intersected the sample size gets too small to trust the values.

The survey is repeated at intervals. How can we tell if a movement from one wave to another is significant or not?

Solution Highlights

- Database structure to store many waves of survey
- Computation Engine to allow us to compute "on the fly"
- Most popular metrics precomputed and cached in OLAP cubes
- Statistical methods used to add extra information such as weighted sample size and hypothesis testing
- Compute wave on wave deltas
- Online dashboard to show the results
- Custom dashboard widgets to warn users about small sample sizes and results of hypothesis testing

The Solution (Continued)

Once we had our data stored in an appropriate structure, we built some data manipulation tools to pivot the data into a format that was easier to use. Each respondent has many different attributes. For example, the country where they work, their educational background, the industry they work in and size of company. Each of these attributes is found as the answer to a specific question and we scan the responses to the survey to allocate the attributes against each respondent. This makes it much easier to calculate results for a set of attribute filters on the fly.

The shaped data was then loaded into an OLAP cube and metrics were defined.

Metric Value (d) and sample size (e)

$$\bar{d}_k = \frac{\sum_{i=1}^{n_k} w_{ki} d_{ki}}{\sum_{i=1}^{n_k} w_{ki}} \quad e_k = \frac{\left(\sum_{i=1}^{n_k} w_{ki}\right)^2}{\sum_{i=1}^{n_k} w_{ki}^2}$$

Variance

$$s_k^2 = \frac{1}{\sum_{i=1}^{n_k} w_{ki} - 1} \left(\sum_{i=1}^{n_k} w_{ki} d_{ki}^2 - \frac{\left(\sum_{i=1}^{n_k} w_{ki} d_{ki}\right)^2}{\sum_{i=1}^{n_k} w_{ki}} \right)$$

T test of significance in difference in the result

$$T = \frac{\bar{t}_1 - \bar{t}_2}{\sqrt{\left(\frac{(e_1 - 1)s_1^2 + (e_2 - 1)s_2^2}{e_1 + e_2 - 2}\right) \cdot \left(\frac{1}{e_1} + \frac{1}{e_2}\right)}}$$

Finally, we produced a custom Power BI dashboard to show the results of the survey along with color coding to warn about small sample sizes and statistically significant movements.

Extra Value

By creating activity logs the central team were able to perform an analysis of query activity. This told us the degree to which the system was being used but also told us what the business was interested in knowing. This great feedback was used to plan for the next surveys so that we could control costs in reducing unwanted data and focus on the facts in which the business had the most interest.

Our Approach

Partner led project management

Multidisciplinary teams with Data Engineers, Data Analysts and Business Analysts

New Capabilities

New custom survey dashboard

Data Mining

System warns against unsafe conclusions

Technologies Used

SQL Database

SQL Server Analysis Services

R Scripts

Custom Java Script

PowerBI

Secret Sauce

Use different data structures and make sure they are great for the task:

- Database structure for survey data
- Database structure for loading to OLAP
- OLAP structure for computation

Use custom reporting to color code the resulting metrics when sample sizes are too small, or delta movements are significant

Log the accesses to the report and save users' queries to complete a further analysis of what users want to know