

# WCC Business Solutions

## Data Quality Analytics

Improve Matching Results by Improving Data Quality

# Data Quality Analytics

## Introduction

For both staffing agencies and public employment services, helping job seekers find the best job and helping companies find the best candidate are critical for success. ELISE offers powerful technology to find the best match and an incredible range of control-mechanisms to optimize matching, but a crucially important success factor is the quality of the data being matched. Consequently, WCC has developed Data Quality Analytics, a set of tools, reports, and best practices, as well as consulting expertise, to help you find and resolve issues in data quality.

Staffing agencies and public employment services work with large amounts of data – often millions of job seeker profiles and hundreds of thousands of job descriptions. It's incredibly difficult to keep this data consistent. There are often errors that occur with data conversion or replication; there may be data that is simply missing; and there are almost always errors that occur during data entry. Each of these can reduce the effectiveness of the matching process, so finding them and fixing them, or even preventing them in the first place, are great ways to improve the quality of your matches. Working in cooperation with our customers, WCC consultants have identified several types of issues and have developed specific Data Quality Analytics for addressing them.

**Staffing agencies and public employment services work with large amounts of data – often millions of job seeker profiles and thousands of job descriptions. It's incredibly difficult to keep this data consistent.**

## Fill Rate Analytics

### Discover Missing Data

One of the most important quality issues is missing data. If critical data is missing, it can completely invalidate match results. In one recent case, the customer designed a match around location information. When they checked the results, they found it wasn't working. Using the ELISE Fill Rate Analytics, WCC consultants discovered that the fill rate for the location was 0% (See **Figure 1**). This was particularly strange since the main database clearly contained the location data.



Figure 1: Fill Rate Report

The answer was found in the way data was being replicated into ELISE, so this simple analysis quickly identified the underlying issue and allowed the customer to easily improve match results.

### Modeling Matching Weight Factors

The Fill Rate Report can also be used to help model matching weight factors. As you can also see in **Figure 1**, Skills have very low fill rate in the customer's database. In practice, both public employment services and staffing agencies often model skills to have a high impact on the match score because skills are so important in finding the right job. However if job seekers do not provide their skills, weighting the match criteria

# Data Quality Analytics

more heavily on skills will have a negative impact on the quality of the jobs found. A pattern like this usually means that the organization needs to improve the data entry process or they need to adjust the weighting-factors model, in this case, decreasing the importance of skills in the matching process.

## Data Distribution Analytics

### Discover Inconsistent Data

Another issue with data quality results from inconsistent data. To help solve this problem, WCC has developed Data Distribution Analytics that provide statistics on the specific values that are present in every search property.

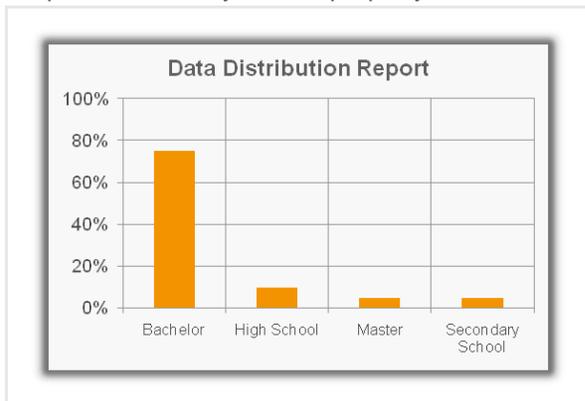


Figure 2: Data Distribution Report

Another customer example shows this in action: the customer reported that users were not able to find the correct educational profile of job seekers for a specific job opening. After using the Fill Rate Analytics WCC's consultants found that all job seekers had the Education field filled in – so missing data was definitely not the problem.

But looking at the Data Distribution for Education (see **Figure 2**), it was clear that three quarters of the candidates had a Bachelors degree. This did not make sense because the customer's business unit was specialized for temporary placements of

blue collar workers such as truck drivers or cooks. Recruiters complained that they could not find the right candidates because most job seekers were registered with a bachelor degree.

...analyzing the quality of the data helped find and resolve the issue and improved the quality of the match results.

This turned out to be a data entry problem: On the screen where the candidate's data was entered, a drop down menu was used to select the education level. Unfortunately, the screen designers used "Bachelor" as the first and default choice. Since most users never bothered entering their education and simply left the default option, there were a disproportionate and inaccurate number of candidates with Bachelor degrees. While the issue turned out to be a user interface design problem, analyzing the quality of the data helped find and resolve the issue and improved the quality of the match results.

### Identify Obsolete Data

Data Distribution Analytics can also help identify obsolete data. For example, one staffing customer had a taxonomy with thousands of occupations that they first created years ago but kept adding to as new occupations were identified. The Data Distribution Analytics showed that only a small fraction of the taxonomy was actually used in the system, but that users still had to navigate thousands of obsolete choices. By removing unused occupations, data entry became easier and faster, and users were also more likely to enter the correct occupation.

## Supply vs. Demand Analytics

### Measure Supply versus Demand

Data Quality Analytics can also be used for business intelligence purposes. For example, Supply vs. Demand Analytics examine the correlation between jobs being sought and all available jobs.

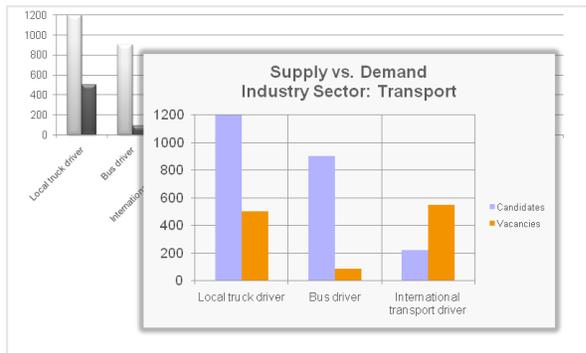


Figure 3: Supply vs. Demand Report

**Figure 3** shows how many job seekers are applying for specific occupations and how many job openings are available. As you can see, there are a disproportionate number of candidates registered as bus drivers compared with the number of openings for bus drivers. Unfortunately, many candidates will remain unemployed in this scenario.

This chart was actually generated when European air traffic was shut down and many companies needed to send goods by truck instead of by air. As a result, there are an unusually high number of vacancies for international transport drivers. Knowing this, it is possible that some bus drivers might be temporarily placed as international transport drivers. With Supply vs. Demand Analytics, staffing companies can manage placements more effectively, and government agencies can identify imbalances between supply and demand.

This report can also identify occupations (in this case bus drivers) where it may be difficult for an unemployed person to obtain a new job. In this way, Supply vs. Demand Analytics can suggest an alternative path such as matching on skills to identify an alternative career or using gap analysis to identify appropriate training to advance a career.

### Analytics Find Business Process Issues

Another business intelligence role for Supply vs. Demand Analytics is to pinpoint business process issues. In one recent customer situation, two branch offices had very different supply vs. demand graphs as shown in **Figure 4**.

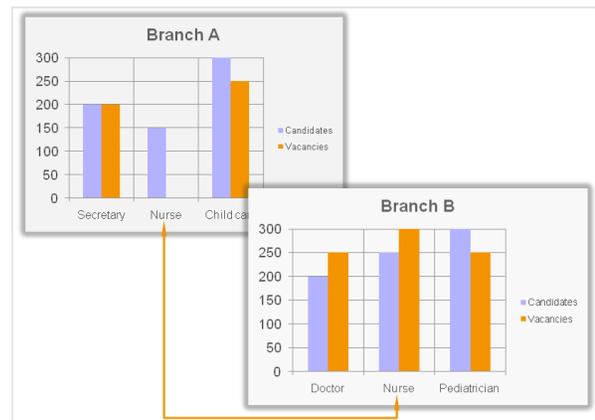


Figure 4: Supply vs. Demand Report

As you can see, Branch A clearly has an oversupply of candidates for nurse positions but without any vacancies for this position. Branch B, however, shows a well balanced relationship between supply and demand for the same nurse position.

This analysis identified a business process issue: nurses were often registered at Branch A whereas they all should have been registered at Branch B, which specializes in healthcare jobs.

## Search Analytics

### Identify Most Used Search Criteria

In addition to analyzing the data in the database itself, much can be learned by looking at how the system is being used – in other words, which fields are job seekers and employers using most often to specify their search.

For example, in **Figure 5** the criteria that are used most often are measured. In this case, users most often searched on keywords, then on location and so on. This insight can be used to further optimize search weights by making the most frequently used properties more important in calculating the match score.

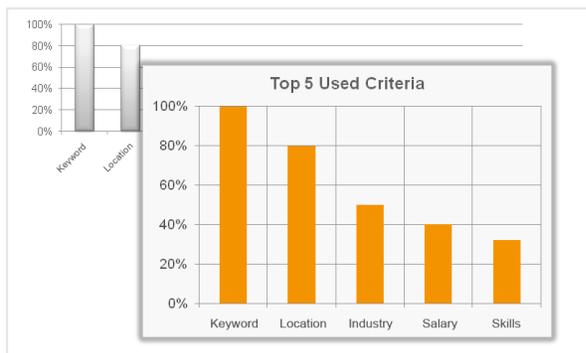


Figure 5: Search Analytics – Most Used Search Criteria

Likewise, if some search criteria are used very rarely it might be sensible to remove them completely, so that they do not lower matching scores artificially.

Such analysis can also be used to help design the user interface for the search system. Knowing which criteria are used most often may indicate which ones should be most prominent in the user interface.

### Identify Most Used Values

Understanding which words or values are used search most frequently can provide useful insight into the way users are actually searching. The first graph in **Figure 6** shows the most popular search keywords and the second graph shows the most frequently sought industries .

This detailed information can help to further optimize the matching process and can result in even better matches for job seekers. For example, at one of our customers we discovered that a significant percentage of users typed the name of the city where they lived in the Keyword search field. However, the system was configured in such a way that keyword searches were only being matched against Job Titles and Company names. As a consequence, many users were dissatisfied with the search results. Search Analytics helped identify this issue and after reconfiguring the Keyword search to search in Job Locations as well, job seeker satisfaction with the match results increased significantly.

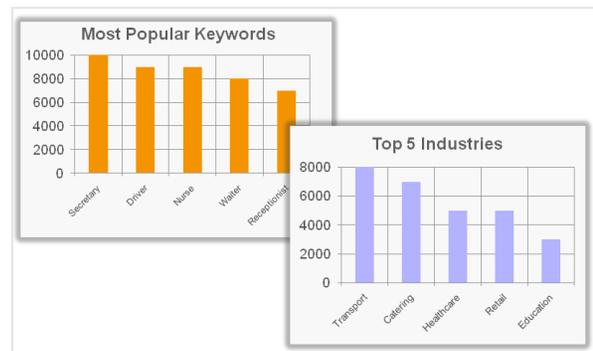


Figure 6: Search Analytics – Mostly Used Values

### Optimize Affinities

Detailed analysis of search usage patterns can also help in compensating for common spelling errors. By creating a list of words most likely to be spelled incorrectly, it is possible to return the right

# Data Quality Analytics

vacancies to job seekers despite incorrectly spelled keywords. For example, in **Figure 7**, an affinity table is shown in which several alternative spellings of “receptionist” are specified in order to return the correct match results despite spelling errors.

	Receptionist	Receptionest	Receptonist	Resepionist
Receptionist	100%	90%	90%	90%

Figure 7: Search Analytics - Typo Affinities

Another direct benefit of analyzing search keywords, is the ability to define affinities between words: For example, job seekers looking for a “receptionist” position may also type in other keywords like “office manager” or “office assistant” or perhaps even “secretary” as shown in **Figure 8**. By defining affinities between alternative keywords, more jobs can be presented to job seekers that are related to their search, thereby increasing the likelihood of a successful match.

	Receptionist	Office Manager	Office Assistant	Secretary
Receptionist	100%	90%	50%	60%
Office Manager	90%	100%	70%	60%
Office Assistant	70%	70%	100%	40%
Secretary	60%	60%	40%	100%

Figure 8: Search Analytics - Keyword Affinities

## Identify Imbalances

From a business perspective, it might be very interesting to know whether job seekers are indeed able to find jobs they are looking for or whether there is a mismatch and exactly in what

type of jobs. A staffing company could use such input to improve on their vacancy and/or candidate pool and government agency could use it to identify imbalances and implement proper policies to address those imbalances.

**Data Quality Analytics can provide immediate, tangible benefits to your business in a variety of ways – all of which lead to better matching and more satisfied customers.**

## Next Steps

As you’ve seen in the preceding real-life examples, Data Quality Analytics can provide immediate, tangible benefits to your business in a variety of ways – all of which lead to better matching and more satisfied customers.

To learn more about Data Quality Analytics services from WCC and how they can benefit your business, contact us at:

[consulting@wcc-group.com](mailto:consulting@wcc-group.com)

WCC Services B.V.  
Savannahweg 17  
Utrecht ■ 3542 AW ■ NL  
Tel. +31 30 7503200  
Fax. +31 30 7503299  
consulting@wcc-group.com  
www.wcc-group.com

Washington DC Office:  
Two Fountain Square  
Reston Town Center  
11921 Freedom Drive ■ Suite 550  
Reston ■ VA 20190  
Tel. +1 703-904-4320  
info@wcc-group.com

WCC is a market leader in developing and licensing high-performance search & match software helping businesses / governments and their customers find people, products or services that best fit their requirements in sub-second response time. WCC's technology is used by many of the largest staffing firms, job boards and Departments of Labor in the world. Its flagship product ELISE is also utilized in the security and content management industries.

WP60-0102-2