



HR Data Remediation Through Automation

(Private Equity Firm)

Cleaned up, updated & streamlined client's existing database (DB) of prospects for hiring executives through a combination of manual searching, automated offline scraping & Python based de-duplication algorithms

HR DATA REMEDIATION THROUGH AUTOMATION

ABOUT THE CLIENT

Client is one of the largest **Private Equity firms** headquartered in the U.S. with investments in companies across various industries

SITUATION



- Client's current database (DB) of prospects for hiring senior executives for their portfolio companies had opportunities to improve data quality and data completeness. These opportunities had increased over time, due to 'multiple sources', 'data loss during transfer from old DB', 'using different variants of nomenclature' etc.
- Merilytics partnered with the client to clean up, update & streamline the existing database through a combination of manual search, automated offline scraping and de-duplication algorithms.

VALUE ADDITION



- Mined external databases (such as Capital IQ etc.) to address mapping of industries/sectors and data completeness issues
- Developed an offline scraping bot to automate the search process and extract required information from the external databases
- Built Python based de-duplication algorithms to match the Positions/Titles with Company master database leading to reduction in redundancies and higher accuracy of information
- Utilized the Python based algorithms to group Positions/Titles with similar company names and manually analyzed the groups to identify companies that can be added to the database
- Once the above companies are added, they can be mapped to the corresponding Positions/Titles

IMPACT



- Achieved data completeness of 'critical' fields for ~98% of the companies in the company master database
- Increased the number of correct Position-Company mapping by ~50%
- Offline scraping automation and de-duplication algorithms reduced manual efforts by 40%
- Updated database improved the filtering capabilities of the tool (by Industry/Company) and helped the client in making better hiring decisions

APPROACH & METHODOLOGY

MANUAL SEARCH PROCESS

Divided the entire process into following steps:

- **Step 1** - Search for the missing company details in online company databases, such as Capital IQ
- **Step 2** - Select the most appropriate company from the search results (also based on the details of the linked 'Contacts')
- **Step 3** - Copy the information on all the missing fields from Capital IQ into client's database
- **Step 4** - Map every company with appropriate Industry and Sector Codes based on the description of the company

AUTOMATION PROCESS

Automated Step 3 in the Manual Search Process by creating an **offline scraper** (the other three steps involved a degree of subjective judgment)

- Downloaded company data from the company databases and created a macro to scrape data from those reports and extract required information into single table

TEXT & REVERSE MATCHING

- Developed **Python based 'Text / Word Matching algorithms'** to search for similar names in the Company Master for every 'Experience detail' (**Position, Company etc.**) in the Position master database, after removing repetitive / common words
- **Ran mapping in reverse for each company**, to generate a list of companies (from Position Master) that have similar company name as in the company master database.
- The methodology used for text matching was **'% partial string match in unordered string'**

Example of downloaded report

[Click here to generate General Info compilation](#)

[Click here to generate Key Facts compilation](#)

[Click here to generate Final compilation](#)

Macro can consolidate
required data from
hundreds of Excel tabs into
single tab in < 2 minutes

OUTPUT

4

REVERSE MATCHING ALGORITHM

ILLUSTRATIVE

Position master

Name	Position	Company Name
Person 1	Director	Merilytics Company
Person 1	Manager	Business Merilytics
Person 2	Analyst	Merilytics Inc.
Person 3	Manager	Merilytics Inc Ltd
Person 4	PE	Merilytics Limited
Person 5	IT Team	Merilytics IT
Person 6	CEO	Meritus

Reverse Matching Algorithm:
Comparing Companies in Position Master,
against company in Company Master, e.g.,
'Merilytics'

'Reverse Matching Algorithm' searches for similar
matches from positions list for every company and
generates results in separate columns

Option to change confidence interval of 'match
percentage' to increase / decrease the number of
matches

List was checked during manual search process and
the matches were added as aliases in Company
Master

Company Name	Match 1	Match 2	Match 3	Match 4	Match 5	Match 6
Merilytics	Merilytics Company	Business Merilytics	Merilytics Inc.	Merilytics Inc Ltd	Merilytics Limited	Merilytics IT

LEARNINGS

- We have learned how to combine different reports and extract required information from hundreds of excel tabs to single tab by using macro
- We gained knowledge on 'Text Matching Algorithms' and used them for the de-duplication of the Position list and to link them to the *Company Master*
- We have improved our understanding on Industry & Sector codes