Strategies to Maximize Batch Locating Service Data in Longitudinal Studies using the Survey of Doctorate Recipients
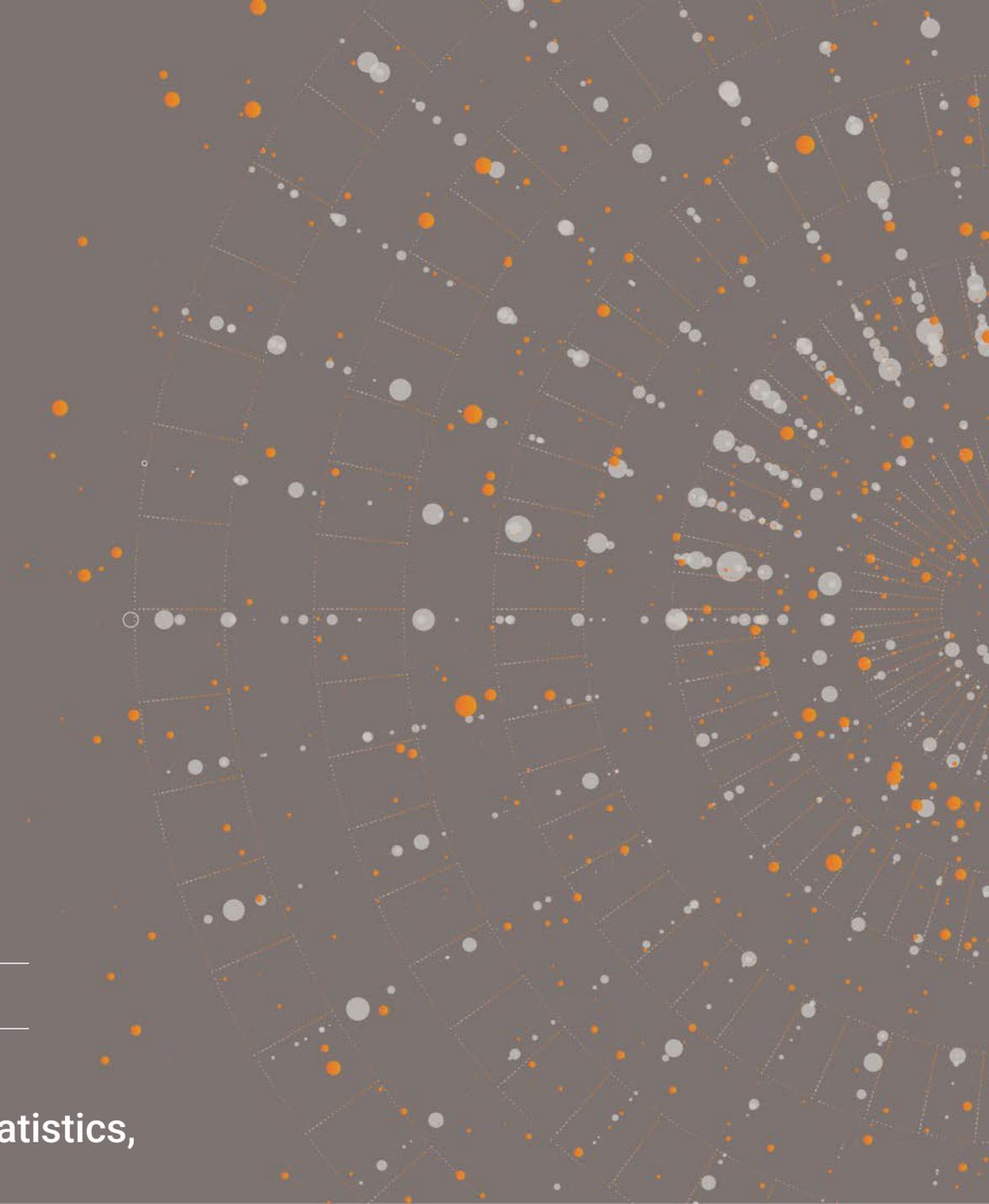
2024 FCSM Research and Policy Conference

10.24.2024

Christopher Wong, NORC at the University of Chicago
Flora Lan, National Center for Science and Engineering Statistics,
U.S. National Science Foundation

Disclaimer

This presentation provides results of exploratory research sponsored by the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF). This information is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any views expressed are those of the authors and not necessarily those of NCSES or NSF.

This product has been reviewed for unauthorized disclosure of confidential information under NCSES-DRN24-077.

Agenda

01 Background on Batch Locating and the SDR

02 Challenges Posed by Batch Locating

03 Methods for Assessing and Using Data from Batch Locating

04 Results from the SDR's Initial Batch Locating Efforts

05 Discussion

Longitudinal studies need quality contact information

Attrition risk if on-file contact information is outdated, incomplete, or missing.

Studies often conduct locating efforts ("tracking" and "tracing") to verify existing or identify new contact information.

- Case-level locating: Conducted manually for each sample unit. Very resource intensive.

- Batch-level locating: Conducted in an automated manner for many sample units at the same time.

Couper MP, Ofstedal MB. 2009. Keeping in Contact with Mobile Sample Members. In Lynn P, editor, Methodology of Longitudinal Surveys, pp. 183–203.

2023 Survey of Doctorate Recipients (SDR)

- Survey sponsored by the National Center for Science and Engineering Statistics within the National Science Foundation and by the National Institutes of Health.

- Sample of U.S.-trained research doctoral degree-holders in a science, engineering, or health field. Each cycle collects key demographic, education, and career history information.

- Began in 1973. Conducted biennially.

- Locating considerations
  - ~125,000 sample members in 2023.
  - Sample members vary by location (domestic/international), career stage (early career through retirement), and sector (academia, industry, government, etc.).
  - Sample members are highly mobile.

Batch locating on the 2023 SDR

- Two batch locating services vendors.
  - NORC held preexisting agreements with each vendor; both vendors had been used on previous NORC-administered studies.
  - The vendors use different data sources.

- NORC identified cases that needed locating and submitted case information to the vendors for matching.

- Depending on NORC's request specifications, vendors returned name, address, telephone number, e-mail address, and mortality information from their database.

- Fees varied by vendor, batch size, and request type.

Ideally batch locating is fast, affordable, and high quality
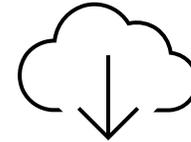

Last Known Sample Member Info.csv

- Name
- DOB
- SSN
- Last known address

Upload file to vendor

Vendor programmatically matches known info to individuals in their database, for whom they have current info

Download updated file

Best Contact Info for Sample Member.csv

- Current name
- Current address
- Current phone
- Current e-mail

Reality of batch locating: Fast, affordable, yet not always high quality

- **Not all returned contact information is fit for data collection outreach.**
  - Vendors serve many clients and may not tune their matching criteria for each client request.
  - Vendor's database may not be effective for locating certain types of individuals (e.g., international residents; people with limited/outdated utility, credit, or financial footprints).

- **Risks of using poor-quality contact information.**
  - Wrong person could be sent survey invitation.
  - Prompts could be sent to outdated information.

- **The SDR needed to develop a system for addressing quality concerns.**

The SDR's process: Laying the groundwork for quality

- Worked with vendor to obtain best documentation.

- Connected with a member of the vendor's data team about process, questions, and wish list items.

- Conducted tests on small batches before ordering full submissions.

- Developed SAS script for selecting best contact information to use in submission files.

The SDR's process: Assessing quality of returned records

- Developed a SAS script to assign match quality variables.

- Focused on
  - Accuracy (i.e., sample person verification).
  - Currentness (i.e., contact information is still in use).

- Used indicators to assess and direct case management decisions.
  - What information could be used for sample member prompting as delivered?
  - What cases should be sent to individual manual locating (i.e., case-level locating, different from batch locating)?

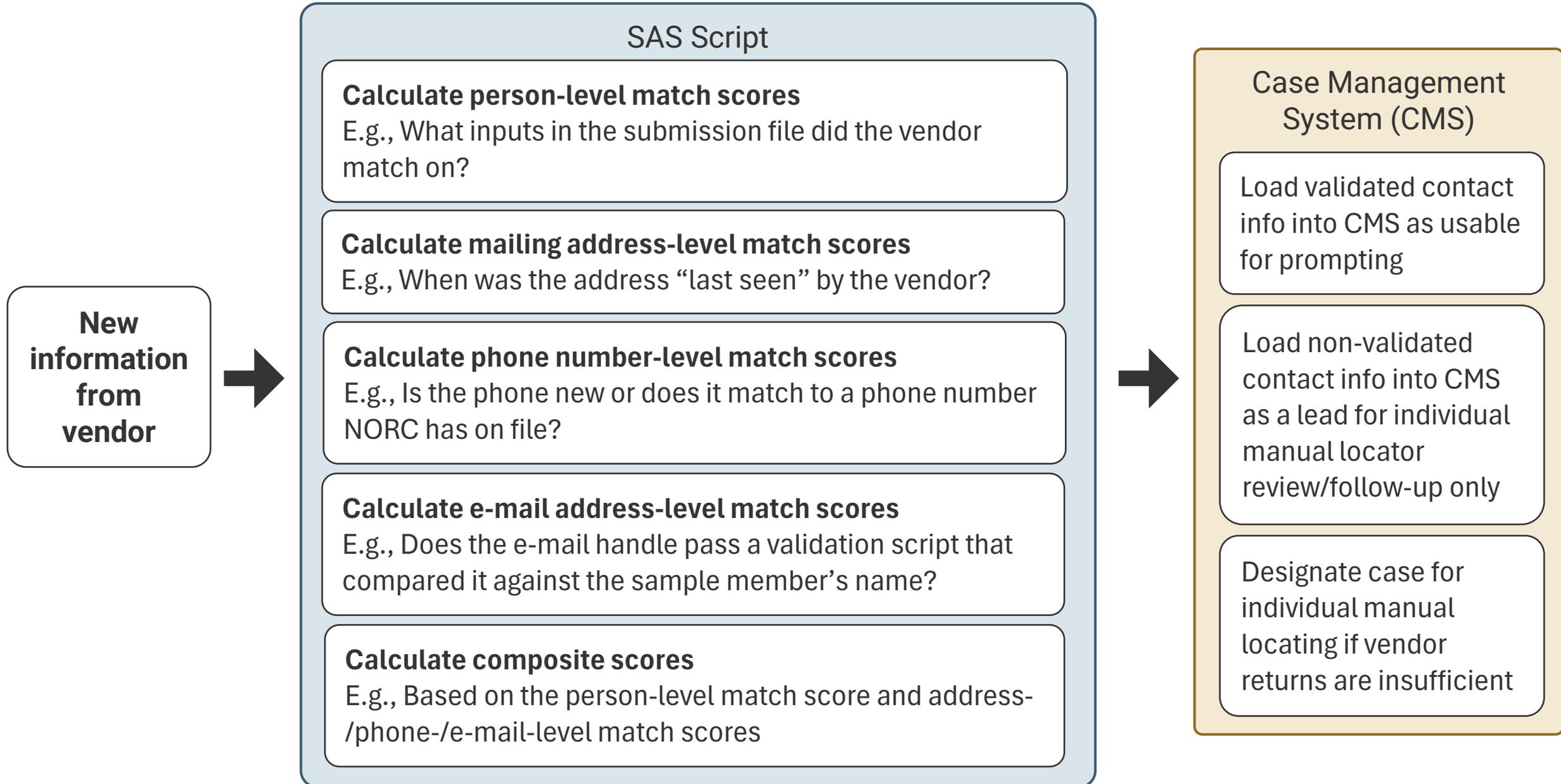| Name	Description	Values
Addr_First_Seen	Date vendor first saw address	[DATE]
Addr_Last_Seen	Date vendor last saw address	[DATE]
Addr_New	Address already on file	Y/N
Addr_Status	Status of on-file address	Active/Inactive

Objective variables (dates, input/output indicator flags)

Used to create

Name	Description	Value	Value Label	Assign If
Addr_Score	Quality score for vendor-provided address	4	High quality	Vendor saw the address within past 1 year. Address is not already on file or, if it was on file, it was not previously coded as inactive.
		3	Medium-high quality	Vendor saw the address within past 2 years. Address is not already on file or, if it was on file, it was not previously coded as inactive.
		2	Medium quality	Vendor saw the address within past 1 year. However, address was coded as inactive during previous SDR round.
		1	Low quality	All others.

Subjective variables (quality scores)

New information from vendor

SAS Script

Calculate person-level match scores
E.g., What inputs in the submission file did the vendor match on?

Calculate mailing address-level match scores
E.g., When was the address "last seen" by the vendor?

Calculate phone number-level match scores
E.g., Is the phone new or does it match to a phone number NORC has on file?

Calculate e-mail address-level match scores
E.g., Does the e-mail handle pass a validation script that compared it against the sample member's name?

Calculate composite scores
E.g., Based on the person-level match score and address-/phone-/e-mail-level match scores

Case Management System (CMS)

Load validated contact info into CMS as usable for prompting

Load non-validated contact info into CMS as a lead for individual manual locator review/follow-up only

Designate case for individual manual locating if vendor returns are insufficient

Initial batch locating returns were high in quantity but mixed in quality

Group	#	%
Submitted to Vendor A for address/phone/e-mail	42,104	100.0
Address returned	36,901	87.6
High-quality address returned	14,658	34.8
Phone number returned	34,988	83.1
High-quality phone returned	33,178	78.8
E-mail returned	26,522	63.0
High-quality e-mail returned	13,842	32.9
Eligible for manual locating post-batch search	33,775	80.2

One vendor was used to conduct all batch searches during the 2023 SDR pre-field period in spring/summer 2023.

The vendor returned some contact information for most cases, but much was assessed as relatively lower quality.

80% of submitted cases were eligible to receive individual manual locating after the batch search.

Cases that were initially assessed as successful in batch locating were more likely to go on to complete the survey than cases that were not assessed as successful

2023 Final Response Outcome	Initial Batch Search Assessed as Successful; Never Received Manual Locating	Initial Batch Search Assessed as Successful; Later Received Manual Locating[a]	Initial Batch Search Not Assessed as Successful; Received Manual Locating
Overall	7,684	645	31,250
Confirmed Located	3,531	223	12,803
Respondent	2,901	124	9,393
Ineligible[b]	8	37	408
Eligible Nonrespondent	622	62	3,002
Final Unlocatable/Uncontacted	4,153	422	18,447

38% who received only the initial pre-field batch search went on to complete the survey.

30% who received both the initial pre-field batch search *and* manual locating completed the survey.

[a] Cases were assigned to manual locating based on assessments of returned mail, bounced e-mail, and telephone prompting outcomes.
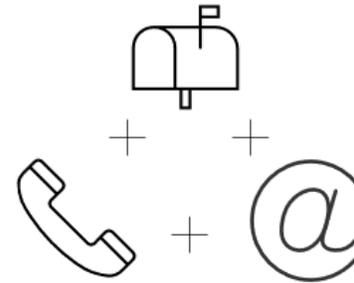
[b] All individuals who were identified by a batch search as deceased were assigned to manual locating for verification.

Individual manual locating efforts remained important

For verification

- Investigating borderline quality information
- Deceased confirmation

For supplementing

- Searching for additional modes of contact information
- Searching for employer information

Other design considerations

Flexibility: Is it easy to adjust scoring criteria to account for different vendors or subgroup needs?

Repeatability: Is it easy to swap inputs when batch searches are repeated?

Accessible output: Is it easy for staff to access the quality indicators even after initial updates to the case management system have been made?

Discussion and next steps

- Batch locating services searches were a powerful tool for SDR locating efforts. But returns had to be carefully assessed, and manual locating efforts were still valuable.

- Future challenges
  - Keeping up with new batch services and changes in batch services.
  - Optimal ways of looking at quality in future SDR rounds may be different than in 2023.

- Future research
  - Further assessing accuracy of quality indicators.
  - Systematic comparison of vendor returns against one another.
  - Cost analysis of batch locating savings over fully manual locating.

Thank you.

Christopher Wong
Research Director
wong-christopher@norc.org

Flora Lan
Survey Statistician
flan@nsf.gov

✳ Research You Can Trust™

✳ NORC at the University of Chicago

Additional results: Follow-up batch locating outcomes during the main data collection period

Vendor	Search Type	N Submitted	High-Quality Address Returned		High-Quality Telephone Returned		High-Quality E-mail Returned	
			#	%	#	%	#	%
A	Address, Phone, Email	22,213	16,306	73.4	20,399	91.8	9,747	43.9
A	Address, Phone	12,588	5,551	44.1	2,560	20.3	-	-
B	Address, Email	27,024	23,867	88.3	-	-	8,574	31.7

- Multiple vendors were used to conduct batch searches during the main data collection period.

- Submitted cases included newly identified locating problems but also partially located or potentially located cases.
  - Contributed to a better high-quality return rate compared to pre-field searches.
  - Even if the vendor's best address matched what NORC believed to be the best address, this was valuable data for informing case management decisions.

Example: Incorporating batch contact variables as part of a broader suite of variables for making outreach decisions

Case ID	Address Key	CMS Status	Original Source	Count of Prompts Mailed	Date of Most Recent Prompt	Ever Matched by Vendor A	Date Matched by Vendor A	Date Last Seen by Vendor A	Ever Matched by Vendor B	Date Matched by Vendor B	Date Last Seen by Vendor B	Vendor B Address Score	Include in Next Mailing?
99990001	1	Active	2021 Quex	3	4/1/2023	Y	1/8/2023	12/1/2022	Y	3/28/2023	3/28/2023	100	Y
99990001	2	Active	Vendor	1	3/12/2023	Y	1/8/2023	5/1/2020	Y	3/28/2023	2/1/2023	95	N
99990001	3	Inactive	Web Search	0		Y	1/8/2023	12/1/2018	N				N
99990002	4	Active	Vendor	1	4/1/2023	N			Y	3/28/2023	3/28/2023	100	Y
99990002	5	Inactive	2015 Quex	0		N			N				N