Column Name	Expected Type	Data Type	Empty Allowed?	Missing Values Handling	Description
comp_name	Text / String	varchar(255)	Yes[1] , but not preferably	Blank cells → stored as ""; other tokens ("NA", "N/A", etc.) stored verbatim (as is).	Company name, stored as `name` in the database.
comp_domain	Text / String	varchar(255)	Yes[1] , but not preferably	Blank cells → stored as ""; other tokens stored verbatim (as is).	Company website domain (e.g. `example.com`), used for matching.
annrev	Integer	numeric	Yes (blank→0)[2]	Values in ["", "none", "unknown", "n/a"] → 0; other tokens (e.g. "NA") not treated as missing (parse or error). Text values are allowed as long as they have a numeric value in the text (this will be extracted (e.g. 100M-1000M)	Annual revenue; parsed via `extract_revenue_lower_bound`.
comp_linkedin	Text / String	varchar(255)	Yes[1] , but not preferably	Blank cells → stored as None; strings like "NA"/"N/A" remain literal.	Company LinkedIn URL or handle; prefixes stripped.

Column Name	Expected Type	Data Type	Empty Allowed?	Missing Values Handling	Description
firstname	Text / String	varchar(255)	Yes, but fullname will be missing this value, creating an extra space most probably	Blank cells → stored as ""; other tokens remain literal.	Contact's first name; later combined with `lastname`.
lastname	Text / String	varchar(255)	Yes	Blank cells → stored as ""; other tokens remain literal.	Contact's last name, merged with firstname
jobtitle	Text / String	varchar(255)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Contact job title, mapped against `dim_jobtitles`.
manlevel	Text / String	varchar(100)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Management level, mapped against `dim_manlevels`.
empemail	Text / String	varchar(255)	Yes	Blank cells → stored as None; strings like "NA"/"N/A" remain literal.	Contact email; used for matching.
emplinkedin	Text / String	varchar(255)	No[4]	Blank or null → row dropped; strings like "NA"/"N/A" are non-empty and ingested as literal.	Contact LinkedIn URL or handle; required.

Column Name	Expected Type	Data Type	Empty Allowed?	Missing Values Handling	Description
comp_phone	Text / String	varchar(100)	Yes	Blank cells → stored as ""; strings like "NA"/"N/A" remain literal.	Company phone number; formatting preserved.
comp_street	Text / String	varchar(255)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Street address, mapped against dim_addresses.
comp_city	Text / String	varchar(255)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	City name, mapped against dim_cities.
comp_state	Text / String	varchar(255)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	State name, mapped against dim_states.
comp_country	Text / String	varchar(100)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Country name, mapped against dim_countries.
comp_zipcode	Text / String	varchar(100)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Postal/ZIP code, mapped against dim_postalcodes.

Column Name	Expected Type	Data Type	Empty Allowed?	Missing Values Handling	Description
qa_disposition	Text / String	varchar(100)	Yes	Only exact "qualified" → status 1; blank, "NA", "N/A", etc. → status 4.	Qualification status; mapped to `emailstatus_id`.
empsize	Integer	Integer	Yes (blank→0)[2]	Values in ["", "none", "unknown", "n/a"] → 0; other tokens (e.g. "NA") not treated as missing (parse or error).	Employee size; parsed via `extract_lower_bound`.
comp_industry	Text / String	varchar(100)	Yes (blank→"Unknown")[3]	Nulls/NaN → "Unknown"; strings like "NA"/"N/A" remain literal.	Industry name, mapped against dim_industries

- [1] Blank stored as empty string; "NA", "N/A" not treated as missing and are stored as is. Best to keep these empty.
- [2] Numeric fields treat ["", "none", "unknown", "n/a"] as missing → 0; others not recognized.
- [3] Mapped text fields replace only nulls/NaN with "Unknown"; other tokens remain literal.
- [4] Rows with empty/null `emplinkedin` are dropped; non-empty strings like "NA" are ingested.