# Title: Dedupe effectiveness

## Use case

|  |  |  |
| --- | --- | --- |
| **Scenarios** | **An example** | **Likely duplicates that will be detected** |
| Shortened Names & Use of nicknames | Robert Andrews | Bob Andrews; Rob Andrews |
| Katherine Johnson | Cathy Johnson; Kathy Johnson |
| Short forms, punctuation, abbreviations | Freshdesk | Freshdesk Inc, Freshdesk Corp. |
| Johnson and Johnson | Johnson & Johnson |
| Change in the sequence of words | University of Manchester | Manchester University |
| Phonetic matches and verb forms | Ritchie Rich | Rishi Rich |
| Foreign Languages | Akmaral Tulegenova    Фридрих Платонов | [Акмарал Тулегенова](https://teststage2.fsales.io/leads/587453)    Fridrich Platonov |
| Multiple email addresses from the same person | [beckiemoris@freshworks.com](mailto:beckiemoris@freshworks.com) | [beckiemoris@gmail.com](mailto:beckiemoris@gmail.com) |
| Typos: Misplaced letters | Freshdesk | Freshdeks |
| Typos: Spaces within a word | Freshworks | Fresh works; Freshwork s |
| Typos: QWERTY keys | FreshWorks | FreshEorks |
| Typos: Missing Letter | FreshWorks | Freshorks |
| Phone numbers with and without codes | 987654321 | +91-987654321 0987654321  (987) 654-321 |

## Metric

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | What does it tell us ? | ES | Estate|Forest | Garden|Blossom |
| Acceptance rate (Number of merges / Number of times dedupe was called) | Given that the dedupe system is called, how many merges happened on each call on an average. Higher this ratio, it would indicate that our model is showing better suggestions on every call | ~.9% | ~25.6% | ~29.9% |

#### Trend we expect when a new set of accounts go live is there will high number of merges initially (old leads getting deduped) and it will settle down with high acceptance rate

|  |  |  |
| --- | --- | --- |
|  | ES Merges (week before) | Freddy Merges |
| Estate and Forest | 184 | 526 |
| Blossom and Garden | 263 | 526 |

### Merges over time ( Freddy-dedupe went live on 30th May )

# 

Y-axis is number of merges

|  |  |  |
| --- | --- | --- |
| Metric | What does it tell us ? | Freddy |
| Number of merges / Number of suggestions | Given that the merge event has happened for some lead ids, how many merges happened for every suggestion shown for those lead ids. Closer this ratio to 1, its says, almost every suggestion is merged which indicates our model is working well | ~92% merge |

## Graph Database

Graph database is used to store duplicates (vertex as leads, edge as similarity score). This graph can be further extended to store more meaningful relations between leads and contacts.