Match de Endereços

* Colocar tratamento da base do IBGE:
  + Como é a base original
  + Como é as bases transformadas
* Cep + tipo de logradouro (por extenso) + nome do logradouro
  + Ex: 40210904AVENIDAANITAGARIBALDI
* Casos comuns de erro:
  + Numeração:
    - 40080003AVENIDASETEDESETEMBRO
    - 40080003AVENIDA7DESETEMBRO
  + Abreviações
    - AVENIDADOUTORVITALBRASIL
    - AVENIDADRVITALBRASIL
  + Erros pequenos no CEP:
    - Um único dígito errado:
      * 41825900 AVENIDAANTONIOCARLOSMAGALHAES
      * 41825000AVENIDAANTONIOCARLOSMAGALHAES
    - Três dígitos errados:
      * 40210904AVENIDAANITAGARIBALDI
      * 40210750AVENIDAANITAGARIBALDI
  + Erros grandes no CEP
    - 41820915AVENIDATANCREDONEVES (não tem nenhum CEP nem parecido com esse)
  + Erros de digitação:
    - Colocar exemplos aqui.
  + Colocar o tipo de logradouro no campo do logradouro:
    - 41635150AVENIDADORIVALCAYMMI
    - 41635150AVENIDAAVENIDADORIVALCAYMMI
  + Colocar o nome do edifício ou condomínio:
    - 41730101AVENIDALUISVIANAFILHOWALLSTREETEMPRESARIAL
* Preparação de dados
  + Tudo maiúsculas
  + Sem pontuação (vírgula, traços, pontos).
  + Sem acentos
  + Sem espaço
  + Remover pronomes de tratamento (DOUTOR, SENHOR, CORONEL, etc).
  + Substituir erros comuns:
    - Numerais
    - Tipo de endereço dobrado
* Buscas sempre dentro do município.
* Match 1: mesmo município, correspondências Exata cep + endereço:
  + Colocar porcentagem de correspondências exatas
  + Pode ser sem ou com Pronome de tratamento.
  + Logradouros com quatro letras ou menos (excluindo o tipo) só aceitam correspondência exata.
    - Ex: Travessa TW30
* Match 1: mesmo município, correspondências Exata cep + endereço:
* Match 2: mesmo município, mesmo cep, pequena divergência de endereço.
  + Distância máxima de 5.
  + Porcentagem
* Match 3: mesmo município, mesmo endereço, pequenas divergências no CEP.
  + Distância máxima de 2
  + Porcentagem
* Match 4: mesmo município, cep + endereço sem o tipo e máximo de 10 caracteres.
  + Ex:
    - De: 41730101AVENIDALUISVIANAFILHOWALLSTREETEMPRESARIAL
    - Para: 41730101LUISVIANAF
  + Distância máxima de 2
  + Porcentagem
* Match 5: não encontrando correspondência nos casos acima, se usa a média de latlon dos endereços do município.
  + Porcentagem
* Número:
  + Casos de endereço sem número (SN):
    - Quando estiver registrado como sem número na Receita
    - Quando estiver número zero.
    - Quando o número estiver maior que o máximo do IBGE
  + Quando houver casos de sem-número, usar a média de latlon do endereço
  + Se não houver correspondência exata, usar a média dos dois números mais próximos.
* Classificação dos Matchs
  + Match 1Ex:
    - Correspondência exata dos endereços.
    - Não é SN:
      * Correspondência exata dos números OU
      * Média dos dois mais próximos.
  + Match 1SN:
    - Correspondência exata dos endereços.
    - Sem número detectado, média do latlon do endereço.
  + Match 2Ex:
    - mesmo município, mesmo cep, pequena divergência de endereço.
    - Não é SN
  + Match 2SN:
    - mesmo município, mesmo cep, pequena divergência de endereço.
    - Sem número detectado, média do latlon do endereço.
  + Match 3Ex:
    - mesmo município, mesmo endereço, pequenas divergências no CEP.
    - Não é SN
  + Match 3SN:
    - mesmo município, mesmo endereço, pequenas divergências no CEP.
    - Sem número detectado, média do latlon do endereço.
  + Match 4Ex:
    - mesmo município, cep + endereço sem o tipo e máximo de 10 caracteres.
    - Não é SN
  + Match 4SN:
    - mesmo município, cep + endereço sem o tipo e máximo de 10 caracteres.
    - Sem número detectado, média do latlon do endereço.
  + Match 5:
    - Endereço não encontrado, usar média de latlon dos endereços do município.
* Casos difíceis de resolver:
  + Mudanças de denominação de Rua
    - Às vezes a OSC declara um nome antigo
      * Rua Madre Rita
      * Rua Beata Rita Amada de Jesus
    - Às vezes o nome registrado é novo, mas não foi registrado.
      * Rua Assembleia de Deus
    - Às vezes se declara um CEP que existe no município, mas é de outra rua.
      * O que fazer então? Considerar o CEP ou a rua?
  + Possíveis casos de falso positivo
    - Nomes parecidos dentro do mesmo CEP:
      * 45256000AVENIDACENTRAL
      * 45266000ALAMEDACENTRAL
      * \*-Match 2, Caetanos-BA
* Colocar tabelas de frequência Relativa de cada match:

**Somente novos endereços:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Somente novos endereços** | | | **Todas OSC ativas** | | |
| **Correspondência Endereço:** | **Ex** | **SN** | **Total novas** | **Ex** | **SN** | **Total**  **ativas** |
| Match 1 | % | % | % | % | % | % |
| Match 2 | % | % | % | % | % | % |
| Match 3 | % | % | % | % | % | % |
| Match 4 | % | % | % | % | % | % |
| Match 5 | % | % | % | % | % | % |
| **Total** | **%** | **%** | **100%** | **%** | **%** | **100%** |

* Colocar tempo de execução da rotina