**Arquivo “histórico\_SEF” ( São procedimentos para todas cidades, ou seja, todo o arquivo)**

1. Na coluna “complemento1” realizar a limpeza para restar apenas o andar do apartamento. Será criada uma coluna com essa informação. O importante nesse arquivo não é ter o número do apartamento, mas seu andar. Dessa forma, quero ter ao final uma coluna com o número do andar, tipo “1”, “2” e não “101”, “102)....

Sugestões de filtros para se chegar a isso:

1. Primeiramente aplicar o filtro na coluna espécie para “AP”;
2. Em seguida, antes das palavras “º.Pavimento ou o.Pavimento”, pegar o número existente e já considerar como andar;
3. Antes de “º.Andar ou o.Andar”, pegar o número existente e já considerar como andar;
4. Em seguida fazer um filtro com os caracteres “ap” para os textos da coluna “complemento1”. Depois, pegar os números de 1 a 4 digitos existentes após esse resultado; Com isso, adotam-se as seguintes regras:

d.1- Para apartamentos com 4 números ou mais.

- De 1000 a 2999 – Pega-se os dois primeiros números como andar; Por exemplo, para “ap. 1204”, retorna-se o número 12 para a coluna andar

- De 3000 a 9999- Pega-se o primeiro numero como andar (pois acho muito difícil um prédio de mais de 30 andares).

d.2- Apartamentos com 3 números. Esse quase não tem erro. Pega-se o primeiro número;

d.3- Apartamentos com 2 números. Pega-se o primeiro número (considerando os que começam com 1, se começar com zero vai para opção seguinte).;

d.4- Apartamentos com 1 número – Considera-se todos como andar 1

d.5- Apartamentos sem números – Considera-se todos como andar 1

1. Os próximos procedimentos são para as colunas “Bairros” e “Logradouro”, onde se deve passar todos os caracteres para maiúsculo, sem cedilha, sem til, sem acentos, etc...
2. Outro passo importante é retirar todos os duplos espaços entre palavras.
3. Na coluna logradouro, existem vários casos em que escreveram as iniciais do tipo de logradouro. Precisamos excluir essas palavras, pois os tipos de logradouro já estão em outra coluna própria para isso. A ideia é até unir tipo de logradouro com logradouro, mas se não fizermos essa exclusão, ela ficará dobrada. Dessa forma, sugiro excluir a UNICAMENTE A PRIMEIRA palavra da coluna logradouro, quando elas forem iguais a:

RUA ; AL. ; AL ; ALAMEDA ; AV. ; AV ; AVENIDA ; AVEN. ; AVEN ; BECO ; CONJ. HAB. ; ESTRADA ; LOT. ; LOT ; RODOVIA ; TRAV. ; TRAV ; VIA; R. ; TVA ; AVN ; R ; PCA ; PCA. ; CH ; CONJ. ; CONJ ; PQ ; PQ. ; PARQ. ; PARQ ; TR. ; TR ; VL. ; VL ; ALAM. ; ALAM ; RV. ; RV ; ROD. ; ROD ; RODOV. ; RODOV ; RDV. ; RDV ; JD. ; JD ; JARD. ; JARD ; QD. ; QD

1. As seguintes abreviaturas devem ser substituídas (ATENÇÃO, pois em ambos os casos, com e sem ponto, a palavra só deve ser substituída se estiver separada por espaço, caso contrário ao fazer o comando para excluir “VER”, a palavra “VEREADOR” viraria “EADOR”. Entretanto, se algumas das situações abaixo estiver difícil, não tem muito problema, pois são poucos casos

DR. ou DR = Doutor

DRA. ou DRA = Doutora

CEL. ou CEL = Coronel

PROF. OU PROF = Professor

PROFA. OU PROFA = PROFESSORA

VER. OU VER= Vereador

DEP. OU DEP = deputado

PE. = Padre

CAP. ou CAP = CAPITAO

SGT. OU SGT = SARGENTO

PRES. OU PRESID. = PRESIDENTE

PREF. OU PREF - PREFEITO

GOV. OU GOV = GOVERNADOR

J.K ou JK ou JK. ou Juscelino K. ou “Juscelino K ou J. Kubitschek ou J Kubitschek = Juscelino Kubitschek

SRA. OU SRA = SENHORA

N.SRA OU N. SRA OU NSA. SRA. OU NSA.SRA = NOSSA SENHORA

N.S = NOSSA SENHORA

N . S = NOSSA SENHORA

RESID. OU RESID = RESIDENCIAL

COMEND. OU COMEND = COMENDADOR

COMTE. OU COMTE= COMANDANTE

EXP. OU EXPED. OU EXP OU EXPED = EXPEDICIONARIO

FARM. OU FARMAC. OU FARM OU FARMAC = FARMACEUTICO

JD. OU JD = JARDIM

MAJ. OU MAJ = MAJOR

MAL. OU MAL OU MAREC. = MARECHAL

MARIA AP. = MARIA APARECIDA

MONS. OU MONS OU MON. OU MON = MONSENHOR

1. As seguintes palavras/expressões devem ser excluídas (o ponto e virgula é para separar cada uma que relacionei):

NO. ; No. ; no. ; Nº ; nº; “ ; ‘

Tudo que estiver entre parênteses (inclusive os parênteses)

Tudo que tiver e estiver depois de C/; ESQ. ; ESQ ; ESQUINA ; /

Tudo que tiver e estiver depois de virgula ,

Tudo que tiver e estiver depois de hifen -, (mas somente se tiver um espaço logo após o hífen). Por exemplo: BR-040 fica.

Tudo que tiver e estiver depois de “LT.” ou “lt.” ou ”Lt.” ou “LOTE”

1. Na coluna logradouro, as palavras que possuem um ponto depois, existem casos em que estão sem espaço depois delas. Por exemplo, “C. Manoel de Carvalho” está escrito “C.Manoel de Carvalho”. Com isso seria bom adicionar um espaço após todos os pontos que não tiverem espaço, pois caso contrário, teríamos um duplo espaço.
2. Repetir o processo do item 5

FIM

Passo a passo cruzamento arquivo Historico\_SEF acima saneado com arquivo CEP\_Completo

Pré condição – O arquivo Historico\_SEF já deverá estar saneado por Logradouro e Bairro.

Em seguida faremos os seguintes procedimentos:

1. No arquivo HISTORICO\_SEF, concatenação entre a coluna TP\_logradouro (é a que contém rua, avenida) com a coluna logradouro.
2. Realizar join entre os dois arquivos para as variáveis CEP, Cidade (ou município), Bairro e Logradouro;
3. Os que baterem tudo são separados, enquanto os que não tiverem batido, adota-se o seguinte procedimento: Cruza-se Cidade, Bairro e Logradouro por semelhança. O que bater igualzinho, retorna-se para o arquivo Historico\_CEP, o CEP constante no arquivo CEP\_Completo.
4. Já os que não tiverem batido igualzinho, tentar-se-á realizar por semelhança procedimento ainda em estudo, mas onde:
5. Cidade deverá ser igualzinho
6. Bairro poderá ser semelhante, mas deverá ser atendido o critério anterior.
7. Rua poderá ser semelhante, mas deverá ter atendido os critérios anteriores