VET FILE DATA SCRUB

**What this does:**

The purpose of this program is to validate the RHS VET file and scrub the data if the file is error free. The program picks up the file in xls format and converts it to xlsx. In xlsx file each column is validated against concern constraints, if the file passes with zero error then its data is inserted into the database and an email is sent notifying stating file was successfully uploaded. If the file fails validation an email with the error file attached is sent.

**Input File:**

The input file is placed on “/hthjav1/RH1\_Vet/inbound/”and the name for the file to be passed by “”

The input file provide is in the xls format, the file contains 66 columns in total and all the columns are defined in the system in same order as they were provided for first time in a file. To mark and highlight the errors the file needs to be converted in xslx format and further processing is done. Each error is highlighted with different color code.

**Scheduled Process:**

NONE

**Menu Option:**

NONE

**Files Impacted:**

* The original file is not affected by the java program.
* The original file is converted to the xlsx format and validation is performed on it. If the validation fails the errors are highlighted in the xlsx file with different colors.
* If the file passes validation with zero error the data form xlsx file is inserted into “EVERELG” table in the database.

**Errors and Exceptions:**

|  |  |  |
| --- | --- | --- |
| Description of the error | Example wording | Fix |
| The SSN is invalid(blue color cell) | The SSN provided has more than 9 digits. This cell will be highlighted with blue color. | Check for the correct SSN, the SSN must be of 9 digits and without dashes. |
| Invalid format date(green color cell) | The date provided is in correct format. This cell will be highlighted with green color. | All the Dates must be in yyyy-mm-dd format. |
| Invalid gender type(orange color cell) | The gender field is empty or the notations are incorrect. This cell will be highlighted with orange color. | The Gender field cannot be left empty. The cell must contain either ‘M’ or ‘F’ as an entry. Any other entry than this two alphabets will return as an error. |
| Invalid address entry(red color cell) | The Zip Code or the State in the field is incorrect. | The Zip Code and the State field cannot be left empty. The zip code must be valid and at least 5 digit. The state in the field must also be valid and of 2 alphabets e.g. New Jersey as NJ for state and 85208-7273 or 85208 for zip code. |
| The field length is more than what is expected(gold color cell) | The cells highlighted with gold color indicate the size of the data in the field is longer that than the normal set of data. | Each data field is set to the specific size and the system cannot take input for that field if the size of provided data exceeds. E.g. the following fields are set to the corresponding size:   1. SSN, 9 2. First Name, 25 3. Last Name, 25 4. Address, 30 |

All the above errors will be encoded into the xlsx file and will be attached in an email and sent to the client for the corrections.

**Output file:**

The output file is in the xlsx format and is stored at the path described in the java program.

Also the contents of the xlsx file are inserted into the database if the file is error file the data is inserted to the CBLIB/EVERELG. To maintain the history from all previous uploads the data is also inserted in the history table @ CBLIB/EVERELGHST.

Along with the above file an email is also sent indicating the status of the file. The email might contain either of following the subject:

1. Your file submitted on "date" has successfully uploaded.
2. Your file submitted on "date" has failed to upload. The file was not found or is in an incorrect format.
3. Your file submitted on "date" has failed to upload. Please review the Excel File attached to this email, correct the errors indicated, and resubmit.

The no. 3 email will have an error file attached, this file will have the fields highlighted that failed the validation

**Miscellaneous Notes:**

The following are the color codes for the errors:

* Blue color indicates the error with the SSN
* Green color indicates the error with the Date format
* Orange Color indicates the error with the gender field
* Red color indicates the error with the Zip code or State field
* Gold color indicates the size of the data in corresponding field is greater than expected.