
Point-in-Time Count Data Cleaning Guide

Resource for Everyone Counts 2024 and Beyond – December 2024



Housing, Infrastructure
and Communities Canada

Logement, Infrastructures
et Collectivités Canada

Canada

Aussi disponible en français sous le titre : **Guide sur le nettoyage des données de dénombrement ponctuel**

Information contained in this publication or product may be reproduced, in part or in whole, and by any means, for personal or public non-commercial purposes without charge or further permission, unless otherwise specified. Commercial reproduction and distribution are prohibited except with written permission from Infrastructure Canada.

For more information, contact:

Housing, Infrastructure and Communities Canada

180 Kent Street, Suite 1100

Ottawa, Ontario K1P 0B6

info@inf.gc.ca

© His Majesty the King in Right of Canada, as represented by the Minister of Housing, Infrastructure and Communities, 2025.

Cat. No. T94-83/2025E-PDF

ISBN 978-0-660-75567-0

Contents

Introduction.....	3
For non-HIFIS Users.....	3
For HIFIS Users.....	3
Reviewing Completed and Incomplete Surveys.....	4
Reviewing Screening Questions and Survey Eligibility.....	4
Addressing Duplicate Observations.....	4
Creating Unique Identifiers.....	5
Overnight Locations and Permanent Residence.....	5
Active and Recent Shelter Use.....	7
Sociodemographic Variables.....	7
Age.....	7
Connecting Families.....	8
Gender, Sexual Orientation and 2SLGBTQI+ Identity.....	10
Newcomer Experience.....	10
Migration and Time in Community.....	11
Indigenous Status and Racial Identity.....	11
Veteran Status.....	12
Questions with Multiple Response Options.....	12
Annex A. Special Considerations for HIFIS Instances.....	14

Introduction

For non-HIFIS Users

This document aims to guide Community Entities (CE) through the data cleaning process to prepare PiT survey data for reporting to Housing, Infrastructure and Communities Canada. The procedures and suggested formulas outlined in this document are intended for use in Excel.

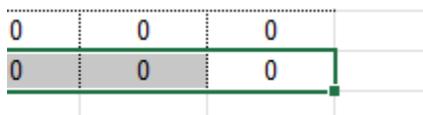
The accompanying Excel file, named “HPD-Template-PiT-Core_OptionalQuestionsDataTemplate-202411.xlsx”, is intended to guide you during the processing of cleaning and submitting your PiT survey data to HICC if you are not using HIFIS to do so. If you are using HIFIS 4 or HIFIS Lite to enter your survey data (either live data entry and/or after the fact), you do not need to use this template and you can export your data directly through the HIFIS PiT module. It can be accessed in the PiT group on GCcollab, under Files > Phase 4: Post-Count, or by downloading the attached file.

The first sheet in the Excel file provides a sample template for reference, with rows exemplifying survey respondents and the columns corresponding to the questions for which a response is required, based on family role.

The second sheet provides a data dictionary with the value labels and codes, as well as variable descriptions.

The third sheet is a blank template for you to use to enter survey data. Core and optional questions are included in this template, though it can be modified to add/remove columns based on the structure of the questionnaire your community is using. Response options are provided as a drop-down list.

The fourth sheet uses formulas that reference the data dictionary to auto-populate the survey dataset with the value codes linked to the responses given, which can facilitate data analysis. To add rows to this sheet (representing additional survey respondents or dependents), select cells A500:EK500, then left-click the green square at the lower right-hand side of the selection (pictured below), and drag down to fill the desired number of rows.



If you are using paper surveys during your PiT Count and do not intend to use HIFIS 4 or HIFIS Lite to enter your data after the fact, you can enter responses directly into the Excel template.

When you save your PiT survey data in the template provided, the default file format is an Excel Workbook (.xlsx). If you are using a statistical software program such as SPSS or Stata to conduct your analyses, you may wish to save your dataset as a CSV Comma delimited file (.csv).

For HIFIS Users

If you are using HIFIS to enter and submit your PiT Count survey data, you do not need to use the Excel template to submit survey data. You may wish to review the data dictionary on the second sheet in this file, or you can review the full [data dictionary for all HIFIS export files](#) from the PiT module. Additionally, please see [Annex A](#) for information on specific data cleaning considerations depending on the version of HIFIS you are using.

Reviewing Completed and Incomplete Surveys

This would be the first step in cleaning the survey data that is collected for the PiT Counts. It investigates cases when surveys are abandoned by the respondents or when it is not possible for a surveyor to complete a survey. In general, the data fields are defined by the term “Unclear/Blank Response” in the data. When creating computed variables, these entries should be excluded from the calculation.

“Unclear/Blank Responses” can be reflective of when the surveyor skipped a question or if a survey ended early. These differ from “Decline to Answer” and “Don’t Know” responses, which can be analyzed post-Count to assess the response rate of the survey question and identify opportunities for improving survey design. They can also highlight if a significant proportion of survey respondents avoided a particular question, signaling a need for revision or removal of that question.

Reviewing Screening Questions and Survey Eligibility

For the optional screening questions, when a respondent indicates they are staying in one of the following locations - someone else’s house, hotel/motel self-funded, hospital, treatment centre, jail, prison, remand centre - and they indicate that they also have a safe and permanent residence, this would indicate that the respondent should have been screened out of the survey. In such cases, these specific observations should be removed from the survey dataset, as the respondents are not considered to be experiencing homelessness at the time of the Count. This can be done by filtering for “Yes” responses in the “c1_permanent residence” variable. The data that are shown following this filtering should be deleted from the survey dataset and moved to another file with a descriptive name (e.g., 2024 PiT Data deleted cases).

Addressing Duplicate Observations

Measures that can be taken to avoid duplication begin during the training and implementation phases of the PiT Count. First, one of the screening questions (“Have you answered a survey with a person with this (identifier)?”) is intended to screen out people who have already responded. If your community uses a longer survey period, however, some potential respondents may forget that they have already responded or have other reasons to want to participate again. To further identify and reduce duplicates, we recommend that you create a unique identifier for each respondent, which is explained in greater detail in the following section.

Finally, you can check your survey dataset for overly similar observations (e.g., two individuals with the same age, gender, racial identity, and age of first homeless experience) to flag any other suspected duplicates. Once suspected duplicates are flagged, they can be reviewed in more detail to validate whether they are in fact duplicates and to decide which observations should be excluded from the analysis.

Consider the survey dates/times and completeness of the survey responses when deciding which respondent (row) to keep. For example, if a survey was abandoned very early and the same individual completed the full survey later within the survey period, it would be better to keep the most recent and complete response. Do not combine answers from separate observations suspected to be collected from the same survey respondent.

Creating Unique Identifiers

Unique identifiers reduce the risk of encountering data quality issues, including when de-duplicating observations to ensure that each person's survey responses are recorded only once. Unique identifiers can be randomly generated, or non-random: an example of a unique identifier is the first two letters of the respondent's first name, the first two letters of their last name, and the day of their birth (e.g. for John Smith, born April 1, it would be JOSM0401). This helps to flag potential duplicate observations using basic personal information.

Individual privacy is of the utmost importance. If your community is collecting names through its PiT Count (e.g., to create or update a Unique Identifier List or By-Name List), please ensure that the data submitted to the Government of Canada is anonymized, such that no individual person is identifiable.

Overnight Locations and Permanent Residence

In screening question C, respondents are asked where they will be staying on the night of the count, whose responses are captured in columns "c_overnight location" and "c_overnight other location" in the dataset. In order to ensure that the proper locations are captured, it is recommended to review the write-in responses in the "c_overnight other location" variable and determine if they can be re-categorized into the pre-existing fields in the "c_overnight location" variable. This can be done through the following steps:

1. Sort the data in ascending order by "c_overnight other location".
2. Review the list of text input responses for exact or approximate matches to one of the pre-existing fields (see responses for screening question C in the data dictionary). Review the list also for written responses that may be logically categorized as this overnight location.
3. Repeat this process for each of the overnight location response options.
4. Once filtered, copy the response to the "c_overnight location" column.
5. It is important not to delete the write-in response, in order to conserve the original information from the raw data.
6. Clear filters.

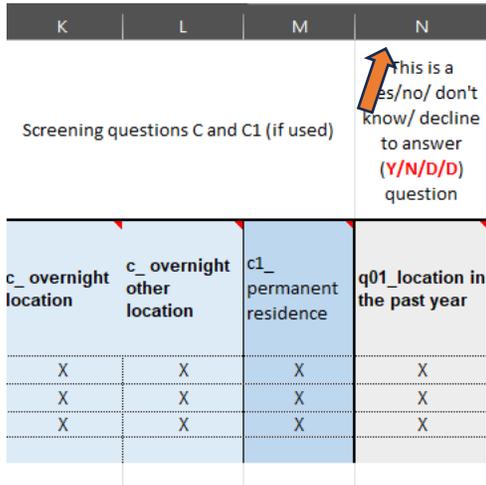
Standardized Location Variable

The responses from this question are also used to create the location variable which has standardized categories, namely- sheltered, unsheltered, transitional housing, hotels/motels, systems, hidden homelessness. The steps to create the variable are as follows:

1. On a separate worksheet, create a list of the standardized responses (see below)

Location			
Sheltered			
Unsheltered			
Encampments			
Transitional Housing			
Hotels/Motels			
Systems			
Hidden			

2. Create a new variable adjacent to that of the “c1_permanent residence” variable by inserting a new column by right-clicking first at the column letter at the top (see picture below) and then selecting “Insert”.



3. Rename the column to “Location”.
4. Select all cells below the variable name (shortcut keys: CTRL+SHIFT+Down Arrow).
5. Format these cells as a drop down menu, by going to Data>Data Validation>List.
6. For the source, select the list of standard responses that were created in step 1, and then click “OK”.
7. The table below can be used to determine how to standardize the responses from the “c_overnight location” variable, by filtering for the responses and choosing the standard response in the “Location” variable.

Response in “c_overnight location”	Standardized response in “Location”
HOMELESS SHELTER (e.g. emergency, family or domestic violence shelter, warming centre, drop-in)	Sheltered
UNSHeltered IN A PUBLIC SPACE (e.g. street, park, bus shelter, forest, or abandoned building)	Unsheltered
VEHICLE (e.g. car, van, recreational vehicle (RV), truck, boat)	
ENCAMPMENT (e.g. group of tents, makeshift shelters, or other long-term outdoor settlement)	Encampments
TRANSITIONAL SHELTER/HOUSING	Transitional Housing
HOTEL/MOTEL FUNDED BY CITY OR HOMELESS PROGRAM	Hotel/Motel
HOSPITAL	Systems
TREATMENT CENTRE	
JAIL, PRISON, REMAND CENTRE	
SOMEONE ELSE’S PLACE	Hidden Homelessness

Active and Recent Shelter Use

This section connects cases when a respondent indicates staying in an emergency shelter on the day of the Count and the response given for emergency shelter use. In these cases, the respondent are automatically considered to have spent one night in an emergency shelter in the 12 months.

When the overnight location on the day of the count is in an Emergency shelter, this should be reflected on the recent shelter use question as well. In other words, they are automatically considered to have spent one night in an emergency shelter in the past 12 months. To ensure this, the following steps can be followed:

1. Filter “c_overnight location” for the response “Homeless Shelter (Emergency, Family or Domestic Violence Shelter)”.
2. Confirm that all the responses for variable “q01_location in the past year” are “Yes”.
 - a. If there are other responses here, change to “Yes”.
3. Confirm that there is a “1” in the “q01_homeless shelter” variable.
 - a. If blank or “0”, change response to “1”.
4. Clear all filters.

It is recommended to repeat the above steps for all locations (unsheltered, encampments, transitional, etc.), so that the respondent’s location on the night of the count is reflected accordingly.

Sociodemographic Variables

Age

In question 3, survey respondents are asked to indicate their age in years. For the purposes of analysis, respondents are grouped according to the following mutually-exclusive age ranges: children (aged 0-12), youth (aged 13-24), adults (aged 25-49), older adults (aged 50-64), and seniors (aged 65+). To create this categorical variable, you can add a column with a formula of nested IF statements in the Excel template or you can compute a new variable if you are using a statistical software such as SPSS or Stata.

If you choose to generate the age range variable in sheet 3 or 4 of the Excel template (with a column name of AgeRange, for example), this formula might look something like the following:

```
=IF(AO3="", "", IF(AO3<13, "Children (0-12)", IF(AO3<25, "Youth (13-24)", IF(AO3<50, "Adults (25-49)", IF(AO3<65, "Older Adults (50-64)", IF(AO3>=65, "Seniors (65+)"))))))
```

If you are using SPSS, sample syntax to compute this variable might look like the following:

```
IF (q03aageyears>64) AgeRange=5.  
IF (q03aageyears<65) AgeRange=4.  
IF (q03aageyears<50) AgeRange=3.  
IF (q03aageyears<25) AgeRange=2.  
IF (q03aageyears<13) AgeRange=1.  
EXECUTE.
```

```
VALUE LABELS AgeRange  
1 'Children (0-12)'  
2 'Youth (13-24)'
```

3 'Adults (25-49)'
4 'Older Adults (50-64)'
5 'Seniors (65+)'.
EXECUTE.

In order to validate that the age ranges are inclusive of all responses, mutually exclusive, and aligned with the age ranges specified for reporting, you can create a pivot table in Excel (or a crosstabulation in SPSS or other statistical software program) to compare the q03aageyears and AgeRange.

Connecting Families

*Relevant columns in the sample template for reference (sheet 1): **C** (survey number), **D** (Family Head survey number), **J** (family role), **AI:AN** (binary indicators for types of family members reported by the respondent), and **AP** (age, applicable for dependent children).*

In question 2 of the survey, respondents are asked if anyone else is staying with them tonight. The response options include: partner, child(ren), pet(s), or other adult (which can include other family or friends).

People commonly experience homelessness as a single individual. In this case, the code 1111 is used in the Family Head survey number field (column **D**), to indicate that this field is not applicable. The family role in column **J** should indicate "Single". In column **AI**, the main response to the question about family, the field for this survey should indicate "None", and columns **AJ:AN** for this survey should be equal to 0. This applies to every respondent who answers the survey on their own; unaccompanied youth aged 13 and up, if surveyed, would have their data entered in this manner.

When multiple people are staying together, they are considered to be experiencing family homelessness. In this case, column **AI** should indicate "Yes (answered)". Family makeup can be further disaggregated into the following three categories: single parents/single adults with children, couples/multiple adults with children, and couples/multiple adults without children.

When surveying individuals who are experiencing homelessness as part of a family, each adult is surveyed separately. The first person to be surveyed will be considered the Family Head for the purposes of linking surveys to each other. When subsequent adults are surveyed, you can record the Family Head survey number on the adult family members' questionnaires.

During data entry, the survey number of the Family Head should be indicated under column **D** for all family members, including the Family Head themselves (i.e., the first adult surveyed in a family unit should have identical inputs in column **C** and **D** of the template). The Family Head indicator in column **AJ** of the template should equal 1 for the Family Head (i.e., the first adult surveyed within the family), and should equal 0 for all other family members. The count of Family Heads should reflect the number of family units experiencing homelessness on the reference night of the survey. However, due to challenges accurately collecting data on family composition, this is not a reliable count of families experiencing homelessness and should not be reported as such.

The field for family role in column **J** should indicate "Family Head" for the first adult family member surveyed. For subsequent respondents within the family, column **J** should indicate "Partner" or "Other adult" as the family role.

The indicators for family type can be generated with the columns **AK:AN** in the template, which are equal to 1 if the respondent reports that they are staying with a partner, other adult, kid(s), and/or pet(s). A given field should be equal to 0 if the respondent does not indicate staying with anyone in the family role corresponding to that column. For all adult members of a family, the input for columns **AK:AN** should be identical.

An important note to avoid duplicate observations in families of multiple adults with children. If multiple parents/guardians within a family unit reported their children and their ages, it is important to ensure only one row is added for each child. However, it is acceptable to indicate in the template that each adult has children in column **AM**.

In 2024, pets were incorporated into the family question, to better reflect the diverse living situations of people experiencing homelessness with companion animals, who can be considered chosen family. Pet(s) are reflected as a factor of the respondent's living situation in column **AN** of the template. Unlike dependent children, pets are not to be recorded as separate observations in new rows.

If multiple adults within a family report having pet(s), it is acceptable to indicate in the template that each adult has pet(s) in column **AN**. If a single individual reports having pet(s), they are still considered to be a single individual and should not be categorized as a Family Head, as there is no need to link surveys or create new rows in this case. In the rows that record single survey respondents with pets, column **D** = 1111, column **AI** = "Yes (Answered)", columns **AJ:AM** = 0, and column **AN** = 1.

Your community decided the minimum age to be surveyed. Children under the minimum survey age accompanied by an adult should be recorded as separate respondents, with completed fields for their age (column **AP** in the template) and the survey number of the Family Head (column **D**). The field for family role in Column **J** should indicate "Child / Dependent". Columns **AI:AN** remain blank for the new rows generated to record dependent children.

If dependent children exceed the minimum age required to submit a survey, they will have been recorded by their parent or guardian as an accompanied youth (i.e., a new row will have been created with Family role = "Child / Dependent", but age range = "Youth (13-24)"). If these individuals consented to be surveyed, this would result in duplicate observations for accompanied youth. For analyses disaggregated by age range of the survey respondent, dependent youth should be excluded to avoid double-counting.

In order to review family data, you can create a new column with a filter variable that equals 1 where Survey number = Family Head survey number, then create crosstabulations between the new column and columns **J** and **AI**. Where **J** = "Family Head", **AI** = 1, and survey number = Family Head survey number.

Crosstabulations of column **D** and **AP** can be used to check for duplicate rows observing dependent children. If multiple children of the same age appear within a family, refer to the survey response(s) of the adult(s) in the family to check whether this is a duplication error.

Gender, Sexual Orientation and 2SLGBTQI+ Identity

Gender Variable

The “Gender” variable can be created using the responses in “q12_gender identity”, where the responses could be categorized into Woman, Man or Gender Diverse. It would also be recommended to format the variable as a list, such as the “Location” variable (see section on Overnight location on steps to create list). The responses from “q12_gender identity” can be categorized based on the following:

Responses from q12_gender identity	Gender Variable
Man	Man
Woman	Woman
Two-Spirit, Trans Woman, Trans Man, Non-Binary (Genderqueer), Not Listed	Gender Diverse

2SLGBTQI+ Variable

A “2SLGBTQI+” variable can be created for analysis using a combination of responses from the Gender (q12_gender identity) and Sexual Identity (q13_sexual orientation) question. The “2SLGBTQI+” variable is determined based on the following:

- Respondents who indicate that their gender identity is Non-binary, Trans man/woman, Two-Spirit or not listed would be categorized as 2SLGBTQI+. (In this case, compute 2SLGBTQI+ variable = 1).
- Respondents who identify as having a sexual orientation as Gay, Lesbian, Bisexual, Two-Spirit, Pansexual, Asexual, Queer, Questioning, Not Listed would be categorized as being 2SLGBTQI+. (In this case, compute 2SLGBTQI+ variable = 1).
- Respondents who indicate a gender identity of Man or Woman and a sexual orientation of Heterosexual are categorized as non-2SLGBTQI+. (In this case, compute 2SLGBTQI+ variable = 0.)
- Respondents do not need to respond to both questions in order to meet the criteria for 2SLGBTQI+ = 1. However, it is necessary that both questions are completed in order to meet the criteria for 2SLGBTQI+ = 0. In cases where responses to both questions are missing, you can leave the 2SLGBTQI+ variable as a missing value.

Newcomer Experience

In question 6, survey respondents are asked if they had come to Canada as a newcomer, and also when they had arrived in Canada.

Respondents who have indicated "No" are categorized as people without a newcomer experience coming to Canada. The term "newcomer" applies to respondents who have identified as having the experience of arriving in Canada under one of the following statuses:

- Immigrant;
- Refugee;
- Asylum claimant;
- Temporary foreign worker;
- Other work permit;
- Student/Study permit;
- Temporary resident;
- Other.

For the purposes of your local analysis, you may wish to group some or all of the latter five statuses together in a single category.

For Newcomers, the response for the follow-up question for question 6, can also be used to inform on the following questions. If the respondent chose “Always been here” in question 7, this response can be replaced by the duration that was indicated by the respondent from question 6. This is because respondents may interpret “Always been here” to mean that they have been in the community since arriving in Canada.

Migration and Time in Community

This question asks respondents whether they had moved from another community and also how long they have been in the current community where the PiT Count is taking place.

In cases when the respondent answered “Always been here”, but then provides the name of the previous city, the approach for cleaning the data would be the following:

1. Ensure “Answered” is marked for variable “q07_answered name of previous community”.
2. Replace the “Always been here” response and mark as “Unclear/Blank Response” for variable “q07_answered duration in community”.

If possible, it is also recommended to go through all the responses provided for the name of previous communities, and ensure that the Country and Province fields are entered as well. This will help when conducting analyses looking at interprovincial migrations, migrations within a province, as well as migrations into Canada from another country.

Indigenous Status and Racial Identity

Question 8 and 8b asks respondents if they identify as First Nations, Métis, or Inuit, and for their racial identity. For the first part of the question, Indigenous Identity, the respondent can only select one of the three Indigenous Identities (First Nations, Métis or Inuit), or they can indicate that they have Indigenous ancestry or are not Indigenous. For the second part of the question, belonging to a racialized group, the respondent can choose multiple responses. The following are considerations for cleaning the data related to the Racial Identity question.

1. In cases when an Indigenous Identity is chosen, but the Racial Identity question is unanswered. The blank response for Racial Identity should be changed to “Indigenous Only”.
2. If a respondent indicates having Indigenous Ancestry, and left the Racial Identity question unanswered, the response to question 8b should remain as an “Unclear/Blank Response”.
3. If possible, it is recommended to review the responses that are set as “Not listed”, which includes the write-in responses, to check if it can be categorized into existing response options. This can be done by searching the responses in the “q08b_Not listed variable” for exact or approximate/logical matches to each of the pre-existing racial identities. The re-categorized responses should not be deleted.

Similar to the Canadian Census, Indigenous identity and racial identity are assessed through two separate questions. In alignment with the Employment Equity Act, respondents who identify as Indigenous are categorized as “Not a visible minority”. However, when representing all population groups, they are included in a distinct category labeled “Indigenous peoples”.

Veteran Status

In question 9, survey respondents are asked if they ever served in the Canadian Military or RCMP. The response options are: “Yes, Military”, “Yes, RCMP”, “Both Military and RCMP”, “No”, “Don’t Know”, and “Decline to Answer”.

For the purposes of analysis, the first three responses are aggregated into a binary variable where 1 = Veteran and 0 = Non-veteran. To create this binary variable, you can add a column with a formula of nested IF statements in the Excel template or you can compute a new variable if you are using a statistical software such as SPSS or Stata.

If you choose to generate the veteran status variable in sheet 3 or 4 of the Excel template (with a column name of VeteranBinary, for example), this formula might look something like the following:

```
=IF(BX3="Unclear / Blank Response","",IF(BX3="","", IF(BX3="Yes, Military",1,IF(BX3="Yes, RCMP",1,IF(BX3="Both Military and RCMP",1,IF(BX3="No",0,IF(BX3="Don't Know",3,IF(BX3="Decline to Answer",3))))))))))
```

If you are using SPSS and using the value lookup codes provided in the data dictionary, sample syntax to compute this variable might look like the following:

```
IF (q09avetrn=83501) VeteranBinary=1.  
IF (q09avetrn=83502) VeteranBinary=1.  
IF (q09avetrn=83503) VeteranBinary=1.  
IF (q09avetrn=83551) VeteranBinary=0.  
IF (q09avetrn=83581) VeteranBinary=3.  
IF (q09avetrn=83582) VeteranBinary=3.  
EXECUTE.
```

```
VALUE LABELS VeteranBinary  
0 'Non-veteran'  
1 'Veteran'  
3 'Don't know/Decline to answer'  
EXECUTE.
```

In order to validate that the binary veteran variable is inclusive of all responses, mutually exclusive, and aligned with the categories specified for reporting, you can create a pivot table in Excel (or a crosstabulation in SPSS or other statistical software program) to compare the q09avetrn and VeteranBinary.

Questions with Multiple Response Options

These data cleaning considerations apply to survey questions that permit multiple responses, namely: sources of income, reasons for housing loss, racial identity, and overnight locations in the past year. For these questions, there is one column which indicates whether that question was answered and a binary column corresponding to each response option, as well as a column corresponding to the text input field if the response is not listed.

1. If sum of the binary columns ≥ 1 for a given multiple-response question, then that question should be marked as “Answered”.
2. If possible, it is recommended to review the responses that are set as “Not listed”, which includes the write-in responses, to check if it can be categorized into existing response options. This can be done by searching the responses in the Other/Not listed column of that question for exact or approximate/logical matches to each of the pre-existing response options. The re-categorized responses should not be deleted. The steps to re-categorize “Not listed” responses are similarly discussed in the sections on Active and Recent Shelter Use and Indigenous Status and Racial Identity.

Annex A. Special Considerations for HIFIS Instances

If your community is using HIFIS 4, you may encounter some additional data cleaning issues to address, depending on the version of your HIFIS instance. If you are using HIFIS Lite, these considerations aren't necessary, as bug fixes have already been applied.

The following list represents the obsolete columns included in the data export, for which the solution is simply to drop these variables by deleting the obsolete columns, or to disregard these columns.

- q01afamily
- q02aansage
- q02aageyrs
- q03aanshaa
- q03ahlsage
- q05asheltr
- q11dmental
- q11esubstc
- q14bcovid
- q04banseps
- q04bepisod
- q04cptadd
- q05ansnst
- q05bansnst
- q11fbrain

The table below provides more detailed information on the issues in the PIT Count module data export for each version of HIFIS, and data cleaning considerations to respond to these issues.

HIFIS Version	Issues	Data Cleaning Considerations / Steps to Resolve in Excel
4.0.60.3	<p>Language selection - In Live Surveys – Scrolling mode, toggling between English and French clears numerical fields.</p> <p>Optional Question 1b - Users are unable to select an option for the optional question on reasons for shelter avoidance.</p> <p>Optional Question 8a - Selecting an Indigenous community of origin clears other responses to Q8.</p>	<p>Language selection - If using Live Surveys – Scrolling mode, select the language of the survey (depending on the language needs of the respondent and the surveyor) prior to entering responses.</p> <p>Optional Question 1b - Analysis of reasons for shelter avoidance is not possible in this instance, unless included as a local question.</p> <p>Optional Question 8a - Avoiding use of the optional question on Indigenous community of origin circumvents this issue.</p>

	<p>Question 13 (q13asexorn) - Order of response options is not aligned with paper survey.</p>	<p>Question 13 (q13asexorn) - Carefully checking response selection will circumvent issues in this case.</p>
4.0.60.4 & 4.0.60.4.1	<p>Question 1b (q01b_83772) - When Other (Not in list) is selected, the text input is not saved in the export.</p> <p>Question 2 (familyrel) - When Pet(s) is selected, this does not appear in the export.</p> <p>Questions 3, 5, 6, 7 (q03aageyrs, q05ahlsage, q06birdays, q07abcdays) - In Live Surveys – Paging mode, navigating between questions or editing after saving may cause the “Age” and “Duration” fields to reset to zero.</p> <p>Questions 14, 15 - Custom responses for Reasons for Housing Loss and Sources of Income are not reflected in PiTLnkMult.csv.</p>	<p>Question 1b (q01b_83772) - If q01b_83772=1, this means “Other (not in list)” was selected. If you wish to review the text input responses specifying custom reasons for avoiding shelter, you may filter on this variable and review the associated surveys (by survey number).</p> <p>Question 2 (familyrel) - Analysis of Pet(s) data is not possible in this instance, unless included as a local question.</p> <p>Questions 3, 5, 6, 7 (q03aageyrs, q05ahlsage, q06birdays, q07abcdays) – Using Live Surveys - Scrolling mode or entering data after the fact circumvents this issue. Otherwise, this should be treated as a missing value.</p> <p>Questions 14, 15 - Analysis of custom responses to Reasons for Housing Loss and Sources of Income question is not possible in this instance.</p>
4.0.60.4.2	<p>Question 2 (familyrel) - Partners / Other Adults may be erroneously listed as Pet(s), if the family composition includes pets. Correct family role is listed</p>	<p>Question 2 (familyrel) -</p> <p>1. Flag dependent rows generated for adults Insert a column named partnercheck in column J with the formula: =IF(AND(ISNUMBER(SEARCH("*>*",B2)),OR(I2="Partner",I2="Other adult (can include other family or friends)")), "Incorrect", "Correct")</p>

	<p>in a new row linked to survey number of family head. Single individuals with pets are also listed as a family head.</p> <p><i>Question 8 (q08cansabc) - Answered / Don't Know / Decline to Answer not appearing in data export.</i></p>	<p>2. Flag accurately labeled pets Insert a column named ispet in column K with the formula: =IF(AND(ISNUMBER(SEARCH("*>*",B2)),I2="Pet(s)),1,0)</p> <p>3. Flag adults labeled as pets Insert a column named petcheck in column L with the formula: =IF(AND(K2=0,I2="Pet(s)),"Incorrect","Correct")</p> <p>4. Recode family role for adults mislabeled as pets Insert a column named familyrel2 in column M with the formula: =IF(L2="Incorrect",INDEX(1:I,MATCH(C2&L2,C:C&J:J,0)),I2)</p> <p>5. Flag single adults with pets Insert a column named singlewpet in column N with the formula: =IF(COUNTIF(C:C,C2)>(COUNTIFS(C:C,C2,M:M,"="&"Pet(s)"+1),0,1)</p> <p>6. Recode family role for single adults with pets Insert a column named familyrel3 in column O with the formula: =IF(AND(N2=1,M2="Family Head"),"Single",M2)</p> <p>7. Flag duplicates Insert a column named dropif1 in column P with the formula: =IF(COUNTIFS(C\$2:C2,C2,M\$2:M2,M2)>1,1,"")</p> <p><i>Question 8 (q08cansabc) -</i> 1. Recode q08cansabc as Answered if q08cabcomn is not blank In q08cansabc column, use formula: =IF(O2="", "", "Answered")</p>
4.0.60.4.3	No known issues	N/A