



*Teaching
Genealogists AI™*
Instructor: Sylvia
Vargas

Census Records for Puerto Rican Genealogy

Module 10 — Using AI for Census Analysis

Course Instructor: Sylvia Vargas
Organization: Puerto Rican Genealogy Group
Website: <https://puertoricangenealogy.org>
Date Created: 2026-02-27
Estimated Time: 40–50 minutes

Learning Objectives

By the end of this module, you will:

- Identify what AI tools can and cannot do in genealogical census research

- Apply practical AI workflows for extracting, comparing, and organizing census data
 - Write effective prompts for census analysis tasks
 - Recognize and guard against AI hallucination
 - Maintain GPS compliance in AI-assisted research
-

Section 1 — What AI Can and Cannot Do

What AI Tools Do Well

AI tools are most useful in census research for tasks that involve organizing, formatting, and pattern-spotting across text you provide. Specifically, they can:

- Extract structured data from transcribed census entries and format it as a table
- Compare multiple census entries for the same family and flag age discrepancies
- Calculate implied birth years from reported ages across multiple census years
- Translate Spanish-language census column headers and genealogical terminology
- Suggest alternative spellings and anglicizations for Spanish names
- Explain the historical context of census categories (race classification, occupational terms)
- Help draft a narrative summary from a data table you have already built
- Suggest logical next steps in a research strategy (with important caveats — see Section 5)

What AI Tools Cannot Do

AI tools have fundamental limitations that are especially important in genealogical work:

- **They do not search databases.** AI tools cannot search FamilySearch, Ancestry, or AGPR for records. They only work with text you provide.
 - **They cannot read images reliably.** Do not paste census images and ask AI to transcribe them without verifying every character against the original.
 - **They do not know what records exist.** An AI tool cannot tell you whether a specific record survives for a given municipality.
 - **They hallucinate.** AI tools generate plausible-sounding but sometimes completely false information — including invented collection IDs, fabricated records, and wrong historical facts. See Section 5 for a full discussion.
 - **They are not genealogical sources.** AI output cannot be cited. Every fact in your research narrative must trace back to a verifiable original or derivative source.
-

Section 2 — GPS Compliance When Using AI

Any conclusion reached with AI assistance must still meet all five elements of the Genealogical Proof Standard:

1. **Reasonably exhaustive research** — AI does not search for you. It can help you organize what you have found, but the search itself must be conducted through proper repositories.
2. **Complete and accurate citations** — AI-generated text is not a source. Cite the original census record, not the AI output.
3. **Thorough analysis of each source** — AI can help organize data, but the Three-Layer analysis (source type, information quality, evidence value) is your responsibility.
4. **Resolution of conflicting evidence** — AI can flag discrepancies, but the reasoning and judgment about how to resolve them must come from you.
5. **A soundly reasoned, coherently written conclusion** — AI can help draft language, but the logic, interpretation, and final conclusion are yours.

Core rule: AI is a data organization tool. It is not a research authority. The moment you treat AI output as a conclusion rather than as a starting point, your GPS compliance is at risk.

Section 3 — Practical AI Workflows for Census Analysis

Workflow 1: Multi-Year Data Extraction

When to use: You have transcribed census entries for the same person or family from multiple years and want to build a structured comparison table.

How to do it:

1. Transcribe the census entries (name, age, household members, barrio, etc.) from the original images
2. Paste all entries into the AI chat with a clear prompt (see Section 4)
3. Ask the AI to build a comparison table with one row per census year
4. Review every cell against your original transcriptions before using the table

Workflow 2: Age Discrepancy Analysis

When to use: You have ages from multiple census years and want to calculate implied birth years and identify inconsistencies.

How to do it:

1. Provide the name, census year, and reported age for each entry
2. Ask the AI to calculate the implied birth year for each entry and flag years where the discrepancy exceeds 3 years

3. Use the flagged discrepancies as starting points for further investigation — not as conclusions

Workflow 3: Spanish Terminology Translation

When to use: You encounter Spanish-language column headers, occupation terms, or status categories in colonial or civil records.

How to do it:

1. List the terms you need translated
2. Ask the AI to provide translations and, where relevant, the genealogical or historical context for each term
3. Verify unfamiliar historical terms against a reference source such as Elizabeth Shown Mills' *Evidence Explained* or the FamilySearch wiki

Workflow 4: Name Variation Generation

When to use: A name search has failed and you need a list of possible alternate spellings or anglicizations.

How to do it:

1. Provide the original Spanish name
2. Ask the AI to list possible anglicizations, abbreviations, and phonetic misspellings that might appear in U.S. federal census records
3. Use the resulting list to run additional searches — but verify every result against the actual record

Workflow 5: Migration Pattern Summary

When to use: You have migration-related fields from the 1940 census and want a plain-language summary.

How to do it:

1. Provide the “where living 5 years ago” field for each household member
2. Ask the AI to summarize the migration pattern and identify any household members who relocated
3. Treat the summary as an organizational aid — not as evidence

Worked Example: A Complete Census Comparison

The following example shows the full sequence for Workflow 1 (multi-year data extraction) from raw transcription through AI output through researcher action.

Step 1 — The researcher's transcriptions (from original census images):

1910 U.S. Federal Census, Rincón, Puerto Rico, Barrio Ensenada
Juan García, head, male, age 35; wife María López, age 30; children: Pedro age 8, Carmen age 5, Rosa age 2

1920 U.S. Federal Census, Rincón, Puerto Rico, Barrio Ensenada
Juan García, head, male, age 48; wife María López, age 42; children: Pedro age 18, Carmen age 15, Rosa age 12, Luis age 7

1930 U.S. Federal Census, Rincón, Puerto Rico, Barrio Partido
Juan García, head, male, age 57; wife María López, age 50; children in household: Luis age 17

Step 2 — The researcher’s prompt:

“I am doing genealogical research on a Puerto Rican family from Rincón. Below are my transcriptions of U.S. Federal Census entries for Juan García across three census years. Please build a comparison table with columns for: census year, name as recorded, age, implied birth year, barrio, and household members listed. Then flag any implied birth year that differs by more than 3 years from another entry. Use only the information I provide. If a field is missing from my data, leave it blank.”

Step 3 — The AI’s output:

Census Year	Name as Recorded	Age	Implied Birth Year	Barrio	Household Members
1910	Juan García	35	1875	Ensenada	Wife: María López (30); Pedro (8), Carmen (5), Rosa (2)
1920	Juan García	48	1872	Ensenada	Wife: María López (42); Pedro (18), Carmen (15), Rosa (12), Luis (7)
1930	Juan García	57	1873	Partido	Wife: María López (50); Luis (17)

Discrepancy flagged: The 1910 implied birth year (1875) differs from the 1920 implied birth year (1872) by 3 years — at the threshold. The 1920 and 1930 entries are consistent (1872 and 1873). The 1910 entry warrants additional investigation.

Step 4 — What the researcher does next:

The researcher notes two findings from the table:

1. *Birth year discrepancy:* The 1910 entry implies a birth year 3 years later than 1920 and 1930 suggest. The researcher will search Puerto Rico civil registration and, where accessible, baptismal records to establish a documented birth year

for Juan García. The discrepancy does not prove a different person — it reflects normal census age-reporting variation — but it must be resolved before drawing conclusions about identity.

2. *Barrio change*: Juan appears in Barrio Ensenada in 1910 and 1920, then Barrio Partido in 1930. Both barrios are within the *municipio* of Rincón. This is not evidence of migration out of Rincón, but the researcher will note it in the research log and look for property or tax records that might explain a within-municipio move.

The children’s ages are internally consistent across census years (Pedro: 8 → 18; Carmen: 5 → 15; Rosa: 2 → 12; Luis first appears at age 7 in 1920 and 17 in 1930, consistent with a birth around 1913), which strengthens confidence that these three entries describe the same household.

Section 4 — Prompt Strategies

Effective AI prompts for census analysis share three elements: context, task, and constraints.

Context: Tell the AI what it is working with.

“I am doing Puerto Rican genealogical research. I have census entries for the same individual from 1910, 1920, 1930, and 1940.”

Task: Tell the AI exactly what you want it to do.

“Build a comparison table with one row per census year. Include columns for: name as recorded, age, implied birth year, household members, and municipality/barrio.”

Constraints: Tell the AI what not to do.

“Use only the information I provide. Do not add details from outside sources. If a field is missing from my data, leave it blank.”

Example prompt (complete):

“I am doing genealogical research on a Puerto Rican family. Below are my transcriptions of census entries for Juan García from Rincón, Puerto Rico in 1910, 1920, and 1930. Please build a comparison table with columns for: census year, name as recorded, age, implied birth year, barrio, and household members listed. Use only the information I provide. If a field is blank in my data, leave it blank in the table.”

Section 5 — The Hallucination Problem

AI tools generate text by predicting what words are likely to follow other words based on patterns in training data. They do not look things up. They do not know when they are wrong. As a result, they sometimes generate plausible-sounding but false information — a problem known as hallucination.

In genealogical research, hallucination is especially dangerous because:

- Invented facts look identical to real facts in AI output
- AI tools may generate fake FamilySearch collection IDs, Ancestry record numbers, or archive finding aids
- AI tools may invent census entries, birth dates, or family relationships that do not exist in any record
- Wrong information introduced into a family tree spreads to other researchers who copy it

How to guard against hallucination:

- Never ask AI to tell you what records exist for a specific person or place — use FamilySearch, AGPR, and other repositories directly
- Never cite AI output as a source
- Verify every fact the AI returns against the original document you provided
- If the AI introduces a name, date, or fact you did not provide, investigate it — do not assume it is correct
- Treat AI output as a draft to be checked, not a finding to be published

Practical rule: If you did not put it in the prompt, and the AI added it to the output, verify it before trusting it.

Section 6 — Building an AI-Assisted Research Practice

Integrating AI tools into a GPS-compliant workflow requires clarity about where AI belongs in the research process.

Research Stage	AI Role	Your Role
Searching repositories	None — AI cannot search	You search FamilySearch, Ancestry, AGPR
Transcribing records	Verification only — AI may help format	You transcribe from original images
Organizing data	Strong — AI excels at table building	You verify every cell
Translating terminology	Helpful — but verify unfamiliar terms	You confirm against reference sources
Analyzing discrepancies	Helpful — AI can flag inconsistencies	You reason through the resolution
Writing citations		You write every citation

Research Stage	AI Role	Your Role
	None — AI cannot cite records	
Drawing conclusions	None — AI does not have genealogical judgment	You write the reasoned conclusion

Section 7 — AI Tools for Genealogical Research

Three AI tools are most commonly used in genealogical research contexts:

Tool	Access	Notes
<i>Claude (Anthropic)</i>	Free tier available; subscription for extended use	Strong at structured data tasks and following precise instructions
<i>ChatGPT (OpenAI)</i>	Free tier available; subscription for GPT-4	Widely used; good at translation and table formatting
<i>Gemini (Google)</i>	Free tier available	Useful for translation; integrates with Google Workspace

All three share the same fundamental limitation: they hallucinate, and none can search genealogical databases on your behalf. Choose based on your comfort level with each interface. For more on evaluating AI tools for genealogical work, see the [Teaching Genealogists AI](#) resources on this website.

Worksheets Included

- WS_07 — AI Census Analysis Worksheets (Comparison Template, Data Extraction Table, Correlation Summary Builder)

Reflection Questions

1. You ask an AI tool to “find all census records for my family in Rincón from 1900 to 1950 and describe what each one contains.” The AI produces a detailed response naming specific FamilySearch collections and describing what each contains about Puerto Rican families in that period. What is the fundamental problem with this output, and how would you reformulate the request to get something genuinely useful?

2. You paste your transcriptions of a family's 1910, 1920, and 1930 census entries into an AI tool and ask for a comparison table. When you review the result, you notice the table includes a household member named "Josefina," age 14 in 1920, who does not appear in any of your transcriptions. What do you do next, and what does this situation illustrate about the nature of AI output?
 3. Looking at the workflow table in this module, a colleague argues that "AI can't draw conclusions or write citations, so it's not really useful for serious genealogical research." Using specific examples from this module, explain what AI can genuinely contribute to GPS-compliant census research and where its boundaries lie.
-

Course Complete

This is the final module of *Census Records for Puerto Rican Genealogy*. You have now covered:

- The political history of Puerto Rican censuses from Spanish colonial padrones through the U.S. federal era
- The 1887 Spanish census and 1899 U.S. Military Census and their correct research use
- Household-level analysis of the five U.S. federal census years (1910–1950)
- Municipal and barrio records that fill the gaps between federal enumerations
- Race and social classification across census systems
- Migration patterns and how to track families across the island and to the mainland
- Correlating census data with civil registration records
- Handling access limitations and building alternative source strategies
- Using AI tools responsibly within a GPS-compliant research practice

Return to the [Course Overview](#) for a complete module list and downloadable worksheets.
