

Singularidades en las Curvas Poblacionales en el AMBA 1991-2022

Fernando Meseri

10 de marzo de 2024

Fernando Meseri

1. Introducción

La información estadística que brindan las proyecciones de población es un insumo principal en la implementación de políticas estatales. Las proyecciones poblaciones es común encontrarlas a nivel País o Provincia, pero resulta particularmente importante poder contar con dichas proyecciones con un mayor nivel de desagregación, municipio/departamento. En este caso se analiza en particular los municipios del AMBA, Argentina para el período 1991-2022. Las proyecciones del Instituto Nacional de estadísticas y Censos (INDEC) han estimado la población dichos municipios. El valor arrojado para el CENSO 2022 en el caso de La Matanza presenta una desviación importante respecto al error promedio encontrado entre las proyecciones y el valor relevando en el CENSO 2022. A partir del análisis de datos censales y variables indirectas se analiza las curvas poblacionales, el error en las proyecciones INDEC respecto a lo arrojado en el CENSO 2022 así como el caso puntual de La Matanza.

2. Revisión bibliográfica

2.1. Marco Conceptual

La información estadística que brindan las proyecciones de población constituye una herramienta fundamental para la planificación de políticas públicas de corto, mediano y largo plazo. Permite estimar demanda potencial de bienes y servicios en distintas áreas como Salud, Educación, entre otras - Instituto Nacional de Estadísticas y Censos (INDEC,2013) [1]. El estado puede de esta forma determinar los recursos presupuestarios necesarios para satisfacer estas demandas. En la provincia de Buenos Aires ciertos aspectos del presupuesto son asignados en base a la población de cada municipio. Es necesario entonces contar con la información en un alto nivel de desagregación espacial(municipios).

2.2. Marco Teórico

La elaboración de proyecciones de población es una tarea compleja que debe ser realizada a través de un análisis exhaustivo que permita considerar los censos anteriores como también registros vitales y estimaciones de migración. (INDEC,2013) [1]. En general se ha utilizado en método de las componentes para elaborar dichas proyecciones. Mas esta metodología no ha podido ser replicada al nivel de las jurisdicciones más elementales, departamentos, por cuanto la información no es suficientemente confiable y la inestabilidad de la migración interna no admite formulación de hipótesis a mediano plazo. (Álvarez,2001) [2]. Una forma de realizar estas predicciones ha sido mediante métodos matemáticos de extrapolación en base a la información censal previa. (Álvarez,2001) [2]. El INDEC provee proyecciones de población por departamento para el período 2010-2025(INDEC, 2015) [3], particularmente para todos los municipios del AMBA. Se destaca que el crecimiento de la población en Argentina observado entre 2001 y 2010 a nivel departamental pone en evidencia las diferencias geográficas que existen en la dinámica poblaciones, con un comportamiento heterogéneo.

2.3. Estado del arte

Históricamente se observa conceso en la utilización del Método de las Componentes para la determinación de proyecciones poblacionales a nivel País o Provincia. El mismo contempla el crecimiento poblacional intercensal y proyecta cada una de las variables determinantes de forma independiente -fecundidad, mortalidad y migración (Álvarez,2001) [2]. Ciertamente la Serie de Análisis Demográfico de INDEC utiliza este método para la proyecciones Nacionales y Provinciales (INDEC, 2015) [3]. Asimismo, para población de países desarrollados, también se ha utilizado el modelo

de regresión logística en este tipo de predicciones (Gupta, Bhattacharya,Chattyopadhyay ,2012) [4]. Pero este modelo tiene ciertas limitaciones cuando se aplica a data censal dispersa en el tiempo, especialmente para países en desarrollo. Generalmente en estos casos las tasas de crecimiento relativo presentan tendencias inusuales, distintas a la tendencia decreciente de la regresión logística. Gupta et al., (2012) proponen modelos simplificados y variantes de Tsoularis and Wallace Model (TWM) que han proporcionado mejores resultados. A mayor nivel de desagregación se trabaja con métodos alternativos, como puede ser extrapolación matemática, Ratio- Correlation Method, Housing Unit Method, entre otros (Hoque, 2012)[5]. Por otra parte el centro Latinoamericano y Caribeño de Demografía (CELADE) ha promovido la utilización de otras técnicas para mejorar las estimaciones poblaciones derivadas de la extrapolación matemática. Se utiliza la metodología de variables sintomáticas, que permite establecer correlaciones a las tendencias poblacionales con información de variables indirectamente asociadas al fenómeno de crecimiento poblacional, a saber: nacimientos y defunciones, matrícula escolar, permisos de construcción, otros (Álvarez,2001)[2]. En lo que respecta a técnicas propias de ciencias de datos para análisis de información censal, se desatacan los siguientes usos: la utilización de data mining para búsqueda de patrones en la información censal, predicciones y forecasting utilizando modelos ARIMA e inducción con árboles de decisión (Chawda, Rane, Giri, 2018) [6]. También se destaca el uso de árboles regresión y clasificación para el agrupamiento o clustering en distintas clases, tomando como input información censal.

3. Definición del problema

El municipio de La Matanza una singularidad en su curva de crecimiento poblacional, tanto número de habitantes como tasas intercensales, respecto a los municipios aledaños del AMBA para el período 1991-2022.-

4. Justificación del estudio

La información estadística que brindan las proyecciones de población constituye una herramienta fundamental para la planificación de políticas públicas de corto, mediano y largo plazo. Permite estimar demanda potencial de bienes y servicios en distintas áreas como Salud, Educación, entre otras - Instituto Nacional de Estadísticas y Censos (INDEC,2013) [1]. Las proyecciones del INDEC en la serie demográfica N °38(INDEC, 2015) [3] han estimado la población para los municipios del AMBA. El valor arrojado para el censo 2022 en el caso de La Matanza presenta una desviación importante respecto al error promedio encontrado para el resto de los municipios. Es por esto que se pretende analizar la singularidad en la curva poblacional de La Matanza.

5. Alcances del trabajo y limitaciones

El alcance del trabajo es básicamente el análisis y estudio de datos censales. Comprende principalmente la utilización de datos censales del AMBA para el período 1991-2022. Analizar las curvas poblacionales, su comparación y el error en las proyecciones INDEC respecto a lo arrojado en el CENSO 2022 (INDEC ,2022) [7]. El trabajo se limita a demostrar la singularidad o no en el dato poblacional de la Matanza, sin intentar explicar las causas del hecho, particularmente en lo que se refiere al fenómeno demográfico que pudiese estar detrás de esta singularidad.

6. Hipótesis

Es posible demostrar la curva de crecimiento poblacional de la Matanza presenta una singularidad respecto a los municipios aledaños (AMBA) en situaciones socio-demográficas similares para el período 1991-2022.

Preguntas

1. ¿Se puede estimar una tasa de crecimiento promedio de la población urbana/suburbana en base a los 4 Censos anteriores?
2. ¿Se puede individualizar tasas de crecimiento distintas por municipio??
3. ¿Considerando las curvas de crecimiento reales de los municipios y las proyecciones realizadas, dado un intervalo de confianza se puede realmente afirmar que la Matanza presenta una singularidad?
4. ¿Es esperable que municipios aledaños crezcan poblacionalmente en el mismo orden de magnitud?

5. La apertura de los datos por edad, sexo y nivel de habitantes en el hogar muestra cierto comportamiento esperable, algún patrón. Se asemeja a los valores encontrado para La Matanza para Censo 2022.-

Respuestas

1. Sí, es posible determinar la tasa de crecimiento en función a los censos anteriores, utilizando estadística descriptiva.
2. Sí, es posible, ya que el censo incluye la desagregación por departamento/municipio. Luego estos se pueden agrupar en sectores de interés, ej. AMBA.
3. Si, en principio. Las causas de este fenómeno no son de interés del presente trabajo. El fenómeno demográfico es complejo, multicausal. Aun así, se pueden establecer valores de referencia acordes a la región, donde las características socio-demográficas son similar (conurbano).
4. Sí, municipios urbanos suburbanos con iguales características socio-demográficas es esperable que tengan tasas de crecimiento similares.
5. Es esperable que estos factores socio-demográficos en grandes muestras muestren patronales similares, o valores semejantes.

7. Objetivos

El objetivo del presente trabajo consiste en demostrar que la tasa de crecimiento y curva poblacional de la Matanza desde 1991 hasta 2022 presenta una singularidad respecto a aquella desarrollada por los otros municipios de conurbano bonaerense. Tomando como base los datos obtenidos en los 4 últimos censos nacionales 1991-2001-2010 y 2022.

OBJETIVOS ESPECÍFICOS

1. Procesar y unificar la información censal desde 1991 a 2022 para distintos niveles de granularidad, ya que en algunos casos no se presentan todas las variables. Se requiere un pre procesamiento y limpieza de datos importante.
2. Exploratory Data Analysis. Construir las curvas poblaciones y comparaciones geográficas de las mismas. Analizar Data mining patterns.
3. Realizar proyecciones para población censo 2022 con metodologías tradicionales modernas utilizando las variables en forma individual o combinadas, tomando como base los 3 censos anteriores (1991-2001-2010).
4. Utilizar data mining para llenar los tiempos intercensales de las variables de interés.
5. Realizar proyecciones para el censo 2022 mediante data mining (ARIMA, decision tree regression algoritm, others), utilizando los 3 censos anteriores (1991-2001-2010)
6. Comparar las curvas censales y su ajuste con las proyecciones realizadas para 2022.-
7. Determinar el error de las distintas proyecciones para 2022 para todos los municipios, comparar con la situación de la Matanza.
8. Inferir si existe alguna metodología que ajuste mejor las proyecciones para el caso del AMBA.

Variables Se trata de una hipótesis multivariable donde las variables podrían tener una relación causa-efecto sobre el fenómeno analizado.

1. **Población Total.** Variable Dependiente. Cuantitativa. Cantidad de personas que habitan un determinado sector del territorio, en distintos niveles de agregación. INDEC
2. **Tasa de Natalidad.** Independiente. Cuantitativa. Refiere a la relación que existe entre el número de nacimientos ocurridos en un cierto periodo y la cantidad total de población existente en el área geográfica.
3. **Tasa de mortalidad.** Independiente. Cuantitativa. Es la proporción de personas que fallecen respecto al total de la población en un período de tiempo, usualmente expresada en tanto por mil (%) por año.

4. **Sexo.** Contextual. Categórica. Sexo de las personas, por nivel de desagregación. INDEC
5. **Edad.** Contextual. Cuantitativa. Edad de las personas, por nivel de desagregación. INDEC
6. **Nivel educativo.** Contextual. Categórica Ordinal. Se refiere al nivel de educación máximo alcanzado por un personal hasta el momento. Primario, secundario, terciario, Universitario. INDEC
7. **Cantidad de habitantes en el hogar.** Contextual. Cuantitativa. Se determina como la cantidad de personas conviviendo de forma permanente bajo el mismo techo. INDEC
8. **Asistencia escolar.** Contextual. Categórica Ordinal. Refiere a la asistencia o no a establecimientos educativos, desagregados por edad.

8. Metodologías

8.1. Técnicas

Para dominio del problema, particularmente debido al nivel de desagregación analizado-departamentos- se utilizarán las siguientes metodologías en el trabajo.

Metodologías Tradicionales

Metodología de variables sintomáticas (distribución por prorratoe; dist. Proporcional; método de tasas variables, correlación de razones, correlación de diferencias, correlación de tasas) (Álvarez,2001) [2].

El método más sencillo es la **distribución por prorratoe**, que se basa en el supuesto de que la distribución de la población en un área mayor se distribuya entre sus áreas menores en forma idéntica a la distribución de una variable sintomática. La población del área menor i en el momento t+n está expresada por:

$$P_{(i,t+n)} = \frac{S_{(i,t+n)}}{S_{(t+n)}} \times P_{(t+n)} \quad (1)$$

Donde:

t+n es el momento n años posteriores al último censo (momento t)

P (i,t+n) población del área menor i en el momento t+n;

P (t+n) estimación disponible de la población del área mayor en el momento t+n;

S (i, t+n) valor de la variable sintomática del área menor i en el momento t+n;

S (t+n) valor de la variable sintomática del área mayor en el momento t+n.

El método de **distribución proporcional** supone que la población varía durante un período en la misma proporción que lo hace la variable sintomática en el mismo período. La población del área menor i en el momento t+n se deduce de esta relación contemplando un factor de ajuste (Fa) para que la suma de las áreas menores coincida con el área mayor:

$$P_{(i,t+n)} = \frac{S_{(i,t+n)}}{S_{(i,t)}} \times P_{(i,t)} \times F_{(a)} \quad (2)$$

$$F_{(a)} = \frac{P_{(i,t+n)}}{\frac{S_{(i,t+n)}}{S_{(i,t)}} \times P_{(i,t)}} \quad (3)$$

Donde:

t es el momento del último censo;

t+n es el momento n años posterior al último censo;

P (i,t+n) población del área menor i en el momento t+n;

P (i,t) población del área menor i en el momento t ;

S (i,t) valor de la variable sintomática del área menor i en el momento t ;

S (i,t+n) valor de la variable sintomática del área menor i en el momento t+n;

F a factor de ajuste entre suma de áreas menores estimadas y área mayor.

El método de las **tasas vitales de Bogue** consiste en estimar la población de áreas menores utilizando solamente datos de los registros vitales. Bogue (1950) tomó como supuesto la existencia de una relación inversa entre las variaciones anuales de las tasas de natalidad y de mortalidad de las áreas menores y que el desvío de las tasas propias en relación a las tasas del área mayor permanece constante a través del tiempo. La expresión del modelo es:

$$P_{(i,t+n)} = a \times \frac{B_{(i,t+n)}}{b_{(i,t+n)}} + (1 - a) \times \frac{D_{(i,t+n)}}{d_{(i,t+n)}} \quad (4)$$

Donde:

a es el coeficiente de ponderación entre las dos estimaciones;

b $(i,t+n) = [(B(i,t) / P(i,t)) / (B(\bullet,t) / P(\bullet,t))] * (B(\bullet,t+n) / P(\bullet,t+n))$

d $(i,t+n) = [(D(i,t) / P(i,t)) / (D(\bullet,t) / P(\bullet,t))] * (D(\bullet,t+n) / P(\bullet,t+n))$

B es la cantidad de nacidos vivos;

b es la tasa bruta de natalidad;

D es la cantidad de defunciones;

d es la tasa bruta de mortalidad.

EN DESARROLLO —————— 4

Metodologías Data Mining CART, ARIMA, decision tree regression and classification algorithms (Chawda, Rane, Giri, 2018) [6].

8.2. Herramientas

Como herramientas se contemplan los lenguajes Python y SQL. Los datos se trabajarán en base de datos POSTGRES, para visualización geográfica se utilizará QGIS y OpenJump. En cuanto al scripting se gestionará mediante Jupyter Notebooks.

9. Referencias bibliográficas

1. Instituto Nacional de Estadística y Censos - INDEC. (2013). Proyecciones provinciales de población por sexo y grupo de edad 2010-2014. - Instituto Nacional de Estadística y Censos - INDEC, 2013. E-Book. (Vol. 1a ed.).
2. Instituto Nacional de Estadística y Censos - INDEC. (2015). Estimaciones de población por sexo, departamento y año calendario 2010-2025. (1a. ed.).
3. Gupta, Arindam , Gupta, Arindam , Gupta, Arindam (2012).Exploring New Models for Population Prediction in Detecting Demographic Phase Change for Sparse Census Data
4. Hoque, Nazrul(2012). Evaluation of small area population estimates produced by Housing Unit, Ratio-correlation, and Component Method II compared to 2000 Census counts. Canadian Studies in Population 39, No. 1–2, pg. 91-108.
5. Chawda, Manan, Rane, Rutuja, Giri, Skrinanth(2018) 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT) .Demographic Progress Analysis of Census Data Using Data Mining , pg 1894-1897

10. DESARROLLO

11. Obtención de datos desde las fuentes. Consolidación Dataset

En primera instancia se busca recopilar los datos de los CENSOS poblacionales correspondientes desde 1991. Es decir Censos 1991, 2001, 2010 y 2022. En general el INDEC dispone la información en formatos CSV o bien archivos Microsoft Excel de difícil procesado para su ingestión por parte de un proceso de ETL. En primera instancia los archivos

no cuentan con estructura de encabezados en la primera línea, presentan descripciones y comentarios en las primeras líneas y no respetan formatos uniformes según la dimensión de análisis de los datos dentro del mismo periodo censal. Por otra parte, el formato de presentación y las variables relevadas se han visto modificadas desde el censo 1991 hasta el 2022. Una vez analizada las variables comunes a todos estos censos, fue necesario un pre-proceso manual de los archivos para lograr uniformidad y consistencia de forma de que sean tomados por el proceso de ETL . En cuanto a las variables censales comunes al período analizado se toman las indicadas en el cuadro 1.

Cuadro 1: Descripción de Variables

Variable	Description
Censo	Año del censo indicado como entero de 4 dígitos.
CodigoDept	Código de 4 dígitos con el INDEC indica los municipios / departamentos censales.
Población	Cantidad de personas que habitan un determinado sector del territorio, en distintos niveles de agregación.
Varones	Cantidad de personas de sexo Masculino.
Mujeres	Cantidad de personas de sexo Femenino.
VivPartTot	Cantidad de viviendas particulares totales para ese nivel de agregación.
VivColectTot	Cantidad de viviendas Colectivas totales para ese nivel de agregación.
IndMasc	Índice de masculinidad.
Superficie	Superficie total del partido a la fecha del censo.

Es importante aclarar que existe una base consolidada -REDATAM [9]- con toda la información censal para los censos 2001 y 2010, la misma no está disponible para los censos de 1991 y 2022. REDATAM [9] es un software para procesamiento estadístico especializado en microdatos de censos de población y vivienda, encuestas y estadísticas vitales, desarrollado por el CELADE - División de Población de la CEPAL, de las Naciones Unidas. A su vez esta base sólo puede accederse mediante este software específico, no disponiendo de back ups y en general las salidas de las consultas son también archivos CSV no normalizados que requieren post-proceso.

11.1. Dimensión Departamento- Slowly Changing Dimension

Al analizar los distintos censos desde 1991 hasta 2022 los departamentos del AMBA han modificado su división político – administrativa. Es decir, se han creado varios partidos nuevos, así como fusionado otros. Esto general que la dimensión departamento cambia a través de los distintos censos. Es por esto que se adopta un enfoque tipo Slowly Changing dimensión tipo 3. Donde se impacta el cambio de departamento, conservando en el registro el valor anterior, de que departamento proviene y la superficie correspondiente asociada. Esto es necesario para poder luego procesar información consistente en cuenta a la granularidad analizada. Mismo el INDEC consigna :"Nota: con el fin de posibilitar la comparación entre los Censos 1991 y 2001, los datos que corresponden al año 1991 fueron reprocesados según la división político - administrativa vigente al año 2001." [8] INDEC, 2001

En el cuadro 2 puede verse la dimensión departamento con los campos agregados y los comentarios correspondientes en cada caso.

11.2. Diccionario de Departamentos

Debido a las inconsistencias y falta de normalización de los archivos CSV que provee el INDEC, fue necesaria la creación de un diccionario de departamento asociando las distintas acepciones del nombre de departamento con su correspondiente código para que el proceso de ETL pudiese utilizar los códigos de departamento como clave foránea común a todos los Censos Nacionales. Se encontraron casos donde el mismo departamento figura nombrado con tilde , sin tilde , con espacios o abreviaturas en el campo. De esta forma en la ingesta (ETL) se buscaba el código de departamento correspondiente, que funciona como clave única.

Cuadro 2: Dimensión Departamento

Codigo	Departamento	PartidoFrom	Codigo From	Sup1991	Sup2001	Comentarios
6028	Almirante Brown			122.0	122.0	
6035	Avellaneda			55.0	55.0	
6091	Berazategui			188.0	188.0	
6260	Esteban Echeverría			377.0	120.0	(1) Superficie modificada, cede tierras a Cañuelas y San Vicente y para la creación de Ezeiza y Presidente Perón. Leyes provinciales 11.550 del 20/10/1994 y 11.480 del 25/11/1993.
6270	Ezeiza	Esteban Echeverría	6260	0.0	223.0	(2) Se crea con tierras del partido de Esteban Echeverría. Ley provincial 11.550 del 20/10/1994.
6274	Florencio Varela			206.0	190.0	(3) Sup.modificada, cede tierras a Presidente Perón. Ley 11.480 del 25/11/1993.
6371	General San Martín			56.0	56.0	
6408	Hurlingham	Morón	6568	0.0	36.0	(4) Se crea con tierras del partido de Morón. Ley provincial 11.610 del 28/12/1994.
6410	Ituzaingó	Morón	6568	0.0	39.0	(5) Se crea con tierras del partido de Morón. Ley provincial 11.610 del 28/12/1994.
6412	José C. Paz	General Sarmiento	6005	0.0	50.0	(6) Se crea con tierras del partido de General Sarmiento. Ley provincial 11.551 del 20/10/1994.
6427	La Matanza			323.0	323.0	
6434	Lanús			45.0	45.0	
6490	Lomas de Zamora			89.0	89.0	
6515	Malvinas Argentinas	General Sarmiento	6005	0.0	63.0	(7) Se crea con tierras del partido de General Sarmiento e incorpora un sector del partido de Pilar. Ley provincial 11.551 del 20/10/1994.
6539	Merlo			170.0	170.0	
6560	Moreno			180.0	180.0	
6568	Morón			131.0	56.0	(8) Superficie modificada, cede tierras para la creación de los partidos de Hurlingham e Ituzaingó. Ley 11.610 del 28/12/1994.
6658	Quilmes			125.0	125.0	
6749	San Fernando			924.0	924.0	
6756	San Isidro			48.0	48.0	
6760	San Miguel	General Sarmiento	6005	0.0	80.0	(9) Se crea con tierras del partido de General Sarmiento. Ley provincial 11.551 del 20/10/1994.
6805	Tigre			360.0	360.0	
6840	Tres De Febrero			46.0	46.0	
6861	Vicente López			39.0	39.0	
6005	General Sarmiento			196.0	0.0	Desaparece para el año 2001

11.3. Proyecciones INDEC 2010-2025

Se incorpora también al conjunto de datos que se analizarán proyecciones realizadas por el INDEC [3] que indican la “Población estimada al 1 de julio de cada año calendario por sexo, según partido. Provincia de Buenos Aires. Años 2010-2025”.

11.4. Componente geográfico

Con el objetivo de enriquecer el dataset se incorpora una tabla con los polígonos geográficos de cada departamento. La misma se obtiene del Instituto Geográfico Nacional (IGM) [10], de donde se obtiene la capa SIG para todos los departamentos de la República Argentina. Luego se cruza con el listado de departamentos del AMBA , generando un dataset con las características geográficas sumadas a los datos censales recopilados desde 1991 a 2022.

Cuadro 3: Descripción de Variables Geográficas

Variable	Nombre	Descripción
geom	geometry	Polígono WKT con los límites del departamento
fna	Nombre geográfico	Nombre completo que se utiliza para designar un objeto en un mapa o carta. Está formado por el término genérico y el término específico. Ejemplo: río Mendoza.
gna	Término genérico	Parte del nombre geográfico que indica el tipo de objeto que identifica. Ejemplo: río, monte, glaciar, establecimiento.
nam	Término específico	Parte de un nombre geográfico que acompaña al término genérico y que identifica e individualiza un objeto geográfico determinado. Ejemplo: Paraná en río Paraná; Upsala en glaciar Upsala; Las Marías en establecimiento Las Marías; Esperanza en el caso de bahía Esperanza.
in1	Código INDEC	Código único de vías de circulación asignado por el Instituto Nacional de Estadística y Censos de la República Argentina.
fdc	Fuente de captura	Identificación del nombre y tipo de fuente utilizada para capturar la información. Puede incluir fecha y otros datos adicionales.
sag	Autoridad de fuente	Nombre de la autoridad responsable de la información utilizada.

En base

11.5. Herramientas

Para el guardado y manipulación de los datos se va a trabajar con un base de datos POSTGRES (Local). Se desarrolla justamente un proceso de ETL en Python y SQL mediante en primera crear todas las tablas necesarias, para luego producir la ingesta ,transformar en caso de que sea necesario, de los diferentes inputs que van a componer el modelo de datos.

11.6. Diagrama de entidad relación

12. Metodologías Aplicadas

..— TODAVIA SIN DESARROLLO.—

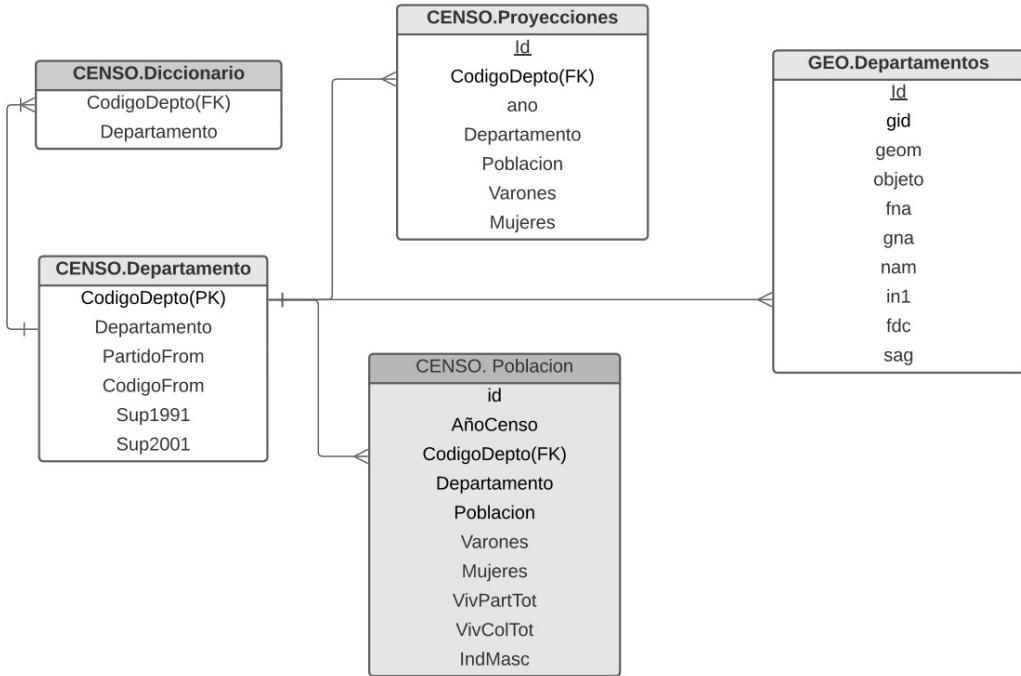


Figura 1: Diagrama de Entidad Relación

13. Análisis exploratorio de datos

13.1. Población

Con sustento en la base de datos ".^MBA"(POSTGRES/POSTGIS) se analizarán los distintos datos recopilados según nivel de agregación. En primera instancia se generó un dataset con todos los departamentos del AMBA, sus características geográficas así como los valores de las variables correspondientes a los censos 1991-2001-2010-2022.2

En el cuadro 4 se puede observar el esquema del dataset .

nam	coddepto	anio	pob	var	muj	vivpart	vivtotal	sup	indmasc	denspob
Almirante Brown	06028	1991	450698.0	222042.0	228656.0	nan	nan	157.87	97.1	2854.87
Almirante Brown	06028	2010	552902.0	270247.0	282655.0	156218.0	78.0	157.87	95.6	3502.26
Almirante Brown	06028	2022	585852.0	281842.0	301779.0	184403.0	60.0	157.87	93.4	3710.98
Almirante Brown	06028	2001	515556.0	252454.0	263102.0	143543.0	88.0	157.87	96.0	3265.70
Avellaneda	06035	2001	328980.0	155450.0	173530.0	117200.0	59.0	68.54	89.6	4799.82
Avellaneda	06035	1991	344991.0	164243.0	180748.0	nan	nan	68.54	90.9	5033.43
Avellaneda	06035	2010	342677.0	162264.0	180413.0	121307.0	68.0	68.54	89.9	4999.66
Avellaneda	06035	2022	370939.0	174572.0	194911.0	144988.0	64.0	68.54	89.6	5412.01
Berazategui	06091	2010	324244.0	158608.0	165636.0	96029.0	37.0	268.91	95.8	1205.77
Berazategui	06091	2022	360582.0	173892.0	186255.0	117699.0	37.0	268.91	93.4	1340.90
.....										

Cuadro 4: Dataset Censos AMBA

Se observa en el cuadro 5 que existen valores nulos para el valor poblaciones, que se corresponden, con los casos de municipios de desaparecen en la reorganización administrativa desde 1991 a 2001.

A saber : Esteban Echeverría' , 'Ezeiza' , 'Florencio Varela' , 'Hurlingham' , 'Ituzaingó' , 'José C. Paz' , 'Malvinas Argentinas' , 'Morón' , 'San Miguel' y 'General Sarmiento' .

Se destaca también que para el censo 1991 no se detalla el total de viviendas particulares y colectivas .

Asimismo se analiza la evolución de la densidad de población [hab/km²], la misma puede observarse en la figura 3. Las mismas consideraciones aplican en el caso de densidad respecto a los departamentos que sufren modificaciones administrativas entre 1991 y 2001, donde los valores de densidad se ven fuertemente afectados y no pueden asociarse al fenómeno poblacional.

Cuadro 5: Summary de columnas de Censos AMBA

Column	Non-Null Count	Dtype
nam	96	object
cod_dept	96	object
anio	96	object
pob	90	float64
var	90	float64
muj	90	float64
vivpart	72	float64
vivtotal	72	float64
sup	96	object
ind_masc	90	object
dens_pob	90	object

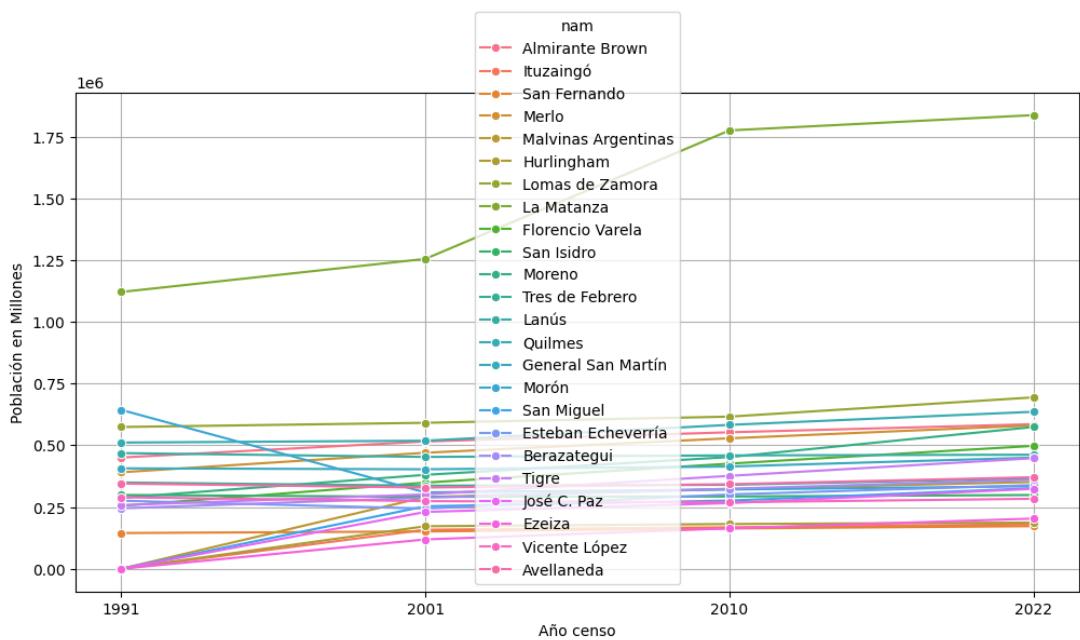


Figura 2: Población por departamento 1991-2022

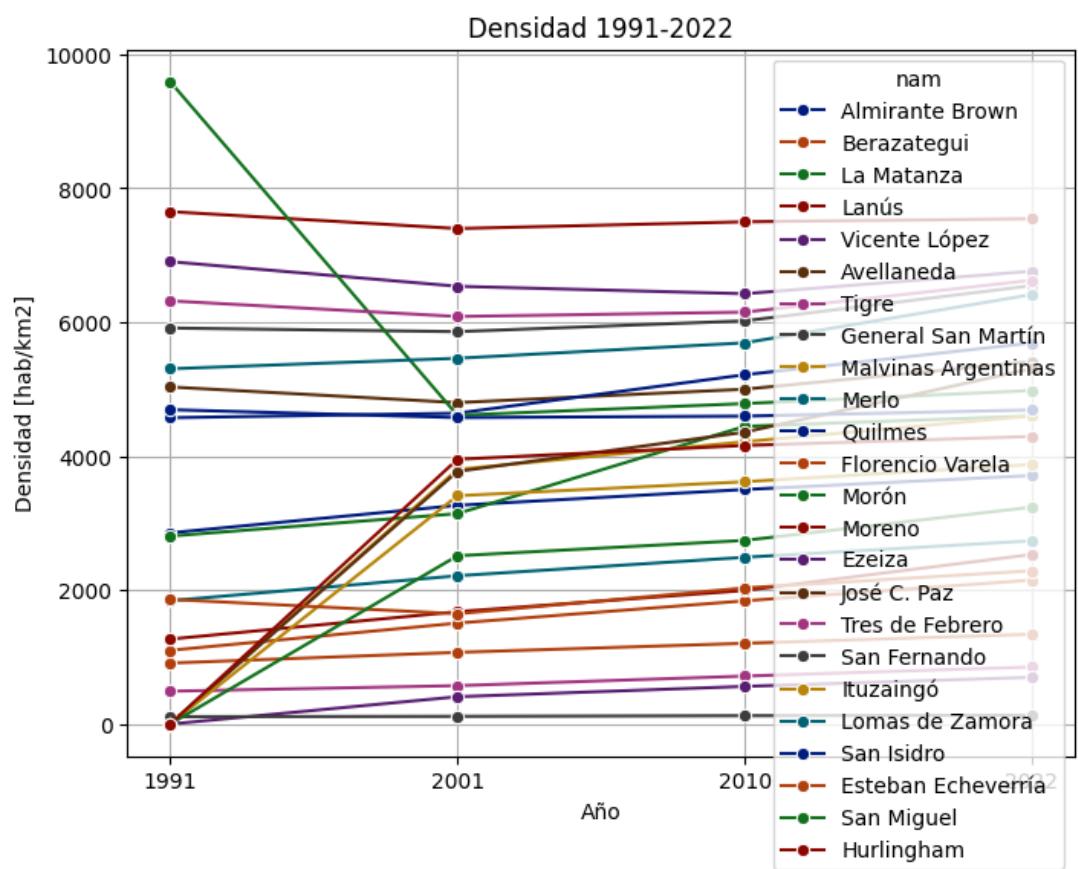


Figura 3: Densidad de Población por departamento 1991-2022

13.1.1. Población y densidad en vista geográfica

En este caso se busca una visualización geográfica del dataset. Utilizando el software QGIS conectado directamente a la base de datos AMBA (POSTGRES) se pudo disponer la información de cada Departamento con su geolocalización. Se detalla en la figura 4 la población total para cada departamento (Censos 1991 -2022).Se observa claramente que el distrito más poblado es La Matanza. Debido su extensión no es el departamento con mayor densidad poblacional en [hab/km²] figura 5 , los distritos más densamente poblados son Lanús y Vicente López.

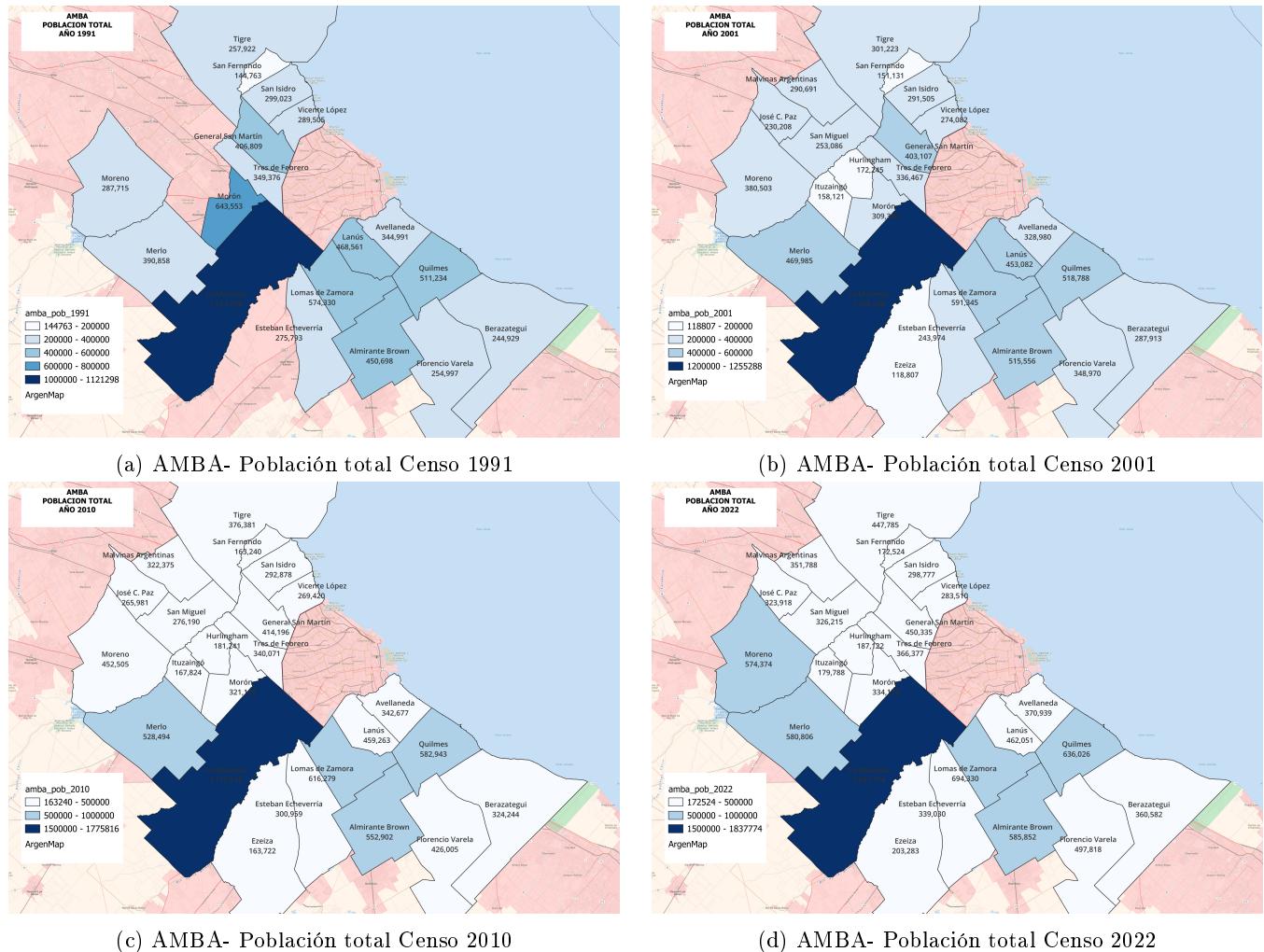


Figura 4: Población Total Censos 1991 2022

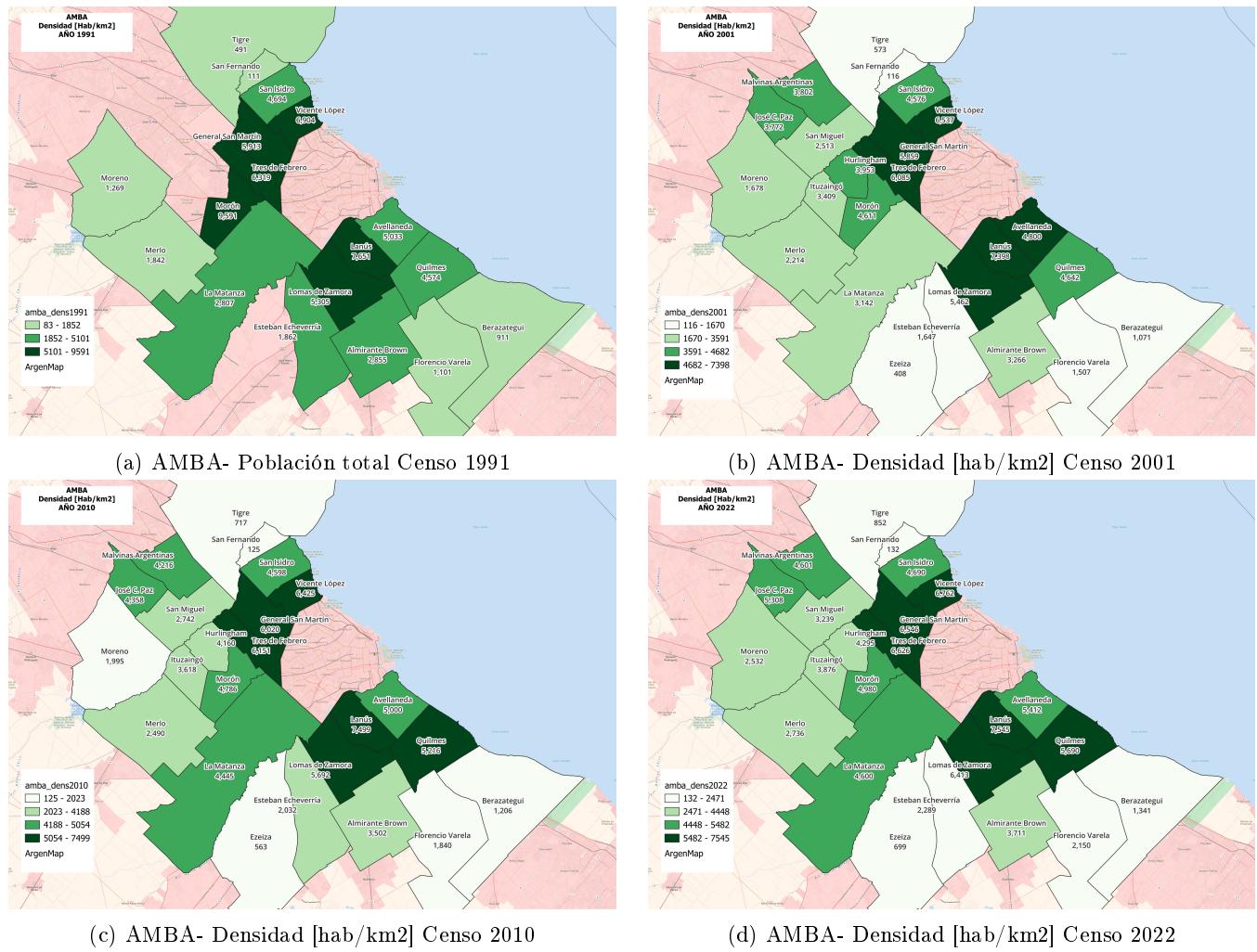


Figura 5: AMBA- Densidad [hab/km²] Censos 1991 2022

13.1.2. Población - Ratios de crecimiento

En este caso se determinaron los ratios de crecimiento entre censos consecutivos para cada departamento. Para este análisis se trata de manera diferenciada aquellos municipios que han sufrido modificaciones administrativas desde 1991 a 2001, donde se que al desmembrarse en varios partidos, resignando superficie, algunas tasas o ratios de crecimiento se muestran negativos. Asimismo partidos que ha visto incrementada su superficie presentan tasas de crecimiento no representativas del fenómeno demográfico.

Por otra parte partidos que desaparecen o se crean en el año 2001 no se consideran el valor del ratio para el intervalo 1991-2001.

En el cuadro 6 se puede observar los valores obtenidos. Para el año en cuestión la tasa surge de dividir el valor poblacional del censo de este año sobre el valor poblacional del censo anterior, expresado en porcentaje. Es decir:

$$\text{Growth Ratio} = \left(\frac{\text{Población}_{(\text{Censo } X)}}{\text{Población}_{(\text{Censo } X - 1)}} - 1 \right) \times 100 \quad (5)$$

nam	anio	pob	growth_ratio
Almirante Brown	2001	515556.0	14.39
Almirante Brown	2022	585852.0	5.96
Almirante Brown	2010	552902.0	7.24
Avellaneda	2001	328980.0	-4.64
Avellaneda	2010	342677.0	4.16
Avellaneda	2022	370939.0	8.25
Berazategui	2022	360582.0	11.21
Berazategui	2001	287913.0	17.55
Berazategui	2010	324244.0	12.62
Esteban Echeverría	2001	243974.0	-11.54
Esteban Echeverría	2010	300959.0	23.36
Esteban Echeverría	2022	339030.0	12.65
Ezeiza	2022	203283.0	24.16
Ezeiza	2010	163722.0	37.81
....			

Cuadro 6: Extracto. Tasa de crecimiento Intercensal

A partir de análisis estadístico simple sobre estas tasas se observa una gran disparidad en los valores, siendo que se trata de un sector geográfico de similares características demográficas. Cuadro 7

Estadístico	Valor
Count	66.000000
Mean	9.286961
Standard Deviation	13.111467
Minimum	-51.926259
25 % Percentile	3.033150
50 % Percentile	8.129835
75 % Percentile	16.476163
Maximum	41.466819

Cuadro 7: Summary Statistics for Growth Ratio

A partir de la elaboración de un gráfico boxplot se determinaron los **Outliers** para estas tasas de crecimiento.

En este caso en el cuadro 6 se observan 4 outliers, tanto en valores positivos como negativos. Como se comentó anteriormente no se analizarán los casos con modificaciones administrativas desde 1991 a 2001. Cuadro 2 .Se desestima

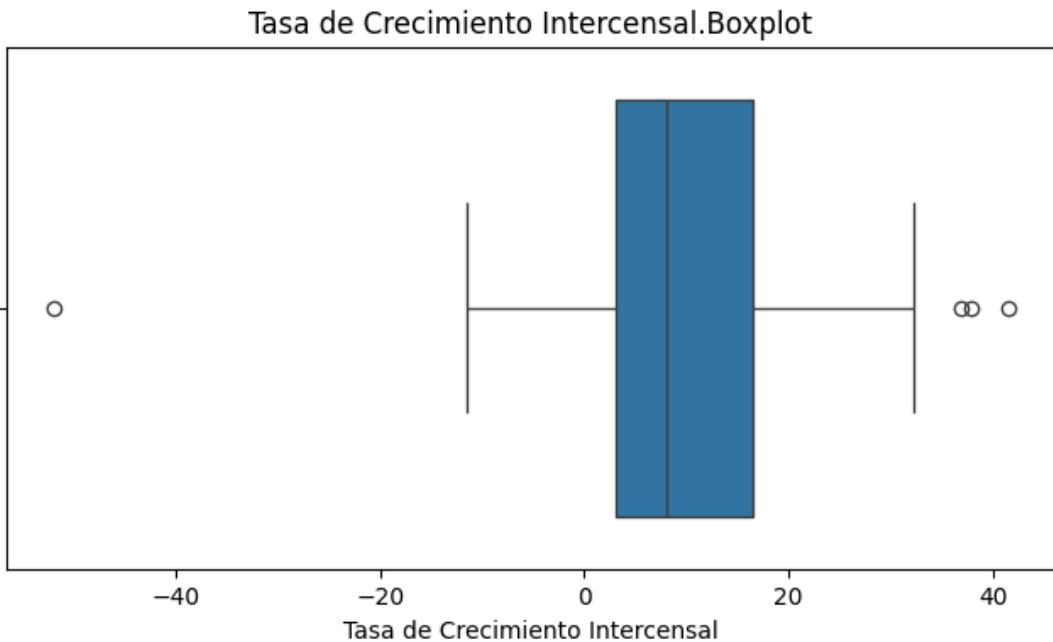


Figura 6: Box plot Tasas de crecimiento intercensa 1991-2022

el caso de Morón para el periodo 1991-2001 donde la tasa es negativa de 51.9, debido a la cesión de tierras para la creación de distintos partidos.

id	nam	codDept	anio	pob	growthRatio
54	Morón	06568	2001	309380.0	-51.92
78	Florencio Varela	06274	2001	348970.0	36.85
13	La Matanza	06427	2010	1775816.0	41.47
29	Ezeiza	06270	2010	163722.0	37.80

Cuadro 8: Outliers para Tasas de crecimiento intercensal

Se grafica a continuación (Figura 7) la curva población de los 3 outliers destacados :Florencio Varela ,La Matanza y Ezeiza.

13.1.3. Tasas de crecimiento intercensal. coeficiente de Variación

Con el objetivo de profundizar el análisis de las tasas de crecimiento se procedió a calcular luego la desviación estandar y el coeficiente de variación por Departamento.

En primera instancia se agrupa por departamento para determinar la media y la desviación estándar de cada subset de datos. Se define el coeficiente de variación (CV) como:

$$CV_{(Departamento)} = \left(\frac{\text{Std.Dev}_{(Departamento)}}{\text{Mean}_{(Departamento)}} \right) \times 100 \quad (6)$$

El coeficiente de variación es una medida de la dispersión alrededor de la media de la población. En este caso sería curvas poblacionales con gran dispersión en su tasas de crecimiento para los censos del periodo analizado año 1991 -2022.-

En la figura 8 se observa el resultado .Se destaca en particular los valores extremos tanto negativos como positivos.A saber : Vincente López, Morón, Tres de Febrero, Avellaneda ,La Matanza , Esteban Echeverría, General San Martín. Cabe recordar en este instancia que los partidos de Esteban Echeverría , Florencio Varela y Morón sufrieron modificaciones administrativa en el período 1991-2001.-

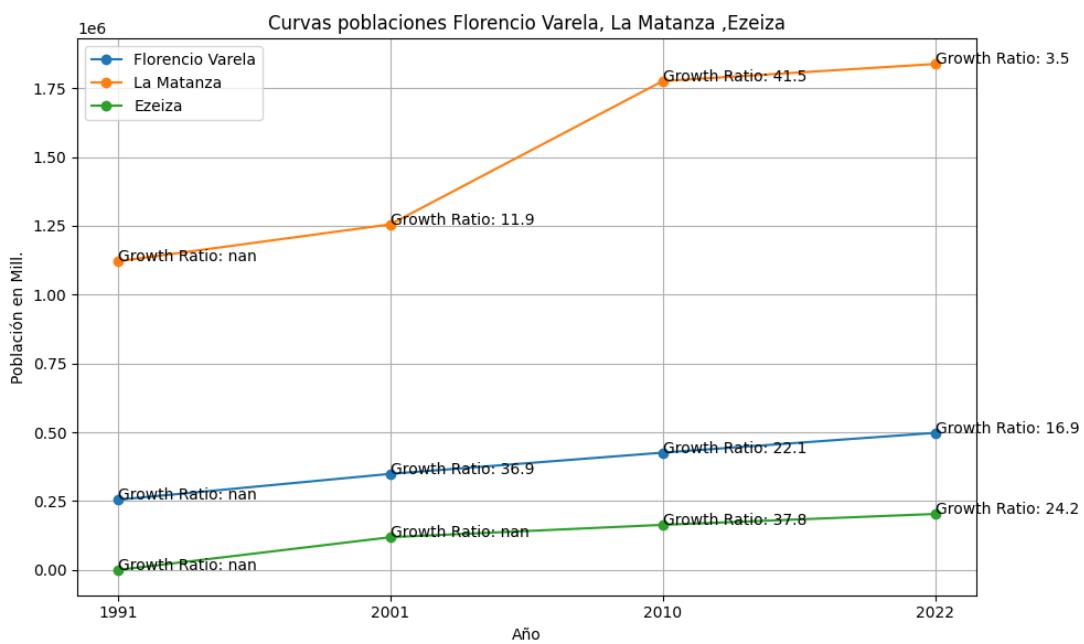


Figura 7: Curvas de Población por departamento 1991-2022

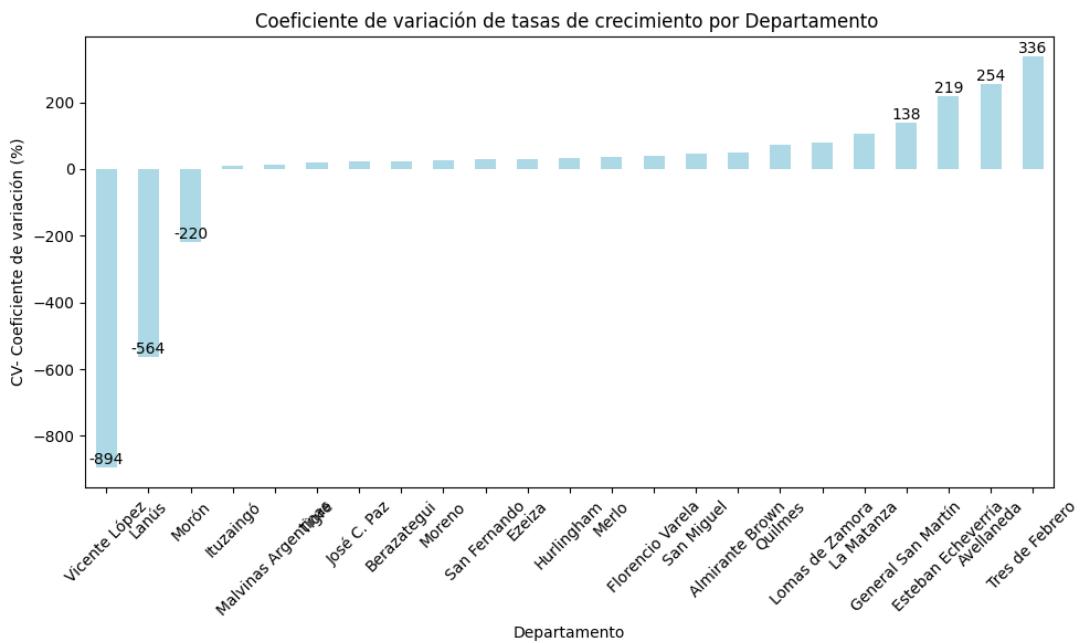


Figura 8: CV. Coeficiente de variación por departamento -1991-2022

En las figuras 9 y 10 puede observarse la curvas poblacionales de los departamentos destacados.

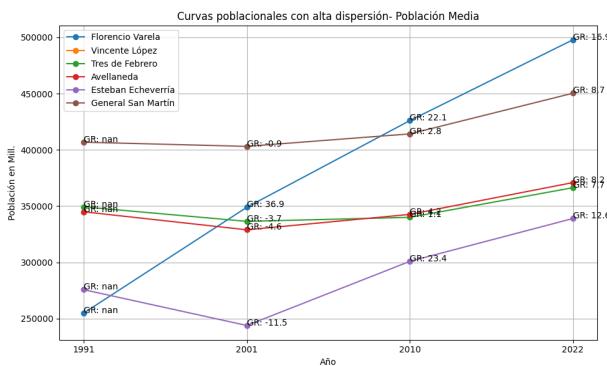


Figura 9: Curvas poblacionales de alta dispersión - Población Media

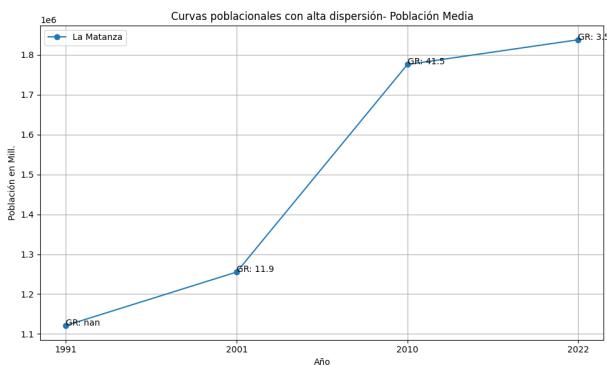


Figura 10: Curvas poblacionales de alta dispersión - Población Alta