

Una caracterización de la estructura socioeconómica de la Argentina en 2023 basada en la Encuesta permanente de datos (EPH)

Objetivos

El objetivo de este trabajo es presentar en forma práctica para los usuarios las principales características sociodemográficas y de ingreso de la población argentina tal como las releva la EPH.

Esto nos permite, entre muchas otras cosas que iremos describiendo en este trabajo,

Calcular la mediana de ingresos de la población relevada por la EPH (más sobre el alcance de los datos debajo), en función de características sociodemográficas, para esta entrega, elegí analizar los ingresos para los distintos grupos:

- a) Por sexo: mujeres y varones.
- b) Por nivel educativo.
- c) Por aglomerado relevado y por región del país.
- d) Por categoría de actividad: trabajadores asalariados, y dentro de estos, formales e informales, privado vs estatal, patrones y cuentapropistas, jubilados y pensionados, perceptores de ayuda social.
- e) Caracterizar el nivel educativo de la población argentina, y la relación con su condición ocupacional, uno de los objetivos fundamentales del campo de la Economía Laboral. Los niveles educativos relevados por la EPH son,
- f) La relación entre ingresos y otras características demográficas, como lo es el tamaño de establecimiento.

El trabajo comienza con una descripción general de la muestra y la distribución del ingreso en las dos primeras páginas del dashboard y concluye con 2 estudios de caso puntuales, un análisis de las diferencias por género y por aglomerado de residencia

El motivo por el cual utilizamos la mediana en lugar de la media es porque, en mi opinión, las medianas de ingresos están menos influidas por valores extremos y permiten una representación más adecuada de la tendencia central, eso no implica que desconozcamos la distribución del ingreso, de hecho, uno de nuestras páginas se centra exclusivamente en la distribución.

Hipótesis

La hipótesis es que las características demográficas repercuten en forma fundamental en el nivel de ingreso de la población argentina, que el nivel educativo es la variable demográfica más fuertemente correlacionada con los ingresos, en particular en el nivel Universitario Completo, que existen diferencias geográficas muy marcadas que exceden el poder explicativo de las características demográficas, lo que requeriría ampliar el análisis al nivel de la estructura económica de cada provincia y aglomerado, lo cual excede el alcance del trabajo y la base de datos.

También veremos como la variable género impacta en los ingresos, una conclusión de este trabajo al menos, y en función de la muestra que tenemos, es que el género está más correlacionado con la cantidad de horas trabajadas que con el ingreso por hora, y es a través del canal de las horas

trabajadas donde se explican fundamentalmente las diferencias de ingresos entre varones y mujeres.

Una última hipótesis, que surge de una lectura de la literatura sobre encuestas de hogares, ingresos y pobreza, es que la EPH es un instrumento muy útil para entender los ingresos laborales, por jubilaciones y transferencias desde el Estado, pero inadecuado para captar los ingresos de dueños de grandes empresas privadas e individuos con muy elevados ingresos por renta de la propiedad, que se concentran en el 1% más rico de la población.

Usuarios:

Los usuarios de esta información, son, **a nivel operativo, todas aquellas personas interesadas en conocer la distribución del ingreso y la demografía de Argentina**, cuales son los ingresos medianos como varían en función de la región y aglomerado en el que viven, su nivel educativo, a cuánto ascienden los ingresos salariales, los de los jubilados y pensionados, los de aquellos que derivan su ingreso de rentas de la propiedad, negocios propios, alquiler de viviendas, etc. E incluso poder comparar su propio ingreso con las medianas publicadas y poder entender donde se sitúa cada uno en la pirámide social.

Además, teniendo en cuenta que en el contexto de inflación de Argentina los salarios, tanto públicos como privados, las jubilaciones, los alquileres, las tarifas de servicios públicos y otras fuentes de ingresos son renegociados y actualizados con una alta frecuencia (tendiente a ser trimestral e incluso mensual), **a nivel táctico** conocer la distribución del ingreso y su composición debería ser de suma utilidad para ayudar a clarificar esas renegociaciones en las que participan **trabajadores, sindicatos, entidades patronales y funcionarios públicos**.

Fuente de datos

La base de datos es la base individual de la Encuesta permanente de hogares (EPH) del primer trimestre de 2023, la última publicada a agosto de 2023.

El archivo ***usu_individual_T123*** que forma parte de esta entrega permite ver la base de datos en su estado natural, es decir, tal como es publicada por el Indec, a partir de la cual se generaron las distintas tablas y se hicieron las transformaciones en Power BI que se detallan más adelante en este documento.

La EPH es una base de datos que surge de entrevistas que lleva adelante el Indec para relevar la situación ocupacional y de ingresos de la población argentina, junto con características personales como edad, nivel educativo, vivienda, etc.

Es de destacar que la cobertura de la EPH no es nacional, sino sobre un conjunto de aglomerados urbanos (detallados en el campo aglomerados), que representan alrededor del 60% de la población argentina. Como puede verse en la suma del campo Cantidad de habitantes en la tabla 1 (campo PONDERA en la base original) del dashboard, los registros de la EPH representan a 29.3 millones personas, si la población del país es de 46 millones de personas según el censo 2022, entonces la EPH releva la situación del 63% de la población. Lamentablemente, según mi conocimiento, el Indec no indaga sobre la situación ocupacional o de ingresos de los habitantes de localidades pequeñas o rurales que representan al restante 37% de la población.

El campo CODUSU es un número de identificación para cada registro, que permite relacionar una vivienda con los hogares y personas que la componen a lo largo de los cuatro trimestres en que

participa. Y el componente es el componente del hogar, combinando ambos tenemos un registro único de cada persona relevada en la muestra, que a su vez, a través del campo PONDERA, permite que una muestra de 48 mil observaciones represente a 29 millones de habitantes.

Finalmente, quiero describir las que en mi opinión son las principales debilidades de la EPH, en primer lugar, todos los datos dependen de las respuestas de las personas contactadas por el Indec, muchas personas contactadas no responden la encuesta y, entiendo, es altamente probable que las características de las personas que no responden la encuesta sean distintas de las que sí responden, creando una imagen relativamente sesgada de las características personales, ocupacionales y de ingresos de la población, dado el estado de la disponibilidad de datos, no tenemos forma de cuantificar ese sesgo.

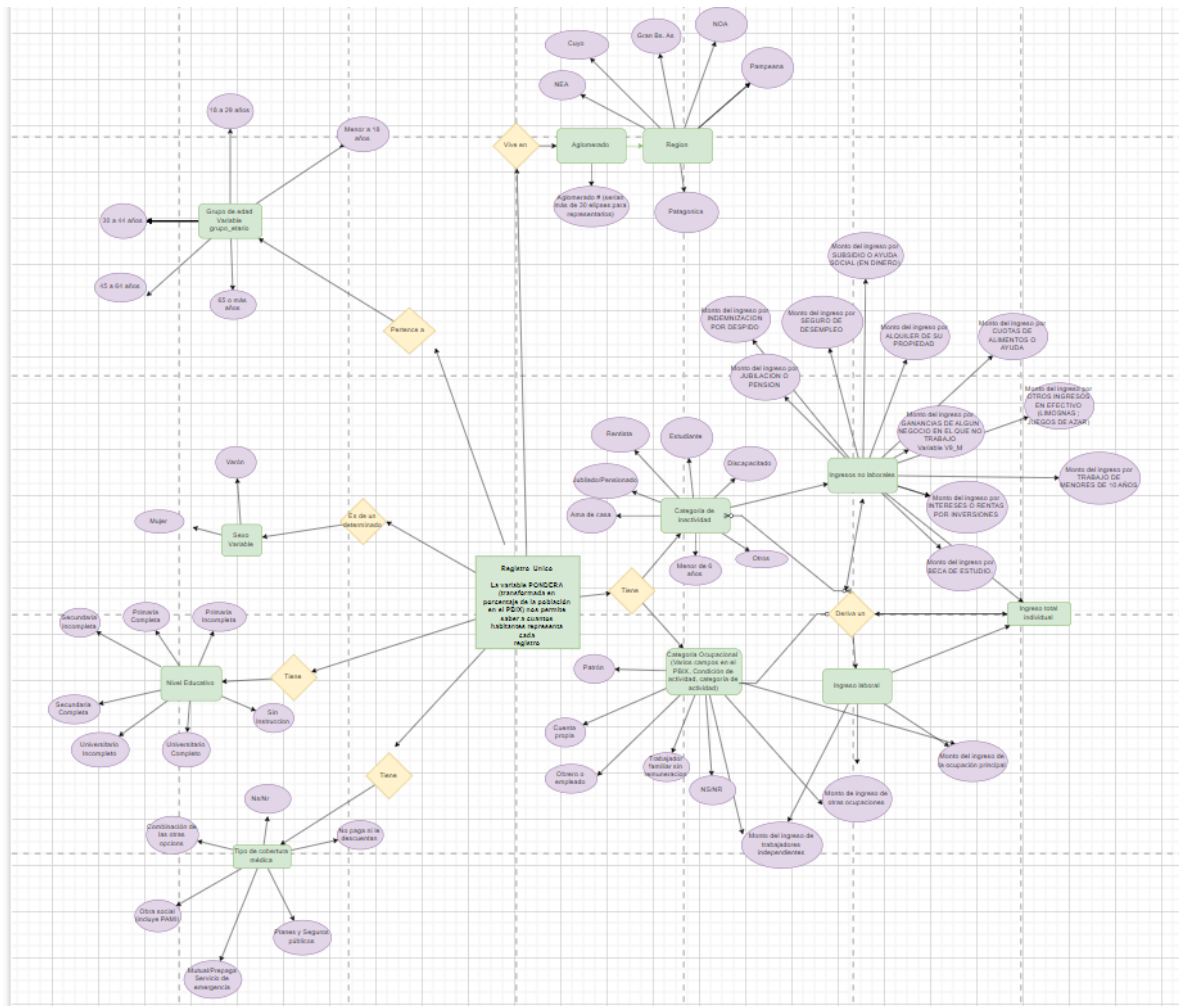
En todo caso, en mi opinión esto no representa un problema mayor, dado que las discusiones públicas sobre el índice de pobreza e ingresos se basan en datos elaborados a partir de la EPH combinados con los valores de la canasta básica, de modo que la sociedad reconoce el valor de este instrumento a la hora de analizar la situación laboral y de ingresos de la población.

En segundo lugar, es ampliamente reconocido en la literatura que existe una subestimación de ingresos en las declaraciones de las encuestas de hogares, este es un problema que existe en todos los países en los que se utilizan estos instrumentos para evaluar las condiciones de vida, ocupacionales y de ingresos, y también es sabido que esa subdeclaración es más elevada cuanto mayor es el nivel de ingresos del entrevistado, y para ingresos no laborales como renta de las inversiones, es decir, las encuestas de hogar presentan un cuadro de la distribución del ingreso que tiende a ser más igualitario y más centrado en la distribución del ingreso laboral que la población subyacente que se está evaluando.

En parte, este sesgo de las encuestas de hogares puede ser atenuado estudiando los ingresos elevados (generalmente al interior del percentil más rico de la población) que surgen de la información tributaria, o tratando de estimar el ingreso de los miembros más ricos y famosos de la sociedad en base al valor de sus empresas (siempre y cuando coticen en bolsa y pueda calcularse dicho valor) pero, hasta donde tengo conocimiento, no hay estudios de ese estilo para Argentina.

Diagrama Entidad Relación:

Para una vista detallada se puede consultar el archivo en formato draw.io Diagrama Entidad – Relación que forma parte de la entrega:



Tablas y Variables:

Tabla 1 - Características de la muestra

Clave	Nombre del Campo	Descripción	Tipo de dato
	CODUSU	Código para distinguir viviendas: permite aparear hogares y personas. Además, permite realizar el seguimiento de hogares y personas a través de los trimestres	VARCHAR
	ANO4	Año de relevamiento (4 dígitos)	Int
	COMPONENTE	Numero de componente del hogar relevado	Int
PK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observación de la muestra	VARCHAR
	REGION	Region del país a la que pertenece el aglomerado	CHAR
	AGLOMERADO	Aglomerado en el que vive la gente relevada en cada línea	CHAR
	Cantidad de habitantes	Cantidad de habitantes que representa cada línea de la muestra (PONDERA en el original)	Int
FK	Categoría Ocupacional	Categoría ocupacional que representa la observación.	CHAR
	Categoría de Inactividad	Categoría de inactividad que representa la observación, (Jubilado, Ama de casa, discapacitado, etc)	CHAR
FK	Condición de actividad	Variable construida a partir de la categoría ocupacional y de inactividad para tener una única variable que agrupe a ocupados e inactivos	CHAR
	Porcentaje de la población	Variable construida a partir de cantidad de habitantes para entender que porcentaje de la población total relevada representa cada línea de la base de datos	Int

Tabla 2 - Características de los miembros del Hogar

Clave	Nombre del Campo	Descripción	Tipo de dato
PK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observación de la muestra	VARCHAR
	Sexo	Sexo de la observación (Varones y Mujeres)	CHAR
	Fecha_Nacimiento	Fecha de nacimiento de la observación	INT
	Edad	Edad en años cumplidos	INT
	Situación Conyugal	Estado civil (Solteros, Casados, Unidos, Divorciados, etc)	VARCHAR
	Tipo de cobertura médica	Varias opciones, Obra Social, Pre-paga, sin cobertura	VARCHAR
	Sabe Leer y escribir	Campo que capta el grado de alfabetización	VARCHAR
	Asiste_establecimiento_educativo	¿Asiste o asistió a algún establecimiento educativo? (colegio, escuela, universidad) Sí, asiste No asiste, pero asistió Nunca asistió	VARCHAR
	Ese establecimiento es:	Ese establecimiento es... Público Privado Ns/Nr	VARCHAR
	Donde nació	¿Dónde nació? En esta localidad En otra localidad de esta provincia En otra provincia (especificar) En un país limítrofe (especificar Brasil, Bolivia, Chile, Paraguay, Uruguay) En otro país (especificar) Ns/Nr	VARCHAR
	Donde vivía hace 5 años	¿Dónde vivía hace 5 años? En esta localidad En otra localidad de esta provincia En otra provincia (especificar) En un país limítrofe (especificar Brasil, Bolivia, Chile, Paraguay, Uruguay) En otro país (especificar) Ns/Nr	VARCHAR
	Nivel Educativo	Nivel educativo alcanzado, las opciones son: Primaria Incompleta (incluye educación especial) Primaria Completa Secundaria Incompleta Secundaria Completa Superior Universitaria Incompleta Superior Universitaria Completa Sin instrucción Ns/Nr	VARCHAR
FK	Condición de actividad	Variable construida a partir de la categoría ocupacional y de inactividad para tener una única variable que agrupe a ocupados e inactivos	VARCHAR

Continuacion Tabla 2

Clave	Nombre del Campo	Descripcion	Tipo de dato
	Grupo de edad	Variable construida agrupando observaciones del campo Edad para facilitar el análisis: Menor a 18 a años De 18 a 29 años de 30 a 44 años 45 a 59 años Mayor a 60 años	VARCHAR
	Codigo Nivel Educativo	Variable construida a partir del Nivel educativo que lo re-convierte en una variable numerica para facilitar algunos calculos	INT

Tabla 3 - Ingreso per Cápita Familiar

Clave	Nombre del Campo	Descripcion	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observacion de la muestra	VARCHAR
PK	IPCF	MONTO DEL INGRESO PER CÁPITA FAMILIAR.	INT
	Nº de decil del IPCF	Numero de decil del IPCF en el que se encuentra el ho	INT
	PONDERA	Ponderador	INT
	IPCF Ajustado	IPCF ajustado por inflacion.	INT

Tabla 4 - Ingreso Total Individual

Clave	Nombre del Campo	Descripcion	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observacion de la muestra	VARCHAR
PK	Ingreso total Individual	MONTO DE INGRESO TOTAL INDIVIDUAL	INT
	Nº de decil de ingreso total individual	Nº de decil de ingreso total individual del TOTAL EPH	INT
	Ponderador para ingreso total individual	Ponderador para ingreso total individual (PONDIIIO en el original)	INT
	PONDERA	Ponderador de la observacion e el total de la muestra	INT
PK	Ingreso total Individual Ajustado	MONTO DE INGRESO TOTAL INDIVIDUAL ajustado por inflacion	INT
	Porcentaje de la poblacion con ingresos	Porcentaje de la poblacion (con ingresos) relevada que representa cada observacion	INT
	Grupo de Ingresos	Construida a partir de los deciles, para identificar al 10% más rico de la poblacion, al 50% más pobre, a los deciles intermedios	VARCHAR
	Suma de ingresos totales	Ingreso multiplicado el ponderador	INT
	Porcentaje del ingreso	Porcentaje del ingreso total captado por cada observacion de la muestra	INT

Tabla 5 - Ingresos de la ocupacion principal

Clave	Nombre del Campo	Descripcion	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observacion de la muestra	VARCHAR
PK	MONTO DE INGRESO DE LA OCUPACIÓN PRINCIPAL	MONTO DEL INGRESO DE LA OCUPACION PRINCIPAL	INT
	Nº de decil de ingreso de la ocupación principal	Nº de decil de ingreso de la ocupacion principal del TOTAL EPH	INT
	Ponderador del ingreso de la ocupación principal	Ponderador para ingreso total individual (PONDII en el original)	INT
	PONDERA	Ponderador de la observacion e el total de la muestra	INT
	Ingreso de la Ocupacion Principal Ajustado	Monto del ingreso de la ocupacion principal ajustado por inflacion	INT
	Suma de ingresos de la ocupacion principal	Ingreso multiplicado el ponderador	INT

Tabla 6 - Ingresos de la ocupacion principal de T. Independientes

Clave	Nombre del Campo	Descripcion	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observacion de la muestra	VARCHAR
	¿En ese negocio/empresa/actividad ¿tiene socios o familiares	Pregunta de la EPH que permite identificar si los trabajadores independientes tienen socios	VARCHAR
	Monto de ingreso de patrones y cta.propia sin socios	Monto de ingreso de patrones y cta.propia sin socios	INT
	Monto de ingreso de patrones y cta.propia con socios	Monto de ingreso de patrones y cta.propia con socios	INT
FK	PONDERA	Ponderador de la observacion e el total de la muestra	INT
	Monto de ingreso de patrones y cta.propia Ajustado	Monto de ingreso de patrones y cta.propia ajustado por inflacion	INT
PK	Suma de los ingresos de patrones y cuentapropistas	Ingreso multiplicado el ponderador	INT

Tabla 7 - Ingresos de otras ocupaciones

Clave	Nombre del Campo	Descripcion	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observacion de la muestra	VARCHAR
PK	MONTO DE INGRESO DE OTRAS OCUPACIONES	Monto de ingresos por otras ocupaciones	INT
FK	PONDERA	Ponderador de la observacion e el total de la muestra	INT
	MONTO DE INGRESO DE OTRAS OCUPACIONES Ajustado	MONTO DE INGRESO DE OTRAS OCUPACIONES Ajustado por inflacion	INT

Tabla 8 - Ingresos no laborales

Clave	Nombre del Campo	Descripción	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observación de la muestra	VARCHAR
	Monto del ingreso por JUBILACION O PENSION	Monto del ingreso por JUBILACION O PENSION	INT
	Monto del ingreso por INDEMNIZACION POR DESPIDO	Monto del ingreso por INDEMNIZACION POR DESPIDO	INT
	Monto del ingreso por SEGURO DE DESEMPLEO	Monto del ingreso por SEGURO DE DESEMPLEO	INT
	Monto del ingreso por SUBSIDIO O AYUDA SOCIAL (EN DINERO)	Monto del ingreso por SUBSIDIO O AYUDA SOCIAL (EN DINERO)	INT
	Monto del ingreso por ALQUILER DE SU PROPIEDAD	Monto del ingreso por ALQUILER DE SU PROPIEDAD	INT
	Monto del ingreso por GANANCIAS DE ALGUN NEGOCIO EN EL QUE NO TRABAJA	Monto del ingreso por GANANCIAS DE ALGUN NEGOCIO EN EL QUE NO TRABAJA	INT
	Monto del ingreso por INTERESES O RENTAS POR INVERSIONES	Monto del ingreso por INTERESES O RENTAS POR INVERSIONES	INT
	Monto del ingreso por BECA DE ESTUDIO.	Monto del ingreso por BECA DE ESTUDIO.	INT
	Monto del ingreso por CUOTAS DE ALIMENTOS O AYUDA	Monto del ingreso por CUOTAS DE ALIMENTOS O AYUDA	INT
	Monto del ingreso por OTROS INGRESOS EN EFECTIVO (LIMOSNAS ; JUEGOS DE AZAR)	Monto del ingreso por OTROS INGRESOS EN EFECTIVO (LIMOSNAS ; JUEGOS DE AZAR)	INT
	Monto del ingreso por aguinaldo.	Monto del ingreso por aguinaldo.	INT
	Monto total de ingreso no laboral percibido en ese mes	Monto total de ingreso no laboral percibido en ese mes	INT
FK	PONDERA	Ponderador de la observación e el total de la muestra	INT
	Monto del ingreso por aguinaldo ajustado	Monto del ingreso por aguinaldo ajustado por inflación	INT
	Monto del ingreso por ALQUILER DE SU PROPIEDAD Ajustado	Monto del ingreso por ALQUILER DE SU PROPIEDAD Ajustado ajustado por inflación	INT
	Monto del ingreso por BECA DE ESTUDIO Ajustado	Monto del ingreso por BECA DE ESTUDIO Ajustado ajustado por inflación	INT
	Monto del ingreso por CUOTAS DE ALIMENTOS O AYUDA Ajustado	Monto del ingreso por CUOTAS DE ALIMENTOS O AYUDA Ajustado ajustado por inflación	INT
	Monto del ingreso por GANANCIAS DE ALGUN NEGOCIO EN EL QUE NO TRABAJA Ajustado	Monto del ingreso por GANANCIAS DE ALGUN NEGOCIO EN EL QUE NO TRABAJA Ajustado ajustado por inflación	INT
	Monto del ingreso por INDEMNIZACION POR DESPIDO Ajustado	Monto del ingreso por INDEMNIZACION POR DESPIDO Ajustado ajustado por inflación	INT
	Monto del ingreso por INTERESES O RENTAS POR INVERSIONES AJUSTADO	Monto del ingreso por INTERESES O RENTAS POR INVERSIONES AJUSTADO ajustado por inflación	INT
	Monto del ingreso por JUBILACION O PENSION Ajustado	Monto del ingreso por JUBILACION O PENSION Ajustado ajustado por inflación	INT
	Monto del ingreso por OTROS INGRESOS EN EFECTIVO Ajustado	Monto del ingreso por OTROS INGRESOS EN EFECTIVO Ajustado ajustado por inflación	INT
	Monto del ingreso por SEGURO DE DESEMPLEO Ajustado	Monto del ingreso por SEGURO DE DESEMPLEO Ajustado ajustado por inflación	INT
	Monto del ingreso por SUBSIDIO O AYUDA SOCIAL Ajustado	Monto del ingreso por SUBSIDIO O AYUDA SOCIAL Ajustado ajustado por inflación	INT
PK	Monto total de ingreso no laboral percibido	Monto total de ingreso no laboral percibido	INT
	Ingreso por rentas y alquileres ajustado	Ingreso por rentas y alquileres ajustado ajustado por inflación	INT
	Suma de los ingresos por jubilación	Ingreso por jubilación multiplicado por el ponderador	INT
	Suma de los ingresos por Ayuda social	Ingreso por ayuda social multiplicado por el ponderador	INT
	Suma de los ingresos de renta de la propiedad y alquileres	Ingreso de la propiedad multiplicado por el ponderador	INT

Tabla 9 - Indicadores de formalidad de asalariados

Clave	Nombre del Campo	Descripción	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observación de la muestra	VARCHAR
PK	¿Por ese trabajo tiene descuento jubilatorio?	Variable que capta si la observación tiene descuento jubilatorio en su trabajo	VARCHAR
PK	¿Aporta por sí mismo a algún sistema jubilatorio?	Variable que capta si la observación aporta a un sistema jubilatorio	VARCHAR
	PONDERA	Ponderador de la observación e el total de la muestra	INT
	Trabajador Formal	Variable construida a partir de las respuestas sobre descuento jubilatorio para saber si una observación representa a trabajadores formales o informales	VARCHAR

Tabla 10 - Características de los ocupados

Clave	Nombre del Campo	Descripción	Tipo de dato
FK	Registro_Unico	Variable que combina CODUSU y COMPONENTE que permite asignar un código a cada observación de la muestra	VARCHAR
	Cantidad de ocupaciones.	Cantidad de ocupaciones de la observación	INT
PK	Total de horas que trabajó en la semana en la ocupación principal	Horas trabajadas en la ocupación principal	INT
	Total de horas que trabajó en la semana en otras ocupaciones	Horas trabajadas en otras ocupaciones	INT
PK	El negocio/empresa/institución/actividad en la que trabaja es	Variable que capta si la observación trabaja en una entidad privada, estatal o de otro tipo	VARCHAR
	A qué se dedica o produce el negocio/ empresa/	Rama del negocio/empresa donde trabaja la observación	INT
	Si presta servicio doméstico en hogares particulares	Variable que capta si la observación trabaja en servicio doméstico y de que tipo (1=Casa de familia)	INT
	Antigüedad en años	Antigüedad laboral	INT
PK	¿Cuántas personas, incluido...trabajan allí en total?	Personas que trabajan en el establecimiento	INT
	¿Dónde realiza principalmente sus tareas?	Lugar donde se realiza el trabajo	INT
	PONDERA	Ponderador de la observación e el total de la muestra	INT
	Tamaño establecimiento	Variable construida para dar cuenta del tamaño del establecimiento donde trabaja la observación, donde, si en el establecimiento trabajan más de 100 personas, se lo denomina 'Establecimiento Grande', caso contrario, 'Establecimiento Chico'	VARCHAR

Transformaciones de datos realizadas

- La primera transformación, y la única realizada por fuera de Power BI, es la creación de un registro único de observaciones: Dado que dentro de un hogar (identificado por CODUSU) puede haber varios componentes, concatenamos CODUSU y Componente usando la función de Excel Concatenate para tener un campo 'Registro único' que identifica cada línea de la muestra.
- Para facilitar la legibilidad, se reemplazaron nombres de variables por otros mucho más descriptivos, por ejemplo,

P47T es el monto del ingreso total individual, es decir, el ingreso total de cada individuo por todas las fuentes.

P21 es el ingreso de la ocupación principal, etc.

El archivo complementario 'Glosario Variables EPH' contiene los nombres originales de los campos junto a su descripción que permite llegar al nombre de los campos en el archivo PBIX.

- Siguiendo con lo anterior, muchas categorías tienen valores numéricos para describir variables cuantitativas, la transformación en Power BI implicó reemplazar los valores numéricos, que hubieran requerido que un usuario deba revisar el glosario de variables constatemente, por valores que describen la variable, por ejemplo, la variable Nivel_Ed, o Nivel educativo, originalmente tenía valores numéricos para cada tipo de respuesta,

Campo	Descripcion
NIVEL_ED	Nivel educativo
	1 = Primaria Incompleta (incluye educación especial)
	2 = Primaria Completa
	3 = Secundaria Incompleta
	4 = Secundaria Completa
	5 = Superior Universitaria Incompleta
	6 = Superior Universitaria Completa
	7 = Sin instrucción
	9 = Ns/Nr

En este caso, lo que hice fue reemplazar los valores numéricos por el texto descriptivo.

Otro ejemplo son las categorías de actividad e inactividad:

Campo	Descripcion
CAT_INAC	Categoría de inactividad
	1 = Jubilado / Pensionado
	2 = Rentista
	3 = Estudiante
	4 = Ama de casa
	5 = Menor de 6 años
	6 = Discapacitado
CAT_OCUP	7 = Otros
	Categoría ocupacional (para ocupados y desocupados con ocupación anterior)
	1 = Patrón
	2 = Cuenta propia
	3 = Obrero o empleado
	4 = Trabajador familiar sin remuneración
	9 = Ns/Nr

Para entender todas estas transformaciones, se debe consultar el archivo complementario “Glosario Variables EPH” que se adjunta con la entrega, ya que detallarlas todas en este documento tomaría mucho espacio y sería poco práctico de leer.

Además, hay una columna creada que agrupa categorías de actividad e inactividad, ya que en el original todos los no ocupados figuran como ‘Sin Ocupación’ mientras que la categoría de inactividad está en otra columna, agruparlas en una única columna será muy útil para el análisis como se verá más adelante, el código es:

Condición de actividad =

```
if ('Tabla 1 - Características de la muestra'[Categoría Ocupacional]="Sin Ocupación",  
'Tabla 1 - Características de la muestra'[Categoría de Inactividad], 'Tabla 1 -  
Características de la muestra'[Categoría Ocupacional]  
  
)
```

Ajuste de las variables de ingreso por inflación

Como la base de datos es del primer trimestre de 2023, se ajustaron todos los valores de montos de ingresos de la población por el IPC entre enero de 2023 y junio de 2023, usando la variación del Índice de Precios al Consumidor (IPC) publicada por el Indec. Esto da como resultado un ajuste del 42.2% de todas las variables de ingreso

Total nacional	Nivel general	IPC - Enero 2023=100
Jun-23	6.0	142.2
May-23	7.8	134.2
Apr-23	8.4	124.5
Mar-23	7.7	114.8
Feb-23	6.6	106.6
Jan-23	6.0	100.0

Todas las variables de ingreso que incluyen la palabra ‘ajustado’ en la base de datos fueron transformadas mediante este procedimiento. El código simplemente (para tomar un ejemplo) es:

```
Ingreso total Individual Ajustado = 'Tabla 4 - Ingreso Total Individual'[Ingreso total  
Individual]*1.422
```

Suponer que todos los ingresos de la población se movieron junto con el IPC en la misma medida es un supuesto muy importante, el motivo por el que lo hacemos es para intentar traer los valores (desactualizados) del primer trimestre de 2023 a precios que resulten familiares a un usuario de la base de datos que utiliza el dashboard en la segunda mitad de 2023. De todos modos, dado el contexto inflacionario, incluso el nivel de precios a junio resulta difícil de comparar con el nivel de

ingresos y gastos que un usuario pueda tener en septiembre de 2023, el valor principal del trabajo radica en la identificación de los ratios entre distintos ingresos, la distribución del mismo y su relación con variables demográficas, sociológicas y geográficas, que son relativamente más estables que el nivel de precios.

A principios de 2024 cuando el Indec publique los resultados de la EPH del tercer trimestre de 2023 podremos ver en cuanto se desviaron las estimaciones arrojadas por este método de la realidad.

Creación de grupos de edad

El campo edad detalla las edades de cada observación, para poder facilitar el análisis se creó el campo Grupo de edad con los siguientes valores:

Menor a 18 años.

18 a 29 años.

30 a 44 años.

45 a 59 años.

60 años y más.

El código es:

Grupo de edad =

```
IF(
  [Edad] < 18, "Menor a 18 años",
  IF(
    [Edad] < 30, "18 a 29 años",
    IF(
      [Edad] < 45, "30 a 45 años",
      IF(
        [Edad] >= 60, "Mayor a 60 años",
        "45 a 59 años"
      )
    )
  )
)
```

Grupos de ingresos

La propia EPH nos da los deciles a los que pertenece cada observación, a través del campo **Nº de decil de ingreso total individual**.

Se procedió a elaborar los siguientes estratos de distribución del ingreso:

- 10% más rico de la población: Decil 10.
- Deciles 5 a 9, que representan al 40% intermedio de la distribución del ingreso.
- 50% más pobre de la población.

Ingresos totales

Cuando analicemos la distribución del ingreso, nos va a interesar, entre otras cosas, qué porcentaje del ingreso total se lleva cada grupo, para eso necesitamos multiplicar los ingresos por el total de población que representa cada observación, un ejemplo representativo de estas transformaciones en diversas categorías de ingreso es el siguiente:

```
Suma de ingresos totales = 'Tabla 4 - Ingreso Total Individual'[Ingreso total Individual Ajustado]*'Tabla 4 - Ingreso Total Individual'[PONDERA]
```

Formalidad laboral

Una categoría fundamental para analizar el mercado laboral y la sociedad argentinas es la informalidad laboral, habida cuenta el elevadísimo porcentaje de trabajadores informales y el hecho de que, como demuestra el dashboard, tienen características personales muy disímiles respecto a los trabajadores formales. La EPH no pregunta directamente por la formalidad, pero en este trabajo usamos la pregunta sobre descuento jubilatorio para construir un indicador de formalidad en la tabla 9:

Trabajador Formal =

```
IF(
  [¿Por ese trabajo tiene descuento jubilatorio?] = "N/A", "Inactivo",
  IF(
    [¿Por ese trabajo tiene descuento jubilatorio?] = "Si", "Formal" ,

    IF('Tabla 9 - Indicadores de formalidad de asalariados'[¿Por ese trabajo tiene
    descuento jubilatorio?]="No", "Informal"

  )
)
```

Medidas creadas

Las medidas creadas se pueden agrupar, en su mayoría, en 2 conjuntos, medianas de ingresos y porcentaje de la población que percibe un determinado tipo de ingreso.

Para no sobrecargar este archivo de documentación, voy a mostrar el ejemplo de las jubilaciones, que puede generalizarse a las restantes medianas de ingreso y porcentajes de la población que percibe un determinado tipo de ingreso:

Mediana de las jubilaciones =

```
CALCULATE(MEDIAN('Tabla 8 - Ingresos no laborales'[Monto del ingreso por JUBILACION O PENSION Ajustado])
, 'Tabla 8 - Ingresos no laborales'[Monto del ingreso por JUBILACION O PENSION Ajustado]>0
)
```

Poblacion que percibe jubilación =

```
CALCULATE(SUM('Tabla 1 - Características de la muestra'[Cantidad de habitantes])
,
'Tabla 8 - Ingresos no laborales'[Monto del ingreso por JUBILACION O PENSION Ajustado]>0
)
)
```

Porcentaje de jubilados =

```
[Poblacion que percibe jubilación]/SUM('Tabla 1 - Características de la muestra'[Cantidad de habitantes])
```

Los otros 3 tipos de medidas calculadas que ayudarán en nuestro análisis son:

Horas Trabajadas en la ocupacion principal =

```
MEDIAN('Tabla 10 - Características de los ocupados'[Total de horas que trabajó en la semana en la ocupacion principal])
```

Ingreso por hora trabajada =

```
[Mediana de los ingresos de la ocupacion principal]/[Horas Trabajadas en la ocupacion principal]
```

Y finalmente la mediana de la edad, como veremos más adelante, algunos distritos tienen una población bastante más envejecida (CABA en particular), o menos joven, que otras, y la edad está fuertemente correlacionada con el nivel de ingresos:

Mediana de la Edad = `MEDIAN('Tabla 2 - Características de los miembros del Hogar'[Edad])`

Finalmente, aproveché que la EPH pregunta por el número de personas que trabajan en el establecimiento donde se desempeñan los trabajadores, construyendo un indicador a partir de la pregunta '¿Cuántas personas, incluido...trabajan allí en total?' en la tabla 10, donde establecimiento grande agrupa a los establecimientos de más de 100 personas,

Tamaño establecimiento =

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="12", "Establecimiento Grande",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="11", "Establecimiento Grande",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="10", "Establecimiento Grande",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="9", "Establecimiento chico",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="8", "Establecimiento chico",

if('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="7", "Establecimiento chico",

if('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="6", "Establecimiento chico",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="5", "Establecimiento chico",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="4", "Establecimiento chico",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="3", "Establecimiento chico" ,

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="2", "Establecimiento chico",

IF('Tabla 10 - Características de los ocupados'[¿Cuántas personas, incluido...trabajan allí en total?]="1", "Establecimiento chico",

"No aplica / Inactivo / Sin Respuesta"

Los códigos numéricos en la EPH son:

¿Cuántas personas, incluido...trabajan allí en total?

- 1 = 1 persona
- 2 = 2 personas
- 3 = 3 personas
- 4 = 4 personas
- 5 = 5 personas
- 6 = de 6 a 10 personas
- 7 = de 11 a 25 personas
- 8 = de 26 a 40 personas
- 9 = de 41 a 100 personas
- 10 = de 101 a 200 personas
- 11 = de 201 a 500 personas
- 12 = más de 500 personas
- 99 = Ns/Nr

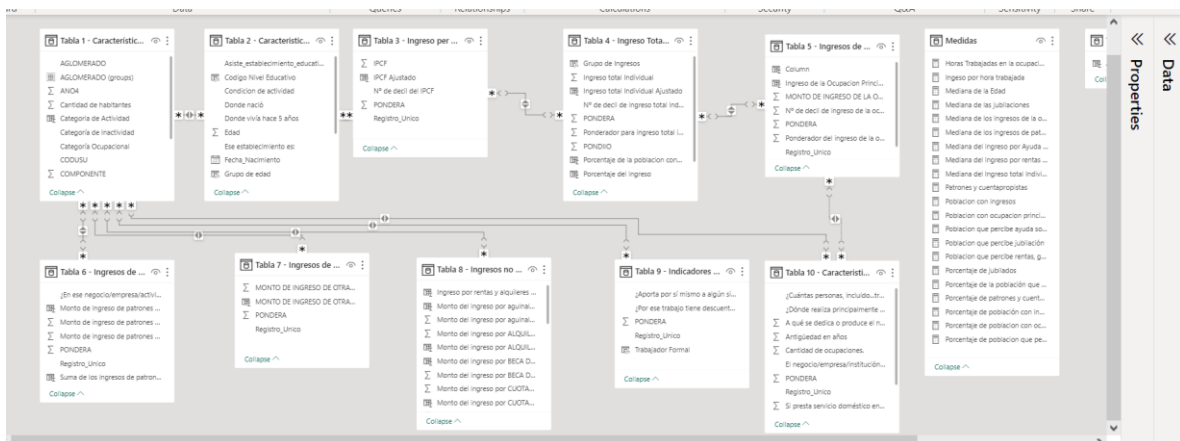
Tabla calendario

La variable temporal que tenemos en la serie es el año de nacimiento, para ello se creó una tabla calendario, por ejemplo, para el año de nacimiento (meses y días siguen el mismo procedimiento)

```
Anio = YEAR('Tabla calendario'[Fecha_Nacimiento].[Date] )
```

Modelo relacional

Habiendo descrito el modelo entidad-relación, nuestras tablas, campos y medidas creadas, estamos en condiciones de ver las relaciones entre tablas:



Como puede verse, todas las tablas están relacionadas entre sí a través de numerosas variables, entre las que se destaca el registro único de cada observación.

Segmentaciones elegidas

Las principales segmentaciones están relacionadas a las características demográficas y sociológicas, ya que el objetivo del trabajo es comparar el nivel de ingresos de distintos grupos de la población y su participación en la distribución del ingreso.

La base tiene una gran riqueza en este aspecto, que nos permite analizar el ingreso desde una variedad de ángulos, veremos más adelante que todas estas segmentaciones son relevantes y muy significativas en su correlación con los ingresos:

Aglomerado, Sexo, Lugar de nacimiento ('Donde nació'), Migraciones (Donde vivía hace 5 años), Condición de actividad, Grupo de edad y Grupo de ingresos

AGLOMERADO

All

Search

- ☒ Select all
- ☒ Bahía Blanca
- ☒ CABA
- ☒ Comodoro ...
- ☒ Concordia
- ☒ Corrientes

Sexo

All

[Back to report](#)

Donde nació

All

Search

- ☒ Select all
- ☒ En esta localidad
- ☒ En otra localidad de esta provincia
- ☒ En otra provincia (especificar)
- ☒ En otro país (especificar)
- ☒ En un país limítrofe

Sexo

All

- ☐ Mujer
- ☐ Varón

Condición de actividad

All

Search

- ☐ Select all
- ☐ Activo
- ☐ Ama de casa
- ☐ Cuenta propia
- ☐ Discapacitado
- ☐ Estudiante

Amo de casa

Cuenta propia

Donde vivía hace 5 años

All

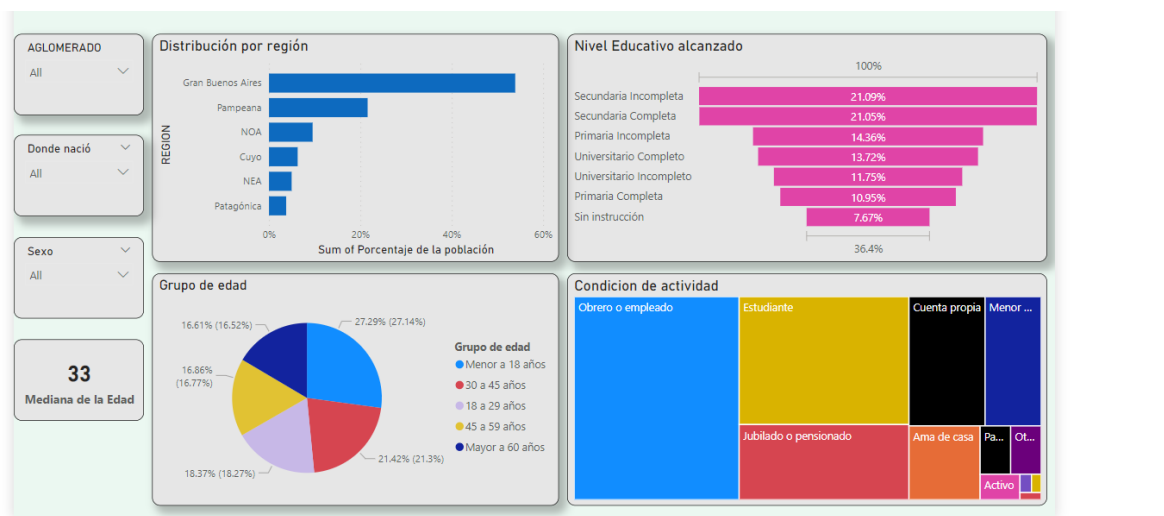
- ☐ Select all
- ☐ En otra localidad de esta provincia
- ☐ En esta localidad
- ☐ En otra provincia
- ☐ En otro país
- ☐ s/Nr
- ☐ Sin respuesta

Explicación del tablero y principales conclusiones

En esta sección vamos a recorrer las 4 páginas del tablero y destacar algunas de las conclusiones a mi juicio más relevantes, pero de ningún modo esto agota todas las posibilidades de análisis que ofrece el dashboard dada la gran cantidad de variables y filtros posibles:

➤ Demografía

Esta página da una visión general de la composición de la muestra y una caracterización a grandes rasgos de variables sociodemográficas.



En el aspecto geográfico, puede observarse la absoluta preponderancia del Gran Buenos Aires y la región pampeana, el primero en particular concentra el 53% de la población relevada. Recordemos que la EPH sólo encuesta los aglomerados más poblados y deja fuera al 37% de la población del país, de modo que el Gran Buenos Aires en particular está sobrerrepresentado, según los datos del censo 2022, CABA y los 24 partidos del GBA en realidad concentran el 30% de la población.

Yendo a la distribución por edades, vemos que la Argentina está atravesando un proceso de envejecimiento, ya que el 16% de la muestra tiene más de 60 años, frente a un 27% de menores de 18 años. La edad mediana es de 33 años, para poner en el contexto internacional, Italia, como modelo de sociedad que concluyó la transición demográfica, tiene una mediana de 47 años, en el otro extremo, Nigeria tienen una edad mediana de 18 años, en el contexto regional, las medianas de Brasil, Chile y Uruguay están en los 35 años, ver fuente (sección UN Figures)

https://en.wikipedia.org/wiki/List_of_countries_by_median_age

Nótese que en esta fuente la mediana para argentina es de 31.5 años, algo ligeramente a nuestra estimación, tal vez la población no representada en la EPH es algo más joven que la población de grandes centros urbanos, tal vez la diferencia se origine en que el último dato de Naciones Unidas es de 2020, etc.

Con respecto al nivel educativo, predominan el secundario completo e incompleto, lo cual tiene cierta lógica si tenemos en cuenta que el 27% de la muestra es menor a 18 años, si excluimos ese segmento de edad, vemos que el porcentaje de personas 'Sin instrucción' baja del 7.67% al 0.35%, mientras que, por caso, el universitario completo (que va a ser muy relevante cuando analicemos ingresos) pasa del 13.7% a casi 19% de la población.

En la categoría de actividad, hay un predominio de trabajadores en relación de dependencia, seguido de estudiantes (recordemos que es una muestra de toda la población, incluyendo un 27% de menores de 18 años), jubilados y pensionados, mucho más atrás aparecen cuentapropistas y amas de casa. Patrones y empleados son el 1.56% de la población.

Nótese la amplia cobertura del sistema previsional, el 13.5% de la muestra es jubilado o pensionado, compárese con el 16.5% de personas mayores de 60 años que señalábamos en párrafos anteriores, si se tiene en cuenta que para varones la edad de jubilación es de 65 años, podemos concluir que la cobertura del sistema previsional es casi total.

➤ **Distribución del ingreso**

En esta sección vamos a analizar la distribución del ingreso, he dividido a la población en 3 grupos,

- 10% más rico de la población, que tiene una mediana de ingresos totales individuales de **\$426.000**, que capta el **30%** del ingreso total.
- 50% más pobre de la población, con un ingreso mediano de **\$75.000**, que percibe el **23%** del ingreso total.
- Deciles intermedios, el 40% de la población, con un ingreso de **\$191.000**, **47%** del total

La mediana de ingresos de la población por todo tipo de fuente es de **\$115.964**. Recordemos que las cifras en pesos **están ajustadas al nivel del IPC de junio de 2023**.

Entonces, el 10% de la población tiene un ingreso mediano que es 5.6 más alto que el 50% más pobre de la población.

¿Es este un nivel elevado de desigualdad?, tomando en cuenta las dificultades de hacer comparaciones internacionales, y de acuerdo a esta fuente:

<https://ourworldindata.org/grapher/income-share-of-the-top-10-pip>

el 30% que percibe el 10% más rico de la población en Argentina es similar al porcentaje de Estados Unidos, algo más alto que en Europa (el índice de Alemania, por caso, es de 25%) y bastante inferior al de los países más grandes de Latinoamérica, como Brasil (41%), México y Chile (ambos en 35%)

Podría decirse, tentativamente, que el nivel de desigualdad en Argentina se encuentra en un punto intermedio en la comparación internacional.

Más allá de la desigualdad, ¿Cuáles son las fuentes de ingresos de los argentinos?

Fundamentalmente salarios o ingresos de la ocupación principal si son independientes y jubilaciones.

En la introducción y más adelante en las conclusiones mencioné que las encuestas de hogares no captan adecuadamente los ingresos de sectores ligados a la propiedad de grandes empresas, activos

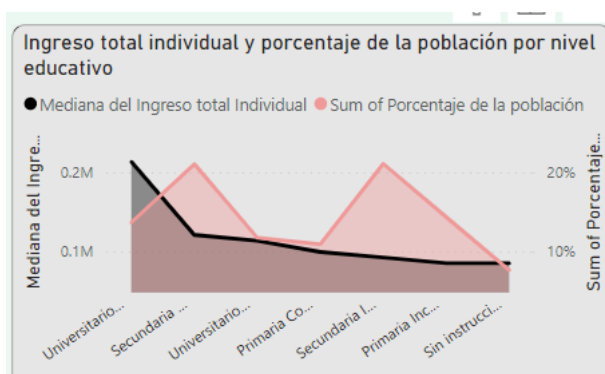
financieros e inmobiliarios, etc. Es en esta sección donde podemos ver por qué. En la tarjeta de ingresos de patrones y cuentapropistas vemos que la mediana del ingreso es de \$85.000, bastante inferior a la mediana general, y representan un porcentaje muy reducido de la población (no llega al 1% de la muestra). Uno esperaría encontrar en esta categoría a muchos de los individuos de mayores ingresos de la sociedad, pero 'Patrones y Cuentapropistas' es una categoría demasiado general que agrupa a individuos en lo alto de la pirámide social junto a trabajadores independientes del sector informal. En el apartado final de este trabajo volveremos en detalle sobre este punto.

Una observación interesante es que la distribución de las jubilaciones es menos desigual que los ingresos laborales, la mediana de las jubilaciones del 10% más rico (\$300.000) es 4.2 veces más alta que las jubilaciones del 50% más pobre (\$71.000), mientras que para los ingresos por todo concepto principal el ratio es de más de 5.

En cuanto a la ayuda social, vemos que el 6% de la población declara percibir transferencias de este tipo, y el monto mediano (\$28.000) resulta muy inferior a otras fuentes de ingresos, uno puede concluir en base a esto que la ayuda social complementa, más que reemplazar, los ingresos laborales o por jubilaciones.

En el resto de la página comparamos los ingresos con algunas variables demográficas.

En primer lugar, nótese el fuerte gradiente de ingresos en el nivel Universitario completo, básicamente, el 10% más rico de la población se encuentra conformado por personas con este nivel educativo.



Luego observamos el tipo de cobertura médica, este indicador sintetiza la informalidad laboral, y vemos que el 32% de la población no tiene cobertura más allá del servicio público, si hacemos foco en este sector, vemos que su ingreso es de \$71.000, muy inferior a la mediana global (\$115.000).

Finalmente, aproveché que la EPH pregunta por el número de personas que trabajan en el establecimiento donde se desempeñan los trabajadores, construyendo un indicador a partir de la pregunta '¿Cuántas personas, incluido...trabajan allí en total?' en la tabla 10, donde establecimiento grande agrupa a los establecimientos de más de 100 personas, como se detalla en la sección de Medidas más arriba. Vemos que los ingresos de personas que trabajan en establecimientos grandes son bastante más altos que los ingresos de personas que trabajan en establecimientos chicos.

Esto concuerda con la literatura de la economía laboral, los establecimientos más grandes suelen representar a las empresas más productivas y exitosas de la economía, que suelen emplear a los

trabajadores más calificados y ofrecer las mejores remuneraciones, a esto se suma que en establecimientos grandes se encuentran la mayoría de los empleados estatales, que, como veremos más adelante, suele tener remuneraciones más altas que el promedio.

➤ **Genero**

El primero de nuestros estudios de caso antaño a las diferencias de género.

Sin agotar todas las posibilidades que ofrece el dashboard, podemos concluir en una primera vista de la página que,

1. Las mujeres tienen a tener un nivel educativo más elevado que los varones, obsérvese por caso la muy significativa diferencia en el nivel Universitario completo.
2. Las mujeres tienen una participación laboral menor a la de los varones, y, en particular, si hacemos foco en el rubro de patrones y cuentapropistas, vemos que hay el triple de varones, mientras que en el nivel de amas de casa el predominio femenino es total.
3. El nivel de ingresos laborales, por ocupación principal, de las mujeres, es significativamente inferior al de los varones, concretamente un 27% más bajo:

Ingresos de la ocupación principal				
Sexo	Sum of Porcentaje de la poblacion con ingresos	Mediana de los ingresos de la ocupacion principal	Horas Trabajadas en la ocupacion principal	Ingreso por hora trabajada
Varón	49%	135,090	40	3,377
Mujer	52%	99,540	30	3,318
Total	100%	113,760	40	2,844

Sin embargo, aprovechando que la EPH pregunta por las horas trabajadas en la ocupación principal, pude estimar una mediana de horas trabajadas en función del género, vemos que las mujeres trabajan un 25% de horas menos que los varones, lo cual en cierto modo enlaza con la participación laboral que discutíamos en el apartado anterior, si combinamos las horas trabajadas con el ingreso declarado, vemos que el ingreso por hora trabajada es muy similar entre varones y mujeres, a simple vista, esto implicaría que el género afecta los resultados económicos fundamentalmente a través de la participación laboral. No obstante, más trabajo estadístico y de recolección de datos se requiere en este punto, puesto que hemos visto anteriormente que a) el nivel educativo de las mujeres es más alto que el de los varones y b) que el nivel educativo está muy fuertemente correlacionado con los ingresos, si no hubiera diferencias en el mercado laboral, deberíamos entonces observar un ingreso laboral horario más alto para las mujeres, en lugar de la paridad que muestra el dashboard.

En mi opinión, se necesitan estudios con un nivel de detalle y desagregación más grandes que lo que permite la muestra para arribar a conclusiones definitivas sobre la relación entre género y mercado laboral.

➤ **Diferencias geográficas**

Finalmente, aprovechamos la información a nivel de aglomerado para estudiar las diferencias geográficas.

Podemos observar que las medianas de ingreso total individual son muy variables entre aglomerados, con la Patagonia y CABA siendo los lugares más ricos, las ciudades de la región central del país en una posición intermedia y las ciudades de NOA y NEA en los puestos más bajos.

El nivel de desigualdad geográfica es extraordinariamente significativo, en Ushuaia los ingresos (\$227.000) casi triplican los de Resistencia (\$85.000).

La estructura económica (que no es relevada en la EPH) y la presencia del estado, ya sea como empleador o a través de transferencias como jubilaciones y ayuda social, son muy importantes a la hora de explicar estos resultados.

Esta página del dashboard incluye además gráficos por nivel educativo, formalidad laboral (ambos podrían funcionar como proxy de la estructura económica, hasta cierto punto) y segmentaciones por género, edad, condición de actividad y migraciones internas y externas ('Donde vivía hace 5 años') que permiten al usuario explorar la enorme heterogeneidad que presentan las distintas regiones argentinas.

En este punto, y para ilustrar la utilidad de los gráficos presentados, quiero hacer una comparación entre dos aglomerados geográficamente muy cercanos, que de hecho forman una única área metropolitana, pero que resultan muy disímiles, como lo son CABA y los Partidos del GBA.

Desde el punto de vista de los ingresos, en CABA la mediana es de \$195.000, un 72% más alto que los \$113.000 de los partidos del GBA, donde el ingreso mediano está muy cerca del total nacional.

Veamos algunas características demográficas relacionadas,

- La edad mediana es de 40 años en CABA frente a los 34 años del GBA, si volvemos a nuestro análisis de ingreso y edad en CABA observamos mayor presencia de los grupos de 30 a 45 y 45 a 60 años que tienen mayores ingresos.
- Mucho más relevante aún, el nivel educativo de los residentes de CABA es significativamente más alto que en el GBA, por caso, para el nivel universitario completo, vemos que el 30% de los residentes de CABA alcanzan ese nivel, (3% de la población total cuando seleccionamos CABA, mientras que el distrito tiene el 10% de la población de la muestra). Para el GBA, la cifra es de 10%, ya que los profesionales representan 4.19% de la población total cuando seleccionamos el aglomerado, que a su vez concentra el 43% de la población relevada. El promedio nacional, recordemos, es de 13.72% para esta categoría. Y recordemos también que los individuos con este nivel educativo son quienes pueblan fundamentalmente el decil más rico, como se vio en un apartado anterior.
- Desde el punto de vista del empleo público, el 8% de la muestra nacional trabaja en el Estado, si tenemos en cuenta que los activos son el 45% de la población relevada, entonces el peso de esta categoría es de casi 18%. En CABA, el 19% de la población trabaja en el sector público (sacar la proporción de empleo público sobre la población total del distrito restando a los inactivos), mientras que en el GBA dicha proporción es de menos del 14%. Se puede constatar en la página de distribución del ingreso que las remuneraciones son más altas en el sector público, y si exploráramos las opciones que nos ofrece el dashboard, constataremos que el empleo público tiene una mayor proporción de personas de nivel

educativo alto, que la formalidad es casi total, mientras que el sector informal de la economía, que concentra los ingresos bajos, es un fenómeno fundamentalmente privado. Más en general, observando esta categoría no parece ser cierto que el empleo estatal tenga más peso en el interior o fuera de las capitales provinciales donde se concentran las administraciones, en los aglomerados más pobres del país, más que el empleo público, se observa una preponderancia de la informalidad laboral.

- Un tema que me resulta interesante destacar es el peso de las migraciones, a nivel nacional, podemos ver que el 85% de la población nació en la misma provincia donde vive, sea en la misma localidad (casi 80%), o bien en otra localidad de la provincia, el 10% nació en otra provincia respecto de donde reside, mientras que el 5% de la muestra nació en el extranjero.

A nivel GBA (es decir, CABA más partidos del Conurbano), los migrantes internos representan también cerca del 10% de la población, en donde observamos una participación mayor bastante mayor que el promedio nacional es en la categoría de migrantes internacionales, que representan el 7.2% de la población de la región, contrástese esta cifra con el 5% del promedio nacional, y concluiremos que la región Gran Buenos Aires es el destino preferido de los migrantes internacionales.

Conclusiones

A lo largo de este trabajo hemos recorrido la estructura sociodemográfica y de ingresos de la población argentina tal como es relevada por la EPH. Las conclusiones a las que puede arribar un usuario del tablero son innumerables dada la gran cantidad de información y slicers que tenemos, aquí sólo me enfocaré en sintetizar lo más brevemente posible las conclusiones que rápidamente surgen de una primera observación de las distintas páginas:

- La EPH sólo cubre al 63% de la población del país, esto redundo en una sobrerrepresentación de la población urbana y en particular del Area Metropolitana de Buenos Aires.
- La población argentina está atravesando un proceso de envejecimiento, estando a mitad de camino entre los países más jóvenes del mundo, fundamentalmente africanos, y los más envejecidos, como los países europeos. A nivel latinoamericano está dentro del promedio.
- El nivel educativo de la población fundamentalmente se concentra a nivel secundario, sea completo e incompleto, sólo un cuarto de la población de 30 a 60 años tiene Universitario completo.
- El nivel educativo está fuertemente correlacionado con los ingresos, hemos visto que para esa categoría los ingresos casi duplican la media nacional.
- La EPH es insuficiente para estudiar los ingresos de lo más alto de la pirámide, fundamentalmente cuando se trata de rentas de la propiedad, sea de sociedades, inmuebles, campos, depósitos bancarios en el país y en el extranjero, etc.

- En cambio, la EPH es un excelente instrumento para estudiar los ingresos laborales y las transferencias estatales, como son jubilaciones y ayuda social.
- Los argentinos derivan fundamentalmente sus ingresos de su trabajo, donde mayormente se desempeñan como empleados, con un alto grado de informalidad (no menos de un tercio de la población no tiene descuento jubilatorio en su empleo). La segunda fuente de ingresos son las jubilaciones, habida cuenta que un sexto de la población tiene más de 60 años y la cobertura previsional es casi universal. En cambio, la ayuda social sólo abarca al 6% de la muestra y está muy por debajo de los ingresos laborales medianos, incluyendo los informales.
- El 10% más rico de la población concentra el 30% del ingreso y gana más de 5 veces más el promedio del 50% más pobre de la población. En la comparativa internacional, la impresión que queda es que Argentina ocupa un lugar intermedio entre los países más equitativos del norte de Europa y los más desiguales, como Brasil, Colombia o Sudáfrica. En todo caso esta es un área de análisis que requiere de más y mejores datos para llegar a conclusiones definitivas.
- Las jubilaciones tienden a tener una distribución algo más igualitaria que los ingresos laborales.
- Los ingresos en el sector público son significativamente más altos que en el sector privado, en particular en relación a los trabajadores informales.
- El análisis de género arroja, al menos, dos aristas interesantes, por un lado, las mujeres tienen un nivel educativo más alto que los hombres pero perciben menores ingresos. Si bien la diferencia total de ingresos parece estar relacionada con las horas trabajadas, ya que el ingreso horario es casi idéntico, resulta sugestivo que el mayor nivel educativo de las mujeres no se traduzca en un mayor ingreso laboral por hora. En este punto creo que se necesitan análisis adicionales que exceden el alcance de la muestra y de este trabajo, como, por caso, qué tipo de carreras eligen hombres y mujeres y cuáles son sus perspectivas de ingresos, cuál es la influencia del tiempo dedicado al cuidado de los hijos por ambos géneros, etc.
- Existen muy marcadas diferencias geográficas. En los distritos más ricos del país los ingresos más que duplican a los de los distritos más pobres. El análisis que hicimos comparando CABA con partidos del GBA parecen apuntar a una correlación con el nivel educativo, el empleo público y la formalidad laboral en cada distrito, pero, desde mi punto de vista, necesitamos una buena encuesta que analice en detalle la estructura económica de cada provincia para terminar de comprender este fenómeno. Hasta donde llega mi conocimiento, el Indec no publica encuestas de ese estilo, los datos que disponemos son muy generales y no son fáciles de agrupar con la muy detallada EPH.

Futuras líneas de investigación

Habiendo recorrido nuestra base de datos y las posibilidades de análisis que brinda, existen, a mi juicio, 4 direcciones en las que se necesita mejorar, o directamente implementar, una recolección de datos que nos permita comprender más acabadamente la composición de la población argentina y sus ingresos:

1. Cobertura geográfica

Como se ha descrito anteriormente, la EPH releva apenas algo más del 60% de la población argentina, aquella que vive en los aglomerados más grandes. Muy poco es lo que sabemos de la composición, características y situación de ingresos de la población rural (casi 10% del total según el censo 2010) y aquellas personas que viven en aglomerados 'chicos', que representan casi 30% de la población. Es urgente extender la EPH para cubrir a la totalidad de la población para tener una representación más clara de la sociedad.

2. Ingresos de la propiedad y rentas

Como vimos, la EPH resulta un instrumento relativamente pobre para captar los ingresos de los propietarios de grandes empresas y otras rentas de la propiedad como pueden ser intereses por plazos fijos, tenencia de bonos de deuda pública, distribución de utilidades entre los propietarios de acciones de empresas, etc. Más aún, es ampliamente conocido que buena parte de la riqueza propiedad de argentinos está radicada fuera del país, indudablemente los dueños de esos activos reciben algún tipo de renta por su propiedad, pero la EPH no permite distinguirlo.

En mi opinión, esto se debe a que la EPH descansa sobre las respuestas de personas contactadas por el Indec en forma aleatoria, por un lado, existe una amplia subdeclaración de ingresos en los estratos más altos y en general en aquellos ingresos no relacionados con salarios o jubilaciones que son más fáciles de calcular en forma mensual. Por otro lado, el porcentaje de población que percibe ingresos muy significativos por rentas de la propiedad es bastante reducido, una muestra aleatoria de cerca de 48 mil observaciones, como tiene nuestra base, para representar a 29 millones de personas (que, como vimos, tampoco es la totalidad del país, sólo el 63%) tiene chances muy bajas de contactar en forma consistente a personas con ingresos muy elevados relacionados con rentas de la propiedad, la evidencia internacional indica que estaríamos hablando de bastante menos del 1% de la población. Estas debilidades son muy conocidas a nivel internacional, es decir, no es un defecto de la EPH sino de las encuestas de hogares en general. Hasta donde tengo conocimiento, en países desarrollados se intenta, con diversos grados de éxito, analizar los ingresos de la parte más alta de la distribución a través de información administrativa y tributaria, como pueden ser los pagos por impuesto a las ganancias de las sociedades, no tengo conocimiento de que existan esfuerzos tendientes a construir bases de datos de ese estilo en Argentina.

3. Distribución de la riqueza

Como se puede apreciar, la base de datos y todo el análisis se centró en los ingresos, mientras que poco sabemos de la distribución de la riqueza, es decir, de la propiedad de empresas privadas, inmuebles para vivienda, comerciales, campos, tenencias financieras

etc. Esto se relaciona directamente con el punto anterior, la experiencia internacional indica que la riqueza está mucho más desigualmente distribuida que los salarios, y más aún que las transferencias estatales como son las jubilaciones, y, dado que la propiedad es fuente de ingresos, nuestro análisis de la distribución de los ingresos se enriquecería en gran medida.

4. Grupos poblacionales no definidos en la EPH y minorías

Si bien la base de datos nos ha permitido analizar la composición de la población y su situación de ingresos en forma muy variada, podrían introducirse preguntas adicionales que nos permitan ahondar más en las características sociales y demográficas, por ejemplo, no tenemos información sobre personas que se perciben como no binarias, tampoco tenemos datos sobre minorías étnicas y religiosas o de otro tipo.