

DATA SCIENCE

INFORME FINAL ANDEMSOC ANÁLISIS Y VISUALIZACIÓN DE DATOS

Analía Elizabeth Gomory:

- ae.gomory@gmail.com
- <http://www.linkedin.com/in/elizabeth-gomory-9240b1217>
- <https://ae-gomory.itch.io>
- <https://github.com/ElizabethGomory>
- <https://www.credly.com/users/analia-gomory>
- <https://orcid.org/0009-0003-0498-7658>

INTRODUCCIÓN

Una empresa multinacional decidió instalarse en el país y necesita entender sus distintos aspectos económicos y sociales.

Para ello, contamos con bases de datos públicos de la oficina nacional (INDEC).

Nos decidimos a separar el siguiente informe en distintos reportes. El primero, será sobre el perfil productivo y exportador del país, con especial foco en los grandes rubros de exportaciones y en las provincias de origen.

En el segundo reporte, trabajaremos con un análisis demográfico y social, en la que crearemos una Base de Datos con los archivos proporcionados, con los que reconocemos y analizamos la densidad poblacional, rasgos geográficos, económicos, sanitarios y de género.

A partir de estos resultados, se llegaron a distintas conclusiones partiendo del punto de, según el tipo de empresa que se quiera montar en nuestro país, va a convenir una u otra provincia.

Para poder fundamentar esta afirmación, nos basamos en el estudio guiando a la empresa contratante sobre el comportamiento de las exportaciones, los puertos que tendrán en cuenta en caso de necesitarlos, la variable de las exportaciones según el tipo de explotación de recursos que se realice, y en caso que quieran producir en nuestro país, hicimos un reconocimiento de nuestra geografía y demografía que les serán útiles, al momento de tomar decisiones.

Metodología

Para realizar los análisis correspondientes se utilizaron bases de datos entregadas por el INDEC.

Para lograr una mejor visualización al momento de observar y hacer un análisis preliminar de los datos, se utilizaron los Software de Tableau, Jupyter y Visual Studio Code.

Los mismos serán utilizados para poder validar un análisis preliminar. Luego, continuaremos con la creación de nuevas Bases de Datos donde se complementen los distintos archivos iniciales, permitiéndonos crear parámetros de correlación, comparación e información, complementadas con gráficas acordes para su visualización y entendimiento.

Análisis preliminar

Observando y estudiando los datos del INDEC – en crudo – pudimos llegar a distintas conclusiones, las cuales, corroboraremos en reportes posteriores.

Para el segundo reporte, se necesitó crear una nueva base de datos que nos permitiera analizar, en una sola tabla, la conjugación de todos los datos proporcionados. Para ello, se decidió trabajar con los datos anuales a partir del 2015 en periodos de 5 años, la cual es información compartida y congruente entre las distintas tablas de información. También, se determinó que no aportaba datos relevantes el “Total País” en ninguna de las tablas, ya que nos disponemos a analizar individualmente datos de cada provincia.

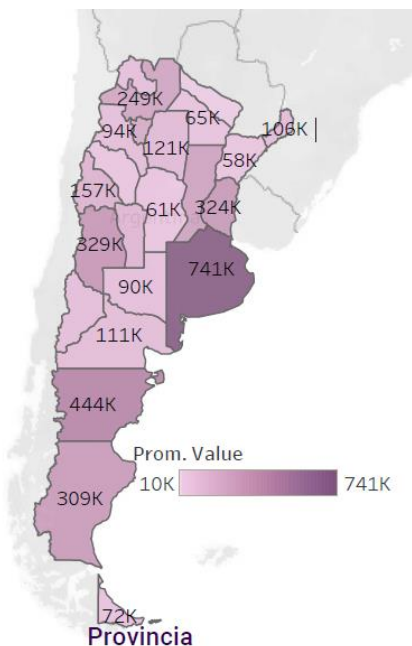
Hipótesis:

1. El Rubro de “combustibles y energía” es mayor que el de “manufactura de origen industrial” en las exportaciones del país. Esto puede deberse a que todo el territorio argentino explota este rubro, mientras que no todas las provincias realizan manufactura industrial.
2. La provincia de Buenos Aires es el mayor exportador comercial. Los de menos valor en exportaciones, son provincias sin puerto ni límites fronterizos internacionales.
3. La diferencia de esperanza de vida entre géneros ronda entre los 5 a 6 años, esta brecha se acorta en los próximos 15 años.
4. La situación geográfica y sanitaria impacta no solo en la esperanza de vida, sino también, marca la brecha habitacional y laboral del país. Por lo que las provincias de mayor superficie en Km², tienen menor densidad poblacional y se encuentran entre los mayores productores y mano de obra.

Análisis Exploratorio de Datos. Validación de Análisis Preliminar.

Reporte 1. Exportaciones País.

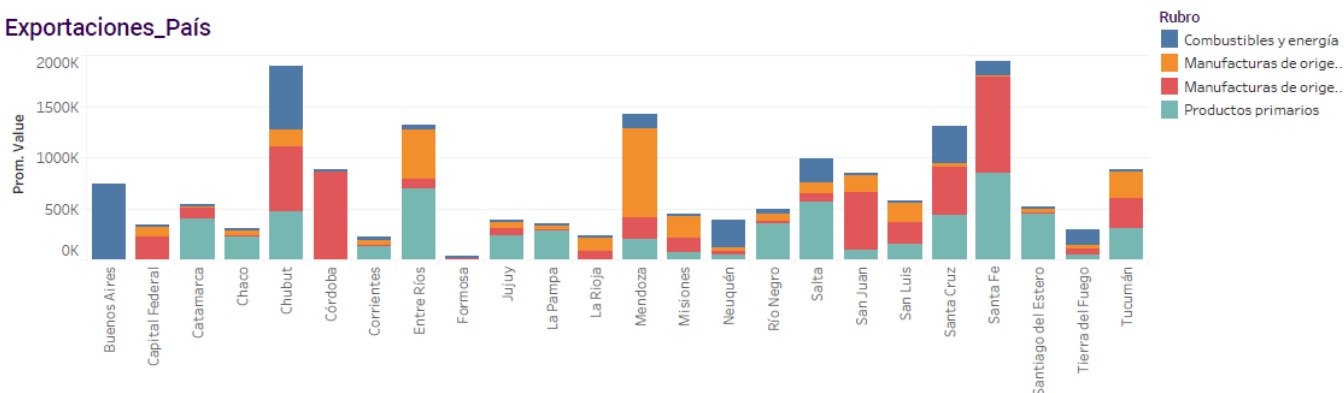
Los primeros datos evaluados fueron los que se encontraban en el archivo preliminar de exportaciones. Una primera vista de nuestro mapa político, nos permitió observar los valores de exportaciones por provincia.



Al analizarla, se puede observar que la mayor parte de las exportaciones son realizadas por las provincias de Buenos Aires – a la cual se le sumó los valores de Capital Federal – Chubut y Santa Cruz. Estas tres provincias, además de ser las que, a simple vista, son las que ocupan mayor superficie del territorio argentino, cuentan con puertos comerciales internacionales. La provincia de Buenos Aires, con 741K en exportaciones, tiene el puerto homónimo, principal de la República, por ello tiene el valor más alto.

Para el siguiente análisis, se utilizaron los parámetros de valores de exportación y provincias en los ejes principales y se filtraron por rubros, así, podremos observar si existe alguna coincidencia entre la gráfica anterior y los rubros explotados por provincia y los valores que las preceden.

Exportaciones_País



Como se puede ver, Buenos Aires, a pesar de contar con el valor más alto en las exportaciones, solo explota combustibles y energía con un valor a 741.410 en este rubro. Sin embargo, Chubut se encuentra en primer lugar con un valor promedio de 618.79 en el mismo rubro, pero, sumando los valores de manufactura de origen industrial, productos primarios y – en menor medida – manufactura de origen agropecuario.

Otra observación que podemos tomar en cuenta, es la presencia de manufactura industrial y agropecuaria en el país. Que toman gran parte en provincias portuarias y aledañas a las mismas. Y en mayor medida, los productos primarios rigen con mayor explotación en esas mismas provincias, incluyendo en los primeros puestos a la provincia de Salta. Cabe destacar que la producción de productos primarios es el común denominador en casi todo el país, salvo aquellas provincias que sólo explotan un solo rubro (Buenos Aires y Córdoba).

Es importante destacar qué provincias trabajan en mayor cantidad cada rubro, de esta manera, la empresa contratante, podrá conocer qué lugares se destacan según sus intereses. *(Las valuaciones se encuentran promediadas)*

- Combustibles y Energía: Buenos Aires con Capital Federal (741.410), Chubut (618,79), Santa Cruz (355.633)
- Manufactura de Origen Industrial: Santa Fe (964.291), Córdoba (886.243), Chubut (637,690), San Juan (562.720), Santa Cruz (476,726)
- Manufactura de Origen Agropecuario: Mendoza (875.642), Entre Ríos (483.302)
- Productos Primarios: Santa Fe (858.952), Entre Ríos (705.463), Salta (574.351)

Reporte 2. Análisis Demográfico y Social.

Los archivos aportados por el INDEC no se encuentran listos para poder realizar este análisis más allá de una observación superficial. Para poder completar este informe se decidió crear una nueva base de datos, integrando las distintas tablas de información, agregando nuevos campos calculados que nos permitan reconocer y analizar a profundidad la densidad poblacional, la diferencia entre géneros y cómo éstos aportan datos para el reconocimiento de la situación sanitaria, demográfica y social de nuestro país.

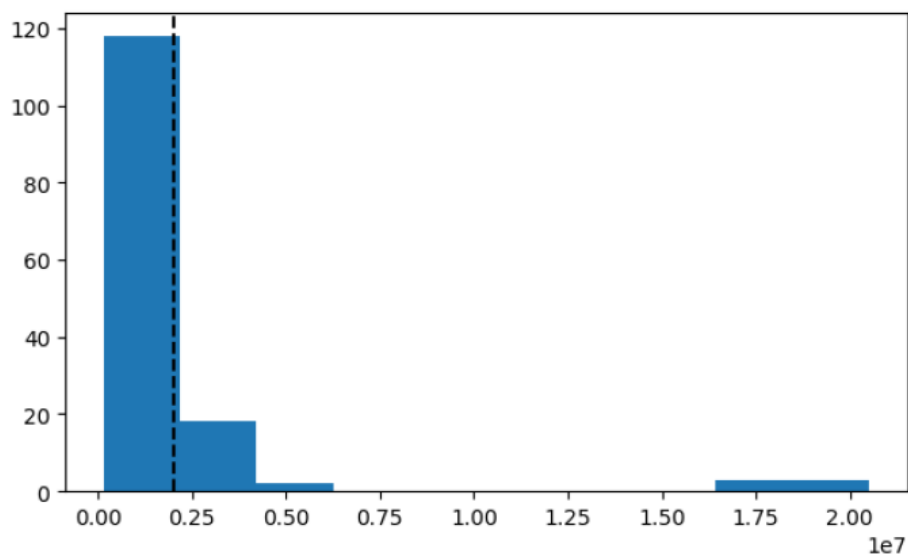
	provincia	anio	pob_total	pob_hombres	pob_mujeres	ev_mujeres	ev_hombres	diferencia_ev	sup_km2	densidad_pob
0	Capital Federal	2015	3054267	1426582	1627685	81.68	75.41	6.27	200	15271.335000
1	Capital Federal	2020	3075646	1444545	1631101	82.60	76.36	6.24	200	15378.230000
2	Capital Federal	2025	3086680	1456560	1630120	83.41	77.21	6.20	200	15433.400000
3	Capital Federal	2030	3084450	1461355	1623095	84.14	77.96	6.18	200	15422.250000
4	Capital Federal	2035	3069488	1459479	1610009	84.77	78.63	6.14	200	15347.440000
...
139	Tierra del Fuego	2020	173432	88412	85020	82.20	76.44	5.76	1002445	0.173009
140	Tierra del Fuego	2025	194926	99244	95682	83.05	77.29	5.76	1002445	0.194451
141	Tierra del Fuego	2030	216380	110000	106380	83.84	78.08	5.76	1002445	0.215852
142	Tierra del Fuego	2035	237433	120498	116935	84.54	78.66	5.88	1002445	0.236854
143	Tierra del Fuego	2040	258020	130721	127299	85.14	79.27	5.87	1002445	0.257391

144 rows x 10 columns

Con esta nueva base de datos, graficaremos distintas instancias que nos permitirán analizar distintos parámetros de interés. Por ello comenzamos con una descripción rápida de los datos cuantificables, para reconocer los valores relevantes con los que se trabajarán en distintas gráficas.

```
count    144.00
mean     655.43
std      3073.82
min       0.15
25%       4.49
50%       9.65
75%      20.62
max     15433.40
Name: densidad_pob, dtype: float64
```

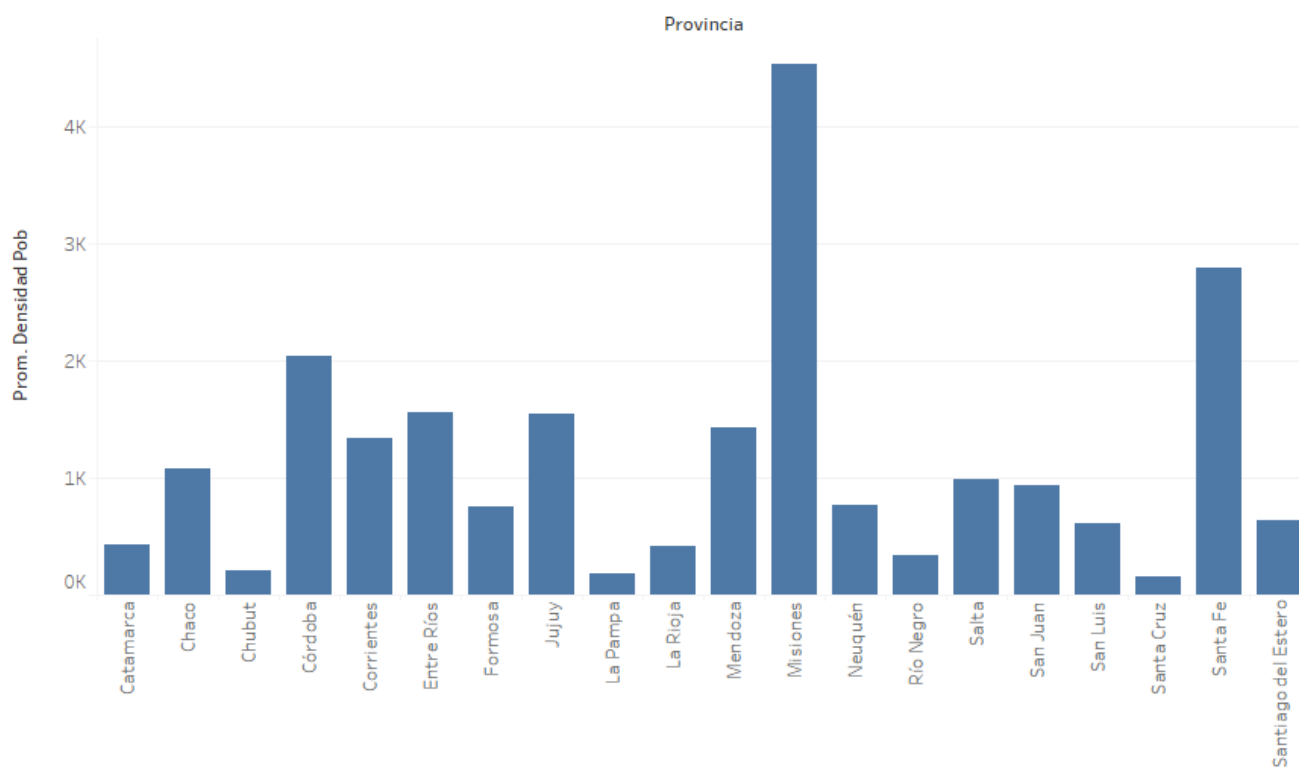
Hisograma de Población.



En primer lugar, con la máxima expresión, tenemos a Capital Federal con 15345,2 habitantes por Km², seguido de la provincia de Tucumán con 81,18 habitantes por Km² y, en tercer lugar, la Provincia de Buenos Aires con una densidad poblacional de 60,71 habitantes por Km². Esto se puede deber a que las grandes poblaciones se encomunan en las grandes ciudades universitarias, lo que puede explicar que la provincia de Tucumán se encuentre entre los primeros puestos. A su vez, también se puede deber a que las poblaciones de provincias rurales, turísticas o de bajos recursos, los habitantes más jóvenes emigren a las grandes ciudades en busca de una mejor escala salarial por la variedad de oportunidades laborales que éstas ofrecen.

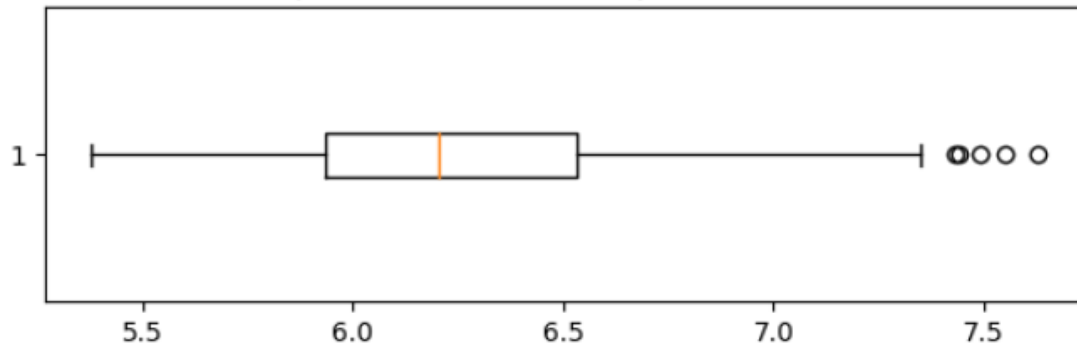
En esta misma escala de densidad, las provincias con menor densidad de población son: Tierra del Fuego con 0,20 habitantes por Km² y Santa Cruz con 1,78 habitantes por Km². En el Caso de Santa Cruz, podemos corroborar lo que hablábamos en un principio, a pesar de su superficie, muchas de éstas no son habitables por su geografía, las pequeñas “grandes ciudades” en estas provincias, se concentran muchos turistas, por lo que su ocupación habitacional autóctona es pequeña a comparación de las comunidades rurales, aisladas de este circuito.

Para que esta gráfica sea más entendible en cuanto sus valores, filtraremos estas provincias que se encuentran fuera de la correlación (datos atípicos) en ambos extremos.



En Términos generales, se observa una gran variabilidad en la densidad de población entre las provincias argentinas, lo que refleja las diferencias en la distribución demográfica a lo largo del país, a pesar de la superficie territorial.

El análisis de esta misma base de datos, nos arroja distintos resultados según la variable en estudio. Uno de ellos es la Diferencia en Esperanza de Vida en Argentina. En primer lugar, se decidió estudiar los valores a partir de un **Boxplot**.



A Primera vista, tenemos valores atípicos dentro de la Esperanza de vida, estos pueden corresponderse a la gran diferencia marcada por las provincias que superan el valor extremo de 7.00. Veamos cuáles son en una lista donde promediamos los valores de los años por provincia:

provincia	
Buenos Aires	6.49
Capital Federal	6.19
Catamarca	5.74
Chaco	6.57
Chubut	7.00
Corrientes	6.00
Córdoba	6.43
Entre Ríos	7.07
Formosa	5.90
Jujuy	6.07
La Pampa	6.23
La Rioja	6.14
Mendoza	5.91
Misiones	6.46
Neuquén	6.37
Río Negro	5.90
Salta	6.30
San Juan	5.56
San Luis	6.10
Santa Cruz	7.17
Santa Fe	7.11
Santiago del Estero	6.57
Tierra del Fuego	5.78
Tucumán	5.95

Antes de continuar con en el análisis, vale la observación que se esperaba encontrar entre los primeros puestos, provincias con temperaturas de templadas a cálidas, con mayor vegetación y fauna – lugares donde se explotan, principalmente, recursos agropecuarios – que son las que se encuentran al centro y norte del país. Ya que, al ser ricos en comida saludable, natural y una vida menos estresante a comparación de las grandes ciudades donde hay más consumo de alimentos procesados y rutinas poco saludables, como ser Capital Federal y Buenos Aires. Sorprende encontrarnos con que estas zonas, se

encuentran en la media de esperanza de vida. Por lo que tuvimos que estudiar otros factores que podrían estar influenciando.

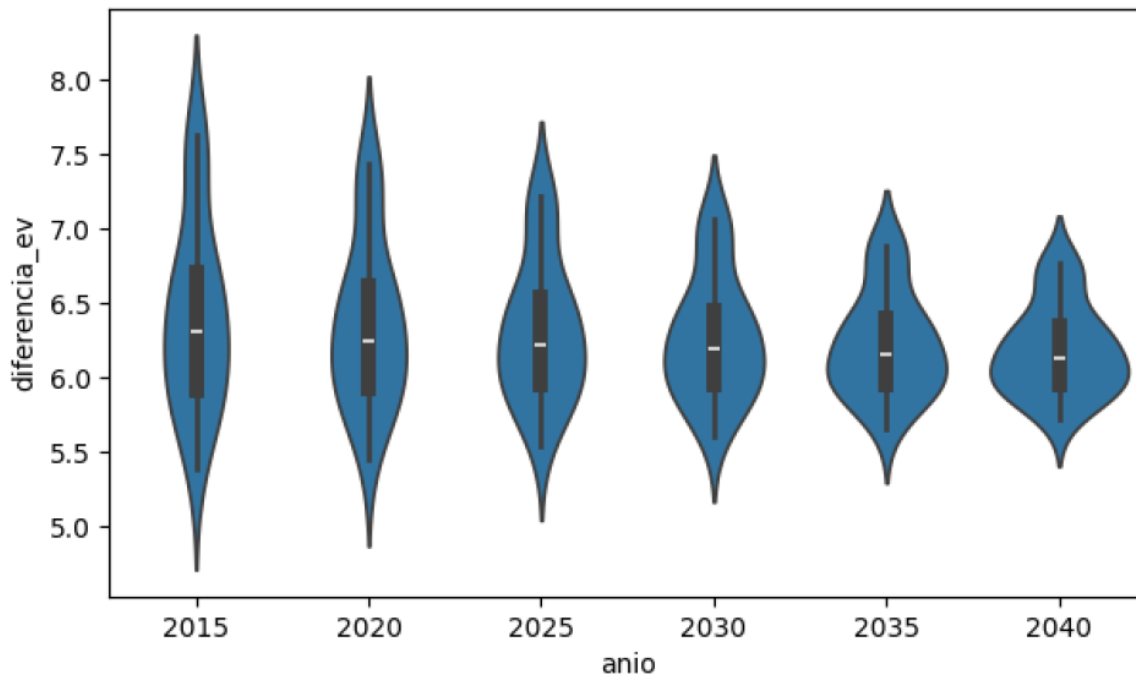
Así nos encontramos con los servicios de Salud y factores ambientales, los cuales, combinados explican por qué una región fría como Chubut y Santa Cruz, se encuentran entre los primeros lugares, estas son provincias reconocidas por tener mejor calidad de agua potable y un aire más limpio que al norte argentino, por ejemplo: Salta, Jujuy y Tucumán; que es de público conocimiento que existen regiones, localidades completas, que no cuentan con estos servicios ni una campaña fuerte de concientización del cuidado de salud y medio ambiental. Esto impacta directamente en la esperanza de vida. Santa Fe, en cambio, se puede tomar en cuenta que tiene uno de los mejores accesos a los servicios de salud (al igual que Córdoba) del país, hasta se puede mencionar que tiene reconocimientos a nivel mundial por ello. Por otro lado, las campañas en el cuidado del bienestar personal y concientización del cuidado de los espacios públicos, hizo de esta provincia, ser reconocida por una cultura mucho más saludable en canto alimentación, deportes y un gran cuidado de los espacios verdes, históricos y culturales. Siendo así, podría compararse Santa Fe con Mendoza ya que cuentan con estilos de vida y campañas de bienestar y cuidado muy parecidos, aun así, es una región propensa a cambios climáticos muy fuertes y no cuentan con un acceso a los servicios de Salud pública como Santa Fe, por lo que prefieren curarse solos, con el tiempo y con medicinas naturales. Por ello, es posible entender por qué se encuentran en extremos opuestos.

En el caso de San Juan, no sorprende que se encuentre en este puesto, ya que, aunque tiene características geográficas diferentes a Chubut o Santa Cruz, con un clima más cálido, y una geografía menos accidentada, la calidad del agua, del aire y del entorno en general no contribuyen a una mejor salud y esperanza de vida. Por último, Tierra del Fuego, a pesar de su cercanía con los puestos más altos, tiene uno de los valores más bajos. Aquí influyen muchos factores en medidas diferentes. Debido a su clima extremo y una geografía muy pobre en materia prima, hace que el estilo de vida sea aún más laboriosa que en el resto del país. Por lo que, en su historia, la migración de un gran porcentaje de la población joven en busca de un mejor estilo de vida en tierras más fértiles, hace de esta provincia tener una mayor población de gente mayor que jóvenes que puedan dedicarse a la producción y cuidado del adulto mayor. Por lo que este último grupo poblacional, tiene una tendencia a morir más joven que en el resto de Argentina.

Por último, se decidió hacer un análisis para analizar si nuestra observación inicial sobre la Diferencia en Esperanza de Vida Entre Géneros. Y si la proyección que hicimos está en lo cierto, y de ser así ¿a qué puede deberse?

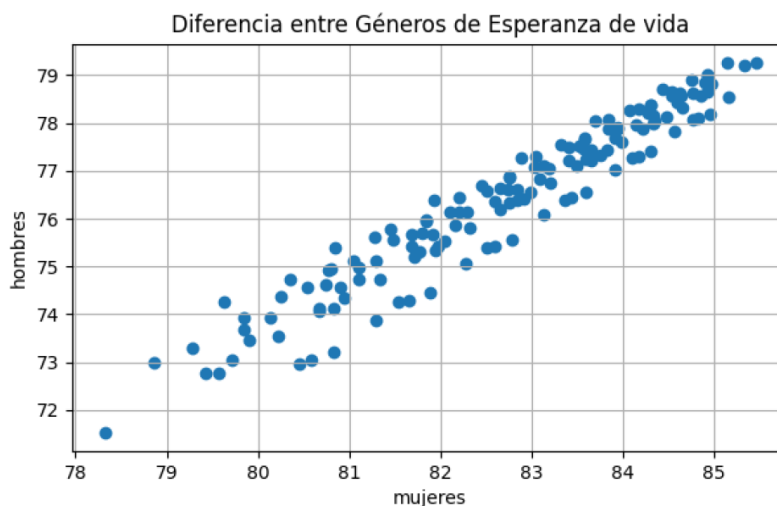
En promedio, podemos encontrarnos con que la diferencia es de 6,29. Según la provincia, varía entre 5 y 6 años de diferencia. Puede deberse a que los hombres, por una cuestión biológica, sean más propensos a problemas de salud que mujeres y tienen una tendencia muy marcada a no buscar atención médica inmediata. La mayor parte de Argentina, por cuestiones socio culturales, son más los hombres que obtienen una carga más pesada de trabajo que las mujeres: cargar bultos, ser choferes de vehículos grandes, trabajos pesados en la construcción, explotación minera, rurales, entre otros. Lo que provoca un deterioro de su bienestar más rápido que en el de la mujer.

A partir de nuestra base de datos, concluimos que, para poder estudiarlos, una mejor visualización de los datos será un **Gráfico de Violín**.



Con esta gráfica, podemos corroborar la distribución quinquenal de la Diferencia en Esperanza de Vida Entre Géneros en Argentina. Entre el 2015 y el 2020 podemos encontrarnos con datos casi similares que superan ambos extremos de la caja, principalmente la superior. Pero a partir del 2025 al 2040, según los datos observados, comienzan a agruparse dentro del primer cuartil (Q1) y el tercer cuartil (Q3). Los cuales culminan con una caja de distribución más pequeña entre los 5,8 y 6,2 - la cual es la media en 2015 - obteniendo una gran densidad dentro del segundo cuartil (Q2) que se encuentra con un valor aproximado a 6.

Esto puede llegar a darse en un futuro ya que la influencia política de los últimos años del género femenino, logró en gran medida que las mujeres puedan ejercer las mismas obligaciones que el hombre, por lo que comienza a verse una nueva corriente en el que las mujeres se preocupan más por lograr cumplir objetivos laborales para igualarse con los hombres (trabajos pesados, más horas, etc.) y dejan de lado el cuidado personal.



El **Gráfico De Dispersión** nos muestra una fuerte correlación positiva en el que la métrica de los hombres aumenta a la par que el valor de las mujeres incrementa.

Conclusiones y Recomendaciones.

A partir de estos resultados podemos concluir que según el tipo de empresa que se quiera montar en nuestro país, va a convenir una provincia o región u otra.

Si la empresa contratante busca solamente importar sus productos, el puerto Buenos Aires, de la provincia homónima, es el puerto más seguro. Si su importación es sobre productos explotados en su país de origen, para luego, ser procesados en una industria argentina y después, expandirse en el país, quizá le sea más conveniente, reconocer en las primeras gráficas la provincia más cercana al puerto que explote ese mismo producto. A modo de ejemplo, si sus productos son de origen industrial para luego ser repartidos en distintos comercios del país o su producción tiene que ver con combustibles o energía, elegir Buenos Aires; pero, si su producción es de productos primarios u origen agropecuario y estos necesitan ser procesados en el país, el puerto Rawson de Chubut sería su mejor embajada. Si bien en las métricas no se tuvo en cuenta su existencia, si sus productos tienen que ver con la industria de la investigación que se lleva a cabo en la Antártida y éstos necesitan resguardos especiales hasta que se puedan transportar a la misma, el puerto de Tierra del Fuego, es su destino.

Avanzando en los análisis, es importante la observación de los rubros explotados en nuestro territorio, la densidad poblacional y la esperanza de vida. Este conjunto de datos al unísono, van a ayudarlos a comprender que en el caso que quieran instalarse en el país para la explotación de alguno de ellos, va a ser importante seguir las tendencias, y reconocer qué provincia va a contar con personas más preparadas o con experiencia en su rubro al momento de contratar.

En el caso de los productos primarios y agropecuario, según el producto, va a ser importante realizar un estudio más detallado sobre el clima y tipo de región que se adapte mejor al mismo. Por ejemplo, si la empresa contratante es de la India o África y quieren producir especias (condimentos) tendrían que pensar en el Norte Argentino, zona cordillerana, en caso que sean especias dulces y picantes, pero si son especias minerales les convendrá instalarse en zona cordillerana de Mendoza o Santa Cruz. Por otro lado, si es cría de animales para alimentos, lo mejor es pensar en el sur del país, para vinos según los sabores y texturas que quieran conseguir, deberán decidir entre Mendoza, Salta y San Juan y por último si es cría de animales para confección de textiles, será entre Santa Cruz, Misiones, Salta y Córdoba.

Referencias.

INDEC: archivos ".csv":

- Empleo
- Hogares_viviendas_superficie
- Esperanza_de_vida
- Poblacion
- Exportaciones

<https://www.argentina.gob.ar/produccion/segemar/geologia-y-recursos-minerales/recursos-minerales>

<https://www.argentina.gob.ar/transporte/puerto-ba/el-puerto#:~:text=Puerto%20Buenos%20Aires%2C%20con%20una,carga%20de%20contenedores%20del%20pa%C3%ADs.>

<https://cdsa.aacademica.org/000-106/234.pdf>