



Universidad del Desarrollo
Facultad de Ingeniería

MIGRACIÓN INTERNA POR REGIÓN

Para años desde el 2020, 2021 y 2022

POR: JORGE WILSON

Capstone project presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. Loreto Bravo

Dr. Leo Ferres

Enero 2023

SANTIAGO

AGRADECIMIENTO

Agradecimiento a Loreto Bravo y Leo Ferres por su aporte constante como guías y fuentes de información, también en específico para Big Data a Víctor Navarro.

Agradecimientos también a el líder del proyecto que me encuentro, Gerson Bastos por brindarme el tiempo para el desarrollo

Agradecimientos como siempre a mi familia y amigos por brindarme sus opiniones y puntos de vista para reforzar y retroalimentar este trabajo.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS.....	4
4. DATOS Y METODOLOGÍA.....	5
4.1. DATOS	5
4.2. METODOLOGÍA	6
5. RESULTADOS	9
6. CONCLUSIÓN Y TRABAJO FUTURO.....	11
7. BIBLIOGRAFÍA	17

Resumen

El presente documento señala bajo la información que brinda los datos provenientes de registros de detalles extendidos (XDR) correspondientes a una mezcla de activación por el usuario y del dispositivo, el traslado de dispositivos por Chile en los años 2020, 2021 y 2022. Con ello, se realiza el análisis bajo el traslado dentro de un mes entre regiones, considerando la inmigración y emigración de una región. Para ello se utiliza el trabajo con Python y un servidor para el trabajo con grandes volúmenes de datos de la Universidad del desarrollo y también de la optimización constante del código para poder procesar y lograr el análisis. La motivación de este estudio concierne a las distintas modalidades que han ocurrido debido a la pandemia, que causa y brinda el potencial de la remotización de variados trabajos y una eventual dispersión de personas de Santiago a regiones, la cual se busca confirmar con los datos anteriormente indicados.

1. Introducción

Últimamente han existido bastantes modificaciones al estilo de forma de trabajo para todos en Chile, debido al fin de la pandemia, a la remotización de bastantes trabajos que han abierto posibilidades de posibles migraciones interna y externa en Chile. En el presente estudio se abordará las inmigraciones y emigraciones dentro de regiones en Chile para los años 2020, 2021 y 2022. Ello bajo el conjunto de datos desde los dispositivos celulares que se conectan a las antenas, brindándonos información de ubicación y temporalidad. Con esta información se crea un algoritmo que demuestra la movilización y filtra las personas que con su dispositivo realizaron un traslado de región. Para ello y por la cantidad de datos se utiliza el servidor de la Universidad del desarrollo, el cual permite el trabajo con spark y la totalidad de datos.

El trabajo con spark corresponde a una ejecución con grandes volúmenes de datos, lo que involucra un extenso tiempo de procesamiento, limitaciones y el correcto procedimiento de ejecución para llevarlo a cabo de la manera más óptima posible.

El análisis XDR de los dispositivos móviles señalan en que lugar descargan datos o existe alguna interacción con las antenas a la cual están vinculados los dispositivos, lo que genera un registro y con lo cual podemos ver la localización del dispositivo y si agregamos la variable tiempo, su movimiento o traslado dentro de Chile.

El actual estudio busca entregar la movilización en las regiones de Chile, lo cual también se puede interpretar según zonas de Chile -Norte, centro y Sur- y una posible toma de decisiones entorno a búsquedas raíces de estas movilizaciones o cambios definitivos de residencia en Chile, ya sea por acontecimientos o razones con un grado mayor de profundidad que requiere evaluación y eventualmente implementar soluciones.

2. Trabajo Relacionado

1. Con base en el trabajo realizado por Luca Pappalardo, Leo Ferres, Manuel Sacasa, Ciro Cattuto and Loreto Bravo. (Junio de 2021). *Evaluation of home detection algorithms on mobile phone data using individual-level*. Se logra el traspaso de información de definición y caso de uso de algoritmo a utilizar entre (CDR, XDR y CPR). Además, se demuestra que los CDR, el cual es el flujo más utilizado en literatura para la detección de hogar, es de baja precisión y estabilidad, en comparación con los XDR, algoritmo utilizado para el presente estudio. Finalmente también se demuestra cambio en estabilidad y precisión según horas en el día.

3. Hipótesis y Objetivos

Desde hace un tiempo, y por diferentes razones incluyendo como significativa la pandemia, se comenzó el trabajo remoto, innumerables variedades de trabajos que se pueden realizar de manera virtual, esto representa una oportunidad para la migración externa e interna en Chile, con mayor probabilidad está última en caso de que se necesite de manera esporádica el trabajo presencial también. La hipótesis en este caso es el aumento de migración interna en Chile desde Santiago al resto de Chile, incluyendo también la migración interna entre todas las regiones, la cual como hipótesis esperamos que sea menor en caso de regiones entre regiones.

El objetivo general de la presente investigación es ver el flujo de migración interna ocurrida en Chile en los años 2020, 2021 y 2022. También lograr ver el movimiento de las personas en Chile y conseguir la ponderación de las ciudades entorno a migración para la ciudad y también para las visitas a ciudades durante esos años.

Obtener la cantidad de personas que hicieron cambio de domicilio dentro de Chile comparando un movimiento de región en la cual se trasladaron para los años mencionados

4. Datos y Metodología

4.1. Datos

Se considera un total de 3.472.657.540 datos, provenientes de registros de detalles extendidos (XDR) correspondientes a una mezcla de activación por el usuario y del dispositivo, ya sea solicitando explícitamente una dirección http o descargando automáticamente contenidos de Internet, XDR corresponde a una tupla (n, t, A, k) , siendo A: antena involucrada, n: identificador de la persona, t es una marca de tiempo de cuando se crea el registro y k es la cantidad de información descargada en kilobytes (Luca Pappalardo, Leo Ferres, Manuel Sacasa, Ciro Cattuto and Loreto Bravo, 2021). En este caso se provee la base de datos con algunos cambios a la tupla, considerando:

- Device: Identificador de celular, similar a n
- Datetime: Momento en el cuándo se crea el registro, similar a t
- Lat: Latitud desde donde proviene el registro, descomposición de ubicación de A
- Lon: Longitud desde donde proviene el registro, descomposición de ubicación de A

Considerando que, para el presente estudio, la cantidad de información descargada en kilobytes sobre migración interna no nos señala alguna clase de relación (k).

Consecuentemente contamos con información XDR, y con la tupla $(n, t, A(\text{lat}), A(\text{lon}))$, para el trabajo y análisis de migración interna en Chile.

Para ello se analiza por variable: device, datetime, lat y lon.

Correspondiente a todos los datos, como primera asignación es remover datos nulos del dataset, quedando un total de 2.903.775.797 de datos. Se realiza para todas las variables, entrando en el detalle de cada una:

1. Device, ya que si no posee un ID válido, no se puede corroborar sus movimientos o traslados en la ciudad y/o país.
2. Datetime, en caso de valor nulo, no se podrá comparar sus movimientos y traslados con el resto de ellos en los años anteriormente señalados
3. Latitud y longitud, si no posee la latitud o longitud, es improbable sus movimientos respecto a ellas en la ubicación con los demás datos que se posee de ese device.

En el caso de limpieza por variable **Device**, se permite datos del tipo string, entendiendo influencias entre la mezcla entre números y letras, exceptuando el caso anteriormente nombrado de valor null.

Para el caso de **datetime**, se permite datos entre los años anteriormente señalados considerando todos los meses de esos años, y corroborar de que existan dichas fechas. Por ejemplo: que no exista meses con fechas de día mayor al que realmente tienen.

Respecto a la **variable de latitud**, confirma una **latitud limite mayor** de: -18,33 correspondiente al límite entre Arica y Tacna. Sin embargo, Chile hacia la derecha posee territorio aún más al norte con el tema de Visviri (que también se incluye en la definición del límite Visviri hasta -17.53) y para el sur se define la **latitud limite menor** de -56 en el sur de Puerto Williams.

Respecto a la **variable de longitud**, se confirma **longitud mayor** de: -67,5 y **longitud menor** de: -74 corroborado en mapas, en donde se encuentra el límite de Chile entorno al océano y respecto a la frontera con Argentina

Definiendo estos límites, no se encontró ningún dato superior o que rebase estos límites de territorio.

4.2. Metodología

En el presente estudio se logran analizar 3.472.657.540 para los años 2020, 2021 y 2022. Realizando los filtros anteriormente señalados queda un total de 2.903.775.797

Como se mencionó anteriormente y en búsqueda de validar la hipótesis, se estudia y analiza la migración interna la cual corresponde al desplazamiento de una persona o grupo de personas desde su lugar de residencia hacia otro punto para radicarse de manera semipermanente o definitiva. Se conoce como migración interna al desplazamiento de personas dentro de un mismo país. Esta migración puede ser temporal o definitiva.

De forma de análisis y para visualizar posibles traslados que señalan migración interna a diferencia de vacaciones u otros motivos, debe existir un mínimo que respalde y asegure con algún nivel de confianza el traslado de la persona, para ello entraremos en el análisis de las principales variables que intervienen en la definición de migración interna, correspondiente al tiempo y ubicación

$$\text{Migración interna} = f(\text{temporalidad}, \text{ubicación})$$

Variable temporal

Esto exige un análisis en diferentes temporalidades, si bien no hay un tiempo definido para la migración interna, existe un nivel en las vacaciones por trabajadores, según el código del trabajo “Los trabajadores chilenos con más de un año de servicio tienen derecho a 15 días hábiles de vacaciones al año, lo que en la práctica son 21 días corridos,

o tres semanas”, y trabajos correspondientes a la industria minera que exigen harta movilización de trabajadores, consideran un máximo de 14 días por persona para el traslado. Por ello, y bajo tiempo de traslado para considerarlo significativo en otras áreas como trabajo, salud, educación, construcción, infraestructura, entre otras para la ciudad. Se considera el respaldo en caso de que exista un traslado mayor a 1 mes entre ciudades.

En este caso y para el análisis se necesita como mínimo 3 datos por dispositivo para asegurar un traslado mayor a 1 mes entre ciudades obteniendo un mayor nivel de confianza de migración interna por sobre otros motivos como vacaciones o trabajos fuera de la ciudad.

Para ello y por términos computacionales, se remite al uso de diferencia entre datos ordenados por fecha. Es decir, para un device se obtendrá la resta dentro de su categoría, entregando una diferencia de un total-1.

Para nomenclatura se define lo siguiente:

Número de datos por device = n; Device = d

Max $\Delta(X_i - X_{i+1}, \dots, -X_n)$: Máximo entre las diferencias

Max $\Delta X_i =$ Diferencia mayor sub_i

El algoritmo para observar y respaldar la migración interna con una temporalidad mayor a un mes, correspondiente al tiempo hábil de vacaciones señalado anteriormente, considera los siguientes requisitos de variable tiempo para que el dispositivo pueda ser analizado y considerado:

Requisito: $n \geq 3$

Como primer paso entonces, se filtra la variable device, en caso de que no posea 3 datos o más. Junto con ello, también hay que considerar la variable tiempo, para asegurar que estos 3 datos efectivamente provienen de meses diferentes.

Posteriormente para el cálculo de la diferencia de tiempo entre los datos de un mismo device, debe existir como mínimo 2 valores que superen el mes para chequear la movilización del dispositivo.

Max₁ $\Delta(X_i - X_{i+1}, \dots, -X_n)_1$,

Max₂ $\Delta(X_i - X_{i+1}, \dots, -X_n)_2$,

Notando que la lista de

Max ΔX_2 corresponde a $\{X_i, \dots, n\}$ sin $Max \Delta X_1$

Finalmente,

$$\Delta Max_1 \geq 30 \text{ y } \Delta Max_2 \geq 30$$

Variable de ubicación

Además de la medición del tiempo, es necesario verificar el movimiento del dispositivo para cerciorar una eventual migración interna, para ello además del análisis en diferentes meses hay que chequear si es que existe un traslado de región

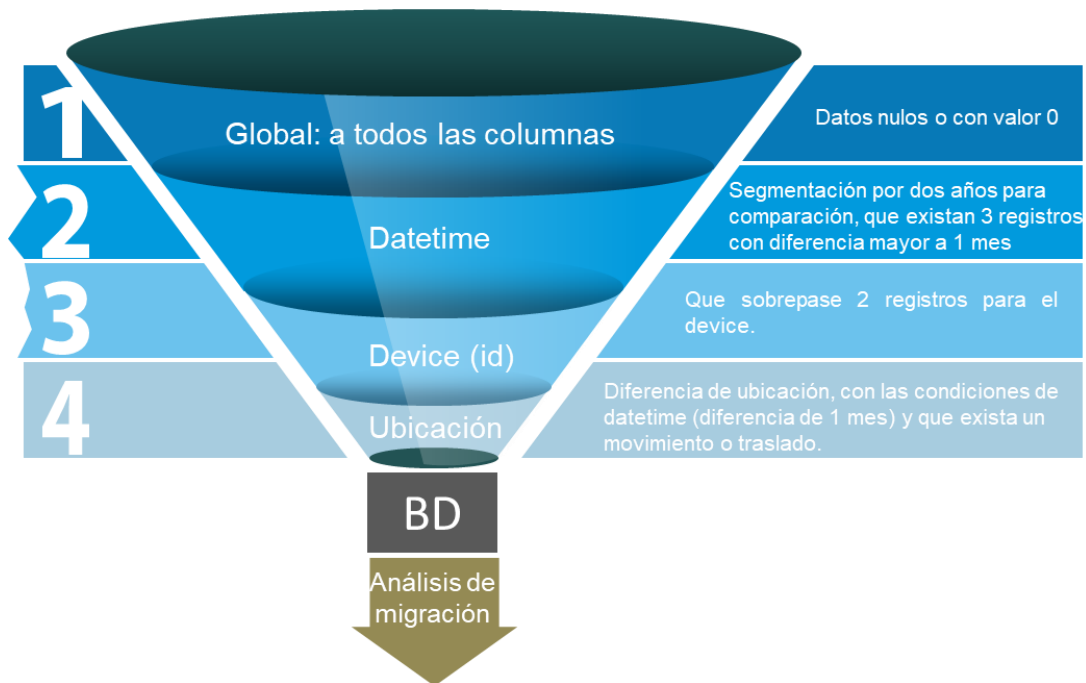
En el presente estudio y debido a la migración interna se establece un análisis de movimiento de región.

Como procedimiento entonces se establece los dispositivos que sufrieron movimientos de región analizando así una posible migración interna, y además si cumple con la variable temporal anteriormente señalada, se puede tomar una conclusión de dicho dispositivo.

Realizando el filtro de datos que sufrieron cambios de región, correspondió a un 8%, quedando en total de 246.820.942 registros que sufrieron traslados de región.

Procedimiento de metodología

De manera de resumir se entrega un esquema de los principales filtros en conjunto con el procesamiento en orden para abordarlo debido al extenso dataset que se proporciona para el análisis y la lentitud del análisis.



5. Resultados

Como se mencionó en el punto anterior, tomando la función y teniendo consideración de la temporalidad y de la ubicación, quedo un total de 246.820.942 para los años 2020, 2021 y 2022, en los cuales existió un cambio de región. Sin embargo, esto en conjunto con el análisis temporal provoca un nuevo filtro de estos datos, de los cuales solo 238.546.684 supera ambos puntos. Considerando específicamente como valor único el device, corresponde a 858.699 dispositivos que sufrieron este traslado.

Adentrándonos en el detalle de estos datos, también es relevante inspeccionar el punto de origen y destino para la migración interna por región

Migración de Región 2020 a 2021

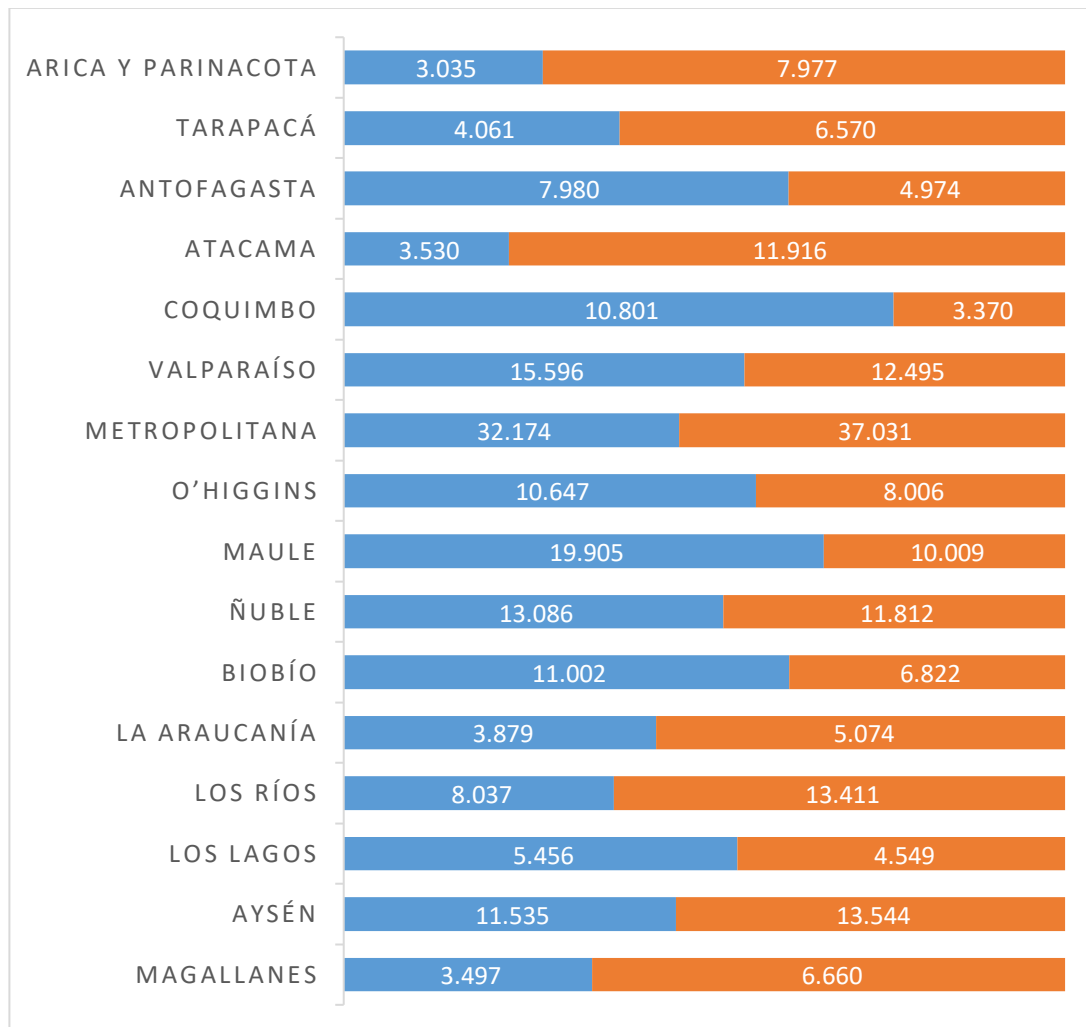


Gráfico I. de Movilización del dispositivo por regiones considerando filtros anteriormente mencionados en metodología 2020 – 2021

Migración de Región 2021 a 2022

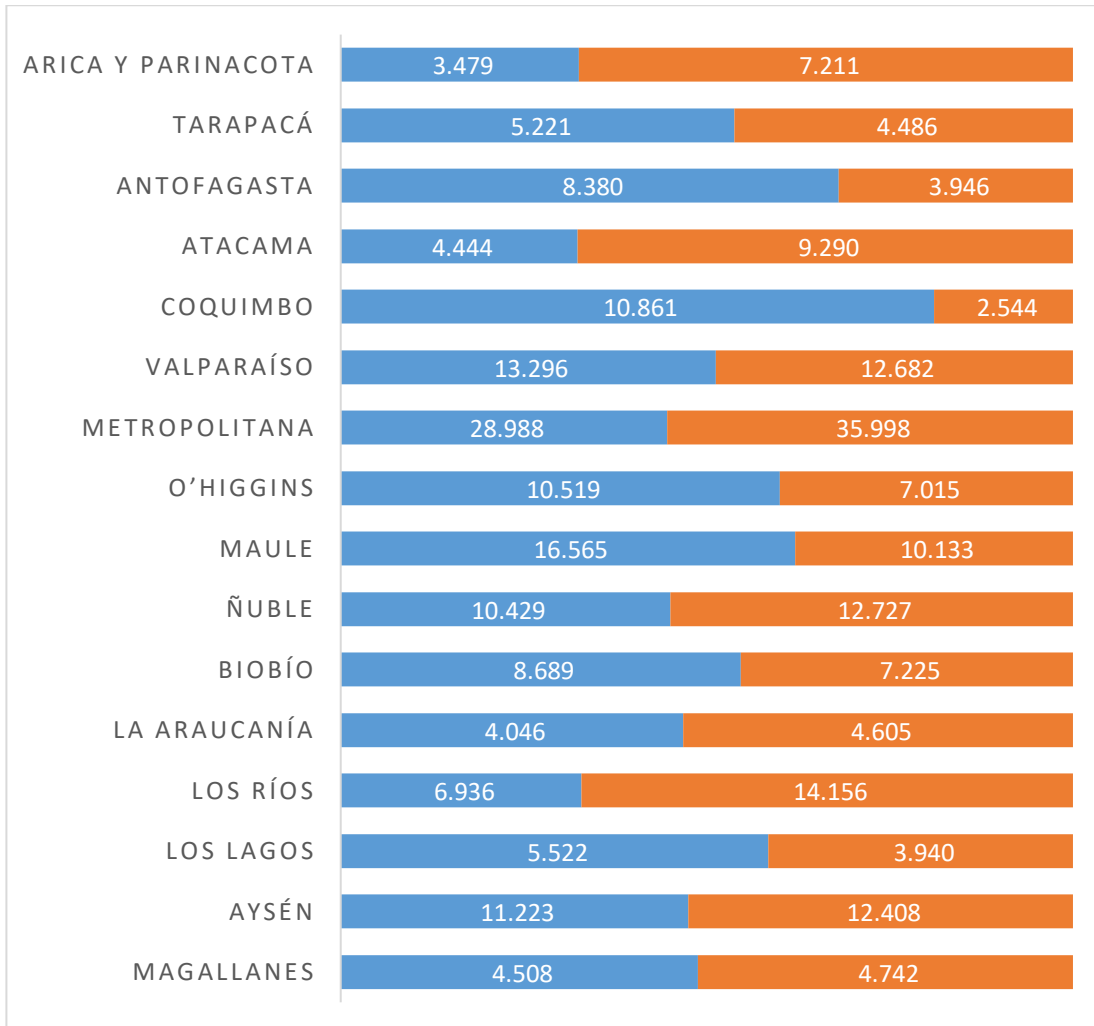


Gráfico II. de Movilización del dispositivo por regiones considerando filtros anteriormente mencionados en metodología 2021 – 2022

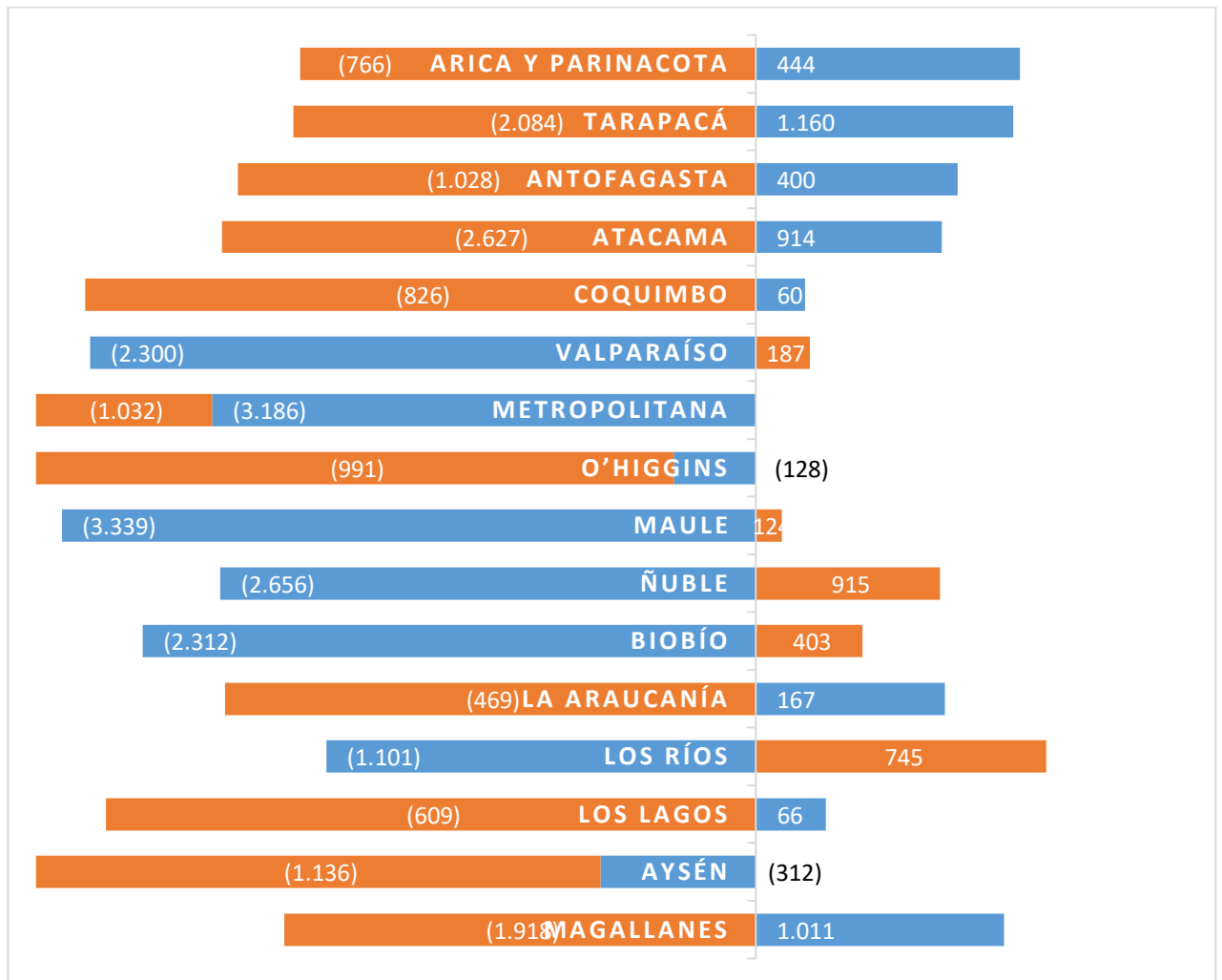


Gráfico III. Diferencia de migraciones entre 2020 y 2021 con 2021 y 2022.

6. Conclusiones

De manera general, se evidencia que el XDR nos puede brindar información respecto a la movilización de las personas en Chile, además y como continuación se puede realizar un estudio con una función que se establezca la movilidad según kilometraje, observar también la movilidad dentro de una región. Como punto a mejorar se debe considerar también la posibilidad de que una persona lleve más de un dispositivo, por ejemplo, en algunas empresas entregan celulares destinados solo al trabajo y además la persona llevaría el propio, lo cual puede ajustar la conclusión y resultado del análisis.

De manera más específica del análisis anterior mente señalado Santiago posee el mayor traslado entorno a inmigrantes y emigrantes en Chile, además la zona norte queda con menor traslados de inmigración.

Existían bastantes hipótesis respecto a la migración, por el término de la pandemia, la vuelta a la normalidad de bastantes trabajadores y la remotización de bastantes trabajos. Sin embargo, no se logra encontrar un dato significativo que demuestre estas condiciones, según otros reportes de migración Santiago suele ser la sede de mayor movilidad entorno a inmigración y emigración. Junto a ello, también se ve una inmigración menor a la zona norte y sur de Chile, indicando una percepción de las personas menor calidad en comparación con el resto de las ciudades en Chile, debido a diferentes aconteceres y noticias que han surgido en el último tiempo, además de infraestructura, migración, seguridad, entre otras. Notando además que ambas zonas poseen diferentes aconteceres y problemas. Exceptuando para la zona norte: Antofagasta y Coquimbo que poseen hartos inmigrantes a dichas regiones. Para la zona sur se exceptúa: Los Lagos y Biobío

Por el lado de Santiago notamos una emigración mayor a la inmigración, cosa que podría ser explicada por la remotización del trabajo, esto demuestra un movimiento paulatino a regiones por parte de personas que hace poco residían en Santiago.

Por otro lado, y en específico por región notamos un movimiento mayor de inmigrantes que emigrantes para: Antofagasta, Coquimbo, Valparaíso, O'Higgins, Maule, Ñuble, Biobío y los Lagos. Destacando Coquimbo, Maule y Biobío las cuales poseen hartas personas que se trasladaron para dicha zona.

En el caso contrario, para las regiones en específico que consiguieron un mayor movimiento de emigrantes que de inmigrantes corresponde a: Arica y Parinacota, Tarapacá, Atacama, Metropolitana, La Araucanía, Los Ríos, Aysén y Magallanes. Destacando Atacama, Arica y Parinacota y Magallanes.

Para el caso de la diferencia, entre los años anteriormente mencionados:

En las regiones de la zona norte (Arica y Parinacota, Tarapaca, Antofagasta, Atacama y Coquimbo) muestran una disminución a **la emigración que se venía dando en el 2020 y 2021, disminuyendo en 7.331 personas**. Además por el lado de **Inmigración señala un aumento de 2.979 personas**.

Con respecto a la **zona central (Valparaíso, Biobío, Maule, O'Higgins, Ñuble, Metropolitana)** se entrega como resultado de diferencia entre los años mencionados de **-394 emigrantes y 13.921 inmigrantes** a esta zona. **Principalmente debido a pocas emigraciones desde la Región Metropolitana y O'Higgins**, ya que de todas las demás existe mayor cantidad de emigrantes.

Para la zona sur (Magallanes, Aysén, Los Lagos, Los Ríos, La Araucanía) entrega un resultado de **-170 Emigrantes** de diferencia y **-3.387 Inmigrantes, exceptuando Los Ríos** que posee 745 inmigrantes.

Como punto a desarrollar también se establece una función que permita según el kilometraje trasladado y tiempo (en nuestro caso consideramos 1 mes) ver la movilidad de las personas, ya sea por la misma ciudad, diferentes ciudades y también diferente temporalidad. Por ejemplo, si

logramos ver un traslado de una persona mayor a 1 año, podría indicar un traslado más permanente en dicha ciudad, región o provincia..

7. Bibliografía

Luca Pappalardo, Leo Ferres, Manuel Sacasa, Ciro Cattuto and Loreto Bravo. (Junio de 2021). *Evaluation of home detection algorithms on mobile phone data using individual-level.*