

Samir Oussaid Padrón

Mejora en la distribución de los pasajeros del transporte público respecto a las matrices OD.

Improvement in the distribution of passengers in public transportation with respect to the OD matrices.

Trabajo Fin de Grado
Grado en Matemáticas
La Laguna, abril de 2023

DIRIGIDO POR
Ginés León Rodríguez
Carlos Pérez González

Ginés León Rodríguez

*Dpto. de Big Data y Data Science
TITSA
38111 Sta. María del Mar, Tenerife*

Carlos Pérez González

*Dpto. de Matemáticas, Estadística
e Investigación Operativa
Universidad de La Laguna
38200 La Laguna, Tenerife*

Agradecimientos

En primer lugar, deseo agradecer a mis compañeros de estudios, quienes estuvieron a mi lado en este camino académico. Su apoyo, colaboración y motivación fueron fundamentales para superar los desafíos y alcanzar los objetivos propuestos. Juntos compartimos momentos de estudio, discusiones enriquecedoras y experiencias inolvidables que han dejado una huella significativa en mi formación.

También quiero expresar mi profundo agradecimiento a mi madre, cuyo amor incondicional y apoyo constante han sido mi mayor fortaleza. Su aliento, comprensión y sacrificio han sido un pilar fundamental en mi vida y en la consecución de mis metas. Agradezco su amor infinito y su dedicación incansable.

Por último, agradezco a mi padre por su contribución y por brindarme las oportunidades para seguir adelante en mi educación. Aunque nuestros caminos han sido diferentes, valoro su presencia en mi vida y reconozco su papel en mi crecimiento y desarrollo personal.

A todos ustedes, compañeros y familia, les estoy profundamente agradecido por ser parte de mi camino y por su inmenso apoyo. Sus palabras de aliento, ánimo y confianza me han impulsado a esforzarme aún más y a superar los desafíos que se presentaron en esta etapa de mi vida. Gracias por creer en mí y ser una parte integral de mi éxito.

Samir Oussaid Padrón
La Laguna, 9 de julio de 2023

Resumen · Abstract

Resumen

La isla de Tenerife se encuentra dividida en 216 zonas de transporte público según la redacción del Plan Territorial Especial de Ordenación del Transporte de Tenerife, se han identificado errores de baja correlación entre la matriz origen-destino (OD) de telefonía y la matriz OD de transporte público. Estos errores plantean un desafío significativo para la empresa TITSA. En el presente trabajo de fin de grado, nos proponemos abordar este problema mediante la generación de una proyección que nos permita comprender el desplazamiento de los pasajeros de transporte público al finalizar su trayecto entre zonas cercanas, basándonos en la matriz OD de telefonía. Expondremos detalladamente el proceso y el algoritmo final empleado, así como su implementación en lenguaje Python. Además, presentaremos la solución de casos particulares de forma gráfica con el fin de ilustrar nuestros resultados.

Palabras clave: *Matriz OD – Correlación – Proyección – Algoritmo.*

Abstract

The island of Tenerife is divided into 216 public transportation zones according to the drafting of the Special Territorial Plan for Transport Organization in Tenerife. Low correlation errors have been identified between the Telephony Origin-Destination (OD) matrix and the Public Transportation OD matrix. These errors present a significant challenge for the company TITSA. In this undergraduate thesis, we aim to address this problem by generating a projection that allows us to understand the movement of public transportation passengers when completing their journeys between nearby zones, based on the the Telephony OD matrix. We will provide a detailed explanation of the process and the final algorithm used, along with its implementation in Python. Additionally, we will present graphical solutions for specific cases to illustrate our findings.

Keywords: *OD Matrix – Correlation – Projection – Algorithm.*

Contenido

Agradecimientos	III
Resumen/Abstract	V
Introducción	IX
1. Presentación del problema general y objeto de estudio	1
1.1. Introducción al problema	1
1.1.1. ¿Qué es una matriz Origen-Destino?	1
1.1.2. Estructura de la matriz origen y destino	2
1.1.3. Matriz origen destino de telefonía móvil (MOD-TEL)	3
1.1.4. Matriz origen destino de transporte público (MOD-TP)	4
1.2. Presentación del problema general.	5
1.3. Zonas de estudio particulares	7
1.3.1. Zonas de Aeropuertos	7
1.3.2. Zonas de Intercambiadores	10
1.3.3. Zonas de Hospitales	11
2. Fundamentos teóricos previos y tratamiento de datos	13
2.1. Artificial Intelligence & Machine learning	13
2.1.1. Fundamentos para construir un buen sistema de Machine Learning	14
2.2. Tratamiento de datos en TITSA	15
2.2.1. Modelos predictivos	15
2.2.2. Técnicas de visualización de datos	15
2.2.3. Entorno de programación	17
2.3. Matrices OD empleadas	17
3. Resolución del problema general	23
3.1. Objetivo de los algoritmos	23
3.2. Algoritmo previo: de zonas colindantes (destino)	24

3.3. Algoritmo final: de zonas colindantes (origen-destino)	29
3.4. Resolución con Python	33
4. Resolución de zonas particulares	37
4.1. Zonas de aeropuertos	37
4.1.1. Zona Aeropuerto del sur (1029)	37
4.1.2. Zona Aeropuerto del norte (102)	39
4.2. Zona de intercambiadores	40
4.2.1. Par OD: Intercambiador La Laguna - Intercambiador Santa Cruz	41
4.2.2. Par OD: Intercambiador Santa Cruz - Intercambiador La Laguna	42
4.3. Zonas de hospitales	43
4.3.1. Hospital Universitario de Canarias (zona 77)	43
4.4. Estudio de coeficiente de correlación de Pearson	44
Bibliografía	49
Poster	51

Introducción

En este trabajo de fin de grado, trataremos de resolver un problema que surge a la empresa de Transportes Interurbanos de Tenerife S.L.(TITSA). Que se podría resumir tal que la división de la isla de Tenerife en 216 zonas de transporte público según la redacción del Plan Territorial Especial de Ordenación del Transporte de Tenerife, ha generado una baja correlación entre la matriz OD de telefonía (MOD-TEL) y la matriz OD de transporte público (MOD-TP). Siendo el objetivo de este proyecto generar una proyección que estime el la distribución de los pasajeros de transporte público al finalizar sus trayectos entre sus zonas cercanas para mejorar la correlación entre ambas matrices. Este problema se ha considerado de alta complejidad debido al gran tamaño de las matrices OD, y este proyecto consistiría en el primer avance hacia una solución optima.

En el primer capítulo, comenzaremos por definir y explicar la obtención de estas matrices, así como analizar las causas que generan esta falta de correlación. Además, presentaremos los casos particulares que utilizaremos a lo largo del proyecto para ilustrar nuestras soluciones.

A continuación, en el segundo capítulo exploraremos las aplicaciones y utilidades que ofrecen las matrices OD, así como el valioso uso que les brinda la empresa TITSA. Destacaremos especialmente su relevancia en métodos de predicción, gracias al empleo de técnicas de inteligencia artificial y aprendizaje automático. Además, presentaremos las matrices origen-destino utilizadas en este proyecto y realizaremos visualizaciones con Power BI que permitirán comparar y contrastar ambas matrices de manera gráfica.

Posteriormente, en el tercer capítulo abordaremos la resolución del problema general, presentando un algoritmo preliminar y el algoritmo final que nos llevará a obtener una solución satisfactoria. Asimismo, detallaremos la implementación del algoritmo final en el lenguaje de programación Python y explicaremos los resultados obtenidos.

En el cuarto capítulo presentaremos de manera gráfica la solución de los casos particulares seleccionados en el primer capítulo, lo que nos permitirá com-

prender mejor el desplazamiento de los pasajeros de transporte público entre zonas cercanas al finalizar su trayecto.

Para finalizar, realizaremos unas conclusiones que nos describirán la eficacia de los resultados obtenidos y una acción de mejora.

Presentación del problema general y objeto de estudio

En este capítulo, daremos inicio a nuestro estudio al abordar diferentes aspectos relacionados con las matrices origen-destino. En primer lugar, nos adentraremos en la definición y estructura de una matriz origen-destino, comprensión fundamental para entender el alcance de nuestro proyecto.

A continuación, nos centraremos en analizar detalladamente la matriz OD de telefonía y la matriz OD de transporte público. Exploraremos su obtención y cómo reflejan el flujo de movimientos entre diferentes zonas. Estas matrices son piezas clave en nuestro estudio, ya que permiten capturar y representar la interacción entre los usuarios de los servicios de telefonía y transporte público.

En el siguiente apartado, nos adentraremos en el problema general que nos ha impulsado a realizar esta investigación. Nos enfocaremos en las causas que generan la baja correlación entre ambas matrices. Este desafío plantea interrogantes que requerirán de un análisis exhaustivo para comprender las razones subyacentes a esta discrepancia.

Para concluir este capítulo, presentaremos casos particulares cuidadosamente seleccionados. Estos casos nos servirán como ejemplos concretos para ilustrar y explicar el proceso llevado a cabo en nuestro proyecto. A través de estos casos, buscaremos brindar una visión más clara y detallada de cómo se aborda la resolución del problema planteado.

1.1. Introducción al problema

1.1.1. ¿Qué es una matriz Origen-Destino?

Una matriz origen-destino es una herramienta utilizada en la planificación del transporte y la movilidad urbana. Esta matriz muestra las relaciones entre los lugares de origen y destino de los viajes realizados por la población en una determinada área geográfica. Permite entender las necesidades de transporte de la población, planificar rutas eficientes, identificar áreas de congestión, analizar

patrones de tráfico y calcular la demanda de transporte en diferentes momentos del día.

En el contexto de la movilidad, esta matriz es fundamental para comprender los patrones de desplazamiento de las personas dentro de una ciudad o región. Proporciona información sobre cómo se utilizan los diferentes modos de transporte para llegar a destinos como el trabajo, la escuela, el comercio o los centros de salud. Esto es especialmente relevante para el transporte público, ya que permite a las autoridades y planificadores tomar decisiones informadas sobre la asignación de recursos y la planificación de servicios, con el objetivo de mejorar la eficiencia de las rutas, aumentar la capacidad de los servicios y mejorar la accesibilidad a lugares importantes.

Además, la matriz origen-destino también es útil para evaluar el impacto de cambios en la infraestructura de transporte, como la construcción de nuevas carreteras o líneas de transporte público. Al conocer los patrones de movilidad existentes, es posible identificar oportunidades para mejorar la movilidad y reducir los tiempos de viaje. Esto puede tener efectos positivos tanto en la calidad de vida de la población como en la economía local. En resumen, la matriz origen-destino es una herramienta clave en la evaluación y el diseño de sistemas de transporte y movilidad urbana.

Gracias al análisis de las dos Matrices Origen Destino, y al desarrollo de algoritmos de optimización combinatoria y de programación de entera, aplicado a los problemas VPR (Vehicle Routing Problem), la empresa dispone de una serie de algoritmos que continuamente están testeando la red, dando alertas allí donde se producen ineficiencias, y proponiendo escenarios que se pueden aplicar para solucionar las ineficiencias detectadas.

1.1.2. Estructura de la matriz origen y destino

En la tarea de proporcionar un servicio de transporte público eficaz, resulta crucial analizar la movilidad de la región en cuestión. Para este fin, los datos de geolocalización recopilados por los servicios de telefonía móvil se convierten en una valiosa fuente de información, ya que la gran mayoría de las personas poseen un teléfono móvil y lo llevan consigo durante sus desplazamientos. Por tanto, al comparar los datos proporcionados por estos servicios con los datos de movilidad del transporte público almacenados por el proveedor, es posible revelar el nivel de cobertura ofrecido y facilitar la mejora de la red.

En esta sección, se describirán dos matrices origen-destino (MOD) que representan la movilidad de las personas entre las 216 zonas de transporte público en las que esta dividida Tenerife (por la redacción del Plan Territorial Especial de Ordenación del Transporte de Tenerife). La primera matriz se construye a

partir de los datos de los servicios de telefonía móvil, mientras que la segunda se basa en los datos proporcionados por el proveedor de transporte público. Estas matrices ofrecen una visión detallada de los patrones de desplazamiento y permiten comprender cómo se relacionan los lugares de origen y destino en términos de movilidad.

1.1.3. Matriz origen destino de telefonía móvil (MOD-TEL)

La MOD-TEL ha sido obtenida en colaboración con el Cabildo de Tenerife y generada por el Área de Movilidad y Carreteras de dicho organismo. Esta es una tabla de datos que registra los desplazamientos de personas desde una zona de telefonía a otra. Cada entrada en la tabla representa un desplazamiento. Como podemos observar en la tabla de la figura 1.1 se incluye información como la fecha (**Fechas_Utilizadas**), el código origen (**origen**), el código de destino (**destino**), y la cantidad de personas que realizan dicho desplazamiento (**viajes**). Se considera un desplazamiento cuando una persona supera un tiempo predefinido en la zona de destino.

Fechas_Utilizadas	origen	destino	viajes
22, 23 y 24 de marzo 2022	1	59	2,23
22, 23 y 24 de marzo 2022	3	59	13,24
22, 23 y 24 de marzo 2022	5	59	5,17
22, 23 y 24 de marzo 2022	6	59	2,59
22, 23 y 24 de marzo 2022	7	59	5,76
22, 23 y 24 de marzo 2022	39	59	38,19
22, 23 y 24 de marzo 2022	40	59	60,83
22, 23 y 24 de marzo 2022	42	59	106,68
22, 23 y 24 de marzo 2022	43	59	105,53
22, 23 y 24 de marzo 2022	44	59	132,83
22, 23 y 24 de marzo 2022	45	59	36,89
22, 23 y 24 de marzo 2022	46	59	304,91
22, 23 y 24 de marzo 2022	47	59	17,78
22, 23 y 24 de marzo 2022	48	59	46,95
22, 23 y 24 de marzo 2022	49	59	329,87
22, 23 y 24 de marzo 2022	50	59	56,31
22, 23 y 24 de marzo 2022	51	59	117,59
22, 23 y 24 de marzo 2022	52	59	644,90

Figura 1.1. Ejemplo de matriz OD de Telefonía visualizada en Power BI.

Observamos en la figura 1.1 la tabla de una matriz OD de telefonía que recoge datos del 22,23 y 24 de marzo de 2022, que consta de 33.437 filas, que reflejan las combinaciones entre las 216 zonas de transporte que recogieron datos no nulos durante esas fechas.

Los datos de movilidad de telefonía móvil se obtienen a partir de las antenas ubicadas en las diferentes zonas. La localización de los dispositivos móviles se logra mediante la multilateralización (triangulación, combinación) de las señales

de radio entre varias torres de radio de la red y el teléfono. El número de dispositivos localizados en una zona se considera un indicador del número de personas presentes en dicha zona. Sin embargo, en algunas ocasiones, los dispositivos de algunas personas pueden estar conectados a antenas ubicadas en zonas adyacentes de mayor altura o potencia, lo que puede ocasionar que en los datos de telefonía aparezcan personas ubicadas en una zona en la que no se encuentran realmente. Se asume que en los datos de la MOD-TEL se ha realizado un proceso de calibración previo para solucionar este problema y mejorar la precisión de los resultados.

1.1.4. Matriz origen destino de transporte público (MOD-TP)

La Matriz de Origen y Destino de Transporte Público (MOD-TP) es una estructura de datos que almacena de manera organizada toda la información relevante sobre los viajes realizados por una persona en transporte público, desde la parada inicial hasta la final, teniendo en cuenta los transbordos efectuados. Los datos registrados en la matriz incluyen el identificador de la tarjeta utilizada, la fecha del viaje, la parada de origen, la hora de inicio, la parada de destino y la hora de llegada. Sin embargo, para efectos de comparación con otras matrices de movilidad, solo se consideran los desplazamientos que van desde la parada de origen hasta la de destino.

En la MOD-TP, se considera que un viaje de una persona desde una parada de origen hasta una parada de destino constituye un desplazamiento una vez que ha transcurrido el tiempo límite de trasbordo establecido por el proveedor de transporte (en este caso, se ha fijado en 90 minutos a partir de la subida en la parada de origen).

PARADA_ORIGEN	PARADA_DESTINO	ZONA TELEFONIA_ORIGEN	ZONA TELEFONIA_DESTINO	Pasajeros
9181	1519	57	59	27
2625	2384	96	59	107
30400	1520	66	59	34
30300	33500	66	59	43
30200	33500	66	59	71
30400	33500	66	59	82
31400	33500	76	59	97
33100	1520	76	59	30
33100	1515	76	59	26
33100	33500	76	59	386
33200	33500	76	59	74
30500	33500	67	59	84
32100	1519	98	59	30
31300	33500	58	59	98
31000	33500	75	59	76
31200	33500	58	59	78
1763	1569	55	59	45
33300	33500	58	59	75
32100	33500	98	59	120
31600	33500	79	59	66
31900	33500	97	59	82

Figura 1.2. Ejemplo de matriz OD de Transporte Público visualizada en Power BI.

Observamos en la figura 1.2 la tabla de una matriz OD de transporte público de TITSA que recoge datos del 22, 23 y 24 de marzo de 2022, que consta de 70.939 filas, y obtenemos información de los códigos de parada origen (**PARADA_ORIGEN**), parada destino (**PARADA_DESTINO**), la zona de telefonía de origen (**ZONA TELEFONIA_ORIGEN**) y destino (**ZONA TELEFONIA_DESTINO**), y la cantidad de pasajeros en cada movimiento (**Pasajeros**).

En la actualidad, la MOD-TP se utiliza para calcular la distancia recorrida por cada pasajero, lo que ha permitido obtener información sobre el par de paradas de origen y destino para el 94 % de los pasajeros. Gracias a ello, se ha reducido significativamente el porcentaje de paradas de salida desconocidas, pasando del 80 % al 6 %. Para lograrlo, se han empleado diversos métodos estadísticos y técnicas de aprendizaje automático (Machine Learning), que han posibilitado predecir la parada de bajada para la mayoría de los pasajeros en función de sus patrones de uso. Basándose en distintos algoritmos, incluyendo aquellos basados en Inteligencia Artificial, se ha logrado determinar con precisión en qué parada se subió y bajó cada usuario en cada etapa de su viaje.

1.2. Presentación del problema general.

La división de la isla de Tenerife en **216 zonas de transporte público** plantea un desafío en cuanto a la correlación de datos entre la Matriz de Origen y Destino de Telefonía Móvil (MOD-TEL) y la Matriz de Origen y Destino de Transporte Público (MOD-TP), dado que la correlación entre los movimientos de personas en ambas matrices es muy baja. La zonificación es consistente con la utilizada en la redacción del Plan Territorial Especial de Ordenación del Transporte de Tenerife que se observa en la figura 1.3.



Figura 1.3. Representación gráfica de la división en zonas de Transporte de Tenerife.

Este fenómeno puede ser atribuido a diversos factores de influencia, entre los cuales se destacan los siguientes aspectos:

- En primer lugar, es importante señalar que las zonas de telefonía móvil presentan una variabilidad en términos de tamaño y forma. Por ejemplo, la región de Santa Cruz se compone de numerosas zonas pequeñas y asimétricas como se observa en el mapa de la figura 1.4. Mientras que otras zonas como Arico por ejemplo, se caracterizan por zonas más extensas y homogéneas. Asimismo, es relevante destacar que la zonificación adoptada solo guarda parcial coherencia con la división territorial de la isla. Por ejemplo, la región de Fasnia se divide en dos zonas, una ubicada por encima y otra por debajo de la autopista TF-1. Esta subdivisión se explica por la influencia adicional de la densidad poblacional, dado que la parte más habitada de Fasnia, al igual que en muchos otros municipios, se encuentra en la zona costera. Estas diferencias en la zonificación son un factor determinante que dificulta la comparación de los datos de movilidad provenientes de la telefonía móvil y del transporte público.

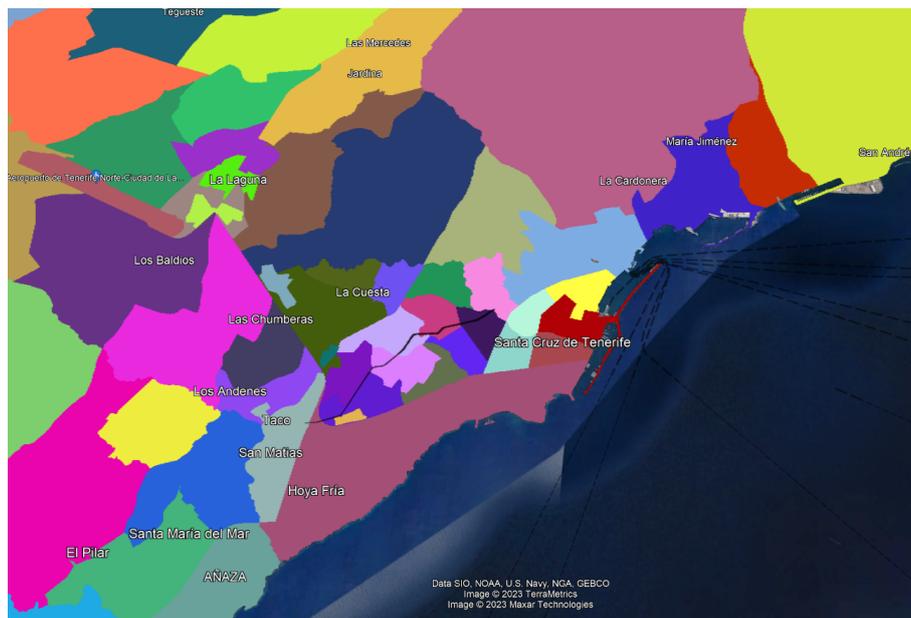


Figura 1.4. Representación gráfica de las zonas de transporte en la región de Santa Cruz.

- Otro aspecto relevante a considerar son las paradas de transporte público. En algunos casos, se observan desplazamientos hacia zonas que registran un número significativo de pasajeros en la MOD-TEL y, sin embargo, no se registran movimientos en la MOD-TP. Esto se debe a la inexistencia de paradas de transporte público en dichas zonas. Por otro lado, ocurre lo contrario, donde

se registran una cantidad considerable de pasajeros en la MOD-TP y ningún desplazamiento en la MOD-TEL. Este escenario se debe a situaciones como la presencia de intercambiadores, que representan una zona importante en el sistema de transporte público y, en consecuencia, generan un gran número de pasajeros que finalizan sus desplazamientos. Sin embargo, estos viajes no son registrados en la MOD-TEL, ya que los pasajeros no pasan el tiempo suficiente dentro de dicha zona, sino que se desvían hacia zonas cercanas, ya sean colindantes o accesibles a pie.

Estas diferencias antes mencionadas constituyen uno de los principales factores que dificultan la comparación de los datos de movilidad provenientes de la telefonía móvil y del transporte público.

1.3. Zonas de estudio particulares

Vamos a enumerar los casos particulares han sido seleccionados con el fin de ilustrar de manera clara y concisa el funcionamiento de nuestro enfoque:

- Zona de transporte: 1029 (**Aeropuerto Sur**)
- Zona de transporte: 102 (**Aeropuerto Norte**)
- Zona de transporte: 57 (**Intercambiador SC**)
- Zona de transporte: 96 (**Intercambiador LL**)
- Zona de transporte: 77 (**Hospital Universitario de Canarias**)

A lo largo de nuestra investigación, analizaremos detalladamente cada uno de estos casos, considerando sus características y propiedades específicas. Mediante su estudio exhaustivo, podremos comprender mejor las particularidades que presentan y cómo nuestra solución se adapta a cada uno de ellos.

1.3.1. Zonas de Aeropuertos

El estudio de los dos aeropuertos de la isla como casos de análisis se justifica por varias razones de interés matemático.

- En primer lugar, los aeropuertos cuentan con una zona geográfica específica asignada exclusivamente para su funcionamiento, como podemos observar en las figuras [1.5](#) y [1.6](#).



Figura 1.5. Representación gráfica de zona del aeropuerto del sur (zona 1029).



Figura 1.6. Representación gráfica de zona del aeropuerto del norte (zona 102).

- Además, como se observa en la tabla de la figura 1.7 existe una curiosa particularidad en la variación de la cantidad de personas entrantes y salientes en relación con la hora del día. Esta variabilidad temporal es de especial relevancia y representa un desafío interesante para los análisis matemáticos, ya que permite investigar patrones y tendencias que puedan estar presentes en los flujos de pasajeros.

Hora	Salientes Tel	Salientes TP	Entrantes Tel	Entrantes TP
0	67,65	0	46,42	0
1	29,10	1	37,65	0
2	35,74	0	35,15	0
3	26,91	0	48,15	0
4	27,56	0	69,39	0
5	41,50	0	142,47	9
6	55,37	2	255,38	23
7	87,16	2	238,28	40
8	135,45	20	244,58	66
9	114,28	16	225,72	45
10	185,07	40	143,77	43
11	215,79	39	189,22	32
12	201,92	47	198,39	50
13	192,67	41	187,90	41
14	262,50	40	239,65	69
15	193,13	67	225,71	66
16	224,60	58	215,24	70
17	276,95	70	126,52	38
18	292,15	44	99,33	34
19	249,75	84	97,89	24
20	187,91	74	103,18	31
21	119,49	27	82,67	5
22	75,97	10	63,44	11
23	85,76	2	62,40	0
Total	3384,45	646	3378,61	697

Figura 1.7. Tabla de zona del aeropuerto del sur con datos obtenidos por la MOD-TEL Y MOD-TP.

En el caso de **la zona 102 (aeropuerto norte)** carece de paradas de transporte público. Esta particularidad tiene como consecuencia que los datos relacionados con la llegada y salida de personas a través del transporte público en dicha zona sean inexistentes. Esto hace que sea un buen ejemplo de estudio puesto que existe un gran porcentaje de pasajeros de transporte público que acceden a sus zonas cercanas, donde si existen paradas de transporte público, para terminar accediendo a al aeropuerto.

Mediante el estudio detallado de estos dos aeropuertos y considerando sus características geográficas y su variabilidad temporal, se espera contribuir al avance de los conocimientos matemáticos relacionados con el análisis de flujos de

Otro caso a estudiar es la zona que contiene a **el intercambiador de Santa Cruz (zona 57)**, que se como se ilustra en la figura 1.9 se observa que esta en la esquina inferior derecha (representado con un círculo blanco), siendo toda la zona 57 la representada en color morado. A partir de esta observación, se ha llegado a la conclusión de que los datos relacionados con dicha zona son de gran valor debido a que el intercambiador comparte espacio con puntos de alto interés, como zonas industriales y centros comerciales, entre otros.

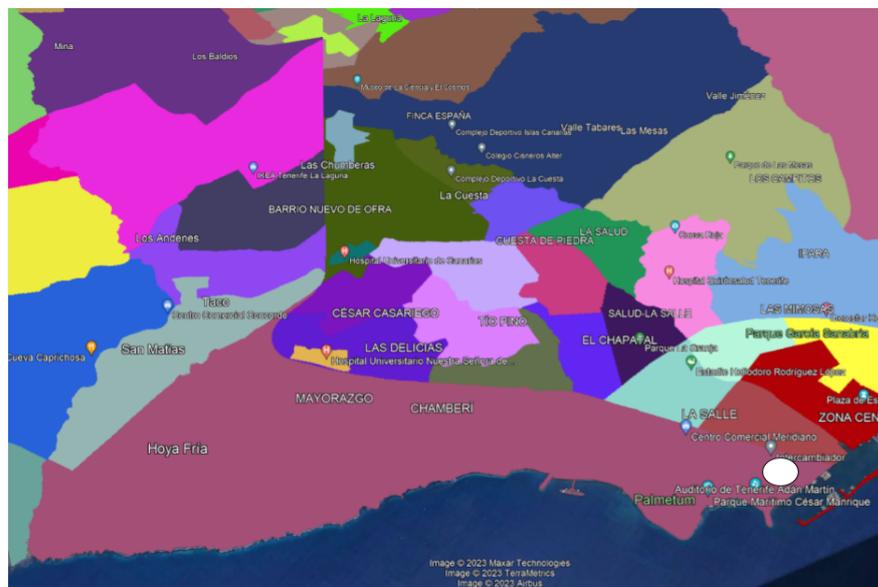


Figura 1.9. Representación gráfica de la región del Intercambiador de Santa Cruz (zona 57).

En resumen, se puede afirmar que ambos intercambiadores son casos de estudio relevantes, donde debido al funcionamiento de nuestras matrices OD se registra un mayor número de pasajeros utilizando el transporte público en comparación con la movilidad telefónica en la mayoría de los desplazamientos.

1.3.3. Zonas de Hospitales

El **Hospital Universitario de Canarias (código zona 77)**, observamos su zona representada en la figura 1.10, que lo escogemos debido a que no existen paradas de transporte público en toda su zona. Por lo que es uno de los motivos de realizar este estudio ya que los datos de la MOD-TP son nulos en este caso, y se busca generar una proyección que teniendo en cuenta los pasajeros que entran

Fundamentos teóricos previos y tratamiento de datos

En primer lugar, introduciremos los conceptos fundamentales de la inteligencia artificial (Artificial Intelligence, AI) y el aprendizaje automático (Machine Learning, ML). Exploraremos los pasos necesarios para construir un sólido sistema de ML, comprendiendo su importancia y aplicaciones en diversos campos.

A continuación, nos adentraremos en el tratamiento de datos de TITSA, y examinaremos cómo se utilizan las matrices OD en su departamento de Big Data & Data Science. Observaremos el valor que estas matrices aportan y cómo pueden ser aprovechadas para la creación de modelos predictivos. Además, presentaremos Power BI, la herramienta utilizada por TITSA para realizar visualizaciones, y describiremos el entorno de programación empleado en este proyecto: Python, junto con la plataforma Visual Studio Code.

Por último, nos sumergiremos en las matrices OD empleadas específicamente en este proyecto. A través de visualizaciones elaboradas con Power BI, compararemos los datos entre estas matrices, ofreciendo una perspectiva gráfica y comprensible de su relación.

2.1. Artificial Intelligence & Machine learning

La Inteligencia Artificial (IA) y el Aprendizaje Automático (ML) son dos campos interrelacionados en el ámbito de la informática. La IA se centra en la creación de sistemas capaces de realizar tareas que requieren inteligencia humana, como reconocimiento de voz, toma de decisiones y aprendizaje. Por su parte, el ML se enfoca en desarrollar algoritmos y modelos que permiten a las computadoras aprender y mejorar automáticamente a través de la experiencia y los datos.

La empresa de Transportes Interurbanos de Tenerife comenzó a implementar el departamento de Big Data y Data Science a finales de 2017. Esto demuestra que, aunque los sistemas siempre recopilarán anomalías en los datos, existen técnicas, procesos y herramientas que permiten limpiar y depurar los datos, brindándoles una fiabilidad sin precedentes.

En este sentido, la IA y el ML desempeñan un papel fundamental al permitir analizar y resolver problemas complejos mediante el uso de algoritmos y modelos matemáticos. Estas tecnologías tienen aplicaciones en diversas áreas, como la predicción de patrones, la optimización de procesos y la clasificación de datos.

2.1.1. Fundamentos para construir un buen sistema de Machine Learning

Para crear buenos sistemas de Machine Learning se requiere tener en cuenta varios aspectos fundamentales:

1. Datos de calidad: Es crucial contar con conjuntos de datos completos, limpios y representativos del problema que se desea resolver. Los datos deben ser relevantes y estar correctamente etiquetados para entrenar y evaluar los modelos de Machine Learning.
2. Selección y preprocesamiento de características: Es importante identificar las características relevantes que aportan información para el problema en cuestión. Además, se deben realizar técnicas de preprocesamiento de datos, como la normalización, la eliminación de ruido o la reducción de dimensionalidad, para mejorar la calidad de las características utilizadas.
3. Elección del algoritmo adecuado: Existen diversos algoritmos de Machine Learning, cada uno con sus propias fortalezas y debilidades. Se debe seleccionar el algoritmo más apropiado para abordar el problema y los datos disponibles, considerando factores como la naturaleza del problema (clasificación, regresión, agrupamiento, etc.) y el tamaño del conjunto de datos.
4. Entrenamiento y validación rigurosos: El proceso de entrenamiento del modelo debe ser cuidadoso y riguroso. Es esencial dividir los datos en conjuntos de entrenamiento, validación y prueba para evaluar y ajustar el rendimiento del modelo. Esto permite detectar posibles problemas de sobreajuste o subajuste, y mejorar la generalización del modelo a nuevos datos.
5. Optimización de hiperparámetros: Muchos algoritmos de Machine Learning tienen parámetros ajustables, llamados hiperparámetros, que influyen en el rendimiento del modelo. La elección óptima de estos hiperparámetros es crucial y se logra mediante técnicas de búsqueda sistemática, como la validación cruzada o la búsqueda en cuadrícula.
6. Evaluación y ajuste continuo: Es importante evaluar regularmente el rendimiento del modelo en datos no vistos y realizar ajustes si es necesario. El aprendizaje continuo y la retroalimentación constante permiten mejorar y adaptar los modelos de Machine Learning a medida que se obtienen nuevos datos y se enfrentan nuevos desafíos.
7. Interpretación de resultados: Comprender y explicar los resultados obtenidos por los modelos de Machine Learning es esencial. Esto implica analizar y

visualizar las predicciones del modelo, así como interpretar las características más relevantes para comprender su impacto en las decisiones tomadas por el modelo.

2.2. Tratamiento de datos en TITSA

Las Matrices OD desempeñan un papel esencial en el análisis de redes de transporte, proporcionando datos fundamentales para abordar los problemas de eficiencia en la red. Los procesos asociados a la recopilación de información de la Matriz Origen Destino generan diversos conjuntos de datos que permiten comprender el comportamiento de los usuarios al utilizar los servicios de la empresa. Estos conjuntos de datos son utilizados como base para desarrollar modelos predictivos que a continuación se explican.

2.2.1. Modelos predictivos

TITSA utiliza modelos predictivos, que son herramientas que permiten realizar predicciones basadas en el análisis de datos. Estos modelos utilizan información histórica para predecir eventos futuros o comportamientos de interés. Para cada predicción, se tienen en cuenta diversos datos, como la edad y el género del usuario, el día de la semana, la fecha, la hora, las paradas de subida y bajada, el lugar de residencia, el tipo de tarjeta, las condiciones meteorológicas, el tipo de día (laboral, festivo, etc.), entre otros.

Al construir estos procesos junto con técnicas matemáticas, TITSA puede obtener datos como los que vamos a enumerar a continuación:

- Predicción de las ventas totales o la posible pérdida de clientes
- Predicción de la ocupación de vehículos en tiempo real
- Predicción de la ocupación del bus en tiempo real por viaje
- Predicción del número de pasajeros anuales
- Algoritmos de detección de averías tempranas
- Algoritmos de optimización, teniendo en cuenta factores como el impacto ambiental, el horro económico y el beneficio social

Algunas técnicas matemáticas comúnmente utilizadas para complementar estos procesos incluyen los modelos ARIMA y SARIMA, que son modelos dinámicos de series temporales, el modelo lineal, que utiliza una variable predictora y una respuesta, y el modelo no lineal, que busca obtener los valores de los parámetros asociados a la mejor "curva de ajuste".

2.2.2. Técnicas de visualización de datos

TITSA ha incorporado la herramienta **Power BI** en su departamento de Big Data & Data Science, específicamente utilizando Power BI Desktop. Esta aplicación de escritorio descargable y gratuita forma parte de Power BI y

proporciona un entorno más completo y potente para crear informes y paneles interactivos con funcionalidades avanzadas de modelado y análisis de datos. Power BI Desktop se utiliza para diseñar, desarrollar y publicar informes en Power BI.

Power BI Desktop brinda a TITSA una plataforma integral que les permite conectar, transformar y visualizar sus datos de manera interactiva. Pueden conectar y combinar diversas fuentes de datos, como bases de datos, archivos Excel, servicios en la nube, entre otros, para obtener una visión completa de la información. La herramienta ofrece una amplia gama de opciones de visualización, como gráficos, tablas y mapas interactivos, que facilitan la exploración y comprensión de los datos.

Una de las ventajas destacadas de Power BI Desktop es su capacidad para crear informes y paneles personalizados. TITSA puede diseñar informes con visualizaciones interactivas y dinámicas, combinando diferentes elementos visuales para contar una historia con los datos. Estos informes pueden ser compartidos con otros usuarios de Power BI, lo que facilita la colaboración eficiente y la toma de decisiones basada en datos.

Además, Power BI Desktop ofrece funcionalidades avanzadas, como la creación de medidas personalizadas, análisis de datos utilizando el lenguaje DAX y la capacidad de realizar consultas y transformaciones complejas en el Editor de consultas. Estas características permiten a TITSA llevar a cabo análisis en profundidad y obtener información valiosa de sus datos.

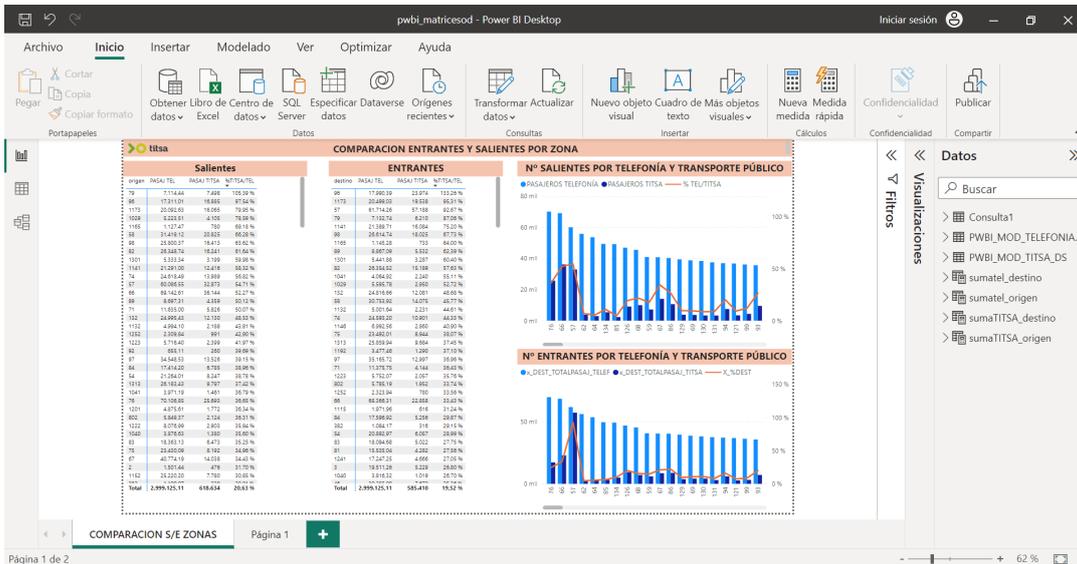


Figura 2.1. Panel de visualizaciones en Power BI desktop sobre pasajeros entrantes y salientes a cada zona por MOD-TEL y MOD-TP.

2.2.3. Entorno de programación

El estudio se ha implementado utilizando el lenguaje de programación **Python**, debido a su facilidad de lectura y escritura gracias a su similitud con el lenguaje humano. Python es un lenguaje multiplataforma de código abierto y gratuito, lo que brinda la posibilidad de desarrollar software sin restricciones.

Para el desarrollo de nuestro código se ha usado **Visual Code Studio**, que es un entorno de desarrollo de código abierto y multiplataforma creado por Microsoft. Es ampliamente utilizado por su interfaz intuitiva, soporte para múltiples lenguajes de programación y amplia personalización. Es una herramienta poderosa para el desarrollo de software.

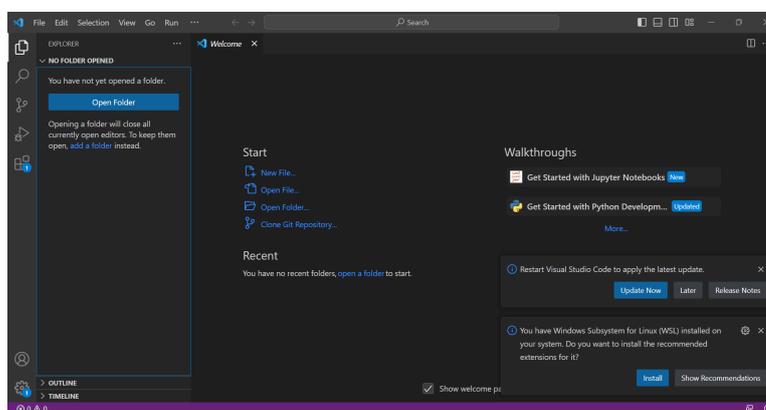


Figura 2.2. Entorno Visual Studio Code.

2.3. Matrices OD empleadas

Nos centraremos en la descripción de las matrices OD de telefonía y transporte público que utilizaremos para abordar y resolver el problema planteado. Nuestro objetivo es analizar detalladamente los datos obtenidos de ambas fuentes, con el fin de demostrar las posibles causas subyacentes del problema que nos proponemos solucionar.

Vamos a trabajar con la tabla que observamos en la figura 2.3, que contiene en cada fila los datos del código de zona origen (**origen**) y destino (**destino**), la hora del periodo que va desde las 00:00 hasta las 23:00 (**periodo**), además de los pasajeros en cada movimiento por telefonía (**pasajeros telefonía**) y transporte público (**pasajeros transporte público**).

origen	destino	periodo	pasajeros telefonía	pasajeros transporte público
96	57	14	27,267	115
96	57	18	4,243	104
96	96	14	11,027	99
96	57	13	16,2	91
57	57	14	304,461	73
57	57	19	301,558	73
57	96	20	6,938	73
96	57	12	13,777	70
96	57	15	24,624	68
126	132	14	37,344	66
101	57	9	28,688	65
57	96	19	17,669	64
96	57	16	17,409	63
96	57	19	5,778	63
129	132	14	90,678	62
126	126	14	278,695	59
57	96	21	7,942	58
126	132	13	18,413	58
57	96	13	26,652	57
1	1	14	68,672	56
54	57	7	63,518	56

Figura 2.3. Parte de la tabla de la matriz OD empleada.

Cantidad de movimientos en cada periodo

La tabla que se refleja en la figura 2.3 contiene **270.795 filas** que representan los movimientos no nulos de telefonía entre las 216 zonas de transporte de Tenerife en un día fraccionado por periodos de una hora. De estas filas solo existen movimientos de transporte público con pasajeros no nulos en **34.146 filas**. Observamos la diferencia gráfica de datos en la figura 2.4. .

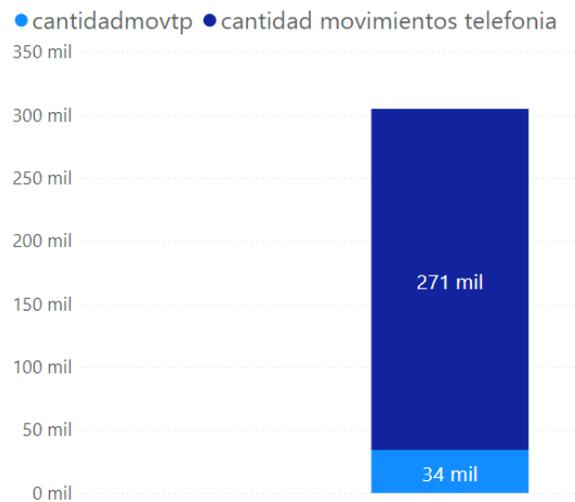


Figura 2.4. Representación gráfica de cantidad de movimientos de telefonía frente a la cantidad de movimientos de transporte.

Ahora vamos a observar unos histogramas que representan la cantidad de movimientos no nulos de telefonía por periodo (figura 2.5) y también de la cantidad de movimientos no nulos de telefonía por hora (figura 2.6).

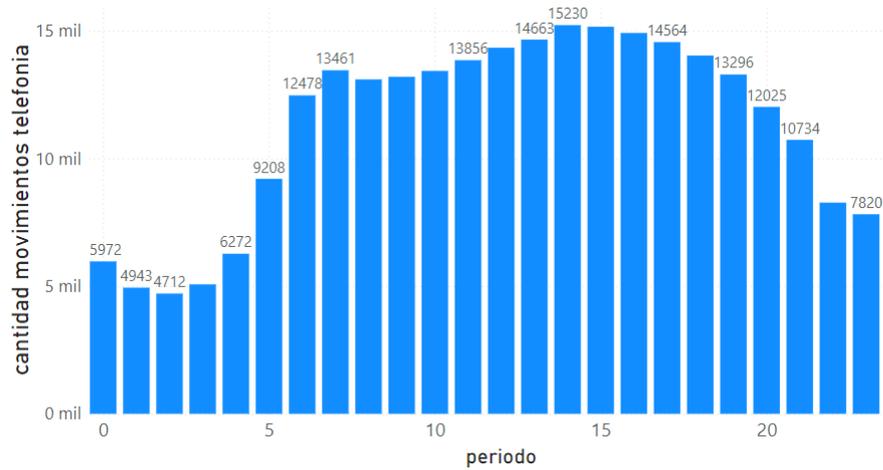


Figura 2.5. Gráfico que mide el número de movimientos por telefonía por periodo obtenido por la MOD-TEL.

En la figura 2.5 obtenemos que la hora con mayor cantidad de movimientos no nulos de telefonía es a las 14:00 horas con **15230** y la que tiene menor cantidad de movimientos es a las 02:00 con **4712**.

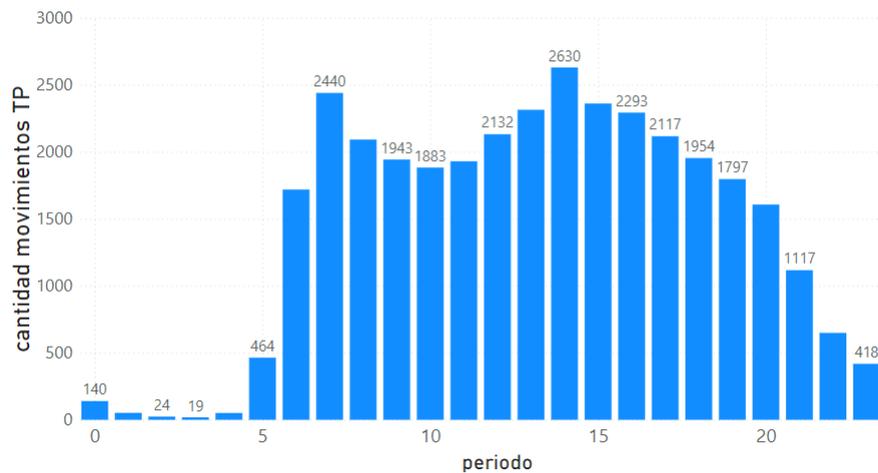


Figura 2.6. Gráfico que mide el número de movimientos por transporte público por periodo obtenido por la MOD-TP.

En la figura 2.6 obtenemos que la hora con mayor cantidad de movimientos no nulos de transporte es a las 14:00 horas con **2630** y la que tiene menor cantidad de movimientos es a las 03:00 con **19**.

Cantidad de pasajeros por periodo

La tabla de la figura 2.3 contiene un total de **2.714.655,34** pasajeros por movilidad de telefonía y **127.720** pasajeros por movilidad de transporte en todo el día. Observamos la diferencia gráfica de datos en la figura 2.7. .

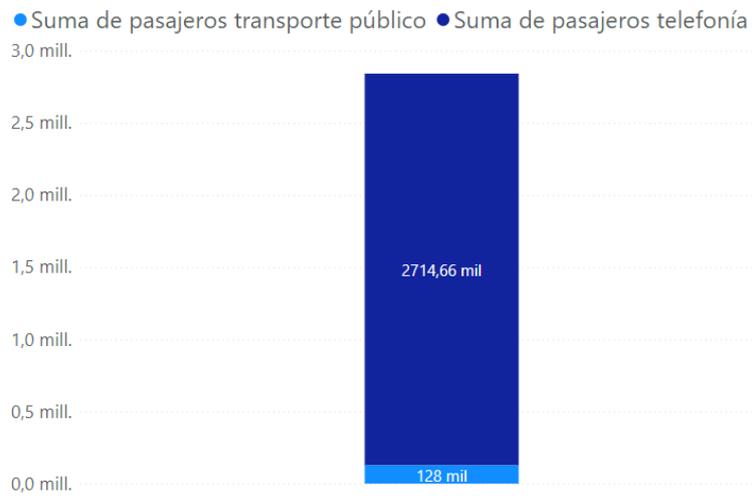


Figura 2.7. Representación gráfica del número de pasajeros de telefonía frente al número de pasajeros de transporte.

Ahora vamos a observar unos histogramas que representan la cantidad de pasajeros de telefonía por periodo(figura 2.8) y también de la cantidad de pasajeros de transporte público por periodo (figura 2.9). .

En la figura 2.8 obtenemos que la hora con mayor cantidad de pasajeros de telefonía es a las 14:00 horas con **166.990,992** y la que tiene menor cantidad de pasajeros es a las 03:00 con **28.643**.

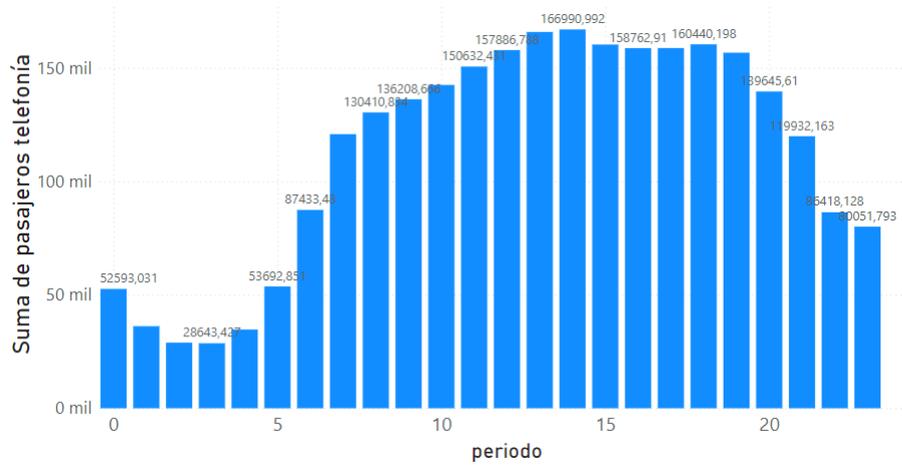


Figura 2.8. Representación gráfica del número de pasajeros de telefonía por periodo.

En la figura 2.9 obtenemos que la hora con mayor cantidad de pasajeros de transporte es a las 14:00 con **12.726** y la que tiene menor cantidad de pasajeros es a las 03:00 con **22**.

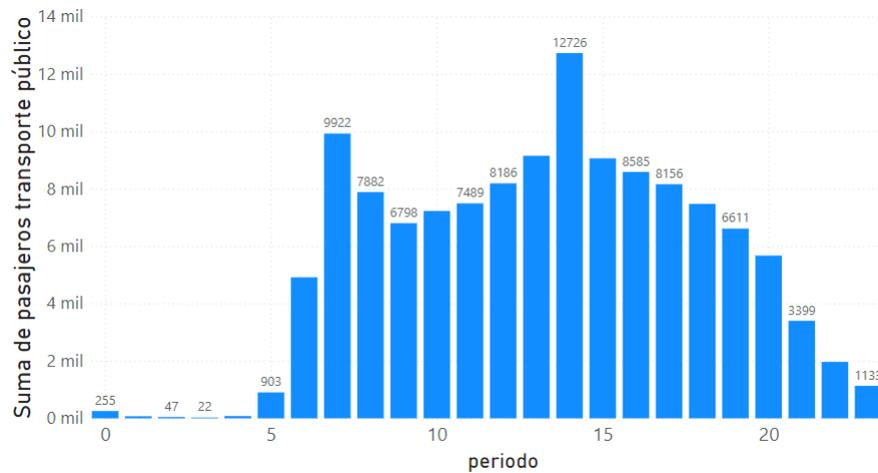


Figura 2.9. Representación gráfica del número de pasajeros de transporte público por periodo.

Resolución del problema general

En este capítulo, exploraremos el desarrollo de nuestra proyección. En primer lugar, definiremos el objetivo principal de nuestra proyección, detallando claramente lo que buscamos lograr con este enfoque. Luego, presentaremos una metodología para agrupar los resultados obtenidos, estableciendo un marco claro para el análisis de datos.

Posteriormente, abordaremos un algoritmo preliminar basado en zonas colindantes (solo considerando el origen) y lo respaldaremos con la resolución de un ejemplo gráfico. Este ejemplo será de gran ayuda para comprender cómo funciona el algoritmo y cómo se refleja en la distribución de pasajeros de transporte público.

Luego, presentaremos el algoritmo final basado en zonas colindantes (considerando tanto el origen como el destino). Acompañaremos este algoritmo con la resolución de otro ejemplo gráfico, que nos permitirá visualizar cómo se produce el cambio en la distribución de pasajeros de transporte público.

A continuación, nos adentraremos en la implementación del algoritmo final en Python y explicaremos detalladamente los resultados obtenidos. Analizaremos las ventajas y limitaciones de esta implementación y evaluaremos la eficacia de la proyección generada.

3.1. Objetivo de los algoritmos

Nuestro objetivo es generar un algoritmo que distribuya los pasajeros de transporte público (TP) entre las distintas zonas de TP, los pasajeros no se crean ni se destruyen, se proyectan a zonas colindantes en base al porcentaje de movilidad de telefonía que aporte cada par OD. Tras realizar la proyección, una zona quedará con alguno de los siguientes tres estados:

- **Zonas de atracción:** Se trata de una zona que en un principio contaba con una serie de movimientos de pasajeros (en la matriz OD de transporte

público) y dicho número se ha visto incrementado debido a que según la matriz de telefonía existe una mayor movilidad entre dicha zona y otras. Por lo general son zonas que no poseen paradas de transporte público y son el destino final de los pasajeros.

- **Zona de cesión:** Se trata de una zona que en un principio contaba con una serie de movimientos de pasajeros (en la matriz OD de transporte público) y dicho número se ha visto disminuido debido a que según la matriz de telefonía existe una menor movilidad entre dicha zona y otras. Por lo general son zonas que poseen paradas de transporte público, pero no es el destino final de los pasajeros.
- **Zona neutral:** Son zonas donde la movilidad no se ha visto afectada.

Ahora vamos a enumerar un algoritmo previo, como antecedente a la obtención de nuestro algoritmo final.

3.2. Algoritmo previo: de zonas colindantes (destino)

En el presente estudio, nos centraremos en el análisis de zonas colindantes. Las zonas colindantes se definen como áreas adyacentes o contiguas que comparten una frontera común.

La consideración de zonas colindantes puede proporcionar una medida aproximada de la influencia espacial, pero no necesariamente refleja con total precisión la conectividad o accesibilidad real desde el punto de vista del usuario. Es decir, la facilidad con la que los usuarios pueden llegar desde las paradas de transporte público a las zonas cercanas puede variar y no ser completamente precisa al utilizar únicamente la proximidad geográfica.

En este proceso vamos a tener en cuenta solo el destino de los movimientos obtenidos por las matrices OD, y hallaremos las proyecciones sobre las zonas colindantes a nuestro destino central.

Algorithm 1 Algoritmo zonas colindantes (destino)

Require: Zona de destino central. Número de pasajeros de MOD-TP y MOD-TEL

Ensure: $proyeccionI \equiv$ proyección de los pasajeros de MOD-TP ajustada a los porcentajes de MOD-TEL para cada destino perteneciente a la macrozona de destino (D).

1: Obtener las zonas colindantes a nuestra zona de destino central, generando así la macrozona de destino (D).

2: Obtener la movilidad de pasajeros entrantes a cada uno de los destinos de nuestra macrozona de destino por telefonía (tel_i) y por transporte público (tp_i).

3: Calculamos el total de personas entrantes a cada uno de los destinos (D) por movilidad de telefonía.

$$TotalTel = \sum_{i \in D} tel_i$$

4: Se calcula el porcentaje de aplicación de la movilidad de telefonía sobre cada uno de los destinos.

$$porcentajeTel_i = \frac{tel_i}{TotalTel} \cdot 100, \forall i \in D$$

5: Calculamos el total de personas entrantes a cada uno de los destinos (D) por movilidad de transporte público.

$$TotalTp = \sum_{i \in D} tp_i$$

6: Para finalizar obtenemos la proyección de la siguiente forma:

$$proyeccionI = porcentajeTel_i \cdot TotalTp, \forall i \in D$$

Ejemplo resuelto:

Vamos a realizar una representación de como funcionaria nuestro algoritmo resolviendo un ejemplo con datos ficticios.

Nuestro algoritmo recibe un destino central, como observamos en la figura 3.1.

Zona de destino



Figura 3.1. Representación gráfica de zona de destino central.

En primer lugar, halla sus zonas colindantes (zona 2,3,4 y 5) obteniendo así la macrozona de destino (siendo la unión del destino central y sus zonas colindantes, y denotada por D), y el número de personas que entran a cada una de ellas por la movilidad de telefonía(tel_i), como observamos en la figura 3.2.

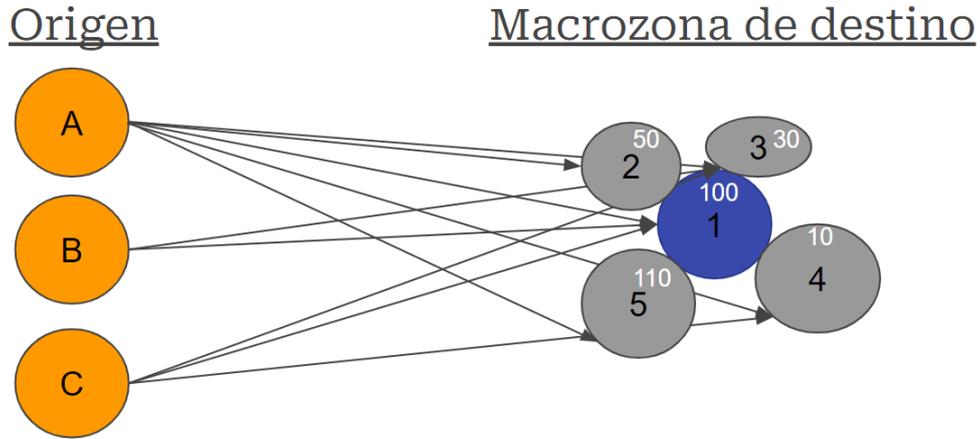


Figura 3.2. Representación gráfica de macrozona de destino (D).

En segundo lugar, vamos a hallar el total de personas que entran en la macrozona de destino (D) por movilidad de telefonía, como observamos en la figura 3.3.

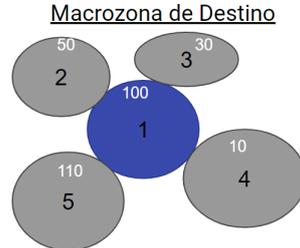


Figura 3.3. Representación gráfica de macrozona de destino (D) con datos obtenidos por MOD-TEL.

En cada zona de la figura 3.3 el número que acompaña al identificador de la zona es la cantidad de personas entrantes por movilidad telefónica (tel_i), por lo que hallamos el total:

$$TotalTel = \sum_{i \in D} tel_i = 100 + 50 + 30 + 10 + 110 = 300$$

En tercer lugar, se calcula el porcentaje de aplicación de la movilidad de telefonía sobre cada uno de los destinos pertenecientes a la macrozona D .

$$porcentajeTel_i = \frac{tel_i}{TotalTel} \cdot 100, \forall i \in D$$

$$\text{Zona 1: } porcentajeTel_1 = \frac{tel_1}{TotalTel} \cdot 100 = 100/300 = 0,33$$

$$\text{Zona 2: } porcentajeTel_2 = \frac{tel_2}{TotalTel} \cdot 100 = 50/300 = 0,17$$

$$\text{Zona 3: } porcentajeTel_3 = \frac{tel_3}{TotalTel} \cdot 100 = 30/300 = 0,10$$

$$\text{Zona 4: } porcentajeTel_4 = \frac{tel_4}{TotalTel} \cdot 100 = 10/300 = 0,03$$

$$\text{Zona 5: } porcentajeTel_5 = \frac{tel_5}{TotalTel} \cdot 100 = 110/300 = 0,37$$

En cuarto lugar, sabiendo la cantidad de personas que entran a cada una de nuestras zonas pertenecientes a la macrozona de destino (D) por transporte público, vamos a hallar el total de estas.

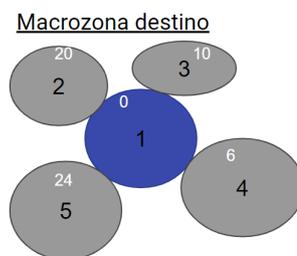


Figura 3.4. Representación gráfica de macrozona de destino (D) con datos obtenidos por MOD-TP.

En cada zona de la figura 3.4 el número que acompaña al identificador de la zona es la cantidad de pasajeros entrantes por transporte público (tp_i), por lo que hallamos el total:

$$TotalTp = \sum_{i \in D} tp_i = 0 + 20 + 10 + 6 + 24 = 60$$

En quinto y último lugar, multiplicamos el total de pasajeros de transporte público en nuestra macrozona por el porcentaje hallado previamente para cada zona y obtenemos la proyección de pasajeros de transporte público para cada zona perteneciente a D .

$$proyeccionI_i = porcentajeTel_i \cdot TotalTp, \forall i \in D$$

$$\text{Zona 1: } proyeccionI_1 = porcentajeTel_1 \cdot TotalTp = 60 \cdot 0,33 = 19,80$$

$$\text{Zona 2: } proyeccionI_2 = porcentajeTel_2 \cdot TotalTp = 60 \cdot 0,17 = 10,20$$

$$\text{Zona 3: } proyeccionI_3 = porcentajeTel_3 \cdot TotalTp = 60 \cdot 0,10 = 6,00$$

$$\text{Zona 4: } \textit{proyeccionI}_4 = \textit{porcentajeTel}_4 \cdot \textit{TotalTp} = 60 \cdot 0,03 = 1,80$$

$$\text{Zona 5: } \textit{proyeccionI}_5 = \textit{porcentajeTel}_5 \cdot \textit{TotalTp} = 60 \cdot 0,37 = 22,20$$

Resumiendo todo el proceso en la tabla 3.5:

Zonas Destino	tel _i	tp _i	porcentajeTel _i	ProyeccionI _i
1	100	0	0,33	19,80
2	50	20	0,17	10,20
3	30	10	0,10	6,00
4	10	6	0,03	1,80
5	110	24	0,37	22,20
Totales	300	60	1	60

Figura 3.5. Tabla que resume el proceso del algoritmo 1.

3.3. Algoritmo final: de zonas colindantes (origen-destino)

En este nuevo enfoque, consideramos el origen de los movimientos capturados por las matrices OD para generar proyecciones específicas entre nuestro origen central y sus zonas colindantes hacia el destino central y sus zonas colindantes. Al incorporar esta información adicional, logramos obtener una estimación más precisa y detallada de la movilidad de transporte. Esta metodología nos brinda una visión más completa de la facilidad con la que los usuarios pueden llegar desde las paradas de transporte público a zonas cercanas.

Algorithm 2 Algoritmo zonas colindantes (origen-destino)

Require: Zona de origen y destino central. Número de pasajeros de MOD-TP y MOD-TEL

Ensure: $proyeccionI_{ij} \equiv$ proyección de los pasajeros de MOD-TP ajustada a los porcentajes de MOD-TEL para cada movimiento entre las macrozonas de origen (O) y de destino (D).

1: Obtener las zonas colindantes a nuestra zona de origen y destino central, así generando las macrozonas de origen (O) y de destino (D).

2: Para cada par de la macrozona origen a la macrozona destino ($O_i - D_j$), se obtiene la movilidad existente tanto en transporte (tp_{ij}) como en telefonía (tel_{ij}):

3: Se calcula la movilidad total de telefonía de la macrozona origen a la macrozona destino:

$$TotalTel = \sum_{i \in O} \sum_{j \in D} tel_{ij}$$

4: Se calcula el porcentaje de aplicación de la movilidad de telefonía para cada uno de los pares ($O_i - D_j$):

$$porcentajeTel_{ij} = \frac{tel_{ij}}{TotalTel} \cdot 100, \forall i \in O, \forall j \in D$$

5: Se calcula la movilidad total de transporte de la macrozona origen a la macrozona destino:

$$TotalTp = \sum_{i \in O} \sum_{j \in D} tp_{ij}$$

6: Para finalizar obtenemos la proyección de pasajeros de transporte público para cada par ($O_i - D_j$) de la siguiente forma:

$$proyeccionI_{ij} = porcentajeTel_{ij} \cdot TotalTp, \forall i \in O, \forall j \in D$$

Funcionamiento del algoritmo:

Vamos a realizar una representación de como funcionaria nuestro algoritmo resolviendo un ejemplo con datos ficticios.

Nuestro algoritmo recibe un par de origen central (O_0) y destino central (D_0), como observamos en la figura 3.6.

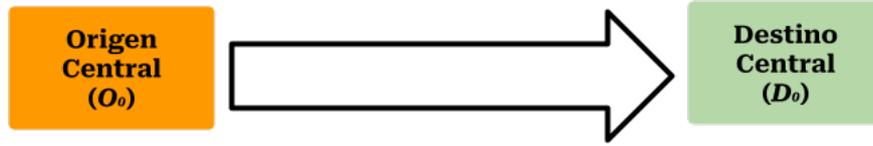


Figura 3.6. Representación gráfica de zona de origen y destino central.

En primer lugar, se hallan las zonas colindantes al origen central y al destino central, obteniendo así las macrozonas de origen y destino que observamos en la figura 3.7, denotando el origen y destino central por O_0 y D_0 , a las zonas colindantes al origen central por O_1, O_2 y O_3 , a las zonas colindantes al destino central por D_1, D_2 y D_3 , a la unión de la zona de origen central y sus zonas colindantes por macrozona de origen O ($\{O_0, O_1, O_2, O_3\}$), a la unión de la zona de destino central y sus zonas colindantes por macrozona de destino D ($\{D_0, D_1, D_2, D_3\}$).

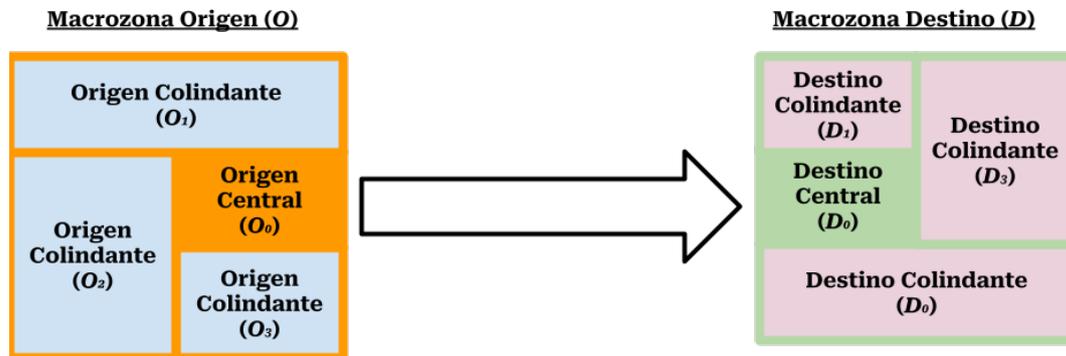


Figura 3.7. Representación gráfica de las macrozonas de origen (O) y destino (D).

En segundo lugar, para cada par de la macrozona origen a la macrozona destino ($O_i - D_j$), se obtiene la movilidad existente tanto en transporte(tp_{ij}) como en telefonía(tel_{ij}), y escogemos los pares en los que su movilidad en telefonía no sean nulos, como observamos en la tabla de la figura 3.8.

O_i	D_j	tp_{ij}	tel_{ij}
O_0	D_0	0	20
O_0	D_1	7	5
O_0	D_2	6	30
O_1	D_2	30	5
O_1	D_3	3	42
O_2	D_0	2	13
O_2	D_3	15	17
O_3	D_0	2	15
O_3	D_1	3	28

Figura 3.8. Datos de movilidad de telefonía y transporte obtenidos por la MOD-TEL y la MOD-TP.

En tercer lugar, sabiendo la cantidad de personas que hay para cada par $(O_i - D_j)$, se calcula el total de la movilidad por telefonía (tel_{ij}):

$$TotalTel = \sum_{i \in O} \sum_{j \in D} tel_{ij} = 20 + 5 + 30 + 5 + 42 + 13 + 17 + 15 + 28 = 175$$

En cuarto lugar, se calcula el porcentaje de aplicación de la movilidad de telefonía para cada uno de los pares $(O_i - D_j)$, como se refleja en la figura 3.9.

$$porcentajeTel_{ij} = \frac{tel_{ij}}{TotalTel} \cdot 100, \forall i \in O, \forall j \in D$$

O_i	D_j	tp_{ij}	tel_{ij}	porcentajeTel _{ij}
O_0	D_0	0	20	0,11
O_0	D_1	7	5	0,03
O_0	D_2	6	30	0,17
O_1	D_2	30	5	0,03
O_1	D_3	3	42	0,24
O_2	D_0	2	13	0,07
O_2	D_3	15	17	0,10
O_3	D_0	2	15	0,09
O_3	D_1	3	28	0,16

Figura 3.9. Tabla que refleja el porcentaje de aplicación de la movilidad de telefonía.

En quinto lugar, se calcula el total de la movilidad por transporte:

$$TotalTp = \sum_{i \in O} \sum_{j \in D} tp_{ij} = 0 + 7 + 6 + 30 + 3 + 2 + 15 + 2 + 3 = 70$$

En sexto y último lugar, se calcula la proyección de pasajeros de transporte para cada uno de los pares $(O_i - D_j)$, como observamos en la tabla de la figura 3.10.

$$proyeccionI_{ij} = porcentajeTel_{ij} \cdot TotalTp, \forall i \in O, \forall j \in D$$

O_i	D_j	tp_{ij}	tel_{ij}	$porcentajeTel_{ij}$	$proyeccionI_{ij}$
O_0	D_0	0	20	0,11	7,70
O_0	D_1	7	5	0,03	2,10
O_0	D_2	6	30	0,17	11,90
O_1	D_2	30	5	0,03	2,10
O_1	D_3	3	42	0,24	16,80
O_2	D_0	2	13	0,07	4,90
O_2	D_3	15	17	0,10	7,00
O_3	D_0	2	15	0,09	6,30
O_3	D_1	3	28	0,16	11,20
Totales		70	175	1	70

Figura 3.10. Tabla que refleja la proyección del total de la movilidad de transporte en cada par ($O_i - D_j$).

Para finalizar vamos a clasificar los resultados obtenidos en la tabla 3.10 de la forma explicada en el comienzo de este capítulo:

- **Zonas de atracción:** D_0, D_1 y D_3
- **Zona de cesión:** D_2
- **Zona neutral**

Vamos a mostrar otro enfoque sobre los resultados en el que vamos sacar unas conclusiones tomando solo los movimientos que toman como origen a nuestro origen central. De O_0 parten 13 personas de transporte público y se reparten de la forma que se observa en la figura 3.11.

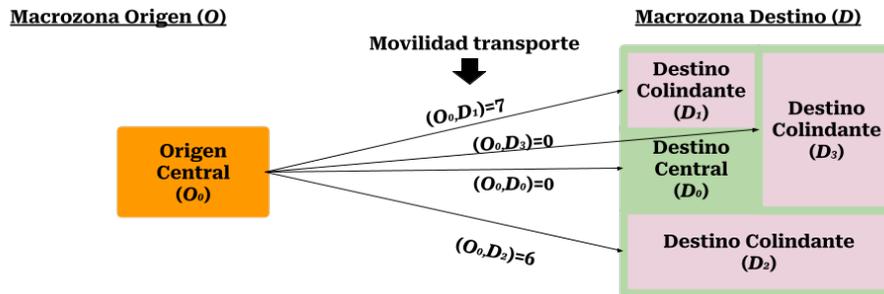


Figura 3.11. Representación gráfica de movilidad de telefonía entre O_0 y D .

Comparando los datos con la figura 3.12, obtenemos que de nuestro destino central existen 21,7 movimientos. Y se refleja que las zonas en amarillo son zonas de atracción, la roja es zona de cesión y la gris es zona neutral.

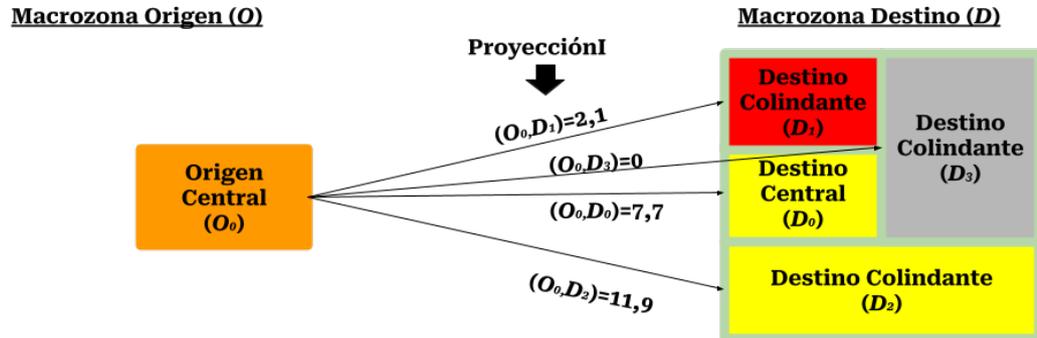


Figura 3.12. Representación gráfica de la proyección obtenida por el algoritmo entre O_0 y D .

3.4. Resolución con Python

Partiendo de la tabla mencionada en el capítulo 2.3., que se trata de un documento en formato CSV, y basándonos en el enfoque y los resultados obtenidos a lo largo del proceso, hemos desarrollado un código en Python que implementa el algoritmo final (lo podemos encontrar en la siguiente ubicación: https://github.com/samirop8/Proyecci-nfinal_matricesOD). Ahora vamos a describir el producto final de nuestro código.

Una vez que se ejecutó el programa en Python el resultado se fundamenta en un nuevo archivo CSV que incluye proyecciones de pasajeros de transporte público para cada par OD del documento original 2.3. Estas proyecciones representan los desplazamientos desde la macrozona de origen (que abarca el origen central y sus zonas colindantes) hacia la macrozona de destino (que abarca el destino central y sus zonas colindantes).

En la figura 3.13 presentamos un proceso gráfico que ilustra el procedimiento desde un par OD del documento original hasta el resultado final. Se tomó como ejemplo el par OD con los siguientes valores: origen (código zona origen): **57**, destino (código zona destino): **1029** y periodo: **17** (correspondiente a las 17:00 horas) del documento original.

Una vez ejecutado el código que implementa el algoritmo final, se obtiene la siguiente tabla que se observa en la figura 3.13. Donde cada fila representa un movimiento entre la macrozona de origen con **origen central 57** y la macrozona de destino con **destino central 1029** en el **periodo de las 17:00**. En esta tabla se presentan las siguientes columnas:

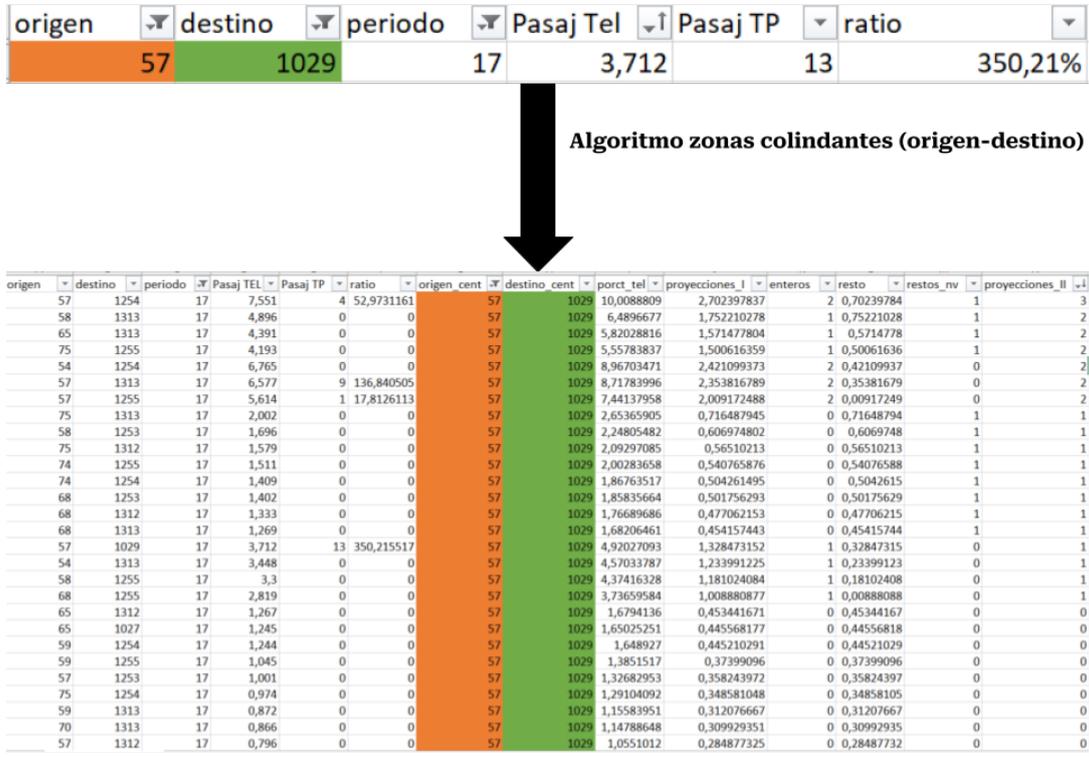


Figura 3.13. Representación gráfica de la proyección obtenida por el algoritmo entre O_0 y D .

- **Origen:** conjunto de orígenes pertenecientes a la macrozona de origen.
- **Destino:** conjunto de destinos pertenecientes a la macrozona de destino.
- **Periodo:** Hora del día.
- **Pasaj TEL:** cantidad de pasajeros por movilidad de telefonía obtenidos por MOD-TEL.
- **Pasaj TP:** cantidad de pasajeros por movilidad de transporte obtenidos por MOD-TP.
- **Ratio:** Pasajeros transporte público/Pasajeros telefonía.
- **Origen_cent:** Origen central.
- **Destino_cent:** Destino central.
- **Porct_tel:** ratios de movilidad de telefonía explicado en 3.9.
- **Proyecciones_I:** proyección explicada en 3.10.
- **Enteros:** parte entera de la proyección.
- **Resto:** parte decimal de la proyección.
- **Resto_nv:** si $resto \geq 0,5 \implies Resto_nv = 1$; si no $Resto_nv = 0$.
- **Proyecciones_II= Entero + Resto_nv.** Obteniendo así una aproximación entera de **Proyecciones_I**.

Como resultado de nuestra solución para cada par OD del documento original, se generan numerosas filas en la solución. Esto hace que, al aplicar el código en Python al documento completo, obtenemos un número de filas que excede el límite máximo de una hoja de cálculo. Por lo tanto, para fines prácticos, nos enfocaremos únicamente en los casos de zonas particulares mencionadas en el capítulo **1.3**.

Resolución de zonas particulares

Para finalizar, en este capítulo presentaremos los resultados específicos obtenidos para las zonas particulares mencionadas en el primer capítulo de este proyecto. Utilizaremos visualizaciones gráficas para ilustrar de manera clara el cambio en la distribución de pasajeros del transporte público en estas áreas. Estas representaciones serán esenciales para comprender y visualizar el impacto de nuestra proyección en el movimiento de pasajeros dentro de estas zonas específicas.

4.1. Zonas de aeropuertos

A continuación, estudiaremos las zonas del Aeropuerto del Sur (zona 1029) y del Aeropuerto del Norte (zona 102), considerando estas zonas como destinos centrales. Seleccionaremos un origen central y un período apropiado para cada caso.

4.1.1. Zona Aeropuerto del sur (1029).

Vamos a estudiar el par OD central entre la zona 1029 y la zona 57, siendo esta la que contiene al intercambiador de Santa Cruz, en el periodo 17:00.

Observamos en la figura 4.1 una representación gráfica que contiene los datos de movilidad de transporte público entre ambas macrozonas, donde las zonas con un color mas oscuro son nuestros origen y destino centrales. En el caso de la macrozona de origen vemos la cantidad de pasajeros salientes en cada zona hacia la macrozona de destino. Y en el caso de la macrozona de destino obtenemos la cantidad de pasajeros entrantes en cada zona desde la macrozona de origen.

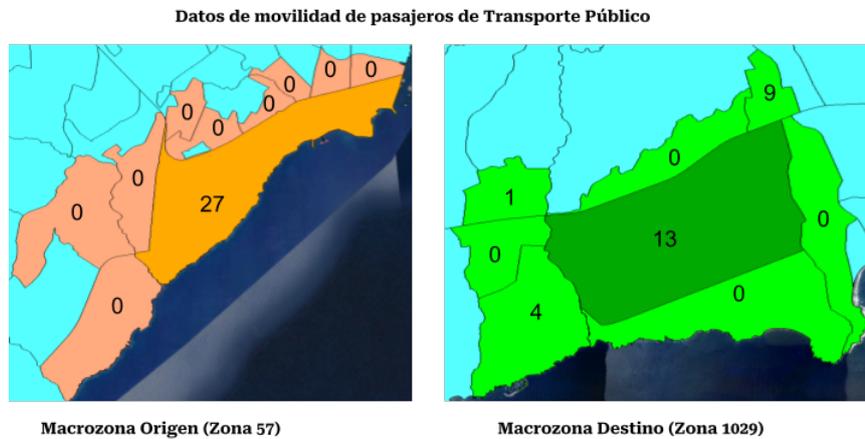


Figura 4.1. Representación gráfica de los datos de movilidad entre las zonas 57 y 1029.

En la figura 4.2 observamos una representación gráfica de la proyección final de los pasajeros de transporte público.

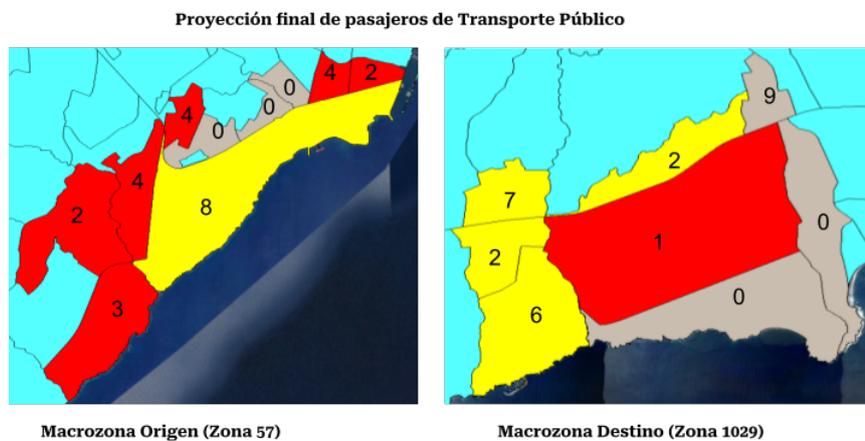


Figura 4.2. Representación gráfica de la proyección final entre las zonas 57 y 1029.

Vamos a explicar los resultados obtenidos para cada macrozona:

- **Macrozona Origen:** En un principio, observamos que desde nuestro origen central salieron 27 personas, mientras que ninguna de las zonas colindantes presentó movimientos de pasajeros. Sin embargo, en nuestra proyección, vemos que las zonas colindantes también generan movimientos hacia nuestro origen central para utilizar el transporte público. Esto convierte a nuestro origen central en una zona de atracción (representada en color amarillo), a

las zonas colindantes con mayor número de pasajeros salientes en zonas de cesión (color rojo), y a las zonas que siguen sin presentar pasajeros salientes en zonas neutrales (color gris oscuro), como se muestra en la representación gráfica.

- Macrozona Destino:** En esta observación, notamos que inicialmente había 13 personas dirigiéndose a nuestro destino central. Sin embargo, en la proyección, vemos que se convierte en una zona de cesión (representada en color rojo), ya que una persona termina ingresando a esta zona y las restantes 12 se distribuyen en algunas de sus zonas colindantes, las cuales se convierten en zonas de atracción (representadas en color amarillo). Las zonas con la misma cantidad de personas que ingresan y salen se consideran zonas neutrales (representadas en color gris oscuro).

4.1.2. Zona Aeropuerto del norte (102).

Vamos a estudiar el par OD central entre la zona 102 y la zona 96, siendo esta la que contiene al intercambiador de La Laguna, en el periodo 14:00. Cabe destacar que en este caso coinciden algunas de las zonas colindantes de nuestro origen central y de nuestro destino central.

Observamos en la figura 4.3 una representación gráfica que contiene los datos de movilidad de transporte público entre ambas macrozonas, donde las zonas con un color mas oscuro son nuestros origen y destino centrales. En el caso de la macrozona de origen vemos la cantidad de pasajeros salientes en cada zona hacia la macrozona de destino. Y en el caso de la macrozona de destino obtenemos la cantidad de pasajeros entrantes en cada zona desde la macrozona de origen.

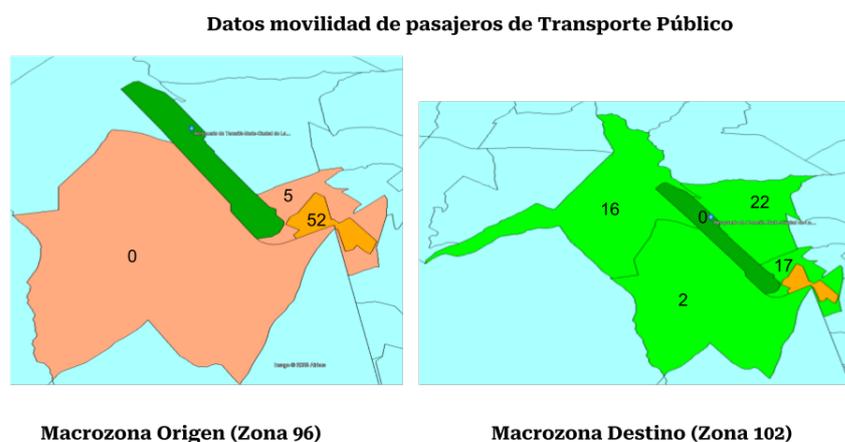


Figura 4.3. Representación gráfica de los datos de movilidad entre las zonas 57 y 1029.

En la figura 4.4 observamos una representación gráfica de la proyección final de pasajeros de transporte público.

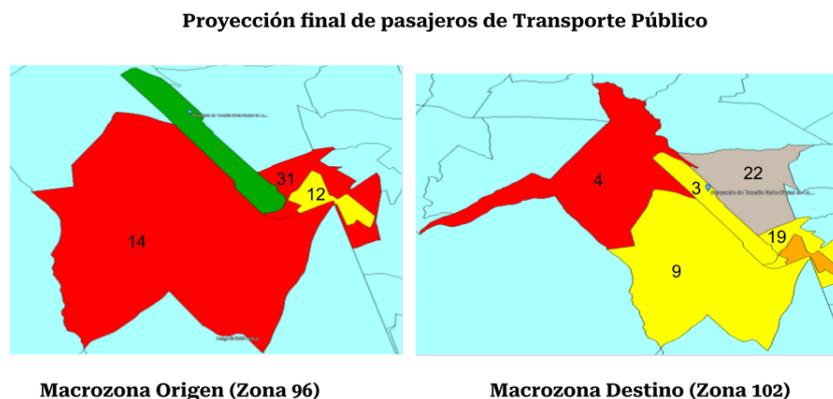


Figura 4.4. Representación gráfica de la proyección final entre las zonas 57 y 1029.

Vamos a analizar los resultados obtenidos en ambas macrozonas.:

- **Macrozona Origen:** Podemos observar que inicialmente salen 52 personas de nuestro origen central, 5 de la zona colindante y ninguna de la zona de la izquierda. Sin embargo, en nuestra proyección se muestra que nuestro origen central se convierte en una zona de atracción (color amarillo), ya que muchas personas se desplazan hacia ella desde las zonas colindantes de cesión (color rojo) para utilizar el transporte público. Además, podemos observar nuestro destino central en color verde..
- **Macrozona Destino:** Observamos que inicialmente no había ninguna persona que se dirigiera a nuestro destino central, y las 57 personas que salían de la macrozona de origen se distribuían en sus zonas colindantes. Sin embargo, después de la proyección, podemos observar que tres personas entran en la zona del aeropuerto, convirtiéndola en una zona de atracción (color amarillo) y generando que una de sus zonas colindantes se convierta en una zona de cesión (color rojo). Además, dos zonas colindantes experimentan un aumento en el número de pasajeros que entran por transporte público, también convirtiéndose en zonas de cesión, mientras que una zona mantiene el mismo número de pasajeros, siendo una zona neutral.

4.2. Zona de intercambiadores

En este caso vamos a estudiar ambos Intercambiadores juntos, donde primero vamos a tomar el intercambiador de La Laguna (zona 96) como origen

central y el intercambiador de Santa Cruz (zona 57) como destino central, y viceversa posteriormente.

4.2.1. Par OD: Intercambiador La Laguna - Intercambiador Santa Cruz

Vamos a tomar el origen central 96 y el destino central 57, en el periodo 14:00 para realizar este estudio.

Podemos observar que las dos macrozonas están en la figura 4.5, por lo que en la imagen de la izquierda, la macrozona de origen va a ser la zona naranja, siendo la que tiene un color naranja mas oscuro el origen central, y la macrozona de destino la que tiene un color verde, siendo la zona con un color verde mas oscuro el destino central.

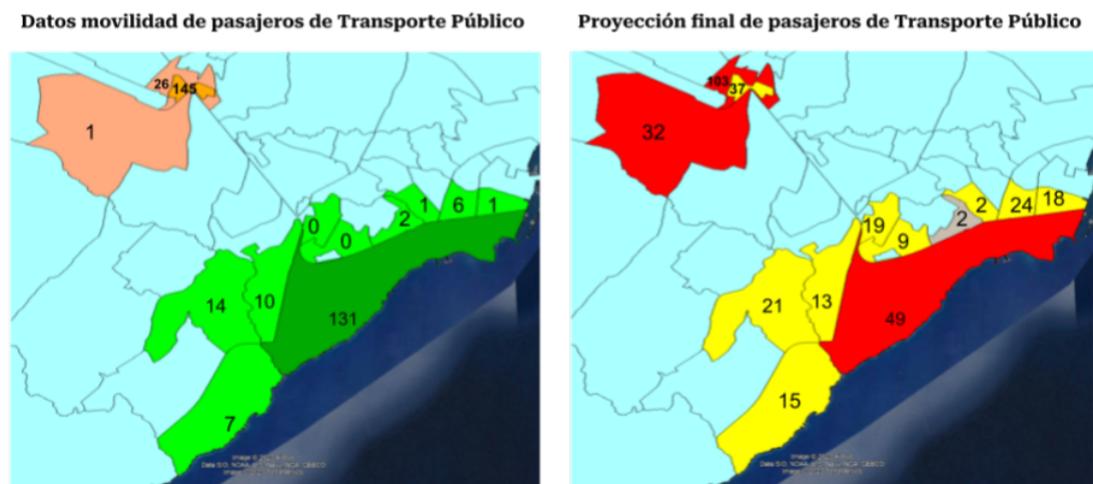


Figura 4.5. Representación gráfica de la movilidad y proyección final de pasajeros de transporte público entre las zonas 96 y 57.

A partir de los resultados mostrados en la figura 4.5, procederemos a explicar lo obtenido para cada macrozona de manera detallada.:

- Macrozona Origen (Naranja):** Obtenemos que inicialmente hay 145 personas saliendo de nuestro origen central, 26 personas saliendo de la zona colindante y 1 persona saliendo de la zona de la izquierda. Sin embargo, en nuestra proyección se observa que nuestro origen central se convierte en una zona de atracción (color amarillo), ya que un gran número de personas se desplazan hacia ella desde las zonas colindantes que se convierten en zonas de cesión (color rojo).

- Macrozona Destino (Verde):** En la figura 4.5, podemos observar que inicialmente hay 172 personas saliendo de la macrozona de origen, con 131 de ellas dirigiéndose a nuestro destino central y el resto distribuyéndose en las zonas colindantes. Sin embargo, en nuestra proyección se observa que a la zona del intercambiador de Santa Cruz entran 49 personas, convirtiéndola en una zona de cesión (color rojo), y como resultado la mayoría de sus zonas colindantes se convierten en zonas de atracción (color amarillo), excepto una que mantiene 2 personas entrantes y se considera zona neutral (color gris).

4.2.2. Par OD: Intercambiador Santa Cruz - Intercambiador La Laguna

Vamos a tomar el origen central 57 y el destino central 96, en el periodo 07:00 para realizar este estudio. Podemos observar que las dos macrozonas están en la figura 4.5, por lo que en la imagen de la izquierda, la macrozona de origen va a ser la zona naranja, siendo la que tiene un color naranja mas oscuro el origen central, y la macrozona de destino la que tiene un color verde, siendo la zona con un color verde mas oscuro el destino central.

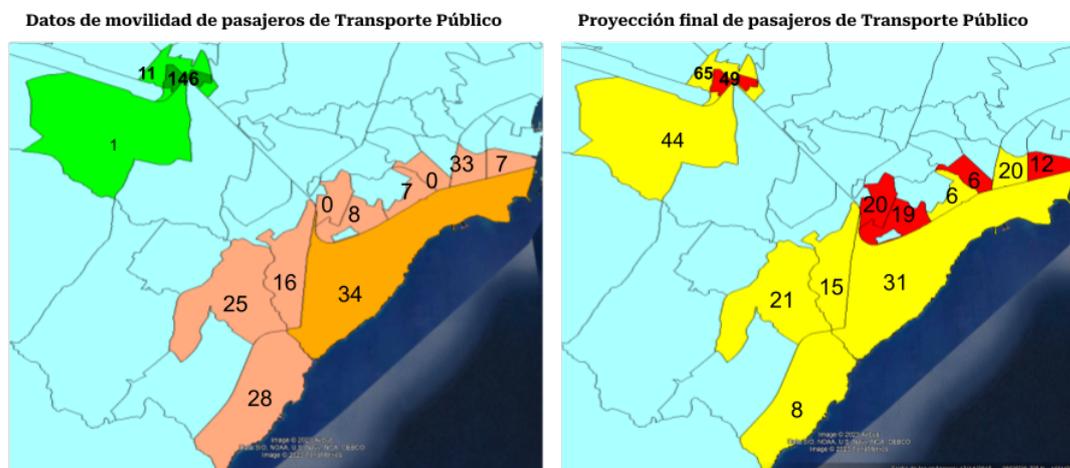


Figura 4.6. Representación gráfica de la movilidad y proyección final de pasajeros de transporte público entre las zonas 57 y 96.

En base a los resultados mostrados en la figura 4.6, analizaremos lo obtenido para cada macrozona:

- Macrozona Origen(Naranja):** Se observa que inicialmente salen 34 personas de nuestro origen central, mientras que de las zonas colindantes salen 124 personas. En la proyección, podemos ver que la mayoría de las zonas se

convierten en zonas de atracción, ya que reciben personas de las zonas colindantes que utilizan el transporte público para dirigirse hacia la macrozona de destino.

- **Macrozona Destino(Verde):** A partir de la figura 4.6, obtenemos que inicialmente había 146 personas dirigiéndose a nuestro destino central, mientras que las otras 158 personas que salían de la macrozona de origen se distribuían en sus zonas colindantes. Después de la proyección, se observa que a la zona del intercambiador de La Laguna ingresan 49 personas, convirtiéndola en una zona de cesión (color rojo), mientras que sus dos zonas colindantes se convierten en zonas de atracción (color amarillo).

4.3. Zonas de hospitales

4.3.1. Hospital Universitario de Canarias (zona 77)

En este caso vamos a estudiar la zona que contiene el Hospital Universitario de Canarias (zona 77), para ello vamos a estudiar el par OD tomando como origen central la zona 96, siendo el intercambiador de La Laguna, y el destino central la zona 77, en el periodo 14:00.

Podemos observar que las dos macrozonas están en la figura 4.7, por lo que en la imagen de la izquierda, la macrozona de origen va a ser la zona naranja, siendo la que tiene un color naranja mas oscuro el origen central, y la macrozona de destino la que tiene un color verde, siendo la zona con un color verde mas oscuro el destino central.

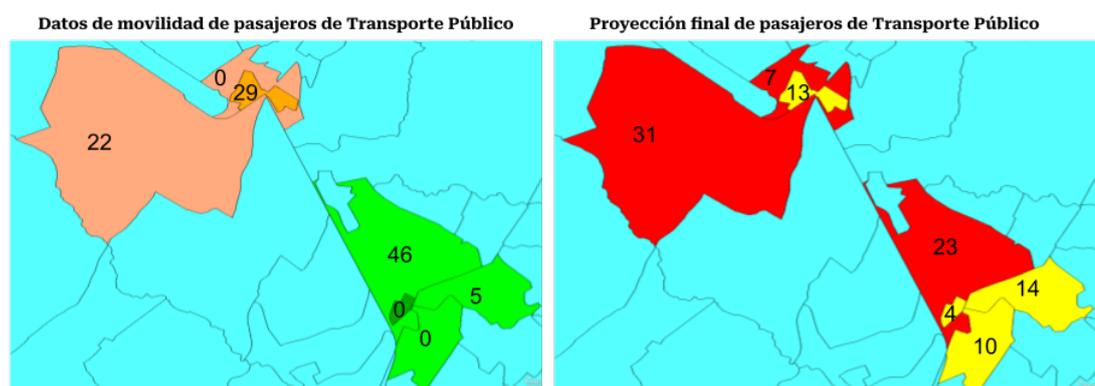


Figura 4.7. Representación gráfica de la movilidad y proyección final de pasajeros de transporte público entre las zonas 96 y 77.

A continuación, analizaremos los resultados obtenidos para cada macrozona, basándonos en la información presentada en la figura 4.7.

- **Macrozona Origen(Naranja):** Analizando los datos presentados en la figura 4.7, se puede observar que inicialmente se registra un flujo de salida de 34 personas desde nuestro origen central, mientras que las zonas colindantes contribuyen con 124 personas. Sin embargo, al realizar la proyección, se destaca que la mayoría de las zonas actúan como zonas de atracción, ya que reciben pasajeros de sus zonas colindantes con el propósito de utilizar el transporte público para dirigirse hacia la macrozona de destino.
- **Macrozona Destino(Verde):** Basándonos en la información proporcionada en la figura 4.7, se destaca el flujo inicial de 146 personas hacia nuestro destino central, mientras que el resto de las 158 personas que salían de la macrozona de origen se distribuían en las zonas colindantes. Sin embargo, al realizar la proyección, se observa que la zona del intercambiador de La Laguna registra la entrada de 49 personas, convirtiéndose en una zona de cesión (color rojo), mientras que sus dos zonas colindantes se transforman en zonas de atracción (color amarillo).

4.4. Estudio de coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson es una medida de dependencia lineal entre dos variables aleatorias cuantitativas.

De manera menos formal, podemos definir el coeficiente de correlación de Pearson como un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas y continuas.

Dado un par de variables aleatorias (X,Y) , el coeficiente de correlación poblacional de Pearson (también denotado por $\rho_{X,Y}$) se define como:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov_{XY}}{\sqrt{Var_X Var_Y}}$$

Donde

σ_{XY} es la covarianza de $((X, Y))$

σ_X es la desviación típica de la variable X

σ_Y es la desviación típica de la variable Y

Sabiendo que el coeficiente de correlación de Pearson oscila entre -1 y $+1$.

- **Sí $\rho < 0$,** existe una correlación negativa, es decir, que las dos variables están asociadas en sentido inverso. Cuánto más se acerca a -1 , mayor es la fuerza de esa relación invertida (cuando el valor en una sea muy alto, el valor en la otra será muy bajo). Cuando es exactamente -1 , eso significa que tienen una correlación negativa perfecta.

- Sí $\rho > 0$, existe una correlación positiva. En este caso las variables estarían asociadas en sentido directo. Cuanto más cerca de +1, más alta es su asociación. Un valor exacto de +1 indicaría una relación lineal positiva perfecta.
- Sí $\rho = 0$, o próxima a 0, indica que no hay relación lineal entre las dos variables.

Ahora, tomando la tabla de la figura 4.8 que muestra los resultados del par OD central 97-77 en el periodo 14:00, estudiado previamente, como ejemplo para mostrar la correlación entre los datos de movilidad de telefonía (**Pasaj TEL**) y los datos de movilidad de transporte público (**Pasaj TP**), y la mejora de correlación obtenida al comparar los datos de movilidad de telefonía con la proyección final (**Proyecciones_II**).

Origen	Destino	Pasaj TEL	Pasaj TP	Proyecciones_I	Proyecciones_II
97	77	5,97	0	1,94	2
61	77	2,18	0	0,71	1
96	75	3,01	0	0,98	1
96	77	1,52	0	0,49	1
96	76	21,20	25	6,90	7
97	78	24,19	1	7,87	8
96	78	10,85	4	3,53	4
97	75	25,96	0	8,45	8
61	75	4,25	0	1,38	1
97	76	40,99	21	13,34	13
61	78	7,05	0	2,29	2
61	76	9,40	0	3,06	3

Figura 4.8. Tabla que refleja la proyección de la movilidad de TP para el par OD central 96-77 en el periodo 14:00.

Siendo X la variable que toma los datos de **Pasaj TEL**, e Y la variable que toma los datos de **Pasaj TP**, obtenemos que el coeficiente de relación de Pearson es $\rho_{X,Y} = 0,66$. Ahora, tomando Z a la variable que toma los datos de **Proyecciones_II**, al calcular su coeficiente de correlación, $\rho_{X,Z} = 0,99$, observamos una casi perfecta correlación que refleja claramente una mejoría.

Conclusiones

Finalmente se ha logrado obtener una proyección que permite distribuir los pasajeros de transporte público de manera acorde a la distribución de nuestra matriz OD de telefonía (MOD-TEL) para la gran mayoría de los pares OD de nuestro documento inicial. Se realizaron estudios de correlación entre los datos de movilidad de telefonía y la proyección final, lo cual reveló un alto coeficiente de correlación para la mayoría de los pares OD. Sin embargo, es importante mencionar que, debido a la metodología de generación de la proyección, existe un pequeño porcentaje de casos en los que no hay movilidad de transporte público entre determinadas macrozonas, a pesar de existir movilidad de telefonía. Estos casos particulares resultan en una proyección nula y una correlación baja entre ambos conjuntos de datos.

Además los datos obtenidos brindan múltiples posibilidades de acción. El enfoque escogido, se centra en los movimientos entre macrozonas, pero se podría estudiar cada movimiento e incluso compararlos en diferentes momentos del día. Es importante destacar la cantidad considerable de tiempo invertido en la empresa, donde mi función consistió en contrastar ideas con mis compañeros, aportando mi conocimiento matemático hasta llegar al algoritmo final que se ha reflejado en este proyecto. Sin embargo, quiero mencionar que también planteamos un algoritmo alternativo que podría ser más preciso, basado en el uso de zonas isócronas en lugar de zonas colindantes. Es decir, considerar las zonas a las que se puede llegar caminando desde nuestras zonas de origen y destino centrales para realizar el estudio.

A pesar de todo estamos muy satisfechos con los resultados obtenidos. La idea final se basa en automatizar el proceso para que pueda recibir distintas matrices OD tanto de telefonía como de transporte público y generar la proyección de forma directa.

Bibliografía

- [1] DIAZ DE SANTOS, ED. y SALAZAR GONZÁLEZ, J.J. *Programación Matemática*. 2001.
- [2] TITSA S.L., Departamento Big Data & Data Science. *Algoritmos y procesos*. 2023.
- [3] TITSA S.L., Departamento Big Data & Data Science. *MATRIZ ORIGEN Y DESTINO - OPEN DATA*. 2023.
- [4] VANDERPLAS, J. *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly, 2016.
- [5] LACHEV, Teo. *Applied Microsoft Power BI. Bring your data to life*. Editor: Edward Price, 2016.
- [6] TITSA S.L., Departamento Big Data & Data Science. *Power BI*. 2023.
- [7] GOOGLE EARTH PRO. *Zonas transporte público de Tenerife [Mapa]*. [Fecha de consulta: 08-07-2023]. 2023.
- [8] STUDOCU. *Correlación de Pearson* [en línea]. [Fecha de consulta: 08-07-2023]. Disponible en: <https://www.studocu.com/en-us/document/university-of-northern-iowa/analisis-y-desarrollo-de-sistemas-de-informacion/correlacion-de-pearson/13267068>.

Improvement in the distribution of passengers in public transportation with respect to the OD matrices.

Samir Oussaid Padron

Facultad de Ciencias • Sección de Matemáticas
Universidad de La Laguna
alu0101278650@ull.edu.es

Abstract

The island of Tenerife is divided into 216 public transportation zones according to the drafting of the Special Territorial Plan for Transport Organization in Tenerife. Low correlation errors have been identified between the Telephony Origin-Destination (OD) matrix and the Public Transportation OD matrix. These errors present a significant challenge for the company TITSA. In this undergraduate thesis, we aim to address this problem by generating a projection that allows us to understand the movement of public transportation passengers when completing their journeys between nearby zones, based on the the Telephony OD matrix. We will provide a detailed explanation of the process and the final algorithm used, along with its implementation in Python. Additionally, we will present graphical solutions for specific cases to illustrate our findings.

1. Introduction



Figure 1: Graphical representation of the zoning division of Transportation in Tenerife.

Firstly, the process of obtaining the matrices will be explained, and the causes of the lack of correlation will be analyzed, being one of them the zoning division of Transportation in the island. Specific cases will also be presented to illustrate the proposed solutions. The applications and utilities of the OD matrices will be explored, highlighting their use in prediction methods through artificial intelligence and machine learning. Visualizations using Power BI will be created to compare both matrices. Next, the resolution of the problem will be addressed by presenting preliminary algorithms and the final algorithm. Implementation in Python will be detailed, and the obtained results will be explained. Finally, the solution to selected specific cases will be graphically shown, allowing the understanding of the movement of public transportation passengers between nearby zones.

2. Resolution with Python

The final algorithm has been implemented in Python, and this will generate a solution for all cases. However, we have chosen a few particular cases to show the final product. These zones are the ones that includes the airports of the island, the main public transport facility of Santa cruz and La Laguna, and the University Hospital of the Canary Islands.

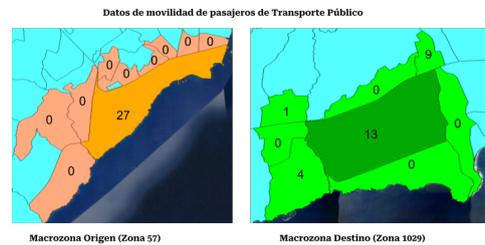


Figure 2: Graphical representation of public transport passengers movements for the OD pair 96-57.

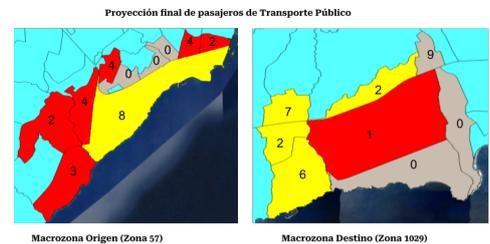


Figure 3: Graphical representation of the solution for the OD pair 96-57.

In the first image, we can observe the number of passengers traveling between both zones using public transportation. In the second image, we obtain the final projection that displays the movement made by public transportation passengers at the end of their journey.

References

- [1] TITSA, Departamento Big Data & Data Science. *Algoritmos y procesos*. 2022.
- [2] TITSA, Departamento Big Data & Data Science. *MATRIZ ORIGEN Y DESTINO - OPEN DATA*. 2022.
- [3] DIAZ DE SANTOS, ED. y SALAZAR GONZÁLEZ, J.J. *Programación Matemática*. 2001.