Universidad Nacional de San Agustin de Arequipa

Año de la recuperación y consolidación de la economía peruana

Escuela Profesional de Ciencia de la Computación

Pipeline Ciencia de Datos

Informe Pipeline

A-2025

Docente: Dra. Ana Maria CuadrosValdivia

Javier Wilber Quispe Rojas

Índice

1. Introdución

En este informe, se presentan los resultados del análisis aplicado al dataset de movilidad y el dataset social de Boston y la movilidad de este . El objetivo es identificar los problemas presentes en los datos, los patrones encontrados y cómo se relacionan con la segregación racial y socioeconómica en la ciudad. Se describen los métodos empleados para resolver los problemas en los datos y las hipótesis planteadas.

2. ¿Qué problemas se identifican en el dataset?

El análisis del dataset reveló varios problemas clave:

2.1. Problema 1: Diferentes unidades geográficas en los datasets

Los datasets de movilidad y social estaban en diferentes unidades geográficas: el dataset social utilizaba unidades de `CBG` (Census Block Group), mientras que el dataset de movilidad utilizaba unidades de `CT` (Census Tract).

	GEOID	white	black	asian	hispanic
0	250250002011	0.678457	0.015273	0.152733	0.061897
1	250250002012	0.695652	0.000000	0.097826	0.043478
2	250250002013	0.892231	0.027569	0.033417	0.046784
3	250250002014	0.569620	0.142405	0.055907	0.159283
4	250250002021	0.792929	0.030303	0.037598	0.130752

Figura 1: white

	geoid_o	geoid_d	visitor_flows
0	25025000201	25025000201	670
1	25025000201	25025000202	643
2	25025000201	25025000301	53
3	25025000201	25025000302	205
4	25025000201	25025000401	96
...	...	...	...

Figura 2: white_inflow

Solución: Se aplicaron funciones para convertir las unidades de `CBG` a `CT`, lo que permitió la armonización de las unidades geográficas en ambos datasets. Esto facilitó la comparación y análisis conjunto de los datos.

	tract_id	white_CT	black_CT	asian_CT	hispanic_CT	wh
0	25025000201	0.761211	0.039623	0.072337	0.068144	
1	25025000202	0.730743	0.026785	0.117305	0.118214	
2	25025000301	0.632481	0.083329	0.113383	0.142914	
3	25025000302	0.669512	0.047004	0.133369	0.101856	
4	25025000401	0.710345	0.028156	0.139056	0.095679	
...	...	...	...	...	...	
145	25025140300	0.137746	0.552668	0.004981	0.300058	
146	25025140400	0.065716	0.644845	0.001411	0.257817	
147	25025981100	0.215085	0.433291	0.006806	0.326670	
148	25025981202	0.858921	0.070539	0.000000	0.037344	
149	25025981800	0.687500	0.312500	0.000000	0.000000	

150 rows x 32 columns

Figura 3

	geoid_o	geoid_d	visitor_flows
0	25025000201	25025000201	670
1	25025000201	25025000202	643
2	25025000201	25025000301	53
3	25025000201	25025000302	205
4	25025000201	25025000401	96
...	...	...	...

Figura 4

2.2. Problema 2: Valores nulos en el dataset de movilidad

El dataset de movilidad contenía una cantidad significativa de valores nulos, que afectaban el análisis de los flujos de movilidad y otros aspectos relevantes.

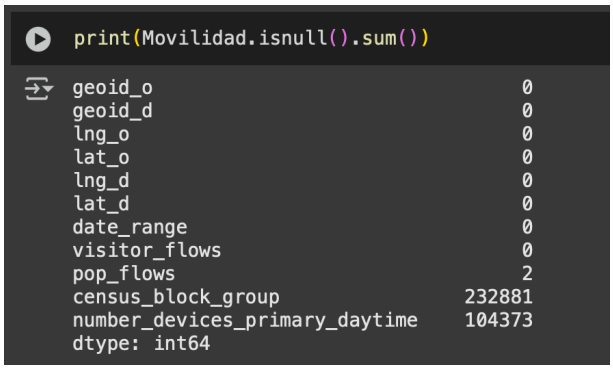
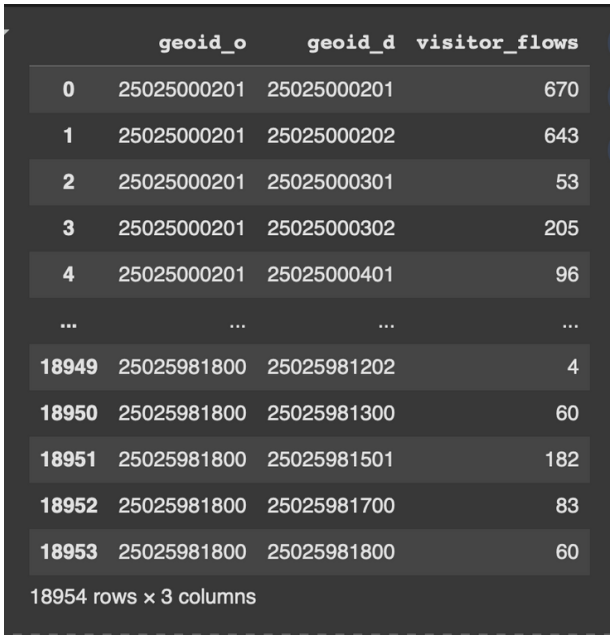


Figura 5: Caption

Solución: Se eliminaron las columnas con más del 50 % de valores nulos, lo que permitió simplificar el dataset y mejorar la precisión de los resultados sin perder información esencial.



	geoid_o	geoid_d	visitor_flows
0	25025000201	25025000201	670
1	25025000201	25025000202	643
2	25025000201	25025000301	53
3	25025000201	25025000302	205
4	25025000201	25025000401	96
...	...	...	...
18949	25025981800	25025981202	4
18950	25025981800	25025981300	60
18951	25025981800	25025981501	182
18952	25025981800	25025981700	83
18953	25025981800	25025981800	60

18954 rows × 3 columns

Figura 6: Caption

3. ¿Qué descubrieron al analizar los datos?

3.1. Distribución racial desigual

Algunas áreas de Boston presentan una alta concentración de población blanca, mientras que otras áreas tienen una mayor proporción de población negra, asiática o hispana. Esto puede reflejar la segregación residencial o patrones históricos de asentamiento en la ciudad.

Figura 7

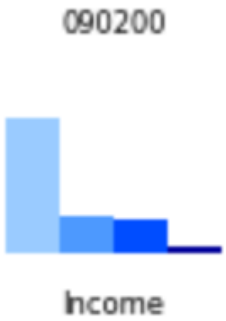


Figura 8

3.2. Valores atípicos (outliers)

Se identificaron valores atípicos en los datos de movilidad, especialmente en columnas relacionadas con las concentraciones de población y actividades en ciertas zonas. Estos outliers no eran errores, sino representaciones válidas de fenómenos sociales y económicos, como grandes concentraciones de grupos sociales o actividades económicas.

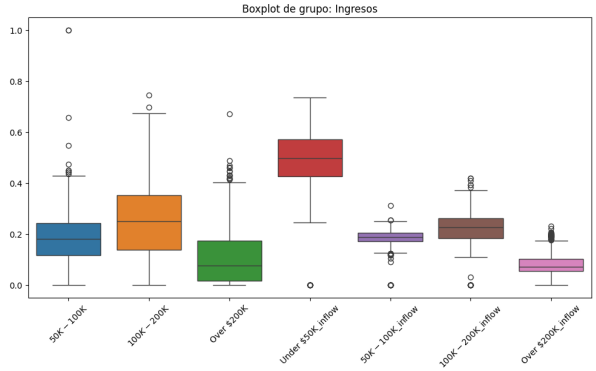


Figura 9

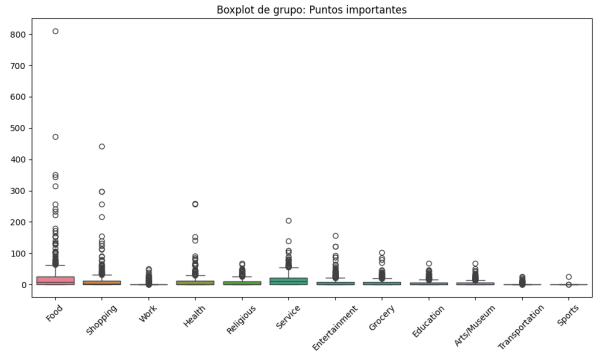


Figura 10

4. ¿Qué reflejan los patrones de tendencias?

Los patrones de tendencias en los datos reflejan varias dinámicas sociales y económicas clave:

4.1. Segregación racial y socioeconómica

Los patrones de segregación racial y socioeconómica en Boston reflejan que ciertos grupos sociales están concentrados en áreas específicas, mientras que las áreas con mayor diversidad tienen mayores flujos de personas, indicando movimientos hacia zonas con mejores oportunidades económicas o de vivienda.

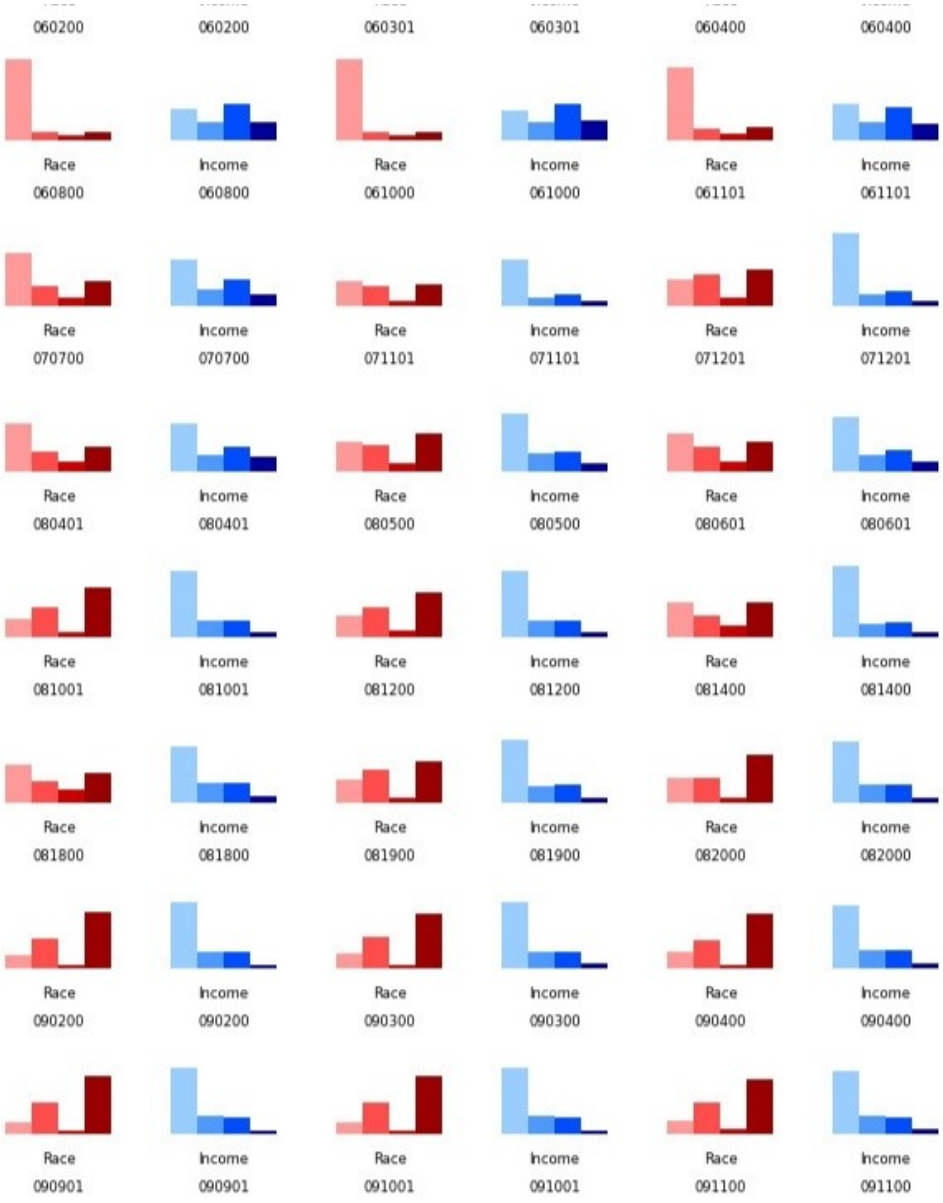


Figura 11

4.2. Flujos de movilidad

Los flujos de movilidad en las zonas con mayores valores de `visitor_flows` indican que existen rutas de transporte clave y áreas de trabajo que atraen a personas de diferentes orígenes, lo que también podría reflejar la falta de oportunidades en otras áreas más segregadas.

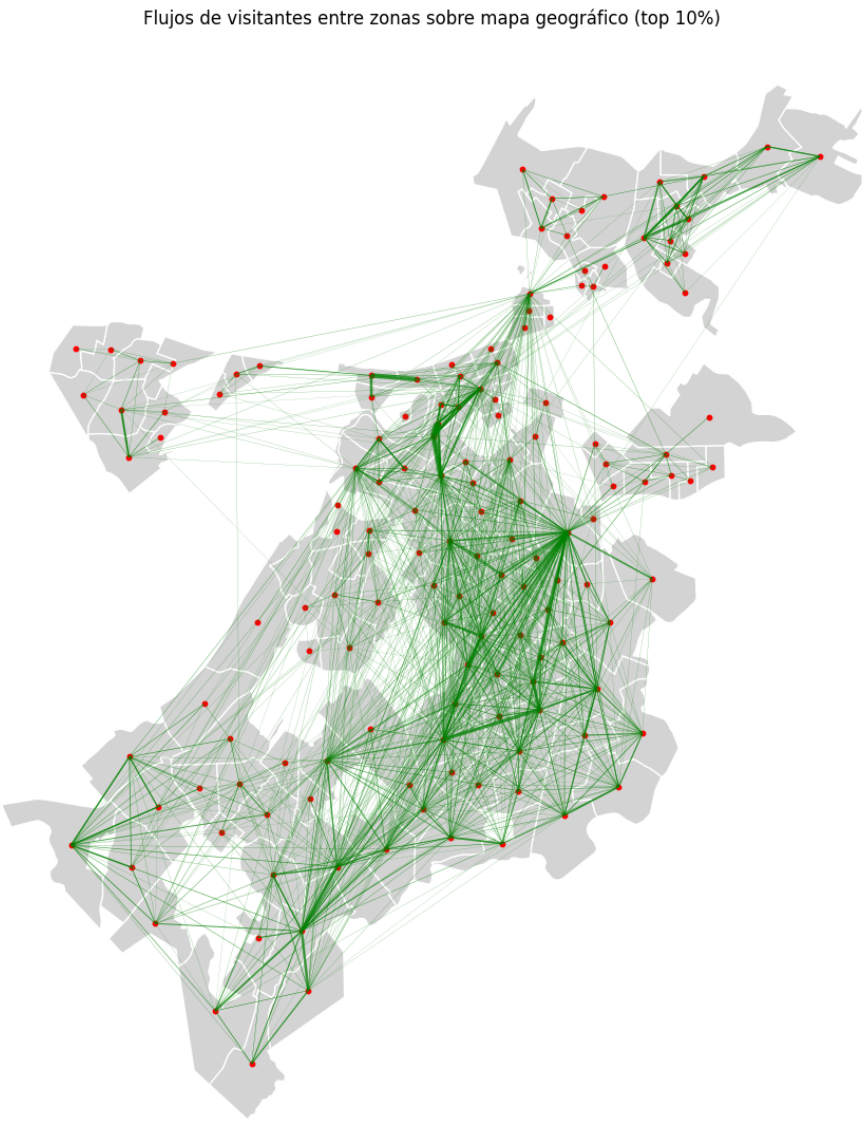


Figura 12

5. Hipótesis

Hipótesis: Se plantea la hipótesis de que los niveles de segregación racial y socioeconómica en un área específica afectan no solo la composición poblacional, sino también los flujos de movilidad hacia esas áreas. Se espera que las áreas con una mayor concentración de un grupo racial o socioeconómico presenten patrones de movilidad más segregados y menos interconexión con otras zonas de la ciudad.

Respuesta a la Hipótesis: Al analizar los datos, se observó que la hipótesis se cumple en gran medida. Las áreas con una alta concentración de grupos sociales específicos, ya sea en términos raciales o de ingresos, presentan patrones de movilidad más segregados. Esto se refleja en los flujos de movilidad hacia y desde esas áreas, que muestran una menor interconexión con otras zonas de la ciudad. Además, las rutas de transporte y las zonas de trabajo parecen estar más concentradas en ciertas áreas, lo que refuerza la segregación espacial.

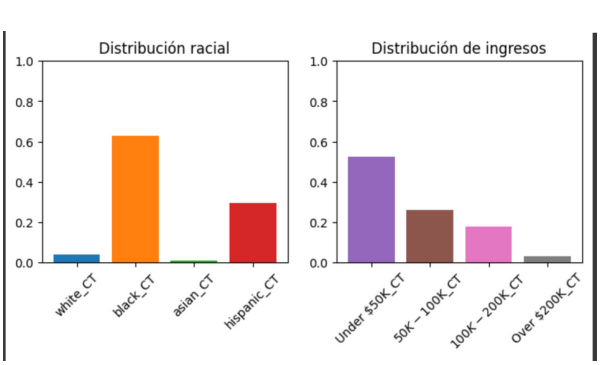


Figura 13

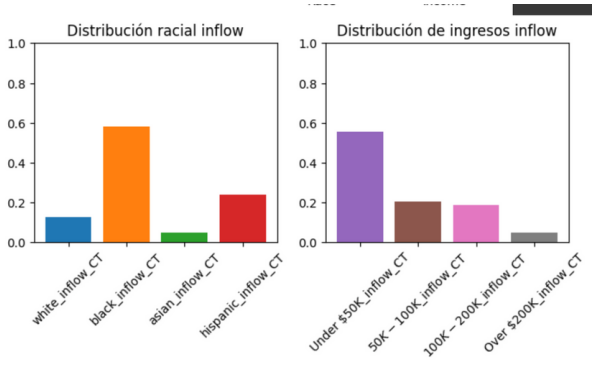


Figura 14

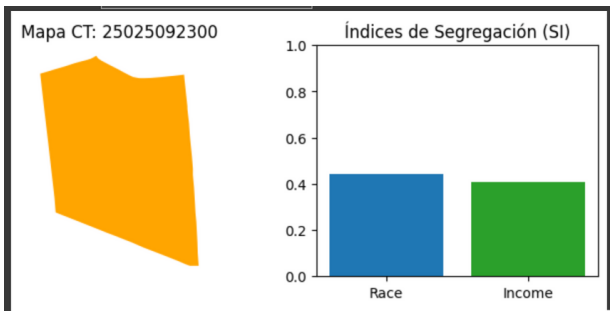


Figura 15: Caption