



**Aplicando pipeline a mi
Dataset**

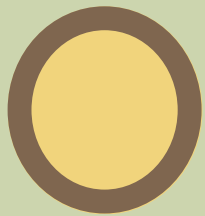
Javier Quispe Rojas

1. ¿ QUE PROBLEMAS SE IDENTIFICAN EN EL DATASET?

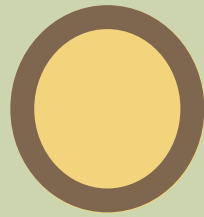
	GEOID	white	black	asian	hispanic
0	250250002011	0.678457	0.015273	0.152733	0.061897
1	250250002012	0.695652	0.000000	0.097826	0.043478
2	250250002013	0.892231	0.027569	0.033417	0.046784
3	250250002014	0.569620	0.142405	0.055907	0.159283
4	250250002021	0.792929	0.030303	0.037598	0.130752

	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
...

- Los Geoid estan en diferentes unidades grograficas el dataset de social esta en unidad CBG (Census Block Group) mientras que el de movilidad esta en unidades CT(Census Tract)



Como lo resolvi

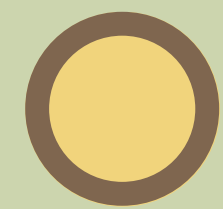


	tract_id	white_CT	black_CT	asian_CT	hispanic_CT	wh
0	25025000201	0.761211	0.039623	0.072337	0.068144	
1	25025000202	0.730743	0.026785	0.117305	0.118214	
2	25025000301	0.632481	0.083329	0.113383	0.142914	
3	25025000302	0.669512	0.047004	0.133369	0.101856	
4	25025000401	0.710345	0.028156	0.139056	0.095679	
...	
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	

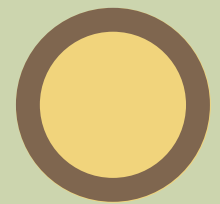
150 rows x 32 columns

	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
...

Aplicamos funciones para poder volver de la unidad CBG a la CT



1. ¿ QUE PROBLEMAS SE IDENTIFICAN EN EL DATASET?



El dataset de movilidad , poseia mchos valores nulos , el cual decidimos eliminarlo ya que no era información indispensable y abos pasaban el 50 %.

```
▶ print(Movilidad.isnull().sum())
```

```
⇒ geoid_o      0  
   geoid_d      0  
   lng_o       0  
   lat_o       0  
   lng_d       0  
   lat_d       0  
   date_range  0  
   visitor_flows 0  
   pop_flows   2  
   census_block_group 232881  
   number_devices_primary_daytime 104373  
   dtype: int64
```

Como lo resolvi

Eliminamos las columnas que no que
tenian valores nuevos y también
columnas que no nos ayudaban para
ver mas claro la movilidad.

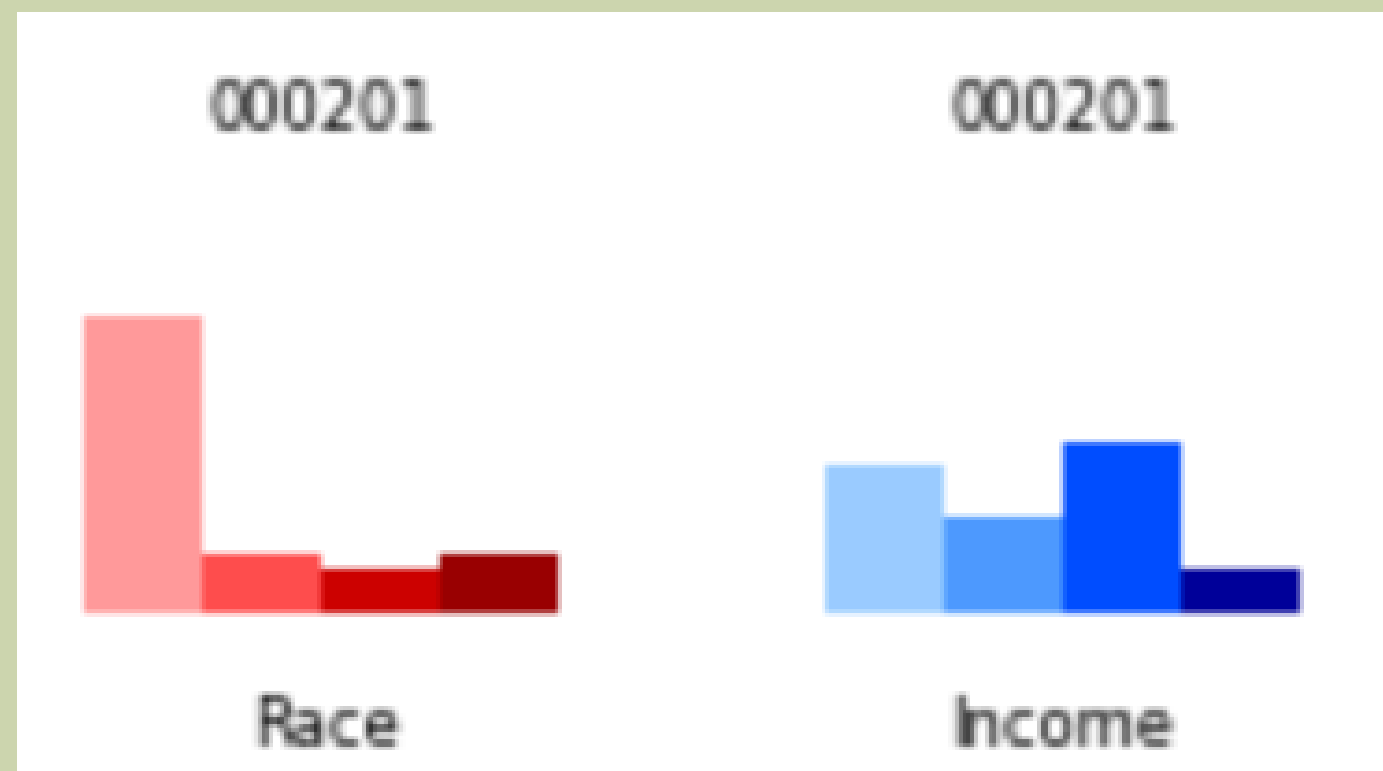
	geoid_o	geoid_d	visitor_flows
0	25025000201	25025000201	670
1	25025000201	25025000202	643
2	25025000201	25025000301	53
3	25025000201	25025000302	205
4	25025000201	25025000401	96
...
18949	25025981800	25025981202	4
18950	25025981800	25025981300	60
18951	25025981800	25025981501	182
18952	25025981800	25025981700	83
18953	25025981800	25025981800	60
18954 rows x 3 columns			



2.- ¿ QUÉ DESCUBRIERON AL ANALIZAR LOS DATOS?



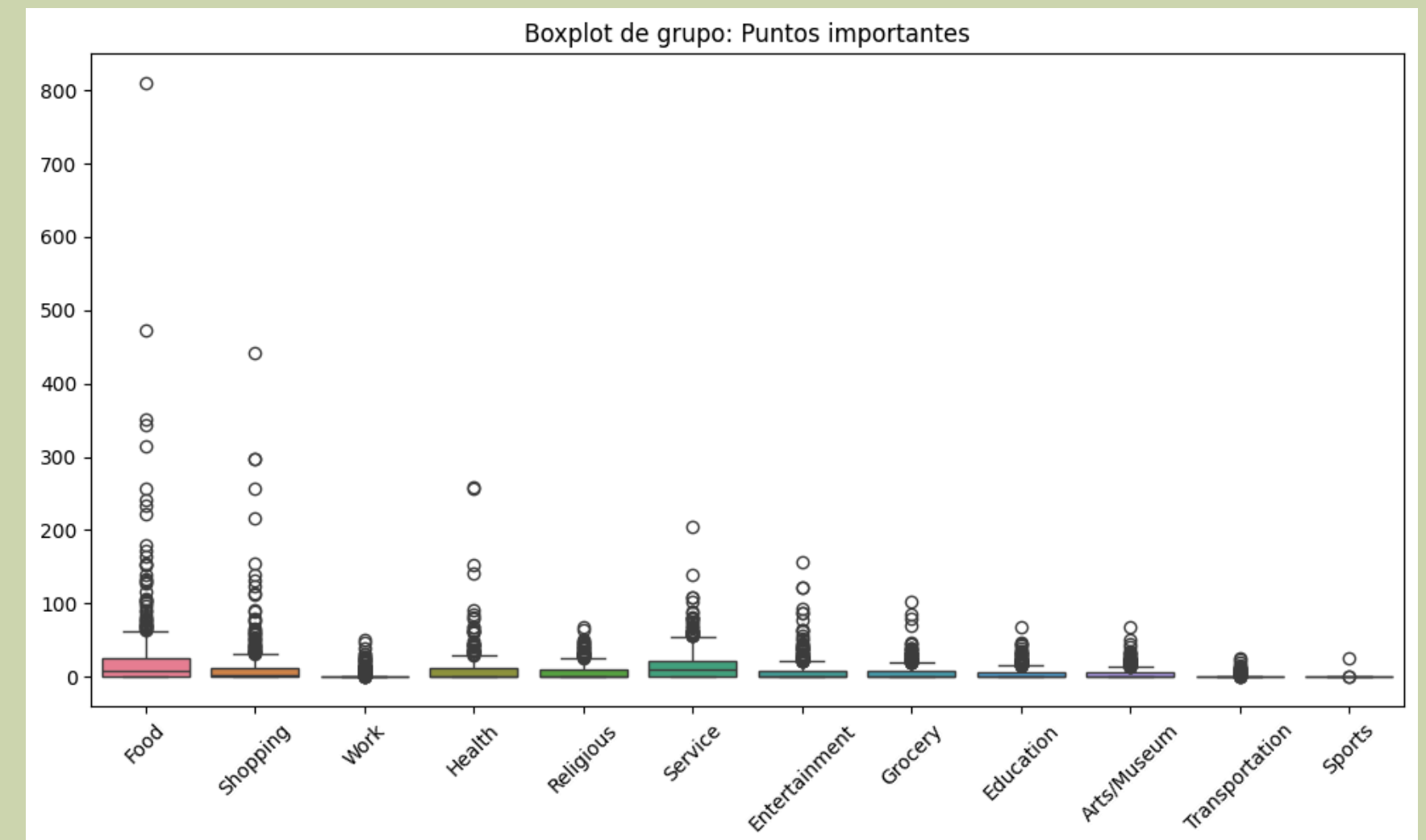
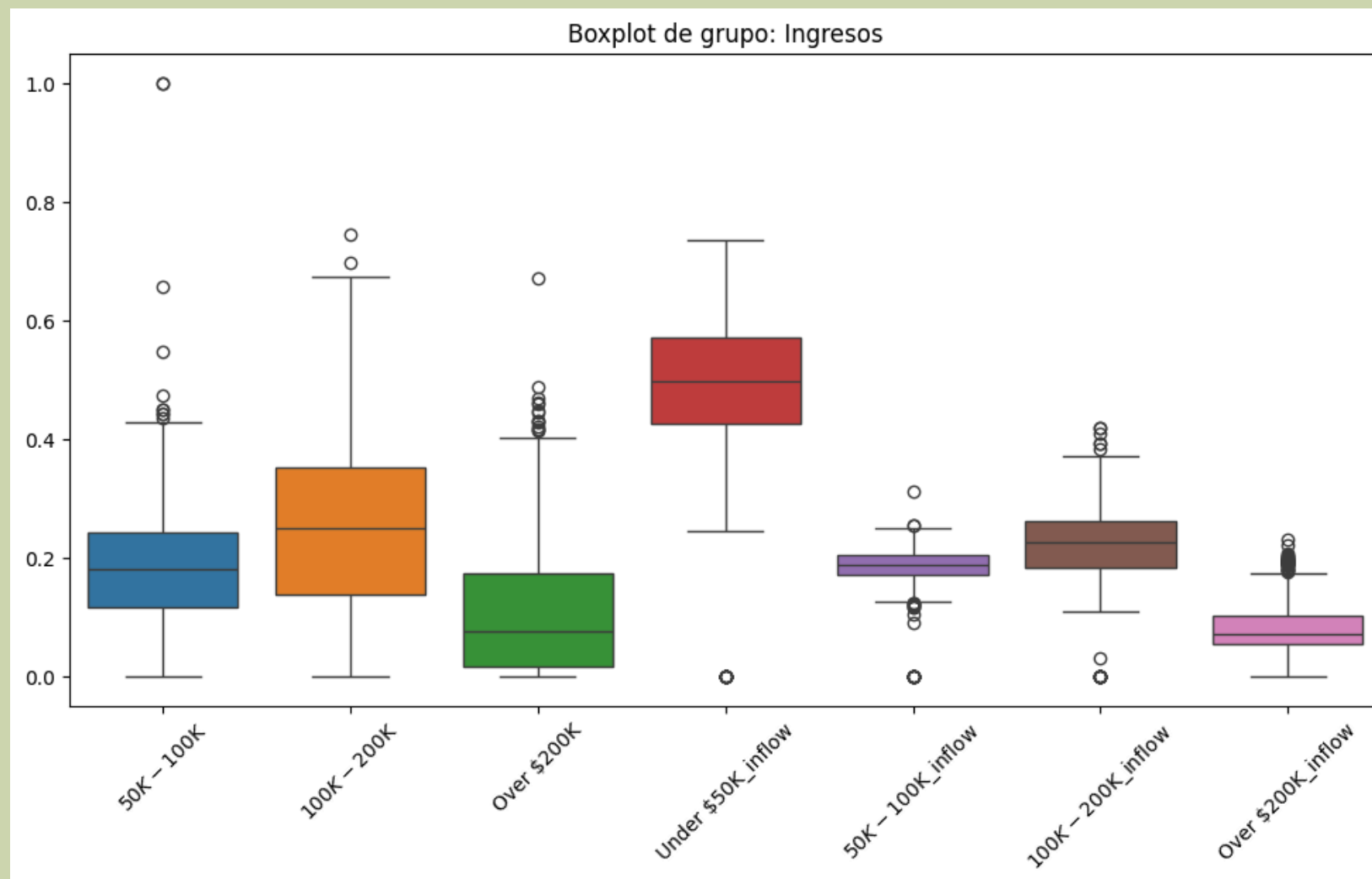
Se identificaron valores atípicos en los datos que representan características reales y relevantes, como zonas con concentraciones elevadas de ciertos grupos sociales o actividades específicas, más que errores o ruido.



2.- ¿ QUÉ DESCUBRIERON AL ANALIZAR LOS DATOS?



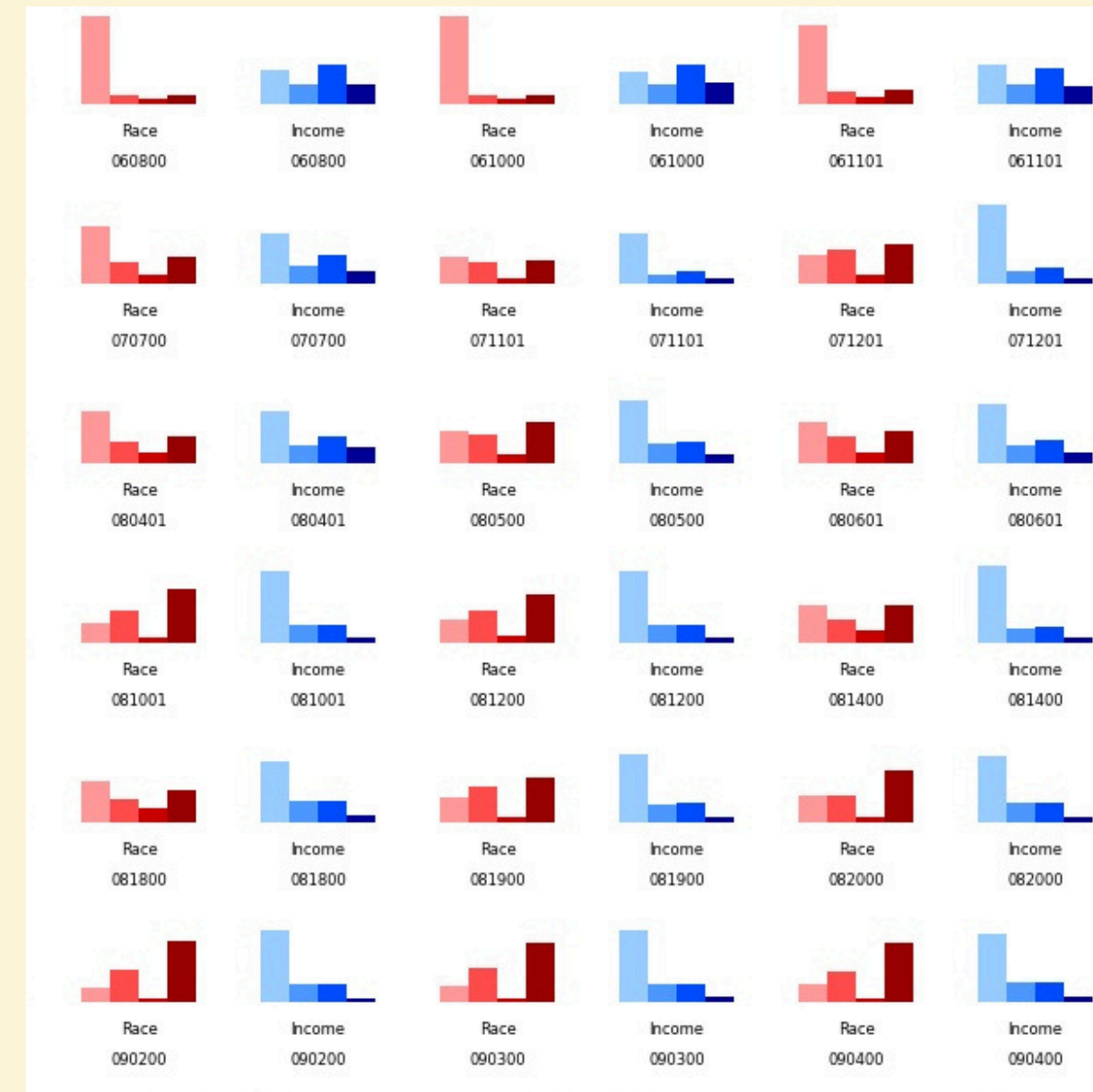
Se identificaron valores atípicos en los datos de movilidad, especialmente en columnas relacionadas con las concentraciones de población y actividades en ciertas zonas. Estos outliers no eran errores, sino representaciones válidas de fenómenos sociales y económicos, como grandes concentraciones de grupos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



3. ¿Que reflejan los patrones de tendencias?

Distribución racial desigual:

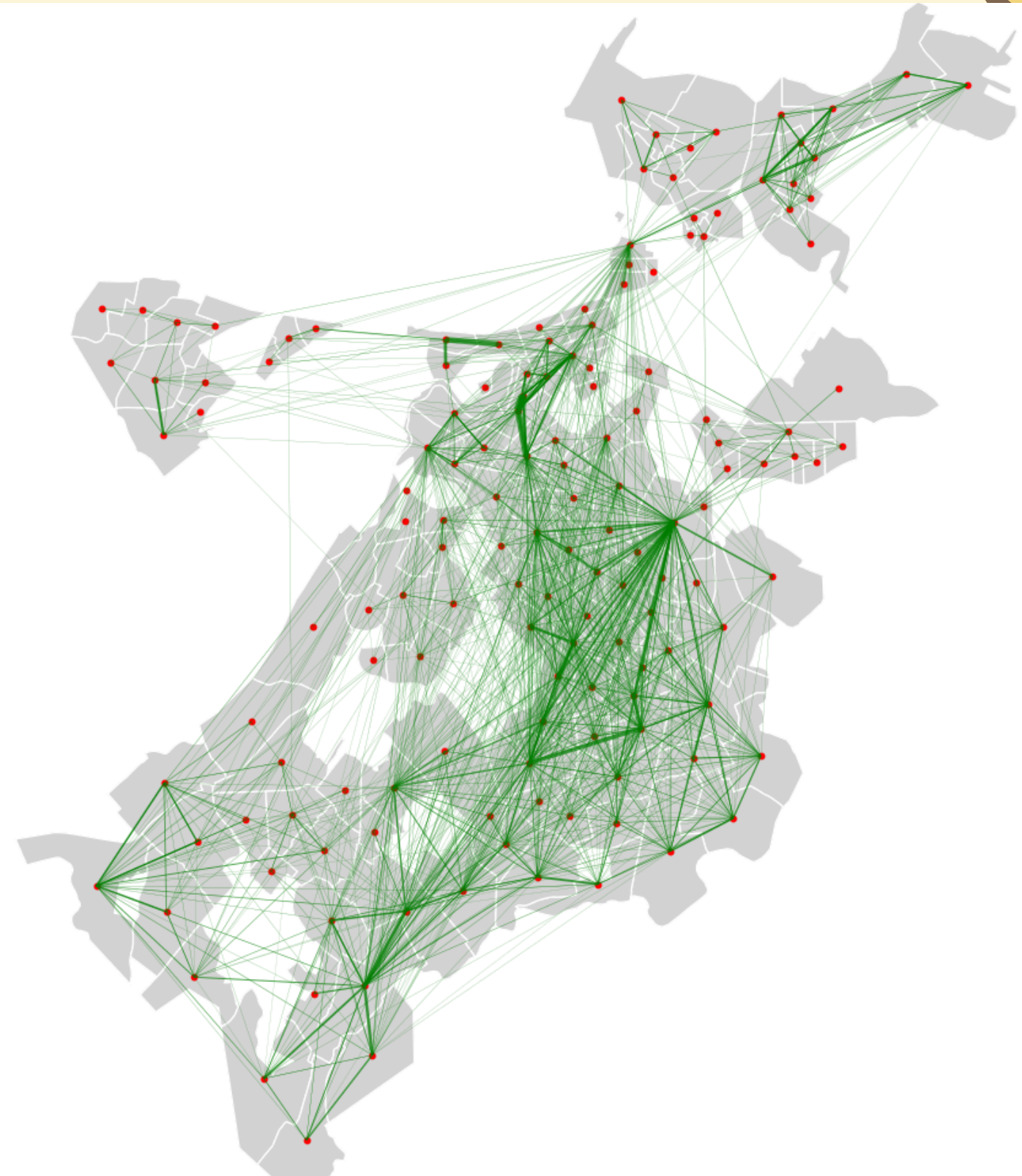
- Algunas áreas pueden tener una alta concentración de población blanca, mientras otras áreas presentan mayor proporción de población negra, asiática o hispana.
- Esto puede reflejar segregación residencial o patrones históricos de asentamiento.



3. ¿ Que reflejan los patrones de tendencias?

Conectividad y relaciones entre zonas:

Las áreas con mayores valores en visitor_flows muestran conexiones o trayectos más frecuentes entre zonas de origen y destino. Esto puede indicar rutas de transporte principales, zonas de trabajo o actividad económica que atraen personas.





4- ¿Cómo se manifiestan los niveles de segregación racial y socioeconómica en un área específica, y cómo se reflejan en la composición poblacional y en los flujos de movilidad hacia esa área?



Los índices de segregación muestran que en el área hay una concentración desigual de grupos raciales y niveles socioeconómicos. Tanto la composición poblacional como los flujos de movilidad hacia esa zona reflejan estos patrones segregados. Esto indica que la segregación afecta no solo dónde viven las personas, sino también cómo y hacia dónde se movilizan dentro de la ciudad.

Mapa CT: 25025092300



Índices de Segregación (SI)

