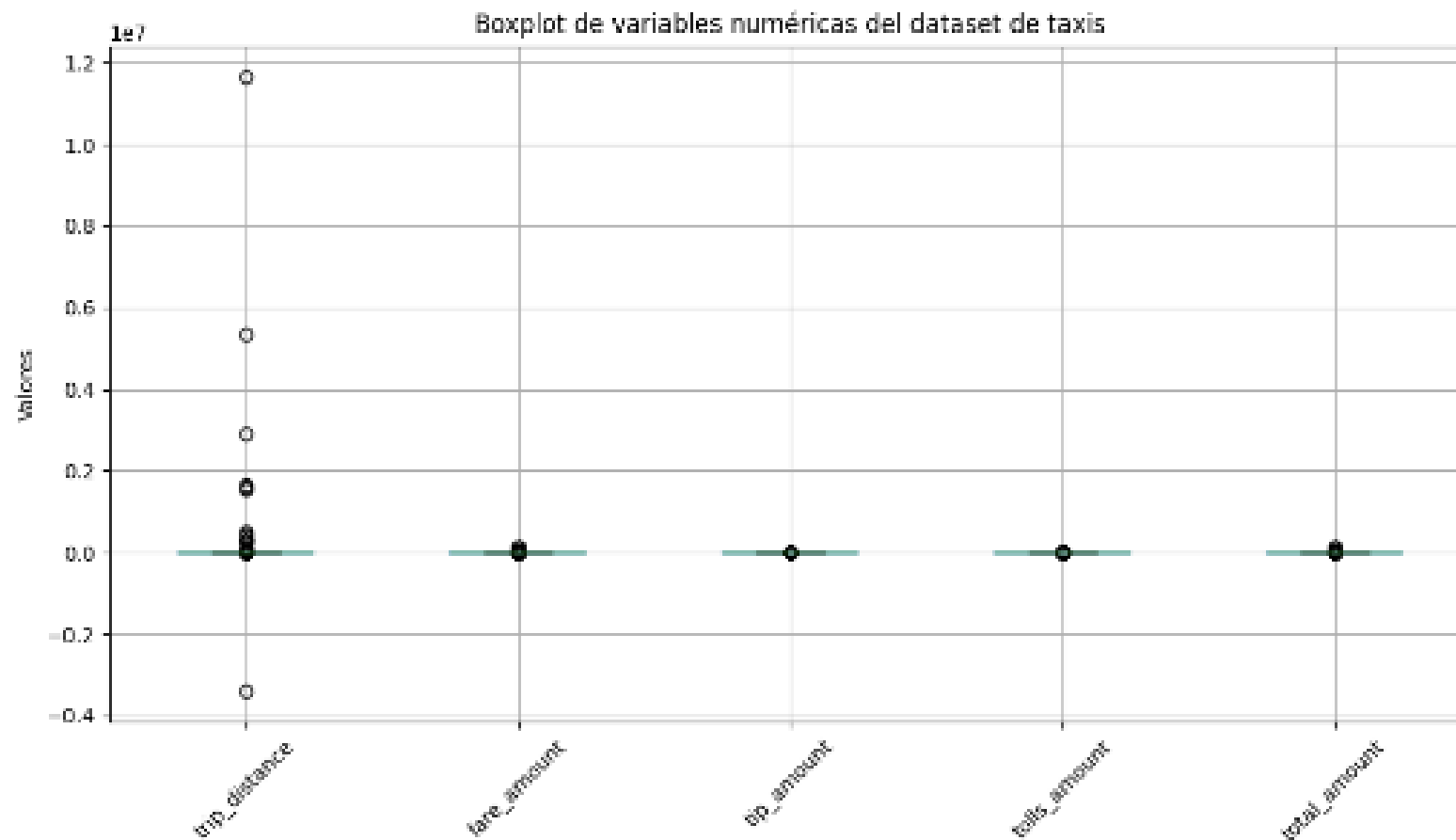


# Pipeline de Cs. de Datos

---

**Saul Condori Machaca**



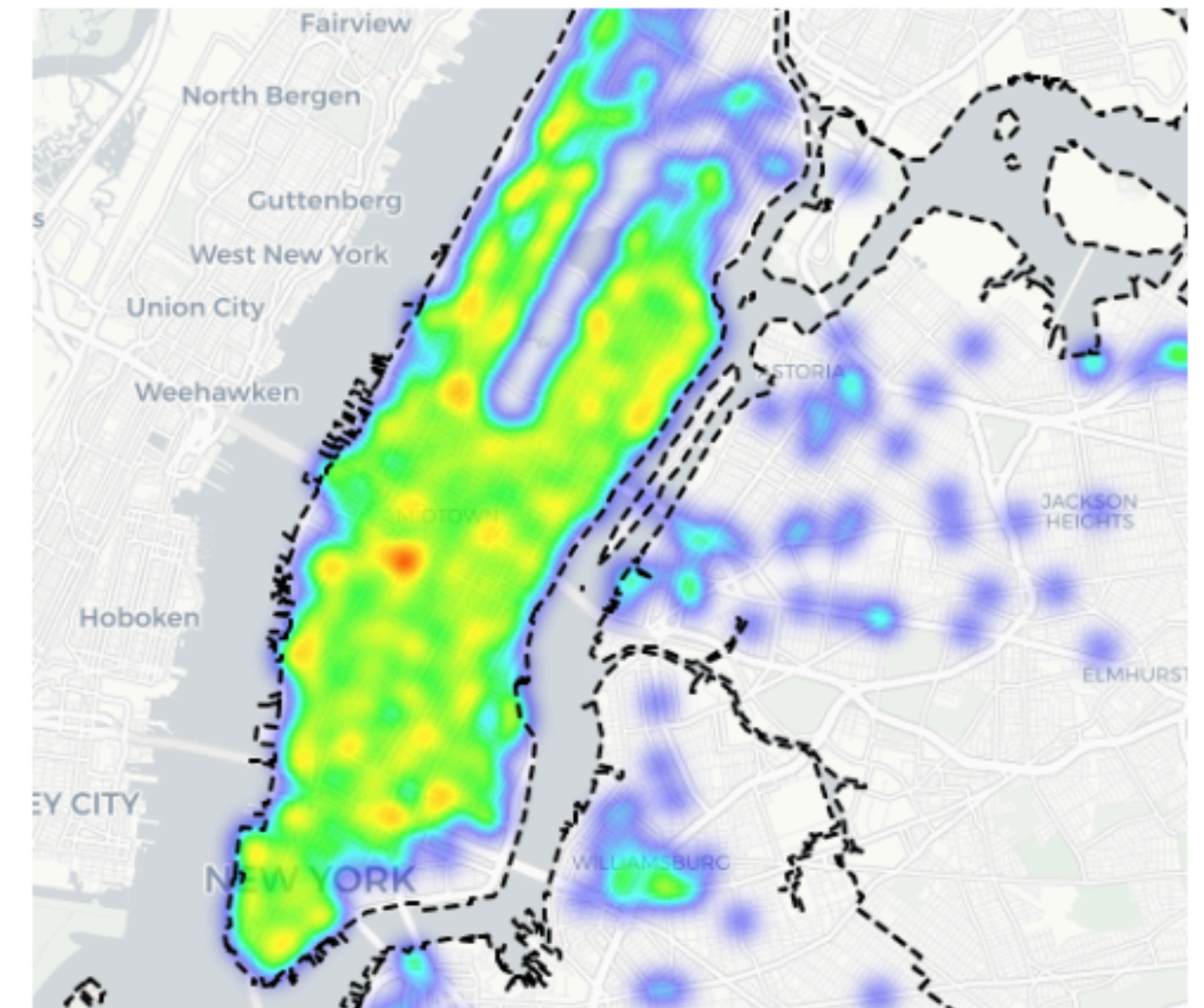
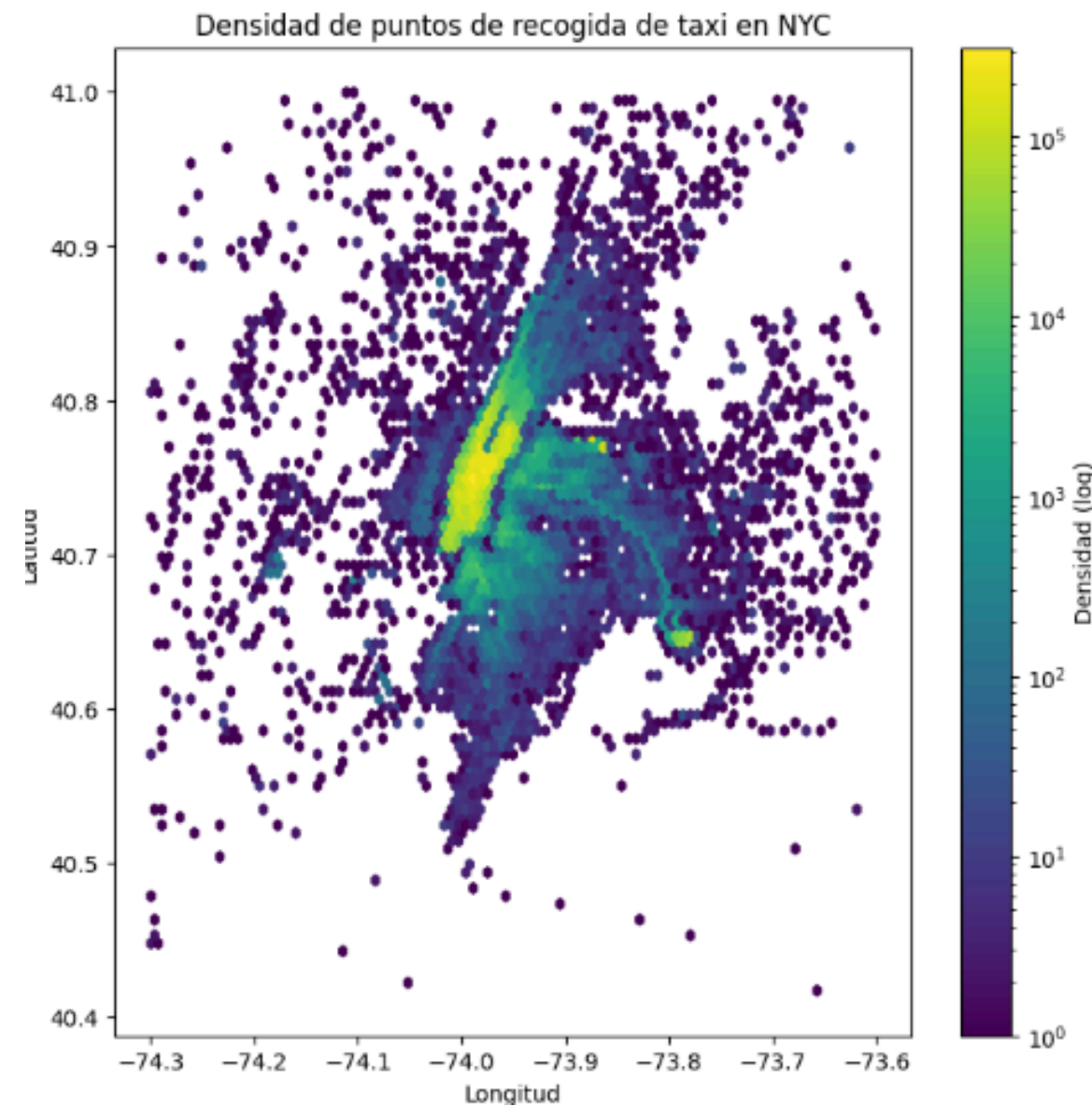
# Problemas en los Datos

- Presencia de valores atípicos extremos (distancias y montos negativos).
- Coordenadas geográficas fuera del rango real de NYC.
- Registros duplicados (taxis) y valores nulos (sismos).
- Importancia de la limpieza para garantizar coherencia y precisión.



# ¿Qué descubrimos al analizar los datos?

- Mayor número de viajes en franjas horarias de tarde y noche (actividad urbana intensa).
- Alta densidad de recogidas de taxi en zonas como Manhattan y Midtown.
- Concentración espacial de trayectorias permite identificar zonas de alto tránsito.



# ¿Qué reflejan los patrones de tendencia?

Comparación de actividad urbana vs. frecuencia sísmica

- La actividad de taxis crece durante el día: mañana → tarde → noche.
- La frecuencia de sismos descende: mayor en la mañana, menor en la noche.

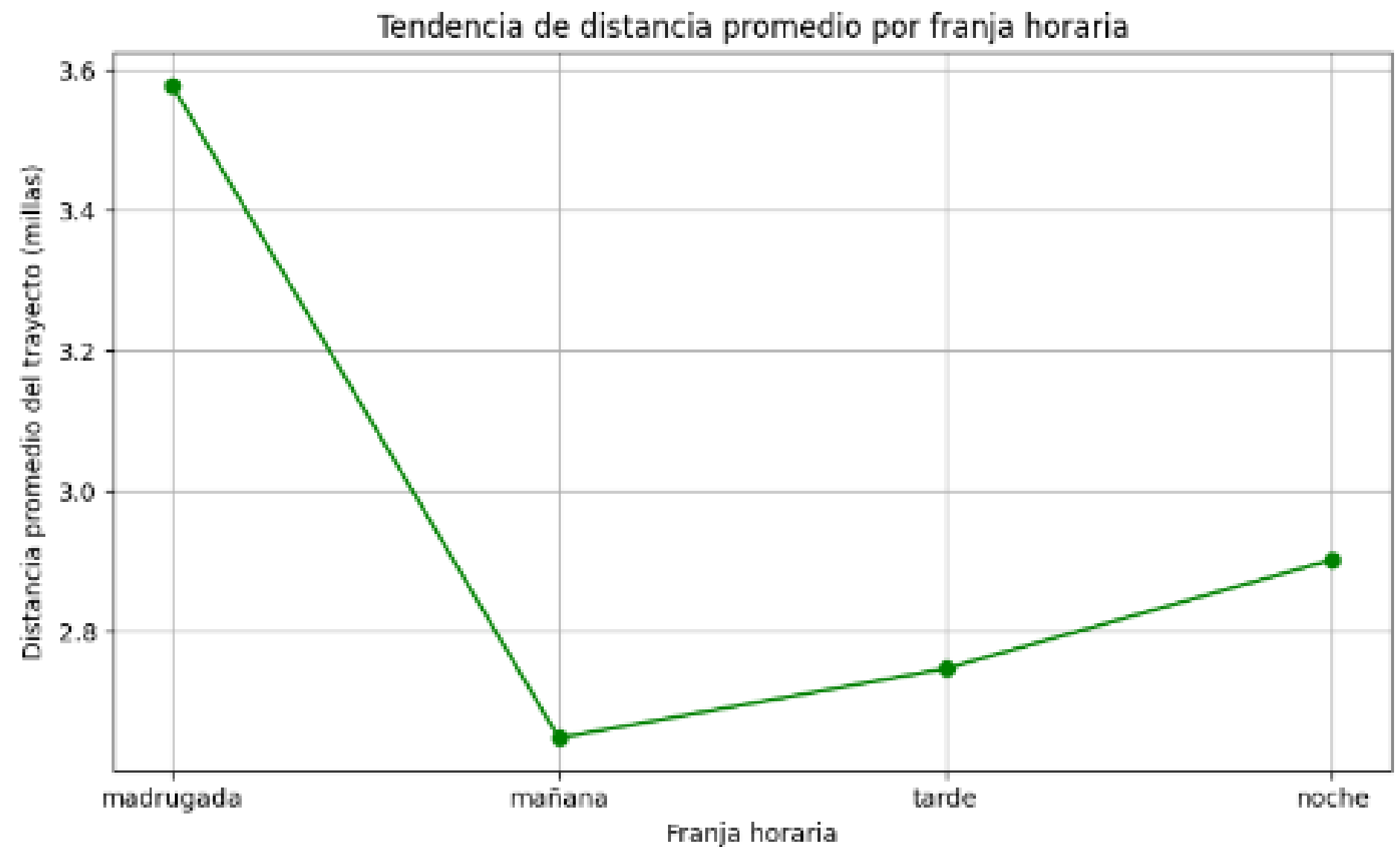




# ¿Qué reflejan los patrones de tendencia?

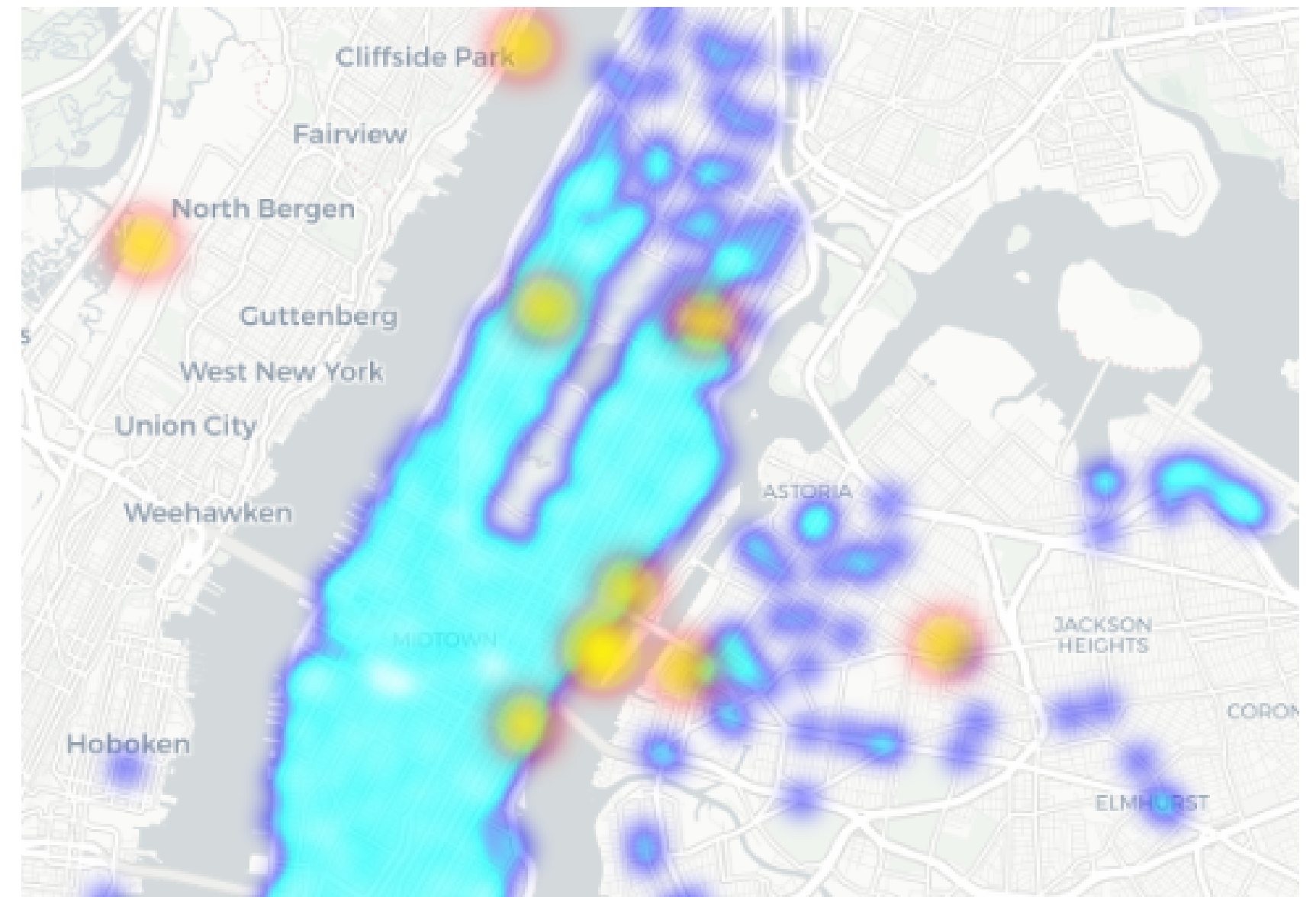
Distancia promedio de trayectos según franja horaria

- Trayectos más largos en la madrugada (desplazamientos periféricos).
- Menor distancia en la mañana (viajes laborales intraurbanos).
- Tendencia ayuda a anticipar rutas de evacuación más extensas o congestionadas.



# ¿Zonas de alta movilidad coinciden con zonas sísmicas?

- Se superpusieron trayectorias urbanas con epicentros sísmicos.
- Coincidencias visibles en Manhattan y sectores de Queens.
- Identificación de zonas críticas urbanas con doble exposición: tránsito + sismos.
- Relevancia para planificar evacuaciones inteligentes y focalizadas.



---

# GRACIAS

---