

EDA Travel Analysis Pandas

Federico Barbarroja | Tecnicatura Universitaria en Inteligencia Artificial - FCEIA - UNR

Descripción General

El dataset fue elegido desde los datos proporcionados por la ciudad de Buenos Aires abarcando todo el año 2018 y seleccionado por la variedad de datos que tiene.

El mismo contiene datos de una muestra de personas encuestadas en diferentes zonas del Gran Buenos Aires y los viajes que realizan en el día a día a lo largo de una año

Tipos de datos

Cuantitativos: duración, cantidad de etapas, códigos.

Categoricos: motivos del viaje, tipo de viaje, tipo de transporte.

Geográficos: partido de origen/destino, radios censales.

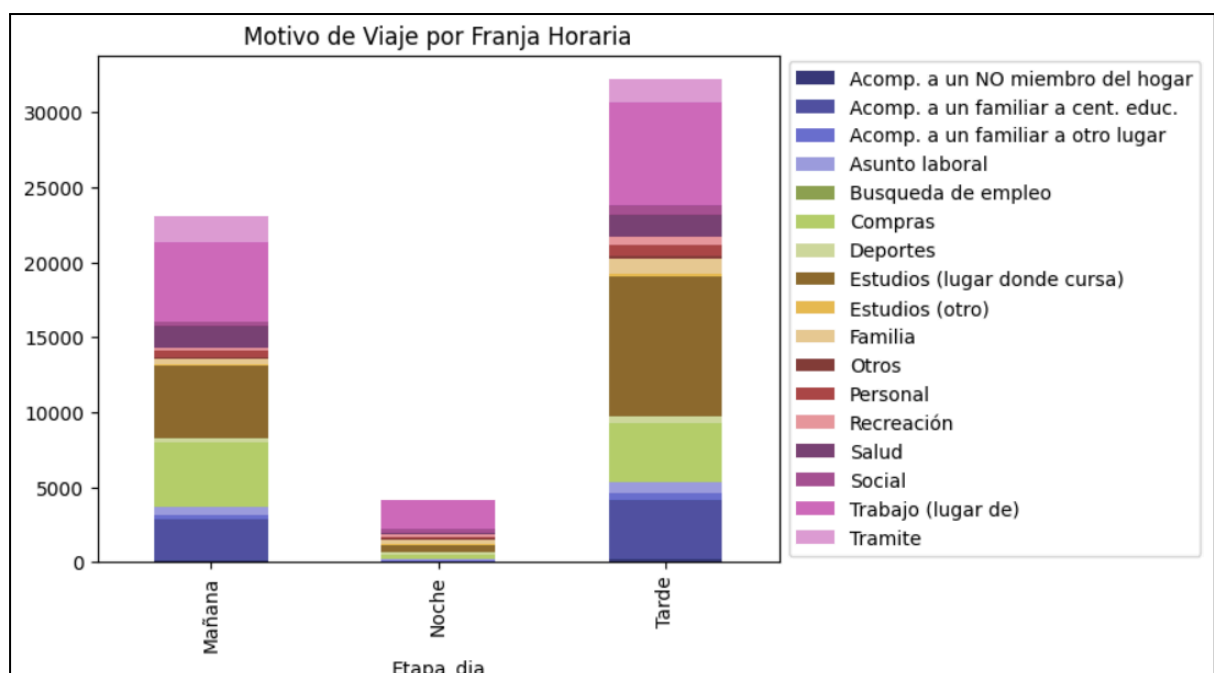
Horarios: hora inicio y final.

Análisis

Distribución horaria

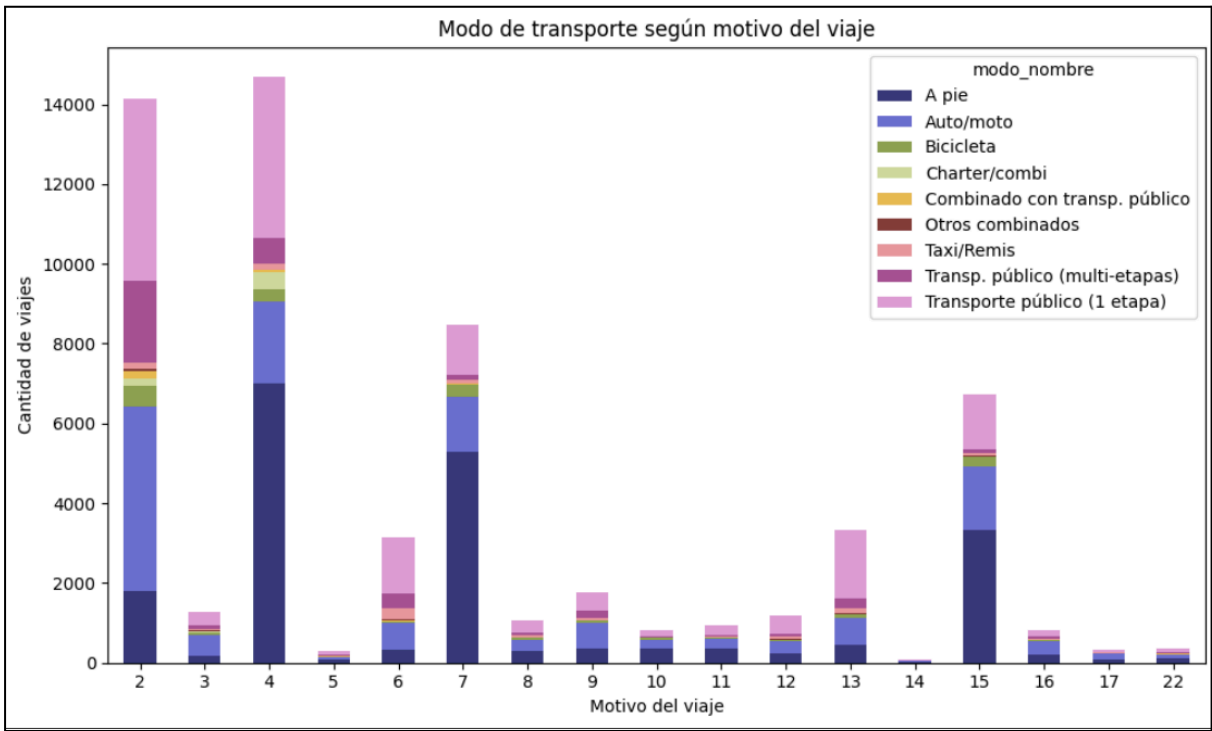
Hallazgos:

- Tanto en la mañana como en la tarde predominan los viajes a recintos educativos y laborales (coincide con horarios de entrada y de salida)
- Los viajes con motivos sociales o de recreación aumentan a la tarde
- Las compras prevalecen a la mañana y a la tarde, con la diferencia que a la mañana se realizan más trámites.



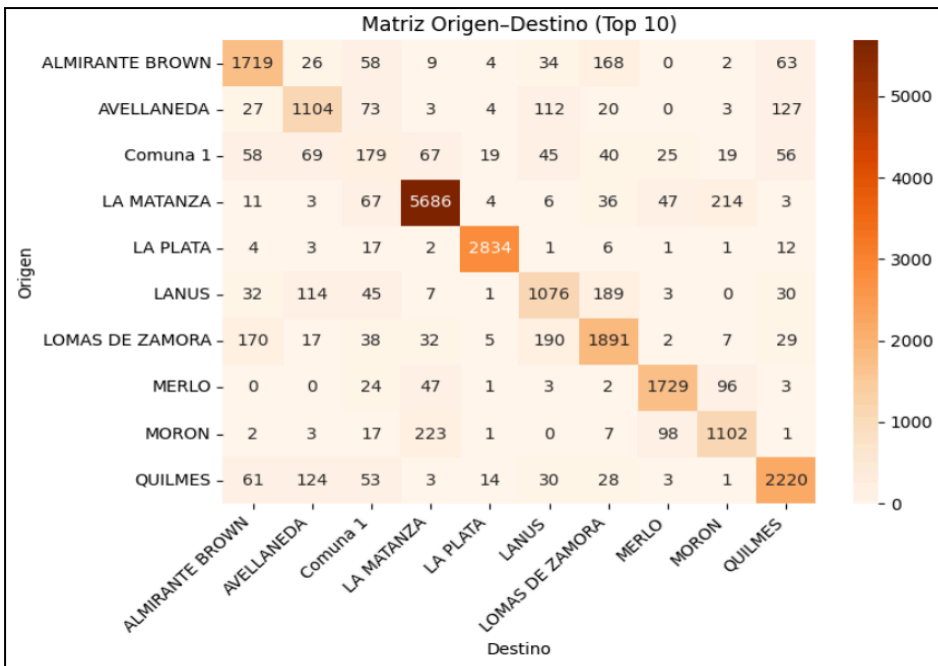
Medio de transporte

Se observa que para los viajes por motivos laborales o educativos predominan los modos que incluyen caminata, transporte público, automóvil o combinaciones de estos. En cambio, para motivos recreativos o compras, se usan más caminatas u otros tipos de transporte como bicicleta o taxi.



Origen / Destino

Se observa una notable paridad entre los gráficos, por lo que podemos suponer que la mayoría de los viajes se realizan internamente (dentro del mismo partido). **Vamos a investigar si en verdad es así** con una matriz de frecuencia cruzada.



Ahora podemos ver que, efectivamente, la mayor cantidad de viajes de los orígenes del top 10, son internos (desde un partido hacia el mismo).

Tiempo

Podemos observar que el 90% de los viajes (según el cuantil calculado) dura a lo sumo 1 hora, mientras que la mediana es de 20 minutos, un tiempo bastante inferior al de la media, por lo que esto nos indica una presencia de fuertes outliers que afectan el valor de la media.

Distancia

Podemos observar que como ya vimos en el apartado del tiempo, en la distancia también tiene una fuerte presencia de outliers, haciendo que su media (5,23 km) sea más del doble que su mediana (2,08 km). También vemos que el 70% de los viajes tiene una distancia como máximo de 4,5 k)

Conclusiones

A lo largo del trabajo vimos los diferentes comportamientos de las variables que nos resultaban de mayor interés, y pudimos obtener algunas conclusiones:

- Distribución horaria: La mayoría de los viajes se hacían durante la franja de la tarde, seguida por la mañana. Esto nos indica posibles picos en los horarios de ingreso/egreso a los trabajos y recintos educativos.
- Medio de transporte: Se observa una fuerte relación entre el motivo del viaje y el medio de transporte utilizado. Por ejemplo para fines laborales y educativos predominan transportes públicos, caminatas y auto/moto. Mientras que para fines recreativos es más común taxis y caminatas.
- Movilidad entre partidos: A través de la matriz Origen–Destino vimos que ciertos partidos actúan como grandes emisores y receptores de viajes. En el análisis pudimos ver que la mayoría de los viajes son internos (dentro del mismo partido), tomando como ejemplo que los 10 partidos de mayor origen, son los mismo que los 10 partidos de mayor destino.
- Tiempo: La mayoría de los viajes tiene una duración inferior a 21 minutos. Sin embargo, existen valores extremos que podrían responder a trayectos interurbanos o viajes con múltiples etapas.
- Distancia. La mayoría de los viajes tiene una distancia menor a 4,5 km. Sin embargo, también hay valores atípicos que indican un trayecto interpartidario o de múltiples etapas.