



Tecnológico de Monterrey

Campus Santa Fe

Desarrollo de proyectos de análisis de datos (Gpo 307)

ETAPA 3 – Análisis Exploratorio de Datos (EDA)

Nombre:

German Gedovius Dalloz - A01782383

Samuel Aranzola Monroy - A01786577

ETAPA 3 – Análisis Exploratorio de Datos (EDA)

Proyecto: Factores que influyen en el precio y desempeño de alojamientos de Airbnb

Después de generar un dataset limpio en la Etapa 2, en esta fase realizamos un análisis exploratorio de datos (EDA) con el objetivo de entender cómo se distribuyen las variables principales y qué relaciones existen entre los factores que podrían influir en el precio, ocupación e ingresos de los alojamientos. El análisis se realizó con el archivo `listings_clean.csv`.

Primero se describieron las variables numéricas centrales, como `price`, `price_per_guest`, `accommodates`, `availability_365`, `estimated_occupancy_I365d` y `estimated_revenue_I365d`. El precio mostró una distribución asimétrica aun después de filtrar outliers, lo cual es típico en mercados donde existen alojamientos premium que elevan la cola superior. La capacidad `accommodates` tuvo valores principalmente entre 1 y 6, indicando que la mayoría de propiedades están diseñadas para grupos pequeños.

En las variables categóricas, `room_type` mostró que la oferta está dominada por “Entire home/apt”, seguida de “Private room”. Además, `neighbourhood_cleansed` permitió identificar los barrios con mayor concentración de alojamientos y comparar sus precios promedio, revelando diferencias geográficas importantes.

Posteriormente se estudiaron relaciones entre variables. La relación entre capacidad y precio fue positiva: a mayor número de huéspedes, el precio tiende a aumentar, aunque con variabilidad en capacidades altas. El análisis por tipo de habitación confirmó que las propiedades completas tienen precios medianos mucho más altos que las privadas o compartidas. La correlación entre precio e ingresos estimados mostró una relación positiva, aunque no perfecta, indicando que la ocupación también influye en los ingresos.

También revisamos características del anfitrión. `Host_is_superhost` mostró un precio promedio ligeramente mayor frente a anfitriones regulares, lo cual sugiere que la reputación podría estar asociada a la percepción de calidad.

Finalmente, se construyó un subset de columnas clave (`listings_clean_core_eda.csv`) para facilitar análisis posteriores. El EDA permitió concluir que el precio depende principalmente del tipo de alojamiento, la capacidad y la ubicación; y que características del anfitrión y la ocupación también influyen. Estos hallazgos guiarán las siguientes etapas del proyecto.