TALLER 2

Proceso ETL con datasets de Airbnb Ciudad de México

26 DE OCTUBRE DE 2025
YENIFER GONZALEZ QUIRAMA
EMILIANO VÉLEZ SUÁREZ

Introducción

El presente informe describe el desarrollo de un proceso **ETL (Extracción, Transformación y Carga)** aplicado sobre los datasets de **Airbnb Ciudad de México**, con el propósito de construir una base de datos limpia y estructurada que sirva como base para análisis y toma de decisiones.

El proyecto se desarrolló en **Python**, empleando librerías como **pandas**, **sqlite3**, **logging** y **openpyxl**, siguiendo las tres fases fundamentales del proceso ETL: **Extracción** de datos desde MongoDB, **Transformación** mediante limpieza y enriquecimiento, y **Carga** en SQLite y Excel.

Descripción del dataset

El dataset proviene de **Airbnb Ciudad de México** y contiene tres colecciones principales:

- **listings:** Información de los alojamientos (anfitrión, tipo de habitación, precio, ubicación).
- **calendar:** Registros diarios de disponibilidad y precio.
- **reviews:** Opiniones y valoraciones de los huéspedes.

Los datos fueron exportados desde MongoDB a formato **CSV** para su procesamiento.

Análisis exploratorio

El análisis exploratorio se realizó utilizando **Python y Jupyter Notebook**, con el objetivo de identificar patrones generales, posibles inconsistencias y variables relevantes para el proceso ETL.

Durante esta fase se inspeccionaron los tres datasets principales (listings, calendar, reviews), revisando su estructura, cantidad de registros y tipos de datos en una primera instancia. Buscando posteriormente la existencia de valores nulos, duplicados y atípicos. Para lo cual, se encontraron los siguientes resultados:

- Se encontraron valores nulos en todos los datasets, pero los comportamientos fueron diferentes en cada uno:
    - En listings una gran parte de las columnas presenta estos valores, pero ninguno pasaba del umbral del 50%, por lo que no era

necesario eliminarlas. Por el contrario, se requería hacer una imputación de las mismas.
    - En calendar las columnas relacionadas con las noches tienen muy pocos nulos, hasta el punto que es casi indistinguible. En este caso, también es apropiado imputarlos de alguna forma. Adicionalmente, se encuentra la columna de adjusted_price, la cual sí debe ser eliminada. Pues está prácticamente vacía.
    - En reviews pasa algo relativamente similar a calendar: solo tiene unos pocos nulos en la columna de reviewer_name, por lo que se considera la imputación de estos.
- No se encontraron valores duplicados en ningún dataset.
- Se evaluaron los principales campos numéricos de listings y calendar (reviews no presenta este tipo de dato) para responder las preguntas de negocio (temas de ocupación, ventas y calificaciones), dando como resultado que la gran mayoría de columnas sí presentan valores atípicos, de acuerdo a lo obtenido en las gráficas. Sin embargo, no todos los casos son dignos de ser eliminados o corregidos, pues estos tienen sentido dentro del negocio.

Finalmente, se hicieron algunas transformaciones y se comprobó la necesidad de otras:

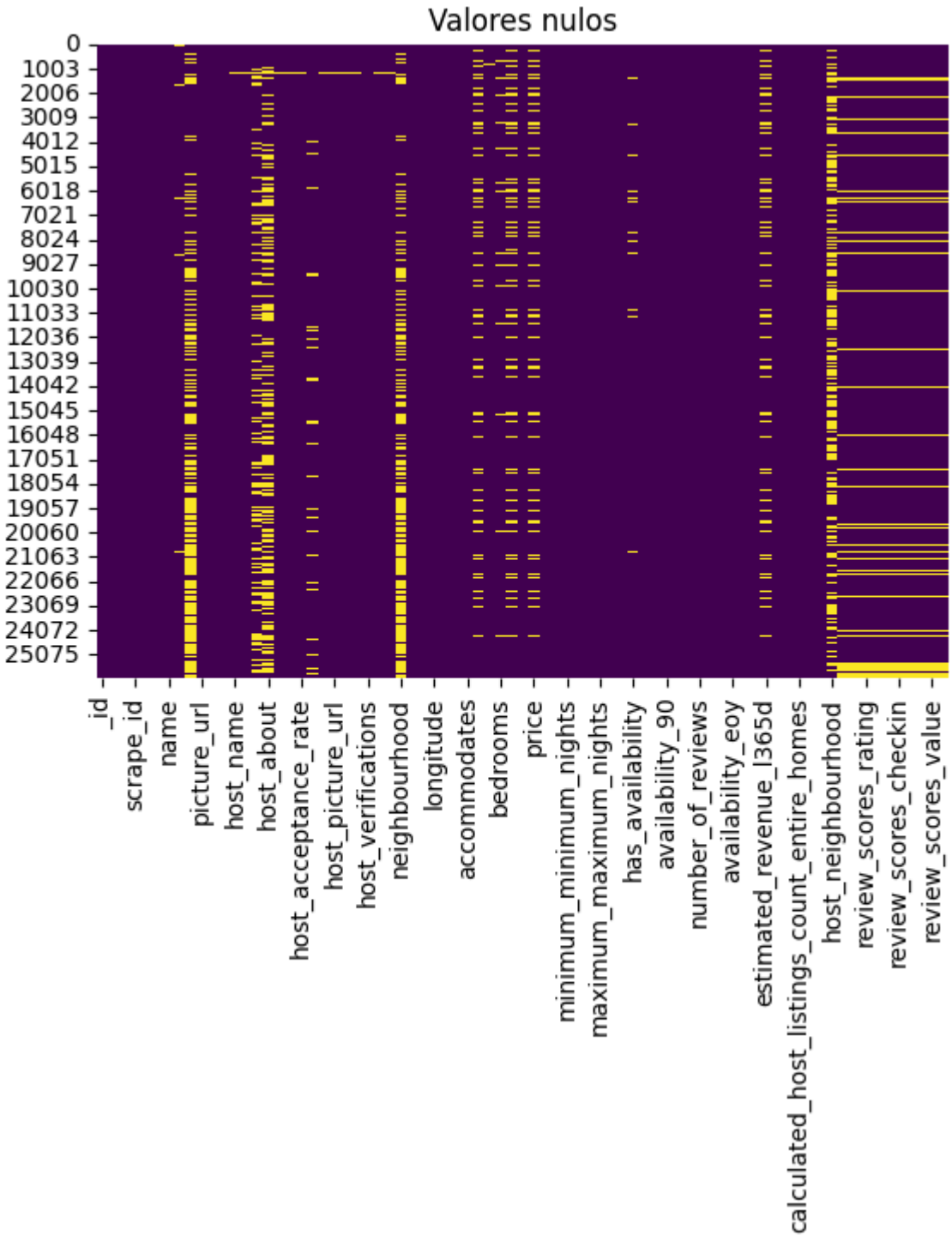
- Estandarización de datos: se estandarizaron las fechas al formato YYYY-MM-DD, los textos a un formato más plano (eliminando el html simple, eliminando espacios y recortando extremos) los precios a valores numéricos, eliminando símbolos como $ y ,; y eliminando valores con porcentaje para pasarlos a numérico.
- Se revisó si había necesidad de desanidar campos, lo que resultó siendo afirmativo. Teniendo las columnas amenities y host_verifications de la colección listings.
- Se revisó si había necesidad de pivotar o agrupar datos y en general no fue el caso. Solo se reconoció que sí hay algunos campos que deberían agruparse para cumplir con las preguntas de negocio, siendo estos los relacionados con la ocupación, los ingresos y las calificaciones.

Gráficas y hallazgos importantes

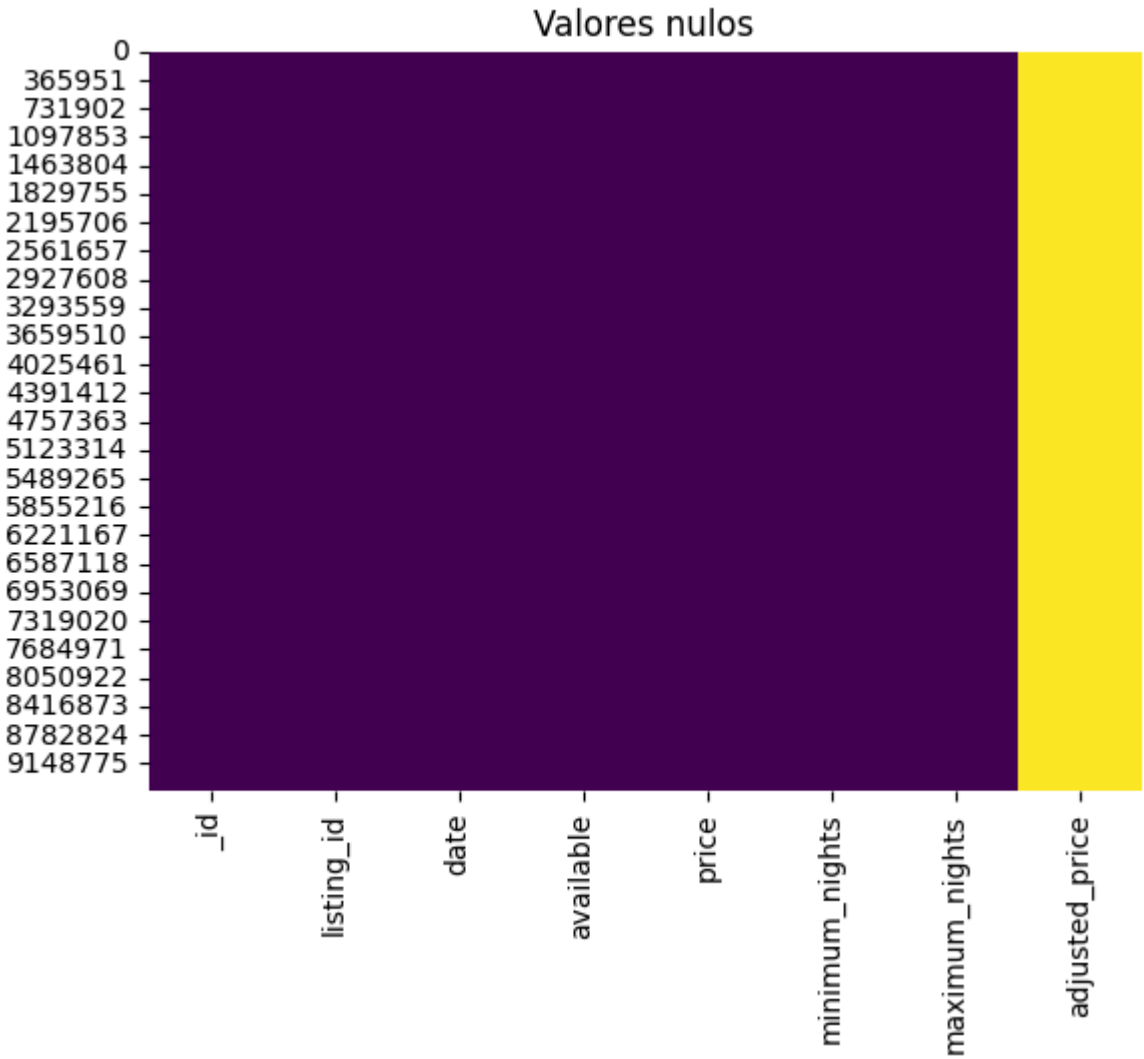
Gráficas:

- Valores nulos:

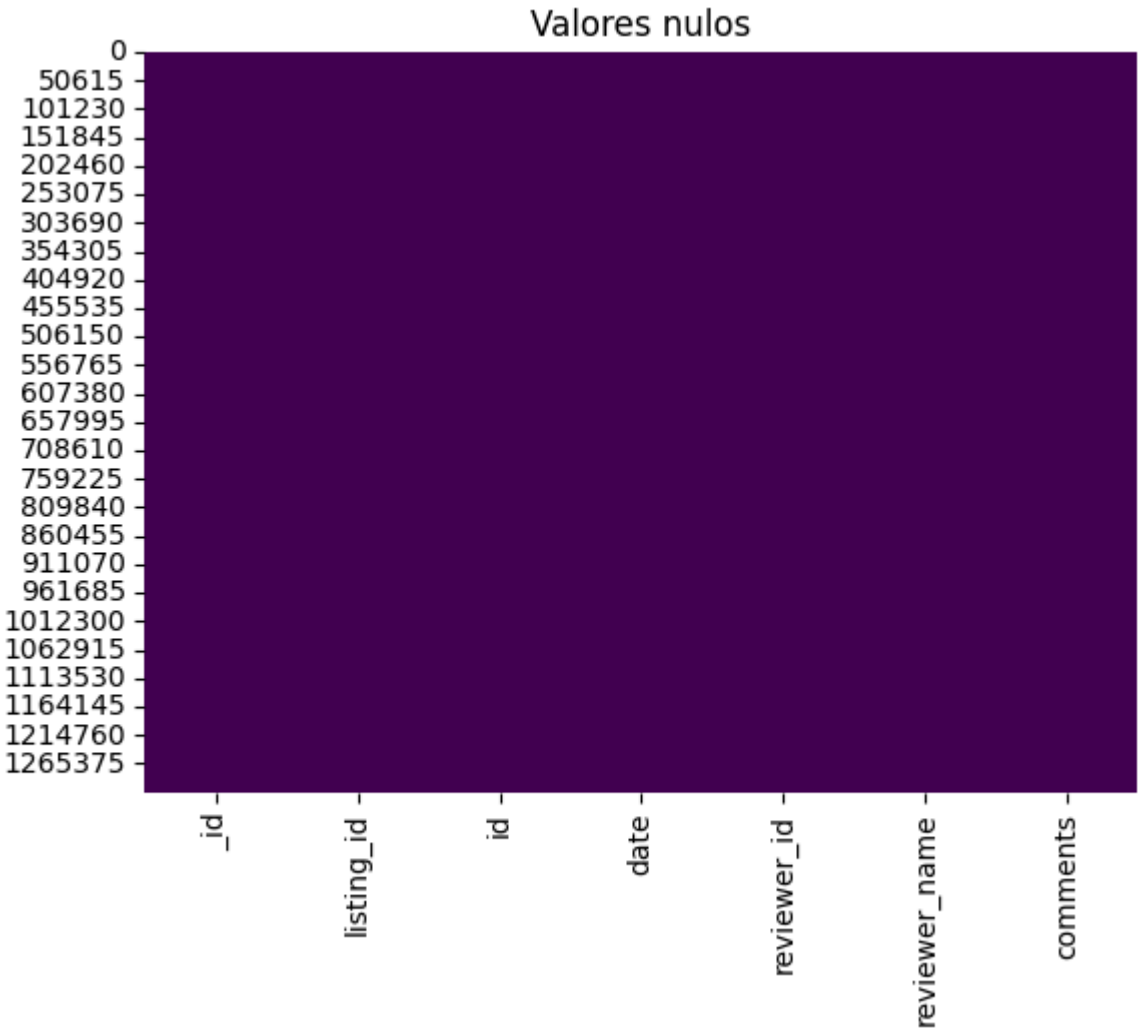
- Listings:


Valores nulos
0
1003
2006
3009
4012
5015
6018
7021
8024
9027
10030
11033
12036
13039
14042
15045
16048
17051
18054
19057
20060
21063
22066
23069
24072
25075
id
scrape_id
name
picture_url
host_name
host_about
host_acceptance_rate
host_picture_url
host_verifications
neighbourhood
longitude
accommodates
bedrooms
price
minimum_minimum_nights
maximum_maximum_nights
has_availability
availability_90
number_of_reviews
availability_eoy
estimated_revenue_l365d
calculated_host_listings_count_entire_homes
host_neighbourhood
review_scores_rating
review_scores_checkin
review_scores_value

- Calendar:

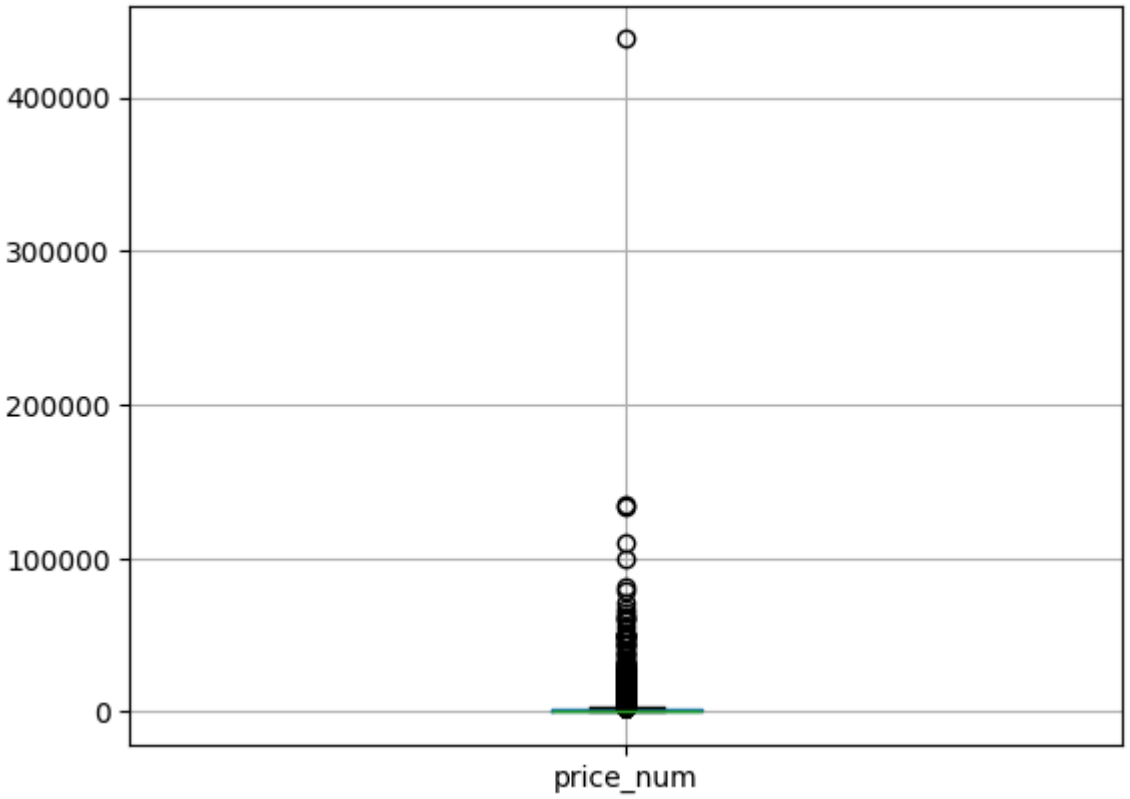

Valores nulos
0
365951
731902
1097853
1463804
1829755
2195706
2561657
2927608
3293559
3659510
4025461
4391412
4757363
5123314
5489265
5855216
6221167
6587118
6953069
7319020
7684971
8050922
8416873
8782824
9148775
_id
listing_id
date
available
price
minimum_nights
maximum_nights
adjusted_price

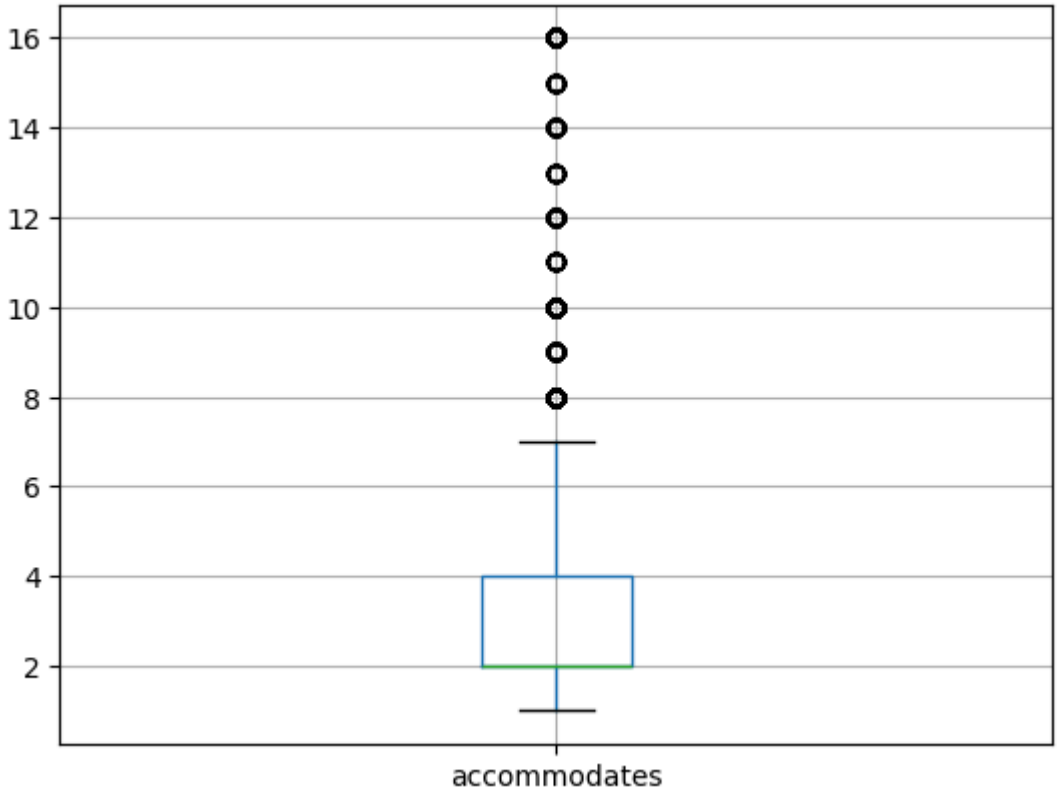
- Reviews:

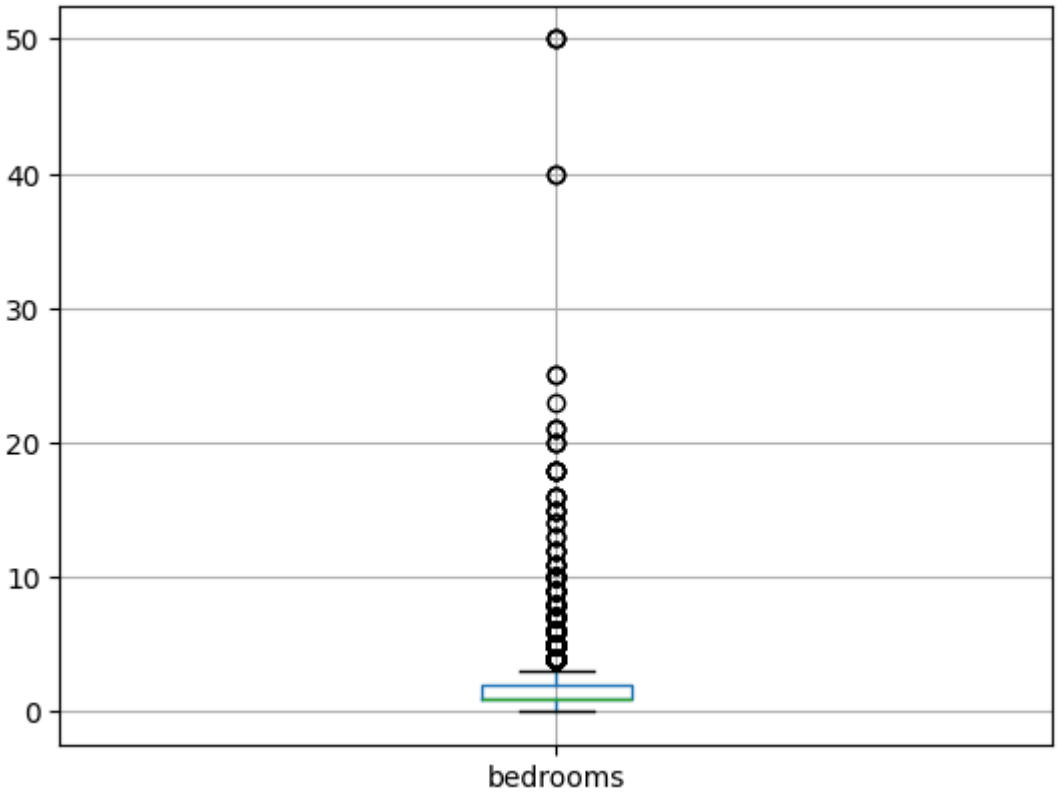


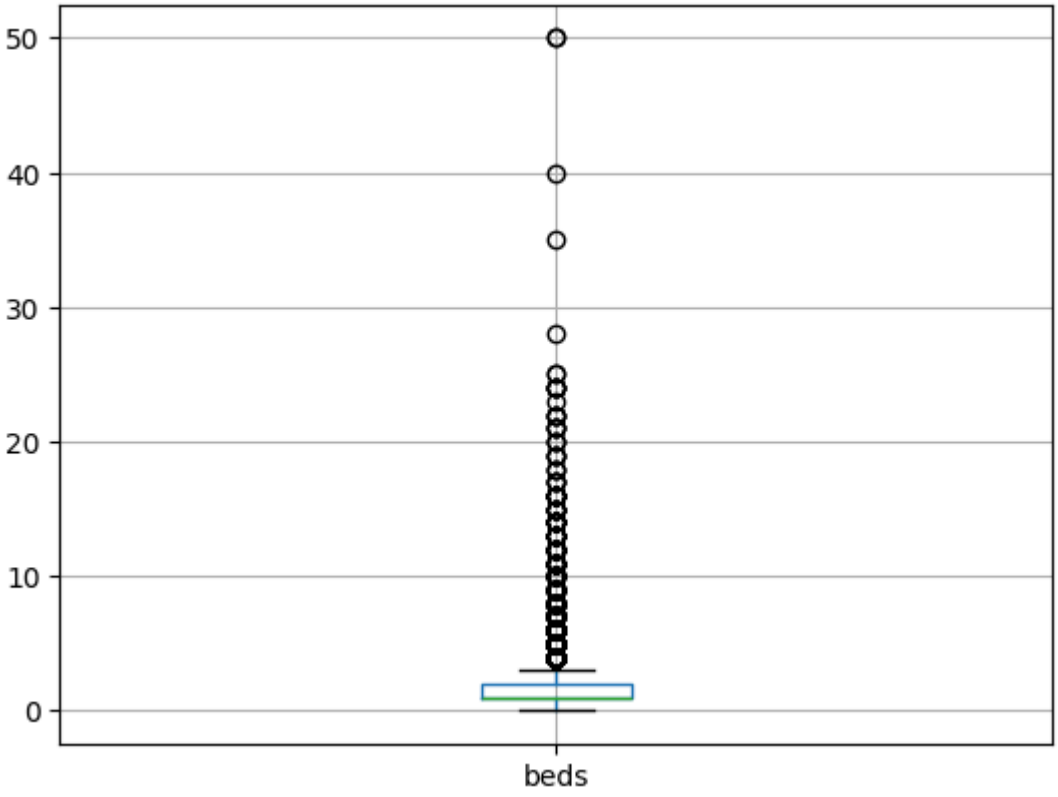
- Valores atípicos:

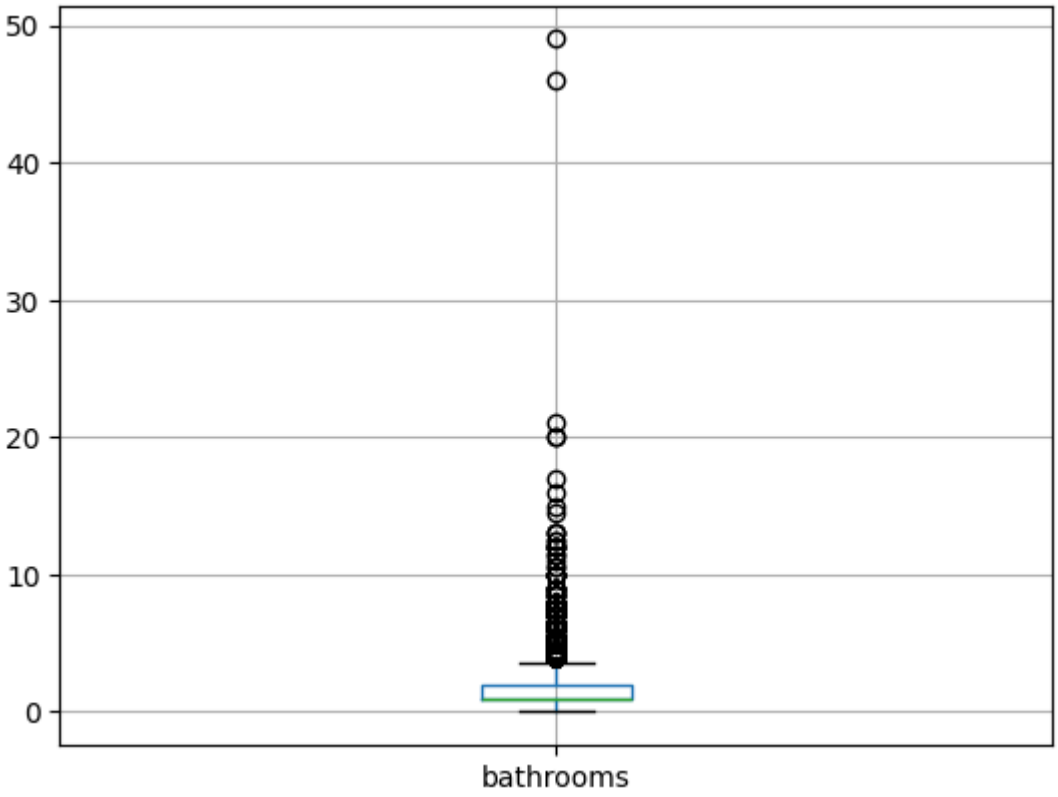
- Listings:

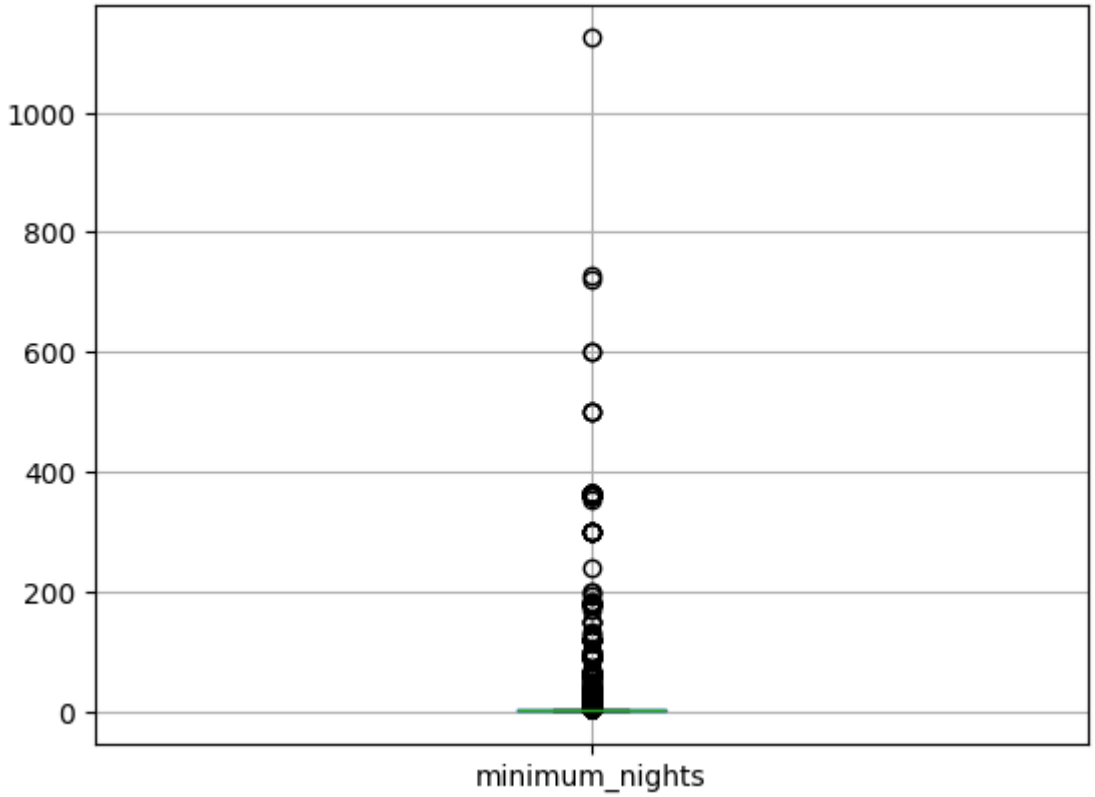

400000
300000
200000
100000
0
price_num

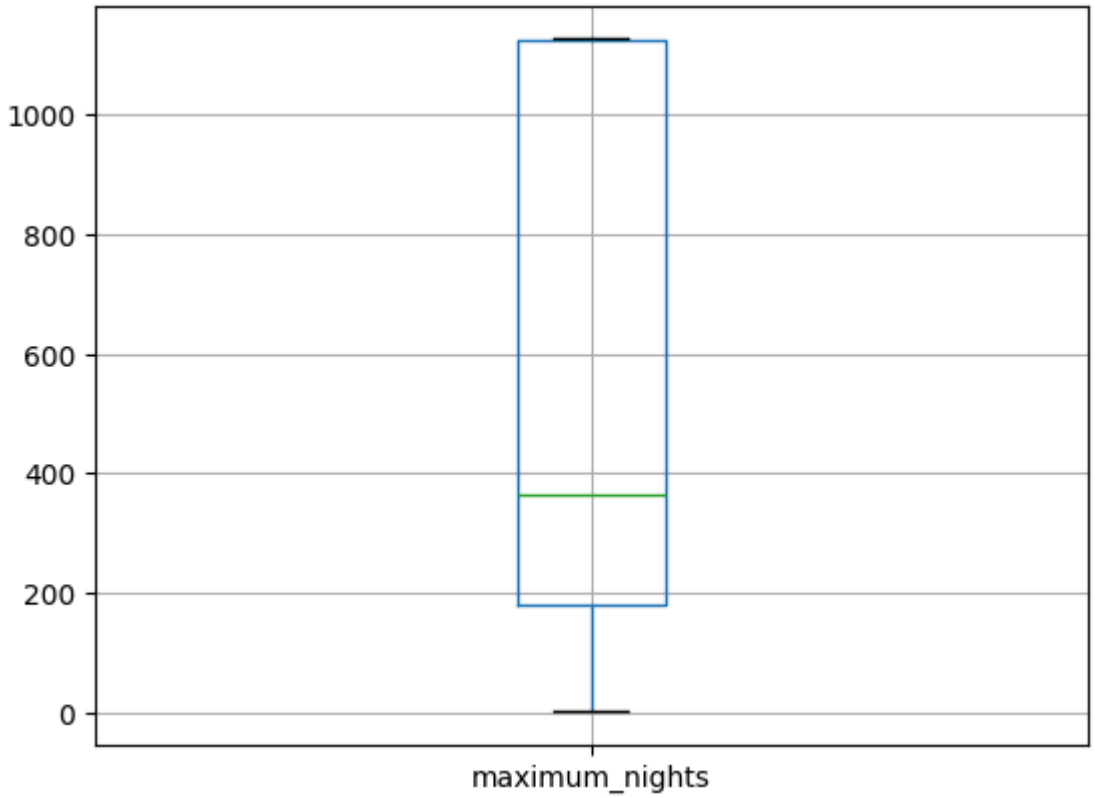

16
14
12
10
8
6
4
2
accommodates

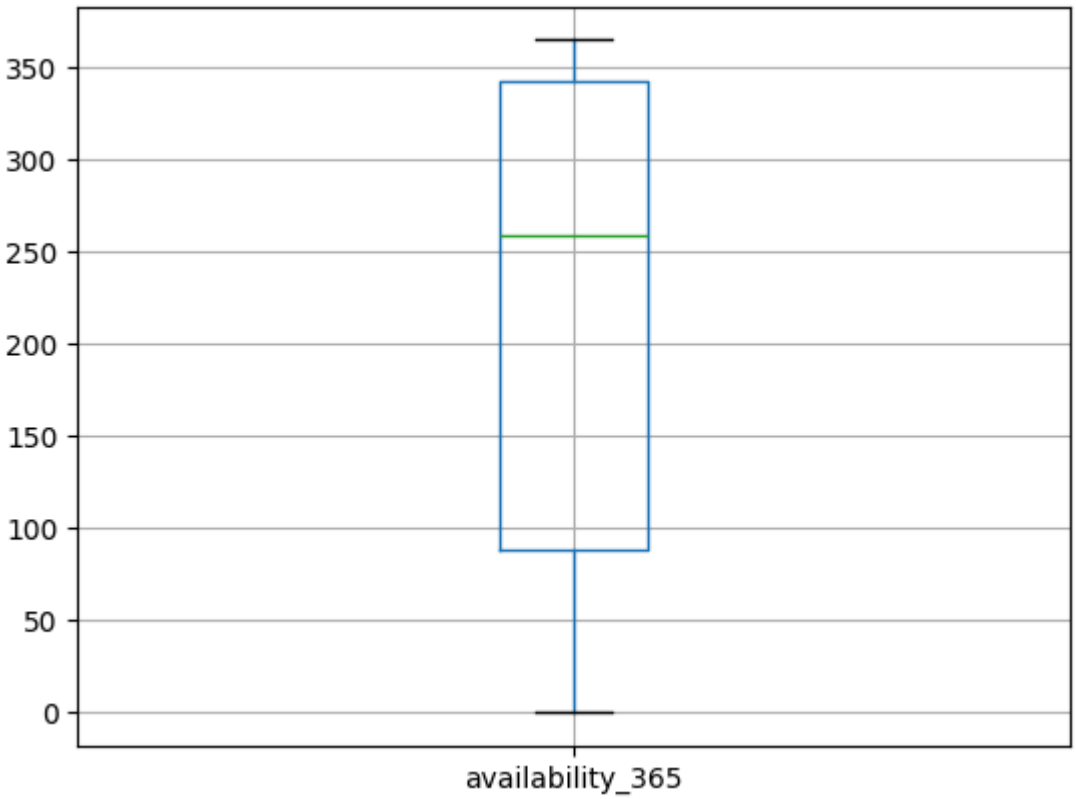

50
40
30
20
10
0
bedrooms

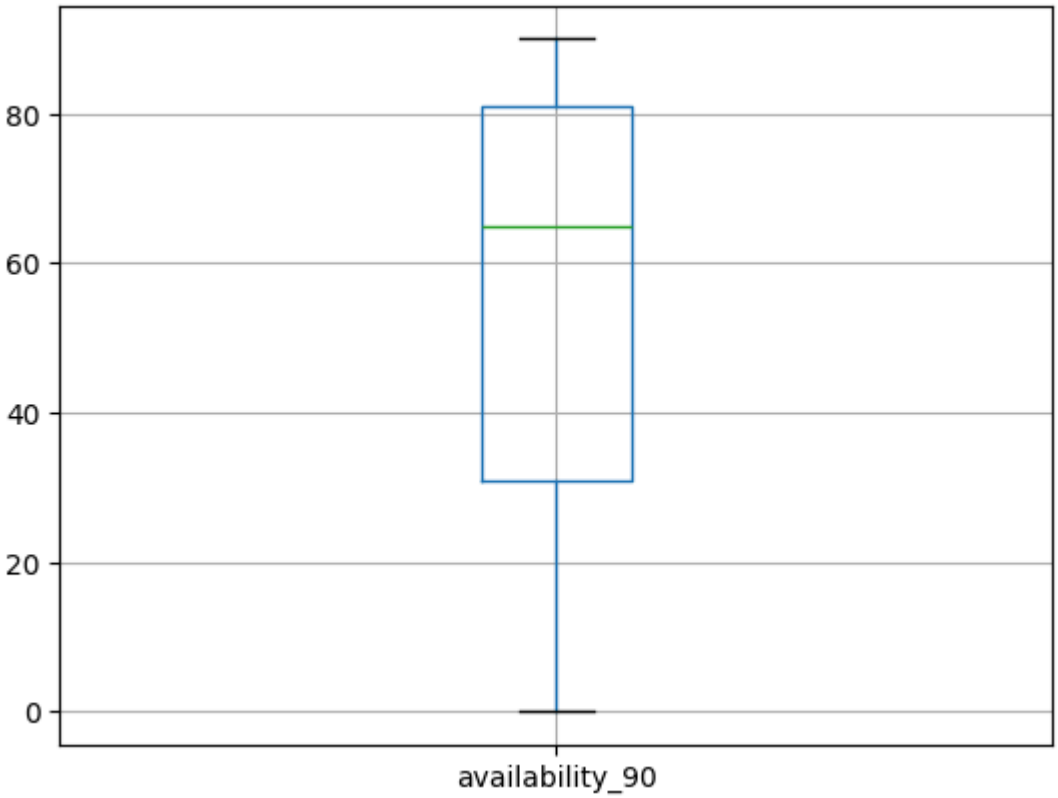

50
40
30
20
10
0
beds

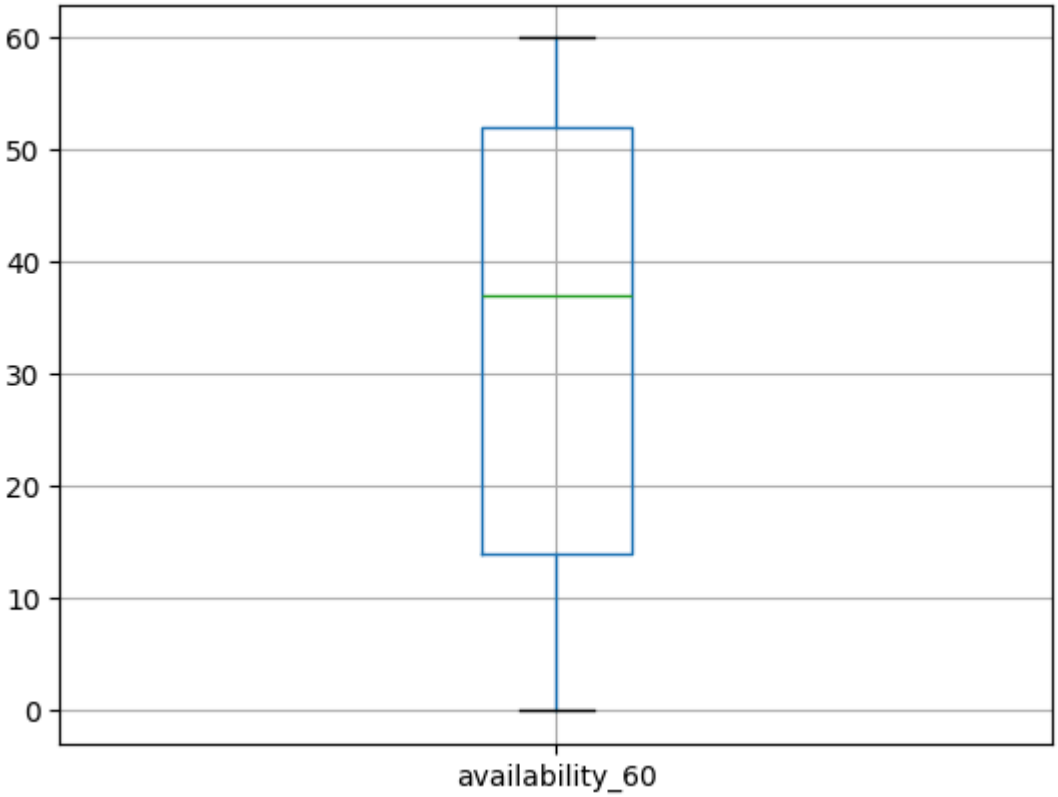

50
40
30
20
10
0
bathrooms

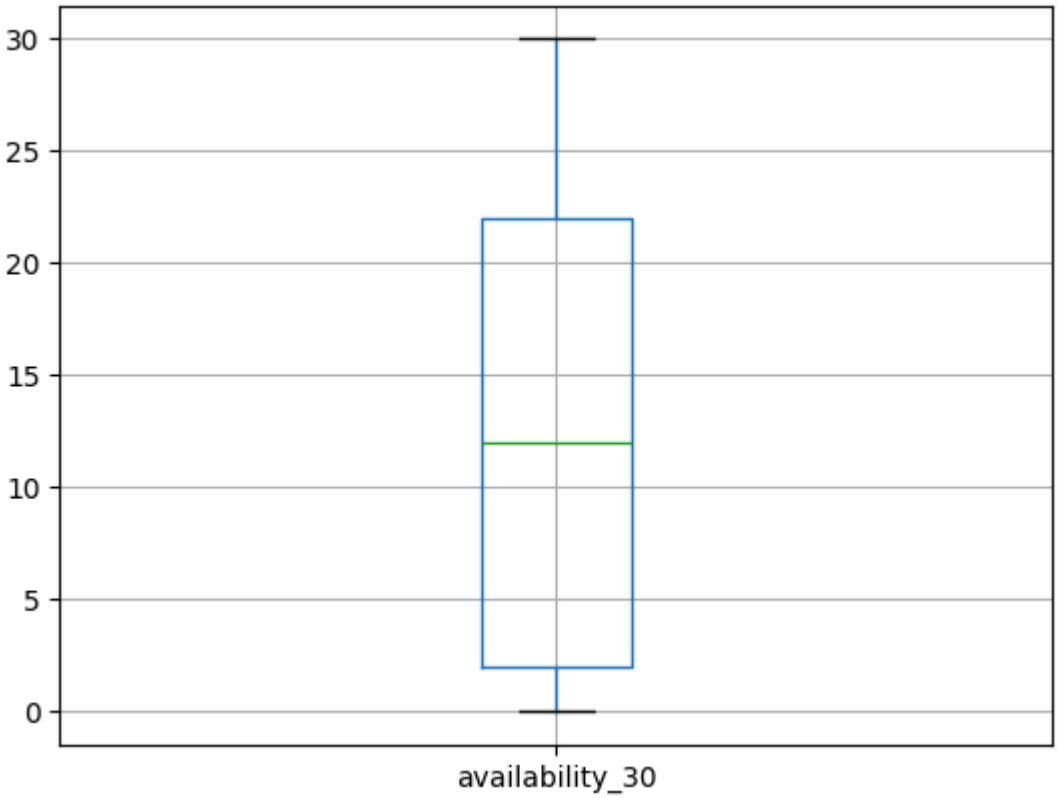

1000
800
600
400
200
0
minimum_nights

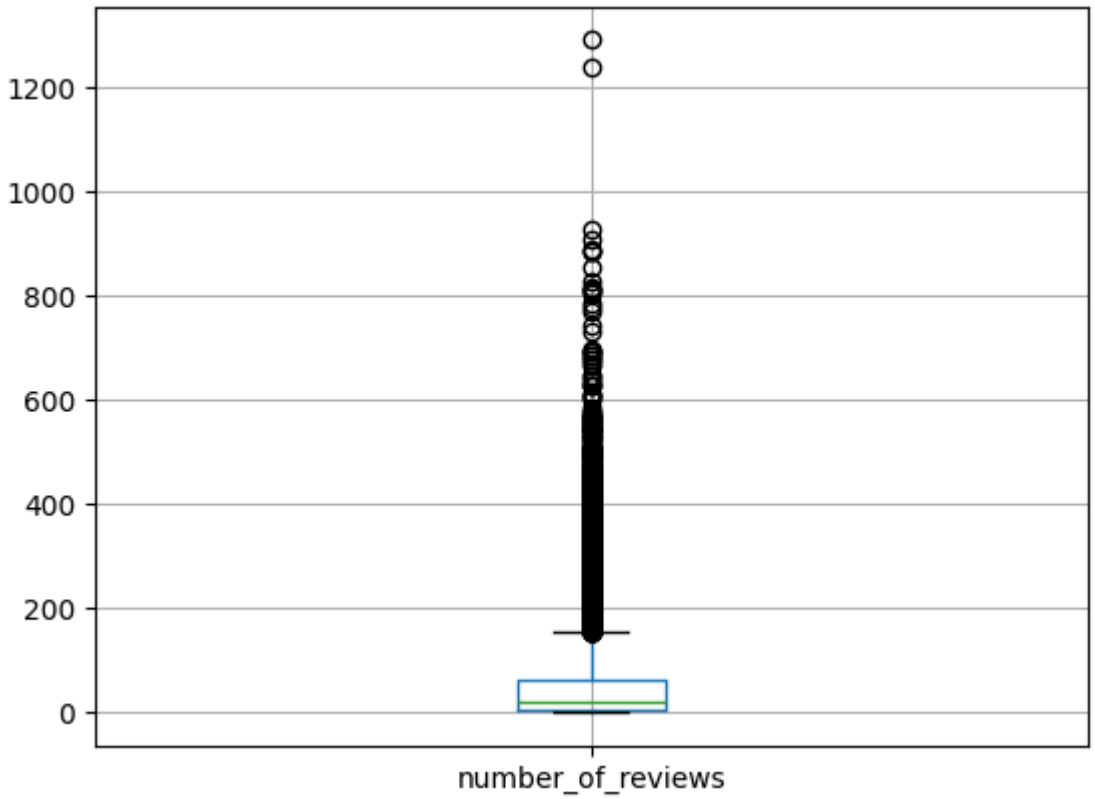

1000
800
600
400
200
0
maximum_nights

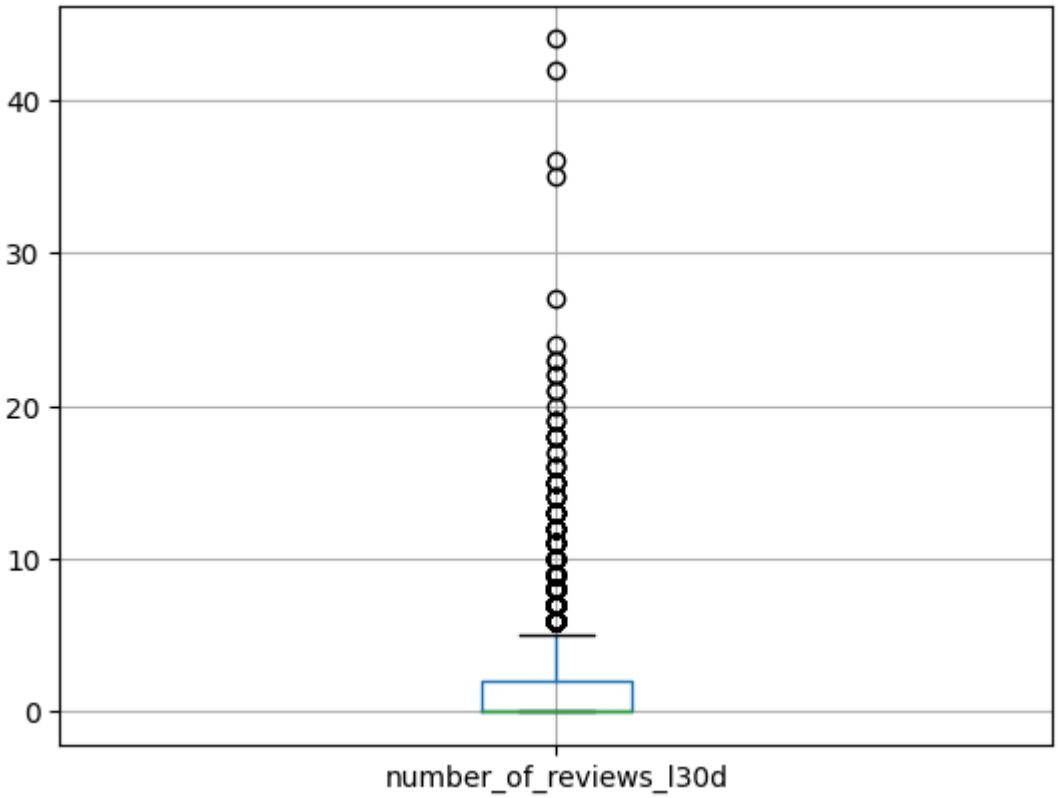

350
300
250
200
150
100
50
0
availability_365

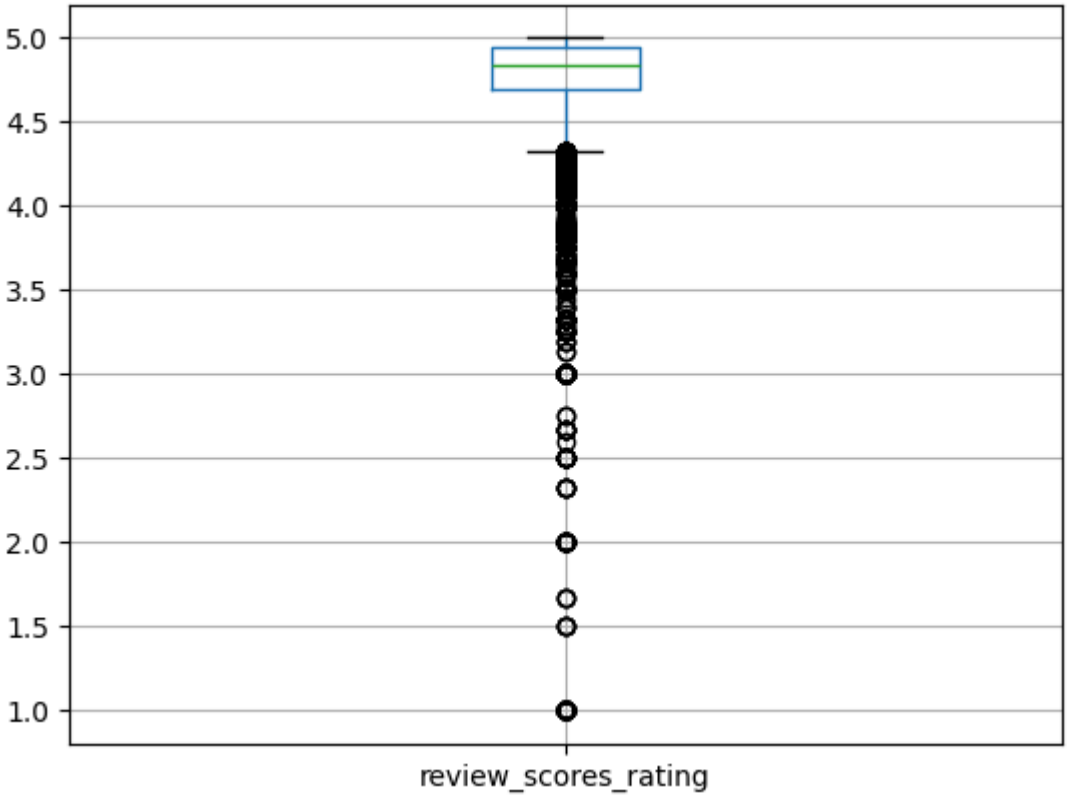

80
60
40
20
0
availability_90

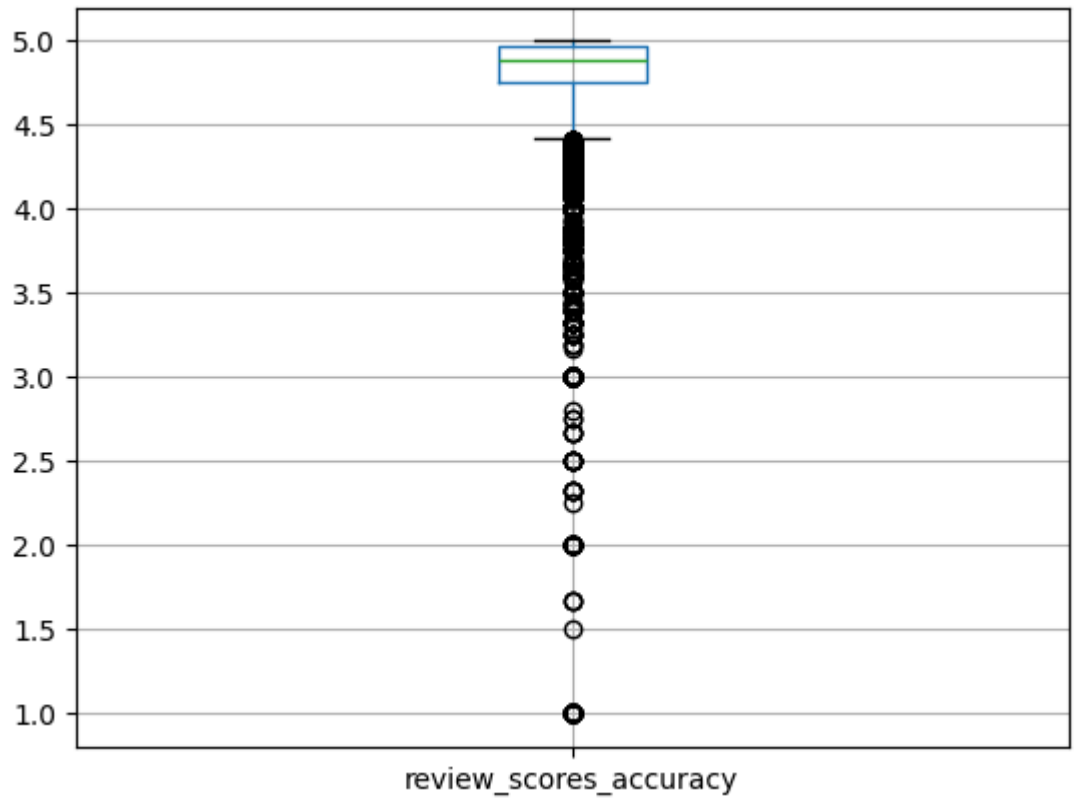

60
50
40
30
20
10
0
availability_60

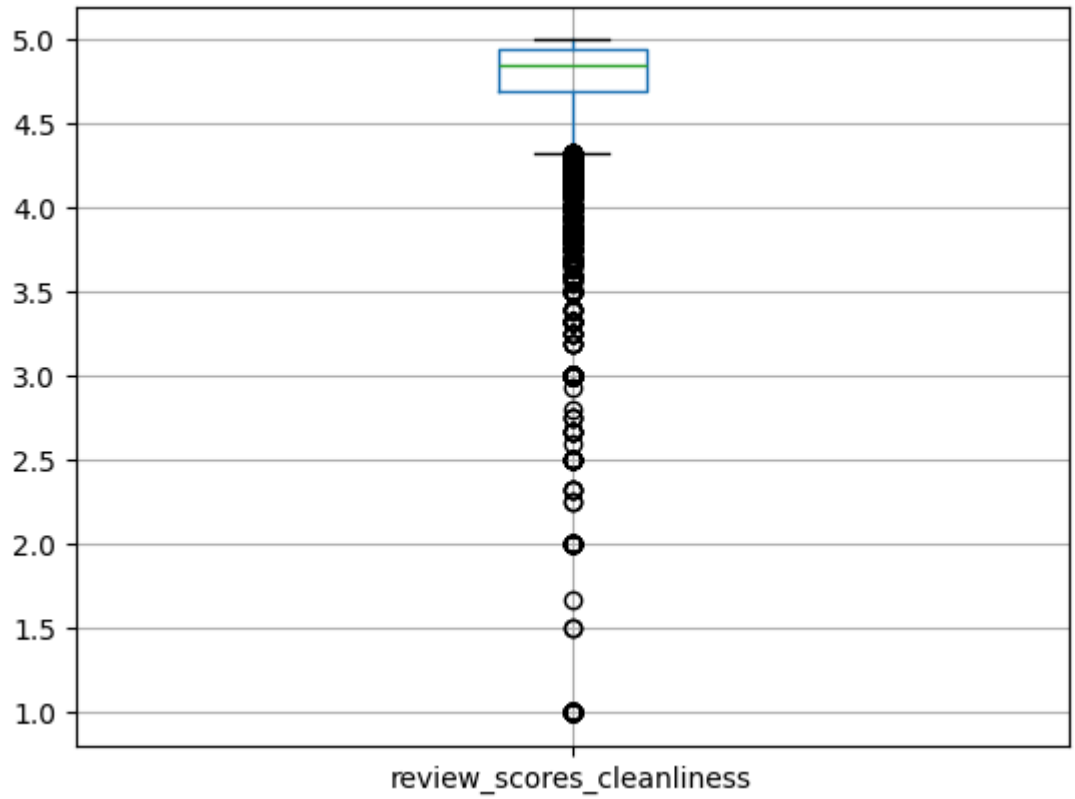

30
25
20
15
10
5
0
availability_30

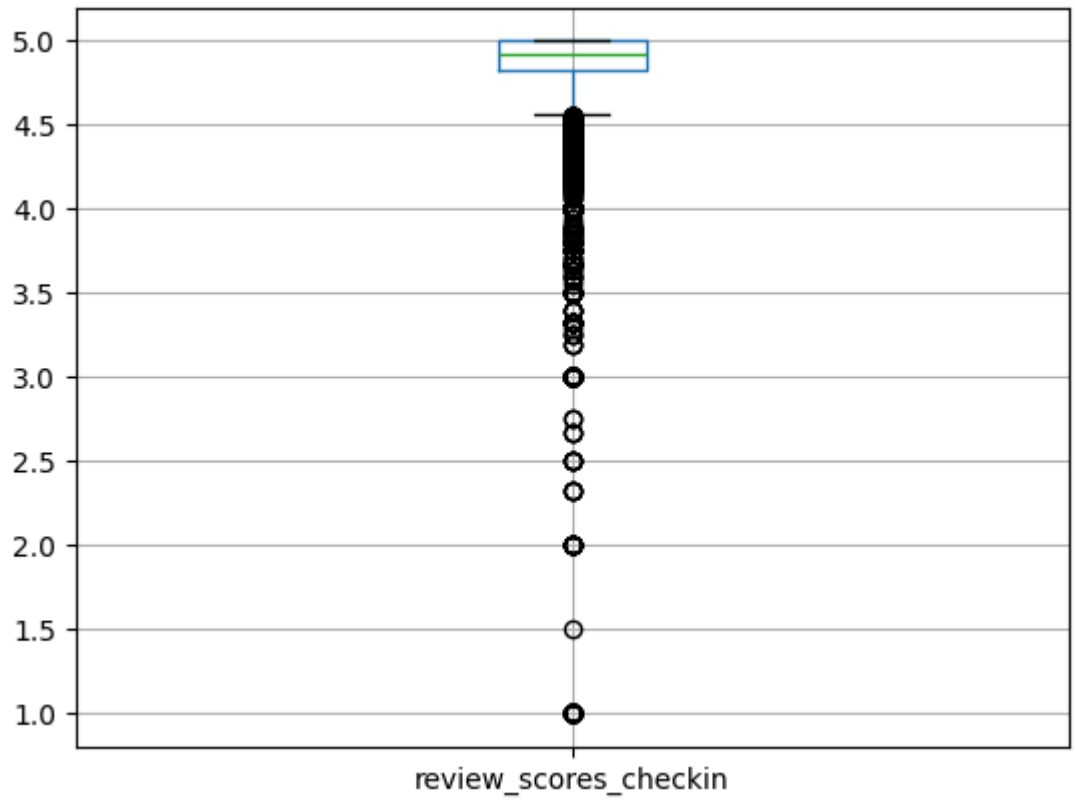

1200
1000
800
600
400
200
0
number_of_reviews

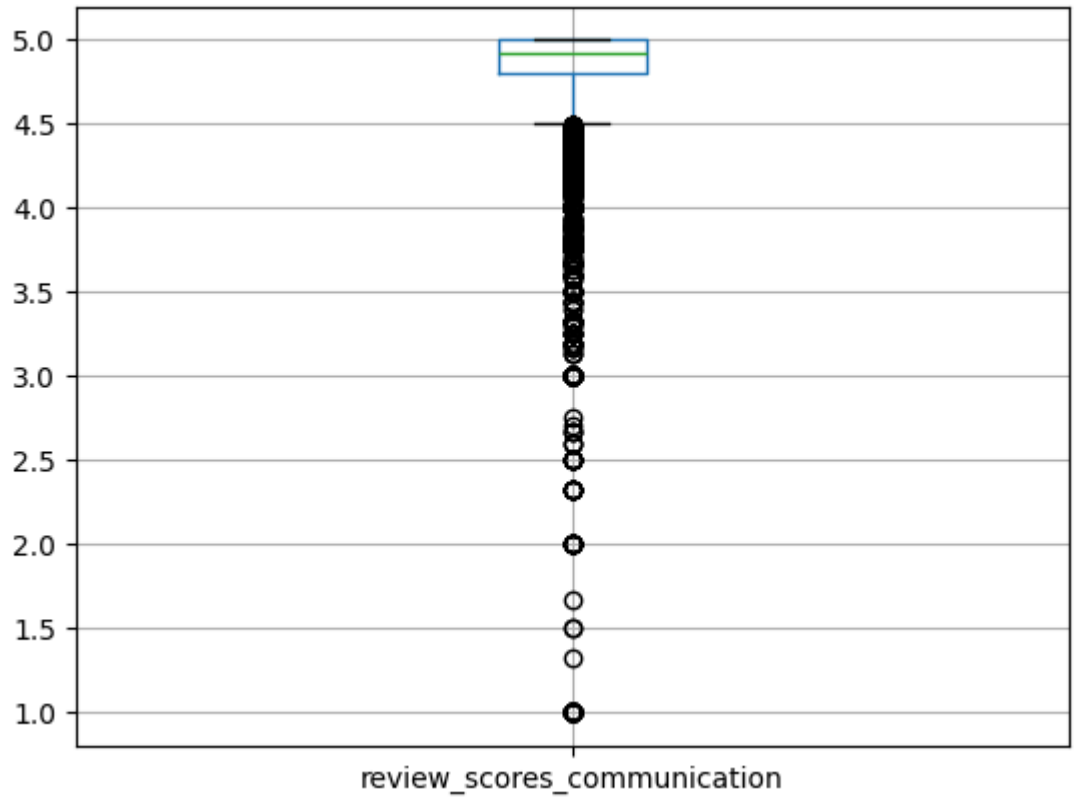

40
30
20
10
0
number_of_reviews_l30d

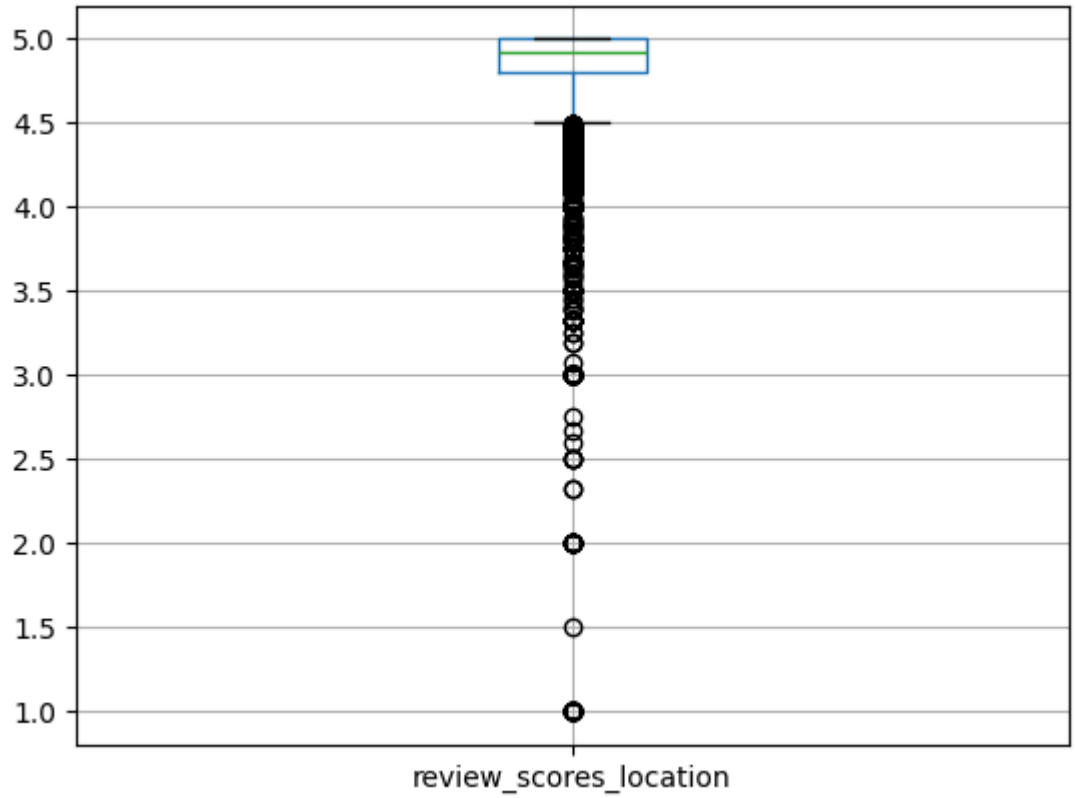

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_rating

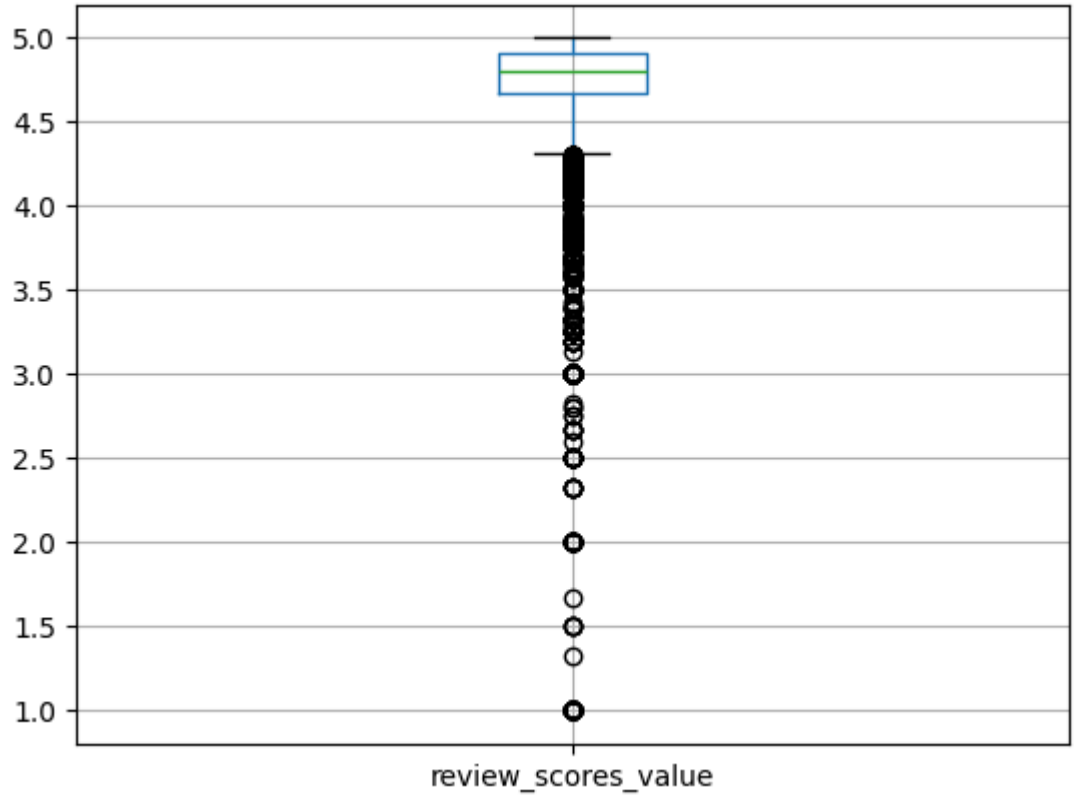

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_accuracy

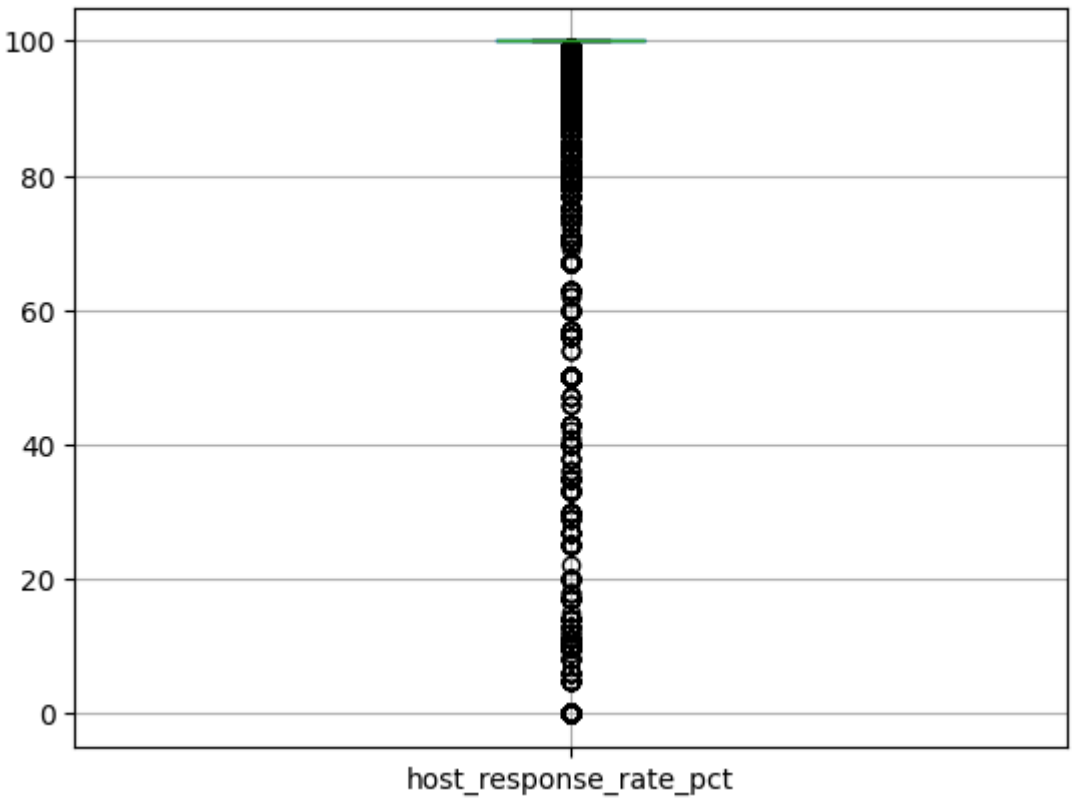

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_cleanliness

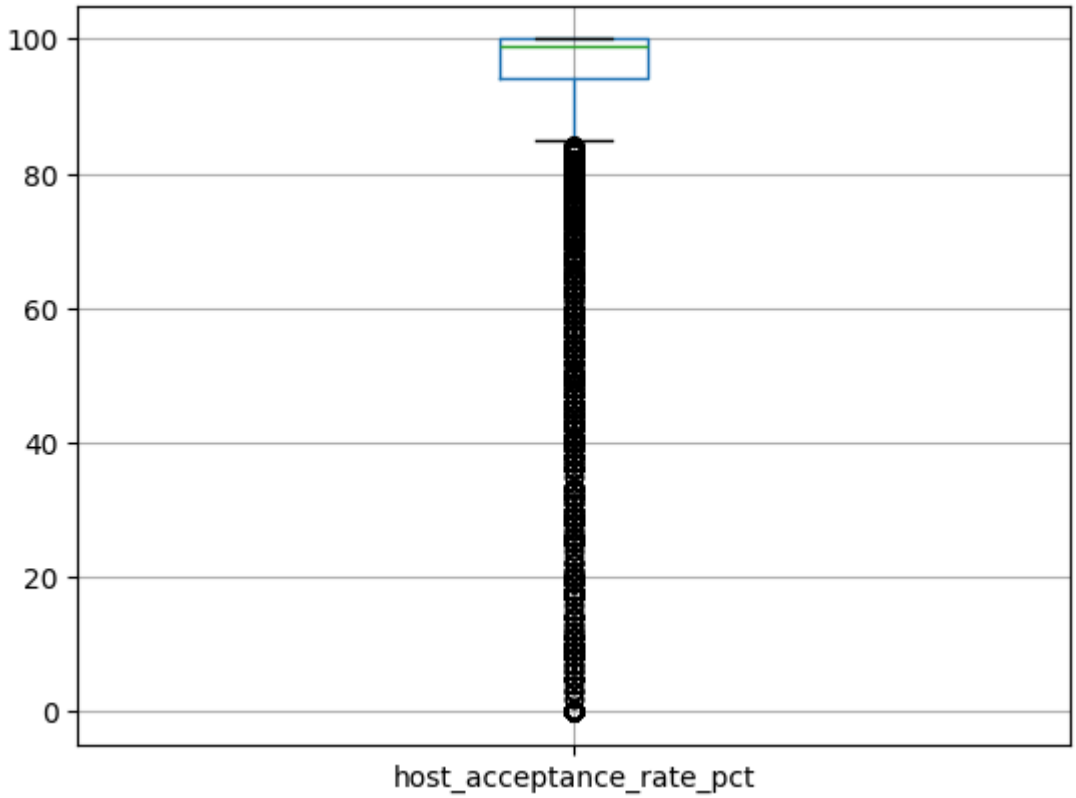

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_checkin

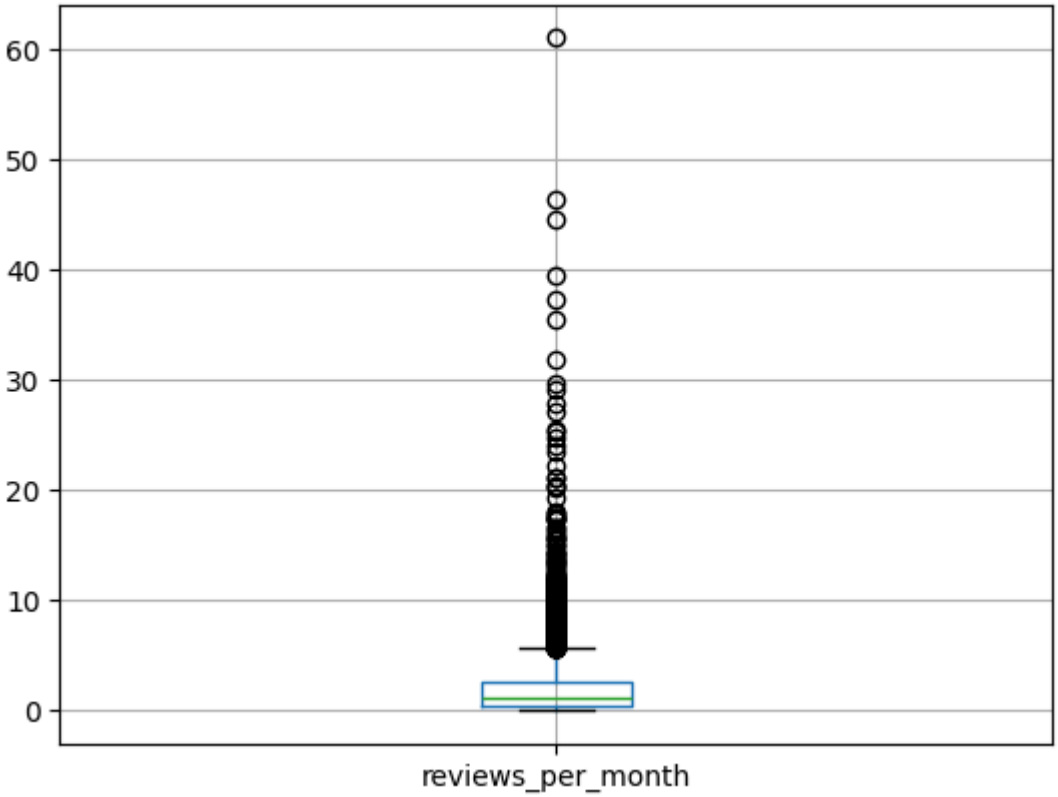

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_communication


5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_location

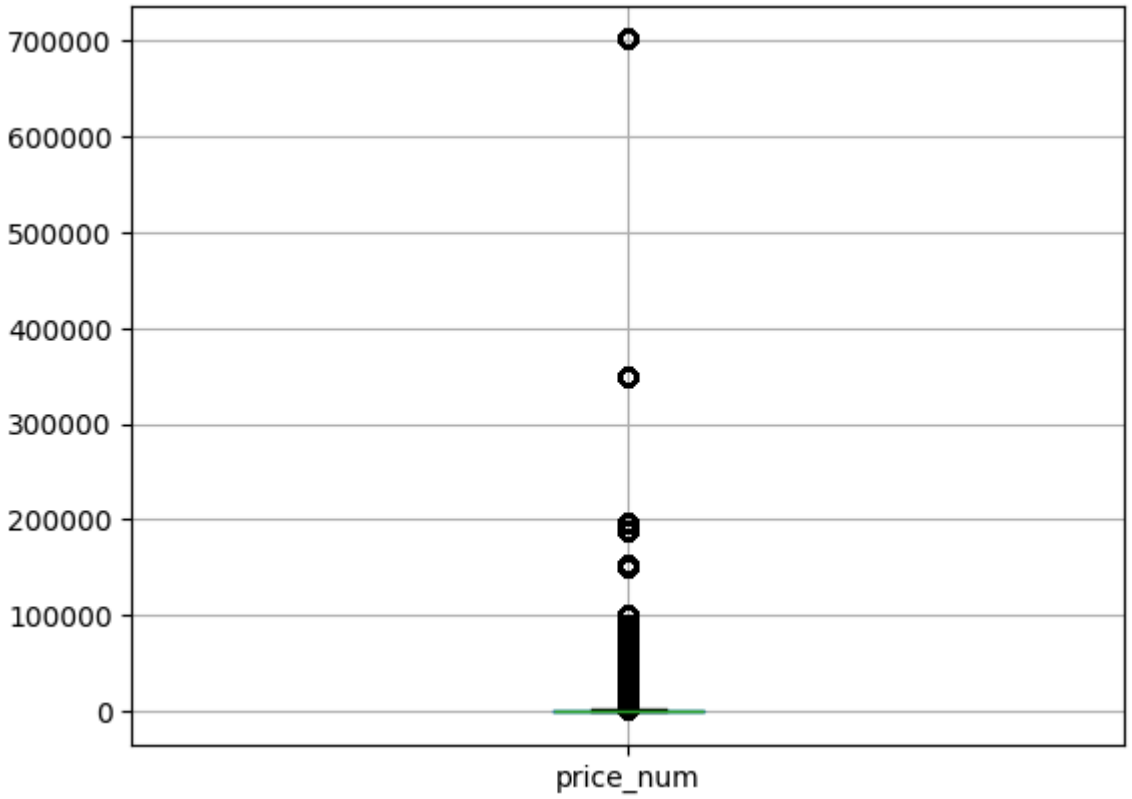

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
review_scores_value

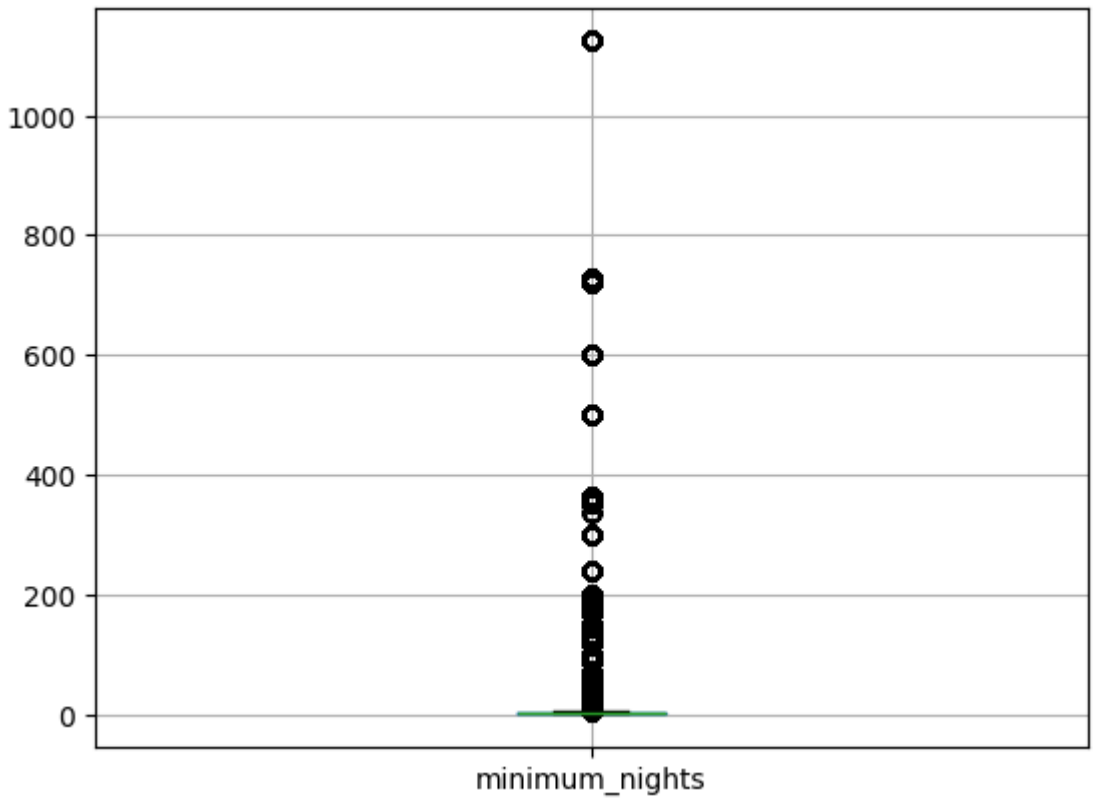

100
80
60
40
20
0
host_response_rate_pct

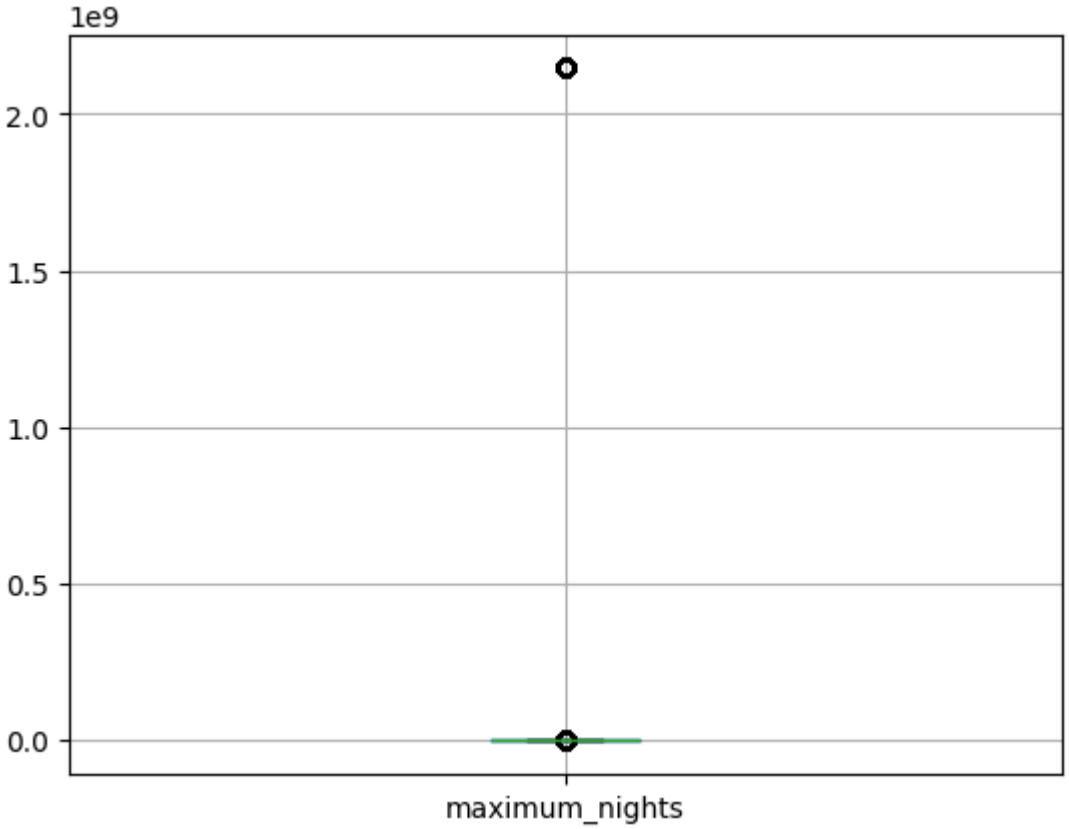

100
80
60
40
20
0
host_acceptance_rate_pct


60
50
40
30
20
10
0
reviews_per_month

- Calendar:


700000
600000
500000
400000
300000
200000
100000
0
price_num


1000
800
600
400
200
0
minimum_nights


1e9
2.0
1.5
1.0
0.5
0.0
maximum_nights

Hallazgos principales:

Aunque mucho de esto ya se cubrió en el resúmen del análisis exploratorio, pero aquí se va a condensar en lo más importante:

- La única colección con valores nulos eliminables es calendar, las otras pueden ser subsanadas por medio de imputación.
- No se encontraron registros duplicados en ninguno de los tres datasets.
- La gran mayoría de columnas numéricas de los datasets (exceptuando a reviews) presentan valores atípicos. Sin embargo, no todos los campos necesitan ser corregidos en este sentido.
- Se detectaron dos campos con estructura anidada en la colección listings: host_verifications y amenities. Los cuales contenían listas dentro de una sola celda.

Descripción de las transformaciones realizadas

Tipo de transformación	**Descripción**
Normalización de los datos	- Se pasaron las fechas a formato ISO date - Se pasaron las columnas price a float y se eliminó el símbolo $ y la coma. - Todos los porcentajes se pasaron a float y se eliminó el símbolo % - Se normalizó todo el texto, eliminando espacios, quitando el HTML y estandarizando unicode.
Limpieza de datos	Se usaron diferentes técnicas como la imputación por mediana para rellenar todos los valores nulos existentes en los 3 datasets.
Derivación de variables	Se extrajeron los campos de día, semana, mes, trimestre y año de las fechas de calendar; las cuales eran importante para la dimensión de tiempo. Adicionalmente, se generaron diferentes grupos para las columnas de precios de listings y calendar. En las que se clasifican por lo caros que son.
Desanidado de campos	Se expandieron los campos host_verifications y amenities en columnas binarias, representando la presencia o ausencia de cada elemento. Aunque en este caso no se mantuvieron

Tipo de transformación	Descripción
	todas, pues la gran mayoría no aportarían realmente nada de valor al negocio. Puesto que no todas las propiedades las poseen y lo único que van a generar es que la sábana de datos sea extremadamente grande (porque serían 100 columnas adicionales que darían de amenities) y tenga mucho ruido en los datos. Razón por la cual, se evaluó cuáles eran las 18 comunidades más presentes y esas fueron las elegidas. En cuanto a host verifications, si se usaron los 3 tipos presentes en la colección original.
Descarte de columnas innecesarias	Desde este punto, se tomó la oportunidad de no usar columnas que no tuvieran relevancia alguna para el negocio. Siendo ejemplos de esto urls, estimaciones, object_id, y los campos de la colección reviews (aunque estos se usaron para un campo calculado). Puesto que estos realmente no tienen influencia para el negocio y no tienen absolutamente nada que ver con temas de ocupación, calificaciones o ingresos. Adicionalmente, esto reduce el tamaño de dataset y lo organiza de mejor manera
Creación de variable derivada	Se generó el campo **reviews_per_month** con base en la relación entre los datasets reviews y calendar, permitiendo calcular la cantidad promedio de reseñas mensuales por alojamiento.

Todas las operaciones fueron registradas en logs para trazabilidad del proceso.

Ejemplo del log generado

2025-10-26 22:41:27,933 - INFO - === INICIO ETL (main_etl.py) ===

2025-10-26 22:41:27,977 - INFO - Conexión exitosa a la base de datos: bi_mx

2025-10-26 22:41:36,189 - INFO - [Origen Mongo] Esperados -> {'listings': 26067, 'calendar': 9514717, 'reviews': 1315986}

2025-10-26 22:41:37,409 - INFO - Colección listings_mx añadida al dataframe exitosamente. #Número de registros: 26067

2025-10-26 22:42:31,546 - INFO - Colección calendar_mx añadida al dataframe exitosamente. #Número de registros: 9514717

2025-10-26 22:42:42,158 - INFO - Colección reviews_mx añadida al dataframe exitosamente. #Número de registros: 1315986

2025-10-26 22:42:42,593 - INFO - Conexión con MongoDB cerrada existosamente

2025-10-26 22:42:42,944 - INFO - [INIT] Recibidos | listings: (26067, 77) cols=77 | calendar: (9514717, 8) cols=8 | reviews: (1315986, 7) cols=7

2025-10-26 22:42:42,944 - INFO - [run] Inicio | price_mode=quantile, price_bins=None, price_labels=None

2025-10-26 22:42:42,944 - INFO - [normalize_types] Inicio

2025-10-26 22:44:05,597 - INFO - [normalize_types] Calendar: drop ['adjusted_price', '_id']

2025-10-26 22:44:05,751 - INFO - [normalize_types] Listings: % → ['host_response_rate_pct', 'host_acceptance_rate_pct']

2025-10-26 22:44:30,649 - INFO - [normalize_types] Limpieza de texto: 11 cols

2025-10-26 22:44:30,649 - INFO - [_fill_review_dates] Inicio

2025-10-26 22:44:33,611 - INFO - [_fill_review_dates] Fin | fuentes: {'reviews_first': 0, 'calendar_first': 3261, 'last_scraped_first': 0, 'fallback_first': 0, 'reviews_last': 0, 'calendar_last': 3261, 'last_scraped_last': 0, 'fallback_last': 0}

2025-10-26 22:44:33,615 - INFO - [normalize_types] Fin | listings=(26067, 77) | calendar=(9514717, 6) | reviews=(1315986, 7)

2025-10-26 22:44:33,615 - INFO - [clean_nulls] Inicio

2025-10-26 22:44:33,682 - WARNING - [clean_nulls] Drops por claves: {'id': 0, 'host_id': 0}

2025-10-26 22:44:34,171 - INFO - [clean_nulls] host_response_rate_pct: host=0 grupo=3480 global=24

2025-10-26 22:44:34,222 - INFO - [clean_nulls] host_acceptance_rate_pct: host=0 grupo=2703 global=36

2025-10-26 22:44:34,282 - INFO - [clean_nulls] review_scores_rating: grupo=3215 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,341 - INFO - [clean_nulls] review_scores_cleanliness: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,400 - INFO - [clean_nulls] review_scores_accuracy: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,460 - INFO - [clean_nulls] review_scores_communication: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,520 - INFO - [clean_nulls] review_scores_checkin: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,580 - INFO - [clean_nulls] review_scores_location: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,637 - INFO - [clean_nulls] review_scores_value: grupo=3216 barrio=46 global=0 winsor p1..p99

2025-10-26 22:44:34,638 - INFO - [clean_nulls] reviews_per_month: 3261 → 0.0

2025-10-26 22:44:34,641 - INFO - [clean_nulls] latitude: barrio=0 global=0

2025-10-26 22:44:34,644 - INFO - [clean_nulls] longitude: barrio=0 global=0

2025-10-26 22:44:34,653 - INFO - [clean_nulls] bathrooms: imputados 3991 desde bathrooms_text (sin redondear).

2025-10-26 22:44:34,653 - WARNING - [clean_nulls] bathrooms: 20 sin fuente → 0.0.

2025-10-26 22:44:34,660 - INFO - [clean_nulls] availability_30: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,665 - INFO - [clean_nulls] availability_60: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,670 - INFO - [clean_nulls] availability_90: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,675 - INFO - [clean_nulls] availability_365: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,680 - INFO - [clean_nulls] number_of_reviews: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,685 - INFO - [clean_nulls] number_of_reviews_ltm: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,690 - INFO - [clean_nulls] number_of_reviews_l30d: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,695 - INFO - [clean_nulls] number_of_reviews_ly: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,700 - INFO - [clean_nulls] accommodates: grupo=0 global=0 cast→int≥0

2025-10-26 22:44:34,723 - INFO - [clean_nulls] bedrooms: grupo=967 global=4 cast→int≥0

2025-10-26 22:44:34,789 - INFO - [clean_nulls] beds: grupo=3992 global=30 cast→int≥0

2025-10-26 22:44:34,797 - INFO - [clean_nulls] host_total_listings_count: grupo=5 global=0 cast→int≥0

2025-10-26 22:44:34,826 - INFO - [clean_nulls] host_is_superhost: 1343 → False

2025-10-26 22:44:34,829 - INFO - [clean_nulls] host_has_profile_pic: 5 → False

2025-10-26 22:44:34,832 - INFO - [clean_nulls] host_identity_verified: 5 → False

2025-10-26 22:44:34,866 - INFO - [clean_nulls] host_since: desde last_scraped=5 | a 1970=0

2025-10-26 22:44:35,436 - INFO - [clean_nulls] calendar.minimum_nights: 220 → mediana=2.0

2025-10-26 22:44:35,583 - INFO - [clean_nulls] calendar.maximum_nights: 220 → mediana=1120.0

2025-10-26 22:44:36,889 - INFO - [clean_nulls] Red de seguridad listings: {'description': "str→'unknown' 767", 'neighborhood_overview': "str→'unknown' 11678", 'host_location': "str→'unknown' 5480", 'host_about': "str→'unknown' 10457", 'host_thumbnail_url': "str→'unknown' 5", 'host_picture_url': "str→'unknown' 5", 'host_listings_count': 'num→mediana(4.0) 5', 'neighbourhood': "str→'unknown' 11676", 'bathrooms_text': "str→'unknown' 27", 'has_availability': "str→'unknown' 991", 'estimated_revenue_l365d': 'num→mediana(57234.0) 3999', 'host_neighbourhood': "str→'unknown' 12192", 'price_num': 'num→mediana(1055.0) 3999'}

2025-10-26 22:44:37,704 - INFO - [clean_nulls] Fin | tamaños antes L:26067 C:9514717 R:1315986 | después L:26067 C:9514717 R:1315986

2025-10-26 22:44:37,704 - INFO - [derive_features] Inicio | price_mode=quantile

2025-10-26 22:44:39,910 - INFO - [derive_features] Fin | listings=(26067, 90) | calendar=(9514717, 12) | reviews=(1315986, 7)

2025-10-26 22:44:39,910 - INFO - [expand_nested_fields] Inicio

2025-10-26 22:44:41,506 - INFO - [expand_nested_fields] Fin | amenities=18 verifications=3

2025-10-26 22:44:41,533 - INFO - [build_flat_sheet] Inicio

2025-10-26 22:44:42,185 - INFO - [_impute_listing_prices_and_buckets] Inicio

2025-10-26 22:44:42,367 - INFO - [_impute_listing_prices_and_buckets] Fin | from_calendar=0 | by_neighbourhood=0 | global=0 | bucket_mode=qcut

2025-10-26 22:44:49,246 - INFO - [build_flat_sheet] Join: filas calendar=9514717 | resultado=9514717 | col_listings_dim=62

2025-10-26 22:45:07,658 - INFO - [build_flat_sheet] reviews_in_month: 9412966 → 0

2025-10-26 22:45:13,481 - INFO - [build_flat_sheet] Fin | flat_sheet=(9514717, 77)

2025-10-26 22:45:15,132 - INFO - [run] Fin | flat_sheet=(9514717, 77)

2025-10-26 22:45:15,133 - INFO - [Transformacion] DF final con 9514717 filas y 77 columnas.

2025-10-26 22:45:20,174 - INFO - Carga inicializada con 9514717 registros.

2025-10-26 22:45:20,174 - INFO - === INICIO DE CARGA DE DATOS ===

2025-10-26 22:45:20,185 - INFO - Conectado a SQLite: data/airbnb.db

2025-10-26 22:46:55,809 - INFO - Datos insertados en la tabla 'airbnb_limpio' correctamente.

2025-10-26 22:46:55,817 - INFO - Conectado a SQLite: data/airbnb.db

2025-10-26 22:47:15,152 - INFO - Verificación: 9514717 registros encontrados en 'airbnb_limpio'.

2025-10-26 22:47:16,126 - INFO - [ExcelPart] Inicio | filas=9,514,717 | máx/archivo=200,000 | cols=77

2025-10-26 22:47:16,127 - INFO - [ExcelPart] Truncando columnas de texto a 500 chars en 12 columnas...

2025-10-26 22:47:29,972 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_1.xlsx' para filas 0..199,999 (total 200,000)

2025-10-26 22:49:58,154 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_1.xlsx'

2025-10-26 22:49:58,155 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_2.xlsx' para filas 200,000..399,999 (total 200,000)

2025-10-26 22:52:39,631 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_2.xlsx'

2025-10-26 22:52:39,631 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_3.xlsx' para filas 400,000..599,999 (total 200,000)

2025-10-26 22:54:49,032 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_3.xlsx'

2025-10-26 22:54:49,032 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_4.xlsx' para filas 600,000..799,999 (total 200,000)

2025-10-26 22:56:47,825 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_4.xlsx'

2025-10-26 22:56:47,826 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_5.xlsx' para filas 800,000..999,999 (total 200,000)

2025-10-26 22:59:15,863 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_5.xlsx'

2025-10-26 22:59:15,863 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_6.xlsx' para filas 1,000,000..1,199,999 (total 200,000)

2025-10-26 23:02:01,642 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_6.xlsx'

2025-10-26 23:02:01,642 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_7.xlsx' para filas 1,200,000..1,399,999 (total 200,000)

2025-10-26 23:05:02,254 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_7.xlsx'

2025-10-26 23:05:02,254 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_8.xlsx' para filas 1,400,000..1,599,999 (total 200,000)

2025-10-26 23:07:36,549 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_8.xlsx'

2025-10-26 23:07:36,550 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_9.xlsx' para filas 1,600,000..1,799,999 (total 200,000)

2025-10-26 23:09:35,241 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_9.xlsx'

2025-10-26 23:09:35,241 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_10.xlsx' para filas 1,800,000..1,999,999 (total 200,000)

2025-10-26 23:11:37,223 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_10.xlsx'

2025-10-26 23:11:37,224 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_11.xlsx' para filas 2,000,000..2,199,999 (total 200,000)

2025-10-26 23:13:36,615 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_11.xlsx'

2025-10-26 23:13:36,616 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_12.xlsx' para filas 2,200,000..2,399,999 (total 200,000)

2025-10-26 23:16:07,774 - INFO - [ExcelPart] OK 'data/airbnb_limpio_part_12.xlsx'

2025-10-26 23:16:07,775 - INFO - [ExcelPart] Generando 'data/airbnb_limpio_part_13.xlsx' para filas 2,400,000..2,599,999 (total 200,000)

…

2025-10-26 12:00:00,775 - INFO - [ExcelPart] Total archivos generados: 48

2025-10-26 12:01:00,775 - INFO - === FIN DE CARGA DE DATOS ===

2025-10-26 12:02:00,775 - INFO - === FIN ETL (main_etl.py) ===

Conclusiones sobre la calidad y utilidad de los datos

- Los datos analizados presentan alta consistencia y completitud, sin duplicados relevantes ni valores faltantes críticos.
- Las transformaciones realizadas permitieron mejorar la estructura de los datos para su análisis posterior en herramientas como Power BI o Excel
- La inclusión de variables derivadas (mes, año, rango de precios) facilitó la generación de indicadores y gráficos descriptivos.
- El dataset de Airbnb para Ciudad de México se considera fiable y útil para análisis exploratorios y predictivos, permitiendo obtener una visión clara del comportamiento del mercado de hospedaje.

Referencias

- Inside Airbnb – Dataset Ciudad de México: https://insideairbnb.com/get-the-data/
- Documentación oficial de pandas: https://pandas.pydata.org

- Python Software Foundation – módulo sqlite3
- Guías académicas del curso