



Avance para Aimbridge LATAM



EQUIPO



Carlos Verdaguer
Project Manager



Fortunato Martínez
Director of Operations



Jorge Oviedo
Director of Logistics



Gerardo Barajas
Data Scientist



Sebastian Neira
Data Analyst



Anibal Angulo
Data Analyst

Propósito

Generar valor **SIGNIFICATIVO**
para nuestros clientes.

Contenidos

- Data Understanding
 - Con y sin "datos atípicos"
- Recolección de variables externas
 - Complicaciones de obtención de datos
 - Propuestas de solución
 - Uso de API
- Propuesta para Modelos
 - Utilización para los hoteles

Data Understanding

Datos Obtenidos y Utilizados

- Por el momento se utilizan los datos de las regiones más grandes en México (MTY, GDL y CDMX).
 - Tres venues, uno por región, con datos desde 2020 hasta presente de 2022
- Adicionalmente se tienen 15 venues en 13 distintas entidades de México
 - Con datos desde 2021 hasta presente de 2022

Estadística Descriptiva

CON Datos Atípicos

	venue_id	found	night_id	business_date_id	month	rooms_revenue	rooms_available	rooms_occupied
count	2646.000000	2646.000000	2646.000000	2646.000000	2646.000000	2646.000000	2646.000000	2646.000000
mean	33.000000	0.861678	23961.500000	18703.500000	5.919501	64905.656863	165.677627	47.900983
std	18.058883	0.345303	4408.382416	254.659431	3.458086	90656.325282	69.456067	75.422640
min	16.000000	0.000000	17641.000000	18263.000000	1.000000	-1835.400000	0.000000	0.000000
25%	16.000000	1.000000	18302.250000	18483.000000	3.000000	12379.872500	168.000000	11.000000
50%	25.000000	1.000000	25137.500000	18703.500000	5.000000	45060.050000	180.000000	38.000000
75%	58.000000	1.000000	28444.750000	18924.000000	9.000000	84228.850000	219.000000	63.000000
max	58.000000	1.000000	29106.000000	19144.000000	12.000000	904534.230000	219.000000	940.000000

SIN Datos Atípicos

	venue_id	found	night_id	business_date_id	month	rooms_revenue	rooms_available	rooms_occupied
count	2140.000000	2140.0	2140.000000	2140.000000	2140.000000	2140.000000	2140.000000	2140.000000
mean	29.303738	1.0	23639.045794	18744.836449	5.906542	58944.089290	191.229439	45.247196
std	16.567323	0.0	4705.950956	237.624478	3.447008	44196.248794	21.416989	30.686369
min	16.000000	1.0	17641.000000	18263.000000	1.000000	-1835.400000	162.000000	0.000000
25%	16.000000	1.0	18194.750000	18559.000000	3.000000	28401.215000	180.000000	24.000000
50%	25.000000	1.0	25278.500000	18761.000000	5.000000	49863.705000	180.000000	42.000000
75%	25.000000	1.0	28503.250000	18942.250000	9.000000	83202.847500	219.000000	63.000000
max	58.000000	1.0	29105.000000	19144.000000	12.000000	191762.750000	219.000000	165.000000

Matriz de Correlaciones

CON Datos Atípicos

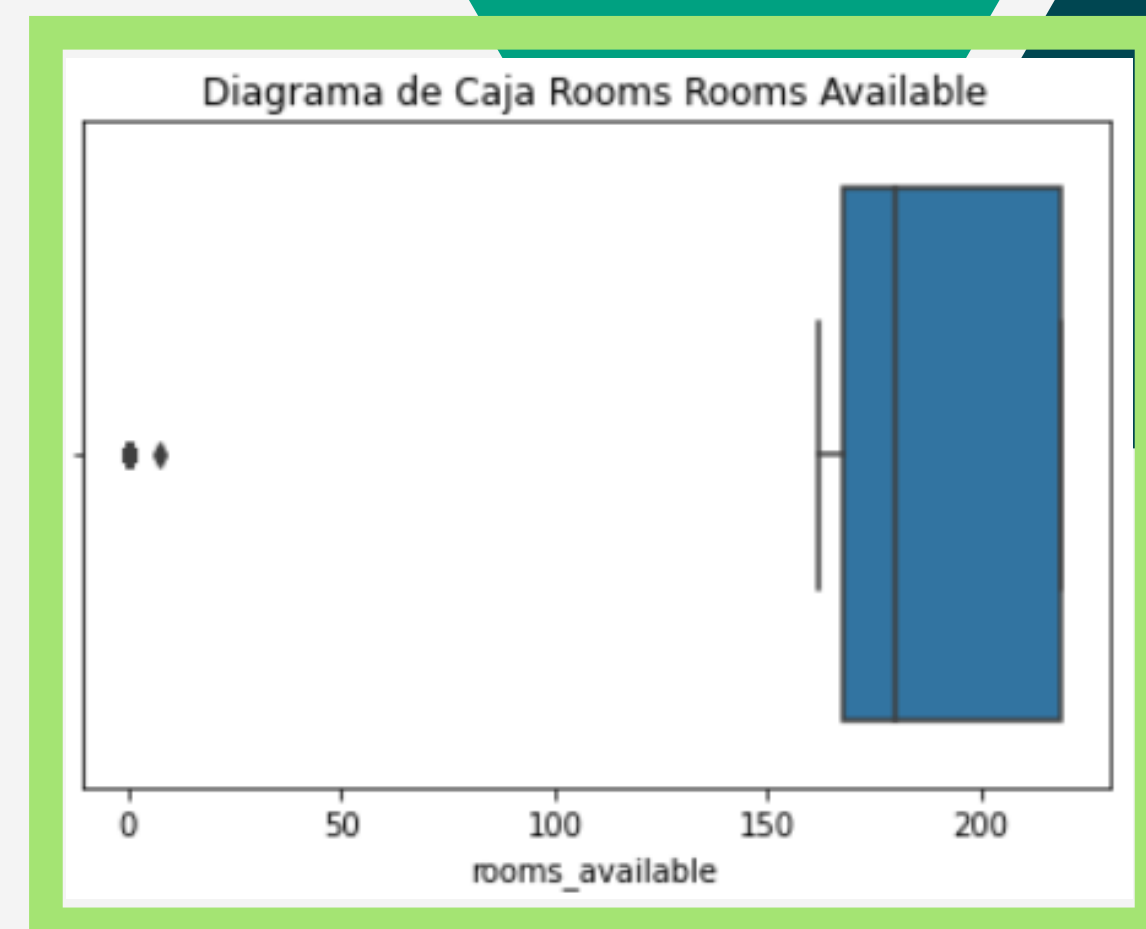
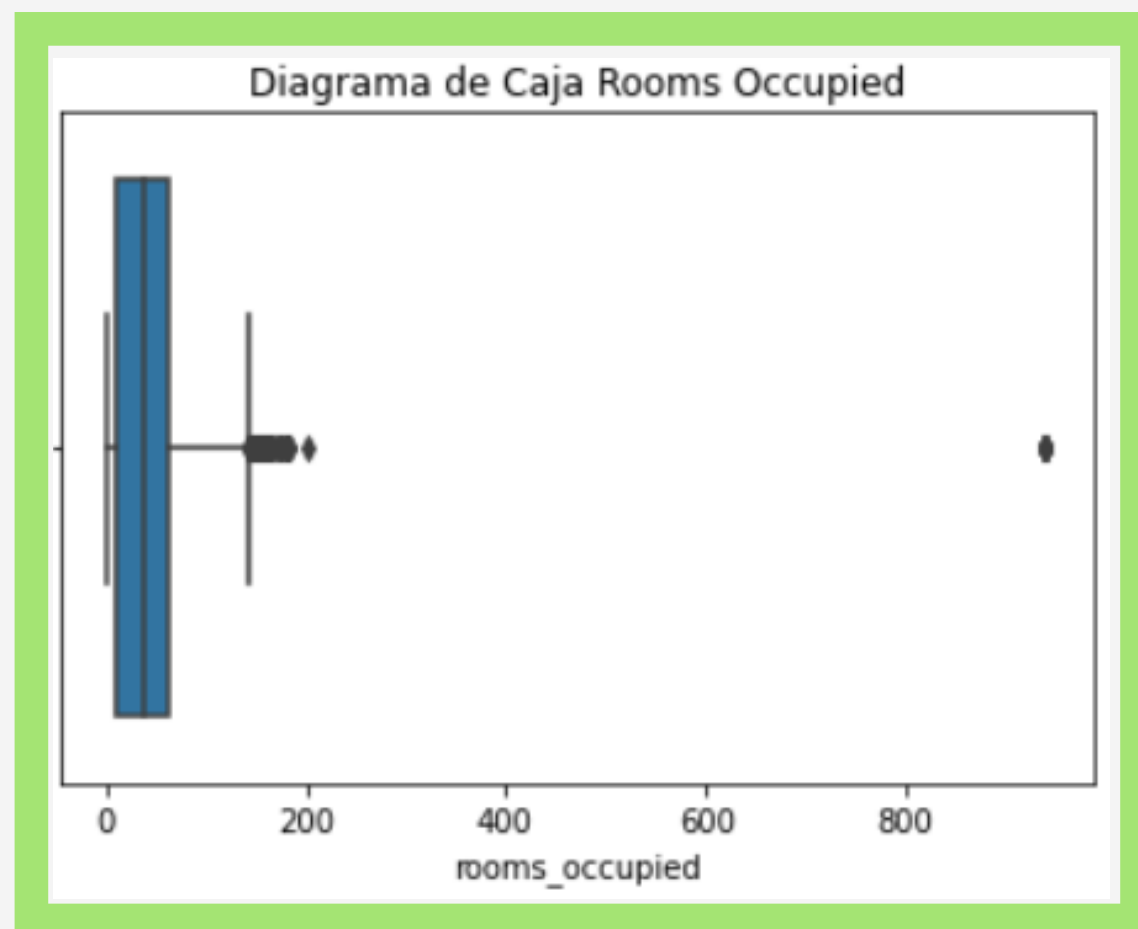
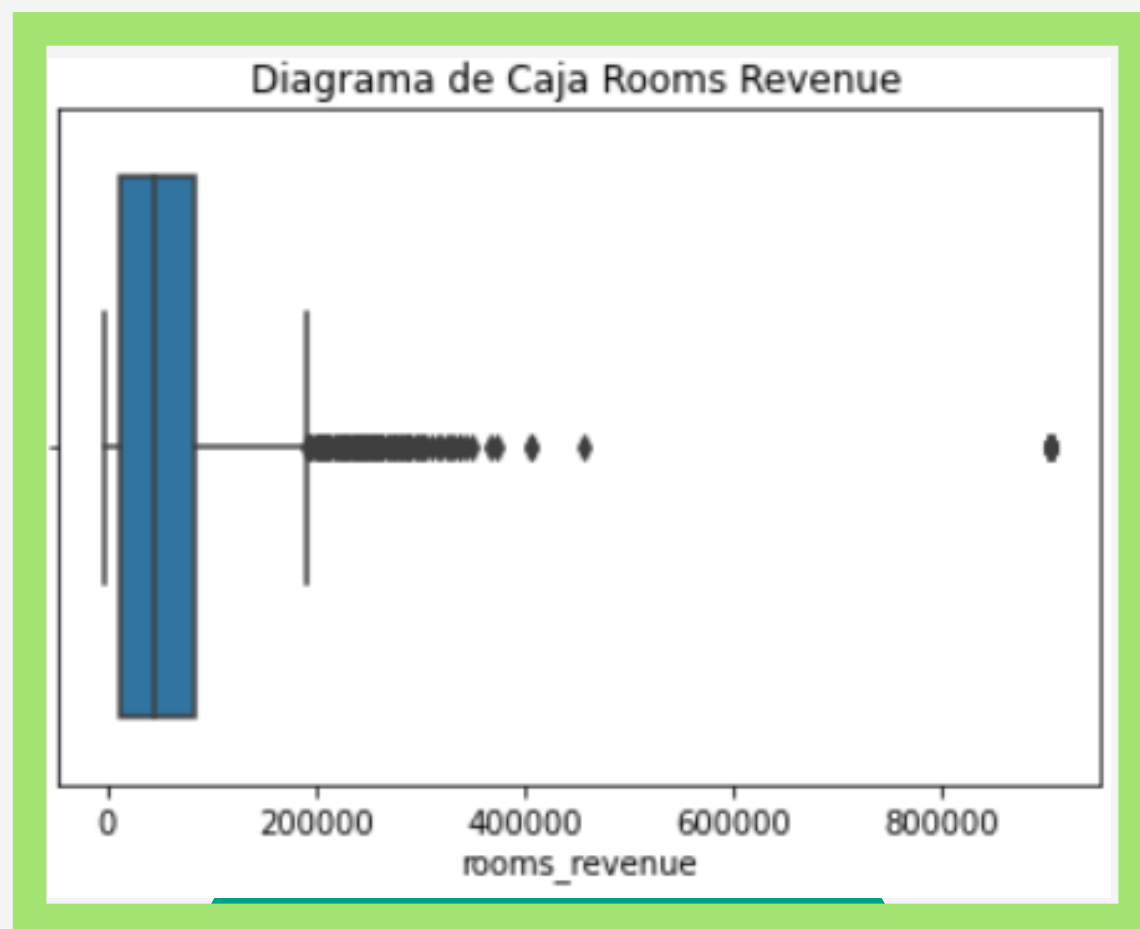
	venue_id	found	night_id	business_date_id	month	rooms_revenue	rooms_available	rooms_occupied
venue_id	1.000000e+00	-0.554759	0.384215	2.922122e-14	-1.477003e-15	-0.244805	-0.625588	-0.218035
found	-5.547590e-01	1.000000	-0.083449	4.059898e-01	-6.885301e-02	0.286906	0.955892	0.254506
night_id	3.842153e-01	-0.083449	1.000000	5.776709e-02	6.071407e-03	0.093583	0.113520	-0.071509
business_date_id	2.922122e-14	0.405990	0.057767	1.000000e+00	1.051015e-01	0.213322	0.339979	0.188545
month	-1.477003e-15	-0.068853	0.006071	1.051015e-01	1.000000e+00	-0.072039	-0.058530	-0.048710
rooms_revenue	-2.448055e-01	0.286906	0.093583	2.133219e-01	-7.203929e-02	1.000000	0.331294	0.906440
rooms_available	-6.255881e-01	0.955892	0.113520	3.399793e-01	-5.853028e-02	0.331294	1.000000	0.249142
rooms_occupied	-2.180350e-01	0.254506	-0.071509	1.885448e-01	-4.870977e-02	0.906440	0.249142	1.000000

SIN Datos Atípicos

	venue_id	found	night_id	business_date_id	month	rooms_revenue	rooms_available	rooms_occupied
venue_id	1.000000	NaN	0.426209	0.316327	-0.061769	-0.067277	-0.403759	-0.065054
found	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
night_id	0.426209	NaN	1.000000	0.086298	0.014708	0.250595	0.654361	-0.017972
business_date_id	0.316327	NaN	0.086298	1.000000	0.060292	0.471996	-0.222184	0.541191
month	-0.061769	NaN	0.014708	0.060292	1.000000	0.047198	0.062199	0.046688
rooms_revenue	-0.067277	NaN	0.250595	0.471996	0.047198	1.000000	0.283377	0.920983
rooms_available	-0.403759	NaN	0.654361	-0.222184	0.062199	0.283377	1.000000	0.007615
rooms_occupied	-0.065054	NaN	-0.017972	0.541191	0.046688	0.920983	0.007615	1.000000

Diagramas de Caja

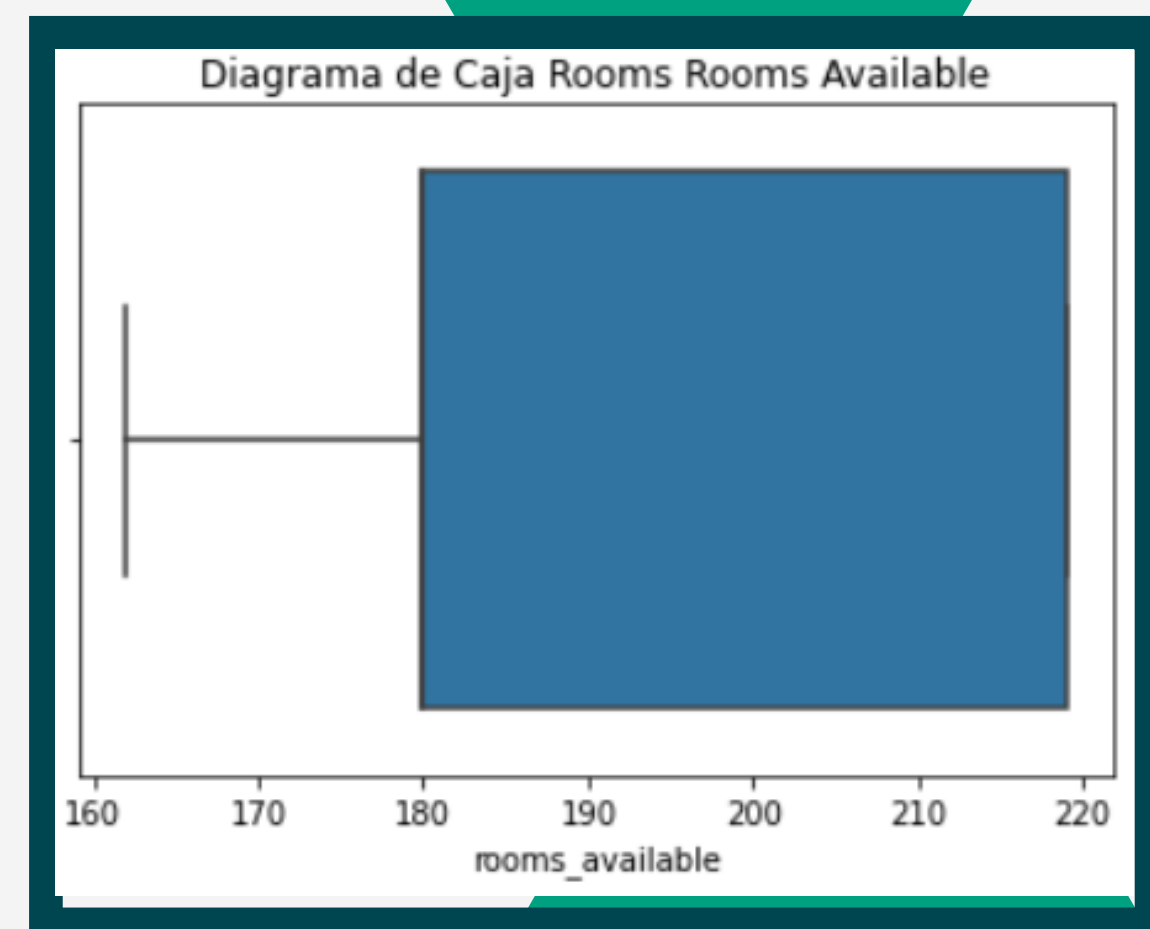
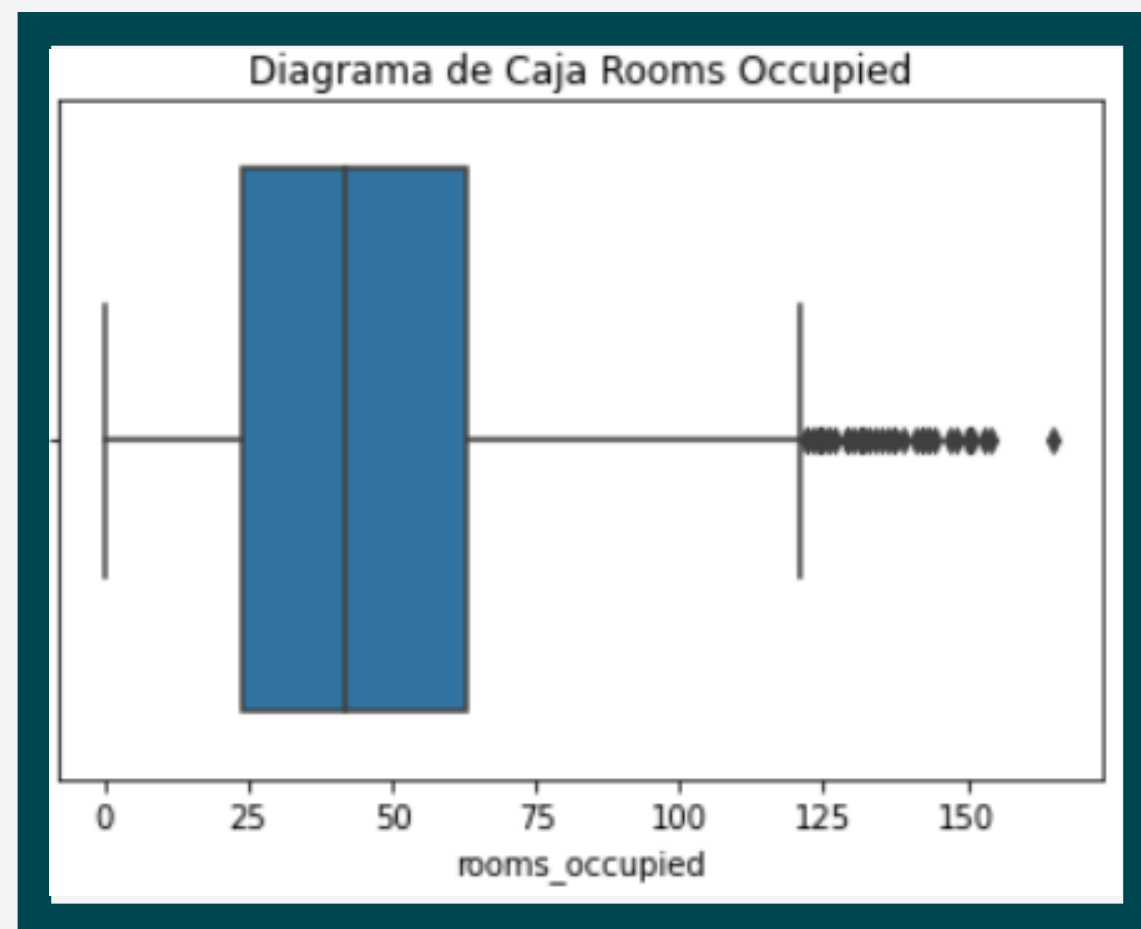
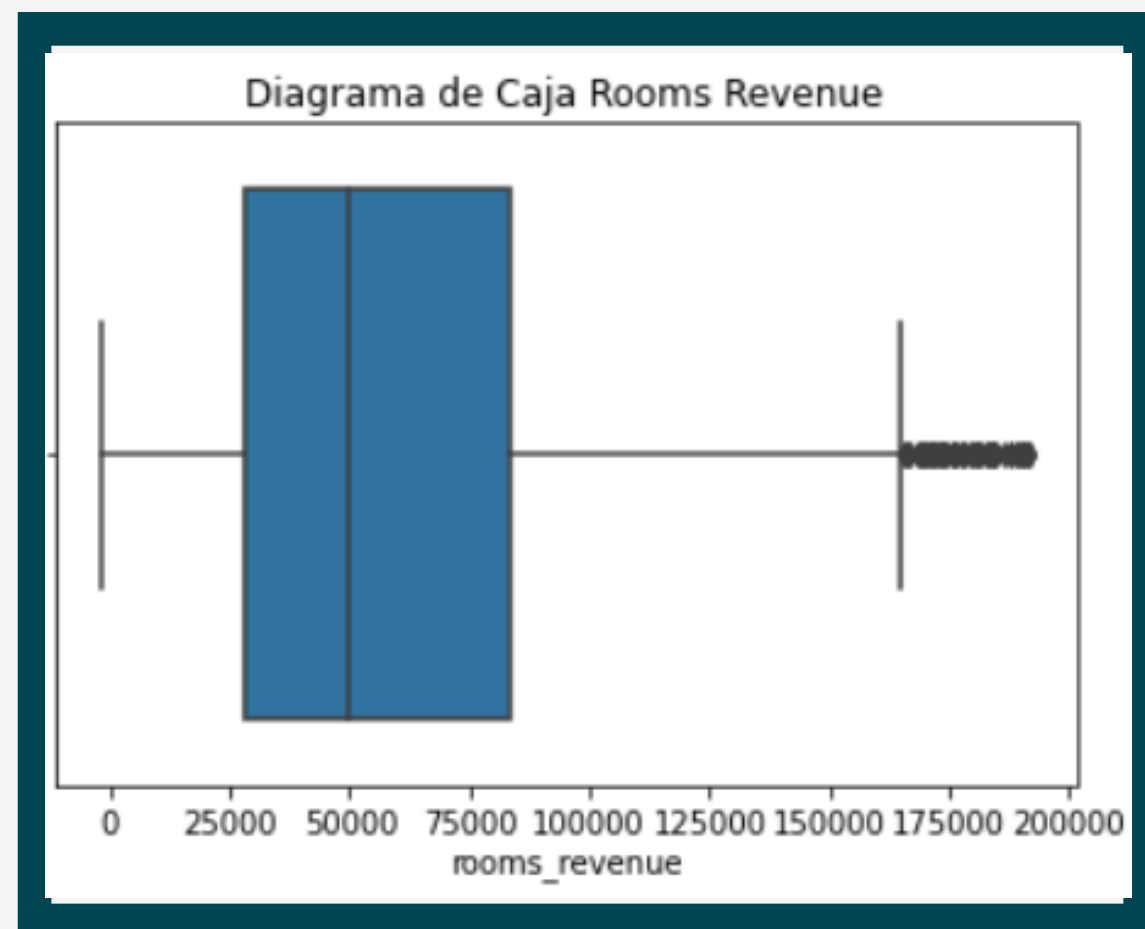
CON Datos Atípicos

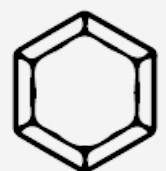


¿Parecen existir consumos excesivos?

Diagramas de Caja

SIN Datos Atípicos



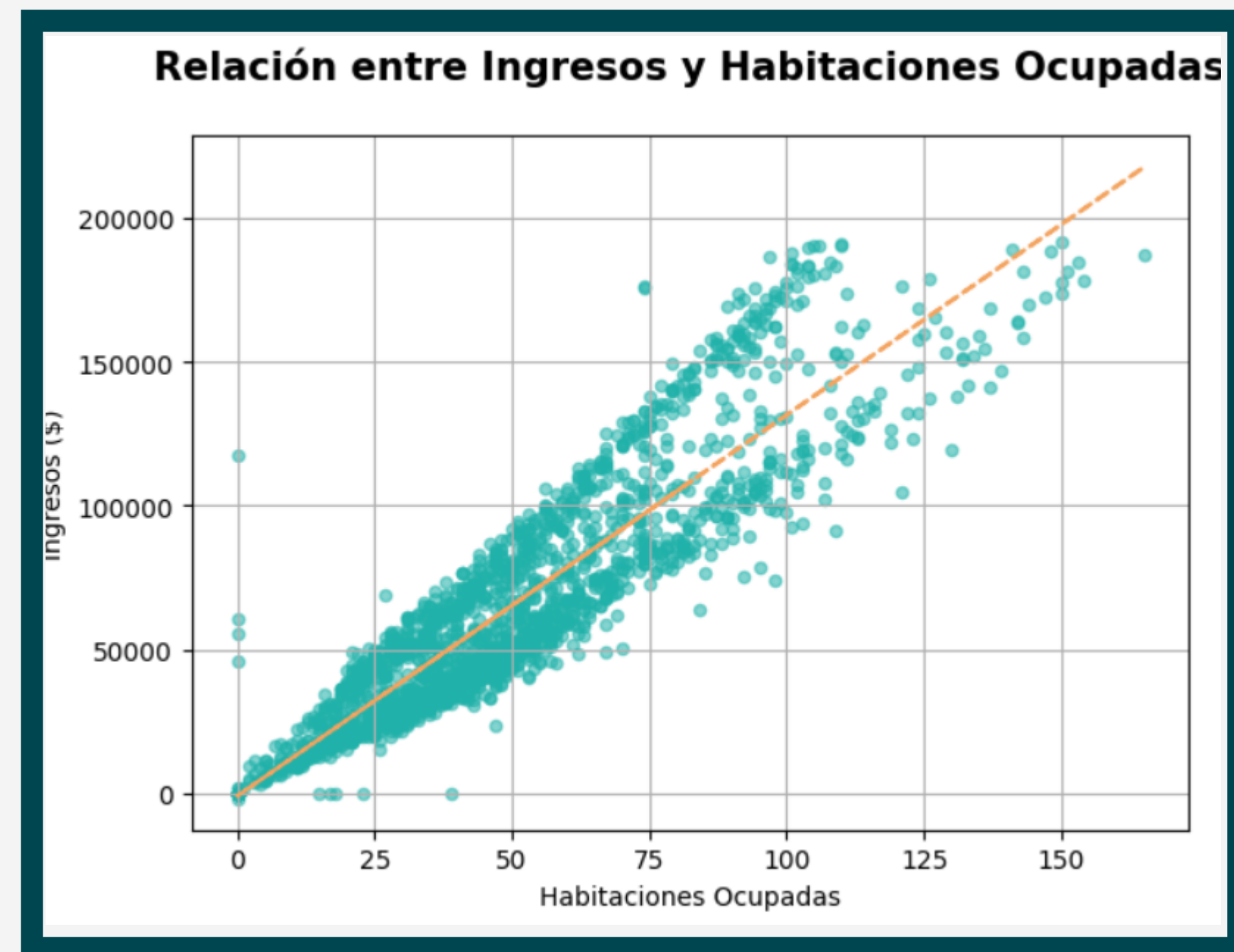
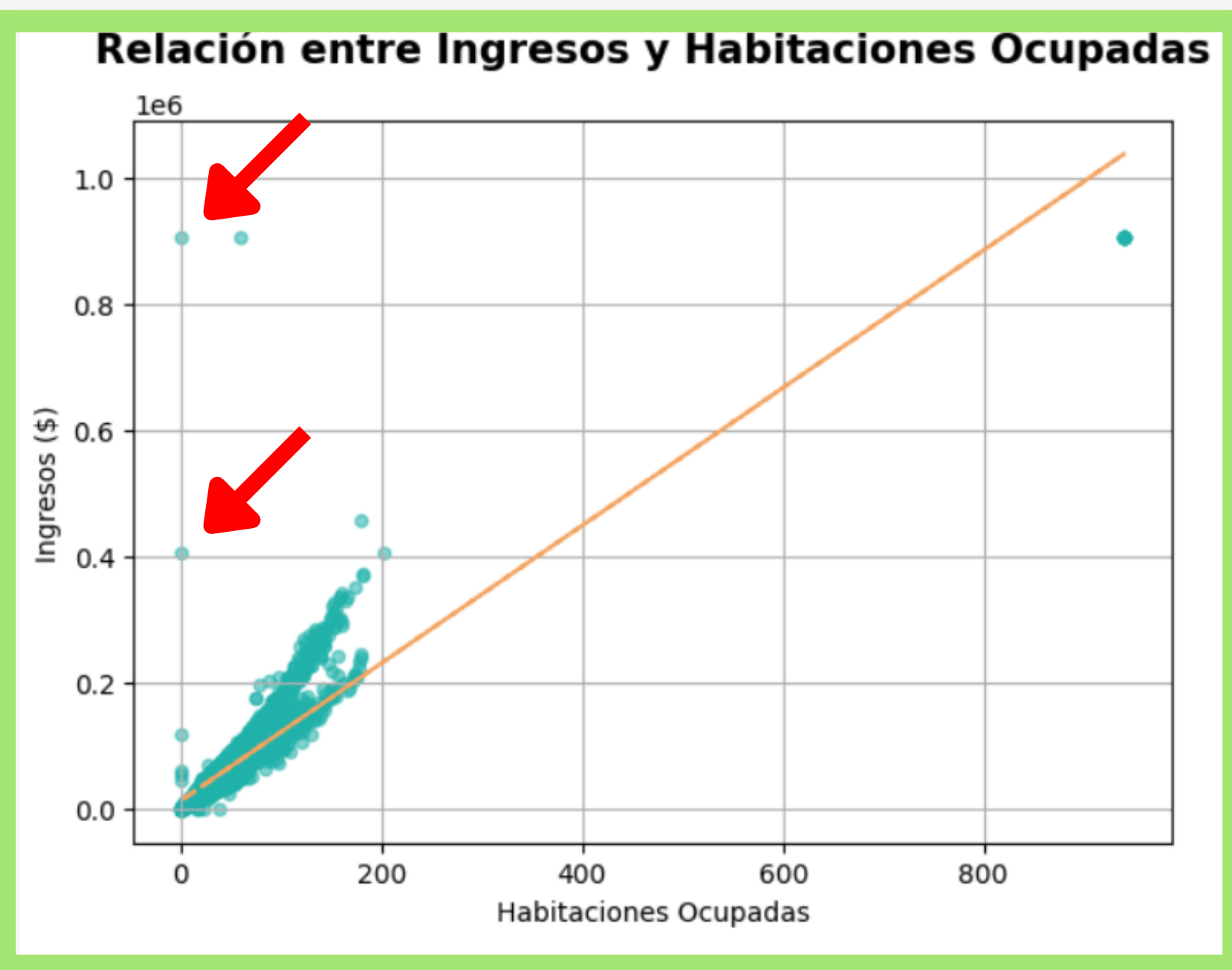


FRACTAL
S O L U T I O N S

Relaciones entre Variables

CON Datos Atípicos

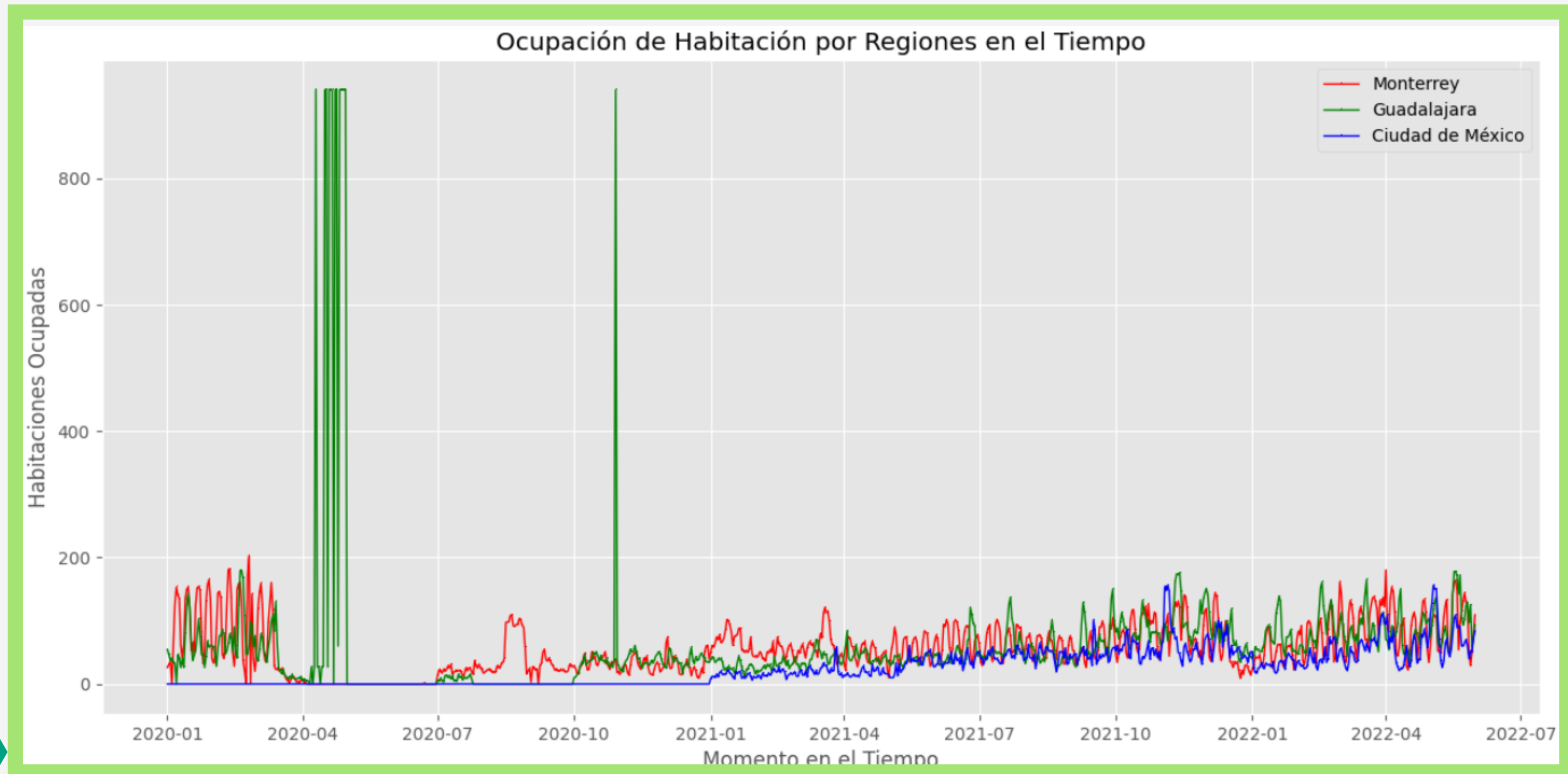
SIN Datos Atípicos



Hay datos que parecen exagerados y poco recurrentes; ¿apoya para un futuro mejor modelo?

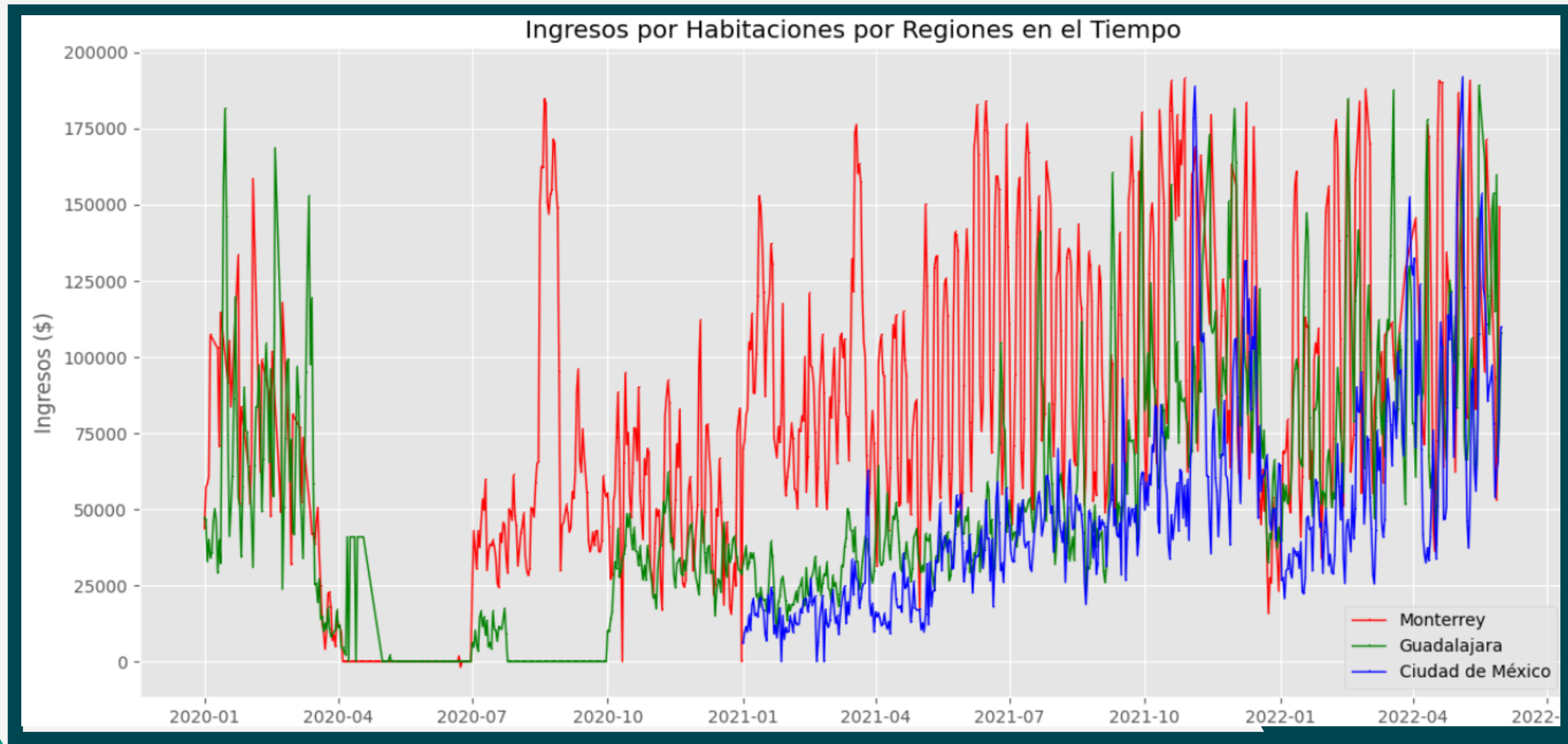
Variables de Interés en el Tiempo

CON Datos Atípicos



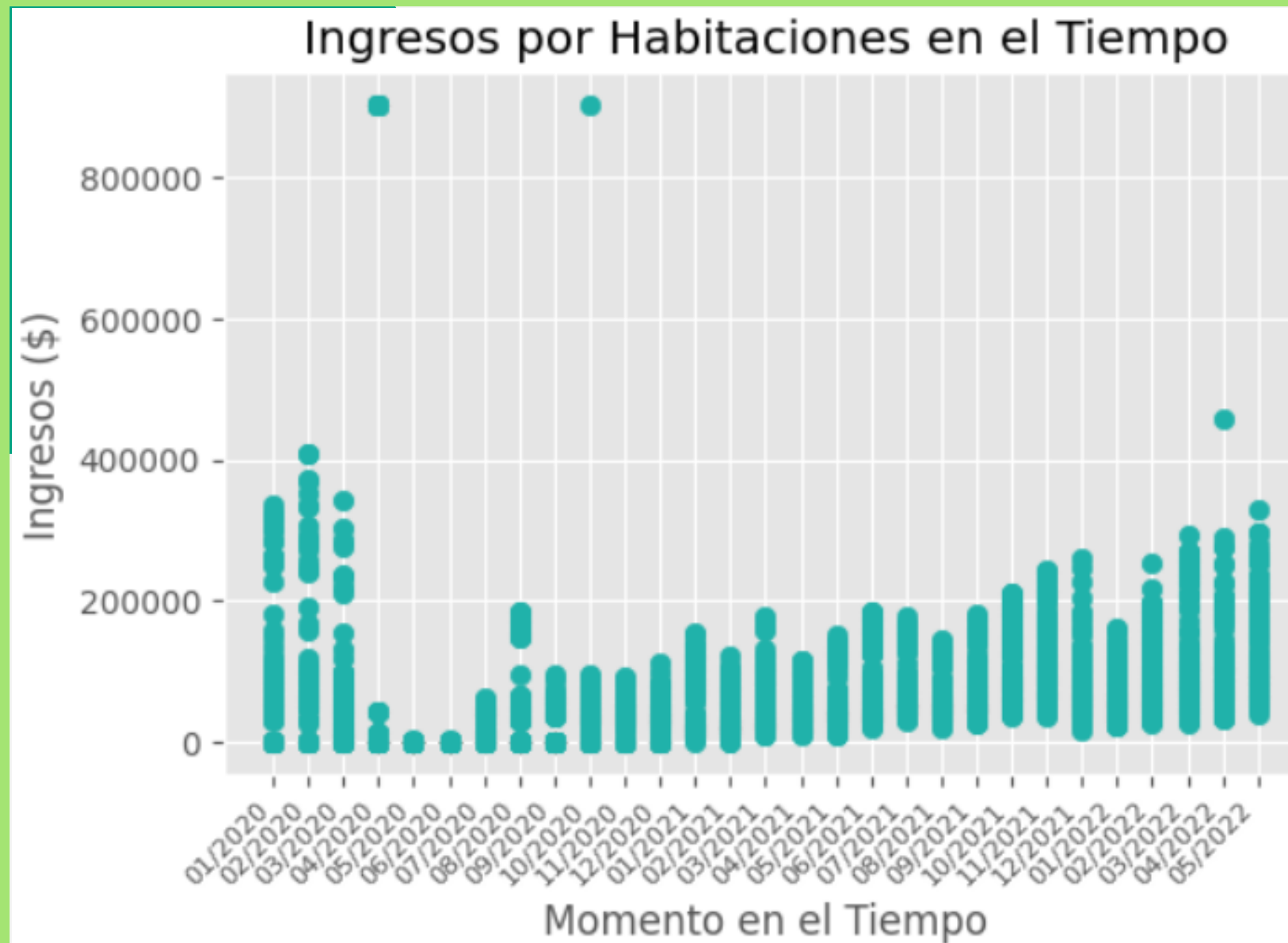
Variables de Interés en el Tiempo

SIN Datos Atípicos

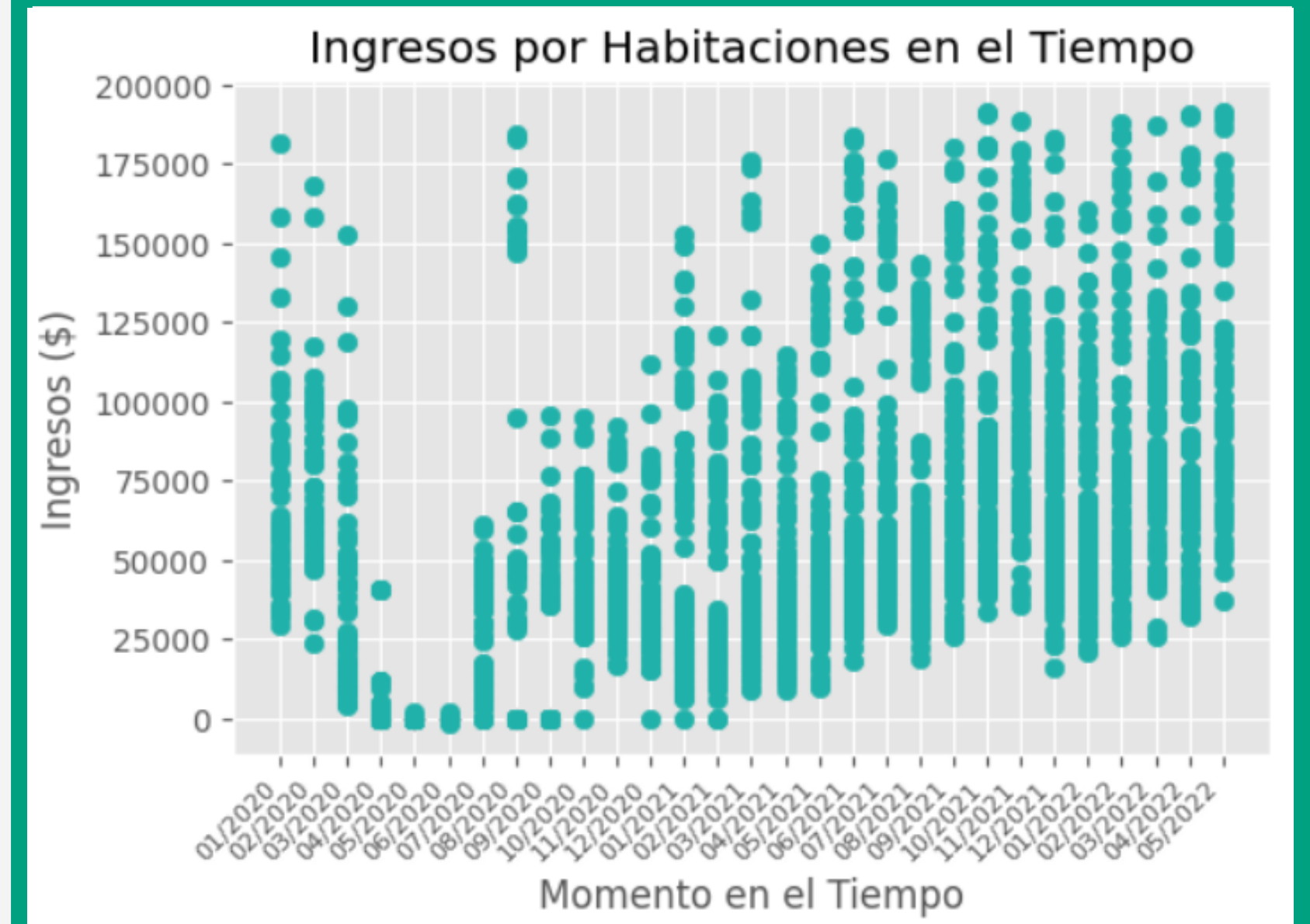


Variables de Interés en el Tiempo

CON Datos Atípicos

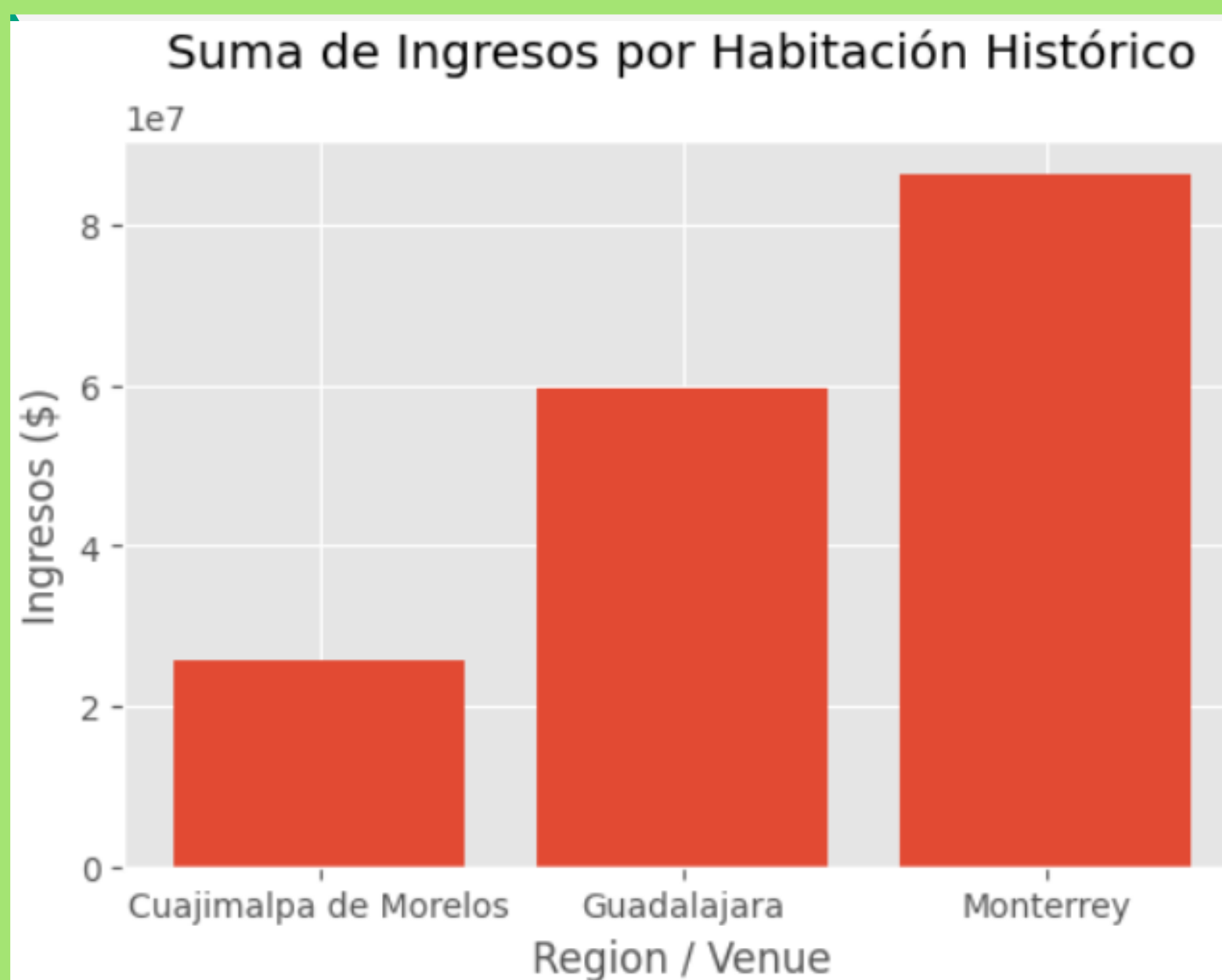


SIN Datos Atípicos

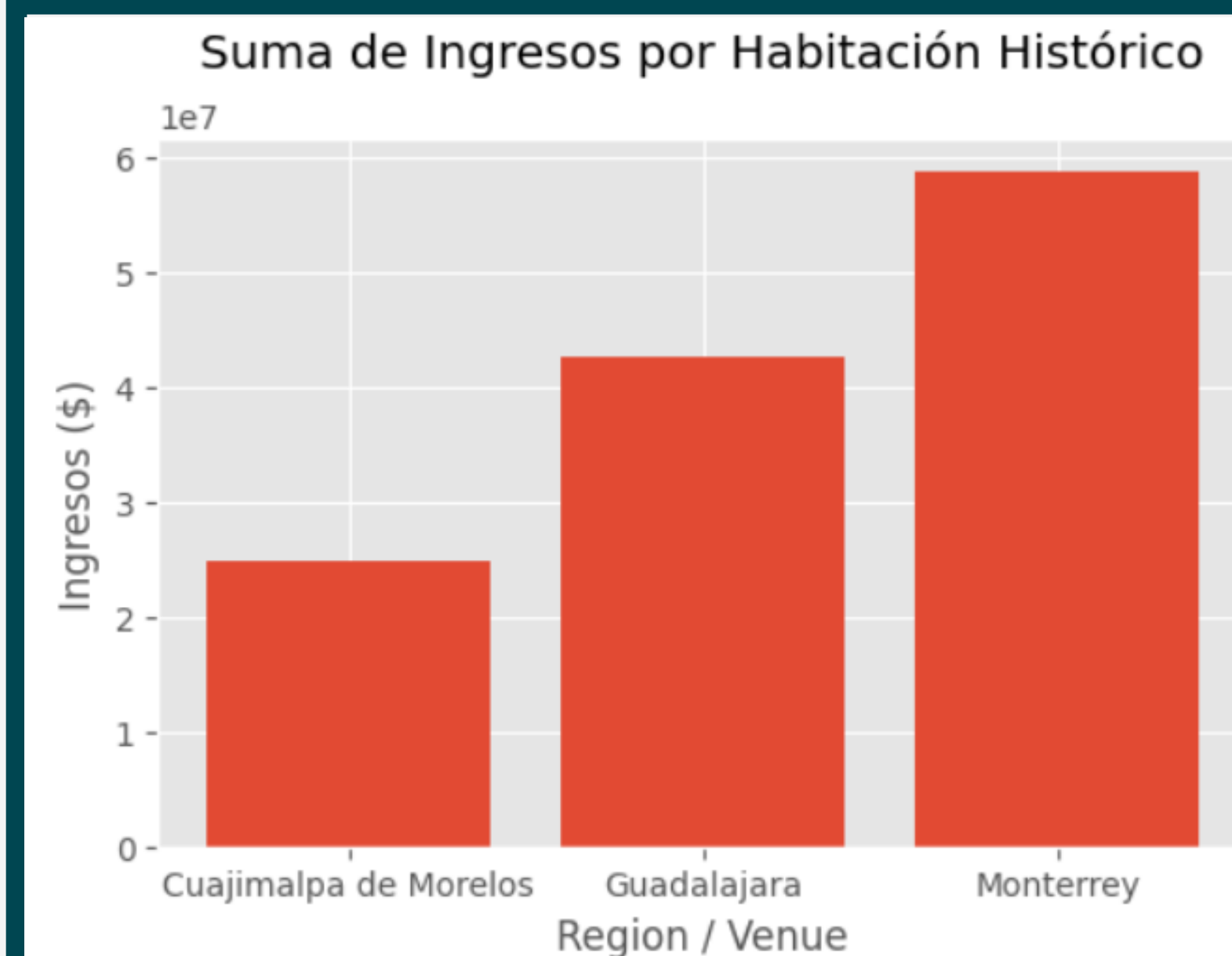


Variables de Interés en el Tiempo

CON Datos Atípicos



SIN Datos Atípicos



Recolección de Variables Externas

Variable Externa a Recolectar

- **Variable:** Eventos Comerciales.
- Complicaciones
 - Demasiada información requerida.
 - Datos históricos sesgados por el COVID-19.
 - Falta de fuente de información macro.
 - Demasiados tipos de eventos para una sola categoría.

Alternativa de Variable

- **Variable:** Eventos musicales
- Razonamiento
 - Fuente de turismo a nivel mundial.
 - No limitada a solo negocios.
 - Facilidad de implementación en el modelo predictivo.
 - Existen API con la información requerida.

Utilización de API

Por medio de una **API** de una aplicación conocida como **songkick**. Podemos obtenerla información de los conciertos o eventos musicales realizados en un rango de fechas específico.

Después se requeriría el **obtener los lugares que coincidan con los hoteles.**

Parameter	Type	Value
apikey	Required	Your API key.
artist_id	Required unless <code>mbid: {music_brainz_id}</code> specified	The ID of the artist to return past events for.
music_brainz_id	Required unless <code>artist_id</code> specified	The MusicBrainz ID of the artist to return past events for.
min_date	Optional	A date in the form YYYY-MM-DD.
max_date	Optional	A date in the form YYYY-MM-DD.
page	Optional	Results are paginated. This specifies the results page number. (First page = 1)
per_page	Optional	The number of results to return in each page. (Max 50)
order	Optional	Results are ordered by date: 'asc' or 'desc'. (Default = 'asc')

Narrow your results

[All results](#)

[Upcoming events](#)

Past events

Search

SEARCH

PAST FESTIVAL

Saturday 10 September 2016

Life in Color - Mexico City 2016

Nicky Romero and Vicetone

Mexico City, Mexico

I WAS THERE

PAST FESTIVAL

Saturday 07 March 2020

Remind GNP - Mexico City 2020

Christopher Cross, Fobia, and Pat Benatar & Neil Giraldo

Mexico City, Mexico

I WAS THERE

Dato de Interés

Formato JSON de la API

```
{
  "resultsPage": {
    "results": {
      "event": [
        {
          "id":11129128,
          "type":"Concert",
          "uri":"http://www.songkick.com/concerts/11129128-wild-flag-at-fillmore?utm_source=PARTNER_ID&utm_medium=organic",
          "displayName":"Wild Flag at The Fillmore (April 18, 2012)",
          "start": {
            "time":"20:00:00",
            "date":"2012-04-18",
            "datetime":"2012-04-18T20:00:00-0800"
          },
          "performance": [
            {
              "artist": {
                "id":29835,
                "uri":"http://www.songkick.com/artists/29835-wild-flag?utm_source=PARTNER_ID&utm_medium=organic",
                "displayName":"Wild Flag",
                "identifier": []
              },
              "id":21579303,
              "displayName":"Wild Flag",
              "billingIndex":1,
              "billing":"headline"
            }
          ]
        }
      ],
    }
  }
}
```

Formato JSON de la API

```
{
  "location": {
    "city": "San Francisco, CA, US",
    "lng": -122.4332937,
    "lat": 37.7842398
  },
  "venue": {
    "id": 6239,
    "displayName": "The Fillmore",
    "uri": "http://www.songkick.com/venues/6239-fillmore?utm_source=PARTNER_ID&utm_medium=p",
    "lng": -122.4332937,
    "lat": 37.7842398,
    "metroArea": {
      "id": 26330,
      "uri": "http://www.songkick.com/metro-areas/26330-us-sf-bay-area?utm_source=PARTNER_I",
      "displayName": "SF Bay Area",
      "country": { "displayName": "US" },
      "state": { "displayName": "CA" }
    }
  },
  "status": "ok",
  "popularity": 0.012763
}, ....
]
},
"totalEntries": 24,
"perPage": 50,
"page": 1,
"status": "ok"
}
```

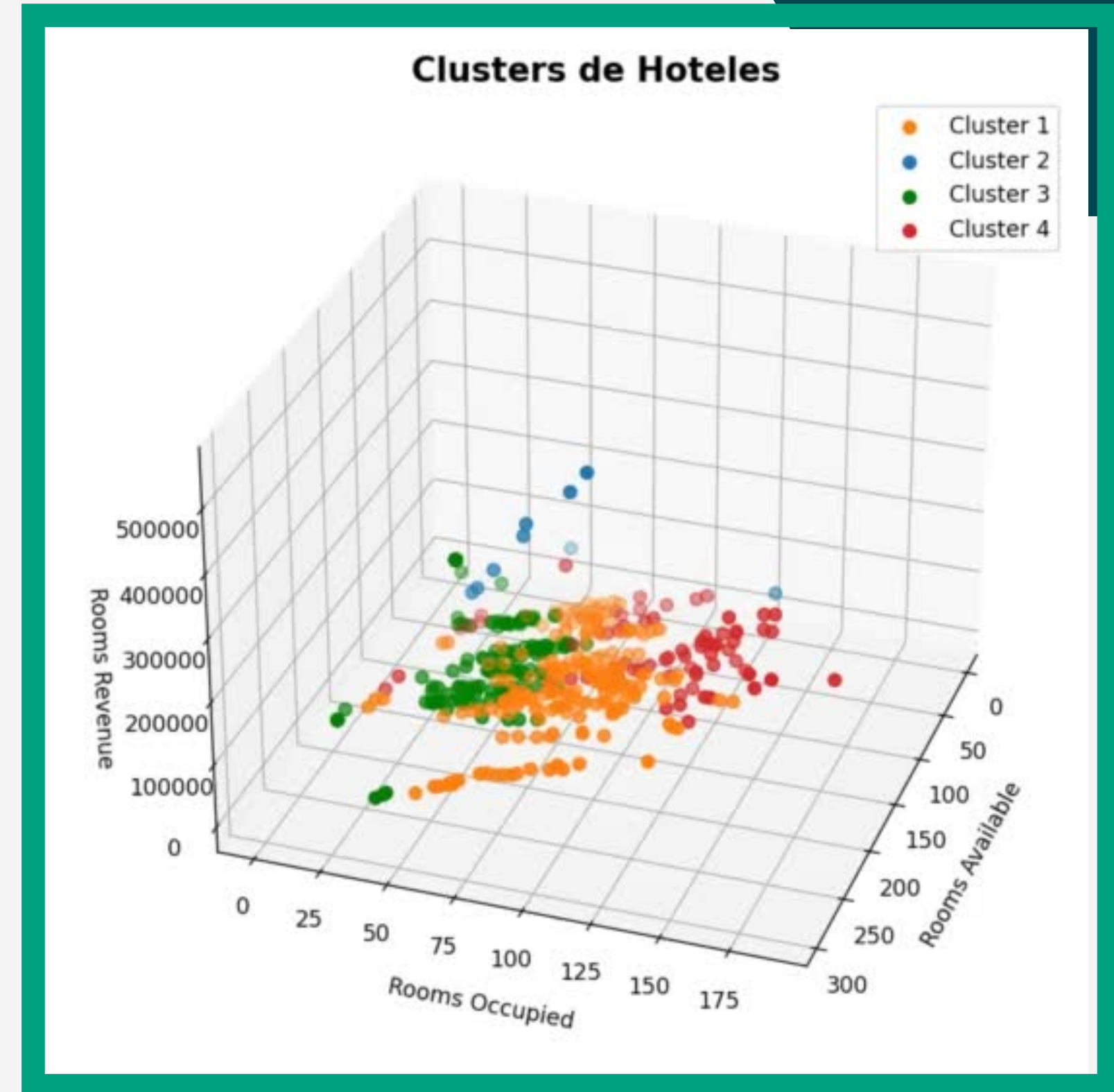
→ Dato de Interés

Propuesta para Modelos

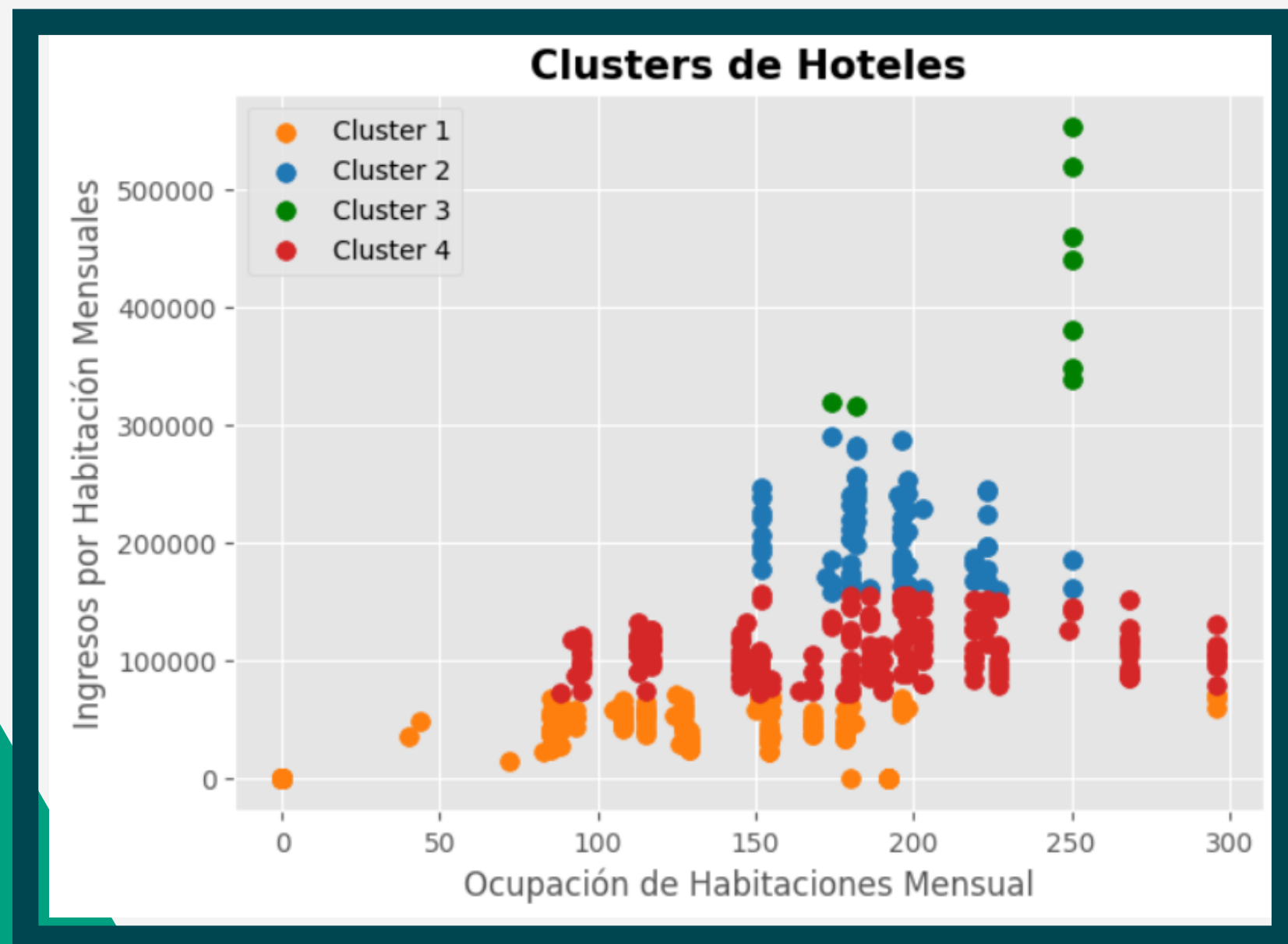
Segmentación de Hoteles

Para hacer más eficiente el uso de los modelos.

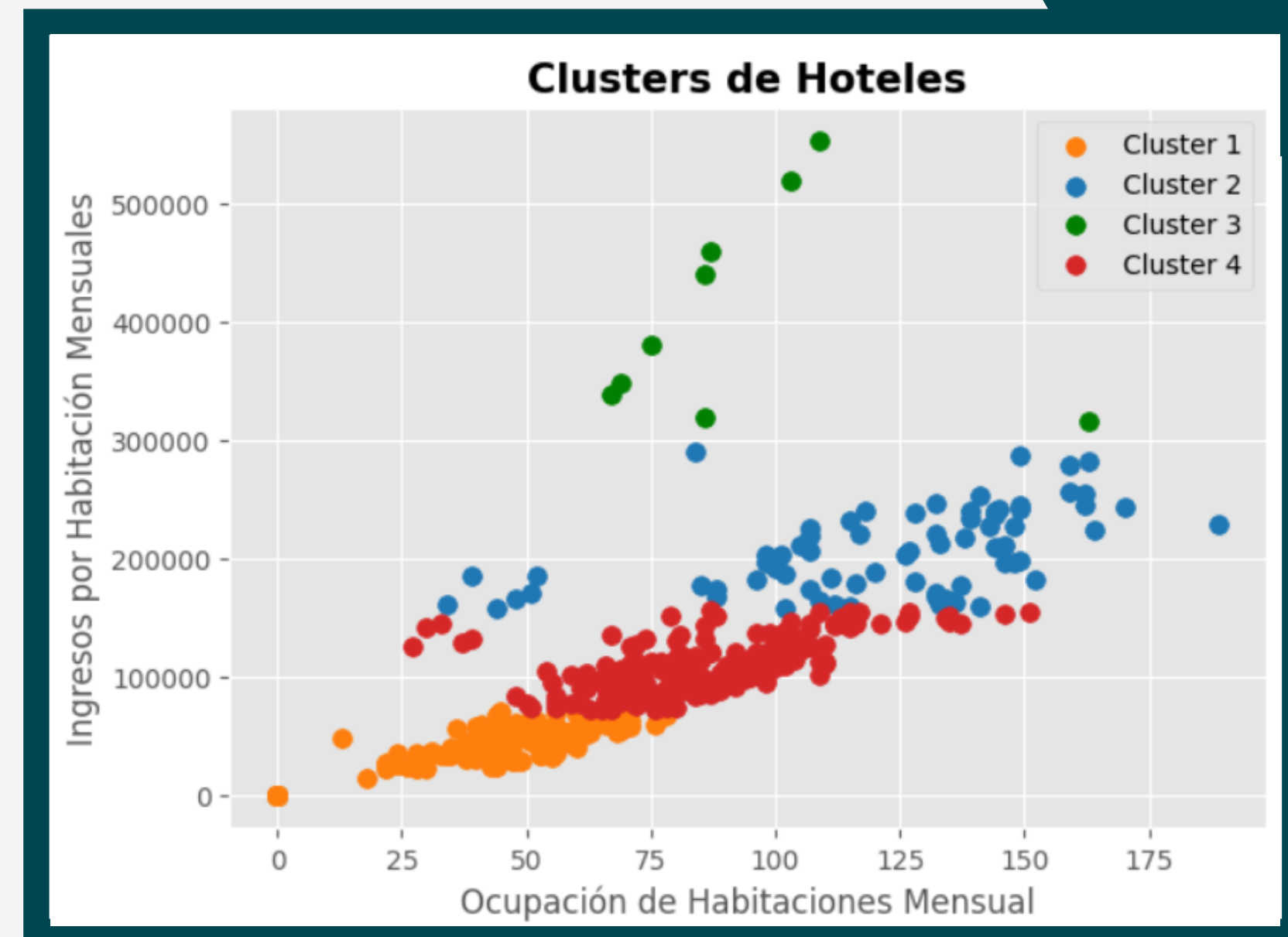
- Agrupar los hoteles con características similares
 - Usar un modelo predictivo por grupo
- Obteniendo los hoteles en México, con base en las habitaciones se generan 4 grupos utilizando K-means clustering



Segmentación de Hoteles



¿ Agrupar por aquellos que tienen mayor ingreso mensual ?



¿ Agrupar hoteles por distintos niveles de ocupación ?

Conclusion

- Pasos a seguir con los datos atípicos?
- Optar por la utilización de modelos en grupos de hoteles
 - Por niveles de ocupación y/o ingreso
 - Por subdivisiones de zonas o entidades

La variable de eventos musicales es una opción más viable para continuar explorando con el fin de aumentar la significancia del modelo predictivo.

Gracias

CONTACTO

+ 52 1 (81) 1077 9523

+ 52 1 (411) 106 6139

fractalsolutions.contacto@gmail.com