

Gestión de propiedades y reservas de apartamentos en un portafolio

Calendar

Muestra la representación detallada de la disponibilidad y el estado de las propiedades (apartamentos) en un determinado período de tiempo. Aquí está una explicación más detallada de cada campo en el bloque "Calendar":

1. **PropertyId (Id de la propiedad)**: Es una identificación única asociada a cada propiedad, que probablemente se refiere a un apartamento específico dentro del portafolio.
2. **CalendarDate (Fecha en el calendario)**: Representa una fecha específica entre el 01/07/2022 y el 31/12/2022.
3. **Blocked (Estado)**:
 - 0 (No reservado): Indica que el apartamento estaba disponible en esa fecha.
 - 1 (Bloqueado): Significa que el apartamento estaba cerrado y no estaba disponible para reservar en esa fecha.
 - 2 (Reservado): Indica que el apartamento estaba ocupado por clientes en esa fecha.
4. **Property_BookingId (Id de reserva de la propiedad)**: Este campo solo está presente si el estado (Blocked) es igual a 2, lo que significa que la propiedad estaba reservada en esa fecha. En este caso, el **Property_BookingId** sería el identificador único asociado a esa reserva específica.

Propiedades (Properties)

Proporciona información estática sobre las propiedades del portafolio, como su identificación única y la capacidad de personas que pueden alojarse en cada apartamento.

1. **PropertyId (Id de la propiedad)**: Es una identificación única asociada a cada propiedad (apartamento) en el portafolio.
2. **PR_Sleeps (Capacidad para dormir)**: Indica la capacidad del apartamento en personas, es decir, cuántas personas pueden dormir en ese apartamento.

Reservas (Bookings)

Proporciona información detallada sobre las reservas realizadas, incluyendo el costo total, duración, fecha de creación y el canal a través del cual se realizó la reserva.

1. **Property_BookingId (Id de reserva de la propiedad)**: Identificación única asociada a una reserva específica.
2. **PB_TotalPrice (Precio total de la reserva)**: Representa el costo total de la reserva.
3. **PB_NumNights (Número de noches de la reserva)**: Indica cuántas noches está programada la reserva.

4. **PB_BookingCreatedDate (Fecha de creación de la reserva)**: Muestra la fecha en que se creó la reserva.
5. **PB_BookingChannelOriginId (Id del canal de origen de la reserva)**: Identificación del canal a través del cual se realizó la reserva.

Visión general de la ocupación mensual

- Calcula la ocupación mensual para cada propiedad en el portafolio. Utiliza los datos del bloque "Calendar" para contar las noches bloqueadas, noches disponibles y las noches reservadas en cada mes.
- La ocupación sera igual :

$$\begin{aligned} \text{Ocupación (\%)} &= (\text{Dias Reservados} / \text{Dias Disponibles}) * 100 \\ &= (\text{Dias Reservados} / ((\text{Date_maxima} - \text{Date_minima}) - \text{Dias} \\ &\quad \text{Bloqueado})) * 100 \end{aligned}$$

Con el fin de abordar esta situación, utilizaremos los datos de 'Calendar'. Al observar la tabla, se nota la ausencia de valores nulos, a excepción de 'Property_BookingId', que puede presentar un valor NaN cuando 'Blocked' es igual a '1' o '0'.

Out[3]:

	Columna	Tipo de Dato	Valores Únicos	Total Valores	Valores Nulos	Porcentaje Nulos
0	PropertyId	int64	223	38893	0	0.000000
1	CalendarDate	datetime64[ns]	184	38893	0	0.000000
2	Blocked	int64	3	38893	0	0.000000
3	Property_BookingId	float64	6160	31168	7725	19.862186

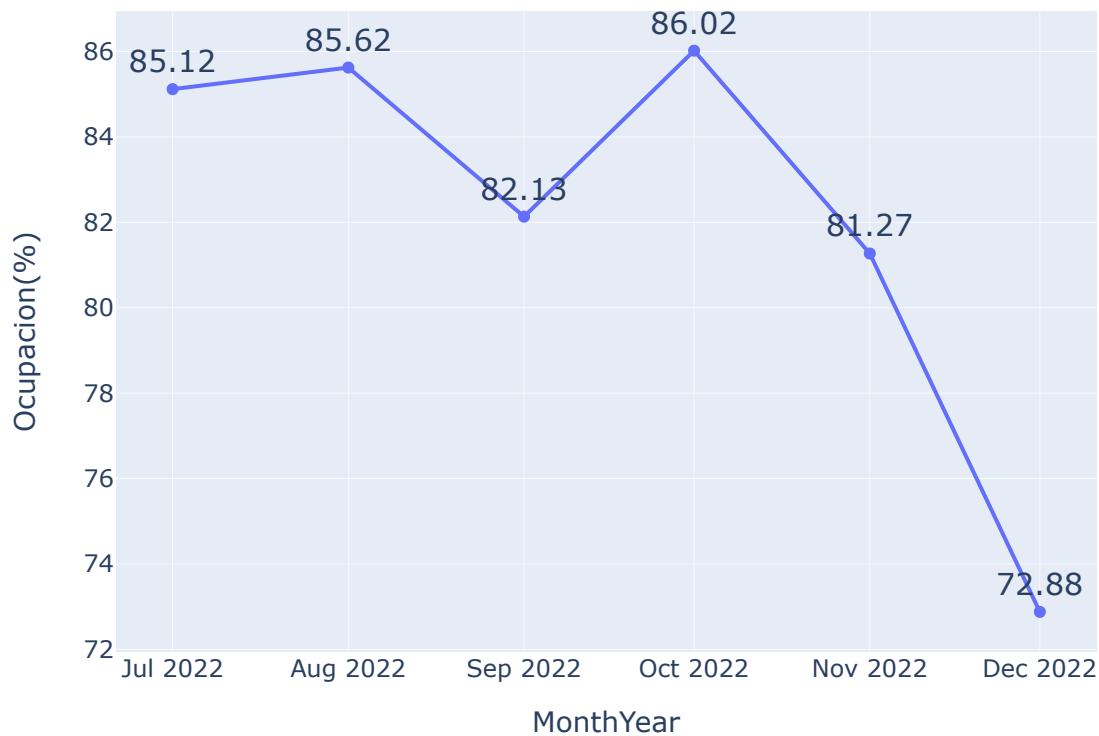
Podemos calcular el porcentaje de ocupación mensual agrupando los datos por mes y utilizando la función pivot en la columna "Blocked". Con este enfoque, logramos obtener la información necesaria para calcular el porcentaje de ocupación.

Out[4]:

	Blocked	0	1	2	suma	ocupacion(%)
MonthYear						
2022-07	921	63	5268	6252	85.118759	
2022-08	916	95	5456	6467	85.624608	
2022-09	1129	71	5190	6390	82.133249	
2022-10	908	83	5587	6578	86.020015	
2022-11	1148	330	4981	6459	81.269375	
2022-12	1744	317	4686	6747	72.877138	

Se puede apreciar una disminución en el porcentaje de ocupación durante los meses de noviembre y diciembre, lo cual se representa de manera más clara en la siguiente gráfica.

Ocupacion(%) del portafolio



Ocupacion por propiedad(%)

Los datos comprenden el período desde el 31 de diciembre de 2022 a las 00:00:00 hasta el 1 de julio de 2022 a las 00:00:00, lo que equivale a un total de 184 días.

Podemos utilizar **'PropertyId'** y, a través del indicador **'Blocked'**, calcular el número de días que la propiedad ha pasado en cada uno de los tres estados.

Se obtiene un total de 223 propiedades diferentes, junto con la cantidad de días que han pasado en cada estado, así como la 'Suma'

Out[7]:

Blocked	0	1	2	Suma
---------	---	---	---	------

PropertyId

2883	23	1	160	184
3963	35	1	148	184
3964	17	1	166	184
4138	21	7	156	184
4181	22	1	161	184

Al aplicar un filtro, podemos identificar 29 propiedades con una suma de días que difiere de 184. Esto podría deberse a la falta de registros para estas propiedades o a que ingresaron o quedaron fuera durante el período en el que se recopilaron los datos.

Out[8]:

Blocked	0	1	2	Suma
---------	---	---	---	------

PropertyId

43330	44	1	136	181
43332	16	3	154	173
43333	23	1	149	173
43334	19	2	150	171
43313	18	5	143	166
43337	24	2	138	164
43340	31	2	124	157
43338	45	11	97	153
43344	60	7	70	137
43347	34	6	97	137

También es posible considerar la fecha máxima y la fecha mínima para cada propiedad, y a partir de estas fechas, calcular el número de días disponibles para cada propiedad. Si hay concordancia en los registros, es muy probable que algunas propiedades hayan quedado fuera o ingresado durante el período en el que se recopilaron los datos y que no sea un error en los datos o datos faltantes.

Out[9]:

	PropertyId	CalendarDate	CalendarDate	MaxCalendarDate	MinCalendarDate	DiasDiferencia
--	------------	--------------	--------------	-----------------	-----------------	----------------

0	42997	2022-07-01	2022-07-01	2022-12-31	2022-07-01	184
1	42997	2022-07-02	2022-07-02	2022-12-31	2022-07-01	184
2	42997	2022-07-03	2022-07-03	2022-12-31	2022-07-01	184
3	42997	2022-07-04	2022-07-04	2022-12-31	2022-07-01	184
4	42997	2022-07-05	2022-07-05	2022-12-31	2022-07-01	184

Filtramos los datos para cuando 'DiasDiferencia' es diferente a 184 para comparar con los datos anteriores

Out[10]:

	PropertyId	DiasDiferencia
35403	43330	181
35584	43332	173
35757	43333	173
35930	43334	171
34317	43313	166
36101	43337	164
36532	43340	157
36265	43338	153
36876	43344	137
37287	43347	137

Las tablas coinciden; podemos asumir que algunas propiedades quedaron fuera o ingresaron durante el período en el que se recopilaron los datos. Con esta información, podemos calcular el porcentaje de ocupación para cada propiedad tomando en cuenta el tiempo que la propiedad formo parte del portafolio.

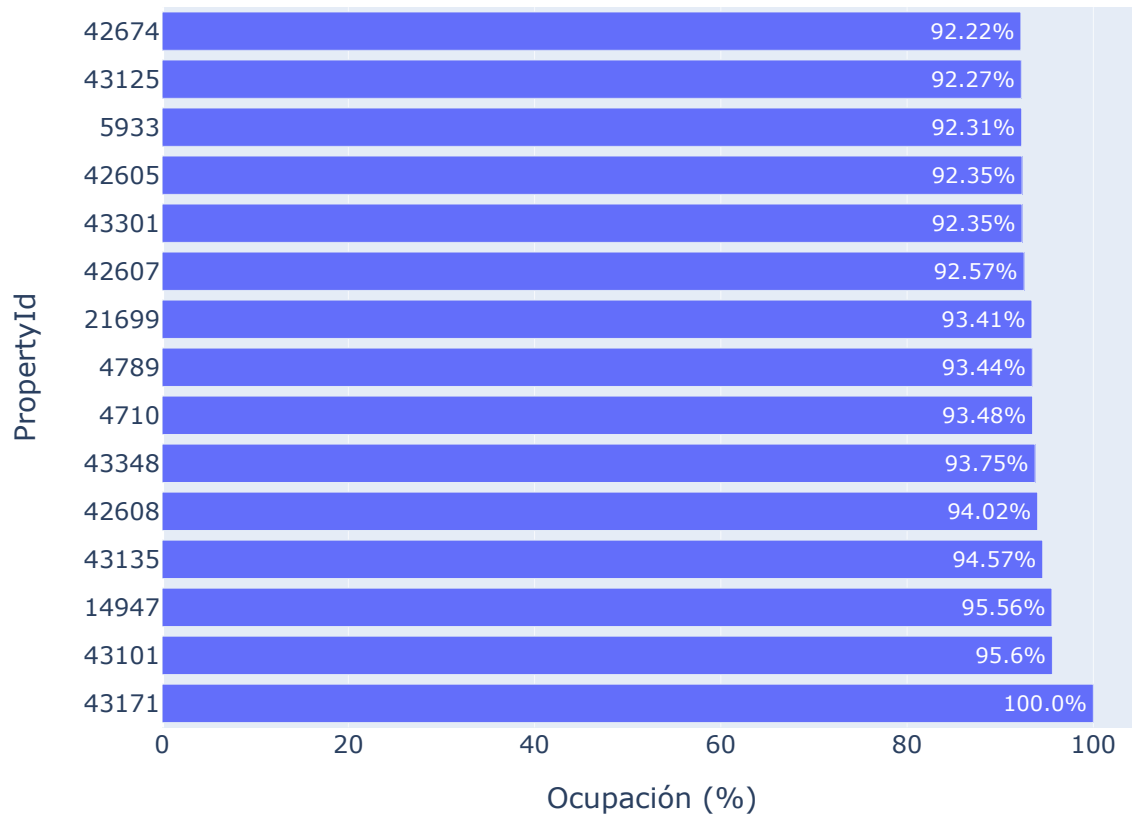
Ocupación (%) = (Dias Reservados / Dias Disponibles) * 100

Out[11]:

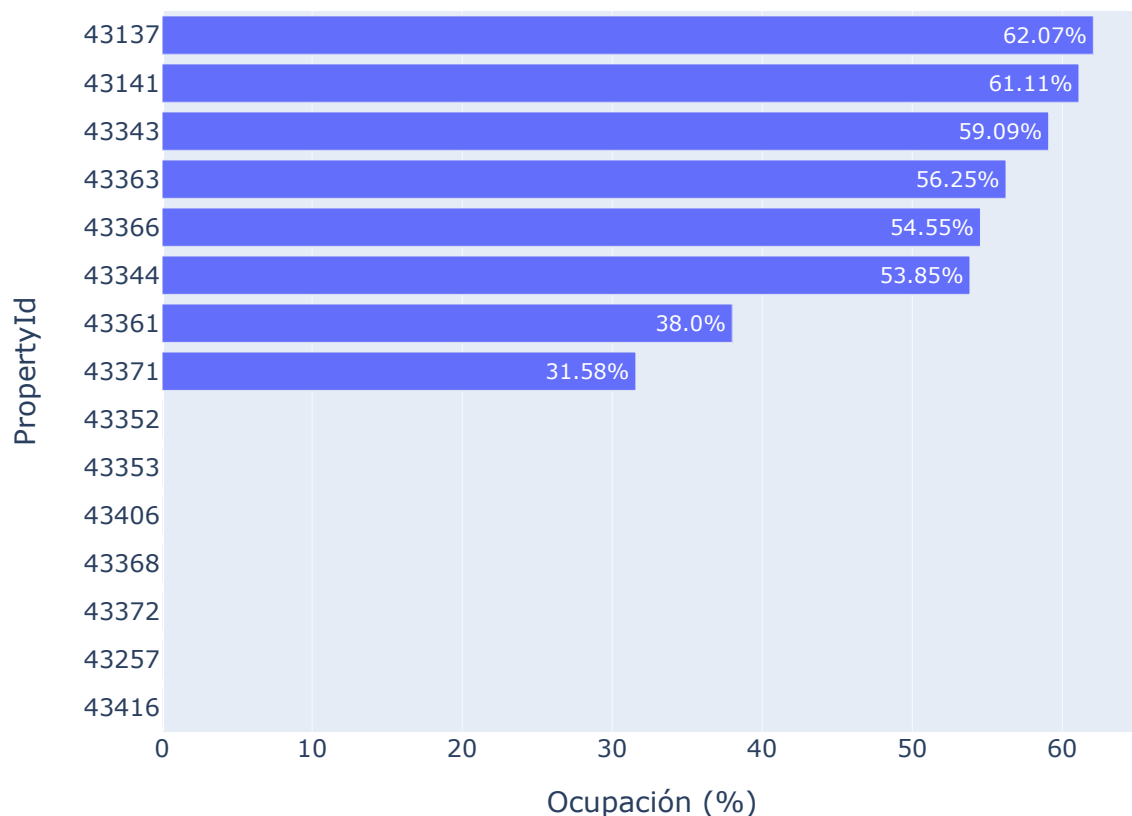
Blocked	0	1	2	Suma	ocupacion(%)
PropertyId					
2883	23	1	160	184	87.431694
3963	35	1	148	184	80.874317
3964	17	1	166	184	90.710383
4138	21	7	156	184	88.135593
4181	22	1	161	184	87.978142
4259	31	2	151	184	82.967033
4710	12	0	172	184	93.478261
4785	19	1	164	184	89.617486
4786	16	3	165	184	91.160221
4787	23	2	159	184	87.362637
4788	30	8	146	184	82.954545
4789	12	1	171	184	93.442623
4790	33	2	149	184	81.868132
4869	18	6	160	184	89.887640
4928	24	0	160	184	86.956522

Con los datos obtenidos, podemos representar gráficamente las propiedades que tienen el porcentaje de ocupación más alto y más bajo.

Top 15 PropertyId por Ocupación



Bottom 15 PropertyId por Ocupación



La gráfica exhibe **9 propiedades** que no alcanzaron el **50% de ocupación** durante los **6 meses** analizados, de las cuales 7 permanecieron desocupadas en todo ese período. Es necesario investigar las razones detrás de esta baja ocupación.

ADR (Tarifa Diaria Promedio)

Para calcular el ADR (Tarifa Diaria Promedio), utilizaré los datos de '**Bookings**', a los cuales agregaremos la variable '**PR_Sleeps**' obtenida de '**Properties**'.

Out[17]:

	Property_BookingId	PB_TotalPrice	PB_NumNights	PB_BookingCreatedDate	PB_BookingChannelOrig
0	103656	492.71	2	2022-06-28 20:10:07	
1	102879	2130.51	11	2022-06-13 14:27:03	
2	104627	348.10	2	2022-07-13 18:44:25	
3	104725	659.71	4	2022-07-15 09:59:53	
4	103466	618.51	3	2022-06-25 15:43:49	

Resumen de los datos obtenidos

Out[18]:

	Columna	Tipo de Dato	Valores Únicos	Total Valores	Valores Nulos	Porcentaje Nulos
0	Property_BookingId	int64	6160	6160	0	0.0
1	PB_TotalPrice	float64	5862	6160	0	0.0
2	PB_NumNights	int64	63	6160	0	0.0
3	PB_BookingCreatedDate	datetime64[ns]	6159	6160	0	0.0
4	PB_BookingChannelOriginId	int64	13	6160	0	0.0
5	ADR	float64	5994	6160	0	0.0
6	PropertyId	int64	216	6160	0	0.0
7	PR_Sleeps	int64	9	6160	0	0.0

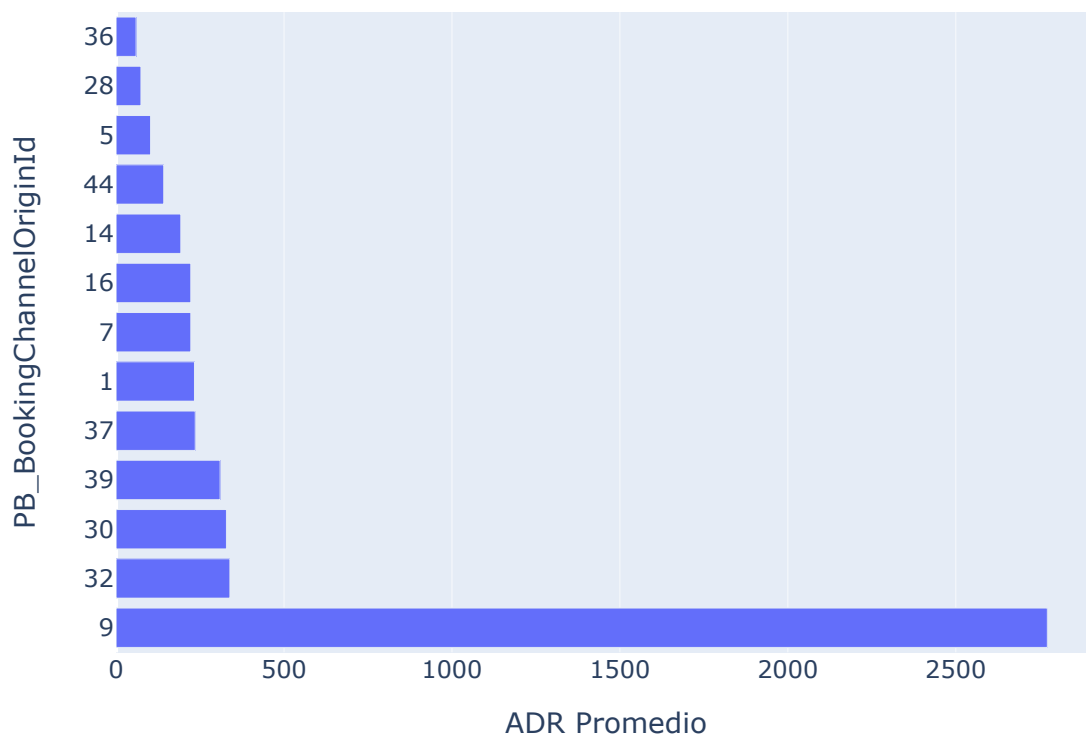
Es posible calcular el **ADR** (Tarifa Diaria Promedio) según el canal agrupando los datos en base a dicho canal y calculando el promedio. Esto nos permitirá obtener una visión segmentada del desempeño del **ADR** en función de los diferentes canales utilizados.

Out[19]:

	PB_BookingChannelOriginId	ADR
3	9	2772.568532
8	32	338.859123
7	30	328.518720
11	39	309.778073
10	37	235.676954
0	1	233.339862
2	7	222.429914
5	16	222.324951
4	14	192.617544
12	44	141.787248
1	5	102.695994
6	28	73.995265
9	36	59.615385

El ADR del **canal 9** sobresale notablemente, como se evidencia en la siguiente gráfica. Es necesario verificar la precisión de los datos y determinar si este resultado excepcional se debe a la exactitud de la información o a algún error en la recopilación de datos.

ADR Promedio por Canal de Reserva



Los datos indican un alquiler de una propiedad por un monto de **1,809,468.00**, lo cual no se corresponde con las demás cifras registradas para el alquiler de esa misma propiedad.

Out[21]:

	Property_BookingId	PropertyId	PB_TotalPrice	PB_NumNights	PB_BookingCreatedDate
1585	108110	42684	1809468.00	6	2022-09-15 19:55:33
4194	104425	43168	1144.63	1	2022-07-10 18:34:59
4895	103373	43230	2890.76	4	2022-06-23 16:40:37
1755	104261	42892	3559.61	6	2022-07-07 16:02:47
5480	108570	43302	562.88	1	2022-09-28 15:09:56

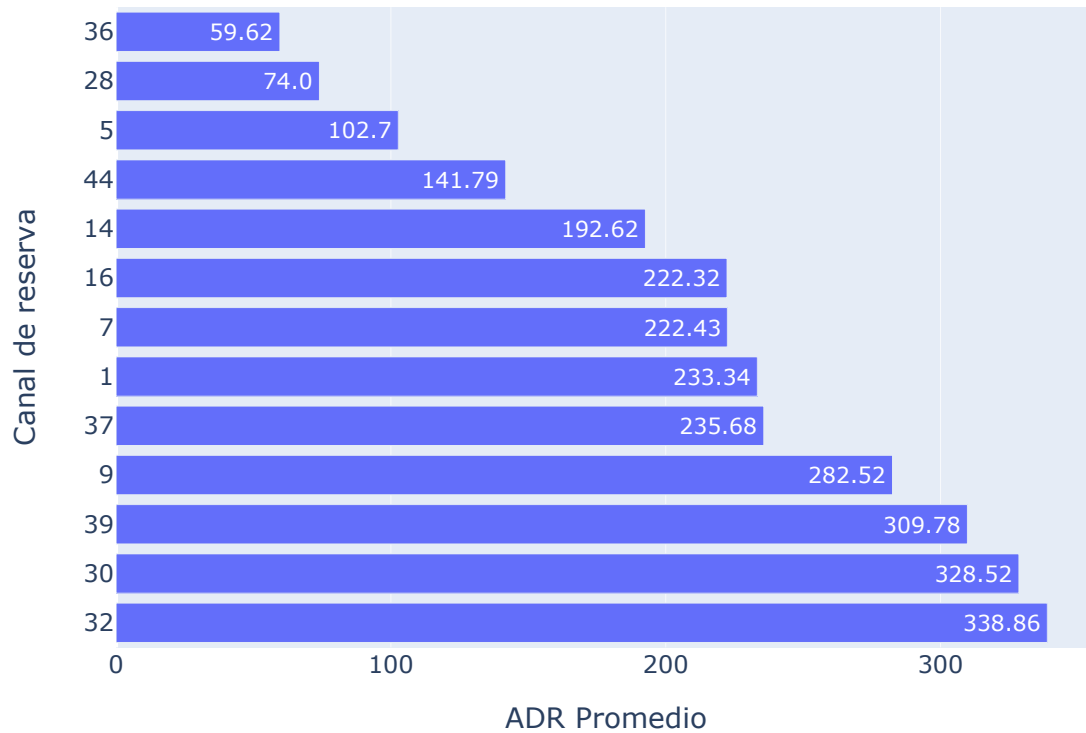
Es necesario eliminar este dato ya que parece ser un error, ya que el monto de alquiler de **1,809,468.00** no concuerda con las demás cifras registradas para el alquiler de esa misma propiedad.

Out[22]:

	Property_BookingId	PropertyId	PB_TotalPrice	PB_NumNights	PB_BookingCreatedDate
1585	108110	42684	1809468.00	6	2022-09-15 19:55:33
1564	102219	42684	1806.50	4	2022-05-31 21:23:57
1575	105824	42684	1559.26	4	2022-08-04 04:27:42
1572	103338	42684	1890.00	5	2022-06-22 22:11:53
1569	104774	42684	1105.26	3	2022-07-15 22:31:26

Una vez eliminados los datos incorrectos, podemos representar gráficamente el ADR en función del canal.

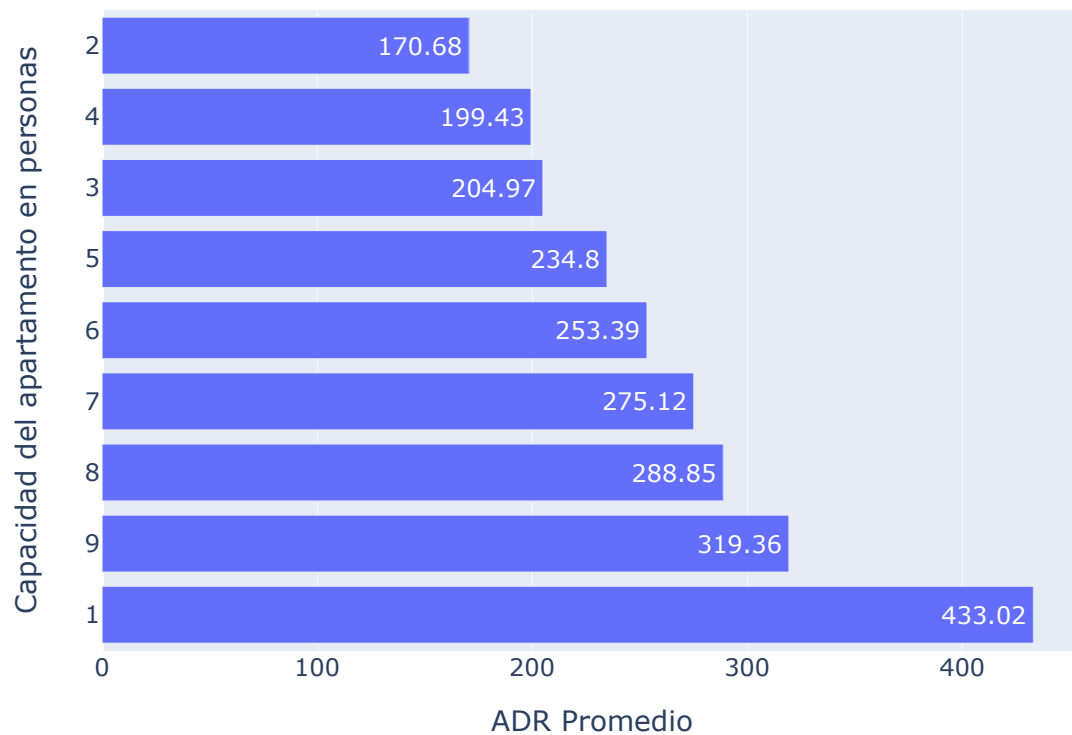
ADR Promedio por Canal de Reserva



Se observa que los canales **(32, 30, 39 y 9)** tienen el ADR más alto, mientras que los canales **(36, 28, 5 y 44)** muestran un ADR significativamente más bajo.

También podemos visualizar gráficamente el **ADR** en relación con la capacidad de personas que pueden hospedarse en la propiedad.

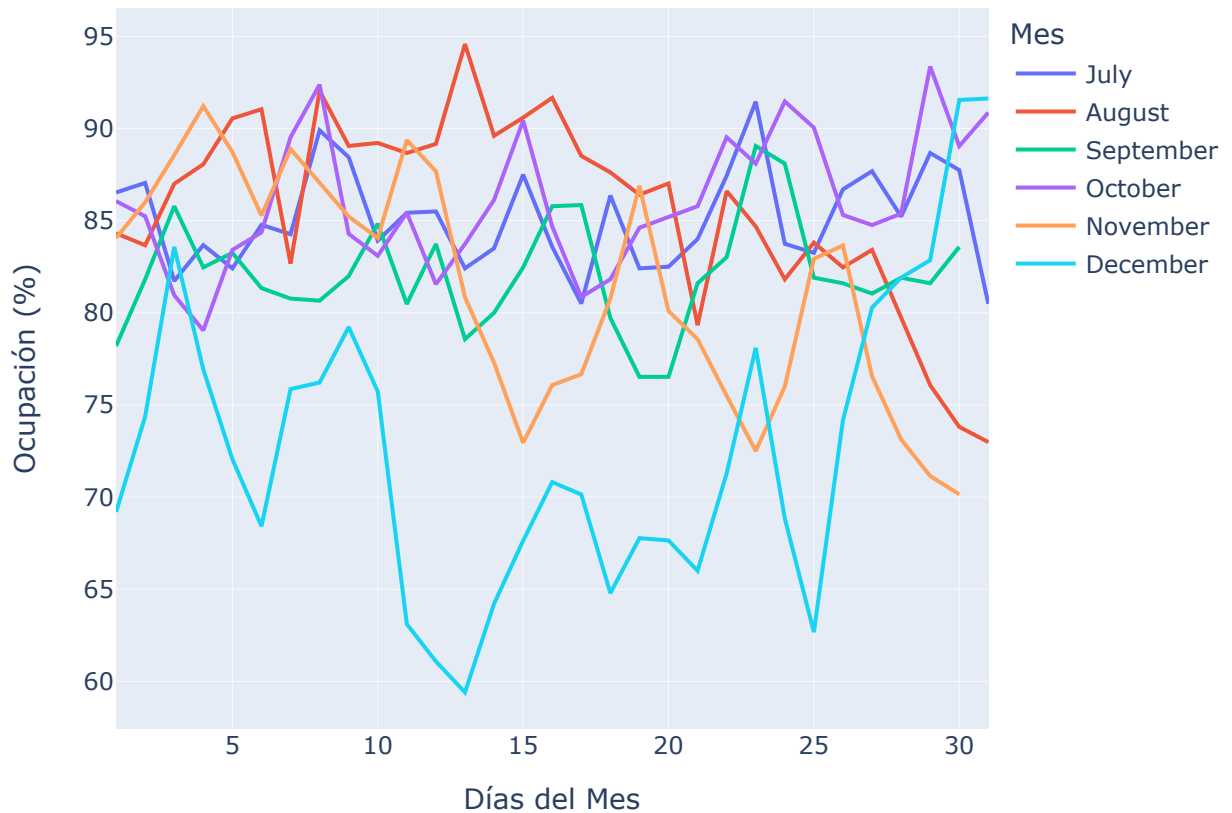
ADR Promedio por Capacidad del apartamento en personas



Comparación del porcentaje de ocupación mes a mes.

Desde '**calendario**', extraeremos la fecha '**CalendarDate**' y la información correspondiente al estado de la propiedad '**Blocked**'. Estos datos se agruparán por día para calcular el porcentaje de ocupación. Una vez obtenidos estos datos, procederemos a representar gráficamente la evolución del porcentaje de ocupación mes a mes.

Ocupación Mensual

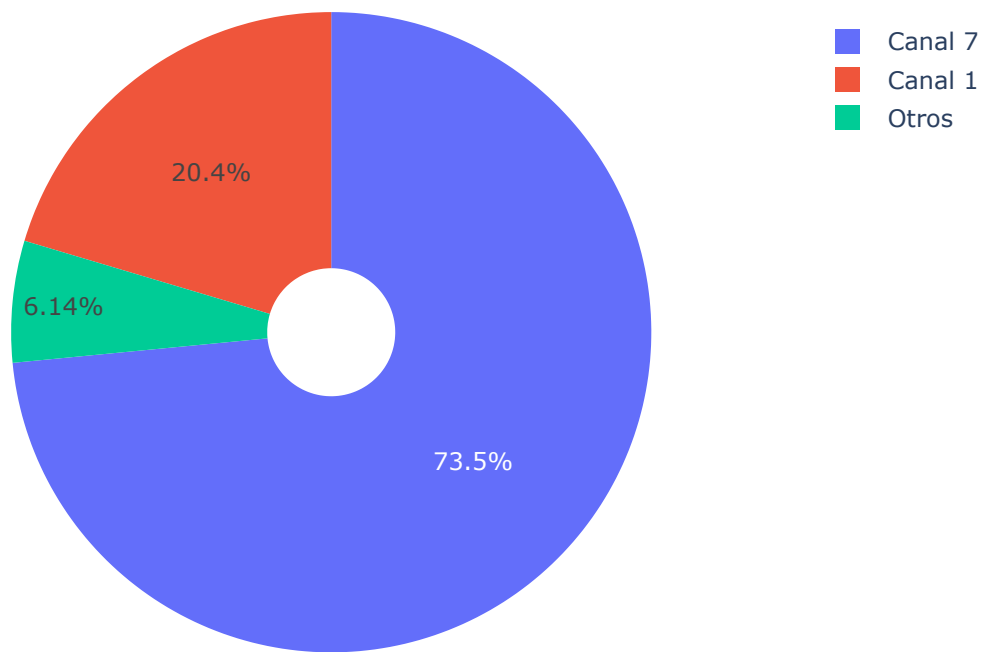


Se observa una disminución en el porcentaje de ocupación que comienza en noviembre y continúa en descenso hacia diciembre, aunque logra recuperarse a finales de dicho mes.

Análisis de Reservas

Los datos indican que el **94%** de las reservas se realizan a través de los canales (**'7' y '1'**), mientras que los otros **11 canales** representan solo el **6%** de las reservas. Podemos ver esto en la siguiente gráfica

Distribución de Reservas por Canal



Puedes encontrar el código en formato de cuaderno Jupyter (IPython Notebook) en el siguiente enlace:
<https://github.com/davidsosaolea/Stay-Unique>.