



Pontificia Universidad Javeriana  
Facultad de Ciencias Económicas y Administrativas  
Analítica de datos

**Factores que afectan la cantidad de check-ins diarios en los hoteles Caesars de Las Vegas**

Valeria Garcia Torres  
Valery Ramírez Mejía

Juan Nicolas Velasquez Rey  
Bogotá D.C, Colombia  
14 de Noviembre de 2025

## Índice

<b>Introducción.....</b>	<b>3</b>
<b>Desarrollo.....</b>	<b>3</b>
I. Descripción de datos.....	3
II. Modelos.....	5
1. MODELO DE REGRESIÓN LINEAL SIMPLE.....	5
2. MODELO DE REGRESIÓN LINEAL MÚLTIPLE.....	6
3. MODELO NO LINEAL.....	8
III. Comparación entre modelos con la matriz de confusión de cada uno.....	10
<b>Conclusión.....</b>	<b>11</b>

## **Introducción**

El caso Caesars Entertainment plantea un problema central en la gestión hotelera: ¿cómo predecir cuántos check-ins habrá cada día para asignar correctamente el personal de recepción y equilibrar servicio y costos? Cuando hay poco personal, las filas son largas y la experiencia del cliente empeora; cuando hay demasiado, aumentan los costos innecesarios. Para responder a esta pregunta, Caesars recopila datos históricos de check-ins diarios, precios de habitaciones, tipos de clientes, eventos especiales y variables de temporada.

El objetivo general es analizar los factores que influyen en las variaciones diarias de los check-ins en los hoteles de Caesars Las Vegas, con el fin de planificar de manera eficiente el personal en recepción. Los objetivos específicos incluyen: (1) Determinar si los días festivos, los tipos de clientes o las variaciones en el precio promedio tienen un impacto significativo en la demanda; (2) Desarrollar un modelo estadístico que explique y prediga las fluctuaciones en el número de check-ins diarios.; (3) Evaluar la precisión del modelo mediante métricas de desempeño, como la matriz de confusión, para medir su capacidad predictiva. Este análisis es importante porque permite tomar decisiones basadas en evidencia para equilibrar servicio y costos. Asimismo, tiene el beneficio de reducir tiempos de espera, mejora la experiencia del huésped, evita gastos innecesarios en personal y aumenta la rentabilidad al alinear la operación con la demanda real.

## **Desarrollo**

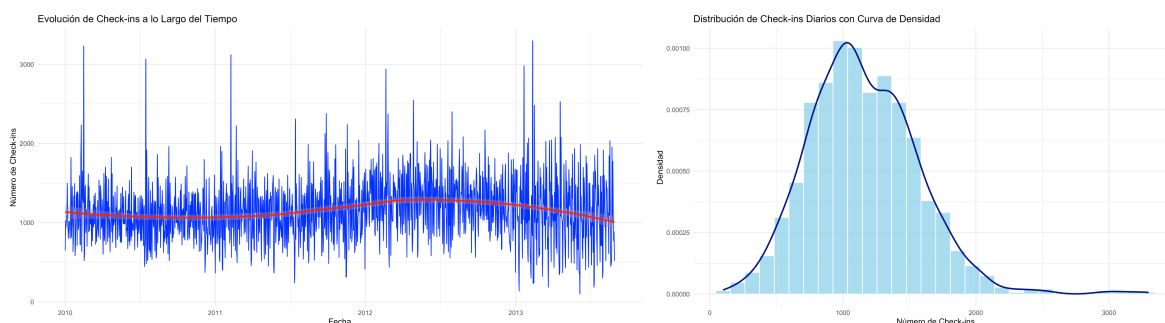
### **I. Descripción de datos**

Para empezar, es importante entender cómo se comportan los check-ins diarios a lo largo del tiempo y qué tan variables son. En la Tabla 1 se evidencia que, aunque el promedio es de 1154 check-ins por día, existe una gran variación entre el mínimo (98) y el máximo (3298), junto con una desviación estándar alta, lo que confirma que la demanda cambia mucho dependiendo de la fecha. Esto se refleja de manera visual en la Gráfica 1, donde se observa una serie de tiempo con picos muy marcados y fluctuaciones constantes, lo que indica que algunos días hay un aumento excepcional de huéspedes, mientras que en otros la ocupación cae de forma considerable. Por otro lado, en la Gráfica 2, donde se muestra el histograma con la curva de densidad, se observa que la cola derecha es más larga, lo que indica un sesgo positivo: es decir, la mayoría de los días se concentran alrededor del nivel promedio, pero existen varios días con cantidades de check-ins excepcionalmente altas (outliers).

Tabla 1. Resumen estadístico de los check-ins

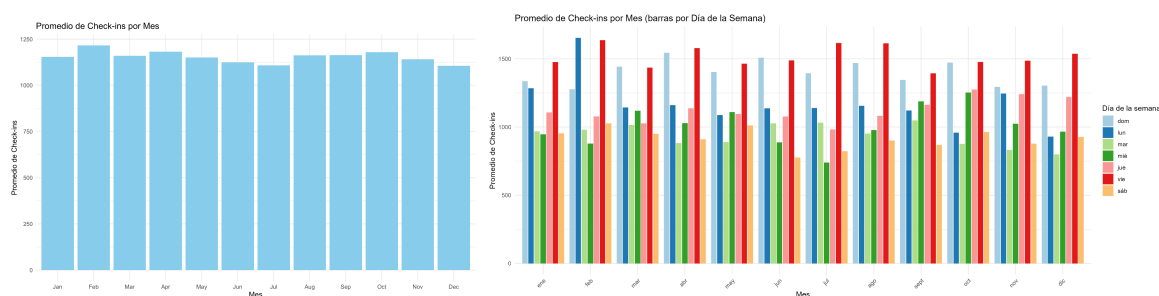
Promedio	Min	Max	Desv Est
1154	98	3298	410

Gráfica 1 y 2. Check-ins diarios y la distribución de los datos

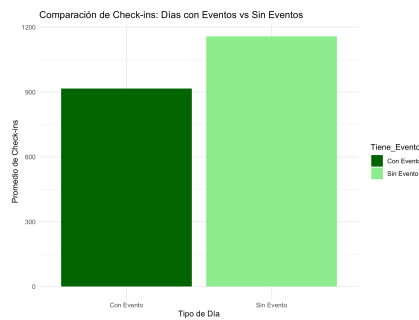


Para entender mejor en qué momentos ocurrían estos picos inusuales de check-ins (outliers), se revisó cómo se comportan los promedios por mes y por día de la semana. En el Gráfico 4 se observa que, aunque la mayoría de los meses mantienen niveles parecidos, febrero, abril y octubre muestran una demanda ligeramente más alta. Luego, el Gráfico 5 revela que los días con más check-ins fueron domingo y viernes, lo que ayuda a explicar por qué ciertos días alcanzan valores mucho más altos que el resto. Después, para comprobar si estos picos coincidían con eventos especiales (como el Super Bowl, Año Nuevo o grandes convenciones) se analizó la comparación del Gráfico 6. Allí se ve que los días con evento no siempre generan más check-ins, e incluso los días sin evento presentan un promedio un poco mayor. Esto sugiere que muchos de los valores extremos no dependen únicamente de celebraciones masivas, sino de patrones propios del turismo en Las Vegas y de la dinámica natural de los fines de semana.

Gráfico 4 y 5. Check-ins promedios por mes (barras por día de la semana)



**Gráfico 6.** Check-ins con evento vs. sin evento



## II. Modelos

Para profundizar en el comportamiento observado en los gráficos y entender qué factores influyen realmente en la cantidad de check-ins, se construyeron tres modelos distintos con el objetivo de predecir la demanda diaria. Estos modelos permiten analizar cómo variables como el precio, los tipos de clientes, los eventos y los días de la semana afectan las llegadas, y así obtener una visión más completa y útil para la planificación del hotel.

### 1. MODELO DE REGRESIÓN LINEAL SIMPLE

El intercepto estima el número promedio de check-ins cuando el precio FIT ADR es cero; aunque ese valor no existe en la práctica, funciona como referencia del nivel base de llegadas. En este modelo, el intercepto es 1.279,3 check-ins y es altamente significativo ( $p < 2e-16$ ), por lo que se concluye que el promedio de llegadas es distinto de cero con un nivel de significancia del 0,1% ( $p < 0,001$ ).

Por otro lado, el coeficiente del FIT ADR es negativo (-0,6916), lo que indica que por cada unidad adicional en el precio, los check-ins disminuyen en aproximadamente 0,69 unidades. Este efecto también es significativo al 0,1% ( $p = 0,000114 < 0,001$ ), por lo que se rechaza la hipótesis nula y se concluye que el precio promedio tiene un efecto real y negativo sobre la cantidad de llegadas.

$$Checkins = \beta_0 + \beta_1(FIT\ ADR) + \varepsilon$$

**Tabla 2.** Resultados del modelo lineal simple

Parámetro	Estimación (Estimate)	Error Estándar (Std. Error)	Valor t	Valor p	Signif.
Intercepto (b0)	1.279,301	34,2887	37,31	< 2e-16	***
FIT ADR (b1)	-0,6916	0,1787	-3,87	0,000114	***
R <sup>2</sup> : 0,0111			R <sup>2</sup> Ajustado: 0,01036		

Nota. Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

El  $R^2$  de 0,0111 muestra que el precio solo explica el 1,11% de la variación en los check-ins. Esto significa que el modelo tiene muy poco poder para predecir la demanda únicamente a partir del precio, por lo que es necesario incluir más variables para entender mejor el comportamiento real de los huéspedes.

## 2. MODELO DE REGRESIÓN LINEAL MÚLTIPLE

El modelo 2 se realizó para mejorar la precisión del análisis inicial y comprobar si variables adicionales, como los tipos de cliente, los eventos específicos y los días de la semana, explicaban mejor las variaciones en los check-ins. Al incluir estos factores, el modelo permite identificar con mayor claridad qué elementos tienen un efecto real y significativo sobre la demanda diaria, ofreciendo una visión más completa y útil para la planificación operativa.

$$\begin{aligned} \text{Checkins} = & \beta_0 + \beta_1(\text{FIT ADR}) + \beta_2(\text{Casino}) + \beta_3(\text{Group}) + \beta_4(\text{SE}) + \beta_5(\text{SuperBowl}) + \\ & \beta_6(\text{Valentines}) + \beta_7(\text{NY}) + \beta_8(\text{Easter}) + \beta_9(\text{LaborDay}) + \beta_{10}(\text{Xmas}) + \\ & \beta_{13}(\text{DiaSemana4}) + \beta_{14}(\text{DiaSemana5}) + \beta_{15}(\text{DiaSemana6}) + \beta_{16}(\text{DiaSemana7}) + \varepsilon \end{aligned}$$

Tabla 3. Resultados del modelo lineal múltiple

Variable	Estimate	Std. Error	t value	Pr(> t )	Signif.
(Intercept)	1137,132	80,02692	14,209	< 2e-16	***
FIT ADR	-0,63208	0,21215	-2,979	0,002941	**
Casino	0,24363	0,0565	4,312	1.74e-05	***
Group	0,10838	0,02847	3,806	0,000147	***
SE	-0,01903	0,07145	-0,266	0,790051	
SuperBowl	-765,118	177,1179	-4,32	1.68e-05	***
Valentines	-361,154	171,9869	-2,1	0,035928	*
NY	-698,279	199,8521	-3,494	0,000492	***
Easter	-203,314	173,7115	-1,17	0,242047	
LaborDay	-135,584	200,4799	-0,676	0,49897	
Xmas	-425,614	202,1348	-2,106	0,035428	*

DiaSemana 2	-198,996	37,00015	-5,378	8.89e-08	***
DiaSemana 3	-411,951	37,73127	-10,918	< 2e-16	***
DiaSemana 4	-364,903	37,36168	-9,767	< 2e-16	***
DiaSemana 5	-275,109	36,2784	-7,583	6.33e-14	***
DiaSemana 6	74,37534	36,04632	2,063	0,039277	*
DiaSemana 7	-536,909	36,36368	-14,765	< 2e-16	***
R <sup>2</sup> : 0.3102			R <sup>2</sup> Ajustado: 0.3018		

Nota. Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

El modelo tiene un R<sup>2</sup> de 0,3102, lo que significa que alrededor del 31% de la variación en los check-ins puede explicarse con las variables incluidas. El R<sup>2</sup> ajustado (0,3018) confirma que el modelo sigue siendo consistente incluso considerando cuántos predictores se usaron.

#### **Las variables de tipo de cliente muestran efectos reales:**

- Casino tiene un coeficiente positivo y altamente significativo ( $p < 0,001$ ), indicando que este segmento aumenta los check-ins.
- Group también es positivo y muy significativo ( $p < 0,001$ ), mostrando que la llegada de grupos eleva la demanda.

#### **Entre los eventos especiales, sólo dos muestran efectos fuertes y reales:**

- Super Bowl presenta una disminución significativa en los check-ins ( $p < 0,001$ ), indicando un impacto negativo claro.
- New Year (Año Nuevo) también reduce de manera significativa las llegadas ( $p < 0,001$ ).

#### **Los días de la semana muestran un patrón consistente:**

- Los días martes (2), miércoles (3), jueves (4), viernes (5) y domingo (7) son significativamente diferentes al lunes (todos  $p < 0,001$ ), presentando una reducción real en los check-ins.
- Esto indica que el lunes es el día con mayor estabilidad, mientras que estos días muestran una demanda menor de manera estadísticamente comprobada.

### 3. MODELO NO LINEAL

El modelo con términos no lineales se desarrolló con el propósito de capturar patrones más complejos en la demanda diaria que no podían ser identificados mediante una especificación lineal tradicional. Al incorporar términos cuadráticos para las variables clave (FIT ADR, Casino, Group y SE), el modelo permite evaluar si el efecto de estos factores cambia en diferentes niveles, revelando relaciones que no son estrictamente lineales. Además, incluye eventos especiales y dummies de día de la semana, lo que posibilita aislar su impacto específico sobre los picos y caídas en los check-ins. Este enfoque enriquecido proporciona una descripción más precisa del comportamiento real de la demanda, mejora la capacidad explicativa del modelo y ayuda a entender mejor cómo interacciones complejas entre precio, tipo de cliente y días de la semana influyen en la operación hotelera.

$$\text{Checkins} = \beta_0 + \beta_1(\text{FIT ADR}) + \beta_2(\text{FIT ADR}^2) + \beta_3(\text{Casino}) + \beta_4(\text{Casino}^2) + \beta_5(\text{Group}) + \beta_6(\text{Group}^2) + \beta_7(\text{SE}) + \beta_8(\text{SE}^2) + \beta_9(\text{Eventos}) + \beta_{10}(\text{DiaSemano}) + \varepsilon$$

**Tabla 4.** Resultados del modelo no lineal

Variable	Estimate	Std. Error	t value	Pr(> t )	Signif.
(Intercept)	747,3	152,8	4,892	1.12e-06	***
FIT ADR	-1,318	0,817	-1,613	0,1069	
FIT ADR2	0,001325	0,001825	0,726	0,46789	
Casino	0,8213	0,1989	4,13	3.86e-05	***
Casino 2	-0,00021	7,31E-05	-2,896	0,00384	**
Group	0,2632	0,06114	4,305	1.80e-05	***
Group2	-4,6E-05	2,35E-05	-1,945	0,05195	.
SE	-0,0181	0,1589	-0,114	0,90933	
SE2	5,38E-05	0,000151	0,357	0,72138	
SuperBowl	-732,4	178,7	-4,097	4.44e-05	***
Valentines	-372,1	170,8	-2,178	0,02958	*



NY	-523,5	202,3	-2,588	0,00977	**
Easter	-172,5	172,4	-1	0,31729	
LaborDay	-109,5	199	-0,55	0,58208	
Xmas	-301	202	-1,491	0,1363	
DiaSemana 2	-174,4	37,21	-4,686	3.07e-06	***
DiaSemana 3	-374,1	38,66	-9,676	<2e-16	***
DiaSemana 4	-334,1	38,06	-8,779	<2e-16	***
DiaSemana 5	-264,2	36,36	-7,265	6.37e-13	***
DiaSemana 6	80,83	35,79	2,258	0,02409	*
DiaSemana 7	-530,3	36,23	-14,636	<2e-16	***
$R^2: 0,325$			$R^2 \text{ Ajustado: } 0,3148$		

El Modelo 3 constituye una evolución metodológica del Modelo 2 al incorporar términos no lineales que permiten capturar dinámicas complejas en la demanda diaria que un modelo estrictamente lineal no puede representar. Al añadir componentes cuadráticos para variables clave como FIT ADR, Casino, Group y SE, el modelo revela patrones de retornos decrecientes, especialmente en los segmentos Casino y Group, cuyos efectos positivos sobre la demanda se atenúan a niveles elevados, lo que ofrece una lectura mucho más precisa del comportamiento real del mercado. Esta ampliación estructural se traduce en una mejora del poder explicativo, pasando de un  $R^2$  del 31% en el Modelo 2 a un 32,5% en el Modelo 3, incremento que también se refleja en el  $R^2$  ajustado y confirma la pertinencia estadística de los nuevos términos sin incurrir en sobreajuste. A pesar de estos avances, los efectos de eventos especiales como Super Bowl, San Valentín y Año Nuevo así como la estacionalidad semanal permanecen consistentes y altamente significativos, demostrando que son patrones estructurales de la operación hotelera. En síntesis, el Modelo 3 ofrece una comprensión más profunda y refinada de los factores que impulsan la variabilidad en los check-ins, consolidándose como una herramienta analítica superior para optimizar la planificación operativa y la asignación eficiente de recursos en entornos hoteleros de alta demanda.

### III. Comparación entre modelos con la matriz de confusión de cada uno

Después de presentar los tres modelos y revisar cómo cada uno intenta predecir correctamente los días de mayor o menor cantidad de check-ins, se construyeron dos tablas para comparar su desempeño. La primera tabla muestra cómo se comportan los modelos al clasificar los días en dos categorías, mientras que la segunda tabla resume las métricas principales que permiten evaluar qué tan bien funciona cada modelo en general.

Tabla 5. Matrices de confusión de los tres modelos.

Modelo 1			Modelo 2			Modelo 3		
	Predicción 0	Predicción 1		Predicción 0	Predicción 1		Predicción 0	Predicción 1
Real 0	331	233	Real 0	586	212	Real 0	575	203
Real 1	384	389	Real 1	129	410	Real 1	140	419

La Tabla 5 muestra cuántas veces cada modelo acertó o se equivocó al clasificar los días. En palabras simples, aquí se ve cuántos días fueron identificados correctamente como de “baja demanda” o “alta demanda”. El Modelo 1, que era el más básico, se equivocó varias veces y no logró distinguir bien entre los dos tipos de días. En cambio, el Modelo 2 y el Modelo 3 hicieron un trabajo mucho mejor, acertando en la mayoría de los casos. Esto indica que, cuando se agregan más variables o una estructura un poco más compleja, los modelos son capaces de reconocer mejor los patrones reales en los datos.

Tabla 6. Métricas comparativas de los modelos

Métrica	Modelo 1 (Simple)	Modelo 2 (Múltiple)	Modelo 3 (No lineal)
Accuracy	0.5385	0.7450	0.7435
Sensibilidad	0.4629	0.8196	0.8042
Especificidad	0.6254	0.6592	0.6736

<b>PPV (Valor Predictivo Positivo)</b>	0.5869	0.7343	0.7391
<b>NPV (Valor Predictivo Negativo)</b>	0.5032	0.7607	0.7496
<b>Kappa</b>	0.0870	0.4829	0.4810
<b>Balanced Accuracy</b>	0.5442	0.7394	0.7389

La Tabla 6 resume qué tan bien funcionó cada modelo de una forma más general. Aquí se ve claramente que el Modelo 2 y el Modelo 3 tienen valores más altos en todas las métricas importantes, lo que significa que logran predecir con mucha más precisión que el Modelo 1. Ambos modelos reconocen mejor los días con mucha demanda y los días con poca demanda, y muestran una mayor estabilidad en sus resultados. En comparación, el Modelo 1 tiene valores más bajos, lo que confirma que no fue suficiente usar solo una variable para explicar el comportamiento de los check-ins. Por esto, los modelos más completos demostraron ser más útiles para entender y anticipar los cambios diarios en la demanda del hotel.

### Conclusión

El análisis realizado permitió responder de manera clara la pregunta central del caso: ¿qué factores influyen en la cantidad diaria de check-ins en los hoteles Caesars y cómo puede utilizar esa información para planear mejor el personal de recepción? A lo largo del estudio se observó que la demanda hotelera no depende de un solo elemento aislado, sino de una combinación de precio, tipo de cliente, eventos especiales y patrones semanales. El uso de tres modelos diferentes —simple, múltiple y no lineal— permitió profundizar gradualmente en esta dinámica y obtener una comprensión más completa del comportamiento real de los huéspedes.

En relación con los objetivos específicos, todos fueron cumplidos. Primero, se identificó qué variables tienen un impacto significativo en los check-ins diarios. Se encontró que los tipos de cliente como Casino y Group aumentan la demanda, mientras que eventos como Super Bowl y Año Nuevo disminuyen las llegadas, y que varios días de la semana presentan reducciones claras respecto al lunes. Segundo, se desarrollaron modelos estadísticos capaces de explicar y predecir las fluctuaciones diarias de la demanda: el Modelo

1 sirvió como punto de partida, pero los Modelos 2 y 3 incorporaron más información y ofrecieron una lectura más realista y precisa del comportamiento de los huéspedes. Finalmente, el tercer objetivo también se cumplió, ya que las métricas de desempeño confirmaron qué tan eficiente era cada modelo para clasificar días de alta o baja demanda. Las matrices de confusión y los indicadores comparativos demostraron que los Modelos 2 y 3 tienen una capacidad predictiva mucho mayor, lo que los convierte en herramientas útiles para la toma de decisiones.

En conjunto, el estudio muestra que es posible anticipar con buena precisión la demanda diaria de check-ins y, por tanto, mejorar la planificación del personal en recepción. Comprender estos patrones permite equilibrar costos operativos y calidad del servicio, reduciendo filas en días de alta demanda y evitando exceso de personal en días de baja ocupación. En consecuencia, el análisis no solo responde a la pregunta del caso, sino que aporta una base sólida para una gestión hotelera más eficiente, estratégica y centrada en la demanda real.