

Sistema de Reservas de Hotel

Informe Predicción 3:

Series Temporales

Mayra Goicochea Neyra

CONTENIDO

CONTENIDO	2
ABSTRACT	3
INTRODUCCION	3
OBJETIVO	3
SERIE TEMPORAL	3
MODELAMIENTO	5
CONCLUSIONES	6
REFERENCIAS BIBLIOGRAFICAS	6
ANEXO	7

ABSTRACT

En los últimos años, el sector de alojamiento se ha solidificado como uno de los más rentables y competitivos debido al aumento del turismo no solo local (dentro de su país) sino al turismo global, dado las nuevas regulaciones de globalización de los gobiernos en muchos países.

Debido a la alta competencia, la gestión de ingresos toma parte relevante en el día a día de los administradores de hoteles. Siempre buscan estar a tiempo de tomar decisiones adecuadas para no perder clientes, ni afectar su imagen social. Es necesario tener la disponibilidad óptima para atender la demanda, que tiene temporadas altas y bajas.

En el presente informe, se presenta una solución para estimar la demanda de reservas realizadas, con la finalidad de medir su desempeño en estos años y que estrategias tomar a cara de la alta competencia, y sobre todo con los nuevos medios de venta que se han abierto por el uso de internet.

INTRODUCCIÓN

La industria de alojamiento se ha solidificado como un sector muy rentable y competitivo. Uno de sus factores de crecimiento ha sido la globalización.

Otro de los factores importantes el Internet, no solo porque permite conectarse a usuarios de todo el mundo, sino que guarda grandes volúmenes de información de los usuarios. Cada vez hay más empresas, como sitios web de comercio electrónico y tiendas en línea que ofrecen recomendaciones de productos para dirigirse a clientes potenciales. Esta tendencia de proporcionar recomendaciones, como ofertas y promociones personalizadas, a los clientes a través de diversos medios, como

sitios web, redes sociales en línea, televisión y teléfonos inteligentes, aumenta día a día.

Es así como el sector de alojamiento se ha convertido en una de las industrias más ricas en datos. Estas propiedades hacen que el análisis de datos en la industria hotelera sea complejo. Satisfacer las expectativas de los clientes es un factor clave en la industria hotelera para captar la lealtad de los clientes. Con ese objetivo, los encargados del área de marketing y los analistas de datos buscan activamente formas para utilizar esta información de la mejor manera posible y avanzar con soluciones analíticas óptimas de datos.

En este informe, se presenta un prototipo basado en modelos predictivos de series temporales. Los modelos formulados son ARIMA y ETS. Siendo el modelo más eficiente al caso del negocio el de ARIMA con un margen de error menor al ETS.

OBJETIVO

El prototipo busca estimar el número de reservas con la finalidad de proporcionar al área comercial y planeamiento, información relevante a sus estrategias para el futuro.

SERIE TEMPORAL

El conjunto de datos contiene 40060 observaciones de reservas realizadas entre el periodo 30 de Junio del 2013 al 31 de Agosto del 2017.

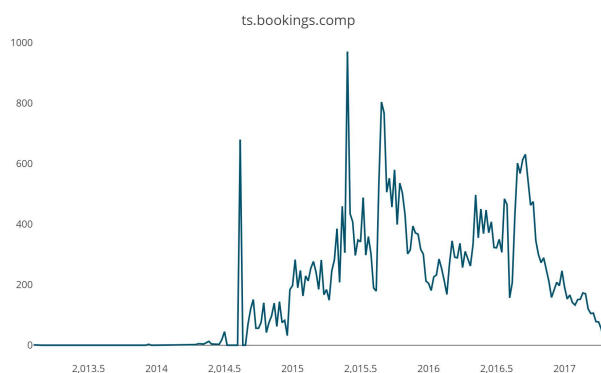
En el conjunto de datos solo se presentaba la fecha de arribo estimado (o realizado) de la reserva. Para calcular la fecha de la reserva se consideró la siguiente formula $ArrivalDate - LeadTime$. Hay 444 observaciones de reservas que tiene un $LeadTime$ (Tiempo de anticipación de la reserva) mayor a 365 días (se puede asumir que debido a error o hay casos atípicos).

Se consideró además que es una serie temporal irregular (dado que no todas las semanas tendrán reservas realizadas), además de que la información es relativa a la restricción del dataset bajo la fecha de arribo. Por ello, en el

análisis exploratorio (Informe 1) se realizó la imputación de las semanas que no tienen reservaciones, porque los modelos (especialmente ARIMA no podrá establecer las diferencias correctas entre los valores con sus valores pasados).

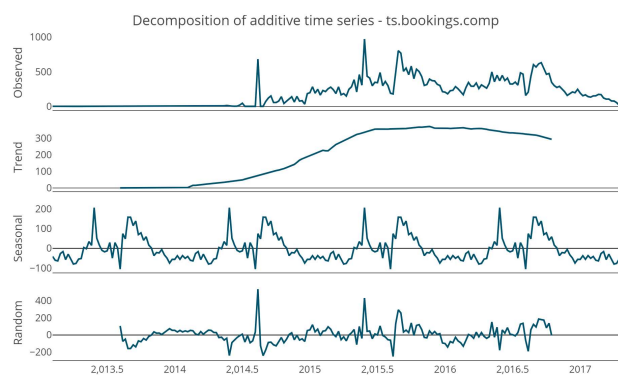
La serie temporal muestra un bajo número entre 2013-2014 (ver Fig.1)

Fig: 1.



En cuanto a las características de la serie, su descomposición muestra lo siguiente: (Ver fig.2)

Fig: 2.



- Parece que hubo un crecimiento en los primeros años, aunque es subjetivo a la restricción de los datos, Porque se observa además que se amortigua así que no se puede determinar que tiene tendencia.

- Tiene además ruido de aleatoriedad (Random). Aunque considero que puede deberse a la restricción de los datos.
- En cuanto a la estacionalidad, el grafico de descomposición muestra que es estacional en los años (repite un patrón cada año). Sin embargo, el grafico 3 (Plot Estacional trimestral) muestra que no es estacional por trimestre. (Ver Fig. 3). Y el grafico 4, también demuestra que no es estacional semanal (Ver Fig.4) por lo que la serie no es estacional para el horizonte de tiempo que estamos analizando.

Fig: 3.

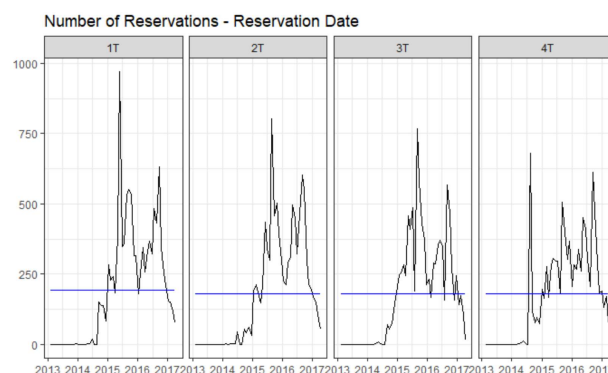
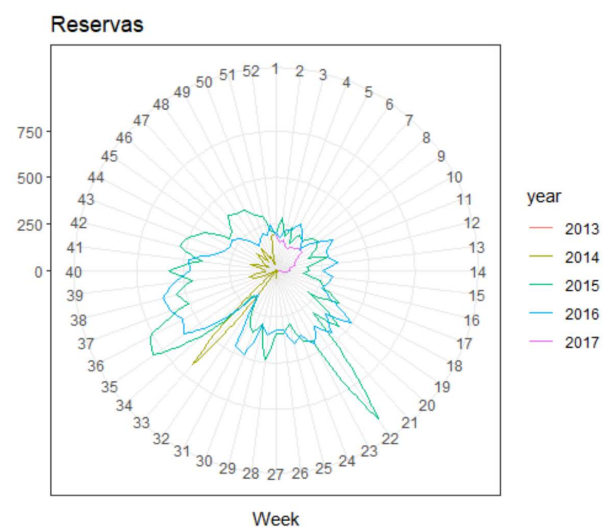
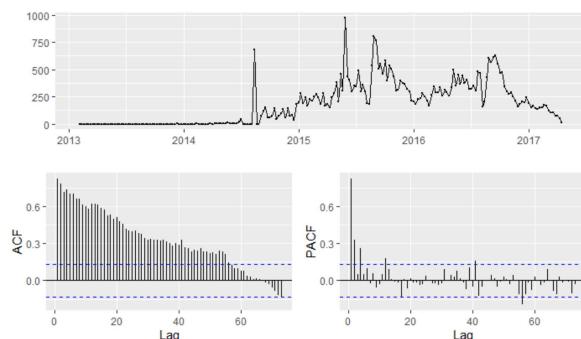


Fig: 4.



El grafico ACF muestra que cuantos más retardos se realicen, la función de auto correlación decae, lo que confirma que no existe una asociación lineal entre las observaciones separadas.

Fig: 4.



MODELAMIENTO

Para la construcción de los modelos predictivos, se dividió la muestra en dos datasets:

- Train: contiene información del periodo: 24 de Junio del 2013 al 14 de Mayo del 2017.
- Test: contiene la información del periodo: 21 de Mayo del 2017 al 31 de Agosto del 2017.

El primer modelo generado es un ETS Aditivo, se obtuvo un AIC de 2912.654. (Ver Fig.5)

Fig: 5.

```
ETS(A,N,N)
Call:
ets(y = bookings.comp.train)

Smoothing parameters:
  alpha = 0.4022

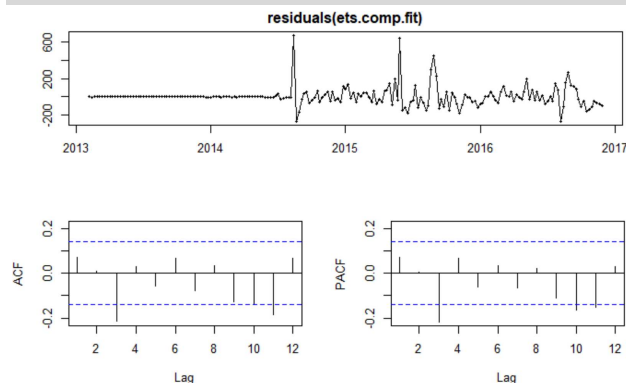
Initial states:
  l = 0.1608

sigma: 105.7998

AIC    AICc    BIC
2912.654 2912.777 2922.534
```

Los residuos del modelo ETS, muestran que en el lag 3 se consigue un valor negativo mayor a -0.2, no es ruido blanco.

Fig: 6.



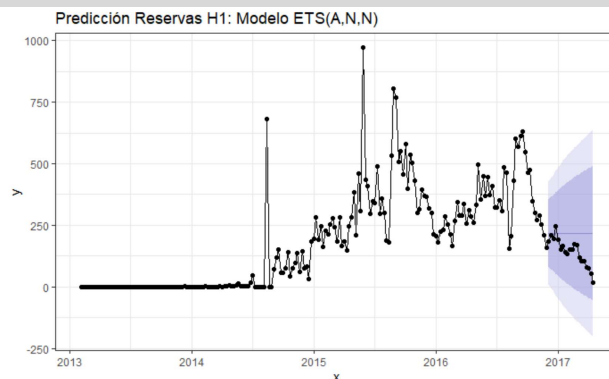
La predicción de este modelo, obtuvo 95.1885 de RMSE (Raíz del error medio cuadrado) y 131.7569 de MAPE (Porcentaje de Error Medio). (Ver Fig. 7)

Fig: 7.

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	2.732244	105.2668	54.38631	-Inf	Inf	0.9543996
Test set	-77.834118	95.1885	80.55071	-130.6526	131.7569	1.4135460

En cuanto al grafico de sus pronósticos, sus valores solo tienen tendencia y se debe a que la serie temporal no es estacional. (Ver Fig.8)

Fig: 8.



Un modelo ARIMA es mas adecuado para la serie temporal. El primer modelo generado es

un ARIMA(2,1,2). Tiene un AIC de 2402.78. (Ver Fig. 9)

Fig. 9.

```
Series: bookings.comp.train
ARIMA(2,1,2)

Coefficients:
      ar1  ar2      ma1      ma2
      -0.5001  0.4  -0.0120  -0.7116
s.e.      0.1367  0.1   0.1223   0.0887

sigma^2 estimated as 10543:  log likelihood=-1196.39
AIC=2402.78  AICc=2403.09  BIC=2419.22
```

También se creó un modelo ARIMA (1,1,3), que obtuvo un mejor AIC de 1924.91 (Ver Fig. 10)

Fig. 10.

```
Series: bookings.comp.train
ARIMA(1,1,3)
Box Cox transformation: lambda= 0.7918145

Coefficients:
      ar1      ma1      ma2      ma3
      -0.8845  0.3575  -0.5563  -0.2237
s.e.      0.1237  0.1355  0.1098  0.0761

sigma^2 estimated as 944.1:  log likelihood=-957.46
AIC=1924.91  AICc=1925.23  BIC=1941.36
```

En cuanto a sus residuos, ambos modelos no presentan correlaciones significativas.

En cuanto a los gráficos de sus predicciones, el que mejor se acerca es el modelo ARIMA (1,1,3). (Ver Fig. 11 y 12)

Fig. 11.

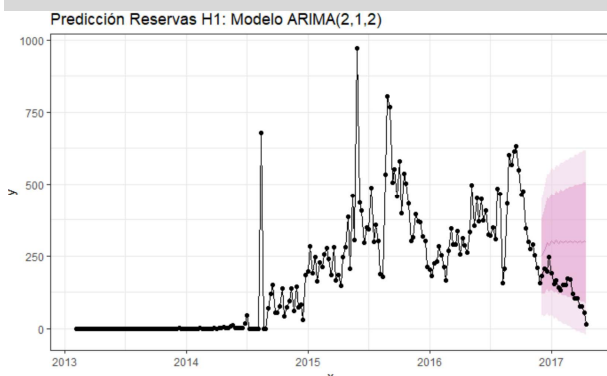
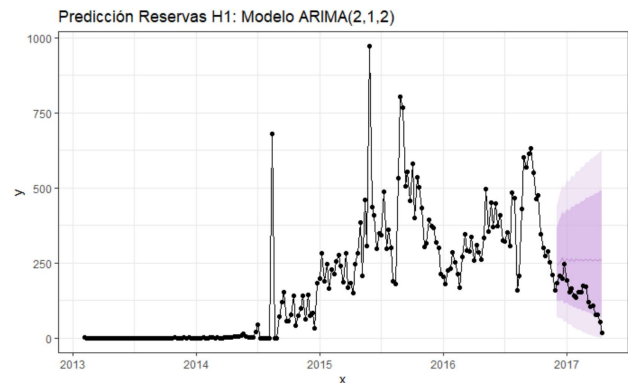


Fig. 12.



Las métricas de Error demuestran que el modelo ARIMA (1,1,3) es con menor error. (Ver Fig. 13)

Fig. 13.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	5.967269	101.3821	51.8596	NaN	Inf	0.3126635	0.0005408376	NA
Test set	-153.079030	164.6943	153.0790	-212.0957	212.0957	0.9229193	0.7165459997	8.022213
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	6.809597	100.508	51.23844	NaN	Inf	0.3089185	0.0506410	NA
Test set	-117.210160	130.019	117.21016	-172.6169	172.6169	0.7066645	0.7110463	6.647606

CONCLUSIONES

- La serie temporal generada guarda información de las reservas con fecha de arribo en el periodo 1 de julio del 2015 al 31 de agosto del 2017, por lo que la estimación de los modelos no es tan exacta para una predicción del número de reservas en forma general. Se podría afinar con un dataset que considere como base la fecha de reserva.
- El modelo ARIMA se ajusta mejor a esta serie temporal, dado mantiene cierta aleatoriedad que el modelo ETS no considera.

REFERENCIAS BIBLIOGRAFICAS

- [1] Analysis de Series temporales en R. ARIMA ([link](#))
- [2] Time series analysis and forecasting with the TSstudio package ([link](#))
- [3] ARIMA modelling in R, ([link](#))

- [4] Build High Performance Time Series Models using Auto ARIMA in Python and R, ([link](#))

ANEXO

- Archivo RMarkdown en [Github](#) (MasterDS2019>Prediccion>Research>PRED-Examen20-Informe3MGN.Rmd).