Time Series Analysis of 51 Chicago Upper Upscale Hotels' ADR (Average Daily Rates), Occupied Rooms and RevPAR(Revenue Per Available Room)

Cheng Chen
Yimei Tang

Table of Contents

Abstract	2
Data Description	3
Data Analysis	5
Conclusions	10
Appendix	16

Abstract

Over the decade, hospitality industry has neglected the importance of data and was slow in applying data analysis in the industry. In the area of hospitality, we have three major players: hotel guests who are the end users of the hotel rooms, hotel sales and operation team who set the rate and deliver the best possible services for guests and hotel investors who make the final decisions whether to build or renovate hotels in certain areas. In order to leverage the use of data, we decided to apply our knowledge in time series analysis in providing answers for these three major players and start our journey from the city of Chicago and in the upper-upscale hotel class.

In this report, we will explain how much it will cost to book an upper-upscale hotel room on average (among these 51 hotels [See Appendix II]), how many occupied rooms we will expect in these 51 hotels across the busiest areas of Chicago and how much an investor could expect to earn from every hotel room he invested in the following 12 months.

Our goal in this project is to find the time series relationship (if any) respectively in ADR(Average Daily Rate), Occupied Rooms and RevPAR(Revenue Per Available Rooms) and if there is time series relationships , what is the ideal model for each factor .

Data Description

1. Data Source:

We are very grateful for Professor Lisa Thomas from School of Hospitality Leadership in providing us with three datasets for this project.

These datasets are originally generated by STR company.

2. Data Explanation:

These datasets are collected by STR company from 50 upper-upscale hotels in downtown Chicago and river north area. To see the full list of these 50 hotels, please refer to our appendix.

After discussing with team members, we decided to use the following data in our project:

(1) Monthly ADR: ADR stands for Average Daily Rate. Monthly ADR are calculated by adding that month's daily rate of each hotel in each day and divide them by the number of days in the month. The daily rate of each hotel is calculated by dividing daily room revenue by rooms sold in transient, group and contract three segments (each segment has different daily rates).

(2) Monthly Occupied Rooms: Monthly Occupied Rooms are calculated by adding all rooms that are sold each month.

(3) Monthly RevPAR: RevPAR stands for Revenue Per Available Room. Monthly RevPAR are calculated by dividing total room revenue generated that month by total supply rooms that month.

(4) Monthly Supply Rooms: Monthly Supply Rooms are calculated by adding all rooms that are for rent each month.

3. Data Cleaning

Although the datasets contain raw data, daily data, weekday and weekend data, monthly data, quarterly data and yearly data, we focus on the monthly data since it is more stable and is a large enough sample size for our project.

Monthly data has 143 month's values, starting from Jan 2005 to Nov 2016 (inclusively).

When we began our data cleaning process, we initially proposed to take out the data before 2010 since the economic crisis has impacted the industry of hospitality significantly. However, we found that we could still build dependable models by using the original dataset that has 11-year span. Therefore, we chose not to eliminate data that were observed before 2010.

Since the ADR range from 100 to 300, while the total occupied rooms and total supply rooms are both above 200,000 range, to read the data easier, we divided total occupied rooms and total supply rooms by 1,000.

Data Analysis

1. Data Exploration

Step One: We transformed three sets of data into Time Series data (Monthly ADR data, Monthly Occupied Rooms data and Monthly RevPAR data) to create time plots.

Fig 1_1 & 2 & 3: Time Plots of Monthly ADR & Occupied Rooms & Monthly RevPAR

Fig 1_1 Monthly ADR TS Plot

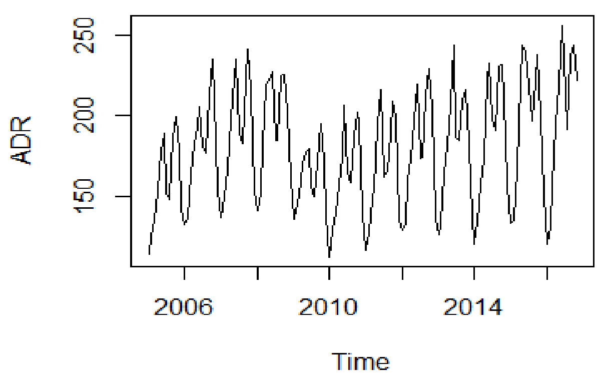


Fig 1_2 Monthly Occupired Rms TS Plot

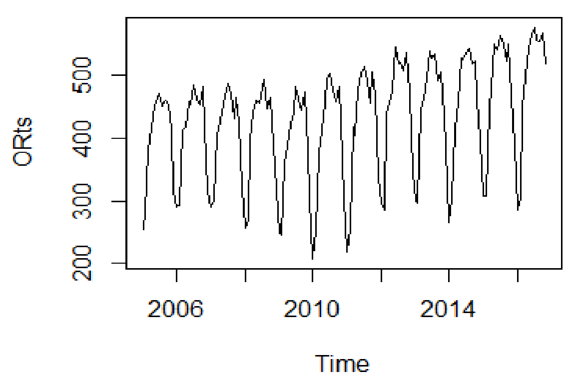
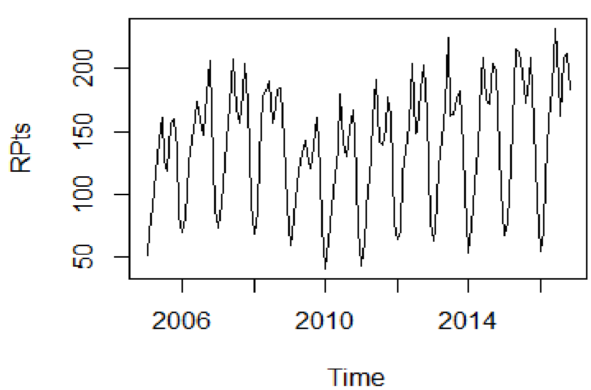


Fig 1_3 Monthly RevPAR TS Plot



From Fig 1_1, we can tell that the ADR data has seasonality behavior. There was a growing trend from 2005 to 2008. Than the ADR drop between 2008 to 2010. After 2010, the ADR starts to grow again till 2016.

From Fig 1_3, we can see the revenues per room has similar trend as ADR.

From Fig 1_2, we can see that the number of occupied rooms kept stable between 2005 and 2010 and has a growing trend after 2010.

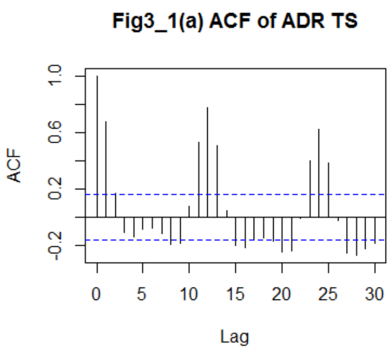
Step Two: Analysis Normal Distribution

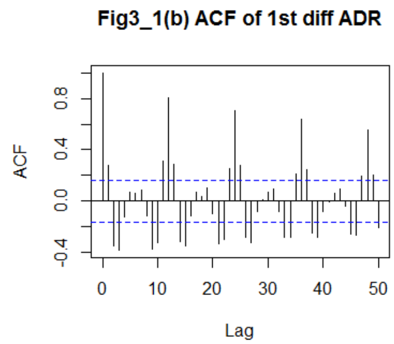
We create histograms and QQ plots to see the distribution of three data sets, and also use JB test to test the data normality. (Please see Fig 2_1, Fig 2_2, and Fig 2_3 in **Appendix I** for Histograms and QQ plots of three data sets)

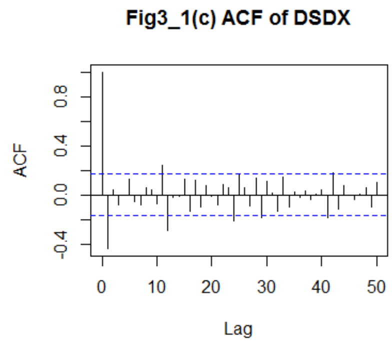
The P-values of JB tests for three data sets are 0.05, 0.004, and 0.04. All P-values are significant in $\alpha = 0.05$ level, which means all three TS data sets are normal distributed.

Step Three: Plot ACF

Since we saw seasonality behavior and trends for all three data sets on the time plots, we decide to use auto-correlation function to analyze the stationarity of X_t (ADR, Occupied Rooms, RevPAR) and its first difference. And we would like to check whether there is evidence for seasonality. Since three sets of data have similar situation, we only describe the ACFs of ADR data set. (Please see Fig 3_2(a)-(c), Fig 3_3(a)-(c) for Occupied Rooms RevPAR data sets in **Appendix I**)







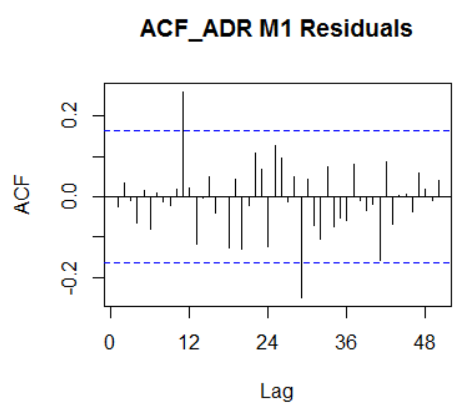
From the ACF plots of ADR data set, we can tell that after de-trending and de-seasonalizing, the data is still not stationary. Both Occupied Rooms and RevPAR data set are also not stationary after de-trending and de-seasonalizing.

2. Model Fitting & Residual Analysis

During the data exploration, we can tell that all three sets of data have similar features. It is explainable since ADR, Occupied Rooms, and RevPAR are all collected in same time range. So we will explain our model fitting processes by taking ADR data set as example.

We used auto.arima function and BIC criterion in R to help primarily select a model for each data set. The output showed that **ARIMA (1,0,1) (0,1,1) [12]** is the best model. So we plotted the ACF (Fig 4_1) of this model's residual as following:

Fig 4_1



From the Fig 4_1, we can see that the auto-correlation at some lags are still big. So we went further conduct JB test and Ljung Box test on the first model's residuals. The result shows that the residuals are independent and normal distributed. So we considered that the model can be used to fit the ADR data.

Based on the first model, we tried a lot of other models and selected out following two models as good ones: **ARIMA (0,1,1) (0,1,1) [12]** and **ARIMA (0,1,1) (1,0,1) [12].** Both models' residuals were all white noises tested through JB test and Ljung Box test.

We also selected some models for each of Occupied Rooms data and RevPAR data. After residual analysis for each possible model, we only found one best fitted model for each data set.

<u>Occupied Rooms</u>

Auto selected: **ARIMA (1,1,1) (0,1,2) [12]** (Since the residuals are not independent, we dropped this model selected through auto.arima function.)

So **ARIMA (0,1,6) (0,1,2) [12]** was the only model with white noise residuals.

<u>RevPAR</u>

Auto selected: **ARIMA (2,0,0) (2,1,1) [12]** (The residuals are independent without normal distribution)

ARIMA (2,0,1) (0,1,1) [12] (The residuals are independent without normal distribution)

ARIMA (1,0,1) (1,1,0) [12] (The residuals are white noise)

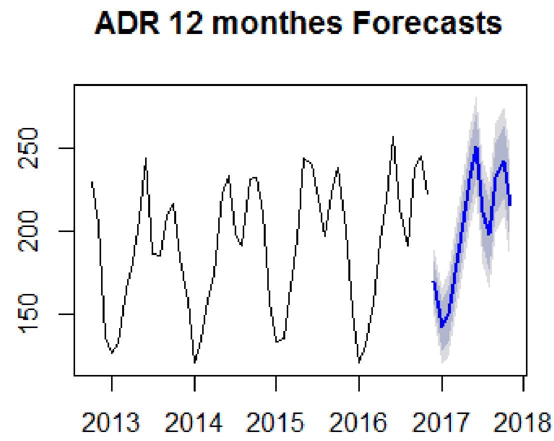
3. Forecast Analysis

After building the models, we computed a one-year (12 month) forecast for ADR, Occupied Rooms, and RevPAR. We have computed forecasts for every fitted model. To save space, we only display the best model's forecast values and intervals as following:

<u>ADR Forecast</u>

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Dec 2016	168.8999	155.7498	182.0499	148.7886	189.0112
Jan 2017	142.2457	127.2577	157.2337	119.3236	165.1678
Feb 2017	149.6229	133.2054	166.0405	124.5145	174.7314
Mar 2017	174.5130	156.9485	192.0775	147.6504	201.3756
Apr 2017	198.8811	180.3785	217.3837	170.5839	227.1784
May 2017	231.2236	211.9436	250.5036	201.7374	260.7099
Jun 2017	250.7444	230.8141	270.6747	220.2636	281.2251
Jul 2017	211.5137	191.0356	231.9917	180.1951	242.8322
Aug 2017	197.3045	176.3625	218.2465	165.2765	229.3325
Sep 2017	231.7582	210.4217	253.0948	199.1268	264.3896
Oct 2017	241.4539	219.7807	263.1271	208.3076	274.6002
Nov 2017	215.1256	193.1644	237.0869	181.5388	248.7125

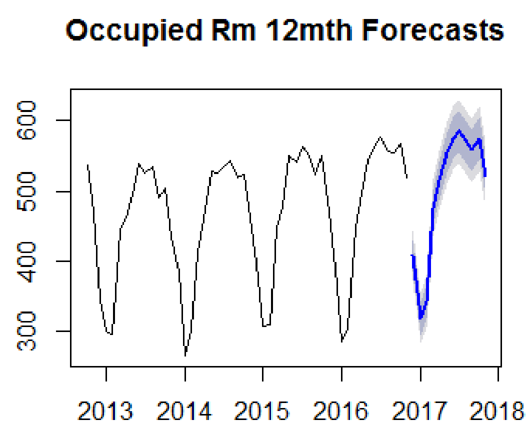
Fig 5_1:



Occupied Room Forecast

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Dec 2016	409.2083	386.2725	432.1441	374.1310	444.2856
Jan 2017	318.7783	294.9276	342.6291	282.3017	355.2549
Feb 2017	342.5483	317.6372	367.4594	304.4501	380.6465
Mar 2017	476.2267	450.0993	502.3541	436.2683	516.1852
Apr 2017	514.6997	487.1897	542.2097	472.6268	556.7726
May 2017	553.8051	524.7364	582.8738	509.3484	598.2618
Jun 2017	575.2929	546.0517	604.5342	530.5723	620.0135
Jul 2017	584.8228	555.4100	614.2356	539.8398	629.8058
Aug 2017	571.8553	542.2714	601.4393	526.6105	617.1001
Sep 2017	558.4677	528.7135	588.2218	512.9626	603.9727
Oct 2017	574.8745	544.9512	604.7978	529.1107	620.6382
Nov 2017	520.4720	490.3806	550.5635	474.4511	566.4929

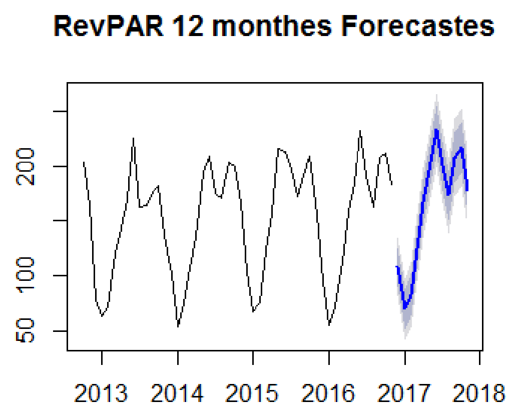
Fig 5_2



RevPAR Forecast

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Dec 2016	108.16744	90.81293	125.52196	81.62600	134.70888
Jan 2017	69.66278	51.20455	88.12101	41.43336	97.89221
Feb 2017	82.33243	62.95066	101.71420	52.69057	111.97429
Mar 2017	125.08829	104.92519	145.25138	94.25149	155.92508
Apr 2017	166.16065	145.33116	186.99013	134.30469	198.01660
May 2017	203.78451	182.38312	225.18589	171.05391	236.51510
Jun 2017	232.86343	210.96885	254.75801	199.37857	266.34830
Jul 2017	200.30844	177.98689	222.62998	166.17058	234.44630
Aug 2017	172.82790	150.13555	195.52024	138.12295	207.53285
Sep 2017	207.85352	184.83831	230.86873	172.65479	243.05224
Oct 2017	216.78843	193.49150	240.08536	181.15885	252.41802
Nov 2017	178.24360	154.70040	201.78680	142.23738	214.24982

Fig 5_3:



4. Models Validation

From the forecast, we can tell that our models are catching the time series trends. To test the accuracy of our models, we used “Back testing” to validate models. We fitted models with 80% of our data and tested with the other 20%. After computing the Mean Absolute Percentage Errors between all candidate models for one data set, we chose a best model for ADR data and Occupied Rooms data.

ADR Model: **ARIMA (1,0,1) (0,1,1) [12]**

The MAPE for this model is 4.96%

Occupied Rooms Model: ARIMA (0,1,6) (0,1,2) [12]
The MAPE for this model is 2.6%

For **RevPAR** data, though the lowest MAPE (6.73%) model is ARIMA (2,0,0) (2,1,1) [12], this model’s residuals are independent without normal distributed. So we decided to choose the model of **ARIMA (1,0,1) (1,1,0) [12]** with white noise residuals whose MAPE is lightly higher at 7.12%.

Conclusion

1. How much am I expected to pay for a hotel room on average in these 51 hotels?

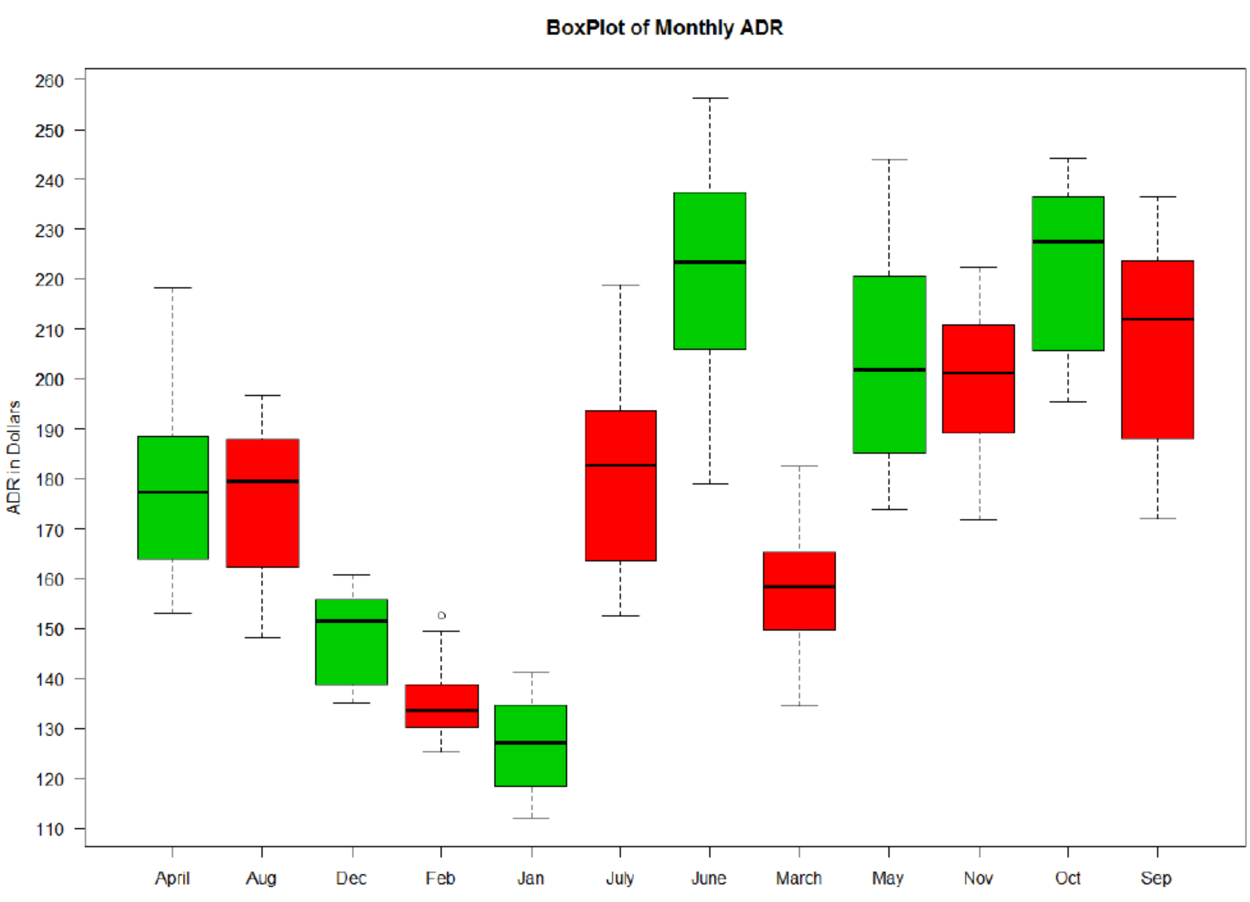
Based on our statistical analysis, the best model (least MAPE) we found for forecasting ADR is ARIMA (1,0,1) (0,1,1) [12]

$$(1-B^{12})(1-\Phi B)X_t = (1-\theta_1 B)(1-\theta_2 B^{12})a_t, \quad a_t \sim WN(0,\sigma^2)$$

```
> coeftest(m1)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.931728   0.040832 22.8185 < 2.2e-16 ***
ma1  -0.384822   0.099908 -3.8518 0.0001173 ***
sma1 -0.681488   0.088885 -7.6671 1.759e-14 ***
```

The model could be written as:

$$(1-B^{12})(1-0.93B)X_t=(1-0.38B)(1-0.68B^{12})a_t$$



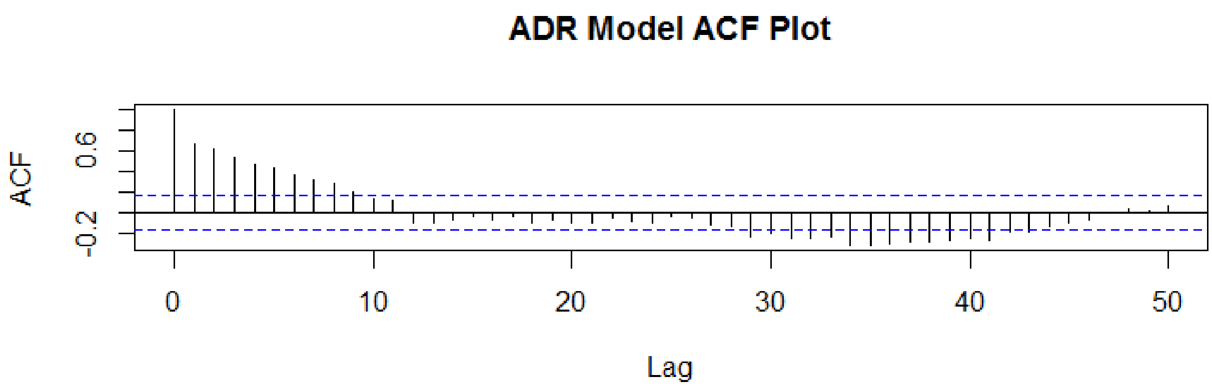
Therefore, if you are going to stay at one of the 51 hotels, your average daily rate will greatly depend on the same month of last year's rate, previous month of last year and the previous month of this year's rate.

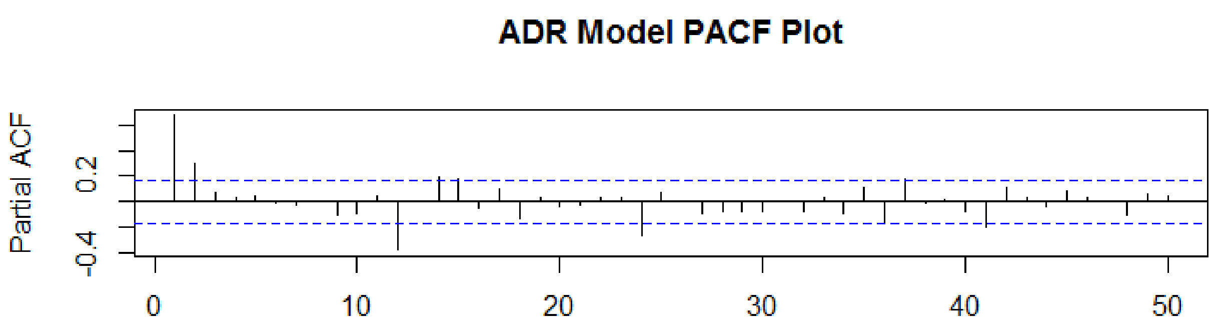
If you are a savvy customer, you will probably pick January, February, March and December. If you are visiting Chicago in June, September and October, you should expect to spend more in hotel rooms than other months. However, July and August will be better choices than June if you want to visit Chicago in the summer. (Fig 6_1 Box plot of Monthly ADR)

For hotel operators, it is not worthwhile to have price war during winter since the price is already low enough. However, hotels that want to capture more market share could lower their ADR within a reasonable range in June, September and October to attract price-sensitive guests.

On the other hand, as we could tell from the ACF plot (Fig 6_2) and the model itself, if the previous month's ADR is above its average, then next month's ADR will increase above its average. However, if last year's month ADR is above its average, then the month's ADR this year will be below its average. Therefore, if we have a higher than average ADR in March last year and February this year's ADR is lower than average, we will expect March this year's ADR will below average as well.

Fig 6_2



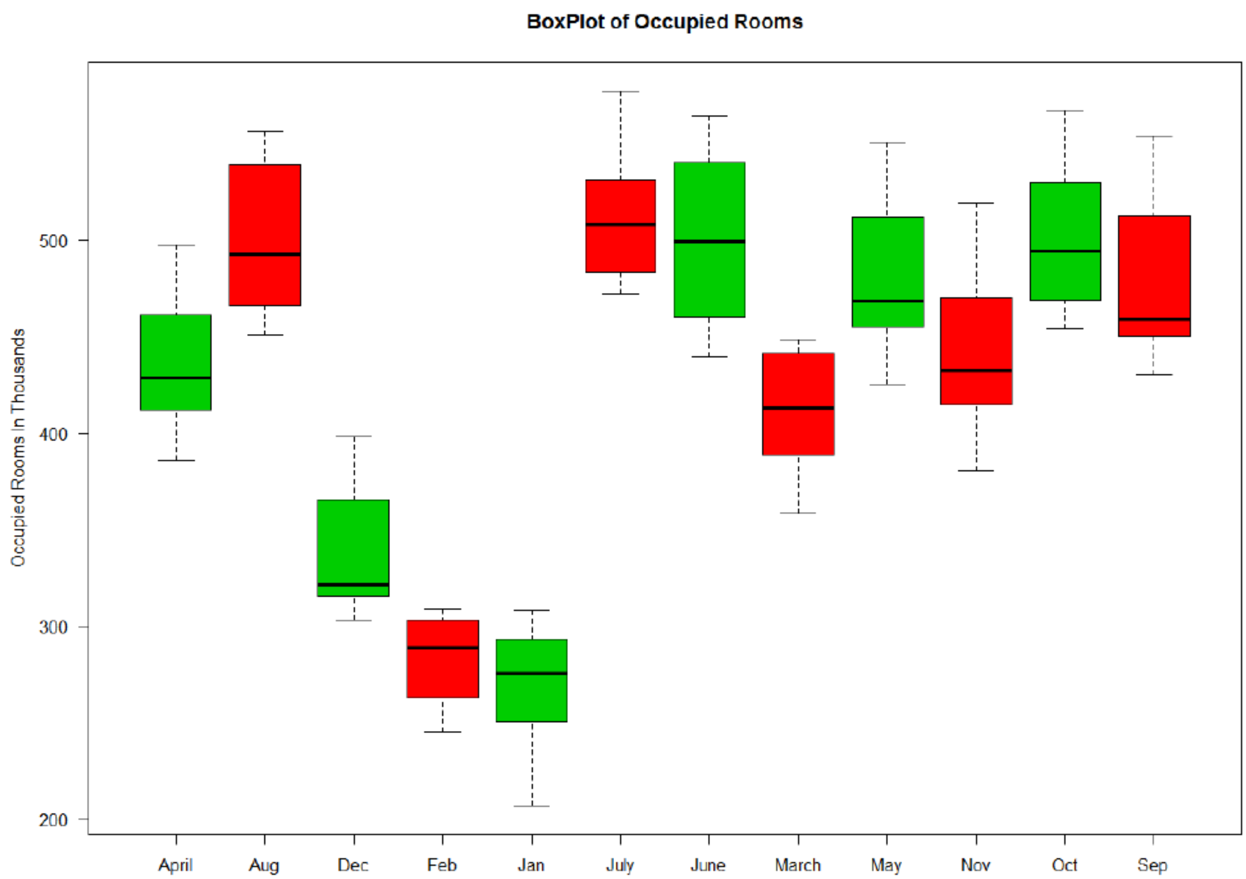



```
> coeftest(m7)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)    
ma1  -0.840143   0.099331 -8.4580 < 2.2e-16 ***
ma6  -0.307137   0.064581 -4.7558 1.976e-06 ***
sma1 -0.461982   0.099224 -4.6559 3.225e-06 ***
sma2 -0.227047   0.095711 -2.3722   0.01768 *  
```

2. How many occupied rooms will hotels expect in the next 12 months?

Based on our statistical analysis, the best model (least MAPE) we found for forecasting Occupied Rooms is ARIMA (0,1,6) (0,1,2) [12].

$$(1-B^{12})(1-B)X_t=(1-0.84B-0.31B^6)(1-0.46B^{12}-0.23B^{24})a_t$$



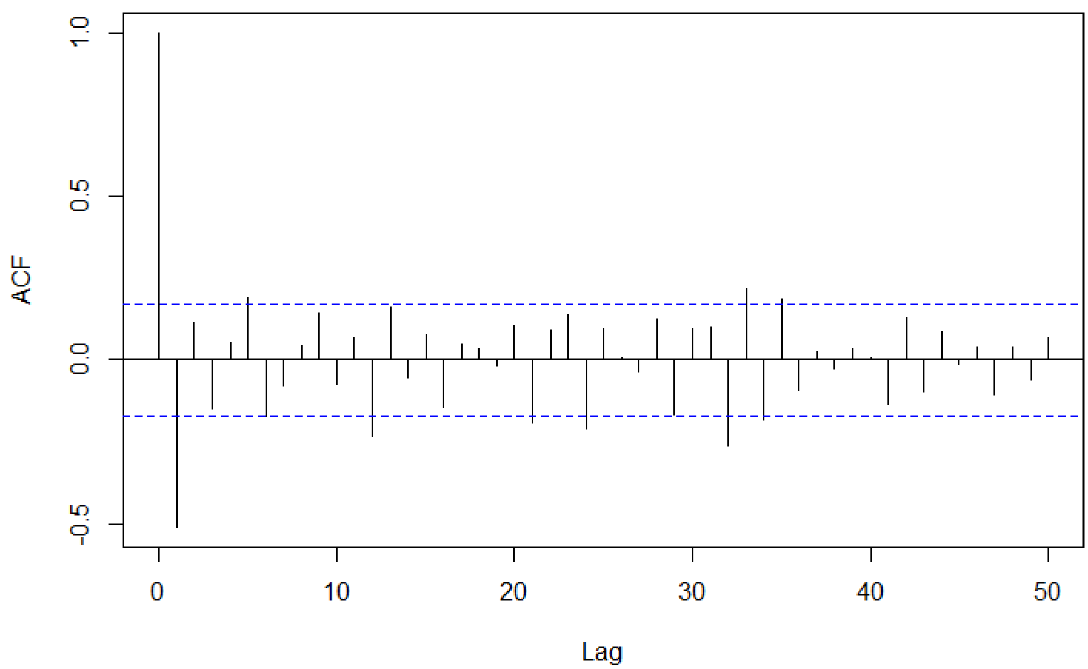
Therefore, the occupied rooms per month is influenced by occupied rooms from the same month last year, previous month last year, previous month this year and 6 months ahead this year's occupied rooms.

For hotel guests, just like the trend of ADR, in January, February and December, the demand (which is quantified in the occupied rooms) is lowest and therefore you will usually enjoy more comprehensive service in these three months compared to other busier months.

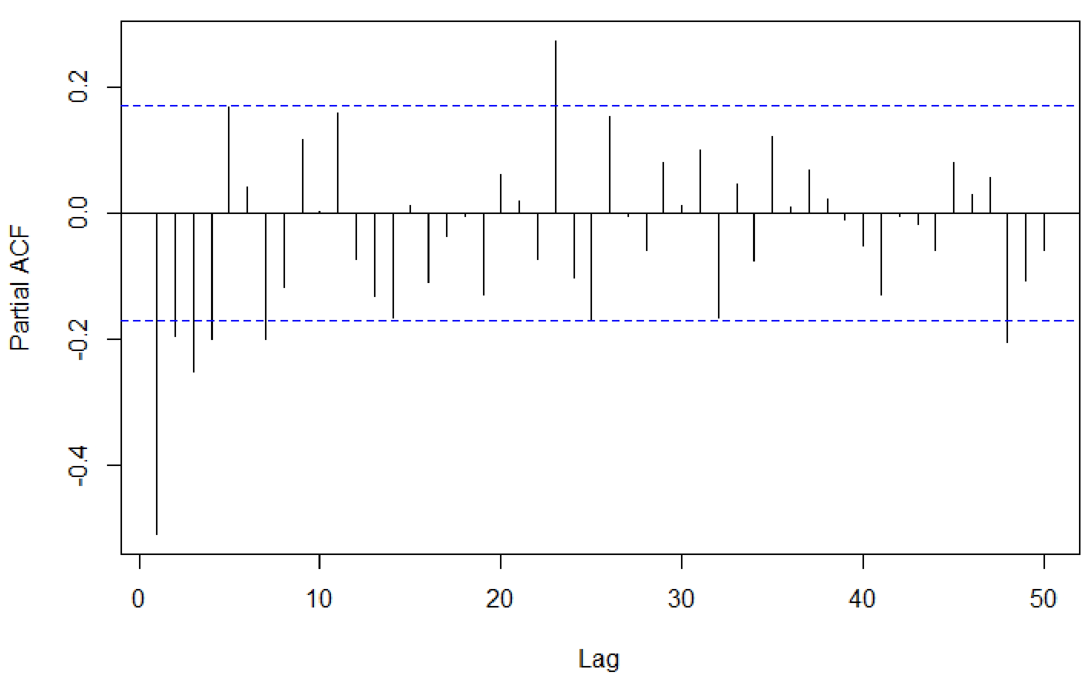
For hotel operations, these are the slow seasons so it might be a good time for hotel renovation and conduct hotel staff training. Also, these three-month's low occupancy should be taken into consideration for marketing strategy and pricing strategy, as well as hotel staffing guidance in doing yearly budget.

From observing the ACF plot and PACF plot below:

Occupied Rooms Model ACF Plot



Occupied Rooms Model PACF Plot



We know that if the previous month's occupied rooms are below its average, we will expect to see current month's occupied rooms to be above its average. It is reasonable in real life. If some tourists are not visiting Chicago in a month, they might choose the month before that month or after than month (usually within 3 months' span). What interesting is that if the month last year has below average occupied rooms, there will be above average occupied rooms in the same month this year. our assumption for this behavior is that tourists usually don't visit Chicago two years in a row.

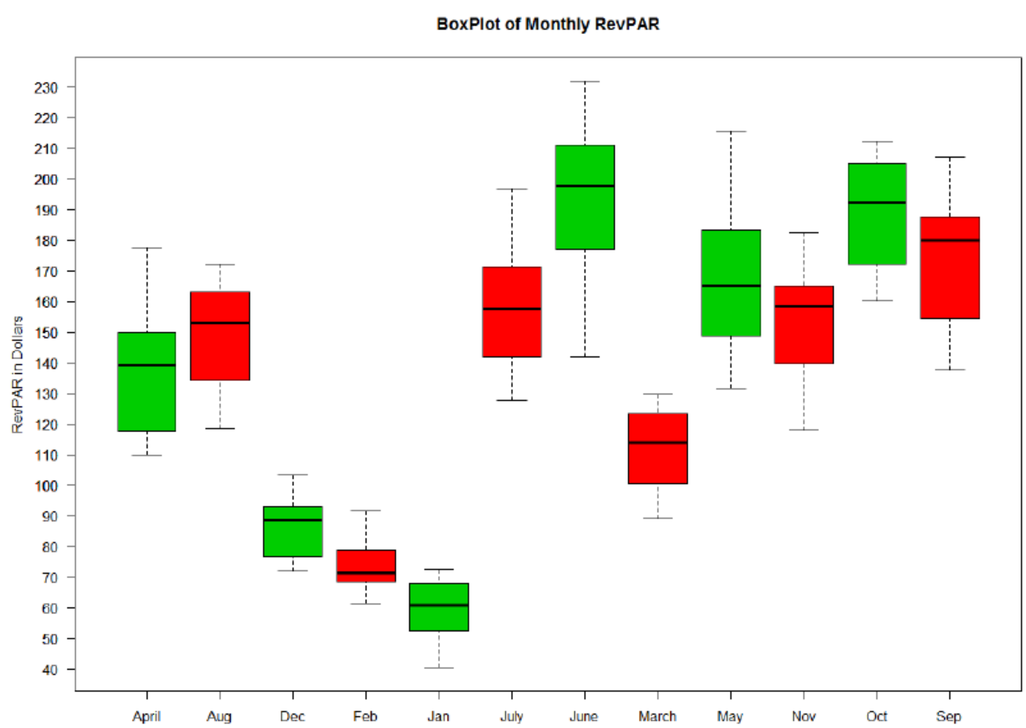
3. How many revenues could hotel investors expect to generate from a hotel room in the next 12 months?

Based on our statistical analysis, the best model (least MAPE) we found for forecasting RevPAR is: **ARIMA (1,0,1) (1,1,0) [12]**

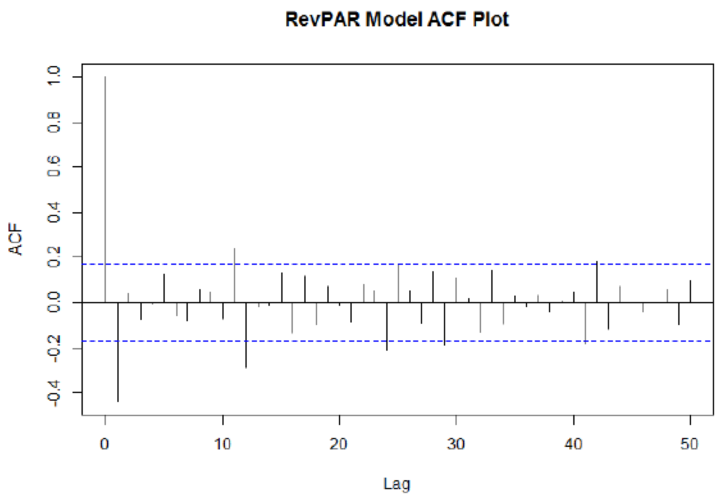
```
> coeftest(m11)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)    
ar1   0.940279   0.036643 25.6609 < 2.2e-16 ***
ma1  -0.578007   0.081020 -7.1341 9.739e-13 ***
sar1 -0.359425   0.086502 -4.1551 3.251e-05 ***
```

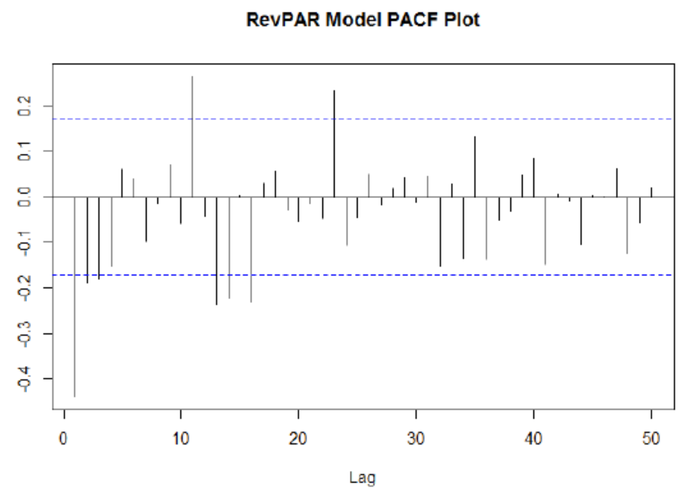
$$\mathbf{(1-B^{12})(1-0.94B)(1+0.36B^{12})X_t=(1-0.57B)a_t}$$

Therefore, the RevPAR per month is influenced by RevPAR of the same month last year, previous month last year and previous month this year.



Like our results from ADR and Occupied rooms, hotel investors are not generating much revenue from hotel rooms they invested in winter. Therefore, hotel investors should consider close its hotels for renovations or major upgrades in winter. Hotel investors should also consider the time it takes to construct a hotel and open new hotels in June, September or October to generate enough cash flows for future operations. It might worth postpone opening dates from February to March since the upside is much bigger.



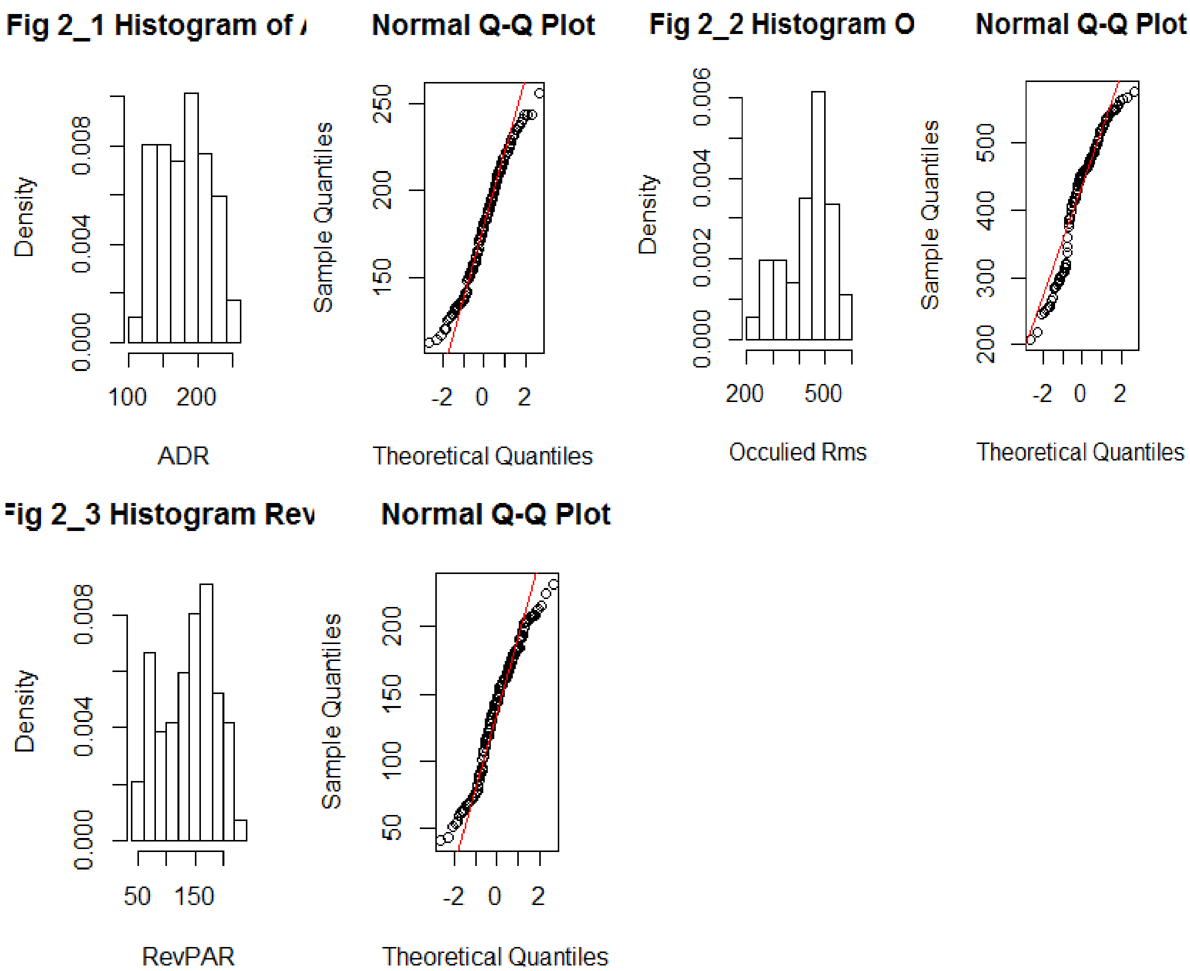


Reading the ACF and PACF chart in addition to the model itself, we know that if the previous month's RevPAR is below its average, this month' s RevPAR will fall below its average as well. This will be a useful sign for hotel operator team: if they fail to meet their monthly revenue goal this month, there is great chance that they can't meet next month's revenue goals. Therefore, they should be alert and take actions to remedy the situation right away. However, if previous month last year's RevPAR fell below its average point, the RevPAR of the month this year might be above the average.

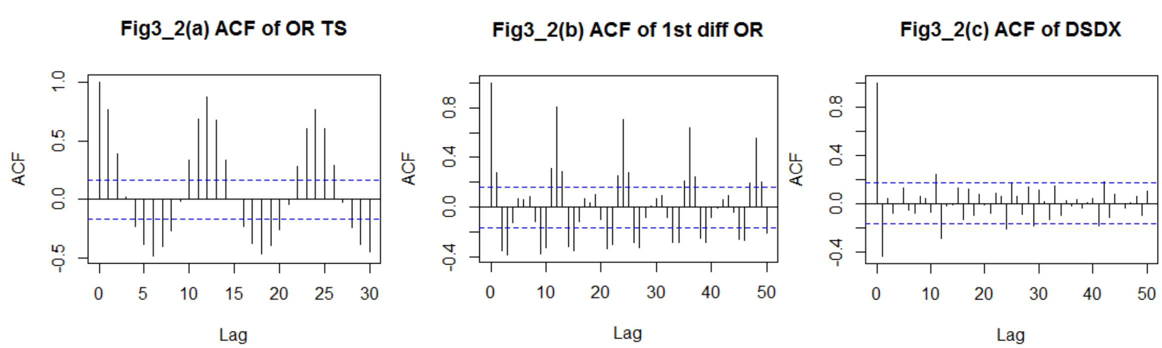
Appendix

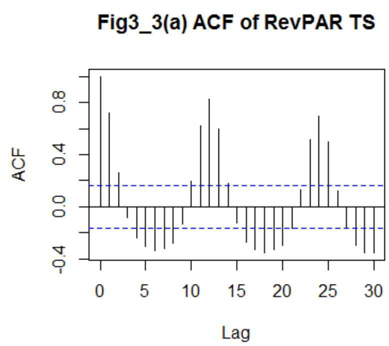
I. Graphs

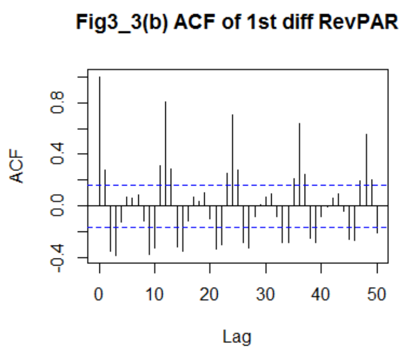
1. Histograms and QQ Plots of TS data

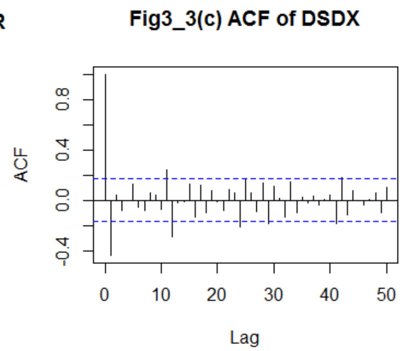


2. ACF plots of Time Series Data









II. List of 50 Hotels in the Dataset

Hotel	Open Date	Rooms	Change in Rooms
Swissotel Chicago	Aug 1988	661	Y
Kimpton Hotel Allegro	Jun 1927	483	Y
Curio Collection LondonHouse Chicago	May 2016	452	Y
Radisson Blu Aqua Hotel Chicago	Oct 2011	334	Y
Hard Rock Hotel Chicago	Jan 2004	381	
Kimpton Hotel Monaco Chicago	Jun 1958	191	Y
Club Quarters Wacker @ Michigan	Jun 2002	350	Y
Wyndham Grand Chicago Riverfront	Jun 1959	334	Y
Hyatt Regency Chicago	Jun 1974	2019	
Renaissance Chicago Downtown Hotel	Nov 1991	560	Y
The Alise Chicago	Oct 1999	122	
The Silversmith Hotel & Suites	Jun 1998	144	Y
Club Quarters Central Loop	Jun 2000	429	Y
Hyatt Centric The Loop Chicago	Apr 2015	257	Y
Chicago Athletic Association Hotel	May 2015	241	Y
Kimpton The Gray	Aug 2016	293	Y
Hilton The Palmer House	May 1925	1641	Y
Hilton Chicago	Mar 1927	1544	
Renaissance Blackstone Chicago Hotel	Mar 2008	332	Y
Marriott Downtown @ Medical District UIC	Sep 1988	113	
Autograph Collection Hotel Chicago Downtown	Oct 1998	354	Y
PUBLIC Chicago	Jun 1926	285	Y
Westin Chicago River North	Oct 1987	429	Y
Dana Hotel	May 2008	216	Y
Kinzie Hotel	Jul 2003	215	
Ivy Hotel	Jun 2012	63	Y

Autograph Collection Chicago Downtown	U/C	195	
Freehand Chicago	Jun 1927	137	Y
Embassy Suites Chicago Downtown Lakefront	Aug 2001	455	
Hyatt Chicago Magnificent Mile	Apr 1999	419	Y
Sheraton Grand Chicago	Mar 1992	1218	Y
The Whitehall Hotel	Jun 1974	222	
Joie De Vivre The Talbott Hotel	Jun 1926	149	
Warwick Allerton Hotel Chicago	Jun 1924	443	
Tremont Hotel	Jun 1976	135	
Raffaello Hotel	Jun 1979	160	Y
Marriott Chicago Downtown Magnificent Mile	May 1978	1200	Y
Hilton The Drake Hotel Chicago	Jun 1920	535	
Westin Michigan Avenue Chicago	Jun 1963	752	
Omni Chicago Hotel	Apr 1990	347	
ACME Hotel Company Chicago	Jun 1925	130	
Millennium Knickerbocker	Jun 1927	305	
Hilton Suites Chicago Magnificent Mile	Jul 1990	345	
The James Hotel Chicago	Jun 1985	294	Y
Seneca Hotel & Suites	Jun 1926	100	Y
Hotel Chicago Illinois Medical District	Apr 2016	116	Y
Joie De Vivre Hotel Lincoln	Jun 1928	184	Y
Kimpton Hotel Palomar Chicago	Mar 2010	261	Y
Embassy Suites Chicago Downtown	Aug 1991	368	Y
State & Grand By Bridgestreet	Mar 2010	145	Y
The Godfrey Hotel Chicago	Feb 2014	221	Y
	Total	21279	