Hotel Optimization

Thomas Atchison - 260986081

Emilie Seliverstoff - 260883141

Elaine Yu - 261030551

Desautels Faculty of Management

MGSC 416: Data-driven Models for Operations Analytics

Professor Rim Hariss

April 13, 2023

I. Introduction

Dynamic pricing refers to using algorithms to determine the optimal price of a good or service by accounting for fluctuations in demand and other factors which may affect pricing strategy (Lacalle, 2021). Seasonality and other factors significantly impact hotel rooms, causing variations in demand throughout the year. In order to maximize their revenues (and profits), hotel managers engage in dynamic pricing. This strategy allows hotel managers to find the best combination that minimizes the number of empty rooms while maximizing the price of each room through a better match of supply and demand (Lacalle, 2021). In fact, managers cannot remain idle by when it comes to pricing decisions as managerial decisions can greatly impact the demand. For instance, the lower the demand for the room, the lower the price of the room (and vice-versa).

We, the Bronfman Consulting Group, have been hired by a large European hotel to implement a dynamic pricing strategy. Our task? Forecasting their demand and creating an optimization model which selects the price for the different types of rooms within the hotel in order to maximize revenues.

II. Data Overview

To attain our objective of predicting demand and setting a dynamic pricing strategy accordingly, we will need a dataset that collects historical reservations of a hotel. We used an online hotel reservation dataset available on Kaggle by Ahsan Raza containing 36,275 observations and 19 columns (features) collected between 2017 and 2019 (Raza, 2023). Critical variables used include arrival year, arrival month, arrival day, lead time, number of weeknights, weekend nights, room type reserved, and average price per room. Although this data pool was initially intended to predict customer room reservation cancellations, some data preprocessing is used to achieve our goal: focusing on the non-cancelled rooms. We cleaned the data by removing any N/A values and rows in which the booking status shows "Canceled" since they are insignificant in determining the demand. We also did not include observations where the average price per room indicates an odd number that can skew our results, like $0. Before determining seasonality and implementing the predictive model for demand, we added new features such as "arrival" and "departure" to the dataset that identify the flow of the reservations. In other words, we want to determine when customers are staying. In addition, the variable "room_type_reserved" shows seven types of rooms available in the selected hotel. We dropped room types 3 and 7 since they do not have enough data for us to make non-bias predictions for them.

Methodology/Model

The first task was to create a predictive model estimating the demand. After previously mentionned data-cleaning, we searched for attributes that may impact demand in the dataset. Our first hypothesis for which variables may affect demand is seasonality. As seen in Figure 1, the demand varies primarily according to the time of the year and more granularly at the monthly level. As a result, our model will include a month factor, in addition to a week factor that accounts for the noticeable changing trend. The second potential attribute is the Room Type, where the demand largely varies according to the type of room selected (Figure 2). As mentioned earlier, we removed room types 3 and 7 due to a lack of data. Hence, we will focus on room types 1, 2, 4, 5 and 6. Then, other attributes were added to the model, such as price (Figure 3) and the price in the previous two weeks, as individuals may alter their traveling days slightly to avoid the higher cost of travel. After splitting the data into a training and a testing dataset, we built the first model (initial attributes can be found in Figure 4). Using backward stepwise regression, the best linear model, with a probability F-Statistic of 3.54e-68 and adjusted R squared of 0.690, included the following attributes: week, avg_price_per_room, peak, and dummy variables for the Room Type (Figures 5 and 6). The peak attribute is a binary variable indicating whether or not the month is during peak demand which ended up being a more significant predictor of seasonality than month. Due to the nature of the data, we realized that across many weeks arrivals were close to 0 for certain Room Types. (Figure 7) As a result, we fitted a log-linear model. (Figure 8) The best non-linear model, with a probability F-Statistic of 1.56e-98 and an adjusted R squared of 0.810, included the same variables besides the peak that was shown as insignificant. (Figure 9) However, when using the test dataset, the linear model performed better on all metrics comprising MSE, SSE, and MAPE. Therefore we will use the linear model to estimate the demand. (Figure 10)

Once the predictive model completed, we can create the optimization model. As mentioned, we aim to find which combination of prices for each room type maximizes the hotel's total revenue. The optimization will occur for 17 weeks (4 months) from 08-15-2018 to 12-15-2018 (our testing dataset). The decision variable is *Pit*, which signifies the Price of Room Type *i* at week *t*. Then, the objective function is to maximize the total revenue where the latter is the multiplication of the price (*Pit*) and the demand where the demand function, found in the previous step, is a function of price (d(*Pit*)) for each room type *i* at each week *t*. Onto the constraints, the first one is a capacity constraint which implies that the demand cannot exceed the maximum number of rooms available per room type at any time for each room type. First, to identify the maximum number of rooms available per room type, the methodology was to start at t = 0, then track the flow of demand (arrival and departure) and assume that the maximum flow at any point in time during this process would be the capacity (Figure 11). The second constraint is a "realistic" constraint where the price of each room at each time must remain between the minimum and maximum price offered by the hotel in the dataset. However, to add more flexibility to the model to better match price, demand, and supply, a 25% variation was allowed in the price. The last constraint was a "price-ladder" constraint to ensure the model remained sane. For hotels, the room types have a particular quality order where certain rooms, such as suites, must be priced higher than standard-type rooms. If not, all the demand would go to the suite. Therefore, the price of each room was found, and the conclusion was that Rooms 2 and 4 must be higher than rooms 1, 5, and 6 (Figure 12). However, again to add flexibility, the price must remain over the minimum price of the room rather than the actual price to account for certain exceptions (e.g., discounts). Of course, the last constraint would be a non-negativity constraint where the price cannot be below $0 (Figure 13).

Results and Insights

Our optimization model yielded the optimal price for room Types 1, 2, 4, 5, and 6 each week between August 15 and December 15, 2018 (our entire test data set) (See Figure 14). The model yielded a total revenue of $713,839.00 across the test data set. This objective value shows the potential maximum revenue the hotel can earn by implementing a dynamic pricing strategy every week which better matches the demand and supply. The price for room types 1 and 4 is constant at $143.01 and $181.63, respectively. Our predictive model's positive correlation between demand and average room prices can explain this phenomenon, which leads these room types to optimize at the maximum allowable price (see Conclusion and Limitations). For room type 2, 5, and 6, it is possible to see some price fluctuations, some of them as big as 64% from one week to another.

The shadow prices of constraints, except for a few, are non-binding; therefore, shows a value of 0, meaning that a change in the parameter representing the right-hand side of such a constraint does not impact the optimal value of this linear optimization model. However, increasing the maximum price threshold for room type 1 increases the overall revenue by $240.23. The same applies for room types 4, 5, and 6 by $94.28, $36.45, and $26.15, respectively. In addition, an increase in the number of rooms available for room type 2 increases the revenue by $215.83. Hence, it would be interesting for the hotel manager to analyze whether the increase in price threshold while being mindful of competitor prices is an idea that should be considered. Furthermore, the hotel can consider increasing room capacity for room type 2, if they believe that this consistent increase of $215.83 per trimesters outweighs the cost of building new rooms or most likely converting an existing room to a room of type 2.

Overall, our linear optimization model outperformed the actual performance of the hotel. In that time frame, the hotel managed to get revenues of $710,900.17, $2,938.83 less than our model. This 0.4% increase can yield the hotel $6,327.83 additional dollars of revenues per year. This may seem like a small number but over the long-run, this may bring interesting supplemental earnings from the hotel, in addition to the insights given by the shadow price.

III. Conclusion and Limitations

We recommend that Hotel Europe implement this dynamic pricing strategy and optimization model to best match supply and demand within its hotel in order to maximize its yearly revenues. Such a pricing strategy would result in prices reflecting better forecasted (therefore actual) demand which could increase their revenues by a projected 0.4%.

Throughout the optimization process, we identified certain limitations that may affect the quality of the model. The first limitation is the inconsistencies in the data, specifically for room prices. During certain weeks, the average price of the room was reported to be between $0 and $30. While this may result from a charity function or a clerical error, as we do not have the hotel's information, we cannot determine which prices are implausible and remove them all from the dataset nor can we appropriately predict them. Our second limitation is the lack of available data for specific rooms. Room types 3 and 7 had fewer entries than the other rooms and were therefore removed from the model. Our model is limited to optimizing revenue for room Types 1, 2, 4, 5, and 6. The third limitation of the model is the fact that it needs to take into account outside variables such as competitor pricing. Depending on the prices offered by competitors, Hotel Europe may lower their prices to be more price competitive and attract more consumers. However, due to the lack of competitor data, our model did not consider these constraints. Our fourth limitation is that our model does not consider consumer cancellations and instead functions under the assumption that no customer will cancel. We did not consider cancellation as we do not have information regarding the hotel's cancellation policies and to project the opportunity cost of a cancellation was out of our project score. As a result, we removed all canceled bookings when cleaning the dataset. Lastly, our model does not appear to be properly accounting for the relationship between price and demand, as it portrays it as a positive relationship in our final model. While the model correctly depicts it as a negative relationship individually, when mixed with other variables, it changes. Yet, while we attempted to de-seasonalize, account for seasonality, and other techniques, none of them gave us a positive relationship. However, since our projections are accurate and our model imposes constraints on prices, the model remains rational.

IV. Appendices

Figure 1: Time Series Line Plot of Demand over Time

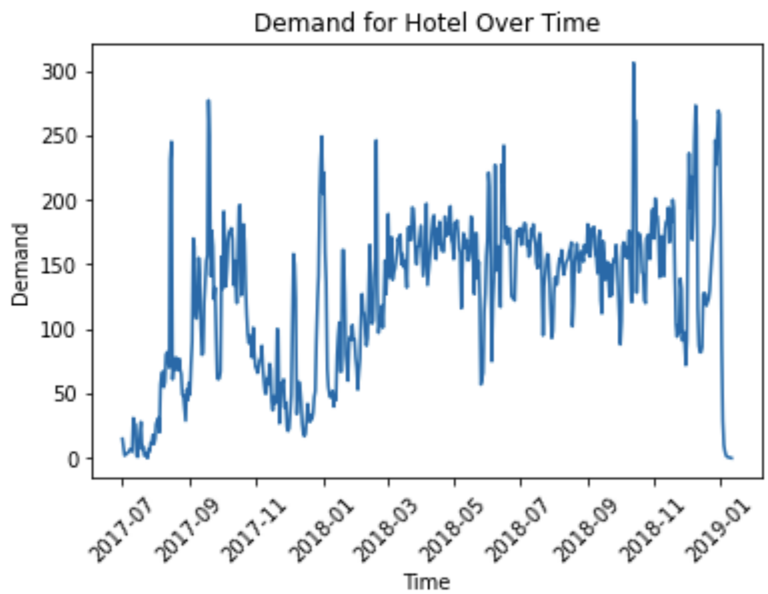

Figure 2: Demand per Room Type

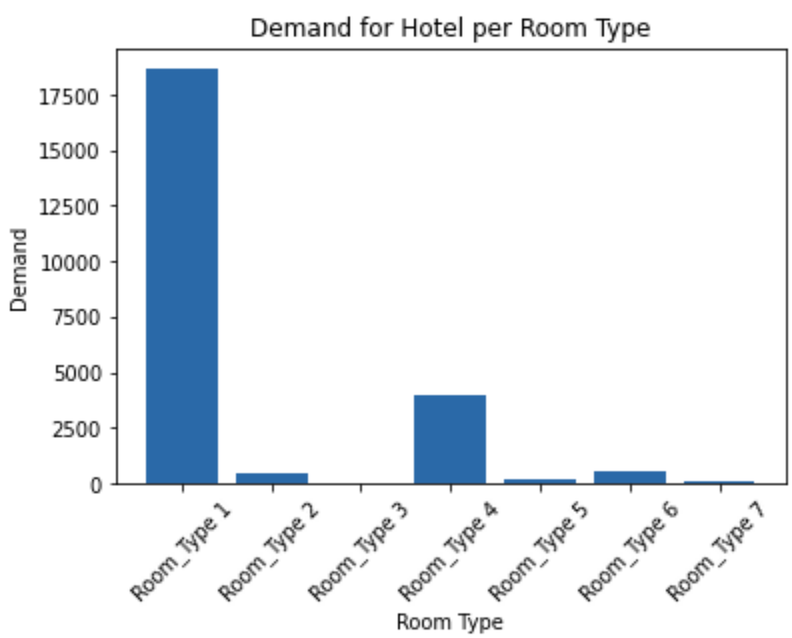

Figure 3: Demand per Price for Hotel (Room Type 1)

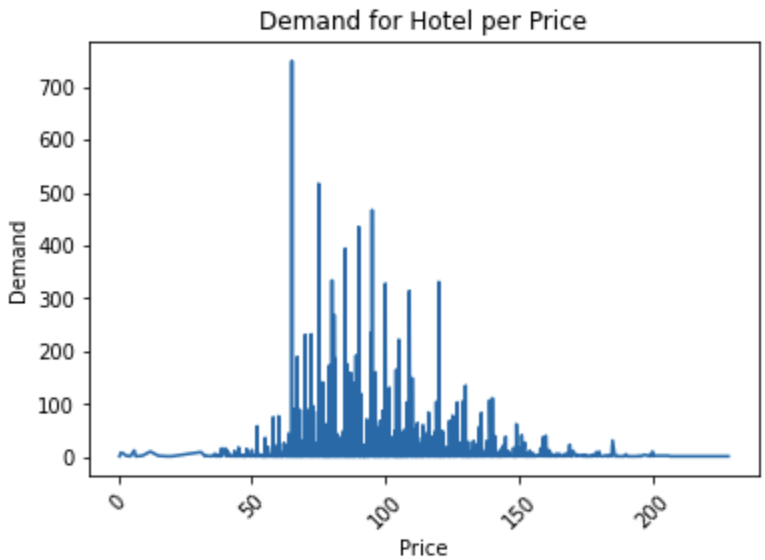

Figure 4: Initial Linear Model Equation

$$Demand = \beta_0 + \beta_1 week + \beta_2 price + \beta_3 price_{t-1} + \beta_4 price_{t-2} + \beta_5 month_m + \beta_6 roomtype_i$$

Figure 5: Variables in Final Linear Model

	coef	std err	t	P>|t|	[0.025	0.975]
const	193.6281	12.970	14.929	0.000	168.096	219.160
week	-0.8361	0.265	-3.152	0.002	-1.358	-0.314
avg_price_per_room	0.2327	0.107	2.178	0.030	0.022	0.443
peak	21.8498	8.293	2.635	0.009	5.524	38.175
Room_Type 2	-193.8957	9.909	-19.568	0.000	-213.401	-174.390
Room_Type 4	-163.9233	9.641	-17.003	0.000	-182.901	-144.946
Room_Type 5	-203.7756	11.575	-17.604	0.000	-226.561	-180.990
Room_Type 6	-214.0758	13.247	-16.161	0.000	-240.152	-188.000
Room_Type 7	-218.4495	16.196	-13.488	0.000	-250.331	-186.568

Figure 6: Final Linear Model Equation

$$Demand = \beta_0 + \beta_1 week + \beta_2 price + \beta_5 peak_k + \beta_6 roomtype_i$$

Figure 7: Count of the Demand per Week

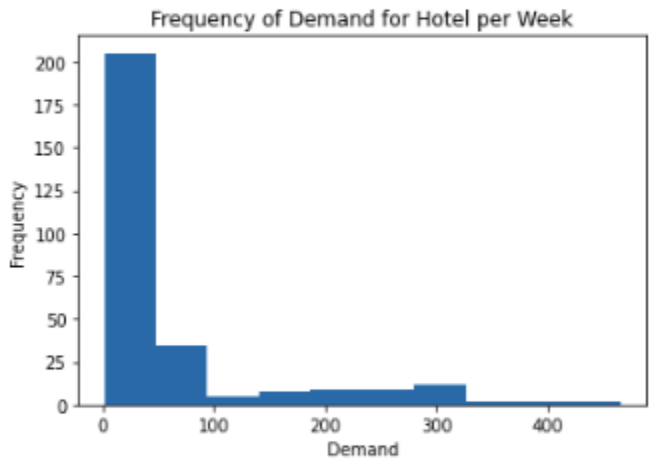

Figure 8: Non-Linear Model Equation

$$\ln(Demand) = \beta_0 + \beta_1 week + \beta_2 price + \beta_6 roomtype_i$$

Figure 9: Variables in Non-Linear Model

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.7289	0.196	24.100	0.000	4.343	5.115
week	-0.0085	0.003	-2.559	0.011	-0.015	-0.002
avg_price_per_room	0.0063	0.002	3.897	0.000	0.003	0.010
Room_Type 2	-3.5384	0.150	-23.536	0.000	-3.834	-3.242
Room_Type 4	-1.8481	0.146	-12.635	0.000	-2.136	-1.560
Room_Type 5	-4.5389	0.175	-25.865	0.000	-4.884	-4.193
Room_Type 6	-4.1283	0.201	-20.545	0.000	-4.524	-3.733
Room_Type 7	-5.2345	0.246	-21.320	0.000	-5.718	-4.751

Figure 10: Comparison of Linear and Non-Linear Model

Model	MSE	SSE	sMAPE
Linear	2,072.55	209,328.06	33.45
Non-Linear	3,354.81	338,835.82	51.48

Figure 11: Capacity of Each Room Type

Room Type	**Capacity**
Room Type 1	267
Room Type 2	15
Room Type 4	78
Room Type 5	13
Room Type 6	11

See Figure 12: Maximum Price of Each Room Type

Room Type	**Max Price ($)**
Room Type 1	143.01
Room Type 2	181.64
Room Type 4	181.64
Room Type 5	143.01
Room Type 6	143.01

Figure 13: Optimization Model

$$
\begin{aligned}
&max \sum_{t=1}^{t}\sum_{i=0}^{5} P_{it} * D(P_{it}) && \\
&s.t.\ D(P_{it}) \leq c_i \quad \forall\, i, \forall\ t && (capacity\ constraint) \\
&\quad P_{it} \leq \max\{P_{i0}, \ldots, P_{it-1}\} * (1 + 0.2) && (maximum\ price\ constraint) \\
&\quad P_{it} \geq \min\{P_{i0}, \ldots, P_{it-1}\} * (1 - 0.2) && (minimum\ price\ constraint) \\
&\quad P_{2t} \geq P_{1t} \quad \forall\, t && (price\ ladder\ 1\ constraint) \\
&\quad P_{2t} \geq P_{5t} \quad \forall\, t && (price\ ladder\ 2\ constraint) \\
&\quad P_{2t} \geq P_{6t} \quad \forall\, t && (price\ ladder\ 3\ constraint) \\
&\quad P_{4t} \geq P_{1t} \quad \forall\, t && (price\ ladder\ 4\ constraint) \\
&\quad P_{4t} \geq P_{5t} \quad \forall\, t && (price\ ladder\ 5\ constraint) \\
&\quad P_{4t} \geq P_{6t} \quad \forall\, t && (price\ ladder\ 6\ constraint) \\
&\quad P_{it} \geq 0 \quad \forall\, i, \forall\ t && (non-negativity\ constraint)
\end{aligned}
$$

Figure 14: Optimal Prices

t	i=0	i=1	i=2	i=3	i=4
33	143.01	174.70	181.63	143.01	51.85
34	143.01	177.78	181.63	143.01	51.85
35	143.01	180.86	181.63	143.01	51.85
36	143.01	98.02	181.63	135.92	143.01
37	143.01	101.10	181.63	139.00	143.01
38	143.01	104.18	181.63	142.08	143.01
39	143.01	107.26	181.63	143.01	143.01
40	143.01	110.34	181.63	143.01	143.01
41	143.01	113.42	181.63	143.01	143.01
42	143.01	116.50	181.63	143.01	143.01
43	143.01	119.58	181.63	143.01	143.01
44	143.01	122.66	181.63	143.01	143.01
45	143.01	181.64	181.63	51.85	51.85
46	143.01	181.64	181.63	51.85	51.85
47	143.01	181.64	181.63	51.85	51.85
48	143.01	181.64	181.63	51.85	51.85
49	143.01	138.06	181.63	143.01	143.01

References

Lacalle, E. (2021, April 26). *What is dynamic pricing in hotels and how does it work?* Mews.com; Mews Systems LTD. https://www.mews.com/en/blog/dynamic-pricing-hotels

Raza, A. (2023). *Hotel Reservations Dataset*. Kaggle.com. https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset