



FRANCES X. FREI
DENNIS CAMPBELL

GuestFirst Hotel (B): Taking Advantage of Panel Data

It was now early Tuesday morning, and Jeff Hausner had spent the better part of the previous day trying to make sense of his analysis of the impact of customer loyalty on RevPar. After running analysis on his yearly data, the only conclusion Hausner could make was that loyalty appeared to have no impact at all on RevPar. However, lingering concerns about the lack of site-location characteristics as control variables in the analysis made Hausner uncomfortable about making this conclusion just yet. Realizing that it would be impossible to obtain additional data before his meeting with Anne Gibson the next morning, Hausner telephoned Gina Morehouse, a former co-worker and expert on statistics and data analysis.

Another Perspective on the Data

Hausner reviewed the notes he had taken during his conversation with Morehouse who had proven incredibly helpful and may have even solved his problem regarding site-location characteristics. Not wanting to lose any of the information he had gathered, he typed up his notes and saved them on his office computer (**Exhibit 1**).

If Hausner understood his notes correctly, then he could combine the datasets he created for 1997-2000 into one large panel dataset rather than treating them as four separate datasets. Morehouse assured him that doing so would allow him to control for most hotel site-location characteristics without actually obtaining data on these characteristics. While this did not make complete sense to Hausner, he recalled Morehouse telling him that for site-location variables that did not change over time, he could instead use hotel dummy variables. She had stressed that using a dummy variable for each hotel might even be better than having the actual data in terms of removing site-location effects. Needing an intuitive explanation of why this was true, Hausner made a note to himself to look this up in his favorite Statistics book later on.

Professors Frances X. Frei and Dennis Campbell prepared this case. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Copyright © 2001, 2002, 2003, 2005, 2017 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to <http://www.hbsp.harvard.edu>. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the permission of Harvard Business School.

The First Look at the Panel Dataset

Hausner combined the four separate datasets he had created for 1997-2000 into one dataset, a panel dataset according to Morehouse, giving him 168 observations (4 observations for each of 42 hotels). He then created 42 dummy variables, one for each hotel, as Morehouse had suggested. In addition to the data he had already collected for 1997-2000, Hausner noticed that GuestFirst's performance reporting system had data on loyalty for each of the 42 hotels in 1996 (unfortunately, there was no data on RevPar or CompRevPar for 1996). Although unsure that he would need it, Hausner decided to add the 1996 loyalty data to his panel dataset. Thus, his final dataset consisted of data on all 42 company-owned hotels for each of the years from 1996-2000, with data on RevPar and CompRevPar missing in 1996. (**Exhibit 2** lists the names and descriptions of the variables in Hausner's new dataset. **Exhibit 3** contains a view of the sample data and summary statistics.)

Hausner sat down and prepared to dive back into his analysis. He hoped his new dataset and Morehouse's insights would help him make some sense of the relationship between loyalty and RevPar before his Wednesday morning meeting with Gibson. Hausner was determined to not only determine whether or not a relationship existed, but to also have a set of recommendations for what GuestFirst should do based on his results.

Exhibit 1 Notes from Discussion with Gina Morehouse on Panel Data

- Panel datasets are datasets that have observations on the same units (e.g., hotels) over multiple time periods.
- Rather than treating the datasets for the years 1997-2000 as four separate datasets, they can be combined into one larger panel dataset.
 - Instead of four datasets each with 42 observations, this will produce one dataset with $(42 \text{ hotels}) \times (4 \text{ years}) = 168$ observations.
 - It is important to order the data by hotel and then by year.
- Many site-location factors such as urban versus resort settings, number and type of tourist attractions nearby the hotel, hotel size, number of competitors, and population remain relatively constant over time. Panel data allows the use of individual hotel dummy variables to control for these and any other factors that affect RevPar that remain relatively constant over time.
 - Dummy variables for each hotel are variables that take on the values of 0 and 1. For each hotel, a dummy variable is created, that will be a 1 for the four time periods of data for that hotel, and a 0 for every other observation in the data. For GuestFirst company-owned hotels, there will be 42 dummy variables named D1-D42. D1=1 for hotel property #1 and D1=0 for all other hotel properties. Similarly, D34=1 for hotel property #34 and D34=0 for all other hotel properties.
 - Re-running the analysis with the panel dataset and including hotel dummy variables should allow the impact of loyalty on RevPar to be isolated from other influences that stay constant over time.
- Remember that panel data has a time-series dimension. This means that factors that affect RevPar and change over time have to be controlled for. These factors include changes in economic conditions, changes in a hotel's local market due to severe weather or a drop in vacation travel. Hotel dummies will not control for these time-varying effects.

Exhibit 2 Variable Names and Descriptions (All Variables Defined at the Individual Hotel Level)

Name	Description
Property	Unique hotel property identification number
Year	Year of observation. Ranges from 1996-2000.
Loyalty	Percentage of customers reporting that they are "likely" or "very likely" to return to the GuestFirst hotel of their most recent visit.
RevPar	Revenue per available room. Defined as annual hotel room revenues divided by total annual available room-days.
CompRevPar	Average revenue per available room for a hotel's competitors.
D1-D42	Dummy variables for each property. For example D1=1 for property #1 and 0 for all other properties. Similarly, D34=1 for property #34 and 0 for all other properties.

Source: Case authors.

Exhibit 3 Summary Statistics from Sample Hotels

Hotel	Year	Loyalty	RevPar	CompRevPar	D1	D2	...	D42
1	1996	41.3	N/A	N/A	1	0	...	0
1	1997	58.7	79.11	109.90	1	0	...	0
1	1998	56.5	79.41	102.30	1	0	...	0
1	1999	63.0	66.06	102.40	1	0	...	0
1	2000	52.5	84.64	111.64	1	0	...	0
2	1996	66.7	N/A	N/A	0	1	...	0
2	1997	69.8	91.66	132.40	0	1	...	0
2	1998	68.5	86.11	129.70	0	1	...	0
2	1999	73.0	85.17	130.50	0	1	...	0
2	2000	74.6	88.57	128.96	0	1	...	0
...
42	1996	38.0	N/A	N/A	0	0	...	1
42	1997	41.3	99.03	98.70	0	0	...	1
42	1998	36.9	100.59	102.40	0	0	...	1
42	1999	48.0	105.25	107.70	0	0	...	1
42	2000	41.5	108.89	95.48	0	0	...	1
Mean		54.6	111.20	107.80				
Standard Deviation		9.7	37.06	13.07				
Minimum		24.9	62.55	90.00				
Maximum		77.0	273.38	172.96				

Source: Case authors.