**DATA SETS AND STORIES**

**Café Data 2.0: New Data From a New and Improved Café**

Concetta A. DePaolo\(^a\), David F. Robinson\(^b\), and Aimee Jacobs\(^c\)

\(^a\)Operations & Supply Chain Management, Scott College of Business, Indiana State University, Terre Haute, Indiana, USA; \(^b\)Department of Management, Scott College of Business, Indiana State University, Terre Haute, Indiana, USA; \(^c\)Management Information Systems and Operations & Supply Chain Management, Scott College of Business, Indiana State University, Terre Haute, Indiana, USA

**ABSTRACT**

In this article, we revisit the café story first introduced in 2011. Recent data from this café run by business students at a Midwestern public university are explored and analyzed. The data were collected using a point-of-sale system over a three-month period during the spring semester of 2015. These data can be used in introductory statistics courses to illustrate the use of time series and forecasting, applications of data mining and visualization, as well as sampling, confidence intervals, and inference using ANOVA and chi-square tests for independence. Since the data pertain to a student-run business, we believe that statistics students, especially those in business disciplines, will find the data’s context to be relevant and interesting. In addition to the technical exercises, we provide background and context for several managerial issues that these data can be used to address, thus emphasizing the importance of data-driven decision making.

1. **Introduction**

We, like many undergraduate statistics instructors, prefer whenever possible to provide real data for our students to analyze. This approach is consistent with the Guidelines for Assessment and Instruction in Statistics Education (GAISE), which state “It is important to use real data in teaching statistics to be authentic, to consider issues related to how and why the data were produced or collected, and to relate the analysis to the problem context” (GAISE College Report 2010, p. 16).

Several authors have discussed benefits of using real, contextualized data for the teaching and learning of statistical topics. Benefits include that real data immediately provide credibility to the usefulness of learning analytical methods as a way to “understand the world” (Mvududu and Kanyongo 2011; Hasbrouck, Deniz and Hodges 2014) and that they promote motivation, interest, and engagement (Neumann, Hood and Neumann 2013). Other studies have shown that use of real-life data and problem-solving contexts promote learning and understanding of statistical concepts (Neumann, Hood, and Neumann 2013) and improve student attitudes (Tsao 2006) and confidence (Hasbrouck, Deniz, and Hodges 2014) with regard to statistics.

In addition, coming from an institution that emphasizes experiential and service learning, we strive to provide our students with consulting and service projects involving data analysis as much as possible. We emphasize that data are not analyzed in a vacuum and that there are always questions to be answered and decisions to be made based on analyses. This helps promote the usefulness of statistics in students’ minds.

With these principles in mind, instructors sought to collaborate with our college’s student-run business—The Executive Express Café—to obtain and analyze sales data. Partnership with the café created a mutually beneficial opportunity to provide meaningful, relevant data for students in our classes to analyze, as well as provide information for managerial decision making for the café.

We provide background and details about the café in Section 2. In Section 3, we discuss the operational challenges and issues that the café management is facing and in Section 4, we describe the data obtained from the point-of-sale (POS) system that are used in these analyses. In Section 5, we detail pedagogical uses of the data within a business context, including applications of time series forecasting, data mining and visualization, sampling, confidence intervals, and inference with ANOVA and chi-square. Finally, in Section 6, we conclude.

2. **The Executive Express Café**

2.1. **Background and History**

The Executive Express Café (the café) was founded in the fall of 2009 and is run by undergraduate business students in our college, which emphasizes experiential education and hands-on learning. Initially, the café carried a limited number of items (sandwiches, cookies, drinks, chips, etc.) and record-keeping was done “by hand” using Excel. Sales data for particular products and overall dollar sales were explored in the original Café Data paper (DePaolo and Robinson 2011). Since its inception,
the café has been operating on a semester-by-semester basis and is completely closed during the summer.

The College of Business relocated into a restored 1930s era Federal Courthouse in August 2012, and the new café opened in September 2012. The new location is a more spacious and professionally designed space in the lower level of the building where most of the classrooms are located. An immediate impact in sales was evident in that daily sales climbed by almost 30% within a few weeks of opening the new location. Sales have improved as new products, including espresso-based coffee drinks such as lattes and mochas, have been introduced into the café and publicity surrounding the café’s location has helped customers find the café despite having no signs visible from outside the building. With comfortable seating at tables, tall counters, and booths, the café area is now a hub of student social activity, team meetings, and studying. The two key goals of the café remain the same: (a) to provide a great service to the college and the university, and (b) as a business-learning laboratory for the students who operate the café and who carry out café research projects in other classes.

In January 2015, the café switched to an improved point-of-sale (POS) system that tracks sales of all items on a per-transaction basis and reports these data in the form of spreadsheet data compatible with QuickBooks® accounting software by Intuit. The new POS system allows the students to have precise data for the first time since opening the café. When sales volumes were smaller, adjustments in stock levels and order sizes could be improvised. As sales have grown, managing the supply chain and keeping waste under control have become more difficult. Keeping the shelves full of product has proven to be the biggest problem. The café is frequently stocked out of items because students try to limit daily waste of expiring food, customer consumption varies, and suppliers experience control problems.

In addition to this concern, teams working to improve the café have identified other operational issues they would like to address. For example, management is very interested in how sales and customer traffic varies throughout the week for staffing and purchasing purposes. Other operational issues, as well as additional management decisions that could be better informed by understanding the café’s data, are described in more detail in Section 3.

### 2.2. Café Operations

Undergraduate students may choose to be involved in the Café as part of their capstone business experience; therefore, the first week of the semester involves orientation to the “course” and short training sessions on café equipment (e.g., coffee machines and the POS system). Training is conducted by volunteers who staffed the café in previous semesters. Teams are formed and leaders are designated during the second week. The café has normally opened to the public during third week of the semester, with further on-the-job training of each shift. Some class time is also devoted to training sessions.

In the most recent semester, a new startup procedure was piloted in which three assistants from the previous semester’s operation were hired to take charge of training and startup. This was done in part to get the café open sooner in the semester and also to ease some of the training burden shouldered by the professor in charge of the course that both operates the café and conducts consulting projects. The pilot startup plan succeeded and the café opened during the second week of classes, a week earlier than usual.

Normal hours for the café are from 8 a.m. to 3 p.m. Staffing consists of two or three students in the café space with one person who picks up and delivers food items such as pizza. All café volunteers operate the POS system, prepare hot drinks, and restock items as they run out. Some volunteers specialize in closing shift operations such as cleaning the equipment and balancing the nightly deposits. Outside of the café shifts, students make strategic decisions on new items, publicity campaigns, managing social media, accounting, work flow, new equipment, and scheduling.

The café operates mainly as a convenience store selling food that is prepackaged or served by the customers themselves. All volunteers are trained in good food handling practices and comply with state regulations for food service. No food items are prepared in the café itself. Drinks are prepared in the form of café lattes, espressos, tea, hot chocolate, and brewed coffee.

### 2.3. Café Menu

The café provides drinks and food at reasonable prices for students, faculty, and staff. Hot drinks feature lattes, mochas, hot chocolate, and hot tea. Brewed coffee and iced tea is also served. The café features Pepsi fountain drinks in compliance with the campus-wide exclusive agreement. There is also a variety of flavors of Gatorade in bottles as well as bottled water, a particularly strong seller. The café has experimented with selling some energy drinks but they have not proven popular.

The café serves cinnamon rolls and bagels as well as cookies and a variety of granola bars. Lunch is by far the most popular meal, serving mainly pizza, sandwiches, chips, and a variety of salads and yogurt parfaits. Several pizza restaurants have provided pizza for the café. The most popular is Papa John’s, which is served daily. Domino’s pizza was featured early on but students showed a strong preference for Papa John’s so Domino’s was discontinued. A local gourmet pizza is quite popular and is currently featured two days a week.

In order to provide a variety of products throughout the week, the supply chain team has arranged to feature Chick-fil-A chicken sandwiches every Wednesday. Their strategy is to use popular national brands and local favorites as featured products. To build interest in the café and to attract customers from neighboring offices and from other locations on campus, using featured brands is a centerpiece of the café’s growth strategy.

### 3. Motivation and Purpose of Data Analyses: Operational Problems and Questions

The student managers are interested in making sense of some of the experiences, trends, and data they have observed while operating the café. Some of the questions that the managers would like to address involve daily sales and payment methods. These operational issues are described in the following section.
3.1. Understanding Daily Sales and Donations

3.1.1. Sales of Specific Items by Day of the Week
Managers do not have a clear picture as to whether there are certain items that seem to sell better on certain days. If there is a weekly cycle, this could be useful in purchasing so that there is less chance of stock out if the café were to run short of an item. On the other hand, having slow sales of particular items on specific days of the week could help the students buy fewer items or delay their purchase to limit the number of items that expire and are wasted.

Also, when negotiating prices and terms of sale, the café is seen as a very small customer by a number of its suppliers. In order to negotiate with these suppliers, it is essential to know how many of any one item the café sells. For example, if there are eight slices of pizza in each large pizza, how many pizzas does the café sell in a day (depending on the day of the week), in a week, a month, a semester? These totals will be helpful in establishing set purchasing schedules and quantities. The total number of pizzas sold will enable the café team to negotiate better pricing as well.

3.1.2. Overall Sales by Day of the Week
The students have noted some differences in the level of sales on various days of the week due to class scheduling in the building. There is anecdotal evidence that more classes meet on Tuesday and Thursday than on other days (most will be either Tuesday and Thursday, or Monday, Wednesday, and Friday at our university), and in fact, it is easy to see that sales on these days tend to be higher than other days of the week. However, managers would like to understand and estimate the fluctuations in sales between days of the week for staffing purposes so that long customer queues can be avoided.

3.1.3. Patterns in Customer Charitable Donations
Currently, a small glass jar sits on the countertop where customers check out. The team does little to promote donations simply leaving the jar in a place where it is visible. This has not always been the case, in fact, several years ago shortly after the earthquake in Haiti, students actively solicited donations for earthquake relief and raised nearly $500 dollars in one semester. In another semester, the Boys and Girls Club received a similar donation. However, in recent semesters, the café has not always had a specific charity for which it has solicited donations long term.

It has been observed that charitable donations have declined over time, in spite of the increased sales volume in the café since moving to the new building in 2012. One possible reason is attributed to the declining use of cash for purchases; in early operations dating back to 2010, cash accounted for approximately 70% of sales. Anecdotal reports suggest that cash sales account for substantially less than 70% of sales now. Managers are wondering if the daily donations are simply a function of cash sales in dollars for that day, or are they related to total sales that include both credit and cash transactions? Marketing team members have varying opinions of why donations have dropped, including loss of novelty, lack of publicity about donations, or even who is staffing the café that day. All of these factors are unrelated to sales. Managers would like to analyze these trends, determine any correlations with sales, and explore if donations fluctuate significantly between days of the week.

The managerial questions (MQs) regarding total daily sales and donations data can be summarized as follows:

MQ1: How do sales of highest selling items fluctuate between days of the week? Are sales of specific items increasing or decreasing over time? How can this information be used for inventory and purchasing decisions?
MQ2: What are the busiest days of the week in terms of sales? How do overall sales fluctuate between days of the week? Are sales increasing or decreasing over time? How can this information be used to inform staffing decisions?
MQ3: Is there a significant downward trend in charitable donations over the semester? Do donations vary by day of the week? Are donations correlated with either cash sales or total sales (cash and card) for that day? How can this information be used to increase customer charitable donations?

3.2 Understanding Payment Methods and Patterns
The café currently accepts both credit/debit cards and cash. The café teams have debated how to alleviate long customer queues and there is a faction pushing for an additional cash register to handle cash transactions. Another group is looking to solve the problem by adding a credit–debit “card only” capable tablet that could accept transactions but not handle cash. Certain purchases could then be directed to the cashier with the tablet, or even a self-service cashier station could be installed if no cash will be exchanged. This would reduce the time customers would spend waiting to purchase. The managers would like to understand how many transactions and how much of their sales are garnered by each payment method to help determine which, if either, of these suggestions might be most beneficial. They would also like to explore any differences in number of items or total dollar amounts between cash and non-cash payments.

The managerial questions (MQs) regarding payment methods can be summarized as follows:

MQ4: Does the percentage of transactions that are made by cash versus card differ depending on the day of the week?
MQ5: Do customers who use cards buy more or less per transaction than those who use cash?
MQ6: Do you recommend that an additional register (handling both cash and credit/debit cards) be purchased? What about a card-only tablet for processing card transactions?

4. Data

4.1. Data Collection
Through 2014, sales data for the café were tracked manually (using Excel), but in late 2014, the café used operational proceeds to purchase a QuickBooks® POS system. Beginning in January of 2015, the café began tracking and reporting sales...
data from the new POS system. The POS provides detailed records of each day’s transactions, including items, quantity, price, and payment method. At this point, the POS system will not track inventory, however, it has the capability to do so and will be used in the future. Sales to students are considered tax exempt; however, faculty and staff must pay sales tax.

The POS software allows a PC to function as a cash register, incorporating a cash drawer from the POS package with a PC already owned by the college. Although the system is capable of operating with a touch-screen display, for the first semester the café used a normal flat-panel monitor to display transactions. Bar codes for most products are kept on a sheet near the register and a handheld barcode scanner is used to log the purchased items into the POS system. The cashier indicates if the transaction is credit-debit or cash. Each item is then recorded in the POS system memory as a sale with the item description, quantity, and dollar amount recorded. Credit/debit transactions require an additional operation and must be “swiped” in a card reader terminal connected via phone line to the credit card company. Charitable donations consist of spare change left by customers as they pay for their purchases. No provision exists at this time to take donations from a credit or debit transaction. Donations are counted daily as part of the daily sales report that is sent to the university Controller.

To keep the café cash account under proper control, a daily reconciliation and report are prepared for the university Controller’s office and café management. At the end of each day, the POS cash register is totaled and a daily report records the day’s sales activity. Students working the last

| Date          | Qty Sold | Total | Payment       | No. Of Items | Tax |
|---------------|----------|-------|---------------|--------------|-----|
| 3/26/2015     | 3        | 4.50  | Credit Card   | 3            | 0.00|
| PPIN 1 Slice  |          |       |               |              |     |
| PPIN 1 Slice  |          |       |               |              |     |
| Gatorade Orange |      |       |               |              |     |
| 3/26/2015     | 3        | 3.00  | Credit Card   | 3            | 0.00|
| PPIN 1 Slice  |          |       |               |              |     |
| PPIN 1 Slice  |          |       |               |              |     |
| CMB Med Drink |          |       |               |              |     |
| 3/26/2015     | 2        | 3.25  | Credit Card   | 2            | 0.00|
| Baesler’s RB Sandwich | |       |               |              |     |
| Medium Soda (20oz) | |       |               |              |     |
| 3/26/2015     | 3        | 3.00  | Credit Card   | 3            | 0.00|
| PPIN 1 Slice  |          |       |               |              |     |
| PPIN 1 Slice  |          |       |               |              |     |
| CMB Med Drink |          |       |               |              |     |
| 3/26/2015     | 1        | 0.75  | Credit Card   | 1            | 0.00|
| Chips Fritos Chili Cheese | |       |               |              |     |
| 3/26/2015     | 3        | 4.28  | Credit Card   | 3            | 0.28|
| Baesler’s RB Sandwich | |       |               |              |     |
| Sun Chips - French Onion | |       |               |              |     |
| Medium Soda (20oz) | |       |               |              |     |
| 3/26/2015     | 3        | 4.15  | Credit Card   | 3            | 0.00|
| Bagel Cinnamon Raisin | |       |               |              |     |
| WP 1 Slice    |          |       |               |              |     |
| WP 1 Slice    |          |       |               |              |     |

Figure 1. POS transaction data.
shift count all of the cash in the register, add in any donations, and hold out the beginning balance used to start the next day. The excess cash remaining is then deposited in the controller’s office.

In order to generate the raw data used in this case study, the café director downloaded records of each transaction from the POS into an Excel spreadsheet. Figure 1 shows what the raw data from the POS system look like before they are cleaned and manipulated for analysis.

Donations were not recorded in the POS system so donations data were gathered from the daily sales reports submitted to the Controller’s office.

### 4.2. Time Series Data

The “Time Series Data” file contains 63 time series records covering the dates January 22 through April 27, 2015. The café sells more than 150 items, but only 14 of the top selling items are shown in this dataset, each in a separate column. Each entry indicates the number of that type of item sold on a particular day. There are several missing values from Friday, March 13, 2015 (t = 37). On this date, the Friday before the university’s Spring break, the café was closed. An excerpt from the data file is shown in Table 1.

This dataset has 17 columns:
- Date,
- DayOfWeek shows which day of the week that date fell on,
- Time_t identifies the period (day),
- 14 additional columns tell how many items of each type were sold each day.

### 4.3. Daily Sales and Charitable Donations Data

The “Daily Sales and Donations Data” file contains 63 time series records showing the total sales and cash sales (in dollars) as well as the amount of customer charitable donations each day from January 22 through April 27, 2015. We note that data for Friday, March 13, the café was closed, and therefore data for that date is completely missing from the file. An excerpt from the data file is shown in Table 1.

There are six columns in this dataset:
- Date is the date of the transaction,
- DayOfWeek indicates the day of the week (Monday through Friday),
- Time_t identifies the period (day),
- TotalSales shows the total dollar amount of sales for that date,
- CashSales shows the total dollar amount of sales paid for in cash that day,
- Donations shows the amount of cash donated by customers.

### 4.4. Transactions Data

The “Transactions Data” file contains 7288 records and displays summary information for each transaction. The first several rows of the dataset are shown in Table 3.

There are six columns in this dataset:
- TransactionNum is an identifier for the rows of this dataset,
- Date is the date of the transaction,
- DayOfWeek tells on which day of the week (Monday through Friday) the transaction occurred,
- Qty Sold is the number of items purchased during the transaction,
- Total is the total dollar amount of the transaction,
- Payment is the method of payment for the transaction, taking four possible values: Cash, Credit Card, Debit Card, or Split (if the customer used a combination of methods to make the purchase).
5. Pedagogical Uses

This section will present example pedagogical uses of the café data. The example data analysis provided in this section will help students better understand the buying habits and preferences of the café clientele and will help inform decision making for the café’s management team. We will use time series forecasting methods as well as data mining and visualization to address the managerial questions proposed in Section 3. We also suggest some examples of using the data to illustrate concepts of sampling, confidence intervals, and inference using ANOVA and chi-squared tests for independence.

5.1. Time Series Analysis and Forecasting Applications

The first pedagogical use involves time series analyses to address managerial questions 1, 2, and 3 regarding sales patterns for the most frequently purchased items, overall sales totals, and customer charitable donations. We begin with the “Time Series Data” file. This file shows café sales over a period of about 12 weeks between January and April of 2015. The café is open 5 days a week (Monday through Friday) and we expect, based on observation, to find seasonality within each week. These daily fluctuations may be attributed to differing class schedules of students and faculty who do not always have the same classes each day of the week and therefore may be in the building some days and not on others. For example, the time series plot of the number of pizza slices is shown in Figure 2. An initial inspection of this graph shows a definite seasonal pattern that repeats every 5-day work week. Specifically, if we examine the first occurrence of Monday through Friday (see circled data points below and recall that this dataset begins on a Thursday), we observe a saw-tooth pattern suggesting sales on Monday, Wednesday, and Friday are lower than Tuesday and Thursday and that this pattern persists throughout the dataset.

HELPFUL HINT: We usually like to initiate a discussion about WHY we would expect to see seasonality in the sales of specific items as well as for overall sales. Since students are familiar with traffic patterns and class schedules on different days of the week, and they themselves may not be in the building every day of the week, they are usually able to explain why seasonality occurs and to make a guess about which days will be higher or lower. Once students recognize the presence of seasonality, we then prompt the class to consider appropriate forecasting methods. In our introductory business statistics course and in our discussion below, we use time series decomposition.

ALTERNATE APPLICATION: Depending on your course’s topic coverage, there may be different approaches that you would like your students to use to model these seasonal time series. A more advanced exercise might involve modeling the series with more than one approach and then evaluating alternatives and recommending a model.

POTENTIAL PITFALL: The first day for which data were collected is on a Thursday, not a Monday. This is important for students to realize when interpreting which seasonal indices are associated with which days of the week.

POTENTIAL PITFALL: We call attention to the fact that there is one day for which data are not reported. On Friday, March 13, 2015, the café was closed as it was the Friday before the university’s spring break.

HELPFUL HINT: We find this dataset provides a good opportunity to discuss missing, zero, and unreported data points in time series.

First, students should recognize the “true” or “actual” values for sales for the day the café was closed are all zeroes. Discussion may then be directed to why these values may be left as unreported or intentionally omitted as opposed to entering the “true” zero data points.
Although not in this particular dataset, the instructor may wish to take this opportunity to discuss data points that are mistakenly entered and obviously not reasonable (e.g., that 1000 sandwiches were sold in a day). Then, students may be prompted to consider how data might be truly “missing” (i.e., there was activity but the values are unknown). This could be caused by human error, computer glitches, or perhaps power was out one day and sales were tracked manually without the POS system.

Usually at least one student will recognize that using the true zero values for this day would be implying in our analysis that the café was open and just did not have any sales. Students usually can intuitively recognize that using zero values in this case would skew results. Similarly, it is not difficult for them to realize that obvious errors/outliers will have similar effects, and that it is reasonable to omit these values from our analysis.

At this point, we find it is important to make sure students understand why deleting or ignoring a missing or unreported observation in a time series is not appropriate, and then discuss options for filling in missing values that will not greatly skew results. Offentimes students will suggest a mean, so the conversation can be directed to the seasonality of these data, and can lead students to conjecture or recognize that replacing missing values with the average of the previous and following same-season values should not have any undesirable effects on the solution and is a justifiable (though not the only) approach.

ALTERNATE APPLICATION: If necessary and depending on the level of the course, the instructor may instead wish to supply the “missing” values prior to supplying the data to students.

### 5.1.1. Sales of Specific Items

Recall that MQ1 stated: “How do sales of the highest selling items fluctuate between days of the week? Are sales of specific items increasing or decreasing over time? How can this information be used for inventory and purchasing decisions?”

Questions for students:

1. Given the context, that items are sold Monday through Friday on a college campus, discuss how and why you MIGHT expect sales to vary from day to day.
2. Refer to the “Time Series Data” file. Examine the data representing the number of pizza slices sold each day (pizza is the top-selling item at the café) and create a time series plot of the data. Qualitatively discuss your observations about the seasonality and trend, if any, in this plot.
3. Identify missing/unreported data points. Discuss and justify how you will address these missing values.
4. Select a time series forecasting method that is appropriate for this time series and create a forecasting model. Describe the attributes of each model, including trend, seasonal indices, and accuracy measures.
5. What other approaches might be used to analyze these data? For example, try separating the data by day of the week and treat the group of observations for a given day as its own time series. Graph these five time series. Do these graphs cause you to reevaluate your findings? Discuss any new insights.
6. Use your results to address MQ1. How can this information be used for inventory and purchasing purposes? Be specific in your recommendations.
7. (OPTIONAL) Repeat this analysis for other top-selling items at the café.

ALTERNATIVE APPLICATIONS: Depending on the course and assignment, you may wish to have students analyze additional or different items (there are 14 time series available). On take-home projects, we have used the approach that different time series are assigned to different students to prevent unauthorized collaboration.

Pizza is the top-selling item at the café, followed by medium sodas, chips, large sodas, sandwiches, and bottled water. All of these items (along with total sales, explored in the next section) display daily seasonality. In our introductory classes, we use seasonal decomposition to forecast seasonal time series, so that is the method we present here. When choosing between multiplicative and additive seasonality, we note that for the majority of the time series in these data (individual items, sales, and donations), a multiplicative model results in smaller mean absolute percent error (MAPE) so these are the models we present here.

We describe results for pizza and sandwiches, which are both perishable items (compared to the others with longer shelf lives) and as such, the café’s staff has a greater interest in accurately forecasting their sales. In addition, we will find that these time series act differently enough so that examining both provides a richer discussion of the data and its implications.

Figure 2 above displayed the time series plot for number of pizza slices sold per day. We note that a visual inspection suggests seasonality and perhaps a slight downward trend. Figure 3 shows the plot for Sandwiches-Assorted. Here, we notice seasonality and possibly an upward trend. The multiplicative decomposition models for both products are shown in Table 4. We filled in missing values for Friday, March 13 using the means of the values for Friday, March 6 and Friday, March 27 and used Minitab 17 to produce the output shown.

Pizza slices sold shows a meaningful downward trend of 1.22 per day ($t = -5.41$, $df = 61$, $p < 0.001$) or about six slices per week. We believe that the decrease can be explained by weather; as the spring semester progresses and warmer weather arrives, students are less likely to want hot pizza. The trend for sandwiches is not significant. There is a great deal of daily seasonality in the data and these two items show similar patterns. The busiest days are Tuesdays or Thursdays, followed by Monday and then Wednesday, with Fridays being the slowest days by far. These variations make sense in terms of the number of classes, and therefore students and faculty in the building, with a larger number of classes on Tuesday and Thursday, followed by Monday and Wednesday, with fewer courses meeting Monday, Wednesday, and Friday.

It is important to take note of the accuracy measures. The sandwiches can be kept refrigerated for 2 to 3 days, so the errors here (less than three sandwiches per day, on average)
are not especially troubling. However, any unsold pizza must be discarded at the end of the day; thus, the MAD of more than 23, indicating that the average errors are almost three pizza pies per day, appear to be more of a concern for the café’s management.

When we examine each day of the week separately, we obtain additional information. Figure 4 shows each product as five separate time series by the day of week. The results for sandwiches provide no additional information; it turns out that when each day is examined separately, we still find no significant trends and can observe the differences in sandwiches sold by day. On the other hand, some interesting information is discovered when pizza is examined separately for each day. Pizza does show a downward trend on most days of the week, as suggested by the full time series model; however, on Thursdays there is not a significant trend. (It can be shown to be positive 4.2 slices per Thursday, with a p value of > 0.10.)

Therefore, when addressing MQ1, conclusions regarding inventory and purchasing might include suggestions such as purchasing more pizza on Tuesdays and Thursdays, less on Mondays and Wednesdays, and very little on Fridays, and using the forecasting model to generate forecasts for each day. Also, we can say that during the spring semester, on days except for Thursdays, the staff should begin to order less pizza as the semester progresses. The trend suggests sales of about six fewer slices per week, which would be about one pie per week less. (Of course, during the fall semester, these patterns may not hold and need to be studied.)

As for sandwiches, many more should be purchased on Tuesdays, followed by Thursdays, with fewer on Mondays and only a very few on Wednesdays and Fridays. Again, the forecasting model can be used to estimate sales and accuracy measures suggest that waste will not be substantial.

### 5.1.2. Total Sales

For the purposes of addressing MQ2, we refer to the “Daily Sales and Donations Data” file. Recall that MQ2 asks: “What are the busiest days of the week in terms of sales? How do overall sales fluctuate between days of the week? Are sales increasing or decreasing over time? How can this information be used to inform staffing decisions?”

Questions for students:
1. Given the context, that the café operates Monday through Friday on a college campus, discuss how and why you MIGHT expect sales data to vary from day to day.
2. Refer to the “Daily Sales and Donations Data” file. Examine the time series representing Total Sales and create a time series plot of the data. Qualitatively discuss your observations about the seasonality and trend, if any, in this plot.

### Table 4. Decomposition models for pizza slices and sandwiches sold.

| Model Type          | Length | Fitted Trend Equation                  | Period Index | Seasonal Indices | Accuracy Measures |
|---------------------|--------|----------------------------------------|--------------|------------------|-------------------|
| Pizza Slices        | 63     | $Y_t = 155.03 - 1.223 \times t$       | 1  $0.81736$ | 2  $1.48217$     | MAPE 25.18        |
|                     |        |                                        |              |                  | MSD 1244.30       |
| Sandwiches-Asst     | 63     | $Y_t = 8.93 + 0.0674 \times t$        | 1  $0.79287$ | 2  $2.00123$     | MAPE 26.7647      |
|                     |        |                                        |              |                  | MSD 26.7467       |
|                     |        |                                        |              |                  |                   |

Figure 3. Time series plot of number of sandwiches sold per day.
3. Identify missing data points. Discuss and justify how you will address these missing values.

4. Select a time series forecasting method that is appropriate for this time series and create a forecasting model for Total Sales. Describe the attributes of the model, including trend, seasonal indices, and accuracy measures.

5. What other approaches might be used to analyze these data? For example, try separating the data by day of the week and treat the group of observations for a given day as its own time series. Graph these five time series. Do these graphs cause you to reevaluate your findings? Discuss any new insights.

6. Use your results to address MQ2. How can this information be used for staffing purposes? Be specific in your recommendations.

We begin this analysis with a time series plot of Total Sales (see Figure 5). We observe visually that there could be a slight positive trend, and that seasonality appears to be present.

The multiplicative decomposition model for total sales (in dollars) is shown in Table 5. We used Minitab 17 for this analysis and, because we expect seasonality to repeat each week, we filled in the missing value for Friday, March 13 using the mean of the values for Friday, March 6 and Friday, March 27.

The trend component for Total Sales is small, and can be shown to be insignificant. However, there is noticeable seasonality. Thursday is the busiest day, followed by Tuesday and Wednesday, which have indices that are very near one another, and Fridays are the slowest days by far. Again, these variations are consistent with what we know about the number of classes, and therefore students and faculty in
the building on various days of the week, with a larger number of classes on Tuesday and Thursday, followed by Monday and Wednesday, with many fewer courses meeting Monday, Wednesday, and Friday. The accuracy measures indicate that errors were about 15% (MAPE) or about $46, on average, which is fairly accurate.

Next, we examine how the time series appears if each day of the week is considered separately. Figure 6 shows the time series plot, by day of week, generated with Minitab 17.

A visual inspection of these graphs suggests that while total sales on Thursdays appear to be trending upward, the other days of the week do not seem to exhibit this same pattern. In fact, it can be shown with trend analysis that Thursdays have a significant, positive trend of about $15 per Thursday, while the other days show no significant trends.

All of the findings for Total Sales can be used to inform staffing decisions (MQ2) as follows. The café should have the most staff on Tuesdays, Wednesdays, and Thursdays. On Mondays, they can expect Sales to be about 28% below the trend, and can plan to cut back on staff. Fridays are much slower than other days, with Total Sales on average of less than half of the trend. This may indicate that only a skeleton crew is necessary on Fridays. In addition, while staffing levels may remain steady on Mondays, Tuesdays, Wednesdays, and Fridays throughout the semester, it may be necessary towards the end of the semester to increase staffing levels on Thursdays as sales tend to increase.

5.1.3. Charitable Donations
Recall that MQ3 asked: “Is there a significant downward trend in charitable donations over the semester? Do donations vary by day of the week? Are donations correlated with either cash sales or total sales (cash and card) for that day? How can this information be used to increase customer charitable donations?”

Questions for students:
1. Given the context, that the café operates Monday through Friday on a college campus, discuss how and why you MIGHT expect the data to vary from day to day.
2. Refer to the “Daily Sales and Donations Data” file. Examine the time series representing Donations and create a time series plot of the data. Qualitatively discuss your observations about the seasonality and trend, if any, in this plot.
3. Identify missing data points. Discuss and justify how you will address these missing values.
4. Select a time series forecasting method that is appropriate for this time series and create a forecasting model for Donations. Describe the attributes of the model, including trend, seasonal indices, and accuracy measures.
5. Create scatterplots to compare Donations versus Cash Sales and Donations versus Total Sales. Run correlations to determine if there is a significant correlation between Donations and either Total Sales or Cash Sales.
6. Use your results to address MQ3.

| Table 5. Decomposition model for total sales. |
|-----------------------------------------------|
| TotalSales                                      |
| Multiplicative Model                            |
| Fitted Trend Equation                           |
| Multiplicative Model                            |
| Length 63                                       |
| $Y_t = 321.2 + 0.472 × t$                       |
| Seasonal Indices                                |
| Period Index                                    |
| 1 0.72467                                       |
| 2 1.26104                                       |
| 3 1.24576                                       |
| 4 1.36319                                       |
| 5 0.40534                                       |
| Accuracy Measures                               |
| MAPE 15.11                                      |
| MAD 45.66                                       |
| MSD 3923.49                                     |
We begin by plotting Donations over time (see Figure 7). We observe visually that there does not appear to be a substantial trend but that seasonality may be present.

The multiplicative decomposition model for Donations (in dollars) is displayed in Table 6. Again, we used Minitab 17 and filled in the missing value for Friday, March 13 using the mean of the values for Friday, March 6 and Friday, March 27.

It can be shown that Donations are decreasing significantly but only slightly over time ($b_1 = -0.03$, $t = -2.47$, $df = 61$, $p = 0.016$), with an estimated decrease of about $0.03$ per day, or about $0.16$ per week. There is also noticeable daily seasonality in donations, which are highest on Thursday (about 25% above the trend), followed by Tuesdays at about 10% above trend. Mondays are in the middle (index very nearly 1.00) with Wednesdays only slightly lower. Again, not surprisingly, Fridays are the slowest days by far, with about one-third less in donations compared to trend. Again, these variations are consistent with traffic patterns within the building. Note that accuracy measures indicate substantial variation in donations.

We next created scatterplots for Donations versus the variables Cash Sales and Total Sales (see Figure 8) and ran bivariate correlations between these three variables (see Table 7).

The scatterplots suggest a possible positive, though not strong, correlation between donations and the sales variables. The correlation analyses indicate that donations are not significantly correlated with total sales but are moderately correlated with cash sales. If we use simple regression to predict donations based on cash sales, we find a significant ($t = 2.75$, $df = 61$, $p = 0.008$) but weak model ($r^2 = 0.11$) estimating that donations $\approx 1.03 + 0.01 \times$CashSales.

When addressing how to increase donations (MQ3), because cash sales are moderately correlated with donations,
increasing cash sales could lead to more donations. However, simply advertising and calling attention to the donations jar, especially for customers paying with cash, could work as well.

5.2. Data Mining and Visualization Applications

These data were gathered over an entire semester and constitute the population of transactions for the spring 2015 session. Therefore, we suggest some data mining and data visualization exercises using the "Transactions Data" file to address several of the managerial questions.

POTENTIAL PITFALL: For the analyses described here, the number of records in the full Transactions Dataset may lead to slow processing times, depending on the speed of individual machines and internet connections. You may wish to warn students of this possibility and, if appropriate, offer suggestions to speed up calculations for those experiencing significant delays.

5.2.1. Payment Methods by Day of Week

Recall that MQ4 stated “Do the number of transactions made by cash or card differ depending on the day of the week?”

Questions for students:
1. Refer to the "Transactions Data" file. How can the data be displayed, in graphical and tabular form, to show the number of transactions made by cash, credit card, debit card, or split payments each day of the week? What information can you discern from these displays?
2. What insights can you gain by examining the summarized data? How would you use these findings to address MQ4?

HELPFUL HINT: This is a nice opportunity for students to see data, observe the data type, and decide how to summarize and compare variables. We find that our students sometimes have trouble with this step. We encourage students to think about the data types of the variables Day of the Week and Payment Type, which are both qualitative, and if necessary lead them to the discovery that a contingency table is an appropriate way to summarize numerically.

ALTERNATIVE APPLICATION: See Section 5.3.4 for alternative approach to address this managerial question assuming that only a random sample, rather than the whole population, is available.

We begin by creating a contingency table to display summary values and also a line chart to visually display the number of transactions using each payment type by day of the week. Results of this analysis are shown in Table 8 and Figure 9. For the payment methods with larger frequencies, we also show the percentage of each day’s transactions in the contingency table.

The table and graph inform conclusions regarding MQ4. The data suggest that there is a small difference in payment patterns by day of the week, with a slightly higher number of cash transactions on Mondays and Fridays compared to Tuesdays and Thursdays when the number of cash and credit

| Table 6. Decomposition model for donations. |
|---------------------------------------------|
| **Donations**                               |
| **Multiplicative Model**                    |
| **Length 63**                               |
| **Fitted Trend Equation**                   |
| $Y_t = 3.806 - 0.0320 \times t$            |
| **Seasonal Indices**                        |
| 1 1.00275                                  |
| 2 1.10212                                  |
| 3 0.97267                                  |
| 4 1.25272                                  |
| 5 0.66974                                  |
| **Accuracy Measures**                       |
| MAPE 88.4010                               |
| MAD 1.3297                                 |
| MSD 3.6589                                 |

Figure 8. Scatterplots of donations versus cash sales and donations versus total sales.
transactions are practically equal. Another interesting finding is that the debit transactions are quite a bit larger on Wednesday than any other day. Within the context of the traffic patterns in this university building, we might conclude that those who had classes in the building on Mondays and Fridays tended to use cash more; and that there may have been a small number of customers who only frequented the café on Wednesdays during this semester who preferred debit transactions.

5.2.2. Amount Purchased with Credit Versus Cash

Recall that MQ5 asks “Do customers who use cards buy more or less per transaction than those who use cash?” Again, looking at the “Transactions Data” file, we want to examine both the number of items per transaction and the dollar amount of each transaction separated by payment type.

Questions for students:
1. Refer to the “Transactions Data” file. How can the data be displayed graphically to show the number of items sold and the dollar amount of sales using each payment type? What information can be gleaned from these graphs?
2. How can summary information about items sold and dollar amount of sales for each payment type be displayed in tabular form? What insights can you gain from this table?
3. How would you use these findings to address MQ5?

HELPFUL HINT: Again, this situation presents a good opportunity for students to see data, observe the data type, and decide how to summarize data and compare variables. We believe our students always need more practice with this task. Students usually are able to determine that they want to look at descriptive statistics for these numerical variables separated by payment type.

ALTERNATIVE APPLICATION: See Section 5.3.4 for alternative approach to address this managerial question assuming that only a random sample, rather than the whole population, is available.

We begin our analysis by displaying boxplots for both the Total Sales and the Quantity Sold separated by payment type (see Figure 10). We also compute descriptive statistics for each payment type and display the results in Table 9.

Examining these results to address MQ5, we see that the boxplots and summary measures do appear to suggest some small differences in the total amount of sales by payment method, with cash sales appearing somewhat lower in dollar amount ($2.61 vs. more than $3 per transaction) than the others. This does seem to agree with intuition that the higher the dollar amount of a transaction, the more likely a customer would be to use credit or debit instead of cash. The differences in number of items sold appear less evident.

5.2.3. Adding an Additional Register

Recall that MQ6 asked “Do you recommend that an additional register (handling both cash and credit/debit cards) be
Questions for students:
Based on what you have discovered about payment methods and patterns in addressing MQ4 and MQ5:
1. At this time, do you recommend adding an additional register to handle both cash and card transactions? Why or why not?
2. At this time, do you recommend adding a card-only tablet for processing card transactions? Why or why not?
3. Consider a hypothetical situation at some point in the future: Suppose that payment patterns have shifted substantially toward either more cash transactions or alternatively toward more card transactions. In these scenarios, what would you recommend to the café staff about adding either an additional register or a tablet?

To address this question, we recall two findings from earlier analyses. First, approximately 51% of transactions were cash sales and the other 49% were either credit, debit, or split transactions that require the use of a card-reading machine (refer back to Table 8). Second, we observe from Table 9 that cash sales tend to consist of slightly smaller total sale amounts and perhaps slightly fewer items sold.

Since the payment methods are fairly evenly split between cash and card transactions, there does not seem to be a clear indication that either a register or a tablet would be most helpful. If at some point in the future, cash transactions were to increase substantially, then an additional register makes sense; on the other hand, if credit transactions were to become a substantially larger portion of the payments, then it would seem that a tablet would be most helpful. Lastly, we suspect that customers tend to spend a little more money and possibly purchase more items if they use cards. One foreseeable conclusion that students may make is that adding a tablet to pay with credit or debit may sway some customers to use cards when they might have used cash. Using a card in order to avoid a long queue at the POS cash register may also influence them to purchase more than they originally intended.

5.3. Sampling, Estimation, and Inference Applications

It is typical for datasets found in textbooks to assume that the data are “nice” random samples from an unknown population. Because the Transactions Data represent the population of all transactions at the café during a semester as well as a non-random sample of all of the café’s transactions over all semesters, the dataset can provide statistics instructors with some teaching opportunities that are not extremely common with traditional teaching materials.

In this section, we suggest possible applications for demonstrating sampling and estimation concepts. Initially, all three café datasets can be used to spur discussion about populations, random samples, and non-random samples and can provide opportunities for students to think about contextual information in assessing the representativeness of samples. The ”Transactions Data” file can also be used for generating random samples to make point and interval estimates and then comparing to the actual population parameters. Finally, we note that random samples from this population can be extracted to provide alternative applications for students to answer real-life managerial questions using statistical inference techniques, if the instructor does not choose to provide the entire population to students.

POTENTIAL PITFALL: Before proceeding, we caution that treating transactional data over this time period as a
population may introduce concerns with some of the analyses. We acknowledge that measures such as the mean may not be stable over time and that treating observations as a population will disregard any time-related information in the data. However, as long as these issues are acknowledged, we believe the benefits of the potential pedagogical uses outweigh the limitations, especially with regard to variables such as Total (dollar amount of transactions) and Qty_Sold (how many items were purchased per transaction) that were found to not trend significantly over time.

Instructors who use these exercises may wish to discuss these limitations with students; and again, it may be helpful to also discuss why café staff might choose to proceed with these analyses to inform decision making even though they are aware of the shortcomings.

5.3.1. Populations, Samples, and Representativeness

Questions for students:

Consider the three datasets provided. The “Time Series Data” file contains information on the quantity of various items bought each day the café was open during the entire spring 2015 semester. The “Daily Sales and Donations Data” totals all of the sales and donations for each day the café was open throughout that semester, and the “Transactions Data” shows how many items, the total dollar amount, and the payment method of all transactions made during that time frame.

1. Do these data (in the three data files provided) constitute a population, a random sample, or a nonrandom sample of the sales/transactions of the café? It is possible that they could be considered a population in one sense or context, and a sample in another? Explain.

2. Consider the number of individual items in the “Time Series Data” file. Do you think these observations may be a representative sample of the number of each item sold per day for all days the café is open? Which items, and why? (Hint: Make sure you consider that some types of items may be purchased more in warmer weather and some in colder weather.)

3. Consider the Total Sales from the “Daily Sales and Donations” data file. Do you think these observations over the spring 2015 semester may be a representative sample of sales totals by day of the week, considering that from semester to semester, the course scheduling is consistent on Monday-Wednesday-Friday and Tuesday-Thursday in terms of the number and times classes are offered in the building?

4. Consider the Cash Sales from the “Daily Sales and Donations” data file. Do you think these observations can be considered a representative sample of cash sales from all days the café has been open? (Hint: What are the long-term trends in retail with respect to cash vs. credit?)

5. Consider the Cash Sales from the “Daily Sales and Donations” data file. Do you think any of these variables may be representative sample of all café transactions? Under what conditions? Explain.

6. Finally, how useful do you think these data can be to the café managers, given its limitations and considering the need for the staff to make decisions based on the data they have?

These questions provide a nice opportunity to discuss populations and samples. These data are a population of all transactions/sales of the café during the spring 2015 semester, but they may also be considered a nonrandom sample of transactions/sales over all days that the café has been open.

Because of the similar class schedules each semester and the fact that the student/faculty mix remains fairly consistent over time, the customers purchasing items during this semester might be representative of past and future customers who are in the building. Therefore, with regard to individual items sold, it may be possible to estimate other semesters based on the current semester data. These data show that products such as pizza, coffee products, and cold drinks tend to show trends over this semester (more pizza and coffee products in colder weather and more cold drinks in warmer weather), whereas other items such as sandwiches, chips, and cinnamon rolls do not tend to vary as much with weather. We might infer that these patterns could hold in future semesters.

Also, with a presumed similar customer base each semester, it is possible that the Total sales figures by day of the week might be a representative sample of sales. This conclusion, of course, is dependent upon the assumption that the café does not raise prices of its items very often or very much. However, with long-term trends toward credit/electronic payment methods and away from cash, the cash sales totals from this semester may not be representative of cash sales in past and future semesters, nor would the payment methods from the Transactions Data file be accurate in estimating percent of cash versus credit in all café transactions.

When considering number of items purchased and total dollar sales of transactions, it is reasonable to assume that in a business such as the café, the number of items purchased (for example, a sandwich and a drink for lunch) is quite stable and probably has not and will not trend greatly over time. Again, provided that the café does not change prices too much and too often, the total sales per transaction may also be a fairly representative depiction of transaction amounts.

HELPFUL HINT: It may be helpful to also discuss that the café managers would choose to use these data to make decisions moving forward, even though they recognize that patterns may not continue, because it is the only data they have to work with.

5.3.2. Sampling Applications

Because we are afforded population data for an entire semester, it is possible for these data to be used to illustrate sampling concepts and sampling error. While instructors sometimes discuss these concepts in abstract, it is helpful to have concrete data and examples to illustrate these ideas to students with hands-on activities. For these applications, we will focus on the “Transactions Data” file.

To demonstrate sampling error, the instructor could provide students with a random sample of transactions from the file. (Major statistical software packages provide options for
randomly selecting rows for analysis.) Students can first compute sample statistics, for example, mean of Quantity Sold or Total transaction amount, and then can be given the entire population and asked to compare the sample statistics to the actual population parameters (see Table 10). This activity could also be done with the Payment variable, with students computing and comparing the sample and population proportions of transactions paid for with cash or with credit.

Questions for students:
1. (Optional) Refer to the “Transactions Data” file. Randomly generate a sample from the population using software.
2. Given the sample, compute the sample mean (or the sample proportion).
3. Using the entire population, compute the sample mean (or proportion).
4. Compute and discuss the sampling error. How large is it? Will every sample give you this error?

ALTERNATIVE APPLICATION: A natural extension of this activity is to have students generate multiple samples from the population to study sampling variation and sampling distributions. Instructors can assign different students or groups different sample sizes, have each repeat the experiment multiple times, and then compute and display the distribution of the sample statistic.

5.3.3. Confidence Interval Exercises

The “Transactions Data” file is a fairly large dataset with 7288 observations and several variables, including number of items (Qty Sold), Total (dollar amount of the transaction), and Payment (indicating if cash, credit, or debit was the method of payment). Due to its size, it is possible to generate many different random samples, allowing instructors to provide different data to different students or student groups. We find this flexibility useful.

Here, we describe the use of samples from the “Transactions Data” to make inferences about the population of transactions for this particular semester. These activities can be used to illustrate confidence intervals for the population mean and population proportion. In addition, if the instructor desires, he or she can follow up these activities with a presentation of the actual population parameters and allow students to discuss the accuracy of their conclusions.

Questions for students:
1. (Optional) Refer to the “Transactions Data” file. Randomly generate a sample from the population using software.
2. From your sample, compute a 95% confidence interval to estimate the population mean number of items purchased per transaction (Qty Sold).
3. Use your sample to compute a confidence interval to estimate the population mean total dollar amount of transactions (Total).
4. Use your sample to compute a confidence interval to estimate the population proportion of transactions that are paid for by cash (Payment variable).
5. (Optional) If given the population by your instructor, compute the true population means and the true population proportion. Discuss the accuracy of your estimates.

ALTERNATIVE APPLICATION: A follow-up activity that instructors may wish to pursue would be to have each student or group compute a confidence interval based on a different sample but using the same sample size and confidence level, and then determine how many and what percentage of the confidence intervals contained the actual population parameter.

ALTERNATIVE APPLICATION: Instead of confidence intervals, the samples can be used in one sample t-tests for the population mean number of items or mean total dollar amount, or for a z-test for the population proportion of cash (or credit) sales.

5.3.4. Inference on Samples Using ANOVA and Chi-Square Tests

We propose two possible ways to use the “Transactions Data” file for inference exercises.

First, the full dataset can be used to generate samples, perhaps differing by student, on which students may perform tests and then compare their conclusions to actual population results. Second, an instructor may provide only samples (instead of the full population, as in the data mining Section 5.2) to address the managerial questions MQ4 (percent of transactions using cash by day of week), MQ5 (amount purchased with cash vs. credit), and MQ6 (additional register or tablet). These managerial questions and backstory provide opportunities for students to apply ANOVA and Chi-square tests for independence to realistic business situations.

Recall that MQ4 stated “Do the number of transactions made by cash or card differ depending on the day of the week?”

Questions for students:
1. (Optional) Randomly generate a sample from the spring 2015 population of transactions in the “Transactions Data” file using software.
2. How can your sample data be displayed to show the number of transactions made by cash, credit card, debit card, or split payments each day of the week?
3. Can you tell by simply observing the sample data whether there is a significant difference in the population number of transactions made by cash or cards among the days of the week? Why or why not?
4. What statistical analysis can be conducted on the sample data to determine if a statistically significant difference exists in payment types by day of the week in the population of transactions for this semester?
5. (Depending on the sample) Are there any categories that are too small? If so, what should you do with those observations to make the analysis valid or reliable?
6. Perform the analysis on the sample and make a conclusion regarding MQ4, the percentage of transactions that are cash versus credit by day of the week, for all of the spring 2015 semester.
7. (Optional) If your instructor has provided you with the population data, compute the true percentages of cash versus card by day of the week for the spring 2015
semester; then discuss the accuracy of the conclusions you reached using only a sample of the data.

HELPFUL HINT: Again, this is a nice opportunity for students to practice examination of data types to determine how to summarize and compare variables. We prompt students to think about the variables Day of the Week and Payment Type and to deduce that, because both qualitative, a contingency table is an appropriate way to summarize.

HELPFUL HINT: We like to have students articulate why just observing the sample frequencies in the table cannot prove whether a significant difference in the population of transactions exists. Then, we segue into which statistical method allows us to test a sample to see if a difference exists in these means in the population.

In Table 11, we show an example of a contingency table for the day of the week versus cash/card transactions for a random sample of 200 transactions. This and several other samples we observed suggest no evidence of a difference in the percent of cash/card sales for different days of the week in the population of transactions for spring 2015. If students use sample data to address MQ4, their sample will likely lead them to conclude no meaningful differences exist in payment patterns from day to day during the semester studied. However, if population data is explored (see Table 12), some small differences are noticed, with a slightly larger percentage of cash transactions on Mondays and Fridays.

Questions for students:
Recall that managerial question MQ5 asks "Do customers who use cards buy more or less per transaction than those who use cash?"

1. (Optional) Randomly generate a sample from the spring 2015 population of transactions in the “Transactions Data” file using software.
2. Based on only summary information from the sample, can you tell if there are significant differences within the population for the mean number of items sold and the mean sale amount for various payment types?
3. What kind of statistical test is appropriate to apply to the sample data to test for a difference in these means in the population?
4. (Depending on the sample) Should you worry about small sample sizes for Split and Debit Cards? If so, how will you address this issue?
5. Perform statistical analyses on the sample data and make inferences about differences in mean items and mean amount of sales by payment type in the population for spring 2015.
6. (Optional) If your instructor has provided you with the population data, compute the actual mean number of items sold (Qty Sold) and the actual mean dollar Total for transactions during the spring 2015 semester, then discuss the accuracy of the conclusions you reached using only a sample of the data.

Table 11. Example contingency table for a random sample of 200 transactions.

|       | Cash | Card | Total |
|-------|------|------|-------|
| Mon   | 17   | 18   | 35    |
|       | 48.6%| 51.4%| 100%  |
| Tue   | 35   | 24   | 59    |
|       | 59.3%| 40.7%| 100%  |
| Wed   | 16   | 17   | 33    |
|       | 48.5%| 51.5%| 100%  |
| Thu   | 29   | 29   | 58    |
|       | 50.0%| 50.0%| 100%  |
| Fri   | 53.3%| 46.7%| 100%  |
| Total | 105  | 95   | 200   |
|       | 52.5%| 47.5%| 100%  |

NOTE. Pearson Chi-Square = 1.681, DF = 4, p value = 0.794

In Table 12, we show a contingency table for population data.

|       | Cash | Card | Total |
|-------|------|------|-------|
| Mon   | 667  | 560  | 1227  |
|       | 54.4%| 45.6%| 100%  |
| Tue   | 954  | 952  | 1906  |
|       | 50.1%| 49.9%| 100%  |
| Wed   | 747  | 767  | 1514  |
|       | 49.3%| 50.7%| 100%  |
| Thu   | 1012 | 1012 | 2024  |
|       | 50.0%| 50.0%| 100%  |
| Fri   | 343  | 274  | 617   |
|       | 55.6%| 44.4%| 100%  |
| Total | 3723 | 3565 | 7288  |
|       | 51.1%| 48.9%| 100%  |

Table 13. Example ANOVA results for quantity sold and total sale by payment type for a random sample of 200 transactions.

|                   | Groups | n | Average | St Dev |
|-------------------|--------|---|---------|--------|
| Qty Sold by Payment Type | Card   | 101 | 2.12    | 1.04   |
|                   | Cash   | 99  | 1.99    | 1.09   |
| Source of variation | SS     | Df | MS      | F      | p value |
| Between groups    | 0.831  | 1  | 0.8308  | 0.73   | 0.394   |
| Within groups     | 225.564| 198| 1.1392  |        |         |
| Total             | 226.395| 199|         |        |         |

|                   | Groups | n | Average | St Dev |
|-------------------|--------|---|---------|--------|
| Total by Payment Type | Card   | 101 | 3.04    | 1.51   |
|                   | Cash   | 99  | 2.79    | 1.52   |
| Source of variation | SS     | Df | MS      | F      | p value |
| Between groups    | 3.107  | 1  | 3.107   | 1.36   | 0.246   |
| Within groups     | 453.562| 198| 2.291   |        |         |
| Total             | 456.668| 199|         |        |         |
In conclusion, we presented a real dataset that is accessible and hopefully of interest to students, especially those in business disciplines. Motivated by actual managerial questions, the data-set should set the stage for applications of statistical analyses within a real-world context. The student managers of the café are interested in making sense of some of the experiences, trends and data, which were categorized as Understanding Daily Sales and Donations (MQ1–MQ3) and Understanding Payment Methods and Patterns (MQ4–MQ6). Whereas the final decision making lies with the students who run the café, the findings and recommendations that follow from our analyses are summarized here.

The investigation of sales of individual items, specifically pizza and sandwiches, by day of the week (MQ1) showed similar patterns of daily seasonality, with Thursdays and Tuesdays being the busiest days, respectively. We found no trend in sales of sandwiches, but did discover a meaningful downward trend in pizza slices sold, which could possibly be attributable to customers having less desire for hot items as the weather gets warmer over the course of the spring semester. It is important to note that the accuracy measures for forecasting item sales vary widely, which is problematic for perishable items like pizza and sandwiches. Interestingly, when pizza sales for each day of the week are considered separately, we do find a downward trend in all days of the week except for Thursdays, suggesting that even toward the end of the semester, a similar number of pizza slices will be sold on Thursdays.

When exploring sales (MQ2), we found no significant trend for Total Sales, but did see noticeable seasonality in the days of week. The busiest day in overall dollar sales is Thursday, followed by Tuesday and Wednesday. Fridays are the slowest days by far. Further analysis of total sales by day of the week suggests that Thursdays have a significant, positive trend of about $15 per Thursday, while the other days show no significant trends when examined by themselves. The recommendation for staffing the café would be to have the most staff on Tuesdays, Wednesdays, and Thursdays. Only minimal staff would be necessary on Mondays and Fridays. In addition, it may be necessary towards the end of the semester to increase staffing levels on Thursdays as sales tend to increase.

The analysis of customer charitable donations (MQ3) showed that Donations are decreasing slightly but significantly over time. Donations are highest on Thursday, followed by Tuesday and Wednesday, and finally with Fridays being the slowest days. However, accuracy measures indicate a great deal of error in forecasting donations. We found donations are moderately correlated with Cash Sales and estimated that Donations = $1.03 + 0.01 CashSales. The recommendation might be to increase cash sales along with advertising to generate more donations.

When investigating payment methods by day of week (MQ4), our analysis of the population provided insight into patterns of payment type. It was discovered that no debit transactions occurred on Mondays or Tuesdays and that there were higher percentages of cash transactions on Mondays and

### Table 14. Population parameters for quantity sold and total sale amount by payment type.

| Variable | Payment | n  | Mean | St Dev |
|----------|---------|----|------|-------|
| Quantity Sold | Card | 3565 | 2.11 | 1.05 |
| Cash | 3723 | 1.86 | 1.04 |
| Total Sale | Card | 3565 | $2.13 | $1.60 |
| Cash | 3723 | $2.61 | $1.55 |

### Table 15. Example Independent sample t-tests for quantity sold and total sale by payment type for a random sample of 200 transactions.

| Variable | Payment | n  | Mean | St Dev | Difference | t stat (df) | p value |
|----------|---------|----|------|-------|------------|-------------|---------|
| Quantity Sold | Card | 101 | 2.12 | 1.04 | 0.129 | 0.85 (198) | 0.394 |
| Cash | 99 | 1.99 | 1.09 |
| Total Sale | Card | 101 | 3.04 | 1.51 | 0.249 | 1.16 (198) | 0.246 |
| Cash | 99 | 2.79 | 1.52 |

HELPFUL HINT: We find that this is a nice opportunity for students to discuss and reiterate that observing differences in sample means is not sufficient to conclude differences in population means. Once ANOVA is identified as an appropriate technique, we discuss the small samples for the Debit and Split categories, especially compared to the cash and credit card. We recommend combining all of the credit, debit, and split transactions into one category we call “Card.”

In Table 13, we show examples of ANOVAs for the quantity sold and total dollar amount by payment type (cash vs. card) for a random sample of 200 transactions. This and several other samples we observed suggest no evidence of a difference in the means of quantity sold or total sale amount between cash and card purchases for the population of transactions in spring 2015. If students use sample data to address MQ5, their sample means of quantity sold or total sale amount between cash and cards will likely lead them to conclude no meaningful differences exist in purchases paid for with cash compared to credit during the semester in question. However, if the population data are explored (see Table 14), some small differences are noticed, with a slightly larger mean number of items and total dollar amounts for card purchases compared to cash. The magnitudes of these differences, though, are probably not large enough to warrant any major conclusions or actions.

ALTERNATIVE APPLICATION: This is also a nice opportunity to discuss how a two-sample t-test with equal variances is a special case of ANOVA. We present the results of the independent samples t-tests in Table 15.

Finally, we consider MQ6, which asks: “Do you recommend that an additional register (handling both cash and credit/debit cards) be purchased? What about a card-only tablet for processing card transactions?” If students are only provided with sample data and asked to make inferences about the population of spring 2015 transactions, they may use the results from analyses of samples in MQ4 and MQ5 above. Because they are not likely to find evidence of significant differences in card versus cash transactions by day of the week, nor in cash versus credit totals or items purchased, students may not have strong recommendations for one or the other of the options based only on their analysis of samples.

6. Conclusions
Fridays compared to the other days. The analysis of amount purchased using cards versus cash (MQ5) found that there is some difference in the average total sales, with cash transactions consisting of smaller dollar totals.

With regard to the question of adding an additional register or a tablet (MQ6), the analysis of population data showed that since payment methods are fairly evenly split between cash and card transactions, there does not seem to be a clear indication to which direction would be more beneficial at this point in time. The recommendation would be to observe future sales. It would be more beneficial to purchase a register if, in the future, cash transactions were to increase substantially, whereas a tablet might be more beneficial to purchase if credit transactions were to increase. This second scenario may be more likely given retail trends. Further, because customers may possibly spend more money when using cards, it is conceivable that adding a tablet to pay with credit or debit may encourage more card use by students by influencing them to purchase more than they originally intended. Also, because card transactions take additional time for credit verification, adding more capacity to take cards may shorten queues even if the number of transactions is stable.

The real-world managerial questions presented here provide a context to teach statistical methods such as forecasting while emphasizing a data-driven approach to decision making. The story provides background and setting and the data is accessible and easily understood by students. The helpful hints should provide sufficient resources for instructors to use these data in their statistics courses and to spark interesting discussions about statistical methods and applications.

References
DePaolo, C. A., and Robinson, D. F. (2011), “Café Data,” Journal of Statistics Education, 19(1), www.amstat.org/publications/jse/v19n1/depaolo.pdf
Eckerson, W. W. (2002), “Data Quality and the Bottom Line,” TDWI Report, The Data Warehouse Institute.
Fan, W. (2012), “Data Quality: Theory and Practice,” in Web-Age Information Management (pp. 1–16), eds. H. Gao, L. Lim, W. Wang, and L. Chen, Berlin Heidelberg: Springer.
GAISE (2010), “Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report,” Retrieved July 2, 2015, from The American Statistical Association (ASA): http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf
Hasbrouck, R. B., Deniz, B., and Hodges, H. (2014), “Examining Potential Latent Constructs in Teaching a Business Statistics Course: Clustering Responses from Attitudinal Survey Data,” Journal of Business and Economics, 5(5), 708–717. DOI: 10.15341/jbe(2155-7950)/05.05.2014/009.
Hernández, M. A., and Stolfo, S. J (1998), “Real-World Data is Dirty: Data Cleansing and the Merge/Purge Problem,” Data Mining and Knowledge Discovery, 2(1), 9–37.
Kumar, A., and Palvia, P. (2001), “Key Data Management Issues in a Global Executive Information System,” Industrial Management & Data Systems, 101(4), 153–164.
Mvududu, N., and Kanyongo, G. Y. (2011), “Using Real Life Examples to Teach Abstract Statistical Concepts,” Teaching Statistics, 33, 12–16. doi: 10.1111/j.1467-9639.2009.00404.x
Neumann, D., Hood, M., and Neumann, M. (2013), “Using Real-Life Data When Teaching Statistics: Student Perception of this Strategy in an Introductory Statistics Course,” Statistics Education Journal, 12(2), 59–70. http://iase-web.org/documents/SERI/SERI12%282%29_Neumann.pdf
Rahm, E., and Do, H. H. (2000), “Data Cleaning: Problems and Current Approaches,” IEEE Data Eng. Bull., 23(4), 3–13
Tsao, Yea-Ling. (2006), “Teaching Statistics With Constructivist-Based Learning Method To Describe Student Attitudes Toward Statistics,” Journal of College Teaching & Learning, 3(4), 59–64.