Data Article

Sandwich analytics: A dataset comprising one year’s weekly sales data correlated with crime, demographies, and weather

Trent J. Spaulding a, b, *, Edgar E. Hassler a, c, Charles H.L. Edwards a, Joseph A. Cazier a, c

a Center for Analytics Research and Education, Appalachian State University, USA
b Department of Nutrition and Healthcare Management, Appalachian State University, USA
c Department of Computer Information Systems, Appalachian State University, USA

1. Data

The data were produced from a point-of-sale (POS) system from a quick-serve sandwich chain over the duration of one year. Data were produced by 10 stores which varied in location and other characteristics (see Table 2). The primary sales data and all secondary data are genuine. Nevertheless...
names, addresses and other key identifying information have been altered to increase anonymity. The secondary data includes demographic data, unemployment data, crime data, weather data, and statistics on other nearby restaurants. Secondary data was primarily collected at the county level, consequently some stores will share the same secondary data. The data set includes three tables: weekly sales, store attributes, and weather reports. All three tables have been anonymized. The datasets contain data for a one-year period from April 2012 to March 2013.

1.1. Weekly sales

The weekly sales data was collected from the chain’s POS system, and contains raw data for each item sold in each store for each week, from April 2012 through March 2013 (see Table 1). Profit calculations are included in the dataset, but the costs do not include labor and other variable costs; therefore, profit is gross profit not necessarily an indication of total net profit. However, the owner gave estimates that can be used for labor and approximate rent/lease values for each store, which are included in store attributes table. These could be used as a supplement to estimate net profit and illustrate cost differences in profitability.

1.2. Store attributes

The store attributes table contains descriptive attributes about each store, including location information, information about nearby schools and other restaurants, and demographic and crime
information for the county each store is located in. The 10 stores were located in 8 cities across 6 counties (see Fig. 1). The stores include a variety of structures and locations (see Table 2). The chain owner also provided several other data points available in this table. First, stores with more traveler clients are so designated. Second, several stores served higher portions of Hispanic and Native American races. Finally, estimated rent and labor cost is provided for each store.
1.3. County data included in the store attributes table

Crime data were collected from the northern state’s data portal [4]. Employment data were collected from the Bureau of Labor Statistics Local Area Unemployment [7]. Census data were retrieved from the US Census and use the 2012 Census estimate [5] (see Table 3).

1.4. Weather data

The weather data were collected from NOAA’s Climatic Data Center [6]. Eight variables were retained, described in the data description below. Observations exist for each day for five weather stations which are linked back to the sales data by the store attributes table. The weather data consist primarily of wind speed, precipitation, and temperature observations. Descriptive statistics, broken down by weather station, as well as a data dictionary can be found in Table 4. Averages across different weather stations can be found in Table 5.

2. Experimental design, materials, and methods

The sales data were captured directly from the subject company’s POS reporting system and are produced and shared with permission of the original data stewards with obfuscation of the original stores.

The secondary data were collected from a number of sources. Weather reports were pulled from NOAA’s Climatic Data Center for the nearest airport to each store. The store attributes table includes both primary and secondary data: the variables describing the location of the store and proximity to schools were provided by the firm. The variables describing nearby restaurants were manually collected with google maps. Variables describing competition were collected in 2018; therefore, there is a time lag between most of the data and the data related to competitive pressure. We recognize this weakness in the data, but note that generally this competitive pressure does not change quickly.

Crime statistics were pulled from the northern state’s data portal, and aggregated into violent, non-violent, property, society, and other crimes. Violent crimes include murder, manslaughter, forcible sex, assault, and kidnapping/abduction. Property crimes were aggregated in the original data, and include arson, bribery, burglary, counterfeiting/forgery, destruction of property, extortion/blackmail, robbery, and theft. Society crimes were also provided in the original data, and include drug violations, pornography, prostitution, weapon violation, and animal cruelty. The remaining reported crimes were aggregated into other crimes and include non-forcible sex and violation of no contact order.
Table 3  
Employment related data and descriptive statistics.

| Variable         | Description                                                      | Mean (SD)                                      |
|------------------|------------------------------------------------------------------|------------------------------------------------|
| Labor_Force      | Total county labor force in 2012                                 | 60,407.00 (37,201.80)                         |
| Employed         | Number of employed persons in 2012                               | 54,592.50 (33,158.93)                         |
| Unemployed       | Number of unemployed persons in 2012                            | 5,814.50 (4,065.29)                           |
| Unemployment_Rate| Percentage of labor force unemployed in 2012                    | 9.28 (1.02)                                   |
| Crime_Pop        | County population from the crime dataset for 2012               | 105,967.80 (60,598.65)                        |
| Total_Crimes     | Total number of crimes reported in the county in 2012           | 8,675.40 (6,207.53)                           |
| Total_Crime_Rate | Total crimes per 1000 residents in 2012                         | 75.34 (22.89)                                 |
| Violent_Crimes   | Number of violent crimes reported in the county in 2012         | 1,551.00 (1,056.97)                           |
| Violent_Rate     | Violent crimes per 1000 residents in 2012                       | 13.46 (3.71)                                  |
| Property_Crimes  | Property crimes reported in the county in 2012                  | 5,832.80 (4,294.81)                           |
| Property_Rate    | Property crimes per 1000 residents in 2012                      | 50.98 (17.37)                                 |
| Society_Crimes   | Society crimes reported in the county in 2012                   | 944.00 (685.33)                               |
| Society_Rate     | Society crimes per 1000 residents in 2012                       | 8.00 (2.04)                                   |
| Other_Crimes     | Other crimes reported in the county in 2012                      | 347.60 (262.24)                               |
| Other_Rate       | Other crimes per 1000 residents in 2012                         | 2.91 (1.10)                                   |
| Total_Census_Pop | County population from the 2012 Census estimate                  | 123,861.17 (74,067.21)                        |
| Non-Hispanic_White | Number of White residents from the Census estimate in 2012 | 72,322.17 (41,309.94)                         |
| Non-Hispanic_Black | Number of Black residents from the Census estimate in 2012  | 1,373.83 (679.67)                             |
| Non-Hispanic_Native_American | Number of Native American residents from the Census estimate in 2012 | 2,437.17 (3,396.23)                          |
| Non-Hispanic_Asian | Number of Asian residents from the Census estimate in 2012 | 1,942.83 (1,705.55)                           |
| Non-Hispanic_Pacific_Islander | Number of Pacific Islander residents from the Census estimate in 2012 | 144.17 (54.70)                              |
| Non-Hispanic_Two_or_more | Number of residents from two or more ethnicities from the Census estimate in 2012 | 2,405.90 (1,420.96)                         |
| Hispanic_White   | Number of White residents from the Census estimate in 2012      | 39,978.83 (33,211.06)                         |
| Hispanic_Black   | Number of Black residents from the Census estimate in 2012      | 630.00 (527.69)                               |
| Hispanic_Native_American | Number of Native American residents from the Census estimate in 2012 | 1,436.67 (1,758.56)                          |
| Hispanic_Asian   | Number of Asian residents from the Census estimate in 2012      | 289.67 (293.56)                               |
| Hispanic_Pacific_Islander | Number of Pacific Islander residents from the Census estimate in 2012 | 123.50 (113.63)                             |
| Hispanic_Two_or_more | Number of residents from two or more ethnicities from the Census estimate in 2012 | 925.83 (902.81)                             |

Note: N = 6 Counties. Crime variables are missing for one county. County variables are found in the store table with the prefix “county_”.
Table 4
Weather related data and descriptive statistics.

| Variable                          | Description                                                                 | Mean (SD)         |
|-----------------------------------|------------------------------------------------------------------------------|-------------------|
| Station                           | Anonymized weather station name                                              |                   |
| Date                              | Date of observation                                                          |                   |
| Avg_Wind                          | Average wind speed in miles per hour for week                                | 6.17(2.41)        |
| Precip                            | Precipitation in inches for week                                             | 0.03(0.04)        |
| Snow                              | Snowfall in inches for week                                                  | 0.02(0.10)        |
| Snow_Depth                        | Snow depth in inches for week                                                | 0.13(0.74)        |
| Max_Temp                          | Maximum observed temperature in degrees Fahrenheit for week                   | 64.85(18.81)      |
| Min_Temp                          | Minimum observed temperature in degrees Fahrenheit for week                   | 40.40(11.43)      |
| Week                              | Week of the year                                                             |                   |
| Days_Precipitated                 | Days in the week where precipitation fell                                    | 1.62(1.65)        |
| Days_Snowed                       | Days in the week which snow fell                                             | 0.06(0.37)        |
| Days_With_Snow_Accumulation       | Days in the week with snow accumulation                                      | 0.08(0.60)        |
| Days_With_Strong_Wind             | Days in the week with wind speed above 25mph (25mph is what the NWS considers to be a strong breeze) | 0.01(0.11)        |
| Cold_Days                         | Days in the week where the temperature was below 45°F Fahrenheit             | 1.04(2.02)        |
| Bad_Weather_Days                  | Days in the week in which it precipitated, snowed, was windy, or below 45°F | 2.34(2.26)        |
| Bad_Weather_Week                  | Whether the week had 5 or more bad weather days (Yes/No)                     |                   |

Table 5
Mean weather statistics by weather station.

| Weather Station         | Avg_Wind | Precip | Snow_Depth | Max_Temp | Min_Temp |
|-------------------------|----------|--------|------------|----------|----------|
| Appleton Airport        | 5.02     | 0.02   | 0.27       | 65.66    | 38.46    |
| Lake City Airport       | 6.37     | 0.02   | 0.00       | 63.57    | 40.09    |
| North Town Airport      | 6.48     | 0.03   | 0.00       | 65.99    | 39.84    |
| Rail City Airport       | 6.05     | 0.02   | 0.00       | 67.00    | 40.04    |
| River City Airport      | 7.38     | 0.05   | 0.00       | 64.73    | 44.58    |

Demographic statistics were retrieved from the US census bureau for the 2012 census estimate on the county level in both states. Unemployment statistics were collected from the Bureau of Labor Statistics’ Local Area Unemployment estimates for 2012.

The sales data were anonymized by renaming menu items and renumbering the stores. Locations were anonymized by changing the names of cities, counties, and states. None of the other data was changed so that any conclusions drawn from the data can still be valid. Keys to de-anonymize the data are held by the authors.

Acknowledgments

We thank Hank Holzwarth for initial work on data cleaning and obfuscation and Brendan Sullivan for data cleaning efforts.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.L. Newman, E. Howlett, S. Burton, Implications of fast food restaurant concentration for preschool-aged childhood obesity, J. Business Res. 67 (8) (2014) 1573–1580.
[2] R. Thomadsen, Product positioning and competition: the role of location in the fast food industry, Mark. Sci. 26 (6) (2007) 792–804.
[3] Q. Shen, P. Xiao, McDonald’s and KFC in China, Competitors or companions, Mark. Sci. 33 (2) (2014) 287–307.
[4] Local State Statistical Analysis Center. Not Listed to Protect Anonymity of the Data, However This Information Can Be Made Available to Researchers by Contacting the Authors.

[5] US Census Demographic Data by County, 2018. https://www.census.gov. (Accessed December 2018).

[6] NOAA Climatic Data Center Climate Data Online, 2018. https://www.ncdc.noaa.gov/cdo-web/datasets. (Accessed April 2019).

[7] Bureau of Labor Statistics Local Area Unemployment, 2018. https://www.bls.gov/lau/. (Accessed January 2019).

[8] T.J. Spaulding, E.E. Hassler, C. Edwards, J.A. Cazier, Sandwich Analytics: A Dataset Comprising One Year’s Weekly Sales Data Correlated with Crime, Demographics, and Weather, Mendeley Data, v1, 2019. https://doi.org/10.17632/6htjnfs78b.1.