ISSU0053 Data Science and Big Data Analytics

# DATASET INFORMATION

# cafedata.csv

This dataset contains data from a café, called Executive Express, run by undergraduate business students at a NY public university. It was collected over a ten-week period from January to April 2010.

Variables include a number of types of items sold and wasted, along with total sales values.

**SOURCES:** The sales and item data were collected by the café’s student volunteer workers each day from January 19 through April 1, 2010. The daily temperature data was obtained from an internet-based historical weather site called Weather Underground, http://www.wunderground.com/history/.

**VARIABLE DESCRIPTIONS:**

A header line containing variable names is included in the data file. Variables are comma delimited.

Each day is one line in the data file. Missing observations are coded as na.

The variables are as follows:

* Location;
* Day Code: 1 = Mon, 2 = Tues, 3 = Wed, 4 = Thurs, 5 = Fri;
* Sandwiches Sold: Number of bread sandwiches sold per day;
* Sandwiches Waste: Number of bread sandwiches expired and thrown away each day;
* Wraps Sold: Number of wrap sandwiches sold per day;
* Wraps Waste: Number of wrap sandwiches expired and thrown away each day;
* Muffins Sold: Number of assorted muffins sold each day;
* Muffins Waste: Number of assorted muffins expired and thrown away each day;
* Cookies Sold: Number of assorted cookies sold each day;
* Cookies Waste: Number of assorted cookies expired and thrown away each day;
* Fruit Cup Sold: Number of fruit cups sold each day;
* Fruit Cup Waste: Number of fruit cups expired and thrown away each day;
* Chips: Number of bags of chips sold each day;
* Juices: Number of assorted bottled juices sold each day;
* Sodas: Number of cups of soda sold each day;
* Coffees: Number of cups of coffee sold each day;
* Sales USD: Total dollar sales indicated on sales register receipt each day;
* Max Daily Temperature (F): Maximum daily temperature recorded for the city each day;

SPECIAL NOTES: One day has missing data for all sale items because the café was closed.

ISSU0053 Data Science and Big Data Analytics

# DATASET INFORMATION

# db\_all

Located in file: SectionD.Rda

The datasets consist of a study and follow up on a series of patients.

Variables including the number of pregnancies the patient has had, their BMI, insulin level, age, and so on, were collected in a patient study.

The diabetes columns records whether the patient went on to develop diabetes in the following five years.

**Variables in the dataset:**

Pregnancies Number of times pregnant

Glucose Plasma glucose conc. over 2 hrs in an oral glucose tolerance test

BloodPressure Diastolic blood pressure (mm Hg)

Triceps Triceps skin fold thickness (mm)

Insulin 2-Hour serum insulin (mu U/ml)

Mass Body mass index (weight in kg/(height in m)^2)

Pedigree Stores result of calculation related to diabetes among relations. Age Age (years)

Diabetes Positive / Negative.

Did the subject go on to develop diabetes in the following 5 years.