Explanation of Dataset

You will see the following worksheets in the dataset.

Name of the Worksheet	Description
Abbeville, LA	Contains the incoming monthly aggregate medical examination volume at the Abbeville, LA HC for cardiovascular exams from January 2006 to December 2013
May-2007 Violet, LA	Includes all the medical exams that were rerouted to the Violet, LA HC during May 2007
May-2007 New Orleans, LA	Includes all the medical exams that were rerouted to the New Orleans, LA HC during May 2007
May-2007 Lafayette, LA	Includes all the medical exams that were rerouted to the Lafayette, LA HC during May 2007
May-2007 Baton Rouge, LA	Includes all the medical exams that were rerouted to the Baton Rouge, LA HC during May 2007
December 2013 Data	Contains the exams that were rerouted from Abbeville HC to nearby HCs during December 2013. The worksheet is at the examination level
Heart Related Condition Codes	Provides the six-digit heart condition codes with December 2013 data
Condition Code Map	All six-digit condition codes

Abbeville, LA Worksheet: The "Incoming Examinations" data field in the worksheet contains the aggregate monthly cardiovascular exam request volume that were received and completed (or rerouted) by Abbeville HC.

May-2007 Violet, LA, May-2007 New Orleans, LA, May-2007 Lafayette, LA, and May-2007 Baton Rouge, LA Worksheets: During the second of May 2007, the Abbeville HC was closed for renovation, so exam requests were routed to four nearby HCs (Violet, New Orleans, Lafayette, and Baton Rouge). The number of those rerouted exams is not reflected in the 107 exams that appear for May, 2007 in the "Abbeville, LA" worksheet.

Note the following regarding the above four worksheets:

- The field "Original Hospital Location" denotes the HC from where the exam was rerouted from. The only location of our interest for the group project is Abbeville HC.
- The field "Examination" contains the type of the examination which was rerouted. For the purpose of this group project, we focus on only heart related exams.
- The "Date" field denotes the date on which the original exam request was sent from the originating LO to the original HC.
- The field "Request ID" is a unique identifier for the examination requests.

In October 2008, the HC received an unusually large number of requests since neighboring New Orleans HC was closed due to hurricane.

From December 2009 to February 2010, the number of incoming examinations was not recorded. According to the Fargo Health administration, there were 5129 requests for cardiovascular exams during that time.

The data for May, June, and July of 2013 are incomplete in the "Abbeville, LA" worksheet since during those months there were rerouted requests from the Abbeville HC and those numbers are not reflected in the numbers in the "Abbeville, LA" worksheet. Those rerouted requests were mixed with other rerouted requests in the "May 2007 New Orleans, LA", "May-2007 Lafayette, LA", and "May-2007 Baton Rouge, LA" worksheets and can be identified by filtering on the "Date" field.

During December 2013, all the exam requests were rerouted from Abbeville HC to other HCs in the vicinity. The data relevant for that month is provided separately in the "December 2013 Data" worksheet. Each row in this worksheet corresponds to a 17-digit SYSID that had been automatically assigned to an examination which was rerouted in December 2013. The following table explains the two pieces of routing SYSIDs:

Type of Exam	**Format**
Rerouted from Abbeville	SYSID starts with "L839" and ends either in "TGU3" or "ROV8"
Heart related condition	Have one of the condition codes listed in the "Heart Related Condition Codes" worksheet

The "Heart Related Condition Codes" worksheet provides the mapping between six-digit condition codes and corresponding heart related medical conditions.

The "Condition Code Map" worksheet provides general mapping between six-digit condition codes and corresponding medical conditions.

Hints

- The forecasting models you need to create are for Abbeville HC location and for heart related (cardiovascular) exams only.
- The dataset has intentional data quality issues (outliers, missing values, inconsistent data, and duplicates). You will need to clean the dataset prior to building the model.
- You will need to justify your data cleaning strategies in the project report.
- You may use techniques such as exponential smoothing, double exponential smoothing, and ARIMA for models. You will need to evaluate three models based on their accuracy utilizing their metrics such as MAPE, MAD, and MSE (review week 8 for guidance).
- You can clean the dataset in Excel. You can test and train the models in Excel and R.