Microsoft

Applied Machine Learning

Lab 1 – Working with Time Series Data

Overview

In this lab, you will use R to create a forecasting model for time series data. Specifically, you will write R code to predict dairy production levels for the next twelve months based on historical data.

What You'll Need

To complete this lab, you will need the following:

- An Azure ML account
- The files for this lab

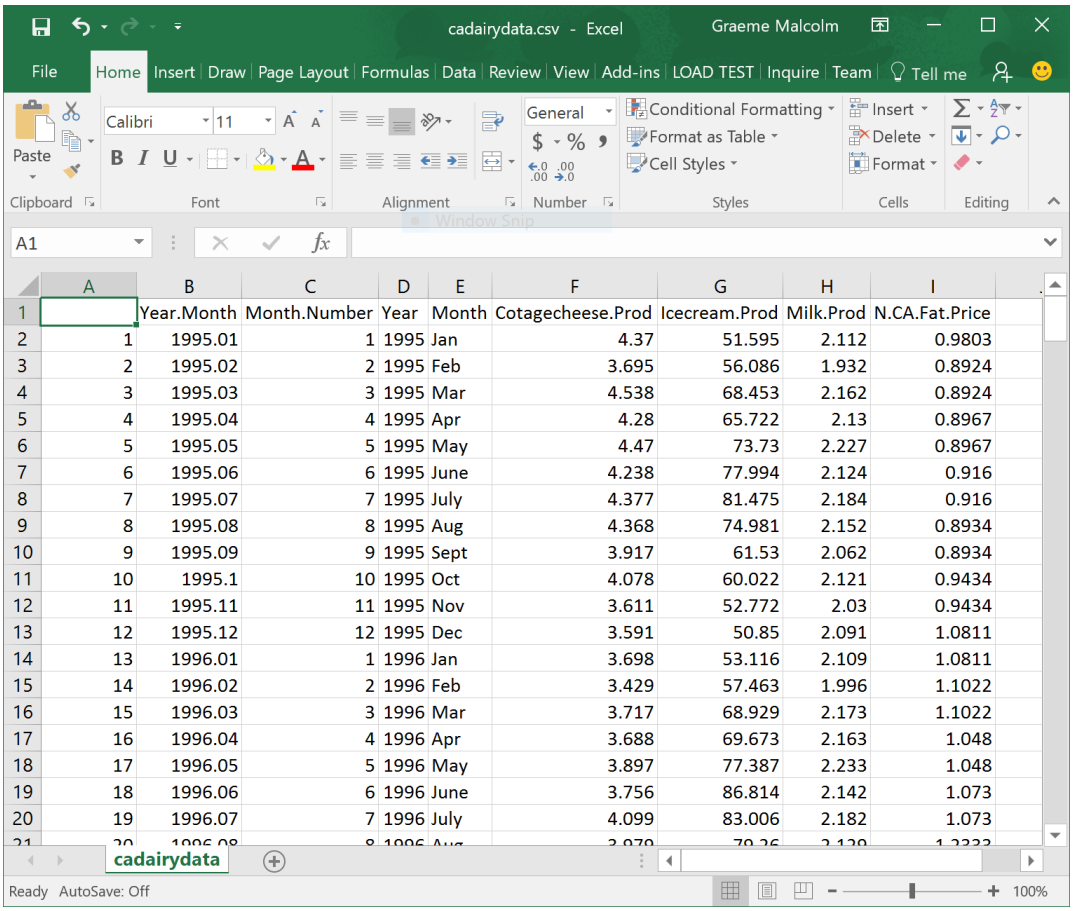
Note: To set up the required environment for the lab, follow the instructions in the Setup Guide for this course.

Exploring and Uploading Historical Data

In this lab, you will create a forecasting model for dairy production. The forecasting model is based on an existing dataset of dairy production history for California.

Explore the Dataset

1. In the folder where you extracted the lab files for this module (for example, C:\DAT203.3x\Lab01), open the **cadairydata.csv** file, using either a spreadsheet application such as Microsoft Excel, or a text editor such as Microsoft Windows Notepad.
2. View the contents of the **cadairydata.csv** file, noting that it contains dairy production data from January 1995 to December 2013, as shown here:



	A	B	C	D	E	F	G	H	I
1		Year.Month	Month.Number	Year	Month	Cotagecheese.Prod	Icecream.Prod	Milk.Prod	N.CA.Fat.Price
2	1	1995.01	1	1995	Jan	4.37	51.595	2.112	0.9803
3	2	1995.02	2	1995	Feb	3.695	56.086	1.932	0.8924
4	3	1995.03	3	1995	Mar	4.538	68.453	2.162	0.8924
5	4	1995.04	4	1995	Apr	4.28	65.722	2.13	0.8967
6	5	1995.05	5	1995	May	4.47	73.73	2.227	0.8967
7	6	1995.06	6	1995	June	4.238	77.994	2.124	0.916
8	7	1995.07	7	1995	July	4.377	81.475	2.184	0.916
9	8	1995.08	8	1995	Aug	4.368	74.981	2.152	0.8934
10	9	1995.09	9	1995	Sept	3.917	61.53	2.062	0.8934
11	10	1995.1	10	1995	Oct	4.078	60.022	2.121	0.9434
12	11	1995.11	11	1995	Nov	3.611	52.772	2.03	0.9434
13	12	1995.12	12	1995	Dec	3.591	50.85	2.091	1.0811
14	13	1996.01	1	1996	Jan	3.698	53.116	2.109	1.0811
15	14	1996.02	2	1996	Feb	3.429	57.463	1.996	1.1022
16	15	1996.03	3	1996	Mar	3.717	68.929	2.173	1.1022
17	16	1996.04	4	1996	Apr	3.688	69.673	2.163	1.048
18	17	1996.05	5	1996	May	3.897	77.387	2.233	1.048
19	18	1996.06	6	1996	June	3.756	86.814	2.142	1.073
20	19	1996.07	7	1996	July	4.099	83.006	2.182	1.073

3. Close the data file without saving any changes.

Upload the Dataset to Azure Machine Learning

1. Browse to https://studio.azureml.net and sign in using the Microsoft account associated with your free Azure ML account.
2. If the **Welcome** page is displayed, close it by clicking the **OK** icon (which looks like a checkmark). Then, if the **New** page (containing a collection of Microsoft samples) is displayed, close it by clicking the **Close** icon (which looks like an X).
4. At the bottom left, click **NEW**; and in the **NEW** dialog box, in the **DATASET** tab, click **FROM LOCAL FILE**. Then in the **Upload a new dataset** dialog box, browse to select the **cadairydata.csv** file from the folder where you extracted the lab files on your local computer. Enter the following details, and then click the ✓icon.
    - **This is a new version of an existing dataset**: Unselected
    - **Enter a name for the new dataset**: cadairydata.csv
    - **Select a type for the new dataset**: Generic CSV file with a header (.csv)
    - **Provide an optional description**: Historical dairy data.
5. Wait for the upload of the dataset to complete, then click **OK** on the status bar at the bottom of the Azure ML Studio page.

Working with Time-Series Data in Jupyter

Now you are ready to use R code in a Jupyter notebook to work with the time-series data and create a forecasting model for dairy production.

Upload a Jupyter Notebook

1. In Azure ML Studio, click **NEW**; and in the **NEW** dialog box, in the **NOTEBOOK** tab, click **Upload**. Then in the **Upload a new notebook** dialog box, browse to select the **TimeSeries.ipynb** file from the folder where you extracted the lab files on your local computer. Enter the following details, and then click the ✓icon.
    - **Enter a name for the new notebook**: TimeSeries
    - **Select a language for the new notebook**: R
2. Wait for the upload of the notebook to complete, then click **OK** on the status bar at the bottom of the Azure ML Studio page.

Use R to Work with the Time Series Data

1. In Azure ML Studio, on the Notebooks tab, open the **TimeSeries** notebook you uploaded in the previous procedure.
2. Follow the instructions in the notebook to work with the time series data.
3. When you have completed all of the coding tasks in the notebook, save your changes and then close and halt the notebook.

Summary

In this lab, you have used R in a Jupyter notebook to work with time-series data.