

## Weather Underground Import

Import Weather Underground data into IBM SPSS Modeler



Product: IBM® SPSS® Modeler

Extension type: Utility

**Table of Contents**

Description.....	3
Requirements.....	3
Installation.....	3
R Packages used.....	3
User Interface.....	4
Example.....	5-8
User Input.....	5
Results.....	8
Important links.....	8
Learn.....	8
Discuss.....	8

**Description:**

This extension imports weather data for a vector of valid 3 or 4 digit Airport codes or a valid Weather Station ID (examples: "BUF", "ORD", "VABB" for Mumbai). The extension will return a record of weather information for a given date at each location.

**Requirements:**

- SPSS Modeler v16.0 or later
- R: <http://www.r-project.org/>
- R Essentials for SPSS Modeler plugin: <https://developer.ibm.com/predictiveanalytics/downloads/>

**Installation:**

Close SPSS Modeler. Save the .cfe file in the CDB folder of the IBM SPSS Modeler installation directory for Windows and Linux. The copy should reside in that same folder and not in a sub-folder.

For example, for Windows 7 the default location is "C:\ProgramData\IBM\SPSS\Modeler\16\CDB". If the ProgramData folder is hidden type the path manually.

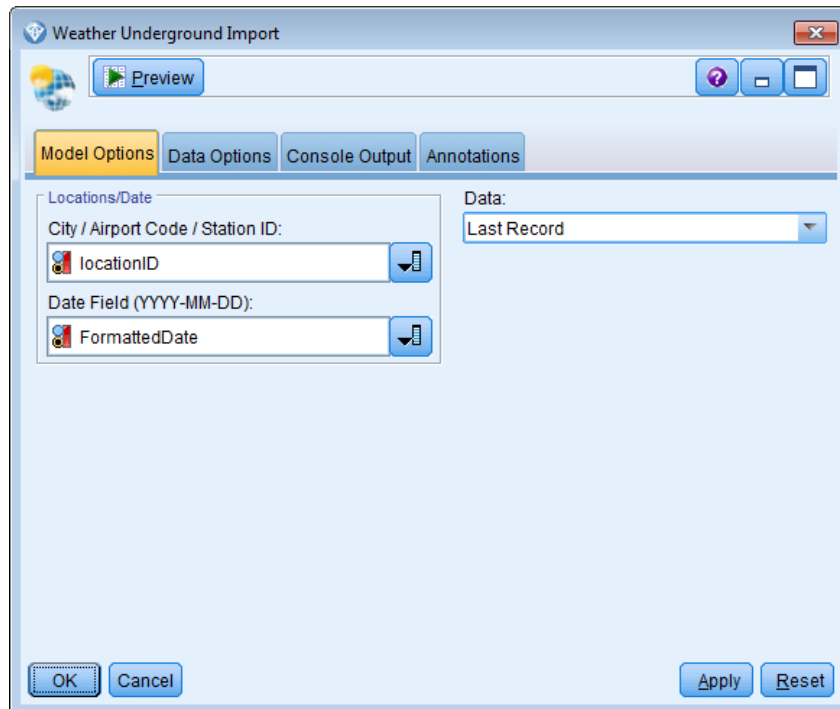
Restart SPSS Modeler: the node will now appear in the Field Ops palette.

**R Packages used:**

- plyr package created by Hadley Wickham Lang <https://cran.r-project.org/web/packages/plyr/>
- weatherData package created by Ram Narasimhan <https://cran.r-project.org/web/packages/weatherData/>

## User Interface

- Double click on the node to get to the options. There are the following fields:
  - City/Airport Code/Station ID: click the drop down to select the column containing the location
  - Data: You have the option to return the last record on file for the date or select the record that is the minimum or maximum for each of the weather features. For example you can return the records for a date that contain the minimum temperature for the date at each location.
  - Date Field (YYYY-MM-DD): Select the field that contains a date in the format YYYY-MM-DD

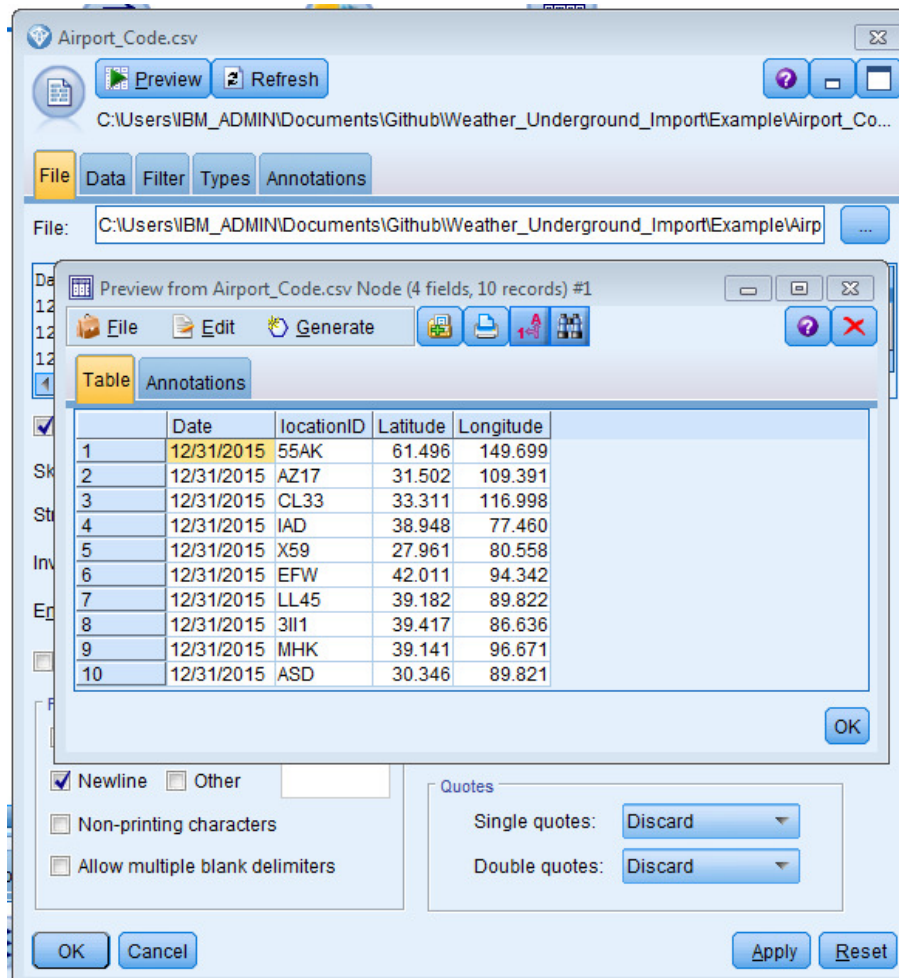


## Example

This example will demonstrate how to import weather data for a column of valid locations in a CSV file. The CSV used in this example can be found in the example folder of this extension's GitHub Repository.

### User Input

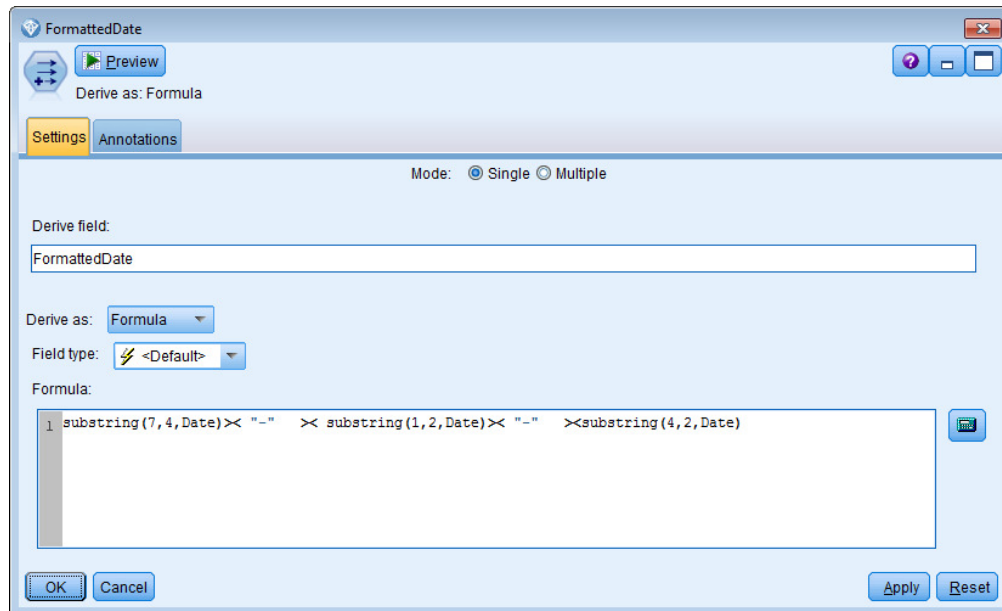
1. From the sources palette, add a Var File node to the stream and read Airport\_Code.csv. The locationID column of this dataset has the valid location identifier required for this extension.



- Before using the extension we need to convert the date from the format mm/dd/yyyy to yyyy-mm-dd. We can do this with a Derive node, found in the Field Ops palette. We will derive a new field based on the following formula:

`substring(7,4,Date)><"-" >< substring(1,2,Date)><"-" ><substring(4,2,Date)`

This means for each record in the column **Date** we will join 4 characters starting at position 7 ('2015') with a dash '-' then 2 characters at position 1 ('12'), followed by a dash '-' and finally 2 characters at position 4 ('31') to create a date of '2015-12-31'.



FormattedDate

Derive as: Formula

Settings Annotations

Mode: ☒ Single ☐ Multiple

Derive field:

FormattedDate

Derive as: Formula

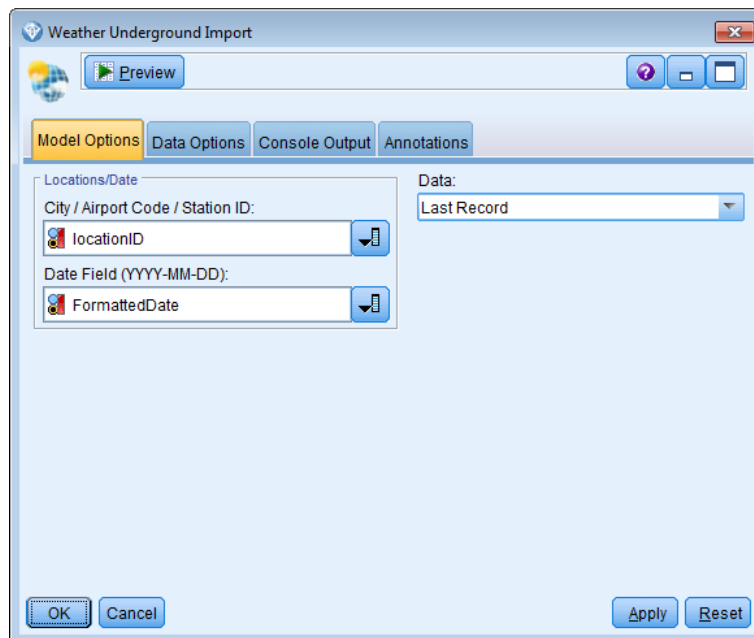
Field type: <Default>

Formula:

```
1 substring(7,4,Date)><"-" >< substring(1,2,Date)><"-" ><substring(4,2,Date)
```

OK Cancel Apply Reset

3. Add the Weather Underground Import extension from the Field Ops palette and connect it to the Var File node. Select locationID for the City/Airport Code/Station ID field, and select FormattedDate for the date field. Finally, choose the type of data to import in the Data field or leave as the default Last Record.



4. From the Output palette, connect a Table to the extension
5. Click on the Table and run the stream

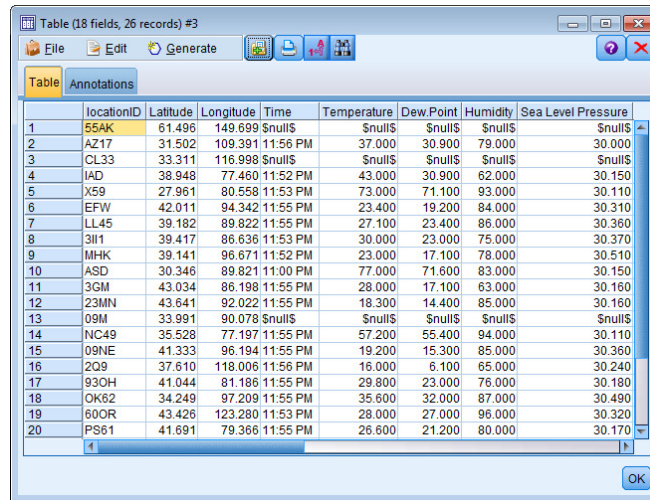


## Results

This extension imports the following fields into IBM SPSS Modeler.

- Time
- Temperature
- Dew.Point
- Humidity
- Sea Level Pressure
- Visibility
- Wind Direction
- Wind Speed
- Gust Speed
- Precipitation
- Events
- Conditions
- Wind Dir Degrees
- DateUTC
- DateTime

If you used the sample data provided, your results should match the table below:



The screenshot shows a table window titled "Table (18 fields, 26 records) #3". The table contains 20 rows of weather data. The first row is highlighted in yellow. The columns are: LocationID, Latitude, Longitude, Time, Temperature, Dew.Point, Humidity, and Sea Level Pressure. The data is as follows:

	LocationID	Latitude	Longitude	Time	Temperature	Dew.Point	Humidity	Sea Level Pressure
1	55AK	61.496	149.699	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$
2	AZ17	31.502	109.391	11:56 PM	37.000	30.900	79.000	30.000
3	CL33	33.311	116.998	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$
4	IAD	38.948	77.460	11:52 PM	43.000	30.900	62.000	30.150
5	X59	27.961	80.558	11:53 PM	73.000	71.100	93.000	30.110
6	EFW	42.011	94.342	11:55 PM	23.400	19.200	84.000	30.310
7	LL45	39.182	89.822	11:55 PM	27.100	23.400	86.000	30.360
8	3I1	39.417	86.636	11:53 PM	30.000	23.000	75.000	30.370
9	MHK	39.141	96.671	11:52 PM	23.000	17.100	78.000	30.510
10	ASD	30.346	89.821	11:00 PM	77.000	71.600	83.000	30.150
11	3GM	43.034	86.198	11:55 PM	28.000	17.100	63.000	30.160
12	23MN	43.641	92.022	11:55 PM	18.300	14.400	85.000	30.160
13	09M	33.991	90.078	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$
14	NC49	35.528	77.197	11:55 PM	57.200	55.400	94.000	30.110
15	09NE	41.333	96.194	11:55 PM	19.200	15.300	85.000	30.360
16	2Q9	37.610	118.006	11:56 PM	16.000	6.100	65.000	30.240
17	93OH	41.044	81.186	11:55 PM	29.800	23.000	76.000	30.180
18	OK62	34.249	97.209	11:55 PM	35.600	32.000	87.000	30.490
19	60OR	43.426	123.280	11:53 PM	28.000	27.000	96.000	30.320
20	PS61	41.691	79.366	11:55 PM	26.600	21.200	80.000	30.170

This weather data can be used to build models or do analysis involving data from different sources.

## Important Links

### Learn

- Learn more about [SPSS software](#).
- Visit [developerWorks Business analytics](#) for more technical analytics resources for developers.
- Create your Bluemix account here: <https://console.ng.bluemix.net/solutions/watson>
- The [Comprehensive R Archive Network](#) is the main site for the R project and each R package. The help pages and manuals that are associated with optimx, nlmrt, and Rcgmin are detailed. Numerous references are provided.
- Read "[Do I need to learn R?](#)" (Catherine Dalzell, developerWorks, September 2013) to learn why R is a valuable tool for data analytics that was expressly designed to reflect the way that statisticians think and work.
- "[Calling R from SPSS](#)" describes how to use R code inside IBM SPSS Modeler 16.
- Read "[Create new nodes for IBM SPSS Modeler 16 using R](#)" to learn how to create new extensions easily.

### Discuss

- Visit the [IBM SPSS Community](#) to share tips and experiences with other IBM SPSS developers.
- Follow [developerWorks on Twitter](#) to be among the first to hear about new resources.