CSC 458 Data Mining and Predictive Analytics I, Spring 2021, Dr. Parson's answers

I will post my output ARFF files in ~parson/DataMine/csc458trees2sp2021.solution.zip.

That is the handout example data for Fall 2022 tutorial on Weka classification. ^^^^^

Dr. Dale E. Parson, Assignment 2, Using Weka rules and trees to correlate several stream-sampling attributes with dissolved oxygen in milligrams per liter. Due by 11:59 PM on Saturday March 13 via D2L. There will be a 10% per-day late penalty after that, and I cannot award points after I go over my solution during the following class period.

Perform the following steps to set up for this semester's projects and to get my handout. Start out in your login directory on csit (a.k.a. acad).

cd $HOME
mkdir DataMine # This should already be there from assignment 1.
cd ./DataMine
cp ~parson/DataMine/csc458trees2sp2021.problem.zip csc458trees2sp2021.problem.zip
unzip csc458trees2sp2021.problem.zip
cd ./csc458trees2sp2021

TO GET THE FILES FROM ACAD TO YOUR Mac OR Linux machine use this command line:

scp YOURLOGIN@acad.kutztown.edu://home/kutztown.edu/parson/DataMine/csc458trees2sp2021.problem.zip csc458trees2sp2021.problem.zip

You can use the reverse command line to copy files from your machine into your acad account.

WINDOWS USERS should log into https://download.kutztown.edu/ , log in, go to Computer Science, and download WinSCP. It appears you can just go to http://winscp.net/eng/download.php
To get it.

EDIT THE SUPPLIED FILE README.txt when the following questions starting at Q1 below. Keep with the supplied format, and do not turn in a Word or PDF or other file format. I will deduct 20% for other file formats, because with this many varying assignments being turned in, I need a way to grade these in reasonable time, which for me is a batch edit run on the **vim** editor. Please turn in your final files README.txt and USGS_PA_STREAM_2012_NOMINAL.arff.gz by the deadline using the D2L Assignment page 2.

Running Weka (Fall 2022 students have already done that for Assignment 2 on regression.)

Open USGS_PA_STREAM_2012.arff.gz via Weka's Preprocess tab to investigate the following attributes. The time-related attributes derive from datetime.

agency_cd	USGS (US Geological Survey) These data are automated water samples from Pennsylvania streams In 2012 obtained as a text file from https://waterdata.usgs.gov/nwis, Water Quality, Historical Observations, then run through my Python data extraction script.
site_no	The USGS site number for the sampling station.

site_name	The USGS name for the site.
datetime	When the sample was taken.
tz_cd	Time zone.
OxygenMgPerLiter	Dissolved oxygen in milligrams per liter will be our target attribute.
pH	Base / acidity pH scale. Low numbers are acidic, with 7 being neutral.
TempCelsius	Water temperature in centigrade.
Conductance	Electrical conductance in microsiemens per centimeter at 25ºC.
GageHt	Stream gage height in feet.
DischargeRate	Flow discharge rate in cubic feet per second.
TimeOfYear	Nominal value for one of the seasons, derived from datetime.
TimeOfDay	Nominal value for one time of day of sample, derived from datetime.
OxygenClass	Nominal range for OxygenMgPerLiter, with the ranges coming from A PA Dept. of Environmental web site. Below is my Python function That shows the mapping from numeric levels to nominal values.
month	Month as a number 2 through 12; there were no January measures.
MinuteOfDay	Number of minutes since the preceding midnight for this sample.
MinuteFromMidnite	Number of minutes to the closest midnight for this sample.
MinuteOfYear	Sampling time in minutes since the previous start of the year.
MinuteFromNewYear	Sampling time in minutes to the closest start of the year.

Here is the Python 3.x function for deriving OxygenClass from OxygenMgPerLiter.

```
def oxygen2Class(paramMap):
    if (not 'OxygenMgPerLiter' in paramMap.keys()):
        return None
    level = paramMap['OxygenMgPerLiter']
    if level == None:
        return None
    level = float(level)
    result = 'Unsafe'
    if level >= 6.0:
        result = 'NorthernPikeExcellent'
    elif level >= 5.5:
        result = 'BlackBassGood'
    elif level >= 4.2:
        result = 'CommonSunfishMedium'
    elif level >= 3.3:
        result = 'BlackBullheadLow'
    return str(result)
```

PART I – Preparing the data. 22% for the correct saved ARFF file.

1. Open ARFF file **USGS_PA_STREAM_2012.arff.gz** in Weka and observe that the attribute names and types in your dataset match those on the previous page; bring the **Edit** Preprocessor Window up, full screen, and scroll around inspecting for missing values that are grayed out in this Editor. You can click on a heading such as datetime to sort the instances on that attribute. Shift-click on a heading gives a descending sort. Close the Edit window when you are done.
2. Run Weka's **unsupervised -> attribute -> StringToNominal** filter to turn strings into sets of values that you can read in Preprocess. (NOTE: I recommend applying StringToNominal, Discretize, or other

to-nominal conversions only when you are ready to start analysis of a dataset, not at the time of its capture or pre-analysis storage. The reason is that adding more instances later may change the **universe** of a nominal set. However, it is necessary to split a dataset into a training set and a testing set **after** conversion of attributes to nominals, so that both datasets have the same underlying nominal sets of values.) After selecting this filter, click its command line display and set the **attributeRange** to **first-last**, then click the **Apply** button in the upper right of the Preprocess window. The **attributeRange** also accepts numeric attribute ranges separated by a "-", and individual attribute numbers separated by a ",", using the attribute numbers in the lower left of the Preprocess window. Some filters require you to be more precise with the **attributeRange** ranges, but StringToNominal converts only strings to sets of symbols, leaving non-string attributes unchanged. I usually save this step until I am ready to analyze a dataset, since adding new instances later on may add strings not in the current nominal set.

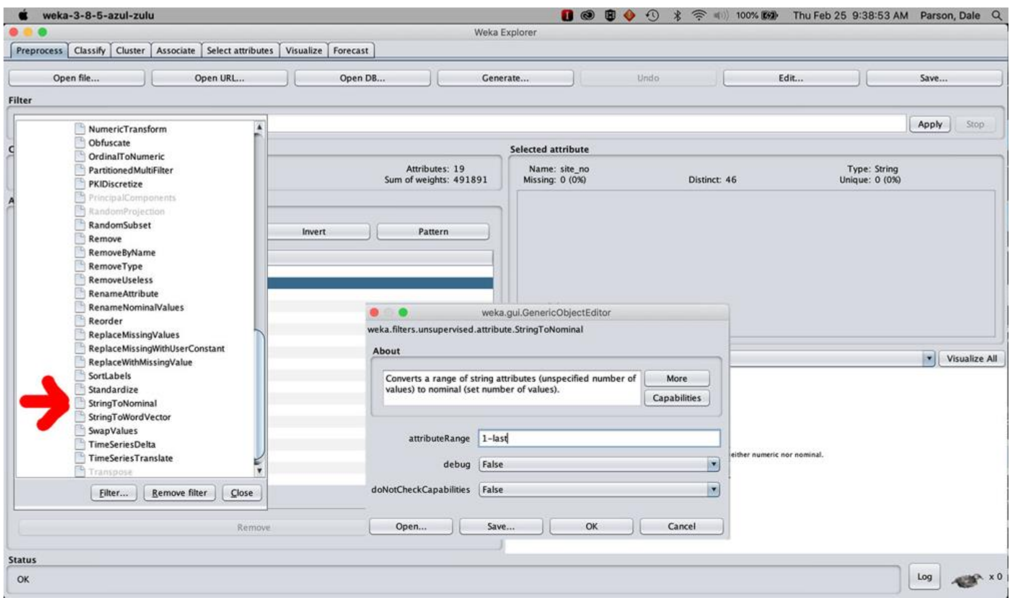


Figure 1: StringToNominal filter

3. Run Weka's **unsupervised** -> **attribute** -> **RemoveUseless** filter.

   Q1 in README.txt: <u>Which attributes did RemoveUseless remove, and why</u>? Read the pop-up RemoveUseless documentation in Weka, and execute Undo in the Weka preprocessor if you need to inspect the pre-RemoveUseless attribute values. Make sure to re-run RemoveUseless if you execute Undo.

   agency_cd is removed because it is single-valued and therefore does not help to distinguish among target attribute values. As noted for to-nominal filters, **I recommend applying RemoveUseless only when you are ready to start analysis of a dataset**, not at the time of its capture or pre-analysis storage. The reason is that adding instances later may add different values for previously single-valued attributes.

4. Click the Visualize tab in Preprocess and pop up a plot of OxygenMgPerLiter on the Y axis as a function of MinuteOfYear on the X axis. It will look like Figure 2 at the top of the next page. The four highest peaks all come from the same site (site_no & site_name), and the two lowest troughs at the lower right come from the same site. These sites provide outlying data that we want to eliminate from our training set and test set of data.

   Q2 in README.txt: What are the **site_no** and **site_name** values for these two outlying sites?

01540500 Susquehanna River at Danville (4 peaks)
01467087 Frankford Creek at Castor Ave, Philadelphia, PA (2 troughs)

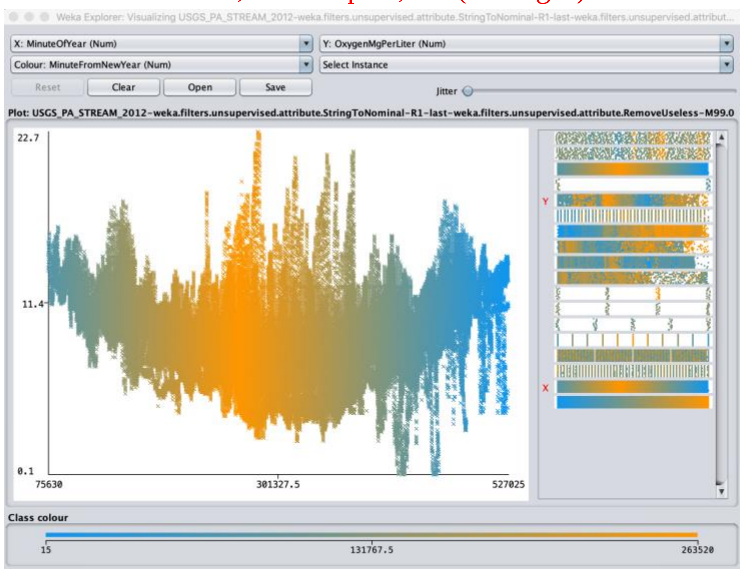
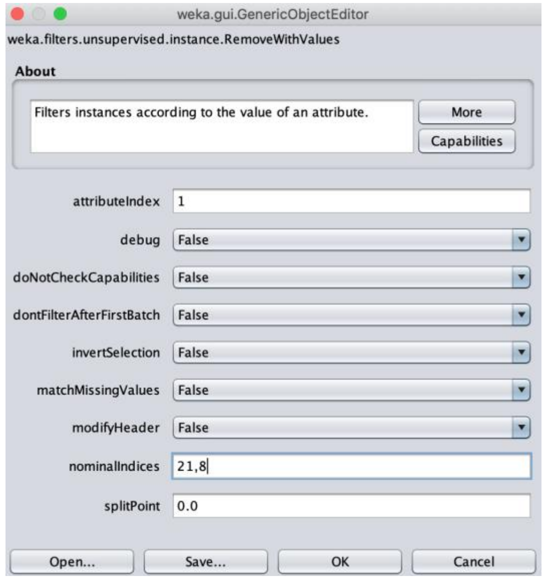


Figure 2: OxygenMgPerLiter on the Y axis as a function of MinuteOfYear on the X axis

5. In the Preprocess tab use the **unsupervised -> instance -> RemoveWithValues** Filter to remove all instances coming from these two sites. We do not want to train our models on these outliers, and we will skip testing them for now. In the RemoveWithValues parameter pop-up you can use the **attributeIndex** of either site_no or site_name (its number to te left of the attribute name) and you can use the two **nominalIndices** values for those sites by scrolling through the **Selected attribute** scrolling region. After clicking **Apply** for **RemoveWithValues** to remove those sites, verify that they are gone using Visualize as in Figure 2 and verifying that those peaks and troughs are gone. There will be new peaks and troughs that do not repeat for the same site.



6. In the Preprocess tab delete attributes so that only these remain:

site_no	The USGS site number for the sampling station.
OxygenMgPerLiter	Dissolved oxygen in milligrams per liter will be our target attribute.
pH	Base / acidity pH scale. Low numbers are acidic, with 7 being neutral.
TempCelsius	Water temperature in centigrade.
Conductance	Electrical conductance in microsiemens per centimeter at 25°C.
DischargeRate	Flow discharge rate in cubic feet per second.
MinuteOfDay	Number of minutes since the preceding midnight for this sample.
MinuteOfYear	Sampling time in minutes since the previous start of the year.

We are deleting datetime, tz_cd, and the more coarse-grain temporal attributes because they are redundant with the remaining temporal attributes, which are numeric and therefore more useful to Weka. We delete OxygenClass because it is redundant with OxygenMgPerLiter. We are deleting GageHt because it has many missing values for some sites. We will delete site information because it is trivial for some machine learning algorithms to simply memorize site+time -to- OxygenMgPerLiter mappings without analyzing underlying patterns in physical attribute correlations. First we must partition the instances into training data and testing data using a Python script that I am supplying.

7. Reorder attributes using the **Unsupervised -> Attribute -> Reorder** filter so that OxygenMgPerLiter moves to the bottom while all others stay the same. Weka expects the target attribute (a.k.a. *class*) being predicted to be the last one; Weka forces you to explicitly set the target attribute if you do not perform this reordering.

site_no	The USGS site number for the sampling station.
pH	Base / acidity pH scale. Low numbers are acidic, with 7 being neutral.
TempCelsius	Water temperature in centigrade.
Conductance	Electrical conductance in microsiemens per centimeter at 25°C.
DischargeRate	Flow discharge rate in cubic feet per second.
MinuteOfDay	Number of minutes since the preceding midnight for this sample.

MinuteOfYear — Sampling time in minutes since the previous start of the year.
OxygenMgPerLiter — Dissolved oxygen in milligrams per liter will be our target attribute.

8. Partition OxygenMgPerLiter into 10 bins using the **Unsupervised -> Attribute -> Discretize** filter. You will have to set config parameter **ignoreClass** to **True**, since this attribute is the class (i.e., the target attribute). Set **useEqualFrequency** to **False** to main the roughly normal, bell-shaped distribution of OxygenMgPerLiter in the 10 bins. **Apply** the filter.

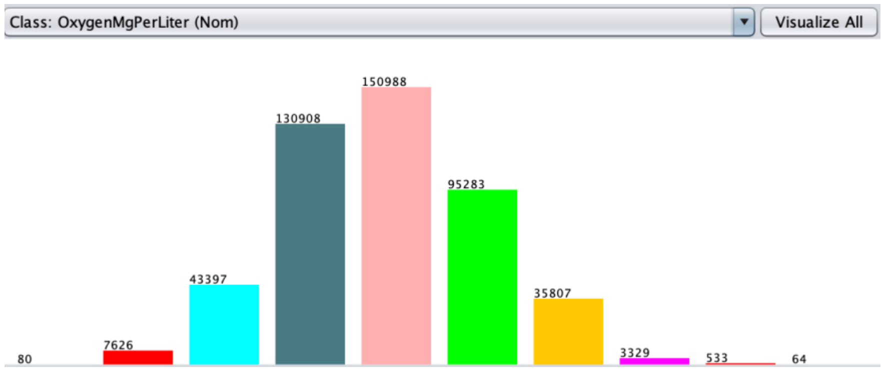
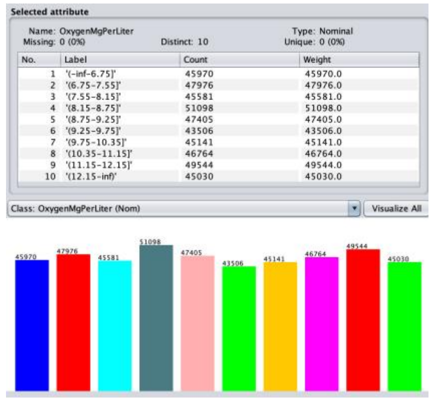


Figure 3: Discretized OxygenMgPerLiter with useEqualFrequency = False

SAVE THIS DATA SET AS file USGS_PA_STREAM_2012_NOMINAL.arff.gz. It must contain all of these attributes & instances. You will perform other temporary attribute deletions, but you will not save those reduced datasets. You must turn in this ARFF file via D2L along with your completed **README.txt** file when you are ready.

NOTE: After considering possible student edit errors based on emails & prior semester experience, I am going to duplicate the analysis steps, highlighted in underlined red, using a variety of discretizations. The additional reason for doing this is to demonstrate those discretizations, defined as follows. However, incorrect edits do cost some points, since they usually skew results throughout the subsequent steps. I will not charge points twice for a single edit error as long as I can sort it out. Please edit carefully.

ALL_NOMINAL – accidentally Discretized all numeric attributes in Unsupervised -> Attribute -> Discretize Step 8 above.
ALLSUPERVISED – intentionally Discretized all numeric attributes in **Supervised** -> Attribute -> Discretize Step 8 above. Supervised Discretization attempts to correlate non-target attribute bins to the target bin values, first applying the non-equal-frequency unsupervised Discretization on OxygenMgPerLiter (denoted as "DO" below for brevity).
DO_USEEQUALFREQ – only DO was Discretized, but **useEqualFrequency** was set to True, making this DO bin distribution for the full dataset:



Next run my supplied script in the directory containing this ARFF file like this:

python splitTrainTest.py

where python is python3.x. If you copy the above ARFF file back onto acad or mcgonagall, you can run this:

/usr/local/bin/python3.7 splitTrainTest.py

That will partition **USGS_PA_STREAM_2012_NOMINAL.arff.gz** into two files:

USGS_PA_STREAM_2012_TRAIN.arff.gz contains training data from 4 training sites determined to have a good cross-section of representative data for the entire dataset. There are 49066 instances.

USGS_PA_STREAM_2012_TEST.arff.gz contains additional testing data from the remaining sites. There are 418949 instances.

PART II – Analyzing the data. 6.5% for each of Q1 through Q12.

Exit & restart Weka and **load USGS_PA_STREAM_2012_TRAIN.arff.gz into Weka. Remove the site_no attribute for all learning.** We do not want Weka to memorize site+time -to- OxygenMgPerLiter mappings as previously discussed.

LINKED READINGS. Read over this essay. It was the best match for the Google query "dissolved oxygen in water".
http://www.fondriest.com/environmental-measurements/parameters/water-quality/dissolved-oxygen/
Here is another, in case that site is down when you go to it. Try to at least skim them both.
https://water.usgs.gov/edu/dissolvedoxygen.html

Also, review the Kappa statistic so you can interpret its significance. For any analysis question in Q3-Q13 for which you paste the Kappa statistic into README.txt, make sure to interpret the Kappa statistic in your analysis.

"Landis and Koch considers 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect. Fleiss considers kappas > 0.75 as excellent, 0.40-0.75 as fair to good, and < 0.40 as poor. It is important to note that both scales are somewhat arbitrary." Read the rest of my Kappa page yourself. We will go over Kappa statistic in class.

EDIT THE SUPPLIED FILE README.txt and answer the following questions. Please read these instructions closely. I will deduct points for missing requirements.

Q3: Run the rule **OneR** classifier. What are the "Correctly Classified Instances" as a percentage correct, and the "Kappa statistic", for class attribute OxygenMgPerLiter?

```
Correctly Classified Instances       27915               56.8928 %
Kappa statistic                          0.4219
Landis and Koch moderate
```

ALL_NOMINAL:

```
        Correctly Classified Instances       27927               56.9172 %
        Kappa statistic                          0.4129
        Kappa degraded insignificantly, but easier rule interpretation in Q4.
```

ALLSUPERVISED:

```
        Correctly Classified Instances       29547               60.2189 %
        Kappa statistic                          0.4723
```

DO_USEEQUALFREQ:

```
        Correctly Classified Instances       17926               36.5345 %
        Kappa statistic                          0.2893
```

Q4: Copy & paste OneR's rule into README.txt. (Use mouse sweep and control-C on Windows, command-C on Mac)What attribute does OneR use to predict OxygenMgPerLiter, and what pattern can you see, if any in the mappings of this rule? If there are a lot of IF rules, just look at a contiguous block of them at a time. There is a pattern.

As MinuteOfYear goes from early in year to mid-year, OxygenMgPerLiter decreases bin value. After mid-year it ramps back up. (NOTE: I inserted line numbers via vim so I can highlight the brief exponential spike in DO during warm weather indicated by the spikes in Figure 2 both before and after deleting the 2 sites.)

```
   1    MinuteOfYear:
   2            < 98446.0       -> '(12.24-14.03]'
   3            < 99016.0       -> '(10.45-12.24]'
   4            < 99766.0       -> '(12.24-14.03]'
   5            < 100396.0      -> '(10.45-12.24]'
   6            < 101536.0      -> '(12.24-14.03]'
   7            < 101776.0      -> '(10.45-12.24]'
   8            < 104476.0      -> '(12.24-14.03]'
   9            < 104746.0      -> '(10.45-12.24]'
  10            < 105810.5      -> '(12.24-14.03]'
  11            < 106455.0      -> '(10.45-12.24]'
  12            < 107295.0      -> '(12.24-14.03]'
  13            < 111285.0      -> '(10.45-12.24]'
  14            < 111555.0      -> '(12.24-14.03]'
  15            < 120465.0      -> '(10.45-12.24]'
```

16	< 123075.0	-> '(8.66-10.45]'
17	< 123255.0	-> '(10.45-12.24]'
18	< 123855.0	-> '(8.66-10.45]'
19	< 127125.0	-> '(10.45-12.24]'
20	< 127245.0	-> '(8.66-10.45]'
21	< 130935.0	-> '(10.45-12.24]'
22	< 131325.0	-> '(8.66-10.45]'
23	< 134175.0	-> '(10.45-12.24]'
24	< 135945.0	-> '(8.66-10.45]'
25	< 136245.0	-> '(10.45-12.24]'
26	< 137415.0	-> '(8.66-10.45]'
27	< 137835.0	-> '(10.45-12.24]'
28	< 138765.0	-> '(8.66-10.45]'
29	< 139245.0	-> '(10.45-12.24]'
30	< 140445.0	-> '(8.66-10.45]'
31	< 140685.0	-> '(10.45-12.24]'
32	< 141735.0	-> '(8.66-10.45]'
33	< 142155.0	-> '(10.45-12.24]'
34	< 143265.0	-> '(8.66-10.45]'
35	< 143625.0	-> '(10.45-12.24]'
36	< 144645.0	-> '(8.66-10.45]'
37	< 145035.0	-> '(10.45-12.24]'
38	< 146085.0	-> '(8.66-10.45]'
39	< 146445.0	-> '(10.45-12.24]'
40	< 147555.0	-> '(8.66-10.45]'
41	< 147705.0	-> '(10.45-12.24]'
42	< 148005.0	-> '(8.66-10.45]'
43	< 148185.0	-> '(6.87-8.66]'
44	< 148875.0	-> '(8.66-10.45]'
45	< 149385.0	-> '(10.45-12.24]'
46	< 150435.0	-> '(8.66-10.45]'
47	< 150855.0	-> '(10.45-12.24]'
48	< 151935.0	-> '(8.66-10.45]'
49	< 152385.0	-> '(10.45-12.24]'
50	< 153645.0	-> '(8.66-10.45]'
51	< 153795.0	-> '(10.45-12.24]'
52	< 154305.0	-> '(8.66-10.45]'
53	< 154635.0	-> '(6.87-8.66]'
54	< 155775.0	-> '(8.66-10.45]'
55	< 156075.0	-> '(6.87-8.66]'
56	< 156645.0	-> '(8.66-10.45]'
57	< 156825.0	-> '(6.87-8.66]'
58	< 158685.0	-> '(8.66-10.45]'
59	< 158925.0	-> '(6.87-8.66]'
60	< 159975.0	-> '(8.66-10.45]'
61	< 160485.0	-> '(6.87-8.66]'
62	< 160785.0	-> '(8.66-10.45]'
63	< 160875.0	-> '(6.87-8.66]'
64	< 161175.0	-> '(8.66-10.45]'
65	< 161895.0	-> '(6.87-8.66]'
66	< 162255.0	-> '(8.66-10.45]'

67	< 162405.0	-> '(6.87-8.66]'
68	< 162825.0	-> '(8.66-10.45]'
69	< 163065.0	-> '(6.87-8.66]'
70	< 163605.0	-> '(8.66-10.45]'
71	< 163995.0	-> '(10.45-12.24]'
72	< 164745.0	-> '(8.66-10.45]'
73	< 164835.0	-> '(10.45-12.24]'
74	< 165015.0	-> '(8.66-10.45]'
75	< 165435.0	-> '(10.45-12.24]'
76	< 166185.0	-> '(8.66-10.45]'
77	< 167205.0	-> '(10.45-12.24]'
78	< 167685.0	-> '(8.66-10.45]'
79	< 168495.0	-> '(10.45-12.24]'
80	< 169035.0	-> '(8.66-10.45]'
81	< 170265.0	-> '(10.45-12.24]'
82	< 170355.0	-> '(8.66-10.45]'
83	< 171555.0	-> '(10.45-12.24]'
84	< 171855.0	-> '(8.66-10.45]'
85	< 172575.0	-> '(10.45-12.24]'
86	< 173295.0	-> '(8.66-10.45]'
87	< 174075.0	-> '(10.45-12.24]'
88	< 174795.0	-> '(8.66-10.45]'
89	**< 175335.0**	**-> '(10.45-12.24]' (NOTE: LAST HI DO 121 days into year 4/30/2012)**
90	< 180825.0	-> '(8.66-10.45]'
91	< 180945.0	-> '(6.87-8.66]'
92	< 181515.0	-> '(8.66-10.45]'
93	< 183675.0	-> '(6.87-8.66]'
94	< 183975.0	-> '(8.66-10.45]'
95	< 184365.0	-> '(6.87-8.66]'
96	< 185025.0	-> '(8.66-10.45]'
97	< 186165.0	-> '(6.87-8.66]'
98	< 186405.0	-> '(8.66-10.45]'
99	< 186555.0	-> '(6.87-8.66]'
100	< 186645.0	-> '(8.66-10.45]'
101	< 187605.0	-> '(6.87-8.66]'
102	< 187695.0	-> '(8.66-10.45]'
103	< 188025.0	-> '(6.87-8.66]'
104	< 193755.0	-> '(8.66-10.45]'
105	< 195705.0	-> '(6.87-8.66]'
106	< 200055.0	-> '(8.66-10.45]'
107	< 200775.0	-> '(6.87-8.66]'
108	< 201405.0	-> '(8.66-10.45]'
109	< 202575.0	-> '(6.87-8.66]'
110	< 202875.0	-> '(8.66-10.45]'
111	< 206595.0	-> '(6.87-8.66]'
112	< 206685.0	-> '(8.66-10.45]'
113	< 221685.0	-> '(6.87-8.66]'
114	< 221865.0	-> '(8.66-10.45]'
115	< 222195.0	-> '(6.87-8.66]'
116	< 222375.0	-> '(8.66-10.45]'
117	< 222615.0	-> '(6.87-8.66]'

118	< 222705.0	-> '(8.66-10.45]'
119	< 223275.0	-> '(6.87-8.66]'
120	< 223485.0	-> '(8.66-10.45]'
121	< 223575.0	-> '(6.87-8.66]'
122	< 223875.0	-> '(8.66-10.45]'
123	< 223965.0	-> '(6.87-8.66]'
124	< 224175.0	-> '(5.08-6.87]'
125	< 224625.0	-> '(6.87-8.66]'
126	< 224805.0	-> '(5.08-6.87]'
127	< 231975.0	-> '(8.66-10.45]'
128	< 232065.0	-> '(6.87-8.66]'
129	< 232215.0	-> '(8.66-10.45]'
130	< 232605.0	-> '(6.87-8.66]'
131	< 233415.0	-> '(8.66-10.45]'
132	< 233505.0	-> '(6.87-8.66]'
133	< 233625.0	-> '(8.66-10.45]'
134	< 234345.0	-> '(6.87-8.66]'
135	< 234465.0	-> '(8.66-10.45]'
136	< 238755.0	-> '(6.87-8.66]'
137	< 238995.0	-> '(8.66-10.45]'
138	< 239895.0	-> '(6.87-8.66]'
139	< 239985.0	-> '(8.66-10.45]'
140	< 245925.0	-> '(6.87-8.66]'
141	< 246225.0	-> '(8.66-10.45]'
142	< 246915.0	-> '(6.87-8.66]'
143	< 247035.0	-> '(8.66-10.45]'
144	< 247155.0	-> '(6.87-8.66]'
145	< 247395.0	-> '(8.66-10.45]'
146	< 247485.0	-> '(6.87-8.66]'
147	< 247665.0	-> '(8.66-10.45]'
148	< 248025.0	-> '(6.87-8.66]'
149	< 248235.0	-> '(8.66-10.45]'
150	< 248415.0	-> '(6.87-8.66]'
151	< 248475.0	-> '(8.66-10.45]'
152	< 248685.0	-> '(6.87-8.66]'
153	< 248835.0	-> '(8.66-10.45]'
154	< 252975.0	-> '(6.87-8.66]'
155	< 253065.0	-> '(8.66-10.45]'
156	< 255945.0	-> '(6.87-8.66]'
157	< 256035.0	-> '(8.66-10.45]'
158	< 256725.0	-> '(6.87-8.66]'
159	< 256815.0	-> '(5.08-6.87]'
160	< 257175.0	-> '(6.87-8.66]'
161	< 257265.0	-> '(8.66-10.45]'
162	< 258645.0	-> '(6.87-8.66]'
163	< 258885.0	-> '(8.66-10.45]'
164	< 259635.0	-> '(6.87-8.66]'
165	< 259815.0	-> '(5.08-6.87]'
166	< 260025.0	-> '(6.87-8.66]'
167	< 260205.0	-> '(8.66-10.45]'
168	< 260925.0	-> '(6.87-8.66]'

169	< 261285.0	-> '(5.08-6.87]'
170	< 261465.0	-> '(6.87-8.66]'
171	< 261885.0	-> '(8.66-10.45]'
172	< 262395.0	-> '(6.87-8.66]'
173	< 262515.0	-> '(5.08-6.87]'
174	< 262845.0	-> '(6.87-8.66]'
175	< 263325.0	-> '(8.66-10.45]'
176	< 263805.0	-> '(6.87-8.66]'
177	< 263925.0	-> '(5.08-6.87]'
178	< 264285.0	-> '(6.87-8.66]'
179	< 264405.0	-> '(8.66-10.45]'
180	**< 264885.0**	**-> '(10.45-12.24]' (NOTE Average DO Spike 184 days 7/2/2012)**
181	< 265065.0	-> '(8.66-10.45]'
182	< 265305.0	-> '(6.87-8.66]'
183	< 265575.0	-> '(5.08-6.87]'
184	< 265755.0	-> '(6.87-8.66]'
185	< 266145.0	-> '(3.29-5.08]'
186	< 266745.0	-> '(6.87-8.66]'
187	< 267045.0	-> '(3.29-5.08]'
188	< 267195.0	-> '(6.87-8.66]'
189	< 267705.0	-> '(3.29-5.08]'
190	< 268155.0	-> '(6.87-8.66]'
191	< 268515.0	-> '(3.29-5.08]'
192	< 268635.0	-> '(6.87-8.66]'
193	< 269325.0	-> '(3.29-5.08]'
194	< 269565.0	-> '(6.87-8.66]'
195	< 269745.0	-> '(5.08-6.87]'
196	< 269955.0	-> '(3.29-5.08]'
197	< 270255.0	-> '(6.87-8.66]'
198	< 270795.0	-> '(3.29-5.08]'
199	< 270975.0	-> '(6.87-8.66]'
200	< 271155.0	-> '(5.08-6.87]'
201	< 271425.0	-> '(3.29-5.08]'
202	< 271545.0	-> '(6.87-8.66]'
203	< 271845.0	-> '(3.29-5.08]'
204	< 272265.0	-> '(6.87-8.66]'
205	< 272475.0	-> '(3.29-5.08]'
206	< 272655.0	-> '(5.08-6.87]'
207	< 272895.0	-> '(3.29-5.08]'
208	< 273045.0	-> '(6.87-8.66]'
209	< 273255.0	-> '(3.29-5.08]'
210	< 273435.0	-> '(6.87-8.66]'
211	< 273705.0	-> '(8.66-10.45]'
212	< 274425.0	-> '(6.87-8.66]'
213	< 274545.0	-> '(8.66-10.45]'
214	< 278745.0	-> '(6.87-8.66]'
215	< 278865.0	-> '(8.66-10.45]'
216	< 279075.0	-> '(3.29-5.08]'
217	< 279195.0	-> '(6.87-8.66]'
218	< 279345.0	-> '(8.66-10.45]'
219	< 279945.0	-> '(6.87-8.66]'

220	< 280155.0	-> '(3.29-5.08]'
221	< 280305.0	-> '(6.87-8.66]'
222	< 280545.0	-> '(3.29-5.08]'
223	< 280935.0	-> '(6.87-8.66]'
224	< 281145.0	-> '(3.29-5.08]'
225	< 281685.0	-> '(6.87-8.66]'
226	< 281895.0	-> '(8.66-10.45]'
227	< 282375.0	-> '(6.87-8.66]'
228	< 282855.0	-> '(5.08-6.87]'
229	< 283125.0	-> '(6.87-8.66]'
230	< 283275.0	-> '(8.66-10.45]'
231	< 283725.0	-> '(6.87-8.66]'
232	< 284205.0	-> '(5.08-6.87]'
233	< 285345.0	-> '(6.87-8.66]'
234	< 285825.0	-> '(5.08-6.87]'
235	< 286605.0	-> '(6.87-8.66]'
236	< 286845.0	-> '(5.08-6.87]'
237	< 287085.0	-> '(3.29-5.08]'
238	< 287205.0	-> '(5.08-6.87]'
239	< 287835.0	-> '(6.87-8.66]'
240	< 288165.0	-> '(5.08-6.87]'
241	< 288525.0	-> '(3.29-5.08]'
242	< 288705.0	-> '(5.08-6.87]'
243	< 289395.0	-> '(6.87-8.66]'
244	< 289875.0	-> '(5.08-6.87]'
245	< 289995.0	-> '(3.29-5.08]'
246	< 291315.0	-> '(5.08-6.87]'
247	< 291435.0	-> '(6.87-8.66]'
248	< 291705.0	-> '(5.08-6.87]'
249	< 292575.0	-> '(6.87-8.66]'
250	< 292845.0	-> '(5.08-6.87]'
251	< 295155.0	-> '(6.87-8.66]'
252	< 295335.0	-> '(5.08-6.87]'
253	< 296595.0	-> '(6.87-8.66]'
254	< 296925.0	-> '(5.08-6.87]'
255	< 298065.0	-> '(6.87-8.66]'
256	< 298425.0	-> '(5.08-6.87]'
257	< 298635.0	-> '(6.87-8.66]'
258	< 298845.0	-> '(5.08-6.87]'
259	< 299325.0	-> '(6.87-8.66]'
260	< 299955.0	-> '(5.08-6.87]'
261	< 301095.0	-> '(6.87-8.66]'
262	< 301425.0	-> '(5.08-6.87]'
263	< 302805.0	-> '(6.87-8.66]'
264	< 303015.0	-> '(5.08-6.87]'
265	< 303915.0	-> '(6.87-8.66]'
266	< 304575.0	-> '(5.08-6.87]'
267	< 305355.0	-> '(6.87-8.66]'
268	< 306075.0	-> '(5.08-6.87]'
269	< 306885.0	-> '(6.87-8.66]'
270	< 307395.0	-> '(5.08-6.87]'

271	< 308385.0	-> '(6.87-8.66]'
272	< 308715.0	-> '(5.08-6.87]'
273	< 311175.0	-> '(6.87-8.66]'
274	< 311265.0	-> '(5.08-6.87]'
275	< 311415.0	-> '(6.87-8.66]'
276	< 311625.0	-> '(5.08-6.87]'
277	< 315645.0	-> '(6.87-8.66]'
278	< 316005.0	-> '(5.08-6.87]'
279	< 317145.0	-> '(6.87-8.66]'
280	< 317295.0	-> '(5.08-6.87]'
281	< 322995.0	-> '(6.87-8.66]'
282	< 323115.0	-> '(5.08-6.87]'
283	< 324405.0	-> '(6.87-8.66]'
284	< 324525.0	-> '(5.08-6.87]'
285	< 325875.0	-> '(6.87-8.66]'
286	< 326205.0	-> '(5.08-6.87]'
287	< 336375.0	-> '(6.87-8.66]'
288	< 336585.0	-> '(8.66-10.45]'
289	< 337815.0	-> '(6.87-8.66]'
290	< 338085.0	-> '(8.66-10.45]'
291	< 339285.0	-> '(6.87-8.66]'
292	< 339525.0	-> '(8.66-10.45]'
293	< 340725.0	-> '(6.87-8.66]'
294	< 340995.0	-> '(8.66-10.45]'
295	< 347385.0	-> '(6.87-8.66]'
296	< 347655.0	-> '(5.08-6.87]'
297	< 347985.0	-> '(6.87-8.66]'
298	< 348075.0	-> '(8.66-10.45]'
299	< 349335.0	-> '(6.87-8.66]'
300	< 349575.0	-> '(8.66-10.45]'
301	< 350745.0	-> '(6.87-8.66]'
302	< 351015.0	-> '(8.66-10.45]'
303	< 352305.0	-> '(6.87-8.66]'
304	< 352425.0	-> '(8.66-10.45]'
305	< 354645.0	-> '(6.87-8.66]'
306	< 355065.0	-> '(5.08-6.87]'
307	< 360165.0	-> '(6.87-8.66]'
308	< 360345.0	-> '(5.08-6.87]'
309	< 361125.0	-> '(6.87-8.66]'
310	< 361395.0	-> '(3.29-5.08]'
311	< 361635.0	-> '(6.87-8.66]'
312	< 362205.0	-> '(3.29-5.08]'
313	< 363135.0	-> '(6.87-8.66]'
314	< 363375.0	-> '(3.29-5.08]'
315	< 363705.0	-> '(6.87-8.66]'
316	< 364245.0	-> '(3.29-5.08]'
317	< 365085.0	-> '(6.87-8.66]'
318	< 365415.0	-> '(8.66-10.45]'
319	< 366525.0	-> '(6.87-8.66]'
320	< 366885.0	-> '(8.66-10.45]'
321	< 367965.0	-> '(6.87-8.66]'

322	< 368355.0	-> '(8.66-10.45]'
323	< 369345.0	-> '(6.87-8.66]'
324	< 369705.0	-> '(8.66-10.45]'
325	< 370875.0	-> '(6.87-8.66]'
326	< 371145.0	-> '(8.66-10.45]'
327	< 372165.0	-> '(6.87-8.66]'
328	< 372645.0	-> '(8.66-10.45]'
329	< 373665.0	-> '(6.87-8.66]'
330	< 374115.0	-> '(8.66-10.45]'
331	< 375195.0	-> '(6.87-8.66]'
332	< 375405.0	-> '(8.66-10.45]'
333	< 378015.0	-> '(6.87-8.66]'
334	< 383055.0	-> '(8.66-10.45]'
335	< 383175.0	-> '(6.87-8.66]'
336	< 383295.0	-> '(8.66-10.45]'
337	< 383415.0	-> '(6.87-8.66]'
338	< 383655.0	-> '(8.66-10.45]'
339	< 383775.0	-> '(6.87-8.66]'
340	< 383865.0	-> '(8.66-10.45]'
341	< 384165.0	-> '(6.87-8.66]'
342	< 384645.0	-> '(8.66-10.45]'
343	< 385005.0	-> '(6.87-8.66]'
344	< 385455.0	-> '(8.66-10.45]'
345	< 385695.0	-> '(6.87-8.66]'
346	< 386085.0	-> '(8.66-10.45]'
347	< 386445.0	-> '(6.87-8.66]'
348	< 387585.0	-> '(8.66-10.45]'
349	< 388035.0	-> '(6.87-8.66]'
350	< 389085.0	-> '(8.66-10.45]'
351	< 389505.0	-> '(6.87-8.66]'
352	< 390435.0	-> '(8.66-10.45]'
353	< 390975.0	-> '(6.87-8.66]'
354	< 391875.0	-> '(8.66-10.45]'
355	< 392355.0	-> '(6.87-8.66]'
356	< 393495.0	-> '(8.66-10.45]'
357	< 393735.0	-> '(6.87-8.66]'
358	< 394905.0	-> '(8.66-10.45]'
359	< 395145.0	-> '(6.87-8.66]'
360	< 396465.0	-> '(8.66-10.45]'
361	< 396765.0	-> '(6.87-8.66]'
362	< 397725.0	-> '(8.66-10.45]'
363	< 397965.0	-> '(6.87-8.66]'
364	< 400905.0	-> '(8.66-10.45]'
365	< 401085.0	-> '(5.08-6.87]'
366	< 401265.0	-> '(8.66-10.45]'
367	< 401535.0	-> '(5.08-6.87]'
368	< 402075.0	-> '(8.66-10.45]'
369	< 402525.0	-> '(5.08-6.87]'
370	< 405645.0	-> '(8.66-10.45]'
371	< 406125.0	-> '(5.08-6.87]'
372	< 406815.0	-> '(8.66-10.45]'

373	< 407085.0	-> '(5.08-6.87]'
374	< 407265.0	-> '(8.66-10.45]'
375	< 407715.0	-> '(5.08-6.87]'
376	< 408345.0	-> '(8.66-10.45]'
377	**< 408435.0**	**-> '(10.45-12.24]' (NOTE: 284 days 10/10/2012 start higher DO)**
378	< 409755.0	-> '(8.66-10.45]'
379	< 409995.0	-> '(10.45-12.24]'
380	< 411195.0	-> '(8.66-10.45]'
381	< 412005.0	-> '(10.45-12.24]'
382	< 412545.0	-> '(8.66-10.45]'
383	< 413625.0	-> '(10.45-12.24]'
384	< 414075.0	-> '(8.66-10.45]'
385	< 414855.0	-> '(10.45-12.24]'
386	< 415935.0	-> '(8.66-10.45]'
387	< 416085.0	-> '(10.45-12.24]'
388	< 417015.0	-> '(8.66-10.45]'
389	< 417195.0	-> '(10.45-12.24]'
390	< 418395.0	-> '(8.66-10.45]'
391	< 418725.0	-> '(10.45-12.24]'
392	< 418995.0	-> '(8.66-10.45]'
393	< 419205.0	-> '(10.45-12.24]'
394	< 419835.0	-> '(8.66-10.45]'
395	< 420105.0	-> '(10.45-12.24]'
396	< 424155.0	-> '(8.66-10.45]'
397	< 424815.0	-> '(10.45-12.24]'
398	< 425535.0	-> '(8.66-10.45]'
399	< 425655.0	-> '(10.45-12.24]'
400	< 425895.0	-> '(8.66-10.45]'
401	< 426315.0	-> '(10.45-12.24]'
402	< 427185.0	-> '(8.66-10.45]'
403	< 427275.0	-> '(10.45-12.24]'
404	< 427365.0	-> '(8.66-10.45]'
405	< 427575.0	-> '(10.45-12.24]'
406	< 428505.0	-> '(8.66-10.45]'
407	< 429015.0	-> '(10.45-12.24]'
408	< 438615.0	-> '(8.66-10.45]'
409	< 438825.0	-> '(10.45-12.24]'
410	< 440145.0	-> '(8.66-10.45]'
411	< 440235.0	-> '(10.45-12.24]'
412	< 440565.0	-> '(8.66-10.45]'
413	< 440985.0	-> '(10.45-12.24]'
414	< 441075.0	-> '(8.66-10.45]'
415	< 449985.0	-> '(10.45-12.24]'
416	< 450525.0	-> '(12.24-14.03]'
417	< 451335.0	-> '(10.45-12.24]'
418	< 452145.0	-> '(12.24-14.03]'
419	< 452865.0	-> '(10.45-12.24]'
420	< 453495.0	-> '(12.24-14.03]'
421	< 454245.0	-> '(10.45-12.24]'
422	< 454845.0	-> '(12.24-14.03]'
423	< 460095.0	-> '(10.45-12.24]'

424	< 460545.0	-> '(12.24-14.03]'
425	< 461595.0	-> '(10.45-12.24]'
426	< 462015.0	-> '(12.24-14.03]'
427	< 463125.0	-> '(10.45-12.24]'
428	< 463425.0	-> '(12.24-14.03]'
429	< 463875.0	-> '(10.45-12.24]'
430	< 464145.0	-> '(12.24-14.03]'
431	< 464595.0	-> '(10.45-12.24]'
432	< 464895.0	-> '(12.24-14.03]'
433	< 465375.0	-> '(10.45-12.24]'
434	< 465615.0	-> '(12.24-14.03]'
435	< 466095.0	-> '(10.45-12.24]'
436	< 466395.0	-> '(12.24-14.03]'
437	< 466665.0	-> '(10.45-12.24]'
438	< 466755.0	-> '(12.24-14.03]'
439	< 466875.0	-> '(10.45-12.24]'
440	< 467205.0	-> '(12.24-14.03]'
441	< 467565.0	-> '(10.45-12.24]'
442	< 468345.0	-> '(12.24-14.03]'
443	< 468465.0	-> '(10.45-12.24]'
444	< 468795.0	-> '(12.24-14.03]'
445	< 469035.0	-> '(10.45-12.24]'
446	< 472545.0	-> '(12.24-14.03]'
447	< 472725.0	-> '(10.45-12.24]'
448	< 474165.0	-> '(12.24-14.03]'
449	< 474285.0	-> '(10.45-12.24]'
450	< 474765.0	-> '(12.24-14.03]'
451	< 475065.0	-> '(10.45-12.24]'
452	< 475545.0	-> '(12.24-14.03]'
453	< 475725.0	-> '(10.45-12.24]'
454	< 487695.0	-> '(12.24-14.03]'
455	< 490125.0	-> '(10.45-12.24]'
456	< 496275.0	-> '(12.24-14.03]'
457	< 497355.0	-> '(10.45-12.24]'
458	< 511845.0	-> '(12.24-14.03]'
459	< 512205.0	-> '(10.45-12.24]'
460	< 524595.0	-> '(12.24-14.03]'
461	>= 524595.0	-> '(14.03-15.82]'

ALL_NOMINAL:

TempCelsius:

'(-inf-3.44]'	-> '(12.24-14.03]'
'(3.44-6.78]'	-> '(12.24-14.03]'
'(6.78-10.12]'	-> '(10.45-12.24]'
'(10.12-13.46]'	-> '(8.66-10.45]'
'(13.46-16.8]'	-> '(8.66-10.45]'
'(16.8-20.14]'	-> '(8.66-10.45]'
'(20.14-23.48]'	-> '(6.87-8.66]'
'(23.48-26.82]'	-> '(6.87-8.66]'
'(26.82-30.16]'	-> '(6.87-8.66]'
'(30.16-inf)'	-> '(8.66-10.45]'

? -> '(10.45-12.24]'

ALLSUPERVISED:

TempCelsius:

'(-inf-2.75]' -> '(14.03-15.82]'
'(2.75-3.35]' -> '(12.24-14.03]'
'(3.35-3.55]' -> '(12.24-14.03]'
'(3.55-3.85]' -> '(12.24-14.03]'
'(3.85-4.45]' -> '(12.24-14.03]'
'(4.45-4.65]' -> '(12.24-14.03]'
'(4.65-4.85]' -> '(12.24-14.03]'
'(4.85-5.05]' -> '(12.24-14.03]'
'(5.05-5.55]' -> '(12.24-14.03]'
'(5.55-5.85]' -> '(12.24-14.03]'
'(5.85-6.15]' -> '(12.24-14.03]'
'(6.15-6.35]' -> '(12.24-14.03]'
'(6.35-6.45]' -> '(12.24-14.03]'
'(6.45-7.25]' -> '(12.24-14.03]'
'(7.25-7.85]' -> '(10.45-12.24]'
'(7.85-8.25]' -> '(10.45-12.24]'
'(8.25-8.35]' -> '(10.45-12.24]'
'(8.35-8.65]' -> '(8.66-10.45]'
'(8.65-9.15]' -> '(10.45-12.24]'
'(9.15-9.45]' -> '(10.45-12.24]'
'(9.45-10.05]' -> '(10.45-12.24]'
'(10.05-10.75]' -> '(10.45-12.24]'
'(10.75-11.15]' -> '(10.45-12.24]'
'(11.15-11.65]' -> '(10.45-12.24]'
'(11.65-12.35]' -> '(10.45-12.24]'
'(12.35-12.55]' -> '(8.66-10.45]'
'(12.55-13.25]' -> '(8.66-10.45]'
'(13.25-13.55]' -> '(8.66-10.45]'
'(13.55-14.05]' -> '(8.66-10.45]'
'(14.05-14.15]' -> '(8.66-10.45]'
'(14.15-14.35]' -> '(8.66-10.45]'
'(14.35-14.75]' -> '(8.66-10.45]'
'(14.75-14.95]' -> '(8.66-10.45]'
'(14.95-15.15]' -> '(8.66-10.45]'
'(15.15-15.95]' -> '(8.66-10.45]'
'(15.95-16.65]' -> '(8.66-10.45]'
'(16.65-16.85]' -> '(8.66-10.45]'
'(16.85-17.15]' -> '(8.66-10.45]'
'(17.15-17.65]' -> '(8.66-10.45]'
'(17.65-17.95]' -> '(8.66-10.45]'
'(17.95-19.05]' -> '(8.66-10.45]'
'(19.05-19.35]' -> '(8.66-10.45]'
'(19.35-19.85]' -> '(8.66-10.45]'
'(19.85-20.75]' -> '(6.87-8.66]'
'(20.75-21.35]' -> '(6.87-8.66]'
'(21.35-21.85]' -> '(6.87-8.66]'
'(21.85-22.55]' -> '(6.87-8.66]'
'(22.55-23.55]' -> '(6.87-8.66]'

'(23.55-24.55]' -> '(6.87-8.66]'
'(24.55-25.05]' -> '(6.87-8.66]'
'(25.05-25.45]' -> '(6.87-8.66]'
'(25.45-25.75]' -> '(6.87-8.66]'
'(25.75-26.25]' -> '(6.87-8.66]'
'(26.25-26.55]' -> '(6.87-8.66]'
'(26.55-27.05]' -> '(3.29-5.08]'
'(27.05-27.45]' -> '(5.08-6.87]'
'(27.45-27.65]' -> '(3.29-5.08]'
'(27.65-27.85]' -> '(3.29-5.08]'
'(27.85-28.35]' -> '(3.29-5.08]'
'(28.35-28.65]' -> '(6.87-8.66]'
'(28.65-29.15]' -> '(6.87-8.66]'
'(29.15-29.85]' -> '(6.87-8.66]'
'(29.85-32.35]' -> '(6.87-8.66]'
'(32.35-inf)' -> '(10.45-12.24]'
? -> '(10.45-12.24]'

DO_USEEQUALFREQ:

1	MinuteOfYear:	
2	< 98446.0	-> '(12.24-14.03]'
3	< 99016.0	-> '(10.45-12.24]'
4	< 99766.0	-> '(12.24-14.03]'
5	< 100396.0	-> '(10.45-12.24]'
6	< 101536.0	-> '(12.24-14.03]'
7	< 101776.0	-> '(10.45-12.24]'
8	< 104476.0	-> '(12.24-14.03]'
9	< 104746.0	-> '(10.45-12.24]'
10	< 105810.5	-> '(12.24-14.03]'
11	< 106455.0	-> '(10.45-12.24]'
12	< 107295.0	-> '(12.24-14.03]'
13	< 111285.0	-> '(10.45-12.24]'
14	< 111555.0	-> '(12.24-14.03]'
15	< 120465.0	-> '(10.45-12.24]'
16	< 123075.0	-> '(8.66-10.45]'
17	< 123255.0	-> '(10.45-12.24]'
18	< 123855.0	-> '(8.66-10.45]'
19	< 127125.0	-> '(10.45-12.24]'
20	< 127245.0	-> '(8.66-10.45]'
21	< 130935.0	-> '(10.45-12.24]'
22	< 131325.0	-> '(8.66-10.45]'
23	< 134175.0	-> '(10.45-12.24]'
24	< 135945.0	-> '(8.66-10.45]'
25	< 136245.0	-> '(10.45-12.24]'
26	< 137415.0	-> '(8.66-10.45]'
27	< 137835.0	-> '(10.45-12.24]'
28	< 138765.0	-> '(8.66-10.45]'
29	< 139245.0	-> '(10.45-12.24]'
30	< 140445.0	-> '(8.66-10.45]'
31	< 140685.0	-> '(10.45-12.24]'
32	< 141735.0	-> '(8.66-10.45]'
33	< 142155.0	-> '(10.45-12.24]'

34	< 143265.0	-> '(8.66-10.45]'
35	< 143625.0	-> '(10.45-12.24]'
36	< 144645.0	-> '(8.66-10.45]'
37	< 145035.0	-> '(10.45-12.24]'
38	< 146085.0	-> '(8.66-10.45]'
39	< 146445.0	-> '(10.45-12.24]'
40	< 147555.0	-> '(8.66-10.45]'
41	< 147705.0	-> '(10.45-12.24]'
42	< 148005.0	-> '(8.66-10.45]'
43	< 148185.0	-> '(6.87-8.66]'
44	< 148875.0	-> '(8.66-10.45]'
45	< 149385.0	-> '(10.45-12.24]'
46	< 150435.0	-> '(8.66-10.45]'
47	< 150855.0	-> '(10.45-12.24]'
48	< 151935.0	-> '(8.66-10.45]'
49	< 152385.0	-> '(10.45-12.24]'
50	< 153645.0	-> '(8.66-10.45]'
51	< 153795.0	-> '(10.45-12.24]'
52	< 154305.0	-> '(8.66-10.45]'
53	< 154635.0	-> '(6.87-8.66]'
54	< 155775.0	-> '(8.66-10.45]'
55	< 156075.0	-> '(6.87-8.66]'
56	< 156645.0	-> '(8.66-10.45]'
57	< 156825.0	-> '(6.87-8.66]'
58	< 158685.0	-> '(8.66-10.45]'
59	< 158925.0	-> '(6.87-8.66]'
60	< 159975.0	-> '(8.66-10.45]'
61	< 160485.0	-> '(6.87-8.66]'
62	< 160785.0	-> '(8.66-10.45]'
63	< 160875.0	-> '(6.87-8.66]'
64	< 161175.0	-> '(8.66-10.45]'
65	< 161895.0	-> '(6.87-8.66]'
66	< 162255.0	-> '(8.66-10.45]'
67	< 162405.0	-> '(6.87-8.66]'
68	< 162825.0	-> '(8.66-10.45]'
69	< 163065.0	-> '(6.87-8.66]'
70	< 163605.0	-> '(8.66-10.45]'
71	< 163995.0	-> '(10.45-12.24]'
72	< 164745.0	-> '(8.66-10.45]'
73	< 164835.0	-> '(10.45-12.24]'
74	< 165015.0	-> '(8.66-10.45]'
75	< 165435.0	-> '(10.45-12.24]'
76	< 166185.0	-> '(8.66-10.45]'
77	< 167205.0	-> '(10.45-12.24]'
78	< 167685.0	-> '(8.66-10.45]'
79	< 168495.0	-> '(10.45-12.24]'
80	< 169035.0	-> '(8.66-10.45]'
81	< 170265.0	-> '(10.45-12.24]'
82	< 170355.0	-> '(8.66-10.45]'
83	< 171555.0	-> '(10.45-12.24]'
84	< 171855.0	-> '(8.66-10.45]'

85	< 172575.0	-> '(10.45-12.24]'
86	< 173295.0	-> '(8.66-10.45]'
87	< 174075.0	-> '(10.45-12.24]'
88	< 174795.0	-> '(8.66-10.45]'
89	**< 175335.0**	**-> '(10.45-12.24]' (NOTE: LAST HI DO 121 days into year 4/30/2012)**
90	< 180825.0	-> '(8.66-10.45]'
91	< 180945.0	-> '(6.87-8.66]'
92	< 181515.0	-> '(8.66-10.45]'
93	< 183675.0	-> '(6.87-8.66]'
94	< 183975.0	-> '(8.66-10.45]'
95	< 184365.0	-> '(6.87-8.66]'
96	< 185025.0	-> '(8.66-10.45]'
97	< 186165.0	-> '(6.87-8.66]'
98	< 186405.0	-> '(8.66-10.45]'
99	< 186555.0	-> '(6.87-8.66]'
100	< 186645.0	-> '(8.66-10.45]'
101	< 187605.0	-> '(6.87-8.66]'
102	< 187695.0	-> '(8.66-10.45]'
103	< 188025.0	-> '(6.87-8.66]'
104	< 193755.0	-> '(8.66-10.45]'
105	< 195705.0	-> '(6.87-8.66]'
106	< 200055.0	-> '(8.66-10.45]'
107	< 200775.0	-> '(6.87-8.66]'
108	< 201405.0	-> '(8.66-10.45]'
109	< 202575.0	-> '(6.87-8.66]'
110	< 202875.0	-> '(8.66-10.45]'
111	< 206595.0	-> '(6.87-8.66]'
112	< 206685.0	-> '(8.66-10.45]'
113	< 221685.0	-> '(6.87-8.66]'
114	< 221865.0	-> '(8.66-10.45]'
115	< 222195.0	-> '(6.87-8.66]'
116	< 222375.0	-> '(8.66-10.45]'
117	< 222615.0	-> '(6.87-8.66]'
118	< 222705.0	-> '(8.66-10.45]'
119	< 223275.0	-> '(6.87-8.66]'
120	< 223485.0	-> '(8.66-10.45]'
121	< 223575.0	-> '(6.87-8.66]'
122	< 223875.0	-> '(8.66-10.45]'
123	< 223965.0	-> '(6.87-8.66]'
124	< 224175.0	-> '(5.08-6.87]'
125	< 224625.0	-> '(6.87-8.66]'
126	< 224805.0	-> '(5.08-6.87]'
127	< 231975.0	-> '(8.66-10.45]'
128	< 232065.0	-> '(6.87-8.66]'
129	< 232215.0	-> '(8.66-10.45]'
130	< 232605.0	-> '(6.87-8.66]'
131	< 233415.0	-> '(8.66-10.45]'
132	< 233505.0	-> '(6.87-8.66]'
133	< 233625.0	-> '(8.66-10.45]'
134	< 234345.0	-> '(6.87-8.66]'
135	< 234465.0	-> '(8.66-10.45]'

136	< 238755.0	-> '(6.87-8.66]'
137	< 238995.0	-> '(8.66-10.45]'
138	< 239895.0	-> '(6.87-8.66]'
139	< 239985.0	-> '(8.66-10.45]'
140	< 245925.0	-> '(6.87-8.66]'
141	< 246225.0	-> '(8.66-10.45]'
142	< 246915.0	-> '(6.87-8.66]'
143	< 247035.0	-> '(8.66-10.45]'
144	< 247155.0	-> '(6.87-8.66]'
145	< 247395.0	-> '(8.66-10.45]'
146	< 247485.0	-> '(6.87-8.66]'
147	< 247665.0	-> '(8.66-10.45]'
148	< 248025.0	-> '(6.87-8.66]'
149	< 248235.0	-> '(8.66-10.45]'
150	< 248415.0	-> '(6.87-8.66]'
151	< 248475.0	-> '(8.66-10.45]'
152	< 248685.0	-> '(6.87-8.66]'
153	< 248835.0	-> '(8.66-10.45]'
154	< 252975.0	-> '(6.87-8.66]'
155	< 253065.0	-> '(8.66-10.45]'
156	< 255945.0	-> '(6.87-8.66]'
157	< 256035.0	-> '(8.66-10.45]'
158	< 256725.0	-> '(6.87-8.66]'
159	< 256815.0	-> '(5.08-6.87]'
160	< 257175.0	-> '(6.87-8.66]'
161	< 257265.0	-> '(8.66-10.45]'
162	< 258645.0	-> '(6.87-8.66]'
163	< 258885.0	-> '(8.66-10.45]'
164	< 259635.0	-> '(6.87-8.66]'
165	< 259815.0	-> '(5.08-6.87]'
166	< 260025.0	-> '(6.87-8.66]'
167	< 260205.0	-> '(8.66-10.45]'
168	< 260925.0	-> '(6.87-8.66]'
169	< 261285.0	-> '(5.08-6.87]'
170	< 261465.0	-> '(6.87-8.66]'
171	< 261885.0	-> '(8.66-10.45]'
172	< 262395.0	-> '(6.87-8.66]'
173	< 262515.0	-> '(5.08-6.87]'
174	< 262845.0	-> '(6.87-8.66]'
175	< 263325.0	-> '(8.66-10.45]'
176	< 263805.0	-> '(6.87-8.66]'
177	< 263925.0	-> '(5.08-6.87]'
178	< 264285.0	-> '(6.87-8.66]'
179	< 264405.0	-> '(8.66-10.45]'
180	**< 264885.0**	**-> '(10.45-12.24]' (NOTE Average DO Spike 184 days 7/2/2012)**
181	< 265065.0	-> '(8.66-10.45]'
182	< 265305.0	-> '(6.87-8.66]'
183	< 265575.0	-> '(5.08-6.87]'
184	< 265755.0	-> '(6.87-8.66]'
185	< 266145.0	-> '(3.29-5.08]'
186	< 266745.0	-> '(6.87-8.66]'

187	< 267045.0	-> '(3.29-5.08]'
188	< 267195.0	-> '(6.87-8.66]'
189	< 267705.0	-> '(3.29-5.08]'
190	< 268155.0	-> '(6.87-8.66]'
191	< 268515.0	-> '(3.29-5.08]'
192	< 268635.0	-> '(6.87-8.66]'
193	< 269325.0	-> '(3.29-5.08]'
194	< 269565.0	-> '(6.87-8.66]'
195	< 269745.0	-> '(5.08-6.87]'
196	< 269955.0	-> '(3.29-5.08]'
197	< 270255.0	-> '(6.87-8.66]'
198	< 270795.0	-> '(3.29-5.08]'
199	< 270975.0	-> '(6.87-8.66]'
200	< 271155.0	-> '(5.08-6.87]'
201	< 271425.0	-> '(3.29-5.08]'
202	< 271545.0	-> '(6.87-8.66]'
203	< 271845.0	-> '(3.29-5.08]'
204	< 272265.0	-> '(6.87-8.66]'
205	< 272475.0	-> '(3.29-5.08]'
206	< 272655.0	-> '(5.08-6.87]'
207	< 272895.0	-> '(3.29-5.08]'
208	< 273045.0	-> '(6.87-8.66]'
209	< 273255.0	-> '(3.29-5.08]'
210	< 273435.0	-> '(6.87-8.66]'
211	< 273705.0	-> '(8.66-10.45]'
212	< 274425.0	-> '(6.87-8.66]'
213	< 274545.0	-> '(8.66-10.45]'
214	< 278745.0	-> '(6.87-8.66]'
215	< 278865.0	-> '(8.66-10.45]'
216	< 279075.0	-> '(3.29-5.08]'
217	< 279195.0	-> '(6.87-8.66]'
218	< 279345.0	-> '(8.66-10.45]'
219	< 279945.0	-> '(6.87-8.66]'
220	< 280155.0	-> '(3.29-5.08]'
221	< 280305.0	-> '(6.87-8.66]'
222	< 280545.0	-> '(3.29-5.08]'
223	< 280935.0	-> '(6.87-8.66]'
224	< 281145.0	-> '(3.29-5.08]'
225	< 281685.0	-> '(6.87-8.66]'
226	< 281895.0	-> '(8.66-10.45]'
227	< 282375.0	-> '(6.87-8.66]'
228	< 282855.0	-> '(5.08-6.87]'
229	< 283125.0	-> '(6.87-8.66]'
230	< 283275.0	-> '(8.66-10.45]'
231	< 283725.0	-> '(6.87-8.66]'
232	< 284205.0	-> '(5.08-6.87]'
233	< 285345.0	-> '(6.87-8.66]'
234	< 285825.0	-> '(5.08-6.87]'
235	< 286605.0	-> '(6.87-8.66]'
236	< 286845.0	-> '(5.08-6.87]'
237	< 287085.0	-> '(3.29-5.08]'

238	< 287205.0	-> '(5.08-6.87]'
239	< 287835.0	-> '(6.87-8.66]'
240	< 288165.0	-> '(5.08-6.87]'
241	< 288525.0	-> '(3.29-5.08]'
242	< 288705.0	-> '(5.08-6.87]'
243	< 289395.0	-> '(6.87-8.66]'
244	< 289875.0	-> '(5.08-6.87]'
245	< 289995.0	-> '(3.29-5.08]'
246	< 291315.0	-> '(5.08-6.87]'
247	< 291435.0	-> '(6.87-8.66]'
248	< 291705.0	-> '(5.08-6.87]'
249	< 292575.0	-> '(6.87-8.66]'
250	< 292845.0	-> '(5.08-6.87]'
251	< 295155.0	-> '(6.87-8.66]'
252	< 295335.0	-> '(5.08-6.87]'
253	< 296595.0	-> '(6.87-8.66]'
254	< 296925.0	-> '(5.08-6.87]'
255	< 298065.0	-> '(6.87-8.66]'
256	< 298425.0	-> '(5.08-6.87]'
257	< 298635.0	-> '(6.87-8.66]'
258	< 298845.0	-> '(5.08-6.87]'
259	< 299325.0	-> '(6.87-8.66]'
260	< 299955.0	-> '(5.08-6.87]'
261	< 301095.0	-> '(6.87-8.66]'
262	< 301425.0	-> '(5.08-6.87]'
263	< 302805.0	-> '(6.87-8.66]'
264	< 303015.0	-> '(5.08-6.87]'
265	< 303915.0	-> '(6.87-8.66]'
266	< 304575.0	-> '(5.08-6.87]'
267	< 305355.0	-> '(6.87-8.66]'
268	< 306075.0	-> '(5.08-6.87]'
269	< 306885.0	-> '(6.87-8.66]'
270	< 307395.0	-> '(5.08-6.87]'
271	< 308385.0	-> '(6.87-8.66]'
272	< 308715.0	-> '(5.08-6.87]'
273	< 311175.0	-> '(6.87-8.66]'
274	< 311265.0	-> '(5.08-6.87]'
275	< 311415.0	-> '(6.87-8.66]'
276	< 311625.0	-> '(5.08-6.87]'
277	< 315645.0	-> '(6.87-8.66]'
278	< 316005.0	-> '(5.08-6.87]'
279	< 317145.0	-> '(6.87-8.66]'
280	< 317295.0	-> '(5.08-6.87]'
281	< 322995.0	-> '(6.87-8.66]'
282	< 323115.0	-> '(5.08-6.87]'
283	< 324405.0	-> '(6.87-8.66]'
284	< 324525.0	-> '(5.08-6.87]'
285	< 325875.0	-> '(6.87-8.66]'
286	< 326205.0	-> '(5.08-6.87]'
287	< 336375.0	-> '(6.87-8.66]'
288	< 336585.0	-> '(8.66-10.45]'

289	< 337815.0	-> '(6.87-8.66]'
290	< 338085.0	-> '(8.66-10.45]'
291	< 339285.0	-> '(6.87-8.66]'
292	< 339525.0	-> '(8.66-10.45]'
293	< 340725.0	-> '(6.87-8.66]'
294	< 340995.0	-> '(8.66-10.45]'
295	< 347385.0	-> '(6.87-8.66]'
296	< 347655.0	-> '(5.08-6.87]'
297	< 347985.0	-> '(6.87-8.66]'
298	< 348075.0	-> '(8.66-10.45]'
299	< 349335.0	-> '(6.87-8.66]'
300	< 349575.0	-> '(8.66-10.45]'
301	< 350745.0	-> '(6.87-8.66]'
302	< 351015.0	-> '(8.66-10.45]'
303	< 352305.0	-> '(6.87-8.66]'
304	< 352425.0	-> '(8.66-10.45]'
305	< 354645.0	-> '(6.87-8.66]'
306	< 355065.0	-> '(5.08-6.87]'
307	< 360165.0	-> '(6.87-8.66]'
308	< 360345.0	-> '(5.08-6.87]'
309	< 361125.0	-> '(6.87-8.66]'
310	< 361395.0	-> '(3.29-5.08]'
311	< 361635.0	-> '(6.87-8.66]'
312	< 362205.0	-> '(3.29-5.08]'
313	< 363135.0	-> '(6.87-8.66]'
314	< 363375.0	-> '(3.29-5.08]'
315	< 363705.0	-> '(6.87-8.66]'
316	< 364245.0	-> '(3.29-5.08]'
317	< 365085.0	-> '(6.87-8.66]'
318	< 365415.0	-> '(8.66-10.45]'
319	< 366525.0	-> '(6.87-8.66]'
320	< 366885.0	-> '(8.66-10.45]'
321	< 367965.0	-> '(6.87-8.66]'
322	< 368355.0	-> '(8.66-10.45]'
323	< 369345.0	-> '(6.87-8.66]'
324	< 369705.0	-> '(8.66-10.45]'
325	< 370875.0	-> '(6.87-8.66]'
326	< 371145.0	-> '(8.66-10.45]'
327	< 372165.0	-> '(6.87-8.66]'
328	< 372645.0	-> '(8.66-10.45]'
329	< 373665.0	-> '(6.87-8.66]'
330	< 374115.0	-> '(8.66-10.45]'
331	< 375195.0	-> '(6.87-8.66]'
332	< 375405.0	-> '(8.66-10.45]'
333	< 378015.0	-> '(6.87-8.66]'
334	< 383055.0	-> '(8.66-10.45]'
335	< 383175.0	-> '(6.87-8.66]'
336	< 383295.0	-> '(8.66-10.45]'
337	< 383415.0	-> '(6.87-8.66]'
338	< 383655.0	-> '(8.66-10.45]'
339	< 383775.0	-> '(6.87-8.66]'

340	< 383865.0	-> '(8.66-10.45]'
341	< 384165.0	-> '(6.87-8.66]'
342	< 384645.0	-> '(8.66-10.45]'
343	< 385005.0	-> '(6.87-8.66]'
344	< 385455.0	-> '(8.66-10.45]'
345	< 385695.0	-> '(6.87-8.66]'
346	< 386085.0	-> '(8.66-10.45]'
347	< 386445.0	-> '(6.87-8.66]'
348	< 387585.0	-> '(8.66-10.45]'
349	< 388035.0	-> '(6.87-8.66]'
350	< 389085.0	-> '(8.66-10.45]'
351	< 389505.0	-> '(6.87-8.66]'
352	< 390435.0	-> '(8.66-10.45]'
353	< 390975.0	-> '(6.87-8.66]'
354	< 391875.0	-> '(8.66-10.45]'
355	< 392355.0	-> '(6.87-8.66]'
356	< 393495.0	-> '(8.66-10.45]'
357	< 393735.0	-> '(6.87-8.66]'
358	< 394905.0	-> '(8.66-10.45]'
359	< 395145.0	-> '(6.87-8.66]'
360	< 396465.0	-> '(8.66-10.45]'
361	< 396765.0	-> '(6.87-8.66]'
362	< 397725.0	-> '(8.66-10.45]'
363	< 397965.0	-> '(6.87-8.66]'
364	< 400905.0	-> '(8.66-10.45]'
365	< 401085.0	-> '(5.08-6.87]'
366	< 401265.0	-> '(8.66-10.45]'
367	< 401535.0	-> '(5.08-6.87]'
368	< 402075.0	-> '(8.66-10.45]'
369	< 402525.0	-> '(5.08-6.87]'
370	< 405645.0	-> '(8.66-10.45]'
371	< 406125.0	-> '(5.08-6.87]'
372	< 406815.0	-> '(8.66-10.45]'
373	< 407085.0	-> '(5.08-6.87]'
374	< 407265.0	-> '(8.66-10.45]'
375	< 407715.0	-> '(5.08-6.87]'
376	< 408345.0	-> '(8.66-10.45]'
377	**< 408435.0**	**-> '(10.45-12.24]' (NOTE: 284 days 10/10/2012 start higher DO)**
378	< 409755.0	-> '(8.66-10.45]'
379	< 409995.0	-> '(10.45-12.24]'
380	< 411195.0	-> '(8.66-10.45]'
381	< 412005.0	-> '(10.45-12.24]'
382	< 412545.0	-> '(8.66-10.45]'
383	< 413625.0	-> '(10.45-12.24]'
384	< 414075.0	-> '(8.66-10.45]'
385	< 414855.0	-> '(10.45-12.24]'
386	< 415935.0	-> '(8.66-10.45]'
387	< 416085.0	-> '(10.45-12.24]'
388	< 417015.0	-> '(8.66-10.45]'
389	< 417195.0	-> '(10.45-12.24]'
390	< 418395.0	-> '(8.66-10.45]'

391	< 418725.0	-> '(10.45-12.24]'
392	< 418995.0	-> '(8.66-10.45]'
393	< 419205.0	-> '(10.45-12.24]'
394	< 419835.0	-> '(8.66-10.45]'
395	< 420105.0	-> '(10.45-12.24]'
396	< 424155.0	-> '(8.66-10.45]'
397	< 424815.0	-> '(10.45-12.24]'
398	< 425535.0	-> '(8.66-10.45]'
399	< 425655.0	-> '(10.45-12.24]'
400	< 425895.0	-> '(8.66-10.45]'
401	< 426315.0	-> '(10.45-12.24]'
402	< 427185.0	-> '(8.66-10.45]'
403	< 427275.0	-> '(10.45-12.24]'
404	< 427365.0	-> '(8.66-10.45]'
405	< 427575.0	-> '(10.45-12.24]'
406	< 428505.0	-> '(8.66-10.45]'
407	< 429015.0	-> '(10.45-12.24]'
408	< 438615.0	-> '(8.66-10.45]'
409	< 438825.0	-> '(10.45-12.24]'
410	< 440145.0	-> '(8.66-10.45]'
411	< 440235.0	-> '(10.45-12.24]'
412	< 440565.0	-> '(8.66-10.45]'
413	< 440985.0	-> '(10.45-12.24]'
414	< 441075.0	-> '(8.66-10.45]'
415	< 449985.0	-> '(10.45-12.24]'
416	< 450525.0	-> '(12.24-14.03]'
417	< 451335.0	-> '(10.45-12.24]'
418	< 452145.0	-> '(12.24-14.03]'
419	< 452865.0	-> '(10.45-12.24]'
420	< 453495.0	-> '(12.24-14.03]'
421	< 454245.0	-> '(10.45-12.24]'
422	< 454845.0	-> '(12.24-14.03]'
423	< 460095.0	-> '(10.45-12.24]'
424	< 460545.0	-> '(12.24-14.03]'
425	< 461595.0	-> '(10.45-12.24]'
426	< 462015.0	-> '(12.24-14.03]'
427	< 463125.0	-> '(10.45-12.24]'
428	< 463425.0	-> '(12.24-14.03]'
429	< 463875.0	-> '(10.45-12.24]'
430	< 464145.0	-> '(12.24-14.03]'
431	< 464595.0	-> '(10.45-12.24]'
432	< 464895.0	-> '(12.24-14.03]'
433	< 465375.0	-> '(10.45-12.24]'
434	< 465615.0	-> '(12.24-14.03]'
435	< 466095.0	-> '(10.45-12.24]'
436	< 466395.0	-> '(12.24-14.03]'
437	< 466665.0	-> '(10.45-12.24]'
438	< 466755.0	-> '(12.24-14.03]'
439	< 466875.0	-> '(10.45-12.24]'
440	< 467205.0	-> '(12.24-14.03]'
441	< 467565.0	-> '(10.45-12.24]'

442	< 468345.0	-> '(12.24-14.03]'
443	< 468465.0	-> '(10.45-12.24]'
444	< 468795.0	-> '(12.24-14.03]'
445	< 469035.0	-> '(10.45-12.24]'
446	< 472545.0	-> '(12.24-14.03]'
447	< 472725.0	-> '(10.45-12.24]'
448	< 474165.0	-> '(12.24-14.03]'
449	< 474285.0	-> '(10.45-12.24]'
450	< 474765.0	-> '(12.24-14.03]'
451	< 475065.0	-> '(10.45-12.24]'
452	< 475545.0	-> '(12.24-14.03]'
453	< 475725.0	-> '(10.45-12.24]'
454	< 487695.0	-> '(12.24-14.03]'
455	< 490125.0	-> '(10.45-12.24]'
456	< 496275.0	-> '(12.24-14.03]'
457	< 497355.0	-> '(10.45-12.24]'
458	< 511845.0	-> '(12.24-14.03]'
459	< 512205.0	-> '(10.45-12.24]'
460	< 524595.0	-> '(12.24-14.03]'
461	>= 524595.0	-> '(14.03-15.82]'

Q5. Does your answer to Q3 and Q4 confirm, refute, or neither, information in the PART II linked readings? Explain why.

CONFIRM, from Reading 1 https://www.fondriest.com/environmental-measurements/parameters/water-quality/dissolved-oxygen/#6 In this case season -> dissolved oxygen level. Seasonal temperature also correlates (higher temperature means on average lower DO), but MinuteOfYear is fine-grain seasonal data.

"Typical Dissolved Oxygen Levels
Dissolved oxygen concentrations are constantly affected by diffusion and aeration, photosynthesis, respiration and decomposition. While water equilibrates toward 100% air saturation, dissolved oxygen levels will also fluctuate with **temperature**, salinity and pressure changes [3]. As such, dissolved oxygen levels can range from less than 1 mg/L to more than 20 mg/L depending on how all of these factors interact. In freshwater systems such as lakes, rivers and streams, dissolved oxygen concentrations will vary by **season**, location and water depth."

ALL_NOMINAL:

Rise in TempCelsius correlates with drop in DO per the reading, so yes. TempCelsius generally correlates with MinuteOfYear, so they are equally valid EXCEPT for the late June / early July exponential spike in plant growth per Figure 2 above. MinuteOfYear can identify that spike, while TempCelsius cannot.

ALLSUPERVISED:

Same as ALL_NOMINAL, more detailed rule is harder to read but a little more accurate.

DO_USEEQUALFREQ: Same as default.

Q6. Does running the **J48** classification tree on these attributes add any accuracy to OneR's results on the same data? Explain your answer. Copy & paste "Correctly Classified %, and the "Kappa statistic" into README.txt.

```
Correctly Classified Instances      46703               95.184  %
Kappa statistic                         0.9374
```

That is a substantial improvement.

ALL_NOMINAL:

Correctly Classified Instances 40723 82.9964 %
Kappa statistic 0.7773
Worse than all discretized for J48, better than OneR on all discretized.

ALLSUPERVISED:

Correctly Classified Instances 44122 89.9238 %
Kappa statistic 0.8685

DO_USEEQUALFREQ:

Correctly Classified Instances 43876 89.4224 %
Kappa statistic 0.8818

Q7: After removing attributes so that only **MinuteOfYear**, **MinuteOfDay** and **OxygenMgPerLiter** remain, run rule **OneR**. What are the "Correctly Classified Instances %" and "Kappa statistic" values?

Correctly Classified Instances 27915 56.8928 %
Kappa statistic 0.4219
Same as previous OneR

ALL_NOMINAL:

Correctly Classified Instances 25359 51.6834 %
Kappa statistic 0.3274

ALLSUPERVISED:

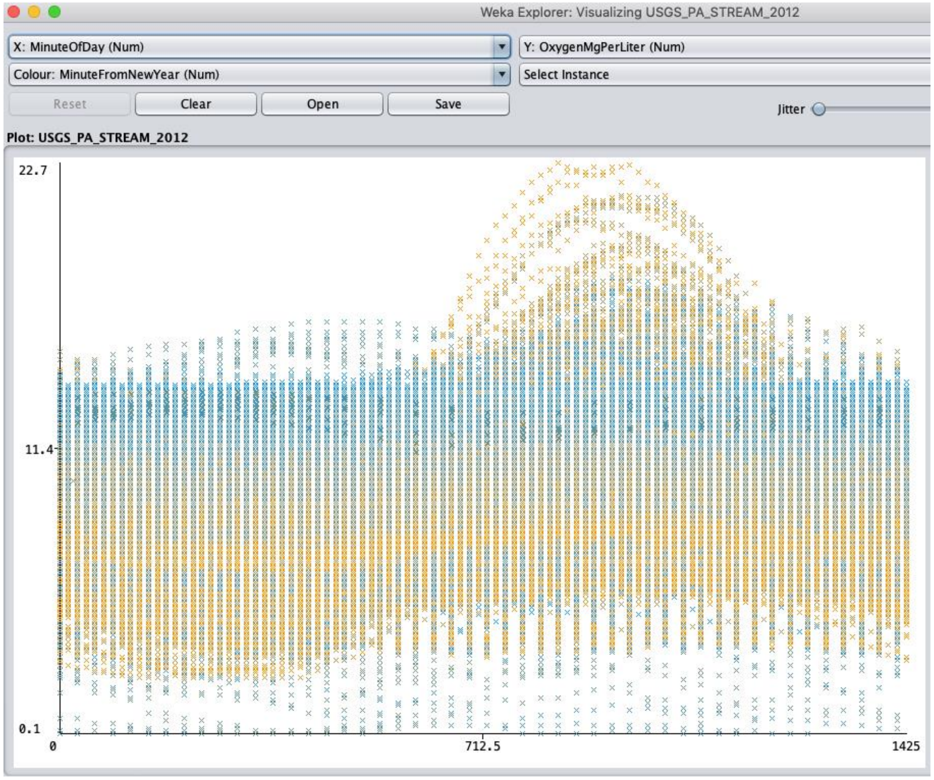
Correctly Classified Instances 29026 59.1571 %
Kappa statistic 0.4488

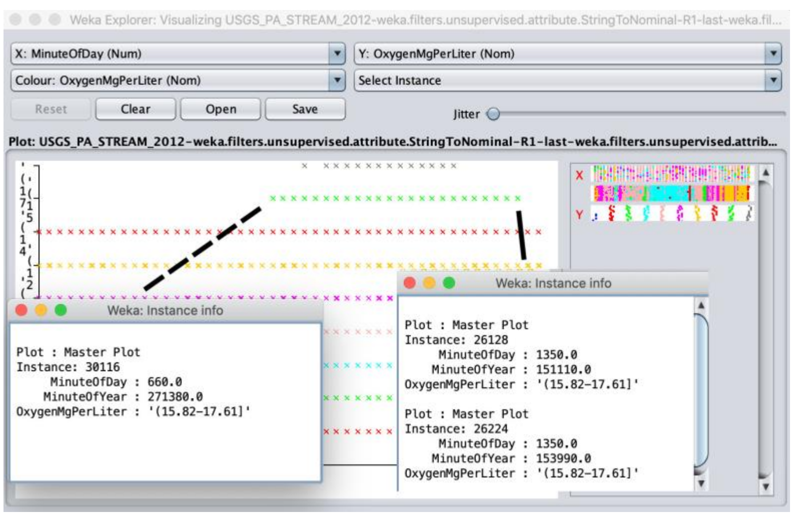
DO_USEEQUALFREQ:

Correctly Classified Instances 17926 36.5345 %
Kappa statistic 0.2893

Q8: Use the Visualize tab to inspect **OxygenMgPerLiter** on the Y axis as a function of **MinuteOfDay** on the X axis. Can you see a pattern of value changes in the afternoon that indicate photosynthesis per your Part II linked readings? Why or why not?

UPDATE March 8: Reading the 2D visualizations after Discretize on OxygenMgPerLiter is difficult. Here is the visualization of the numeric OxygenMgPerLiter value. This is easier to see:

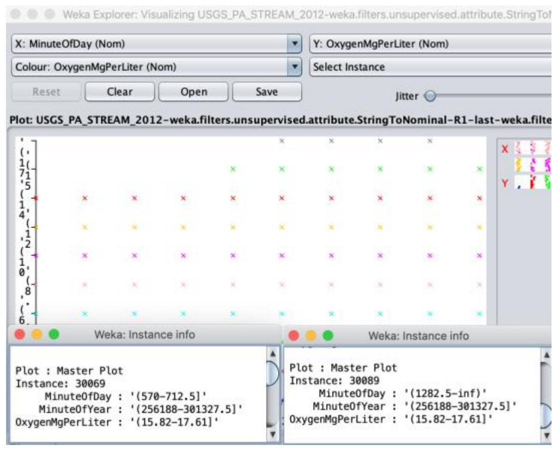




Yes there is a pattern indicating **photosynthesis**.

The Discretized DO level graph above is actually useful. Using the green line of discrete DO values in range (15.82-17.61] mg per liter as appears in the pop-up, you can see that between 660 minutes (11 AM) through 1350 minutes (10:30 PM) there is elevated average DO due to **photosynthesis**.

ALL_NOMINAL:



ALLSUPERVISED: Similar to ALL_NOMINAL
DO_USEEQUALFREQ: Impossible to see this pattern, all flat.

Q9. Repeat Q7 using the J48 tree. Do you see any change in "Correctly Classified Instances %" and "Kappa statistic" values? Show these values in README.txt.

Correctly Classified Instances 29498 60.119 %
Kappa statistic 0.4608
J48 is marginally better than OneR with these attributes, still only a Landis and Koch moderate Kappa value.

ALL_NOMINAL:
Correctly Classified Instances 25517 52.0055 %
Kappa statistic 0.336

ALLSUPERVISED:
Correctly Classified Instances 29137 59.3833 %
Kappa statistic 0.4524

DO_USEEQUALFREQ:
Correctly Classified Instances 19990 40.741 %
Kappa statistic 0.3377

Load your saved file USGS_PA_STREAM_2012_TRAIN.arff.gz **and remove site_no** to get back to a full set of attributes.
Apply **OneR**, **J48**, and **RandomTree** to your file's dataset to predict **OxygenMgPerLiter**.

Q10. What are their respective "Correctly Classified Instances" and "Kappa statistic" values? Which is the most predictive?

OneR:
Correctly Classified Instances 27915 56.8928 %
Kappa statistic 0.4219

J48:
Correctly Classified Instances 46703 95.184 %
Kappa statistic 0.9374

RandomTree:
Correctly Classified Instances 46843 95.4694 %
Kappa statistic 0.9411

RandomTree is most accurate and marginally better than J48. Both trees are hard to interpret manually.

ALL_NOMINAL:

OneR:
Correctly Classified Instances 27927 56.9172 %
Kappa statistic 0.4129
J48:
Correctly Classified Instances 40723 82.9964 %
Kappa statistic 0.7773
RandomTree:
Correctly Classified Instances 40775 83.1024 %
Kappa statistic 0.7789

ALLSUPERVISED:
OneR:
Correctly Classified Instances 29547 60.2189 %
Kappa statistic 0.4723
J48:
Correctly Classified Instances 44122 89.9238 %
Kappa statistic 0.8685
RandomTree:
Correctly Classified Instances 45599 92.934 %
Kappa statistic 0.9081

DO_USEEQUALFREQ:
OneR:
Correctly Classified Instances 17926 36.5345 %
Kappa statistic 0.2893
J48:
Correctly Classified Instances 43876 89.4224 %
Kappa statistic 0.8818
RandomTree:
Correctly Classified Instances 44216 90.1154 %
Kappa statistic 0.8895

Q11. Which of Q12 is the easiest to understand in terms of looking at the rule or tree structure? Why?

OneR because it is possible to see the slopes in DO as a function of MinuteOfYear. The trees are very complicated to try to read.

ALL_NOMINAL:
OneR because it is possible to see the correlation in DO as a function of TempCelsius. The trees are very complicated to try to read.
ALLSUPERVISED: Same as ALL_NOMINAL.
DO_USEEQUALFREQ: Same as above, OneR.

In the Classify tab click the **Supplied test set** radio button and set it to USGS_PA_STREAM_2012_TEST.arff.gz. We are changing from cross-validation on the training data to validation against USGS_PA_STREAM_2012_TEST.arff.gz test data, which contains some non-representative sampling sites. A decrease in kappa accuracy indicates some degree of over-fitting the learned models to the training data.

You may get a warning because the site_no attribute in USGS_PA_STREAM_2012_TEST.arff.gz is missing from USGS_PA_STREAM_2012_TRAIN.arff.gz from which your models are learned. Just click through the warning and proceed with testing. Weka will show this attribute mapping:

Attribute mappings:

```
Model attributes              Incoming attributes
----------------------------  ----------------
(numeric) pH                  --> 2 (numeric) pH
(numeric) TempCelsius         --> 3 (numeric) TempCelsius
(numeric) Conductance         --> 4 (numeric) Conductance
(numeric) DischargeRate       --> 5 (numeric) DischargeRate
(numeric) MinuteOfDay         --> 6 (numeric) MinuteOfDay
(numeric) MinuteOfYear        --> 7 (numeric) MinuteOfYear
(nominal) OxygenMgPerLiter    --> 8 (nominal) OxygenMgPerLiter
```

Q12. Apply **OneR**, **J48**, and **RandomTree** to predict **OxygenMgPerLiter**. What are their respective "Correctly Classified Instances" and "Kappa statistic" values? Did they increase, decrease, or stay about the same? Do you think that there was some over-fitting to the training data compared with cross-validation? Explain.

OneR:
    Correctly Classified Instances 27915 47.0284 % (down from 56.8928 %)
    Kappa statistic 0.2957 (way down from 0.4219, moderate to fair)
J48:
    Correctly Classified Instances 46703 46.1935 % (way down from 95.184 %)
    Kappa statistic 0.2929 (down from 0.9374, actually worse than OneR)
RandomTree:
    Correctly Classified Instances 46843 48.108 % (down from 95.4694 %)
    Kappa statistic 0.3037 (down from 0.9411)

Yes, the substantial drop in the % correct and Kappa for the decision trees indicates over-fitting to the training dataset.

ALL_NOMINAL:
  OneR:
    Correctly Classified Instances 216266 51.6211 %
    Kappa statistic 0.3381
  J48:
    Correctly Classified Instances 191936 45.8137 %
    Kappa statistic 0.2855
  RandomTree:
    Correctly Classified Instances 195355 46.6298 %
    Kappa statistic 0.2962

ALLSUPERVISED:
  OneR:
    Correctly Classified Instances 197904 47.2382 %
    Kappa statistic 0.303
  J48:
    Correctly Classified Instances 195837 46.7448 %
    Kappa statistic 0.2961
  RandomTree:
    Correctly Classified Instances 195782 46.7317 %
    Kappa statistic 0.2959
DO_USEEQUALFREQ:
  OneR:
    Correctly Classified Instances 103480 24.6999 %
    Kappa statistic 0.1642
  J48:
    Correctly Classified Instances 104190 24.8694 %
    Kappa statistic 0.1646
  RandomTree:
    Correctly Classified Instances 98519 23.5158 %
    Kappa statistic 0.1503

NOTE: Building the trees takes some time. You will see "building model for fold 1" through "fold 10" at the bottom left during its run. Weka uses a subset of the data set instances for training and the other instances for testing. It is important to separate training data from test data, so as not to pollute the tests with over-fitting. In this case Weka is using *ten-fold cross-validation*. It randomly picks 10 equal-size, distinct partitions (folds) of the instances. It uses 9 for training and 1 for testing, then swaps the 1 testing fold into the training set and pulls an unused fold for testing an learns again, and so on, until each fold has appeared as the test data set once, and in the training data 9 times. Weka can maximize the generality of its learned structures on moderate data set sizes this way. It is also possible to use distinct training and testing files. We may do so later in the semester.

When you have completed all of your work and double-checked the assignment requirements, make sure that both **USGS_PA_STREAM_2012_NOMINAL.arff.gz** saved at the end of Part I, and your **README.txt** that answers Q1 through Q12, turned in to the D2L Assignment 2 page Late assignments lose 10% per day late, and I will not accept an assignment after I go over its solution in class.

Filters
Attribute Unsupervised
  Reorder (to make the target the last attribute at the bottom)
  StringToNominal (to turn a limited number of strings into a usable
          set of symbols, for example, we turned
          "d" "r" "l" "p" for lsTOarff files into
          nominal set {d, r, l, p} which means
          directory, regular-file, symbolic-link, or
          named-pipe. If you added more file types by
          getting more data (e.g., "b" for block IO, and
          "c" for character IO device files), you would
          have to extend the arff file nominals to include
          {d, r, l, p, b, c}.
   In addition to StringToNominal, just remove the date fields.
   If we had time intervals, we could compute a numeric time interval

(onedate - otherdate), in terms of hours, days, etc. in Python.

Discretize (to turn numeric attributes into nominal BINS of values, e.g., like turning numeric grades into A, Aminus, etc.)

useEqualFrequency of FALSE distributes bins across numeric range of the attribute.

useEqualFrequency of TRUE tries to distribute bins across equals sizes.

If the filter such as Discretize has no effect, try setting ignoreClass to TRUE and run it again. When ignoreClass is false, some of the filters try to correlate the attribute being filtered with the target attribute, also known as the class.

Instance Unsupervised filters

RemoveWithValues to eliminate two unrepresentative sites.

Rules

ZeroR just to see it. It just picks the most popular target value.

OneR maps the most predictive attribute to the target attribute.

We iterated and used OneR to pick the next most predictive attribute, removed it temporarily, and did this again, until this process became less predictive, or we got enough attributes.

This process is partially redundant with Weka's "Select attributes" Tab.

Trees

J48 -- top-down partition (splitting) of attribute-to-target mappings

Sometimes setting unpruned to TRUE increases accuracy at the cost of human intelligibility of the tree.

RandomTree -- bottom-up partition (splitting) of attribute-to-target mappings