

Project: Business Intelligence

Due: Tuesday, Sept. 5

Business Intelligence (BI) technologies provide historical, current and predictive views of business operations. Common functions of business intelligence technologies include reporting, online analytical processing, analytics, and data mining.

BI technologies can handle large amounts of structured and sometimes unstructured data to help identify, develop and otherwise create new strategic business opportunities.

This is all about converting **raw data** into usable **knowledge**. Management does not want the details. They are far too busy. They need the key takeaways only.

The Challenge

Your task is to write a BI application to compute several summary statistics of a critical data set. The attached file contains a critical monthly operating statistic for a U.S. nuclear plant that was commissioned in 1950. Each number is the amount of nuclear fuel consumed each month. And you have 60 years worth of data.

Management needs to know if the nuclear plant is still at peak operating performance. The following statistics are needed to make a well informed assessment:

SUMMARY STATS

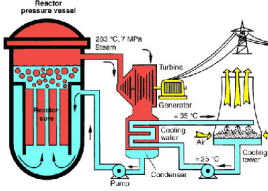
- Average value for all months
- Historical minimum, and the month and year it occurred.
- Historical maximum, and the month and year it occurred.

TREND ANALYSIS

- Average value for each decade. Assume a decade begins in 1950 and ends in 1959, for example.

SEASONALITY

- Average values for January, April, July, and October



The Task

Write a program that can read in a large data set from a text file, and populate an appropriate Java data structure. From there you can unleash your classic, powerful algorithms to mine the data for the critical management information needed. You will then write all the above statistics to a text file for submission.

Your file is a comma separated file, or CSV file, named **Uranium Fuel.CSV**

Verifying your data is a critical step. Often called “scrubbing” your data, it is vital to the integrity of your final output. Garbage In, Garbage Out, as they say. Input files are rarely structured in a way to be used directly by an application. Processing always involves parsing all the data into elements, removing unwanted labels or extraneous columns, and then populating data structures.

This is a large input file, so you will need to first work on a smaller sample, to confirm the robustness of your file operations and your analytics. Here are the key statistics for the first decade:

1950 through 1959

average:	106,554.4
min:	100,207.0
max:	111,947.0
decade average:	should be same as above
January:	106,722.6
April:	107,593.7
July:	106,769.1
October:	104,910.1

After testing on the sample file is running smoothly, run your program to read the entire input file, process it, and write your output to a **space** delimited text file. The naming convention of your output file should be consistent with your input file.

Submission

Upload your source code and your output text file to Canvas.