

# CPSC 327

## In Class Project

### Problem

You have an input file that looks like this

name	age	t-shirt
name1	,14	,small
name2	,14	,small
name3	,16	,small
name4	,	,small
name5	,16	,small
name6	,16	,medium
name7	,16	,medium
name8	,	,medium
name9	,17	,medium
name10	,	,medium
name11	,18	,large
name12	,18	,large
name13	,19	,large
name14	,19	,large
name15	,19	,large

Notice that there is missing data for age. Often you have to impute (approximate) values for this data.

Of course it will not be perfect, but there are ways to make it close. Here are a couple of standard ways;

- fill in with the average for the column
- fill in with the average of a subset of rows for that column (for instance if the average for 'small' is 15, so set missing small rows to 15)

### Requirements

Design an app the takes 2 file names as command line parameters;

- infile where initial data comes from
- outfile where the cleaned data goes

Please provide a library that handles file IO