Analyzing Basket Data

- ## How to Run

  In order to run the program, you should supply two mandatory parameters:

  data_file_path, output_file_path.

  Both will take string to the path of the files. The output file doesn't have to be existed; but the program will check if it is possible to create it.

  The third parameter is number_of_items_in_combinations which is 2 by default and it couldn't be less than 2. This will address the need if we want to look for occurrence of more than 2 products in baskets.

  There is a progress report while application is running and it will report every 2000 items passed for each step.

- ## Why it computes the result correctly

  The application will read the source data file ad it will store every basket data in a separate file. However, it will not just save the products in a file, instead it will create the possible combinations of each basket and will store each baskets combination to its corresponding file.

  In the next phase, program will iterate throw all generated files and will store each combination from every file in a dictionary in which the key is the combination itself. If the key existed in the dictionary it will simply add to its value which in this case would be the occurrence of the combination. So, every time it finds a matching key it will be one more time for its occurrence.

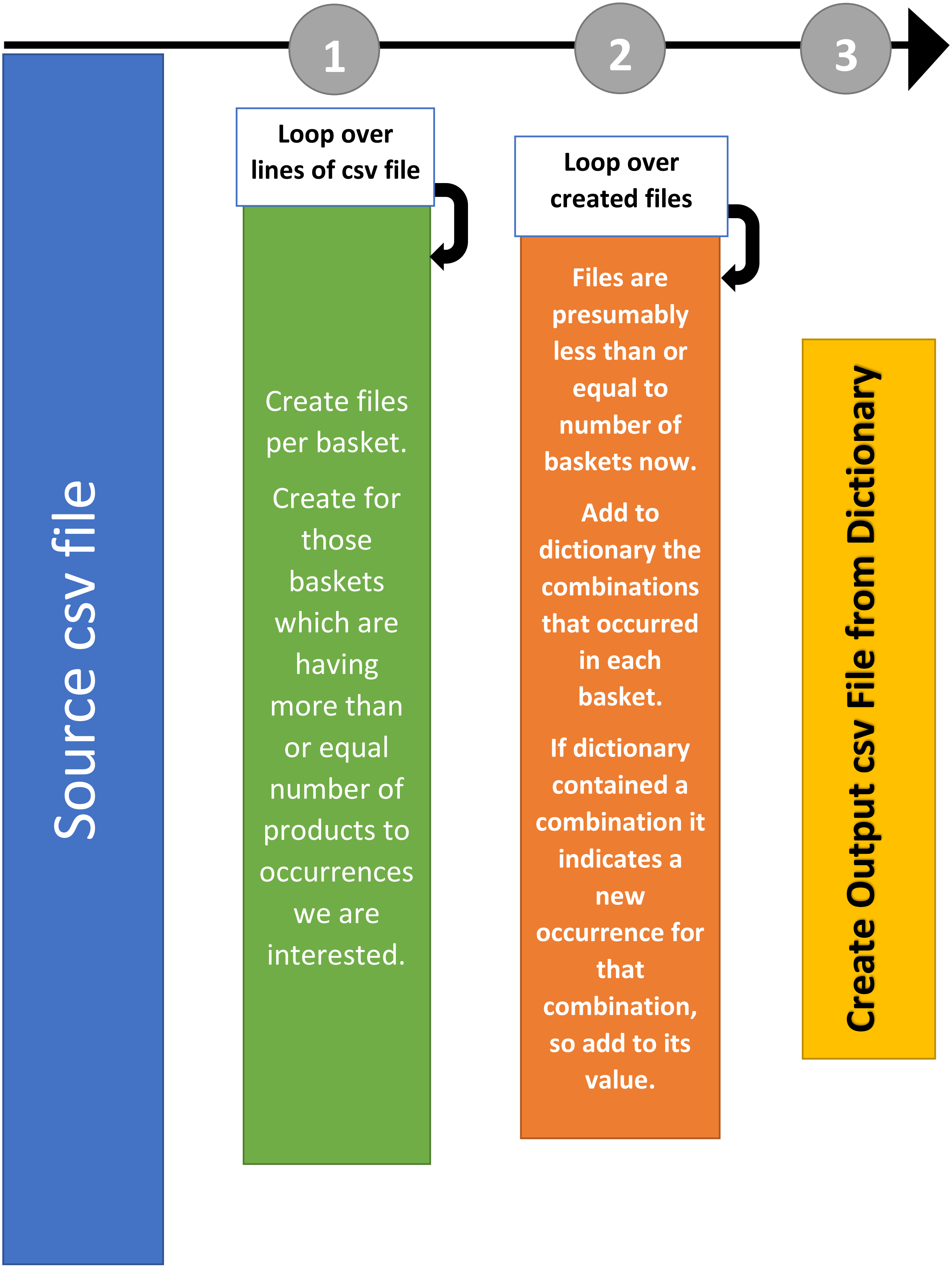
  Finally, by simply writing the dictionary to a csv file we should have the result.

- ## Why it works within the memory constraints

  Because there is no loading of file into memory it will just iterate throw lines which is a fast operation. Also, because intermediate files are very small and simple, it does not take a huge memory or processing time.

Bonus Items:

- Program tested with different scales and it was behaving as desired
- The application will work regardless of how many times a combination occurs, so there would be no difficulty in having such a case.
- For the ability to test different number of occurrences the default parameter (number_of_items_in_combinations) is created. Changing the number would result in processing for different number of occurrences. The point is it is limited to numbers greater than or equal to 2, because occurrences of one is just meaningless!



Source csv file

Loop over lines of csv file

Create files per basket.

Create for those baskets which are having more than or equal number of products to occurrences we are interested.

Loop over created files

Files are presumably less than or equal to number of baskets now.

Add to dictionary the combinations that occurred in each basket.

If dictionary contained a combination it indicates a new occurrence for that combination, so add to its value.

Create Output csv File from Dictionary