

Advanced Java – Fall 2021

Assignment 1

January 17, 2022

Contents

0.1	Introduction	2
0.2	Overview of Implementation	2
0.2.1	Classes	2
0.2.2	Dataset Reading	2
0.2.3	Statistics	2
0.2.4	Data Structure + Results	3

0.1 Introduction

To read the live covid dataset of the world and so some basic statistics over this data.

This will be a command-line based application where the values for which the statistics needs to be shown will be passed via command line arguments.

The CSV file contains a header line, followed by (as of now) 151K lines of data.

0.2 Overview of Implementation

0.2.1 Classes

- **Controller** - The Main Class which does the processing, reading and creating entities.
- **Row** - The entity class which holds the structure of the data to be stored from the dataset file.
- **Pair** - A Pair class of String, double. Used while creating the Heap.

0.2.2 Dataset Reading

- Read each line of the dataset after skipping the header using File Stream.
- Split each line using my custom split function. `mySplit()`
- For each line created an instance of required attributes needed for the statistics.

0.2.3 Statistics

- Created a Map based on which display parameter is required i.e **COUNTRY**, **DATE**, **CONTINENT**
- Aggregated values for each unique entry of the above based on the **by** value passed via command line.
- **NC**, **ND**, **TD**, **NCS**, **NDS**
- Created a function `getNewValue()` which calculates the updated aggregate based on which **by** value has been passed.

0.2.4 Data Structure + Results

- Created a Priority Queue with a custom Comparator.
- Either Min Heap or Max Heap based on command line argument `stat - min / max`
- Inserted the aggregated map values into the created Heap.
- Extracted top/bottom `limit` values to display the result.

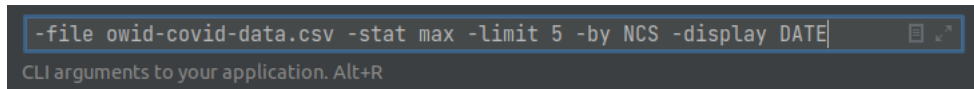


Figure 1: Command Line Arguments

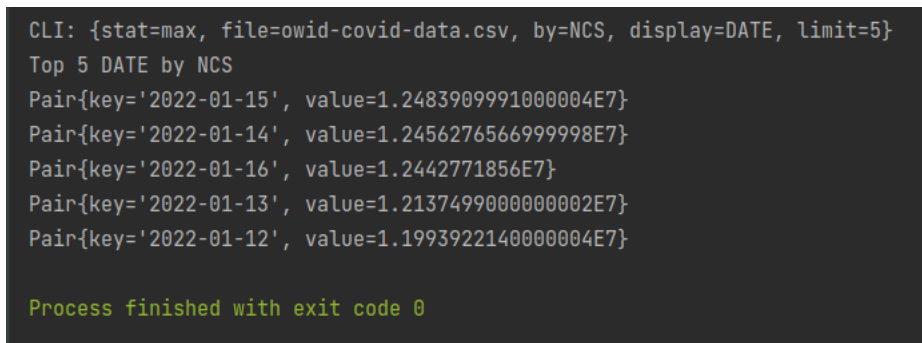


Figure 2: Sample Results