

Status Report

PROJECT SUMMARY

Report Date 23/01/2022

Project Name

Prepared By Kostandin Vllahu & Flavian Sokolli

January 18, 2022

statistical processing on Covid dataset

STATUS SUMMARY

This project is developed to perform statistical operation on given Covid-Dataset .

The CSV file contains a header line, 151K lines of data. Each data line contains COVID information about a particular country on a particular day. While the CSV file contains 60+ columns, you will only focus on the following

1. iso_code (COD)
2. continent (CNT)
3. location (LOC)
4. date (DT)
5. total_cases (TC)
6. new_cases (NC)
7. new_cases_smoothed (NCS)
8. total_deaths (TD)
9. new_deaths (ND)
10. new_deaths_smoothed (NDS)
11. reproduction_rate (RR)
12. new_tests (NT)
13. total_tests (TT)
14. stringency_index (SI)
15. population (POP)
16. median_age (MA)

Working of program:

1. Step-1: Take input from user and validate its pattern
2. Step-2: After taking valid input parse into tokens, (token. is command, which is broken. Into pieces)
3. Step-3 : Different tokens have different meanings
4. Step-4: Read csv file from given path
5. Step-5: finally display the result

Input Command structure and parameters :

The command line interface accept the format :

-file pathToFile -param1 value1 -param2 value2 ... -paramN valueN.

The parameters (param1 .. paramN) can be stat, limit, by and display. Values that of these parameters as follow :

- 1. file: path to covid data file
- 2. stat: the statistics to be computed, can assume value either min or max
- 3. limit: integer in the range [1..100].
- 4. by: the field on which the statistics will be computed. It can have value NC, NCS, ND, NDS, NT, NDPC .
- 5. Display: The type of data to be displayed. Values can be DATE, COUNTRY, CONTINENT..

DataSet:

FileHomeInsertPage LayoutFormulasDataReviewViewTell me what you want to do...

Calibri11A

Code:

- 1. Language: Java-8
- 2. IDE:Netbeans

Covid_DataSet_Stats - NetBeans IDE 8.2

File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help

Search (Ctrl+F)

Projects Files Services

Covid_DataSet_Stats

Source Packages

Covid_DataSet_Stats

Test Packages

<default package>

Libraries

Test Libraries

Navigator

Members

Covid_DataSet_Stats

main(String[] args)

Get_Data_From_CSV

Get_Data_From_CSV(String path)

Read_Data()

display(int range, String by, String pr)

continent : List<String>

date : List<String>

fileName : String

iso_code : List<String>

location : List<String>

median_age : List<String>

new_cases : List<String>

new_cases_smoothed : List<String>

new_deaths : List<String>

new_deaths_smoothed : List<String>

Source History

Covid_DataSet_Stats.java

46

47

//read column-1 using stream for iteration using lambda

48 Stream.iterate(0, start -> start + 1)

49 .limit(values.size())

50 .forEach(CLO -> iso_code.add(values.get(CLO).get(0)));

51

52 Stream.iterate(0, start -> start + 1)

53 .limit(values.size())

54 .forEach(CLO -> continent.add(values.get(CLO).get(1)));

55

56 Stream.iterate(0, start -> start + 1)

57 .limit(values.size())

58 .forEach(CLO -> location.add(values.get(CLO).get(2)));

59

60

61 Stream.iterate(0, start -> start + 1)

62 .limit(values.size())

63 .forEach(CLO -> date.add(values.get(CLO).get(3)));

64

65

66 Stream.iterate(0, start -> start + 1)

67 .limit(values.size())

68 .forEach(CLO -> total_cases.add(values.get(CLO).get(4)));

69

70 Stream.iterate(0, start -> start + 1)

71 .limit(values.size())

72 .forEach(CLO -> new_cases.add(values.get(CLO).get(5)));

73

74

75 Stream.iterate(0, start -> start + 1)

76 .limit(values.size())

77 .forEach(CLO -> new_cases_smoothed.add(values.get(CLO).get(6)));

78

79

80

81 Stream.iterate(0, start -> start + 1)

82 .limit(values.size())

83 .forEach(CLO -> total_deaths.add(values.get(CLO).get(7)));

84

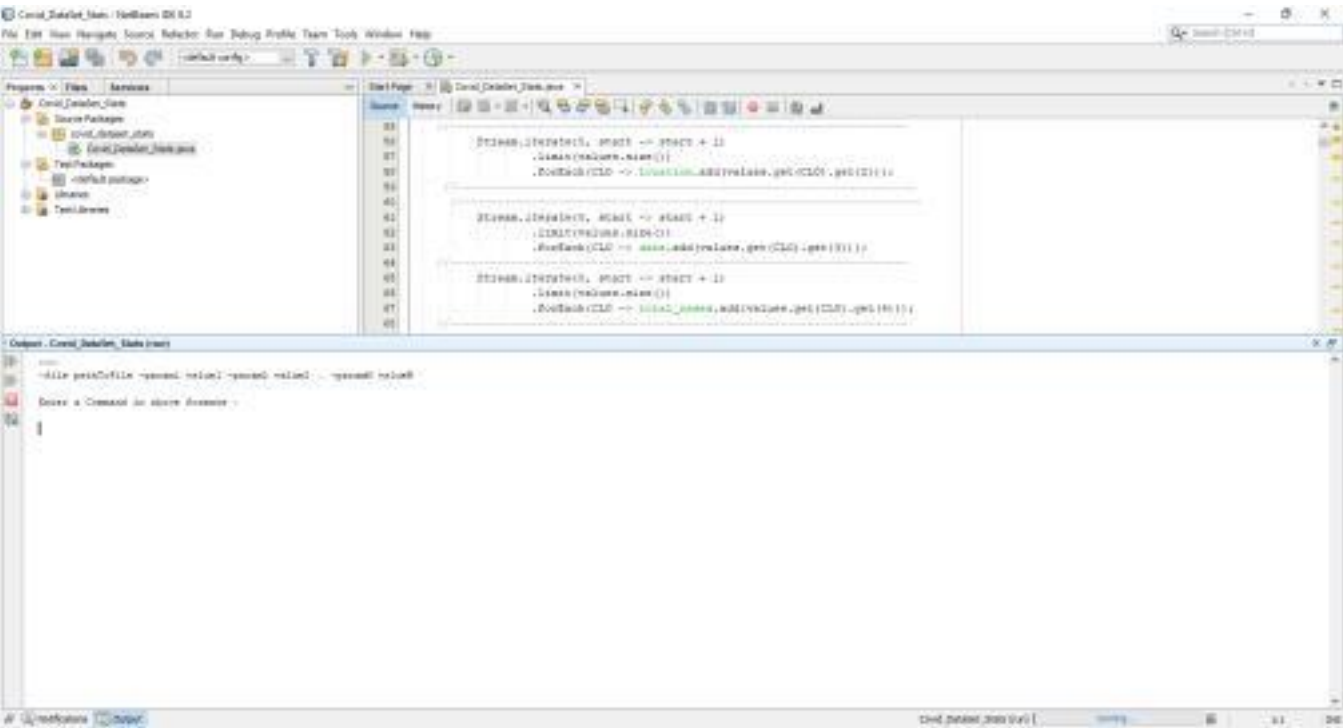
85

1:1

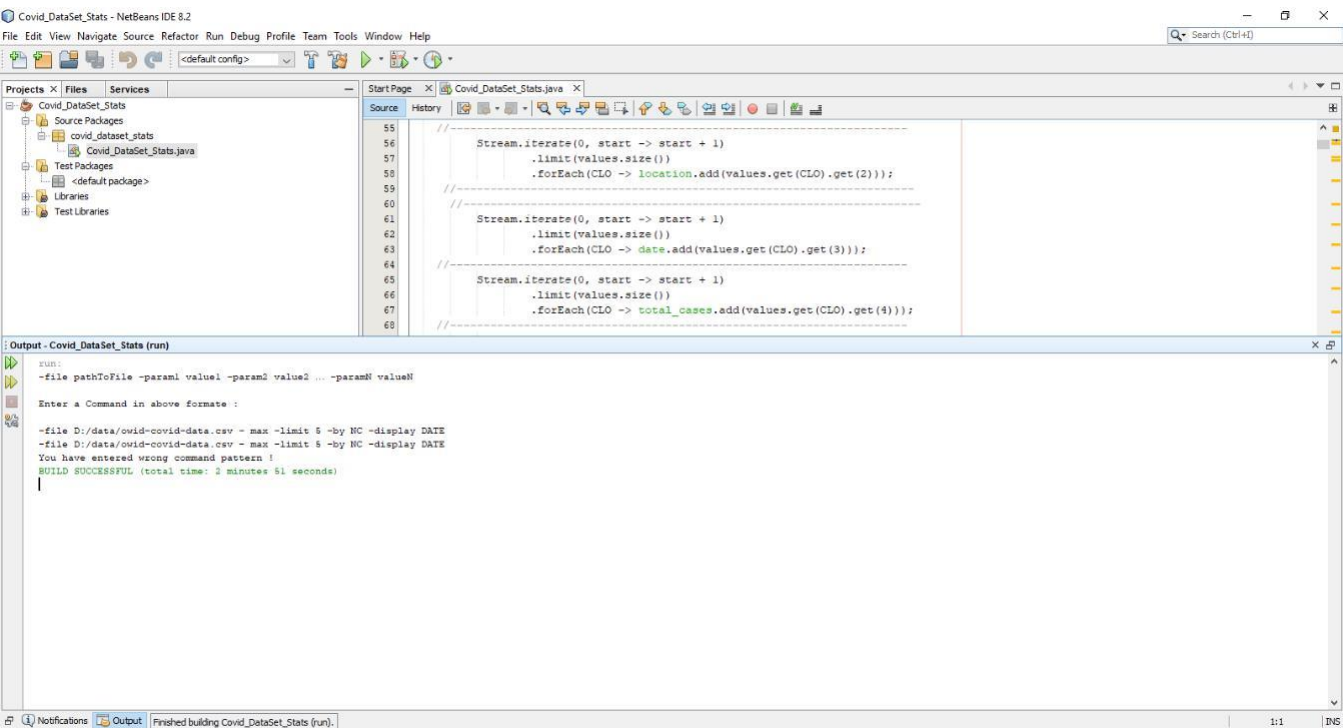
INS

How program work:

1. On Execute:

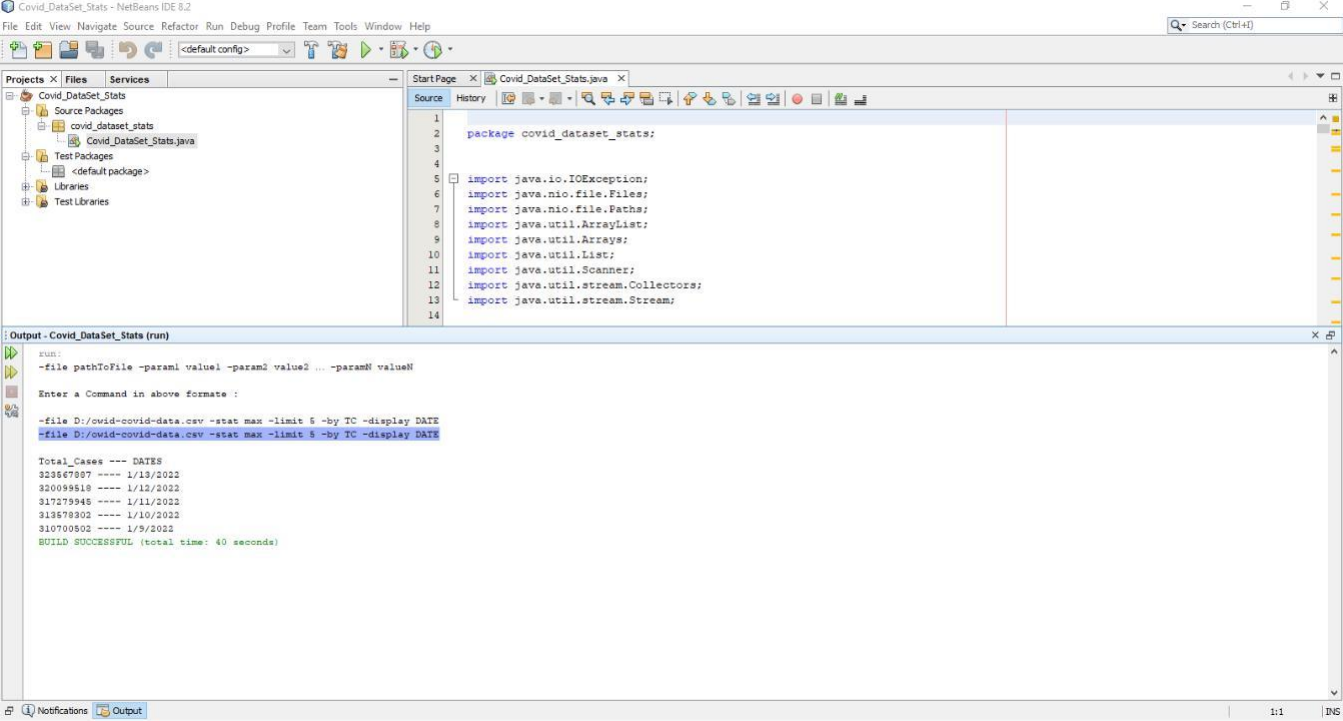


2. IF wrong command is entered :

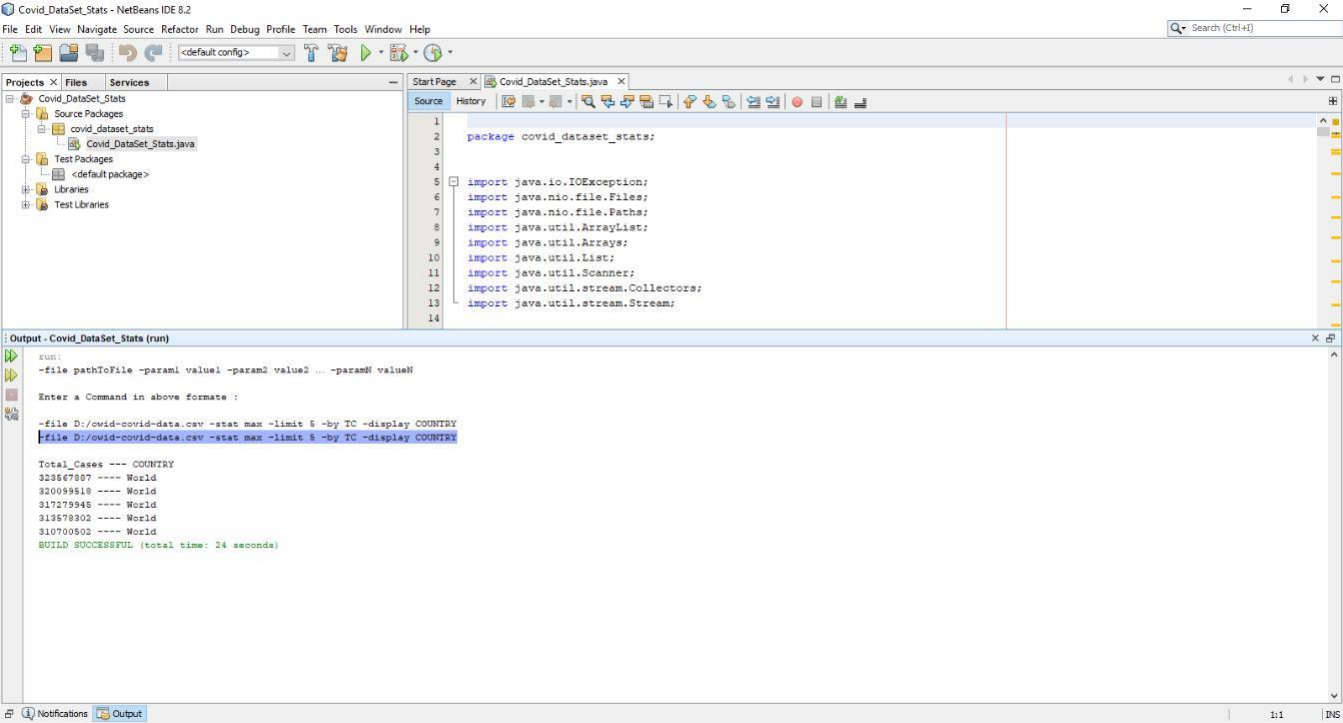


3. Let's execute correct command :

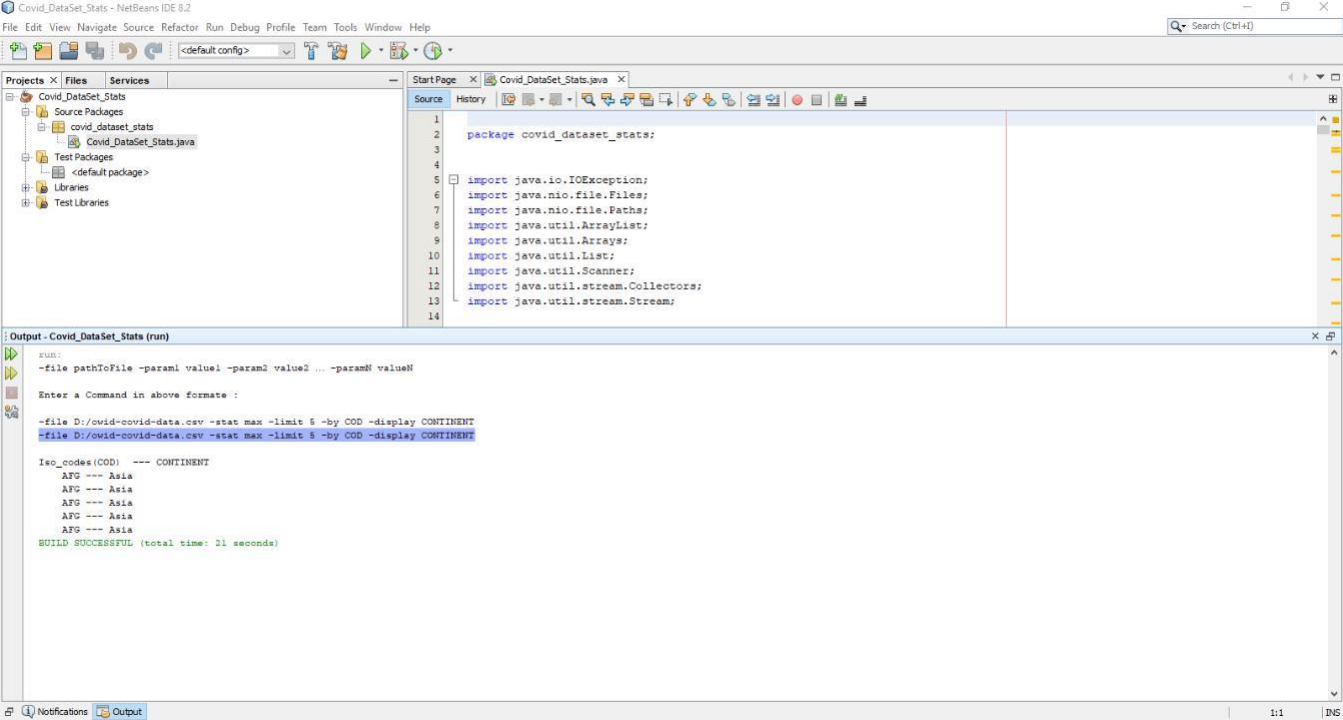
-file D:/owid-covid-data.csv -stat max -limit 5 -by TC -display DATE



-file D:/owid-covid-data.csv -stat max -limit 5 -by TC -display COUNTRY



-file D:/owid-covid-data.csv -stat max -limit 5 -by COD -display CONTINENT



There were some problems with the excel csv file where many data were missing, we had to edit the csv file so our program could run properly. We couldn't fix few commands like this two commands:

-file D:\owid-covid-data.csv -stat max -limit 5 -by POP -display COUNTRY

-file D:\owid-covid-data.csv -stat min -limit 5 -by POP -display DATE

And the error coming out is **Index -1 out of bounds for length 154682**

The github repository url: https://github.com/kostandinvllahu/Advance_Java_Project

