Documentation

Fall 2022, Healthcare Analytics: AMC
By: Bhavya Kashyap, Rida Fathima

*This document provides information about the location of raw-data files, code, and resultant data. The folder locations are with respect to the current folder **FinalDocumentation**. The contents of which are:*

| Name	Status	Date modified	Type	Size
Documentation.pdf	⊘	05-12-2022 23:05	Adobe Acrobat D...	772 KB
AMC_Code.ipynb	⊘	05-12-2022 23:00	IPYNB File	199 KB
chromedriver.exe	⊘	27-09-2022 19:50	Application	11,894 KB
Crosswalk_data	⊘	05-12-2022 23:03	File folder	
PubmedFinalData	⊘	05-12-2022 23:01	File folder	
PubmedRawData	⊘	05-12-2022 22:17	File folder	
CrosswalkRawData	⊘	05-12-2022 21:53	File folder	
.ipynb_checkpoints	⊘	05-12-2022 21:17	File folder	
EDA	⊘	05-12-2022 21:14	File folder	
Journals	⊘	05-12-2022 14:21	File folder	

Note: **AMC_Code.ipynb** is the master code file.

1. *Extracting journals according to SJR rank:*
   Raw Data has been downloaded and saved in folder ***Journals***.

FinalDocumentation > Journals

Name	Status	Date modified	Type	Size
Biochemistry.csv	⊘	02-09-2022 19:32	Microsoft Excel C...	116 KB
HealthPolicy.csv	⊘	02-09-2022 19:34	Microsoft Excel C...	15 KB
journals.csv	⊘	24-11-2022 04:35	Microsoft Excel C...	5 KB
Medicine.csv	⊘	02-09-2022 19:35	Microsoft Excel C...	135 KB
Multidisciplinary.csv	⊘	02-09-2022 19:36	Microsoft Excel C...	6 KB
Nursing.csv	⊘	02-09-2022 19:36	Microsoft Excel C...	7 KB
Pharmacy.csv	⊘	02-09-2022 19:37	Microsoft Excel C...	37 KB

The final data file containing the names of all required journal names has been created and saved in folder ***Journals***.

FinalDocumentation > Journals

Name	Status	Date modified	Type	Size
Biochemistry.csv	⊘	02-09-2022 19:32	Microsoft Excel C...	116 KB
HealthPolicy.csv	⊘	02-09-2022 19:34	Microsoft Excel C...	15 KB
journals.csv	⊘	24-11-2022 04:35	Microsoft Excel C...	5 KB
Medicine.csv	⊘	02-09-2022 19:35	Microsoft Excel C...	135 KB
Multidisciplinary.csv	⊘	02-09-2022 19:36	Microsoft Excel C...	6 KB
Nursing.csv	⊘	02-09-2022 19:36	Microsoft Excel C...	7 KB
Pharmacy.csv	⊘	02-09-2022 19:37	Microsoft Excel C...	37 KB

2. The ***Exploratory Data Analysis*** section of code imports two raw data files which are saved in the ***EDA*** folder as mentioned in the screenshot below.

FinalDocumentation > EDA

Name	Status
georgiadata.csv	⊘
uncdata.csv	⊘

3. The ***Creating crosswalk*** section of the code requires a ***chromedriver.exe*** present which is present in the same folder as the code. Please ensure the exe file stays in the same folder as the code (this is necessary for the code to run).

FinalDocumentation

Name	Status
Documentation.docx	⊘
AMC_Code.ipynb	⊘
chromedriver.exe	⊘

4. The ***Creating crosswalk*** section of the code requires raw data of each state to be imported those raw data files are present in ***CrosswalkRawData*** folder.

FinalDocumentation > CrosswalkRawData   Search CrosswalkRawData

Name	Status	Date modified	Type	Size
Alabama.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	7 KB
Alaska.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	1 KB
Arizona.csv	⊘	24-11-2022 05:16	Microsoft Excel C...	11 KB
Arkansas.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	5 KB
California.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	33 KB
Colorado.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	6 KB
Connecticut.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	4 KB
Delaware.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	2 KB
District_of_Columbia.csv	⊘	15-10-2022 22:17	Microsoft Excel C...	3 KB
Florida.csv	⊘	14-10-2022 18:34	Microsoft Excel C...	32 KB
Hawaii.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	3 KB
Idaho.csv	⊘	11-10-2022 23:51	Microsoft Excel C...	3 KB
Illinois.csv	⊘	11-10-2022 23:56	Microsoft Excel C...	14 KB
Indiana.csv	⊘	11-10-2022 23:43	Microsoft Excel C...	5 KB
Iowa.csv	⊘	11-10-2022 23:47	Microsoft Excel C...	3 KB
Kansas.csv	⊘	12-10-2022 00:01	Microsoft Excel C...	4 KB
Kentucky.csv	⊘	12-10-2022 00:04	Microsoft Excel C...	5 KB
Louisiana.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	8 KB
Maine.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	2 KB
Maryland.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	8 KB
Massachusetts.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	9 KB
Michigan.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	18 KB
Minnesota.csv	⊘	16-10-2022 02:53	Microsoft Excel C...	7 KB

5. The *Creating crosswalk* section of the code creates crosswalk data of each state, result data files are present in *CrosswalkData* folder.

FinalDocumentation > Crosswalk_data   ∨ ↻   🔎 Search Crosswalk_dat

Name	Status	Date modified	Type	Size
Alabama_crosswalk.csv	⊘	17-10-2022 02:40	Microsoft Excel C...	27 KB
Alaska_crosswalk.csv	⊘	14-10-2022 23:44	Microsoft Excel C...	2 KB
Arizona_crosswalk.csv	⊘	14-10-2022 21:46	Microsoft Excel C...	44 KB
Arkansas_crosswalk.csv	⊘	14-10-2022 23:44	Microsoft Excel C...	18 KB
California_crosswalk.csv	⊘	15-10-2022 20:11	Microsoft Excel C...	125 KB
Colorado_crosswalk.csv	⊘	15-10-2022 20:12	Microsoft Excel C...	23 KB
Conneticut_crosswalk.csv	⊘	15-10-2022 20:50	Microsoft Excel C...	16 KB
Delaware_crosswalk.csv	⊘	15-10-2022 22:06	Microsoft Excel C...	8 KB
Doc_crosswalk.csv	⊘	15-10-2022 22:53	Microsoft Excel C...	10 KB
Florida_crosswalk.csv	⊘	16-10-2022 01:29	Microsoft Excel C...	112 KB
Georgia_crosswalk.csv	⊘	11-10-2022 03:00	Microsoft Excel C...	43 KB
Hawaii_crosswalk.csv	⊘	16-10-2022 02:01	Microsoft Excel C...	12 KB
Idaho_crosswalk.csv	⊘	12-10-2022 01:28	Microsoft Excel C...	8 KB
Illinois_crosswalk.csv	⊘	12-10-2022 04:36	Microsoft Excel C...	55 KB
Indiana_crosswalk.csv	⊘	12-10-2022 01:54	Microsoft Excel C...	19 KB
Iowa_crosswalk.csv	⊘	12-10-2022 02:16	Microsoft Excel C...	8 KB
Kansas_crosswalk.csv	⊘	12-10-2022 02:33	Microsoft Excel C...	13 KB
Kentucky_crosswalk.csv	⊘	12-10-2022 02:54	Microsoft Excel C...	18 KB
Louisiana_crosswalk.csv	⊘	16-10-2022 02:32	Microsoft Excel C...	25 KB
Maine_crosswalk.csv	⊘	16-10-2022 02:33	Microsoft Excel C...	6 KB
Maryland_crosswalk.csv	⊘	16-10-2022 03:11	Microsoft Excel C...	29 KB
Massachusetts_crosswalk.csv	⊘	16-10-2022 12:19	Microsoft Excel C...	34 KB

6. *Downloading bibliographical data for articles of pubmed* section of code requires raw data files of universities.csv and journals.csv which are stored in *PubmedRawData* folder.

FinalDocumentation > PubmedRawData

Name

journals.csv
universities.csv

7. ***Downloading bibliographical data for articles of pubmed*** section of code creates result data files for each university, these are stored in ***PubmedFinalData*** folder.

FinalDocumentation > PubmedFinalData

Name
ALBANY MEDICAL COLLEGE_FULLRECORD.CSV
ALBERT EINSTEIN COLLEGE OF MEDICINE_FULLRECORD.CSV
BAYLOR UNIVERSITY_FULLRECORD.CSV
BOSTON UNIVERSITY_FULLRECORD.CSV
BROWN UNIVERSITY_FULLRECORD.CSV
CAMPBELL UNIVERSITY_FULLRECORD.CSV
CASE WESTERN RESERVE UNIVERSITY_FULLRECORD.CSV
CENTRAL MICHIGAN UNIVERSITY_FULLRECORD.CSV
CITY UNIVERSITY OF NEW YORK_FULLRECORD.CSV
COLUMBIA UNIVERSITY_FULLRECORD.CSV
CORNELL UNIVERSITY_FULLRECORD.CSV
CREIGHTON UNIVERSITY_FULLRECORD.CSV
DARTMOUTH COLLEGE_FULLRECORD.CSV
DARTMOUTH_FULLRECORD.CSV
DREXEL UNIVERSITY_FULLRECORD.CSV
DUKE UNIVERSITY_FULLRECORD.CSV
EAST CAROLINA UNIVERSITY_FULLRECORD.CSV
EAST TENNESSEE STATE UNIVERSITY_FULLRECORD.CSV
EASTERN VIRGINIA MEDICAL SCHOOL_FULLRECORD.CSV
EDWARD VIA COLLEGE OF OSTEOPATHIC MEDICINE_FULLRECORD.CSV
EMORY UNIVERSITY_FULLRECORD.CSV

NOTE: For universities for which FullRecord.csv (final csv) is not obtained, the filtered data wasn't available on pubmed.