**Note about the code:**

The code done in simple way too easy to understand not optimized for speed or memory.

**This is the list of columns names before and after cleaning**

|  |  |  |
| --- | --- | --- |
|  | Name of column after cleaning | Name of column before cleaning |
| 1 | "year" | العام |
| 2 | "month" | الشهر |
| 3 | "city" | المدينة |
| 4 | "district", | الأحياء |
| 5 | "price\_riyals", | السعر بالريال السعودي |
| 6 | "area\_m2", | المساحة بالمتر المربع |
| 7 | "no\_deals", | عدد الصفقات |
| 8 | "status", | حالة العملية(معتمدة/ملغية/..) |
| 9 | "realstate\_cls", | تصنيف العقار |
| 10 | "realstate\_type" | نوع العقار |

**This is the list of excel files names output in “xlsx’ folder and its names before and after cleaning**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name of Sheet after cleaning in “xlsx’ | Name of column before cleaning | Notes |
| 1 | all.xlsx | All possible data | All possible information in the source sheet |
| 2 | all\_m.xlsx | All data | No missing values except for "realstate\_cls", "realstate\_type", "district", |
| 3 | Dammam.xlsx | مدينة الدمام | All possible information in the source sheet |
| 4 | Dammam\_m.xlsx | مدينة الدمام | No missing values except for "realstate\_cls", "realstate\_type" |
| 5 | Jeddah.xlsx | مدينة جدة | All possible information in the source sheet |
| 6 | Jeddah\_m.xlsx | مدينة جدة | No missing values except for "realstate\_cls", "realstate\_type" |
| 7 | Makkah.xlsx | مدينة مكة المكرمة | All possible information in the source sheet |
| 8 | Makkah\_m.xlsx | مدينة مكة المكرمة | No missing values except for "realstate\_cls", "realstate\_type" |
| 9 | Medina.xlsx | المدينة المنورة | All possible information in the source sheet |
| 10 | Medina\_m.xlsx | المدينة المنورة | No missing values except for "district", |
| 11 | Riyadh.xlsx | مدينة الرياض | All possible information in the source sheet |
| 11 | Riyadh\_m.xlsx | مدينة الرياض | No missing values except for "realstate\_cls", "realstate\_type" |

**This is the list of JSON files names output in “json’ folder and its names before and after cleaning**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name of Sheet after cleaning in “xlsx’ | Name of column before cleaning | Notes |
| 1 | all.json | All possible data | All possible information in the source sheet |
| 2 | all\_m. json | All data  Ready to be import in firebase project without any error | No missing values except for "realstate\_cls", "realstate\_type", "district", |
| 3 | Dammam. json | مدينة الدمام | All possible information in the source sheet |
| 4 | Dammam\_m. json | مدينة الدمام | No missing values except for "realstate\_cls", "realstate\_type" |
| 5 | Jeddah. json | مدينة جدة | All possible information in the source sheet |
| 6 | Jeddah\_m. json | مدينة جدة | No missing values except for "realstate\_cls", "realstate\_type" |
| 7 | Makkah. json | مدينة مكة المكرمة | All possible information in the source sheet |
| 8 | Makkah\_m. json | مدينة مكة المكرمة | No missing values except for "realstate\_cls", "realstate\_type" |
| 9 | Medina. json | المدينة المنورة | All possible information in the source sheet |
| 10 | Medina\_m. json | المدينة المنورة | No missing values except for "district", |
| 11 | Riyadh. json | مدينة الرياض | All possible information in the source sheet |
| 11 | Riyadh\_m. json | مدينة الرياض | No missing values except for "realstate\_cls", "realstate\_type" |

**YouTube references:**

1- Import firebase database data from excel or csv file in node child tree type structure

<https://www.youtube.com/watch?v=RO0LMwpoGPg>

2- Python Pandas Tutorial (Part 9): Cleaning Data - Casting Datatypes and Handling Missing Values

<https://www.youtube.com/watch?v=KdmPHEnPJPs>

3- Python Pandas Tutorial (Part 11): Reading/Writing Data to Different Sources - Excel, JSON, SQL, Etc

<https://www.youtube.com/watch?v=N6hyN6BW6ao&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS&index=11>

4- Python Pandas Tutorial (Part 6): Add/Remove Rows and Columns From DataFrames

<https://www.youtube.com/watch?v=HQ6XO9eT-fc&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS&index=6>

5- Python Pandas Tutorial (Part 5): Updating Rows and Columns - Modifying Data Within DataFrames

<https://www.youtube.com/watch?v=DCDe29sIKcE&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS&index=5>

6- Python Pandas Tutorial (Part 3): Indexes - How to Set, Reset, and Use Indexes

<https://www.youtube.com/watch?v=W9XjRYFkkyw&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS&index=3>

**Project URL references:**

Database link

<https://data.gov.sa/Data/ar/dataset/real-estate-detail-for-district-in-2018-2019-2020-and-2021>

read excel files

<https://pandas.pydata.org/docs/reference/api/pandas.read_excel.html>

Insert column into DataFrame at specified location.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.insert.html>

Indexing and Selecting Data with Pandas

<https://www.geeksforgeeks.org/indexing-and-selecting-data-with-pandas/>

<https://www.sharpsightlabs.com/blog/pandas-index/>

<https://stackoverflow.com/questions/11346283/renaming-column-names-in-pandas>

Rename columns in pandas

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rename.html>

<https://www.geeksforgeeks.org/rename-column-by-index-in-pandas/>

<https://stackoverflow.com/questions/11346283/renaming-column-names-in-pandas>

How to change the order of DataFrame columns?

<https://stackoverflow.com/questions/13148429/how-to-change-the-order-of-dataframe-columns>

Add column to dataframe with constant value

<https://stackoverflow.com/questions/29517072/add-column-to-dataframe-with-constant-value>

Selecting multiple columns in a Pandas dataframe

<https://stackoverflow.com/questions/11285613/selecting-multiple-columns-in-a-pandas-dataframe>

Drop row by index

[https://thispointer.com/python-pandas-how-to-drop-rows-in-dataframe-by-index-labels/#:~:text=Delete%20a%20Multiple%20Rows%20by%20Index%20Position%20in%20DataFrame&text=drop()%20function%20accepts%20only,pass%20it%20to%20drop().&text=rows%20a%20%26%20b-,As%20default%20value%20of%20inPlace%20is%20false%2C%20so%20contents,dfObj%20will%20not%20be%20modified](https://thispointer.com/python-pandas-how-to-drop-rows-in-dataframe-by-index-labels/#:~:text=Delete%20a%20Multiple%20Rows%20by%20Index%20Position%20in%20DataFrame&text=drop()%20function%20accepts%20only,pass%20it%20to%20drop().&text=rows%20a%20%26%20b-,As%20default%20value%20of%20inPlace%20is%20false%2C%20so%20contents,dfObj%20will%20not%).

How to get column names in Pandas dataframe

<https://www.geeksforgeeks.org/how-to-get-column-names-in-pandas-dataframe/>

Saving json file:

<https://appdividend.com/2022/01/18/python-dict-to-json/#:~:text=To%20convert%20Python%20dict%20to,dictionary%20to%20a%20string%20object>.

<https://docs.python.org/3/library/json.html>

<https://realpython.com/python-json/>

<https://stackoverflow.com/questions/55529775/uploading-json-to-firebase-from-dataframe>

<https://datatofish.com/export-pandas-dataframe-json/>

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_json.html>

Saving csv contain Arabic characters

<https://stackoverflow.com/questions/43321243/save-pandas-dataframe-containing-chinese-character-to-file>

Validate json

<https://jsonlint.com/>

<https://jsonformatter.curiousconcept.com/>

<https://codebeautify.org/csv-to-xml-json>

Using conditions in python

<https://www.w3schools.com/python/python_conditions.asp>

Using pandas Data Frame

Generate descriptive statistics.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>

Pandas delete rows based on multiple column values

<https://stackoverflow.com/questions/29017525/deleting-rows-based-on-multiple-conditions-python-pandas>

Column of strings to floats in Python

<https://stackoverflow.com/questions/16729483/converting-strings-to-floats-in-a-dataframe/33687073>

<https://datatofish.com/convert-string-to-float-dataframe/>

<https://stackoverflow.com/questions/43956335/convert-float64-column-to-int64-in-pandas>

Select data frame row by index

<https://www.geeksforgeeks.org/select-rows-columns-by-name-or-index-in-pandas-dataframe-using-loc-iloc/>

<https://stackoverflow.com/questions/28256761/select-pandas-rows-by-excluding-index-number>

<https://stackoverflow.com/questions/19155718/select-pandas-rows-based-on-list-index>

row loop operation in data frame

<https://stackoverflow.com/questions/19790790/splitting-dataframe-into-multiple-dataframes>

<https://www.w3schools.com/python/ref_func_range.asp>

Remove missing values.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.dropna.html>

How to add sheet to existing excel file with pandas?

<https://stackoverflow.com/questions/48280260/how-to-add-sheet-to-existing-excel-file-with-pandas>

<https://soulsinporto.medium.com/how-to-add-new-worksheets-to-excel-workbooks-with-pandas-47122704fb75>

<https://thinkingneuron.com/add-new-sheet-to-excel-using-pandas/>

<https://stackoverflow.com/questions/42370977/how-to-save-a-new-sheet-in-an-existing-excel-file-using-pandas>

Saving excel files

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_excel.html>

Append rows of other to the end of caller, returning a new object.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.append.html>

<https://stackoverflow.com/questions/49916371/how-to-append-new-row-to-dataframe-in-pandas>

Zipping file online

<https://www.ezyzip.com/zip-folder-online.html>

How do I store Excel data into Firebase?

<https://stackoverflow.com/questions/40467777/how-do-i-store-excel-data-into-firebase#:~:text=For%20example%2C%20on%20Excel%202013,%3E%20Change%20File%20Type%20%3E%20CSV.&text=Finally%2C%20take%20that%20.,Voila>!

***Project descriptions:***

I need a data analyst to clean my spreadsheet of data which contains A list of cities with information about it  
  
\*\* mandatory part:   
  
1-it is your job to provide youtube tutorials about the method you used for data cleaning and any tutorial you come across that helped you\*\*  
At least 3 to 4 directly related tutorials and +  
  
2-You are also required to screen record the code and explain each function written and the reason behind it.

The requirement:

1- You want related tutorials in English

2- Screenshot with explaining the code and the reason behind it.

3- output format json and Xls file

**Steps to upload country list in firebase:**

1. Download Excel spreadsheet and clean it up in such a way that the first row contains the headers and everything else is the actual data and convert all Arabic language to English and UTF-8 encoded which means that it cannot contain. $ # [ ] / or ASCII control characters 0-31 or 127.
2. Convert it into csv or convert it directly to convert csv into json
3. Save in a file say test.json.
4. Open firebase with login.
5. Select database.
6. click on (...) in verticle, beside + - sign,
7. Click on import.
8. browse and select test.json file

**Concept of cleaning dataset**

1 - read excel file, sheet number   
2- explore sheet  
3- save some information before removing unnecessary rows  
4- extract every year information and

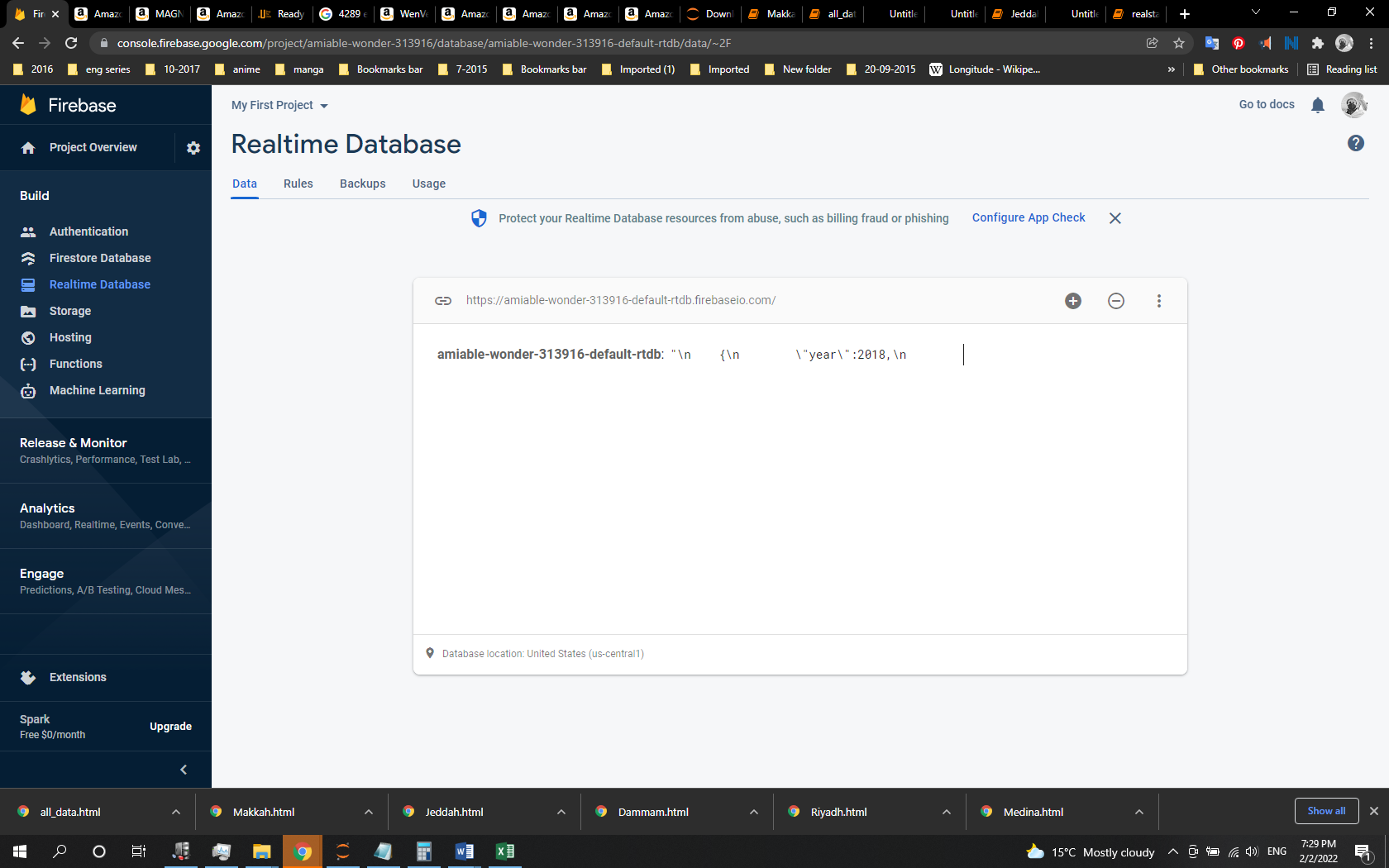
5- Rename columns and remove unnecessary rows  
5- repeat this steps for every year  
6- join all years together  
7- reorganize columns and modify data types  
8- remove Nan values  
9- save database in a different format

10- Repeat all the above for every city then

11 – Join all data together and remove missing values and save data as all\_m.json file

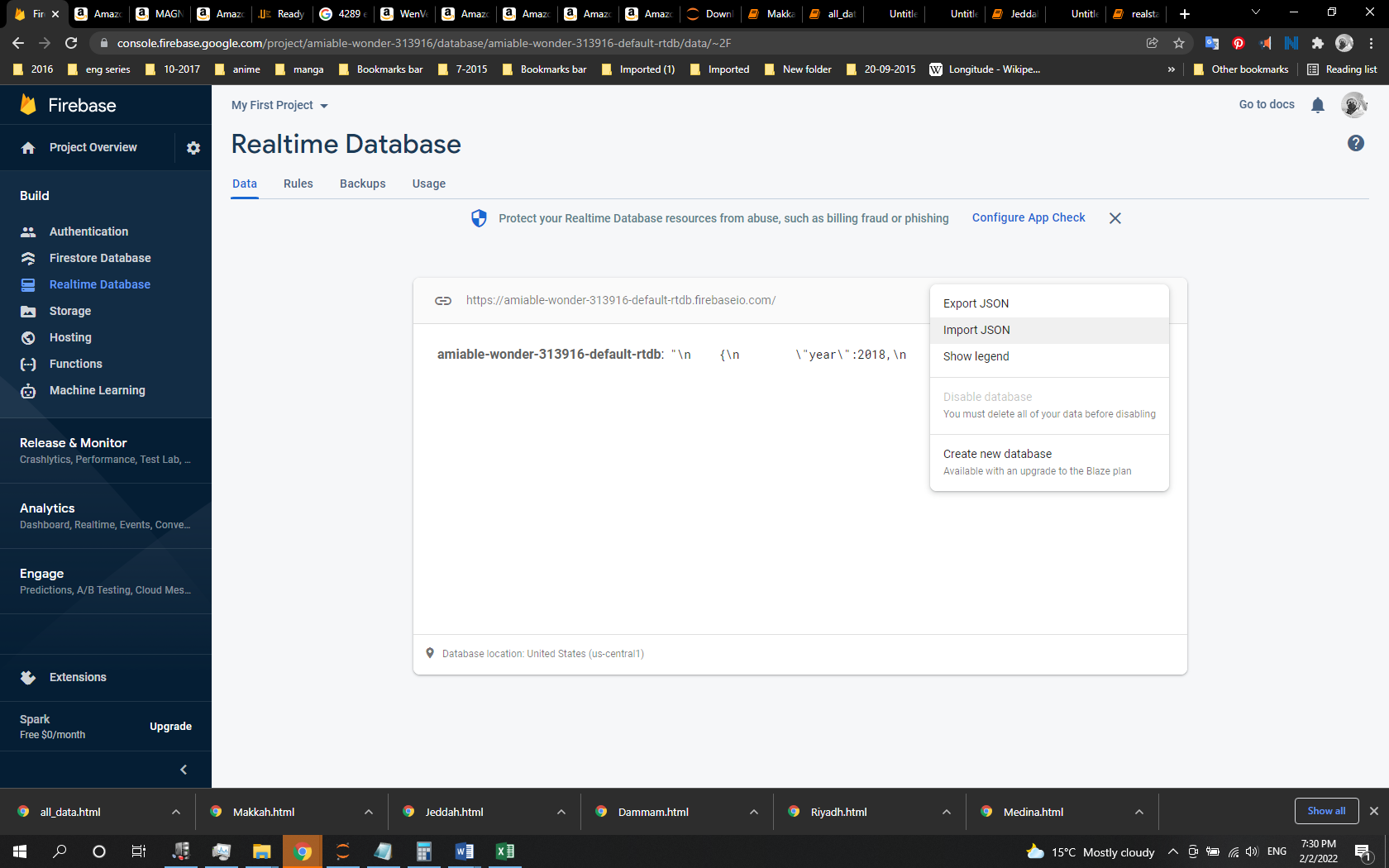
12- Open firebase with login.

13 – in Realtime Database tab in firebase project

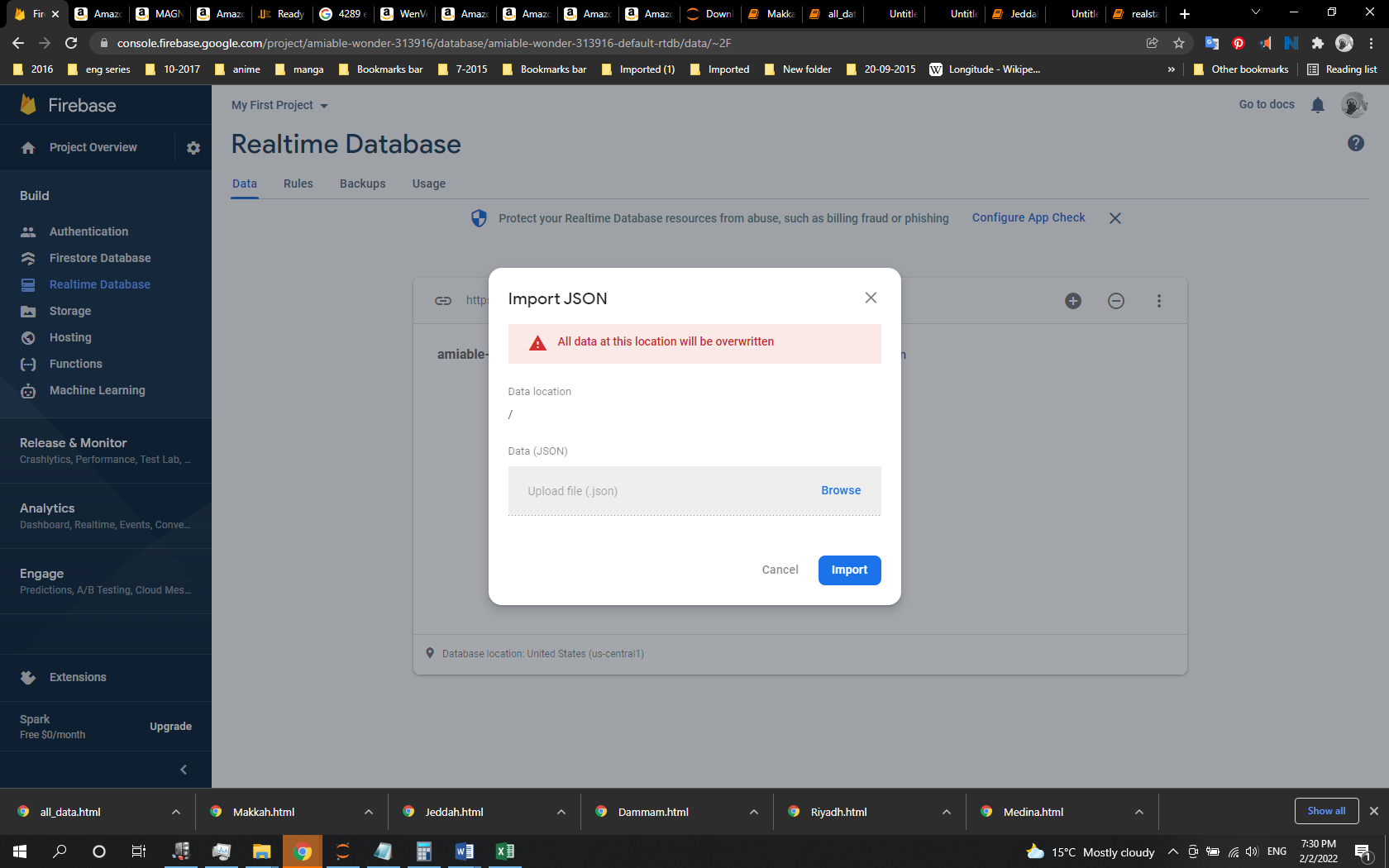


14- Click on (...) in verticle, beside + - sign,

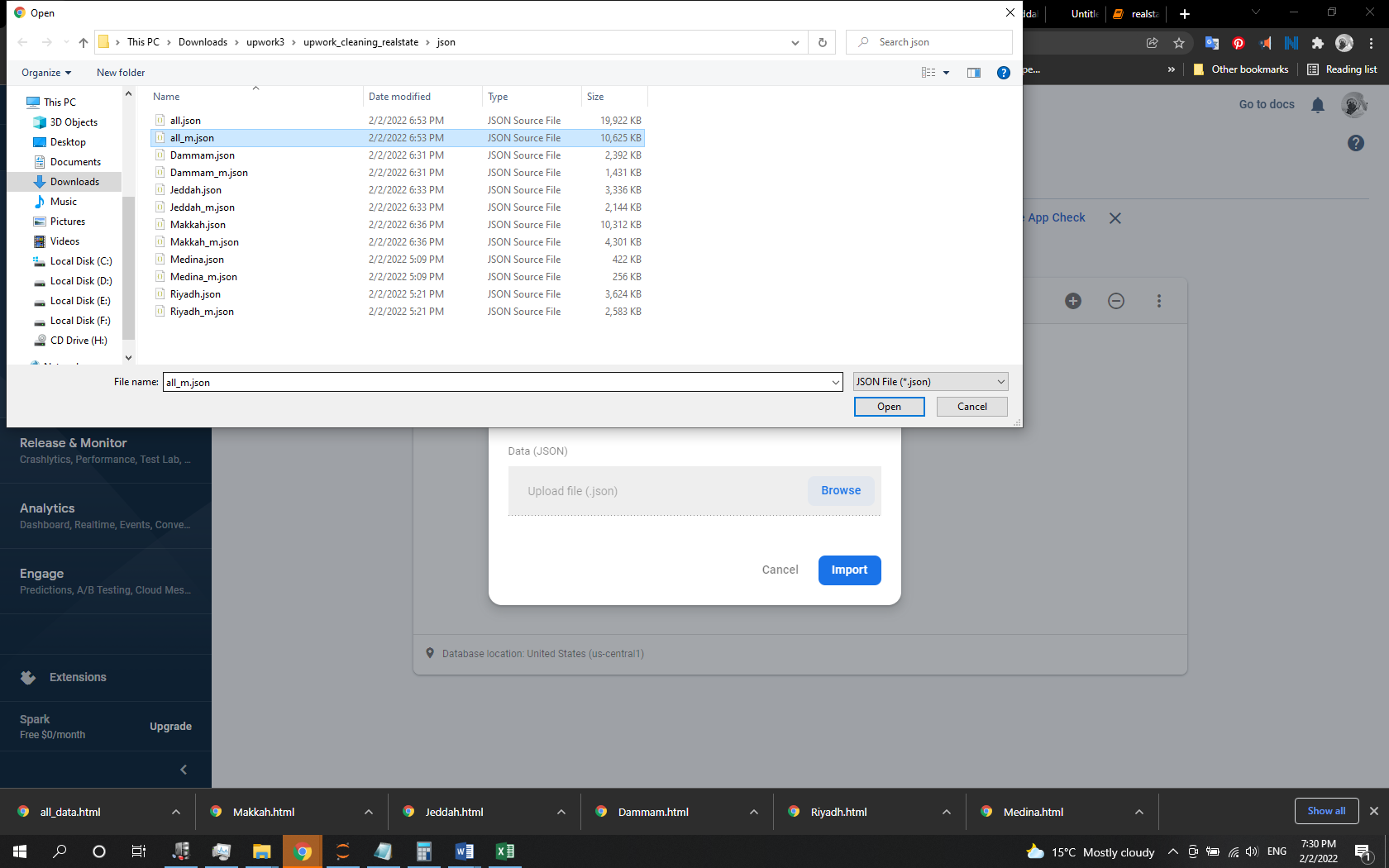
15- Click on import JSON.



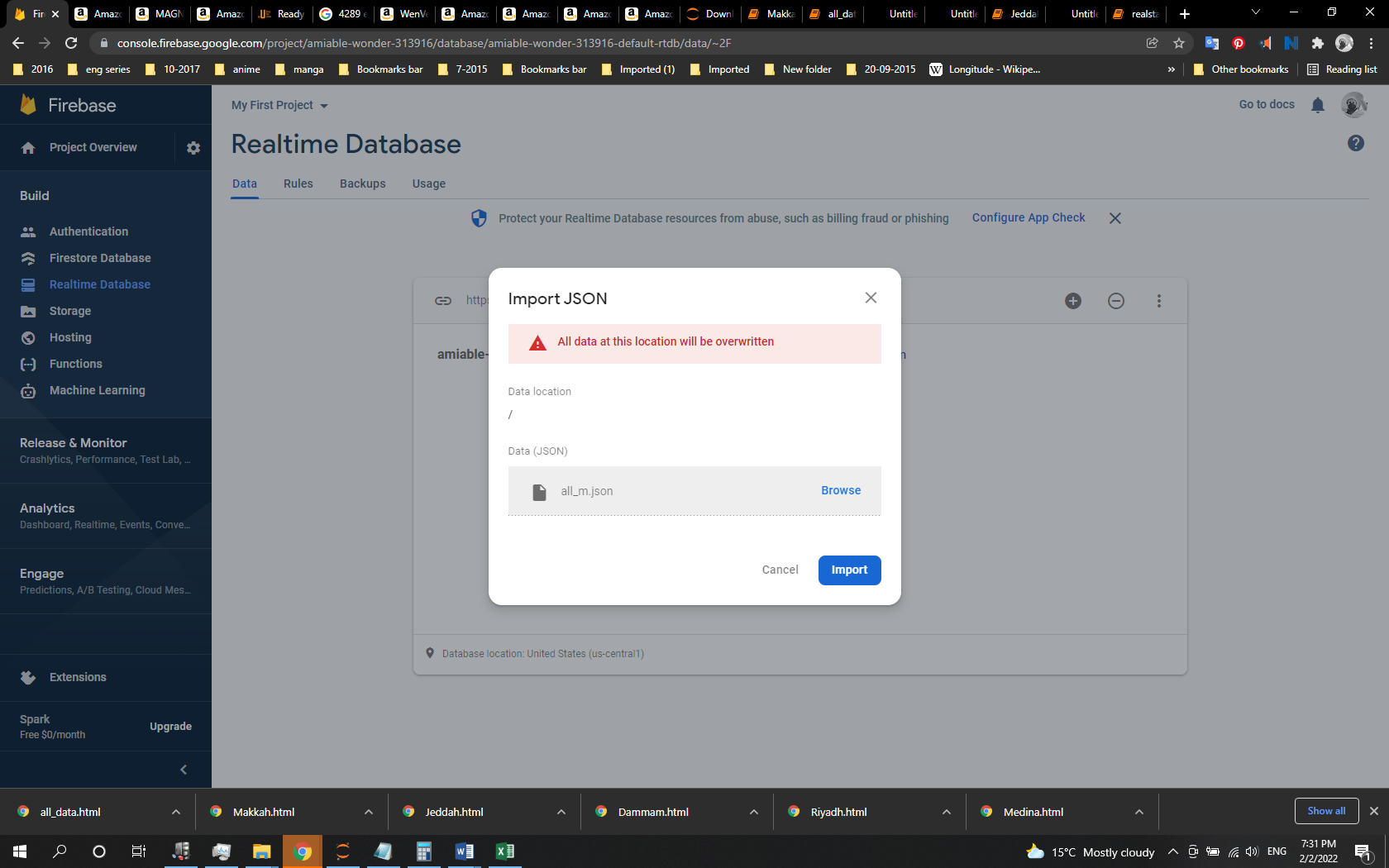
16- Click on import.

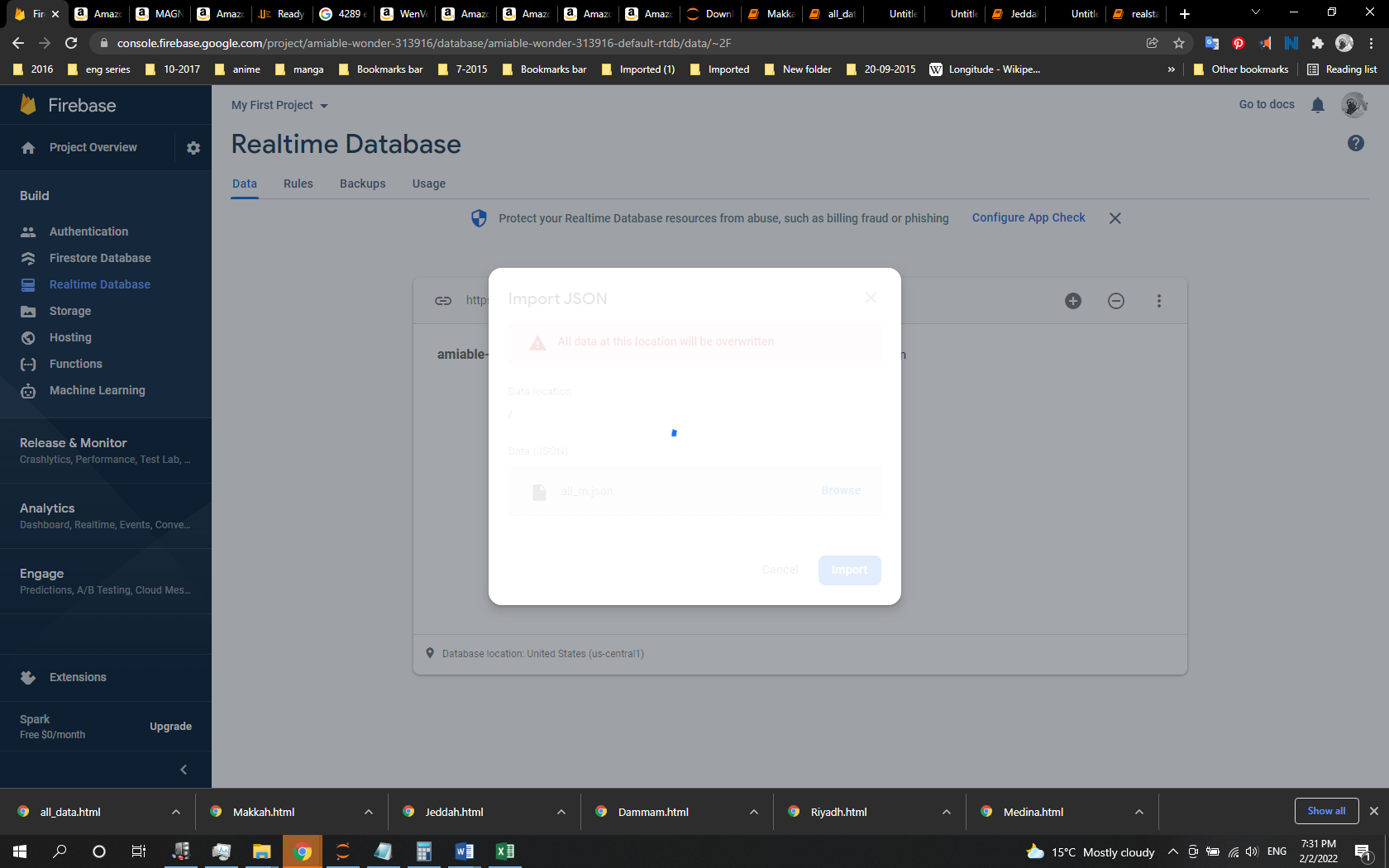


17- Browse xlsx folder and select all\_m.xlsx.



18- Click on import and wait…





19- until notification says data imported successfully.

