**Laurent Meeting:**

* Files are in Databricks/Workspace/Shared
* Go to Shared/matching\_code/Matching\_foreign\_partners\_Mexico
* Create a matching\_foreign\_partners\_India, create a notebook for each country.
* So the only thing that I need to change is the first 7 lines

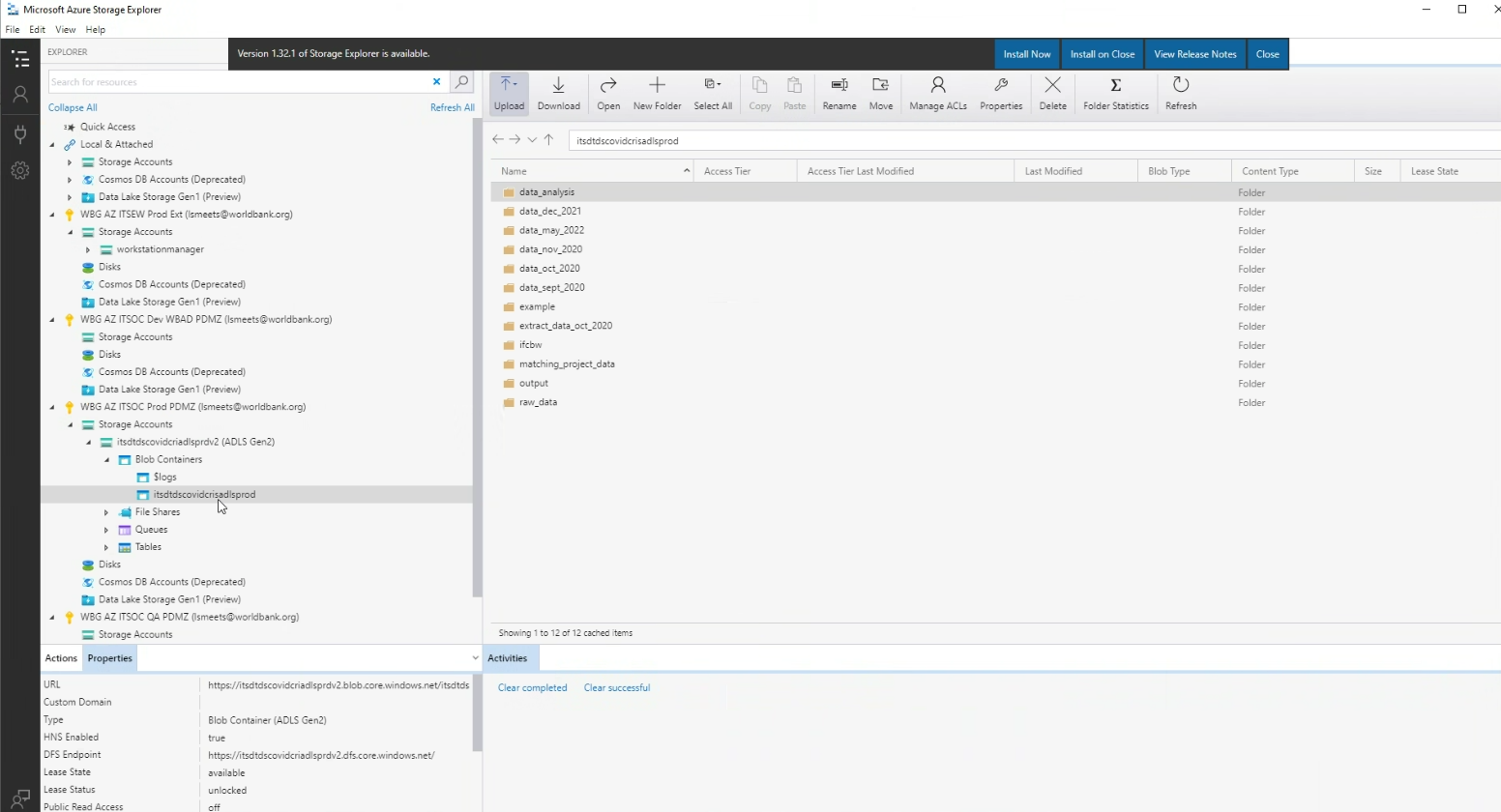
****

The only part that would change is line 4 and line 5, instead of MEX it will become IND and 5 instead of MEX\_foreign\_ you will add IND\_foreign\_ . Aberdeen data does not need to be updated.

TAKE THE SAMPLE CODE FROM match\_code\_Mexico\_partners\_SA (starting with S because in the other codes the library collapse changed its syntax so it will not work)

* The only way to get the data out of the databricks environment (once is ready) is in the Microsoft Azure Storage Explorer. You will not see the final data in the databricks folder, you will see the data in the Microsoft Azure Storage Explorer.

The data in the azure explorer is located here:

Then you go to matching\_project\_data/output/MEX\_foreign\_partners/ that is located the data that the matching codes produce 

Ask Ana if the names are changing?

Create a correspondence between old IDS and new IDS.

List of questions for Laurent:

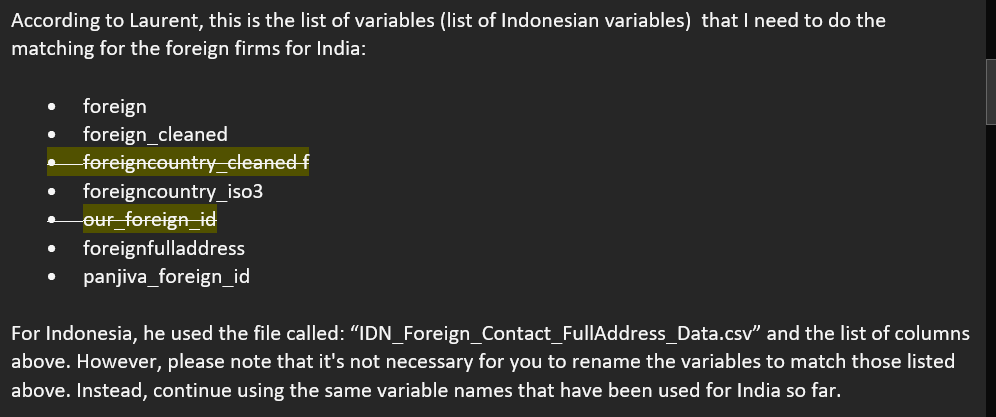
1. What’s the specific objective of the matching that we are doing here? Get for the foreign firms of India the information in Panjiva and Builtwith?
2. Why the country codes that you use are different for Aberdeen and the current\_country\_code string?

A: Use Mexico ones

If you want to overwrite something a file you need to delete that (the folder) first in the Microsoft azure explorer.

1. From where do you take the list of countries that you need to match? There is no list
2. How were you able to start the foreign firms matching for India if you did not have the file with the following variables for India:

* Based on our last meeting, we needed the following variables for India (not necessarily the same names but this is based on names used for the matching of foreign firms of Indonesia):



* You used the file IDN\_Foreign\_Contact\_FullAddress\_Data.csv (provided by Ana) to match the foreign firms for Indonesia.
* Ana has already provided me these file for the foreign firms of India, but do not know exactly at what moment this file plays a role, even more knowing that you already started the matching without the need of that file.
* **Answer:** Request an updated file of the file IND\_foreign\_names\_for\_laurent.csv in C:\Users\wb602786\OneDrive - WBG\Trade and Tech Project - Laurent Files\India\raw\_data\Panjiva. If you are to update a file it should be a file of trade outcomes, but not others.
* **Meeting with Ana:** She said that it is possible to create a correspondence between these new IDs and the old IDs using the raw firm name. The point is that this foreign id is useless because it changed in the new versions and everything in Ana’s work is now based on the raw firm names, no more on the IDs. She asked me to join the files IND\_foreign\_names\_for\_laurent.csv and the new file that Alejandro sent me using the raw firm names in principle (or the raw firm name and the isocode if the previous thing does not work). What she expects is that the difference between the two files is that the newest file has more rows (more firms because they cover 2022 now) and the previous file did not cover that year. So if the proportion of firms that did not match is high probably we will need to do the long process that Laurent said he did before the matching. Otherwise, we can just ignore these new rows and stick to the old file but creating a correspondence between the old IDs and the new IDs (that are not ready yet)

1. Ask 5. If this is not clear. What’s the difference between files of Panjiva provided by Ana and Panjiva\_data\_clean\_wide\_IND\_partners.parquet? Show an example: IDN\_Exports\_Monthly\_foreign\_domestic\_level\_dropdup.dta

* Two files, one if od trade outcomes and other is about other variables scraped.

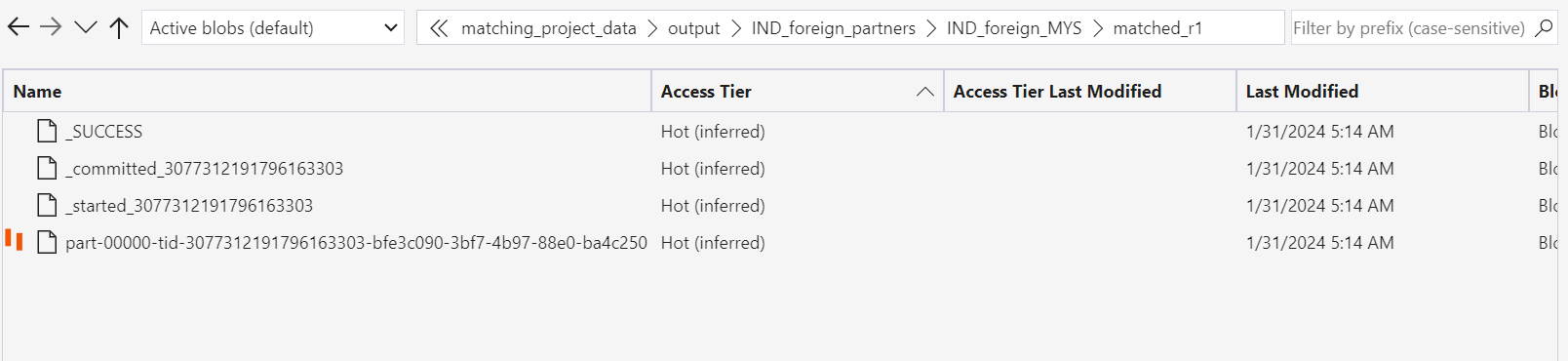
1. I understand that the Aberdeen data does not need to be modified. But is it the same for the Panjiva data given that Ana is giving us new versions of this data? Does not need the file Panjiva\_data\_clean\_wide\_IND\_partners.parquet to be updated based on these new versions?

* Please no

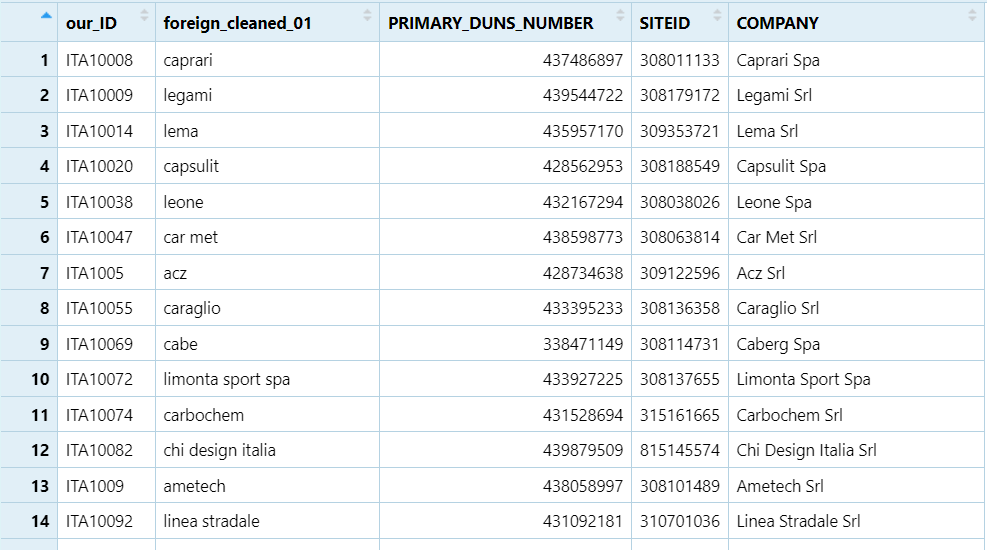
1. In which part do you match the data to Builtwith?

* Matching with Builtwith is done, but he was not able to find it. He does not see the file where he did it.
* shared
* “match vuiltwith Mexico.R” change to IND and India all parts where Mexico appears and this will run the builtiwtih matching for india.
* Will check if the matching is done.

1. Once the code runs (I have not run any code though), what I understand is that it produces the following files in the Microsoft Azure Storage Explorer



Red is the one with the following data:



But is this the matched data that we need? I was expecting something more like the file matched\_foreign\_firms\_indonesia\_v2.csv

1. Finally, I see that you use different rounds to match the data, you matched it using by blocks. In which part do you join these blocks? (matched\_r1, matched\_r2, matched\_r3, etc.)

* Use “load and combine Panjiva to Aberdeen matches from server.R” to join the files. This is located in Tech and Trade C:\Users\wb602786\OneDrive - WBG\Trade and Tech Project - Laurent Files\Mexico.
* Download the folder not each file.

Error in China’s notebook

RserveException: eval failed



**Instructions to RUN matching to Builtwith for example for mexico**

1. Use code “match Builtwith Mexico” in Databricks
2. Use “combine matched builtwith data with Panjiva data.R”