

Tutorial: Merging two tables in Open Refine

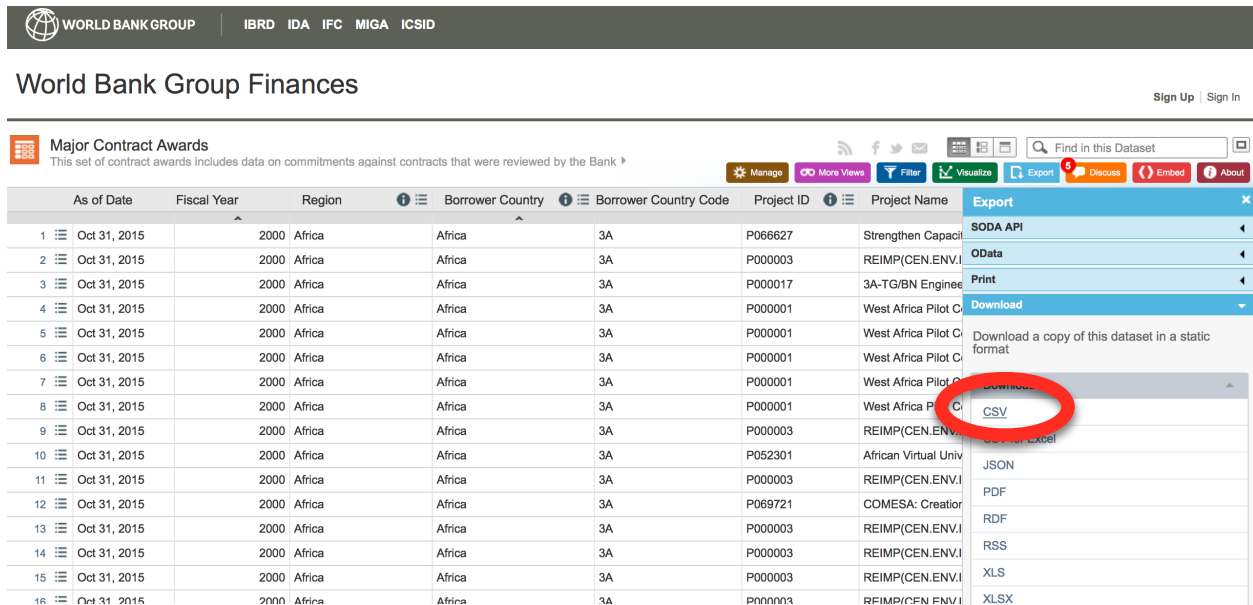
- The data:

Go to Major Contract Awards:

<https://finances.worldbank.org/Procurement/Major-Contract-Awards/kdui-wcs3>

Filter by Region: Latin America

Go to Export and choose CSV to export all the data in csv format.



The screenshot shows the World Bank Group Finances website. The page title is "World Bank Group Finances" and the sub-page is "Major Contract Awards". The table below shows a list of contract awards with columns for As of Date, Fiscal Year, Region, Borrower Country, Borrower Country Code, Project ID, and Project Name. The "Export" button is visible, and the export menu is open, showing options like SODA API, OData, Print, Download, CSV, JSON, PDF, RDF, RSS, XLS, and XLSX. The "CSV" option is circled in red.

	As of Date	Fiscal Year	Region	Borrower Country	Borrower Country Code	Project ID	Project Name
1	Oct 31, 2015	2000	Africa	Africa	3A	P066627	Strengthen Capacit
2	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
3	Oct 31, 2015	2000	Africa	Africa	3A	P000017	3A-TG/BN Enginee
4	Oct 31, 2015	2000	Africa	Africa	3A	P000001	West Africa Pilot C
5	Oct 31, 2015	2000	Africa	Africa	3A	P000001	West Africa Pilot C
6	Oct 31, 2015	2000	Africa	Africa	3A	P000001	West Africa Pilot C
7	Oct 31, 2015	2000	Africa	Africa	3A	P000001	West Africa Pilot C
8	Oct 31, 2015	2000	Africa	Africa	3A	P000001	West Africa P
9	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
10	Oct 31, 2015	2000	Africa	Africa	3A	P052301	African Virtual Univ
11	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
12	Oct 31, 2015	2000	Africa	Africa	3A	P069721	COMESA: Creator
13	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
14	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
15	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI
16	Oct 31, 2015	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENVI

b. Data #2:



Merging two tables in Open Refine

Giannina Segnini
gs2799@columbia.edu

USE THE CLEAN FILE FROM
<http://ieg.worldbankgroup.org/data>

- 2- Open the two files in separate browser tabs as shown below:
- Open OpenRefine from the browser by typing <http://127.0.0.1:3333>
- Select file: Major_Contract_Awards.csv
- Click Next
- CLICK Create Project.

The screenshot shows the OpenRefine interface with a data table containing 5 rows. The table has the following columns: All, As of Date, Fiscal Year, Region, Borrower Count, Borrower Count, Project ID, Project Name, Procurement Type, and Procurement Category. The data rows are as follows:

All	As of Date	Fiscal Year	Region	Borrower Count	Borrower Count	Project ID	Project Name	Procurement Type	Procurement Category
1	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P066627	Strengthen Capacity of ACBF to Manage & Operate the PACT	Implementation Activity	Consultant Services
2	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000003	REIMP(CEN.ENV.INFO)	Implementation Activity	Consultant Services
3	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000017	3A-TG/BN Engineering TAL (FY92)	Equipment Information Technology	Goods
4	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000001	West Africa Pilot Community-based Natural Resources & Wildlife	Feasibility studies, Design Studies	Consultant Services
5	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000001	West Africa Pilot Community-based Natural Resources &	Feasibility studies, Design Studies	Consultant Services

Open another tab

(type <http://127.0.0.1:3333> in the new tab to open OpenRefine)

Open file: IEG_World_Bank_Project_Performance_Ratings_asof_Jan182017_download.xlsx

- 3- Go to the file “Major_Contract_Awards.csv” and find column Project ID
- Click Edit Columns
- “Add Column based on this column”

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

All	As of Date	Fiscal Year	Region	Borrower Count	Borrower Count	Project ID	Project Name	Procurement Type	Procurement C
1.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A		on Capacity of Manage & the PACT	Implementation Activity	Consultant Services
2.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A				Services
3.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A				
4.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A				Services
5.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000001	West Af Commu Natural Wildli		Services
6.	10/31/2015 12:00:00 AM	2000	Africa	Africa	3A	P000001	West Af Community-based Natural Resources & Wildli	design solutions	Services

a- Name the new column as OUTCOME

b- Go to the empty box and replace the word "value" with the following script

**if (value!
='null',cell.cross("IEG_World_Bank
_Project_Performance_Ratings_asof_J
an182017_download_xlsx", "Project
ID").cells["IEG_Outcome"].value
[0],")**



Merging two tables
in Open Refine

Giannina Segnini
gs2799@columbia.edu