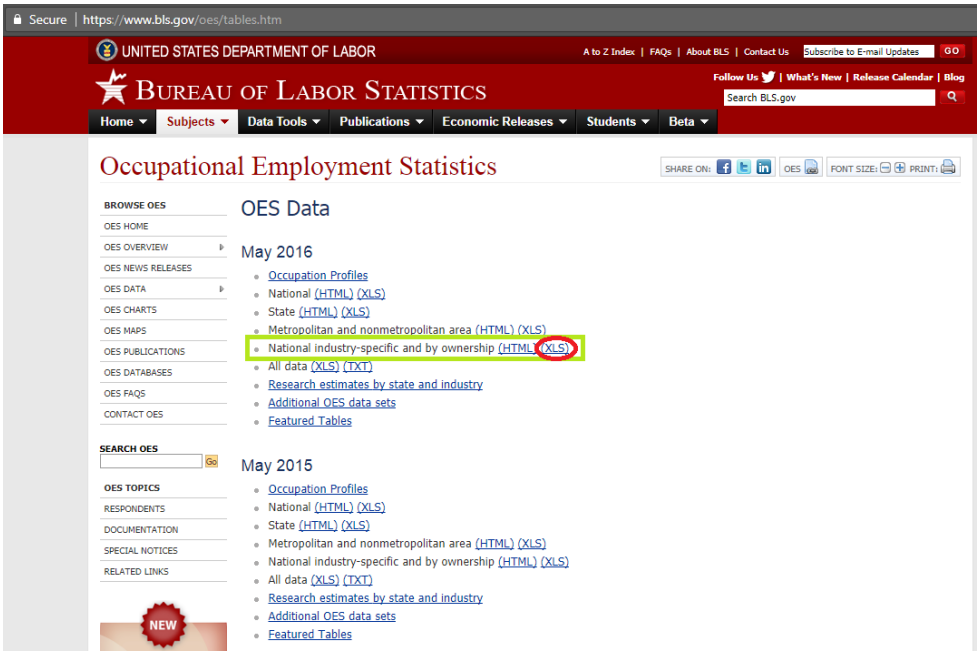
Workforce Data Quality Initiative | Indiana

Licensing Occupation Assignment Instructions

1. Download the Licensing Code package in the language of your choice
    a. Python
    b. R
    c. SAS
    d. SQL
    e. Stata

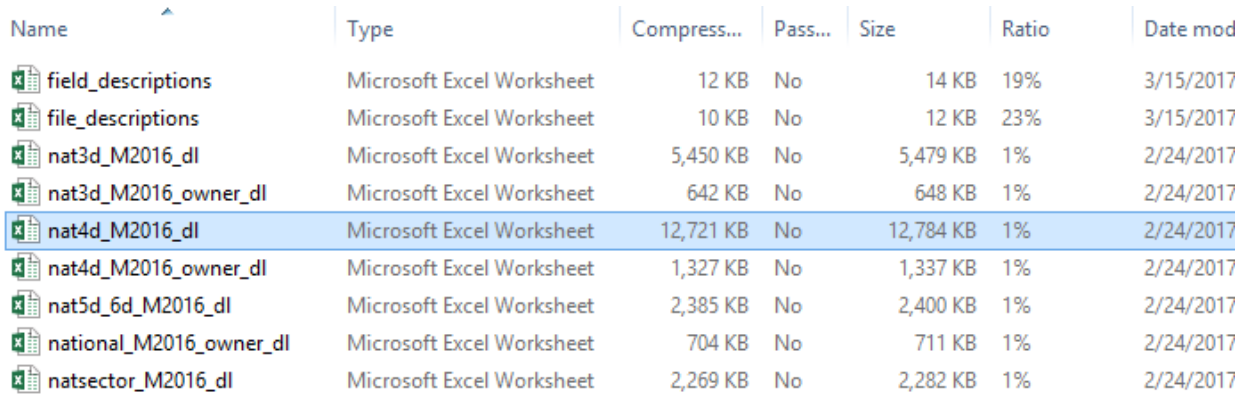
2. Extract the code package files to a folder of your choosing

3. Download and compile the accompanying OES industry-specific staffing patterns file for each year of interest
    a. Visit https://www.bls.gov/oes/tables.htm
    b. Select the Excel version of the National industry-specific and by ownership file for the year of interest.



c. Unzip the downloaded file(s)
d. Copy the nat4d_MYYYY_dl (where YYYY represents the year of the file) file(s) and paste into the code package folder.

NOTE: some years have this file split into two separate files based upon the industry. These files must be merged together prior to running the code.



Name	Type	Compress...	Pass...	Size	Ratio	Date mod
field_descriptions	Microsoft Excel Worksheet	12 KB	No	14 KB	19%	3/15/2017
file_descriptions	Microsoft Excel Worksheet	10 KB	No	12 KB	23%	3/15/2017
nat3d_M2016_dl	Microsoft Excel Worksheet	5,450 KB	No	5,479 KB	1%	2/24/2017
nat3d_M2016_owner_dl	Microsoft Excel Worksheet	642 KB	No	648 KB	1%	2/24/2017
nat4d_M2016_dl	Microsoft Excel Worksheet	12,721 KB	No	12,784 KB	1%	2/24/2017
nat4d_M2016_owner_dl	Microsoft Excel Worksheet	1,327 KB	No	1,337 KB	1%	2/24/2017
nat5d_6d_M2016_dl	Microsoft Excel Worksheet	2,385 KB	No	2,400 KB	1%	2/24/2017
national_M2016_owner_dl	Microsoft Excel Worksheet	704 KB	No	711 KB	1%	2/24/2017
natsector_M2016_dl	Microsoft Excel Worksheet	2,269 KB	No	2,282 KB	1%	2/24/2017

4. Prepare the licensing micro data from your state or other local area

a. Using the same software as the code package you downloaded, clean and transform your local record-level licensing data to match the fields listed in Table 1 below. Table 3 gives some examples of records post-processing.

NOTE: Do not attempt to process the micro data using Excel, as there will likely be too many records to properly load in Excel, and your data will be truncated.

Table 1: Micro license data fields

FIELD NAME	FIELD DESCRIPTION
universal_id	Numeric identifier of a person in the licensing data. Individuals may hold more than one license, so there may be multiple identical universal_ids within the same dataset, but they should all represent the same person. This universal_id must also be a direct match to the same individual within the unemployment insurance quarterly wage records.
license_board	The issuing licensing board or committee
license_type	The name/title of the license. *This field must contain license types/titles that are exact matches (including spelling, spaces, and case) to the license_type field in the license_soc_crosswalk file.*
issue_date	The calendar date the license was issued. This field must be in the following format (Year-Month-Date) as: YYYY-MM-DD
expiration_date	The calendar date the license expired or is scheduled to expire. This field must be in the following format (Year-Month-Date) as: YYYY-MM-DD

status_date	The calendar date on which the status of the license was last updated in the system. This field must be in the following format (Year-Month-Date) as: YYYY-MM-DD
license_status	The most recent status of the license. This determines whether the license was legally able to be used by the individual for the year in question. Due to the high volume of license status values, IBRC developed 10 standardized license status categories. Your data will need to be aligned with the standardized 10 categories listed in License status values table.
licensernk	The priority rank for each standardized license status category. These rankings are listed in Table 2.

Table 2: License status values

STANDARDIZED LICENSE STATUS	PRIORITY RANK	NON-STANDARD STATUS VALUES ASSIGNED TO CATEGORY
ACTIVE	1	Active
CONDITIONALLY ACTIVE	2	Conditional, Probation, Probation/Referral, Valid to Practice While Reviewed
SUPERSEDED	3	Superseded
NOT PRACTICING	4	Current/Not Practicing, Retired
EXPIRED	5	Dead, Expired, Expired/Holding Application, Expired/Non-Renewable, Inactive, Inactive/Expired, Inactive/Probation, Null and Void, Probation/Expired, Referral/Expired, Registry History, Storage, Unassigned/Expired, Voluntary Surrender
SUSPENDED	6	Emergency Suspension, Suspended
REVOKED	7	Finding, Rescinded, Revoked
NOT AWARDED	8	Abandoned Application, Application Denied, Cancelled, Deleted Application, Failed Exam, Renewal Denied, Withdrawn Application
PENDING APPLICATION	9	Pending Application, Reinstatement Pending
UNASSIGNED	99	Unassigned

Table 3: Examples of micro license data records post-processing

universal_id	license_board	license_type	issue_date	expiration_date	status_date	license_status	licensernk
11111111	Nursing Board	Licensed Practical Nurse	2003-06-28	2017-06-31	2016-10-30	Active	1
11111112	Pharmacy Board	Pharmacy Tech In-Training	2011-08-20	2012-08-20	2012-06-01	Superseded	3
11111112	Pharmacy Board	Pharmacy Technician	2012-07-01	2018-06-30	2012-07-01	Active	1
11111113	Real Estate Commission	Real Estate Instructor Permit	2008-04-18	2012-05-30	2012-06-01	Expired	5

11111113	Real Estate Commission	Real Estate Broker	2005-10-15	2018-09-29	2005-10-15	Active	1
11111113	Certified Residential Appraiser	Appraiser Board	2013-03-16	2019-03-31	2013-03-16	Active	1

b. Once the data have been processed, save the data as a csv file in the code package folder, using "|" as the delimiter and name the file "microlicensedata.csv". DO NOT use a comma as a delimiter.

5. Open the license_soc_crosswalk.xlsx file and ensure that there are no licenses contained in your micro license data that are not listed in the crosswalk file.

a. Add any licenses available in the micro licensing data that are not included in the crosswalk. You must fill in the SOC code(s) associated with each type of license.
b. It is not necessary to delete any records from the crosswalk file, even though they may not be in your micro licensing data.
c. Ensure that the license titles from the crosswalk file match the license titles from your micro licensing data.
d. Save the file using the license_soc_crosswalk.xlsx name.

6. Prepare the wage records data from your state or other local area

a. Using the same software as the code package you downloaded, clean and transform your local record-level quarterly wage records data, for the year in question, to match the fields listed in Table 4 below. This format allows for the incorporation of as many occupations per individual as area available in your source data. However, due to diminishing marginal returns, and size constrains, more than 3 jobs per individual per quarter is not recommended.

Table 4: Quarterly wage records data fields

FIELD NAME	FIELD DESCRIPTION
universal_id	Numeric identifier of a person in the wage records data. Individuals may be on more than one payroll in the state, but the data should be in wide format for the Stata code to run properly, so the wage records should only contain one row per universal_id. This universal_id must also be a direct match to the same individual within the micro licensing data.
naicsYYYYqQ_jJ *example: naics2016q3_j1*	The 6-digit NAICS code of the employer for the year, quarter, and job in question. YYYY represents the 4 digit year, Q represents the number of the quarter, and J represents the job rank (individuals on only one payroll for the quarter will have only job number 1 in the wage records. Individuals on more than one payroll in the

	quarter will have the highest paying job listed as job number 1, and the second highest paying job listed as job number 2, etc.)
wgYYYYqQ_jJ *example: wg2016q3_j1*	The wage of the employee for the year, quarter, and job in question. YYYY represents the 4 digit year, Q represents the number of the quarter, and J represents the job rank (individuals on only one payroll for the quarter will have only job number 1 in the wage records. Individuals on more than one payroll in the quarter will have the highest paying job listed as job number 1, and the second highest paying job listed as job number 2, etc.)

b. Once the data have been processed, save the data as a csv file in the code package folder, using "|" as the delimiter and name the file "wagerecsYYYY.csv", with YYYY as the 4-digit year of the wage records. For example: wagerecs2016.csv. DO NOT use a comma as a delimiter.

7. Using the same software as the code package you downloaded, open the Licensing_SOC_Assignment code file (the file extension will depend upon the language of the code package you downloaded), and update the following inputs:

a. File pathway to folder with the transformed data files
b. Year of licensing data
c. SOC code vintage year
d. Number of jobs to assign in wage records per person
e. Indicator of bartenders in micro licensing data
f. Indicator of bartenders in micro licensing data
g. Indicator of engineers in micro licensing data
h. Indicator of asbestos licenses in micro licensing data
i. Indicator of nurses in micro licensing data
j. Indicator of nurse practitioners in micro licensing data
k. Indicator of nurse prescriptive licenses in micro licensing data
l. Indicator of cosmetologists in micro licensing data

8. Run the code.

9. The output from the model will contain the same fields as Table 5.

Table 5: Output fields for licensing occupation assignment module

FIELD NAME	FIELD DESCRIPTION
universal_id	Numeric identifier of a person in the wage and licensing data.
year	The year of the occupation assignment.
job	The job number of the individual. The job number indicates the quarterly wage ranking of the job, with 1 being the highest paying job.

quarter	The quarter of the occupation assignment.
naics	The standard NAICS code for the industry of employment associated with the occupation assignment
wg	The quarterly wage for the industry of employment associated with the occupation assignment
soc	The standard occupation code (SOC) as estimated by the model
soclvl	The level/methodological explanation of the SOC assignment made by the model
licensetot	The number of active licenses held by the individual during the year
NAICSmatchq	The number of quarters and jobs during the year with a match between the type of license and the industry of employment
assigntot	The number of quarters and jobs during the year for which a SOC assignment was made