

# MANUAL TÉCNICO DEL USO DE TESSERACT UI Y DESCARGA DE DATOS A PYTHON

Introducción a conceptos básicos de Tesseract UI

Versión 2

(16 de marzo de 2021)



## TABLA DE CONTENIDO

<b>1. Manejo de datos mediante Tesseract UI .....</b>	<b>1</b>
1.1. Drilldowns .....	2
1.2. Measures.....	3
1.3. Cuts .....	4
<b>2. Visualización y descarga de datos.....</b>	<b>6</b>
2.1. Exportar datos a Python .....	6

## 1. Manejo de datos mediante Tesseract UI

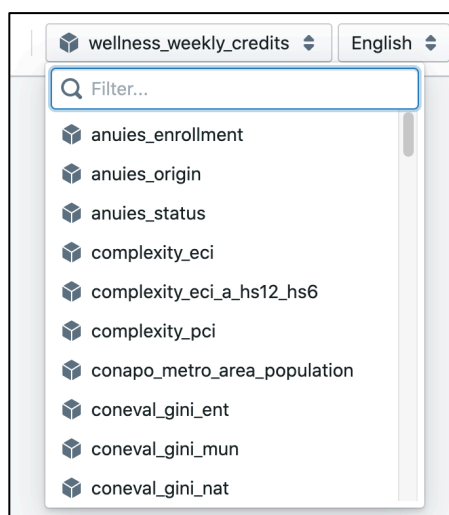
Para acceder de forma rápida y dinámica a los datos disponibles en dataMéxico se utiliza Tesseract UI, un explorador dinámico de API's disponible en [api.datamexico.org/ui](https://api.datamexico.org/ui).

Esta herramienta permite acceder a las distintas bases de datos existentes en dataMéxico. Su interfaz facilita el acceso y el filtrado de información en base a elementos como drilldowns, measures y cortes. La figura 1, muestra una visión general del API Explorer.

	Occupation ID	Occupation	Quarter ID	Quarter	Number of Records	Workforce	Monthly Wage
1	1111	Top Government and...	20101	2010-Q1	4	3488	3896.714
2	1111	Top Government and...	20102	2010-Q2	7	4313	9065.152
3	1111	Top Government and...	20103	2010-Q3	6	4714	15614.213
4	1111	Top Government and...	20104	2010-Q4	5	4577	24369.674
5	1111	Top Government and...	20111	2011-Q1	7	5196	1084.357
6	1111	Top Government and...	20112	2011-Q2	4	3242	4672.733
7	1111	Top Government and...	20113	2011-Q3	4	2520	11525.794
8	1111	Top Government and...	20114	2011-Q4	1	839	10000
9	1111	Top Government and...	20121	2012-Q1	4	2549	9178.894
10	1111	Top Government and...	20122	2012-Q2	7	4024	9670.974
11	1111	Top Government and...	20123	2012-Q3	11	1700	14095.294
12	1111	Top Government and...	20124	2012-Q4	15	3760	16848.936
13	1111	Top Government and...	20131	2013-Q1	13	2496	5426.282
14	1111	Top Government and...	20132	2013-Q2	10	1393	3783.202
15	1111	Top Government and...	20133	2013-Q3	17	1474	22198.948
16	1111	Top Government and...	20134	2013-Q4	10	3451	6461.895
17	1111	Top Government and...	20141	2014-Q1	15	2141	19871.008
18	1111	Top Government and...	20142	2014-Q2	14	1284	36707.632
19	1111	Top Government and...	20143	2014-Q3	10	702	32358.974
20	1111	Top Government and...	20144	2014-Q4	9	702	6598.291
21	1111	Top Government and...	20151	2015-Q1	12	1453	13264.281
22	1111	Top Government and...	20152	2015-Q2	8	829	27032.569
23	1111	Top Government and...	20153	2015-Q3	14	2311	11625.79
24	1111	Top Government and...	20154	2015-Q4	10	1328	8539.157
25	1111	Top Government and...	20161	2016-Q1	12	2946	2252.206
26	1111	Top Government and...	20162	2016-Q2	9	2989	84275.477
27	1111	Top Government and...	20163	2016-Q3	11	2143	6359.309
28	1111	Top Government and...	20164	2016-Q4	3	268	34029.851
29	1111	Top Government and...	20171	2017-Q1	7	621	11690.821

Figura 1. Visión general del API Explorer

En la barra superior es posible acceder a todos los cubos de datos disponibles. Se debe seleccionar un cubo y luego seleccionar *drilldowns* y *measures* para hacer la consulta de datos. La figura 2 muestra la ventana que emerge para seleccionar un cubo de datos.



**Figura 2. Selección de cubos**

Los principales elementos para seleccionar los datos de interés se encuentran en la barra lateral izquierda de la pantalla. Para ejecutar una consulta debe haber seleccionado, como mínimo, un *drilldown* y un *measure*. La consulta se ejecuta con el botón *Execute query* ubicado en la parte inferior de la barra lateral izquierda.

A continuación se entrega una breve descripción de los elementos disponibles.

### 1.1. Drilldowns

Los *drilldowns* corresponden a las variables que se pueden agregar a la salida de datos. Algunas variables presentan opciones agrupadas las cuales se pueden ir desglosando según el objetivo deseado. Las variables agrupadas más comunes son fechas (mes, trimestre, año) y geografías (municipios, metroáreas, entidades federativas, nación), pero dependiendo del cubo pueden existir otras variables con este comportamiento. La figura 3 muestra las opciones de *drilldowns* para un cubo en particular. Es posible seleccionar todas las variables de interés que se desplegarán en la consulta.

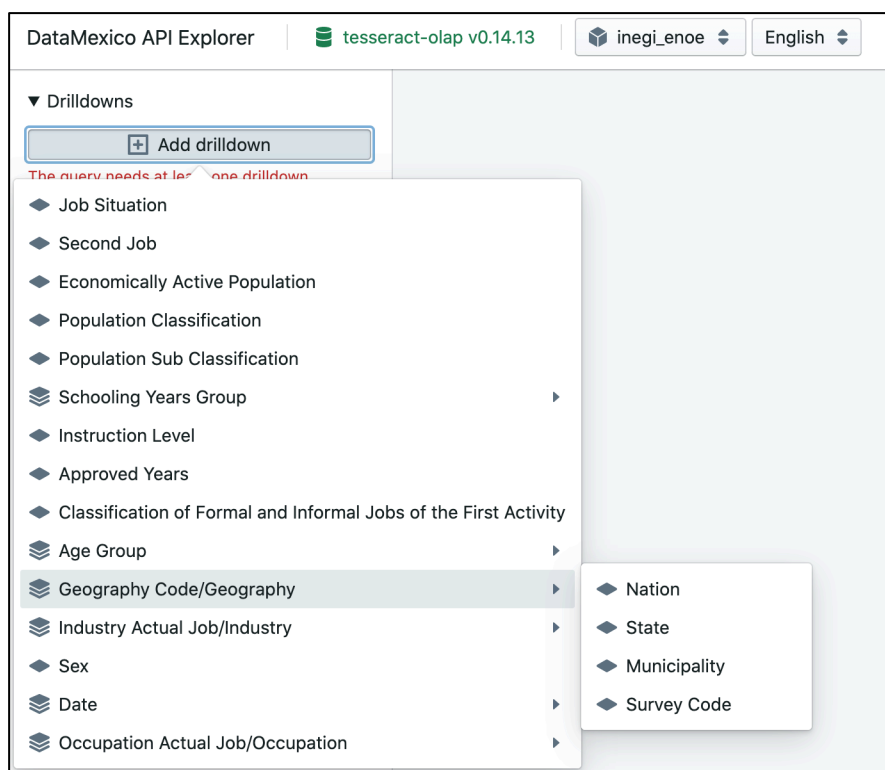


Figura 3. Visión general de Drilldowns

## 1.2. Measures

Cada cubo contiene diferentes *measures*. Un measure es básicamente una columna de datos numérica que se desplegará al hacer la consulta. La cantidad y tipo de *measure* dependerá del cubo seleccionado. Debe seleccionar al menos un *measure* para ejecutar la consulta. La figura 4 muestra un listado de seis *measures* donde dos de ellas han sido seleccionadas.

▼ Drilldowns

The query needs at least one drilldown.

▼ Measures

- ☒ Number of Records
- ☒ Workforce
- ☐ Worked Hours Week
- ☐ Worked Days Week
- ☐ Monthly Wage
- ☐ Workforce MOE

▼ Cuts

**Figura 4. Visión general de Measures**

### 1.3. Cuts

Los *cuts* o cortes permiten filtrar los datos que se desplegarán en la consulta. Todas las variables disponibles en los *drilldowns* están disponibles en *cuts* para hacer filtros.

Para usar correctamente esta propiedad:

1. Seleccione la variable del listado disponible al presionar *Add cut* (figura 5a).
2. La variable aparecerá disponible bajo *Cuts* (figura 5b).
3. Posicione el mouse sobre el nombre de la variable para que se desplieguen los valores disponibles.
4. Seleccione todos los valores necesarios para ejecutar el filtrado de datos (figura 5c).

A modo de ejemplo, las figuras 5a, 5b y 5c muestran cómo filtrar los datos de la entidad federativa de Aguascalientes.

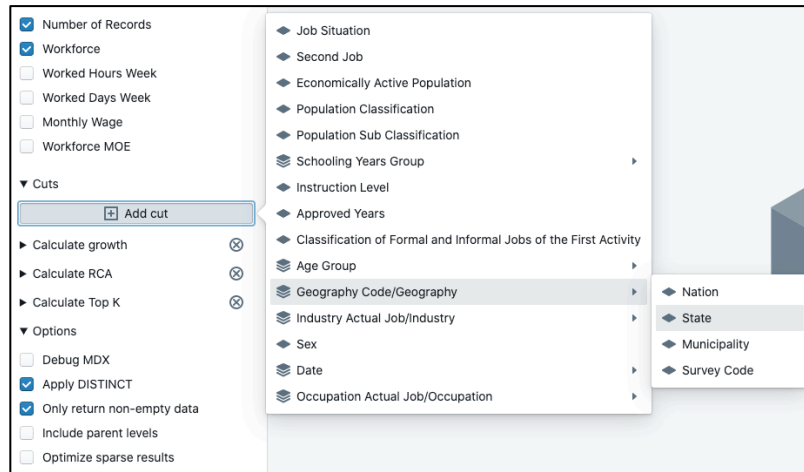


Figura 5a. Selección de variables para filtrado



Figura 5b. La variable seleccionada está disponible para filtrar

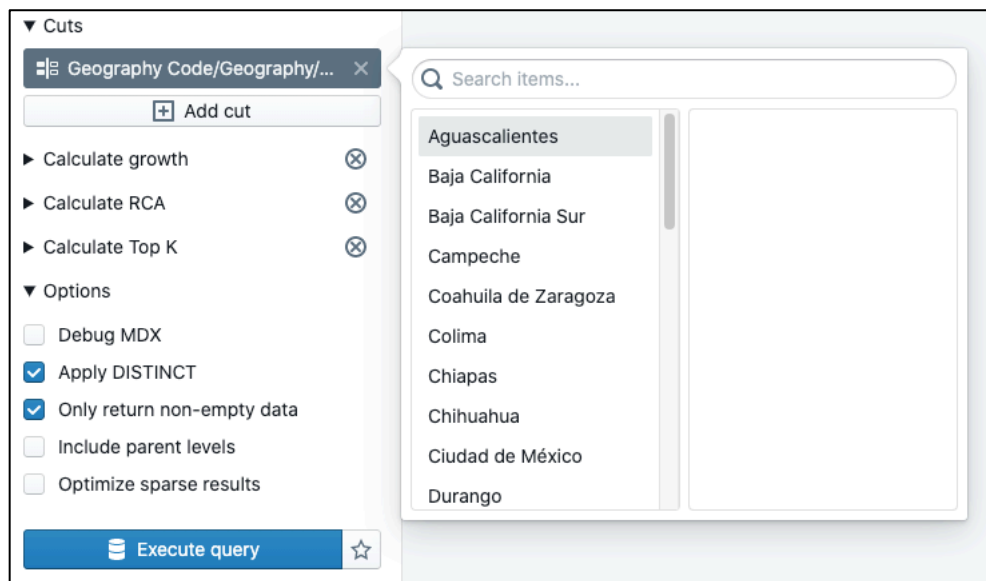


Figura 5c. Opciones disponibles para filtro y selección de Aguascalientes

## 2. Visualización y descarga de datos

En la parte superior de Tesseract UI se encuentran diferentes opciones para ver los datos de la consulta ejecutada y descargarlos. La figura 6 muestra las opciones.

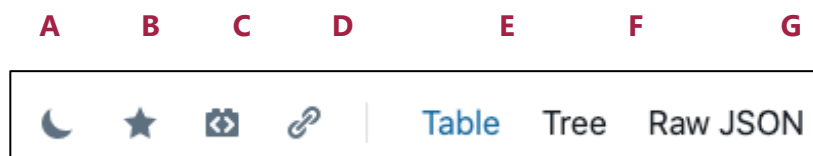



Figura 6 Opciones de visualización y descarga de datos

- a) **Modo Diurno/Nocturno:** Cambia los colores de la interfaz a una paleta de colores claras u oscuras.
- b) **Queries destacadas:** Repositorio de queries guardadas previamente.
- c) **Parámetros de debug:** Permite descargar datos en formato CSV, JSON Tidy y JSON Arrays. Además, entrega información acerca de los parámetros utilizados en la query en la sección *olap-client* y *query state*.
- d) **Copiar al portapapeles:** Facilita la copia de la URL, pasándola directo al portapapeles.
- e) **Table:** permite ver los datos de la query ejecutada en formato de tabla.
- f) **Tree:** permite ver los datos de la query ejecutada en formato de árbol.
- g) **Raw JSON:** permite ver los datos de la query ejecutada en formato JSON.

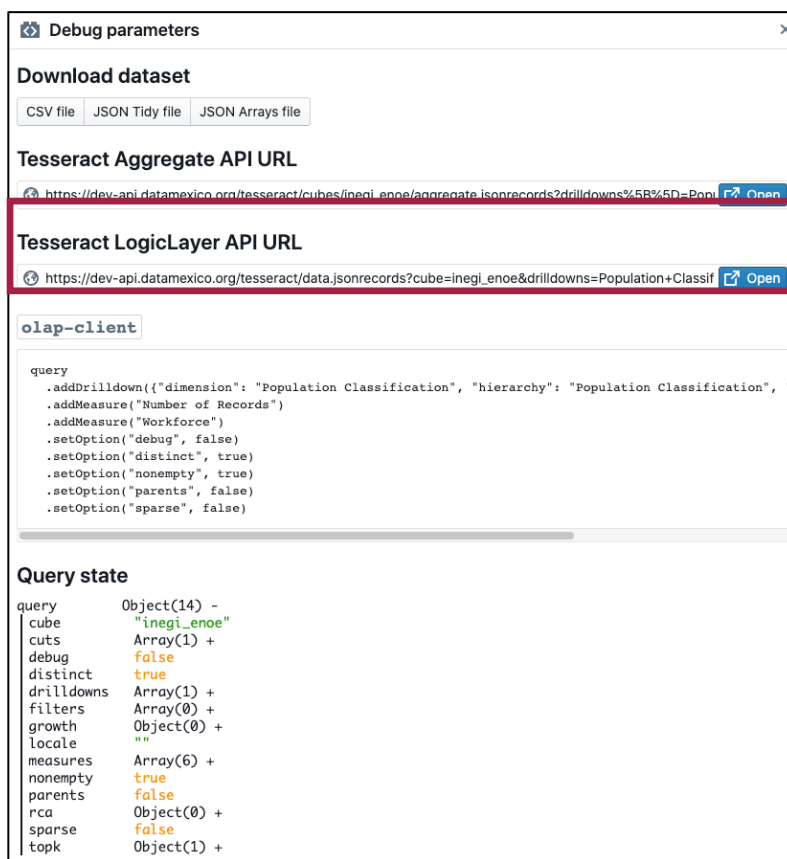
### 2.1. Exportar datos a Python

Para hacer sus propios análisis con los datos disponibles en Tesseract UI, se recomienda exportar los datos a Python y trabajarlos en un *dataFrame* utilizando la librería Pandas.

Para ello, una vez que ya ejecutó la consulta, debe ir al botón *Debug* (  ) mencionado en la sección anterior y que se ubica en la barra superior derecha de la pantalla. Ahí debe copiar la URL que se encuentra bajo el título *Tesseract LogicLayer API URL*. En estricto rigor puede utilizar cualquiera de las URL mostradas en *Debug Parameters*, pero la mencionada anteriormente es más fácil de entender por un humano y puede cambiar los parámetros con



facilidad. La figura 7 muestra la ventana desplegada al presionar el botón *Debug* y destaca la URL que se debe copiar para luego leer los datos con Python.



**Figura 7. Ventana desplegada al presionar el botón Debug y selección de URL**

Para hacer consultas más exactas, puede agregar parámetros adicionales al final de la URL de la consulta ejecutada. Las opciones disponibles permiten ordenar los datos de manera ascendente o descendente, traer un número establecido de filas o excluir determinados registros de la consulta. La forma de agregar estas opciones a la URL es la siguiente:

- Limit:** *&limit=value*, limita la consulta solo al número de resultados establecidos en *value*. Por ejemplo, si para determina consulta solo necesita 5 registros, bastará con agregar al final de la URL *&limit=5*
- Sort:** *&sort=<measure>.<asc/desc>*, ordena los resultados en base a una medida establecida. Por ejemplo, si tiene los datos de salario por ocupaciones, puede

ordenar las ocupaciones de mayor a menor salario agregando **&sort=Salario.desc**, donde salario es la measure que se ordena de forma descendente (más a menos).

- c) **Exclude:** **&exclude = <drilldown>:<value>**, excluye resultados que cumplan con la condición indicada. Por ejemplo, si un **drilldown** de rango de edad tiene 8 categorías, y la categoría 1 corresponde a menores de 15 años para los cuales no necesita los datos, podría interesarle trabajar con las restantes categorías y omitir la categoría número 1. Para ello debe agregar a la URL **&exclude=Age+Range:1**.

Una vez abierto un notebook de Python, solo se deben ejecutar las siguientes líneas para tener los datos en un dataframe:

```
import pandas as pd
import requests
data = requests.get("copie aquí la URL")
df = pd.DataFrame(data.json()["data"])
```

Si no tiene instaladas las librerías *pandas* y *requests* se recomienda revisar la documentación en línea en [https://pandas.pydata.org/pandas-docs/stable/getting\\_started/install.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/install.html) para el caso de *pandas* y en <https://requests.readthedocs.io/en/master/> para el caso de *requests*.

Luego de ejecutar estas líneas y con un poco de conocimiento de Python, puede procesar los datos y hacer sus propios cálculos, descargar varios set de datos y unirlos, analizar diferentes métricas y contar sus propias historias a través de los datos. La figura 8 muestra el resultado de ejecutar las líneas de código en Python.

df

	Year	Group ID	Group	State ID	State	Number of Records	Workforce	Monthly Wage
0	2010	11	Officials and Senior Officials from the Public...	1	Aguascalientes	138	14282	3658.661952
1	2010	12	Directors and Managers in Financial, Administr...	1	Aguascalientes	47	4919	4446.692417
2	2010	14	Directors and Sales Managers, Restaurants, Hot...	1	Aguascalientes	127	12750	2690.590510
3	2010	21	Specialists in Economic and Administrative Sci...	1	Aguascalientes	149	16492	5681.430390
4	2010	41	Traders in Stores	1	Aguascalientes	20	1634	1776.175031
...	...	...	...	...	...	...	...	...
457	2020	94	Aides in Food Preparation	1	Aguascalientes	33	3901	3004.232761
458	2020	95	Peddlers	1	Aguascalientes	105	12494	2773.161117
459	2020	96	Domestic Workers, Cleaners, Launderers and oth...	1	Aguascalientes	315	36852	3157.538424
460	2020	97	Parcel Workers, Support for Shows and Delivery...	1	Aguascalientes	25	2914	4416.612560
461	2020	99	Unspecified Occupations	1	Aguascalientes	3	369	0.000000

462 rows × 8 columns

**Figura 8. DataFrame generado con datos de API Explorer**