

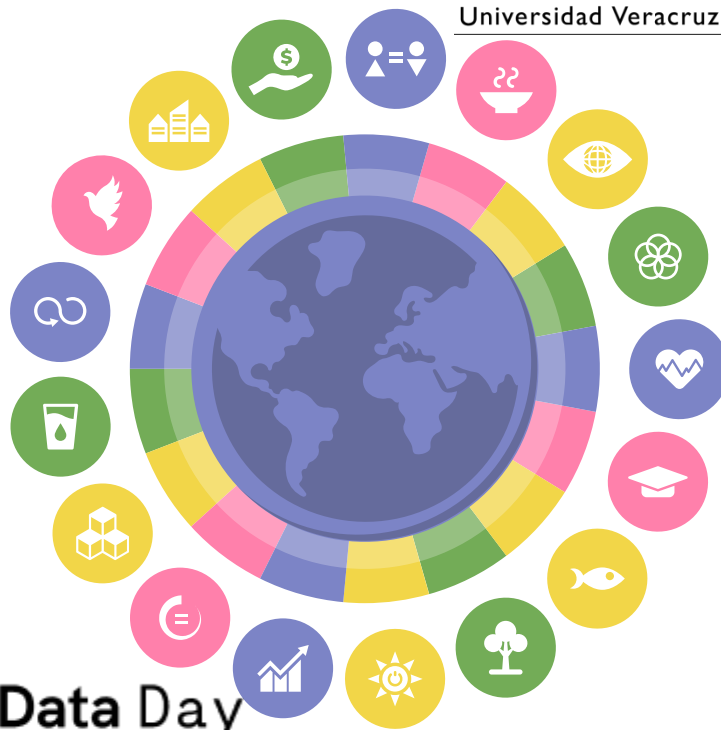
Datathon para la visualización de los ODS utilizando Datos Abiertos



Universidad Veracruzana



Día de los Datos Abiertos, 2024



Open Data Day



Programa



01

**ODS de la Agenda
2030**

MCD Lorena López Lozada

02

**Introducción al
manejo de datos**

Dr. Ángel Juan Sánchez García

03

**Visualización
espacial de datos**

Dr. Ángel Fernando Argüello Ortiz

04

**Introducción al
Aprendizaje Automático**

MIA Juan Luis López Herrera

Dr. Ángel Juan Sánchez García

• • • • •
angesanchez@uv.mx



02

Introducción al manejo de datos

Anaconda



Distribución de Python que está diseñada para simplificar el proceso de instalación y gestión de paquetes y entornos de Python.

Contiene una amplia variedad de paquetes preinstalados que son comúnmente utilizados en ciencia de datos, aprendizaje automático, análisis de datos y otras áreas relacionadas.

<https://www.anaconda.com/download>

Anaconda Navigator



Anaconda Navigator



Upgrade Now

Home

Environments

Learning

Community

Anaconda Toolbox
Supercharged local notebooks. Click the Toolbox tile to install.

Read the Docs

Documentation

Anaconda Blog

All applications
on
base (root)
Channels

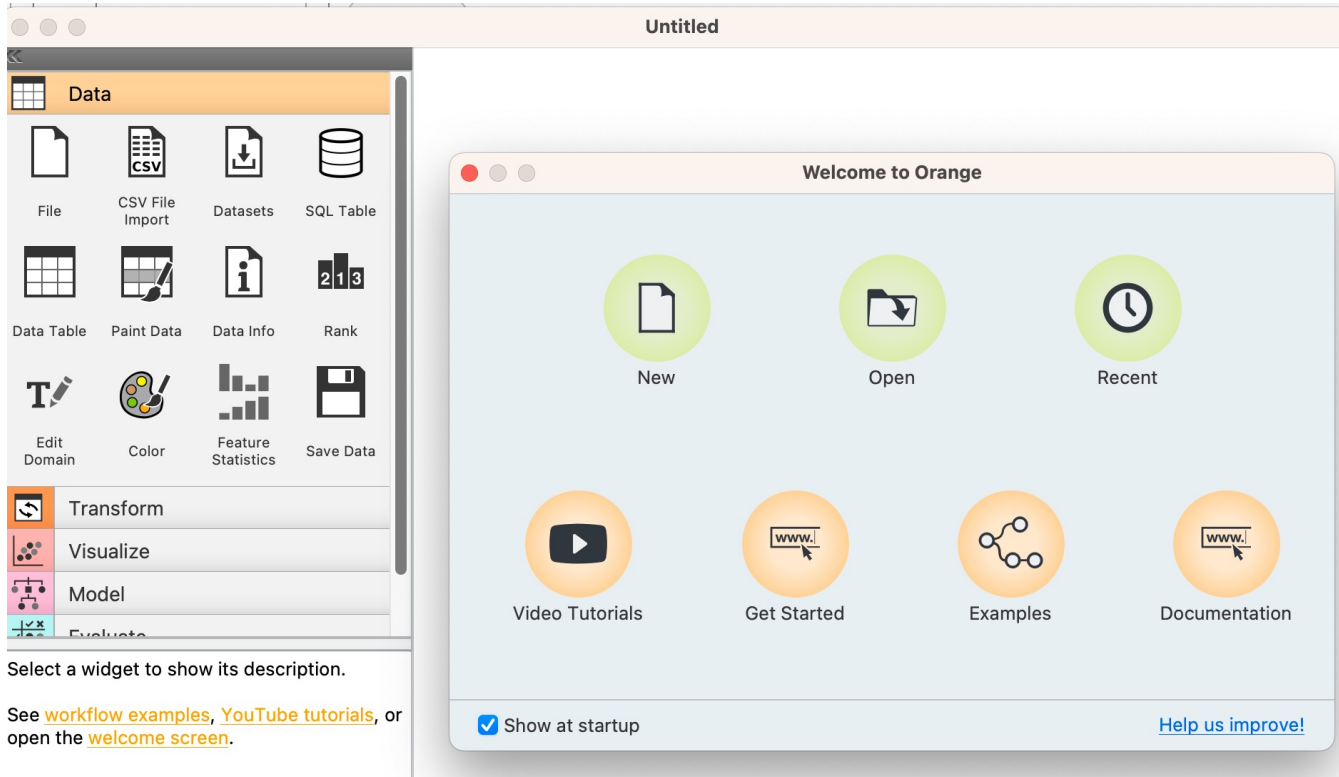
<div style="text-align: center;"></div> <p>DataSpell</p> <p>DataSpell is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.</p> <div style="text-align: center; border: 2px solid red; padding: 5px; width: fit-content; margin: 0 auto;">Install</div>	<div style="text-align: center;"></div> <p>Anaconda Notebooks</p> <p>Cloud-hosted notebook service from Anaconda. Launch a preconfigured environment with hundreds of packages and store project files with persistent cloud storage.</p> <div style="text-align: center; margin-top: 10px;">Launch</div>	<div style="text-align: center;"></div> <p>JupyterLab</p> <p>3.6.3</p> <p>An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.</p> <div style="text-align: center; margin-top: 10px;">Launch</div>	<div style="text-align: center;"></div> <p>Jupyter Notebook</p> <p>6.5.4</p> <p>Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.</p> <div style="text-align: center; margin-top: 10px;">Launch</div>
<div style="text-align: center;"></div> <p>Orange 3</p> <p>3.34.0</p> <p>Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.</p> <div style="text-align: center; border: 2px solid green; padding: 5px; width: fit-content; margin: 0 auto;">Launch</div>	<div style="text-align: center;"></div> <p>Qt Console</p> <p>5.4.2</p> <p>PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.</p> <div style="text-align: center; margin-top: 10px;">Launch</div>	<div style="text-align: center;"></div> <p>Spyder</p> <p>5.4.3</p> <p>Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features</p> <div style="text-align: center; margin-top: 10px;">Launch</div>	<div style="text-align: center;"></div> <p>Anaconda on AWS Graviton</p> <p>Running your Anaconda workloads on AWS Graviton-based processors could provide up to 40% better price performance</p> <div style="text-align: center; margin-top: 10px;">Launch</div>

¿Por qué Orange?

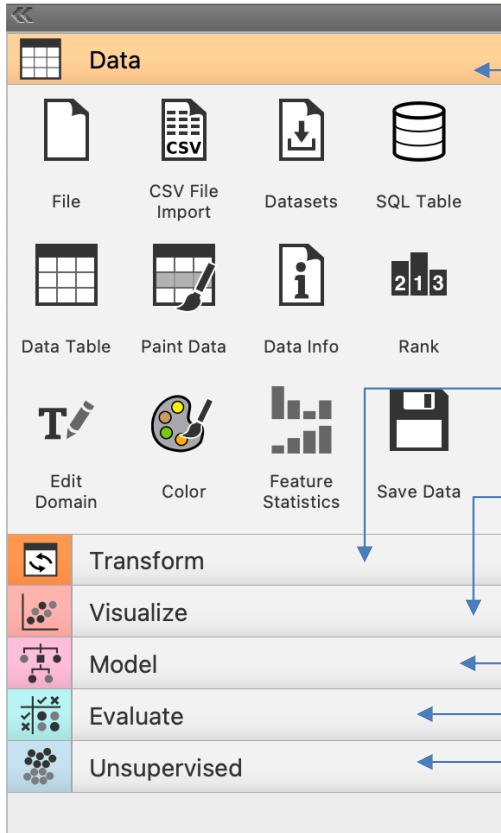
1. Interfaz gráfica intuitiva que facilita la creación y ejecución de flujos de trabajo de análisis de datos sin necesidad de escribir código.
2. Amplia variedad de herramientas para la manipulación de datos, visualización, modelado predictivo entre otras (arrastrar y soltar) .
3. Visualización de datos interactiva para explorar y comprender los datos de manera rápida y eficiente.
4. Integración con Python para aprovechar las bibliotecas y herramientas de análisis de datos de Python



Orange



Orange (Widgets)



Conectar fuentes de datos, analizar y desempeñar operaciones sobre datos.

Analizar visualmente a través de gráficos (barras, dispersión, etc).

Modelar y administrar la persistencia de los modelos

Predecir datos y calificar modelos

Modelar no supervisadamente

Datos para trabajar

Internacional

<https://biblioguias.cepal.org/odsdata>

Nacional

<https://datamx.io/dataset/indicadores-ods>

Agenda 2030

<https://agenda2030.mx/#/home>

INEGI 2020

https://www.inegi.org.mx/programas/ccpv/2020/#datos_abiertos

Visualización



Scatter Plot

Visualizar la relación entre dos variables cuantitativas, datos atípicos, patrones.



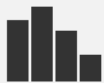
Box Plot

Visualizar la distribución de un conjunto de datos y resumir varias características como la mediana, los cuartiles, el rango intercuartílico y los valores atípicos potenciales



Bar Plot

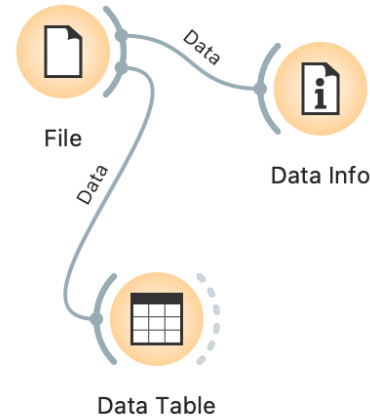
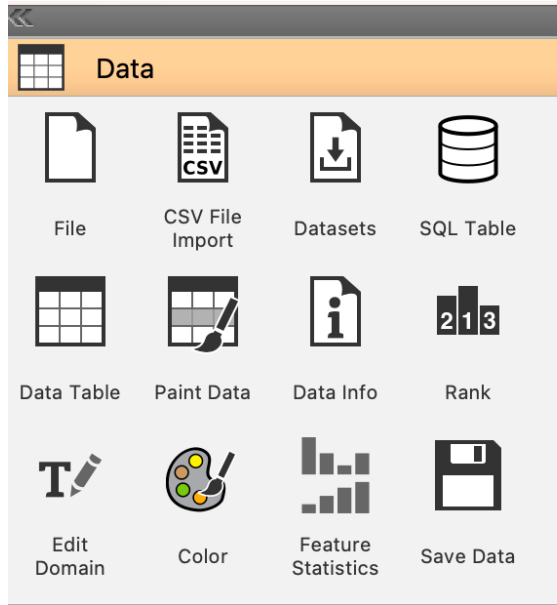
Visualizar la relación entre una variable categórica y una variable numérica. Estos gráficos muestran la frecuencia o la suma de la variable numérica para cada categoría de la variable categórica.



Distributions

Visualizar la distribución de los datos y comprender la forma y la dispersión de los datos, así como para identificar patrones, tendencias y posibles valores atípicos en un conjunto de datos

Cargar datos en Orange



¡Vamos a practicar con los widgets!



Datos INEGI 2020

⊖ Título ↓	Periodo ↓	Formatos	
⊖ Archivos para descarga			
⊖ Principales resultados por localidad (ITER)			
Estados Unidos Mexicanos	2020	i	CSV 30.6 MB
Aguascalientes	2020	i	CSV 289 KB
Baja California	2020	i	CSV 532 KB
Baja California Sur	2020	i	CSV 195 KB
Campeche	2020	i	CSV 300 KB
Coahuila de Zaragoza	2020	i	CSV 435 KB
Colima	2020	i	CSV 129 KB
Chiapas	2020	i	CSV 3.20 MB
Chihuahua	2020	i	CSV 1.27 MB
Ciudad de México	2020	i	CSV 149 KB
Durango	2020	i	CSV 826 KB
Guanajuato	2020	i	CSV 1.62 MB
Guerrero	2020	i	CSV 1.43 MB

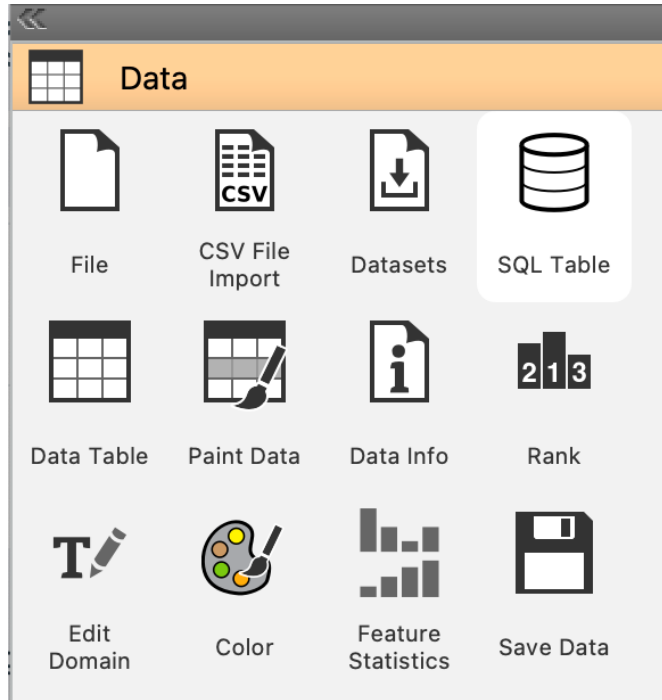
Datos INEGI 2020 (datos)

A	B	C	D	E	F	G	H	I	J	K	L	M	N
	NOM_ENT	MUN	NOM_MUN	LOC	NOM_LOC	AGEB	MZA	POBTOT	POBFEM	POBMAS	P_0A2	P_0A2_F	P_0A2_M
30	Veracruz de I	000	Total de la en	0000	Total de la en	0000	000	8062579	4190805	3871774	342746	169702	173044
30	Veracruz de I	001	Acajete	0000	Total del mur	0000	000	9701	4781	4920	514	245	269
30	Veracruz de I	001	Acajete	0001	Total de la lo	0000	000	1810	882	928	84	41	43
30	Veracruz de I	001	Acajete	0001	Total AGEB u	0028	000	1441	707	734	65	32	33
30	Veracruz de I	001	Acajete	0001	Acajete	0028	001	2 *	*	*	*	*	*
30	Veracruz de I	001	Acajete	0001	Acajete	0028	002	0	0	0	0	0	0
30	Veracruz de I	001	Acajete	0001	Acajete	0028	003	27	13	14 *	*		0
30	Veracruz de I	001	Acajete	0001	Acajete	0028	004	14	7	7 *	*		0
30	Veracruz de I	001	Acajete	0001	Acajete	0028	005	25	10	15	0	0	0
30	Veracruz de I	001	Acajete	0001	Acajete	0028	006	119	59	60	7	5 *	
30	Veracruz de I	001	Acajete	0001	Acajete	0028	007	133	67	66	7	5 *	
30	Veracruz de I	001	Acajete	0001	Acajete	0028	008	22	12	10 *		0 *	

Datos INEGI 2020 (Diccionario de datos)

A		B
1 CENSO DE POBLACIÓN Y VIVIENDA 2020		
2 Relación de indicadores		
		3 Tercera edición.
4 Núm.	Indicador	Descripción
5 1	Clave de enti	Código que identifica a la entidad federativa. El código 00 identifica a los registros con los totales a nivel nacional.
6 2	Entidad fede	Nombre oficial de la entidad federativa.
7 3	Clave de mur	Código que identifica al municipio o demarcación territorial al interior de una entidad federativa, conforme al Marco Geoestadísti
8 4	Municipio o c	Nombre oficial del municipio o demarcación territorial en el caso de la Ciudad de México.
9 5	Clave de locz	Código que identifica a la localidad al interior de cada municipio o demarcación territorial conforme al Marco Geoestadístico. El c
0 6	Localidad	Nombre con el que se reconoce a la localidad dado por la ley o la costumbre.
1 7	Clave del AG	Clave que identifica al AGEB urbana, al interior de una localidad, de acuerdo con la desagregación del Marco Geoestadístico.
2 8	Clave de mar	Clave que identifica a la manzana, al interior de una AGEB, de acuerdo a la desagregación del Marco Geoestadístico.
3 1	Población tot	Total de personas que residen habitualmente en el país, la entidad federativa, el municipio o la demarcación territorial y la localid.
4 2	Población fer	Total de mujeres que residen habitualmente en el país, la entidad federativa, el municipio o la demarcación territorial y la localida
5 3	Población mz	Total de hombres que residen habitualmente en el país, la entidad federativa, el municipio o la demarcación territorial y la localidz
6 4	Población de	Personas de 0 a 2 años de edad.
7 5	Población fer	Mujeres de 0 a 2 años de edad.
8 6	Población mz	Hombres de 0 a 2 años de edad.
9 7	Población de	Personas de 3 a 130 años de edad.
0 8	Población fer	Mujeres de 3 a 130 años de edad.
1 9	Población mz	Hombres de 3 a 130 años de edad.
2 10	Población de	Personas de 5 a 130 años de edad.
3 11	Población fer	Mujeres de 5 a 130 años de edad.
4 12	Población mz	Hombres de 5 a 130 años de edad.




Cargar datos en Orange




File


Cargar datos en Orange

Source

File: conjunto_de_datos_ageb_urbana_30_cpv2020.csv   ...  Reload

URL: 








File Type

Automatically detect type 

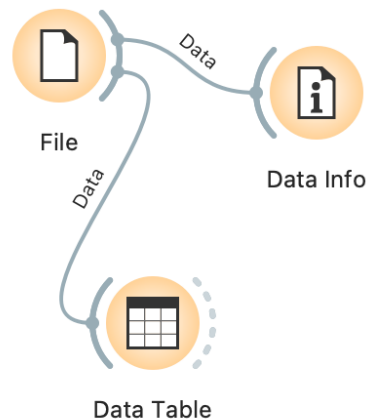
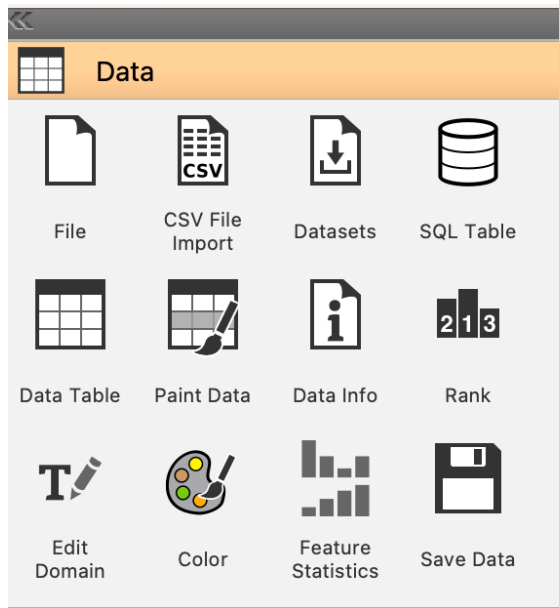
Info

97957 instances
7 features (no missing values)
Data has no target variable.
223 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	ENTIDAD	 numeric	feature	
2	NOM_ENT	 categorical	feature	Veracruz de Ignacio de la Llave
3	MUN	 numeric	feature	
4	LOC	 numeric	feature	
5	MZA	 numeric	feature	
6	POBTOT	 numeric	feature	
7	VIVTOT	 numeric	feature	

Cargar datos en Orange



¡Vamos a practicar con los widgets!

Cargar datos en Orange

- Con un editor de texto, buscar y reemplazar los '*' y 'N/P' con espacio en blanco
- Recargar la entrada de datos

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0400	005	33	28	27			0		31
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0400	006	2 *	*	*	*	*	*	*	*
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0400	007	9	5	4	0	0	0	0	9
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0400	008	37	22	15	3 *	*	*	*	34
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0400	009	21	11	10	N/D	N/D	N/D	N/D	
30 Veracruz de l	065	Emiliano Zap	0169	Total AGEB u	0434	000	2219	1185	1034	65	29	36	2154	
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0434	001	44	27	17 *	0 *			43	
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0434	002	60	29	31	0	0	0	60	
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0434	003	72	36	36 *	*		0	71	
30 Veracruz de l	065	Emiliano Zap	0169	Jacarandas	0434	004	53	33	20 *	*		0	52	

Indicadores de los datos

Indicador	Rango de variables
Datos demográficos	1-3
Edad y género	4-47
Nacimiento en la entidad	49-54
Edad y entidad federativa	55-60
Hablantes de lenguas indígenas	61-72
Afrodecendencia	74-76
Discapacidad	77-92
Escolaridad	93-134
Laborales	135-146
Derecho habientes	147-156
...	...

¿Cómo empezar?

Plantear un área que te gustaría investigar.

Plantear una o varias preguntas de investigación que te gustaría presentar o responder

Ejemplo

Área: Discapacidad

Identificar Variables:

Poblacion con discapacidad

Poblacion con discapacidad para caminar

Poblacion con discapacidad ver, aun con lentes

Poblacion con discapacidad para hablar

Poblacion con discapacidad para oir

.....

Hacer un análisis exploratorio.

Ejemplo

76	Población masculina que se considera afroamericana o afrodescendiente	Hombres que se consideran afroamericanos o afrodescendientes.
77	Población con discapacidad	Personas que realizan con mucha dificultad o no pueden hacer al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usa
78	Población con discapacidad para caminar, subir o bajar	Personas que tiene mucha dificultad o no pueden caminar, subir o bajar.
79	Población con discapacidad para ver, aun usando lentes	Personas que tienen mucha dificultad o no pueden ver, aun cuando usen lentes.
80	Población con discapacidad para hablar o comunicarse	Personas que tienen mucha dificultad o no pueden hablar o comunicarse.
81	Población con discapacidad para oír, aun usando aparato auditivo	Personas que tienen mucha dificultad o no pueden oír, aun usando aparato auditivo.
82	Población con discapacidad para vestirse, bañarse o comer	Personas que tienen mucha dificultad o no pueden bañarse, vestirse o comer.
83	Población con discapacidad para recordar o concentrarse	Personas que tienen mucha dificultad o no pueden recordar o concentrarse.
84	Población con limitación	Personas que realizan con poca dificultad al menos una de las siguientes actividades: ver, aun usando lentes; oír, aun usando aparato auditivo
85	Población con limitación para caminar, subir o bajar	Personas que tienen poca dificultad para caminar, subir o bajar.
86	Población con limitación para ver, aun usando lentes	Personas que tienen poca dificultad para ver, aun usando lentes.
87	Población con limitación para hablar o comunicarse	Personas que tienen poca dificultad para hablar o comunicarse.
88	Población con limitación para oír, aun usando aparato auditivo	Personas que tienen poca dificultad para oír, aun usando aparato auditivo.
89	Población con limitación para vestirse, bañarse o comer	Personas que tienen poca dificultad para bañarse, vestirse o comer.
90	Población con limitación para recordar o concentrarse	Personas que tienen poca dificultad para recordar o concentrarse.
91	Población con algún problema o condición mental	Personas con algún problema o condición mental, como las relacionadas con el trastorno de la conciencia, retraso mental y las alteraciones c
92	Población sin discapacidad, limitación, problema o condición mental	Personas que no tienen dificultad para realizar alguna actividad cotidiana como: ver, aun usando lentes; oír aun usando aparato auditivo; cam

Ejemplo

En el widget “select columns” de la Sección “Transform” vamos a ignorar todas excepto las de interés.

The screenshot displays a data transformation tool interface with two main panels: "Ignored (218)" and "Features (7)".

Ignored (218) Panel: Contains a list of 218 columns, each with a red 'N' icon. The visible columns are:

- ENTIDAD
- NOM_ENT
- MUN
- LOC
- MZA
- POBTOT
- POBFEM
- POBMAS
- P_0A2
- P_0A2_F
- P_0A2_M
- P_3YMAS
- P_3YMAS_F
- P_3YMAS_M
- P_5YMAS
- P_5YMAS_F
- P_5YMAS_M
- P_12YMAS
- P_12YMAS_F
- P_12YMAS_M
- P_15YMAS
- P_15YMAS_F

Features (7) Panel: Contains a list of 7 columns, each with a red 'N' icon. The visible columns are:

- PCON_DISC
- PCDISC_MOT
- PCDISC_VIS
- PCDISC LENG
- PCDISC_AUD
- PCDISC_MOT2
- PCDISC_MEN

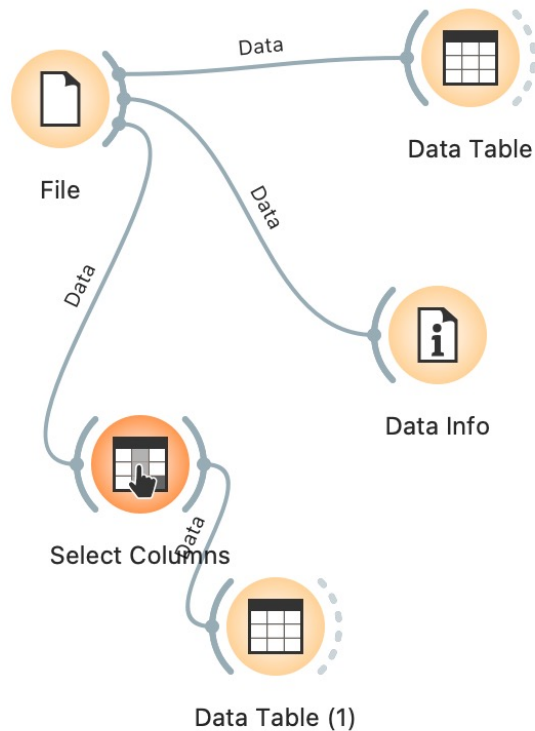
Target Panel: Contains a text input field for defining the target variable.

Metas (5) Panel: Contains a list of 5 columns, each with a black 'S' icon. The visible columns are:

- NOM_MUN
- NOM_LOC
- AGEB
- REL_H_M
- PROM_HNV

Navigation arrows (>) are present between the panels to move items between them.

Ejemplo



¿Cómo seguir?

Plantear una o varias preguntas de investigación que te gustaría presentar o responder

¿La población con discapacidad está relacionada con alguna en especial?

La población con discapacidad para comunicarse en Veracruz está relacionada con la población discapacidad visual en veracruz?

Las que sen de tu interés.....

Otro ejemplo

1. Trabajar con un conjunto de registros, por ejemplo ciudad de Xalapa o acajete.
2. ¿Debería haber una relación entre la población con discapacidad y los servicios de derechohabencia?
3. Las demás preguntas dependen de ti... Todo al alcance de un clic y un flujo

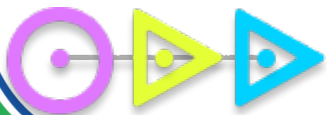
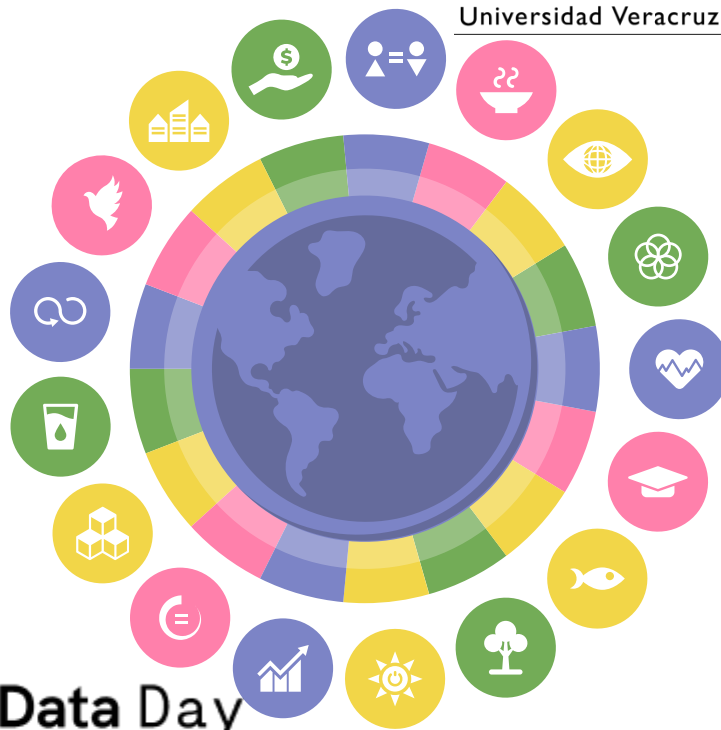
Datathon para la visualización de los ODS utilizando Datos Abiertos



Universidad Veracruzana



Día de los Datos Abiertos, 2024



Open Data Day