

TRABAJO DE GRADO
Opción Seminario-Diplomado

APRENDIZAJE DE NUEVAS HERRAMIENTAS DE IA PARA EL ANÁLISIS Y
VISUALIZACIÓN DE DATOS

Corporación Universitaria Remington.

Facultad de Ingeniería

Seminario de Análisis y Visualización de Datos

Ángela María Arcila Gálvez

Tutor: John Fredy Mira Mejía

Opción de Trabajo de grado Seminario-Diplomado.

2024

TRABAJO DE GRADO
Opción Seminario-Diplomado

DEDICATORIA

A mis hermosos hijos,

El presente documento es el resultado y finalización de un gran esfuerzo, dedicación y vocación por la dirección que le he dado a mi carrera. Quiero dedicar esto a ustedes, que son el motor de mi vida, la mayor motivación y fuente de inspiración. Son los que hacen que cada día quiera ser mejor y alcanzar más metas, imponiéndome propósitos más altos cada vez.

Gracias por ser mi apoyo y comprender los tiempos dedicados a la culminación de este esfuerzo. Espero que este esfuerzo los motive en sus propias carreras, a seguir sus sueños, alcanzar las metas que se han trazado y a no rendirse ante ningún obstáculo; no importa la situación, la edad o el tiempo, con esfuerzo y dedicación podemos lograr alcanzar nuestros propósitos.

Los amo.

Ángela María Arcila Gálvez

TRABAJO DE GRADO
Opción Seminario-Diplomado

TABLA DE CONTENIDO

1. RESUMEN.....	8
2. INTRODUCCIÓN.....	10
3. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.....	13
3.1. Formulación del problema.....	13
3.2. Problemas específicos.....	14
4. OBJETIVOS DE INVESTIGACIÓN.....	15
4.1. Objetivo General.....	15
4.2. Objetivos Específicos.....	15
5. JUSTIFICACIÓN.....	16
6. MARCO CONCEPTUAL Y CONTEXTUAL.....	17
6.1. Marco conceptual.....	17
6.2. Contexto Educativo.....	17
6.3. Orange.....	18
7. METODOLOGÍA.....	20
8. DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE.....	22
9. APLICACIÓN DE LA HERRAMIENTA DE INTELIGENCIA ARTIFICIAL: ORANGE.....	23
9.1. Fuentes de extracción.....	23
9.2. Criterios de Selección.....	23
9.3. Definición de variables.....	24

TRABAJO DE GRADO
Opción Seminario-Diplomado

10. ANÁLISIS Y VISUALIZACIÓN DE LA BASE DE DATOS MEDIANTE ORANGE.....	26
10.1. Aprendiendo qué es Orange.....	26
10.2. Aprendiendo para qué sirve Orange.....	26
10.3. Inicio.....	27
10.4. El lienzo.....	27
11. TRAER OTRA BASE DE DATOS PARA COMPARAR.....	37
12. INTEGRACIÓN DE AMBAS BASES DE DATOS PARA PODER REALIZAR EL COMPARATIVO ENTRE AMBAS.....	43
13. INTEGRACIÓN DE GRÁFICOS AL PROCESO DE DATOS.....	45
13.1. Visualización Por distribución.....	45
13.2. Visualización Por bar plot.....	48
13.3. Visualización por scatter plot.....	49
13.4. Visualización por line plot.....	50
14. UTILIZACIÓN DE OTRAS DE LAS HERRAMIENTAS QUE PRESENTA ORANGE.....	51
14.1. Unsupervised correlaciones.....	51
14.2. Unsupervised distances.....	53
15. RESULTADOS ESPERADOS.....	58
16. CONCLUSIONES.....	59
17. REFERENCIAS BIBLIOGRAFICAS.....	62

TRABAJO DE GRADO
Opción Seminario-Diplomado

TABLA DE FIGURAS

Figura 1.....	28
Figura 2.....	29
Figura 3.....	29
Figura 4.....	30
Figura 5.....	31
Figura 6.....	32
Figura 7.....	32
Figura 8.....	33
Figura 9.....	33
Figura 10	34
Figura 11.....	35
Figura 12.....	36
Figura 13.....	36
Figura 14.....	37
Figura 15.....	38
Figura 16.....	39
Figura 17.....	40
Figura 18.....	40
Figura 19.....	41
Figura 20.....	41
Figura 21.....	42

TRABAJO DE GRADO
Opción Seminario-Diplomado

Figura 22.....	43
Figura 23.....	44
Figura 24	44
Figura 25.....	45
Figura 26	46
Figura 27.....	47
Figura 28.....	47
Figura 29.....	48
Figura 30.....	48
Figura 31.....	49
Figura 32.....	49
Figura 33.....	50
Figura 34.....	50
Figura 35.....	51
Figura 36.....	52
Figura 37	53
Figura 38.....	53
Figura 39.....	54
Figura 40.....	54
Figura 41.....	55
Figura 42.....	56
Figura 43.....	57
Figura 44.....	57

TRABAJO DE GRADO
Opción Seminario-Diplomado

TRABAJO DE GRADO
Opción Seminario-Diplomado

1. RESUMEN

El presente documento se enfoca en el aprendizaje del análisis y visualización de datos utilizando una herramienta de inteligencia artificial llamada Orange. Durante el desarrollo del mismo se exploran las diferentes funcionalidades de Orange para procesar, analizar, integrar y visualizar datos de forma rápida y efectiva.

El principal objetivo es el de adquirir habilidades en el manejo de datos y en la utilización de técnicas de análisis y visualización para la toma de decisiones en cualquier área.

Para alcanzar este objetivo, se realizan distintas actividades como la importación de un conjunto específico de datos, la limpieza y procesamiento de la información, la aplicación de aprendizaje automático y la creación de visualizaciones creativas, atractivas e interactivas para una mejor toma de decisiones.

En el presente desarrollo, aprendemos a utilizar Orange para realizar tareas como la clasificación, clusterización, regresión y detección de anomalías en los datos. Además se profundiza en las diferentes herramienta para la visualización de los datos como por ejemplo gráficos por distribución en diferentes opciones, por bar plot, scatter plot; también se realiza el aprendizaje de herramientas de aprendizaje sin supervisión proporcionadas por Orange.

TRABAJO DE GRADO
Opción Seminario-Diplomado

Este documento brinda la oportunidad de adquirir nuevos conocimientos avanzados en el análisis y visualización de datos mediante el uso de herramientas de inteligencia artificial, en este caso Orange, lo que permite mejorar las habilidades en la toma de decisiones basadas en datos y en la resolución de problemas complejos en distintos ámbitos profesionales.

Palabras clave:

Análisis, visualización, inteligencia artificial, datos, aprendizaje.

TRABAJO DE GRADO
Opción Seminario-Diplomado

2. INTRODUCCIÓN

El presente documento se enfoca en el aprendizaje del análisis y visualización de datos utilizando una herramienta de inteligencia artificial en este caso: Orange. Durante el desarrollo del mismo se exploran las diferentes funcionalidades de Orange para procesar, analizar, integrar y visualizar datos de forma rápida y efectiva.

El principal objetivo es el de adquirir habilidades en el manejo de datos y en la utilización de técnicas de análisis y visualización para la toma de decisiones en cualquier área. Para llegar a un verdadero conocimiento y alcanzar nuestros objetivos se realizan diferentes tareas como la importación de bases de datos, la limpieza y procesamiento de la información, la aplicación de aprendizaje automático no supervisado y la creación de visualizaciones creativas, atractivas e interactivas para una mejor toma de decisiones.

Se espera con este proyecto, ayudar a otros a adquirir las habilidades necesarias para utilizar Orange de manera eficiente y efectiva en la realización de análisis de datos en diferentes contextos, teniendo en cuenta que no es solo el conocimiento de una herramienta de inteligencia artificial en específico, que existen muchas de ellas en el mercado tanto de código libre como es nuestro caso como de pago, lo importante para el presente proyecto es ayudar a que

TRABAJO DE GRADO
Opción Seminario-Diplomado

se conozco que existen herramientas con las funcionalidades necesarias para realizar un excelente análisis de datos y su respectiva toma de decisiones.

Actualmente, la inteligencia artificial se ha convertido en una herramienta muy importante y muy utilizada para el análisis de datos y la toma de decisiones en diversos sectores. Es por eso que debemos reconocer la necesidad de adquirir nuevas habilidades en el manejo de los diferentes tipos de herramientas de inteligencia artificial disponibles, ya que cada una nos ofrece diferentes capacidades y funciones específicas que nos pueden ayudar en la resolución de problemas y en el mejoramiento de procesos.

Una de las mayores virtudes de las técnicas de aprendizaje automático no supervisado con inteligencia artificial, permite a las organizaciones identificar patrones y tendencias en grandes volúmenes de datos de forma rápida y eficiente, lo que facilita la toma de decisiones basadas en datos con mayor precisión.

De igual forma, las herramientas de visualización de los datos para ser analizados y procesados, es de gran utilidad en la detección de anomalías, en la optimización de procesos o en la mejora en la adquisición de la información; por lo tanto, conocer los diferentes tipos de herramientas de inteligencia artificial que están disponibles, nos permite que podamos aprovechar al máximo su potencial y obtener resultados más precisos, no sólo por aquellas personas dedicadas o con conocimientos en programación, sino también

TRABAJO DE GRADO
Opción Seminario-Diplomado

aquellos que carecen de esta experiencia, ya que sabemos que el análisis de la información y la toma de decisiones no es exclusiva de determinados saberes.

TRABAJO DE GRADO
Opción Seminario-Diplomado

3. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

a. Formulación del problema:

Se podría decir que aún hoy en día el aprendizaje de nuevas herramientas de inteligencia artificial para el análisis y visualización de los datos es un desafío constante para los profesionales y estudiantes en general. A medida que la tecnología avanza surgen nuevas necesidades y con ello nuevas herramientas, es crucial poder llegar a estar al tanto de la mayoría de ellas y de como utilizarlas de forma efectiva.

Uno de los grandes inconvenientes radica en la falta de educación en el tema, así como los recursos disponibles; a menudo los profesionales y estudiantes se enfrentan a dificultades para adquirir los conocimientos necesarios y así poder desarrollar la habilidades para manejar las herramientas de minería de datos de forma correcta y efectiva.

Se hace completamente necesario investigar y analizar en profundidad los procesos educativos de nuevas herramientas de minería de datos, establecer los retos y obstáculos comunes para el desarrollo de habilidades, proponer nuevos procedimientos para la obtención de este conocimiento y la aplicación del mismo optimizando el proceso de aprendizaje y su aplicación en el área educativa y laboral.

TRABAJO DE GRADO
Opción Seminario-Diplomado

b. Problemas específicos:

No siempre es fácil para cualquiera comprender la mayoría de los conceptos técnicos que utilizan las herramientas de IA y que es el análisis de datos.

Falta de divulgación de las mismas herramientas de IA y como su funcionamiento de manera general.

La rapidez con la surgen y evolucionan las nuevas herramientas de IA puede llegar a ser un desafío.

Carencia en la capacidad de procesamiento y almacenamiento de datos necesarios para el aprendizaje y aplicación de las IA.

Escasez de habilidades técnicas específicas para la aplicación de las herramientas de IA.

Resistencia al cambio sobre nuevas tecnologías, técnicas y herramientas para explorar y experimentar nuevas tendencias de visualización y análisis de los datos.

Problemas para la integración de los conocimientos previos con los nuevos al momento de hacer nuevos análisis de la información y su aplicabilidad.

TRABAJO DE GRADO
Opción Seminario-Diplomado

4. OBJETIVOS DE INVESTIGACIÓN

a. Objetivo General:

El objetivo general del presente desarrollo es aprender y aplicar nuevas habilidades en la utilización de herramientas de inteligencia artificial para el análisis y visualización de los datos

b. Objetivos Específicos:

Aprender a utilizar herramientas de inteligencia artificial para el análisis de datos, con el fin de mejorar las capacidades de procesamiento e interpretación de la información.

Aprender a aplicar las habilidades adquiridas en la utilización de herramientas de inteligencia artificial para la generación de conocimientos en diversos ámbitos profesionales y académicos.

TRABAJO DE GRADO
Opción Seminario-Diplomado

5. JUSTIFICACIÓN

Los nuevos aprendizajes sobre las herramientas de IA para el análisis y la visualización de datos es un tema de gran importancia en la actualidad por diversos motivos, entre ellos se encuentran:

Aumento en la eficacia debido a este tipo de herramientas automatizan tareas lentas repetitivas y de gran volumen de datos, permitiendo que los analistas puedan enfocarse en desarrollar las estrategias para la aplicación de la información obtenida.

Las nuevas herramientas de IA ayudan a procesar grandes volúmenes de dato de forma mucho mas eficiente, dando como resultado análisis mas precisos y completos, identificando patrones y tendencias que de otra forma se podrían pasar por alto.

Debido a que se obtienen una mayor información y comprensión de la misma, se pueden tomar mejores decisiones, las cuales irán respaldadas de una mejor visualización y análisis de los datos.

Mediante el aprendizaje de estas nuevas herramientas, se pueden mejorar o innovar en nuevos procesos que nos lleven a nuevos descubrimientos.

El aprendizaje de nuevas herramientas de IA puede ser un factor clave para la competitividad de las empresas, los profesionales y los estudiantes sacando el mejor provecho de sus recursos en un mundo cada vez más competitivo.

TRABAJO DE GRADO
Opción Seminario-Diplomado

6. MARCO CONCEPTUAL Y CONTEXTUAL

a. Marco conceptual:

Dentro del presente desarrollo nos centraremos en el aprendizaje de la visualización y análisis de los datos a través de la utilización de la herramienta de inteligencia artificial Orange. Debemos enfatizar en la importancia de adquirir nuevas y mejores habilidades en el manejo y procesamiento de los datos, así como también en la aplicación de técnicas de análisis y visualización de los datos en diferentes contextos.

La relevancia de actividades como la importación de los datos, la limpieza y el procesamiento de los mismos, la aplicación del aprendizaje automático y autónomo, la creación de visualizaciones que sean creativas e interactivas es de vital importancia. En los tiempos en los que ya nos encontramos debemos adquirir conocimientos sobre este tipo de herramientas que son aprovechables en cualquier momento y contexto. Pero también es importante aprender la conceptualización del por qué debemos aprender, qué es y cómo se usan las herramientas para la minería de datos en tiempos de la inteligencia artificial.

b. Contexto Educativo:

Este texto está destinado a un ambiente no solo laboral, sino principalmente educativo, donde podamos desarrollar habilidades para el análisis y

TRABAJO DE GRADO
Opción Seminario-Diplomado

visualización de datos a través de una herramienta de inteligencia artificial. Nos enfocamos en la importancia en el manejo de datos, su análisis, visualización y procesamiento para posteriormente poder realizar una toma de decisiones informada. Se resaltan actividades como la importación de los datos, la limpieza, el procesamiento y la clasificación de la información. Aquí se proporciona una forma de aprendizaje y una visión de cómo las herramientas de inteligencia artificial pueden ayudarnos a mejorar nuestras habilidades y procesos en la toma de decisiones basadas en análisis de base de datos, ayudando así a resolver problemas complejos y específicos en diferentes ámbitos profesionales.

c. Orange:

La herramienta que trabajaremos en este documento es de inteligencia artificial con código abierto, es utilizada para el análisis y visualización de datos, para la construcción y valoración de modelos de machine Learning. Esta herramienta nos permite realizar actividades como la clasificación, regresión, clustering, detección de anomalías, creación de gráficos fáciles de entender que son creativos y efectivos; esta herramienta utiliza una interfaz visual muy fácil de aprender y utilizar.

Esta herramienta es útil para científicos de datos, investigadores y profesionales que deseen analizar grandes cantidades de datos de forma eficiente y rápida, así como para aquellos que deseen crear y probar modelos de machine Learning de manera sencilla, realizando procesos de

TRABAJO DE GRADO
Opción Seminario-Diplomado

programación con Python o para aquellos que no tienen el conocimiento y puedan utilizarlo sin necesidad de programar.

TRABAJO DE GRADO
Opción Seminario-Diplomado

7. METODOLOGÍA

Con el fin de realizar un buen desarrollo del presente texto, se implementaron las siguientes etapas:

1. Análisis de la situación o del problema: lo primero que siempre vamos a necesitar es identificar y definir el problema en cuestión o identificar el conjunto de datos que vamos a analizar y visualizar teniendo en cuenta su finalidad, en esta ocasión con la ayuda de la herramienta de inteligencia artificial Orange.
2. Recolección de datos: para llevar a cabo el análisis, debemos realizar la recolección de los datos, asegurando no solo su calidad y relevancia sino también su confiabilidad y procedencia.
3. Importación de datos en Orange: Importar los datos recopilados en Orange para comenzar con el análisis y visualización.
4. Limpieza y procesamiento de datos: Realizar la limpieza y procesamiento de los datos para eliminar valores atípicos, errores y completar datos faltantes.

TRABAJO DE GRADO
Opción Seminario-Diplomado

5. aprendizaje y aplicación de nuevas habilidades adquiridas: se hará uso de las herramientas y funciones proporcionadas por Orange, para aprender nuevas técnicas de procesamiento de datos.

6. Visualización de datos: se utilizarán las funcionalidades ofrecidas por Orange para presentar la información de manera gráfica, que sea atractiva y fácil de visualizar para su correcto análisis.

7. Análisis de resultados: Analizar los resultados obtenidos a través del análisis y visualización de datos en Orange para extraer conclusiones y tomar decisiones informadas.

Con la presente metodología, se pretende desarrollar de manera estructurada el aprendizaje del análisis y la visualización de datos utilizando la herramienta de inteligencia artificial Orange, teniendo en cuenta que esta herramienta es solo una de muchas que se investigaron y de descartaron para poder incluir en el proceso de aprendizaje a aquellos que no tienen conocimientos en programación.

TRABAJO DE GRADO
Opción Seminario-Diplomado

8. DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE

En el presente documento se realiza el análisis de varios aspectos importantes del objetivo a alcanzar. Se tratarán temas como el alcance y los objetivos del presente documento, la metodología utilizada para realizar el aprendizaje, los resultados obtenidos además de las conclusiones que nos aporte el desarrollo del presente proyecto, además de las recomendaciones que puedan surgir para un mejor aprendizaje y aplicación de herramientas de inteligencia artificial para la toma de decisiones basadas en bases de datos.

El documento se estructura de forma clara y concisa, presentando la información de manera ordenada y fácil de entender. Se utilizan imágenes de los gráficos, tablas y datos utilizados para alcanzar el aprendizaje.

TRABAJO DE GRADO
Opción Seminario-Diplomado

9. APLICACIÓN DE LA HERRAMIENTA DE INTELIGENCIA
ARTIFICIAL: ORANGE

a. Fuentes de extracción

El sector social es una de las áreas que genera grandes volúmenes de información y captura de datos sobre la población colombiana. Por esta razón, para recopilar los datos de estudio del presente proyecto de manera adecuada, se toman en consideración las siguientes fuentes de extracción:

Departamento Administrativo Nacional de Estadística (DANE)

El Ministerio de Salud y Protección Social

Registro Civil

Datos Abiertos de Colombia

Banco Mundial

Centro Nacional de Memoria Histórica

b. Criterios de Selección

Teniendo en cuenta las fuentes mencionadas anteriormente y con el propósito de establecer cuáles y cómo se van a seleccionar las fuentes de información que serán trabajadas, se consideran los siguientes criterios de selección enfocados en la calidad de los datos:

- i. **Credibilidad:** Medida en que él usuario puede utilizar los datos como una entrada de decisión que tenga

TRABAJO DE GRADO
Opción Seminario-Diplomado

validez.(R. Y. Wang et al., 1995) . En la ejecución de los algoritmos de clustering, se identifican patrones con los datos que se están asignando como entrada.

- ii. **Accesibilidad:** Se refiere a la capacidad del usuario de llegar a los datos, teniendo los medios y los privilegios necesarios para obtenerlos (R.Y. Wang, Reddy, & Kon, 1995) Por ejemplo, se solicitó información a Datos Abiertos de Colombia sobre la base de datos de la población registrada en el Valle del Cauca en el último censo y se pudo descargar de la página sin restricciones.

- iii. **Cantidad:** se refiere a la capacidad de almacenar una gran cantidad de datos en una fuente de información para poder realizar el análisis requerido para el proyecto.

- iv. **Precisión:** se refiere a la presencia de detalles suficientes y adecuados en los datos (Oms y Ops, 2014).

TRABAJO DE GRADO
Opción Seminario-Diplomado

c. Definición de variables

Para el desarrollo del presente documento, se tuvieron en cuenta variables de información que no fueran sensibles a comprometer la identidad e integridad de los participantes del último censo en el Valle del Cauca, como la estratificación. Los datos con los cuales se trabajará durante el presente proyecto fueron obtenidos de las fuentes mencionadas.

TRABAJO DE GRADO
Opción Seminario-Diplomado

10. ANÁLISIS Y VISUALIZACIÓN DE LA BASE DE DATOS
MEDIANTE ORANGE

a. Aprendiendo qué es Orange

Orange es una plataforma de código abierto que facilita el análisis y la visualización de datos de manera intuitiva. Con Orange, los usuarios pueden crear flujos de trabajo para la minería de datos utilizando una amplia gama de herramientas visuales. Esta herramienta permite realizar análisis de datos simples con visualizaciones inteligentes, explorando distribuciones estadísticas, diagramas de caja, diagramas de dispersión y diversas técnicas avanzadas como árboles de decisión, agrupación jerárquica, mapas de calor, MDS y proyecciones lineales.

b. Aprendiendo para qué sirve Orange

Orange es una plataforma de código abierto que facilita el análisis y la visualización de datos de manera intuitiva. Con Orange, los usuarios pueden crear flujos de trabajo para la minería de datos utilizando una amplia gama de herramientas visuales. Esta herramienta permite realizar análisis de datos simples con visualizaciones inteligentes, explorando distribuciones

TRABAJO DE GRADO
Opción Seminario-Diplomado

estadísticas, diagramas de caja, diagramas de dispersión y diversas técnicas avanzadas como árboles de decisión, agrupación jerárquica, mapas de calor, MDS y proyecciones lineales.

Orange es una plataforma de código abierto que facilita el análisis y la visualización de datos de manera intuitiva. Con Orange, los usuarios pueden crear flujos de trabajo para la minería de datos utilizando una amplia gama de herramientas visuales. Esta herramienta permite realizar análisis de datos simples con visualizaciones inteligentes, explorando distribuciones estadísticas, diagramas de caja, diagramas de dispersión y diversas técnicas avanzadas como árboles de decisión, agrupación jerárquica, mapas de calor, MDS y proyecciones lineales.

c. Inicio:

Lo primero que haremos será realizar un análisis descriptivo para el procesamiento de la base de datos obtenida del último censo en Colombia para el territorio del Valle del Cauca. Esta información puede encontrarse fácilmente en la plataforma de datos abiertos de Colombia.

Orange es un software de código libre que nos permite hacer minería de datos y contar con alternativas de aplicación de algoritmos para realizar análisis, visualizaciones y/o predicciones.

d. El lienzo:

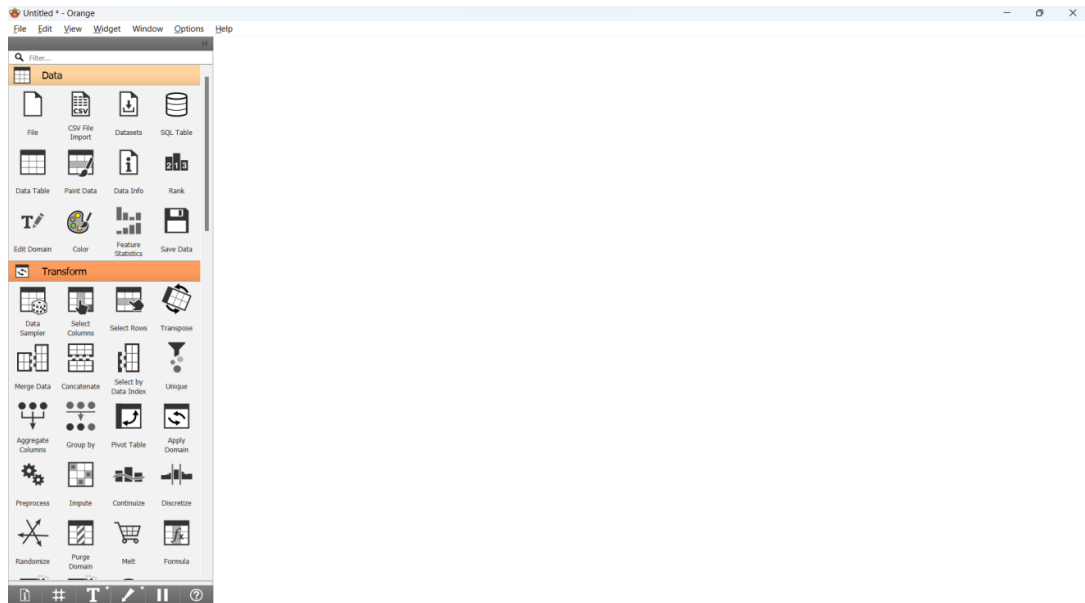
TRABAJO DE GRADO
Opción Seminario-Diplomado

Si observamos en la parte izquierda de la pantalla existen segmentos o tramos en donde encontramos widgets, estos representan las diferentes funciones:

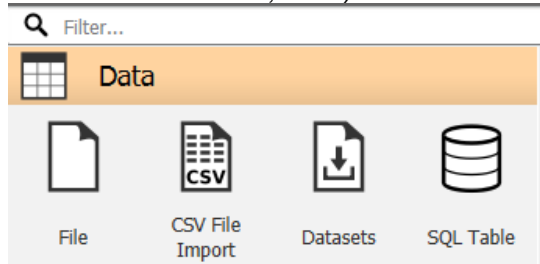
1. Data: es para la importación y procesamiento de datos.
2. Visualización: visualización de los datos importados.
3. Modelos: me permite generar modelos según la necesidad de cada situación.
4. Evaluación: me permite realizar la evaluación de los modelos creados.
5. Aprendizaje sin supervisión: me permite realizar análisis no supervisado.
6. Procesamiento de texto: permite realizar diversas tareas como la limpieza y preprocesamiento de texto, la extracción de características textuales, la clasificación de textos, la visualización de datos textuales
7. Reglas de asociación: son utilizadas para descubrir patrones o relaciones entre diferentes conjuntos de datos.

(figura 1, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

TRABAJO DE GRADO Opción Seminario-Diplomado



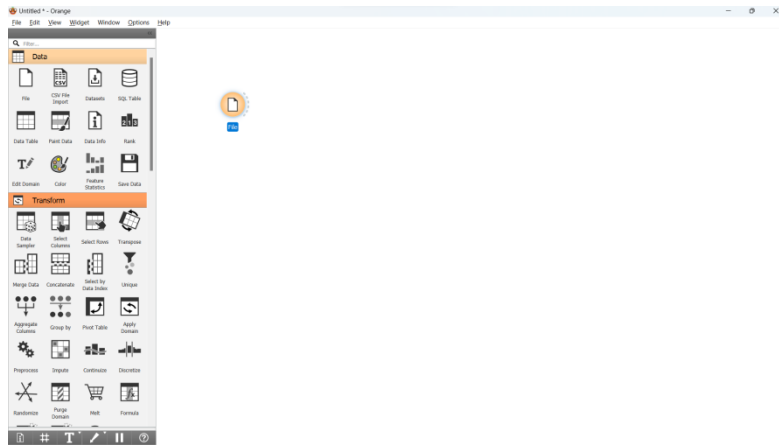
(figura 2, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



En la figura 3 se inicia con la importación de la base de datos:

(figura 3, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

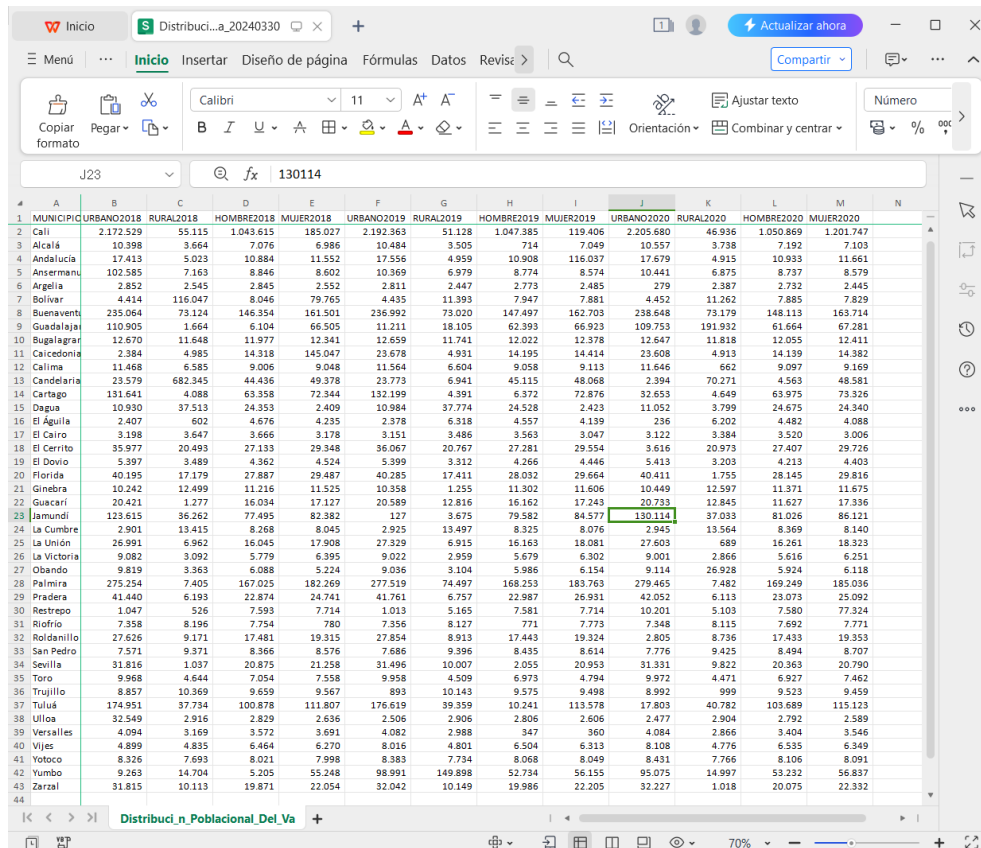
TRABAJO DE GRADO Opción Seminario-Diplomado



En la figura 4 se ve una vista desde Excel de la base de datos que trabajaremos:

(figura 4, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

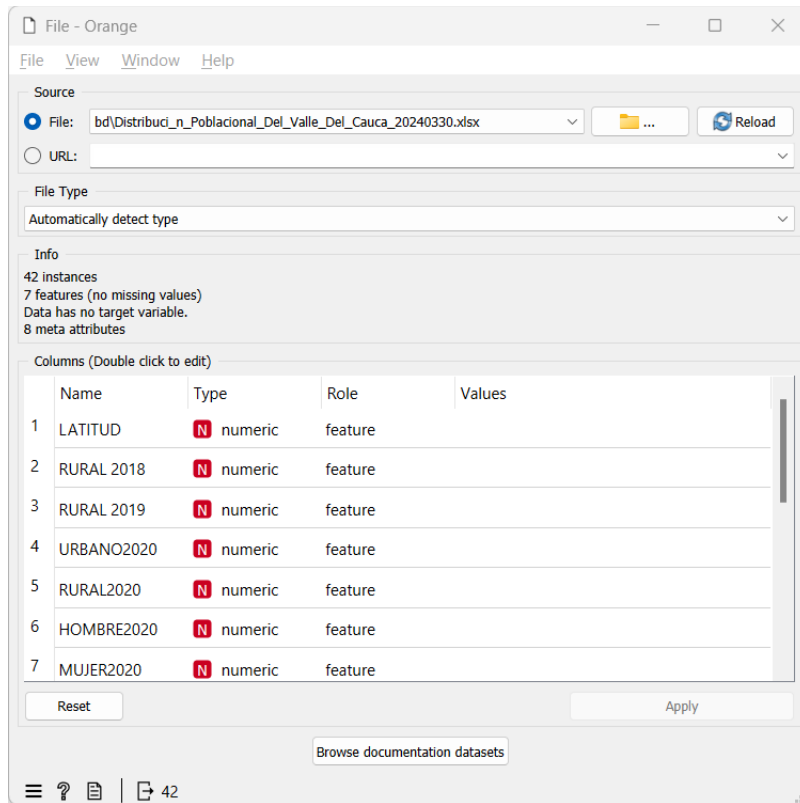
TRABAJO DE GRADO
Opción Seminario-Diplomado



MUNICIPIO	URBANO2018	RURAL2018	HOMBRE2018	MUJER2018	URBANO2019	RURAL2019	HOMBRE2019	MUJER2019	URBANO2020	RURAL2020	HOMBRE2020	MUJER2020
1 Cali	2.172.529	55.115	1.043.615	185.027	2.192.363	51.128	1.047.385	119.406	2.205.680	46.936	1.050.869	1.201.747
3 Alcalá	10.398	3.664	7.076	6.986	10.484	3.505	7.14	7.049	10.557	3.738	7.192	7.103
4 Andalucía	17.413	5.023	10.894	11.552	17.556	4.959	10.908	116.937	17.679	4.915	10.933	11.561
5 Ansermanu	102.585	7.163	8.846	8.602	10.369	6.979	8.774	8.574	10.441	6.875	8.737	8.579
6 Argelia	2.852	2.545	2.845	2.552	2.811	2.447	2.773	2.485	2.79	2.387	2.732	2.445
7 Bolívar	4.414	116.047	8.046	79.765	4.435	11.393	7.947	7.881	4.452	11.262	7.885	7.829
8 Buenaventura	235.064	73.124	146.354	161.501	236.992	73.020	147.497	162.703	238.648	73.179	148.113	163.714
9 Guadalupe	110.905	1.664	6.104	66.505	11.211	18.105	62.393	66.923	109.753	191.932	61.664	67.281
10 Bugalagraz	12.670	11.648	11.977	12.341	12.659	11.741	12.022	12.378	12.647	11.818	12.055	12.411
11 Calcedonia	2.384	4.985	14.318	145.047	23.678	4.931	14.195	14.414	23.608	4.913	14.199	14.382
12 Callina	11.468	6.585	9.006	9.048	11.564	6.604	9.058	9.113	11.646	662	9.097	9.169
13 Candelaria	23.579	682.345	44.436	49.378	23.773	6.941	45.115	48.068	2.394	70.271	4.563	48.581
14 Cartago	131.641	4.088	63.358	72.344	132.199	4.391	6.372	72.876	32.653	4.649	63.975	73.326
15 Dagua	10.930	37.513	24.353	2.409	10.984	37.774	24.528	2.423	11.052	3.799	24.675	24.340
16 El Águila	2.407	602	4.676	4.235	2.378	6.318	4.557	4.139	236	6.202	4.482	4.088
17 El Cairo	3.198	3.647	3.666	3.178	3.151	3.486	3.563	3.047	3.122	3.384	3.520	3.006
18 El Cerrito	35.977	20.493	27.133	29.348	36.067	20.767	27.281	29.554	3.616	20.973	27.407	29.726
19 El Dovio	5.397	3.489	4.362	4.524	5.399	3.312	4.266	4.446	5.413	3.203	4.213	4.403
20 Florida	40.195	17.179	27.887	29.487	40.285	17.411	28.032	29.664	40.411	1.755	28.145	29.816
21 Ginebra	10.242	12.499	11.216	11.525	10.358	1.255	11.302	11.606	10.449	12.597	11.371	11.675
22 Guacarí	20.421	1.277	16.034	17.127	20.589	12.816	16.162	17.243	20.733	12.845	11.627	17.396
23 Jamundí	123.615	36.262	77.495	82.382	127	3.675	79.582	84.577	130.114	37.033	81.026	86.121
24 La Cumbre	2.901	13.415	8.268	8.045	2.925	13.497	8.325	8.076	2.945	13.564	8.369	8.140
25 La Unión	26.991	6.562	16.045	17.908	27.329	6.915	16.163	18.081	27.603	689	16.261	18.323
26 La Victoria	9.082	3.092	5.779	6.395	9.022	2.959	5.679	6.302	9.001	2.866	5.616	6.251
27 Obando	9.819	3.363	6.088	5.224	9.036	3.104	5.986	6.154	9.114	26.928	5.924	6.118
28 Palmira	275.254	7.405	167.025	182.269	277.519	74.497	168.253	183.763	279.465	7.482	169.249	185.036
29 Pradera	41.440	6.193	22.874	24.741	41.761	6.757	22.987	26.931	42.052	6.113	23.073	25.092
30 Restrepo	1.047	526	7.593	7.714	1.013	5.165	7.581	7.714	10.201	5.103	7.580	77.324
31 Riofrio	7.358	8.196	7.754	7.80	7.356	8.127	7.71	7.773	7.348	8.115	7.692	7.771
32 Roldanillo	27.626	9.171	17.481	19.315	27.854	8.913	17.443	19.324	2.805	8.736	17.433	19.353
33 San Pedro	7.571	9.371	8.366	8.576	7.686	9.396	8.435	8.614	7.776	9.425	8.494	8.707
34 Sevilla	31.816	1.037	20.875	21.258	31.496	10.007	2.055	20.953	31.331	9.822	20.363	20.790
35 Toro	9.968	4.644	7.054	7.558	9.958	4.509	6.973	4.794	9.972	4.471	6.927	7.462
36 Trujillo	8.857	10.369	9.659	9.567	893	10.143	9.575	9.498	8.992	999	9.523	9.459
37 Tuludá	174.951	37.734	100.878	111.807	176.619	39.359	10.241	113.578	17.803	40.782	103.689	115.123
38 Ulloa	32.549	2.916	2.829	2.636	2.906	2.806	2.606	2.477	2.904	2.792	2.589	2.589
39 Versalles	4.094	3.169	3.572	3.691	4.082	2.988	347	360	4.084	2.866	3.404	3.546
40 Vijes	4.899	4.835	6.464	6.270	4.801	6.504	6.313	6.313	4.776	6.535	6.349	6.349
41 Yotoco	8.326	7.693	8.021	7.998	8.383	7.734	8.068	8.049	8.431	7.766	8.106	8.091
42 Yumbo	9.263	14.704	5.205	55.248	98.991	149.898	52.734	56.155	95.075	14.997	53.232	56.837
43 Zerral	31.815	10.113	19.871	22.054	32.042	10.149	19.986	22.205	32.227	1.018	20.075	22.332

(figura 5, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

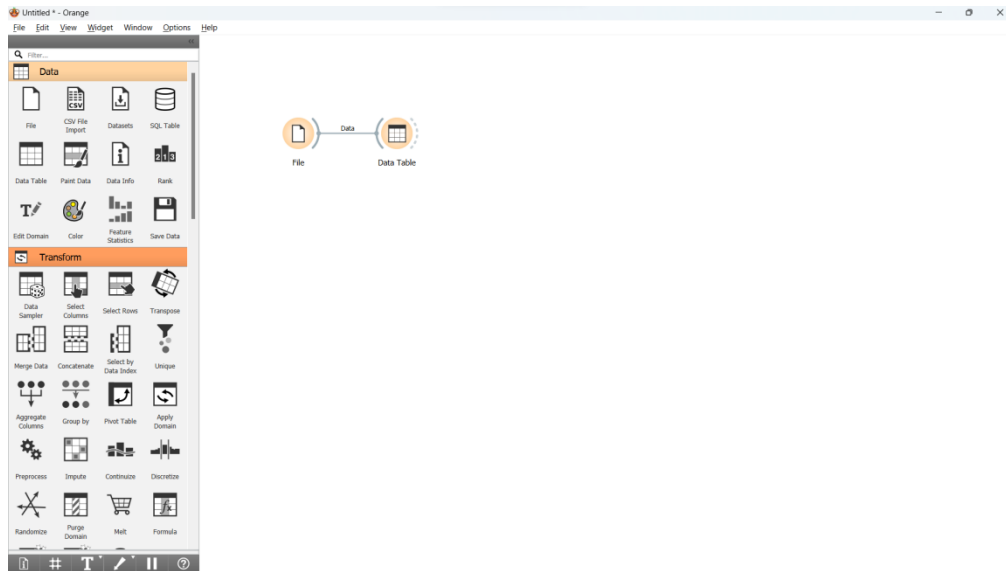
TRABAJO DE GRADO Opción Seminario-Diplomado



Después de realizar la importación, se utilizará una Datatable para visualizar la base de datos como se muestra en la figura 6:

(figura 6, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

TRABAJO DE GRADO Opción Seminario-Diplomado



La figura 7 es la visualización de la base de datos general desde orange:

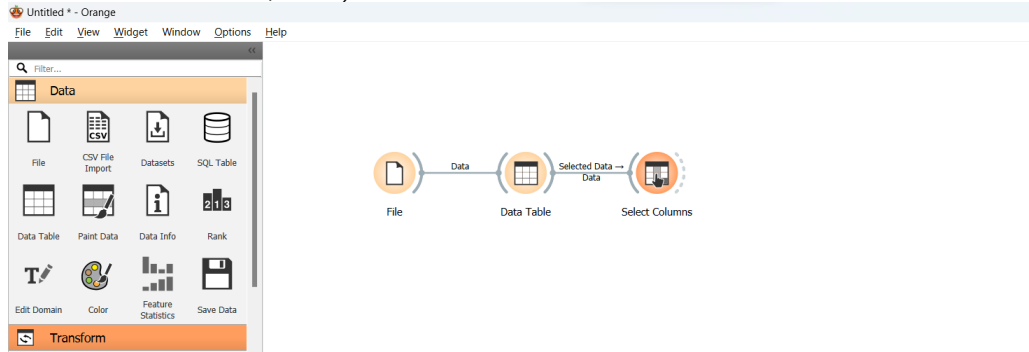
(figura 7, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

	MUNICIPIO	URBANO2018	RURAL2018	HOMBRE2018	MUJER2018	URBANO2019	RURAL2019	HOMBRE2019	MUJER2019	URBANO2020	RURAL2020	HOMBRE2020	MUJER2020
1	...	2172529	55115	1043615	185027	2192363	51128	1047385	119406	2205680	46936	1050869	1201747
2	cañá	10398	3664	7076	6986	10484	3505	714	7049	10557	3738	7192	7103
3	edificia	17413	5023	10684	11552	17556	4959	10908	116037	17679	4915	10933	11661
4	sierranuevo	102525	7163	8846	8602	10389	6979	2774	8574	10441	6875	8737	8579
5	peño	2852	2545	2845	2552	2811	2447	2773	2485	279	2387	2732	2445
6	álar	4414	116047	8046	79765	4435	11393	7947	7881	4452	11262	7885	7829
7	enaventura	239064	73124	146354	161501	236992	73020	147497	162703	238648	73179	148113	163714
8	adajajara de ...	110905	1664	6104	66505	11211	18105	62393	66923	109753	191932	61664	67281
9	gatagrande	12670	11648	11977	12341	12659	11741	12022	12378	12647	11818	12055	12411
10	icedonia	2384	4985	14318	145047	23678	4931	14195	14414	23608	4913	14139	14382
11	lima	11468	6585	9606	9048	11564	6604	9058	9113	11646	662	9097	9169
12	indolara	23579	68245	44436	49378	23773	6941	45115	48068	2394	70271	4565	48581
13	itago	131641	4088	63358	72344	132199	4391	6372	72876	32653	4649	63975	73326
14	igua	10930	37513	24553	2409	10984	37774	24528	2423	11052	3799	24675	24340
15	Aguala	2407	602	4676	4235	2378	6318	4557	4139	236	6202	4482	4088
16	Cairo	3198	3647	3666	3178	3151	3486	3563	3047	3122	3384	3520	3006
17	Cerrito	35977	20493	27133	29348	36067	20767	27281	29554	3616	20973	27407	29726
18	Dovio	5397	3489	4362	4524	5399	3312	4266	4446	5413	3203	4213	4403
19	iriba	40195	11719	27887	29487	40285	13411	28032	29654	40411	1755	28145	28816
20	niebra	10242	12499	11216	11525	10358	1255	11302	11606	10449	12597	11371	11675
21	acari	20421	1277	16034	17127	20589	12816	16162	17243	20733	12845	11627	17336
22	mundí	123615	36262	77495	82382	127	3675	79582	84577	130114	37033	81026	86121
23	Cumbre	2901	13415	8268	8045	2925	13497	8325	8076	2945	13564	8369	8140
24	Unión	26991	6962	16045	17908	27329	6915	16163	18081	27603	689	16261	18323
25	Victoria	9082	3092	5779	6395	9022	2959	5679	6302	9001	2866	5616	6251
26	ando	9819	3363	6088	5224	9036	3104	5986	6154	9114	26928	5924	6118
27	lima	275254	7465	167025	182059	27519	74497	168253	183763	279455	7462	169249	185036
28	iderra	41440	6193	22874	24741	41761	6757	22987	26931	42052	6113	23073	25092
29	strepo	1047	526	7593	7714	1013	5165	7581	7714	10201	5103	7580	77324
30	ñico	7358	8196	7754	780	7356	8127	771	7773	7348	8115	7692	7771
31	idanillo	27626	9171	17481	19315	27854	8913	17443	19324	2805	8736	17433	19353
32	n Pedro	7571	9371	8366	8576	7686	9396	8435	8614	7776	9425	8494	8707
33	villa	31816	1037	20875	21258	31496	10007	2055	20953	31331	9822	20363	20790
34	io	9968	4644	7054	7558	9958	4509	6973	4794	9972	4471	6927	7462
35	ñico	8857	10369	9659	9567	893	10143	9575	9498	8992	999	9523	9459
36	lúa	174951	37734	100878	111807	176619	39359	10241	113578	17803	40782	103689	115123
37	loa	32549	2916	2829	2636	2506	2906	2806	2606	2477	2904	2792	2589

En la figura 8 se utiliza un Select column para empezar la limpieza de los datos con los que se va a trabajar.

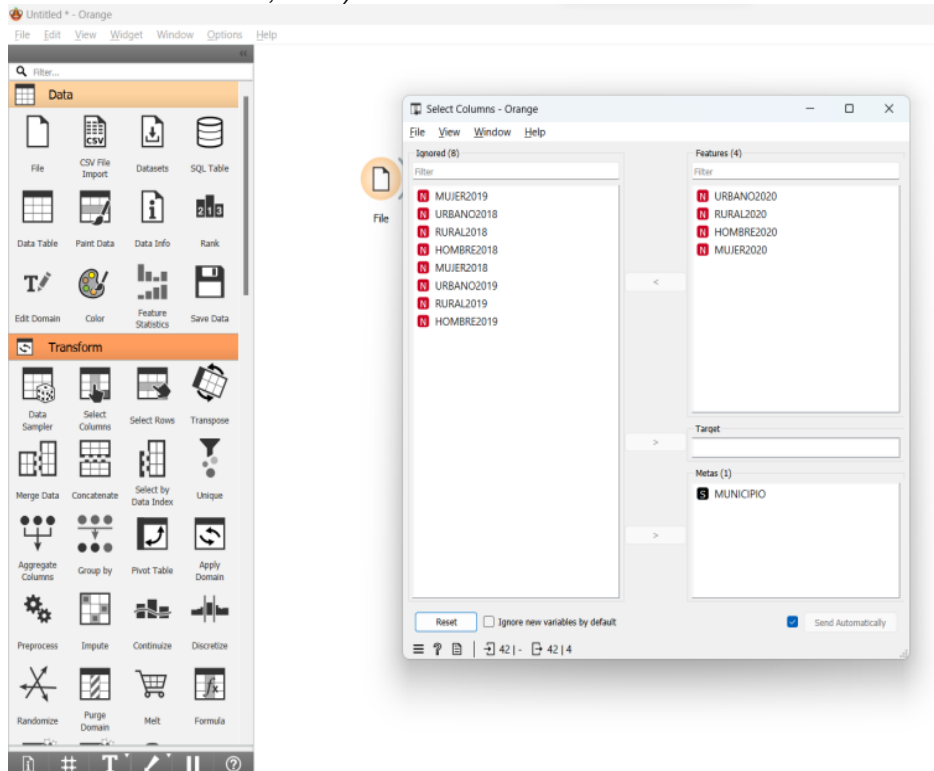
TRABAJO DE GRADO Opción Seminario-Diplomado

(figura 8, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



En la figura 9 se seleccionan las variables con las que se van a trabajar:

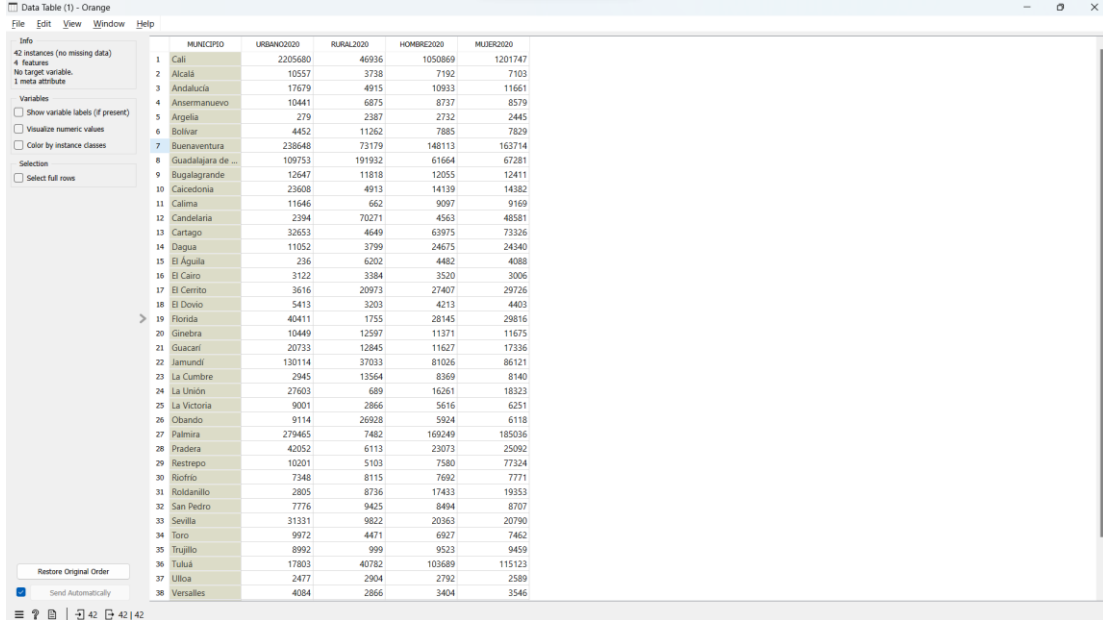
(figura 9, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



En la figura 10 se selecciona otro Datatable para visualizar la base de datos:

TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 10, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



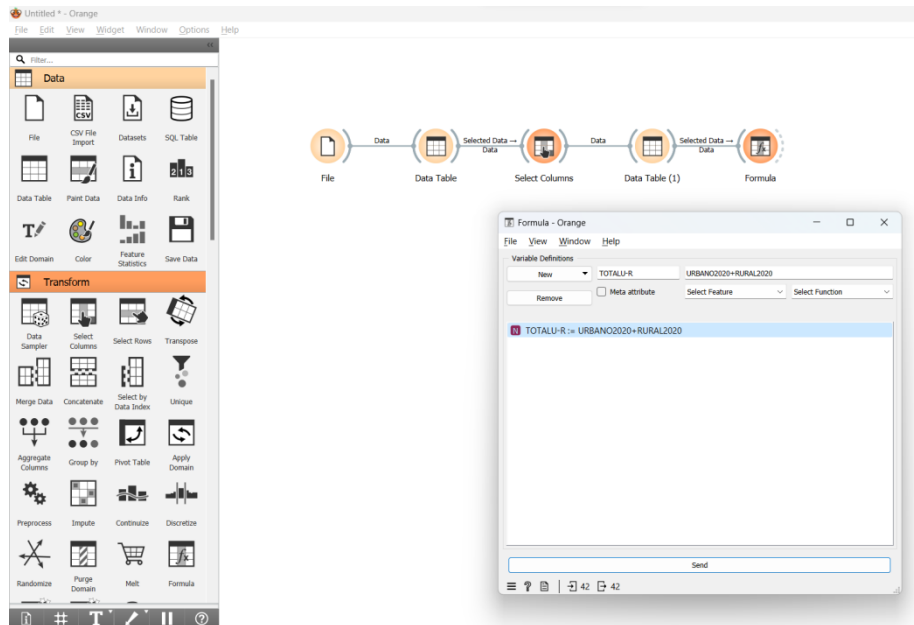
MUNICIPIO	URBANO2020	RURAL2020	HOMBRE2020	MUJER2020
1 Cali	2205600	46936	1050069	1201747
2 Alcalá	10557	3738	7192	7103
3 Andalucía	17679	4915	10993	11661
4 Ansermanuevo	10441	6875	8737	8579
5 Argelia	279	2387	2732	2445
6 Bolívar	4452	11262	7885	7829
7 Buenaventura	238648	73179	148113	163714
8 Guadalupe de ...	109753	191932	61664	67281
9 Bugalagrande	12647	11818	12055	12411
10 Calcedonia	23608	4913	14139	14382
11 Calima	11646	662	9097	9169
12 Candelaria	2394	70271	4563	48581
13 Cartago	32653	4649	63975	73326
14 Dagua	11052	3799	24675	24340
15 El Águila	236	6202	4482	4088
16 El Cairo	3122	3384	3520	3006
17 El Cerrito	3616	20973	27407	29726
18 El Dovio	5413	3203	4213	4403
19 Florida	40411	1755	28145	29816
20 Ginebra	10449	12597	11371	11675
21 Guacarí	20733	12845	11627	17336
22 Jamundí	130114	37033	81026	86121
23 La Cumbre	2945	13564	8369	8140
24 La Unión	27603	689	16261	18323
25 La Victoria	9001	2866	5616	6251
26 Obando	9114	26928	5924	6118
27 Palmira	279465	7482	169249	185036
28 Pradera	42052	6113	23073	25092
29 Restrepo	10201	5103	7580	77324
30 Rofrío	7348	8115	7692	7771
31 Roldanillo	2805	8736	17433	19353
32 San Pedro	7776	9425	8494	8707
33 Sevilla	31331	9822	20363	20790
34 Toro	9972	4471	6927	7462
35 Trujillo	8992	999	9523	9459
36 Tulúá	17803	40782	103689	115123
37 Ulloa	2477	2904	2792	2589
38 Versalles	4084	2866	3404	3546

Como se puede ver en la imagen superior, solo se encuentran las variables que se seleccionaron anteriormente.

En la figura 11 se agrega un widget de fórmula, esto para poder construir otra variable que no está en la base de datos, en este caso será un campo numérico:

(figura 11, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

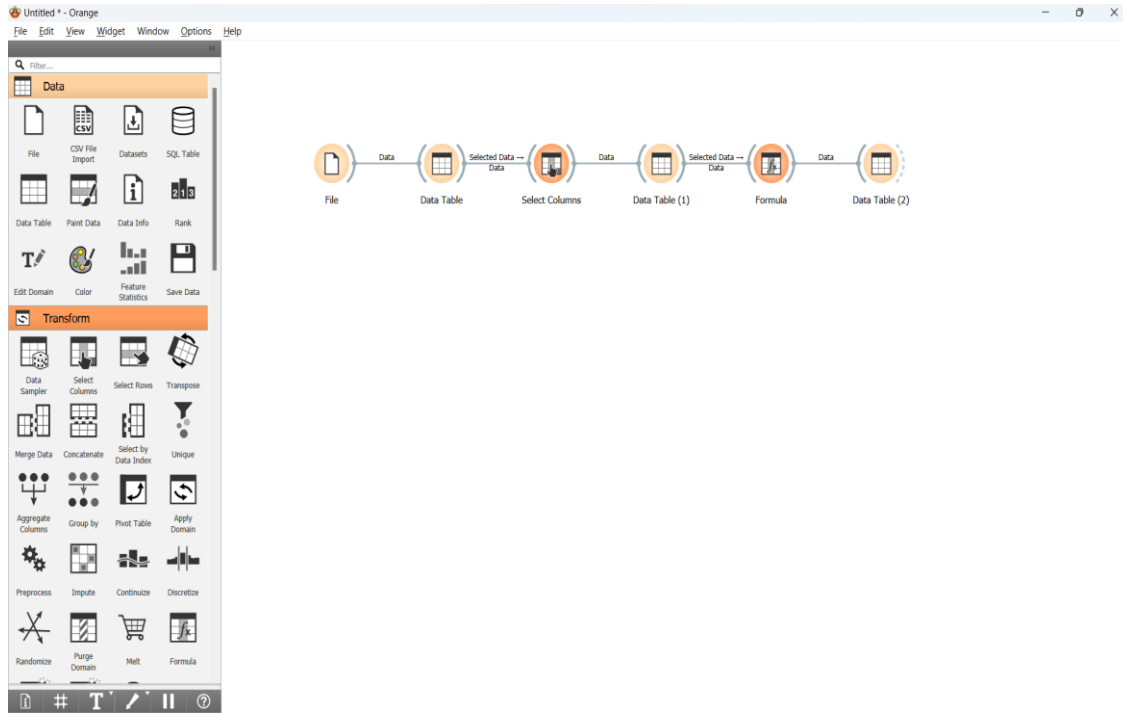
TRABAJO DE GRADO Opción Seminario-Diplomado



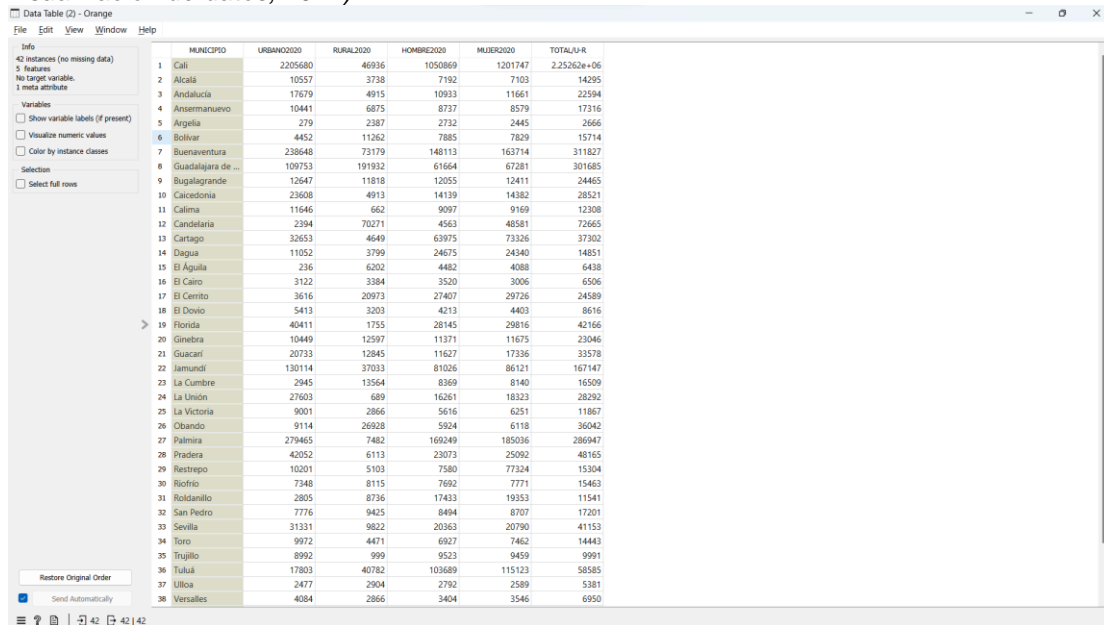
En la figura 12 y 13, nuevamente se realiza la visualización con data table de cómo va el procesamiento de la información:

(figura 12, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

TRABAJO DE GRADO
Opción Seminario-Diplomado



(figura 13, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



The screenshot shows the 'Data Table (2) - Orange' window. The table contains demographic data for various municipalities. The columns are: MUNICIPIO, URBANO2020, RURAL2020, HOMBRE2020, MUJER2020, and TOTAL/U-R. The rows are numbered 1 to 38.

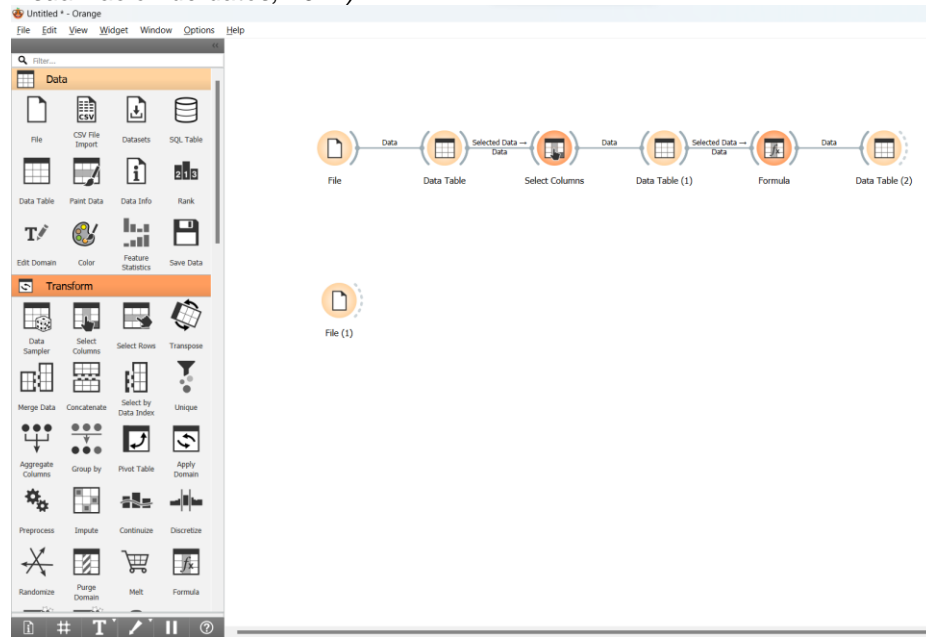
	MUNICIPIO	URBANO2020	RURAL2020	HOMBRE2020	MUJER2020	TOTAL/U-R
1	Cali	2205680	46936	1050869	1201747	2.25262e+06
2	Alcalá	10557	3738	7192	7103	14295
3	Andalucía	17679	4815	10933	11661	22594
4	Ansomannuevo	10441	6875	8737	8579	17316
5	Argelia	279	2387	2732	2445	2666
6	Bolívar	4452	11262	7885	7829	15714
7	Buenaventura	238648	73179	148113	163714	311827
8	Guadalajara de ...	109753	191932	61664	67281	301685
9	Bugalagrande	12647	11818	12055	12411	24465
10	Caicedonia	23608	4913	14139	14382	28521
11	Callima	11646	662	9097	9169	12308
12	Candelaria	2394	7071	4563	48581	72665
13	Cartago	33653	4649	63975	73326	37302
14	Dagua	11052	3799	24675	24340	14851
15	El Águila	236	6202	4482	4088	6438
16	El Cairo	3122	3384	3520	3006	6506
17	El Cerrito	3616	20973	27407	29726	24589
18	El Dovio	5413	3203	4213	4403	8616
19	Florida	40411	1755	28145	29816	42166
20	Ginebra	10449	12597	11371	11675	23046
21	Guacarí	20733	12845	11627	17336	33578
22	Jamundí	130114	37033	81026	86121	167147
23	La Cumbre	2945	13554	8369	8140	16509
24	La Unión	27603	689	16261	18323	28292
25	La Victoria	9001	2866	5616	6251	11867
26	Obando	9114	26928	5924	6118	36042
27	Palmira	279465	7482	169249	185036	286947
28	Pradera	42052	6113	23073	25092	48165
29	Restrepo	10201	5103	7580	77324	15304
30	Riofrío	7348	8115	7692	7771	15463
31	Roldanillo	2805	8736	17433	19353	11541
32	San Pedro	7776	9425	8484	8707	17201
33	Sovilla	31331	9822	20563	20790	41153
34	Toro	9972	4471	6927	7462	14443
35	Trujillo	8992	999	9533	9459	9991
36	Tulúa	17803	40782	103689	115123	58585
37	Ulloa	2477	2904	2792	2589	5381
38	Versalles	4084	2866	3404	3546	6950

11. TRAER OTRA BASE DE DATOS PARA COMPARAR

TRABAJO DE GRADO
Opción Seminario-Diplomado

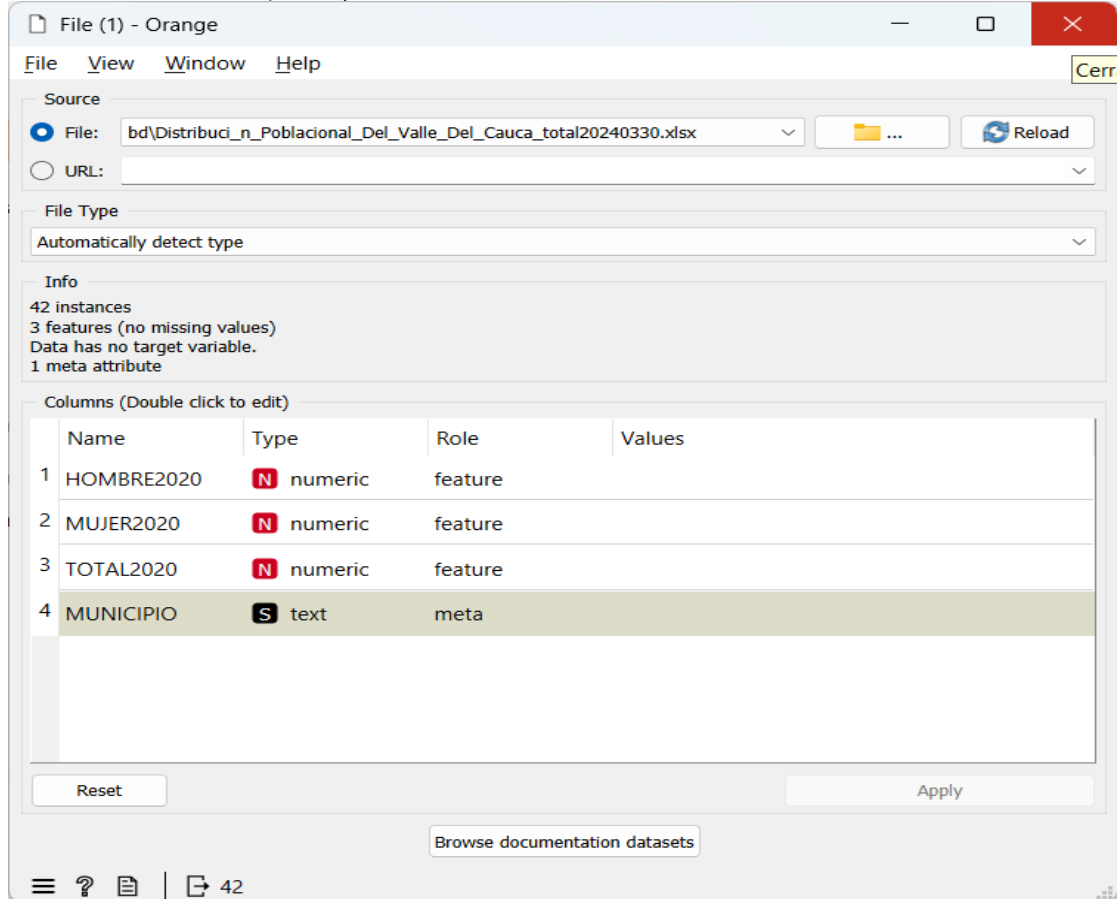
En la figura 14 y 15 se traen otra base de datos para realizar la comparación de que se desea entre datos

(figura 14, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 15, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



The screenshot shows the 'File (1) - Orange' window. The 'Source' section has 'File' selected with the path 'bd\Distribuci_n_Poblacional_Del_Valle_Del_Cauca_total20240330.xlsx'. The 'File Type' is set to 'Automatically detect type'. The 'Info' section shows: 42 instances, 3 features (no missing values), Data has no target variable, and 1 meta attribute. The 'Columns' table is as follows:

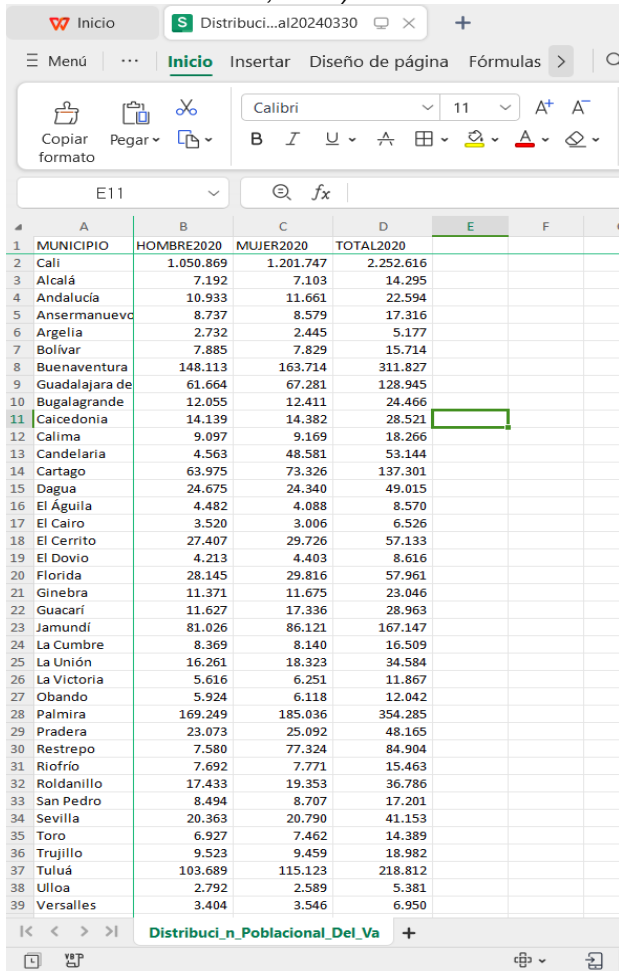
	Name	Type	Role	Values
1	HOMBRE2020	N numeric	feature	
2	MUJER2020	N numeric	feature	
3	TOTAL2020	N numeric	feature	
4	MUNICIPIO	S text	meta	

Buttons for 'Reset', 'Apply', and 'Browse documentation datasets' are visible at the bottom.

TRABAJO DE GRADO
Opción Seminario-Diplomado

En la figura 16 se visualiza la segunda base de datos que integraremos al análisis:

(figura 16, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

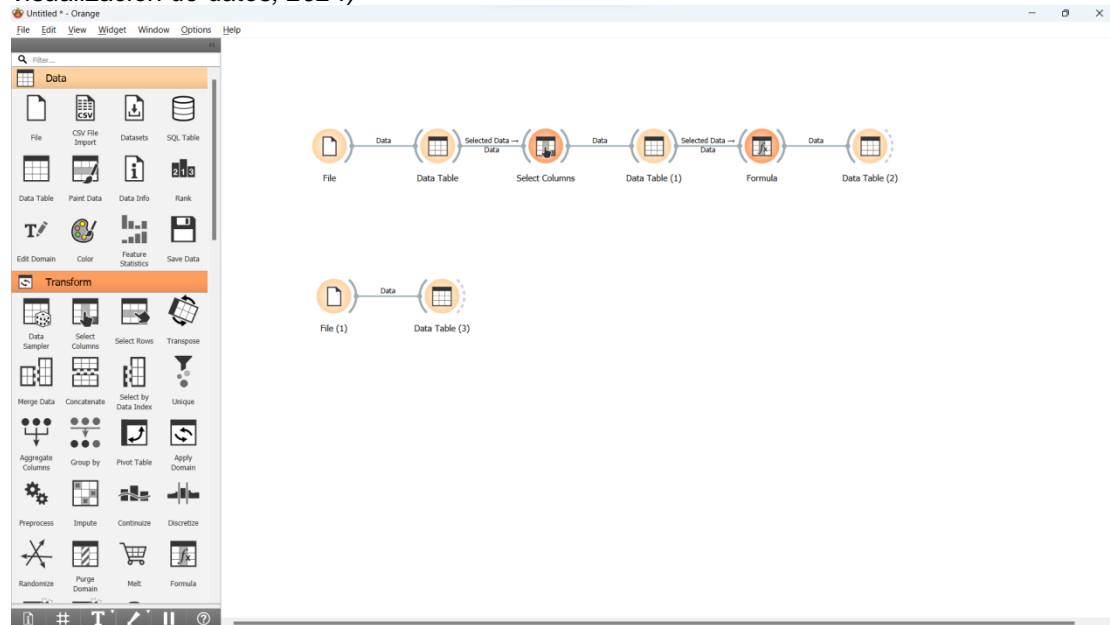


#	A	B	C	D	E	F	G
	MUNICIPIO	HOMBRE2020	MUJER2020	TOTAL2020			
1	Cali	1.050.869	1.201.747	2.252.616			
2	Alcalá	7.192	7.103	14.295			
3	Andalucía	10.933	11.661	22.594			
4	Ansermanuevc	8.737	8.579	17.316			
5	Argelia	2.732	2.445	5.177			
6	Bolívar	7.885	7.829	15.714			
7	Buenaventura	148.113	163.714	311.827			
8	Guadalajara de	61.664	67.281	128.945			
9	Bugalagrande	12.055	12.411	24.466			
10	Caicedonia	14.139	14.382	28.521			
11	Calima	9.097	9.169	18.266			
12	Candelaria	4.563	48.581	53.144			
13	Cartago	63.975	73.326	137.301			
14	Dagua	24.675	24.340	49.015			
15	El Águila	4.482	4.088	8.570			
16	El Cairo	3.520	3.006	6.526			
17	El Cerrito	27.407	29.726	57.133			
18	El Dovio	4.213	4.403	8.616			
19	Florida	28.145	29.816	57.961			
20	Ginebra	11.371	11.675	23.046			
21	Guacarí	11.627	17.336	28.963			
22	Jamundí	81.026	86.121	167.147			
23	La Cumbre	8.369	8.140	16.509			
24	La Unión	16.261	18.323	34.584			
25	La Victoria	5.616	6.251	11.867			
26	Obando	5.924	6.118	12.042			
27	Palmira	169.249	185.036	354.285			
28	Pradera	23.073	25.092	48.165			
29	Restrepo	7.580	77.324	84.904			
30	Riofrio	7.692	7.771	15.463			
31	Roldanillo	17.433	19.353	36.786			
32	San Pedro	8.494	8.707	17.201			
33	Sevilla	20.363	20.790	41.153			
34	Toro	6.927	7.462	14.389			
35	Trujillo	9.523	9.459	18.982			
36	Tuluá	103.689	115.123	218.812			
37	Ulloa	2.792	2.589	5.381			
38	Versalles	3.404	3.546	6.950			
39							

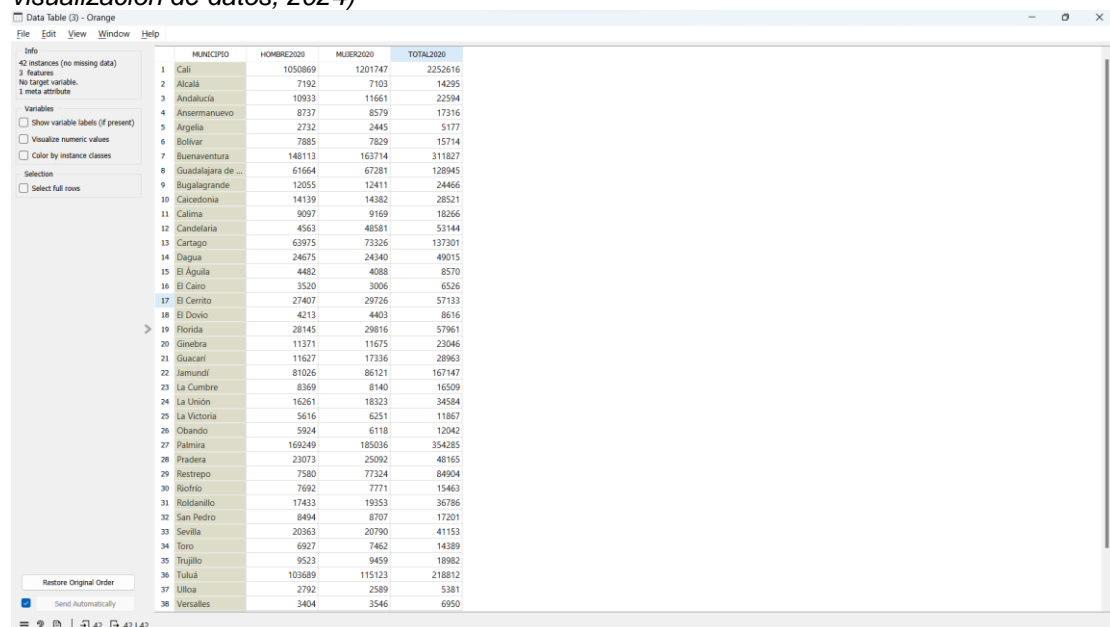
TRABAJO DE GRADO
Opción Seminario-Diplomado

En la figura 17 y 18 se visualiza con data table

(figura 17, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 18, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



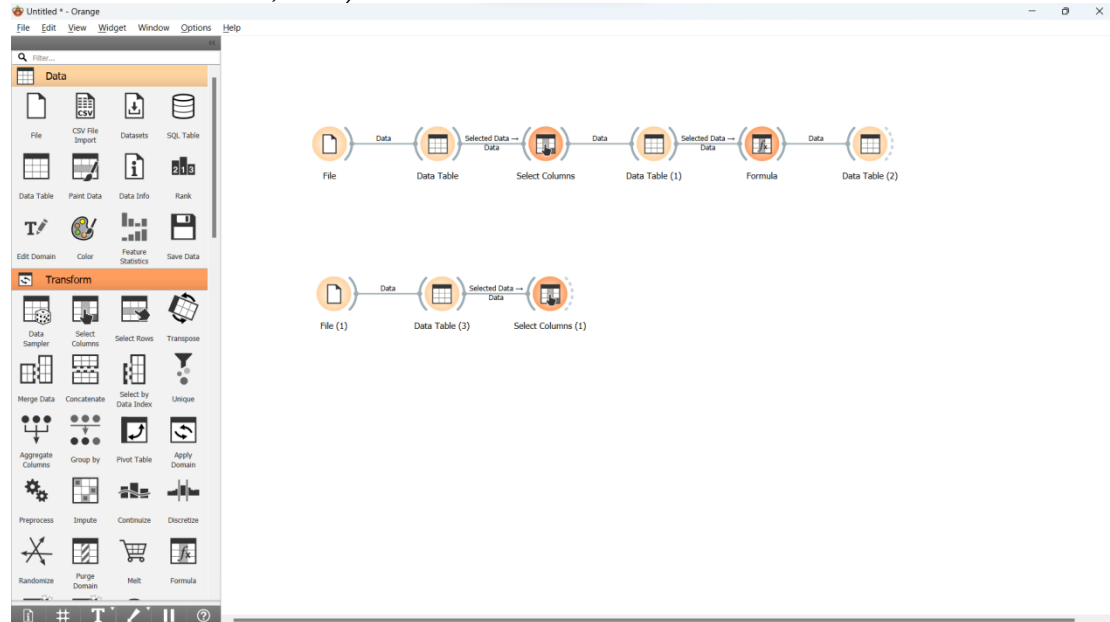
The screenshot shows a data table widget in Orange software. The table has 42 instances and 3 features. The data is as follows:

	MUNICIPIO	HOMBRE2020	MUJER2020	TOTAL2020
1	Cali	1050869	1201747	2252616
2	Alcalá	7192	7103	14295
3	Andalucía	10933	11661	22594
4	Ansermanuevo	8737	8579	17316
5	Argelia	2732	2445	5177
6	Bolívar	7885	7829	15714
7	Buenaventura	148113	163714	311827
8	Guadalajara de...	61664	67281	128945
9	Bugalagrande	12055	12411	24466
10	Caicedonia	14139	14382	28521
11	Calmo	9097	9169	18266
12	Candelaria	4563	48581	53144
13	Cartago	63975	73326	137301
14	Dagua	24675	24340	49015
15	El Águila	4482	4088	8570
16	El Cairo	3520	3006	6526
17	El Cerrito	27407	29726	57133
18	El Dovio	4213	4403	8616
19	Florida	28145	29816	57961
20	Ginebra	113711	11675	23046
21	Guacarí	11627	17336	28963
22	Jamundí	81026	86121	167147
23	La Cumbre	8369	8140	16509
24	La Unión	16261	18323	34584
25	La Victoria	5616	6251	11867
26	Obando	5924	6118	12042
27	Palmira	169249	185036	354285
28	Pradera	23073	25092	48165
29	Restrepo	7580	7724	15304
30	Rafiro	7692	7771	15463
31	Roldanillo	17433	19353	36786
32	San Pedro	8494	8707	17201
33	Sevilla	20363	20790	41153
34	Toro	6927	7462	14389
35	Trujillo	9523	9459	18982
36	Tuluá	103689	115123	218812
37	Ulloa	2792	2589	5381
38	Versalles	3404	3546	6950

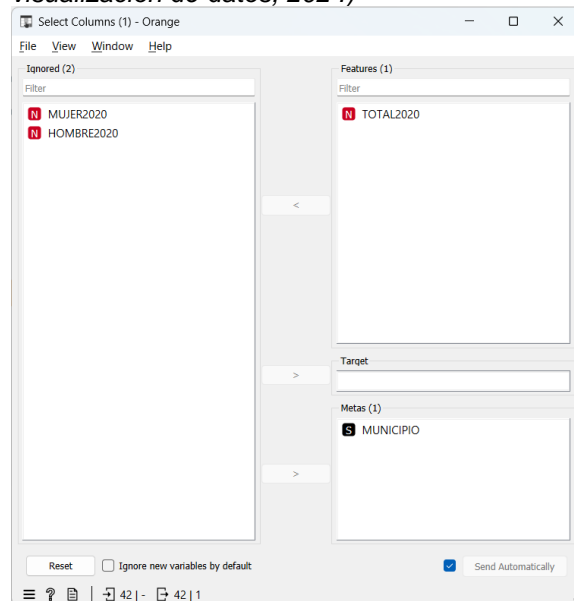
TRABAJO DE GRADO Opción Seminario-Diplomado

En la figura 19, 20 y 21 se va a realizar la selección de las columnas con las que se desea trabajar

(figura 19, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 20, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 21, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024

Data Table (4) - Orange

File Edit View Window Help

Info

42 instances (no missing data)
1 feature
No target variable.
1 meta attribute

Variables

Show variable labels (if present)

Visualize numeric values

Color by instance classes

Selection

Select full rows

	MUNICIPIO	TOTAL2020
1	Cali	2252616
2	Alcalá	14295
3	Andalucía	22594
4	Ansermanuevo	17316
5	Argelia	5177
6	Bolívar	15714
7	Buenaventura	311827
8	Guadalajara de ...	128945
9	Bugalagrande	24466
10	Caicedonia	28521
11	Calima	18266
12	Candelaria	53144
13	Cartago	137301
14	Dagua	49015
15	El Águila	8570
16	El Cairo	6526
17	El Cerrito	57133
18	El Dovio	8616
19	Florida	57961
20	Ginebra	23046
21	Guacarí	28963
22	Jamundí	167147
23	La Cumbre	16509
24	La Unión	34584
25	La Victoria	11867
26	Obando	12042
27	Palmira	354285
28	Pradera	48165
29	Restrepo	84904
30	Riofrío	15463
31	Roldanillo	36786
32	San Pedro	17201
33	Sevilla	41153
34	Toro	14389
35	Trujillo	18982
36	Tuluá	218812
37	Ulloa	5381
38	Versalles	6950

Restore Original Order

Send Automatically

☰ ? 📄 | ↶ 42 ↷ 42 | 42

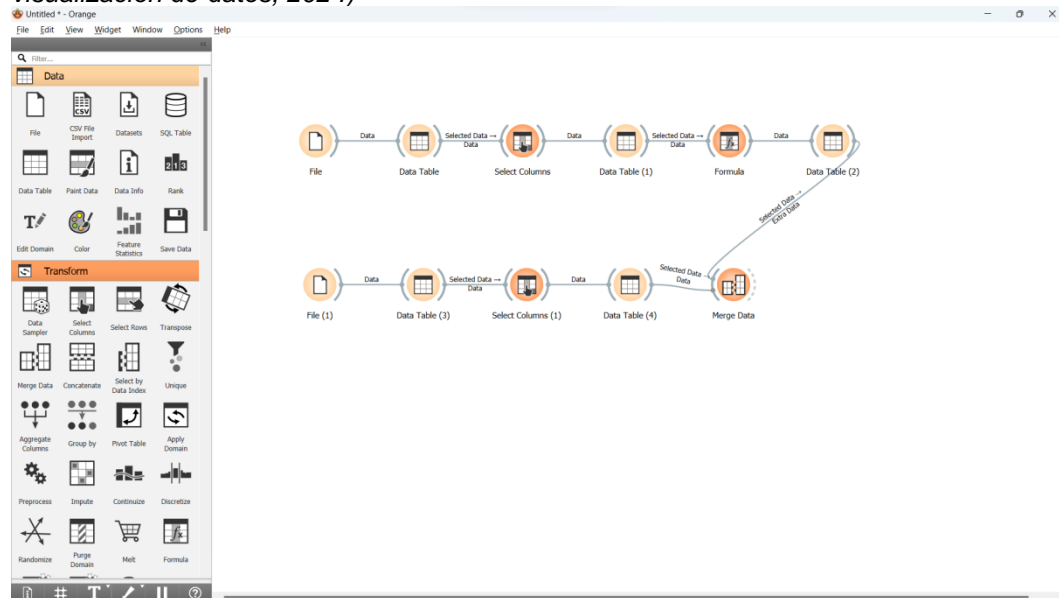
)

TRABAJO DE GRADO
Opción Seminario-Diplomado

**12. INTEGRACIÓN DE AMBAS BASES DE DATOS PARA PODER
REALIZAR EL COMPARATIVO ENTRE AMBAS**

Para realizar esta acción, contamos en el widget Marge Data, el cual me realizará la integración de los datos escogidos, lo que se muestra en la figura 22:

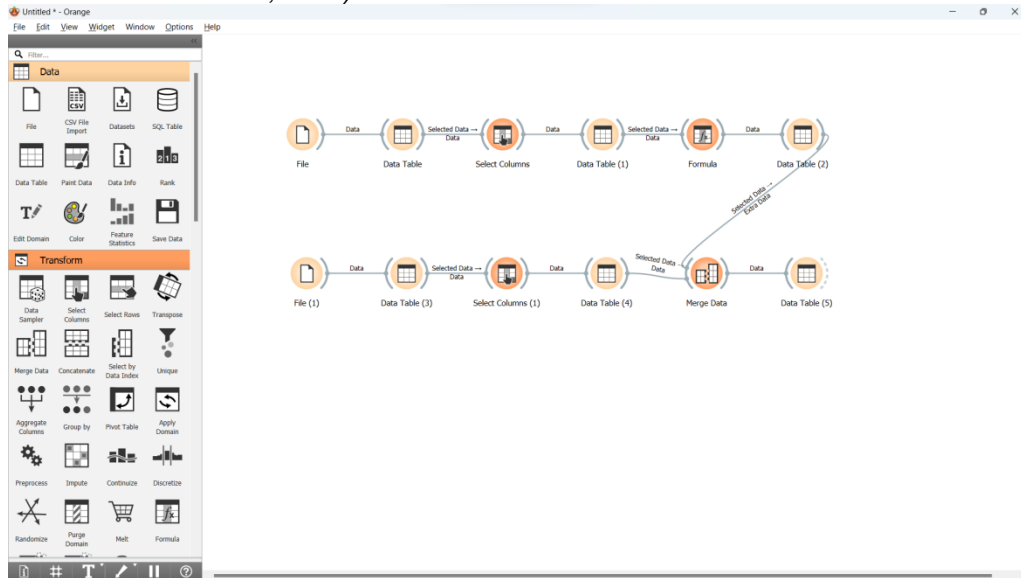
(figura 22, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



En la figura 23 se asegura que la integración de la información sea el que se necesitaba, por lo que se vuelve a realizar una visualización con data table:

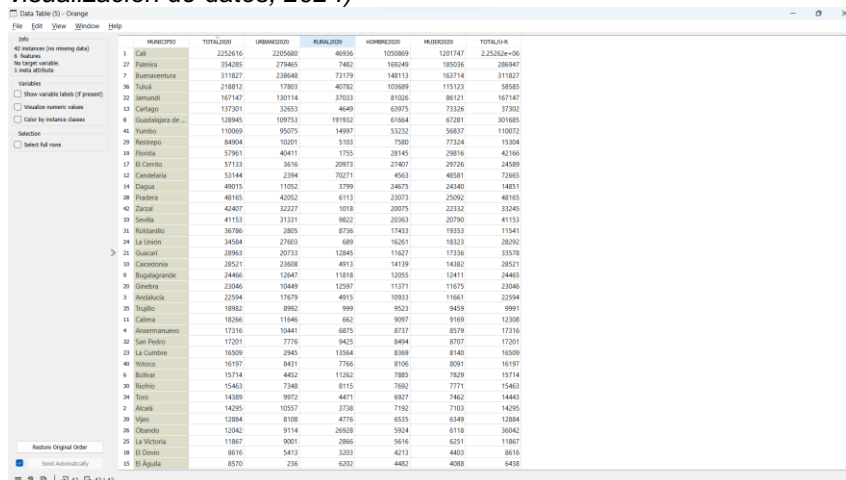
TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 23, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



En la figura 24 se ve cómo se suma a las demás columnas el total 2020 que se encontraba en la segunda base de datos, aquí se podrá comparar esta columna con el total/u-r. Información que debería ser igual para ser correcta.

(figura 24, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



MUNICIPIO	TOTAL 2020	URBANO 2020	RURAL 2020	HOMBRES 2020	MUJERES 2020	TOTAL U-R
1 Cali	2252316	2205660	46556	1050069	1201247	2,250216=406
27 Palmira	354285	279465	74820	160249	185036	389847
7 Buenaventura	311827	238648	73179	148113	163714	311827
6 Talulá	218812	17803	40782	105069	110123	30565
22 Jamundí	167147	130114	37033	81036	86121	167147
13 Cartago	137301	32653	4649	63975	73326	37302
4 Guadalupe de...	120845	109753	19932	61664	67281	307685
41 Venecia	119069	99575	19697	53532	56937	110052
29 Restrepo	84904	10201	5103	7580	77324	15304
19 Florida	57961	40411	1755	28145	29816	42166
17 El Cerrito	57133	3616	20073	27407	29726	24809
12 Candelería	53144	2394	70271	4563	48581	72665
14 Dagua	49015	11062	3799	24675	24340	14851
28 Pradera	48165	42062	6113	23073	25092	48165
42 Zarzal	42407	32227	1018	20075	22332	33245
30 Sevilla	41153	31331	9822	20963	20790	41153
21 Riobanlo	38786	2865	8736	17463	19353	11561
24 La Unión	34584	27603	689	16261	18323	28292
21 Guacarí	28963	20733	12345	11627	17336	33578
10 Calcedonia	28521	23668	4853	14139	14382	28521
9 Buenaprenda	24466	12647	11818	12055	12411	24466
30 Genovea	23046	10489	12597	11371	11675	23046
3 Andalucía	22594	17679	4915	10933	11661	22594
30 Tugalo	19362	8962	999	9523	9839	9991
11 Calima	18266	11646	662	9097	9169	12308
4 Ansermanuevo	17216	10461	6875	8737	8579	17316
12 San Pedro	17201	7776	9425	8484	8717	17201
23 La Cumbre	16509	2545	13564	8369	8140	16509
40 Volcavo	16197	8431	7766	8106	8091	16197
4 Bellostr	15714	4462	11262	7883	7829	15714
30 Riochío	15463	7348	8115	7692	7771	15463
30 Tiro	14389	9972	4417	6927	7462	14483
2 Alcañes	14295	10527	3768	7182	7103	14295
39 Vajó	13284	8108	4776	6335	6349	12884
28 Obando	12042	9114	2628	5924	6118	36042
18 La Victoria	11867	9061	2886	5616	6251	11867
18 El Envoyé	8916	5413	3303	4213	4403	8916
15 El Águila	8570	236	6302	4482	4088	6438

TRABAJO DE GRADO
Opción Seminario-Diplomado

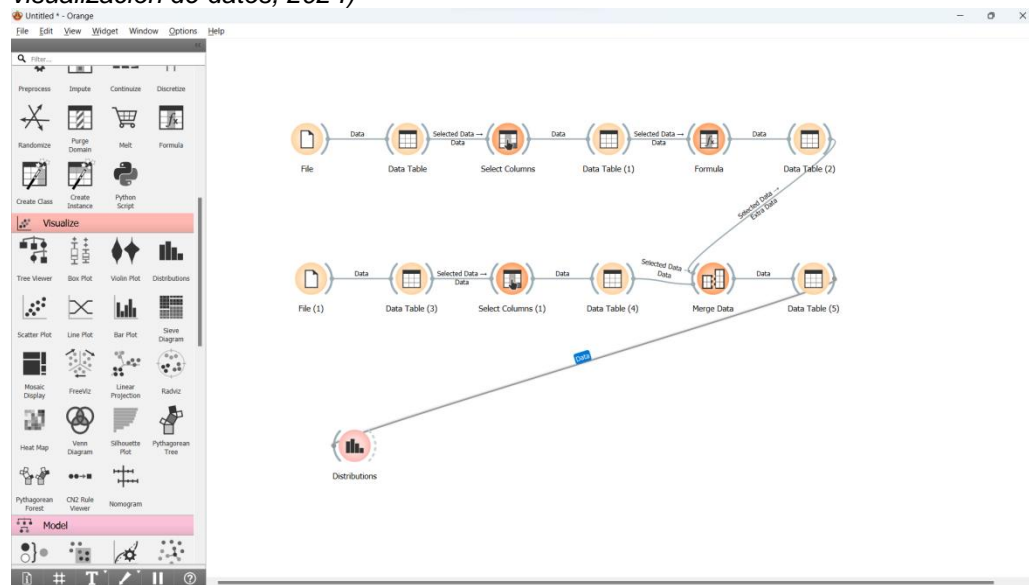
13. INTEGRACIÓN DE GRÁFICOS AL PROCESO DE DATOS:

Ahora se puede empezar a visualizar el análisis de la información con varios tipos de gráficos y todo dependiendo de la información que se requieren.

a. Visualization Por distribution:

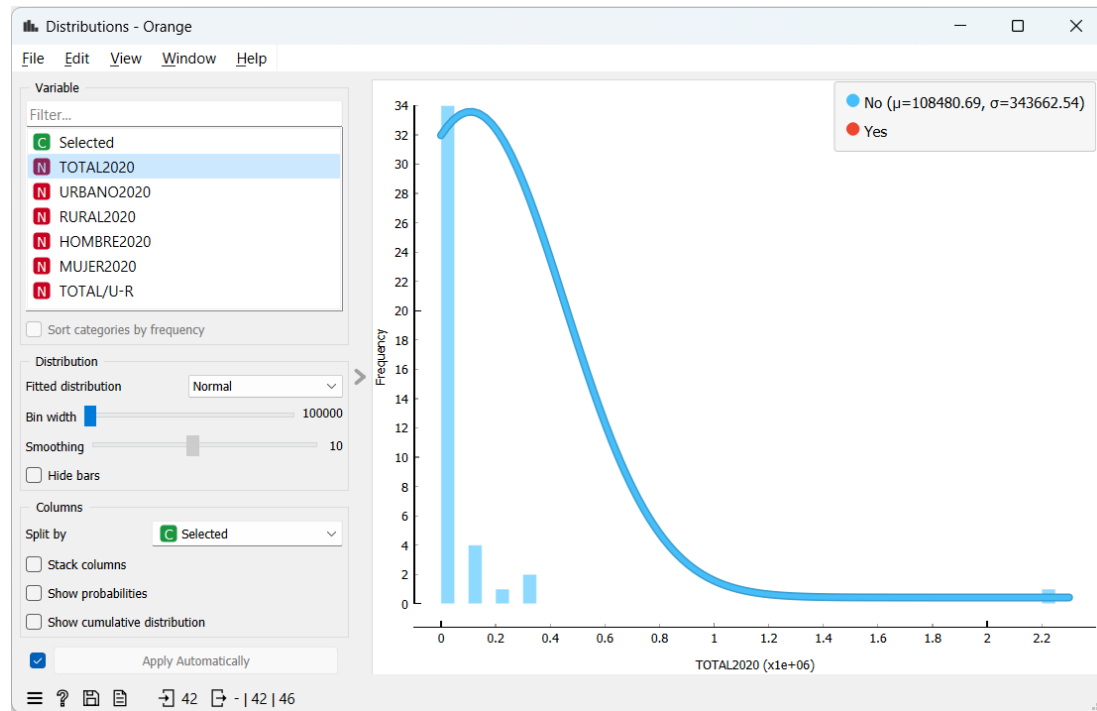
En la figura 25, 26, 27 y 28 se utiliza la herramienta Visualization Por distribution

(figura 25, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



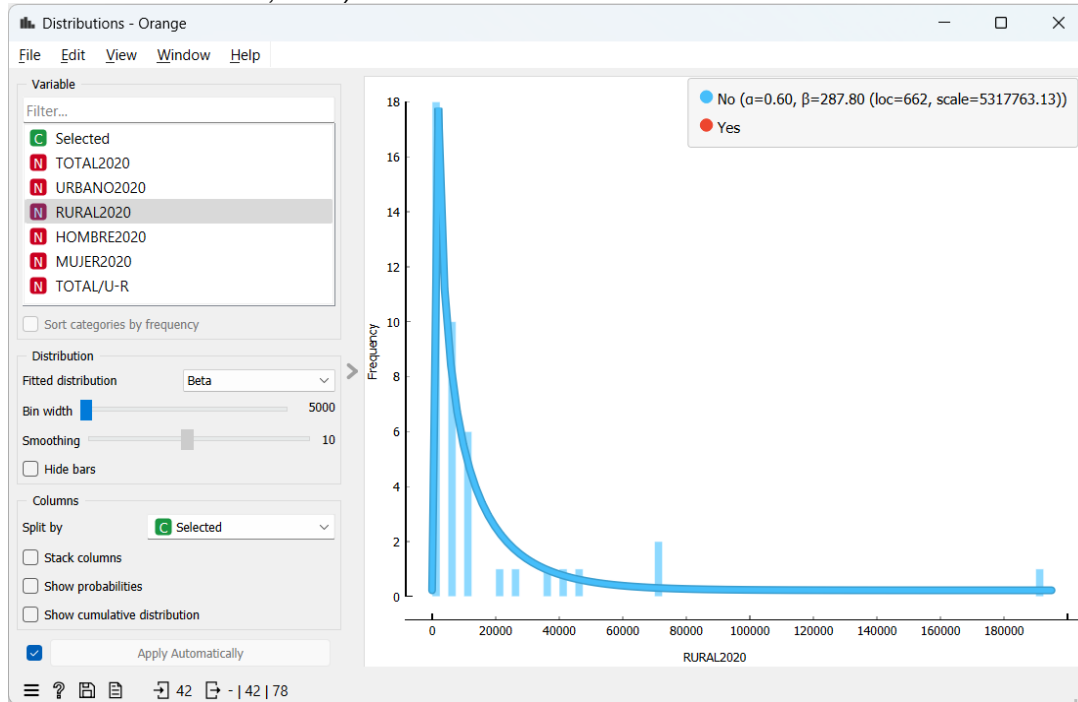
TRABAJO DE GRADO Opción Seminario-Diplomado

(figura 26, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

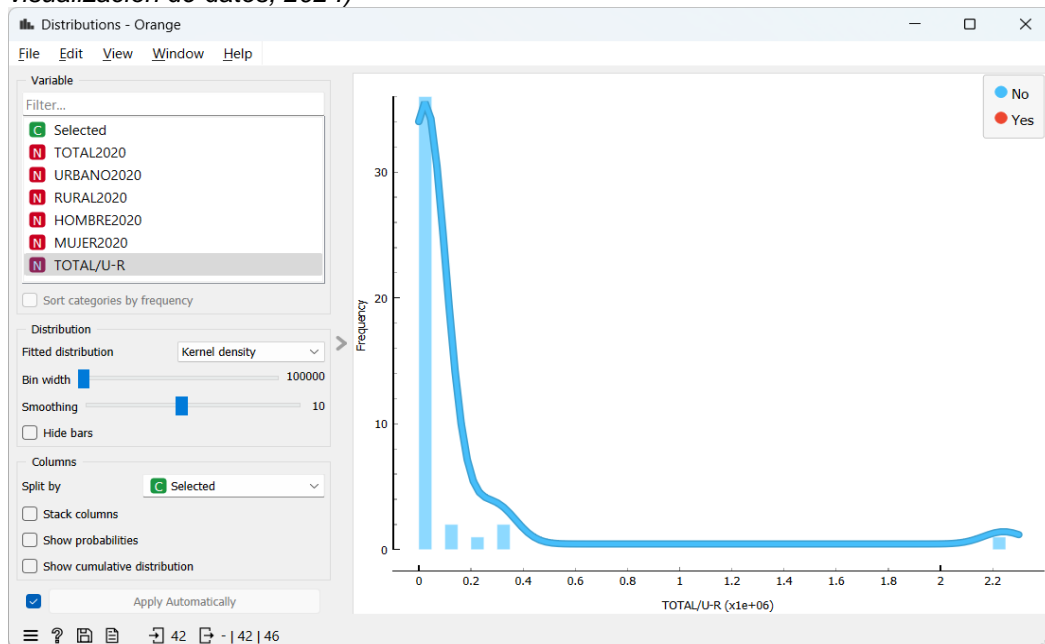


TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 27, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 28, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

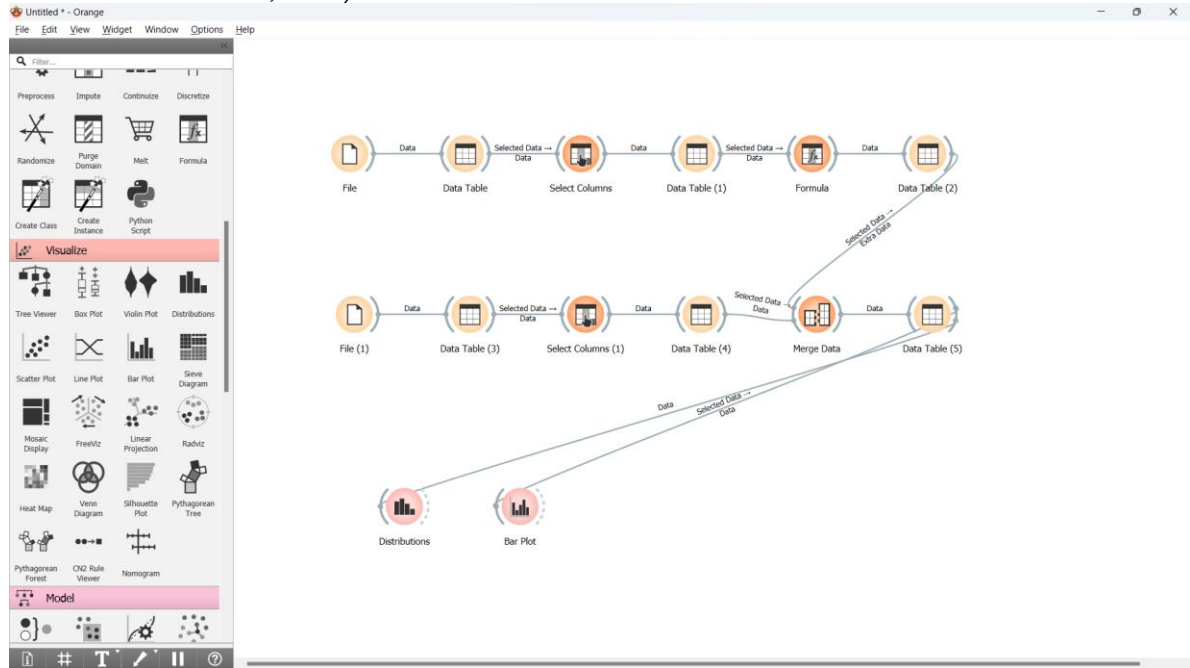


TRABAJO DE GRADO
Opción Seminario-Diplomado

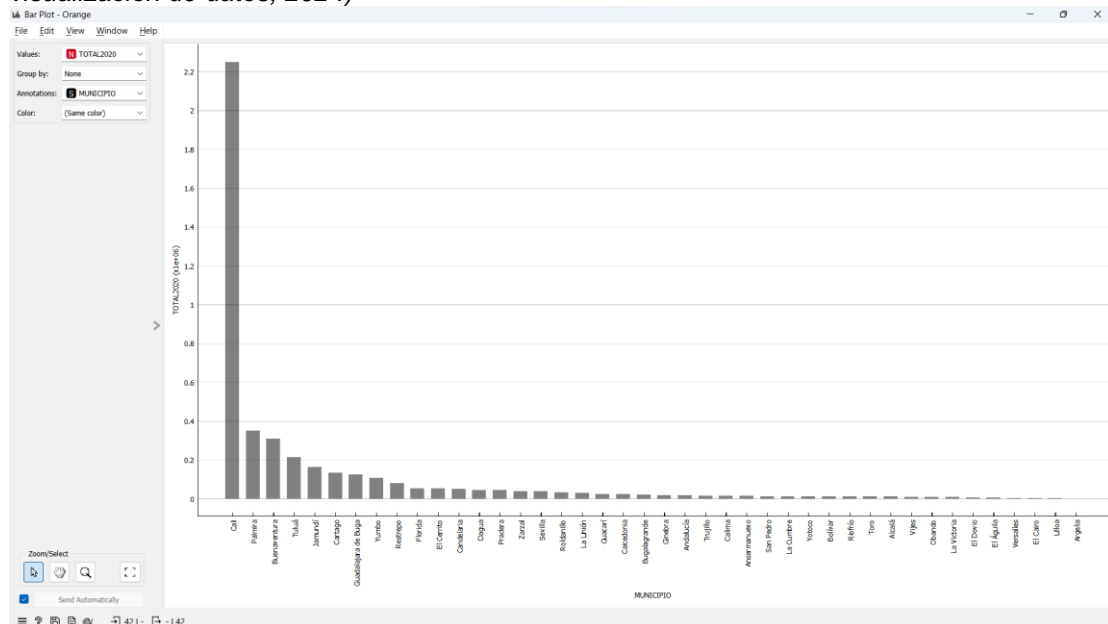
b. Visualización Por bar plot:

En la figura 29 y 30 se utiliza la herramienta Visualización Por bar plot

(figura 29, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 30, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

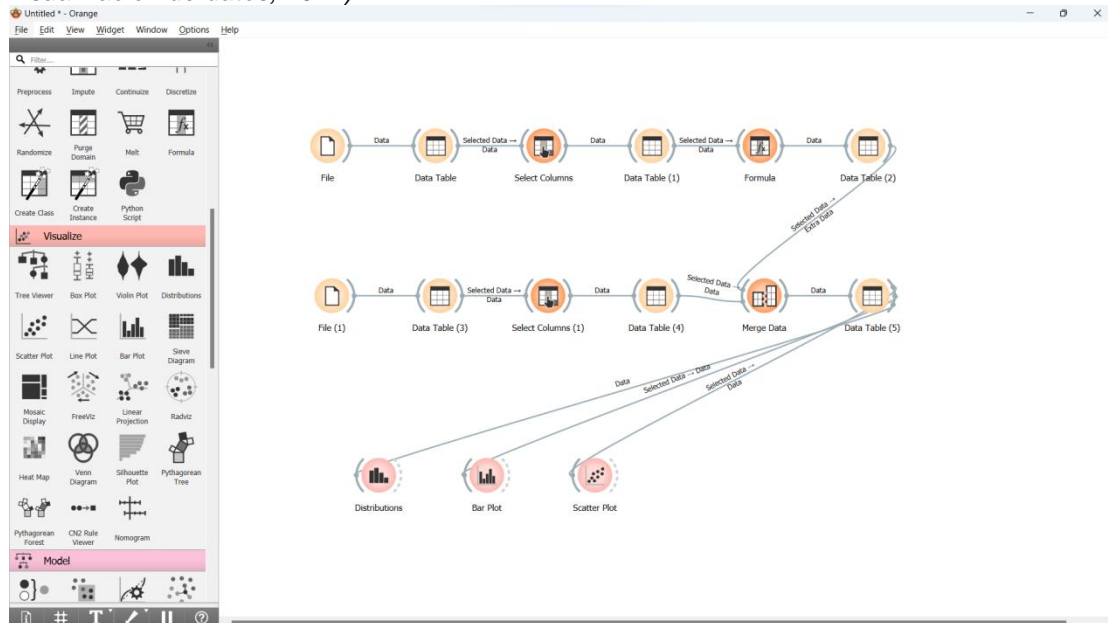


TRABAJO DE GRADO
Opción Seminario-Diplomado

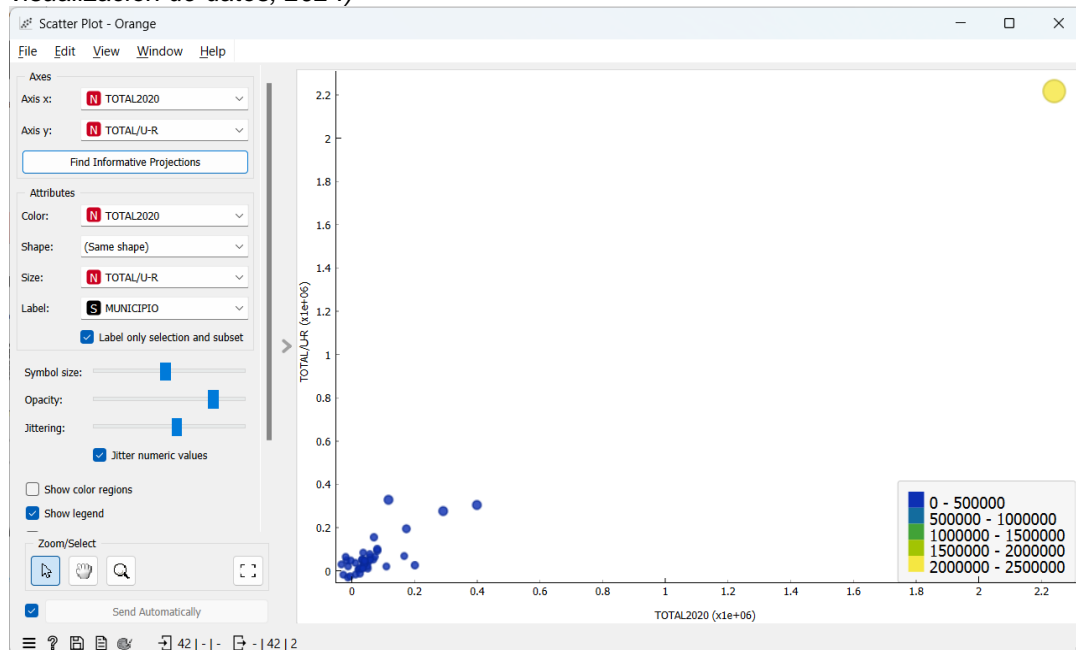
c. Visualización por scatter plot:

En la figura 31 y 32 se utiliza la herramienta Visualización por scatter plot

(figura 31, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 32, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

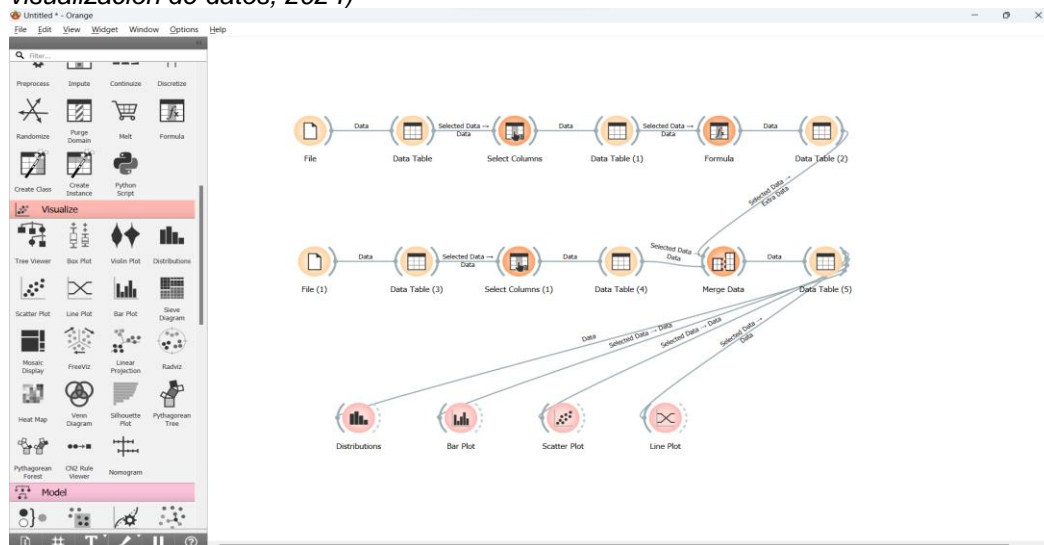


TRABAJO DE GRADO
Opción Seminario-Diplomado

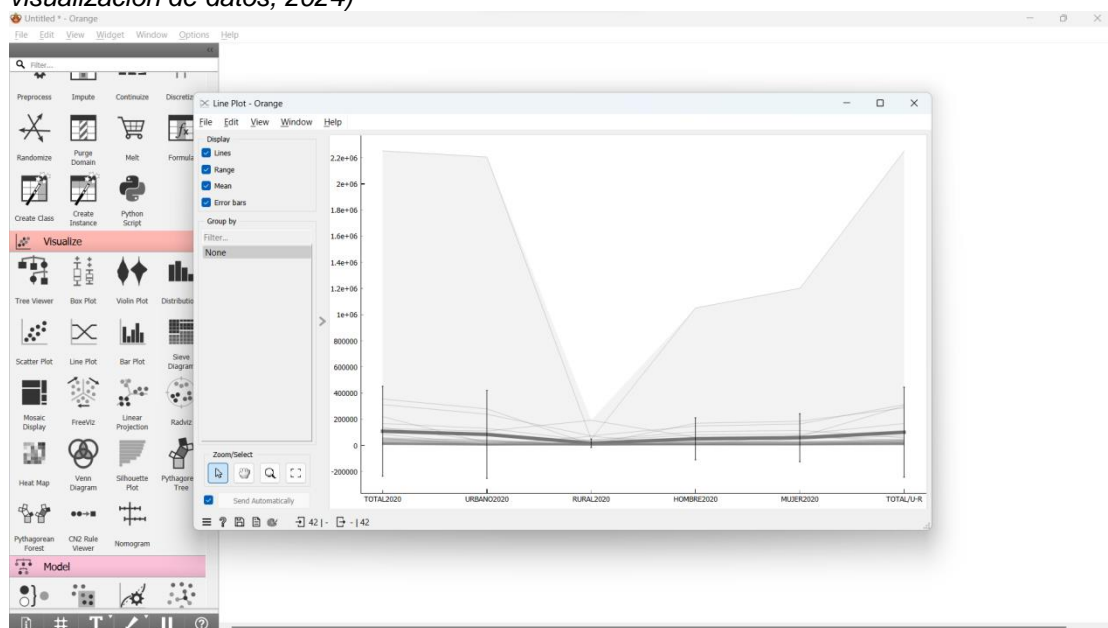
d. Visualización por line plot:

En la figura 33 y 34 se utiliza la herramienta Visualización por line plot

(figura 33, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 34, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

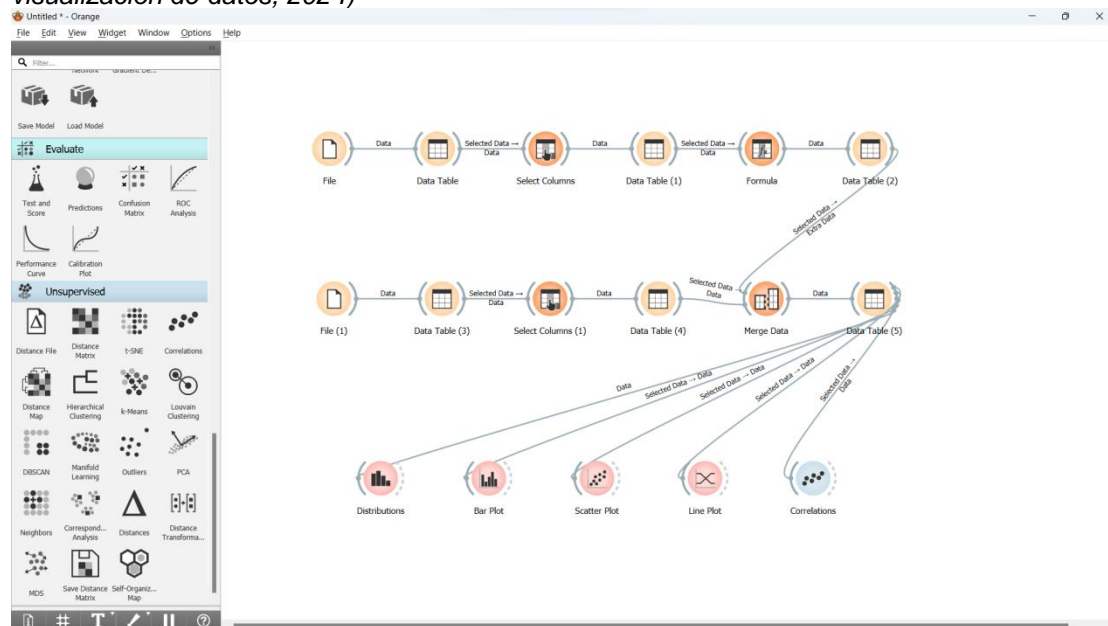
14. UTILIZACIÓN DE OTRAS DE LAS HERRAMIENTAS QUE PRESENTA ORANGE:

Los widgets de aprendizaje sin supervisión:

Unsupervised correlaciones:

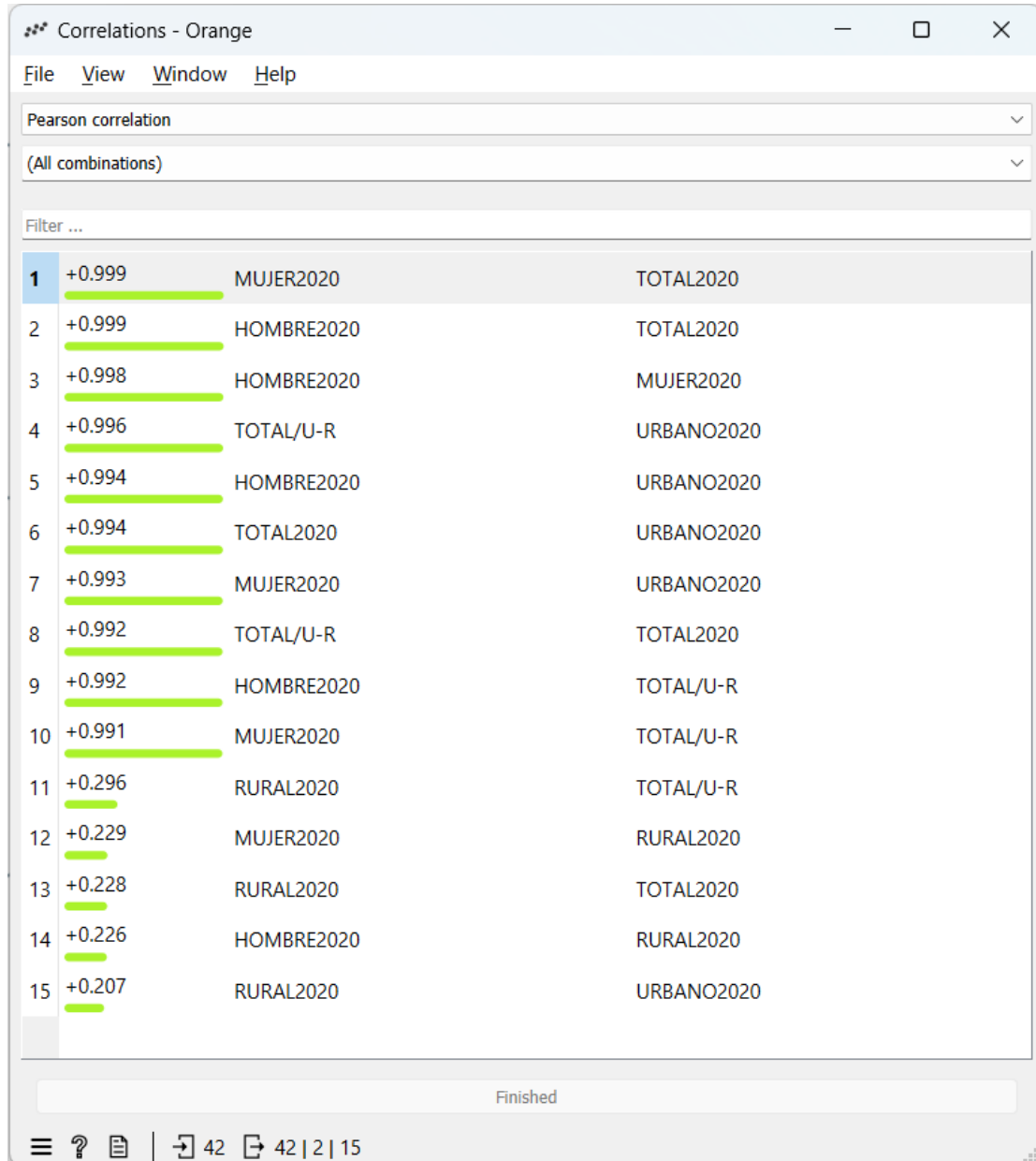
En la figura 35 y 36 se utiliza la herramienta Unsupervised correlaciones

(figura 35, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 36, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

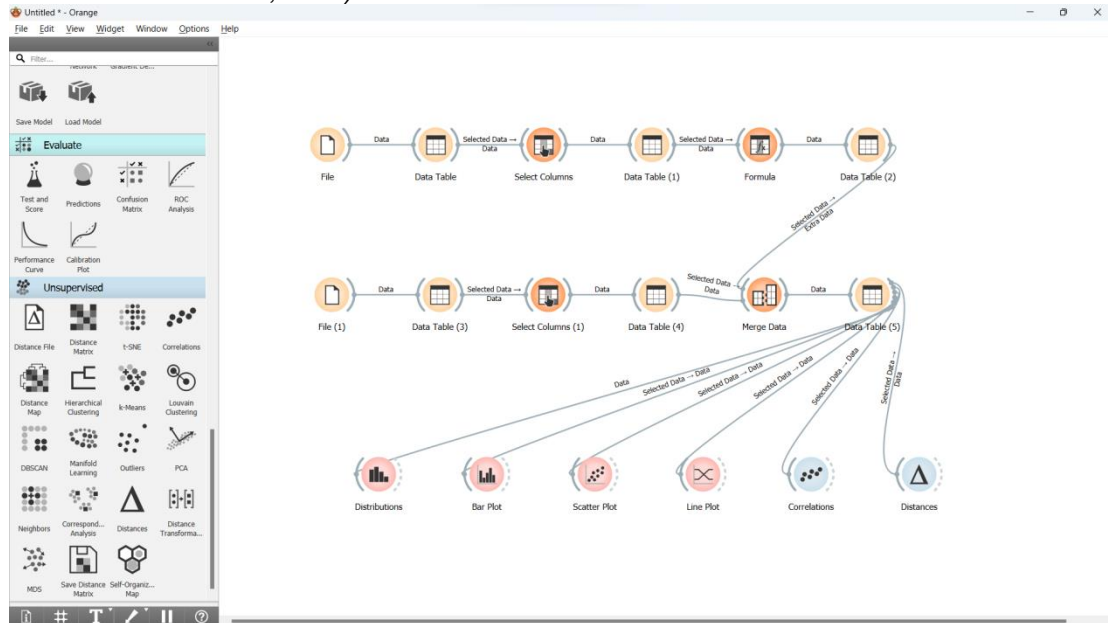


TRABAJO DE GRADO
Opción Seminario-Diplomado

15.1. Unsupervised distances:

En la figura 37 y 38 se utiliza la herramienta Unsupervised correlaciones

(figura 37, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 38, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

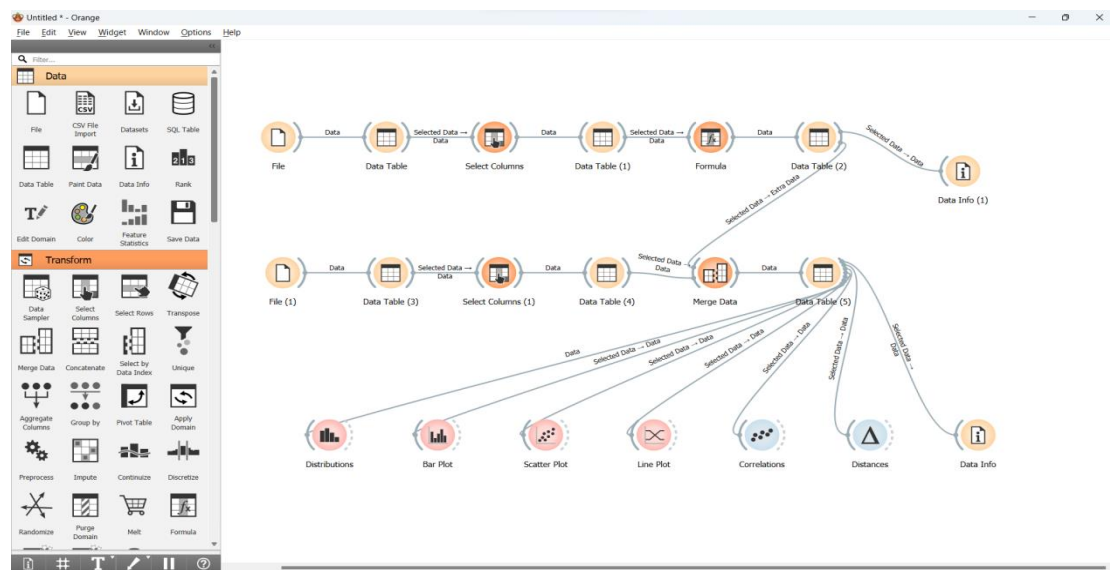
Distances: 42x42 distance matrix

	Cali	Palmira	Buenaventura	Tuluá	Jamundí	Cartago	Ibadalajara de Buga	Yumbo	Restrepo	Florida	El C
Cali		431,602	435,980	470,135	463,390	483,787	480,968	477,579	492,554	491,877	492,688
Palmira	431,602		17,457	47,601	39,835	52,185	61,493	48,024	60,952	60,275	64
Buenaventura	435,980	17,457		41,301	34,556	54,953	44,988	48,745	63,720	63,043	63
Tuluá	470,135	47,601	41,301		18,747	15,675	47,858	22,804	22,418	24,822	22
Jamundí	463,390	39,835	34,556	18,747		20,397	32,905	14,189	29,164	28,487	29
Cartago	483,787	52,185	54,953	15,675	20,397		44,020	11,945	9,403	9,604	13
Ibadalajara de Buga	480,968	61,493	44,988	47,858	32,905	44,020		36,093	52,355	50,391	51
Yumbo	477,579	48,024	48,745	22,804	14,189	11,945	36,093		17,602	14,298	16
Restrepo	492,554	60,952	63,720	22,418	29,164	9,403	52,355	17,602		9,458	8,699
Florida	491,877	60,275	63,043	24,822	28,487	9,604	50,391	14,298	9,458		6,044
El Cerrito	492,688	64,760	63,854	22,552	29,298	13,346	51,201	16,736	8,699	6,044	
Candelaria	491,021	70,163	55,833	23,884	30,328	22,069	43,181	22,139	15,244	16,717	12
Dagua	495,822	64,220	66,988	25,687	32,432	12,035	54,336	18,243	6,291	4,501	4
Pradera	491,948	60,346	63,114	25,116	28,558	10,859	50,461	14,368	9,728	2,044	6

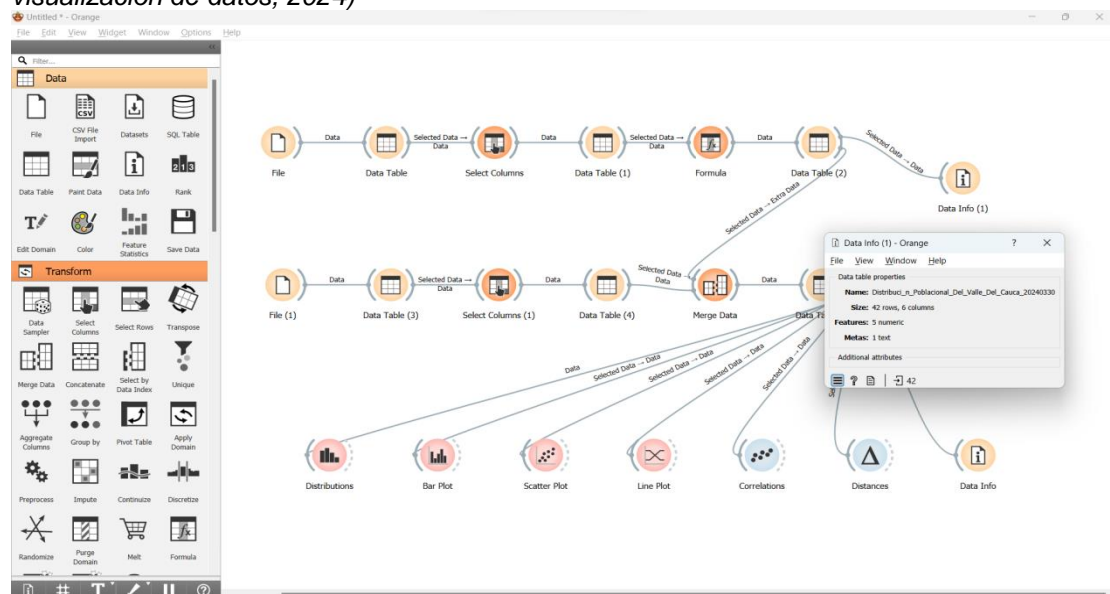
TRABAJO DE GRADO Opción Seminario-Diplomado

También se obtiene la información interna de cualquiera de los widgets sin tener que abrirlos y en el momento en que se requiera con data info como se puede visualizar en las figuras 39 y 40:

(figura 39, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



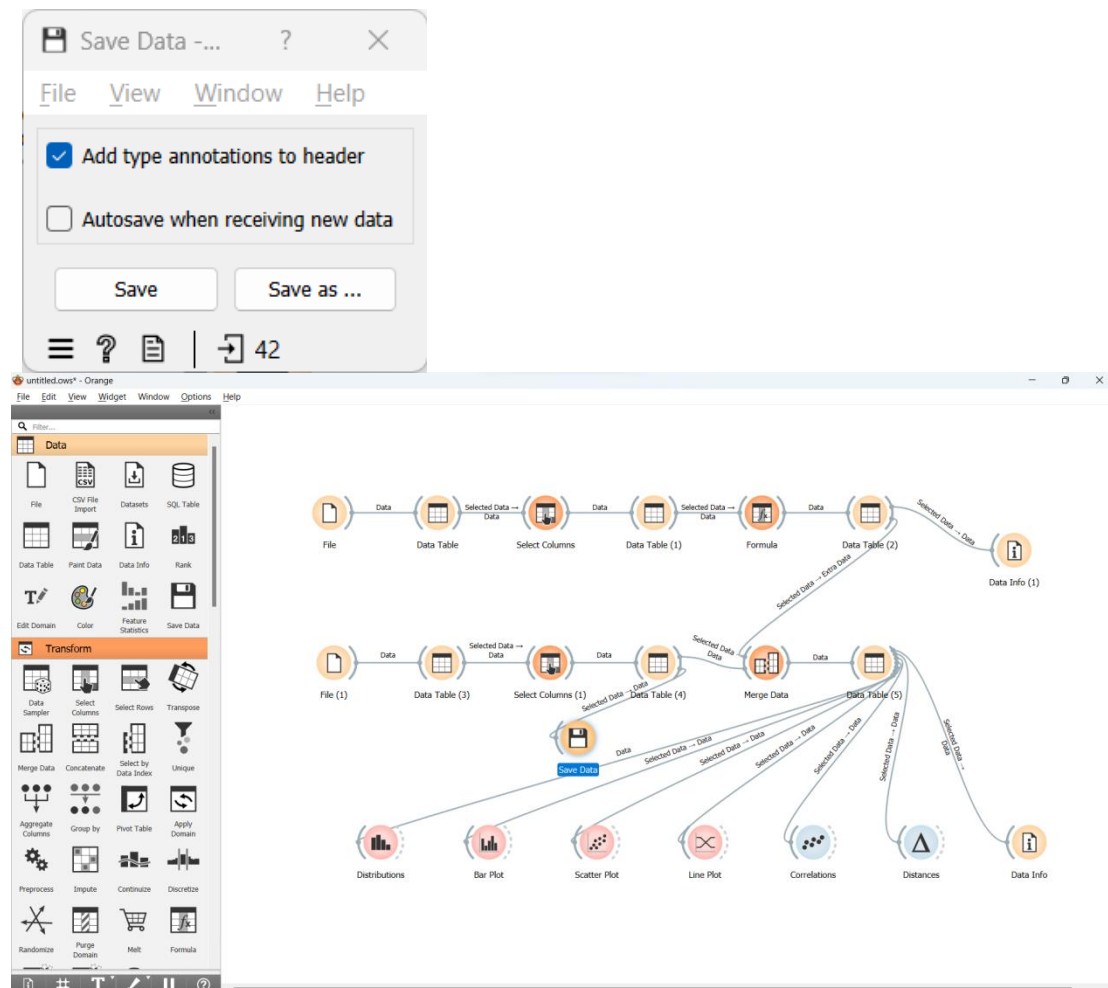
(figura 40, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO Opción Seminario-Diplomado

Se puede ir guardando el análisis por partes si se va a realizar cambios que no se desean perder y luego volver a utilizar en algún punto del análisis según como se evidencia en las imágenes 41 y 42:

(figura 41, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 42, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)

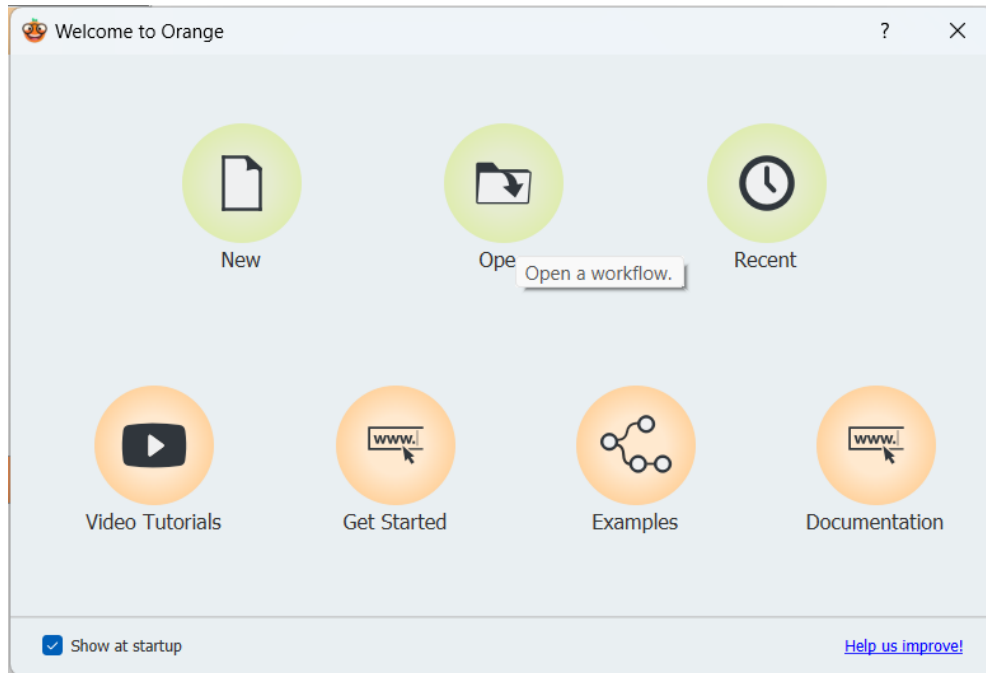
Data: Distribución Poblacional Del Valle Del Cauca total20240330: 42 instances, 2 variables
Features: numeric (no missing values)
Metas: string

	MUNICIPIO	TOTAL2020
1	Cali	2252616
2	Alcalá	14295
3	Andalucía	22594
4	Ansermanuevo	17316
5	Argelia	5177
6	Bolívar	15714
7	Buenaventura	311827
8	Guadalajara de ...	128945
9	Bugalagrande	24466
10	Caicedonia	28521
11	Calima	18266
12	Candelaria	53144
13	Cartago	137301
14	Dagua	49015
15	El Águila	8570
16	El Cairo	6526

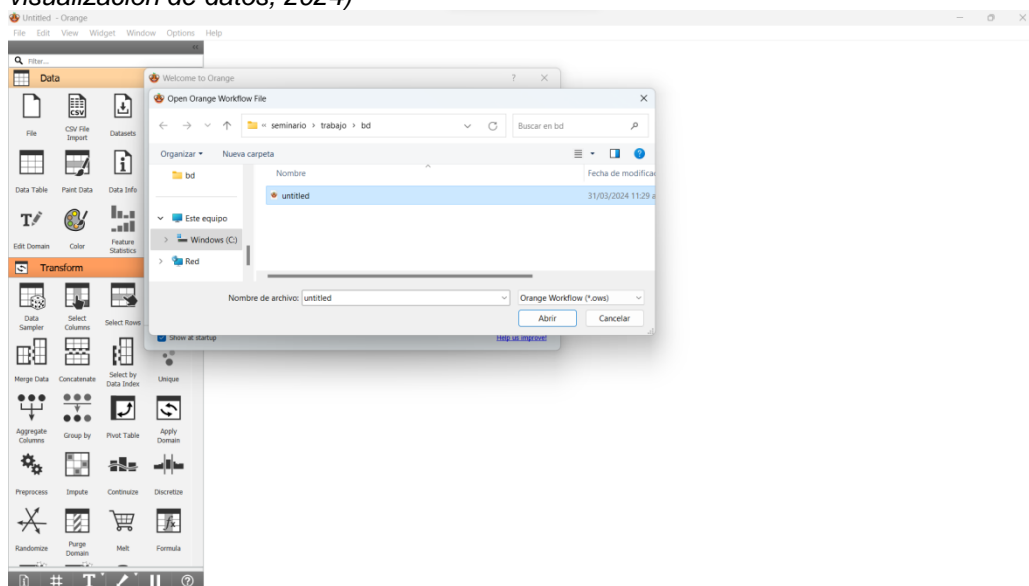
El trabajo de visualización se puede cerrar y abrir en cualquier momento sin temor a que se pierdan los datos analizados así como se ve en la figura 43 y 44:

TRABAJO DE GRADO
Opción Seminario-Diplomado

(figura 43, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



(figura 44, creación del autor, Aprendizaje de nuevas herramientas de ia para el análisis y visualización de datos, 2024)



TRABAJO DE GRADO
Opción Seminario-Diplomado

16. RESULTADOS ESPERADOS

Dentro de los resultados esperados para el presente trabajo se esperan una mayor eficiencia en el análisis de los datos, todo debido al uso de las nuevas herramientas para la visualización y análisis de los datos, pudiendo contar con una integración de información mucho mas confiable, rápida y precisa. Obteniendo también una considerable mejora en la toma de decisiones basados en datos.

Un mayor reconocimiento a la necesidad de generar mayores y mejores conocimientos sobre IA y la aplicación en las diferentes áreas laborales y educativas, desarrollando las habilidades necesarias para la generación de habilidades que generen procesos innovadores para la resolución de problemas complejos, haciendo indiscutible la mejora en el desempeño laboral y educativo.

TRABAJO DE GRADO
Opción Seminario-Diplomado

17. CONCLUSIONES

A través del proceso de aprendizaje de análisis y visualización de datos con herramientas de inteligencia artificial, hemos logrado adquirir varios conocimientos y habilidades que nos permiten comprender en profundidad la importancia de utilizar estas tecnologías en la toma de decisiones basadas en datos.

Antes que nada, hemos podido aprender a utilizar varias técnicas de análisis de datos, como la exploración de datos, la limpieza de datos, la transformación de variables, etc. Estas técnicas nos permiten extraer información valiosa para la toma de decisiones.

Además, hemos desarrollado habilidades en la visualización de datos, utilizando herramientas como gráficos y tablas para representar de manera clara y comprensible la información contenida en los datos. La visualización de datos nos ha permitido comunicar de manera efectiva los resultados de nuestros análisis y facilitar la interpretación de los mismos por parte de otras personas.

También hemos explorado el potencial de las herramientas de inteligencia artificial en el análisis y visualización de datos, como algoritmos de aprendizaje

TRABAJO DE GRADO
Opción Seminario-Diplomado

automático no supervisados. Estas herramientas nos brindan la posibilidad de realizar análisis más avanzados y obtener predicciones más precisas a partir de los datos, lo que ha ampliado nuestras capacidades analíticas y toma de decisiones.

El aprendizaje de análisis y visualización de datos con herramientas de inteligencia artificial ha sido fundamental para fortalecer nuestras habilidades en el manejo de grandes volúmenes de datos, la identificación de patrones y tendencias, la comunicación de resultados y la toma de decisiones informadas. Estamos preparados para aplicar estos conocimientos y habilidades en situaciones reales y contribuir al éxito de proyectos y organizaciones a través del análisis de datos y la inteligencia artificial.

Por lo anterior podemos concluir que los resultados del aprendizaje del análisis y visualización de los datos con Orange nos da la oportunidad de conocer las diferentes oportunidades en el uso de las herramientas de inteligencia artificial, además de poder aplicar una variedad de funcionalidades según la complejidad de los conjuntos de datos utilizados y las técnicas aplicadas. Podríamos decir que algunas de esas nuevas habilidades estarían dirigidas a:

Análisis estadístico descriptivo de un conjunto de datos, incluyendo medidas de tendencia, dispersión y distribución.

TRABAJO DE GRADO
Opción Seminario-Diplomado

Creación de gráficos y visualizaciones para explorar patrones y tendencias en los datos, como gráficos por distribución, por bar plot, scatter plot, line plot, Unsupervised correlaciones, Unsupervised distances. Identificación de relaciones entre variables utilizando técnicas de correlación y regresión. Clasificación de datos en grupos o categorías utilizando algoritmos de clustering.

El aprendizaje del análisis y visualización de datos con Orange pudo proporcionarnos una comprensión más profunda y significativa de la información contenida en los conjuntos de datos, permitiendo tomar decisiones más informadas y realizar análisis más efectivos.

TRABAJO DE GRADO
Opción Seminario-Diplomado

18. REFERENCIAS BIBLIOGRAFICAS

<https://www.hpe.com/lamerica/es/what-is/ai-data-management.html>

<https://www.unite.ai/es/analistas-de-datos-de-herramientas-de-inteligencia-artificial/>

<https://www.ibm.com/es-es/topics/data-visualization>

<https://www.questionpro.com/blog/es/analisis-de-inteligencia-artificial/>

<https://educacioncontinua.uniandes.edu.co/es/programas/herramientas-de-inteligencia-artificial-y-analisis-de-datos-en-el-diseno-de-productos>

<https://www.oracle.com/es/business-analytics/what-is-data-visualization/>

<https://orangedatamining.com/>

<https://www.comparasoftware.co/orange-mineria-de-datos>

<https://culturacrm.com/crm/recursos-crm/orange-data-mining-analisis-datos>

TRABAJO DE GRADO
Opción Seminario-Diplomado

- Willy marcelo maurer, march 2019, minería de datos fácil minería de datos para economistas y profesionales de negocios
- Alberto díaz ribón, curso académico 2021/2022 , máster en estadística aplicada, software disponible en técnicas de minería de datos. Comparativa y aplicación a datos reales, universidad de granada.
- Sergio pio alvarez, 2018, minería de calidad de datos: aplicación de técnicas de minería de datos para la evaluación de la calidad de los datos, universidad de la república montevideo – uruguay.
- Silvia mariana haro rivera, 2016-2017, tecnicas de clasificación de minería de datosy software orange, universidad nueva granada.