****

**GUÍA PRÁCTICA**

**LECTURA DE FICHERO DE DATOS Y VISUALIZACIÓN DE DATOS**

1. **Datos Generales**

|  |  |
| --- | --- |
| **Carrera:** | **Tecnología Superior en Big Data** |
| **Período académico:** | **Octubre 2023 – Febrero 2024** |
| **Asignatura:** | **Minería de Datos - Laboratorio** |
| **Unidad Nº:** | 1. **Conceptos Básicos** |
| **Tema:** | **Lectura de ficheros de datos en Python para la Visualización de Datos.** |
| **Ciclo-Paralelo:** | **M3A** |
| **Fecha de inicio de la Unidad:** | **27/10/2023** |
| **Fecha de fin de la Unidad** | **10/11/2023** |
| **Práctica Nº**: | **1** |
| **Horas:** | **9** |
| **Docente:** | **Mgtr. Verónica Paulina Chimbo Coronel** |

1. **Contenido** 
   1. **Fundamentos**

|  |
| --- |
| La gestión eficiente y la comprensión profunda de los datos son fundamentales en cualquier proyecto de análisis y toma de decisiones. Esta guía práctica tiene como objetivo proporcionar una introducción detallada sobre cómo leer ficheros de datos en Python y visualizar la información de manera efectiva. Exploraremos el uso de las librerías estándar como Pandas y Matplotlib/Seaborn para manipulación y visualización de datos, así como la integración de herramientas especializadas como "Y-Data Profiling" y "PowerClient - Quick Report" para análisis más avanzados y generación de informes rápidos.  En los primeros pasos, aprenderemos los fundamentos de la lectura de diferentes tipos de ficheros, desde CSV hasta JSON, utilizando Pandas. A continuación, nos sumergiremos en el análisis exploratorio de datos, abordando conceptos como limpieza, manipulación y estadísticas descriptivas para obtener una comprensión sólida de los conjuntos de datos.  Posteriormente, exploraremos técnicas de perfilado de datos utilizando herramientas especializadas como "Y-Data Profiling" para obtener métricas detalladas y visualizaciones avanzadas. Finalmente, nos sumergiremos en la creación y representación de instancias de visualización rápida utilizando "PowerClient - Quick Report" y otras librerías de visualización, permitiéndonos comunicar de manera efectiva los resultados obtenidos.  Esta guía está diseñada para aquellos que buscan mejorar sus habilidades en la lectura de ficheros de datos, manipulación y visualización en Python, integrando herramientas especializadas para un análisis más avanzado y la generación de informes eficientes. |

* 1. **Objetivos de la Guía**

|  |
| --- |
| * Dominar la Lectura de Ficheros: Proporcionar habilidades sólidas para la lectura de diferentes formatos de ficheros de datos, utilizando la biblioteca Pandas en Python, permitiendo a los usuarios cargar, explorar y entender datos de manera eficiente. * Análisis Exploratorio Efectivo: Enseñar técnicas fundamentales de análisis exploratorio de datos, incluida la identificación y manejo de valores faltantes, así como la exploración de estadísticas descriptivas, para obtener una comprensión inicial de los conjuntos de datos. * Visualización Impactante: Introducir conceptos clave de visualización de datos utilizando Matplotlib y Seaborn, y explorar la integración de "PowerClient - Quick Report" para generar visualizaciones impactantes que resalten patrones y tendencias de manera clara y efectiva. |

* 1. **Evaluación del Aprendizaje**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rúbrica de Evaluación de la Guía Práctica   |  |  | | --- | --- | | **Criterios de Evaluación** | **Puntuación Máxima** | | Paso 1: Instalación de Python y Bibliotecas | /1 | | Paso 2: Importar Bibliotecas | /2 | | Paso 3: Carga de Datos | /5 | | Paso 4: Exploración de Datos | /1 | | Paso 5: Exportar Resultados | /1 | | Puntuación Total | /10 | |

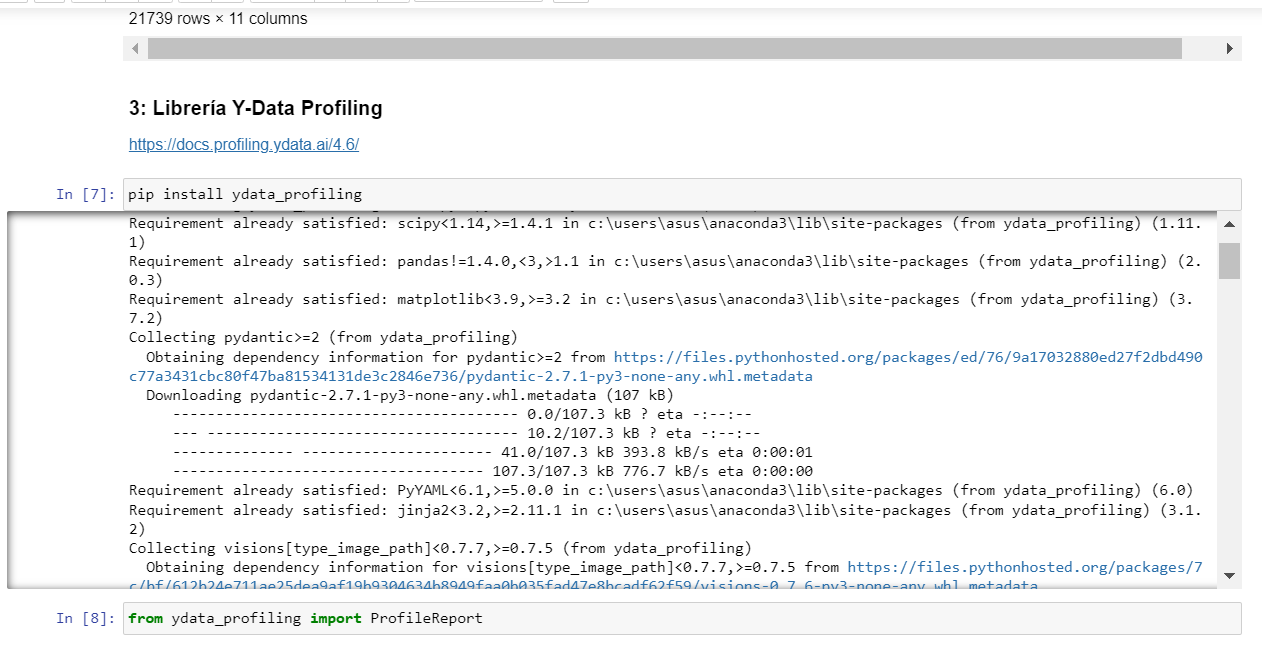
* 1. **Preparación previa, materiales, herramientas, equipos y software**

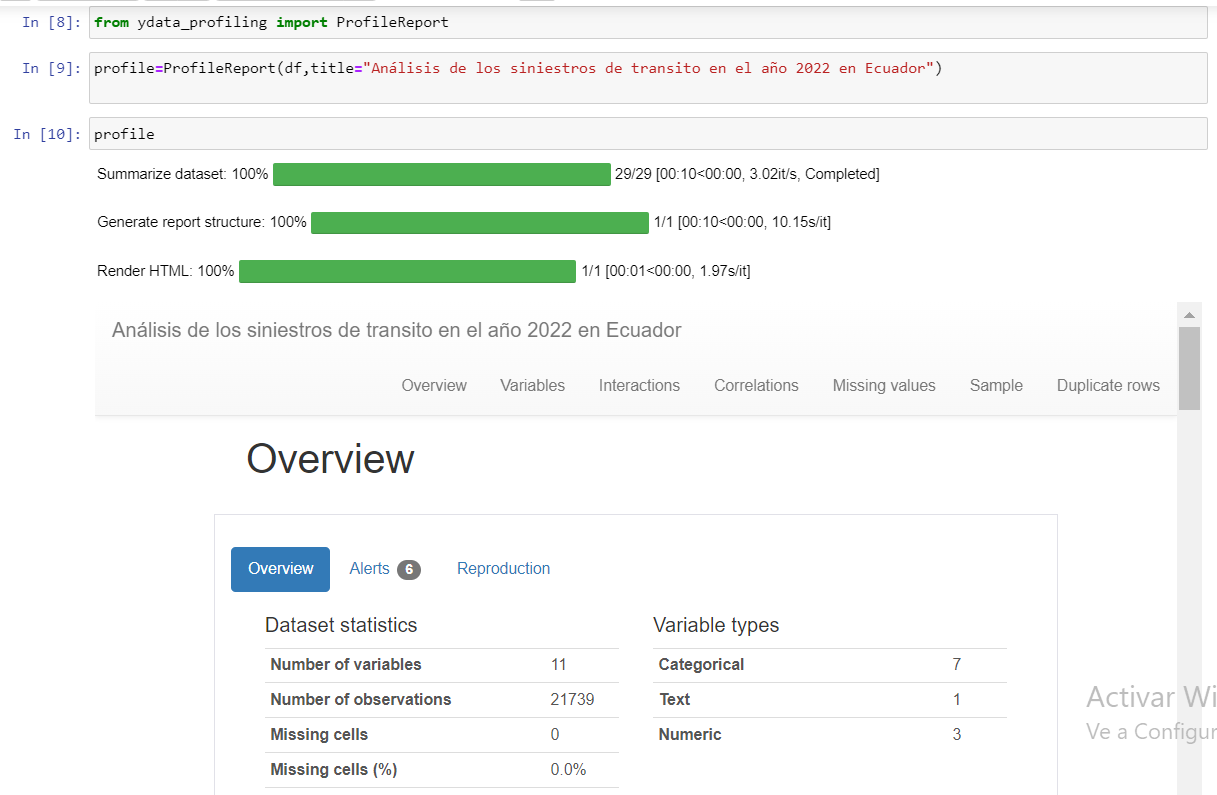
|  |
| --- |
| Antes de comenzar con la guía, es esencial realizar la preparación previa para asegurar una experiencia fluida. A continuación, se detallan los elementos necesarios:  **Materiales:**  Conjunto de Datos: Asegúrate de tener un conjunto de datos relevante para tu práctica. Puede ser un archivo CSV, Excel, JSON u otro formato común.  **Herramientas:**  Anaconda Navigator: Instala Anaconda Navigator, una plataforma que facilita la gestión de entornos y paquetes de Python. Puedes descargarlo desde aquí.  Jupyter Notebook: Utiliza Jupyter Notebook, que generalmente viene incluido con Anaconda. Proporciona un entorno interactivo para el desarrollo paso a paso.  Librerías Python:  Pandas: Para la manipulación de datos.  Matplotlib y Seaborn: Para la visualización de datos.  Librerías Específicas (si es necesario):  Y-Data Profiling y PowerClient - Quick Report  **Equipos y Software:**  Computadora: Asegúrate de tener acceso a una computadora con suficiente capacidad de procesamiento y almacenamiento.  Sistema Operativo: La guía está diseñada para ser compatible con sistemas operativos Windows, macOS o Linux.  Espacio en Disco: Verifica que tienes suficiente espacio en disco para almacenar conjuntos de datos y posiblemente los informes y visualizaciones generadas. |

* 1. **Procedimientos a emplear**

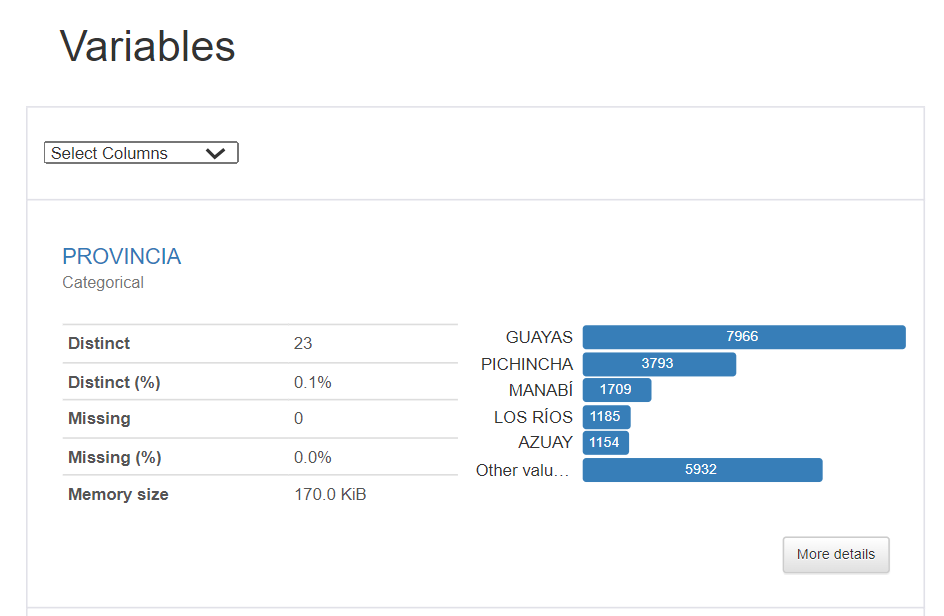
|  |
| --- |
| 1. **Inicio del Entorno Virtual:**   Abre Anaconda Navigator y selecciona el entorno virtual que creaste para esta guía. Asegúrate de que esté activado.     1. **Lanzamiento de Jupyter Notebook:**   Abre una terminal en tu entorno virtual y escribe jupyter notebook. Esto abrirá la interfaz de Jupyter en tu navegador.     1. **Creación de un Nuevo Cuaderno (Notebook):**   En la interfaz de Jupyter, crea un nuevo cuaderno. Este será tu espacio de trabajo para seguir la guía paso a paso.    1.- Seleccionar un directorio para guardar el archivo  2.- Crear el notebook con la siguiente estructura de nombre practica1\_nombres\_apellidos     1. **Importación de Librerías:**   En la primera celda de tu cuaderno, importa las librerías necesarias utilizando los comandos:   1. **Lectura del Conjunto de Datos:** 2. **Visualización Básica:**   **Documenta tus pasos, observaciones y decisiones en el cuaderno de Jupyter. Reflexiona sobre los resultados y las visualizaciones generadas.**  **Guardado del Cuaderno y Resultados:**  **Guarda tu cuaderno de Jupyter con todos los resultados y visualizaciones. También, guarda cualquier informe o gráfico generado durante la práctica.**  **Para la Data se debe de seleccionar de los repositorios existentes para descargar datos a nivel nacional.**   1. [**Instituto Nacional de Estadística y Censos (INEC)**](https://www.ecuadorencifras.gob.ec/directoriodeempresas/)**:** El INEC de Ecuador proporciona una amplia gama de datos estadísticos relacionados con población, economía, agricultura y otros temas. Puedes encontrar datos en su sitio web oficial. 2. [**Portal de Datos Abiertos del Gobierno de Ecuador**](https://datosabiertos.gob.ec/dataset/base-de-datos-de-matricula-de-uep-ano-2021)**:** El gobierno de Ecuador tiene un portal de datos abiertos que ofrece una variedad de conjuntos de datos en áreas como salud, educación, transporte y más. Puedes acceder a estos datos de forma gratuita. 3. [**Banco Central del Ecuador:**](https://contenido.bce.fin.ec/documentos/Estadisticas/)El Banco Central del Ecuador publica datos económicos, financieros y estadísticas relacionadas con la economía del país. Su sitio web proporciona acceso a informes y datos económicos. 4. [**World Bank Data**](https://datos.bancomundial.org/pais/ecuador?view=chart)**:** El Banco Mundial ofrece una plataforma de datos donde puedes encontrar información sobre el desarrollo de Ecuador, incluyendo indicadores económicos, sociales y ambientales. 5. [**Observatorio Laboral de Ecuador:**](https://www.ecuadorencifras.gob.ec/estadisticas-laborales-abril-2023-2/) Este portal proporciona datos relacionados con el mercado laboral en Ecuador, incluyendo tasas de empleo, salarios y más. 6. [**Universidades y Centros de Investigación:**](https://siau.senescyt.gob.ec/estadisticas-de-educacion-superior-ciencia-tecnologia-e-innovacion/) Muchas universidades y centros de investigación en Ecuador publican datos relacionados con sus investigaciones. Puedes explorar los sitios web de estas instituciones para acceder a datos académicos. 7. [**Repositorio de datos abiertos de la UNAE**](https://unae.edu.ec/observatorio/publicaciones_interes/bases-de-datos/).- Presenta un sin numero de datos que pueden ser descargados para analizar. 8. **Organizaciones Internacionales:** Organizaciones como la ONU, la OMS y el FMI suelen publicar datos relacionados con Ecuador en sus sitios web. 9. [**Datos Abiertos del GAD Cuenca.**](https://cuencaparticipa.cuenca.gob.ec/processes/Datosabiertos)-Acceder a datos abierto del GAD de cuenca sobre datos relacionados a turismo, transporte entre otros.   La data a usar se llama 2022\_SINIESTROS\_TRÁNSITO\_BDD y se obtuvo del [**Instituto Nacional de Estadística y Censos (INEC)**](https://www.ecuadorencifras.gob.ec/directoriodeempresas/)   1. CARGA Y VISUALIZACION DE LA DATA |

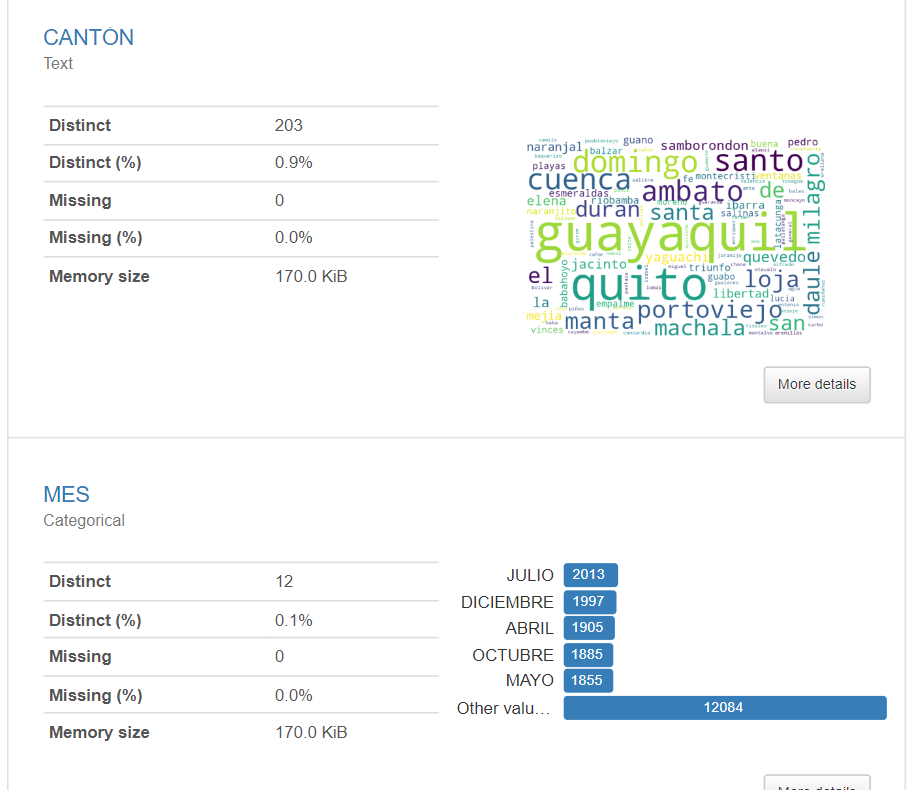
1. CARGA Y USO DE LA LIBRERÍA Y-DATA PROFILING

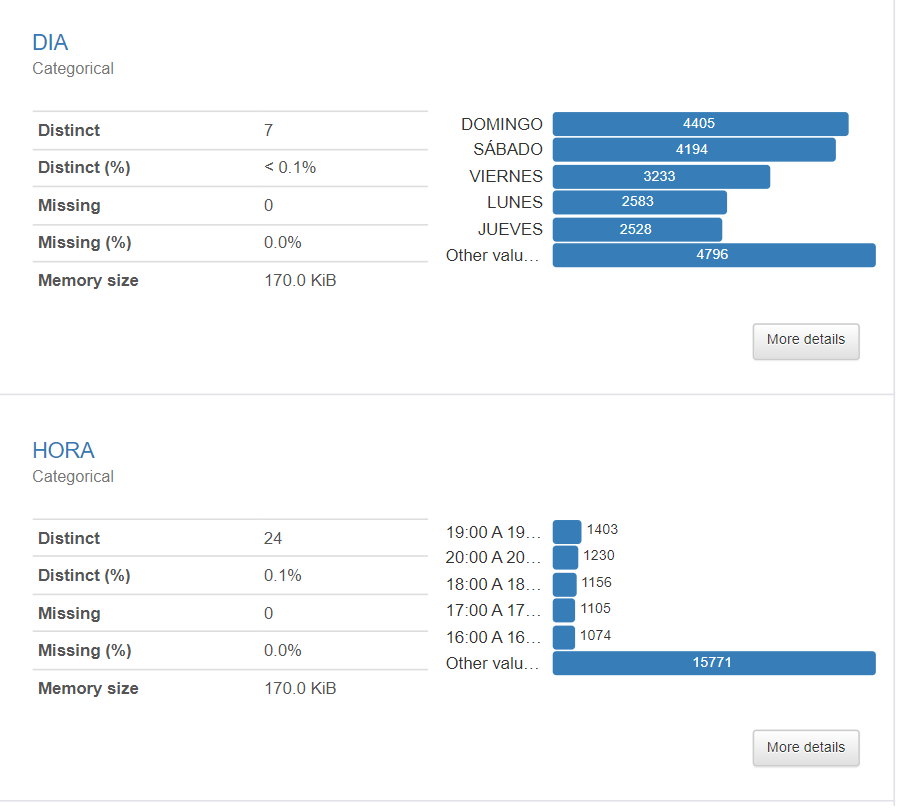
****

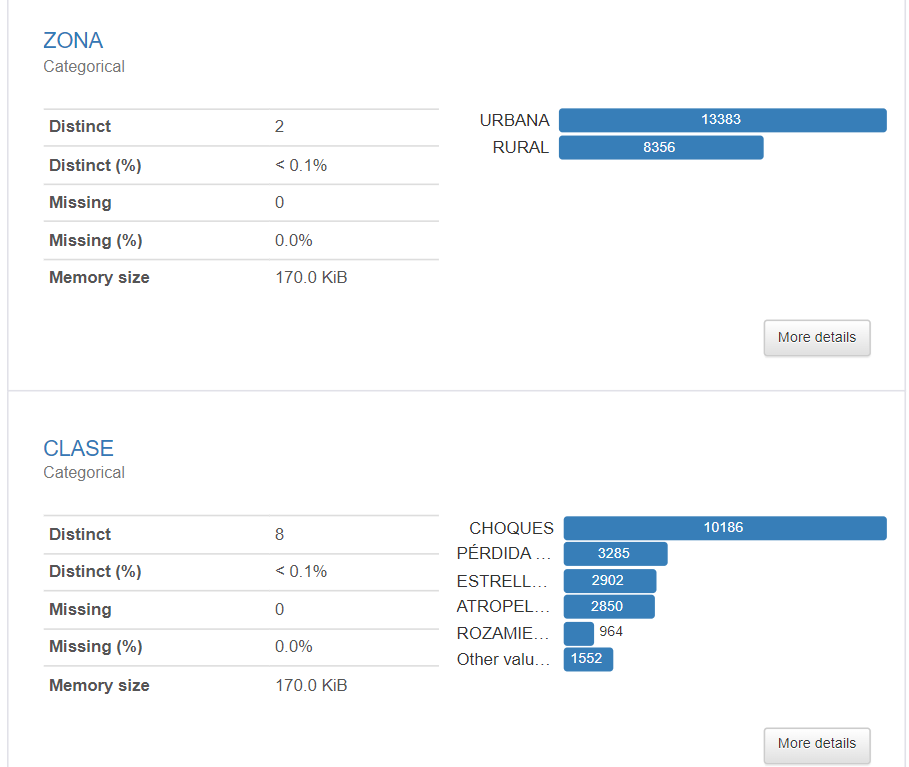
****

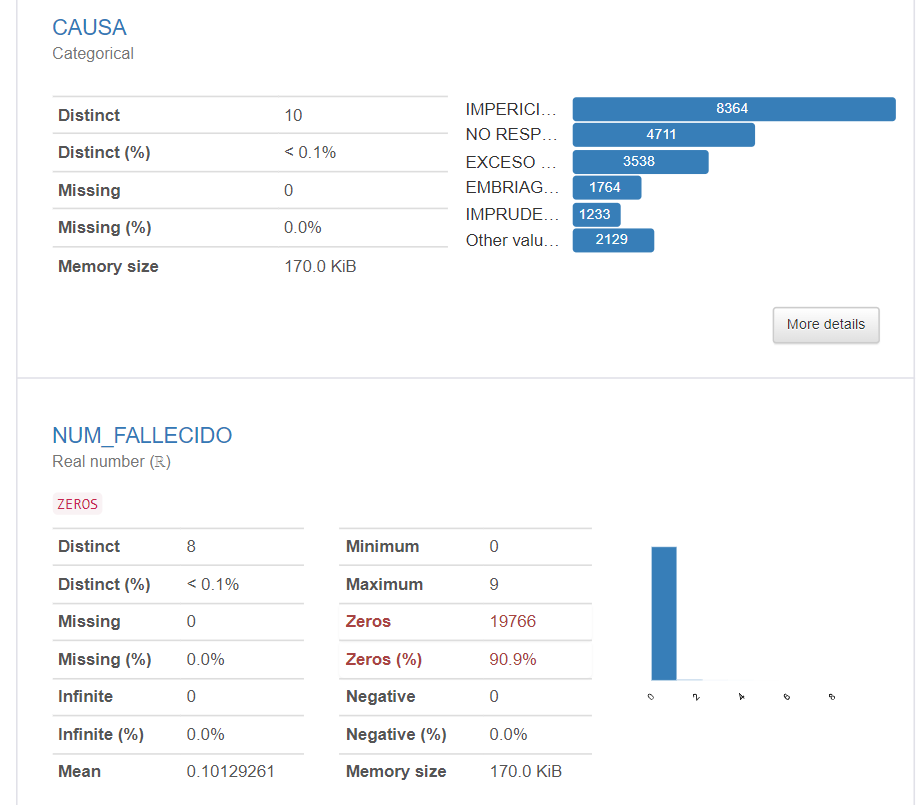
****

****

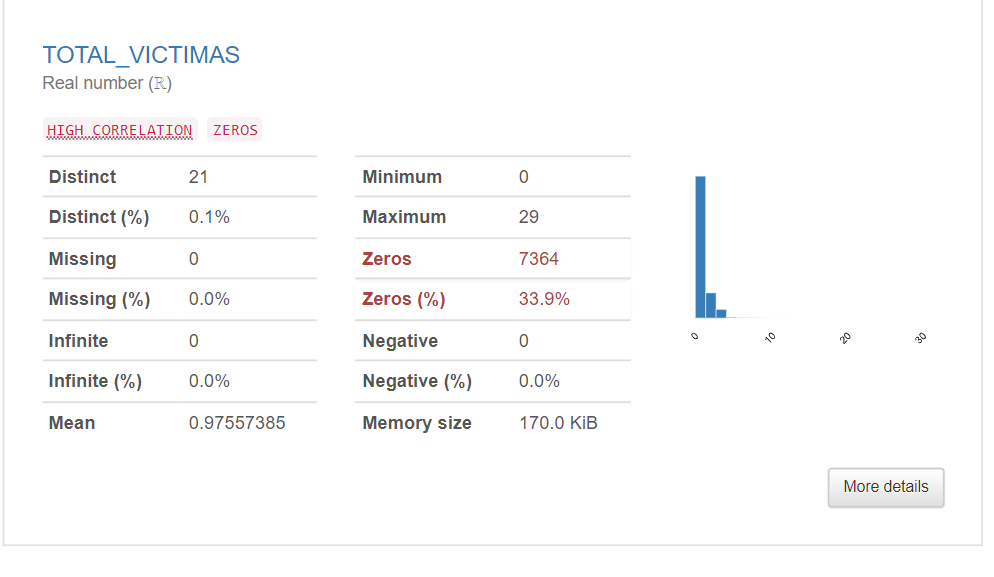
****

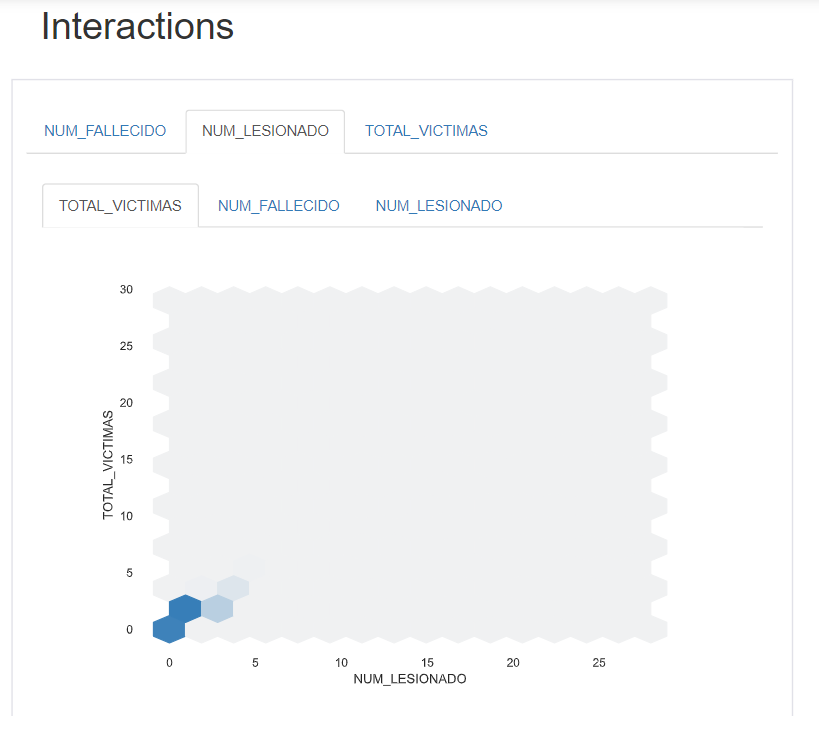
****

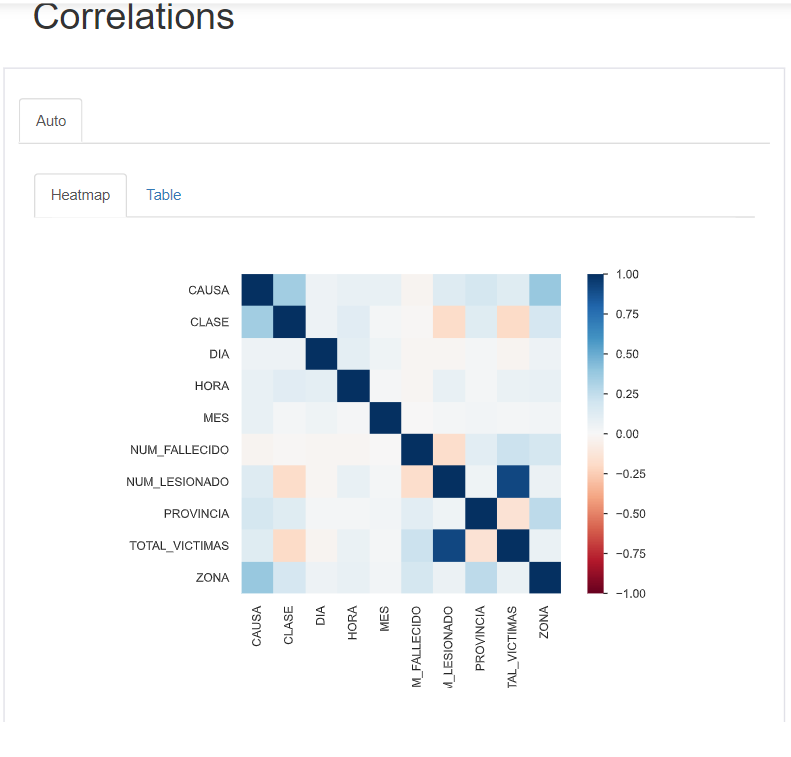
****

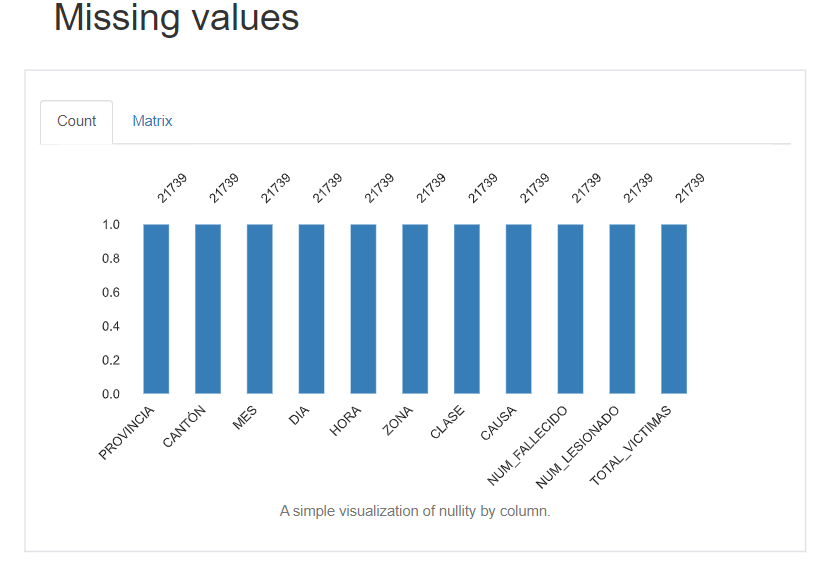
****

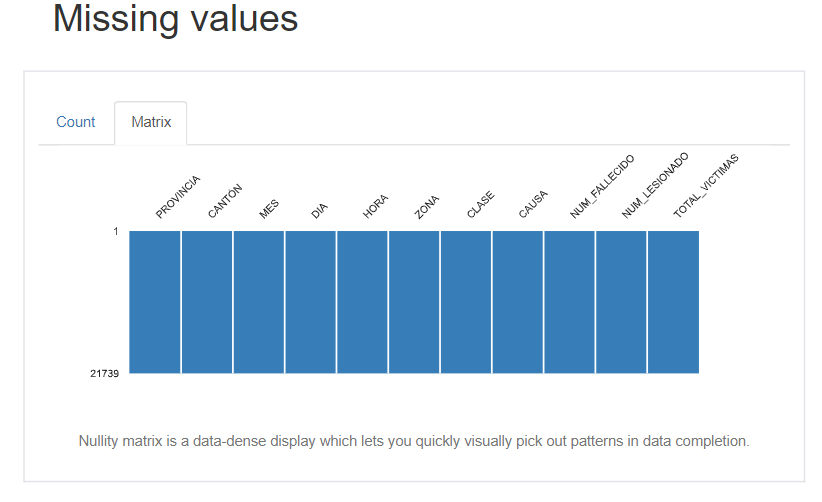
****

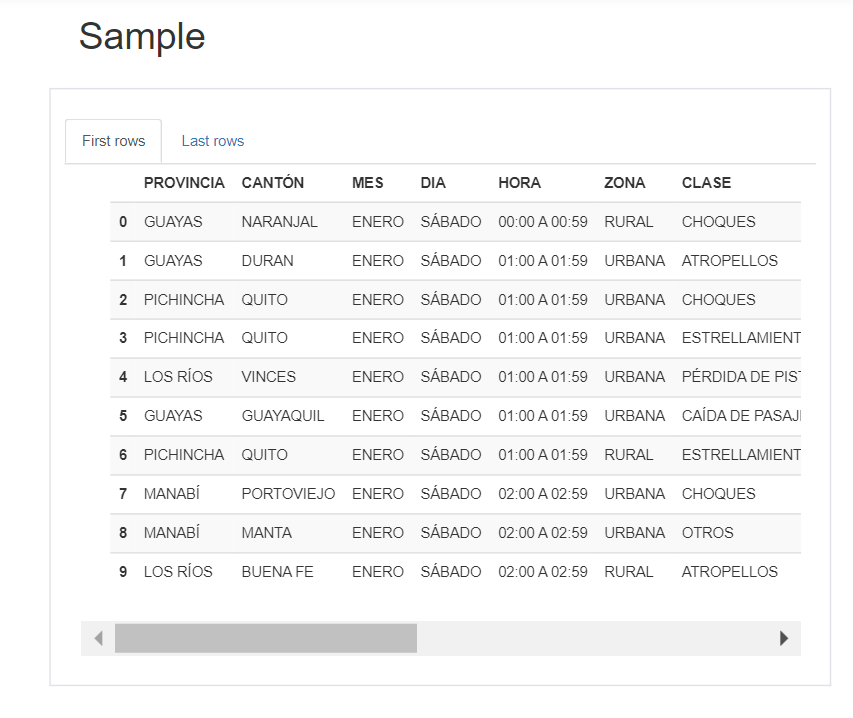
****

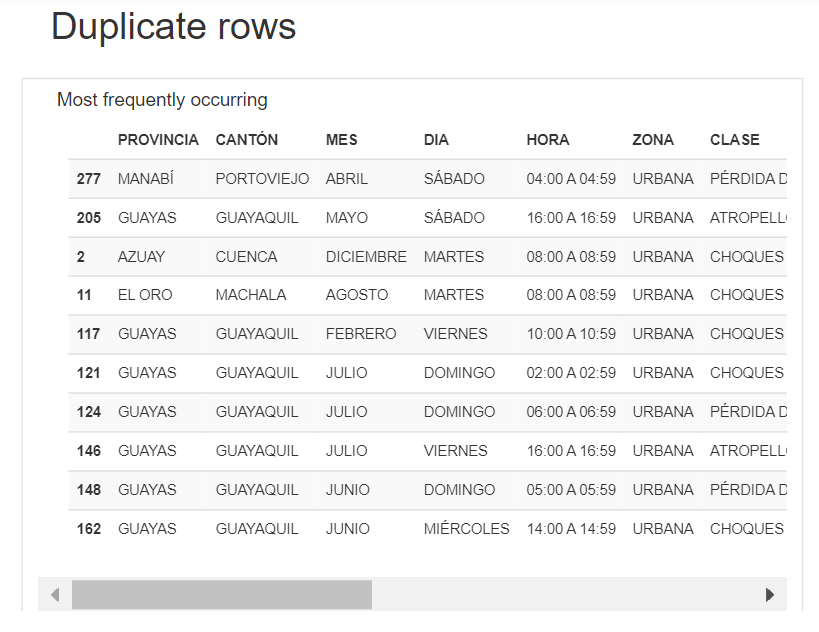
****

****

****

****

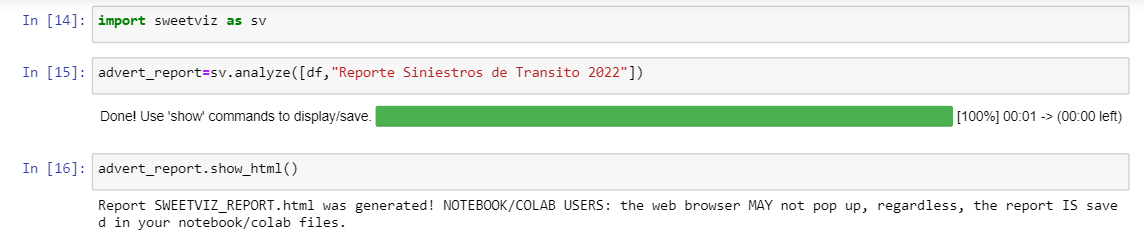
****

****

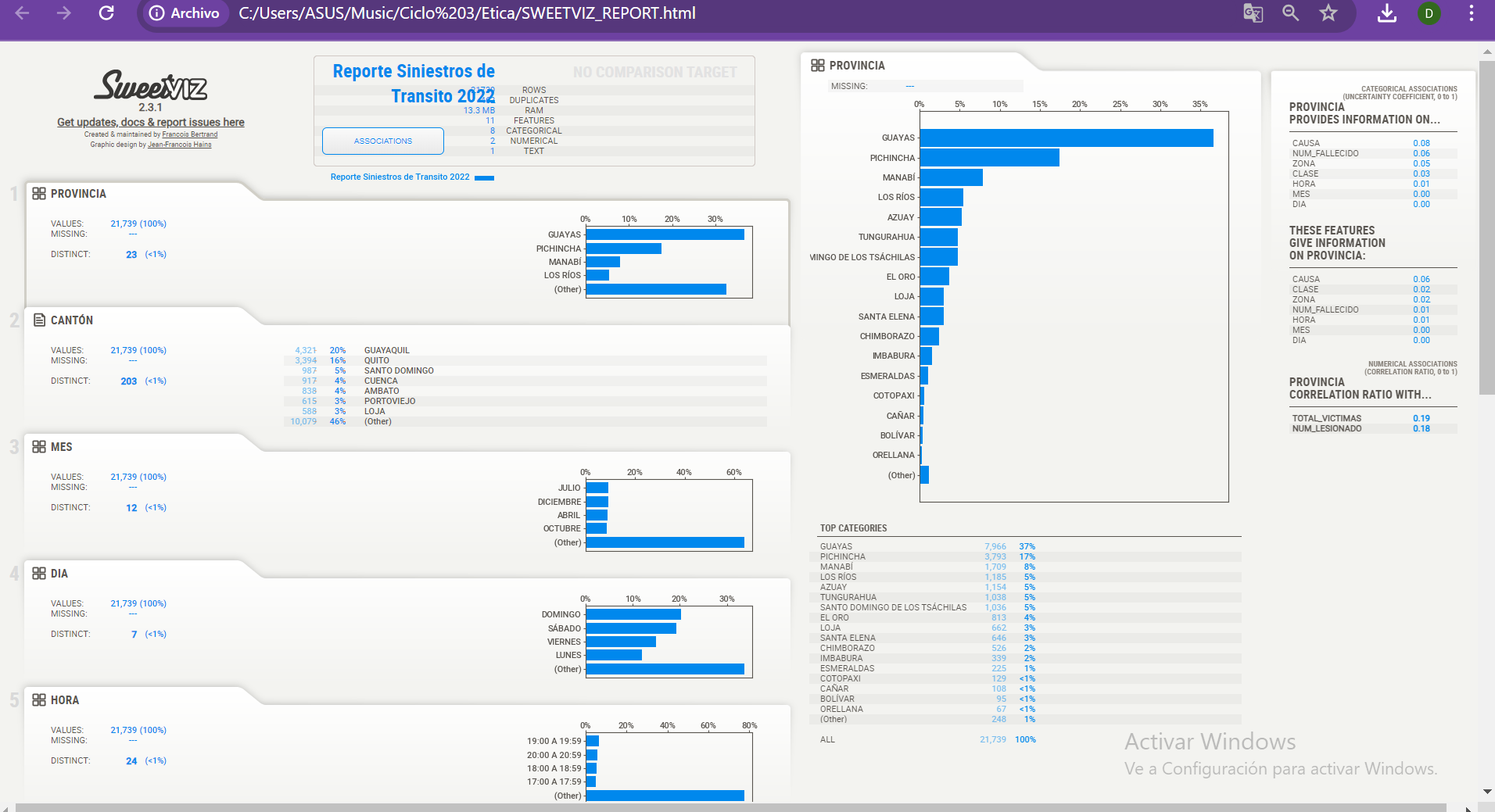
1. CARGA DE LA LIBRERÍA SWEETVIZ

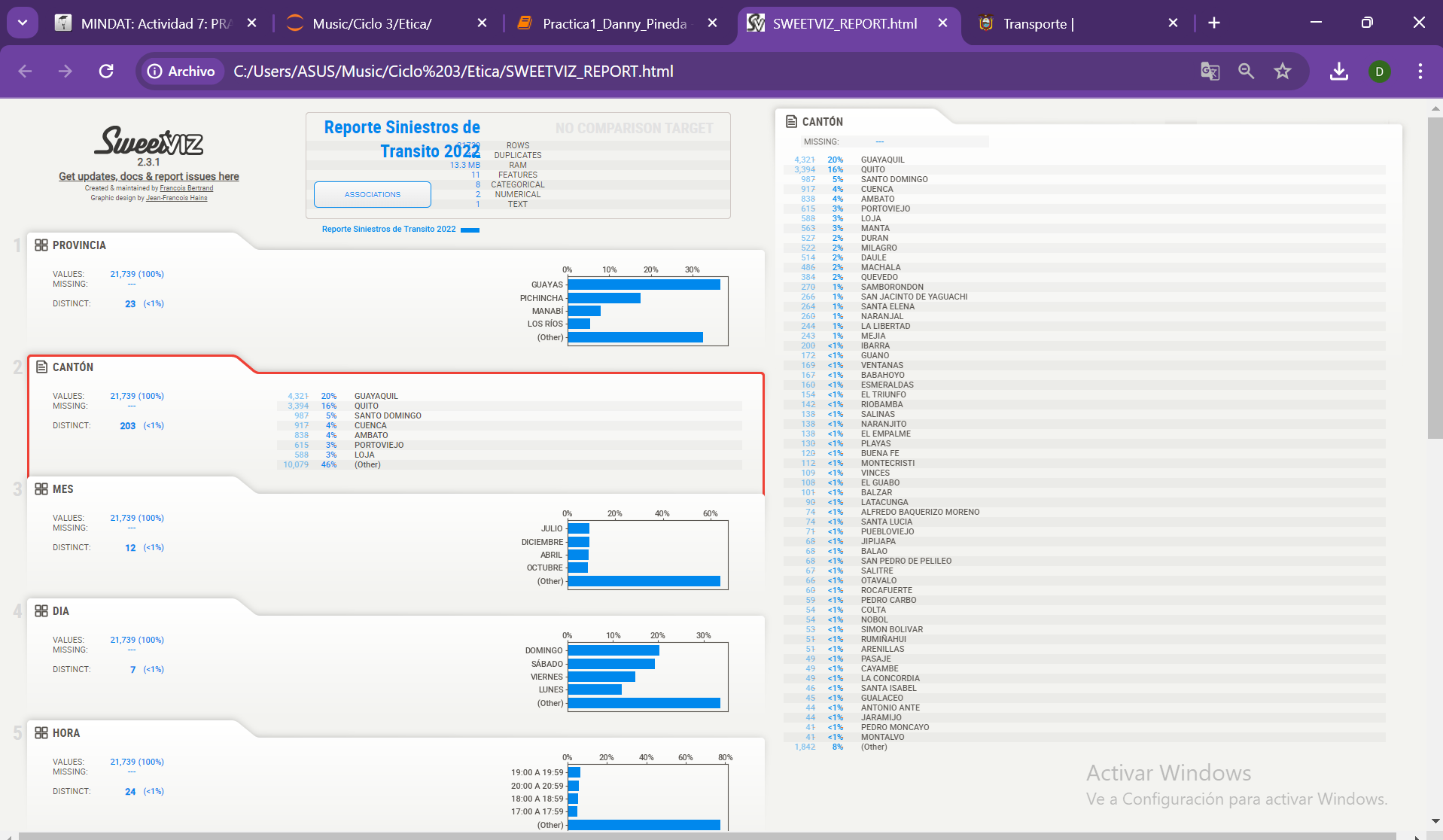
****

1. USO DE LA LIBRERÍA SWEETVIZ Y VISUALIZACION DE RESULTADOS

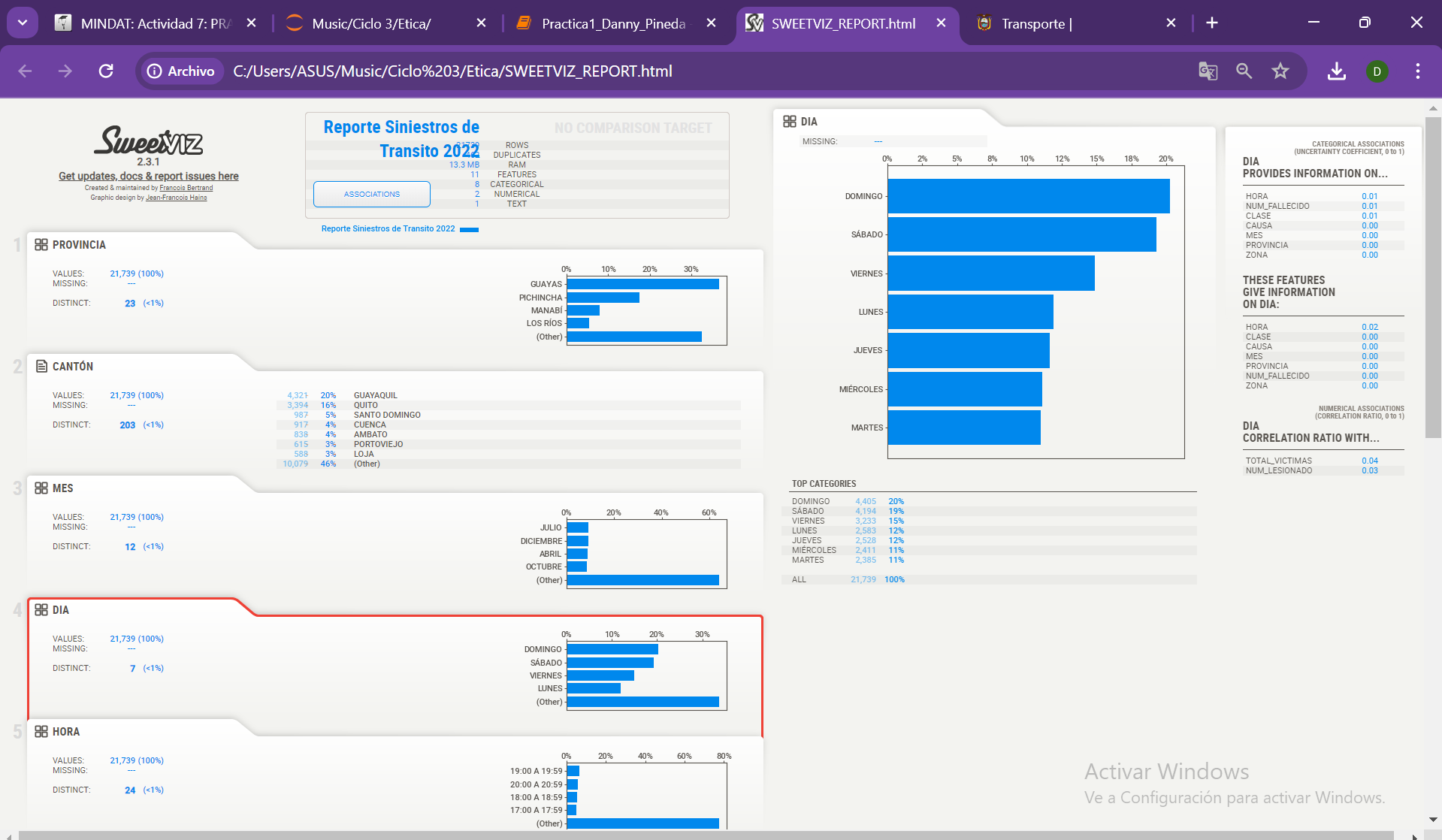


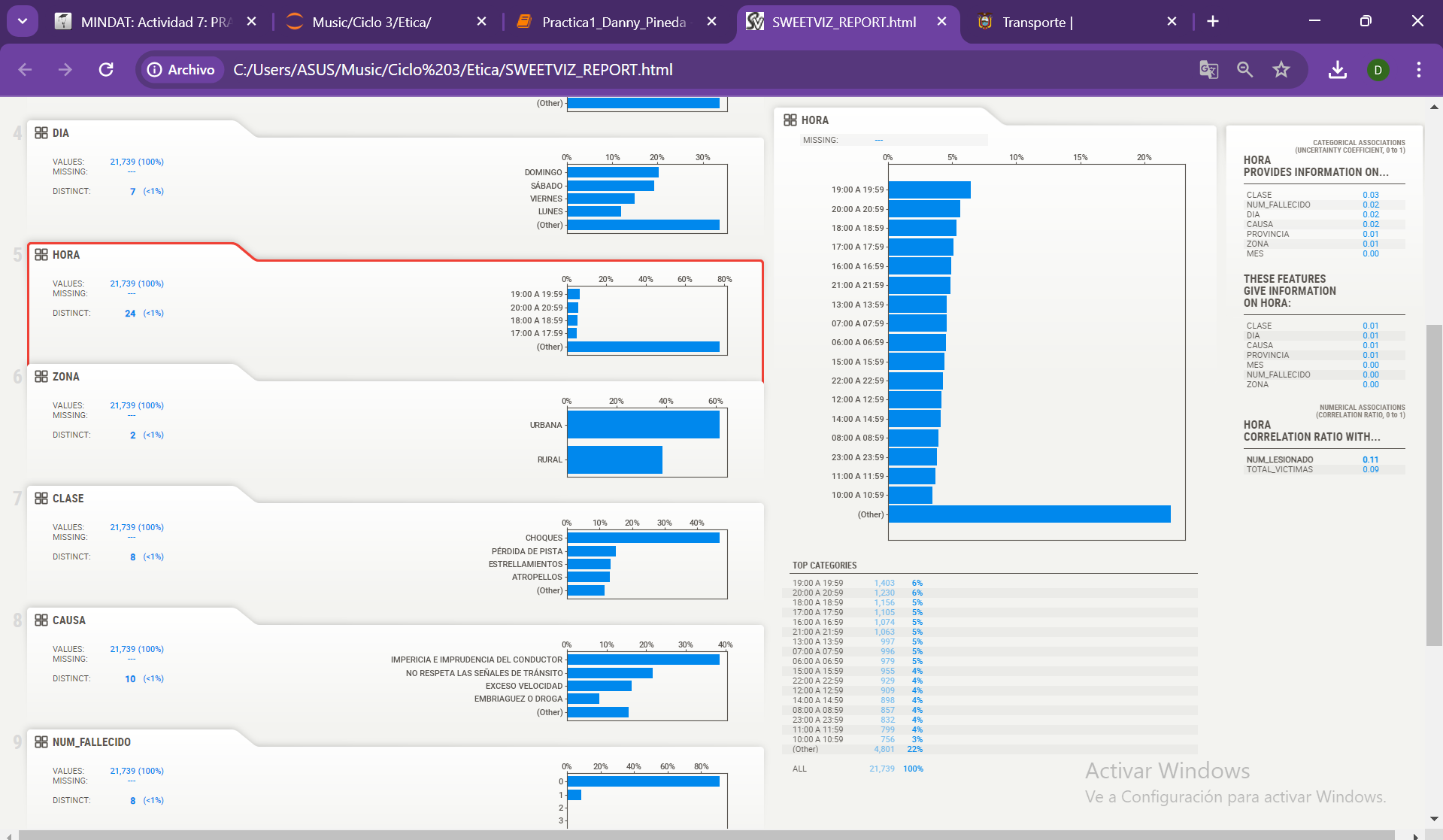
* VISUALIZACION DE RESULTADOS

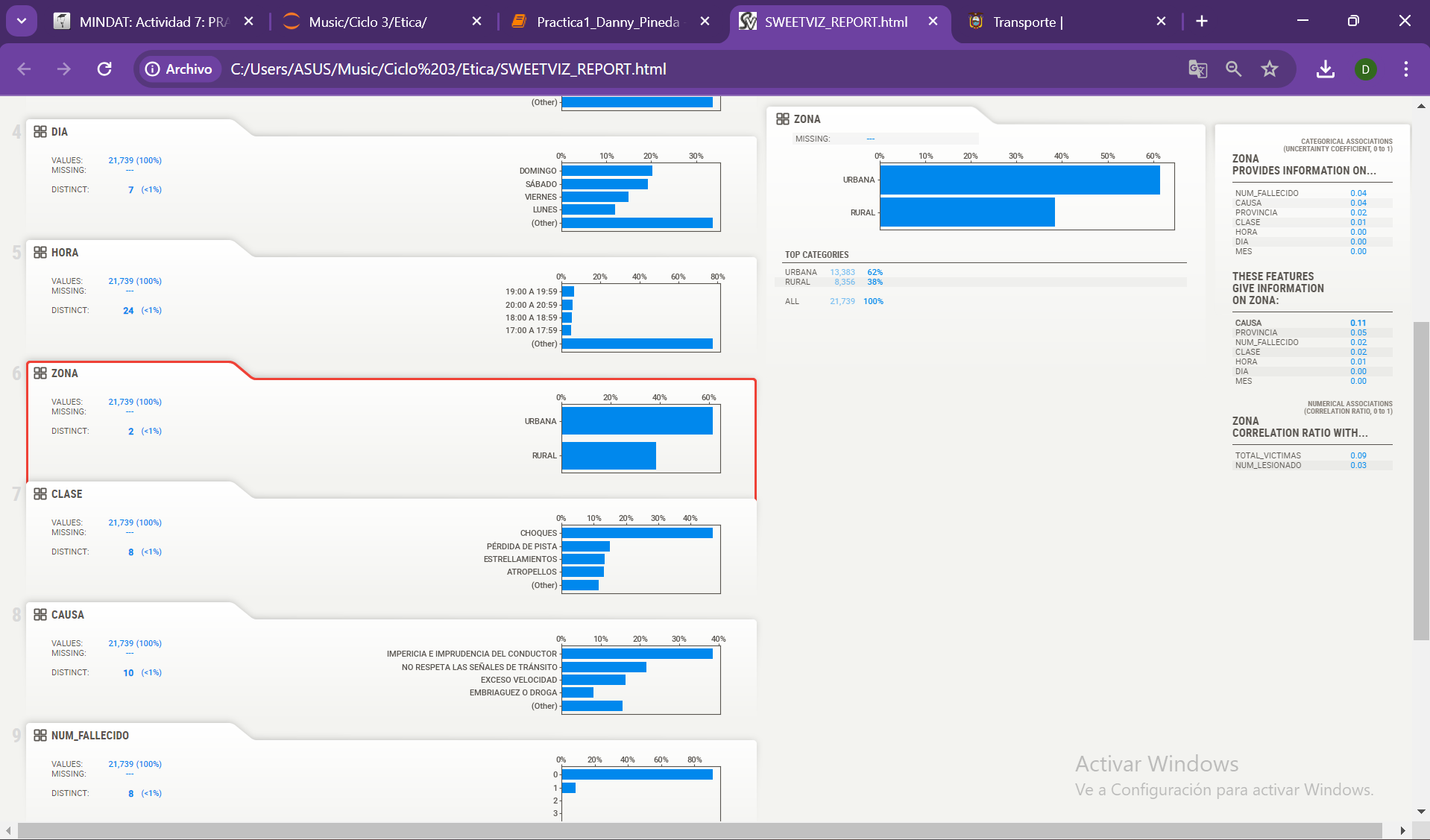
****

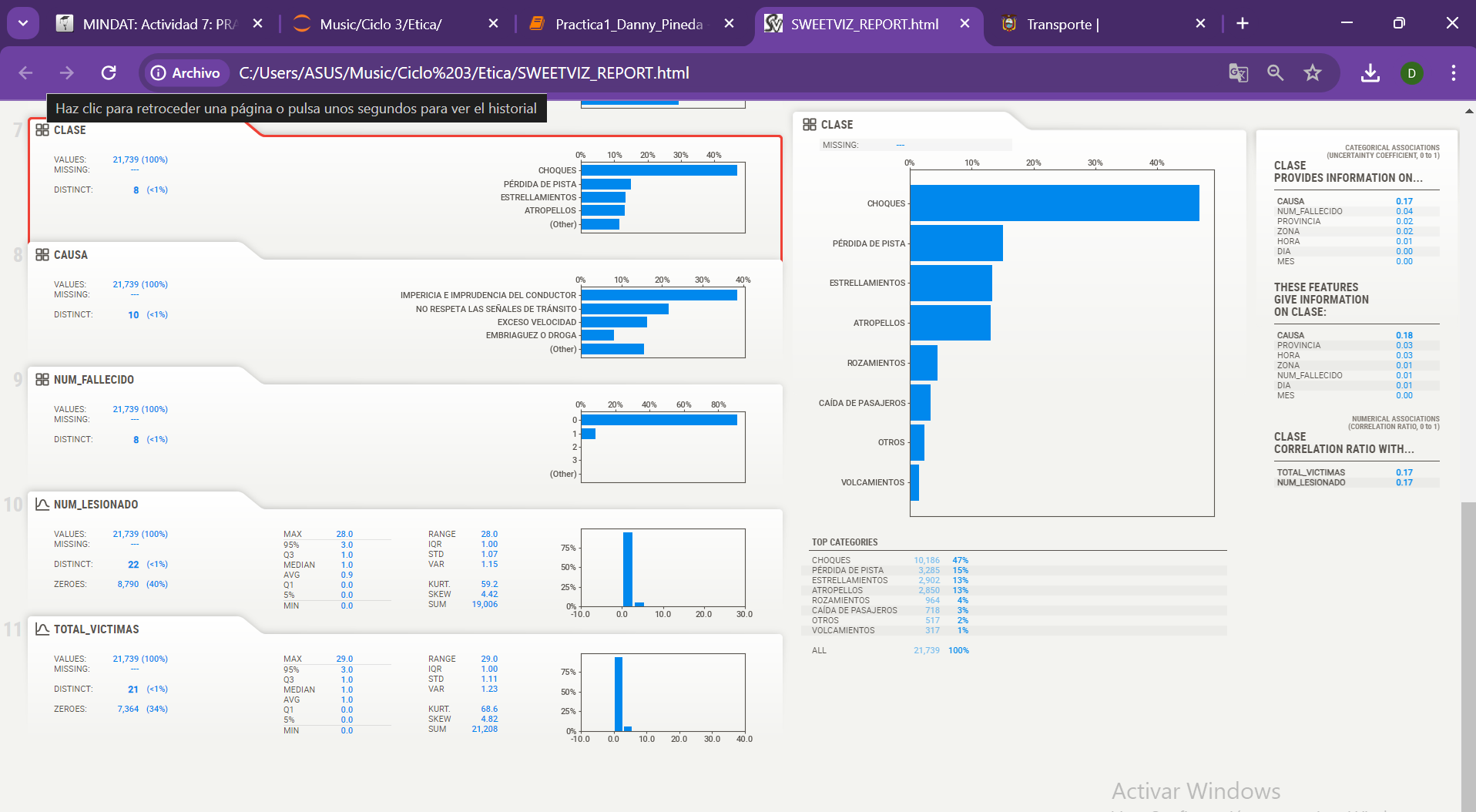
****

****

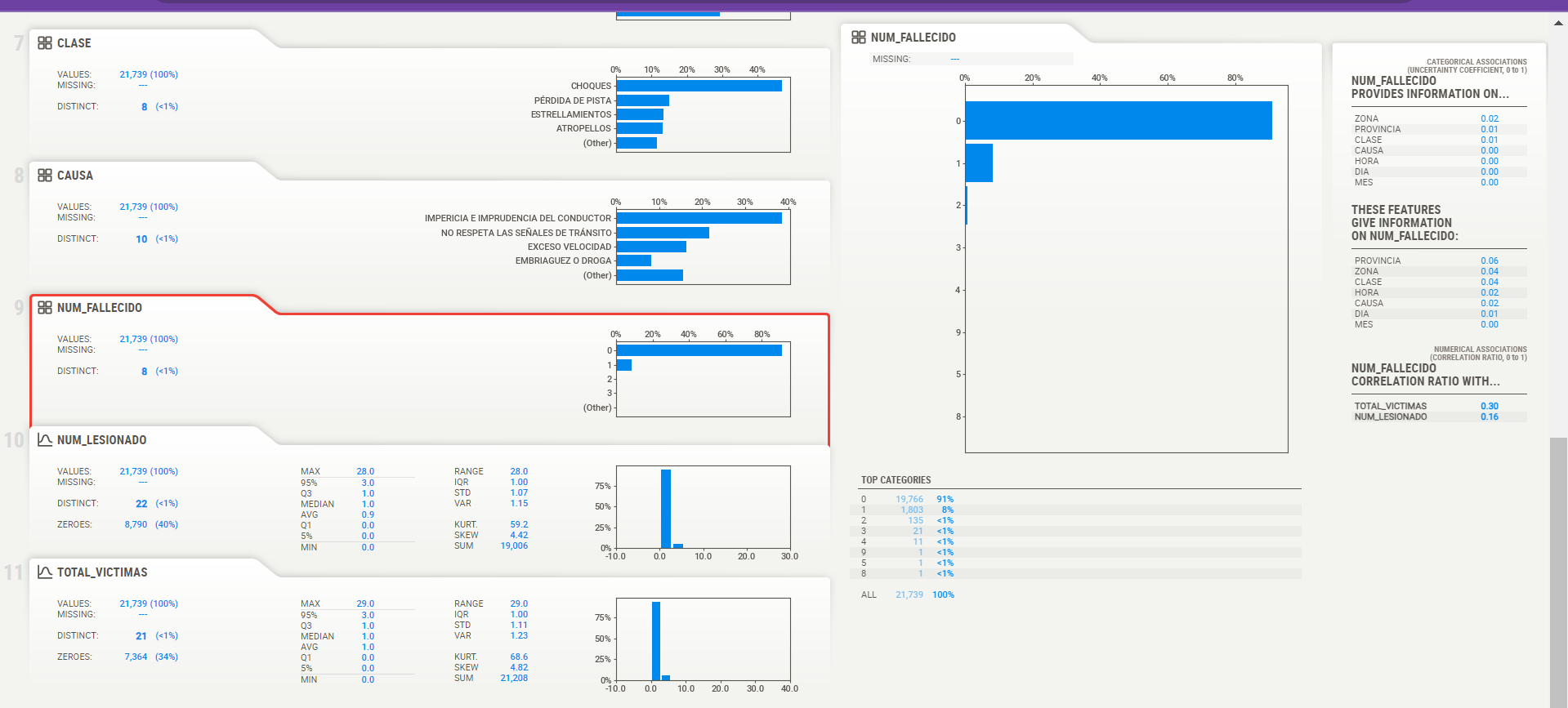
****

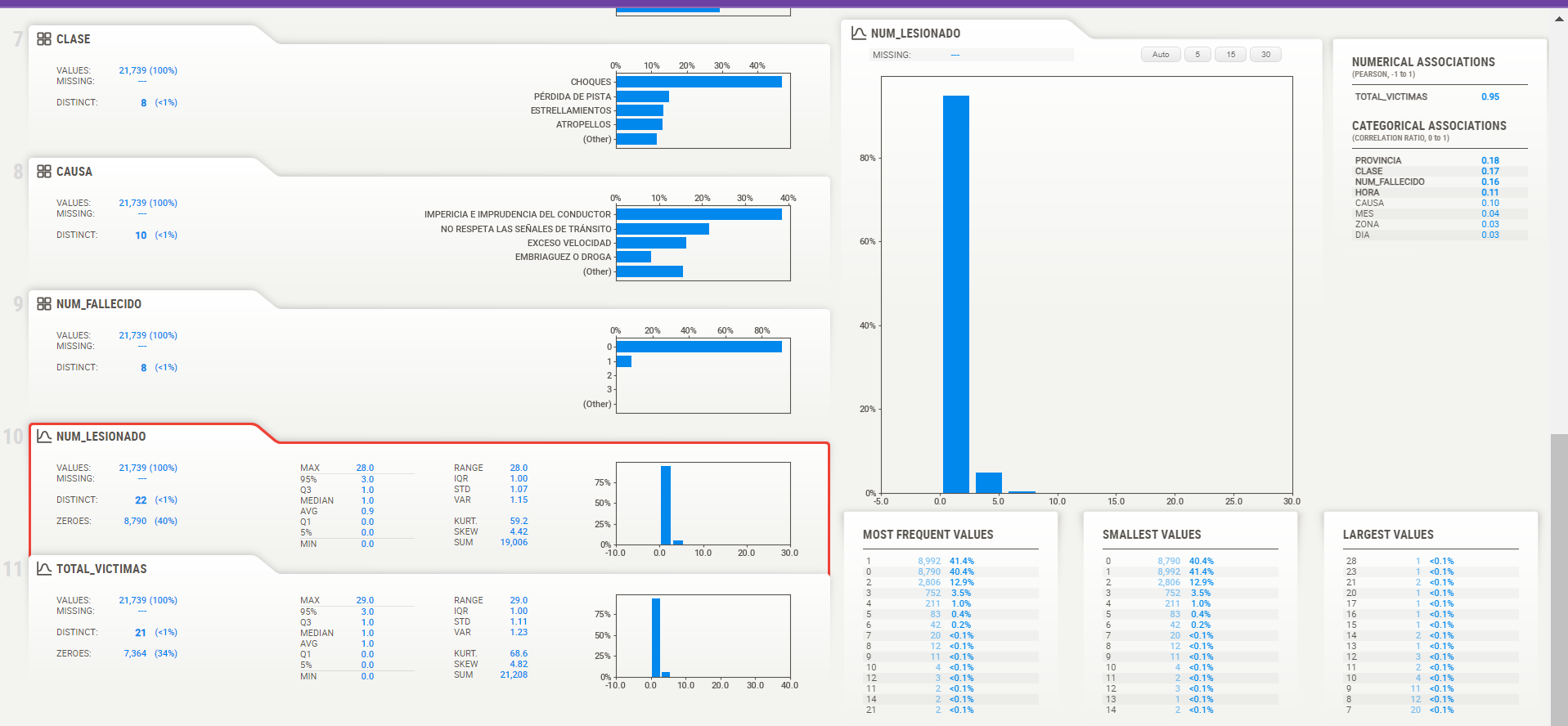
****

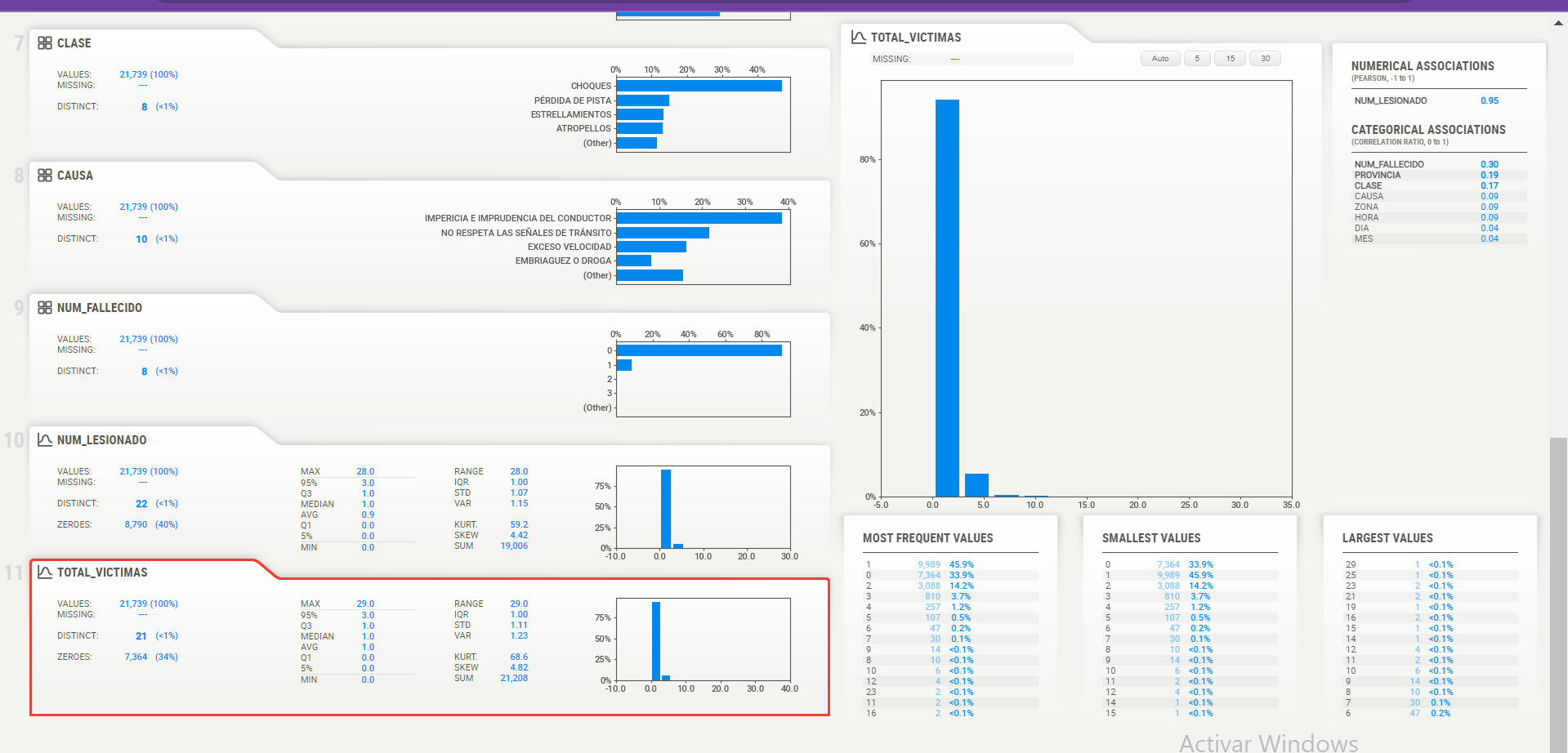
****

****

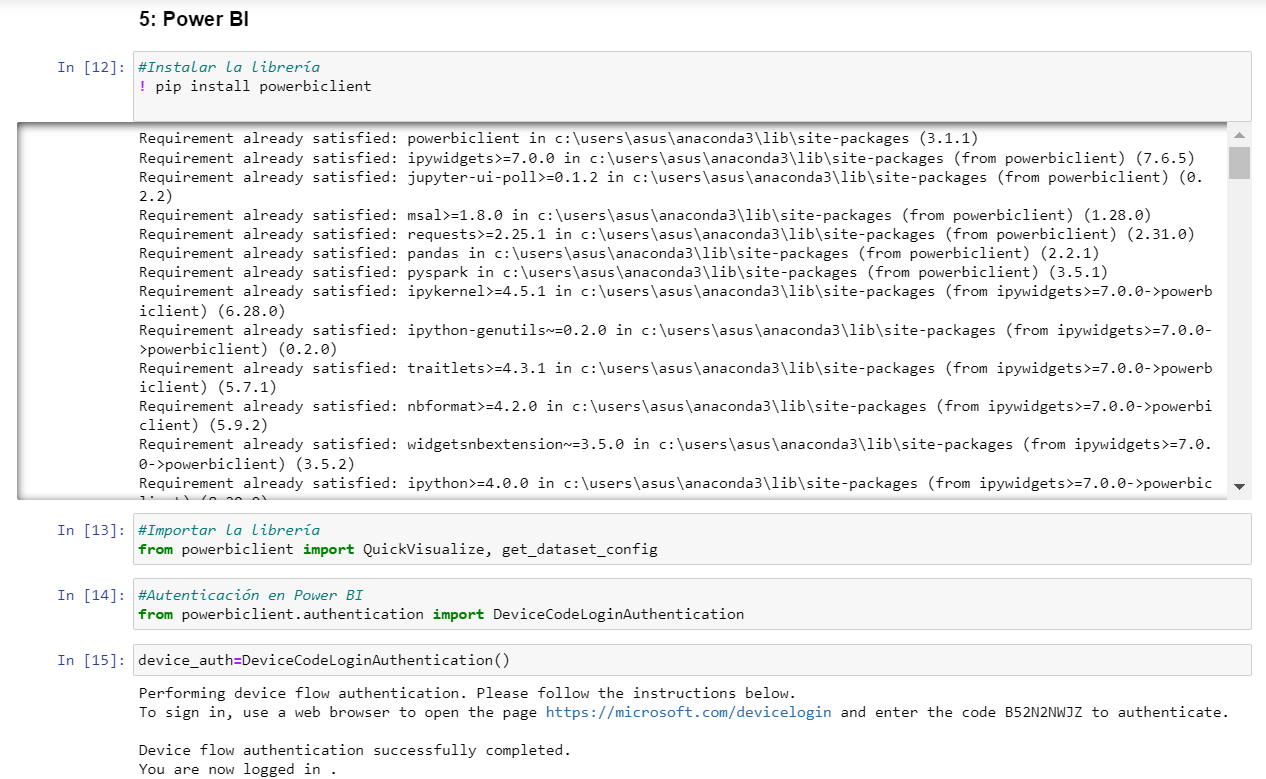
****

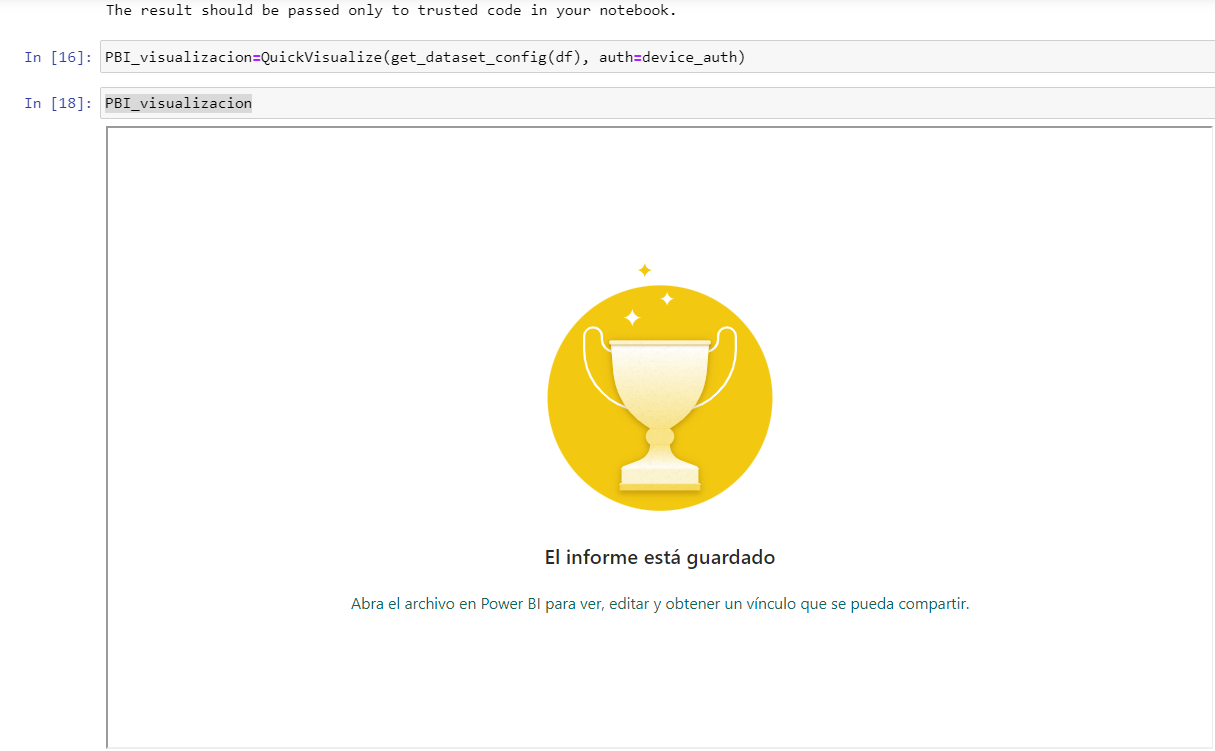
****

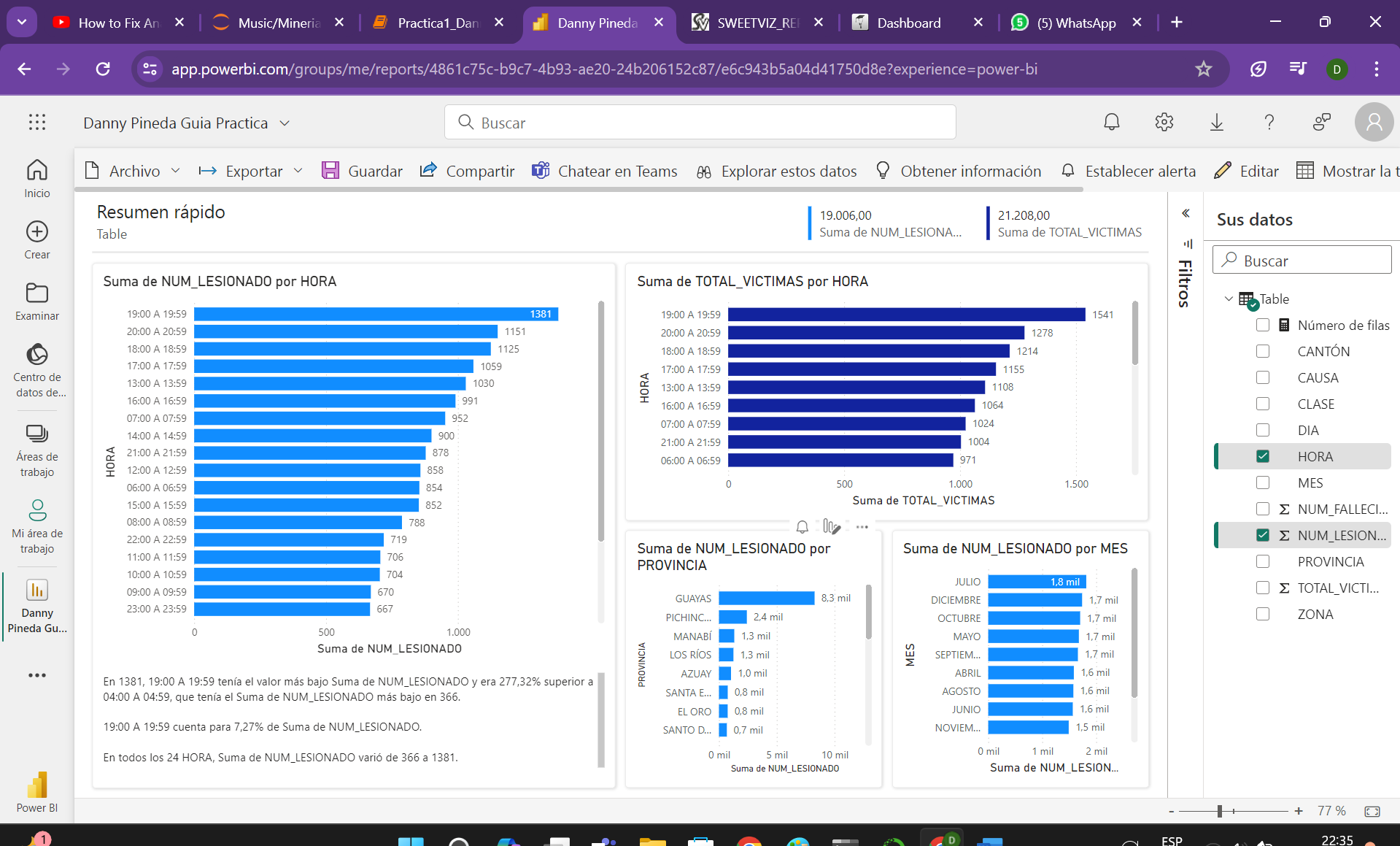
****

****

1. INSTALACION Y USO DE LA LIBRERÍA POWER BI





****

* LINK DEL INFORME DE POWER BI ENVIADO A WHATSAPP
* <https://app.powerbi.com/view?r=eyJrIjoiMGFkZTBhOWMtNjVkYi00OGQ2LTg1MzQtOTAyZTAwODdjZjM4IiwidCI6IjI1NzM5YzY1LWM3OWYtNDAxYy1iYWIwLWQ3NTVlOTBhNjY2MiIsImMiOjR9>
  1. **Normas de Seguridad**

|  |
| --- |
| Las normas de seguridad se han tomado del reglamento general de seguridad para el uso de los talleres, aulas y laboratorios del Instituto Superior Universitario Tecnológico del Azuay.  El estudiante, al ingresar a los talleres o laboratorios, está sujeto a este reglamento; y, tendrá la supervisión del profesor y del personal técnico; será responsable de:  a) Usar los EquipoS de Protección Personal(EPP) de acuerdo con lo establecido en la “Matriz de equipos de protección individual (EPP`s) requeridos para el ingreso de estudiantes y profesores a los laboratorios y talleres del INSTITUTO”;  b) Al inicio de cada práctica, recibir y revisar el material y herramientas requeridas para la Práctica, serán responsables de su buen uso.  c) La operación de los equipos por los estudiantes deberá ser con el conocimiento de su funcionamiento y bajo las directrices del profesor o personal técnico del laboratorio o taller; bajo ninguna circunstancia el estudiante podrá trabajar solo y sin vigilancia;  d) Seguir las instrucciones dadas por el docente o el personal técnico de apoyo;  e) Al término de la práctica, entregar limpio tanto el material como su área de trabajo;  f) Informar inmediatamente al profesor o personal técnico de apoyo, cualquier desperfecto que se localice en los equipos e instalaciones. |

* 1. **Resultados esperados**

|  |
| --- |
| * Tiene habilidades sólidas para la lectura de diferentes formatos de ficheros de datos, utilizando la biblioteca Pandas en Python, permitiendo a los usuarios cargar, explorar y entender datos de manera eficiente. * Aplica técnicas fundamentales de análisis exploratorio de datos, incluida la identificación y manejo de valores faltantes, así como la exploración de estadísticas descriptivas, para obtener una comprensión inicial de los conjuntos de datos. * Visualiza datos utilizando Matplotlib y Seaborn, y explorar la integración de "PowerClient - Quick Report" para generar visualizaciones impactantes que resalten patrones y tendencias de manera clara y efectiva. |

* 1. **Bibliografía**

|  |
| --- |
| **Descripción en norma APA** |
| Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. Morgan Kaufmann.  * Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann. * Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-54. * Tan, P. N., Steinbach, M., & Kumar, V. (2013). Introduction to Data Mining. Pearson. * Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer. |

1. **Firmas de Responsabilidad**

|  |  |  |
| --- | --- | --- |
| **ESTUDIANTE** | **DOCENTE** | **DIRECTORA DE CARRERA** |
| **Nombre: Danny Pineda**  **Firma** | **Nombre:** Mgtr. Verónica Chimbo  **Firma** | **Nombre:** Mgtr. Verónica Segarra  **Firma** |
| **Fecha: ( )** | **Fecha:** (04/06/2023) | **Fecha: ( )** |