

TAREA UNIDAD 2 TALLER

SOFTWARE PARA DATA SCIENCE | UCIF 1042

**Profesor**: Patricio Galdames

**Alumno**: Maria Paz Cisternas Pardo

19.242.027-K

**Fecha**: 07/06/2024

Tabla de contenido

[Tabla de Ilustraciones 1](#_Toc168220494)

[1.1 CONTEXTO: 2](#_Toc168220495)

[1.2 OBJETIVO: 2](#_Toc168220496)

[2. REQUERIMIENTOS Y PROGRESO 3](#_Toc168220497)

[2.1 MODULO: "DAL.PY" 3](#_Toc168220498)

[2.2 MODULO: "DAO.PY" 3](#_Toc168220499)

[2.1 MODULO: "DTO.PY" 4](#_Toc168220500)

[2.3 PROBLEMAS PRESENTADOS: 6](#_Toc168220501)

[3.1 MAIN.PY 7](#_Toc168220502)

[3.2 GENERADOR\_ENCUESTA.PY 7](#_Toc168220503)

[3.2 PROCESADOR \_ENCUESTA.PY 8](#_Toc168220504)

# Tabla de Ilustraciones

[Ilustración 1 Método de filtrado por medio de Dataframe 5](#_Toc164644481)

[Ilustración 2 main.py 6](#_Toc164644482)

[Ilustración 3 generador\_encuestas.py ; generación de encuesta 6](#_Toc164644483)

[Ilustración 4 generador\_encuestas.py ; Archivo csv generado desde app de consola 7](#_Toc164644484)

[Ilustración 5 procesador\_encuestas.py ; menú de inicio 7](#_Toc164644485)

[Ilustración 6 procesador\_encuestas.py ; Función Estadísticas Generales 8](#_Toc164644486)

[Ilustración 7 procesador\_encuestas.py ; Estadísticas Generales con filtro de Sexo 8](#_Toc164644487)

[Ilustración 8 procesador\_encuestas.py ; Estadísticas Generales con filtro de rango de edad 9](#_Toc164644488)

[Ilustración 9 procesador\_encuestas.py ; archivo de texto con los datos procesados 9](#_Toc164644489)

1. PROYECTO

## 1.1 CONTEXTO:

Una tienda de electrónica desea realizar un análisis de sus datos de ventas y clientes para obtener información relevante y tomar decisiones estratégicas. Los datos de ventas incluyen información sobre el ID de venta, el ID de cliente, el nombre del producto, la categoría, el precio unitario y la cantidad vendida. Los datos de clientes incluyen información sobre el ID de cliente, el nombre, la edad, el género y la ubicación.

## 1.2 OBJETIVO:

Aplicar los conceptos de manipulación y análisis de datos utilizando la biblioteca Pandas de Python para procesar los datos de ventas y clientes de la tienda, cruzarlos y generar informes que respondan a las preguntas planteadas.

# 2. REQUERIMIENTOS Y PROGRESO

Se entiende como requerimiento lo solicitado en el documento “Tarea Unidad 2 Taller de Software para Data Science”, donde se decidió trabajar con arquitectura MVC con el propósito de simplificar la mantenibilidad e interacción en la ejecución en consola, donde **datos.py** generara los archivos csv tanto para clientes como para ventas, **limpiar\_datos.py** se encargará de preparar los datos para su posterior análisis, **modelado.py** para procesar los datos, **visualización.py** presentará por medio de gráficos los análisis estadísticos y **main.py** será la capa de presentación al consumir los módulos.

## 2.1 MODULO: "DATOS.PY"

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Detalle** | **Estado** |
| **1** | Genera archivo extensión csv con mínimo 50 filas, debe ser ingresad por el usuario y validada esta condición. |  |
| **2** | En el archivo generado, el delimitador debe ser espacio (“ ”) y la primera fila de etiquetas. |  |
| **3** | Las filas deben ser con los siguientes valores:   * Sexo (str) : “Femenino” o “Masculino” * Edad(int) : entre 18 y 45 años * Respuesta (str): “Si”,”No”,”Tal vez” |  |
| **4** | Generar datos aleatoriamente en las 3 columnas con el módulo *random()* |  |

## 2.2 MODULO: "LIMPIAR\_DATOS.PY"

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Detalle** | **Estado** |
| **1** | Entrada de usuario que solicite el archivo a procesar |  |
| **2** | Importar los módulos necesarios para leer el archivo CSV y realizar análisis de datos básicos |  |
| **3** | Generar Diccionario que contenga los datos con cabeceras como claves y sus respectivos datos. |  |
| **4** | Acondicionar datos para su correcto funcionamiento |  |
| **5** | Calcular estadística descriptiva utilizando ciclos y condicionantes. |  |
| **6** | Contar frecuencia de Respuesta utilizando ciclos y condicionantes. |  |
| **7** | Generar menú interactivo utilizando ciclos:   1. Leer archivo de datos 2. Mostrar estadísticas generales (media, mediana, conteo de respuestas por tipo) 3. Filtrar datos por sexo y mostrar estadísticas, permita al usuario especificar el sexo o el rango de edad y muestre las estadísticas 4. Filtrar datos por rango de edad y mostrar estadísticas, permita al usuario especificar el sexo o el rango de edad y muestre las estadísticas 5. Guardar resultados en un archivo (todos aquellos generados) 6. Salir Para las opciones de filtrado, genere un archivo de texto que contenga las estadísticas generales y los resultados de los filtros aplicados. |  |

## 2.3 MODULO: "MODELADO.PY"

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Detalle** | **Estado** |
| **1** | Genera archivo extensión csv con mínimo 50 filas, debe ser ingresad por el usuario y validada esta condición. |  |
| **2** | En el archivo generado, el delimitador debe ser espacio (“ ”) y la primera fila de etiquetas. |  |
| **3** | Las filas deben ser con los siguientes valores:   * Sexo (str) : “Femenino” o “Masculino” * Edad(int) : entre 18 y 45 años * Respuesta (str): “Si”,”No”,”Tal vez” |  |
| **4** | Generar datos aleatoriamente en las 3 columnas con el módulo *random()* |  |

## 2.4 MODULO: "VISUALIZACION.PY"

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Detalle** | **Estado** |
| **1** | Genera archivo extensión csv con mínimo 50 filas, debe ser ingresad por el usuario y validada esta condición. |  |
| **2** | En el archivo generado, el delimitador debe ser espacio (“ ”) y la primera fila de etiquetas. |  |
| **3** | Las filas deben ser con los siguientes valores:   * Sexo (str) : “Femenino” o “Masculino” * Edad(int) : entre 18 y 45 años * Respuesta (str): “Si”,”No”,”Tal vez” |  |
| **4** | Generar datos aleatoriamente en las 3 columnas con el módulo *random()* |  |

## 2.3 PROBLEMAS PRESENTADOS:

Intenté usar filtros con diferentes enfoques: usé Filter en el diccionario de datos, lambda en el mismo y for anidados. Sin embargo, obtuve la columna asociada en lugar de las filas con el filtro utilizado.

 Para solucionarlo, se usó la librería Pandas, convirtiendo el diccionario en un DataFrame y aplicando los filtros con éxito para posteriormente convertir el Dataframe resultante en un diccionario, lo cual simplificó los cálculos al reutilizar la función de cálculos estadísticos generales.

Texto

Descripción generada automáticamente

Ilustración 1 Método de filtrado por medio de Dataframe

3. PRESENTACIÓN DE LA SOLUCIÓN

## 3.1 MAIN.PY

La siguiente imagen presenta la ejecución de main.py donde solicita al usuario ingresar una opción de tarea que desea ejecutar.

Texto

Descripción generada automáticamente

Ilustración 2 main.py

## 3.2 GENERADOR\_ENCUESTA.PY

A continuación se visualiza la llamada al módulo generador\_encuesta, probando la validación “filas > 50” y luego generando el archivo encuesta\_habitos.csv

Texto

Descripción generada automáticamente

Ilustración 3 generador\_encuestas.py ; generación de encuesta

Aquí se puede ver el archivo csv generado con la función Random.

Interfaz de usuario gráfica

Descripción generada automáticamente

Ilustración 4 generador\_encuestas.py ; Archivo csv generado desde app de consola

## 3.2 PROCESADOR \_ENCUESTA.PY

En esta imagen se visualiza la llamada al módulo procesador\_encuesta, desplegando el menú, donde la opción a) solicita el nombre del archivo a procesar, almacenándolo en la memoria temporal.

Texto

Descripción generada automáticamente

Ilustración 5 procesador\_encuestas.py ; menú de inicio

En esta imagen se ve las estadísticas generales calculadas desde el archivo leído; en este caso se incluyó la moda para las columnas Sexo y Respuestas.

Forma

Descripción generada automáticamente con confianza media

Ilustración 6 procesador\_encuestas.py ; Función Estadísticas Generales

Aquí se puede visualizar las estadísticas generales calculadas con el filtro de Sexo. Para reducir el riesgo de error de tipeo se utilizó un sistema de opciones

Texto

Descripción generada automáticamente con confianza media

Ilustración 7 procesador\_encuestas.py ; Estadísticas Generales con filtro de Sexo

En este caso se visualiza las estadísticas generales calculadas con el filtro por rango de Edad.

Forma

Descripción generada automáticamente con confianza media

Ilustración 8 procesador\_encuestas.py ; Estadísticas Generales con filtro de rango de edad

A continuación, el archivo escrito con la información almacenada.

Texto

Descripción generada automáticamente

Ilustración 9 procesador\_encuestas.py ; archivo de texto con los datos procesados

4. CONCLUSIÓN

Python es una herramienta versátil y eficiente, demostrando que posee la capacidad de analizar datos y procesarlos para generar informes utilizando librerías básicas. Tiene la capacidad de analizar granes volúmenes de datos y existen librerías para realizar cálculos complejos además de hacer vistas atractivas.