

## **Proyecto Beatrix: Manual de Usuario**

Andres.P Olivar, Andrés.F Pino y Wilmer.U Velasco

TEDESOF, Universidad del Valle

Estadística y Probabilidad

Danny Calvo

06 de Junio de 2024

## **Introducción**

Beatrix-statistics es un paquete estadístico desarrollado en el lenguaje de programación Python que proporciona una gran variedad de herramientas y funciones para el análisis de datos. Este proyecto se desarrolló con el objetivo de analizar grandes cantidades de datos a través de un archivo de formato CSV, aplicando todos los conceptos vistos en clases y llevándolos a un nivel funcional. También se creó con un diseño amigable e intuitivo para el usuario, con el fin de facilitar tareas de análisis estadísticos comunes.

## **Funcionalidades principales**

**1. Analizar cantidades masivas de datos:** Beatrix cuenta con un algoritmo que le permite analizar una cantidad extremadamente grande de datos que se puede encontrar distribuida entre 6 columnas que no tienen que estar relacionadas entre sí directamente. Ej. una columna de temperatura de un cuerpo y otra de presión atmosférica.

**2. Cálculo de medidas de tendencia central:** Capacidad de calcular la media, mediana, moda y definir los cuartiles.

**3. Cálculo de medidas de variabilidad:** Beatrix tiene la capacidad de calcular también la varianza, la desviación estándar, los percentiles, contando también con la capacidad para calcular el porcentaje de datos que se encuentre entre 2 números en el teorema de Chebyshev a partir de las desviaciones estándar entre los números.

**4. Generación de tablas y gráficos:** Capacidad de graficar tablas de frecuencia, diagramas de barras, gráficos del teorema de Chebyshev y diagramas de caja.

**5. Aplicación de teoremas estadísticos:** Utilización de la regla empírica, el teorema de Chebyshev, entre otros conceptos estadísticos para analizar conjuntos de datos.

**6. Almacenar los resultados del análisis en un archivo de texto:** Con el fin de conservar los resultados de los análisis estadísticos para un futuro uso, Beatrix almacena todos los cálculos que se hagan en un archivo, recalando que el usuario es capaz de definir el nombre de este archivo.

## **Ingreso de datos por un archivo .CSV**

Un ejemplo de la forma en que se ingresan los datos a Beatrix sería la siguiente imagen,

NO.	Fecha/hora	Intervalo	Temperatura interior	Humedad interior	Temperatura
1192	21/09/2018 08:57	5	24.6	64	24.6
1193	21/09/2018 09:00	5	24.1	60	23.6
1194	21/09/2018 09:02	5	24.7	64	24.6
1195	21/09/2018 09:05	5	24.1	60	23.6
1196	21/09/2018 09:07	5	24.7	64	24.6

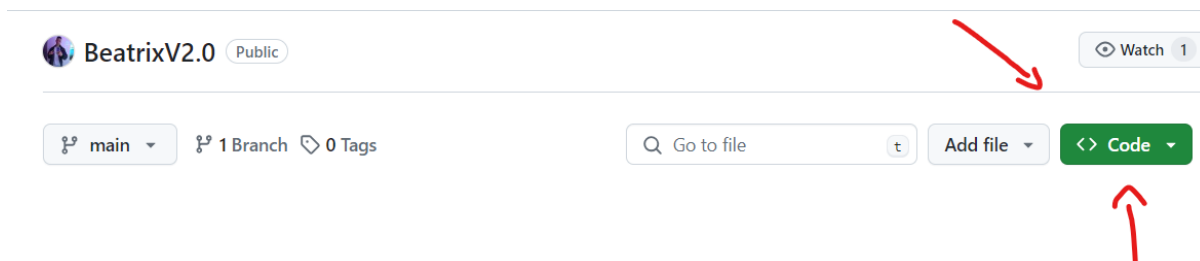
El ingreso de datos se realiza a través de un archivo CSV, que se caracteriza por tener todos los datos de una fila separado por comas ",", Beatrix analiza el documento fila por fila, omitiendo la primera fila que contiene los títulos, y almacena los datos entre las posiciones 3 y 8 (índices 2 y 7) en listas separadas. Luego, el usuario puede elegir a cuáles de las columnas o datos quiere aplicar el análisis.

## Paso a Paso para usar Beatrix

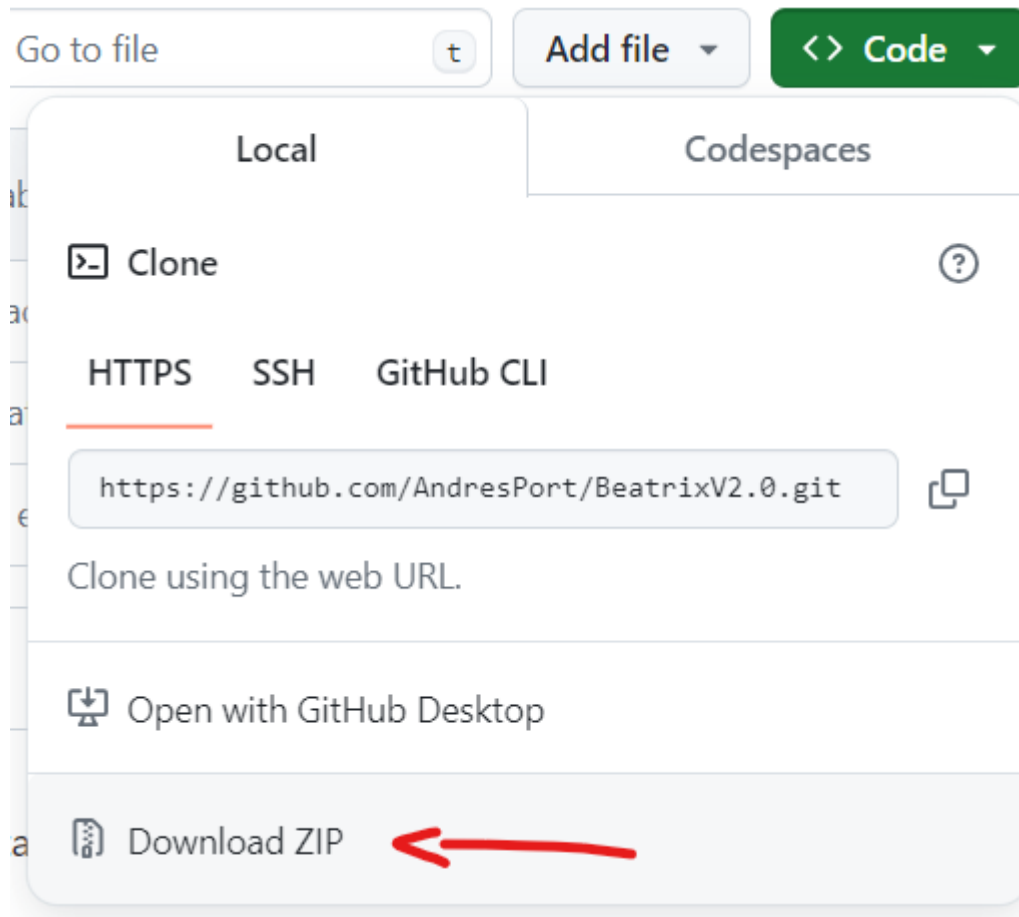
### 1.Descarga e Instalación

Lo primero es descargar el paquete, para esto, ingresa a

<https://github.com/AndresPort/BeatrixV2.0.git>. Una vez allí, da click en el botón "<code>"



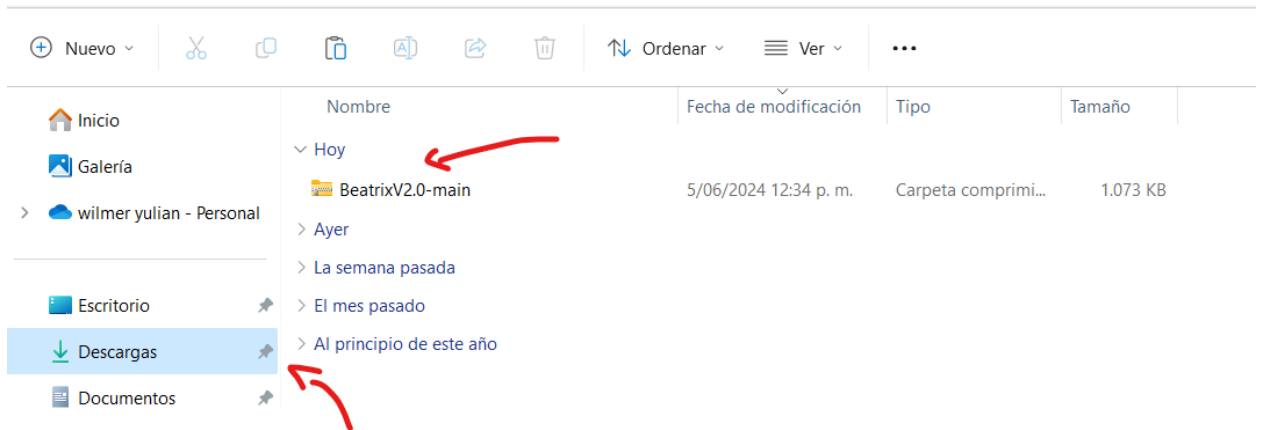
El cual desplegará unas opciones, de la cual vas a escoger "Download ZIP"



Una vez hecho eso automáticamente se iniciará la descarga.

## 1.2 Descomprimir

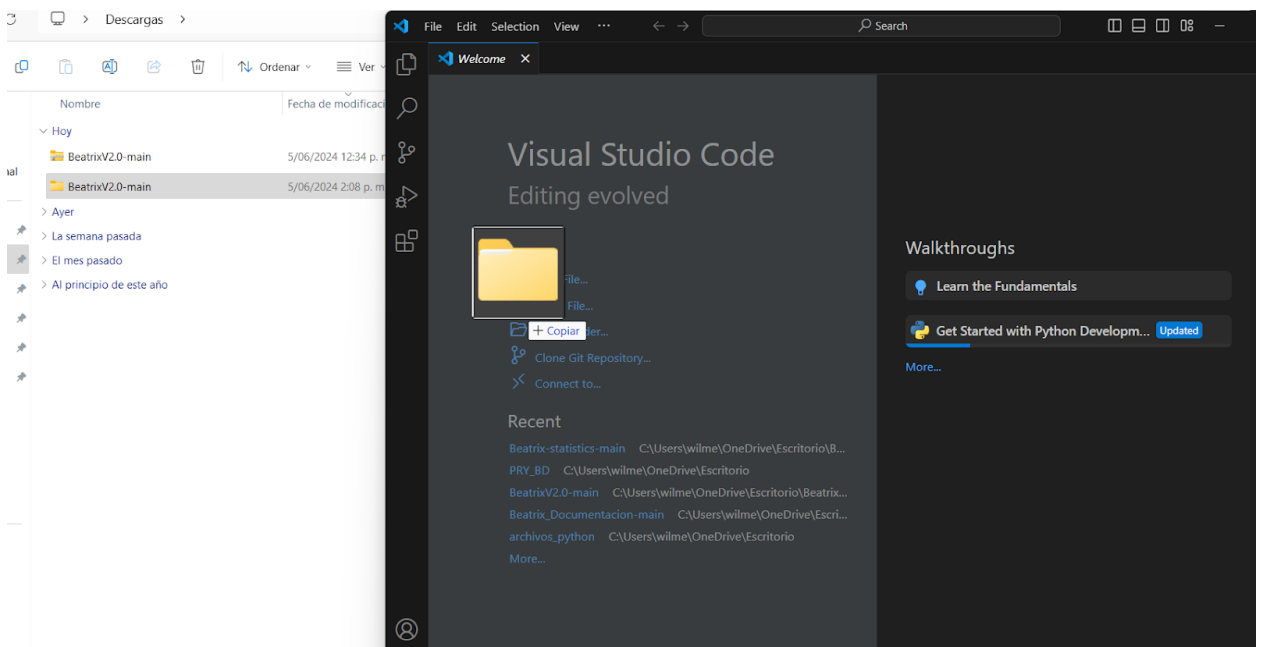
Dirijase a sus archivos, navegue hasta descargas y descomprima el archivo mediante alguna herramienta(como WinRAR):



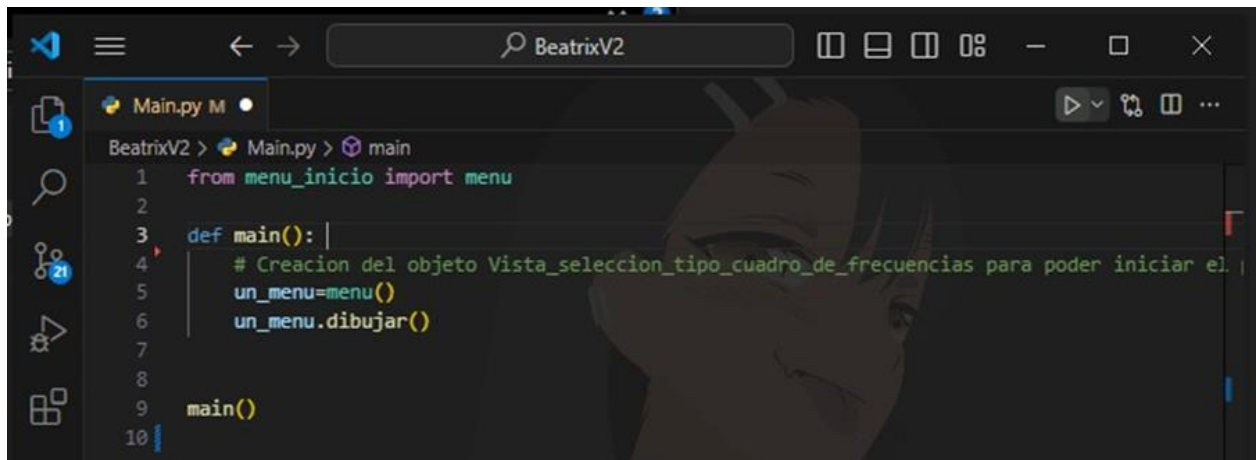
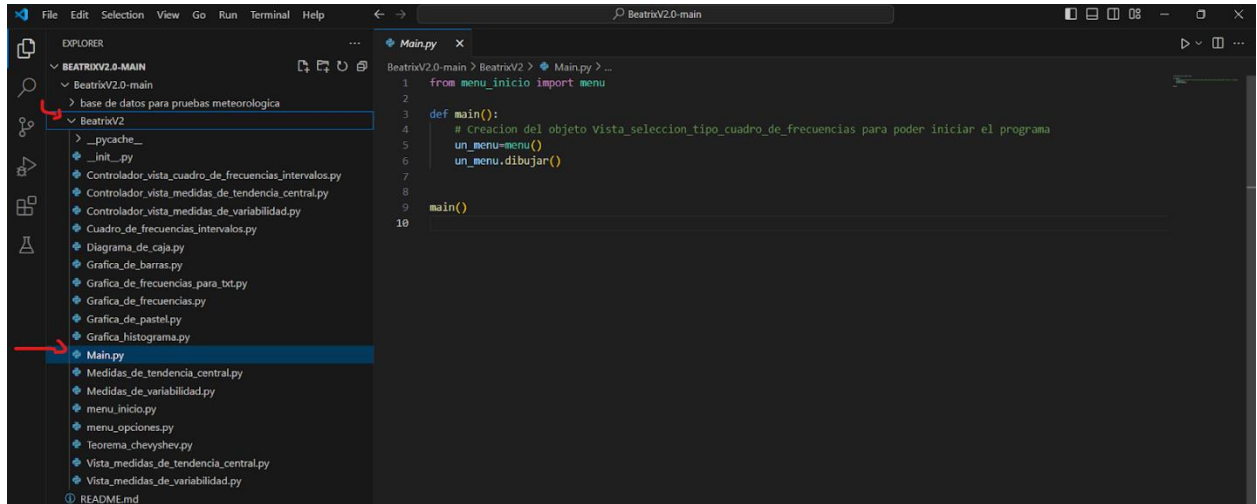
## 2 Ejecución y funcionamiento del programa

Para ejecutar Beatrix existen 2 formas: Ejecutar el archivo Main desde un IDE o ejecutarlo desde un directorio para que funcione como un programa.

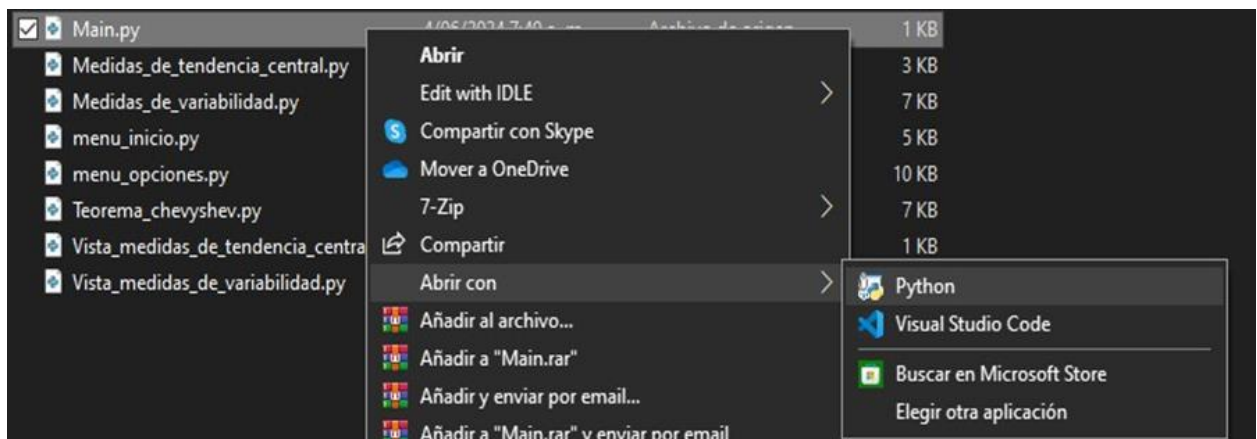
**Desde un IDE (por ejemplo, Visual Studio Code):** Abra Visual Studio Code y arrastre o abra la carpeta que descomprimió en el paso 1.2:



Luego despliegue y busque el **main.py**:



Desde un directorio para que funcione como un programa:



Tras ejecutar el archivo se verá de la siguiente manera si se ejecutó desde un IDE:





Y se verá de esta forma si se ejecutó desde un directorio en una terminal externa o del sistema:



The image shows a Windows terminal window titled "C:\Windows\py.exe". The terminal has a black background with yellow text. The text is as follows:

```
WELCOME  
TO  
BEATRIX
```

---

```
POR FAVOR INGRESE EL NOMBRE QUE QUIERA QUE  
TENGA EL ARCHIVO FINAL CON LOS RESULTADOS
```

On the left side of the terminal, there is a vertical list of 20 entries, each consisting of a left arrow, the text "[7;30;45m", and a right arrow. On the right side, there is a vertical list of 20 entries, each consisting of a left arrow, the text "[0m", and a right arrow. The terminal window has standard Windows window controls (minimize, maximize, close) in the top right corner.

## 2.1 Ingreso del nombre del archivo

- De ambas formas le pedirá que ingrese el nombre que se le asignará al archivo de texto



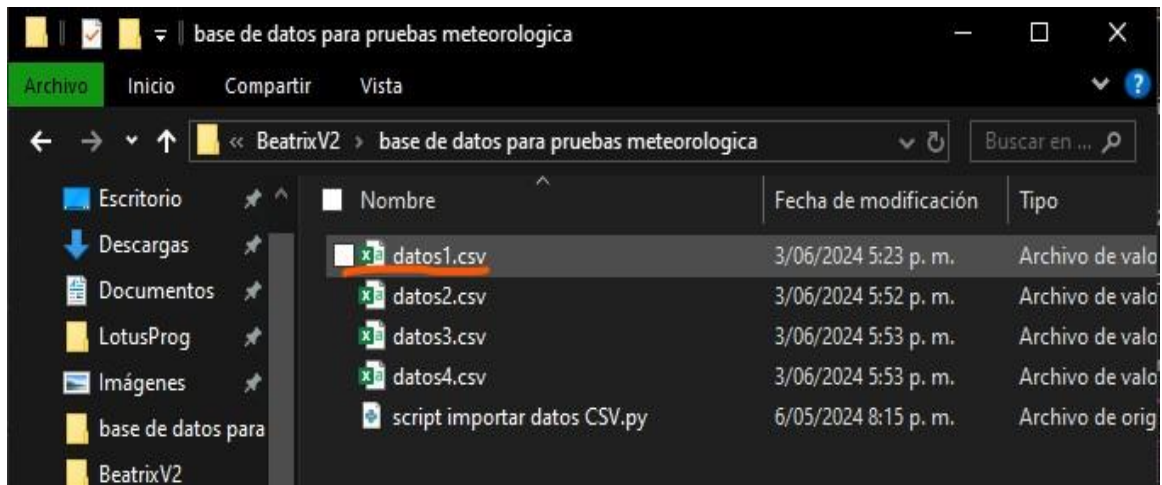
en el que se guardarán los resultados del análisis, por lo que solo queda ingresarlo y presionar ENTER.

Tras ingresar el nombre y presionar ENTER le pedirá que ingrese la dirección raíz del archivo con finalización .csv para empezar el análisis.

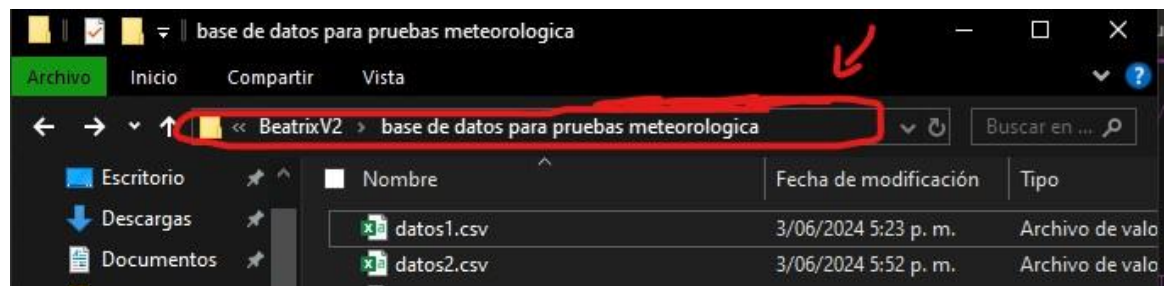
## 2.2 Ingresar la dirección raíz del archivo CSV

Para obtener la dirección raíz del archivo deberá hacer lo siguiente:

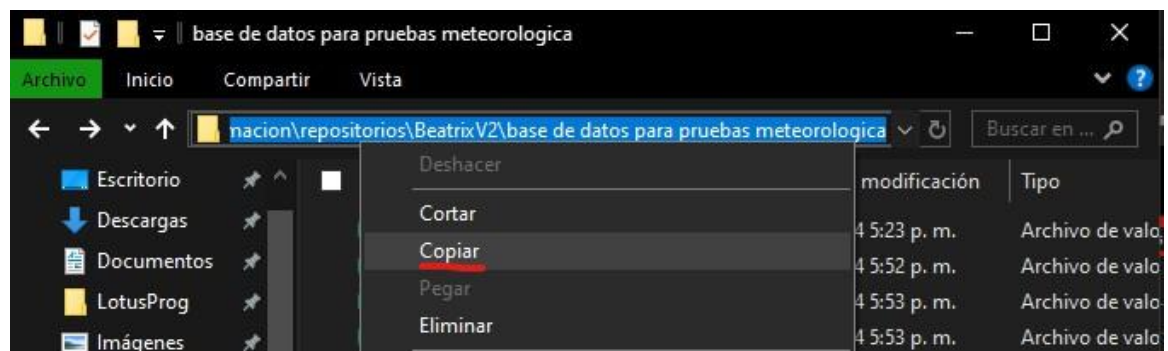
- Posicionarse en la carpeta en donde se encuentra el archivo. En este caso se van a usar los datos que se encuentran en el archivo datos1.csv



-Dar click izquierdo en la parte superior para mostrar la dirección de esa carpeta



-Al presionar en la parte indicada se marcará en azul la dirección raíz de la carpeta, por lo que solo deberá presionar click derecho y copiarla



Pegar la dirección raíz en la terminal en donde se ejecute el programa. En el caso en que no pueda usar el click derecho presione las teclas Ctrl y la tecla V al mismo tiempo

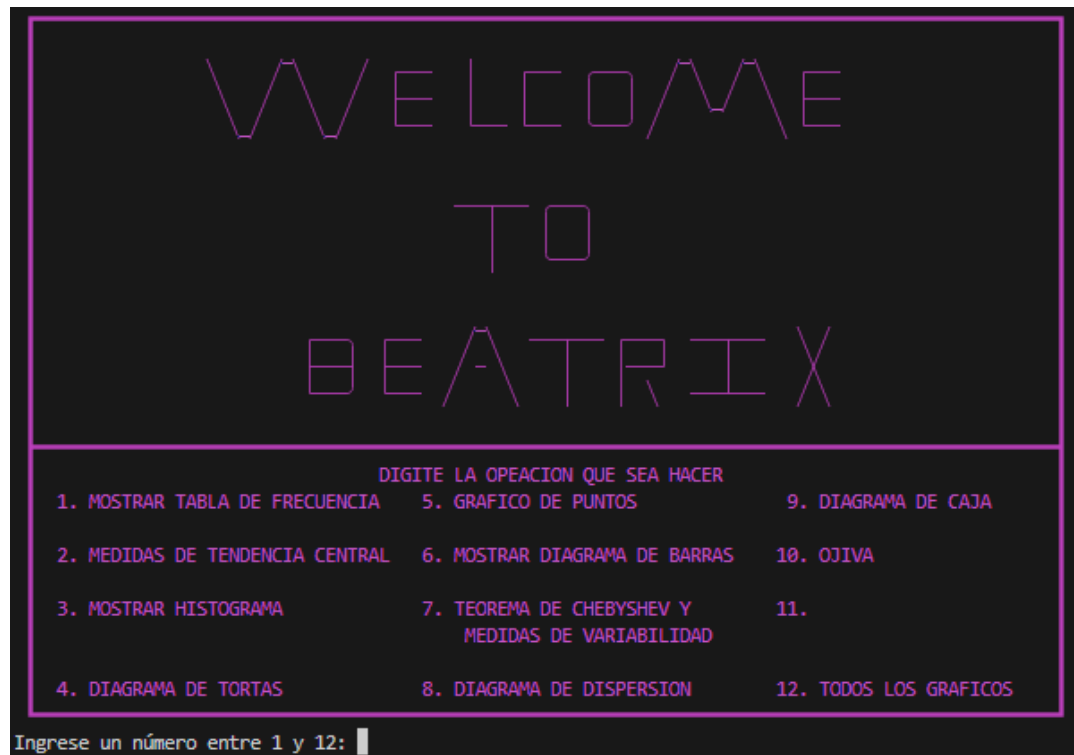
```
resultados analisis 1
Por favor ingrese la raiz del archivo .csv a evaluar, asegurese de que los datos se encuentren seaparados por comas ',' Ej:
D://Programacion//repositorios//BeatrixV2//base de datos para pruebas meteorologica//datos1.csv
D:\Programacion\repositorios\BeatrixV2\base de datos para pruebas meteorologica
```

5. Cambiar todos los caracteres "\" por "/"y al final agregar el nombre del archivo junto con la extensión .csv y luego presionar ENTER

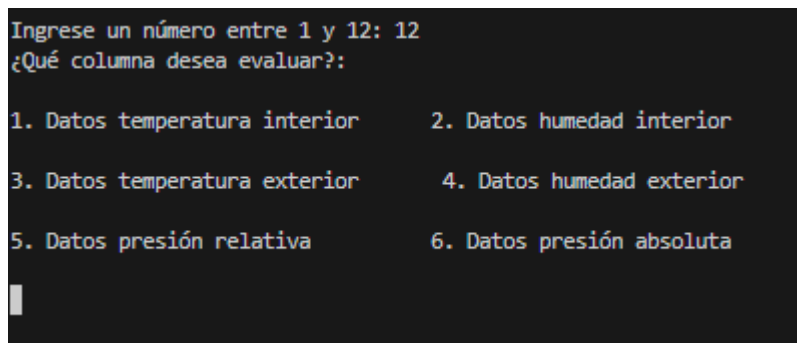
```
resultados analisis 1
Por favor ingrese la raiz del archivo .csv a evaluar, asegurese de que los datos se encuentren seaparados por comas ',' Ej:
D://Programacion//repositorios//BeatrixV2//base de datos para pruebas meteorologica//datos1.csv
D://Programacion//repositorios//BeatrixV2//base de datos para pruebas meteorologica//datos1.csv
```

## 2.3 Ejecución

Luego de ingresar la dirección raíz y presionar ENTER aparecerá un menú con las opciones a elegir, permitiéndole ingresar un número entre 1 y 12:



- Luego de ingresar un número entre 1 y 12 le permitirá elegir a cuál de las 6 columnas desea emplear para el análisis



5. Ingreso de la cantidad de intervalos a emplear

- A continuación el programa le pedirá al usuario que ingrese la cantidad de intervalos

que vea conveniente para el análisis de los datos, contando con la restricción de que el número que se va a ingresar debe ser mayor o igual a 5

```

4. DIAGRAMA DE TORTAS      8. DIAGRAMA DE DISPERSION      12. TODOS LOS GRAFICOS
Ingrese un número entre 1 y 12: 12
¿Qué columna desea evaluar?:
1. Datos temperatura interior    2. Datos humedad interior
3. Datos temperatura exterior    4. Datos humedad exterior
5. Datos presión relativa        6. Datos presión absoluta
1
Por favor ingrese el numero de intervalos con los que se va a analizar los datos, mínimo 5:

```

## 6. Resultados

7. Finalmente le mostrará la cantidad total de datos analizados del documentos, en este caso datos1, junto con las operaciones que seleccionó previamente

Total de datos analizados: 36673						
INTERVALO	FRECUENCIA	FR	FRA	FRP	FRPA	F°

## Pruebas de Funcionamiento

A continuación se muestran pruebas de que funciona cada una de las opciones que se dan en el menú:

1 Se utiliza el primer data set, opción 1: tablas de frecuencia, y se escogió los datos de la temperatura interior con 10 intervalos:

Resultado:

1

Por favor ingrese el numero de intervalos con los que se va a analizar los datos, minimo 5: 10  
 Total de datos analizados: 36673

INTERVALO	FRECUENCIA	FR	FRA	FRP	FRPA	F°
21.5-22.1	276	0.0075	0.007	0.75	0.75	2.7
22.2-22.8	1641	0.0447	0.052	4.47	5.22	16.09
22.9-23.5	3202	0.0873	0.14	8.73	13.95	31.43
23.6-24.2	9395	0.2562	0.396	25.62	39.57	92.23
24.3-24.9	11300	0.3081	0.704	30.81	70.38	110.92
25.0-25.6	7792	0.2125	0.916	21.25	91.63	76.5
25.7-26.3	2366	0.0645	0.981	6.45	98.08	23.22
26.4-27.0	477	0.013	0.994	1.3	99.38	4.68
27.1-27.7	192	0.0052	0.999	0.52	99.9	1.87
27.8-28.4	32	0.0009	1.0	0.09	99.99	0.32

Ln 10, Col 9, Spaces: 4, UTF-8

Si nos dirigimos a nuestra carpeta donde tenemos los archivos, podemos ver el documento txt que genero el programa con los resultados:

Nombre	Fecha de modificacion	tipo	tamaño
<div> <div> <div></div> <div>Hoy</div> </div> <div> <div></div> <div>prueba_tabla_frecuencia</div> </div> </div>	5/06/2024 10:51 p. m.	Archivo	12 KB
<div> <div></div> <div>BeatrixV2.0-main</div> </div>	5/06/2024 2:08 p. m.	Carpeta de archivos	

Lo abrimos y vemos que efectivamente se guardaron los archivos:

Archivo    Editar    Ver						
numero de intervalos seleccionado: 10						
Total de datos analizados: 36673						
INTERVALO	FRECUENCIA	FR	FRA	FRP	FRPA	F°
21.5-22.1	276	0.0075	0.007	0.75	0.75	2.7
22.2-22.8	1641	0.0447	0.052	4.47	5.22	16.09
22.9-23.5	3202	0.0873	0.14	8.73	13.95	31.43
23.6-24.2	9395	0.2562	0.396	25.62	39.57	92.23
24.3-24.9	11300	0.3081	0.704	30.81	70.38	110.92
25.0-25.6	7792	0.2125	0.916	21.25	91.63	76.5
25.7-26.3	2366	0.0645	0.981	6.45	98.08	23.22
26.4-27.0	477	0.013	0.994	1.3	99.38	4.68
27.1-27.7	192	0.0052	0.999	0.52	99.9	1.87
27.8-28.4	32	0.0009	1.0	0.09	99.99	0.32

2. Se utiliza el primer data set, opción 2: medidas de tendencia central y los datos de humedad interior, 5 intervalos:

```

1. Datos temperatura interior      2. Datos humedad interior
3. Datos temperatura exterior      4. Datos humedad exterior
5. Datos presión relativa          6. Datos presión absoluta

2
Por favor ingrese el numero de intervalos con los que se va a analizar los datos, minimo 5: 5
Total de datos analizados: 36673

Media: 63.11
Mediana: 64
Moda: La moda es (60-68) con 13612 repeticiones

Cuartil 1: 57.0
Cuartil 2: 64.0
Cuartil 3: 69.0
Cuartil 4: 82

PS C:\Users\wilme\OneDrive\Escritorio\BeatrixV2.0-main_ultimo>

```



Se registra correctamente en el archivo de texto:

```
grafico seleccionado: 2
Que columna desea evaluar?:
1. Datos temperatura interior      2. Datos humedad interior
3. Datos tempeatura exterior      4. Datos humedad exterior
5. Datos presion relativa          6. Datos presion_ bsoluta
numero de columna seleccionado: 2
numero de intervalos seleccionado: 5
Total de datos analizados: 36673
Medidas de tendencia central
Media: 63.11
Mediana: 64
Moda: La moda es (60-68) con 13612 repeticiones
Cuartil 1: 57.0
Cuartil 2: 64.0
Cuartil 3: 69.0
Cuartil 4: 82
```

3. se utiliza el primer data set, opción 3: mostrar histograma, datos de la temperatura exterior y 8 intervalos:

5. Datos presión relativa

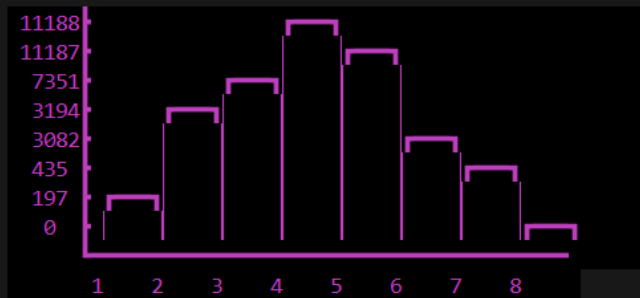
6. Datos presión absoluta

3

Por favor ingrese el numero de intervalos con los que se va a analizar los datos, minimo 5: 8

Total de datos analizados: 36634

Grafica histograma



1 = 21.2-22.0  
2 = 22.1-22.9  
3 = 23.0-23.8  
4 = 23.9-24.7  
5 = 24.8-25.6  
6 = 25.7-26.5  
7 = 26.6-27.4  
8 = 27.5-28.3

Se guarda correctamente en el archivo de texto:

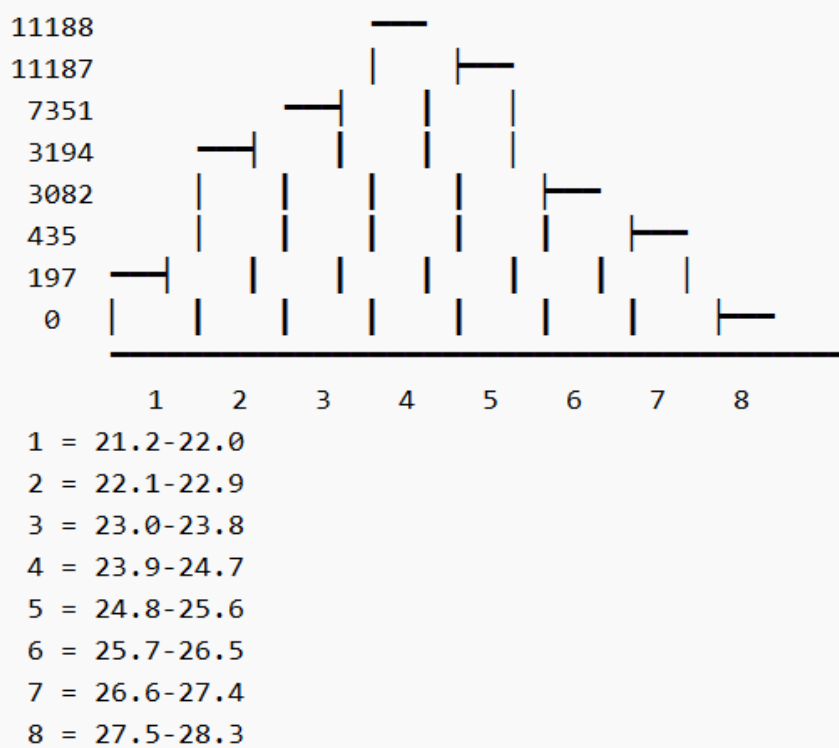
Que columna desea evaluar?:

- |                               |                           |
|-------------------------------|---------------------------|
| 1. Datos temperatura interior | 2. Datos humedad interior |
| 3. Datos temperatura exterior | 4. Datos humedad exterior |
| 5. Datos presión relativa     | 6. Datos presión absoluta |

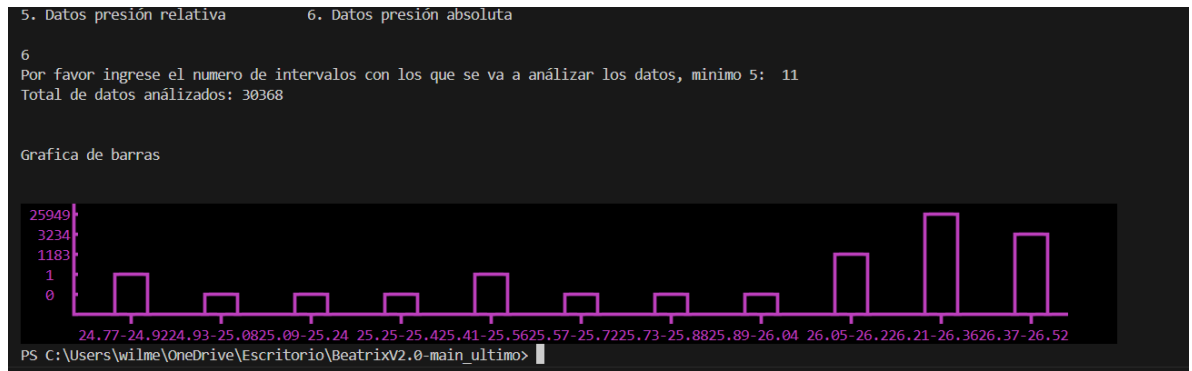
numero de columna seleccionado: 3

numero de intervalos seleccionado: 8

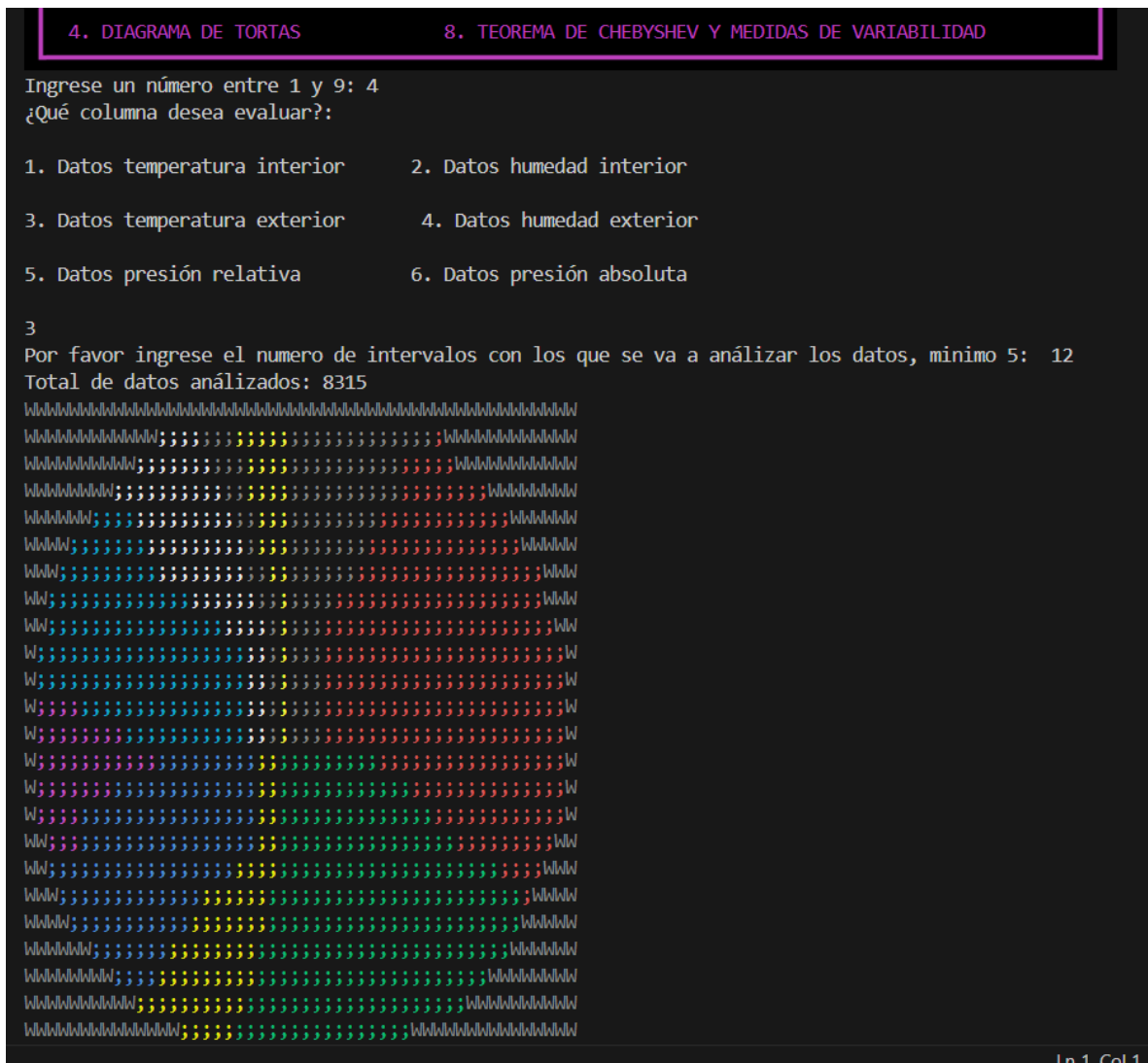
Total de datos analizados: 36634



4. se utiliza el segundo data set, opción 6:mostrar diagrama de barras, datos de presión absoluta y 11 intervalos:



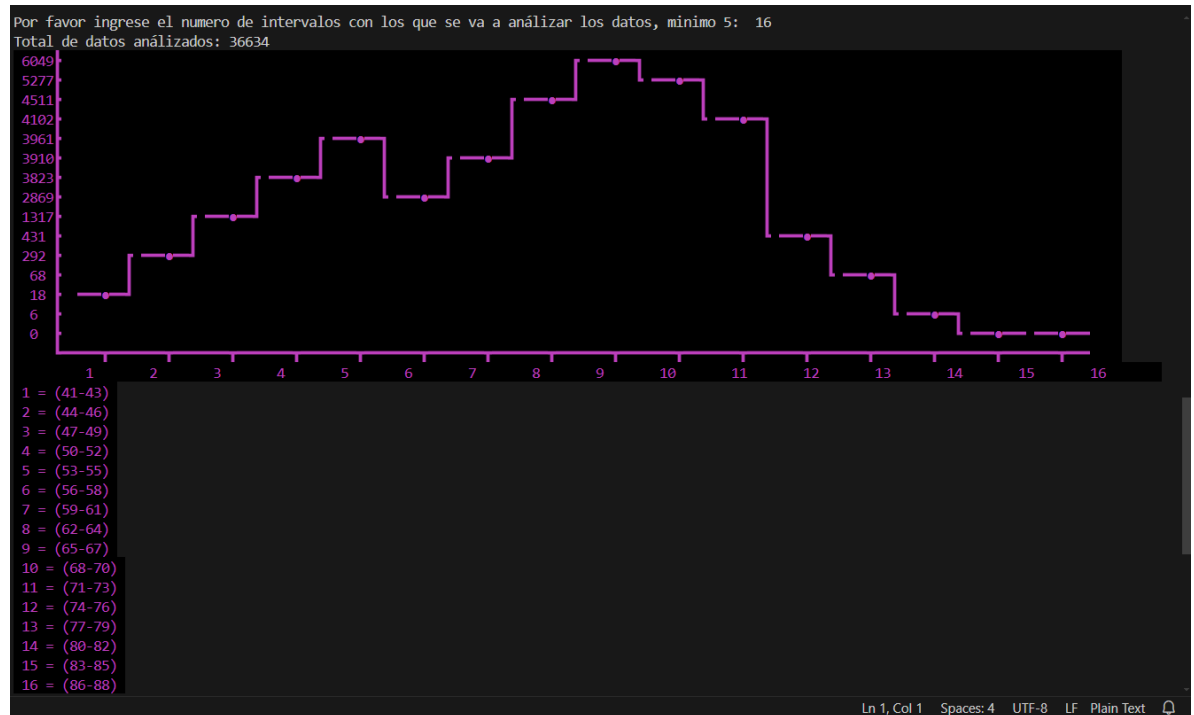
5. Se utiliza el data tercer data set, opción 4: diagrama de torta, los datos de temperatura exterior y 12 intervalos:



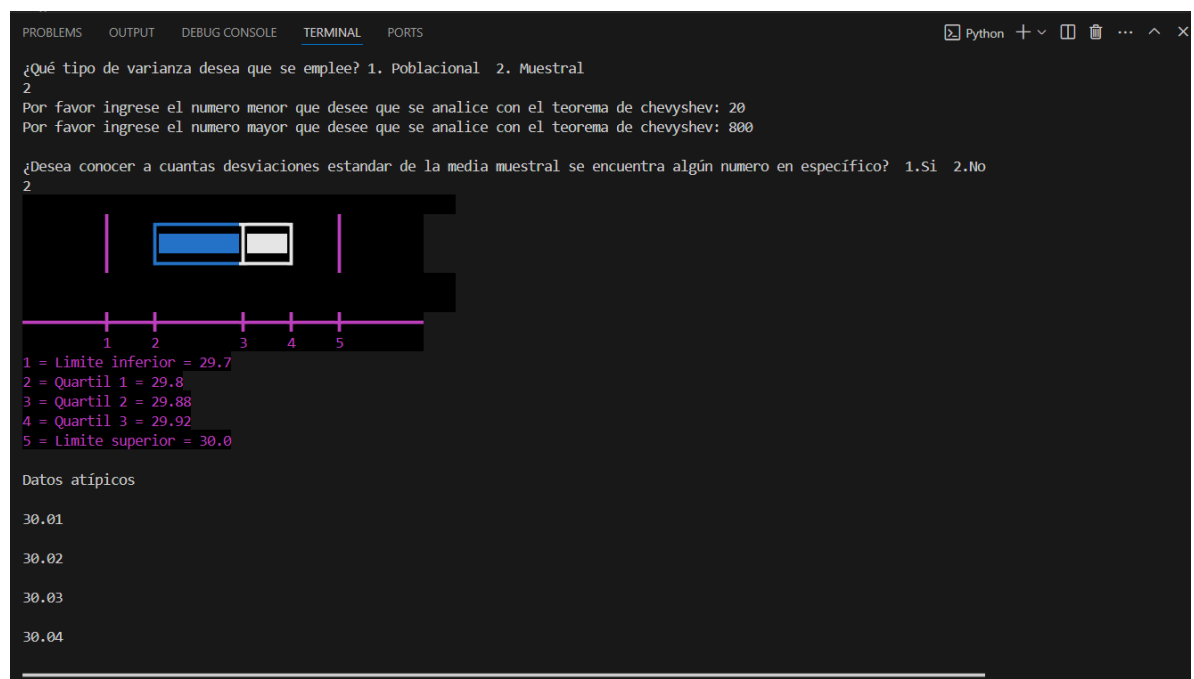
COLOR	■	NOMBRE = 18.6-20.1	VALOR EN GRADOS = 15.55°
COLOR	■	NOMBRE = 20.2-21.7	VALOR EN GRADOS = 92.12°
COLOR	■	NOMBRE = 21.8-23.3	VALOR EN GRADOS = 73.04°
COLOR	■	NOMBRE = 23.4-24.9	VALOR EN GRADOS = 30.85°
COLOR	■	NOMBRE = 25.0-26.5	VALOR EN GRADOS = 35.28°
COLOR	■	NOMBRE = 26.6-28.1	VALOR EN GRADOS = 28.91°
COLOR	■	NOMBRE = 28.2-29.7	VALOR EN GRADOS = 37.33°
COLOR	■	NOMBRE = 29.8-31.3	VALOR EN GRADOS = 32.22°
COLOR	■	NOMBRE = 31.4-32.9	VALOR EN GRADOS = 11.09°
COLOR	■	NOMBRE = 33.0-34.5	VALOR EN GRADOS = 3.17°
COLOR	■	NOMBRE = 34.6-36.1	VALOR EN GRADOS = 0.36°
COLOR	■	NOMBRE = 36.2-37.7	VALOR EN GRADOS = 0.07°

Dado que un archivo de texto no puede procesar código ASCII (que fue lo que se empleó para realizar este diagrama de torta) no se puede mostrar la gráfica en el archivo de texto.

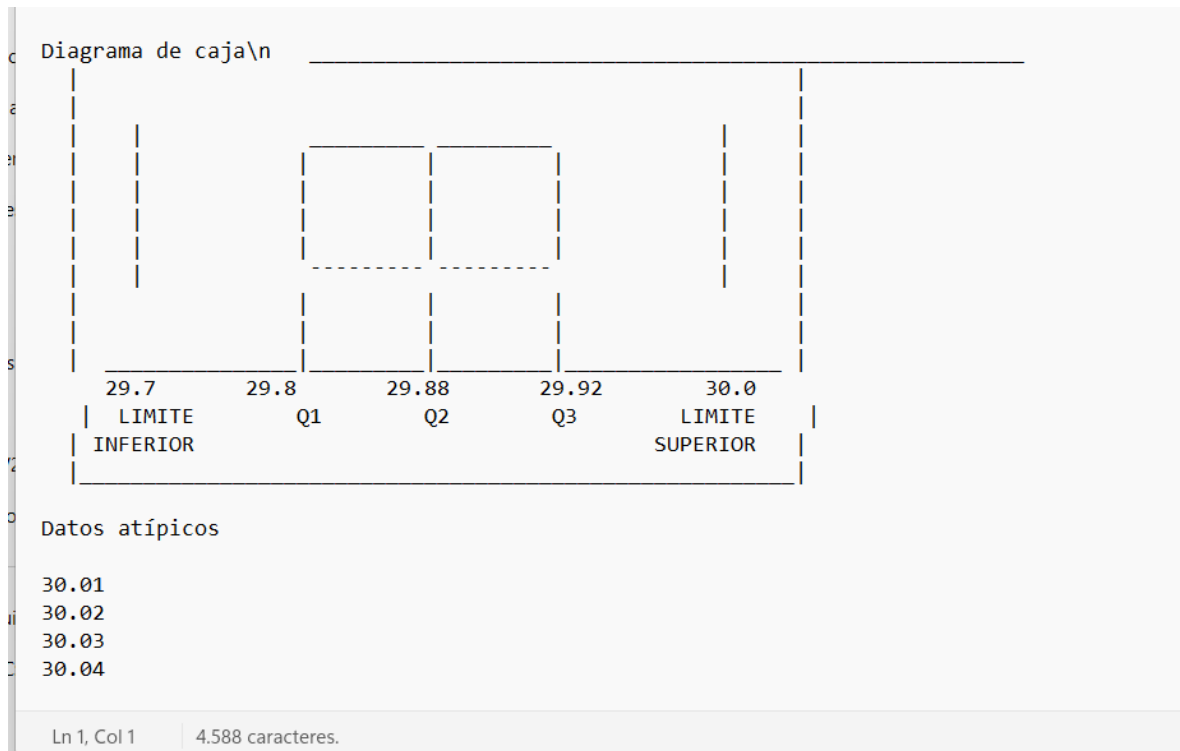
6. Se utiliza el primer data set, opción 5: ojiva, datos de humedad exterior y 16 intervalos:



7. Se utiliza el primer data set, opción 7: diagrama de caja, datos de presión relativa  
 y 20 intervalos:



Así se registra en el archivo de texto:



### Explicación general del código

Para realizar este código se optó por utilizar el paradigma de la programación orientada a eventos y el patrón de diseño MVC(Modelo Vista Controlador), con el fin de tener una estructura mas organizada. Inicialmente, el programa lee los datos del archivo csv mediante un script de lectura de datos, en el que se leen los datos fila por fila y se almacena en una lista para poder acceder a ellos con más facilidad.

El método principal lo que hace es ejecutar un menú en el cual se pide ingresar la ruta del archivo csv, luego hace un llamado a una clase()que se encarga de almacenar los datos de las columnas del archivo csv en listas, ya que los archivos cuentan con diferentes campos como datos de la temperatura interior, exterior, humedad interior, exterior, presión

relativa y presión absoluta, una vez llenos, se muestra el menú que contiene las opciones, como los diagramas y las medidas de tendencia central, para la cual, hay una clase que se encarga específicamente de calcular la media, media, moda y los cuartiles.

Las clases modelo son el esquema de los programas, como los de medidas de tendencia central y medidas de variabilidad(rango, media muestra, varianza, etc.), las vistas es como tal las gráficas estadísticas(histograma, barras, ojiva,etc.) incluyendo la parte del menú que es la que directamente ve el usuario e interactúa con el, es ahí donde entran las clases controlador, que cumplen el rol de intermediario entre las vistas y el modelo haciendo validaciones y demás cálculos.

## **Conclusiones**

Beatrix es una herramienta poderosa para el análisis estadístico de grandes volúmenes de datos, facilitando cálculos complejos y la generación de gráficos y tablas de manera eficiente. Su diseño amigable e intuitivo permite a los usuarios realizar análisis detallados y almacenar los resultados para su uso futuro.