

Estudiantes

Mynell Jemuel Myers Hall

Andrew Denilson López Herrera

Profesor

Allan Rodriguez Davila

I Proyecto

Lenguajes de programación

Código del curso: IC4700

Ingeniería en Computación

I Semestre 2024

Fecha de entrega: 18 de marzo del 2024

**Manual de Usuario**

Este proyecto ha sido construído y probado en Linux, en la distribución Ubuntu. Por lo que, todos los pasos del manual de usuario serán descritos únicamente para ser utilizado en Ubuntu.

**1) Compilación del programa**

1. Descargamos el proyecto de su repositorio de GitHub.
2. Una vez descargado el proyecto, entramos a la carpeta **programa**.
3. En la carpeta **programa** abriremos la terminal de Linux.
4. En la terminal escribiremos el comando **gcc main.c -o Aplicacion -lm**.
5. Con esto el programa ya estará compilado.

**2) Ejecución del programa**

1. Para ejecutar el programa, debemos abrir la terminal en la carpeta **programa** del proyecto.
2. Se deberá ejecutar el comando **./Aplicacion**.
3. El programa mostrará el menú principal.

**3) Uso del programa**

1. **Menú principal –** El menú principal mostrará 9 opciones que puede seleccionar el usuario:

1. Gestión de regiones

2. Gestión de datos climáticos

3. Procesamiento de datos

4. Análisis de datos

5. Análisis de correlación

6. Visualización de datos

7. Búsqueda de datos

8. Estadísticas

9. Salir

El usuario deberá ingresar cualquiera de los números del menú para ingresar a las opciones. Cualquier otro número ingresado será rechazado por el sistema.

1. **Gestión de regiones – Opción 1:** Esta opción le permite al usuario crear, eliminar y mostrar las regiones creadas. Estas se mostrarán en el menú **Menu Gestion de Regiones**.

**-Crear región – Opción 1:** Esta opción permite al usuario crear nuevas regiones, el usuario solicitará cuatro datos para crear la región: el nombre, la descripción, la posición **x** y la posición **y**. El nombre no debe estar en blanco y no se permiten nombres repetidos. La descripción no debe estar en blanco. La posición **x** debe ser un numero flotante. La posición **Y** debe ser un número flotante.

**-Eliminar región – Opción 2:** Esta opción permite al usuario eliminar una región que haya sido creada. Para eliminar la región solamente se solicita el nombre de la región. El nombre de no debe estar en blanco y debe coincidir con el nombre de alguno de las regiones.

**-Mostrar regiones – Opción 3:** Esta región permite al usuario ver las regiones que han sido creadas.

1. **Gestión de datos climáticos – Opción 2:** Esta opción permite al usuario ingresar nuevos registros de datos climáticos desde un archivo csv. Esta opción solicita la URL del archivo para poder abrir el archivo y extraer los datos. Se recomienda que el archivo se encuentre en la carpeta **programa** para que solo se tenga que escribir el nombre del archivo. Cuando el programa solicite la URL del nombre se debe poner en el siguiente formato “**nombreprograma**.csv” sin las comillas. Se indicará si los datos fueron cargados.
2. **Procesamiento de datos – Opción 3:** Esta opción consta de tres opciones: completar datos faltantes, eliminar datos duplicados y eliminar valores atípicos. Todas estas opciones son automáticas, por lo que el usuario solo debe usar las opciones.

**-Completar datos faltantes – Opción 1**: Esta opción permite al usuario completar los datos faltantes de los datos climáticos. El usuario solo debe usar la opción.

**-Eliminar datos duplicados – Opción 2:** Esta opción permite al usuario eliminar los datos duplicados de los registros climáticos. El usuario solo debe usar la opción.

**-Eliminar valores atípicos – Opción 3:** Esta opción permite al usuario eliminar los valores atípicos de los datos climáticos. El usuario solo debe usar la opción.

1. **Análisis de datos – Opción 4:** Esta opción le permite al usuario ver al usuario la media, la mediana, la moda y la desviación estándar por cada uno de los datos de temperatura, humedad, presión atmosférica, velocidad del viento y precipitación. Este análisis necesita los siguientes datos:

**-Nombre de la región:** Este nombre puede ser opcional, indica de cual región se desea hacer el análisis. Si se ingresa ALL, entonces se hará el análisis de todas las regiones.

**-Fecha inicial:** Indica al usuario que los registros que se utilicen deben de ser de esa fecha en adelante.

**-Fecha final:** Indica al usuario que los registros que se utilicen deben de ser de esa fecha hacia atrás.

Al final, el programa muestra el análisis según los datos ingresados.

1. **Análisis de correlación – Opción 5:** Esta opción permite al usuario ver el análisis de correlación de Pearson de los datos climáticos. Para realizar esta operaci+on se necesita de los siguientes datos.

**-Nombre de la región:** Este nombre puede ser opcional, indica de cual región se desea hacer el análisis. Si se deja el vacío el espacio el programa entiende que se deben utilizar todas las regiones, entonces se hará el análisis de todas las regiones.

**-Dato uno**: Es el grupo de datos que se utilizara para realizar el análisis, este grupo de datos pueden ser las siguientes: temperatura, humedad, presión atmosférica, velocidad del viento y la precipitación. El programa pregunta cual de esos datos desea utilizar.

**-Dato dos:** Es el grupo de datos que se utilizara para realizar el análisis, este grupo de datos pueden ser las siguientes: temperatura, humedad, presión atmosférica, velocidad del viento y la precipitación. El programa pregunta cual de esos datos desea utilizar.

**-Fecha inicial:** Indica al usuario que los registros que se utilicen deben de ser de esa fecha en adelante.

**-Fecha final:** Indica al usuario que los registros que se utilicen deben de ser de esa fecha hacia atrás.

Al final, el programa muestra el análisis según los datos ingresados.

1. **Visualización de datos – Opción 6:** Esta opción le permite al usuario ver el rango y la varianza de los datos climáticos de una región o de todas las regiones. Esta opción le solicita lo siguiente al usuario:

**-Nombre de la región:** Este nombre puede ser opcional, indica de cual región se desea hacer el análisis. Si se deja el vacío el espacio el programa entiende que se deben utilizar todas las regiones, entonces se hará el análisis de todas las regiones.

**-Posición X:** La posición **x** aproximada de una región.

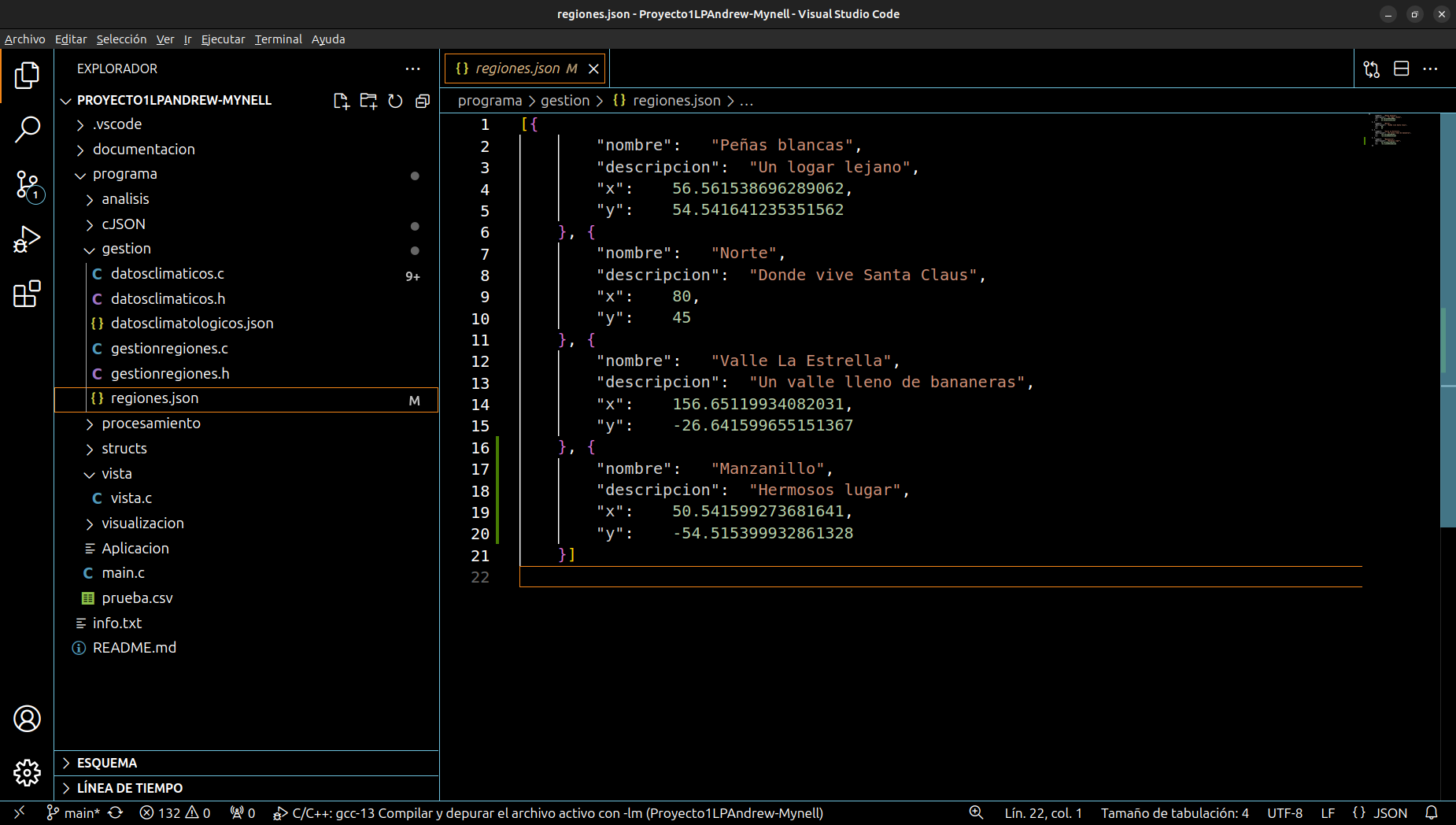
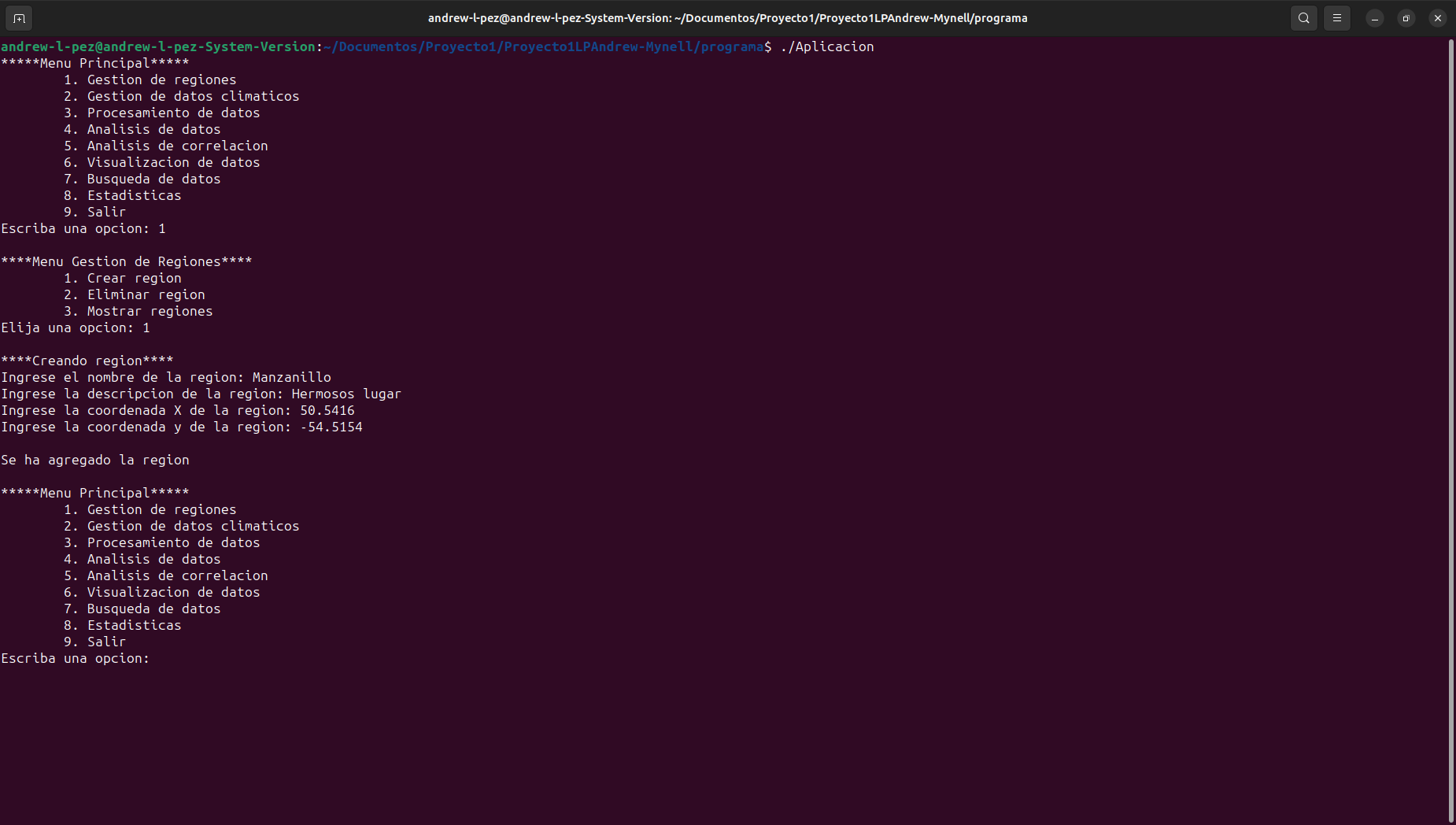
**-Posición Y:** La posición **Y** aproximada de una región.

1. **Búsqueda de datos – Opción 7:**
2. **Estadísticas – Opción 8:** Esta opción le muestra al usuario las siguientes estadísticas: top 3 de zonas con mayor precipitación, top 5 de registros con mayor velocidad del viento y top 5 de mes-año con mayor cantidad de registros.
3. **Salir – Opción 9:** Para cerrar el programa el usuario debe ingresar la opción 9 en el menú principal, con esto el programa termina su ejecución y todos los datos quedan guardados en archivos JSON.

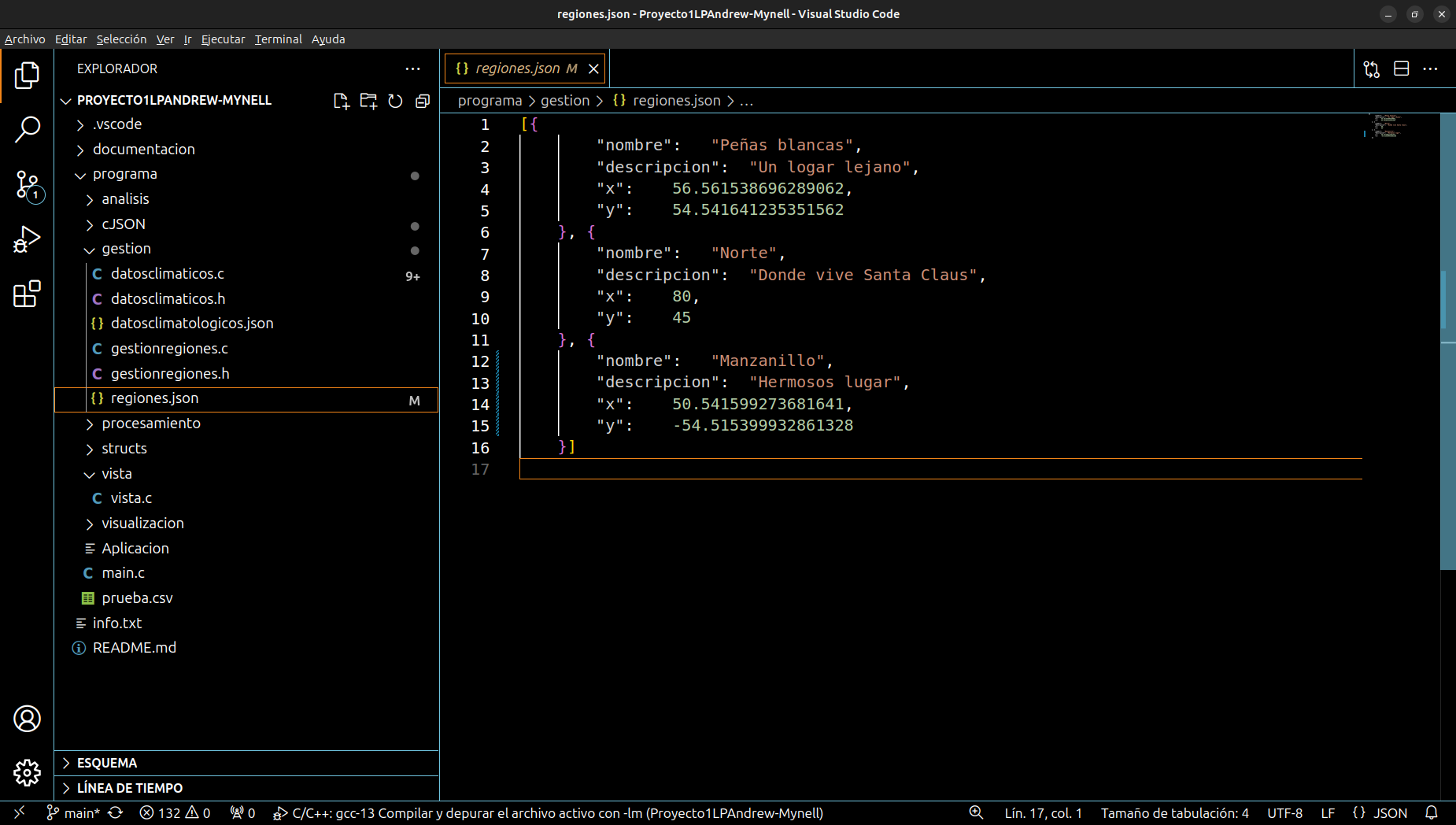
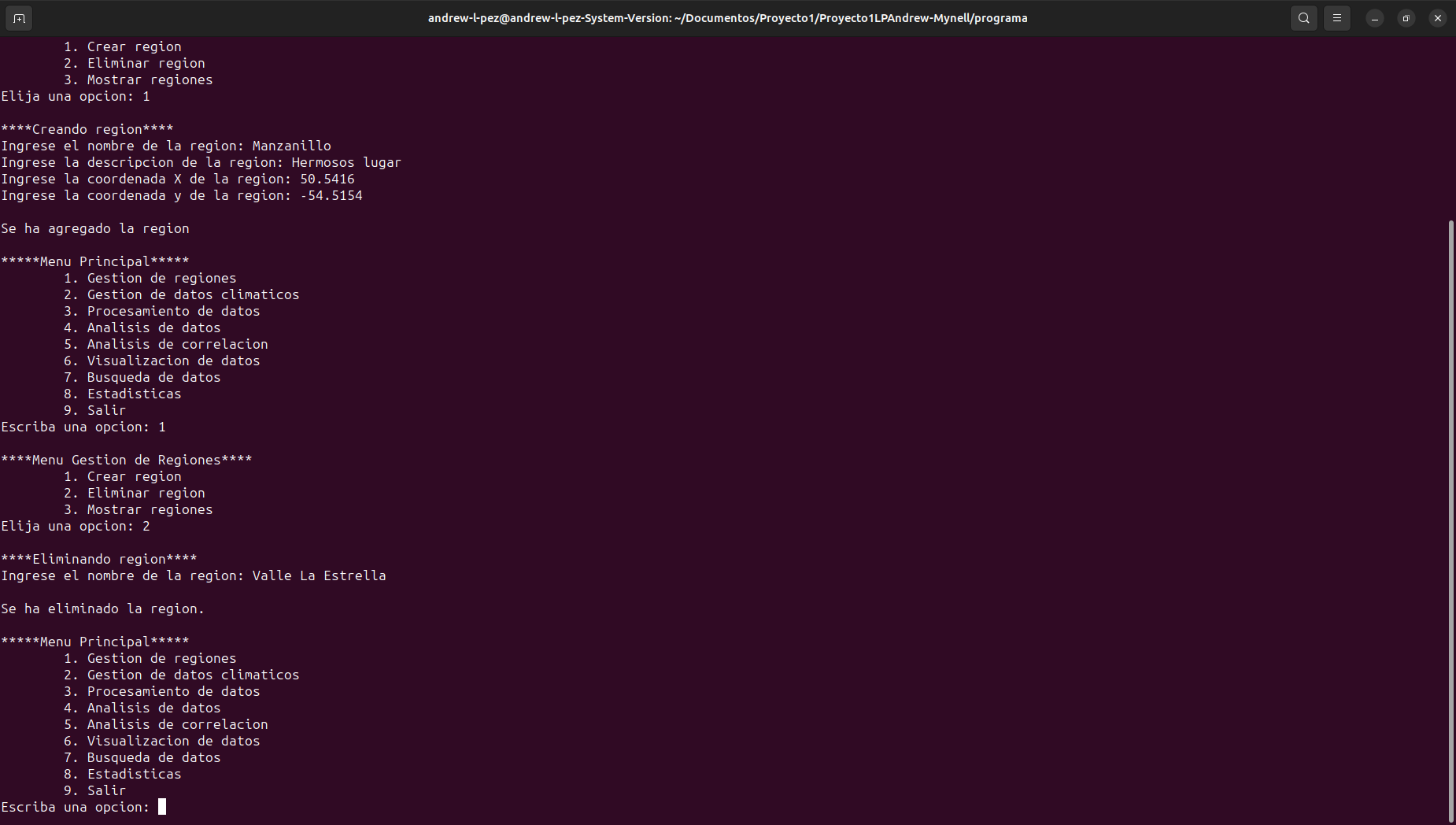
**Pruebas de funcionalidad**

1. **Gestón de regiones**

**Ingresando región nueva**

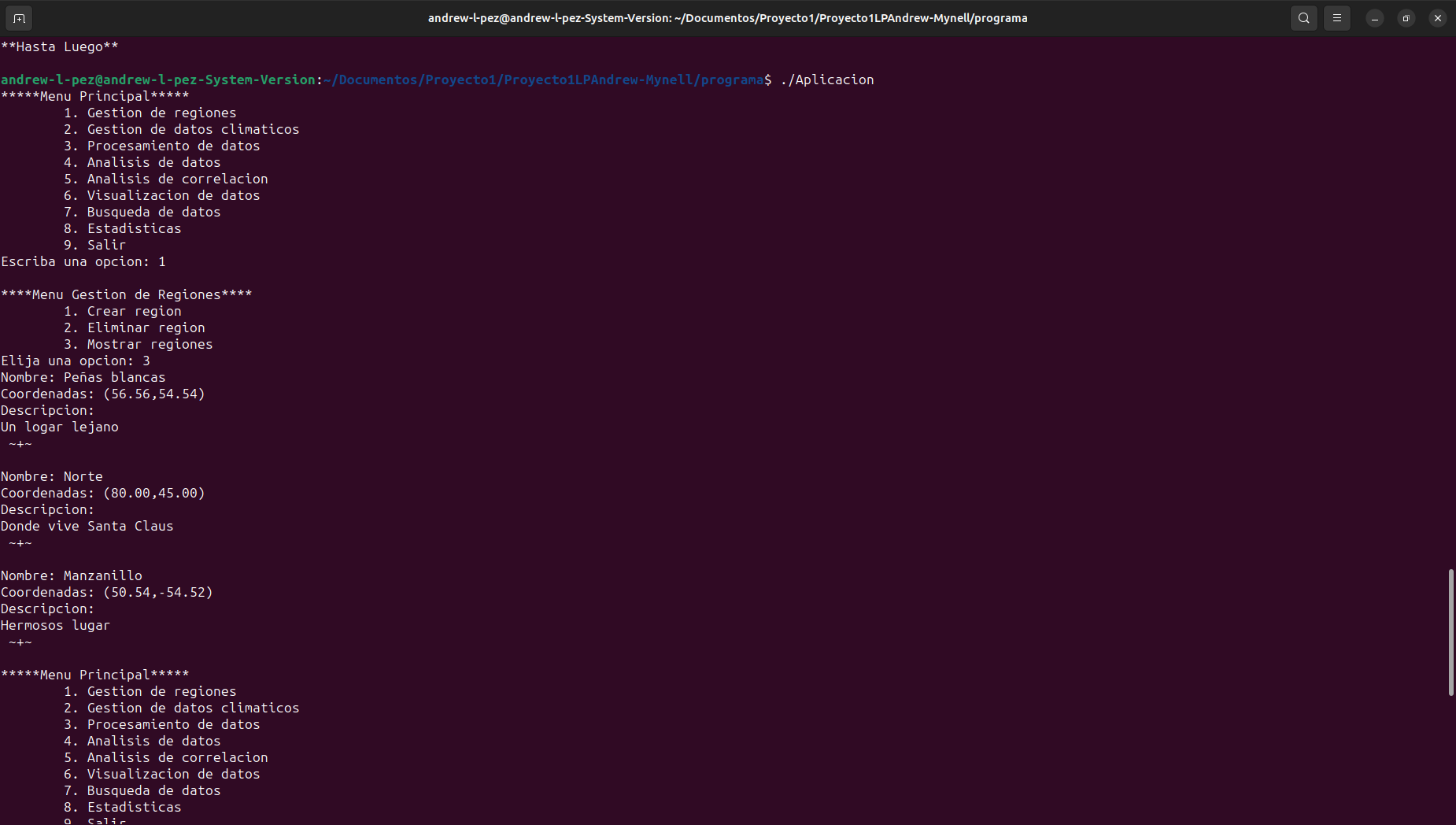
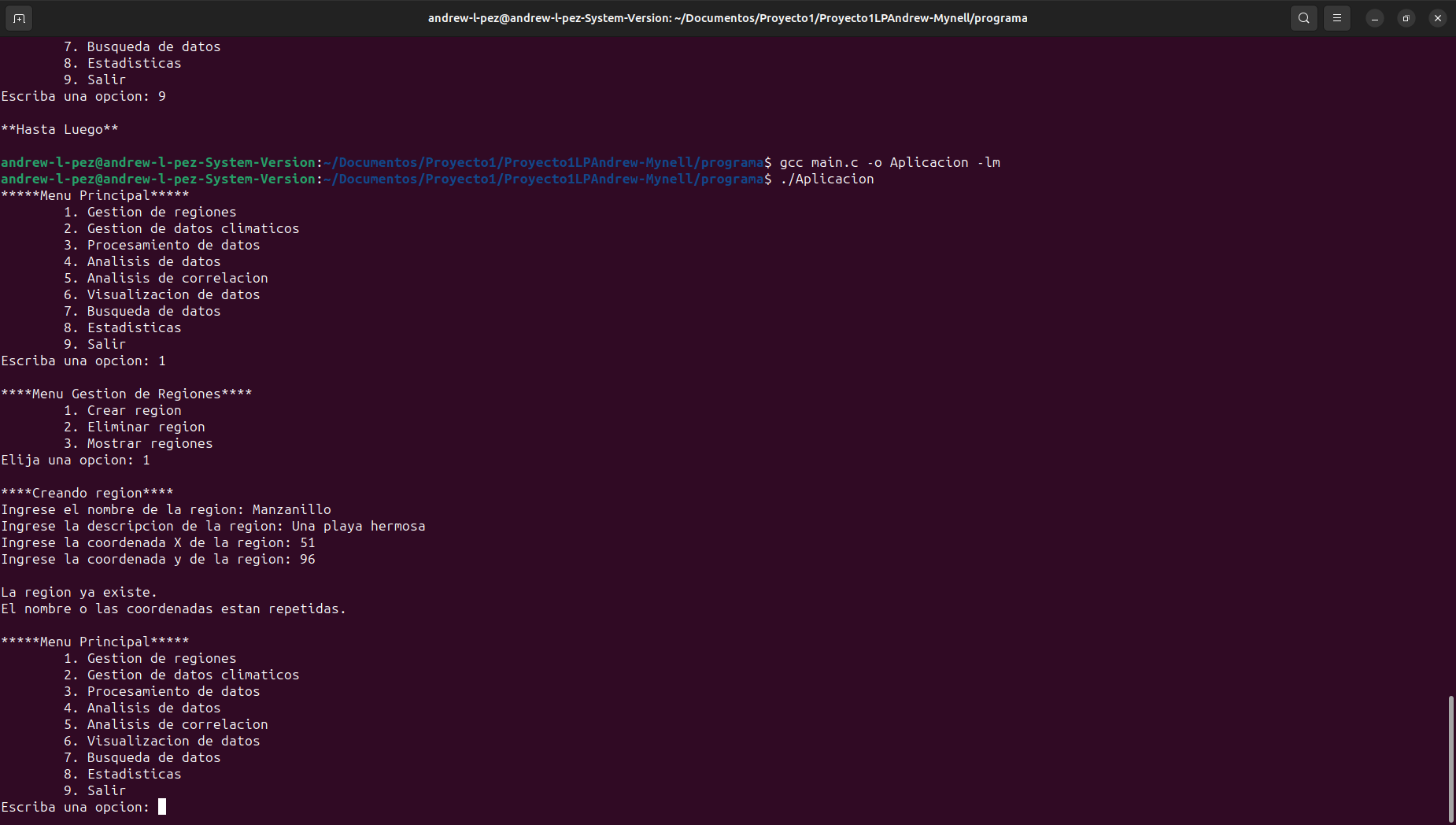
****

**Eliminando región**



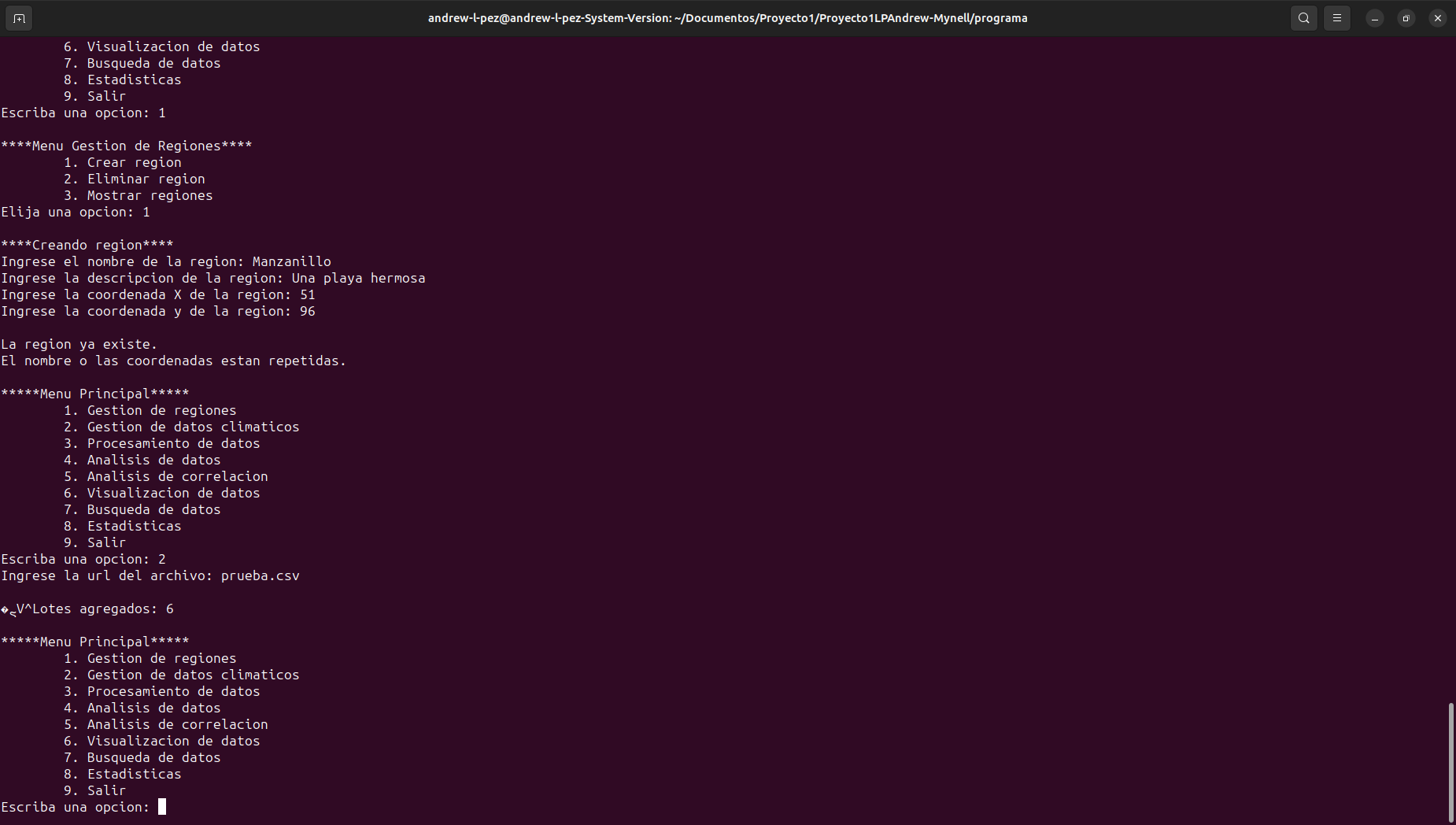
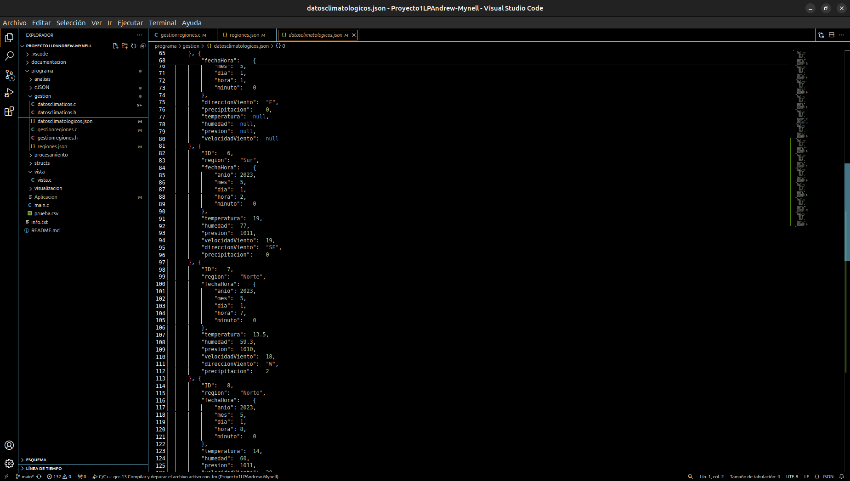
**Mostrando región**

**Validando datos**

****

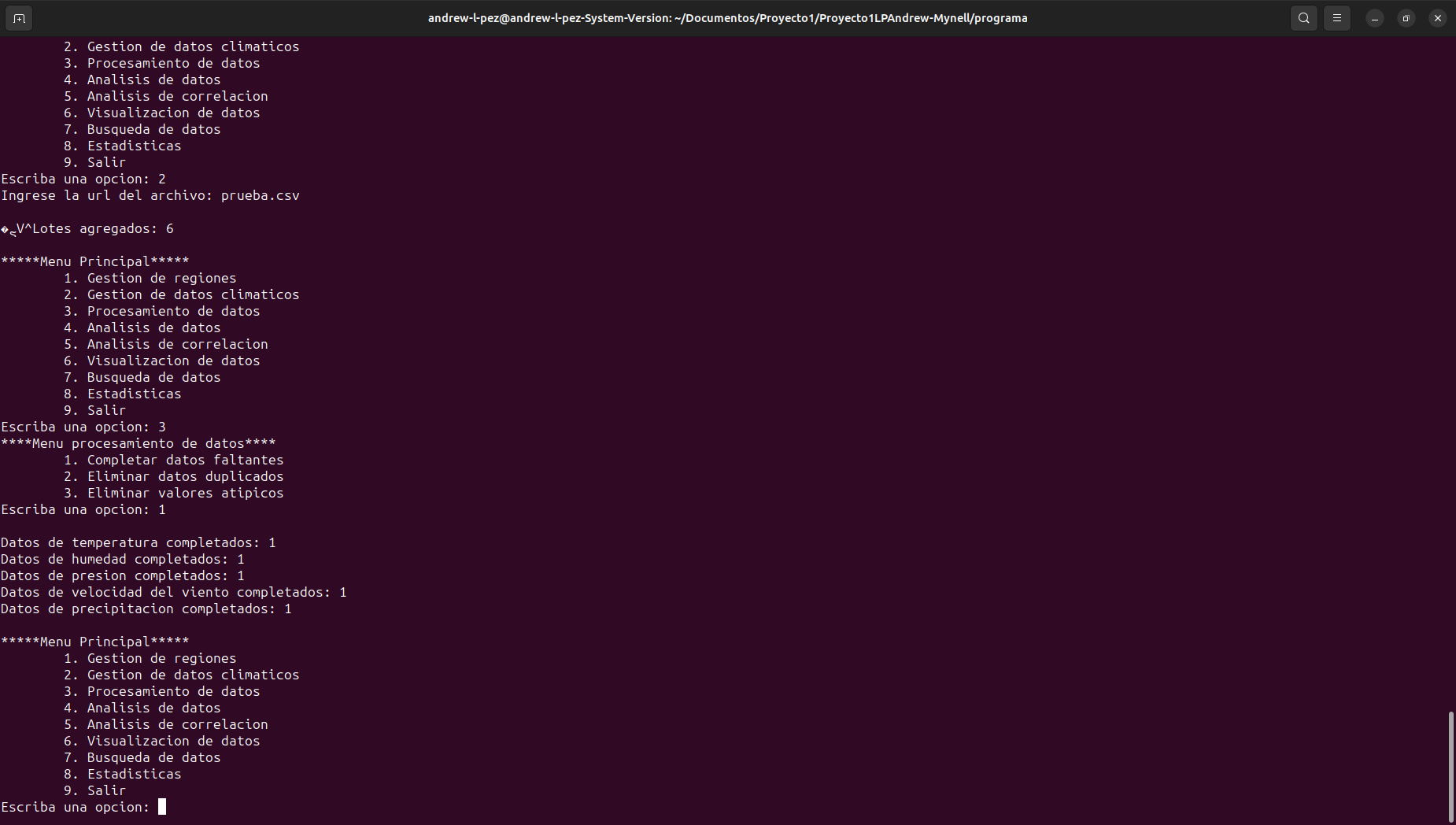
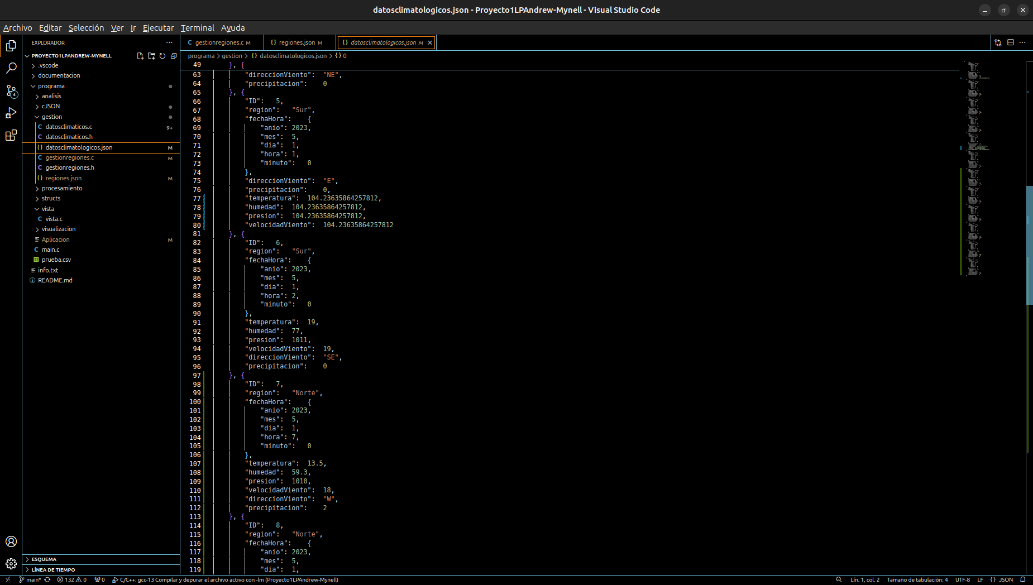
1. **Gestión de datos climáticos**

**Ingresando nuevo lote**

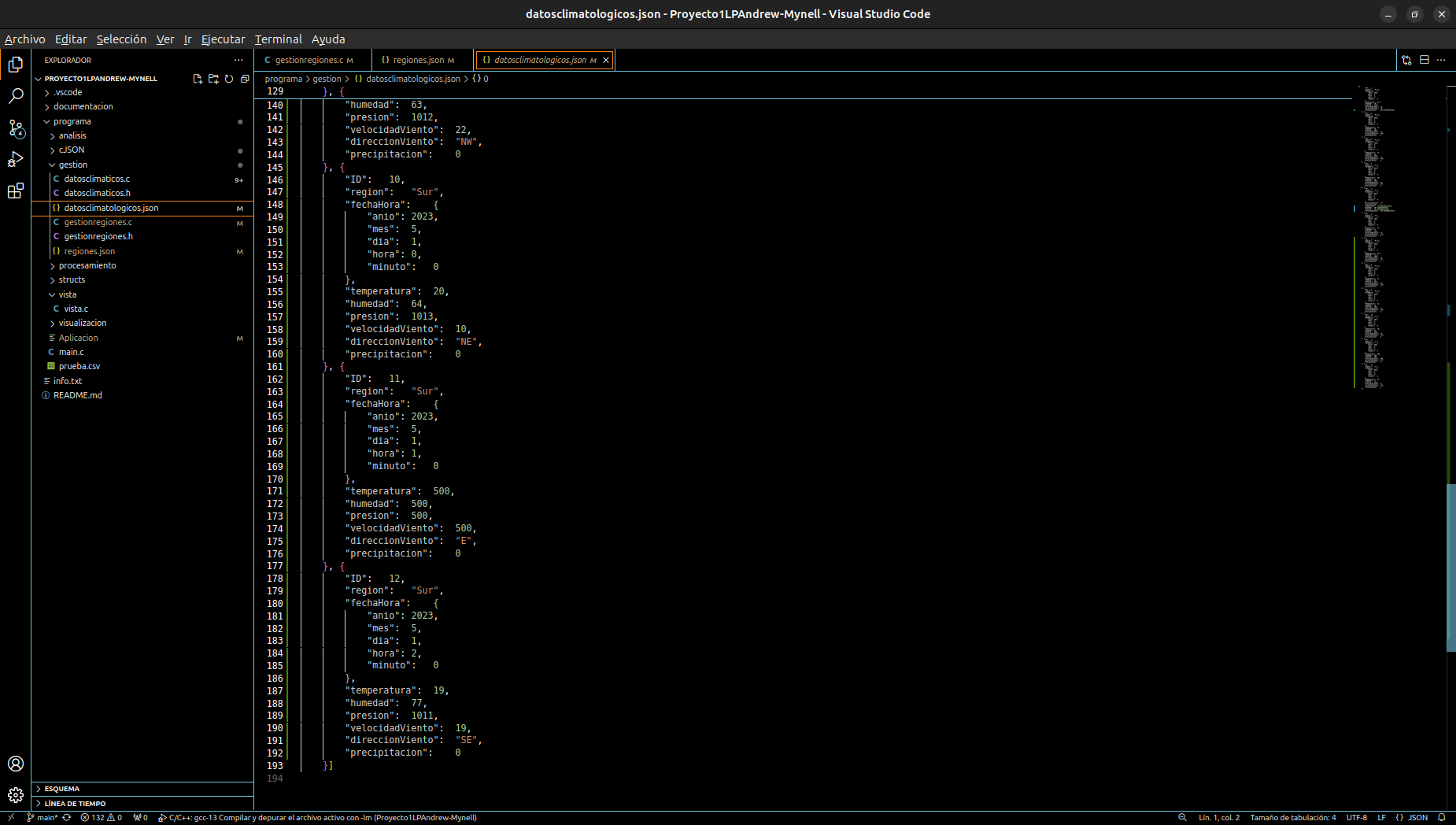


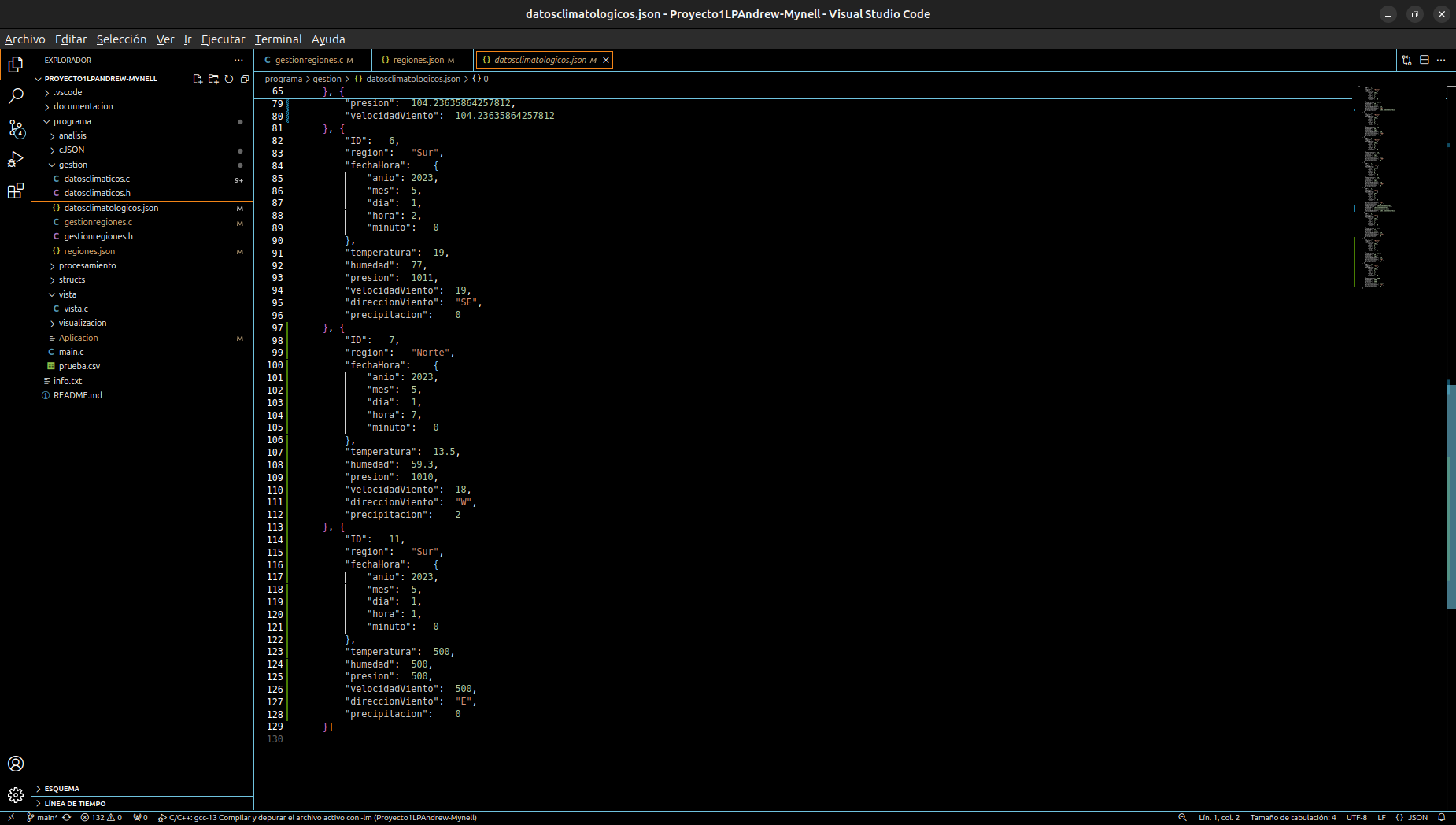
1. **Procesamiento de datos**

**Completando datos faltantes**

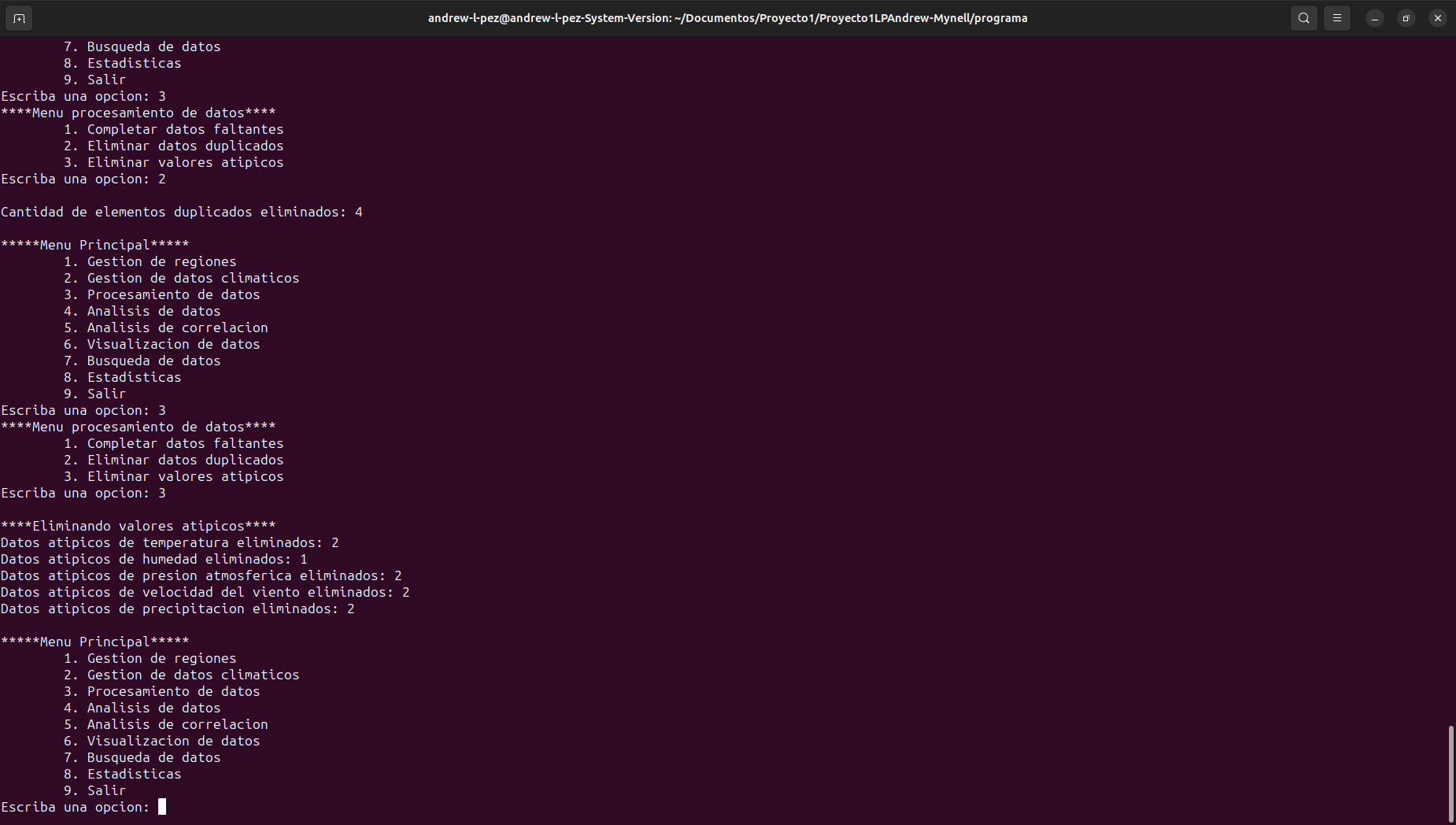
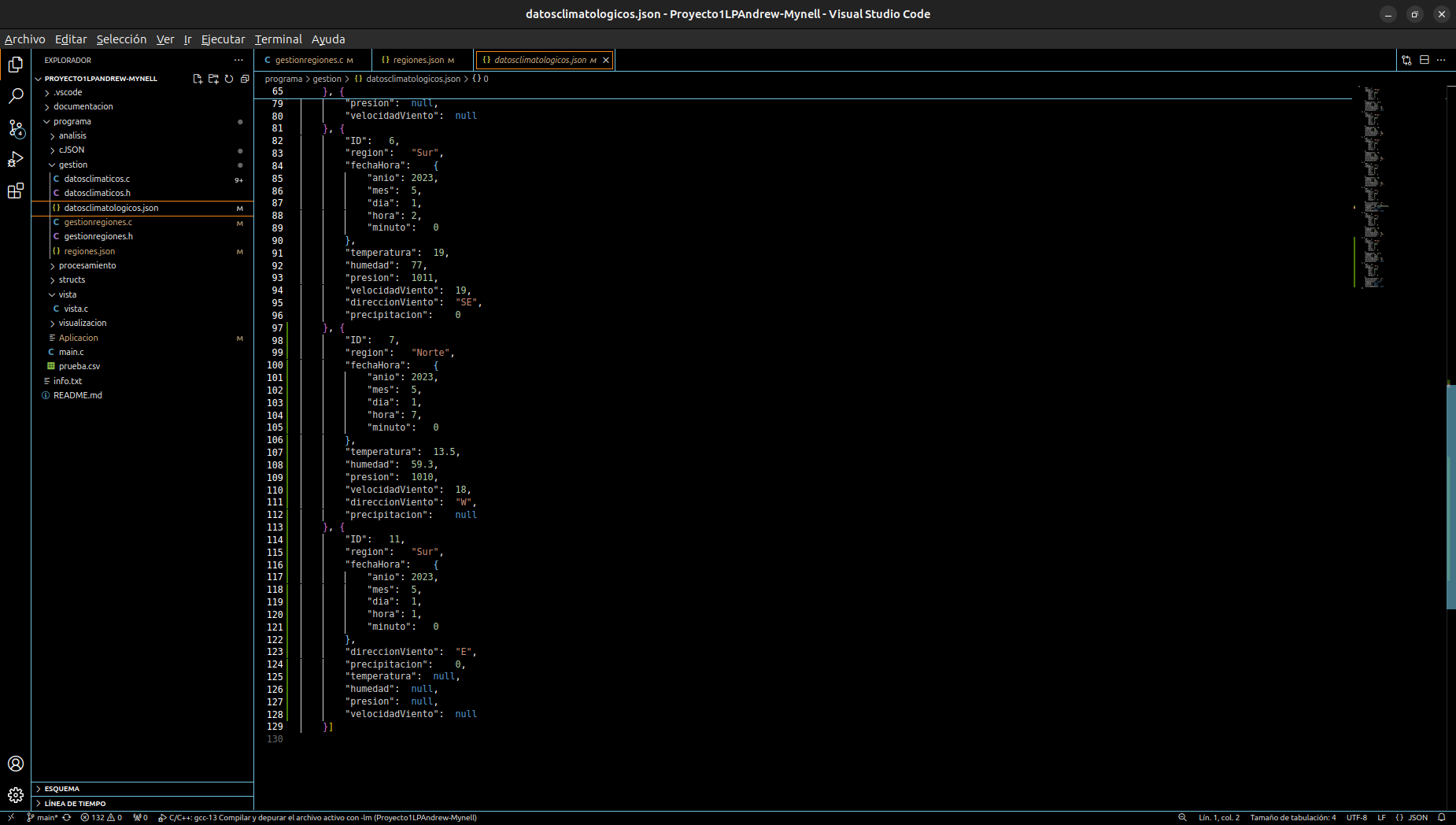


**Eliminando datos duplicados**

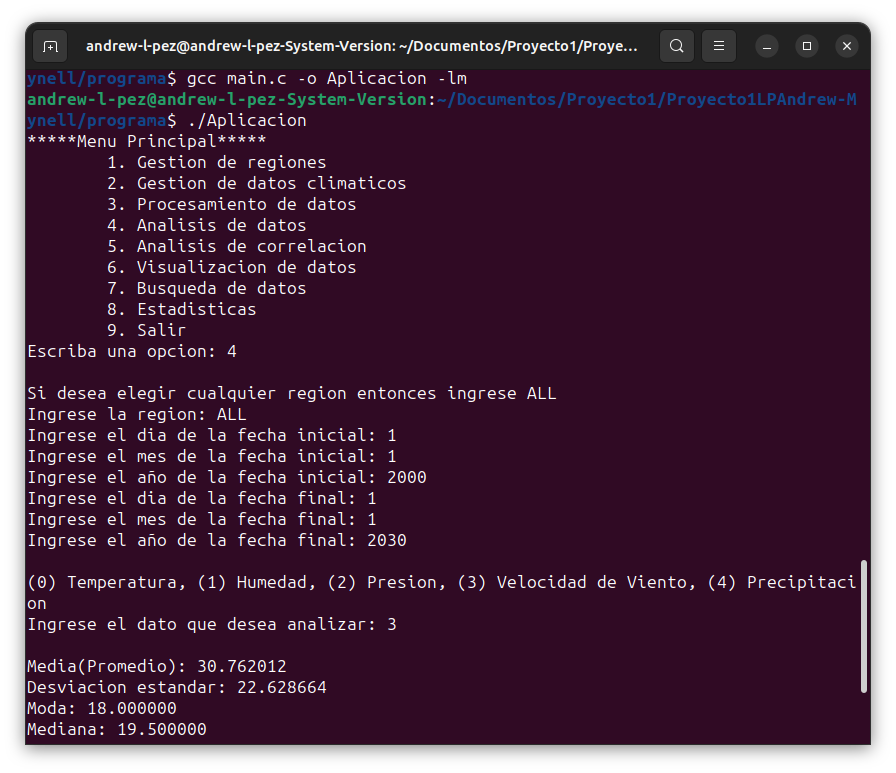
Texto

Descripción generada automáticamente

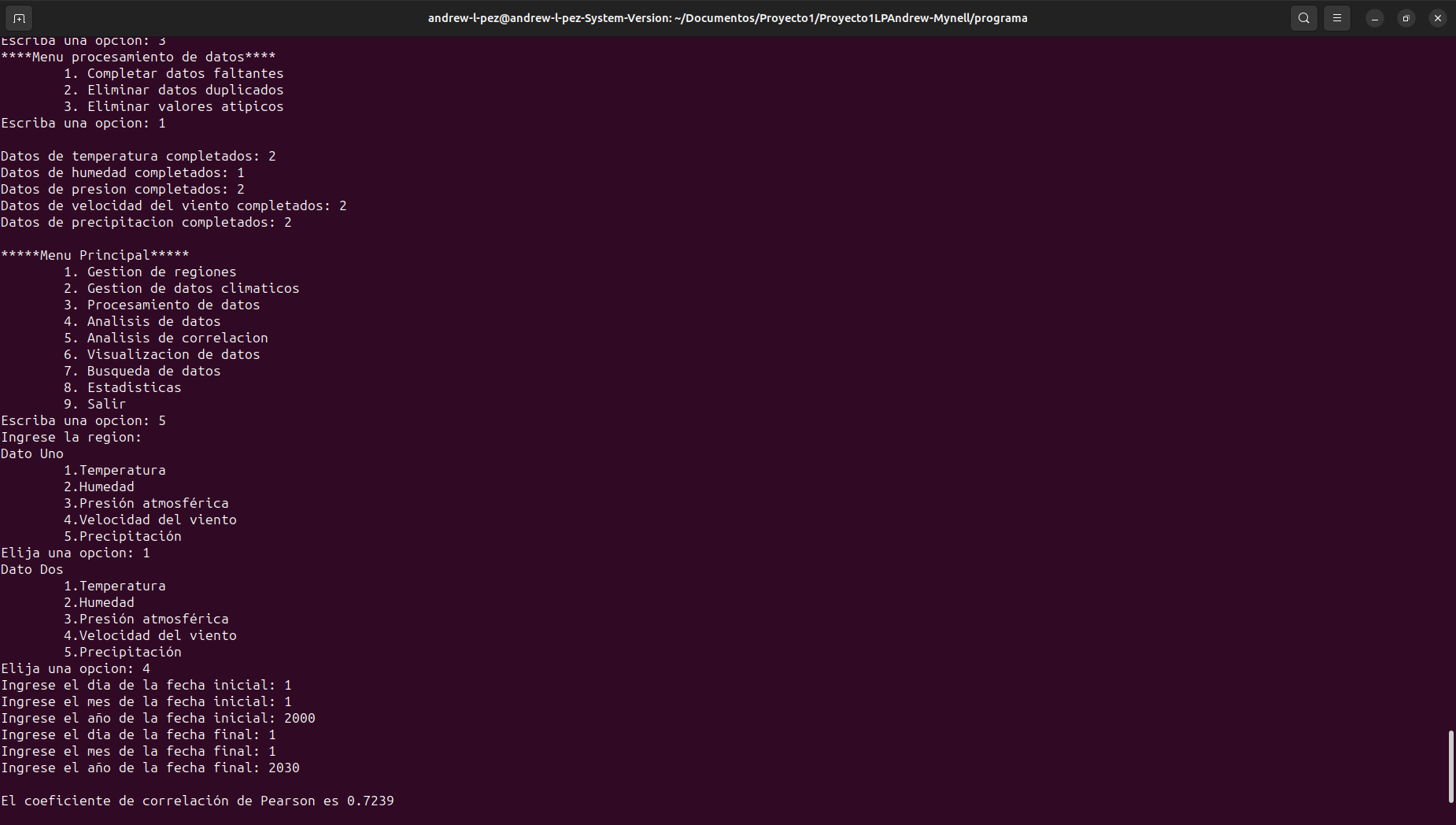
**Eliminando valores atípicos**

****

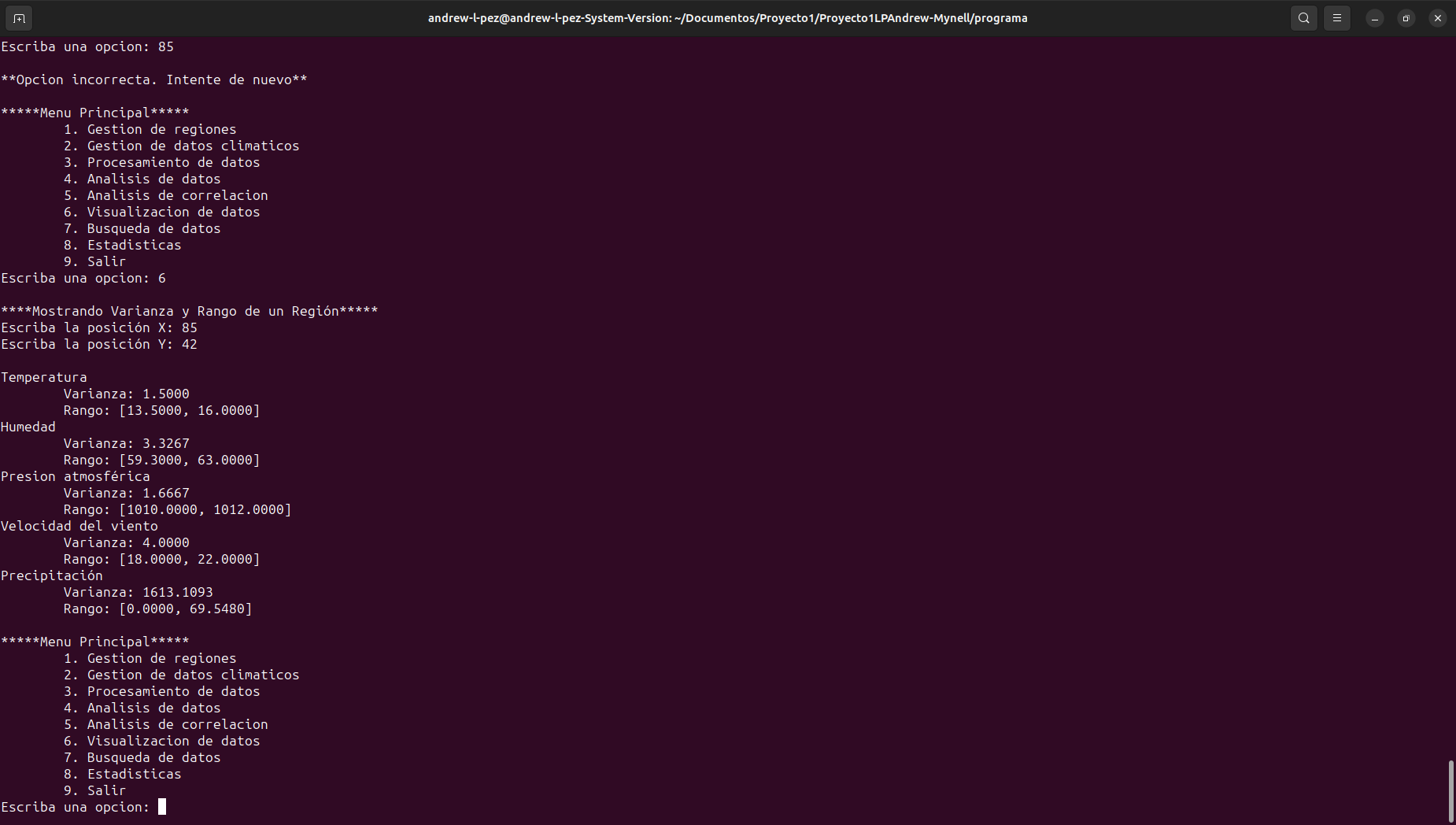


****

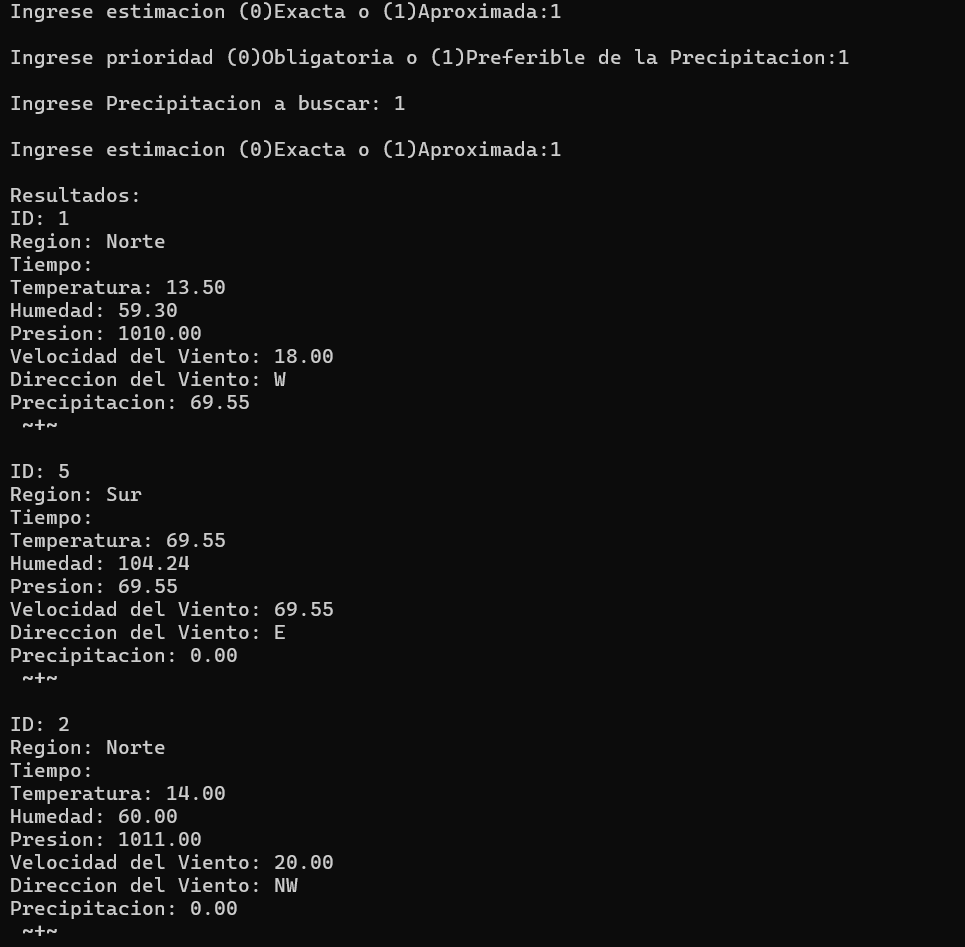
1. **Análisis de correlación de Pearson**

****

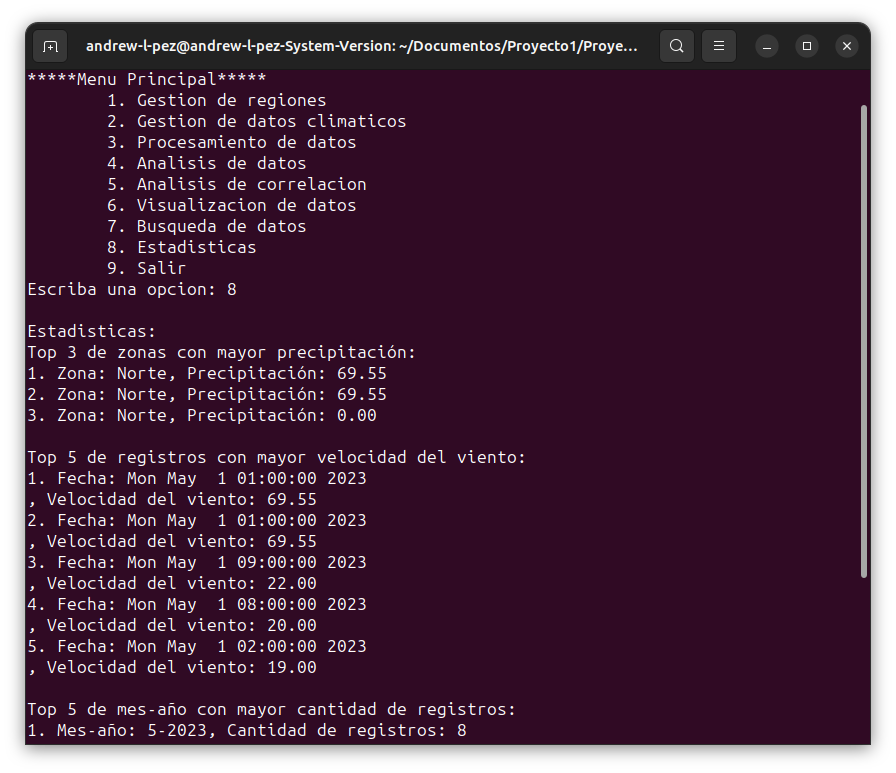
1. **Visualización de datos**

****

1. **Busqueda de datos**



**8) Estadísticas**

****

**Diseño del programa**

**a) Algoritmos utilizados**

1. **Gestión de regiones**

**-Creación de regiones:** Para la creación de regiones se verifica que no se repita los nombres ni los puntos cardinales, si no hay nada repetido, se guarda en el archivo JSON y se vuelve a cargar el archivo.

**-Eliminación de regiones:** Para la eliminación de las regiones se compara cada región para saber si su nombre coincide, en caso de que el nombre coincida, se procede a eliminar del archivo JSON y se vuelve a cargar el archivo.

**-Mostrar regiones**: Se recolectan los datos de región y luego se envían a mostrar.

1. **Gestión de datos climáticos**

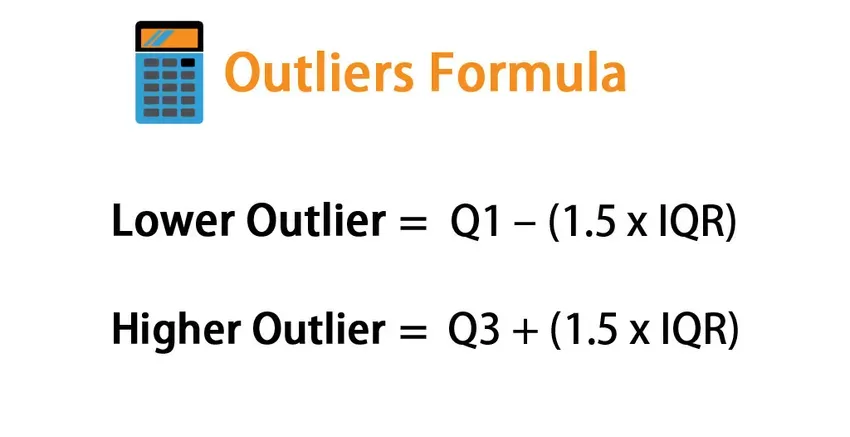
Para cargar los elementos de un csv se recorre línea por línea el csv, a cada línea se le aplica un split utilizando como separador las comas que posee. Luego, los datos se convierten en su respectivo tipo de datos, región y dirección del viento se quedan como strings, cada parte de la fecha se convierte en int y el resto de los datos se convierte en float. Estos se guardan de una vez en un objeto JSON para luego ser guardados en el archivo JSON y ser cargados en memoria.

1. **Procesamiento de datos**

**-Completar datos faltantes:** Para completar los datos faltantes se utiliza la media de todos los datos que si existen.

**-Eliminar datos duplicados:** Para eliminar los datos duplicados se compara una región con otra, pero se compara todos los datos de la región. Si se demuestra que una región está duplicada, entonces se procede a guardar el id de esa región en un arreglo, esto con el objetivo de marcarlo y cuando se termina todas las comparaciones, las regiones se eliminan directamente del archivo JSON. Las regiones que se encuentran en el arreglo, no se utilizan para comparar con otros. Al final, los datos se vuelven a cargar desde el archivo JSON.

**-Eliminar valores atípicos:** Para este proceso se decidió utilizar una fórmula para calcular un rango de valores típicos, todo número fuera de este este rango se considera un valor atípico. La fórmula es la siguiente:



Los datos se deben ordenar de menos a mayor, luego se debe calcular la mediana y luego a este se le multiplica por 3, seguidamente se debe calcular el primer cuartil “Q1” y el tercer cuartil “Q3”, al primer cuartil se le resta la media, este da como resultado el límite inferior, luego al tercer cuartil se le suma la media y este da como resultado el límite superior. Con esto dos límites se puede saber si un número es típico o atípico.

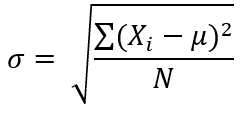
1. **Análisis de datos**

**Media:** Se uso la fórmula de media, que es sumar todos los datos del conjunto y dividirlos por la cantidad de individuos del conjunto.

**Mediana:** Se empleo un ordenamiento qsort con una función auxiliar para ordenar el conjunto de menor a mayor y se extrae el valor de en medio en caso de que sea un conjunto impar, en caso de que sea par se obtiene el promedio de los dos números centrales.

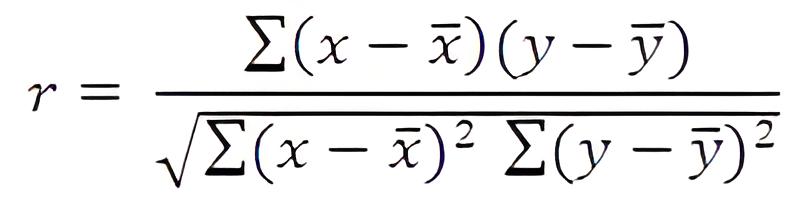
**Moda:** Con moda se busco el elemento que más se repite.

**Desviación estándar:** Se uso la siguiente formula teniendo en cuenta que (μ)Media, (Xi) Elemento de índice i y (N) número de elementos en el conjunto:



1. **Análisis de correlación**

Para este análisis se sigue la siguiente fórmula:

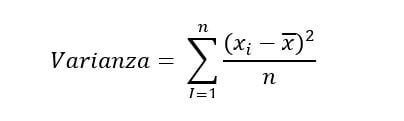


Para realizar esta fórmula, los dos grupos de datos deben ser del mismo tamaño.

1. **Visualización de datos**

Para realizar la operación de rangos, primero se ordena los datos y luego se escoge el primero y el último dato.

Para realizar la operación de varianza se utilizó la siguiente fórmula:



1. **Búsqueda de datos**

Se hicieron búsquedas comunes de exactitud (obteniendo todos los elementos con el valor exacto solicitado) y aproximación (obteniendo el elemento más cercano al valor buscado), pero en especial al buscar por aproximación una región se utilizo la siguiente formula: d = √ ((x2 - x1) ² + (y2 - y1) ²).

También se hizo un método para evitar la duplicación de búsquedas.

1. **Estadísticas**

Se uso qsort() para organizar cada datos que se necesitase para su respectivo top, y se presentaron los datos que quedaron arriba de la lista.

En el caso especifico del top mes-año se empleo un conteo con respecto al mes y año utilizando la técnica de memoize.

**Librerías usadas**

* **cJSON.h:** Se utiliza cJSON para el manejo de archivos JSON.
* **math.h:** Se utiliza math para hacer el cálculo de raíces y saber si un número es NAN.

**Análisis de resultados**

|  |  |  |  |
| --- | --- | --- | --- |
| **Objetivos** | **Sí** | **No** | **Observaciones** |
| **Documentación interna** | **X** |  |  |
| **Documentación externa** | **X** |  |  |
| **Gestión de Regiones** | **X** |  |  |
| **Gestión de datos climáticos** | **X** |  |  |
| **Procesamiento de datos** | **X** |  |  |
| **Análisis de datos** | **X** |  |  |
| **Análisis de correlación** | **X** |  |  |
| **Visualización de datos** | **X** |  |  |
| **Búsqueda de datos** | **X** |  |  |
| **Estadísticas** | **X** |  |  |
| **Unicidad** | **X** |  |  |
| **Uso de Structs** | **X** |  |  |

**Link del repositorio:** [**https://github.com/Mynell2022/Proyecto1LPAndrew-Mynell.git**](https://github.com/Mynell2022/Proyecto1LPAndrew-Mynell.git)