**[35 puntos] Ejercicio 1 – Análisis Exploratorio de Datos**

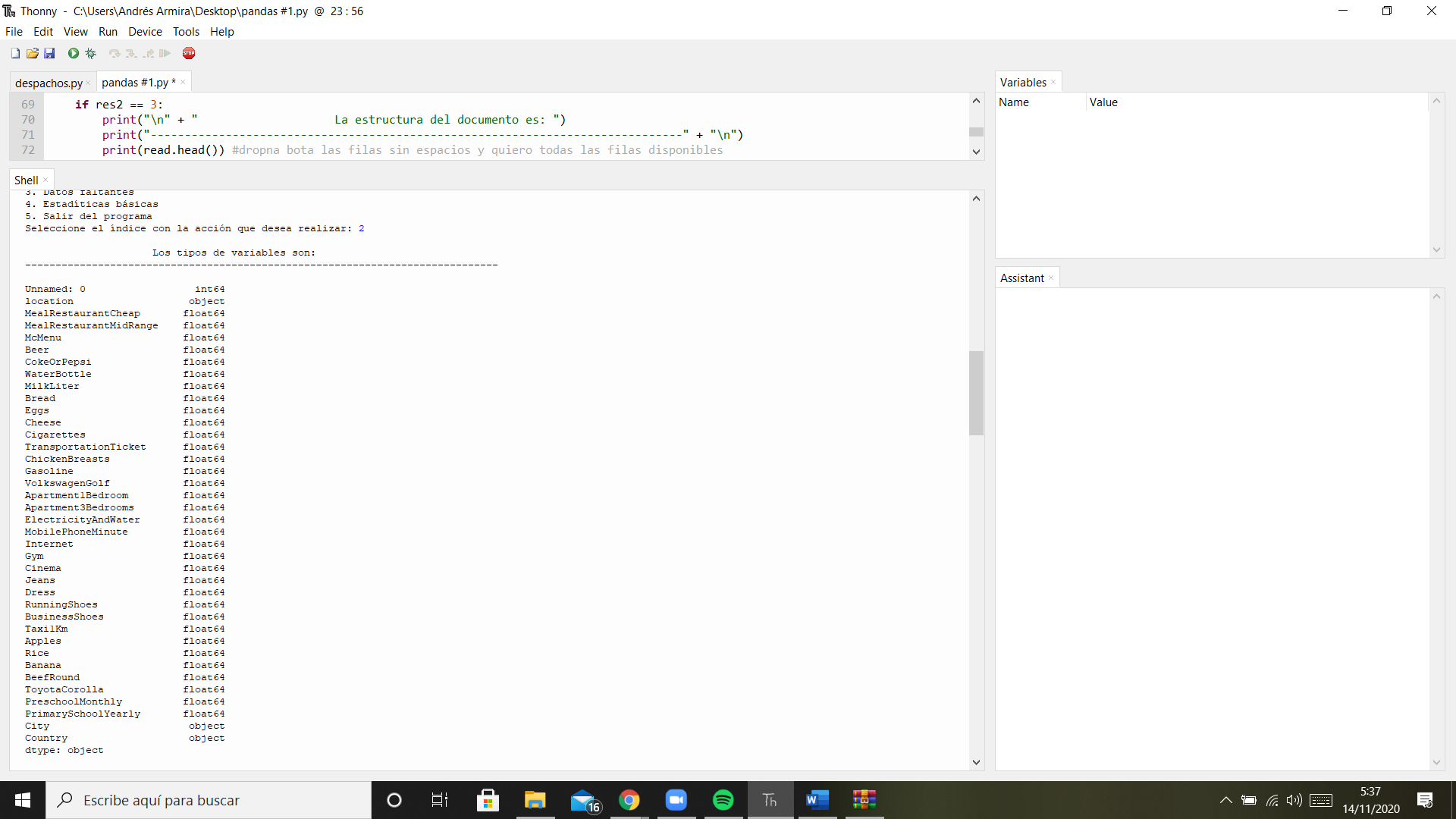
**Lo primero que se debe hacer es conocer qué datos hay en el archivo de datos. Usted decide realizar un programa utilizando el módulo de Pandas para hacer un análisis exploratorio de datos (EDA por sus siglas en inglés) buscando la siguiente información:**

**- ¿Cuál es la forma de los datos? Es decir:**

**1. ¿Cuántos registros (filas) hay?** R/ 160 filas.

**2. ¿Cuántos datos (columnas) hay en cada registro?** R/ 38 columnas

**2. ¿Cuáles son los datos que hay en cada registro y de qué tipo son?**



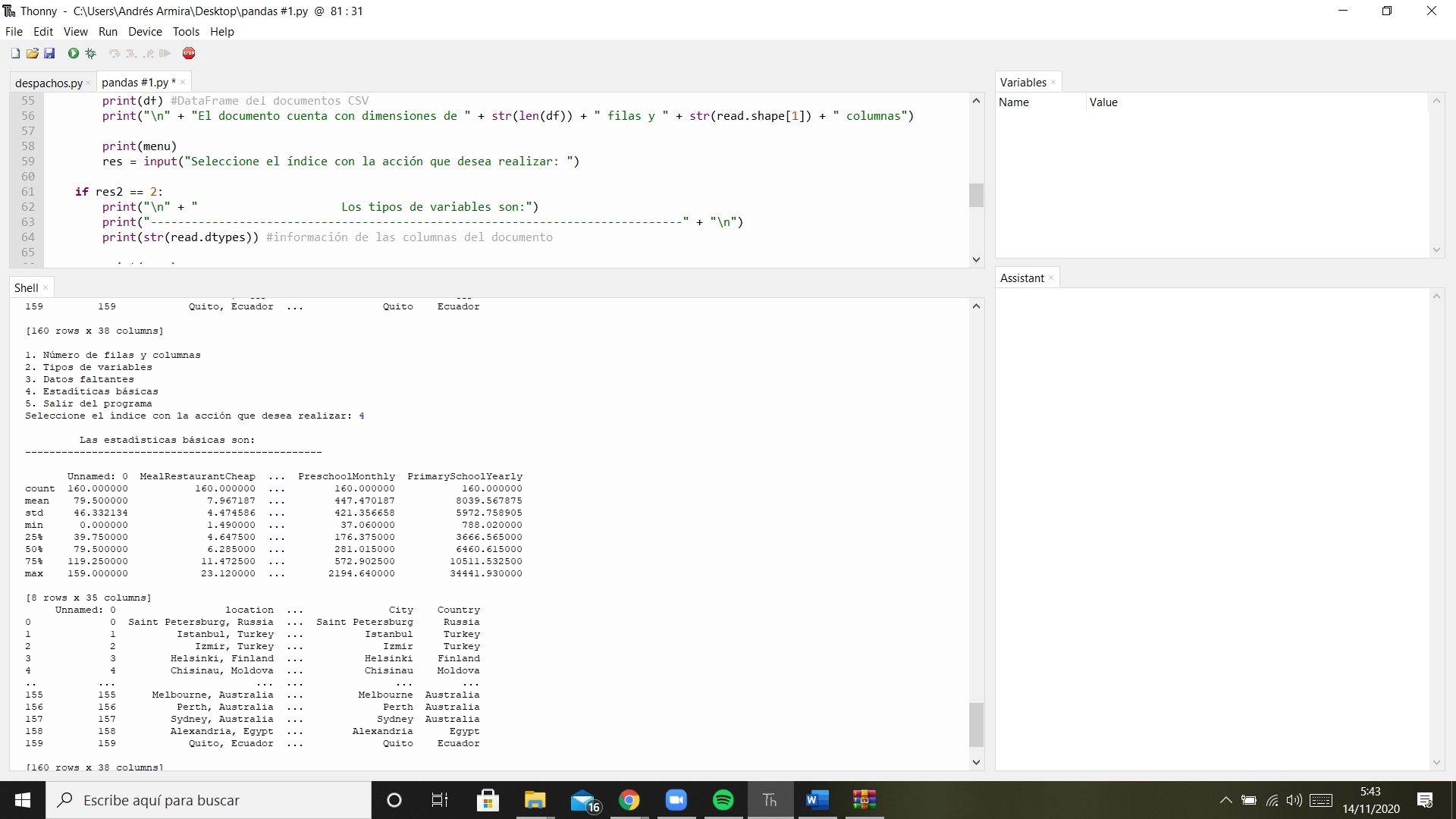
**3. ¿Hay variables categóricas?**

R/ Existen variables de comida del hogar, comida rápida, bebidas, transporte, abarrotería, ropa y entretenimiento.

**4. ¿Hay datos faltantes en los registros?**

R/ Ninguno, el CSV ya está limpio.

**5. ¿Cuáles son las estadísticas (media, mínimo, máximo, etc.) descriptivas de cada dato? Si hay variables categóricas, ¿qué valores las describen?**



**Evaluación:**

**[35 pts.] Debe realizar el análisis, diseño (algoritmo) y programa.**

**a. [15 pts.] Análisis:**

**i. [05 pts.] ¿Qué acciones debe poder hacer su programa? Enumérelas.**

1. Mostrar información del dataframe.

2. Mostrar la estadística descriptiva básica.

3. Mostrar datos faltantes.

**ii. [05 pts.] ¿Con qué va a trabajar (datos)?**

R/ Archivo CSV, “costo\_de\_vida”.

**iii. [05 pts.] ¿Qué requisitos y limitaciones debe tomar en cuenta? ¿Qué cálculos debe hacer? Incluya**

**cualquier explicación de cómo diseñará su solución.**

R/ El programa debe mostrar: filas, columnas, tipos de variables, datos y la estadística descriptiva de cada uno. Para mostrar los datos anteriores es necesario utilizar el módulo “Pandas” con las funciones de .info() y .describe(). La única limitación dentro del código es que no muestra las 160 filas.

**b. [5 pts.] Diseño:**

**i. ¿Qué funciones de PANDAS utilizará para resolver el problema? Explique cómo esas funciones le**

**ayudarán a resolver lo que se le pide.**

* pd.read\_csv(archivo) lee el archivo CSV como un archivo dentro de pandas.
* pd.DataFrame(read) encuentra el dataframe del archivo.
* (read.dtypes) brinda la información de las columnas del documento.
* read.head() dropna bota las filas sin espacios y obtiene todas las filas disponibles.
* read.describe() muestra la estadística descriptiva del CSV.
* No. de columnas: len(read\_csv).
* No. de filas: read\_csv.shape[1].
* Tipos de variables: .info() y .dtypes().
* Datos faltantes: "NaN" y .dropna().

d. [5 pts] Solución

i. Conteste a las preguntas planteadas en la descripción del ejercicio

e.

[55 puntos] Ejercicio 2 – trabajando con los datos

a. ¿Cuál es la variable que más influye en el Costo de Vida Total? ¿Cuál es la segunda variable más influyente?

3. Una de las herramientas más importantes en el análisis de datos es la visualización de estos a través de gráficas:

a. Realice una gráfica de barras mostrando, para los 10 primeros países del DataFrame, el Costo de Vida Total.

b. Realice una gráfica que muestre la relación que hay entre todas las variables, tomadas de dos en dos.

[55 pts.] Debe realizar el análisis, diseño (algoritmo) y programa para resolver el problema anterior.

a. [10 pts] Análisis:

i. [05 pts] ¿Qué acciones adicionales debe poder hacer su programa? Enumérelas.

ii. [05 pts.] ¿Con qué va a trabajar (datos)?

iii. [05 pts.] ¿Qué requisitos y limitaciones debe tomar en cuenta? ¿Qué cálculos debe hacer? Incluya

cualquier explicación de cómo diseñará su solución.

b. [15 pts.] Diseño:

i. ¿Qué funciones de PANDAS utilizará para resolver el problema? Explique cómo esas funciones le ayudarán a resolver lo que se le pide.

c. [30 pts.] Programa:

i. [03 pts.] Encabezado y documentación.

ii. [05 pts.] ¿El programa establece correctamente cinco grupos países en orden de mayor a menor Costo de Vida Total?

iii. [05 pts.] ¿El programa muestra los rangos en los que se agruparon los Países?

iv. [07 pts.] ¿El programa calcula correctamente el Costo de Vida Total de los países?

v. [03 pts.] ¿El programa despliega correctamente las variables más influyentes en el Costo de Vida Total?

vi. [07 pts.] ¿El programa muestra las gráficas requeridas?