Instructivo

1. Identificar y cargar las bases de datos de SIPSA que encontrará en la carpeta compartida, los nombres de los datos son: “microdato-abastecimiento-20xx”.

|  |
| --- |
| RESPUESTA |
| Para para los cálculos se creo una base de datos **compilada** con cada uno de los “microdato-abastecimiento-20xx” en Excel, esto con el objetivo de hacer más ágil el trabajo de limpieza de la base, además del ahorro en esfuerzo computacional. En este orden de ideas, se eliminaron los índices, se agregaron las columnas de día, mes, año y semestre para cada año (2020-2022); sin contar que se eliminaron las decoraciones y solo se conservaron los valores de cada data set |
| **La base a ser trabajada quedo de siguiente manera** |
|  |
| **Para integar esta base de datos en Python se uso el siguiente codigo** |
| df = pd.read\_excel(r'C:\Users\huawei\OneDrive - ANLA - Autoridad Nacional de Licencias Ambientales\Escritorio\wetransfer\_prueba\_2023-07-26\_1451\prueba\compilado.xlsx') |

1. Eliminar las categorías “Código departamento” y “Código municipio” (“las categorías” hace referencia a los nombres de las columnas)

|  |
| --- |
| RESPUESTA |
| Para eliminar las columnas descritas en el enunciado se requiere que el programa busque en cada hoja del libro de Excel y las borre, lo que debido a la cantidad de datos toma una cantidad importante de tiempo. No obstante, mediante el uso de un loop se logro reducir el tiempo de espera en la mitad |
| **Para realizar lo anteriormente descrito se empleó el siguiente código** |
| dfs = []  file\_path = r'C:\Users\huawei\OneDrive - ANLA - Autoridad Nacional de Licencias Ambientales\Escritorio\wetransfer\_prueba\_2023-07-26\_1451\prueba\compilado.xlsx'  # Read the Excel file with multiple sheets outside the loop  df = pd.read\_excel(file\_path, sheet\_name=['2020\_1', '2020\_2', '2021\_1', '2021\_2', '2022\_1', '2022\_2'])  # Store individual DataFrames in the dfs list  for sheet\_name in ['2020\_1', '2020\_2', '2021\_1', '2021\_2', '2022\_1', '2022\_2']:      dfs.append(df[sheet\_name])  col\_to\_delete = ['Código Departamento', 'Código Municipio']  # Remove the specified columns from each DataFrame  for df in dfs:      # Check if the columns exist before attempting to drop them      columns\_to\_drop = [col for col in col\_to\_delete if col in df.columns]      df.drop(labels=columns\_to\_drop, axis=1, inplace=True, errors='ignore') |
| **De este modo, la base inicial pasa de esto** |
|  |
| **A esto** |
|  |

1. Procesar la base manteniendo todas las categorías resultantes del paso anterior pero cambiando la periodicidad de diaria a mensual (para agregar categorías de diarias a mensuales utilizar la “suma de las cantidades en kilos”)

|  |
| --- |
| RESPUESTA |
| Para agrupar las categorías por la suma de las cantidades en kilos (variable Cant Kg) según el mes en que se encuentre, se emplea la columna agregada al inicio con el número del mes (enero=1, diciembre =12). De esta forma se suman los valores de Cant Kg, dado el numero del mes en que se encuentre de la siguiente manera |
| combined\_df = pd.concat(dfs, ignore\_index=True)  sum\_by\_month = combined\_df.groupby('mes ')['Cant Kg'].sum().reset\_index()  sum\_by\_month.head(10) |
| Lo que nos arroja como resultado un nuevo data frame de (nx1) de dimensión, de la siguiente forma |
|  |
| Posteriormente agregamos esta nueva columna a nuestro data frame original, con el siguiente código |
| combined\_df = combined\_df.merge(sum\_by\_month, on='mes ', suffixes=('', '\_sum'))  combined\_df.head(10) |
| Dejando nuestro nuevo data frame de la siguiente manera |
|  |

1. Con la base procesada vamos a extraer la siguiente información:
   1. Variación anual de la producción de “Verduras y Hortalizas” (Categoría: Grupo) para los meses de mayo 2022 **y** mayo 2021

|  |
| --- |
| RESPUESTA |
| Para el cálculo de la variación anual de la producción de Verduras y Hortalizas se deben aplicar varios filtros al data frame. Los filtros son por: mes, año, grupo y verduras y hortalizas (valor o categoría en la variable grupo); el código empleado fue el siguiente |
| # se filtra el data frame para que solo me muestre los datos para el mes de mayo  may\_df = combined\_df[combined\_df['mes '] == 5]  # de la base anterrior se filtra para  VERDURAS Y HORTALIZAS en la columna grupo  verduras\_hortalizas\_df = may\_df[may\_df['Grupo'] == 'VERDURAS Y HORTALIZAS']  # de la base anterior se separa los datos para los años 2021 y 2022  year\_2021\_df = verduras\_hortalizas\_df[verduras\_hortalizas\_df['año '] == 2021]  year\_2022\_df = verduras\_hortalizas\_df[verduras\_hortalizas\_df['año '] == 2022]  # ya se ha filtrado por mes, grupo, año y por la categoria VERDURAS Y HORTALIZAS, ahora se calcula la cantidad producida en ese tiempo empleando la suma de la cantidaes en kilos para cada año  sum\_2021 = year\_2021\_df['Cant Kg\_sum'].sum()  sum\_2022 = year\_2022\_df['Cant Kg\_sum'].sum()  # se calcula la variacion anual para los dos años  annual\_variation = ((sum\_2022 - sum\_2021) / sum\_2021) \* 100 |
| **Como resultado de la operación tenemos** |
| **Cantidad de verduras y hortalizas para el mes de mayo en el 2021** = 80431321378379.42 kg  **Cantidad de verduras y hortalizas para el mes de mayo en el 2022** = 95913550770086.67 kg  **Variación en la cantidad de kg producidos en el mes de mayo para los años 2021 y 2022** = 19.249% |

* 1. Variación anual de la producción de “Yuca” (Categoría: Alimentos) para los meses de mayo 2022 **y** mayo 2021

|  |
| --- |
| RESPUESTA |
| De forma similar a la pregunta anterior se emplea la misma rutina, solo que se cambian los parámetros, por lo que el código empleado tiene la siguiente forma |
| # se filtra el data frame para que solo me muestre los datos para el mes de mayo  #may\_df2 = combined\_df[combined\_df['mes '] == 5]  # de la base anterrior se filtra para Yuca en la columna Alimento  yuca\_df = may\_df[may\_df['Alimento'] == 'Yuca']  # de la base anterior se separa los datos para los años 2021 y 2022  yuca\_2021\_df = yuca\_df[yuca\_df['año '] == 2021]  yuca\_2022\_df = yuca\_df[yuca\_df['año '] == 2022]  # ya se ha filtrado por mes, año, alimento y por la categoria VERDURAS Y, ahora se calcula la cantidad producida en ese tiempo empleando la suma de la cantidaes en kilos para cada año  sum\_yuca\_2021 = yuca\_2021\_df['Cant Kg\_sum'].sum()  sum\_yuca\_2022 = yuca\_2022\_df['Cant Kg\_sum'].sum()  # se calcula la variacion anual para los dos años  annual\_variation\_yuca = ((sum\_yuca\_2022 - sum\_yuca\_2021) / sum\_yuca\_2021) \* 100 |
| **Como resultado de la operación tenemos** |
| **Cantidad de yuca para el mes de mayo en el 2021** = 3610474596725.991kg  **Cantidad de yuca para el mes de mayo en el 2022** = 3490661194487.411kg  **Variación en la cantidad de kg producidos en el mes de mayo para los años 2021 y 2022** = -3.318% |

* 1. Variación mensual de la producción de “Frutas” (Categoría: Grupo) **tomando solamente** tipos de “Banano” (Categoría: Alimento) para el mes de mayo 2022

|  |
| --- |
| RESPUESTA |
| Para este ejercicio es necesario filtrar los datos para el mes de mayo y abril del 2022, de ese modo encontraremos la variación mensual. En ese orden de ideas se filtran los datos por mes y año, luego se filtra por la característica frutas de la columna grupo y los tipos diversos de banano ('Banano bocadillo', 'Banano criollo', 'Banano Urabá'), finalmente se calcula la cantidad producida en ese periodo de tiempo. El código empleado para los cálculos anteriores es el siguiente |
| # se filtra el data frame para solo mostrar valores para el mes de mayo de 2022  may\_2022\_df = combined\_df[(combined\_df['mes '] == 5) & (combined\_df['año '] == 2022)]  # seguidamente se filtra el anterior data frame segun la caracteristica frutas de la colunma grupo y los tipos diversos de banano  banana\_types = ['Banano bocadillo', 'Banano criollo', 'Banano Urabá']  frutas\_banano\_may\_2022\_df = may\_2022\_df[(may\_2022\_df['Grupo'] == 'FRUTAS') & (may\_2022\_df['Alimento'].isin(banana\_types))]  # se calcula la cantidad de banano y sus variedades producida en mayo del 2022  sum\_may\_2022 = frutas\_banano\_may\_2022\_df['Cant Kg\_sum'].sum()  ################################### se repite el anteior procedimiento para el mes de abril de 2022 para comparar ###############################3  april\_2022\_df = combined\_df[(combined\_df['mes '] == 4) & (combined\_df['año '] == 2022)]  frutas\_banano\_april\_2022\_df = april\_2022\_df[(april\_2022\_df['Grupo'] == 'FRUTAS') & (april\_2022\_df['Alimento'].isin(banana\_types))]  sum\_april\_2022 = frutas\_banano\_april\_2022\_df['Cant Kg\_sum'].sum()  # se calcula la variacion mensual  monthly\_variation = ((sum\_may\_2022 - sum\_april\_2022) / sum\_april\_2022) \* 100 |
| **Como resultado de la operación tenemos** |
| **Cantidad de yuca para el mes de mayo en el 2021** = 5.387.220.000.000 kg  **Cantidad de yuca para el mes de mayo en el 2022** = 5,386,217,000,000 kg  **Variación en la cantidad de kg producidos en el mes de mayo para los años 2021 y 2022** = 0.018% |

* 1. Variación anual de la producción de “Frutas” (para el mes de abril 2022) **excluyendo** las frutas importadas (las frutas importadas lo dicen de forma explícita en la categoría de “Alimento”; por ejemplo, puedes encontrar “ciruela” y “ciruela importada”)

|  |
| --- |
| RESPUESTA |
| Para este ejercicio iniciamos filtrando el data frame para que solo muestre información para abril de 2022, seguidamente se excluyen las frutas catalogadas como importadas del DF y se calculan las cantidades de fruta producidas para ese periodo de tiempo. Se repite el mismo proceso anterior para abril del 2021, por lo que el código empleado es el siguiente |
| # se filtra el data frame para que solo muestre informacion para abril de 2022  april\_2022\_df = combined\_df[(combined\_df['mes '] == 4) & (combined\_df['año '] == 2022)]  # se excluyen las frutas catalogadas como importadas del DF  excluded\_items = ['Arveja seca importada', 'Ciruela importada', 'Durazno importado', 'Frutas importadas otras', 'Manzana importada', 'Pera importada', 'Uva importada']  frutas\_april\_2022\_filtered\_df = april\_2022\_df[~april\_2022\_df['Alimento'].isin(excluded\_items)]  # se calculan las candidades producidas para ese periodo de tiempo  sum\_april\_2022\_excluded = frutas\_april\_2022\_filtered\_df['Cant Kg\_sum'].sum()  ############################################### se repite el mismo proceso anterior para abril del 2021 ############################  april\_2021\_df = combined\_df[(combined\_df['mes '] == 4) & (combined\_df['año '] == 2021)]  frutas\_april\_2021\_filtered\_df = april\_2021\_df[~april\_2021\_df['Alimento'].isin(excluded\_items)]  sum\_april\_2021\_excluded = frutas\_april\_2021\_filtered\_df['Cant Kg\_sum'].sum()  # se calcula la variacion anual de produccion de frutas menos las importadas  annual\_variation = ((sum\_april\_2022\_excluded - sum\_april\_2021\_excluded) / sum\_april\_2021\_excluded) \* 100  # se imprimen los resultados  print("Total 'Cant Kg\_sum' for Frutas in April 2022 (excluding specified items):", sum\_april\_2022\_excluded)  print("Total 'Cant Kg\_sum' for Frutas in April 2021 (excluding specified items):", sum\_april\_2021\_excluded)  print("Annual variation in 'Cant Kg\_sum' for Frutas between April 2021 and April 2022 (excluding specified items):", annual\_variation, "%") |
| **Como resultado de la operación tenemos** |
| **Cantidad de yuca para el mes de mayo en el 2021** = 220636200000000.kg  **Cantidad de yuca para el mes de mayo en el 2022** = 219512400000000 kg  **Variación en la cantidad de kg producidos en el mes de mayo para los años 2021 y 2022** = -0.509% |

* 1. Variación bienal de la producción de todos los tipos de “papa” que tienen como departamento de procedencia a Nariño para el mes de febrero 2022.

|  |
| --- |
| RESPUESTA |
| En este punto se realizo un filtrado por mes y año (febrero y años 2020 -2022), luego se seleccionaron los datos para todas aquellas variedades de papa para los años de 2020 y 2022.  En ese orden de ideas, los datos obtenidos se filtran una vez mas para el depto. de Nariño, se calculan las cantidades producidas para esos periodos de tiempo, para luego estimar la variación bienal, de forma que el código empleado es el siguiente |
| # se filtran los datos de modo que se muestre informacion solo para febrero del 2022  february\_2022\_df = combined\_df[(combined\_df['mes '] == 2) & (combined\_df['año '] == 2022)]  # se filtran los datos de modo que se muestre informacion solo para febrero del 2020  february\_2020\_df = combined\_df[(combined\_df['mes '] == 2) & (combined\_df['año '] == 2020)]  # se filtran las filas para "papa" en la columna 'Grupo' y tipos específicos de "papa" en la columna 'Alimento' para febrero de 2022  papa\_types = ['Papa capira', 'Papa suprema', 'Papa criolla', 'Papa superior', 'Papa única', 'Papa R-12', 'Papa parda pastusa', 'Papa rubí', 'Papa Betina', 'Papa nevada', 'Papa Morasurco']  papa\_narino\_feb\_2022\_df = february\_2022\_df[(february\_2022\_df['Grupo'] == 'TUBERCULOS, RAICES Y PLATANOS') & (february\_2022\_df['Alimento'].isin(papa\_types))]  # se filtran las filas para "papa" en la columna 'Grupo' y tipos específicos de "papa" en la columna 'Alimento' para febrero de 2020  papa\_types = ['Papa capira', 'Papa suprema', 'Papa criolla', 'Papa superior', 'Papa única', 'Papa R-12', 'Papa parda pastusa', 'Papa rubí', 'Papa Betina', 'Papa nevada', 'Papa Morasurco']  papa\_narino\_feb\_2020\_df = february\_2020\_df[(february\_2020\_df['Grupo'] == 'TUBERCULOS, RAICES Y PLATANOS') & (february\_2020\_df['Alimento'].isin(papa\_types))]  # seguidamente se filtran las filas para el depto de Nariño en la columna de depto de procedencia  papa\_narino\_feb\_2022\_df = papa\_narino\_feb\_2022\_df[papa\_narino\_feb\_2022\_df['Departamento Proc.'] == 'NARIÑO']  papa\_narino\_feb\_2020\_df = papa\_narino\_feb\_2020\_df[papa\_narino\_feb\_2020\_df['Departamento Proc.'] == 'NARIÑO']  # se separa el data frame para los dos años del periodo bienal (2020-2022)  year\_2022\_df = papa\_narino\_feb\_2022\_df[papa\_narino\_feb\_2022\_df['año '] == 2022]  year\_2020\_df = papa\_narino\_feb\_2020\_df[papa\_narino\_feb\_2020\_df['año '] == 2020]  # se calcula la suma de las cantidades producidas para cada año  sum\_2022 = year\_2022\_df['Cant Kg\_sum'].sum()  sum\_2020 = year\_2020\_df['Cant Kg\_sum'].sum()  # se calcula la variacion bienal  biennial\_variation = ((sum\_2022 - sum\_2020) / sum\_2020) \* 100  # resultados  print("Total 'Cant Kg\_sum' for all types of papa from Nariño in February 2022:", sum\_2022)  print("Total 'Cant Kg\_sum' for all types of papa from Nariño in February 2020:", sum\_2020)  print("Biennial variation in 'Cant Kg\_sum' for all types of papa from Nariño between February 2020 and February 2022:", biennial\_variation, "%") |
| **Como resultado de la operación tenemos** |
| **Cantidad de yuca para el mes de mayo en el 2020** = 5,723,106,000,000 kg  **Cantidad de yuca para el mes de mayo en el 2022** = 4,095,697,000,000 kg  **Variación en la cantidad de kg producidos en el mes de mayo para los años 2021 y 2022** = -28.435% |

Favor adjuntar todo lo que se considere necesario en la justificación de los resultados. El script usado en documento aparte y resultados explícitos por pregunta.