

Reto 4 – Fundamentos de programación

Contexto: Los datos del siguiente archivo son una muestra aleatoria del registro de casos de covid-19 que podemos encontrar en: <https://www.datos.gov.co/Salud-y-Proteccion-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data>

Hipotéticamente, usted forma parte de un grupo de investigación el cual trabaja en la evaluación del impacto de esta enfermedad en la sociedad. Esto implica que se debe realizar un estudio estadístico de los casos de COVID en Colombia y así identificar patrones que nos permitan tomar medidas más eficientes para minimizar este impacto negativo. Para llevar a cabo esta tarea se le suministro la información necesaria en un archivo CSV, usted debe realizar una función en Python que lea estos datos y entregue los siguientes indicadores:

- Promedio de edades de los infectados por departamentos
- Promedio de porcentaje de infectados masculinos y femeninos
- Porcentaje de la procedencia de los infectados

Como ya sabemos, los datos leídos se organizan en un DataFramer que cuentan con la siguiente estructura:

Index	D de casc	Fecha de notificación	igo DIVIP	Ciudad de ubicación	Departamento o Distrito	atención	Edad	Sexo	Tipo	Estado	País de procedencia	FIS	Fecha de m
83	84	2020-03-16T00:00:00.000	76001	Cali	Valle del Cauca	Recuperado	51	M	Importado	Leve	ECUADOR	2020-03-10T00:00:00.000	nan
84	85	2020-03-17T00:00:00.000	76001	Cali	Valle del Cauca	Recuperado	42	F	Relacionado	Asintomático	nan	nan	nan
85	86	2020-03-15T00:00:00.000	66001	Pereira	Risaralda	Recuperado	29	M	Importado	Leve	FRANCIA	2020-03-13T00:00:00.000	nan
86	87	2020-03-16T00:00:00.000	25126	Cajicá	Cundinamarca	Recuperado	45	M	Importado	Leve	ESPAÑA	2020-03-11T00:00:00.000	nan
87	88	2020-03-15T00:00:00.000	73001	Ibagué	Tolima	Recuperado	36	M	Relacionado	Leve	nan	2020-03-11T00:00:00.000	nan
88	89	2020-03-14T00:00:00.000	68001	Bucaramanga	Santander	Recuperado	39	F	Importado	Leve	ESPAÑA	2020-03-14T00:00:00.000	nan
89	90	2020-03-15T00:00:00.000	41001	Neiva	Huila	Recuperado	53	F	Relacionado	Leve	nan	2020-03-15T00:00:00.000	nan
90	91	2020-03-17T00:00:00.000	25126	Cajicá	Cundinamarca	Recuperado	32	M	Importado	Leve	ESPAÑA	2020-03-13T00:00:00.000	nan
91	92	2020-03-11T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	50	M	Relacionado	Leve	nan	2020-03-10T00:00:00.000	nan
92	93	2020-03-18T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	26	F	Importado	Leve	ESPAÑA	2020-03-11T00:00:00.000	nan
93	94	2020-03-17T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	53	M	Importado	Leve	ESTADOS UNIDOS	2020-03-10T00:00:00.000	nan
94	95	2020-03-17T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	69	M	Importado	Leve	TURQUÍA	2020-03-16T00:00:00.000	nan
95	96	2020-03-17T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	37	F	Importado	Leve	ESTADOS UNIDOS	2020-03-15T00:00:00.000	nan
96	97	2020-03-16T00:00:00.000	13001	Cartagena de Indias	Cartagena D.T. y C.	Recuperado	24	M	Importado	Leve	ESTADOS UNIDOS	2020-03-15T00:00:00.000	nan
97	98	2020-03-16T00:00:00.000	13001	Cartagena de Indias	Cartagena D.T. y C.	Recuperado	21	M	Importado	Leve	ESPAÑA	2020-03-13T00:00:00.000	nan
98	99	2020-03-16T00:00:00.000	73001	Ibagué	Tolima	Recuperado	16	M	Relacionado	Leve	nan	2020-03-13T00:00:00.000	nan
99	100	2020-03-16T00:00:00.000	66001	Pereira	Risaralda	Recuperado	21	M	Importado	Leve	ESPAÑA	2020-03-12T00:00:00.000	nan
100	101	2020-03-18T00:00:00.000	66170	Dosquebradas	Risaralda	Recuperado	18	M	Importado	Leve	ESPAÑA	2020-03-11T00:00:00.000	nan
101	102	2020-03-16T00:00:00.000	66001	Pereira	Risaralda	Recuperado	79	M	Importado	Leve	ESPAÑA	2020-03-09T00:00:00.000	nan
102	103	2020-03-18T00:00:00.000	13001	Cartagena de Indias	Cartagena D.T. y C.	Recuperado	24	F	Importado	Leve	ESPAÑA	2020-03-11T00:00:00.000	nan
103	104	2020-03-19T00:00:00.000	8001	Barranquilla	Barranquilla D.E.	Recuperado	48	F	Importado	Leve	ESPAÑA	2020-03-15T00:00:00.000	nan
104	105	2020-03-19T00:00:00.000	76520	Palma	Valle del Cauca	Recuperado	23	F	Importado	Leve	ESPAÑA	2020-03-09T00:00:00.000	nan
105	106	2020-03-19T00:00:00.000	5001	Medellín	Antioquia	Recuperado	64	M	Importado	Leve	ESPAÑA	2020-03-13T00:00:00.000	nan
106	107	2020-03-12T00:00:00.000	5001	Medellín	Antioquia	Recuperado	56	M	Importado	Leve	ESPAÑA	2020-03-10T00:00:00.000	nan
107	108	2020-03-17T00:00:00.000	5001	Medellín	Antioquia	Recuperado	57	F	Importado	Leve	ESPAÑA	2020-03-14T00:00:00.000	nan

Debe tener en cuenta que, en la columna del sexo del infectado, se cometieron errores en el ingreso de datos, por lo tanto, debemos limpiar la base de datos haciendo que los únicos valores en esa columna sean “F” para femenino y “M” para masculino, ambas letras en mayúsculas.

Orden de la información

La información debe estar organizada en 3 diccionarios uno para cada indicador:

Diccionario 1: Debe contener los porcentajes por sexos, este diccionario debe tener como llaves las letras “F” y “M” para femenino y masculino respectivamente, y como valor de estas llaves el porcentaje correspondiente redondeado con 4 cifras decimales.

Ejemplo: {F: 55.5, M: 44.5}

Diccionario 2: Debe contener el promedio de edades por departamento, las llaves de este diccionario deben ser el nombre del departamento, mientras que los valores de estas llaves es el promedio de las edades de los casos en el respectivo departamento. Estos valores deben estar redondeados con 4 cifras decimales.

Ejemplo: {Antioquía: 23.4554, Risaralda: 35.33, Amazonas: 31.4432}

Diccionario 3: El diccionario debe contar con el porcentaje de la procedencia de los infectados, es decir, de aquellos casos que llegaron a Colombia desde otro país y calcular el porcentaje correspondiente de cada país. Las llaves de este diccionario deben ser los países de procedencia de los infectados y los valores de estas debe ser el porcentaje redondeado con 4 cifra decimales de infectado procedentes de ese país.

Ejemplo {España: 20.32, Ecuador: 53.33, Francia: 26.35}

Nota: Los departamentos y países dependen de la información en el documento.

Una vez organizada la información estos diccionarios la función debe retornar una lista con estos diccionarios:

[Diccionario1, Diccionario2, Diccionario3]

Esqueleto de la función

```
def casos(ruta_archivo: str)-> dict:
    pass
```

La respuesta esperada es la siguiente:

```
[
{
  'F': 45.01,
  'M': 54.99
},
{
  'Amazonas': 38.4545,
  'Antioquia': 40.5,
  'Archipiélago de San Andrés Providencia y Santa Catalina': 44.6667,
  'Atlántico': 37.3636,
  'Barranquilla D.E.': 39.2708,
  'Bogotá D.C.': 40.8306,
  'Bolívar': 47.2308,
  'Boyacá': 43.2459,
  'Buenaventura D.E.': 39.9615,
  'Caldas': 40.2553,
  'Caquetá': 37.3333,
  'Cartagena D.T. y C.': 42.5385,
  'Casanare': 37.8,
  'Cauca': 40.5122,
  'Cesar': 39.9143,
  'Chocó': 40.4444,
  'Cundinamarca': 38.5761,
  'Córdoba': 43.6129,
  'Huila': 46.5714,
  'La Guajira': 41.7333,
  'Magdalena': 37.8429,
  'Meta': 35.7033,
  'Nariño': 37.8577,
  'Norte de Santander': 43.2444,
  'Quindío': 44.1692,
  'Risaralda': 38.8411,
  'Santa Marta D.T. y C.': 43.8095,
  'Santander': 41.3, 'Sucre': 28.0,
  'Tolima': 35.3592,
  'Valle del Cauca': 41.641
}
```

```
},  
  
{  
  'ALEMANIA': 1.23,  
  'ANTILLAS NEERLANDESAS': 0.984,  
  'ARABIA SAUDÍ': 0.123,  
  'ARGENTINA': 0.369,  
  'ARUBA': 0.369,  
  'AUSTRALIA': 0.123,  
  'BRASIL': 4.428,  
  'BÉLGICA': 0.123,  
  'CANADÁ': 0.369,  
  'CHILE': 0.738,  
  'COSTA RICA': 0.246,  
  'CROACIA': 0.123,  
  'CUBA': 0.492,  
  'ECUADOR': 5.9041,  
  'EGIPTO': 1.353,  
  'EMIRATOS ARABES UNIDOS': 0.123,  
  'ESPAÑA': 31.3653,  
  'ESTADOS UNIDOS': 24.7232,  
  'ESTADOS UNIDOS DE AMÉRICA': 0.123,  
  'FRANCIA': 2.091,  
  'GRECIA': 0.492,  
  'GUATEMALA': 0.123,  
  'HONDURAS': 0.123,  
  'IRLANDA': 0.123,  
  'ISRAEL': 0.738,  
  'ITALIA': 1.107,  
  'JAMAICA': 1.968,  
  'KENYA': 0.123,  
  'MARRUECOS': 0.246,  
  'MÉXICO': 5.5351,  
  'PANAMA': 4.182,  
  'PERU': 1.23,  
  'PUERTO RICO': 1.599,  
  'REINO UNIDO DE GRAN BRETAÑA E IRLANDA DEL NORTE': 1.722,  
  'REPÚBLICA DOMINICANA': 1.107,  
  'SUECIA': 0.123,  
  'SUIZA': 0.123,  
  'TURQUÍA': 3.321,  
  'URUGUAY': 0.123,  
  'VENEZUELA': 0.492  
}]
```