

Analisis de Covid 19

Nota: el dataset contiene informacion sobre los casos de Covid que se presentaron en Colombia durante el año 2020.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Importacion de dataset

```
datos = pd.read_csv('covid.csv', low_memory = False)
datos.drop(columns = ['Código DIVIPOLA departamento', 'Código DIVIPOLA
municipio', 'Unidad de medida de edad', 'Código ISO del
país', 'Pertenencia étnica', 'Nombre del grupo étnico'], inplace = True)
datos.rename(columns={'Nombre departamento': 'Departamento', 'Nombre
municipio': 'Municipio', 'Recuperado': 'Evolucion'}, inplace = True)
datos['fecha reporte web'] = pd.to_datetime(datos['fecha reporte
web'])
datos['Mes'] = datos['fecha reporte web'].dt.month
```

Analisis Descriptivo

Variables

Categoricas:

- Departamento
- Municipio
- Sexo
- Estado
- Evolucion
- Tipo de Recuperación

Numericas:

- Edad

```
v = ['Departamento', 'Municipio', 'Sexo', 'Evolucion', 'Tipo de
recuperación']
for variable in v:
    datos[variable] = datos[variable].str.capitalize()

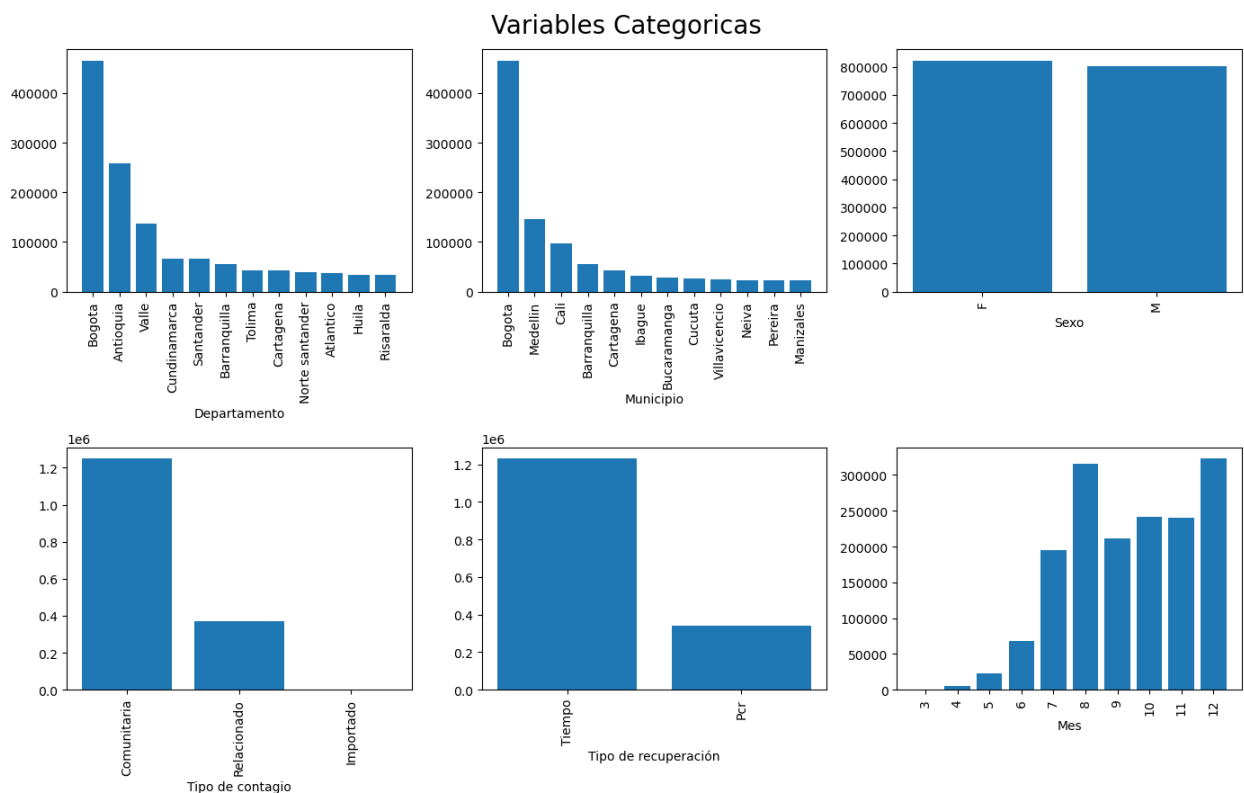
fig, axs = plt.subplots(2, 3, figsize = (14,9))
fig.suptitle('Variables Categoricas', fontsize = 20)
variables = ['Departamento', 'Municipio', 'Sexo', 'Tipo de contagio',
'Tipo de recuperación', 'Mes']
```

```

for i, var in enumerate(variables):
    a, b = np.unravel_index(i, (2, 3))
    conteo = datos[var].value_counts().head(12)
    axs[a, b].bar(conteo.index, conteo)
    axs[a, b].set_xlabel(var)
    axs[a, b].set_xticks(conteo.index) # Establecer las posiciones de
las etiquetas del eje x
    axs[a, b].set_xticklabels(conteo.index, rotation=90) # Rotar
etiquetas del eje x

plt.tight_layout() # Ajustar diseño
plt.show()

```



```

datos[['Edad', 'Mes']].describe().round(3)#mean()#value_counts()

```

	Edad	Mes
count	1623865.000	1623865.000
mean	39.972	9.405
std	17.476	1.934
min	3.000	3.000
25%	27.000	8.000
50%	37.000	9.000
75%	52.000	11.000
max	89.000	12.000

Tras realizar graficos de barras para las variables categoricas y resumen de las varibales numericas se obtiene lo siguiente:

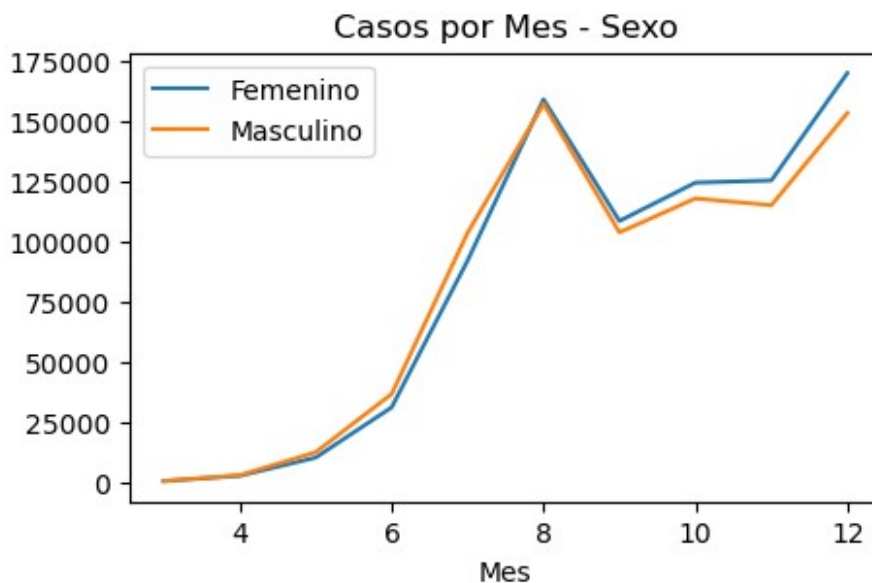
- Bogota, Medellin y Cali representan las ciudades con mayor cantidad de casos reportados.
- La variable Sexo no es relevante por si sola ya que se ve como existe una similitud en casos por sexo.
- La mayoría de casos tienen un tipo de contagio Comunitario y el tipo de recuperacion es por el tiempo.
- El mes con mayor cantidad de casos es Diciembre y el comportamiento de esta variable es ascendente, resalta el mes Agosto donde hay un pico.

```
tb = datos.groupby(['Mes', 'Sexo'])['Sexo'].count()
tb = tb.reset_index(name='Cantidad')

tb_femenino = tb[tb['Sexo']=='F']
tb_masculino = tb[tb['Sexo']=='M']

plt.figure(figsize = (5,3))
plt.plot(tb_femenino['Mes'],tb_femenino['Cantidad'],label =
'Femenino')
plt.plot(tb_masculino['Mes'],tb_masculino['Cantidad'],label =
'Masculino')
plt.xlabel('Mes')
plt.title('Casos por Mes - Sexo')

plt.legend()
plt.show()
```



```

tbl = datos.groupby(['Mes','Tipo de contagio'])['Tipo de
contagio'].count()
tbl = tbl.reset_index(name = 'Cantidad')

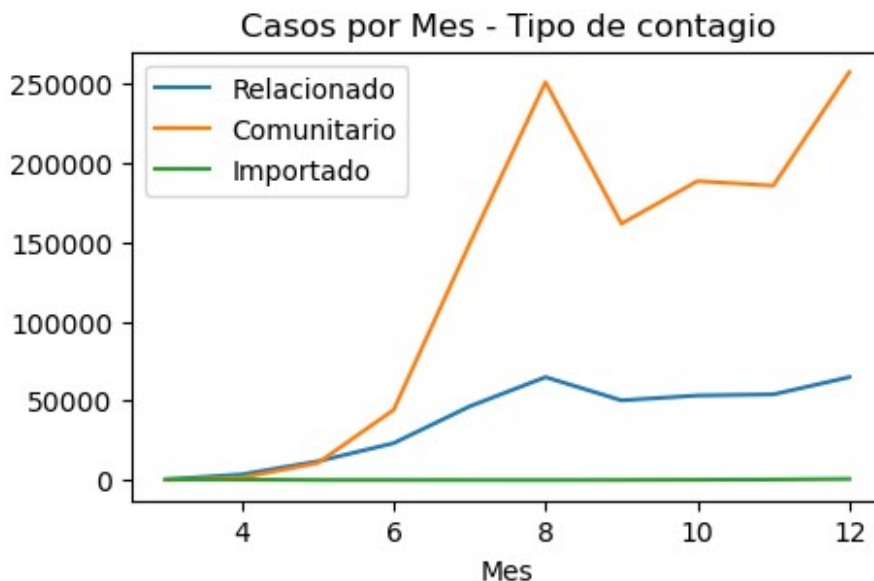
tbl_relacionado = tbl[tbl['Tipo de contagio']== 'Relacionado']
tbl_comunitaria = tbl[tbl['Tipo de contagio']== 'Comunitaria']
tbl_importado = tbl[tbl['Tipo de contagio']== 'Importado']

plt.figure(figsize =(5,3))

plt.plot(tbl_relacionado['Mes'],tbl_relacionado['Cantidad'], label =
'Relacionado')
plt.plot(tbl_relacionado['Mes'],tbl_comunitaria['Cantidad'],label =
'Comunitario')
plt.plot(tbl_relacionado['Mes'],tbl_importado['Cantidad'],label =
'Importado')

plt.legend()
plt.title('Casos por Mes - Tipo de contagio')
plt.xlabel('Mes')
plt.show()

```



```

tb2 = datos.groupby(['Mes','Tipo de recuperación'])['Tipo de
recuperación'].count()
tb2 = tb2.reset_index(name = 'Cantidad')

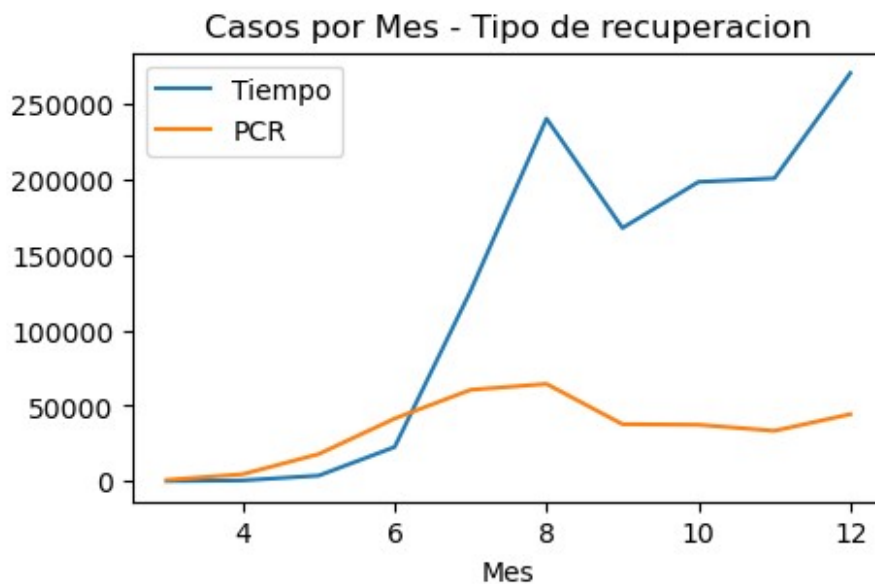
tb2_tiempo = tb2[tb2['Tipo de recuperación']== 'Tiempo']
tb2_pcr = tb2[tb2['Tipo de recuperación']== 'Pcr']

plt.figure(figsize =(5,3))

```

```
plt.plot(tb2_tiempo['Mes'],tb2_tiempo['Cantidad'], label = 'Tiempo')
plt.plot(tb2_pcr['Mes'],tb2_pcr['Cantidad'],label = 'PCR')

plt.legend()
plt.title('Casos por Mes - Tipo de recuperacion')
plt.xlabel('Mes')
plt.show()
```



Con las graficas agrupadas se puede determinar lo siguiente:

- Para el sexo se puede ver que en el mes 8 hay un cambio de comportamiento donde los casos de sexo femenino empiezan a superar los casos de sexo masculino, justo en el mes que sucede el pico.
- respecto a los tipos de contagio tanto el comunitario como el relacionado presentan una baja en el mes 8 y un aumento para el mes 12, sucede con mayor fuerza para el contagio comunitario.
- El tipo de recuperacion tiene el mismo comportamiento del tipo de contagio, la mayoría de pacientes se les determino su recuperacion segun la variable tiempo. Cabe resaltar como del mes 8 a 11 PCR mantiene una tendencia a la baja.

Probabilidades

```
densidad, bins = np.histogram(datos['Edad'], bins=15, density=True)
resultados = pd.DataFrame({'Clase inferior': bins[:-1],
                           'Clase superior': bins[1:],
                           'Densidad P Edad': densidad})
resultados
```

	Clase inferior	Clase superior	Densidad P Edad
0	3.000000	8.733333	0.003572

1	8.733333	14.466667	0.004893
2	14.466667	20.200000	0.009925
3	20.200000	25.933333	0.018166
4	25.933333	31.666667	0.027167
5	31.666667	37.400000	0.024407
6	37.400000	43.133333	0.020694
7	43.133333	48.866667	0.013140
8	48.866667	54.600000	0.015089
9	54.600000	60.333333	0.013235
10	60.333333	66.066667	0.009134
11	66.066667	71.800000	0.005366
12	71.800000	77.533333	0.004560
13	77.533333	83.266667	0.003240
14	83.266667	89.000000	0.001831

```
densidad1, bins1 = np.histogram(datos['Mes'], bins = 15, density = True)
```

```
resultados1 = pd.DataFrame({'Clase inferior':bins1[:-1],
                             'Clase superior':bins1[1:],
                             'Densidad P Mes':densidad1})
```

```
resultados1
```

	Clase inferior	Clase superior	Densidad P Mes
0	3.0	3.6	0.000924
1	3.6	4.2	0.005614
2	4.2	4.8	0.000000
3	4.8	5.4	0.023165
4	5.4	6.0	0.000000
5	6.0	6.6	0.069314
6	6.6	7.2	0.200439
7	7.2	7.8	0.000000
8	7.8	8.4	0.323949
9	8.4	9.0	0.000000
10	9.0	9.6	0.217472
11	9.6	10.2	0.248262
12	10.2	10.8	0.000000
13	10.8	11.4	0.246305
14	11.4	12.0	0.331222