

Analisis descriptivo de fallecidos por COVID-19

Caso de Estudio

Durante la pandemia de COVID-19, Colombia enfrentó grandes desafíos en términos de salud pública, con miles de personas fallecidas en distintas regiones del país.

Este análisis busca entender los factores clave que influyeron en las tasas de mortalidad, considerando aspectos temporales, geográficos y demográficos. El país necesita una comprensión más profunda de cómo la pandemia afectó a diferentes regiones y grupos poblacionales, para identificar los factores de riesgo y mejorar las respuestas futuras ante emergencias sanitarias.

Tecnologias a Usar

- VsCode
 - Python y Librerias para el Analisis de datos
 - Power BI para las visualizaciones
-

Obtencion, preparacion y analisis exploratorio de los datos

- Se tiene un dataset procedente de la pagina del gobierno nacional de colombia (Fuente Confiable)
 - El dataset de 143.125 registros y 23 columnas
 - existen registros en blanco
-

Manipulacion y Limpieza de datos

Conversion de tipos de datos, llenado de registros vacios y normalizacion de texto

```
# Cambiar el nombre de las columnas capitalizando la primera
df = df.rename(columns=lambda x: x.capitalize())

# Cambio de tipo de dato a columnas de fecha
columnas_fecha = ['Fecha de notificación', 'Fecha de muerte',
for columnas in columnas_fecha:
    df[columnas] = pd.to_datetime(df[columnas], errors='coerce')

#Seleccionar columnas de tipo fecha aleatoriamente
df.select_dtypes(include='datetime64[ns]').sample(10)

# Llenado de valores nan con la moda o valor que mas se repite
df['Nombre del grupo étnico'] = df['Nombre del grupo étnico'].value_counts(dropna=False)

# Llenado de valores nan con una categoría 'Desconocido y N/A'
df['Código iso del país'] = df['Código iso del país'].fillna('Desconocido')
df['Nombre del país'] = df['Nombre del país'].fillna('Desconocido')
df['Fecha de diagnóstico'] = df['Fecha de diagnóstico'].fillna('Desconocido')
df['Tipo de recuperación'] = df['Tipo de recuperación'].fillna('Desconocido')

# Correccion de valores de la columna 'Recuperado'
df['Recuperado'] = df['Recuperado'].str.replace('fallecido', 'fallecido por COVID-19')
```

Codigo del Analisis

```
# Media de edades por genero
media_edades_sexo = df.groupby('Sexo')['Edad'].mean()
media_edades_sexo

# Medidas de tendencia Central
print('La Mediana de la edad es: ', df['Edad'].median())
```

```

print('La Moda de la edad es: ', df['Edad'].mode()[0])

# Desviacion estandar para la muestra -> ddof = 1
print('Desviacion estandar columnas numericas: \n', df.std(nu
print('Promedio de las edades: ', df['Edad'].median())

# Mortalidad por genero
contagio_por_genero = df.groupby('Sexo')['Recuperado'].count(
contagio_por_genero

# Contagio mas recurrentes
tipo_contagio = df['Tipo de contagio'].value_counts()
tipo_contagio

```

Analisis descriptivo de fallecidos

¿Media de las edades de los fallecidos?

La media muestral indica que hay en promedio un 69.68% de fallecidos Femeninos y 67.69% de fallecidos Hombres.

¿mediana y moda para la distribucion de las edades?

- La Mediana general de la edad es: 70.0
- La Moda general de la edad es: 72

Esto refleja que las personas de una edad avanzada son mas propensos a contraer la enfermedad.

¿Desviacion estandar para evaluar la variabilidad?

- Desviacion estandar para la edad: 15.572765411243607

Esto quiere decir que existe una Heterogeneidad alta en la muestra, en pocas palabras las edades de las personas son bastante dispersas. hay personas muy mayores y personas muy por debajo de la media de 70 años que son afectados por la enfermedad.

¿Hubieron mas contagios de hombres o mujeres?

Genero	# Fallecidos
F	56383
M	86742

Segun estos datos la tasa de mortalidad es mas alta en Hombres que mujeres, esto supone que el hombre era el mas expuesto a la enfermedad.

¿Que tipo de contagio fue mas recurrente en los casos?

Tipo de contagio	# Casos
Comunitaria	112423
Relacionada	30608
Importado	94

Esto Evidencia que los contagios fueron en su mayoria por comunidad, esto quiere decir que hubieron focos de propagacion y no se tomaron las medidas adecuadas para mitigar la enfermedad.

De esta tabla puede desprenderse analisis de contagios por region y evidenciar cuales de las medidas para la propagacion del virus fue escaza.

Graficos en Power BI

