

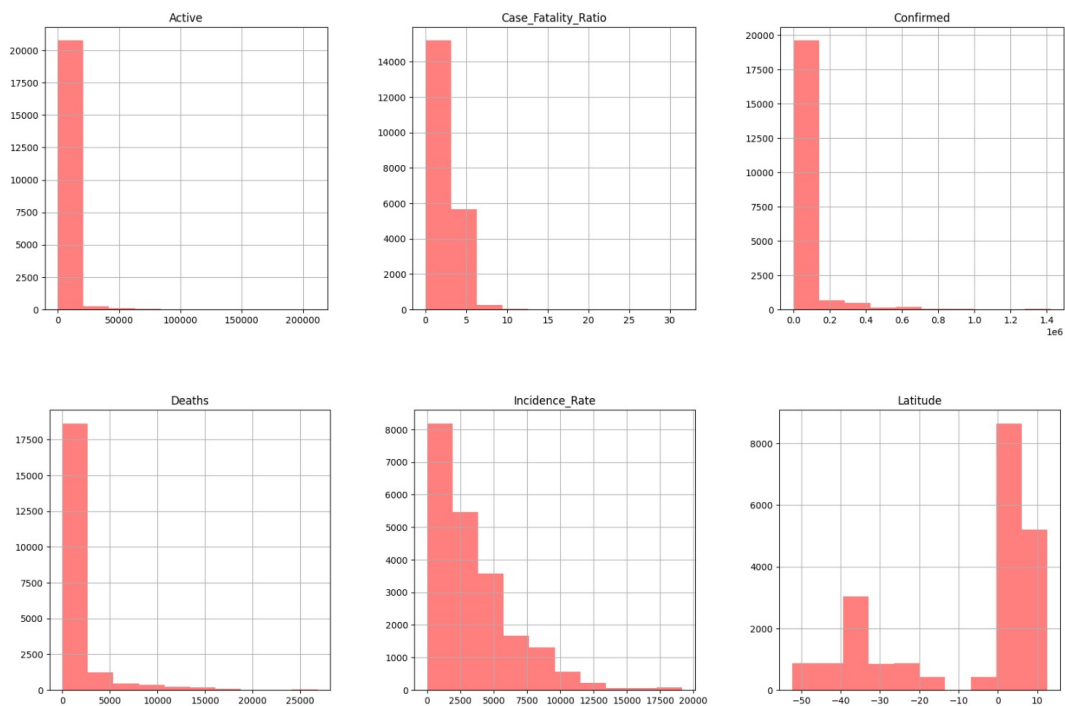
## ANÁLISIS DESCRIPTIVO

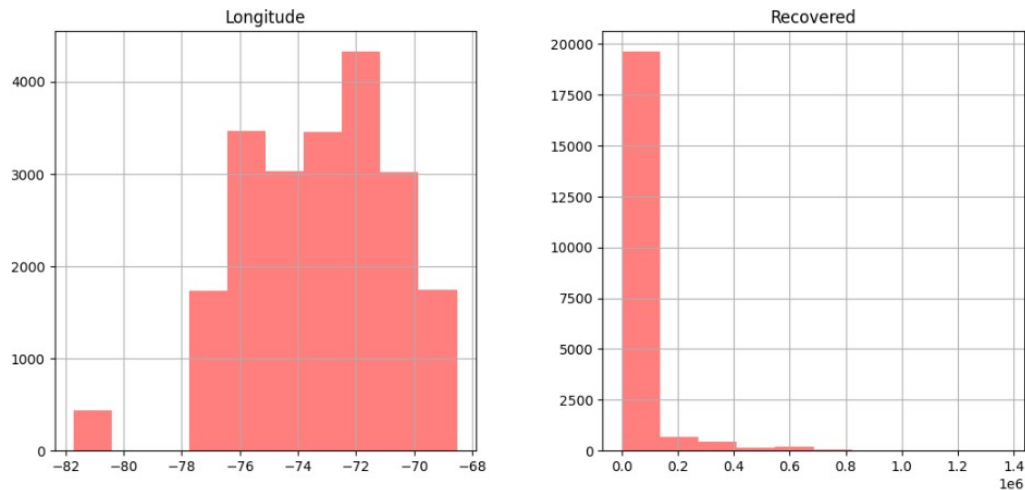
Para la realización de este paso, se tomo el data set “casosdiarios”. Con el fin de ver la distribución e los datos. Se realiza una limpieza retirando las columnas de fecha, latitud, longitud y provincia ya que no son relevantes para la investigación.

```
[15] dfu_shapiro = df[['Active', 'Case_Fatality_Ratio', 'Confirmed', 'Deaths', 'Incidence_Rate', 'Recovered']]
dfu_shapiro.head()
```

	Active	Case_Fatality_Ratio	Confirmed	Deaths	Incidence_Rate	Recovered
0	632	2.449031	23887	585	3931.796410	22670
1	1114	1.342774	23161	311	2419.600846	21736
2	226	2.072014	10859	225	4803.421979	10408
3	79	1.326291	8520	113	2948.668254	8328
4	137	0.944584	1588	15	1539.386184	1436

Como resultado se adjuntan las imágenes de las gráficas.





Como resultado de se puede inferir que la distribución de datos no tienen un comportamiento normal o acampanado. Para confirmar este hecho, se procede a realizar una prueba Shapiro Wilk, para corroborar si afirmado.

```
[28] for col in dfu_shapiro.select_dtypes(include=['float']).columns:
      dfu_shapiro[col] = dfu_shapiro[col].astype(int)

dfu_shapiro.dtypes

Active          int64
Case_Fatality_Ratio  int64
Confirmed        int64
Deaths           int64
Incidence_Rate    int64
Recovered        int64
dtype: object

stat, p = shapiro(dfu_shapiro.iloc[:, :-1])
print("El valor p de la prueba Shapiro-Wilk es: ", p)

El valor p de la prueba Shapiro-Wilk es: 0.0
```

Podemos inferir que la prueba de Shapiro Wilk confirma que los datos no son acampanados o normales.