

ANÁLISIS DE SEÑALES

Julián Andrés Rueda Gutierrez^{1*}

Abstract

El presente trabajo tiene como objetivo realizar un algoritmo , haciendo usos de tecnologías tales como : Machine Learning , Inteligencia artificial y derivados tales como Redes Neuronales para poder realizar la extraccion de datos de la web y poder almacenar dicha data y posteriormente realizar un analisis de la misma y poder obtener informacion relevante a partir de graficas, adicionalmente a partir de un modelo de machine learning haciendo uso de la regresion lineal poder predecir datos futuros en base a la data actual recolectada y de esta manera tener un panorama mas clara del contexto actual en relacion al covid en colombia

Keywords

Analisis de datos — Inteligencia Artificial — Redes Neuronales – Machine Learning – Big data – Covid-19

¹ Ingeniería electronica, Universidad Sergio Arboleda, Analisis de señales

² Departamento de Ingeniería, Universidad Sergio Arboleda, Bogotá, Colombia

*Autor: julian.rueda@correo.usa.edu.co

Contents

Introducción	1
1 Procedimiento	1
1.1 Web Scrapping	1
1.2 Analisis de datos	3
2 Resultados	3
2.1 Graficas	3
Graficos de barras Colombia • Graficos de torta Colombia • Graficos de dispersión Colombia • Graficos Curvas de Aprendizaje • Graficos de barras Bogota • Matriz de correlación • Mapa de calor • Funcion de entrenamiento y prediccion (Regresion Lineal)	
2.2 Conclusiones	3
References	3

Introducción

Mucho se ha dicho respecto a los uso de las nuevas tecnologías tales como la inteligencia artificial , Machine Learning, Redes Neuronales entre otras que hacer parte de una nueva revolucion tecnologica en los tiempos contemporáneos. Desde los revolucionarios aportes de John McCarthy en 1956 en donde se acuño del termino inteligencia artificial, hasta las recientes experiencias en el analisis de datos y el uso de los mismos para crear desde un asistente virtual hasta proveer recomendaciones a los usuarios en base a una serie de preferencias, nuestro entendimiento de estas tecnologías ha variado tanto, en sentido a la variedad de usos tan impredecibles que puede tener en la actualidad , Tal como se muestra en el presente proyecto cuya finalidad es realizar una analisis predictivo e ilustrativo de los datos extraidos a partir de un

algoritmo de machine learning para posteriormente realizar una correspondiente evaluacion de los mismos, y poder identificar diferentes aspectos a partir de las graficas realizadas , analizar aspectos tales como identificar en que edades es mayor en la region de colombia , entre otros muchos aspectos a evaluar a partir de un previo analisis.

1. Procedimiento

Paso 1: Creacion de algoritmo de extraccion de datos de la web para posteriormente ser almacenados y analizados.

Paso 2 : Analisis de la informacion extraida .

Paso 3: Realizacion de graficas descriptivas.

Paso 4 : Creacion de algoritmo para realizar la regresion lineal .

Paso 5 : Realizar el entrenamiento de l modelo propuesto.

Paso 6: Validar el modelo creado mediante un umbral estipulado.

Paso 7: Verificacion de curva de aprendizaje.

- 1 Tabla de casos en ciudades de Colombia
- 7 Tabla de casos por localidades en Bogota
- 3 Tabla de casos por departamento en Colombia
- 9 Tabla de casos por edad en Colombia

1.1 Web Scrapping

El primer paso a realizar para poder realizar el correspondiente analisis de la data fue realizar un algoritmo de extraccion de datos de la web haciendo uso de la libreria en python denominada BeautifulSoup la cual permite identificar los elementos de una pagina web haciendo uso de un metodo denominado Find el cual tiene como objetivo

	Ciudad	Casos
0	Ciudad	Casos
1	ABEJORRAL	96
2	ABREGO	96
3	ABRIAQUI	23
4	ACACIAS	2380
...
1107	ZARZAL	690
1108	ZETAQUIRA	19
1109	ZIPACON	26
1110	ZIPAQUIRA	4412
1111	ZONA BANANERA	329

1112 rows x 2 columns

Figure 1. Tabla de casos en ciudades

```
[[ 'Departamento', 'Casos'],  
 [ 'AMAZONAS', '3159'],  
 [ 'ANTIOQUIA', '220859'],  
 [ 'ARAUCA', '4201'],  
 [ 'ATLANTICO', '33373'],  
 [ 'BARRANQUILLA', '46532'],  
 [ 'BOGOTA', '387418'],  
 [ 'BOLIVAR', '7301'],  
 [ 'BOYACA', '22862'],  
 [ 'CALDAS', '26489'],  
 [ 'CAQUETA', '13619'],  
 [ 'CARTAGENA', '32999'],  
 [ 'CASANARE', '7296'],  
 [ 'CAUCA', '16206'],  
 [ 'CESAR', '30391'],  
 [ 'CHOCO', '4563'],  
 [ 'CORDOBA', '28486'],  
 [ 'CUNDINAMARCA', '54753'],  
 [ 'GUAINIA', '1210'],  
 [ 'GUAJIRA', '12357'],
```

Figure 2. Tabla de casos por departamento

	Localidad	Femenino	Masculino
Id			
1	Usaquén	12123	10657
2	Chapinero	4312	4251
3	Santa Fe	3110	3268
4	San Cristobal	9704	8969
5	Usme	7846	6673
6	Tunjuelito	5057	4872
7	Bosa	16120	14213
8	Kennedy	25701	23375
9	Fontibón	9201	8290
10	Engativá	20777	18953
11	Suba	26839	23523
12	Barrios Unidos	4282	3211

Figure 3. Tabla de casos por localidades en Bogotá

	Edad	Mujeres	Hombres
2	0 a 9	22034	23138
3	10 a 19	43412	43753
4	20 a 29	155295	145124
5	30 a 39	156810	161717
6	40 a 49	111484	110654
7	50 a 59	92832	89590
8	60 a 69	54972	55538
9	70 a 79	28658	30594
10	80 a 89	15418	14795
11	90 a 99	3426	2756
12	100 a 109	149	97
13	110 a 119	1	2

Figure 4. Tabla de casos por edad en Colombia

encontrar segmentos de la pagina web a partir de identificadores de los elementos de la pagina tales como "id"

, "class", por los cuales podemos tener acceso a dicha informacion .

1.2 Analisis de datos

Basados en la data extraida anteriormente y almacenada de manera posterior en una base de datos interna poder identificar y realizar los anlisis respectivos de la manera que se presenta a continuacion

2. Resultados

Posteriormente a la extraccion , validacion y analisis de los datos se pudo identificar como el modelo planteado para realizar una prediccion de la cantidad de hombres y mujeres recuperados de covid entorno a la variable Target "Recuperados" escogida a causa de realizar una pre- via validacion en torno a la matriz de correlacion realizada como se puede identificar en la figura 15 en la cual se puede identificar que es la feature cuya correlacion con las demas variables es mayor , teniendo en cuenta esto podemos identificar como a partir del algoritmo de entrenamiento creado para realizar la regresion lineal correspondiente fue el adecuado , adicionalmente teniendo en cuenta los valores de error dados como se puede identificar en la figura 11 y teniendo en cuenta la grafica dada en la curva de aprendizaje podemos identificar como se va alineando en torno al valor esperado.

2.1 Graficas

Posteriormente a la extraccion de los datos proveidos mediante enlaces externos en la web haciendo uso de machine learning se procede a la realizacion de los graficos entorno a las variables extraidas para realizar un analisis correcto en base a los resultados de las mismas

2.1.1 Graficos de barras Colombia

2.1.2 Graficos de torta Colombia

2.1.3 Graficos de dispersión Colombia

2.1.4 Graficos Curvas de Aprendizaje

2.1.5 Graficos de barras Bogota

2.1.6 Matriz de correlación

2.1.7 Mapa de calor

2.1.8 Funcion de entrenamiento y prediccion (Regresion Lineal)

2.2 Conclusiones

Se puede observar la relevancia que tiene realizar una validacion de los datos perdidos para de esta manera consecuentemente darle tratamiento correcto a los datos y posteriormente realizar un analisis adecuado en torno al modelo realizado de manera tal que al realizar las graficas correspondientes no se encuentren datos perdidos o faltantes y el analisis posterior sea el correcto.

Podemos identificar como el analisis de datos por medio de regresiones lineales , coeficiente de dispersion y graficas descriptivas , ayudan a tener un panorama más claro frente al proceso de darle un valor agregado a la información resultante, asegurando así el debido proceso.

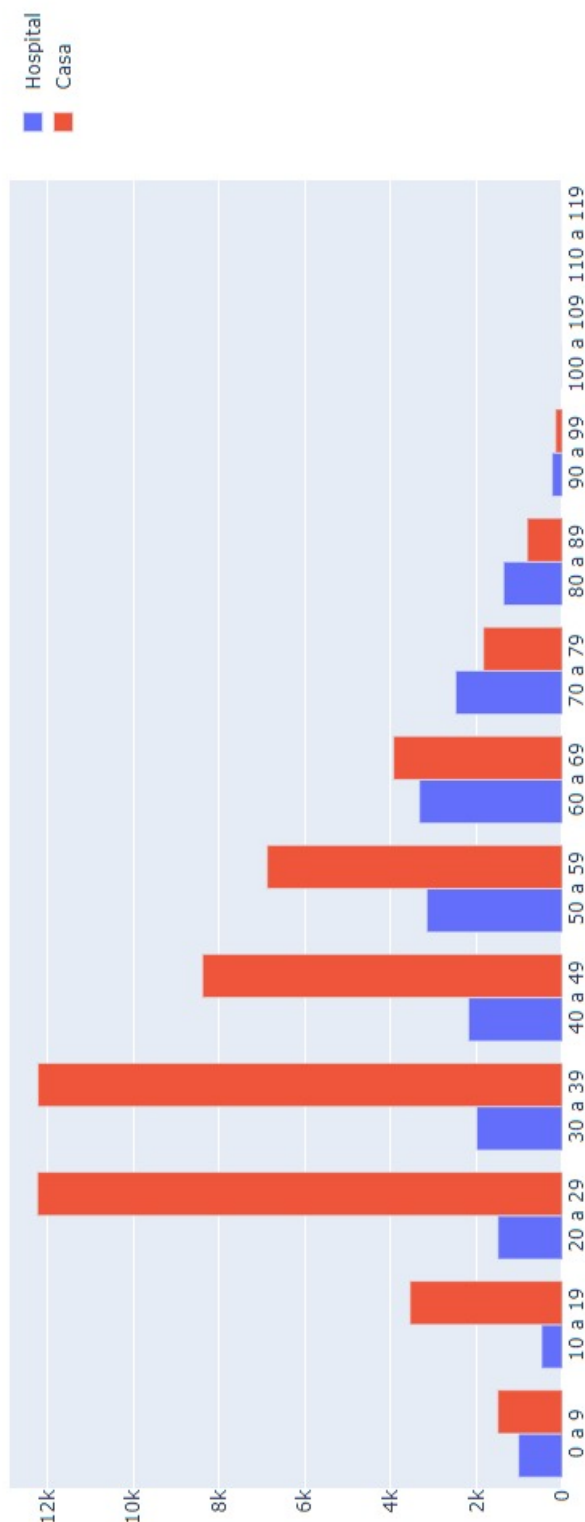


Figure 5. Tabla de casos por edad en casa y hospital

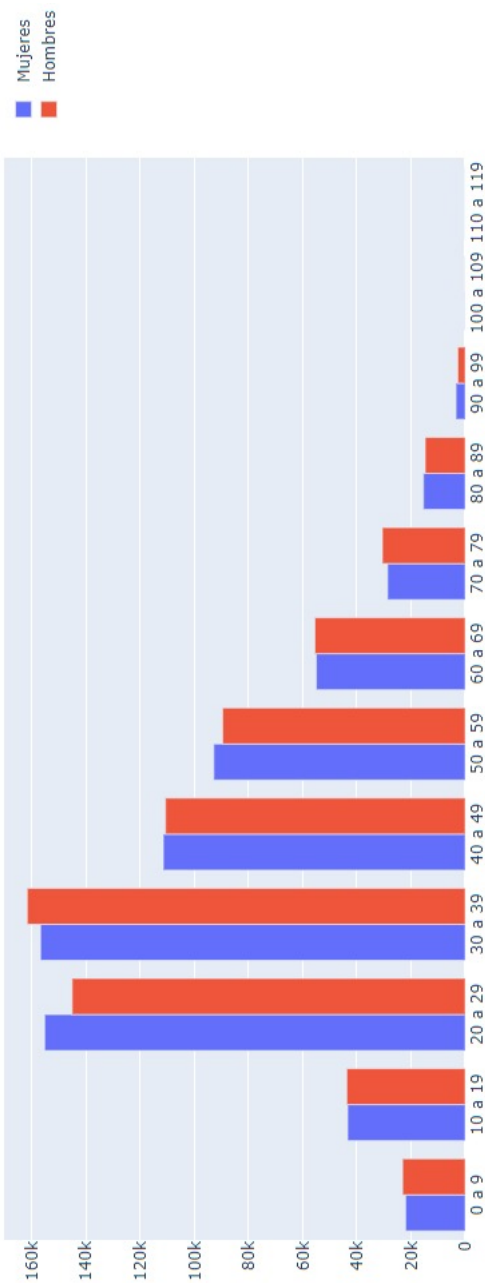


Figure 6. Gráfica Covid por edad y genero

References

[1] wilkinsonpc.
<https://coronavirus.wilkinsonpc.com.co/>

[2] wilkinsonpc.
<https://coronavirus.wilkinsonpc.com.co/casos-por-edad.html>

[3] datosmacro.
<https://datosmacro.expansion.com/otros/coronavirus/colombia>

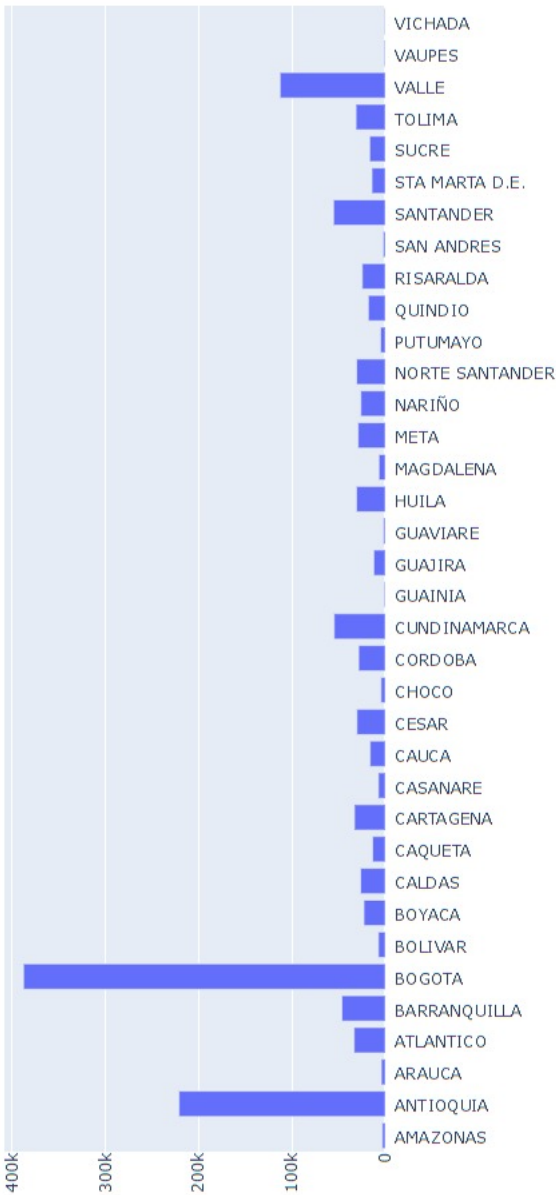


Figure 7. Gráfica Covid por Departamento

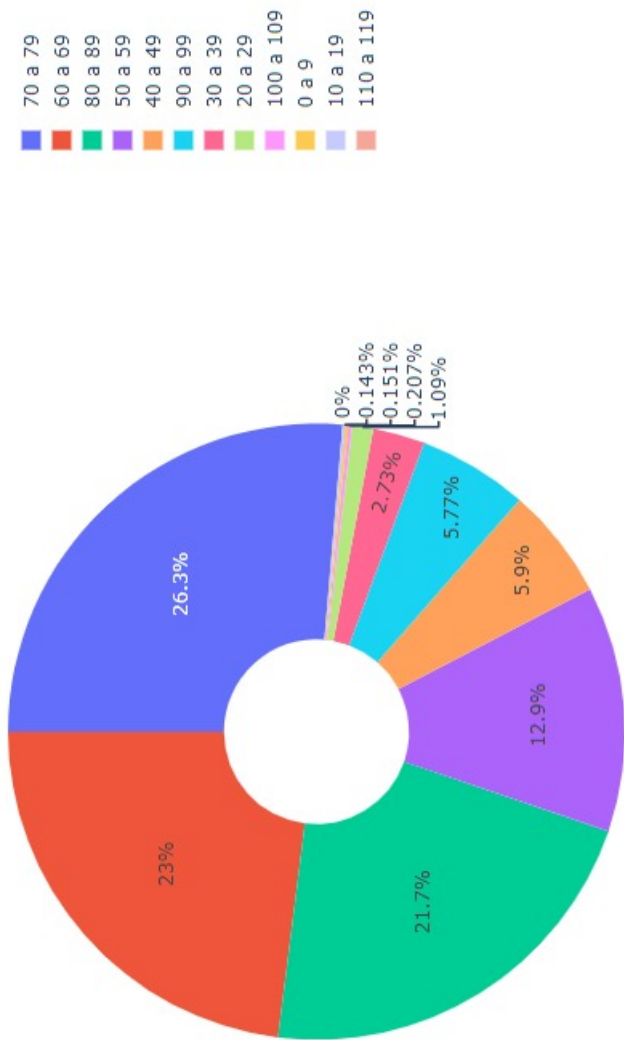


Figure 8. Tabla de fallecidos por edad en colombia

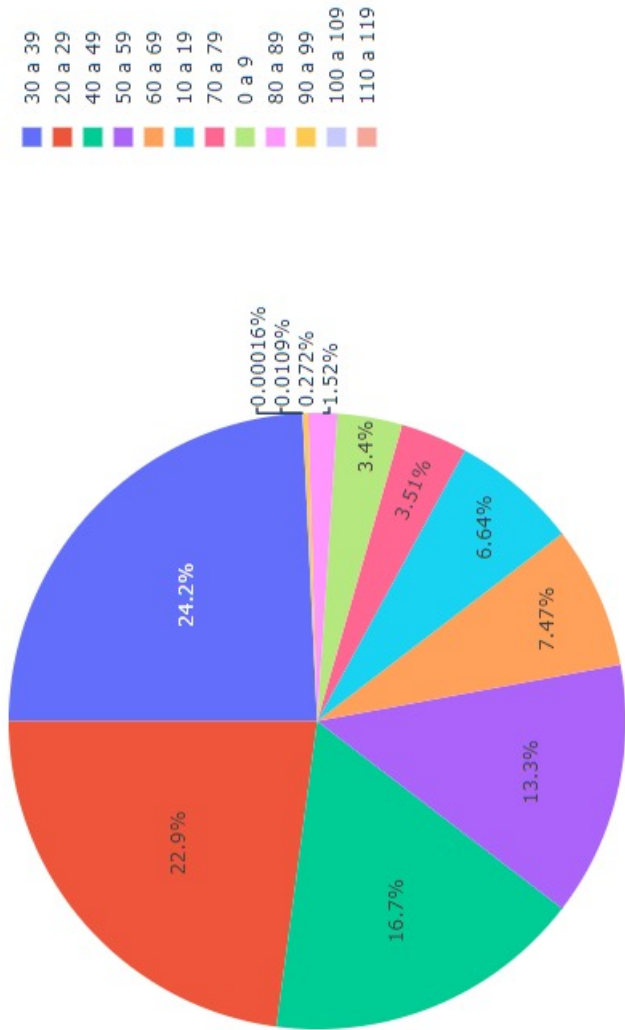


Figure 9. Tabla de casos por edad Recuperados

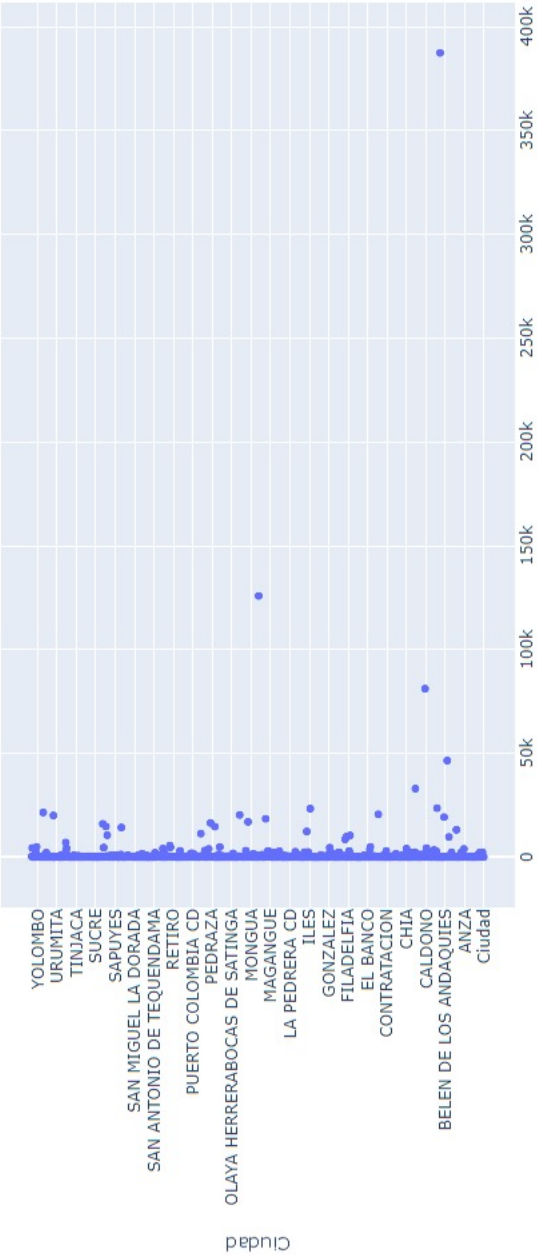


Figure 10. Gráfica Covid por Ciudad

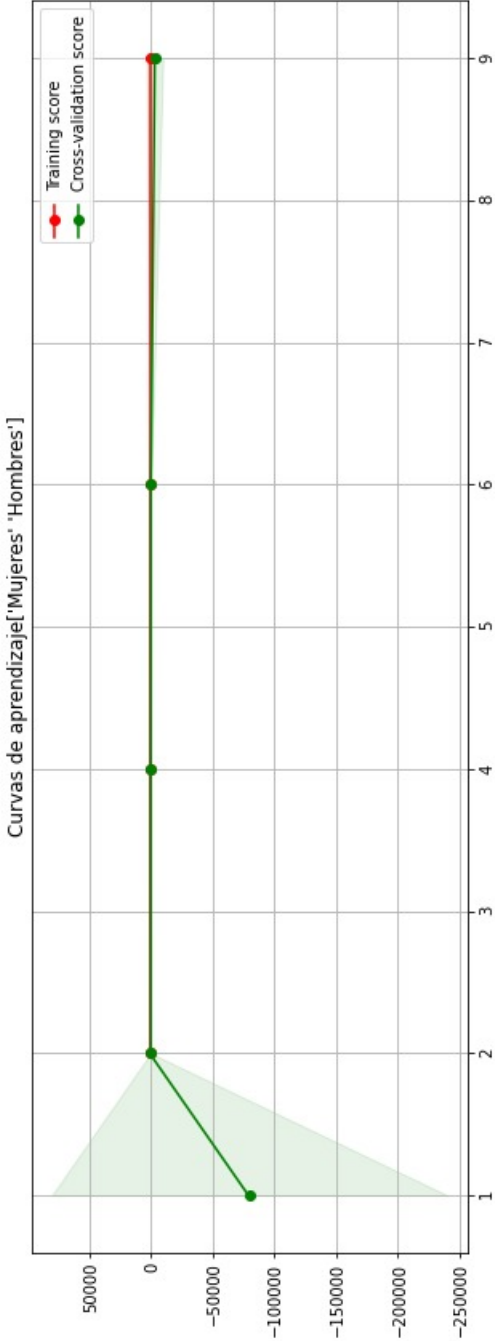


Figure 11. Curva de aprendizaje

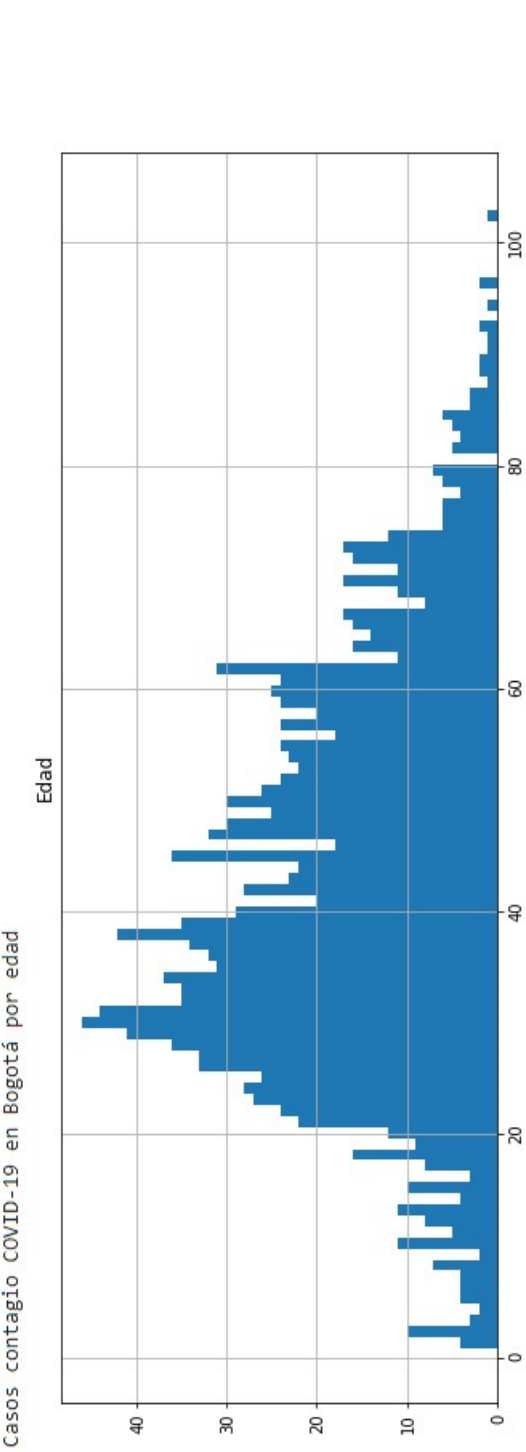


Figure 12. Tabla de casos por edad en Bogota

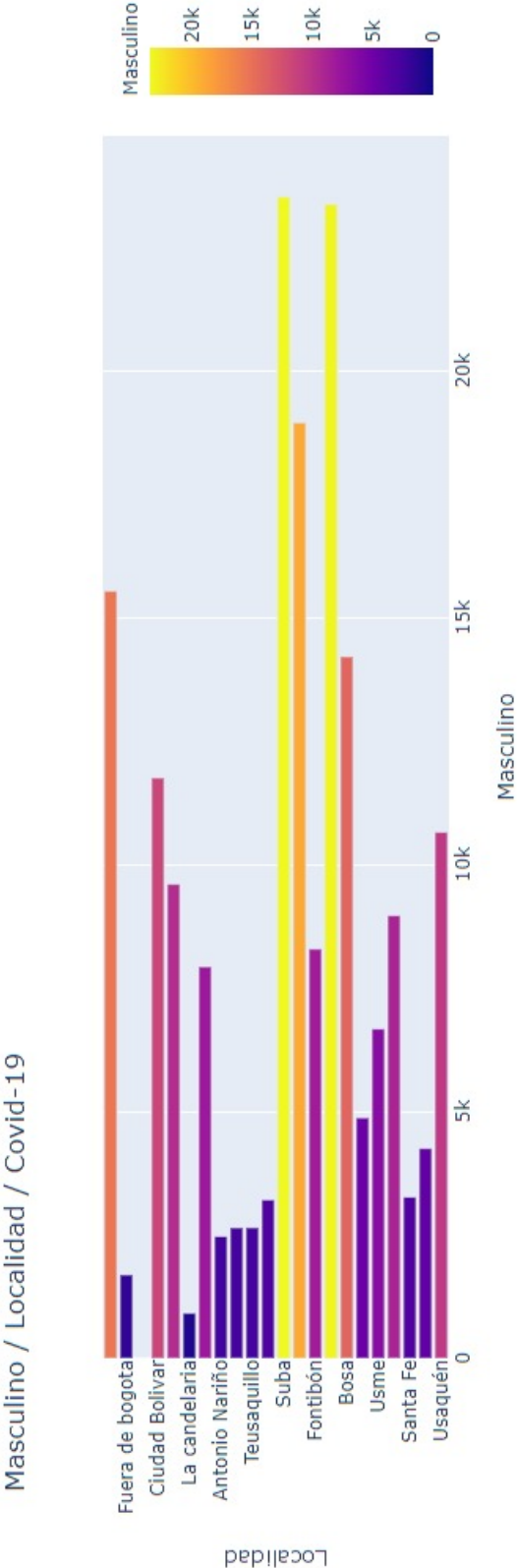


Figure 13. Tabla de casos por localidad y genero Masculino

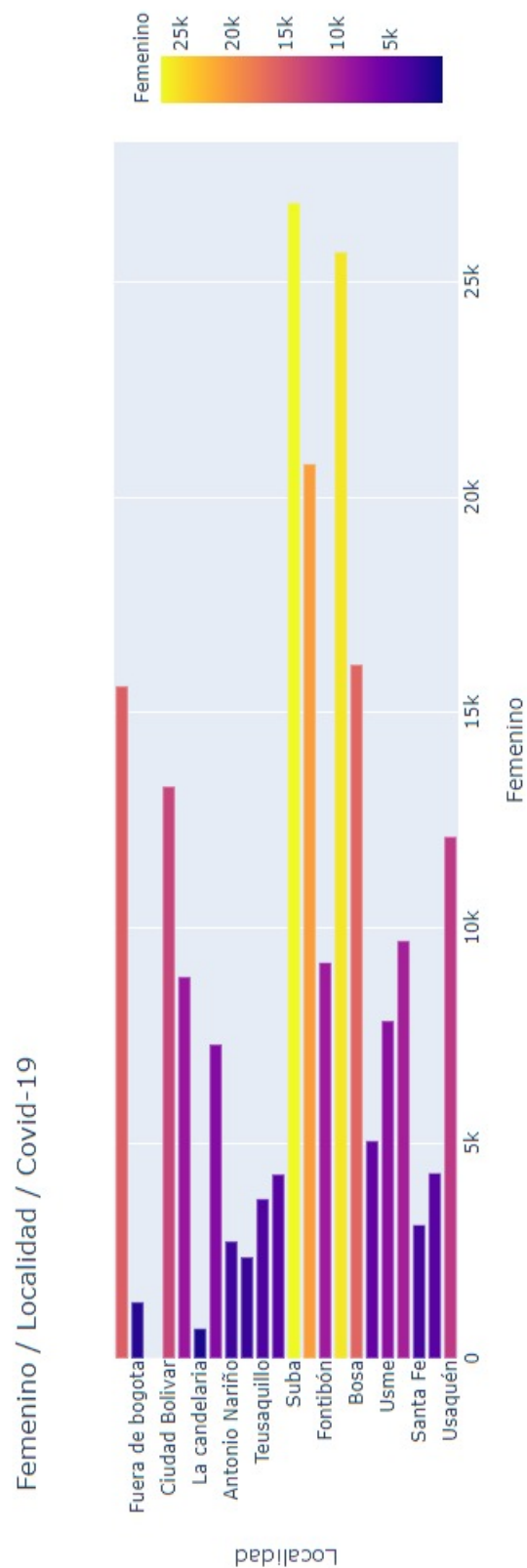


Figure 14. Tabla de casos por localidad y genero Femenino

	Mujeres	Hombres	Recuperados	Fallecidos	Críticos	Hospital
Mujeres	1.000000	0.998219	0.998713	-0.135089	0.286347	0.518631
Hombres	0.998219	1.000000	0.998436	-0.128177	0.296045	0.526960
Recuperados	0.998713	0.998436	1.000000	-0.175199	0.253334	0.488038
Fallecidos	-0.135089	-0.128177	-0.175199	1.000000	0.759141	0.664935
Críticos	0.286347	0.296045	0.253334	0.759141	1.000000	0.955835
Hospital	0.518631	0.526960	0.488038	0.664935	0.955835	1.000000

Figure 15. Matriz de Correlacion

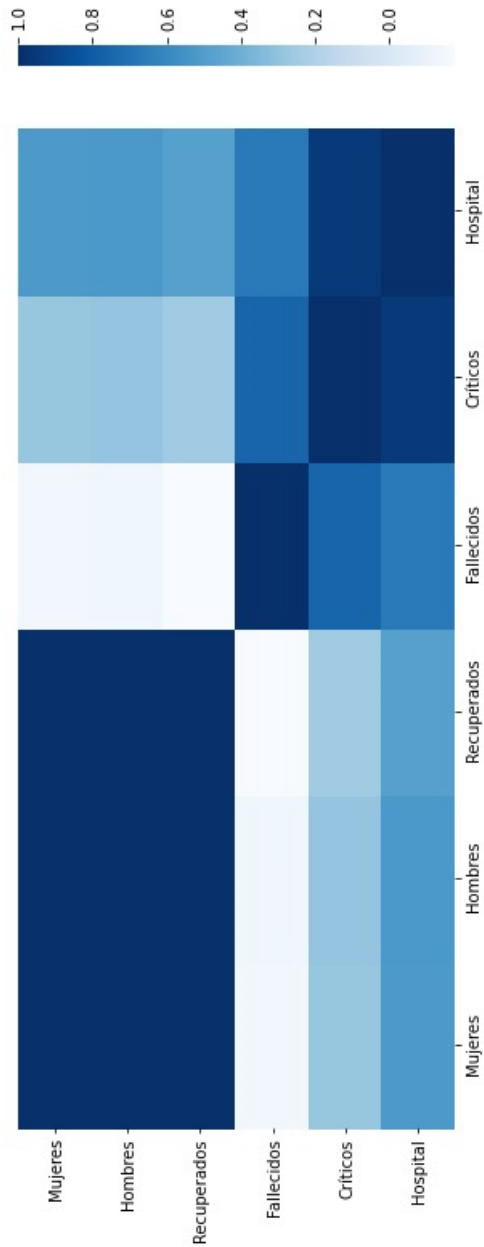


Figure 16. Tabla de casos por localidad y genero Femenino

```
def entrenamiento(dataCorrelacion):
    X = dataCorrelacion.drop('Recuperados', axis=1)
    y = dataCorrelacion['Recuperados']

    #Dividir en conjunto de train y test
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

    # Selección del Modelo Lineal
    modelo = LinearRegression()

    # Entrenar el modelo
    modelo.fit(X_train, y_train)

    # Variables de predicción
    y_predict_train = modelo.predict(X_train)
    y_predict = modelo.predict(X_test)
    return y_predict, y_test
```

Figure 17. Funcion de entrenamiento Regresion Lineal