

NOMBRE: ALEX BENAVIDEZ

CARRERA: ING EN SISTEMAS

MATERIA: SIMULACION

PROFESOR: DIEGO QUISI

- Diseñe y desarrolle un modelo y/o script que permita simular el siguiente caso real:
- Investigar los datos de los países contagiados por COVID-19, especialmente de Latinoamérica (menos Ecuador), deberán escoger uno y que no se repita, para ello se va a seleccionar el orden en el que publique dentro del foro "Tema prueba 1", con estos datos obtener los siguientes modelos:
 - Generar un modelo matemático de predicción para regresión lineal, exponencial, polinómico y logarítmico, de los nuevos contactos en la próxima semana (7 días después).
 - Generar un modelo probabilístico con los datos.
 - Finalmente, contrarrestar los modelos matemáticos y generar las siguientes conclusiones
 - Cual tiene una mejor predicción
 - Ventajas y desventajas de los modelos.
 - Cuál es el principal problema del modelo probabilístico
- El proceso de simulación desarrollado deberá considerar los siguientes aspectos:
 - Se debe establecer un modelo basado en modelos matemáticos y probabilísticos.
 - El programa deberá generar gráficas que indiquen la ecuación matemática y probabilística de tendencias
 - Deben calcularse las siguientes métricas:
 - Total de infectados dentro de 7 días (matemático y probabilístico).

In [81]:

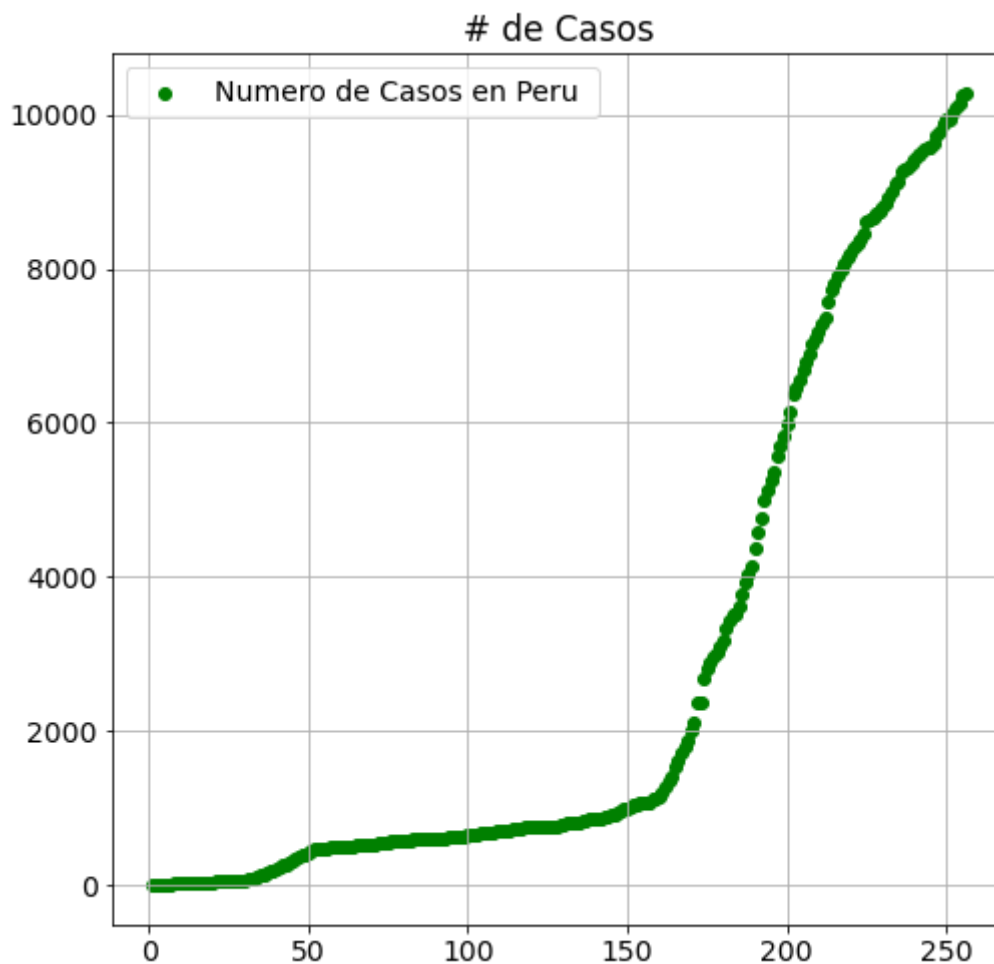


```
1 import pandas as pd
2 import numpy as np
3 from datetime import datetime, timedelta
4 from sklearn.metrics import mean_squared_error
5 from scipy.optimize import curve_fit
6 from scipy.optimize import fsolve
7 from sklearn import linear_model
8 import matplotlib.pyplot as plt
```

In [90]:

```
1 url = 'owid-covid-data.csv'
2 df = pd.read_csv(url).fillna(0)
3 df = df.loc[(df['location'] == 'Jamaica') & (df['total_cases'] != 0)]
4 df = df.loc[:, ['date', 'total_cases']]
5 x = np.arange(1, len(df)+1, 1)
6 print("Total de dias que llevamos es:" , len(x))
7 y = np.array(df.values[:,1], dtype='float')
8
9 plt.figure(figsize=(8, 8))
10 plt.scatter(x,y,label='Numero de Casos en Peru ', color='Green')
11 plt.grid(True)
12 plt.legend()
13 plt.title('# de Casos');
```

Total de dias que llevamos es: 256



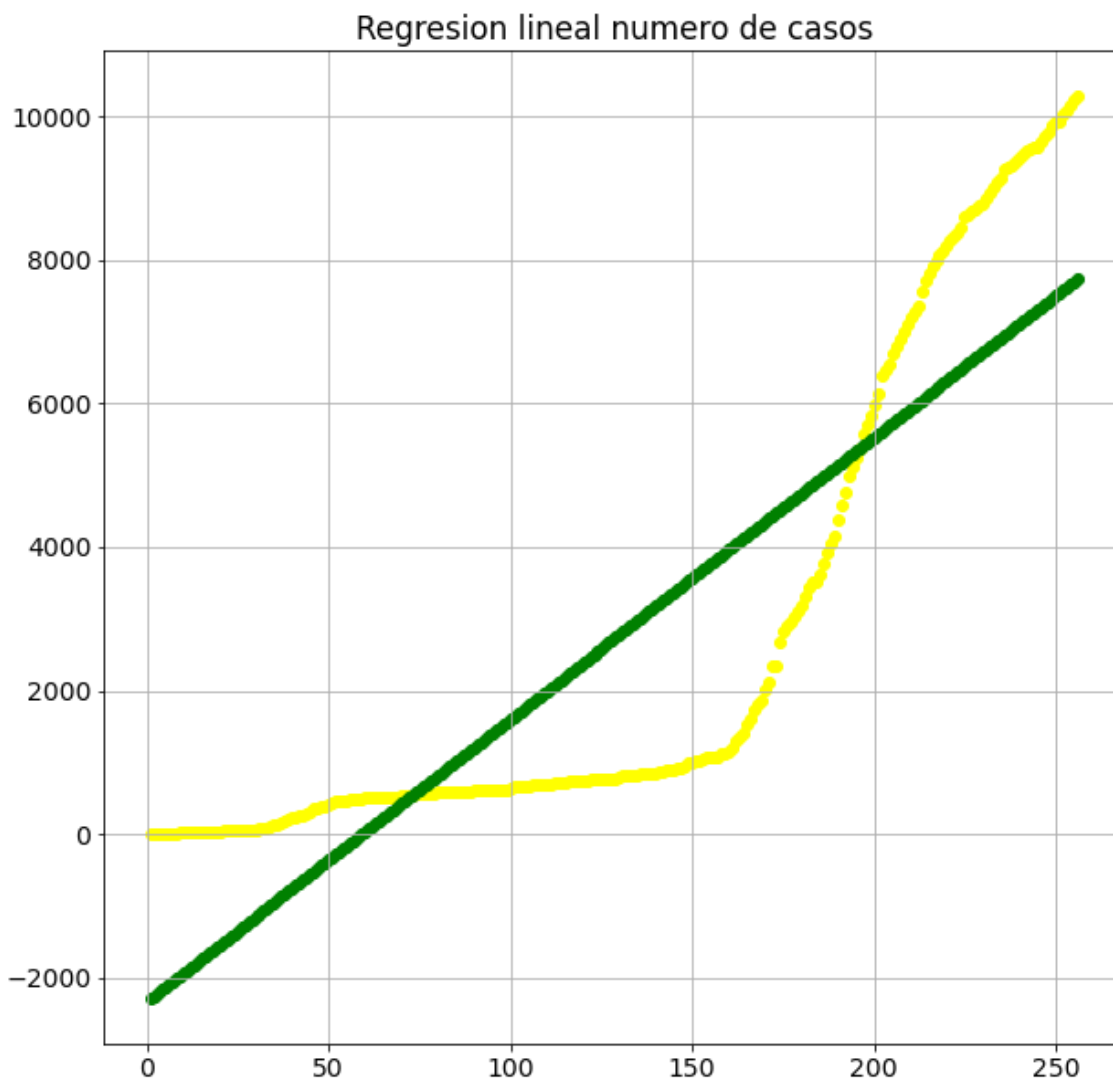
1. UTILIZACION DE REGRESION LINEAL PARA PREDICCION DE NUMERO DE CONTAGIADOS EN JAMAICA

In [91]:

```

1 x_li=x
2 y_li=y
3 dfl=df
4 modelo = linear_model.LinearRegression()
5 modelo.fit(x_li.reshape((-1,1)), y_li)
6 y_pred = modelo.predict(x_li.reshape((-1,1)))
7
8 plt.figure(figsize=(10, 10))
9 plt.scatter(x_li,y_li,color='yellow')
10 plt.title("Regresion lineal numero de casos")
11 plt.scatter(x_li,y_pred,color='green')
12 plt.grid(True)
13 plt.show()
14 #numero = input('Ingrese el dia que desee predecir cuantos contagiados habra de covid e
15 #nuevo_x = np.array([int(numero)])
16
17 if (modelo.intercept_ < 0):
18     ecua='y = {}x {}'
19 else:
20     ecua='y = {}x + {}'
21 print("La ecuacion es la siguiente: ", ecua.format(modelo.coef_[0],modelo.intercept_))
22
23
24 prediccion = modelo.predict([[len(x)+7]])
25 print("LA PREDICCION UTILIZANDO REGRESION LINEAL ES IGUAL: " , prediccion[0])

```



La ecuación es la siguiente: $y = 39.298220774204644x - 2326.3487132352966$
 LA PREDICCIÓN UTILIZANDO REGRESIÓN LINEAL ES IGUAL: 8009.083350380524

2. UTILIZACION DE REGRESION EXPONENCIAL PARA PREDICCIÓN DE NUMERO DE CONTAGIADOS EN JAMAICA

In [92]:

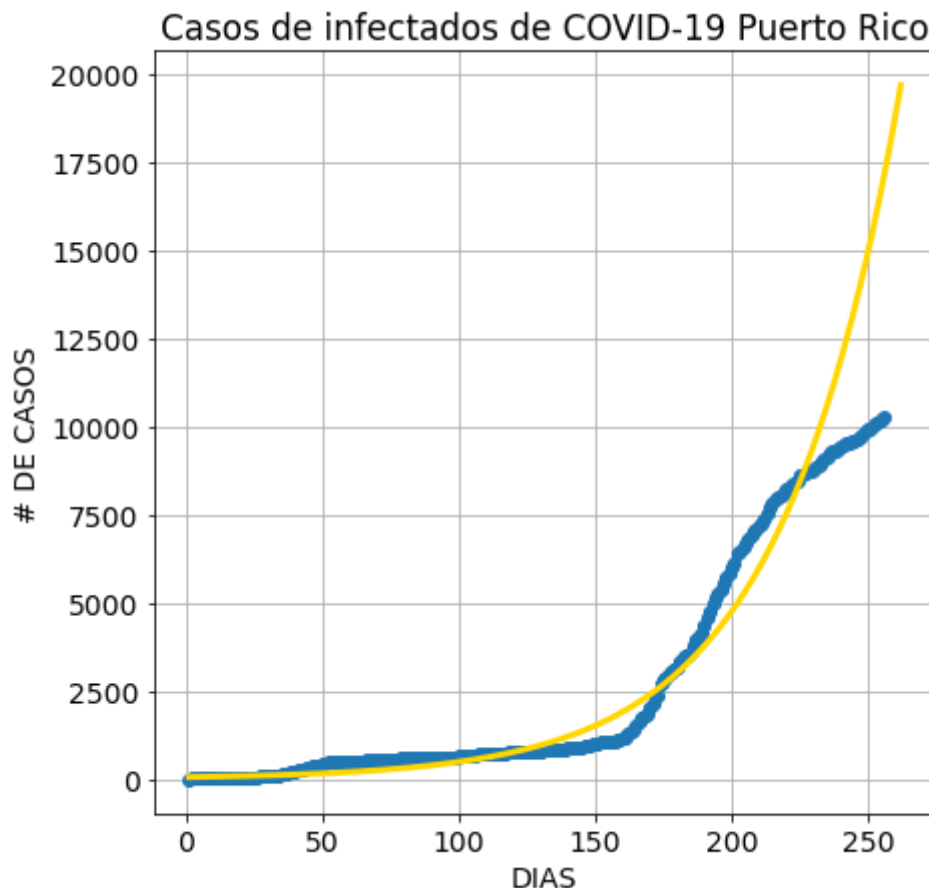
```

1
2
3 curve_fit=np.polyfit(xexp,np.log(yexp),1)
4 print(curve_fit)
5 pred_xe=np.array(list(range(min(xexp),max(xexp)+7)))
6 yx=np.exp(curve_fit[1])*np.exp(curve_fit[0]*pred_xe)
7 plt.title('Casos de infectados de COVID-19 Puerto Rico')
8 plt.plot(xexp,yexp,"o")
9 plt.plot(pred_xe,yx,color='gold',linewidth=3.0)
10 plt.xlabel('DIAS')
11 plt.ylabel('# DE CASOS')
12 plt.grid(True)
13
14 print("LA PREDICCIÓN CON EL MODELO EXPONENCIAL ES IGUAL A: ", yx[len(yx)-1])
15

```

[0.02292083 3.8820383]

LA PREDICCIÓN CON EL MODELO EXPONENCIAL ES IGUAL A: 19678.74780339439



3. UTILIZACION DE REGRESION POLINOMIAL PARA

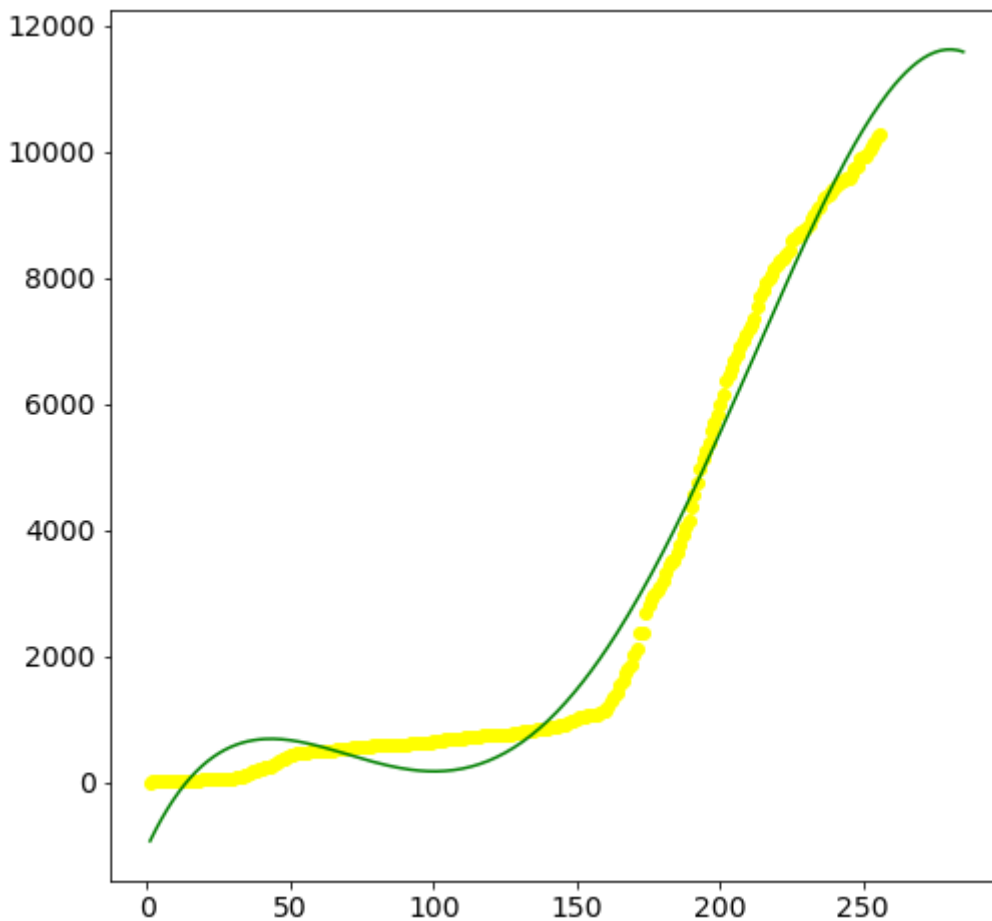
PREDICCION DE NUMERO DE CONTAGIADOS EN JAMAICA

In [93]:

```

1 x_pl=x
2 y_pl=y
3 df_pol=df
4
5 funcion = np.poly1d(np.polyfit(x_pl, y_pl, 4))
6
7 print(funcion)
8
9 plt.figure(figsize=(8, 8))
10 plt.scatter(x_pl, y_pl, color = "yellow")
11 x1=np.arange(1,len(df)+30,1, dtype='float')
12 plt.plot(x1, funcion(x1), color='green')
13 plt.show()
14
15 print(funcion)
16
17
18 print("LA PREDICCION UTILIZANDO REGRESION POLINOMIAL ES IGUAL: " , round(funcion(len(x)
```

$$-2e-05 x^4 + 0.0113 x^3 - 1.783 x^2 + 97.23 x - 1030$$



$$-2e-05 x^4 + 0.0113 x^3 - 1.783 x^2 + 97.23 x - 1030$$
 LA PREDICCION UTILIZANDO REGRESION POLINOMIAL ES IGUAL: 11161.78

4. UTILIZACION DE REGRESION LOGARITMICA PARA PREDICCION DE NUMERO DE CONTAGIADOS EN JAMAICA

In [94]:

```
1 from scipy.optimize import curve_fit
2 from sklearn.linear_model import LogisticRegression
3
4 x_log= x
5 y_log=y
6 df_log=df
```

In [95]:

```
1 def modelo_logistico(x,a,b):
2     return a+b*np.log(x)
3
4 exp_fit = curve_fit(modelo_logistico,x_log,y_log) #Extraemos los valores de los parámetros
5 print(exp_fit)
6
```

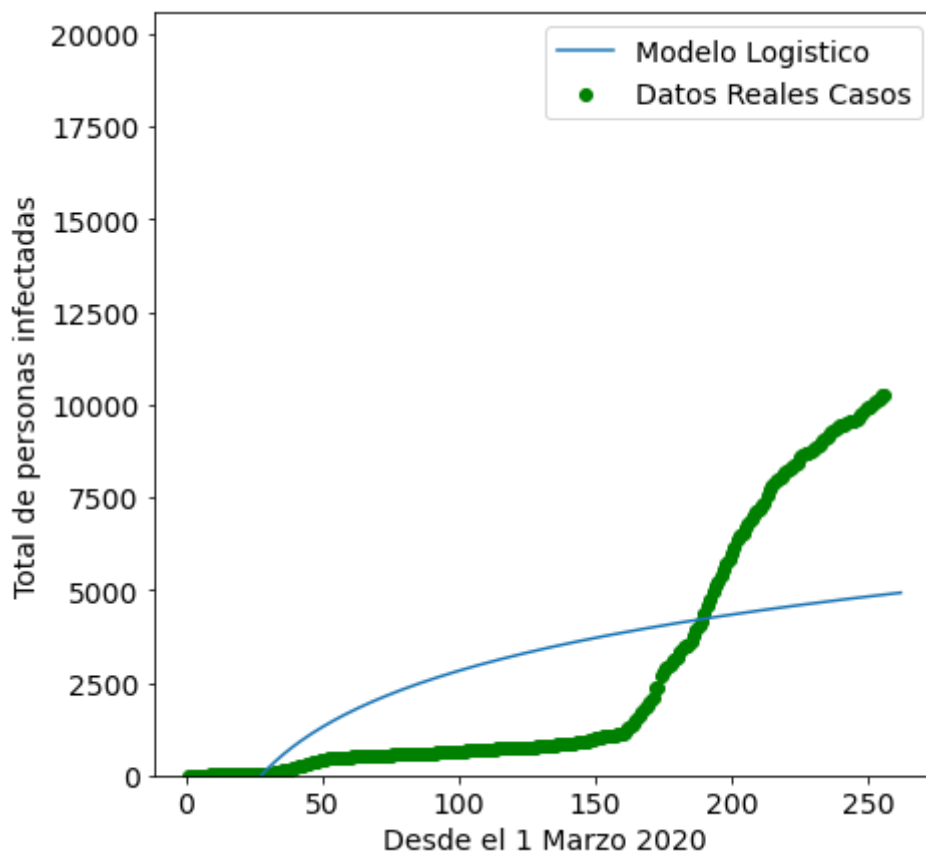
```
(array([-7232.39981968, 2183.49750359]), array([[ 634905.45182802, -133339.3793394 ],
        [-133339.3793394 , 29243.66541603]]))
```

In [98]:

```

1 pred_x = list(range(min(x_log),max(x_log)+7)) # Predecir 7 dias mas
2 plt.rcParams['figure.figsize'] = [7, 7]
3 plt.rc('font', size=14)
4 # Real data
5 plt.scatter(x_log,y_log,label="Datos Reales Casos",color="green")
6 # Predicted exponential curve
7 valores = [modelo_logistico(i,exp_fit[0][0],exp_fit[0][1]) for i in pred_x]
8 plt.plot(pred_x, [modelo_logistico(i,exp_fit[0][0],exp_fit[0][1]) for i in pred_x], label="Modelo Logistico")
9 plt.legend()
10 plt.xlabel("Desde el 1 Marzo 2020")
11 plt.ylabel("Total de personas infectadas")
12 plt.ylim(0,max(y)*2) # Definir Los Limites de Y
13 plt.show()
14
15 print("PREDICCION UTILIZANDO REGRESION LOGARITMICA ES : ", valores[len(pred_x)-1])

```



PREDICCION UTILIZANDO REGRESION LOGARITMICA ES : 4926.066503413828

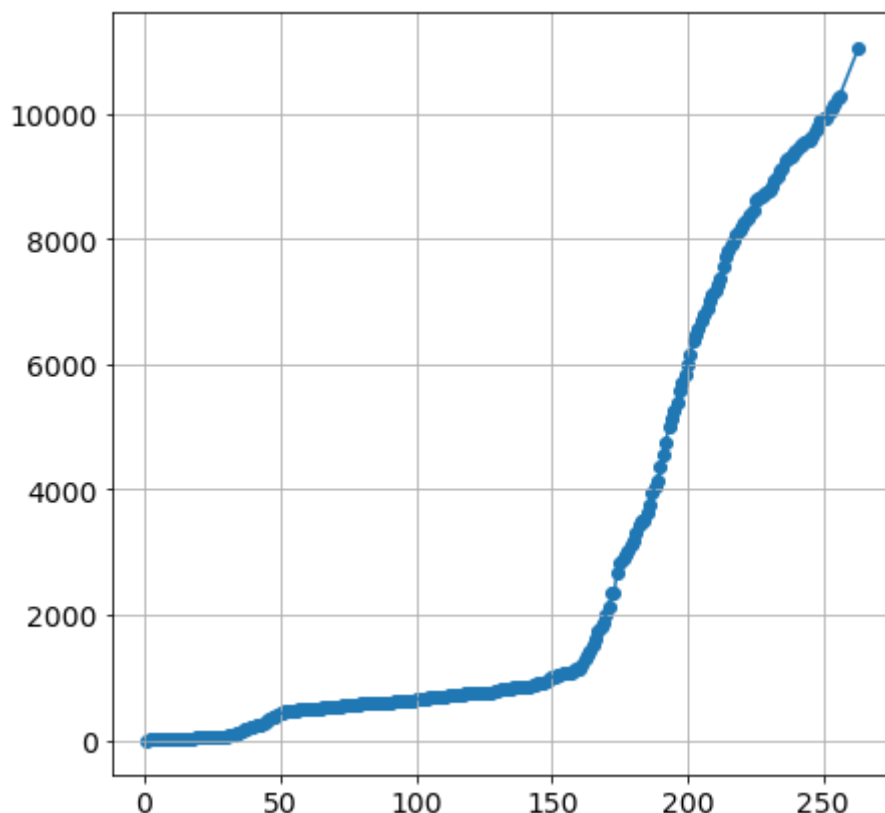
5. UTILIZACION DE REGRESION PROBABILISTICO PARA PREDICCION DE NUMERO DE CONTAGIADOS EN JAMAICA

In [100]:

```
1 x_pro= x
2 y_pro=y
3 df_pro=df
4
5 media = df_pro.values[:,1].mean()
6 mediana = np.median(df_pro.values[:,1])
7 print("MEDIA: ",media)
8 print("MEDIANA: ",mediana)
9
10 for it in range(int(x_pro[-1]), int(x_pro[-1]+7)):
11     x_pro=np.append(x,(it+1))
12     y_pro=np.append(y,y[-1] + mediana)
13
14 prediccion_siguiente = int(y_pro[-1]+mediana)
15 plt.plot(x_pro,y_pro)
16 plt.scatter(x_pro,y_pro)
17 plt.grid(True)
18 plt.show()
19 print("LA PREDICCION UTILIZANDO REGRESION PROBABILISTICA ES IGUAL:", prediccion_siguiente)
20
```

MEDIA: 2723.47265625

MEDIANA: 771.0



LA PREDICCION UTILIZANDO REGRESION PROBABILISTICA ES IGUAL: 11826

In [99]:

```
1 print("LA PREDICCION UTILIZANDO REGRESION LINEAL ES IGUAL: " , prediccion[0])
2 print("LA PREDICCION CON EL MODELO EXPONENCIAL ES IGUAL A: ", yx[len(yx)-1])
3 print("LA PREDICCION UTILIZANDO REGRESION POLINOMIAL ES IGUAL: " , round(funcion(len(x)
4 print("LA PREDICCION UTILIZANDO REGRESION LOGARITMICA ES : ", valores[len(pred_x)-1])
5 print("LA PREDICCION UTILIZANDO REGRESION PROBABILISTICA ES IGUAL:", prediccion_siguier
```

LA PREDICCION UTILIZANDO REGRESION LINEAL ES IGUAL: 8009.083350380524
 LA PREDICCION CON EL MODELO EXPONENCIAL ES IGUAL A: 19678.74780339439
 LA PREDICCION UTILIZANDO REGRESION POLINOMIAL ES IGUAL: 11161.78
 LA PREDICCION UTILIZANDO REGRESION LOGARITMICA ES : 4926.066503413828
 LA PREDICCION UTILIZANDO REGRESION PROBABILISTICA ES IGUAL: 11826

Cual tiene una mejor prediccion

- El modelo polinomial tiene una mejor prediccion, se asemeja mucho a la realidad

Modelo Lineal

- Ventajas

1. Es un modelo facil de entener e implementar, esto puede ser veneficioso para la implementacion en diferentes ambitos laborales.
2. Es facil de implementar, es muy util cuando la realcion a modelar no es demasiado compleja.

- Desventajas

1. No es posible modelar relaciones demasiado complejas
2. No es posible obtener realciones no lineales sin transformar primeramente la entrada.

Modelo Logarítmico

-Ventajas

1. Es facil su implementacion y es facil entenderlo
2. Su entrenamiento es rapido

-Desventajas

1. A veces es es muy simppe para poder captar relaciones complejas entre variables

Modelo Exponencial

-Ventajas

1. Nos da imformacion adecuada ya que estan diseñadas para valores y crecimientos rapidos.

-Desventajas

1. No disminuye en el futuro, sino sigue aumentando
2. Matematicamente es difícil su desarrollo, también en algunos casos hace complejo todo el proceso matemático para poder obtener resultados

Modelo Polinomial

-Ventajas

1. Funciona con cualquier muestra
2. Permite trabajar con datos no lineales

-Desventajas

1. En este modelo se requiere elegir el grado correcto del polinomio para una buena relación sesgo/varianza

Cual es el principal problema del modelo probabilístico

- El problema que tiene el modelo probabilístico es que con valores pequeños trabaja de manera correcta, en cambio cuando se trabaja con valores grandes o que existe una variación amplia entre cada uno, este modelo toma un error muy grande, lo que hace que la predicción no sea la correcta.