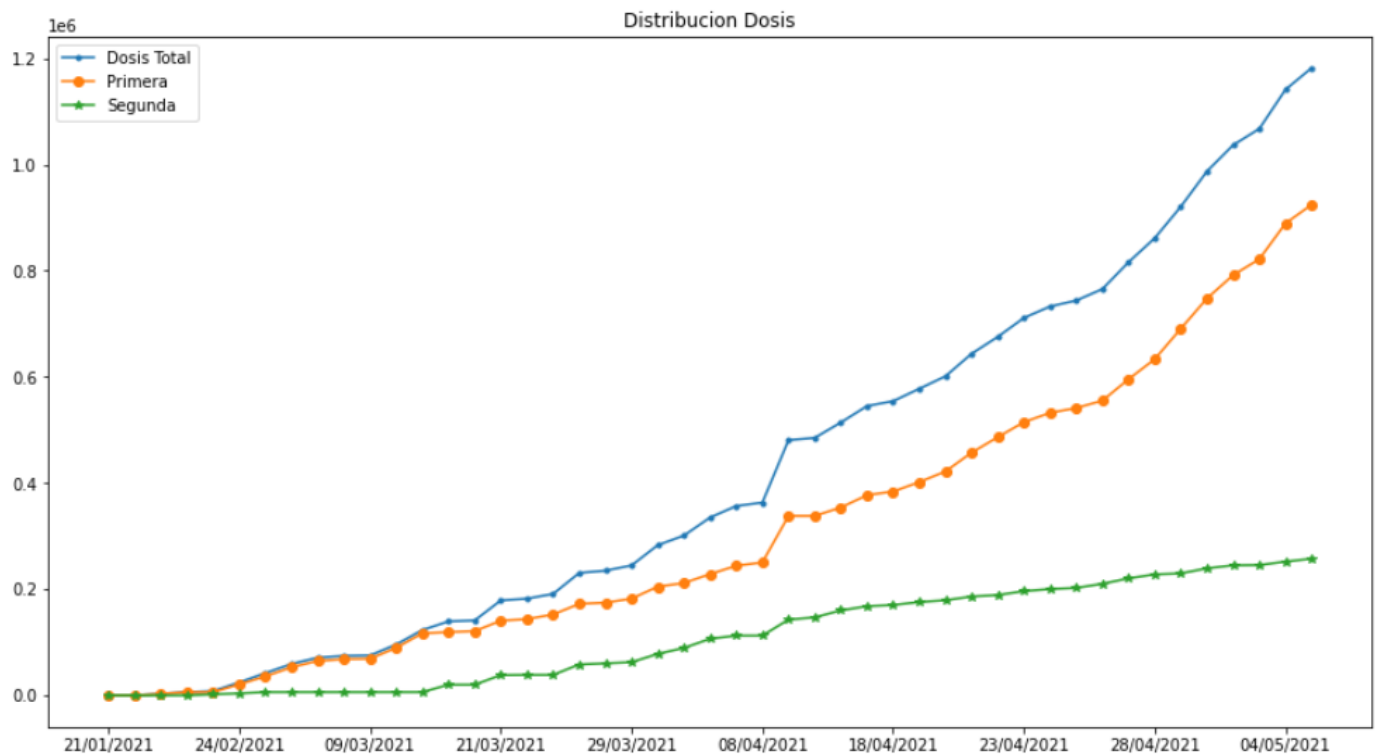


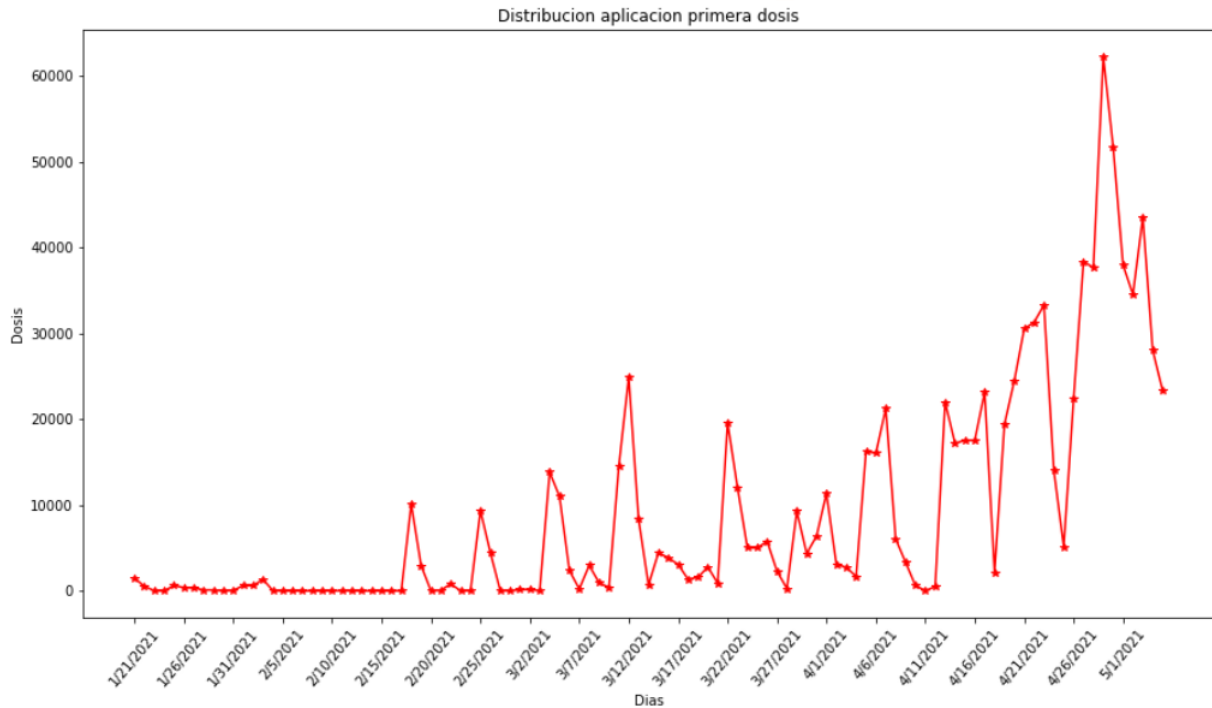
Universidad Politécnica Salesiana

Estudiante: Javier Vazquez

1. Generar graficas para entender y procesar los datos:
 - Generar gráficas y reportes del total de personas vacunadas.

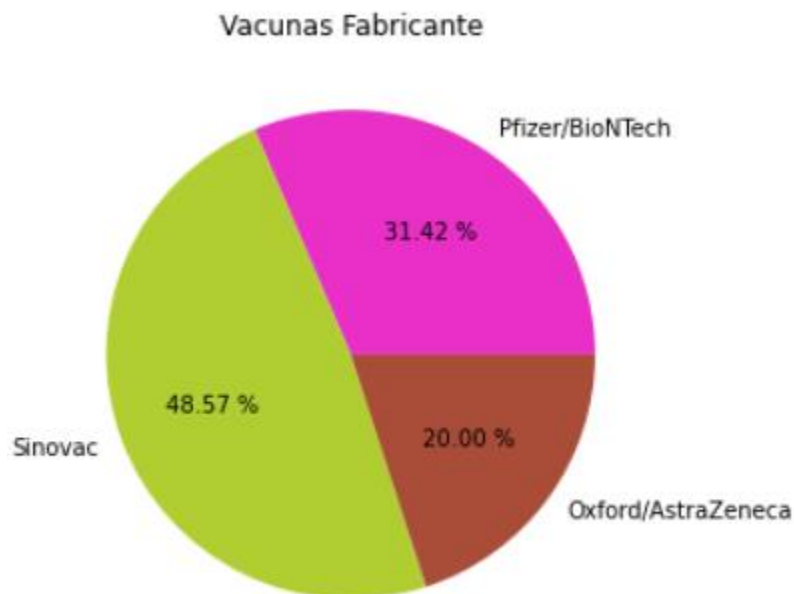


Existe un total casi de 1200000 dosis que han llegado al Ecuador, de las cuales claramente se puede observar como la mayor parte se han utilizado como primera dosis para la población, mientras la segunda dosis existe una muy pequeña población que cuenta con esta, no mas de 250k.



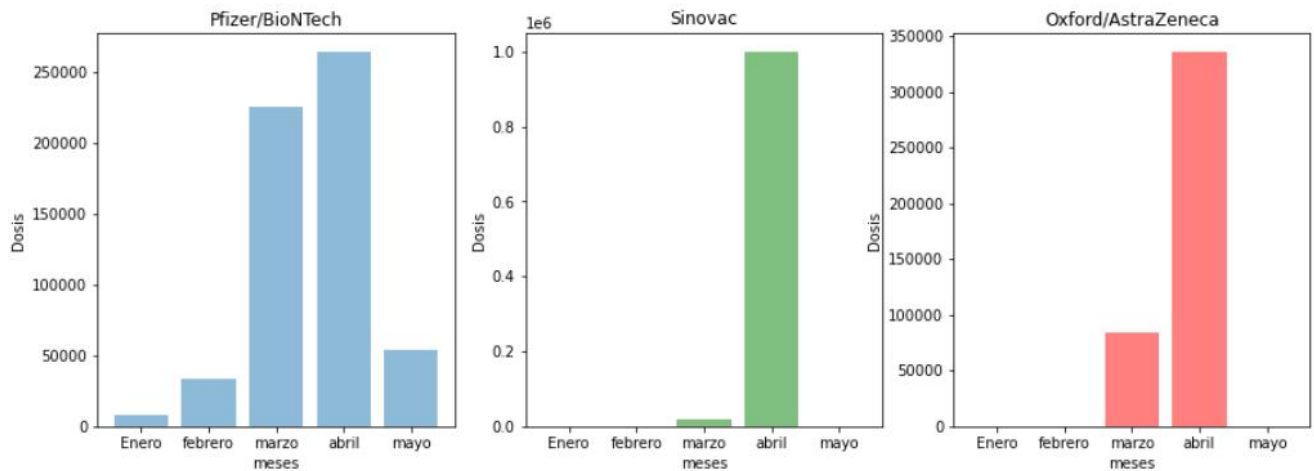
En el grafico podemos observar cual a ido siendo las dosis de vacunas aplicadas a las personas en un solo día, se aprecia claramente como a ido variando este valor inclusive en algunos días en los cuales no se realizo la vacunación.

- Generar grafico de pie por fabricante de la vacuna.



Observamos que el fabricante Sinovac es el que mayor dosis a suministrado al Ecuador

- **Generar histogramas de vacunas por mes de llega y fabricante.**



Analizando el grafico claramente el fabricante Pfizer/BioNTech es el que a estado no en mucha cantidad suministrando vacunas a Ecuador desde enero, ya desde el mes de marzo comenzaron nuevos fabricantes a llegar al Ecuador como es el caso de Sinovac que únicamente en el mes de abril suministro cerca de un millón de vacunas, mucho más que el fabricante Pfizer que ha venido desde meses anteriores. El fabricante Oxford también empezó desde marzo con entregas de vacunas e igual en abril tuvo un gran suministro de dosis para el Ecuador.

2. Generar un reporte parametrizado que pueda ingresar los datos de las fechas inicio y fin para obtener la información de las gráficas vistas en el primer punto.

Utilizando la librería de Papermill y mediante etiquetas dentro del cuaderno de Jupyter podemos parametrizar las variables que necesitamos, en este caso serian las fechas de inicio y fin que quisiéramos analizar

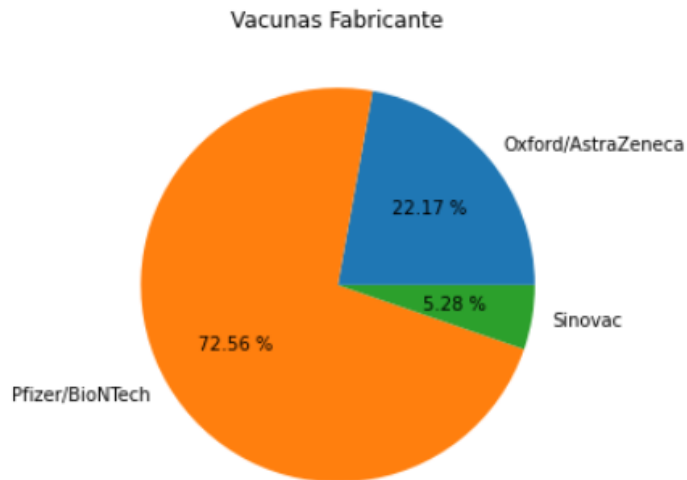
```
In [1]: parameters ✕
# Parameters
fecha_inicio=''
fecha_fin=''

In [2]: injected-parameters ✕
# Parameters
fecha_inicio = "2021-01-20"
fecha_fin = "2021-03-30"
```

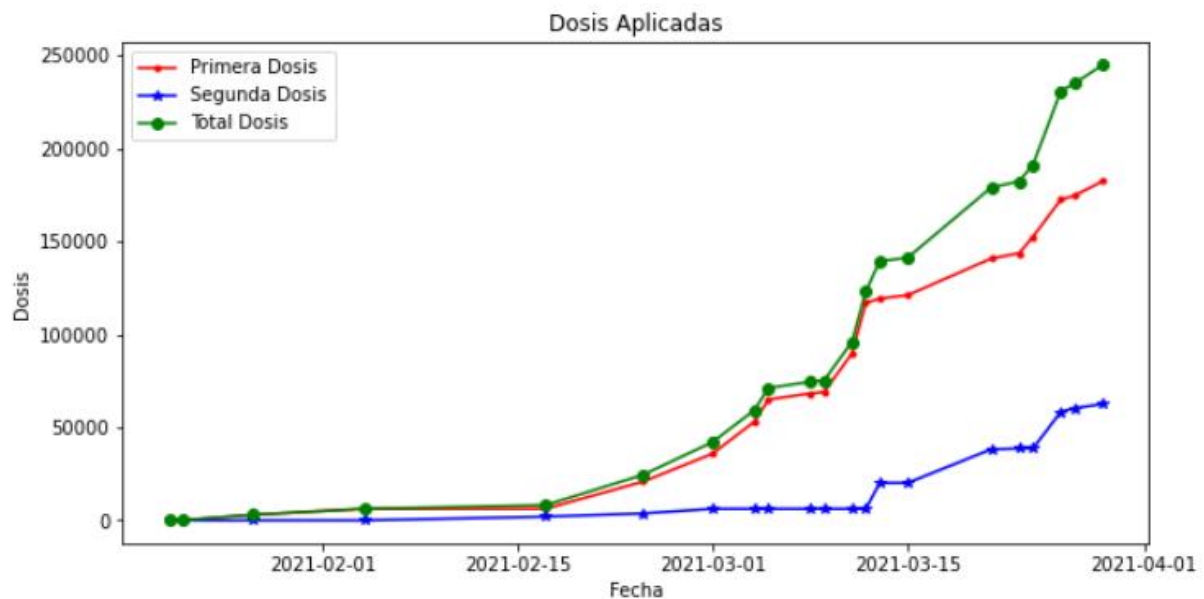
Fecha Inicio: 2021-01-20

Fecha Fin: 2021-03-30

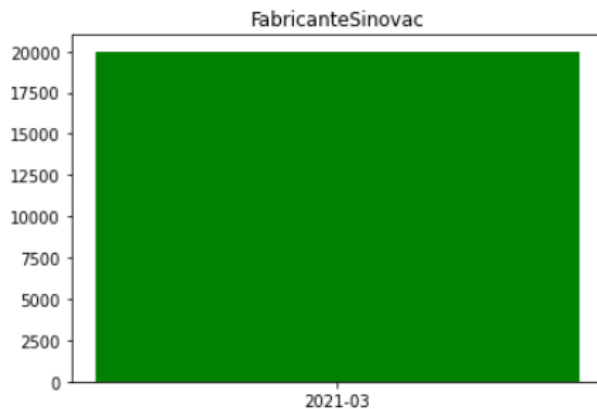
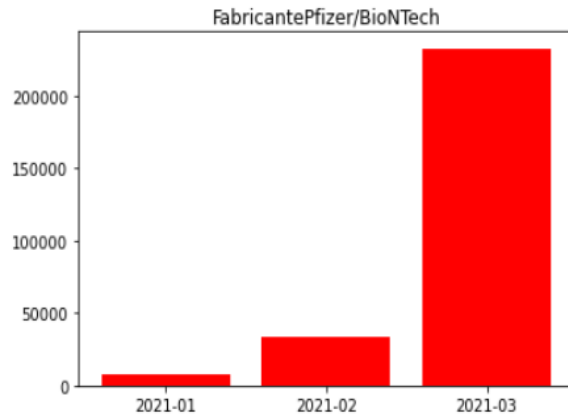
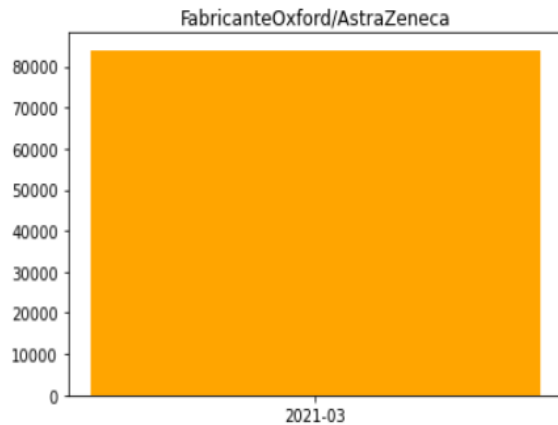
Fabricantes de vacunas que suministraron al Ecuador en el rango de fechas propuesto



Analizando el grafico podemos darnos cuenta que hasta inicios de marzo todas las dosis que ha recibido Ecuador se han aplicado como primera dosis a las personas, cerca del 10 marzo es cuando se comenzó aplicar en mayor cantidad la segunda dosis.

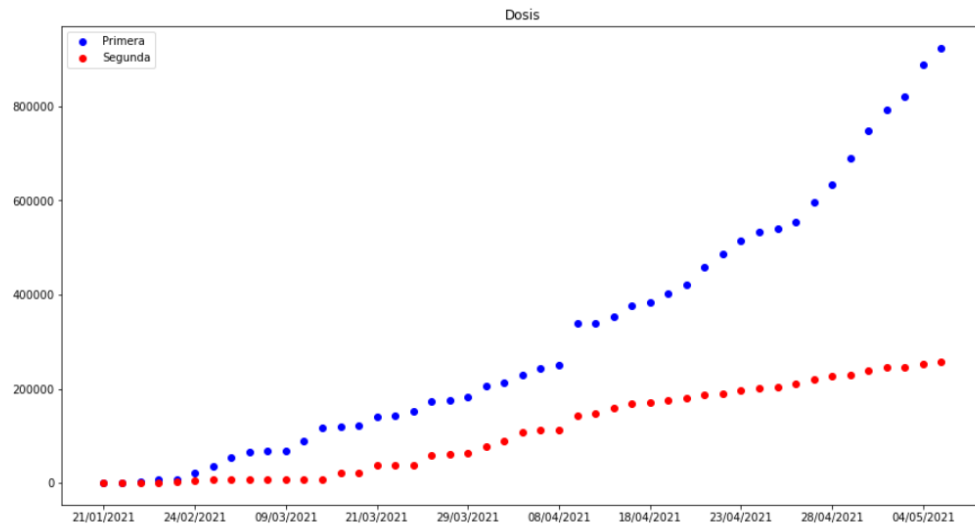


En los diferentes histogramas de los fabricantes que han llegado a Ecuador, claramente Pfizer es el que mas dosis suministro en este rango de fechas, seguido por Oxford.

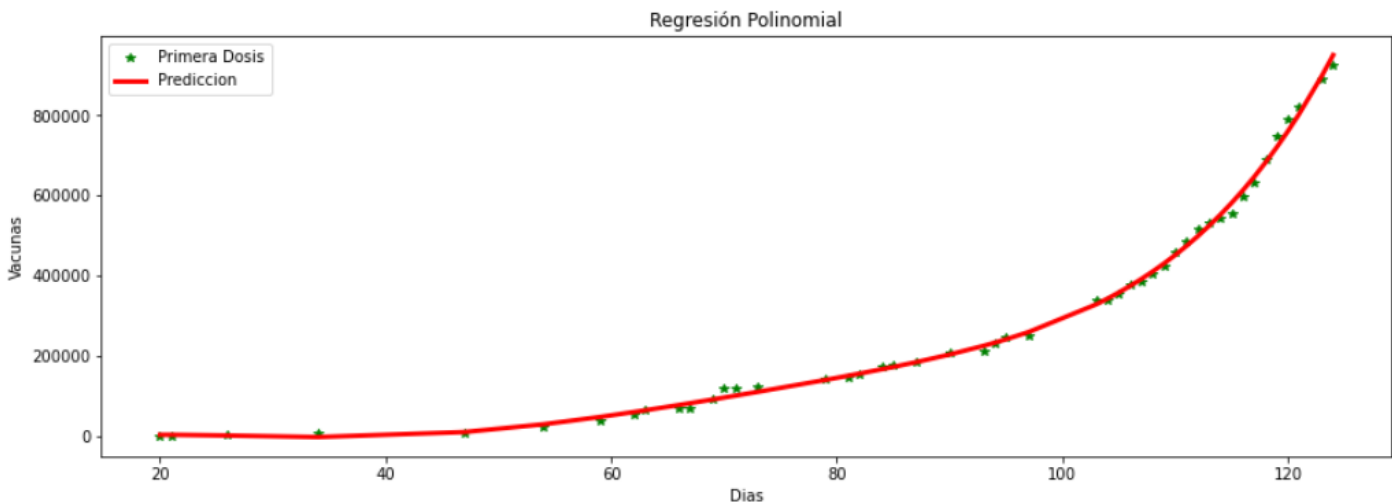


3. Generar un modelo matemático de predicción basado en regresión, del procesos de vacunación en base al número actual de vacunados (1 y 2 dosis) y a la llegada de nuevas vacunas.

Como observamos en el siguiente grafico de puntos, notamos como una regresión lineal no puede ser aplicada para predecir la primera dosis, necesitaríamos un modelo polinomial que se ajustaría de mejor manera, mientras que para la segunda dosis parece que un modelo lineal seria el correcto.



En base a la primera dosis se ajustó una regresión polinomial de grado 5 que abarca la mayoría de puntos y no deja muchos datos residuos

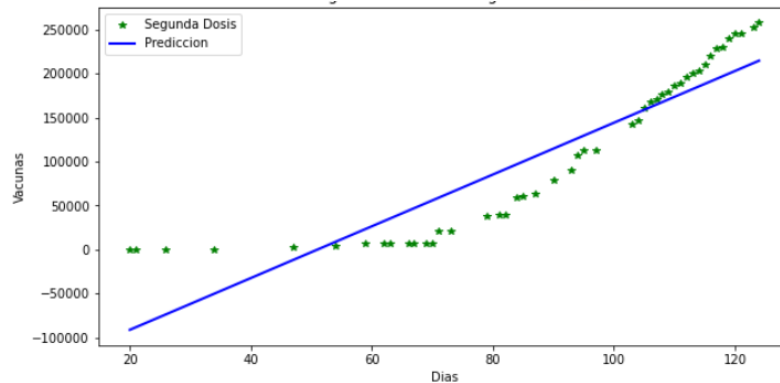


Fecha tentativa en la cual serán vacunados con la primera dosis toda la población de Ecuador será el 15 de septiembre.

```
In [55]: fecha_fin='15/09/2021'
fecha_fin=(datetime.strptime(fecha_fin, '%d/%m/%Y') - datetime.strptime("01/01/2021", '%d/%m/%Y')).days
pol_pred=pol_reg.predict(poly_reg.fit_transform([[fecha_fin]]))
pol_pred

Out[55]: array([[1.8609199e+08]])
```

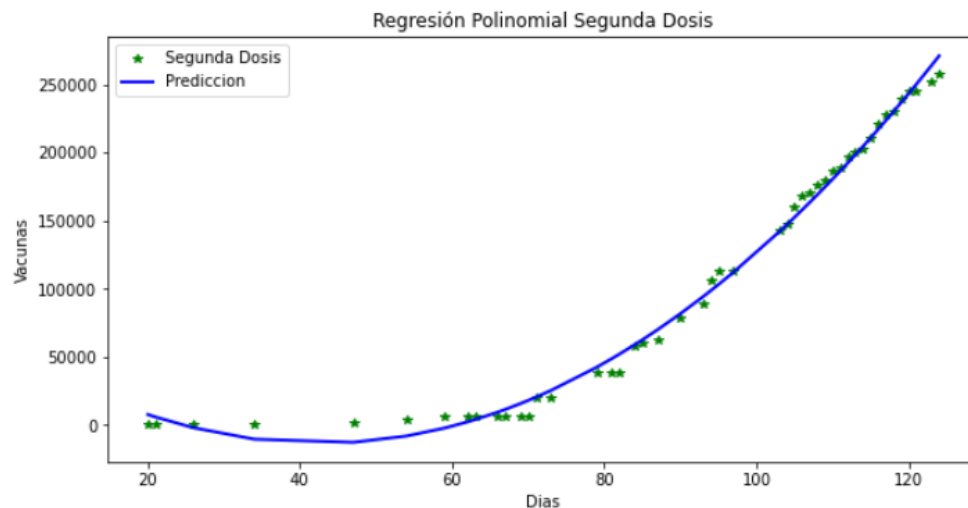
Como se mostro al inicio de este punto como estaban distribuidos los datos de la segunda dosis, a simple viste daba por entendido aplicar una regresión lineal, pero al crear este modelo claramente no es la mejor opción realizar esto, ya que deja muchos datos sueltos, e inclusive predice hasta valores negativos.



```
In [154]: print("Score : ",pol_reg_segunda.score(X_poly,y_segunda))
```

Score : 0.8292522681365518

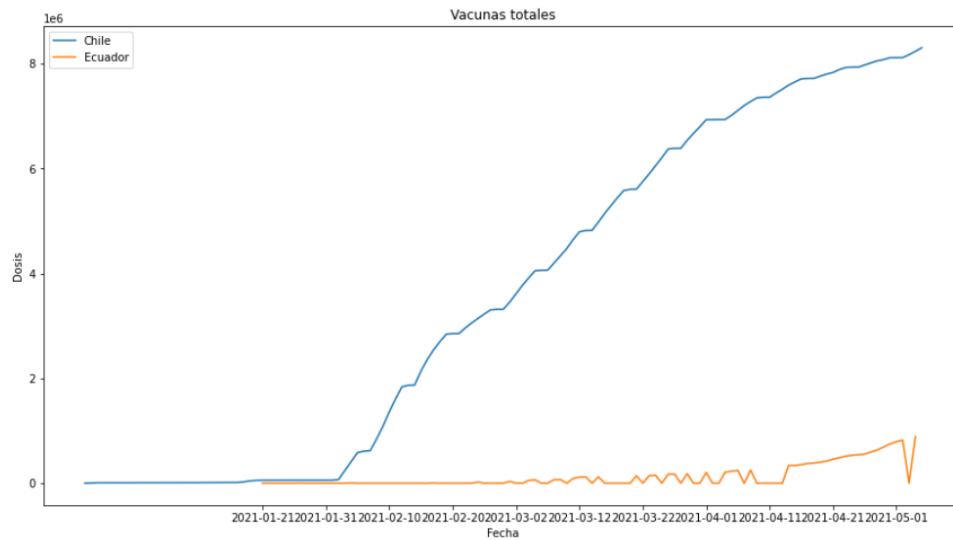
Utilizando un regresión polinomial de grado 3 fue la que mas se ajusto a los datos, se probó también con un polinomio de grado 4 pero después de los 175 días haber empezado la vacunación, el modelo empezó a predecir valores negativos, por lo que se volvió analizar el problema y la reducción del grado del polinomio.



4. **Desarrollar y generar un proceso de comparación con al menos dos países (1. Latinoamérica, 1. Europa).**

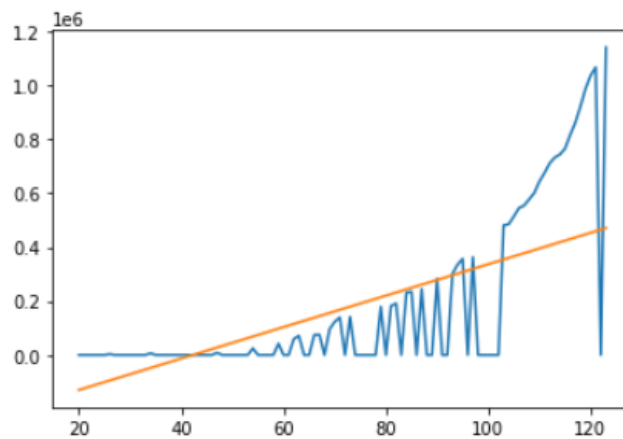
ECUADOR VS CHILE

La primera comparación con un país Latinoamericano fue con Chile que cuenta con un población similar a la de Ecuador

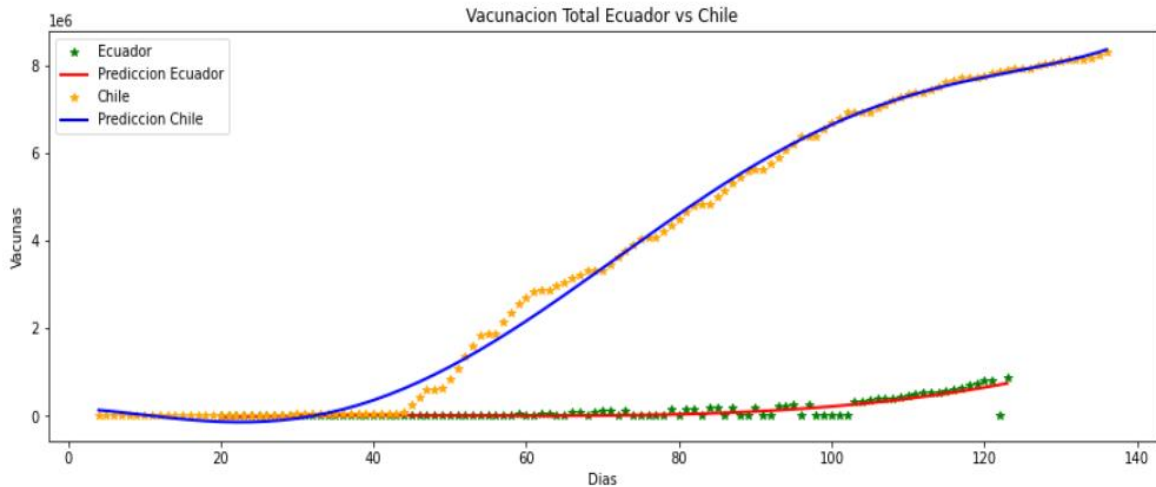


Donde observamos claramente como Chile cuenta con un alto suministro de vacunas a su población.

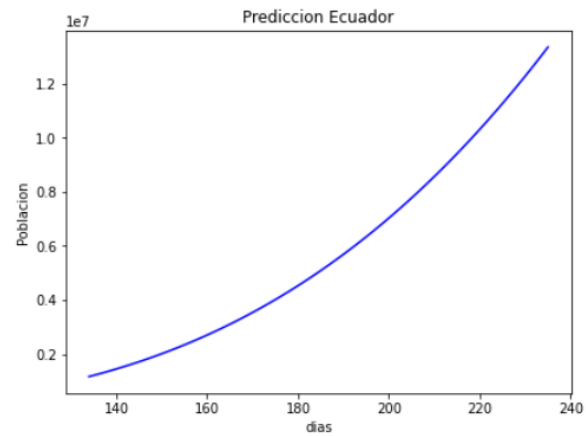
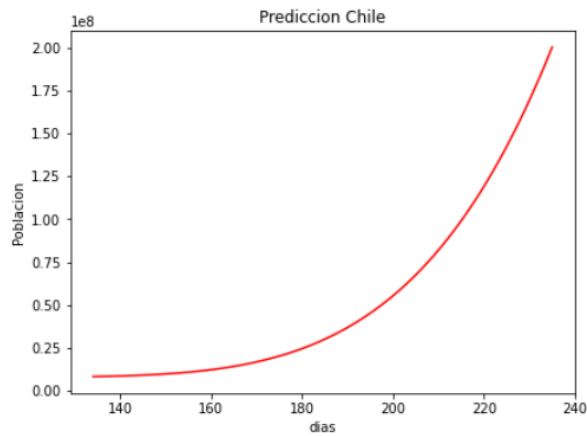
Claramente observamos como una regresión lineal seria la adecuada para los datos de Ecuador, pero al realizar este modelo lineal no abarca la mayoría de los datos, lo que nos da a entender que una regresión polinomial seria la que se ajuste de mejor manera.



Personas vacunas en los respectivos países.

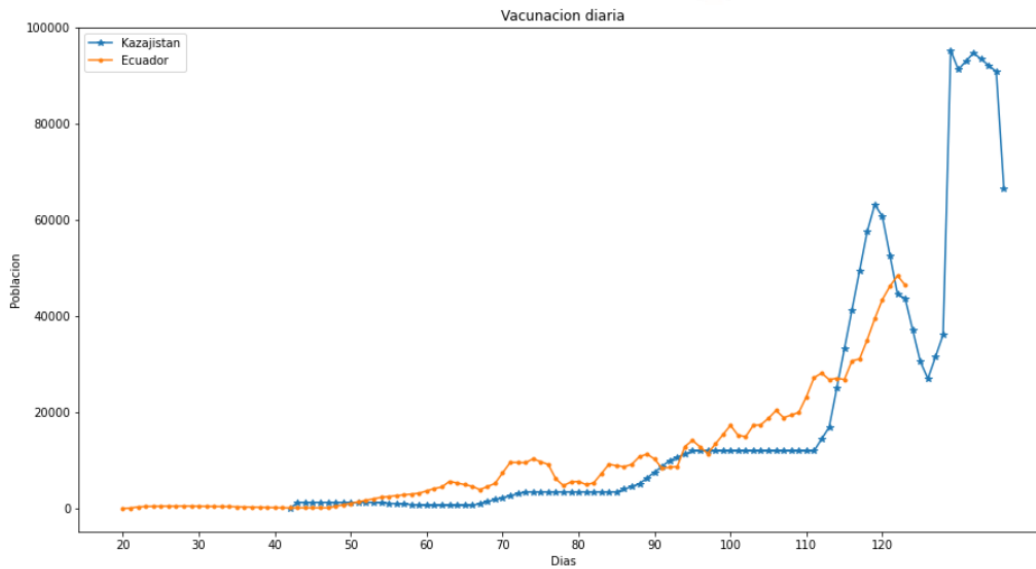


Para la predicción de los dos países, se hasta el 25 de agosto del 2021 donde claramente Chile ya contaría con toda su población vacunada mientras Ecuador estaría recién sobre los 12 millones de personas vacunadas.

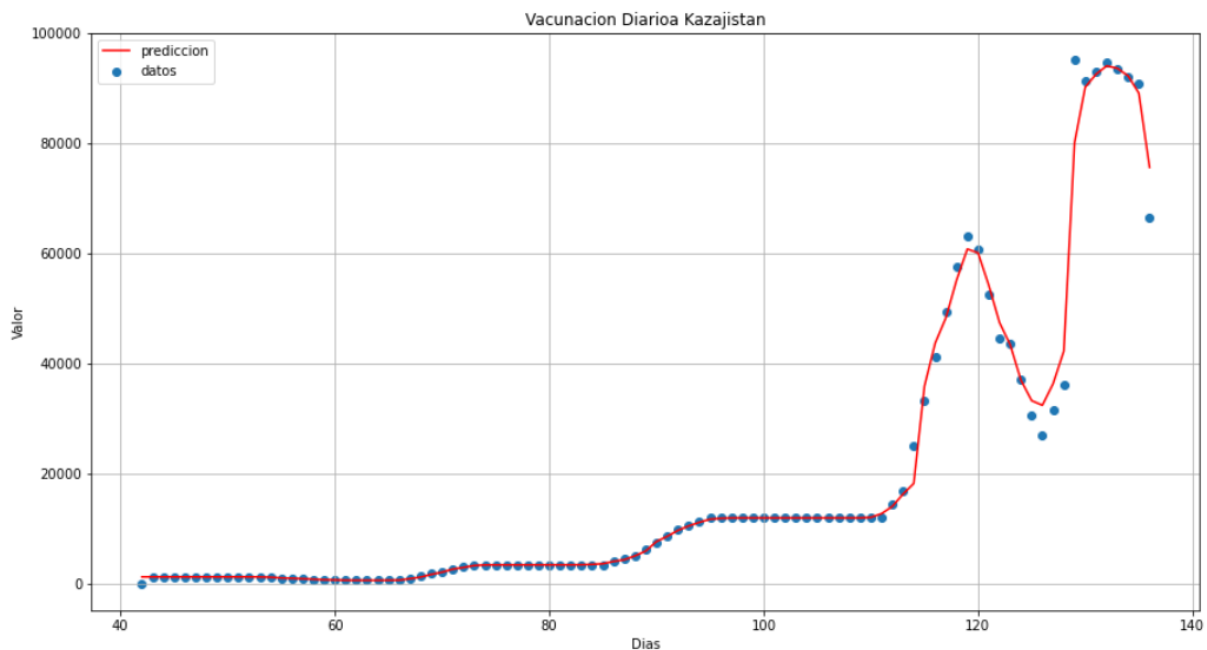


ECUADOR VS KAZAJISTAN

Para esta comparación al existir diferencias entre las poblaciones de ambos países, lo que se tomo en cuenta es el plan de vacunación diaria de cada país y poder predecir cuando podrá ser vacunada toda sus poblaciones



Analizando los datos de Kazajistan, una regresión polinomial no importa el grado que sea no se ajusta a los datos ya que no se encuentran bastantes dispersos, una regresión lineal tampoco seria la adecuada, aquí la mejor opción seria aplicar una regresión Random Forest que claramente observamos en el siguiente grafico se nota como abarca la mayor cantidad de datos.



Contando con una precisión muy cercana a los datos reales.

```

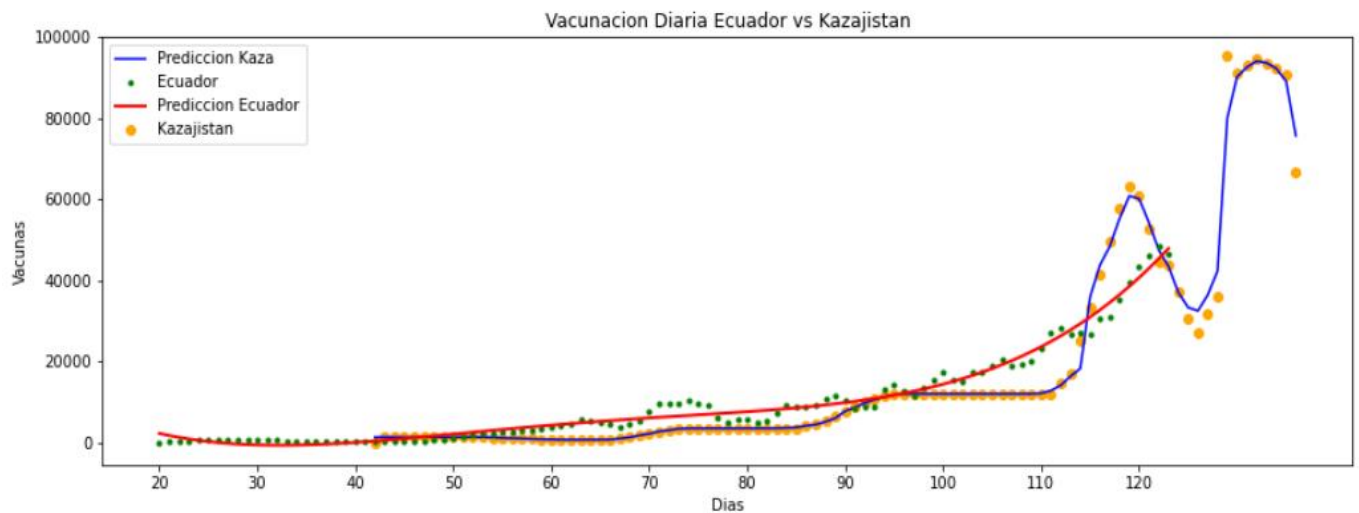
|:
print("Precision del modelo, valores Train = ",r2_score(y_train_k, pred_train_rf))

pred_test_rf = model_rf.predict(X_test_k)
#print(np.sqrt(mean_squared_error(y_test_k,pred_test_rf)))
print("Precision del modelo, valores test = ",r2_score(y_test_k, pred_test_rf))

Precision del modelo, valores Train = 0.9931855013257307
Precision del modelo, valores test = 0.971517109082715

```

Comparación de los modelos de regresión utilizados.



La fecha tentativa en la cual Kazajistan vacunaría a toda su población es para el 20 de Diciembre del 2020

Resultados

```

In [152]: fecha_fin='10/01/2022'
          fecha_fin=(datetime.strptime(fecha_fin, '%d/%m/%Y') - datetime.strptime("01/01/2021", '%d/%m/%Y')).days
          dia_kaza=np.arange(40,fecha_fin, dtype=int)
          pred_total_kaza=model_rf.predict(dia_kaza.reshape(-1,1))
          pred_total_kaza.sum()

Out[152]: 19665224.199999996

In [153]: fecha_fin_ecu='10/12/2021'
          fecha_fin_ecu=(datetime.strptime(fecha_fin_ecu, '%d/%m/%Y') - datetime.strptime("01/01/2021", '%d/%m/%Y')).days
          pol_ecu_diaria.predict(poly_ecu_diaria.fit_transform([[fecha_fin_ecu]]))

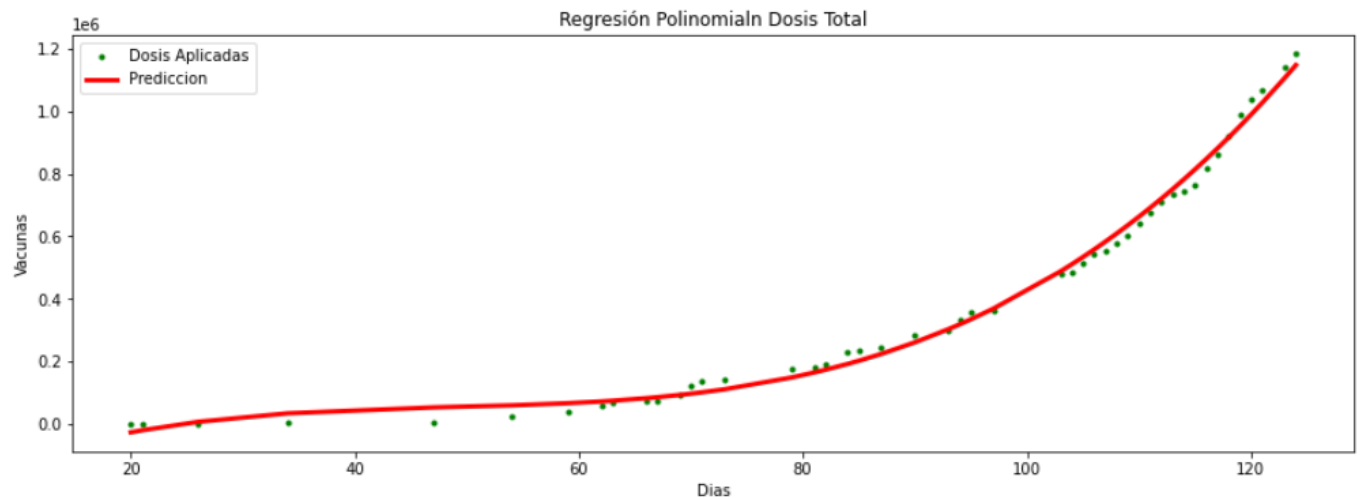
Out[153]: array([[18735510.69655354]])

```

In []:

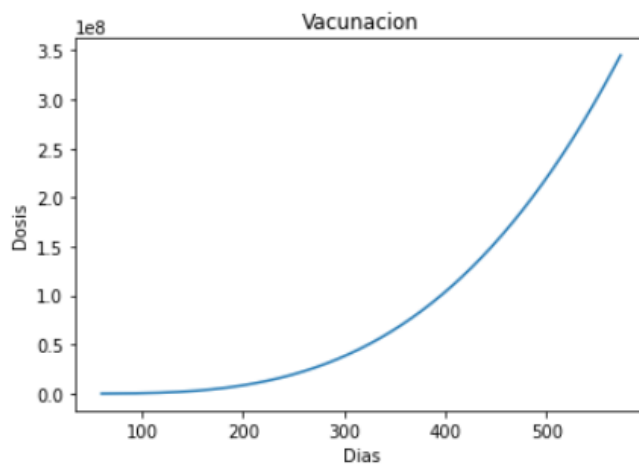
5. Identificar cual es la fecha tentativa en la que todos los ecuatorianos podrán ser

Para este análisis se realizó con la aplicación de dosis total, mediante una regresión polinomial de grado 3.



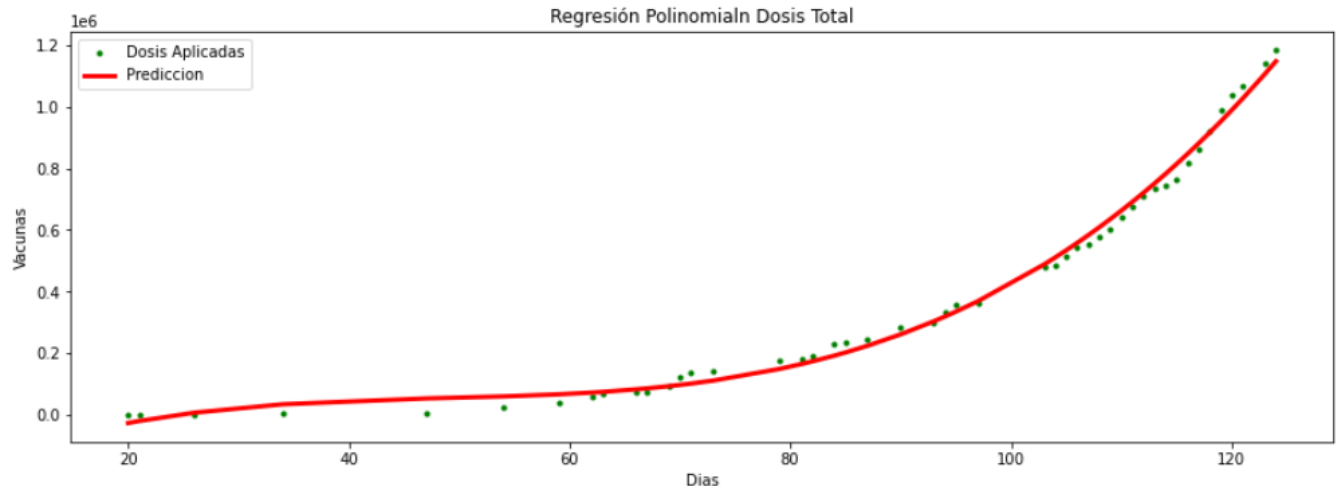
Y la fecha tentativa en la que todos los ecuatorianos serán vacunados con ambas dosis será el 29 de Julio del 2022, esto en base al total de dosis aplicadas, por lo que por una dosis tendría que aplicarse un poco más de 17 millones y esto multiplicados por 2 que sería la segunda dosis nos daría un valor por encima de los 34 millones.

```
] nueva_fecha='29/07/2022'
fec = (datetime.strptime(nueva_fecha, '%d/%m/%Y') - datetime.strptime("01/01/2021", '%d/%m/%Y')).days
prediccion_fecha= pol_reg.predict( poly_reg.fit_transform([[fec]]))
prediccion_fecha
[3.47025744e+08]
```

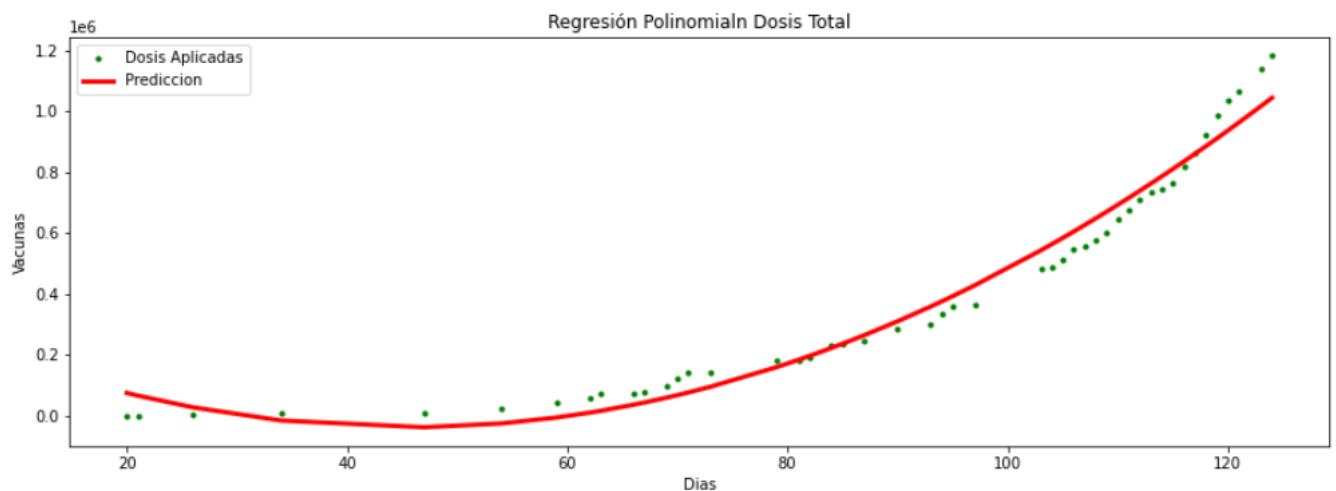


6. vacunados con las dos dosis.

Para poder realizar este cálculo, se realizó el análisis de la cantidad de dosis totales aplicadas en Ecuador.



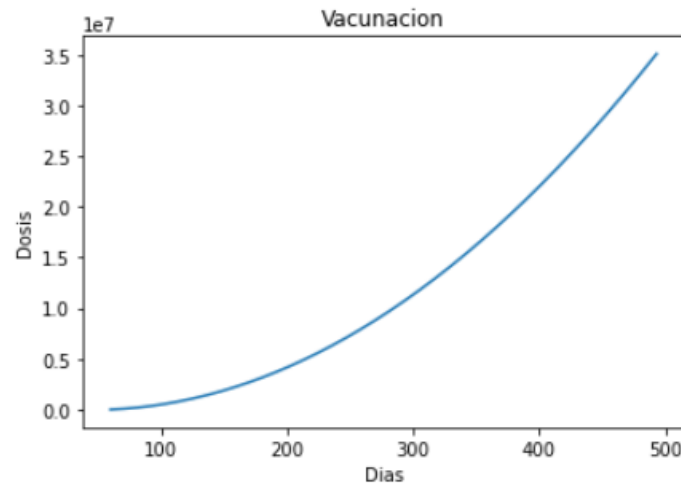
Claramente se utilizó una regresión polinomial de grado 3 que se ajusta de mejor manera a los datos, se probó también con una polinomial de grado 2 pero como observamos en el siguiente gráfico, existen muchos datos que no cubre el modelo de regresión.



La fecha tentativa en la que todos los ecuatorianos podrán ser vacunados con las 2 dosis es para el 10 de mayo del 2022

```
In [81]: nueva_fecha='10/05/2022'
fec = (datetime.strptime(nueva_fecha, '%d/%m/%Y') - datetime.strptime("01/01/2021", '%d/%m/%Y')).days
prediccion_fecha= pol_reg.predict( poly_reg.fit_transform([[fec]]))
prediccion_fecha

Out[81]: array([[35238751.18242434]])
```



Ventajas y Desventajas de los modelos

El modelo polinomial se ajusta de a los datos cuando estos tienden a incrementar cada vez mas y más, formando una curvatura sensible a valores atípicos.

El modelo de regresión Random Forest se ajusta de mejor manera ante datos que están en constante cambio, como en el caso que se utilizó para la vacunación diaria, donde claramente hay días que se suministran mas vacunas que otros días.

Una de las principales desventajas del modelo lineal se vuelve creciente o decreciente conforme al tiempo y es muy sensible a valores atípicos.

Opinión

Resulta bastante interesante crear este tipo de modelos que se ajustan bastante a la realidad por la que esta atravesando tanto el país como el mundo entero. Pudiendo así evidenciar como nuestro país esta enfrentando esta situación y cual es su postura frente a otros países, evidenciando claramente como Ecuador cuenta con una mala gestión de vacunación frente a otros países tanto de Latinoamérica como Europeos.

Conclusión

La cantidad de los datos influye en gran parte al entrenamiento del modelo, es por eso que a veces también resulta un poco difícil elegir el modelo correcto, asumiendo que los datos pueden mantenerse en esa constante que se han venido dando. Se debe también ir modificando los parámetros de los modelos de regresión hasta obtener el mejor ajuste para nuestros datos.

Recomendación

Al trabajar en comparación de datos con otros países, y las graficas se observan como se pueden

aplicar tanto regresiones polinomiales o lineales, se deben analizar por separados los gráficos ya que a simple vista puede engañar al ojo humano y aplicar una regresión lineal en vez de una polinomial que se ajuste de mejor manera a los datos.