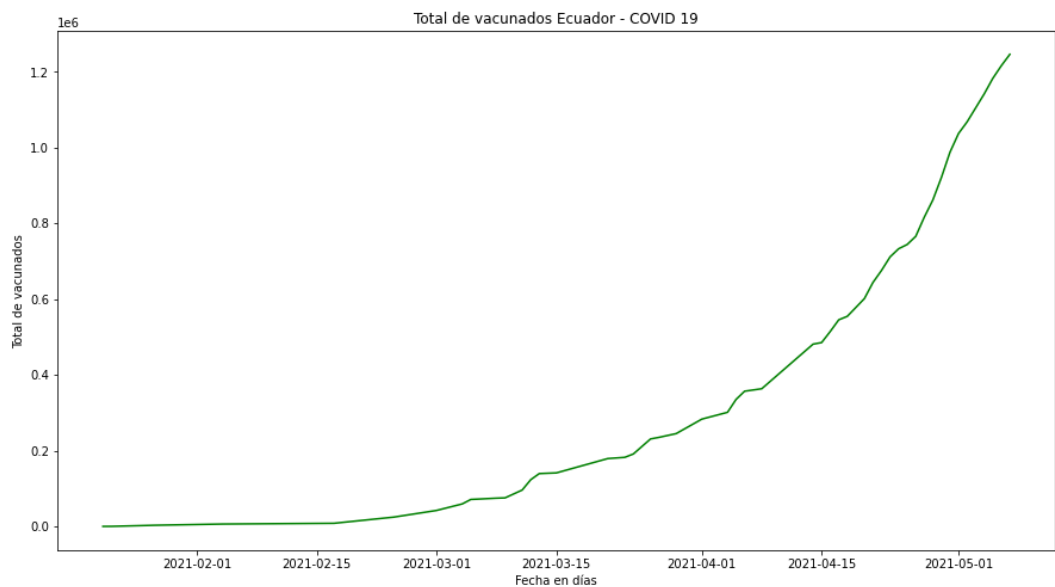


## PRUEBA 1

**Realizado por:** Tatiana Doménica Cardenas Jho

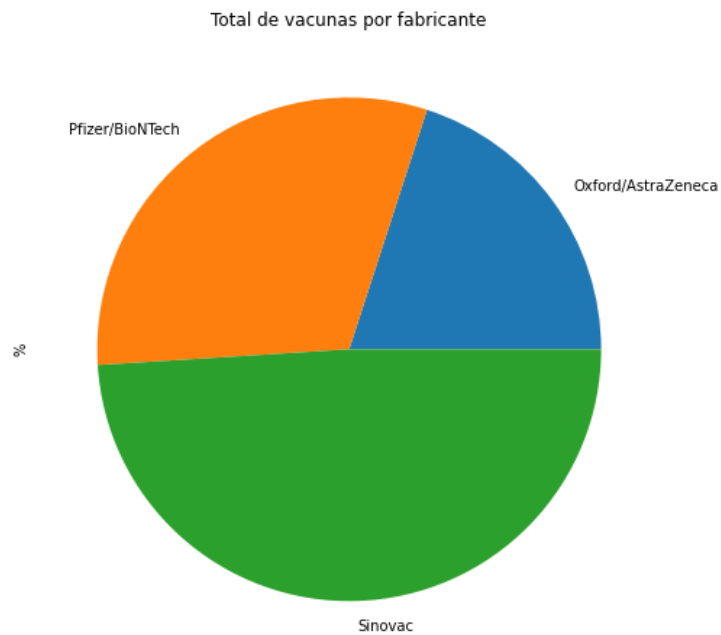
### Enunciado:

- Diseñe y desarrolle un modelo y/o *script* que permita simular el siguiente caso real:
  - Se tiene los datos del ecuador ([https://github.com/andrab/ecuacovid/tree/master/datos\\_crudos](https://github.com/andrab/ecuacovid/tree/master/datos_crudos)). En base a ello obtener los siguientes modelos:
    - Generar graficas para entender y procesar los datos:
      - Generar gráficas y reportes del total de personas vacunadas.



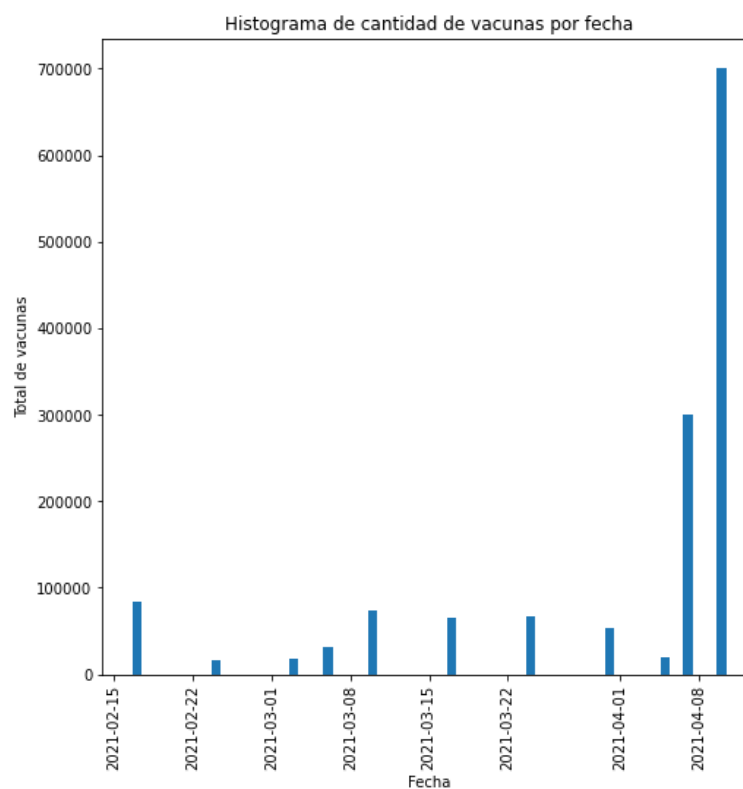
- Generar grafico de pie por fabricante de la vacuna.

A continuación, se presenta el grafico de pastel por fabricante de la vacuna para el COVID-19 que ha llegado al Ecuador. De manera inicial se graficaron los datos que se obtuvo del dataset propuesto por el docente y posteriormente se procedió a realizar un porcentaje para que este sea graficado en este diagrama.



*Ilustración 1 Diagrama de pastel con el total de las vacunas por fabricante sin datos parametrizados*

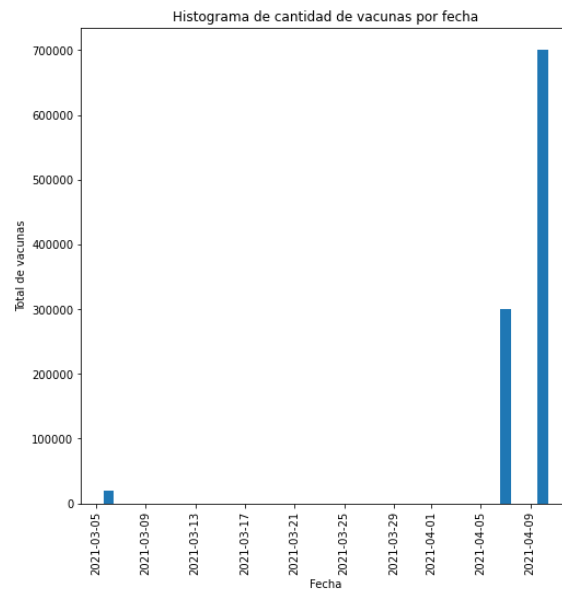
- Generar histogramas de vacunas por mes de llega y fabricante.



*Ilustración 2 Histograma con todas las vacunas por mes*

```
#Cantidad de vacunas por la fabrica de Sinovac
df_prov_s=df_prov_hist_graph.loc[(df_prov_hist_graph['vaccine'] == 'Sinovac')]
date = df_prov_s['arrived_at'].tolist()
vaccine=df_prov_s['total'].tolist()
fig = plt.figure(figsize=(8,8))
plt.title('Histograma de cantidad de vacunas por fecha')
plt.bar(date, vaccine)
plt.xlabel('Fecha')
plt.xticks(rotation=90)
plt.ylabel('Total de vacunas')

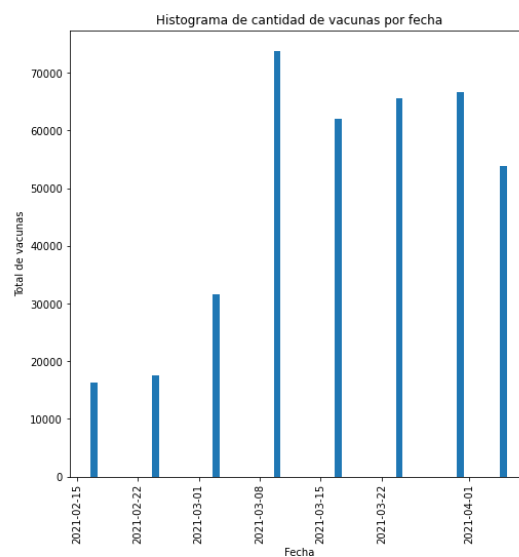
Text(0, 0.5, 'Total de vacunas')
```



*Ilustración 3* Histograma vacuna Sinovac

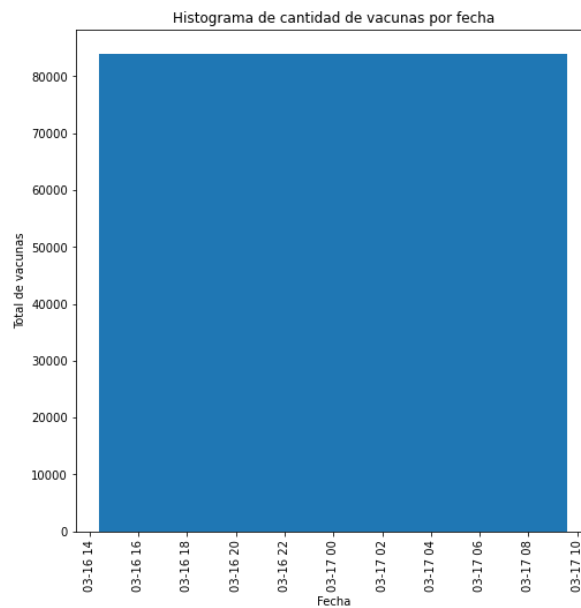
```
#Cantidad de vacunas por la fabrica de Pfizer
df_prov_p=df_prov_hist_graph.loc[(df_prov_hist_graph['vaccine'] == 'Pfizer/BioNTech')]
date = df_prov_p['arrived_at'].tolist()
vaccine=df_prov_p['total'].tolist()
fig = plt.figure(figsize=(8,8))
plt.title('Histograma de cantidad de vacunas por fecha')
plt.bar(date, vaccine)
plt.xlabel('Fecha')
plt.xticks(rotation=90)
plt.ylabel('Total de vacunas')

Text(0, 0.5, 'Total de vacunas')
```



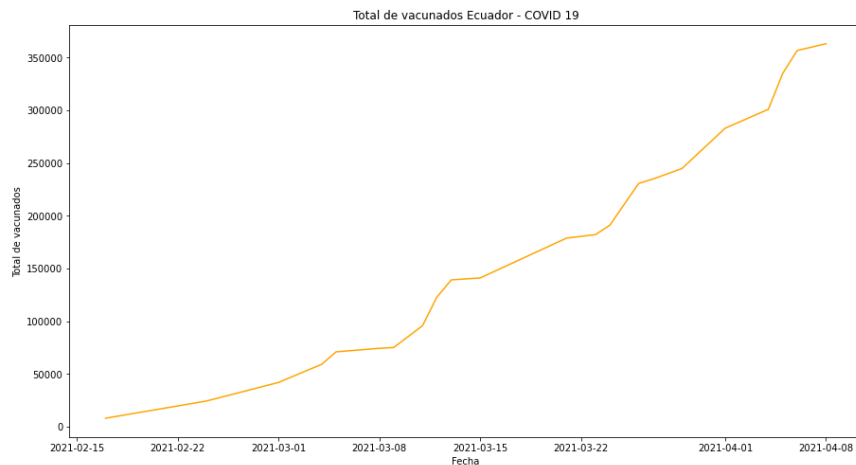
*Ilustración 4* Histograma vacuna pfizer

```
#Cantidad de vacunas por la fabrica de Astrazeneca
df_prov_o=df_prov_hist_graph.loc[(df_prov_hist_graph['vaccine'] == 'Oxford/AstraZeneca')]
date = df_prov_o['arrived_at'].tolist()
vaccine=df_prov_o['total'].tolist()
fig = plt.figure(figsize=(8,8))
plt.title('Histograma de cantidad de vacunas por fecha')
plt.bar(date, vaccine)
plt.xlabel('Fecha')
plt.xticks(rotation=90)
plt.ylabel('Total de vacunas')
Text(0, 0.5, 'Total de vacunas')
```



*Ilustración 5 Historial vacuna AstraZeneca*

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



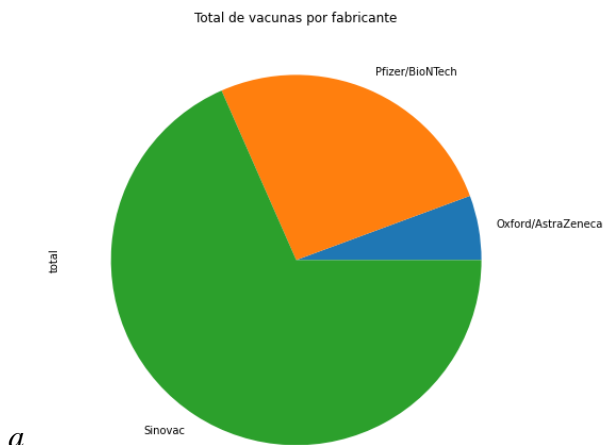
*Ilustración 6 Diagrama total vacunados con datos parametrizados*

```

: #Grafica total de vacunas por fabricante parametrizado
fig = plt.figure(figsize=(15,8))
plt.title("Total de vacunas por fabricante")
graph = df_prov_pie
graph.plot.pie()

: <AxesSubplot:title={'center':'Total de vacunas por fabricante'}, ylabel='total'>

```



a

*Ilustración 7 Diagrama pastel con datos parametrizados*

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

Primera dosis:

```

#Variables
totaldate = df['fecha'].values.reshape(-1,1)
primera_dosis = df['primera_dosis'].values.reshape(-1,1)

#Split de datos
X_train, X_test, y_train, y_test = train_test_split(totaldate, primera_dosis, test_size=0.2, random_state=42)

#Entrenamiento
linear_regressor = LinearRegression()
linear_regressor.fit(X_train, y_train)

LinearRegression()

print('Pendiente:', linear_regressor.coef_)
print('Intersección:', linear_regressor.intercept_)

Pendiente: [[6.32599515e-11]]
Intersección: [-1.02100732e+08]

#prediccion1_dosis = linear_regressor.predict(X_test)

#PRUEBA 1
m = linear_regressor.coef_[0][0]
c = linear_regressor.intercept_[0]
#Prediccion de x dias desde la fecha inicial del Dataset 2020-03-01
dias=700
label1 = m*dias-c
print("Numero de casos a los ",dias," de la fecha inicial del dataset, se obtiene una prediccion de : ",label1)

Numero de casos a los 700 de la fecha inicial del dataset, se obtiene una prediccion de : 102100732.1242712

```

*Ilustración 8 Modelo regresión 1era dosis*

Segunda dosis:

```
#Variables
totaldate = df['fecha'].values.reshape(-1,1)
segunda_dosis = df['segunda_dosis'].values.reshape(-1,1)

#Split de datos
X_train2, X_test2, y_train2, y_test2 = train_test_split(totaldate, segunda_dosis, test_size=0.2, random_state=42)

#Entrenamiento
linear_regressor2 = LinearRegression()
linear_regressor2.fit(X_train2, y_train2)

LinearRegression()

print('Pendiente:', linear_regressor2.coef_)
print('Intersección:', linear_regressor2.intercept_)

Pendiente: [[3.55126789e-11]]
Intersección: [-57350681.62325501]

#prediccion2_dosis = linear_regressor2.predict(X_test2)

#PRUEBA 1
m = linear_regressor2.coef_[0][0]
c = linear_regressor2.intercept_[0]
#Prediccion de x dias desde la fecha inicial del Dataset 2020-03-01
dias=700
label1 = m*dias-c
print("Numero de casos a los ",dias," de la fecha inicial del dataset, se obtiene una prediccion de : ",label1)

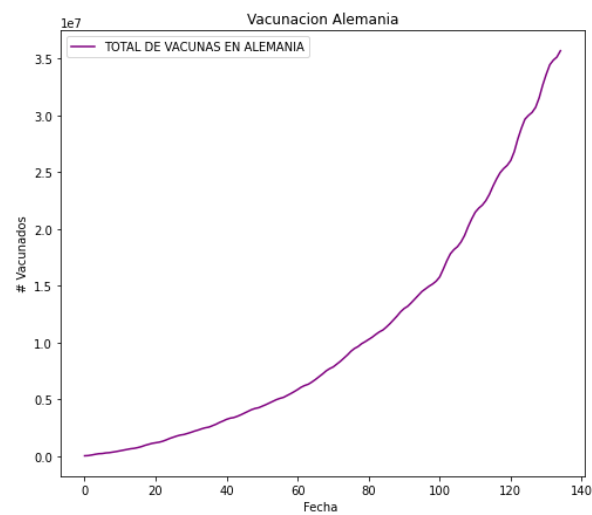
Numero de casos a los 700 de la fecha inicial del dataset, se obtiene una prediccion de : 57350681.62325504
```

### *Ilustración 9 Modelo regresión 2da dosis*

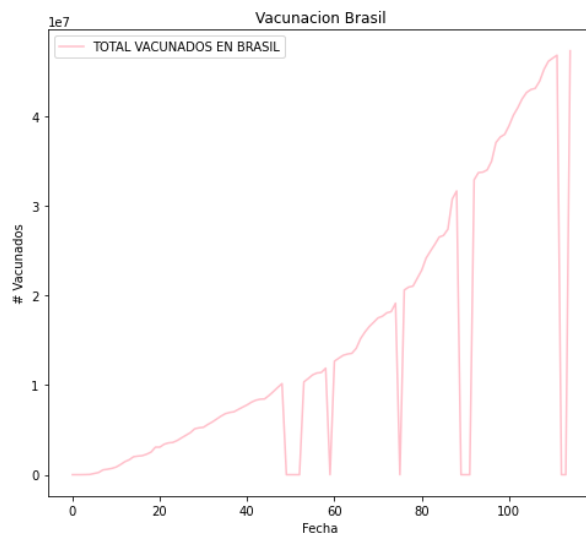
Como se puede apreciar en la prueba 1 de "PRIMERA DOSIS" vs "SEGUNDA DOSIS" existe una gran diferencia de predicción de cantidad de contagios con una dosis y con dos dosis en el Ecuador. De igual manera se puede apreciar este cambio con la prueba 2 y 3 tanto de primera dosis como de segunda dosis

- Desarrollar y generar un proceso de comparación con al menos dos países (1. Latinoamérica, 1. Europa).
  - Generar las graficas de regresión y comparar.

Para poder generar un proceso de comparación de los datos de COVID 19 entre los distintos países del mundo se ha escogido a Alemania y Grecia como países de Europa y como representante de Latinoamérica se trabajará con Brasil.

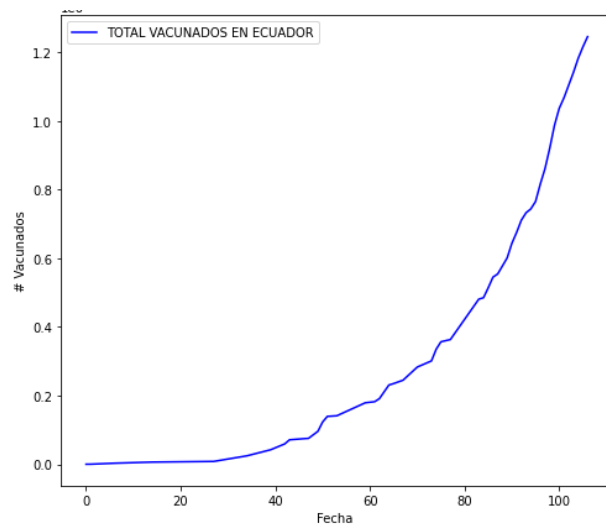


*Ilustración 10 Vacunación en Alemania*

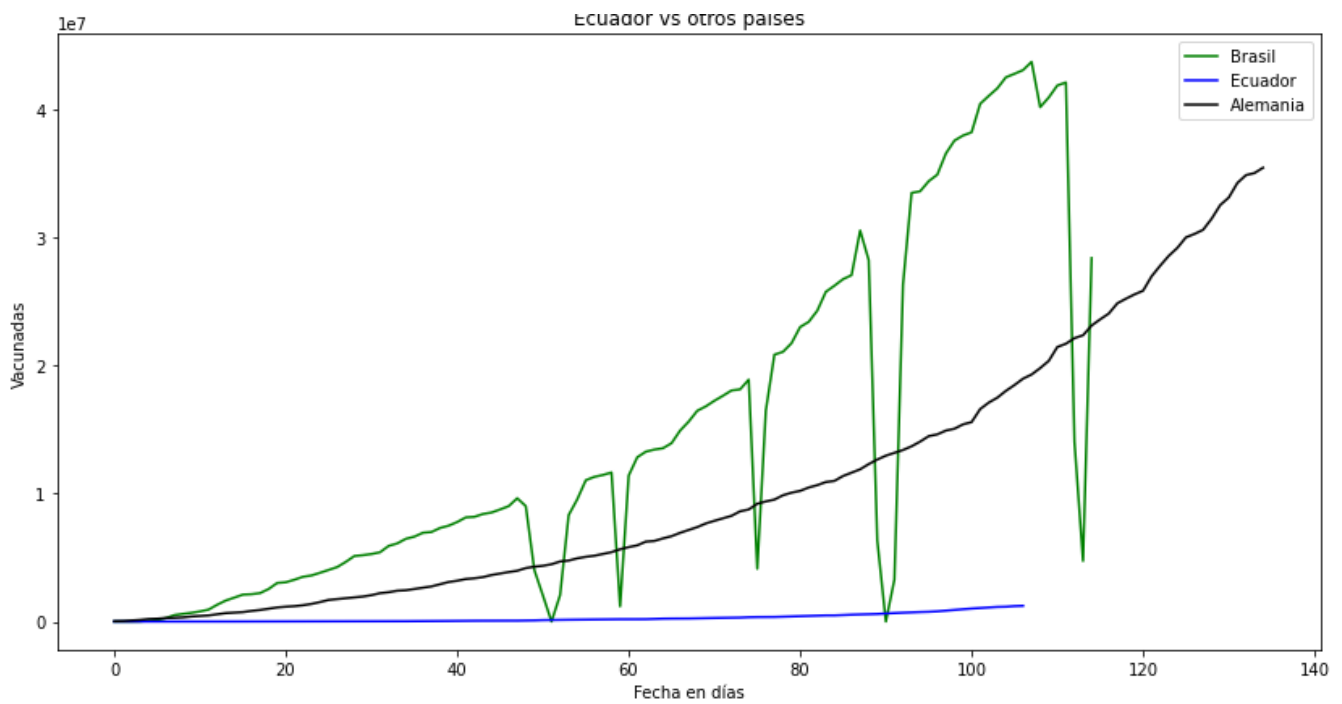


*Ilustración 10 Vacunación en Brasil*

Como se puede observar en la grafica anterior existen caidas a 0 y esto ocurrio al momento de limpiar el total de vacunas por fechas puesto que habian fechas que no tuvieron vacunas, por lo que se igualo a cero estas fechas



*Ilustración 10 Vacunación en Ecuador*



*Ilustración 11 Vacunación en Ecuador*

Como se puede presenciar en la ilustración, Ecuador se encuentra bastante inferior al resto de países, sin embargo cabe resaltar que el dataset tanto de Alemania como de Brasil comienzan en puntos diferentes, no obstante, si se tiene un dataset actualizado se pueden reflejar mejor los resultados.



## **Conclusión:**

- Podemos diferir que entre las ventajas de aplicar el Random Forest para poder hacer la comparación entre la vacunación de Ecuador versus otros países de America latina y Europa podemos encontrar que en este modelo Existen muy pocas suposiciones y por lo tanto la preparación de los datos es mínima, además puede manejar hasta miles de variables de entrada e identificar las más significativas. Método de reducción de dimensionalidad.
- La visualización de la data en diferentes graficas permite tener una noción distinta del comportamiento de la data en el dataframe, la grafica de pie es muy útil si se tiene pocas columnas, ya que al momento de ser varias este puede ser inentendible para el usuario. Para este proceso es mejor hacer una limpieza o filtrado de data que es útil para el ejercicio.