



PROYECTO 2 FINAL

GITHUB:

<https://github.com/KatiaDaniela/KATIADANIELAGARCIA-02-PROYECTO>

Katia Daniela Garcia Valenzuela

27/septiembre/2020

Grupo 2

ÍNDICE

ÍNDICE	1
Introducción	2
Definición del código	3
Solución del problema	7
Conclusión	8
Anexos	9
Definición de variables	9
Reportes	9

Introducción

A continuación se presenta el desarrollo del proyecto final del curso de análisis de datos, el cual tiene como objetivo poner en práctica los conocimientos adquiridos de Python en el curso para la clasificación de datos y realización de programas que nos permitan analizarlos.

En este caso se trata de bases de datos de una empresa cuyas operaciones se resumen en la realización de exportaciones e importaciones.

Se realizó una serie de procesos con sus respectivos códigos para poder así, extraer de los múltiples datos, información que pueda ser de utilidad para la toma de decisiones, que se reflejan en una serie de reportes emitidos.

Definición del código

La descripción detallada del código de encuentra dentro del código mismo.

El código cuenta con ciclos que permiten dar la resolución a los problemas planteados. A continuación se bosqueja el esquema general de estos.

ciclo:

```
for linea in lector:
    lista_datos.append(linea)
```

- Convierte el contenido de la colección csv en una colección lista (**lista_datos**).

```
for exp in lista_datos:
    if exp[1] == "Exports":
        rutaOrigen= exp[2]
        rutaDestino = exp[3]
```

Se procesa la **lista_datos** para seleccionar las líneas que representa exportaciones.

```
for Fre in Frecuencia:
    if Fre[0]==rutaOrigen and Fre[1]==rutaDestino:
        estaEnFrecuencia=True
        break
```

por cada línea que recupera en **exp** realiza una búsqueda sobre la lista **Frecuencia** para ir acumulando los datos de exportación.

```
for exp10 in Frecuencia:
    print(f'{exp10[0]:20} {exp10[1]:20} {exp10[2]:10} {(exp10[3]/valorExpTotal*100):5.2f}%')
    sumaPorcentajes10+=exp10[3]/valorExpTotal*100
    contador10+=1
    if contador10==10:
        break
print(sumaPorcentajes10)
```

De la colección **frecuencia** se imprimen las 10 primeras líneas debido a que representan las 10 exportaciones más concurridas.

```

for imp in lista_datos:
    if imp[1] == "Imports":
        rutaOrigen= imp[2]
        rutaDestino = imp[3]

```

Se procesa la **lista_datos** para seleccionar las líneas que representa importaciones.

```

for Fre in FrecuenciaImp:
    if Fre[0]==rutaOrigen and Fre[1]==rutaDestino:
        estaEnFrecuencia=True
        break

```

por cada línea que recupera en **imp** realiza una búsqueda sobre la lista Frecuencia para ir acumulando los datos de **imp**.

```

for imp10 in FrecuenciaImp:
    print(f'{imp10[0]:20} {imp10[1]:20} {imp10[2]:10} {(imp10[3]/valorImpTotal*100):5.2f}%')
    sumaPorcentajes10+=imp10[3]/valorImpTotal*100
    contador10+=1
    if contador10==10:
        break
print(sumaPorcentajes10)

```

De la colección **Frecuenciaimp** se imprimen las 10 primeras líneas debido a que representan las 10 importaciones más concurridas.

```

diccionario_transporte={}
for linea in lista_datos:
    if linea[1]=='Exports':
        transporte=linea[-3]
        valorExp=int(linea[-1])
        if transporte not in diccionario_transporte:
            diccionario_transporte[transporte]=[1,valorExp]
        else:
            listaLocal=diccionario_transporte[transporte]
            listaLocal[0]=listaLocal[0]+1
            listaLocal[1]=listaLocal[1]+valorExp

```

Se procesan las lista_datos para recuperar las exportaciones para generar el **diccionario_transportes** donde se acumula la información referente al medio de transporte.

```

lista_transporte=[]
for transporte in diccionario_transporte:
    listaLocal=diccionario_transporte[transporte]
    listaLocal.append(transporte)
    lista_transporte.append(listaLocal)

```

El **diccionario_transporte** se convierte en la **lista_transporte** de exportaciones.

```

for medio in lista_transporte:
    print(f'{medio[2]:5}    {medio[0]:10}    {medio[1]:19.2f}')
    contador+=1
    if contador==3:
        break

```

De la **lista_transporte** se imprimen los 3 medios de transporte con más importes por exportación.

```

diccionario_transporte={}
for linea in lista_datos:
    if linea[1]=='Imports':
        transporte=linea[-3]
        valorImp=int(linea[-1])
        if transporte not in diccionario_transporte:
            diccionario_transporte[transporte]=[1,valorImp]
        else:
            listaLocal=diccionario_transporte[transporte]
            listaLocal[0]=listaLocal[0]+1
            listaLocal[1]=listaLocal[1]+valorImp

```

La **lista_datos** se procesa para seleccionar las importaciones y generar un diccionario de datos que acumula la información para medios de transporte de exportación.

```

lista_transporte.sort(reverse=True, key=lambda x:x[1])
for medio in lista_transporte:
    print(f'{medio[2]:5}    {medio[0]:10}    {medio[1]:19.2f}')
    contador+=1
    if contador==3:
        break

```

De la **lista_transporte** se imprimen los 3 medios de transporte con más importes por exportación.

```

sumaPorcentajes=0
Frecuencia.sort(reverse=True, key=lambda x:x[3])
for exp in Frecuencia:
    print(f'{exp[0]:20}    {exp[1]:20}    {exp[2]:10}    {(exp[3]/valorExpTotal*100):5.2f}%')
    sumaPorcentajes+=exp[3]/valorExpTotal*100
    if sumaPorcentajes>=80:
        break

print(f'{sumaPorcentajes:70.2f}%')

```

De la lista **Frecuencia** se imprimen las exportaciones que representan el 80% del total de exportaciones.

```

FrecuenciaImp.sort(reverse=True, key=lambda x:x[3])
sumaPorcentajes=0
for imp in FrecuenciaImp:
    print(f'{imp[0]:20}    {imp[1]:20}    {imp[2]:10}    {(imp[3]/valorImpTotal*100):5.2f}%')
    sumaPorcentajes+=imp[3]/valorImpTotal*100
    if sumaPorcentajes>=80:
        break

print(f'{sumaPorcentajes:70.2f}%')

```

De la lista **FrecuenciaImp** se imprimen las importaciones que representan el 80% del total de importaciones.

```

listaImp=[]
sumaPorcentajes=0.0
for linea in FrecuenciaImp:
    listaImp.append(linea[0]+"-"+linea[1])
    sumaPorcentajes+=linea[3]/valorImpTotal*100
    if sumaPorcentajes>=80:
        break
conjuntoImp=set(listaImp)

```

De la lista **FrecuenciaImp** se genera una lista de importaciones con origen-destino. Después se convierte en el **conjuntoImp** (cuyo contenido es origen-destino).

```

listaExp=[]
sumaPorcentajes=0.0
for linea in Frecuencia:
    listaExp.append(linea[0]+"-"+linea[1])
    sumaPorcentajes+=linea[3]/valorExpTotal*100
    if sumaPorcentajes>=80:
        break
conjuntoExp=set(listaExp)
interseccion=conjuntoExp.intersection(conjuntoImp)

```

De la lista **FrecuenciaExp** se genera una lista de importaciones con origen-destino. Después se convierte en el **conjuntoExp** (cuyo contenido es origen-destino).

Solución del problema

Para desarrollar la solución del problema se utilizaron las listas que representan la información de una empresa así como el desarrollo de nuevas listas, diccionarios y conjuntos para el mejor manejo de la información. Se utilizó el lenguaje de programación Python para programar la solución software para cada uno de los requisitos planteados. Se logró cumplir exitosamente el 100% del objetivo del proyecto.

Conclusión

En definitiva, encontrar un método adecuado para la extracción y selección de datos con el fin de dar conclusiones para la toma de decisiones en un negocio no es tarea fácil. Creo que hay que hacer uso de más de un método y tomar en cuenta el objetivo que se quiere alcanzar.

Dicho esto, me parece que el mejor método (de los establecidos dentro del problema) de análisis de datos es el tercero: rutas que representan el 80% de los ingresos de la compañía. Intentando hacer uso de Pareto, enfocando esfuerzos a aquellas rutas que en verdad son las que aportan más ingresos.

El prestar atención a las rutas que más le generan valor a la empresa es una decisión que se toma pensando en el objetivo final de toda compañía: generar ingresos. Se trata de generar valor para el cliente pero también generar valor para la empresa. Sin utilidades no hay forma de que se pueda continuar operando.

Así mismo, realicé un análisis extra basado en el anterior: rutas que se encuentran dentro del pareto y coinciden en que se importa y exporta. Lo anterior para prestar extra atención ya que la mayoría de las actividades de valor se generan en estas tres rutas: Japón-México, Alemania-China y China-Japón. Además, dos de estas rutas (Japón-México y China- Japón) se encuentran dentro de las 10 más demandadas.

El hecho de que estas coincidan en exportación e importación, además de coincidir dentro de las rutas más demandadas, nos lleva a decisiones como aumentar inversión en estas rutas y enfocar esfuerzos de colocación de transporte, vendedores, instalaciones, presupuesto y demás, en ellas, reduciendo así costos y perdidas en inversion no redituable.

Igualmente, los resultados arrojados por este análisis de datos invitan a realizar otro más a profundidad: ¿qué es lo que se está haciendo en estas rutas para que tengan tanto éxito? y ¿qué de eso se puede replicar en las otras para que sean igual de exitosas? Y de este modo lograr un crecimiento sostenido para todas las rutas y en consecuencia, para la empresa en general.

Anexos

Definición de variables

[Las variables se explican en el código.](#)

Reportes

Titulo de la figura se encuentra en la parte inferior de la misma

10 EXPORTACIONES MÁS CONCURRIDAS			
Origen	Destino	Frecuencia	%Ingresos
South Korea	Vietnam	497	4.29%
Netherlands	Belgium	437	2.02%
USA	Netherlands	436	0.64%
China	Mexico	330	7.65%
Japan	Brazil	306	2.10%
Germany	France	299	1.81%
South Korea	Japan	279	2.87%
Australia	Singapore	273	0.31%
Canada	Mexico	261	5.28%
China	Spain	250	1.16%
28.1379726583802			
10 IMPORTACIONES MÁS CONCURRIDAS			
Singapore	Thailand	273	7.23%
Germany	China	233	2.39%
China	Japan	210	5.83%
Japan	Mexico	206	7.06%
China	Thailand	200	6.90%
Malaysia	Thailand	195	6.27%
Spain	Germany	142	3.68%
Mexico	USA	122	4.26%
China	United Arab Emirates	114	6.05%
Brazil	China	113	0.74%
50.41060365941507			

Pantalla #1. 10 exportaciones e importaciones más concurridas

EXPORTACIÓN: 3 MEDIOS DE TRANSPORTE CON MÁS IMPORTES		
Trasnporte	Frecuencia	Importe Total
Sea	8434	65592622000.00
Rail	2780	34505043000.00
Air	2045	32785147000.00
IMPORTACIÓN: 3 MEDIOS DE TRANSPORTE CON MÁS IMPORTES		
Trasnporte	Frecuencia	Importe Total
Sea	2254	34938000000.00
Rail	601	9123000000.00
Road	449	5990000000.00

Pantalla #2. 3 medios de transportes más usados para exportaciones e importaciones

EXPORTACIÓN: RUTAS REPRESENTANTES DEL 80% DE LOS INGRESOS			
Origen	Destino	Frecuencia	Porcentaje
China	Mexico	330	7.65%
Canada	Mexico	261	5.28%
South Korea	Vietnam	497	4.29%
France	Belgium	223	3.46%
France	United Kingdom	147	3.39%
China	South Korea	189	2.99%
USA	Mexico	194	2.94%
South Korea	Japan	279	2.87%
Germany	Italy	130	2.84%
China	Germany	142	2.55%
China	USA	111	2.37%
Russia	Belorussia	219	2.20%
France	USA	103	2.12%
Japan	Brazil	306	2.10%
Netherlands	Belgium	437	2.02%
USA	Argentina	139	1.97%
Germany	France	299	1.81%
China	Japan	133	1.66%
Japan	China	196	1.59%
USA	Canada	136	1.51%
USA	Brazil	101	1.49%
USA	Belgium	211	1.48%
Russia	China	130	1.44%
Germany	China	95	1.35%
Russia	Germany	120	1.27%
Japan	USA	123	1.23%
Russia	Netherlands	110	1.22%
France	Germany	109	1.21%
Japan	Mexico	179	1.19%
China	Spain	250	1.16%
South Korea	China	94	1.06%
Belgium	United Kingdom	146	1.02%
USA	Singapore	75	1.00%
France	Spain	116	0.91%
Russia	Turkey	80	0.90%
Italy	Germany	109	0.88%
United Kingdom	Spain	176	0.86%
China	Argentina	143	0.77%
South Korea	USA	77	0.77%
Japan	South Korea	62	0.77%
Spain	Russia	99	0.68%
			80.29%

Pantalla #3. Rutas que representan el 80% de los ingresos de exportaciones

IMPORTACIONES: RUTAS REPRESENTANTES DEL 80% DE LOS INGRESOS			
Origen	Destino	Frecuencia	Porcentaje
Singapore	Thailand	273	7.23%
Japan	Mexico	206	7.06%
China	Thailand	200	6.90%
Malaysia	Thailand	195	6.27%
China	United Arab Emirates	114	6.05%
China	Japan	210	5.83%
Mexico	USA	122	4.26%
Japan	United Arab Emirates	76	4.03%
Spain	Germany	142	3.68%
Germany	Mexico	70	3.25%
South Korea	United Arab Emirates	51	3.08%
USA	Germany	58	2.92%
USA	Thailand	109	2.82%
Germany	China	233	2.39%
South Korea	Mexico	66	2.32%
China	Singapore	55	2.28%
Mexico	Japan	65	2.06%
USA	India	57	2.04%
France	Poland	83	1.85%
Spain	Mexico	41	1.72%
Australia	Japan	102	1.66%
Russia	India	42	1.53%
			81.25%

Pantalla #4. Rutas que representan el 80% de los ingresos de importaciones

```
_____païse con exportaciones e importaciones_____
{'Japan-Mexico', 'Germany-China', 'China-Japan'}
```

Pantalla #5. Intersección de pantalla 3 y 4