
REPORTE 02

PROYECTO FINAL DE INTRODUCCIÓN AL ANÁLISIS
DE DATOS CON PYTHON

BECAS SANTANDER TECNOLOGÍA:

DESARROLLA COMPETENCIAS PARA LA REVOLUCIÓN
DIGITAL

Óscar Hernández Terán

[https://github.com/oscarteran/
EmTech_Reporte02](https://github.com/oscarteran/EmTech_Reporte02)

Índice

1. Objetivo	3
2. Introducción	3
2.1. Descripción del caso	3
3. Desarrollo	3
3.1. Primer inspección	3
4. Resultados	4
4.1. Creación de rutas y separación de base de datos	4
4.2. 10 rutas más demandadas.	4
4.3. Medios de transporte más utilizados	6
4.4. Valor total	7
4.4.1. Importaciones	8
4.4.2. Exportaciones	8
5. Gráficos	9
5.1. Top 10 rutas más importantes.	9
5.1.1. Importaciones	9
5.1.2. Exportaciones	9
5.2. Demanda por medio de transporte	10
5.2.1. Importaciones	10
5.2.2. Exportaciones	10
5.3. Países con 80 % de la demanda.	11
5.3.1. Importaciones	11
5.3.2. Exportaciones	11
6. Conclusiones	12

Índice de figuras

1. Estructura original de la base de datos	3
2. Estructura modificada de las BD.	4
3. Top 10 rutas Importaciones.	5
4. Top 10 rutas Importaciones.	5
5. Aportación porcentual al total del top 10 Rutas.	6
6. Transportes más utilizados en ambas direcciones.	7
7. Países y sus respectivas aportaciones al valor total en Importaciones.	8
8. Países y sus respectivas aportaciones al valor total en Exportaciones.	8
9. Top 10 rutas de importación	9
10. Top 10 rutas de exportación	9
11. Demanda de transportes en importación	10
12. Demanda de transportes en importación	10
13. Países con mayor concentración de demanda	11

14. Países con mayor concentración de demanda	11
---	----

1. Objetivo

Poner en práctica las herramientas esenciales de Python para el análisis de datos a partir de la clasificación y manejo de los mismos, mediante la creación y el uso de archivos y estructuras de datos, implementación y creación de funciones, uso de módulos importados y destructuring.

2. Introducción

2.1. Descripción del caso

Synergy Logistics es una empresa dedicada a la intermediación de servicios de importación y exportación de diferentes productos. Actualmente la empresa cuenta con una base de datos que refleja las rutas más importantes que opera desde el año 2015, con su respectivo origen y destino, año, producto, modo de transporte y valor total. Su propósito, es que a partir de estos datos se genere un análisis que sirva de la base para la estructuración de su estrategia operativa.

3. Desarrollo

3.1. Primer inspección

Como primer paso en cualquier trabajo de análisis de datos, vamos tanto a cargar como a dar una primer inspección a la estructura de nuestra base de datos. Como se nos permitió el uso de la biblioteca Pandas, haremos uso de esta para almacenar el archivo en una estructura de tipo DataFrame. El resultado de cargar e imprimir en consola se muestra en la figura 2.

```
----- Forma general de la Base de datos (BD) -----
```

	register_id	direction	origin	destination	year	date	product	transport_mode	company_name	total_value
0	1	Exports	Japan	China	2015	31/01/15	Cars	Sea	Honda	33000000
1	2	Exports	Japan	China	2015	01/02/15	Cars	Sea	Honda	16000000
2	3	Exports	Japan	China	2015	02/02/15	Cars	Sea	Honda	29000000
3	4	Exports	Japan	China	2015	03/02/15	Cars	Sea	Honda	14000000
4	5	Exports	Japan	China	2015	04/02/15	Cars	Sea	Honda	17000000
5	6	Exports	Japan	China	2015	05/02/15	Cars	Sea	Honda	20000000
6	7	Exports	Japan	China	2015	06/02/15	Cars	Sea	Honda	27000000
7	8	Exports	Japan	China	2015	07/02/15	Cars	Sea	Honda	33000000
8	9	Exports	Japan	USA	2015	08/02/15	Machinery and electronics	Air	ReachExt J.J	15000000
9	10	Exports	Japan	USA	2015	09/02/15	Machinery and electronics	Air	ReachExt J.J	31000000

Figura 1: Estructura original de la base de datos

4. Resultados

4.1. Creación de rutas y separación de base de datos

Como primer punto, separamos la base de datos por la categoría de dirección, teniendo ahora dos DataFrames, uno para *Importaciones* y otro para *Exportaciones*. Esta separación se realizó con ayuda de pandas. Una vez que se encuentran separados los datos, realizamos una función que nos crea las rutas por las cuales se envían los productos.

Por último en este apartado, creamos una columna que únicamente contiene el valor de 1 para todos los renglones. Esto lo hicimos para ayudarnos a contar cuantas veces se repiten los valores de interés. Las nuevas bases de datos tienen la siguiente estructura.

```
----- Creación de la columna de rutas -----
```

IMPORTACIONES												
	register_id	direction	origin	destination	year	date	product	transport_mode	company_name	total_value	Total_Use	Routes
0	15409	Imports	Mexico	USA	2015	31/01/15	Cars	Rail	Honda	3000000	1	Mexico-USA
1	15410	Imports	Mexico	USA	2015	01/02/15	Cars	Rail	Honda	13000000	1	Mexico-USA
2	15411	Imports	Mexico	USA	2015	02/02/15	Cars	Rail	Honda	40000000	1	Mexico-USA
3	15412	Imports	Mexico	USA	2015	03/02/15	Cars	Rail	Honda	48000000	1	Mexico-USA
4	15413	Imports	Mexico	USA	2015	04/02/15	Cars	Rail	Honda	44000000	1	Mexico-USA

EXPORTACIONES												
	register_id	direction	origin	destination	year	date	product	transport_mode	company_name	total_value	Total_Use	Routes
0	1	Exports	Japan	China	2015	31/01/15	Cars	Sea	Honda	33000000	1	Japan-China
1	2	Exports	Japan	China	2015	01/02/15	Cars	Sea	Honda	16000000	1	Japan-China
2	3	Exports	Japan	China	2015	02/02/15	Cars	Sea	Honda	29000000	1	Japan-China
3	4	Exports	Japan	China	2015	03/02/15	Cars	Sea	Honda	14000000	1	Japan-China
4	5	Exports	Japan	China	2015	04/02/15	Cars	Sea	Honda	17000000	1	Japan-China

Figura 2: Estructura modificada de las BD.

4.2. 10 rutas más demandadas.

En este punto resumimos las 10 rutas que más demanda tienen, tanto en importaciones como exportaciones para tener mayor claridad de los datos. Es importante mencionar que la demanda aquí se priorizó por el valor total, ya que una mayor ocurrencia de la ruta en la base de datos, no se refleja en mayor valor comercial, por esto, para maximizar ganancias y minimizar gastos, se decidió ordenar a partir de la columna **total value**.

Las 10 rutas más importantes en importaciones se observan en [3](#) mientras que las 10 más importantes en exportaciones en [4](#).

Algo que nos ayudará a tomar una decisión más informada es, *¿Qué tanto contribuyen al valor total estas 10 rutas?*. Podemos dar respuesta a esta pregunta haciendo un conteo porcentual de las aportaciones del top 10 de rutas contra el resto de ellas. El resultado se observa en [5](#).

```

-----
----- Top 10 rutas -----

```

IMPORTACIONES		
	total_value	Total_Use
Routes		
Singapore-Thailand	4017000000	273
Japan-Mexico	3918000000	206
China-Thailand	3831000000	200
Malaysia-Thailand	3482000000	195
China-United Arab Emirates	3357000000	114
China-Japan	3237000000	210
Mexico-USA	2365000000	122
Japan-United Arab Emirates	2238000000	76
Spain-Germany	2044000000	142
Germany-Mexico	1804000000	70

Figura 3: Top 10 rutas Importaciones.

EXPORTACIONES		
	total_value	Total_Use
Routes		
China-Mexico	12250000000	330
Canada-Mexico	8450000000	261
South Korea-Vietnam	6877007000	497
France-Belgium	5538069000	223
France-United Kingdom	5427000000	147
China-South Korea	4790000000	189
USA-Mexico	4710000000	194
South Korea-Japan	4594000000	279
Germany-Italy	4541000000	130
China-Germany	4090000000	142

Figura 4: Top 10 rutas Importaciones.

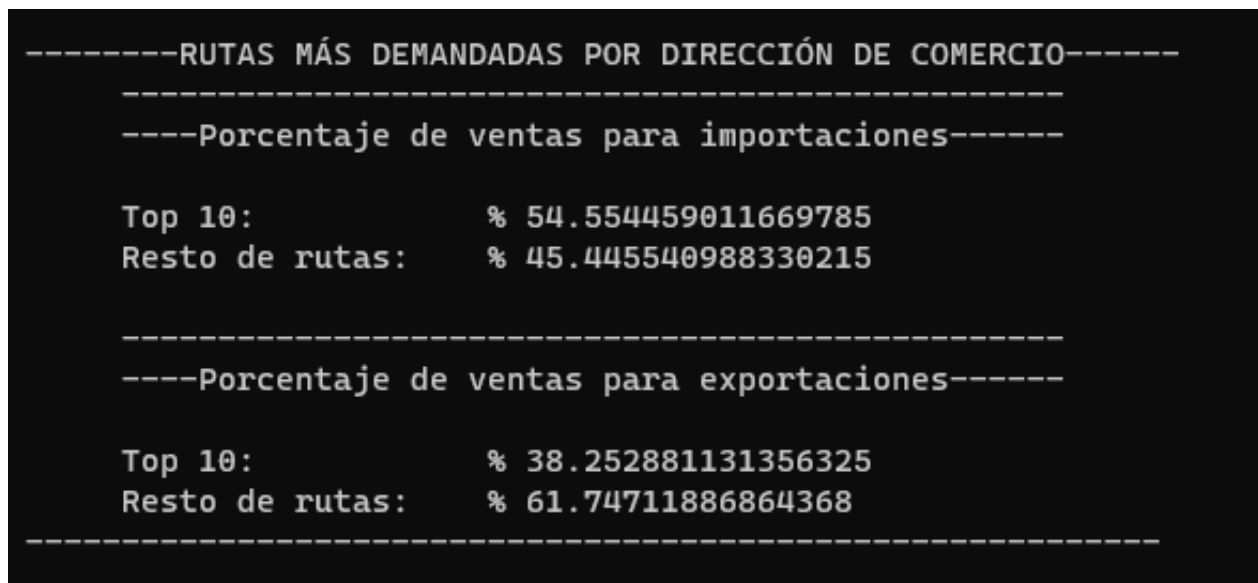


Figura 5: Aportación porcentual al total del top 10 Rutas.

4.3. Medios de transporte más utilizados

Para cumplir con este punto seguimos haciendo uso de la misma lógica de programación funcional, así también explotando el uso de los métodos integrados en la biblioteca Pandas.

Podemos observar de la base de datos que tenemos 4 tipos de transportes, estos son:

- Sea
- Rail
- Road
- Aire

Como se hizo anteriormente, se ordenaron los modos de transportes por el valor total del mismo, sin embargo, para este caso podemos notar que se cumple el orden descendente tanto para *total value* así como para *Total Use*. El orden de importancia comercial de los medios de transportes se describen en la figura 6.

----- Transportes -----		
IMPORTACIONES		
	total_value	Total_Use
transport_mode		
Sea	34938000000	2254
Rail	9123000000	601
Road	5990000000	449
Air	5477000000	344
EXPORTACIONES		
	total_value	Total_Use
transport_mode		
Sea	65592622000	8434
Rail	34505043000	2780
Road	27280486000	2149
Air	32785147000	2045

Figura 6: Transportes más utilizados en ambas direcciones.

4.4. Valor total

Un aspecto bastante importante para la priorización y toma de decisiones inteligentes es resaltar que países son los que concentran una mayor demanda, y por lo tanto, mayor comercialización y beneficios para la empresa **Synergy Logistics**. En las figuras 7 y 8 listamos aquellos países, tanto para importaciones como para exportaciones, que concentran en ellos solos el 80 % del valor total de las operaciones de la empresa, respectivamente para cada dirección de operación.

4.4.1. Importaciones

```

-----
----- Países con aportación del 80% -----
-----
IMPORTACIONES
Countries  Total_value  Percent
0      Thailand 13745000000 24.753278
1      Mexico  8894000000 40.770422
2  United Arab Emirates 7848000000 54.903832
3      Japan  5828000000 65.399438
4      Germany 5109000000 74.600202
5      USA    3758000000 81.367959

```

Figura 7: Países y sus respectivas aportaciones al valor total en Importaciones.

4.4.2. Exportaciones

```

EXPORTACIONES
Countries  Total_value  Percent
0      Mexico 28315015000 17.678841
1      USA    13159409000 25.895086
2      Belgium 13099380000 34.073851
3      Germany 11303451000 41.131305
4      China  11259992000 48.161625
5      Japan  8777152000 53.641752
6  United Kingdom 8150673000 58.730729
7      Brazil 7508016000 63.418454
8  South Korea  7321668000 67.989831
9      Vietnam 6877007000 72.283578
10     Italy   6109329000 76.098016
11     Spain   5934191000 79.803104
12     France  5850549000 83.455969

```

Figura 8: Países y sus respectivas aportaciones al valor total en Exportaciones.

5. Gráficos

A continuación, se muestra un total de 6 figuras, que corresponden a cada punto de los antes descritos, creadas con la herramienta *Matplotlib* para la mejor apreciación de las características de nuestra base de datos.

5.1. Top 10 rutas más importantes.

5.1.1. Importaciones

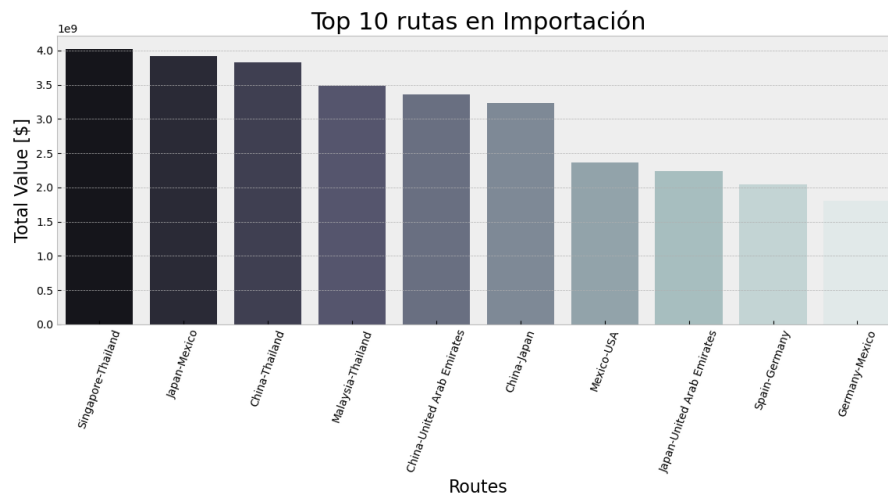


Figura 9: Top 10 rutas de importación

5.1.2. Exportaciones

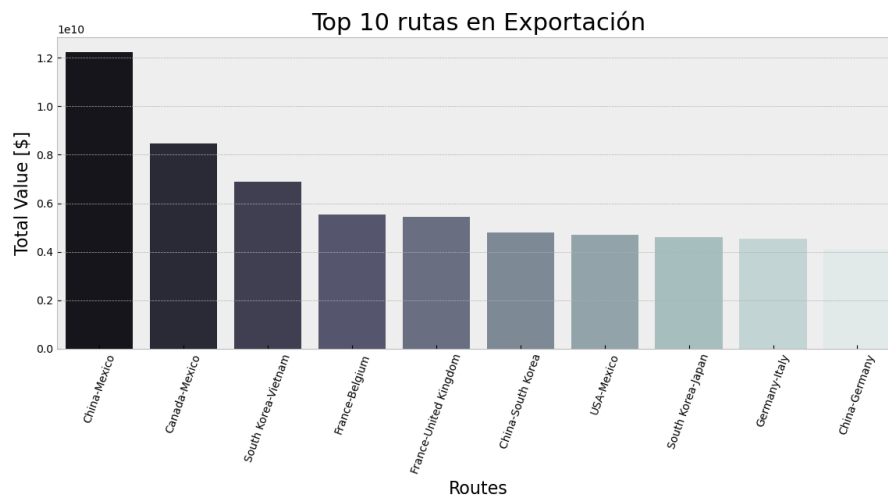


Figura 10: Top 10 rutas de exportación

5.2. Demanda por medio de transporte

5.2.1. Importaciones

Porcentaje del valor de las Importaciones según transporte

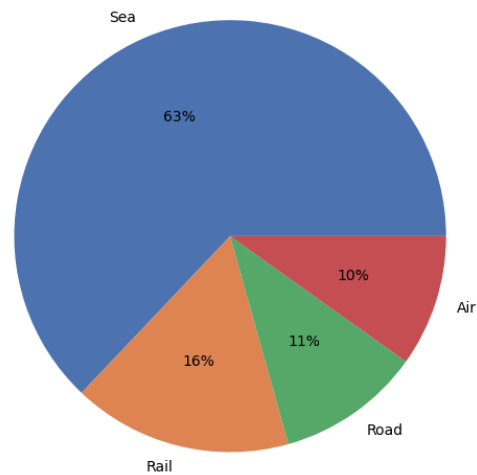


Figura 11: Demanda de transportes en importación

5.2.2. Exportaciones

Porcentaje del valor de las Exportaciones según transporte

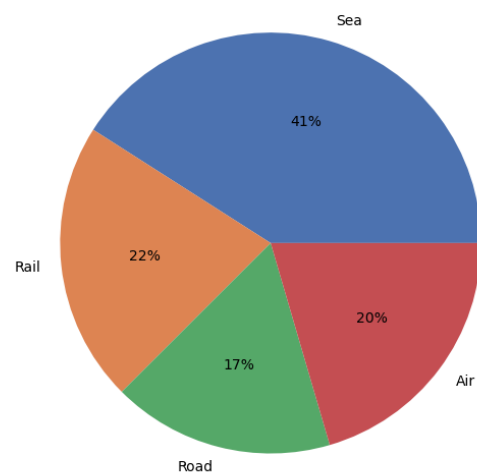


Figura 12: Demanda de transportes en importación

5.3. Países con 80 % de la demanda.

5.3.1. Importaciones

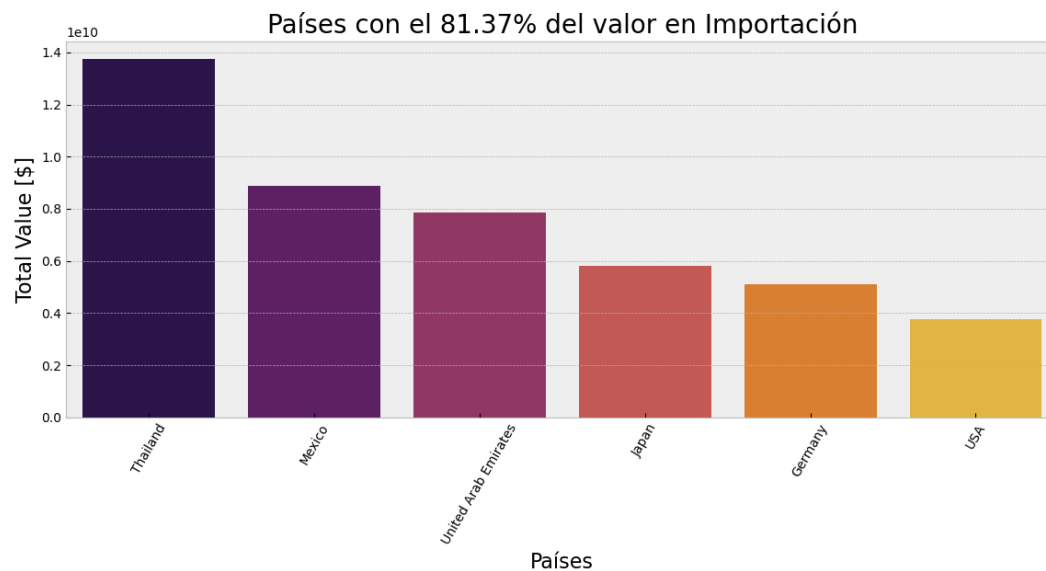


Figura 13: Países con mayor concentración de demanda

5.3.2. Exportaciones

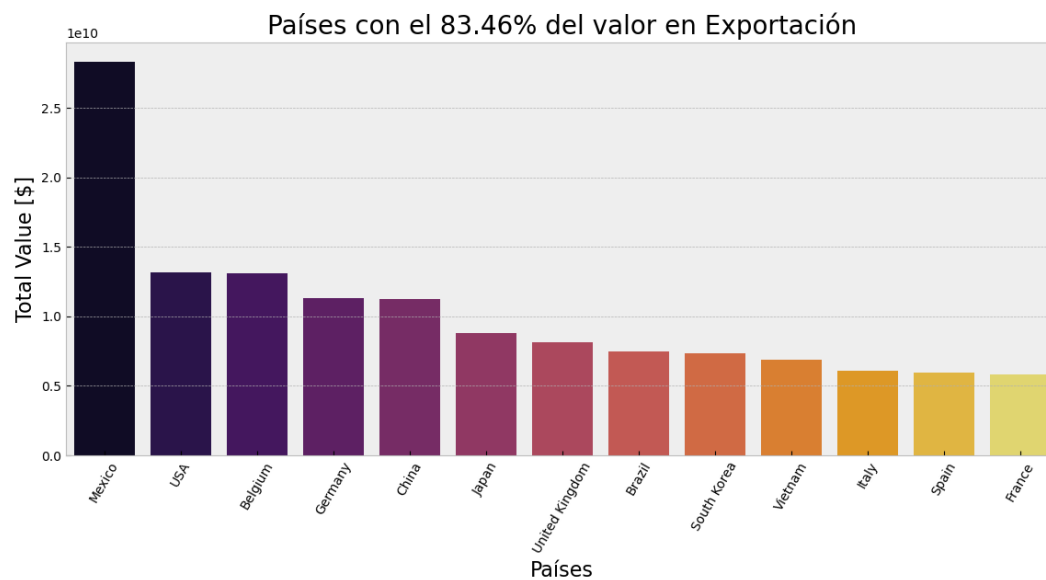


Figura 14: Países con mayor concentración de demanda

6. Conclusiones

Opción 1) Rutas de importación y exportación. Como se menciona *Synergy Logistics* está considerando implementar los esfuerzos en las 10 rutas más importantes que maneja actualmente. En las figuras 9 y 10 observamos cuales son estas rutas y su respectivo valor comercial. Para mejorar el análisis realizamos un análisis de aportaciones porcentuales en conjunto del top 10 de rutas con el resto de ellas. Vemos en 5 que tanto aportan estas rutas.

Podemos concluir que en *importaciones* es bastante rentable enfocar los esfuerzos en estas, ya que solo 10 rutas se reflejan en el 54 % de los ingresos.

Sin embargo, en cuanto a exportaciones, se recomienda aumentar esfuerzos en estas rutas pero sin dejar de lado el resto de ellas, como vemos, el top 10 representa únicamente el 38 % del valor total, por lo que un enfoque total se reflejaría en posibles pérdidas en el resto de rutas.

Opción 2) Medio de transporte utilizado. La empresa *Synergy Logistics* cuenta con 4 medios de transporte para mover sus productos. La administración se ha preguntado si es necesario reducir el uso del medio de transporte con menor demanda a fin de tener ahorros, y por lo tanto mejores utilidades.

Tanto en la figura 11 como en 12 podemos ver un Gráfico de pastel en el que se detallan los porcentajes de ocupación de cada medio.

En *Importaciones* es claro el dominio que tiene el transporte por medio marítimo, con un total de 63 %, los medios terrestres y aéreos están casi empatados con 11 % y 10 %, respectivamente.

En *Exportaciones* vemos también un dominio por parte del medio marítimo, pero ciertamente no tan drástico, en último lugar tenemos al medio terrestre con 17 %.

Las recomendaciones son que en *importaciones* se haga un énfasis en el medio marítimo, y una reducción leve en los medios terrestres y aéreos.

En cuanto a las *exportaciones*, se recomienda una reducción en el transporte terrestre, enfocando esfuerzos en los 3 restantes para el aumento de las utilidades.

Opción 3) Valor total de importaciones y exportaciones. *Synergy Logistics* ha decidido enfocarse en los países que concentran el 80 % de sus utilidades. El equipo de *data analyst* ha realizado un análisis y encontró los países que representan este valor, estos países se muestran en las figuras 13 y 14 para importaciones y exportaciones, respectivamente.

El equipo también considera que el enfoque en estos países se reflejará en un aumento de la productividad, por lo que recomienda redoblar esfuerzo en este grupo.