

EMTECH INSTITUTE: Proyecto #2

Darío Chaires Arciniega

September 28, 2020



Contents

1	Introducción	3
2	Solución al problema	4
3	Conclusión	6
4	Referencias	7

1 Introducción

Este proyecto tiene como función poner en práctica los aspectos básicos de programación con python con un enfoque en análisis y clasificación de datos, utilizando adecuadamente el uso de variables, números, listas, operadores lógicos, sentencias, controles de flujo y funciones previamente estudiados en el curso de data science.

El objetivo principal es proponer una solución a un determinado problema basándose en las herramientas estadísticas de un conjunto de datos. El caso a resolver tiene como protagonista a una empresa de logística internacional llamada Synergy logistics , la cual tiene una base de datos en la que cada movimiento es registrado, así como información sobre el ingreso generado y fecha de realización. Con los datos brindados se nos dan tres consideraciones para una futura estrategia operativa en el 2021:

1) Rutas de importación y exportación: Basándonos en los datos que nos brindan, se nos pide hacer un análisis de las 10 rutas más demandadas para enfocar esfuerzos en estas últimas.

2) Medio de transporte utilizado: Buscar los tres medios de transporte con mayor demanda basándonos en el total ingresos. El fin es encontrar aquél medio de transporte con menor impacto en el ingreso neto.

3) Valor total de importaciones y exportaciones: Con los ingresos netos, la empresa busca enfocarse en los países que generan el 80% del valor de importaciones y exportaciones.

El objetivo tras estas opciones es escribir un código que nos ayude a la clasificación de los datos y facilitar los análisis sugeridos en cada una de las opciones.

2 Solución al problema

Yo opté por basarme principalmente en la tercera opción, el análisis sobre las rutas que generan el 80% del ingreso neto. Trabajar directamente en el análisis de ingresos me parece la forma más simple e intuitiva para proponer una estrategia efectiva en el siguiente año. A su vez, realicé el análisis de las 10 rutas más demandadas como herramienta complementaria. El análisis en general decidí dividirlo por importaciones y exportaciones dada que hay una diferencia significativa tanto en el número de movimientos como en el ingreso por dirección.

Para iniciar el análisis, me pareció que el uso de diccionarios sería más útil para evitar confusión en el código, de modo que cada línea en el archivo *csv* es almacenada como un diccionario cuyas claves son cada una de las columnas presentes en la primera fila, esta fila representa el encabezado de la base de datos.

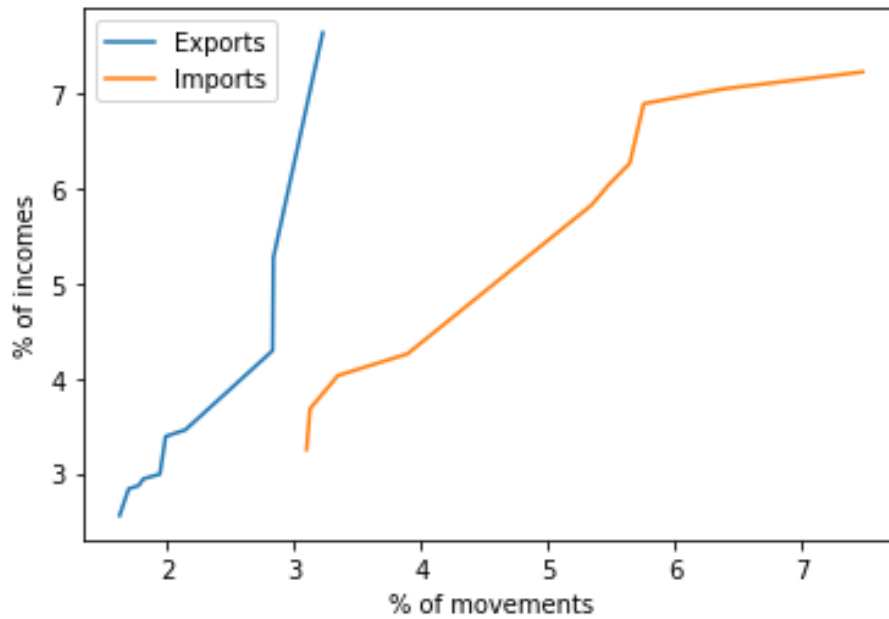


Figure 1: En estas gráficas podemos ver la evolución del porcentaje de ingresos con respecto al porcentaje de movimientos de cada ruta. En el eje y tenemos el porcentaje individual de ingresos por ruta, en el x el porcentaje individual de movimientos, la gráfica azul representa las exportaciones y amarilla las importaciones

El siguiente paso fue crear una función, que llamé *f1*, la cual enlista todas las rutas por frecuencia de movimientos y total de ingreso por orden. El único parámetro que la función incluye una variable que llamé *direction* que toma como argumentos los elementos asociados a las claves *Exports* e *Imports*. La función regresa una tupla con tres elementos: una lista de las rutas con su frecuencia ordenadas de mayor a menor, la suma total de ingresos y, otra lista con las rutas y sus ingresos totales también de mayor a menor.

Posteriormente creo seis variables globales a las que podré acceder a lo largo del código. Para verificar cuáles son las principales rutas que generan el 80% del total de ingresos y el 60% más demandado. El código suma individualmente los porcentajes y cuando llega al 80% para los ingresos y al 60% para las demandas, aplico un `break` para detener el ciclo `for` que enlista las rutas.

Como bloque final, grafico la relación entre los ingresos y movimientos de las primeras 10 rutas, tanto para exportaciones como importaciones, como se puede ver en la figura 1. En las graficas se puede observar que la relación entre el porcentaje de ganancias y movimientos es siempre creciente, al menos para las 10 primeras rutas. Sin embargo la gráfica de exportaciones tiene una mayor afinidad en la relación ingreso-movimientos, de modo que se le podría dar enfoque a esas transacciones. Por otro lado, la gráfica de importaciones resulta más irregular, aún así cabe recalcar que el rango de porcentajes entre 3.5% y el 6% es el que presenta un mayor incremento en los ingresos con respecto al porcentaje de movimientos, incluso es posible percibir un cambio de concavidad que posteriormente se estabiliza en las rutas de importación con mayor demanda.

Este análisis simple pone en evidencia la correlación que existe entre el total de ingresos y los movimientos de cada ruta, hace más fácil intuir que las 10 primeras rutas seleccionadas con mayor demanda suponen una parte esencial en el número total de ingresos.

3 Conclusión

Al ver una relación creciente entre las 10 primeras rutas enlistadas, podemos verificar que los esfuerzos en el 2021 deberían estar particularmente enfocados en las rutas con mayor número de movimientos. Por otro lado, la cuestión del medio de transporte podría tener también una importante relación en el porcentaje de ingresos, cuyos efectos también deberían ser tomados en cuenta, por simplicidad y eficacia de los resultados decidí omitir esta parte en mi análisis al considerarla secundaria.

Como trabajo posterior, se podría hacer un filtro sobre los ingresos generados en los últimos tres años, para en este caso tener una estrategia más adecuado a corto plazo y considerar el medio de transporte menos utilizado para utilizar los ingresos de una mejor forma.

Como nota final, en este proyecto me fue posible ver el potencial que tiene python como lenguaje de programación y como herramienta de análisis de datos con las librerías que están incluidas en *Spider*. Análogamente, pude usar el potencial de las funciones, tanto funciones lambda como las propiamente definidas, para mejorar la organización del código. El uso de de librerías para graficar también fueron de ayuda en este análisis para hacer una síntesis más autocontenida de los resultados obtenidos. En general, este proyecto me hizo precenciar el potencial del uso de análisis de datos con un enfoque más laboral, diferente al enfoque académico en el que me ido desarrollando este último año de mi carrera.

4 Referencias

El código escrito y la base de datos utilizada en este proyecto están disponibles en el repositorio público de GitHub anexado aquí:

<https://github.com/dchezz/Python-project2/upload/master>