

# Análisis de Distribución en una Cafetería para la Variable "Money"

Sebastian Jimenez Bauer - A01708830

## 1. Propósito del Análisis

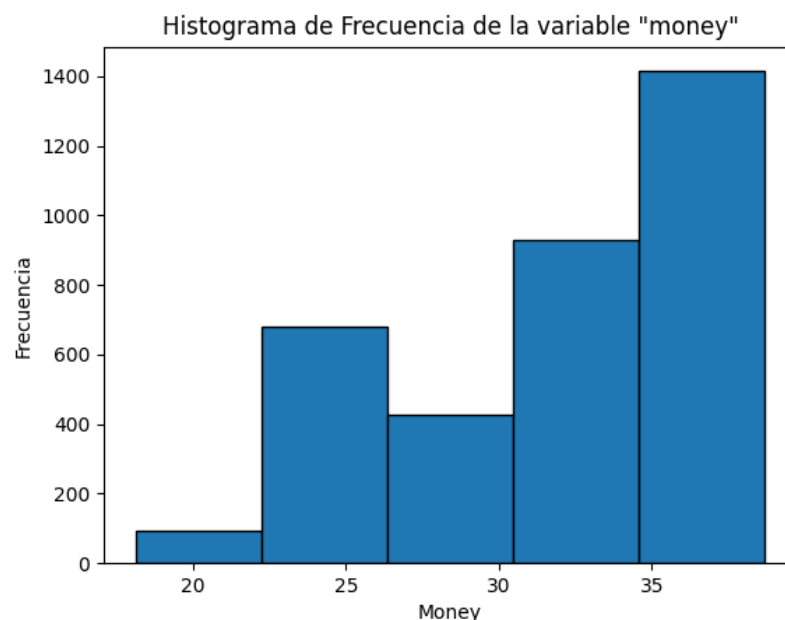
En este reporte se va a analizar la variable de dinero para poder modelar el comportamiento utilizando una distribución de probabilidad. Utilizando herramientas como la librería Fitter en Python se identificaran rangos de valor y la utilidad de este tipo de modelado.

## 2. Descripción del Proceso y los Datos

La variable "Money", representa la frecuencia de la cantidad de dinero que se gastan los clientes en una cafetería. Los datos fueron recolectados utilizando una base de datos en Kaggle y se utilizan para observar la frecuencia y distribución en la cafetería. El conjunto de datos viene de un registro de transacciones de una cafetería que no se menciona directamente por cuestiones de privacidad. Esta información contiene detalles sobre las ventas, tipos de pago, hora de la compra y la preferencia de los clientes. Para el uso de este ejemplo se utilizo solamente el costo y la frecuencia de gasto.

## 3. Análisis Visual: Histograma de Frecuencia

Para la primera aproximación se creo un histograma de frecuencia utilizando la librería de Fitter. Este Histograma nos permite visualizar, el centro y la dispersión: Multimodalidad:

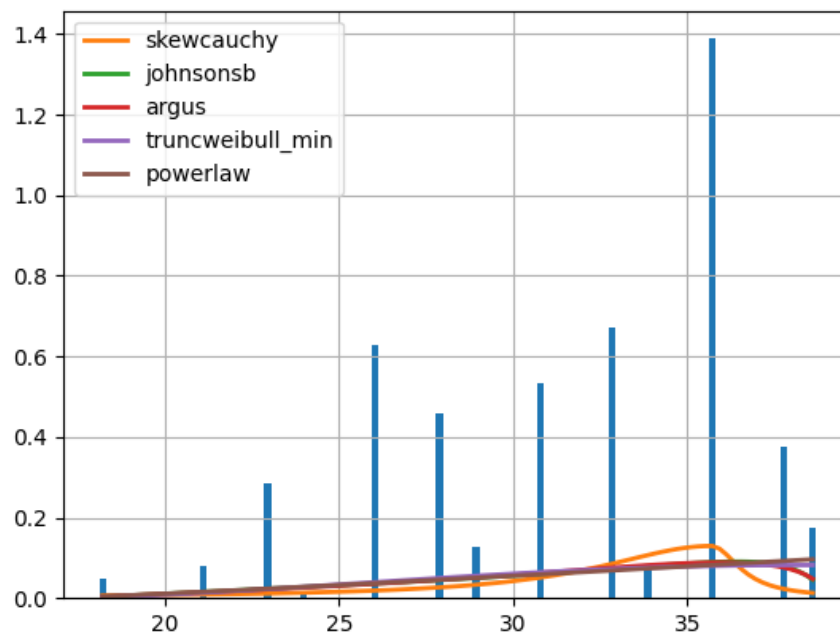


Se encontraron concentraciones entre los valores 25 - 32 y en el rango de 35 a 38. La distribución no es simétrica, tiene un sesgo significativo a la izquierda. El rango definido donde hay mayoría de los valores se encuentran entre 18 y 38

Esto significa que los datos no tienen una única tendencia clara, sino hay una dependencia de varios procesos y valores.

#### 4. Modelado: Distribuciones de Probabilidad

Se utilizó la librería Fitter de Python para poder comparar la distribución empírica de los datos con distribuciones de probabilidad teóricas.



El resultado muestra las distribuciones continuas estándar skewcauchy, johnsonsb, argus, powerlaw y cruncweibull no representan correctamente los datos de la cafetería. Las curvas teóricas son demasiado diferentes mostrar los cambios agudos y altos en el histograma.

Este ajuste indica que la información medida no sigue una distribución continua simple.

## 5. Conclusión y Reflexiones Finales

Esta información nos permite observar que los datos de la cafetería, específicamente la frecuencia en el que los clientes gastan cierta cantidad de dinero no es una distribución empírica. La distribución es multimodal y tiene un sesgo significativo hacia la izquierda, y el rango de los valores mas frecuentes es entre 35 y 38. Esto significa que la gente que va a la cafetería en promedio gasta entre 35 y 38.

Este ejercicio demuestra la importancia de modelar fenómenos reales, ya que con estos datos la cafeteria puede tomar acciones de marketing y estrategia de negocio para poder capturar la atención de sus clientes. Aunque no se haya encontrado una distribución teórica, el análisis nos permitió identificar que el proceso no es homogéneo y tiene varios puntos de concentración de valor. A pesar de no tener un modelo formal, sabemos qué rangos de valores son los mas comunes. Al final este conocimiento es importante para la toma de decisiones de negocio para entender que estrategias se podrían implementar para mejorar el negocio.

El modelado de distribuciones es una herramienta poderosa para poder convertir los datos crudos en información útil y accionables. Normalmente la realidad es mucho más compleja que los modelos teóricos estándar pero con datos crudos y conclusiones basadas en esos datos siempre hay acciones que se pueden tomar y observaciones que valen la pena estudiar y entender.

## 6. Referencias

Cokelaer, T. (2017). Fitter: Fit distributions to data. Zenodo. <https://doi.org/10.5281/zenodo.10397459>

Kaushal, N. (Año de publicación). Título del conjunto de datos [Conjunto de datos]. Kaggle. Recuperado el 28 de septiembre de 2025, de <https://www.kaggle.com/docs/datasets>