# Práctica 2: Limpieza y validación de los datos

1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

El dataset contiene información financiera de las 200 más importantes criptomonedas y fue elaborado por CoinMarketCap.com. Presenta los siguientes campos:

**Currency**, nombre de la criptomoneda, por ejemplo Bitcoin, la criptomoneda mas popular.

**Date**, es el dia de registro de los valores.

**Open**, es el precio en dólares con que empezó el día la criptomoneda.

**High**, es el precio más alto durante el día.

**Low**, es el precio más bajo durante el día.

**Close**, es el precio en dólares con que cerró el día la criptomoneda.

**Volume**, es la cantidad total en dólares que se comercio durante el día.

**Marketcap**, es la capitalización de mercado de la criptomoneda.

Es importante porque en la actualidad el uso y comercio de las monedas virtuales está en auge, y supone un mercado por explotar para nuevas posibilidades de inversión. En este trabajo se busca responder las siguientes preguntas:

**¿Cuáles fueron las cinco criptomonedas con mayor capitalización de mercado durante el 2017?**

**¿Cuál fue el mejor día de la semana para comprar cada criptomoneda durante el 2017?**

2. Limpieza de los datos.

**Selección de los datos de interés a analizar. ¿Cuáles son los campos más relevantes para responder al problema?**

Los campos más relevantes para responder el problema son:

**Currency, High, Low, Date**

¿Los datos contienen ceros o elementos vacíos? ¿Y valores extremos? ¿Cómo gestionarías cada uno de estos casos?

En este dataset, no se encontró datos con elementos vacios o ceros, pero de todas maneras la verificación en el código fue puesta por si se actualiza el dataset. Lo que si se encontró en el dataset fueron valores de cadena en un campo de tipo numérico ("**Market Cap**") y se procedió a su remoción, así como su conversión ("**MarketCap**") a formato numérico ya que el campo estaba siendo considerado de tipo cadena al contener comas. De forma similar el campo "Date" que representa el día registrado no estaba siendo considerado como tipo fecha, por lo tanto también se procedió a su conversión.

3. Análisis de los datos.

**Selección de los grupos de datos que se quieren analizar/comparar.**

**Comprobación de la normalidad y homogeneidad de la varianza. Si es necesario (y posible), aplicar transformaciones que normalicen los datos.**

Se determino seleccionar las cinco criptomonedas con mayor capitalización de mercado durante el 2017, debido a la volatilidad de este mercado. En base a esto, se seleccionaron las criptomonedas Bitcoin, ethereum, bitcoin-cash, ripple y litecoin.

Se creó un dataset para cada una de las criptomonedas con las columnas "Diff" (diferencia entre el mayor valor y el menor valor de venta) y "**WeekDay**" (día de la semana).

Se aplicó el test ANOVA de un factor, para determinar si la distribución de cada uno de los grupos de días de la semana, para cada una de las criptomonedas. Los resultados se muestran a continuación:

Bitcoin ANOVA One-Way P Value is: 0.8691538820823383

Ethereum ANOVA One-Way P Value is: 0.8133476152425794

Bitcoin-cash ANOVA One-Way P Value is: 0.8568052682129781

Ripple ANOVA One-Way P Value is: 0.6556193305015214

Litecoin ANOVA One-Way P Value is: 0.8902377563037155

Para la moneda Bitcoin el resultado fue :

Día Ganancia

Viernes 406.911569

Jueves 348.483529

Miércoles 337.615962

Sábado 312.480980

Domingo 281.660577

Martes 275.825962

Lunes 262.574615

4. Representación de los resultados a partir de tablas y gráficas.

A continuación se presentan los gráficos obtenidos para responder a las dos preguntas mencionadas anteriormente.

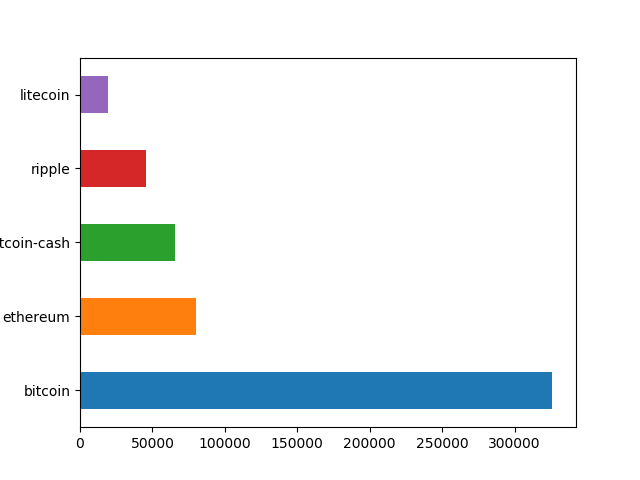


Figura 1. Las cinco monedas con mayor capitalización de mercado durante el 2017.

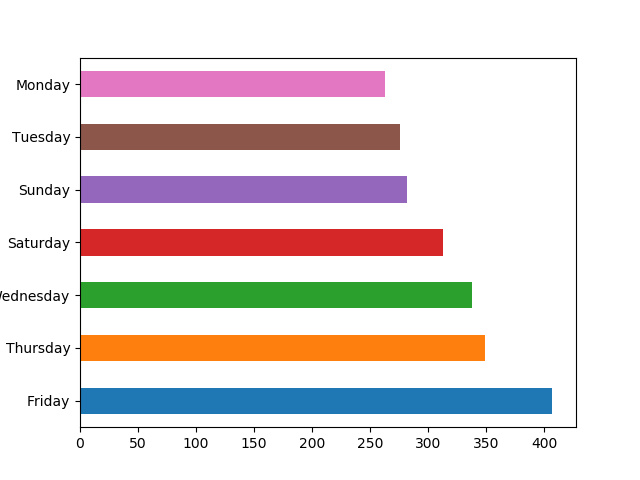


Figura 2. Viernes fue el mejor día para comprar Bitcoin durante el 2017.

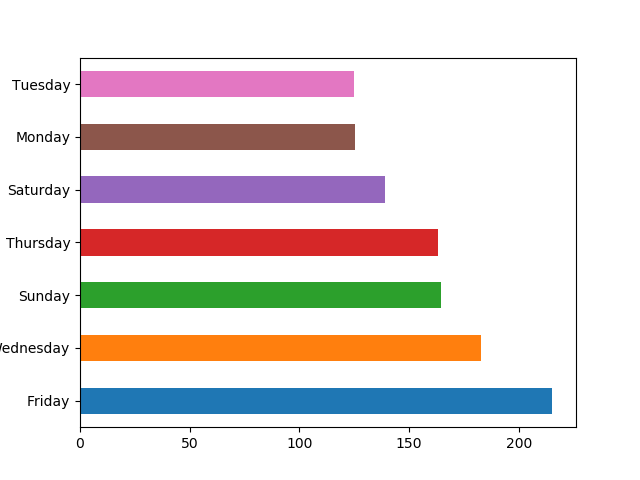


Figura 3. Viernes fue el mejor día para comprar Ethereum durante el 2017.

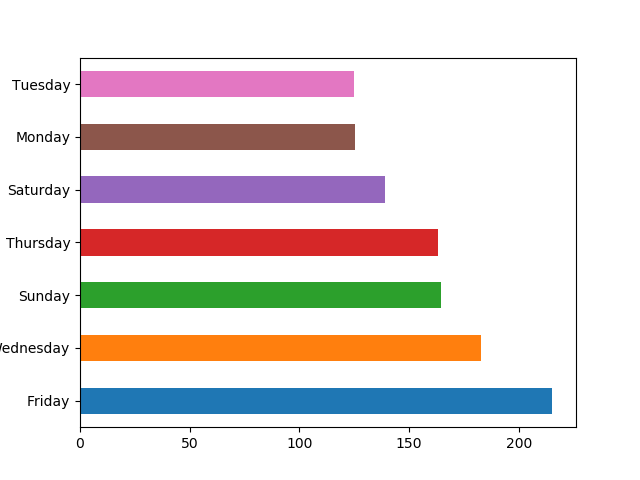


Figura 4. Viernes fue el mejor día para comprar Bitcoin-Cash durante el 2017.

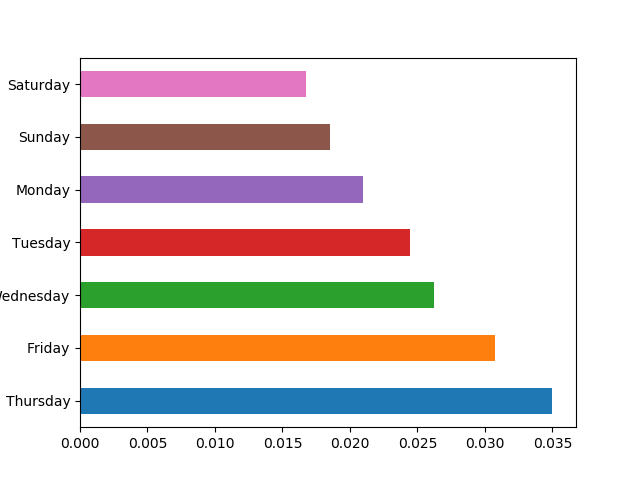


Figura 5. Jueves fue el mejor día para comprar Ripple durante el 2017.

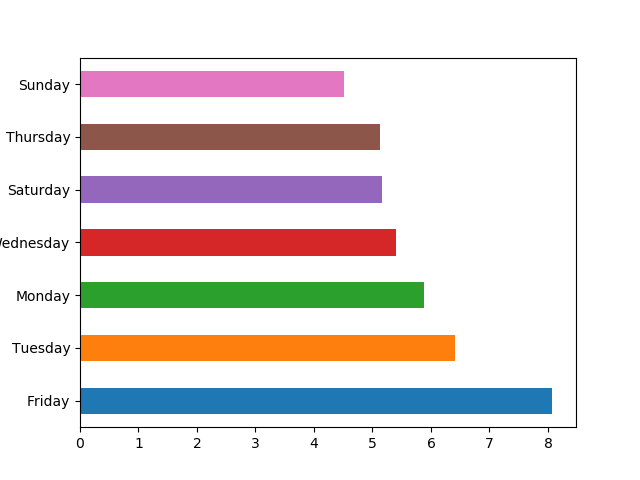


Figura 6. Viernes fue el mejor día para comprar Litecoin durante el 2017.

5. Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

A partir de los datos se puede inferir lo siguiente, para las monedas con mayor valor como Bitcoin, Bitcoin-dash, Ethereum y Litecoin el viernes fue el mejor día para comprar mientras que para Ripple fue el día Jueves.

6. Código: Hay que adjuntar el código, preferiblemente en R, con el que se ha realizado la limpieza, análisis y representación de los datos. Si lo preferís, también podéis trabajar en Python.