

## **Practica 2.**

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

Es un dataset con las cotizaciones de diferentes criptomonedas a lo largo de un periodo máximo de 7 días. Además, el dataset contiene un total de 100 coins siendo los primeros 100 coins en cotización.

El análisis que vamos a llevar a cabo es importante porque nos puede ayudar a determinar las criptomonedas más importantes en cotización así como aquellas más volátiles y en donde la inversión es más arriesgada.

### **2. Integración y selección de los datos de interés a analizar.**

La selección de los datos ha sido después de una búsqueda en internet y de localizar algunos trabajos en Kaggle interesantes sobre el tema. Además, el interés del autor porque la temática me ha llevado a la página web <https://coinmarketcap.com/es/> y a realizar la integración de los mismos mediante un API call. El fichero se ha descargado en Json de acuerdo con los estándares habituales en comunicación de datos mediante servidores visto hasta ahora en la asignatura.

### **3. Limpieza de los datos.**

#### **3.1. ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos? 3.2. Identificación y tratamiento de valores extremos.**

El dataset contiene elementos vacío que se registran como NAN en el csv, sin embargo, en nuestro supuesto nos decantamos por eliminarlos ya que no nos aportan valor significativo en el análisis. Sin embargo, cabe destacar que existen métodos alternativos para tratar estos supuestos como rellenar las observaciones con la media de los mismos en dicha feature o con el más frecuente de estos.

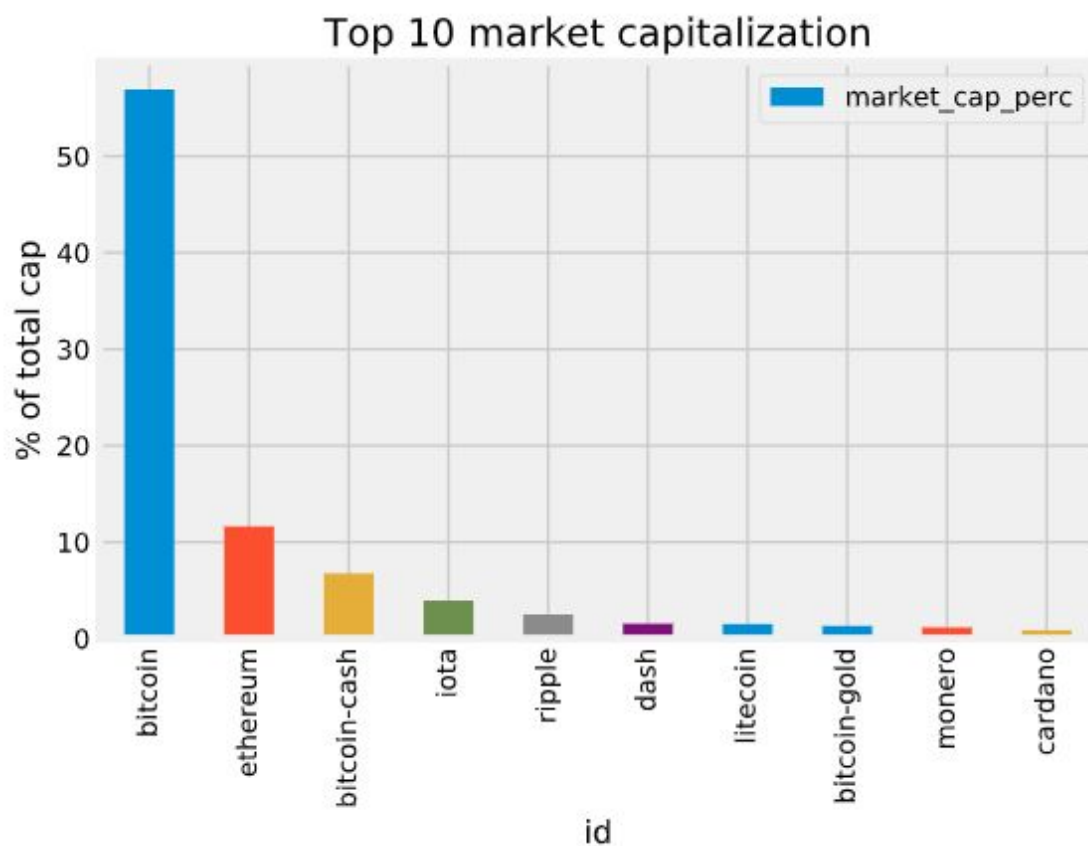
Como se puede apreciar en el notebook, en cuanto al bitcoin nos surge un supuesto de valores extremos debido a su alta capitalización en relación con otros coins. Hemos decidido tratar este supuesto con un método de normalización a través de la aplicación de un algoritmo logarítmico con Log10 en su plotting para hacerlo más intuitivo. No obstante, existen múltiples alternativas.

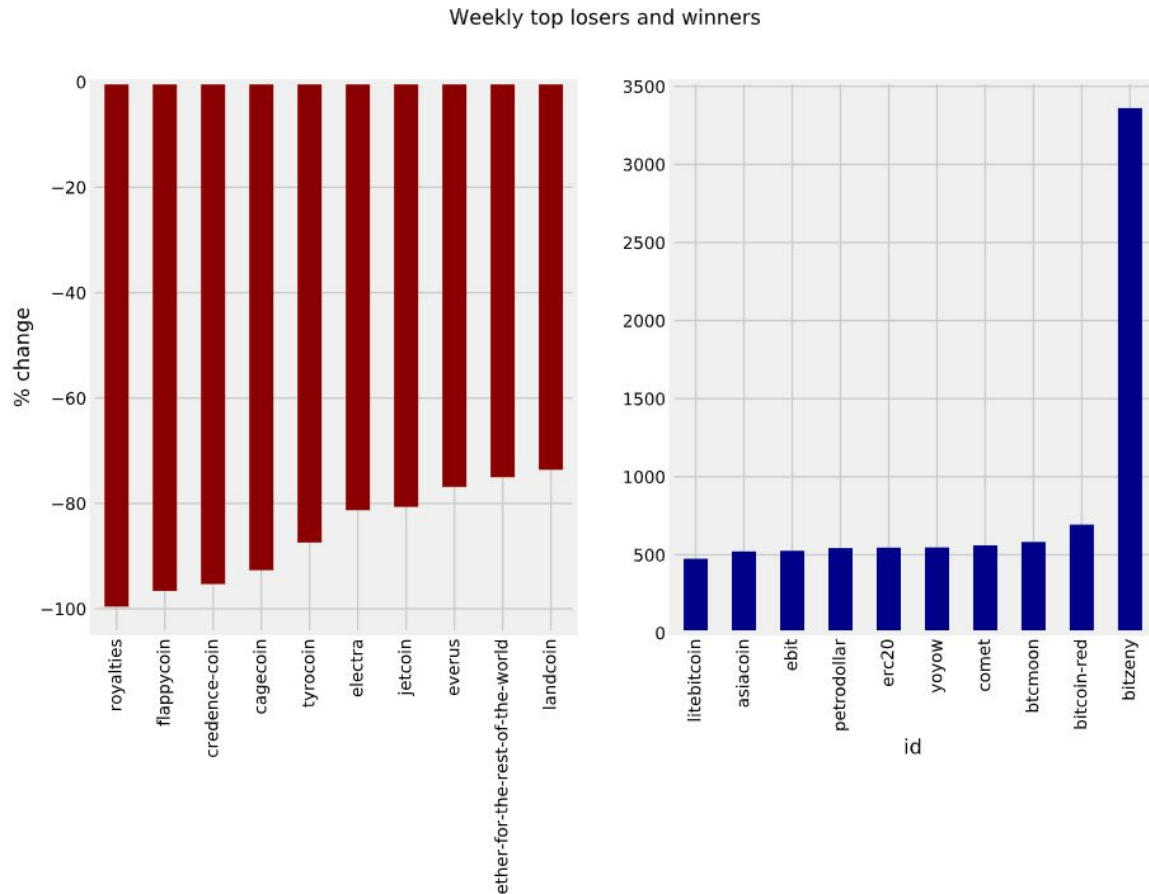
### **4. Análisis de los datos.**

**4.1. Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar). 4.2. Comprobación de la normalidad y homogeneidad de la varianza. 4.3. Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc.**

Hemos hecho un análisis de la distribución de los datos, un sorting de los mismos y un filtro de aquellos que destacan más. Este análisis previo nos ha ayudado a tener una visión más clara de nuestro objetivo y del camino a trazar. Llegamos a este punto, aplicar métodos estadísticos básicos como el análisis de la distribución o pruebas estadísticas para comparar coins distintos nos acabado de aportar una intuición adecuada para sacar conclusiones. Por último, hacer un análisis de los porcentajes de capitalización con una función lambda nos acaba de ayudar a determinar con exactitud el objetivo de nuestro estudio.

## 5. Representación de los resultados a partir de tablas y gráficas.





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

Una de la conclusión que se obtienen del análisis es que las coins con menos capitalización son más fluctuantes y conllevan más riesgos que sus hermanas más capitalizadas (como bitcoin). Además, se aprecia la gran disparidad que existe hoy en día, su riesgo y la gran variedad.

**7. 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.**

Se adjunta un enlace al github:

Notebook en python, requirements, cover, wiki and dataset