1. **Contexto. Por qué el sitio web elegido proporciona dicha información.**

En enero de 2020, el famoso Covid-19 llegó a España forzando la decisión de confinamiento para todas las comunidades desde mediados de marzo, cuando el número de contagios empezó a aumentar drásticamente. Debido a esta situación, la mayoría de las empresas se han visto afectadas por la imposibilidad de abrir sus oficinas, tiendas, o establecimientos en general dado que los empleados deben mantenerse en casa. En este sentido, ante la imposibilidad de muchas empresas de ofrecer sus servicios/productos (muchos de ellos presenciales), la economía ha recibido un importante impacto.

Este impacto se distribuye a lo largo de un continuo en el que influye directamente la capacidad de las empresas para continuar produciendo/ofreciendo/distribuyendo sus servicios/productos a distancia.

Si la empresa cotiza en bolsa, una manera de observar mejor el impacto de esta pandemia es estudiar las variaciones en el precio de las acciones. De esta forma, se pueden analizar (dado que ya ha pasado unas cuantas semanas desde el inicio del confinamiento), los cambios que se han producido, por ejemplo, con respecto al inicio del año o las mismas épocas de años anteriores.

En este sentido, el código extrae un dataset con información (en euros) sobre las cotizaciones en bolsa (apertura, máximo, mínimo, cierre y cierre ajustado) de la empresa Telefónica, desde el 2015-03-30 hasta la fecha actual 2020-03-28, utilizando las librerías: pandas\_datareader, argparse y datetime.

El sitio web elegido es yahoo por su confiabilidad y por facilitar el campo cierre ajustado.

1. **Definir un título para el dataset**

EVOLUCIÓN DE LAS COTIZACIONES DE TELEFÓNICA DURANTE LOS ÚLTIMOS 5 AÑOS

1. **Descripción del dataset.**

El dataset tiene como objetivo principal servir de base para comprender cuál, en términos de cantidad y dinero, ha sido el impacto que ha tenido la pandemia, concretamente en la empresa Telefónica.

Es por ello que contiene datos de cotizaciones de la empresa desde el 30 de marzo de 2015 hasta el 28 de marzo de 2020 (fecha actual del proyecto). El periodo abarca un periodo temporal de aproximadamente 5 años. Hay que tener en cuenta, sin embargo, que el número de registros (días) del dataset no se corresponde con 1.825 (el correspondiente a aproximadamente 5 años), dado que la bolsa no está activa durante los fines de semana y algunos festivos.

Como ventajas del programa en general, cabe destacar la flexibilidad que aporta a la extracción de información ya que permite consultar diferentes periodos de tiempo y diferentes empresas de interés. Además, está preparado para extraer la información y dejarla preparada para su uso sin procesos de limpieza posterior.

Por otro lado, si es cierto que la información que ofrece está limitada a estudiar un objetivo en particular (impacto en las cotizaciones de una empresa), y ofrece exclusivamente los campos descritos más abajo, sin posibilidad de ampliación. No limita, sin embargo, el análisis de otras cuestiones siempre que se puedan llevar a cabo a través de la información proporcionada en este dataset.

1. **Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente**

Imagen que contiene captura de pantalla, gente, grupo, diferente

Descripción generada automáticamente

1. **Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.**

El dataset está compuesto por 1280 registros (filas) con los siguientes campos (columnas):

- Date: dato de tipo fecha que indica el día en la que fueron recogidos los demás datos.

- High: dato de tipo numérico que indica la cotización máxima que alcanzaron las acciones de Telefónica en la fecha indicada.

- Low: dato de tipo numérico que indica la cotización mínima que alcanzaron las acciones de Telefónica en la fecha indicada.

- Open: dato de tipo numérico que indica la cotización de apertura (inicial) de las acciones de Telefónica en la fecha indicada.

- Close: dato de tipo numérico que indica la cotización de cierre (final) de las acciones de Telefónica en la fecha indicada.

- Volume: dato de tipo numérico que indica el volumen de operaciones que se llevaron a cabo con las acciones de Telefónica en la fecha indicada.

- Adj Close: dato de tipo numérico que indica la cotización final de las acciones de Telefónica en la fecha indicada. El cierre ajustado es el precio de cierre después de los ajustes para todas las distribuciones de splits y dividendos aplicables. Los datos se ajustan utilizando los multiplicadores de splits y dividendos correspondientes, de conformidad con las Center for Research in Security Prices (CRSP).

(Fuente: https://es.ayuda.yahoo.com/kb/%C2%BFQu%C3%A9-es-el-cierre-ajustado-sln28256.html visitada 29/03/2020)

El periodo recogido incluye aproximadamente los 5 años anteriores a la fecha actual: desde el 30 de marzo de 2015 hasta el 28 de marzo de 2020.

Durante el trabajo de extracción se ha optado por el uso de la librería pandas\_datareader dado que ofrece claras ventajas como: sencillez del código y, por tanto, mejor legibilidad, además de una extracción más precisa a través del uso de parámetros concretos de símbolo (empresa en bolsa) y fechas específicas, así como parámetros por defecto.

También se ha hecho uso de la librería argparse, con el fin de posibilitar la interacción con personas para que estas puedan extraer los datos requeridos en cada momento.

Finalmente se ha incorporado además la librería datetime para el manejo de las fechas de manera adecuada.

1. **Agradecimientos**

Los datos han sido recogidos de la base de datos online de yahoo finance. Utilizando para ello el lenguaje de programación Python y la librería: pandas\_datareader.

1. **Inspiración**

Dada la situación actual en la que se ve sumergida el mundo entero por el COVID-19, se ha dispuesto el siguiente dataset para evaluar el impacto económico que ha tenido en la empresa Telefónica. Esto se ha llevado a cabo seleccionando el periodo de los 5 años anteriores.

Este dataset puede ser utilizado como base para elaborar modelos predictivos, con la intención de predecir cotizaciones futuras.

1. **Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:**

Se han de tener en cuenta las siguientes consideraciones:

* La información extraída de la página de Yahoo Finance u otras entidades asociadas, no puede utilizarse con fines comerciales por ninguno de sus usuarios.
* Zenodo, la página de destino para el dataset de esta práctica está certificada por OPENAire, lo que lo convierte en Open Data Source.
* Teniendo esto en cuenta, la publicación del dataset en Zenodo no se establecería como una actividad comercial.
* En base a lo anterior se establece que la posible licencia para aplicar es: CC-BY-NC-SA 4.0, que permite la reproducción y distribución del dataset, pero prohíbe su comercialización comercialización, a la vez que obliga a replicar esta licencia en todas sus redistribuciones.

1. **Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.**

**# Import libraries**

import pandas\_datareader

import datetime

import argparse

import pandas as pd

import plotly.graph\_objects as go

**# Parse command line arguments**

parser = argparse.ArgumentParser()

parser.add\_argument("--ticker", help="Enter corporation ticker symbol",

default = 'TEF.MC')

parser.add\_argument("--startDate", help="Enter start date of interval YY-MM-DD",

default = str(datetime.date.today()-datetime.timedelta(days = 365)))

parser.add\_argument("--endDate", help="Enter end date of interval YY-MM-DD",

default = str(datetime.date.today()))

args = parser.parse\_args()

**# Variables**

ticker = args.ticker

startDate = datetime.datetime.strptime(str(args.startDate), "%Y-%m-%d")

endDate = datetime.datetime.strptime(str(args.endDate),"%Y-%m-%d")

**# Read data from yahoo**

df = pandas\_datareader.DataReader(ticker, 'yahoo', startDate, endDate)

**# Get Date into a column value**

df.reset\_index(inplace=True,drop=False)

**# Print dataframe head**

print(df.head)

**# Export dataframe to csv**

df.to\_csv(ticker + '.csv')

**# This code is adapted from https://plot.ly/python/ohlc-charts/**

**# OHLC Chart without Rangeslider**

fig = go.Figure(data=go.Ohlc(x=df['Date'],

open=df['Open'],

high=df['High'],

low=df['Low'],

close=df['Close']))

fig.update(layout\_xaxis\_rangeslider\_visible=False)

**# Adding customized text**

fig.update\_layout(

title=ticker,

yaxis\_title='Cotización')

**# Show OHLC Chart**

fig.show()

**# Save OHLC Chart html**

fig.write\_html("file.html")

1. **Dataset. Publicación del dataset en formato CSV en Zenodo con una pequeña descripción.**