

Análisis Del Mercado De Acciones Y Predicción Utilizando LSTM

Introducción

Una serie de tiempo es una secuencia de datos u observaciones, medidos en determinados momentos y ordenados cronológicamente. Visualmente, es una curva que evoluciona en el tiempo.

Una serie de tiempo es un conjunto de observaciones sobre los valores que toma una variable (cuantitativa) a través del tiempo. Por tanto, una serie de tiempo es una forma estructurada de representar datos.

Los datos se pueden comportar de diferentes maneras a través del tiempo: puede que se presente una tendencia, estacionalidad o simplemente no presenten una forma definida.

Representar los datos del negocio como series de tiempo suele ayudar a las empresas a visualizar la actividad del negocio. A su vez, usualmente las series de tiempo se utilizan para predecir el comportamiento futuro de la variable medida.

La predicción de series de tiempo significa que se extienden los valores históricos de la serie al futuro, donde aún no se han hecho mediciones.

Para llevar a cabo el pronóstico se definen dos variables principales: cantidad de períodos y horizonte de predicción. La cantidad de períodos representa el nivel de agregación de los datos. Usualmente los datos se encuentran por meses, semanas o días, permitiendo obtener el grado necesario de desagregación para sacar correctas conclusiones. Por su parte, el horizonte de planificación representa el número de períodos futuros o alcance a pronosticar.

Todo pronóstico tiene asociado un alcance, pudiendo ser de corto, mediano o largo plazo. A modo general se presenta la siguiente tabla resumen, no obstante el alcance varía de acuerdo a la industria.

En esta tarea se realiza un análisis de los datos del mercado de valores, particularmente algunas acciones de tecnología, en este caso de las empresas: Oracle, Google, Microsoft y Tesla. La información de stock de estas compañías proviene de yahoo finance, importando la librería `zfinance`. Se realizan diferentes gráficos para visualizar estos datos en diferentes perspectivas y se analiza el riesgo de un stock, basado en su historial de rendimiento anterior. Finalmente se realiza una predicción de los precios futuros de las acciones a través de un método de memoria a corto plazo (LSTM), en este caso para la compañía Oracle.

A lo largo de esta tarea, se plantean seis preguntas a resolver:

- ¿Cuál fue el cambio en el precio de las acciones a lo largo del tiempo?
- ¿Cuál fue el rendimiento diario de las acciones en promedio?

- ¿Cuál fue el promedio móvil de las diversas acciones?
- ¿Cuál fue la correlación entre las distintas poblaciones?
- ¿Cuánto valor ponemos en riesgo al invertir en una acción en particular?
- ¿Cómo podemos intentar predecir el comportamiento futuro de las acciones?

Datos

Los registros cuentan con 8 variables. A continuación se mencionan cada una de las 8 con una breve descripción:

Atributo	Descripción
Date	Fecha en días hábiles.
Close	Valor del stock al cierre de la bolsa de valores.
Open	Valor del stock a la apertura de la bolsa de valores.
High	Valor máximo del stock durante el día.
Low	Valor mínimo del stock durante el día.
Volume	Número de acciones negociadas en el stock en un periodo de tiempo específico.
Adj Close	El precio de cierre ajustado.
Company_Name	Nombre de la compañía.

Metodología, Modelo y Resultados

¿Cuál fue el cambio en el precio de las acciones a lo largo del tiempo?

Para los siguientes gráficos, se consideraron el precio de las acciones, de las compañías mencionadas anteriormente, desde el 1 de junio de 2021 hasta el 27 de mayo de 2022.

En este gráfico, se observa una serie de tiempo para cada compañía, tomando en cuenta el precio de cierre ajustado.

El precio de cierre ajustado se modifica en una acción para reflejar con precisión el valor de esta población después de contabilización de sus acciones corporativas. Se considera que es el precio real de esa acción y se usa a menudo al examinar los rendimientos históricos o al realizar un análisis detallado de dichos rendimientos.

Se puede observar que en este rango de fechas, las series de tiempo no son estacionarias, es decir, son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

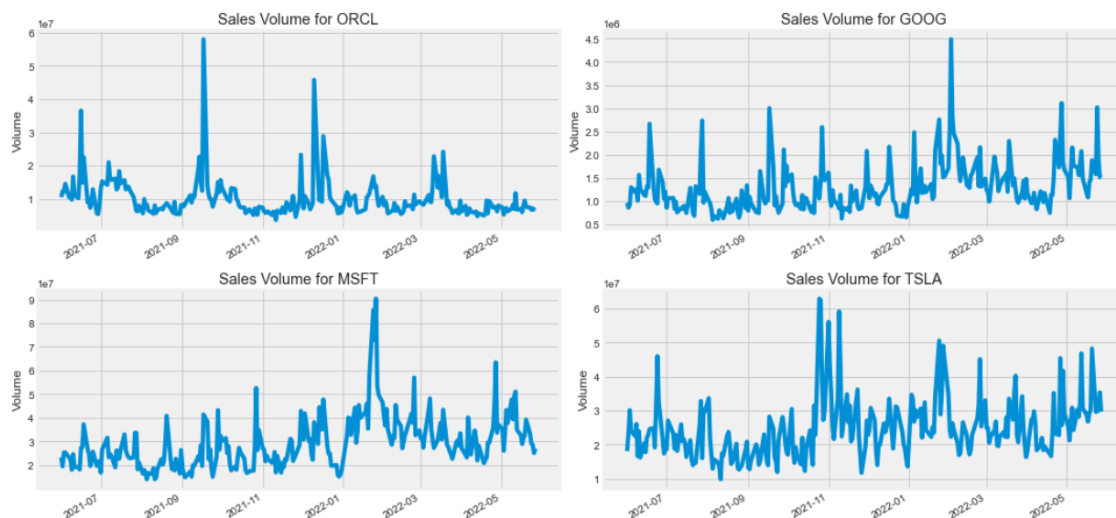
En las cuatro se observa una tendencia decreciente en este último mes.



A diferencia del gráfico anterior, en el siguiente gráfico, se observa una serie de tiempo para cada compañía, tomando en cuenta el volumen.

Los cambios en el volumen de una acción son importantes, esto debido a que un incremento del volumen nos dirá que hay más agentes en el mercado interesados en negociar la acción, o bien hay agentes que están operando por cuantías mayores.

No se observa claramente alguna tendencia creciente o decreciente por lo que se consideran series estacionarias, ya que cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.



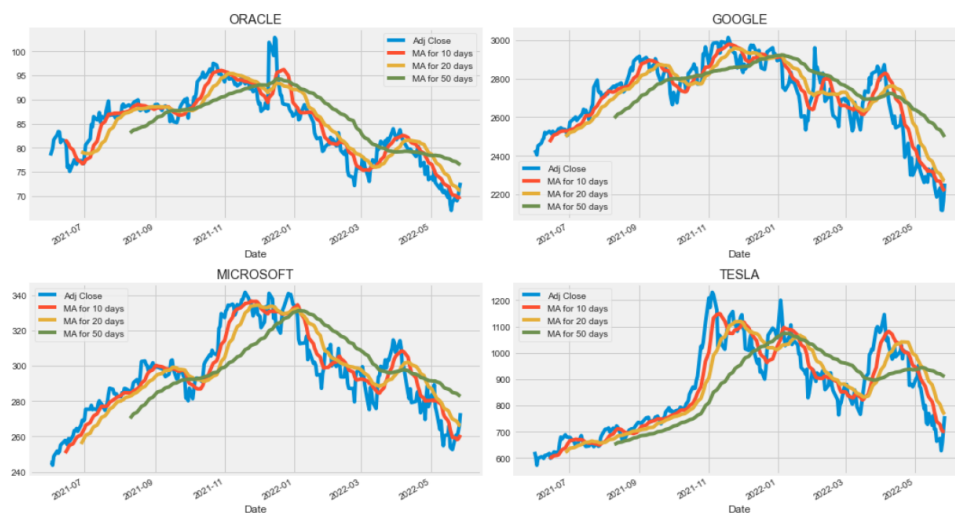
¿Cuál fue el promedio móvil de las diversas acciones?

Para los siguientes gráficos, se consideraron el precio de las acciones, de las compañías mencionadas anteriormente, desde el 1 de junio de 2021 hasta el 27 de mayo de 2022.

En este gráfico, se observa una serie de tiempo para cada compañía, tomando en cuenta el precio de cierre ajustado y se observa el MA(Moving Average) de 10, 20 y 50 días.

En estadística, un promedio móvil (promedio móvil o promedio corriente) es un cálculo para analizar puntos de datos mediante la creación de una serie de promedios de diferentes subconjuntos del conjunto de datos completo. También se denomina media móvil (MM) o media rodante y es un tipo de filtro de respuesta de impulso finito. Las variaciones incluyen: formas simples, acumulativas o ponderadas. Dada una serie de números y un tamaño de subconjunto fijo, el primer elemento de la media móvil se obtiene tomando el promedio del subconjunto fijo inicial de la serie de números. A continuación, el subconjunto se modifica por "desplazamiento hacia adelante", es decir, excluyendo el primer número de la serie e incluyendo el siguiente valor en el subconjunto. Un promedio móvil se usa comúnmente con datos de series temporales para suavizar las fluctuaciones a corto plazo y resaltar las tendencias o ciclos a largo plazo. El umbral entre corto y largo plazo depende de la aplicación, y los parámetros de la media móvil se establecerán en consecuencia. Por ejemplo, se utiliza a menudo en el análisis técnico de datos financieros, como precios de las acciones, rendimientos o volúmenes de negociación. También se utiliza en economía para examinar el producto interno bruto, el empleo u otras series cronológicas macroeconómicas.

En las tres series de tiempo con el MA se observan más suavizados los datos, y se observa para las empresas Oracle y Microsoft como a principios del año 2022 va en tendencia decreciente. Para la empresa Tesla, se observan picos crecientes y decrecientes después de noviembre 2021, cuando antes de este mes, iba en tendencia creciente. Y para Google, caso parecido a Tesla, solo que a partir de septiembre 2021.



¿Cuál fue el rendimiento diario de las existencias en promedio?

Para los siguientes gráficos, se consideraron el precio de las acciones, de las compañías mencionadas anteriormente, desde el 1 de junio de 2021 hasta el 27 de mayo de 2022.

En este gráfico, se observa una serie de tiempo para cada compañía, tomando en cuenta el precio de cierre ajustado y se observa el porcentaje de retorno diario.

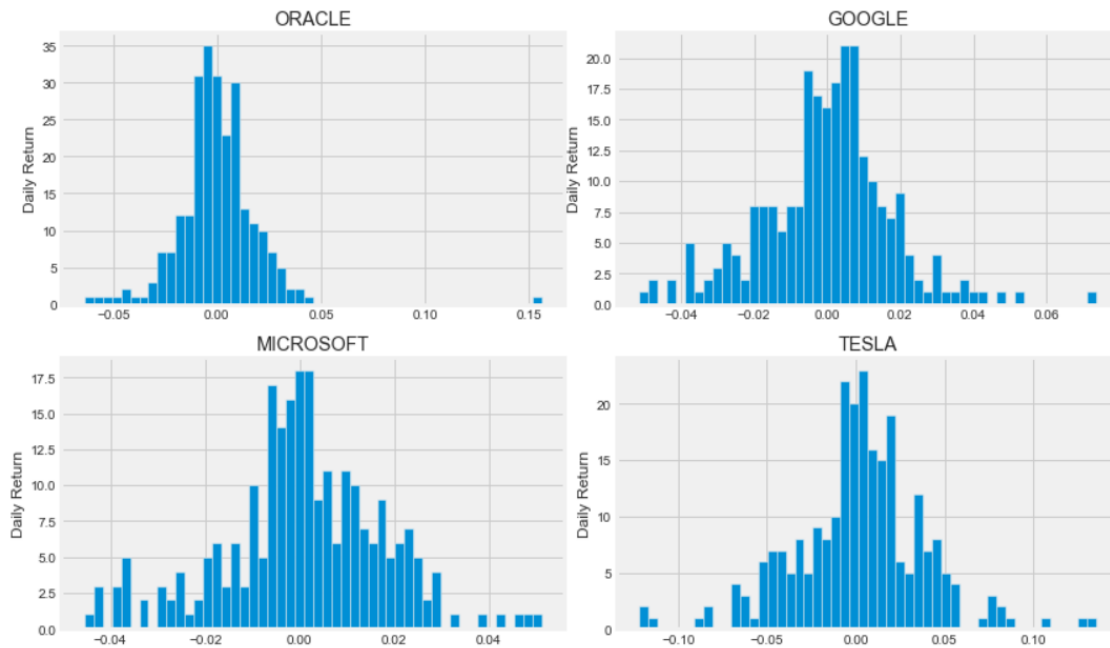
El rendimiento diario de una acción se utiliza para medir el funcionamiento diario de la acción, es el precio de la acción en el cierre de hoy comparado al precio de la misma acción en el cierre de ayer. Rendimiento diario positivo significa apreciación en el precio de las acciones en la comparación diaria.

El porcentaje de retorno diario es el cambio porcentual entre el elemento actual y un elemento anterior. Se calcula el cambio porcentual de la fila inmediatamente anterior de forma predeterminada. Esto es útil para comparar el porcentaje de cambio en una serie temporal de elementos.

En todas las series se observa una serie estacionaria, sin clara tendencia creciente o decreciente.



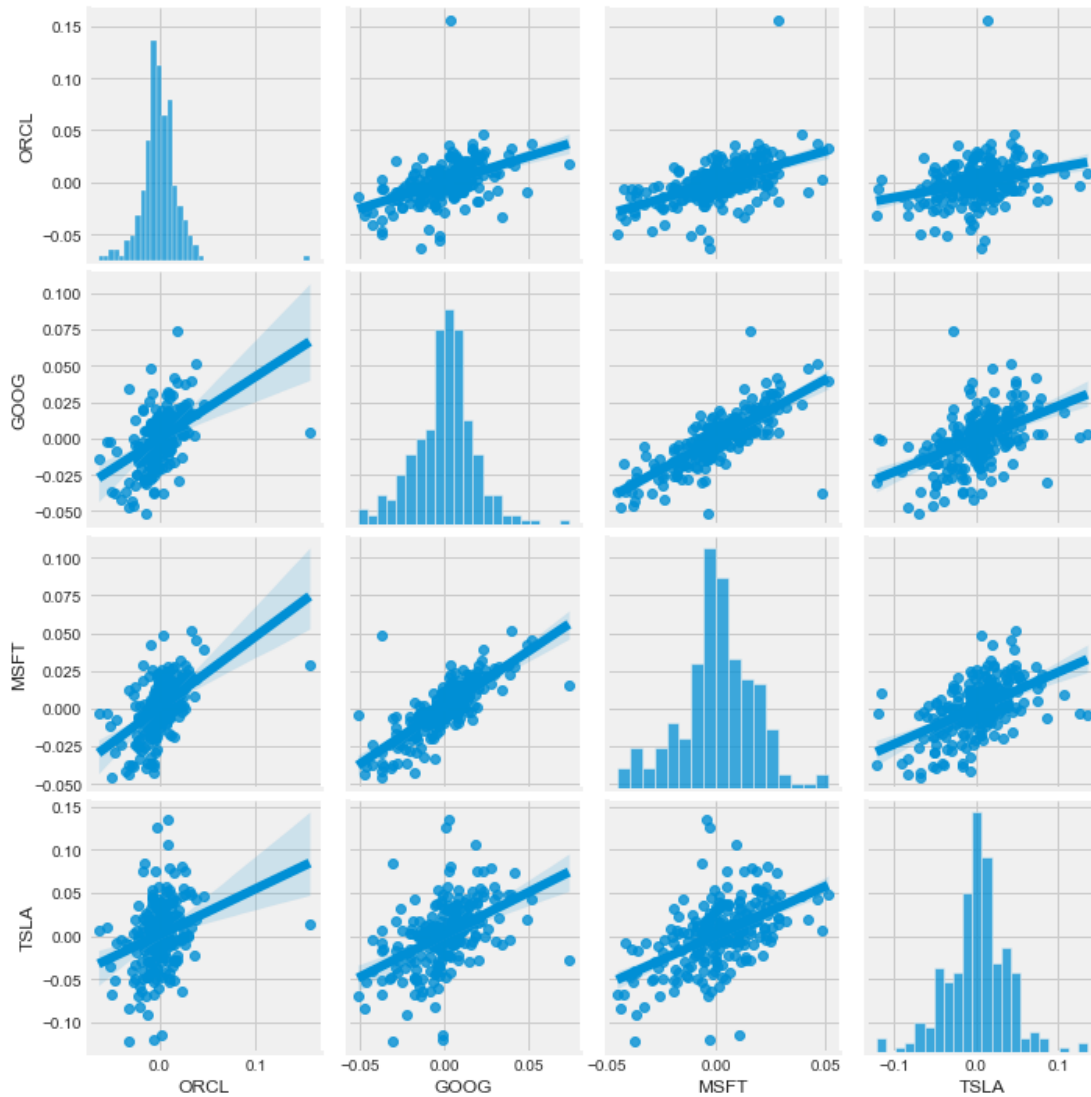
En este siguiente gráfico se observa el rendimiento diario promedio usando un histograma para cada compañía. Se puede observar mayor rendimiento en la compañía Tesla, ya que se observan barras altas después de 0. A diferencia de la compañía Oracle, no se observan tantas barras altas después de 0, además de mostrar un gran sesgo a la derecha. El resto de las gráficas aparentan una distribución normal.



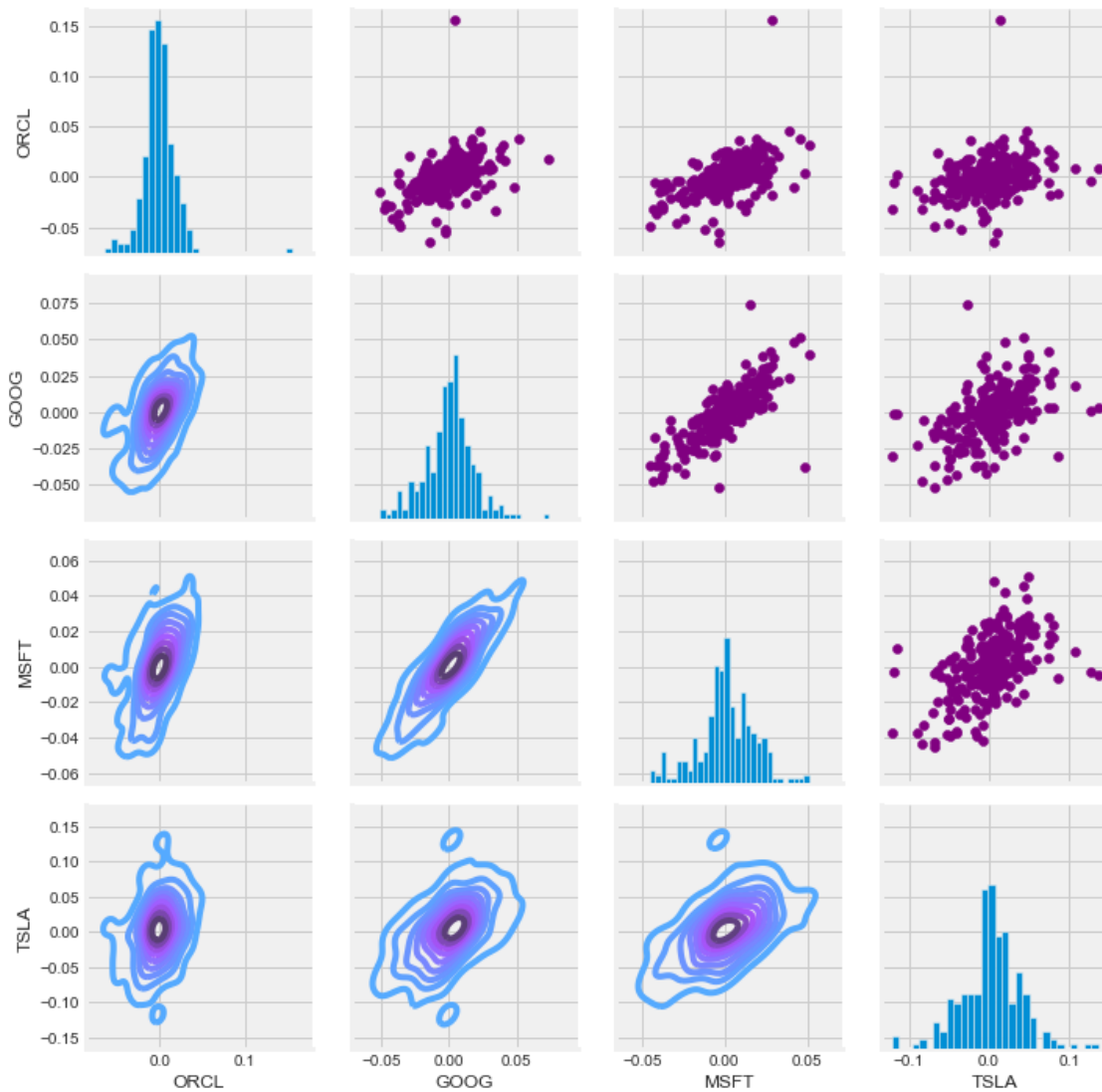
¿Cuál fue la correlación entre los diferentes precios de cierre de las acciones?

Para los siguientes gráficos, se consideraron el precio de las acciones, de las compañías mencionadas anteriormente, desde el 1 de junio de 2021 hasta el 27 de mayo de 2022.

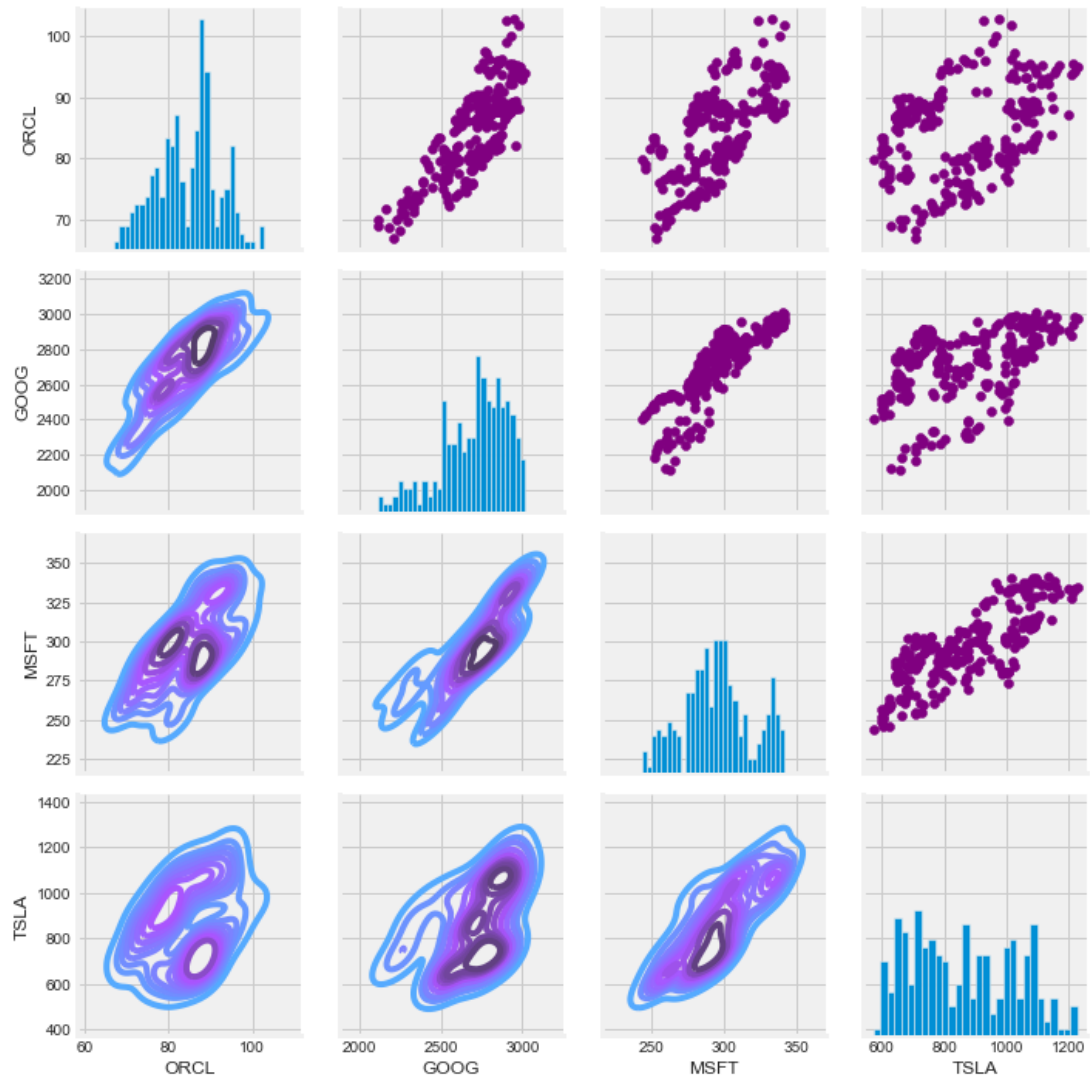
En este gráfico se observa la correlación del rendimiento diario(calculado en la pregunta 3) de cada compañía con el resto. Se puede observar una fuerte correlación entre las compañías Microsoft y Google. A diferencia de Tesla y Google, Tesla y Microsoft, una baja correlación.



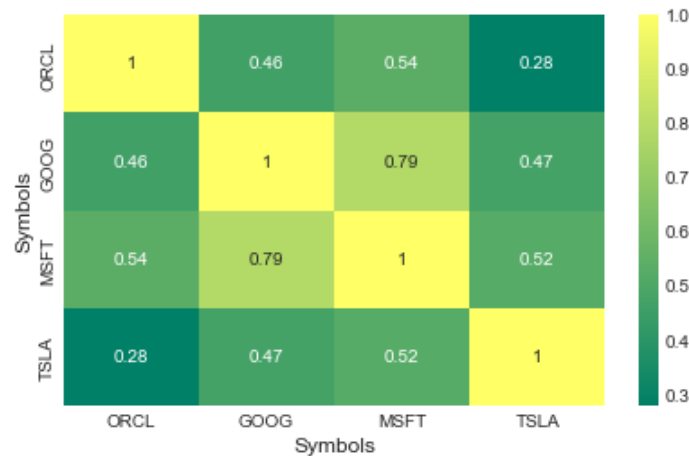
En este siguiente gráfico, se pueden observar las mismas correlaciones, agregando las Bivariate Seaborn Kdeplot.



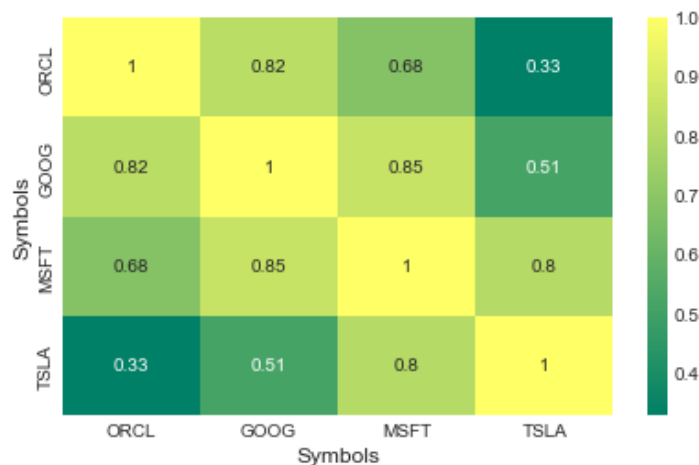
En este gráfico se observa la correlación del precio de cierre ajustado de cada compañía con el resto. Se puede observar una fuerte correlación entre las compañías Microsoft y Google. A diferencia de Tesla y Oracle donde existe una baja correlación.



Aquí se observa mejor la correlación del rendimiento diario de cada compañía con el resto. Se puede observar una fuerte correlación entre las compañías Microsoft y Google. A diferencia de Tesla y Oracle donde existe una baja correlación.



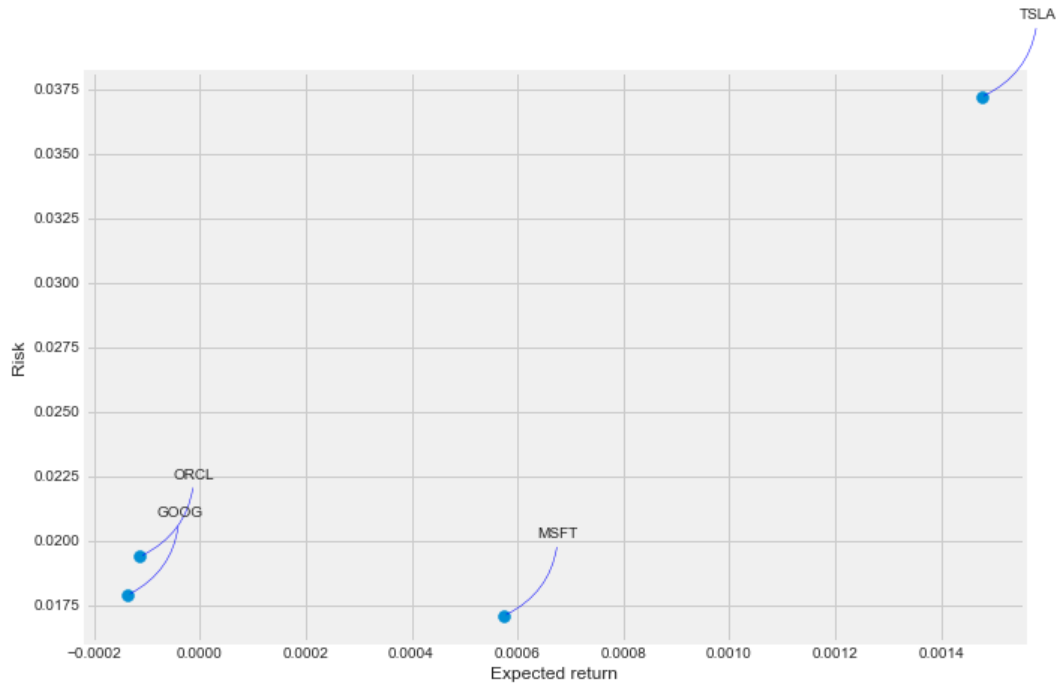
Aquí se observa mejor la correlación del precio de cierre ajustado de cada compañía con el resto. Se puede observar una fuerte correlación entre las compañías Microsoft y Google. A diferencia de Tesla y Oracle donde existe una baja correlación.



¿Cuánto valor ponemos en riesgo al invertir en una acción en particular?

Hay muchas maneras en que podemos cuantificar el riesgo, una de las formas más básicas de usar la información que hemos recopilado sobre los rendimientos porcentuales diarios es comparando el rendimiento esperado con la desviación estándar de los rendimientos diarios.

Al observar la siguiente gráfica, se concluye que no hay mucho riesgo en invertir en Microsoft, ya que el riesgo es menor y el rendimiento esperado no es tan bajo. A diferencia de Tesla, en el cual, el riesgo es mucho mayor aunque el rendimiento esperado es mucho mayor también.



Predicción del precio de cierre de las acciones de Oracle

Para la siguiente serie de tiempo, se consideró el precio de las acciones de cierre de la compañía Oracle, elegí esta en particular, porque me pareció interesante realizar el análisis para la compañía en la cual laboro desde hace 7 años. El rango de fechas consta desde el 1 de enero de 2012 hasta el 27 de mayo de 2022.

Se observa una tendencia creciente, sin embargo, hay un cambio drástico a partir de 2022, se observa claramente una tendencia decreciente. También se observa un alto crecimiento a partir del 2020, por lo cual considero que la pandemia del covid tuvo alguna relación en el precio de cierre de las acciones. Y el hecho de que van bajando a partir del año 2022, debido a que la pandemia ha bajado.



Uno de los modelos más avanzados para pronosticar series de tiempo es la Red Neuronal de Memoria de Corto Plazo Larga (LSTM). Según Korstanje en su libro, Predicción avanzada con Python:

”La celda LSTM añade memoria a largo plazo de una manera aún más performante porque permite que se aprendan aún más parámetros. Esto hace que sea la más poderosa [Red Neuronal Recurrente] para hacer pronósticos, especialmente cuando tiene una tendencia a largo plazo en sus datos. Los LSTM son uno de los modelos de vanguardia para la predicción en este momento.” (2021).

Esa es la buena noticia. La mala noticia es, y lo sabes si has trabajado con el concepto en TensorFlow, diseñar e implementar un modelo LSTM útil no siempre es sencillo.

Para realizar la predicción, se utilizó este modelo LSTM, de la librería tensorflow.keras, de un total de 2619 registros, se utilizó para el entrenamiento el 95 %, dando el total de 2489 registros.

Después todos los registros, específicamente el precio de cierre, se escalaron con la función MinMaxScaler de la librería sklearn. Este estimador escala y traduce cada característica individualmente de tal manera que se encuentre en el rango dado en el conjunto de entrenamiento, en nuestro caso entre 0 y 1.

Luego, se escalaron los registros de entrenamiento y se realizaron conjuntos de 60 registros para el `x_train` y el valor 61 como el `y_train`.

Para el modelo LSTM, se utilizaron 128 neuronas en la primera capa, 64 en la segunda, 25 en la tercera y 1 en la capa de salida, ya que se espera un valor individual por día para predecir.

Para compilar el modelo se utilizó el optimizador Adam.

Además, en la compilación, se agregó una función de pérdida.

La función de pérdida (L), nos dice qué tan bien está funcionando nuestra red neuronal. Hay muchas funciones de costo diferentes. Se utilizó el error cuadrado medio como una función de costo, que se puede definir como la media de la diferencia cuadrada entre la salida real y la salida prevista:

$$J = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

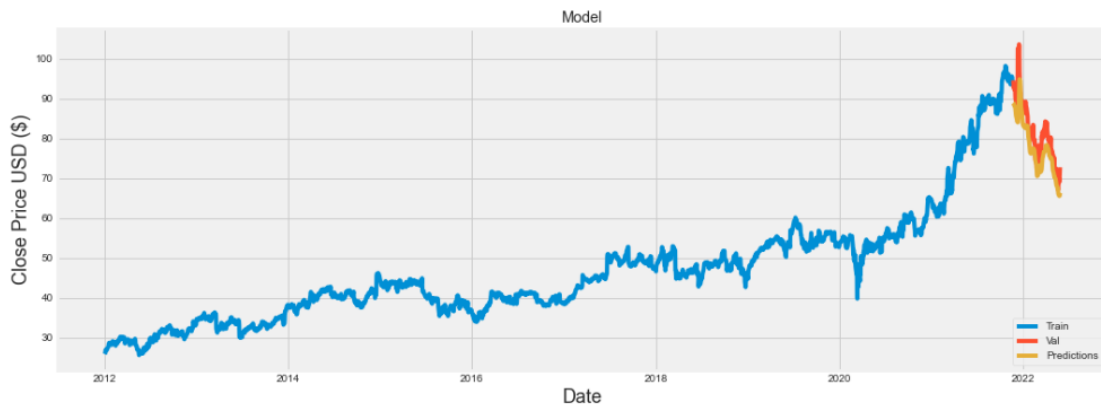
Para el entrenamiento, se definió una época en tamaño de batch 1.

Para las pruebas utilizando el modelo, se seleccionó el resto 5 % del total de los registros, es decir, las x_{test} y y_{test} .

Para la predicción se realizó la inversa de los valores escalados para tomar los valores cercanos a los reales.

También se obtuvo el valor del error cuadrado medio, dando un total de 4.846418427666557, lo cual considero bajo, podría mejorar, tal vez utilizando un mejor modelo o realizando cambios en los valores de la cantidad de neuronas, capas del modelo y el optimizador en la compilación.

En la gráfica de abajo, se observa la serie en color azul, los registros de entrenamiento, la serie en color rojo, los registros de prueba y la serie en color amarillo, las predicciones. Se nota que no hay mucha diferencia entre las dos últimas, por lo que los resultados fueron favorables.



Conclusiones

Es importante realizar este tipo de análisis en el mercado de acciones, debido a que hay muchos inversionistas interesados en comprar y/o vender acciones, sobretodo, en las compañías en su mayoría de tecnología. También el hecho de conocer el riesgo al invertir, la correlación entre varias compañías, el rendimiento y el volumen de las acciones, tomando en cuenta datos históricos.

Para la serie de tiempo en los precios de cierre de las acciones de Oracle, se podría considerar, el hecho de que en la clara tendencia creciente a partir de 2020 hasta 2022, la crisis económica derivada del COVID-19 tomó por sorpresa a la mayoría de compañías a nivel mundial. Muchas de estas experimentaron una considerable caída en sus ingresos por las restricciones

gubernamentales y por los cambios en los patrones de los consumidores. No obstante, de acuerdo con Financial Times, durante esta coyuntura también existieron un grupo de empresas, vinculadas al sector de tecnología y de comercio electrónico, que aumentaron el valor de sus acciones, como en este caso la compañía Oracle.

La mayoría de estas empresas se diferencian de su competencia por tener un alto grado de digitalización, innovación, agilidad y resiliencia. Es importante indicar que la mayoría de estos factores se fueron construyendo desde antes de la pandemia, razón por la cuál pudieron responder de mejor manera ante esta.

No obstante, el coronavirus ha logrado impulsar la digitalización e incluso cambios en el modelo de negocios.

Para la predicción, los valores históricos ayudaron mucho para realizar este tipo de pronósticos en series no estacionarias y aunque las predicciones fueron muy cercanas a los valores reales, para los valores futuros no hay mucha certeza de que vayan a ser reales.

Referencias

- [1] GITHUB (2022). *ProcesamientoDeDatos* [Online]. Available: https://github.com/vickyrnz24/ProcesamientoDeDatos/blob/main/Procesamiento_de_datos_---Tarea_3.pdf
- [2] GITHUB (2022). *ProcesamientoDeDatos* [Online]. Available: <https://github.com/vickyrnz24/ProcesamientoDeDatos/blob/main/Tarea3.ipynb>
- [3] SAYAH, F. (2022). *Stock Market Analysis + Prediction using LSTM* [Online]. Available: <https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm>