



Análisis del precio histórico de las acciones de McDonalds

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

MARTÍNEZ RUGERIO LUCÍA

INTRODUCCIÓN A CIENCIA DE DATOS

M.C. JAIME ALEJANDRO ROMERO SIERRA

25 DE NOVIEMBRE DE 2024

Introducción

Objetivo del Proyecto

El propósito de este proyecto es analizar las acciones de McDonalds para identificar patrones que describen su comportamiento, además se pretende encontrar los periodos en que las acciones de McDonalds tienen su precio mínimo y máximo, para así guiar a quien esté interesado en las acciones de McDonalds.

Justificación y contexto: ¿Por qué es importante resolver o estudiar esta problemática?

Para los accionistas es importante reconocer cuándo es bueno comprar o vender una acción. Desafortunadamente, si no se tiene un conocimiento previo sobre el comportamiento de las acciones, a los accionistas les puede suponer una pérdida, ya que las acciones pueden verse afectadas de último momento por múltiples factores.

Por lo tanto, conocer sobre sus patrones de comportamiento ayuda a los accionistas a anticipar sus movimientos y tomar decisiones informadas que conlleven un menor riesgo. Este proyecto supone una guía para aquellos que estén interesados en la compra de acciones de McDonalds.

Fuentes de datos:

- Origen: El dataset consiste de los datos históricos del precio de las acciones de McDonalds recolectadas de Yahoo Finance
- Cantidad de Datos: El dataset original consiste de 7 columnas que comprenden Fecha, precio de apertura, Precio más alto, Precio más bajo, Precio al cierre, Precio de cierre ajustado y el Volumen. Y 14651 renglones de información.
- Principales características: Contiene los precios históricos de las acciones de McDonalds. Empezando en julio de 1966 hasta septiembre de 2024.

Metodología

Proceso de limpieza de datos:

Dado a que se retomó la base de datos original, no se contó con datos ausentes, ni duplicados, se tomó la decisión de dejar los valores atípicos.

Análisis Exploratorio de Datos (EDA):

1. Descripción General de los Datos

Visión General

Total de registros: 14652

Date: Los datos van desde 05-07-1966 hasta 19-09-2024

Open: El precio en que abren las acciones van desde 0 hasta 300.95 dependiendo el día.

Low y High: El mínimo, máximo y promedio de estas dos variables tienen valores muy similares.

Adj Close: Cabe recalcar que este valor es importante porque proporciona un reflejo más preciso del valor de la acción, permite a los inversores comparar con mayor precisión el rendimiento de una acción a lo largo del tiempo.

Volume: Este es el que contuvo más valores atípicos, después de eliminarlos su valor máximo fue de 86,981,300 y mínimo de 48,600

Tipos de Variables:

```
df.dtypes
✓ 0.0s
Date      object
Open      float64
High      float64
Low       float64
Close     float64
Adj Close float64
Volume    int64
dtype: object
```

Date: La fecha de negociación, formato de fecha.

Open: El precio al que abrió la acción en un día determinado, formato float.

High: El precio más alto de la acción durante la sesión de negociación, formato float.

Low: El precio más bajo de la acción durante la sesión de negociación, formato float.

Close: El precio de la acción al cierre del mercado, formato float.

Adj Close: El precio de cierre ajustado por dividendos y divisiones de acciones, formato float.

Volume: El número de acciones negociadas durante el día, formato integer.

Resumen Estadístico:

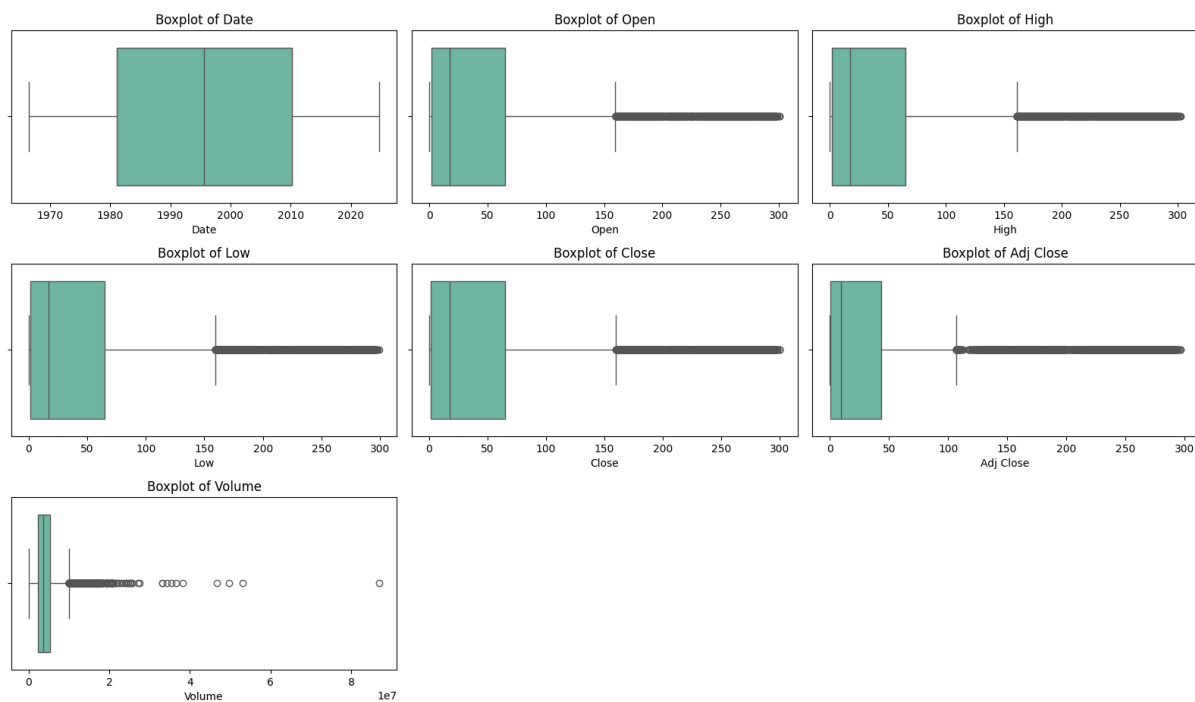
```
df.describe()
```

✓ 0.1s

	Open	High	Low	Close	Adj Close	Volume
count	14652.000000	14652.000000	14652.000000	14652.000000	14652.000000	1.465200e+04
mean	50.710007	51.143266	50.329801	50.746652	41.162607	4.270940e+06
std	73.844970	74.349949	73.292929	73.833827	69.263743	3.102599e+06
min	0.000000	0.156379	0.152263	0.154321	0.066407	4.860000e+04
25%	1.586420	1.607253	1.567901	1.589506	0.702369	2.357100e+06
50%	17.219999	17.474999	17.056250	17.265000	9.334254	3.588300e+06
75%	64.832502	65.440001	64.615000	64.980001	43.152880	5.393350e+06
max	300.950012	302.390015	299.309998	300.529999	296.690002	8.698130e+07

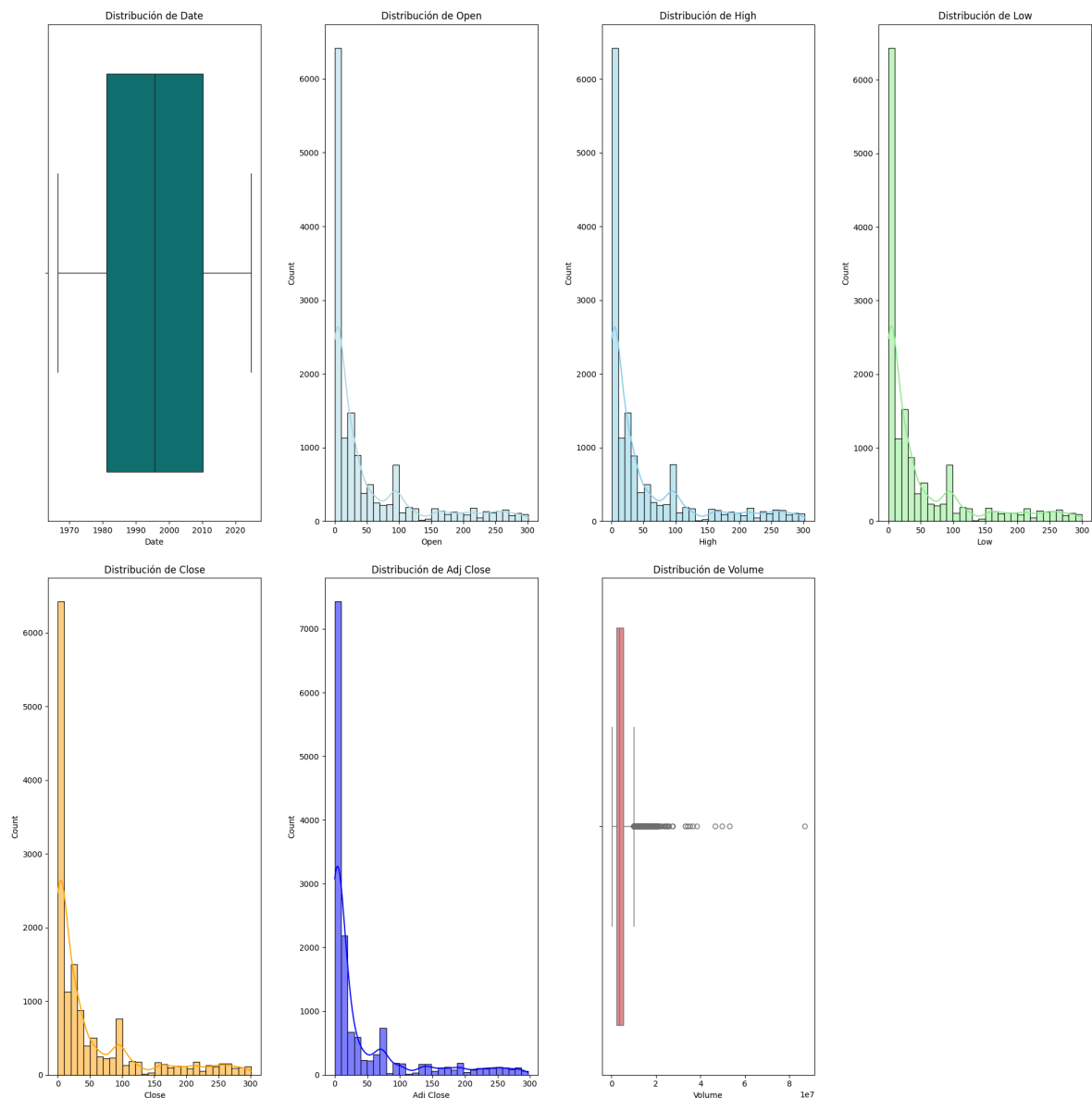
2. Visualización y Distribución de Variables Individuales

Variables numéricas:



Estos boxplots muestran que las variables de ‘Volumen’ y ‘Adj Close’ cuentan con más outliers. Esto se debe a que ambas variables son influenciadas por cambios repentinos del mercado y cambios de la empresa. Por lo tanto los outliers de ‘Volume’ y ‘Adj Close’ nos brindan una visión de eventos que han afectado al mercado entre los años de 1966 y 2024.

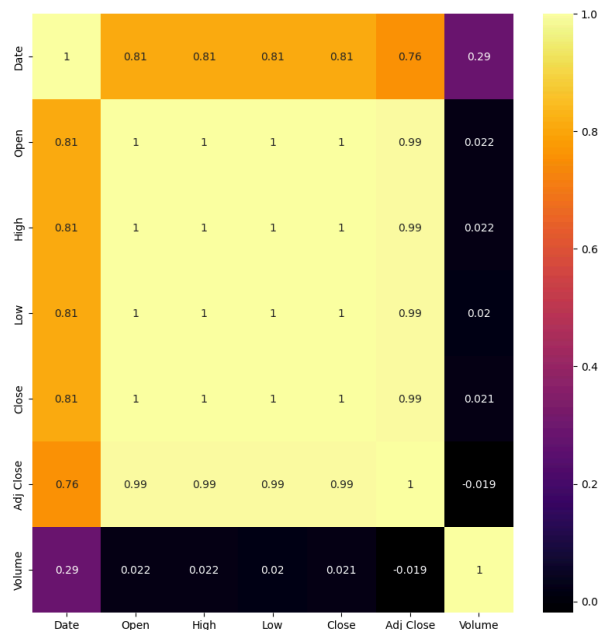
Variables Categóricas:



Aquí podemos observar como las variables se comportan de manera similar y en el caso de ‘Volume’ vemos que tiene valores atípicos y donde se encuentra el 50% de los datos.

3. Correlación entre Variables

Matriz de Correlación:



Debido a que son acciones, podemos observar que están completamente relacionadas entre sí. La variable con menor relación es 'Volume', aunque sí se ve influenciado por las otras, su crecimiento también se ve influenciado por factores externos.

En el caso de 'Adj Close', se esperaría que esté directamente relacionado a las demás, ya que esta variable es vulnerable a las decisiones del mercado.

4. Correlación entre Variables

- **Identificación de Outliers:**

Para este proyecto no se eliminaron Outliers ya que representan eventos extraordinarios clave para analizar el comportamiento del mercado, por ello, eliminar estos valores significa perder tendencias fundamentales para el análisis más preciso.

- **Tratamiento de Outliers:** No se eliminaron Outliers.

5. Análisis de Valores Faltantes

```
df.isnull().sum()
✓ 0.0s
Date      0
Open      0
High      0
Low       0
Close     0
Adj Close 0
Volume    0
dtype: int64
```

- **Identificación de Datos Faltantes:** Dado a que se retomó la base de datos original, no hay datos faltantes.

- **Estrategia de Imputación o Eliminación:** No se requirió de ninguna estrategia de Imputación.

6. Relación entre variables Categóricas y Numéricas

- **Análisis Comparativo:** Ya que se trata de acciones, el dataset no cuenta con variables Categóricas.

7. Observaciones y Hallazgos Importantes

El día 5 de Octubre de 2006 se presentó la máxima negociación de acciones por día, con un total de 86,981,300 negociaciones. Al hacer una investigación más profunda, descubrí que en los años previos al 2006, el mercado inmobiliario en EE.UU presentó un auge y los bancos facilitaron el acceso a préstamos hipotecarios sin verificar la capacidad de pago de los prestatarios. Finalmente la burbuja inmobiliaria explotó y la bolsa de valores reflejó esta inestabilidad.

Índices como el de Dow Jones, encargado de medir el crecimiento económico de EE.UU. a través de 30 empresas cotizadas en la bolsa de valores de Nueva York, presentó una ligera caída, lo que significa un descenso en el valor total de esas 30 empresas, esto provocó desconfianza entre los inversionistas y un aumento en los volúmenes de negociación de acciones de las empresas que lo componen.

Modelo de Machine Learning

- Descripción del modelo

Modelo: Regresión polinomial

- Justificación

Las acciones sufren de variaciones que no se ajustan a un comportamiento lineal. Por lo tanto, se requiere un modelo más flexible que pueda adaptarse a los cambios. Entonces un modelo de regresión polinomial se ajusta a este dataset.

- Implementaciones y Entrenamiento

```
#ANUAL

# Dividir los datos en conjunto de entrenamiento y prueba
X = df[['Date']].values
y = df['Adj Close'].values # Variable dependiente
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)

# Crear el modelo de características polinómicas de grado 2
poly = PolynomialFeatures(degree=6)
X_train_poly = poly.fit_transform(X_train)
X_test_poly = poly.transform(X_test)

# Crear el modelo de regresión lineal
model = LinearRegression()
model.fit(X_train_poly, y_train)

# Predecir con el conjunto de prueba
y_pred = model.predict(X_test_poly)

# Evaluación del modelo
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error (MSE): {mse}")
print(f"R^2 Score: {r2}")
```

LinearRegression()

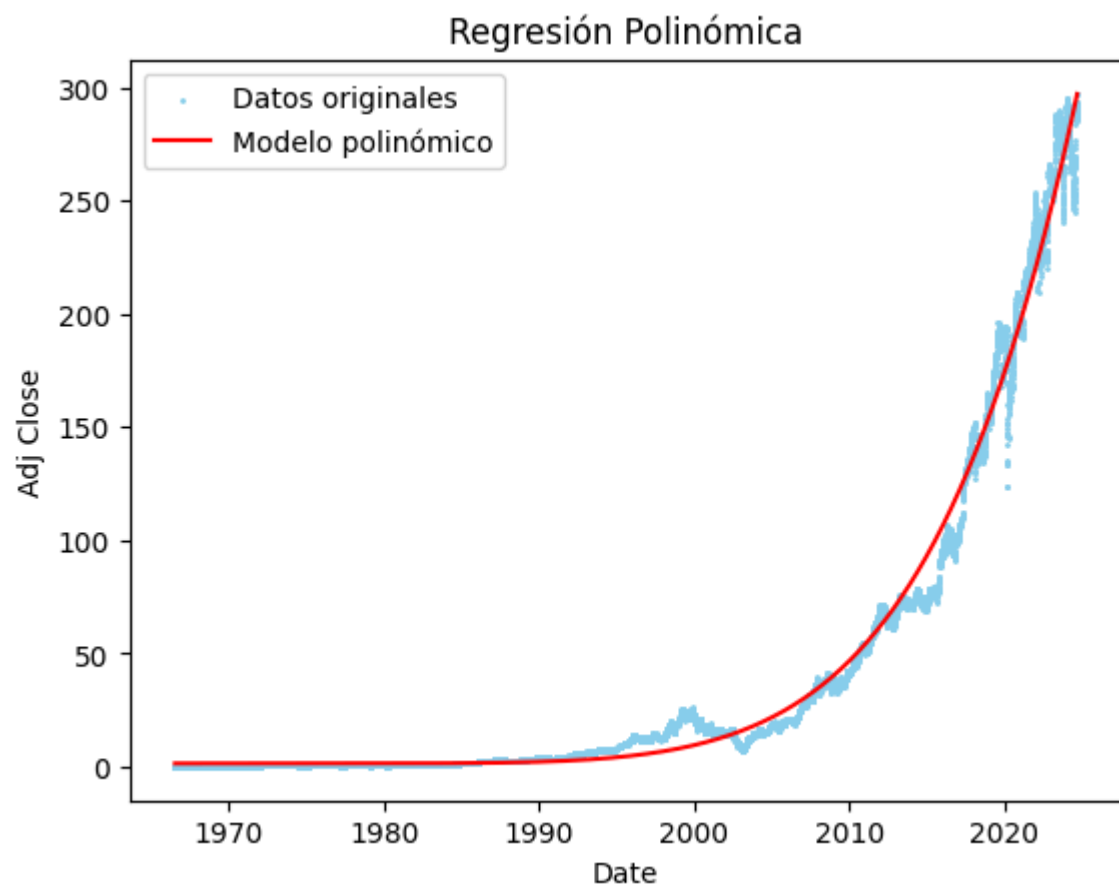
Mean Squared Error (MSE): 57.64124424046876
R^2 Score: 0.987801203238435

```
# Gráfica de los datos originales y el modelo ajustado
plt.scatter(X, y, s=1, color='skyblue', label='Datos originales')

# Ordenar X_test para la visualización
sorted_indices = np.argsort(X_test.flatten()) # Obtener índices de ordenamiento
X_test_sorted = X_test[sorted_indices]
y_pred_sorted = y_pred[sorted_indices]

plt.plot(X_test_sorted, y_pred_sorted, color='red', label='Modelo polinómico')
plt.title('Regresión Polinómica')
plt.xlabel('Date')
plt.ylabel('Adj Close')
plt.legend()
plt.show()
```


- **Resultados**



Dashboard

El dashboard ayuda a que sea más fácil para el cliente visualizar los hallazgos y comprender los resultados. El dashboard de este proyecto pretende ayudar al inversionista a tener una idea de los precios en promedio de las acciones de McDonalds.



Conclusiones y futuras líneas de trabajo

Hallé que las acciones no tienen un comportamiento lineal, sino que se ven afectadas por cambios repentinos en el mercado, y aunque ya existen índices que ayudan a las empresas a reducir el riesgo de pérdida vendiendo las acciones, a mi parecer, para lograr definir los periodos en que conviene comprar o vender acciones va a depender mucho del contexto económico del país y la situación en que se encuentra la empresa.

Sin lugar a duda este proyecto puede mejorar, necesita de un análisis más profundo y conocimiento previo del comportamiento que suelen tener las acciones en determinado tiempo. Además de continuar con los modelos de Machine Learning de acuerdo al día, mes y semana para así predecir con mayor exactitud con los datos que se tiene.

Referencias

Dow Jones sets record close again. 06/10/2006. ABC News Online. (2006, 10 junio).

<https://web.archive.org/web/20061024071855/http://www.abc.net.au/news/newsitems/200610/s1756737.htm>

Gonzalez, L. (2022, 7 septiembre). *Regresión polinomial – teoría. Aprende IA.*

<https://aprendeia.com/algorithmo-regresion-polinomial-machine-learning/>

IBM Cognos Analytics 11.1.x. (s. f.).

<https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=terms-r2>

Jornada, L. (2007, 2 abril). *La economía de Estados Unidos en 2006 (1/2) - La Jornada.*

<https://www.jornada.com.mx/2007/04/02/index.php?section=economia&article=020n1eco>

Precio de cierre (definición). (s. f.). IG.

<https://www.ig.com/latam/glosario-trading/definicion-de-precio-de-cierre#:~:text=El%20precio%20de%20cierre%20ajustado,ajustar%C3%A1%20a%2020%2C90%24.>

Time series / date functionality — pandas 2.2.3 documentation. (s. f.).

https://pandas.pydata.org/docs/user_guide/timeseries.html

Top 10 fast food giants: Stock price dataset(2024). (2024, 20 septiembre). Kaggle.

<https://www.kaggle.com/datasets/nguyentienhhan/stock-prices-of-the-10-largest-fast-food-companies?select=MCD.csv>

Wikipedia contributors. (s. f.). *Portal:Current events/2006 October 5 - Wikipedia.*

https://en.wikipedia.org/wiki/Portal:Current_events/2006_October_5