

Predicción de Series Temporales Financieras con Machine Learning

Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones - Edición 2024

Mentor: Emmanuel Tassone

Integrantes:

- Gonzalez, Juan Cruz
- Fullana Jornet, Marcelo Fernando
- Herrador, Emanuel Nicolás
- Itovich, Griselda
- Pereira, Ariel Maximiliano

FAMAF

Facultad de Matemática,
Astronomía, Física y
Computación



UNC

Universidad
Nacional
de Córdoba



CCAD

CÓRDOBA
CLUSTER



Ministerio de
CIENCIA Y
TECNOLOGÍA



Parte 1


Presentación del proyecto y de los datos: *primeros análisis y visualización*

- Objetivos
 - Conceptos preliminares
 - Datasets
 - Visualizaciones
 - Histórico de los precios de cierre
 - Actual con gráfico de velas
 - Métodos de *downsampling* y *upsampling*
 - Riesgos de inversión
 - Análisis histórico de la volatilidad
 - ¿Cuándo una acción es cara/barata?
 - Caída de precios diaria tolerable
-

Objetivos generales

- 1) Visualizar y analizar valores históricos de las acciones de *Tesla* (2010 - 2024), realizando cálculos de métricas financieras..
- 2) Realizar transformaciones y un análisis detallado de los datos, observando dependencias, tendencias, patrones y relaciones con nuestro *objetivo de predicción* (precios de cierre).
- 3) Utilizar técnicas avanzadas de Machine Learning para desarrollar modelos predictivos que nos ayuden a anticipar los movimientos en los mercados financieros.

En esta presentación vamos a mencionar los avances en los objetivos 1 y 2



Conceptos preliminares: *Tesla* y el mercado



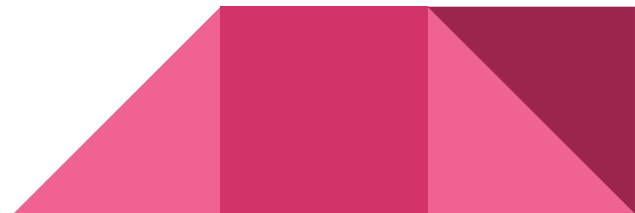
¿Qué es **TESLA**?

Tesla es una empresa dedicada a la fabricación de autos eléctricos y soluciones de energía renovable.



NASDAQ (*National Association of Securities Dealers Automated Quotation*) es el segundo mercado de valores y bolsa de valores automatizada y electrónica más grande de los Estados Unidos.

Las acciones de *Tesla* comenzaron a cotizar en este mercado desde el año 2010.



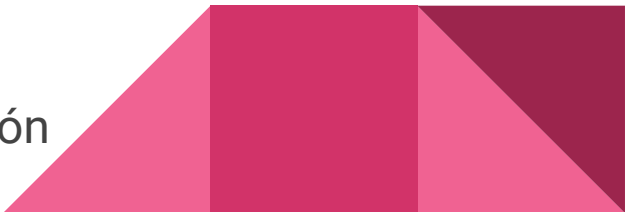
Conceptos preliminares: *Acción*

¿Qué es una acción bursátil?

Una acción representa una participación en la propiedad de una empresa, que incluye un derecho sobre las ganancias y los activos de la empresa.

¿Cómo se miden las acciones?

- *Métricas diarias:*
 - Open: precio al inicio del día
 - High: precio más alto del día
 - Low: precio más bajo del día
 - Close: precio al final del día
 - Volumen: cantidad de acciones negociadas
 - Retorno: cambio porcentual del cierre respecto al día anterior
- *Volatilidad:* grado de variación del precio de una acción
 - Indica nivel de riesgo asociado



Datasets: *Presentación*

date	open	high	low	close	volume	change_percent	avg_vol_20d
07/27/2010	20.91	21.18	20.2599	20.55	619675	-1.91	74326218.5
07/28/2010	20.55	20.9001	20.5101	20.7201	467183	0.82	60589147.1
07/29/2010	20.7699	20.88	20.0001	20.3499	615910	-1.78	48155285.4
07/30/2010	20.1999	20.4399	19.5501	19.9401	426830	-2.02	42303010.9
08/02/2010	20.4999	20.97	20.3331	20.9199	719145	4.92	38986014.6
08/03/2010	21	21.9501	20.82	21.9501	1231022	4.92	34749810
08/04/2010	21.9501	22.1799	20.85	21.2601	920755	-3.14	30246691.7
08/05/2010	21.54	21.5499	20.0499	20.4501	796379	-3.81	25054321.9
08/06/2010	20.1	20.16	19.5201	19.59	742138	-4.2	22566971.1
08/09/2010	19.8999	19.98	19.4499	19.5999	812655	0.05	21523785.6
08/10/2010	19.65	19.65	18.8199	19.0299	1281285	-2.91	20474705.1
08/11/2010	18.69	18.8799	17.85	17.9001	797649	-5.94	17925860.2
08/12/2010	17.7999	17.9001	17.3901	17.6001	691281	-1.68	15635348.3
08/13/2010	18.18	18.45	17.6604	18.3201	634513	4.09	14145325.7
08/16/2010	18.45	18.8001	18.2616	18.78	486211	2.51	12645117.3

Trabajamos con este conjunto de datos que cubren los registros diarios del mercado de Tesla desde su entrada en NASDAQ el 29 de junio de 2010 hasta el 15 de abril de 2024 (total de 3472 casos/registros).

Datasets: *Particularidades*

- *No es continuo*
 - Ya que el mercado no opera los fin de semanas o feriados.
- *Tenemos dos datasets (raw y ajustado por split):*
 - Los datos han sido ajustados para que los cambios en los precios de las acciones no se vean afectados por los splits que Tesla haya hecho a lo largo de la historia.
 - Trabajaremos con el dataset ajustado.



Visualizaciones

1. Gráficos: date (rango completo) vs. open, close, high, low, resultan ser similares.
2. **Con escala logarítmica**, se distinguen **tres periodos** de tiempo, con dinámicas particulares (Gráficos: date (rango completo) vs volumen. En tiempos de subas notorias de la acción, aumenta significativamente el volumen negociado.

Precio de cierre de Tesla



Visualizaciones avanzadas: *Gráficos de velas*

Gráficos de velas

1. **Uso habitual en el mundo financiero**
2. Emplean los datos: open, close, high y low.
3. Los colores de las velas en verde o rojo muestran periodos de ganancia o de pérdida respectivamente.

Gráfico de velas de las acciones de Tesla entre 2020 y 2024



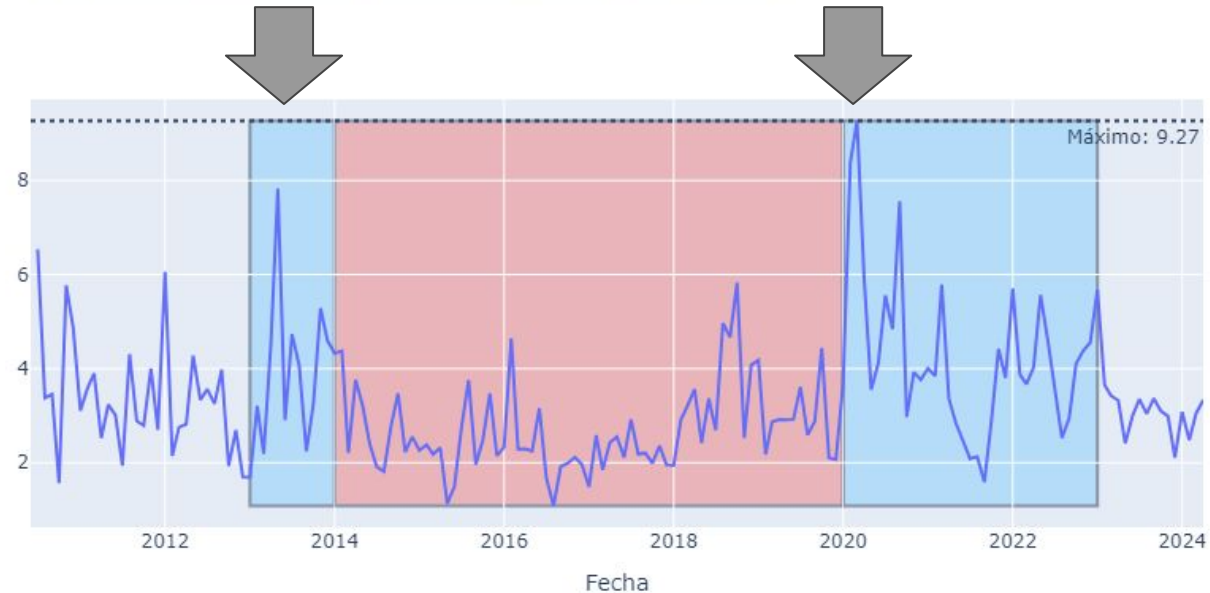
Otras visualizaciones avanzadas:

Downsampling: agrupamiento temporal de datos (disminuir frecuencia)
Upsampling: aumentar granularidad en los datos (mayor frecuencia)

Riesgos de inversión: *Volatilidad histórica*

1. Qué observar del **precio de la acción** cuando se quiere **invertir**?
2. **Change_percent (retorno diario)**. Mide el cambio de precios de cierre relativo porcentual. **El desvío estándar es la VOLATILIDAD. Es una medida de RIESGO clásico.**
3. Volatilidad mensual histórica
4. Máximos en 2013: 8% y en 2020: 9.3%.

Volatilidad mensual de las acciones de Tesla



Riesgos de inversión: *¿Cuándo una acción es cara o barata?*

1. La temporalidad es clave!

2. Periodo 2020 - 2024

3. Valores de cierre 'p':

Promedio= 207.9

Mediana= 219.93

Q1=166.16

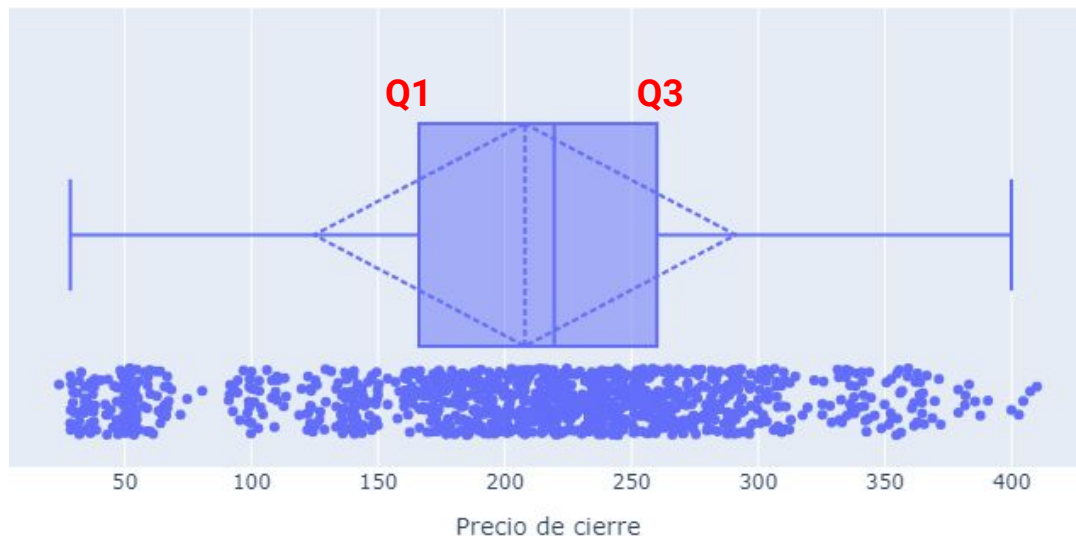
Q3=259.93

4. Un criterio posible:

CARO: $p > Q3$

BARATO: $p < Q1$

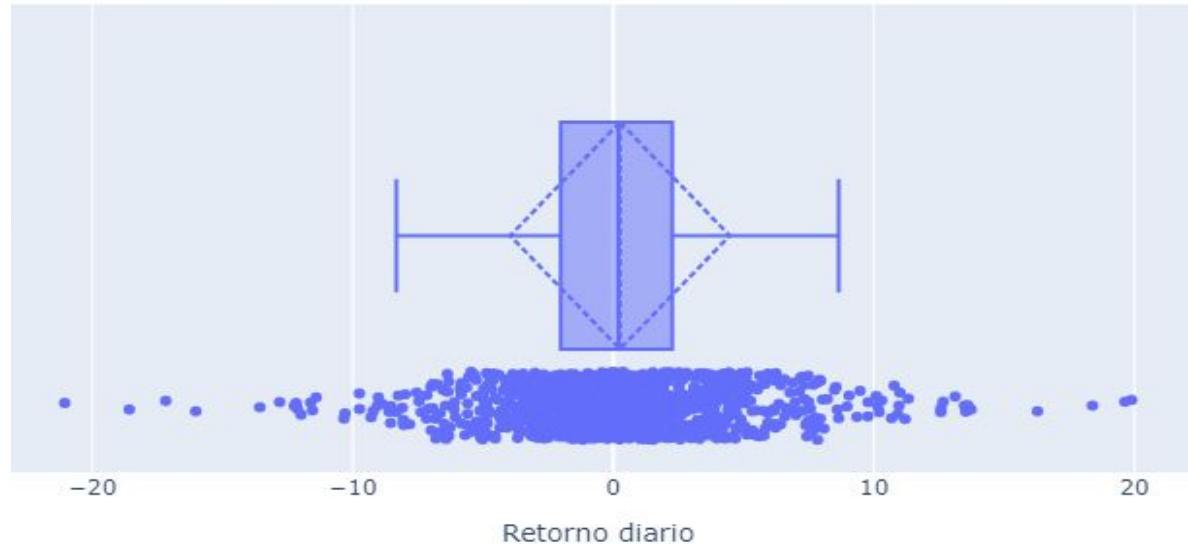
Distribución de precios de cierre de Tesla (2020-2024)



Riesgos de inversión: Hasta donde nos estamos arriesgando?

1. **Análisis de retornos diarios**
(change_percent)
2. **Periodo 2020-2024**
3. Valores de resumen:
Promedio= 0.25%
Mediana= 0.2%
Desvío= 4.2%
4. **Pérdida "esperada"**
aceptable: hasta Q1=-2%
5. **Variabilidad de precios**
"esperada": 8.4%

Distribución de retornos diarios de Tesla (2020-2024)




Parte 2

Transformación y análisis
detallado de los datos

- Transformación de los datos
 - Normalización
 - Tratamiento de outliers
 - Manejo de valores faltantes
 - Suavizamiento de series temporales
 - Dependencia de las variables
 - Análisis de Correlación
 - AutoCorrelación y AutoCorrelación Parcial
 - Análisis de estacionalidad
-

Transformación de los datos: *Normalización*

- Para que la escala de los datos no *influya* en el modelo y que la importancia de cada *feature* sea por su relación con el *target*.
 - Features involucradas:
 - Precios de cierre
 - Retornos diarios
 - Métodos usados:
 - Min-Max (escalado): ajuste en el rango $[0, 1]$.
 - Z-Score (estandarización): ajuste para media 0 y desvío 1.
 - Nos permite ver cómo se relacionan los datos respecto a la media y el desvío
- 

Serie de tiempo de *close*

Original



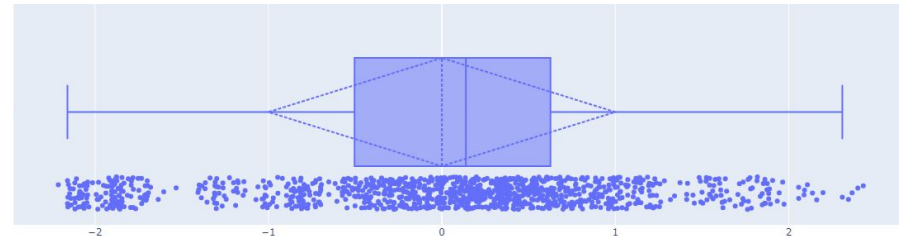
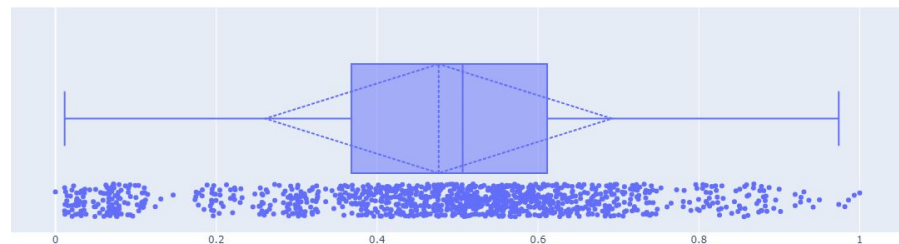
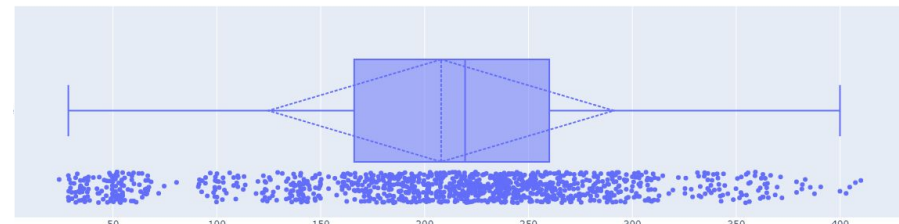
Min-Max



Z-Score



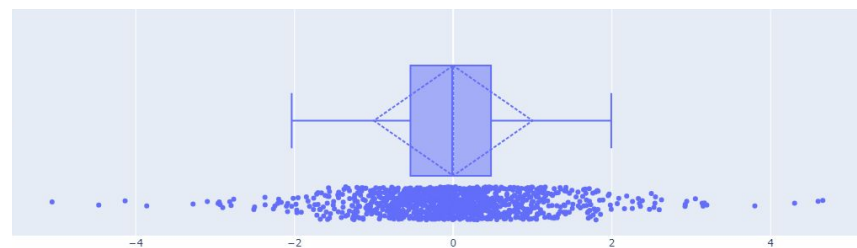
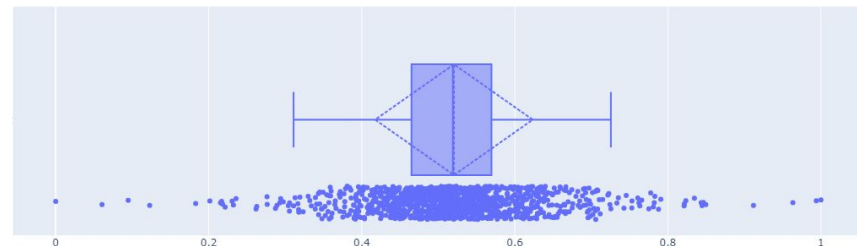
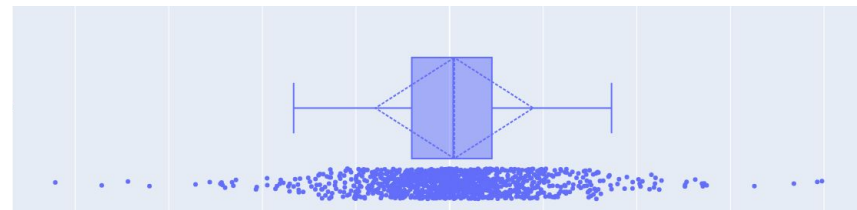
Distribución de los *precios de cierre*




Serie de tiempo de *change_percent*



Distribución de los *retornos diarios*



Transformación de los datos: *Outliers*

- La limpieza de *outliers* es necesaria para mejorar la eficacia y eficiencia de los *modelos* dado que tienen un gran impacto en las medidas sensibles a los extremos e introducen un sesgo.
 - Features involucradas:
 - Precios de cierre
 - Retornos diarios
 - Métodos usados y el rango aceptado:
 - Rango Intercuartílico (*el elegido*): $Q1 - 1.5 \times IQR \leq x \leq Q3 + 1.5 \times IQR$
 - Z-Score: $\mu - 3\sigma \leq x \leq \mu + 3\sigma$
- 

Serie de tiempo de *close*

Original

— Originales
— Precio de cierre

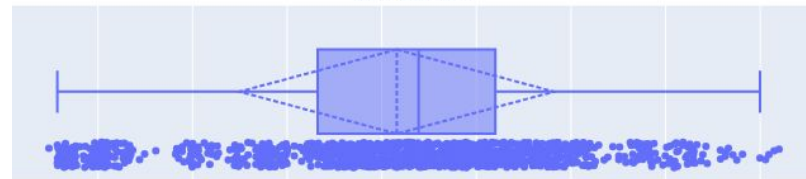
IQR

Z-Score

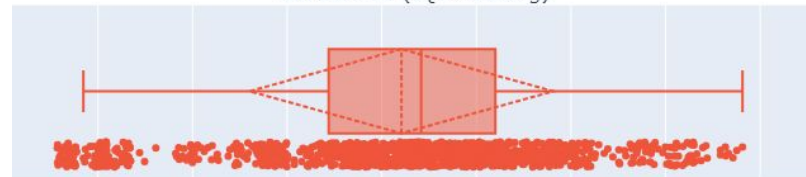


Distribución de los *precios de cierre*

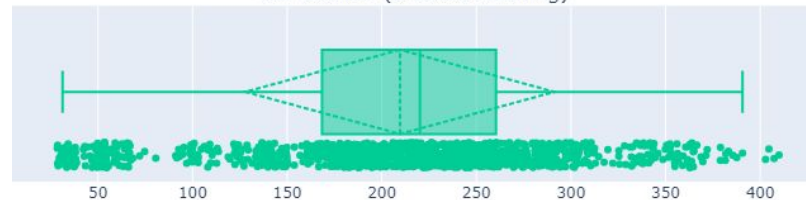
Con outliers



Sin outliers (IQR cleaning)



Sin outliers (Z-score cleaning)



Serie de tiempo de *change_percent*

Original

— Originales
— Precio de cierre

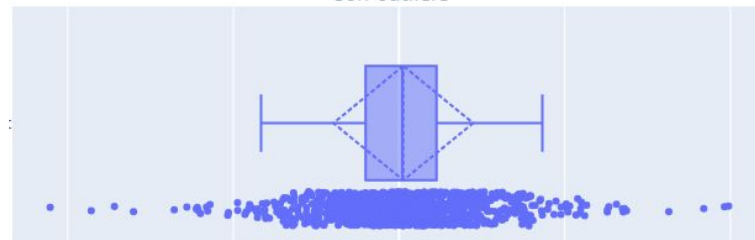
IQR

Z-Score

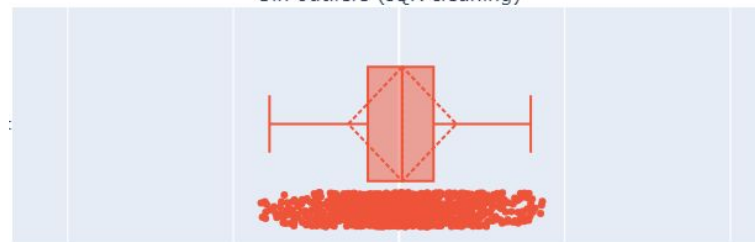


Distribución de los *retornos diarios*

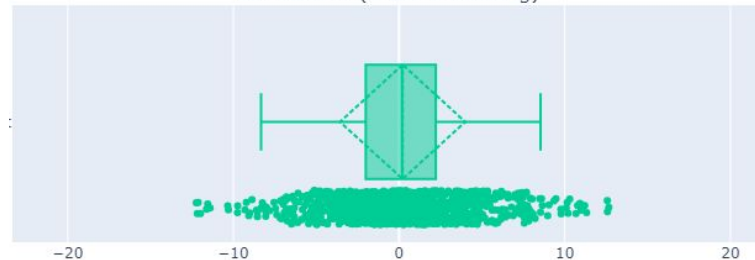
Con outliers



Sin outliers (IQR cleaning)

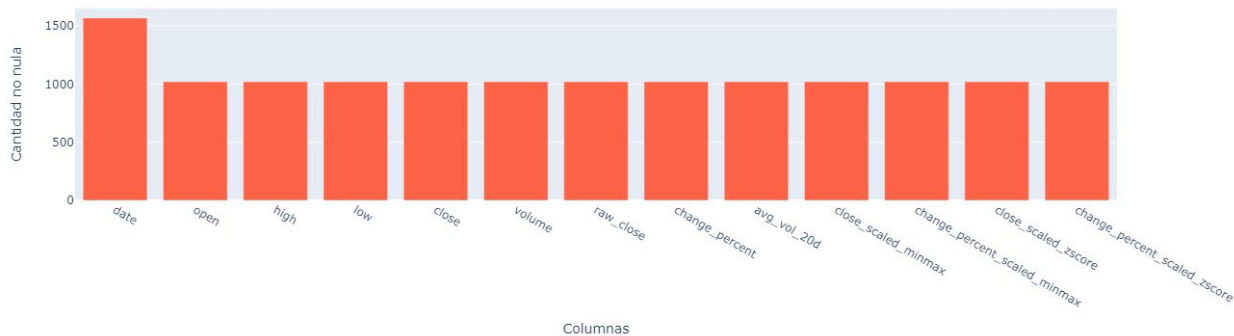


Sin outliers (Z-score cleaning)



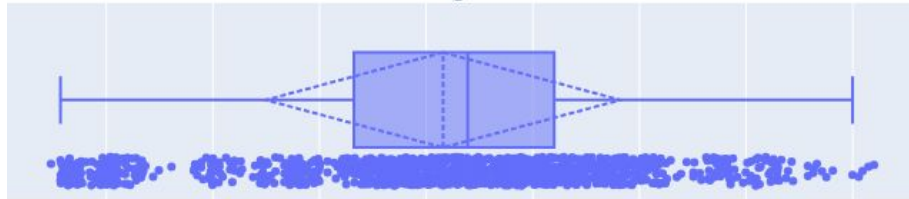
Transformación de los datos: *Imputación*

- Para completar los días que quedaron sin valores por la limpieza de *outliers* con el *método IQR*, hacemos interpolación lineal.
- Visualización de cantidad de campos con valores por columna:

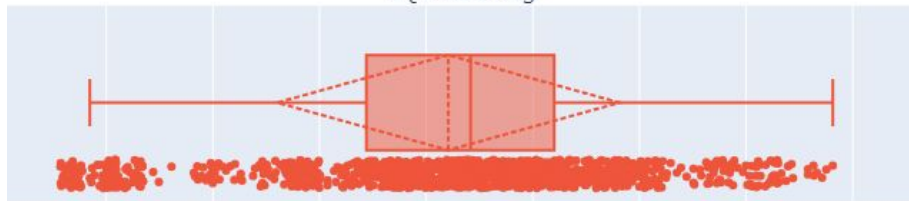


Distribución de los *precios de cierre*

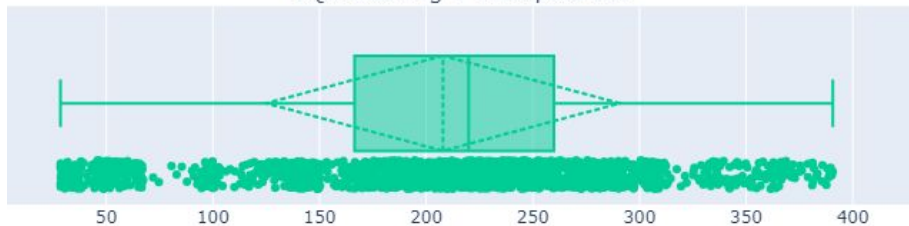
Original



IQR cleaning

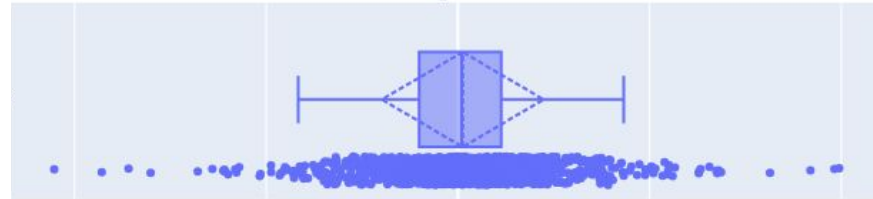


IQR cleaning + Interpolación

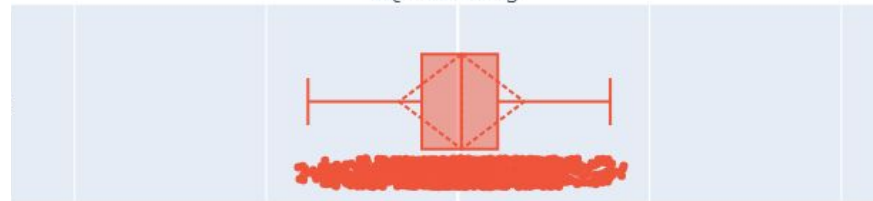


Distribución de los *retornos diarios*

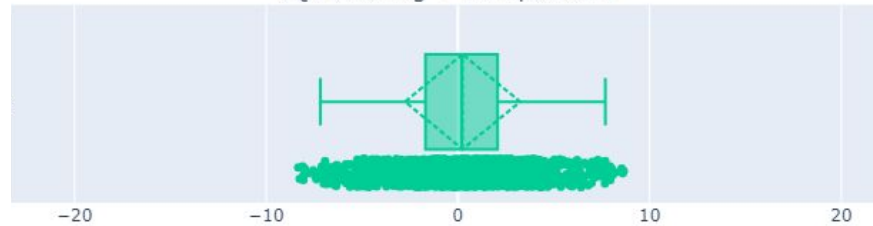
Original



IQR cleaning



IQR cleaning + Interpolación

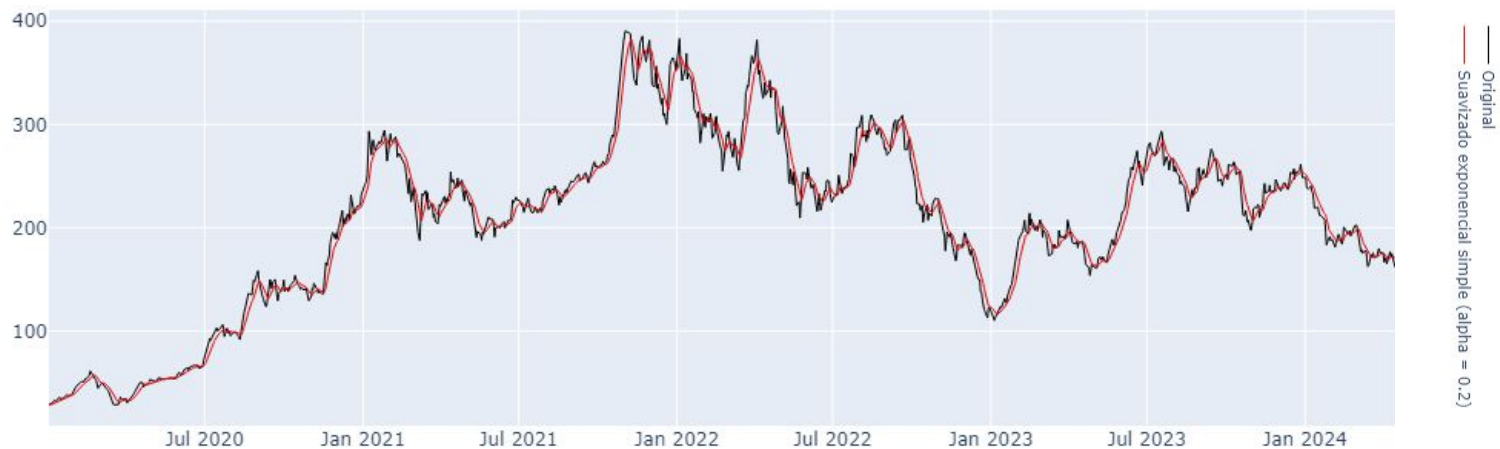


Transformación de los datos: *Suavizamiento*

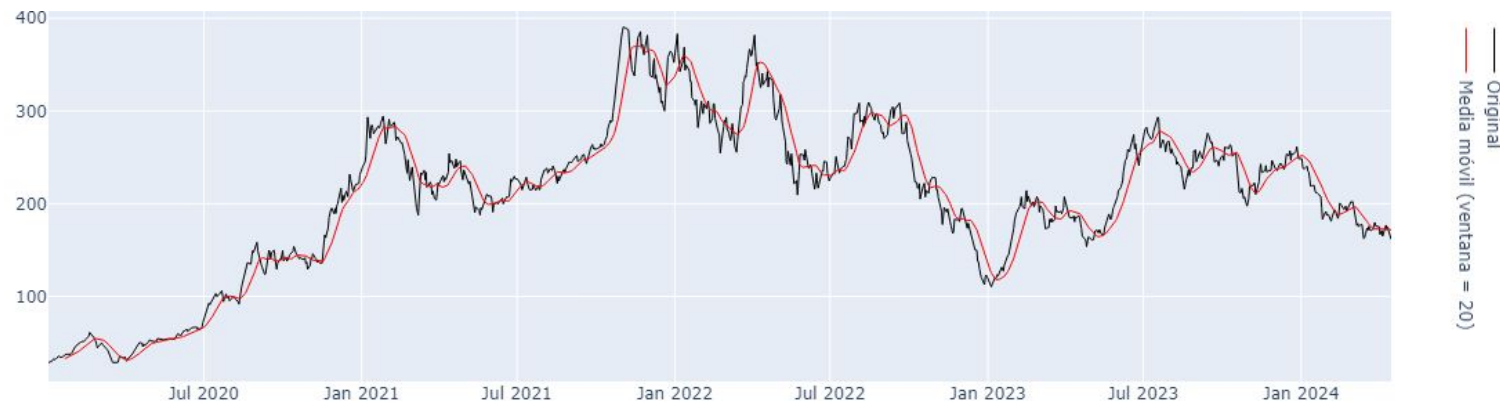
- El *suavizamiento de series temporales* se utiliza para reducir la variabilidad y resaltar las tendencias subyacentes.
 - Revela tendencias o ciclos a *largo* plazo.
 - Suaviza el ruido y las fluctuaciones a *corto* plazo.
- Métodos usados:
 - Exponencial (*cobra sentido con el análisis de AutoCorrelación y AutoCorrelación Parcial*)
 - Se basa en que el valor futuro se calcula como una combinación ponderada de los valores previos, ponderando a los más recientes sobre los valores anteriores.
 - Mediante medias móviles
 - No realiza ponderación de algún valor, sino que es el promedio de los valores de una cantidad fija de días anteriores
 - Consideramos la ventana de 20 días.

Serie de tiempo de *close*

Exponencial simple



Medias móviles



Dependencias de las variables: Entre variables

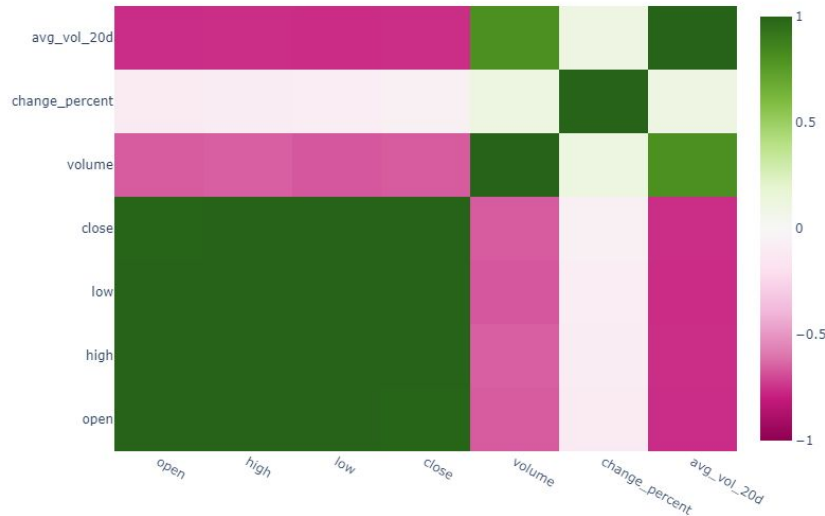
Análisis de correlación

El análisis de correlación se utiliza en ciencia de datos para medir y evaluar la relación entre dos variables cuantitativas. Específicamente, se enfoca en determinar si existe una relación estadística entre las variables y, de ser así, la fuerza y la dirección de dicha relación.



Dependencias de las variables: Análisis de correlación

Correlación entre las variables del dataset de Tesla (2020-2024)



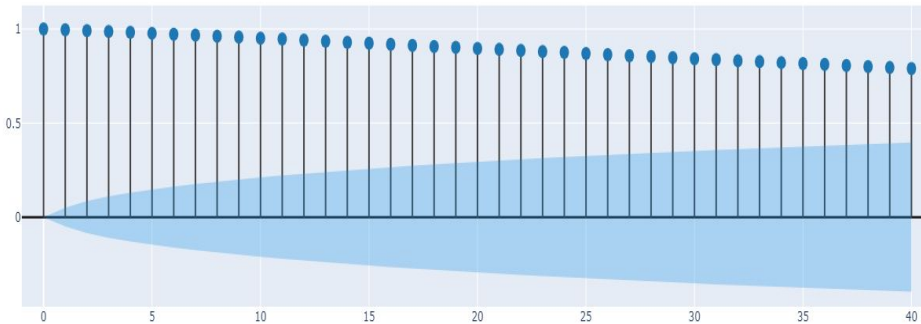
Medición de la fuerza de la relación: La correlación cuantifica qué tan fuerte es la relación entre dos variables, con un valor que generalmente oscila entre -1 y 1.

- **Correlación positiva (+1):** A medida que una variable aumenta, la otra también lo hace. Podemos observar que esto pasa con las variables open, close, low y high, dado que representan los valores de los precios diarios de las acciones de Tesla.
- **Correlación negativa (-1):** A medida que una variable aumenta, la otra disminuye. Podemos ver que esto ocurre con las variables avg_vol_20 y volumen con open, close, low y high.
- **Correlación cero (0):** No hay una relación lineal entre las variables. Podemos ver que esto ocurre entre change_percent y cada una de las variables, a excepción de ella misma

Dependencias de las variables: Temporal intrínseca

Análisis de Autocorrelación (ACF)

Autocorrelación de close

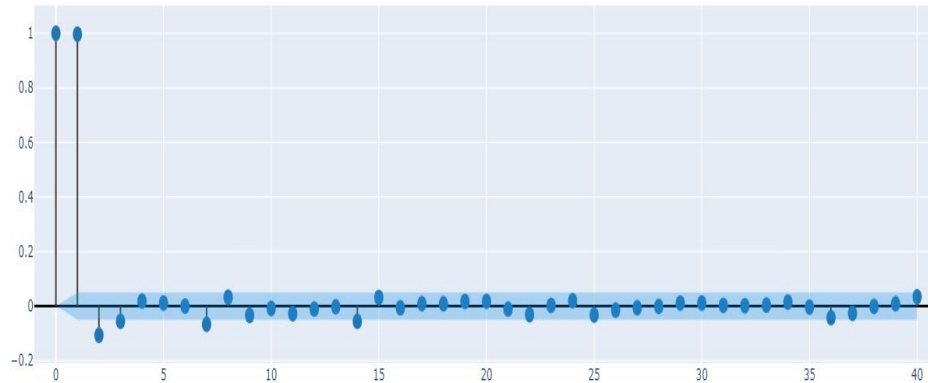


La **autocorrelación** mide la correlación de una serie temporal consigo misma en diferentes rezagos (lags). En otras palabras, evalúa la relación entre los valores actuales de la serie y los valores anteriores en distintos puntos en el tiempo. Todas las correlaciones con los primeros 40 retardos se encuentran por encima del intervalo de confianza (región sombreada), lo que nos indica que los valores actuales de los precios de cierre de Tesla se encuentran fuertemente correlacionados con sus valores pasados

Dependencias de las variables: Temporal intrínseca

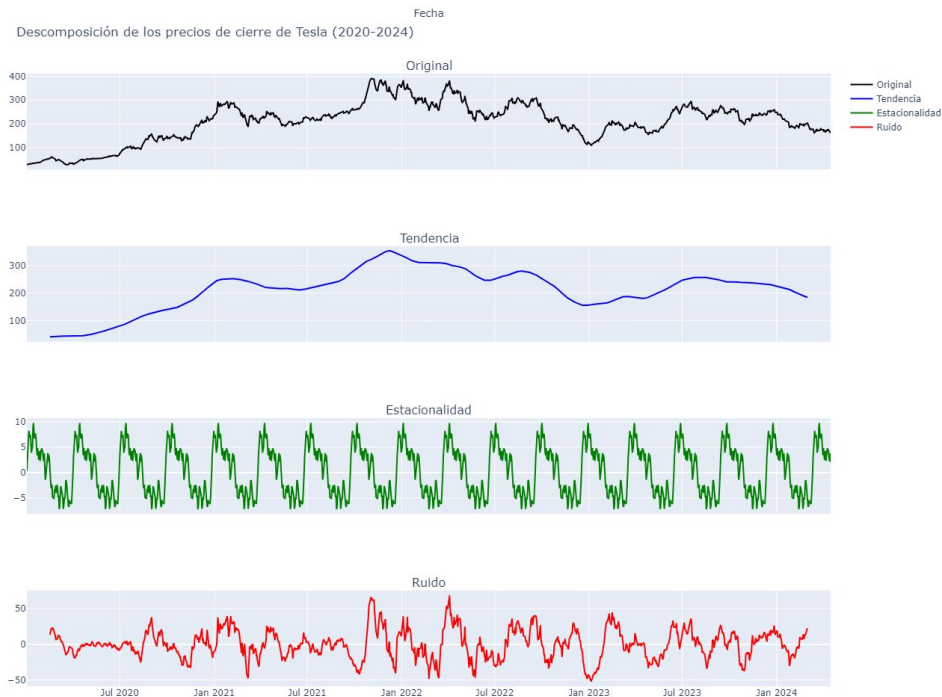
Análisis de Autocorrelación Parcial (PACF)

Autocorrelación Parcial de cierre



La **autocorrelación parcial** mide la relación entre la serie y un valor anterior en un lag específico, eliminando la influencia de los valores intermedios. Es decir, mide la correlación entre los valores de la serie y un lag determinado, después de ajustar por los lags intermedios. Podemos notar que las únicas correlaciones fuertes que se encuentran son con el primer retardo (casi 1) y el segundo (-0.1), lo cual nos indica que los valores actuales de los precios de cierre de Tesla están fuertemente correlacionados con sus valores inmediatamente anteriores

Análisis de Estacionalidad



El análisis de estacionalidad se utiliza para identificar y entender patrones recurrentes en los datos a lo largo del tiempo, que se repiten en intervalos regulares (por ejemplo, semanal, mensual, trimestral o anual). Este análisis es especialmente importante en series temporales, donde es crucial separar las variaciones estacionales de otras tendencias y ciclos para hacer predicciones precisas y tomar decisiones informadas. En el gráfico, podemos notar que la tendencia actual es bajista, que la estacionalidad es mayormente positiva y que el ruido es irregular a lo largo del tiempo y no sigue un patrón claro (hay momentos de mayor y menor volatilidad).

Conclusiones

Hemos llegado al final de nuestra presentación y consideramos que hemos cumplido con los objetivos que nos propusimos inicialmente, los cuales detallamos

- 1) Visualizar y analizar valores históricos de las acciones de *Tesla* (2010 - 2024), realizando cálculos de volatilidad, distribución de retornos diarios y otras métricas.
- 2) Realizar transformaciones y un análisis detallado de los datos, observando dependencias, tendencias, patrones y relaciones con nuestro *objetivo de predicción* (precios de cierre).

Ahora nos queda prepararnos para la segunda etapa, en donde buscaremos cumplir con el siguiente objetivo

Utilizar técnicas avanzadas de Machine Learning para desarrollar modelos predictivos que nos ayuden a anticipar los movimientos en los mercados financieros.

