



Facultad de Matemática,  
Astronomía, Física y  
Computación



UNC

Universidad  
Nacional  
de Córdoba

DIPLOMATURA DE DATOS 2024

# BOLETÍN OFICIAL

MENTOR: EMMANUEL TASSONE

## Trabajo Práctico 2 - Exploración y curación de datos

Una vez que entendimos el dataset y visualizamos su conjunto de datos, estamos listos para adentrarnos en su exploración. Esto es, interpretar los datos sacando conclusiones más complejas y menos visibles a simple vista. Iremos descubriendo más información de nuestro dataset a lo largo del trabajo práctico y sus preguntas.

### Actividades

#### 1. Análisis exploratorio inicial

- Obtener un resumen estadístico de las columnas *close* y *change\_percent*.
- A partir de esta información, ¿cuál diría que es un precio caro para la acción tesla y un precio barato? ¿Cuál es la caída de precios diaria máxima que un inversor tendría que estar dispuesto a soportar para invertir en esta acción?
- Hacer un gráfico del precio de cierre, mostrando la media, la desviación estándar y los cuartiles más importantes.

#### 2. Normalización de Series Temporales

- Normalice la serie temporal utilizando min-max scaling
- Normalice la serie temporal utilizando z-score

#### 3. Valores atípicos

- Visualizar los valores atípicos en las columnas *close* y *change\_percent*. Para esto, deberás ayudarte con diagramas de caja y la función `boxplot` de la librería `seaborn`.
- Utilice el método del rango intercuartílico (IQR) para descartar los valores atípicos. ([https://en.wikipedia.org/wiki/Interquartile\\_range](https://en.wikipedia.org/wiki/Interquartile_range))
- Utilice el método de Z-score para descartar los valores atípicos (utilizar 3 sigmas).

#### 4. Manejo de Valores Faltantes

- Calcular los valores faltantes de las columnas *change\_percent* y *avg\_vol\_20d*, utilizando las funciones `isnull()` y `sum()`.
- Eliminar las filas del dataset donde falten datos en algunas de las columnas mencionadas.

#### 5. Análisis de correlación

- Calcula la matriz de correlación para las columnas open, high, low, close, volume, y avg\_vol\_20d. Visualiza la matriz de correlación utilizando un mapa de calor (heatmap).
- Realiza un análisis de autocorrelación y autocorrelación parcial para la serie temporal de precios de cierre utilizando las funciones `plot_acf` y `plot_pacf` de la librería `statsmodels`.

#### 6. Suavizamiento de Series Temporales

- Aplica las técnicas de suavizamiento exponencial y suavizamiento mediante medias móviles a la columna de precios de cierre para reducir la variabilidad y resaltar las tendencias.

#### 7. Análisis de Estacionalidad (opcional)

- Descompon la serie temporal de los precios de cierre para identificar componentes estacionales mediante la librería `statsmodels.api`. En particular, utiliza la función `seasonal_decompose()` para descomponer la serie en las tres componentes  $T[t] + S[t] + e[t]$ , donde  $T[t]$  es la tendencia,  $S[t]$  la componente estacional y  $e[t]$  el ruido.
- Visualiza los componentes calculadas.



### Características que debe cumplir el entregable

- El proyecto debe ser escrito en un *jupyter notebook* (.ipynb), siguiendo las convenciones PEP8 (<https://peps.python.org/pep-0008/>).
- El Notebook debe ser claro y estar bien organizado: debe contar con un índice, apartados, código fácil de leer, probado y comentado (no abusar de los comentarios).
- El archivo .ipynb debe compartirse a través de un Google Collab con permisos de edición habilitados al mentor. A su vez, una copia del archivo debe enviarse a la casilla de correo [emmanuel.tassone@unc.edu.ar](mailto:emmanuel.tassone@unc.edu.ar)
- Tener en cuenta que el archivo entregable debe contener solo las resoluciones pedidas, dejando de lado análisis adicionales que se hayan hecho.
- Es importante que luego de cada código y gráfico **haya una conclusión o interpretación** de lo obtenido.

