UNIVERSIDAD SAN FRANCISCO DE QUITO

FUNDAMENTOS DE CIENCIA DE DATOS

PROYECTO 1

TEMA: IMPACTO DE INSIDER TRADING FRENTE A LAS VARIACIONES DEL PRECIO DE LA ACCIÓN DE APPLE QUE COTIZAN EN LA BOLSA DE NUEVA YORK DESDE EL AÑO 1 DE ENERO DEL 2015 HASTA EL 27 DE MARZO 2025.

ALUMNO: SANTIAGO RODRIGUEZ

MARZO 2025

Contenido

1.	DATOS EXTRAIDOS DE YAHOO FINANCE	3
2.	DATOS EXTRAIDOS DE OPENINSIDER.COM	5
3.	MERGE DE TABLAS	11
Ilus	stración 1: primera visualización de datos curados de AAPL	3
Ilus	stración 2: tratamiento inicial como borrado de columnas	3
Ilus	stración 3: redondear a 2 decimales	4
Ilus	stración 4: creación de columna movimiento	4
Ilus	stración 5: Tabla de cotizaciones de la accion AAPL final	5
Ilus	stración 6: primera visualización de tabla curada de OPENINSIDER.COM	5
	stración 7: eliminación de coumnas 0, X, 1D, 1W, 1M, 6M	
Ilus	stración 8: separación de fecha y hora	6
Ilus	stración 9: filtrar información desde el año 2015	7
Ilus	stración 10: verificación de cargos duplicados	7
Ilus	stración 11: filtrado de nombres de directivos que poseen múltiples cargos	8
Ilus	stración 12: reemplazo de cargos	8
	stración 13: estandarización de tipos de transacción	
Ilus	stración 14: verificación de tipos de transacción	9
	stración 15: creación de tabla dinámica para evaluar consolidación de fechas	
Ilus	stración 16 verificación de fechas duplicadas	10
Ilus	stración 17: unión de tablas	11
Ilus	stración 18 tabla fct definitiva	12
Ilus	stración 19: exportación a un archivo .csv	12

REPORTE DATA_WRANGLING

- 1. DATOS EXTRAIDOS DE YAHOO FINANCE
- Se importan los datos de la nueva carpeta denominada "curated"

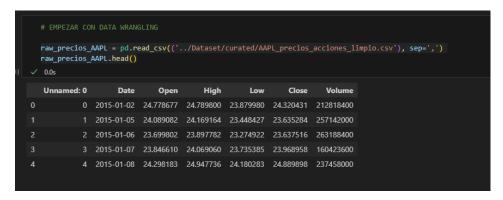


Ilustración 1: primera visualización de datos curados de AAPL

 Se reinician los índices, eliminamos las columnas index, unnamed:0 y aplicamos un .info()

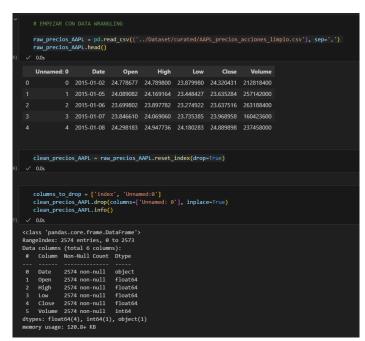


Ilustración 2: tratamiento inicial como borrado de columnas

Se redondean los decimales a 2

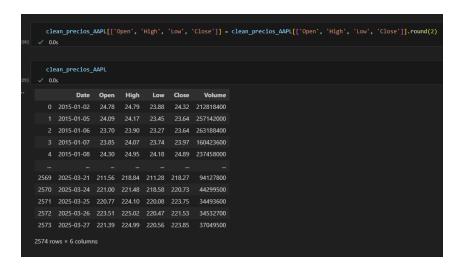


Ilustración 3: redondear a 2 decimales

 Creamos una columna donde se observa la diferencia entre el precio de apertura y el precio de cierre, con esto, se muestra si al final del día esta acción se movió. A continuación, una visualización de como quedaría:

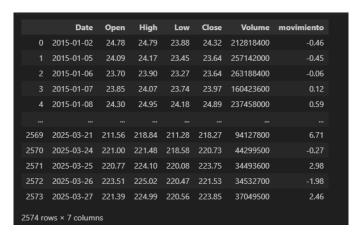


Ilustración 4: creación de columna movimiento

A continuación, se va a realizar la eliminación de las columnas High, Low y Volume, a continuación, las razones:

- High y Low: en base a la información extraída con la API no puedo observar por hora el comportamiento de la acción en el momento que se negociaron las acciones por parte de los directivos de Apple. Solamente muestra los picos que tuvo esa acción durante ese día pero a nivel macro, no me aporta nada para el propósito de mi análisis.
- Volume: Es el float de acciones circulando en el mercado, es decir, la cantidad de acciones de esa compañía que está disponible al público que pueden ser negociadas.
 No es significativo para el propósito de mi análisis.

Así quedaría:

	Date	Open	Close	movimiento	
0	2015-01-02	24.78	24.32	-0.46	
	2015-01-05	24.09	23.64	-0.45	
2	2015-01-06	23.70	23.64	-0.06	
	2015-01-07	23.85	23.97	0.12	
4	2015-01-08	24.30	24.89	0.59	
2569	2025-03-21	211.56	218.27	6.71	
2570	2025-03-24	221.00	220.73	-0.27	
2571	2025-03-25	220.77	223.75	2.98	
2572	2025-03-26	223.51	221.53	-1.98	
2573	2025-03-27	221.39	223.85	2.46	
2574 rows × 4 columns					

Ilustración 5: Tabla de cotizaciones de la accion AAPL final

2. DATOS EXTRAIDOS DE OPENINSIDER.COM

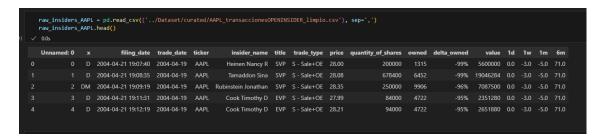


Ilustración 6: primera visualización de tabla curada de OPENINSIDER.COM

Se eliminan las siguientes columnas por los siguientes motivos:

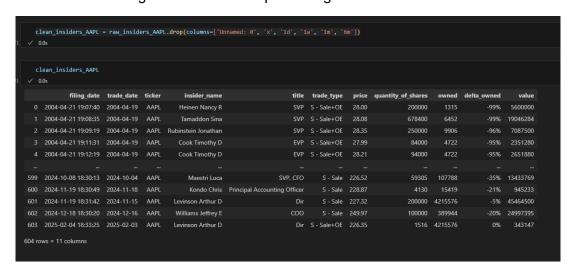


Ilustración 7: eliminación de coumnas 0, X, 1D, 1W, 1M, 6M

- X = datos faltantes que se presentan en la columna X que hace referencia a la forma en la que se reportó el formulario ante la SEC, algunos registros tienen valores vacíos.
- (1d, 1w, 1m, 6m) = como se puede evidenciar que en el año 2004 estas columnas si
 poseían datos, pero conforme ha ido avanzando el tiempo se dejó de reportar esta
 información, entiendo que por las cabeceras de la base de datos querían denotar la
 frecuencia de transacciones negociadas, pero esta información no tiene relevancia para
 el propósito de mi análisis.
- Separación de la fecha y hora en la columna filing date:

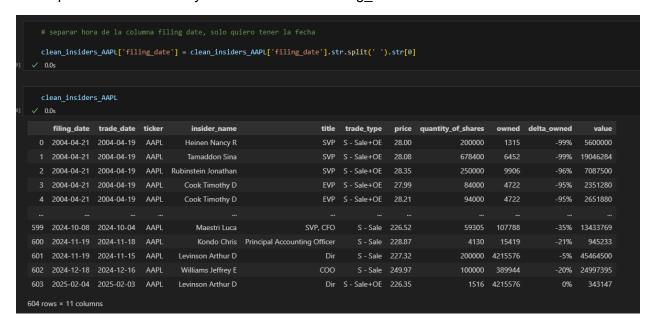


Ilustración 8: separación de fecha y hora

 Filtrado de información solo del período que tengo información de los datos extraídos de Yahoo finance, es decir, solo filtro la fecha desde el 2015-01-01; reduciendo el número de registros de 604 líneas a 221. A continuación, un detalle:

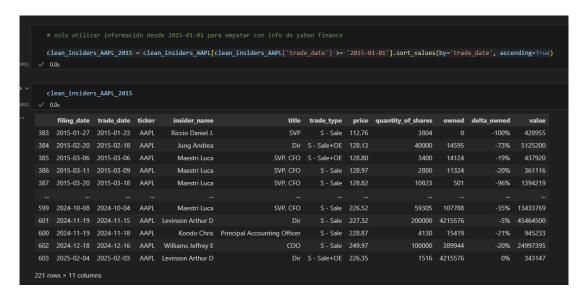


Ilustración 9: filtrar información desde el año 2015

Unificar los cargos que poseen los directivos

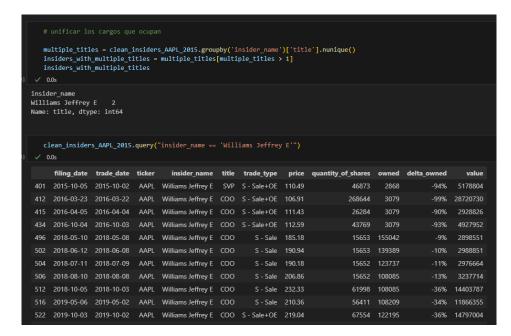


Ilustración 10: verificación de cargos duplicados

• Este código no me sirve ya que esta persona a lo largo de su carrera en Apple mantuvo 2 cargos distintos. Lo que quería llegar es a las personas que tienen más de 2 cargos en una misma línea, así que tomaré otro enfoque.

Ilustración 11: filtrado de nombres de directivos que poseen múltiples cargos

 Con esto se observa que 3 personas tienen varios títulos respecto al llenado del formulario, sin embargo, voy a colocar solo un título a estos 3 colaboradores. Kondo Chris no lo modifico ya que su cargo supera los 3 caracteres, pero se mantiene en un solo cargo. Se procede a modificar a los que tienen varias iniciales.

Ilustración 12: reemplazo de cargos

Se ajustan los cargos que ocupan a uno solo. A continuación, el significado:

CEO: Gerente general

• CFO: Gerente financiero

COO: Gerente de Operaciones

Dir: Director ejecutivo

- PAO : Contador general
- SVP : Vicepresidente Senior

Estandarizar tipos de transacción.

```
clean_insiders_AAPL_2015['trade_type'] = clean_insiders_AAPL_2015['trade_type'].replace({
     'S - Sale': 'S',
'S - Sale+OE': 'S',
  clean_insiders_AAPL_2015
     filing_date trade_date ticker insider_name title trade_type price quantity_of_shares owned delta_owned
                                                                                              value
384 2015-02-20 2015-02-18 AAPL Jung Andrea Dir
385 2015-03-06 2015-03-06 AAPL Maestri Luca CFO S 128.80 3400 14124 -19%
                                                                                             437920
387 2015-03-20 2015-03-18 AAPL Maestri Luca CFO S 128.82 10823 501
                                                                                     -96% 1394219
                                              ...
A<u>PL</u> Maestri Luca CFO
599 2024-10-08 2024-10-04 AAPL
    2024-11-19 2024-11-18 AAPL Kondo Chris PAO
                                                   S 228.87
                                                                    100000 389944
 603 2025-02-04 2025-02-03 AAPL Levinson Arthur D Dir S 226.35 1516 4215576
221 rows × 11 columns
```

Ilustración 13: estandarización de tipos de transacción

 Sale + OE: es una venta que fue ejecutada a través de un contrato de opciones, si lo vemos de forma general, sigue siendo una operación de venta. No interfiere la forma en la que se negoció esa acción.

Ilustración 14: verificación de tipos de transacción

Eliminación de filing_date: Se elimina filing date ya que es una columna donde muestra la fecha de reporte del formulario a la entidad de control. Filing date vs trade date tenían una brecha de 2 días de diferencia, generalmente se reportaba la transacción luego de 2 a 3 días y esto representaría un sesgo por lo que decido eliminar esta columna. Me quedo con trade_date porque registra el día exacto donde tuvo lugar la transacción y esto se conectaría con la tabla de cotizaciones de yahoo.

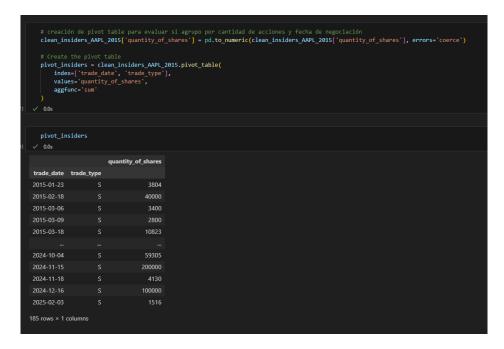


Ilustración 15: creación de tabla dinámica para evaluar consolidación de fechas

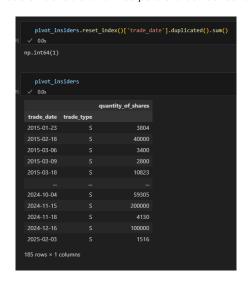


Ilustración 16 verificación de fechas duplicadas

 Aquí estaba evaluando si debía agrupar por trade_date, cantidad de acciones negociadas por día y por tipo de transacción para evitar tener fechas duplicadas, sin embargo, pensando en cómo avanzaría el modelo más adelante quisiera asignarle un peso por el cargo que ocupan los directivos y evaluar si por eso es por lo que se ve afectado el precio de la acción.

3. MERGE DE TABLAS

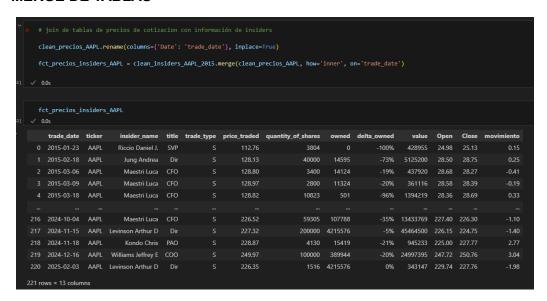


Ilustración 17: unión de tablas

Al momento de realizar el cruce de información me doy cuenta de que la columna de price_traded difiere bastante con las columnas de Open y Close, lo lógico hubiera sido que el precio negociado tenga relación con el precio del día.

La explicación para esta inconsistencia es que en junio del 2014¹ Apple realizó una división (Split) de sus acciones de 7 a 1, esto quiere decir que el valor de sus acciones en ese momento se dividió para 7. Por ejemplo, si una acción costaba \$70, luego de la división (Split), habrían 7 acciones que valen \$10.

Las compañías realizan esta estrategia con el objetivo de captar más inversionistas para aumentar su capitalización de mercado, aumentar dividendos, recompra de acciones para accionistas internos y aumentar su beneficio por acción (EPS), entre los más comunes.

Entonces, la información extraída con el API de Yahoo ya obtiene las cotizaciones de las acciones con el efecto ajustado por splits; sin embargo, la información de OPEN INSIDERS, registra tal cual el precio que tuvo lugar la transacción en ese momento.

Para denotar esta diferencia tomamos la primera fila en donde el price_traded es 112.76 y si tomamos el precio de cierre que es 24.98, la división da 4.54. Tomando en cuenta que el split tuvo lugar en el 2014, los valores no van a coincidir porque en el año 2015 se arrastra el efecto que tuvo el split.

Más adelante, en agosto de 2020 se realiza un nuevo split de 4 a 1, por lo que de nuevo price_traded no empataría con la información de precios de apertura y cierre.

¹ Fuente: página web de Apple de relación con inversionistas. Enlace: https://investor.apple.com/faq/default.aspx#:~:text=Apple's%20stock%20has%20split%20five,%2C%20and %20June%2016%2C%201987.

Es por este motivo que voy a eliminar la columna price_traded y ejecutaré de nuevo los códigos para obtener la tabla fct definitiva.

Luego de esta explicación, el resultado de la tabla final sería:



Ilustración 18 tabla fct definitiva

Con esto se concluye el data wrangling, exportamos los datos a una nueva carpeta llamada clean y renombramos al archivo.csv.

Ilustración 19: exportación a un archivo .csv