

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

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

```
# EMPEZAR CON DATA WRANGLING

raw_precios_AAPL = pd.read_csv('../Dataset/curated/AAPL_precios_acciones_limpio.csv'), sep=',')
raw_precios_AAPL.head()
```

Unnamed: 0	Date	Open	High	Low	Close	Volume
0	2015-01-02	24.778677	24.789800	23.879980	24.320431	212818400
1	2015-01-05	24.089082	24.169164	23.448427	23.635284	257142000
2	2015-01-06	23.699802	23.897782	23.274922	23.637516	263188400
3	2015-01-07	23.846610	24.069060	23.735385	23.968958	160423600
4	2015-01-08	24.298183	24.947736	24.180283	24.889898	237458000

Ilustración 1: primera visualización de datos curados de AAPL

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

```
# EMPEZAR CON DATA WRANGLING

raw_precios_AAPL = pd.read_csv('../Dataset/curated/AAPL_precios_acciones_limpio.csv'), sep=',')
raw_precios_AAPL.head()
```

Unnamed: 0	Date	Open	High	Low	Close	Volume
0	2015-01-02	24.778677	24.789800	23.879980	24.320431	212818400
1	2015-01-05	24.089082	24.169164	23.448427	23.635284	257142000
2	2015-01-06	23.699802	23.897782	23.274922	23.637516	263188400
3	2015-01-07	23.846610	24.069060	23.735385	23.968958	160423600
4	2015-01-08	24.298183	24.947736	24.180283	24.889898	237458000

```
clean_precios_AAPL = raw_precios_AAPL.reset_index(drop=True)
```

Unnamed: 0	Date	Open	High	Low	Close	Volume
0	2015-01-02	24.778677	24.789800	23.879980	24.320431	212818400
1	2015-01-05	24.089082	24.169164	23.448427	23.635284	257142000
2	2015-01-06	23.699802	23.897782	23.274922	23.637516	263188400
3	2015-01-07	23.846610	24.069060	23.735385	23.968958	160423600
4	2015-01-08	24.298183	24.947736	24.180283	24.889898	237458000

```
columns_to_drop = ['index', 'Unnamed: 0']
clean_precios_AAPL.drop(columns=['Unnamed: 0'], inplace=True)
clean_precios_AAPL.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2574 entries, 0 to 2573
Data columns (total 6 columns):
#   Column  Non-Null Count  Dtype
---  -
0   Date    2574 non-null      object
1   Open    2574 non-null      float64
2   High    2574 non-null      float64
3   Low     2574 non-null      float64
4   Close   2574 non-null      float64
5   Volume  2574 non-null      int64
dtypes: float64(4), int64(1), object(1)
memory usage: 120.8+ KB
```

Ilustración 2: tratamiento inicial como borrado de columnas

- Se redondean los decimales a 2

```
clean_precios AAPL[['Open', 'High', 'Low', 'Close']] = clean_precios AAPL[['Open', 'High', 'Low', 'Close']].round(2)
```

✓ 0.0s

clean_precios AAPL

✓ 0.0s

	Date	Open	High	Low	Close	Volume
0	2015-01-02	24.78	24.79	23.88	24.32	212818400
1	2015-01-05	24.09	24.17	23.45	23.64	257142000
2	2015-01-06	23.70	23.90	23.27	23.64	263188400
3	2015-01-07	23.85	24.07	23.74	23.97	160423600
4	2015-01-08	24.30	24.95	24.18	24.89	237458000
...
2569	2025-03-21	211.56	218.84	211.28	218.27	94127800
2570	2025-03-24	221.00	221.48	218.58	220.73	44299500
2571	2025-03-25	220.77	224.10	220.08	223.75	34493600
2572	2025-03-26	223.51	225.02	220.47	221.53	34532700
2573	2025-03-27	221.39	224.99	220.56	223.85	37049500

2574 rows × 6 columns

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:

	Date	Open	High	Low	Close	Volume	movimiento
0	2015-01-02	24.78	24.79	23.88	24.32	212818400	-0.46
1	2015-01-05	24.09	24.17	23.45	23.64	257142000	-0.45
2	2015-01-06	23.70	23.90	23.27	23.64	263188400	-0.06
3	2015-01-07	23.85	24.07	23.74	23.97	160423600	0.12
4	2015-01-08	24.30	24.95	24.18	24.89	237458000	0.59
...
2569	2025-03-21	211.56	218.84	211.28	218.27	94127800	6.71
2570	2025-03-24	221.00	221.48	218.58	220.73	44299500	-0.27
2571	2025-03-25	220.77	224.10	220.08	223.75	34493600	2.98
2572	2025-03-26	223.51	225.02	220.47	221.53	34532700	-1.98
2573	2025-03-27	221.39	224.99	220.56	223.85	37049500	2.46

2574 rows × 7 columns

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
1	2015-01-05	24.09	23.64	-0.45
2	2015-01-06	23.70	23.64	-0.06
3	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

```
raw_insiders_AAPL = pd.read_csv('../Dataset/curated/AAPL_transaccionesOPENINSIDER_limpio.csv', sep=',')
raw_insiders_AAPL.head()
```

0.0s

Unnamed: 0	x		filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owed	value	1d	1w	1m	6m
0	0	D	2004-04-21 19:07:40	2004-04-19	AAPL	Heinen Nancy R	SVP	S - Sale+OE	28.00	200000	1315	-99%	5600000	0.0	-3.0	-5.0	71.0
1	1	D	2004-04-21 19:08:35	2004-04-19	AAPL	Tamaddon Sina	SVP	S - Sale+OE	28.08	678400	6452	-99%	19046284	0.0	-3.0	-5.0	71.0
2	2	DM	2004-04-21 19:09:19	2004-04-19	AAPL	Rubinstein Jonathan	SVP	S - Sale+OE	28.35	250000	9906	-96%	7087500	0.0	-3.0	-5.0	71.0
3	3	D	2004-04-21 19:11:31	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	27.99	84000	4722	-95%	2351280	0.0	-3.0	-5.0	71.0
4	4	D	2004-04-21 19:12:19	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	28.21	94000	4722	-95%	2651880	0.0	-3.0	-5.0	71.0

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

- Se eliminan las siguientes columnas por los siguientes motivos:

```
clean_insiders_AAPL = raw_insiders_AAPL.drop(columns=['Unnamed: 0', 'x', '1d', '1w', '1m', '6m'])
```

```
clean_insiders_AAPL
```

	filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owed	value
0	2004-04-21 19:07:40	2004-04-19	AAPL	Heinen Nancy R	SVP	S - Sale+OE	28.00	200000	1315	-99%	5600000
1	2004-04-21 19:08:35	2004-04-19	AAPL	Tamaddon Sina	SVP	S - Sale+OE	28.08	678400	6452	-99%	19046284
2	2004-04-21 19:09:19	2004-04-19	AAPL	Rubinstein Jonathan	SVP	S - Sale+OE	28.35	250000	9906	-96%	7087500
3	2004-04-21 19:11:31	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	27.99	84000	4722	-95%	2351280
4	2004-04-21 19:12:19	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	28.21	94000	4722	-95%	2651880
...
599	2024-10-08 18:30:13	2024-10-04	AAPL	Maestri Luca	SVP, CFO	S - Sale	226.52	59305	107788	-35%	13433769
600	2024-11-19 18:30:49	2024-11-18	AAPL	Kondo Chris	Principal Accounting Officer	S - Sale	228.87	4130	15419	-21%	945233
601	2024-11-19 18:31:42	2024-11-15	AAPL	Levinson Arthur D	Dir	S - Sale	227.32	200000	4215576	-5%	45464500
602	2024-12-18 18:30:20	2024-12-16	AAPL	Williams Jeffrey E	COO	S - Sale	249.97	100000	389944	-20%	24997395
603	2025-02-04 18:33:25	2025-02-03	AAPL	Levinson Arthur D	Dir	S - Sale+OE	226.35	1516	4215576	0%	343147

604 rows × 11 columns

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:

```
# separar hora de la columna filing date, solo quiero tener la fecha
```

```
clean_insiders_AAPL['filing_date'] = clean_insiders_AAPL['filing_date'].str.split(' ').str[0]
```

✓ 0.0s

```
clean_insiders_AAPL
```

✓ 0.0s

	filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owned	value
0	2004-04-21	2004-04-19	AAPL	Heinen Nancy R	SVP	S - Sale+OE	28.00	200000	1315	-99%	5600000
1	2004-04-21	2004-04-19	AAPL	Tamaddon Sina	SVP	S - Sale+OE	28.08	678400	6452	-99%	19046284
2	2004-04-21	2004-04-19	AAPL	Rubinstein Jonathan	SVP	S - Sale+OE	28.35	250000	9906	-96%	7087500
3	2004-04-21	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	27.99	84000	4722	-95%	2351280
4	2004-04-21	2004-04-19	AAPL	Cook Timothy D	EVP	S - Sale+OE	28.21	94000	4722	-95%	2651880
...
599	2024-10-08	2024-10-04	AAPL	Maestri Luca	SVP, CFO	S - Sale	226.52	59305	107788	-35%	13433769
600	2024-11-19	2024-11-18	AAPL	Kondo Chris	Principal Accounting Officer	S - Sale	228.87	4130	15419	-21%	945233
601	2024-11-19	2024-11-15	AAPL	Levinson Arthur D	Dir	S - Sale	227.32	200000	4215576	-5%	45464500
602	2024-12-18	2024-12-16	AAPL	Williams Jeffrey E	COO	S - Sale	249.97	100000	389944	-20%	24997395
603	2025-02-04	2025-02-03	AAPL	Levinson Arthur D	Dir	S - Sale+OE	226.35	1516	4215576	0%	343147

604 rows x 11 columns

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:

```
# solo utilizar información desde 2015-01-01 para empatar con info de yahoo finance
```

```
clean_insiders_AAPL_2015 = clean_insiders_AAPL[clean_insiders_AAPL['trade_date'] >= '2015-01-01'].sort_values(by='trade_date', ascending=True)
```

```
clean_insiders_AAPL_2015
```

	filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owned	value
383	2015-01-27	2015-01-23	AAPL	Riccio Daniel J.	SVP	S - Sale	112.76	3804	0	-100%	428955
384	2015-02-20	2015-02-18	AAPL	Jung Andrea	Dir	S - Sale+OE	128.13	40000	14595	-73%	5125200
385	2015-03-06	2015-03-06	AAPL	Maestri Luca	SVP, CFO	S - Sale+OE	128.80	3400	14124	-19%	437920
386	2015-03-11	2015-03-09	AAPL	Maestri Luca	SVP, CFO	S - Sale	128.97	2800	11324	-20%	361116
387	2015-03-20	2015-03-18	AAPL	Maestri Luca	SVP, CFO	S - Sale	128.82	10823	501	-96%	1394219
...
599	2024-10-08	2024-10-04	AAPL	Maestri Luca	SVP, CFO	S - Sale	226.52	59305	107788	-35%	13433769
601	2024-11-19	2024-11-15	AAPL	Levinson Arthur D	Dir	S - Sale	227.32	200000	4215576	-5%	45464500
600	2024-11-19	2024-11-18	AAPL	Kondo Chris	Principal Accounting Officer	S - Sale	228.87	4130	15419	-21%	945233
602	2024-12-18	2024-12-16	AAPL	Williams Jeffrey E	COO	S - Sale	249.97	100000	389944	-20%	24997395
603	2025-02-04	2025-02-03	AAPL	Levinson Arthur D	Dir	S - Sale+OE	226.35	1516	4215576	0%	343147

221 rows × 11 columns

Ilustración 9: filtrar información desde el año 2015

- Unificar los cargos que poseen los directivos

```
# unificar los cargos que ocupan
```

```
multiple_titles = clean_insiders_AAPL_2015.groupby('insider_name')['title'].nunique()
```

```
insiders_with_multiple_titles = multiple_titles[multiple_titles > 1]
```

```
insiders_with_multiple_titles
```

```
insider_name
Williams Jeffrey E    2
Name: title, dtype: int64
```

```
clean_insiders_AAPL_2015.query("insider_name == 'Williams Jeffrey E'")
```

	filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owned	value
401	2015-10-05	2015-10-02	AAPL	Williams Jeffrey E	SVP	S - Sale+OE	110.49	46873	2868	-94%	5178804
412	2016-03-23	2016-03-22	AAPL	Williams Jeffrey E	COO	S - Sale+OE	106.91	268644	3079	-99%	28720730
415	2016-04-05	2016-04-04	AAPL	Williams Jeffrey E	COO	S - Sale+OE	111.43	26284	3079	-90%	2928826
434	2016-10-04	2016-10-03	AAPL	Williams Jeffrey E	COO	S - Sale+OE	112.59	43769	3079	-93%	4927952
496	2018-05-10	2018-05-08	AAPL	Williams Jeffrey E	COO	S - Sale	185.18	15653	155042	-9%	2898551
502	2018-06-12	2018-06-08	AAPL	Williams Jeffrey E	COO	S - Sale	190.94	15653	139389	-10%	2988851
504	2018-07-11	2018-07-09	AAPL	Williams Jeffrey E	COO	S - Sale	190.18	15652	123737	-11%	2976664
506	2018-08-10	2018-08-08	AAPL	Williams Jeffrey E	COO	S - Sale	206.86	15652	108085	-13%	3237714
512	2018-10-05	2018-10-03	AAPL	Williams Jeffrey E	COO	S - Sale	232.33	61998	108085	-36%	14403787
516	2019-05-06	2019-05-02	AAPL	Williams Jeffrey E	COO	S - Sale	210.36	56411	108209	-34%	11866355
522	2019-10-03	2019-10-02	AAPL	Williams Jeffrey E	COO	S - Sale+OE	219.04	67554	122195	-36%	14797004

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.

```

filtered_insiders = clean_insiders_AAPL_2015[clean_insiders_AAPL['title'].str.len() > 3]
filtered_insiders[['trade_date', 'insider_name', 'title']].drop_duplicates()

0.0s

C:\Users\santil\AppData\Local\Temp\ipykernel_17232\1727297835723.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  filtered_insiders = clean_insiders_AAPL_2015[clean_insiders_AAPL['title'].str.len() > 3]

   trade_date  insider_name  title
385  2015-03-06    Maestri Luca  SVP, CFO
386  2015-03-09    Maestri Luca  SVP, CFO
387  2015-03-18    Maestri Luca  SVP, CFO
390  2015-04-16    Maestri Luca  SVP, CFO
391  2015-04-20    Maestri Luca  SVP, CFO
...         ...           ...    ...
593  2024-08-09      Kondo Chris  Principal Accounting Officer
594  2024-08-15      Kondo Chris  Principal Accounting Officer
597  2024-10-02  Adams Katherine L  SVP, GC, Secretary
599  2024-10-04    Maestri Luca  SVP, CFO
600  2024-11-18      Kondo Chris  Principal Accounting Officer

74 rows x 3 columns

# To get unique rows in the DataFrame
filtered_insiders.drop_duplicates()

# Alternatively, to get unique values from a specific column, e.g., 'insider_name'
filtered_insiders['insider_name'].unique()

0.0s

array(['Maestri Luca', 'Sewell D Bruce', 'Kondo Chris',
      'Adams Katherine L.'], dtype=object)

```

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.

```

#reemplazar las líneas que contengan varios títulos con uno solo, esto se evidenció en el EDA
clean_insiders_AAPL_2015['title'] = clean_insiders_AAPL_2015['title'].replace({
    ....'SVP, CFO': 'CFO',
    ....'SVP, GC, Secretary': 'SVP',
    ....'Principal Accounting Officer': 'PAO',
    ....'SVP, Gen'l Counsel, Secretary': 'SVP',
    ....'VP, Corporate Controller': 'VP',
    ....'VP, Controller, PAO': 'VP'
})

0.0s

clean_insiders_AAPL_2015['title'].unique()

0.0s

array(['SVP', 'Dir', 'CFO', 'CEO', 'COO', 'PAO'], dtype=object)

```

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.

```
# estandarizar los tipos de transacción
clean_insiders_AAPL_2015['trade_type'] = clean_insiders_AAPL_2015['trade_type'].replace({
    'S - Sale': 'S',
    'S - Sale+OE': 'S',
    'P - Purchase': 'P',
})
```

✓ 0.0s

clean_insiders_AAPL_2015

✓ 0.0s

	filing_date	trade_date	ticker	insider_name	title	trade_type	price	quantity_of_shares	owned	delta_owned	value
383	2015-01-27	2015-01-23	AAPL	Riccio Daniel J.	SVP	S	112.76	3804	0	-100%	428955
384	2015-02-20	2015-02-18	AAPL	Jung Andrea	Dir	S	128.13	40000	14595	-73%	5125200
385	2015-03-06	2015-03-06	AAPL	Maestri Luca	CFO	S	128.80	3400	14124	-19%	437920
386	2015-03-11	2015-03-09	AAPL	Maestri Luca	CFO	S	128.97	2800	11324	-20%	361116
387	2015-03-20	2015-03-18	AAPL	Maestri Luca	CFO	S	128.82	10823	501	-96%	1394219
...
599	2024-10-08	2024-10-04	AAPL	Maestri Luca	CFO	S	226.52	59305	107788	-35%	13433769
601	2024-11-19	2024-11-15	AAPL	Levinson Arthur D	Dir	S	227.32	200000	4215576	-5%	45464500
600	2024-11-19	2024-11-18	AAPL	Kondo Chris	PAO	S	228.87	4130	15419	-21%	945233
602	2024-12-18	2024-12-16	AAPL	Williams Jeffrey E	COO	S	249.97	100000	389944	-20%	24997395
603	2025-02-04	2025-02-03	AAPL	Levinson Arthur D	Dir	S	226.35	1516	4215576	0%	343147

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.

```
# verificamos que solo existan 2 letras en trade_type
clean_insiders_AAPL_2015['trade_type'].unique()
```

✓ 0.0s

array(['S', 'P'], dtype=object)

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.

```

# creación de pivot table para evaluar si agrupo por cantidad de acciones y fecha de negociación
clean_insiders_AAPL_2015['quantity_of_shares'] = pd.to_numeric(clean_insiders_AAPL_2015['quantity_of_shares'], errors='coerce')

# Create the pivot table
pivot_insiders = clean_insiders_AAPL_2015.pivot_table(
    index=['trade_date', 'trade_type'],
    values='quantity_of_shares',
    aggfunc='sum'
)

```

✓ 0.0s

pivot_insiders

✓ 0.0s

trade_date	trade_type	quantity_of_shares
2015-01-23	S	3804
2015-02-18	S	40000
2015-03-06	S	3400
2015-03-09	S	2800
2015-03-18	S	10823
...
2024-10-04	S	59305
2024-11-15	S	200000
2024-11-18	S	4130
2024-12-16	S	100000
2025-02-03	S	1516

185 rows × 3 columns

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

```

pivot_insiders.reset_index()['trade_date'].duplicated().sum()

```

✓ 0.0s

np.int64(1)

pivot_insiders

✓ 0.0s

trade_date	trade_type	quantity_of_shares
2015-01-23	S	3804
2015-02-18	S	40000
2015-03-06	S	3400
2015-03-09	S	2800
2015-03-18	S	10823
...
2024-10-04	S	59305
2024-11-15	S	200000
2024-11-18	S	4130
2024-12-16	S	100000
2025-02-03	S	1516

185 rows × 3 columns

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

```
# join de tablas de precios de cotización con información de insiders
clean_precios_AAPL.rename(columns={'Date': 'trade_date'}, inplace=True)

fct_precios_insiders_AAPL = clean_insiders_AAPL_2015.merge(clean_precios_AAPL, how='inner', on='trade_date')
```

✓ 0.0s

fct_precios_insiders_AAPL

✓ 0.0s

	trade_date	ticker	insider_name	title	trade_type	price_traded	quantity_of_shares	owned	delta_owned	value	Open	Close	movimiento
0	2015-01-23	AAPL	Riccio Daniel J.	SVP	S	112.76	3804	0	-100%	428955	24.98	25.13	0.15
1	2015-02-18	AAPL	Jung Andrea	Dir	S	128.13	40000	14595	-73%	5125200	28.50	28.75	0.25
2	2015-03-06	AAPL	Maestri Luca	CFO	S	128.80	3400	14124	-19%	437920	28.68	28.27	-0.41
3	2015-03-09	AAPL	Maestri Luca	CFO	S	128.97	2800	11324	-20%	361116	28.58	28.39	-0.19
4	2015-03-18	AAPL	Maestri Luca	CFO	S	128.82	10823	501	-96%	1394219	28.36	28.69	0.33
...
216	2024-10-04	AAPL	Maestri Luca	CFO	S	226.52	59305	107788	-35%	13433769	227.40	226.30	-1.10
217	2024-11-15	AAPL	Levinson Arthur D	Dir	S	227.32	200000	4215576	-5%	45464500	226.15	224.75	-1.40
218	2024-11-18	AAPL	Kondo Chris	PAO	S	228.87	4130	15419	-21%	945233	225.00	227.77	2.77
219	2024-12-16	AAPL	Williams Jeffrey E	COO	S	249.97	100000	389944	-20%	24997395	247.72	250.76	3.04
220	2025-02-03	AAPL	Levinson Arthur D	Dir	S	226.35	1516	4215576	0%	343147	229.74	227.76	-1.98

221 rows x 13 columns

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:

fct_precios_insiders_AAPL
✓ 0.0s

	trade_date	ticker	insider_name	title	trade_type	quantity_of_shares	owned	delta_owned	value	Open	Close	movimiento
0	2015-01-23	AAPL	Riccio Daniel J.	SVP	S	3804	0	-100%	428955	24.98	25.13	0.15
1	2015-02-18	AAPL	Jung Andrea	Dir	S	40000	14595	-73%	5125200	28.50	28.75	0.25
2	2015-03-06	AAPL	Maestri Luca	CFO	S	3400	14124	-19%	437920	28.68	28.27	-0.41
3	2015-03-09	AAPL	Maestri Luca	CFO	S	2800	11324	-20%	361116	28.58	28.39	-0.19
4	2015-03-18	AAPL	Maestri Luca	CFO	S	10823	501	-96%	1394219	28.36	28.69	0.33
...
216	2024-10-04	AAPL	Maestri Luca	CFO	S	59305	107788	-35%	13433769	227.40	226.30	-1.10
217	2024-11-15	AAPL	Levinson Arthur D	Dir	S	200000	4215576	-5%	45464500	226.15	224.75	-1.40
218	2024-11-18	AAPL	Kondo Chris	PAO	S	4130	15419	-21%	945233	225.00	227.77	2.77
219	2024-12-16	AAPL	Williams Jeffrey E	COO	S	100000	389944	-20%	24997395	247.72	250.76	3.04
220	2025-02-03	AAPL	Levinson Arthur D	Dir	S	1516	4215576	0%	343147	229.74	227.76	-1.98

221 rows × 12 columns

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.

```
# tomar el dataframe limpio y guardarlo como un nuevo csv
fct_precios_insiders_AAPL.to_csv('../Dataset/clean/AAPL_fct_precios_insider.csv')
```

✓ 0.0s

Ilustración 19: exportación a un archivo .csv