NTAsset Internship Assignment Logic Explanation

Paramaa Sawanpanyalert

1. Alignment of Ownership Analysis

We download the ownership information from GitHub. The first dataframe (df) contains the provided data for task 1 (Alignment of Ownership), while the second dataframe (df2) is utilized for task 2 (Insider Activity), which consists of 5 sheets of CSV.

The next five steps will outline the process of aligning ownership analysis.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 342 entries, 0 to 342
Data columns (total 26 columns):
 #   Column                                                 Non-Null Count  Dtype
---  ------                                                 --------------  -----
 0   SP_COMPANY_NAME                                        342 non-null    object
 1   ciqticker                                              342 non-null    object
 2   SHARES_OWNED_INSIDERS                                  342 non-null    int64
 3   SP_NO_FLOAT_SHARES                                     342 non-null    int64
 4   SHARES_OWNED_INSTN_BANKS/INVEST_BANKS                  342 non-null    int64
 5   SHARES_OWNED_INSTN_CHARITABLE_FOUNDATIONS              342 non-null    int64
 6   SHARES_OWNED_INSTN_CORP_PENSION_SPONSORS               342 non-null    int64
 7   SHARES_OWNED_INSTN_EDUCATIONAL/CULTURAL_ENDOWMENTS     342 non-null    int64
 8   SHARES_OWNED_INSTN_FAMILY_OFFICES/TRUSTS               342 non-null    int64
 9   SHARES_OWNED_INSTN_GOVT_PENSION_SPONSORS               342 non-null    int64
 10  SHARES_OWNED_INSTN_INSURANCE_COMPANIES                 342 non-null    int64
 11  SHARES_OWNED_INSTN_INVEST_MGRS                         342 non-null    int64
 12  SHARES_OWNED_INSTN_REIT                                342 non-null    int64
 13  SHARES_OWNED_INSTN_UNCLASSIFIED                        342 non-null    int64
 14  SHARES_OWNED_INSTN_UNION_PENSION_SPONSORS              342 non-null    int64
 15  SHARES_OWNED_INSTN_HEDGE_FUND_MGRS_LESS5PCT_STAKE      342 non-null    int64
 16  SHARES_OWNED_INSTN_SOVEREIGN_WEALTH_FUNDS_LESS5PCT_STAKE  342 non-null  int64
 17  SHARES_OWNED_INSTN_VC/PE_FIRMS_LESS5PCT_STAKE          342 non-null    int64
 18  SHARES_OWNED_STRATEGIC_HEDGE_FUND_MGRS_MORE_EQUAL5SHARES_STAKE  342 non-null  int64
 19  SHARES_OWNED_STRATEGIC_VC/PE_FIRMS_MORE_EQUAL5SHARES_STAKE  342 non-null  int64
 20  SHARES_OWNED_STRATEGIC_SOVEREIGN_WEALTH_FUNDS_MORE_EQUAL5SHARES_STAKE 342 non-null  int64
 21  SHARES_OWNED_STRATEGIC_COMPANY_CONTROLLED_FOUNDATIONS  342 non-null    int64
 22  SHARES_OWNED_STRATEGIC_ESOP                            342 non-null    int64
 23  SHARES_OWNED_STRATEGIC_STATE_OWNER                     342 non-null    int64
 24  SHARES_OWNED_STRATEGIC_CORPORATIONS_PRIVATE            342 non-null    int64
 25  SHARES_OWNED_STRATEGIC_CORPORATIONS_PUBLIC             342 non-null    int64
dtypes: int64(24), object(2)
memory usage: 72.1+ KB
```

1. Since the data is stored in Python object format, we need to convert it to an integer (int64) type using the convert_to_int function. Also, NaN values should be replaced with 0 to make calculations easier.

2. We find out the total number of companies' shares by adding up how many each shareholder has.

- *Creating new column name 'total' to collect number of shares owned by every stakeholder*

ciqticker	SP_COMPANY_NAME	SHARES_OWNED_INSIDERS	total
PSE:UPSON	Upson International Corp.	2445580300	3125001300
SGX:5CP	Silverlake Axis Ltd.	1958070394	2568790352
SEHK:6993	Blue Moon Group Holdings Limited	4449569000	5913962724
IDX:CMRY	PT Cisarua Mountain Dairy Tbk	5965562100	8151304122
NSEI:PDSL	PDS Limited	101430014	138643214
...	...	...	...
TPEX:5530	Lungyen Life Service Corporation	0	14376188
HNX:IDC	IDICO Corporation - JSC	0	352689887
IDX:INTP	PT Indocement Tunggal Prakarsa Tbk	0	3673147414
HOSE:VRE	Vincom Retail Joint Stock Company	0	280580208
SET:CPNREIT	CPN Retail Growth Leasehold REIT	0	2866463012

342 rows × 3 columns

3. To figure out the insider ownership percentage, we divide the number of shares owned by insiders by the total number of shares issued.

$$\%Insider\ ownership = \frac{\#\ Shares\ insider\ held}{\#\ Outstanding\ shares}$$

4. Then, we make these percentages easier to compare by putting them on a scale from 1 to 10 using the score_convertor function.

Score convertor function

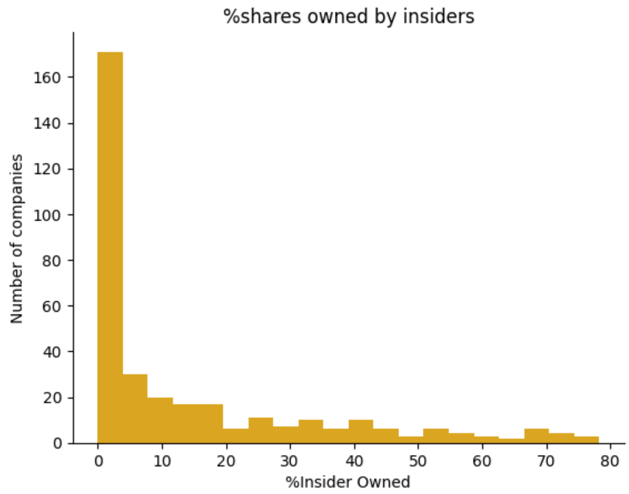
$$Score = 1 + \frac{(ownership\ percentage - \min percentage) * 9}{(\max\ \ \text{percentage} - \min\ \ percentage)}$$

With this formula we can normalize ownership percentage to score range from 1 to 10, which derives from typical normalization formula, as shown in figure below.

$$x_{normalized} = \frac{(x - x_{minimum})}{(x_{maximum} - x_{minimum})}$$

The reason why we use **<u>normalization</u>** over standardization is that the data <u>is not normally distributed</u>, as shown in the figure below.

Standardization is useful when the distribution of the data is Gaussian (normal distribution) or when the algorithm used in the analysis assumes that the data is normally distributed, as it helps to center the data around 0 with a standard deviation of 1.



Normalization is useful when the distribution of the data does not follow a Gaussian distribution (non-normal distribution) or when the algorithm used in the analysis does not assume any specific distribution of the data.

5. After that, we calculate the ownership score and add it to the last column of the dataframe. We set up DFtask1 with just the ticker, company name, and ownership score for later merging.

	SP_COMPANY_NAME	ownership_percentage	ownership_score
ciqticker			
PSE:UPSON	Upson International Corp.	78.258537	10.000000
SGX:5CP	Silverlake Axis Ltd.	76.225387	9.766181
SEHK:6993	Blue Moon Group Holdings Limited	75.238367	9.652670
IDX:CMRY	PT Cisarua Mountain Dairy Tbk	73.185370	9.416568
NSEI:PDSL	PDS Limited	73.159018	9.413538
...	...	...	...
TPEX:5530	Lungyen Life Service Corporation	0.000000	1.000000
HNX:IDC	IDICO Corporation - JSC	0.000000	1.000000
IDX:INTP	PT Indocement Tunggal Prakarsa Tbk	0.000000	1.000000
HOSE:VRE	Vincom Retail Joint Stock Company	0.000000	1.000000
SET:CPNREIT	CPN Retail Growth Leasehold REIT	0.000000	1.000000

342 rows × 3 columns

2. Insider Activity Analysis

We can explain the entire analysis in 8 main steps, which are:

1. The ownership information is imported into dataframe 2 (df2).
2. Data cleaning involves replacing NaN values with 0 to facilitate calculations.
3. Insider holdings for each company across multiple time periods are aggregated, treating all insiders collectively as if they exchange shares among themselves, while still maintaining a constant total share count.
4. Holding position data from five distinct time periods are integrated into a unified dataframe (df3) to enhance clarity and visualization.

	period Q1	period Q2	period Q3	period Q4	period Q5
SEHK:425	4.500720e+08	4.500720e+08	4.500920e+08	4.501940e+08	4.501940e+08
SEHK:1992	2.553945e+07	2.553945e+07	2.553945e+07	2.587275e+07	2.601792e+07
SGX:5CP	1.952552e+09	1.957952e+09	1.957952e+09	1.957750e+09	1.957750e+09
SEHK:3315	3.646054e+08	3.646054e+08	3.646054e+08	3.646054e+08	3.646054e+08
SET:FPT	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	2.030820e+07
...	...	...	...	...	...
SEHK:1836	8.113329e+07	8.348329e+07	8.665329e+07	8.665329e+07	8.742279e+07
NSEI:ACCELYA	1.751150e+05	1.751150e+05	1.176300e+05	1.176300e+05	1.176300e+05
HOSE:GEG	3.063516e+07	3.063516e+07	1.987902e+06	1.987902e+06	1.987902e+06
NSEI:QUESS	2.288637e+07	2.348637e+07	2.348637e+07	2.348637e+07	2.348637e+07
SET:PRINC	2.583987e+09	2.610200e+09	2.610200e+09	2.610200e+09	2.610200e+09

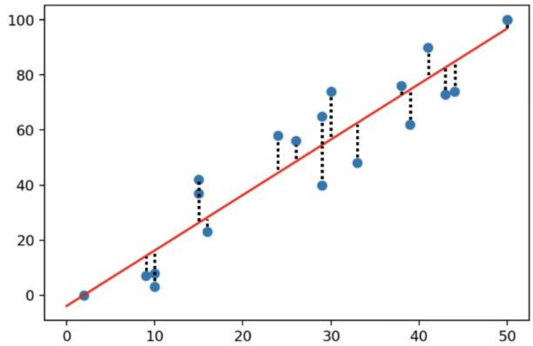
347 rows × 5 columns

5. The insider_activity function facilitates the examination of the holding position value for a specific company.

6. To recognize trends in insider activity, the slope of the **best fit line** of provided numerical data serves as a reliable indicator, obtained through the ***"polyfit" function***, which determines the optimal fit line for the given dataset.

Polyfit function

In NumPy, the polyfit function is used to fit a polynomial of a specified degree to a set of data points using the method of least squares. It returns the coefficients of the polynomial that best fits the data.



Here's the syntax of the polyfit function:

$$numpy.polyfit(x,\ y,\ deg,\ rcond = None,\ full = False,\ w = None,\ cov = False)$$

- *x:* The independent variable, or the x-coordinates of the data points.
- *y:* The dependent variable, or the y-coordinates of the data points.
- *deg:* The degree of the polynomial to be fitted.
- *rcond:* Relative condition number of the fit. Default is len(x)*eps, where eps is the machine precision.
- *full:* If True, returns additional outputs. Default is False.
- *w:* Weights to apply to the y-coordinates of the data points.
- *cov:* If True, return the covariance matrix.

The solution minimizes the squared error

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

in the polynomial equations

```
x[0]**n * p[0] + ... + x[0] * p[n-1] + p[n] = y[0]
x[1]**n * p[0] + ... + x[1] * p[n-1] + p[n] = y[1]
...
x[k]**n * p[0] + ... + x[k] * p[n-1] + p[n] = y[k]
```

In this case, we use only 1- degree linear equations fitting the given points

$$y = mx + c$$

We focus on m as being a slope of the best fit line.

7. Subsequently, the slope is normalized to a scale ranging from 1 to 10 using the score_convertor function.

8. The insider score is then assigned to dftask2, preparing it for integration with dftask1 into the final dataframe.

ciqticker	slope	insider_activity_score
SGX:8AZ	1.630333e+08	10.000000
IDX:HEAL	8.994061e+07	8.945874
SET:MINT	3.539680e+07	8.159255
IDX:CSAP	2.120656e+07	7.954606
NSEI:PAYTM	1.984053e+07	7.934906
...	...	...
IDX:MTDL	-5.463846e+07	6.860786
PSE:UPSON	-1.368484e+08	5.675173
IDX:BUKA	-1.545171e+08	5.420359
IDX:ULTJ	-3.126204e+08	3.140228
IDX:GOTO	-4.610229e+08	1.000000

347 rows × 2 columns

3. Aggregation of Ownership Factors:

- The final scores are aggregated using the **median** method.
- It is important to regard the score as ordinal data due to the inconsistent intervals between each score, albeit retaining an ordered structure. Given that the rating or score represents ordinal data, the suitable statistical measures include mode and median.
- The selection of the median over the mode is justified as it offers a more robust measure of central tendency. Then, we also apply normalization to the final score column making it ranged from 1 to 10 using the score_convertor function.

Here's the final result table, sorted by final score(ranget right column), consisting of 5 columns, including

1. Company name
2. CIQ ticker
3. Alignment of ownership score
4. Insider activity score
5. Ownership score

ciqticker	SP_COMPANY_NAME	ownership_score	insider_activity_score	final_score
SGX:8AZ	Aztech Global Ltd.	8.915041	10.000000	10.000000
SGX:5CP	Silverlake Axis Ltd.	9.766181	7.663472	9.188949
SEHK:6993	Blue Moon Group Holdings Limited	9.652670	7.648810	9.118964
NSEI:PDSL	PDS Limited	9.413538	7.645357	8.986508
IDX:HEAL	PT Medikaloka Hermina Tbk	8.045342	8.945874	8.949554
...	...	...	...	...
IDX:CLEO	PT Sariguna Primatirta Tbk	1.119961	7.231476	4.232066
IDX:SMRA	PT Summarecon Agung Tbk	1.166952	7.122967	4.198476
IDX:ULTJ	PT Ultrajaya Milk Industry & Trading Company Tbk	4.788986	3.140228	4.001524
IDX:BUKA	PT Bukalapak.com Tbk.	1.227283	5.420359	3.301761
IDX:GOTO	PT GoTo Gojek Tokopedia Tbk	1.432113	1.000000	1.000000

342 rows × 4 columns

4. Other factors need to be considered for LT investment perspective:

1. **Share Buybacks**: When a company repurchases its own shares from the market. Share buybacks can signal confidence from management and may indicate that the company believes its stock is undervalued. This can be calculated by ***comparing number of outstanding shares*** in different period of time.

$$\Delta \text{ in share buyback} = \#\text{ Outstanding shares}_{\text{Current}} - \#\text{ Outstanding shares}_{\text{previos}}$$

- Number of outstanding shares of 2 different periods of time are required for compute this factors.

---

2. **Short Interest**: The percentage of a company's shares that have been sold short by investors betting that the stock price will decline. High short interest can indicate bearish sentiment, while low short interest may suggest bullish sentiment. It derives from number of the number of shorted shares divided by the number of shares outstanding

$$\text{Short Interest} = \frac{\#\text{ Shares sold short}}{\#\text{ Outstanding shares}}$$

- This will require number of shares sold short for calculating short interest value.

---

3. **Float**: The number of shares available for trading by the public. Low float stocks can be more volatile due to limited supply, while high float stocks may have more stable prices

$$\%\text{ Free float} = \frac{\#\text{ Free float}}{\#\text{ Outstanding shares}}$$

- This value can be calculated based on given data.
- For long-term investments, we might prefer price stability over short spikes in price due to a lack of float.

---

4. **Ownership of institutional investors**: The percentage of a company's shares held by institutional investors, such as mutual funds, pension funds, and hedge funds. High institutional ownership can indicate confidence from professional investors and may influence stock price movements.

$$\%\text{Institutional holding} = \frac{\#\text{ Shares held by institutions}}{\#\text{ Outstanding shares}}$$

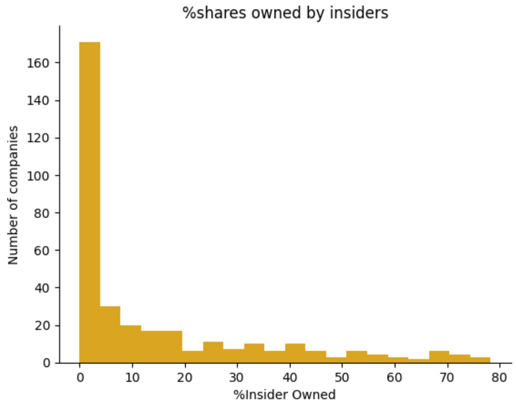
- This value can be calculated based on given information.

In addition, I will show calculations of both float and institutional holding score also in the python notebook.

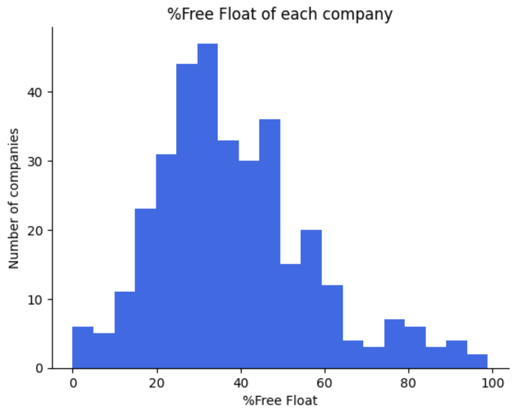
Here's a result of combining those 4 factors into final score:

ciqticker	SP_COMPANY_NAME	ownership_score	insider_activity_score	float_score	inst_score	final_score
SGX:8AZ	Aztech Global Ltd.	8.915041	10.000000	3.504305	1.228908	10.000000
SET:SNNP	Srinanaporn Marketing Public Company Limited	7.216953	7.722741	4.942569	1.233949	9.757830
SET:PSP	P.S.P. Specialties Public Company Limited	6.009204	7.648770	6.134324	1.000000	9.742923
HOSE:TVS	Thien Viet Securities Joint Stock Company	5.809179	7.675252	6.292539	1.000000	9.703954
KOSDAQ:A129890	Abko Co., Ltd.	5.415023	7.640440	6.568008	1.035913	9.593331
...	...	...	...	...	...	...
IDX:BRIS	PT Bank Syariah Indonesia Tbk	1.008027	7.689366	1.879098	1.168091	1.264639
PSE:FB	San Miguel Food and Beverage, Inc.	1.001176	7.648779	1.988881	1.005004	1.214957
IDX:PANI	PT Pantai Indah Kapuk Dua Tbk	1.000000	7.648770	1.982232	1.003735	1.207577
IDX:ADCP	PT Adhi Commuter Properti Tbk	1.000131	7.649436	1.909585	1.000000	1.136507
IDX:ADMF	PT Adira Dinamika Multi Finance Tbk	1.000000	7.648770	1.717699	1.045559	1.000000

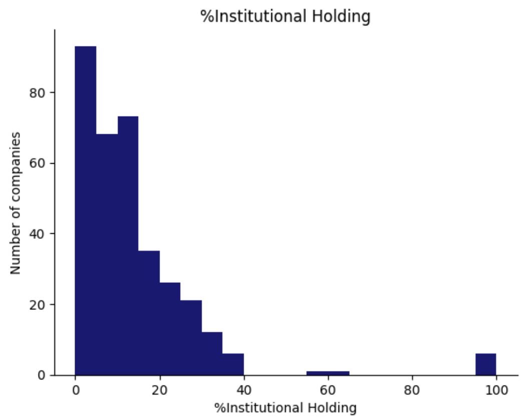
342 rows × 6 columns



Distribution of number of company categorized by percentage of insider ownership.



Distribution of number of company categorized by percentage of free float



Distribution of number of company seperated by percentage of institutional holding