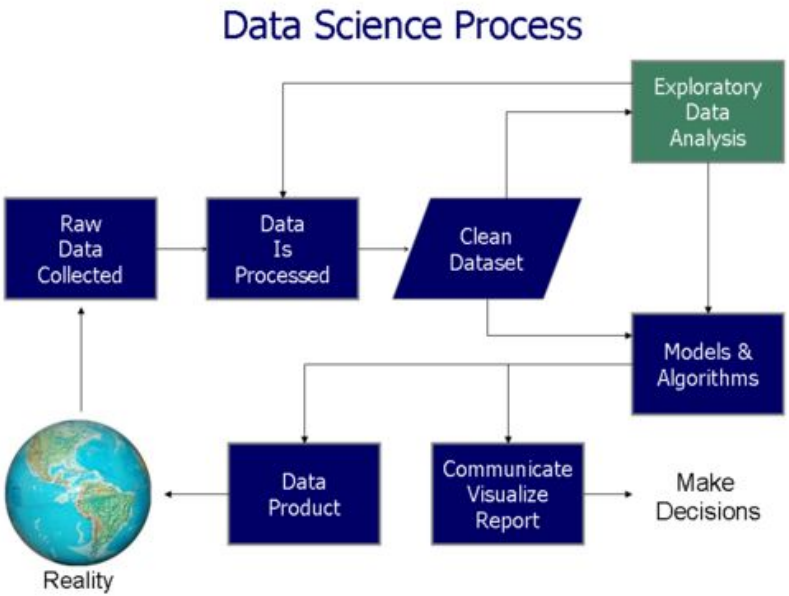
Exploring Market Data

By John Lee

What is Exploratory Data Analysis (EDA)?



- An initial analysis of the data set that summarized its main characteristic
- Detects anomalies or unexpected o observations in the data
- Handles missing or incomplete data
- Ensure well-engineered features
- Guides which statistical techniques to use
- Understands data assumption

Simple (Arithmetic) Returns vs Log Returns

Simple (Arithmetic) Returns:

- The proportional profit of a trade or investment.
- Can be used to calculate daily, weekly, monthly, and yearly change in price.

$$\text{Simple Return} = \frac{P_t - P_{t-1}}{P_{t-1}}, \text{ where } P_t \text{ is the price at time t}$$

Log Returns:

- Always smaller than simple returns.
- Better reflects portfolio changes in reality.
- Best used to calculate minute, hourly, daily, or any other short-term change in price.

$$\text{Log Return} = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1}), \text{ where } P_t \text{ is the price at time t}$$

Why Log Return?

- Because it has the convenient property that a k-period return is simply the sum of the single-period log returns.

$$\text{Log return for the most recent } k \text{ periods}$$

$$\begin{aligned}
&= (\log(P_t) - \log(P_{t-1})) + (\log(P_{t-1}) - \log(P_{t-2})) + \ldots + (\log(P_{t-k+1}) - \log(P_{t-k})) \\
&= \log(P_t) - \log(P_{t-k}) \\
&= \log\left(\frac{P_t}{P_{t-k}}\right)
\end{aligned}$$

Why Log Return?

- Better describes reality.

Example:

You've invested $100 in year 1, it grows to $200 in year 2, then comes back to year $100 in year 3.

$$\begin{aligned}\text{Average Simple Return} &= \frac{(\frac{\$200-\$100}{\$100} + \frac{\$100-\$200}{\$200})}{2} * 100\% \\ &= \left(\frac{1-0.5}{2}\right) * 100\% \\ &= 25\%\end{aligned}$$

$$\begin{aligned}\text{Log Return } &= (\log(\frac{\$200}{\$100}) + \log(\frac{\$100}{\$200}) * 100\% \\ &= (\log(\$200) - \log(\$100) + \log(\$100) - \log(\$200)) * 100\% \\ &= 0\%\end{aligned}$$

Types of Visualizations

- Time Plots
- Scatter Plots
- Correlation Heatmap
- Box Plots
- Histograms

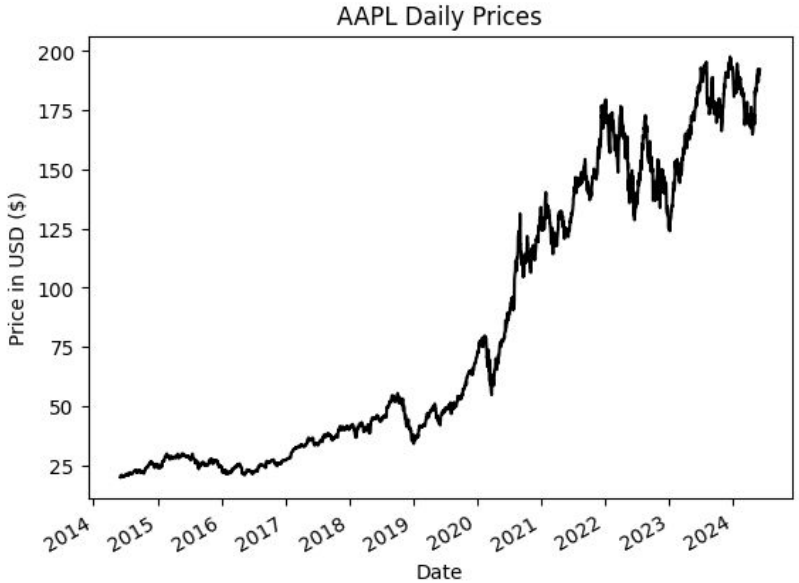
Time Plots

- Observe price trend over time.
- Identify recurring patterns or seasonal effects in financial data.
- Assess how volatility of the asset varies through time.

Time Plots

Example:

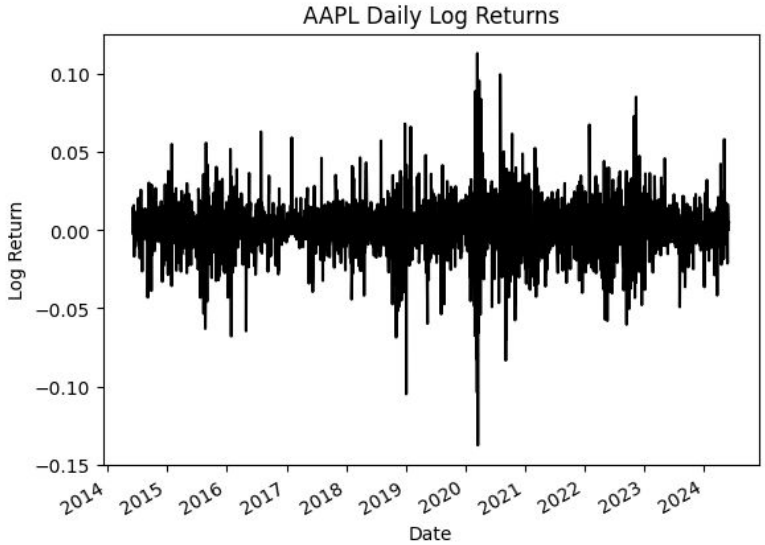
- Apple's stock (Ticker: AAPL) price from 6/1/2014 to 6/1/2024, approximately 2516 trading days.
- Shows price trend over time formed by impulses and corrections.



Time Plots

Example:

- Apple's stock (Ticker: AAPL) log returns from 6/1/2014 to 6/1/2024, approximately 2520 trading days.
- More stable relative to price trend.
- Shows volatility clustering - higher returns tends to stay high, vice versa.



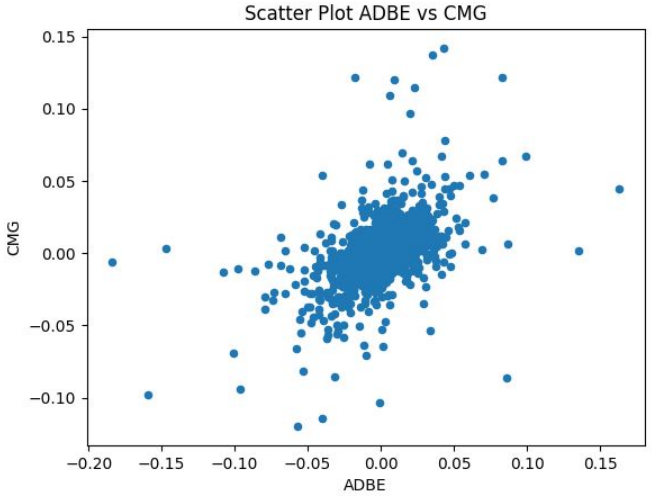
Scatter Plots

- Shows the relationship between two assets' performances.
    - **Direction:** describes whether an asset will react positively, negatively, or neutral relative to other assets changes.
    - **Linearity:** describes whether the data points approximately form a straight line.
    - **Strength:** describes how closely the data points adhere to the overall straight line/ trend pattern.
    - **Outliers:** observations that fall far away from the general trend or cluster of data points.
    - **Clusters:** spot whether the points form subgroups due to similar characteristics or behaviors.

Scatter Plots

Example:

- Would Adobe's (Ticker: ADBE) stock can be a good hedge for an investment in Chipotle (Ticker: CMG)?
- We plotted Adobe's log return against Chipotle's log return.
- The positive relationship suggests that it is relatively risky to hold both stocks at the same time b/c if one stock crash, the other would likely crash as well.



Exploring Market Data

By John Lee