

# NASDAQ Stock Market Analysis Report (2021)

## 1. Introduction

The objective of this analysis was to examine the performance of the top 10 NASDAQ-listed companies during the fiscal year 2021. The project involved fetching historical daily data, visualizing price trends, and applying unsupervised machine learning (K-Means clustering) to group companies based on their stock price performance on a monthly basis.

## 2. Implementation Steps

### Step A: Data Retrieval (Historical Stock Values)

**Objective:** Retrieve daily historical stock values for the top 10 NASDAQ companies for the year 2021.

- **Implementation:** A Python script was developed using the yfinance library to download historical market data.
- **Parameters:**
  - **Timeframe:** January 1, 2021, to December 31, 2021.
  - **Companies Selected:** AAPL, MSFT, AMZN, NVDA, GOOGL, META, TSLA, NFLX, AVGO, and ASML.
  - **Data frequency:** Daily.
- **Outcome:** Successful retrieval of "Adjusted Close" prices for all 10 companies, ensuring dividends and splits were accounted for.

### Step B: Trend Analysis (Line Charts)

**Objective:** Visualizing the stock value trends for each company.

- **Implementation:** Using matplotlib, individual line charts were generated for each company.
- **Features:**

- The X-axis represents the timeline (2021), and the Y-axis represents the Adjusted Close Price (\$).
- Key statistics were calculated and overlaid on each chart, including Minimum Price, Maximum Price, and Percentage Change over the year.
- A normalized comparison chart was also generated to compare relative growth, base-lining all stocks to 100 at the start of the year.
- **Observation:** Companies like NVDA and GOOGL showed significant positive trends (approx +125% and +69% respectively), while others like AMZN remained relatively flat (+5.8%).

## Step C: K-Means Clustering

**Objective:** Group companies based on stock market values using the K-Means algorithm.

### *i. Data Preparation and Grouping*

- **Data Extraction:** The script resampled the daily data to extract the last trading price of every month for each company.
- **Standardization:** To ensure the clustering was not biased by the absolute magnitude of stock prices (e.g., AMZN at ~\$170 vs. ASML at ~\$700), the StandardScaler from sklearn was applied to normalize the features before clustering.
- **Clustering Configuration:** The K-Means algorithm was configured with n\_clusters=3 (Optimal K) to categorize stocks into three distinct groups (e.g., Low, Medium, and High price tiers relative to the group).
- **Validation:** Silhouette scores were calculated for each month to quantify the separation quality of the clusters. Scores consistently ranged between 0.490 and 0.672, indicating a reasonable structure.

### *ii. Visualization (Bubble Charts)*

**Objective:** Create bubble charts per month where bubbles represent clusters.

- **Implementation:**
  - A grid of bubble charts was generated for all 12 months of 2021.
  - **Representation:** Each bubble represents a specific Cluster (C0, C1, C2).

- **Bubble Size:** The size of the bubble is proportional to the number of companies within that cluster (e.g., in Jan 2021, Cluster C0 contained 5 companies, making it the largest bubble).
- **Axes:** The X-axis displays the Average Price (\$) of the cluster, while the Y-axis separates the Cluster IDs for clarity.
- **Detailed View:** Detailed scatter plots were also generated to identify exactly which companies belonged to which cluster. For example, in the early months, "high price" stocks like ASML often formed their own cluster, while lower-priced stocks (at the time) like AAPL, NVDA, and CSCO grouped together.

### 3. Conclusion

The automated script successfully processed the financial data for 2021. The K-Means analysis revealed distinct price-tier groupings among the top NASDAQ companies. The bubble charts effectively visualized the shifting market dynamics, showing how companies migrated between clusters as their stock prices fluctuated throughout the year.