## Work Trial Task: Cross-Impact Analysis of Order Flow Imbalance (OFI)

**Objective:**

The goal of this task is to assess your ability to process high-frequency equity market data, compute Order Flow Imbalance (OFI) metrics across multiple levels of the Limit Order Book (LOB), and evaluate cross-asset impacts on short-term price changes. You will implement the methodologies outlined in the paper **"Cross-Impact of Order Flow Imbalance in Equity Markets"** to complete this task. Additionally, you will demonstrate your skills in creating a well-documented GitHub repository and a concise LaTeX-based report summarizing your approach and findings.

---

## Task Description:

You are provided with a subset of equity market data containing order book updates and trades for several highly liquid stocks. Your objective is to:

1. **Compute OFI Metrics:**
    - Derive multi-level OFI metrics (up to 5 levels) for each stock in the dataset.
    - Integrate these multi-level OFIs into a single metric using Principal Component Analysis (PCA) or another dimensionality reduction method.
2. **Analyze Cross-Impact:**
    - Examine the contemporaneous cross-impact of OFI on short-term price changes across stocks.
    - Evaluate the predictive power of lagged cross-asset OFI on future price changes (e.g., 1-minute and 5-minute horizons).
3. **Quantify Results:**
    - Use regression models to assess the explanatory power of contemporaneous OFI and predictive power of lagged OFI.
    - Compare self-impact (within the same stock) vs. cross-impact (between stocks) in your models.
4. **Visualization and Reporting:**
    - Create visualizations (e.g., heatmaps, scatter plots) to illustrate cross-impact relationships and OFI trends.
    - Summarize your findings in a concise report.

---

## Dataset Description:

We recommend you work with high-frequency data for 5 Nasdaq 100 stocks, representing various sectors:

**Example Stocks:**

- **Tech:** AAPL
- **Healthcare:** AMGN
- **Consumer Discretionary:** TSLA
- **Financials:** JPM
- **Energy:** XOM

**Data To Use:**

In order to implement this paper you will need at least 5 levels of order book depth (L2), over a time period of 1 week / 1 month. There are a few data sources that can provide this data, but we recommend Databento. In order to get this data:

1. Navigate to https://databento.com/ and sign up for an account
2. Once you have an account you can either
    a. To download files individually → Navigate to the "Nasdaq TotalView-ITCH" catalog, specify the "MBP-10" schema, select a time frame, and click
    b. To download via API → Navigate to API key from the side bar, and follow documentation to pull the data above from "Nasdaq TotalView-ITCH"
3. Check out data documentation for MBP-10 schema [here]

*Note: When creating an account you will be asked for your payment information, but you will get $125 in free credits to use, which should be more than enough for this dataset.*

---

## Deliverables:

1. **Public GitHub Repository**
    - **Code and Scripts:**
        - Python scripts for data preprocessing, OFI calculation, and cross-impact analysis.
        - Modularized code with clear documentation and comments.

**Folder Structure:**

```
project/
├── data/              # Placeholder or sample datasets (no proprietary data)
├── notebooks/         # Jupyter notebooks for analysis and results
├── scripts/           # Python scripts for modular implementations
├── results/           # Outputs (e.g., figures, tables, analysis results)
├── README.md          # Detailed instructions on how to run the code
```

    - └── requirements.txt  # List of Python packages used in the project
    - **README File:**
        - Description of the task and steps to run the analysis.

        - Brief summary of the findings.
2. **Research Write-Up in LaTeX:**
    - A **1-2 page PDF** summarizing the methodology, findings, and key insights.
    - **Sections to Include:**
        - **Title Page:** Project title, your name, and date.
        - **Abstract:** Overview of the task and key findings.
        - **Methodology:** Explanation of OFI calculation, PCA (if used), and cross-impact analysis.
        - **Results:** Key findings with relevant visualizations (e.g., heatmaps, scatter plots).
        - **Discussion:** Interpretation of results and implications for trading strategies.
        - **Conclusion:** Summary of results and proposed next steps.
        - **References:** List of any research papers or sources used.
3. **Visualizations:**
    - Include clear and informative charts/plots for:
        - OFI metrics for individual stocks.
        - Cross-impact relationships between stocks.

---

## Submission Guidelines:

- **GitHub Repository:**
    - Provide a public GitHub link to your repository with the folder structure outlined above.
    - Ensure all code is reproducible, and dependencies are clearly listed in `requirements.txt`.
- **PDF Report:**
    - Submit a well-formatted LaTeX-based PDF summarizing your methodology and findings.
- **Deadline:**
    - The task must be completed and submitted within **5 days** of receiving the dataset.
- **Submission:**
    - Email your GitHub link and PDF report to [submission email].

---

## Evaluation Criteria:

1. **Data Handling and Preprocessing:**
    - Ability to clean and manage high-frequency datasets effectively.
2. **Feature Engineering:**

    - Accuracy and quality of multi-level OFI calculations and integration.
3. **Cross-Impact Analysis:**
    - Depth and rigor of contemporaneous and predictive analysis.
4. **Code Quality:**
    - Modularity, readability, and adherence to best practices.
5. **Reporting:**
    - Clarity, structure, and professionalism of the LaTeX report.
6. **Bonus (Optional):**
    - Sector-level insights or additional predictive modeling using advanced techniques (e.g., random forests, LSTMs).

---

## Bonus Task (Optional):

Explore how cross-impact varies by stock sector (e.g., tech vs. healthcare) or investigate whether combining OFI with other features improves prediction accuracy.