Justification

In designing the relational database schema for this project, our primary objective was to ensure that the structure was logically sound, efficient for querying, and for the database to be able to be easily extendable to accommodate future needs. To achieve this, we followed the principles of database normalization and ensured that our schema adheres all the way up to Third Normal Form (3NF). This level of normalization helps reduce redundancy, prevent anomalies during data operations, and improve overall data consistency and integrity.

To begin, the schema satisfies First Normal Form (1NF) because all tables contain only atomic values with no repeating groups or multivalued attributes. Each row is uniquely identified by a primary key, such as id in the Stocks, News, and Sentiment tables, as well as ticker in the StockMetaData table.

Additionally, the schema meets Second Normal Form (2NF) since all non key attributes are fully dependent on the entire primary key. Each table uses a single column primary key, so there are no partial dependencies. For instance, in the Stocks table, fields like date, openPrice, closePrice, highPrice, lowPrice, and volume all depend on the full primary key id, not on just a part of it. Because each table uses a single column primary key, there are no partial dependencies, and all attributes are fully functionally dependent on their table's primary key.

Lastly, the design complies with Third Normal Form (3NF) because there are no transitive dependencies. Each non key attribute is directly dependent only on the primary key and not on other non key fields. For example, in the StockMetaData table, attributes such as company, industry, sector, marketCap, and hq all directly describe the ticker, which serves as the table's primary key. There are no intermediate non key attributes acting as dependencies for others, which confirms that the schema is fully normalized up to 3NF.

Overall, the schema is well normalized and designed to avoid data duplication, promote data integrity, and support efficient querying and relational operations. This normalized structure also supports future extensibility. If we wanted to include additional information such as quarterly earnings reports, analyst ratings, or user generated comments on articles, we could add new tables and reference existing keys without needing to redesign our entire schema. This collective effort is a direct benefit of maintaining a 3NF design and reinforces its importance in relational databases.