1. What automated approaches can you use?

* Data Integration: Automate the merging of multiple datasets if the active ingredients, buyer information, and contract details come from different sources.
* Data Cleaning: Automate tasks such as filling missing values, correcting data types (e.g., converting dates), and removing or flagging outliers based on predefined rules. Automate logic based columns.
* Reporting: Develop automated reports and dashboards using tools like Tableau or Power BI.
* Predictive Modelling: Develop models to forecast price based on historical data, and automate the retraining and validation of these models over time.
* Anomaly Detection: Implement automated systems to flag unusual changes in price that could indicate errors.

1. What manual tasks would you perform?

* Quality Assurance: Periodically review automated outputs for accuracy and relevance.
* Complex Decision Making: While analysis can be automated, the decisions based on this analysis often require human judgment, considering factors beyond the data.
* Data Exploration and Custom Analysis: Engage in exploratory data analysis manually to identify potential areas of opportunity or concern that automated tools might not pick up.

1. How would you improve this process long term and how would you build your roadmap?

* Data Architecture: Organizing data into distinct datasets based on unique identifiers, allowing for easy integration with other data sources as needed including a data dictionary. E.g., contract data, participant data etc.
* Incorporate Machine Learning: Enhance your data analysis capabilities by incorporating more advanced machine learning models that can learn from data trends over time.
* Upgrade Infrastructure: As data volume grows, consider upgrading your data storage and processing infrastructure to handle larger datasets more efficiently.
* Continuous Learning and Adaptation: Keep the system adaptive to new data sources and analytical methods, incorporating best practices and emerging technologies.

1. Would you change anything if you would need to scale this process from a few SKU's to hundreds and thousands.

* Cloud Services and Big Data Technologies: Use scalable cloud services (like AWS, Google Cloud) and big data technologies (like Hadoop or Spark) to handle increased data volume and processing needs.
* Enhanced Automation: Increase the level of automation in data processing and analysis to handle larger datasets without a proportional increase in manual effort.
* Robust Data Governance: Implement strict data governance frameworks to ensure data quality and compliance as you scale up.