

Data Modelling

We need to predict the stock levels of products based on sales data and sensor data on an hourly basis in order to more intelligently procure products from suppliers. So we will be using timestamp ,product_id for sure as its on an hourly basis

Strategic Planning

- We will be using transaction_id, timestamp, product_id, category, unit_price, quantity , payment_type from sales model , and temperature from sensor storage temperature, estimated_stock_pct from sensor stock levels model respectively
- Mainly time_stamp and product_id is important to be included in dataframe which is to be accessed later
- Here features will be transaction_id, timestamp, product_id, temperature, estimated_stock_pct ,category, unit_price, quantity, payment_type
- We ensure that categories are fairly uniformly distributed for better classification if needed via payment_type
- Else we are to predict the stock levels we classify or group them by estimated stock pct with respective product_id and timestamp

Summary

- We can observe that timestamp , product_id are the basis for grouping and to predict the estimated stock levels we can use linear regression with polynomial features as we have many features like category, unit_price, quantity etc...

Product_id	timestamp	category	Unit_price	Transaction_id	quantity	Estimated_stock_pct	Payment_type	temperatu re