DATA CONSULTANT CASE STUDY

NEWTON

THE SCENARIO

You have just started a diagnostic for one of the biggest food retailers in the UK. Over the past decade they have been incredibly successful and grown enormously, but they are facing difficulties and finding it hard to turn a profit.

Every year they reduce or throw out over £500 million of products that has come to the end if its life. This represents close to 10% of their annual turnover. Alongside this, up to 10% of the time a product a customer is after is not available in a store.

Reducing waste and improving product availability would greatly improve profitability. However, no one knows what the biggest causes of these two issues are and where they need to begin to solve the issues. From the time a product leaves a supplier, goes through the retailers supply chain and ends up in store, it will have been processed through a number of different systems, owned by a number of different teams and available in a number of different format.

Your role is to bring all this data together to paint a complete picture of the journey a product makes through the supply chain and then apply logic to identify the biggest causes of waste and unavailability. The data is large and messy and often ambiguous.

THE PROCESS

All products start with a forecast. This is the amount of product expected to sell within each store on a daily basis. This forecast is compared with the amount of stock currently held within stores (the stockfile) and an order is generated to purchase the difference. Quite often the stockfile is incorrect, so an accurate picture of what it is in store is not available.

Stock must be ordered at a minimum level. This is called a tray size and is set for each product. For example, BLT sandwiches come on a tray size of 18.

The order is placed with suppliers. Some suppliers cannot deliver every day, so we need to make sure we order enough to cover the entire delivery period. Sometimes, suppliers will over or under-deliver.

These orders are delivered into various distribution centres across the country. From here the product is allocated to store based on its need (the difference between the forecast and the amount they already have) and the volume delivered by the supplier. This allocation is used to pick the amount for each store, loaded onto lorries and dispatched. Quite often the amount picked will be incorrect.

When a lorry arrives in store checks are done to make sure the amount delivered is correct, often it is not. The product is unloaded and either put straight onto shelves or stored within a backroom. In the store processes exist to check and confirm the stockfile matches the actual product they have.

On the day a product is due to expire, it is considered wasted. The price will be marked down in stages to try and recover some of the lost value, but if it doesn't sell by the end of the day it will be thrown in the bin.

THE DATA

You have data from several different systems. Most of the data is for an item, store, date grouping. The information you have is:

1. Daily sales (value and quantity)
2. Gross waste – before recovery (value and quantity)
3. Recovery (value)
4. Closing stock level
5. Target closing stock level
6. Forecast sales
7. Forecast waste
8. Amount delivered
9. Amount ordered
10. Tray size
11. Whether the product was ranged (ie, was it being discontinued)
12. Whether the product stocked out (ie, was not available to purchase)
13. The shelf life of the product
14. The lead time for a product
15. How much the stockfile was adjusted by following a stock check

The data is split between four tab delimited files:

1. Case Study – In Store Data
2. Case Study – Forecast Data
3. Case Study – Depot Data
4. Case Study – Closing Inventory

THE LOGIC

We need to be able to assign a reason for each waste and stockout event. To do this, the following reasons and logic have been identified:

Reason	Waste Explanation	Waste logic	Stockout Explanation	Stockout logic
Product is not ranged	Product no longer ranged so stock wasted	Waste qty > 0 on days when ranging indicator is 0	Product is no longer ranged so not available	Stockout indicator 1 on a day when ranging indicator is 0
Depot to store delivery inaccurate	Store receives more from the depot than the amount allocated to them	Received_units (store) > Allocation (store) Shelf life before and including waste event	Store receives less from the depot than the amount allocated to them	Received_units (store) < Allocation (store) Allocation to store lead time before and including stockout
Sales Forecast Error	We predicted we would sell more than we did	(Sum(Forecast) - Sum(actual sales + Lost Sales)) > forecast_tolerance over Shelf life before and including Waste event AND ABS(forecast_demand_qty_over_shelf_life-sales_qty_over_shelf_life)>1	We predicted we would sell less than we did	(Sum(Forecast) - Sum(actual sales + Lost Sales)) < (forecast_tolerance *-1) over Allocation to store lead time before and including day of stockout AND ABS(forecast_demand_qty_over_lead_time-sales_qty_over_lead_time)>1
Waste Prediction	We predicted we would waste more than we did	Sum(waste_Forecast) > Sum(actual waste) + waste_tolerance over Shelf life before and including Waste event	We predicted we would waste less than we did	Sum(waste_Forecast) < Sum(actual waste) - waste_tolerance over Allocation to store lead time before and including day of stockout
Stock file adjustment	Store had more stock than it reported	Stockfile adjustment > 0 on either the day of or day after the waste event	Store had less stock than reported	Stockfile adjustment < 0 On event day or event day minus 1
Negative Stockfile	Store had more stock than it reported	closing inventory < 0 on day before or day of the waste event		

Forecast_tolerance is between 75% of forecast and 125% of forecast. Waste tolerance is between 75% of waste forecast and 125% of waste forecast

YOUR ROLE

Your role is to take the various data sources and stitch them together. Be warned that some are incomplete or don't make sense.
Once you have stitched them together and filtered out the invalid data, write a script that will apply the logic and attribute reasons for waste and availability.
Once these reasons have been attributed, visualise the data in a way that makes it easy for someone to see where the biggest problems are.

NOTES

When presenting back you will be expected to talk through the approach you have taken, the technology you chose (and why) and the kind of people you might want to speak with to validate your understanding of the data and your findings.