Facts:

Facts	Explanation	Relevant Questions
No. of Units Sold (From Quantity column)	The fact can be directly gathered from the quantity column since it represents quantity sold of that product for that sale.	**How many units of products has a vendor sold over time across their various outlets?** The Product_ID uniquely identifies each product and therefore can be directly used from the original operational data to get the number of units sold from quantity columns group by product ID. Since we need the sum of quantities across different outlets we will use sum aggregation over different outlets. **Which staff and which sales locations are the best performing in terms of both the number of units sold, and the total profit?** The result of the question can be directly achieved by the aggregation of summing of the number of units sold and the total profit. This aggregation will be done upon Staff_ID and Postcode attribute. **Do some products sell better than others with respect to brand, supplier, or whether they offer loyalty points?** The result of the question can be obtained by summing the number of units sold for each product_ name, brand, supplier and the Loyalty_Status. **What are the worst performing products (in terms of number of units sold) in each outlet, city, and country?** Number of units sold can be aggregated (summed) for different product ID's, outlet, city and country. **Do the customers of some vendors buy more products than those of other vendors?** Vendor_Name and Customer_ID can be used to make the appropriate comparison for aggregated value of number of units

		sold between different vendors for a particular customer.
Total Profit	Total Profit can be created using the Unit_Cost and Price from the Sale_Line Table, as a separate field in the data warehouse. Total profit can be calculated as Sale_Price – Unit_Cost. The original database will contain the facts Unit_Cost and Sale_Price, but our require fact which is total profit will be calculated as we load it into the data warehouse.	**Which staff and which sales locations are the best performing in terms of both the number of units sold, and the total profit?** This field can be aggregated (SUM) by location to answer the relevant question.
unit_sale_price	This is used to calculate the total profit and would make sense to keep it in the fact table for future reference. It will be gathered from price field in sale_line table.	
unit_cost	This is also used to calculate the total profit and would make sense to keep it in the fact table for future reference. It will be gathered from Unit_cost field in the Product Table.	
Total_revenue_sales	Total revenue of sales field can be calculated by using the Price, Quantity and Discount (quantity*price (from Sale_Line table) - Discount) for a given sale and product. Each sale will now have a revenue	**How does the total revenue of sales vary across customers' gender, age, and location?** The result of the question can be calculated by summing the total revenue of sales per customer. **How are customers paying, and what is the total revenue through each payment method?**

	associated with a customer.	The result of the question can be calculated by aggregating(summing) total revenue sales for each different payment type. How are customers paying can be answered by aggregating(count) for each Payment_Type. **For each outlet, what are the busiest hours of the day in terms of number of sales?** Total Revenue of sales can be aggregated over sale_hour and sale_day.
discount	This measure represents the discount offered on different sales.	**Are there any trends in the amount of discount offered by different vendors?** The result of the question can be directly calculated using aggregation (MEAN) over discount amount by Sales_Date and Vendor_Name.

Dimensions:

Dimension Table	Dimensions	Explanation	Relevant Questions
Customer	customer_ID		**Do the customers of some vendors buy more products than those of other vendors?**
	Age	The age can be calculated using the Date_Of_Birth which is calculated by the current year field subtracting the current year from the Date_Of_Birth field.	**How does the total revenue of sales vary across customers' gender, age, and location?**
	Gender	This can be used to answer the relevant questions around demographics. As needed to filter by Q3 in the analysis question.	**How does the total revenue of sales vary across customers' gender, age, and location?**

	Location	This is calculated as a mix of attributes city, postcode and country. But for the purpose we will keep it by post_code.	**How does the total revenue of sales vary across customers' gender, age, and location?**
Product	product_id		**How many units of products has a vendor sold over time across their various outlets?**
	product_name		**Do some products sell better than others with respect to brand, supplier, or whether they offer loyalty points?**
	brand_type		
	Supplier_Name	This is assumed to represent the supplier	**Do some products sell better than others with respect to brand, supplier, or whether they offer loyalty points?**
	Loyalty_status	The Loyalty_Status will be calculated for each product brand and supplier based on their loyalty points greater then 0, this column can be referred to as 1 or 0, indicating whether the supplier offers loyalty points or not. Each Product may have different loyalty value based on a particular supplier and brand which makes the loyalty status different for each of the product brand and their supplier.	**Do some products sell better than others with respect to brand, supplier, or whether they offer loyalty points?**
Staff	Staff_id	Staff ID is considered to best represent the staff.	**Which staff and which sales locations are the best performing in terms of both the number of units sold, and the total profit?**
	Staff_name		**Which staff and which sales locations are the best**

			performing in terms of both the number of units sold, and the total profit?
Outlet	outlet_id		**How many units of products has a vendor sold over time across their various outlets?** Aggregation would be count, time. Since the time of which we are storing the information in the database we are calculating the count of the products sold by each outlet. Since Outlet_ID represents an outlet, and the Product_ID represents different products which are stored in our product table, the aggregation over Product_ID upon different outlets will give us the products sold by venders across different outlets. **Are there any trends in the amount of discount offered by different vendors?**
	vendor_name	This represents the vendor under which the sale took place.	**Are there any trends in the amount of discount offered by different vendors?** The vendor name can be used to calculate aggregated (SUM) of the discount. **Do the customers of some vendors buy more products than those of other vendors?**
	outlet_name		**For each outlet, what are the busiest hours of the day in terms of number of sales?**
	outlet_location (from postcode in Outlet Table)	Here we assume that the location can be well represented by the postcode which makes Postcode a major dimension for performing this aggregation.	**Which staff and which sales locations are the best performing in terms of both the number of units sold, and the total profit?**

	outlet_city	This represents outlet city	**What are the worst performing products (in terms of number of units sold) in each outlet, city, and country?**
	outlet_country	This represents outlet country	**What are the worst performing products (in terms of number of units sold) in each outlet, city, and country?**
Time	sales_date		**How many units of products has a vendor sold over time across their various outlets?** The relevant question require to calculate the no of products sold over time. So, this dimensions along with the vendor name and outlet ID will be needed to aggregate (summing) the number of units sold. **Are there any trends in the amount of discount offered by different vendors?** Because we are looking for trend by different vendors it is imperative that we aggregate the fact grouped by sales_date.
	sales_hour	extracting an hour from Sale_Date field	**For each outlet, what are the busiest hours of the day in terms of number of sales?**
	sales_day	extract day from Sale_Date field	**For each outlet, what are the busiest hours of the day in terms of number of sales?**
payment	Payment_Type	Since this is a single field used for filtering, we can also keep it in the fact table but here we are considering creating a separate dimension.	**How are customers paying, and what is the total revenue through each payment method?**
	Sales_ID	To link the payment dimension with the fact table we are including sales_id to gather the	

		payment type information for each sale.	