

Data Normalization Task

Puranjoy Patra,

Well done with the previous task. I have a new one for you.

One of our client projects requires minor support. The dataset is currently in a de-normalized form, and we need your help to transform it into a normalized form.

Your task will involve creating proper fact and dimension tables based on the dataset, which can be found in the "fact_order_lines.csv" file.

Here are a few screenshots of the dataset to give you a better idea of what it looks like and how the final output should be presented:

Input Data

Denormalized Table

fact_order_lines

order_id	order_placement_date	mmm_yy	week_no	customer_id	customer_name	city	product_id	product_name	category	order_qty	delivery_qty
FMR342036	01-03-2022	Mar_22	W 10	789203	Rel Fresh	Vadodara	25891601	AM Tea 500	beverages	110	110
FMR342206	01-03-2022	Mar_22	W 10	789220	Atlas Stores	Surat	25891601	AM Tea 500	beverages	78	78
FMR325016	01-03-2022	Mar_22	W 10	789501	Viveks Stores	Surat	25891601	AM Tea 500	beverages	139	139
FMR331026	01-03-2022	Mar_22	W 10	789102	Vijay Stores	Ahmedab	25891601	AM Tea 500	beverages	180	180
FMR329026	01-03-2022	Mar_22	W 10	789902	Elite Mart	Ahmedab	25891601	AM Tea 500	beverages	93	88
FMR339036	01-03-2022	Mar_22	W 10	789903	Elite Mart	Vadodara	25891601	AM Tea 500	beverages	93	84
FMR324216	01-03-2022	Mar_22	W 10	789421	Lotus Mart	Ahmedab	25891601	AM Tea 500	beverages	108	103
FMR344216	01-03-2022	Mar_22	W 10	789421	Lotus Mart	Ahmedab	25891601	AM Tea 500	beverages	75	75

Output Data

Normalized Table

fact_orders

order_id	order_placement_date	customer_id	product_id	order_qty	delivery_qty
FMR342036	01-03-2022	789203	25891601	110	110
FMR342206	01-03-2022	789220	25891601	78	78
FMR325016	01-03-2022	789501	25891601	139	139
FMR331026	01-03-2022	789102	25891601	180	180
FMR329026	01-03-2022	789902	25891601	93	88
FMR339036	01-03-2022	789903	25891601	93	84
FMR324216	01-03-2022	789421	25891601	108	103
FMR344216	01-03-2022	789421	25891601	75	75
FMR346216	01-03-2022	789621	25891601	170	170
FMR335206	01-03-2022	789520	25891601	182	146
FMR327026	01-03-2022	789702	25891601	93	74
FMR341226	01-03-2022	789122	25891601	68	61
FMR321016	01-03-2022	789101	25891601	124	124

Normalized Tables

dim_customers

customer_id	customer_name	city
789201	Rel Fresh	Surat
789202	Rel Fresh	Ahmedabad
789203	Rel Fresh	Vadodara
789301	Expression Stores	Surat
789303	Expression Stores	Vadodara
789101	Vijay Stores	Surat
789102	Vijay Stores	Ahmedabad

dim_dates

date	mmm_yy	week_no
01-03-2022	Mar_22	W 10
02-03-2022	Mar_22	W 10
03-03-2022	Mar_22	W 10
04-03-2022	Mar_22	W 10
05-03-2022	Mar_22	W 10
06-03-2022	Mar_22	W 11
07-03-2022	Mar_22	W 11
08-03-2022	Mar_22	W 11

dim_products

product_id	product_name	category
25891101	AM Milk 500	Dairy
25891102	AM Milk 250	Dairy
25891103	AM Milk 100	Dairy
25891201	AM Butter 100	Dairy
25891202	AM Butter 250	Dairy
25891203	AM Butter 500	Dairy
25891301	AM Ghee 250	Dairy
25891302	AM Ghee 150	Dairy

Please complete this task within the next 2 days.

Best regards,



Hem

Head of Data Analytics | AtliQ Technologies Pvt. Ltd.

Website: www.atliq.com