

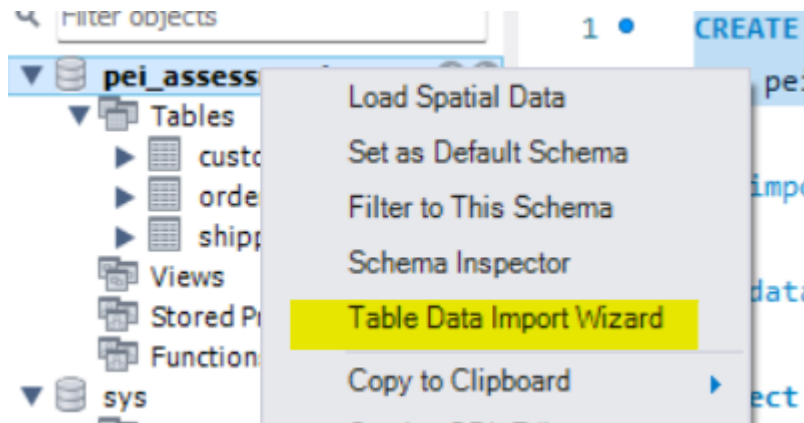
PEI Assessment

Data provided was in Json, CSV and Excel format for ease of operation. I converted all the other data formats to CSV using Python (the code for which is available here https://github.com/BharathNandan036/PEI/blob/main/Files_to_csv.ipynb) as I could easily load the csv data to a table in mysql server easily.

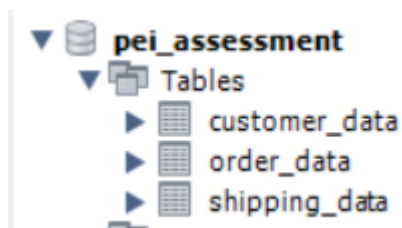
I created a database called PEI_Assessment using the below SQL:

```
CREATE DATABASE pei_assessment;  
USE pei_assessment;
```

Then i imported the data (already converted to CSV) using the table import function in MySQL Workbench as shown below:



Imported all the 3 CSV files and the data base not looked like below:



Customer Data Sanity check:

Started off with basic sanity checks and I understood that there were 250 rows of data and even the unique customer IDs were 250 this ensured that there were no duplicates. Checked for null values in any column and also for special characters or numbers in any categoric variable columns.

I could understand that there were few rows in which either first name or last name had either a special character or a number within the data.

Code for the above is available here :

https://github.com/BharathNandan036/PEI/blob/main/Customer_data_sanity_checks.sql

Created a script to understand what is the incorrect field. This would help to debug and fix the issues.

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
	customer_id	correction_needed	incorrect_value
▶	6	first name has a special character	N!cole
	14	first name has a special character	N!cole
	109	first name has a number	R0bert
	113	last name has a number	R0berts
	118	first name has a number	R0bert
	162	first name has a special character	N!cole
	171	first name has a special character	L@rry
	198	first name has a number	R0bert
	211	first name has a number	Al1cia
	214	first name has a special character	N!cole
	236	first name has a number	Al1cia
	242	last name has a number	R0berts

Orders data sanity checks

Checked for similar pattern of any incorrect values in any field or any record but the orders data was clean and required no changes. The scripts used to perform the checks is available here : https://github.com/BharathNandan036/PEI/blob/main/order_data_sanity_checks.sql

Shipping data sanity checks

Even Shipping data did not have any issues, the scripts used to perform the checks is available here :

https://github.com/BharathNandan036/PEI/blob/main/shipping_data_sanity_checks.sql

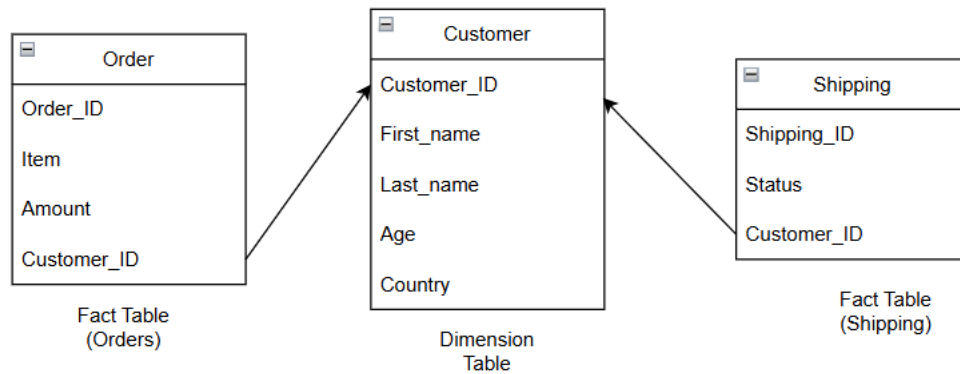
Context Checks

Checked for the context of customer data in every other fact table. I checked if the customers who placed the orders and customers to whom the shipments were booked were all a subset of customers from customer data. The queries used to check the context are a part of each of the files mentioned above.

I performed the entire task in python as we can incorporate these checks as a part of data ingestion and the above errors can be eliminated once for all. The python code is here :

https://github.com/BharathNandan036/PEI/blob/main/PEI_Data_Sanity_Checks.ipynb

Data Model



I did not have tool in which I could build a detailed data model so i have used <https://app.diagrams.net/> to draw a rough data model sketch for reference.

Reports

The script for the reports are available here :

<https://github.com/BharathNandan036/PEI/blob/main/Report.sql>

1. Country wise total amount of the pending shipments. (`select distinct country , 'Status', sum(amount) as pending_amount from customer_Data cd join order_data od on cd.customer_id = od.customer_id join shipping_data sd on sd.Customer_ID where `status` <> 'Delivered' group by 1,2;)`

	country	Status	pending_amount
▶	UK	Pending	36652500
	USA	Pending	35730000
	UAE	Pending	7492500

2. Customer wise number of orders placed , number of items purchased, total sales and items purchased. (`select distinct cd.customer_id , count(order_id) as number_of_orders , count(item) as no_of_items , sum(amount) as amount_spent, group_concat(item separator ',') as items from customer_Data cd join order_data od on cd.customer_id = od.customer_id group by 1 order by 4 desc;)` too big output for screenshot.
3. Country wise maximum products purchased and quantity. (`with rank_sales as (select distinct country as country , item , count(item) as qunatity , rank() over(partition by country order by count(item) desc) as ran from customer_data cd join order_data od on cd.Customer_ID = od.Customer_ID group by 1,2) select country ,item, qunatity`

from rank_sales where ran = 1 order by 2 desc;)

	country	item	qunatity
▶	UK	Mousepad	24
	USA	Mousepad	18
	UAE	Keyboard	12

4. Age category wise most items purchased and quantity. (with age_cat as (select case when age < 30 then 'age<30' else 'age>30' end as age_cat , item , count(item) as qunatity , rank() over(partition by case when age < 30 then 'age<30' else 'age>30' end order by count(item) desc) as ran from customer_data cd join order_data od on cd.Customer_ID = od.Customer_ID group by 1,2) select age_cat ,item, qunatity from age_cat where ran = 1 order by 2 desc;)

	age_cat	item	qunatity
▶	age<30	Mousepad	17
	age>30	Keyboard	35

5. Country with minimum transactions and sales. (select country , count(order_id) as orders from customer_data cd join order_data od on cd.Customer_ID = d.Customer_ID group by 1 order by 2 asc limit 1 ;

select country , sum(amount) as sales from customer_data cd join order_data od on cd.Customer_ID = od.Customer_ID group by 1 order by 2 asc limit 1 ;

select country , count(order_id) as orders , sum(amount) as sales from customer_data cd join order_data od on cd.Customer_ID = od.Customer_ID group by 1 order by 2 asc , 3 asc limit 1;)

	country	orders	sales
▶	UAE	40	49950