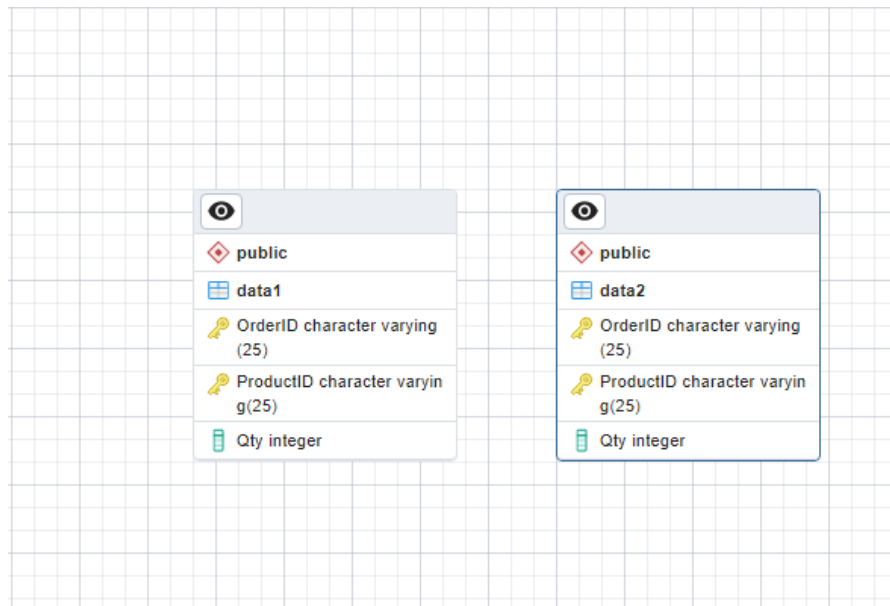


NeenOpal Task-1

Database Creation:

Created a Postgres database with the following schema, there are 2 tables `data1` and `data2`

```
CREATE TABLE data1 (  
  "OrderID" varchar(25),  
  "ProductID" varchar(25),  
  "Qty" integer,  
  PRIMARY KEY("OrderID", "ProductID"));  
CREATE TABLE data2 (  
  "OrderID" varchar(25),  
  "ProductID" varchar(25),  
  "Qty" integer,  
  PRIMARY KEY("OrderID", "ProductID"));
```



Question1: How to identify the Records (Order ID + Product ID combination) present in data1 but missing in data2



```
SELECT * FROM data1
LEFT JOIN data2
ON data1."OrderID" = data2."OrderID" AND data1."ProductID" = data2."ProductID"
WHERE data2."OrderID" IS null AND data2."ProductID" IS null;
```

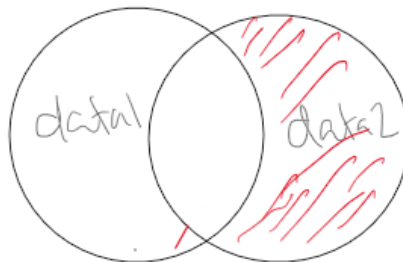
<https://github.com/MASHOD0/neenopel/blob/main/task1/q1values.csv>

To get the count of the number of missing records

```
SELECT COUNT(*) FROM data1
LEFT JOIN data2
ON data1."OrderID" = data2."OrderID" AND data1."ProductID" = data2."ProductID"
WHERE data2."OrderID" IS null AND data2."ProductID" IS null;
```

Number of records: 507

Question2: How to identify the Records (Order ID + Product ID combination) missing in data1 but present in data2



```
SELECT * FROM data2
LEFT JOIN data1
ON data2."OrderID" = data1."OrderID" AND data2."ProductID" = data1."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null;
```

alternative query using **RIGHT JOIN**

```
SELECT * FROM data1
RIGHT JOIN data2
ON data1."OrderID" = data2."OrderID" AND data1."ProductID" = data2."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null;
```

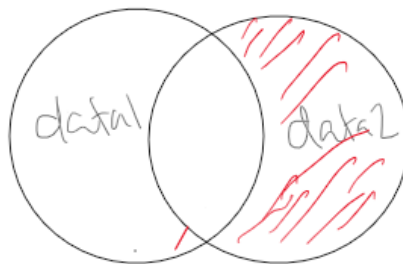
<https://github.com/MASHOD0/neenopel/blob/main/task1/q2values.csv>

To get the count of the number of missing records

```
SELECT COUNT(*) FROM data2
LEFT JOIN data1
ON data2."OrderID" = data1."OrderID" AND data2."ProductID" = data1."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null;
```

Number of records = 508

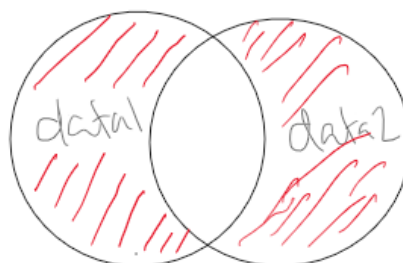
Question3: Find the Sum of the total Qty of Records missing in data1 but present in data2



```
SELECT SUM(data2."Qty") FROM data2
LEFT JOIN data1
ON data2."OrderID" = data1."OrderID" AND data2."ProductID" = data1."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null;
```

answer = 1956

Question4: Find the total number of unique records (Order ID + Product ID combination) present in the combined dataset of data1 and data2



```
SELECT * FROM data1
FULL OUTER JOIN data2
ON data1."OrderID" = data2."OrderID" AND data1."ProductID" = data2."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null OR data2."OrderID" IS null AND data2."ProductID" IS null ;
```

<https://github.com/MASHOD0/neenopel/blob/main/task1/q4values.csv>

To get the count of the missing records

```
SELECT COUNT(*) FROM data1
FULL OUTER JOIN data2
ON data1."OrderID" = data2."OrderID" AND data1."ProductID" = data2."ProductID"
WHERE data1."OrderID" IS null AND data1."ProductID" IS null OR data2."OrderID" IS null AND data2."ProductID" IS null ;
```

answer = 1015