**2.B Exploring the Northwind Database**

3. A record in the table represents a record for that specific table for example for the table categories it will give me records of different categories, for customercustomerdemo it will give customer demographics and customerdemographics might give me more details regarding the customer demographics while customers will give records of each company who is a customer such as location and contact info. The employees’ table would give me names of the employees and their information and home address and job title. Employee territory will give me all the location that Northwind has employees in. Order details would give me details about each order that was purchased, and orders would give me orders that were purchased on a broader level like where it is, each order is being shipped to and who its shipping it. Products would give me the names of each product that Northwind has and quantity, price, and if it’s in stock, or has been discontinued. The region would give me a description of the region where the product is being shipped. Shippers would give me more information regarding the shipper and contact information for them. Suppliers would give me more information regarding the companies who come together to make the final product and their contact information. Territories would provide me with the territories that each product was sent to like the city or state.

4. It is consistent in what I thought the data represented as it would list the different categories and upon seeing the row count, I can assume there are 8 different categories and 93 different customers. There are no records for customercustomerdemo table and cusotmerdemographics table, so it looks like the tables were not actually needed to draw any insights. However, there are more employee territories than there are employees, so I’m not sure what data employee territories represent…. maybe territories where employees have sold products or the territory/city they manage. I also see that there are more order details than there are orders. This is because orders have each order and who that order belongs to which customer and information on when that customer ordered, required date when they need the order and when the ordered shipped. It also gives us information about who it shipped with and the ship to name or customer name again and their address while order details have repeated orders with different products so the customer could have made a single purchase or order with multiple products or items and gives us the price of the item, quantity ordered, and if there were any discounts. The products table has product name who it is supplied by and what category it falls under. It also gives us the amount per unit, unit price, and if the item is in stock. Region gives us cardinal directions of the region so a total of 4 records one for eastern, one for western, one for northern, and one for southern.

Shippers gave us shipper names and their phone number which was consistent with what I thought the data represented. Suppliers give us 29 rows so 29 records for each supplier’s name and their main point of contact information.

Finally, territories have 53 records or cities of where the product is shipped to and their cardinal direction of the location of the city.

6. The primary key of the categories table is CategoryID and this table does not have parent table. The primary key of the customers table is CustomerID and this table does not have a parent table. Th primary key for the employees’ table is Employee ID and this table also does not have a parent table. The employeeterritories table has two keys one foreign key referencing the parent employees table (employeeid) and the second foreign key that is referencing the parent table of territories. (territoriesID). The order details table is made up of two foreign keys as well, one is referencing the orders table as (ordered) and the other key is referencing the products table as(productid). The orders table has a primary key of OrderID but has no parent table. The products table has a primary key to productid with no parent table. Region has primary key of RegionID with no parent table. Shippers table has a primary key of shipperID with no parent table. Suppliers have a primary key of supplierID and no parent table. Finally, territories have a primary key of territoryID, which has a parent table referencing region tables with (RegionID).

7. On the categories table the value for the column categoryID represents each unique category and I would see most likely some numbers or integers. This column is part of the primary key to this table. I think this column would not be valuable to bring into our PowerBI Model because it’s just numbers and would not give a full name of what the category is. I think the end user will not have knowledge to refer to and see what that category ID is when they can just quickly glance at the name instead. I also do not think there could be any calculations done to this column data. The category name would be a string value where it represents the actual names of the categories. Although this is not a part of the primary key to this table or a foreign key, I think it would be valuable to bring into our Power BI model because it clearly states the name of the category, and the end user will understand what it represents. I think this column is appropriately named for Data Analysis purposes. I do not think the description or picture columns would be valuable to bring into our Power BI Model as they are used to better understand or describe the categories.

For the customer’s table. I the value in the customerID column would also be a number or integer representing each unique company. This column is a part of the primary key to this table. However, I do not think this would be valuable to bring into our Power BI Model if it could be hidden, I would hide it because I would like to just show the next column which is company name to display the actual customer’s name that northwind for a better understanding of who the customer is. I think the company name is not appropriately named. A more appropriate name for this column would be customer name. I don’t think there would be any calculations where this column data might be used but the data type for this column would be a string and format for this column would be lowercase. The next columns would be contact name and contact tile which would also be data types of string and contain character string values of the point of contact for the customer and their job title within that customer. There would not be any calculations where this column data might be used and would not be valuable to bring into our Power BI model. For the columns address, region, postal code, phone and fax would not be valuable to bring into the powerBI model but just hide them as they would be string character values and are not primary keys or any foreign keys referencing parent tables. However, for city, country would be string type character values but I could do a count calculation of these column data to see how many customers we have in a particular city or a particular country.

For the employees table the employeeID would probably be a integer value number identifying the unique employees that northwind has. This is a primary key to this table but would not bring any value to my Power BI model because I would rather display each employee by their first and last name for a better understanding of who they are. The last name and first name columns would be string character values that represent the actual employees. I believe this column could be renamed to employee first name or employee last name. I may not be able to calculate data from this column by itself but maybe I could use group by to calculate what territory they manage or sell the most in. The column for title of courtesy would also be a string of character values. It is not a primary key column but might give us more insight information on who might be married and hold a higher degree of study. I would keep and rename the column employeehonorifics. I could maybe do a count on the employee titles. For the birthdate and hire date columns I would expect string value characters as well because I would not be able to perform any calculations on these columns as they are unique to each employee. I would say maybe a count or calculation of who shares the same birthday for an office celebration. I think I would keep these and maybe hide them, if necessary, in the PowerBI model. For the columns for address would be a string value character along with city, region, postal code, country, phone, extension, notes, and reports to. I think city and region would be valuable to keep because we could count who lives in that city, region, or country. I don’t think home phone, extension, photo, and photo path would be very valuable so maybe if I can hide them or not bring them into the powerBI model as I would not be able to perform a calculation on these columns nor do they describe something relevant. However, I would keep the notes column as they describe each employee’s education and the reportsto column as they give us information who who supervises who or holds a higher position in the company. I would keep salary as I would expect this to be an integer value it would be valuable to bring into the Power BI model because we can perform a sum, min, max, or mean calculation on the salaries of all employees combined. Would be able to see who gets the least or most and who gets paid an average salary among the employees.

On the employeeterritories table, for the employeeID column I would expect an integer value and this column references the parent table of employees and is made up of the terrirotyID or cities that each employee may manage. I could maybe perform a calculations of count with how many territories each employee has. I think it would be valuable to bring into the Power BI Model because it would give us some insight on which city each employee handles.

On the order details tables, for the orderID column it is a primary key referencing the parent table of orders and another primary key referencing the parent table of products both expected to be integers I think these would be valuable to identify the order and the product being order however it might be hard to decipher without using a legend of which order is whose and which product is what. I also would not be able to do any calculations with these columns. The next three columns are unit price, quantity, and discount, all three expected to be integer or float values representing the price per item, the quantity of each item ordered and any discount amount. I think the unit price and quantity would be valuable to bring into the PowerBI model because there could be a calculation done by multiplying the unity price by quantity or adding them together. I think both columns have appropriate names. From first glance, it looks like there are only zero values and would not be able to do any calculations this column. Therefore, it would not be valuable to bring into the PowerBI model and would need to hide the column if possible.

On the orders table, for OrderID it’s expected for it to be an integer string value to identify each individual order from each individual customer. Which would not be valuable to bring into my PowerBI model because I would not be able to do any calculations on this column. This column is also the primary key for this table. The next column is customerID and I would expect an integer string value character, but it is a string character value with abbreviation for the customer’s name. This column is not the primary key to this table, but it refers to the parent table of customers. I think this column would be valuable to bring into the Power BI model to clearly identify who the customer is in this table. We also have the employeeID in the next column which is an integer value representing each employee who might have taken the customer’s order. I might be able to do a count calculation on the employee’s who took the most orders based on their employeeID so this column could be valuable to bring into the PowerBI model without renaming the column. The next three columns are orderdate, requiredate, and shippeddate which have date values representing the date the order was placed, the date that the order is required at the customer and the date when the order shipped. I could probably do a calculation on what date most orders are ordered if for more near a special holiday or see when orders are more in demand by looking at the orderdate and required date. I think these columns would be valuable to bring into the Power BI model because they would give us some insights on the demand dates. However, I would not rename these columns as I think they are appropriate names but would not bring over or I would hide the shipped date column. The next column is shipvia which would be referencing the shipper table. It is an integer value representing the shipper who shipped, who they shipped the order with. I think this would be a valuable column to bring into our PowerBI model because I could do a count of which shipper is used most often. The next column in this table is freight and I think this is the weight of the shipment. This is a float value and would only be useful if I were giving a rate based on the weight to do a certain calculation with it. However, this column is not appropriately named. It would be better if it were named “shipment weight”. I don’t think this column would be valuable to bring into our Power BI model because I would not do a calculation on it. The next column is shipname which is a string character value and would not be valuable to bring into the PowerBI model because its redundant information since we already have the customer’s name under customerID. I would also not bring the ship address as it would not be a column I could do a calculation from. For the columns ship city, ship region, postal, code and ship country, each containing string value character would help me analyze where most orders are being shipped so I would bring these into the PowerBI model if I had not already extracted the information from another table. I would also rename these columns as shiptoaddress, shiptocity, shiptoregion,shipto postalcode, and shiptocountry for more clarity.

For the next table which is products, I would keep productname, supplierID, quantityperunit, Unit,price, unitsonOder, and reorderlevel, I would then drop the column for discontinued items. I think the others are valuable to bring into the powerBI model because I would be able to make the sum of the unit price to find an average price, lowest price and highest priced item.I could then compare if the unit is nearing the restock order and calculate the sum, units on hand, and recorder level across all products.

For the region tables it only has two columns which regionID is an integer value representing each unique region and being the primary key to this table. I would not be able to make any calculations on this column unless it’s being referenced by another table. The second column is a regiondescription which has the four cardinal directions. I think the column has an appropriate name and although I may not be able to calculate any data in these columns if another table references this region table I could probably perform a count calculation and I think it would be valuable to bring into the Power BI model.

Shippers is the next table which only has three columns. The first column has an integer value representing each unique shipper, the second column is the shippername which has string character values of shipper’s name and the last column is phone which is most likely a string value. Although, I may not be able to perform calculations on these three columns if any of the other tables reference this shipper name I would rather use the company name and rename the column as shipper name. I would then bring this shipper name column into the Power BI model.

For suppliers table, the first column is an integer value to uniquely identify the company. It is the primary key to this table, and I would not be able to calculate data in this column. I do not think it would be valuable to bring to the Power BI model since we could display the companyname column instead for better readability and this column would be string value characters. I would also rename the column from companyname to shipper name. For contactname column and address column I would omit as they are string value characters and would not be able to perform calculations on these columns. I would probably be able to do a count calculation on the city,region, postal code, and country to see where most of the shippers are located. I would rename these columns as shippercity, shipperregion, shipperpostalcode, and shippercountry. The next few columns are a mix of integers representing phone numbers in different formats, fax numbers and homepage website links.These last three columns I would hide or not bring into our PowerBI model at all as I would not be able to calculate on these columns or extract useful insights as I already have city, region, and country.

The last table is territories, and it only has three columns. The first column is territoryID and has an integer value representing each unique territory description. The territoryID is also the primary key to this table. The second column is territorydescription which city names which would be a string value, and the third column is the regionID which has the integer value to represent the direction of where each territory description is at. This is a foreign key referencing the parent table region. I can’t do a calculation on the territoryID which would not be valuable to bring into the PowerBI model. However, I would rename the column “territorydescription” to “territorycity” and could possibly do a calculation of how many territory cities belong to the same region. I would bring territorycity and regionID into the PowerBI model as it would provide more insights as to where the cities are located within the country or continent.