2/2/2024

SQL Assessment -Part 2

1. **SCENARIO**
   1. You just started as a new Data Analyst at a shipping company called Northwind.  There is a new initiative to better understand customers and  trends within the beverages, confections, and dairy products . Your manager has called a meeting to discuss a new project with a deadline of February 10th. The manager doesn’t know too much about the database other than it has sales data recorded by the sales team, so they are depending on you to help educate them and deliver a report on your preliminary findings from the data.
2. **Challenge**
   1. You will need to create a report from the Northwind database that will help your manager understand the data.  You will need to determine the best way to present the data.  This may be several separate queries or 1 or 2 larger queries.  Either way you will need to combine the results in a report in Excel.  You will also need to determine fields that would be relevant. Remember the manager hasn’t given you any guidance. You are teaching them about the data.
   2. Your report should include but not limited to the following:
      1. **Product Insight:**
         1. Number of products available by each category
         2. Average price and quantity sold by product and category
         3. Identify the most expensive product and the least expensive in each category (try to include this in a single query)
         4. Think of at least 2 additional metrics that may be of interest to highlight product performance
      2. Products **Customer Insight:**
         1. Sales performance by some time frame (you choose) and location
            1. Sales performance is how well or poorly a product or category performed over a given period of time
         2. Any insight into the shipments?
         * Orders shipped on time?
         * Shipping company?- shippers table
         1. Are there any trends in customer spending?
         2. Think of at least 2 additional metrics that may be of interest to highlight customer trends
   3. Now pull this report together in an Excel workbook. You can add charts and pivot tables to extend your analysis further. Make sure to provide a summary of your findings, pre and post analysis. Remember this is for your manager so make it as professional as possible.
3. **Submit Project deliverables to Canvas SQL Assessment**
4. Submit final report in Excel with the above analysis
5. Submit  all queries in a single file. This can be a Word or a \*.sql script
6. Make sure to make comments at the beginning of each query for context.. It doesn’t have to be too detailed. *Ex. Average price of products by category and location*

--a. Number of products available by each category

**SELECT** *c*.category\_name **AS** *Category*,-- p.discontinued, p.product\_name ,

**count**(*p*.product\_name) **AS** *Product*

**FROM** categories *c*

**LEFT** **JOIN** products *p* **ON** *c*.category\_id = *p*.category\_id

**WHERE** *c*.category\_name **IN** (**'Beverages'**, **'Confections'**, **'Dairy Products'**)

**GROUP** **BY** *c*.category\_name--, p.discontinued , p.product\_name

-- Average price and quantity sold by product and category

**SELECT** *c*.category\_name **AS** *Category*,

*p*.product\_name **AS** *Product*,

**round**(**avg**(**cast**(*od*.unit\_price **AS** **NUMERIC**)),2) **AS** *Avg\_price*,

**sum**(*od*.quantity) **AS** *quantity\_sold*

**FROM** order\_details *od*

**LEFT** **JOIN** products *p* **ON** *p*.product\_id = *od*.product\_id

**LEFT** **JOIN** categories *c* **ON** *c*.category\_id = *p*.category\_id

**WHERE** *c*.category\_name **IN** (**'Beverages'**, **'Confections'**, **'Dairy Products'**)

**GROUP** **BY** *c*.category\_name, *p*.product\_name

**ORDER** **BY** *c*.category\_name

-- Identify the most expensive product and the least expensive in each category

-- from canva module RANK() OVER(ORDER BY TotalSpent DESC) AS SalesRank

**WITH** cte **as**

(

**SELECT** c.category\_name **AS** Category,

p.product\_name **AS** Product,

p.unit\_price **AS** price,

--round(avg(cast(od.unit\_price AS NUMERIC)),2) AS Avg\_price,

**sum**(od.quantity) **AS** quantity\_sold,

**RANK**() **OVER**(**partition** **BY** c.category\_name **ORDER** **BY** p.unit\_price **DESC**) **AS** most\_expensive,

**RANK**() **OVER**(**partition** **BY** c.category\_name **ORDER** **BY** p.unit\_price **asc**) **AS** least\_expensive

**FROM** order\_details od

**LEFT** **JOIN** products p **ON** p.product\_id = od.product\_id

**LEFT** **JOIN** categories c **ON** c.category\_id = p.category\_id

**WHERE** c.category\_name **IN** (**'Beverages'**, **'Confections'**, **'Dairy Products'**)

**GROUP** **BY** c.category\_name, p.product\_name, p.unit\_price

**ORDER** **BY** c.category\_name

)

**SELECT** ***Category***, ***product***, ***price***, ***quantity\_sold***,

**CASE**

**WHEN** ***most\_expensive*** = 1 **THEN** **'Most expensive'**

**WHEN** ***least\_expensive*** = 1 **THEN** **'Least Espensive'**

**ELSE** **'N/a'**

**END** **AS** *Max\_or\_min*

**FROM** ***cte***

**WHERE** ***most\_expensive*** = 1 **OR** ***LEAST\_expensive*** = 1

--d. Think of at least 2 additional metrics that may be of interest to highlight product performance

-- Date range from July 1996 to May 1998

--sales trends month/year

**SELECT** *c*.category\_name **AS** *Category*,

*p*.product\_name **AS** *Product*,

**date\_part**(**'month'**,*o*.order\_date) **AS** *Month\_ordered*,

**date\_part**(**'year'**, *o*.order\_date) **AS** *YEAR\_ordered*,

**round**(**CAST**(**sum**(*od*.quantity \* *od*.unit\_price)**AS** **NUMERIC**),2) **AS** *Total\_sales*

**FROM** order\_details *od*

**LEFT** **JOIN** orders *o* **ON** *o*.order\_id = *od*.order\_id

**LEFT** **JOIN** products *p* **ON** *p*.product\_id = *od*.product\_id

**LEFT** **JOIN** categories *c* **ON** *c*.category\_id = *p*.category\_id

**WHERE** *c*.category\_name **IN** (**'Beverages'**, **'Confections'**,**'Dairy Products'**)

**GROUP** **BY** *c*.category\_name, *p*.product\_name, *o*.order\_date

**order** **BY** ***total\_sales*** **DESC**

--2. total sales by product and how many times a product has been ordered

**SELECT** *p*.product\_name , *c*.category\_name,**count**(*od*.product\_id) **AS** *order\_count*,

**round**(**CAST**(**sum**(*od*.quantity \* *od*.unit\_price)**AS** **NUMERIC**),2) **AS** *Total\_sales*

**FROM** order\_details *od*

**LEFT** **JOIN** products *p* **ON** *p*.product\_id = *od*.product\_id

**LEFT** **JOIN** categories *c* **ON** *c*.category\_id = *p*.category\_id

**WHERE** *c*.category\_name **IN** (**'Beverages'**, **'Confections'**,**'Dairy Products'**)

**GROUP** **BY** *p*.product\_name, *c*.category\_name

**ORDER** **BY** **count**(*od*.product\_id) **desc**

--

--a. Sales performance by some time frame (you choose) and location

--i. Sales performance is how well or poorly a product or category performed over a given period of time

**SELECT** *c*.category\_name **AS** *Category*,

*o*.ship\_country **AS** *Ship\_Country*,

*o*.ship\_city **AS** *city*,

*p*.product\_name **AS** *Product*,

*o*.shipped\_date **AS** *Shipped\_Date*,

**date\_part**(**'month'**,*o*.shipped\_date) **AS** *shipped\_month*,

**date\_part**(**'year'**, *o*.shipped\_date) **AS** *shipped\_yr*,

**round**(**CAST**(**sum**(*od*.quantity \* *od*.unit\_price)**AS** **NUMERIC**),2) **AS** *Total\_sales*

**FROM** categories *c*

**LEFT** **JOIN** products *p* **ON** *p*.category\_id = *c*.category\_id

**LEFT** **JOIN** order\_details *od* **ON** *od*.product\_id = *p*.product\_id

**LEFT** **JOIN** orders *o* **ON** *o*.order\_id = *od*.order\_id

**LEFT** **JOIN** employees *e* **ON** *e*.employee\_id = *o*.employee\_id

**LEFT** **JOIN** employee\_territories *et* **ON** *et*.employee\_id = *e*.employee\_id

**WHERE** **date\_part**(**'year'**, *o*.shipped\_date) = **'1997'** **AND** category\_name = **'Confections'**

**GROUP** **BY** *c*.category\_name , *od*.unit\_price,*p*.product\_name,*o*.shipped\_date,*o*.ship\_country, *o*.ship\_city

**ORDER** **BY** *c*.category\_name

--b. Any insight into the shipments? shipped /not shipped/shipped late

--location orders/country

**WITH** cte **as**(

**SELECT** **DISTINCT** o.order\_id , o.order\_date, o.ship\_country, o.required\_date , o.shipped\_date ,

**date\_part**(**'day'**,o.required\_date::**timestamp** - o.order\_date::**timestamp**) **as** required\_num\_days,

**date\_part**(**'day'**,o.shipped\_date::**timestamp** - o.order\_date::**timestamp**) **as** actual\_days\_to\_ship,

s.company\_name **AS** shipped\_by

**FROM** orders o

**LEFT** **JOIN** shippers s **ON** s.shipper\_id = o.ship\_via

**LEFT** **JOIN** order\_details od **ON** od.order\_id = o.order\_id

**LEFT** **JOIN** products p **ON** p.product\_id = od.product\_id

**LEFT** **join** categories c **ON** c.category\_id = p.category\_id

**WHERE** c.category\_name **IN** (**'Beverages'**, **'Confections'**,**'Dairy Products'**)

**ORDER** **BY** order\_id )

**SELECT** \*,

**CASE** **WHEN** ***actual\_days\_to\_ship*** <= ***required\_num\_days*** **THEN** **'shipped\_on\_time'**

**WHEN** ***actual\_days\_to\_ship*** >= ***required\_num\_days*** **THEN** **'shipped\_late'**

**ELSE** **'not\_shipped'** **END** **AS** *shipped*

**FROM** ***cte***

-- shipping company used & avg cost

**SELECT** **count**(*o*.order\_id) ,

**round**(**cast**(**avg**(*o*.freight) **AS** **NUMERIC**),2) **AS** *avg\_shipping\_cost*,

*s*.company\_name **AS** *shipped\_by*

**FROM** orders *o*

**LEFT** **JOIN** shippers *s* **ON** *s*.shipper\_id = *o*.ship\_via

**GROUP** **BY** *s*.company\_name

**ORDER** **BY** **count**(*o*.order\_id) **DESC**

--c. Are there any trends in customer spending?

-- What are they buying – products- order size/spending

-- Top customers- total spent-

**SELECT** *c*.customer\_id,

*c*.company\_name **AS** *Company\_Name*,

*c*.country, *cc*.category\_name,

**round**(**CAST**(**sum**(*od*.quantity \* *od*.unit\_price)**AS** **NUMERIC**),2) **AS** *total\_spent*,

**count**(**DISTINCT** od.order\_id) **AS** *total\_orders*,

**count**(*p*.product\_id) **AS** *Total\_products\_purchased*

**FROM** customers *c*

**JOIN** orders *o* **ON** *o*.customer\_id = *c*.customer\_id

**JOIN** order\_details *od* **ON** *od*.order\_id = *o*.order\_id

**JOIN** products *p* **ON** *od*.product\_id = *p*.product\_id

**JOIN** categories *cc* **ON** *cc*.category\_id = *p*.category\_id

**WHERE** *cc*.category\_name **IN** (**'Beverages'**, **'Confections'**,**'Dairy Products'**)

**GROUP** **BY** *c*.customer\_id, *c*.company\_name, *c*.country, *cc*.category\_name

**ORDER** **BY** *c*.company\_name , ***total\_spent*** **desc**

---frequent customers, total orders, month they have placed an order,

--divide total orders by how many months they have ordered to get avg order per month

**WITH** cte **as**(

**SELECT** c.customer\_id , c.company\_name,

**count**( o.order\_id) **AS** total\_orders,

**count**(**DISTINCT** **date\_part**(**'month'**,o.shipped\_date)) **AS** order\_month,

**count**( o.order\_id) /**count**(**DISTINCT** **date\_part**(**'month'**,o.shipped\_date)) **AS** avg\_orders\_month

**FROM** customers c

**JOIN** orders o **ON** o.customer\_id = c.customer\_id

**JOIN** order\_details od **ON** od.order\_id = o.order\_id

**JOIN** products p **ON** od.product\_id = p.product\_id

**JOIN** categories cc **ON** cc.category\_id = p.category\_id

**WHERE** cc.category\_name **IN** (**'Beverages'**, **'Confections'**,**'Dairy Products'**)

**GROUP** **BY** c.customer\_id , c.company\_name

**ORDER** **BY** total\_orders **DESC** )

**SELECT** \*,

**CASE** **WHEN** order\_month **BETWEEN** 7 **AND** 12 **AND** total\_orders >= 20 **THEN** **'Frequent\_Customer'**

**WHEN** order\_month **BETWEEN** 4 **AND** 7 **AND** total\_orders >= 10 **THEN** **'Occasional\_customer'**

**WHEN** order\_month **BETWEEN** 1 **AND** 4 **AND** total\_orders <= 10 **THEN** **'Infrequent\_customer'**

**ELSE** **'Infrequent\_customer'** **END** **AS** Customer\_frequency

**FROM** cte