**Andy\_Chen\_Data\_assignment**

*File Orientation:*

* Items\_sold tab is a line item detail of all orders that our customers placed over a 2 month period
  + Product\_id = parent product id
  + Variant\_id = id of the specific variant within the product
  + SKU = internal identifier of item
  + Title = name of the parent product
  + Variant\_title = name of the variant product within the parent product
  + Name = Title combined with Variant\_title within a string
  + Quantity = Quantity Ordered
  + orderName = Order number
* Ids\_orders tab is a customer and order number association
  + Customer\_id = hashed id of the customer that placed the order, these numbers may appear more than once
  + orderName = Order number that will match with the orderName column in Items\_sold tab
  + Order\_created\_at = the date that the customer placed the order on our site
  + Cancelled\_at = reflects the date that the customer requested their order to be cancelled and no items shipped
* sku\_subcategory
  + Sku = will match SKU column in the Items\_sold tab
  + Product\_type = general sku category
  + Sub\_category = granular sku category

*Data Analysis Tasks:*

* Generally, which sku subcategories are a part of the first (sometimes only) order by the customer?

***Answer: Decorative Pillows, Duvet Covers, Sheet Sets***

SELECT d.sub\_category, count(d.sub\_category) AS [sum]

FROM (SELECT c.customer\_id, sku.sub\_category, count(sku.sub\_category)

FROM (SELECT \*

FROM (SELECT a.customer\_id, a.first\_time, orderName

FROM (SELECT customer\_id, MIN(order\_created\_at) AS first\_time

FROM ids\_orders AS ids WHERE ids.cancelled\_at IS NULL GROUP BY customer\_id) AS a

INNER JOIN ids\_orders AS ids ON a.first\_time = ids.order\_created\_at AND a.customer\_id = ids.customer\_id WHERE ids.cancelled\_at IS NULL) AS b

INNER JOIN items\_sold AS item ON b.orderName = item.orderName) AS c

INNER JOIN sku\_subcategory AS sku ON c.sku=sku.sku

GROUP BY c.customer\_id, sku.sub\_category) AS d

GROUP BY d.sub\_category

ORDER BY count(d.sub\_category) DESC;

* When people buy a duvet cover, comforter, or sheet set that is full, full/queen, or queen, what other item subcategories are also a part of that order?

***Answer: Decorative Pillows, Throws***

SELECT sub\_category, count(sub\_category) AS [sum]

FROM (SELECT b.customer\_id, sku.sub\_category, count(sku.sub\_category) AS sum

FROM(SELECT \*

FROM(SELECT \*

FROM (SELECT ids.customer\_id

FROM (SELECT \*

FROM items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName = ids.orderName WHERE ids.cancelled\_at IS NULL) AS a

INNER JOIN sku\_subcategory sku ON a.sku=sku.sku WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Full/Queen','Full','Queen')) AS b

INNER JOIN ids\_orders AS ids ON b.customer\_id = ids.customer\_id) AS c

INNER JOIN items\_sold AS item ON c.orderName=item.orderName) AS d

INNER JOIN sku\_subcategory AS sku ON sku.sku=d.sku

GROUP BY b.customer\_id, sku.sub\_category) AS [%$##@\_Alias]

GROUP BY sub\_category

ORDER BY count(sub\_category) DESC;

* When people buy a duvet cover, comforter, or sheet set that is full, full/queen, or queen, what other item subcategories are part of their **next** order?

***Answer: Decorative Pillows, Throws***

SELECT sub\_category, count(sub\_category) AS [sum]

FROM (SELECT ids.customer\_id, sub\_category, count(sub\_category)

FROM (((SELECT customer\_id

FROM (items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON item.sku=sku.sku

WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Full/Queen','Full','Queen')) AS a

INNER JOIN ids\_orders AS ids ON ids.customer\_id = a.customer\_id)

INNER JOIN items\_sold AS item ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON sku.sku=item.sku

WHERE item.orderName NOT IN (SELECT item.orderName

FROM (items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON item.sku=sku.sku

WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Full/Queen','Full','Queen'))

GROUP BY ids.customer\_id, sub\_category)

GROUP BY sub\_category

ORDER BY count(sub\_category) DESC;

* When people buy a duvet cover, comforter, or sheet set that is twin or twin XL what other item subcategories are also a part of that order?

***Answer: Decorative Pillows, Throws, Duvet Inserts***

SELECT sub\_category, count(sub\_category) AS [sum]

FROM (SELECT b.customer\_id, sku.sub\_category, count(sku.sub\_category) AS sum

FROM(SELECT \*

FROM(SELECT \*

FROM (SELECT ids.customer\_id

FROM (SELECT \*

FROM items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName = ids.orderName WHERE ids.cancelled\_at IS NULL) AS a

INNER JOIN sku\_subcategory sku ON a.sku=sku.sku WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Twin','Twin XL')) AS b

INNER JOIN ids\_orders AS ids ON b.customer\_id = ids.customer\_id) AS c

INNER JOIN items\_sold AS item ON c.orderName=item.orderName) AS d

INNER JOIN sku\_subcategory AS sku ON sku.sku=d.sku

GROUP BY b.customer\_id, sku.sub\_category) AS [%$##@\_Alias]

GROUP BY sub\_category

ORDER BY count(sub\_category) DESC;

* When people buy a duvet cover, comforter, or sheet set that is twin or twin XL what other item subcategories are part of their **next** order?

***Answer: Decorative Pillows, Wall Photo Display, Duvet Inserts, Throws, Wall Décor***

SELECT sub\_category, count(sub\_category) AS [sum]

FROM (SELECT ids.customer\_id, sub\_category, count(sub\_category)

FROM (((SELECT customer\_id

FROM (items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON item.sku=sku.sku

WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Twin','Twin XL')) AS a

INNER JOIN ids\_orders AS ids ON ids.customer\_id = a.customer\_id)

INNER JOIN items\_sold AS item ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON sku.sku=item.sku

WHERE item.orderName NOT IN (SELECT item.orderName

FROM (items\_sold AS item

INNER JOIN ids\_orders AS ids ON item.orderName=ids.orderName)

INNER JOIN sku\_subcategory AS sku ON item.sku=sku.sku

WHERE sub\_category IN('Duvet Covers','Sheet Sets','Comforters') AND variant\_title IN('Twin','Twin XL'))

GROUP BY ids.customer\_id, sub\_category)

GROUP BY sub\_category

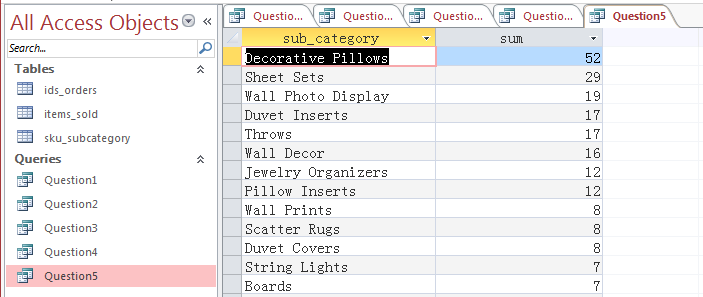
ORDER BY count(sub\_category) DESC;

*My Method:*

I built a database for storing the three sheets in Access as three tables. And then I wrote five SQL queries for each question. I got the above answers after I run the five SQL queries.

I attached my Access Database file: Andy\_Chen\_Data\_assignment.accdb. All the three tables and the five queries are in that file. And please check it through Microsoft Access. You can get all the results after you run all five queries.

The interface is as following:



Please check it. Thank you!