

# 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:

sub_category	sum
Decorative Pillows	52
Sheet Sets	29
Wall Photo Display	19
Duvet Inserts	17
Throws	17
Wall Decor	16
Jewelry Organizers	12
Pillow Inserts	12
Wall Prints	8
Scatter Rugs	8
Duvet Covers	8
String Lights	7
Boards	7

Please check it. Thank you!