You might be aware that CoDS-COMAD 2025, the prestigious international data science conference is happening by the end of the year. As a part of the conference, there is an associated data challenge to predict the key attributes from product images. Given an image of a dress item (in 5 categories: Men Tshirts,Sarees,Kurtis,Women Tshirts,Women Tops & Tunics), the task is to predict the attributes of the image such as color, sleeve_styling, transparency, fit_shape, pattern, length, etc. Each category of the dress item may have a different number of dress attributes. To download the data, you may have to fill a form to get access to the competition page. We are going to use our knowledge from manifold learning and dimension reduction lectures to visualize the dataset and discover interesting patterns and their association with product attributes.

```python
import pandas as pd
train_data=pd.read_csv("/kaggle/input/visual-taxonomy/train.csv")

import pyarrow.parquet as pa
attributes=pa.read_table("/kaggle/input/visual-taxonomy/category_attri
butes.parquet")
attributes
```

```
pyarrow.Table
Category: string
No_of_attribute: int64
Attribute_list: list<item: string>
  child 0, item: string
----
Category: [["Men Tshirts","Sarees","Kurtis","Women Tshirts","Women
Tops & Tunics"]]
No_of_attribute: [[5,10,9,8,10]]
Attribute_list:
[[["color","neck","pattern","print_or_pattern_type","sleeve_length"],
["blouse_pattern","border","border_width","color","occasion","ornament
ation","pallu_details","pattern","print_or_pattern_type","transparency
"],
["color","fit_shape","length","occasion","ornamentation","pattern","pr
int_or_pattern_type","sleeve_length","sleeve_styling"],
["color","fit_shape","length","pattern","print_or_pattern_type","sleev
e_length","sleeve_styling","surface_styling"],
["color","fit_shape","length","neck_collar","ocassion","pattern","prin
t_or_pattern_type","sleeve_length","sleeve_styling","surface_styling"]
]]
```

```python
train_data
```

```
          id             Category  len      attr_1   attr_2  
attr_3  \
0          0          Men Tshirts    5     default    round  printed  

1          1          Men Tshirts    5  multicolor     polo    solid  

2          2          Men Tshirts    5     default     polo    solid  
```

```
3          3          Men Tshirts    5  multicolor     polo    solid

4          4          Men Tshirts    5  multicolor     polo    solid

...      ...                 ...  ...         ...      ...      ...

70208  70374  Women Tops & Tunics   10  multicolor   fitted  regular

70209  70375  Women Tops & Tunics   10      yellow  regular     crop

70210  70376  Women Tops & Tunics   10      maroon   fitted     crop

70211  70377  Women Tops & Tunics   10         NaN      NaN      NaN

70212  70378  Women Tops & Tunics   10        pink     boxy     crop

            attr_4         attr_5   attr_6      attr_7         attr_8
\
0          default  short sleeves      NaN         NaN            NaN

1            solid  short sleeves      NaN         NaN            NaN

2            solid  short sleeves      NaN         NaN            NaN

3            solid  short sleeves      NaN         NaN            NaN

4            solid  short sleeves      NaN         NaN            NaN

...            ...            ...      ...         ...            ...

70208  square neck         casual  printed     default  short sleeves

70209   round neck         casual  default     default  short sleeves

70210   round neck         casual    solid       solid  short sleeves

70211         high            NaN      NaN         NaN  short sleeves

70212       v-neck         casual  printed  typography  short sleeves

                attr_9  attr_10
0                  NaN      NaN
1                  NaN      NaN
2                  NaN      NaN
3                  NaN      NaN
4                  NaN      NaN
...                ...      ...
70208  regular sleeves  ruffles
```

```
70209  regular sleeves  knitted
70210  regular sleeves  knitted
70211              NaN      NaN
70212  regular sleeves      NaN

[70213 rows x 13 columns]

train_data['Category'].unique

<bound method Series.unique of 0                Men Tshirts
1                Men Tshirts
2                Men Tshirts
3                Men Tshirts
4                Men Tshirts
                ...
70208    Women Tops & Tunics
70209    Women Tops & Tunics
70210    Women Tops & Tunics
70211    Women Tops & Tunics
70212    Women Tops & Tunics
Name: Category, Length: 70213, dtype: object>
```

Task 1

The challenge dataset contains ~70k training image in 5 categories with the respective attributes. For each category, pick any two attributes of your choice (say color or length or pattern, or any) to form baskets. Each basket is a <category, attribute> tuple. You will create two tuples per category, so in total you should have 10 baskets. Draw 100 samples from each basket. If your basket does not have 100 samples, reconfigure your basket by changing the attribute.

```
categories = ["Men Tshirts", "Sarees", "Kurtis", "Women Tshirts",
"Women Tops & Tunics"]
no_of_attributes = [5, 10, 9, 8, 10]

attribute_list = [
    ["color", "neck", "pattern", "print_or_pattern_type",
"sleeve_length"],
    ["blouse_pattern", "border", "border_width", "color", "occasion",
"ornamentation", "pallu_details", "pattern", "print_or_pattern_type",
"transparency"],
    ["color", "fit_shape", "length", "occasion", "ornamentation",
"pattern", "print_or_pattern_type", "sleeve_length",
"sleeve_styling"],
    ["color", "fit_shape", "length", "pattern",
"print_or_pattern_type", "sleeve_length", "sleeve_styling",
"surface_styling"],
    ["color", "fit_shape", "length", "neck_collar", "ocassion",
"pattern", "print_or_pattern_type", "sleeve_length", "sleeve_styling",
"surface_styling"]
]

mapped_attributes = []
for i, category in enumerate(categories):
    no_attrs = no_of_attributes[i]
    mapped_attrs = [f'attr_{j+1}' for j in range(no_attrs)]
    mapped_attributes.append(mapped_attrs)

baskets = []
for i, category in enumerate(categories):
    mapped_attrs = mapped_attributes[i]
    selected_attributes = random.sample(mapped_attrs, 2)
    for attribute in selected_attributes:
        baskets.append((category, attribute))

samples = {}
target_values = {}
target_counter = 0

for basket in baskets:
    category, attribute_column = basket
    basket_df = train_data[(train_data['Category'] == category) &
train_data[attribute_column].notna()]
    if len(basket_df) < 100:
        basket_df = train_data[train_data['Category'] ==
category].sample(100, replace=True)
    else:
        basket_df = basket_df.sample(100)
    target_values[basket] = target_counter
    basket_df['target'] = target_counter
    target_counter += 1
    samples[basket] = basket_df

for basket, sample_df in samples.items():
    print("\n")
    print(f"Basket: {basket}, Target: {target_values[basket]}")
    print("\n")
    print(sample_df)
```

```
Basket: ('Men Tshirts', 'attr_5'), Target: 0


        id     Category  len      attr_1 attr_2   attr_3
attr_4  \
7148  7272  Men Tshirts    5         NaN    NaN      NaN          NaN

3192  3192  Men Tshirts    5       white  round  printed   typography
```

```
1092  1092  Men Tshirts    5     default   polo    solid       solid

3182  3182  Men Tshirts    5       black   polo    solid       solid

4233  4233  Men Tshirts    5     default  round      NaN     default

...    ...          ...  ...         ...    ...      ...         ...

5162  5162  Men Tshirts    5     default   polo    solid       solid

1754  1754  Men Tshirts    5  multicolor   polo    solid       solid

3550  3550  Men Tshirts    5     default   polo    solid       solid

522    522  Men Tshirts    5     default  round      NaN     default

1048  1048  Men Tshirts    5     default  round  printed     default


             attr_5 attr_6 attr_7 attr_8 attr_9 attr_10  target
7148  short sleeves    NaN    NaN    NaN    NaN     NaN       0
3192  short sleeves    NaN    NaN    NaN    NaN     NaN       0
1092  short sleeves    NaN    NaN    NaN    NaN     NaN       0
3182  short sleeves    NaN    NaN    NaN    NaN     NaN       0
4233  short sleeves    NaN    NaN    NaN    NaN     NaN       0
...             ...    ...    ...    ...    ...     ...     ...
5162  short sleeves    NaN    NaN    NaN    NaN     NaN       0
1754  short sleeves    NaN    NaN    NaN    NaN     NaN       0
3550  short sleeves    NaN    NaN    NaN    NaN     NaN       0
522   short sleeves    NaN    NaN    NaN    NaN     NaN       0
1048  short sleeves    NaN    NaN    NaN    NaN     NaN       0

[100 rows x 14 columns]


Basket: ('Men Tshirts', 'attr_3'), Target: 1


        id     Category  len      attr_1 attr_2   attr_3
attr_4  \
1993  1993  Men Tshirts    5       white   polo    solid       solid

2945  2945  Men Tshirts    5       black  round  printed  typography

947    947  Men Tshirts    5  multicolor   polo    solid       solid

3354  3354  Men Tshirts    5     default   polo    solid       solid

3307  3307  Men Tshirts    5       white  round  printed     default

...    ...          ...  ...         ...    ...      ...         ...
```

```
2261  2261  Men Tshirts    5  multicolor   polo    solid       solid

6658  6661  Men Tshirts    5         NaN    NaN  printed     default

3173  3173  Men Tshirts    5       black   polo    solid       solid

452    452  Men Tshirts    5     default   polo  printed     default

3871  3871  Men Tshirts    5     default   polo    solid       solid


             attr_5 attr_6 attr_7 attr_8 attr_9 attr_10  target
1993  short sleeves    NaN    NaN    NaN    NaN     NaN       1
2945  short sleeves    NaN    NaN    NaN    NaN     NaN       1
947   short sleeves    NaN    NaN    NaN    NaN     NaN       1
3354  short sleeves    NaN    NaN    NaN    NaN     NaN       1
3307  short sleeves    NaN    NaN    NaN    NaN     NaN       1
...             ...    ...    ...    ...    ...     ...     ...
2261  short sleeves    NaN    NaN    NaN    NaN     NaN       1
6658  short sleeves    NaN    NaN    NaN    NaN     NaN       1
3173  short sleeves    NaN    NaN    NaN    NaN     NaN       1
452   short sleeves    NaN    NaN    NaN    NaN     NaN       1
3871  short sleeves    NaN    NaN    NaN    NaN     NaN       1

[100 rows x 14 columns]


Basket: ('Sarees', 'attr_2'), Target: 2


          id Category  len          attr_1         attr_2
attr_3  \
12578  12743   Sarees   10  same as border   woven design    big
border
12011  12176   Sarees   10   same as saree   woven design  small
border
19640  19805   Sarees   10   same as saree   woven design  small
border
21276  21441   Sarees   10   same as saree   woven design  small
border
20323  20488   Sarees   10             NaN           zari  small
border
...      ...      ...  ...             ...            ...           ..
.
7722    7887   Sarees   10   same as saree   woven design    big
border
10714  10879   Sarees   10  same as border  temple border    big
border
12441  12606   Sarees   10   same as saree   woven design    big
```

```
border
16054  16219   Sarees   10   same as saree   woven design    big
border
16058  16223   Sarees   10   same as saree   woven design  small
border

           attr_4       attr_5    attr_6         attr_7      attr_8  \
12578     default        daily       NaN  same as saree  zari woven
12011  multicolor        party  jacquard   woven design  zari woven
19640  multicolor        party  jacquard   woven design  zari woven
21276  multicolor        party  jacquard   woven design  zari woven
20323       white  traditional       NaN            NaN  zari woven
...           ...          ...       ...            ...         ...
7722   multicolor        party  jacquard   woven design  zari woven
10714   navy blue        party       NaN     zari woven       solid
12441  multicolor        party  jacquard   woven design  zari woven
16054  multicolor        party  jacquard   woven design  zari woven
16058  multicolor        party  jacquard   woven design  zari woven

             attr_9 attr_10  target
12578  ethnic motif      no       2
12011      applique      no       2
19640      applique      no       2
21276           NaN      no       2
20323       peacock      no       2
...             ...     ...     ...
7722            NaN      no       2
10714         solid      no       2
12441           NaN      no       2
16054           NaN      no       2
16058           NaN      no       2

[100 rows x 14 columns]


Basket: ('Sarees', 'attr_7'), Target: 3


          id Category  len         attr_1        attr_2        attr_3
\
15609  15774   Sarees   10  same as saree  woven design  small border

20271  20436   Sarees   10  same as saree          zari  small border

12154  12319   Sarees   10  same as saree  woven design  small border

11167  11332   Sarees   10        default     no border     no border

15456  15621   Sarees   10  same as saree          zari  small border
```

```
...      ...      ...  ...            ...           ...           ...

17301  17466   Sarees   10  same as saree  woven design  small border

19037  19202   Sarees   10        default  woven design  small border

14953  15118   Sarees   10        default     no border     no border

18967  19132   Sarees   10  same as saree           zari  small border

9454    9619   Sarees   10  same as saree  woven design  small border


           attr_4 attr_5    attr_6         attr_7      attr_8
attr_9  \
15609  multicolor  party  jacquard   woven design  zari woven
applique
20271       cream  party  jacquard   woven design  zari woven
floral
12154  multicolor  party  jacquard   woven design  zari woven
applique
11167     default  party       NaN     zari woven  zari woven
default
15456       cream  party  jacquard   woven design  zari woven
peacock
...           ...    ...       ...            ...         ...
...
17301  multicolor  party  jacquard   woven design  zari woven
NaN
19037  multicolor  daily       NaN  same as saree  zari woven
default
14953         NaN    NaN       NaN  same as saree         NaN
default
18967       cream  party  jacquard   woven design  zari woven  ethnic
motif
9454   multicolor  party  jacquard   woven design  zari woven
NaN

      attr_10  target
15609      no       3
20271     yes       3
12154      no       3
11167      no       3
15456      no       3
...       ...     ...
17301      no       3
19037      no       3
14953      no       3
18967      no       3
9454       no       3
```

```
[100 rows x 14 columns]


Basket: ('Kurtis', 'attr_4'), Target: 4


          id Category  len      attr_1    attr_2       attr_3 attr_4
attr_5  \
29475  29640   Kurtis    9        blue       NaN  knee length  daily
net
30010  30175   Kurtis    9       black       NaN          NaN  daily
net
32088  32254   Kurtis    9  multicolor       NaN          NaN  daily
NaN
31379  31544   Kurtis    9   navy blue       NaN          NaN  daily
net
29901  30066   Kurtis    9         red       NaN  calf length  daily
net
...      ...      ...  ...         ...       ...          ...    ...
...
28322  28487   Kurtis    9         red       NaN  knee length  daily
net
27529  27694   Kurtis    9      yellow    a-line  calf length  daily
default
27014  27179   Kurtis    9       black  straight  knee length  daily
net
27673  27838   Kurtis    9       green    a-line  calf length  daily
NaN
26947  27112   Kurtis    9      maroon       NaN          NaN  party
default

        attr_6   attr_7                  attr_8   attr_9 attr_10
target
29475      NaN    solid  three-quarter sleeves  regular     NaN
4
30010    solid    solid  three-quarter sleeves  regular     NaN
4
32088      NaN      NaN  three-quarter sleeves  regular     NaN
4
31379      NaN      NaN          short sleeves  regular     NaN
4
29901      NaN      NaN  three-quarter sleeves  regular     NaN
4
...        ...      ...                    ...      ...     ...      ..
.
28322    solid    solid  three-quarter sleeves  regular     NaN
4
27529    solid    solid  three-quarter sleeves  regular     NaN
4
```

```
27014  default  default  three-quarter sleeves  regular     NaN
4
27673    solid    solid  three-quarter sleeves  regular     NaN
4
26947  default  default  three-quarter sleeves      NaN     NaN
4

[100 rows x 14 columns]


Basket: ('Kurtis', 'attr_3'), Target: 5


          id Category  len     attr_1    attr_2       attr_3 attr_4
attr_5  \
27570  27735   Kurtis    9        NaN       NaN  calf length  daily
default
26839  27004   Kurtis    9     maroon  straight  knee length  daily
net
29554  29719   Kurtis    9        red    a-line  knee length  daily
NaN
28313  28478   Kurtis    9     purple    a-line  knee length  daily
net
29991  30156   Kurtis    9  navy blue       NaN  knee length  daily
NaN
...      ...      ...  ...        ...       ...          ...    ...
...
27797  27962   Kurtis    9        red    a-line  calf length  daily
NaN
29934  30099   Kurtis    9      black  straight  knee length  daily
NaN
27709  27874   Kurtis    9       grey    a-line  knee length  daily
NaN
29452  29617   Kurtis    9     purple       NaN  calf length  daily
NaN
26392  26557   Kurtis    9      black  straight  knee length  daily
net

        attr_6   attr_7                 attr_8   attr_9 attr_10
target
27570  default  default  three-quarter sleeves  regular     NaN
5
26839  default  default  three-quarter sleeves  regular     NaN
5
29554    solid      NaN         short sleeves  regular     NaN
5
28313    solid    solid  three-quarter sleeves  regular     NaN
5
29991    solid    solid         short sleeves  regular     NaN
5
```

```
...        ...      ...                    ...      ...     ...     ..
.
27797    solid    solid  three-quarter sleeves  regular     NaN
5
29934      NaN      NaN  three-quarter sleeves  regular     NaN
5
27709  default  default  three-quarter sleeves  regular     NaN
5
29452    solid      NaN  three-quarter sleeves  regular     NaN
5
26392  default  default  three-quarter sleeves  regular     NaN
5

[100 rows x 14 columns]


Basket: ('Women Tshirts', 'attr_7'), Target: 6


          id       Category  len      attr_1   attr_2   attr_3
attr_4  \
34936  35102  Women Tshirts    8     default      NaN  regular
printed
37164  37330  Women Tshirts    8       black  regular     long
printed
47661  47827  Women Tshirts    8       white  regular  regular
printed
42752  42918  Women Tshirts    8     default  regular     crop
printed
42929  43095  Women Tshirts    8        pink  regular  regular
printed
...      ...            ...  ...         ...      ...      ...      ..
.
33449  33615  Women Tshirts    8  multicolor  regular     long
default
43509  43675  Women Tshirts    8       white  regular  regular
printed
41511  41677  Women Tshirts    8      yellow  regular     crop
printed
40351  40517  Women Tshirts    8     default  regular  regular
printed
42075  42241  Women Tshirts    8       black  regular  regular
printed

            attr_5         attr_6           attr_7 attr_8 attr_9
attr_10  \
34936      default  short sleeves  regular sleeves    NaN    NaN
NaN
37164       quirky  short sleeves  regular sleeves    NaN    NaN
NaN
```

```
47661  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
42752       quirky  short sleeves  regular sleeves    NaN    NaN
NaN
42929      default  short sleeves  regular sleeves    NaN    NaN
NaN
...            ...            ...              ...    ...    ...     .
..
33449      default  short sleeves  regular sleeves    NaN    NaN
NaN
43509       quirky  short sleeves  regular sleeves    NaN    NaN
NaN
41511       quirky  short sleeves  regular sleeves    NaN    NaN
NaN
40351  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
42075      graphic  short sleeves  regular sleeves    NaN    NaN
NaN

       target
34936       6
37164       6
47661       6
42752       6
42929       6
...       ...
33449       6
43509       6
41511       6
40351       6
42075       6

[100 rows x 14 columns]


Basket: ('Women Tshirts', 'attr_1'), Target: 7


          id       Category  len  attr_1   attr_2   attr_3   attr_4  \
41120  41286  Women Tshirts    8  yellow  regular     crop  printed
44666  44832  Women Tshirts    8   white  regular  regular  printed
47032  47198  Women Tshirts    8   white  regular  regular  printed
48701  48867  Women Tshirts    8   white  regular     crop  printed
40120  40286  Women Tshirts    8   black  regular  regular  printed
...      ...            ...  ...     ...      ...      ...      ...
35680  35846  Women Tshirts    8  maroon  regular  regular    solid
37292  37458  Women Tshirts    8   white    loose     long  printed
40462  40628  Women Tshirts    8    pink  regular  regular  printed
37564  37730  Women Tshirts    8   black  regular     crop  printed
33346  33512  Women Tshirts    8    pink  regular  regular  printed
```

```
            attr_5         attr_6           attr_7 attr_8 attr_9
attr_10  \
41120  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
44666  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
47032  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
48701   typography  short sleeves  regular sleeves    NaN    NaN
NaN
40120       quirky  short sleeves  regular sleeves    NaN    NaN
NaN
...            ...            ...              ...    ...    ...      .
..
35680        solid  short sleeves  regular sleeves    NaN    NaN
NaN
37292      graphic  short sleeves  regular sleeves    NaN    NaN
NaN
40462  funky print  short sleeves  regular sleeves    NaN    NaN
NaN
37564  funky print   long sleeves  regular sleeves    NaN    NaN
NaN
33346      graphic   long sleeves   cuffed sleeves    NaN    NaN
NaN

       target
41120       7
44666       7
47032       7
48701       7
40120       7
...       ...
35680       7
37292       7
40462       7
37564       7
33346       7

[100 rows x 14 columns]


Basket: ('Women Tops & Tunics', 'attr_10'), Target: 8


          id             Category  len     attr_1   attr_2   attr_3  \
52470  52636  Women Tops & Tunics   10    default   fitted  regular
64327  64493  Women Tops & Tunics   10        NaN   fitted      NaN
56783  56949  Women Tops & Tunics   10  navy blue   fitted     crop
59872  60038  Women Tops & Tunics   10      white  regular     crop
```

```
51495  51661  Women Tops & Tunics   10     maroon   fitted  regular
...      ...                  ...  ...        ...      ...      ...
67178  67344  Women Tops & Tunics   10      white  regular     crop
54321  54487  Women Tops & Tunics   10      peach   fitted  regular
51957  52123  Women Tops & Tunics   10     yellow  default      NaN
63027  63193  Women Tops & Tunics   10      black   fitted  regular
60341  60507  Women Tops & Tunics   10        NaN  regular  regular

            attr_4  attr_5   attr_6      attr_7                 attr_8
\
52470   round neck  casual    solid       solid          short sleeves

64327  square neck  casual    solid         NaN           long sleeves

56783   round neck  casual    solid       solid          short sleeves

59872   round neck  casual  printed  typography          short sleeves

51495  square neck  casual    solid       solid  three-quarter sleeves

...            ...     ...      ...         ...                    ...

67178   round neck  casual  printed      quirky          short sleeves

54321   round neck  casual    solid       solid          short sleeves

51957      default     NaN  default      floral  three-quarter sleeves

63027         high  casual    solid       solid          short sleeves

60341      default  casual    solid       solid  three-quarter sleeves


                attr_9  attr_10  target
52470     puff sleeves  ruffles       8
64327              NaN  knitted       8
56783          default  knitted       8
59872  regular sleeves  default       8
51495          default  knitted       8
...                ...      ...     ...
67178  regular sleeves  default       8
54321              NaN  default       8
51957  regular sleeves  default       8
63027              NaN  knitted       8
60341  regular sleeves  default       8

[100 rows x 14 columns]


Basket: ('Women Tops & Tunics', 'attr_5'), Target: 9
```

```
          id             Category  len      attr_1   attr_2  
attr_3  \
51996  52162  Women Tops & Tunics   10        blue  regular  regular

69282  69448  Women Tops & Tunics   10      yellow  default     crop

62491  62657  Women Tops & Tunics   10   navy blue  default  regular

64240  64406  Women Tops & Tunics   10        pink     boxy     crop

59670  59836  Women Tops & Tunics   10  multicolor  regular  regular

...      ...                  ...  ...         ...      ...      ...

52591  52757  Women Tops & Tunics   10         red     boxy  regular

53003  53169  Women Tops & Tunics   10        blue  regular  regular

62617  62783  Women Tops & Tunics   10       black  default  regular

65628  65794  Women Tops & Tunics   10      yellow   fitted     crop

68001  68167  Women Tops & Tunics   10       black     boxy     crop


           attr_4  attr_5   attr_6      attr_7         attr_8  \
51996  round neck  casual  default     default  short sleeves
69282      v-neck   party    solid       solid  short sleeves
62491        high  casual    solid       solid  short sleeves
64240  round neck  casual  printed  typography   long sleeves
59670     default  casual  printed     default  short sleeves
...           ...     ...      ...         ...            ...
52591  round neck  casual  printed     default   long sleeves
53003  round neck  casual  printed     graphic  short sleeves
62617  round neck  casual    solid       solid   long sleeves
65628     default  casual    solid       solid  short sleeves
68001  round neck  casual  printed  typography  short sleeves

                attr_9        attr_10  target
51996  regular sleeves            NaN       9
69282  regular sleeves        knitted       9
62491          default       applique       9
64240  regular sleeves            NaN       9
59670  regular sleeves        tie-ups       9
...                ...            ...     ...
52591  regular sleeves  waist tie-ups       9
53003  regular sleeves       applique       9
62617  regular sleeves       applique       9
65628  regular sleeves            NaN       9
68001  regular sleeves            NaN       9
```

```
[100 rows x 14 columns]
```

```
baskets
```

```
[('Men Tshirts', 'attr_5'),
 ('Men Tshirts', 'attr_3'),
 ('Sarees', 'attr_2'),
 ('Sarees', 'attr_7'),
 ('Kurtis', 'attr_4'),
 ('Kurtis', 'attr_3'),
 ('Women Tshirts', 'attr_7'),
 ('Women Tshirts', 'attr_1'),
 ('Women Tops & Tunics', 'attr_10'),
 ('Women Tops & Tunics', 'attr_5')]
```

```
import pandas as pd
import random

categories = ["Men Tshirts", "Sarees", "Kurtis", "Women Tshirts",
"Women Tops & Tunics"]
no_of_attributes = [5, 10, 9, 8, 10]

attribute_list = [
    ["color", "neck", "pattern", "print_or_pattern_type",
"sleeve_length"],
    ["blouse_pattern", "border", "border_width", "color", "occasion",
"ornamentation", "pallu_details", "pattern", "print_or_pattern_type",
"transparency"],
    ["color", "fit_shape", "length", "occasion", "ornamentation",
"pattern", "print_or_pattern_type", "sleeve_length",
"sleeve_styling"],
    ["color", "fit_shape", "length", "pattern",
"print_or_pattern_type", "sleeve_length", "sleeve_styling",
"surface_styling"],
    ["color", "fit_shape", "length", "neck_collar", "occasion",
"pattern", "print_or_pattern_type", "sleeve_length", "sleeve_styling",
"surface_styling"]
]

mapped_attributes = []
for i, category in enumerate(categories):
    no_attrs = no_of_attributes[i]
    mapped_attrs = [f'attr_{j+1}' for j in range(no_attrs)]
    mapped_attributes.append(mapped_attrs)

baskets = []
for i, category in enumerate(categories):
    mapped_attrs = mapped_attributes[i]
    selected_attributes = random.sample(mapped_attrs, 2)
    for attribute in selected_attributes:
        baskets.append((category, attribute))

samples = []
target_values = {}
target_counter = 0

for basket in baskets:
    category, attribute_column = basket
    basket_df = train_data[(train_data['Category'] == category) &
train_data[attribute_column].notna()]

    if len(basket_df) < 100:
        basket_df = train_data[train_data['Category'] ==
category].sample(100, replace=True)
    else:
        basket_df = basket_df.sample(100)

    target_values[basket] = target_counter
    basket_df['target'] = target_counter
    target_counter += 1

    samples.append(basket_df)

final_df = pd.concat(samples, ignore_index=True)
final_df
```

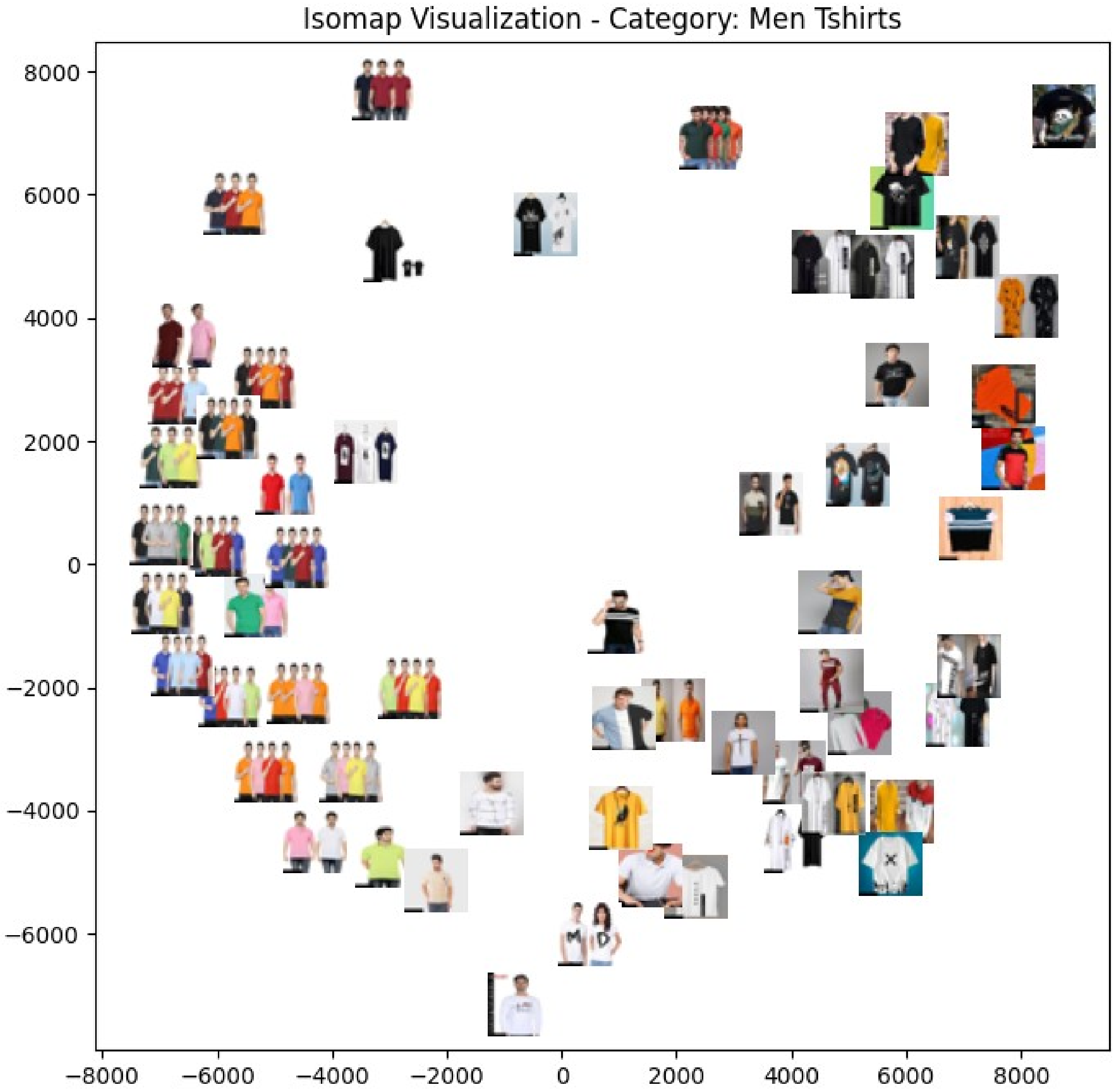
```
        id              Category  len   attr_1   attr_2   attr_3
attr_4  \
0     5641           Men Tshirts    5      NaN    round    solid
NaN
1     6257           Men Tshirts    5      NaN    round      NaN
NaN
2     2795           Men Tshirts    5    white    round  printed
default
3     1135           Men Tshirts    5  default     polo    solid
solid
4     5453           Men Tshirts    5  default    round      NaN
default
..     ...                   ...  ...      ...      ...      ...
...
995  64666  Women Tops & Tunics   10    white     boxy     crop  round
neck
996  60174  Women Tops & Tunics   10    green  default      NaN
NaN
997  59481  Women Tops & Tunics   10    white  regular      NaN
NaN
998  59400  Women Tops & Tunics   10    white   fitted      NaN
NaN
999  65473  Women Tops & Tunics   10     pink   fitted  regular  round
```

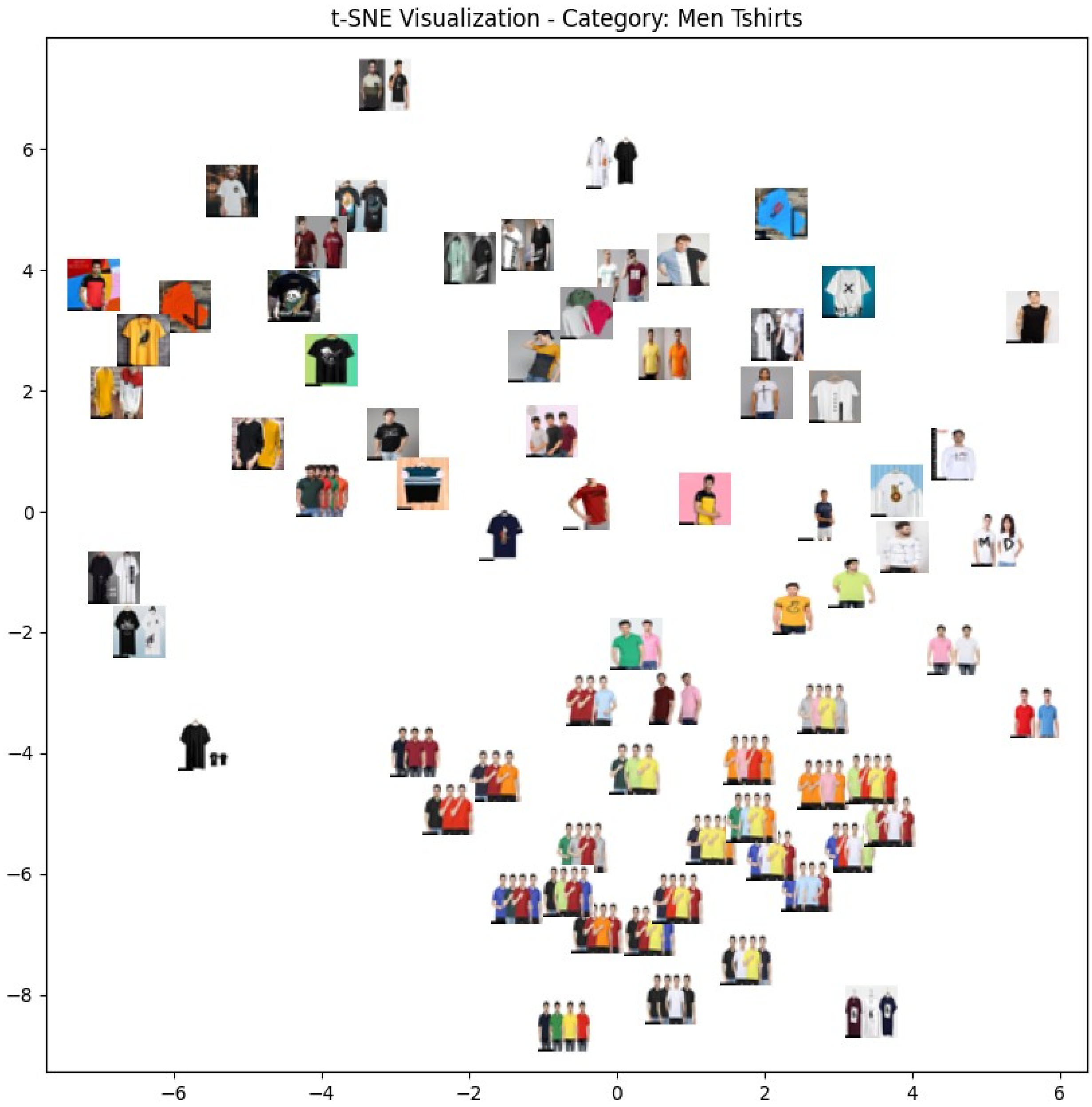
```
neck

            attr_5   attr_6      attr_7                    attr_8  \
0              NaN      NaN         NaN                       NaN
1              NaN      NaN         NaN                       NaN
2    short sleeves      NaN         NaN                       NaN
3    short sleeves      NaN         NaN                       NaN
4    short sleeves      NaN         NaN                       NaN
..             ...      ...         ...                       ...
995         casual  printed  typography              long sleeves
996            NaN  default      floral  three-quarter sleeves
997            NaN    solid       solid                       NaN
998            NaN    solid         NaN                       NaN
999         casual    solid       solid             short sleeves

              attr_9  attr_10  target
0                NaN      NaN       0
1                NaN      NaN       0
2                NaN      NaN       0
3                NaN      NaN       0
4                NaN      NaN       0
..               ...      ...     ...
995  regular sleeves      NaN       9
996              NaN      NaN       9
997              NaN      NaN       9
998              NaN      NaN       9
999              NaN  knitted       9

[1000 rows x 14 columns]
```

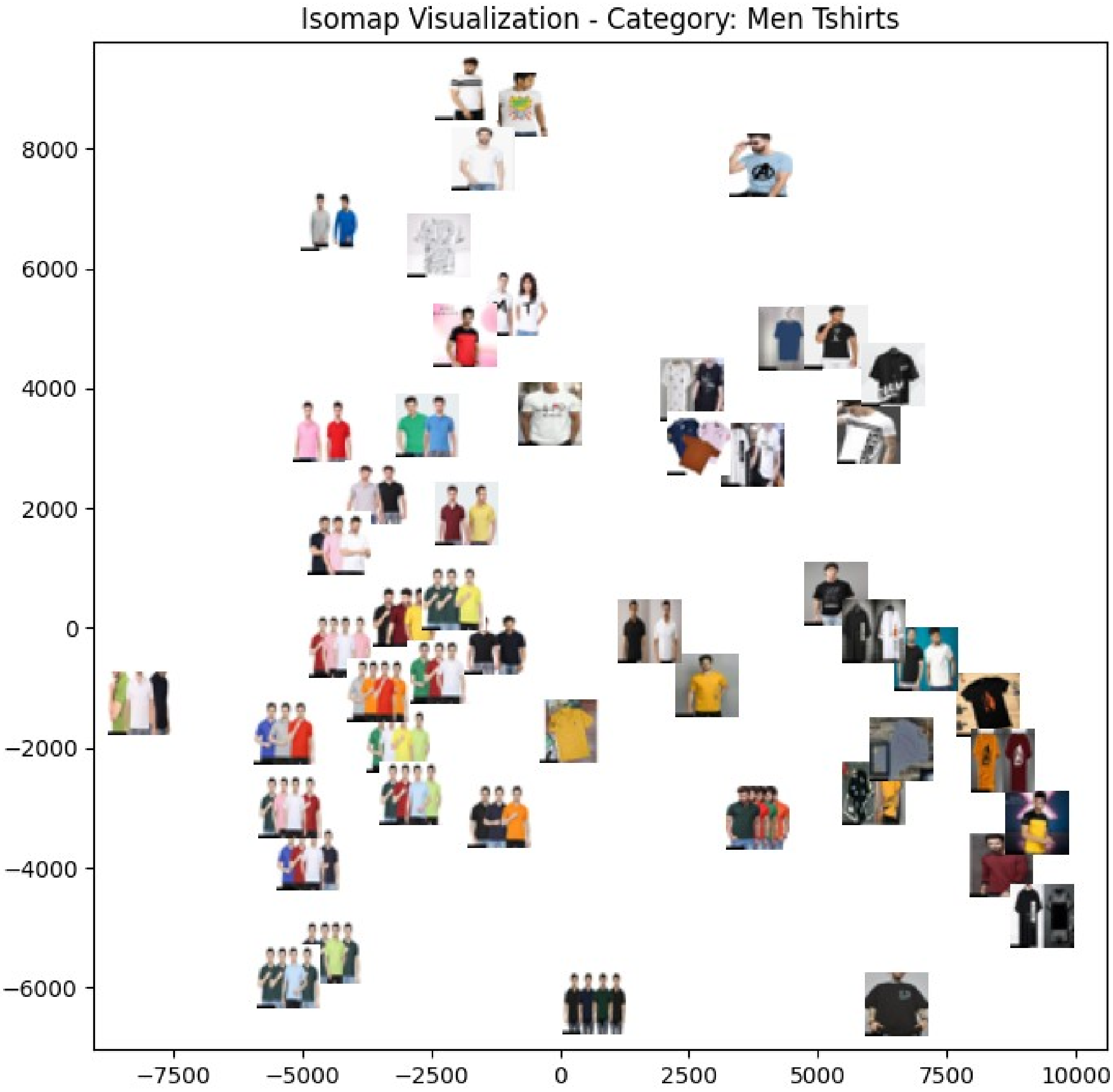
Task 2

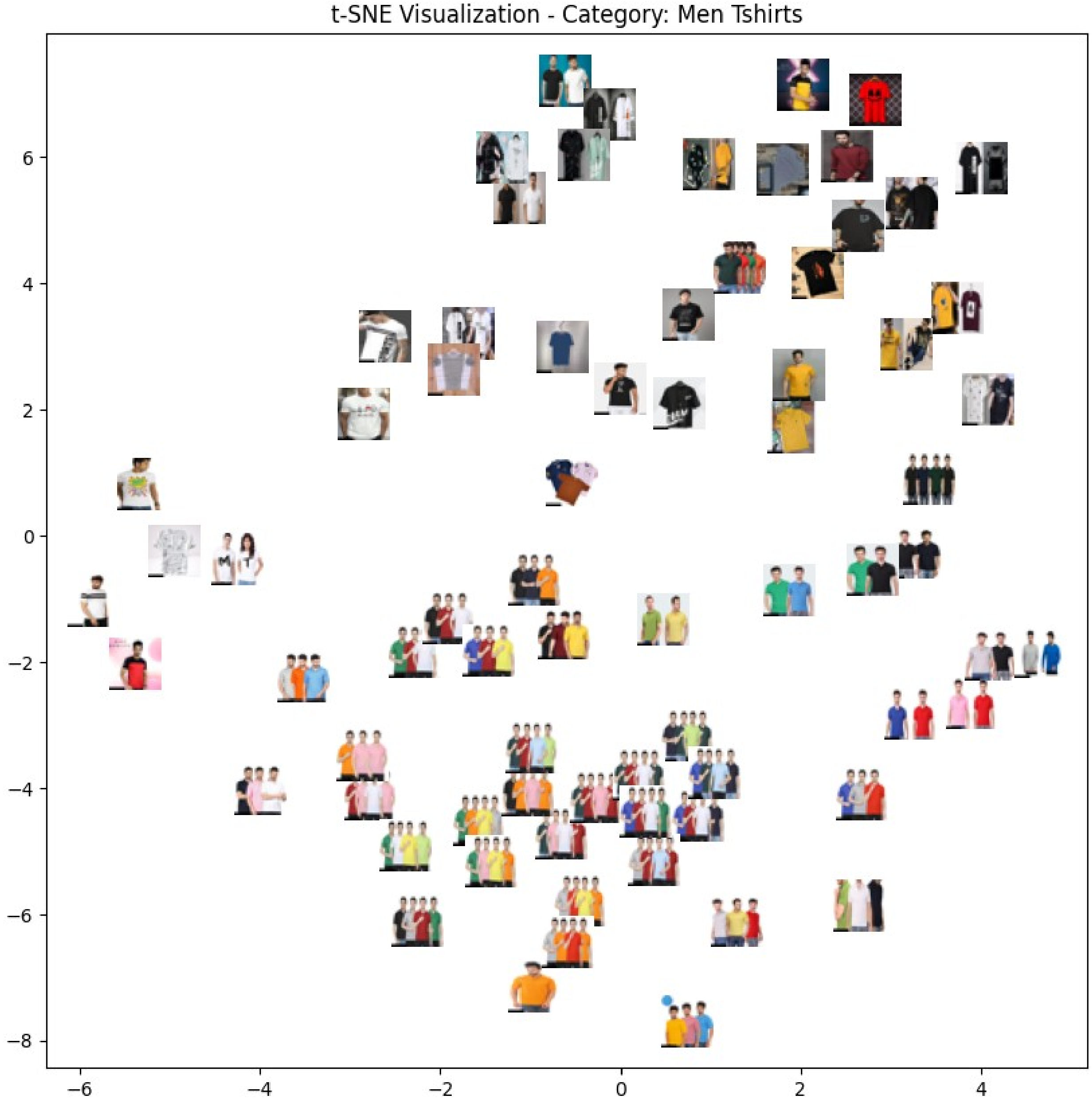
Create the visualization like below (which we reviewed in the class) for each basket. You should use Isomap and tSNE with two components, which would represent the intrinsic dimensions of the manifold on which the dataset resides. You will have 10 visuals using Isomap and another 10 visuals vis tSNE.

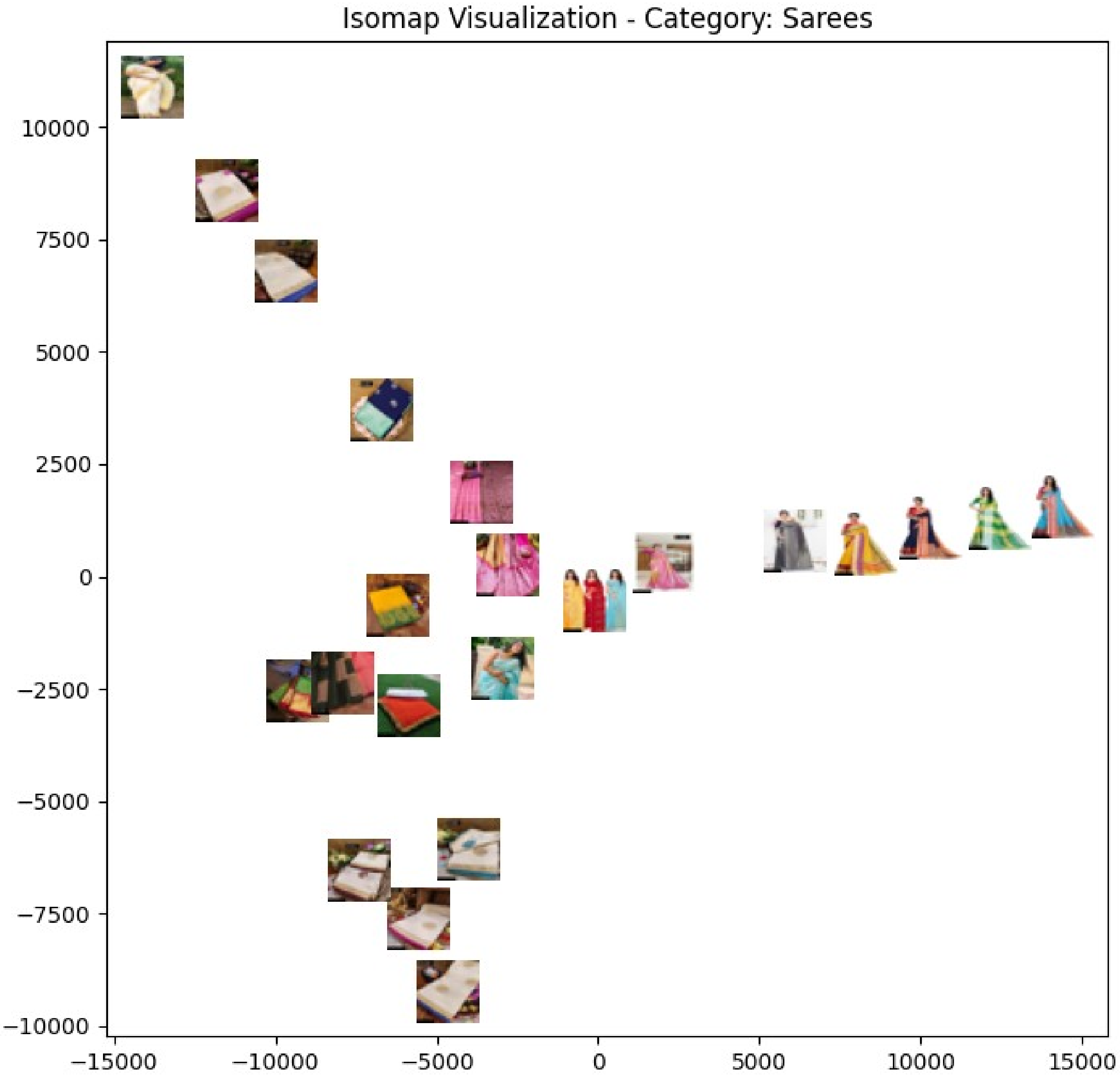
```python
import matplotlib.pyplot as plt
from sklearn.manifold import Isomap, TSNE
from matplotlib.offsetbox import OffsetImage, AnnotationBbox
from PIL import Image
import numpy as np
import os

image_dir = '/kaggle/input/visual-taxonomy/train_images/'

def plot_components(data, model, images, ax=None, thumb_frac=0.05):
    proj = model.fit_transform(data)
    ax = ax or plt.gca()
    ax.plot(proj[:, 0], proj[:, 1], '.', alpha=0)
    min_dist = (proj.max(0) - proj.min(0)) * thumb_frac
    shown_images = np.array([2 * min_dist])
    for i in range(proj.shape[0]):
        dist = np.sum((proj[i] - shown_images) ** 2, axis=1)
        if np.min(dist) < min_dist[0] ** 2:
            continue
        shown_images = np.vstack([shown_images, proj[i]])
        imagebox = OffsetImage(images[i])
        ab = AnnotationBbox(imagebox, proj[i], frameon=False)
        ax.add_artist(ab)

num_samples_per_category = 100
num_categories = len(final_df) // num_samples_per_category

for i in range(num_categories):
    start_idx = i * num_samples_per_category
    end_idx = start_idx + num_samples_per_category
    sample_df = final_df.iloc[start_idx:end_idx]
    category_name = sample_df['Category'].iloc[0]
    images = []
    for image_id in sample_df['id']:
        padded_id = str(image_id).zfill(6)
        image_path = os.path.join(image_dir, f"{padded_id}.jpg")
        image = np.array(Image.open(image_path).resize((28, 28)))
        images.append(image)
    images = np.stack(images)
    flattened_images = images.reshape(len(images), -1)

    fig, ax = plt.subplots(figsize=(8,8))
    isomap = Isomap(n_neighbors=5, n_components=2)
    plot_components(flattened_images, isomap, images=images, ax=ax,
thumb_frac=0.05)
    ax.set_title(f"Isomap Visualization - Category: {category_name}")
    plt.show()

    fig, ax = plt.subplots(figsize=(10, 10))
    tsne = TSNE(n_components=2, random_state=0)
    plot_components(flattened_images, tsne, images=images, ax=ax,
thumb_frac=0.05)
    ax.set_title(f"t-SNE Visualization - Category: {category_name}")
    plt.show()
```

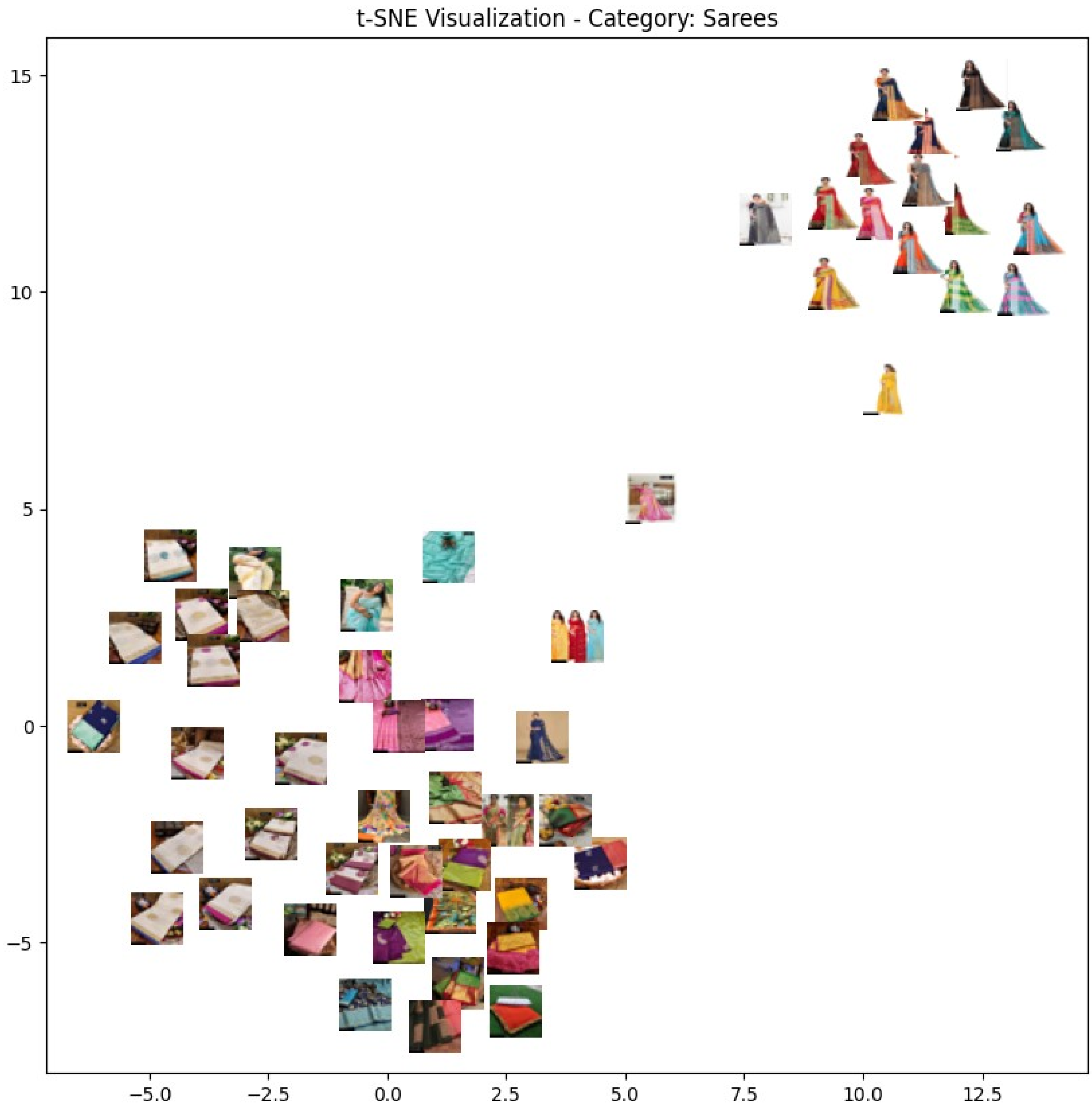

Isomap Visualization - Category: Men Tshirts
8000
6000
4000
2000
0
−2000
−4000
−6000
−8000
−6000
−4000
−2000
0
2000
4000
6000
8000


t-SNE Visualization - Category: Men Tshirts
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6

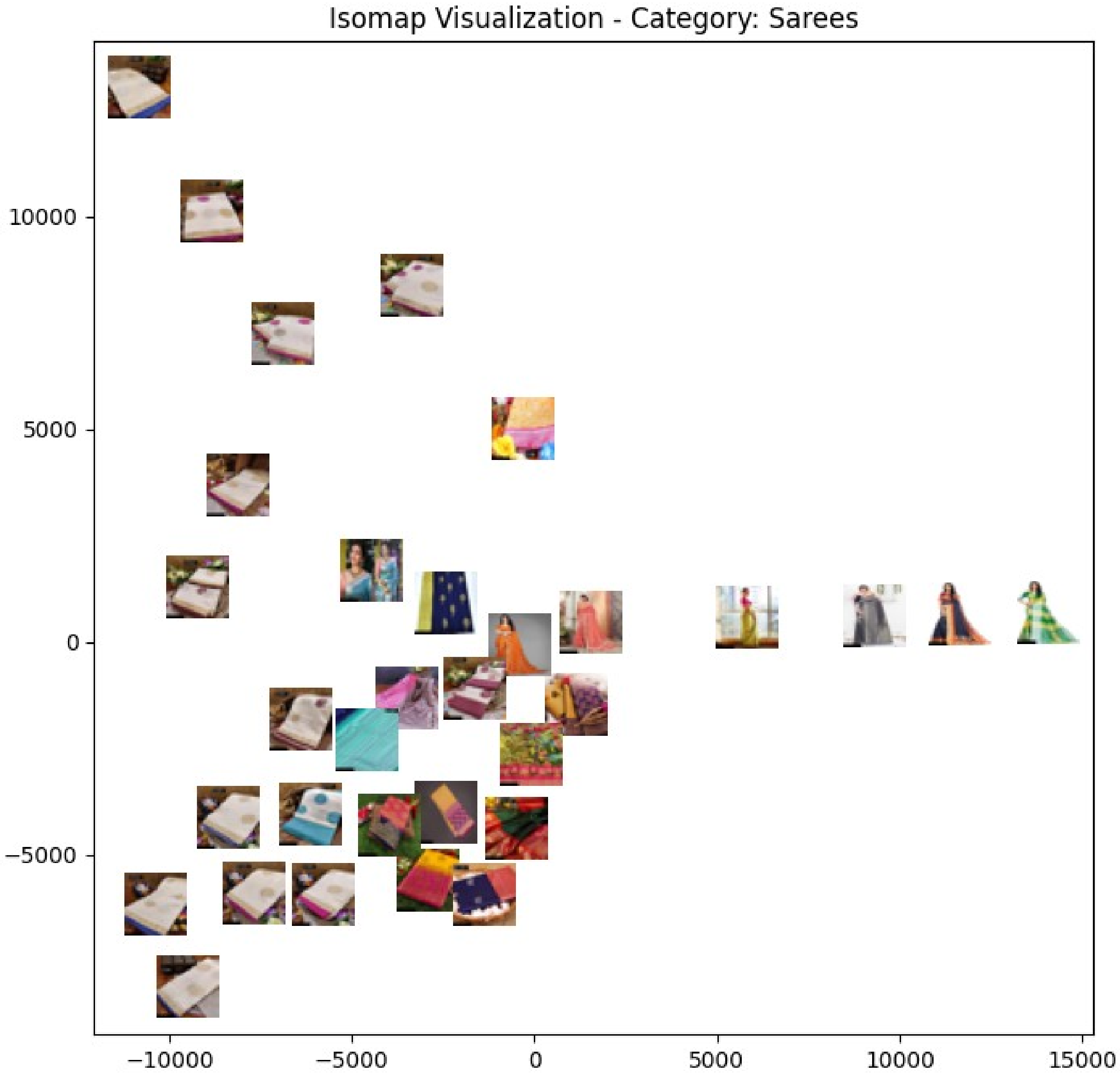

Isomap Visualization - Category: Men Tshirts
8000
6000
4000
2000
0
−2000
−4000
−6000
−7500
−5000
−2500
0
2500
5000
7500
10000

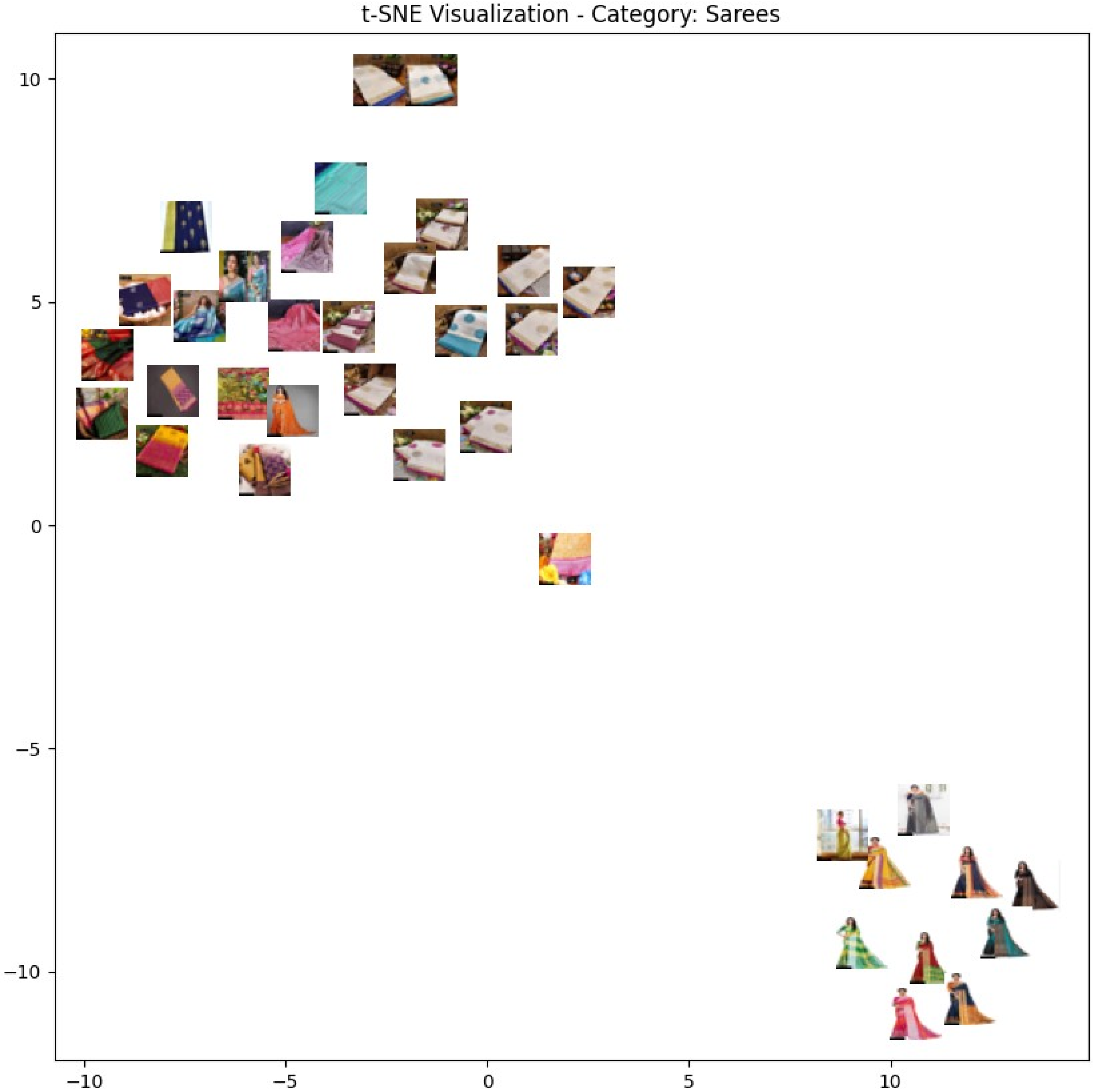

t-SNE Visualization - Category: Men Tshirts
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4

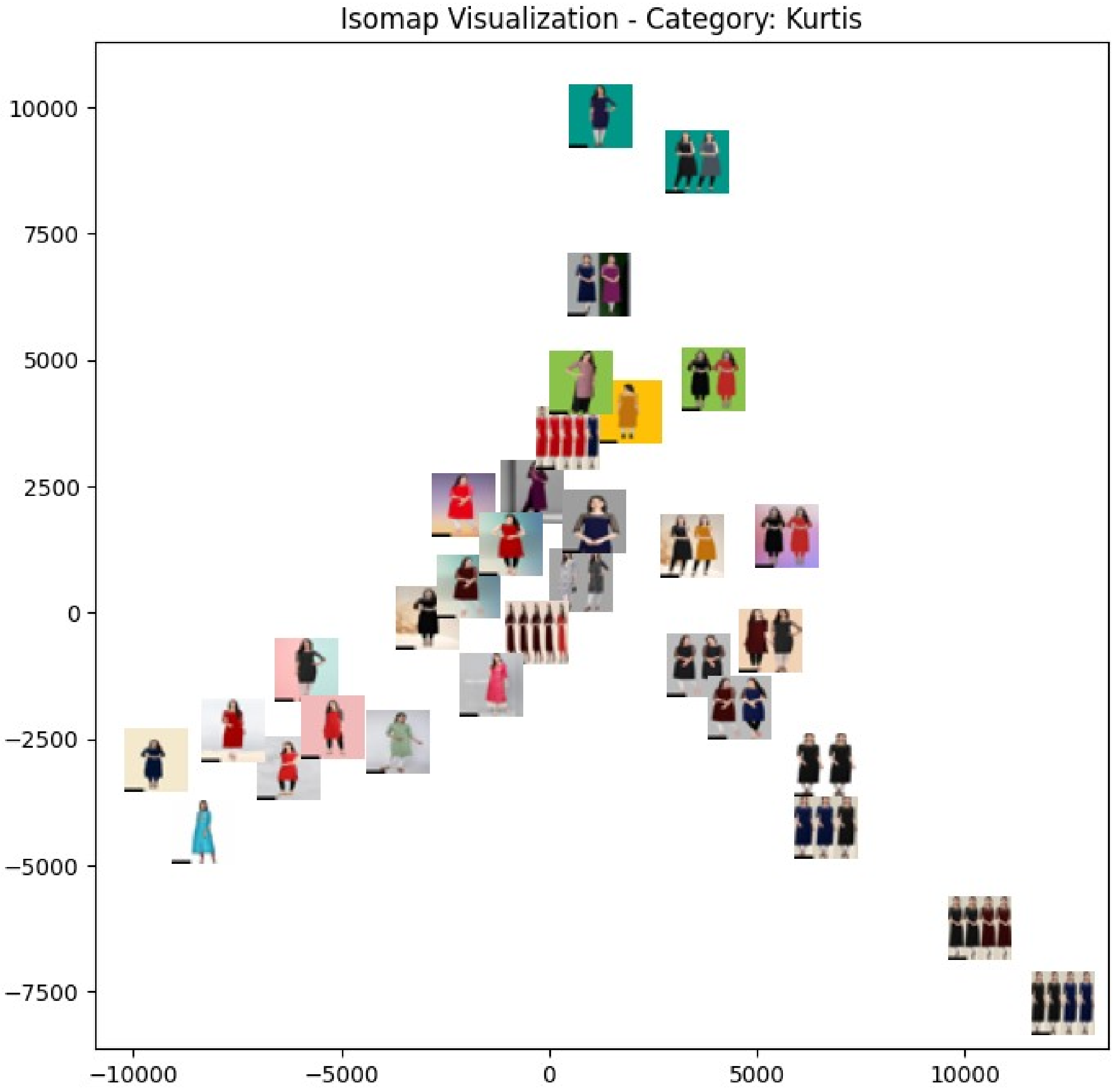

Isomap Visualization - Category: Sarees
10000
7500
5000
2500
0
−2500
−5000
−7500
−10000
−15000
−10000
−5000
0
5000
10000
15000

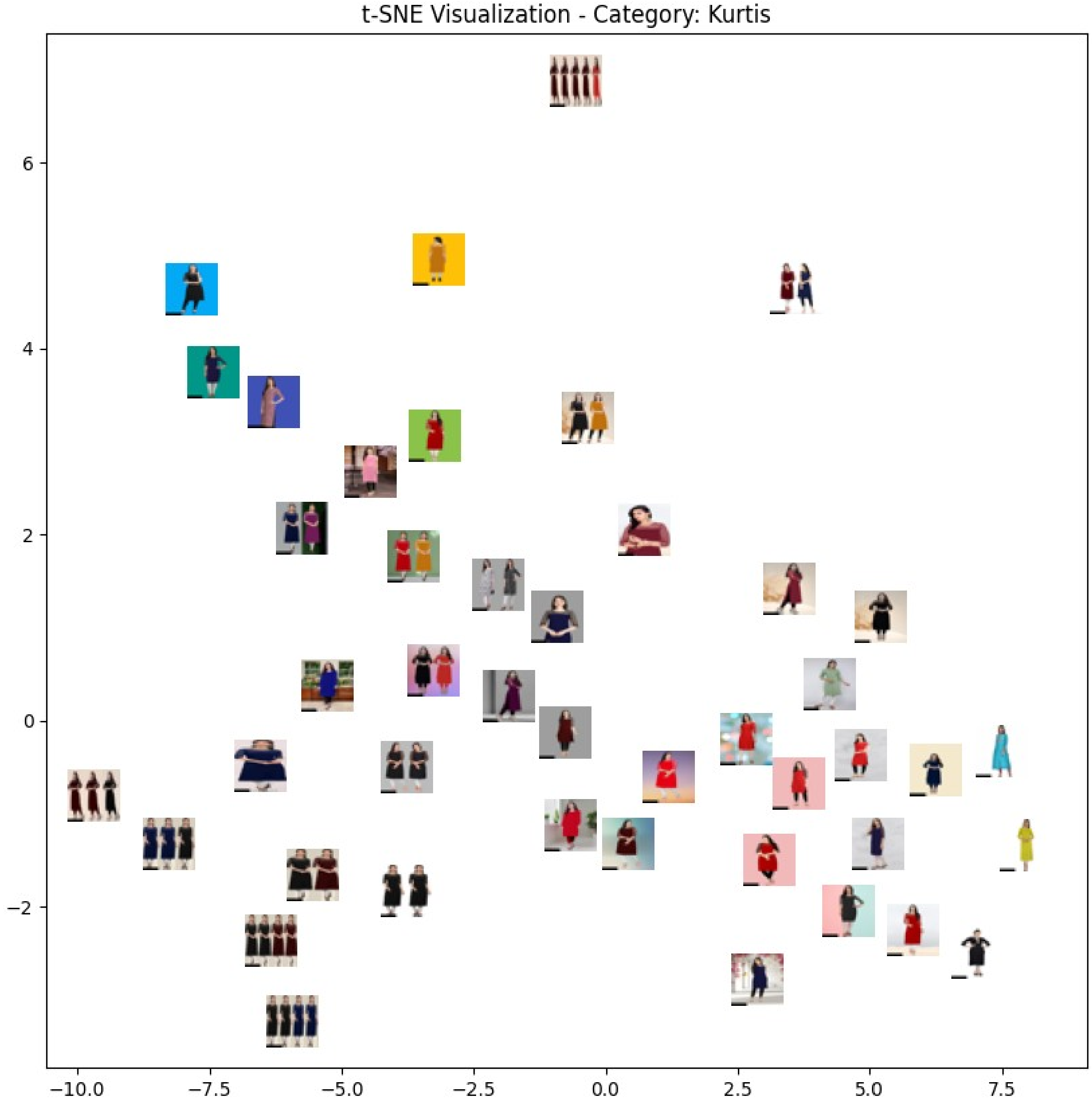


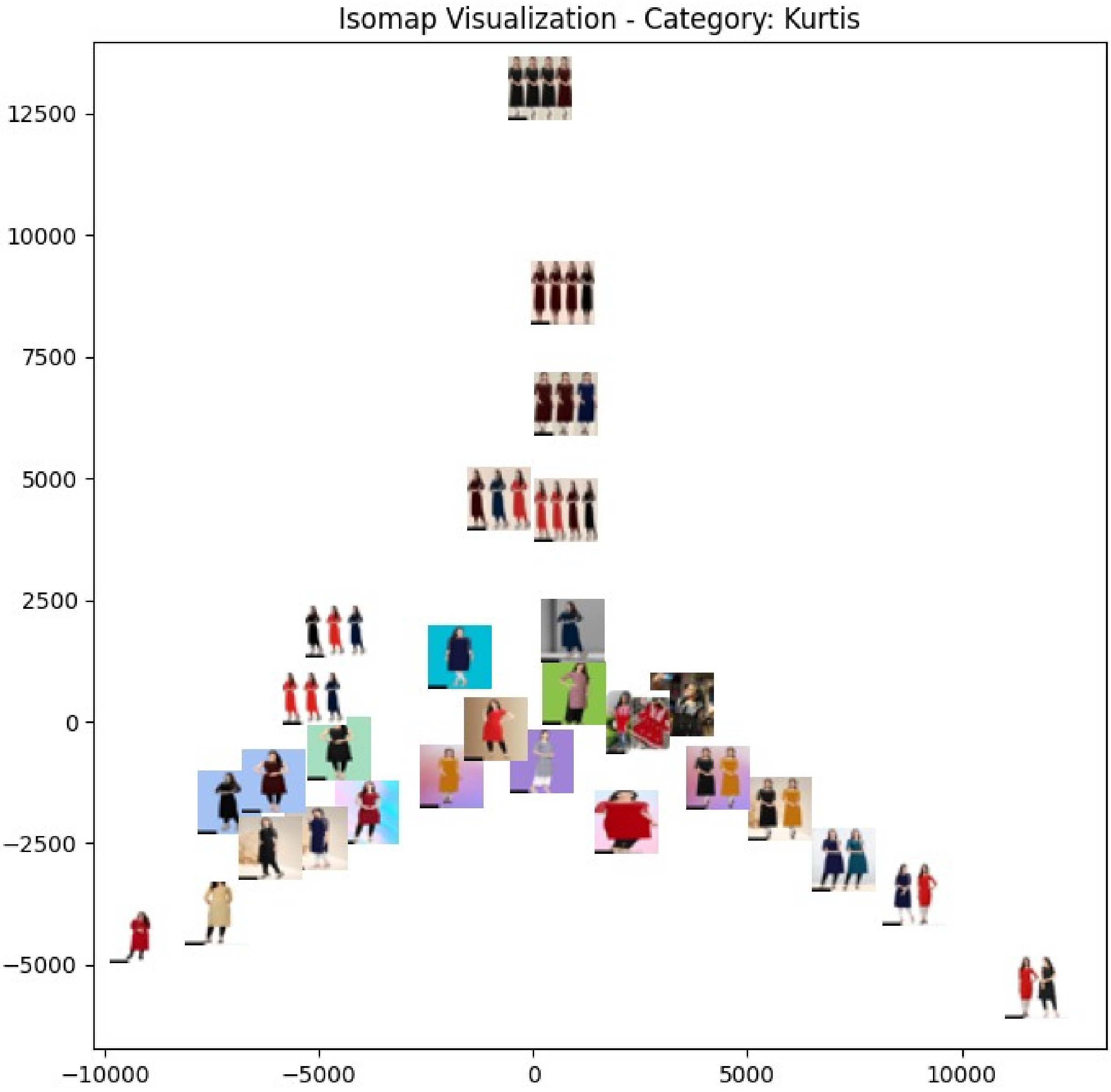
```
/opt/conda/lib/python3.10/site-packages/sklearn/manifold/
_isomap.py:373: UserWarning: The number of connected components of the
neighbors graph is 2 > 1. Completing the graph to fit Isomap might be
slow. Increase the number of neighbors to avoid this issue.
  self._fit_transform(X)
/opt/conda/lib/python3.10/site-packages/scipy/sparse/_index.py:108:
SparseEfficiencyWarning: Changing the sparsity structure of a
csr_matrix is expensive. lil and dok are more efficient.
  self._set_intXint(row, col, x.flat[0])
```

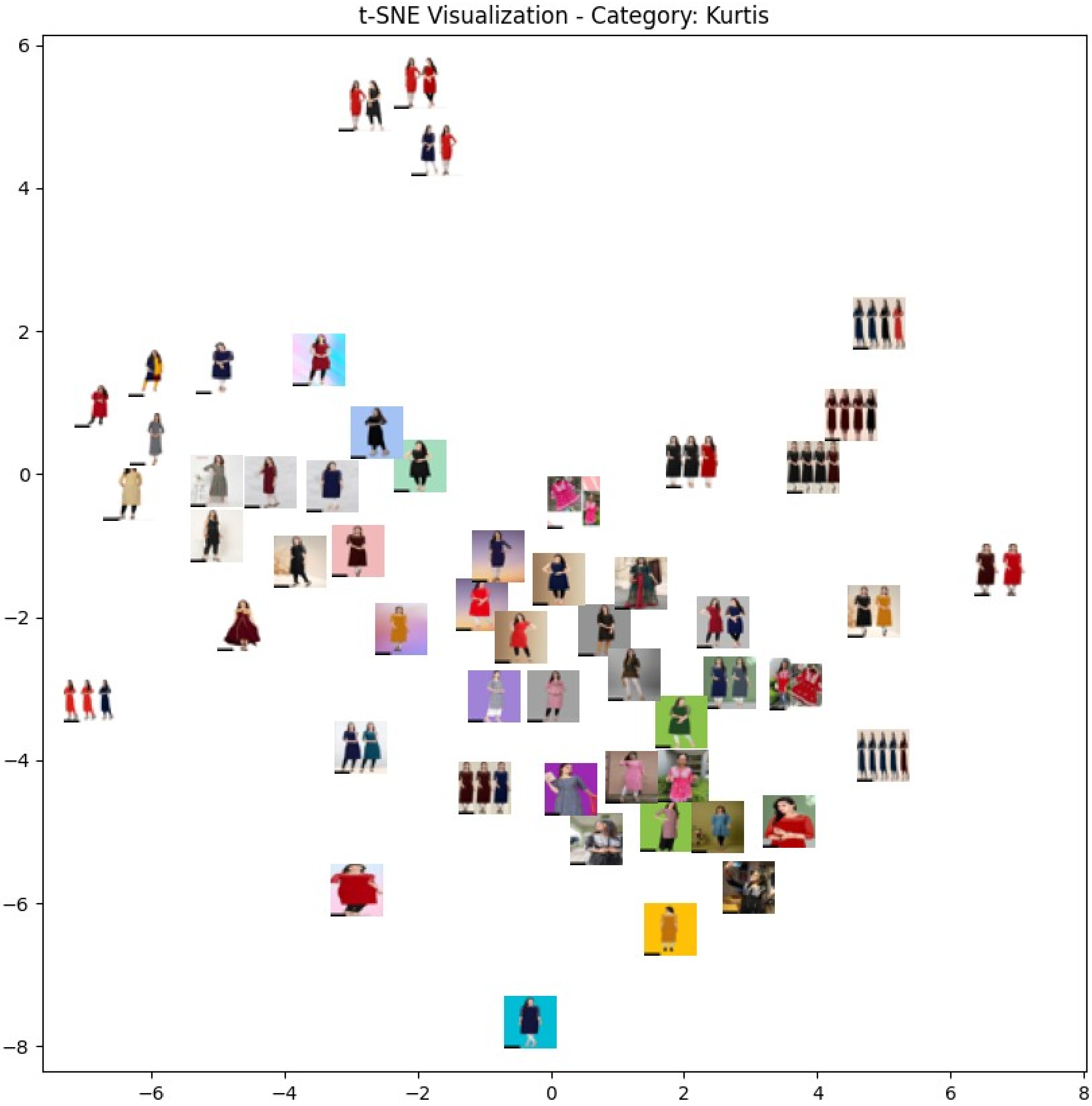

Isomap Visualization - Category: Sarees
10000
5000
0
−5000
−10000
−5000
0
5000
10000
15000

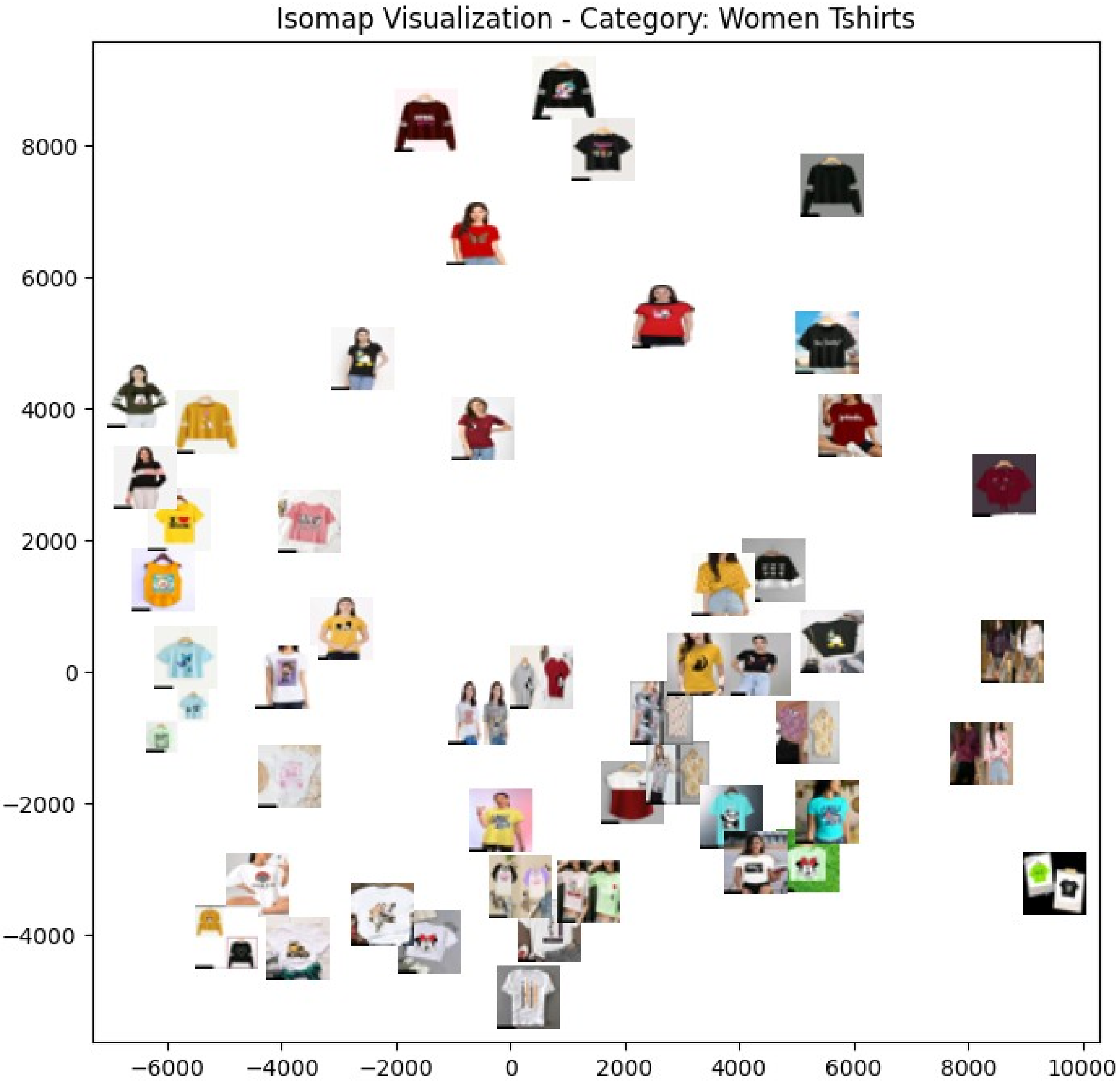

t-SNE Visualization - Category: Sarees
10
5
0
−5
−10
−10
−5
0
5
10


Isomap Visualization - Category: Kurtis
10000
7500
5000
2500
0
−2500
−5000
−7500
−10000
−5000
0
5000
10000

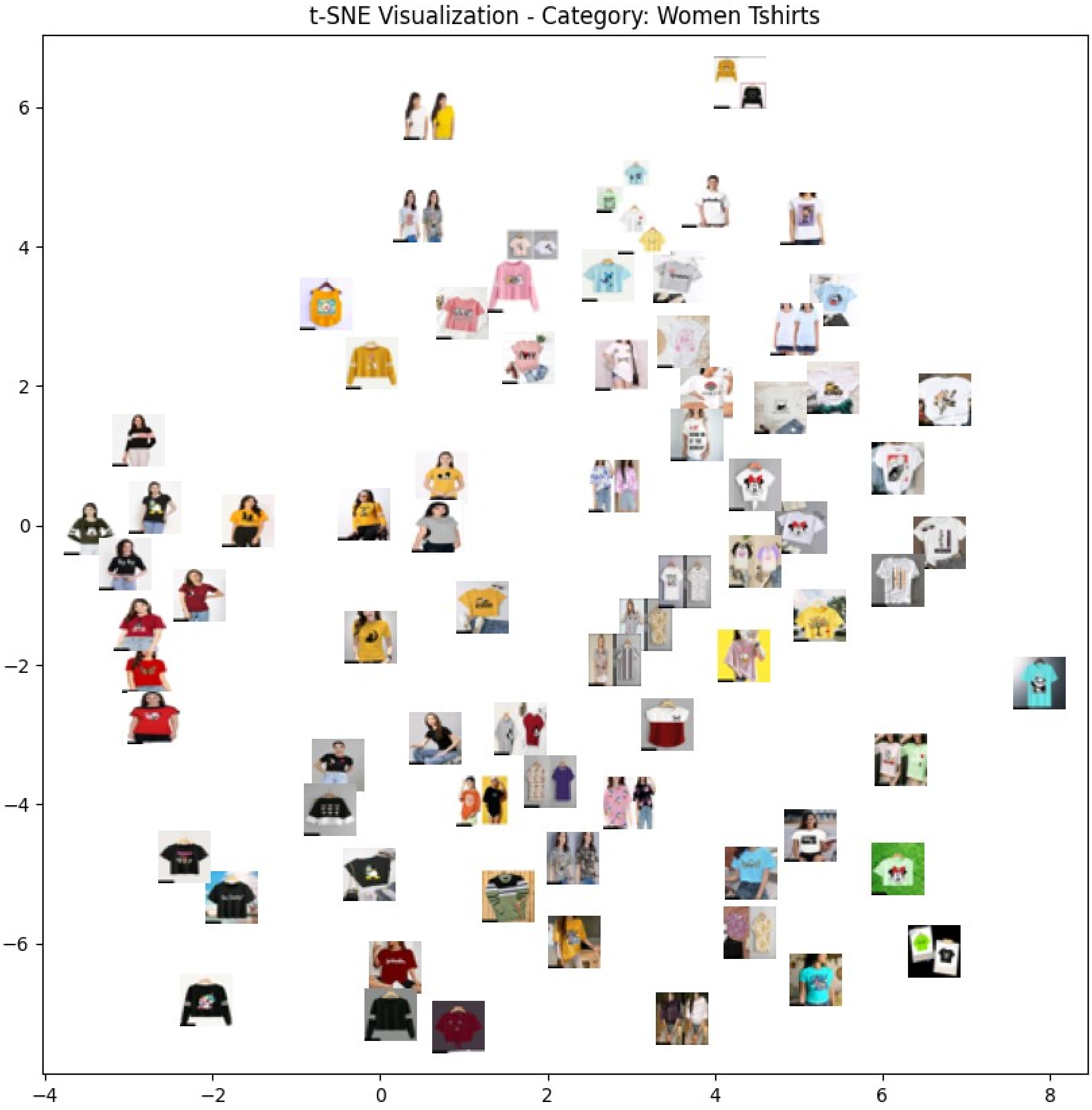

t-SNE Visualization - Category: Kurtis
6
4
2
0
−2
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

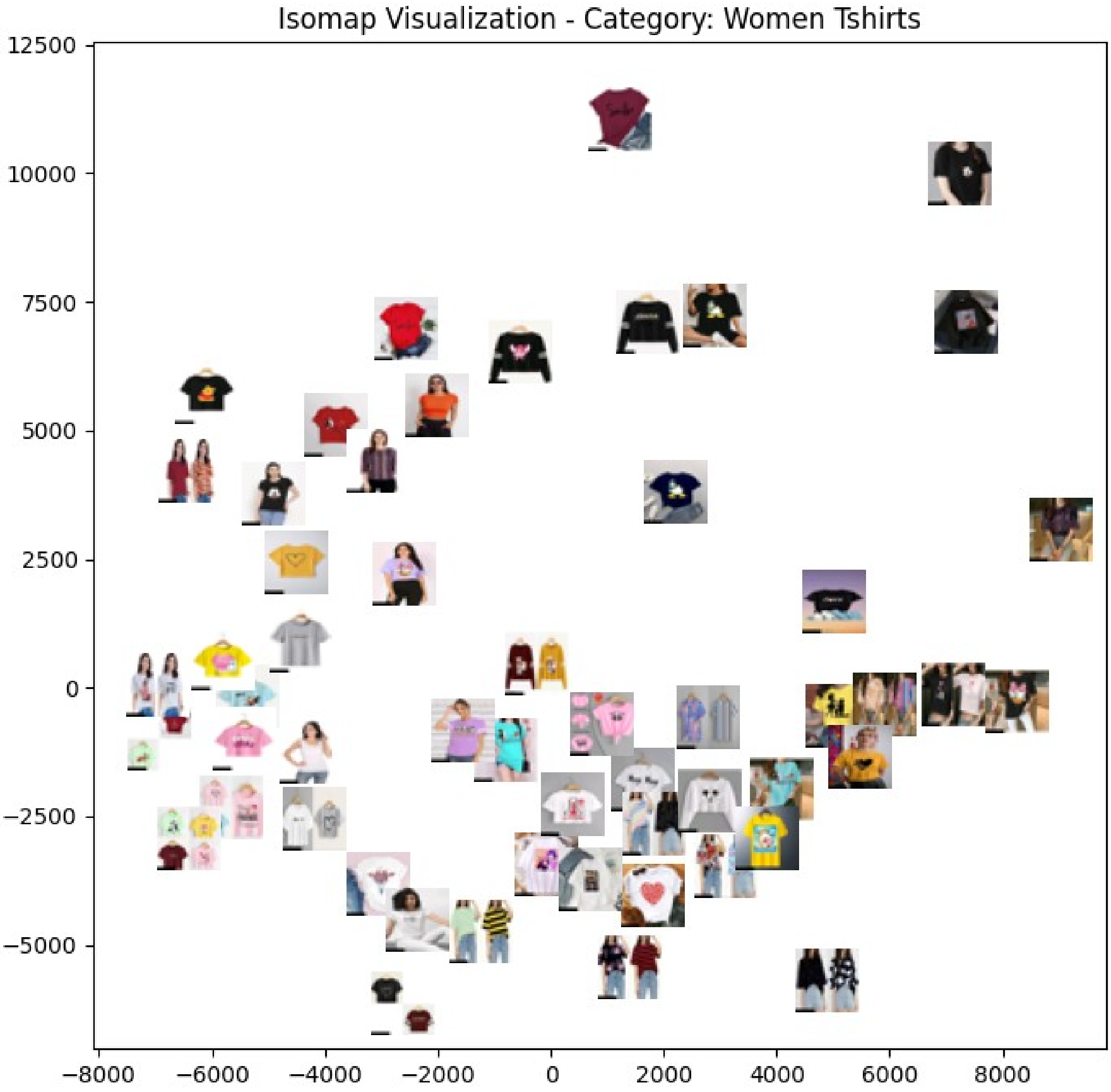

Isomap Visualization - Category: Kurtis
12500
10000
7500
5000
2500
0
−2500
−5000
−10000
−5000
0
5000
10000

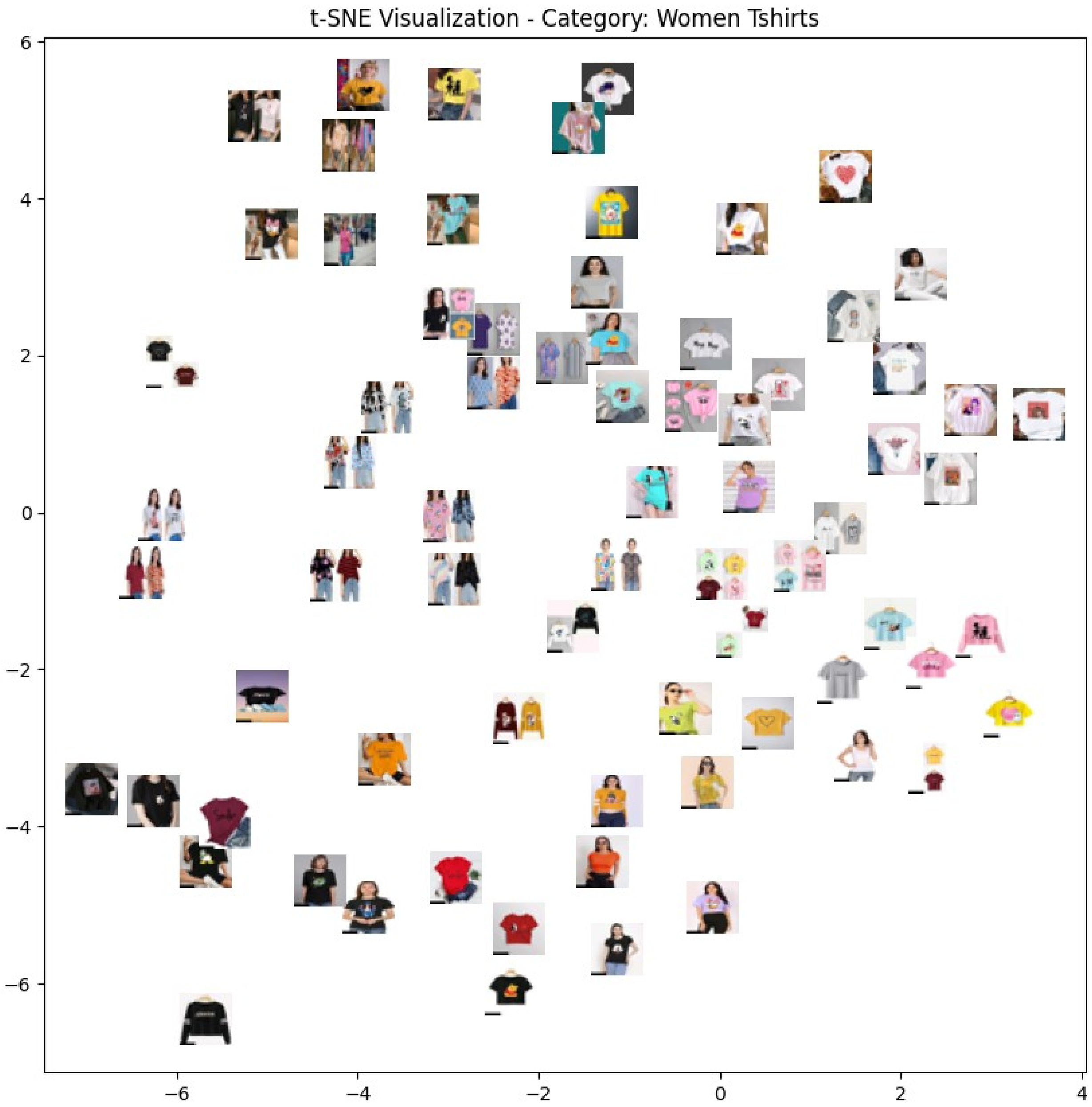

t-SNE Visualization - Category: Kurtis
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8

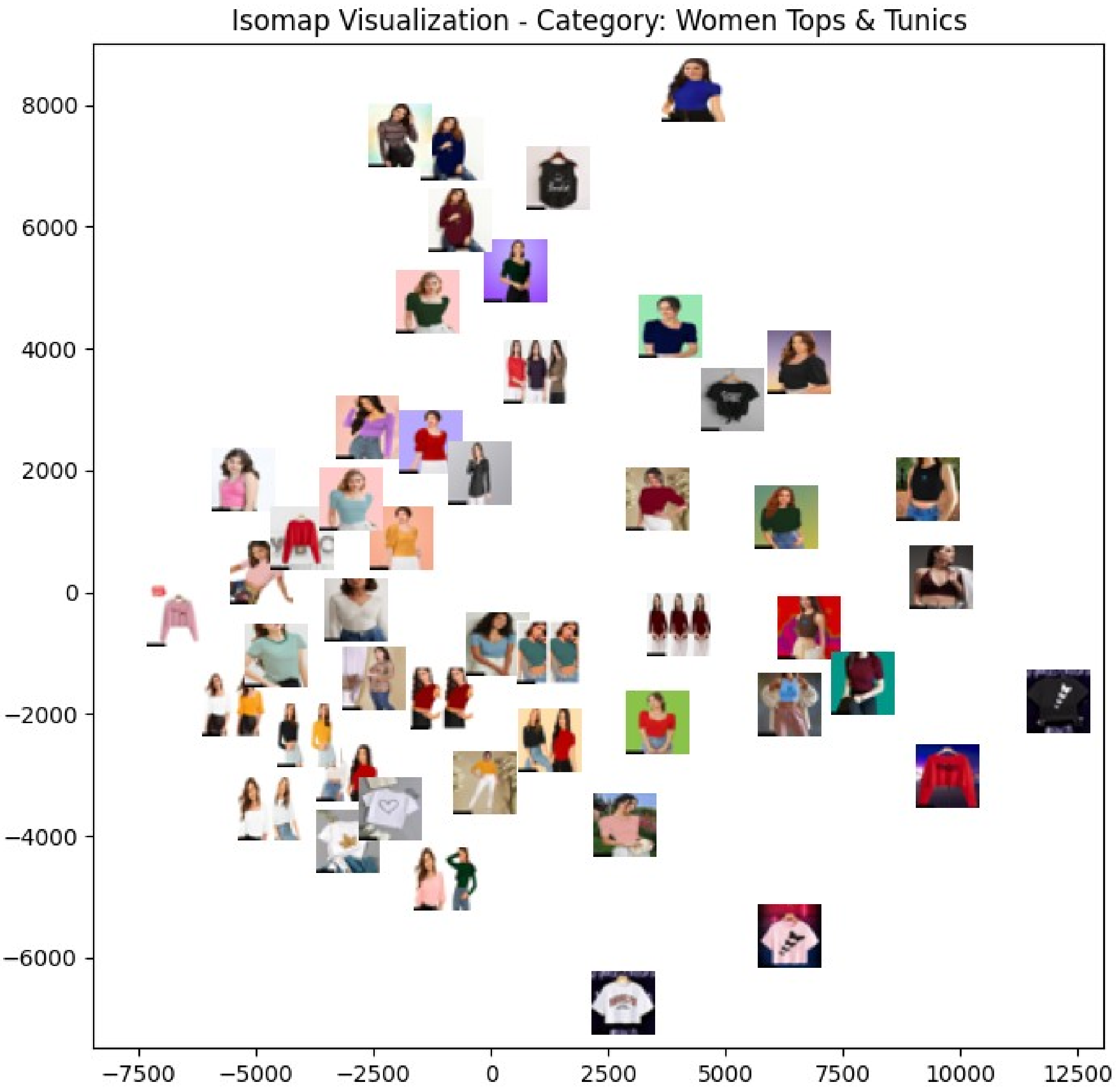


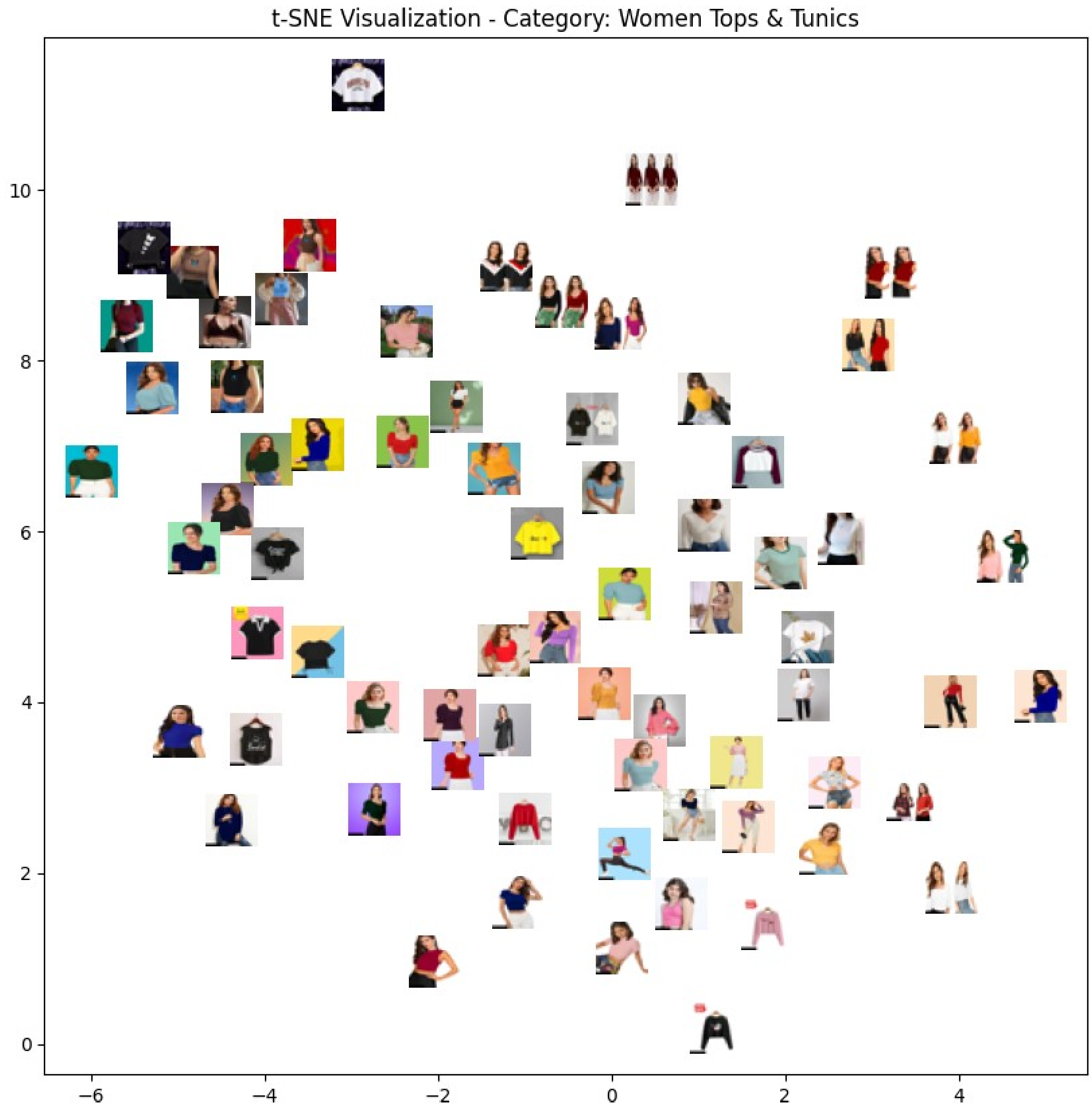
Isomap Visualization - Category: Women Tshirts
8000
6000
4000
2000
0
−2000
−4000
−6000
−4000
−2000
0
2000
4000
6000
8000
10000

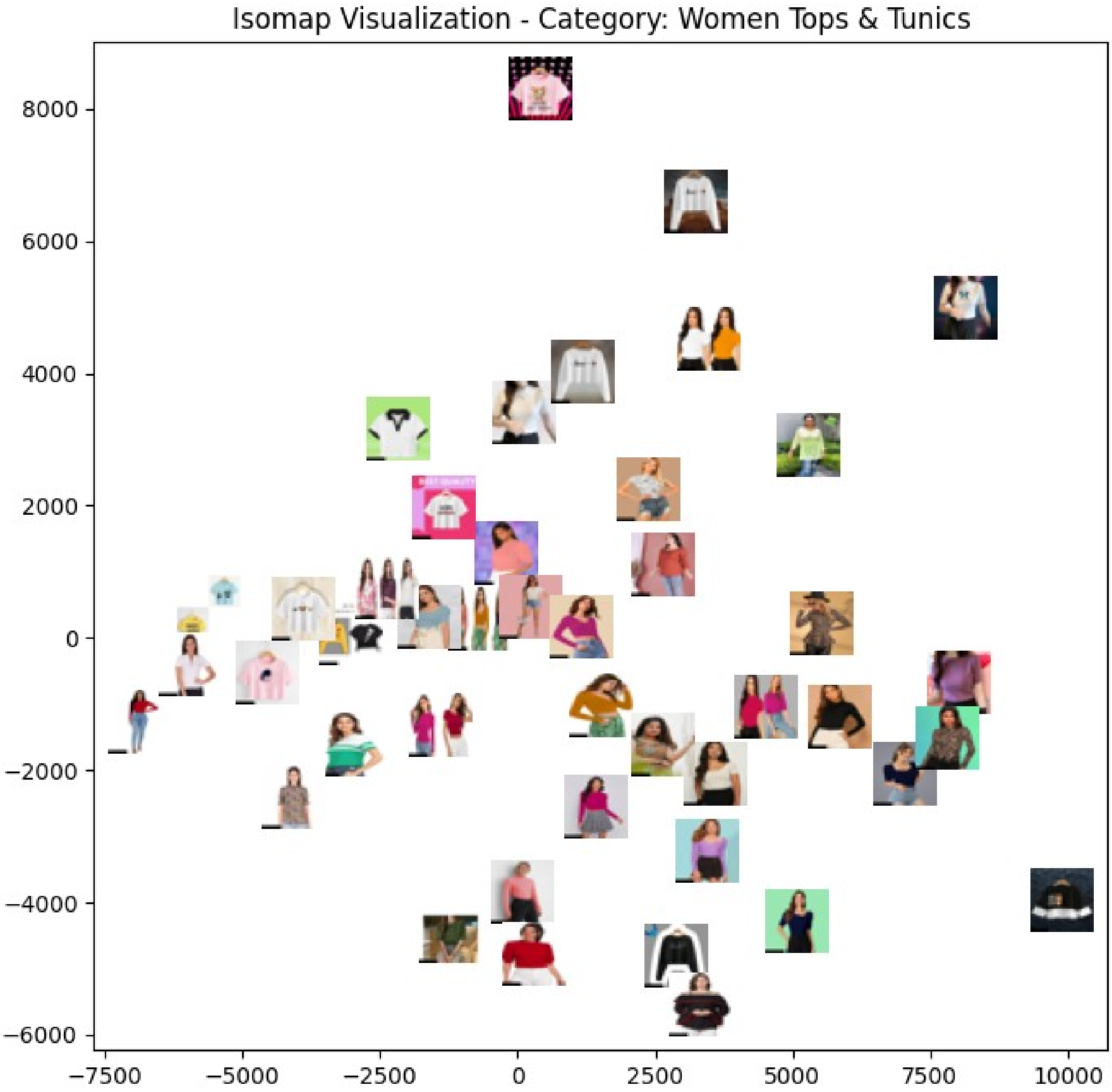

t-SNE Visualization - Category: Women Tshirts
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6
8

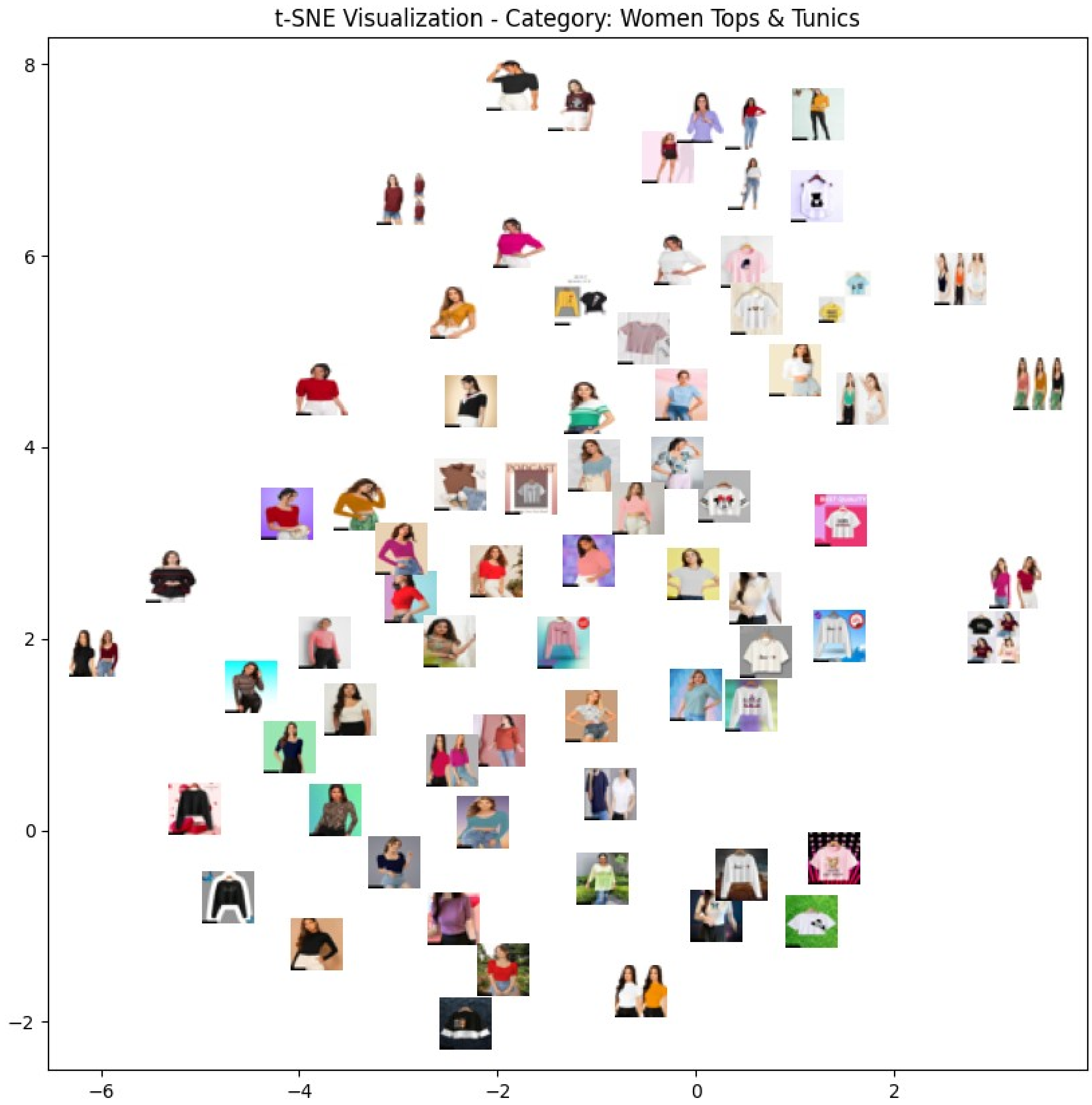

Isomap Visualization - Category: Women Tshirts
12500
10000
7500
5000
2500
0
−2500
−5000
−8000
−6000
−4000
−2000
0
2000
4000
6000
8000


t-SNE Visualization - Category: Women Tshirts
6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4


Isomap Visualization - Category: Women Tops & Tunics
8000
6000
4000
2000
0
−2000
−4000
−6000
−7500
−5000
−2500
0
2500
5000
7500
10000
12500


t-SNE Visualization - Category: Women Tops & Tunics
10
8
6
4
2
0
−6
−4
−2
0
2
4


Isomap Visualization - Category: Women Tops & Tunics
8000
6000
4000
2000
0
−2000
−4000
−6000
−7500
−5000
−2500
0
2500
5000
7500
10000



Task 3

Now comes the interesting part. Recognize the patterns and figure out a name for the components your manifold learning methods have discovered. You should also reason your choice of the name to the discovered manifold dimension.

In the Isomap and t-SNE visualizations for each category, images are plotted based on specific attributes that represent intrinsic differences in the appearance, style, and features of each product. The basis on which images are organized in the manifold space for each category:

1. **Men's T-Shirts**

- **Basis**: The clustering in Isomap and t-SNE visualizations is primarily influenced by **fit type** (slim vs. regular fit), **color** (dark vs. light tones), and **pattern** (plain vs. graphic).
- **Isomap**: Likely organizes images along a gradient where clusters reflect changes in fit and color tones, with distinct groupings for solid and patterned T-shirts.
- **t-SNE**: Creates clusters based on color and pattern, showing clear separations between plain and graphic tees, as well as light and dark colors.

2. **Sarees**

- **Basis**: Images are organized based on **occasion** (casual vs. festive), **ornamentation** (plain vs. embroidered), and **color vibrancy** (pastels vs. bold colors).
- **Isomap**: Likely captures occasion and ornamentation, creating separations for casual vs. heavily decorated sarees.
- **t-SNE**: Emphasizes color and design intricacies, clustering sarees by levels of ornamentation and color intensity.

3. **Kurtis**

- **Basis**: The visualizations are organized according to **length** (short vs. long), **sleeve style** (sleeveless, short, or full sleeves), and **pattern** (solid vs. printed).
- **Isomap**: Likely arranges images based on kurtis' lengths and sleeve types, showing clusters for shorter, sleeveless styles vs. longer, sleeved styles.
- **t-SNE**: Separates the kurtis by pattern and color details, with distinct clusters for printed vs. solid kurtis.

4. **Women's T-Shirts**

- **Basis**: Attributes such as **fit type** (loose vs. fitted), **neckline** (round vs. v-neck), and **color** are primary factors in how images are plotted.
- **Isomap**: Organizes images by fit and neckline, grouping loose fits vs. more fitted designs, with further subdivision by neckline style.
- **t-SNE**: Clusters images by color and pattern, creating separations between plain and patterned T-shirts, and variations in color tone.

5. **Women's Tops & Tunics**

- **Basis**: Length (cropped vs. tunic-length), **sleeve style** (sleeveless vs. full sleeves), and **surface styling** (plain vs. embellished) are the main organizing attributes.
- **Isomap**: Likely captures length and embellishment, grouping plain, shorter tops separately from longer, embellished tunics.
- **t-SNE**: Clusters images by sleeve style and surface detail, showing clear distinctions between sleeveless styles and more elaborate, styled options.