Weeks 7 & 8 Exercises

Name: Gyan Kannur Course: DSC540 - Data Preparation

```
import pandas as pd
import warnings
warnings.filterwarnings("ignore")

## Reading the "Hierarchy Survey Response" file into pandas dataframe
original_df = pd.read_excel('./datasets/BOING-BOING-CANDY-HIERARCHY-
2016-SURVEY-Responses.xlsx')
##Create a copy of the loaded dataframe for future transformations
candy_hier_df = original_df.copy()

## Printing the total number of rows and columns present in the
dataframe
rlength, clength = candy_hier_df.shape
print(f"Total rows {rlength}, Total columns = {clength}")
```

```
Total rows 1259, Total columns = 123
```

```
## Printing few records in the dataframe using head() method
candy_hier_df.head(3)
```

```
                 Timestamp  \
0 2016-10-24 05:09:23.033
1 2016-10-24 05:09:54.798
2 2016-10-24 05:13:06.734

  Are you going actually going trick or treating yourself? Your
gender:  \
0                                                  No
Male
1                                                  No
Male
2                                                  No
Female

  How old are you? Which country do you live in?  \
0               22                          Canada
1               45                             usa
2               48                              US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                      Ontario              JOY
1                                           il              MEH
2                                     Colorado              JOY
```

```
   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                 DESPAIR
1                                                     MEH
2                                                 DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
2                         JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [JK Rowling]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                    2

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher
```

```
1                                                3 or higher

2                                                3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                                3 or higher

1                                                3 or higher

2                                                3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                               Friday
1                                               Friday
2                                               Sunday

  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                           South to North

1                                             East to West

2                                             East to West


  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                  Science: Latest News and Headlines

1                  Science: Latest News and Headlines

2                  Science: Latest News and Headlines


   [York Peppermint Patties] Ignore
0                                NaN
1                                NaN
2                                NaN

[3 rows x 123 columns]
```

Chapter 7 Transformations

Transformation 1: Filter out missing data

```
## Filtering out the data if any of the column has missing values

candy_hier_missing_data_df = candy_hier_df.dropna()

## Printing the rows present in the missing dataframe

candy_hier_missing_data_df.shape[0]
```

```
0
```

From the above output, we see total number of row as zero record. This means that all the records present in the dataset have one or more data columns with missing or null values

Transformation 2: Fill in missing values

```
## Filling the missing values with forward fill and axis=1

candy_hier_clean_df = candy_hier_df.fillna(method='ffill',axis=1)
candy_hier_clean_df.shape
```

```
(1259, 123)
```

```
## After populating the values, checking if the dataframe has any null values

candy_hier_clean_missing_data_df = candy_hier_clean_df.dropna()

## Printing the number of rows and columns of the cleaned dataframe

candy_hier_clean_missing_data_df.shape
```

```
(1259, 123)
```

```
##Printing sample records from the dataframe using head() function

candy_hier_clean_missing_data_df.head(3)
```

```
                    Timestamp  \
0  2016-10-24 05:09:23.033000
1  2016-10-24 05:09:54.798000
2  2016-10-24 05:13:06.734000

  Are you going actually going trick or treating yourself? Your
gender:  \
0                                                 No
Male
1                                                 No
Male
2                                                 No
```

```
Female

  How old are you? Which country do you live in?  \
0               22                          Canada
1               45                             usa
2               48                              US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                        Ontario              JOY
1                                             il              MEH
2                                       Colorado              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                 DESPAIR
1                                                     MEH
2                                                 DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
2                         JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [JK Rowling]  \
0                                       3 or higher

1                                       3 or higher

2                                       3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                2

1                                       3 or higher

2                                       3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                       3 or higher

1                                       3 or higher

2                                       3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
```

```
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                         Friday
1                                         Friday
2                                         Sunday

  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                      South to North

1                                        East to West

2                                        East to West


  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                 Science: Latest News and Headlines

1                 Science: Latest News and Headlines

2                 Science: Latest News and Headlines


     [York Peppermint Patties] Ignore
0  Science: Latest News and Headlines
```

```
1  Science: Latest News and Headlines
2  Science: Latest News and Headlines

[3 rows x 123 columns]
```

Transformation 3: Removing duplicates from the dataframe/dataset

```
## Removing the duplicates from the dataframe using drop_duplicate
method available in pandas

candy_hier_dup_df = candy_hier_df.drop_duplicates()

## Printing the shape of dataframe to show number of rows and columns

candy_hier_dup_df.shape

(1259, 123)
```

From the above output it looks like the size of the new DataFrame after duplicate deletion function is same as the original one. This is performed before replacing the null values.

```
### Performing the duplicate check after populating the missing/null
values

candy_hier_dup_clean_df = candy_hier_clean_df.drop_duplicates()

## Printing the shape of dataframe to show number of rows and columns

candy_hier_dup_clean_df.shape

(1259, 123)
```

From the above output it looks like the size of the new DataFrame after duplicate deletion function is same as the original one. So, there is no duplicate present in the dataset.

Chapter 8 Transformations

Transformation 1: Create Hierarchical index

```
## Printing the default index value of the dataframe

candy_hier_df.index

RangeIndex(start=0, stop=1259, step=1)
```

From the above result, we could see the dataframe is indexed on the rows with integers ranging 0 thru 1259 with step of 1.

```
## Printing few records using head() command to show the index values

candy_hier_df.head(3)
```

```
                 Timestamp  \
0 2016-10-24 05:09:23.033
1 2016-10-24 05:09:54.798
2 2016-10-24 05:13:06.734

  Are you going actually going trick or treating yourself? Your
gender:  \
0                                                    No
Male
1                                                    No
Male
2                                                    No
Female

  How old are you? Which country do you live in?  \
0               22                          Canada
1               45                             usa
2               48                              US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                       Ontario              JOY
1                                            il              MEH
2                                      Colorado              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                 DESPAIR
1                                                     MEH
2                                                 DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
2                         JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [JK Rowling]  \
0                                       3 or higher

1                                       3 or higher

2                                       3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                  2

1                                       3 or higher

2                                       3 or higher
```

```
  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                        Friday
1                                        Friday
2                                        Sunday

  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                      South to North

1                                        East to West

2                                        East to West
```

```
  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                  Science: Latest News and Headlines

1                  Science: Latest News and Headlines

2                  Science: Latest News and Headlines


   [York Peppermint Patties] Ignore
0                              NaN
1                              NaN
2                              NaN

[3 rows x 123 columns]
```

```
##check index
candy_hier_df.index
```

```
RangeIndex(start=0, stop=1259, step=1)
```

```
## Creating hierarchical index based on Timestamp and How old are you?
columns

Hier_candy_hier_df = candy_hier_df.set_index(['Timestamp', 'How old
are you?'])

##print details of index
Hier_candy_hier_df.index[0:1]
```

```
MultiIndex([('2016-10-24 05:09:23.033000', 22)],
           names=['Timestamp', 'How old are you?'])
```

From the above result, we observe its a multi index after setting index

```
## Printing few records using head() command to show the index values

Hier_candy_hier_df.head(3)
```

```
                                          Are you going actually going
trick or treating yourself?  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
No
2016-10-24 05:09:54.798 45
No
2016-10-24 05:13:06.734 48
No
```

```
                                          Your gender:  \
Timestamp               How old are you?
2016-10-24 05:09:23.033 22                        Male
2016-10-24 05:09:54.798 45                        Male
2016-10-24 05:13:06.734 48                      Female

                                          Which country do you live in?
\
Timestamp               How old are you?

2016-10-24 05:09:23.033 22                                      Canada

2016-10-24 05:09:54.798 45                                         usa

2016-10-24 05:13:06.734 48                                          US


                                          Which state, province, county
do you live in?  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
Ontario
2016-10-24 05:09:54.798 45
il
2016-10-24 05:13:06.734 48
Colorado

                                           [100 Grand Bar]  \
Timestamp               How old are you?
2016-10-24 05:09:23.033 22                            JOY
2016-10-24 05:09:54.798 45                            MEH
2016-10-24 05:13:06.734 48                            JOY

                                           [Anonymous brown globs that
come in black and orange wrappers]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
DESPAIR
2016-10-24 05:09:54.798 45
MEH
2016-10-24 05:13:06.734 48
DESPAIR

                                           [Any full-sized candy
bar]  \
Timestamp               How old are you?
```

```
2016-10-24 05:09:23.033 22                                                JOY

2016-10-24 05:09:54.798 45                                                JOY

2016-10-24 05:13:06.734 48                                                JOY


                                              [Black Jacks]  \
Timestamp               How old are you?
2016-10-24 05:09:23.033 22                             MEH
2016-10-24 05:09:54.798 45                             JOY
2016-10-24 05:13:06.734 48                             MEH

                                              [Bonkers (the candy)]  \
Timestamp               How old are you?
2016-10-24 05:09:23.033 22                                 MEH
2016-10-24 05:09:54.798 45                               DESPAIR
2016-10-24 05:13:06.734 48                                 MEH

                                              [Bonkers (the board
game)]  ...  \
Timestamp               How old are
you?                                 ...
2016-10-24 05:09:23.033 22
MEH  ...
2016-10-24 05:09:54.798 45
MEH  ...
2016-10-24 05:13:06.734 48
JOY  ...

                                            Please estimate the degree(s)
of separation you have from the following celebrities [JK Rowling]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
3 or higher
2016-10-24 05:09:54.798 45
3 or higher
2016-10-24 05:13:06.734 48
3 or higher

                                            Please estimate the degree(s)
of separation you have from the following celebrities [JJ Abrams]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
2
2016-10-24 05:09:54.798 45
3 or higher
2016-10-24 05:13:06.734 48
```

```
3 or higher

                                          Please estimate the degree(s)
of separation you have from the following celebrities [Beyoncé]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
3 or higher
2016-10-24 05:09:54.798 45
3 or higher
2016-10-24 05:13:06.734 48
3 or higher

                                          Please estimate the degree(s)
of separation you have from the following celebrities [Bieber]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
3 or higher
2016-10-24 05:09:54.798 45
3 or higher
2016-10-24 05:13:06.734 48
3 or higher

                                          Please estimate the degree(s)
of separation you have from the following celebrities [Kevin Bacon]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
3 or higher
2016-10-24 05:09:54.798 45
3 or higher
2016-10-24 05:13:06.734 48
3 or higher

                                          Please estimate the degree(s)
of separation you have from the following celebrities [Francis Bacon
(1561 - 1626)]  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
3 or higher

2016-10-24 05:09:54.798 45
3 or higher

2016-10-24 05:13:06.734 48
3 or higher
```

```
                                          Which day do you prefer,
Friday or Sunday?  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
Friday
2016-10-24 05:09:54.798 45
Friday
2016-10-24 05:13:06.734 48
Sunday

                                          Do you eat apples the correct
way, East to West (side to side) or do you eat them like a freak of
nature, South to North (bottom to top)?  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
South to North

2016-10-24 05:09:54.798 45
East to West

2016-10-24 05:13:06.734 48
East to West


                                          When you see the above image
of the 4 different websites, which one would you most likely check out
(please be honest).  \
Timestamp               How old are you?

2016-10-24 05:09:23.033 22                                  Science:
Latest News and Headlines

2016-10-24 05:09:54.798 45                                  Science:
Latest News and Headlines

2016-10-24 05:13:06.734 48                                  Science:
Latest News and Headlines


                                           [York Peppermint Patties]
Ignore
Timestamp               How old are you?

2016-10-24 05:09:23.033 22
NaN
2016-10-24 05:09:54.798 45
NaN
2016-10-24 05:13:06.734 48
```

```
NaN

[3 rows x 121 columns]
```

From the above result, we could see Timestamp and how old are you? fields have been removed from the actual dataset and added as index

```
## Restoring the original index columns

Hier_candy_hier_df.reset_index()
```

```
                   Timestamp How old are you?  \
0    2016-10-24 05:09:23.033               22
1    2016-10-24 05:09:54.798               45
2    2016-10-24 05:13:06.734               48
3    2016-10-24 05:14:17.192               57
4    2016-10-24 05:14:24.625               42
...                      ...              ...
1254 2016-10-29 16:53:52.516               52
1255 2016-10-30 06:53:54.735               33
1256 2016-10-30 11:06:10.827              NaN
1257 2016-10-30 16:07:26.539               48
1258 2016-10-30 17:06:45.660               44

     Are you going actually going trick or treating yourself? Your
gender:  \
0                                                    No
Male
1                                                    No
Male
2                                                    No
Female
3                                                    No
Male
4                                                   Yes
Male
...                                                 ...
...
1254                                                 No
Female
1255                                                 No
Male
1256                                                 No
Male
1257                                                 No
Male
1258                                                Yes
Female

     Which country do you live in?  \
```

```
0                            Canada
1                               usa
2                                US
3                               usa
4                               USA
...                             ...
1254                            USA
1255                  united states
1256                            NaN
1257                         canada
1258                             Us
```

```
     Which state, province, county do you live in?  [100 Grand Bar]  \
0                                          Ontario              JOY
1                                               il              MEH
2                                         Colorado              JOY
3                                               il              JOY
4                                     South Dakota              MEH
...                                            ...              ...
1254                                            TX              JOY
1255                                     minnesota              JOY
1256                                           NaN              JOY
1257                                            BC              NaN
1258                                            Nh              JOY

      [Anonymous brown globs that come in black and orange
wrappers]  \
0                                                DESPAIR

1                                                    MEH

2                                                DESPAIR

3                                                    MEH

4                                                DESPAIR

...                                                  ...

1254                                             DESPAIR

1255                                             DESPAIR

1256                                                 MEH

1257                                             DESPAIR

1258                                                 MEH


      [Any full-sized candy bar]  [Black Jacks]  ...  \
```

```
0                              JOY            MEH  ...
1                              JOY            JOY  ...
2                              JOY            MEH  ...
3                              JOY            MEH  ...
4                              JOY        DESPAIR  ...
...                            ...            ...  ...
1254                           JOY            MEH  ...
1255                           JOY        DESPAIR  ...
1256                           JOY            NaN  ...
1257                           JOY        DESPAIR  ...
1258                           JOY            JOY  ...

     Please estimate the degree(s) of separation you have from the
following celebrities [JK Rowling]  \
0                                                3 or higher

1                                                3 or higher

2                                                3 or higher

3                                                3 or higher

4                                                3 or higher

...                                                        ...

1254                                             3 or higher

1255                                 Actually, that's me.

1256                                                   NaN

1257                                                     1

1258                                             3 or higher


     Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                        2

1                                                3 or higher

2                                                3 or higher

3                                                3 or higher

4                                                3 or higher

...                                                        ...
```

```
1254                                                3 or higher

1255                                                3 or higher

1256                                                        NaN

1257                                                          2

1258                                                3 or higher


     Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                                   3 or higher

1                                                   3 or higher

2                                                   3 or higher

3                                                   3 or higher

4                                                   3 or higher

...                                                         ...

1254                                                3 or higher

1255                                                3 or higher

1256                                                        NaN

1257                                                3 or higher

1258                                                3 or higher


     Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                                   3 or higher

1                                                   3 or higher

2                                                   3 or higher

3                                                   3 or higher

4                                                   3 or higher

...                                                         ...

1254                                                3 or higher
```

```
1255                                                3 or higher

1256                                                        NaN

1257                                                3 or higher

1258                                                3 or higher


     Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                                   3 or higher

1                                                   3 or higher

2                                                   3 or higher

3                                                   3 or higher

4                                                   3 or higher

...                                                         ...

1254                                                          2

1255                                                3 or higher

1256                                                        NaN

1257                                                          2

1258                                                3 or higher


     Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                                   3 or higher

1                                                   3 or higher

2                                                   3 or higher

3                                                   3 or higher

4                                                   3 or higher

...                                                         ...

1254                                                3 or higher

1255                                  Actually, that's me.
```

```
1256                                                    NaN

1257                                          3 or higher

1258                                          3 or higher


     Which day do you prefer, Friday or Sunday?  \
0                                           Friday
1                                           Friday
2                                           Sunday
3                                           Sunday
4                                           Friday
...                                            ...
1254                                        Friday
1255                                        Friday
1256                                        Sunday
1257                                        Sunday
1258                                        Sunday

     Do you eat apples the correct way, East to West (side to side) or
do you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                           South to North

1                                             East to West

2                                             East to West

3                                           South to North

4                                             East to West

...                                                    ...

1254                                          East to West

1255                       Sinusoidally around the equator

1256                          nne to east to nnw to s to n

1257                                          East to West

1258                                          East to West


     When you see the above image of the 4 different websites, which
one would you most likely check out (please be honest).  \
0                       Science: Latest News and Headlines

1                       Science: Latest News and Headlines
```

```
2                     Science: Latest News and Headlines

3                     Science: Latest News and Headlines

4                                                   ESPN

...                                                  ...

1254                  Science: Latest News and Headlines

1255                  Science: Latest News and Headlines

1256                  Science: Latest News and Headlines

1257                  Science: Latest News and Headlines

1258                                           Daily Dish


      [York Peppermint Patties] Ignore
0                                    NaN
1                                    NaN
2                                    NaN
3                                    NaN
4                                    NaN
...                                  ...
1254                                 NaN
1255                                 NaN
1256                                 NaN
1257                                 NaN
1258                                 NaN

[1259 rows x 123 columns]
```

The output shows that the index for the dataframe has been restored as original (1259 rows X 123 columns)

Transformation 2: Pivot the data

```
## Before applying pivot, we have consider alter the text values to
integer.
## Lets apply 1 for DESPAIR, 2 for JOY and 3 for MEH

print("The columns present in the dataframe:")
print(candy_hier_df.columns)
candy_subset_df = candy_hier_df[['How old are you?','Your gender:','
[100 Grand Bar]',
                                 ' [Anonymous brown globs that come in
black and orange wrappers]',
```

```
                                ' [Any full-sized candy bar]',
                                ' [Black Jacks]']

The columns present in the dataframe:
Index(['Timestamp', 'Are you going actually going trick or treating
yourself?',
       'Your gender:', 'How old are you?', 'Which country do you live
in?',
       'Which state, province, county do you live in?', ' [100 Grand
Bar]',
       ' [Anonymous brown globs that come in black and orange
wrappers]',
       ' [Any full-sized candy bar]', ' [Black Jacks]',
       ...
       'Please estimate the degree(s) of separation you have from the
following celebrities [JK Rowling]',
       'Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]',
       'Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]',
       'Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]',
       'Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]',
       'Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]',
       'Which day do you prefer, Friday or Sunday?',
       'Do you eat apples the correct way, East to West (side to side)
or do you eat them like a freak of nature, South to North (bottom to
top)?',
       'When you see the above image of the 4 different websites,
which one would you most likely check out (please be honest).',
       ' [York Peppermint Patties] Ignore'],
      dtype='object', length=123)

## Printing few values using head command

candy_subset_df.head(3)

  How old are you? Your gender:  [100 Grand Bar]  \
0               22         Male              JOY
1               45         Male              MEH
2               48       Female              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                  DESPAIR
1                                                      MEH
2                                                  DESPAIR

   [Any full-sized candy bar]  [Black Jacks]
```

```
0                          JOY            MEH
1                          JOY            JOY
2                          JOY            MEH
```

```
## Printing the shape of the dataframe

print(f'The shape of the dataframe before filtering null values:
{candy_subset_df.shape}')
```

```
The shape of the dataframe before filtering null values: (1259, 6)
```

```
## Renaming the columns

candy_subset_df = candy_subset_df.rename(columns={'How old are
you?':'Age','Your gender:':'Gender',' [100 Grand Bar]': 'Grand_Bar', '
[Anonymous brown globs that come in black and orange wrappers]':
'Black_Orange_Wrappers',' [Any full-sized candy bar]':'Candy_Bar','
[Black Jacks]':'Black_Jacks'})
candy_subset_df.columns
```

```
Index(['Age', 'Gender', 'Grand_Bar', 'Black_Orange_Wrappers',
'Candy_Bar',
       'Black_Jacks'],
      dtype='object')
```

```
## Applying pivot function to the above dataset

table = pd.pivot_table(data=candy_subset_df,index=['Gender'],
aggfunc='count')
table
```

```
                     Age  Black_Jacks  Black_Orange_Wrappers  Candy_Bar
\
Gender

Female               401          334                    405        406

I'd rather not say    24           23                     26         27

Male                 779          678                    777        785

Other                 17           14                     17         16


                     Grand_Bar
Gender
Female                     381
I'd rather not say          26
Male                       751
Other                       16
```

```
## We will use 'Timestamp' as our key value , since this operation
requires column with Unique values / No Duplicates.
## pd.melt() method along with the key as gender displays each column
name under variable column and corresponding
## values from each of the rows are displayed in 'value' column

melted = pd.melt(candy_hier_df, ['Timestamp'])

## Display the Pivoted Data, along the Timestamp column

melted
```

```
                     Timestamp  \
0      2016-10-24 05:09:23.033
1      2016-10-24 05:09:54.798
2      2016-10-24 05:13:06.734
3      2016-10-24 05:14:17.192
4      2016-10-24 05:14:24.625
...                        ...
153593 2016-10-29 16:53:52.516
153594 2016-10-30 06:53:54.735
153595 2016-10-30 11:06:10.827
153596 2016-10-30 16:07:26.539
153597 2016-10-30 17:06:45.660

                                                 variable value
0       Are you going actually going trick or treating...    No
1       Are you going actually going trick or treating...    No
2       Are you going actually going trick or treating...    No
3       Are you going actually going trick or treating...    No
4       Are you going actually going trick or treating...   Yes
...                                                   ...   ...
153593                  [York Peppermint Patties] Ignore   NaN
153594                  [York Peppermint Patties] Ignore   NaN
153595                  [York Peppermint Patties] Ignore   NaN
153596                  [York Peppermint Patties] Ignore   NaN
153597                  [York Peppermint Patties] Ignore   NaN

[153598 rows x 3 columns]
```

Chapter 10 Transformations

Transformation 1: Grouping with Dicts/Series

```
## Copying the content of hierarchy dataframe into another dataframe
for processing

candy_hier_trans_df = candy_hier_df
```

```
## Create a new column "Date" based on timestamp

candy_hier_trans_df['Date'] = candy_hier_trans_df['Timestamp'].dt.date

## Displaying few records after adding date column

candy_hier_trans_df.head(3)
```

```
                 Timestamp  \
0 2016-10-24 05:09:23.033
1 2016-10-24 05:09:54.798
2 2016-10-24 05:13:06.734

  Are you going actually going trick or treating yourself? Your
gender:  \
0                                                   No
Male
1                                                   No
Male
2                                                   No
Female

  How old are you? Which country do you live in?  \
0               22                          Canada
1               45                             usa
2               48                              US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                      Ontario              JOY
1                                           il              MEH
2                                     Colorado              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                             DESPAIR
1                                                 MEH
2                                             DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
2                         JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                 2

1                                       3 or higher

2                                       3 or higher
```

```
  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                        Friday
1                                        Friday
2                                        Sunday

  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                      South to North

1                                        East to West

2                                        East to West
```

```
  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                 Science: Latest News and Headlines

1                 Science: Latest News and Headlines

2                 Science: Latest News and Headlines


   [York Peppermint Patties] Ignore        Date
0                            NaN  2016-10-24
1                            NaN  2016-10-24
2                            NaN  2016-10-24

[3 rows x 124 columns]
```

```
## Renaming the column "Your Gender:" to "Gender"
## Displaying the few sample records after renaming the column

candy_hier_trans_df = candy_hier_trans_df.rename({'Your gender:':
'Gender'}, axis=1)
candy_hier_trans_df.head(3)
```

```
                Timestamp  \
0 2016-10-24 05:09:23.033
1 2016-10-24 05:09:54.798
2 2016-10-24 05:13:06.734

  Are you going actually going trick or treating yourself?  Gender  \
0                                                 No          Male
1                                                 No          Male
2                                                 No        Female

  How old are you? Which country do you live in?  \
0               22                         Canada
1               45                            usa
2               48                             US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                       Ontario              JOY
1                                            il              MEH
2                                      Colorado              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                             DESPAIR
1                                                 MEH
2                                             DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
```

```
2                          JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [JJ Abrams]  \
0                                                    2

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                       Friday
1                                       Friday
2                                       Sunday
```

```
  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                      South to North

1                                        East to West

2                                        East to West


  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                 Science: Latest News and Headlines

1                 Science: Latest News and Headlines

2                 Science: Latest News and Headlines


   [York Peppermint Patties] Ignore        Date
0                                NaN  2016-10-24
1                                NaN  2016-10-24
2                                NaN  2016-10-24

[3 rows x 124 columns]
```

```
## Taking count of records group by based on Date column that we
created

candy_hier_trans_df.groupby('Date').count()
```

```
            Timestamp  \
Date
2016-10-24        781
2016-10-25        150
2016-10-26         35
2016-10-27        255
2016-10-28         27
2016-10-29          7
2016-10-30          4

            Are you going actually going trick or treating yourself?
Gender  \
Date

2016-10-24                                                      781
775
2016-10-25                                                      150
150
2016-10-26                                                       35
```

```
34
2016-10-27                                                    255
253
2016-10-28                                                     27
27
2016-10-29                                                      7
7
2016-10-30                                                      4
4

            How old are you?  Which country do you live in?  \
Date
2016-10-24               760                               770
2016-10-25               147                               148
2016-10-26                35                                35
2016-10-27               248                               249
2016-10-28                26                                26
2016-10-29                 7                                 7
2016-10-30                 3                                 3

            Which state, province, county do you live in?   [100 Grand
Bar]  \
Date

2016-10-24                                                 734
734
2016-10-25                                                 139
127
2016-10-26                                                  34
35
2016-10-27                                                 237
249
2016-10-28                                                  26
26
2016-10-29                                                   7
7
2016-10-30                                                   3
3

             [Anonymous brown globs that come in black and orange
wrappers]  \
Date

2016-10-24                                                    763

2016-10-25                                                    146

2016-10-26                                                     35

2016-10-27                                                    251
```

```
2016-10-28                                                    27

2016-10-29                                                     7

2016-10-30                                                     4


             [Any full-sized candy bar]   [Black Jacks]  ...  \
Date                                                      ...
2016-10-24                          770             658  ...
2016-10-25                          148             114  ...
2016-10-26                           35              27  ...
2016-10-27                          253             224  ...
2016-10-28                           27              23  ...
2016-10-29                            5               7  ...
2016-10-30                            4               3  ...

            Please estimate the degree(s) of separation you have from
the following celebrities [JK Rowling]  \
Date

2016-10-24                                                   727

2016-10-25                                                   141

2016-10-26                                                    34

2016-10-27                                                   240

2016-10-28                                                    26

2016-10-29                                                     6

2016-10-30                                                     3


            Please estimate the degree(s) of separation you have from
the following celebrities [JJ Abrams]  \
Date

2016-10-24                                                   728

2016-10-25                                                   140

2016-10-26                                                    33

2016-10-27                                                   240

2016-10-28                                                    26
```

```
2016-10-29                                                          6

2016-10-30                                                          3


            Please estimate the degree(s) of separation you have from
the following celebrities [Beyoncé]  \
Date

2016-10-24                                                        727

2016-10-25                                                        141

2016-10-26                                                         34

2016-10-27                                                        238

2016-10-28                                                         26

2016-10-29                                                          6

2016-10-30                                                          3


            Please estimate the degree(s) of separation you have from
the following celebrities [Bieber]  \
Date

2016-10-24                                                        726

2016-10-25                                                        141

2016-10-26                                                         34

2016-10-27                                                        239

2016-10-28                                                         26

2016-10-29                                                          6

2016-10-30                                                          3


            Please estimate the degree(s) of separation you have from
the following celebrities [Kevin Bacon]  \
Date

2016-10-24                                                        730

2016-10-25                                                        141
```

```
2016-10-26                                                    34

2016-10-27                                                   239

2016-10-28                                                    26

2016-10-29                                                     6

2016-10-30                                                     3


            Please estimate the degree(s) of separation you have from
the following celebrities [Francis Bacon (1561 - 1626)]  \
Date

2016-10-24                                                   724

2016-10-25                                                   141

2016-10-26                                                    34

2016-10-27                                                   240

2016-10-28                                                    26

2016-10-29                                                     6

2016-10-30                                                     3


            Which day do you prefer, Friday or Sunday?  \
Date
2016-10-24                                                 766
2016-10-25                                                 148
2016-10-26                                                  35
2016-10-27                                                 249
2016-10-28                                                  27
2016-10-29                                                   6
2016-10-30                                                   4

            Do you eat apples the correct way, East to West (side to
side) or do you eat them like a freak of nature, South to North
(bottom to top)?  \
Date

2016-10-24                                                   758

2016-10-25                                                   145

2016-10-26                                                    33
```

```
2016-10-27                                                          241

2016-10-28                                                           27

2016-10-29                                                            6

2016-10-30                                                            4


            When you see the above image of the 4 different websites,
which one would you most likely check out (please be honest).  \
Date

2016-10-24                                                          761

2016-10-25                                                          145

2016-10-26                                                           35

2016-10-27                                                          244

2016-10-28                                                           27

2016-10-29                                                            6

2016-10-30                                                            4


              [York Peppermint Patties] Ignore
Date
2016-10-24                                        0
2016-10-25                                        0
2016-10-26                                        0
2016-10-27                                        0
2016-10-28                                        0
2016-10-29                                        0
2016-10-30                                        0

[7 rows x 123 columns]
```

```
## Taking count of records group by Date and gender columns (2
columns) that are present in the dataset

candy_hier_trans_df.groupby(['Date','Gender']).count()
```

```
                                Timestamp  \
Date       Gender
2016-10-24 Female                     231
           I'd rather not say          15
           Male                       520
           Other                        9
```

```
2016-10-25 Female                      85
           I'd rather not say           4
           Male                        60
           Other                        1
2016-10-26 Female                      13
           Male                        19
           Other                        2
2016-10-27 Female                      61
           I'd rather not say           9
           Male                       179
           Other                        4
2016-10-28 Female                      15
           Male                        11
           Other                        1
2016-10-29 Female                       4
           Male                         3
2016-10-30 Female                       1
           Male                         3

                                 Are you going actually going trick or
treating yourself?  \
Date       Gender

2016-10-24 Female
231
           I'd rather not say
15
           Male
520
           Other
9
2016-10-25 Female
85
           I'd rather not say
4
           Male
60
           Other
1
2016-10-26 Female
13
           Male
19
           Other
2
2016-10-27 Female
61
           I'd rather not say
9
```

```
            Male
179
            Other
4
2016-10-28 Female
15
            Male
11
            Other
1
2016-10-29 Female
4
            Male
3
2016-10-30 Female
1
            Male
3

                                  How old are you?  \
Date       Gender
2016-10-24 Female                                225
           I'd rather not say                     13
           Male                                  510
           Other                                   9
2016-10-25 Female                                 84
           I'd rather not say                      4
           Male                                   58
           Other                                   1
2016-10-26 Female                                 13
           Male                                   19
           Other                                   2
2016-10-27 Female                                 59
           I'd rather not say                      7
           Male                                  177
           Other                                   4
2016-10-28 Female                                 15
           Male                                   10
           Other                                   1
2016-10-29 Female                                  4
           Male                                    3
2016-10-30 Female                                  1
           Male                                    2

                                  Which country do you live in?  \
Date       Gender
2016-10-24 Female                                          227
           I'd rather not say                               15
           Male                                            516
```

```
            Other                                                   9
2016-10-25 Female                                                  84
            I'd rather not say                                      4
            Male                                                   59
            Other                                                   1
2016-10-26 Female                                                  13
            Male                                                   19
            Other                                                   2
2016-10-27 Female                                                  59
            I'd rather not say                                      7
            Male                                                  178
            Other                                                   4
2016-10-28 Female                                                  15
            Male                                                   10
            Other                                                   1
2016-10-29 Female                                                   4
            Male                                                    3
2016-10-30 Female                                                   1
            Male                                                    2

                                  Which state, province, county do you
live in?  \
Date        Gender

2016-10-24 Female
215
            I'd rather not say
13
            Male
494
            Other
9
2016-10-25 Female
78
            I'd rather not say
3
            Male
57
            Other
1
2016-10-26 Female
13
            Male
19
            Other
1
2016-10-27 Female
55
            I'd rather not say
```

```
7
            Male
171
            Other
3
2016-10-28 Female
15
            Male
10
            Other
1
2016-10-29 Female
4
            Male
3
2016-10-30 Female
1
            Male
2

                                  [100 Grand Bar]  \
Date       Gender
2016-10-24 Female                             216
           I'd rather not say                  14
           Male                               492
           Other                                8
2016-10-25 Female                              74
           I'd rather not say                   3
           Male                                49
           Other                                1
2016-10-26 Female                              13
           Male                                19
           Other                                2
2016-10-27 Female                              58
           I'd rather not say                   9
           Male                               176
           Other                                4
2016-10-28 Female                              15
           Male                                10
           Other                                1
2016-10-29 Female                               4
           Male                                 3
2016-10-30 Female                               1
           Male                                 2

                                  [Anonymous brown globs that come in
black and orange wrappers]  \
Date       Gender

2016-10-24 Female
```

```
227
            I'd rather not say
13
            Male
509
            Other
9
2016-10-25 Female
85
            I'd rather not say
4
            Male
56
            Other
1
2016-10-26 Female
13
            Male
19
            Other
2
2016-10-27 Female
60
            I'd rather not say
9
            Male
176
            Other
4
2016-10-28 Female
15
            Male
11
            Other
1
2016-10-29 Female
4
            Male
3
2016-10-30 Female
1
            Male
3

                                    [Any full-sized candy bar]   [Black
Jacks]  \
Date       Gender

2016-10-24 Female                                          229
```

```
192
           I'd rather not say                                14
12
           Male                                             514
443
           Other                                              8
7
2016-10-25 Female                                            84
63
           I'd rather not say                                 4
4
           Male                                              59
46
           Other                                              1
1
2016-10-26 Female                                            13
10
           Male                                              19
14
           Other                                              2
2
2016-10-27 Female                                            61
51
           I'd rather not say                                 9
7
           Male                                             177
160
           Other                                              4
4
2016-10-28 Female                                            15
13
           Male                                              11
10
           Other                                              1
0
2016-10-29 Female                                             3
4
           Male                                               2
3
2016-10-30 Female                                             1
1
           Male                                               3
2

                                  [Bonkers (the candy)]  ...  \
Date       Gender                                        ...
2016-10-24 Female                                  183  ...
           I'd rather not say                       12  ...
           Male                                    438  ...
```

```
            Other                                              7  ...
2016-10-25 Female                                             62  ...
            I'd rather not say                                 3  ...
            Male                                              43  ...
            Other                                              1  ...
2016-10-26 Female                                             10  ...
            Male                                              18  ...
            Other                                              1  ...
2016-10-27 Female                                             50  ...
            I'd rather not say                                 7  ...
            Male                                             158  ...
            Other                                              4  ...
2016-10-28 Female                                             13  ...
            Male                                              10  ...
            Other                                              0  ...
2016-10-29 Female                                              4  ...
            Male                                               2  ...
2016-10-30 Female                                              1  ...
            Male                                               3  ...

                                  Please estimate the degree(s) of
separation you have from the following celebrities [JK Rowling]  \
Date       Gender

2016-10-24 Female
209
            I'd rather not say
14
            Male
492
            Other
9
2016-10-25 Female
79
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
13
            Male
18
            Other
2
2016-10-27 Female
57
            I'd rather not say
```

```
6
            Male
172
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Please estimate the degree(s) of
separation you have from the following celebrities [JJ Abrams]  \
Date       Gender

2016-10-24 Female
209
            I'd rather not say
14
            Male
493
            Other
9
2016-10-25 Female
78
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
12
            Male
18
            Other
2
2016-10-27 Female
58
            I'd rather not say
```

```
6
            Male
171
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Please estimate the degree(s) of
separation you have from the following celebrities [Beyoncé]  \
Date       Gender

2016-10-24 Female
209
            I'd rather not say
14
            Male
492
            Other
9
2016-10-25 Female
79
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
13
            Male
18
            Other
2
2016-10-27 Female
58
            I'd rather not say
```

```
6
            Male
169
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Please estimate the degree(s) of
separation you have from the following celebrities [Bieber]  \
Date       Gender

2016-10-24 Female
209
            I'd rather not say
14
            Male
491
            Other
9
2016-10-25 Female
79
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
13
            Male
18
            Other
2
2016-10-27 Female
58
            I'd rather not say
6
```

```
            Male
170
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Please estimate the degree(s) of
separation you have from the following celebrities [Kevin Bacon]  \
Date       Gender

2016-10-24 Female
209
            I'd rather not say
14
            Male
495
            Other
9
2016-10-25 Female
79
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
13
            Male
18
            Other
2
2016-10-27 Female
58
            I'd rather not say
7
```

```
            Male
169
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Please estimate the degree(s) of
separation you have from the following celebrities [Francis Bacon
(1561 - 1626)]  \
Date        Gender

2016-10-24 Female
207
            I'd rather not say
13
            Male
492
            Other
9
2016-10-25 Female
79
            I'd rather not say
4
            Male
57
            Other
1
2016-10-26 Female
13
            Male
18
            Other
2
2016-10-27 Female
58
            I'd rather not say
```

```
6
            Male
171
            Other
4
2016-10-28 Female
14
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
2

                                  Which day do you prefer, Friday or
Sunday?  \
Date        Gender

2016-10-24 Female
228
            I'd rather not say
14
            Male
510
            Other
9
2016-10-25 Female
84
            I'd rather not say
3
            Male
60
            Other
1
2016-10-26 Female
13
            Male
19
            Other
2
2016-10-27 Female
60
            I'd rather not say
```

```
8
            Male
176
            Other
4
2016-10-28 Female
15
            Male
11
            Other
1
2016-10-29 Female
3
            Male
3
2016-10-30 Female
1
            Male
3

                                  Do you eat apples the correct way, East
to West (side to side) or do you eat them like a freak of nature,
South to North (bottom to top)?  \
Date       Gender

2016-10-24 Female
225

            I'd rather not say
15

            Male
504

            Other
9

2016-10-25 Female
81

            I'd rather not say
4

            Male
59

            Other
1

2016-10-26 Female
```

```
13

            Male
18

            Other
1

2016-10-27 Female
59

            I'd rather not say
7

            Male
170

            Other
4

2016-10-28 Female
15

            Male
11

            Other
1

2016-10-29 Female
3

            Male
3

2016-10-30 Female
1

            Male
3


                                When you see the above image of the 4
different websites, which one would you most likely check out (please
be honest).  \
Date        Gender

2016-10-24 Female
227

            I'd rather not say
```

```
14

            Male
506

            Other
9

2016-10-25 Female
80

            I'd rather not say
4

            Male
60

            Other
1

2016-10-26 Female
13

            Male
19

            Other
2

2016-10-27 Female
59

            I'd rather not say
8

            Male
173

            Other
3

2016-10-28 Female
15

            Male
11

            Other
1

2016-10-29 Female
```

```
3

            Male
3

2016-10-30 Female
1

            Male
3


                                  [York Peppermint Patties] Ignore
Date       Gender
2016-10-24 Female                                                0
           I'd rather not say                                    0
           Male                                                  0
           Other                                                 0
2016-10-25 Female                                                0
           I'd rather not say                                    0
           Male                                                  0
           Other                                                 0
2016-10-26 Female                                                0
           Male                                                  0
           Other                                                 0
2016-10-27 Female                                                0
           I'd rather not say                                    0
           Male                                                  0
           Other                                                 0
2016-10-28 Female                                                0
           Male                                                  0
           Other                                                 0
2016-10-29 Female                                                0
           Male                                                  0
2016-10-30 Female                                                0
           Male                                                  0

[22 rows x 122 columns]
```

Transformation 2: Grouping with Index Levels

Grouping with Single Index

```
## Set Gender and Date as index column for the group by function

candy_hier_trans_df = candy_hier_trans_df.set_index(['Date','Gender'])

## Displaying few records using head command

candy_hier_trans_df.head()
```

```
                                  Timestamp  \
Date       Gender
2016-10-24 Male   2016-10-24 05:09:23.033
           Male   2016-10-24 05:09:54.798
           Female 2016-10-24 05:13:06.734
           Male   2016-10-24 05:14:17.192
           Male   2016-10-24 05:14:24.625

                  Are you going actually going trick or treating
yourself?  \
Date       Gender

2016-10-24 Male                                                  No

           Male                                                  No

           Female                                                No

           Male                                                  No

           Male                                                 Yes


                  How old are you? Which country do you live in?  \
Date       Gender
2016-10-24 Male                 22                          Canada
           Male                 45                             usa
           Female               48                              US
           Male                 57                             usa
           Male                 42                             USA

                  Which state, province, county do you live in?  \
Date       Gender
2016-10-24 Male                                           Ontario
           Male                                                il
           Female                                         Colorado
           Male                                                il
           Male                                      South Dakota

                   [100 Grand Bar]  \
Date       Gender
2016-10-24 Male               JOY
           Male               MEH
           Female             JOY
           Male               JOY
           Male               MEH

                   [Anonymous brown globs that come in black and
orange wrappers]  \
Date       Gender
```

```
2016-10-24 Male                                                    DESPAIR

           Male                                                        MEH

           Female                                                  DESPAIR

           Male                                                        MEH

           Male                                                    DESPAIR


                    [Any full-sized candy bar]  [Black Jacks]  \
Date       Gender
2016-10-24 Male                            JOY            MEH
           Male                            JOY            JOY
           Female                          JOY            MEH
           Male                            JOY            MEH
           Male                            JOY        DESPAIR

                    [Bonkers (the candy)]  ...  \
Date       Gender                          ...
2016-10-24 Male                       MEH  ...
           Male                    DESPAIR  ...
           Female                     MEH  ...
           Male                       MEH  ...
           Male                       MEH  ...

                   Please estimate the degree(s) of separation you have
from the following celebrities [JK Rowling]  \
Date       Gender

2016-10-24 Male                                            3 or higher

           Male                                            3 or higher

           Female                                          3 or higher

           Male                                            3 or higher

           Male                                            3 or higher


                   Please estimate the degree(s) of separation you have
from the following celebrities [JJ Abrams]  \
Date       Gender

2016-10-24 Male                                                    2

           Male                                            3 or higher
```

```
           Female                                                3 or higher

           Male                                                  3 or higher

           Male                                                  3 or higher


                   Please estimate the degree(s) of separation you have
from the following celebrities [Beyoncé]  \
Date       Gender

2016-10-24 Male                                                  3 or higher

           Male                                                  3 or higher

           Female                                                3 or higher

           Male                                                  3 or higher

           Male                                                  3 or higher


                   Please estimate the degree(s) of separation you have
from the following celebrities [Bieber]  \
Date       Gender

2016-10-24 Male                                                  3 or higher

           Male                                                  3 or higher

           Female                                                3 or higher

           Male                                                  3 or higher

           Male                                                  3 or higher


                   Please estimate the degree(s) of separation you have
from the following celebrities [Kevin Bacon]  \
Date       Gender

2016-10-24 Male                                                  3 or higher

           Male                                                  3 or higher

           Female                                                3 or higher

           Male                                                  3 or higher

           Male                                                  3 or higher
```

```
                  Please estimate the degree(s) of separation you have
from the following celebrities [Francis Bacon (1561 - 1626)]  \
Date       Gender

2016-10-24 Male                                                3 or higher

           Male                                                3 or higher

           Female                                              3 or higher

           Male                                                3 or higher

           Male                                                3 or higher


                  Which day do you prefer, Friday or Sunday?  \
Date       Gender
2016-10-24 Male                                            Friday
           Male                                            Friday
           Female                                          Sunday
           Male                                            Sunday
           Male                                            Friday

                  Do you eat apples the correct way, East to West
(side to side) or do you eat them like a freak of nature, South to
North (bottom to top)?  \
Date       Gender

2016-10-24 Male                                            South to North

           Male                                              East to West

           Female                                            East to West

           Male                                            South to North

           Male                                              East to West


                  When you see the above image of the 4 different
websites, which one would you most likely check out (please be
honest).  \
Date       Gender

2016-10-24 Male                    Science: Latest News and Headlines

           Male                    Science: Latest News and Headlines

           Female                  Science: Latest News and Headlines

           Male                    Science: Latest News and Headlines
```

```
           Male                                                          ESPN


                    [York Peppermint Patties] Ignore
Date       Gender
2016-10-24 Male                                  NaN
           Male                                  NaN
           Female                                NaN
           Male                                  NaN
           Male                                  NaN

[5 rows x 122 columns]
```

```
## Grouping with Index levels
## Display the Counts of records in each of the 'Date' groups

candy_hier_trans_df.groupby('Date')[['How old are you?']].count()
```

```
            How old are you?
Date
2016-10-24               760
2016-10-25               147
2016-10-26                35
2016-10-27               248
2016-10-28                26
2016-10-29                 7
2016-10-30                 3
```

Grouping with multiple Indices

```
## Grouping with Index levels
## Display the Counts of records in each of the 'Date' and 'Your
gender:' as a combined group
## This will help us drill down to next elevl of information like how
many people of each gender group responded on a given day

candy_hier_trans_df.groupby(['Date', 'Gender'])[['How old are
you?']].count()
```

```
                                How old are you?
Date       Gender
2016-10-24 Female                           225
           I'd rather not say                13
           Male                             510
           Other                              9
2016-10-25 Female                            84
           I'd rather not say                 4
           Male                              58
           Other                              1
2016-10-26 Female                            13
```

```
            Male                                19
            Other                                2
2016-10-27  Female                              59
            I'd rather not say                   7
            Male                               177
            Other                                4
2016-10-28  Female                              15
            Male                                10
            Other                                1
2016-10-29  Female                              4
            Male                                3
2016-10-30  Female                              1
            Male                                2
```

Transformation 3: Cross Tabs

```
## Crosstab operation to display the group frequencies / counts
between the combinations of various Dates and gender values
## Selected this combination to ensure the earlier results are
validated / cross-verified as well
## As an initial step, we have to reset the index value

candy_hier_trans_df = candy_hier_trans_df.reset_index()
pd.crosstab(candy_hier_trans_df.Date, candy_hier_trans_df['Gender'],
margins=True)
```

Gender	Female	I'd rather not say	Male	Other	All
Date					
2016-10-24	231	15	520	9	775
2016-10-25	85	4	60	1	150
2016-10-26	13	0	19	2	34
2016-10-27	61	9	179	4	253
2016-10-28	15	0	11	1	27
2016-10-29	4	0	3	0	7
2016-10-30	1	0	3	0	4
All	410	28	795	17	1250

```
## Crosstab operation to display the group frequencies / counts
between the combinations of
## various Dates and responses received for ' [Black Jacks]' column
## Selected this column because we had filled in default values for
this column

pd.crosstab(candy_hier_trans_df['Date'], candy_hier_trans_df[' [Black
Jacks]'], margins=True)
```

[Black Jacks]	DESPAIR	JOY	MEH	All
Date				
2016-10-24	333	40	285	658
2016-10-25	44	8	62	114

```
2016-10-26           15    4    8    27
2016-10-27          108   20   96   224
2016-10-28           10    2   11    23
2016-10-29            1    1    5     7
2016-10-30            2    1    0     3
All                 513   76  467  1056
```

Chapter 11 Transformations

Transformation 1: Convert between string and date time

```
## Considering the original dataframe and check the datatypes for each
of the field

datatypes = candy_hier_df.dtypes
print("The datatype for each of the field is given below: \n")
print(datatypes)
```

```
The datatype for each of the field is given below:

Timestamp
datetime64[ns]
Are you going actually going trick or treating yourself?
object
Your gender:
object
How old are you?
object
Which country do you live in?
object

...
Which day do you prefer, Friday or Sunday?
object
Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to top)?
object
When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).
object
 [York Peppermint Patties] Ignore
float64
Date
object
Length: 124, dtype: object
```

```
## We could see datatype for timestamp field is datetime.
## So, as part of this transformation, we will change the datatype for
this field to string
## Add the result into a new column
```

```
candy_hier_df['Timestamp_String'] =
candy_hier_df['Timestamp'].apply(lambda x: x.strftime('%c'))

## Displaying few records in the dataframe using head command

candy_hier_df.head(3)
```

```
                 Timestamp  \
0 2016-10-24 05:09:23.033
1 2016-10-24 05:09:54.798
2 2016-10-24 05:13:06.734

  Are you going actually going trick or treating yourself? Your
gender:  \
0                                                   No
Male
1                                                   No
Male
2                                                   No
Female

  How old are you? Which country do you live in?  \
0               22                          Canada
1               45                             usa
2               48                              US

  Which state, province, county do you live in?  [100 Grand Bar]  \
0                                      Ontario              JOY
1                                           il              MEH
2                                     Colorado              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                   DESPAIR
1                                                       MEH
2                                                   DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  ...  \
0                         JOY            MEH  ...
1                         JOY            JOY  ...
2                         JOY            MEH  ...

  Please estimate the degree(s) of separation you have from the
following celebrities [Beyoncé]  \
0                                      3 or higher

1                                      3 or higher

2                                      3 or higher
```

```
  Please estimate the degree(s) of separation you have from the
following celebrities [Bieber]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Kevin Bacon]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Please estimate the degree(s) of separation you have from the
following celebrities [Francis Bacon (1561 - 1626)]  \
0                                          3 or higher

1                                          3 or higher

2                                          3 or higher


  Which day do you prefer, Friday or Sunday?  \
0                                        Friday
1                                        Friday
2                                        Sunday

  Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to
top)?  \
0                                      South to North

1                                        East to West

2                                        East to West


  When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).  \
0                 Science: Latest News and Headlines

1                 Science: Latest News and Headlines

2                 Science: Latest News and Headlines
```

```
   [York Peppermint Patties] Ignore        Date
Timestamp_String
0                                  NaN  2016-10-24  Mon Oct 24 05:09:23
2016
1                                  NaN  2016-10-24  Mon Oct 24 05:09:54
2016
2                                  NaN  2016-10-24  Mon Oct 24 05:13:06
2016

[3 rows x 125 columns]
```

```
## Printing the datatype of each of the field after adding the field

datatypes = candy_hier_df.dtypes
print("The datatype for each of the field after transformation is
given below: \n")
print(datatypes)
```

```
The datatype for each of the field after transformation is given
below:

Timestamp
datetime64[ns]
Are you going actually going trick or treating yourself?
object
Your gender:
object
How old are you?
object
Which country do you live in?
object

...
Do you eat apples the correct way, East to West (side to side) or do
you eat them like a freak of nature, South to North (bottom to top)?
object
When you see the above image of the 4 different websites, which one
would you most likely check out (please be honest).
object
 [York Peppermint Patties] Ignore
float64
Date
object
Timestamp_String
object
Length: 125, dtype: object
```

From the above result, we could see the newly added field "Timestamp_String" is created as object instead of DateTime datatype

Transformation 2: Generate date range

Since the data available for 7 days (2016-10-24 thru 2016-10-30) in the dataset, we will generate Hour range and minute range index. Starting point would be the Timestamp from the first and last occurrences from 'Timestamp' column - the data is already sorted on the Timestamp column.

```
## Create Date range on Hour basis using the Timestamps from '0'
occurrence and '-1' (last) occurrence from DataFrame

hour_range = pd.date_range(candy_hier_df['Timestamp'].iloc[0],
candy_hier_df['Timestamp'].iloc[-1], freq='H')

## Display the hour_range index generated

hour_range

DatetimeIndex(['2016-10-24 05:09:23.033000', '2016-10-24
06:09:23.033000',
               '2016-10-24 07:09:23.033000', '2016-10-24
08:09:23.033000',
               '2016-10-24 09:09:23.033000', '2016-10-24
10:09:23.033000',
               '2016-10-24 11:09:23.033000', '2016-10-24
12:09:23.033000',
               '2016-10-24 13:09:23.033000', '2016-10-24
14:09:23.033000',
               ...
               '2016-10-30 07:09:23.033000', '2016-10-30
08:09:23.033000',
               '2016-10-30 09:09:23.033000', '2016-10-30
10:09:23.033000',
               '2016-10-30 11:09:23.033000', '2016-10-30
12:09:23.033000',
               '2016-10-30 13:09:23.033000', '2016-10-30
14:09:23.033000',
               '2016-10-30 15:09:23.033000', '2016-10-30
16:09:23.033000'],
              dtype='datetime64[ns]', length=156, freq='h')

## Create Date range on minutes basis using the Timestamps from '0'
occurrence and '-1' (last) occurrence from DataFrame

minute_range = pd.date_range(candy_hier_df['Timestamp'].iloc[0],
candy_hier_df['Timestamp'].iloc[-1], freq='T')

minute_range

DatetimeIndex(['2016-10-24 05:09:23.033000', '2016-10-24
05:10:23.033000',
               '2016-10-24 05:11:23.033000', '2016-10-24
```

```
05:12:23.033000',
               '2016-10-24 05:13:23.033000', '2016-10-24
05:14:23.033000',
               '2016-10-24 05:15:23.033000', '2016-10-24
05:16:23.033000',
               '2016-10-24 05:17:23.033000', '2016-10-24
05:18:23.033000',
               ...
               '2016-10-30 16:57:23.033000', '2016-10-30
16:58:23.033000',
               '2016-10-30 16:59:23.033000', '2016-10-30
17:00:23.033000',
               '2016-10-30 17:01:23.033000', '2016-10-30
17:02:23.033000',
               '2016-10-30 17:03:23.033000', '2016-10-30
17:04:23.033000',
               '2016-10-30 17:05:23.033000', '2016-10-30
17:06:23.033000'],
              dtype='datetime64[ns]', length=9358, freq='min')
```

Transformation 3: Frequencies and date offsets

Creating offset date / time ranges with predefined offset of 4 hours duration

```
## Create Date range on 4 hour basis using the above mentioned
Timestamps range known to us

offset_hour_range = pd.date_range(candy_hier_df['Timestamp'].iloc[0],
candy_hier_df['Timestamp'].iloc[-1], freq='4H')

## Display the 4 hour range index generated

offset_hour_range

DatetimeIndex(['2016-10-24 05:09:23.033000', '2016-10-24
09:09:23.033000',
               '2016-10-24 13:09:23.033000', '2016-10-24
17:09:23.033000',
               '2016-10-24 21:09:23.033000', '2016-10-25
01:09:23.033000',
               '2016-10-25 05:09:23.033000', '2016-10-25
09:09:23.033000',
               '2016-10-25 13:09:23.033000', '2016-10-25
17:09:23.033000',
               '2016-10-25 21:09:23.033000', '2016-10-26
01:09:23.033000',
               '2016-10-26 05:09:23.033000', '2016-10-26
09:09:23.033000',
               '2016-10-26 13:09:23.033000', '2016-10-26
17:09:23.033000',
               '2016-10-26 21:09:23.033000', '2016-10-27
```

```
01:09:23.033000',
               '2016-10-27 05:09:23.033000', '2016-10-27
09:09:23.033000',
               '2016-10-27 13:09:23.033000', '2016-10-27
17:09:23.033000',
               '2016-10-27 21:09:23.033000', '2016-10-28
01:09:23.033000',
               '2016-10-28 05:09:23.033000', '2016-10-28
09:09:23.033000',
               '2016-10-28 13:09:23.033000', '2016-10-28
17:09:23.033000',
               '2016-10-28 21:09:23.033000', '2016-10-29
01:09:23.033000',
               '2016-10-29 05:09:23.033000', '2016-10-29
09:09:23.033000',
               '2016-10-29 13:09:23.033000', '2016-10-29
17:09:23.033000',
               '2016-10-29 21:09:23.033000', '2016-10-30
01:09:23.033000',
               '2016-10-30 05:09:23.033000', '2016-10-30
09:09:23.033000',
               '2016-10-30 13:09:23.033000'],
              dtype='datetime64[ns]', freq='4h')
```

Creating offset date / time ranges with predefined offset of 30 minutes duration

```
## Create Date range on 30 minutes basis using the above mentioned
Timestamps range known to us

offset_minute_range =
pd.date_range(candy_hier_df['Timestamp'].iloc[0],
candy_hier_df['Timestamp'].iloc[-1], freq='30min')

## Display the 30 minute range index generated

offset_minute_range
```

```
DatetimeIndex(['2016-10-24 05:09:23.033000', '2016-10-24
05:39:23.033000',
               '2016-10-24 06:09:23.033000', '2016-10-24
06:39:23.033000',
               '2016-10-24 07:09:23.033000', '2016-10-24
07:39:23.033000',
               '2016-10-24 08:09:23.033000', '2016-10-24
08:39:23.033000',
               '2016-10-24 09:09:23.033000', '2016-10-24
09:39:23.033000',
               ...
               '2016-10-30 12:09:23.033000', '2016-10-30
12:39:23.033000',
```

```
               '2016-10-30 13:09:23.033000', '2016-10-30
13:39:23.033000',
               '2016-10-30 14:09:23.033000', '2016-10-30
14:39:23.033000',
               '2016-10-30 15:09:23.033000', '2016-10-30
15:39:23.033000',
               '2016-10-30 16:09:23.033000', '2016-10-30
16:39:23.033000'],
              dtype='datetime64[ns]', length=312, freq='30min')
```