DSC540-T303-Data-Preparation-Week7-8

January 27, 2025

[1]:
```
# Weeks 7 & 8: Data Cleaning and Transforming
```

[2]:
```
# select at least two methods from each chapter to perform on one of the␣
↪datasets.
# 1. •        Chapter 7
# o         Filter out missing data
# o         Fill in missing data
# o         Remove duplicates
# o         Transform data using either mapping or a function
# o         Replace values
# o         Discretization and Binning
# o         Manipulate Strings
```

[3]:
```
#Load the necessary libraries
import numpy as np
import pandas as pd
```

[4]:
```
# Load the .xlsx file
file_path = "CANDY-HIERARCHY-2015-SURVEY-Responses.xlsx"
df = pd.read_excel(file_path)
```

[5]:
```
df.head(10)
```

[5]:
```
                Timestamp         How old are you?  \
0 2015-10-23 08:46:20.451                       35
1 2015-10-23 08:46:51.583                       41
2 2015-10-23 08:47:34.285                       33
3 2015-10-23 08:47:58.964                       31
4 2015-10-23 08:48:11.719                       30
5 2015-10-23 08:49:06.808                       38
6 2015-10-23 08:50:08.918                       48
7 2015-10-23 08:52:14.267                       39
8 2015-10-23 08:52:22.112  8999999999999995805696
9 2015-10-23 08:53:30.967                       54

  Are you going actually going trick or treating yourself?  [Butterfinger]  \
0                                                 No                 JOY
1                                                 No                 JOY
```

```
2                                                          No             DESPAIR
3                                                          No                 JOY
4                                                          No                 NaN
5                                                          No                 JOY
6                                                          No                 JOY
7                                                          No             DESPAIR
8                                                         Yes             DESPAIR
9                                                          No                 JOY

  [100 Grand Bar]  \
0             NaN
1             JOY
2         DESPAIR
3             JOY
4             JOY
5             JOY
6             JOY
7             JOY
8         DESPAIR
9             JOY

  [Anonymous brown globs that come in black and orange wrappers]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                            DESPAIR
5                                            DESPAIR
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
9                                            DESPAIR

  [Any full-sized candy bar]  [Black Jacks]  [Bonkers]  [Bottle Caps]  …  \
0                        JOY            NaN        NaN            NaN  …
1                        JOY        DESPAIR    DESPAIR            JOY  …
2                        JOY        DESPAIR    DESPAIR        DESPAIR  …
3                        JOY        DESPAIR    DESPAIR            JOY  …
4                        JOY            NaN        NaN            NaN  …
5                        JOY        DESPAIR        JOY            JOY  …
6                        JOY            JOY    DESPAIR            JOY  …
7                        NaN            NaN        NaN            JOY  …
8                    DESPAIR        DESPAIR    DESPAIR        DESPAIR  …
9                        JOY            JOY    DESPAIR            JOY  …

  [Necco Wafers] Which day do you prefer, Friday or Sunday?  \
0           NaN                                       NaN
```

```
1         DESPAIR                                        NaN
2         DESPAIR                                        NaN
3         DESPAIR                                        NaN
4             NaN                                        NaN
5             JOY                                        NaN
6         DESPAIR                                        NaN
7             JOY                                        NaN
8         DESPAIR                                        NaN
9             JOY                                        NaN

  Please estimate the degrees of separation you have from the following folks
[Bruce Lee]  \
0                                               NaN
1                                               NaN
2                                               NaN
3                                               NaN
4                                               NaN
5                                               NaN
6                                               NaN
7                                               NaN
8                                               NaN
9                                               NaN

  Please estimate the degrees of separation you have from the following folks
[JK Rowling]  \
0                                               NaN
1                                               NaN
2                                               NaN
3                                               NaN
4                                               NaN
5                                               NaN
6                                               NaN
7                                               NaN
8                                               NaN
9                                               NaN

  Please estimate the degrees of separation you have from the following folks
[Malala Yousafzai]  \
0                                               NaN
1                                               NaN
2                                               NaN
3                                               NaN
4                                               NaN
5                                               NaN
6                                               NaN
7                                               NaN
8                                               NaN
```

```
9                                                NaN

  Please estimate the degrees of separation you have from the following folks
[Thom Yorke]  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
5                                                NaN
6                                                NaN
7                                                NaN
8                                                NaN
9                                                NaN

  Please estimate the degrees of separation you have from the following folks
[JJ Abrams]  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
5                                                NaN
6                                                NaN
7                                                NaN
8                                                NaN
9                                                NaN

  Please estimate the degrees of separation you have from the following folks
[Hillary Clinton]  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
5                                                NaN
6                                                NaN
7                                                NaN
8                                                NaN
9                                                NaN

  Please estimate the degrees of separation you have from the following folks
[Donald Trump]  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
```

```
4                                                NaN
5                                                NaN
6                                                NaN
7                                                NaN
8                                                NaN
9                                                NaN

  Please estimate the degrees of separation you have from the following folks
[Beyoncé Knowles]
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN
5                                                NaN
6                                                NaN
7                                                NaN
8                                                NaN
9                                                NaN

[10 rows x 124 columns]
```

[6]:
```
# Remove the last 100 columns
df = df.iloc[:, :-100]
df.head(10)
```

[6]:
```
                Timestamp         How old are you?  \
0 2015-10-23 08:46:20.451                       35
1 2015-10-23 08:46:51.583                       41
2 2015-10-23 08:47:34.285                       33
3 2015-10-23 08:47:58.964                       31
4 2015-10-23 08:48:11.719                       30
5 2015-10-23 08:49:06.808                       38
6 2015-10-23 08:50:08.918                       48
7 2015-10-23 08:52:14.267                       39
8 2015-10-23 08:52:22.112  8999999999999995805696
9 2015-10-23 08:53:30.967                       54

  Are you going actually going trick or treating yourself?  [Butterfinger]  \
0                                                 No                 JOY
1                                                 No                 JOY
2                                                 No             DESPAIR
3                                                 No                 JOY
4                                                 No                 NaN
5                                                 No                 JOY
6                                                 No                 JOY
7                                                 No             DESPAIR
```

```
8                                                Yes              DESPAIR
9                                                 No                  JOY

   [100 Grand Bar]  \
0              NaN
1              JOY
2          DESPAIR
3              JOY
4              JOY
5              JOY
6              JOY
7              JOY
8          DESPAIR
9              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                            DESPAIR
5                                            DESPAIR
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
9                                            DESPAIR

   [Any full-sized candy bar]  [Black Jacks]  [Bonkers]  [Bottle Caps]  …  \
0                         JOY            NaN        NaN            NaN  …
1                         JOY        DESPAIR    DESPAIR            JOY  …
2                         JOY        DESPAIR    DESPAIR        DESPAIR  …
3                         JOY        DESPAIR    DESPAIR            JOY  …
4                         JOY            NaN        NaN            NaN  …
5                         JOY        DESPAIR        JOY            JOY  …
6                         JOY            JOY    DESPAIR            JOY  …
7                         NaN            NaN        NaN            JOY  …
8                     DESPAIR        DESPAIR    DESPAIR        DESPAIR  …
9                         JOY            JOY    DESPAIR            JOY  …

   [Candy Corn]  \
0           NaN
1       DESPAIR
2           JOY
3       DESPAIR
4           NaN
5       DESPAIR
6       DESPAIR
```

```
7          NaN
8          JOY
9      DESPAIR

  [Vials of pure high fructose corn syrup, for main-lining into your vein]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                                NaN
5                                                JOY
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
9                                                JOY

  [Candy that is clearly just the stuff given out for free at restaurants]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                                NaN
5                                            DESPAIR
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
9                                            DESPAIR

  [Cash, or other forms of legal tender]  [Chiclets]  [Caramellos]  \
0                                    JOY         NaN           NaN
1                                    JOY     DESPAIR       DESPAIR
2                                    JOY     DESPAIR           JOY
3                                    JOY     DESPAIR           JOY
4                                    JOY         NaN           JOY
5                                    JOY         JOY           JOY
6                                    JOY     DESPAIR           JOY
7                                    JOY         JOY           JOY
8                                DESPAIR     DESPAIR       DESPAIR
9                                    JOY     DESPAIR           JOY

  [Snickers]  [Dark Chocolate Hershey]  [Dental paraphenalia]   [Dots]
0        JOY                       JOY                DESPAIR      JOY
1        JOY                   DESPAIR                DESPAIR      JOY
2        JOY                       JOY                DESPAIR  DESPAIR
3        JOY                       JOY                DESPAIR  DESPAIR
4        NaN                       NaN                    NaN      JOY
5    DESPAIR                   DESPAIR                DESPAIR      JOY
```

```
6        JOY                       JOY                DESPAIR  DESPAIR
7        JOY                       NaN                    JOY      NaN
8    DESPAIR                   DESPAIR                DESPAIR  DESPAIR
9        JOY                       JOY                DESPAIR      JOY

[10 rows x 24 columns]
```

[7]:
```
df.shape
df.head(10)
```

[7]:
```
                Timestamp        How old are you?  \
0 2015-10-23 08:46:20.451                      35
1 2015-10-23 08:46:51.583                      41
2 2015-10-23 08:47:34.285                      33
3 2015-10-23 08:47:58.964                      31
4 2015-10-23 08:48:11.719                      30
5 2015-10-23 08:49:06.808                      38
6 2015-10-23 08:50:08.918                      48
7 2015-10-23 08:52:14.267                      39
8 2015-10-23 08:52:22.112  8999999999999995805696
9 2015-10-23 08:53:30.967                      54

  Are you going actually going trick or treating yourself?  [Butterfinger]  \
0                                                 No                  JOY
1                                                 No                  JOY
2                                                 No              DESPAIR
3                                                 No                  JOY
4                                                 No                  NaN
5                                                 No                  JOY
6                                                 No                  JOY
7                                                 No              DESPAIR
8                                                Yes              DESPAIR
9                                                 No                  JOY

   [100 Grand Bar]  \
0              NaN
1              JOY
2          DESPAIR
3              JOY
4              JOY
5              JOY
6              JOY
7              JOY
8          DESPAIR
9              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                                         DESPAIR
1                                                         DESPAIR
2                                                         DESPAIR
3                                                         DESPAIR
4                                                         DESPAIR
5                                                         DESPAIR
6                                                         DESPAIR
7                                                         DESPAIR
8                                                         DESPAIR
9                                                         DESPAIR

  [Any full-sized candy bar]  [Black Jacks]  [Bonkers]  [Bottle Caps]  …  \
0                        JOY            NaN        NaN            NaN  …
1                        JOY        DESPAIR    DESPAIR            JOY  …
2                        JOY        DESPAIR    DESPAIR        DESPAIR  …
3                        JOY        DESPAIR    DESPAIR            JOY  …
4                        JOY            NaN        NaN            NaN  …
5                        JOY        DESPAIR        JOY            JOY  …
6                        JOY            JOY    DESPAIR            JOY  …
7                        NaN            NaN        NaN            JOY  …
8                    DESPAIR        DESPAIR    DESPAIR        DESPAIR  …
9                        JOY            JOY    DESPAIR            JOY  …

  [Candy Corn]  \
0          NaN
1      DESPAIR
2          JOY
3      DESPAIR
4          NaN
5      DESPAIR
6      DESPAIR
7          NaN
8          JOY
9      DESPAIR

  [Vials of pure high fructose corn syrup, for main-lining into your vein]  \
0                                                         DESPAIR
1                                                         DESPAIR
2                                                         DESPAIR
3                                                         DESPAIR
4                                                             NaN
5                                                             JOY
6                                                         DESPAIR
7                                                         DESPAIR
8                                                         DESPAIR
9                                                             JOY
```

```
   [Candy that is clearly just the stuff given out for free at restaurants]  \
0                                            DESPAIR                          
1                                            DESPAIR                          
2                                            DESPAIR                          
3                                            DESPAIR                          
4                                                NaN                          
5                                            DESPAIR                          
6                                            DESPAIR                          
7                                            DESPAIR                          
8                                            DESPAIR                          
9                                            DESPAIR                          

   [Cash, or other forms of legal tender]  [Chiclets]  [Caramellos]  \
0                                     JOY         NaN           NaN   
1                                     JOY     DESPAIR       DESPAIR   
2                                     JOY     DESPAIR           JOY   
3                                     JOY     DESPAIR           JOY   
4                                     JOY         NaN           JOY   
5                                     JOY         JOY           JOY   
6                                     JOY     DESPAIR           JOY   
7                                     JOY         JOY           JOY   
8                                 DESPAIR     DESPAIR       DESPAIR   
9                                     JOY     DESPAIR           JOY   

   [Snickers]  [Dark Chocolate Hershey]  [Dental paraphenalia]   [Dots]
0         JOY                       JOY                DESPAIR      JOY
1         JOY                   DESPAIR                DESPAIR      JOY
2         JOY                       JOY                DESPAIR  DESPAIR
3         JOY                       JOY                DESPAIR  DESPAIR
4         NaN                       NaN                    NaN      JOY
5     DESPAIR                   DESPAIR                DESPAIR      JOY
6         JOY                       JOY                DESPAIR  DESPAIR
7         JOY                       NaN                    JOY      NaN
8     DESPAIR                   DESPAIR                DESPAIR  DESPAIR
9         JOY                       JOY                DESPAIR      JOY

[10 rows x 24 columns]
```

[8]:
```
# o        Filter out missing data
# o        Fill in missing data
```

[9]:
```
# Drop column with missing data
df.dropna(axis='columns',how='all')

# replace space from column name with '_'
df.columns = df.columns.str.replace(' ', '_')
df.head(10)
```

```
[9]:                  Timestamp        How_old_are_you?  \
     0 2015-10-23 08:46:20.451                      35
     1 2015-10-23 08:46:51.583                      41
     2 2015-10-23 08:47:34.285                      33
     3 2015-10-23 08:47:58.964                      31
     4 2015-10-23 08:48:11.719                      30
     5 2015-10-23 08:49:06.808                      38
     6 2015-10-23 08:50:08.918                      48
     7 2015-10-23 08:52:14.267                      39
     8 2015-10-23 08:52:22.112  8999999999999995805696
     9 2015-10-23 08:53:30.967                      54

       Are_you_going_actually_going_trick_or_treating_yourself? _[Butterfinger]  \
     0                                                 No                   JOY
     1                                                 No                   JOY
     2                                                 No               DESPAIR
     3                                                 No                   JOY
     4                                                 No                   NaN
     5                                                 No                   JOY
     6                                                 No                   JOY
     7                                                 No               DESPAIR
     8                                                Yes               DESPAIR
     9                                                 No                   JOY

       _[100_Grand_Bar]  \
     0              NaN
     1              JOY
     2          DESPAIR
     3              JOY
     4              JOY
     5              JOY
     6              JOY
     7              JOY
     8          DESPAIR
     9              JOY

       _[Anonymous_brown_globs_that_come_in_black_and_orange_wrappers]  \
     0                                            DESPAIR
     1                                            DESPAIR
     2                                            DESPAIR
     3                                            DESPAIR
     4                                            DESPAIR
     5                                            DESPAIR
     6                                            DESPAIR
     7                                            DESPAIR
     8                                            DESPAIR
     9                                            DESPAIR
```

```
  _[Any_full-sized_candy_bar] _[Black_Jacks] _[Bonkers] _[Bottle_Caps]  …  \
0                         JOY           NaN        NaN            NaN  …
1                         JOY       DESPAIR    DESPAIR            JOY  …
2                         JOY       DESPAIR    DESPAIR        DESPAIR  …
3                         JOY       DESPAIR    DESPAIR            JOY  …
4                         JOY           NaN        NaN            NaN  …
5                         JOY       DESPAIR        JOY            JOY  …
6                         JOY           JOY    DESPAIR            JOY  …
7                         NaN           NaN        NaN            JOY  …
8                     DESPAIR       DESPAIR    DESPAIR        DESPAIR  …
9                         JOY           JOY    DESPAIR            JOY  …

  _[Candy_Corn]  \
0           NaN
1       DESPAIR
2           JOY
3       DESPAIR
4           NaN
5       DESPAIR
6       DESPAIR
7           NaN
8           JOY
9       DESPAIR

  _[Vials_of_pure_high_fructose_corn_syrup,_for_main-lining_into_your_vein]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                                NaN
5                                                JOY
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
9                                                JOY

  _[Candy_that_is_clearly_just_the_stuff_given_out_for_free_at_restaurants]  \
0                                            DESPAIR
1                                            DESPAIR
2                                            DESPAIR
3                                            DESPAIR
4                                                NaN
5                                            DESPAIR
6                                            DESPAIR
7                                            DESPAIR
8                                            DESPAIR
```

```
9                                             DESPAIR

  _[Cash,_or_other_forms_of_legal_tender] _[Chiclets] _[Caramellos]  \
0                                     JOY        NaN          NaN
1                                     JOY    DESPAIR      DESPAIR
2                                     JOY    DESPAIR          JOY
3                                     JOY    DESPAIR          JOY
4                                     JOY        NaN          JOY
5                                     JOY        JOY          JOY
6                                     JOY    DESPAIR          JOY
7                                     JOY        JOY          JOY
8                                 DESPAIR    DESPAIR      DESPAIR
9                                     JOY    DESPAIR          JOY

  _[Snickers] _[Dark_Chocolate_Hershey] _[Dental_paraphenalia]  _[Dots]
0         JOY                       JOY                DESPAIR      JOY
1         JOY                   DESPAIR                DESPAIR      JOY
2         JOY                       JOY                DESPAIR  DESPAIR
3         JOY                       JOY                DESPAIR  DESPAIR
4         NaN                       NaN                    NaN      JOY
5     DESPAIR                   DESPAIR                DESPAIR      JOY
6         JOY                       JOY                DESPAIR  DESPAIR
7         JOY                       NaN                    JOY      NaN
8     DESPAIR                   DESPAIR                DESPAIR  DESPAIR
9         JOY                       JOY                DESPAIR      JOY

[10 rows x 24 columns]
```

[10]:
```
# 2. remove [] from the column name
df.columns = [col.replace("[", "") for col in df.columns]
df.columns = [col.replace("]", "") for col in df.columns]

# rename column 'How_old_are_you' to 'age'
df.columns = [col.replace("How_old_are_you?", "age") for col in df.columns]
```

[11]:
```
# Create a new DataFrame with only Age and Timestamp
df = df[["Timestamp",␣
↪"age","Are_you_going_actually_going_trick_or_treating_yourself?
↪","_Any_full-sized_candy_bar","_100_Grand_Bar"]]

df.head(10)
```

[11]:
```
                Timestamp                      age  \
0 2015-10-23 08:46:20.451                       35
1 2015-10-23 08:46:51.583                       41
2 2015-10-23 08:47:34.285                       33
3 2015-10-23 08:47:58.964                       31
```

```
4 2015-10-23 08:48:11.719                       30
5 2015-10-23 08:49:06.808                       38
6 2015-10-23 08:50:08.918                       48
7 2015-10-23 08:52:14.267                       39
8 2015-10-23 08:52:22.112  89999999999999995805696
9 2015-10-23 08:53:30.967                       54

  Are_you_going_actually_going_trick_or_treating_yourself?  \
0                                                      No
1                                                      No
2                                                      No
3                                                      No
4                                                      No
5                                                      No
6                                                      No
7                                                      No
8                                                     Yes
9                                                      No

  _Any_full-sized_candy_bar _100_Grand_Bar
0                       JOY            NaN
1                       JOY            JOY
2                       JOY        DESPAIR
3                       JOY            JOY
4                       JOY            JOY
5                       JOY            JOY
6                       JOY            JOY
7                       NaN            JOY
8                   DESPAIR        DESPAIR
9                       JOY            JOY
```

[12]:
```
#Fill in missing data
```

[13]:
```
df.fillna('JOY')
```

[13]:
```
                    Timestamp age  \
0     2015-10-23 08:46:20.451  35
1     2015-10-23 08:46:51.583  41
2     2015-10-23 08:47:34.285  33
3     2015-10-23 08:47:58.964  31
4     2015-10-23 08:48:11.719  30
…                           …  ..
5625 2015-10-31 05:23:40.526  50
5626 2015-10-31 05:29:26.937  43
5627 2015-10-31 06:13:29.083  35
5628 2015-10-31 06:26:52.566  38
5629 2015-10-31 06:41:31.904  44
```

```
     Are_you_going_actually_going_trick_or_treating_yourself?  \
0                                                    No
1                                                    No
2                                                    No
3                                                    No
4                                                    No
…                                                 …
5625                                                 No
5626                                                 No
5627                                                Yes
5628                                                 No
5629                                                 No

     _Any_full-sized_candy_bar _100_Grand_Bar
0                          JOY            JOY
1                          JOY            JOY
2                          JOY       DESPAIR
3                          JOY            JOY
4                          JOY            JOY
…                        …              …
5625                       JOY       DESPAIR
5626                       JOY            JOY
5627                       JOY            JOY
5628                       JOY            JOY
5629                       JOY            JOY

[5630 rows x 5 columns]
```

[14]:
```
# •         Chapter 8
# o         Create hierarchical index
# o         Combine and Merge Datasets (you will have to either create a new␣
 ↪dataset from your existing data or create a relationship between the data I␣
 ↪have provided)
# o         Reshape
# o         Pivot the data
```

[15]:
```
# Set hierarchical index
# use the set_index() method to create a hierarchical index based on multiple␣
 ↪columns.
# For instance, let's use Timestamp and How old are you? as the hierarchical␣
 ↪index:
df_hierarchical = df.set_index(["Timestamp","age"])

print(df_hierarchical)
```

```
Are_you_going_actually_going_trick_or_treating_yourself?  \
Timestamp                   age
```

```
2015-10-23 08:46:20.451 35                                                  No
2015-10-23 08:46:51.583 41                                                  No
2015-10-23 08:47:34.285 33                                                  No
2015-10-23 08:47:58.964 31                                                  No
2015-10-23 08:48:11.719 30                                                  No
…                                                                         …
2015-10-31 05:23:40.526 50                                                  No
2015-10-31 05:29:26.937 43                                                  No
2015-10-31 06:13:29.083 35                                                 Yes
2015-10-31 06:26:52.566 38                                                  No
2015-10-31 06:41:31.904 44                                                  No

                            _Any_full-sized_candy_bar _100_Grand_Bar
Timestamp               age
2015-10-23 08:46:20.451 35                        JOY            NaN
2015-10-23 08:46:51.583 41                        JOY            JOY
2015-10-23 08:47:34.285 33                        JOY        DESPAIR
2015-10-23 08:47:58.964 31                        JOY            JOY
2015-10-23 08:48:11.719 30                        JOY            JOY
…                                               …            …
2015-10-31 05:23:40.526 50                        JOY        DESPAIR
2015-10-31 05:29:26.937 43                        JOY            JOY
2015-10-31 06:13:29.083 35                        JOY            JOY
2015-10-31 06:26:52.566 38                        JOY            JOY
2015-10-31 06:41:31.904 44                        JOY            JOY

[5630 rows x 3 columns]
```

[16]:
```python
# Access all data for a specific timestamp and age
print(df_hierarchical.loc[("10/23/2015 8:46:52", 41)])

# Access only a specific level
print(df_hierarchical.xs(41, level="age"))
```

```
Empty DataFrame
Columns: [Are_you_going_actually_going_trick_or_treating_yourself?, _Any_full-
sized_candy_bar, _100_Grand_Bar]
Index: []
                        Are_you_going_actually_going_trick_or_treating_yourself?
\
Timestamp
2015-10-23 08:46:51.583                                                     No
2015-10-23 09:24:36.675                                                     No
2015-10-23 09:57:47.100                                                    Yes
2015-10-23 10:11:44.776                                                     No
2015-10-23 10:36:35.401                                                     No
…                                                                         …
2015-10-29 13:23:02.941                                                     No
```

```
2015-10-29 14:29:33.212                                                 No
2015-10-29 19:52:22.161                                                 No
2015-10-30 16:59:44.172                                                 No
2015-10-30 20:04:25.653                                                 No

                        _Any_full-sized_candy_bar _100_Grand_Bar
Timestamp
2015-10-23 08:46:51.583                       JOY            JOY
2015-10-23 09:24:36.675                       JOY            JOY
2015-10-23 09:57:47.100                       JOY            JOY
2015-10-23 10:11:44.776                       JOY            JOY
2015-10-23 10:36:35.401                       JOY            JOY
…                                           …            …
2015-10-29 13:23:02.941                       NaN            NaN
2015-10-29 14:29:33.212                       JOY            JOY
2015-10-29 19:52:22.161                       JOY        DESPAIR
2015-10-30 16:59:44.172                       JOY            JOY
2015-10-30 20:04:25.653                       JOY        DESPAIR

[116 rows x 3 columns]
```

[17]:
```python
# Resetting the Index
#df_reset = df_hierarchical.reset_index()
```

[18]:
```python
# Use three columns as the index
df_hierarchical = df.set_index(["Timestamp", "age", "_100_Grand_Bar"])
print(df_hierarchical)
```

```
Are_you_going_actually_going_trick_or_treating_yourself?  \
Timestamp               age _100_Grand_Bar
2015-10-23 08:46:20.451 35  NaN
No
2015-10-23 08:46:51.583 41  JOY
No
2015-10-23 08:47:34.285 33  DESPAIR
No
2015-10-23 08:47:58.964 31  JOY
No
2015-10-23 08:48:11.719 30  JOY
No
…
…
2015-10-31 05:23:40.526 50  DESPAIR
No
2015-10-31 05:29:26.937 43  JOY
No
2015-10-31 06:13:29.083 35  JOY
Yes
```

```
2015-10-31 06:26:52.566 38  JOY
No
2015-10-31 06:41:31.904 44  JOY
No

                                            _Any_full-sized_candy_bar
Timestamp               age _100_Grand_Bar
2015-10-23 08:46:20.451 35  NaN                                      JOY
2015-10-23 08:46:51.583 41  JOY                                      JOY
2015-10-23 08:47:34.285 33  DESPAIR                                  JOY
2015-10-23 08:47:58.964 31  JOY                                      JOY
2015-10-23 08:48:11.719 30  JOY                                      JOY
…                                                                  …
2015-10-31 05:23:40.526 50  DESPAIR                                  JOY
2015-10-31 05:29:26.937 43  JOY                                      JOY
2015-10-31 06:13:29.083 35  JOY                                      JOY
2015-10-31 06:26:52.566 38  JOY                                      JOY
2015-10-31 06:41:31.904 44  JOY                                      JOY

[5630 rows x 2 columns]
```

[19]:
```python
# Pivot the data
pivot_table = df.pivot(index="Timestamp", columns="age")

print(pivot_table)
```

```
                        Are_you_going_actually_going_trick_or_treating_yourself?
\
age                                                                          NaN
Timestamp
2015-10-23 08:46:20.451                                                   NaN
2015-10-23 08:46:51.583                                                   NaN
2015-10-23 08:47:34.285                                                   NaN
2015-10-23 08:47:58.964                                                   NaN
2015-10-23 08:48:11.719                                                   NaN
…                                                                       …
2015-10-31 05:23:40.526                                                   NaN
2015-10-31 05:29:26.937                                                   NaN
2015-10-31 06:13:29.083                                                   NaN
2015-10-31 06:26:52.566                                                   NaN
2015-10-31 06:41:31.904                                                   NaN

                                                                    …  \
age                        0 0.62    5    6    7    8    9   10   11 …
Timestamp                                                           …
2015-10-23 08:46:20.451  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN …
2015-10-23 08:46:51.583  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN …
2015-10-23 08:47:34.285  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN …
```

```
2015-10-23 08:47:58.964  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
2015-10-23 08:48:11.719  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
…                       …   …   …   …   …   …   …   …   …   …
2015-10-31 05:23:40.526  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
2015-10-31 05:29:26.937  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
2015-10-31 06:13:29.083  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
2015-10-31 06:26:52.566  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …
2015-10-31 06:41:31.904  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  …

                                   _100_Grand_Bar                      \
age                      old enough to know better old enough to party
Timestamp
2015-10-23 08:46:20.451                        NaN                NaN
2015-10-23 08:46:51.583                        NaN                NaN
2015-10-23 08:47:34.285                        NaN                NaN
2015-10-23 08:47:58.964                        NaN                NaN
2015-10-23 08:48:11.719                        NaN                NaN
…                                            …                  …
2015-10-31 05:23:40.526                        NaN                NaN
2015-10-31 05:29:26.937                        NaN                NaN
2015-10-31 06:13:29.083                        NaN                NaN
2015-10-31 06:26:52.566                        NaN                NaN
2015-10-31 06:41:31.904                        NaN                NaN

                                                                              \
age                      older than dirt over 40  too too old too old for this
Timestamp
2015-10-23 08:46:20.451              NaN     NaN  NaN     NaN              NaN
2015-10-23 08:46:51.583              NaN     NaN  NaN     NaN              NaN
2015-10-23 08:47:34.285              NaN     NaN  NaN     NaN              NaN
2015-10-23 08:47:58.964              NaN     NaN  NaN     NaN              NaN
2015-10-23 08:48:11.719              NaN     NaN  NaN     NaN              NaN
…                                  …      …   …       …                …
2015-10-31 05:23:40.526              NaN     NaN  NaN     NaN              NaN
2015-10-31 05:29:26.937              NaN     NaN  NaN     NaN              NaN
2015-10-31 06:13:29.083              NaN     NaN  NaN     NaN              NaN
2015-10-31 06:26:52.566              NaN     NaN  NaN     NaN              NaN
2015-10-31 06:41:31.904              NaN     NaN  NaN     NaN              NaN


age                      very    x
Timestamp
2015-10-23 08:46:20.451  NaN  NaN  NaN
2015-10-23 08:46:51.583  NaN  NaN  NaN
2015-10-23 08:47:34.285  NaN  NaN  NaN
2015-10-23 08:47:58.964  NaN  NaN  NaN
2015-10-23 08:48:11.719  NaN  NaN  NaN
…                       …   …   …
```

```
2015-10-31 05:23:40.526  NaN  NaN  NaN
2015-10-31 05:29:26.937  NaN  NaN  NaN
2015-10-31 06:13:29.083  NaN  NaN  NaN
2015-10-31 06:26:52.566  NaN  NaN  NaN
2015-10-31 06:41:31.904  NaN  NaN  NaN

[5628 rows x 441 columns]
```

[20]:
```
# •         Chapter 10
# o         Grouping with Dicts/Series
# o         Grouping with Functions
# o         Grouping with Index Levels
# o         Split/Apply/Combine
# o         Cross Tabs
```

[21]:
```
# Define grouping function

grouped = df.groupby("age")
grouped.describe()
```

[21]:
```
                 Timestamp                                \
                     count                          mean
age
0                        1      2015-10-30 05:41:53.620000
0.62                     1      2015-10-28 19:49:06.308000
5                        1   2015-10-24 19:28:30.211000064
6                        4   2015-10-26 00:59:51.196250112
7                        2   2015-10-27 10:44:31.238000128
…                      …                             …
too old                  6   2015-10-29 00:16:38.813333248
too old for this         1      2015-10-24 13:25:54.240000
very                     5   2015-10-25 06:15:00.264600064
x                        1      2015-10-28 19:43:31.164000
                        1  2015-10-24 23:08:46.592999936

                                                                      \
                                       min                          25%
age
0                 2015-10-30 05:41:53.620000     2015-10-30 05:41:53.620000
0.62              2015-10-28 19:49:06.308000     2015-10-28 19:49:06.308000
5                 2015-10-24 19:28:30.211000  2015-10-24 19:28:30.211000064
6                 2015-10-23 10:44:29.949000  2015-10-23 10:44:42.758249984
7                 2015-10-25 00:59:04.891000  2015-10-26 05:51:48.064499968
…                                       …                             …
too old           2015-10-28 16:53:47.520000  2015-10-28 17:36:52.894249984
too old for this  2015-10-24 13:25:54.240000     2015-10-24 13:25:54.240000
very              2015-10-23 09:49:02.195000  2015-10-23 16:13:08.492999936
```

```
x                 2015-10-28 19:43:31.164000    2015-10-28 19:43:31.164000
                2015-10-24 23:08:46.593000  2015-10-24 23:08:46.592999936

                                                       \
                                             50%
age
0                   2015-10-30 05:41:53.620000
0.62                2015-10-28 19:49:06.308000
5                 2015-10-24 19:28:30.211000064
6                 2015-10-26 00:37:41.600499968
7                 2015-10-27 10:44:31.237999872
…                                           …
too old           2015-10-28 18:18:36.985999872
too old for this    2015-10-24 13:25:54.240000
very              2015-10-24 12:35:56.211000064
x                   2015-10-28 19:43:31.164000
                2015-10-24 23:08:46.592999936


                                             75%                           max
age
0                   2015-10-30 05:41:53.620000  2015-10-30 05:41:53.620000
0.62                2015-10-28 19:49:06.308000  2015-10-28 19:49:06.308000
5                 2015-10-24 19:28:30.211000064  2015-10-24 19:28:30.211000
6                 2015-10-28 14:52:50.038499840  2015-10-28 15:59:31.635000
7                 2015-10-28 15:37:14.411500032  2015-10-29 20:29:57.585000
…                                           …                          …
too old           2015-10-29 05:33:50.965250048  2015-10-29 17:26:57.566000
too old for this    2015-10-24 13:25:54.240000  2015-10-24 13:25:54.240000
very              2015-10-25 07:45:08.985999872  2015-10-29 08:51:45.438000
x                   2015-10-28 19:43:31.164000  2015-10-28 19:43:31.164000
                2015-10-24 23:08:46.592999936  2015-10-24 23:08:46.593000

[146 rows x 7 columns]
```

[22]:
```
#Cross Tabulation
from io import StringIO
df.head()
```

[22]:
```
                Timestamp age  \
0 2015-10-23 08:46:20.451  35
1 2015-10-23 08:46:51.583  41
2 2015-10-23 08:47:34.285  33
3 2015-10-23 08:47:58.964  31
4 2015-10-23 08:48:11.719  30

  Are_you_going_actually_going_trick_or_treating_yourself?  \
0                                                 No
1                                                 No
2                                                 No
3                                                 No
4                                                 No

  _Any_full-sized_candy_bar _100_Grand_Bar
0                       JOY            NaN
1                       JOY            JOY
2                       JOY        DESPAIR
3                       JOY            JOY
4                       JOY            JOY
```

[23]:
```
pd.crosstab(df["age"],df["Timestamp"],margins=True)
```

[23]:
```
Timestamp         2015-10-23 08:46:20.451000  2015-10-23 08:46:51.583000  \
age
0                                          0                           0
0.62                                       0                           0
5                                          0                           0
6                                          0                           0
7                                          0                           0
…                                        …                           …
too old for this                           0                           0
very                                       0                           0
x                                          0                           0
                                         0                           0
All                                        1                           1

Timestamp         2015-10-23 08:47:34.285000  2015-10-23 08:47:58.964000  \
age
0                                          0                           0
0.62                                       0                           0
5                                          0                           0
6                                          0                           0
7                                          0                           0
…                                        …                           …
too old for this                           0                           0
very                                       0                           0
x                                          0                           0
                                         0                           0
All                                        1                           1

Timestamp         2015-10-23 08:48:11.719000  2015-10-23 08:49:06.808000  \
age
0                                          0                           0
0.62                                       0                           0
```

```
5                                   0                          0
6                                   0                          0
7                                   0                          0
…                                ...                        ...
too old for this                    0                          0
very                                0                          0
x                                   0                          0
                                   0                          0
All                                 1                          1

Timestamp          2015-10-23 08:50:08.918000  2015-10-23 08:52:14.267000  \
age
0                                   0                          0
0.62                                0                          0
5                                   0                          0
6                                   0                          0
7                                   0                          0
…                                ...                        ...
too old for this                    0                          0
very                                0                          0
x                                   0                          0
                                   0                          0
All                                 1                          1

Timestamp          2015-10-23 08:52:22.112000  2015-10-23 08:53:30.967000  …  \
age                                                                        …
0                                   0                          0  …
0.62                                0                          0  …
5                                   0                          0  …
6                                   0                          0  …
7                                   0                          0  …
…                                ...                        ...  …
too old for this                    0                          0  …
very                                0                          0  …
x                                   0                          0  …
                                   0                          0  …
All                                 1                          1  …

Timestamp          2015-10-31 00:50:36.363000  2015-10-31 02:14:53.178000  \
age
0                                   0                          0
0.62                                0                          0
5                                   0                          0
6                                   0                          0
7                                   0                          0
…                                ...                        ...
too old for this                    0                          0
```

```
very                                  0                          0
x                                     0                          0
                                    0                          0
All                                   1                          1

Timestamp         2015-10-31 05:15:32.494000  2015-10-31 05:21:54.299000  \
age
0                                     0                          0
0.62                                  0                          0
5                                     0                          0
6                                     0                          0
7                                     0                          0
…                                   …                          …
too old for this                      0                          0
very                                  0                          0
x                                     0                          0
                                    0                          0
All                                   1                          1

Timestamp         2015-10-31 05:23:40.526000  2015-10-31 05:29:26.937000  \
age
0                                     0                          0
0.62                                  0                          0
5                                     0                          0
6                                     0                          0
7                                     0                          0
…                                   …                          …
too old for this                      0                          0
very                                  0                          0
x                                     0                          0
                                    0                          0
All                                   1                          1

Timestamp         2015-10-31 06:13:29.083000  2015-10-31 06:26:52.566000  \
age
0                                     0                          0
0.62                                  0                          0
5                                     0                          0
6                                     0                          0
7                                     0                          0
…                                   …                          …
too old for this                      0                          0
very                                  0                          0
x                                     0                          0
                                    0                          0
All                                   1                          1
```

```
Timestamp         2015-10-31 06:41:31.904000   All
age
0                                          0     1
0.62                                       0     1
5                                          0     1
6                                          0     4
7                                          0     2
…                                      …   …
too old for this                           0     1
very                                       0     5
x                                          0     1
                                         0     1
All                                        1  5431

[147 rows x 5430 columns]
```

[24]:
```
# •         Chapter 11
# o         Convert between string and date time
# o         Generate date range
# o         Frequencies and date offsets
# o         Convert timestamps to periods and back
# o         Period Frequency conversions
```

[25]:
```
# Frequencies and Date Offsets
```

[26]:
```
# Add 7-day offset to Timestamp
df["Timestamp_Offset"] = df["Timestamp"] + pd.DateOffset(days=7)
print("Added 7-day offset to Timestamp:")
print(df)
```

```
Added 7-day offset to Timestamp:
                   Timestamp age  \
0    2015-10-23 08:46:20.451  35
1    2015-10-23 08:46:51.583  41
2    2015-10-23 08:47:34.285  33
3    2015-10-23 08:47:58.964  31
4    2015-10-23 08:48:11.719  30
…                        …  ..
5625 2015-10-31 05:23:40.526  50
5626 2015-10-31 05:29:26.937  43
5627 2015-10-31 06:13:29.083  35
5628 2015-10-31 06:26:52.566  38
5629 2015-10-31 06:41:31.904  44

     Are_you_going_actually_going_trick_or_treating_yourself?  \
0                                                    No
1                                                    No
2                                                    No
```

```
3                                                     No
4                                                     No
…                                                  …
5625                                                  No
5626                                                  No
5627                                                 Yes
5628                                                  No
5629                                                  No

     _Any_full-sized_candy_bar _100_Grand_Bar        Timestamp_Offset
0                          JOY            NaN 2015-10-30 08:46:20.451
1                          JOY            JOY 2015-10-30 08:46:51.583
2                          JOY        DESPAIR 2015-10-30 08:47:34.285
3                          JOY            JOY 2015-10-30 08:47:58.964
4                          JOY            JOY 2015-10-30 08:48:11.719
…                         …             …                      …
5625                       JOY        DESPAIR 2015-11-07 05:23:40.526
5626                       JOY            JOY 2015-11-07 05:29:26.937
5627                       JOY            JOY 2015-11-07 06:13:29.083
5628                       JOY            JOY 2015-11-07 06:26:52.566
5629                       JOY            JOY 2015-11-07 06:41:31.904

[5630 rows x 6 columns]
```

[27]:
```
# Convert Timestamps to Periods and Back
```

[28]:
```
# Convert Timestamp to Period
df["Month_Period"] = df["Timestamp"].dt.to_period("M")

# Convert back to Timestamp
df["Month_Start"] = df["Month_Period"].dt.to_timestamp()
print("\nConverted Timestamp to period and back:")
print(df["Month_Period"])
```

```
Converted Timestamp to period and back:
0       2015-10
1       2015-10
2       2015-10
3       2015-10
4       2015-10
         …
5625    2015-10
5626    2015-10
5627    2015-10
5628    2015-10
5629    2015-10
Name: Month_Period, Length: 5630, dtype: period[M]
```