Assignment_week_7 & 8_Raghuwanshi_Prashant_DSC540

October 23, 2021

[1]:
```
##### Assignment: Week 7 & Week 8 Exercise, Understanding Packages
##### Name: Prashant Raghuwanshi
##### Date: 10/21/2021
##### Course: DSC540-T301 Data Preparation (2221-1)
```

[2]:
```
# Import libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Chapter 7

[66]:
```
# read MetObjects source file into dataframe- for chapter 7
MetObjects_df = pd.read_csv("C:/Users/dell/Documents/docker/Weeks7&8Data/
 ↪MetObjects1.csv",sep=",")
MetObjects_df.head(5)
```

[66]:

	Object Number	Is Highlight	Is Public Domain	Object ID	Department	\
0	1979.486.1	FALSE	FALSE	1	The American Wing	
1	1980.264.5	FALSE	FALSE	2	The American Wing	
2	67.265.9	FALSE	FALSE	3	The American Wing	
3	67.265.10	FALSE	FALSE	4	The American Wing	
4	67.265.11	FALSE	FALSE	5	The American Wing	

	Object Name	Title	Culture	Period	Dynasty	…	\
0	Coin	One-dollar Liberty Head Coin	NaN	NaN	NaN	…	
1	Coin	Ten-dollar Liberty Head Coin	NaN	NaN	NaN	…	
2	Coin	Two-and-a-Half Dollar Coin	NaN	NaN	NaN	…	
3	Coin	Two-and-a-Half Dollar Coin	NaN	NaN	NaN	…	
4	Coin	Two-and-a-Half Dollar Coin	NaN	NaN	NaN	…	

	Locale	Locus	Excavation	River	Classification	Rights and Reproduction	\
0	NaN	NaN	NaN	NaN	Metal	NaN	
1	NaN	NaN	NaN	NaN	Metal	NaN	
2	NaN	NaN	NaN	NaN	Metal	NaN	
3	NaN	NaN	NaN	NaN	Metal	NaN	
4	NaN	NaN	NaN	NaN	Metal	NaN	

```
                                        Link Resource Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1           NaN
1  http://www.metmuseum.org/art/collection/search/2           NaN
2  http://www.metmuseum.org/art/collection/search/3           NaN
3  http://www.metmuseum.org/art/collection/search/4           NaN
4  http://www.metmuseum.org/art/collection/search/5           NaN

                                  Repository Tags
0  Metropolitan Museum of Art, New York, NY  NaN
1  Metropolitan Museum of Art, New York, NY  NaN
2  Metropolitan Museum of Art, New York, NY  NaN
3  Metropolitan Museum of Art, New York, NY  NaN
4  Metropolitan Museum of Art, New York, NY  NaN

[5 rows x 44 columns]
```

[67]:
```
# Filter out missing data
MetObjects_df_filter = MetObjects_df.dropna()
MetObjects_df_filter.head()
```

[67]:
```
Empty DataFrame
Columns: [Object Number, Is Highlight, Is Public Domain, Object ID, Department,
Object Name, Title, Culture, Period, Dynasty, Reign, Portfolio, Artist Role,
Artist Prefix, Artist Display Name, Artist Display Bio, Artist Suffix, Artist
Alpha Sort, Artist Nationality, Artist Begin Date, Artist End Date, Object Date,
Object Begin Date, Object End Date, Medium, Dimensions, Credit Line, Geography
Type, City, State, County, Country, Region, Subregion, Locale, Locus,
Excavation, River, Classification, Rights and Reproduction, Link Resource,
Metadata Date, Repository, Tags]
Index: []

[0 rows x 44 columns]
```

[68]:
```
# filter the records which contans all N.A
MetObjects_df_filter_na_only = MetObjects_df.dropna(how='all')
MetObjects_df_filter_na_only.head()
```

[68]:
```
  Object Number Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1        FALSE            FALSE         1  The American Wing
1    1980.264.5        FALSE            FALSE         2  The American Wing
2      67.265.9        FALSE            FALSE         3  The American Wing
3     67.265.10        FALSE            FALSE         4  The American Wing
4     67.265.11        FALSE            FALSE         5  The American Wing

  Object Name                          Title Culture Period Dynasty  …  \
0        Coin  One-dollar Liberty Head Coin     NaN    NaN     NaN  …
1        Coin  Ten-dollar Liberty Head Coin     NaN    NaN     NaN  …
```

```
2        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
3        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
4        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …

  Locale Locus Excavation River Classification Rights and Reproduction  \
0    NaN   NaN        NaN   NaN          Metal                      NaN
1    NaN   NaN        NaN   NaN          Metal                      NaN
2    NaN   NaN        NaN   NaN          Metal                      NaN
3    NaN   NaN        NaN   NaN          Metal                      NaN
4    NaN   NaN        NaN   NaN          Metal                      NaN

                                      Link Resource Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1           NaN
1  http://www.metmuseum.org/art/collection/search/2           NaN
2  http://www.metmuseum.org/art/collection/search/3           NaN
3  http://www.metmuseum.org/art/collection/search/4           NaN
4  http://www.metmuseum.org/art/collection/search/5           NaN

                                Repository Tags
0  Metropolitan Museum of Art, New York, NY  NaN
1  Metropolitan Museum of Art, New York, NY  NaN
2  Metropolitan Museum of Art, New York, NY  NaN
3  Metropolitan Museum of Art, New York, NY  NaN
4  Metropolitan Museum of Art, New York, NY  NaN

[5 rows x 44 columns]
```

[69]:
```
#Fill in missing data fill in the NaN values by using ffill
MetObjects_df_ffill = MetObjects_df_filter_na_only.fillna(method='ffill')
MetObjects_df_ffill.head()
```

[69]:
```
  Object Number Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1        FALSE            FALSE         1  The American Wing
1    1980.264.5        FALSE            FALSE         2  The American Wing
2      67.265.9        FALSE            FALSE         3  The American Wing
3     67.265.10        FALSE            FALSE         4  The American Wing
4     67.265.11        FALSE            FALSE         5  The American Wing

  Object Name                         Title Culture Period Dynasty  …  \
0        Coin  One-dollar Liberty Head Coin     NaN    NaN     NaN  …
1        Coin  Ten-dollar Liberty Head Coin     NaN    NaN     NaN  …
2        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
3        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
4        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …

  Locale Locus Excavation River Classification Rights and Reproduction  \
0    NaN   NaN        NaN   NaN          Metal                      NaN
```

```
1    NaN   NaN        NaN   NaN          Metal                      NaN
2    NaN   NaN        NaN   NaN          Metal                      NaN
3    NaN   NaN        NaN   NaN          Metal                      NaN
4    NaN   NaN        NaN   NaN          Metal                      NaN

                                    Link Resource Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1           NaN
1  http://www.metmuseum.org/art/collection/search/2           NaN
2  http://www.metmuseum.org/art/collection/search/3           NaN
3  http://www.metmuseum.org/art/collection/search/4           NaN
4  http://www.metmuseum.org/art/collection/search/5           NaN

                                Repository Tags
0  Metropolitan Museum of Art, New York, NY  NaN
1  Metropolitan Museum of Art, New York, NY  NaN
2  Metropolitan Museum of Art, New York, NY  NaN
3  Metropolitan Museum of Art, New York, NY  NaN
4  Metropolitan Museum of Art, New York, NY  NaN

[5 rows x 44 columns]
```

[60]:
```
#Fill in missing data fill in the NaN values by using bfill
MetObjects_df_bfill = MetObjects_df_ffill.fillna(method='bfill')
MetObjects_df_bfill.head()
```

[60]:
```
  Object Number Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1        FALSE            FALSE         1  The American Wing
1    1980.264.5        FALSE            FALSE         2  The American Wing
2      67.265.9        FALSE            FALSE         3  The American Wing
3     67.265.10        FALSE            FALSE         4  The American Wing
4     67.265.11        FALSE            FALSE         5  The American Wing

  Object Name                        Title  Culture         Period Dynasty  \
0        Coin  One-dollar Liberty Head Coin  Mexican  United States    1850
1        Coin  Ten-dollar Liberty Head Coin  Mexican  United States    1850
2        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850
3        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850
4        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850

   …        Locale Locus                                          Excavation  \
0  …  Architecture   NaN  http://www.metmuseum.org/art/collection/search…
1  …  Architecture   NaN  http://www.metmuseum.org/art/collection/search…
2  …  Architecture   NaN  http://www.metmuseum.org/art/collection/search…
3  …  Architecture   NaN  http://www.metmuseum.org/art/collection/search…
4  …  Architecture   NaN  http://www.metmuseum.org/art/collection/search…

  River Classification                        Rights and Reproduction  \
```

```
0   NaN          Metal    2019 Artists Rights Society (ARS), New York
1   NaN          Metal    2019 Artists Rights Society (ARS), New York
2   NaN          Metal    2019 Artists Rights Society (ARS), New York
3   NaN          Metal    2019 Artists Rights Society (ARS), New York
4   NaN          Metal    2019 Artists Rights Society (ARS), New York

                                      Link Resource             Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1  2019-02-01T10:50:49.477Z
1  http://www.metmuseum.org/art/collection/search/2  2019-02-01T10:50:49.477Z
2  http://www.metmuseum.org/art/collection/search/3  2019-02-01T10:50:49.477Z
3  http://www.metmuseum.org/art/collection/search/4  2019-02-01T10:50:49.477Z
4  http://www.metmuseum.org/art/collection/search/5  2019-02-01T10:50:49.477Z

                                Repository         Tags
0  Metropolitan Museum of Art, New York, NY  Birds|Coins
1  Metropolitan Museum of Art, New York, NY  Birds|Coins
2  Metropolitan Museum of Art, New York, NY  Birds|Coins
3  Metropolitan Museum of Art, New York, NY  Birds|Coins
4  Metropolitan Museum of Art, New York, NY  Birds|Coins

[5 rows x 44 columns]
```

[70]:
```
#Fill in missing data fill in the NaN values by 0
MetObjects_df_filldata = MetObjects_df_bfill.fillna(0)
MetObjects_df_filldata.head()
```

[70]:
```
  Object Number Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1        FALSE            FALSE         1  The American Wing
1    1980.264.5        FALSE            FALSE         2  The American Wing
2      67.265.9        FALSE            FALSE         3  The American Wing
3     67.265.10        FALSE            FALSE         4  The American Wing
4     67.265.11        FALSE            FALSE         5  The American Wing

  Object Name                         Title  Culture         Period Dynasty  \
0        Coin  One-dollar Liberty Head Coin  Mexican  United States    1850
1        Coin  Ten-dollar Liberty Head Coin  Mexican  United States    1850
2        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850
3        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850
4        Coin    Two-and-a-Half Dollar Coin  Mexican  United States    1850

   …        Locale Locus                                          Excavation  \
0  …  Architecture   0.0  http://www.metmuseum.org/art/collection/search…
1  …  Architecture   0.0  http://www.metmuseum.org/art/collection/search…
2  …  Architecture   0.0  http://www.metmuseum.org/art/collection/search…
3  …  Architecture   0.0  http://www.metmuseum.org/art/collection/search…
4  …  Architecture   0.0  http://www.metmuseum.org/art/collection/search…
```

```
  River Classification                          Rights and Reproduction  \
0   0.0          Metal    2019 Artists Rights Society (ARS), New York
1   0.0          Metal    2019 Artists Rights Society (ARS), New York
2   0.0          Metal    2019 Artists Rights Society (ARS), New York
3   0.0          Metal    2019 Artists Rights Society (ARS), New York
4   0.0          Metal    2019 Artists Rights Society (ARS), New York

                                      Link Resource             Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1  2019-02-01T10:50:49.477Z
1  http://www.metmuseum.org/art/collection/search/2  2019-02-01T10:50:49.477Z
2  http://www.metmuseum.org/art/collection/search/3  2019-02-01T10:50:49.477Z
3  http://www.metmuseum.org/art/collection/search/4  2019-02-01T10:50:49.477Z
4  http://www.metmuseum.org/art/collection/search/5  2019-02-01T10:50:49.477Z

                                 Repository         Tags
0  Metropolitan Museum of Art, New York, NY  Birds|Coins
1  Metropolitan Museum of Art, New York, NY  Birds|Coins
2  Metropolitan Museum of Art, New York, NY  Birds|Coins
3  Metropolitan Museum of Art, New York, NY  Birds|Coins
4  Metropolitan Museum of Art, New York, NY  Birds|Coins

[5 rows x 44 columns]
```

[62]:
```
# Remove duplicates based on second and third columns value
MetObjects_df_rm_dup = MetObjects_df_filldata.drop_duplicates(['Department',␣
↪'Object Name', 'Title', 'Culture'], keep='last')
MetObjects_df_rm_dup.head()
```

[62]:
```
   Object Number Is Highlight Is Public Domain Object ID         Department  \
0     1979.486.1        FALSE            FALSE         1  The American Wing
1     1980.264.5        FALSE            FALSE         2  The American Wing
8      67.265.15        FALSE            FALSE         9  The American Wing
9     1979.486.3        FALSE            FALSE        10  The American Wing
10    1979.486.2        FALSE            FALSE        11  The American Wing

   Object Name                                    Title  Culture  \
0         Coin             One-dollar Liberty Head Coin  Mexican
1         Coin             Ten-dollar Liberty Head Coin  Mexican
8         Coin               Two-and-a-Half Dollar Coin  Mexican
9         Coin   Two-and-a-half-dollar Indian Head Coin  Mexican
10        Coin  Two-and-a-half-dollar Liberty Head Coin  Mexican

           Period Dynasty  …        Locale Locus  \
0   United States    1850  …  Architecture   0.0
1   United States    1850  …  Architecture   0.0
8   United States    1850  …  Architecture   0.0
9   United States    1850  …  Architecture   0.0
```

```
10  United States    1850  …  Architecture   0.0

                                          Excavation River Classification  \
0   http://www.metmuseum.org/art/collection/search…   0.0          Metal
1   http://www.metmuseum.org/art/collection/search…   0.0          Metal
8   http://www.metmuseum.org/art/collection/search…   0.0          Metal
9   http://www.metmuseum.org/art/collection/search…   0.0          Metal
10  http://www.metmuseum.org/art/collection/search…   0.0          Metal

                         Rights and Reproduction  \
0    2019 Artists Rights Society (ARS), New York
1    2019 Artists Rights Society (ARS), New York
8    2019 Artists Rights Society (ARS), New York
9    2019 Artists Rights Society (ARS), New York
10   2019 Artists Rights Society (ARS), New York

                                    Link Resource  \
0    http://www.metmuseum.org/art/collection/search/1
1    http://www.metmuseum.org/art/collection/search/2
8    http://www.metmuseum.org/art/collection/search/9
9   http://www.metmuseum.org/art/collection/search/10
10  http://www.metmuseum.org/art/collection/search/11

               Metadata Date                                Repository  \
0   2019-02-01T10:50:49.477Z  Metropolitan Museum of Art, New York, NY
1   2019-02-01T10:50:49.477Z  Metropolitan Museum of Art, New York, NY
8   2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY
9   2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY
10  2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY

                  Tags
0          Birds|Coins
1          Birds|Coins
8   Eagles|Men|Profiles
9   Eagles|Men|Profiles
10  Eagles|Men|Profiles

[5 rows x 44 columns]
```

[71]:
```
#Transform data using either mapping or a function
y_n_to_num = {'TRUE':1, 'FALSE': 0}
MetObjects_df_rm_dup['Is Highlight'] = MetObjects_df_rm_dup['Is Highlight'].
 ↪map(y_n_to_num)
```

[65]:
```
MetObjects_df_rm_dup.head()
```

[65]:
```
  Object Number  Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1           0.0            FALSE         1  The American Wing
1    1980.264.5           0.0            FALSE         2  The American Wing
2      67.265.9           0.0            FALSE         3  The American Wing
3     67.265.10           0.0            FALSE         4  The American Wing
4     67.265.11           0.0            FALSE         5  The American Wing

  Object Name                          Title Culture Period Dynasty  …  \
0        Coin  One-dollar Liberty Head Coin     NaN    NaN     NaN  …
1        Coin  Ten-dollar Liberty Head Coin     NaN    NaN     NaN  …
2        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
3        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …
4        Coin    Two-and-a-Half Dollar Coin     NaN    NaN     NaN  …

  Locale Locus Excavation River Classification Rights and Reproduction  \
0    NaN   NaN        NaN   NaN          Metal                      NaN
1    NaN   NaN        NaN   NaN          Metal                      NaN
2    NaN   NaN        NaN   NaN          Metal                      NaN
3    NaN   NaN        NaN   NaN          Metal                      NaN
4    NaN   NaN        NaN   NaN          Metal                      NaN

                                    Link Resource Metadata Date  \
0  http://www.metmuseum.org/art/collection/search/1           NaN
1  http://www.metmuseum.org/art/collection/search/2           NaN
2  http://www.metmuseum.org/art/collection/search/3           NaN
3  http://www.metmuseum.org/art/collection/search/4           NaN
4  http://www.metmuseum.org/art/collection/search/5           NaN

                               Repository Tags
0  Metropolitan Museum of Art, New York, NY  NaN
1  Metropolitan Museum of Art, New York, NY  NaN
2  Metropolitan Museum of Art, New York, NY  NaN
3  Metropolitan Museum of Art, New York, NY  NaN
4  Metropolitan Museum of Art, New York, NY  NaN

[5 rows x 44 columns]
```

[72]:
```
#Replace values
MetObjects_df_rm_dup['Is Public Domain'] = MetObjects_df_rm_dup['Is Public␣
 ↪Domain'].replace(['TRUE','FALSE'],[1, 0])
MetObjects_df_rm_dup.head()
```

[72]:
```
   Object Number  Is Highlight Is Public Domain Object ID         Department  \
0    1979.486.1           0.0                 0         1  The American Wing
1    1980.264.5           0.0                 0         2  The American Wing
8     67.265.15           0.0                 0         9  The American Wing
9    1979.486.3           0.0                 0        10  The American Wing
```

```
10    1979.486.2          0.0                0        11  The American Wing

   Object Name                                    Title  Culture  \
0         Coin             One-dollar Liberty Head Coin  Mexican
1         Coin             Ten-dollar Liberty Head Coin  Mexican
8         Coin               Two-and-a-Half Dollar Coin  Mexican
9         Coin   Two-and-a-half-dollar Indian Head Coin  Mexican
10        Coin  Two-and-a-half-dollar Liberty Head Coin  Mexican

           Period Dynasty  …        Locale Locus  \
0   United States    1850  …  Architecture   0.0
1   United States    1850  …  Architecture   0.0
8   United States    1850  …  Architecture   0.0
9   United States    1850  …  Architecture   0.0
10  United States    1850  …  Architecture   0.0

                                          Excavation River Classification  \
0   http://www.metmuseum.org/art/collection/search…   0.0          Metal
1   http://www.metmuseum.org/art/collection/search…   0.0          Metal
8   http://www.metmuseum.org/art/collection/search…   0.0          Metal
9   http://www.metmuseum.org/art/collection/search…   0.0          Metal
10  http://www.metmuseum.org/art/collection/search…   0.0          Metal

                         Rights and Reproduction  \
0    2019 Artists Rights Society (ARS), New York
1    2019 Artists Rights Society (ARS), New York
8    2019 Artists Rights Society (ARS), New York
9    2019 Artists Rights Society (ARS), New York
10   2019 Artists Rights Society (ARS), New York

                                      Link Resource  \
0    http://www.metmuseum.org/art/collection/search/1
1    http://www.metmuseum.org/art/collection/search/2
8    http://www.metmuseum.org/art/collection/search/9
9   http://www.metmuseum.org/art/collection/search/10
10  http://www.metmuseum.org/art/collection/search/11

              Metadata Date                                 Repository  \
0   2019-02-01T10:50:49.477Z  Metropolitan Museum of Art, New York, NY
1   2019-02-01T10:50:49.477Z  Metropolitan Museum of Art, New York, NY
8   2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY
9   2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY
10  2019-07-31T03:00:40.447Z  Metropolitan Museum of Art, New York, NY

                  Tags
0           Birds|Coins
1           Birds|Coins
```

```
8   Eagles|Men|Profiles
9   Eagles|Men|Profiles
10  Eagles|Men|Profiles

[5 rows x 44 columns]
```

[84]:
```
# Drop bad data from Object ID columns
MetObjects_df_rm_dup = MetObjects_df_rm_dup[pd.
 ↪to_numeric(MetObjects_df_rm_dup['Object ID'], errors='coerce').notnull()]
```

[88]:
```
MetObjects_df_rm_dup['Object ID'].unique()
```

[88]:
```
array(['1', '2', '9', …, '11815', '11816', '11817'], dtype=object)
```

[92]:
```
# Discretization and Binning
bins = [2000,4000,8000,10000,11817]
ObjectID_bin = pd.cut(pd.to_numeric(MetObjects_df_rm_dup['Object ID']), bins)
```

[93]:
```
pd.value_counts(ObjectID_bin)
```

[93]:
```
(10000, 11817]    1456
(4000, 8000]       882
(2000, 4000]       499
(8000, 10000]      302
Name: Object ID, dtype: int64
```

Chapter 8

[16]:
```
# read BOING-CANDY-HIERARCHY-2016-SURVEY-Responses source file into dataframe␣
 ↪for chapter 8
boing_candy_df = pd.read_excel (r"C:/Users/dell/Documents/docker/Weeks7&8Data/
 ↪BOING-BOING-CANDY-HIERARCHY-2016-SURVEY-Responses.xlsx")
# display firt 5 records
boing_candy_df.head(5)
```

[16]:
```
                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

  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
```

	How old are you?	Which country do you live in?	\
0	22	Canada	
1	45	usa	
2	48	US	
3	57	usa	
4	42	USA	

	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	

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

	[Any full-sized candy bar]	[Black Jacks]	…	\
0	JOY	MEH	…	
1	JOY	JOY	…	
2	JOY	MEH	…	
3	JOY	MEH	…	
4	JOY	DESPAIR	…	

	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	

	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	

	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

  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

  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

  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

  Which day do you prefer, Friday or Sunday?  \
0                                     Friday
1                                     Friday
2                                     Sunday
3                                     Sunday
4                                     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)?  \
0                                     South to North
1                                       East to West
2                                       East to West
3                                     South to North
4                                       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

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

[5 rows x 123 columns]
```

[17]:
```
#before doing hierarchical index , will quickly fix the missing data issue &
 ↪data quality issue as below
# filter the records which contans all N.A
boing_candy_df_filter_na_only = boing_candy_df.dropna(how='all')
#Fill in missing data fill in the NaN values by using ffill
boing_candy_df_ffill = boing_candy_df_filter_na_only.fillna(method='ffill')
#Fill in missing data fill in the NaN values by using ffill
boing_candy_df_bfill = boing_candy_df_ffill.fillna(method='bfill')
#Fill in missing data fill in the NaN values by 0
boing_candy_df_filldata = boing_candy_df_bfill.fillna(0)
```

[18]:
```
# transforming the countries name to country code
# find the list of unique countries present in country columns
unique_countries = boing_candy_df_filldata['Which country do you live in?'].
 ↪unique()
unique_countries
```

[18]:
```
array(['Canada', 'usa', 'US', 'USA', 'UK', 'United States of America',
       'uSA', 'Japan', 'united states', 'USA ', 'canada', 'United States',
       'us', 'france', 'USSA', 'United States of America ', 'U.S.A.',
       'A tropical island south of the equator', 'england', 'uk',
       'Switzerland', 'Murica', 'United Kingdom', 'Neverland', 'USA!',
       'this one',
       "USA (I think but it's an election year so who can really tell)",
       'Korea', 51, 'Usa', 'U.S.', 'Us', 'America ', 'Units States',
       'belgium', 'croatia', 'United states', 'Portugal', 'England',
       'USA USA USA', 'the best one - usa', 'USA! USA! USA!', 47,
       'united states ', 'Cascadia', 'españa', 'u.s.',
       "there isn't one for old men", 'United States ', 'Panama',
       'one of the best ones', 'The Yoo Ess of Aaayyyyyy',
       'United Kindom', 'France', 'America', 'Australia', 'hungary',
       'united states of america', 'UK ', 'Austria', 'Somewhere',
```

```
        'New Zealand', 54, 'Germany', 'Mexico', 44, 'Brasil',
        "god's country", 'South Korea', 'USA!!!!!!', 'Philippines',
        ' United States', 'EUA', 'USA! USA!', 45, 'sweden', 'Canada ',
        'United Sates', "Sub-Canadian North America… 'Merica",
        'The Netherlands', 'Finland', 'Trumpistan', 'U.s.', 'Merica',
        'China', 'germany', 'See above', 'UNited States', 'kenya', 30,
        'Netherlands', 'The republic of Cascadia ', 'United Stetes',
        'america', 'Not the USA or Canada', 'USA USA USA USA',
        'New Zealand ', 'United  States of America', 'netherlands',
        'Denial', 'United State'], dtype=object)
```

[19]:
```
# create function to verify the digits in string
def contains_digit(s):
    isdigit = str.isdigit
    return any(map(isdigit,str(s)))
```

[20]:
```
# create function to normalizing the countries name
import pycountry
def fix_country_name(txt):
    txt1 = ''
    if contains_digit(txt) == False:
        for country in pycountry.countries:
            if country.name in txt:
                #Country's code
                return(country.alpha_2)
        return('US')
```

[21]:
```
# transforming the country name to country code
boing_candy_df_filldata['Which country do you live in?'] =␣
↪boing_candy_df_filldata['Which country do you live in?'].apply(lambda x:␣
↪fix_country_name(x))
```

[22]:
```
# list out unique transforms country codes
boing_candy_df_filldata['Which country do you live in?'].unique()
```

[22]:
```
array(['CA', 'US', 'JP', 'CH', 'GB', None, 'PT', 'PA', 'FR', 'AU', 'AT',
       'NZ', 'DE', 'MX', 'PH', 'NL', 'FI', 'CN'], dtype=object)
```

[23]:
```
#Hierarchical Indexing
hir_boing_candy_df = boing_candy_df_filldata.set_index(['Which country do you␣
↪live in?', 'Which state, province, county do you live in?'])
```

[24]:
```
hir_boing_candy_df.head()
```

[24]:
```
          Timestamp  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
```

```
2016-10-24 05:09:23.033
US                              il
2016-10-24 05:09:54.798
                                Colorado
2016-10-24 05:13:06.734
                                il
2016-10-24 05:14:17.192
                                South Dakota
2016-10-24 05:14:24.625

                                                                                Are
you going actually going trick or treating yourself?  \
Which country do you live in? Which state, province, county do you live in?
CA                              Ontario
No
US                              il
No
                                Colorado
No
                                il
No
                                South Dakota
Yes

                                                                                Your
gender:  \
Which country do you live in? Which state, province, county do you live in?
CA                              Ontario
Male
US                              il
Male
                                Colorado
Female
                                il
Male
                                South Dakota
Male

                                                                                How
old are you?  \
Which country do you live in? Which state, province, county do you live in?
CA                              Ontario
22
US                              il
45
                                Colorado
48
```

```
                              il
57
                              South Dakota
42

[100 Grand Bar]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
JOY
US                            il
MEH
                              Colorado
JOY
                              il
JOY
                              South Dakota
MEH

[Anonymous brown globs that come in black and orange wrappers]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
DESPAIR
US                            il
MEH
                              Colorado
DESPAIR
                              il
MEH
                              South Dakota
DESPAIR

[Any full-sized candy bar]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
JOY
US                            il
JOY
                              Colorado
JOY
                              il
JOY
                              South Dakota
JOY

[Black Jacks]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
```

```
MEH
US                              il
JOY
                                Colorado
MEH
                                il
MEH
                                South Dakota
DESPAIR

[Bonkers (the candy)]  \
Which country do you live in? Which state, province, county do you live in?
CA                              Ontario
MEH
US                              il
DESPAIR
                                Colorado
MEH
                                il
MEH
                                South Dakota
MEH

[Bonkers (the board game)]  \
Which country do you live in? Which state, province, county do you live in?
CA                              Ontario
MEH
US                              il
MEH
                                Colorado
JOY
                                il
DESPAIR
                                South Dakota
JOY


                                                                                …
\
Which country do you live in? Which state, province, county do you live in?  …
CA                              Ontario                                      …
US                              il                                           …
                                Colorado                                     …
                                il                                           …
                                South Dakota                                 …

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

```
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
3 or higher
US                            il
3 or higher
                              Colorado
3 or higher
                              il
3 or higher
                              South Dakota
3 or higher

Please estimate the degree(s) of separation you have from the following
celebrities [JJ Abrams]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
2
US                            il
3 or higher
                              Colorado
3 or higher
                              il
3 or higher
                              South Dakota
3 or higher

Please estimate the degree(s) of separation you have from the following
celebrities [Beyoncé]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
3 or higher
US                            il
3 or higher
                              Colorado
3 or higher
                              il
3 or higher
                              South Dakota
3 or higher

Please estimate the degree(s) of separation you have from the following
celebrities [Bieber]  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
3 or higher
US                            il
3 or higher
```

```
                            Colorado
3 or higher
                            il
3 or higher
                            South Dakota
3 or higher

Please estimate the degree(s) of separation you have from the following
celebrities [Kevin Bacon]  \
Which country do you live in? Which state, province, county do you live in?
CA                          Ontario
3 or higher
US                          il
3 or higher
                            Colorado
3 or higher
                            il
3 or higher
                            South Dakota
3 or higher

Please estimate the degree(s) of separation you have from the following
celebrities [Francis Bacon (1561 - 1626)]  \
Which country do you live in? Which state, province, county do you live in?
CA                          Ontario
3 or higher
US                          il
3 or higher
                            Colorado
3 or higher
                            il
3 or higher
                            South Dakota
3 or higher

Which day do you prefer, Friday or Sunday?  \
Which country do you live in? Which state, province, county do you live in?
CA                          Ontario
Friday
US                          il
Friday
                            Colorado
Sunday
                            il
Sunday
                            South Dakota
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)?  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
South to North
US                            il
East to West
                              Colorado
East to West
                              il
South to North
                              South Dakota
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).  \
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
Science: Latest News and Headlines
US                            il
Science: Latest News and Headlines
                              Colorado
Science: Latest News and Headlines
                              il
Science: Latest News and Headlines
                              South Dakota
ESPN

[York Peppermint Patties] Ignore
Which country do you live in? Which state, province, county do you live in?
CA                            Ontario
0.0
US                            il
0.0
                              Colorado
0.0
                              il
0.0
                              South Dakota
0.0

[5 rows x 121 columns]
```

[25]:
```
# 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)
#creating new dataframe contains country code:
df_country_cd = pd.DataFrame({'Which country do you live in?': ['CA', 'US',␣
↪'JP', 'CH', 'GB', None, 'PT', 'PA', 'FR', 'AU', 'AT',
       'NZ', 'DE', 'MX', 'PH', 'NL', 'FI', 'CN'], 'cnt_code': range(18)})
df_country_cd.head()
```

[25]:

	Which country do you live in?	cnt_code
0	CA	0
1	US	1
2	JP	2
3	CH	3
4	GB	4

[26]:
```
#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)
# Merge Datasets
pd.merge(df_country_cd, hir_boing_candy_df, on='Which country do you live in?').
↪tail()
```

[26]:

	Which country do you live in?	cnt_code	Timestamp	Are you going actually going trick or treating yourself?	Your gender:	How old are you?	[100 Grand Bar]	[Anonymous brown globs that come in black and orange wrappers]
1254	PH	14	2016-10-24 20:52:29.673	No	Male	44	MEH	DESPAIR
1255	NL	15	2016-10-27 04:21:22.486	No	Male	25	MEH	MEH
1256	NL	15	2016-10-27 11:56:38.098	No	Male	50+	MEH	MEH
1257	FI	16	2016-10-27 05:33:18.091	No	Male	34	MEH	DESPAIR
1258	CN	17	2016-10-27 09:51:48.159	Yes	Other	67	DESPAIR	

```
1258                                                JOY

      [Any full-sized candy bar]  [Black Jacks]  …  \
1254                         MEH            MEH  …
1255                         JOY        DESPAIR  …
1256                         JOY            MEH  …
1257                         JOY            MEH  …
1258                         JOY            MEH  …

     Please estimate the degree(s) of separation you have from the following
celebrities [JK Rowling]  \
1254                                       3 or higher
1255                                       3 or higher
1256                                       3 or higher
1257                                       3 or higher
1258                                                 2

     Please estimate the degree(s) of separation you have from the following
celebrities [JJ Abrams]  \
1254                                       3 or higher
1255                                       3 or higher
1256                                       3 or higher
1257                                       3 or higher
1258                                       3 or higher

     Please estimate the degree(s) of separation you have from the following
celebrities [Beyoncé]  \
1254                                       3 or higher
1255                                       3 or higher
1256                                       3 or higher
1257                                       3 or higher
1258                                       3 or higher

     Please estimate the degree(s) of separation you have from the following
celebrities [Bieber]  \
1254                                       3 or higher
1255                                       3 or higher
1256                                       3 or higher
1257                                                 2
1258                                       3 or higher

     Please estimate the degree(s) of separation you have from the following
celebrities [Kevin Bacon]  \
1254                                       3 or higher
1255                                                 2
1256                                       3 or higher
1257                                       3 or higher
```

```
1258                                                  2

     Please estimate the degree(s) of separation you have from the following
celebrities [Francis Bacon (1561 - 1626)]  \
1254                                          3 or higher
1255                                          3 or higher
1256                                          3 or higher
1257                                 Actually, that's me.
1258                                                  2

     Which day do you prefer, Friday or Sunday?  \
1254                                        Friday
1255                                        Friday
1256                                        Sunday
1257                                        Friday
1258                                        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)?  \
1254                                         East to West
1255                               scoops with sharp spoon.
1256                                         East to West
1257                                         East to West
1258                                           Applesauce

     When you see the above image of the 4 different websites, which one would
you most likely check out (please be honest).  \
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
1254                                0.0
1255                                0.0
1256                                0.0
1257                                0.0
1258                                0.0

[5 rows x 123 columns]
```

[27]:
```
#Reshape
hir_boing_candy_df.stack().head()
```

[27]:
```
Which country do you live in?  Which state, province, county do you live in?
CA                             Ontario
```

```
Timestamp                                                    2016-10-24
05:09:23.033000
Are you going actually going trick or treating yourself?
No
Your gender:
Male
How old are you?
22
[100 Grand Bar]
JOY
dtype: object
```

[28]:
```
# Pivot the data
hir_boing_candy_df.pivot_table(index='Which country do you live in?',␣
 ↪columns='Your gender:', values='How old are you?', aggfunc=list)
```

[28]:
```
Your gender:                                                          Female
\
Which country do you live in?
AT                                                                       NaN
AU                                                                  [66, 45]
CA                              [44, 37, 38, 37, 58, 64, 37, 33, 47, 32, 29, 3…
CH                                                                      [35]
CN                                                                       NaN
DE                                                                  [43, 38]
FI                                                                       NaN
FR                                                                      [18]
GB                                                                  [31, 38]
JP                                                                       NaN
MX                                                                       NaN
NL                                                                       NaN
NZ                                                                       NaN
PA                                                                       NaN
PH                                                                       NaN
PT                                                                       NaN
US                              [48, 41, 46, 31, 33, 16, 51, 33, 37, 35, 32, 4…

Your gender:                                                I'd rather not say
\
Which country do you live in?
AT                                                                       NaN
AU                                                                [over 40]
CA                                              [older than I want to be, 47, 35]
CH                                                                       NaN
CN                                                                       NaN
DE                                                                       NaN
FI                                                                       NaN
```

```
FR                                                                  NaN
GB                                                                  NaN
JP                                                                  NaN
MX                                                                  NaN
NL                                                                  NaN
NZ                                                                 [34]
PA                                                                  NaN
PH                                                                  NaN
PT                                                                  NaN
US                          [31, 33, 35, 78, old enough, 50s, 63, 29, Old …

Your gender:                                                       Male
\
Which country do you live in?
AT                                                                 [33]
AU                                                                  NaN
CA                          [22, 28, 34, 58, 38, 48, 31, 42, 53, 17, 35, 3…
CH                                                                  NaN
CN                                                                  NaN
DE                                                             [47, 43]
FI                                                                 [34]
FR                                                                  NaN
GB                                                             [41, 27]
JP                                                          [45, 47, 34]
MX                                                          [54, 37, 38]
NL                                                             [25, 50+]
NZ                                                          [64, 49, 33]
PA                                                                 [47]
PH                                                                 [44]
PT                                                                 [58]
US                          [45, 57, 42, 41, 47, 44, 41, 40, 35, 49, 44, 4…

Your gender:                                                       Other
Which country do you live in?
AT                                                                  NaN
AU                                                                 [42]
CA                                                                  NaN
CH                                                                  NaN
CN                                                                 [67]
DE                                                                  NaN
FI                                                                  NaN
FR                                                                  NaN
GB                                                                  NaN
JP                                                                 [54]
MX                                                                  NaN
NL                                                                  NaN
NZ                                                                  NaN
```

```
PA                                                                    NaN
PH                                                                    NaN
PT                                                                    NaN
US                                [34, 23, 17, 28, 19, 32, 22, 26, 29, 26, 44, 4…
```

Chapter 10

[29]:
```
# read candyhierarchy2017.xlsx source file into dataframe- for chapter 10
candy_2017_df = pd.read_excel (r"C:/Users/dell/Documents/docker/Weeks7&8Data/
 ↪candyhierarchy2017.xlsx")
# display firt 5 records
candy_2017_df.head(5)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\openpyxl\worksheet\_reader.py:312:
UserWarning: Unknown extension is not supported and will be removed
  warn(msg)
```

[29]:

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY
0	90258773	NaN	NaN	NaN	NaN
1	90272821	No	Male	44	USA
2	90272829	NaN	Male	49	USA
3	90272840	No	Male	40	us
4	90272841	No	Male	23	usa

	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar
0	NaN	NaN
1	NM	MEH
2	Virginia	NaN
3	or	MEH
4	exton pa	JOY

	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)
0	NaN
1	DESPAIR
2	NaN
3	DESPAIR
4	DESPAIR

	Q6 | Any full-sized candy bar	Q6 | Black Jacks	…	Q8: DESPAIR OTHER
0	NaN	NaN	…	NaN
1	JOY	MEH	…	NaN
2	NaN	NaN	…	NaN
3	JOY	MEH	…	NaN
4	JOY	DESPAIR	…	NaN

	Q9: OTHER COMMENTS	Q10: DRESS
0	NaN	NaN

```
1  Bottom line is Twix is really the only candy w…  White and gold
2                                              NaN             NaN
3                            Raisins can go to hell  White and gold
4                                              NaN  White and gold

  Unnamed: 113 Q11: DAY Q12: MEDIA [Daily Dish] Q12: MEDIA [Science]  \
0          NaN      NaN                     NaN                   NaN
1          NaN   Sunday                     NaN                   1.0
2          NaN      NaN                     NaN                   NaN
3          NaN   Sunday                     NaN                   1.0
4          NaN   Friday                     NaN                   1.0

  Q12: MEDIA [ESPN] Q12: MEDIA [Yahoo] Click Coordinates (x, y)
0               NaN                NaN                      NaN
1               NaN                NaN                (84, 25)
2               NaN                NaN                      NaN
3               NaN                NaN                (75, 23)
4               NaN                NaN                (70, 10)

[5 rows x 120 columns]
```

[30]:
```
#before doing grouping , will quickly fix the missing data issue & data quality␣
↪issue as below
# filter the records which contans all N.A
candy_2017_df_filter_na_only = candy_2017_df.dropna(how='all')
#Fill in missing data fill in the NaN values by using ffill
candy_2017_df_ffill = candy_2017_df_filter_na_only.fillna(method='ffill')
#Fill in missing data fill in the NaN values by using ffill
candy_2017_df_bfill = candy_2017_df_ffill.fillna(method='bfill')
#Fill in missing data fill in the NaN values by 0
candy_2017_df_filldata = candy_2017_df_bfill.fillna(0)
# transforming the country name to country code
candy_2017_df_filldata['Q4: COUNTRY'] = candy_2017_df_filldata['Q4: COUNTRY'].
↪apply(lambda x: fix_country_name(x))
```

[31]:
```
candy_2017_df_filldata.head()
```

[31]:
```
   Internal ID Q1: GOING OUT? Q2: GENDER Q3: AGE Q4: COUNTRY  \
0     90258773             No       Male      44          US
1     90272821             No       Male      44          US
2     90272829             No       Male      49          US
3     90272840             No       Male      40          US
4     90272841             No       Male      23          US

  Q5: STATE, PROVINCE, COUNTY, ETC Q6 | 100 Grand Bar  \
0                               NM               MEH
1                               NM               MEH
```

```
2                         Virginia                MEH
3                               or                MEH
4                         exton pa                JOY

  Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a.
Mary Janes)  \
0                                              DESPAIR
1                                              DESPAIR
2                                              DESPAIR
3                                              DESPAIR
4                                              DESPAIR

  Q6 | Any full-sized candy bar Q6 | Black Jacks  …  \
0                           JOY              MEH  …
1                           JOY              MEH  …
2                           JOY              MEH  …
3                           JOY              MEH  …
4                           JOY          DESPAIR  …

                                  Q8: DESPAIR OTHER  \
0  when the little kids get the big chocolate bar…
1  when the little kids get the big chocolate bar…
2  when the little kids get the big chocolate bar…
3  when the little kids get the big chocolate bar…
4  when the little kids get the big chocolate bar…

                                Q9: OTHER COMMENTS      Q10: DRESS  \
0  Bottom line is Twix is really the only candy w…  White and gold
1  Bottom line is Twix is really the only candy w…  White and gold
2  Bottom line is Twix is really the only candy w…  White and gold
3                           Raisins can go to hell  White and gold
4                           Raisins can go to hell  White and gold

                                     Unnamed: 113 Q11: DAY  \
0  dress (https://survey.ubc.ca/media/assets/user…   Sunday
1  dress (https://survey.ubc.ca/media/assets/user…   Sunday
2  dress (https://survey.ubc.ca/media/assets/user…   Sunday
3  dress (https://survey.ubc.ca/media/assets/user…   Sunday
4  dress (https://survey.ubc.ca/media/assets/user…   Friday

  Q12: MEDIA [Daily Dish] Q12: MEDIA [Science] Q12: MEDIA [ESPN]  \
0                     1.0                  1.0               1.0
1                     1.0                  1.0               1.0
2                     1.0                  1.0               1.0
3                     1.0                  1.0               1.0
4                     1.0                  1.0               1.0
```

```
  Q12: MEDIA [Yahoo] Click Coordinates (x, y)
0                1.0                 (84, 25)
1                1.0                 (84, 25)
2                1.0                 (84, 25)
3                1.0                 (75, 23)
4                1.0                 (70, 10)

[5 rows x 120 columns]
```

[32]:
```
# Grouping with Dicts/Series
#Hierarchical Indexing
hir_candy_2017_df = candy_2017_df_filldata.set_index(['Q4: COUNTRY','Q2:␣
↪GENDER'])
```

[33]:
```
mapping_media= {'Q12: MEDIA [Daily Dish]':'web', 'Q12: MEDIA [Science]':
↪'tv','Q12: MEDIA [ESPN]':'tv','Q12: MEDIA [Yahoo]': 'web'}
```

[34]:
```
a= hir_candy_2017_df.groupby(mapping_media, axis=1)
```

[35]:
```
a.sum().head()
```

[35]:
```
                          tv  web
Q4: COUNTRY Q2: GENDER
US          Male         2.0  2.0
            Male         2.0  2.0
            Male         2.0  2.0
            Male         2.0  2.0
            Male         2.0  2.0
```

[36]:
```
#Grouping with Functions
a1 = hir_candy_2017_df.groupby(len).sum()
```

[37]:
```
a1.head()
```

[37]:
```
    Internal ID  Q12: MEDIA [Daily Dish]  Q12: MEDIA [Science]  \
2  222089196038                   2460.0                2460.0

   Q12: MEDIA [ESPN]  Q12: MEDIA [Yahoo]
2             2460.0               2460.0
```

[38]:
```
# pivot table
data1 = hir_candy_2017_df.pivot_table(index='Q4: COUNTRY', columns='Q2:␣
↪GENDER', values='Q3: AGE', aggfunc=sum)
data1.head()
```

[38]:
```
Q2: GENDER          Female I'd rather not say Male Other
Q4: COUNTRY
AU          Q3: AGE     84                NaN   73   NaN
```

```
CA          Q3: AGE   3159                233  NaN   379
CH          Q3: AGE     36                 30   45   NaN
CN          Q3: AGE     11                NaN   59   NaN
CR          Q3: AGE    NaN                NaN   40   NaN
```

[39]:
```
#Cross Tabs
#pd.crosstab()
pd.crosstab(data1.Female, data1.Male)
```

[39]:
```
Male    32   34   45   46   59   73   119  229  236  345
Female
11       0    0    0    0    1    0    0    0    0    0
19       0    1    0    0    0    0    0    0    0    0
35       0    0    0    1    0    0    0    0    0    0
36       0    0    1    0    0    0    0    0    0    0
49       0    0    0    0    0    0    0    0    1    0
51       0    0    0    0    0    0    1    0    0    0
69       0    0    0    0    0    0    0    1    0    0
84       0    0    0    0    0    1    0    0    0    0
119      1    0    0    0    0    0    0    0    0    0
140      0    0    0    0    0    0    0    0    0    1
```

Chapter 11

[40]:
```
# read CANDY-HIERARCHY-2015-SURVEY-Responses source file into dataframe- for␣
 ↪chapter 7
ts_candy_df = pd.read_excel (r"C:/Users/dell/Documents/docker/Weeks7&8Data/
 ↪CANDY-HIERARCHY-2015-SURVEY-Responses.xlsx")
# display firt 5 records
ts_candy_df.head(5)
```

[40]:
```
                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

  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

   [100 Grand Bar]  \
0              NaN
1              JOY
```

```
2          DESPAIR
3              JOY
4              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                           DESPAIR
1                                           DESPAIR
2                                           DESPAIR
3                                           DESPAIR
4                                           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  …

   [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

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

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

```
4                                                   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

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

  Please estimate the degrees of separation you have from the following folks
[Donald Trump]  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   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 rows x 124 columns]
```

[41]:
```
from datetime import datetime
```

[42]:
```
# Convert between string and date time
ts_candy_df['Timestamp'] = pd.to_datetime(ts_candy_df['Timestamp'],␣
 ↪format='%Y%m%d%H%M%S')
ts_candy_df.head()
```

[42]:

```
                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

  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

   [100 Grand Bar]  \
0              NaN
1              JOY
2          DESPAIR
3              JOY
4              JOY

   [Anonymous brown globs that come in black and orange wrappers]  \
0                                             DESPAIR
1                                             DESPAIR
2                                             DESPAIR
3                                             DESPAIR
4                                             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  …

   [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
```

```
  Please estimate the degrees of separation you have from the following folks
[Bruce Lee]  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   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

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

  Please estimate the degrees of separation you have from the following folks
[JJ Abrams]  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   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
```

```
  Please estimate the degrees of separation you have from the following folks
[Donald Trump]  \
0                                                 NaN
1                                                 NaN
2                                                 NaN
3                                                 NaN
4                                                 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 rows x 124 columns]
```

[43]:
```
#Generate date range
pd.date_range('2015-10-31', periods=10)
```

[43]:
```
DatetimeIndex(['2015-10-31', '2015-11-01', '2015-11-02', '2015-11-03',
               '2015-11-04', '2015-11-05', '2015-11-06', '2015-11-07',
               '2015-11-08', '2015-11-09'],
              dtype='datetime64[ns]', freq='D')
```

[44]:
```
#Frequencies and date offsets
pd.date_range('2015-10-31', periods=10, freq='30min')
```

[44]:
```
DatetimeIndex(['2015-10-31 00:00:00', '2015-10-31 00:30:00',
               '2015-10-31 01:00:00', '2015-10-31 01:30:00',
               '2015-10-31 02:00:00', '2015-10-31 02:30:00',
               '2015-10-31 03:00:00', '2015-10-31 03:30:00',
               '2015-10-31 04:00:00', '2015-10-31 04:30:00'],
              dtype='datetime64[ns]', freq='30T')
```

[45]:
```
#Convert timestamps to periods and back
ts_candy_df['Timestamp'] = ts_candy_df['Timestamp'].dt.to_period('M')
ts_candy_df['Timestamp'].head()
```

[45]:
```
0    2015-10
1    2015-10
2    2015-10
3    2015-10
4    2015-10
Name: Timestamp, dtype: period[M]
```

[ ]: