Name : Balakrishna Mupparaju

Assignment: Week 7 & DSC540

In [228…

```
import pandas as pd
import numpy as np
df=pd.read_excel('/Users/balakrishnamupparaju/Downloads/Weeks 7 & 8 Data/BOI
print(df.head())
```

```
               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 ce
lebrities [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 ce
lebrities [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]
```

Chapter 7: Data Cleaning & Transformation

In [231…

```python
print("Before cleaning:", df.columns.tolist())

# Clean column names: remove leading/trailing whitespace and literal square
df.columns = df.columns.str.strip().str.replace(r'[\[\]]', '', regex=True)

print("After cleaning:", df.columns.tolist())
```

Before cleaning: ['Timestamp', 'Are you going actually going trick or treating yourself?', 'Your gender:', 'How old are you?', 'Which country do you live in?', 'Which state, province, county do you live in?', ' [100 Grand Bar]', ' [Anonymous brown globs that come in black and orange wrappers]', ' [Any full-sized candy bar]', ' [Black Jacks]', ' [Bonkers (the candy)]', ' [Bonkers (the board game)]', ' [Bottle Caps]', " [Box'o'Raisins]", ' [Broken glow stick]', ' [Butterfinger]', ' [Cadbury Creme Eggs]', ' [Candy Corn]', ' [Candy that is clearly just the stuff given out for free at restaurants]', ' [Caramellos]', ' [Cash, or other forms of legal tender]', ' [Chardonnay]', ' [Chick-o-Sticks (we don’t know what that is)]', ' [Chiclets]', ' [Coffee Crisp]', ' [Creepy Religious comics/Chick Tracts]', ' [Dental paraphenalia]', ' [Dots]', ' [Dove Bars]', ' [Fuzzy Peaches]', ' [Generic Brand Acetaminophen]', ' [Glow sticks]', ' [Goo Goo Clusters]', " [Good N' Plenty]", ' [Gum from baseball cards]', ' [Gummy Bears straight up]', ' [Hard Candy]', ' [Healthy Fruit]', ' [Heath Bar]', " [Hershey's Dark Chocolate]", ' [Hershey’s Milk Chocolate]', " [Hershey's Kisses]", ' [Hugs (actual physical hugs)]', ' [Jolly Rancher (bad flavor)]', ' [Jolly Ranchers (good flavor)]', ' [JoyJoy (Mit Iodine!)]', ' [Junior Mints]', ' [Senior Mints]', ' [Kale smoothie]', ' [Kinder Happy Hippo]', ' [Kit Kat]', ' [LaffyTaffy]', ' [LemonHeads]', ' [Licorice (not black)]', ' [Licorice (yes black)]', ' [Lindt Truffle]', ' [Lollipops]', ' [Mars]', ' [Mary Janes]', ' [Maynards]', ' [Mike and Ike]', ' [Milk Duds]', ' [Milky Way]', ' [Regular M&Ms]', ' [Peanut M&M’s]', " [Blue M&M's]", " [Red M&M's]", " [Third Party M&M's]", ' [Minibags of chips]', ' [Mint Kisses]', ' [Mint Juleps]', ' [Mr. Goodbar]', ' [Necco Wafers]', ' [Nerds]', ' [Nestle Crunch]', " [Now'n'Laters]", ' [Peeps]', ' [Pencils]', ' [Person of Interest Season 3 DVD Box Set (not including Disc 4 with hilarious outtakes)]', ' [Pixy Stix]', ' [Reese’s Peanut Butter Cups]', " [Reese's Pieces]", ' [Reggie Jackson Bar]', ' [Rolos]', ' [Skittles]', ' [Smarties (American)]', ' [Smarties (Commonwealth)]', ' [Snickers]', ' [Sourpatch Kids (i.e. abominations of nature)]', ' [Spotted Dick]', ' [Starburst]', ' [Sweet Tarts]', ' [Swedish Fish]', ' [Sweetums (a friend to diabetes)]', ' [Tic Tacs]', ' [Those odd marshmallow circus peanut things]', ' [Three Musketeers]', ' [Tolberone something or other]', ' [Trail Mix]', ' [Twix]', ' [Vials of pure high fructose corn syrup, for main-lining into your vein]', ' [Vicodin]', ' [Whatchamacallit Bars]', ' [White Bread]', ' [Whole Wheat anything]', ' [York Peppermint Patties]', 'Please list any items not included above that give you JOY.', 'Please list any items not included above that give you DESPAIR.', 'Please leave any witty, snarky or thoughtful remarks or comments regarding your choices.', 'Guess the number of mints in my hand.', 'Betty or Veronica?', '"That dress* that went viral a few years back - when I first saw it, it was _____"', 'What is your favourite font?', 'Please estimate the degree(s) of separation you have from the following celebrities [JK Rowling]', 'Please estimate the degree(s) of separation you have from the following celebrities [JJ Abrams]', 'Please estimate the degree(s) of separation you have from the following celebrities [Beyoncé]', 'Please estimate the degree(s) of separation you have from the following celebrities [Bieber]', 'Please estimate the degree(s) of separation you have from the following celebrities [Kevin Bacon]', 'Please estimate the degree(s) of separation you have from the following celebrities [Francis Bacon (1561 - 1626)]', 'Which day do you prefer, Friday or Sunday?', 'Do you eat apples the correct way, East to West (side to side) or do you eat them like a freak of nature, South to North (bottom to top)?', 'When you see the above image of the 4 different websites, which one would you most likely check out (please be honest).', ' [York Peppermint Patties] Ignore']
After cleaning: ['Timestamp', 'Are you going actually going trick or treating yourself?', 'Your gender:', 'How old are you?', 'Which country do you live

```
in?', 'Which state, province, county do you live in?', '100 Grand Bar', 'Ano
nymous brown globs that come in black and orange wrappers', 'Any full-sized
candy bar', 'Black Jacks', 'Bonkers (the candy)', 'Bonkers (the board gam
e)', 'Bottle Caps', "Box'o'Raisins", 'Broken glow stick', 'Butterfinger', 'C
adbury Creme Eggs', 'Candy Corn', 'Candy that is clearly just the stuff give
n out for free at restaurants', 'Caramellos', 'Cash, or other forms of legal
tender', 'Chardonnay', 'Chick-o-Sticks (we don't know what that is)', 'Chicl
ets', 'Coffee Crisp', 'Creepy Religious comics/Chick Tracts', 'Dental paraph
enalia', 'Dots', 'Dove Bars', 'Fuzzy Peaches', 'Generic Brand Acetaminophe
n', 'Glow sticks', 'Goo Goo Clusters', "Good N' Plenty", 'Gum from baseball
cards', 'Gummy Bears straight up', 'Hard Candy', 'Healthy Fruit', 'Heath Ba
r', "Hershey's Dark Chocolate", 'Hershey's Milk Chocolate', "Hershey's Kisse
s", 'Hugs (actual physical hugs)', 'Jolly Rancher (bad flavor)', 'Jolly Ranc
hers (good flavor)', 'JoyJoy (Mit Iodine!)', 'Junior Mints', 'Senior Mints',
'Kale smoothie', 'Kinder Happy Hippo', 'Kit Kat', 'LaffyTaffy', 'LemonHead
s', 'Licorice (not black)', 'Licorice (yes black)', 'Lindt Truffle', 'Lollip
ops', 'Mars', 'Mary Janes', 'Maynards', 'Mike and Ike', 'Milk Duds', 'Milky
Way', 'Regular M&Ms', 'Peanut M&M's', "Blue M&M's", "Red M&M's", "Third Part
y M&M's", 'Minibags of chips', 'Mint Kisses', 'Mint Juleps', 'Mr. Goodbar',
'Necco Wafers', 'Nerds', 'Nestle Crunch', "Now'n'Laters", 'Peeps', 'Pencil
s', 'Person of Interest Season 3 DVD Box Set (not including Disc 4 with hila
rious outtakes)', 'Pixy Stix', 'Reese's Peanut Butter Cups', "Reese's Piece
s", 'Reggie Jackson Bar', 'Rolos', 'Skittles', 'Smarties (American)', 'Smart
ies (Commonwealth)', 'Snickers', 'Sourpatch Kids (i.e. abominations of natur
e)', 'Spotted Dick', 'Starburst', 'Sweet Tarts', 'Swedish Fish', 'Sweetums
(a friend to diabetes)', 'Tic Tacs', 'Those odd marshmallow circus peanut th
ings', 'Three Musketeers', 'Tolberone something or other', 'Trail Mix', 'Twi
x', 'Vials of pure high fructose corn syrup, for main-lining into your vei
n', 'Vicodin', 'Whatchamacallit Bars', 'White Bread', 'Whole Wheat anythin
g', 'York Peppermint Patties', 'Please list any items not included above tha
t give you JOY.', 'Please list any items not included above that give you DE
SPAIR.', 'Please leave any witty, snarky or thoughtful remarks or comments r
egarding your choices.', 'Guess the number of mints in my hand.', 'Betty or
Veronica?', '"That dress* that went viral a few years back - when I first sa
w it, it was _____"', 'What is your favourite font?', 'Please estimate th
e degree(s) of separation you have from the following celebrities JK Rowlin
g', 'Please estimate the degree(s) of separation you have from the following
celebrities JJ Abrams', 'Please estimate the degree(s) of separation you hav
e from the following celebrities Beyoncé', 'Please estimate the degree(s) of
separation you have from the following celebrities Bieber', 'Please estimate
the degree(s) of separation you have from the following celebrities Kevin Ba
con', 'Please estimate the degree(s) of separation you have from the followi
ng celebrities Francis Bacon (1561 - 1626)', 'Which day do you prefer, Frida
y or Sunday?', 'Do you eat apples the correct way, East to West (side to sid
e) or do you eat them like a freak of nature, South to North (bottom to to
p)?', 'When you see the above image of the 4 different websites, which one w
ould you most likely check out (please be honest).', 'York Peppermint Pattie
s Ignore']
```

In [241…

```
# ============================================================================
# Chapter 7 — Data Cleaning & Transformation
# ============================================================================

# 1. Filter out missing data: Drop rows where *all* columns are missing.
df_cleaned = df.dropna(how='all')
print("Chapter 7.1 - Shape after dropping rows with all missing values:", df
```

```python
# 2. Fill in missing data:
#    (a) Forward fill all columns.
df_filled = df_cleaned.ffill()
#    (b) For numeric columns, fill any remaining missing values with the col
numeric_cols = df_filled.select_dtypes(include=[np.number]).columns
df_filled[numeric_cols] = df_filled[numeric_cols].fillna(df_filled[numeric_c
print("Chapter 7.2 - Missing values after filling (sample):")
print(df_filled.isnull().sum().head())

# 3. Remove duplicates:
df_nodup = df_filled.drop_duplicates()
print("Chapter 7.3 - Shape after removing duplicates:", df_nodup.shape)

# 4. Transform data using mapping or a function:
#    Example: Map gender responses: 'Male' -> 1, 'Female' -> 0.
gender_map = {'Male': 1, 'Female': 0}
df_transformed = df_nodup.copy()
df_transformed['Your gender:'] = df_transformed['Your gender:'].map(gender_m
print("Chapter 7.4 - Transformed 'Your gender:' column sample:")
print(df_transformed['Your gender:'].head())

# 5. Replace values:
#    Example: Replace 'MEH' with 'Neutral' across the dataframe.
df_replaced = df_transformed.replace('MEH', 'Neutral')
print("Chapter 7.5 - Sample after replacing 'MEH' with 'Neutral':")
print(df_replaced.head())

# 6. Discretization and Binning:
#    Ensure again that "How old are you?" is numeric (in case further operat
df_replaced['How old are you?'] = pd.to_numeric(df_replaced['How old are you
#    Bin the "How old are you?" column into age groups.
df_replaced['Age_Group'] = pd.cut(df_replaced['How old are you?'].dropna(),
                                  bins=[0, 30, 50, 100],
                                  labels=['Young', 'Middle-aged', 'Senior'])
print("Chapter 7.6 - Sample Age Groups:")
print(df_replaced[['How old are you?', 'Age_Group']].head())

# 7. Manipulate Strings:
#    Example: Convert text in the "Betty or Veronica?" column to uppercase,
if "Betty or Veronica?" in df_replaced.columns:
    df_replaced["Betty or Veronica?"] = df_replaced["Betty or Veronica?"].st
    print("Chapter 7.7 - 'Betty or Veronica?' column in uppercase (sample):"
    print(df_replaced["Betty or Veronica?"].head())
```

```
Chapter 7.1 - Shape after dropping rows with all missing values: (1259, 123)
Chapter 7.2 - Missing values after filling (sample):
Timestamp                                                   0
Are you going actually going trick or treating yourself?    0
Your gender:                                                0
How old are you?                                            0
Which country do you live in?                               0
dtype: int64
Chapter 7.3 - Shape after removing duplicates: (1259, 123)
Chapter 7.4 - Transformed 'Your gender:' column sample:
0    1.0
1    1.0
2    0.0
3    1.0
4    1.0
Name: Your gender:, dtype: float64
Chapter 7.5 - Sample after replacing 'MEH' with 'Neutral':
                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                 1.0
1                                                 No                 1.0
2                                                 No                 0.0
3                                                 No                 1.0
4                                                Yes                 1.0

  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       Neutral
2                                      Colorado           JOY
3                                            il           JOY
4                                  South Dakota       Neutral

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

  Any full-sized candy bar Black Jacks  ...  \
0                     JOY     Neutral  ...
1                     JOY         JOY  ...
2                     JOY     Neutral  ...
```

```
3                     JOY     Neutral  ...
4                     JOY     DESPAIR  ...

  Please estimate the degree(s) of separation you have from the following ce
lebrities 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 ce
lebrities 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 ce
lebrities 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 ce
lebrities 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 ce
lebrities 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 ce
lebrities 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 e
at 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]
Chapter 7.6 - Sample Age Groups:
   How old are you?    Age_Group
0              22.0        Young
1              45.0  Middle-aged
2              48.0  Middle-aged
3              57.0       Senior
4              42.0  Middle-aged
Chapter 7.7 - 'Betty or Veronica?' column in uppercase (sample):
0       BETTY
1       BETTY
2    VERONICA
3       BETTY
4    VERONICA
Name: Betty or Veronica?, dtype: object
```

In [243…

```python
# ================================================================================
# Chapter 8 — Reshaping & Indexing
# ================================================================================

# 8.1 Create hierarchical index using location-based columns.
location_cols = ['Which country do you live in?', 'Which state, province, co
if set(location_cols).issubset(df_replaced.columns):
    df_hier = df_replaced.copy()
    df_hier.set_index(location_cols, inplace=True)
    print("Chapter 8.1 - Hierarchical index applied (sample):")
    print(df_hier.head())
else:
    print("Chapter 8.1 - Required location columns not found; hierarchical i

# 8.2 Pivot Table (Replacing pivot() to prevent duplicate index errors)
```

```
df_pivot_table = df_replaced.pivot_table(index='Which country do you live ir
                                         columns='Which day do you prefer, F
                                         values='How old are you?',
                                         aggfunc='mean')
print("Chapter 8.2 - Pivot table (sample):")
print(df_pivot_table.head())
```

Chapter 8.1 - Hierarchical index applied (sample):

```
                                                                  Timestamp  \
Which country do you live in? Which state, province, county do you live in?
Canada                        Ontario
2016-10-24 05:09:23.033
usa                           il
2016-10-24 05:09:54.798
US                            Colorado
2016-10-24 05:13:06.734
usa                           il
2016-10-24 05:14:17.192
USA                           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?
Canada                        Ontario
No
usa                           il
No
US                            Colorado
No
usa                           il
No
USA                           South Dakota
Yes


                                                  Your gender:  \
Which country do you live in? Which state, province, county do you live in?
Canada                        Ontario
1.0
usa                           il
1.0
US                            Colorado
0.0
usa                           il
1.0
USA                           South Dakota
1.0


                                                  How old are you?  \
Which country do you live in? Which state, province, county do you live in?
Canada                        Ontario
22.0
usa                           il
45.0
US                            Colorado
48.0
usa                           il
57.0
USA                           South Dakota
42.0
```

100 Grand Bar \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
JOY
usa   il
Neutral
US   Colorado
JOY
usa   il
JOY
USA   South Dakota
Neutral

Anonymous brown globs that come in black and orange wrappers \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
DESPAIR
usa   il
Neutral
US   Colorado
DESPAIR
usa   il
Neutral
USA   South Dakota
DESPAIR

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

Black Jacks \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
Neutral
usa   il
JOY
US   Colorado
Neutral
usa   il
Neutral
USA   South Dakota
DESPAIR

Bonkers (the candy) \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
Neutral
usa   il
DESPAIR
US   Colorado
Neutral
usa   il
Neutral
USA   South Dakota
Neutral

Bonkers (the board game) \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
Neutral
usa   il
Neutral
US   Colorado
JOY
usa   il
DESPAIR
USA   South Dakota
JOY

... \
Which country do you live in? Which state, province, county do you live in?
...
Canada   Ontario
...
usa   il
...
US   Colorado
...
usa   il
...
USA   South Dakota
...

Please estimate the degree(s) of separation you have from the following cele
brities JJ Abrams \
Which country do you live in? Which state, province, county do you live in?
Canada   Ontario
2
usa   il
3 or higher
US   Colorado
3 or higher
usa   il
3 or higher

USA 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?
Canada Ontario
3 or higher
usa il
3 or higher
US Colorado
3 or higher
usa il
3 or higher
USA 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?
Canada Ontario
3 or higher
usa il
3 or higher
US Colorado
3 or higher
usa il
3 or higher
USA 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?
Canada Ontario
3 or higher
usa il
3 or higher
US Colorado
3 or higher
usa il
3 or higher
USA 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?
Canada Ontario
3 or higher
usa il
3 or higher

US Colorado
3 or higher
usa il
3 or higher
USA 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?
Canada Ontario
Friday
usa il
Friday
US Colorado
Sunday
usa il
Sunday
USA 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?
Canada Ontario
South to North
usa il
East to West
US Colorado
East to West
usa il
South to North
USA South Dakota
East to West

When you see the above image of the 4 different websites, which one would yo u most likely check out (please be honest). \
Which country do you live in? Which state, province, county do you live in?
Canada Ontario
Science: Latest News and Headlines
usa il
Science: Latest News and Headlines
US Colorado
Science: Latest News and Headlines
usa il
Science: Latest News and Headlines
USA South Dakota
ESPN

York Peppermint Patties Ignore \
Which country do you live in? Which state, province, county do you live in?
Canada Ontario
NaN

```
usa                             il
NaN
US                              Colorado
NaN
usa                             il
NaN
USA                             South Dakota
NaN


Age_Group
Which country do you live in? Which state, province, county do you live in?
Canada                          Ontario
Young
usa                             il
Middle-aged
US                              Colorado
Middle-aged
usa                             il
Senior
USA                             South Dakota
Middle-aged

[5 rows x 122 columns]
Chapter 8.2 - Pivot table (sample):
Which day do you prefer, Friday or Sunday?  Friday  Sunday
Which country do you live in?
30                                             NaN    40.0
44                                            54.0     NaN
45                                            35.0     NaN
47                                            37.0     NaN
51                                            43.0     NaN
```

In [245…

```python
# ================================================================================
# Chapter 10 — Grouping & Aggregation
# ================================================================================

# 10.1 Grouping with Dicts/Series:
df_group = df_replaced.copy()
gender_dict = {1: 'Men', 0: 'Women'}
df_group['Gender_Group'] = df_group['Your gender:'].map(gender_dict)
grouped_gender = df_group.groupby('Gender_Group')['How old are you?'].mean()
print("Chapter 10.1 - Mean age by Gender Group:")
print(grouped_gender)

# 10.2 Grouping with Functions:
def age_category(age):
    if age < 30:
        return 'Youth'
    elif age < 60:
        return 'Adult'
    else:
        return 'Senior'

df_replaced['Age_Category'] = df_replaced['How old are you?'].apply(age_cate
grouped_age = df_replaced.groupby('Age_Category')['How old are you?'].count(
```

```
print("Chapter 10.2 - Count by Age Category:")
print(grouped_age)

# 10.3 Split/Apply/Combine:
split_apply = df_replaced.groupby('Which country do you live in?')['How old
print("Chapter 10.3 - Summary stats by country:")
print(split_apply.head())

# 10.4 Cross Tabs:
crosstab_cols = ['Your gender:', 'Which day do you prefer, Friday or Sunday?
if set(crosstab_cols).issubset(df_replaced.columns):
    crosstab = pd.crosstab(df_replaced['Your gender:'], df_replaced['Which d
    print("Chapter 10.4 - Crosstab (Gender vs Preferred Day):")
    print(crosstab)
else:
    print("Chapter 10.4 - Required columns for crosstab not found.")
```

```
Chapter 10.1 - Mean age by Gender Group:
Gender_Group
Men      1.283697e+15
Women    3.834314e+01
Name: How old are you?, dtype: float64
Chapter 10.2 - Count by Age Category:
Age_Category
Adult     991
Senior     57
Youth     176
Name: How old are you?, dtype: int64
Chapter 10.3 - Summary stats by country:
                               mean  median  count
Which country do you live in?
30                             40.0    40.0      1
44                             54.0    54.0      1
45                             35.0    35.0      1
47                             37.0    37.0      1
51                             43.0    43.0      1
Chapter 10.4 - Crosstab (Gender vs Preferred Day):
Which day do you prefer, Friday or Sunday?  Friday  Sunday
Your gender:
0.0                                            260     156
1.0                                            523     275
```

In [251…

```
# Chapter 11 - Dates and Times

# 11.1 Convert between string and datetime
df_replaced['Timestamp'] = pd.to_datetime(df_replaced['Timestamp'], errors='
print("Chapter 11.1 - 'Timestamp' converted to datetime (sample):")
print(df_replaced['Timestamp'].head())

# 11.2 Generate date range starting from the earliest timestamp
date_range = pd.date_range(start=df_replaced['Timestamp'].min(), periods=10,
print("Chapter 11.2 - Generated Date Range:")
print(date_range)

# 11.3 Convert timestamps to periods and back (Fixing previous error)
df_replaced['Period'] = df_replaced['Timestamp'].dt.to_period('M')
```

```
df_replaced['Timestamp_from_period'] = df_replaced['Period'].apply(lambda p:
print("Chapter 11.3 - Timestamp, Period, and Timestamp-from-Period (sample):
print(df_replaced[['Timestamp', 'Period', 'Timestamp_from_period']].head())

# 11.4 Convert monthly periods to quarterly periods (fixing previous error)
df_replaced['Quarter'] = df_replaced['Period'].apply(lambda p: p.asfreq('Q')
print("Chapter 11.4 - Period conversion (Monthly to Quarterly):")
print(df_replaced[['Period', 'Quarter']].head())
```

```
Chapter 11.1 - 'Timestamp' converted to datetime (sample):
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
Name: Timestamp, dtype: datetime64[ns]
Chapter 11.2 - Generated Date Range:
DatetimeIndex(['2016-10-24 05:09:23.033000', '2016-10-25 05:09:23.033000',
               '2016-10-26 05:09:23.033000', '2016-10-27 05:09:23.033000',
               '2016-10-28 05:09:23.033000', '2016-10-29 05:09:23.033000',
               '2016-10-30 05:09:23.033000', '2016-10-31 05:09:23.033000',
               '2016-11-01 05:09:23.033000', '2016-11-02 05:09:23.033000'],
              dtype='datetime64[ns]', freq='D')
Chapter 11.3 - Timestamp, Period, and Timestamp-from-Period (sample):
                Timestamp   Period Timestamp_from_period
0 2016-10-24 05:09:23.033  2016-10            2016-10-01
1 2016-10-24 05:09:54.798  2016-10            2016-10-01
2 2016-10-24 05:13:06.734  2016-10            2016-10-01
3 2016-10-24 05:14:17.192  2016-10            2016-10-01
4 2016-10-24 05:14:24.625  2016-10            2016-10-01
Chapter 11.4 - Period conversion (Monthly to Quarterly):
    Period Quarter
0  2016-10  2016Q4
1  2016-10  2016Q4
2  2016-10  2016Q4
3  2016-10  2016Q4
4  2016-10  2016Q4
```

In [ ]: