Untitled

September 6, 2020

HW 1

Name: J.Mo Yang, Joanne Lee, Sejin Park

```
[2]: import pandas as pd
```

```
[3]: data = pd.read_csv('./data/nyc_rest_health_inspections.csv')
     data.head(5)
```

```
[3]:        CAMIS                   DBA          BORO BUILDING          STREET  \
     0   40511702       NOTARO RESTAURANT     MANHATTAN      635  SECOND AVENUE
     1   40511702       NOTARO RESTAURANT     MANHATTAN      635  SECOND AVENUE
     2   50046354                VITE BAR        QUEENS     2507       BROADWAY
     3   50061389  TACK'S CHINESE TAKE OUT  STATEN ISLAND   11C     HOLDEN BLVD
     4   41516263              NO QUARTER      BROOKLYN     8015        5 AVENUE

         ZIPCODE        PHONE CUISINE DESCRIPTION INSPECTION DATE  \
     0   10016.0  2126863400              Italian      06/15/2015
     1   10016.0  2126863400              Italian      11/25/2014
     2   11106.0  3478134702              Italian      10/03/2016
     3   10314.0  7189839854              Chinese      05/17/2017
     4   11209.0  7187019180             American      03/30/2017

                                            ACTION VIOLATION CODE  \
     0  Violations were cited in the following area(s).       02B
     1  Violations were cited in the following area(s).       20F
     2  Violations were cited in the following area(s).       10F
     3  Violations were cited in the following area(s).       02G
     4  Violations were cited in the following area(s).       04M

                              VIOLATION DESCRIPTION
     0       Hot food item not held at or above 140Âº F.
     1             Current letter grade card not posted.
     2  Non-food contact surface improperly constructe…
     3  Cold food item held above 41Âº F (smoked fish …
     4  Live roaches present in facility's food and/or…
```

```
[4]: data1=data.drop(columns=["VIOLATION CODE", "VIOLATION DESCRIPTION", "ACTION",␣
     ↪"INSPECTION DATE"])
     data1
```

```
[4]:             CAMIS                                       DBA          BORO  \
     0        40511702                         NOTARO RESTAURANT     MANHATTAN
     1        40511702                         NOTARO RESTAURANT     MANHATTAN
     2        50046354                                  VITE BAR        QUEENS
     3        50061389                    TACK'S CHINESE TAKE OUT  STATEN ISLAND
     4        41516263                                NO QUARTER      BROOKLYN
     ...          ...                                       ...           ...
     212826   50012545                                   WENDY'S     MANHATTAN
     212827   41084277  EAT A BAGEL (On the Guy V. Molinari Ferry)    MANHATTAN
     212828   41427673             SHERATON NEW YORK HOTEL & TOWERS   MANHATTAN
     212829   50011238                            WAN ZHONG WANG      BROOKLYN
     212830   40949462                         CUCINA & COMPANY      MANHATTAN

              BUILDING                   STREET   ZIPCODE        PHONE  \
     0             635            SECOND AVENUE   10016.0   2126863400
     1             635            SECOND AVENUE   10016.0   2126863400
     2            2507                 BROADWAY   11106.0   3478134702
     3             11C              HOLDEN BLVD   10314.0   7189839854
     4            8015                 5 AVENUE   11209.0   7187019180
     ...           ...                      ...       ...          ...
     212826        938                  8TH AVE   10019.0   2129774788
     212827          0  WHITEHALL FERRY TERMINAL  10004.0   7182738400
     212828        811                  7 AVENUE  10019.0   2128416507
     212829        773                  60TH ST   11220.0   7186808886
     212830        200              PARK AVENUE   10003.0   2126822700

              CUISINE DESCRIPTION
     0                     Italian
     1                     Italian
     2                     Italian
     3                     Chinese
     4                    American
     ...                       ...
     212826             Hamburgers
     212827          Bagels/Pretzels
     212828               American
     212829                Chinese
     212830                   Amer

     [212831 rows x 8 columns]
```

```
[9]: data2 = data.drop(columns=['BORO', 'DBA',␣
     ↪'ZIPCODE','STREET','BUILDING','PHONE','CUISINE DESCRIPTION'])
```

```
#data2
data3= data2.drop_duplicates()
#dropped duplicates
data3
```

[9]:   CAMIS INSPECTION DATE \
        0   40511702   06/15/2015
        1   40511702   11/25/2014
        2   50046354   10/03/2016
        3   50061389   05/17/2017
        4   41516263   03/30/2017
        ...   .
        212826 50012545   08/21/2014
        212827 41084277   02/04/2016
        212828 41427673   08/08/2016
        212829 50011238   06/11/2015
        212830 40949462   NaN

                                                    ACTION VIOLATION CODE \
        0   Violations were cited in the following area(s).   02B
        1   Violations were cited in the following area(s).   20F
        2   Violations were cited in the following area(s).   10F
        3   Violations were cited in the following area(s).   02G
        4   Violations were cited in the following area(s).   04M
        ...   .
        212826 Violations were cited in the following area(s).   04N
        212827 Violations were cited in the following area(s).   10H
        212828 Violations were cited in the following area(s).   08A
        212829 Violations were cited in the following area(s).   04L
        212830   NaN   NaN

                                  VIOLATION DESCRIPTION
        0   Hot food item not held at or above 140Âº F.
        1   Current letter grade card not posted.
        2   Non-food contact surface improperly constructe…
        3   Cold food item held above 41Âº F (smoked fish …
        4   Live roaches present in facility's food and/or…
        ...   …
        212826 Filth flies or food/refuse/sewage-associated (…
        212827 Proper sanitization not provided for utensil w…
        212828 Facility not vermin proof. Harborage or condit…
        212829 Evidence of mice or live mice present in facil…
        212830   NaN

        [212827 rows x 5 columns]
```

1) What are the rules of tidy data?

a.Each variable must have its own column.

b.Each observation must have its own row.

c.Each type of observational unit is a table.

2) If you were to make the data tidy, how many tables would you have, and which columns would exist on which tables?

The First table is above, holding information regarding the restaurant and it has 8 columns. Similarly, the last table holds information about the inspection and it has 5 columns.

[ ]: