```
import pandas as pd


#run this for bathroom products


data_B = pd.read_excel("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/delivery/bathroom_rearrange_05032021.xlsx")


#run this for kitchen products


data_B = pd.read_excel("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/delivery/kitchen_rearrange_05042021.xlsx")


df_B = data_B.copy()
df_B.shape
#bathroom data shape (402, 37)

df_B.head(3)
```

	id	task1	t1p1	t1p2	t1p3	t1p4	t1p5	task2	t2p1	t2p2	...	t5p2	t5p3	t5p4	t5p5	task6	t6p1	t6p2	t6p3	t6p4	t6p5
0	4001	5.0	NaN	1000	NaN	NaN	NaN	2	5712	NaN	...	NaN	NaN	NaN	NaN	6.0	NaN	NaN	NaN	NaN	NaN
1	4008	4.0	8313	NaN	NaN	NaN	NaN	3	11400	NaN	...	NaN	NaN	NaN	NaN	7.0	NaN	NaN	NaN	NaN	NaN
2	10002	3.0	8312	4110	NaN	NaN	NaN	5	4110	1B10	...	1000	NaN	NaN	NaN	6.0	1000	6111	NaN	NaN	NaN

3 rows × 37 columns

```
#check for duplicate id
df_B.id.nunique()
```

```
402
```

```
#resolve missing - change to 0
df_B.fillna(-1,inplace=True)
df_B.head()
```

	id	task1	t1p1	t1p2	t1p3	t1p4	t1p5	task2	t2p1	t2p2	...	t5p2	t5p3	t5p4	t5p5	task6
0	4001	5.0	-1	1000	-1	-1.0	-1.0	2	5712	-1	...	-1	-1	-1.0	-1.0	6.0
1	4008	4.0	8313	-1	-1	-1.0	-1.0	3	11400	-1	...	-1	-1	-1.0	-1.0	7.0
2	10002	3.0	8312	4110	-1	-1.0	-1.0	5	4110	1B10	...	1000	-1	-1.0	-1.0	6.0
3	10005	1.0	8310	8610	9312B	-1.0	-1.0	3	9312	4110	...	-1	-1	-1.0	-1.0	7.0
4	10007	2.0	-1	-1	-1	-1.0	-1.0	3	11800	8410	...	9311	-1	-1.0	-1.0	7.0

5 rows × 37 columns

```
#change format
for i in range(1,7):
    i = str(i)
    df_B["task"+i] = df_B["task"+i].astype(int)
    for j in range(1,6):
        j = str(j)
        try:
            df_B["t"+i+"p"+j] = df_B["t"+i+"p"+j].astype(int)
        except:
            pass
df_B.head(3)
```

	id	task1	t1p1	t1p2	t1p3	t1p4	t1p5	task2	t2p1	t2p2	...	t5p2	t5p3	t5p4	t5p5	task6
0	4001	5	-1	1000	-1	-1	-1	2	5712	-1	...	-1	-1	-1	-1	6
1	4008	4	8313	-1	-1	-1	-1	3	11400	-1	...	-1	-1	-1	-1	7
2	10002	3	8312	4110	-1	-1	-1	5	4110	1B10	...	1000	-1	-1	-1	6

3 rows × 37 columns

```
#bathroom run this
tasks = {1:"sink", 2:"toilet", 3:"shower",4:"door",5:"floor",6:"window/mirror"}
df_new = {i:{"id":[],"task":[],"prod1":[],"prod2":[],"prod3":[],"prod4":[],"prod5":[]} for i in range (1,7)}
df_new
```

```
{1: {'id': [],
  'task': [],
  'prod1': [],
  'prod2': [],
  'prod3': [],
  'prod4': [],
```

```
 'prod5': []},
2: {'id': [],
 'task': [],
 'prod1': [],
 'prod2': [],
 'prod3': [],
 'prod4': [],
 'prod5': []},
3: {'id': [],
 'task': [],
 'prod1': [],
 'prod2': [],
 'prod3': [],
 'prod4': [],
 'prod5': []},
4: {'id': [],
 'task': [],
 'prod1': [],
 'prod2': [],
 'prod3': [],
 'prod4': [],
 'prod5': []},
5: {'id': [],
 'task': [],
 'prod1': [],
 'prod2': [],
 'prod3': [],
 'prod4': [],
 'prod5': []},
6: {'id': [],
 'task': [],
 'prod1': [],
 'prod2': [],
 'prod3': [],
 'prod4': [],
 'prod5': []}}


#kitchen run this
tasks = {1:"extractor hood", 2:"fridge/shelves", 3:"burner area",4:"oven",5:"countertop",6:"floor"}
df_new = {i:{"id":[],"task":[],"prod1":[],"prod2":[],"prod3":[],"prod4":[],"prod5":[]} for i in range (1,7)}
df_new

    {1: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []},
     2: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []},
     3: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []},
     4: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []},
     5: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []},
     6: {'id': [],
      'task': [],
      'prod1': [],
      'prod2': [],
      'prod3': [],
      'prod4': [],
      'prod5': []}}


# traversal - to search each record and fill in the new list
for idx in df_B.index:
    for i in range(1,7):
        i = str(i)
        task = 'task'+i
        try:
            df_new[df_B[task][idx]]['id'].append(df_B['id'][idx])
```

```
            df_new[df_B[task][idx]]['task'].append(tasks[df_B[task][idx]])
            for j in range(1,6):
                j = str(j)
                df_new[df_B[task][idx]]['prod'+j].append(df_B['t'+i+'p'+j][idx])
        except:
            pass


# generate complete dataframe
def toDF(task):
    return pd.DataFrame(df_new[task])

df1 = toDF(1)
for i in range(2,7):
    df1 = df1.append(toDF(i))

df1.index = range(len(df1))
df1
```

	id	task	prod1	prod2	prod3	prod4	prod5
0	4001	extractor hood	11700	-1	-1	-1	-1
1	10002	extractor hood	11100	1000	-1	-1	-1
2	10005	extractor hood	8310	8610	9312B	-1	-1
3	10007	extractor hood	11800	8410	-1	-1	-1
4	10012	extractor hood	5721	-1	-1	-1	-1
...	...	...	...	...	...	...	...
2007	60057	floor	12300	-1	-1	-1	-1
2008	60058	floor	12300	-1	-1	-1	-1
2009	60059	floor	6610	-1	-1	-1	-1
2010	60060	floor	-1	-1	-1	-1	-1
2011	200117	floor	6510	-1	-1	-1	-1

2012 rows × 7 columns

```
#check again for duplicate id
df1.id.nunique()
```

```
402
```

```
# bathroom - export to file
df1.to_csv("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/delivery/bathroom_05052021.csv")


# kitchen - export to file
df1.to_csv("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/delivery/kitchen_05052021.csv")


# code ends here


# the following section is for bathroom section


#filter task 1 = "bathroom sink" and create a new dateframe
df1 = df_B[df_B['task1'] == 1.0]
#df1.shape 106
#bathroom sink data shape (106, 37)
df_BSink1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BSink1

#filter task 2 = "bathroom sink" and create a new dateframe
df2 = df_B[df_B['task2'] == 1.0]
#df2 data shape (68,37)
df_BSink2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
# change the column name for future merge
df_BSink2.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_BSink2

#filter task 3 = "bathroom sink" and create a new dateframe
df3 = df_B[df_B['task3'] == 1.0]
df_BSink3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_BSink3.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_BSink3
#df_BSink3.shape (136,6)

#filter task 4 = "bathroom sink" and create a new dateframe
df4 = df_B[df_B['task4'] == 1.0]
df_BSink4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
```

```
df_BSink4.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_BSink4.head()
#data shape (59,6)

#filter task 5 = "bathroom sink" and create a new dateframe
df5 = df_B[df_B['task5'] == 1.0]
df_BSink5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_BSink5.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_BSink5.head()
#data shape (25,6)

#filter task 6 = "bathroom sink" and create a new dateframe
df6 = df_B[df_B['task6'] == 1.0]
df_BSink6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_BSink6.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_BSink6.head()
#data shape (0,6)

b1 = df_BSink1.append(df_BSink2,ignore_index=True, sort=False)
b2 = b1.append(df_BSink3,ignore_index=True, sort=False)
b3 = b2.append(df_BSink4,ignore_index=True, sort=False)
b4 = b3.append(df_BSink5,ignore_index=True, sort=False)
b5 = b4.append(df_BSink6,ignore_index=True, sort=False)
b5
```

	id	t1p1	t1p2	t1p3	t1p4	t1p5
1	4008	1135	11400	NaN	NaN	NaN
20	10072	1121	6310	NaN	NaN	NaN
34	10123	11400	NaN	NaN	NaN	NaN
37	10130	NaN	5700	NaN	NaN	NaN
44	10149	4210	NaN	NaN	NaN	NaN

```
df_BSink.id.nunique()
```

```
392
```

```
# merge all dataframes
frames1 = [df_BSink1, df_BSink2, df_BSink3,df_BSink4,df_BSink5,df_BSink6]
df_BSink = pd.concat(frames1)
df_BSink

#checked, no duplicates found
#df_BSink.drop_duplicates()
#df_BSink[df_BSink.duplicated()]



# bathroom toilet
#filter task 1 = "bathroom toilet" and create a new dateframe
df1 = df[df_B['task1'] == 2.0]
df_BTlt1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BTlt1.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt1 shape (90,6)

#filter task 2 = "bathroom toilet" and create a new dateframe
df2 = df[df_B['task2'] == 2.0]
df_BTlt2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
#df_BTlt2.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt2 shape (161,6)

#filter task 3 = "bathroom toilet" and create a new dateframe
df3 = df[df_B['task3'] == 2.0]
df_BTlt3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_BTlt3.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt3 shape (92,6)

#filter task 4 = "bathroom toilet" and create a new dateframe
df4 = df[df_B['task4'] == 2.0]
df_BTlt4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_BTlt4.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt4 shape (34,6)

#filter task 5 = "bathroom toilet" and create a new dateframe
df5 = df[df_B['task5'] == 2.0]
df_BTlt5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_BTlt5.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt5 shape (19,6)

#filter task 6 = "bathroom toilet" and create a new dateframe
```

```
df6 = df[df_B['task6'] == 2.0]
df_BTlt6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_BTlt6.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]
# df_BTlt6 shape (3,6)

# merge all bathroom toilet dataframes
frames2 = [df_BTlt1, df_BTlt2, df_BTlt3,df_BTlt4,df_BTlt5,df_BTlt6]
df_BTlt = pd.concat(frames2,sort=False)
df_BTlt

#checked, no duplicates found
#df_BTlt.drop_duplicates()
#df_BTlt[df_BTlt.duplicated()]
```

	id	t2p1	t2p2	t2p3	t2p4	t2p5
16	10052	8314	NaN	NaN	NaN	NaN
19	10060	8311	5721	NaN	NaN	NaN
21	10073	5721	NaN	NaN	NaN	NaN
24	10087	1410	NaN	NaN	NaN	NaN
40	10140	1420	1430	5721	NaN	NaN
...	...	...	...	...	...	...
377	60025	NaN	NaN	NaN	NaN	NaN
380	60029	NaN	NaN	NaN	NaN	NaN
43	10148	NaN	NaN	NaN	NaN	NaN
178	20143	6111	NaN	NaN	NaN	NaN
314	50019	6610	6A00	NaN	NaN	NaN

399 rows × 6 columns

```
# bathroom Tub/Shower
#filter task 1 = "bathroom tub or shower" and create a new dateframe
df1 = df[df_B['task1'] == 3.0]
df_BTbSw1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BTbSw1.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw1 shape (173,6)

#filter task 2 = "bathroom tub or shower" and create a new dateframe
df2 = df[df_B['task2'] == 3.0]
df_BTbSw2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_BTbSw2.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw2 shape (111,6)

#filter task 3 = "bathroom tub or shower" and create a new dateframe
df3 = df[df_B['task3'] == 3.0]
df_BTbSw3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
#df_BTbSw2.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw3 shape (83,6)

#filter task 4 = "bathroom tub or shower" and create a new dateframe
df4 = df[df_B['task4'] == 3.0]
df_BTbSw4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_BTbSw4.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw4 shape (29,6)

#filter task 5 = "bathroom tub or shower" and create a new dateframe
df5 = df[df_B['task5'] == 3.0]
df_BTbSw5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_BTbSw5.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw5 shape (7,6)

#filter task 6 = "bathroom tub or shower" and create a new dateframe
df6 = df[df_B['task6'] == 3.0]
df_BTbSw6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_BTbSw6.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw6 shape (2,6)

# merge all bathroom tub and shower dataframes
frames3 = [df_BTbSw1, df_BTbSw2, df_BTbSw3,df_BTbSw4,df_BTbSw5,df_BTbSw6]
df_BTbSw = pd.concat(frames3,sort=False)
df_BTbSw

#checked, 2 duplicates found
#df_BTbSw.drop_duplicates()
df_BTbSw[df_BTbSw.duplicated()]
```

	id	t3p1	t3p2	t3p3	t3p4	t3p5
274	40014	11930	NaN	NaN	NaN	NaN

```
# Bathroom Door
#filter task 1 = "bathroom door" and create a new dateframe
df1 = df[df_B['task1'] == 4.0]
df_BDr1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BDr1.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr1 shape (6,6)

#filter task 2 = "bathroom door" and create a new dateframe
df2 = df[df_B['task2'] == 4.0]
df_BDr2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_BDr2.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr2 shape (40,6)

#filter task 3 = "bathroom door" and create a new dateframe
df3 = df[df_B['task3'] == 4.0]
df_BDr3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_BDr3.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr3 shape (41,6)

#filter task 4 = "bathroom door" and create a new dateframe
df4 = df[df_B['task4'] == 4.0]
df_BDr4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
#df_BDr4.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr4 shape (73,6)

#filter task 5 = "bathroom door" and create a new dateframe
df5 = df[df_B['task5'] == 4.0]
df_BDr5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_BDr5.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr5 shape (43,6)

#filter task 6 = "bathroom door" and create a new dateframe
df6 = df[df_B['task6'] == 4.0]
df_BDr6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_BDr6.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
# df_BDr6 shape (21,6)

# merge all bathroom Door dataframes
frames4 = [df_BDr1, df_BDr2, df_BDr3,df_BDr4,df_BDr5,df_BDr6]
df_BDr = pd.concat(frames4,sort=False)
df_BDr

#checked, no duplicates found
#df_BDr.drop_duplicates()
#df_BDr[df_BDr.duplicated()]
```

	id	t4p1	t4p2	t4p3	t4p4	t4p5
1	4008	8313	NaN	NaN	NaN	NaN
23	10077	8310	NaN	NaN	NaN	NaN
68	10225	12300	NaN	NaN	NaN	NaN
108	10299	6111	1113	1000	NaN	NaN
260	30114	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...
244	30082	6111	NaN	NaN	NaN	NaN
245	30084	NaN	NaN	NaN	NaN	NaN
304	40058	NaN	NaN	NaN	NaN	NaN
312	50013	6710	11700	NaN	NaN	NaN
328	50081	1121	6710	NaN	NaN	NaN

224 rows × 6 columns

```
#Bathroom Floor
#filter task 1 = "bathroom floor" and create a new dateframe
df1 = df[df_B['task1'] == 5.0]
df_BFlr1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BFlr1.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
# df_BFlr1 shape (3,6)

#filter task 2 = "bathroom floor" and create a new dateframe
df2 = df[df_B['task2'] == 5.0]
df_BFlr2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_BFlr2.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
# df_BFlr2.shape (2,6)
```

```
#filter task 3 = "bathroom floor" and create a new dateframe
df3 = df[df_B['task3'] == 5.0]
df_BFlr3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_BFlr3.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
#df_BFlr3.shape (11,6)

#filter task 4 = "bathroom floor" and create a new dateframe
df4 = df[df_B['task4'] == 5.0]
df_BFlr4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_BFlr4.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
# df_BFlr4.shape (50,6)

#filter task 5 = "bathroom floor" and create a new dateframe
df5 = df[df_B['task5'] == 5.0]
df_BFlr5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
#df_BFlr5.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
#df_BFlr5.shape (160,6)

#filter task 6 = "bathroom floor" and create a new dateframe
df6 = df[df_B['task6'] == 5.0]
df_BFlr6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_BFlr6.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]
# df_BFlr6.shape (174, 6)

# merge all bathroom Door dataframes
frames5 = [df_BFlr1, df_BFlr2, df_BFlr3,df_BFlr4,df_BFlr5,df_BFlr6]
df_BFlr = pd.concat(frames5,sort=False)
df_BFlr
# df_BFlr shape (400,6)

#checked, no duplicates found
#df_BFlr.drop_duplicates()
#df_BFlr[df_BFlr.duplicated()]
```

id	t5p1	t5p2	t5p3	t5p4	t5p5

```
#Bathroom window/mirror
#filter task 1 = "bathroom window/mirror" and create a new dateframe
df1 = df[df_B['task1'] == 6.0]
df_BMrWd1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_BMrWd1.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 2 = "bathroom window/mirror" and create a new dateframe
df2 = df[df_B['task2'] == 6.0]
df_BMrWd2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_BMrWd2.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 3 = "bathroom window/mirror" and create a new dateframe
df3 = df[df_B['task3'] == 6.0]
df_BMrWd3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_BMrWd3.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 4 = "bathroom window/mirror" and create a new dateframe
df4 = df[df_B['task4'] == 6.0]
df_BMrWd4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_BMrWd4.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 5 = "bathroom window/mirror" and create a new dateframe
df5 = df[df_B['task5'] == 6.0]
df_BMrWd5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_BMrWd5.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 6 = "bathroom window/mirror" and create a new dateframe
df6 = df[df_B['task6'] == 6.0]
df_BMrWd6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]

# merge all dataframes
frames6 = [df_BMrWd1, df_BMrWd2, df_BMrWd3,df_BMrWd4,df_BMrWd5,df_BMrWd6]
df_BMrWd = pd.concat(frames6,sort=False)
df_BMrWd

#checked, no duplicates found
#df_BMrWd.drop_duplicates()
#df_BMrWd[df_BMrWd.duplicated()]
```

	id	t6p1	t6p2	t6p3	t6p4	t6p5
47	10165	12300	NaN	NaN	NaN	NaN
54	10187	11400	5700	NaN	NaN	NaN
67	10224	5721	7111	11400	NaN	NaN
105	10294	11400	12300	NaN	NaN	NaN
111	10302	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...
336	50110	6111	6610	NaN	NaN	NaN
338	50114	9311	NaN	NaN	NaN	NaN

```
# merge all dataframe into a full dataset
b1 = df_BSink.set_index('id').join(df_BTlt.set_index('id'))
b2 = b1.join(df_BTbSw.set_index('id'))
b3 = b2.join(df_BDr.set_index('id'))
b4 = b3.join(df_BFlr.set_index('id'))
bathroom_full = b4.join(df_BMrWd.set_index('id'))

bathroom_full_clean = bathroom_full.drop_duplicates()
bathroom_full_clean

#please note there's potential duplicates IDs since the obsevation should be <= 402.
```

id	t1p1	t1p2	t1p3	t1p4	t1p5	t2p1	t2p2	t2p3	t2p4	t2p5	...	t5p1	t5p2	t5p3	t5p4	t5p
4001	NaN	1000	NaN	NaN	NaN	5712	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
4008	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	6710	6610	NaN	NaN	Na
10002	8312	4110	NaN	NaN	NaN	4110	1B10	NaN	NaN	NaN	...	11100	1000	NaN	NaN	Na
10005	6111	6512	6710E	NaN	NaN	9312	4110	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
10007	NaN	NaN	NaN	NaN	NaN	11800	8410	NaN	NaN	NaN	...	11800	8410	NaN	NaN	Na
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
60053	11400	NaN	NaN	NaN	NaN	11400	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
60055	8110	baby oil	NaN	NaN	NaN	11400	1B10	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
60056	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
60060	NaN	NaN	NaN	NaN	NaN	1B10	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
200117	5721	NaN	NaN	NaN	NaN	6510	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na

403 rows × 30 columns

```
#save the bathroom sink data to cvs
bathroom_full_clean.to_csv("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/bathroom_05042021.csv")


# the bathroom section has ended


# the following is the kitchen section


#run this for kitchen products


data_K = pd.read_excel("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/kitchen_rearrange_05042021.xlsx")


df_K = data_K.copy()
df_K.shape
#kitchen data shape (402, 37)

df_K.head(3)
```

	id	task1	t1p1	t1p2	t1p3	t1p4	t1p5	task2	t2p1	t2p2	...	t5p2	t5p3	t5p4	t5p5	task6
0	4001	5.0	NaN	1000	NaN	NaN	NaN	2	5712	NaN	...	NaN	NaN	NaN	NaN	6.0
1	4008	4.0	8313	NaN	NaN	NaN	NaN	3	11400	NaN	...	NaN	NaN	NaN	NaN	7.0
2	10002	3.0	8312	4110	NaN	NaN	NaN	5	4110	1B10	...	1000	NaN	NaN	NaN	6.0

3 rows × 37 columns

```
#filter task 1 = "kitchen extractor hood" and create a new dateframe
df1 = df[df_K['task1'] == 1.0]
df1.head()
```

```
#kitchen extractor hood data shape (39,37)
df_KEHd1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_KEHd1.head()

#filter task 2 = "kitchen extractor hood" and create a new dateframe
df2 = df[df_K['task2'] == 1.0]
df2.head()
#kitchen extractor hood data shape (64,37)
df_KEHd2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
#df_KEHd2.shape (64,6)
# change the column name for future merge
df_KEHd2.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_KEHd2

#filter task 3 = "kitchen extractor hood" and create a new dateframe
df3 = df[df_K['task3'] == 1.0]
df_KEHd3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_KEHd3.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_KEHd3
#df_KEHd3 data shape (62,6)

#filter task 4 = "kitchen extractor hood" and create a new dateframe
df4 = df[df_K['task4'] == 1.0]
df_KEHd4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_KEHd4.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_KEHd4.head()
#df_KEHd4 data shape (51,6)

#filter task 5 = "kitchen extractor hood" and create a new dateframe
df5 = df[df_K['task5'] == 1.0]
df_KEHd5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_KEHd5.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_KEHd5.head()
#data shape (23,6)

#filter task 6 = "kitchen extractor hood" and create a new dateframe
df6 = df[df_K['task6'] == 1.0]
df_KEHd6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_KEHd6.columns =["id","t1p1","t1p2","t1p3","t1p4","t1p5"]
df_KEHd6.head()
#data shape (8,6)

# merge all kichen extractor hood dataframes
frames1 = [df_KEHd1, df_KEHd2, df_KEHd3,df_KEHd4,df_KEHd5,df_KEHd6]
df_KEHd = pd.concat(frames1)
df_KEHd

#checked, no duplicates found
#df_KEHd.drop_duplicates()
#df_KEHd[df_KEHd.duplicated()]
```

	id	t1p1	t1p2	t1p3	t1p4	t1p5
3	10005	8310	8610	9312B	NaN	NaN
16	10052	8314	NaN	NaN	NaN	NaN
22	10074	4110	10200	NaN	NaN	NaN
23	10077	8310	NaN	NaN	NaN	NaN
25	10088	5710	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...
158	20098	8313	NaN	NaN	NaN	NaN
167	20114	6610	NaN	NaN	NaN	NaN
179	20145	8314	NaN	NaN	NaN	NaN
184	20152	6910	NaN	NaN	NaN	NaN
263	30119	1121	NaN	NaN	NaN	NaN

247 rows × 6 columns

```
# kithchen Refrigerator (including shelves)
#filter task 1 and create a new dateframe
df1 = df[df_K['task1'] == 2.0]
df_KRfgS1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_KRfgS1.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

#filter task 2 and create a new dateframe
df2 = df[df_K['task2'] == 2.0]
df_KRfgS2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
#df_BTlt2.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

#filter task 3 and create a new dateframe
```

```
df3 = df[df_K['task3'] == 2.0]
df_KRfgS3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_KRfgS3.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

#filter task 4 and create a new dateframe
df4 = df[df_K['task4'] == 2.0]
df_KRfgS4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_KRfgS4.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

#filter task 5 and create a new dateframe
df5 = df[df_K['task5'] == 2.0]
df_KRfgS5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_KRfgS5.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

#filter task 6 and create a new dateframe
df6 = df[df_K['task6'] == 2.0]
df_KRfgS6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_KRfgS6.columns =["id","t2p1","t2p2","t2p3","t2p4","t2p5"]

# merge all dataframes
frames2 = [df_KRfgS1, df_KRfgS2, df_KRfgS3,df_KRfgS4,df_KRfgS5,df_KRfgS6]
df_KRfgS = pd.concat(frames2,sort=False)
df_KRfgS
# df_KRfgS data shape (328,6)

#checked, 1 duplicates found
#df_KRfgS.drop_duplicates()
df_KRfgS[df_KRfgS.duplicated()]
```

	id	t2p1	t2p2	t2p3	t2p4	t2p5
11	10030	12100	NaN	NaN	NaN	NaN

```
# kitchen burner area
#filter task 1 and create a new dateframe
df1 = df[df_K['task1'] == 3.0]
df_Kbrnr1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_Kbrnr1.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]
# df_BTbSw1 shape (173,6)

#filter task 2 and create a new dateframe
df2 = df[df_K['task2'] == 3.0]
df_Kbrnr2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_Kbrnr2.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]

#filter task 3 and create a new dateframe
df3 = df[df_K['task3'] == 3.0]
df_Kbrnr3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
#df_BTbSw2.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]

#filter task 4 and create a new dateframe
df4 = df[df_K['task4'] == 3.0]
df_Kbrnr4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_Kbrnr4.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]

#filter task 5 and create a new dateframe
df5 = df[df_K['task5'] == 3.0]
df_Kbrnr5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_Kbrnr5.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]

#filter task 6 and create a new dateframe
df6 = df[df_K['task6'] == 3.0]
df_Kbrnr6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_Kbrnr6.columns =["id","t3p1","t3p2","t3p3","t3p4","t3p5"]

# merge all dataframes
frames3 = [df_Kbrnr1, df_Kbrnr2, df_Kbrnr3,df_Kbrnr4,df_Kbrnr5,df_Kbrnr6]
df_Kbrnr = pd.concat(frames3,sort=False)
df_Kbrnr

#checked, no duplicates found
#df_Kbrnr.drop_duplicates()
#df_Kbrnr[df_Kbrnr.duplicated()]
```

id	t3p1	t3p2	t3p3	t3p4	t3p5

```
# kitchen Oven
#filter task 1 and create a new dateframe
df1 = df[df_K['task1'] == 4.0]
df_KOvn1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_KOvn1.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]
```

```
#filter task 2 and create a new dateframe
df2 = df[df_K['task2'] == 4.0]
df_KOvn2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_KOvn2.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]

#filter task 3 and create a new dateframe
df3 = df[df_K['task3'] == 4.0]
df_KOvn3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_KOvn3.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]

#filter task 4 and create a new dateframe
df4 = df[df_K['task4'] == 4.0]
df_KOvn4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
#df_BDr4.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]

#filter task 5 and create a new dateframe
df5 = df[df_K['task5'] == 4.0]
df_KOvn5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_KOvn5.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]

#filter task 6 and create a new dateframe
df6 = df[df_K['task6'] == 4.0]
df_KOvn6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_KOvn6.columns =["id","t4p1","t4p2","t4p3","t4p4","t4p5"]

# merge all dataframes
frames4 = [df_KOvn1, df_KOvn2, df_KOvn3,df_KOvn4,df_KOvn5,df_KOvn6]
df_KOvn = pd.concat(frames4,sort=False)
df_KOvn

#checked, 1 duplicates found
#df_KOvn.drop_duplicates()
df_KOvn[df_KOvn.duplicated()]
```

	id	t4p1	t4p2	t4p3	t4p4	t4p5
242	30080	8313	NaN	NaN	NaN	NaN

```
#Kitchen Meson
#filter task 1 and create a new dateframe
df1 = df[df_K['task1'] == 5.0]
df_KMsn1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_KMsn1.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

#filter task 2 and create a new dateframe
df2 = df[df_K['task2'] == 5.0]
df_KMsn2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_KMsn2.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

#filter task 3 and create a new dateframe
df3 = df[df_K['task3'] == 5.0]
df_KMsn3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_KMsn3.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

#filter task 4 and create a new dateframe
df4 = df[df_K['task4'] == 5.0]
df_KMsn4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_KMsn4.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

#filter task 5 and create a new dateframe
df5 = df[df_K['task5'] == 5.0]
df_KMsn5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
#df_BFlr5.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

#filter task 6 and create a new dateframe
df6 = df[df_K['task6'] == 5.0]
df_KMsn6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]
df_KMsn6.columns =["id","t5p1","t5p2","t5p3","t5p4","t5p5"]

# merge all dataframes
frames5 = [df_KMsn1, df_KMsn2, df_KMsn3,df_KMsn4,df_KMsn5,df_KMsn6]
df_KMsn = pd.concat(frames5,sort=False)
df_KMsn

#checked, no duplicates found
#df_KMsn.drop_duplicates()
#df_KMsn[df_KMsn.duplicated()]
```

	id	t5p1	t5p2	t5p3	t5p4	t5p5
0	4001	NaN	1000	NaN	NaN	NaN
12	10041	11100	11600	NaN	NaN	NaN
15	10047	11800	4410	NaN	NaN	NaN
21	10073	5721	NaN	NaN	NaN	NaN
26	10089	7111	11400	NaN	NaN	NaN
...	...	...	...	...	...	...
360	50159	1113	7111	NaN	NaN	NaN
376	60024	11800	1610	12300	NaN	NaN

```
#Kitchen Floor
#filter task 1 and create a new dateframe
df1 = df[df_K['task1'] == 6.0]
df_KFlr1 = df1[["id","t1p1","t1p2","t1p3","t1p4","t1p5"]]
df_KFlr1.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 2 and create a new dateframe
df2 = df[df_K['task2'] == 6.0]
df_KFlr2 = df2[["id","t2p1","t2p2","t2p3","t2p4","t2p5"]]
df_KFlr2.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 3 and create a new dateframe
df3 = df[df_K['task3'] == 6.0]
df_KFlr3 = df3[["id","t3p1","t3p2","t3p3","t3p4","t3p5"]]
df_KFlr3.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 4 and create a new dateframe
df4 = df[df_K['task4'] == 6.0]
df_KFlr4 = df4[["id","t4p1","t4p2","t4p3","t4p4","t4p5"]]
df_KFlr4.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 5 and create a new dateframe
df5 = df[df_K['task5'] == 6.0]
df_KFlr5 = df5[["id","t5p1","t5p2","t5p3","t5p4","t5p5"]]
df_KFlr5.columns =["id","t6p1","t6p2","t6p3","t6p4","t6p5"]

#filter task 6 and create a new dateframe
df6 = df[df_K['task6'] == 6.0]
df_KFlr6 = df6[["id","t6p1","t6p2","t6p3","t6p4","t6p5"]]

# merge all dataframes
frames6 = [df_KFlr1, df_KFlr2, df_KFlr3,df_KFlr4,df_KFlr5,df_KFlr6]
df_KFlr = pd.concat(frames6,sort=False)
df_KFlr

#checked, no duplicates found
#df_KFlr.drop_duplicates()
#df_KFlr[df_KFlr.duplicated()]
```

	id	t6p1	t6p2	t6p3	t6p4	t6p5
51	10183	NaN	NaN	NaN	NaN	NaN
199	20183	11700	6A00	NaN	NaN	NaN
276	40016	NaN	11930	NaN	NaN	NaN
97	10285	6310	6310	10110	NaN	NaN
147	20077	6512	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...
362	50161	6A00	NaN	NaN	NaN	NaN
366	60008	6111	6512	NaN	NaN	NaN
371	60017	6110	9310	NaN	NaN	NaN
376	60024	12300	11700	NaN	NaN	NaN
390	60046	6610	6310	6100	NaN	NaN

399 rows × 6 columns

```
# merge all dataframe into a full dataset
k1 = df_KEHd.set_index('id').join(df_KRfgS.set_index('id'))
k2 = k1.join(df_Kbrnr.set_index('id'))
k3 = k2.join(df_KOvn.set_index('id'))
k4 = k3.join(df_KMsn.set_index('id'))
kitchen_full = k4.join(df_KFlr.set_index('id'))
kitchen_full
```

```
#show duplicate id
kitchen_full[kitchen_full.duplicated()]
#drop duplicate row
kitchen_full_clean = kitchen_full.drop_duplicates()
kitchen_full_clean
```

id	t1p1	t1p2	t1p3	t1p4	t1p5	t2p1	t2p2	t2p3	t2p4	t2p5	...	t5p1	t5p2	t5p3	t5p4	t5p5	t6p1	t6p2	t6p3	t6p4	t6p5
10030	12100	NaN	NaN	NaN	NaN	12100	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	6111	NaN	NaN	NaN	NaN
30080	5721	NaN	NaN	NaN	NaN	5721	NaN	NaN	NaN	NaN	...	11400	NaN	NaN	NaN	NaN	6110	NaN	NaN	NaN	NaN

2 rows × 30 columns

```
#save the bathroom sink data to cvs
kitchen_full_clean.to_csv("/Users/apple/Dropbox/Safe and Just Cleaning/Yannan/recoding products to tasks/kitchen_05042021.csv")


# the kitchen section has ended
```