model_building

December 10, 2018

```
In [114]: import pandas as pd
          import pulp as lp
          import numpy as np
          from operator import iadd
          from functools import reduce
          from itertools import chain
          from typing import Sequence, Any, DefaultDict, List
```

The below functions assist in creating an lp model problem using `pulp` package

```
In [115]: def create_prob(prob_name: str, sense: int) -> lp.LpProblem:
              return lp.LpProblem(prob_name, sense)


          def add_obj_fn(lp_prob: lp.LpProblem, dvar: lp.LpAffineExpression) -> lp.LpProblem:
              return iadd(lp_prob, dvar)


          def add_constraint(lp_prob: lp.LpProblem, constrs: Sequence[lp.LpConstraint]) -> lp.
             return reduce(iadd, constrs, lp_prob)


          def head(x: Sequence) -> Any:
              return x[0]

          def to_str(indnum, activity) -> str:
              return f'{indnum} - {activity}'
```

1 Loading The Prepared Data

We load our dataset into `activity_df`. `activity_df` contains all the observations related to individuals, their activities, consumption per unit and the relavant sources activities use.

```
In [116]: activity_df = pd.read_csv('data/dataset.csv')

          activity_df.drop(['Unnamed: 0', 'Group'], axis=1, inplace=True)
```

We load the sources carbon footprint dataset into `source_df`. `source_df` contains all the observations related to an carbon footprint an activity contains given that they are used with a specific source.

```
In [117]: source_df = pd.read_csv('data/dataset_source_cf.csv')

          source_df.drop(['Unnamed: 0', 'X__1'], axis=1, inplace=True)
```

```
In [118]: source_df.head()
```

Out[118]:

	Activity	Per	solar_powered_water_heater	gas_water_heater	electric_water_heater_peak_hour	electric_water_heater_off_peak	gas	natural_gas	jetfuel	waste management	hybrid	electric_peak_hours	electric_off_peak_hours
0	Household heating => 70F	hour	0.000000	0.000000	0.000000	0.000000	0.0	0.000436	0.0	0.0	0.0	0.000650	0.000542
1	Household heating < 70F	hour	0.000000	0.000000	0.000000	0.000000	0.0	0.000872	0.0	0.0	0.0	0.000923	0.000901
2	Use of heat pump	hour	0.000000	0.000000	0.000000	0.000000	0.0	0.001074	0.0	0.0	0.0	0.001229	0.001188
3	Use of air conditioner	hour	0.000000	0.000000	0.000000	0.000000	0.0	0.000598	0.0	0.0	0.0	0.007980	0.000721
4	shower - short	activity	0.000012	0.000102	0.000232	0.000199	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000

2 Building The Model

`lp_prob` is the linear programming problem we have to formulize. The problem is one in which we have to minimize the objective function.

```
In [119]: lp_prob = create_prob('Wells Fargo Challenge', lp.LpMinimize)
```

2.1 Defining The Decision Variables

The `individuals` variable represents all the indviduals that we want to focus on, here we are looking at the first **10** individuals. `MAX_NUM_INDV` is used to change the maximum number of individuals the linear programming model should solve for.

The `activities` variable is the list of activities that all indiviuals do.
The `sources` variables is the list of all sources that an activity can use.

```
In [120]: MAX_NUM_INDV = 2
```

```
In [121]: individuals = activity_df.Indnum.unique()[0:MAX_NUM_INDV]
          activities = activity_df.Activity.unique()

          M = 100 # The big number M

          sources = [
            "solar_powered_water_heater",
            "gas_water_heater",
            "electric_water_heater_peak_hour",
            "electric_water_heater_off_peak",
            "gas",
            "natural_gas",
            "hybrid",
            "electric_peak_hours",
            "electric_off_peak_hours",
            "jetfuel",
            "waste management"
          ]
```

Our decision variable is called

2.1.1 S_{ijk}

where * i is individual * j is activity * k is source

It is binary variable that can have either 0 or 1. Meaning that S_{ijk} represents wether an individual uses a particular source or not.

```
In [122]: source_indexes = [
              (f'{indv} - {activity}', source)
              for indv in individuals
              for activity in activities
              for source in sources
          ]

          S_ijk = lp.LpVariable.dicts('S_ijk', source_indexes, lowBound=0, cat='Binary')
```

2.2 The Objective Function

2.2.1 $Z_{min} = \sum S_{ijk} * SCF_{ijk} * C_{ij}$

where * i = 1...n * j = 1...27 * k = 1...10

`C` is the consumption per unit of an activity while `SCF` is the carbon footprint per source.

C_{ij} is assumed to be **constant** for each individual and activity and SCF_{ijk} is assumed to be **constant** for per source.

In the case that C_{ij} is 0 we will use the big `M` method to enforce a big penalty, this will ensure that the linear programming model won't choose that particular source

```
In [123]: d_vars = []

          for indv in individuals:
              for activity in activities:
                  consumption: np.ndarray = activity_df.loc[activity_df['Activity'] == activity

                  for source in sources:
                      source_cf: np.ndarray = source_df.loc[source_df['Activity'] == activity,
                      source_cf = M if head(source_cf) == 0.0 else head(source_cf)
                      d_vars.append(S_ijk[(to_str(indv, activity), source)] * float(source_cf)


          obj_fn = lp.lpSum(sum(d_vars))
```

2.3 The Constraints

The constraint below ensures that an activity can only use one source. The goal of this constraint is to use a single source that has the lowest carbon footprint.

2.3.1 $\sum_k^m S_{ijk} == 1$

where * n = m = 10 * i = 1...n * j = 1...27

```
In [124]: sum_source_conds = []

          for indv in individuals:
              for activity in activities:
                  sum_source = []
                  for source in sources:
                      sum_source.append(S_ijk[(to_str(indv, activity), source)])

                  sum_source_conds.append(lp.lpSum(sum_source) == 1)
```

We add the objective function and the contraints to the model

```
In [125]: lp_prob = add_obj_fn(lp_prob, obj_fn)
          lp_prob = add_constraint(lp_prob, sum_source_conds)

In [126]: lp_prob.solve()
          lp.LpStatus[lp_prob.status]

Out[126]: 'Optimal'
```

The minimum carbon footprint that can be obtained for 10 individuals is

```
In [127]: lp.value(lp_prob.objective)
```

```
Out[127]: 3.238865999999999

In [128]: columns = list(chain(['Individual'], ["Activity"], sources))

In [134]: rows = []

          for indv in individuals:
              for activity in activities:
                  row = [indv, activity]

                  row.extend([lp.value(S_ijk[(to_str(indv, activity), source)]) for source in s

                  rows.append(row)

          solution = pd.DataFrame(rows, columns=columns)

In [135]: solution

Out[135]:     Individual                                          Activity  \
          0            1                          Household heating => 70F
          1            1                           Household heating < 70F
          2            1                                  Use of heat pump
          3            1                            Use of air conditioner
          4            1                                    shower - short
          5            1                           shower - long (> 3 min)
          6            1                                              bath
          7            1                                           wash-up
          8            1                                use of dishwasher
          9            1                             use of clothes washer
          10           1                              use of clothes dryer
          11           1                              use of cooking range
          12           1                                      use of  oven
          13           1          use of self-clean feature of electric oven
          14           1                 Small kitchen appliance in the home
          15           1                                    TV/computer use
          16           1                           air travel - large plane
          17           1               air travel - small  plane (<50 seats)
          18           1                                car trips- self only
          19           1                         car trips - driver and self
          20           1  car trips - 2+ people with multiple end points
          21           1           trips using public ground transportation
          22           1                           bags of garbage disposed
          23           1           bags of recycling deposited (negative CF)
          24           1             bags of compost deposited (negative CF)
          25           1                hazardous or electric items disposed
          26           1                                large items disposed
          27           2                          Household heating => 70F
          28           2                           Household heating < 70F
          29           2                                  Use of heat pump
```

```
30          2                          Use of air conditioner
31          2                                  shower - short
32          2                         shower - long (> 3 min)
33          2                                            bath
34          2                                         wash-up
35          2                              use of dishwasher
36          2                            use of clothes washer
37          2                             use of clothes dryer
38          2                             use of cooking range
39          2                                     use of  oven
40          2      use of self-clean feature of electric oven
41          2             Small kitchen appliance in the home
42          2                                  TV/computer use
43          2                        air travel - large plane
44          2           air travel - small  plane (<50 seats)
45          2                              car trips- self only
46          2                     car trips - driver and self
47          2  car trips - 2+ people with multiple end points
48          2        trips using public ground transportation
49          2                         bags of garbage disposed
50          2       bags of recycling deposited (negative CF)
51          2         bags of compost deposited (negative CF)
52          2            hazardous or electric items disposed
53          2                              large items disposed

    solar_powered_water_heater  gas_water_heater  \
0                          0.0               0.0
1                          0.0               0.0
2                          0.0               0.0
3                          0.0               0.0
4                          1.0               0.0
5                          1.0               0.0
6                          0.0               0.0
7                          1.0               0.0
8                          0.0               0.0
9                          1.0               0.0
10                         0.0               0.0
11                         0.0               0.0
12                         0.0               0.0
13                         0.0               0.0
14                         0.0               0.0
15                         0.0               0.0
16                         0.0               0.0
17                         0.0               0.0
18                         0.0               0.0
19                         0.0               0.0
20                         0.0               0.0
21                         0.0               0.0
```

22	0.0	0.0
23	0.0	0.0
24	0.0	0.0
25	0.0	0.0
26	0.0	0.0
27	0.0	0.0
28	0.0	0.0
29	0.0	0.0
30	0.0	0.0
31	1.0	0.0
32	1.0	0.0
33	0.0	0.0
34	1.0	0.0
35	0.0	0.0
36	1.0	0.0
37	0.0	0.0
38	0.0	0.0
39	0.0	0.0
40	0.0	0.0
41	0.0	0.0
42	0.0	0.0
43	0.0	0.0
44	0.0	0.0
45	0.0	0.0
46	0.0	0.0
47	0.0	0.0
48	0.0	0.0
49	0.0	0.0
50	0.0	0.0
51	0.0	0.0
52	0.0	0.0
53	0.0	0.0

	electric_water_heater_peak_hour	electric_water_heater_off_peak	gas	\
0	0.0	0.0	0.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	
5	0.0	0.0	0.0	
6	0.0	0.0	0.0	
7	0.0	0.0	0.0	
8	0.0	0.0	0.0	
9	0.0	0.0	0.0	
10	0.0	0.0	0.0	
11	0.0	0.0	0.0	
12	0.0	0.0	0.0	
13	0.0	0.0	0.0	

14	0.0	0.0	0.0
15	0.0	0.0	0.0
16	0.0	0.0	0.0
17	0.0	0.0	0.0
18	0.0	0.0	0.0
19	0.0	0.0	0.0
20	0.0	0.0	0.0
21	0.0	0.0	1.0
22	0.0	0.0	0.0
23	0.0	0.0	0.0
24	0.0	0.0	0.0
25	0.0	0.0	0.0
26	0.0	0.0	0.0
27	0.0	0.0	0.0
28	0.0	0.0	0.0
29	0.0	0.0	0.0
30	0.0	0.0	0.0
31	0.0	0.0	0.0
32	0.0	0.0	0.0
33	0.0	0.0	0.0
34	0.0	0.0	0.0
35	0.0	0.0	0.0
36	0.0	0.0	0.0
37	0.0	0.0	0.0
38	0.0	0.0	0.0
39	0.0	0.0	0.0
40	0.0	0.0	0.0
41	0.0	0.0	0.0
42	0.0	0.0	0.0
43	0.0	0.0	0.0
44	0.0	0.0	0.0
45	0.0	0.0	0.0
46	0.0	0.0	0.0
47	0.0	0.0	0.0
48	0.0	0.0	1.0
49	0.0	0.0	0.0
50	0.0	0.0	0.0
51	0.0	0.0	0.0
52	0.0	0.0	0.0
53	0.0	0.0	0.0

	natural_gas	hybrid	electric_peak_hours	electric_off_peak_hours	\
0	1.0	0.0	0.0	0.0	
1	1.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	
3	1.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	
5	0.0	0.0	0.0	0.0	

6	0.0	0.0	0.0	1.0
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	1.0
9	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	1.0
11	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	1.0
13	0.0	0.0	0.0	1.0
14	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	1.0
16	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0
18	0.0	0.0	1.0	0.0
19	0.0	0.0	1.0	0.0
20	0.0	0.0	1.0	0.0
21	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0
27	1.0	0.0	0.0	0.0
28	1.0	0.0	0.0	0.0
29	0.0	0.0	0.0	1.0
30	1.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0
32	0.0	0.0	0.0	0.0
33	0.0	0.0	0.0	0.0
34	0.0	0.0	0.0	0.0
35	0.0	0.0	0.0	1.0
36	0.0	0.0	0.0	0.0
37	0.0	0.0	0.0	1.0
38	0.0	0.0	0.0	1.0
39	0.0	0.0	0.0	1.0
40	0.0	0.0	0.0	1.0
41	0.0	0.0	0.0	0.0
42	0.0	0.0	0.0	1.0
43	0.0	0.0	0.0	0.0
44	0.0	0.0	0.0	1.0
45	0.0	0.0	1.0	0.0
46	0.0	0.0	1.0	0.0
47	0.0	0.0	1.0	0.0
48	0.0	0.0	0.0	0.0
49	0.0	0.0	0.0	0.0
50	0.0	0.0	0.0	0.0
51	0.0	0.0	0.0	0.0
52	0.0	0.0	0.0	0.0
53	0.0	0.0	0.0	0.0

	jetfuel	waste management
0	0.0	0.0
1	0.0	0.0
2	0.0	1.0
3	0.0	0.0
4	0.0	0.0
5	0.0	0.0
6	0.0	0.0
7	0.0	0.0
8	0.0	0.0
9	0.0	0.0
10	0.0	0.0
11	0.0	1.0
12	0.0	0.0
13	0.0	0.0
14	0.0	1.0
15	0.0	0.0
16	1.0	0.0
17	0.0	1.0
18	0.0	0.0
19	0.0	0.0
20	0.0	0.0
21	0.0	0.0
22	0.0	1.0
23	0.0	1.0
24	0.0	1.0
25	0.0	1.0
26	0.0	1.0
27	0.0	0.0
28	0.0	0.0
29	0.0	0.0
30	0.0	0.0
31	0.0	0.0
32	0.0	0.0
33	0.0	1.0
34	0.0	0.0
35	0.0	0.0
36	0.0	0.0
37	0.0	0.0
38	0.0	0.0
39	0.0	0.0
40	0.0	0.0
41	0.0	1.0
42	0.0	0.0
43	1.0	0.0
44	0.0	0.0
45	0.0	0.0

```
46      0.0                0.0
47      0.0                0.0
48      0.0                0.0
49      0.0                1.0
50      0.0                1.0
51      0.0                1.0
52      0.0                1.0
53      0.0                1.0
```

In [138]:
```
original = pd.read_csv('data/dataset_without_source_cf.csv')

original.drop(['Unnamed: 0'], axis=1, inplace=True)

original
```

Out[138]:

	Indnum	Activity
0	1	Household heating => 70F
1	1	Household heating < 70F
2	1	Use of heat pump
3	1	Use of air conditioner
4	1	shower - short
5	1	shower - long (> 3 min)
6	1	bath
7	1	wash-up
8	1	use of dishwasher
9	1	use of clothes washer
10	1	use of clothes dryer
11	1	use of cooking range
12	1	use of oven
13	1	use of self-clean feature of electric oven
14	1	Small kitchen appliance in the home
15	1	TV/computer use
16	1	air travel - large plane
17	1	air travel - small plane (<50 seats)
18	1	car trips- self only
19	1	car trips - driver and self
20	1	car trips - 2+ people with multiple end points
21	1	trips using public ground transportation
22	1	bags of garbage disposed
23	1	bags of recycling deposited (negative CF)
24	1	bags of compost deposited (negative CF)
25	1	hazardous or electric items disposed
26	1	large items disposed
27	2	Household heating => 70F
28	2	Household heating < 70F
29	2	Use of heat pump
...	...	...
27024	1001	bags of compost deposited (negative CF)

```
27025    1001            hazardous or electric items disposed
27026    1001                          large items disposed
27027    1002                       Household heating => 70F
27028    1002                        Household heating < 70F
27029    1002                               Use of heat pump
27030    1002                         Use of air conditioner
27031    1002                                 shower - short
27032    1002                       shower - long (> 3 min)
27033    1002                                          bath
27034    1002                                       wash-up
27035    1002                            use of dishwasher
27036    1002                          use of clothes washer
27037    1002                           use of clothes dryer
27038    1002                           use of cooking range
27039    1002                                   use of  oven
27040    1002     use of self-clean feature of electric oven
27041    1002             Small kitchen appliance in the home
27042    1002                                TV/computer use
27043    1002                       air travel - large plane
27044    1002          air travel - small  plane (<50 seats)
27045    1002                           car trips- self only
27046    1002                    car trips - driver and self
27047    1002  car trips - 2+ people with multiple end points
27048    1002       trips using public ground transportation
27049    1002                       bags of garbage disposed
27050    1002      bags of recycling deposited (negative CF)
27051    1002        bags of compost deposited (negative CF)
27052    1002           hazardous or electric items disposed
27053    1002                          large items disposed

       solar_powered_water_heater  gas_water_heater  \
0                               0                 0
1                               1                 1
2                               0                 0
3                               0                 0
4                               0                 0
5                               0                 0
6                               0                 0
7                               0                 1
8                               0                 0
9                               0                 0
10                              0                 0
11                              0                 0
12                              0                 0
13                              0                 0
14                              0                 0
15                              0                 0
16                              0                 0
```

```
17                                0                 0
18                                0                 0
19                                0                 0
20                                0                 0
21                                0                 0
22                                0                 0
23                                0                 0
24                                0                 0
25                                0                 0
26                                0                 0
27                                0                 0
28                                0                 0
29                                0                 0
...                             ...               ...
27024                             0                 0
27025                             0                 0
27026                             0                 0
27027                             0                 0
27028                             0                 0
27029                             0                 0
27030                             0                 0
27031                             1                 0
27032                             0                 0
27033                             0                 0
27034                             0                 0
27035                             0                 0
27036                             0                 0
27037                             1                 0
27038                             0                 0
27039                             0                 0
27040                             0                 0
27041                             0                 0
27042                             0                 0
27043                             0                 0
27044                             0                 0
27045                             0                 0
27046                             0                 0
27047                             0                 0
27048                             0                 0
27049                             0                 0
27050                             0                 0
27051                             0                 0
27052                             0                 0
27053                             0                 0

       electric_water_heater_peak_hour  electric_water_heater_off_peak  gas  \
0                                    0                               0    0
1                                    0                               0    0
```

2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	1
19	0	0	1
20	0	0	1
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
...	...	...	...
27024	0	0	0
27025	0	0	0
27026	0	0	0
27027	1	0	0
27028	0	1	0
27029	0	0	0
27030	0	0	0
27031	0	0	0
27032	0	1	0
27033	0	0	0
27034	0	0	0
27035	0	0	0
27036	0	0	0
27037	0	0	0
27038	0	0	0
27039	0	0	0
27040	0	0	0
27041	0	0	0
27042	0	0	0

```
27043                          0                          0    0
27044                          0                          0    0
27045                          0                          0    1
27046                          0                          0    1
27047                          0                          0    1
27048                          0                          0    0
27049                          0                          0    0
27050                          0                          0    0
27051                          0                          0    0
27052                          0                          0    0
27053                          0                          0    0

       natural_gas  hybrid  electric_peak_hours  electric_off_peak_hours  \
0                1       0                    0                        0
1                1       0                    0                        0
2                0       0                    0                        0
3                0       0                    0                        0
4                0       0                    0                        0
5                0       0                    0                        0
6                0       0                    0                        0
7                0       0                    0                        0
8                0       0                    0                        0
9                0       0                    0                        0
10               0       0                    0                        0
11               0       0                    0                        0
12               0       0                    0                        0
13               0       0                    0                        0
14               0       0                    0                        0
15               0       0                    0                        0
16               0       0                    0                        0
17               0       0                    0                        0
18               0       0                    0                        0
19               0       0                    0                        0
20               0       0                    0                        0
21               0       0                    0                        0
22               0       0                    0                        0
23               0       0                    0                        0
24               0       0                    0                        0
25               0       0                    0                        0
26               0       0                    0                        0
27               1       0                    0                        0
28               1       0                    0                        0
29               0       0                    0                        0
...            ...     ...                  ...                      ...
27024            0       0                    0                        0
27025            0       0                    0                        0
27026            0       0                    0                        0
27027            1       0                    0                        0
```

27028	1	0	0	0
27029	0	0	0	0
27030	0	0	0	0
27031	0	1	0	0
27032	0	0	0	0
27033	0	0	0	0
27034	0	0	0	0
27035	0	0	0	0
27036	0	0	0	0
27037	0	0	0	1
27038	0	0	0	1
27039	0	0	0	1
27040	0	0	0	0
27041	0	0	0	0
27042	0	0	0	0
27043	0	0	0	0
27044	0	0	0	0
27045	0	0	0	0
27046	0	0	0	0
27047	0	0	0	0
27048	0	0	0	0
27049	0	0	0	0
27050	0	0	0	0
27051	0	0	0	0
27052	0	0	0	0
27053	0	0	0	0

	jetfuel
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0

```
20           0
21           0
22           0
23           0
24           0
25           0
26           0
27           0
28           0
29           0
...        ...
27024        0
27025        0
27026        0
27027        0
27028        0
27029        0
27030        0
27031        0
27032        0
27033        0
27034        0
27035        0
27036        0
27037        0
27038        0
27039        0
27040        0
27041        0
27042        0
27043        0
27044        0
27045        0
27046        0
27047        0
27048        0
27049        0
27050        0
27051        0
27052        0
27053        0

[27054 rows x 12 columns]
```