Week_2_Assignment_Raj_Ponnam

January 12, 2023

```
[1]: # Import libraries
     import csv
     #import xlrd
     import pandas as pd
     import matplotlib.pyplot as plt
     from datetime import datetime as dt
```

```
[18]: # Read world population data
      fileData1 = 'world-population.xlsm'
      population = pd.read_excel(fileData1)

      # Read AEP data
      fileData2 = 'AEP_hourly.csv'
      aep = pd.read_csv(fileData2)

      # Read sample superstore data
      fileData3 = 'Superstore.xls'
      superstore = pd.read_excel(fileData3)

      print(population.head())
      print(aep.head())
      print(superstore.head())
```

```
   Year  Population
0  1960  3028654024
1  1961  3068356747
2  1962  3121963107
3  1963  3187471383
4  1964  3253112403
             Datetime   AEP_MW
0  2004-12-31 01:00:00  13478.0
1  2004-12-31 02:00:00  12865.0
2  2004-12-31 03:00:00  12577.0
3  2004-12-31 04:00:00  12517.0
4  2004-12-31 05:00:00  12670.0
   Row ID        Order ID Order Date  Ship Date      Ship Mode Customer ID  \
0       1  CA-2016-152156 2016-11-08 2016-11-11   Second Class   CG-12520
1       2  CA-2016-152156 2016-11-08 2016-11-11   Second Class   CG-12520
```

```
2        3  CA-2016-138688  2016-06-12  2016-06-16    Second Class     DV-13045
3        4  US-2015-108966  2015-10-11  2015-10-18  Standard Class     SO-20335
4        5  US-2015-108966  2015-10-11  2015-10-18  Standard Class     SO-20335

      Customer Name      Segment        Country                City  …  \
0       Claire Gute     Consumer  United States          Henderson  …
1       Claire Gute     Consumer  United States          Henderson  …
2    Darrin Van Huff    Corporate  United States        Los Angeles  …
3    Sean O'Donnell     Consumer  United States    Fort Lauderdale  …
4    Sean O'Donnell     Consumer  United States    Fort Lauderdale  …

   Postal Code  Region       Product ID         Category Sub-Category  \
0        42420   South  FUR-BO-10001798        Furniture    Bookcases
1        42420   South  FUR-CH-10000454        Furniture       Chairs
2        90036    West  OFF-LA-10000240  Office Supplies       Labels
3        33311   South  FUR-TA-10000577        Furniture       Tables
4        33311   South  OFF-ST-10000760  Office Supplies      Storage

                                        Product Name     Sales  Quantity  \
0                   Bush Somerset Collection Bookcase   261.9600         2
1  Hon Deluxe Fabric Upholstered Stacking Chairs,…   731.9400         3
2  Self-Adhesive Address Labels for Typewriters b…    14.6200         2
3     Bretford CR4500 Series Slim Rectangular Table   957.5775         5
4                   Eldon Fold 'N Roll Cart System    22.3680         2

   Discount     Profit
0      0.00    41.9136
1      0.00   219.5820
2      0.00     6.8714
3      0.45  -383.0310
4      0.20     2.5164

[5 rows x 21 columns]
```
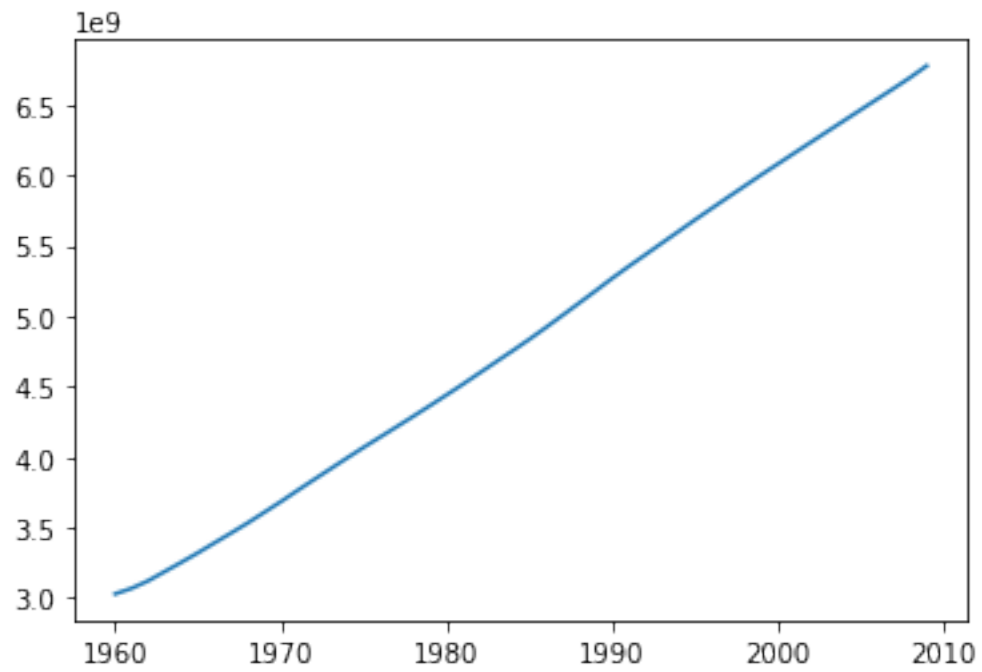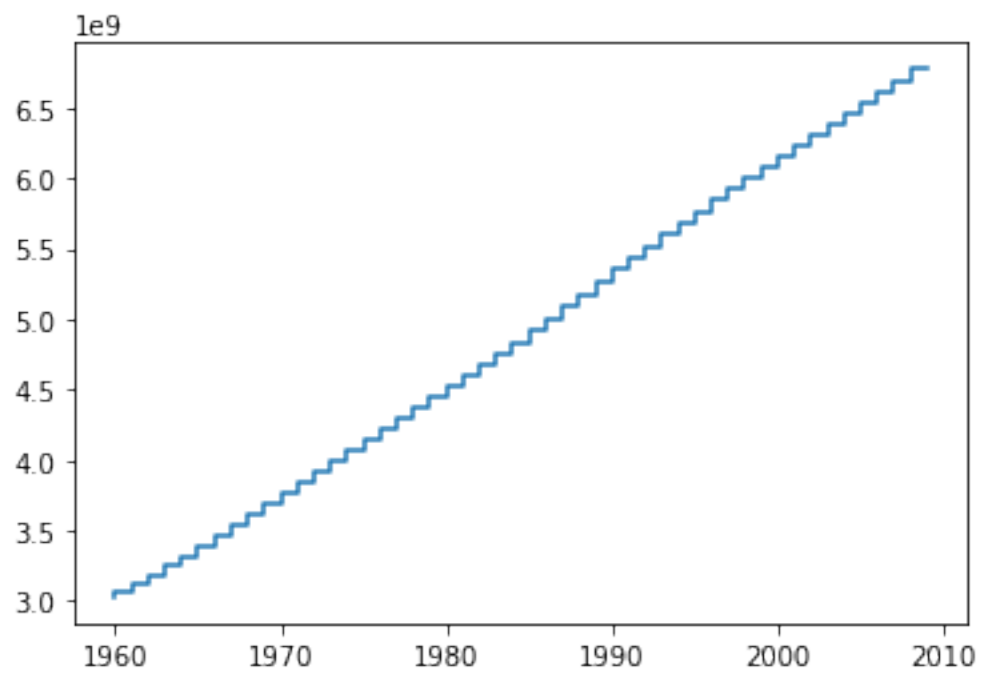
1 World Population data

Line and Step graph

```python
[9]: X = population['Year']
     Y = population['Population']
     plt.plot(X, Y)
     plt.show()
```

```
[10]: plt.step(X, Y)
      plt.show()
```

2 AEP Data

Line and Step graph

For the AEP data, there are too many observations to put in a line plot. Hence I calculated the total of AEP_MW for each year and plotted them

```python
[11]: # Convert pandas column to datetime
      aep['Datetime'] = pd.to_datetime(aep['Datetime'])

      # Extract Year and store in a new column
      aep['Year'] = aep['Datetime'].dt.year

      # Calculate sum of MW for each year
      aep_aggr = aep.groupby(['Year'])['AEP_MW'].agg('sum').
       →reset_index(name='Total_MW')
      aep_aggr.head()
```
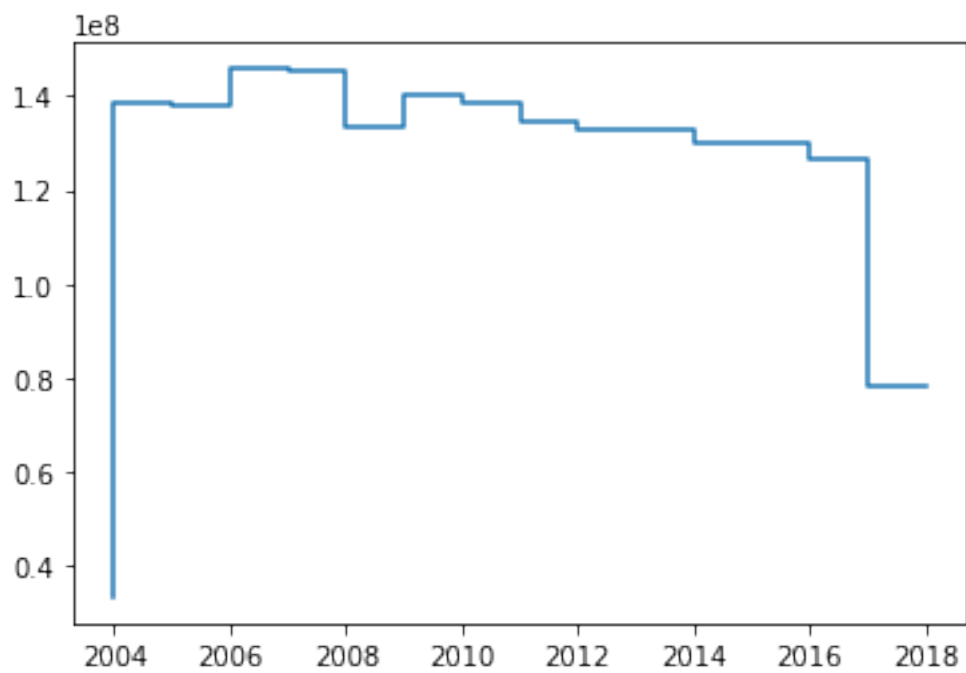
```
[11]:    Year       Total_MW
      0  2004    33479854.0
      1  2005   138752914.0
      2  2006   137826610.0
      3  2007   145781458.0
      4  2008   145224910.0
```

```python
[12]: X = aep_aggr['Year']
      Y = aep_aggr['Total_MW']
      plt.plot(X, Y)
      plt.show()
```
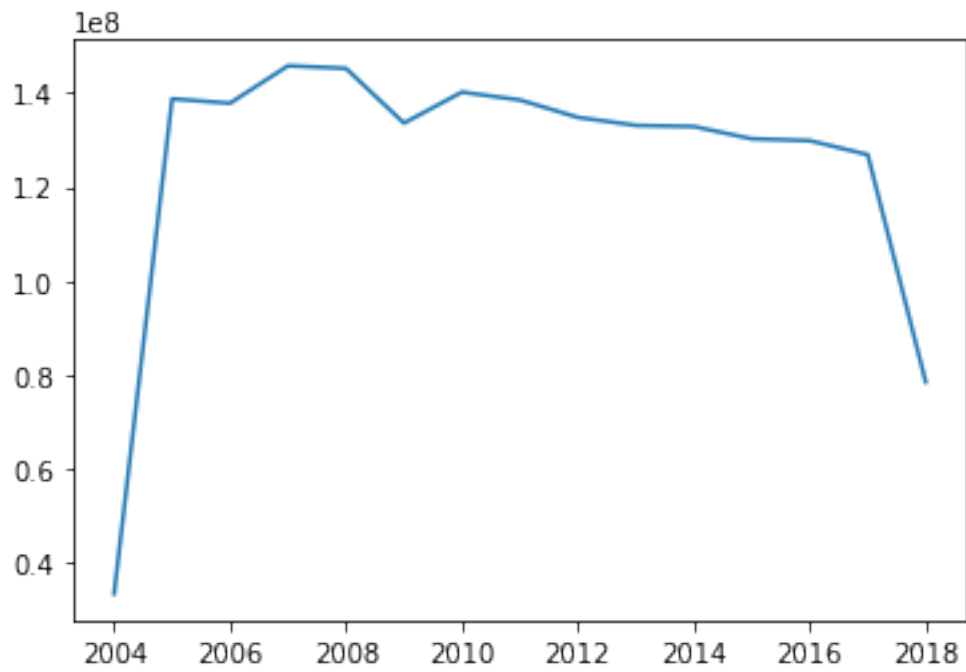
```
[13]: plt.step(X, Y)
      plt.show()
```

3 Sample Superstore data

Line and Step graph

For this data set, I plan to plot the profit by order year and segment. This needs some additional data preparation as below.

```
[19]: # Extract order year
      superstore['Order Year'] = superstore['Order Date'].dt.year

      # Calculate total profit for each segment each year
      superstore_aggr = superstore.groupby(['Order Year'])['Profit'].agg('sum').
       →reset_index(name='Total Profit')

      # Check data
      superstore_aggr.head()
```
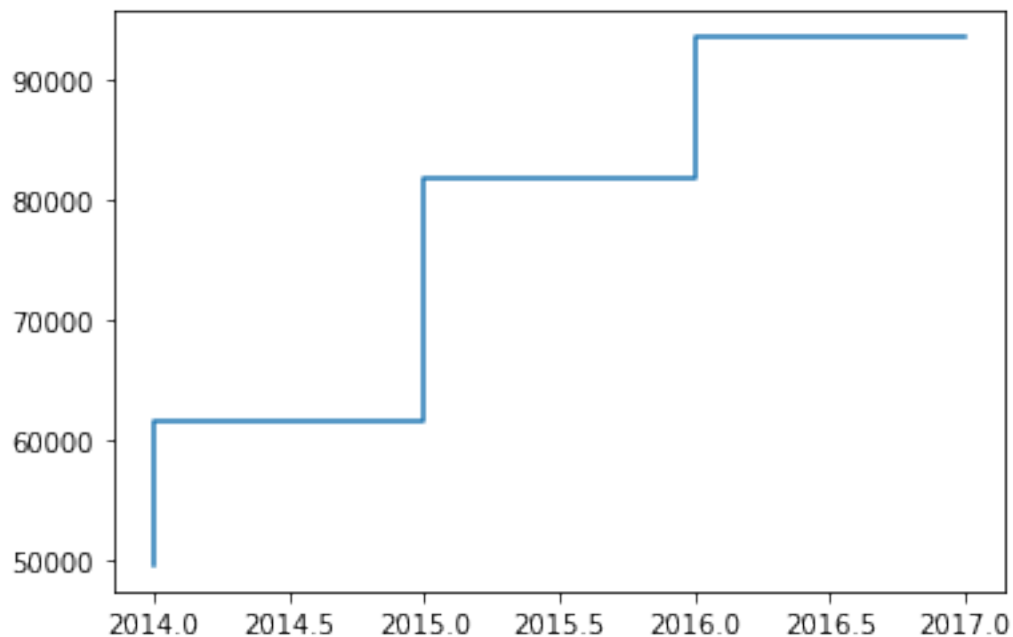
```
[19]:    Order Year  Total Profit
      0        2014    49543.9741
      1        2015    61618.6037
      2        2016    81795.1743
      3        2017    93439.2696
```

```
[20]: X = superstore_aggr['Order Year']
      Y = superstore_aggr['Total Profit']
      plt.plot(X, Y)
      plt.show()
```
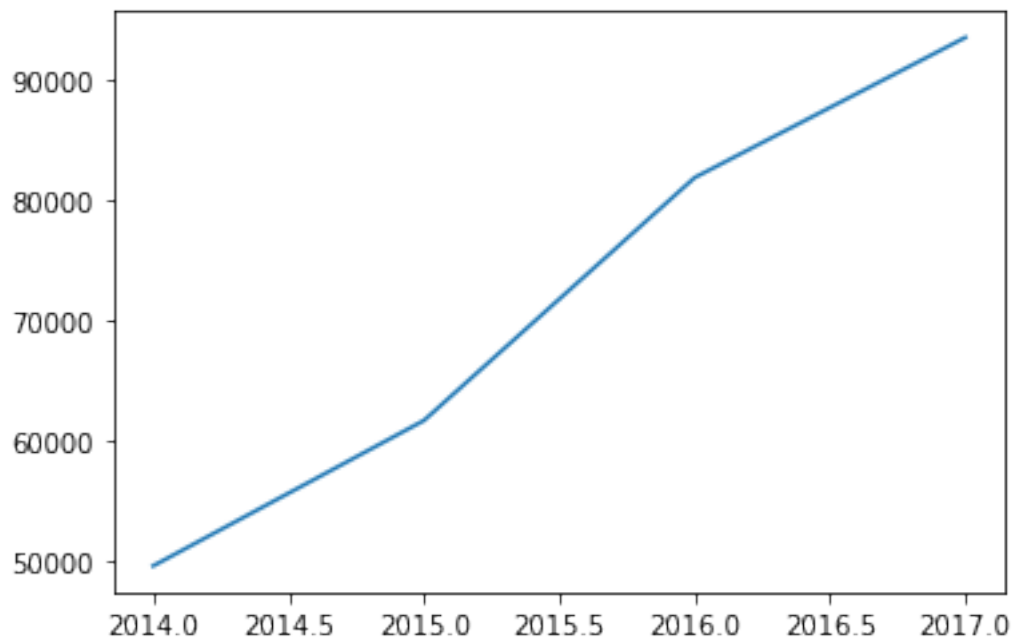
```
[21]: plt.step(X, Y)
      plt.show()
```



```
[ ]: # Import required packages
     library('magrittr')

     # Import data to be used for visualization
     fileData1 = paste(getwd(), '/Users/rajponnam/Documents/ds_course/dsc640/Week2/
       ↪world-population.xlsm', sep = '')
     population = xlsx::read.xlsx(fileData1, sheetIndex = 1, stringsAsFactors =␣
       ↪FALSE)

     fileData2 = paste(getwd(), '/Users/rajponnam/Documents/ds_course/dsc640/Week2/
       ↪AEP_hourly.csv', sep = '')
     aep = read.csv2(fileData2, sep=',', stringsAsFactors = FALSE) %>%
             as.data.frame()

     fileData3 = paste(getwd(), '/Users/rajponnam/Documents/ds_course/dsc640/Week2/
       ↪Superstore.xls', sep = '')
     superstore = xlsx::read.xlsx(fileData3, sheetIndex = 1, stringsAsFactors =␣
       ↪FALSE)

     # Examine data
     print(head(population))
     print(head(aep))
```

```
print(head(superstore))
```