

```
In [ ]:
Week 1 & 2 Exercises
"""
from future import print_function
from itertools import zip_longest

import csv
import logging
import sys
import numpy as np
import pandas as pd
import random
import thinkplot
import thinkstats2
import datetime
import regression
import statsmodels.formula.api as smf
import statsmodels.api as sm
import matplotlib.pyplot as plt
import math

def ReadData(filename='world-population.xlsx'):
    """Reads filename and returns project data in thousands

    filename: string

    returns: pandas Series of projectdata in thousands
    """
    projectData = pd.read_excel(filename)
    return projectData

def convert_to_number(number_stack):
    final_number = 0
    for i in range(0, len(number_stack)):
        final_number += (number_stack.pop() * (math.pow(10, i)))
    return final_number

def return_dict_from_csv_line(header, line):
    # zip them
    zipped_line = zip_longest(header, line, fillvalue=None)
    # Use Dict comprehension to generate the final dict
    ret_dict = {kv[0]: kv[1] for kv in zipped_line}
    return ret_dict

def Draw_LineChart(data):
    # to read the population Data
    Year_column_name = "Year"
    Population_column_name = "Population"
    year_data = data[Year_column_name]
    population_data = data[Population_column_name]
    plt.plot(year_data, population_data)

def DataWranglingwithPython_Activity1():
    LIMIT = 100
    random_number_list = [random.randint(0, LIMIT) for x in range(0, LIMIT)]
    print ("random_number_list : ", random_number_list)

    list_with_divisible_by_3 = [a for a in random_number_list if a % 3 == 0]
    print ("list_with_divisible_by_3 : ", list_with_divisible_by_3)

    length_of_random_list = len(random_number_list)
    length_of_3_divisible_list = len(list_with_divisible_by_3)
    difference = length_of_random_list - length_of_3_divisible_list
    print("difference : ", difference)

    NUMBER_OF_EXPERIMENTS = 10
    difference_list = []
    for i in range(0, NUMBER_OF_EXPERIMENTS):
        random_number_list = [random.randint(0, LIMIT) for x in range(0, LIMIT)]
        list_with_divisible_by_3 = [a for a in random_number_list if a % 3 == 0]
        length_of_random_list = len(random_number_list)
        length_of_3_divisible_list = len(list_with_divisible_by_3)
        difference = length_of_random_list - length_of_3_divisible_list
        difference_list.append(difference)

    print("difference_list : ", difference_list)
    avg_diff = sum(difference_list) / float(len(difference_list))
    print("avg_diff : ", avg_diff)

def DataWranglingwithPython_Activity2():

    multiline_text = """ It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.
    However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered the rightful property of some c
    "My dear Mr. Bennet," said his lady to him one day, "have you heard that Netherfield Park is let at last?"
    Mr. Bennet replied that he had not.
    "But it is," returned she, "Mr. Long has just been here, and she told me all about it."
    Mr. Bennet was so odd a mixture of quick parts, sarcastic humour, reserve, and caprice, that the experience of three-and-twenty years had been insufficient to make his wife understand his character. Her mind was less difficult
    type(multiline_text)
    length = len(multiline_text)
    print("String Length : ", length)
    multiline_text = multiline_text.replace('\n', '')
    print("multiline_text :", multiline_text)

    # remove special chars, punctuation etc.
    cleaned_multiline_text = ""
    for char in multiline_text:
        if char == " ":
            cleaned_multiline_text += char
        elif char.isalnum(): # using the isalnum() method of strings.
            cleaned_multiline_text += char
        else:
            cleaned_multiline_text += " "
    print("cleaned_multiline_text :", cleaned_multiline_text)

    list_of_words = cleaned_multiline_text.split()
    print("list_of_words :", list_of_words)
    list_of_words_length = len(list_of_words)
    print("list_of_words_length :", list_of_words_length)

    # Use set to get unique words
    unique_words_as_list = list(set(list_of_words))
    unique_words_as_list_length = len(unique_words_as_list)
    print("unique_words_as_list_length :", unique_words_as_list_length)

def DataWranglingwithPython_Activity3():
    from itertools import permutations, dropwhile
    permutations(range(3))

    for number_tuple in permutations(range(3)):
        print("number_tuple :", number_tuple)
        assert isinstance(number_tuple, tuple)

    for number_tuple in permutations(range(3)):
        print(list(dropwhile(lambda x: x <= 0, number_tuple)))

    for number_tuple in permutations(range(3)):
        number_stack = list(dropwhile(lambda x: x <= 0, number_tuple))
        print("number_stack :", convert_to_number(number_stack))

def DataWranglingwithPython_Activity4():
    with open("sales_record.csv", "r") as fd:
        first_line = fd.readline()
        header = first_line.replace("\n", "").split(",")
        for i, line in enumerate(fd):
            # Here we loop over the first 10 lines in order to not to make the output too big
            line = line.replace("\n", "").split(",")
            d = return_dict_from_csv_line(header, line)
            print(d)
            if i > 10:
                break

def main():
    print('Inside Main function')
    data = ReadData()
    #print(data):
    #Create a line chart with Matplotlib using Data file: world-population.xlsx
    Draw_LineChart(data);

    # Activity 1: Handling Lists
    #1. Create a list of 100 random numbers.
    #2. Create a new list from this random list, with numbers that are divisible by 3.
    #3. Calculate the length of these two lists and store the difference in a new variable.
    #4. Using a loop, perform steps 2 and 3 and find the difference variable three times.
    #5. Find the arithmetic mean of these three difference values.
    DataWranglingwithPython_Activity1()

    # Activity 2: Analyze a Multiline String and Generate the Unique Word Count
    #1. Get multiline text and save it in a Python variable
    #2. Get rid of all new lines in it using string methods
    #3. Get all the unique words and their occurrences from the string
    #4. Repeat the step to find all unique words and occurrences, without considering case sensitivity
    DataWranglingwithPython_Activity2()

    #1. Look up the definition of permutations and dropwhile from itertools.
    #2. Write an expression to generate all the possible three-digit numbers using 0, 1, and 2.
    #3. Loop over the iterator expression you generated before. Print each element that's returned by the iterator. Use assert and isinstance to make sure that the elements are of the tuple type.
    #4. Write the loop again using dropwhile with a lambda expression to drop any leading zeros from the tuples. As an example, (0, 1, 2) will become (0, 2). Also, cast the output of dropwhile to a list.
    #5. Check the actual type that dropwhile returns.
    #6. Combine the preceding code into one block, and this time write a separate function where you will pass the list generated from dropwhile, and the function will return the whole number contained in the list. As an example,
    #if you pass [1, 2] to the function, it will return 12. Make sure that the return type is indeed a number and not a string. Although this task can be achieved using other tricks, we
    #require that you treat the incoming list as a stack in the function and generate the number by reading the individual digits from the stack.
    DataWranglingwithPython_Activity3()

    #1. Import zip_longest from itertools. Create a function to zip header, line and fillvalue=None.
    #2. Open the accompanying sales_record.csv file from the GitHub Link by using rmode inside a with block and first check that it is opened.
    #3. Read the first line using methods to generate a list of all the column names.
    #4. Start reading the file. Read it line by line.
    #5. Read each line and pass that line to a function, along with the list of the headers. The work of the function is to construct a dict out of these two and fill up the
    #keyvalues. Keep in mind that a missing value should result in None.
    DataWranglingwithPython_Activity4()

if __name__ == "__main__":
    main()

Inside Main function
random_number_list : [67, 25, 44, 25, 15, 12, 12, 48, 68, 32, 100, 59, 74, 83, 7, 67, 15, 66, 25, 62, 33, 88, 69, 13, 28, 93, 64, 32, 85, 40, 69, 11, 57, 15, 66, 89, 88, 36, 90, 73, 80, 87, 41, 62, 7, 45, 52, 0, 67, 16, 99, 16,
84, 97, 91, 24, 91, 56, 62, 31, 60, 45, 89, 63, 62, 4, 80, 32, 61, 70, 95, 67, 9, 92, 48, 33, 0, 23, 55, 17, 9, 28, 79, 5, 39, 36, 30, 28, 57, 16, 19, 37, 46, 61, 9, 21]
list_with_divisible_by_3 : [15, 12, 12, 48, 15, 66, 33, 69, 93, 69, 57, 15, 66, 36, 90, 87, 45, 0, 99, 84, 24, 9, 60, 45, 63, 9, 48, 33, 0, 9, 36, 30, 57, 9, 21]
difference : 65
avg_diff : 66.6
String Length : 1102
multiline_text : It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife. However little known the feelings or views of such a man may be on his first entering a nei
ghbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered the rightful property of some one or other of their daughters. "My dear Mr. Bennet," said his lady to him one day, "have y
ou heard that Netherfield Park is let at last?" Mr. Bennet replied that he had not. "But it is," returned she, "for Mrs. Long has just been here, and she told me all about it." Mr. Bennet was so odd a mixture of quick pa
rts, sarcastic humour, reserve, and caprice, that the experience of three-and-twenty years had been insufficient to make his wife understand his character. Her mind was less difficult to develop. She was a woman of mean understan
ding, little information, and uncertain temper. When she was discontented, she fancied herself nervous. The business of her life was to get her daughters married; its solace was visiting and news.
cleaned_multiline_text : It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife However little known the feelings or views of such a man may be on his first enter
ing a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful property of some one or other of their daughters My dear Mr Bennet said his lady to him one day
have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned she for Mrs Long has just been here and she told me all about it Mr Bennet was so odd a mixture of qu
ick parts sarcastic humour reserve and caprice that the experience of three and twenty years had been insufficient to make his wife understand his character Her mind was less difficult to develop She was a woman of mean und
erstanding little information and uncertain temper When she was discontented she fancied herself nervous The business of her life was to get her daughters married its solace was visiting and news
list_of_words : ['It', 'is', 'a', 'truth', 'universally', 'acknowledged', 'that', 'a', 'single', 'man', 'in', 'possession', 'of', 'a', 'good', 'fortune', 'must', 'be', 'in', 'want', 'of', 'a', 'wife', 'However', 'little', 'know
n', 'the', 'feelings', 'or', 'views', 'of', 'such', 'a', 'man', 'may', 'be', 'on', 'his', 'first', 'entering', 'a', 'neighbourhood', 'this', 'truth', 'is', 'so', 'well', 'fixed', 'in', 'the', 'minds', 'of', 'the', 'surrounding',
'families', 'that', 'he', 'is', 'considered', 'the', 'rightful', 'property', 'of', 'some', 'one', 'or', 'other', 'of', 'their', 'daughters', 'My', 'dear', 'Mr', 'Bennet', 'said', 'his', 'lady', 'to', 'him', 'one', 'day', 'have',
'you', 'heard', 'that', 'Netherfield', 'Park', 'is', 'let', 'at', 'last', 'Mr', 'Bennet', 'replied', 'that', 'he', 'had', 'not', 'But', 'it', 'is', 'returned', 'she', 'for', 'Mrs', 'Long', 'has', 'just', 'been', 'here', 'and', 's
he', 'told', 'me', 'all', 'about', 'It', 'Mr', 'Bennet', 'was', 'so', 'odd', 'a', 'mixture', 'of', 'quick', 'parts', 'sarcastic', 'humour', 'reserve', 'and', 'caprice', 'that', 'the', 'experience', 'of', 'three', 'and', 'twenty',
'years', 'had', 'been', 'insufficient', 'to', 'make', 'his', 'wife', 'understand', 'his', 'character', 'Her', 'mind', 'was', 'less', 'difficult', 'to', 'develop', 'She', 'was', 'a', 'woman', 'of', 'mean', 'understanding', 'littl
e', 'information', 'and', 'uncertain', 'temper', 'When', 'she', 'was', 'discontented', 'she', 'fancied', 'herself', 'nervous', 'The', 'business', 'of', 'her', 'life', 'was', 'to', 'get', 'her', 'daughters', 'married', 'its', 'sol
ace', 'was', 'visiting', 'and', 'news']
list_of_words_length : 194
unique_words_as_list_length : 128
number_tuple : (0, 1, 2)
number_tuple : (0, 2, 1)
number_tuple : (1, 0, 2)
number_tuple : (1, 2, 0)
number_tuple : (2, 0, 1)
number_tuple : (2, 1, 0)
[1, 2]
[2, 1]
[1, 0, 2]
[1, 2, 0]
[2, 0, 1]
[2, 1, 0]
number_stack : 12.0
number_stack : 21.0
number_stack : 102.0
number_stack : 120.0
number_stack : 201.0
number_stack : 210.0
('Region': 'Central America and the Caribbean', 'Country': 'Antigua and Barbuda ', 'Item Type': 'Baby Food', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date': '12/20/2013', 'Order ID': '957081544', 'Ship Date': '1/1
7/2014', 'Units Sold': '552', 'Unit Price': '255.28', 'Unit Cost': '159.42', 'Total Revenue': '140914.56', 'Total Cost': '87999.84', 'Total Profit': '52914.72')
('Region': 'Central America and the Caribbean', 'Country': 'Panama', 'Item Type': 'Snacks', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '7/5/2010', 'Order ID': '301644504', 'Ship Date': '7/26/2010', 'Units So
ld': '12167', 'Unit Price': '152.58', 'Unit Cost': '97.44', 'Total Revenue': '1854591.20', 'Total Cost': '1058663.76', 'Total Profit': '798727.44')
('Region': 'Europe', 'Country': 'Czech Republic', 'Item Type': 'Beverages', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '9/12/2011', 'Order ID': '478051030', 'Ship Date': '9/29/2011', 'Units Sold': '4778', 'U
nit Price': '47.45', 'Unit Cost': '31.79', 'Total Revenue': '226716.10', 'Total Cost': '151892.62', 'Total Profit': '74823.48')
('Region': 'Asia', 'Country': 'North Korea', 'Item Type': 'Cereal', 'Sales Channel': 'Offline', 'Order Priority': 'L', 'Order Date': '5/13/2010', 'Order ID': '892599952', 'Ship Date': '6/15/2010', 'Units Sold': '9016', 'Unit Pri
ce': '205.70', 'Unit Cost': '117.11', 'Total Revenue': '1854591.20', 'Total Cost': '1058663.76', 'Total Profit': '798727.44')
('Region': 'Asia', 'Country': 'Sri Lanka', 'Item Type': 'Snacks', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '7/20/2015', 'Order ID': '571902596', 'Ship Date': '7/27/2015', 'Units Sold': '7542', 'Unit Pri
ce': '152.58', 'Unit Cost': '97.44', 'Total Revenue': '1150758.36', 'Total Cost': '734892.48', 'Total Profit': '415865.88')
('Region': 'Middle East and North Africa', 'Country': 'Morocco', 'Item Type': 'Personal Care', 'Sales Channel': 'Offline', 'Order Priority': 'L', 'Order Date': '11/8/2010', 'Order ID': '412882792', 'Ship Date': '11/22/2010', 'Uni
ts Sold': '48', 'Unit Price': '181.73', 'Unit Cost': '56.67', 'Total Revenue': '8523.04', 'Total Cost': '2720.16', 'Total Profit': '1202.88')
('Region': 'Australia and Oceania', 'Country': 'Federated States of Micronesia', 'Item Type': 'Clothes', 'Sales Channel': 'Offline', 'Order Priority': 'H', 'Order Date': '3/28/2011', 'Order ID': '932776868', 'Ship Date': '5/10/20
11', 'Units Sold': '8258', 'Unit Price': '109.28', 'Unit Cost': '35.84', 'Total Revenue': '902434.24', 'Total Cost': '295966.72', 'Total Profit': '606467.52')
('Region': 'Europe', 'Country': 'Bosnia and Herzegovina', 'Item Type': 'Clothes', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date': '10/14/2013', 'Order ID': '919133651', 'Ship Date': '11/4/2013', 'Units Sold': '92
0', 'Unit Price': '109.28', 'Unit Cost': '35.84', 'Total Revenue': '966144.48', 'Total Cost': '316861.44', 'Total Profit': '649283.04')
('Region': 'Europe', 'Country': 'Bosnia and Herzegovina', 'Item Type': 'Clothes', 'Sales Channel': 'Offline', 'Order Priority': 'M', 'Order Date': '8/27/2016', 'Order ID': '579814469', 'Ship Date': '10/5/2016', 'Units
Sold': '8841', 'Unit Price': '109.28', 'Unit Cost': '35.84', 'Total Revenue': '966144.48', 'Total Cost': '316861.44', 'Total Profit': '649283.04')
('Region': 'Sub-Saharan Africa', 'Country': 'Ethiopia', 'Item Type': 'Baby Food', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date': '4/13/2015', 'Order ID': '192993152', 'Ship Date': '5/7/2015', 'Units Sold': '981
7', 'Unit Price': '255.28', 'Unit Cost': '159.42', 'Total Revenue': '2506083.76', 'Total Cost': '1565026.14', 'Total Profit': '94057.62')
('Region': 'Middle East and North Africa', 'Country': 'Turkey', 'Item Type': 'Office Supplies', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '9/25/2013', 'Order ID': '557156026', 'Ship Date': '10/15/2013', 'Un
its Sold': '3704', 'Unit Price': '651.21', 'Unit Cost': '524.96', 'Total Revenue': '2412081.84', 'Total Cost': '1944451.84', 'Total Profit': '467630.00')
('Region': 'Middle East and North Africa', 'Country': 'Oman', 'Item Type': 'Cosmetics', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date': '5/12/2013', 'Order ID': '741101920', 'Ship Date': '5/17/2013', 'Units Sold':
'7382', 'Unit Price': '437.20', 'Unit Cost': '263.33', 'Total Revenue': '3227410.40', 'Total Cost': '1943902.06', 'Total Profit': '1283508.34'}
```

