Weeks 1 & 2 Exercises

Name : Ragunath Gunasekaran

1. Create a Jupyter notebook where you create a list, iterate over the list and sort your results, generate random numbers, add to the list, and then print your results

In [26]:

```
# Create a list
listname1 = [1,60,80,35,49,23,68,85,24,76,38,92]
print("\n Elements in the listname1 List are :")
print(listname1)

# iterate over the list by using while and Forloop
# Geting the length of the list
lengthoflist = len(listname1)
i = 0

# Sorting the List
listname1.sort()

print("############ While Loop - Iteration ##########################")
while i < lengthoflist:
    print(listname1[i])
    i += 1
'''
print("############ For Loop ##########################")
for i in range(lengthoflist):
    print(listname1[i])
'''
```

```
 Elements in the listname1 List are :
[1, 60, 80, 35, 49, 23, 68, 85, 24, 76, 38, 92]
############ While Loop - Iteration ##########################
1
23
24
35
38
49
60
68
76
80
85
92
```

Out[26]:

```
'\nprint("############ For Loop ##########################")   \nfor i in ra
nge(lengthoflist):\n    print(listname1[i])\n'
```

In [36]:
```
import random

#generate random numbers
limitofnumbers = 30
listname2 = [random.randint(0, limitofnumbers) for x in range(0, limitofr

# using list.extend() to concat
listname1.extend(listname2)
listname1.sort()

# Printing concatenated list
print ("Concatenated the two lists - with Duplicates")
print (str(listname1))
# Removing Duplicates
print ("Concatenated the two lists - Removed Duplicates")
print (list(dict.fromkeys(listname1)))
```

```
Concatenated the two lists - with Duplicates
[0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5,
5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8,
8, 8, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11,
11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13, 13, 14,
14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 17, 17,
17, 18, 18, 18, 18, 18, 18, 18, 18, 18, 19, 19, 19, 20, 20, 20, 20, 21, 21,
21, 21, 21, 21, 22, 22, 22, 22, 22, 22, 23, 23, 23, 23, 24, 24, 24, 24, 24,
24, 24, 24, 25, 25, 25, 25, 25, 25, 25, 26, 26, 26, 26, 27, 27, 27, 27, 27,
27, 27, 28, 28, 28, 28, 28, 28, 28, 28, 29, 29, 29, 29, 30, 30, 30, 35, 38,
49, 60, 68, 76, 80, 85, 92]
Concatenated the two lists - Removed Duplicates
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 38, 49, 60, 68, 76, 80, 85, 92]
```

2. Create a line chart with Matplotlib and the following data file.

a. Data file: world-population.xlsm

b. (Hint: Python for Data Analysis: Page 19-50 & Data Wrangling with Python: Preface)

In [45]:
```
import pandas as pd

# Data worldpopulation Data
WorldPopulation_DF = pd.read_excel('C:/Users/ragun/Documents/GitHub/DSC54

#verify any Nulls
WorldPopulation_DF.isnull().sum()
```

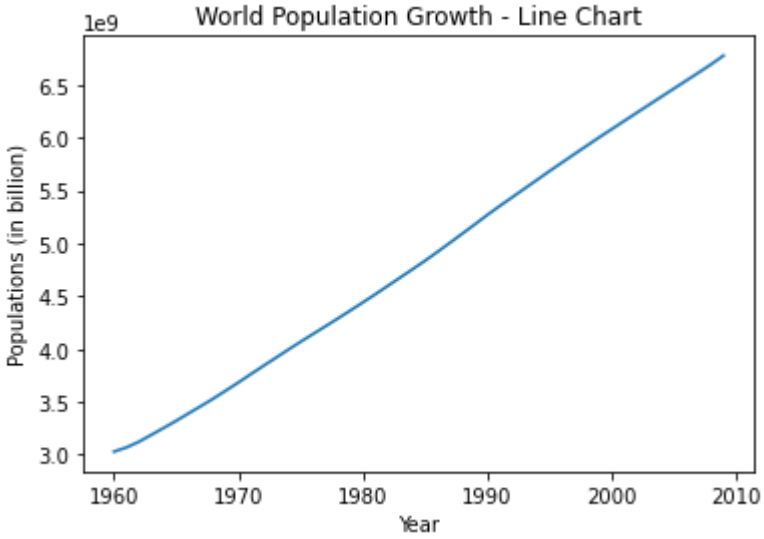
Out[45]:
```
Year          0
Population    0
dtype: int64
```

In [44]:

```
import matplotlib.pyplot as plt

plt.plot(WorldPopulation_DF.Year, WorldPopulation_DF.Population)
plt.xlabel("Year")  # add X-axis label
plt.ylabel("Populations (in billion)")  # add Y-axis label
plt.title("World Population Growth - Line Chart")  # add title
plt.show()

```



3. Complete the following activities:

a. Data Wrangling with Python: Activity 1 page 17

Create a list of random numbers and then create another list from this one whose elements are divisible by three. Also repeat the experiment few times (at least three times) and calculate the arithimetic mean of the differenence of length of the two lists

Task 1 - Create list of 100 Random Numbers

In [67]:
```
1 LimitNumbers = 100
2 ListRandomNumber = [random.randint(0, LimitNumbers) for x in range(0, Li
3 ListRandomNumber
```

Out[67]:
```
[46,
 52,
 36,
 62,
 49,
 91,
 37,
 93,
 88,
 6,
 94,
 92,
 10,
 9,
 89,
 26,
 70,
 79,
 43,
 15,
 36,
 76,
 92,
 82,
 22,
 72,
 68,
 22,
 37,
 61,
 77,
 87,
 29,
 31,
 59,
 88,
 91,
 39,
 54,
 8,
 10,
 88,
 98,
 23,
 20,
 47,
 81,
 95,
 28,
 44,
 5,
 54,
 17,
```

```
10,
5,
7,
26,
43,
59,
31,
14,
84,
68,
74,
77,
58,
80,
18,
72,
20,
87,
25,
86,
16,
21,
77,
67,
21,
91,
100,
54,
8,
53,
43,
8,
72,
93,
48,
98,
50,
20,
42,
33,
25,
25,
94,
20,
19,
61,
36]
```

Task-2

Generate a second list from the above one. The condition of membership in the second list is divisibility by 3.

In [68]:
```
ListDivByThree=[number for number in ListRandomNumber if number % 3 == 0]
ListDivByThree
```

Out[68]:
```
[36,
 93,
 6,
 9,
 15,
 36,
 72,
 87,
 39,
 54,
 81,
 54,
 84,
 18,
 72,
 87,
 21,
 21,
 54,
 72,
 93,
 48,
 42,
 33,
 36]
```

Task-3

- Use the `len` function to measure the length of the first list and the second list
- Store both in two different variables
- Calculate the difference of length between them

In [69]:
```
DiffOfLen = len(ListRandomNumber) - len(ListDivByThree)
DiffOfLen
```

Out[69]:
```
75
```

Task-4

- Pack `Task-2` and `Task-3` in a single while loop and perform them few times in such a way that at the end you have a list with difference of length
- End the while loop when desired number of experiments are finished (at least three, please feel free to do more)
- Calculate the arithmetic mean (common average) on the difference of length that you have. (How to sum all values of a list?)

In [72]:
```
No_of_expts=10

ListDifference =[]

for i in range(0, No_of_expts):
    ListRandom = [random.randint(0, No_of_expts) for x in range(0, No_of_
    ListDivByThree = [number for number in ListRandom if number % 3 == 0]
    difference = len(ListRandom) - len(ListDivByThree)
    ListDifference.append(difference)

ListDifference
```

Out[72]: [6, 6, 6, 8, 7, 7, 8, 8, 7, 7]

In [95]:
```
from statistics import mean

# Calculate the mean of the list - ListDifference
mean(ListDifference)

# Ref - https://www.geeksforgeeks.org/find-average-list-python/
```

Out[95]: 7

b. Data Wrangling with Python: Activity 2 - Analyze Multiline String and Generate the unique Word Count

Task 1 - Create multiline_text and assign p and p file data

In [76]:

```
multiline_text = """It is a truth universally acknowledged, that a single

However little known the feelings or views of such a man may be on his fi

"My dear Mr. Bennet," said his lady to him one day, "have you heard that

Mr. Bennet replied that he had not.

"But it is," returned she; "for Mrs. Long has just been here, and she tol

Mr. Bennet made no answer.

"Do you not want to know who has taken it?" cried his wife impatiently.

"You want to tell me, and I have no objection to hearing it."

This was invitation enough.

"Why, my dear, you must know, Mrs. Long says that Netherfield is taken by

"What is his name?"

"Bingley."

"Is he married or single?"

"Oh! Single, my dear, to be sure! A single man of large fortune; four or

"How so? How can it affect them?"

"My dear Mr. Bennet," replied his wife, "how can you be so tiresome! You

"Is that his design in settling here?"

"Design! Nonsense, how can you talk so! But it is very likely that he may

"I see no occasion for that. You and the girls may go, or you may send th

"My dear, you flatter me. I certainly have had my share of beauty, but I

"In such cases, a woman has not often much beauty to think of."

"But, my dear, you must indeed go and see Mr. Bingley when he comes into

"It is more than I engage for, I assure you."

"But consider your daughters. Only think what an establishment it would b

"You are over-scrupulous, surely. I dare say Mr. Bingley will be very gla

"I desire you will do no such thing. Lizzy is not a bit better than the o

"They have none of them much to recommend them," replied he; "they are al

"Mr. Bennet, how can you abuse your own children in such a way? You take


```

```
"You mistake me, my dear. I have a high respect for your nerves. They are

"Ah, you do not know what I suffer."

"But I hope you will get over it, and live to see many young men of four

"It will be no use to us, if twenty such should come, since you will not

"Depend upon it, my dear, that when there are twenty, I will visit them a

Mr. Bennet was so odd a mixture of quick parts, sarcastic humour, reserve
```

Task 2 - Find the length and type of the multiline_text string

In [79]:
```
type(multiline_text)
```

Out[79]: 4475

In [78]:
```
len(multiline_text)
```

Out[78]: 4475

Task 3 - Remove all new lines by using Replace function

In [80]:
```
multiline_text = multiline_text.replace('\n', "")
```

Task 4 - Find all words in string by using split function

In [82]:
```
import re
#Removing special chars
multiline_text = re.sub(r'[?|$|.|!|"|,|;|:]',r'',multiline_text)

#Split the words based on the Space

list_of_words = multiline_text.split()
list_of_words

```

```
'humour',
'reserve',
'and',
'caprice',
'that',
'the',
'experience',
'of',
'three-and-twenty',
'years',
'had',
'been',
'insufficient',
'to',
'make',
'his',
'wife',
'understand',
'his',
'character',
```

Task 5 - create list only contains unique values (Remove duplicates)

In [90]:
```
unique_words = dict.fromkeys(list_of_words)
len(list(unique_words.keys()))
```

Out[90]: 349

Task 6 - Count number of times the unique word apperared

In [91]:

```
for word in list_of_words:
    if unique_words[word] is None:
        unique_words[word] = 1
    else:
        unique_words[word] += 1
unique_words

#
```

```
'lastMr': 1,
'replied': 3,
'had': 3,
'notBut': 1,
'it': 9,
'returned': 1,
'she': 6,
'for': 12,
'Mrs': 2,
'Long': 2,
'has': 5,
'just': 1,
'been': 2,
'here': 1,
'and': 16,
'told': 1,
'me': 5,
'all': 2,
'about': 1,
```

Task 7 - Top 25

In [94]:
```
top_words = sorted(unique_words.items(), key=lambda key_val_tuple: key_va
top_words[:25]

# Ref - https://www.geeksforgeeks.org/python-n-largest-values-in-diction
```

Out[94]:
```
[('of', 28),
 ('you', 23),
 ('to', 22),
 ('a', 20),
 ('the', 17),
 ('and', 16),
 ('that', 15),
 ('I', 15),
 ('is', 12),
 ('for', 12),
 ('in', 11),
 ('be', 11),
 ('his', 11),
 ('he', 11),
 ('my', 10),
 ('it', 9),
 ('will', 9),
 ('so', 8),
 ('dear', 8),
 ('was', 8),
 ('are', 8),
 ('them', 8),
 ('must', 7),
 ('have', 7),
 ('no', 7)]
```

c. Data Wrangling with Python: Activity 3 page 49

Task 1

Look up the definition of `permutations` and `dropwhile` from itertools.

In [96]:
```
from itertools import permutations, dropwhile
```

In [ ]:
```
permutations?
```

In [98]:
```
dropwhile?
```

Task 2

Write an expression to generate all the possible three digit numbers using 0, 1, and 2

In [99]:
```
permutations(range(3))
```

Out[99]:
```
<itertools.permutations at 0x21d21a76e00>
```

Task 3

Loop over the iterator expression you generated before. Use print to print each element returned by the iterator. Use `assert` and `isinstance` to make sure that the elements are of type tuple

In [100]:
```
for number_tuple in permutations(range(3)):
    print(number_tuple)
    assert isinstance(number_tuple, tuple)
```

```
(0, 1, 2)
(0, 2, 1)
(1, 0, 2)
(1, 2, 0)
(2, 0, 1)
(2, 1, 0)
```

Task 4

Write the loop again. But this time use `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 the dropwhile to a list.

Extra task can be to check the actual type that dropwhile returns without the casting asked above

In [101]:
```
for number_tuple in permutations(range(3)):
    print(list(dropwhile(lambda x: x <= 0, number_tuple)))
```

```
[1, 2]
[2, 1]
[1, 0, 2]
[1, 2, 0]
[2, 0, 1]
[2, 1, 0]
```

d. Data Wrangling with Python: Activity 4 - Activity 04 Design your own CSV parser

In [115]:
```
from itertools import zip_longest
```

In [116]:
```
def 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
```

In [114]:

```
### For this Activity we are not gonna make smaller subtasks like the one
### Write the whole activity's code bellow this line

with open("C:/Users/ragun/Documents/GitHub/DSC540/Weeks 1 & 2/sales_recor
    firstLine = fd.readline()
    header = firstLine.replace("\n", "").split(",")
    for i, line in enumerate(fd):
        # Here we loop over the first 10 lines in order to not to make th
        line = line.replace("\n", "").split(",")
        dict_output = dict_from_csv_line(header, line)
        print(dict_output)
        if i > 10:
            break
```

```
{'Region': 'Central America and the Caribbean', 'Country': 'Antigua and B
arbuda ', 'Item Type': 'Baby Food', 'Sales Channel': 'Online', 'Order Pri
ority': 'M', 'Order Date': '12/20/2013', 'Order ID': '957081544', 'Ship D
ate': '1/11/2014', 'Units Sold': '552', 'Unit Price': '255.28', 'Unit Cos
t': '159.42', 'Total Revenue': '140914.56', 'Total Cost': '87999.84', 'To
tal Profit': '52914.72'}
{'Region': 'Central America and the Caribbean', 'Country': 'Panama', 'Ite
m Type': 'Snacks', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Or
der Date': '7/5/2010', 'Order ID': '301644504', 'Ship Date': '7/26/2010',
'Units Sold': '2167', 'Unit Price': '152.58', 'Unit Cost': '97.44', 'Tota
l Revenue': '330640.86', 'Total Cost': '211152.48', 'Total Profit': '1194
88.38'}
{'Region': 'Europe', 'Country': 'Czech Republic', 'Item Type': 'Beverage
s', 'Sales Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '9/1
2/2011', 'Order ID': '478051030', 'Ship Date': '9/29/2011', 'Units Sold':
'4778', 'Unit Price': '47.45', 'Unit Cost': '31.79', 'Total Revenue': '22
6716.10', 'Total Cost': '151892.62', 'Total Profit': '74823.48'}
{'Region': 'Asia', 'Country': 'North Korea', 'Item Type': 'Cereal', 'Sale
s Channel': 'Offline', 'Order Priority': 'L', 'Order Date': '5/13/2010',
'Order ID': '892599952', 'Ship Date': '6/15/2010', 'Units Sold': '9016',
'Unit Price': '205.70', 'Unit Cost': '117.11', 'Total Revenue': '1854591.
20', 'Total Cost': '1055863.76', 'Total Profit': '798727.44'}
{'Region': 'Asia', 'Country': 'Sri Lanka', 'Item Type': 'Snacks', 'Sales
Channel': 'Offline', 'Order Priority': 'C', 'Order Date': '7/20/2015', 'O
rder ID': '571902596', 'Ship Date': '7/27/2015', 'Units Sold': '7542', 'U
nit Price': '152.58', 'Unit Cost': '97.44', 'Total Revenue': '1150758.3
6', 'Total Cost': '734892.48', 'Total Profit': '415865.88'}
{'Region': 'Middle East and North Africa', 'Country': 'Morocco', 'Item Ty
pe': 'Personal Care', 'Sales Channel': 'Offline', 'Order Priority': 'L',
'Order Date': '11/8/2010', 'Order ID': '412882792', 'Ship Date': '11/22/2
010', 'Units Sold': '48', 'Unit Price': '81.73', 'Unit Cost': '56.67', 'T
otal Revenue': '3923.04', 'Total Cost': '2720.16', 'Total Profit': '1202.
88'}
{'Region': 'Australia and Oceania', 'Country': 'Federated States of Micro
nesia', 'Item Type': 'Clothes', 'Sales Channel': 'Offline', 'Order Priori
ty': 'H', 'Order Date': '3/28/2011', 'Order ID': '932776868', 'Ship Dat
e': '5/10/2011', 'Units Sold': '8258', 'Unit Price': '109.28', 'Unit Cos
t': '35.84', 'Total Revenue': '902434.24', 'Total Cost': '295966.72', 'To
tal Profit': '606467.52'}
{'Region': 'Europe', 'Country': 'Bosnia and Herzegovina', 'Item Type': 'C
lothes', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date':
'10/14/2013', 'Order ID': '919133651', 'Ship Date': '11/4/2013', 'Units S
```

```
old': '927', 'Unit Price': '109.28', 'Unit Cost': '35.84', 'Total Revenu
e': '101302.56', 'Total Cost': '33223.68', 'Total Profit': '68078.88'}
{'Region': 'Middle East and North Africa', 'Country': 'Afghanistan', 'Ite
m Type': 'Clothes', 'Sales Channel': 'Offline', 'Order Priority': 'M', 'O
rder Date': '8/27/2016', 'Order ID': '579814469', 'Ship Date': '10/5/201
6', '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': 'Bab
y Food', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order Date':
'4/13/2015', 'Order ID': '192993152', 'Ship Date': '5/7/2015', 'Units Sol
d': '9817', 'Unit Price': '255.28', 'Unit Cost': '159.42', 'Total Revenu
e': '2506083.76', 'Total Cost': '1565026.14', 'Total Profit': '941057.6
2'}
{'Region': 'Middle East and North Africa', 'Country': 'Turkey', 'Item Typ
e': 'Office Supplies', 'Sales Channel': 'Offline', 'Order Priority': 'C',
'Order Date': '9/25/2013', 'Order ID': '557156026', 'Ship Date': '10/15/2
013', 'Units Sold': '3704', 'Unit Price': '651.21', 'Unit Cost': '524.9
6', 'Total Revenue': '2412081.84', 'Total Cost': '1944451.84', 'Total Pro
fit': '467630.00'}
{'Region': 'Middle East and North Africa', 'Country': 'Oman', 'Item Typ
e': 'Cosmetics', 'Sales Channel': 'Online', 'Order Priority': 'M', 'Order
Date': '5/12/2013', 'Order ID': '741101920', 'Ship Date': '5/17/2013', 'U
nits Sold': '7382', 'Unit Price': '437.20', 'Unit Cost': '263.33', 'Total
Revenue': '3227410.40', 'Total Cost': '1943902.06', 'Total Profit': '1283
508.34'}
```

In [ ]:

```
1
```