```
Before you turn this problem in, make sure everything runs as expected. First, restart the kernel (in the menubar, select
             Kernel \rightarrow Restart) and then run all cells (in the menubar, select Cell \rightarrow Run All).
             Make sure you fill in any place that says YOUR CODE HERE or "YOUR ANSWER HERE", as well as your name and
             collaborators below:
   In [ ]: NAME = "Aasrith Varma"
              COLLABORATORS = "None"
             General Instructions
             1 - Start by downloading this jupyter notebook to your local machine
             2 - Open a tab in your browser and type <a href="https://colab.research.google.com/">https://colab.research.google.com/</a>
             3 - This will open a small window. Choose the last option on the upper menu, "Upload". Then choose the jupyter notebook you
             have saved in step 1
             4 - You can start working on your assignment by answering the questions in the corresponding cells.
             5 - If you have any questions, please reach out to your instructor and TAs
             MAT115: Statistics (MAT115/116) - Assignment 1
             Introduction to Variables Location Based Assignment
             This assignment is a location based-assignment that will require you to interact with the city around in you in a new way. Simply
             put, the objective is to measure a variable in a city in the Guntur district. You will identify a measurable variable in the city and
             then create an estimate using the Fermi estimation technique. Next, you will complete the data collection, calculate descriptive
             statistics on the data, and create relevant data visualizations. You will also have a chance to apply your knowledge of
             probability and simulation to solve a problem. This is an individual assignment. Everything you submit should be your own
             words and reflect your own understanding of the material.
             NOTES:
             Anything marked as optional will only be scored if it is completed correctly. You must upload two files:
               • Primary Resource: A PDF of your entire assignment. Run all cells before converting the notebook to a PDF, and double
                 check to make sure that the PDF is complete with all sections visible. Email attachments will not be accepted. If you're
                 having difficulty converting your notebook to a PDF, try the tips available here
               • Secondary Resource: A zipped folder containing the .ipynb file and your original photo files.
             PART 1: VARIABLE SELECTION [#variables]
             Select a neighborhood in a city in the Guntur district. Visit this neighborhood and spend at least 30 minutes exploring the
             neighborhood to find your variable.
             Important notes:
               • The variable must be something that can be measured at different locations in the city. You need to make at least 10
                 different measurements of this variable, one for each location. The locations must be at least 100 meters away from each

    You must be able to calculate the mean, median, mode, and standard deviation of the variable.

    Be clear about your choice of locations to make the variable measurements.

    Get creative! Try to choose an interesting and informative variable and make sure to justify why the variable you have

                 chosen is interesting.
             1. Define and operationalize your variable here.
             Describe how you selected your variable. Specifically identify the type of variable, and whether you will be measuring a total,
             proportion, or average. Also identify the units it will be measured in and explain in detail how you will measure it. Make sure
             that your explanation is clear enough that another student would understand how to make the same measurement. Give the
             address of the 10 or more locations where you will conduct your measurement and provide an image that clearly identifies
             these locations on a map. (<150 words)
The variable I have selected is the number of parcels taken from ahotel in one day. I have selected this variable in order to know the average
food parcled from the hotel in one day. It helps us to know how many people prefer taking food parcled from hotel. My variable is a Quantitative
discrete variable as the number of parcels taken from the hotels are countable and has a fixed whole number as value. I will be measuring the
average of my data as i want to know how many people prefer taking food home. PROCEDURE:- 1)I went to Guntur District to measure the
values for my variable. Then I visited 10 hotels as referred below. 2) After going to the hotel I have interacted with the manager of the hotel and
told them to keep the data of how many parcels the people are taking from their hotel. 3)I have done the same thing for the other 9 hotels. 4)I
have noted the information about the food parcels from each hotel. The 10 locations I have used to measure the variables are. 1)Drunken
Monkey:-5-87-26, Ground Floor, Chandrika Appartments 6TH Lane, Main Road 2)Hangout:-Dr no: 5-87-40/2 Lakshmipuram main road,
Guntur, Andhra Pradesh 522007 3) The Box:-KLP School Play Ground, road, lane, beside thancos naturals ice-cream, S.V.N Colony,
Navabharath Nagar, Guntur, Andhra Pradesh 522006 4) Burger Stories:-D no: 4-20-23 Ground Floor Beside Thickshake Factory, Ring road,
Siddharth Nagar, Guntur, Andhra Pradesh 522006 5) CINNAMON:-5-87-57/1 Main Road Opposite Kalanikethan, Lakshmipuram, Guntur,
Andhra Pradesh 522007 6) APSARA ice cream:- Happy Foods, Shop No.2, Next to Viva School Siddhartha, Krishna Nagar Main Rd,
Gujjanagundla, Guntur, Andhra Pradesh 522007 7) THICK SHAKE FACTORY:-Ring Rd, Nalanda Nagar, Navabharath Nagar, Guntur, Andhra
Pradesh 522006 8) Dominos: - Ground Floor Annapurna Commercial Complex, Lakshmipuram Main Rd, Guntur, Andhra Pradesh 522007
9)BISTRO:-Brindavan Gardens, Guntur, Andhra Pradesh 522006 10) GISMAT:-Amaravati Road, 8th Line, beside Hindu Pharmacy, Bharathpet,
Guntur, Andhra Pradesh 522007 I measured the data by directly asking the the number of number of parcels taken from the hotels.It
doesn'tany units as it has constant whole number or numerical.
             2. Discuss variable relationships.
               • 2.1 (<150 words)
                   • A. Describe a scenario in which your variable could be an independent variable.
                   B. What could be the dependent variable(s)?
                   • C. What are some possible extraneous or confounding variables in this scenario?
             The variable i chose can be independent variable for the variables like"income of the chef" etc.the variables that could be
             dependent on my variable are the "income of the hotel owner" and "income of the people working in the hotel" it is dependent
             on my variable as the increment in their salary occurs when the food parcels are more.confounding variables are "infrastructure"
             of the hotel" if the infrastructure is good many people will like eat in the hotel. which effect the above mentioned dependent
             and independent variable.
               • 2.2 (< 150 words)
                   • A. Describe a scenario in which your variable could be a dependent variable.
                   • B. What could be the independent variable(s)?
                   • C. What are some possible extraneous or confounding variables in this scenario?
             The number of parcels taken from the hotel are dependent upon some variables like"cost of the food", "taste of the food". it is
             dependent as the taste increases the number of parcles taken increases the confounding variables are "advertisement of
             hotel".
             PART 2: ESTIMATION AND MEASUREMENT [#variables]
             Important note: if there is any reason to believe that you did not authentically complete the location based portion of this
             assignment, this will be refered to the Academic Committee, and you risk receiving zeros in all your grades (as per the course
             policy in the syllabus). Please follow the instructions here carefully and include the original photo files in the zip folder along
              with the ipynb.
               1. Go to a Cafe in the neighborhood of your choice to produce a Fermi estimate of your variable. Use a napkin at a cafe to
                 begin your Fermi estimate. You may not (yet) make any measurements. Your estimate should aim to involve at least 5
                 steps where you compute intermediate values. You will have to describe each step clearly, show your work, state any
                 assumptions you're making, and discuss whether your answer seems plausible (but it's not necessary to do so on the
                 napkin; see step 4 below).
               2. Take some photos to document this experience. You must include:
                   • A photo just of your "back of the napkin" estimate (it can and should be quite rough at this point). You will properly
                     format the calculation later.
                   • A selfie in the cafe in which you constructed your Fermi estimate. Clearly show your face, your Fermi estimate, and
                     some of the interior of the cafe.
                   • A selfie outside of the cafe showing your face and the exterior of the cafe, including the name. Bonus points if you are
                      also holding your completed Fermi estimate in the photo too.
               3. Typeset your full estimation in the Python notebook. Here, be sure to clearly explain all steps, justify all assumptions, and
                 comment on whether the answer seems plausible.
               4. It's time to collect your data! Once again, take some photos to document your experience. Include at least two photos of
                 your variable collection process. At least one photo should include your face and the variable you are counting.
             Follow the instructions in this link to upload your pictures to the jupyter notebook:
    In [1]: from IPython.display import Image, display
              Image(filename="AA1.jpeg", height=400, width=400)
   Out[1]:
   In [2]: from IPython.display import Image, display
              Image (filename="AA2.jpeg", height=400, width=400)
                                      #eat #play #party
    In [3]: from IPython.display import Image, display
              Image(filename="AA3.jpeg", height=400, width=400)
   Out[3]:
                                                Stories from the
    In [4]: from IPython.display import Image, display
              Image(filename="AA4.jpeg", height=400, width=400)
    Out[4]:
                 CINNAMON
    In [5]: from IPython.display import Image, display
              Image(filename="AA5.jpeg", height=400, width=400)
   Out[5]:
                  Domino's
    In [6]: from IPython.display import Image, display
              Image(filename="AA6.jpeg", height=400, width=400)
   Out[6]:
               Apsara Ice Creams
    In [7]: from IPython.display import Image, display
              Image(filename="AA7.jpeg",height=400,width=400)
   Out[7]:
    In [8]: from IPython.display import Image, display
              Image(filename="AA8.jpeg", height=400, width=400)
   Out[8]:
                                           MONKE
    In [9]: from IPython.display import Image, display
              Image(filename="AA9.jpeg", height=400, width=400)
   Out[9]:
    In [2]: from IPython.display import Image, display
              Image(filename="aa12.jpeg", height=400, width=400)
    Out[2]:
                                           LE BISTRO
    In [1]: from IPython.display import Image, display
              Image(filename="aat1.jpeg", height=400, width=400)
    Out[1]:
                 Variable selected: - The Variable I have selected
                 is number of Porcels taken from a
                 hotels in one day.
                 Fermi estimation:
                                     Number of Porcels Token
                Name of hotel
                Hangout
                 The Box
                Burger Stories
                 CINNAMON
                Dominos
                                   25
                APSARA
                 TSF
    In [2]: from IPython.display import Image, display
              Image(filename="aat2.jpeg", height=400, width=400)
    Out[2]:
                  Drunken monkey
                 Bistro
                GISMAT
                termi Estimation: 14+16+20+21
             PART 3: ANALYSIS
               1. Analyze the data in Python [#pythonprogramming]:
                   • 1.1 Use any method to import your collected data into Python. You can simply type the data directly into a Python list
                     or numpy array. Or, you can put the data in a Google sheet, export to a .cvs file, and import into Python. Print your
                     data in the cell below.
   In [19]: import pandas as pd
              df1=pd.read csv("AASRITH DATA.csv")
   Out[19]:
                 NAME OF THE HOTEL NUMBER OF PARCELS
              0
                          HANGOUT
                                                      14
                           THE BOX
                                                      16
                   BURGER STORIES
                                                      20
              3
                          CINNAMON
                                                      21
                           DOMINOS
                                                      25
                            APSARA
                                                      11
                       THICK SHAKE
                                                      25
                  DRUNKEN MONKEY
                                                      22
                             BISTRO
                                                      30
                             GISMAT
              9
                                                      26
               • 1.2 Using Python, calculate the mean, median, mode, range, and standard deviation of your variable. Print these values. If
                 you use a library function, you need to explain how it works with detailed comments. Do not blindly use library functions!
             Note: Round your final answers up to 2 decimals.
    In [6]: def mean(observations):
                  x=len(observations) #Determines the number of the observations. Here x=10.
                  total_packets=sum(observations) #Adding up all the elements of the data that is Total no. of pac
              kets sold in 10 consecutive days.
                  mean=total packets/x #Formula for mean.
                  print("Mean : " + str(mean))
              packets=[14,16,20,21,25,11,25,22,30,26] #My data
             print(mean(parcels)) #Calling the mean function.
             Mean : 21.0
             None
   In [11]: a = [14, 16, 20, 21, 25, 11, 25, 22, 30, 26]
                                                                 # data
              n = len(a)
              a.sort()
              if n % 2 == 0:
                  median1 = a[n//2]
                  median2 = a[n//2 - 1]
                  median = (median1 + median2)/2
              else:
                  median = a[n//2]
              print("median is: " + str(median))
             median is: 21.5
   In [12]: #standard library function, for every number it finds out how many number of times it is repeated. I
              n that it finds the higher times the number repeated and prints that number
              import statistics
             set1 =[14,16,20,21,25,11,25,22,30,26]
             print("Mode of given data set is % s" % (statistics.mode(set1)))
             Mode of given data set is 25
    In [5]: def range(observations):
                 min val=min(observations) #Minimum value in my observations collection.
                 max_val=max(observations) #Maximum value in my observation collection.
                 range=max val-min val #Range formula
                 return range
             data=[14,16,20,21,25,11,25,22,30,26] #My data
              print(range(data)) #Calling Range function
             19
    In [3]: #first it finds out the mean for every observation mean is subtracted from the observation and calcu
              lates the sum and the standard deviation is given by square of that sum divided by under square root
              import statistics
              sample = [14,16,20,21,25,11,25,22,30,26]
              print("standard Deviation of sample is % s "
                                % (statistics.stdev(sample)))
             standard Deviation of sample is 5.9066817155564495
    In []: # Please ignore this cell. This cell is for us to implement the tests
              # to see if your code works properly.
               • 1.3 Create a histogram for your data, properly formatting your figure.
    In [9]: import matplotlib.pyplot as pltwwwwq
              data = [14, 16, 20, 21, 25, 11, 25, 22, 30, 26]
              plt.hist(data,bins=10,color="yellow")
              plt.show
              from scipy.stats import skew
              print("The skewness of the data is", round(skew(data), 2))
```

The skewness of the data is -0.28

#pythonprogramming] (<200 words)</pre>

12.5 15.0 17.5 20.0 22.5 25.0 27.5 30.0

#pythonprogramming, #codereadability]

2. Interpret the descriptive stats: What can you say about the neighborhood based on these values? Is the distribution skewed? Is your visualization in agreement with the descriptive statistics? Explain. [#professionalism, #descriptivestats,

From the above cell I calculated skewness of my lst. My skewness of the distribution is -0.28. My distribution is negatively skewed. I calculated skewness of my distribution from scipy.stats library and imported skew in that library. It is left skewed. When the output is positive then we can say that the distribution is positively skewed whereas the output is negative then we can say that the distribution is negatively skewed. With the help of descriptive statistics and python programming I calculated skewness

1. Can the mean of your data be interpreted as the expected value of a random variable? Explain why or why not in detail.

2. Suppose something unfortunate happened: you stole too many napkins for your Fermi estimate, so you decided to write all of your variable measurements on separate napkins, one napkin for each location. On your way back to the campus, the wind picked up and blew them all away! Luckily, you managed to collect all of the napkins, but now the data is totally randomly reordered, meaning that you have no idea which napkin corresponds to which location. Suppose that you tried to just guess

* What is the probability that you are unlucky, and sadly NONE of the napkins are matched to the correct location (you guessed all of them wrong)? Estimate this probability using a simu

3. [Optional]: What is the expected number of napkins that will be correctly matched to the corresponding location? Estimate

4. [Optional]: Determine the probability distribution as a function of the number of correctly matched napkins and create a

6. [Optional]: Compute the probability or expected value found above or both analytically (without a simulation).

A random permutation of this list can be created with the following code: rand_napkins =

You want to check whether rand_napkins[i] == i, for each value of i from 0 to 9.

this variable in order to know the average food parcled from the hotel in one day.

PART 5: REFLECTION[#probability, #variables]

• To simplify the problem, you can disregard your actual variable data if you wish, and simply make a new list in Python consisting of the numbers 0 through 9: napkins = [0,1,2,3,4,5,6,7,8,9]. Pretend that this is your stack of napkins with the variable measurements in the correct order. Notice that this data satisfies napkins[i] == i, for all values of i from 0 to 9.

np.random.choice(napkins,10,replace=False). You should be able to explain how this function works and why it is relevant

· You'll need to use a loop to create many random lists and repeat the checking procedure, keeping track of the number of

Reflect on your application of the LOs in this assignment. How are the connections in the city mapped to the connections

between the different LOs. Also reflect on how your prediction and estimation from parts 1 and 2 compare to the results. (<200

1. Variable selection: The variable I have selected is the number of parcels taken from ahotel in one day. I have selected

#import the numpy librart

#taking another loop

calculating unlucky probability

#taking the loop to iterate 1000 times

#arranging the numbers in arandom order

if both numbers are same in the lists t

defining the data

randomly which napkin goes with which location. In other words, you randomly assign each napkin to a given location.

PART 4: PROBABILITY CONSIDERATIONS [#probability,

yes, mean of my data is interpreted as the expected value of a random variable. mean of my data is 21.0

lation. Be sure to interpret the result appropriately. See hints below.

3.0

2.5

2.0

1.5

1.0

0.5

0.0

of my data.

(~50 words)

In [5]: import numpy as np

sum=0

a=[14,16,20,21,25,11,25,22,30,26]

b=np.random.choice(a,10,replace = False)

hen count is increased to 1 then loop will break

this probability using a simulation and interpret the result appropriately.

5. [Optional]: Interpret the distribution based on your previous results.

for i **in** range (1000):

for j in range (0,9):

break
unlucky_probability=1-(sum/1000)

print(unlucky_probability)

raise NotImplementedError()

raise NotImplementedError()

YOUR ANSWER HERE

YOUR ANSWER HERE

for the problem.

Think of the index i as the location label.

0.3329999999999999

In []: # YOUR CODE HERE

visualization.

In []: # YOUR CODE HERE

Hints:

words)

if(b[j]==a[j]):