1. **DESCRIPTION**

80% of people who purchase car insurance are men. If the owners of 9 car insurance are randomly selected, then find the probability using **binomial distribution**that exactly**X** out of them are men

* Read a number X from a line of input
* Print the output rounded till 4 decimal point

**Example:**

**Sample Input:**

6

**Sample Output:**

0.1762

1. **DESCRIPTION**

 If the probability of a profit or loss in an investment is equal, find the probability using geometric distribution that an investor’s **k** investment is his first profit

* Take input from the user **k**
* Print the ouput and round up the output till three decimal points

**Example:**

**Sample Input:**

4

**Sample Output:**

0.062

1. **DESCRIPTION**

Conditional Probability

* The probability of an event which is conditioned or dependent on another event is a **Conditional Probability**
* Conditional Probability = P(A|B) = P(A and B)/P(B)
* P(A|B) is the probability of event A occurring, given that event B occurs

You have the Member dataset, which is an input data file **Members.csv**present at the location **/data/training/blackfriday.csv**

This dataset contains information about information related to the people . Here’s a brief description of the columns in the sample dataset

**Dataset Description:**

The dataset contains data of **8** rows and 4 different columns. The columns are:

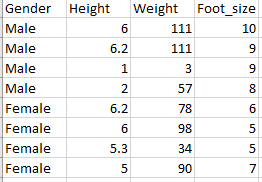
**Gender:**whether the particular person is male of female

**Height:**Height of the person

**Weight:** Weight of the person

**Foot-size:**Foot-size of the person

This is a preview of the data under consideration:



***Question:***

Calculate the probability of members height being more than 5 inches, given that member is female

**Input Format**:

 The file to be read will be **Members.csv**, which contains the data as mentioned above. This file is in **.csv** format.

**Example:**

**Sample Input:**

https://media-doselect.com/Members.csv

**Sample Output:**

0.52

**EXECUTION TIME LIMIT**

1. **DESCRIPTION**

Write a program to perform the following operations:

1. Read a number **X** from a line of input, where X must be a float value
2. Input **X** represents the probability of a person being hit by a falling meteorite
3. Calculate the odds of a person being hit by a falling meteorite
4. Print the output and round up till 3 decimal points

**Example:**

**Sample Input:**

0.07

**Sample Output:**

7.527

1. **DESCRIPTION**

Average number of apples in a carton is 25 with variance of 36. Calculate the probability using normal distribution of number of apples less than**X**.

* Read a number **X** from a line of input
* Print the output and round up the till four decimal points

**Example:**

**Sample Input:**

28

**Sample Output:**

0.6915

1. **DESCRIPTION**

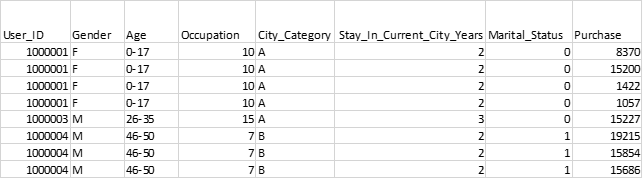
Black Friday falls on the Friday following the ‘Thanksgiving Day’ and is used as an occasion by many stores to offer highly promoted Sales.

You have the Black Friday dataset, which is an input data file **blackfriday.csv**present at the location **/data/training/blackfriday.csv**

This dataset contains information about purchases made in a retail store on Black Friday sale. Here’s a brief description of the columns in the sample dataset:

* **USER\_ID**: ID of the user
* **Gender**: F or M
* **Age**: Age group to which the customer belongs
* **Occupation**: ID of occupation of the customer
* **City\_Category**: A or B or C
* **Stay\_In\_Current\_City\_Years**: 0 to 4+
* **Marital\_Status**: 0: Unmarried, 1: Married
* **Purchase**: Purchase amount in dollars

This is a preview of the data under consideration:



The retailer wants to analyse this data and improve its future sales based on the analysis. In all the questions of this Assignment, we have to perform analysis on this data.

* Purchases made by customers on Black Friday sale are stored in the column named **Purchase**
* **Age**represents the age group the customer belongs to out of 0-17, 18-25, 26-35, 36-45, 46-50, 51-55 and 55+
* **Gender**represents the gender of the customer as F or M
* **City\_Category** represents the category of city the customer belongs to as A, B or C

In this question, we have to perform calculations on the above data as explained below.

**Question:**

Given that the age is 18-25, Calculate the probability of the number of people who have purchased above 10000

**Input Format**:

 The file to be read will be **blackfriday.csv**, which contains the data as mentioned above. This file is in **.csv** format.

**Hint:**

Avoid using repeatative customers

**Example:**

**Sample Input:**

https://media-doselect.s3.amazonaws.com/generic/3M8qkrpOgMEwqevMR5kPon3v/blackfriday.csv

**Sample Output:**

0.3276

1. **DESCRIPTION**

Write a Python code to perform the following operations:

1. Create a list having **10 elements that are positive integer values**

* Read 10 input values on each line of input

2. Convert both the lists into series

3. Find the **population mean**and**population standard deviation** of the series using pandas

* On *first* output line:***Print*** the population mean and population standard deviation values rounded up to **3 decimal places** and separated by a space

4. Draw a**sample** of 5 from the series

* Use **pandas.DataFrame.sample** with the following parameters **n**=sample\_size, **random\_state**=1

5. Find the **sample mean**and**sample standard deviation** of the series using pandas

* On *second* output line:***Print*** the sample mean and sample standard deviation values rounded up to **3 decimal places** and separated by a space

**Example:**

**Sample Input:**

98

63

23

697

136

35

09

343

23

1

**Sample Output:**

142.8 219.589

53.4 60.111

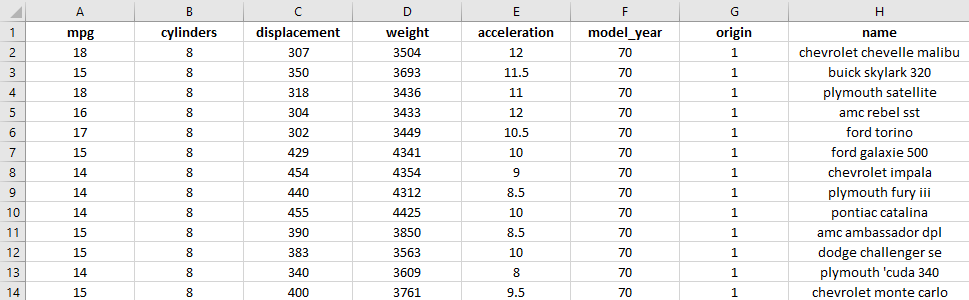
1. **DESCRIPTION**

**Dataset:** mpg.csv

**Dataset Description:**

Data set contains 398 observations containing 8 variables.

Here’s a preview of the data under consideration:



**Problem Statement**

Based on this data set, write a Python code to perform the following operations:

 1. **Load the data set** from the location of the file provided as input using pandas

2. Read a string on the second input line which specifies a **quantitative data column name** in the data set

3. Find the **population mean**and**population standard deviation** of the specified quantitative data column using pandas

4. Draw a**sample** of 200 from the specified quantitative data column

* Use **pandas.DataFrame.sample** with the following parameters **n**=sample\_size, **random\_state**=1

5. Find the **sample mean**and**sample standard deviation** of the specified quantitative data column using pandas

6. Find the **difference** between the sample mean & population mean as well as sample standard deviation & population standard deviation

* On *first* output line: ***Print*** the difference as **<sample mean> -** **<population mean>** rounded up to **3 decimal places**
* On *second* output line: ***Print*** the difference as **<sample std deviation> -** **<population std deviation>** rounded up to **3 decimal places**

**Example:**

**Sample Input:**

https://media-doselect.com/mpg.csv

﻿weight

**Sample Output:**

22.07

34.815

1. **DESCRIPTION**

A food delivery company gets cancellations on ***x*** orders in a day out of 900 total orders. Each customer can make only one cancellation in a day. The company assumes that all customers are independent of each other.

Write a Python code to perform the following operations:

1. Readan integer input which specifies the **number of cancelled orders**

2. Find out the **margin of error**using**scipy.stats.norm.ppf**

* On *first* output line: ***Print***the margin of error value rounded up to **5 decimal places**

3. Determine an approximate **95% confidence interval** for the proportion of orders cancelled in a day

* On *second* output line:***Print*** the confidence interval values rounded up to **5 decimal places** and separated by a space

**Note:**

* Margin of Error = Critical Value\*Standard Error of Statistic
* Confidence Interval = Sample Statistic **±** Margin of Error

**Example:**Let's say 300 out of 900 orders were cancelled

**Sample Input:**

300

The margin of error & confidence interval values should be printed as -

**Sample Output:**

0.02585

0.30749 0.35918

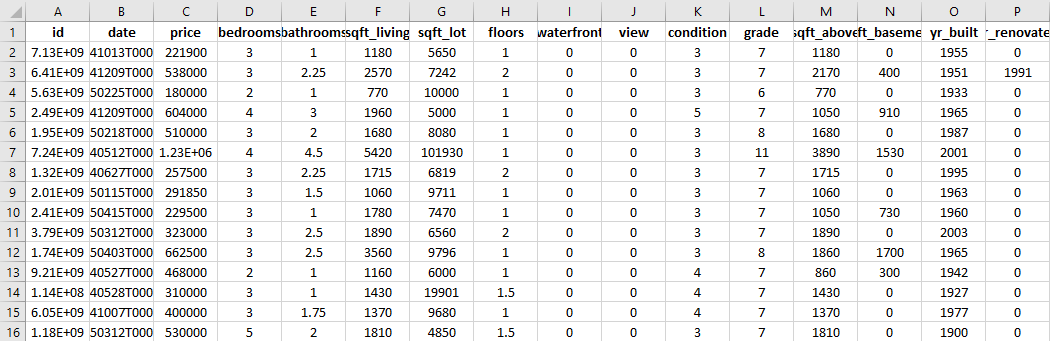
1. **DESCRIPTION**

**Dataset:** Property.csv

**Dataset Description:**

Data set contains 21613 observations containing 21 variables.

Here’s a preview of the data under consideration:



**Problem Statement**

Based on this data set, write a Python code to perform the following operations:

 1. **Load the data set** from the location of the file provided as input using pandas

2. Read a string input which specifies a **quantitative data column name** in the data set

3. Find the **population mean**and**population standard deviation** of the specified quantitative data column using pandas

* On *first* output line: ***Print*** the(1)population mean and (2)population standard deviation values rounded up to **3 decimal places**andseparated by a space

4. Draw a**sample** of 100 from the specified quantitative data column

* Use **pandas.DataFrame.sample** with the following parameters **n**=sample\_size, **random\_state**=4

5. Find the **sample mean**and**sample standard deviation** of the specified quantitative data column using pandas

* On *second* output line: ***Print*** the(1)sample mean and (2)sample standard deviation values rounded up to **3 decimal places**andseparated by a space

6. Check if the sample mean differs from the population mean using **Hypothesis Testing**

a) The hypothesis is stated as follows:

* **Null hypothesis** = sample mean does not differ from the population mean
* **Alternate hypothesis** = sample mean differs from the population mean

b) Perform a test at 95% confidence level and find out the **z-statistic** and **critical value**

* On *third* output line: ***Print***the (1)z-statistic and (2)critical value rounded up to **3 decimal places**andseparated by a space

c) Conclude the relationship between the sample mean and the population mean

* On *fourth* output line: ***Print***the hypothesis that holds true as per Point 1 in the **Note** given below

**Note:**

Point 1:

Z-statistics is

* Lesser than critical value: **fail to reject the null hypothesis**
* Greater than critical value: **reject the null hypothesis**

Point 2:

Make sure your code prints the hypothesis exactly as given above (i.e., lowercase letters and space between words)

**Example:**

**Sample Input:**

https://media-doselect.s3.amazonaws.com/generic/RkzkY87b8Y1QNRwG3QKwe94v/Property.csv

﻿price

**Sample Output:**

540088.142 367127.196

515254.41 280175.923

**-0**.676 1.645

**fail** **to** **reject** **the** **null** **hypothesis**

1. **DESCRIPTION**

Write a Python code to perform the following operations:

1. Readthe following list defined below:

* 763, 667, 593, 402, 348, 278, 123

2. Create another list having **7 elements that are positive integer values**

* Read 7 input values on each line of input

3. Check if there exists a relationship between means of the two lists using **Hypothesis Testing**

a) The hypothesis is stated as follows:

* **Null hypothesis** = there is no relationship (independent)
* **Alternate hypothesis** = there is a relationship

b) Perform a t-test using **stats.ttest\_ind** and find out the **p-value**

* On *first* output line: ***Print***the p-value rounded up to **5 decimal places**

c) Conclude the relationship between means of the two lists

* On *second* output line: ***Print***the hypothesis that holds true as per Point 1 in the **Note** given below

**Note:**

Point 1:

P-value is

* Lesser than significance level (0.05): **there is a relationship**
* Greater than significance level (0.05): **there is no relationship (independent)**

Point 2:

Make sure your code prints the hypothesis exactly as given above (i.e., lowercase letters and space between words)

**Example:**

**Sample Input:**

23

56

86

99

116

294

366

**Sample Output:**

0.00976

there **is** a relationship