

Data Science II with python (Class notes)

STAT 303-2

Arvind Krishna

9/20/22

Table of contents

Preface	6
I Python	7
1 Simple Linear Regression	8
2 Variables, expressions and statements	14
2.0.1 Practice exercise 1	14
2.1 Constants and Variables	15
2.1.1 Variable names	16
2.1.2 Practice exercise 2	17
2.2 Assignment statements	17
2.3 Expressions	17
2.3.1 Practice exercise 3	18
2.4 Converting datatypes	19
2.5 User input	20
2.5.1 Practice exercise 4	20
2.6 Commenting code	21
2.6.1 Practice exercise 5	21
2.7 Programming errors	21
2.7.1 Syntax errors	22
2.7.2 Run-time errors	22
2.7.3 Semantic errors	22
2.7.4 Practice exercise 6	23
2.8 Practice exercise 7	23
3 Control flow statements	25
3.0.1 Comparison operators	25
3.0.2 Logical operators	26
3.0.3 if-elif-else statement	26
3.0.4 Practice exercise 1	27
3.0.5 Try-except	28
3.0.6 Practice exercise 2	28
3.1 Loops	30
3.1.1 for loop	30

3.1.2	while loop	32
3.1.3	Practice exercise 3	33
3.2	break statement	35
3.2.1	Practice exercise 4	36
3.3	continue statement	36
3.3.1	Practice exercise 5:	37
3.4	Loops with strings	38
3.4.1	Practice exercise 6	39
4	Data structures	40
4.0.1	Practice exercise 1	41
4.0.2	Concatenating tuples	41
4.0.3	Unpacking tuples	42
4.0.4	Practice exercise 2	43
4.0.5	Tuple methods	44
4.1	List	45
4.1.1	Adding and removing elements in a list	45
4.1.2	List comprehensions	47
4.1.3	Practice exercise 3	48
4.1.4	Concatenating lists	49
4.1.5	Sorting a list	49
4.1.6	Slicing a list	49
4.1.7	Practice exercise 4	51
4.2	Dictionary	54
4.2.1	Adding and removing elements in a dictionary	54
4.2.2	Iterating over elements of a dictionary	55
4.2.3	Practice exercise 5	56
4.3	Functions	57
4.3.1	Global and local variables with respect to a function	58
4.3.2	Practice exercise 6	58
4.4	Practice exercise 7	60
5	Object-Oriented Programming	63
5.1	Class	64
5.1.1	Creating your own class	64
5.1.2	Example: A class that analyzes a string	66
5.1.3	Practice exercise 1	67
5.2	Inheritance	69
5.2.1	Practice exercise 2	70
5.2.2	Practice exercise 3	71

Appendices	71
A Assignment 1 (Control flow)	72
B 1	73
B.1 1(a)	73
B.2 1(b)	73
C 2	74
C.1 2(a)	74
C.2 2(b)	74
C.3 2(c)	74
C.4 2(d)	74
C.5 3	75
D Assignment 2 (Functions)	76
D.1 1(a)	76
D.2 1(b)	77
D.3 2(a)	77
D.4 2(b)	77
D.5 2(c)	77
D.6 3(a)	78
D.7 3(b)	78
D.8 3(c)	79
D.9 3(d)	79
E Assignment 3 (Data structures)	80
F 1 GDP of The USA	81
F.1 1(a)	81
F.1.1 1(a)(i)	81
F.1.2 1(a)(ii)	81
F.1.3 1(a)(iii)	81
F.2 1(b)	82
F.2.1 1(b)(i)	82
F.2.2 1(b)(ii)	82
F.2.3 1(b)(iii)	82
G 2 Ted Talks	83
G.1 2(a)	83
G.2 2(b)	83
G.3 2(c)	83
G.4 2(d)	83
G.5 2(e)	84

G.6 2(f)	84
G.7 2(g)	84
H 3 Poker	85
I Assignment 4 (Classes and Objects)	87
I.1 Password Manager	87
I.2 SMS store manager	89
I.3 Personalized SMS store manager	91
I.4 Bonus question: Creating a new datatype - list of dictionaries	93
J Datasets	95

Preface

These are class notes for the course STAT303-2. This is not the course text-book. You are required to read the relevant sections of the book as mentioned on the course website.

Part I

Python

1 Simple Linear Regression

```
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
import seaborn as sns
import matplotlib.pyplot as plt
```

Develop a simple linear regression model that predicts car price based on engine size. Datasets to be used: *Car_features_train.csv*, *Car_prices_train.csv*

```
trainf = pd.read_csv('Car_features_train.csv')
trainp = pd.read_csv('Car_prices_train.csv')
train = pd.merge(trainf, trainp)
train.head()
```

	carID	brand	model	year	transmission	mileage	fuelType	tax	mpg	engineSize	price
0	18473	bmw	6 Series	2020	Semi-Auto	11	Diesel	145	53.3282	3.0	37980
1	15064	bmw	6 Series	2019	Semi-Auto	10813	Diesel	145	53.0430	3.0	33980
2	18268	bmw	6 Series	2020	Semi-Auto	6	Diesel	145	53.4379	3.0	36850
3	18480	bmw	6 Series	2017	Semi-Auto	18895	Diesel	145	51.5140	3.0	25998
4	18492	bmw	6 Series	2015	Automatic	62953	Diesel	160	51.4903	3.0	18990

```
#Using the ols function to create an ols object. 'ols' stands for 'Ordinary least squares'
ols_object = smf.ols(formula = 'price~engineSize', data = train)
```

```
#Using the fit() function of the 'ols' class to fit the model
model = ols_object.fit()
```

```
model.summary()
```


Table 1.2: OLS Regression Results

Dep. Variable:	price	R-squared:	0.390
Model:	OLS	Adj. R-squared:	0.390
Method:	Least Squares	F-statistic:	3177.
Date:	Tue, 04 Jan 2022	Prob (F-statistic):	0.00
Time:	17:50:56	Log-Likelihood:	-53949.
No. Observations:	4960	AIC:	1.079e+05
Df Residuals:	4958	BIC:	1.079e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-4122.0357	522.260	-7.893	0.000	-5145.896	-3098.176
engineSize	1.299e+04	230.450	56.361	0.000	1.25e+04	1.34e+04

Omnibus:	1271.986	Durbin-Watson:	0.517
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6490.719
Skew:	1.137	Prob(JB):	0.00
Kurtosis:	8.122	Cond. No.	7.64

```
#Printing model summary which contains among other things, the model coefficients
model.summary()
```

Table 1.5: OLS Regression Results

Dep. Variable:	price	R-squared:	0.390
Model:	OLS	Adj. R-squared:	0.390
Method:	Least Squares	F-statistic:	3177.
Date:	Tue, 04 Jan 2022	Prob (F-statistic):	0.00
Time:	14:34:43	Log-Likelihood:	-53949.
No. Observations:	4960	AIC:	1.079e+05
Df Residuals:	4958	BIC:	1.079e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
--	------	---------	---	------	--------	--------

Intercept	-4122.0357	522.260	-7.893	0.000	-5145.896	-3098.176
engineSize	1.299e+04	230.450	56.361	0.000	1.25e+04	1.34e+04

Omnibus:	1271.986	Durbin-Watson:	0.517
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6490.719
Skew:	1.137	Prob(JB):	0.00
Kurtosis:	8.122	Cond. No.	7.64

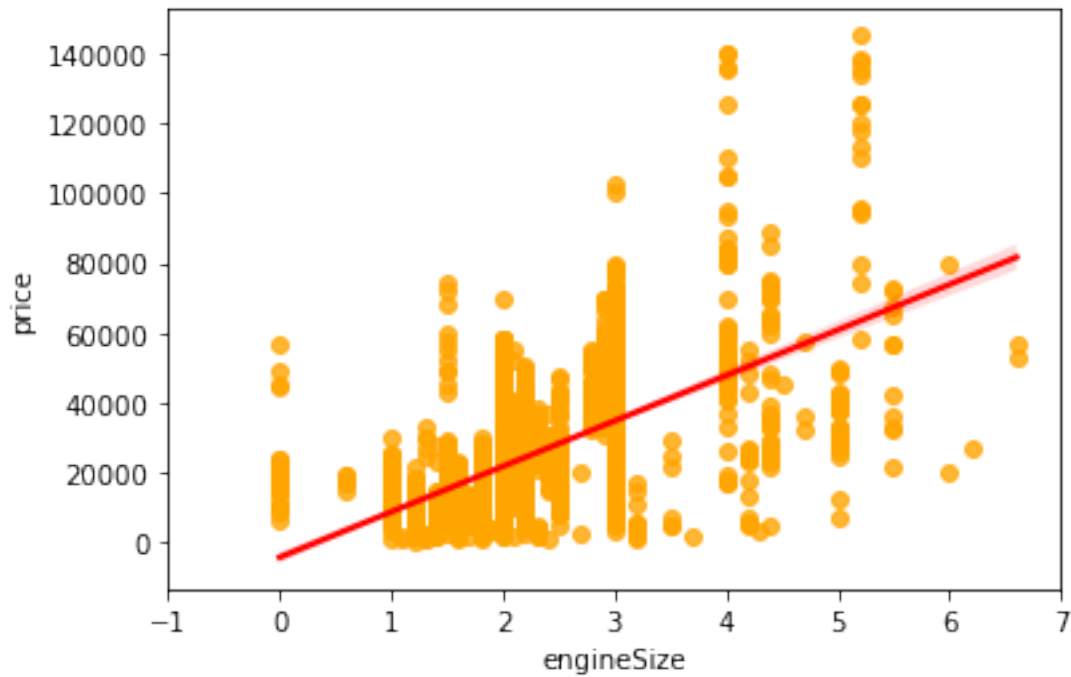
The model equation is: car price = -4122.0357 + 12990 * engineSize

Visualize the regression line

```
sns.regplot(x = 'engineSize', y = 'price', data = train, color = 'orange', line_kws={"color": "red", "dash": [5, 5]})
plt.xlim(-1,7)
```

#Note that some of the engineSize values are 0. They are incorrect, and should ideally be

(-1.0, 7.0)



Predict the car price for the cars in the test dataset. Datasets to be used:
Car_features_test.csv, Car_prices_test.csv

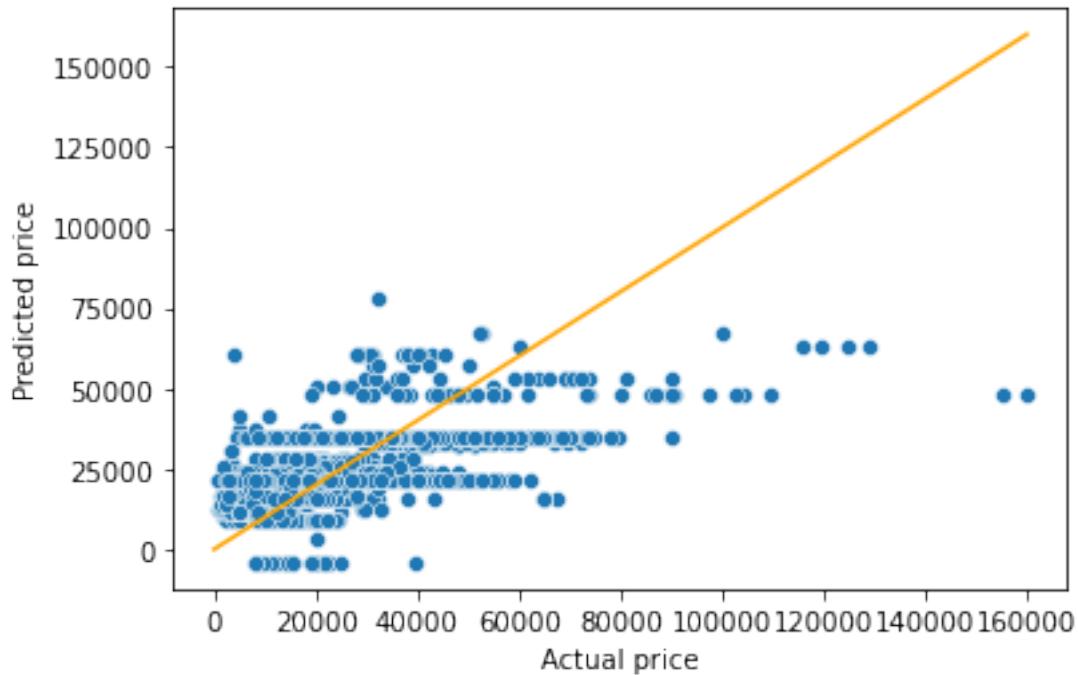
```
testf = pd.read_csv('Car_features_test.csv')
testp = pd.read_csv('Car_prices_test.csv')

#Using the predict() function associated with the 'model' object to make predictions of car
pred_price = model.predict(testf)#Note that the predict() function finds the predictor 'engine'
```

Make a visualization that compares the predicted car prices with the actual car prices

```
sns.scatterplot(x = testp.price, y = pred_price)
#In case of a perfect prediction, all the points must lie on the line x = y.
sns.lineplot(x = [0,testp.price.max()], y = [0,testp.price.max()],color='orange') #Plotting the line x = y
plt.xlabel('Actual price')
plt.ylabel('Predicted price')
```

Text(0, 0.5, 'Predicted price')



The prediction doesn't look too good. This is because we are just using one predictor - engine size. We can probably improve the model by adding more predictors when we learn multiple linear regression.

What is the RMSE of the predicted car price?

```
np.sqrt(((testp.price - pred_price)**2).mean())
```

12995.1064515487

The root mean squared error in predicting car price is around \$13k.

What is the residual standard error based on the training data?

```
np.sqrt(model.mse_resid)
```

12810.109175214136

The residual standard error on the training data is close to the RMSE on the test data. This shows that the performance of the model on unknown data is comparable to its performance on known data. This implies that the model is not overfitting, which is good! In case we overfit a model on the training data, its performance on unknown data is likely to be worse than that on the training data.

Find the confidence and prediction intervals of the predicted car price

```
#Using the get_prediction() function associated with the 'model' object to get the intervals
intervals = model.get_prediction(testf)
```

```
#The function requires specifying alpha (probability of Type 1 error) instead of the confidence
intervals.summary_frame(alpha=0.05)
```

	mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper
0	34842.807319	271.666459	34310.220826	35375.393812	9723.677232	59961.937406
1	34842.807319	271.666459	34310.220826	35375.393812	9723.677232	59961.937406
2	34842.807319	271.666459	34310.220826	35375.393812	9723.677232	59961.937406
3	8866.245277	316.580850	8245.606701	9486.883853	-16254.905974	33987.396528
4	47831.088340	468.949360	46911.740050	48750.436631	22700.782946	72961.393735
...
2667	47831.088340	468.949360	46911.740050	48750.436631	22700.782946	72961.393735

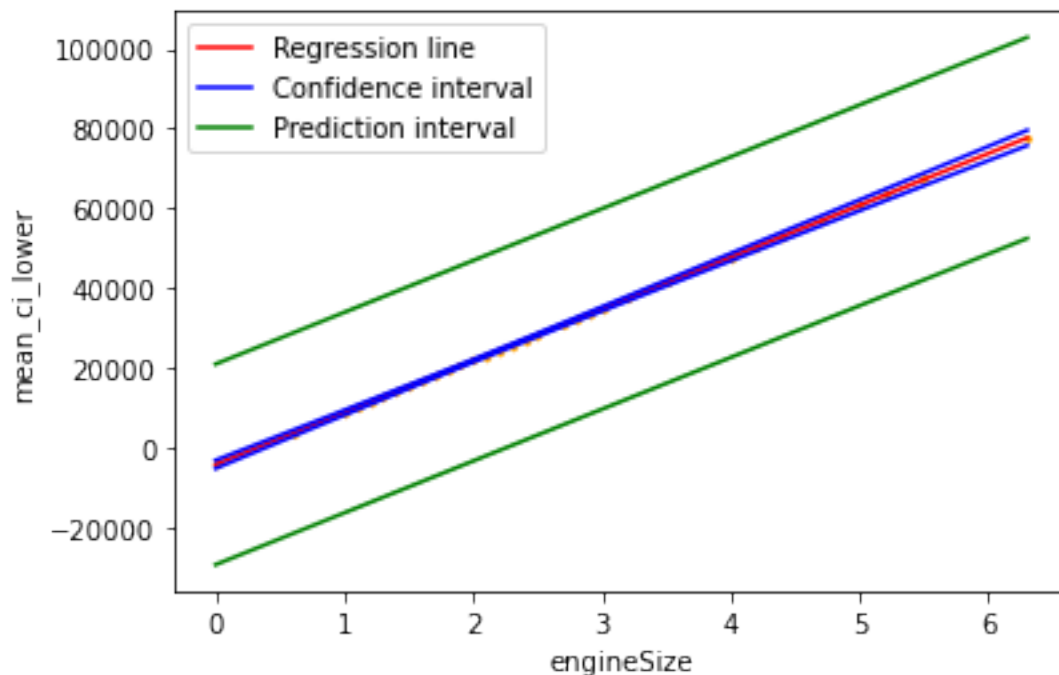
	mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper
2668	34842.807319	271.666459	34310.220826	35375.393812	9723.677232	59961.937406
2669	8866.245277	316.580850	8245.606701	9486.883853	-16254.905974	33987.396528
2670	21854.526298	184.135754	21493.538727	22215.513869	-3261.551421	46970.604017
2671	21854.526298	184.135754	21493.538727	22215.513869	-3261.551421	46970.604017

Show the regression line predicting car price based on engine size for test data.
Also show the confidence and prediction intervals for the car price.

```
interval_table = intervals.summary_frame(alpha=0.05)
```

```
sns.scatterplot(x = testf.engineSize, y = pred_price,color = 'orange', s = 10)
sns.lineplot(x = testf.engineSize, y = pred_price, color = 'red')
sns.lineplot(x = testf.engineSize, y = interval_table.mean_ci_lower, color = 'blue')
sns.lineplot(x = testf.engineSize, y = interval_table.mean_ci_upper, color = 'blue',label=
sns.lineplot(x = testf.engineSize, y = interval_table.obs_ci_lower, color = 'green')
sns.lineplot(x = testf.engineSize, y = interval_table.obs_ci_upper, color = 'green')
plt.legend(labels=["Regression line","Confidence interval", "Prediction interval"])
```

<matplotlib.legend.Legend at 0x1dcf5ccc280>



2 Variables, expressions and statements

Some of the commonly used objects in python are numbers - integer and float, strings and bool (true/false). The data type of the object can be identified using the in-built python function `type()`. For example, see the following objects and their types:

```
type(4)
```

int

```
type(4.4)
```

float

```
type('4')
```

str

```
type(True)
```

bool

2.0.1 Practice exercise 1

What is the datatype of the following objects?

1. 'This is False'
2. "This is a number"
3. 1000
4. 65.65
5. False

Solution: 1. String 2. String 3. Integer 4. Float 5. Bool

2.1 Constants and Variables

A constant is a value that cannot be changed. It may be a number, string or any other datatype. Below are some examples of printing a constant:

```
print(4)
```

4

```
print("This is a string and also a constant")
```

This is a string and also a constant

```
print(False)
```

False

A variable is an object whose value can be changed. For example, consider the object below:

```
x = 2
```

In the above code the variable `x` has been assigned a value of 2. However, the value of `x` can be changed:

```
x = 3  
print("x =", x)
```

x = 3

Thus, the object `x` in the above code is a variable that refers to a memory location storing the constant value of 3.

2.1.1 Variable names

There are a some rules for naming variables:

1. A variable name must start with a letter or underscore `_`
2. A variable name may consist of letters, numbers, and underscores only

For example, some of the valid variable names are `*salary`, `text10`, `__varname`. *Some of the invalid variable names are `salary% 10text`, `varname)`**

3. Variable names are case-sensitive. For example, the variable `Varname` will be different from `varname`.
4. There are certain *reserved words* in python that have some meaning, and cannot be used as variable names. These reserved words are:

```
<IPython.core.display.Image object>
```

Best coding practice: Variables should be named such that they are informative of the value they are storing. For example, suppose we wish to compute the income tax a person has to pay based on their income and tax rate. Below are two ways of naming variables to do this computation:

```
income = 80000
tax_rate = 0.15
print("Income tax = ",income*tax_rate)
```

```
Income tax = 12000.0
```

```
a = 80000
b = 0.15
print("Income tax = ",a*b)
```

```
Income tax = 12000.0
```

The former code chunk is better than the latter one as it makes the code easy to read and understand.

2.1.2 Practice exercise 2

2.1.2.1

In the statements below, classify the objects as variables or constants: 1. `value = "name"` 2. `constant = 7` 3. `another_const = "variable"` 4. `True_False = True`

The objects on the left hand side of the '=' sign are variables, and those on the right hand side are constants.

Which of the following variable names are valid: 1. `var.name` 2. `var9name` 3. `__varname` 4. `varname*`

Variable names 2 and 3 are valid because they contain only letters, number and and/or underscore.

2.2 Assignment statements

Values are assigned to variables with the assignment statement (`=`). An assignment statement may have a constant or an expression on the right hand side of the (`=`) sign, and a variable name on the left hand side.

For example, the code lines below are assignment statements

```
var = 2
var = var + 3
```

2.3 Expressions

The mathematical operations and their corresponding operators are as follows:

1. Exponent: `**`
2. Remainder: `%`
3. Multiplication: `*`
4. Division: `/`
5. Addition: `+`
6. Subtraction: `-`

The operators above are in decreasing order of precedence, i.e., an exponent will be evaluated before a remainder, a remainder will be evaluated before a multiplication, and so on.

For example, check the precedence of operators in the computation of the following expression:

```
2+3%4*2
```

8

In case an expression becomes too complicated, use of parenthesis may help clarify the precedence of operators. Parenthesis takes precedence over all the operators listed above. For example, in the expression below, the terms within parenthesis are evaluated first:

```
2+3%(4*2)
```

5

2.3.1 Practice exercise 3

Which of the following statements is an assignment statement:

1. `x = 5`
2. `print(x)`
3. `type(x)`
4. `x + 4`

Solution:

Only (1) as the constant 5 is being assigned to the variable `x`

What will be the result of the following expression:

```
1%2**3*2+1
```

3

2.4 Converting datatypes

Sometimes a value may have a datatype that is not suitable for using it. For example, consider the variable called *annual_income* in the code below:

```
annual_income = "80000"
```

Suppose we wish to divide *annual_income* by 12 to get the monthly income. We cannot use the variable *monthly_income* directly as its datatype is a string and not a number. Thus, numerical operations cannot be performed on the variable *annual_income*.

We'll need to convert *annual_income* to an integer. For that we will use the python's in-built `int()` function:

```
annual_income = int(annual_income)
monthly_income = annual_income/12
print("monthly income = ", monthly_income)
```

```
monthly income = 6666.666666666667
```

Similarly, datatypes can be converted from one type to another using in-built python functions as shown below:

```
#Converting integer to string
str(9)
```

```
'9'
```

```
#Converting string to float
float("4.5")
```

```
4.5
```

```
#Converting bool to integer
int(True)
```

```
1
```

Sometimes, conversion of a value may not be possible. For example, it is not possible to convert the variable *greeting* defined below to a number:

```
greeting = "hello"
```

However, in some cases, mathematical operators such as `+` and `*` can be applied on strings. The operator `+` concatenates multiple strings, while the operator `*` can be used to concatenate a string to itself multiple times:

```
"Hi" + " there!"
```

```
'Hi there!'
```

```
"5" + '3'
```

```
'53'
```

```
"5"*8
```

```
'55555555'
```

2.5 User input

Python's in-built `input()` function can be used to accept an input from the user. For example, suppose we wish the user to input their age:

```
age = input("Enter your age:")
```

```
Enter your age:34
```

The entered value is stored in the variable `age` and can be used for computation.

2.5.1 Practice exercise 4

Ask the user to input their year of birth, and print their age.

Solution:

```
age = int(input("Input year of birth:"))  
print("Age=", 2022-age)
```

Input year of birth:1990

Age= 32

2.6 Commenting code

The # symbol can be used to comment the code. Anything after the # sign is ignored by python. Commenting a code may have several purposes, such as:

- Describe what is going to happen in a sequence of code
- Document who wrote the code or other ancillary information
- Turn off a line of code - perhaps temporarily

For example, below is code with a comment to describe the purpose of the code:

```
#Computing number of hours of lecture in this course  
print("Total lecture hours of STAT201=", 10*3*(5/6))
```

Total lecture hours of STAT201= 25.0

2.6.1 Practice exercise 5

Which of the following lines is a comment:

1. #this is a comment
2. ##this may be a comment
3. A comment#

Solution: (1) and (2) are comments as they begin with #.

2.7 Programming errors

There are 3 types of errors that can occur in a program - syntax errors, run-time errors, and semantic errors.

2.7.1 Syntax errors

Syntax errors occur if the code is written in a way that it does not comply with the rules / standards / laws of the language (python in this case). For example, suppose a values is assigned to a variable as follows:

```
9value = 2
```

The above code when executed will indicate a syntax error as it violates the rule that a variable name must not start with a number.

2.7.2 Run-time errors

Run-time errors occur when a code is syntactically correct, but there are other issues with the code such as:

- Misspelled or incorrectly capitalized variable and function names
- Attempts to perform operations (such as math operations) on data of the wrong type (ex. attempting to subtract two variables that hold string values)
- Dividing by zero
- Attempts to use a type conversion function such as int on a value that can't be converted to an int

For example, suppose a number is multiplied as follows:

```
multiplication_result = x * 4
```

The above code is syntactically correct. However, it will generate an error as the variable `x` has not been defined as a number.

2.7.3 Semantic errors

Semantic errors occur when the code executes without an error being indicated by the compiler. However, it does not work as intended by the user. For example, consider the following code of multiplying the number 6 by 3:

```
x = '6'  
x * 3
```

```
'666'
```

If it was intended to multiply the number 6, then the variable `x` should have been defined as `x=6` so that `x` has a value of type `integer`. However, in the above code `6` is a `string` type value. When a `string` is multiplied by an integer, say n , it concatenates with itself n times.

2.7.4 Practice exercise 6

Suppose we wish to compute tax using the income and the tax rate. Identify the type of error from amongst syntax error, semantic error and run-time error in the following pieces of code.

```
income = 2000
tax = .08 * Income
print("tax on", income, "is:", tax)
```

```
income = 2000
tax = .08 x income
print("tax on", income, "is:", tax)
```

```
income = 2000
tax = .08 ** income
print("tax on", income, "is:", tax)
```

The errors from top to bottom are:

1. Run-time error (The variable `Income` is not defined)
2. Syntax error (Here `x` is not interpretable for the python compiler. Thus, the code does not comply with the rules of the language)
3. Semantic error (Instead of multiplying the income by tax rate, the tax rate is raised to the power of `income`. The code works but not as intended).

2.8 Practice exercise 7

The formula for computing final amount if one is earning compound interest is given by:

$$A = P \left(1 + \frac{r}{n} \right)^{nt},$$

where:

P = Principal amount (initial investment),

r = annual nominal interest rate,

n = number of times the interest is computed per year,
 t = number of years

Write a Python program that assigns the principal amount of \$10000 to variable P , assign to n the value 12, and assign to r the interest rate of 8%. Then have the program prompt the user for the number of years t that the money will be compounded for. Calculate and print the final amount after t years.

What is the amount if the user enters t as 4 years?

Solution:

```
t = input("Number of years:")
r = 0.08; n = 12; P = 10000;
print("The amount is: $" + str(P*(1+r/n)**(n*int(t))))
```

Number of years:4

The amount is: \$13756.6610043379

3 Control flow statements

A [control flow statement](#) in a computer program determines the individual lines of code to be executed and/or the order in which they will be executed. In this chapter, we'll learn about 3 types of control flow statements:

1. if-elif-else
2. for loop
3. while loop

The first type of control flow statement is **if-elif-else**. This statement helps with conditional execution of code, i.e., the piece of code to be executed is selected based on certain condition(s).

3.0.1 Comparison operators

For testing if conditions are true or false, first we need to learn the operators that can be used for comparison. For example, suppose we want to check if two objects are equal, we use the `==` operator:

```
5==6
```

False

```
x = "hi"  
y = "hi"  
x==y
```

True

Below are the python comparison operators and their meanings.

Python code	Meaning
<code>x==y</code>	Produce True if ... x is equal to y
<code>x != y</code>	... x is not equal to y
<code>x > y</code>	... x is greater than y
<code>x < y</code>	... x is less than y
<code>x >= y</code>	... x is greater than or equal to y
<code>x <= y</code>	... x is less than or equal to y

3.0.2 Logical operators

Sometimes we may need to check multiple conditions simultaneously. The logical operator **and** is used to check if all the conditions are true, while the logical operator **or** is used to check if either of the conditions is true.

```
#Checking if both the conditions are true using 'and'
5==5 and 67==68
```

False

```
#Checking if either condition is true using 'or'
x = 6; y = 90
x<0 or y>50
```

True

3.0.3 if-elif-else statement

The **if-elif-else** statements can check several conditions, and execute the code corresponding to the condition that is true. Note that there can be as many **elif** statements as required.

Syntax: Python uses indentation to identify the code to be executed if a condition is true. All the code indented within a condition is executed if the condition is true.

Example: Input an integer. Print whether it is positive or negative.

```
number = input("Enter a number:") #Input an integer
number_integer = int(number)      #Convert the integer to 'int' datatype
if number_integer>0:              #Check if the integer is positive
    print("Number is positive")
```

```
else:
    print("Number is negative")
```

```
Enter a number:-9
Number is negative
```

In the above code, note that anything entered by the user is taken as a string datatype by python. However, a string cannot be positive or negative. So, we converted the number input by the user to integer to check if it was positive or negative.

There may be multiple statements to be executed if a condition is true. See the example below.

Example: Input a number. Print whether it is positive, negative or zero. If it is negative, print its absolute value.

```
number = input("Enter a number:")
number_integer = int(number)
if number_integer>0:
    print("Number is positive")
elif number_integer==0:
    print("Number is zero")
else:
    print("Number is negative")
    print("Absolute value of number = ", abs(number_integer))
```

```
Enter a number:0
Number is zero
```

3.0.4 Practice exercise 1

Input a number. Print whether its odd or even.

Solution:

```
num = int(input("Enter a number: "))
if num%2==0:          #Checking if the number is divisible by 2
    print("Number is even")
else:
    print("Number is odd")
```

```
Enter a number: 5
Number is odd
```

3.0.5 Try-except

If we suspect that some lines of code may produce an error, we can put them in a **try** block, and if an error does occur, we can use the **except** block to instead execute an alternative piece of code. This way the program will not stop if an error occurs within the **try** block, and instead will be directed to execute the code within the **except** block.

Example: Input an integer from the user. If the user inputs a valid integer, print whether it is a multiple of 3. However, if the user does not input a valid integer, print a message saying that the input is invalid.

```
num = input("Enter an integer:")

#The code lines within the 'try' block will execute as long as they run without error
try:
    #Converting the input to integer, as user input is a string
    num_int = int(num)

    #checking if the integer is a multiple of 3
    if num_int%3==0:
        print("Number is a multiple of 3")
    else:
        print("Number is not a multiple of 3")

#The code lines within the 'except' block will execute only if the code lines within the '
except:
    print("Input must be an integer")
```

```
Enter an integer:hi
Input must be an integer
```

3.0.6 Practice exercise 2

3.0.6.1

Ask the user to enter their exam score. Print the grade based on their score as follows:

Score	Grade
(90,100]	A
(80,90]	B
[0,80]	C

If the user inputs a score which is not a number in [0,100], print invalid entry.

Solution:

```
score = input("Enter exam score:")
try:

    #As exam score can be a floating point number (such as 90.65), we need to use 'float'
    score_num = float(score)
    if score_num>90 and score_num<=100:
        print("Grade: A")
    elif score_num>80 and score_num<=90:
        print("Grade: B")
    elif score_num>=0 and score_num<=80:
        print("Grade: C")
    else:
        print("Invalid score")           #If a number is less than 0 or more than 100
except:
    print("Invalid input")              #If the input is not a number
```

Enter exam score:90

Grade: B

3.0.6.2

Nested if-elif-else statements: This question will lead you to create nested if statements, i.e., an if statement within another if statement.

Think of a number in [1,5]. Ask the user to guess the number.

- If the user guesses the number correctly, print “Correct in the first attempt!”, and stop the program. Otherwise, print “Incorrect! Try again” and give them another chance to guess the number.
- If the user guesses the number correctly in the second attempt, print “Correct in the second attempt”, otherwise print “Incorrect in both the attempts, the correct number is:”, and print the correct number.

Solution:

```
#Let us say we think of the number. Now the user has to guess the number in two attempts.
rand_no = 3
guess = input("Guess the number:")
if int(guess)==rand_no:
    print("Correct in the first attempt!")

#If the guess is incorrect, the program will execute the code block below
else:
    guess = input("Incorrect! Try again:")
    if int(guess)==rand_no:
        print("Correct in the second attempt")
    else:
        print("Incorrect in the both the attempts, the correct number was:", rand_no)
```

3.1 Loops

With loops, a piece of code can be executed repeatedly for a fixed number of times or until a condition is satisfied.

3.1.1 for loop

With a **for** loop, a piece of code is executed a fixed number of times.

We typically use **for** loops with an in-built python function called **range()** that supports **for** loops. Below is its description.

range(): The **range()** function returns a sequence of evenly-spaced integer values. It is commonly used in **for** loops to define the sequence of elements over which the iterations are performed.

Below is an example where the **range()** function is used to create a sequence of whole numbers upto 10. Ignore the **list()** function in the code below, as it will be introduced later.

```
print(list(range(10)))
```

[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

Note that the last element is one less than the integer specified in the `range()` function.

Using the `range()` function, the `for` loop can iterate over a sequence of numbers. See the example below.

Example: Print the first n elements of the [Fibonacci sequence](#), where n is an integer input by the user, such that $n > 2$. In a fibonacci sequence, each number is the sum of the preceding two numbers, and the sequence starts from 0,1. The sequence is as follows:

0,1,1,2,3,5,8,13,....

```
n=int(input("Enter number of elements:"))

#Initializing the sequence to start from 0, 1
n1=0;n2=1

#Printing the first two numbers of the sequence
print(n1)
print(n2)

for i in range(n-2): #Since two numbers of the sequence are already printed, n-2 numbers

    #Computing the next number of the sequence as the summation of the previous two number
    n3 = n1+n2
    print(n3)

    #As 'n3' is already printed, it is no longer the next number of the sequence.
    #Thus, we move the values of the variables n1 and n2 one place to the right to compute
    n1 = n2
    n2 = n3

print("These are the first", n, "elements of the fibonacci series")
```

Enter number of elements:6

0
1
1
2
3
5

These are the first 6 elements of the fibonacci series

As in the `if-elif-else` statement, the `for` loop uses indentation to indicate the piece of code to be run repeatedly.

Note that we have used an in-built python function

3.1.2 while loop

With a **while** loops, a piece of code is executed repeatedly until certain condition(s) hold.

Example: Print all the elements of the [Fibonacci sequence](#) less than n , where n is an integer input by the user, such that $n > 2$. In a fibonacci sequence, each number is the sum of the preceding two numbers, and the sequence starts from 0,1. The sequence is as follows:

0,1,1,2,3,5,8,13,....

```
n=int(input("Enter the value of n:"))

#Initializing the sequence to start from 0, 1
n1=0;n2=1

#Printing the first number of the sequence
print(n1)

while n2<n:

    #Print the next number of the sequence
    print(n2)

    #Computing the next number of the sequence as the summation of the previous two number
    n3 = n1+n2

    #As n2 is already printed, assigning n2 to n3, so that the next number of the sequence
    #Assigning n1 to n2 as n1 has already been used to compute the next number of the sequence
    n1 = n2
    n2 = n3
print("These are all the elements of the fibonacci series less than", n)
```

Enter the value of n:50

0
1
1
2
3
5
8

13
21
34

These are all the elements of the fibonacci series less than 50

3.1.3 Practice exercise 3

3.1.3.1

Write a program that identifies whether a number input by the user is prime or not.

Solution:

```
number = int(input("Enter a positive integer:"))

#Defining a variable that will have a value of 0 if there are no divisors
num_divisors=0

#Checking if the number has any divisors from 2 to half of the number
for divisor in range(2,int(number/2+1)):
    if number%divisor==0:

        #If the number has a divisor, setting num_divisors to 1, to indicate that the
        num_divisors = 1

        #If a divisor has been found, there is no need to check if the number has more
        #Even if the number has a single divisor, it is not prime. Thus, we 'break' out
        #If you don't 'break', your code will still be correct, it will just do some u
        break

#If there are no divisors of the number, it is prime, else not prime
if num_divisors==0:
    print("Prime")
else:
    print("Not prime")
```

Enter a positive integer:97
Prime

3.1.3.2

Update the program above to print the prime numbers starting from 2, and less than n where n is a positive integer input by the user.

Solution:

```
n = int(input("Enter a positive integer:"))

#Defining a variable - number_iterator. We will use this variable to iterate over all integers
#While iterating over each integer from 2 to n, we will check if the integer is prime or not
number_iterator = 2

print(number_iterator) #Since '2' is a prime number, we can print it directly (without checking)

#Continue to check for prime numbers until n (but not including n)
while(number_iterator<n):

    #After each check, increment the number_iterator to check if the next integer is prime
    number_iterator=number_iterator+1

    #Defining a variable that will have a value of 0 if there are no divisors
    num_divisors=0

    #Checking if the integer has any divisors from 2 to half of the integer being checked
    for divisor in range(2,int(number_iterator/2+1)):
        if number_iterator%divisor==0:

            #If the integer has a divisor, setting num_divisors to 1, to indicate that the integer is not prime
            num_divisors=1

            #If a divisor has been found, there is no need to check if the integer has more divisors
            #Even if the integer has a single divisor, it is not prime.
            #Thus, we 'break' out of the loop that checks for divisors
            break

    #If there are no divisors of the integer being checked, the integer is a prime number,
    if num_divisors==0:
        print(number_iterator)
```

```
Enter a positive integer:100
2
```

3
5
7
11
13
17
19
23
29
31
37
41
43
47
53
59
61
67
71
73
79
83
89
97

3.2 break statement

The `break` statement is used to unconditionally exit the innermost loop.

For example, suppose we need to keep asking the user to input year of birth and compute the corresponding age, until the user enters 1900 as the year of birth.

```
#The loop will continue to run indefinitely as the condition 'True' is always true
while True:
    year = int(input("Enter year of birth:"))
    if year==1900:
        break          #If the user inputs 1900, then break out of the loop
    else:
        print("Age = ",2022-year)    #Otherwise compute and print the age
```

Enter year of birth:1987

```
Age = 35
Enter year of birth:1995
Age = 27
Enter year of birth:2001
Age = 21
Enter year of birth:1900
```

3.2.1 Practice exercise 4

Write a program that finds and prints the largest factor of a number input by the user. Check the output if the user inputs 133.

Solution:

```
num = int(input("Enter an integer:"))

#Looping from the half the integer to 0 as the highest factor is likely to be closer to half
for i in range(int(num/2)+1,0,-1):
    if num%i==0:
        print("Largest factor = ",i)

        #Exiting the loop if the largest integer is found
        break
```

```
Enter an integer:133
Largest factor = 19
```

3.3 continue statement

The `continue` statement is used to continue with the next iteration of the loop without executing the lines of code below it.

For example, consider the following code:

```
for i in range(10):
    if i%2==0:
        continue
    print(i)
```

1
3
5
7
9

When the control flow reads the statement `continue`, it goes back to the beginning of the `for` loop, and ignores the lines of code below the statement.

3.3.1 Practice exercise 5:

Write a program that asks the user the question, “How many stars are in the Milky Way (in billions)?”. If the user answers 100, the program should print correct, and stop. However, if the user answers incorrectly, the program should print “incorrect”, and ask them if they want to try again. The program should continue to run until the user answers correctly, or they want to stop trying.

```
#Defining an infinite while loop as the loop may need to run indefinitely if the user keeps trying
while True:
    answer = input("How many stars are there in the Milky Way? ")
    if answer=='100':
        print("Correct")

        #Exiting the loop if the user answers correctly
        break
    else:
        print("Incorrect")
        try_again = input("Do you want to try again? (Y/N) ")
        if try_again=='Y':

            #Continuing with the infinite loop if the user wants to try again
            continue
        else:

            #Exiting the infinite loop if the user wants to stop trying
            break
```

```
How many stars are there in the Milky Way? 101
Incorrect
Do you want to try again? (Y/N) Y
How many stars are there in the Milky Way? 100
Correct
```

3.4 Loops with strings

Loops can be used to iterate over a string, just like we used them to iterate over a sequence of integers.

Consider the following string:

```
sentence = "She sells sea shells on the sea shore"
```

The i^{th} character of the string can be retrieved by its index. For example, the first character of the string `sentence` is:

```
sentence[0]
```

'S'

Slicing a string:

A part of the string can be sliced by passing the starting index (say `start`) and the stopping index (say `stop`) as `start:stop` to the index operator `[]`. This is called slicing a string. For a string `S`, the characters starting from the index `start` upto the index `stop`, but not including `stop`, can be sliced as `S[start:stop]`.

For example, the slice of the string `sentence` from index 4 to index 9, but not including 9 is:

```
sentence[4:9]
```

'sells'

Example:

Input a string, and count and print the number of "t"s.

```
string = input("Enter a sentence:")

#Initializing a variable 'count_t' which will store the number of 't's in the string
count_t = 0

#Iterating over the entire length of the string.
#The length of the string is given by the len() function
for i in range(len(string)):
```

```

        #If the ith character of the string is 't', then we count it
        if string[i]=='t':
            count_t = count_t+1

print("Number of 't's in the string = ",count_t)

```

Enter a sentence:Getting a tatto is not a nice experience
 Number of 't's in the string = 6

3.4.1 Practice exercise 6

Write a program that asks the user to input a string, and print the number of “*the*”s in the string.

```

string = input("Enter a sentence:")

#Defining a variable to store the count of the word 'the'
count_the = 0

#Looping through the entire length of the string except the last 3 letters.
#As we are checking three letters at a time starting from the index 'i', the last 3 letters
for i in range(len(string)-3):

    #Slicing 3 letters of the string and checking if they are 'the'
    if string[i:(i+3)]=='the':

        #Counting the words that are 'the'
        count_the = count_the+1
print("Number of 'the's in the string = ",count_the)

```

Enter a sentence:She sells the sea shells on the sea shore in the spring
 Number of 'the's in the string = 3

4 Data structures

In this chapter we'll learn about the python data structures that are often used or appear while analyzing data.

Tuple is a sequence of python objects, with two key characteristics: (1) the number of objects are fixed, and (2) the objects are immutable, i.e., they cannot be changed.

Tuple can be defined as a sequence of python objects separated by commas, and enclosed in rounded brackets (). For example, below is a tuple containing three integers.

```
tuple_example = (2,7,4)
```

We can check the data type of a python object using the *type()* function. Let us check the data type of the object *tuple_example*.

```
type(tuple_example)
```

tuple

Elements of a tuple can be extracted using their index within square brackets. For example the second element of the tuple *tuple_example* can be extracted as follows:

```
tuple_example[1]
```

7

Note that an element of a tuple cannot be modified. For example, consider the following attempt in changing the second element of the tuple *tuple_example*.

```
tuple_example[1] = 8
```

TypeError: 'tuple' object does not support item assignment

The above code results in an error as tuple elements cannot be modified.

4.0.1 Practice exercise 1

USA's GDP per capita from 1960 to 2021 is given by the tuple T in the code cell below. The values are arranged in ascending order of the year, i.e., the first value is for 1960, the second value is for 1961, and so on. Print the years in which the GDP per capita of the US increased by more than 10%.

```
T = (3007, 3067, 3244, 3375, 3574, 3828, 4146, 4336, 4696, 5032, 5234, 5609, 6094, 6726, 7226, 78
```

Solution:

```
#Iterating over each element of the tuple
for i in range(len(T)-1):

    #Computing percentage increase in GDP per capita in the (i+1)th year
    increase = (T[i+1]-T[i])/T[i]

    #Printing the year if the increase in GDP per capita is more than 10%
    if increase>0.1:
        print(i+1961)
```

```
1973
1976
1977
1978
1979
1981
1984
```

4.0.2 Concatenating tuples

Tuples can be concatenated using the + operator to produce a longer tuple:

```
(2,7,4) + ("another", "tuple") + ("mixed","datatypes",5)
```

```
(2, 7, 4, 'another', 'tuple', 'mixed', 'datatypes', 5)
```

Multiplying a tuple by an integer results in repetition of the tuple:

```
(2,7,"hi") * 3
```

```
(2, 7, 'hi', 2, 7, 'hi', 2, 7, 'hi')
```

4.0.3 Unpacking tuples

If tuples are assigned to an expression containing multiple variables, the tuple will be unpacked and each variable will be assigned a value as per the order in which it appears. See the example below.

```
x,y,z = (4.5, "this is a string", (("Nested tuple",5)))
```

```
x
```

4.5

```
y
```

```
'this is a string'
```

```
z
```

```
('Nested tuple', 5)
```

If we are interested in retrieving only some values of the tuple, the expression `*_` can be used to discard the other values. Let's say we are interested in retrieving only the first and the last two values of the tuple:

```
x,*_,y,z = (4.5, "this is a string", (("Nested tuple",5)), "99", 99)
```

```
x
```

4.5

```
y
```

```
'99'
```

4.0.4 Practice exercise 2

USA's GDP per capita from 1960 to 2021 is given by the tuple T in the code cell below. The values are arranged in ascending order of the year, i.e., the first value is for 1960, the second value is for 1961, and so on.

Write a function that has two parameters:

1. Year : which indicates the year from which the GDP per capita are available in the second parameter
2. Tuple of GDP per capita's: Tuple consisting of GDP per capita for consecutive years starting from the year mentioned in the first parameter.

The function should return a tuple of length two, where the first element of the tuple is the number of years when the increase in GDP per capita was more than 5%, and the second element is the most recent year in which the GDP per capita increase was more than 5%.

Call the function to find the number of years, and the most recent year in which the GDP per capita increased by more than 5%, since the year 2000. Assign the **number of years** returned by the function to a variable named **num_years**, and assign the most recent year to a variable named **recent_year**. Print the values of **num_years** and **recent_year**.

```
T = (3007, 3067, 3244, 3375,3574, 3828, 4146, 4336, 4696, 5032,5234,5609,6094,6726,7226,78
```

```
def gdp_inc(year,gdp_tuple):
    count=0
    for i in range(len(gdp_tuple)-1):

        #Computing the increase in GDP per capita for the (i+1)th year
        increase = (gdp_tuple[i+1]-gdp_tuple[i])/gdp_tuple[i]
        if increase>0.05:
            print(year+i)

        #Over-writing the value of recent_year if the increase in GDP per capita for a
        recent_year = year+i+1

        #Counting the number of years for which the increase in GDP per capita is more
        count = count+1
```

```

    return((count,recent_year))

num_years, recent_year = gdp_inc(2000,T[40:])
print("Number of years when increase in GDP per capita was more than 5% = ", num_years)
print("The most recent year in which the increase in GDP per capita was more than 5% = ",

```

2003

2004

2020

Number of years when increase in GDP per capita was more than 5% = 3

The most recent year in which the increase in GDP per capita was more than 5% = 2021

4.0.5 Tuple methods

A couple of useful tuple methods are `count`, which counts the occurrences of an element in the tuple and `index`, which returns the position of the first occurrence of an element in the tuple:

```
tuple_example = (2,5,64,7,2,2)
```

```
tuple_example.count(2)
```

3

```
tuple_example.index(2)
```

0

Now that we have an idea about tuple, let us try to think where it can be used.

<IPython.core.display.HTML object>

4.1 List

List is a sequence of python objects, with two key characteristics that differentiates it from tuple: (1) the number of objects are variable, i.e., objects can be added or removed from a list, and (2) the objects are mutable, i.e., they can be changed.

List can be defined as a sequence of python objects separated by commas, and enclosed in square brackets []. For example, below is a list consisting of three integers.

```
list_example = [2,7,4]
```

4.1.1 Adding and removing elements in a list

We can add elements at the end of the list using the *append* method. For example, we append the string 'red' to the list *list_example* below.

```
list_example.append('red')
```

```
list_example
```

```
[2, 7, 4, 'red']
```

Note that the objects of a list or a tuple can be of different datatypes.

An element can be added at a specific location of the list using the *insert* method. For example, if we wish to insert the number 2.32 as the second element of the list *list_example*, we can do it as follows:

```
list_example.insert(1,2.32)
```

```
list_example
```

```
[2, 2.32, 7, 4, 'red']
```

For removing an element from the list, the *pop* and *remove* methods may be used. The *pop* method removes an element at a particular index, while the *remove* method removes the element's first occurrence in the list by its value. See the examples below.

Let us say, we need to remove the third element of the list.

```
list_example.pop(2)
```

7

```
list_example
```

```
[2, 2.32, 4, 'red']
```

Let us say, we need to remove the element 'red'.

```
list_example.remove('red')
```

```
list_example
```

```
[2, 2.32, 4]
```

```
#If there are multiple occurrences of an element in the list, the first occurrence will be removed
list_example2 = [2,3,2,4,4]
list_example2.remove(2)
list_example2
```

```
[3, 2, 4, 4]
```

For removing multiple elements in a list, either `pop` or `remove` can be used in a `for` loop, or a `for` loop can be used with a condition. See the examples below.

Let's say we need to remove integers less than 100 from the following list.

```
list_example3 = list(range(95,106))
list_example3
```

```
[95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105]
```

```
#Method 1: For loop with remove
list_example3_filtered = list(list_example3) #
```

```

for element in list_example3:
    if element<100:
        list_example3_filtered.remove(element)
print(list_example3_filtered)

```

[100, 101, 102, 103, 104, 105]

Q1: What's the need to define a new variable `list_example3_filtered` in the above code?

A1: Replace `list_example3_filtered` with `list_example3` and identify the issue.

```

#Method 2: For loop with condition
[element for element in list_example3 if element>100]

```

[101, 102, 103, 104, 105]

4.1.2 List comprehensions

List comprehension is a compact way to create new lists based on elements of an existing list or other objects.

Example: Create a list that has squares of natural numbers from 5 to 15.

```

sqrt_natural_no_5_15 = [(x**2) for x in range(5,16)]
print(sqrt_natural_no_5_15)

```

[25, 36, 49, 64, 81, 100, 121, 144, 169, 196, 225]

Example: Create a list of tuples, where each tuple consists of a natural number and its square, for natural numbers ranging from 5 to 15.

```

sqrt_natural_no_5_15 = [(x,x**2) for x in range(5,16)]
print(sqrt_natural_no_5_15)

```

[(5, 25), (6, 36), (7, 49), (8, 64), (9, 81), (10, 100), (11, 121), (12, 144), (13, 169), (14, 196)]

4.1.3 Practice exercise 3

Below is a list consisting of responses to the question: “At what age do you think you will marry?” from students of the STAT303-1 Fall 2022 class.

```
exp_marriage_age=['24','30','28','29','30','27','26','28','30+','26','28','30','30','30','
```

Use list comprehension to:

4.1.3.1

Remove the elements that are not integers - such as *‘probably never’*, *‘30+’*, etc. What is the length of the new list?

Hint: The built-in python function of the `str` class - `isdigit()` may be useful to check if the string contains only digits.

```
exp_marriage_age_num = [x for x in exp_marriage_age if x.isdigit()==True]
print("Length of the new list = ",len(exp_marriage_age_num))
```

Length of the new list = 181

4.1.3.2

Cap the values greater than 80 to 80, in the clean list obtained in (1). What is the mean age when people expect to marry in the new list?

```
exp_marriage_age_capped = [min(int(x),80) for x in exp_marriage_age_num]
print("Mean age when people expect to marry = ", sum(exp_marriage_age_capped)/len(exp_marriage_age_capped))
```

Mean age when people expect to marry = 28.955801104972377

4.1.3.3

Determine the percentage of people who expect to marry at an age of 30 or more.

```
print("Percentage of people who expect to marry at an age of 30 or more =", str(100*sum([x for x in exp_marriage_age_capped if x >= 30])/len(exp_marriage_age_capped)))
```

Percentage of people who expect to marry at an age of 30 or more = 37.01657458563536 %

4.1.4 Concatenating lists

As in tuples, lists can be concatenated using the `+` operator:

```
import time as tm

list_example4 = [5,'hi',4]
list_example4 = list_example4 + [None,'7',9]
list_example4
```

```
[5, 'hi', 4, None, '7', 9]
```

For adding elements to a list, the `extend` method is preferred over the `+` operator. This is because the `+` operator creates a new list, while the `extend` method adds elements to an existing list. Thus, the `extend` operator is more memory efficient.

```
list_example4 = [5,'hi',4]
list_example4.extend([None, '7', 9])
list_example4
```

```
[5, 'hi', 4, None, '7', 9]
```

4.1.5 Sorting a list

A list can be sorted using the `sort` method:

```
list_example5 = [6,78,9]
list_example5.sort(reverse=True) #the reverse argument is used to specify if the sorting i
list_example5
```

```
[78, 9, 6]
```

4.1.6 Slicing a list

We may extract or update a section of the list by passing the starting index (say `start`) and the stopping index (say `stop`) as `start:stop` to the index operator `[]`. This is called *slicing* a list. For example, see the following example.

```
list_example6 = [4,7,3,5,7,1,5,87,5]
```

Let us extract a slice containing all the elements from the 3rd position to the 7th position.

```
list_example6[2:7]
```

```
[3, 5, 7, 1, 5]
```

Note that while the element at the **start** index is included, the element with the **stop** index is excluded in the above slice.

If either the **start** or **stop** index is not mentioned, the slicing will be done from the beginning or until the end of the list, respectively.

```
list_example6[:7]
```

```
[4, 7, 3, 5, 7, 1, 5]
```

```
list_example6[2:]
```

```
[3, 5, 7, 1, 5, 87, 5]
```

To slice the list relative to the end, we can use negative indices:

```
list_example6[-4:]
```

```
[1, 5, 87, 5]
```

```
list_example6[-4:-2:]
```

```
[1, 5]
```

An extra colon (':') can be used to slice every *i*th element of a list.

```
#Selecting every 3rd element of a list  
list_example6[::3]
```

```
[4, 5, 5]
```

```
#Selecting every 3rd element of a list from the end  
list_example6[::-3]
```

[5, 1, 3]

```
#Selecting every element of a list from the end or reversing a list  
list_example6[::-1]
```

[5, 87, 5, 1, 7, 5, 3, 7, 4]

4.1.7 Practice exercise 4

Start with the list [8,9,10]. Do the following:

4.1.7.1

Set the second entry (index 1) to 17

```
L = [8,9,10]  
L[1]=17
```

4.1.7.2

Add 4, 5, and 6 to the end of the list

```
L = L+[4,5,6]
```

4.1.7.3

Remove the first entry from the list

```
L.pop(0)
```

8

4.1.7.4

Sort the list

```
L.sort()
```

4.1.7.5

Double the list (concatenate the list to itself)

```
L=L+L
```

4.1.7.6

Insert 25 at index 3

The final list should equal [4,5,6,25,10,17,4,5,6,10,17]

```
L.insert(3,25)  
L
```

```
[4, 5, 6, 25, 10, 17, 4, 5, 6, 10, 17]
```

Now that we have an idea about lists, let us try to think where it can be used.

<IPython.core.display.HTML object>

Now that we have learned about lists and tuples, let us compare them.

Q2: A list seems to be much more flexible than tuple, and can replace a tuple almost everywhere. Then why use tuple at all?

A2: The additional flexibility of a list comes at the cost of efficiency. Some of the advantages of a tuple over a list are as follows:

1. Since a list can be extended, space is over-allocated when creating a list. A tuple takes less storage space as compared to a list of the same length.

2. Tuples are not copied. If a tuple is assigned to another tuple, both tuples point to the same memory location. However, if a list is assigned to another list, a new list is created consuming the same memory space as the original list.
3. Tuples refer to their element directly, while in a list, there is an extra layer of pointers that refers to their elements. Thus it is faster to retrieve elements from a tuple.

The examples below illustrate the above advantages of a tuple.

```
#Example showing tuples take less storage space than lists for the same elements
tuple_ex = (1, 2, 'Obama')
list_ex = [1, 2, 'Obama']
print("Space taken by tuple =",tuple_ex.__sizeof__()," bytes")
print("Space taken by list =",list_ex.__sizeof__()," bytes")
```

```
Space taken by tuple = 48 bytes
Space taken by list = 64 bytes
```

```
#Examples showing that a tuples are not copied, while lists can be copied
tuple_copy = tuple(tuple_ex)
print("Is tuple_copy same as tuple_ex?", tuple_ex is tuple_copy)
list_copy = list(list_ex)
print("Is list_copy same as list_ex?",list_ex is list_copy)
```

```
Is tuple_copy same as tuple_ex? True
Is list_copy same as list_ex? False
```

```
#Examples showing tuples takes lesser time to retrieve elements
import time as tm
tt = tm.time()
list_ex = list(range(1000000)) #List containinig whole numbers upto 1 million
a=(list_ex[::-2])
print("Time take to retrieve every 2nd element from a list = ", tm.time()-tt)

tt = tm.time()
tuple_ex = tuple(range(1000000)) #tuple containinig whole numbers upto 1 million
a=(tuple_ex[::-2])
print("Time take to retrieve every 2nd element from a tuple = ", tm.time()-tt)
```

```
Time take to retrieve every 2nd element from a list = 0.03579902648925781
Time take to retrieve every 2nd element from a tuple = 0.02684164047241211
```

4.2 Dictionary

A dictionary consists of key-value pairs, where the keys and values are python objects. While values can be any python object, keys need to be immutable python objects, like strings, integers, tuples, etc. Thus, a list can be a value, but not a key, as elements of list can be changed. A dictionary is defined using the keyword `dict` along with curly braces, colons to separate keys and values, and commas to separate elements of a dictionary:

```
dict_example = {'USA': 'Joe Biden', 'India': 'Narendra Modi', 'China': 'Xi Jinping'}
```

Elements of a dictionary can be retrieved by using the corresponding key.

```
dict_example['India']
```

```
'Narendra Modi'
```

4.2.1 Adding and removing elements in a dictionary

New elements can be added to a dictionary by defining a key in square brackets and assigning it to a value:

```
dict_example['Japan'] = 'Fumio Kishida'
dict_example['Countries'] = 4
dict_example
```

```
{'USA': 'Joe Biden',
 'India': 'Narendra Modi',
 'China': 'Xi Jinping',
 'Japan': 'Fumio Kishida',
 'Countries': 4}
```

Elements can be removed from the dictionary using the `del` method or the `pop` method:

```
#Removing the element having key as 'Countries'
del dict_example['Countries']

dict_example
```

```
{'USA': 'Joe Biden',  
 'India': 'Narendra Modi',  
 'China': 'Xi Jinping',  
 'Japan': 'Fumio Kishida'}
```

```
#Removing the element having key as 'USA'  
dict_example.pop('USA')
```

```
'Joe Biden'
```

```
dict_example
```

```
{'India': 'Narendra Modi', 'China': 'Xi Jinping', 'Japan': 'Fumio Kishida'}
```

New elements can be added, and values of existing keys can be changed using the `update` method:

```
dict_example = {'USA': 'Joe Biden', 'India': 'Narendra Modi', 'China': 'Xi Jinping', 'Countries': 3}  
dict_example
```

```
{'USA': 'Joe Biden',  
 'India': 'Narendra Modi',  
 'China': 'Xi Jinping',  
 'Countries': 3}
```

```
dict_example.update({'Countries': 4, 'Japan': 'Fumio Kishida'})
```

```
dict_example
```

```
{'USA': 'Joe Biden',  
 'India': 'Narendra Modi',  
 'China': 'Xi Jinping',  
 'Countries': 4,  
 'Japan': 'Fumio Kishida'}
```

4.2.2 Iterating over elements of a dictionary

The `items()` attribute of a dictionary can be used to iterate over elements of a dictionary.

```
for key,value in dict_example.items():
    print("The Head of State of",key,"is",value)
```

```
The Head of State of USA is Joe Biden
The Head of State of India is Narendra Modi
The Head of State of China is Xi Jinping
The Head of State of Countries is 4
The Head of State of Japan is Fumio Kishida
```

4.2.3 Practice exercise 5

The GDP per capita of USA for most years from 1960 to 2021 is given by the dictionary D given in the code cell below.

Find:

1. The GDP per capita in 2015
2. The GDP per capita of 2014 is missing. Update the dictionary to include the GDP per capita of 2014 as the average of the GDP per capita of 2013 and 2015.
3. Impute the GDP per capita of other missing years in the same manner as in (2), i.e., as the average GDP per capita of the previous year and the next year. Note that the GDP per capita is not missing for any two consecutive years.
4. Print the years and the imputed GDP per capita for the years having a missing value of GDP per capita in (3).

```
D = {'1960':3007,'1961':3067,'1962':3244,'1963':3375,'1964':3574,'1965':3828,'1966':4146,'
```

Solution:

```
print("GDP per capita in 2015 =", D['2015'])
D['2014'] = (D['2013']+D['2015'])/2

#Iterating over all years from 1960 to 2021
for i in range(1960,2021):

    #Imputing the GDP of the year if it is missing
    if str(i) not in D.keys():
        D[str(i)] = (D[str(i-1)]+D[str(i+1)])/2
        print("Imputed GDP per capita for the year",i,"is $",D[str(i)])
```


GDP per capita in 2015 = 56763
Imputed GDP per capita for the year 1969 is \$ 4965.0
Imputed GDP per capita for the year 1977 is \$ 9578.5
Imputed GDP per capita for the year 1999 is \$ 34592.0

4.3 Functions

If an algorithm or block of code is being used several times in a code, then it can be separately defined as a function. This makes the code more organized and readable. For example, let us define a function that prints prime numbers between *a* and *b*, and returns the number of prime numbers found.

```
#Function definition
def prime_numbers (a,b=100):
    num_prime_nos = 0

    #Iterating over all numbers between a and b
    for i in range(a,b):
        num_divisors=0

        #Checking if the ith number has any factors
        for j in range(2, i):
            if i%j == 0:
                num_divisors=1;break;

        #If there are no factors, then printing and counting the number as prime
        if num_divisors==0:
            print(i)
            num_prime_nos = num_prime_nos+1

    #Return count of the number of prime numbers
    return num_prime_nos
```

In the above function, the keyword `def` is used to define the function, `prime_numbers` is the name of the function, *a* and *b* are the arguments that the function uses to compute the output.

Let us use the defined function to print and count the prime numbers between 40 and 60.

```
#Printing prime numbers between 40 and 60
num_prime_nos_found = prime_numbers(40,60)
```

41
43
47
53
59

```
num_prime_nos_found
```

5

If the user calls the function without specifying the value of the argument `b`, then it will take the default value of 100, as mentioned in the function definition. However, for the argument `a`, the user will need to specify a value, as there is no value defined as a default value in the function definition.

4.3.1 Global and local variables with respect to a function

A variable defined within a function is local to that function, while a variable defined outside the function is global to that function. In case a variable with the same name is defined both outside and inside a function, it will refer to its global value outside the function and local value within the function.

The example below shows a variable with the name `var` referring to its local value when called within the function, and global value when called outside the function.

```
var = 5
def sample_function(var):
    print("Local value of 'var' within 'sample_function()' = ",var)

sample_function(4)
print("Global value of 'var' outside 'sample_function()' = ",var)
```

```
Local value of 'var' within 'sample_function()' = 4
Global value of 'var' outside 'sample_function()' = 5
```

4.3.2 Practice exercise 6

The object `deck` defined below corresponds to a deck of cards. Estimate the probability that a five card hand will be a [flush](#), as follows:

1. Write a function that accepts a hand of 5 cards as argument, and returns whether the hand is a flush or not.
2. Randomly pull a hand of 5 cards from the deck. Call the function developed in (1) to determine if the hand is a flush.
3. Repeat (2) 10,000 times.
4. Estimate the probability of the hand being a flush from the results of the 10,000 simulations.

You may use the function `shuffle()` from the `random` library to shuffle the deck everytime before pulling a hand of 5 cards.

```
deck = [{'value':i, 'suit':c}
for c in ['spades', 'clubs', 'hearts', 'diamonds']
for i in range(2,15)]
```

Solution:

```
import random as rm

#Function to check if a 5-card hand is a flush
def chck_flush(hands):

    #Assuming that the hand is a flush, before checking the cards
    yes_flush =1

    #Storing the suit of the first card in 'first_suit'
    first_suit = hands[0]['suit']

    #Iterating over the remaining 4 cards of the hand
    for j in range(1,len(hands)):

        #If the suit of any of the cards does not match the suit of the first card, the ha
        if first_suit!=hands[j]['suit']:
            yes_flush = 0;

        #As soon as a card with a different suit is found, the hand is not a flush and
        break;
    return yes_flush

flush=0
for i in range(10000):
```


4.4.1.2

If the object in (1) is a dictionary, what is the datatype of the elements within the values of the dictionary?

```
print("Datatype=",type(starbucks_drinks_nutrition[list(starbucks_drinks_nutrition.keys())])
```

```
Datatype= <class 'dict'>
```

4.4.1.3

How many calories are there in Iced Coffee?

```
print("Calories = ",starbucks_drinks_nutrition['Iced Coffee'][0]['value'])
```

```
Calories = 5
```

4.4.1.4

Which drink(s) have the highest amount of protein in them, and what is that protein amount?

```
#Defining an empty dictionary that will be used to store the protein of each drink
protein={}

for key,value in starbucks_drinks_nutrition.items():
    for nutrition in value:
        if nutrition['Nutrition_type']=='Protein':
            protein[key]=(nutrition['value'])

#Using dictionary comprehension to find the key-value pair having the maximum value in the
{key:value for key, value in protein.items() if value == max(protein.values())}
```

```
{'Starbucks® Doubleshot Protein Dark Chocolate': 20,
'Starbucks® Doubleshot Protein Vanilla': 20,
'Chocolate Smoothie': 20}
```

4.4.1.5

Which drink(s) have a fat content of more than 10g, and what is their fat content?

```
#Defining an empty dictionary that will be used to store the fat of each drink
fat={}
for key,value in starbucks_drinks_nutrition.items():
    for nutrition in value:
        if nutrition['Nutrition_type']=='Fat':
            fat[key]=(nutrition['value'])

#Using dictionary comprehension to find the key-value pair having the value more than 10
{key:value for key, value in fat.items() if value>=10}
```

```
{'Starbucks® Signature Hot Chocolate': 26.0, 'White Chocolate Mocha': 11.0}
```

5 Object-Oriented Programming

In Python, everything is an object, which makes it an object-oriented programming language.

Object-oriented programming is the one in which a program is based on *objects*. An object is an independent entity within the program and can cooperatively work with other objects. A program can be made up of one or more objects, which can leverage the functionality and information contained in other objects.

An object consists of two items:

1. **Attributes** - Attributes are the data stored within the object.
2. **Methods** - Methods are the functions defined within the object. Methods can use the object attributes (*or data stored within the object*) as well as accept additional data as arguments.

We have already seen several in-built python objects such as string objects, integer objects, float objects, list objects, tuple objects and dictionary objects, in previous chapters. Each of these objects have attributes and methods associated with them.

For example, consider a *integer* object named as `integer_example`.

```
integer_example = 5
```

The attributes and methods of this *integer* object can be seen by putting a `.` next to its name, and pressing the *tab* key. A dropdown menu consisting of the attributes and methods will appear as shown below.

```
<IPython.core.display.Image object>
```

A list of all attributes and methods associated with an object can be obtained with the `dir()` function. Ignore the ones with underscores - these are used by Python itself. The rest of them can be used to perform operations.

```
#This code is not executed to avoid printing a long list
dir(integer_example)
```

For example, an attribute of `integer_example` is `real`, which contains the real part of the number:

```
integer_example.real
```

5

A example of a method of `integer_example` is `as_integer_ratio()`, which returns a tuple containing the numerator and denominator of the integer when it is expressed as a fraction.

```
integer_example.as_integer_ratio()
```

(5, 1)

Note that attributes do not have any parenthesis after them as they are just data, and cannot accept arguments. On the other hand methods have parenthesis after them as they are functions that may or may not have arguments.

5.1 Class

A *class* is a template for objects. It contains the attributes and methods associated with the object of the class. As an analogy, the *class* `Cat` will consist of characteristics (or *attributes*) shared by all cats such as breed, fur color, etc., as well as capability to perform functions (or *methods*) such as run, meow, etc.

Instance: An *instance* is a specific realization of the object of a particular class. Continuing with the `Cat` analogy of a class, a particular cat is an *instance* of the class `Cat`. Similarly, in the example above, the object `integer_example` is an instance of the class *integer*. The words *object* and *instance* are often used interchangeably.

Creating an *instance* of a class is called **Instantiation**.

5.1.1 Creating your own class

Until now we saw examples of in-built Python classes, such as *integer*, *List*, etc. Now, we'll learn to create our own class that serves our purpose.

Below is a toy example of a class.


```

class ToyClass:
    def __init__(self, x, y):
        self.x = x
        self.y = y

    def add(self):
        return self.x + self.y

    def multiply(self):
        return self.x*self.y

```

We'll use the example above to explain the following terms:

- **The class statement:** We use the `class` statement to create a class. The [Python style guide](#) recommends to use CamelCase for class names.
- **The constructor (or the `__init__()` method):** A class typically has a method called `__init__`. This method is called a constructor and is automatically called when an object or instance of the class is created. The constructor initializes the attributes of the class. In the above example, the constructor accepts two values as arguments, and initializes its attributes `x` and `y` with those values.
- **The `self` argument:** This is the first argument to every method in the class. Whenever the class refers to one of its attributes or methods, it must precede them by `self`. The purpose of `self` is to distinguish the class's attributes and methods from other variables and functions in the program.

The class `ToyClass` consists of two attributes `x` and `y`, a constructor `__init__()`, and two methods `add()` and `multiply()`.

To create an *object* or *instance* of the class `ToyClass`, we'll use the class name with the values to be passed as argument to the constructor for initializing the *object* / *instance*.

```

toy_instance = ToyClass(6,12)

```

The `x` *attribute* of the class `ToyClass` can be called using the `.` operator with the object name:

```

toy_instance.x

```

To use the `multiply()` *method* of the class `ToyClass`, we'll use the `.` operator with the object name:

```
toy_instance.multiply()
```

72

5.1.2 Example: A class that analyzes a string

Let us create a class that analyzes a string.

```
class AnalyzeString:

    #Constructor
    def __init__(self, s):
        s = s.lower()
        self.words = s.split()

    #This method counts the numebr of words
    def number_of_words(self):
        return (len(self.words))

    #This method counts the number of words starting with the string s
    def starts_with(self,s):
        return len([x for x in self.words if x[:len(s)]==s])

    #This method counts the number of words of length n
    def words_with_length(self,n):
        return len([x for x in self.words if len(x)==n])

    #This method returns the frequency of the word w
    def word_frequency(self,w):
        return self.words.count(w)
```

Let us create an instance of the class `AnalyzeString()` to analyze a sentence.

```
#Defining a string
sentence = 'This sentence in an example of a string that we will analyse using a class we
```

```
#Creating an instance of class AnalyzeString()
sentence_analysis = AnalyzeString(sentence)
```

```
#The attribute 'word' contains the list of words in the sentence
sentence_analysis.words
```

```
['this',
 'sentence',
 'in',
 'an',
 'example',
 'of',
 'a',
 'string',
 'that',
 'we',
 'will',
 'analyse',
 'using',
 'a',
 'class',
 'we',
 'have',
 'defined']
```

```
#The method 'word_frequency()' provides the frequency of a word in the sentence
sentence_analysis.word_frequency('we')
```

2

```
#The method 'starts_with()' provides the frequency of number of words starting with a part
sentence_analysis.starts_with('th')
```

2

5.1.3 Practice exercise 1

Write a class called `PasswordManager`. The class should have a list called `old_passwords` that holds all of the user's past passwords. The last item of the list is the user's current

password. There should be a method called `get_password` that returns the current password and a method called `set_password` that sets the user's password. The `set_password` method should only change the password if the attempted password is different from all the user's past passwords. It should either print *'Password changed successfully!'*, or *'Old password cannot be reused, try again.'* Finally, create a method called `is_correct` that receives a string and returns a boolean `True` or `False` depending on whether the string is equal to the current password or not.

To initialize the object of the class, use the list below.

After defining the class:

1. Check the attribute `old_passwords`
2. Check the method `get_password()`
3. Try re-setting the password to `'ibiza1972'`, and then check the current password.
4. Try re-setting the password to `'oktoberfest2022'`, and then check the current password.
5. Check the `is_correct()` method

```
past_passwords = ['titanic1911','ibiza1972','montecarlo799']

class PasswordManager:
    def __init__(self,past_passwords):
        self.old_passwords = past_passwords

    def get_password(self):
        return self.old_passwords[len(self.old_passwords)-1]

    def set_password(self,new_password):
        if new_password not in self.old_passwords:
            self.old_passwords.append(new_password)
            print("Password changed!")
        else:
            print("Old password cannot be reused, try again.")

    def is_correct(self,password):
        if password == self.old_passwords[len(self.old_passwords)-1]:
            return True
        return False

passwd = PasswordManager(past_passwords)
```

5.2 Inheritance

In object-oriented programming there is a concept called *inheritance* where we can create a new class that builds off of another class. The new class gets all of the variables and methods of the class it is inheriting from (called the base class). It can then define additional variables and methods that are not present in the base class, and it can also override some of the methods of the base class. That is, it can rewrite them to suit its own purposes. Here is a simple example:

```
class Parent:

    def __init__(self, a, b):
        self.a = a

    def method1(self):
        return self.a+' should study!'

    def method2(self):
        return self.a+' does not study enough '

class Child(Parent):

    def __init__(self, a,b):
        self.a = a
        self.b = b

    def method1(self):
        return self.a+' should play with ' + self.b

    def method3(self):
        return self.a + ' does not play enough'
```

Note that when inheriting from a class, we indicate the parent class in parentheses in the class statement.

We see that `method1` is present in both the `Parent` and `Child` classes, while `method2` is only present in the `Parent` class. Let us understand how does the `Child` class use the methods of the `Parent` class, and what happens if a method with the same name is present in both the parent and child classes.

```
p = Parent('Sam', 'John')
c = Child('Sam','Pam')
```

```

print('Parent method 1: ', p.method1())
print('Parent method 2: ', p.method2())
print()
print('Child method 1: ', c.method1())
print('Child method 2: ', c.method2())
print('Child method 3: ', c.method3())

```

```

Parent method 1:  Sam should study!
Parent method 2:  Sam does not study enough

```

```

Child method 1:  Sam should play with Pam
Child method 2:  Sam does not study enough
Child method 3:  Sam does not play enough

```

We see in the example above that the child has overridden the parent's `method1`. The child has inherited the parent's `method2`, so it can use it without having to define it. The child also adds some features to the parent class, namely a new variable `b` and a new method, `method3`.

5.2.1 Practice exercise 2

Define a class that inherits the in-built Python class `list`, and adds a new method to the class called `nunique()` which returns the number of unique elements in the list.

Define the following list as an object of the class you created. Then:

1. Find the number of unique elements in the object using the method `nunique()` of the inherited class.
2. Check if the `pop()` method of the parent class works to pop an element out of the object.

```

list_ex = [1,2,5,3,6,5,5,5,12]

class list_v2(list):
    def nuinque(self):
        unique_elements = []
        for x in self:
            if x not in unique_elements:
                unique_elements.append(x)
        return len(unique_elements)

list_ex = list_v2(list_ex)

```

```
print("Number of unique elements = ", list_ex.nunique())
print("Checking the pop() method, the popped out element is", list_ex.pop())
```

Number of unique elements = 6

Checking the pop() method, the popped out element is 12

5.2.2 Practice exercise 3

Define a class named `PasswordManagerUpdated` that inherits the class `PasswordManager` defined in Practice exercise 1. The class `PasswordManagerUpdated` should have two methods, other than the *constructor*:

1. The method `set_password()` that sets a new password. The new password must only be accepted if it does not have any punctuations in it, and if it is not the same as one of the old passwords. If the new password is not acceptable, then one of the appropriate messages should be printed - (a) *Cannot have punctuation in password, try again*, or (b) *Old password cannot be reused, try again*.
2. The method `suggest_password()` that randomly sets and returns a password as a string comprising of 15 randomly chosen letters. Letters may be repeated as well.

A Assignment 1 (Control flow)

1. You may talk to a friend, discuss the questions and potential directions for solving them. However, you need to write your own solutions and code separately, and not as a group activity.
2. Do not write your name on the assignment.
3. Write your code in the *Code* cells of the Jupyter notebook. Ensure that the solution is written neatly enough to understand and grade.
4. Use [Quarto](#) to print the *.ipynb* file as HTML. You will need to open the command prompt, navigate to the directory containing the file, and use the command: `quarto render filename.ipynb --to html`. Submit the HTML file.
5. There are 5 points for clealiness and organization. The code should be commented and clearly written with intuitive variable names. For example, use variable names such as `number_input`, `factor`, `hours`, instead of `a`, `b`, `xyz`, etc.
6. The assignment is worth 100 points, and is due on **5th October 2022 at 11:59 pm**.

B 1

B.1 1(a)

You look at the clock and it is exactly 2pm. You set an alarm to go off in 510 hours. At what time does the alarm go off? If the answer is say, 4 pm, then your code should print - “The alarm goes off at 4 pm”.

(5 points)

B.2 1(b)

Write a program to solve the general version of the above problem. Ask the user for - (1) the time now (in hours), and (2) the number of hours for the alarm to go off. Your program should output the time at which the alarm goes off. Both the user inputs must be in {0, 1, 2..., 22, 23}. If the answer is, say 14:00 hours, then your program should print - “The alarm goes off at 14:00 hours.

Show the output of your program when the user inputs 7 as the current time, and 95 as the number of hours for the alarm to go off.

(10 points)

C 2

C.1 2(a)

Write a program that checks if a positive integer is prime or not. Show the output when the program is used to check if 89 is prime or not.

(5 points)

C.2 2(b)

Prompt the user to input a positive integer. Write a program that prints the **factors** of the positive integer input by the user. Show the output of the program if the user inputs 190.

(10 points)

C.3 2(c)

Prompt the user to input a positive integer. Update the program in 2(b) to print the **prime factors** of the positive integer input by the user. Show the output of the program if the user inputs 190.

(15 points)

C.4 2(d)

Update the program in 2(c), so that it prints “Incorrect input, please enter positive integer” if the user does not enter a positive integer, and then prompts the user to input a positive integer. The program should continue to prompt the user to enter a positive integer until the user successfully enters a positive integer. Show the output of the program if the user enters “seventy” in the first attempt, “#70” in the second attempt, and 70 in the third attempt.

(30 points)

C.5 3

Prompt the user to input an english sentence. Write a program that counts and prints the number of words in the sentence input by the user. The program should continue to run until the user inputs the sentence - “end program”

(20 points)

D Assignment 2 (Functions)

1. You may talk to a friend, discuss the questions and potential directions for solving them. However, you need to write your own solutions and code separately, and not as a group activity.
2. Do not write your name on the assignment.
3. Write your code in the *Code* cells and your answer in the *Markdown* cells of the Jupyter notebook. Ensure that the solution is written neatly enough to understand and grade.
4. Use [Quarto](#) to print the *.ipynb* file as HTML. You will need to open the command prompt, navigate to the directory containing the file, and use the command: `quarto render filename.ipynb --to html`. Submit the HTML file.
5. There are 5 points for clealiness and organization. The code should be commented and clearly written with intuitive variable names. For example, use variable names such as `number_input`, `factor`, `hours`, instead of `a`, `b`, `xyz`, etc.
6. The assignment is worth 100 points, and is due on **14th October 2022 at 11:59 pm**.

D.1 1(a)

Write a function that accepts a word, and a sentence as arguments, and returns the number of times the word occurs in the sentence.

Call the function, and print the returned value if the word is “*sea*”, and the sentence is “*She sells sea shells on the sea shore when the sea is calm.*” Note that this is just an example to check your function. Your function should work for any word and sentence.

(10 points)

D.2 1(b)

Ask the user to input a sentence. Use the function in 1(a) to find the word that occurs the maximum number of times in the sentence. Print the word and its number of occurrences. If multiple words occur the maximum number of times, then you can print any one of them.

Check your program when the user inputs the sentence, *“She sells sea shells on the sea shore when the sea is calm.”*. Your program must print, *“The word with the maximum number of occurrences is ‘sea’ and it occurs 3 times.”* Note that this is just an example to check your program. Your program must work for any sentence”

(20 points)

D.3 2(a)

Write a function that checks if an integer is prime. The function must accept the integer as an argument, and return **True** if the integer is prime, otherwise it must return **False**.

Call your function with the argument as 197.

(4 points)

D.4 2(b)

Write a function that checks if an integer is a factor of another integer. The function must accept both the integers as arguments, and return **True** if the integer is a factor, otherwise it must return **False**.

Call your function with the arguments as (19,85).

(3 points)

D.5 2(c)

Prompt the user to input a positive integer. Use the functions in 2(a) and 2(b) to print the prime factors of the integer. Your program should be no more than 4 lines (excluding the comments)

Check your program is the user inputs 190

(8 points)

D.6 3(a)

The tuple below named as `tuple_of_words` consists of words. Write a function that accepts a word, say `word_to_search` and the `tuple_of_words` as arguments, and finds if the `word_to_search` occurs in the `tuple_of_words` or not. This is very simple to do with the code `word_to_search in tuple_of_words`. However, this code is unfortunately very slow.

As the words in the `tuple_of_words` are already sorted in alphabetical order, we can search using a faster way, called binary search. To implement binary search in a function, start by comparing `word_to_search` with the middle entry in the `tuple_of_words`. If they are equal, then you are done and the function should return `True`. On the other hand, if the `word_to_search` comes before the middle entry, then search the first half of `tuple_of_words`. If it comes after the middle entry, then search the second half of `tuple_of_words`. Then repeat the process on the appropriate half of the `tuple_of_words` and continue until the word is found or there is nothing left to search, in which case the function should return `False`. The `<` and `>` operators can be used to alphabetically compare two strings.

You may write just one function or multiple functions to solve this problem.

Check your function if the `word_to_search` is:

1. `'rocket'`
2. `'rest'`

(25 points)

D.7 3(b)

Update the function in 3(a) to also print the number of iterations it took to find the `word_to_search` or fail to find the `word_to_search`.

Check your function if the `word_to_search` is:

1. `'rocket'`
2. `'rest'`

(10 points)

D.8 3(c)

Update the function in 3(b) to also print the index of `word_to_search` in `tuple_of_words` if the word is found in the tuple. For example, the index of 'abacus' is 0, the index of 'abdomen' is 1, and so on.

Check your function if the 'word_to_search' is:

1. 'rocket'
2. 'rest'

(10 points)

D.9 3(d)

What is the maximum number of iterations it may take for your function to search the `word_to_search`. You may either write a program to answer this question, or answer it analytically.

(5 points)

E Assignment 3 (Data structures)

1. You may talk to a friend, discuss the questions and potential directions for solving them. However, you need to write your own solutions and code separately, and not as a group activity.
2. Do not write your name on the assignment.
3. Write your code in the *Code* cells and your answer in the *Markdown* cells of the Jupyter notebook. Ensure that the solution is written neatly enough to understand and grade.
4. Use [Quarto](#) to print the *.ipynb* file as HTML. You will need to open the command prompt, navigate to the directory containing the file, and use the command: `quarto render filename.ipynb --to html`. Submit the HTML file.
5. There are 5 points for cleanliness and organization. The code should be commented and clearly written with intuitive variable names. For example, use variable names such as `number_input`, `factor`, `hours`, instead of `a`, `b`, `xyz`, etc.
6. The assignment is worth 100 points, and is due on **21st October 2022 at 11:59 pm**.

F 1 GDP of The USA

USA's GDP per capita from 1960 to 2021 is given by the tuple **T** in the code cell below. The values are arranged in ascending order of the year, i.e., the first value is for 1960, the second value is for 1961, and so on.

```
T = (3007, 3067, 3244, 3375, 3574, 3828, 4146, 4336, 4696, 5032, 5234, 5609, 6094, 6726, 7226, 78
```

F.1 1(a)

F.1.1 1(a)(i)

Use list comprehension to produce a list of the gaps between consecutive entries in **T**, i.e, the increase in GDP per capita with respect to the previous year. The list with gaps should look like: [60, 177, ...].

(6 points)

F.1.2 1(a)(ii)

Use the list developed in *1(a)(i)* to find the maximum gap size, i.e, the maximum increase in GDP per capita.

(2 points)

F.1.3 1(a)(iii)

Using list comprehension with the list developed in *1(a)(i)*, find the percentage of gaps that have size greater than \$1000.

(6 points)

F.2 1(b)

F.2.1 1(b)(i)

Create a dictionary `D`, where the **key** is the year, and **value** for the **key** is the increase in GDP per capita in that year with respect to the previous year, i.e., the gaps computed in $1(a)(i)$.

(6 points)

F.2.2 1(b)(ii)

Use the dictionary `D` to find the year when the GDP per capita increase was the maximum as compared to the previous year. Use the list comprehension method.

(6 points)

Hint: `[..... for in D.items() if]`

F.2.3 1(b)(iii)

Use the dictionary `D` to find the years when the GDP per capita decreased with respect to the previous year. Use the list comprehension method.

(6 points)

G 2 Ted Talks

G.1 2(a)

Read the file *TED_Talks.json* on ted talks using the code below. You will get the data in the object `TED_Talks_data`. Just look at the data structure of `TED_Talks_data`. You will need to know how the data is structured in lists/dictionaries to answer the questions below.

(2 points)

```
import json
with open("TED_Talks.json", "r") as file:
    TED_Talks_data=json.load(file)
```

G.2 2(b)

Find the number of talks in the dataset.

(2 points)

G.3 2(c)

Find the `headline`, `speaker` and `year_filmed` of the talk with the highest number of `views`.

(6 points)

G.4 2(d)

What are the mean and median number of `views` for a talk? Can we say that the majority of talks (i.e., more than 50% of the talks) have less `views` than the average number of `views` for a talk? Justify your answer.

(6 points)

G.5 2(e)

Do at least 25% of the talks have more `views` than the average number of `views` for a talk? Justify your answer.

(4 points)

G.6 2(f)

Find the `headline` of the talk that received the highest number of votes in the *Confusing* category.

(8 points)

G.7 2(g)

Find the `headline` and the `year_filmed` of the talk that received the highest percentage of votes in the *Fascinating* category.

Percentage of *Fascinating* votes for a ted talk =
$$\frac{\text{Number of votes in the Fascinating category}}{\text{Total votes in all categories}}$$

(10 points)

H 3 Poker

The object `deck` defined below corresponds to a deck of cards. Estimate the probability that a five card hand will be:

1. Straight
2. Three-of-a-kind
3. Two-pair
4. One-pair
5. High card

You may check the meaning of the above terms [here](#).

(25 points)

Hint:

Estimate these probabilities as follows.

1. Write a function that accepts a hand of 5 cards as argument, and returns relevant characteristics of a hand, such as the number of distinct card values, maximum occurrences of a value etc. Using the values returned by this function (may be in a dictionary), you can compute if the hand is of any of the above types (*Straight / Three-of-a-kind / two-pair / one-pair / high card*).
2. Randomly pull a hand of 5 cards from the `deck`. Call the function developed in (1) to get the relevant characteristics of the hand. Use those characteristics to determine if the hand is one of the five mentioned types (*Straight / Three-of-a-kind / two-pair / one-pair / high card*).
3. Repeat (2) 10,000 times.
4. Estimate the probability of the hand being of the above five mentioned types (*Straight / Three-of-a-kind / two-pair / one-pair / high card*) from the results of the 10,000 simulations.

You may use the function `shuffle()` from the library `random` to shuffle the deck everytime before pulling a hand of 5 cards.

You don't need to stick to the hint if you feel you have a better way to do it. In case you have a better way, you can claim 10 bonus points for this assignment.

```
deck = [{'value':i, 'suit':c}
for c in ['spades', 'clubs', 'hearts', 'diamonds']
for i in range(2,15)]
```

I Assignment 4 (Classes and Objects)

1. You may talk to a friend, discuss the questions and potential directions for solving them. However, you need to write your own solutions and code separately, and not as a group activity.
2. Do not write your name on the assignment.
3. Write your code in the *Code* cells and your answer in the *Markdown* cells of the Jupyter notebook. Ensure that the solution is written neatly enough to understand and grade.
4. Use [Quarto](#) to print the *.ipynb* file as HTML. You will need to open the command prompt, navigate to the directory containing the file, and use the command: `quarto render filename.ipynb --to html`. Submit the HTML file.
5. There are 6 points for cleanliness and organization. The code should be commented and clearly written with intuitive variable names. For example, use variable names such as `number_input`, `factor`, `hours`, instead of `a`, `b`, `xyz`, etc.
6. The assignment is worth 100 points, and is due on **5th November 2022 at 11:59 pm**.
7. There is a Bonus question worth 25 points. You can score 125 out of 100 in the assignment if all is correct. However, there is no partial credit for the bonus question.

I.1 Password Manager

Define a class named `PasswordManagerUpdated` that inherits the class `PasswordManager` defined in [Practice exercise 1](#). The class `PasswordManagerUpdated` should have two methods, other than the constructor:

The method `set_password(new_password)` that sets a new password. The new password must only be accepted if it does not have any punctuations in it, and if it is not the same as one of the old passwords. If the new password is not acceptable, then one of the appropriate messages should be printed - (a) Cannot have punctuation in password, try again, or (b) Old password cannot be reused, try again. The new password is a parameter to this method.

The method `suggest_password()` that randomly sets and returns a password as a string comprising of 15 randomly chosen letters of the English language. Letters may be repeated as well.

```
past_passwords = ['titanic1911', 'ibiza1972', 'montecarlo799']
```

Hint:

1. For generating letters, use `str.lowercase()`
2. For randomly selecting letters, use `random.randint()`
3. For iterating through punctuations, use `from string import punctuation`.

After defining the class, instantiate an object of this class with `past_passwords`. Then:

I.1.1

Check the attribute `old_passwords`.

(2 points)

I.1.2

Check the method `get_password()`

(2 points)

I.1.3

Try re-setting the password to 'ibiza1972', and then check the current password.

(4 points)

I.1.4

Try re-setting the password to 'oktoberfest%2022', and then check the current password.

(6 points)

I.1.5

Try re-setting the password to 'oktoberfest2022', and then check the current password.

(2 points)

I.1.6

Use the method `suggest_password()` to set a new password.

(6 points)

I.1.7

Check the `is_correct()` method by entering the correct password as set in the previous question.

(2 points)

I.2 SMS store manager

Create a new class, named as *SMS_store_manager*. This class will be used to store and manage SMSs of a person's cellphone. An object of this class, say *my_inbox*, will be initialized with a list of existing messages. The class attribute `messages` will store the existing messages during instantiation:

```
my_inbox(existing_messages)
```

Each message in the list will be represented as a dictionary. A sample message is shown below:

```
{'has_been_viewed':True, 'from_number':9348593356, time_arrived='19:50', 'date':'2022-10-2
```

The class should provide these methods:

```
my_inbox.add_new_arrival(from_number, time_arrived, text_of_SMS)
#Makes new SMS dictionary, inserts it after other messages
#in the store, i.e, in the list of messages. When creating this message, its
#'has_been_viewed' status is set False.

my_inbox.message_count()
#Returns the number of sms messages in my_inbox

my_inbox.get_unread_messages()
#Returns unread messages, i.e., messages with 'has_been_viewed' status as False
#Also changes the status of 'has_been_viewed' to True for all messages returned
#While returning unread messages, the 'has_been_viewed' status must not be returned
```

```
my_inbox.delete(i)      # Delete the message at index i
my_inbox.clear()        # Delete all messages from inbox
```

Once you define the class, instantiate an object of this class, and call it *harry_messages*. Initialize the object with the existing messages below.

```
existing_messages = [{'has_been_viewed':False, 'from_number':8769038451, 'time':'09:30','d
                    {'has_been_viewed':False, 'from_number':9579038373, 'time':'19:30'
                    {'has_been_viewed':True, 'from_number':8639568726, 'time':'10:30',
                    {'has_been_viewed':False, 'from_number':4567653456, 'time':'11:50'
                    {'has_been_viewed':False, 'from_number':5646786643, 'time':'18:50'
                    {'has_been_viewed':False, 'from_number':9845543492, 'time':'17:10'
                    {'has_been_viewed':True, 'from_number':8793450987, 'time':'13:20',
                    {'has_been_viewed':True, 'from_number':874556445, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':998456435, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':8769038451, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':7739984533, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':3443498738, 'time':'07:20',
```

Use the object *harry_messages* to:

I.2.1

Add a new message below:

```
from_number=8749373884;
time='07:25';
date='2022-10-29'
text_of_SMS='Hey, I want my bike back.'
```

(4 points)

I.2.2

Count the number of messages.

(2 points)

I.2.3

Print all the unread messages.

(6 points)

I.2.4

Try printing all the unread messages again (*Nothing should show up as all messages were read in in the previous question*).

(4 points)

I.2.5

Clear the inbox, and then count the number of messages.

(4 points)

I.3 Personalized SMS store manager

Inherit the class developed in the previous question to create a new class *SMS_personalized_store_manager*. This class will inherit all the methods of the class *SMS_store_manager*. However, it will have the following differences as compared to the parent class:

1. Instantiation

- a. During instantiation, it will initialize two additional attributes - **spam_words**, and **update_words**, along with the **messages** attribute which stores existing messages. Use the lists below to initialize the attributes of the class.
- b. Each message will be tagged as **spam**, **update**, or **personal** as follows. If the message contains any word / phrase in the list **spam_words**, then it will be tagged as **spam**. If the message is not tagged as spam, and it contains any word / phrase in the list **update_words**, then it will be tagged as **update**. If the message has not been tagged as **spam** or **update**, then it will be tagged as **personal**. The tag will appear as an additional key-value pair in the dictionary of each message, where the key will be **tag**, and the value will be 'spam', 'update' or 'personal'.

A message after tagging may look like:

```
{'has_been_viewed':True, 'from_number':9348593356, 'time_arrived':'19:50', 'date':'2022-10-2
```

```
#Lists for initializing the attributes of the class
existing_messages = [{'has_been_viewed':False, 'from_number':8769038451, 'time':'09:30','d
                    {'has_been_viewed':False, 'from_number':9579038373, 'time':'19:30'
                    {'has_been_viewed':True, 'from_number':8639568726, 'time':'10:30',
                    {'has_been_viewed':False, 'from_number':4567653456, 'time':'11:50',
                    {'has_been_viewed':False, 'from_number':5646786643, 'time':'18:50',
                    {'has_been_viewed':False, 'from_number':9845543492, 'time':'17:10',
                    {'has_been_viewed':True, 'from_number':8793450987, 'time':'13:20',
                    {'has_been_viewed':True, 'from_number':874556445, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':998456435, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':8769038451, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':7739984533, 'time':'07:20',
                    {'has_been_viewed':True, 'from_number':3443498738, 'time':'07:20',
spam_words=['100% more', '100% free', '100% satisfied', 'Additional income', 'Be your own
update_words = ['Your order', 'appointment', 'Reservation confirmed', 'Card Not Present',
```

2. The class will have two additional methods:

- a. `get_unread_messages_by_category()`: This method will return all the **unread** messages of a particular category, i.e, 'spam', 'update' or 'personal'. The method will accept the category as an argument. If no argument is specified by the user for the category, then all the **unread** messages must be displayed. Once unread messages are returned, they will be marked as read. While returning unread messages, the `has_been_viewed` status must be changed to True, but the status itself must not be returned.
- b. `get_messages_by_category()`: This method will return all the messages (both read or unread) of a particular category, i.e, 'spam', 'update' or 'personal'. The method will accept the category as an argument. If no argument is specified by the user for the category, then all the messages must be returned.

3. The class will modify the method `add_new_arrival(self,from_number, time, date, text_of_SMS)` of the parent class to tag a new message as 'spam', 'update' or 'personal'.

Once you define the class, instantiate an object of this class with the lists `existing_messages`, `spam_words` and `update_words`, and call it *ron_messages*:

I.3.1

Add a new message below:

```
from_number=8749373884;  
time='07:25';  
date='2022-10-29'  
text_of_SMS='Hey, I want my bike back.'
```

(6 points)

I.3.2

Print all the unread messages tagged as 'personal'. Use the method `get_unread_messages_by_category()`.

(10 points)

I.3.3

Print all the unread messages tagged as 'update'. Use the method `get_unread_messages_by_category()`.

(10 points)

I.3.4

Print all the unread messages. Use the method `get_unread_messages()`.

(10 points)

I.3.5

Print all the messages tagged as 'spam'. Use the method `get_messages_by_category()`.

(7 points)

I.3.6

Print all the messages tagged as 'update'. Use the method `get_messages_by_category()`.

(7 points)

I.4 Bonus question: Creating a new datatype - list of dictionaries

Read `movie_data` with the code below.

```
import json
with open("movies.json", encoding="utf8") as file:
    movie_data=json.load(file)
```

Inherit the in-built python class `list()` to create a new class `list_dict()`. This class will be used for objects that are a list of dictionaries, where all the dictionaries in the list have the same keys. Add a method in this class, named as `sort_by_dict_value()` that sorts the dictionaries of the list based on the values of the desired key in the dictionaries. Sorting can be done in ascending or descending order depending on the user. The key to be used for sorting and the order (ascending / descending) will be parameters to the method `sort_by_dict_value()`. If the sorting order is unspecified, use ascending as default.

Instantiate an object of the class `list_dict()` with `movie_data`.

If the name of the object is `mov`, then the method `sort_by_dict_value()` may be called as:
`mov.sort_by_dict_value(movie_parameter, ascending = True)`

where `movie_parameter` can be any key of the dictionaries, using the values of which the list of dictionaries has to be sorted.

1.4.1

Use the method `sort_by_dict_value()` to sort the list of dictionaries in increasing order of **Production Budget**. What is the name of the 45th movie in the sorted list of dictionaries?

1.4.2

Use the method `sort_by_dict_value()` to sort the list of dictionaries in decreasing order of **Worldwide Gross**. What is the name of the 2nd movie in the sorted list of dictionaries?

(25 points - No partial credit)

J Datasets

Datasets used in the book can be found [here](#)