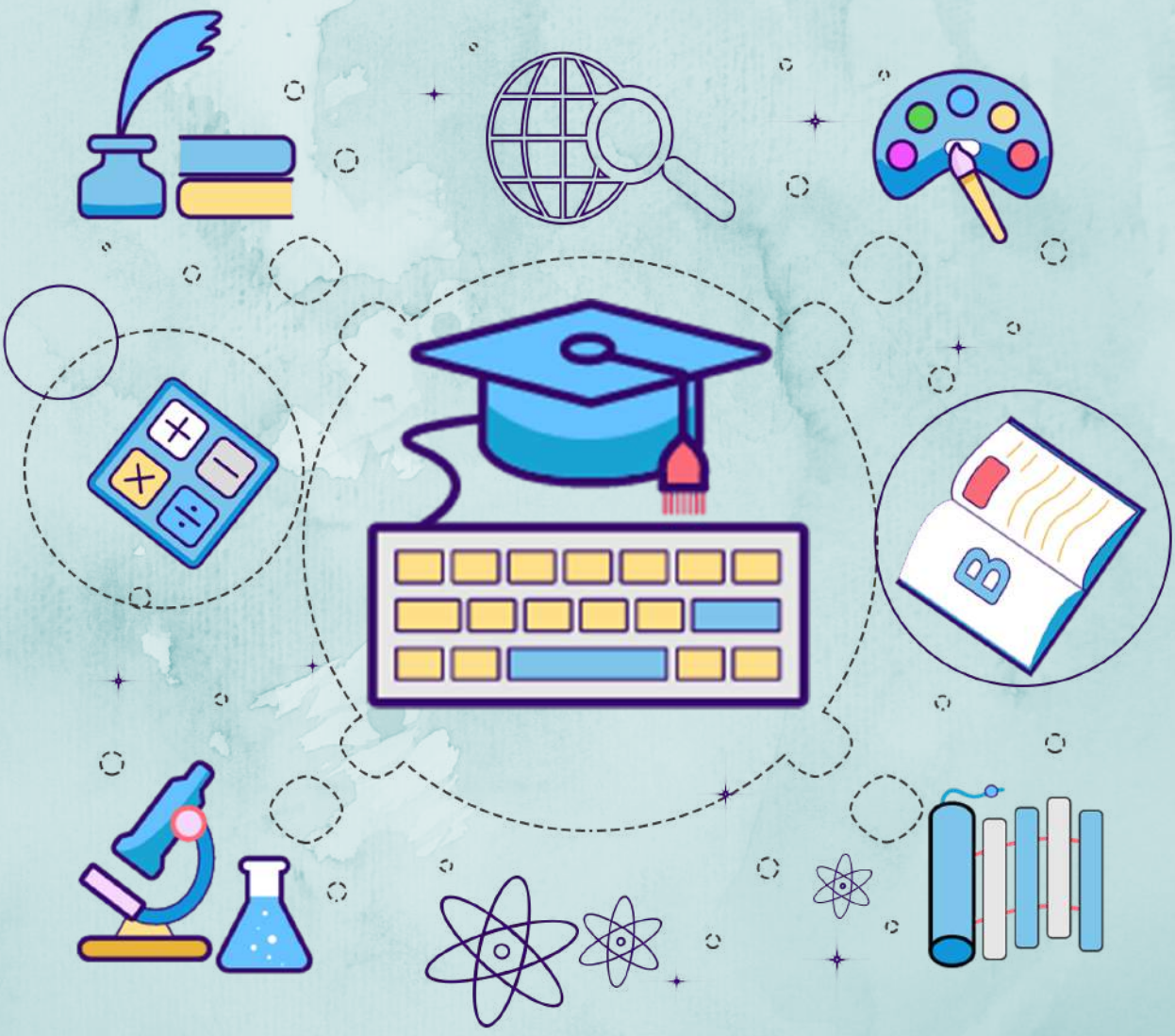
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

# KeralaNotes



SYLLABUS | STUDY MATERIALS | TEXTBOOK

PDF | SOLVED QUESTION PAPERS

www.keralanotes.com

KeralaNotes

# KTU STUDY MATERIALS

# PROGRAMMING IN PYTHON

## CST 362

# Module 5

Related Link :

- KTU S6 CSE NOTES 2019 SCHEME
- KTU S6 SYLLABUS CSE COMPUTER SCIENCE
- KTU PREVIOUS QUESTION BANK S6 CSE SOLVED
- KTU CSE TEXTBOOKS S6 B.TECH PDF DOWNLOAD
- KTU S6 CSE NOTES | SYLLABUS | QBANK | TEXTBOOKS DOWNLOAD

# TUTORIAL QUESTIONS

## MODULE-V

1. Add two matrix and find the transpose of the result ( university question)

Ans:

```
def readmatrix(x,r,c):
    for i in range(r):
        for j in range(c):
            x[i][j]=int(input('enter elements row by row'))
import numpy as np
r1=int(input('rows of a'))
c1=int(input('columns of a'))
r2=int(input('rows of b'))
c2=int(input('columns of b'))
if r1!=r2 or c1!=c2:
    print("cant add matrices")
else:
    A=np.zeros((r1,c1))
    print("Enter the elements of A")
```

```
readmatrix(A,r1,c1)

B=np.zeros((r2,c2))

print("Enter the elements of B")

readmatrix(B,r2,c2)

print("Matrix A")

print(A)

print("Matrix B")

print(B)

C=A+B

print("sum")

print(C)

print("transpose of sum")

print(C.T)
```

2. Create a dataframe from a list of data and set the index.

ANS:

```
import pandas as pd

df = pd.DataFrame(

[[21, 'Amol', 72, 67],[23, 'Lini', 78, 69],[32, 'Kiku', 74, 56],[52, 'Ajit', 54, 76]],
columns=['rollno', 'name', 'physics', 'botony'])
```

```
print('DataFrame with default index\n', df)
#set column as index
df = df.set_index('rollno')
print('\nDataFrame with column as index\n',df)
```

3. Write data to an excel file.

ANS:

```
import pandas as pd
# create dataframe
df_marks = pd.DataFrame({'name': ['Somu', 'Kiku', 'Amol', 'Lini'],
    'physics': [68, 74, 77, 78],
    'chemistry': [84, 56, 73, 69],
    'algebra': [78, 88, 82, 87]})
# create excel writer object
writer = pd.ExcelWriter('output.xlsx')
# write dataframe to excel
df_marks.to_excel(writer)
# save the excel
writer.save()
```

```
print('DataFrame is written successfully to Excel File.')
```

4. Read data from an excel file.

ANS:

```
# Program to extract a particular row value
import xlrd
loc = ("stud.xlsx")
wb = xlrd.open_workbook(loc)
sheet = wb.sheet_by_index(0)
#extracting column names
print(sheet.cell_value(0, 0),sheet.cell_value(0, 1),sheet.cell_value(0, 2))
for i in range(1,sheet.nrows):
    print(sheet.row_values(i))
```

5. Write Python program to write the data given below to a CSV file.(university question)

| SN | Name | Country | Contribution | Year |
|---|---|---|---|---|
| 1 | Linus Torvalds | Finland | Linux Kernel | 1991 |

2 Tim Berners-Lee England World Wide Web 1990

3 Guido van Rossum Netherlands Python 1991

Ans:

```
import pandas as pd
# dictionary of lists
# creating a dataframe from a dictionary
df = pd.DataFrame([[1,' Linus Torvalds','Finland','Linux Kernel ',1991],
    [2,'Tim Berners-Lee','England','World Wide Web',1990],
    [3,'Guido van Rossum','Netherlands','Python',1991]],
columns=['SN','Name','Country','Contribution','Year'])
print("data frame with defaut index=",df)
df=df.set_index('SN')
print("data frame with SN as index=",df)
print(df)
df.to_csv('inventors.csv')
```

6. Create a data frame from the dictionary of lists.

Ans:

```
    import pandas as pd
    # dictionary of lists
```

```
dict = {'name':["aparna", "pankaj", "sudhir", "Geeku"],
 'degree': ["MBA", "BCA", "M.Tech", "MBA"],
'score':[90, 40, 80, 98]}
# creating a dataframe from a dictionary
df = pd.DataFrame(dict)
print(df)
```

7. Given a file "auto.csv" of automobile data with the fields index, company, body-style, wheel-base, length, engine-type, num-of-cylinders, horsepower average-mileage, and price, write Python codes using Pandas to

   1) Clean and Update the CSV file

   2) Find the most expensive car company name

   3) Print all toyota car details

   4) Print total cars of all companies

   5) Find the highest priced car of all companies

   6) Find the average mileage of all companies

   7) Sort all cars by Price column ( university question)

Ans:

Reading the data file and showing the first five records

```
import pandas as pd

df = pd.read_csv("Automobile_data.csv")

df.head(5)
```

| | index | company | body-style | wheel-base | length | engine-type | num-of-cylinders | horsepower | average-mileage | price |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | alfa-romero | convertible | 88.6 | 168.8 | dohc | four | 111 | 21 | 13495.0 |
| 1 | 1 | alfa-romero | convertible | 88.6 | 168.8 | dohc | four | 111 | 21 | 16500.0 |
| 2 | 2 | alfa-romero | hatchback | 94.5 | 171.2 | ohcv | six | 154 | 19 | 16500.0 |
| 3 | 3 | audi | sedan | 99.8 | 176.6 | ohc | four | 102 | 24 | 13950.0 |
| 4 | 4 | audi | sedan | 99.4 | 176.6 | ohc | five | 115 | 18 | 17450.0 |

```
#This will show last 7 rows

df.tail(7)
```

1) Clean and Update the CSV file

```
import pandas as pd

df = pd.read_csv("Automobile_data.csv",

    na_values={

'price':["?","n.a"],

'stroke':["?","n.a"],

'horsepower':["?","n.a"],

'peak-rpm':["?","n.a"],

'average-mileage':["?","n.a"]})

print (df)
```

```
df.to_csv("Automobile_data.csv")
```

2) Find the most expensive car company name

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
df = df [['company','price']][df.price==df['price'].max()]
print(df)
```

3) Print all toyota car details

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
print(df[df['company']=='toyota'])
```

OR

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
car_Manufacturers = df.groupby('company')
toyotaDf = car_Manufacturers.get_group('toyota')
toyotaDf
```

4)Print total cars of all companies

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
df.groupby('company')['company'].count()
```

OR

```
import pandas as pd
df['company'].value_counts()
```

5) Find the highest priced car of all companies

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
df.groupby('company')[['company','price']].max()
```

6) Find the average mileage of all companies

```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
df.groupby('company')[['company','average-mileage']].mean()
```

7) Sort all cars by Price column

```
import pandas as pd

df = pd.read_csv("Automobile_data.csv")

df.sort_values(by=['price', 'horsepower'], ascending=False)[['company','price']]
```

8. Create a stud.csv file containing rollno, name, place and mark of students. Use this file and do the following

   a) Read and display the file contents

   b) Set rollno as index

   c) Display name and mark

   d) rollno,Name and mark in the order of name

   e) Display the rollno,name, mark in the descending order of mark

   f) Find the average mark,median and mode

   g) Find minimum and maximum marks

   h) variance and standard deviation of marks

   i) display the histogram of marks

   j) remove the place column ( university question)

ANS:

a)

```
import pandas as pd
```

```
df = pd.read_csv("stud.csv")

print(df)

    rollno   name      place          mark
0   101      binu      ernkulam       45

1   103      ashik     alleppey       35

2   102      faisal    kollam         48

3   105      biju      kotayam        25

4   106      anu       thrisur        25

5   107      padma     kylm           25
```

b)Set rollno as index

```
df=df.set_index('rollno')

print(df)

            name place mark

rollno

101         binu ernkulam 45

103         ashik alleppey 35

102         faisal kollam 48

105         biju kotayam 25
```

106 anu thrisur 25

107 padma kylm 25

c)Display name and mark

```
df=df[['name','mark']]
print(df)
```

```
name    mark
binu    45
ashik   35
faisal  48
biju    25
anu     25
padma   25
```

d) rollno,Name and mark in the order of name

```
df=df[['name','mark']]
df=df.sort_values('name')
print(df)
```

```
    name mark
```

```
rollno
106      anu      25
103      ashik   35
105     biju      25
101      binu     45
102 faisal 48
107 padma 25
```

e) Display the rollno,name, mark in the descending order of mark

```
df=df.sort_values(by='mark',ascending=False)
print(df)
            name   mark
rollno
102         faisal    48
101         binu     45
103         ashik     35
106         anu       25
105         biju       25
107         padma   25
```

f) Find the average mark,median and mode

```
print(df['mark'].mean())

print(df['mark'].median())

print(df['mark'].mode())
```

```
33.833333333333336

30.0

25
```

g)Find minimum and maximum marks

```
print(df['mark'].min())

print(df['mark'].max())
```

```
25

48
```

h)variance and standard deviation of marks

```
print(df['mark'].var())

print(df['mark'].std())
```

```
112.16666666666667

10.59087657687817
```

i) display the histogram of marks

```
import matplotlib.pyplot as plt
```

```
plt.hist(df['mark'])
```

j) remove the place column

```
df.drop(['place'],axis=1,inplace=True)
print(df)
```

```
  rollno name mark
0 101    binu 45
1 103    ashik 35
2 102    faisal 48
3 105    biju 25
4 106   ann 25
5 107   padma 25
```

9. Given the sales information of a company as CSV file with the following fields month_number, facecream, facewash, toothpaste, bathingsoap, shampoo, moisturizer, total_units, total_profit. Write Python codes to visualize the data as follows
    1) Toothpaste sales data of each month and show it using a scatter plot.
    2) Face cream and face wash product sales data and show it using the bar chart.
    3) Calculate total sale data for last year for each product and show it using a Pie chart. ( university question)

ANS:

1)

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("sales_data.csv")
monthList  = df ['month_number'].tolist()
toothPasteSalesData = df ['toothpaste'].tolist()
plt.scatter(monthList, toothPasteSalesData, label = 'Tooth paste Sales data')
plt.xlabel('Month Number')
plt.ylabel('Number of units Sold')
plt.legend(loc='upper left')
plt.title(' Tooth paste Sales data')
plt.xticks(monthList)
plt.grid(True, linewidth= 1, linestyle="--")
plt.show()
```

2)

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("sales_data.csv")
monthList  = df ['month_number'].tolist()
```

```
faceCremSalesData   = df ['facecream'].tolist()

faceWashSalesData   = df ['facewash'].tolist()


plt.bar([a-0.25 for a in monthList], faceCremSalesData, width= 0.25, label =
'Face Cream sales data', align='edge')

plt.bar([a+0.25 for a in monthList], faceWashSalesData, width= -0.25, label
= 'Face Wash sales data', align='edge')

plt.xlabel('Month Number')

plt.ylabel('Sales units in number')

plt.legend(loc='upper left')

plt.title(' Sales data')


plt.xticks(monthList)

plt.grid(True, linewidth= 1, linestyle="--")

plt.title('Facewash and facecream sales data')

plt.show()
```

3)

```
import pandas as pd

import matplotlib.pyplot as plt
```

```
df = pd.read_csv("sales_data.csv")

monthList  = df ['month_number'].tolist()

labels = ['FaceCream', 'FaseWash', 'ToothPaste', 'Bathing soap', 'Shampoo',
'Moisturizer']

salesData   = [df ['facecream'].sum(), df ['facewash'].sum(), df
['toothpaste'].sum(), df ['bathingsoap'].sum(), df ['shampoo'].sum(), df
['moisturizer'].sum()]

plt.axis("equal")

plt.pie(salesData, labels=labels, autopct='%1.1f%%')

plt.legend(loc='lower right')

plt.title('Sales data')

plt.show()
```