Final Assessment Test Decemeber 2020

Programme	:	**B.Tech (CSE)**	Semester	:	**FALL2021-22**
Course Title	:	**Foundations of Data Analytics Lab**	Code	:	**CSE3505**
			Class Nbr(s)	:	CH2021221001839
Faculty(s)	:	**Dr. B. Radhika Selvamani**	Slot		**L43+44 L55+56**
Time	:	**1.40 hours**	Max. Marks	:	**50**

Answer all the Questions

Question 1

Q.No.		Question Description	Marks
1.	a)	Read the given table. Print the number of rows and colums. Order the table based on volume. Remove duplicates and print the number of rows and columns	**[5]**
	b)	Add a new column "grade" which categorizes the tree based on the volume as below 20 (SMALL) Below 30 and 20 and above (MEDIUM), above 30 (LARGE)	**[10]**
	c)	Group by the table by column "grade" and and calculate the mean girth and height within each category . Draw a box plot for the same.	**[10]**
2.	a)	For the given table airtravels some values of missing. Please fill them using mean imputation and linear regression imputation. Identify the closest match among both	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[10]**
	c)	**Rmd File and Outcome**	**[10]**

Programme	:	**B.Tech (CSE)**	Semester	:	**FALL2021-22**
Course Title	:	**Foundations of Data Analytics Lab**	Code	:	**CSE3505**
			Class Nbr(s)	:	CH2021221001839
Faculty(s)	:	**Dr. B. Radhika Selvamani**	Slot		**L43+44 L55+56**
Time	:	**1.40 hours**	Max. Marks	:	**50**

Answer all the Questions

Question 2

Q.No.		Question Description	Marks
1.	a)	Read the given table. Print the number of rows and colums. Order the table based on year of joining. Remove duplicates and print the number of rows and columns	**[5]**
	b)	Add a new column "age during join" based on the date of birth and the date of joining.	**[10]**
	c)	Group by the table by column "region" and find the average salary within each region. Plot the violin plot for the same. Write your observations	**[10]**
2.	a)	In the given table create a new column "year of join " by obtaining the year from the date of join. Calculate the correlation between age and salary and write your observations. Visualize using suitable plots the correlation between region of the person and his salary.	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[10]**
	c)	**Rmd File and Outcome**	**[10]**

Final Assessment Test Decemeber 2020

Programme	:	**B.Tech (CSE)**	Semester	:	**FALL2021-22**
Course Title	:	**Foundations of Data Analytics Lab**	Code	:	**CSE3505**
			Class Nbr(s)	:	CH2021221001839
Faculty(s)	:	**Dr. B. Radhika Selvamani**	Slot		**L43+L44+L55+L56**
Time	:	**1.40 hours**	Max. Marks	:	**50**

Answer all the Questions

Question 3

Q.No.		Question Description	Marks
1.	a)	Clean the table and split the "range(yearfrom-yearto)" column in to two columns "min" and "max". Remove all special character and make them numeric. Handle missing values.	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[10]**
	c)	**Rmd File and Outcome**	**[10]**
2.	a)	Plot suitable chart to visualize the drop /rise of cases of HIV across year for different countries	**[10]**
	b)	Provide a chart to identify the regions which have been successful in bringing down the HIV cases from 1990to 2005	[5]
	c)	Identify the top 5 countries which have least drop rate of the disease over the years.	[5]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Final Assessment Test December 2020

Programme	: B.Tech (CSE)	Semester	: FALL2021-22
Course Title	: **Foundations of Data Analytics Lab**	Code	: **CSE3505**
		Class Nbr(s) :	CH2021221001839
Faculty(s)	: **Dr. B. Radhika Selvamani**	Slot	**L43+44 L55+56**
Time	: **1.40 hours**	Max. Marks :	**50**

Answer all the Questions

Question 4

Q.No.		Question Description	Marks
1.	a)	Find suitable method to fill the missing values	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[10]**
	c)	**Rmd File and Outcome**	**[10]**
2.	a)	Read the data homes and order the data based on taxes	**[5]**
	b)	Group the data based on the interval of 10 years for the age data. Plot the bar chart for the average tax	**[5]**
	c)	Fond the summary statistics of the acres and taxes data for the interval age calculated in b	**[10]**

Final Assessment Test Decemeber 2020

Programme	:	**B.Tech (CSE)**	Semester	:	**FALL2021-22**
Course Title	:	**Foundations of Data Analytics Lab**	Code	:	**CSE3505**
			Class Nbr(s)	:	CH2021221001839
Faculty(s)	:	**Dr. B. Radhika Selvamani**	Slot		**L43+44 L55+56**
Time	:	**1.40 hours**	Max. Marks	:	**50**

Answer all the Questions

Question 4

Q.No.		Question Description	Marks
1.	a)	For the cars data, find the suitable 3 correlated columns for the column "acceleration". Visualize the same using suitable charts	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[10]**
	c)	**Rmd File and Outcome**	**[10]**
2.	a)	Find the summary statistics of score within the interval of 5 years and compare them with each other using bar chart or violin plot	**[10]**
	b)	**Algorithms Steps and/or Explanations**	**[5]**
	c)	**Rmd File and Outcome**	**[5]**