CIS3389 Final Exam Spring 2020

Instructions

1. This is a take home exam. Use jupyter notebook to write the programs.
2. Download the question paper and the 'Titanic.csv' file from TRACS>Assignments>Final Exam.
3. Create separate jupyter notebook file for each question and name each file with your first name followed by the corresponding question number. For example, Question 1 file should be named as 'xxx_Q1.ipynb', where xxx is your first name.
4. Codes for each question should have the proper part numbers in each file. For example, a., b. etc. If no part numbers, points will be deducted based on how many are missing.
5. Submit answer files in TRACS>Assignments>Final Exam by 05/04/2020 9:00am.

Deliverables

Three .ipynb files for three questions and one pdf file from Question2**.**

Total points: 100 (points for each part mentioned after each question in parenthesis)

Question 1 (20 points)

a. Create a 5*3 NumPy array with values ranging from 0 to 15(15 excluded), name it ar1 and display it. (**2**)
b. Add 5 to each elements of this array. (**1**)
c. Slice array ar1. We only want last two elements of 2^{nd} and 3^{rd} rows (2^{nd} and 3^{rd} are usual row numbers, not index numbers). (**4**)
d. Change the first element of the last row of ar1 to 25. Now, display the first column only to show the number has been changed successfully. (**3+3**)
e. Create a 3*4 NumPy array with random integers between 40 and 120, name it ar2. Display the array. (**3**)
f. Change the datatype of ar2 to float. (**1**)
g. Find the minimum and maximum values in ar2. (**3**)

Question 2 (40 points)

a. Create a pandas data frame of the monthly temperature data for two cities (shown in the table below) (**6**)
b. Draw line plot for city A to show how high and low temperatures change over a year. Chart title should be 'City A Weather Trend'. (Two lines for high and low temperatures should be in the same chart). Use solid line style and circle marker for the chart. (**6**) What can you tell about the high and low temperature trends of City A from this chart? (write this answer in words and comment out) (**2**)

c. Draw line plot for city B to show how high and low temperatures change over a year. Chart title should be 'City B Weather Trend'. (Two lines for high and low temperature should be in the same chart). Use dashed line style and triangle-down marker for the chart. (**6**)
d. Draw stacked bar plots to show the comparison of high temperature of city B vs city A. Chart title should be 'City B vs City A High Temperature'. (**6**)
e. Draw a pie chart for CityA high temperature. Use month names as labels. Explode the slice for the month of 'October'. Show the percentages for each slice. Chart title should be 'Pie Chart for City A High Temperature' (**8**)

Make sure you have proper labels for x and y axes and legends for charts b, c, and d.

-arrange all the 4 charts in figure (size = (20,10)) in two rows and three columns and display. Save the figure as pdf file through your program and name it 'Q2charts.pdf'. (**6**)

Month	CityA_High	CityA_Low	CityB_High	CityB_Low
Jan	33	19	54	26
Feb	37	22	59	30
Mar	44	28	67	37
Apr	56	38	76	46
May	67	47	83	56
Jun	75	57	91	65
Jul	79	62	93	69
Aug	78	60	94	67
Sep	71	53	84	59
Oct	60	42	75	48
Nov	50	35	63	36
Dec	38	24	55	28

Question 3 (40 points)

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

"Titanic.csv" contains data describing the relevant information about the passengers (dictionary is provided below).[1]

Variable	Definition

Survived	Survival (0 = No, 1 = Yes)
Pclass	Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
Gender	Gender
Age	Age in years
Sibsp	# of siblings / spouses aboard Titanic
Parch	# of parents / children aboard Titanic
Fare	Ticket fare
Cabin	Cabin number
Embarked	Port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

a. Read the dataset 'Titanic.csv' using pandas and create a dataframe. (**2**)
b. Check the shape of the data. (**2**)
c. Calculate mean and standard deviation of Fare for each Ticket class (calculate mean and standard deviation together, not in two separated codes). (**5**)
d. Drop the rows with missing values. (**3**)
   How many rows are left in the dataframe after dropping rows with missing values? (write this answer in words and comment out). (**1**)
e. Make sure you work on the dataframe that you have created in the previous step after dropping missing values for rest of the questions till the end.
   Replace the values 0 and 1 in 'Survived' column with 'no' and 'yes' respectively. (**3**)
f. Create a bar chart to check how 'Age' differs by 'Gender'. Use yellow color for the chart. Make sure you have a proper title and axes labels for the chart.
   Which age bar is higher here, between male and female?(write this answer in words and comment out)? (**4+1**)
g. Create a bar chart to check how 'Age' differs by 'Survived'. Use green color for the chart. Make sure you have a proper title and axes labels for the chart.
   Can you say anything regarding the age of the passengers and their chance of survival at that time?(write this answer in words and comment out) (**4+2**)
h. Generate a new variable 'Family_members' which is the sum of SibSp and Parch. Add the variable as a new column to the dataframe. (**3**)
i. Drop the columns 'PassengerId', 'Cabin', 'SibSp', and 'Parch' from the dataframe. (**3**)
j. Print the top 10 oldest passengers' data from the dataframe (show all the columns in your output that are left after dropping 'PassengerId', 'Cabin', 'SibSp', and 'Parch' in the previous step). (**5**) Calculate the max of 'Family_members' for these top 10 oldest passengers. (**2**)