**R Midterm Test**

**Instructions**

1. ***Total Marks = 50 + 5***
2. ***Download Titanic (from Google link shared) + Iris + Diamonds (from Google link shared)***
3. ***Test Duration : 2hrs***

**1.** Find the details of the dead person who is half the age of the oldest person in the ship and whose ‘PassengerId’ is closest to that oldest person. (Hint: Remove the NA’s and then subset. Find the function for that). [Removing NA’s - 2 + Fn - 3]

**2.** Order the data frame on the following order (Apply all these conditions together). [2\*3]

* + Descending order of Pclass
  + Descending order or Fare
  + Ascending order or Age

**a.** What is the Ticket number of the first record now? [3]

**b. Optional:** In the ordered data frame, what is the position of the male person (whose age being even number) who pays the maximum price for 1st class ticket (Pclass == 1) - For brownie points. [5]

**3.** Extract the survived male passenger(s) of Pclass 3 embarked from “S”. Combine this with dead female passenger(s) of Pclass 1 embarked from “C”. What is the unique number of “Tickets” in this data frame? [3]

**4.** Conditional [4]

If the passenger’s name contains ‘Miss’, change the sex to ‘Girl’ else if the passenger’s name contains ‘Master’, change the sex to ‘Boy’, else keep it unchanged.

**5**. Factor Levels [3]

Change the C,Q,S levels of the Embarked column to C = Cherbourg; Q = Queenstown; S = Southampton such that Cherbourg < Queenstown < Southampton.

**6.** Instructions for question Q6 a & b:

* Copy the original data into another variable called “Data”. Use this dataset for Q6.
* Replace the Pclass column in this data set with "First", "Second" or "Third".

**a**. Melting and casting [3]

Transform the data using melting and casting so as to obtain the average fare per class, based on gender (Gender in rows and PClass in columns).

**b**. Using the same data set (“Data”) as above, transform the data so as to obtain the average age and fare of people who survived, based on class (Class in rows & age and fare in columns) [3]

**7.** Paste0 & next/break [6]

Subset the data for all the passengers who did not survive. Run a loop on this dataset to print an output as follows:

**"The passenger named <name of passenger>, who was born in <date>, did not survive "**

* Assume that the accident happened in 1912 and calculated their birth year.
* If the passenger's age is not present, skip that person.
* Prefix a Mr / Mrs. to the name of the passenger based on their gender.
* Print only for the first 50 males. After 50 males, stop the loop.

**8.** Use apply family only! Other solutions won’t carry marks.

* Use Iris data set

**a.** Create a column named all\_measurements. This column should be the sum of all the measurements (length, width for sepal & petal). [3]

**b**. Create a column which should have concatenation of all measurements separated by underscore (use paste + apply). [4]

**9**. read.csv

**a**. Read the diamond dataset. Store as Diamond\_All. [3]

**b**. Read the diamond dataset where price <3700 and certification is HRD. Store as Diamond\_Special. [4]