**Congestive Heart Failure Dataset Analysis Assignment** (30 points)

This week kicks off with a continuation of our work in analyzing clinical datasets, initiated as part of the introductory course in healthcare data science. This exercise will focus on reinforcing our existing knowledge, as well as advancing your understanding of the health data and operational skills in Excel. As in the case with your last dataset analysis assignment, you will be working with a dataset provided in Excel, and will need to show your work in Excel. There is a stronger emphasis on showing all your work now.

The best way to show your work is to perform each operation within a separate tab in Excel. It is easy to achieve: just download the dataset to your PC or Mac, open the file, and every time you start a new task – copy the first tab into a new tab and rename it appropriately to indicate which task number you are working on. This way, we avoid submissions of multiple files and contain all our work within a single Excel file. Sometimes you can combine a few tasks that do not conflict with each other for the same affected columns into a single tab, in which case your tab name should indicate task numbers covered by the corresponding tab. Should you run out of the limits of what Blackboard is willing to accept for a maximum file size, it is time to split some of the answers into a separate Excel file. The alternative to save space is to copy only those columns you need for each particular data manipulation when you initiate new tabs. Whichever method you choose, own your dataset: you are the analyst.

Your tasks are as follows:

1. Sort the spreadsheet by facility number to show how many facilities you are in the dataset. In a separate Word document, provide your straight answer as to how many facilities are present in the dataset. **2 points.**
2. You have columns for the patient admission and discharge dates, but we need to know the length of stay. Create a new column called LOS, for the length of stay, next to the DischargeDate column and use a formula to calculate LOS for each patient encounter. Your output measure will be number of days. There are about 11,000 rows in the file, making manual work impossible to even begin considering -☺ Time to reinforce your advanced Excel skills! **2 points.**
3. We continue filling in the blanks the for necessary basics in the dataset. Whoever pulled this dataset from the data warehouse, forgot to provide race codes. Race descriptors are not very useful for performing statistical operations, so we need to fill the missing values in. Your PatientRaceDesc column and the input table with pre-defined numerical race codes at the end of these assignment instructions will help guide you in the process. Again, the only way to accomplish this for some 11,000 patient encounters is automation. How, is up to you to determine. **2 points.**
4. In the Word document you initiated as part of answering Q1, copy the input table from race codes into it, and add a new column where you will specify the number of patient encounters for each of the race codes specified in the table. Additionally and still working within the same table in the Word document, break each entry for each race code in the table by gender, specifying how many male and female patients are within each category. **2 points.**
5. How many patients have died? Show your work in the Excel file and provide an answer in the Word document you are maintaining along the way. **2 points.**

Hint: review values in the DischargeDispositionDesc and DischargeDispositionCode columns in the dataset.

1. Show 5 doctors with the sickest patients. Provide an answer as a table in Word, with explanation. Show your work in Excel. **2 points.**
2. Is your ranking of five doctors with the sickest patients reliable? Why yes or why not? Answer in your developing Word document. **2 points.**
3. Which hospital facility had the most Medicaid/Public Aid patients? How many of such patients? Show your work in Excel and provide the answer within your Word document. **2 points.**
4. Which physician had the most Medicare patients? **2 points.**
5. Which hospital facility encountered the greatest number of readmissions? Is your answer reliable? Why yes or why not? Show your work in Excel and provide answers in Word. **2 points.**
6. Among these readmissions at a hospital facility that had the greatest number, what was the most common principal diagnosis and what was the most common source of admission? As usual, show your work and answer concisely in Word. **2 points.**
7. Which 5 physicians had the greatest number of readmissions, if all hospitals are equally considered (no divide between top hospital generating readmissions records and all other hospitals: the entire dataset is within the field of your consideration)? **2 points.**
8. Set up a pivot table to display every physician’s statistics on the number of readmitted patients, broken down by race description. So, for every physician in the spreadsheet, I am looking to determine how many readmitted patients he/she had of each race description. **2 points.**
9. Refine the pivot table you have just initiated under question (13) to add ability to filter by patient sex. **2 points.**
10. Refine the pivot table you have initiated under question (13) and enhanced under question (14) to help determine how many patients treated by each physician have died (still broken down by race code description). **2 points**.

Your deliverables for this assignment are:

1. Excel spreadsheet with work shown for all requirements of this assignment.
2. Word or PDF document explaining work and providing answers where necessary and as outlined in the assignment instructions.

**Input table for race codes:**

|  |  |
| --- | --- |
| **Race** | **Numerical Code** |
| African American / Black | 10 |
| American Indian / Alaska Native | 11 |
| Asian / Filipino | 12 |
| Caucasian / White | 13 |
| Declined | 14 |
| Eastern Indian | 15 |
| Hispanic | 16 |
| Middle Eastern | 17 |
| Other | 18 |
| Pacific Islander / Hawaiian | 19 |
| Unknown | 20 |