Ferrer, Ma. Angelika C.
Garcia, Lorenzo Miguel C.
Ogena, Graciel Zaireen C.

CMSC 176: EDA Write up

The given data set comprises encounter ID, patient number, gender, race, age, weight, admission type, discharge disposition, admission source, time-in, payer code, medication, specialty of attending doctor, readmission, etc. of diabetic patients from 130 US hospitals in the year 1999 to 2008. Some attributes of the data set like weight have so many missing values and thus were not included in the exploratory data analysis.

Patients were categorized by race, gender, and age group since most of the values from these categories were defined. There were 76 099 Caucasians, 19 210 African-American, 2273 unidentified races, 2037 Hispanic, 1506 other races, and 641 Asians. It was also shown (in Figure 2) that there were more female than male patients. According to age, there were 161 0 to 9 year olds, 691 10 to 19 year olds, 1657 20 to 29 year olds, 3775 31 to 39 year olds, 9685 40 to 49 year olds, 17 256 50 to 59 year olds, 22 483 60 to 69 year olds, 26 068 70 to 79 year olds, 17 197 80 to 89 year olds, and 2793 90 to 99 year olds. From the given numbers, most of the patients were Caucassians and a large part of the group are ages 70 to 79.

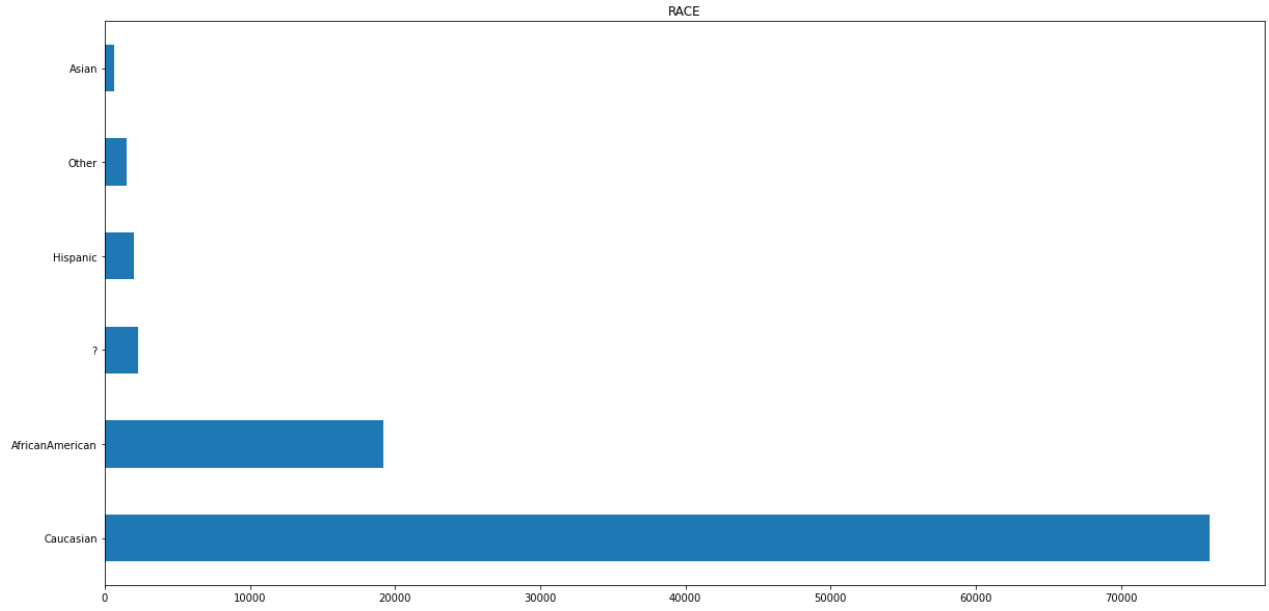

Figure 1. Patients according to race

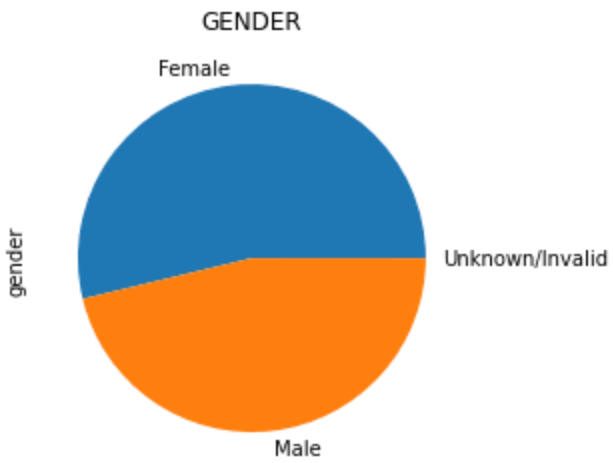

Figure 2. Patients according to gender

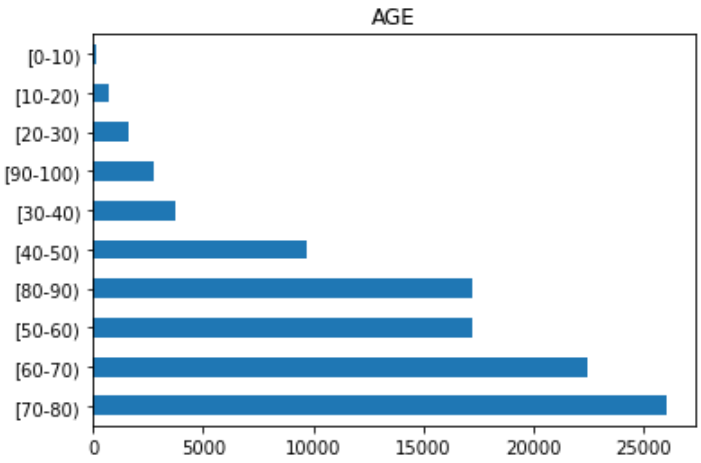

Figure 3. Patients according to age

Similarly, most patients fall in the age group 70 to 80 for both male and female patients. As we can see in Figure 4, the graph is skewed to the right; signifying that more older people retracts diabetes than younger ones

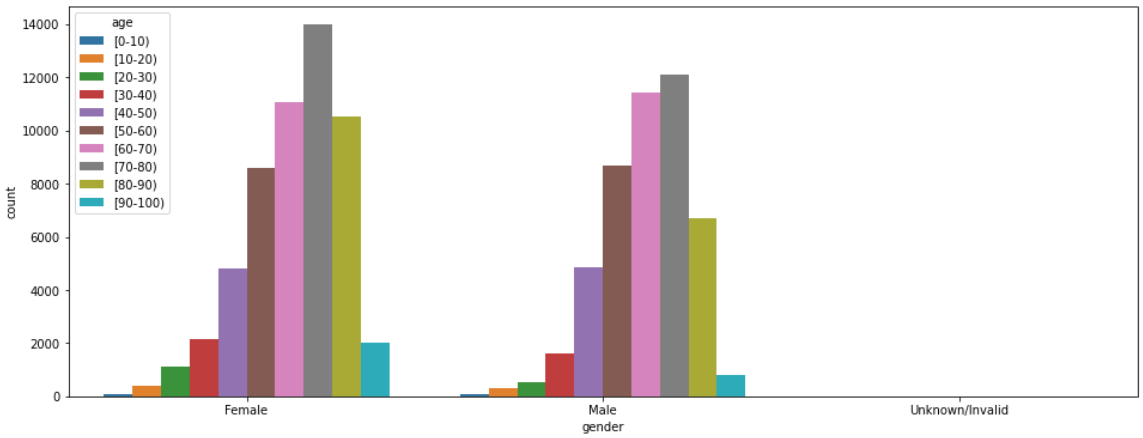

Figure 4. Patients according to age and gender

Figure 5 shows the distribution of gender among the races of the diabetic patients in the data. For Caucassians, there are 39 689 females and 36 410 males. For African-Americans, 11728 are female while 7482 are male. For the unidentified race, there are 1133 females, 1138 males, and 2 unknown genders. For the “Other” races, there are 748 females and 757 males, and 1 with unknown gender. There are also 318 female Asians and 323 male Asians in the data set. Lastly, there are 1092 female Hispanic patients and 945 male Hispanic patients. In comparison, there are more female patients that are Caucasians, African-Americans, and Hispanics while there are more male patients that belong to the Asian, Other, and unidentified race.

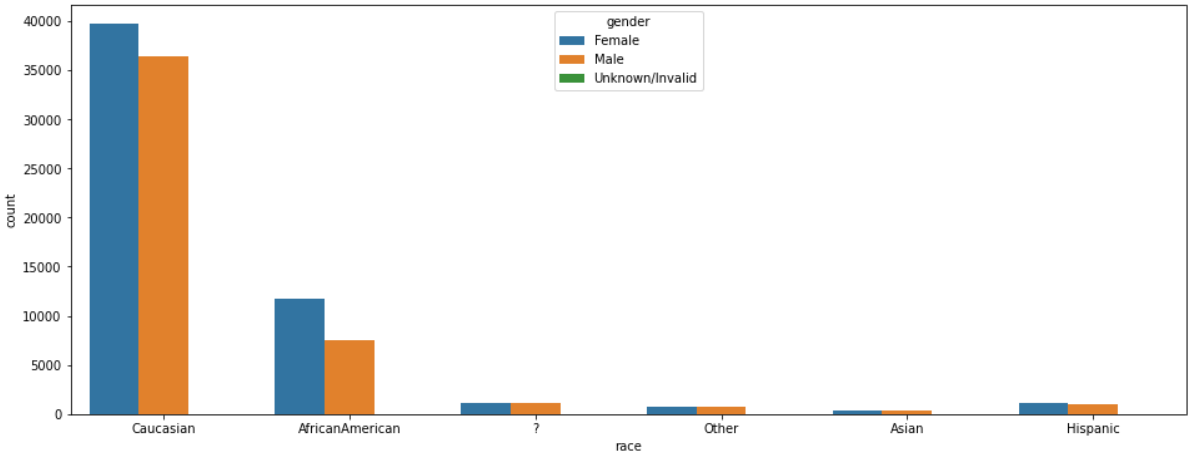

Figure 5. Patients according to race and gender

Attending doctors can also be identified by their medical specialty. Figure 5 below shows the number of doctors based on their field of specialty, with Internal Medicine garnering the highest count of 14 635.

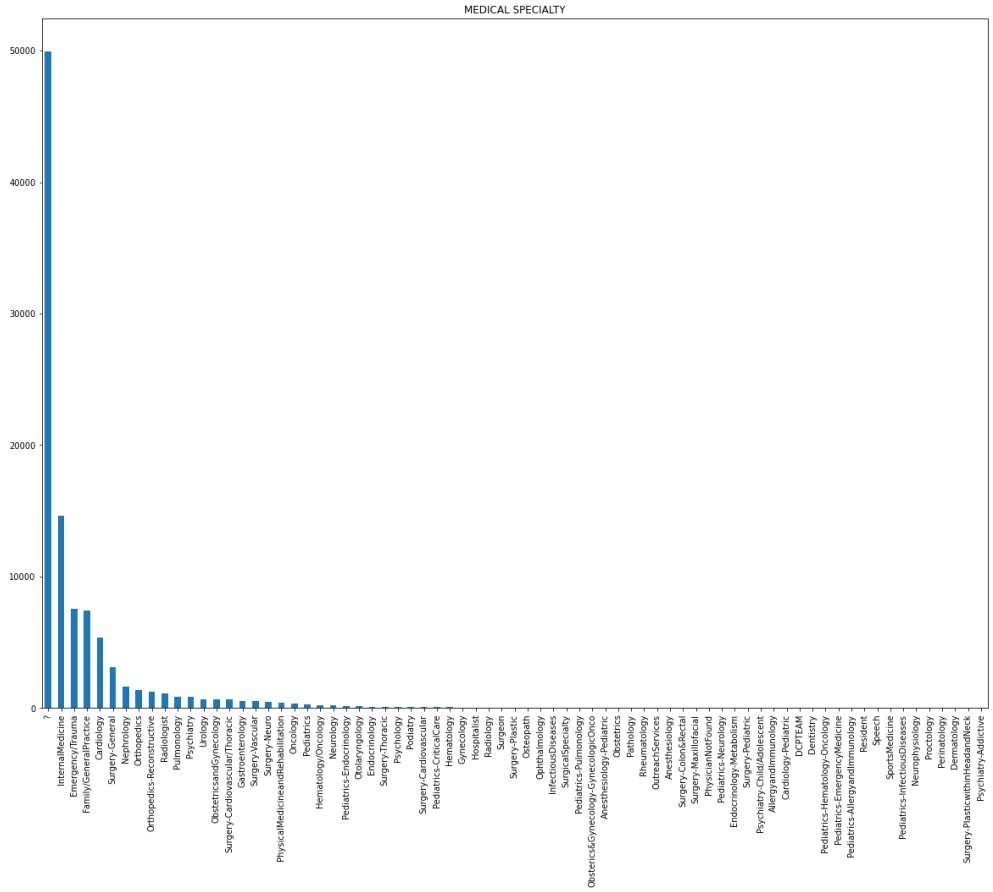

Figure 6. Doctors according to medical specialty

Most patients were admitted via the Emergency Unit. From 1999 to 2008, there were 53990 diabetic patients from Emergency. On the other hand, most patients were discharged to their homes with a total of 60 234 patients.

Based on the count of values, the patients underwent an average of 43 laboratory procedures. It has a standard deviation of 19.67. These patients also took an average of 16 medications, with a minimum of 1 and a maximum of 81. Most medications were not taken by the patients but there were some which was indicated as “steady” like glyburide-metformin, insulin, acarbose, rosiglitazone, pioglitazone, glyburide, glipizide, glimepiride, nateglinide, repaglinide, and metformin. Insulin also went “down” for some of the patients in the data set.

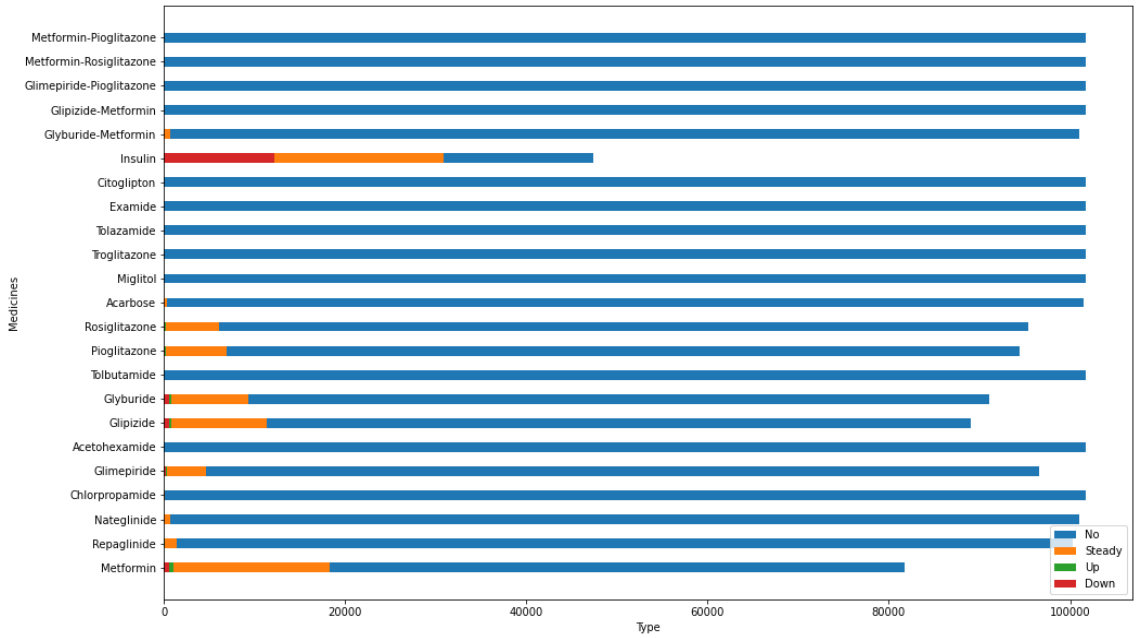

Figure 7. Medications

The repaglinide drug stood out among other medications because it was given in different amounts to the patients. According to the data, 100 227 did not receive this drug, 1384 received a steady dose of this drug, 110 patients' dose of this drug went up, and 45 patients' dose of this drug went down.

Number of medications per age group also varies. For ages 0 to 9 and 10 to 19, most patients take 5 medications. For ages 20 to 29, most patients take 7 medications. For ages 30 to 29, most patients take 10 medications. For ages 40 to 49 and 80 to 89, most patients take 11 medications. For ages 90 to 99, most patients take 12 medications. For ages 50 to 59 and 70 to 79, most patients take 13 medications. For ages 60 to 69, most patients take 15 medications.

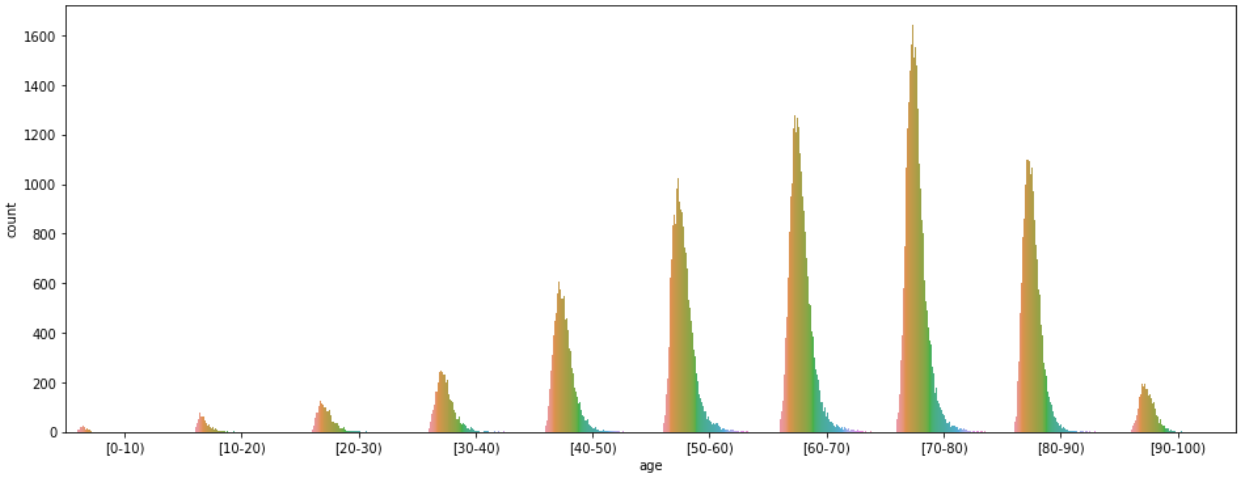

Figure 8. Number of medication per age group

Number of lab procedures undergone by patients per age group differ from one another as well. Figure 8 shows how many patients undergo a certain number of lab procedures. Most patients ages 0 to 9 took 47 lab procedures. Most patients ages 10 to 19 patients took 1 lab procedure followed closely by 50 lab procedures. Likewise, most patients ages 20 to 69 and 80 to 89 undergo only 1 lab procedure. On the other hand, most patients ages 70 to 79 undergo 43 lab procedures while most patients ages 90 to 99 undergo 62 lab procedures.

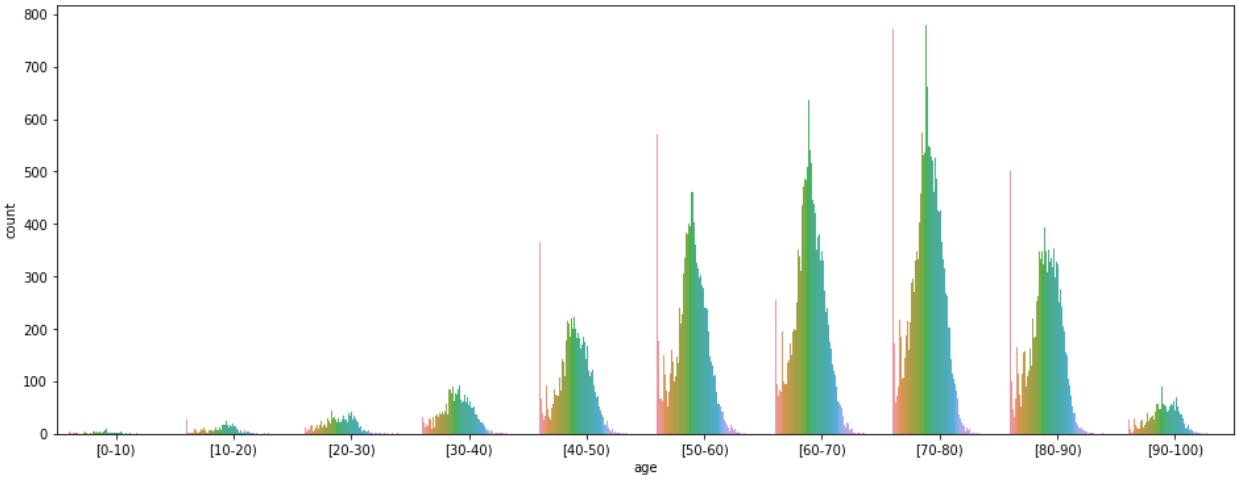

Figure 9. Number of lab procedures per age group