DIABETES DATA SET

FERRER, Ma. Angelika C.
GARCIA, Lorenzo Miguel C.
OGENA, Graciel Zaireen C.

Diabetes Data Set

› 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 <u>130 US hospitals in the year 1999 to 2008</u>

› Some attributes of the data set like weight have so many missing values and thus were not included in the exploratory data analysis
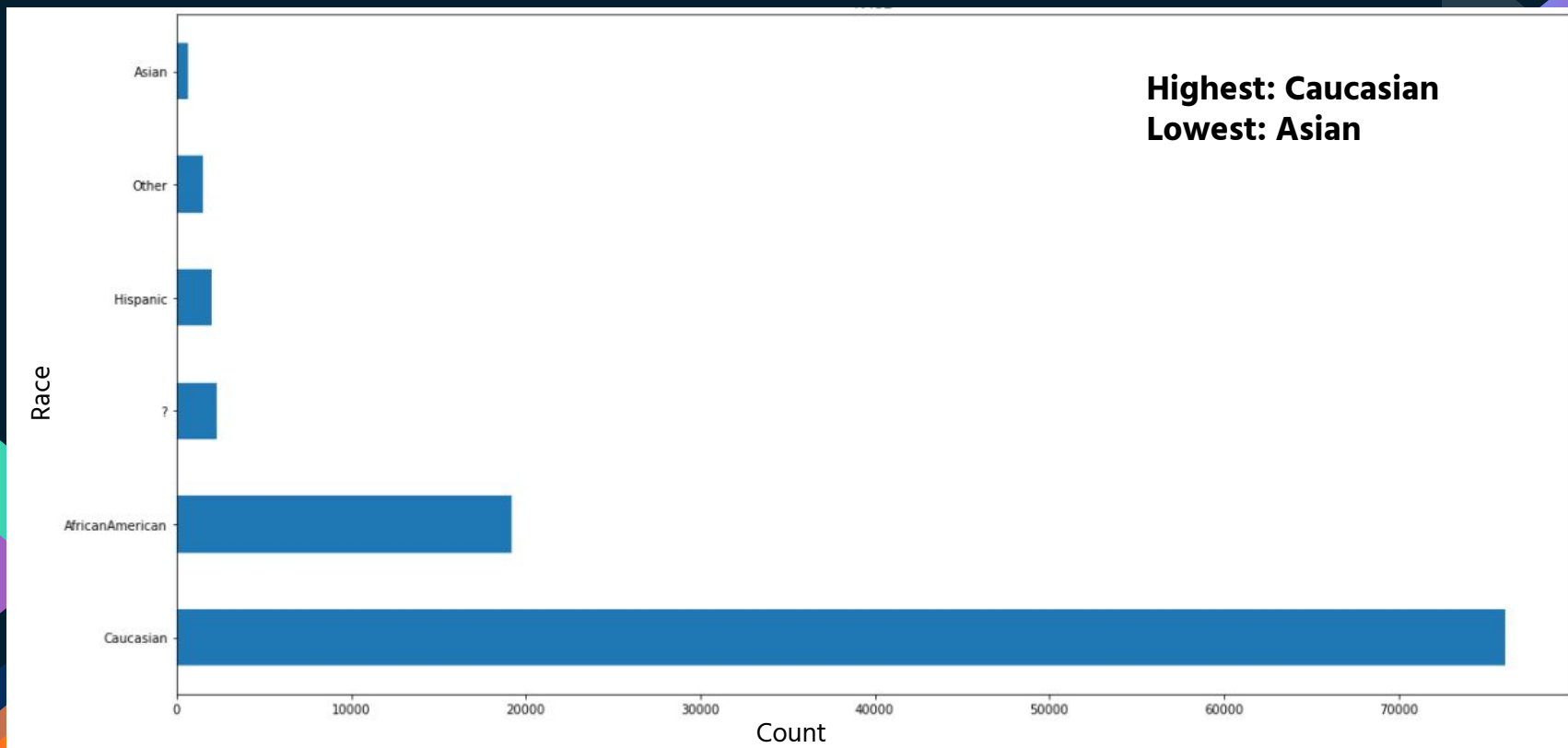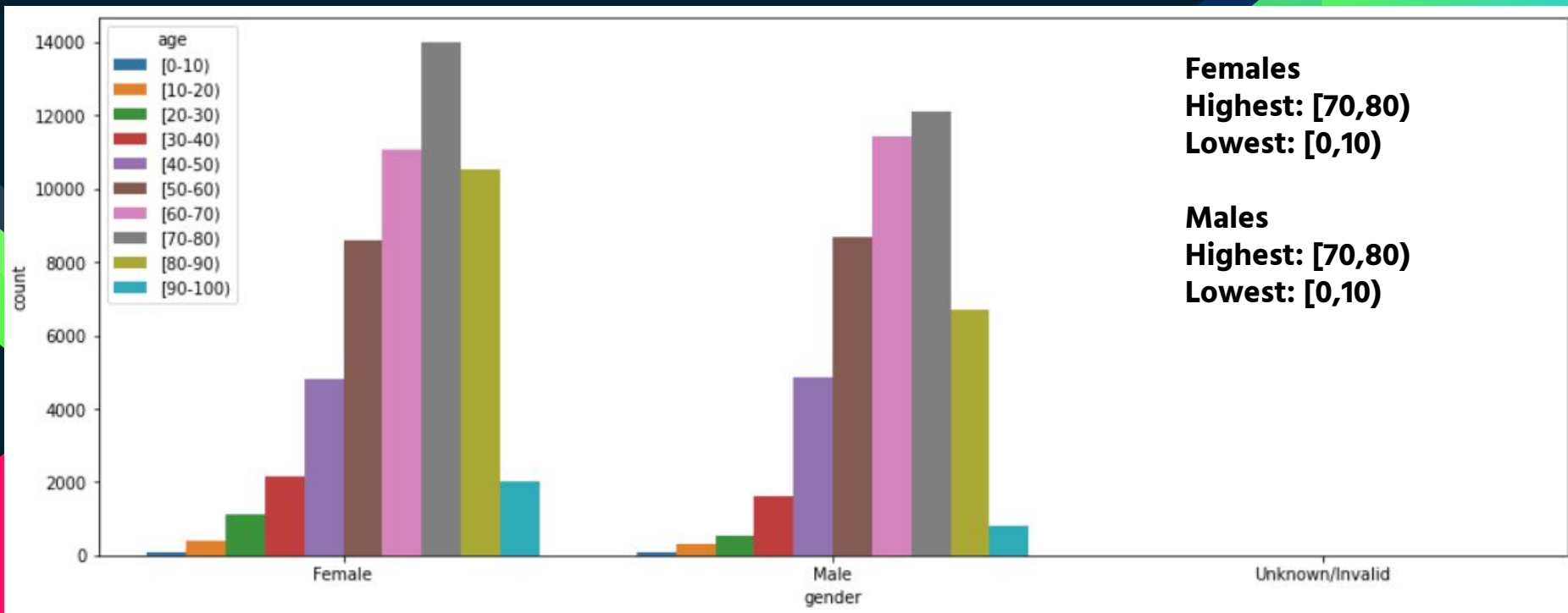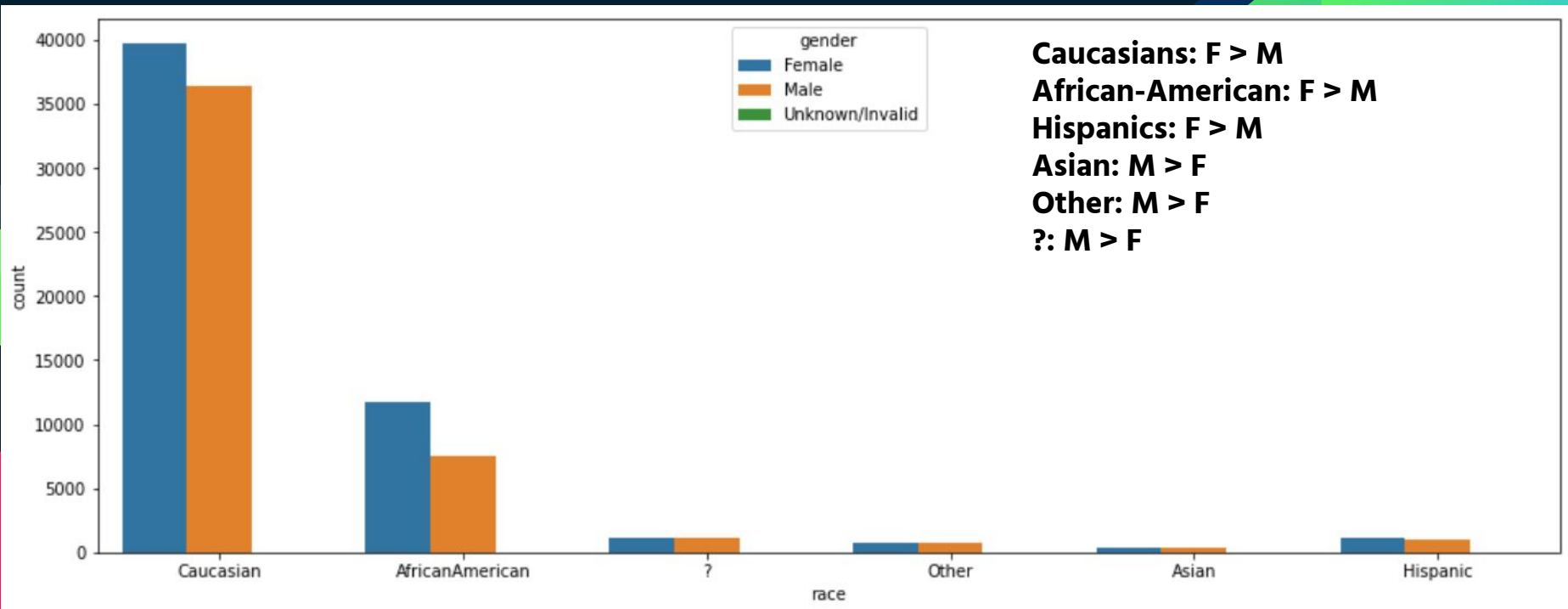
Patients according to race
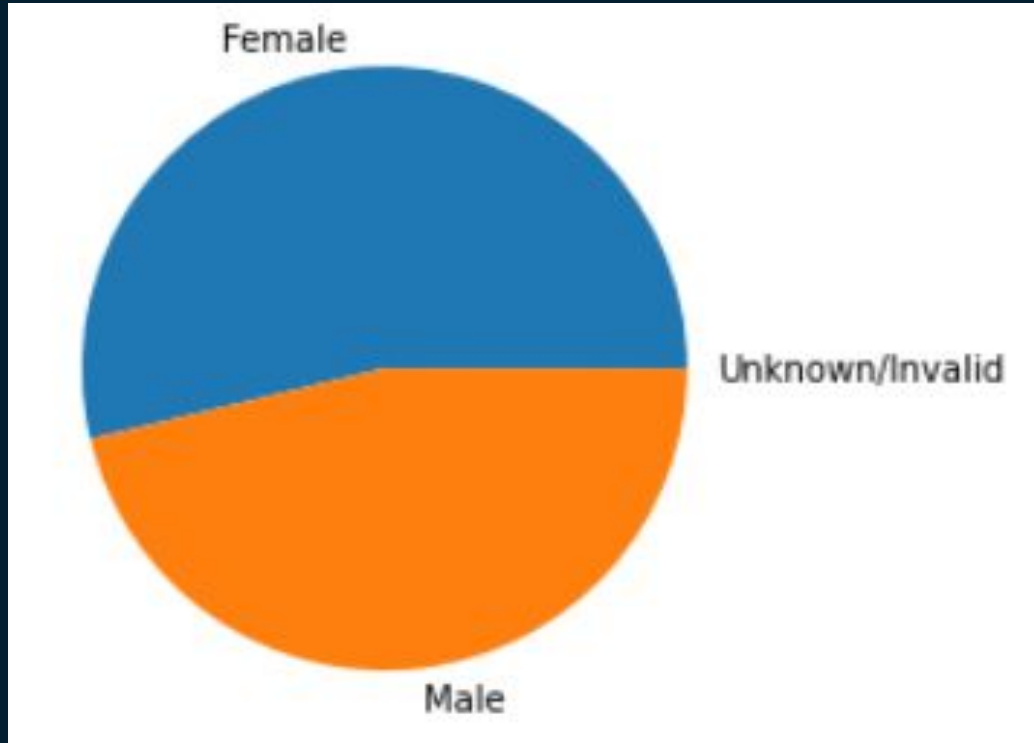
Patients according to gender
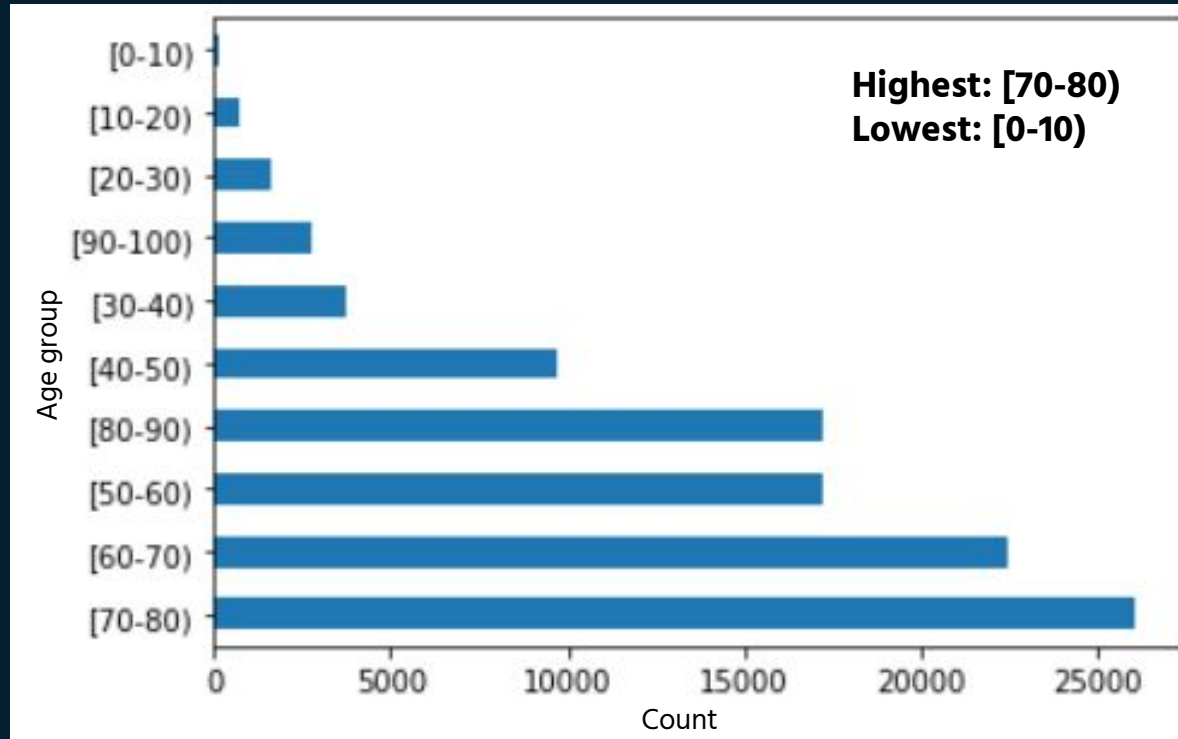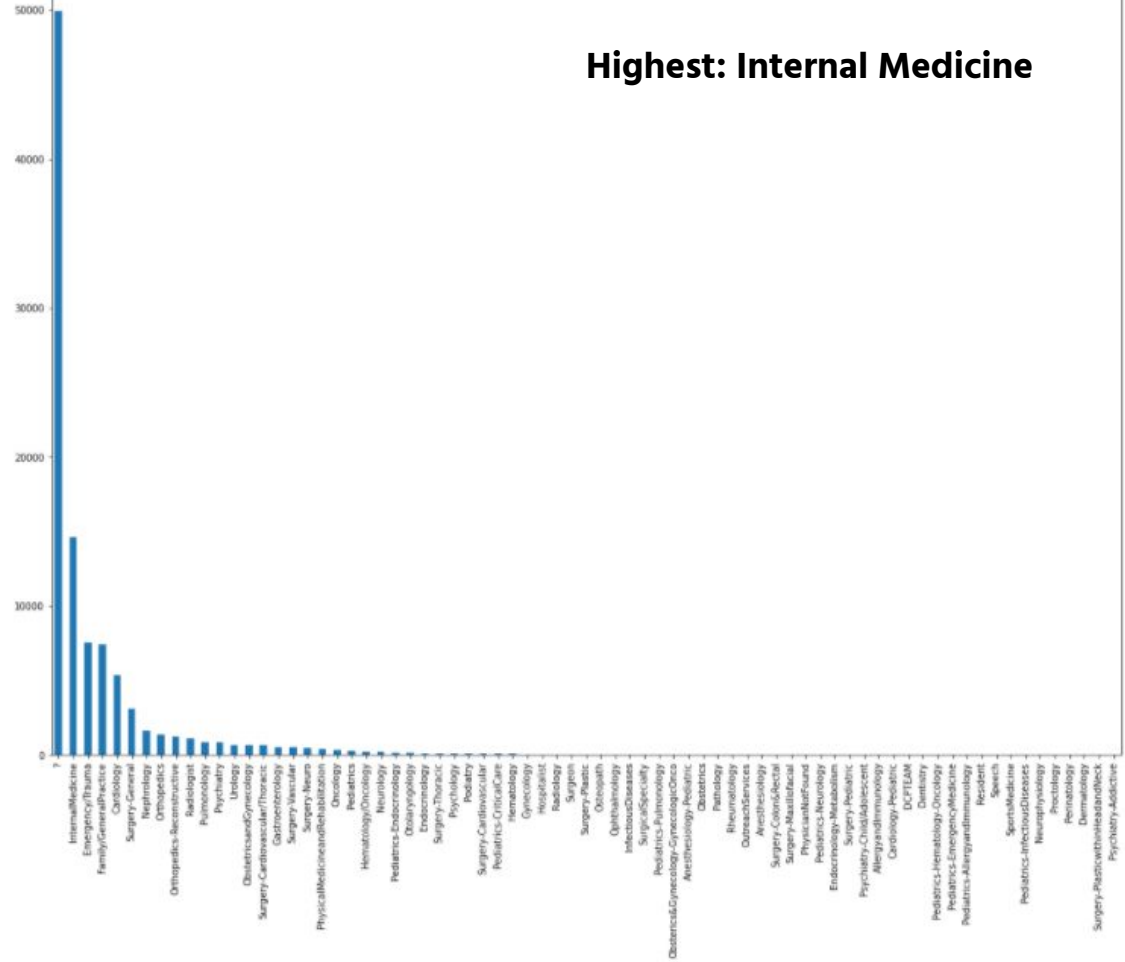
Patients according to age

Patients according to age and gender

Patients according to race and gender

Doctors according to medical specialty

	num_lab_procedures	num_procedures	num_medications
count	101766.000000	101766.000000	101766.000000
mean	43.095641	1.339730	16.021844
std	19.674362	1.705807	8.127566
min	1.000000	0.000000	1.000000
25%	31.000000	0.000000	10.000000
50%	44.000000	1.000000	15.000000
75%	57.000000	2.000000	20.000000
max	132.000000	6.000000	81.000000

admission_type_id

1	53990
3	18869
2	18480
6	5291
5	4785
8	320
7	21
4	10

1	Emergency
2	Urgent
3	Elective
4	Newborn
5	Not Available
6	NULL
7	Trauma Center
8	Trauma Center

discharge_disposition_id

1	60234
3	13954
6	12902
18	3691
2	2128
22	1993
11	1642
5	1184
25	989
4	815
7	623
23	412
13	399

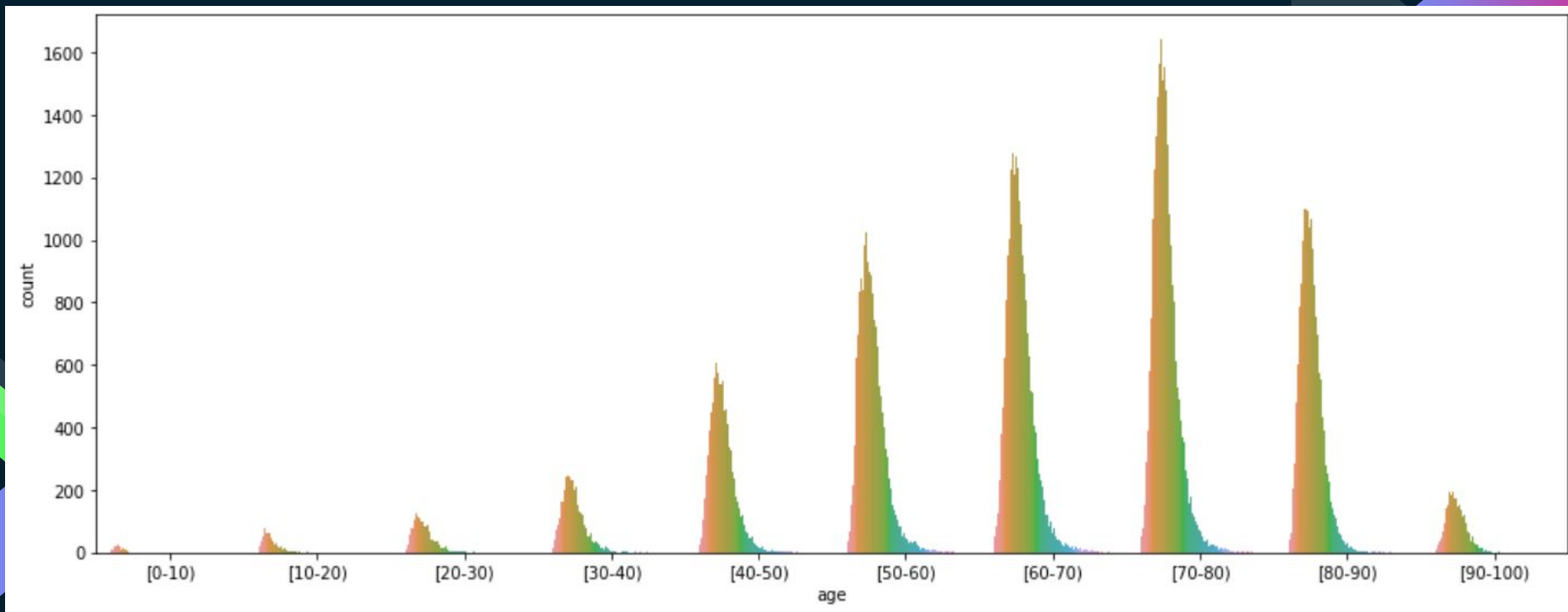|	
14	372
28	139
8	108
15	63
24	48
9	21
17	14
16	11
19	8
10	6
27	5
12	3
20	2

Highest: ID 1 = Discharged to home

Lowest: ID 20 = Expired in a medical facility. Medicaid only, hospice.
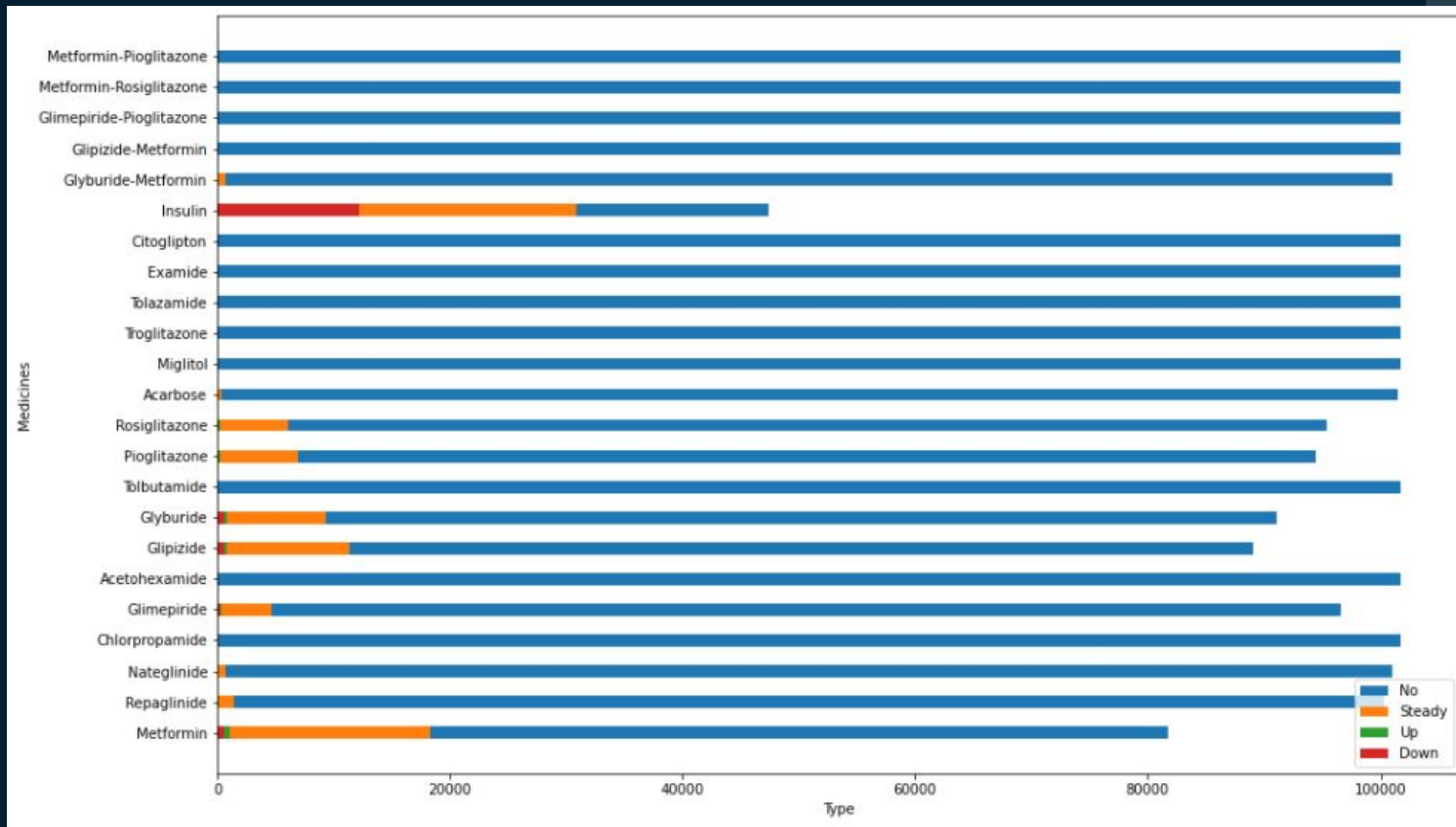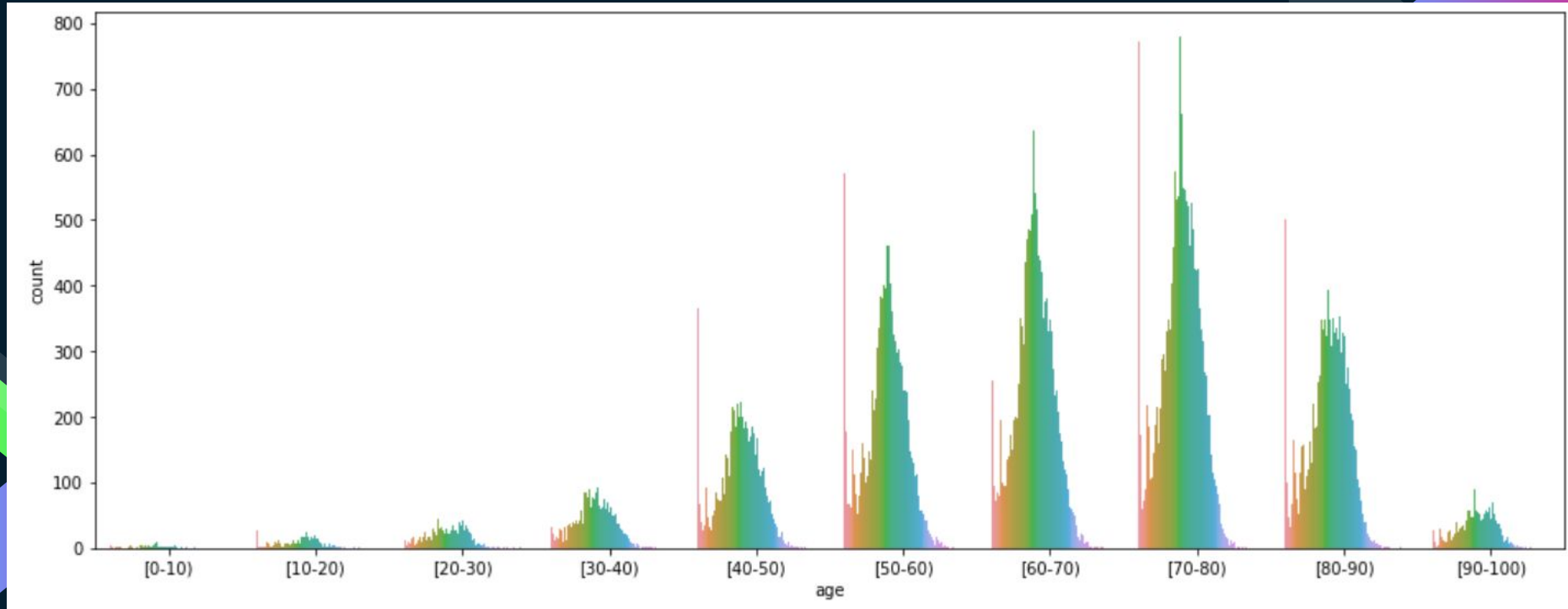
Medications

```
No          100227
Steady        1384
Up             110
Down            45
Name: repaglinide, dtype: int64
```

Number of medication per age group

Number of lab procedures per age group