Ugurcan Sencan

Case – Exploring Electronic Health Records

- **Load the data**

I loaded all data files one by one using directory path and file extension. I preferred reading and assigning them to different variables (data frames). You can see one of them below.

```
patients = pd.read_csv(os.path.join(directory_path, os.listdir(directory_path)[10]))
```

- **Explore the data and make a visualization of a single patient trajectory as she transitions through the medical care system over time**

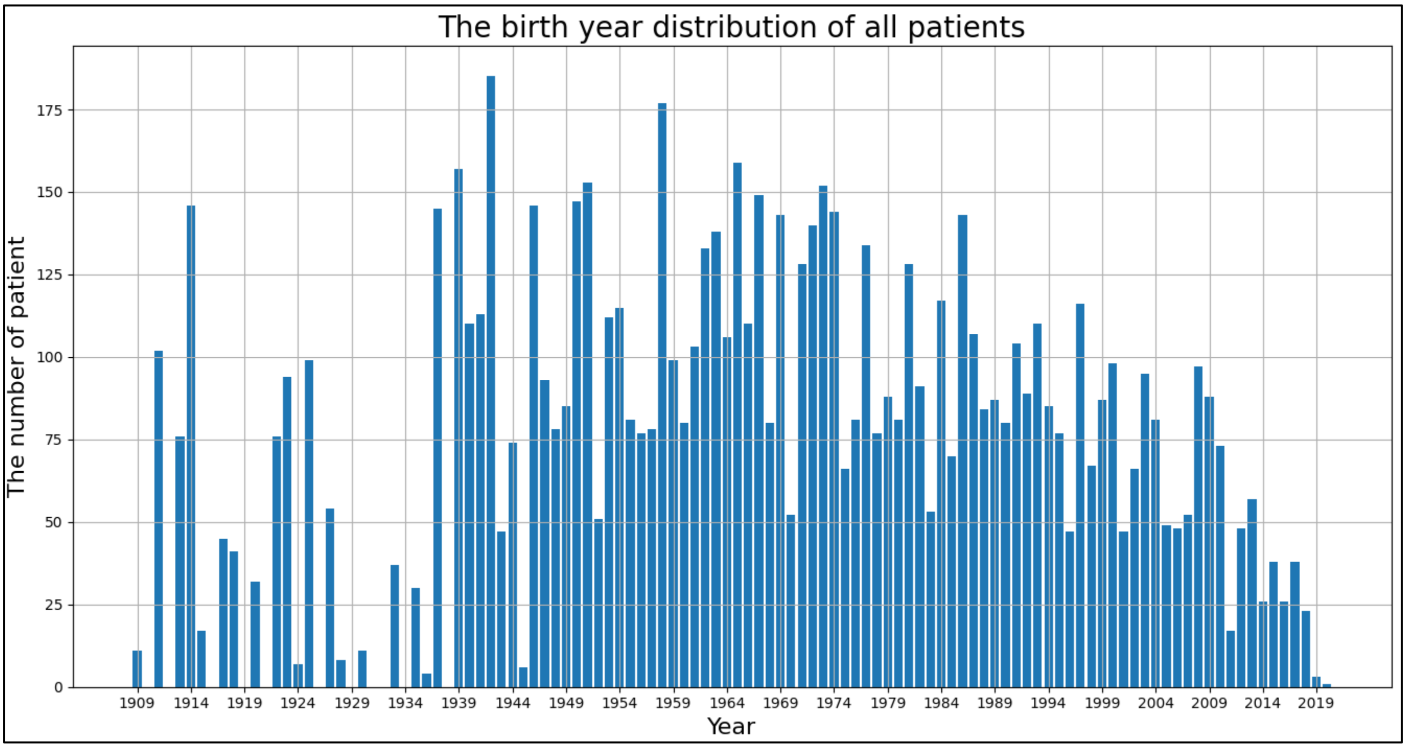


Figure 1. A graph of the birth year distribution of all patients in the data.

The figure 1 shows the birth year distribution of all patients. The youngest patient is **3** years old and the oldest one was **114**. The average of this population is **55.673.** In this data set, there are **8376** patient's detailed information. **1658** of them are recorded as died.

Also, all patients are from different cities, races, ethnicities. The total number of cities where all patients live is **235**. With these data, there can be made many **descriptive analyses** to explore insights.

! The questions that I come up while working on this study are below. However, because of my time constraints in this week, I have only the limited analyses shown in the figures. For example,

- in which cities what medical conditions are encountered the most?
- in which ethnicities/race what conditions are encountered the most?
- in married or single patients what medical conditions are encountered the most? Any specific things?
- which medical conditions triggered each other in a single patient over time?

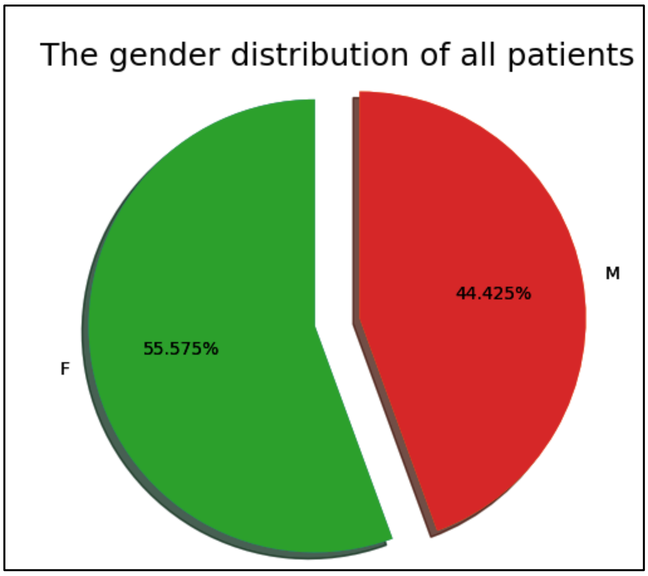


Figure 2. A graph of the birth year distribution of all patients in the data.

The figure 2 shows the gender distribution of all patients. The data consists of only two genders. Nearly %55 of patients are recorded as male, M and the remain is recorded as female, F.

I selected a patient who has the maximum number of conditions (**21**) in the dataset and whose id is below.

6ec18ddf-e9ee-421a-9033-456f558c7b4b

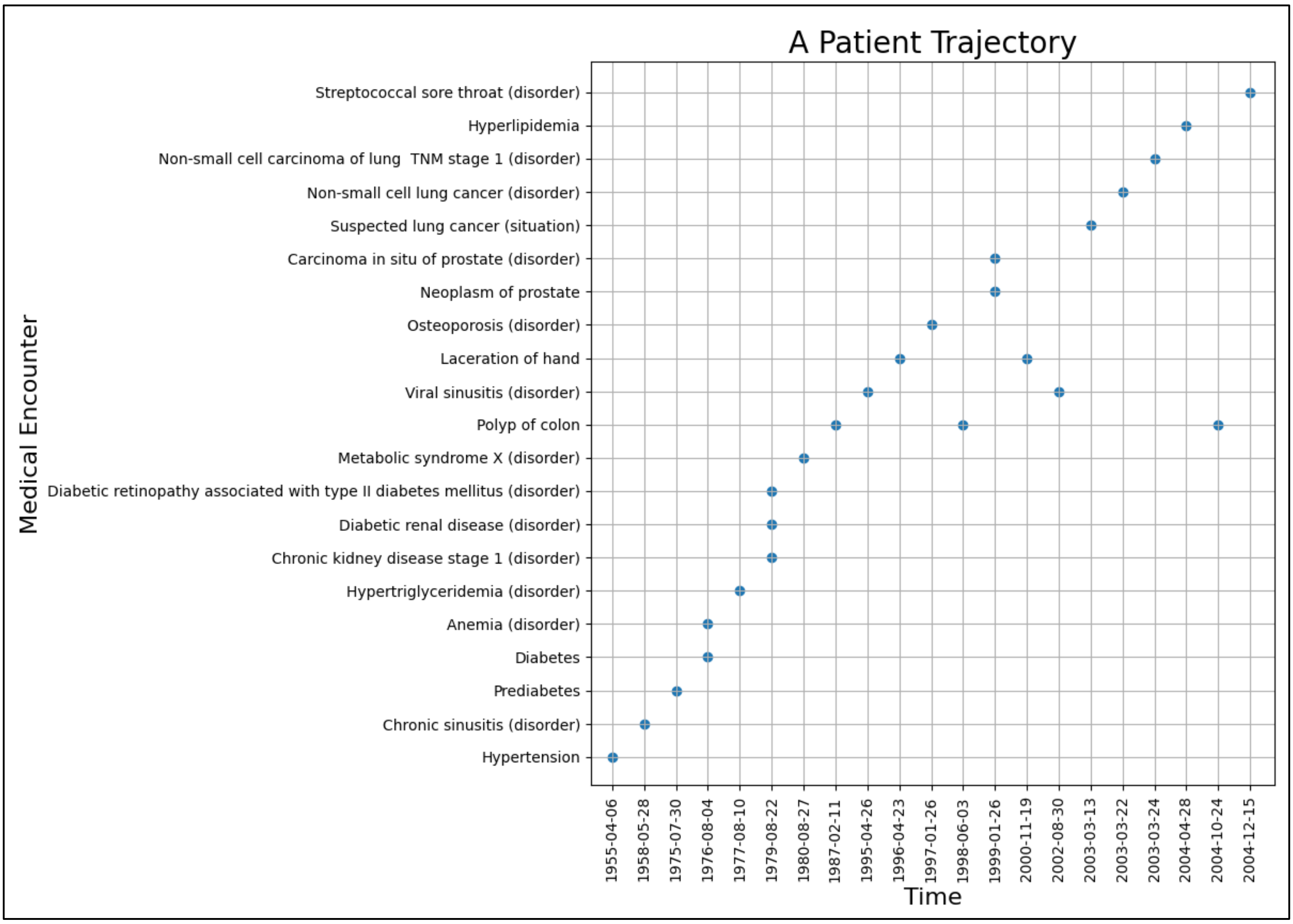


Figure 3. A graph of the single patient trajectory over time who has the maximum number of conditions in the data.

This figure shows a patient's all conditions over time. The patient has 21 conditions and 25 encounters in the dataset with various conditions as you can see from the y-axis in the Figure 3. It also shows that the patient has the same conditions more than one over time. Also, the patient has three different conditions on August 22nd, 1979. This patient was born on February 10th, 1937, and passed away on March 23rd, 2005.

- **Explore the data to find and present patterns of patients with the same conditions such as:**
  - **Which are the three most common conditions (present graphs and numbers)?**

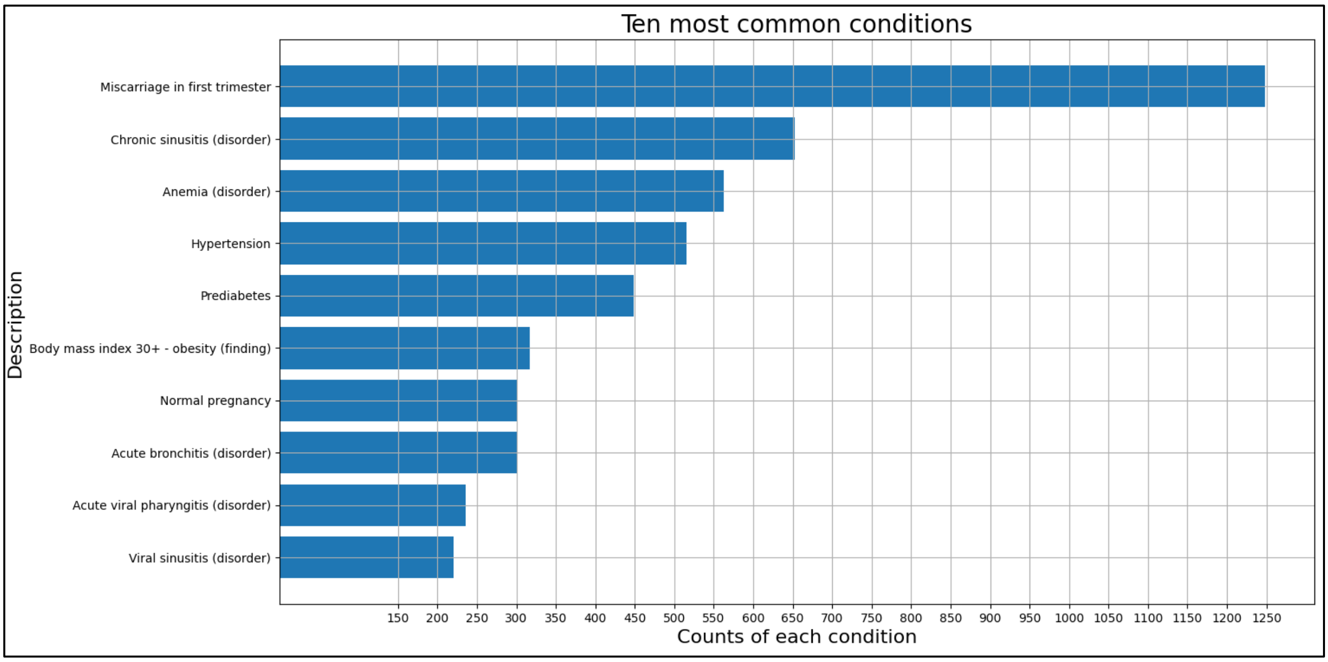
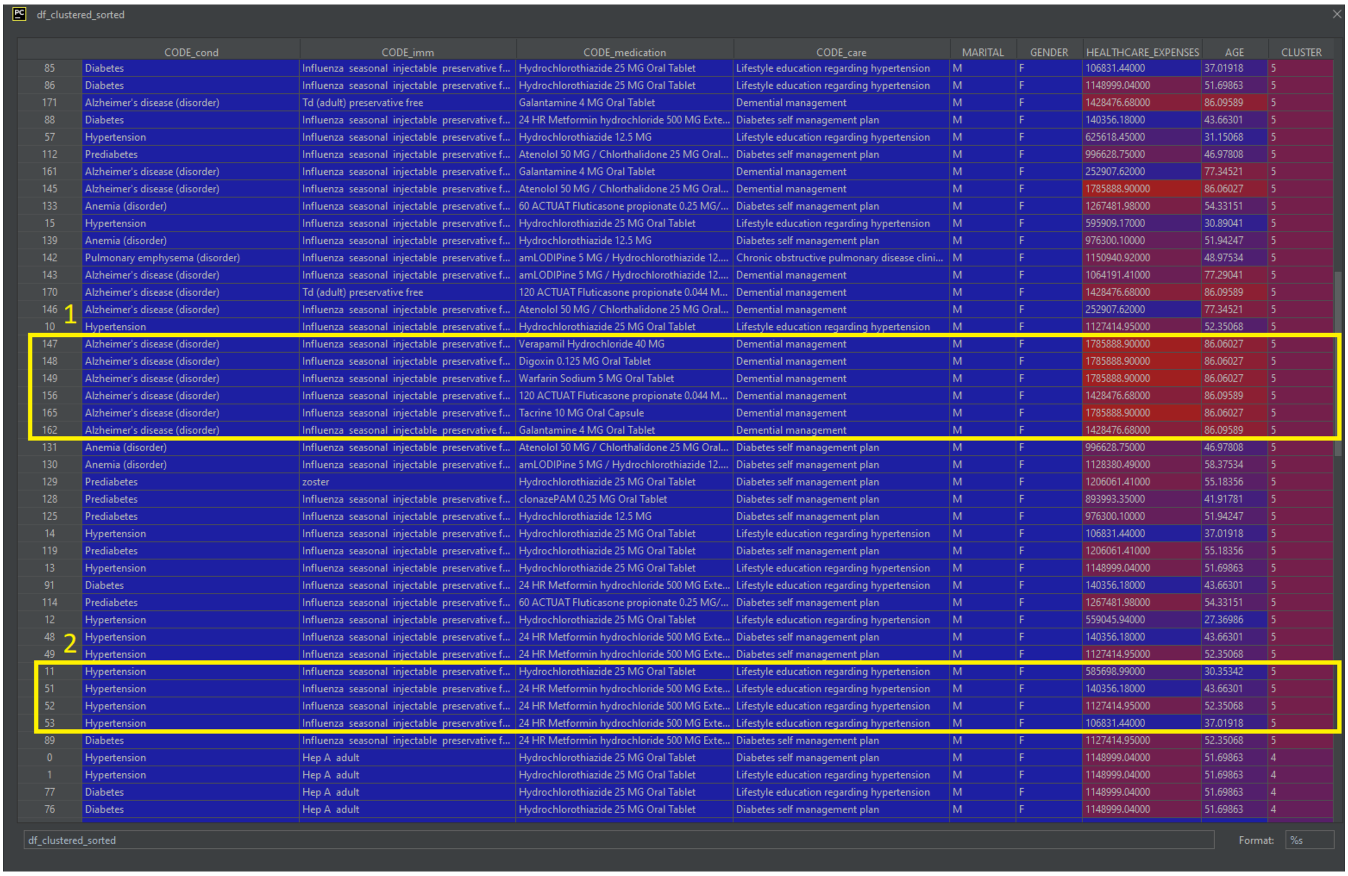


Figure 4. A graph of the ten most conditions in the data.

The figure 4 shows the ten most conditions. It shows that **miscarriage in first trimester** is the most encountered condition and almost 1250 people encountered with this condition.

- **Are there similarities in how the three conditions are treated? Showcase examples**.



df_clustered_sorted

	CODE_cond	CODE_imm	CODE_medication	CODE_care	MARITAL	GENDER	HEALTHCARE_EXPENSES	AGE	CLUSTER
85	Diabetes	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	106831.44000	37.01918	5
86	Diabetes	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	1148999.04000	51.69863	5
171	Alzheimer's disease (disorder)	Td (adult) preservative free	Galantamine 4 MG Oral Tablet	Demential management	M	F	1428476.68000	86.09589	5
88	Diabetes	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Diabetes self management plan	M	F	140356.18000	43.66301	5
57	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 12.5 MG	Lifestyle education regarding hypertension	M	F	625618.45000	31.15068	5
112	Prediabetes	Influenza seasonal injectable preservative f...	Atenolol 50 MG / Chlorthalidone 25 MG Oral...	Diabetes self management plan	M	F	996628.75000	46.97808	5
161	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Galantamine 4 MG Oral Tablet	Demential management	M	F	252907.62000	77.34521	5
145	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Atenolol 50 MG / Chlorthalidone 25 MG Oral...	Demential management	M	F	1785888.90000	86.06027	5
133	Anemia (disorder)	Influenza seasonal injectable preservative f...	60 ACTUAT Fluticasone propionate 0.25 MG/...	Diabetes self management plan	M	F	1267481.98000	54.33151	5
15	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	595909.17000	30.89041	5
139	Anemia (disorder)	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 12.5 MG	Diabetes self management plan	M	F	976300.10000	51.94247	5
142	Pulmonary emphysema (disorder)	Influenza seasonal injectable preservative f...	amLODIPine 5 MG / Hydrochlorothiazide 12....	Chronic obstructive pulmonary disease clini...	M	F	1150940.92000	48.97534	5
143	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	amLODIPine 5 MG / Hydrochlorothiazide 12....	Demential management	M	F	1064191.41000	77.29041	5
170	Alzheimer's disease (disorder)	Td (adult) preservative free	120 ACTUAT Fluticasone propionate 0.044 M...	Demential management	M	F	1428476.68000	86.09589	5
146	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Atenolol 50 MG / Chlorthalidone 25 MG Oral...	Demential management	M	F	252907.62000	77.34521	5
10	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	1127414.95000	52.35068	5
147	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Verapamil Hydrochloride 40 MG	Demential management	M	F	1785888.90000	86.06027	5
148	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Digoxin 0.125 MG Oral Tablet	Demential management	M	F	1785888.90000	86.06027	5
149	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Warfarin Sodium 5 MG Oral Tablet	Demential management	M	F	1785888.90000	86.06027	5
156	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	120 ACTUAT Fluticasone propionate 0.044 M...	Demential management	M	F	1428476.68000	86.09589	5
165	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Tacrine 10 MG Oral Capsule	Demential management	M	F	1785888.90000	86.06027	5
162	Alzheimer's disease (disorder)	Influenza seasonal injectable preservative f...	Galantamine 4 MG Oral Tablet	Demential management	M	F	1428476.68000	86.09589	5
131	Anemia (disorder)	Influenza seasonal injectable preservative f...	Atenolol 50 MG / Chlorthalidone 25 MG Oral...	Diabetes self management plan	M	F	996628.75000	46.97808	5
130	Anemia (disorder)	Influenza seasonal injectable preservative f...	amLODIPine 5 MG / Hydrochlorothiazide 12....	Diabetes self management plan	M	F	1128380.49000	58.37534	5
129	Prediabetes	zoster	Hydrochlorothiazide 25 MG Oral Tablet	Diabetes self management plan	M	F	1206061.41000	55.18356	5
128	Prediabetes	Influenza seasonal injectable preservative f...	clonazePAM 0.25 MG Oral Tablet	Diabetes self management plan	M	F	893993.35000	41.91781	5
125	Prediabetes	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 12.5 MG	Diabetes self management plan	M	F	976300.10000	51.94247	5
14	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	106831.44000	37.01918	5
119	Prediabetes	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Diabetes self management plan	M	F	1206061.41000	55.18356	5
13	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	1148999.04000	51.69863	5
91	Diabetes	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Lifestyle education regarding hypertension	M	F	140356.18000	43.66301	5
114	Prediabetes	Influenza seasonal injectable preservative f...	60 ACTUAT Fluticasone propionate 0.25 MG/...	Diabetes self management plan	M	F	1267481.98000	54.33151	5
12	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	559045.94000	27.36986	5
48	Hypertension	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Diabetes self management plan	M	F	140356.18000	43.66301	5
49	Hypertension	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Diabetes self management plan	M	F	1127414.95000	52.35068	5
11	Hypertension	Influenza seasonal injectable preservative f...	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	585698.99000	30.35342	5
51	Hypertension	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Lifestyle education regarding hypertension	M	F	140356.18000	43.66301	5
52	Hypertension	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Lifestyle education regarding hypertension	M	F	1127414.95000	52.35068	5
53	Hypertension	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Lifestyle education regarding hypertension	M	F	106831.44000	37.01918	5
89	Diabetes	Influenza seasonal injectable preservative f...	24 HR Metformin hydrochloride 500 MG Exte...	Diabetes self management plan	M	F	1127414.95000	52.35068	5
0	Hypertension	Hep A adult	Hydrochlorothiazide 25 MG Oral Tablet	Diabetes self management plan	M	F	1148999.04000	51.69863	4
1	Hypertension	Hep A adult	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	1148999.04000	51.69863	4
77	Diabetes	Hep A adult	Hydrochlorothiazide 25 MG Oral Tablet	Lifestyle education regarding hypertension	M	F	1148999.04000	51.69863	4
76	Diabetes	Hep A adult	Hydrochlorothiazide 25 MG Oral Tablet	Diabetes self management plan	M	F	1148999.04000	51.69863	4

df_clustered_sorted Format: %s

Figure 5. A graph of the ten most conditions in the data.

After using DBSCAN algorithm to make clusters, I chose two examples to show the similarities between in how conditions are treated in Figure 5. When I look at the first example, there are 6 rows belonging to two patients. We can check it from the AGE column. For the same conditions, **the same** care plan, **Demential management,** is made but the medications used for the treatment are different. In another example, you can see 4 different patient's information. All of them has the same conditions, **Hypertension**. Although all of them are at different age, **the same** care plan, **Lifestyle education regarding hypertension,** is made. Additionally, different medications may be given to the younger patients.

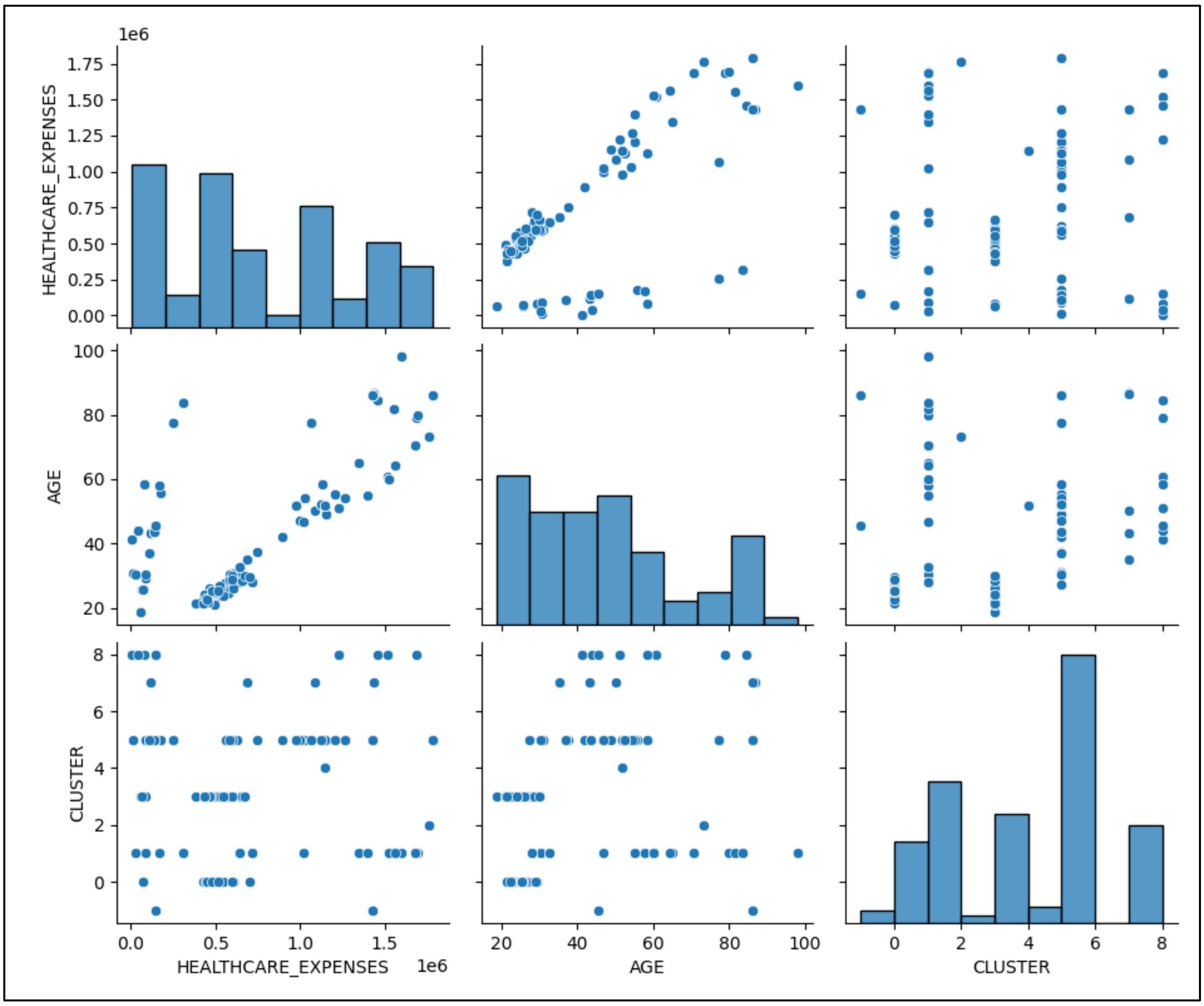


Figure 6. A graph of the ten most conditions in the data.

As an additional figure, Figure 6 shows different charts showing the relationship between different variables. For example, when we look at the healthcare expenses vs age chart, the scatter plot shows that some of patients have less expenses than most of them. Many young patients have the similar expenses of health cares whereas most of older patients have different range of expenses in health care. The older patients familiar with paying more compared to the younger patients. Another plot shows the distribution of clusters in health care expenses.

- **What other common pattern characteristics can be found for the three groups of conditions?**

Other common pattern characteristics include demographic information, medicine prescriptions, and test findings.

• **Formulate three other questions that could be interesting from a machine learning perspective using this data (that could potentially be used in a clinical setting to improve care)**

1. **Can we forecast which individuals are at greater risk of problems following surgery?**

A machine learning model might be developed to predict which patients are more likely to have problems, such as infections or readmissions, using data on patient demographics, medical history, and surgical procedures. This might aid physicians in implementing preventative steps and identifying at-risk individuals to lower the possibility of problems.

2. **Can we create a model that predicts which patients are likely to skip their follow-up appointments?**

Missing follow-up appointments may delay diagnosis and treatment and this results in worse patient outcomes. Aiming to forecast which patients are most likely to skip their visits, a machine learning model might be built after collecting data on prior appointment histories, patient demographics, and medical conditions. This information can be used by clinicians to take proactive efforts in order to ensure that at-risk patients receive the care they need.

3. **Can we identify patients who will benefit from a specific treatment or intervention?**

In order to identify which patients are most likely to benefit from a specific treatment or intervention, a machine learning model might be built by evaluating EHR data on patient demographics, medical history, and treatment outcomes. This could help clinicians in tailoring treatment approaches to specific patients and it may increase the likelihood of effective outcomes while decreasing the risk of adverse events.