# CLAIMS DATA REPORT(NH)

By

Nisha Muthukumaran

Data cleaning:

We began the data analysis by first by importing the 17Million rows into Python. We then sampled 1M unique rows containing unique claim IDs. Next, we merged the one million rows with the original data set of 17Million. We ended up with approximately 2.5Million rows of data. The dataset contains 63 columns of which we require only a few 26 columns. Further, we removed the rows that contained the AMT_PAID and AMT_BILLED as negative values. We also made new columns from the existing data where we grouped the ages into categories of four and we categorized people as instate or outstate residents depending on the whether they belong to New Hampshire or not.

1) Analysis of demographics:

a. Perform a thorough analysis of differences in total costs of healthcare by gender, age groups, county of residence, and in-state vs out-of-state resident patients.

Ans:
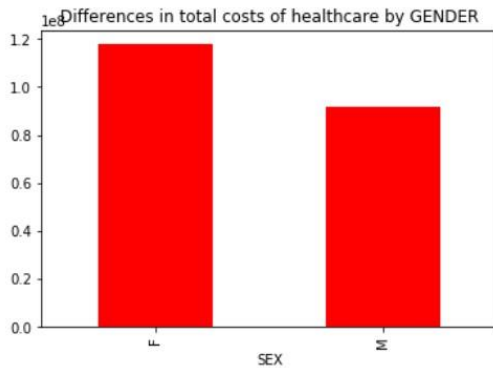
Total cost of healthcare by Gender when plotted shows that majority of the amount paid is maximum in Females is almost 0.12 Billion dollars when compared to males which almost sums up to 0.1Billion dollars. The difference in the expenditure for males and females is almost a solid 0.04 Billion dollars.

Total cost of healthcare when compared between age groups shows that majority was spent on middle aged people who were between the age group of 35 and 65, followed by old and youth which show that the same amount of money was spent on these two. This also shows that teens didn't spend much on healthcare probably because they fall in the "healthier" age group and visited the hospital less frequently.
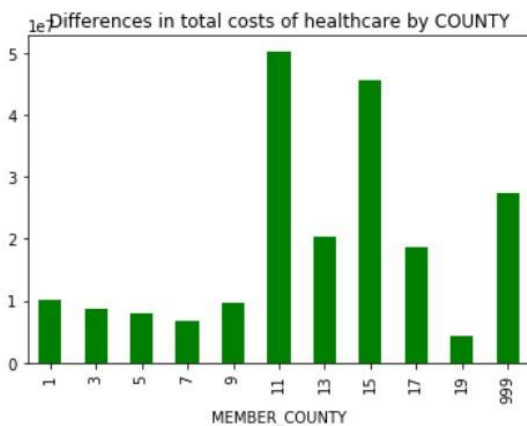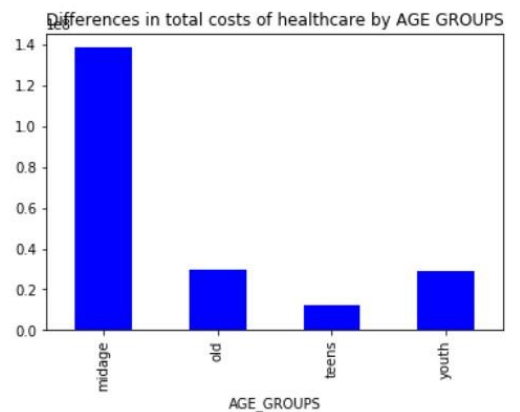

The total cost of healthcare within New Hampshire across the counties shows that Hillsborough had the highest expense, followed by Rockingham. This shows that Hillsborough had a total expense of 50 Million. The lowest expenditure in healthcare amongst the counties was in Sullivan. Non-New Hampshire states had the third highest expenditure of approx. 25 Million.

The total cost of healthcare within New Hampshire was 175 Million approx. This is because the claims data listed is for New Hampshire region mainly. Few of the neighboring counties from the states of Maine and Vermont might be included in this data which is why the value listed for them is very low, almost near 25Million.
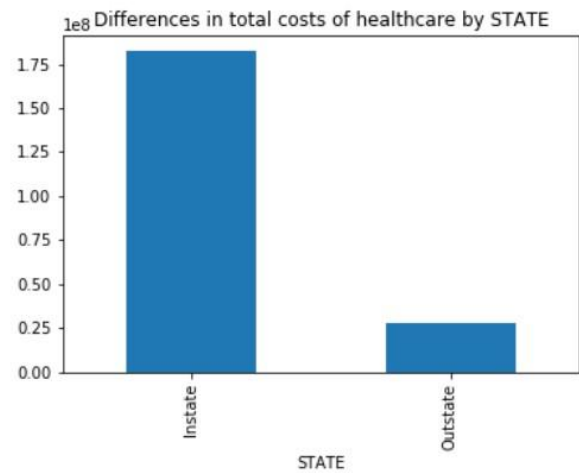
b) What health conditions, based on the frequency of primary ICD are most prevalent by state of residence and also by gender?
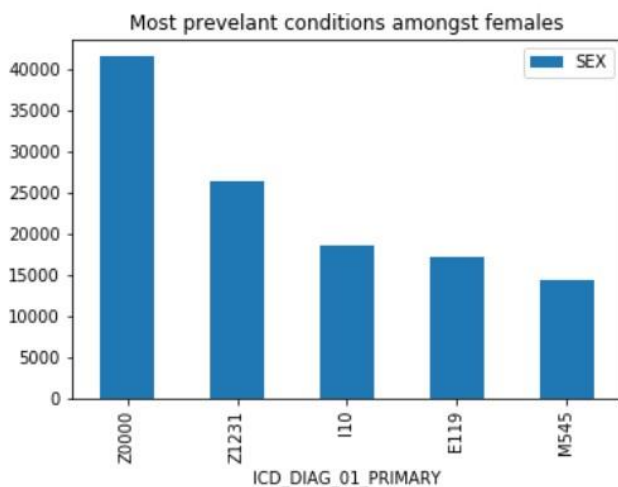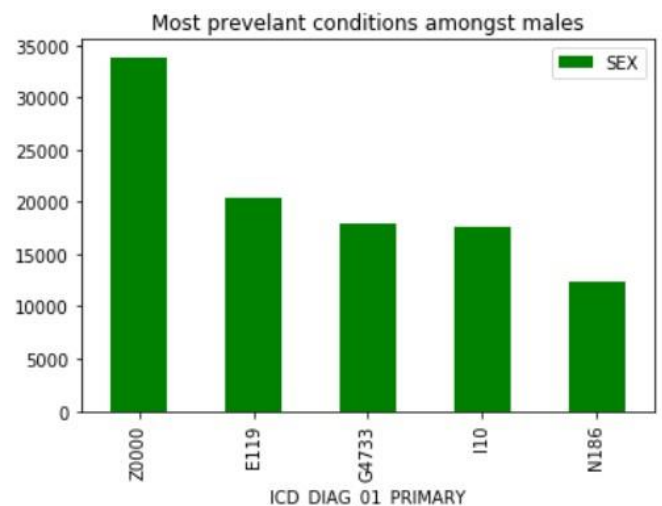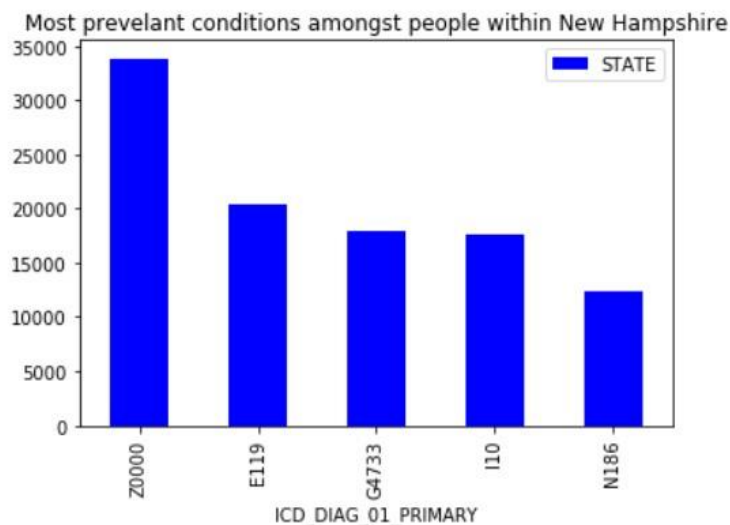

Women:

When observing the most prevalent condition amongst females from the claims data we can see that **Z0000** which corresponds to 'Encounter for general adult medical exam w/o abnormal findings' topped the list. Followed by **Z1231** which was 'Encounter screen mammogram for malignant neoplasm of breast', then **I10** which is 'Essential (primary) hypertension' and **E119** which is 'Type 2 diabetes mellitus without complications' and **M545** which is 'Low back pain'.

Men:

When the same analysis was performed on men, we can see that **Z0000** which corresponds to 'Encounter for general adult medical exam w/o abnormal findings' was the most common diagnosis in the claims data. Followed by **E119** i.e. 'Type 2 diabetes mellitus without complications'. Third most prevalent condition was **G4733** i.e. 'Obstructive sleep apnea (adult) (pediatric)', followed by **I20** i.e. 'Essential (primary) hypertension' and **N186** i.e. End stage renal disease.

When comparing the male and female population we can observe that the most common condition was general medical exam. This can also mean that adults are going for their physical check-ups regularly, since it is covered by many premiums. Another condition for which both males and females are visiting are for Diabetes diagnosis. Which shows that adults are becoming more aware and getting themselves checked for the condition to prevent future complications.

Most prevalent conditions amongst people within New Hampshire

When we compare the conditions across the State of New Hampshire, we found that the most prevalent condition here was **Z000** which corresponds to 'Encounter for general adult medical exam w/o abnormal findings'. This seems to follow the pattern similar to that of men.

Outstate:

Amongst people who apply for claims in New Hampshire even they belong to other states, the most prevalent condition seems to be **Z000** which corresponds to ' Encounter for general adult medical exam w/o abnormal findings', followed by **I10** which is ' Essential (primary) hypertension ' then **Z23** which is Encounter for immunization and **E119** i.e. 'Type 2 diabetes mellitus without complications'

Most prevalent conditions amongst people from other states who get diagnosed in New Hampshire
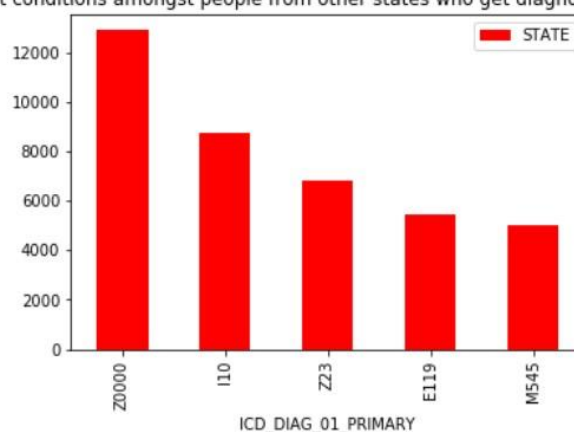
Top 5 prevalent health conditions for women belonging instate
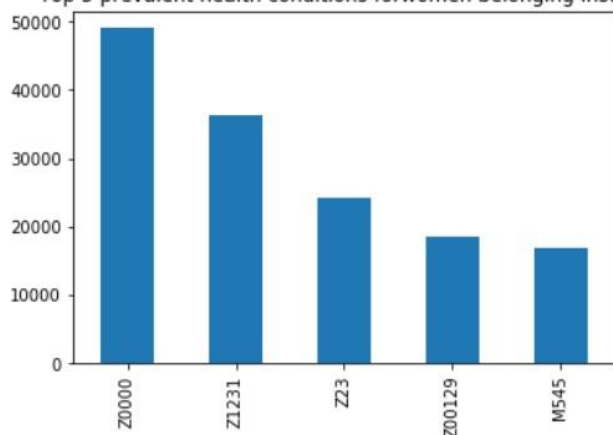
Women and instate:

It was observed that the top five conditions for women within the state of New Hampshire were **Z000**(Encounter for general adult medical exam w/o abnormal findings), **Z1231**(Encounter screen mammogram for malignant neoplasm of breast), **Z23**(Encounter for immunization), **Z00129**(Encounter for routine child health exam w/o abnormal findings), and **M545**(Low back pain)
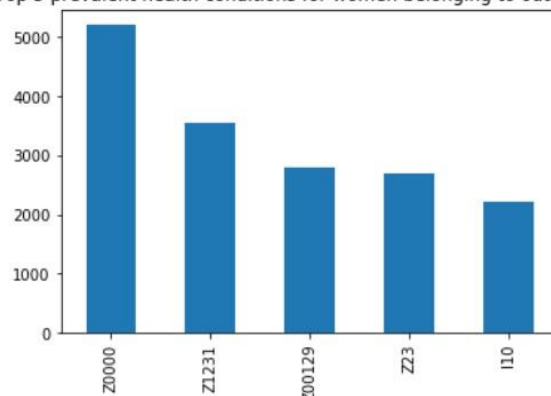

Top 5 prevalent health conditions for women belonging to out of state
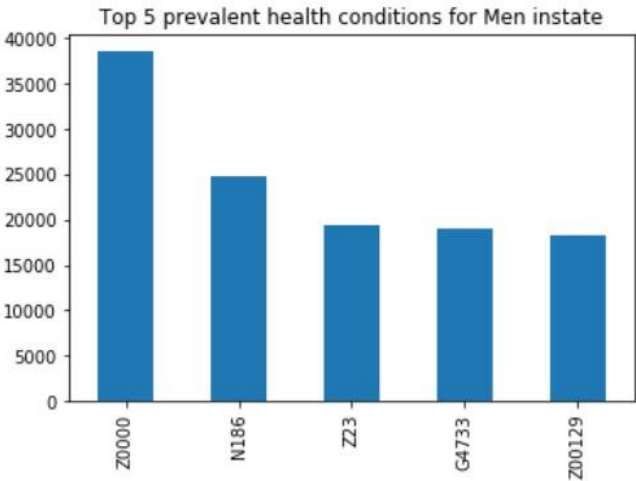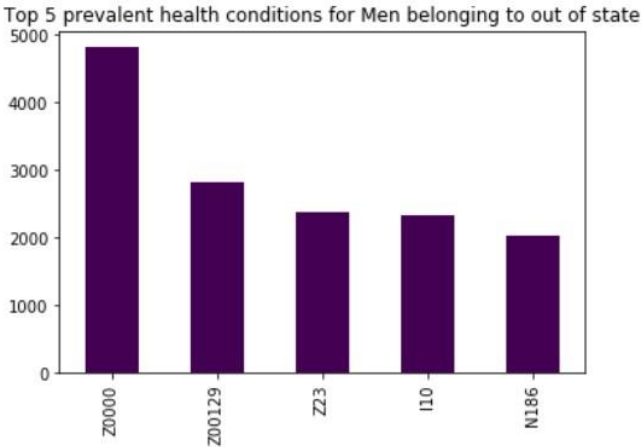
Women and Outstate:

It was observed that the top five conditions for women within the state of New Hampshire were **Z000**(Encounter for general adult medical exam w/o abnormal findings), **Z1231**(Encounter screen mammogram for malignant neoplasm of breast), **Z00129**(Encounter for routine child health exam w/o abnormal findings), **Z23**(Encounter for immunization) and by **I10**(Essential (primary) hypertension)

Top 5 prevalent health conditions for Men instate

Men and Instate:

Top conditions for the males in New Hampshire were **Z000**(Encounter for general adult medical exam w/o abnormal findings), **N186**(End stage renal disease), **Z23**(Encounter for immunization), **G4733**(Obstructive sleep apnea (adult) (pediatric)), **Z00129**(Encounter for routine child health exam w/o abnormal findings)

Men and Outstate:

Top conditions for the males from other states and diagnosed in New Hampshire were **Z000**(Encounter for general adult medical exam w/o abnormal findings), **Z00129**(Encounter for routine child health exam w/o abnormal findings), **Z23**(Encounter for immunization), **I10**(Essential (primary) hypertension), **N186**(End stage renal disease)

Top 5 prevalent health conditions for Men belonging to out of state

From this analysis, what we infer is that irrespective of a person being male or female, instate or out of state, everyone does their regular medical exam. Men have a more common pattern for claims for certain cases like renal disease and women have claims for checkups for breast cancer.

c) What health conditions, based on the total cost of care (amount paid) of primary ICD are most prevalent by state of residence and also by gender?

Female and Instate:

Amount paid for Top 5 prevalent health conditions for Female belonging to instate
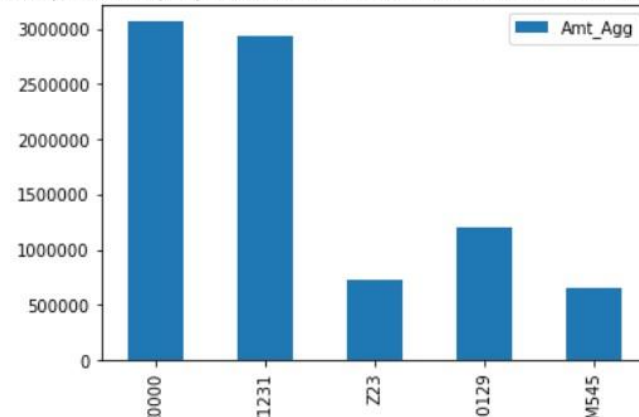
	ICD_DIAG_01_PRIMARY	Amt_Agg	ICD_DIAG_DESC
0	Z0000	3070774.00	Encntr for general adult medical exam w/o abno...
1	Z1231	2933415.16	Encntr screen mammogram for malignant neoplasm...
2	Z23	730108.42	Encounter for immunization
3	Z00129	1197378.51	Encntr for routine child health exam w/o abnor...
4	M545	654500.01	Low back pain

Female and Outstate:

Amount paid for Top 5 prevalent health conditions for Female belonging to out of state

	ICD_DIAG_01_PRIMARY	Amt_Agg	ICD_DIAG_DESC
0	Z0000	281130.32	Encntr for general adult medical exam w/o abno...
1	Z1231	229434.51	Encntr screen mammogram for malignant neoplasm...
2	Z00129	159897.31	Encntr for routine child health exam w/o abnor...
3	Z23	78246.68	Encounter for immunization
4	I10	55469.26	Essential (primary) hypertension

Male and Instate:
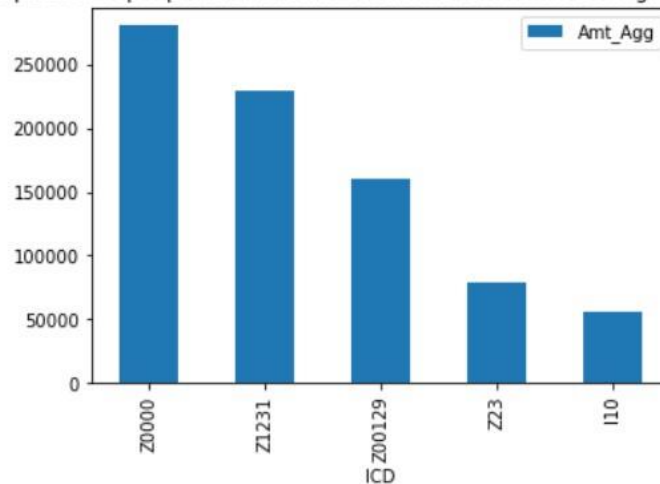

Amount paid for Top 5 prevalent health conditions for Men belonging to instate

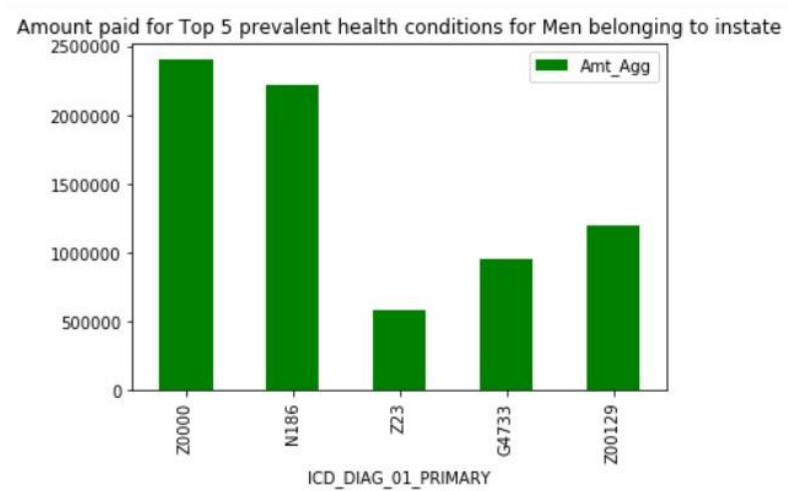|	ICD_DIAG_01_PRIMARY	Amt_Agg	ICD_DIAG_DESC
0	Z0000	2406101.67	Encntr for general adult medical exam w/o abno...
1	N186	2218627.12	End stage renal disease
2	Z23	587233.12	Encounter for immunization
3	G4733	949118.60	Obstructive sleep apnea (adult) (pediatric)
4	Z00129	1194350.01	Encntr for routine child health exam w/o abnor...

Men and Outstate:

Amount paid for Top 5 prevalent health conditions for Men belonging to out of state
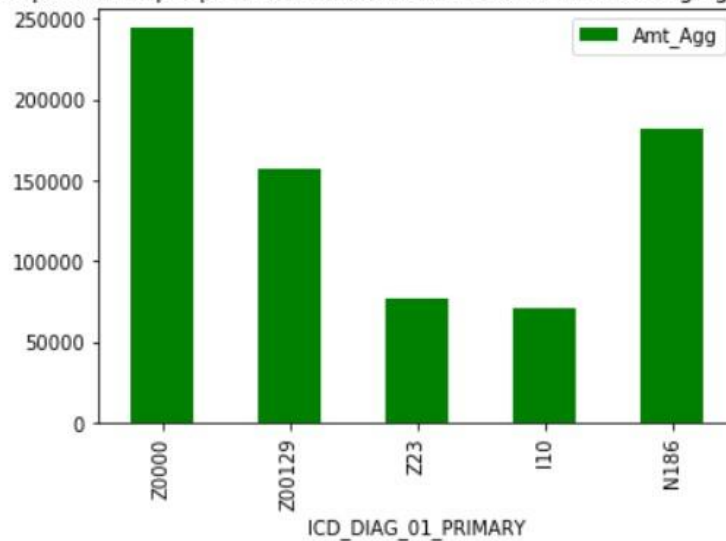
	ICD_DIAG_01_PRIMARY	Amt_Agg	ICD_DIAG_DESC
0	Z0000	244615.32	Encntr for general adult medical exam w/o abno...
1	Z00129	157121.53	Encntr for routine child health exam w/o abnor...
2	Z23	77300.65	Encounter for immunization
3	I10	70920.82	Essential (primary) hypertension
4	N186	182071.00	End stage renal disease

2) Analysis of health conditions: Find, state, and analyze the three most expensive health conditions (based on primary ICD codes along with their definitions) based on total billed amount for NH and out of state residents. You are expected to theorize in a written narrative, the main drivers for such high cost.

Instate residents:

	AMT_BILLED	ICD_DIAG_01_PRIMARY	ICD_DIAG_DESC
0	382801.77	C8339	Diffuse large B-cell lymphoma, extrnod and sol...
1	375455.00	J9602	Acute respiratory failure with hypercapnia
2	348981.51	O364XX0	Maternal care for intrauterine death, not appl...

We found that the most expensive procedure was 'Diffuse large B-cell lymphoma, extrnod and solid organ sites' within the residents of New Hampshire. Lymphoma is a type of blood cancer originating in the lymphatic system and occurs when the cells behave abnormally. We also know that treating blood cancer takes many visits to the doctor for chemotherapy and other abnormal behaviors that need to be reported immediately. This person maybe anemic too which may call for more iron supplements. Each siting for chemotherapy is very expensive too. Sometimes the patient might have to stay over at the hospital for many days, which might increase the cost of stay. The disadvantage of this disease is that the patient may suffer other side effects due to the treatment which may get diagnosed as comorbidities which may increase the overall cost. If we observe the pharmacy claim for the patient going through this treatment, it may be expensive too.

Second most expensive is Acute respiratory failure with hypercapnia. In this case, fluid tends to build up within the lungs and the lungs cannot remove carbon dioxide from the blood. Treatment for this includes oxygen tanks which the patient needs to depend on, which are very expensive. Or the doctor may attach a ventilator to one's throat which is also an expensive process. This may mean that one must get admitted to get this process done and some more time for one to recover at the hospital. This condition can cause long-term damage to one's lungs which may mean frequent visits to the hospital and lot of medication. This condition may trigger emergency events too. The other conditions that this patient has to undergo is **atrial fibrillation and the person may also have diabetes.**

Third most expensive procedure was Maternal care for intrauterine death. This occurs when the fetus that was expected to survive dies during the birth or in the last half of the pregnancy. Since this involves the cost for many departments such as pediatrics and other organ related departments, the cost is high. The complications that can occur to the mother during and after this are high and hence the prevention methods for it are high too. Ultrasound scans need to be performed on the patient to monitor the unborn fetus's behavior which are added costs to the procedure. The other conditions that arise with this is that the mother should be screened for genetic and chromosomal anomalies too. All this adds up to a high cost for the provider.

Outstate residents:

	AMT_BILLED	ICD_DIAG_01_PRIMARY	ICD_DIAG_DESC
0	231206.00	Z3800	Single liveborn infant, delivered vaginally
1	184608.00	Z3800	Single liveborn infant, delivered vaginally
2	194210.63	K5641	Fecal impaction
3	135283.82	Z421	Encounter for breast reconstruction following ...

Patients belong from other states who file their claim in New Hampshire may be filling in New Hampshire as they might that come to visit but have resulted in an emergency which might have triggered them to visit the hospital. The claim for which they were billed the most is Single liveborn infant, delivered vaginally. This probable means that the lady might have been in labour when she was travelling to/from/through New Hampshire, which might have caused her to get admitted in a hospital as an emergency. This also means that it was a normal birth but when admitted as an emergency case, the costs shoot up may fold. The comorbidities associated with this has to be care for the child, as the child has to be immunized to prevent any disease from attacking it and also immunizing the mother. This may result in staying at the hospital for an extra day which is also expensive. The majority of the cases were for emergency child birth situations.

The next condition for which providers were charged heavily is Fecal impaction. This is more serious version of constipation which may have happened to the person during travel since they are not New Hampshire residents. The person has to have scans, they might have been dehydrated which may call for drips and other related conditions. This might also be due to diabetes.

The third most common condition is breast reconstruction after mastectomy. Women tend to surgically remove their breasts to avoid breat cancer and those with breast cancer remove this to avoid the cancer cells from spreading to other regions of the body. To get back the appearance of the breast post the surgery is what this is. This involves two major fields oncology and plastic surgery. The related conditions to this might be disproportionate breasts which may cause the doctor to do rework. This is related to cancer, which is rather an "expensive" disease only because of the expenses that come with the chemotherapy, surgery and medications which traumatise the patients.

3) Analysis of the sickest patient: Identify and describe, using written description supported by numerical analysis, the sickest patients for both NH and out of residents separately.

We defined the sickest patient to be a person who has a lot of claims. Indirectly it means that he has had a lot of visits to the doctor and hospitals. He may have had procedured done on him that require him to come often. We took the SV_LINE column from the data, since it means connecting procedures for an individual, and saw the highest value it can take. And then saw the corresponding CLAIM_ID_KEY for it which represents an individual. This resulted in an induvidual visiting 206 times but the sum of the AMT_BILLED column came upto 22K only. Since this was not a claim that was high enough, we went for a different approach.

We grouped the table by the CLAIM_ID_KEY, and summed all the related AMT_BILLED for that individual and found their diagnosis. Repeated this to the instate dataframe and then to the outstate dataframe. This resulted in fairer results.

Instate:

	CLAIM_ID_KEY	ICD_DIAG_01_PRIMARY	AMT_BILLED	ICD_DIAG_DESC
0	119396347	E840	1710110.28	Cystic fibrosis with pulmonary manifestations

This person seemed to be the "sickest" patient in the instate data. This person was of 27 year old male. He has had over 30 visits to the doctor/hospital. In cycstic fibrosis, the fluids in your body tend to thicken, this cause blockage in many organs. The only organ that doesn't get infected is the lungs. This person will end up having too many comorbidities, each of which relates to an organ in the body since almost all of them get affected. Even if doesn't affect the lung directly, it indirectly cause disorders of the lung. It also causes pneumonia which is a life-threatening disease. It causes dysfunction to the heart too. We can imagine that if this person has become so ill, he would have had to undergo many procedures, some of which may cause him to be hospitalised for a long time. It makes sense to call him the "sickest" patient. The claims that are caused due to these procedures might have been intricate as the disease kills from the inside.

Outstate:

	CLAIM_ID_KEY	ICD_DIAG_01_PRIMARY	AMT_BILLED	ICD_DIAG_DESC
0	15748474	T827XXA	3115189.50	Infect/inflm react d/t oth cardi/vasc dev/impl...

This person seemed to be a 68 year old male who was from Maryland. He had 72 claims under his name. He had Infection and inflammatory reaction due to other cardiac and vascular devices, implants and grafts, subsequent encounter. This means that he had already had procedures done on him in the previous years. He either had an implant or a graft inside his body to help his heart function normally through artificial means. The data also says that he had severe sepsis which is a life-threatening condition that arises when the body's response to infection causes injury to its own tissues and organs. He also had an acute kidney failure with tubular necrosis. He had too many ongoing disease in his body and he definitely needed constant care and procedures to cure his disease. Even though his provider was paying a lot, this man also had to put in quite a bit for all the procedures. If it wasn't for his health insurance he might not have been able to afford this care