

Exploratory Data Analysis of Drug Consumption

Sowmiya K S, Saranya. K, Sumathi. V. P

Abstract-Drug utilization is the consumption of the drug in all the state or country. In the rushing world, technology development is peaked. Human beings are fully addicted with technology, intake of food and that leads to easy attack on diseases results. To cure the disease, intake of medicines at over dosage of chemical compounds leads to lot of side effects. In previous analysis, it tells about the maximum and minimum utilization of drugs and increase in number of enrolls and the mean number of prescription can be found. In addition to that, amount spent for drugs for Medicaid enrolls has been increased in the last decade dramatically Exploratory data analysis (EDA) is an approach for analyzing data sets to summarize their main characteristics, often with visual methods. The methodologies used for this analysis are statistical analysis - collecting, exploring and presenting large amounts of data in a graphical view-pie chart, bar, line graph can be plotted. This analysis speaks the maximum and minimum utilization of drug with limited dosage in a particular state, comparison of the drug with maximum utilization in every quarter can be analysed with the count of prescriptions given in the data set. Maximum drug used in particular time period and the survey of disease is shown.

Keywords: Drug consumption, Cost, National drug code, FFSU

I. INTRODUCTION

In recent years, the contribution of medical to the treatment of diseases has been increased most rapidly than other pharmaceutical approaches. This increase will result in intensification of expenditure of medical drugs. Medicaid is the top payer of medical services in the United States. Therefore, the Center of Medicare and Medicaid services of United States is releasing State Drug Utilization data every year. This paper analyses about the State Drug Utilization Data in 2016. The data contains the National Drug Code (NDC), utilization type, state code, drug used in number of prescriptions, location of the state and the reimbursement of Medicaid and Non - Medicaid agencies in dollars and also the name of the product approved by Food and Drug Administration (FDA). The utilization type is of two types: MCOU (Managed Care Organization Utilization) and FFSU (Fee for Service Utilization). The statement is that the drug with maximum number of prescriptions has to be found in both utilization types. The drug with the minimum amount of reimbursement during the quarter/year period has to be found.

Manuscript published on 30 December 2019.

*Correspondence Author(s)

Sowmiya K, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India,
S,Saranya.K, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India,
Dr.Sumathi.V.P, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India,

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

In the previous analysis, the state with maximum drug utilization, maximum enrolls and the mean number of prescription per enroll has been found among all states in a country. In this analysis, we have extended to the state with maximum number of MCOU or FFSU type has to be found. The top 30 states with the maximum number of prescriptions has found. The states with no utilization of drug is to be found. These analyses have been done with the statistical functions applied to the particular fields to be involved. Data for the study has taken from the Center of Medicare and Medicaid services of United States. It is previously known as Health Care Financing Administration, a federal agency associated with United States Department of Health and Human Services. Medicaid is the health insurance program for poor or low income people of all the ages of the country and disabled individuals. The Medicaid program is a means-tested program run by both state and federal governments hence the fund comes from the collaboration of state and federal dollars. But there is a variation of fund from one state to another. The individual or family which is not eligible for Medicaid are serviced under Non- Medicaid agencies. Medicare program is a national social insurance program and it is administered by only federal government using some private insurance companies all over US under contract. It is funded by a payroll tax, surtaxes from beneficiaries and general revenue and is consistent all over the country. The individual with age 65 and more who worked through payroll tax and young people with disabilities are provided the health insurance. The Center of Medicare and Medicaid Services of United States is taking survey of the reimbursement/ dispense done for the product and the utilization of the drugs every year. This data contains the Drug Utilization in the year 2016 in every state. The data set indicates the two type of utilization applied in every state of United States. They are MCOU (Managed Care Organization Utilization) and FFSU (Fee for Service Utilization). The Managed Care Organization plans is a health insurance type having contracts with health care providers and medical facilities to the poor people and disabled people in the state to the reduced cost. The Fee for Service Utilization plans offer Medicaid benefits through managed care plans or both. Every state pays providers directly for each service got by a Medicaid beneficiary. The state code consists of two characters (for post office) for a single state otherwise 'XX'. The NDC is of 11 digits which is segmented into labeler code, product code and package size code. The number of prescription field is to denote the particular drug utilized by number of patients. The product FDA list name is the first 10 characters of the name of the product approved by FDA. The latitude and longitude represents the location within the state to create maps.



The drug is prepared by the health care organization and the imbursement is done by the Medicaid as well as the non – Medicaid agencies. The unit reimbursed field is the unit of drugs reimbursed or dispensed in a quarter/year period.

The quarter is the particular period of time the NDC is paid for by the state government in FFSC and dispensed as well as by the state in MCOU. The suppression used is that the direct identifiers have been removed as per the Privacy Act and the data suppressed is displayed in that field. In the section 2, the overview and analysis of all the previous papers on drug consumption has been explained with the comparison table. In the section 3, the various analysis of the dataset has been done and the resultant observation has been explained for every analysis separately. In the section 4, all the analysis are put together and the resultant data has been explained. In the section 5, the major result obtained has been told and the future works apart from this analysis has been told.

II. LITERATURE REVIEW

The paper deals with the identification of the drug to cure disease in the specific area. Analyses of maximum and minimum utilization of drug in the state. Payable and nonpayable drug can be found by using the analysis. The drug that can be used for old aged people and orphanages that can be sold freely for them. The figures 1.1 and 1.2 shows the comparison of all the review papers corresponding to the drug consumption domain. The paper “Trends and Current Drug Utilization Patterns of Medicaid Beneficiaries” by Terry R. Lied, Ph.D., Julio Gonzalez, M.P.H., Wendy Taparanskas, Ph.D., and Tejas Shukla, M.S. in 2006 say that Increase in drug utilization and expenditure, results in increase in Medicaid enrolment, the mean number of prescriptions per enrol, mean nominal and inflation-adjusted reimbursement per prescription, and the tendency for increased use of new and more expensive drugs. The paper “Changes in Drug Utilization during a Gap in Insurance Coverage: An Examination of the Medicare Part D Coverage Gap” by Jennifer M.Polinski, William H. Shrank, Haiden A. Huskamp, Robert J. Glynn, Joshua N. Liberman, Sebastian Schneeweiss in 2011 analyses the beneficiaries who reached the coverage gap spending threshold, there were covariate imbalances across beneficiary groups. There was a high prevalence of cardiovascular conditions and diabetes across groups.

S.N	AUTHOR NAME	YEAR PUBLISHED	OBJECTIVE	METHOD	Results obtained	ORIGIN
1.	Terry R. Lied, Ph.D., Julio Gonzalez, M.P.H., Wendy Taparanskas, Ph.D., and Tejas Shukla, M.S.	2006	To investigate trends in noninstitutional drug utilization and expenditures in the Medicaid Program.	Statistical Technique	Increase in drug utilization and expenditure, results in increase in (1) Medicaid enrolment, (2) the mean number of prescriptions per enrollee, (3) mean nominal and inflation-adjusted reimbursement per prescription, and (4) the tendency for increased use of new and more expensive drugs.	National Medicaid data from 1994-2003 State Drug Utilization Data Files for 1994-2003 (Medicaid Drug Rebate Program) and the Master Drug Database of MediSpan.
2.	Jennifer M.Polinski, William H. Shrank, Haiden A. Huskamp, Robert J. Glynn, Joshua N. Liberman, Sebastian Schneeweiss	August 16, 2011	It examined the impact of entering the coverage gap on drug discontinuation, switching to another drug for the same indication, and drug adherence. While increased discontinuation of and adherence to essential medications to a regrettable response. Increased switching to less expensive but therapeutically interchangeable medications to a positive response to minimize costs.	Statistical Technique	Beneficiaries who reached the coverage gap spending threshold, there were covariate imbalances across beneficiary groups. There was a high prevalence of cardiovascular conditions and diabetes across groups.	Medicare Part D Coverage Gap in US Government.

Fig 1.1., Comparison of Review papers

The paper “Comparison of drug utilization patterns in observational data: antiepileptic drugs in pediatric

patients” by Florence T Bourgeois, MD, MPH Karen L Olson, Ph.D., Annapurna Poduri, MD and Kenneth D Mandl, MD, MPH in 2015 analyses show that the Comparing of drug utilization patterns in a paediatric population using observational data, we found similar rates of retention and therapeutic changes. These findings are consistent with available comparative data and demonstrate an approach that could be extended to other drug classes and conditions in paediatric populations to examine drug effectiveness.

S.NO	AUTHOR NAME	YEAR PUBLISHED	OBJECTIVE	METHODOLOGY	Results obtained	ORIGIN
3.	Florence T Bourgeois, MD, MPH Karen L Olson, Ph.D., Annapurna Poduri, MD and Kenneth D Mandl, MD, MPH	2015	To use observational data to measure and compare medication utilization patterns in a pediatric patient population.	Statistical Technique	Comparing drug utilization patterns in a pediatric population using observational data, we found similar rates of retention and therapeutic changes. These findings are consistent with available comparative data and demonstrate an approach that could be extended to other drug classes and conditions in pediatric populations to examine drug effectiveness.	Observational Data on drug utilization between 1994 and 2009
4.	David K. Baugh, M.A., Penelope L. Pine, Steve Blackwell, Ph.D., J.D., R.Ph., and Gary Ciborowski, M.A.	2004	To target cost-effective coverage and drug therapies by containing information on drug spending for dually eligible beneficiaries to policymakers.	Statistical Technique	Spending for prescription drugs for aged and disabled Medicaid enrollees has increased dramatically in the last decade. The increases are a result of many factors (and possibly interactions among the factors), including: increasing enrolment (particularly for disabled persons), increased use of drugs, cost inflation, new drug therapies (e.g., blockbuster drugs), the use of single source versus generic drugs, and state policies to control drug utilization.	Medicaid Prescription Drug Spending in the 1990s in US.
5.	Richard G. Abramson, M.D., Catherine A. Harrington, Pharm. D., Ph.D., Raad Missmar, Susan P. Li, and Daniel N. Mendelson, M.P.P.	2004	To optimize their efforts include focusing on pricing for drugs with high sales volumes, ensuring that MAC lists include prices for all forms and dosages of listed drug.	Statistical Technique	There is variability in MAC program breadth, depth, and pricing aggressiveness. MAC lists contain prices for as many available forms and dosages of a listed product as possible. Our data demonstrated surprising variability in depth for non-FUL drugs, indicating that MAC lists often do not include some of	The Federal upper limit (FUL) program and State maximum allowable cost (MAC) programs

Fig 1.2., Comparison of Review papers

The paper “Medicaid Prescription Drug Spending in the 1990s: A Decade of Change” by David K. Baugh, M.A., Penelope L. Pine, Steve Blackwell, Ph.D., J.D., R.Ph., and Gary Ciborowski, M.A. in 2004 analyses speak about spending for prescription drugs for aged and disabled Medicaid enrollees has increased dramatically in the last decade. The paper analyses speak increases are a result of many factors (and possibly interactions among the factors), including: increasing enrolment (particularly for disabled persons), increased use of drugs, cost inflation, new drug therapies (e.g., blockbuster drugs), the use of single source versus generic drugs and State policies to control drug utilization. The paper “Generic Drug Cost Containment in Medicaid: Lessons from Five State MAC Programs” by Richard G. Abramson, M.D., Catherine A. Harrington, Pharm. D., Ph.D., Raad Missmar, Susan P. Li, and Daniel N. Mendelson, M.P.P. in 2004 tells about that there is variability in MAC program breadth, depth, and pricing aggressiveness. MAC lists contain prices for as many available forms and dosages of a listed product as possible. Our data demonstrated surprising variability in depth for non-FUL drugs, indicating that MAC lists often do not include some of the available strengths and forms (i.e., GCNs) of listed drug entities.

III. IMPLEMENTATION ANALYSIS

Exploratory data analysis is a methodology for approaching a data sets to analyse their main characteristics with visual methods. It is for seeing the data beyond the formal or hypothesis testing task. Its techniques have been adopted into data mining, as well as big data analytic. There are number of tools that are useful for EDA, but EDA is characterized by the attitude more than by techniques. Some graphical techniques used in EDA are histogram, line chart, bar chart, pie chart etc.



In statistical technique, it is used for collecting, exploring and presenting the large amounts of data. This analyses speaks the EDA methodology for analysing the data sets in graphical and statistical view. In this dataset, the state labelled XX has the maximum utilization of drugs, maximum number of labellers and the maximum amount of reimbursement for drugs (11 digit NDC). Hence, the dataset consisting of the record with the State XX has been included for an analysis and the record with the State XX has been dropped for another analysis. The comparison of the two datasets has been taken place for every data analysis.

ANALYSIS 1

The Centre of Medicare and Medicaid agencies are reimbursing amount for the 11 digit NDC drug every year in the period covered. A year is being split into four periods.

The fig 2.1 shows the comparison analysis of the drug consumption in every quarters for the top 12 utilized drugs with the record XX.

The first period is named as Quarter 1 which is the duration of 3 months (January 1 – March 31). The second period is considered as Quarter 2 also the duration of 3 months (April 1 – June 30). The third and fourth period is named as Quarter 3 and Quarter 4 each of which is 3 months (July 1 – September 30) and (October 1 – December 31). The 10 drugs (11 digit NDC) with the maximum utilization has been calculated. The NDC is varied for a product by different Product code,

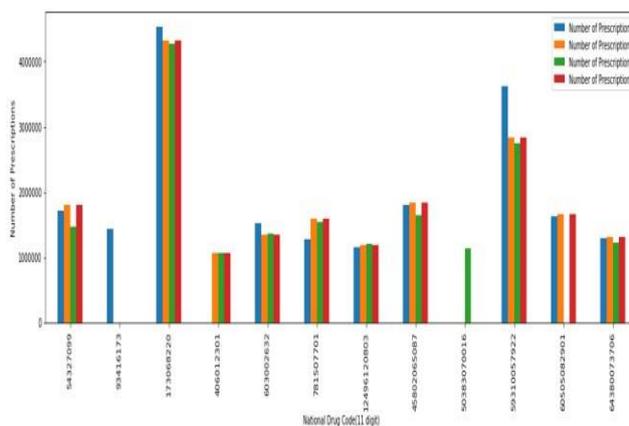


Fig 2.1., Analysis with record XX

Package size code and the Labeller code. Its utilization had been compared for all the four quarters with the count of prescriptions in all the states. It infers that the NDC 173068220 (VENTOLIN H) had been mostly used in all the states. Most of the people in all states are suffered from a breathing disease which is bronchospasm. It has been utilized by most of the people. It suffers the people as a side effect of the drugs used to treat hypertension, and other deadly nightshade diseases.

The fig 2.1 shows the comparison analysis of the drug consumption in every quarters for the top 12 utilized drugs without the record XX.

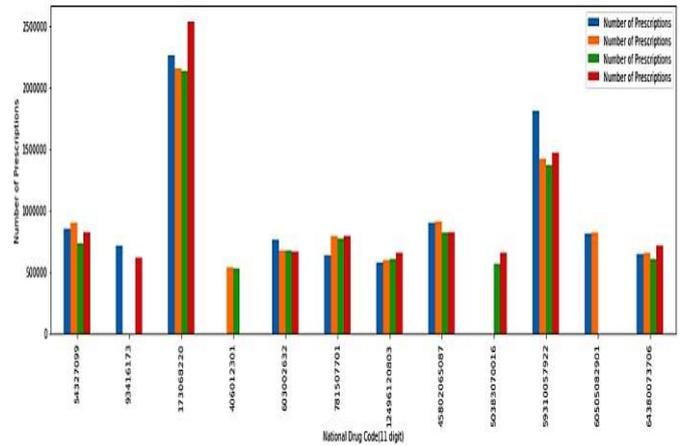


Fig 2.2., Analysis without the record XX

In comparison of the quarters in the first dataset, most of the drugs had been utilized in the first quarter i.e., in the duration of January to March. And in the second dataset, drugs have been mostly utilized in the fourth quarter i.e., in the duration of October to December. The only record XX changes the result analysis for various quarters.

ANALYSIS 2

The Centre of Medicare and Medicaid Agencies are reimbursing amount every year in a period covered. This analysis compares the amount reimbursed by Medicaid and Non Medicaid agencies for every state in the year 2016. This graph compared the least amount spent by both the agencies in various states. Both the figures 2.3 and 2.4 infer that the Medicaid agencies are reimbursing money more than Non Medicaid agencies. The very least amount is spent in WY (Wyoming) state by the Medicaid agencies and also by the Non Medicaid Agencies.

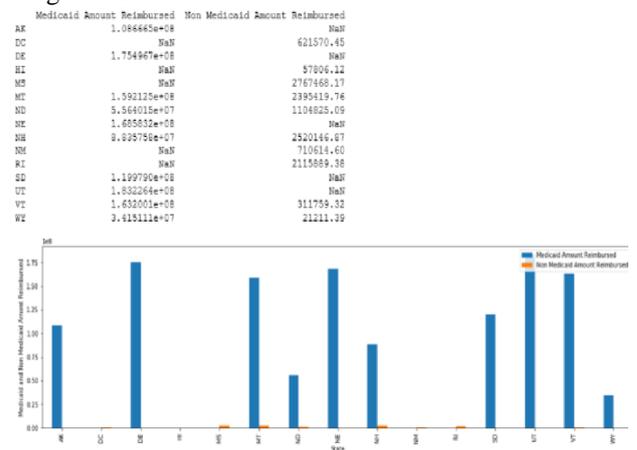


Fig 2.3., Analysis without the record XX

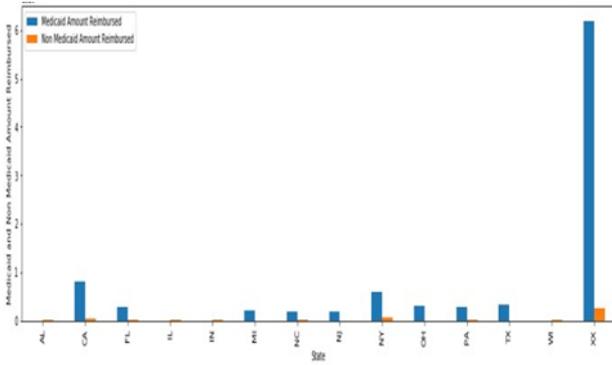


Fig 2.4., Analysis without the record XX

ANALYSIS 3

This analysis compares the amount reimbursed by Medicaid and Non Medicaid agencies for every state in the year 2016. This graph compared the largest amount spent by both the agencies in various states. The Medicaid agencies are investing money for the low or poor income people and the disabled persons. And the Non Medicaid agencies are investing money for those who are not eligible for Medicaid agencies. Both the graphs infer that the Medicaid agencies are reimbursing money more than Non Medicaid agencies. The figure 2.5 says that the very largest amount reimbursed in the state representation XX (which is not an US state, not an Union territory, not Canada and the country is known) by both Medicaid and Non Medicaid agencies

State	Medicaid Amount Reimbursed	Non Medicaid Amount Reimbursed
AK	1.086650e+08	NaN
DC	NaN	621570.45
DE	1.754967e+08	NaN
HI	NaN	57006.12
MD	NaN	2767460.17
MT	1.592125e+08	2395419.76
ND	5.564015e+07	1104625.09
NE	1.665832e+08	NaN
NH	8.095750e+07	2520146.87
NM	NaN	710614.60
RI	NaN	2115889.28
SD	1.190760e+08	NaN
VT	1.622261e+08	NaN
WY	1.632001e+08	311750.13
XX	9.125111e+07	21211.29

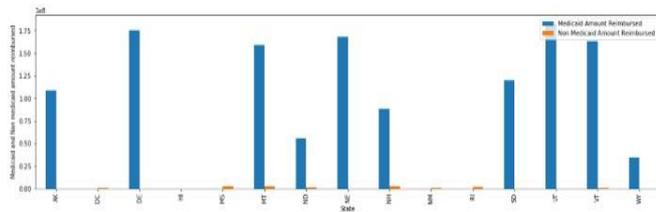


Fig 2.5., Analysis with the record XX

State	Medicaid Amount Reimbursed	Non Medicaid Amount Reimbursed
AL	7.983591e+09	6.585127e+07
CA	2.815700e+09	2.591087e+08
FL	2.815700e+09	9.410328e+07
IL	1.602363e+09	7.587066e+07
IN	NaN	9.301486e+07
MI	1.944591e+09	NaN
MO	1.737224e+09	7.226443e+07
MT	1.696080e+09	NaN
NE	5.390611e+09	5.392320e+08
NH	3.045803e+09	NaN
OH	NaN	5.388199e+07
PA	2.770619e+09	1.7471000e+08
TX	3.338447e+09	NaN
WI	NaN	5.739941e+07

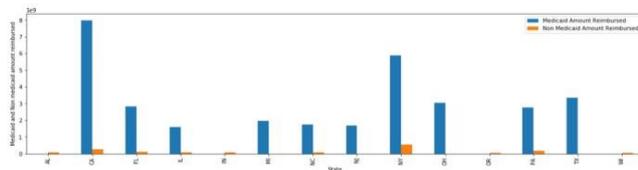


Fig 2.6., Analysis without the record XX

The figure 2.6 infers that the very largest amount is spent in CA (California) state by both the Medicaid agencies and the Non Medicaid Agencies.

Comparison of two graphs infer that the different result is obtained i.e., the state XX is reimbursing the amount more than California for the drugs. This may say that the population or Patients in the XX or CA may be high.

ANALYSIS 4

This analysis shows that the state with the maximum usage of a particular drug. The 11 digit National Drug Code is taken as the input from the user. For the given input, it counts the number of prescriptions mentioned for the particular drug in various states and displays the top 30 states with the maximum utilization of that drug. Here,

the input is given as 64380073706 (NDC) which is VITAMIN D2, Ergocalcif, VIT D2 1.2 drug composition. In the figure 2.7, it says that this drug has been mostly utilized by the state representation XX and then by NY (New York) and then CA (California) and so on.

Enter the National Drug Code:64380073706

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:10: FutureWarning: by argument to sort_index is deprecated, pls use .sort_values(by=...)
# Remove the CMD from sys.path while we load stuff.
```

['Number of Prescriptions']

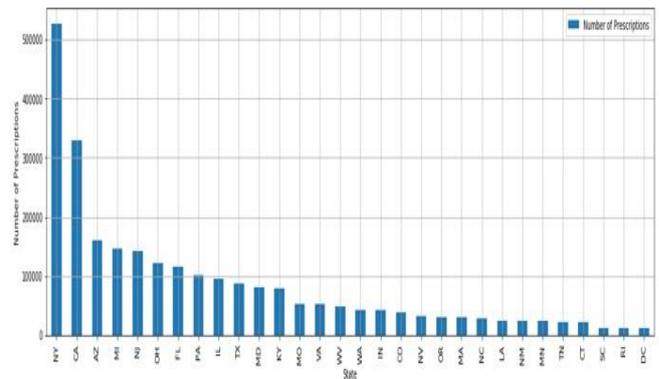


Fig 2.7., Analysis with the record XX

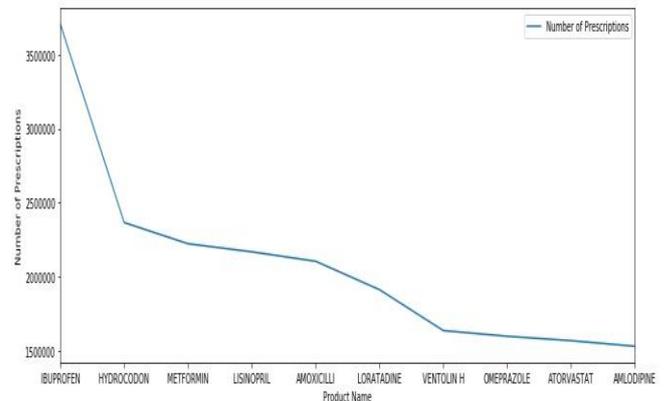


Fig 2.8., Analysis with the record XX



The figure 2.8 infers that NY (New York) is the leading state in the utilization of the particular given drug.

This analysis has been done for other particular drugs, that infers the state representation XX is utilizing most of the drugs in peak and then the other states. This may conclude that the representation XX consists of huge population and the diseases are spread over everywhere.

ANALYSIS 5

This analysis shows that the drug with the maximum usage in a particular given state. The State name is taken as the input from the user. For the given input, it counts the number of prescriptions for every drugs in the given input state and displays the top 10 drugs with the maximum utilization in that state.

```
Enter the National Drug Code:64980073706
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:9: FutureWarning: by argument to sort_index is deprecated, pls use .sort_values(by=...)
if __name__ == '__main__':
['Number of Prescriptions']
```

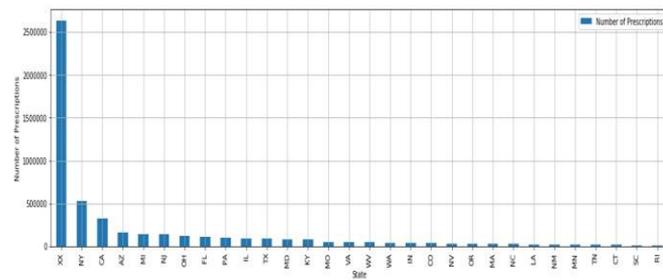


Fig 2.9 Analysis without the record XX

In the figure 2.9, the state representation XX (which is not an US state, not an Union territory, not Canada and the country is known) is taken as input. The number of prescription for very drug utilized in that state has been calculated and the output is display in the form of line graph. The drug AMOXICILLI is utilized by most of the people in the particular area and secondly IBUPROFEN and then LISINOPRIL and so on.

ANALYSIS 6

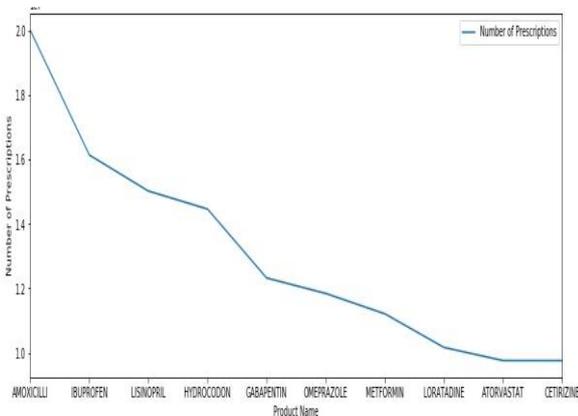


Fig 2.10 ., Analysis with the record XX

In the figure 2.10, the state representation CA (California) is taken as the input. The drug IBUPROFEN had been mostly utilized by the people in the state and secondly HYDROCODON and then METFORMIN and so on. It infers that the people in that state are mostly attacked by fever, headache, back pain and arthritis. This analysis can be

done for all the other states to identify the disease which is mostly attacking the people in that particular state.

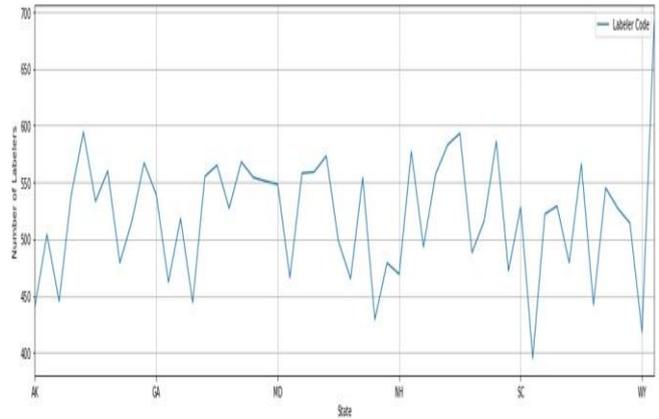


Fig 2.11., Analysis with the record XX

In the figure 2.11, it infers that the state representation XX has the maximum number of labellers in it. In this particular state, they are introducing more number of products in the market.

ANALYSIS 7:

In the figure 2.12, it infers that the state CA (California) has the maximum number of labellers for the packaging and introducing products in the market. It results that this state has the maximum number of patients with the maximum number of diseases. Labelling is the process of compressing a product into the market. This analysis states that labellers are working in the every state of US.

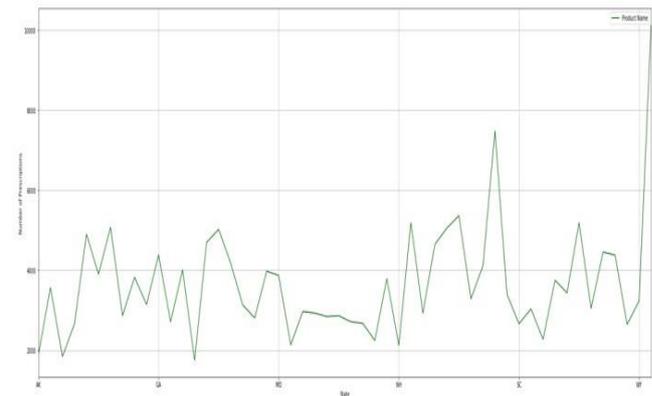


Fig 2.12., Analysis without the record XX

The various drugs have been utilized in every state of US. The number of products for which have been reimbursby e In the figure 2.13, it has been analysed that the state representation XX reimburses for more number of drugs.

In the figure 2.14, the analysis shows that the state PA (Pennsylvania) had utilized more number of drugs for the patients.



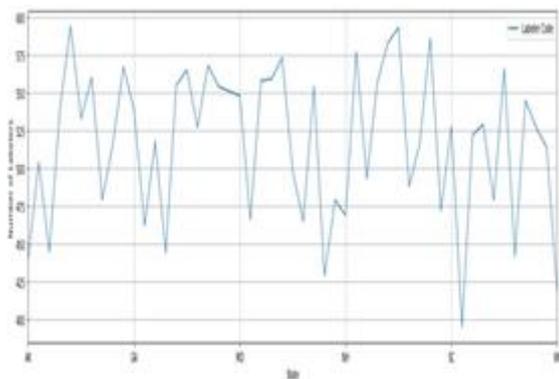


Fig 2.14., Analysis with the record XX

IV. IMPLEMENTATION RESULT

The figures 1.1, 1.3, 1.5, 1.7, 1.9, 1.11, 1.13 obtained in every analysis resulted that the state representation XX (which is not an US state, not an Union territory, not Canada and the country is known) is exhibiting the highest drug utilization rate, highest labeller count rate, highest amount reimbursement rate and hence it confirms that this state is containing the huge population in it. In the analysis of the figures 1.2, 1.4, 1.6, 1.8, 1.10, 1.12, 1.14 the state record XX has been dropped and analysed. In that analysis, the state CA (California) is reimbursing high amount for the drugs. And the number of labellers employing the state California is high. In the result of all the analysis, it infers that the drug VENTOLIN H is in high utilization in all the quarter periods. The utilization of that drug in every quarter varies in analysis of two datasets. And also the amount reimbursed for the drugs is lower in the state WY (Wyoming). The top 10 drugs utilized by every state can be found.

V. CONCLUSION AND FUTURE WORK

This paper infers that the most of the people in all states are suffered from a breathing disease which is bronchospasm. The drug VENTOLIN H has been utilized by most of the people. It suffers the people as a side effect of the drugs used to treat hypertension, and other deadly nightshade diseases. This tells people that the utilization of the drugs for curing the diseases can be avoided. Instead of that, they can increase the intake of Vitamin D rich foods like leafy greens, milk, carrots and eggs. And the count of labellers working are very high. The reduction in the working labellers can also be done when the drug consumption gets low.

REFERENCES

1. The paper "Trends and Current Drug Utilization Patterns of Medicaid Beneficiaries" by Terry R. Lied, Ph.D., Julio Gonzalez, M.P.H., Wendy Taparanskas, Ph.D., and Tejas Shukla, M.S. in 2006. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194949/>
2. The paper "Generic Drug Cost Containment in Medicaid: Lessons from Five State MAC Programs" by Richard G. Abramson, M.D., Catherine A. Harrington, Pharm. D., Ph.D., Raad Missmar, Susan P. Li, and Daniel N. Mendelson, M.P.P. in 2004. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194860/>
3. The paper "Medicaid Prescription Drug Spending in the 1990s: A Decade of Change" by David K. Baugh, M.A., Penelope L. Pine, Steve Blackwell, Ph.D., J.D., R.Ph., and Gary Ciborowski, M.A. in 2004. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194863/>
4. The paper "Comparison of drug utilization patterns in observational data: antiepileptic drugs in pediatric patients" by Florence T

Bourgeois, MD, MPH Karen L Olson, Ph.D., Annapurna Poduri, MD and Kenneth D Mandl, MD, MPH in 2015 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573831/>

5. The paper "Changes in Drug Utilization during a Gap in Insurance Coverage: An Examination of the Medicare Part D Coverage Gap" by Jennifer M. Polinski, William H. Shrank, Haiden A. Huskamp, Robert J. Glynn, Joshua N. Liberman, Sebastian Schneeweiss in 2011 <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001075>
6. <https://www.doctorshealthpress.com/food-and-nutrition-articles/alternative-remedies-food-and-nutrition-articles/bronchospasm-treatment/>