Statistical
Data
Mining
Project – Fall 2021

Health Care Analytics

Effects of Health care expenditure on Mortality and Pharma
USA | UK | Australia | Israel | Canada

Team Members:

1. **Suryanarayana Aneesh Prayaga – U 93559651**
2. **Raj Phani Kakarla – U 04426707**
3. **Vinay Ratnala – U 12127232**
4. **Karna Harsha Vardhan Reddy- U 73857073**

Executive Summary

Executive Summary: Every year many countries around the world adds millions of dollars for the health care yet they show very little effect on the citizens. The Per-capita expenditure on health care has increased 31-fold for the past 4 decades and its showing signs of increase for the upcoming years. The Covid-19 pandemic is a wakeup call for all the countries to optimize its resources and invest on the health of its population. Many countries went into debt as there has been a national lockdown in most of the parts of the world. With the economy being hit and very little resources with the governments of the countries it's critical that they have a lens on the health care expenditure and use the existing resources in the best possible way.

Numerous countries with little financial reserves suffered huge deaths due to lack of hospitals and health care professionals. In the Developed world as we call it these days, we were not able to foresee a pandemic which has clouded and brought the world to a standstill. Countries with professional populations and skilled workers leveraged statistics not only to curb the spread of covid but also to make the best use of the resources as there is very little time to raise funds.On every country rests the responsibility of saving its citizens and making the most out of situation.

Many statisticians and medical professionals teamed up to tackle this crisis, some started and working rigorously on vaccine development while others to stop the transmission.

The role of statisticians and scientific modelling and the responsibility they took to stand up for the situation and serve the best was commended in many countries by saving millions of lives throughout the world. Organizations like WHO are working relentlessly for many decades to address and outline emphasis on health care and helped developing countries to ramp up their health infrastructures.

We wanted to do our part to the countries by analyzing the Heath care data form OECD partnered countries and reveal interesting trends and provide them with a blueprint of how their resources are being spent and make them diligent for the future situations and crisis. Through this project we intend to Analyze 40 years of health care data provided by the countries to OECD and engineer time series models to account for the past and plan for future funding. We also want to shed light on which metrics are affecting or consuming most of the health care and what can the governments do to intelligently to harness their resources at the right places to make the most of the effort.

We engineered models which account for the health expenditure of a set of countries and we choose to evaluate and project the expenditure based on Population, Pharma Sales, Employment of the health care industry, Life expectancy for the past four decades, Mortality, Number of hospitals, Number of nurses and government insurance. We collected data for four decades from 1980-2020.A time series regression was performed on the data of the countries and the results are tabulated accordingly. We performed a descriptive analysis initially and then engineered Time series models for the same. We also performed a multi-level model using LMER to pin point the areas to focus and generated actionable insights form the data.

Problem Definition & Significance:

The target clients of the project will be the governments or any company which has a presence in the health care industry. The other stakeholders would be print and digital media which evaluate the health of the population and audit the resources, their respective governments and outline the shortcomings. The main problem we are trying to address is providing them with a scientific way of spending of health care funds and serve best for their citizens. Numerous countries employed experts and committees to survey innovative ways to optimize the budget and come up with new campaigns to improve the health of the population. Very little progress has been made in this area because the solutions that committees came up with were terminated after the pilot phase and they never got to test them with the population to know their real effects. Our analysis outlines a couple of major problems, The first step to progress is to introspect and learn from the short comings and preparing our best for unexpected situations. After researching on the problem, we discovered that the data is available in abundance, but it’s not harnessed and used to its full potential. We want to make use of the existing data from OECD and leverage our knowledge of statistics and engineer models which would solve questions like

Population, Pharma Sales, Employment of the health care industry, Life expectancy of the 4 decades, Mortality, Number of hospitals, Number of nurses and government insurance

- Funds to be invested in the health care?
- Increase in population for the next 5 years and which campaign/Immunizations should a country come up with to grow a healthy state
- Where should we harness the existing funds to make the best use?
- Did a country do a great job increasing the life expectancy?
- Did health care industry create more jobs?
- How many hospitals are needed or to be built to meet the consulting needs and provide enough health infra to the population.
- Should we increase or decrease the Pharma production to meet the needs of people?
- Should we decrease the price of the Drugs/Medicines – this would increase the expenditure of government to provide subsidies but how would this effect the health of population?
- How is the Mortality increasing/Decreasing with the amount od dollars spent on health care?
- Did we increase the Number of health care providers such as Nurses or Physicians and does this influence health care indicators?
- Is the increase in expenditure increase/decrease due to aging population or Government provided insurance (in the US it will be Medicare and MEDIC AID)
- If there would be a situation or unexpected outbreak such as Covid 19 , we will be ready with a fundamental analysis on the resources we have and which place to develop to tackle the situation in a informed way.

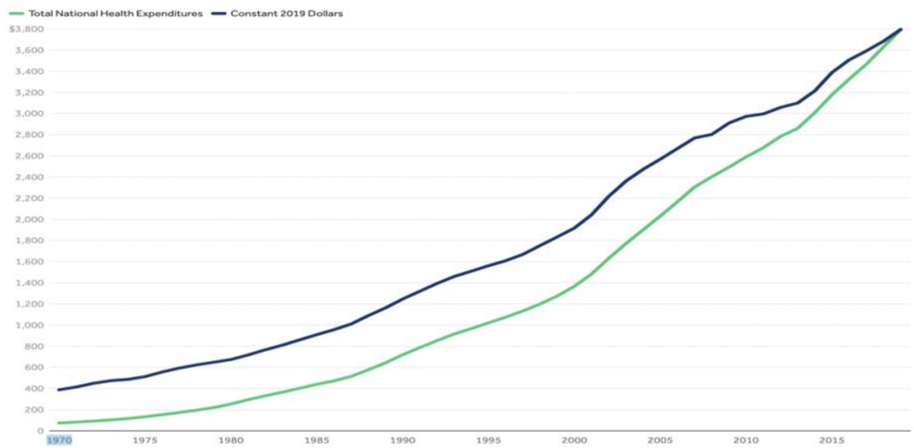
If we have granular data such as states, Hospitals, bills etc. we can perform predictive analysis to or experimental study to make the country a better health optimized country.

Prior Literature:

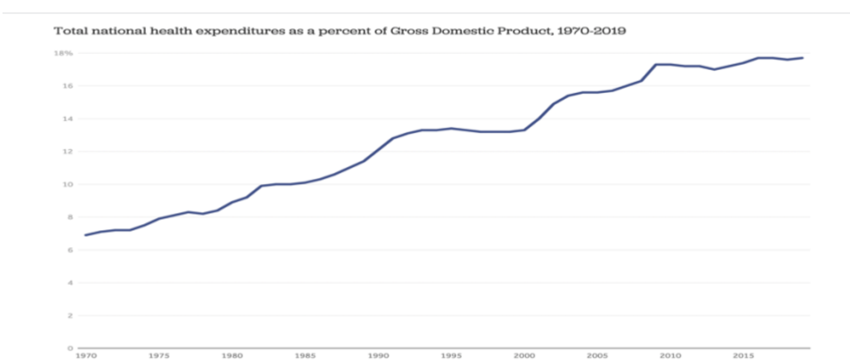
The US housing spent 8% of their income on health for Health related issues which is significant spending for the year. Health Expenditure has soared than overall spending compared to any other sector in the US. The expenditure increased 92% from 2004-2018 which is alarming, and the total expenditure has risen 41%. Even in the great recission in 2008-2009 the health expenditure rose steadily despite the decrease in consumer spending. The Revenue for the companies increased steadily as well.

The total health care expenditure was 74.1b $ in 1970.it increased to $1.5 trillion by 2000.More recently the health expenditure more than doubled in 2019 to 3.8 trillion. This includes all the health care activates in both public and private financing.17.7% of GDP is spent on health care in 2019

The irony is the expenditure on health care is increasing faster than the US economy and it showed no signs of decrease for the past decade.



Health Expenditure



Health Expenditure outplaces GDP

After knowing these facts, the alarming rate at which the Health care expenditure is increasing poses a huge challenge to the administrations and countries to deal with this seriously. Having a bird's eye view on health care expenditure enables leaders and health care decision makers to control the spending and provide the right resources and educate the population on how being in good health can save them thousands of dollars.

Diligent planning, fast action , early prevention and value optimizations are some of the key values solving this problem or having more control over the expenditure can enable a country. This makes it a valid case to solve and a game changer if applied intelligently.

Data Source/Preparation:

The data we used to engineer this project comes from **Organization for Economic Co-operation and Development** which is a multi-national organization which strives to build policies to for better lives of countries with equality, prosperity and well-being.

The data is formatted in multiples tables from different sections such as Health Expenditure, Health Care indicators, Health Status, Health Employment, Health Care resources, Demographic, Economic and Health Care utilization.

We have merged the tables with corresponding country and year and formatted them into a single sheet so that we can perform a multi-level analysis.

Country	Year	Unit	PowerCod	Expenditure_In_Millions	Employment_Head_Count	Number_of_Hospitals	LifeExp_Y	Total_deat
Australia	2014	Australian Dollar	Millions	146798.47	1478711.23	1322	82.4	159052
Canada	2014	Canadian Dollar	Millions	204447.535	1998020	720	81.8	258821
United State	2014	US Dollar	Millions	2848747.961	19576000	5627	78.9	2626418
United Kingc	2014	Pound Sterling	Millions	185376.983	3837573	1568	81.4	602781
Israel	2014	New Israeli Sheqel	Millions	79235	394799.64	83	82.2	42170

Our Dependent variable is Mortality, Pharma sales and how is it effected by the diverse selection of independent variables. The metrics were chosen in such a way that it covers diverse areas and human head counts. The data is segregated according to the country currency as exchanges rates changed significantly over the years. The scale of the variables is also taken into consideration to make the analysis robust and minimize the scale errors.

The Independent variables are:
Population, Pharma Sales, Employment of the health care industry, Life expectancy of the 4 decades, Mortality, Number of hospitals, Number of nurses and government insurance.

The reason for choosing the set of variables that they cover divergent areas of the health care and its expenditure.

Choosing the Head counts of nurses and hospitals are to account for logistical reasons. Pharma sales and government insurance are taken as they can be controlled by the government to take off the burden from the average citizen. Mortality is the classic metric for quality indicator. Number of nurses per population is a measure of the effectiveness of acute care and capability of the country to handle pandemics and outbreaks. Life expectancy is another quality indicator of the country. The idea behind selecting different variables is to increase the explaining ability and building an entire information system as every country's strength is built on the foundations of togetherness and inter-operability of multiple organizations within the country.

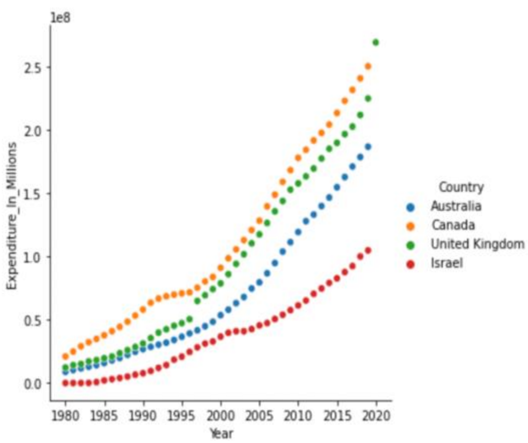
Table of Predictors:

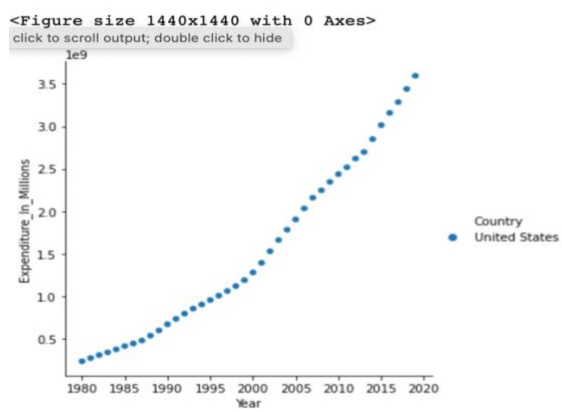
S.NO	Predictor	Rationale	Effect
1	Population	The expenditure on health depends on the % of healthy population in a country, so increase in population need not necessarily mean increase in health care expenditure.	+ve or _ve
2	Pharma Sales	Sale of drugs can be a function of 1)per-capta drug consumption. 2) Population numbers consuming drugs	+ve or -ve

3	Employment	The employment can be direct and indirect employment. Increasing the health care employment may be correlated with the expenditure in terms of salary. We can check the effectiveness of care by modelling interaction of number of employees with population.	+ve or -ve
4	Life Expectancy	The classic indicator of health for any country	+ve or -ve
5	Mortality	Health care indicator, Acquiring less mortality is the goal of any country.	+ve or -ve
6	Hospitals	The total number of hospitals in country will have a direct effect on health care expenditure.	+ve or -ve
7	Nurses	The number of nurses/nursing index. This can be interacted with population and expenditure to make a new feature to evaluate the quality of care. Number of nurses will have a direct effect on the expenditure. This can be used to answer questions like should or can we recruit more nurses in our projected budget.	+ve or _-ve
8	Govt Insurance	The thousands of persons with Govt provided health care insurance, the more people opting for insurance may reflect the general morale of the population. If people are concerned of expenditure costs they opt for insurance. This may have an effect of health expenditure.	+ve or -ve

Descriptive Analysis & Data Visualizations:

Expenditure:

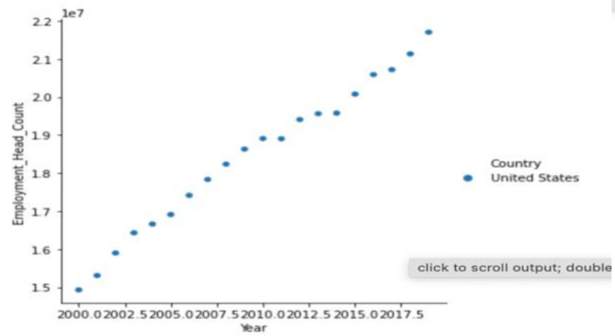


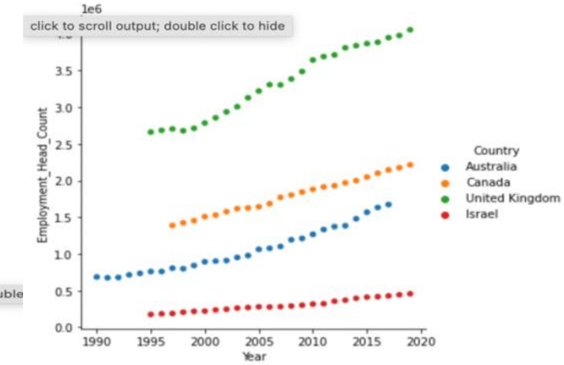


- The Expenditure in the US and Non-US countries see an increasing trend in spending on. Health care. They increased 3-fold for the US and the increase is very steep.

Health Care Employment:

<Figure size 1440x1440 with 0 Axes>

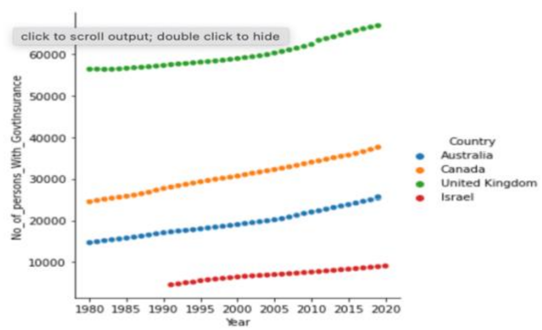


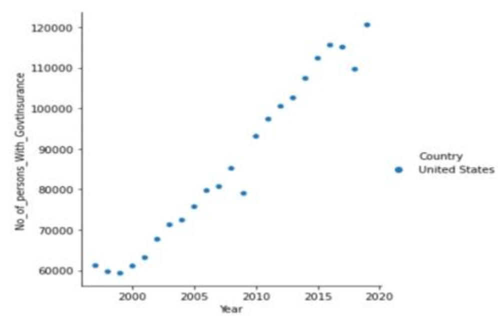


- The US has been recruiting health care professionals at a high rate, this can be attributed to the increase in health care workers and international Visas issued for the previous decade.
- The great recission 2008 is the only time there has been a loss of jobs.
- Israel is showing stability with the recruitment
- UK is adding more than 10000 workers every year for the past decade.
- Canada and Australia are moving at a slower phase

Government Insurance :

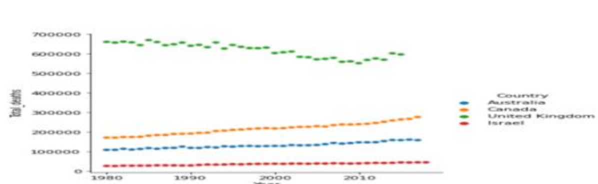
- The number of people opting for govt insurance has been increasing in all the countries, with the USA leading the table. This can also be attributed to Obama cares and Medicare/Aid insurance the Federal govt is covering for individuals over 45.
- Israel is study with the GOVT insurance because there are very little private players.
- Canada,Uk provides Free health care for all, so the increase is related to population.

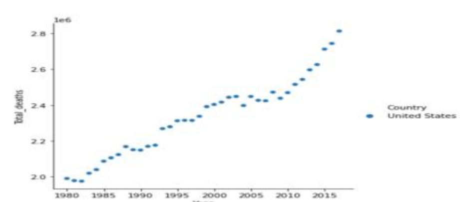




Deaths:

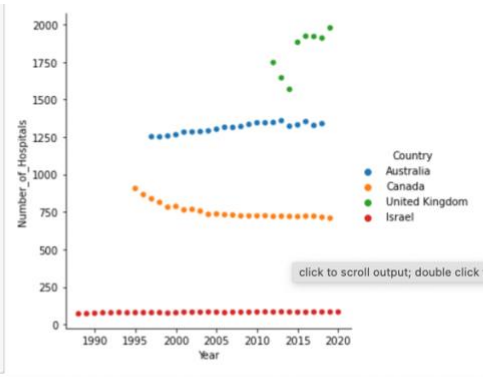
- With the Advancement in Health care and research we are hoping to see less deaths every year but conversely the deaths were increasing at an alarming rate since 1980.

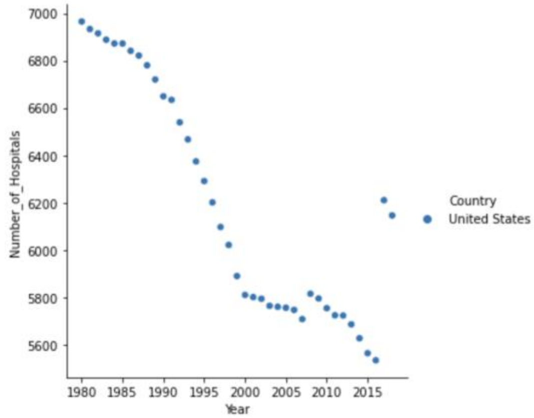




- The UK is the only country which is experiencing lesser deaths than other countries.
- The us is seeing an alarming increase in deaths and we need to control for the Death rates.

Hospitals:

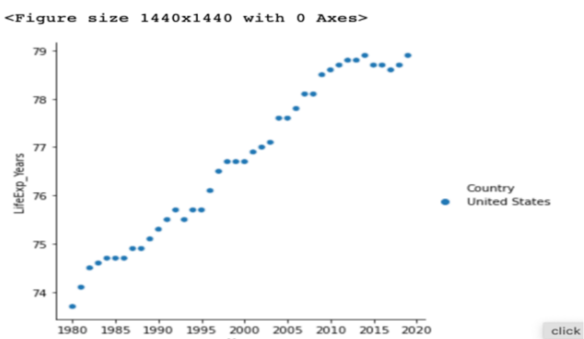


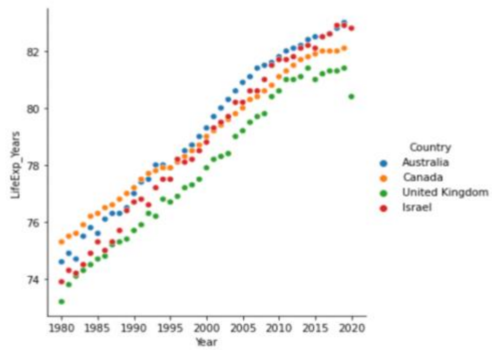


- The Us shows a decreasing trend in number of hospitals and the same trend follows in Canada.
- The UK shows an interesting trend, the hospitals decreased in 2015 and then it increased significantly.

Life expectancy:

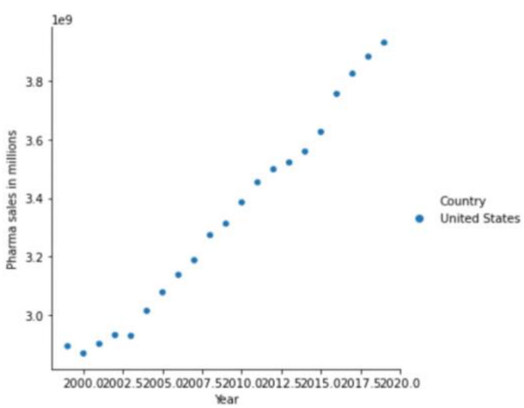
- The mean life expectancy is 73 and all the countries are showing an increase in life expectancy.

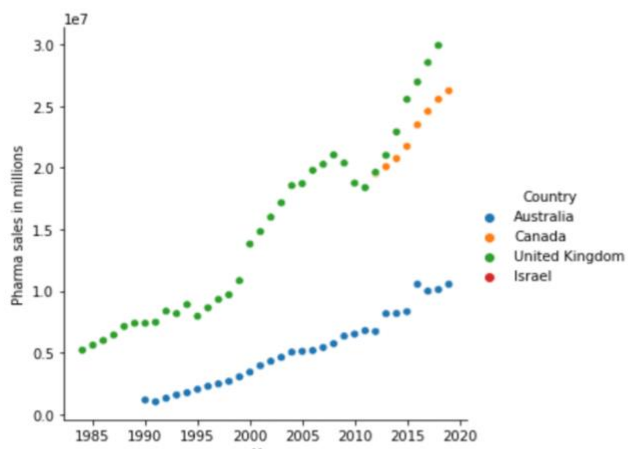




Pharma Sales :

- The amount of Drugs and medicines show an incensing trend.

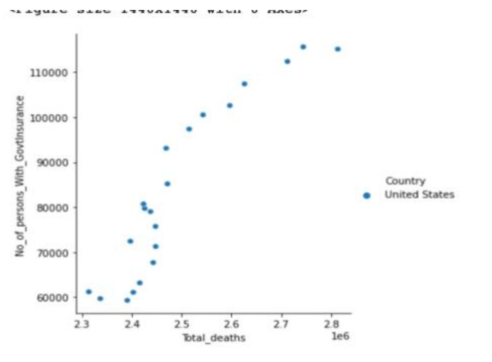


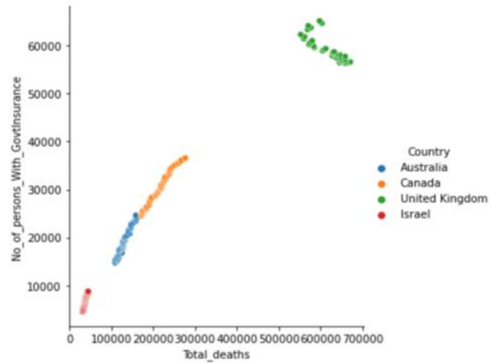


Deaths Vs Insurance:

- The people opting insurance are increasing every year in all the countries except UK.

- With the increase in pollution an diseases the spend on insurance is increasing every year.





Models:

Now we are taking our analysis in 2 parts,

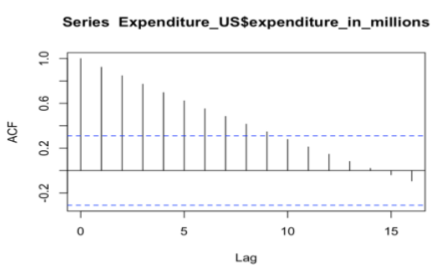
1. time series Regression to predict the future expenditure on health care and build a spending plan accordingly to make the best use of resources.
2. The second model will attribute for the yearly trends and factors attributing/effecting the expenditure Our goal is to find the metrics which are harnessing more expenditure and showing a very little effect on the heath care indicators such as mortality, patient care etc.

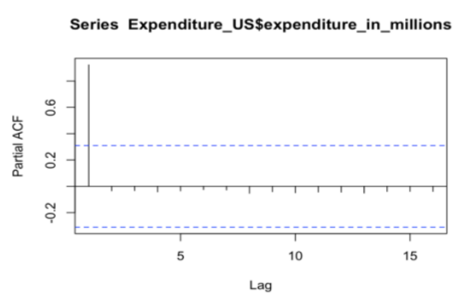
We also want to recommend or point areas which could be improved given the cost the governments invest in them and come up with auditing procedures or quality indicators.

Time Series Analysis:

Methodology:

1. We have collected data for the past 40 years for the countries we selected.
2. An auto regression model will be applied to predict the tentative Expenditure for the following years and put it forward for the governments to raise funds and put forward a budget according.





3.From the above plots, the errors are not Correlated so much, and we have the PACF 1. After researching about the Lag, we found that a lag more than 3 is not practical. So, we are taking a lag of 3.

4. To rescale and eliminate exponential numbers, these two columns are converted into Thousands.

```
Expenditure$expenditure_in_thousands=Expenditure$expenditure_in_millions*1000
Expenditure$pharma_sales_in_thousands=Expenditure$pharma_sales_in_millions*1000
```

5. Train and Test Split data : we will take the 1980-2014 for trainset and 2015-2019 as test set

```
ExpenditureTrain = Expenditure[0:35,]
```

ExpenditureTest = Expenditure[36:38,]

Interpretation:

Firstly, to observe variations of different independent features, time series model has been built.

```
> summary(ExpenditureTrainModel1With1LAg)

Call:
lm(formula = expenditure_in_thousands ~ year + expenditure_in_thousands_lag1 +
    no_of_persons_with_govtinsurance + lifeexp_years + total_deaths,
    data = ExpenditureTrain)

Residuals:
     Min       1Q   Median       3Q      Max
-1951727  -287377   -62320   487301  2115681

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -1.258e+09  7.903e+08  -1.592   0.1226
year                              6.225e+05  4.422e+05   1.408   0.1703
expenditure_in_thousands_lag1     1.119e+00  3.549e-02  31.532   <2e-16 ***
no_of_persons_with_govtinsurance -5.264e+03  1.585e+03  -3.321   0.0025 **
lifeexp_years                     1.222e+06  1.021e+06   1.196   0.2416
total_deaths                      1.159e+02  7.741e+01   1.498   0.1454
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 904200 on 28 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9995
F-statistic: 1.414e+04 on 5 and 28 DF,  p-value: < 2.2e-16
```

From the above output, it can be inferred that for increase in every year expenditure increases by 62250 thousand.
We have designed 3 time series models using Lag1,Lag1 and Lag2(model2),Lag1 and Lag2 and Lag3(Model3).

Stargazer output:

```
                                                                 Dependent variable:
                                   ------------------------------------------------------------------------------
-------------------
                                                               expenditure_in_thousands
                                              (1)                                (2)
 (3)
----------------------------------------------------------------------------------------------------------------
-------------------
year                                232,705.100** (96,830.380)        178,704.600* (102,203.800)        183,812.200
 (120,274.500)
expenditure_in_thousands_lag1            0.947*** (0.041)                  1.424*** (0.164)                  1.381*
** (0.194)
expenditure_in_thousands_lag2                                              -0.477*** (0.158)                  -0.35
2 (0.322)
expenditure_in_thousands_lag3                                                                                 -0.08
5 (0.187)
Constant                           -461,007,471.000** (192,286,887.000) -354,231,440.000* (202,967,271.000) -364,418,547.00
0 (238,912,080.000)
----------------------------------------------------------------------------------------------------------------
-------------------
Observations                                 34                                 33
 32
R2                                          0.998                              0.999
 0.999
Adjusted R2                                 0.998                              0.999
 0.999
Residual Std. Error                978,734.500 (df = 31)             882,953.200 (df = 29)              911,491.2
00 (df = 27)
F Statistic                      10,313.110*** (df = 2; 31)        8,047.608*** (df = 3; 29)          5,362.596**
* (df = 4; 27)
================================================================================================================
===================
Note:                                                                                                  *p<0.1;
 **p<0.05; ***p<0.01
>
```

In a similar way, time series models are built for other 4 countries.
We used this model to predict what expenditure going to be in upcoming years and decided the best model as per the root mean square error.

Country	2021	2022	2023	RMSE
Australia	192309.12	198534.12	204759.12	1259615
United States	3594887	3596052	3597217	41036674
Canada	259288.64	267992.64	276696.64	380222.4
UK	230715.69	236236.69	241757.69	2294325
Israel	112315	119668	127021	37029.68

Fixed/Random effects Model:
Methodology:

- We will apply fixed and random effects model to account for the variance in different sectors/metrics we selected, and we want to extract information.
- We will try to answer questions like if we increase expenditure which sector will have the most effect or which metric will increase if we increase expenditure.

We choose to perform Multi modeling with 2 Levels for mortality:
The first one would be country and the second level will be years.

Interpretation:

The Final Model is taken after comparing the AIC of different Models we engineered
Model 1:

final_Total_deaths1 <- lmer(Total_deaths ~Employment_Head_Count+Number_of_Hospitals+Total_deaths+Expenditure_In_Millions +LifeExp_Years+ Number_of_persons + (1 | Country)+(1|Year), data=my_data, REML=FALSE)

	Dependent variable:		
	Total_deaths		
	(1)	(2)	(3)
Employment_Head_Count	0.229*** (0.027)		0.280*** (0.027)
Number_of_Hospitals	50.771** (24.246)	50.213* (25.874)	
Expenditure_In_Millions	-0.001*** (0.0002)	-0.0005*** (0.0001)	-0.001*** (0.0002)
LifeExp_Years	-2,940.321** (1,254.589)	-2,393.126 (1,819.081)	
Number_of_persons	0.063 (0.383)		0.307 (0.489)
Constant	152,827.800 (126,509.800)	924,782.200 (654,870.700)	-154,297.200* (86,711.810)
Observations	97	97	97
Log Likelihood	-1,374.846	-1,383.906	-1,377.764
Akaike Inf. Crit.	2,767.692	2,781.811	2,769.528
Bayesian Inf. Crit.	2,790.864	2,799.834	2,787.551
Note:			*p<0.1; **p<0.05; ***p<0.01

- Considering this as the best model based on AIC, we can infer the Following Interpretations:
- The mortality is increasing constantly across all countries till 2016, in 2016 it decreased and in 2018 we have the least mortality. Mortality decreased almost 100000 in 2018 which is the best year with less deaths.
- The Mortality decreases with increase in Employee head count.
- Considering every variable constant, the mortality decreases with increase in Employee count.
- The Mortality decreases with increase in life expectancy, it is constant with the real-world logic.
- Conversely, the mortality or number of deaths increased with increase in hospitals. With increase in one hospital the number of deaths increase by 50.
- The mortality decreases with increase in health Expenditure. Though this decrease is not significant it can be attributed to better care and high spending on health.

- The Mortality showed increasing trend with the number of Nurses maybe we need to increase the level of care or give better training for the nurses.

The Second Level: **Country**

- Compared to an average death of 152827 in the countries, The number of deaths/Mortality is less in Australia.
- The Mortality is 1360 less in Australia than other countries.
- United Kingdom has the best mortality Rate, the mortality in UK is -21593 less compared to an average country.
- Israel stands at the top of mortality, the mortality in Israel is 15683 more compared to an average state.
- United States comes after Israel with mortality 7802 greater than average.
- Canada fares below the average mortality by 532.
- In a given country on average 105500 people are losing lives.

First Level: **Year**

- The mortality showed increasing trend till 2016 with more people dying each year.
- But in 2016 it showed a decreasing trend and then it increased again for a year.
- 2018 has the lowest number of deaths in the entire data.
- In an average year 549200 people loose their lives

Pharma Sales:

	Dependent variable:		
	Pharma_sales_in_millions (1)	Pharma_sales_in_millions/1000 (2)	Pharma_sales_in_millions (3)
Employment_Head_Count	159.848*** (7.323)		132.896*** (7.966)
Number_of_Hospitals	-25,876.120*** (5,086.525)	-7.953*** (1.945)	
Total_deaths	19.509 (17.218)	0.017*** (0.006)	-55.332*** (20.381)
LifeExp_Years	-1,073,213.000** (501,649.200)	-563.479*** (187.289)	
Number_of_persons	-36.097 (104.858)		-118.144 (134.045)
Expenditure_In_Millions		0.001*** (0.00001)	
Constant	338,209,171.000 (309,128,360.000)	561,265.600 (373,599.000)	63,576,128.000 (228,592,782.000)
Observations	97	97	97
Log Likelihood	-1,900.725	-1,123.523	-1,907.781
Akaike Inf. Crit.	3,819.451	2,263.047	3,829.562
Bayesian Inf. Crit.	3,842.623	2,283.644	3,847.585
Note:			*p<0.1; **p<0.05; ***p<0.01

final_Expenditure2 <- lmer(Pharma_sales_in_millions/1000 ~ Number_of_Hospitals+Total_deaths+LifeExp_Years+Expenditure_In_Millions+
(1 | Country)+(1|Year), data=my_data, REML=FALSE)

- The average Pharma sales for a year are 5462Million in National currency.
- If the Number of hospitals increase the pharma sale Decreases. They tentatively decrease 7000 dollars for increase in one hospital.
- The pharma sales increase with increase in mortality, which is consistent, if more people are sick more people will be dying and more medications will be used to save them.
- The Pharma sales decrease with increase in life expectancy. The sale decreases -563.5 million dollars with 1 year increase in Life expectancy.
- The pharma sales increase with increase in Health Care expectancy.

Level 1 year:

- The pharma Sales were booming in the 2000's. They kept on increasing till 2007 and it decreased -1107 in 2007.
- The sale decreased again in 2009 and then saw a steep increase.

- The pharma Sales started decreasing from 2013 and followed a decreasing trend since then.
- The pharma industry has its lowest sale in 2018 then 2019, followed by 2015.
- The highest sales for pharma industry are in 2001 and 2002.
- With the resurgence of generic medicines and good food habits by people the pharma sales are decreasing and this a good sign for the countries.

Level 2 Country:

- United Kingdom has the lowest pharma sales in our data it averages less then -582459 compared to an average country.
- Australia is trailing behind UK with an average less than -563008.
- Israel stands third lowest with 504400 compared to the average sale.
- Canada spends more than average by 62827 and is second highest in the data set.
- United States stands paramount with expensive pharma industry and health care.
- United States spend 1587 million dollars more than the average country pharma sale and consumption.

- Health campaigns needs to be established by the state, obesity is a huge problem in the US and health benefits should be inculcated to students.
- Community based public health and prevention programs needs to be organized.
- More health clinics and health care workshops need to be conducted like 2019,2018 which lowered the costs significantly.

Recommendations:

Recommendations to decrease mortality:

- Rather than increasing the number of nurses we need to invest in quality of care. More trainings need to be provided to the nurses and certifications needs to be mandated.
- More employees need to be recruited in health care such as doctors, Helpers, Researchers who can invent new vaccines and treatments for people to lower the chances of death.
- More funds need to be invested in Health care for training professionals and more dollars needs to be invested in research and building hospitals.

Recommendations on pharma and health care costs:

- United States needs to work with the pharma industry and health care professionals to check the expenditure as it starts highest in all the sectors.
- People spending a fortune are not able to stay alive in united states.
- It needs to implement social health and hygiene campaigns and mass disease vaccinations like they do in Australia and UK.
- Canada while performing good in Mortality is expensive for people who needs to consume a lot of medicine.
- Uk needs to keep up the great work with decreasing mortality and decreasing healthcare costs.

For the US:

- Need to adapt health care depending on need rather than ability to pay.
- Prescriptions need to be capped rather than leaving it to manufacturer for profit.
- Private health care should be an option too, the costs in govt health care systems needs to reduced.
- Emphasis needs to be on preventive care than cure.
- Regular checkup and preventive medicine need to be taken to lessen the expenses.
- We should decrease co-pay or deductible for public hospitals, this will decrease costs and become affordable.
- Lab tests need to be decreased which is a burden to the patient.

Quality checks:

Based on adj R2 and residual std error, we concluded the best fit, However, this could be misleading, and we may have an overfitted model. Instead of evaluating a model using the same data that we used to fit the model, we should #' evaluate the model based on an independent (test) data set.

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set

ExpenditureTrain = Expenditure_ISR[0:35,]
ExpenditureTest=Expenditure_ISR[36:38,]

Expenditure in million	2015	2016	2017
Predicted values	153596.48	162179.15	170749.38
Actual values	154904.6	162986.7	171302.8

Country	RMSE
Australia	1259615
United States	41036674
Canada	380222.4
UK	2294325
Israel	37029.68

Values from predicted and actual tend to have less error.

References:

https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_STAT
https://www.kff.org/statedata/
https://www.healthsystemtracker.org/chart-collection/u-s-life-expectancy-compare-countries/
https://www.aeaweb.org/articles?id=10.1257/jep.20.3.97
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3633404/
https://www.cdc.gov/brfss/data_documentation/index.htm
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900819/
https://www.publichealth.columbia.edu/research/comparative-health-policy-library/united-kingdom-summary
https://www.healthcarecan.ca/our-work/advocacy/research/
kff.org/other/state-indicator/health-insurance-coverage-of-the-total-population-cps/?currentTimeframe=0&sortModel=%7B"colId":"Location","sort":"asc"%7D
https://www.oecd.org/els/health-systems/Table-of-Content-Metadata-OECD-Health-Statistics-2021.pdf

Appendix: R Code.

```
#' Multi-Level Models using lme4
rm(list=ls())
library("readxl")
library(dplyr)
library(nlme)
library(lme4)
library(merTools)
library(stargazer)
#' Read data sets from nlme library
#'#' Merge the two data sets into a common data frame
#'my_data represents all the countries
```

```
my_data <- read_excel("OECD_Cleaned.xlsx",sheet="Sheet2")
#we have taken different sheets for different countries
Expenditure <- read_excel("MixedData.xlsx", sheet="AUS")
Expenditure_ISR <- read_excel("MixedData.xlsx", sheet="Israil")
Expenditure_UK <- read_excel("MixedData.xlsx", sheet="UK")
Expenditure_CND <- read_excel("MixedData.xlsx", sheet="Canada")
Expenditure_US <- read_excel("MixedData.xlsx", sheet="USA")

#code to test using MultilevelModelling
#View(my_data)
my_data <- my_data[my_data$Year>2000, ] #filtering column to take data after 2000
my_data[is.na(my_data)] = 0 #replacing na with 0
my_data$Expenditure_In_Millions <- my_data$Expenditure_In_Millions * 1000 #Scaling
the Data to thousands
my_data$Pharma_sales_in_millions <- my_data$Pharma_sales_in_millions *1000
#Scaling the Data to thousands


# 3 models with taking  total_deaths predictor as independent variable
final_Total_deaths1 <- lmer(Total_deaths ~
Employment_Head_Count+Number_of_Hospitals+Total_deaths+Expenditure_In_Million
s
                 +LifeExp_Years+  Number_of_persons+
                  (1 | Country)+(1|Year), data=my_data, REML=FALSE)

summary(final_Total_deaths1)

confint(final_Total_deaths1)
AIC(final_Total_deaths1)
fixef(final_Total_deaths1)                        # Magnitude of fixed effect
ranef(final_Total_deaths1)                        # Magnitude of random effect
coef(final_Total_deaths1)                     # Multicolliearity
library(lmtest)



final_Total_deaths2 <- lmer(Total_deaths ~
LifeExp_Years+Number_of_Hospitals+Expenditure_In_Millions+LifeExp_Years+
                  (1 | Country)+(1|Year), data=my_data, REML=FALSE)
```

```
summary(final_Total_deaths2)
confint(final_Total_deaths2)
AIC(final_Total_deaths2)
fixef(final_Total_deaths2)                              # Magnitude of fixed effect
ranef(final_Total_deaths2)                               # Magnitude of random effect
coef(final_Total_deaths2)


final_Total_deaths3 <- lmer(Total_deaths ~
Employment_Head_Count+Total_deaths+Expenditure_In_Millions
                  +  Number_of_persons+Number_of_persons+
                   (1 | Country)+(1|Year), data=my_data, REML=FALSE)

summary(final_Total_deaths3)
confint(final_Total_deaths3)
AIC(final_Total_deaths3)
fixef(final_Total_deaths3)                              # Magnitude of fixed effect
ranef(final_Total_deaths3)                               # Magnitude of random effect
coef(final_Total_deaths3)


#star gazer output for all the three models.
stargazer(final_Total_deaths1, final_Total_deaths2, final_Total_deaths3, type="text",
single.row=TRUE)


# 3 models with taking expenditure as dpepndent variable
final_Expenditure1 <- lmer(Pharma_sales_in_millions
 ~
Employment_Head_Count+Number_of_Hospitals+Total_deaths+Total_deaths+LifeExp_
Years+Number_of_persons+Number_of_persons+
                   (1 | Country)+(1|Year), data=my_data, REML=FALSE)

summary(final_Expenditure1)

confint(final_Expenditure1)
AIC(final_Expenditure1)
fixef(final_Expenditure1)                              # Magnitude of fixed effect
ranef(final_Expenditure1)                               # Magnitude of random effect
coef(final_Expenditure1)
```

```
final_Expenditure2 <- lmer(Pharma_sales_in_millions/1000 ~
 Number_of_Hospitals+Total_deaths+LifeExp_Years+Expenditure_In_Millions+
(1 | Country)+(1|Year), data=my_data, REML=FALSE)



summary(final_Expenditure2)
confint(final_Expenditure2)
AIC(final_Expenditure2)
fixef(final_Expenditure2)                        # Magnitude of fixed effect
ranef(final_Expenditure2)                         # Magnitude of random effect
coef(final_Expenditure2)



final_Expenditure3 <- lmer(Pharma_sales_in_millions
 ~ Employment_Head_Count+Total_deaths+Total_deaths
  +  Number_of_persons+Number_of_persons+(1 | Country)+(1|Year), data=my_data,
REML=FALSE)

summary(final_Expenditure3)
confint(final_Expenditure3)
AIC(final_Expenditure3)
fixef(final_Expenditure3)                        # Magnitude of fixed effect
ranef(final_Expenditure3)                         # Magnitude of random effect
coef(final_Expenditure3)                         # Magnitude of Total effect


stargazer(final_Expenditure1, final_Expenditure2, final_Expenditure3, type="text",
single.row=TRUE)



#time series model for Australia
colnames(Expenditure)=tolower(make.names(colnames(Expenditure)))
```

```
colSums(is.na(Expenditure))
Expenditure <- subset(Expenditure, Expenditure$year!=2020)
acf(Expenditure$expenditure_in_millions)
pacf(Expenditure$expenditure_in_millions)

#To rescale and eliminate exponential numbers,these two columns are converted into
Thousands.
Expenditure$expenditure_in_thousands=Expenditure$expenditure_in_millions*1000
Expenditure$pharma_sales_in_thousands=Expenditure$pharma_sales_in_millions*100
0

#The errors are not Correlated soo much and we have the PACF 1.
#After researching about the Lag we found that a lag more than 3 is not practical. So we
are taking a lag of 3
#here we are taking 20 years of data , we know it is a time series even though it looks
linear.So we need to use Time Series Techniques
n = nrow(Expenditure)#Numbr of rows
Expenditure$expenditure_in_thousands_lag1 <- c(NA,
Expenditure$expenditure_in_thousands[1:n-1])#One lag
Expenditure$expenditure_in_thousands_lag2 <- c(NA,NA,
Expenditure$expenditure_in_thousands[1:38])#Second Lag
Expenditure$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
Expenditure$expenditure_in_thousands[1:37])#Third Lag

#Now we have added the lag values to account for the Time Series/Effect of the
pervious year on current year.

#Now we will run a time series model to check for the expenditure next year.

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set
ExpenditureTrain = Expenditure[0:35,]
ExpenditureTest = Expenditure[36:38,]

colnames(ExpenditureTrain)
#nurses_counts+
ExpenditureTrainModelWith1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1+employment_hea
d_count
```

```
+no_of_persons_with_govtinsurance+number_of_hospitals+lifeexp_years+
                              total_deaths+pharma_sales_in_thousands,data = ExpenditureTrain)
summary(ExpenditureTrainModelWith1LAg)#one LAg

ExpenditureTrainModel1With1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
                             +no_of_persons_with_govtinsurance+lifeexp_years+
                              total_deaths,data = ExpenditureTrain)
summary(ExpenditureTrainModel1With1LAg)
predicted=predict(ExpenditureTrainModel1With1LAg,ExpenditureTest)
predicted


ExpenditureTrainModelWithTwoLAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+no_of_persons_with_govtinsurance+lifeexp_years+
                               total_deaths,data = ExpenditureTrain)
summary(ExpenditureTrainModelWithTwoLAg)
predicted=predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest)
predicted


ExpenditureTrainModelWithThreeLags =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+expenditure_in_thousands_lag3+no_of_persons_wit
h_govtinsurance+lifeexp_years+
                                 total_deaths,data = ExpenditureTrain)
summary(ExpenditureTrainModelWithThreeLags)
predicted=predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest)
predicted

#Lets Evaluate the model with Root means Square error.
RMSE_LAg1 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModel1With1LAg,ExpenditureTest))^2)
RMSE_LAg1
```

```
RMSE_LAg2 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest))^2)
RMSE_LAg2

RMSE_LAg3 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest))^2)
RMSE_LAg3




#time series model for USA
colnames(Expenditure_US)=tolower(make.names(colnames(Expenditure_US)))


colSums(is.na(Expenditure_US))
Expenditure_US <- subset(Expenditure_US, Expenditure_US$year!=2020)
acf(Expenditure_US$expenditure_in_millions)
pacf(Expenditure_US$expenditure_in_millions)

Expenditure_US$expenditure_in_thousands=Expenditure_US$expenditure_in_millions*
1000
#Expenditure_US$pharma_sales_in_thousands=Expenditure_US$pharma_sales_in_milli
ons*1000


n = nrow(Expenditure_US)#Numbr of rows
Expenditure_US$expenditure_in_thousands_lag1 <- c(NA,
Expenditure_US$expenditure_in_thousands[1:n-1])#One lag
Expenditure_US$expenditure_in_thousands_lag2 <- c(NA,NA,
Expenditure_US$expenditure_in_thousands[1:38])#Second Lag
Expenditure_US$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
Expenditure_US$expenditure_in_thousands[1:37])#Third Lag


#n = nrow(ExpenditureTrain)#Numbr of rows
#ExpenditureTrain$expenditure_in_thousands_lag1 <- c(NA,
ExpenditureTrain$expenditure_in_thousands[1:n-1])#One lag
#ExpenditureTrain$expenditure_in_thousands_lag2 <- c(NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:33])#Second Lag
```

```
#ExpenditureTrain$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:32])#Third Lag

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set
ExpenditureTrain = Expenditure_US[0:35,]
ExpenditureTest = Expenditure_US[36:38,]

colnames(ExpenditureTrain)
#nurses_counts+
ExpenditureTrainModelWith1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1+
                    +number_of_hospitals+lifeexp_years+
                    total_deaths,data = ExpenditureTrain)
summary(ExpenditureTrainModelWith1LAg)#one LAg


predicted=predict(ExpenditureTrainModelWith1LAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag2
ExpenditureTrainModelWithTwoLAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+no_of_persons_with_govtinsurance+lifeexp_years+
                     total_deaths+expenditure_in_thousands_lag2,data =
ExpenditureTrain)
summary(ExpenditureTrainModelWithTwoLAg)

predicted=predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag3
ExpenditureTrainModelWithThreeLags =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+no_of_persons_with_govtinsurance+lifeexp_years+

total_deaths+expenditure_in_thousands_lag2+expenditure_in_thousands_lag3,data =
ExpenditureTrain)
```

```
summary(ExpenditureTrainModelWithThreeLags)
predicted=predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest)
predicted

#Lets Evaluate the model with Root means Square error.
RMSE_LAg1 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWith1LAg,ExpenditureTest))^2)
RMSE_LAg1

RMSE_LAg2 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest))^2)
RMSE_LAg2

RMSE_LAg3 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest))^2)
RMSE_LAg3




#time series model for UK
colnames(Expenditure_UK)=tolower(make.names(colnames(Expenditure_UK)))


colSums(is.na(Expenditure_UK))
Expenditure_UK <- subset(Expenditure_UK, Expenditure_UK$year!=2020)
acf(Expenditure_UK$expenditure_in_millions)
pacf(Expenditure_UK$expenditure_in_millions)

Expenditure_UK$expenditure_in_thousands=Expenditure_UK$expenditure_in_millions*
1000
#Expenditure_UK$pharma_sales_in_thousands=Expenditure_UK$pharma_sales_in_milli
ons*1000


n = nrow(Expenditure_UK)#Numbr of rows
Expenditure_UK$expenditure_in_thousands_lag1 <- c(NA,
Expenditure_UK$expenditure_in_thousands[1:n-1])#One lag
```

```
Expenditure_UK$expenditure_in_thousands_lag2 <- c(NA,NA,
Expenditure_UK$expenditure_in_thousands[1:38])#Second Lag
Expenditure_UK$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
Expenditure_UK$expenditure_in_thousands[1:37])#Third Lag


#n = nrow(ExpenditureTrain)#Numbr of rows
#ExpenditureTrain$expenditure_in_thousands_lag1 <- c(NA,
ExpenditureTrain$expenditure_in_thousands[1:n-1])#One lag
#ExpenditureTrain$expenditure_in_thousands_lag2 <- c(NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:33])#Second Lag
#ExpenditureTrain$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:32])#Third Lag

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set
ExpenditureTrain = Expenditure_UK[0:35,]
ExpenditureTest = Expenditure_UK[36:38,]

colnames(ExpenditureTrain)
#nurses_counts+
ExpenditureTrainModelWith1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
                              ,data = ExpenditureTrain)
summary(ExpenditureTrainModelWith1LAg)#one LAg


predicted=predict(ExpenditureTrainModelWith1LAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag2
ExpenditureTrainModelWithTwoLAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
                              +expenditure_in_thousands_lag2+
                                expenditure_in_thousands_lag2,data = ExpenditureTrain)
summary(ExpenditureTrainModelWithTwoLAg)

predicted=predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest)
predicted
```

```
#expenditure_in_thousands_lag3
ExpenditureTrainModelWithThreeLags =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+expenditure_in_thousands_lag3,data =
ExpenditureTrain)
summary(ExpenditureTrainModelWithThreeLags)
year=c(2020,2021,2022,2023)
TestFrame = data.frame(year)
predicted=predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest)
predicted
x=coef(ExpenditureTrainModelWithThreeLags)

#Lets Evaluate the model with Root means Square error.
RMSE_LAg1 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWith1LAg,ExpenditureTest))^2)
RMSE_LAg1

RMSE_LAg2 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest))^2)
RMSE_LAg2

RMSE_LAg3 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest))^2)
RMSE_LAg3




#time series model for Israel
colnames(Expenditure_ISR)=tolower(make.names(colnames(Expenditure_ISR)))


colSums(is.na(Expenditure_ISR))
Expenditure_ISR <- subset(Expenditure_ISR, Expenditure_ISR$year!=2020)
acf(Expenditure_ISR$expenditure_in_millions)
pacf(Expenditure_ISR$expenditure_in_millions)

Expenditure_ISR$expenditure_in_thousands=Expenditure_ISR$expenditure_in_millions
*1000
```

```
#Expenditure_ISR$pharma_sales_in_thousands=Expenditure_ISR$pharma_sales_in_mill
ions*1000


n = nrow(Expenditure_ISR)#Numbr of rows
Expenditure_ISR$expenditure_in_thousands_lag1 <- c(NA,
Expenditure_ISR$expenditure_in_thousands[1:n-1])#One lag
Expenditure_ISR$expenditure_in_thousands_lag2 <- c(NA,NA,
Expenditure_ISR$expenditure_in_thousands[1:38])#Second Lag
Expenditure_ISR$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
Expenditure_ISR$expenditure_in_thousands[1:37])#Third Lag


#n = nrow(ExpenditureTrain)#Numbr of rows
#ExpenditureTrain$expenditure_in_thousands_lag1 <- c(NA,
ExpenditureTrain$expenditure_in_thousands[1:n-1])#One lag
#ExpenditureTrain$expenditure_in_thousands_lag2 <- c(NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:33])#Second Lag
#ExpenditureTrain$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:32])#Third Lag

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set
ExpenditureTrain = Expenditure_ISR[0:35,]
ExpenditureTest = Expenditure_ISR[36:38,]

colnames(ExpenditureTrain)
#nurses_counts+,+number_of_hospitals+lifeexp_years+total_deaths
ExpenditureTrainModelWith1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
                    ,data = ExpenditureTrain)
summary(ExpenditureTrainModelWith1LAg)#one LAg


predicted=predict(ExpenditureTrainModelWith1LAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag2
ExpenditureTrainModelWithTwoLAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
```

```
                          +expenditure_in_thousands_lag2
                           ,data = ExpenditureTrain)
summary(ExpenditureTrainModelWithTwoLAg)

predicted=predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag3
ExpenditureTrainModelWithThreeLags =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+expenditure_in_thousands_lag3,data =
ExpenditureTrain)
summary(ExpenditureTrainModelWithThreeLags)
predicted=predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest)
predicted

#Lets Evaluate the model with Root means Square error.
RMSE_LAg1 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWith1LAg,ExpenditureTest))^2)
RMSE_LAg1

RMSE_LAg2 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest))^2)
RMSE_LAg2

RMSE_LAg3 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest))^2)
RMSE_LAg3

stargazer(ExpenditureTrainModelWith1LAg, ExpenditureTrainModelWithTwoLAg,
ExpenditureTrainModelWithThreeLags, type="text", single.row=TRUE)




#time series model for Canada
colnames(Expenditure_CND)=tolower(make.names(colnames(Expenditure_CND)))
```

```
colSums(is.na(Expenditure_CND))
Expenditure_CND <- subset(Expenditure_CND, Expenditure_CND$year!=2020)
acf(Expenditure_CND$expenditure_in_millions)
pacf(Expenditure_CND$expenditure_in_millions)

Expenditure_CND$expenditure_in_thousands=Expenditure_CND$expenditure_in_millio
ns*1000
#Expenditure_CND$pharma_sales_in_thousands=Expenditure_CND$pharma_sales_in_
millions*1000


n = nrow(Expenditure_CND)#Numbr of rows
Expenditure_CND$expenditure_in_thousands_lag1 <- c(NA,
Expenditure_CND$expenditure_in_thousands[1:n-1])#One lag
Expenditure_CND$expenditure_in_thousands_lag2 <- c(NA,NA,
Expenditure_CND$expenditure_in_thousands[1:38])#Second Lag
Expenditure_CND$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
Expenditure_CND$expenditure_in_thousands[1:37])#Third Lag


#n = nrow(ExpenditureTrain)#Numbr of rows
#ExpenditureTrain$expenditure_in_thousands_lag1 <- c(NA,
ExpenditureTrain$expenditure_in_thousands[1:n-1])#One lag
#ExpenditureTrain$expenditure_in_thousands_lag2 <- c(NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:33])#Second Lag
#ExpenditureTrain$expenditure_in_thousands_lag3 <- c(NA,NA,NA,
ExpenditureTrain$expenditure_in_thousands[1:32])#Third Lag

#Train Test Split
#we will take the 1980-2014 for trainset and 2015-2019 as test set
ExpenditureTrain = Expenditure_CND[0:35,]
ExpenditureTest = Expenditure_CND[36:38,]

colnames(ExpenditureTrain)
#nurses_counts+
ExpenditureTrainModelWith1LAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1+
                          +number_of_hospitals+lifeexp_years+
                          total_deaths,data = ExpenditureTrain)
summary(ExpenditureTrainModelWith1LAg)#one LAg
```

```
predicted=predict(ExpenditureTrainModelWith1LAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag2
ExpenditureTrainModelWithTwoLAg =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1
                              +expenditure_in_thousands_lag2+
                               expenditure_in_thousands_lag2,data = ExpenditureTrain)
summary(ExpenditureTrainModelWithTwoLAg)

predicted=predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest)
predicted

#expenditure_in_thousands_lag3
ExpenditureTrainModelWithThreeLags =
lm(expenditure_in_thousands~year+expenditure_in_thousands_lag1

+expenditure_in_thousands_lag2+expenditure_in_thousands_lag3,data =
ExpenditureTrain)
summary(ExpenditureTrainModelWithThreeLags)
predicted=predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest)
predicted

#Lets Evaluate the model with Root means Square error.
RMSE_LAg1 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWith1LAg,ExpenditureTest))^2)
RMSE_LAg1

RMSE_LAg2 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithTwoLAg,ExpenditureTest))^2)
RMSE_LAg2

RMSE_LAg3 = sqrt(mean(ExpenditureTest$expenditure_in_thousands-
predict(ExpenditureTrainModelWithThreeLags,ExpenditureTest))^2)
RMSE_LAg3
```