Section 1: Business Need and Importance

Life expectancy strongly indicates a country's health, development, and impact on its citizens' well-being. Additionally, life expectancy can highlight which countries may need help from outside world development organizations such as the World Health Organization or UNICEF. Collecting this data on life expectancy annually is vital to this business, who can utilize the data analysis on life expectancy and its factors to understand better where to place their resource best, advocate for policy changes, and intervene with targeted programs. However, to truly make an impact all data must be collected annually, consistently, or at least weak analysis may be emerged. A public dataset from the World Bank analyzes different factors that impact the life expectancy of 174 countries. In this analysis, businesses can fully understand the life expectancy of different countries and the variables that impact life expectancy, such as GDP healthcare expenditure, the percentage of people living in poverty, and government corruption rates[i]. Utilizing cluster analysis of life expectancy, businesses can get a comprehensive understanding of countries that need specific assistance, such as Doctors without Borders, compared to the World Food Project. With more insights into the variables that impact life expectancy, organizations can put together specific plans to increase life expectancy. Moreover, forecasting analysis will allow businesses to see the impact of their program. Annual life expectancy data allows world organizations to make long-lasting impacts on health crises and programs for all countries.

Section 2: Statistical Methodology

The first unsupervised data mining technique used was agglomerative clustering. In the dataset, each country is given an associated life expectancy and year. Agglomerative clustering was chosen because it can handle mixed datasets and small datasets, along with the ability to handle outliers well. After analyzing the dataset, the year 2013 has the most countries[ii] (23) to analyze life expectancy. To get the best data, all rows with null values are excluded, and the columns "Year" and "Country Code" are removed. The columns "Region" and "Income Group" are also changed into factor variables to ensure the best analysis. The region has six factors: "East Asia & Pacific," "Europe & Central Asia," "Latin America & Caribbean," "Middle East & North Africa," "South Asia," and "And Sub-Saharan Africa." The Income Group has three factors: "Low income," "Lower middle income," and "Upper middle income." Next, the dissimilarity matrix is calculated using the "daisy" function of the R package "cluster." The "Gower" metric[iii] is utilized due to the mixed variable types, which allow each variable to be analyzed appropriately. The dissimilar matrix[iv] compares each instance (i.e., country) against each other, and when the dissimilarity coefficient is close to 0, the countries are like each other. On the other hand, the closer to 1, the less similar the countries are to each other. The lowest coefficient produced is 0.077, and the highest is 0.648. With the "cluster" package, agglomerative clustering can then be performed and evaluated.

The second method deployed was time series cross-validation forecasting for life expectancy. Time series cross-validation forecasting was chosen because of its sensitivity to new data and for keeping it in timeline order. To do this forecasting, two countries were selected, Bangladesh and Tanzania, specifically because Bangladesh has missing data inputs (2010[v], 2014[vi], 2015[vii], 2017[viii], and 2018[ix]) and Tanzania has all inputs from 2006-2019. Visualizations were created to understand better the two countries' most updated life expectancy data[x]. Then, all frequencies are set to one for the three-time series objects because the data is collected annually. The model was then split into training and validation sets for each country: 66% training (6 years) and 33% (3 years) validation for Bangladesh, and 78% training (11 years) and 21% validation for Tanzania (3 years). With the object, three-time series models were created for each country: linear, quadric, and cubic. Then, using the forecasting package, each model was evaluated for the accuracy of each model and looked at both the training and test sets using metrics such as MAE, RMSE, MAPE, and MSE. Finally, using the same package, forecasted life expectancy is project based on the best time series model and the visual models, to further understand the evaluation.

Section 3: Results and Interpretation

The first unsupervised method, agglomerative clustering, was performed utilizing the "Ward" method, which produced an agglomerative coefficient of 0.8592044. This indicates that the cluster structure is strong and well-defined, with different clusters. The result of the clustering also provided a summary of the cluster heights, which represent the dissimilarity between objects. The minimum height was 0.06764, meaning there were some similar objects, and the maximum was 0.92846, meaning the last merge was two highly different clusters[xi]. Based on the banner plot[xii] and dendrogram plots[xiii] produced with the clustering, five distinct clusters can be formed. Each cluster is individually analyzed after cutting the tree and combining it with the original data. Cluster 1[xiv] has the highest average life expectancy (73.23). Notably, this cluster consists of two counties in Europe and Central Asia, both in upper-middle income groups. Additionally, they have CO2 emissions ranging between 550 and 7840 kT and high expenditure for education and health compared to other clusters[xv]. Comparatively, cluster 2[xvi] contains seven countries with the second-highest mean life expectancy at 69.87. All these countries are located between "East Asia & Pacific" and "Latin America & Caribbean" and are low-middle income. Additionally, it has extreme C02 emissions, an average of 18,140 kT. Additionally, this cluster has high undernourishment and unemployment rates, the averages being 10.53% and 8.667%, respectively[xvii]. Next, Cluster 3[xviii], consisting of 8 countries from Sub-Saharan Africa, has the lowest life expectancy at 57.23. In this cluster, there are incredibly high DALYs[xix] of communicable disease, average is 8,619,877 years and the lowest health and education expenditures compared to the other clusters, the averages being 6.111% and 3.409%[xx]. Cluster 4[xxi] has the second lowest average life expectancy at 61.35 when looking at the five countries from Sub-Saharan Africa. While many variables vary significantly, the cluster features exceptionally high rates of unemployment, averaging 11.188%, and the highest corruption level (3.0-4.0) in the dataset[xxii]. Finally, cluster 5 features a singular outlier country, India, with a life expectancy of 67.93 years. One of the defining features of this cluster is its high DALYs of injuries and communicable and noncommunicable diseases (53,868,775 years, 187,804,149 years, and 242,690,489 years, respectively)[xxiii]. Along with low health and education expenditure and a sanitation rate of 31.55%, India should have a lower life expectancy, like clusters 3 and 4. However, its moderately high life expectancy makes it an outlier in this dataset. Overall, clusters 1 and 2 have better economic status, health statistics, and government infrastructure than clusters 3 and 4, which lead to a higher life expectancy.

The second method utilized was time series cross-validation forecasting, specifically for the countries Bangladesh and Tanzania. First, the three models (linear, quadric, and cubic) for each country must be evaluated to determine which is best used to forecast. Thus, there must be accuracy evaluations on all models. Both countries' training models and accuracy functions demonstrate that there is minimal bias for all six models. For Bangladesh, the linear model has a RMSE of 3.6045 and an MAE of 3.1681, which does not format the model well[xxiv]. Then, a quadric model improved the model slightly with better fit and accuracy, demonstrated by the RMSE of 3.4770 and MAE of 3.0697[xxv]. Finally, the cubic model has the best fit with RMSE of 3.2601 and MAE of 2.4710 on the validation set[xxvi]. Similarly, the accuracy statistics further prove the cubic model is the best fit with Theil's U statistic (1.950905)[xxvii] compared to 2.089464 and 2.166977. For Tanzania, the same process was performed. Based on the three models, the cubic model performed the best with a RMSE of 0.1842 and MAE of 0.1663[xxviii], compared to RMSE = 0.5203 and MAE = 0.4554 for linear[xxix], and RMSE = 0.7556, MAE = 0.6855 for quadric[xxx]. However, unlike Bangladesh, the quadric model has the worst accuracy with a Theil's U statistic of 1.5235, compared to the cubic model, which has a Theil U's statistic of 0.3728681.

In the 2029, the 10-year prediction, Bangladesh's life expectancy will reach 87.72289[xxxi] and Tanzania will reach 66.34186[xxxii]. Immediately, Bangladesh seems to have a better model because it has a higher life expectancy, but this would be incorrect because the result is volatile. With all accuracy metrics, Tanzania performs better. Additionally, analyzing the confidence interval, it becomes clear that Bangladesh's forecast shows an extremely weak performance. In 2029, the 95% confidence intervals range between 30.12055 to 145.32524 years, a 115.20469 difference. Comparatively, the Tanzania model has a 95% confidence interval range of 65.05265 to 67.63107. This is shown through projection graphs[xxxiii] where the confidence graph in blue is much wider, as well.

Section 4: Alternative Approaches

Divisive analysis hierarchal cluster was not utilized because it is sensitive to outliers, does not handle mixed data, and is worse for smaller datasets compared to agglomerative clustering where there were clear outliers, combined categorical and numerical data, and a small dataset of 23 countries. At the beginning of the analysis, the number of clusters was unknown; thus, agglomerative clustering was superior to K-Means clustering, where you must specify the number of clusters at the beginning. Additionally, Holt-Winters forecasting was not used because there is no seasonality in life expectancy data and limited the use of historical data, unlike time cross-validation forecasting. Similarly, smoothing forecasting techniques often look at the most recent data, putting less emphasis on historical data, which makes time series cross-validation superior. Finally, the linear trend with seasonality would not work for this dataset because it is not linear, nor is there seasonality, which both are incorporated in the utilized forecasting method.

Section 5: Conclusions

Organizations such as UNICEF or the World Health Organization look at life expectancy and its trajectory as a great indicator of how a country is developing, along with variables that impact life expectancy, such as CO_2 emissions. Utilizing cluster analysis, these organizations can group countries based on specific variables, which can help identify the needs of these countries, such as creating legislation for a cleaner environment or allocating vaccinations to lower the communicable diseases rate. As demonstrated through clusters 3 and 4, implement policies that increase spending on healthcare and education. With all this information, organizations such as the World Health Organization must be consistent with gathering each year, as demonstrated by Bangladesh and Tanzania's life expectancy forecasts, to make a lasting impact. For all accuracy measurements, having consistent data leads to a better forecast for countries' life expectancy. This model can also be used for any variable that contributes to life expectancy. Having accurate and updated models will allow these organizations to make better decisions on where to send doctors, respond to crises, and track the progress of their programs in each country. The impact of these models will implore world organizations to keep consistent annual life expectancy data for all countries. This business can create tailored programs, policies, and health initiatives with accurate analysis and predictions from cluster analysis and forecasting.

Appendix

---

i
```
 [1] "Country.Name"                    "Region"
 [3] "IncomeGroup"                     "Life.Expectancy.World.Bank"
 [5] "Prevelance.of.Undernourishment"  "CO2"
 [7] "Health.Expenditure.."            "Education.Expenditure.."
 [9] "Unemployment"                    "Corruption"
[11] "Sanitation"                      "Injuries"
[13] "Communicable"                    "NonCommunicable"
```

[ii] Please see Appendix B—CSV file called "20143Data".

[iii] The "Manhattan" and "Euclidean" distances were also analyzed, but the "Gower" metric was ultimately chosen because it can handle mixed variables compares to the distance matrix can only handle numerical data, which would exclude the country name and region.

iv
```
> d=daisy(data[,2:14], metric="gower")
> d
Dissimilarities :
            1          2          3          4          5          6          7          8
2  0.39742711
3  0.36940861 0.26357756
4  0.48590024 0.34403041 0.42259733
5  0.44932866 0.30130081 0.31613469 0.18457698
6  0.07701697 0.34759586 0.32470473 0.43645658 0.39322645
7  0.39574920 0.28122767 0.24858854 0.28547751 0.18295592 0.34505239
8  0.30574558 0.25409663 0.10275889 0.42014671 0.33391873 0.29270496 0.27498252
9  0.64817071 0.37150866 0.48062834 0.65418867 0.53731888 0.59206849 0.48206245 0.51163439
10 0.39351345 0.41150407 0.26826975 0.35109888 0.36095406 0.34566792 0.26304644 0.29158943
11 0.49307171 0.30142813 0.40923752 0.13061004 0.13404361 0.44342299 0.28244840 0.42712516
12 0.38794637 0.31309005 0.36575379 0.14188183 0.10814494 0.34949030 0.17032327 0.35319794
13 0.35741532 0.18405543 0.28244323 0.37880297 0.33105574 0.36964057 0.28155841 0.27944797
14 0.32874067 0.22197946 0.14908983 0.39633207 0.28698891 0.29222500 0.19612777 0.17671112
15 0.39282607 0.34562258 0.34724983 0.19069887 0.16570566 0.37897484 0.24968068 0.28774539
16 0.33378583 0.09839329 0.20427744 0.34361313 0.26938898 0.28239393 0.20240829 0.21172617
17 0.36844454 0.24960358 0.18560673 0.28546645 0.18930278 0.31234233 0.08327758 0.21211061
18 0.40471888 0.38143440 0.41690362 0.12691752 0.17968815 0.39714235 0.28708258 0.37154552
19 0.30682637 0.28255924 0.18948466 0.34394765 0.25879320 0.25236646 0.15242033 0.17610070
20 0.50187181 0.33373803 0.41272133 0.09581629 0.13564495 0.44690680 0.27019366 0.43067081
21 0.45764010 0.31090379 0.32786840 0.11999849 0.06763787 0.40186801 0.19548444 0.34590821
22 0.41689834 0.20795909 0.24012580 0.20368455 0.14572857 0.36079612 0.12017045 0.25480376
23 0.27870822 0.28880400 0.20723762 0.45378679 0.36862610 0.26647021 0.19759626 0.18882772
```

```
            9         10         11         12         13         14         15         16
2
3
4
5
6
7
8
9
10 0.61282633
11 0.63600034 0.45140244
12 0.56723318 0.34505726 0.17480974
13 0.52401387 0.43815937 0.31757542 0.31360420
14 0.43314958 0.31432849 0.37993824 0.30369393 0.14713746
15 0.62338043 0.28080385 0.22434530 0.16064721 0.37956354 0.33793452
16 0.37139212 0.35933568 0.32475913 0.24724699 0.18418364 0.16423158 0.33608071
17 0.45043835 0.20733582 0.28224415 0.19480967 0.27806978 0.15104221 0.21877728 0.19254148
18 0.66588175 0.29763950 0.18860233 0.15119571 0.41620695 0.40125819 0.12663742 0.34942314
19 0.48802279 0.15257831 0.35133831 0.26766865 0.33146516 0.18470056 0.21874167 0.22487711
20 0.64467921 0.38014410 0.07249988 0.12999317 0.34746398 0.38342707 0.19890580 0.32921482
21 0.59549389 0.33521851 0.11656447 0.11129797 0.33056479 0.31261774 0.11668060 0.28828574
22 0.46660332 0.25965807 0.20494397 0.16086507 0.23982675 0.21766645 0.22490863 0.16276335
23 0.50788021 0.33263894 0.46117121 0.33924791 0.20074536 0.08961655 0.37400481 0.21540845
```

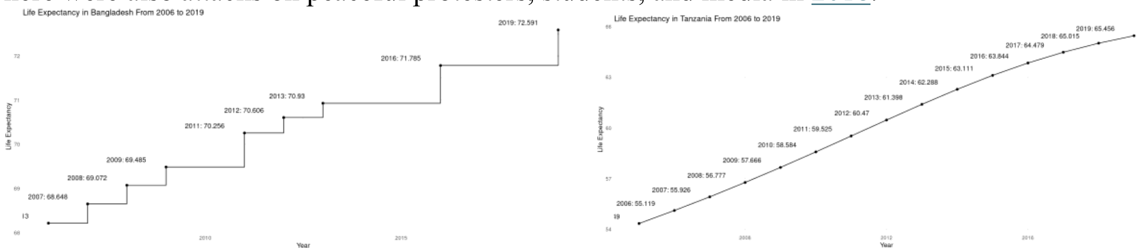
```
           17         18         19         20         21         22
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18 0.29676224
19 0.07323262 0.30396890
20 0.28572796 0.12982813 0.35476440
21 0.21095465 0.15264679 0.26879782 0.08929040
22 0.11573733 0.20947760 0.17967822 0.18807383 0.15551525
23 0.20627355 0.43945017 0.20132271 0.46459730 0.37863071 0.28951112

Metric :  mixed ;  Types = N, N, I, I, I, I, I, I, I, I, I, I, I
Number of objects : 23
```

[v] In 2010, in Bangladesh, there were several acts of impunity by security forces who engaged in killings, acts of torture, and illegal detentions. Additionally, according to the Human Rights Report, "military and police regularly employ torture and cruel, inhuman, or degrading punishment against detainees, despite constitutional guarantees against torture and Bangladesh's ratification of the United Nations Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment. The government failed to investigate the causes of numerous deaths in custody, and there was little action to hold accountable those responsible for the deaths of alleged mutineers from the Bangladesh Rifles border force." Moreover, there were serval cases of border killings.

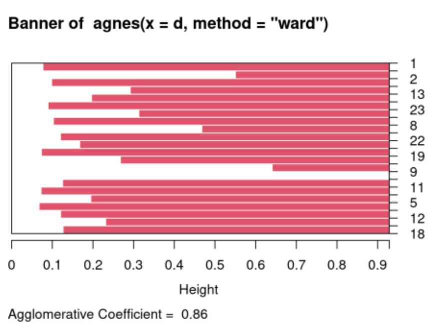
[vi] In 2014, the general election resulted in violence, including killing people, arresting opposition parties, destruction of personal property, and burning of polling places.

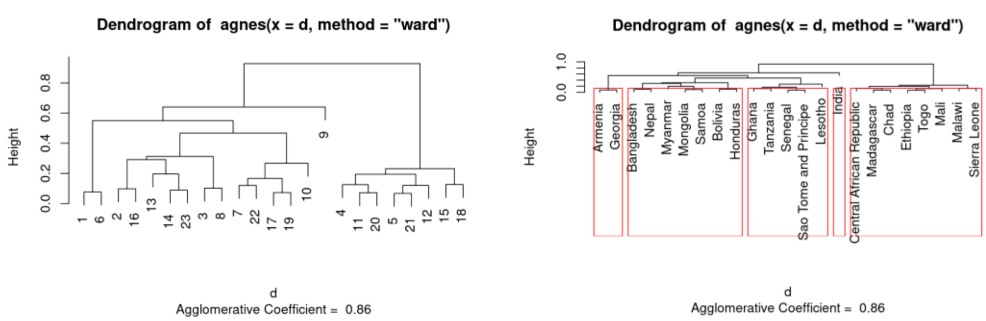
vii There was a Bangladeshi political crisis in 2015, which resulted in the death of 75 people. The Prime Minster, who belonged to the Awani League political party, declared the Bangladesh National Party, the most significant political party, terrorists.
viii In 2017, monsoon rain and river flooding affected over a third of the country. This had lasting impacts on livestock, crops, and water sources, and 40,000 people were displaced from their homes.
ix Once again, an unfair general election was held where many political leaders were detained or killed. There were also attacks on peaceful protesters, students, and media in 2018.

x 


```
> aResult <- agnes(d, method = "ward")
> aResult
Call:    agnes(x = d, method = "ward")
Agglomerative coefficient:  0.8592044
Order of objects:
 [1]  1  6  2 16 13 14 23  3  8  7 22 17 19 10  9  4 11 20  5 21 12 15 18
Height (summary):
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.06764 0.09948 0.14702 0.24196 0.28564 0.92846

Available components:
[1] "order"  "height" "ac"     "merge"  "diss"   "call"   "method"
```
xi

xii 

xiii 

xiv Armenia and Georgia

[xv]

```
> summary(subset(myData, aClusters == 1))
 Country.Name                                 Region                  IncomeGroup
 Length:2           East Asia & Pacific        :0   Low income          :0
 Class :character   Europe & Central Asia      :2   Lower middle income:0
 Mode  :character   Latin America & Caribbean :0   Upper middle income:2
                    Middle East & North Africa:0
                    South Asia                 :0
                    Sub-Saharan Africa         :0
 Life.Expectancy.World.Bank Prevelance.of.Undernourishment      CO2        Health.Expenditure..
 Min.   :72.41              Min.   :3.200                   Min.   :5500   Min.   : 7.883
 1st Qu.:72.82              1st Qu.:4.375                   1st Qu.:6085   1st Qu.: 8.493
 Median :73.23              Median :5.550                   Median :6670   Median : 9.103
 Mean   :73.23              Mean   :5.550                   Mean   :6670   Mean   : 9.103
 3rd Qu.:73.64              3rd Qu.:6.725                   3rd Qu.:7255   3rd Qu.: 9.713
 Max.   :74.06              Max.   :7.900                   Max.   :7840   Max.   :10.323
 Education.Expenditure..  Unemployment     Corruption    Sanitation       Injuries
 Min.   :2.650           Min.   :16.18   Min.   :3.5   Min.   :38.69   Min.   : 82888
 1st Qu.:2.709           1st Qu.:16.99   1st Qu.:3.5   1st Qu.:43.63   1st Qu.: 98712
 Median :2.768           Median :17.80   Median :3.5   Median :48.58   Median :114535
 Mean   :2.768           Mean   :17.80   Mean   :3.5   Mean   :48.58   Mean   :114535
 3rd Qu.:2.828           3rd Qu.:18.61   3rd Qu.:3.5   3rd Qu.:53.52   3rd Qu.:130358
 Max.   :2.887           Max.   :19.42   Max.   :3.5   Max.   :58.46   Max.   :146182
  Communicable     NonCommunicable     aClusters
 Min.   : 79985   Min.   : 769112   Min.   :1
 1st Qu.: 89386   1st Qu.: 891590   1st Qu.:1
 Median : 98786   Median :1014069   Median :1
 Mean   : 98786   Mean   :1014069   Mean   :1
 3rd Qu.:108187   3rd Qu.:1136548   3rd Qu.:1
 Max.   :117588   Max.   :1259027   Max.   :1
```

[xvi] Bangladesh, Bolivia, Honduras, Myanmar, Mongolia, Nepal, and Samoa

```
> summary(subset(myData, aClusters == 2))
 Country.Name                                 Region                  IncomeGroup
 Length:7           East Asia & Pacific        :3   Low income          :0
 Class :character   Europe & Central Asia      :0   Lower middle income:7
 Mode  :character   Latin America & Caribbean :2   Upper middle income:0
                    Middle East & North Africa:0
                    South Asia                 :2
                    Sub-Saharan Africa         :0
 Life.Expectancy.World.Bank Prevelance.of.Undernourishment      CO2         Health.Expenditure..
 Min.   :64.92              Min.   : 3.70                   Min.   :  200   Min.   :2.654
 1st Qu.:68.69              1st Qu.: 6.45                   1st Qu.: 7730   1st Qu.:3.633
 Median :69.47              Median :11.40                   Median :13600   Median :4.630
 Mean   :69.87              Mean   :10.53                   Mean   :18140   Mean   :5.008
 3rd Qu.:71.64              3rd Qu.:14.60                   3rd Qu.:18560   3rd Qu.:6.012
 Max.   :74.05              Max.   :16.50                   Max.   :60600   Max.   :8.477
 Education.Expenditure..  Unemployment     Corruption      Sanitation       Injuries
 Min.   :1.966           Min.   :0.747   Min.   :2.500   Min.   :31.58   Min.   :   6312
 1st Qu.:2.559           1st Qu.:2.495   1st Qu.:2.750   1st Qu.:37.78   1st Qu.: 282319
 Median :4.670           Median :4.230   Median :3.000   Median :45.67   Median : 613162
 Mean   :4.307           Mean   :4.102   Mean   :3.143   Mean   :44.11   Mean   :1068403
 3rd Qu.:5.408           3rd Qu.:5.040   3rd Qu.:3.500   3rd Qu.:47.49   3rd Qu.:1396037
 Max.   :7.580           Max.   :8.667   Max.   :4.000   Max.   :60.98   Max.   :3502633
  Communicable       NonCommunicable      aClusters
 Min.   :    8715   Min.   :   40239   Min.   :2
 1st Qu.:  343537   1st Qu.: 1106245   1st Qu.:2
 Median : 1188492   Median : 2000514   Median :2
 Mean   : 4025533   Mean   : 6560924   Mean   :2
 3rd Qu.: 5389977   3rd Qu.: 8497462   3rd Qu.:2
 Max.   :15514496   Max.   :24678301   Max.   :2
```

[xvii]

[xviii] Central African Republic, Ethiopia, Madagascar, Mali, Malawi, Sierra Leone, Chad, Togo

[xix] Injuries, Communicable and Non-Communicable diseases are all counted in DALYs, which represents the loss of one year of full health.

```
> summary(subset(myData, aClusters == 3))
 Country.Name                                    Region              IncomeGroup
 Length:8           East Asia & Pacific        :0   Low income          :8
 Class :character   Europe & Central Asia      :0   Lower middle income:0
 Mode  :character   Latin America & Caribbean :0   Upper middle income:0
                    Middle East & North Africa:0
                    South Asia                 :0
                    Sub-Saharan Africa         :8
 Life.Expectancy.World.Bank Prevelance.of.Undernourishment      CO2         Health.Expenditure..
 Min.   :49.37             Min.   : 3.70                   Min.   :  120   Min.   : 3.739
 1st Qu.:52.22             1st Qu.:16.98                   1st Qu.: 1025   1st Qu.: 4.042
 Median :57.87             Median :22.40                   Median : 1970   Median : 4.509
 Mean   :57.23             Mean   :22.48                   Mean   : 2789   Mean   : 6.111
 3rd Qu.:60.94             3rd Qu.:27.45                   3rd Qu.: 3002   3rd Qu.: 6.606
 Max.   :64.71             Max.   :40.20                   Max.   :10080   Max.   :11.579
 Education.Expenditure..  Unemployment     Corruption      Sanitation        Injuries
 Min.   :1.777           Min.   :0.995   Min.   :2.000   Min.   : 5.225   Min.   : 229444
 1st Qu.:2.574           1st Qu.:1.824   1st Qu.:2.375   1st Qu.: 8.038   1st Qu.: 373966
 Median :3.068           Median :3.380   Median :2.500   Median :10.661   Median : 546676
 Mean   :3.409           Mean   :3.586   Mean   :2.562   Mean   :12.004   Mean   : 767644
 3rd Qu.:4.442           3rd Qu.:5.625   3rd Qu.:3.000   3rd Qu.:15.246   3rd Qu.: 687662
 Max.   :5.416           Max.   :6.350   Max.   :3.000   Max.   :22.444   Max.   :2729927
  Communicable       NonCommunicable       aClusters
 Min.   : 2568922   Min.   :  935570   Min.   :3
 1st Qu.: 3541632   1st Qu.: 1196396   1st Qu.:3
 Median : 6271542   Median : 2158513   Median :3
 Mean   : 8619877   Mean   : 3174509   Mean   :3
 3rd Qu.: 8910272   3rd Qu.: 2977939   3rd Qu.:3
 Max.   :27856068   Max.   :11553683   Max.   :3
```
xx

xxi Ghana, Tanzania, Senegal, Sao Tome and Principe, Lesotho

```
> summary(subset(myData, aClusters == 4))
 Country.Name                                    Region              IncomeGroup
 Length:5           East Asia & Pacific        :0   Low income          :0
 Class :character   Europe & Central Asia      :0   Lower middle income:5
 Mode  :character   Latin America & Caribbean :0   Upper middle income:0
                    Middle East & North Africa:0
                    South Asia                 :0
                    Sub-Saharan Africa         :5
 Life.Expectancy.World.Bank Prevelance.of.Undernourishment      CO2         Health.Expenditure..
 Min.   :48.66             Min.   : 6.6                    Min.   :  120   Min.   :4.386
 1st Qu.:61.40             1st Qu.:12.0                    1st Qu.:  610   1st Qu.:4.623
 Median :62.06             Median :12.9                    Median : 8010   Median :4.646
 Mean   :61.35             Mean   :14.7                    Mean   : 7036   Mean   :6.228
 3rd Qu.:65.94             3rd Qu.:19.3                    3rd Qu.:10810   3rd Qu.:7.937
 Max.   :68.69             Max.   :22.7                    Max.   :15630   Max.   :9.547
 Education.Expenditure..  Unemployment      Corruption     Sanitation       Injuries
 Min.   :3.358           Min.   : 2.930   Min.   :3.0   Min.   :10.06   Min.   :   3762
 1st Qu.:4.576           1st Qu.: 6.203   1st Qu.:3.5   1st Qu.:19.63   1st Qu.: 168864
 Median :5.687           Median : 8.578   Median :3.5   Median :20.02   Median : 305516
 Mean   :5.371           Mean   :11.188   Mean   :3.5   Mean   :22.21   Mean   : 536514
 3rd Qu.:5.940           3rd Qu.:13.649   3rd Qu.:3.5   3rd Qu.:26.82   3rd Qu.: 785815
 Max.   :7.292           Max.   :24.580   Max.   :4.0   Max.   :34.52   Max.   :1418615
  Communicable       NonCommunicable     aClusters
 Min.   :   23237   Min.   :  28865   Min.   :4
 1st Qu.: 1252935   1st Qu.: 439624   1st Qu.:4
 Median : 3473915   Median :1943532   Median :4
 Mean   : 5675585   Mean   :2878823   Mean   :4
 3rd Qu.: 7640207   3rd Qu.:4330459   3rd Qu.:4
 Max.   :15987630   Max.   :7651635   Max.   :4
```
xxii

```
> summary(subset(myData, aClusters == 5))
 Country.Name                                    Region              IncomeGroup
 Length:1           East Asia & Pacific        :0   Low income          :0
 Class :character   Europe & Central Asia      :0   Lower middle income:1
 Mode  :character   Latin America & Caribbean :0   Upper middle income:0
                    Middle East & North Africa:0
                    South Asia                 :1
                    Sub-Saharan Africa         :0
 Life.Expectancy.World.Bank Prevelance.of.Undernourishment      CO2
 Min.   :67.93             Min.   :14.9                   Min.   :1972430
 1st Qu.:67.93             1st Qu.:14.9                   1st Qu.:1972430
 Median :67.93             Median :14.9                   Median :1972430
 Mean   :67.93             Mean   :14.9                   Mean   :1972430
 3rd Qu.:67.93             3rd Qu.:14.9                   3rd Qu.:1972430
 Max.   :67.93             Max.   :14.9                   Max.   :1972430
 Health.Expenditure.. Education.Expenditure..  Unemployment     Corruption     Sanitation
 Min.   :3.749        Min.   :3.845           Min.   :5.424   Min.   :3.5   Min.   :31.55
 1st Qu.:3.749        1st Qu.:3.845           1st Qu.:5.424   1st Qu.:3.5   1st Qu.:31.55
 Median :3.749        Median :3.845           Median :5.424   Median :3.5   Median :31.55
 Mean   :3.749        Mean   :3.845           Mean   :5.424   Mean   :3.5   Mean   :31.55
 3rd Qu.:3.749        3rd Qu.:3.845           3rd Qu.:5.424   3rd Qu.:3.5   3rd Qu.:31.55
 Max.   :3.749        Max.   :3.845           Max.   :5.424   Max.   :3.5   Max.   :31.55
    Injuries          Communicable         NonCommunicable        aClusters
 Min.   :53868775   Min.   :187804149   Min.   :242690489   Min.   :5
 1st Qu.:53868775   1st Qu.:187804149   1st Qu.:242690489   1st Qu.:5
 Median :53868775   Median :187804149   Median :242690489   Median :5
 Mean   :53868775   Mean   :187804149   Mean   :242690489   Mean   :5
 3rd Qu.:53868775   3rd Qu.:187804149   3rd Qu.:242690489   3rd Qu.:5
 Max.   :53868775   Max.   :187804149   Max.   :242690489   Max.   :5
```
xxiii

```
> accuracy(fReg1Az,VAz)
                    ME       RMSE        MAE           MPE      MAPE      MASE       ACF1
Training set  0.000000 0.09434281 0.07734694 -0.0001831126 0.1105405 0.1708066 -0.1470734
Test set     -2.907036 3.60451322 3.16814796 -4.2140930419 4.5750805 6.9962782  0.4442370
             Theil's U
Training set        NA
Test set      2.166977
```
xxiv

```
> accuracy(fReg2Az,VAz)
                        ME       RMSE        MAE           MPE      MAPE      MASE       ACF1
Training set  2.030122e-15 0.09394828 0.07876871 -0.0001798587 0.1127125 0.1739464 -0.1338137
Test set     -2.785119e+00 3.47700421 3.06969048 -4.0390613120 4.4324976 6.7788527  0.4350113
             Theil's U
Training set        NA
Test set      2.089464
```
xxv

```
> accuracy(fReg3Az,VAz)
                   ME      RMSE       MAE           MPE       MAPE      MASE       ACF1
Training set 0.000000 0.0672732 0.0497483 -0.0000924342 0.07156135 0.1098601 -0.5181024
Test set     1.677381 3.2606348 2.4709592  2.3694638614 3.53103275 5.4566636  0.3602392
             Theil's U
Training set        NA
Test set      1.950905
```
xxvi

xxvii Quartic and 5th, and 6th power model were all tested, but results in the same Theil's U statistics.

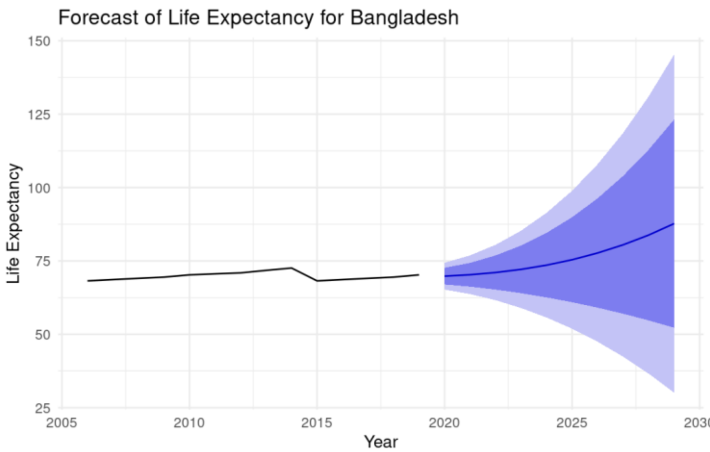
```
> accuracy(fReg3Ta, VTa)
                     ME       RMSE        MAE           MPE       MAPE       MASE         ACF1
Training set  0.0000000 0.01080354 0.01004323 -2.114171e-06 0.01717304 0.01146226  0.173841887
Test set     -0.1662514 0.18420497 0.16625136 -2.570743e-01 0.25707428 0.18974134 -0.001652311
             Theil's U
Training set        NA
Test set     0.3728681
```
xxviii

```
> accuracy(fReg1Ta,VTa)
                        ME       RMSE        MAE           MPE      MAPE       MASE
Training set -6.459479e-16 0.07446632 0.06279339  1.415639e-05 0.1091482 0.07166559
Test set     -4.554455e-01 0.52029182 0.45544545 -7.038776e-01 0.7038776 0.51979623
                     ACF1 Theil's U
Training set  0.510391251        NA
Test set     -0.005737043  1.062334
```
xxix

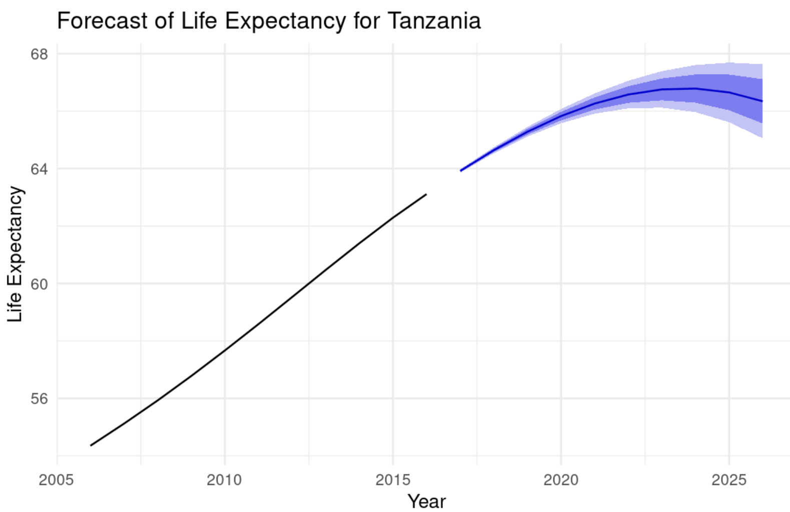
```
> accuracy(fReg2Ta,VTa)
                     ME       RMSE       MAE           MPE       MAPE       MASE         ACF1
Training set  0.0000000 0.05404331 0.0474787 -8.193105e-05 0.08044883 0.05418706  0.359291529
Test set     -0.6855399 0.75563484 0.6855399 -1.060170e+00 1.06016977 0.78240112 -0.004484912
             Theil's U
Training set        NA
Test set      1.523524
```
xxx

```
> RegFinAz <- tslm(newAz ~ poly(trend, degree = 3))
> forecast(RegFinAz, h=10)
     Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
2020       69.81705 67.00004  72.63407 65.24281  74.39130
2021       70.29861 66.24889  74.34832 63.72272  76.87450
2022       71.05543 65.24516  76.86570 61.62076  80.49009
2023       72.13111 64.01804  80.24418 58.95719  85.30504
2024       73.56926 62.58138  84.55714 55.72725  91.41127
2025       75.41347 60.93903  89.88791 51.91001  98.91693
2026       77.70734 59.08954  96.32514 47.47593 107.93875
2027       80.49447 57.02866 103.96027 42.39091 118.59802
2028       83.81845 54.75065 112.88625 36.61843 131.01847
2029       87.72289 52.24890 123.19689 30.12055 145.32524
```
xxxi

```
> RegFinTa <- tslm(TTa ~ poly(trend, degree = 3))
> forecast(RegFinTa, h=10)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2017       63.91602 63.87704 63.95499 63.85088 63.98115
2018       64.63967 64.57603 64.70331 64.53331 64.74602
2019       65.28107 65.18116 65.38098 65.11410 65.44804
2020       65.82683 65.67830 65.97535 65.57861 66.07504
2021       66.26352 66.05289 66.47414 65.91152 66.61552
2022       66.57775 66.29021 66.86528 66.09722 67.05827
2023       66.75610 66.37546 67.13674 66.11998 67.39222
2024       66.78517 66.29382 67.27653 65.96402 67.60633
2025       66.65156 66.03042 67.27270 65.61352 67.68961
2026       66.34186 65.57043 67.11328 65.05265 67.63107
```
xxxii


Forecast of Life Expectancy for Bangladesh
150
125
100
75
50
25
Life Expectancy
2005
2010
2015
2020
2025
2030
Year

xxxiii


Forecast of Life Expectancy for Tanzania
68
64
60
56
Life Expectancy
2005
2010
2015
2020
2025
Year