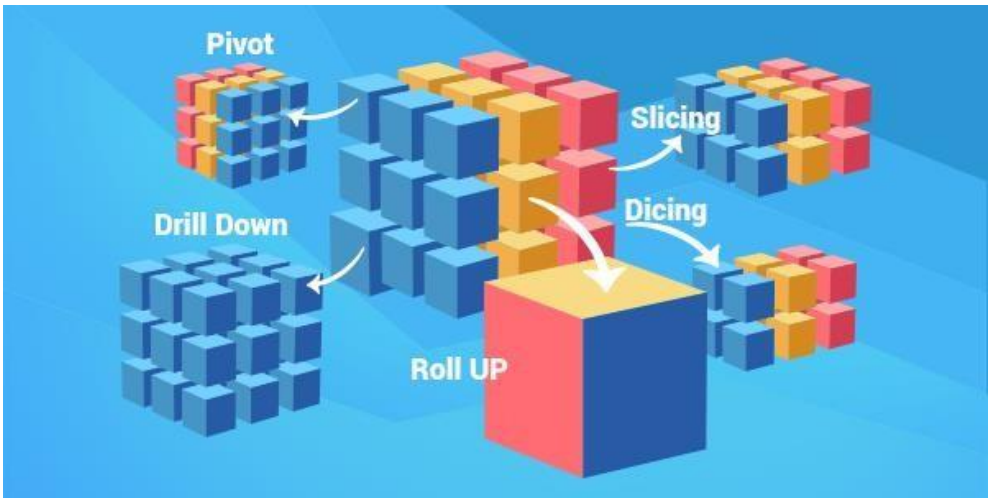
Name	Shubham Golwal
UID no.	2020300**15**
Experiment No.	05

AIM:	OLAP Operations using Tableu Tool
Experiment No. 5	
PROBLEM STATEMENT :	Perform OLAP Operations using Tableu Tool on a dataset
Theory :	**OLAP:** OLAP stands for On-Line Analytical Processing. OLAP is a classification of software technology which authorizes analysts, managers, and executives to gain insight into information through fast, consistent, interactive access in a wide variety of possible views of data that has been transformed from raw information to reflect the real dimensionality of the enterprise as understood by the clients. OLAP implement the multidimensional analysis of business information and support the capability for complex estimations, trend analysis, and sophisticated data modeling. It is rapidly enhancing the essential foundation for Intelligent Solutions containing Business Performance Management, Planning, Budgeting, Forecasting, Financial Documenting, Analysis, Simulation-Models, Knowledge Discovery, and Data Warehouses Reporting. OLAP enables end-clients to perform ad hoc analysis of record in multiple dimensions, providing the insight and understanding they require for better decision making. 

	OLAP Guidelines (Dr.E.F.Codd Rule): **1) Multidimensional Conceptual View:** This is the central features of an OLAP system. By needing a multidimensional view, it is possible to carry out methods like slice and dice. **2) Transparency:** Make the technology, underlying information repository, computing operations, and the dissimilar nature of source data totally transparent to users. Such transparency helps to improve the efficiency and productivity of the users. **3) Accessibility:** It provides access only to the data that is actually required to perform the particular analysis, present a single, coherent, and consistent view to the clients. The OLAP system must map its own logical schema to the heterogeneous physical data stores and perform any necessary transformations. The OLAP operations should be sitting between data sources (e.g., data warehouses) and an OLAP front-end. **4) Consistent Reporting Performance:** To make sure that the users do not feel any significant degradation in documenting performance as the number of dimensions or the size of the database increases. That is, the performance of OLAP should not suffer as the number of dimensions is increased. Users must observe consistent run time, response time, or machine utilization every time a given query is run. **5) Client/Server Architecture:** Make the server component of OLAP tools sufficiently intelligent that the various clients to be attached with a minimum of effort and integration programming. The server should be capable of mapping and consolidating data between dissimilar databases. **6) Generic Dimensionality:** An OLAP method should treat each dimension as equivalent in both is structure and operational capabilities. Additional operational capabilities may be allowed to selected dimensions, but such additional tasks should be grantable to any dimension. **7) Dynamic Sparse Matrix Handling:** To adapt the physical schema to the specific analytical model being created and loaded that optimizes sparse matrix handling. When encountering the sparse matrix, the system must be easy to dynamically assume the distribution of the information and adjust the storage and access to obtain and maintain a consistent level of performance. **8) Multiuser Support:** OLAP tools must provide concurrent data access, data integrity, and access security. **9) Unrestricted cross-dimensional Operations:** It provides the ability for the methods to identify dimensional order and necessarily functions roll-up and drill-down methods within a dimension or across the dimension.

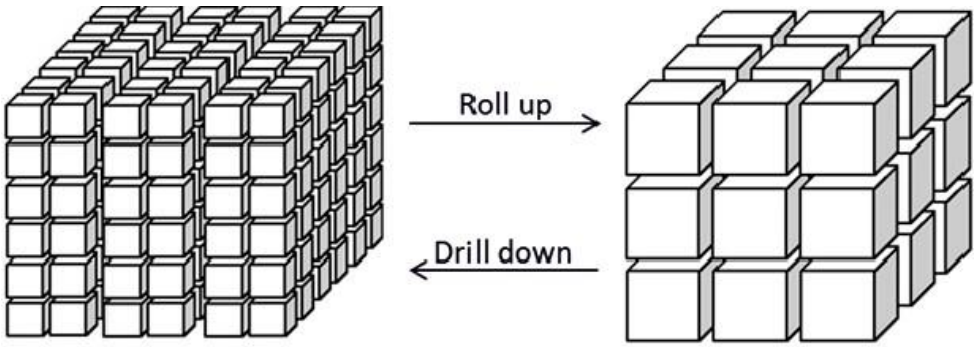
10) Intuitive Data Manipulation: Data Manipulation fundamental the consolidation direction like as reorientation (pivoting), drill-down and roll-up, and another manipulation to be accomplished naturally and precisely via point-and-click and drag and drop methods on the cells of the scientific model. It avoids the use of a menu or multiple tripsto a user interface.
11) Flexible Reporting: It implements efficiency to the business clients to organize columns, rows, and cells in a manner that facilitates simple manipulation, analysis, and synthesis of data.
12) Unlimited Dimensions and Aggregation Levels: The number of data dimensions should be unlimited. Each of these common dimensions must allow a practically unlimited number of customer-defined aggregation levels within any given consolidation path.

Benefits of OLAP:
OLAP holds several benefits for businesses: -
1. OLAP helps managers in decision-making through the multidimensional record views that it is efficient in providing, thus increasing their productivity.
2. OLAP functions are self-sufficient owing to the inherent flexibility support to the organized databases.
3. It facilitates simulation of business models and problems, through extensive management of analysis-capabilities.
4. In conjunction with data warehouse, OLAP can be used to support a reduction in the application backlog, faster data retrieval, and reduction in query drag.

OLAP Operations in the Multidimensional Data Model:
In the multidimensional model, the records are organized into various dimensions, and each dimension includes multiple levels of abstraction described by concept hierarchies. This organization support users with the flexibility to view data from various perspectives.

Several OLAP data cube operation exist to demonstrate these different views, allowing interactive queries and search of the record at hand. Hence, OLAP supports a user-friendly environment for interactive data analysis. Consider the OLAP operations which are to be performed on multidimensional data.

The figure shows data cubes for sales of a shop. The cube contains the dimensions, location, and time and item, where the location is aggregated with regard to city values, time is aggregated with respect to quarters, and an item is aggregated with respect to item types.



Roll-Up:
The roll-up operation (also known as drill-up or aggregation operation) performs aggregation on a data cube, by climbing down concept hierarchies, i.e., dimension reduction. Roll-up is like zooming-out on the data cubes. Figure shows the result of roll-up operations performed on the dimension location. The hierarchy for the location is defined as the Order Street, city, province, or state, country. The roll-up operation aggregates the data by ascending the location hierarchy from the level of the city to the level of the country.

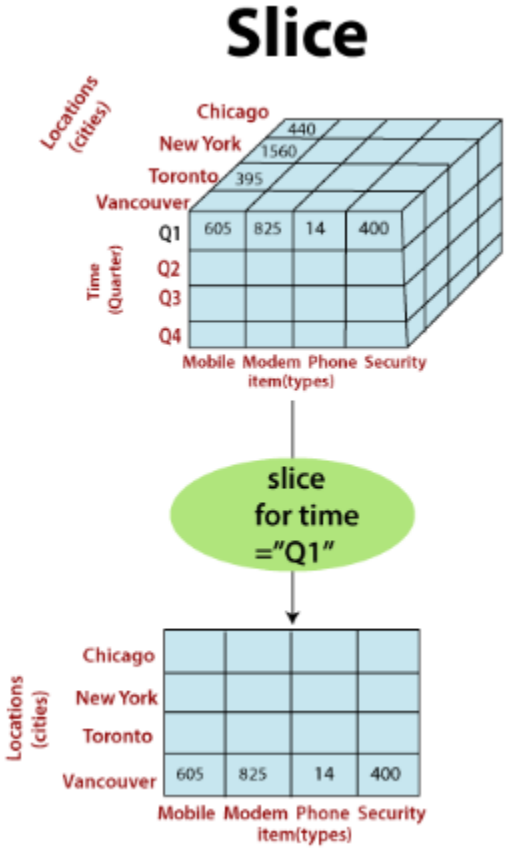
When a roll-up is performed by dimensions reduction, one or more dimensions are removed from the cube. For example, consider a sales data cube having two dimensions, location and time. Roll-up may be performed by removing, the time dimensions, appearing in an aggregation of the total sales by location, relatively than by location and by time.

Drill-Down:
The drill-down operation (also called roll-down) is the reverse operation of roll-up. Drill-down is like zooming-in on the data cube. It navigates from less detailed record to more detailed data. Drill-down can be performed by either stepping down a concept hierarchy for a dimension or adding additional dimensions.
Drill-down appears by descending the time hierarchy from the level of the quarter to a more detailed level of the month. Because a drill-down adds more details to the given data, it can also be performed by adding a new dimension to a cube. For example, a drill-down on the central cubes of the figure can occur by introducing an additional dimension, such as a customer group.

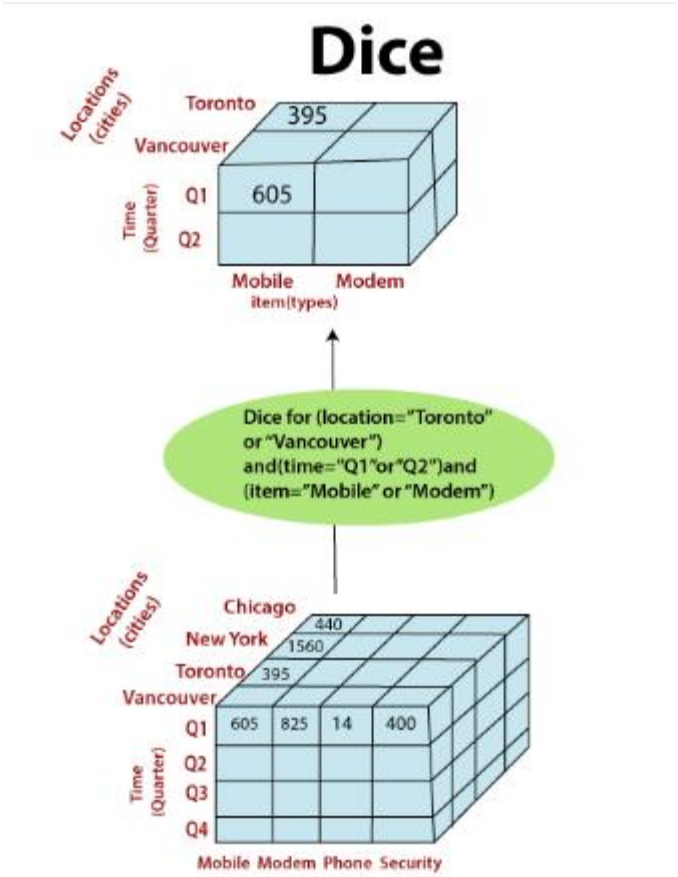
Slice: A slice is a subset of the cubes corresponding to a single value for one or more members of the dimension. For example, a slice operation is

executed when the customer wants a selection on one dimension of a three-dimensional cube resulting in a two-dimensional site. So, the Slice operations perform a selection on one dimension of the given cube, thus resulting in a subcube.

The following diagram illustrates how Slice works.



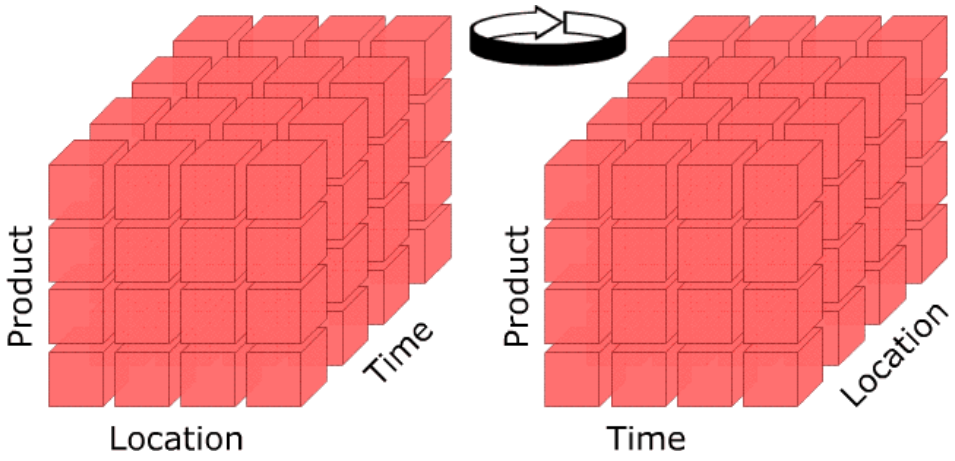
Here Slice is functioning for the dimensions "time" using the criterion time = "Q1". It will form a new sub-cubes by selecting one or more dimensions. Dice The dice operation describes a subcube by operating a selection on two or more dimension.



The dice operation on the cubes based on the following selection criteria involves three dimensions. o (location = "Toronto" or "Vancouver") o (time = "Q1" or "Q2") o (item =" Mobile" or "Modem")

Pivot:

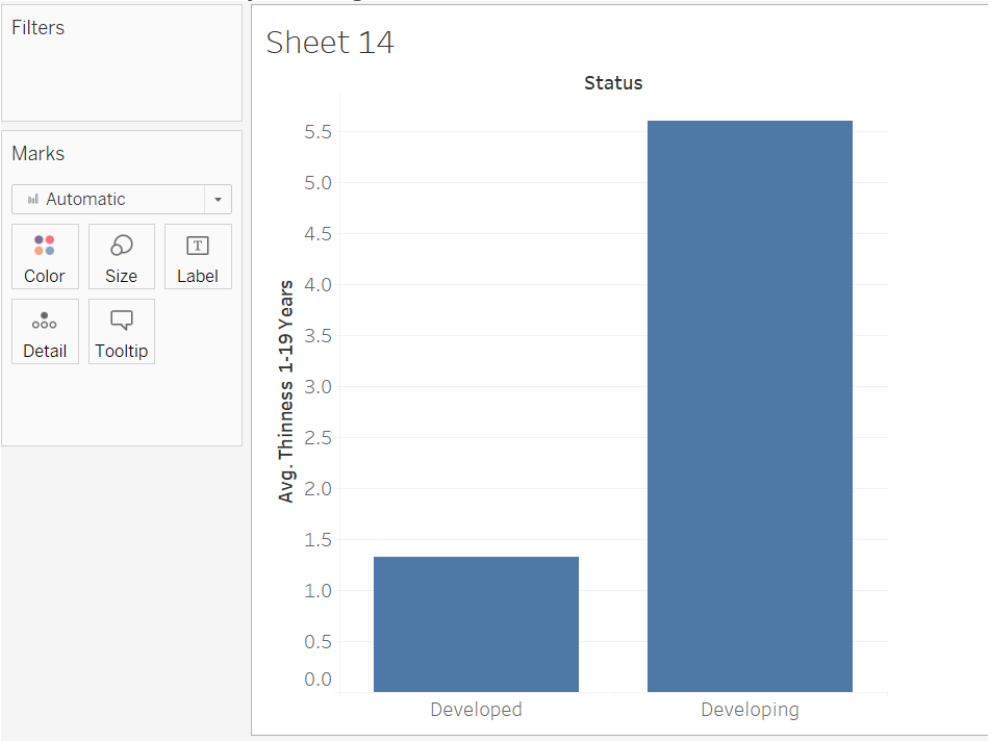
The pivot operation is also called a rotation. Pivot is a visualization operations which rotates the data axes in view to provide an alternative presentation of the data. It may contain swapping the rows and columns or moving one of the row-dimensions into the column dimensions.



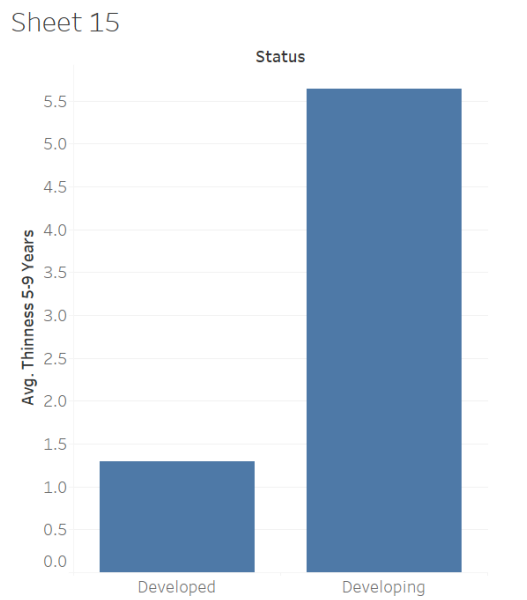
Output

Introduction to dataset:
I have taken WHO Life Expectancy dataset which has 2938 rows and 22 columns for analysis. The few important attributes which we have taken for data analysis are Life expectancy, country names, adult mortality, infant mortality, schooling, alcohol consumption etc.

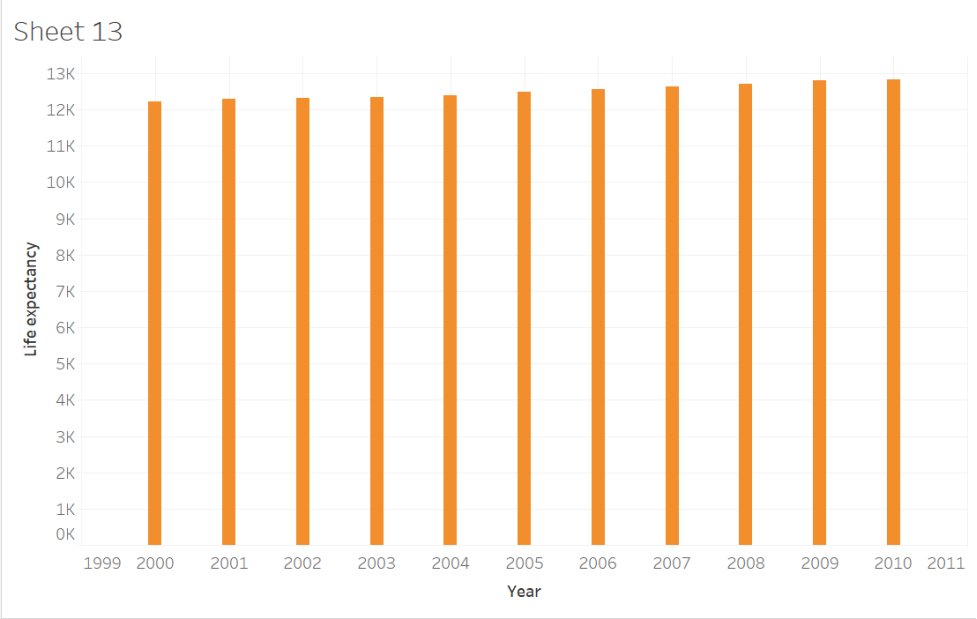
Roll Up and Drill Down:
Status of the Country v/s. avg. thinness 1-19 Yrs



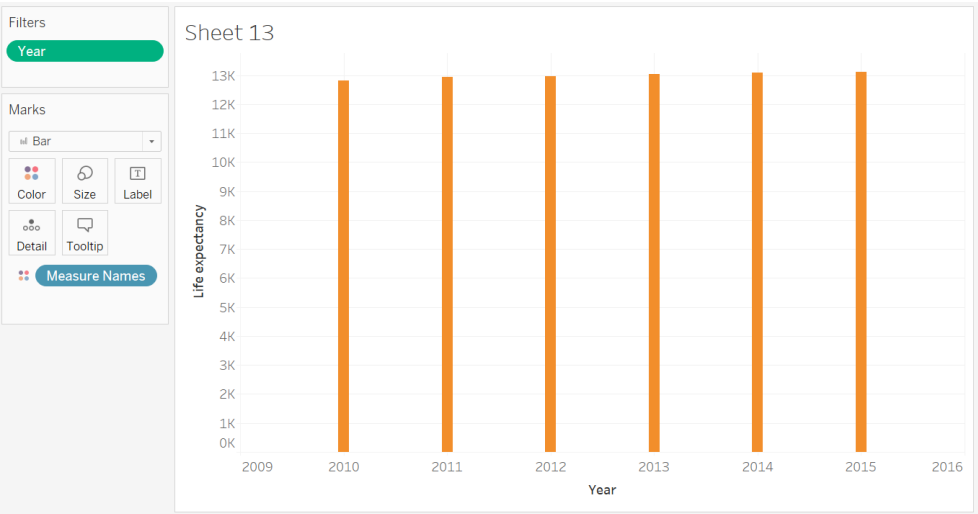
Status of the Country v/s. avg. thinness 5-9 Yrs



Life expectancy v/s. Years (2000-2010)

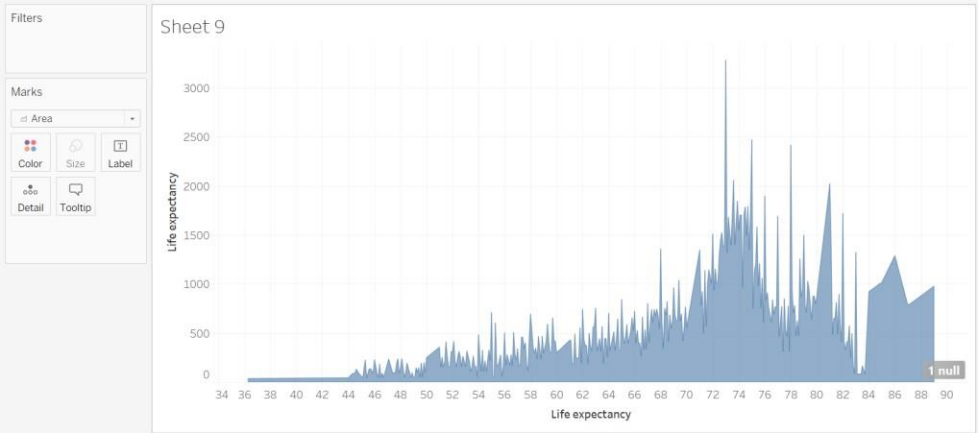


Life expectancy v/s. Years(2010-2016)



This shows that life expectancy over the years has grown gradually but in some years it had affected like in 2008 as of the global economic crisis.

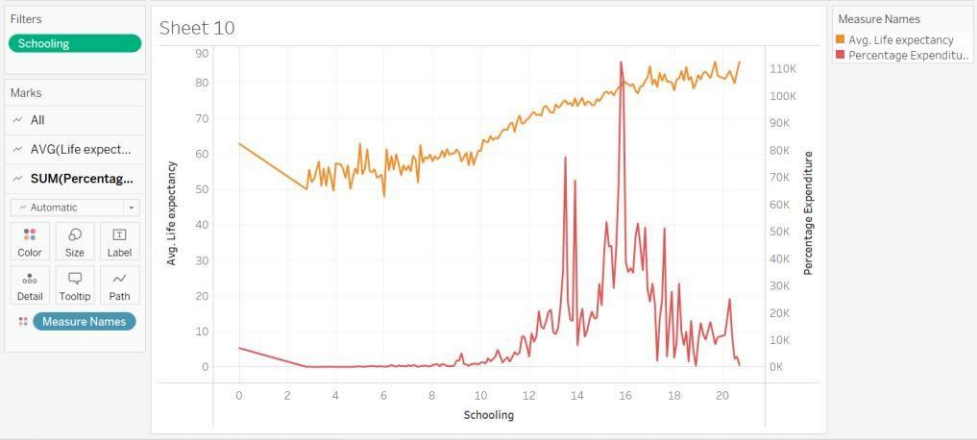
Slicing:



We have sliced the life expectancy column to study where most of the countries' life expectancy lies which comes out to be around 70-78 Yrs.

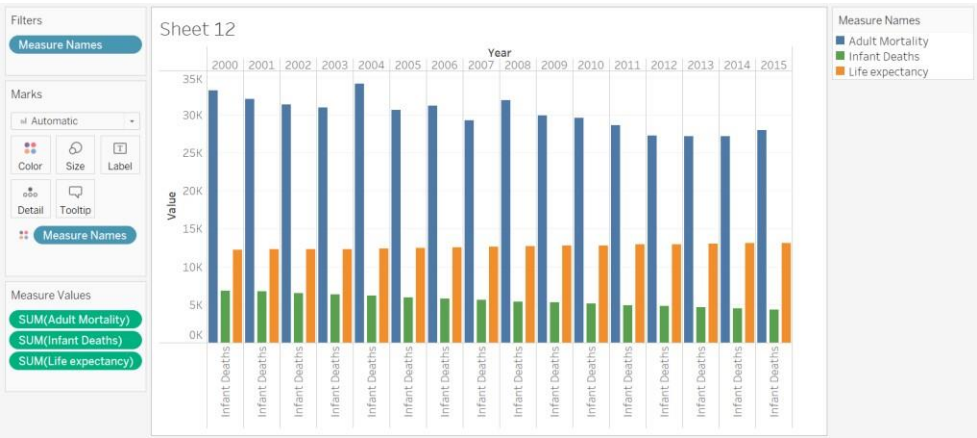
Dicing :

1.



I have plotted Life expectancy/Percentage Expenditure v/s. Schooling which states that as the Education level is high then Life expectancy increases and percentage expenditure is less in higher stages and high in mid stages.

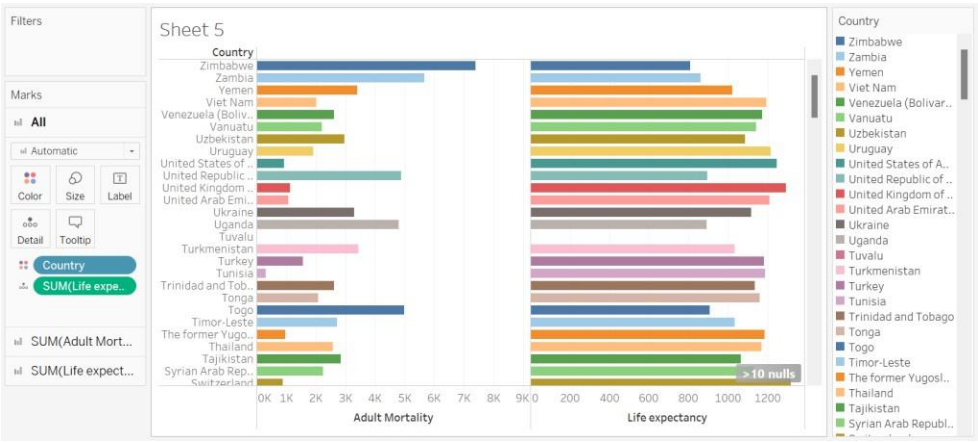
2.



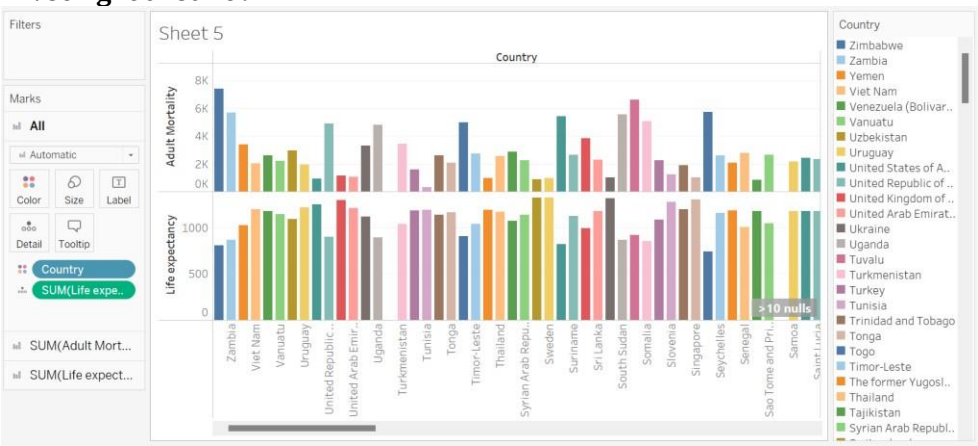
Here year-wise we can see that as the Adult mortality decreases and infant mortality decreases, Life expectancy increases over the year.

Pivot:

This graph shows adult mortality, life expectancy v/s. Country



Pivoting it around:



Conclusion:

1. During our experiment, we discovered that OLAP operations provided us with a quick and efficient way to examine large and intricate data sets. By manipulating data in various ways, we were able to detect valuable insights and patterns that might have gone unnoticed had we employed other techniques.
2. We also determined that different OLAP operations served different purposes in our analyses. For instance, drilling down and rolling up were highly effective for scrutinizing data at different levels of detail, while pivoting proved useful for comparing data across multiple dimensions.
3. Overall, our findings suggested that the integration of OLAP operations in data mining could result in better-informed decisions. By gaining a deeper understanding of the patterns and trends in the data, organizations can enhance their ability to make precise predictions and pinpoint opportunities for improvement or growth.

References:

1. **https://www.geeksforgeeks.org/olap-operations-in-dbms/**
2. **https://www.tutorialspoint.com/dwh/dwh_olap.htm**