

INT217: INTRODUCTION TO DATA MANAGEMENT

Final End term 2020 REPORT

on

Sales, Expense and Revenue Computation Dashboard

Submitted by

Shivani Sharma

Registration No: 11811289

Roll No: A21

Section: RKMO79

Programme Name: B.Tech. CSE

School of Computer Science & Engineering

DECLARATION

I hereby declare that I have completed the project platform from October 10, 2020, to December 4, 2020, under the guidance of Ms. Ashu. I have declared that I have worked full dedication and my learning outcomes fulfill the requirements of training for the award of degree of B.Tech.

CSE, Lovely Professional University, Phagwara.

Date-4, December 2020

Name of Student- Shivani Sharma

Registration No- 11811289

CERTIFICATE

This is to certify that SHIVANI SHARMA bearing Registration no. 11811289 has

completed INT217: INTRODUCTION TO DATA MANAGEMENT project titled,

"Sales, Expense and Revenue Computation Dashboard" under my guidance and

supervision. To the best of my knowledge, the present work is the result of her

original development, effort and study.

SCHOOL OF COMPUTER SCIENCE AND TECHNOLOGY

Lovely Professional University

Phagwara, Punjab.

Date: 04, December, 2020

3

ACKNOWLEDGEMENT

I would like to express my gratitude towards my University for providing me the golden opportunity to do this wonderful project regarding Data Management, which also helped me in doing a lot of homework and learning. As a result, I came to know about so many new things. So, I am thankful to them.

Moreover I would like to thank my friends who helped me a lot whenever I got stuck in some problem related to my course. I am thankful to have such good support from them as they always have my back whenever I need it.

Also, I would like to mention the support system and consideration of my parents who have always been there in my life to make me choose the right thing and oppose the wrong. Without them, I could never have learned and become the person I am now.

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

Table of content

1.	Introd	duction	6
2.	Objec	ctives/Scope of the Analysis	11
3.	Sourc	ce of dataset	15
4.	ETL 1	process	16
5.	Analy	ysis on dataset (for each analysis)	23
	i.	Introduction	
	ii.	General Description	
	iii.	Specific Requirements, functions and formulas	
	iv.	Analysis results	
	v.	Visualization	
6.	List o	of Analysis with results	35
7.	Refer	rences and Bibliography	38

Introduction

Data management refers to a wide range of activities researchers perform as they collect, arrange, describe, store, preserve, and often share materials generated by their research.

Many funding agencies and institutions require researchers to develop data management plans, or DMPs, that outline their data management activities, and to disseminate their data through repositories and archives.

- Data Management Plans links to the DMP Tool, a free web-based resources for writing a data management plan, as well as tools, guides, and best practices for writing DMPs.
- Data Repositories provides suggestions for selecting a repository and briefly discusses data-sharing concerns.
- Summary of Agency Requirements links to a table of federal agency's, their requirements to make research public.
- Statistical Methods & Data Sources is a Research Guide that links to EMU data subscriptions, free data sources, and more.



Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and make the decision based upon the data analysis.

Types of Data Analysis: Techniques and Methods

There are several types of Data Analysis techniques that exist based on business and technology. However, the major types of data analysis are:

- Text Analysis
- Statistical Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

In this project, Statistical Analysis is done

Statistical Analysis shows "What happen?" by using past data in the form of dashboards. Statistical Analysis includes collection, analysis, interpretation, presentation, and modeling of data. It analyses a set of data or a sample of data. There are two categories of this type of Analysis - Descriptive Analysis and Inferential Analysis.

Tool used



Excel is a basic, popular, and widely used analytical tool almost in all industries. Whether one is an expert in SAS, R, or Tableau, will still need to use Excel. Excel becomes important when there is a requirement of analytics on the client's internal data. It analyzes the complex task that summarizes the data with a preview of pivot tables that helps in filtering the data as per client requirement. Excel has the advanced business analytics option which helps in modeling capabilities that have prebuilt options like automatic relationship detection, a creation of DAX measures, and time grouping.



What is Interactive Dashboard?

An interactive dashboard allows you to drill down and filter information so data can be viewed from different perspectives or in more detail. With a condensed and clear presentation of the business figures, dashboards enable data-driven company decisions. People don't need to analyze the whole data set to tell them what's the matter - the headline information is there. People know the color in the chart needs to be green, not red, and so using the interactive features of the dashboard, can find out why this is the case.

In this project interactive dynamic dashboard will be used for data analysis Dynamic dashboards enable each user to see the data they have access to. With a dynamic dashboard, you can control data visibility without having to create a separate dashboard, with its running user and folder, for each level of data access.

Objectives/Scope of the Analysis

Whenever we take any decision in our day-to-day life is by thinking about what happened last time or what will happen by choosing that particular decision. This is nothing but analyzing our past or future and making decisions based on it. For that, we gather memories of our past or dreams of our future. So that is nothing but data analysis. Now the same thing an analyst does for business purposes is called Data Analysis.

In this data analysis project, I have chosen Sales, Expense, and Revenue Computation Dashboard to create a statistical analysis that is done using Microsoft Excel.

- Using data analysis and business intelligence to improve efficiency and streamline everyday operations in real-time.
- A subset of business analytics, operational analytics is supported by data mining, artificial intelligence, and machine learning. It requires a robust team of business and data analysts. And it also requires the right tools, As such, operational analytics is much better suited to large organizations than small businesses—at least for now.

Objectives and benefits of Sales, Expense, and Revenue Computation interactive dynamic Dashboard

1.Faster Decision-Making

Quite simply, businesses that can analyze and react to customer data in real-time can make much faster decisions.

Traditionally, businesses would make adjustments to their operations based on a quarterly or annual data review. In this reactive manner, they might miss out on serious revenue or glaring issues. They'd only become aware after the fact.

On the other hand, companies that embrace an operational analytics platform can make adjustments to processes and workflows in real-time. Or at least close to it. As such, they are in a better position to increase profitability and reduce waste. They can also detect problems and inefficiencies quickly and respond to them rapidly.

One recent study found that improving operations can result in a \$117 billion increase in profitability for global organizations.

2. Enhanced Customer Experiences

Businesses that react to situations in real-time can provide better customer experiences. It's that simple.

For example, imagine an e-commerce company that runs operational analytics. After that, they find that a significant percentage of its users are adding items to their carts but not completing transactions. Armed with that information, they then investigate the issue. It quickly becomes apparent that their website is buggy and checking out is a nuisance.

After identifying what's wrong and fixing it, the company improves the customer experience and drives more online sales.

3. Increased Productivity

Thanks to operational analytics, businesses can see the inefficiencies that exist in their workflows. Accordingly, they can then change their processes to streamline operations.

For example, a company might run analytics and realize that the process for approving a purchase order is too cumbersome. In this case, it requires too many signatures from too many people who are moving around constantly.

This data might encourage them to rethink the process entirely. They may either decide to reduce the number of signatures required to approve a PO.

Or they could opt to move to an online system that eliminates the need for having to track anyone down in person.

Now that you understand some of the benefits of operational analytics, let's take a brief look at a real-world example.

Source of dataset



Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

Link of dataset: https://www.kaggle.com/datatobe/operational-dataset

ETL process

ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.), and finally loads the data into the Data Warehouse system.

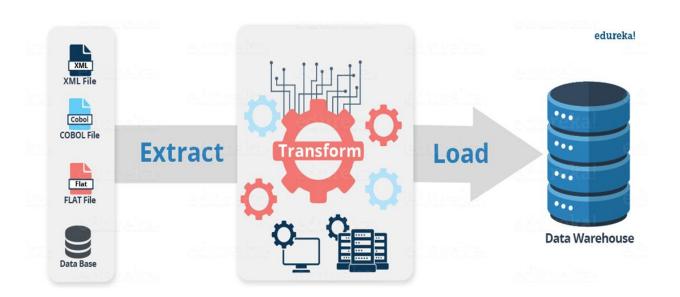
- •ETL full-form is
- •Extract
- •Transform
- •Load

Creating a Data warehouse is not simply extracting data from multiple sources and loading it into the database of a Data warehouse.

- •This requires a complex ETL process.
- •The ETL process requires active inputs from various stakeholders including developers, analysts, testers, top executives, and is technically challenging.
- To maintain its value as a tool for decision-makers, the Data warehouse system needs to change with business changes.
- •ETL is a recurring activity (daily, weekly, monthly) of a Data warehouse system and needs to be agile, automated, and well documented.
- •ETL provides a method of moving the data from various sources into a data warehouse.
- •In the first step extraction, data is extracted from the source system into the staging area.
- •In the transformation step, the data extracted from the source is cleansed and transformed.

•Loading data into the target data warehouse is the last step of the ETL process.





There are many reasons for adopting ETL in the organization:

- •It helps companies to analyze their business data for making critical business decisions.
- •Transactional databases cannot answer complex business questions that can be answered by ETL.
- •A Data Warehouse provides a common data repository
- •ETL provides a method of moving the data from various sources into a data warehouse.
- •As data sources change, the Data Warehouse will automatically update.

- •Well-designed and documented ETL system is almost essential to the success of a Data Warehouse project.
 - •Allow verification of data transformation, aggregation, and calculation rules.
- •ETL process allows sample data comparison between the source and the target system.
- •ETL process can perform complex transformations and requires the extra area to store the data.
- •ETL helps to Migrate data into a Data Warehouse. Convert to the various formats and types to adhere to one consistent system.
 - •ETL is a predefined process for accessing and manipulating source data into the target database.
 - •ETL offers deep historical context for the business.
 - •It helps to improve productivity because it codifies and reuses without a need for technical skills.

Step 1- Extraction

- •In this step, data is extracted from the source system into the staging area.
- •It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also.

Three Data Extraction methods:

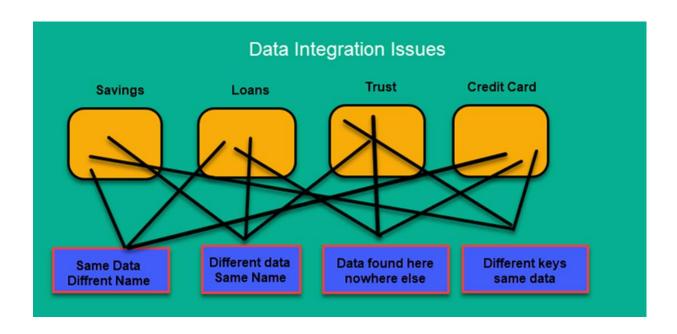
- 1.Full Extraction
- 2. Partial Extraction- without update notification.
- 3. Partial Extraction- with update notification
- •Irrespective of the method used, extraction should not affect performance and response time of the source systems.

Some validations are done during Extraction:

- •Reconcile records with the source data
- •Make sure that no spam/unwanted data loaded
- •Data type check
- •Remove all types of duplicate/fragmented data
- •Check whether all the keys are in place or not

Step 2) Transformation

- •Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped, and transformed.
- •In this step, you apply a set of functions on extracted data. Data that does not require any transformation is called as direct move or pass through data.



Step 3) Loading

- •The third and final step of the ETL process is loading.
- •In this step, the transformed data is finally loaded into the data warehouse.
- •Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals.
- •The rate and period of loading solely depends on the requirements and varies from system to system.
- •Types of Loading:
- •Initial Load populating all the Data Warehouse tables
- •Incremental Load applying ongoing changes as when needed periodically.
- •Full Refresh —erasing the contents of one or more tables and reloading with fresh data.

Best practices ETL process

•Never try to cleanse all the data:

Every organization would like to have all the data clean, but most of them are not ready to pay to wait or not ready to wait. To clean it all would simply take too long, so it is better not to try to cleanse all the data.

•Never cleanse Anything:

Always plan to clean something because the biggest reason for building the Data Warehouse is to offer cleaner and more reliable data.

•Determine the cost of cleansing the data:

Before cleansing all the dirty data, it is important for you to determine the cleansing cost for every dirty data element.

•To speed up query processing, have auxiliary views and indexes:

To reduce storage costs, store summarized data into disk tapes. Also, the trade-off between the volume of data to be stored and its detailed usage is required. Trade-off at the level of granularity of data to decrease the storage costs.

Analysis on dataset (for each analysis)

1. Finding the expense trend using the given dataset

Introduction and general description

Expense Trends Analysis (ETA) is designed specifically for individuals who want to look at summarized expense data over time. Its goal is to be easier than writing your own reports. ETA moves beyond reports to provide analysis. This means you don't have to start with a specific question.

Specific Requirements, functions and formulas

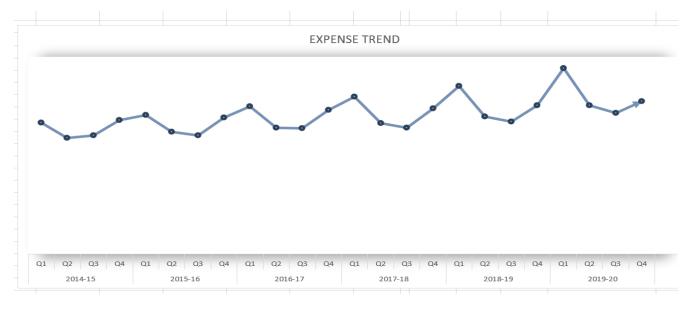
Using power pivot table with columns from Expense dataset and then creating line graph.

Creating Slicer from region using community (parent table) to connect all graphs and charts to end up with dynamic interactive dashboard.

Analysis results

Daniel ababa	C£	C f A	C f C	C (O-t	C N	C (D	C f I	C 4 T- b	C 6 14	C	C f 34	Cf I	Darelan	V	Owenter	Total Foresses	
B AsiaPac	Sum of Jul	Sum of Aug	Sum of Sep	Sum of Oct	Sum of Nov	Sum of Dec	Sum of Jan	Sum of reb	Sum of Iviar	Sum of Apr	Sum of Iviay	Sum of Jun	Region	Year	Quarter	Total Expense 332704578.9	
	107299550	99077838.5	126327191	90777469.9	64392742.9	120240000	CF24C101	04466746.0	150636291	77302671.7	103726744	158699686			Q1	294410301.3	
2014-15	115375860			95555231.4			1	84466746.9						2014-1	Q2	300349139.3	
2015-16 2016-17	117730469	111323414 115961890					•								Q3	339729101.8	
2016-17	130811633	117133222							146390954 142127140		111895086 107591429				Q4	351775700.4	
															Q1	309905580.3	
2018-19	139161311	128717826									120889246			2015-1	Q2		
2019-20	161815478	123767141				175192653			154234552		132845325				Q3	300699353	
Grand Total	772194302	695981331	885649577	613054887	454491536	912764244	445165890	605081207	883374896	479050459	6//653406	1026578029			Q4	345841384.6	
															Q1	374227669.5	
														2016-1	Q2	320235459.3	
															Q3	318808801.9	
													AsiaPa	С	Q4	364690594.1	
															Q1	399058091.6	
														2017-18	Q2	331184222.4	
															Q3	320654204.9	
															Q4	368710761.5	
															Q1	425288759.6	
														2018-19	Q2	348330212.3	
														2020 2	Q3	335146737.7	
															Q4	376888591.6	
															Q1	470770409.5	
														2019-20	Q2	376244890.3	
														2013-20	Q3	357963755.7	
															Q4	387421460.9	

Visualization



2. Finding Operational Effectiveness by Department

Introduction and general description

Operational effectiveness is a core objective of enterprise and IT governance where organizations seek to maximize the efficient use of resources in their business operations and to improve quality, productivity, or competitive positioning in markets in which they participate.

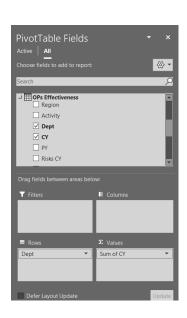
Several financial ratios allow business leaders to analyze their operational efficiency. The basic efficiency ratio is operating expenses divided by revenues, with a lower result indicating a greater level of efficiency.

Specific Requirements, functions and formulas

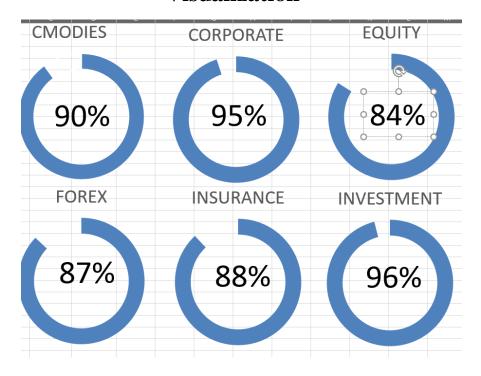
Using power pivot table, selecting following columns from Ops Effectiveness, creating doughnut chart. Also using functions like SUM().

Analysis results

- 1				
2				
3	Row Labels 🔻	Sum of CY	Remaining	
4	Cmodies	90%	10%	
5	Corporate	95%	5%	
6	Equity	84%	16%	
7	FOREX	87%	13%	
8	Insurance	88%	12%	
9	Invtment	96%	4%	
10	Grand Total	540%		
11				



Visualization



3. Finding change in Sales Volumes FY (Fiscal Year)

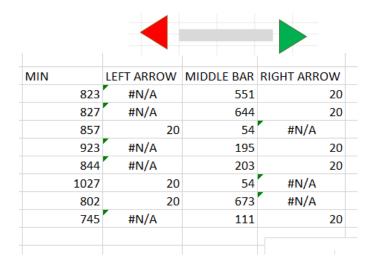
Introduction and general description

A company's fiscal year is its financial year; it is any 12-month period that the company uses for accounting purposes. The fiscal year is expressed by stating the year-end date. A fiscal year-end is usually the end of any quarter, such as March 31, June 30, September 30, or December 31.

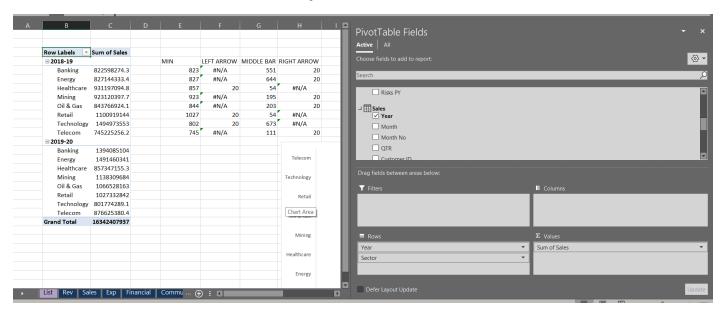
Sales volume is the number of units that are sold in a given time period. This is not to be confused with total sales, which are usually quantified as a monetary value. Sales volume is measured differently: Let's say a cosmetics brand sells 500 units of mascara in Q1. Their sales volume is 500.

Specific Requirements, functions and formulas

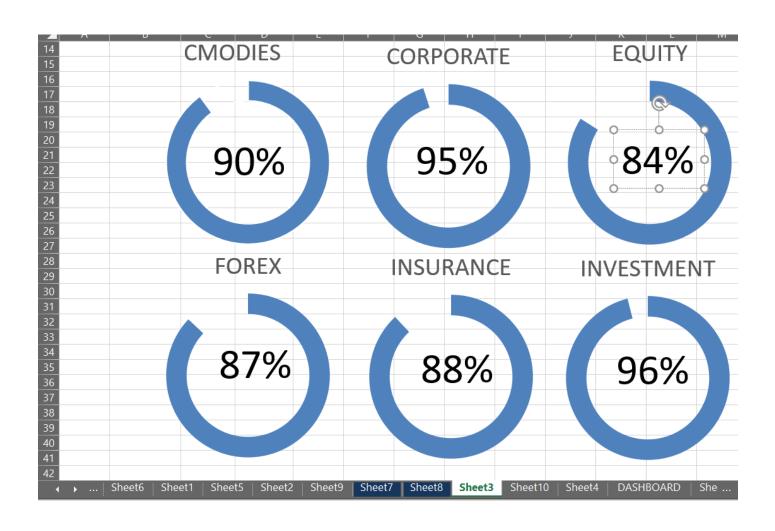
Using power pivot table and creating stack bar graph using shapes and clustered bar graph, also using functions like MIN, MAX,IF functions



Analysis results



Visualization



4.Finding Sales Volumes FY Data CY v PY

Introduction and general description

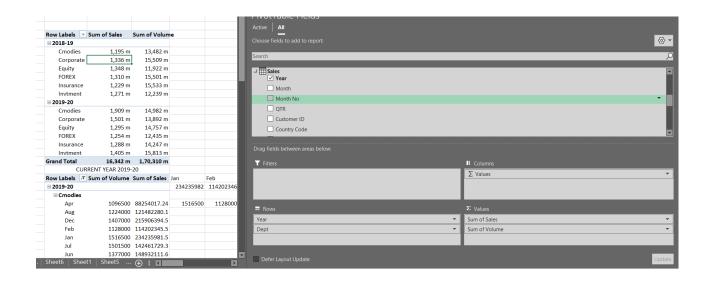
Sales volume is the number of units that are sold in a given time period. Here comparing current year sales volume with previous year sales volume.

Specific Requirements, functions and formulas

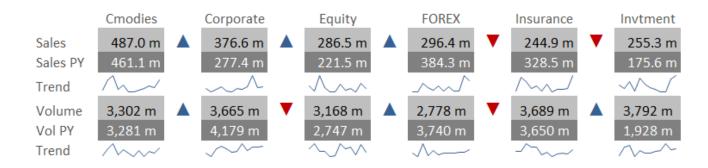
Using power pivot table, from Sales selecting columns and using function IF(logical_test,[Value_if_true],[Value_if_false]) for comparing Previous Year and Current Year Sales Volume.

Using symbols for up and down values for representing in the chart. Also used sparkline for trend.

Analysis results



Visualization



5. Finding the revenue trend using dataset

Introduction and general description

Revenue growth is the increase, or decrease, in a company's sales between two periods. Communicated as a percentage, revenue growth demonstrates the degree to which your company's revenue has grown (or shrunk) over time.

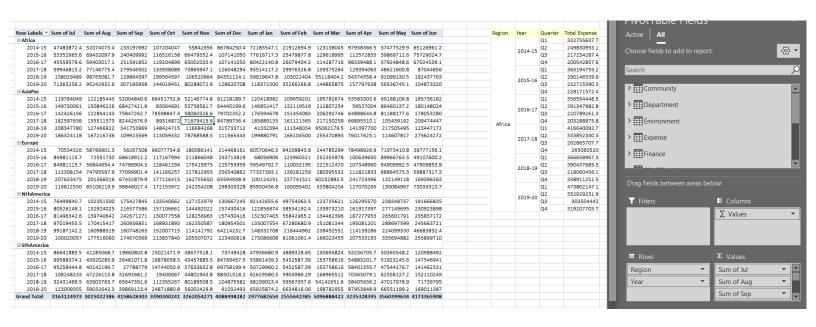
For example, if a company generated \$50 million in revenue during one business year and \$75 million in revenue the next, it saw a 50% revenue growth. If the same company generated \$50 million in annual revenue one year but only \$42 million the next, then it saw a 16% reduction in revenue growth (and might need to think about a new approach to how they coach their sales team!)

Specific Requirements, functions and formulas

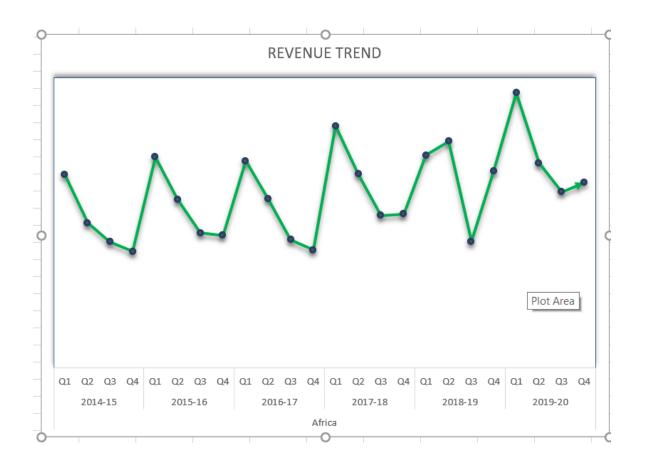
Using power pivot table and creating line graph using Revenue dataset.

And also connecting with slicer with Region

Analysis results

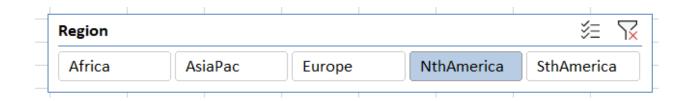


Visualization



Slicer

Region slicer from main table Community connected with every table to make dashboard interactive



List of Analysis with results

SALES VS EXPENSE YOY



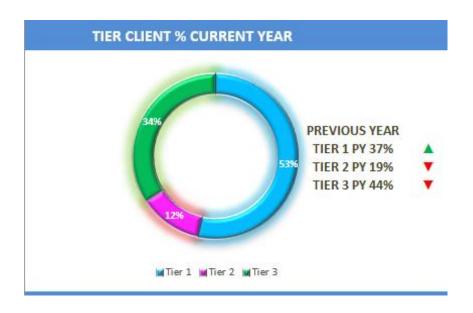
SALES VOLUME TREND CY V PY



RISKS BY DEPARTMENT YOY

Risks by Department YOY										
Department Trend	2015-16	2016-17	2017-18	2018-19	2019-20					
Cmodies	17	11	11	10	8					
Corporate	17	15	19	10	9					
Equity	17	13	20	6	17					
FOREX	11	13	8	9	15					
Insurance	16	12	21	17	18					
Invtment	17	12	14	8	14					
Invtment	17	12	14	8	14					

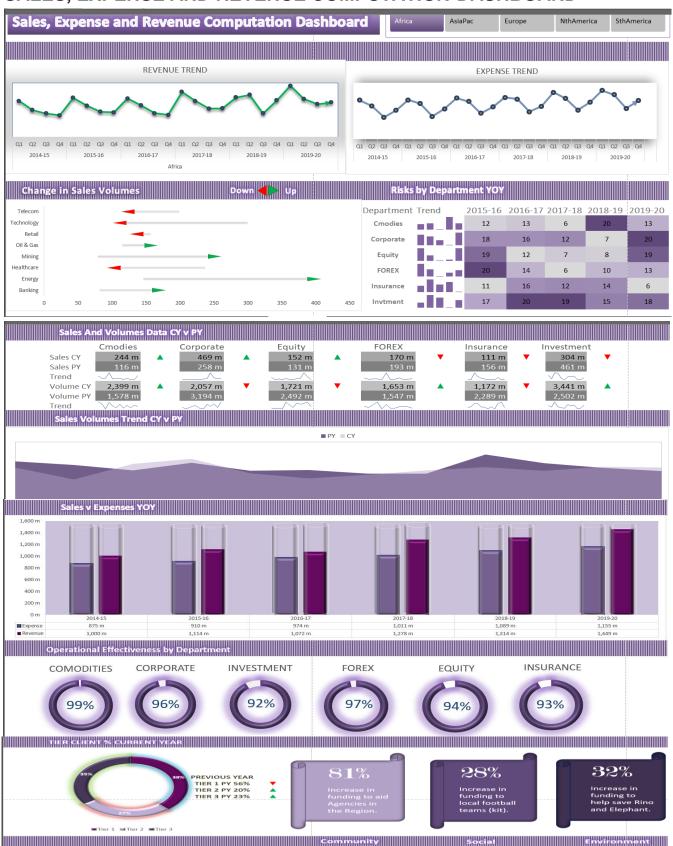
TIER 1 CLIENT % CY V PY



Community Social and Environment analysis region wise



SALES, EXPENSE AND REVENUE COMPUTATION DASHBOARD



References and Bibliography

https://www.investopedia.com/corporat e-finance-4689819

https://www.thebalancesmb.com/

https://courses.lumenlearning.com/

https://www.kaggle.com/

https://www.xplg.com/

https://edu.gcfglobal.org/

https://www.excel-easy.com/

https://www.guru99.com/

https://www.tutorialspoint.com/

https://www.javatpoint.com/