PROJECT DOCUMENTATION

Team ID	PNT2022TMID24015
Project Name	Project - Global Sales Data Analytics

1.INTRODUCTION

1.PROJECT OVERVIEW

To find and analyze the patterns on customers, Service and Transaction, Thus the findings can be utilized to answer the business of the customer.

1.2.PURPOSE

Sales analytics refers to the technology and processes used to **gather sales data and gauge sales performance**. Sales leaders use these metrics to set goals, improve internal processes, find a trend , forecast future sales and revenue more accurately.

2.LITERATURE SURVEY

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to data sets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these data sets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytic, which is the application of advanced analytic techniques on big data. This paper aims to analyze some of the different analytic methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytic in Global sales domains.

2.1 EXISTING PROBLEM

1. Predicting Sales

Predicting sales is of immense importance to organizations as its effects trickle down to critical business processes like inventory management, logistics, production and manpower planning. For instance, buying raw material and maintaining finished goods inventory is fundamentally driven by sales forecasts. Accurately predicting sales helps organizations to make better decisions and ensure the smooth running of processes.

Sales forecasting algorithms use a large amount of diverse data to look for patterns and relationships among various factors that affect sales under changing circumstances, thus predicting sales with a high level of accuracy.

2. Improve lead generation

Analytic has proven to be a great tool to improve lead generation and automate pre-sales processes. Companies are leveraging a vast resource of data to identify the right customers at the right time. Enterprises use a wide array of historical data to get a holistic picture of their prospective sales and many companies are pushing the limit by deploying lead-scoring algorithms which are fueled by granular and segmented data about each of their prospects. A complete 360-degree view of the customer is generated by combining in-house customer data and external data from news reports and social media posts.

These algorithms guide sales strategies by predicting the factors that are pivotal to lead conversion . According to a report by McKinsey, big-data analytics can be used to predict leads that are most likely to close, which is useful in planning the allocation of resources to improve lead conversion rate. By employing intelligent automation into the insight generation process, companies are seeing a significant leap in their ability to identify promising prospects and zero in on the right moment to target them. Enterprises are testing AI-enabled agents powered by predictive analytics and natural language processing to automate pre-sales activities and early lead-generation activities.

3. Analyzing customer sentiment

Sentiment analysis proves useful in understanding the feedback by customers. It employs AI to discern the emotions conveyed by customers and the semantics of the conversation. This is beneficial for businesses to understand how customers perceive their brand.

Sentiment analysis uses text mining algorithms to extract insights from social media websites, blogs or review sites. Automated sentiment analysis tools can be useful in extracting real-time actionable insights.

4. Better Cross-selling and Up-selling

With data analytics, companies can have an understanding of how their up sell and cross-sell strategies will perform

- Association Algorithm
- Classification
- Clustering Algorithm

1. Association Algorithm:

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an item set occurs in a transaction. A typical example is a Marke t Based Analysis.
Market Based Analysis is one of the key techniques used by large relations to show associations between items.It allows retailers to identify relationships between the items that people buy together frequently.
Given a set of transactions, we can find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

2. Classification:

The process of dividing a data set into mutually exclusive groups such that the members of each group are as ”close” as possible to one another, and different groups are as ”far” as possible from one another where distance is measured

with respect to species variable(s) you are trying to predict for example, atypical classification problem is to divide a database of companies into groups that are as homogeneous as possible with respect to a creditworthiness variable with values ”Good” and ”Bad.”

3. Clustering:

The process of dividing a data-set into mutually exclusive groups such that the members of each group are as ”close” as possible to one another, and different groups are as ”far” as possible from one another, where distance is measured with respect to all available variables given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities

•Automated prediction of trends and behaviors

•Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

•Automated discovery of previously unknown patterns Data mining tools sweep through databases and identify previously hidden patterns in one step. DARM discovers rules from various geographically distributed data sets. However, the network connection between those data sets isn’t as fast as in a parallel environment, so distributed mining usually aims to minimize communication costs.

2.2 REFERENCES

1. ”A Fast Distributed Algorithm for Mining Association Rules”,Proc. Parallel and Distributed Information Systems; D.W. Cheung, et al., IEEE CS Press, 1996,pp. 31-42

2. ”Introduction: Recent Developments in Parallel and Distributed Data Mining”,J. Distributed andParallel Databases; M.J. Zaki and Y. Pin, vol. 11, no. 2, 2002,pp. 123-127

3. ”Efficient Mining of Association Rules in Distributed Databases”,IEEE Trans. Knowledge andData Eng.; D.W. Cheung , et al., vol. 8, no. 6, 1996,pp. 911-922

4. ”Communication-Efficient Distributed Mining of Association Rules”; A. Schuster and R. Wolff,Proc. ACM SIGMOD Int’l Conf. Management of Data, ACM Press, 2001,pp. 473-484

5. ”Mining Association Rules Between Sets of Items in Large Databases”; R. Agrawal, T. Imielinski,and A. Swami, Proc. ACMSIGMOD Int’l Conf. Management of Data, , May 1993

6. ”An Optimized Distributed Association Rule Mining Algorithm”; M.Z Ashrafi, Monash UniversityODAM, IEEE DISTRIBUTED SYSTEMS ONLINE 1541-4922 2004.

7. ”The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling”,2nd edn. John Wiley & Sons; Kimball, R., Ross, M., New York (2002)

8. ”Web for Data Mining: Organizing and Interpreting the Discovered Rules Using theWeb”,SIGKDD Explorations; Ma, Y., Li u, B., Wong, C.K., Vol. 2 (1). ACM Press, (2000) 16-23.

9. ”New Algorithm for Fast Discovery of Association Rules”,Technical ReportNo. 261; Zaky, M.J., Parthasarathy,S.,Ogihara,M.,Li,W.,University of Rochester(1997)
http://cs.aue.aau.dk/contribution/projects/datamining/papers/t R651.pdf

2.3 PROBLEM STATEMENT DEFINITION / PROPOSED SOLUTION

• Unlike other algorithms, ODAM offers better performance by minimizing candidate item set generation costs. It achieves this by focusing on two major DARM issues: communication and synchronization. Communication is one of the most important DARM objectives. DARM algorithms will perform better if we can reduce communication (for example, message exchange size) costs. Synchronization forces

• Each participating site waits a certain period until globally frequent item set generation completes.Each site will wait longer if computing support counts takes more time. Hence, we reduce the computation time of candidate itemsets' support counts.

Each method has different aims, expectations, advantages, and disadvantages. For example, the first method exchan ges each candidate item set's support count to generate globally frequent item sets of that pass (CD and FDM are examples of this approach).

3. IDEATION AND PROPOSED SOLUTION:

3.1 Empathy Map Canvas :

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes.

It is a useful tool to help teams better understand their users.
Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

Traditional empathy maps are split into 4 quadrants (Says, Thinks, Does, and Feels), with the user or persona in the middle. Empathy maps provide a

glance into who a user is as a whole and are not chronological or sequential.

The **Says** quadrant contains what the user says out loud. Ideally, it contains verbatim and direct quotes from research.

- I want something reliable
- What size is the best
- What brand do you like
- Where should I start ?

The **Thinks** quadrant captures what the user is thinking throughout the experience. Ask yourself (from the qualitative research gathered): what occupies the user's thoughts? What matters to the user? It is possible to have the same content in both Says and Thinks. However, pay special attention to what users think, but may not be willing to vocalize. Try to understand why they are reluctant to share — are they unsure, self-conscious, polite, or afraid to tell others something?

- I want Something awesome
- Too many acronyms
- what else am i missing

The **Feels** quadrant is the user's emotional state, often represented as an adjective plus a short sentence for context. Ask yourself: what worries the user? What does the user get excited about? How does the user feel about the experience?

- Fear
- Anxious
- Overwhelmed
- inadequate

3.2. Ideation and Brainstorming

We've undoubtedly been in brainstorming sessions. Some of these sessions have likely been fruitful, others disappointing. We often get asked how ideation is different from brainstorming on " Brilliant."- a podcast hosted by Magnani's president. One guest distinguished the two types of sessions by asserting most brainstorms are simply "meetings… with better food." But beyond that perhaps undeserved jab at brainstorming, there are several aspects that separate brainstorms from formal ideation.

First, what is an ideation session, anyway?

Before I jump into the difference between a brainstorm and an ideation session, I should provide some context for anyone unfamiliar with this process. In traditional design-thinking, the ideation phase is often the most exciting step within the process. The ideation session itself is the organized gathering of minds within that step where the litany of ideas is generated against some highly defined problems or desired outcomes. These ideas range from the possible to the seemingly impossible given current organizational constraints.

A time and a place.

They're fun, engaging and often produce creative ideas. They are collaborative and aid in generating new ideas to improve internal processes, develop creative campaigns, share ideas, etc. This is all important work. But let's remember that ideation is the third step in a more formal design-thinking process and should be treated as such. It should be informed by learnings emerging from the Empathize stage, address specific challenges outlined in the Define phase and, finally, create a starting point for the Prototype and Test phases. Ideation is about not only generating ideas but also systematically upending and exploring the mental models surrounding those ideas, assessing recurring themes, evaluating ideas through a variety of lenses and, ultimately, converging and consolidating various branches of thought into manageable

future areas of innovation. Ideation, to that point, also requires more time, commitment, homework and buy-in from stakeholders. Ideation may be utilized for a multitude of business challenges. Some examples include:

- Developing new product or service directions
- Exploring new business strategies and revenue streams
- Finding new business angles by solving complex customer-centric challenges

3.3 Proposed Solution :

The proposed solution should relate the current situation to a desired result and describe the benefits that will accrue when the desired result is achieved. So, begin your proposed solution by briefly describing this desired result.

The Proposed solution is to create a Website with Sign up feature and various feature that are contained in the IBM cog nos Analytics

This also contains code various ways to create a Dashboard and analyze Trend

3.4 Problem Solution Fit :

Shopping Online is currently the need of the hour. Because of this Covid-19, it is not easy to walk into stores and gather Surveys. Every store be it online or offline needs evaluation and analysis to predict daily sales for better sales performance and it is essential to know what kind of goods a customer wants and at times they need it.Global Sales covers all activities involved in selling a product or service to a consumer or business. It is important for sales and marketing teams to review their strategies and performance in order to make improvements. One way to measure performance is with sales analytics. Sales data analytics refers to the use of technology to collect and use sales data to identify actionable insights. It is used to identify, optimize, and increase sales. An efficient sales model that generates higher revenue for the business. This also answers “Who does the problem affects”,”What are the boundaries of the problem ” , “What is the issue” ,etc

4. REQUIREMENT ANALYSIS

4.1 Functional Requirements:

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in use cases.

In this Project there are 6 functional requirement

User Registration - Register through Gmail \ Phone number
User Confirmation - Confirmation Via Email /OTP
User Login - Login Via Email and Password
Dataset - Dataset upload to Database
Visualize / Analyze - Cleaning Data , Analyze the Dataset
Create Dashboards - Create Charts ,Graphs , Tables

4.2 Non-functional Requirements:

Nonfunctional Requirements (NFRs) define system attributes such as security, reliability, performance, maintainability, scalability, and usability. They serve as constraints or restrictions on the design of the system across the different

This Project Contains 6 Non Functional Requirement

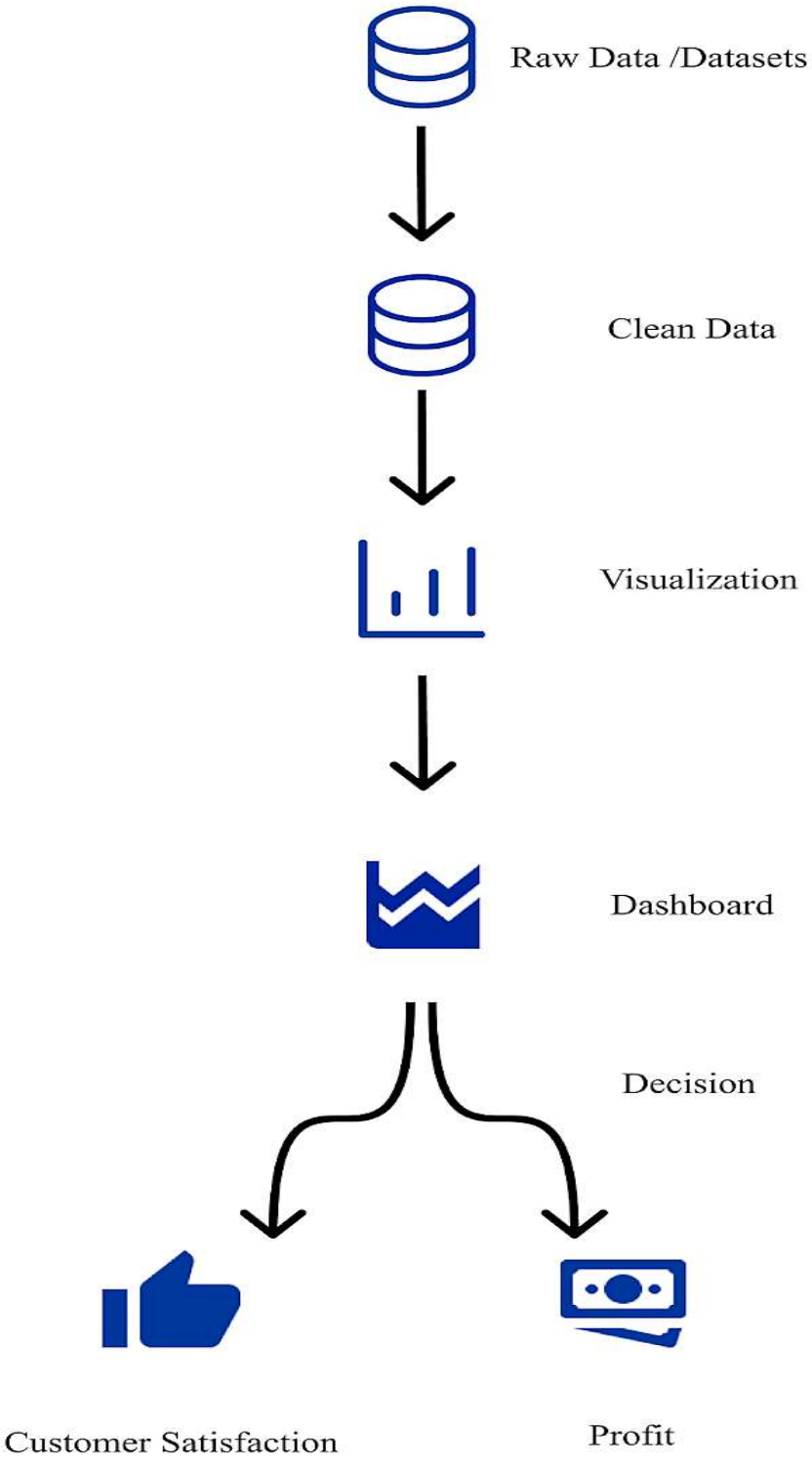
Usability - Optimized resources and it can be used by everyone
Security - It is secured as it has end to end encryption
Reliability - High reliability based on development
Performance - High performance based on the code
Availability - it is available in all platform and websites
Scalability - it is serverless or considered like amazon Lambda so its highly scalable and retractable

5. PRODUCT DESIGN

5.1 Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored. and how it is transformed

This figure contains the DFD :


Raw Data /Datasets
Clean Data
Visualization
Dashboard
Decision
Customer Satisfaction
Profit

5.2 Solution & Technical Architecture:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

Solution Technical Architecture contains many Software HTML , CSS ,JAVASCRIPT ,Flask

IBM Cloud , IBM Cog nos analytical tool , IBM DB2 , Kaggle

5.3 User Stories

The Functional Requirement and Non Functional Requirement can be converted USER STORIES

There must be 10 User stories with accordance to tabulation
USN -1 : As a user, I can register for the application by entering my email / mobile number,password, and confirming my password.

USN -2 : As a user, I will receive confirmation email once I have registered for the application
USN -3 : As a user, I can register for the application through mobile number
USN -4 : As a user, I can register for the application through Gmail
USN -5 : As a user, I can log into the application by entering email & password / phone number and OTP
USN -6 : As a user, I can navigate through the dashboard and view the data visualizations

USN -7 : As an Admin, I can login for the website using my credentials
USN -8 : As an Admin, i can upload the dataset on which the visualization has to be performed
USN -9 : As an Admin,i can customize dashboard if necessary

6 Sprint Planning & Scheduling

6.1 Sprint Planning & Estimation

Sprints have proven to work for thousands, if not millions of teams already, fostering people to get more done and management to fast-forward their inert projects. Once you get it right, the benefits of implementing agile practices won't be long in coming. During the last couple of years, agile has helped many organizations manage changing priorities and improve project visibility among other benefits, such as business alignment, time to market, and increased team productivity. More importantly, 26% of organizations participating in 2020's Annual State of Agile Survey also noted project cost reduction as an important reason for adopting agile.

What is a sprint?

First things first, a sprint is usually a two-week period of time during which specific tasks must be completed based on what the team has prioritized to deliver to the end user soon. In our previous article, we've related a number of reasons why you should work in agile sprints. It's important not to forget that the main purpose of sprints is to deliver frequently. Before cutting to the chase of how you can streamline sprint planning and increase your team's speed, let's look into what sprint planning is.

Sprint planning explained

Sprint planning is a meeting dedicated to one of the Scrum ceremonies, where the team collaborates to agree upon work it'll be able to complete during the upcoming sprint. Sprint planning sessions are needed to set the foundation for

the project and drive the team's boat at the most optimal pace during the next couple of weeks. To run a successful sprint, you'll need to know how to organize work and plan activities that can be finished in a short period of time. The purpose of a sprint planning meeting is to identify the sprint goal and sprint backlog.

- Sprint Goal: This refers to what can be delivered during the sprint
- Sprint Backlog: The list of tasks to be completed during the sprint to achieve that goal.

What happens during the sprint planning meeting? Let's say you're planning to develop a new important product feature or work on a marketing campaign. During the sprint planning session, the team members will have to 'groom' the backlog and say which tasks they'll work on.

You might think that planning for the next couple of weeks is the easiest thing you've ever done. But there's another side of the coin. The work you've planned must fulfill the goal set. This can only happen if you have a healthy backlog. According to Atlassian, a healthy backlog carries out three things:

● Prioritizes each work item, with the most important work listed at the top
● Includes fully-formed user stories the development team can begin to execute on
● Contains an up-to-date estimate for each work item.

Pay attention: To cover a two-week sprint, consider holding a two-hour sprint planning meeting. Ideally, the meeting should be held early in the week so as not to disrupt the team's flow by the weekend. The question is, who should be present at your sprint planning meeting?

Roles involved in agile sprint planning

Generally, the key roles involved in agile sprint planning – the product owner, scrum master, and the team all need to attend, especially if it's part of the agile process in software development companies. Let's take a look at each role.

The Team: Defines the work and effort

Teams also perform a critical role in the sprint planning session, as they actually get work done. They should be in full attendance for the meeting and put their heads together to contribute to the estimation and forecasting process as much as they can. Specifically, they will determine how many of the product backlog items they will be able to complete and elaborate on how they will deliver those product backlog items, so each team member can leave with a clear understanding of priorities and what needs to be accomplished during the next sprint

This SPRINT contains 4 parts :

- Registration & Login
- Dashboard
- Customer care
- Administrator

Each Sprint contains 6 Days from 24- OCT to 14 - Nov

7 . Coding & Solution

7.1 Feature - Authentication

As Flask as Back-end and HTML and CSS as front-end authentication is made easy with the possibility of IBM DB2 . Where it is typically associated with proving a user's identity. Usually, a user proves their identity by providing their credentials, that is, an agreed piece of information shared between the user and the system.

- SIGN-UP with Email
- SIGN-UP with phone number

The name & password is stored in the database with SSL encryption where the authorization factor is send using various functions

SIGN-UP

```
def signup():
    msg=''
    if request.method == "POST":
        username=request.form['username']
        email=request.form['email']
        pw=request.form['password']
        sql='SELECT * FROM register WHERE email =?'
        stmt = ibm_db.prepare(conn, sql)
        ibm_db.bind_param(stmt,1,email)
        ibm_db.execute(stmt)
        acnt=ibm_db.fetch_assoc(stmt)
        print(acnt)

        if acnt:
            msg='Account already exits!!'

        elif not re.match(r'[^@]+@[^@]+\.[^@]+', email):
```

```
        msg='Please enter the avalid email address'
    elif not re.match(r'[A-Za-z0-9]+', username):
        msg='name must contain only character and number'
    else:
        insert_sql='INSERT INTO register VALUES (?,?,?)'
        pstmt=ibm_db.prepare(conn, insert_sql)
        ibm_db.bind_param(pstmt,1,username)
        ibm_db.bind_param(pstmt,2,email)
        ibm_db.bind_param(pstmt,3,pw)
        ibm_db.execute(pstmt)
        msg='You have successfully registered click signin!!'
        return render_template("signin.html")



elif request.method == 'POST':
    msg="fill out the form first!"
return render_template("signup.html",msg=msg)
```

SIGN -IN

```
def signin():
    global userid
    msg = ''
    if request.method == 'POST' :
        un = request.form['username']
        pd = request.form['password']
        sql = "SELECT * FROM register WHERE username =? AND password=?"
        stmt = ibm_db.prepare(conn, sql)
        ibm_db.bind_param(stmt,1,un)
```

```
    ibm_db.bind_param(stmt,2,pd)
    ibm_db.execute(stmt)
    account = ibm_db.fetch_assoc(stmt)
    print (account)
    if account:
        session['loggedin'] = True
        session['id'] = account['USERNAME']
        userid=  account['USERNAME']
        session['username'] = account['USERNAME']
        msg = 'Logged in successfully !'

        return render_template('index.html', msg = msg)
    else:
        msg = 'Incorrect username / password !'
  return render_template('signin.html', msg = msg)

```

7.2 Feature - Automated Dashboard

Automation of various plots , graphs according to need

```
 <div class="Col-xxl-15 col -md-6">
                  <div class="card info-card Data-card">
                    <div class="card-body">
                            <h5 class="card-title">Analytics
</h5>
                             <div class="d-flex align-items-
center">
                               <div class="card-icon rounded-
circle d-flex align-items-center justify-content-center">
                          </div>
                           <div class="embed-responsive embed-
responsive-16by9 ">
                              <iframe class="embed-responsive-
```

```
   item"
   src="https://us1.ca.analytics.ibm.com/bi/?perspective=dashbo
   ard&amp;pathRef=.my_folders%2FGloabal%2Bsuper%2Bdata&amp;clo
   seWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=fa
   lse&amp;shareMode=embedded&amp;action=view&amp;mode=dashboar
   d&amp;subView=model0000018479b19f87_00000002"    width="800"
   height="600"         frameborder="0"         gesture="media"
   allow="encrypted-media" allowfullscreen=""></iframe>
10
11
12
13                        </div>
14                      </div>
15                    </div>
16                  </div>
17                </div>
```

8. TESTING

8.1 Test Cases

Dashboard Design - No of Visualization /Graphs -4
Data Responsive - very Responsive
Amount of Data to Rendered (DB2 Metric) - 3 MB (Global Super Store)
Utilization of Data Filters - Utilized to full effectiveness
Effective User Story No of Scenes Added - 9
Descriptive Reports No of Visualizations / Graphs - 4

8.2 User Acceptance Testing

By Design - 0 servity1,2,3,4
Duplicate - 0 severity 1,2,3,4
External - 0 severity 1,2,3,4
Fixed - 0 severity 1,2,3,4

Not Reproduced - 0 Severity 1,2,3,4
Skipped - 0 severity 1,2,3,4
Wont Fix - 0 severity 1,2,3,4
Total -0 Subtotal

Test Case Analysis

Sign up	Total cases - 4 Pass -4
Sign in	Total cases - 4 Pass -4
Help	Total cases - 2 pass -2
menu	Total cases - 1 pass - 1
Dashboard	Total cases -5 pass - 5

9. Results

Performance metrics are data used to track processes within a business.
This is achieved using activities, employee behavior, and productivity as key metrics.
These metrics are then used by employers to evaluate performance.
This is concerning an established goal such as employee productivity or sales objectives.

Tracking performance metrics is important because they provide valuable information to your business.
The data that these metrics provide can be used to grow your business and increase your profits.
They also help put strategies in place for meeting various objectives. This can be across any aspect of your business.
You can plan for improvements, adjustments, and any changes to the processes your business has in place to meet various goals.

By Design - 0 servity1,2,3,4
Duplicate - 0 severity 1,2,3,4
External - 0 severity 1,2,3,4

Fixed - 0 severity 1,2,3,4
Not Reproduced - 0 Severity 1,2,3,4
Skipped - 0 severity 1,2,3,4
Wont Fix - 0 severity 1,2,3,4
Total -0 Subtotal

10. Advantage & Disadvantage

Advantage :

Easily loaded with data
Signup using IBM
Highly responsive
Highly flexible

Disadvantage :

Loading Time of IBM
User must have account with IBM cognos to View

11. Conclusion

In conclusion this project Global Sales Analytics Shows how to load & mainipulate the data and Visualize it in the desired outcome and shows how to use api and Flask

12 . Future Scope

In the Future the will be various Feature Added

- Dynamic Data Change
- Customer Centre
- Forum
- Data Cleaning
- Data Collection

13. Appendix

Source Code

https://github.com/Az-har/Global-sales

Project Demo Link