GLOBAL SALES DATA ANALYTICS

TEAM MEMBERS:

Team Leader: Harrish Kumar G M

Team Member 1 : Madhumadhi I

Team Member 2 : Ramya S

Team Member 3 : Surya Prakash M

1. INTRODUCTION

1.1 Project Overview:

Shopping online is currently the need of the hour. Because of this COVID, it's not easy to walk in a store randomly and buy anything you want. So, try to understand a few things like, Customer Analysis and Product Analysis of this Global Super Store.In the meantime, business sales are steadily increasing. Companies want to track their sales and progress. The entire data set must be located and saved in the database. They must see the growth of their company. The company wants to know where they stand in relation to other companies. Because this data can be massive, estimating sales predictions is nearly impossible. These terms, as well as the physical estimation, may become time consuming in the near future. In light of this issue, we proposed data analysis and visualisation. Visualization is the graphical representation of data using graphs, tabulations, and charts to determine company sales. We could form an opinion on a company's sales and sales by different salespeople using data visualisation. As a result, we used a web application to integrate both the analysed data and to present that data to the end user through a simple user interface.

1.2 Purpose:

In a sales-driven world, data analytics can be used to identify trends, improve sales success, and increase revenue. It is critical to understand what analytics can do for the business when applying analytics to the sales process. Many businesses have emerged in recent years, and a few have risen to prominence. Some businesses have a large amount of data, while others have a small amount of data. The purpose is to investigate the company's growth in terms of product sales. The first goal is to create a web application that analyses a company's sales data for various products. The primary goal here is to read and analyse data in order to generate insights and increase the company's overall sales. Retail stores make money by selling products. If a company has a network of stores in various geographical locations, the company will find it difficult to understand the customer needs and market potentials in these various locations. In this work, we will use a company's sales data to understand the factors influencing sales, such as the unemployment rate, product popularity, and holidays in different stores located in different locations, so that resources can be managed wisely to maximise sales and earn maximum profit. These analytics will assist businesses in understanding market conditions and the various factors influencing sales. Thus, based on the aforementioned factors, customer demands can be easily observed. When data is properly stored, managed, and analysed, the owners will begin to see patterns, insights, and the big picture of the company, and accordingly, the necessary suitable actions can be taken. This will also aid in optimising the company's operations and maximising sales and profit. Furthermore, this data can be used to forecast a company's future sales so that retailers have a clear understanding of the company's future.

2. LITERATURE SURVEY

2.1 Existing Problem:

Existing systems necessitate the installation of software to analyse and visualise sales data so the data visualisation takes longer. The primary goal is to design and develop a web application that will assist businesses and retailers in analysing and visualising enterprise sales data in the form of graphs. We don't need any storage space because we're creating a web application that can be used by any device with an internet connection once it's deployed. We can quickly see our output by using websites. There is no need to install any software in this application because everything is available on any device with an internet connection. It is not necessary to switch between applications. We can analyse in less time, so the process is faster. The main objective is to create a web application that allows a company or retailer to analyse and visualise sales data as well as compare daily sales by different salespersons.

2.2 References:

- [1] C. Chen, "Information visualization," WIREs Computational Statistics, vol. 2, July/August, 2010, pp. 387-403.
- [2] Liu, Jiaying, Tao Tang, Wei Wang, Bo Xu, Xiangjie Kong, and Feng Xia. "A survey of scholarly data visualization." IEEE Access 6 (2018): 19205-19221.
- [3] Sadiku, Matthew, Adebowale E. Shadare, Sarhan M. Musa, and Cajetan M. Akujuobi. "Data visualization." International Journal of Engineering Research And Advanced Technology (IJERAT) 2, no. 12 (2016)

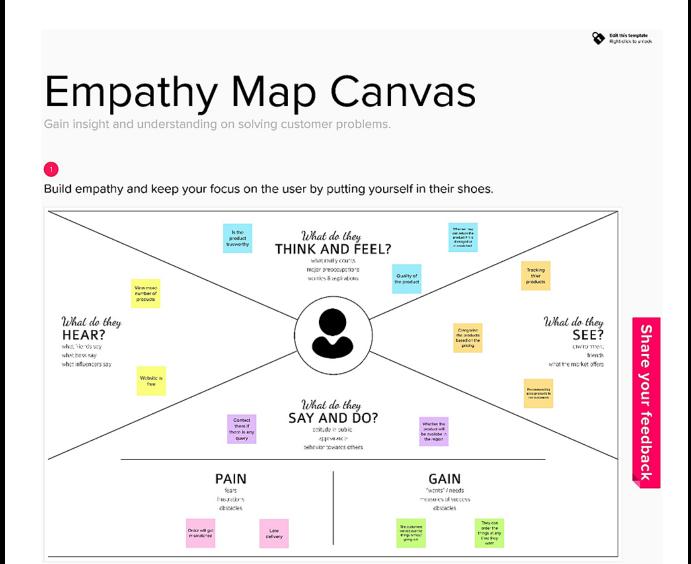
- [4] Jones, Amber Spackman, Jeffery S. Horsburgh, Douglas Jackson-Smith, Maurier Ramírez, Courtney G. Flint, and Juan Caraballo. "A web-based, interactive visualization tool for social environmental survey data." Environmental Modelling & Software 84 (2016): 412-426.
- [5] Ramasubramanian, Laxmi, and Jochen Albrecht. "Essential methods for planning practitioners: Skills and techniques for data analysis, visualization, and communication." (2018).
- [6] Embarak, Ossama. Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems. Apress, (2018).
- [7] K. Manikanta Vamsi et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1042 012019
- [8] Fernando Perez. "The IPython notebook: a historical retrospective".

2.3 Problem Statement Definition:

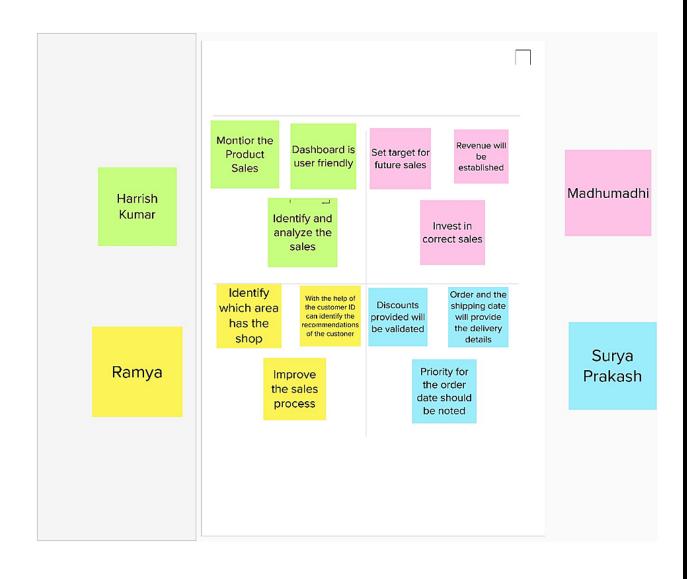
Every company would like a way to get an idea of the sale of different products in the company while also getting analytics and visualisation of sales by different salesperson in the organisation and analysis of whether the total profit matches the number of salesperson in the company. There are some existing softwares that assist businesses in tracking and managing customer interactions and sales. The major issues with almost all of these softwares were that the learning curve was steep, they were expensive, and they required skilled employees to use them.

3. Ideation and Proposed Solution

3.1 Empathy Map Canvas:



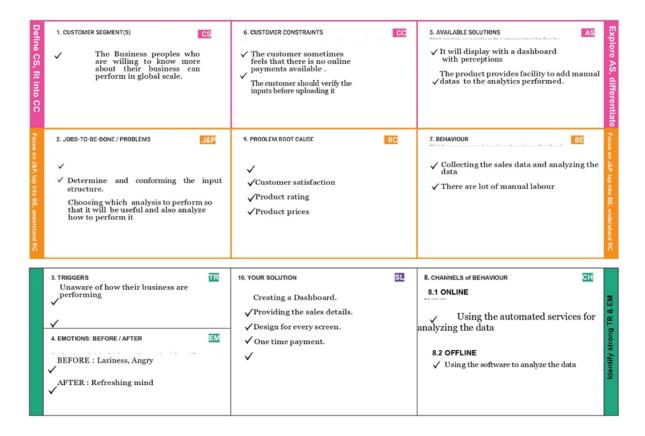
3.2 Ideation & Brainstorming:



3.3 Proposed Solution:

S.N	Parameter	Description
0.		
1.	Problem Statement(Problem to	During the time of pandemic shopping online is
	be solved)	a majorneed.Peoples were not able to walk
		outside and buy the things. There are few things
		like Customer and Product
		AnalysisoftheGlobalStore.
2.	Idea/Solution description	The idea will be done using the IBM
		platformand it will display the previous year
		recordsconsistingofproduct names,andtheir
		categoriesandrepresentitinagraph.
3.	Novelty/Uniqueness	Wewillprovidegifts/couponstothecustomersbuyi ng more than a certain price so that thesales will increase and the door delivery will befaster.
4.	SocialImpact/Customer Satisfaction	All the customers should be aware of the shopshaving the available products and the nearbyshoplocationstobuythethingseasier.
5.	Business Model(RevenueModel)	This is the method which implies salesnumbers. It is useful to determine which aretheproductsthataresellingmore.
6.	Scalability of the Solution	Thepricesofthesimilarproductsallover the world will be kept as same so this willbescalabletoallthecustomers.

3.4 Problem Solution fit:



4 .REQUIREMENT ANALYSIS

4.1 Functional requirement:

S No.	FunctionalRequirement	SubRequirement(Story/Sub-Task)
	(Epic)	
FR-1	User	Registration through Form Registration through Gmail
	Registration	Registration through Linked IN
FR-2	User Confirmation	Confirmation via Email
		Confirmation via OTP
FR-3	Dataset	DatasetuploadtoCognosAnalyticsTool.
FR-4	Visualize/Analyse	To analyse the dataset, drag and drop columns.
FR-5	Create Dashboards	Create Charts, Graphs, Tables, etc.
FR-6	LogOut	When the Dashboard shave beend own loaded,
		logout.

4.2 Non-functionalRequirements:

Following are the non - functional requirements of the proposed solution.

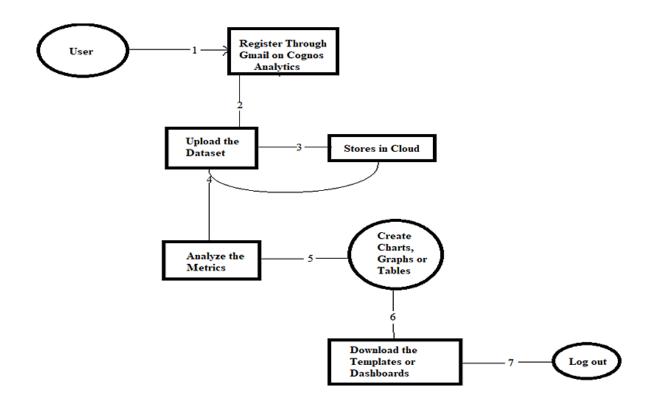
FRNo.	Non-	Description
	FunctionalRequirement	
NFR-1	Usability	The dashboard can be accessed by the user until the proper
		store sales data set is present.
NFR-2	Security	The Dash boards Templates are accessible to anyone
		with the proper Log Incredentials.
NFR-3	Reliability	Templates are dependable because we upload and access them
		via the cloud.
NFR-4	Performance	The usercaneasilydragtoanymetricstheywantto
		view,anditworksasexpected.
NFR-5	Availability	Everyonewhowantsto learnmoreaboutsalesdatacanaccessit
		forfree.
NFR-6	Scalability	Theusercan
		change the metric satany time with dashboards/templates because the
		yarevery
		scalable.

5. PROJECT DESIGN

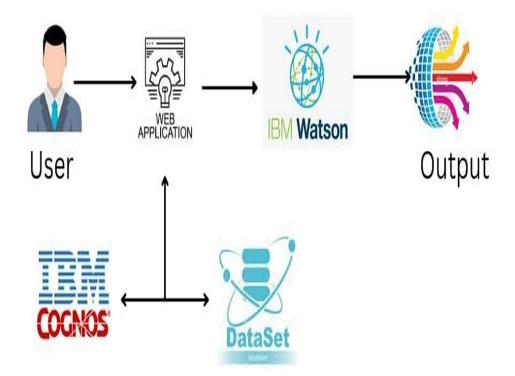
5.1 Data Flow Diagram:

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.

Example: Structural Flow Diagram:



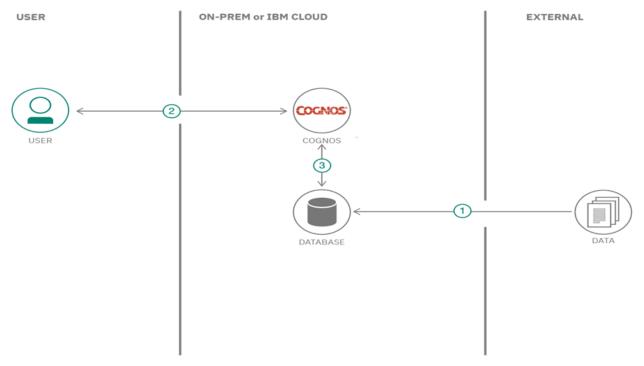
DATA FLOW DIAGRAM FOR GLOBAL SALES DATA ANALYTICS:



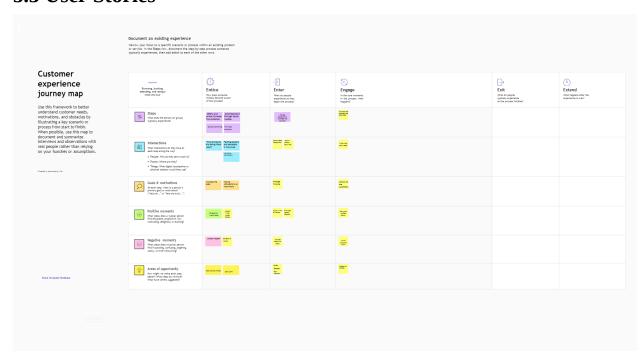
Flow structures:

- Customer Registers themselves through Mail ID on Cognos Analytics.
- Upload various kinds of Datasets.
- Those Datasets are saved in My Content section of Cognos Analytics tool.
- Select the Metrics and Visualize it.
- Create colourful Charts , Graphs or Tables.
- Save the Template and Download it for Business Decision purpose.
- Finally ,Log Out.

5.2 Solution and Technology Architecture: Technical Architecture:



5.3 User Stories



6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

User Type	Functional Requireme nt	User Sto ry	User Story/ Task	Acceptan cecriteria	Priori ty	Relea se
	(Epic)	Number				
Customer (Mobile user)	Registration	USN-1	As a user, Ican registerfor the application by entering my email, password, and confirmi ng my password.	I can access my account /dashboard.	High	Sprint - 1
		USN-2	As a user, I will receive confirmation email onceIhaveregistered for the application.	I can receivemy confirmati onemail & click confirm.	High	Sprint-1
		USN-3	As a user, I can registerfor the application through Facebook.	I can register& access the dashboard with Facebo okLogin .	Low	Sprint - 2
		USN-4	As a user, I can registerfor the application through Gmail.		Medi um	Sprint - 1
	Login	USN - 5	As a user, I can log intothe application by entering email & password.		High	Sprint - 1
	Dashboard					

Customer (Web user)	Login	USN-1	I can sign up for the application as a user byproviding my email address, a password for login confirmation.	I can access my dashboard and account.	High	Sprint-1
Customer CareExecuti ve	Chat Box	USN-1	It is accessible and usable by responsible individuals.	I can easily access using an applicatio n.	High	Sprint- 2
Administrator	Calling	USN-2	It is accessible and usable by responsible individuals.	I can easily access usingan application.	High	Sprint- 2
	Mail	USN-3	It is accessible and usable by responsible individuals.	I can easily access using an applicatio n.	High	Sprint-

6.2 Sprint Delivery Schedule:

Use the below template to create product backlog and sprint schedule

Spri nt	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priori ty	Team Members
Sprin t-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-1	Login	USN-2	As a user,I need validcredentials to log in tomy application.	1	High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM

Sprin t-1	Data Collection		USN-3	N-3 As a user, I need to gatherthe data in the formof CSV/XLS and clean the data		High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-2	Upload dataset US		USN-4	As a user, I can viewthe data of the products	1	Low	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin	Data	US	As a user,	I need	3	High	Harrish Kumar
t-2	Preparation	N-5		for	3		G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-2	Data visualization	US N-6		I can easily the data in the aarts.	4	Medium	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-3	Dashboard	US N-7	summaryo sales by th	As a user,I can view the summaryof theproduct sales by the help dashboard.		Medium	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprint-3	- Dashboard	US 8	visuali waytha insight sales b catego	ser, I must plan zations in a at I'm able to gain as regarding the ased uponthe ry of sales and the tive region	4	High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakash M

Spri nt	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprin t-3	Dashboard	USN-9	As a user, I must be able to gain insights from the charts/graphs through a variety of relationships established in the dashboard.	4	Medium	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t- 4	Prediction	USN-10	As a user, I see the prediction of the specific product's future sales expectation.	4	Medium	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-4	Report	USN-11	As a user, Ican view the listof categorized products and their detailsas a report.	5	High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM
Sprin t-4	Story	USN-12	As a user, I can view theproduct and customer description and more additional information as a story.	5	High	Harrish Kumar G M, Madhumadhi I, Ramya S, SuryaPrakashM

7. CODING & SOLUTIONING

7.1 Feature 1

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
import sklearn.metrics as sm
%matplotlib inline
df = pd.read_csv("Global_Superstore2.csv", encoding = 'unicode_escape', engine = 'python')
df

7.2 Feature 2

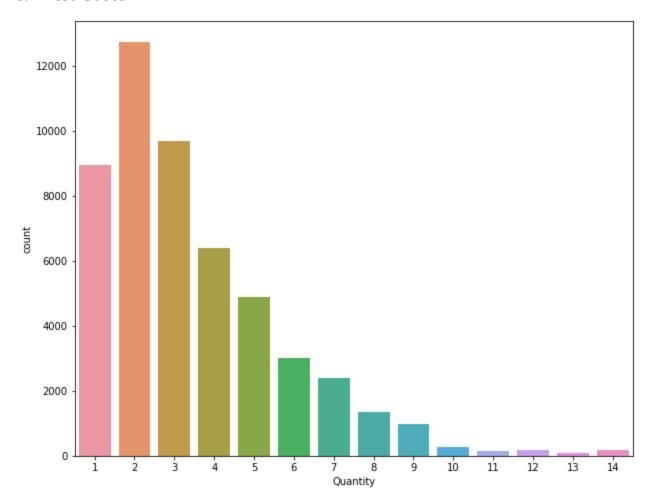
```
df.head()
df.describe()
df.info()
df.columns
df.shape
# Checking missing values
df.isna().sum()
# Generating descriptive statistics summary
df.describe().round()
# Setting the figure size
plt.figure(figsize=(10, 8))
# countplot: Show the counts of observations in each categorical bin using bars
sns.countplot(x='Quantity', data=df)
# Display the figure
plt.show()
col = list(df.columns)
classes = df["Region"].unique()
xdata = df.iloc[:, 0:7].values
ydata = df.iloc[:, 7].values
plt.figure(figsize = (16, 9))
sns.countplot(classes, palette = 'rocket')
plt.xticks(rotation=90)
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

```
from sklearn.linear_model import LogisticRegression
nb_samples = 1000
x, y = make_classification(n_samples=nb_samples, n_features=2,
n_informative=2, n_redundant=0, n_clusters_per_class=1)
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)
model = LogisticRegression()
model.fit(xtrain, ytrain)
print(accuracy_score(ytest, model.predict(xtest)))
```

8. TESTING

Before project deliverables are made available to the client, they must be tested to ensure their accuracy and completeness. It is a crucial component of quality control and aids in ensuring that the finished product lives up to consumer expectations. Typically, there are four phases. a project: acceptability testing, system testing, integration testing, and unit testing. Each of these stages has a distinct goal and concentrates on various facets of the project. A crucial part of quality management is testing. It is the process of ensuring that a project adheres to its specifications and works as predicted. Testing assists in ensuring that the project is appropriate for its intended use and will satisfy the needs of the client or customer.

8.1 Test Cases



8.2 USER ACCEPTANCE TESTING

The last testing phase prior to software production is user acceptability testing. Users who test the software and its user interface utilise it to provide feedback (UI). In manual UAT, people simulate real-world scenarios and evaluate how the software responds and functions. Automated test-case scenarios are another way to simulate a user's experience. It's crucial to plan ahead because UAT is expensive and it can be challenging to combine manual and automated testing during this phase. It's wise to be aware of some of the potential difficulties as a testing plan is developed and a schedule is formed.

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the Global Sales Data Analytics project at the time of the release to User Acceptance Testing(UAT).

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and howthey were resolved.

Resolution	Severity1	Severity2	Severity3	Severity4	Subtotal
ByDesign	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	0	0	1	1
Totals	24	9		1	1

3.Test Case Analysis

11 26 71

This report shows the number of test cases that have passed, failed, and untested

Section	TotalCases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

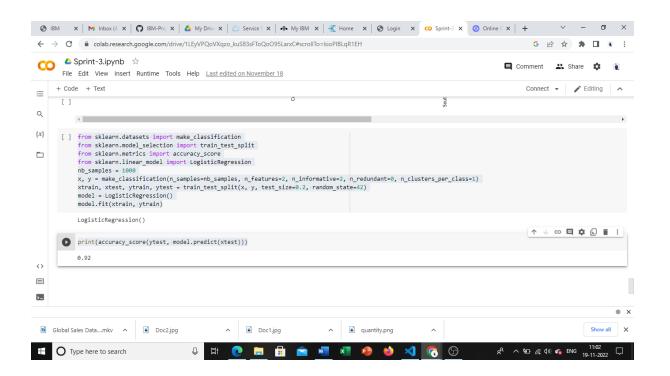
9.RESULTS

9.0.PERFORMANCE METRICS

Performance testing is a testing method that evaluates an application's speed, scalability, and stability under a certain workload. It makes the programme ready for commercial release and aids with ensuring the quality of the software.

- i. Speed The speed at which the application responds.
- ii. Scalability The maximum user load that the application can handle.
- iii. Stability The condition of the application under varying loads.

ACCURACY FOR DATASET:



10.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

• Data analytics helps an organization make better decisions

Decisions within businesses are frequently decided more on intuition than on facts and figures. Lack of access to reliable data that could facilitate improved decision making may be one of the causes of this. In order to help executives make better decisions, analytics can assist in converting the available data into useful information. If less bad decisions are made, this could be a source of competitive advantage since bad decisions can negatively affect a variety of factors, including business growth and profitability.

Increase the efficiency of the work

To assist in achieving certain business goals, analytics can analyse vast amounts of data fast and provide it in a structured way. By enabling the management to communicate with the staff the insights from the analytics results, it promotes an environment of efficiency and cooperation. A company's weaknesses and potential areas for improvement become clear, and steps can be taken to improve workplace efficiency overall and boost productivity.

DISADVANTAGES:

• Lack of alignment within teams

Within an organisation, there is a lack of coordination between several teams or divisions. The insights produced by these teams have little value, though, or have little effect on organisational measures. This might be the result of teams

operating in "silos," each employing its own processes and being isolated from other departments. The analytics team should concentrate on providing the correct answers to the business's questions, and the findings from data analytics teams .

• Lack of commitment and patience

The implementation of analytics solutions is not difficult, but they are expensive and take time to pay for themselves. Setting up protocols and procedures to begin collecting the data may take some time, particularly if there is no previous data. By their very nature, analytics models gain accuracy over time and demand commitment to be put into practise. Because business customers don't always see outcomes right away, they occasionally grow impatient, which undermines their belief in the models.

11. CONCLUSION

A special period in the history of data analysis has resulted from the accessibility of data, low-cost commodity hardware, and innovative information management and analytical tools. We now have the tools necessary to analyse astounding data sets fast and affordably for the first time in history thanks to the convergence of these tendencies. These abilities are neither merely hypothetical nor unimportant. They represent a true advancement and a great chance to achieve significant increases in effectiveness, productivity, income, and profitability. An essential tool for companies all over the world, sales analytics. It keeps our company current. Without this essential component, our company won't survive long in a very cutthroat sector. greater insights are provided by data visualisation. Depending on the business we are operating and locating.

12. FUTURE SCOPE

Global Sales Data Analytics gets rid of human labour and speculation. whether it be selecting the appropriate content, organising marketing initiatives, or creating products. Organizations can use the data analytics insights they uncover to make wise decisions. resulting in improved outcomes and satisfied clients. We can quickly determine the profit and loss for the company by visually representing the data using a bar chart, pie chart, etc.

In the future, we will be able to tell when a consumer buys the following product and how long it takes to deliver it. We will be able to forecast the sales and profit for the upcoming quarter as we have a better understanding of the kinds of products customers seek out, product returns, etc.

1	3	A	P	P	\mathbf{F}	VI	וח	ľX
		\neg			., 1			_

13.1 GITHUBAND PROJECT VIDEOLINK

Github Link: https://github.com/IBM-EPBL/IBM-Project-33101-1660214588

Project Video Link:

https://drive.google.com/file/d/1UDrxm8XjLXtba2 aEremUxtsGctOuKOh/view?usp=share li nk