Name: Alston Fernandes **Roll no.: 19** **Batch: A**

Experiment No: 7 & 8

Topic:	Data Analysis & Data Visualization using Hive/PIG/R/Tableau.
Prerequisite:	Basic knowledge of Data Analysis & data visualization tools like Hive/PIG/R/Tableau etc.
Mapping With COs:	**CSL702.6**
Objective:	Able to be used for data analysis, statistical modeling, and machine learning tasks.
Outcomes:	Analyze the datasets using Data Analysis tools & generate reports using data visualization tools capabilities.
Instructions:	This experiment is a compulsory experiment. All the students are required to perform this experiment individually.
Deliverables:	**Submission:** 1. Installation steps snapshots for Windows 10 systems Step 1: Search **Download Tableau** in Internet Browser, you will be shown with the installation page on the first link or use the following link - https://www.tableau.com/products/desktop/download  You will be redirected to the following page Step 2: Click on Start your free trial button to install in your Windows System.

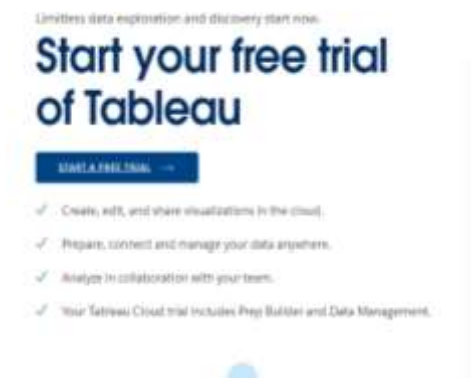
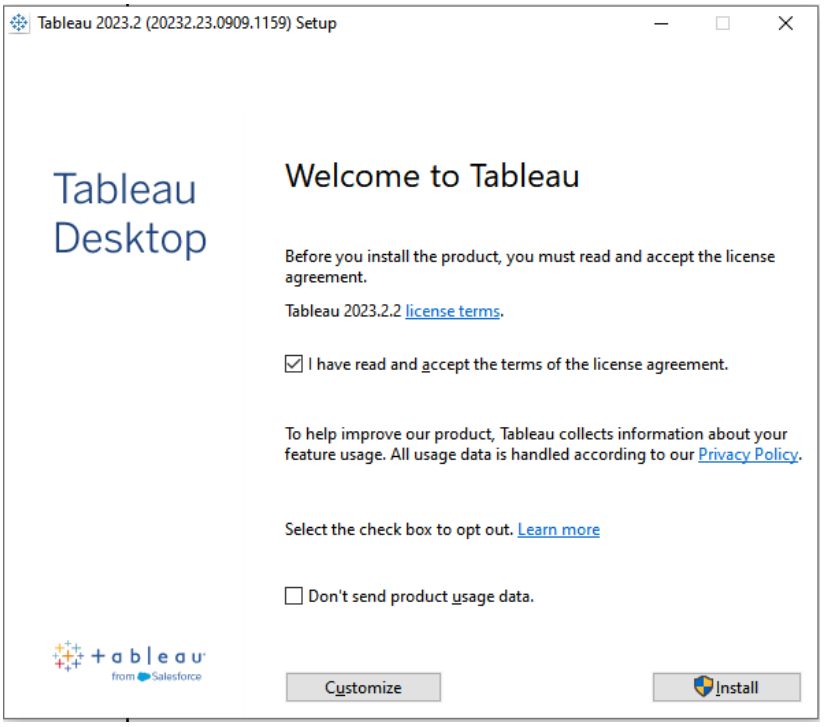


Tableau software will get download into your local system.

Step 3: Installing the software in local machine -



Step 4 - Tableau will be installed in the system and will be redirected with the page below -

2. Detail about Datasets

Dataset used for visualization

1. Customers Database -

customer_code	custmer_name	customer_type
Cus001	Surge Stores	Brick & Mortar
Cus002	Nomad Stores	Brick & Mortar
Cus003	Excel Stores	Brick & Mortar
Cus004	Surface Stores	Brick & Mortar
Cus005	Premium Stores	Brick & Mortar
Cus006	Electricalsara Stores	Brick & Mortar
Cus007	Info Stores	Brick & Mortar
Cus008	Acclaimed Stores	Brick & Mortar
Cus009	Electricalsquipo Stores	Brick & Mortar
Cus010	Atlas Stores	Brick & Mortar
Cus011	Flawless Stores	Brick & Mortar
Cus012	Integration Stores	Brick & Mortar
Cus013	Unity Stores	Brick & Mortar
Cus014	Forward Stores	Brick & Mortar
Cus015	Electricalsbea Stores	Brick & Mortar
Cus016	Logic Stores	Brick & Mortar
Cus017	Epic Stores	Brick & Mortar
Cus018	Electricalslance Stores	Brick & Mortar

2. Market's Database -

markets_code	markets_name	zone
Mark001	Chennai	South
Mark002	Mumbai	Central
Mark003	Ahmedabad	North
Mark004	Delhi NCR	North
Mark005	Kanpur	North
Mark006	Bengaluru	South
Mark007	Bhopal	Central
Mark008	Lucknow	North
Mark009	Patna	North
Mark010	Kochi	South
Mark011	Nagpur	Central
Mark012	Surat	North
Mark013	Bhopal	Central
Mark014	Hyderabad	South
Mark015	Bhubaneshwar	South
Mark097	New York	
Mark999	Paris	

3. Product's Database -

product_code	product_type
Prod001	Own Brand
Prod002	Own Brand
Prod003	Own Brand
Prod004	Own Brand
Prod005	Own Brand
Prod006	Own Brand
Prod007	Own Brand
Prod008	Own Brand
Prod009	Own Brand
Prod010	Own Brand
Prod011	Own Brand
Prod012	Own Brand
Prod013	Own Brand
Prod014	Own Brand
Prod015	Own Brand
Prod016	Own Brand
Prod017	Own Brand
Prod018	Own Brand

4. Transaction's Database -

product_code	customer_code	market_code	order_date	sales_qty	sales_amount	currency
Prod001	Cus001	Mark001	2017-10-10	100	41241	INR
Prod001	Cus002	Mark002	2018-05-08	3	-1	INR
Prod002	Cus003	Mark003	2018-04-06	1	875	INR
Prod002	Cus003	Mark003	2018-04-11	1	583	INR
Prod002	Cus004	Mark003	2018-06-18	6	7176	INR
Prod003	Cus005	Mark004	2017-11-20	59	500	USD
Prod003	Cus005	Mark004	2017-11-22	36	250	USD
Prod003	Cus005	Mark004	2017-11-23	39	21412	INR
Prod003	Cus005	Mark004	2017-11-27	35	19213	INR
Prod003	Cus005	Mark004	2017-11-28	310	170185	INR
Prod003	Cus005	Mark004	2017-11-29	184	101194	INR
Prod003	Cus005	Mark004	2017-11-30	35	19213	INR
Prod004	Cus005	Mark004	2017-11-29	17	9426	INR
Prod004	Cus005	Mark004	2017-12-19	1	218	INR
Prod005	Cus005	Mark004	2018-08-07	5	3093	INR
Prod003	Cus006	Mark004	2017-12-04	58	30306	INR
Prod005	Cus006	Mark004	2018-06-29	38	52319	INR
Prod005	Cus006	Mark004	2018-07-02	93	126296	INR

3. Data Analysis

The data required for this analysis will be sourced from the following tables:

Transactions: Contains detailed information about each sales transaction, including product IDs, customer IDs, market IDs, revenue, and quantity sold.

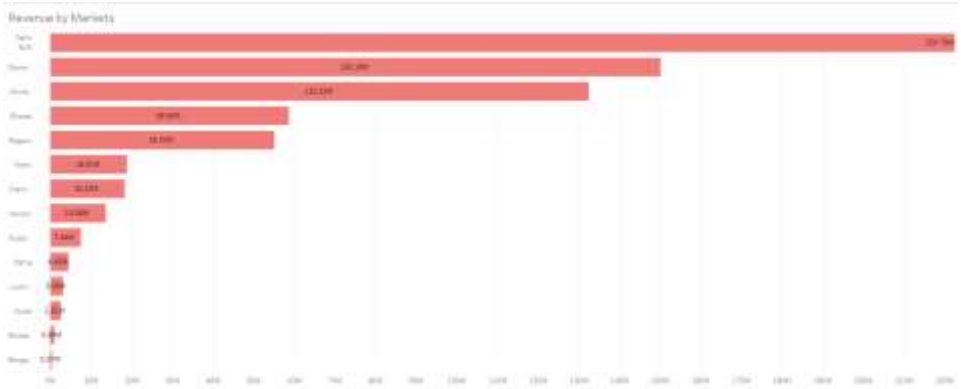
Customers: Contains customer information such as names, addresses, and contact details.

Date: Contains date-related information, allowing for date-based analysis.

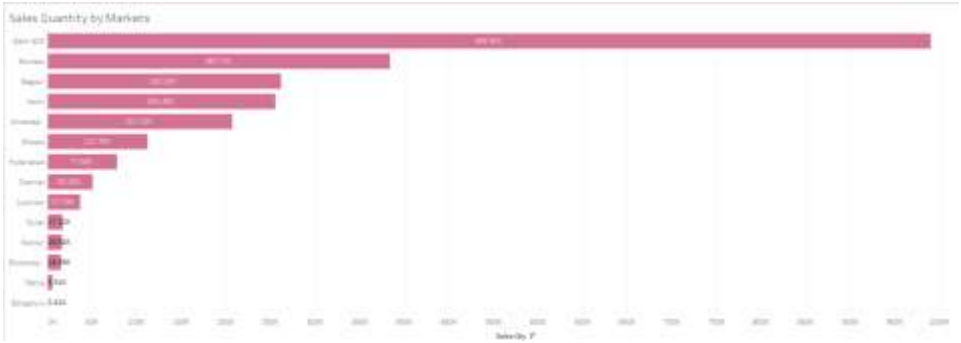
Markets: Provides details about the markets in which Company XYZ operates.

	Products: Contains information about the products Company XYZ sells, including names, descriptions, and pricing. Dashboard will visualize the following requirements - 1. Total Revenue: Calculate the total revenue generated by Company XYZ over the entire dataset period. 2. Total Sales Quantity: Determine the total quantity of products sold by Company XYZ. 3. Revenue by Markets: Display the revenue generated by each market within Company XYZ's operations. 4. Sales Quantity by Markets: Provide a breakdown of the quantity of products sold in each market. 5. Top 5 Customers: Identify and present the top 5 customers who have contributed the most to the company's revenue. 6. Top 5 Products Sold: Identify and present the top 5 products that have been sold the most in terms of quantity. 7. Year-on-Year Sales Insights: ● Calculate and visualize year-on-year total revenue to track revenue growth or decline. ● Analyze year-on-year sales quantity to understand product demand trends over time. ● Provide an option to filter data by specific product IDs and months for more detailed insights. 8. Customer Analysis: Allow users to select specific customers and view their purchase history and contribution to revenue over the years.
	4. Data Visualization **Dashboard -**

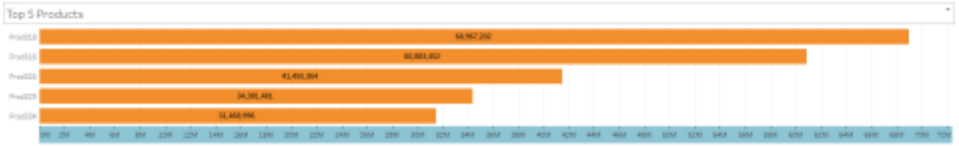
Revenue by Markets -



Sales Quantity by Markets –



Top 5 Customers –

	Top 5 Products Sold –  **Revenue Per Year -** 
Conclusion:	Students will be able to successfully perform Data Analysis & Data Visualization using Hive/PIG/R/Tableau.
References:	Put the reference of resources used to perform this experiment.

Don Bosco Institute of Technology
Department of Computer Engineering

Academic year – 2023-2024

Big Data Analytics

Assessment Rubric for Experiment No. : 07 & 08

Performance Date : 11/10/23
Submission Date : 11/10/23

Title of Experiment : Data Analysis & Data Visualization using Hive/PIG/R/Tableau

Year and Semester : IV^{rh} Year and VIIth Semester
Batch : A
Name of Student : Alston Fernandes
Roll No. : 19

Performance	**Poor**	**Satisfactory**	**Good**	**Excellent**
	1 point	2 points	3 points	4 points
Results and Documentations	**Poor**	**Satisfactory**	**Good**	**Excellent**
	1 point	2 points	3 points	4 points
Viva	**Poor**	**Satisfactory**	**Good**	**Excellent**
	1 point	2 points	3 points	4 points
Timely Submission	**Submission beyond 7 days of the deadline**	**Late submission till 7 days**	**Submission on time**	
	1 points	2 points	3 points	

Faculty Incharge : Ms. Sana Shaikh