CLOUD APPLICATION DEVELOPMENT

Big Data Analysis with IBM Cloud Database

Project 5: Big Data Analysis

**PROBLEM STATEMENT:**

Dive into the world of big data analysis with IBM Cloud Databases. Uncover hidden insights from vast datasets, from climate trends to social patterns. Visualize your findings and derive valuable business intelligence. Embark on data-driven adventures, exploring the endless possibilities of big data!

**OBJECTIVE:**

The objective for the problem statement is to promote IBM Cloud Databases as a platform for big data analysis, highlighting its capabilities to uncover insights from large datasets, such as climate trends and social patterns. The statement encourages users to visualize their findings and leverage the platform for deriving valuable business intelligence, ultimately motivating individuals or businesses to embark on data-driven adventures and explore the potential of big data analytics using IBM Cloud Databases.

**DESIGN THINKING:**

Design thinking is a human-centred approach to problem-solving and project design that focuses on understanding user needs, ideating creative solutions, and iteratively refining those solutions. When applied to a project like "Big Data Analysis with IBM Cloud Databases," it can help ensure that the final solution meets the needs of both data analysts and the organization as a whole. Here's a design thinking approach for such a project:

* **DATA SELECTION:**

Data selection is defined as the process where data relevant to the analysis is decided and retrieved from the data collection.

Examples: **Marketing, Transportation, Government and public**

**administration, Business, Healthcare, Cybersecurity.**

* **DATABASE SETUP:**

Setting up IBM Cloud Databases for storing and managing large datasets involves several steps. Here's a high-level overview:

1. Create an IBM Cloud Account
2. Access IBM Cloud Dashboard
3. Select Databases
4. Choose a Database Service
5. Configure Database Instance
6. Set Up Access and Security
7. Connect and Use the Database
8. Load Data
9. Optimize Performance
10. Monitor and Manage

* **DATA EXPLORATION:**

Data exploration in big data analysis with IBM Cloud Databases involves analyzing large volumes of data to discover patterns, trends, and insights that can drive decision-making. The goal is to understand the data's structure, relationships, and characteristics to make informed choices regarding further analysis or processing. Here's how you can define data exploration in the context of IBM Cloud Databases and big data:

1. Understanding Data Structure and Formats
2. Descriptive Statistics
3. Data Profiling
4. Visualization
5. Sampling and Initial Analysis
6. Pattern Detection
7. Query Optimization and Performance Tuning
8. Feature Engineering
9. Temporal and Spatial Analysis
10. Interactivity and Collaboration

* **ANALYSIS TECHNIQUES:**

In big data analysis with IBM Cloud Databases, applying appropriate analysis techniques involves leveraging statistical analysis and machine learning to derive meaningful insights from the vast amount of data. Here are some techniques and examples:

1. Descriptive Statistics
2. Correlation Analysis
3. Regression Analysis
4. Clustering (e.g., K-means)
5. Classification (e.g., Decision Trees, SVM)
6. Anomaly Detection
7. Natural Language Processing (NLP)
8. Deep Learning
9. Geospatial Analysis
10. Sentiment Analysis
11. Association Rule Mining
12. Graph Analytics
13. Dimensionality Reduction (e.g., PCA)
14. Ensemble Learning

* **VISUALIZATION:**

Designing impactful visualizations to present analysis results in big data analysis with IBM Cloud Databases is crucial for effectively communicating insights. Here are some visualization ideas and techniques:

1. Bar Charts and Column Charts
2. Line Charts
3. Pie Charts
4. Scatter Plots
5. Heatmaps
6. Histograms
7. Area Charts
8. Box Plots
9. Word Clouds
10. Radar Charts

* **BUSINESS INSIGHTS:**

Interpreting analysis findings and deriving valuable business intelligence from big data analysis with IBM Cloud Databases involves extracting actionable insights that can drive business decisions and strategies. Here's how you can derive business intelligence and provide actionable recommendations:

1. Identify Key Trends and Patterns
2. Customer Segmentation
3. Optimize Operational Efficiency
4. Improve Product Development
5. Enhance Marketing Strategies
6. Forecasting and Demand Planning
7. Customer Satisfaction and Retention
8. Risk Management and Fraud Detection
9. Geographic Expansion Strategy
10. Personalized User Experience
11. Real-time Decision Making

**DEVELOPMENT PHASES:**

In development phase we follow the those steps:

**PROCEDURE:**

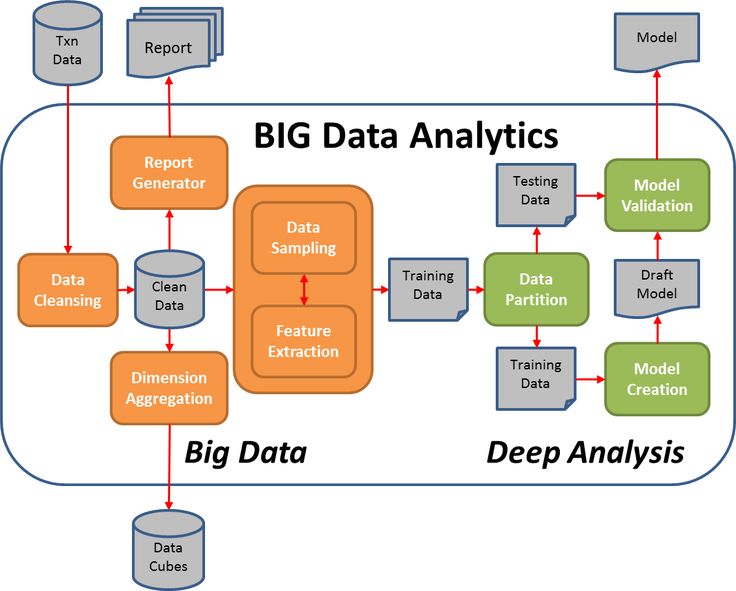
**STEP 1:** Download the shopping trending dataset from the thesis paper.

**STEP 2:** Analyse the data from the given dataset.

**STEP 3:** If any duplicates, null value occurrence in the dataset then those errors are removed immediately from the dataset.

**STEP 4:** To remove those anomalies we are using processes called DATA CLEANING and DATA TRANFORMATION.

**STEP 5:** In this step we will visualize the given dataset by using IBM WATSON STUDIO.



FIRSTLY WE NEED TO KNOW HOW TO ACCESS THE IBM CLOUD ACCOUNT TO DO THE FOLLOWING PROCESS IN THE IBM CLOUD.

To begin building a big data analysis solution using IBM Cloud Databases, follow these steps:

Create an IBM Cloud account. You can create a free account to get started.

➢ Choose the appropriate database service. IBM Cloud offers a variety of database services, including Db2 and MongoDB. Choose the service that is best suited for your needs.

➢ Set up a database instance. Once you have chosen a database service, you need to set up a database instance. This will involve choosing a region and a plan.

➢ Develop queries or scripts to explore and analyze the selected dataset. Once you have set up a database instance, you can start to develop queries or scripts to explore and analyze the selected dataset. You can use the database console to develop and execute queries and scripts.

➢ Perform basic data cleaning and transformation as needed. Before you can analyze your data, you may need to perform some basic data cleaning and transformation. This may involve removing duplicate records, correcting errors, and transforming the data into a format that is compatible with your chosen analysis tools.

**IBM DB2:**

Description: IBM Db2 is a family of data management products, including database servers, developed by IBM.

Role in the Project: Used for storing structured data, providing a reliable and scalable database solution

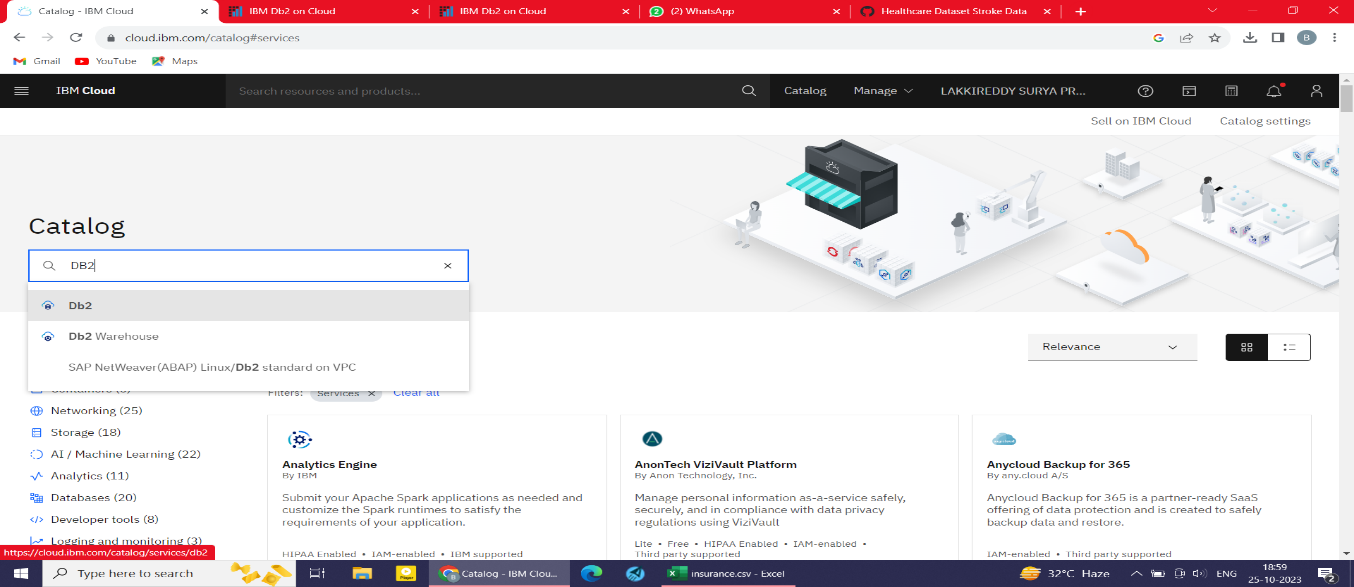
**To set up a database instance after choosing the database, you need to follow these steps:**

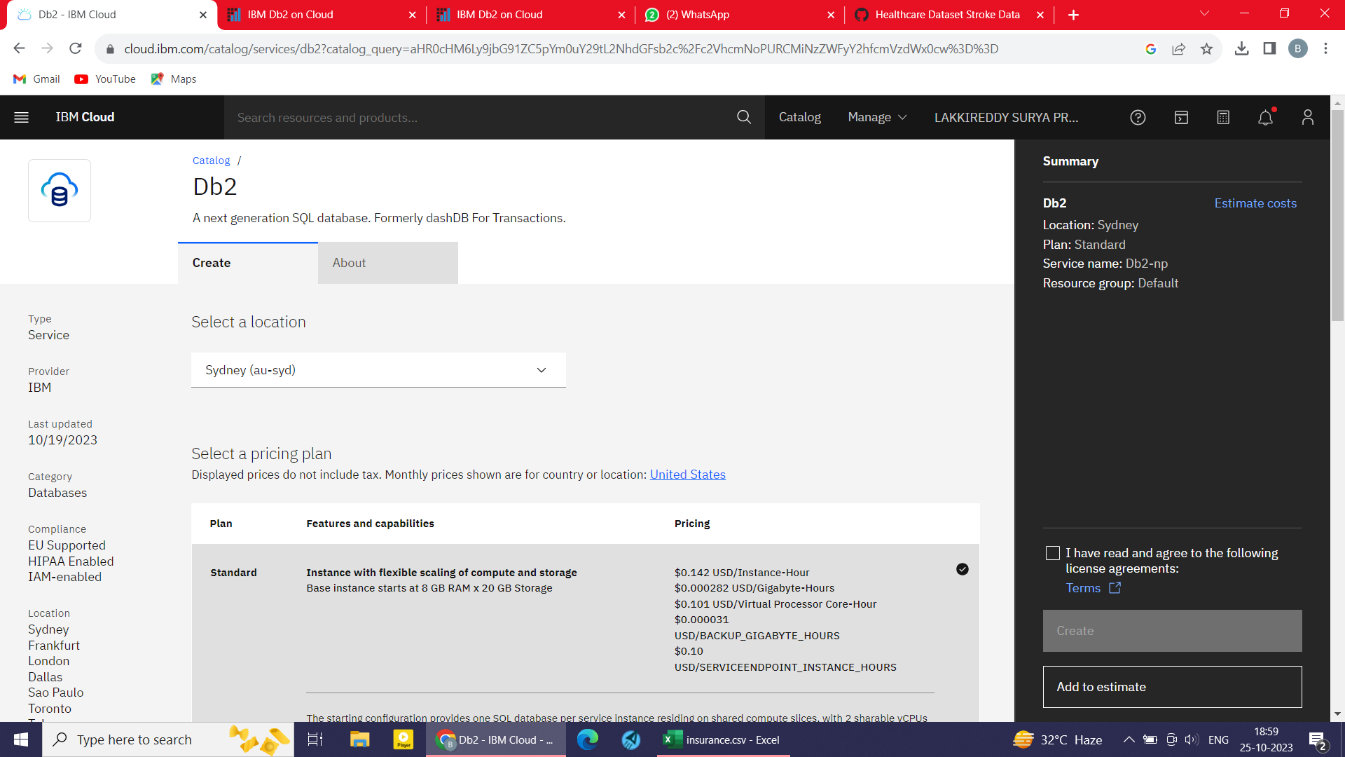
• Create a database instance. This can be done using the database management tool that you are using. For example, to create a database instance in Db2, you would use the CREATE DB command.

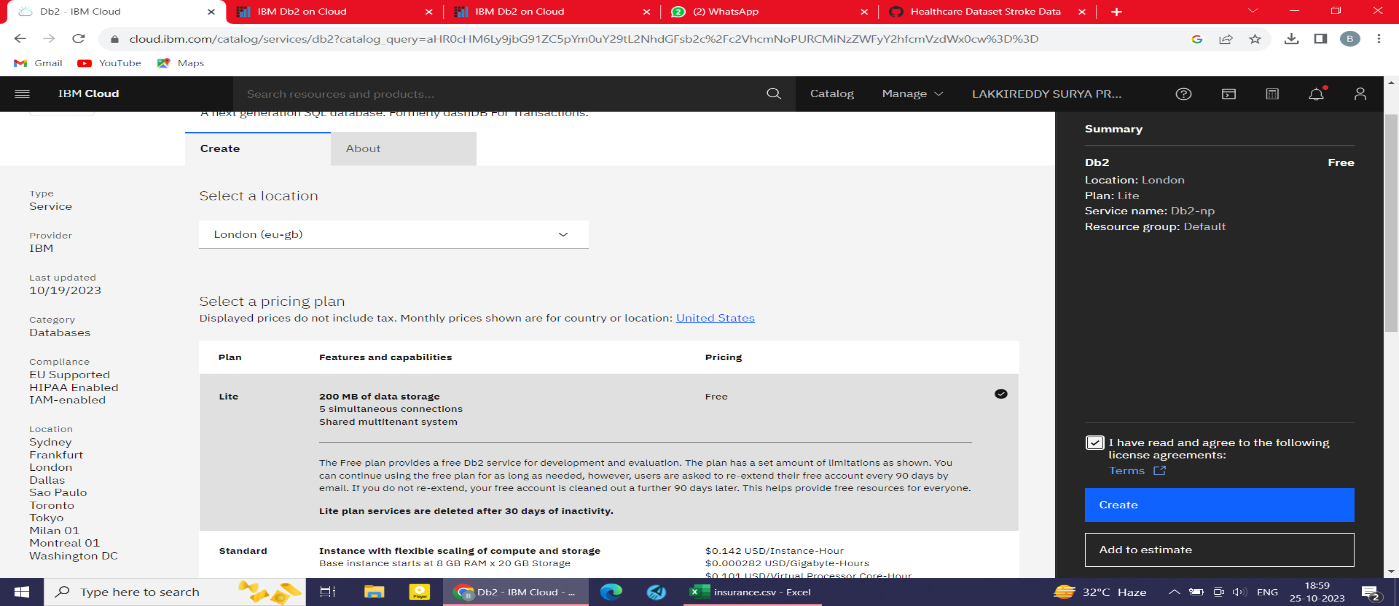
• Configure the database instance. This includes setting things like the database name, the database user accounts, and the database parameters.

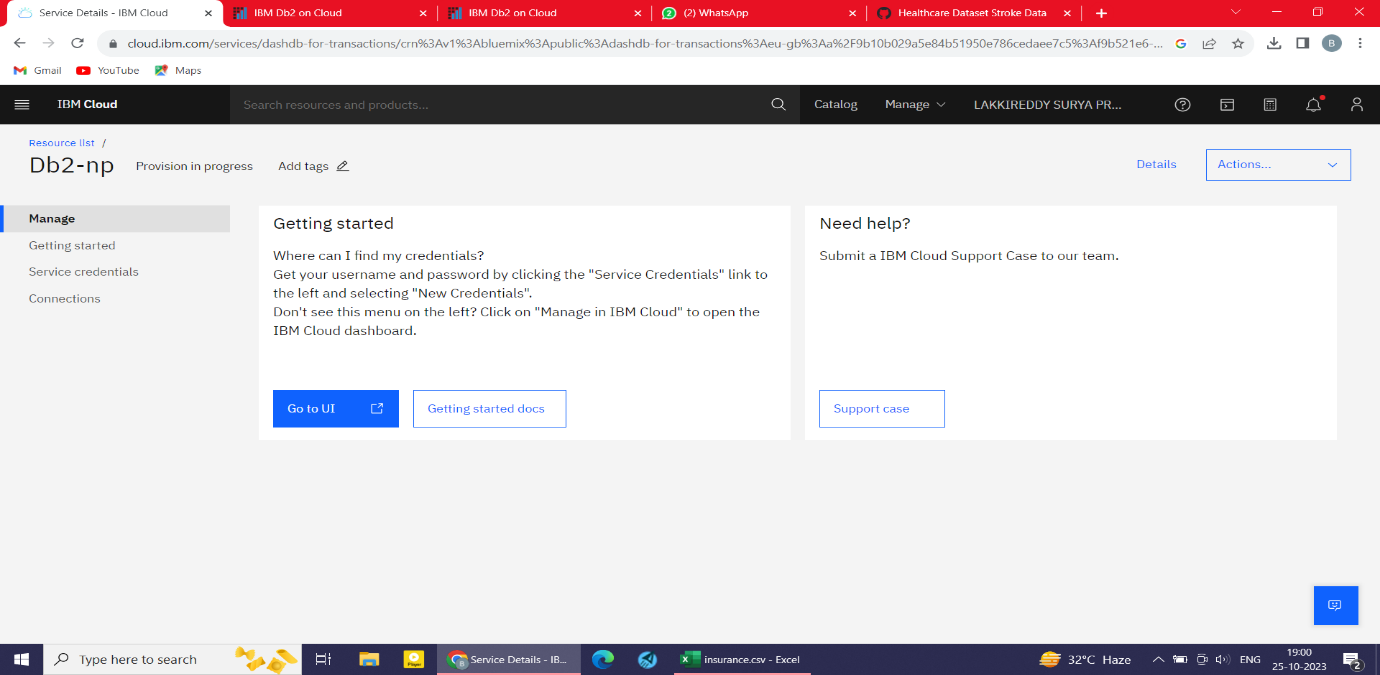
• Start the database instance. This can be done using the database management tool that you are using. For example, to start a database instance in Db2, you would use the START DB command.

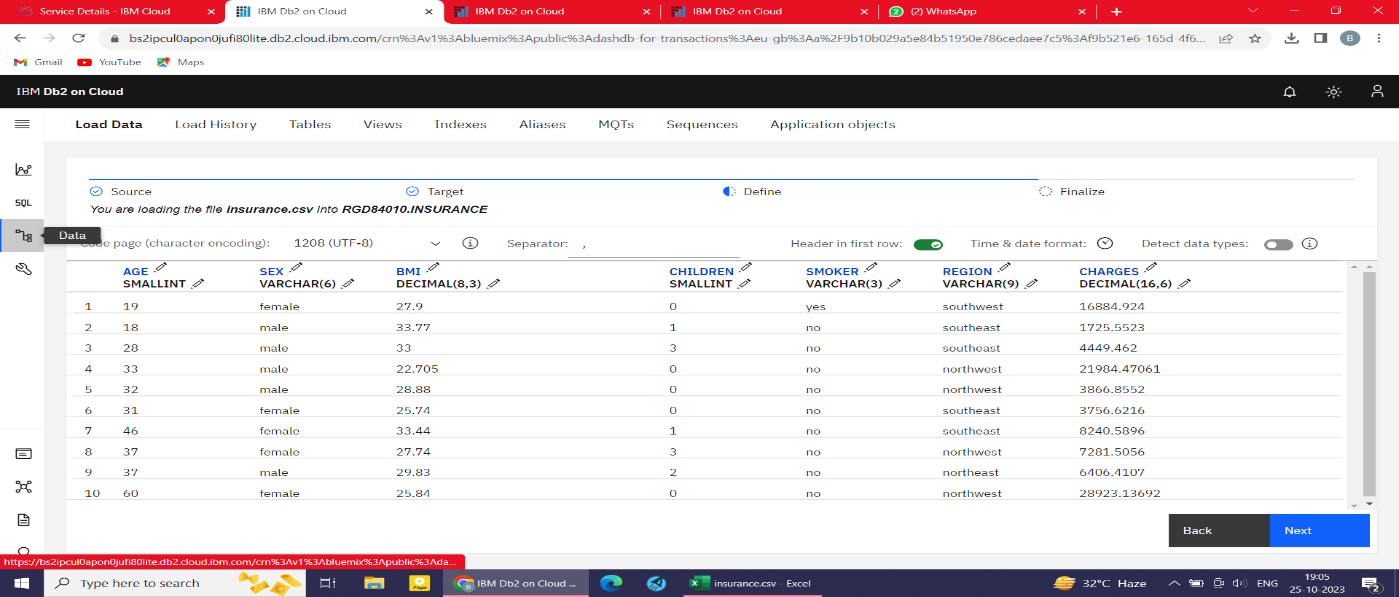
Once the database instance is created, configured, and started, you can start using it to store and manage your data.

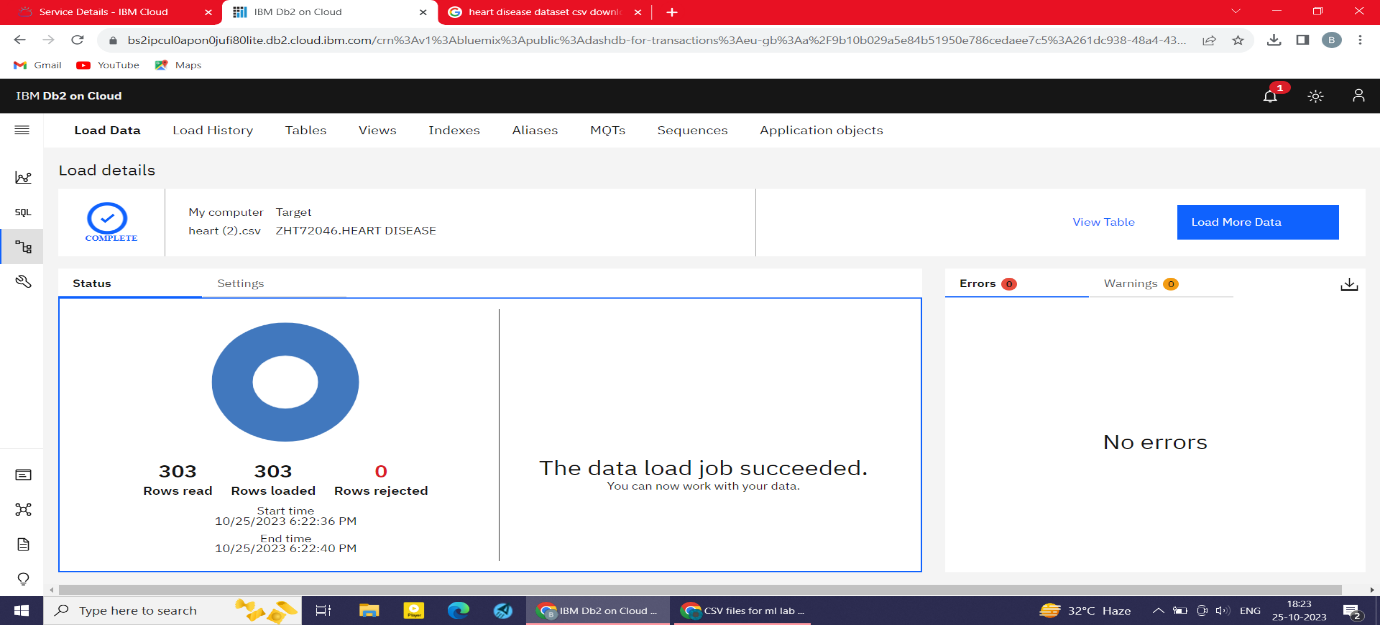












To continue building big data analysis solution using IBM Cloud Databases, we can apply more complex analysis techniques and visualize the results.

Machine Learning:

To use machine learning with IBM Cloud Databases, you can use IBM Cloud Watson Machine Learning. IBM Cloud Watson Machine Learning is a suite of machine learning services that can be used to build and deploy predictive models.

**IBM Watson Studio:**

IBM Watson Studio is a unified platform for data scientists, business analysts, and developers to collaborate on data preparation, machine learning, and data visualization. IBM Watson Studio provides a variety of tools and services that can help you to perform advanced analysis and to visualize your results.

To create an IBM Watson Studio project, follow these steps:

* Go to the IBM Cloud console and sign in to your account.
* Click Catalog and then click Watson Studio.
* Click Create project.
* Enter a name for your project and select a region.
* Click Create.

To add data to your project, you can:

* Upload files from your local computer.
* Connect to a database or cloud storage service.
* Use a Watson Studio service to generate data.

After Creating the IBM Watson Studio. Launch in the IBM Cloud Pak for data to visualize the results in the form of graphs and charts.

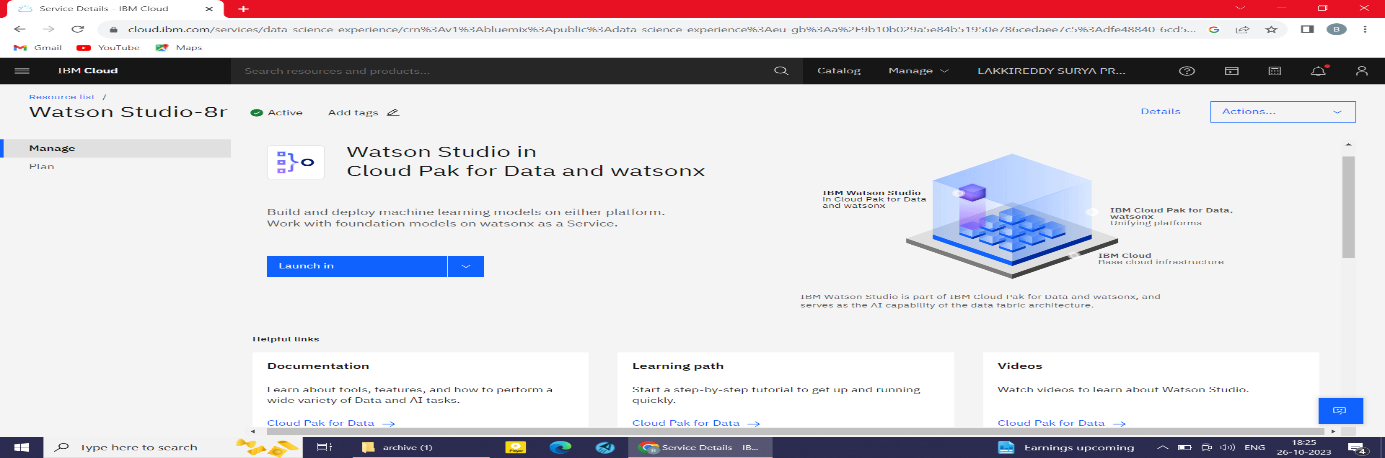
To launch the IBM Cloud Pak for Data to visualize the result in IBM Watson Studio, follow these steps:

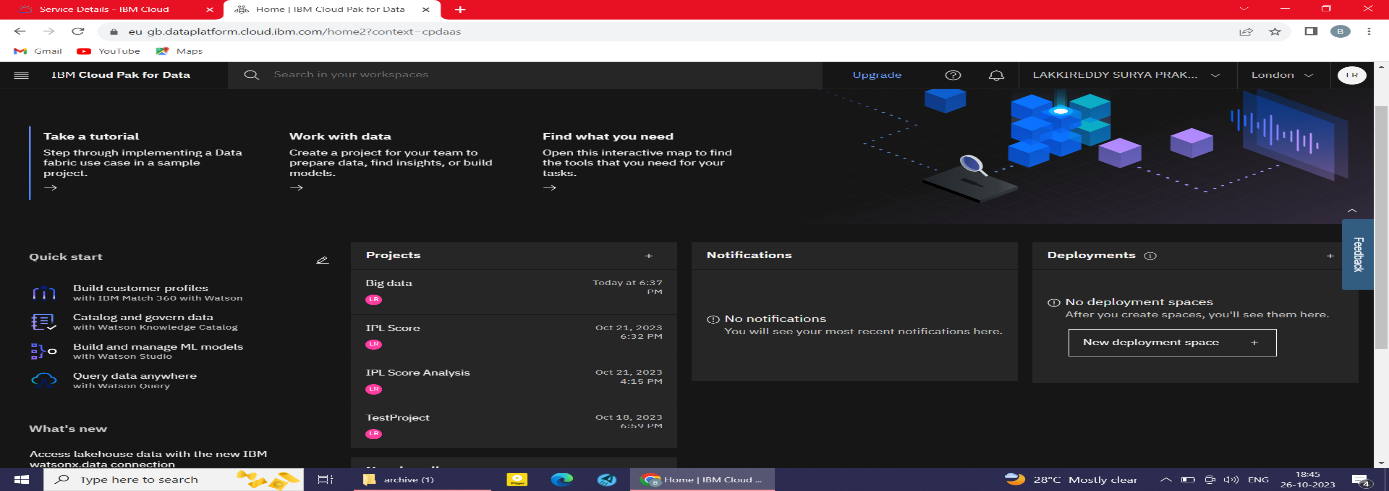
• Go to the Catalog tab in IBM Watson Studio.

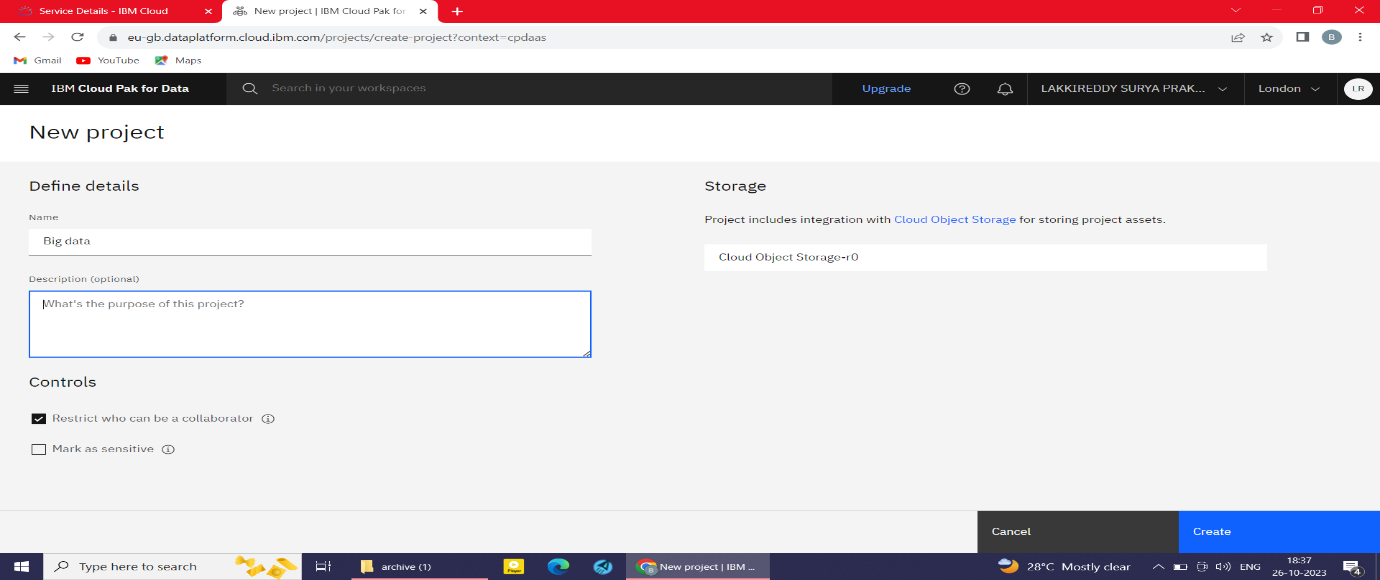
• Click IBM Cloud Pak for Data.

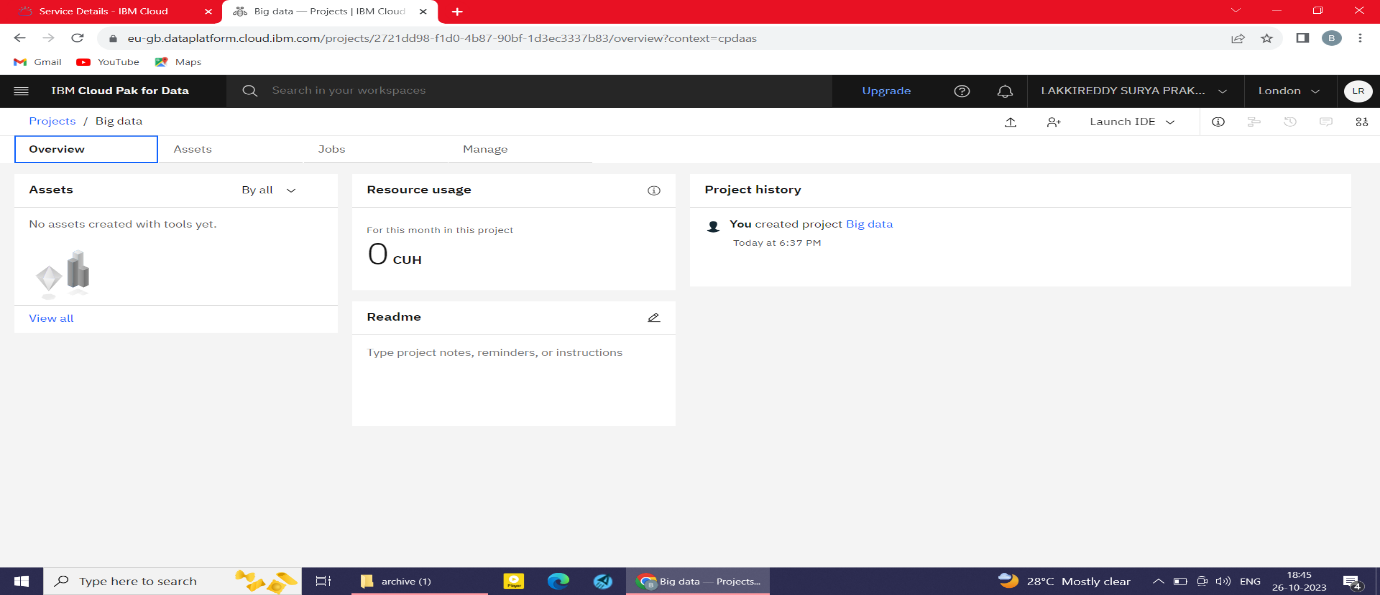
• Click Launch.

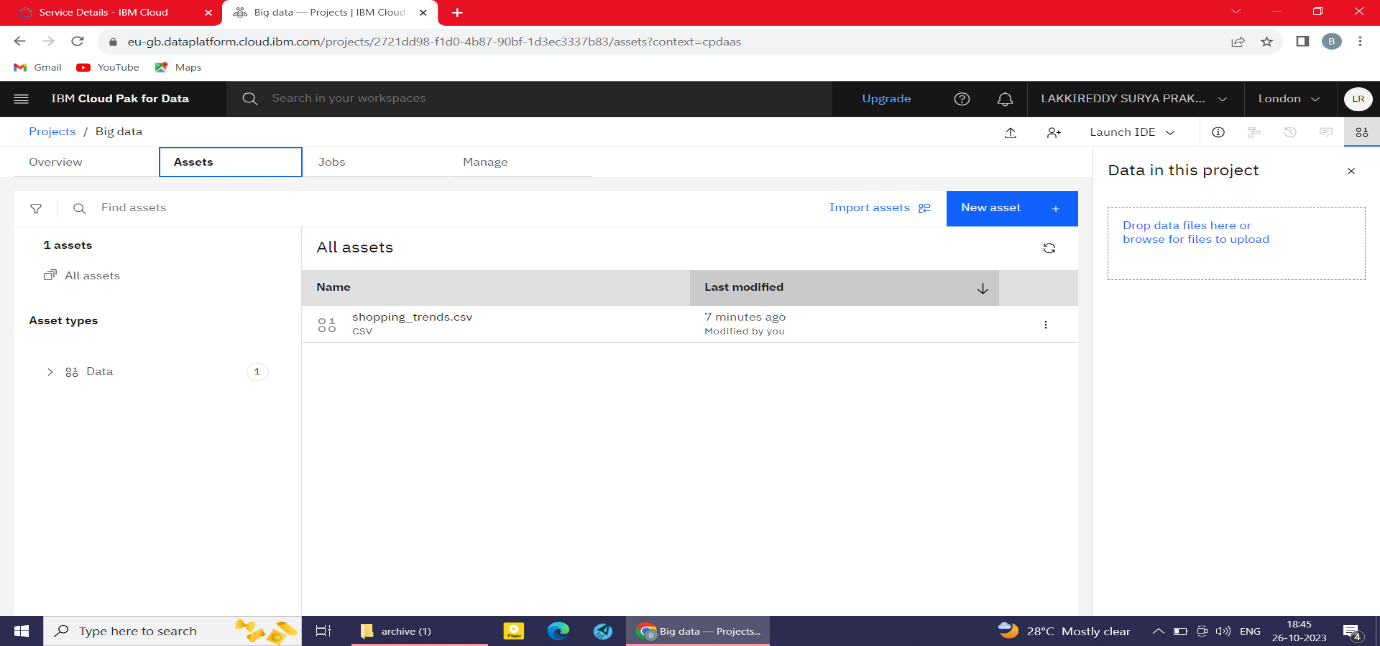
A new window will open with the IBM Cloud Pak for Data user interface.











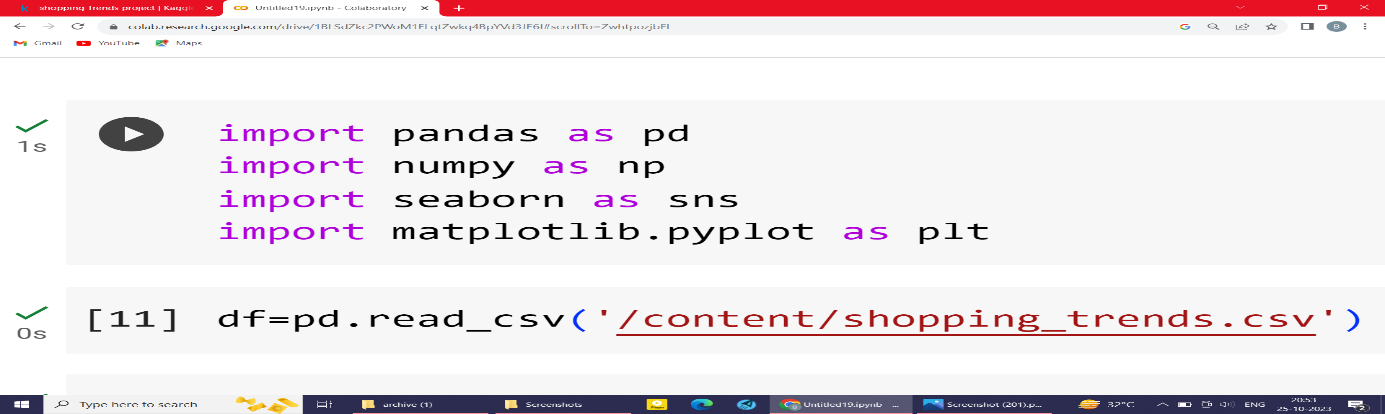
**DESCRIBING OF DATASET:**

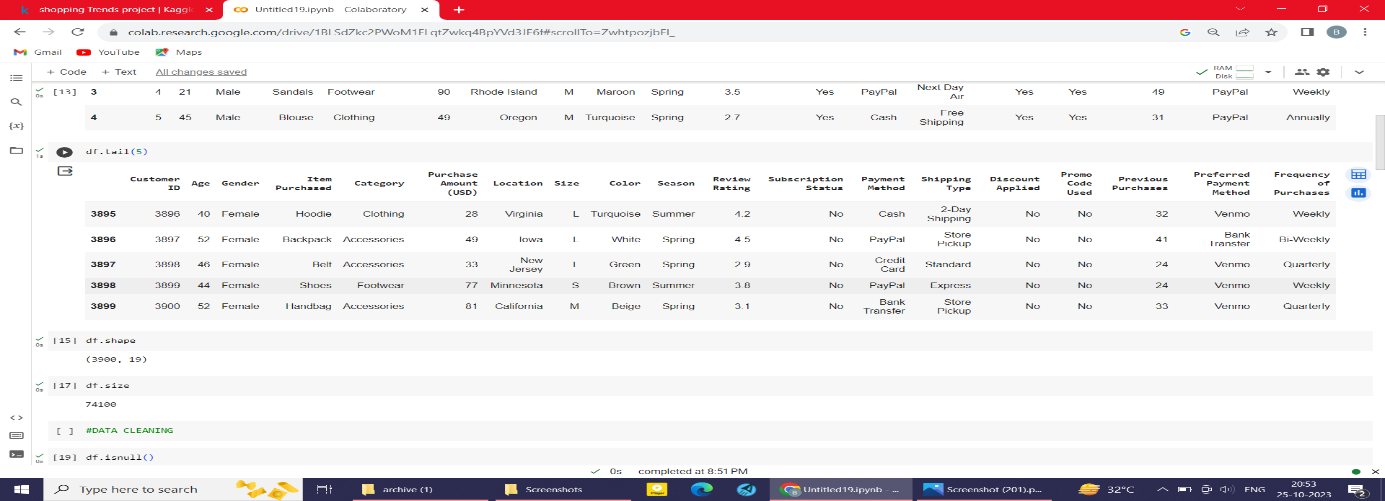
I had chosen the dataset called shopping trends from thesis paper.

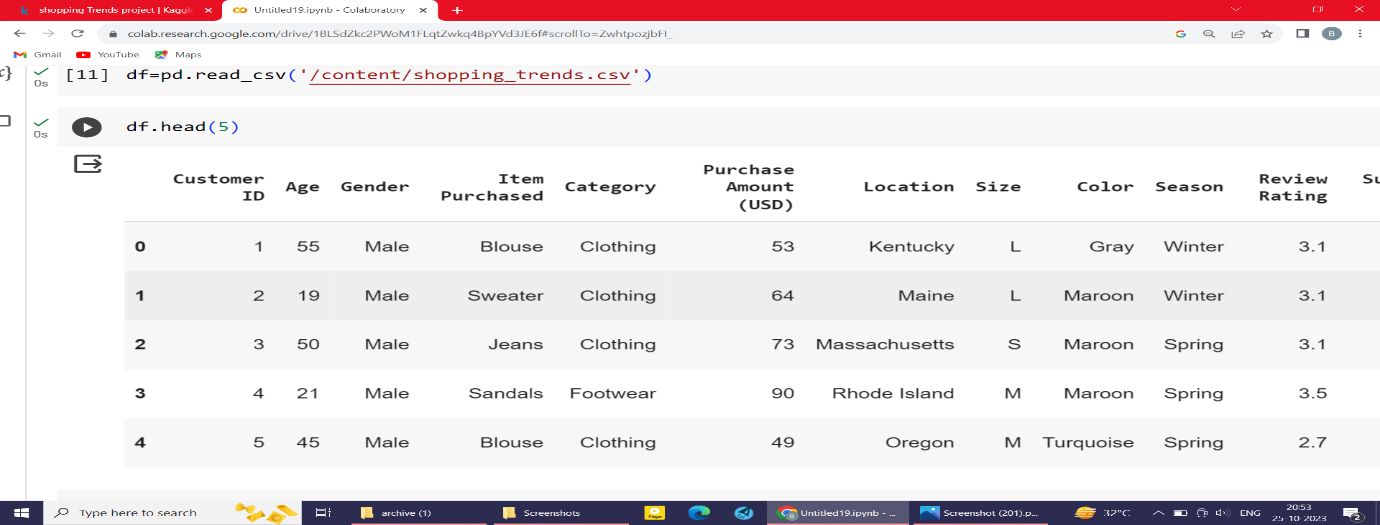
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Customer ID | Age | Gender | Item Purchased | Category | Purchase Amount (USD) | Location |
| 1 | 55 | Male | Blouse | Clothing | 53 | Kentucky |
| 2 | 19 | Male | Sweater | Clothing | 64 | Maine |
| 3 | 50 | Male | Jeans | Clothing | 73 | Massachusetts |
| 4 | 21 | Male | Sandals | Footwear | 90 | Rhode Island |
| 5 | 45 | Male | Blouse | Clothing | 49 | Oregon |
| 6 | 46 | Male | Sneakers | Footwear | 20 | Wyoming |
| 7 | 63 | Male | Shirt | Clothing | 85 | Montana |
| 8 | 27 | Male | Shorts | Clothing | 34 | Louisiana |
| 9 | 26 | Male | Coat | Outerwear | 97 | West Virginia |
| 10 | 57 | Male | Handbag | Accessories | 31 | Missouri |
| 11 | 53 | Male | Shoes | Footwear | 34 | Arkansas |
| 12 | 30 | Male | Shorts | Clothing | 68 | Hawaii |
| 13 | 61 | Male | Coat | Outerwear | 72 | Delaware |
| 14 | 65 | Male | Dress | Clothing | 51 | New Hampshire |
| 15 | 64 | Male | Coat | Outerwear | 53 | New York |
| 16 | 64 | Male | Skirt | Clothing | 81 | Rhode Island |
| 17 | 25 | Male | Sunglasses | Accessories | 36 | Alabama |
| 18 | 53 | Male | Dress | Clothing | 38 | Mississippi |
| 19 | 52 | Male | Sweater | Clothing | 48 | Montana |

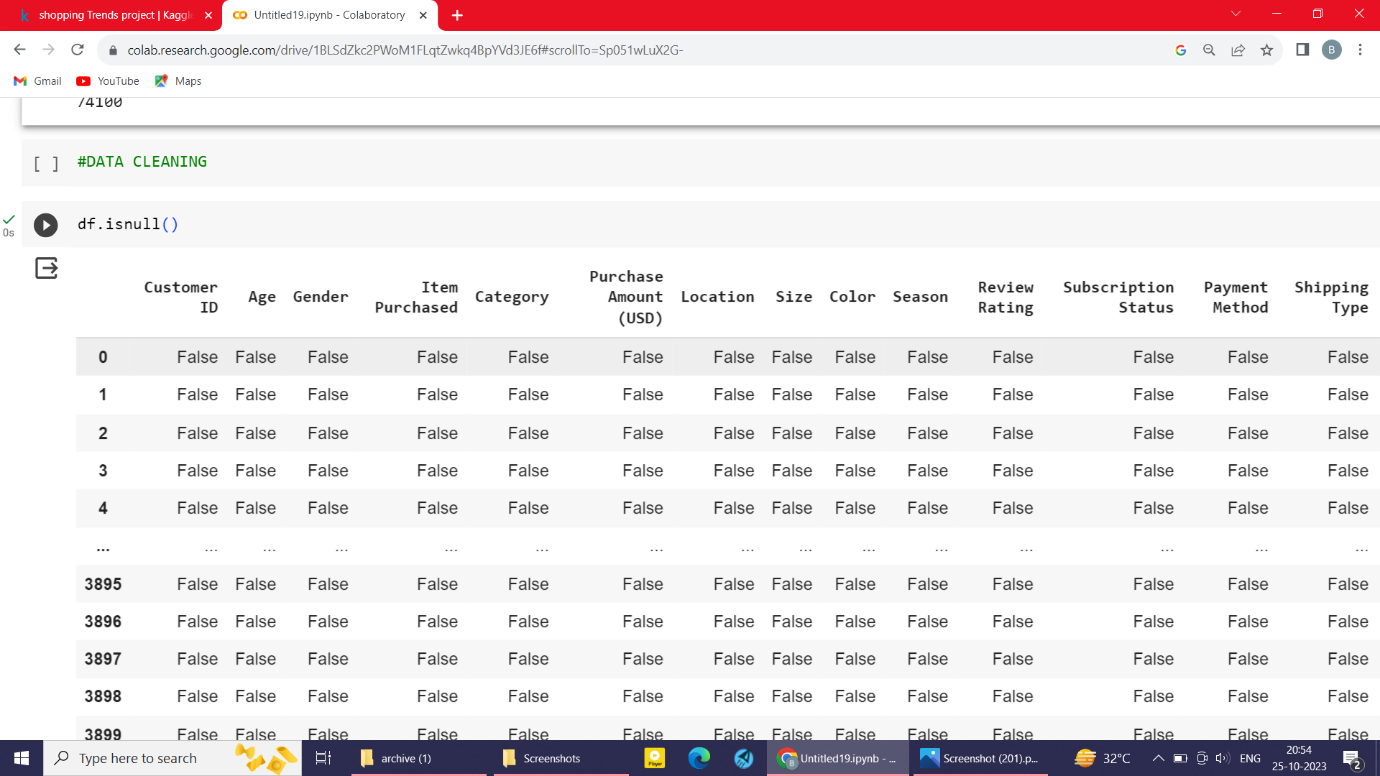
**BY USING THIS DATASET I AM GOING TO PERFORM DATA CLEANING AND DATA TRANSFORMATION AND VISUALIZATION.**

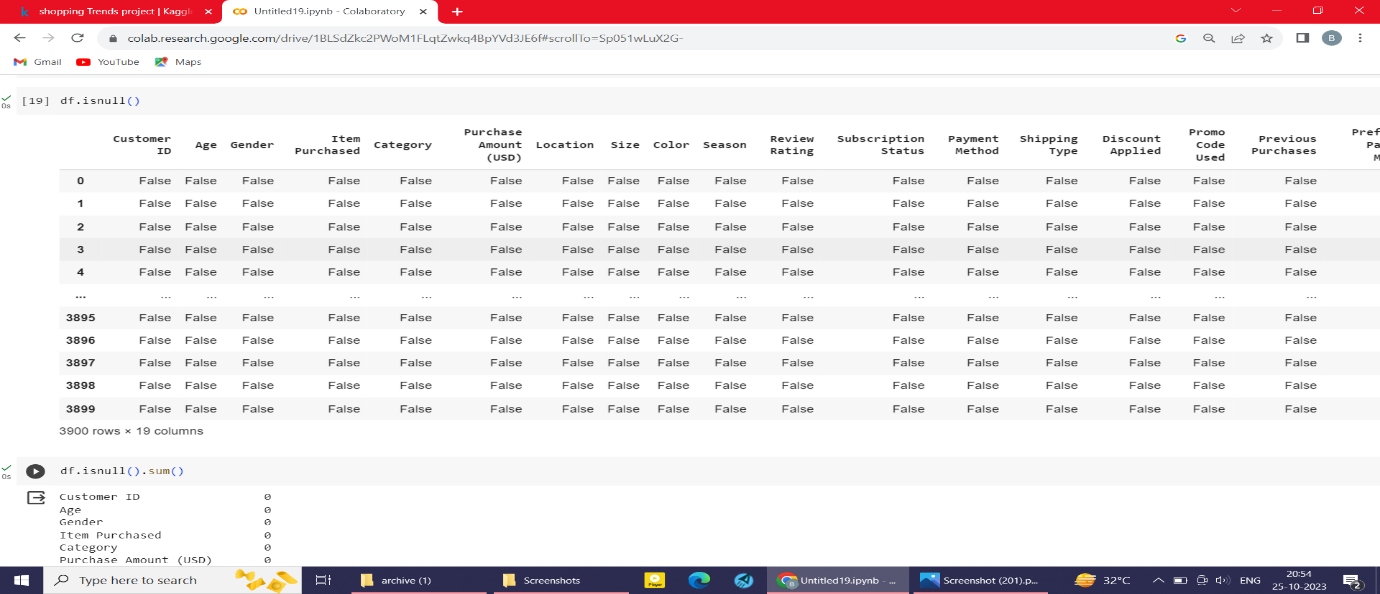
**PRACTICAL IMPLEMENTION OF DATA CLEANING AND DATA TRANSFORMATION.**

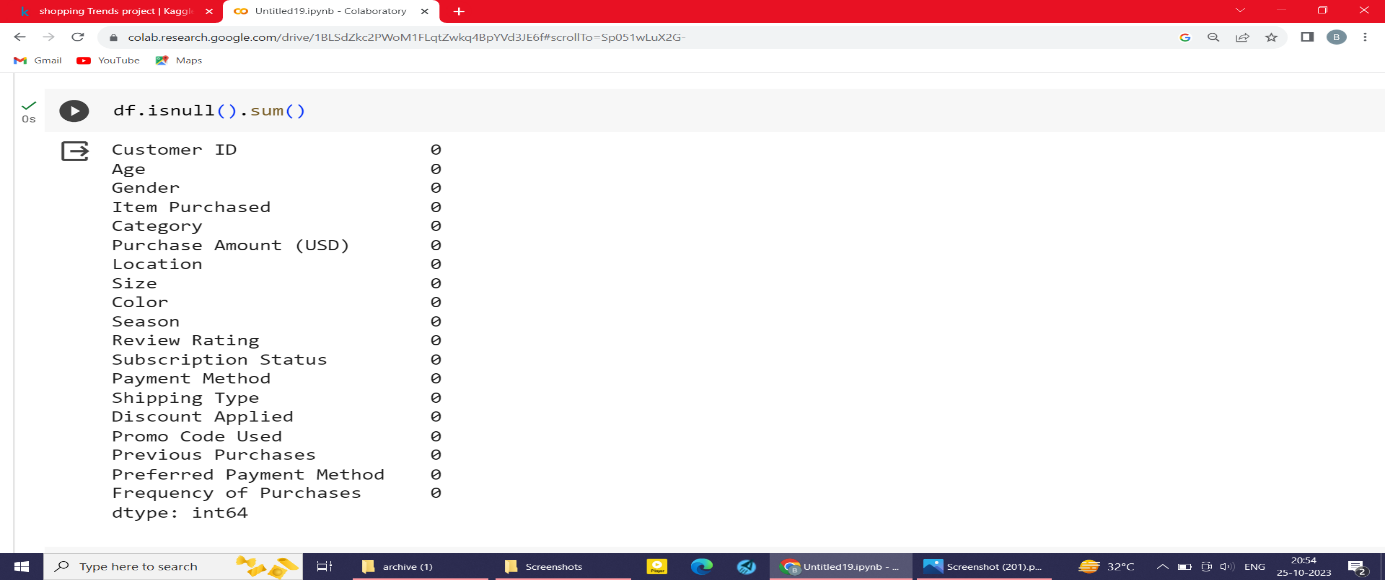


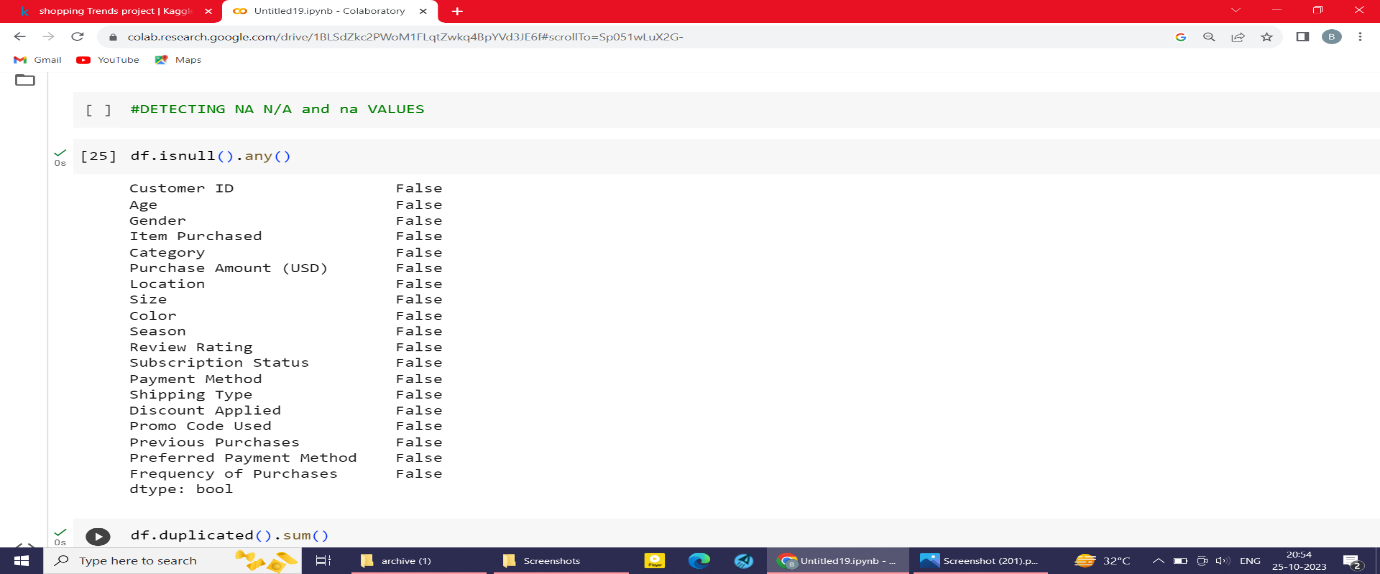


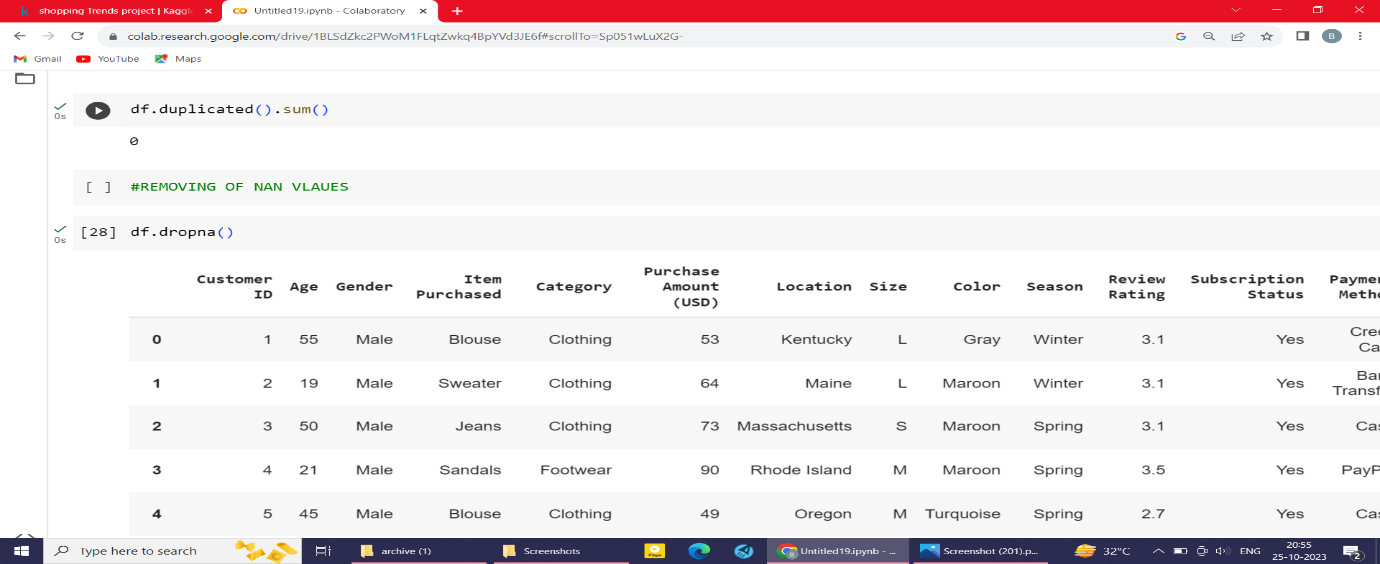


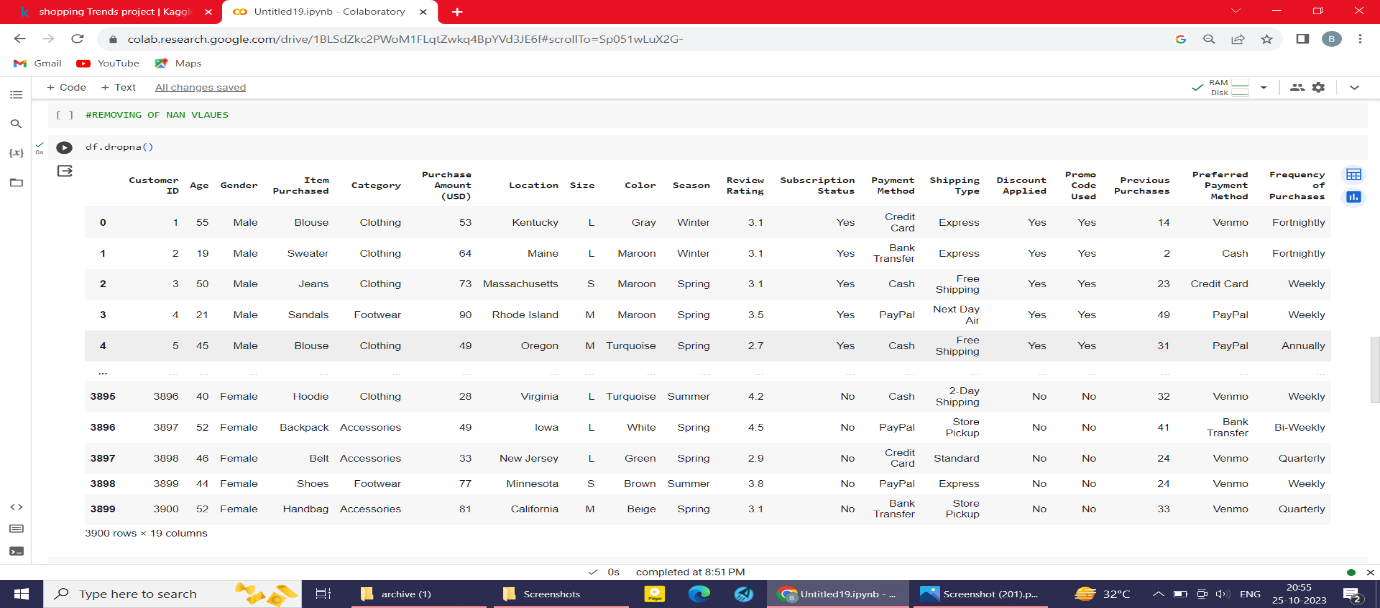


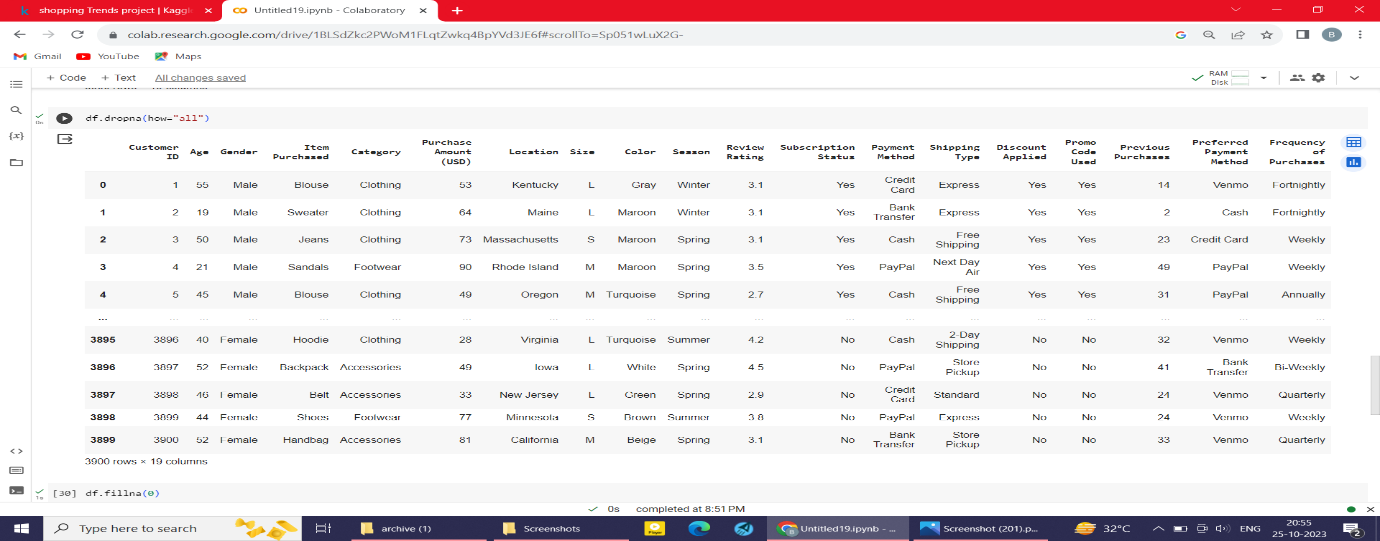


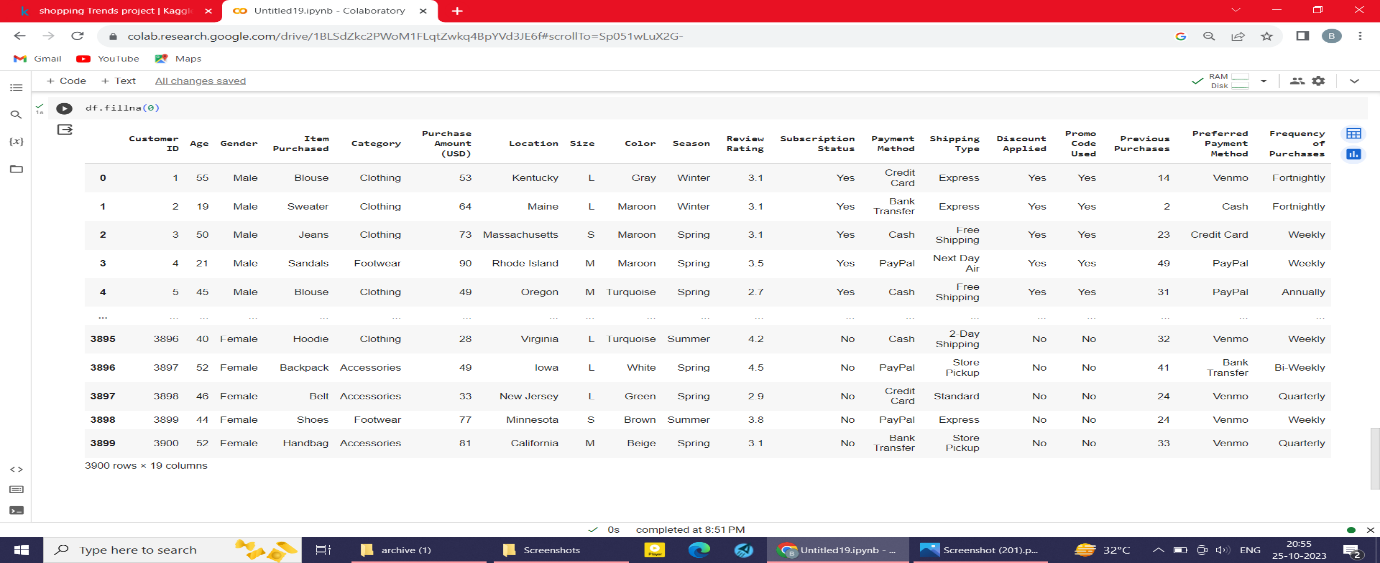


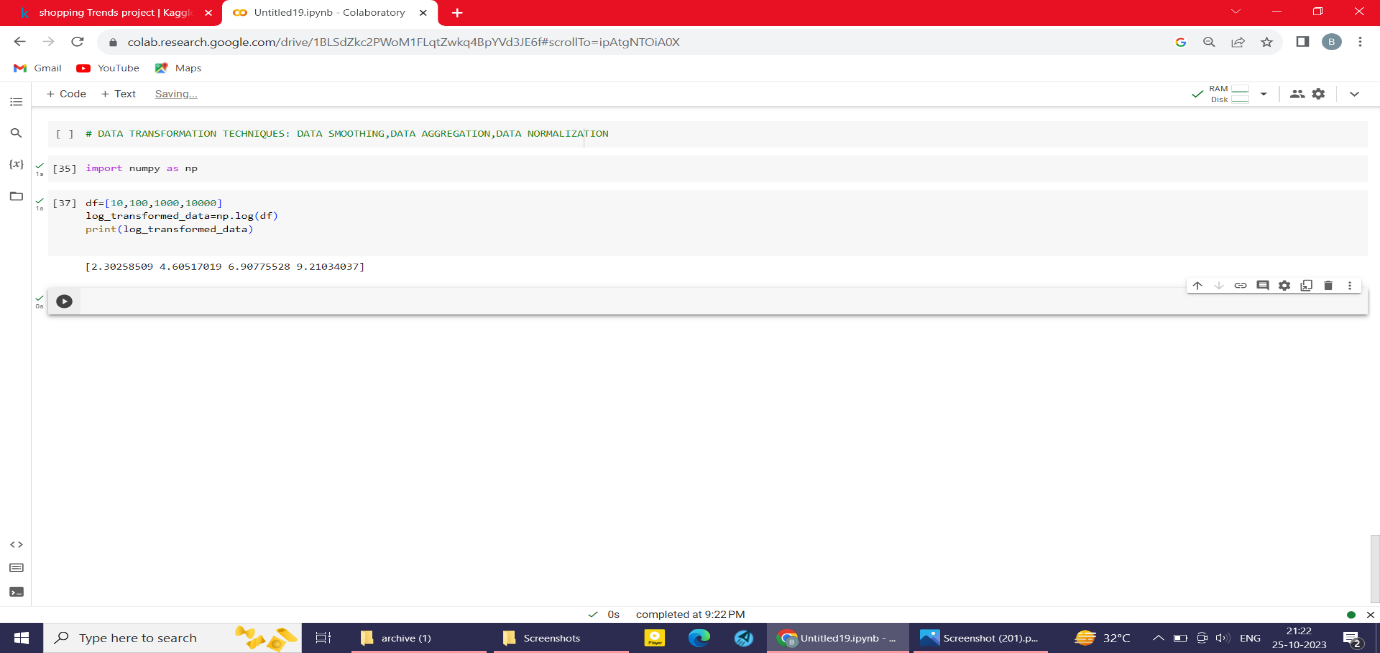












VISUALIZATION:

