ITCS 6100 BIG DATA FOR COMPUTATIONAL ADVANTAGE
GROUP-8
PROJECT DELIVERABLE-1

TEAM MEMBERS:
- Aneela Gannarapu
- Preetham Garre
- Rachana Goli
- Sai Rithwik Reddy Bolla
- Susmitha Dalli

COMMUNICATION PLAN:
- All the Team members decided to discuss perspectives and insights through slack or Gmail chat and exchange their ideas apparently whenever it's required.
- All the Team members planned to gather or meet via Zoom or Google meet whenever required and will work on finishing the tasks accordingly.
- The project's repository can be accessed on GitHub using URL that's given below: https://github.com/rachanagoli/BigDataGroup8

DATA SET SELECTION :
We selected the data set from Kaggle active competitions to analyze. Below are the links to 2 datasets selected.

Preference 1: Telecom churn dataset
Link:
https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets?datasetId=255093&searchQuery=Prop

BUSINESS PROBLEM OR OPPORTUNITY:
The business problem or opportunity for the telecom churn dataset is to identify factors that lead to customer churn, which is when a customer switches to a different telecom service provider. The attributes provided in the dataset can be used to determine which factors have the most significant impact on customer churn. By identifying these factors, the telecom company can take proactive measures to retain customers and prevent them from switching to a different service provider.

Each of the attributes in the dataset can provide us with following opportunities to achieve the KPI which in turn achieves the business objective of the chosen company:

The State, Account length, and Area code attributes may provide insights into regional customer preferences and trends. The International plan and Voice mail plan attributes may indicate customer preferences for additional services. The Number vmail messages, Total day minutes, Total eve minutes, Total night minutes, and Total intl minutes attributes provide information on customer usage patterns. The Total day calls, Total eve calls, Total night calls, and Total intl calls attributes can provide insights into call quality and customer satisfaction. The Total day charge, Total eve charge, Total night charge, and Total intl charge attributes can provide insights into pricing strategies. The Customer service calls attribute can provide insights into customer satisfaction and support needs. The Churn attribute is the target variable that indicates whether a customer has churned or not.

RESEARCH OBJECTIVES AND QUESTIONS:

The chosen dataset consists of various call attributes out of which we are focusing on few attributes to get the correlation between churn rate and the chosen attribute.Furthermore the patterns of the users can be analyzed to predict the churn behavior.The extracted patterns can be can be utilized by the company's to come up with customer retention plans.Data analysis can be every financial quarter so as to match the business quarter of the chosen company.We have opted to employ AWS technologies to implement the solution. Our plan involves implementing appropriate designs and algorithms to identify demand patterns. The crucial task at hand is to determine the most suitable designs and algorithms to analyze the dataset effectively.

We listed out the following questions as an initial attempt towards the project:
- What are the top 5 states that have a greater number of customer churn?
- What is the average account length of customers grouped for each area,and look for hotspots in each state/country?
- How is the international usage for the states with highest churn rates?
- How is the customer service call ratio impacting the user for each state and there by relating it to find relation with churn?
- What is the most frequent period of the day that most of the churn users accessed to make phone calls?
- How do the call durations differ retaining customers vs churned customers?

Preference 2: Customer shopping Dataset

Link:

https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset

BUSINESS PROBLEM OR OPPORTUNITY:

The dataset provided encompasses several attributes, such as age groups and genders, that offer a holistic insight into shopping behaviors. It comprises vital information like invoice numbers, customer IDs, age, gender, payment modes, product categories, quantity, price, order dates, and shopping mall locations. Predictive analytics can be employed to analyze the top-selling product categories in a month or a particular quarter. This analysis can aid in eliminating supply chain challenges and managing inventory for that specific period, preventing products from becoming out of stock or overstocked, which may result in waste if unsold.

RESEARCH OBJECTIVES AND QUESTIONS:

Our objective is to analyze the dataset and devise a solution to the current issue faced by the business in a manner that prevents any future problems. Our primary focus is on providing follow-up support based on demand and facilitating the achievement of business objectives. We have opted to employ AWS technologies to implement the solution. Our plan involves implementing appropriate designs and algorithms to identify demand patterns. The crucial task at hand is to determine the most suitable designs and algorithms to analyze the dataset effectively.

We listed out the following questions as an initial attempt towards the project:
- What are the predicted top selling categories for any upcoming month or quarter?
- Which gender can be targeted to buy a product of a particular category?
- Are there any notable differences in purchase behavior between customers of different age groups and genders?
- Can we identify any trends or seasonal patterns in sales of particular product categories?