# ITCS 6100 Big Data Analytics for Competitive Advantage Spring 2023 - Group 11 – TERM PROJECT
# Project Deliverable 1: Group Formation and Project Understanding

## TEAM

### TEAM MEMBERS

**Nihar Gopidi – 801327998**

**Akhila Chitturi - 801308961**

**Akhil Vadlakonda - 801275406**

**Dinesh Reddy Kankanala-**

**801274360**

**Shashank Patlolla – 801328367**

### COMMUNICATION PLAN

- **Communication Methods**: Google chat group is created for day-to-day conversation and syncing. Email for a more formal and documented way of communication.
- **Response times for communication**: The team members communicate with each other and response time of each member is as minimum as possible. If any team member is absent or away from work place it is informed in the group communication network.
- **Attendance in Meeting**: We have made few fixed timings which is comfortable for everyone and made it mandatory for everyone to attend, but if informed earlier exemption can be granted. we also have made few optional meetings this means that it is not mandatory for all team members to attend every meeting but at least 3 out of 5 members need to be present. The ones not attending shall inform the rest well in advance. It is appreciated if all can attend anyways.
- **Publish control**: All changes made by the team members are pushed to the artifact by raising pull requests which are reviewed by the peers before merging.
- **Team work and work Division**: Work is divided between the team members in the form modules. If anyone faces an issue in the work it will be discussed and come to a deterministic understanding. Any overload can be brought to

the notice of the other members so it can be divided or redistributed.

### PROJECT ARTIFACT REPOSITORY

All our work can be found in the public repository that has been created on GitHub.

The link to the repository is - https://github.com/akhilachitturi1804/BigDataGroup11

# 1/ Car Sales

## BUSINESS PROBLEM, OPPORTUNITY, DOMAIN KNOWLEDGE

### BUSINESS PROBLEM:

The dataset we've selected is from Kaggle, where it was gathered from a well-known website with adverts for vehicles for sale. The dataset is updated weekly and contains information about cars in over 18 columns, including columns for the car's price, brand, fuel type, color, mileage, and power. We can determine some helpful information from this data, such as what price range users are most interested in purchasing in, what type of color appeals to people the most, regardless of cost, what fuel type is valued most, and how to use a correlation graph to understand the correlation between various variables that are closely related. This information reveals which variables in the dataset are closely related so that they can be given more weight.

## SELECTION OF DATASET

### DATASET

https://www.kaggle.com/datasets/ekibee/car-sales-information

## RESEARCH OBJECTIVES AND QUESTIONS

### RESEARCH OBJECTIVE

Finding the variables that affect automobile sales in the US may be the study goal for analyzing the Car Sales Information dataset from Kaggle. Understanding which automotive features—such as make, model, year, price, mileage, fuel type, etc.—have the most effects on sales volume and price as well as how external factors like location and dealer ratings affect sales—could be the specific objective.

**EXAMPLE DESCRIPTIVE QUESTIONS**

- What are the most common attributes of our customers, and how do they relate to their buying behavior?
- How do product reviews or ratings affect sales or customer satisfaction, and what are the most common themes or sentiments in the reviews?
- How does employee productivity vary across different teams or departments, and what factors contribute to the differences?

**EXAMPLE PREDICTIVE QUESTIONS**

- What are the long-term impacts of our business decisions, and how can we use data to make more informed and sustainable choices?
- What are the long-term impacts of our business decisions, and how can we use data to make more informed and sustainable choices?
- How do our products or services compare to those of our competitors in terms of customer satisfaction, and what can we do to improve?

# 2/ Used Cars

## BUSINESS PROBLEM, OPPORTUNITY, DOMAIN KNOWLEDGE

**BUSINESS PROBLEM:**

The automotive industry could benefit from using the Craigslist Cars+Trucks dataset by understanding current market trends and consumer preferences in the used automobile sector. Companies might learn more about the most popular brands and models, the features that are most in demand, and the elements that affect pricing by evaluating this data. Companies might use this data to modify their product lineups and marketing tactics to better cater to the wants and needs of their target market. Making data-driven judgments based on current market trends and consumer behavior by utilizing the Craigslist Cars+Trucks dataset could give automotive companies a competitive edge in the used automobile market.

# SELECTION OF DATASET

## DATASET

https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data

# RESEARCH OBJECTIVES AND QUESTIONS

## RESEARCH OBJECTIVE

Finding important elements that influence the sale of used automobiles, such as make, model, year, mileage, location, and price, through the analysis of the Craigslist Cars+Trucks dataset is the research goal. The investigation might also look at how attributes like color, condition, and features impact used car pricing. In addition to providing useful information for automakers to modify production and marketing strategies to match the needs of their target customers, this research could assist auto dealerships and sellers in setting prices based on demand and supply situations.

## EXAMPLE DESCRIPTIVE QUESTIONS

- What are the most common features that are sought after by buyers in the used car market?
- What is the average mileage for cars being sold in the used car market and how does it impact pricing?
- What is the average price range for the most popular car makes and models in the used car market?

## EXAMPLE PREDICTIVE QUESTIONS

- What is the average mileage for cars being sold in the used car market and how does it impact pricing?
- What is the correlation between the age of a car and its pricing in the used car market?
- How do different body types (e.g. sedan, SUV, truck) impact the pricing of cars being sold in the used car market?