

## PROJECT BRIEF

Greetings from **Elite Bike Store (EBS)**! In your role as a Data Analyst, you will help uncover insights within our bike purchase dataset using **SQL**. EBS is committed to leveraging data-driven decisions to optimize our business strategies and improve customer experiences. Your job is to write and execute SQL queries to analyze the customer and bike purchase data, providing valuable insights into purchasing trends, customer behavior, and business performance.

### Objective:

To develop a series of **SQL queries** that allow the Elite Bike Store team to extract, manipulate, and analyze data from the **Bike Purchase dataset**. Your queries should provide actionable insights into customer segments, sales performance, and trends to aid in data-driven decision-making.

### Task: Data Analysis and Query Writing in SQL

#### Data Preparation:

- **Clean the data** by writing SQL queries to handle missing or inconsistent data.
- Ensure that the data is normalized where necessary to eliminate redundancy and maintain data integrity.
- Use SQL **JOINS** to combine relevant tables if needed (assuming the dataset is in multiple tables).

#### Customer Segmentation:

- Write SQL queries to segment customers based on attributes such as **marital status, income, education, and region**.
- Use SQL **GROUP BY** and **aggregate functions** (e.g., COUNT, AVG, SUM) to summarize data for each customer segment.
- Provide insights on the demographics of customers who are most likely to purchase bikes (e.g., income, age, or region analysis).

#### Sales Performance Analysis:

- Query the data to calculate **total sales, number of bikes purchased**, and the **percentage of customers who purchased a bike**.
- Use **aggregated SQL functions** (e.g., SUM, AVG, MAX, MIN) to calculate **average income of bike buyers**, the **number of children** per buyer, and **other relevant metrics**.
- Analyze the data by **region, occupation**, and **education level** to determine which segments generate the most revenue and are the most profitable.

### Profitability Analysis:

- Write queries to calculate **profit margins** based on hypothetical sales scenarios.
- Use SQL **case statements** to calculate derived metrics such as profit percentage based on customer income, commute distance, or number of cars.
- Provide insights into the factors that lead to higher profitability for the store.

### Advanced Queries (Optional):

- Write advanced SQL queries using **CTEs (Common Table Expressions)**, **subqueries**, and **nested queries** to break down complex analyses into manageable parts.
- Use **window functions** (e.g., RANK, PARTITION BY) to conduct detailed performance ranking of customers or regions.

### Evaluation Criteria:

Participants will be evaluated based on:

- The **complexity and accuracy** of the SQL queries.
- The **clarity** of the results and the ability to provide actionable insights.
- Use of **advanced SQL functions** (e.g., window functions, subqueries, case statements).
- **Creativity** in analyzing the data and extracting valuable information.
- **Documentation** of query logic and methodology.

### Submission Requirements:

#### 1. SQL Query File:

- a. A **SQL file** containing all the queries used for data cleaning, customer segmentation, sales analysis, and profitability analysis.
- b. Queries should be well-structured and formatted for clarity.

## 2. Results:

- a. Submit **query results** as tables or visual outputs (if using a SQL environment that supports visual output).
- b. Include any additional insights drawn from the analysis.

## 3. Documentation:

- a. A brief **report** describing the methodology used, the logic behind the queries, and key insights derived from the dataset.

### Timeline:

- **Deadline:** 15 days from project initiation.
- **Submission Mode:** SQL query file and results to be submitted via a Google Form.

### Data Dictionary:

- **ID:** Unique identifier for each customer.
- **Marital Status:** Indicates whether the customer is single (S) or married (M).
- **Gender:** Customer's gender (M/F).
- **Income:** Customer's annual income.
- **Children:** Number of children.
- **Education:** Level of education (e.g., Bachelors, High School, Partial College).
- **Occupation:** Customer's job type (e.g., Professional, Skilled Manual, Clerical).
- **Home Owner:** Indicates whether the customer owns a home.
- **Cars:** Number of cars owned.
- **Commute Distance:** Distance the customer commutes to work.
- **Region:** Geographic region (e.g., Europe, Pacific).
- **Age:** Customer's age.
- **Purchased Bike:** Indicates whether the customer has purchased a bike (Yes/No).

This **SQL project** will challenge you to develop sophisticated queries that uncover key insights about customer behavior and business performance at Elite Bike Store. We look forward to seeing the insights and recommendations you will derive from the dataset.