Team-Based Learning Tutorial

Instructions

1. This is a group project. Each group may consist of up to 8 members. Please set up your group by October 8.
2. Each group is required to deliver a 3-minute presentation of your project in class.
3. This project is quite open-ended, and you are strongly encouraged to develop innovative ideas and designs!
4. If you have any questions, please reach out to our TAs directly in class.

Topic – LLM-Assisted Innovative Data Visualization

In today's data-driven landscape, effective visualization is essential for interpreting complex information. This project explores how Large Language Models (LLMs) can enhance data visualization techniques, allowing you to uncover insights from diverse datasets. By leveraging LLMs, you can reveal trends and correlations in compelling ways, transforming your data storytelling into engaging and informative visual presentations.

We provide three types of datasets: Titanic, Education Analytics, and Adult Census . You can download these datasets from Canvas -> Files -> Tutorials -> TBL Tutorial -> data, and detailed descriptions are available at the corresponding hyperlinks. For your project, please select one dataset and focus on innovative data visualization with the assistance of large language models. For example, in the Titanic dataset, you could visualize the relationship between gender and survival rates, which might reveal that women had a higher survival rate than men. In the Student Alcohol Consumption dataset, consider visualizing how parents' education levels correlate with final grades, potentially uncovering trends where students with more educated parents achieve better academic results. For the Adult Census Income dataset, you could create visualizations illustrating the connection between education level and income, demonstrating that individuals with higher education generally earn more.

Appendix: Introduction of datasets

1. Titanic Dataset

Overview:

The Titanic dataset records information about the passengers aboard the RMS Titanic during its ill-fated maiden voyage in 1912.

Main Features:

PassengerId: Unique identifier for each passenger

Survived: Survival status (0 = No, 1 = Yes)

Pclass: Passenger class (1 = First class, 2 = Second class, 3 = Third class)

Name: Name of the passenger

Sex: Gender

Age: Age of the passenger

SibSp: Number of siblings/spouses aboard

Parch: Number of parents/children aboard

Ticket: Ticket number

Fare: Ticket fare

Cabin: Cabin number

Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

2. Student Alcohol Consumption Dataset

Overview:

This dataset includes information about alcohol consumption among Portuguese students, often used to analyze the relationship between lifestyle choices and academic performance.

Main Features:

school: School (e.g., 'GP' = Gabriel Pereira, 'MS' = Mousinho da Silveira)

sex: Gender

age: Age

address: Address type ('U' = Urban, 'R' = Rural)

family_size: Size of the family

parents_status: Marital status of parents

study_time: Study time per week

failures: Number of failed courses

activities: Participation in extracurricular activities

alcohol_consumption: Amount of alcohol consumed

G1, G2, G3: Grades for different terms

3. Adult Census Dataset

Overview:

The Adult Census dataset, also known as the "Census Income" dataset, contains demographic information from the U.S. Census, often used to predict whether an individual earns more than $50,000 per year.

Main Features:

age: Age of the individual

workclass: Type of employment

fnlwgt: Final weight (a measure of population representation)

education: Education level

education_num: Number of years of education

marital_status: Marital status

occupation: Occupation of the individual

relationship: Relationship status

race: Race

sex: Gender

capital_gain: Capital gains

capital_loss: Capital losses

hours_per_week: Hours worked per week

native_country: Country of origin

income: Income level (>50K, <=50K)