GUIDE TO GROUP PROJECT MAS291

I. Overview

In this project, you are required to work in a group and present your work to the class.

Each group will look for secondary data (related to housing, finance, health, or any topic of your choice), then use Excel (Python, R, etc.) to perform inferential statistics on the data to obtain useful information.

II. Project Requirements

Groups are expected to write a professional report that describes their data collection, analysis, findings, and implications. Students should apply all of the following techniques to their data:

1. Test a hypothesis and construct confidence intervals for the mean and the proportion of a population.
2. Test a hypothesis and construct a confidence interval for the difference in mean and proportions of two populations.
3. Regression analysis. In this requirement, you are expected to complete the following steps:
   (a) Identify two random variables X and Y in your data. For instance, X= height, Y= weight.
   (b) Construct a scatter plot for the data. Do you observe a linear relationship?
   (c) Compute the sample correlation coefficient.
   (d) Find the equation of the estimated regression line, and use it to predict a future value for Y.

III. Evaluation

Complete all tasks and provide quality results (50%)

- Ensure all assigned tasks are completed thoroughly.
- Results should be accurate and demonstrate comprehensive understanding.

Good documentation (15%)

- Provide a detailed report of all results.
- The report should be well-organized and clearly written.

Solid knowledge on the topic (15%)

- Demonstrate a deep understanding of the subject matter.
- Show the ability to apply knowledge effectively.

Good presentation in class (10%)

- Present findings in a clear and engaging manner.
- Use appropriate visual aids and communication techniques.

Effective teamwork (5%)

- Collaborate effectively with team members.
- Show good communication and coordination within the team.

Creativity and innovation (5%)

- Demonstrate originality and uniqueness in approach.
- Incorporate creative solutions and innovative ideas.

IV. Data sources

Here are several useful websites where you can find high-quality secondary data for your project

General Data Repositories

1. **Kaggle**: Kaggle provides a vast collection of datasets on various topics, contributed by the user community. It's a great place to find well-documented and diverse datasets.
2. **UCI Machine Learning Repository**: A classic source for datasets used in machine learning, including many standard datasets that are widely used in academic research.
3. **Google Dataset Search**: A search engine specifically for datasets, making it easier to find the datasets you need from across the web.

Government and International Organizations

4. **Data.gov**: The U.S. government's open data site, which provides access to datasets on a wide range of topics including health, education, and transportation.
5. **World Bank Open Data**: Offers free and open access to global development data, including economic indicators, social statistics, and more.
6. **United Nations Data**: Provides a single entry point to a wealth of statistical information collected by the United Nations and its specialized agencies.

Specific Domains

Housing

7. **Zillow Research Data**: Provides real estate and housing market data, including home prices, rental prices, and market trends.
8. **U.S. Census Bureau**: Offers a wide range of data, including detailed housing data from the American Community Survey and other programs.

Finance

9. **Yahoo Finance**: Provides historical stock market data, financial news, and financial reports for companies.
10. **Quandl**: A platform for financial, economic, and alternative datasets. Many datasets are available for free, and some require a subscription.

Health

11. **CDC Data & Statistics**: Offers a variety of health-related datasets from the Centers for Disease Control and Prevention.
12. **WHO Global Health Observatory**: Provides health statistics and information from the World Health Organization.

Research and Academic Data

13. **Harvard Dataverse**: A free data repository open to all scientific data from all disciplines. It offers data from Harvard and other research institutions.
14. **ICPSR**: The Inter-university Consortium for Political and Social Research maintains a data archive of research in the social and behavioral sciences.

Data for Specific Analyses

15. **FiveThirtyEight Data**: Provides datasets used in FiveThirtyEight's data journalism articles, covering topics from politics to sports.
16. **Awesome Public Datasets**: A GitHub repository with a curated list of public datasets in various domains.