CS F441
Selected Topics in Computer Science: Data Visualization
BITS Pilani, Pilani, Hyderabad and Goa Campus
Project
Due Date: 29th November 2023 (by Midnight)
Total Marks: 20 (weightage: 10%)

Objective of the Project:

The project plan should be based on the one of the following ideas, by providing specifics of what the project will deliver.

List of Ideas

Students may make project proposals that cater to the following themes:

Idea 1: xxx- Years of BITS-Pilani campuses

Note that BIT-Pilani Goa is celebrating 20 Years of BITS Goa. So, it will be a good time to come up with a few dataviz project plans related to this theme for each of the campuses.

Students may get some ideas from

- BBC page "India economy: Seven years of Modi in seven charts: https://www.bbc.com/news/world-asia-india-57437944" or from
- TOI page: "8 years, 8 indicators: How economy fared under Modi: https://timesofindia.indiatimes.com/business/india-business/8-years-8-reforms-how-economy-fared-under-modi-government/articleshow/91895885.cms"

Idea 2:

The following few suggestions came from our Associate Dean of AUGSD of the Hyderabad Campus:

- Relationship between BITSAT Score and graduated student's CGPA
- Relationship between BITSAT Score and First-year SGPA
- Relationship between SGPA and semester for every student
- Relationship between withdrawals and courses
- CGPA analysis

Idea 3:

BITS students come from different secondary education boards (~ two dozen).

- Visualization of their first sem/year CGPA wrt their board
- Visualizing across last n years for a given board

Idea 4:

Geo-visualization of FD/HD/ Ph.D. students in BITS campuses by location/state/ country/rural/urban/district.

Idea 5:

BITS-Alumni in the world: Many of our BITS-Alumni have moved to many parts of the world and many of them have made names for themselves and for BITS. We could do a simple geo visualization of work places of BITS alumni (data may be accessed from BITS-Pilani alumni center).

- Geo-Visualization of BITS alumni in industry / academia / entrepreneurs.
- Within industry perhaps some categorization in terms of verticals/companies
- Visualization of RCG (recent college graduates) doing Masters / PhD in various countries / universities

Idea 6: Socio-economic aspects of India

Students may look at https://www.indiaspend.com/data-viz for ideas and even propose to reproduce visuals from that page (correcting for any violation of data-viz principles that were taught in the class).

Idea 7: Environmental aspects (Ground-Water levels)

The ground water-level is depleting fast. We need to see how much is water-level across the geographical areas of states in India in different months. Here is one visualization. You can plot the consumption of ground water for different usage, such as agricultural, industry, and urban.

Idea 8: Environmental aspects (Air Pollution)

The air pollution increases in the winter. Visualize the air pollution in different parts of the country for different months. You can plot the sources of pollution for transportation, agriculture, industry, etc.

Idea 9: Energy conservation

The electricity consumption varies in the country in different months. Visualize the energy consumption in different parts of the country for different months. You can plot those for different appliances, such as AC, refrigerators, etc.

Idea 10: Sports

Many of our students are sports (cricket, football, etc…) fans. They are encouraged to propose sports related visualization projects.(see for ex: https://edition.cnn.com/2023/09/20/sport/saudi-arabia-soccer-spl-bin-salman-intl-spt-cmd-dg/index.html)

Idea 11: Medical data visualization

Assuming that our campus hospital maintained a database of records of patient's medical history: (could be the record of health indicators, whether the patient suffered symptoms before, their diagnosis, prognoses and treatment), the record of the daily flow of patients and the symptoms/ diseases they are treated for and their satisfaction survey report; resources available vs resources used, create patient specific visuals to keep them aware of their health trend and take remedial action (say, following healthy habits), for health administrators for improving the health care of the campus residents (say, increasing critical resources).

Idea 12: Academic

- Students may like to implement some recently proposed dataviz articles.
    - For example: Imputing data (filling in missing data) is not easy. See for ex*: Kernel Weighted Least square approach for imputing missing values of metabolomic data*, 27 May 2021, Sci Reports 11, Article number 11108 (2021)
- Network layout algorithms: A number of algorithms have been proposed for network layout. See the article: https://cs.brown.edu/people/rtamassi/gdhandbook/chapters/viz.pdf . It is

chapter from the popular Graph Drawing handbook (Roberto Tamassia, editor, https://cs.brown.edu/people/rtamassi/gdhandbook/).

General Instructions:

1. This assignment can be done in groups of no more than four students.
2. Each group must provide the details of their group in the below given spreadsheet by assigning their group an ID starting from A, B, C, ….Z, A1,B1, C1, .. etc. Please put the group name in front of the group member's name and ID in the appropriate column. https://docs.google.com/spreadsheets/d/1TXYTOuX9FzPYfMt6tNFujg_liCmL11Yy60ro1fw5KQU/edit?usp=sharing
3. The project choice must be submitted by Nov. 10.
4. The assignment must be done using Python Libraries only, using Plotly and/or Matplotlib API (or an API in their ecosystem).
5. The project must have 3 components:
    a. Data acquisition (includes hand creating data table, data scraping from web/pdf file, data accessing from data archival system through a python API, or using an existing excel/csv/json file).
    b. Data preparation (includes data cleaning and combining data from multiple sources).
    c. Data visualization (using Python library method call(s) to create the final visuals).
6. You can discuss with your friends but refrain from copying the code and submitting. Copied codes will receive no credits for the entire assignment.
7. You must demo the code to the instructor on a scheduled date and time assigned to your project (during Dec 1 – 3) after submission.
8. Presence during the demo is a mandatory requirement. Any student absent during the demo will be awarded zero marks.

Submission Instructions:

1. The submission will be through CSF441 Google classroom page.
2. There will be only one submission per group.

3. The name of the file should be id1_CSF441_Project.zip, where id1 refers to the ID of only one member of the group. The zip file must include code, sample visuals, and a project report.
4. The project report (in pdf or word format) must include the goal, detailed instructions for completing each component of the project, and the source(s) of data.
5. **The project effort should be at least n-PersonWeek (where n is the group size).**