CSCE 5310 Methods in Empirical Analysis (Summer 2024)
Project Proposal

Dear All,

The course project is designed as a group project in multiple steps. This course project starts with the first step, that prepare a proposal to study your datasets. There are some requirements to meet in your dataset.

Dataset:

Your dataset should have more than **ten independent numerical variables**. Each group member is responsible for working on **three variables**. Don't forget to normalize your data if it is not. We need to have an Exploratory Data Analysis (EDA) as a complementary part of the proposal, EDA should cover the following list for each one of the variables. Don't forget to interpret all your results.

1. Scatter plot
2. Box plot (individual and together like example)
3. Histogram
4. Normal assessment
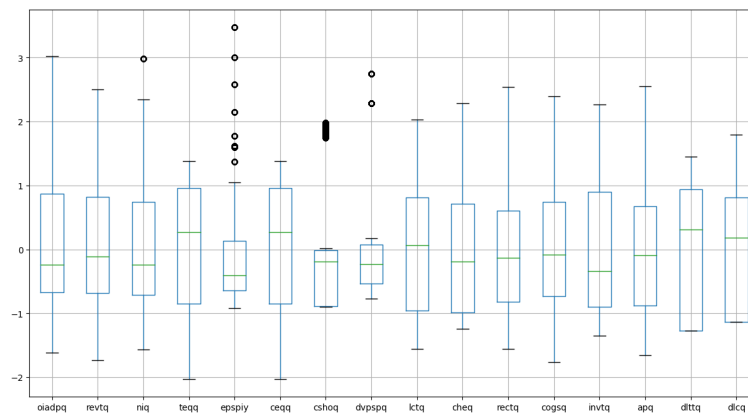5. Parameters estimation (Proportion, mean, and SD)

Figure: Sample box plot to compare all your variables and their outliers.

Topics:

- A tentative Topic can be (any topic needs to be a statistical analysis), but not limited to these topics:
- Stock market Analysis
- Any disease prediction analysis, your analysis could be.
- Measuring attitude towards a topic such as Mental Health, TV shows, etc
- Geospatial Projects: Is salary related to location, crime related to location.
- Student/Employment Performance
- House Pricing
- Fraud transaction detection
- Covid-19: Is the vaccine working? Daily death/recovery.
- Traffic Analysis (website traffic/ holiday road traffic/ road accidents)
- …

Tools:
You may select your most comfortable tools to perform this analysis. However, we suggest using a tool that helps you to interpret your result properly instead of more abstract tools that make it hard to interpret the results.

Template:
Template for submission of project proposal
1. Project title
2. Group members: Can be a maximum of 3 (as it is created in canvas)
3. Overview: A project overview is a detailed description of a project's goals and objectives, the steps to achieve these goals, and the expected outcomes. In addition, a project overview enables you to outline the project schedule, budget, and necessary resources, and status.
4. Critics: if your dataset is old, this section prepared the critics of previous studies and should cover the knowledge gap that may not be studied. However, if the dataset is new (such as active competitions), then this section may be ignored.
5. Goals: The project goal refers to achieving the desired outcome (performance goal) at a specific end date (time goal) by employing a particular amount of resources (resource goal). To be used as a measure for goal achievement, performance goals must be formulated and the degree of accomplishment verifiable.
6. Specification (Tools, Dataset, Implementation): A project specification is a document used for successful project management that defines the management plan of a project as a whole. It lists the needs, objectives, constraints, expected features, deadlines, and budget as accurately as possible.
7. Milestones: A milestone is a specific point within a project's life cycle to measure the progress toward the ultimate goal. Milestones in project management are used as signal posts for a project's start or end date, external reviews or input, budget checks, submission of a major deliverable, etc.
8. Responsibility of each group member: Each member should contribute to analyzing their variable to receive full marks.
9. Reference: Format reference properly.

Submission:

1. Turn into pdf format, one submission per group on Canvas by the deadline.
2. If you use an existing dataset or follow any resource, link them properly with proper descriptions. Otherwise, it will be considered a plagiarized submission.

Integrity:
Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to, cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

Rubrics:

Format	IEEE/ACM	5
Flow	Clear and logical sequence of steps	5
Critics	Reason for your study	5
Brainstorming	Ideas	5
Creativity	Repeated projects will not be accepted	7
Scatter plot	Plot and interpret	10
Box plot (individual)	Plot and interpret	10
Box plot (together)	Plot and interpret	10
Histogram	Plot and interpret	10
Normal assessment	Plot and interpret	10
Parameter estimation, Proportion	Interpret	10
Parameter estimation, Mean	Interpret	10
Parameter estimation, SD	Interpret	10
References		3
Total		**110**