Data Analytics I – Course Syllabus

Description: In this course, students will learn the various methods used to source and load data sets selected for analysis. The features and attributes of data will be reviewed after a data quality assessment has been conducted.

Pre-requisites course(s) or knowledge - None

Topics covered: Data sourcing, feature engineering, and preprocessing

Software: Microsoft Excel

At the end of the course, you will know:

1. **How to perform a data quality assessment.**
2. **How to clean a data set in preparation for data analysis.**
3. **How to identify different statistical data types ideal for different analytical algorithms.**
4. **How to perform feature aggregation and sampling for analytical efficiency.**

Learning Exercises

- Determining business objectives (based on background and success criteria).
- Understanding the business domain of knowledge.
- Acquiring or sourcing data to address business problems.
- Creating or reading a data dictionary.
- Understanding the difference between categorical, continuous, and discrete variables.
- Creating research questions and hypotheses to investigate business problems based on an initial review of the data set.
- Identifying common anomalies found in data and implications for data analysis.
- Create a data quality report.
- Using cleaning techniques during processing to remove noise from the data (duplicates, paragraph or blob columns, erroneous values, contradictions, mislabeled data).
- Resampling imbalanced data using scaling techniques.
- How to handle missing values and outliers (mean/mode/median imputation, delete).
- Creating new features to address goals for analysis.

Class Meeting Night – Wednesdays 6:00 pm – 9:00 pm. Short breaks will be taken throughout the session as determined by the mentor to allow for healthy stretching, refreshment, or to engage in a simple conversation with classmates 😊

Office Hours – Monday evenings 7:00 pm – 8:00 pm