

# Initial and Exploratory Analysis

- **Sayed Ahmed**
- PhD Studies in Electrical and Computer Eng. (McMaster University) (Partially Complete)
- Master of Engineering in Electrical and Computer Engineering (McMaster University)
- MSc in Data Science and Analytics (Toronto Metropolitan University/Ryerson)
- MSc in Computer Science (U of Manitoba)
- BSc. Engineering in Computer Science and Engineering (BUET)
- Extensive experience in Software Development and Engineering (primarily in Canada)
- Significant experience in Teaching
- Taught in Universities, Colleges, and Training Institutes



# Data Analysis Types: for Cleaning and/or Adapting the data

- Initial Analysis
- Exploratory Analysis

# Initial Analysis

## Univariate Analysis

- Deciding/Determining the dependent (target) variable
- Assigning the correct data types, appropriate column names
- **Address:** Inconsistencies, missing values, outliers
- Categorical variables with too many levels (address the issue)
- **(understand)** Distributions of the variables (is it right fit for the project)
- Imbalance in the dependent variable
- Time variables
- Univariate visualizations
- A detailed data dictionary
- Low variance filter

# Initial Analysis

## **Bivariate Analysis**

- Pairwise relations
- Pairwise visualizations
- Correlation analysis

# Initial Analysis

## Multivariate Analysis

- Multivariate relations
- Statistical tools

# Exploratory Analysis

## Normalizing

- Sub-setting the data
- Clustering

## **Others**

- Decision rules, association rules, n grams
- Time series analysis

## **Reference:**

Courses at Ryerson such as  
Research Methodology, Machine Learning,  
Internet