CGT 270 Data Visualization
Module 1
Week 3
Lab 3: Mining Data

The goal of this lab is to identify and implement techniques for mining data. In this lab you will identify patterns, extreme and subtle feature about data. You will identify basic descriptors for the data, and categorize data according to the specifications defined in the Parse Worksheet you completed in Week 2. After completing this lab, you will:

1. List at least three (3) questions you feel you can answer with the data sets you have acquired (Week 1) and parsed (Week 2).
2. Your questions must incorporate ALL three (3) of the data sets you've acquired from Lab 1: Tableau Dataset, Additional Dataset #1, and Additional Dataset #2
3. List any assumptions you are making in this stage of the data visualization process.

What you should be able to do (at the end of this lab):

Understand	***Describe*** the type of techniques to be used to better understand the data.
Apply	***Execute*** techniques and methods (statistical methods) on the data.
Evaluate	***Examine*** the resulting data and determine if it enables you to answer the question being solved.
Analysis	***Identify*** patterns, extreme and subtle features about the data.
Create	***Determine*** if the data can support the question to be answered.

In the table below list each variable in the Tableau dataset, its data type (parsing) and a basic statistical or mining technique that can be applied to better understand the variable. See the examples in the table below.

List the variables from Week 2's parsing lab and provide basic mining procedures.

Variable	Data Type	Basic mining procedure
duration_minutes	Integer	Average, max, min
duration_seasons	integer	String length
type	string	String length
title	string	String length
Date_added	string	String length
Release_year	integer	Average, max, min
rating	string	String length
description	string	String length
Show-id	integer	Average, max, min
director	string	String length
country	string	String length
cast	string	String length

Listed_in	string	String length
Imdb_id	string	String length
Title	string	String length
Plot	string	String length
Type	string	String length
Rated	string	String length
Year	integer	Average, max, min
Released_at	string	String length
Added_at	string	String length
Runtime	string	String length
Genre	string	String length
Director	string	String length
Writer	String	String length
Actors	String	String length
Language	String	String length
County	String	String length
Awards	String	String length
Metascore	Integer	Average, max, min
Imdb_rating	Float	Average, max, min
Imdb_votes	Integer	Average, max, min
S.no	Integer	Average, max, min
Name of the show	String	String length
Year of release	integer	Average, max, min
No of seasons available	Integer	Average, max, min
Language	String	String length
Genre	String	String length
IMDb rating	Float or integer	Average, max, min
Age of viewers	string	String length

Part I: Tableau Data Set

View the following link to complete the Mine Activity Worksheet for the Tableau Data set

https://tinyurl.com/MINE-Activity-Worksheet

Select **Training Data** as the Visualization Assignment

Part II: Additional Data set #1

View the following link to complete the Mine Activity Worksheet for the Tableau Data set

https://tinyurl.com/MINE-Activity-Worksheet

Select **Lab Assignment** as the Visualization Assignment

Part III: Additional Data set #2:

View the following link to complete the Mine Activity Worksheet for the Tableau Data set

https://tinyurl.com/MINE-Activity-Worksheet

Select **Lab Assignment** as the Visualization Assignment

What to submit: Three (3) pdf files. You should receive an email with a copy of your responses to the worksheet. Save the worksheet as a PDF (or print to PDF) as described below.

1. Mine Activity Worksheet for the Tableau Dataset
    a. Save as: **LastnameFirstInitial-CGT27000Fall2021-Lab3-Mine-TrainingData.pdf**
2. Mine Activity Worksheet for Additional Dataset #1
    a. Save as: **LastnameFirstInitial-CGT27000Fall2021-Lab3-Mine-Dataset#1.pdf**
3. Mine Activity Worksheet for Additional Dataset #2
    a. Save as: **LastnameFirstInitial-CGT27000Fall2021-Lab3-Mine-Dataset#2.pdf**

Part IV: Questions and Assumptions

Make sure you

- List **three (3) questions** you feel you can answer using the datasets you have acquired, parsed and mined. You MUST use complete sentences. Your questions MUST incorporate ALL three (3) of the data sets you've acquired.
  What is the most common year of movies/show releases?
  What is the most common rating?
  What is the average length of the title names?
- List three (3) **assumptions** you are making in this stage of the data visualization process where prompted in the worksheets.
  -The average release year of the movies or TV shows will not be the current year
  -The average rating will not be real low because poorly rated content probably won't be on there as much or stay on there as long
  -Most of the TV shows and movies will be from the USA