* [Overview](#gjdgxs)
* [Introduction](#30j0zll)
* [Introduction to Google Colab](#1fob9te)
* [Assignment](#3znysh7)
* [Teacher's Solution](#2et92p0)

# CSE 450 - Machine Learning & Data Mining

## Exploration 01: Netflix Movies

 [*Photo by Mollie Sivaram on Unsplash*](https://unsplash.com/photos/yubCnXAA3H8)

## Overview

In this data exploration, you will:

* Learn how to use Google Colab.
* Learn how to load, explore, and filter data with the Pandas library.
* Perform some basic data visualization with the Seaborn graphing library.

Throughout this exploration, when you're asked to use a new function or library, we'll usually provide a link to that function's documentation, or a tutorial related to it.

## Introduction

Throughout the course, we'll be using [Google Colab](http://colab.research.google.com) for most of our data exploration and modeling.

This assignment will teach you how Google Colab works and give you some practice using it. You'll also get some practice using the Python language, as well as some of the most popular libraries used by professional data scientists.

Learning Python and Friends

The assignments in this course require a fair amount of Python coding as well as the use of a few popular Python-based data science tools. We don't spend much time in this class explicitly learning Python or these tools, for a few reasons:

* In industry, you'll often be asked to complete a task or project that requires quickly coming up to speed on a new technology or software library that you've never used before.
* By the time you start your first job in data science, many of the tools and libraries will be different anyway, so we try to focus on teaching principles and helping you "learn how to learn" on your own so you can keep up with this rapidly changing field.
* As a 400-level student, we expect that at this point you have enough diligence and experience reading technical documentation and tutorials that you can handle this largely on your own.

However, we have put together [a guide containing some good references and tutorials](https://byui-cse.github.io/cse450-course/course/programming-resources.html), as well as [a guide with tips for reading technical documentation](https://byui-cse.github.io/cse450-course/course/reading-technical-documentation.html).

## Introduction to Google Colab

Google Colab is a cloud-based data science development environment. It includes support for the most widely used libraries and tools in data science, as well as providing access to optimized hardware for data analysis.

Google Colab uses "Jupyter Notebooks" for data exploration and analysis. If you've never heard of those before, Jupyter Notebooks are interactive documents that contain a mix of text and python code "cells".

Local Development vs Google Colab

While it's possible to use Jupyter notebooks locally, Google Colab provides a lot of great datasets, and access to cloud-based instances that have all the standard data science libraries pre-installed and configured for use.

You *can* opt to do your development locally, but keep in mind that if you run into problems with your local configuration, you'll be largely on your own to sort that out.

The following video will walk you through how to complete this first assignment, demonstrating how to use Google Colab, some things to watch out for when using it, and how to submit your assignments.

## Assignment

A consumer watchdog group wants to see if Netflix has more movies for adults or children.

Using a dataset containing metadata for all of the movies Netflix had available on their platform in 2019, use the [MPAA movie rating system](https://en.wikipedia.org/wiki/Motion_Picture_Association_film_rating_system#MPAA_film_ratings) to determine if they are correct.

Click on the Open In Colab button below to open a Google Colab notebook with the template for this assignment. Once you've completed the assignment, don't forget to take the corresponding quiz in Canvas.



## Teacher's Solution

Once you have absolutely exhausted all of your best efforts in solving the data exploration problems, and you are stuck on where to go next, you can [view the teacher's solution here](https://github.com/byui-cse/cse450-course/blob/master/notebooks/Exploration_01_Solved.ipynb).