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

# CSE 450 - Machine Learning & Data Mining

## Exploration 02: Cereal Marketing

 [*Photo by Haley Owens on Unsplash*](https://unsplash.com/photos/QdwrSyJV3_4)

## Overview

In this data exploration, you will:

* Practice more data manipulation and aggregation with the Pandas library.
* Practice using descriptive statistics on quantitative data.
* Perform more visualizations 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

As with our previous data explorations, this assignment uses [Google Colab](http://colab.research.google.com). For more information on using Google Colab, including how to submit assignments with it, please see the information in [Data Exploration 01](http://docs.google.com/exploration-01.html)

Seaborn Version

If you're trying to use a visualization in this assignment that you read about in the Seaborn documentation, but which gives you an error about being unavailable, it might be because that function was added in a more recent version than the one Google Colab uses by default.

You can see what version of seaborn you have installed in your Colab runtime with this commands:

import seaborn as sns  
sns.\_\_version\_\_

The [What's New in Each Version Page](https://seaborn.pydata.org/whatsnew.html) tells you what has been added with each version, so you can see if what you're trying to use is too recent.

If you want to upgrade the version of seaborn used by your runtime, which will give you access to the latest features, you can issue a command using the ! prefix:

!pip install seaborn --upgrade

Once this command finishes, the runtime will likely tell you that you need to restart the runtime for that change to take affect. This will also require you to re-run all of your code cells, which can result in reimporting an older version of seaborn again.

To get around this issue, put the upgrade command in your very first cell, restart the runtime, then run all your cells.

If you don't need or want to use the latest version of seaborn, you can ignore this information.

## Assignment

You're working as a data analyst at a cereal marketing company in New York.

In a strategy meeting, the marketing director tells you that in 2018, the US weight loss industry was worth over $72 Billion dollars, growing 4% compared to the previous year. [1](#2et92p0)

In contrast, sales of cold cereal fell 6% to $8.5 billion during the same time period. [2](#tyjcwt)

Cereal executives have approached the marketing company asking how they can somehow tap into the weight loss market growth to boost the sales of their cereal brands.

Your assignment is to analyze a dataset of nutritional information for major US cereals, and calculate some metrics that can be used by the marketing team.

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_02_Solved.ipynb).

1. [Business Wire — Feb 25, 2019](https://www.businesswire.com/news/home/20190225005455/en/72-Billion-Weight-Loss-Diet-Control-Market) [↩](#3dy6vkm)
2. [Wall Street Journal — Aug 20, 2019](https://www.wsj.com/articles/cereal-makers-try-again-to-jump-start-stale-sales-11566293404) [↩](#1t3h5sf)