

# Drexel University

## College of Computing and Informatics

### INFO 212: Data Science Programming I

#### Assignment 2

Due Date: Sunday, Nov. 12, 2023

This assignment counts for 10% of the final grade

#### A. Assignment Overview

This assignment provides the opportunity for you to practice with Python data analysis skills.

#### B. What to Hand In

Submit a completed this Jupyter notebook.

#### C. How to Hand In

Submit your Jupyter notebook file through the course website in the Blackboard Learn system.

#### D. When to Hand In

1. Submit your assignment no later than 11:59pm in the due date.
2. There will be a 10% (absolute value) deduction for each day of lateness, to a maximum of 3 days; assignments will not be accepted beyond that point. Missing work will earn a zero grade.

#### E. Written Presentation Requirements (if applicable)

Images must be clear and legible. Assignments will be judged on the basis of visual appearance, grammatical correctness, and quality of writing, as well as their contents. Please make sure that the text of your assignments is well-structured, using paragraphs, full sentences, and other features of well-written presentation.

Your Name:

#### ▼ Question [100 points]:

Download the data file "winemag-data-130k-v2.csv" and load it into a Pandas DataFrame using the column 0 as index. This data set contains 130k reviews for different wine varieties. It has the following columns:

1. country: The country where the wine is from.
2. description: A description or review of the wine.
3. designation: The vineyard within the winery where the grapes that made the wine are from.
4. points: The number of points WineMag rated the wine on a scale of 1-100.

5. price: The cost for a bottle of the wine.
6. province: The province or state where the wine is from.
7. region\_1: The wine-growing area in the province or state.
8. region\_2: Sometimes there are more specific regions specified within a wine-growing area.
9. taster\_name: Name of the taster.
10. taster\_twitter\_handle: Twitter handle for the taster.
11. title: The title of the wine review, which often contains the vintage.
12. variety: The type of grapes used to make the wine.
13. winery: The winery that made the wine.

## What to do:

You are asked to use the dataset to identify the most promising country, region, and wine variety for launching a new winery. You will make a recommendation supported by visualizations and data analyses. Complete the following activities. **You will write code or small programs in the activities. For each piece of code or program, you must:**

1. **write a description about what you are going to do**
2. **write the code or program with necessary comments**
3. **explain the plots or results in detail after running the code or program**

**Finally, You must back up any selections, recommendations, answers, etc. with visualizations or evidence extracted from data analysis.**

1. Analyze the number of wine reviews by country and region. Describe market size and consumer interest.
2. Analyze the average rating by country and region. Describe insights about reputation of winery locations.
3. Analyze wine varieties that consistently receive higher ratings across different regions. Describe popularity and consumer preference.
4. Analyze the wine prices by country, region, and variety. Describe insights into the potential profitability and market positioning.
5. Based on the data analysis, recommend a country, region, and wine variety that offers the best balance of market size, reputation, profitability, consumer preferences, and competitive positioning for launching a new winery.

