**Assignment 5 (Due per BrightSpace)**

1. Download the following dataset from kaggle.com & import it on your .ipynb notebook

[Glassdoor Data science Jobs - 2024 (kaggle.com)](https://www.kaggle.com/datasets/kuralamuthan300/glassdoor-data-science-jobs/code)

2. For the dataset, complete the following tasks.

1. Change it to satisfy the requirement of 3 NF. Provide ERD and data dictionary. The only tool allowed is **Visual Paradigm**. (Online version is not accepted.)
2. Using Numpy and Pandas, perform your analysis. Your codes must include the following. (2+ means two or more.)
   1. Function, Loop, Conditionals, Read from & Write to csv files
   2. 2+ Exploring data characteristics OR 2+ plots(visualization)
   3. 2+ OOP design elements *OR* Numpy Array and 2+ related function(s)
   4. Pandas DataFrame and 2+ related function(s)
   5. 2+ Data wrangling including dealing with missing values
3. Provide five meaningful questions and answers related to the dataset.
4. You can refer to any codes including generative AI. However, you need **to cite properly and should understand your code**. I may ask you **to modify** it to produce other results during the **code review**. If you fail to show your understanding, you **may lose significant points.** During code review, googling is allowed but Gen AI is **NOT** allowed.

3. Please submit MS word document, and Jupyter notebook file(.ipynb), and Visual Paradigm file. Your assignment report (MS word document) should include the following and follow APA format. You don't have to include everything in the report but should include what is needed to hit the rubric points.

* Cover page
* Introduction
* ERD (3 NF) & data dictionary
* Analysis including five questions and answers
* Conclusion
* References – should include the source if you refer to them in the report

4. In your naming of the variables, functions, classes, data types, and overall structure, you should be consistent and clear.

**Rubric**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 5 | 4 | 3 | 2 | 1 | 0 |
| ERD | - Includes all the necessary entities  - Includes all the relevant relationships  - Consistent and meaningful naming |  | - Includes part of necessary entities  - Includes part of the relevant relationships  - Some consistent and meaningful naming |  | Minimal effort has been made in some areas. | No effort has been made. |
| Python code | All requirements are met. | Four | Three | Two | One | No requirements are met. |
| Questions and answers | Highly Relevant five questions and answers. | Four | Three | Two | One | No meaningful Q & As |
| Introduction & Conclusion & Reference | Comprehensive contents are present & followed APA format. |  | Some relevant contents are present. |  | Minimal relevant contents are present | No relevant contents are present. |