BUST08039 Fundamentals of Programming for Business Applications

**Individual Assessment (50%)**

**Objective:**

Utilize Python to analyse a given data set, address a specific business-related question and visually represent your findings.

**Instructions:**

* **Dataset:** Reflect on the information of the given dataset and consider what types of business questions it can answer.
* **Formulate Business Question:** Develop a clear, concise business question that the chosen dataset can help answer. A good question should be straightforward but meaningful, aimed at uncovering insights that could potentially benefit a business scenario or decision-making process.
* **Data Preparation**: Ensure that the data is adequately and appropriately prepared for analysis. This may include data cleaning and manipulation tasks necessary to format and tailor the data for analytical purposes.
* **Data Analysis:** Use descriptive analysis to examine the dataset and answer your question. Focus on extracting insights through effective visual representations. Advanced models or algorithms are not required.

**Deliverable:**

Submit one Jupyter Notebook containing:

* **Python Code**: Provide the code needed to demonstrate your analysis process and results. Ensure it is well-organized, with concise comments explaining the rationale behind your actions. Focus on the key steps your audience needs to understand with an explanatory approach, while including all the necessary steps leading to your final results.
* **Graphical Representation**: Use one canvas/figure to showcase your findings. This can include a single standalone graph or a combination of subgraphs. You may use additional graphs during intermediary steps in your code, but your mini-report should feature only one final graph that clearly presents your main findings. Including more than one graph in the mini-report may result in a markdown.
* **Mini-report in Markdown**: Include a markdown cell that contains a mini-report (200 words +/- 10%). The report should address the following:

1. **Business Question:** Clearly outline the business question you are addressing.
2. **Relevance:** Explain why this question is important and worth exploring.
3. **Data Justification:** Describe why the specific attributes you selected from the dataset are suitable for addressing the question.
4. **Findings and Insights:** Present the insights your analysis has revealed, directly linking them to your visual representation and how they answer the business question.

You don't need to submit any data files for this assignment. Ensure your code can be executed with the assumption that the data file is stored in the same directory as your Jupyter Notebook.

To add a markdown cell for a report in Jupyter Notebook:

1. Select an empty cell where you plan to insert your report.
2. Press Esc followed by M to change the cell type to Markdown.
3. Alternatively, use the menu: Cell -> Cell Type -> Markdown.
4. Use Markdown syntax to format your text as needed. You may go to this link for more tips: <https://www.earthdatascience.org/courses/intro-to-earth-data-science/file-formats/use-text-files/format-text-with-markdown-jupyter-notebook/>
5. Run the cell to render and display the formatted text.

**Approved Libraries and Restrictions:**

For this assignment, you are restricted to using **only** the following libraries: pandas and matplotlib. The use of any additional libraries not listed here will result in markdown points from the final score.

The use of generative models such as ChatGPT is not allowed for any part of this assignment.

**Marking Rubric:**

The following table outlines the categories for assessment and their corresponding marks breakdown. The final grade for this assignment will be the average of your points in all categories and contributes 50% to your total grade for the course.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criteria & Category** | **Fail (0-40%)** | **Pass (40-50%)** | **Good (50-60%)** | **Very Good (60-70%)** | **Excellent (70%+)** |
| **A) Business Question and Answer:**  how well students define and respond to a relevant business question using data-driven insights | Vaguely defined question and/or answer with minimal logical or data connection. | Clearly defined question with insufficient business relevance; the answer is shallow or overlooks data potential. | Clearly articulated question and answer that leverage data appropriately with sound reasoning. | Well-articulated, clear, and relevant question and answer, demonstrating thorough logical analysis and optimal use of data. | Exceptionally well-articulated, relevant question and answer, with insightful data-driven analysis. |
| **B) Using the Data:** how well students select and prepare relevant data for answering the business question | Ineffective and irrelevant data use with poor preparation or choice of data for the question. | Adequate use of somewhat relevant data, with minimal data preparation for analysis. | Good selection and preparation of relevant data, addressing key requirements for analysis. | Strong choice of data and effective preparation, ensuring thorough readiness for analysis. | Excellent selection and preparation, ensuring data quality and integrity for insightful analysis. |
| **C) Visualization:** how effectively students use visuals to support and clarify their analysis | Graphs/visuals are confusing, misleading, or irrelevant. | Graphs/visuals are adequate but may offer limited clarity or could be misinterpreted. | Good visuals that enhance understanding of the data with minimal errors. | Clear and impactful visuals that strongly support and clarify the argument. | Exceptional visuals that enhance comprehension and strongly reinforce the argument. |
| **D) Code Quality & Structure:** assessment of code structure, readability, efficiency, and conciseness. | Poorly structured, difficult to read, inefficient code with frequent errors, redundancy, and minimal documentation. | Basic organization and readability, with some errors, redundancies, and limited documentation. | Organized, readable, and mostly efficient code with minimal redundancy, clear comments, and minimal errors. | Highly readable and efficient code, with effective handling of redundancies, comprehensive documentation, and strong error management. | Exceptionally well-structured, efficient, and concise code, with flawless readability, no redundancies, detailed documentation, and no errors. |

Guidance for markers:

A: Read mini-report, it is very rare that the business question is irrelevant. Check if the answer is logically correct and comprehensive, a good answer should not just stating the findings, but also insight and implications

B: Look for if the question raised by the student can be well answered by the chosen data attributes, if there is any data transformation, and finally (additional point) if the data has been cleaned

C: For the visual: descriptive and clear title and labels, good use of colors, choice of visual types, any overlapping/misleading/misinterpretation, any clear chance for annotation but missed

D: code conciseness, structure, comments, readability, and efficiency (mark up for the use of self-defined functions, but only if these functions are well-designed and serves a clear purpose).