**Python EDA FINAL PROJECT**

By now you should be comfortable with working on an end to end Exploratory Data Analysis in Python. For this final assignment/project, we would like you to **study and pick a dataset** from one of the links below. **Treat the project as a business case and solve the problem**. A total of **100 points** is allocated to this project.

Datasets:

* HR analytics: <https://www.kaggle.com/datasets/shivan118/hranalysis>
* Bike Share Analytics (geospatial analytics required, take longitude and latitude as numerical variables only): <https://www.kaggle.com/datasets/evangower/cyclistic-bike-share>
* Human Activity Analytics: https://www.kaggle.com/datasets/carlodipaola/bellabeat-dataset
* Industrial quality analytics: https://www.kaggle.com/datasets/edumagalhaes/quality-prediction-in-a-mining-process
* Food Chain Campaign( a bit more on statistics): <https://www.kaggle.com/datasets/chebotinaa/fast-food-marketing-campaign-ab-test>

We would like you to provide the following in your submission. We have also provided a project template to help you get started.

* A **google colab notebook** with all the code, this should include the plots with all the elements of a graph and a few insights in a markdown format just like the example notebooks during the lecture sessions.
* A **document writeup** making your business case with any important insights to solve the business case. You need to set up a business problem and solve it end to end.

You will be graded on the following rubric:

|  |  |  |
| --- | --- | --- |
| Section | Criteria | Points |
| **Project Scope** | Is the project goals and objectives mentioned? | 5 points |
| **Analysis** | Data is cleaned and transformed properly | 5 points |
|  | Basic metadata exploration and data columns to variable type have been mapped out clearly. | 5 points |
|  | At least 10 questions are asked and these are related to the business case. | 5 points |
|  | The student has analyzed the data using univariate analysis and written his insights | 5 points |
|  | The student has analyzed the data using bivariate analysis and written his insights | 5 points |
|  | The student has analyzed the data using multivariate analysis and written his insights | 5 points |
|  | The student has used the correct charts and plots for each type of analysis and interpreted them correctly. At least three charts for each type of analysis. | 10 points |
|  | The student has used different elements to describe the graph( labels for each axis, legend if required, title, aesthetics - styles, colors, size etc.) | 5 points |
|  | Use of statistics - Measures of central tendency(mean, median, mode, quartiles), spread ( standard deviation, variance, range, IQR ), Shape(Kurtosis, Skewness and Modality) and Frequencies, Co-occurrences and/or other statistical analysis method. | 5 points |
| **Code Quality** | No errors produced while running the notebook, code is readable. | 10 points |
| **Overall conclusions** | At least 3 important insights and final conclusions are included.  The insights should be interesting enough instead of just out of common sense. | 20 points |
|  | The conclusions are drawn logically and reasonably without logic fallacy. | 5 points |
|  | The project included potential actions or next steps that could be taken, based on the analysis | 10 points |

**Best of Luck** 🙂