Final assignment Business Statistics –

“Professional product” (7023BST_BP) as per course catalogue

The final assignment for the course Business Statistics contains 2 sub-assignments. In both sub-assignments, you will work in groups of 4 students. The group composition will remain the same in both sub-assignments. In sub-assignment 1, each group must choose one dataset from the 2 datasets (RetailRocket & Tmall) that are made available to you. In sub-assignment 2, you do not have a choice.; the dataset (Tao Yin) is compulsory.

Together sub-assignments 1 & 2 contribute equally to the grade you receive for the “Professional product” (7023BST_BP) as stated in the course catalogue. The grading criteria for both sub-assignments are indicated in the table below. The group composition is made available through Brightspace.

You are not allowed to use any other data than the dataset provided. You will formulate your research questions, conduct exploratory data analysis, develop and test hypotheses, test your assumptions and propose a statistical model or create clusters if/where applicable. In creating clusters, you will have to conduct feature engineering (for example using one-hot encoding).

The deliverables

As mentioned earlier, you will work in groups of 4, and the datasets are made available on Brightspace. You will present your findings in a Jupyter-Python notebook (uploaded as .ipynb) and upload a short video of maximum 10 minutes per group (uploaded as .mp4), where each group member takes turns in explaining the findings. Your individual contribution will be assessed separately. Each team member should indicate clearly the lines he/she has coded.

Grading criteria

Please read the grading criteria and check that you understand how it is calculated.

	Weights	100%	70%	50%	30%
		Good	**Sufficient**	**Insufficient**	**Poor**
1. Translate research questions and data into multivariate data analysis.	15	All formulated research questions have been logically translated into data analyses	Most formulated research questions have been logically translated into data analyses	Some of the formulated research questions have been logically translated into data analysis	There is no clear logic between the formulated questions and the data analysis
2. Organize and document the data analysis process.	20	The analysis performed is thoroughly and exhaustively embedded in the theory of Statistics	The analysis performed is incomplete in its thoroughness and completeness in the theory of Statistics	The analysis performed is neither thorough nor exhaustive in the theory of Statistics	Hardly any analysis undertaken.
3. Program and analyse data in Python.	20	Code is extensively commented and is efficiently written w.r.t. logic, libraries and usage of variables	Code is extensively commented but is inefficiently written w.r.t. logic, libraries and usage of variables	Code is not extensively commented and is inefficiently written w.r.t. logic, libraries and usage of variables	Code lacks logic and is difficult to comprehend
4. Draw conclusions from the outcomes of multivariate data analyses.	15	The right conclusions have been drawn and they are exhaustive	The right conclusions have been drawn but they are not exhaustive	Not all conclusions are correct	Conclusions are very superficial
5. Professionally report the outcomes of multivariate data analyses.	20	The outcomes have been illustrated exhaustively with the right types of graphs & plots	The outcomes are illustrated exhaustively but not always with the right types of graphs & plots	The outcomes illustrated are limited in nature	Less than 5 outcomes illustrated
6. Reflect on processes and issues associated with team work in data analysis settings	10	The reflection is critical and provides recommendations at an individual level	Reflection is critical but the recommendations provided are superficial	Reflection is not crtical enough at an individual level	No reflection provided
	100				