Final course project: grade distribution

PART 1: Data collection and cleaning:

Criterion	description	weight
Automation	If a data source offers an API for data access, you are required to use it. If none is available, you can still automate the data download from your script, but it's not obligatory.	30%
Data cleanliness	-Cleanliness relates to the presence of invalid content; ie, erroneous, duplicated, corrupted, ill-formatted, or missing values. -cleaning must be conservative and not discard large chunks of data at once.	30%
Data Tidiness	Tidiness refers to the structure of the dataset. It should be clear and understandable, with descriptive column names, and structured in a way that best serves the analysis requirements.	30%

PARTS 2 & 3: Descriptive analysis:

Criterion	description	weight
correctness	Your statistics are calculated correctly from the given data.	50%
Eliminating bias	You try your best to eliminate bias, and control for confounding variables while answering statistical questions (eg, calculating percentages of the total population instead of counts, conditional probabilities	20%

	instead of joint…etc). If you suspect the presence of inherent biases you cannot control, due to the nature of the dataset, or the collection process etcetera, you should cite them and explain their effect.	
Visualization	Answers to each analysis question should be accompanied by at least one descriptive visualization. You are not limited as to its type, or content. You are free to combine, separate, transform, or standardize your data in any way that makes your visualization clearer, and more efficient. You will be graded on the quality of your graphs.	15%
Insights/ commentary	Each analysis question is accompanied by a comment where you interpret the outcome, and explain why you see the trends you do, if any.	15%

PART 4: Hypothesis testing:

Criterion	description	weight
Hypothesis formulation	Hypothesis is correctly formulated and adequate to the claim under test.	40%
Testing procedure	Chooses an adequate hypothesis test and implements it correctly.	35%
Result and commentary	Reports the result of the test and interprets the validity of the claim accordingly.	25%

PART 5: Regression analysis:

Criterion	description	weight
Model building	Variables are transformed appropriately, and the model is implemented correctly (by hand or using a code library: statsmodels, sklearn, .. etc)	60%
Assessing your result	Ability to interpret model parameters, and answer questions about the relationships between predictor and target variables.	40%

PART 7: DOCUMENTATION

Criterion	description	weight
comprehensiveness	Your reports document all the components of your work, and provide additional insights into its significance and limitations.	50%
organization	Your reports are well organized and well written. Your business report presents your findings in a cohesive, compelling way with no unnecessary technical detail.	20%
Quality of visualizations (business)	Your visualizations are clear, appropriate to context, and not cluttered or misleading.	10%
Commentary (business)	Your insights should be relevant, non-trivial, and well-researched.	10%
Ability to explain limitations, and unexpected outputs	You should provide well-researched explanations to unexpected results, or give your educated opinion in case no explanation can be found.	10%

PART 8: Presentation:

Criterion	description	weight
Quality of slides	Slides should be neat, concise, comprehensive, and visually compelling.	40%
Presentation skills	Present your findings in a cohesive way without exceeding the time limit.	20%
Ability to answer questions		40%