

Project 1 Submission Page

[Start Assignment](#)

Due Tuesday by 11:59pm **Points** 100 **Submitting** a file upload **Available** until Aug 10 at 11:59pm

Project 1 - Statistical Inference

[Full details of the project here.](#)

Deadline:

August 10, 2021 (Tuesday) 11:59 PM

Percentage:

20% (part of the 40% for Projects)

Deliverables:

Zip file containing:

- Jupyter Notebook file – *ipynb* file
- Other Python 3 files (if necessary) – *py* files
- Dataset files – *csv* files

Submission guidelines:

Filename format:

CSMODEL-Project1-<Section>-Group<#>.zip

Only one member of the group needs to submit the assignment.

[2021T3] Project 1 Rubric

Course Chat

Criteria	Ratings				
Description of Data and Method of Collection	5 pts Complete An overview or description of the data is provided in the Notebook, including the data collection process, and its implications on the types of conclusions that could be made from the data.	2 pts Incomplete An overview or description lacks details, or the description does not include the data collection process or its implications to the conclusions.			
Description of Variables / Observations / Structure of the Data	5 pts Complete A description of the variables, observations, and/or structure of the data is provided. It should be clear to the reader what each part of the dataset represents without having to go through external resources.	2 pts Incomplete A description of observations, variables, or structure is present but is missing key aspects of the data.			
Data Cleaning	10 pts Complete The necessary steps for cleaning are performed, including explanations for every step. If no data cleaning is done, there should be a justification on why it is not needed.	5 pts Incomplete Preprocessing and cleaning steps are performed but lacks explanation, or the cleaning done is insufficient for the dataset.		justification is weak or incorrect.	
Exploratory Data Analysis 1	10 pts Complete The first exploratory data analysis question is sufficiently answered, and both the appropriate numerical summaries and visualizations are presented.	5 pts Incomplete The first exploratory data analysis question is not sufficiently answered, or the appropriate numerical summaries or visualizations is not presented.	0 pts No Marks There is no analysis done for the first exploratory data analysis question.		10 pts
Exploratory Data Analysis 2	10 pts Complete The second exploratory data analysis question is sufficiently answered, and both the appropriate numerical summaries and visualizations are presented.	5 pts Incomplete The second exploratory data analysis question is not sufficiently answered, or the appropriate numerical summaries or visualizations is not presented.	0 pts No Marks There is no analysis done for the second exploratory data analysis question.		10 pts
Exploratory Data Analysis 3	10 pts Complete The third exploratory data analysis question is sufficiently answered, and both the appropriate numerical summaries and visualizations are presented.	5 pts Incomplete The third exploratory data analysis question is not sufficiently answered, or the appropriate numerical summaries or visualizations is not presented.	0 pts No Marks There is no analysis done for the third exploratory data analysis question.		10 pts
Research Question 1	5 pts Complete The first research question is clearly defined. The importance to the researcher and the community is explained convincingly. The research question arose from the EDA.	2 pts Incomplete The first research question is defined but either is not clear, or its significance is not explained convincingly. The research question did not arise from the EDA.	0 pts No Marks The first research question is not defined.		5 pts

Criteria	Ratings			Pts
Research Question 2	5 pts Complete The second research question is clearly defined. The importance to the researcher and the community is explained convincingly. The research question arose from the EDA.	2 pts Incomplete The second research question is defined but either is not clear, or its significance is not explained convincingly. The research question did not arise from the EDA.	0 pts No Marks The second research question is not defined.	5 pts
Statistical Inference 1	10 pts Complete The appropriate statistical inference technique is used to answer the first research question.	5 pts Incomplete The statistical inference technique that is used to answer first research question is applied in an insufficient way. Some preprocessing steps are not performed to prepare the data for statistical inference to answer the first research question.	0 pts No Marks Statistical inference is not performed to answer the first research question.	10 pts
Statistical Inference 2	10 pts Complete The appropriate statistical inference technique is used to answer the second research question.	5 pts Incomplete The statistical inference technique that is used to answer second research question is applied in an insufficient way. Some preprocessing steps are not performed to prepare the data for statistical inference to answer the second research question.	0 pts No Marks Statistical inference is not performed to answer the second research question.	10 pts
Insights and Conclusion 1	10 pts Complete The insights and conclusions to the first research question are stated clearly and backed up with statistical evidence when needed.	5 pts Incomplete The insights and conclusions to the first research question are stated but not clearly enough, or some statistical evidence is lacking.	0 pts No Marks No insights or conclusions are presented for the first research question. The insights or conclusions are based on an inappropriate statistical inference technique applied to answer the first research question.	10 pts
Insights and Conclusion 2	10 pts Complete The insights and conclusions to the second research question are stated clearly and backed up with statistical evidence when needed.	5 pts Incomplete The insights and conclusions to the second research question are stated but not clearly enough, or some statistical evidence is lacking.	0 pts No Marks No insights or conclusions are presented for the second research question. The insights or conclusions are based on an inappropriate statistical inference technique applied to answer the second research question.	10 pts
Total Points: 100				