**Pontifícia Universidade Católica do Paraná (PUCPR)**

**Data Science**

**Computer Science**

**Professor Jean Paul Barddal**

**CHECKPOINT – PEER REVIEW**

In this part of the work, you and your team will peer review another’s team work. Below, you will find details about how you should assess another’s team work. The rubric provides insights on what is expected to reach different grades (0 to 4, 4 to 7, and so forth).

In addition to providing a grade to each item, you should provide feedback (evidence) on why this grade was assigned, as well as suggestions on how the other team should enhance their work.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rubric item** | **Weight (%)** | **0 to 4** | **4 to 7** | **7 to 9** | **9 to 10** |
| **Statistical description of the dataset** | 10% | The report does not provide a statistical description of the dataset. | The report provides a statistical description of the dataset. | The report provides a statistical description of the dataset and it is accompanied by a textual description of at least 5 variables. | The report provides a statistical description of the dataset and it is accompanied by a textual description of at least 5 variables that provided non-obvious insights from data. |
| **Univariate data analysis** | 30% | The report discusses less than 20 variables or their analysis does not include skewness, kurtosis, and statistical descriptions of the variables chosen. | The report discusses 20 or more variables alongside their skewness, kurtosis, and statistical descriptions of the variables chosen. | The report discusses 20 or more variables alongside their skewness, kurtosis, and statistical descriptions of the variables chosen and provides a textual description discussing these variables. | The report discusses 20 or more variables alongside their skewness, kurtosis, and statistical descriptions of the variables chosen and provides a textual description discussing these variables with non-obvious insights. |
| **Multivariate data analysis** | 30% | The report contains less than 8 multivariate data visualizations. | The report contains 8 or more multivariate data visualizations, yet, it does not provide textual descriptions of the main findings. | The report contains 8 or more multivariate data visualizations, accompanied by their respective textual descriptions. | The report contains 8 or more multivariate data visualizations, accompanied by their respective textual descriptions. At least one of the visualizations and textual description account for a newly generated variable. |
| **Effective data visualization** | 30% | The report does not provide at least 3 visualizations, or such visualizations do not appear earlier in the report. | The report includes at least 3 different visualizations, yet, such visualizations lack evidence that they were improved. | The report includes 3 or more visualizations and these have adequate size, colors, and textures. | The report includes 3 or more visualizations and these have adequate visual components and adhere to Evergreen’s checklist. |

|  |  |
| --- | --- |
| **Team name: Jardins** | **Evaluators’ team name: Gustavo Hammerschmidt** |
| |  |  | | --- | --- | | **Rubric Item:** Statistical description of the dataset | **Grade (0 to 10)**: 9 | | **Feedback: Mudanças feitas na base dão poucos indícios de suas aplicabilidades: faltou descrever e organizar melhor as informações da seção “loading your data”; faltou adicionar a coluna ‘SAFRA’ aos dataframes de treinamento e teste.**  **Suggestions for improvements: Organizar e descrever mais o uso de cada instrução ou motivos para as alterações.** | |  |  |  | | --- | --- | | **Rubric Item:** Univariate data analysis | **Grade (0 to 10)**: 7 | | **Feedback: São apresentadas 24 variáveis, juntas de suas informações tais quais: (1) a descrição da variável, mas não o motivo de sua seleção para a apresentação de seu gráfico; (2) os valores skewness, kurtosis, média e desvio padrão(foi adicionada ao conjunto, também, a moda do set: informação não solicitada) foram automaticamente calculados – o que, para skewness e kurtosis, pode não condizer com a disposição dos valores em gráfico(Teorema dos sets de Ambrose), e isso implicou em uma incorreta representação da informações da primeira variável apresentada, TEMPOCPF; e (3) representação de uma mesma variável em muitos gráficos, dando a impressão de não se saber a melhor forma de representar o conjunto de valores em gráfico, e sujeitando os visualizadores a pouca noção do porquê de tal escolha senão apenas a descrição da variável. Vale ressaltar os devidos cuidados que a equipe jardins não teve em separar as variáveis em subseções e padronizar os conteúdos subsequentes de cada uma de forma uniforme.**  **Suggestions for improvements: Organizar e uniformizar as seções por variável analisada, tomar o devido cuidado com as informações apresentadas e apresentar uma descrição do motivo de escolha de cada variável e sua importância para a pesquisa como um todo.** | |  |  |  | | --- | --- | | **Rubric Item:** Multivariate data analysis | **Grade (0 to 10)**: 8 | | **Feedback: As análises feitas, no geral, possuem bons porquês, mas, no entanto, suas hipóteses são binárias ao ponto de serem óbvias: duas análises apenas trouxeram um novo comportamento dos dados antes não vistos nas correlações destes. Importação de bibliotecas foram feitas nas seções, e, por padrão, deveriam ter sido feitas na primeira seção.**  **Suggestions for improvements: Organizar as importações de bibliotecas e formular hipóteses que apontem resultados não esperados antes e que sejam menos óbvios ou simples.** | |  |  |  | | --- | --- | | **Rubric Item:** Effective data visualization | **Grade (0 to 10)**: 5 | | **Feedback: As três apresentações finais feitas advêm das análises multivariadas feitas; contudo, os resultados apresentados não sofreram aprimoramento visual como: mudança de tamanho, cor ou textura, muito menos estão nos conformes da checklist de Evergreen. Ademais, as explicações feitas sobre o motivo das escolhas foram sintetizadas e não parecem ter sido feitas *de forma a apresentar a um público que não detém conhecimento prévio do conjunto de dados*. Recomenda-se a divisão das apresentações finais em subseções.**  **Suggestions for improvements: Organizar as apresentações finais em subseções, melhorar a apresentação dos gráficos e informações mostradas, e conferir maior eloquência as apresentações de modo a explicar a um público leigo os resultados obtidos.** | | | |