BBI Data Science Programming Challenge – Dec 2020

Coding Instructions:

- Two datasets **SpendData.csv** and **TestData.csv** are provided for **Question 1**. You are free to decide how best to use them.
- Refer to **MockSurveyData.xls** for **Question 2.**
- All source codes must be provided with comments along with instructions on how to execute the code.
- You are free to consider other tools. Regardless of the tool you use, source code / workings must be provided and clearly documented
- You do not need to augment this dataset with any external data. However, if you choose to code needs to be properly commented.
- Incomplete or partially commented code will not be evaluated.
- The solution along with the original code and screenshots of the results has to be shared with the recruiting team by responding to the programming challenge email - **recruiting@bbinsight.com.**
- Preference will be given to those programs that have better performance, follow universal coding standards, and are submitted before the deadline.

Question 1

- **Some of the respondents have been tagged as belonging to group 1 – 6. However, due to a data calculation issue, some of the respondents have had their groups (pov6) missing.**

- **Build a model that will classify these respondents back into one of the 6 groups.**

    a. Please explain the choice of metric / evaluation criterion used

    b. What are the assumptions you made when building this model?

    c. What were the approaches you considered? Please explain the reason for the technique / approach used as well as the pros and cons.

    d. Please explain under what conditions will the model you choose be not appropriate

    e. How confident are you of the model's robustness and how would you explain the model's performance?

f. Why is your model performing well / not well?

g. Was any feature engineering required? If yes, what were they. If no, why?

- **Please include your supporting scripts/output from tool of your choice.**

<div style="border:1px solid black; padding:10px;">

<center>**Question 2**</center>

</div>

- **With reference to the MockSurveyData.xls dataset for visitors from a particular country,**

    a. List the areas you would look into when it comes to data preparation

    b. Highlight any data idiosyncrasies that you might have observed

    c. Put together a short presentation/document to tell a story about the key insights and observations, taking into consideration dataset limitations.

- **Please include your supporting scripts/output from a tool of your choice.**

If you need any further information, please feel free to reach out to us at recruiting@bbinsight.com, Thank you!