Data Analyst

Technical Assessment

This assessment is designed to evaluate your conceptual skills in data analysis. Please provide your answers in the space after each question.

The primary objective is to assess your problem-solving approach and methodology, rather than the correctness of results. Each answer will be evaluated based on its relevance to the given context.

Data set used for this code has already been provided. More information can be found in the link below:

<https://open-data.bouldercolorado.gov/datasets/10c2e57f741b45428c492f67aeb98b1b_0/about>

1. You have had a look at the data and done some preliminary exploratory analysis. You now have a meeting with the client representative to discuss the data for spend analysis purposes. What are some of the questions you may ask the client in the meeting? What other data would you ask for, if any, to enhance analysis?
2. What are the expectations from this Spend Analysis?
3. Departmental priority to spend in the company and cost cutting for the departments which are least priority.
4. All department\_desc columns which contains ‘NA’ has fund\_desc as 0. Since we have some inconsistencies in the dataset like department\_desc, it is difficult to plot a dashboard for precisions and taking decisions. Get details about these inconsistencies and if they can be handled.
5. We can observe that November is the month that has spend highest from past 5 years. Why? Is it the seasonal trend, are there regular maintenance at that time of the year?
6. How are you going to use this spend analysis for future savings in terms of department, source, and company priorities?
7. What is the assigned total amount for each department in a month or year?
8. What is the source you are using: Azure, AWS, Oracle, etc.,? How can we reduce the resources more cost efficiently and reliable?
9. Check if the suppliers are over charging and suggest alternate suppliers who can fulfill the work for lesser amount.
10. List down what steps you would take to ensure that data inconsistencies such as duplications and outliers are dealt with.
11. Firstly, understand the data. What each column says about the data.
12. Check for null values, irregular cardinality, and outliers.
13. Remove the null values if not important or small dataset.
14. If large dataset, then impute null values with mean/median (numeric) and mode (categorical).
15. Check for datatype and convert to appropriate one where necessary.
16. The client wants to stay on top of key updates and track progress regularly. What steps would you take to ensure the client is up to date with the project progress?
17. Weekly meetings and update the analysis done on a weekly basis.
18. Use sharepoint to track live progress of the project analysis where they can easily access it at any point of time.
19. Make sure to upload all the files and progress reports at the end of the day to the sharepoint album.
20. Make a live Power BI Dashboard and give access to client for up to date analysis done.
21. The client informed you that the data contains some sensitive information that he wants excluded from the analysis. One such case is where the account description includes similar words to ‘salary’ or ‘wages’. What steps would you take to ensure this sensitive data is removed?

* Check with the client what are the sensitive data that should be excluded from the analysis and we can create a query that filters the keywords related to them in Python Jupyter or Power BI DAX.