



Code Test

Junior Data Analyst

1. Using Python, parse the provided csv file and load the data into a pandas DataFrame.
2. Perform data cleaning and preprocessing tasks on the DataFrame to ensure data quality. This includes handling missing values, correcting data formats, and removing duplicates.
3. Explore the dataset and answer the following questions:
 - a. How many unique clients are there in the dataset?
 - b. Which store has the highest total sales?
 - c. How many transactions were made in each store?
 - d. What is the distribution of product purchases across different stores?
4. Create visualizations to represent the insights gained from the data analysis. This can include bar charts, pie charts, or any other appropriate visualizations.
5. Calculate the total amount spent by each client and identify the top-spending clients.
6. Write a SQL query to find all clients from the store "King St" whose latest purchase was a "cappuccino". Sort the clients by the total amount spent on all purchases. Show the client names and their total spending.
7. Validate the query's optimization by explaining its execution plan. Discuss any indexes or optimizations that can be applied to improve its performance.
8. Execute the query using Python and display the results.
9. Write a summary of the key findings from the data analysis and suggest actionable insights based on the results that would help this customer improve their business.
10. Remember to document your code and provide clear explanations for each step.

Good luck!