Data Scientist II Technical Challenge

Dear Candidate,

Thank you for applying to this vacancy.

Please find below the questions for the written assessment component of the selection process.

Part 1: Exploratory Data Analysis

You will be provided with a dataset containing information about Customer transactions.

Your task is to:

- Load the dataset and perform basic data cleaning (e.g., handling missing values, correcting data types).
- Conduct exploratory data analysis to understand the main characteristics of the data.
- Visualize key insights using appropriate plots (e.g., histograms, bar charts, scatter plots).

Deliverables:

- Summary statistics of the dataset.
- At least three different visualizations with explanations.

Task 2: Predictive Modeling

Using the cleaned Customer Transactions dataset from Task 1:

- Identify a target variable for prediction (e.g., predicting customer churn, transaction amount).
- Develop a predictive model using an appropriate machine learning algorithm.
- Evaluate the model's performance using relevant metrics (e.g., accuracy, precision, recall, RMSE).

Deliverables:

- Explanation of the chosen target variable and model.
- Model training and evaluation process.
- Performance metrics and interpretation.

Task 3: Natural Language Processing

Using the Customer Reviews dataset:

- Preprocess the text data (e.g., tokenization, stopword removal, stemming/lemmatization).
- Perform sentiment analysis on the reviews.
- Visualize the distribution of sentiment scores.

Deliverables:

- Preprocessed text data.
- Sentiment analysis results.
- Visualizations of sentiment distribution.

Task 4: Real-world Scenario

Consider the following business problem:

Your company wants to improve customer satisfaction by understanding the main topics and sentiments expressed in customer reviews. Your task is to:

- Use topic modeling to identify the main topics in the customer reviews.
- Summarize the findings and suggest actionable insights for business improvements.

Deliverables:

- Explanation of the topic modeling process.
- Summary of the main topics identified.
- Actionable insights based on the analysis.

Submission

Please ensure that your **GitHub project** contains all the required deliverables and is organized in a clear and logical manner.

Good luck!