Technical Test: Data Science Coach Role

Test Details

Time Limit: 6–8 hours

Tools Allowed: Python/R, Jupyter Notebook, Excel, SQL, or any preferred tool.

Deliverables:

1. Jupyter Notebook or R script with detailed code and comments.
2. PDF or PowerPoint presentation summarizing findings, models, recommendations, data science trends, and challenges.

Scenario: Retail Data Analysis and Forecasting

You are hired as a data scientist to analyze sales data from a retail company, help them understand their business performance, and predict future sales. Additionally, the company is interested in understanding the latest data science trends and challenges, along with an overview of key algorithms used in the field.

Tasks

1. Data Exploration and Cleaning

- Handle missing or inconsistent values.
- Summarize the dataset (e.g., number of rows, columns, duplicates, etc.).
- Identify trends and patterns (e.g., seasonality, best-selling categories).

2. Data Analysis

Answer the following questions:

1. What are the top 5 best-selling products and categories by revenue?
2. Which region has the highest average revenue per store?
3. How does sales performance vary across different months?

3. Visualization

- Create meaningful visualizations (e.g., bar charts, line plots, heatmaps) to showcase trends and insights.
- Include at least one advanced visualization (e.g., sales heatmap by region and month).

4. Predictive Modeling

- **Objective**: Predict the daily revenue of a store based on historical data.
- **Steps**:
    1. Prepare the data (e.g., feature engineering, splitting into train/test).
    2. Train at least two models (e.g., Linear Regression, Random Forest, XGBoost).
    3. Evaluate models using appropriate metrics (e.g., RMSE, MAE, R^2).
    4. Select the best model and justify your choice.

5. Recommendation

- Based on your findings, provide actionable insights to improve sales.
- Suggest 2–3 strategies for increasing revenue based on your analysis.

---

6. Presentation Task: Data Science Trends and Challenges

Prepare a presentation (5–10 slides) to showcase your understanding of the current landscape of data science. Include the following sections:

1. **Overview of Data Science Trends**
    - Discuss emerging technologies (e.g., generative AI, large language models, AutoML, etc.).
    - Highlight key trends such as ethical AI, data privacy, and real-time analytics.
2. **Challenges in Data Science**
    - Address common challenges such as data quality, scalability, model interpretability, and deployment.
3. **Recap of Key Algorithms in Data Science**
    - Provide an overview of commonly used algorithms (e.g., regression, decision trees, clustering, neural networks, etc.).
    - Include a table summarizing the algorithms, their applications, strengths, and limitations.
4. **Recap of Analysis and Recommendations**
    - Summarize your analysis and predictive modeling outcomes from the technical test.
5. **Personal Perspective**
    - Share your thoughts on the future of data science and your approach to overcoming challenges.

PS : Link to data set : https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset/data