Data Analyst Technical Test - Create a Visualization from a dataset

Tags

Technical Test Instructions for Data Analyst Role

Objective:
You will use the Alzheimer's Disease Deaths dataset from Kaggle to clean, analyze, and visualize the data. Additionally, you will integrate data from an open API using Python to enhance your analysis. Your goal is to derive comprehensive insights and findings from the combined datasets. You can add in any other datasets you find.

Dataset: Death Alzheimer's Dataset on Kaggle

Open API:
Choose one of the following open APIs to integrate with the Alzheimer's Disease Deaths dataset:

1. CDC Wonder API
2. WHO Global Health Observatory API
3. OpenWeatherMap API
4. US Census Bureau API
5. HealthData.gov API

Steps to Complete the Test:

1. **Download the Dataset:**
    - Visit the Kaggle link provided above.
    - Download the dataset to your local machine.
2. **Data Cleaning:**

- Inspect the dataset for any missing or inconsistent data.
- Clean the dataset by handling missing values, correcting any inconsistencies, and ensuring the data is in a usable format.

3. **Data Integration:**
    - Choose one of the open APIs listed above.
    - Retrieve relevant data from the chosen API using Python.
    - Integrate the API data with the Alzheimer's Disease Deaths dataset.
4. **Data Analysis:**
    - Perform exploratory data analysis (EDA) on the combined dataset.
    - Use descriptive statistics to summarize the datasets.
    - Identify any patterns, trends, or anomalies in the data.
5. **Data Visualization:**
    - Use Looker Studio (formerly Google Data Studio) to create visualizations representing the key findings from your analysis.
    - Ensure at least one visualization combines data from both the Alzheimer's Disease Deaths dataset and the API data.
    - Use various types of charts (e.g., bar charts, line graphs, scatter plots) to effectively communicate the insights.
6. **Insights and Findings:**
    - Write a report summarizing your insights and findings from the combined datasets.
    - Highlight any significant trends, correlations, or observations.
    - Provide recommendations or conclusions based on your analysis.
7. **Submission:**
    - Submit your cleaned dataset, analysis code (in a Jupyter notebook or script), Looker Studio report link, and the final written report.
    - Ensure your code is well-documented and easy to follow.

Evaluation Criteria:

- Accuracy and thoroughness of data cleaning.
- Depth and quality of data integration and analysis using Python.
- Clarity and effectiveness of visualizations in Looker Studio.
- Insightfulness and relevance of findings.
- Quality of documentation and reporting.

Deadline:
Please submit your completed work by 30 June 2024

Additional Notes:

- Use Python for your analysis, preferably with libraries such as Pandas, NumPy, Matplotlib, Seaborn, and appropriate libraries for API integration (e.g., requests).
- Ensure your visualizations in Looker Studio are clear and well-labeled.
- Your report should be concise, well-structured, and free of grammatical errors.

If you have any questions or need further clarification, please feel free to reach out.

Good luck!