Data Science Internship Challenge

---

1. What algorithm, machine learning or AI approaches would you take to find anomalies in the duration of a span?

Ans:-

To find anomalies in the duration of a span, I would consider the following approaches:

1. **Statistical Methods**:
    - **Z-Score**: Calculate the z-score for each span duration to identify outliers.
    - **IQR (Interquartile Range)**: Use the IQR to detect outliers by identifying spans that fall below Q1 - 1.5*IQR or above Q3 + 1.5*IQR.
2. **Machine Learning Methods**:
    - **Isolation Forest**: An unsupervised learning algorithm that works well for anomaly detection by isolating anomalies.
    - **One-Class SVM**: A type of SVM that is trained on normal data and can identify anomalies as deviations from the normal pattern.
    - **Autoencoders**: Neural networks trained to reconstruct input data. Anomalies can be detected by high reconstruction errors.
3. **Time Series Analysis**:
    - **ARIMA/SARIMA**: Time series models that can forecast expected durations and identify anomalies as deviations from the forecasted values.
    - **LSTM Networks**: Recurrent neural networks that can learn temporal dependencies and detect anomalies in sequential data.

2. Why do you think that approach is a good approach?

Ans:-

- **Statistical Methods**: Simple and effective for small datasets or when the distribution of data is known.

- **Machine Learning Methods**: Suitable for complex datasets with non-linear relationships. They can handle high-dimensional data and learn patterns without explicit programming.
- **Time Series Analysis**: Effective for data with temporal dependencies, allowing for the detection of anomalies based on historical patterns.

These approaches provide a comprehensive toolkit for detecting anomalies in span durations, each with its strengths depending on the nature of the data and the specific requirements of the analysis.
