Evaluation and Assessment of AIRD-DeepChain Candidates

To get started, follow these steps:

* Read the test instructions: Review the test instructions and requirements carefully to understand what is expected from you.
* Begin working on the test: Use your preferred coding environment or editor to solve the test. You can use internet resources, such as documentation, tutorials, or stack overflow, to help you understand concepts or find solutions to specific problems. However, ensure that you write your own code and avoid directly copying and pasting code without understanding or proper attribution.
* Document your code: As you solve the test, it's important to provide clear and concise documentation for your code. This includes adding comments to explain your thought process, approach, and any relevant details.
* Test your solution: After completing the test, test your code to ensure it functions as expected and meets the requirements specified in the instructions.
* Include proper documentation: Along with your code, include a README file or any other required documentation that provides clear instructions on how to run your solution and any other relevant details.
* Rename the Solution files according to the following naming convention: Q1-Sol.py/Q1-Sol.ipynb or accordingly.
* Send the zipped folder containing containing solution files to the following mail: [Human Resource Department](mailto:hr@deepchain.pk)

Remember to manage your time effectively to ensure you complete the test within the given 4-hour time frame. Good luck with your test!

**Question # 01 : Statistical Analysis of the data - Weightage 30 %**

**Time Required : 40 mins**

The task is to explore the data and present the critical insights using the statistical evidence. Following steps should be performed:

* Explore the distribution of each variable in the dataset and draw the distributions using histograms, box plots, or density plots.
* Calculate and analyze the correlation between each variable in the dataset. Identify any strong positive or negative correlations and draw conclusions based on the findings.
* Identify outliers in the dataset and provide insights about their presence. Determine the underlying reasons for the outliers by examining the context of the data.
* Perform statistical tests to compare the distributions of different variables. Present the results of the statistical tests and interpret their implications for the datasets.
* **Challenge** : Formulate a hypothesis about whether the variable **A** (presumably a predictor) helps in predicting the variable **B** (presumably the target). Provide a justification for the hypothesis based on the previous analysis. Perform relevant statistical tests or modeling techniques to validate the hypothesis.

Explain your reasoning and present your findings.

**Note :** Remember to use libraries such as matplotlib or seaborn to create visually appealing and informative plots to support your analysis and conclusions. Provide clear instructions on the expected input, output, and any assumptions or limitations.

**Question # 02 : Time Series Analysis - Weightage 20 %**

**Time Required : 90 mins**

**Part i: Time Series Decomposition:**

Time series decomposition involves separating a time series into its underlying components, such as trend, seasonality, and residuals. Your task is to decompose the time series and to answer the following questions:

1. Write a code snippet in Python using any preferred time series library to decompose the given time series and plot each component separately to visualize their individual patterns and characteristics.
2. Discuss the significance and interpretation of each component (trend, seasonality, and residuals) in the context of the given time series.
3. **Challenge**: Implement a custom function for time series decomposition without using any built-in library or function specifically designed for decomposition. You may use other libraries for basic mathematical operations or data manipulation.

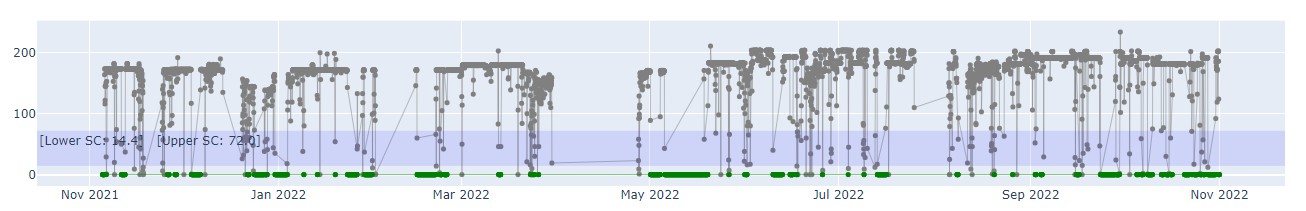
**Part ii: Time Series Annotation :**

Find the samples in time series that satisfies the following conditions:

And label these detected data points as 'U' and all other points as 'D'.

You are required to generate a labeled time series plot where each feature vector is plotted and have colored samples w.r.t to ‘U’ and ‘D’.

**Hint** : Sample plot of one of the features is given below.



**Question No 03: Implementation of a Deep Learning model - Weightage 50 %**

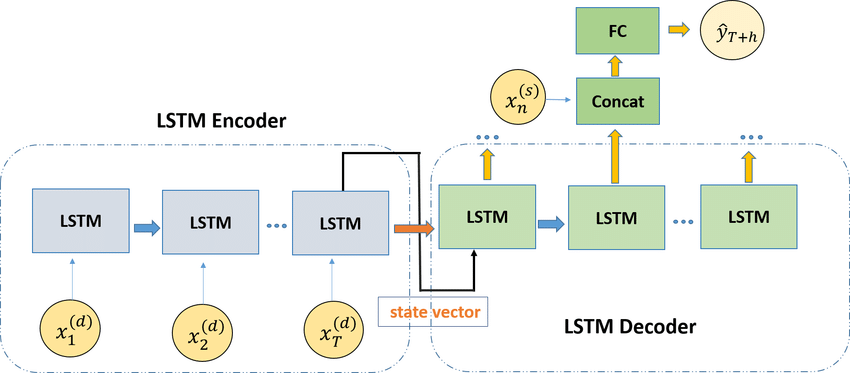
**Time Required : 90 mins**

**Task :** As an AI Researcher, your task is to implement a neural network architecture for time series forecasting. The architecture should include an encoder and decoder component, with a state vector in between. Here are the guidelines for implementing the architecture:

**Encoder**: Learns and captures information from the input sequence, propagating it for further processing.

**State Vector**: Stores sequential information of the input sequence, aiding the decoder in accurate predictions.

**Decoder**: Generates predictions by decoding the encoded information from the encoder in a sequential format.

* 

Consider the following points during implementation:

* Configure the dimensions of the input and output layers to enable accurate multi-horizon forecasting.
* **Challenge**: Make the input window size flexible instead of dynamic. This means the model should be able to handle input windows of any size during the inference stage for forecasting.
* Train the model using the training dataset, validate its performance using a validation dataset, and evaluate the predictions using a separate test dataset.
* Ensure that your implementation is compatible with the chosen framework (Keras, PyTorch, or TensorFlow) and follows best practices for neural network architectures. Document any assumptions made regarding specific layer types or parameter settings.