

## Hack to Hire 2024

### Use Case For Data Scientist:

**Problem Statement:** Develop a state-of-the-art question-answering model leveraging the Quora Question Answer Dataset. The objective is to create an AI system capable of understanding and generating accurate responses to a variety of user queries, mimicking a human-like interaction.

**Data set:** <https://huggingface.co/datasets/toughdata/quora-question-answer-dataset>

### **Important Notes:**

- You can use your Google Colab, or other services for the case study.
- Prepare a report and presentation to showcase your skills to the judges.

### **To Do:**

#### **1. Data Exploration, Cleaning, and Preprocessing:**

- Analyze the structure and content of the dataset. Remove any irrelevant information.
- Apply techniques such as tokenization, stop word removal, and stemming/lemmatization.

#### **2. Model Selection and Evaluation:**

- Test various NLP models such as BERT, T5, and GPT.
- Use metrics like ROUGE, BLEU, and F1-score to assess model performance.

#### **3. Visualization:**

- Create visualizations to show data distribution, feature importance, and model performance.
- Use Matplotlib, Seaborn, Plotly for creating charts and graphs.

#### **4. Insights and Recommendations:**

- Extract meaningful insights from your analysis and model results.
- Suggest **novel improvements** based upon your findings. (Very Important)

#### **5. Presentation and Documentation/Report:**

- Include sections such as Introduction, Literature Survey, Methodology, Results, and Conclusion in your report. Incorporate slides, graphs, and charts to support your findings in the report. Add Github link of your code in the report. Ensure your code and findings are well-documented for reproducibility. Submit your report on the portal. Shortlisted candidates will be invited to the event.
- On the date of event, present your work using a PPT and show a live demo of your work.

- Be ready to answer questions from judges on fundamental concepts and about your approach and results.

### **How to share your solution?**

1. Upload your Code on GitHub Repository.
2. Make your Git repo Public and share the link in the Microsoft Form.
3. Summarize your work in the readme, mention the tech stack front end and backend used in the project and additional tools and library you may have used.