**BASIC REQUIREMENTS:**Have IT-related experience demonstrating EACH of the five competencies listed below:

* **Data Preparation** - Collects, wrangles, and prepares data for analysis from disparate sources, using scripting languages. Facilitates work with IT staff to develop and deploy automated and scalable processes.
* **Data Analysis and Reporting** - Performs applicable statistical analysis and modeling to pull meaningful insights. Produces data visualizations, summaries, and reports. Provides recommendations on analytical approaches, tools, and architectures.
* **Applied Data Science** - Enhances products, services, and systems with design and development of data science projects to support business needs and operations. Ensures long-term sustainability and continuity of projects through documentation and engineering best practices.
* **Problem Solving** - Uses sound judgment and creativity to apply technical knowledge, methodologies, and tools in order to recommend and/or develop new solutions for organizational problems. Able to collaborate with others to resolve challenges, come up with solutions, and make progress using empathy, situational awareness, and tact.
* **Communication**- Explains in-depth or technical concepts using oral, written, and visual mediums, and in ways that different types of audiences can understand. Presents findings, recommendations, and alternatives to help others make a decision or understand the value of organizational IT needs. Communicates relevant project considerations and updates such as risks, estimates, and dependencies, within a team and across stakeholder groups.

**Examples of specialized experience** include

* Experience that demonstrates the **ability to utilize data analysis and data science knowledge and methods** to solve business problems from concept to delivery
* **Collect, wrangle, and prepare data** using scripting languages
* **Apply appropriate analytical techniques** and tools to **produce data visualizations**; summaries, reports, effectively **document code for reproducibility**
* **Communicate technical information** including information about analytical methods and outcomes **in ways that different types of audiences can understand**.

**Assessment prompts**

1. Download, merge, and describe the dataset and its basic characteristics (e.g., shape, variable types, basic stats).
2. Choose several variables and create visualizations to show their distributions. Justify your variable selection.
3. Clean the dataset to handle any missing data and justify your decisions.
4. For building a model, would you rescale any data in this dataset? How and why or  why not?
5. Build a model to identify risk factors for diabetes. Explain your choice of model and  what it can predict. What metrics would you use to assess performance? For this  dataset, how would you know your model is adequate?
6. Using these data, what are some identifiable risk factors for diabetes? How do you  know? Explain as if you were reporting the results to a non-technical stakeholder.