

Análisis de datos

Los candidatos a este examen buscan demostrar conocimientos introductorios sobre cómo manipular, analizar y comunicar de manera responsable los resultados del análisis de datos.

Los candidatos deben tener al menos 150 horas de instrucción o experiencia práctica en manipulación, análisis, visualización y comunicación de datos. Deben estar familiarizados con los conceptos generales de datos, las leyes relacionadas con los datos y las prácticas de análisis responsables.

Para aprobar el examen, también se espera que el candidato tenga los siguientes conocimientos y habilidades previos:

- Habilidades de lectura de octavo grado
- Habilidades de pensamiento crítico y resolución de problemas.
- Habilidades de alfabetización digital, incluida la capacidad de investigar, crear contenido y resolver problemas al usar la tecnología
- Álgebra I

1. Conceptos básicos de datos

1.1 Definir el concepto de datos

1.2 Describir los tipos de variables de datos básicos

- Booleano, numérico, cadena

1.3 Describir las estructuras básicas utilizadas en el análisis de datos.

- Tablas, filas, columnas, listas

1.4 Describir categorías de datos

- Cualitativos, cuantitativos, estructurados, no estructurados, metadatos, big data

2. Manipulación de datos

2.1 Importar, almacenar y exportar datos

- Comprensión fundamental de ETL (extracción, transformación y carga) procesos, herramientas de manipulación de datos (SQL, R, Python, Microsoft Excel, incluidos aspectos de Power Query) y formatos de archivos de almacenamiento de datos comunes (archivos de datos delimitados, XML, JSON)

2.2 Datos limpios

- Propósito y prácticas comunes (manejo de NULL, caracteres especiales, recortar espacios, formatear inconsistentemente, eliminar duplicados, imputar datos, etc.); validando datos

2.3 Organizar datos

- Propósito y prácticas comunes (clasificar, filtrar, dividir, transponer, añadir, truncar, etc.)

2.4 Datos agregados

- Propósito y prácticas comunes (agrupar, unir/fusionar, resumir, pivotante, etc.)

3. Análisis de datos

3.1 Describir y diferenciar entre tipos de análisis de datos

- Análisis descriptivo, análisis diagnóstico, prueba de hipótesis, análisis predictivo, análisis prescriptivo.

3.2 Describir y diferenciar entre métricas de agregación y interpretación de datos.

- Búsqueda, filtrado, valores únicos, funciones agregadas como Sum, Max, Mínimo, recuento, promedio/media, moda, mediana, desarrollo estándar

3.3 Describir y diferenciar entre análisis de datos exploratorios métodos

- Identificar relaciones de datos, describir conceptos de exploración de datos (granularidad, etc.), describir conceptos de minería de datos (anomalías, análisis de correlación, patrones, valores atípicos, etc.)

3.4 Evaluar y explicar los resultados de los análisis de datos.

- Calcular tendencias, determinar valores esperados, interpretar resultados de modelos predictivos, valores p, pruebas t y análisis de regresión.

3.5 Definir y describir el papel de la inteligencia artificial en el análisis de datos

- Definir inteligencia artificial, aprendizaje automático y algoritmos; describir cómo La IA se utiliza en el análisis de datos; describir cómo se utilizan los algoritmos de aprendizaje automático en el análisis de datos (Nota: los algoritmos específicos están fuera de alcance)

4. Visualización y comunicación de datos

4.1 Datos del informe

- Mostrar información de manera efectiva en tablas y gráficos; explicar cuándo y por qué desagregar los datos

4.2 Crear visualizaciones a partir de datos

- Identificar prácticas de visualización de datos que minimicen el potencial de mala interpretación; identifique los tipos de visualización que representan la estructura de datos subyacente y las preguntas de análisis (incluyendo comparación, tiempo/tendencia, parte-todo, relación, distribución, gráficos de correlación, diagrama de caja y bigotes, gráfico de dispersión, diagrama de dispersión, gráfico de barras, diagrama de Sankey, histograma, gráfico circular, gráfico de columnas, etc.)

4.3 Derivar conclusiones de una visualización de datos

- Traducir una representación visual de datos en palabras; identificar diferencias entre reclamaciones basándose en un análisis y su representación gráfica

5. Prácticas analíticas responsables

5.1 Describir las leyes y mejores prácticas de privacidad de datos

- RGPD, FERPA, HIPAA, IRB, PCI, etc.

5.2 Describir las mejores prácticas para el manejo responsable de datos

- Métodos para manejar la PII, asegurar datos y proteger el anonimato dentro de pequeños conjuntos de datos; importancia de anonimizar los datos; compensaciones al equilibrar la interpretabilidad y la precisión; Deficiencias de hacer generalizaciones a nivel de población con datos de muestra limitados.

5.3 Dado un escenario, describa los tipos de sesgo que afectan la recopilación y Interpretación de datos

- Sesgo de confirmación, sesgo cognitivo humano, sesgo motivacional, sesgo de muestreo; Seleccionar visualizaciones/representaciones de datos para evitar sesgos.

