Retos ejes Análisis de datos, Programación y Ciberseguridad Hackathon 05 abril 2025

Competencias tecnológicas programadas

Las competencias tecnológicas para el eje análisis de datos son:

- Limpieza y preparación de datos: detección y tratamiento de datos faltantes, duplicados y outliers.
- Exploratory Data Analysis (EDA): generación de resúmenes estadísticos y gráficos para identificar patrones clave.
- Calidad de datos: identificación y corrección de inconsistencias en el dataset.
- Metodología CRISP-DM (básica): aplicación de las etapas de comprensión del negocio, comprensión de los datos y preparación de datos.
- Visualización de datos: uso de herramientas como Matplotlib, Seaborn o Plotly.
- Estadística descriptiva: análisis de medidas de tendencia central, dispersión y correlaciones.

Para el eje programación, las competencias tecnológicas son:

- Desarrollo Frontend: Uso de HTML, CSS, JavaScript y frameworks como React o Vue.js para construir una interfaz intuitiva.
- Desarrollo Backend: Creación de APIs con Python (Flask/Django) o Node.js para manejar la lógica del negocio.
- Manejo de Bases de Datos: Uso de PostgreSQL, MySQL o SQLite para almacenar información sobre estudiantes y recursos educativos.
- Consumo de APIs: Integración de APIs para acceso a datos educativos y estadísticas.
- Seguridad y autenticación: Implementación de autenticación y control de accesos básicos.

Para el eje ciberseguridad, las competencias tecnológicas son:

- Configuración de redes y rastreo: Análisis de tráfico y detección de posibles vulnerabilidades en redes.
- Mecanismos de autenticación: Propuestas de autenticación multifactor (MFA), manejo de credenciales y gestión de accesos.
- Gestión de la información y eventos de seguridad (SIEM): Diseño de estrategias de monitoreo y detección de amenazas.
- Uso de Kali Linux: Simulación de posibles ataques y evaluación de riesgos en arquitecturas hipotéticas.

Reto Análisis de datos: Análisis Exploratorio para la Priorización de Recursos Educativos

Contexto: El Ministerio de Educación y el Ministerio de Tecnologías de la Información buscan optimizar la asignación de recursos educativos en la Región 5 de Colombia. Para ello, han recopilado datos sobre los puntajes de pruebas estandarizadas y variables sociodemográficas de los estudiantes. Sin embargo, antes de implementar modelos de inteligencia artificial o sistemas en la nube, es fundamental realizar un análisis exploratorio profundo para detectar patrones en los datos y proporcionar una base sólida para la toma de decisiones.

El conjunto de datos que sirve de apoyo para este reto es “Resultados únicos Saber 11” alojado en Datos abiertos (Resultados únicos Saber 11 | Datos Abiertos Colombia), compartido por el Instituto Colombiano para la Evaluación de la Educación.

Reto: Realizar un análisis exploratorio de datos (EDA) y generar un informe detallado que:

- Evalúe la calidad del dataset, detectando y corrigiendo problemas como datos faltantes, valores atípicos y errores en la información.
- Identifique patrones y relaciones clave entre los puntajes académicos y variables sociodemográficas.

- Genere visualizaciones efectivas que ayuden a los tomadores de decisiones a comprender la distribución de los puntajes y las disparidades regionales.
- Proponga hipótesis y recomendaciones basadas en los hallazgos para la priorización de recursos educativos.

Especificaciones del reto:

- **Limpieza y preparación de datos:** Corrección de datos erróneos, normalización y manejo de valores faltantes.
- **EDA detallado:** Análisis estadístico descriptivo para identificar tendencias y relaciones.
- **Visualización de datos:** Creación de dashboards y gráficos explicativos.
- **Informe final:** Presentación estructurada de los hallazgos y recomendaciones.

Solución esperada:

- Un dataset limpio y preparado para futuras etapas de modelado.
- Un informe detallado con resúmenes estadísticos, gráficos y hallazgos clave.
- Visualizaciones que faciliten la interpretación de los datos.
- Recomendaciones basadas en los análisis para guiar la asignación de recursos educativos.
- Presentación final con un resumen de los resultados y conclusiones en un tiempo máximo de 5 minutos.

El reto será desarrollado en 2 sprints de 1.5 horas cada uno.

Backlog del Producto:

- **Épica 1:** Limpieza y Preparación de Datos
    - Cargar y explorar el dataset.
    - Identificar y corregir datos faltantes y valores atípicos.
    - Estandarizar formatos y normalizar variables relevantes.
- **Épica 2:** Análisis Exploratorio de Datos (EDA)
    - Generar estadísticas descriptivas de las variables clave.

    - Analizar correlaciones entre los puntajes y factores sociodemográficos.
    - Identificar diferencias regionales en los puntajes de las pruebas.
- **Épica 3**: Visualización de Datos
    - Crear gráficos de distribución de puntajes por región.
    - Generar mapas de calor y gráficos de dispersión para explorar correlaciones.
    - Construir dashboards con los hallazgos principales.
- **Épica 4:** Generación de Reportes y Presentación
    - Elaborar un informe con los hallazgos clave y conclusiones.
    - Redactar recomendaciones basadas en los datos analizados.
    - Preparar la presentación final con visualizaciones y hallazgos clave.

Distribución por Sprints:

- **Sprint 1 (1.5 horas):** Exploración y Limpieza de Datos
    - Cargar y explorar el dataset.
    - Identificar y corregir valores faltantes, duplicados y errores.
    - Generar primeras estadísticas descriptivas y análisis de calidad de datos.

- **Sprint 2 (1.5 horas):** Análisis y Presentación de Resultados
    - Realizar EDA y análisis de correlaciones.
    - Crear visualizaciones interactivas.
    - Redactar informe con hallazgos y recomendaciones.
    - Preparar la presentación final del análisis.

Reto Programación: Desarrollo de un Prototipo para la Gestión de Recursos Educativos

Contexto: El Ministerio de Educación y el Ministerio de Tecnologías de la Información tienen el objetivo de desarrollar una infraestructura digital para la gestión de recursos educativos en la Región 5. Actualmente, no existe una

plataforma que centralice la información sobre las necesidades de los estudiantes y la asignación de recursos. Para dar el primer paso en esta transformación digital, se requiere un prototipo funcional de una plataforma que permita registrar información básica de estudiantes y visualizar los recursos educativos disponibles en su zona.

Reto: Desarrollar un prototipo funcional (mockup) de una plataforma web que incluya:

1. Un sistema de registro de estudiantes y necesidades educativas, con datos básicos como nombre, edad, zona geográfica y nivel educativo.
2. Una interfaz de consulta de recursos educativos, donde se puedan visualizar estadísticas sobre los recursos asignados por región.
3. Una API backend que permita la comunicación entre la base de datos y la interfaz de usuario.
4. Una implementación básica de autenticación y control de accesos, asegurando que solo usuarios autorizados puedan modificar información.

Especificaciones

- **Frontend:** Interfaz web con formularios para el registro de estudiantes y visualización de datos.
- **Backend:** API REST para gestionar la información en la base de datos.
- **Base de datos:** Uso de SQL para almacenar información sobre estudiantes y recursos educativos.
- **Autenticación básica**: Implementación de control de accesos con usuarios y contraseñas.
- **Visualización de datos:** Uso de gráficos para mostrar la distribución de recursos por región.

Solución esperada:

- Una plataforma funcional con una interfaz amigable para registrar y visualizar información.

- API REST que permita la gestión de estudiantes y recursos educativos.
- Autenticación básica implementada para el control de accesos.
- Visualización de estadísticas mediante gráficos y reportes automatizados.
- Presentación final explicando la arquitectura, funcionalidades y beneficios del prototipo.

El reto será desarrollado en 2 sprints de 1.5 horas cada uno.

Backlog del Producto:

- **Épica 1:** Desarrollo de la Interfaz de Usuario (Frontend)
    - Diseñar la estructura de la interfaz con HTML, CSS y JavaScript.
    - Crear formularios para el registro de estudiantes.
    - Implementar una sección para visualizar estadísticas y recursos educativos.
- **Épica 2:** Desarrollo del Backend y API REST
    - Crear una API REST con Flask/Django o Node.js.
    - Implementar endpoints para manejar el registro de estudiantes y la consulta de datos.
    - Conectar la API con la base de datos.
- **Épica 3**: Gestión de Bases de Datos
    - Diseñar la estructura de la base de datos en PostgreSQL o MySQL.
    - Crear tablas para almacenar información de estudiantes y recursos.
    - Implementar consultas SQL para recuperar datos de manera eficiente.
- **Épica 4**: Seguridad y Autenticación
    - Implementar autenticación básica con usuarios y contraseñas.
    - Configurar permisos para restringir la modificación de datos.
    - Aplicar medidas de seguridad básicas en la API.
- **Épica 5:** Visualización y Presentación de Resultados
    - Generar gráficos interactivos con datos de los recursos educativos asignados.
    - Crear reportes con estadísticas clave sobre la asignación de recursos.
    - Preparar la presentación del proyecto con una demostración funcional.

Distribución por Sprints:

- **Sprint 1 (1.5 horas):** Desarrollo Inicial y Conexión con la Base de Datos
  - Creación de la estructura de la base de datos y conexión con la API.
  - Desarrollo de endpoints básicos para gestionar estudiantes y recursos.
  - Implementación de la interfaz inicial con formularios básicos.

- **Sprint 2 (1.5 horas):** Seguridad, Visualización y Presentación
  - Implementación de autenticación y control de accesos.
  - Creación de gráficos y reportes con estadísticas clave.
  - Ensamble final del prototipo y preparación de la presentación.
  - Generación del informe y preparación de la presentación final.

Reto Ciberseguridad: Diseño de una Estrategia de Seguridad para una Infraestructura de Datos Educativos

Contexto: El Ministerio de Educación y el Ministerio de Tecnologías de la Información tienen la intención de implementar una infraestructura digital segura para gestionar los recursos educativos en la Región 5. Actualmente, no existe una infraestructura establecida, por lo que es necesario diseñar un plan de seguridad integral que contemple los principales riesgos cibernéticos, asegurando que la futura infraestructura esté protegida contra ataques y accesos no autorizados.

Reto: Desarrollar una estrategia de seguridad para la futura infraestructura de datos educativos que:

1. Identifique los principales riesgos de ciberseguridad en una infraestructura de datos educativos en la nube.
2. Diseñe mecanismos de autenticación y control de acceso para garantizar la seguridad de los datos.
3. Proponga un sistema de monitoreo de eventos de seguridad (SIEM) para la detección y mitigación de amenazas en la red.

4. Evalúe potenciales vulnerabilidades mediante simulaciones en un entorno controlado usando Kali Linux.

Especificaciones

- **Identificación de riesgos y vulnerabilidades:** Análisis de posibles amenazas en una infraestructura de datos educativos en la nube.
- **Propuesta de autenticación y gestión de accesos:** Diseño de políticas de seguridad, MFA y control de credenciales.
- **Diseño de monitoreo de seguridad:** Estrategia para la detección y respuesta a incidentes cibernéticos.
- **Evaluación de riesgos mediante simulaciones:** Simulación de posibles ataques en una arquitectura hipotética utilizando herramientas de Kali Linux.

Solución esperada:

- Un plan estratégico de ciberseguridad para la futura infraestructura.
- Propuestas detalladas de autenticación segura y gestión de accesos.
- Recomendaciones sobre herramientas de monitoreo y respuesta a incidentes.
- Simulación de posibles ataques y análisis de impacto.
- Presentación de resultados con una exposición clara del plan de seguridad.

El reto será desarrollado en 2 sprints de 1.5 horas cada uno.

Backlog del Producto:

- **Épica 1:** Análisis de Riesgos y Amenazas
    - Identificar amenazas cibernéticas comunes en infraestructuras de datos educativos.
    - Analizar riesgos en una arquitectura de nube para almacenamiento y gestión de datos.
    - Definir escenarios de ataques cibernéticos y su impacto.
- **Épica 2:** Diseño de Estrategias de Seguridad

    - Diseñar un sistema de autenticación y gestión de accesos (MFA, credenciales seguras).
    - Proponer un esquema de permisos y control de identidades.
    - Evaluar opciones de cifrado para la protección de datos.
- **Épica 3**: Estrategia de Monitoreo y Respuesta a Incidentes
    - Definir estrategias de monitoreo de eventos de seguridad (SIEM).
    - Establecer protocolos de respuesta ante incidentes cibernéticos.
    - Proponer herramientas para detección y mitigación de amenazas.
- **Épica 4**: Simulación y Evaluación de Vulnerabilidades
    - Usar Kali Linux para simular ataques en una infraestructura de prueba.
    - Evaluar el impacto de ataques comunes como fuerza bruta y phishing.
    - Documentar hallazgos y sugerencias de mitigación.
- **Épica 5:** Presentación del Plan de Seguridad
    - Elaborar un informe con los hallazgos y estrategias de seguridad.
    - Crear un resumen ejecutivo con recomendaciones clave.
    - Preparar la presentación con los resultados obtenidos.

Distribución por Sprints:

- **Sprint 1 (1.5 horas):** Análisis de Riesgos y Diseño de Seguridad
    - Identificación de amenazas y vulnerabilidades en una infraestructura educativa en la nube.
    - Diseño de estrategias de autenticación y control de accesos.
    - Definición de protocolos de cifrado y gestión de credenciales.

- **Sprint 2 (1.5 horas):** Simulación, Evaluación y Presentación
    - Simulación de vulnerabilidades en una infraestructura de prueba usando Kali Linux.
    - Evaluación del impacto de ataques potenciales y medidas de mitigación.
    - Generación del informe y preparación de la presentación final.

Criterios de evaluación

Los criterios de evaluación para cada reto son:

Reto Análisis de datos

La siguiente tabla presenta los criterios para el eje análisis de datos y su ponderación.

Criterio de Evaluación	Descripción	Ponderación
Limpieza y calidad de los datos	Se evalúa la capacidad del equipo para corregir datos faltantes, valores atípicos y problemas de formato.	10%
Exploración y análisis estadístico	Se analiza la profundidad del EDA y la identificación de patrones en los datos.	15%
Visualización de datos	Se valora la claridad, precisión e impacto de los gráficos y dashboards generados.	10%
Interpretación y generación de hipótesis	Se evalúa la capacidad del equipo para formular hipótesis y conclusiones basadas en los datos.	15%
Relevancia de las recomendaciones	Se valora si las propuestas son aplicables y sustentadas por el análisis realizado.	20%
Presentación de la solución	Se evalúa capacidad del equipo para explicar sus hallazgos y conclusiones en un tiempo máximo de 5 minutos.	30%

Reto Programación

La siguiente tabla presenta los criterios para el eje programación y su ponderación.

Criterio de Evaluación	Descripción	Ponderación
Funcionalidad de la plataforma	Se evalúa si la plataforma permite registrar y consultar datos de estudiantes y recursos educativos.	20%
Calidad del Backend y API REST	Se mide la correcta implementación de la API y su capacidad para gestionar datos.	15%

Criterio de Evaluación	Descripción	Ponderación
Estructura y eficiencia de la base de datos	Se analiza la organización de los datos y la optimización de consultas.	15%
Autenticación y seguridad	Se valora la implementación de medidas de seguridad básicas para el control de accesos.	10%
Visualización de datos y reportes	Se evalúa la calidad y claridad de los gráficos y estadísticas generados.	10%
Presentación de la solución	Capacidad del equipo para explicar la arquitectura y funcionalidades en un tiempo máximo de 5 minutos.	30%

Reto Ciberseguridad

La siguiente tabla presenta los criterios para el eje ciberseguridad y su ponderación.

Criterio de Evaluación	Descripción	Ponderación
Identificación de riesgos y amenazas	Se evalúa la capacidad del equipo para analizar vulnerabilidades y definir riesgos en la futura infraestructura.	15%
Propuesta de autenticación y control de accesos	Se mide la calidad de las estrategias diseñadas para la protección de credenciales y gestión de permisos.	15%
Estrategia de monitoreo y respuesta a incidentes	Se analiza la viabilidad de los protocolos de detección y mitigación de amenazas.	15%
Evaluación de vulnerabilidades con Kali Linux	Se valora la aplicación de herramientas de ciberseguridad para simular ataques y medir su impacto.	15%
Informe y recomendaciones estratégicas	Se evalúa la claridad y aplicabilidad del plan de seguridad propuesto.	10%
Presentación de la solución	Se evalúa capacidad del equipo para presentar su propuesta en un tiempo máximo de 5 minutos.	30%