

Desafío Profesional Data Analyst

Te damos la bienvenida

El Desafío Profesional se plantea como una forma integral de evaluación dentro de la Certificación en Data Analyst. Es una parte crucial de la formación en Digital House y un requisito indispensable para obtener la certificación.

Este documento tiene como objetivo presentar algunos pilares fundamentales: el desafío propuesto, lo que esperamos de un analista de datos como tú, y las formas de evaluación que implementaremos para confirmar que has completado con éxito la especialización. ¿Listos?



Índice

- 01 ¿Qué es el desafío profesional?
- 02 ¿Qué buscamos?
- 03 Etapas
- 04 Evaluación



01

¿Qué es el desafío profesional?

¿Qué es el desafío profesional?

El Desafío Profesional es una aplicación práctica de los conceptos teóricos estudiados hasta el momento y tiene como objetivo el desarrollo de un proyecto integral de análisis de datos. Funciona como una estrategia didáctica que consiste en la realización de un conjunto de actividades interrelacionadas con un inicio, un desarrollo y un final.

La metodología que utilizamos en Digital House de “aprender haciendo” favorece la articulación de conocimientos teóricos y prácticos, además de acercar a los estudiantes a la realidad del mercado profesional en el campo del Data Analyst.





Beneficios

01

Permite aplicar las herramientas aprendidas en los módulos realizados.

02

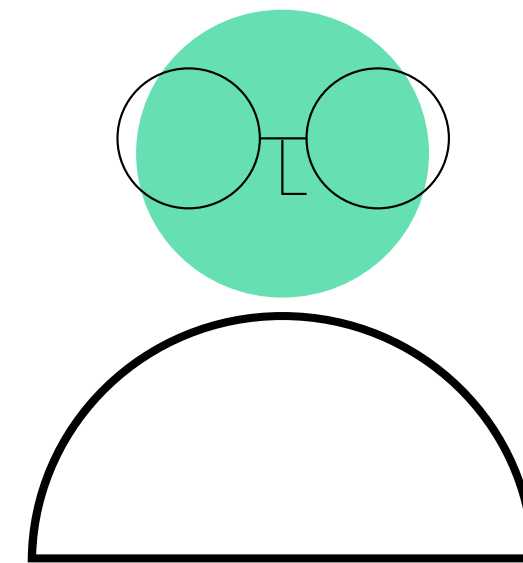
Nos desafía a llevar las ideas a la práctica.

¿Cómo funciona?

En esta Certificación, el desafío está destinado a la creación de un proyecto integral de análisis y visualización de datos.

Te recomendamos que lo **comiences después de haber completado el resto de los cursos** de la certificación, ya que este desafío involucra gran parte de todos los conocimientos adquiridos en los módulos anteriores, con un enfoque especial en el análisis de datos, limpieza, visualización y presentación de insights.

El desafío se divide en **tres etapas principales**, cada una con un conjunto de entregables específicos. Todos los requerimientos deberán ser entregados al final de este curso a través del envío de tu repositorio y la presentación del proyecto.

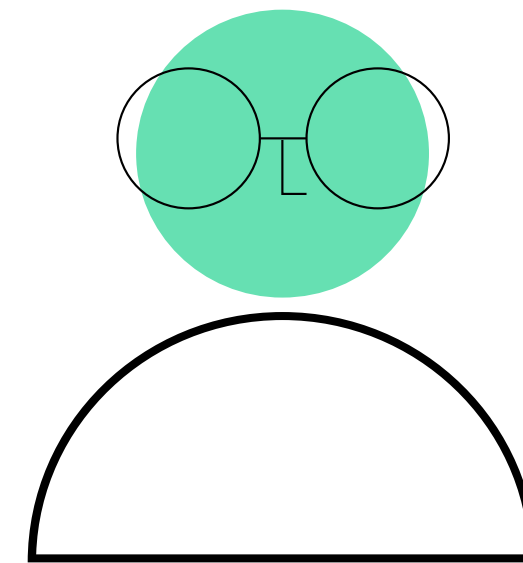


¿De qué se trata el desafío?

En términos generales, el objetivo del proyecto es realizar un análisis integral de un conjunto de datos relacionados con un caso de negocio específico, como parte del Desafío Profesional en Data Analyst.

Este proyecto será un mínimo producto viable para la entrega final, donde el estudiante deberá aplicar **técnicas de exploración, limpieza y visualización de datos** para **extraer insights valiosos**. Los participantes trabajarán con datos reales para realizar tareas como la identificación de patrones, la creación de dashboards interactivos, y la formulación de recomendaciones basadas en los análisis realizados.

El estudiante deberá manejar un conjunto de datos clave, con los que se realizará un seguimiento exhaustivo de las operaciones analíticas, asegurando un registro detallado de todas las transformaciones y análisis (también conocidos como pasos de procesamiento o actividad analítica).



02

¿Qué buscamos?

¿Qué buscamos?



- **Carga y exploración de datos:** Importación y exploración inicial de los datasets proporcionados.
- **Limpieza y transformación de datos:** Procesos de limpieza de datos para asegurar su integridad y consistencia, incluyendo la gestión de valores faltantes y la normalización de datos.
- **Análisis exploratorio de datos (EDA):** Realización de un análisis exploratorio para identificar patrones, tendencias y anomalías en los datos.

¿Qué buscamos?



- **Visualización de datos:** Creación de gráficos y visualizaciones que permitan una interpretación clara de los datos, utilizando herramientas como Matplotlib, Seaborn y Power BI.
- **Construcción de dashboards:** Desarrollo de dashboards interactivos para presentar los insights obtenidos de manera clara y efectiva.
- **Presentación de insights:** Formulación de conclusiones y recomendaciones basadas en el análisis de datos, y su presentación a través de un storytelling efectivo.

03

Etapas

Etapas 1

Exploración Visual de los Datos

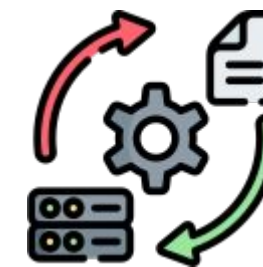
En esta primera etapa, los estudiantes aplicarán los conocimientos adquiridos en los módulos de **Introducción a Python**, **Programación en Python**, **Matemática para IA** y **Visualización de Datos con Python**. El objetivo es realizar una exploración visual de los datos utilizando técnicas de análisis exploratorio de datos (EDA). Los estudiantes crearán visualizaciones iniciales que les permitan identificar patrones, tendencias, y posibles anomalías en el dataset.



Etapa 2

Limpieza y Transformación de Datos

En la segunda etapa, los estudiantes pondrán en práctica los conocimientos de los módulos de **Bases de Datos en MySQL** y el módulo de **Exploración y Limpieza de Datos**. Aquí, el enfoque estará en la limpieza y transformación de los datos para asegurar que estén en condiciones óptimas para el análisis posterior. Se espera que los estudiantes manejen valores faltantes, normalicen variables, y apliquen técnicas de integración de datos.



Etapa 3

Creación de Dashboard y Presentación

La última etapa se centra en los módulos de **Data Analytics con Power BI** y **Visualización y Dashboard Avanzados**. Los estudiantes utilizarán Power BI para crear dashboards interactivos que presenten los insights obtenidos de manera clara y efectiva. Además, deberán preparar una presentación final que resuma el trabajo realizado, incluyendo un video explicativo de 5 minutos.



04

Evaluación



La evaluación del Desafío Profesional tiene como objetivo asegurar que los estudiantes apliquen de manera efectiva los conocimientos adquiridos a lo largo del curso en un proyecto realista y desafiante. Se considerarán diversos aspectos clave para medir tanto la profundidad del análisis como la calidad de las soluciones propuestas.

Profundidad del Análisis Exploratorio de Datos

- **Evaluación:** Se evaluará si el análisis exploratorio cubre todas las variables importantes.
- **Criterios:** Identificación de patrones, tendencias y posibles anomalías en los datos.
- **Técnicas:** Verificación del uso adecuado de técnicas estadísticas (medidas de tendencia central, dispersión, correlaciones).
- **Visualizaciones:** Implementación de visualizaciones como histogramas, boxplots y matrices de dispersión.

Calidad y Limpieza de los Datos Procesados

- **Evaluación:** Se evaluará la detección y tratamiento de datos faltantes y outliers.
- **Transformaciones:** Verificación de la correcta aplicación de transformaciones, normalización o escalado de datos.
- **Documentación:** Asegurar que el proceso de limpieza esté bien documentado y sea reproducible.

Uso de Herramientas

- **Python y Librerías:** Evaluación del uso eficiente de Python, Pandas, NumPy para la manipulación de datos.
- **Visualización:** Uso de Matplotlib y Seaborn para la creación de visualizaciones.
- **Power BI:** Verificación de la correcta conexión a fuentes de datos y la creación de dashboards interactivos.
- **Buenas Prácticas:** Revisión de la legibilidad, organización y buenas prácticas en la programación del código.

Calidad y Claridad de los Dashboards

- **Diseño:** Evaluación de la claridad y estética del diseño del dashboard.
- **Navegación:** Disposición de los elementos y facilidad de navegación.
- **Interactividad:** Valoración de la implementación de filtros, segmentaciones y otras funcionalidades interactivas.
- **Claridad:** Verificación de que la información presentada sea clara, concisa y comprensible para el público general.

Coherencia y Justificación de las Conclusiones

- **Fundamentación:** Evaluación de si las conclusiones y recomendaciones están basadas en el análisis de datos.
- **Storytelling:** Comprobación de la capacidad del estudiante para contar una historia con los datos.
- **Justificación:** Verificación de que todas las decisiones durante el análisis y visualización de datos estén bien justificadas y documentadas.

Performance y Escalabilidad

- **Mantenimiento:** Evaluación de que la solución propuesta sea fácil de mantener y actualizar.
- **Documentación y Código:** Revisión de la documentación y código para soportar futuras modificaciones.
- **Rendimiento:** Comprobación de que el análisis y las visualizaciones soporten cambios y actualizaciones en los datos sin degradar el rendimiento.

¡Muchas gracias!