

INFORME DE CIERRE: DIPLOMADO EN CIENCIA DE DATOS

INFORME DE CIERRE

1. CONTEXTO

La Institución Universitaria Mayor de Cartagena se compromete con una gestión integrada y coordinada entre la Universidad, el sector empresarial y el Estado. A través de su Dirección de Extensión Académica y la modalidad de Proyectos Especiales, la institución trabaja de manera conjunta con diversos actores sociales e institucionales, con el objetivo de promover la transferencia de conocimiento y potenciar la proyección social. De esta manera, este enfoque contribuye al desarrollo del capital social y al crecimiento humano integral. Para materializar este propósito, la U Mayor establece convenios y celebra contratos que permiten el acompañamiento y la ejecución de programas, proyectos y actividades en las áreas académica, administrativa, logística y operativa, apoyando la implementación de planes estratégicos tanto en el sector público como privado.

Dicho lo anterior, el presente documento tiene como objeto presentar el informe final del tipo de evento denominado “Diplomado en ciencia de datos” asignado al programa de “Transformación digital y cibercultura-Módulo competencia digitales” como diplomado, con una duración de 80 horas, con un cupo de 30 personas.

2. CONTENIDO TEMÁTICO

El diplomado de Ciencia de datos tuvo como objetivo, fomentar la toma de decisiones a partir de los conceptos y herramientas que ofrece la ciencia de datos. De esta forma, su alcance radica en entender los principios esenciales que rigen el mundo de la ciencia de datos con una sólida base teórica y práctica, además de comprender las técnicas de análisis y modelados de datos, que permiten interpretar y transformar datos complejos en información significativa.

Para lograr este propósito, el contenido del diplomado fue estructurado de la siguiente manera:

Módulo I. Introducción a la Ciencia de Datos (12 horas)

- Definición y alcance de la ciencia de datos
- Aplicaciones en la vida cotidiana y para la gestión en entidades
- Fundamentos del ciclo de vida de un proyecto de ciencia de datos

Fechas programadas: 03, 05, 10 12 de septiembre del 2024, sesión virtual sincrónica.

Módulo II. Fundamentos de Programación en Python (20 horas)

- Introducción a Python y su entorno de desarrollo
- Variables, tipos de datos y estructuras de control
- Funciones y manejos de archivos en Python

Fechas programadas: 17, 19, 24 y 26 de septiembre del 2024 y 01 y 03 de octubre del 2024, sesión virtual sincrónica.

Módulo III. Manipulación de Datos con Pandas (20 horas)

- Introducción a la biblioteca Pandas
- Importancia y exploración de datos
- Manipulación de datos: filtrado, selección y transformación

Fechas programadas: 08, 10, 15, 17, 22 y 24 de octubre del 2024, sesión virtual sincrónica.

Módulo IV. Visualización de Datos con Matplotlib y Seaborn (12 horas)

- Introducción a la visualización de datos
- Creación de gráficos con Matplotlib
- Visualización avanzada con Seaborn

Fechas programadas: 29, 30 y 31 octubre del 2024, sesión virtual sincrónica.

Módulo V. Introducción a Estadística Descriptiva (8 horas)

- Conceptos básicos de estadística descriptiva
- Medidas de tendencia central y dispersión
- Visualización y análisis exploratorios de datos

Fechas programadas: 12, 14 y 19 de noviembre del 2024, sesión virtual sincrónica.

Módulo VI. Proyecto final (10 horas)

- Desarrollo de un proyecto de ciencia de datos
- Aplicación de conocimientos adquiridos

Fechas programadas: 21, 26 y 28 de noviembre del 2024, sesión virtual sincrónica.

El diplomado fue dictado en la forma planeada, vale decir, cumpliendo con el total del contenido y en las fechas programadas. La Umayor generó canales de comunicación que permitieron la entrega de información de forma expedita y en tiempo real, a fin de que los participantes conocieran el detalle de los aspectos técnicos, metodológicos y logísticos para el desarrollo de cada sesión de trabajo. El docente encargado de dictar el programa fue:

Nombre y apellidos	Cedula	Módulos asignados
JORGE MARIO GAVIRIA HINCAPIE	15507955	TODOS

La comunicación empleada entre el docente y los servidores públicos se trazó vía correo electrónico específico del programa (cienciadedatos@umayor.edu.co) y los correos registrados por el personal asistente. Mediante este correo se asignaron las sesiones virtuales, se envió los talleres y videos pertinentes al tema, se realizaron los formularios de asistencia, los formularios de evaluación de conocimientos previos, el formulario de conocimientos adquiridos y el formulario de nivel de conformidad.

2. CIERRE DE PROGRAMA

2.1 APROBACIÓN DEL PROGRAMA

El curso fue proyectado para máximo 30 servidores públicos, sin embargo, una vez abierta la inscripción, se registraron 35 personas, equivalente a una ocupación del 116.6%.

Para la expedición de la constancia de asistencia al diplomado, se exige que el participante haya asistido a un mínimo del 60% de la programación; partiendo de esto, se verificó la información correspondiente, determinando que 20 personas reunieron los requisitos necesarios, es decir, cumplieron con una asistencia al 60% de las sesiones, lo que es equivalente al 57.14% de los participantes. A continuación, se lista las personas que cumplieron los requisitos para la obtención de la constancia de participación en el diplomado Ciencia de datos:

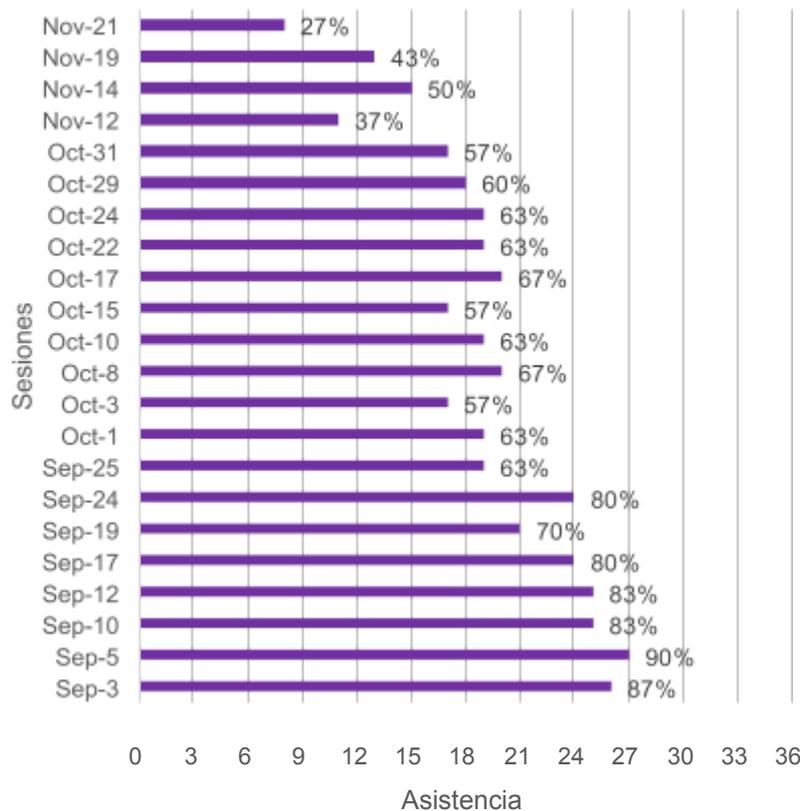
N°	NOMBRE	CÉDULA
1	ADOLFO PARDO MORENO	1047397342
2	ADRIANA RAMIREZ BENITEZ	53116709
3	ALEXANDER ROMERO DELGADO	8853194
4	ANGEL MARIA PEREZ SALGADO	73140508
5	ÁNGEL URIEL VILLEGAS DIAZ	71590193
6	CAMILO ROMÁN TORRES CATALÁN	73181999
7	CARLOS RAFAEL MARTÍNEZ CABARCAS	1102824931
8	DANIEL ALBERTO AMEZQUITA LOPEZ	73139969
9	DAVID MAURICIO ARNEDO GONZÁLEZ	3907730
10	DEYSIS ARAQUE MENCO	1047467537
11	EMILIO ALEJANDRO VALENCIA CASSERES	1047395449
12	FRANKLIN GABRIEL AMADOR HAWKINS	18001896
13	HARMENSON POLO OLIVO	84087526
14	INGRID PAOLA SOLANO BENITEZ	25274077
15	JORGE ELIECER CASTRO MENDOZA	92447945
16	MANUEL ANTONIO LÓPEZ YEPES	1067403736
17	MARCOS MAURICIO VERBEL PEÑARANDA	9104360
18	RICARDO JUAN PUELLO ESTRADA	73129651
19	ROSYANY RODRIGUEZ ACOSTA	45514258
20	YITHZAK DAVID FERNÁNDEZ GUTIÉRREZ	1143380791

NÚMERO DE INSCRITOS	NÚMERO DE PERSONAS CERTIFICADAS
35	23

2.2 ASISTENCIA DE LOS PARTICIPANTES

Es importante recordar que, tal como lo indica el artículo 2.6.6.8. del Decreto 1075 de 2015, por medio del cual se expide el Decreto Único Reglamentario del Sector Educación, la participación en este evento de educación informal brindó oportunidades para complementar, actualizar, perfeccionar, renovar o profundizar conocimientos, habilidades, técnicas y prácticas, pero no conduce a título alguno o certificado de aptitud ocupacional.

Como ya se mencionó, 20 de las 35 personas asistentes cumplieron con el 60% de las actividades propias del diplomado. Sin embargo, se tuvo una participación en promedio de 19 personas por sesión tal como se muestra en la siguiente gráfica:



2.2 DESARROLLO DEL PROGRAMA Y CONOCIMIENTOS

Módulo 1: Introducción a la Ciencia de Datos

El objetivo del módulo fue, introducir a los participantes en los fundamentos de la ciencia de datos, mediante la exploración de enfoques analíticos cuantitativos y cualitativos aplicados a datos relevantes. Esto incluye familiarizarlos con el entorno

estadístico y su implementación práctica, utilizando un lenguaje de programación, estableciendo así las bases para el análisis de datos eficiente y estructurado.

Como diagnóstico inicial, basado en el test de conocimientos previos, se logró establecer una línea base, la cual permitió reconocer que, los participantes no cuentan con conceptos básicos sobre la analítica de datos e inteligencia de negocios. En este sentido, se evidencio que solamente 3 de los participantes conocen el lenguaje de programación Python. A su vez, se identificó que estos no tienen claridad en cuanto a conceptos de análisis bivariado ni multivariado.

Según la evaluación de los conocimientos adquiridos durante el módulo, se observó que los participantes son capaces de distinguir una variable en un conjunto de datos, identificar la diferencia y naturaleza de las variables cuantitativas. Además, tienen claridad sobre las medidas de tendencia central y comprenden la importancia de la normalización de los datos para su análisis. Finalmente, los participantes son capaces de establecer, de manera básica, modelos de aprendizaje automático utilizados en la ciencia de datos.

Módulo 2: Fundamentos de Programación en Python.

El alcance del Módulo 2 se centró en, capacitar a los participantes en el uso de Python como herramienta clave para el análisis de datos, enfocándose en la comprensión de su interfaz, la estructura del lenguaje y las bibliotecas más relevantes. El módulo busco, adaptar estas capacidades a las necesidades específicas de los usuarios, optimizando su aplicación en diversos contextos analíticos y fortaleciendo su competencia en la programación orientada a datos

Durante la evaluación de conocimientos previos, realizada a través de un test de seis preguntas específicas sobre Python y completada por 20 participantes, se

identificó que solo tres funcionarios tenían experiencia o nociones básicas/avanzadas en el uso de este lenguaje de programación.

Al finalizar el módulo, los servidores públicos adquirieron los conocimientos básicos de programación, utilizando Python como una de las herramientas clave para el análisis de datos. Comprendieron la estructura de un script y el proceso de instalación de bibliotecas. Estas habilidades quedaron evidenciadas en un ejercicio individual, en el cual cada participante desarrolló una estructura condicional (if, else) tanto simples como anidadas, así como ciclos (bucles) for y while, implementados en funciones dentro del programa mencionado.

Módulo 3: Manipulación de Datos con Pandas.

El módulo 3 del programa, tuvo como objetivo desarrollar habilidades básicas en la manipulación y análisis de grandes conjuntos de datos utilizando la librería Pandas en Python, con el fin de realizar operaciones de limpieza, transformación, agregación y visualización de datos, para extraer información relevante y facilitar la toma de decisiones basadas en datos.

En línea con los conocimientos previos y los adquiridos durante el módulo, los participantes desarrollaron competencias para establecer distribuciones binomiales, cálculo de probabilidades de distribución normal y realizar cálculos de intervalos de confianza, utilizando herramientas de Excel. Estos conocimientos, quedaron demostrados mediante un ejercicio práctico, en el que los participantes calcularon y estructuraron una tabla de distribución de frecuencias en Excel.

Módulo 4: Visualización de Datos con Matplotlib y Seaborn.

El propósito del módulo se enfocó en, desarrollar habilidades fundamentales en Python orientadas al análisis y visualización de datos mediante el uso de bibliotecas especializadas como NumPy, Pandas, Matplotlib y Seaborn, fomentando la comprensión de estructuras básicas de datos como listas y diccionarios, para facilitar la manipulación, organización y representación gráfica de información en el ámbito laboral o académico.

Con base en los diagnósticos realizados en los módulos anteriores, se identificó que los participantes carecían de habilidades suficientes para abordar los temas propuestos en este módulo. Por ello, se implementó una metodología centrada en la realización de cinco ejercicios grupales, diseñados para explicar de manera detallada la sintaxis necesaria para manipular datos utilizando la biblioteca Pandas. Estos ejercicios se enfocaron en el manejo de estructuras clave como arreglos, diccionarios, tuplas, y series unidimensionales y bidimensionales, promoviendo una comprensión práctica y efectiva de estas herramientas.

Módulo 5: Introducción a Estadística Descriptiva.

El objetivo del módulo fue, proporcionar a los participantes una comprensión fundamental de la estadística descriptiva, abordando conceptos clave y técnicas esenciales para organizar, resumir e interpretar datos de manera efectiva, a fin de establecer una base sólida para el análisis cuantitativo en diversos contextos académicos y profesionales.

Al igual que en el módulo 4, se identificó que los participantes no contaban con conocimientos de nivel medio o avanzado en programación con Python. Esta limitación dificultó la resolución individual de los ejercicios y talleres propuestos en el módulo. Por ello, se adoptó una metodología basada en el desarrollo de los

talleres durante las sesiones de clase, donde se explicó de manera detallada la sintaxis necesaria para realizar análisis, resolver problemas y generar gráficos de los ejercicios planteados. Al finalizar el módulo, se asignó un ejercicio para que los participantes replicaran el código aprendido, consolidando así los conocimientos adquiridos.

Módulo 6: Proyecto final.

El propósito final del módulo fue consolidar las habilidades adquiridas en el manejo de Python y en los fundamentos del análisis de datos, desarrolladas a lo largo de los módulos previos del diplomado en Ciencia de Datos. Esto se logró mediante la aplicación integral de dichos conocimientos al diseño y ejecución de un proyecto final centrado en un objeto de estudio específico.

Como parte de la evaluación de conocimientos previos y los temas impartidos durante el programa, se evidenció que, en el ámbito conceptual, todos los participantes incorporaron adecuadamente terminología clave de estadística descriptiva, comprendiendo conceptos fundamentales como probabilidad, frecuencia y desviación estándar, entre otros.

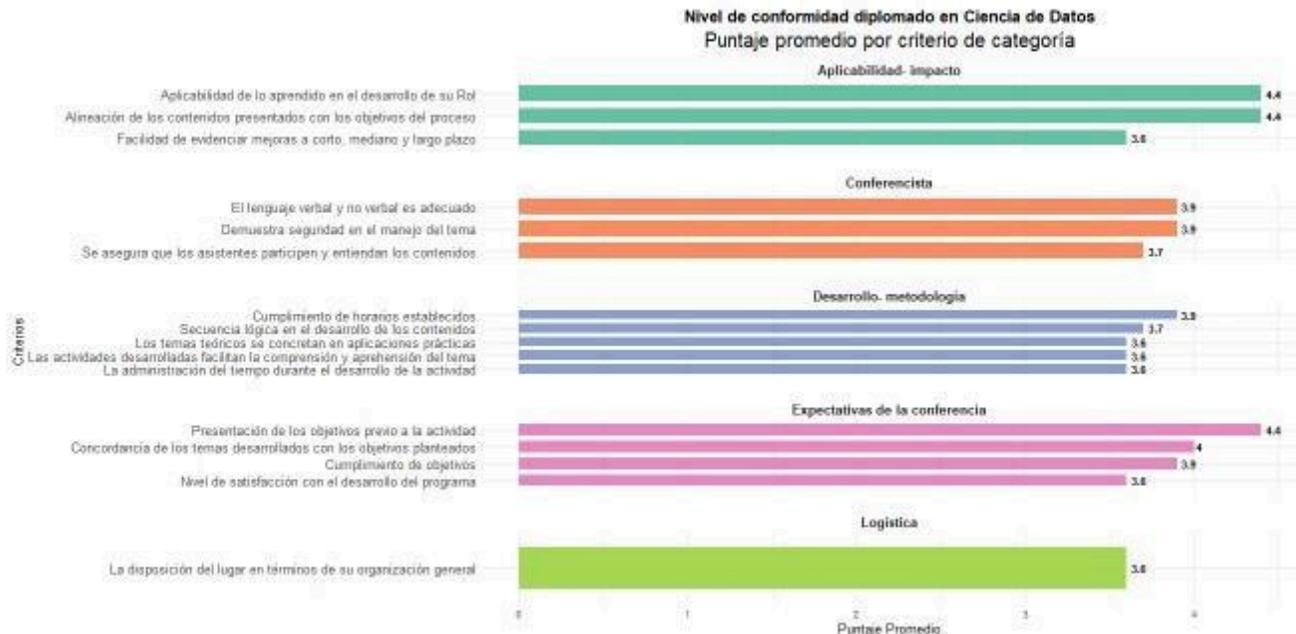
En cuanto al lenguaje de programación en Python, los participantes demostraron capacidad para interpretar y responder a códigos específicos, además de describir de manera estructurada la metodología necesaria para abordar un proyecto de analítica de datos.

En el desarrollo de sus proyectos finales, los participantes lograron realizar un análisis de datos detallado, aplicando técnicas de visualización como histogramas de frecuencia, gráficos de tendencia, diagramas de cajas y bigotes, gráficos de barras y de torta. Asimismo, integraron un análisis de regresión logística,

demostrando competencias en técnicas de análisis multivariado y reforzando su capacidad para abordar problemas complejos en ciencia de datos.

3. NIVEL DE CONFORMIDAD FRENTE AL CURSO

En este apartado se evaluó el nivel de conformidad de los servidores público frente al diplomado Ciencia de datos, para esto, se abordó el tema desde 5 apartados evaluando cada uno de estos en una escala de 0 a 5 siendo 5 la mayor calificación de satisfacción. A continuación, se muestran los resultados:



Nota: La encuesta de conformidad fue diligenciada por 8 servidores públicos.

Finalmente, y acorde a la encuesta de nivel de conformidad frente al diplomado, se apreció los siguientes comentarios:

Aspectos positivos destacados: La actividad fue valorada principalmente por su relevancia práctica, claridad en los temas y capacidad de transmitir estrategias de enseñanza a pesar de las falencias formativas. Los participantes reconocieron el

enfoque participativo e inclusivo, así como la capacidad del diplomado en extrapolar y aplicar los temas impartidos.

Como aspectos a mejorar:

Áreas de mejora: Logística: generar un día extra para atender dudas individuales.

Metodología: Incrementar la interacción presencial y considerar un ajuste en la intensidad horaria para profundizar en los temas clave.

Recomendaciones finales:

1. Fortalecer los recursos tecnológicos y metodológicos para facilitar la participación efectiva en modalidades híbridas.
2. Evaluar la posibilidad de aumentar las horas de clase en los módulos más complejos.
3. Promover un plan estándar para la entrega de materiales complementarios que refuercen el aprendizaje.

Estas sugerencias, junto con el reconocimiento de los logros alcanzados, contribuirán a la mejora continua de actividades formativas futuras.