
PROGRAMA DE ESTUDIOS: CURSO DE MANIPULACIÓN Y VISUALIZACIÓN DE DATOS

KOICA

HGU HANDONG GLOBAL
UNIVERSITY



UNA

Manipulación y Visualización de Datos

Información Básica

Información del curso

1. Título: **Manipulación y Visualización de Datos**
2. Año lectivo: **2024**
3. Semestre: **Primero**
4. Departamento: **Centro de Innovación TIC (FP-UNA)**
5. Año/Nivel: **Intermedio**
6. Formato de clase: **Charlas y sesiones prácticas**

Hora y lugar

1. Días: **Miércoles y Jueves**
2. Hora: **19:00 a 21:30**
3. Ubicación: **Online**

Información del instructor

1. Nombre: **Alexis González**
2. Oficina: **Campus Virtual**
3. Contacto: nagb1992@gmail.com
4. Perfil profesional:



Experiencia gestionando proyectos que generen valor al negocio, con una mirada crítica y analítica desde el nacimiento de la idea hasta implementación, demostrando el aporte a través de los datos. Experiencia en definición e implementación de estrategias comerciales y digitales, enfocando todas las acciones en mejorar la experiencia del cliente ofreciendo una oferta de valor. Como ingeniero de datos tengo experiencia preparando datos y generando valor con los nuevos conocimientos encontrados. Desde mi perfil de líder en ciencia de datos soy responsable de la planeación del roadmap de ejecución de los proyectos; Impactar con las soluciones el margen de los negocios de los clientes. Servir de mentoring, coaching y desarrollar las carreras de los científicos de datos.

Habilidades de liderazgo, pensamiento estratégico, adaptabilidad al cambio, negociación, toma de decisión y aportando al desarrollo del equipo.

Notas adicionales

1. Se anima a los estudiantes a comunicarse con el instructor a través de los datos de contacto proporcionados para cualquier consulta o inquietud relacionada con el curso.
2. Es importante asistir tanto a las sesiones teóricas como a las prácticas para beneficiarse plenamente de la estructura del curso.

Prerrequisito

- Programación en Python.
- Notebook con Sistema Operativo Windows (O Linux, pero el alumno debe tener conocimientos en Linux para avanzar).
- Espacio en disco duro de al menos 20GB.
- Memoria Ram de 8GB (Recomendable 16GB).

Descripción del Curso

En el panorama moderno donde los datos actúan como el eje de la toma de decisiones en todas las industrias, dominar el arte de la manipulación y visualización de datos es crucial. Este curso, "Visualización y manipulación de datos", está meticulosamente elaborado para capacitar a los estudiantes con las habilidades prácticas y el conocimiento teórico necesarios para transformar datos sin procesar en narrativas visuales reveladoras. A través de una experiencia de aprendizaje inmersiva con Python como núcleo, los participantes participarán en el proceso de extremo a extremo de limpieza, manipulación y creación de visualizaciones convincentes de datos. Al aprovechar las poderosas bibliotecas de Python, como pandas, Matplotlib, Seaborn y Plotly, los estudiantes navegarán a través de escenarios de datos del mundo real, preparándolos para los desafíos de la industria de la ciencia de datos.

El plan de estudios está diseñado con un fuerte énfasis en el aprendizaje práctico, asegurando que cada concepto teórico esté basado en una aplicación práctica. Desde el principio, los estudiantes trabajarán en ejercicios y proyectos estructurados que reflejen problemas de datos genuinos que se encuentran en entornos profesionales. Este enfoque no sólo facilita una comprensión más profunda del material, sino que también ayuda al desarrollo de habilidades de resolución de problemas fundamentales para los profesionales de datos.

A medida que avanza el curso, los participantes profundizarán en técnicas avanzadas de manipulación de datos y explorarán un amplio espectro de herramientas de visualización, preparándolos para comunicar de manera efectiva conocimientos de datos complejos.

Una característica distintiva de este curso es su componente de aprendizaje basado en proyectos, que culmina en un proyecto final donde los estudiantes aplican sus habilidades recién adquiridas para abordar un proyecto integral de visualización y análisis de datos. Esto no sólo solidifica su aprendizaje, sino que también les permite crear una cartera sólida, mostrando sus capacidades a posibles empleadores. El curso está estructurado para apoyar a estudiantes de distintos niveles de habilidades, ofreciendo orientación personalizada por parte de instructores experimentados y fomentando un entorno de aprendizaje colaborativo que fomenta la interacción entre pares y el trabajo en equipo.

Dirigido a aspirantes de científicos de datos, profesionales que desean adquirir habilidades en el manejo y visualización de datos y estudiantes de campos relacionados que buscan aumentar sus habilidades analíticas, este curso ofrece un camino para dominar la visualización y manipulación de datos. Al finalizar, los participantes adquirirán una profunda competencia en el uso de Python para la ciencia de datos, estarán listos para abordar desafíos de datos complejos y elevar su carrera en el floreciente campo del análisis de datos. Este curso representa no solo un viaje educativo, sino una experiencia transformadora, que equipa a los alumnos con las habilidades para convertir los datos en

decisiones.

Características clave:

- **Enfoque práctico:** desde la primera sesión, los estudiantes se sumergirán en ejercicios y proyectos de codificación que imitan los desafíos de datos del mundo real. Este enfoque práctico garantiza que el conocimiento teórico se refuerce mediante la aplicación práctica.
- **Cobertura integral:** el plan de estudios cubre cada paso del proceso de recopilación de datos, desde la inspección inicial de datos hasta técnicas de limpieza avanzadas, e introduce a los estudiantes una amplia gama de métodos de visualización, desde gráficos básicos hasta mapas y paneles interactivos.
- **Herramientas relevantes para la industria:** los estudiantes dominarán el uso de Python y sus poderosas bibliotecas, como Pandas para manipulación de datos, Matplotlib y Seaborn para visualizaciones estáticas y Plotly para gráficos interactivos. Estas herramientas se utilizan ampliamente en la industria y preparan a los estudiantes para roles de datos en diversos sectores.
- **Aprendizaje basado en proyectos:** el curso culmina con un proyecto final que desafía a los estudiantes a aplicar sus habilidades en un conjunto de datos del mundo real. Este modelo de aprendizaje basado en proyectos no sólo mejora la comprensión, sino que también ayuda a crear un portafolio que muestre sus capacidades a posibles empleadores.
- **Instrucción y soporte de expertos:** dirigido por instructores experimentados con experiencia en ciencia de datos, el curso ofrece orientación y soporte personalizados, lo que garantiza que los estudiantes puedan progresar en temas complejos a su propio ritmo.
- **Entorno de aprendizaje colaborativo:** se anima a los estudiantes a trabajar en equipos en determinadas tareas y proyectos, simulando la naturaleza colaborativa de proyectos de ciencia de datos del mundo real y mejorando sus habilidades de comunicación y trabajo en equipo
- **Horario de aprendizaje flexible:** diseñado para adaptarse a los horarios ocupados de los estudiantes, el curso ofrece una combinación de sesiones sincrónicas (en vivo) y asincrónicas (pre-grabadas), junto con horarios de oficina flexibles para soporte personalizado.

Objetivo del Curso

Al completar con éxito este curso, los estudiantes podrán:

1. Comprender los principios de la manipulación de datos y la importancia de tener datos limpios.
 2. Aprender a manipular, limpiar y preprocesar datos utilizando Python y sus bibliotecas.
 3. Dominar el arte de visualizar datos utilizando varios tipos de cuadros, gráficos y elementos interactivos.
 4. Desarrollar habilidades para comunicar conocimientos de datos complejos de manera efectiva a través de visualizaciones.
 5. Completar un proyecto final que involucra desafíos de visualización y manipulación de datos del mundo real.
- **Comprender los principios de la manipulación de datos y la importancia de los datos limpio**

Resultado: Los estudiantes obtendrán una comprensión integral del proceso de manipulación de datos, incluidos los desafíos de lidiar con datos del mundo real. Esto incluye reconocer problemas

comunes de calidad de los datos, como valores faltantes, inconsistencias y datos duplicados.

Habilidades: Al final de esta sección, los estudiantes podrán identificar y articular la importancia de la calidad de los datos y el impacto que tiene en los resultados del análisis de datos. Aprenderán las mejores prácticas en técnicas de inspección y auditoría de datos para evaluar la calidad de los datos.

- **Aprender a manipular, limpiar y preprocesar datos utilizando Python y sus bibliotecas:**

Resultado: Los participantes adquirirán competencia en el uso de Python, en particular bibliotecas como pandas y NumPy, para realizar tareas esenciales de preprocesamiento y limpieza de datos. Esto incluye el manejo de datos faltantes, conversiones de tipos de datos, normalización y transformación.

Habilidades: Los estudiantes desarrollarán habilidades prácticas en la aplicación de código Python para limpiar y pre-procesar conjuntos de datos, preparándolos para el análisis. Aprenderán cómo automatizar procesos de limpieza de datos e implementar técnicas eficientes de manipulación de datos.

- **Dominar el arte de visualizar datos utilizando varios tipos de cuadros, gráficos y elementos interactivos:**

Resultado: Los estudiantes explorarán y dominarán una amplia gama de técnicas de visualización, desde gráficos básicos como histogramas y diagramas de dispersión hasta visualizaciones más complejas como paneles interactivos y mapas geoespaciales.

Habilidades: Al interactuar con las bibliotecas Matplotlib, Seaborn y Plotly, los estudiantes adquirirán la capacidad de elegir el tipo apropiado de visualización para diferentes conjuntos de datos y objetivos de análisis. Aprenderán a personalizar las visualizaciones para hacerlas más informativas y atractivas.

- **Desarrollar habilidades para comunicar conocimientos de datos complejos de manera efectiva a través de visualizaciones:**

Resultado: El curso tiene como objetivo dotar a los estudiantes de las habilidades necesarias para traducir hallazgos de datos complejos en historias visuales claras, concisas y convincentes. Esto incluye comprender los principios del diseño de visualización de datos y cómo adaptar presentaciones visuales a diferentes audiencias.

Habilidades: Los estudiantes aprenderán técnicas para crear visualizaciones narrativas y cómo utilizar elementos visuales para resaltar ideas clave. También desarrollarán la capacidad de criticar y mejorar las visualizaciones existentes para lograr claridad e impacto.

- **Completar un proyecto final que implica desafíos de visualización y manipulación de datos del mundo real:**

Resultado: El proyecto final sirve como culminación de las habilidades y conocimientos adquiridos a lo largo del curso. Los estudiantes abordarán un proyecto integral que involucra conjuntos de datos complejos, lo que les exigirá aplicar técnicas avanzadas de visualización y manipulación de datos.

Habilidades: Este objetivo se centra en desarrollar habilidades de gestión de proyectos y resolución de problemas, ya que los estudiantes necesitarán planificar, ejecutar y presentar sus proyectos. También refuerza las habilidades de colaboración y comunicación, ya que los estudiantes pueden trabajar en equipos y presentar sus hallazgos a una audiencia.

A través de estos objetivos detallados, el curso pretende no solo impartir habilidades técnicas sino también fomentar una comprensión profunda del papel fundamental que desempeña la calidad de los datos en el análisis, el poder de la visualización en la narración de datos y la aplicación práctica de estas habilidades para resolver problemas reales. problemas mundiales.

Política de Calificación

La calificación del curso se estructura de la siguiente manera para evaluar la comprensión, el compromiso y la aplicación práctica del material del curso por parte de los estudiantes:

1. Asistencia: 20% de la nota final. La asistencia regular es crucial ya que refleja el compromiso y la participación del estudiante en el proceso de aprendizaje.
2. Tareas y Pruebas: 50% de la nota final. Este componente integral evalúa la comprensión de los estudiantes del material del curso a través de tareas y pruebas regulares. Estas tareas están diseñadas para reforzar el aprendizaje y garantizar un compromiso continuo con el contenido del curso.
3. Proyecto Final: 30% de la nota final. El proyecto final se centra en todo el contenido del curso, poniendo a prueba la comprensión general de los estudiantes y su capacidad para integrar diferentes temas aprendidos a lo largo del curso.

Esta política de calificación está diseñada para evaluar de manera justa el desempeño de los estudiantes en diferentes aspectos del curso, fomentando el esfuerzo constante, la participación activa y una comprensión profunda de la materia.

Libros de Texto y Otros Materiales

Se proporcionarán materiales del curso.

Tareas y Exámenes

Para el curso “Manipulación y Visualización de Datos”, una combinación de tareas, exámenes y un proyecto final puede proporcionar una evaluación integral de la comprensión y aplicación de los materiales del curso por parte de los estudiantes. A continuación se presenta un enfoque detallado para estructurar estos componentes:

Tareas

Objetivo: Reforzar el aprendizaje mediante la aplicación práctica y proporcionar a los estudiantes retroalimentación sobre su progreso.

Tareas semanales: Breves tareas centradas en temas específicos cubiertos en las clases, como ejercicios de limpieza de datos, desafíos de visualización o cuestionarios breves para evaluar conocimientos teóricos. Estas tareas fomentan la práctica regular y garantizan que los estudiantes se mantengan interesados con el material del curso.

Proyecto final: Proyecto más profundo que requieren que los estudiantes apliquen una combinación de habilidades aprendidas durante una semana. Por ejemplo, un proyecto podría implicar limpiar un conjunto de datos desordenado y luego crear una serie de visualizaciones para responder preguntas

específicas sobre los datos.

Proyecto Final

Objetivo: Ofrecer a los estudiantes la oportunidad de aplicar toda la gama de habilidades que han aprendido a un desafío de datos complejo del mundo real.

Descripción del proyecto: El proyecto final debe involucrar un conjunto de datos que presente desafíos genuinos de visualización y manipulación de datos, como datos faltantes, formatos inconsistentes y la necesidad de transformaciones complejas. Los estudiantes tendrán la tarea de limpiar los datos, explorarlos para descubrir ideas y crear una serie de visualizaciones que comuniquen sus hallazgos de manera efectiva.

Entregables:

- Informe de limpieza de datos: documentar los pasos tomados para limpiar y preparar los datos, incluidos los desafíos enfrentados y cómo se superaron.
- Portafolio de visualización de datos: una colección de visualizaciones creadas a partir del conjunto de datos limpios, acompañadas de breves explicaciones de los conocimientos que cada visualización pretende transmitir.
- Presentación final: una presentación ante la clase (y potencialmente invitados externos) donde los estudiantes muestran su proyecto, explican su proceso y discuten sus hallazgos. Esto ayuda a desarrollar las habilidades de comunicación de los estudiantes y permite recibir comentarios de sus compañeros e instructores.

Criterios de Evaluación

Tareas: se califican según la precisión, la integridad y la profundidad de comprensión demostrada en las respuestas. También se consideran la puntualidad en la presentación y el cumplimiento de las pautas de asignación.

Proyecto final: evaluado según la complejidad y minuciosidad del proceso de limpieza de datos, la creatividad y claridad de las visualizaciones, la capacidad de descubrir y comunicar conocimientos a partir de los datos y la presentación y el informe general.

Al integrar estos componentes en el curso, se anima a los estudiantes no solo a aprender nuevas habilidades sino también a aplicarlas en escenarios variados y cada vez más complejos, preparándose para los desafíos de los proyectos de ciencia de datos del mundo real.

Actividad del Curso

Para enriquecer la experiencia de aprendizaje en el curso “Manipulación y Visualización de Datos”, se puede diseñar una variedad de actividades didácticas para involucrar a los estudiantes, mejorar sus habilidades y fomentar una comprensión más profunda del tema. A continuación se ofrece un vistazo más detallado a las actividades potenciales:

Clases Interactivas

- **Objetivo:** Introducir conceptos y demostrar técnicas de una manera dinámica y atractiva.
- **Detalles de la actividad:** Las clases se pueden hacer interactivas a través de sesiones de codificación en vivo, donde los instructores demuestran técnicas de visualización y manipulación de datos en tiempo real usando Python. Se pueden integrar encuestas y cuestionarios para evaluar la comprensión y fomentar la participación. Las conferencias

invitadas de profesionales de la industria también pueden proporcionar información práctica sobre cómo se aplican la visualización y la manipulación de datos en diferentes sectores.

Sesiones de Laboratorio

- **Objetivo:** Proporcionar experiencia práctica con ejercicios guiados.
- **Detalles de la actividad:** Las dedicadas sesiones de laboratorio permiten a los estudiantes aplicar lo que han aprendido en un ambiente controlado con el apoyo de asistentes docentes. Estas sesiones pueden centrarse en tareas específicas, como limpiar un conjunto de datos con valores faltantes o crear un panel con elementos interactivos. Se puede fomentar la programación entre pares para fomentar la colaboración y el intercambio de conocimientos.

Foros de Discusión

- **Objetivo:** Facilitar el aprendizaje entre pares y la retroalimentación del instructor.
- **Detalles de la actividad:** Los foros de discusión en línea se pueden utilizar para que los estudiantes publiquen preguntas, compartan recursos y discutan los desafíos que enfrentan. Los instructores y asistentes de cátedra pueden moderar estos foros y brindar orientación e información adicional. Los temas de debate semanales pueden estar relacionados con el contenido del curso o eventos actuales en ciencia de datos para estimular la participación.

Estudios de Caso

- **Objetivo:** Analizar escenarios del mundo real para aplicar y profundizar la comprensión.
- **Detalles de la actividad:** Presentar a los estudiantes estudios de casos que involucran conjuntos de datos reales de diversas industrias (por ejemplo, atención médica, finanzas, comercio) que les permita abordar problemas prácticos de visualización y manipulación de datos. Estos estudios de caso se pueden discutir en grupos, y cada grupo presenta sus hallazgos y soluciones a la clase, promoviendo el pensamiento crítico y las habilidades de comunicación.

Talleres

- **Objetivo:** Explorar temas y herramientas avanzadas en profundidad.
- **Detalles de la actividad:** La organización de talleres sobre herramientas especializadas (como funciones avanzadas de Plotly) o metodologías (como el preprocesamiento de datos de aprendizaje automático) brinda a los estudiantes la oportunidad de ampliar su conjunto de habilidades. Estos talleres pueden ser dirigidos por expertos y ofrecer una combinación de presentación, demostración, y práctica.

Sesiones de Revisión por Pares

- **Objetivo:** Desarrollar habilidades críticas de evaluación y mejorar el trabajo de proyectos a través de la retroalimentación.
- **Detalles de la actividad:** Los estudiantes pueden enviar tareas o componentes del proyecto para revisión por pares, dando y recibiendo comentarios sobre su trabajo. Este proceso ayuda a los estudiantes a aprender de los enfoques y errores de los demás y a perfeccionar su propio trabajo. Se pueden proporcionar criterios para una retroalimentación efectiva para garantizar comentarios constructivos y valiosos.

Presentaciones del Proyecto Final

- **Objetivo:** Exponer la culminación del aprendizaje y recibir comentarios de una audiencia más amplia.
- **Detalles de la actividad:** Al final del curso, los estudiantes presentan sus proyectos finales a la clase e invitados de la comunidad académica y profesional. Esto no sólo les permite demostrar

sus logros sino también practicar hablar en público y recibir comentarios valiosos. También puede servir como una oportunidad para establecer contactos.

Recursos de Formación Continua

- Objetivo: Fomentar la educación y la exploración constantes más allá del curso.
- Detalles de la actividad: proporcione a los estudiantes una lista seleccionada de recursos para seguir aprendiendo, incluidos cursos en línea, tutoriales, libros y foros. La orientación sobre cómo contribuir a proyectos de código abierto o participar en concursos de ciencia de datos también puede ayudar a mantener el compromiso con el campo una vez finalizado el curso.

La incorporación de estas actividades en el curso de manipulación y visualización de datos no solo enriquece la experiencia de aprendizaje, sino que también prepara a los estudiantes para los desafíos multifacéticos de la profesión de la ciencia de datos. A través de una combinación de aprendizaje teórico, aplicación práctica e interacción con sus compañeros, los estudiantes pueden desarrollar un sólido conjunto de habilidades que son muy valoradas en la industria.

Cronograma del Curso

1. Introducción a la manipulación de datos
2. Técnicas de limpieza de datos
3. Transformación de datos
4. Introducción a la visualización de datos
5. Visualización de datos avanzada

01 - Fundamentos de la Ciencia de Datos en Python		Horas en Clases	Horas de Práctica	Clases
0	Introducción a la Ciencia de Datos	1	1,5	Clase 1
1	Instalación de Anaconda	0		Clase 1
2	Python: Funciones, string, tipos y secuencias	1	3	Clase 2
3	Python: Leyendo y escribiendo en archivos CSV	1		Clase 3
4	Python: Objetos, map(), Lambda y comprensión de listas en Python	1	1,5	Clase 4
5	Python: La librería numérica de Python (NumPy)	1	4	Clase 5 y 6
02 - Procesamiento básico de datos con pandas				
6	Python: Estructura de Datos de Python Series	1	3	Clase 7
7	Python: Consulta de Datos de Python Series	1		Clase 8
8	Python: Estructura de Datos e Indexación y Carga de un "Dataframe"	1	3	Clase 9
9	Python: Consulta de Datos de un "Dataframe" y valores vacíos	1		Clase 10
03 - Procesamiento avanzado con pandas				

10	Python: Combinando Dataframes y Modismos en pandas	1	3	Clase 11
11	Python: Escalas y manejo de fechas	1		Clase 12
04 - Análisis estadístico en Python				
12	Distribuciones y prueba de hipótesis en Python	1	1,5	Clase 13
13	Proyecto Final	0	7,5	Clase 14, 15 y 16
TOTAL DE HORAS:		12	28	40

Contenidos del Curso

Etapa		CONTENIDO
01	Fundamentos de la Ciencia de Datos en Python	En esta etapa el alumno recibirá una introducción al campo de la ciencia de datos, revisará la funcionalidad y las características comunes de Python que usan los científicos de datos, instalará Anaconda y utilizará las libretas Jupyter.
02	Procesamiento básico de datos con pandas	En esta etapa del curso el alumno aprenderá los fundamentos de uno de los juegos de herramientas más importantes que Python tiene para la limpieza y el procesamiento de datos: Pandas. Aprenderá cómo leer datos en estructuras de DataFrame, cómo consultar estas estructuras y se indexan los detalles sobre dichas estructuras.
03	Procesamiento avanzado con pandas	En esta etapa el alumno profundizará su comprensión de la biblioteca de pandas de Python al aprender a combinar DataFrames, generar tablas de resumen, agrupar datos en piezas lógicas y manipular fechas. También actualizará su comprensión de las escalas de datos y se traerá a debate los problemas con la creación de métricas para el análisis.
04	Análisis estadístico en Python	En esta etapa final del curso, al alumno se le presentará una variedad de técnicas estadísticas, tales como distribuciones, y muestreo, además de practicar con librerías de visualización. Luego se le asignará un proyecto de curso, donde participará en una actividad de limpieza de datos del mundo real y proporcionará evidencia a favor (¡o en contra!) de una hipótesis dada. Este proyecto es adecuado para una cartera de ciencia de datos y pondrá a prueba sus conocimientos sobre limpieza, fusión, manipulación y prueba de importancia en los datos.