

Programa de Asignatura
Taller de Obtención y Preparación de Datos



A. Antecedentes Generales

| | | | | | |
|---------------------------------|---|----|------------------|----|----------------------------------|
| 1. Unidad Académica | Facultad de Ingeniería | | | | |
| 2. Carrera | Ingeniería Civil Informática e Innovación Tecnológica | | | | |
| 3. Código | IIT128W | | | | |
| 4. Ubicación en la malla | I año, II semestre | | | | |
| 5. Créditos | UDD | 10 | SCT | 6 | |
| 6. Tipo de asignatura | Obligatorio | X | Electivo | | Optativo |
| 7. Duración | Bimestral | | Semestral | X | Anual |
| 8. Módulos semanales | Clases Teóricas | 1 | Clases Prácticas | 1 | Ayudantía |
| 9. Horas académicas | Clases | 68 | Ayudantía | 34 | Otras horas por periodo completo |
| 10. Pre-requisito | Taller de Programación | | | | |

B. Aporte al Perfil de Egreso

La asignatura **Taller de Obtención y Preparación de Datos**, perteneciente al ciclo de Bachillerato, está diseñado para ayudar al estudiante a adquirir habilidades necesarias para la correcta y eficiente aplicación de técnicas de obtención, limpieza y preparación de datos, utilizando criterios de imputación y manipulando las estructuras de datos a conveniencia para satisfacer las necesidades de información acorde al problema o desafío que se intenta resolver.

Este curso busca que los estudiantes utilicen métodos básicos de exploración en un set de datos mediante estructuras de tipo serie y dataframe de la librería Pandas para la selección, filtrado y sumarización de los datos.

Taller de Obtención y Preparación de Datos tributa al desarrollo de las siguientes competencias Genéricas UDD: Pensamiento Crítico, Transformación Digital y Comunicación, a las competencias específicas de la carrera: Resolución de problemas bajo un enfoque sistémico, Gestión de proyectos tecnológicos, Modelamiento científico y Aprendizaje continuo.



C. Competencias y Resultados de Aprendizaje Generales que desarrolla la asignatura

| Competencias Genéricas | Resultados de Aprendizaje Generales |
|--|---|
| <i>Pensamiento Crítico</i> | Manipula estructuras de datos vectoriales y matriciales utilizando biblioteca Numpy para resolver un problema. |
| <i>Transformación Digital</i> | |
| <i>Comunicación</i> | |
| Competencias Específicas | |
| <i>Resolución de problemas bajo un enfoque sistémico</i> | Utiliza métodos básicos de exploración en un set de datos mediante estructuras de tipo serie y dataframe de la librería Pandas para la selección, filtrado y sumarización de los datos. |
| <i>Innovación</i> | |
| <i>Gestión de proyectos tecnológicos</i> | Recupera datos desde distintas fuentes utilizando librerías utilitarias de Python para su posterior utilización. |
| <i>Modelamiento científico</i> | |
| <i>Aprendizaje Continuo</i> | Aplica técnicas para la identificación y tratamiento de valores perdidos en un set de datos utilizando librerías de Python. |
| | Aplica técnicas de Data Wrangling para la manipulación, enriquecimiento y transformación de datos en una estructura de tipo serie o dataframe. |
| | Aplica técnicas de unión, agrupamiento y pivoteo de datos utilizando la librería Pandas para el reacomodo de la información. |

D. Unidades de Contenidos y Resultados de Aprendizaje

| Unidades de Contenidos | Competencia | Resultados de Aprendizaje |
|---|---|---|
| UNIDAD I: Introducción a Numpy. <ul style="list-style-type: none"> ● Reseña de la librería Numpy. ● Creación de Arreglos de Numpy: <ul style="list-style-type: none"> ○ Vectores ○ Matrices ● Funciones preconstruidas de creación: <ul style="list-style-type: none"> ○ Arreglo con valores (arange) ○ Matrices de ceros y unos ○ Vector con distribución de puntos ○ Matriz identidad ○ Matriz aleatoria ● Redimensionado de un arreglo. ● Indexación y selección: | <i>Transformación Digital</i> <i>Innovación</i> <i>Gestión de proyectos tecnológicos</i> <i>Aprendizaje Continuo</i> | Distingue las características y utilidad de las estructuras de vectores y matrices para resolver un problema. Construye arreglos con datos utilizando funciones preconstruidas de creación de la librería Numpy. Aplica operaciones entre arreglos para resolver un problema. Utiliza selección condicional para la selección de elementos de un arreglo de acuerdo con la correspondiente sintaxis. |



| | | |
|--|---|---|
| <ul style="list-style-type: none"> ○ Selección de elementos de un arreglo. ○ Selección condicional de elementos de un arreglo. ● Referencia y copia de arreglos. ● Operaciones: <ul style="list-style-type: none"> ○ Operaciones entre arreglos ○ Operaciones con escalares ● Aplicando funciones a un arreglo. | | <p>Realiza operaciones matriciales entre arreglos Numpy para resolver un problema.</p> |
| <p>UNIDAD II: Conceptos generales de Álgebra Lineal con Python.</p> <ul style="list-style-type: none"> ● Escalares, vectores, matrices y tensores. ● Operaciones entre vectores y matrices. ● Determinante. ● Ecuaciones lineales. ● Producto escalar y ortogonalidad. ● Vectores linealmente independientes. ● Transformaciones lineales. ● Autovalores. | <i>Pensamiento Crítico</i> <i>Transformación Digital</i> <i>Modelamiento científico</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> <i>Aprendizaje Continuo</i> | <p>Realiza operaciones matriciales y vectoriales para resolver un problema determinado.</p> <p>Identifica el uso del álgebra como parte esencial de la rama informática vinculada a la analítica y la ciencia de datos.</p> <p>Identifica el uso de las matemáticas en la solución de problemas cotidianos.</p> |
| <p>UNIDAD III: Introducción a Pandas.</p> <ul style="list-style-type: none"> ● Reseña de la librería Pandas. ● Para qué se utiliza. ● El tipo de dato Serie: <ul style="list-style-type: none"> - Características del tipo de dato Serie. - Creación de una Serie. - Obtención de datos de una Serie. - Operaciones sobre una Serie. ● El tipo de dato DataFrame: <ul style="list-style-type: none"> - Características del tipo de dato DataFrame. - Creación de un DataFrame. - Selección de filas o columnas en un DataFrame. - Selección de elementos en un DataFrame. - Selección condicional en un DataFrame. | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> <i>Aprendizaje continuo</i> | <p>Distingue las características y utilidad de las estructuras de Series y DataFrames para resolver un problema.</p> <p>Selecciona elementos de una serie y filas o columnas de un dataframe para resolver un problema.</p> <p>Utiliza la sintaxis de selección condicional para el filtrado de elementos de un DataFrame.</p> <p>Utiliza métodos básicos de exploración una serie o dataframe para resolver un problema.</p> |



| | | |
|--|---|--|
| <ul style="list-style-type: none"> Métodos básicos de Exploración: <ul style="list-style-type: none"> - Métodos básicos de exploración (head, tail, info, describe). - Métodos básicos de sumarización (min, max, count, mean, median, sum). - Métodos unique, nunique, value_counts. | | |
| UNIDAD IV: Obtención de datos desde archivos. <ul style="list-style-type: none"> Obtención y guardar datos desde archivos: <ul style="list-style-type: none"> - Archivos CSV. - Archivos Excel. - Archivos Binarios. - El formato HDF. Obtención y guardar datos desde la web: <ul style="list-style-type: none"> - Leer tablas web con Pandas. Obtención y guardar datos desde una Base de Datos Relacional: <ul style="list-style-type: none"> - Librerías de acceso a una base de datos. - Rescate de datos de una tabla. | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Resolución de problemas bajo un enfoque sistémico</i> <i>Modelamiento científico</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> <i>Aprendizaje Continuo</i> | <p>Identifica los distintos métodos para la obtención de datos de diversas fuentes utilizando librerías de Python.</p> <p>Recupera información desde archivos de texto CSV y Excel utilizando librerías Python para su posterior procesamiento.</p> <p>Recupera información desde la web utilizando librerías de Python para su posterior procesamiento.</p> <p>Recupera información desde bases de dato relacionales utilizando librerías Python para su posterior procesamiento.</p> |
| UNIDAD V: Manejo de valores nulos y atípicos. <ul style="list-style-type: none"> Manejo de Valores Perdidos: <ul style="list-style-type: none"> - Qué es un valor perdido. - Identificación de valores perdidos. - Filtrado de la data perdida. - Imputación de datos. - Imputación de valores cualitativos. - Técnicas avanzadas de imputación. Identificación de Outliers: | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Resolución de problemas bajo un enfoque sistémico</i> <i>Modelamiento científico</i> | <p>Utiliza técnicas para la identificación de valores perdidos en un set de datos mediante el uso de librerías.</p> <p>Explica criterios para el tratamiento de los valores perdidos describiendo mecanismos de imputación.</p> <p>Utiliza técnicas para la detección y filtrado de registros outliers en un set de datos.</p> |



| | | |
|--|---|--|
| <ul style="list-style-type: none"> - Qué es un Outlier. - Detección y filtrado de Outliers. | <i>Innovación</i> <i>Gestión de proyectos</i> <i>Tecnológico</i> <i>Aprendizaje Continuo</i> | |
| UNIDAD VI: Data Wrangling. <ul style="list-style-type: none"> ● Qué es data wrangling. ● Principales tareas de data wrangling. ● Ordenamiento y manipulación de datos: <ul style="list-style-type: none"> - Muestreos aleatorios, permutación de la data. - Ordenamiento de la data. - Detección y eliminación de registros duplicados. - Reemplazo de valores. - Discretización y Binning. ● Manipulación de Cadenas de Caracteres: <ul style="list-style-type: none"> - Métodos de la clase String. - Expresiones Regulares ● Enriquecimiento de la data: <ul style="list-style-type: none"> - Utilización de funciones y mapeos. - Aplicación de funciones dentro de una Serie de datos (apply). - Aplicación de funciones dentro de un DataFrame (apply). - Utilización de expresiones lambda en la aplicación de funciones. ● Manipulación de la estructura de un DataFrame: <ul style="list-style-type: none"> - Agregar o eliminar columnas. - Redimensión de un dataframe. - Convertir el tipo de dato de una columna. - Definir y resetear índices. - Renombrar columnas e índices. - Remover duplicados. | <p>Describe el concepto de Data Wrangling identificando sus principales tareas y reconociendo su importancia.</p> <p>Aplica técnicas de ordenamiento y manipulación de datos utilizando la librería Pandas.</p> <p>Aplica técnicas de enriquecimiento de los datos utilizando funciones y expresiones lambda.</p> | |



| | | |
|--|--|---|
| UNIDAD VII: Agrupamiento y Pivoteo de datos. <ul style="list-style-type: none"> • Agrupamiento con indexación jerárquica. • Agrupamiento de datos con la función groupby. • Pivoteo de DataFrames. • Despivoteo de DataFrames (método melt). • Combinación y Merge de datos: <ul style="list-style-type: none"> - Concatenación de dataframes. - Merge de dataframes. | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Comunicación</i> <i>Resolución de problemas bajo un enfoque sistemático</i> <i>Modelamiento científico</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> <i>Aprendizaje Continuo</i> | <p>Utiliza indexación jerárquica para ordenar y agrupar los datos de un dataframe utilizando la librería Pandas.</p> <p>Utiliza la función groupby para facilitar el análisis agregado de la información.</p> <p>Aplica técnicas de pivoteo y de despivoteo de dataframes para facilitar el análisis agregado de la información.</p> <p>Aplica técnicas para la combinación de dos sets de datos para resolver un problema.</p> |
| UNIDAD VIII: Manejo de fecha. <ul style="list-style-type: none"> • Ingesta de datos de fechas. • Manipulación y transformación de datos de fechas. | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Comunicación</i> <i>Resolución de problemas bajo un enfoque sistemático</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> | <p>Aplica técnicas para un eficiente uso de datos del tipo fecha.</p> |



| | | |
|---|--|---|
| | <i>Aprendizaje Continuo</i> | |
| UNIDAD IX: Consideraciones de desempeño. <ul style="list-style-type: none"> • Uso eficiente del almacenamiento. • Uso eficiente del tiempo de procesamiento. | <i>Pensamiento Crítico</i> <i>Transformación digital</i> <i>Comunicación</i> <i>Resolución de problemas bajo un enfoque sistémico</i> <i>Modelamiento científico</i> <i>Innovación</i> <i>Gestión de proyectos Tecnológicos</i> <i>Aprendizaje Continuo</i> | Aplica distintas técnicas para almacenar y procesar los datos, para usar eficientemente los recursos de procesamiento y almacenamiento disponibles. |

E. Estrategias de Enseñanza

El curso será abordado mediante variadas estrategias metodológicas, cada una de ellas formulada sobre la base de los resultados de aprendizaje que se desea transferir y desarrollar en el estudiante, las cuales son:

- Clases expositivas.
- Uso de plataforma de trabajo interactivas (Jupyter Notebooks).
- Formulación de preguntas.
- Aprendizaje Basado en Proyectos (ABP).
- Debates.
- Flipped Learning.
- Resolución de ejercicios y problemas.
- Talleres en sala.
- Talleres en laboratorio.



F. Estrategias de Evaluación

Para las diferentes instancias evaluativas se contará con criterios claros y conocidos por los estudiantes, el curso contará con al menos dos certámenes y exámenes.

- Rúbricas.
- Informes escritos (Bitácora de equipo).
- Certámenes (Project).
- Exámenes (Capstone Project).

G. Recursos de Aprendizaje

Bibliografía Obligatoria:

- McGregor, S (2021). *Practical Python Data Wrangling and Data Quality*. O'Reilly Media, Inc.
- Cohen, M (2022). *Practical Practical Linear Algebra for Data Science*. O'Reilly Media, Inc.
- McKinney, W (2022). *Python for Data Analysis, 3rd Edition*. O'Reilly Media, Inc.
- Johansson, R (2018). *Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*, Apress