

Aspectos generales

| | |
|--|---|
| Título: | Introducción al análisis de datos con Python |
| Programas de posgrado o planes de estudio en donde se ofertará adicionalmente: | Ciencias Biológicas Ciencias Biomédicas Ciencias Médicas Odontológicas y de la Salud. |
| Área del conocimiento: | Genética, genómica y bioinformática |
| Semestre: | 2024-2 |
| Modalidad: | Tópico selecto |
| Horario: | Viernes 9:00-13:00 |
| No. sesiones: | 16 |
| Horas por sesión: | 4.0 |
| Total alumnos PDCB: | 5 |
| Total alumnos: | 20 |
| Videoconferencia: | Si |
| Lugar donde se imparte: | Instituto Nacional de Enfermedades Respiratorias |
| Informes: | maresther.martinez@iibiomedicas.unam.mx |

Métodos de evaluación

| MÉTODO | PORCENTAJE | NOTAS |
|-----------------------------|------------|-------|
| Participación en clase | 20% | |
| Presentación de un proyecto | 40% | |
| Trabajos | 40% | |

Contribución de este curso/tópico en la formación del alumnado del PDCB:

El alumno desarrollará habilidades básicas de análisis y visualización de datos, las cuales le servirán para desarrollar sus proyectos de investigación. Se recomienda fuertemente que el alumno tenga ya un conjunto de datos para trabajar.

Profesor (a) responsable

| | |
|-----------|---|
| Nombre: | Martínez Sanchez Mariana Esther |
| Teléfono: | |
| Email: | maresther.martinez@iibiomedicas.unam.mx |

Profesores (as) participantes

| PARTICIPANTE | ENTIDAD O ADSCRIPCIÓN | SESIONES |
|--------------|-----------------------|----------|
|--------------|-----------------------|----------|

| | | |
|---|----------------------|---|
| MARTÍNEZ SANCHEZ MARIANA ESTHER Responsable | Facultad de Medicina | Análisis de datos |
| | | Análisis de datos |
| | | Análisis de datos |
| | | Exploración y limpieza de datos (8 hrs) |
| | | Exploración y limpieza de datos (8 hrs) |
| | | Graficación básica (12 hrs) |
| | | Graficación básica (12 hrs) |
| | | Graficación básica (12 hrs) |
| | | Introducción a python (8 hrs) |
| | | Introducción a python (8 hrs) |
| | | Proyecto |
| | | Proyecto |
| | | Proyecto |

Introducción

El objetivo del curso es que los alumnos adquieran la capacidad de manejar conjuntos de datos de datos sencillos, por ejemplo una hoja de Excel utilizando el lenguaje de programación python. Aprenderán a cargar, manipular, limpiar, obtener estadísticas descriptivas y generar visualizaciones básicas de conjuntos de datos sencillos usando el lenguaje de programación Python. Se trabajará con un conjunto de datos propuesto por el alumno cuyo análisis se presentará como proyecto final.

Temario

Unidad 1: Introducción a python (8 hrs)

- Instalación de anaconda
- Jupyter notebooks
- Introducción a python
- Introducción a github básico
- Descargar datos abiertos
- ¿Qué es un dato?
- Cargar los datos en python
- Ubicación en la carpeta
- Excel, csv, pickle
- Documentación de python

Unidad 2: Exploración y limpieza de datos (8 hrs)

- Tidy data
- Pasos de un análisis de datos
- Exploración básica - pandas_profiling
- Operaciones básicas de pandas - slicing
- Estructura de una tabla (index, columns, dtype, nan)
- Limpieza de datos
- Guardar datos

Unidad 3: Graficación básica (12 hrs)

- Teoría de la visualización de datos
- Tipos de gráficas (lines, bars, scatter, hist)
- Elementos de una gráfica (title, axis, etc)
- Biblioteca matplotlib
- Biblioteca seaborn (distplot, heatmap, multiples)

Unidad 4: Análisis de datos (12 hrs)

- Estadística básica (count, sum, mean, median, moda, std)
- Estadística intermedia (corr, linear_regression)
- Agrupamiento (groupby, filter, aggregate)
- Tablas pivote (pivot_table)
- Consideraciones del análisis de datos

Unidad 5: Proyecto (24 hrs)

- Exploración, limpieza, visualización y análisis de un conjunto de datos propuesto por el alumno.

Bibliografía

- Downey, A. (2008). How to think like a computer scientist: learning with python. Green Tea Press.
- Westfall, P. H., & Henning, K. S. (2013). Understanding advanced statistical methods (p. 543). Boca Raton, FL, USA:: CRC Press.
- Wickham, Hadley. 2014. "Tidy Data." Journal of Statistical Software 59 (1): 1–23.
- doi:10.18637/jss.v059.i10.
- Ferreira Filho, W. (2017). Computer Science Distilled: Learn the Art of Solving Computational Problems. Code Energy LLC.
- Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures. O'Reilly Media.
- D'ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press.
- Danjou, J. (2016). The Hacker's Guide to Python. Julien Danjou. Danjou, J. (2016). The Hacker's Guide to Python. Julien Danjou.
- How do I ask a good question? - Help Center. (n.d.). Stack Overflow. Retrieved November 3, 2022, from <https://stackoverflow.com/help/how-to-ask>

Documentación en línea constantemente actualizada:

- Anaconda <https://www.anaconda.com/faq>
- Jupyter Project Documentation <https://docs.jupyter.org/en/latest/>
- Guía estilo <https://pep8.org/>
- Selección de colores <http://colorbrewer2.org/>
- Pandas documentation <https://pandas.pydata.org/docs/>
- Seaborn documentation <https://seaborn.pydata.org/>

Observaciones

Se trabajará con un conjunto de datos propuesto por el alumno cuyo análisis se presentará como proyecto final. Es OBLIGATORIO que el alumno contacte a la profesora y describa el conjunto de datos (tema y tamaño) antes de inscribirse al curso.