

Curso “On-line”

1. Nombre del Actividad:

“Big Data: Análisis de datos sociales con Python”

2. Fundamentación:

En la era actual, el manejo de grandes volúmenes de datos (Big Data) se ha convertido en una herramienta fundamental para el sector público. La capacidad de analizar y extraer información valiosa de estos conjuntos de datos permite a las entidades gubernamentales tomar decisiones más informadas, mejorar la eficiencia de sus operaciones y brindar mejores servicios a los ciudadanos.

El presupuesto es una de las principales herramientas de gestión, reflejando políticas, y cuantificando en términos monetarios el plan de gobierno. El procesamiento e interpretación de la información que contiene es vital para poder lograr los objetivos de política y para un mejor monitoreo de la gestión.

Dada la envergadura de las bases del presupuesto público, se necesitan herramientas especialmente diseñada para poder llevar adelante detallados análisis con gran cantidad de datos, pero de una forma sencilla.

Python es un potente lenguaje de programación, utilizado para un sinnúmero de tareas en la industria. Pero en los últimos años, el desarrollo de módulos para el procesamiento de información lo ha situado como la herramienta por defecto para procesar el pago grandes volúmenes de datos.

Python es software libre, lo que permite usarlo sin el pago de licencias. Además, tiene una gran comunidad de desarrolladores que mejoran día a día cada una de sus principales características.

Su relativa facilidad de uso ha hecho posible que usuarios ajenos al ambiente IT lo utilicen y participen de su actividad.

3. Contribución Esperada:

Al finalizar este curso, los participantes podrán:

- Comprender los fundamentos del Big Data y su relevancia para el sector público.
- Instalar y configurar herramientas de desarrollo Python para Big Data.
- Manipular y analizar conjuntos de datos utilizando bibliotecas Python como Pandas y NumPy.
- Visualizar datos de manera efectiva utilizando bibliotecas como Matplotlib y Seaborn.

4. Objetivos Específicos

El curso se propone contribuir a la gestión cotidiana para aquellas personas interesadas en la big data del sector público en general, y el presupuesto en particular. El funcionario obtendrá las herramientas necesarias para:

- Comprender los principios básicos de la programación y procesamiento de datos con python
- Generar reportes analíticos y gráficos con cruces de información presupuestaria con módulos tales como Pandas y Matplotlib.

5. Perfil del Participante:

Este curso está dirigido a funcionarios que deseen adquirir las habilidades necesarias para trabajar con Big Data utilizando Python. No se requieren conocimientos previos de programación, aunque se recomienda tener familiaridad con conceptos básicos de informática.

6. Contenido Temático

Módulo I: Introducción a la programación en Python

- Cómo instalar Python
- Archivos de programa en Python
- Jupyter Notebooks
- Variables y tipos
- Operadores y comparaciones
- Tipos compuestos (cadenas, listas, tuplas)
- Control de flujo
- Ciclos
- Funciones
- Clases
- Módulos
- Excepciones

Módulo II: Exploración y procesamiento de datos

- Librerías principales
- Importación de datos: txt, csv, xlsx, y zip.
- Inspección inicial de la información.
- Manejo básico de la librería pandas
- Selección de filas y/o columnas, selección condicionada.
- Funciones apply, mapping and merge.
- Funciones de manejo de texto.
- Creación de muestras y subconjuntos de data frames: subsets and samples.

Módulo III: Paquetes Gráficos

- Introducción a plotly y matplotlib y demás librerías gráficas.
- Gráficos base: histogramas, scatter plot, bar plot, box plot.
- Otras herramientas para visualización de datos: manejo de mapas.
- Personalización de gráficos: ejes, colores, títulos.
- Exportación de gráficos.

Módulo IV: Análisis de datos ponderados

- Procesamiento de la Encuesta Permanente de Hogares
- Creación de indicadores sociales a partir de muestras ponderadas
- Módulos ad-hoc

Módulo V: Scrapping de datos

- Introducción a BeautifulSoup
- Ventajas y limitaciones del scrapping y de la librería
- Aplicación: Creación de una base de datos con las noticias de la Casa Rosada

Módulo VI: Georreferenciación

- Georreferenciación de información con el módulo Geopandas
- Introducción a las proyecciones y formatos de capas con información espacial (shape, GeoJson).
- Puntos y mapas coropleticos
- Operaciones de análisis espacial
- Isocronas

Módulo VII: Presupuesto en Python

- Análisis de datos presupuestarios de la ONP con Pandas
- Módulo PySupuesto

Módulo VIII: Dashboards

- ¿Qué es un framework y por qué lo utilizamos?
- Presentación de los principales frameworks para programar tableros de control
- Elaboración de una simple demo con Dash o Streamlit

7. Bibliografía y sitios recomendados

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, Wes McKinney. Del creador de la librería Pandas, un libro detallado sobre su uso
- <http://www.python.org> - The official web page of the Python programming language.
- <http://www.python.org/dev/peps/pep-0008> - Guía de estilo para la programación en Python.
- <http://www.greenteapress.com/thinkpython/> - Libro gratuito sobre Python.
- Python Essential Reference - Un buen libro de referencia sobre programación en Python.

Recursos adicionales:

- Kaggle: Sitio con datagrames y notebooks preparados para practicar EDA y machine learning. Bases de datos de todo tipo, con código comentado. Recomendado para practicar con datos reales.
- Google Colab: Editor y ejecutor de notebooks en la nube. No hace falta instalar nada local. Gratis.
- Corey Schafer: Excelente fuente de tutoriales sobre Python.

- Keith Galli: Completo tutorial de Pandas.

8. Estrategia Metodológica

Exposición teórica. Se realizará la exposición teórica de los temas desarrollados en el programa, con el auxilio de recursos audiovisuales (filminas PPT y videos), con un espacio para preguntas en el que se promoverá el debate.

Se enviará material teórico en formato PDF y en filminas formato PPT.

Los participantes podrán presentar para su análisis y resolución, casos propios de su realidad laboral junto a una propuesta de solución, para ser debatida entre los participantes, con el seguimiento del Docente que ejercerá el rol de facilitador.

9. Recursos Didácticos:

Se desarrollará totalmente de manera virtual donde se trabajará a través de la exposición del profesor, con la aplicación de las herramientas virtuales necesarias para que cada uno de los participantes acceda de manera clara y efectiva a los conceptos desarrollados.

Se realizarán intercambios a través del uso de la plataforma que permitan dar respuesta a las inquietudes que surjan sobre cada uno de los temas desarrollados.

Se trabajará con el uso de videoconferencia, placas interactivas y chat con los participantes con el fin de generar la dinámica más adecuada para el acceso a los conceptos desarrollados en la actividad. Se promoverá la discusión y participación.

10. Descripción de la Modalidad

El curso se desarrollará totalmente de manera on-line en siete (7) clases de dos (2) horas cada una, con una duración total de catorce (14) horas.

Durante las mismas se llevarán a cabo actividades consistentes en ejercicios, prácticos y comentarios de documentos a los que accederán vía Internet.

11. Evaluación

Se realizará un cuestionario práctico para evaluar el aprendizaje adquirido.

Al final de la cursada se realizará una encuesta de opinión sobre el docente.

12. Requisitos de Asistencia y Aprobación

Los participantes deberán asistir al 80% de las clases dictadas.

13. Lugar

Para el desarrollo de las clases se usará la plataforma Zoom, y los participantes podrán acceder mediante un link que les será suministrado por ASAP al comienzo de las clases.