

WEBINAR VIRTUAL

INTRODUCCIÓN AL PROCESAMIENTO DE DATOS CON PANDAS

Síntesis

Pandas es una biblioteca de software escrita como extensión de Numpy para manipulación y análisis de datos con el lenguaje de programación Python. Manipula grandes sets o conjunto de datos numéricos, tablas y series de tiempo. Maneja múltiples formatos de archivos de datos como txt, gz, parquet, pickle, csv, xls entre otros.

I. Comenzando con un poco de teórica

- Introducción sobre la data
- Tendencias de crecimiento de la data en el mundo
- ¿Qué es python?
- Un poco de historia de pandas
- ¿Qué es Pandas?

II. Característica y funcionalidades de Pandas

- Características interesantes sobre Pandas
- Comprender el significado y uso de los DataFrames en Pandas
- Ejecutar operaciones básicas con Pandas
- Graficar la data con Matplotlib
- Leer y exportar datos en diferentes formatos de archivos
- Limitaciones sobre Pandas

III. Aplicando pandas a la realidad

- Demo: Analizando un archivo con Dataframes.
- Recomendaciones/conclusiones

Martín David Llauce Santos: Software Engineering en [Teamcore Solutions](https://www.teamcore.net)(www.teamcore.net) con amplia experiencia en desarrollo de software, gestión de proyectos y operaciones de TI. Las áreas de conocimiento incluyen las tecnologías de código abierto, virtualización de infraestructura, tecnologías móviles y arquitecturas en la nube con Google Cloud Platform(GPC) y Amazon Web Service(AWS) .

Experiencia basada en el desarrollo proyectos en python(Django, Flask, Django Rest-Framework), pruebas de software(Unit Testing), desarrollo de pipelines con despliegue continuo basado integración continua y entrega continua(CI/CD), dockerización de aplicaciones para escalamiento horizontal almacenados en Container Registry(GPC) y Elastic Container Registry(AWS), desarrollo y dockerización de scraping con selenium y despliegue en VM Spot(GCP), desarrollo de ETL desplegados en App Engine(GPC), mantenimiento de Data Lake desplegado en AWS Lambda con almacenamiento de archivos parquets en S3 y consultados por Athena(AWS), desarrollo de pipeline con apache beam desplegados en Dataflow(GCP), desacoplamiento de ETL para escalamiento horizontal usando tecnologías como cloud storage, firestore y bigquery.

Linkedin: <https://www.linkedin.com/in/martin-david-llauce-santos/>

GitHub: <https://github.com/davidllauce>

Twitter: [@davidllauce](https://twitter.com/davidllauce)