

# PostgreSQL para Ciencia de Datos: Desbloquea el Potencial de tus Bases de Datos Relacionales 🚀

**Docente:** Yoseph Daniel Ayala Valencia (yoseph.ayala@pucp.edu.pe)

---

## 1. Sumilla

Este curso ha sido diseñado para capacitar a los estudiantes en el uso de PostgreSQL, una de las herramientas de bases de datos más robustas y versátiles del mercado. Exploraremos desde los fundamentos hasta las funcionalidades avanzadas, integrando Python como un aliado estratégico para potenciar el análisis y manipulación de datos. Los estudiantes aprenderán a desarrollar flujos de trabajo eficientes que combinen SQL y Python, ampliando sus capacidades en ciencia de datos y soluciones de bases de datos modernas. 💻 🔗

---

## 2. Descripción del Curso

El curso aborda la creación, gestión y optimización de bases de datos relacionales en PostgreSQL, con un enfoque práctico en resolver desafíos reales de análisis de datos. Aprenderemos a integrar Python para interactuar con PostgreSQL y automatizar procesos, sin perder el enfoque en las potentes funcionalidades nativas de SQL. Al final del curso, los estudiantes estarán capacitados para diseñar proyectos de análisis de datos desde la conceptualización hasta la presentación de resultados, utilizando lo mejor de ambas herramientas. 📊 🔗

---

## 3. Objetivos del Curso




Al finalizar el curso, los estudiantes serán capaces de:

- 📁 Diseñar, implementar y administrar bases de datos relacionales en PostgreSQL.
  - 🛠️ Escribir consultas avanzadas, utilizando funciones de agregación, subconsultas, funciones de ventana, y procedimientos almacenados.
  - 🤖 Desarrollar funciones SQL dentro de scripts Python para automatizar tareas y optimizar flujos de trabajo.
  - 📊 Presentar proyectos analíticos de alto impacto mediante storytelling efectivo con datos.
- 

## 4. Metodología

Este curso combina teoría y práctica, con un enfoque en la resolución de problemas reales.

## Actividades:

-  Clases interactivas con ejemplos prácticos en PostgreSQL.
-  Integración de Python para ejecutar consultas y enriquecer análisis.
-  Proyectos que aplican los conceptos aprendidos, con orientación constante para garantizar el éxito de los participantes.

**Nota:** Este curso no cuenta con clases de práctica. Sin embargo, las clases se caracterizarán por combinar teoría con práctica. Nos enfocaremos en aquello que todo Data Analyst/Data Scientist necesita en su toolkit.

---

## 5. Contenido del Curso

### Clase 1 (22/01): Introducción a PostgreSQL y Primeras Consultas

- Conceptos Básicos de Bases de Datos Relacionales.
  - Instalación y Configuración de PostgreSQL.
  - Creación de Bases de Datos y Tablas.
    - INSERT, UPDATE, DELETE en PostgreSQL.
  - Consultas Básicas en SQL:
    - Uso de SELECT, WHERE, y operadores de comparación.
  - Conexión de PostgreSQL con Python:
    - Uso de **psycopg2** para conectarse a PostgreSQL desde Python.
    - Primeras consultas desde Python para interactuar con la base de datos.
- 

### Clase 2 (29/01): Filtros, Condicionales y fuentes de información

- Operadores y Funciones en WHERE:
    - Operadores matemáticos (+, -, \*, /, %).
    - Operadores de texto: LIKE, ILIKE, SIMILAR TO, y el uso de expresiones regulares.
    - Operadores de tiempo y fecha: Uso de funciones como NOW(), CURRENT\_DATE, y operadores de comparación con fechas.
    - Operadores lógicos: IN, NOT IN, BETWEEN, IS NULL.
  - Uso de la cláusula CASE WHEN para condicionales.
  - Distintas Fuentes de Información:
    - Utilización de distintas fuentes de información: .csv, .xlsx, páginas web, etc.
    - Utilización de Python para subida de distintas fuentes de información a PostgreSQL
- 

### Clase 3 (05/02): Joins, Subconsultas y Optimización de Consultas

- Joins (INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN).
  - Subconsultas en SQL para análisis más complejos.
  - Optimización de consultas.
  - Ejercicio de integración con Python:
    - Uso de Python para ejecutar consultas complejas con JOINS y subconsultas.
    - Implementación de un análisis de datos basado en estas consultas.
- 

#### **Clase 4 (12/02): Funciones de Agregación, Agrupamiento y Funciones de Ventana**



- Funciones de Agregación:
    - Uso de SUM, COUNT, AVG, MIN, MAX.
  - Agrupamiento de Datos:
    - Uso de GROUP BY para agrupar datos.
    - Filtrado de resultados agregados con HAVING.
  - Funciones de Agregación con condicionales:
    - Uso de CASE WHEN con funciones de agregación.
  - Introducción a Funciones de Ventana:
    - Aplicación de funciones de ventana en análisis de datos.
    - Cláusula OVER.
- 

#### **Clase 5 (19/02): Tablas Temporales y Proyecto**



- Creación de tablas temporales y uso de Common Table Expression (CTE)
  - Estructuración del Proyecto Final
    - Estructura de cómo se organiza de los scripts: inputs, queries, análisis, resultados
    - Storytelling - ¿Cómo dar una presentación de impacto usando datos?
    - Ejemplo de caso de estudio que se verá en clase
- 

#### **Clase 6 (26/02): Variables, Procedimientos Almacenados y Funciones**



- Uso de Variables
  - Procedimientos Almacenados y funciones definidas por el usuario.
  - Ejercicio práctico con Python:
    - Creación y ejecución de funciones y procedimientos almacenados desde un script Python, utilizando psycopg2.
- 

#### **Clase 7 (05/03): Vistas, Seguridad y Administración de Bases de Datos**



- Creación y uso de Vistas.

- Administración de usuarios y permisos en PostgreSQL.
  - Seguridad en bases de datos:
    - Gestión de roles y contraseñas.
  - Mejores prácticas en administración de bases de datos.
  - Creación de procedimientos de respaldo y restauración de bases de datos.
- 

### Clase 8 (12/03): Proyecto Final 🎓

- ¡Hora de demostrar lo aprendido! 🚀
  - Exposición Final del proyecto por los alumnos
- 

## 6. Evaluación

- 📄 **Trabajo Final (60%):**
    - El trabajo final (12 puntos).
    - 2 avances (6 puntos cada uno).
      - Primer avance (entrega hasta 15/02).
      - Segundo avance (12/03).
  - 📊 **Evaluación (20%):**
    - 1 evaluación para reforzar conocimientos (entrega hasta 25/02) (4 puntos).
  - 🕒 **Asistencia (10%):**
    - 80% de asistencia (6 clases) para puntaje máximo (2 puntos).
  - 🗣️ **Participación en Clase (10%):**
    - Participación activa en clase (2 puntos).
- 

## 7. Sobre el docente

👤 Yoseph Daniel Ayala Valencia es Economista licenciado de la PUCP. Cuenta con un diplomado en Ciencia de Datos de la PUCP. En cuanto a su experiencia laboral, actualmente trabaja como Científico de Datos en el BCP. Anterior a ello, ha sido Analista de Ciencia de Datos en Financiera Confianza.

Cuenta con amplia experiencia en el sector académico. Ha trabajado para Columbia University, liderando un proyecto en conjunto con la Contraloría General de la República. Además, ha sido asistente de investigación en centros como GRADE y el Departamento de Economía de la PUCP, entre otros.