Informe Semillero Análisis de Datos (AD). Semestre 2021-2.

Enfoque.
El semillero está orientado a aprender, asimilar y aplicar todo lo relacionado con los programas Numpy, Pandas, MatplotLib y Scikit-Learn. Estos programas forman parte de las herramientas esenciales, creadas en Python para el análisis de datos.

Objetivo

Desarrollar las competencias necesarias para realizar análisis de datos exploratorio e inferencial.

Estudiantes.

El semillero está dirigido y puede ser de utilidad para todos los estudiantes de la Universidad EIA en sus diferentes escuelas. No tiene prerrequisitos.

Metodología.

Encuentros semanales (este semestre: miércoles 14-16 y sábados 8-10) con presentación de los temas propuestos, formulación de tareas prácticas para cada tema. El semillero esta dividido en dos partes: herramientas básicas(Numpy, Pandas y MatplotLib) y bases de Machine Learning con Scikit-Learn. Cada una de estas partes esta acompañada de un proyecto de investigación para aplicar lo aprendido en el semillero.

Programa.

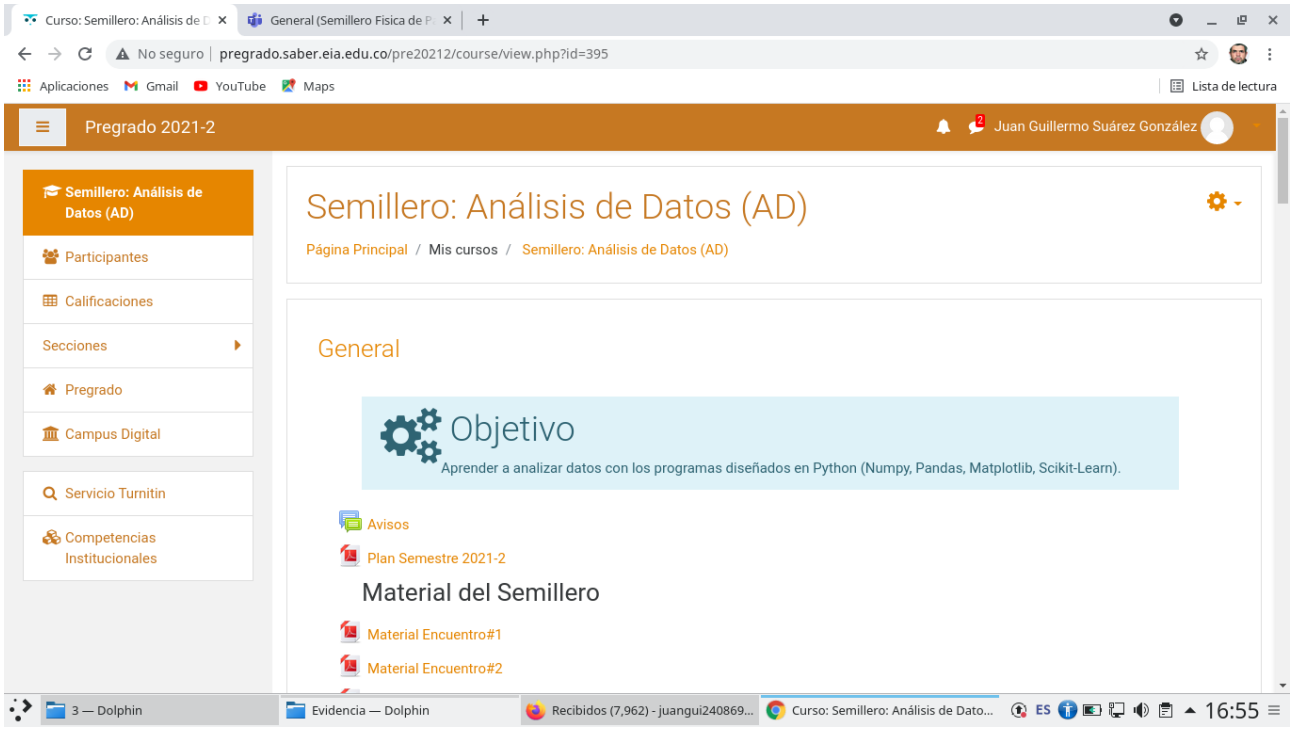
#1. 04.08.21 Python para el AD.
#2. 11.08.21 Numpy I.
#3. 18.08.21 Numpy II.
#4. 25.08.21 Numpy III.
#5. 01.09.21 Numpy IV.
Parciales 04-08.09.21
#6. 15.09.21 Pandas I. Formulación del Proyecto I (EDA).
#7. 22.09.21 Pandas II.
#8. 25.09.21 Pandas III.
#9. 29.09.21 Pandas IV. Proyecto II (IDA).
#10. 06.10.21 Series Temporales.
#11. 09.10.21 Matplotlib I.
#12. 13.10.21 Matplotlib II.
#13. 16.10.21 Scikit_Learn I.
#14. 20.10.21 Scikit-Learn II.
#15. 27.10.21 Hyperparámetros y Validación de modelos.
#16. 03.11.21 Revisión de proyectos 1 y 2.

06.11.21 Fin de clases.

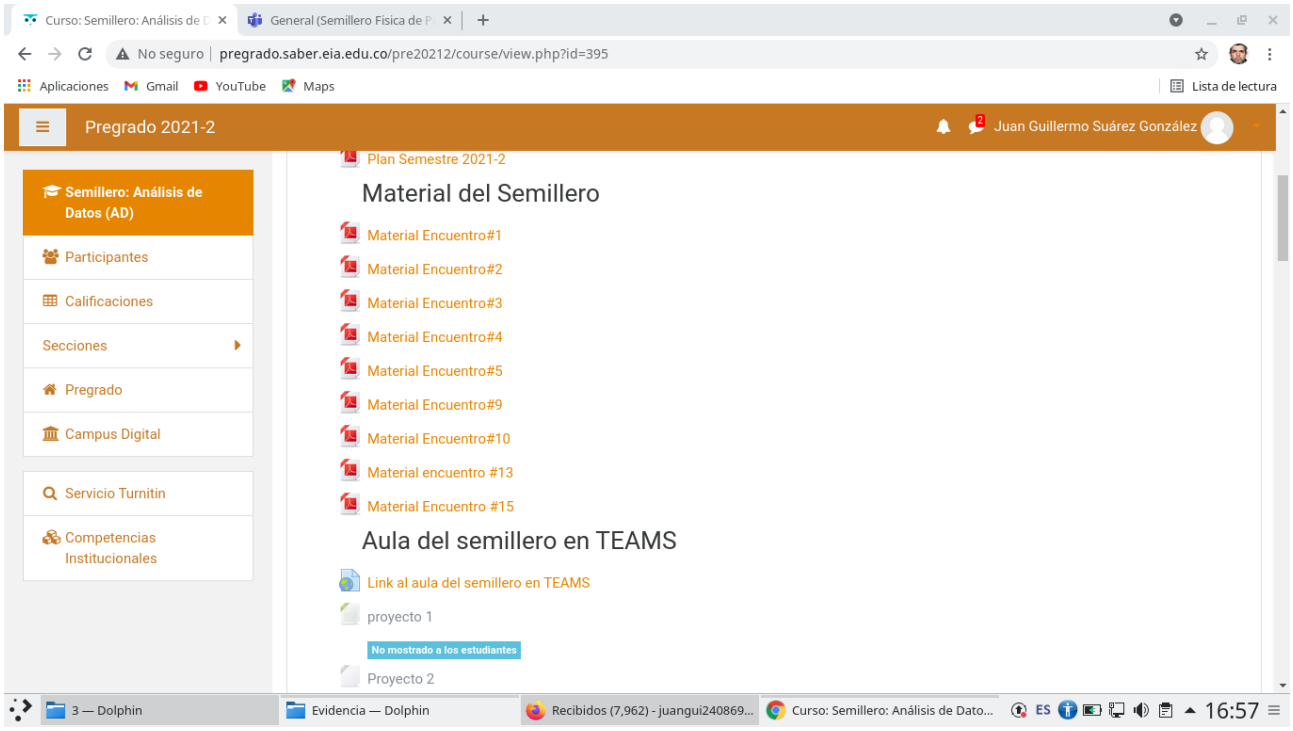
Cada encuentro está acompañado de un notebook de Jupyter. Los notebooks están disponibles en el aula en Moodle del semillero.

Aula en Moodle.
El aula en Moodle incluye:

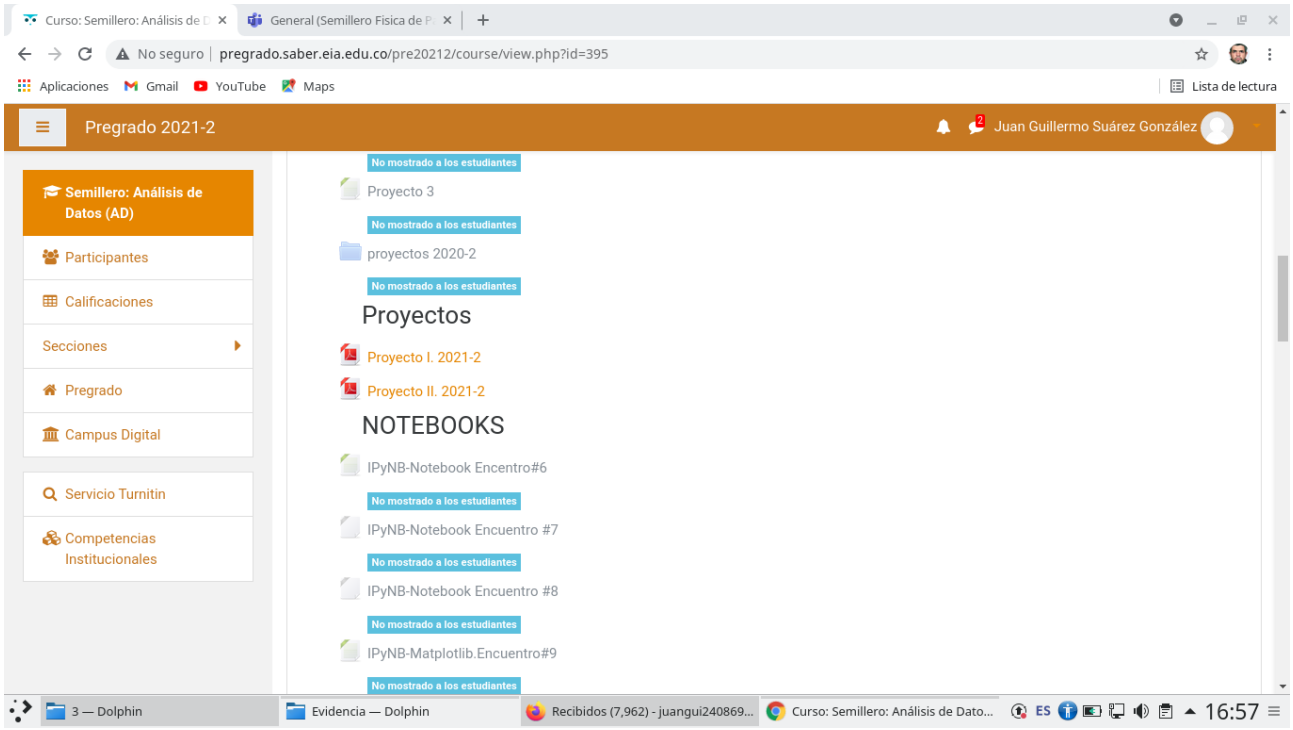
1. Objetivo del semillero y plan.



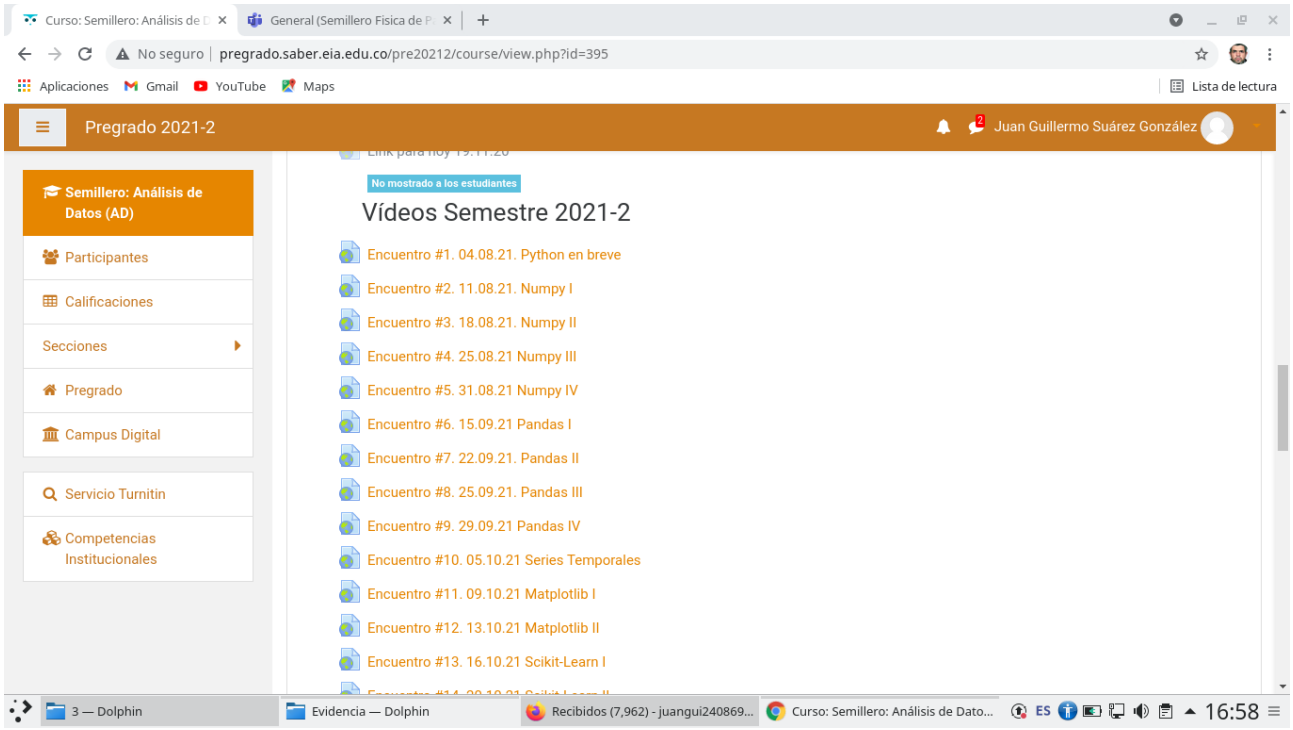
2. Materiales complementarios.



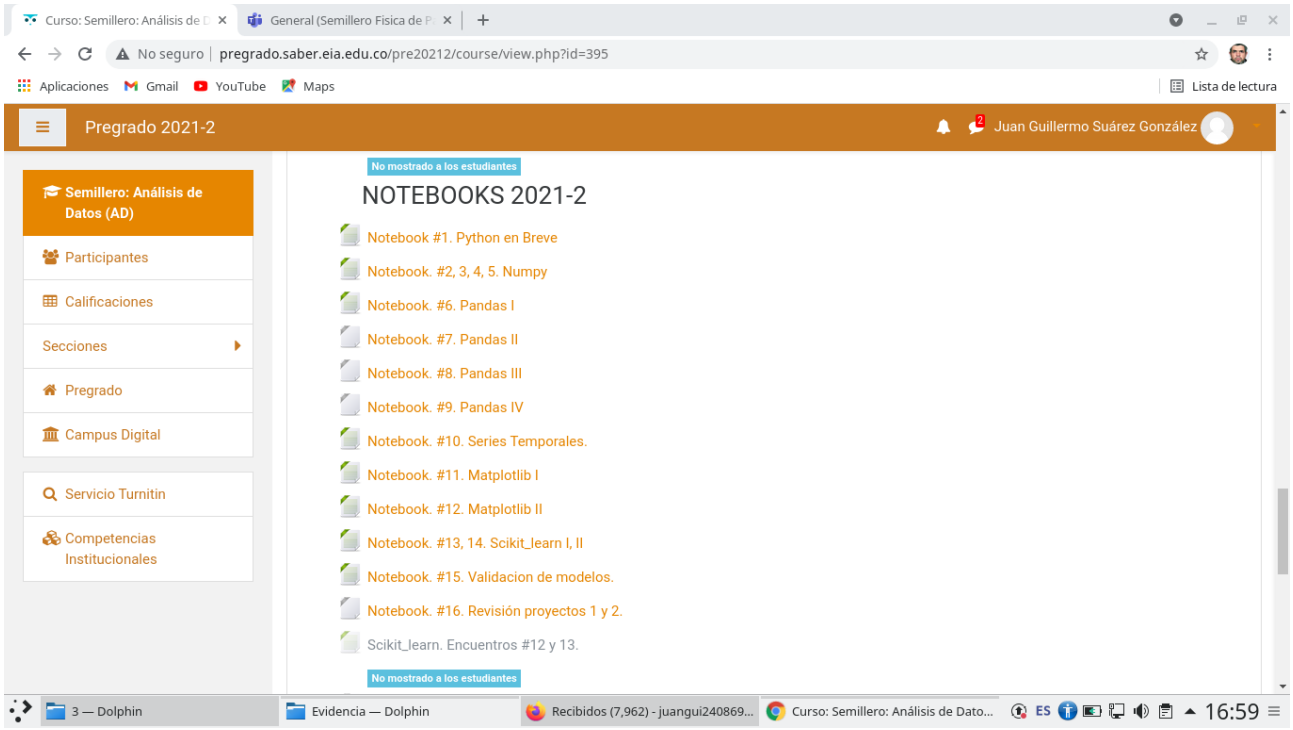
3. Proyectos.



4. Vídeos.



5. Notebooks 2021-2



Certificado.

Para obtener el certificado del semillero es necesario participar en las actividades del semillero durante dos semestres consecutivos y presentar los respectivos proyectos.

Resultados.

Logramos completar en su totalidad el plan propuesto al inicio del semillero. Durante el semillero realizamos dos proyectos de investigación: el primero dedicado al análisis exploratorio de datos y el segundo dedicado al análisis inferencial de datos. Para los proyectos utilizamos la base de datos **UCI_Credit_Card.csv.** Esta base de datos contiene información sobre el historial de pagos de las tarjetas de crédito en Taiwan durante abril – septiembre 2005. El objetivo del proyecto 2 consistía en predecir si un cliente estaría o no en condición de cumplir con las obligaciones de pago de sus cuotas de tarjeta de crédito en meses posteriores con base en los pagos hechos en meses anteriores. Se utilizaron varios algoritmos para las predicciones. La precisión alcanzada fue del 80% en promedio. El cambio de algoritmo permitió reducir el número de falsos negativos (clientes que no cumplieron con sus obligaciones pero fueron clasificados por los algoritmos como personas al orden del día). Los resultados de los proyectos están detallados en un archivo que se agrega a la evidencia del semillero.

Conclusiones.

Al finalizar el semillero logramos:
1. considerar todos los temas planeados,
2. realizar algunas tareas para la asimilación del material expuesto,
3. realizar dos proyectos de investigación con aplicación de los temas estudiados,
4. para el próximo semestre planeamos continuar con el trabajo aplicando métodos estadísticos en computación con ayuda de los programas Scipy y Statsmodels.