

Ciclo de Ciencia de Datos - CCD

PITEI - Comunicación Digital y Big Data

La Bernalesa

- La propuesta de este **Ciclo de ciencia de datos - CCD** - ofrece un conjunto de espacios de aprendizaje que permiten articular trayectorias formativas continuas y trabajar competencias específicas en formatos acotados, buscando ofrecer a los y las estudiantes herramientas que mejoren sus posibilidades de insertarse en el mundo del trabajo de la sociedad del conocimiento.
- El objetivo principal del **CCD** se orienta a contribuir con la democratización de la oferta formativa en el campo del **análisis y la ciencia de datos**, diversificando la propuesta con diversas actividades y cursos.
 - **Divulgación y debate** en el espacio académico y ciudadano.
 - Aprendizaje de habilidades específicas vinculadas con formación profesional y tecnológica para la **analítica de datos** – bases, consistencia, limpieza, transformación, procesamiento y visualización de información.
 - Formación en programación con **Python** para el análisis de datos –sintaxis, principales librerías para la limpieza, orden, clasificación, graficación y visualización de datos
 - Capacitación, dentro de la perspectiva de la inteligencia artificial y el aprendizaje automático, en **Gestión de datos para la toma de decisiones**, orientada a equipos que ya trabajan con datos, promoviendo la cooperación entre experiencias de investigación, y buscando enriquecer las prácticas en los espacios académicos e institucionales del ámbito público.

Ejes didácticos y cursos CCD

- El CCD reúne sus cursos alrededor de **tres ejes formativos**, que si bien no reconocen límites estrictos entre sí, permiten estructurar junto con **cursos complementarios**, una propuesta formativa integral y diversificada.
- Para 2022, se ofrecerán un total de **6 cursos**, y se llevará adelante entre los meses de octubre y diciembre, mientras que la segunda tendrá lugar entre los meses de abril a julio del próximo año.
- La totalidad de sus cursos se ofrecerán en modalidad virtual sincrónica, con actividades asincrónicas y seguimiento por plataforma por parte de los docentes/instructores. Se prevé un encuentro de cierre presencial (no obligatorio) en la sede de La **Escuela de Educación Profesional de La Bernalesa**.
- En 2023, se sumarán los **niveles avanzados** que proyectan trayectorias formativas, para quienes certificaron el primer nivel y se abrirán nuevas cohortes para los niveles iniciales.

Modalidad para la gestión del CCD

- Para acceder a la oferta de cursos del **CCD** lxs postulantes deberán responder un cuestionario de pre inscripción, brindando información de base sobre trayectorias formativas y saberes previos, antecedentes laborales y recursos TICs disponibles. A partir de los datos obtenidos la coordinación del Ciclo ofrecerá orientación para la inscripción a los distintos cursos del ciclo.
- Quienes se hayan preinscripto deberán asistir al menos a dos de los tres **conversatorios de divulgación** inicial para estar habilitados para la inscripción a cursos.
- Lxs estudiantes tomarán sus cursos vía la **plataforma moodle** de la Secretaría de Extensión de la UNQ, para lo cual deberán previamente inscribirse mediante el sistema de gestión académica **SIU Guaraní**.
- Lxs docentes instructores a cargo de los cursos utilizarán la plataforma para el desarrollo de los mismos, utilizando la diversidad de recursos didácticos disponibles – Salas para encuentros sincrónicos, lecciones, foros, cuestionarios, contenidos en formato digital, videos.

Ejes didácticos y cursos

2022

Conversatorios de apertura

■ Eje Analítica de datos

Introducción a la Analítica de datos

■ Eje Competencias computacionales

Introducción a Python

■ Eje Aprendizaje automático / inteligencia artificial

Gestión de datos para la toma de decisiones. Nivel I

■ Cursos complementarios

Sistemas de información Geográfica (GIS)

Big Data & Social Analytics

Procesamiento y Análisis de datos con SPSS

Cursos de nivel avanzado
(II y III) en 2023

Cronograma de actividades 2022

 Cursos Eje
Analítica de datos

 Cursos Eje Competencias
computacionales

 Cursos Eje Aprendizaje automático
/ inteligencia artificial

 Cursos
complementarios

	26-sep	3-oct	10-oct	17-oct	24-oct	31-oct	7-nov	14-nov	21-nov	28-nov	5-dic
Difusión /Propuesta Preinscripción	X										
Conversatorios / Orientación e inscripción a cursos		X									
Introducción a la analítica de datos			X	X	X	X	X	X	X	X	X
Gestión de datos para la toma de decisiones I				X	X	X	X	X	X		
Introducción a Python para el análisis de datos			X	X	X	X	X	X			
Big Data & Social Analytics				X	X	X	X	X			
Sistemas de información Geográfica (GIS)				X	X	X	X	X			
Procesamiento de datos SPSS					X	X	X	X	X		

Días y horarios oferta 2022

 Cursos Eje
Analítica de datos

 Cursos Eje Competencias
computacionales

 Cursos Eje Aprendizaje automático
y la inteligencia artificial

 Cursos
complementarios

	Lunes	Martes	Miércoles	Jueves	Viernes
Introducción a la analítica de datos		De 18 a 19hs. Práctico Comisión I De 19 a 20 hs. Práctico Comisión II		De 18 a 20 hs. Teórico Comisión I y II	
Gestión de datos para la toma de decisiones I	Cohorte cerrada para equipos institucionales del ámbito público				
Introducción a Python para el análisis de datos		Teórico / Práctico Comisión I 19 a 21hs		Teórico / Práctico Comisión II 19 a 21hs	
Big Data & Social Analytics	Teórico / Práctico Comisión I 19 a 21hs		Teórico / Práctico Comisión II 19 a 21hs		
Sistemas de información Geográfica (GIS)			Teórico / Práctico Comisión I 18.30 a 20.30hs		Teórico / Práctico Comisión II 18.30 a 20.30hs
Procesamiento de datos SPSS	Teórico / Práctico Comisión I 18 a 20hs				

Divulgación

Debates académicos y ciudadanos sobre los datos y la sociedad del conocimiento



Conversatorios de Divulgación

Lunes 3 de
Octubre
18 hs

Análisis de redes aplicado a la investigación del crimen organizado

- El análisis reticular y la antropología sociocultural
- Entrevista a Germán Silva – Analista de datos
Coordina: Marcela Grinszpun

Miércoles 5 de
Octubre
19 hs

El análisis de datos aplicado al deporte

- La experiencia profesional y los datos aplicados en las principales ligas profesionales de fútbol
- Entrevista a Matías Conde – Analista de datos
• Coordina: Emilio Cabello

Jueves 6 de
Octubre
18 hs

La era de los datos y las transformaciones digitales

- Presentación del libro del Dr. Sergio Paz La economía digital El futuro ya llegó!
- Coordina: Germán Reynolds

Quienes se hayan preinscripto al CCD deberán asistir al menos a dos de los tres **conversatorios de divulgación** para estar habilitados a la inscripción a cursos.

Analítica de datos

El oficio para la captura, consistencia, transformación y visualización de datos



Principales contenidos. Nivel I

- **Fundamentos del Análisis de datos.** Bases de Datos y datos Estructurados. Obtención, limpieza y transformación de los datos; procesamiento y presentación de datos BI.
Herramientas: SQL, OpenRefine, PowerBI.
- **Fundamentos de Estadística I.** Debates epistemológicos actuales. Conceptos de población y de unidad de análisis. Censos y muestras. Set de datos. Resumiendo los Datos: Medidas de tendencia central, dispersión y posición. Gráficos estadísticos. *Herramientas: Excel, SPSS.*
- **Visualización de Datos.** Fundamentos de diseño para no diseñadores: principios que deben seguir las visualizaciones y los reportes. De los informes operativos a los tableros estratégicos. Storytelling.
Herramientas: Tableau, Power Bi, Google Data Studio, Datawrapper y Flourish

Principales contenidos. Nivel II

- **Pre-procesamiento y almacenamiento de Datos.** Datos abiertos, scrapping, datos privados, encuestas, datos internos, enmascaramiento de la información. Buenas prácticas en la construcción de BBDD. Tipos de conexiones entre tablas. Consultas en lenguaje SQL. Importación y exportación de datos. *Herramientas: SQL Management Studio.*
- **Modelos alternativos para el análisis de datos.** Modelos mecánicos, estadísticos, sistémicos e interpretativos. Modelos predictivos. Las simulaciones como posible superación de la dicotomía entre el modo inductivo y el modo deductivo. Sistemas complejos: Autómatas Celulares y Modelos Basados en Agentes. Modelo Reticular y Análisis de Redes sociales. Principales métricas. *Herramientas: Mirek Celebration, Netlogo, Gephi.*
- **Fundamentos de Estadística II.** Probabilidad básica y funciones de probabilidad. Coeficientes de correlación y contrastación de hipótesis. Regresión lineal. Gráficos multivariados.

Análítica de datos – Equipo docente

Equipo docente



Germán Silva

- **Antropólogo**, experto en **metodologías de investigación** formales, **analista de datos**, analista estadístico. Profesor de grado y postgrado universitario.
- Germán es antropólogo con orientación sociocultural, docente de la materia Introducción a la ciencia de los datos en el Departamento de Economía y Administración de la UNQ.
- Actualmente se desempeña como analista de datos, especialista en metodologías formales de investigación aplicadas para la Corte Suprema de Justicia de la Nación.



Marcela Grinszpun

- Lic. en Sociología (UBA). Maestría en Generación y Análisis de Información Estadística (UNTREF)
- Docente investigadora en Metodología de la Investigación y Estadística aplicada (UBA/UNQ)
- Marcela se desempeña desde hace mas de 15 años como docente investigadora, evaluadora de proyectos y consultora especializada en análisis cuantitativo tanto en el ámbito privado como en proyectos de organismos oficiales nacionales e internacionales .
- Sus principales competencias se orientan al diseño de componentes en estrategias teórico metodológicas cualitativas y cuantitativas, en instrumentos de recolección de información, y es experta en procesamiento de datos y elaboración de informes/presentaciones.



Matías Conde

- Analista de datos. Especialista en Investigación de Mercado, Inteligencia de Negocios y Análisis de Redes Sociales, con formación en Antropología y Data Science.
- Matías tiene mas diez años de experiencia como analista para consultoras de investigación y opinión pública, como también procesando datos, generando contenidos y realizando informes sobre eventos deportivos para diferentes equipos profesionales y medios de comunicación
- <https://wakelet.com/@MatiasConde>
- <https://twitter.com/mmcr85>

Python

Introducción a Python para análisis de datos



Principales contenidos. Nivel I

- **Instalación y armado del entorno de desarrollo y librerías.** Directorio de trabajo y proyectos. Documentación del lenguaje y su sintaxis básica. Diferencias entre los entornos de programación. Librerías. Herramientas del paquete Anaconda, distintos IDEs. Importación de datos.
- **Limpieza y clasificación de datos.** Funciones básicas para la realización de la limpieza y clasificación de los datos para el análisis. Pandas y Numpy. Series y Data Frames, carga de datos, manipulación y estadísticas.
- **Visualización de datos.** Librerías Matplotlib y Seaborn. Principales variables y gráficos, como decidir cual es la mejor visualización para los resultados que se quieren mostrar. Gráficos dinámicos y estáticos.

Principales contenidos. Nivel II

- **Fuentes convencionales.** Carga de datos a partir de fuentes de texto plano (CSV, txt, tsv). Excel y bases de datos relacionales (SQL). Imputación. Conversión de tipo de datos en la lectura. Limpieza y manipulación básica de datos.
- **Interacción con APIs.** Request. Capturando y parseando utilizando JSONPath. APIs de datos abiertos de Argentina para la importación de datos. APIs de Google. Maps: Reverse geocoding.
- **Webscrapping.** Introducción a HTML. Nuevas fuente de datos. XPath y selectores. Como encontrar lo que quiero extraer. Diferencias entre las librerías Scrapy, BeautifulSoup y Selenium
- **Text mining.** Tokenización de texto. Frecuencias de palabras. Stemming. Lemmatization. Stopwords Análisis de sentimiento.

Python – Docente



Gestión de datos para la toma de decisiones

Probabilidades, funciones y modelos predictivos



Principales Contenidos. Nivel I

Módulo 1 – Gestión de datos para la toma de decisiones

El objetivo de este módulo es que los participantes apropien el concepto de Organizaciones Guiadas por Datos (DDO), sus componentes básicos, los roles necesarios y tengan ejemplos concretos de las herramientas utilizadas y los resultados que se obtienen.

Datos 2020. Evolución de los sistemas de información y disponibilidad de datos para la toma de decisiones. Inteligencia Artificial y Aprendizaje Automático como herramientas para el descubrimiento de conocimiento en grandes cantidades de datos. Inteligencia de Negocios: ejemplo el reporting moderno en las organizaciones. Analítica predictiva: ejemplo de modelos basados en evidencia. Data Warehousing Analítico: para asegurar la integridad, calidad, seguridad y disponibilidad de datos. Organizaciones Guiadas por Datos (DDO), su estructura, roles y herramientas. Evaluación y cierre del módulo.

Conocimientos previos requeridos: Manejo de planillas de cálculo, facilidad para navegar por Internet, lectura básica de textos técnicos en inglés. Es deseable el conocimiento de programación y lenguajes de consulta a base de datos.

Principales contenidos. Nivel II

Módulo 2 – Conceptos básicos de analítica de datos

El objetivo de este módulo es que los participantes apropien la metodología de trabajo y algunas herramientas básicas de la analítica avanzada en las organizaciones.

Introducción al aprendizaje automático., Introducción a las herramientas de analítica: construcción de tableros y visualización de datos. Introducción a las herramientas de analítica: modelos predictivos. Métodos de predicción basados en reglas. Métodos de predicción basados en probabilidades. Métodos de predicción basado en similitudes. Métodos de ensamble. Marco Metodológico para la analítica de datos. Gestión de datos para analítica. Calidad de datos para analítica. Metaalgoritmos. Evaluación y cierre del módulo.

Principales contenidos. Nivel III

Módulo 3 – Taller de analítica de datos

El objetivo de este módulo es que los participantes profundicen y apropien los conceptos y herramientas trabajados en los módulos 1 y 2 en el desarrollo de proyectos acotados pero de su propio interés.

Diseño de casos. Métodos de segmentación. Asistencia para el desarrollo del caso.
Introducción al Text Mining. Asistencia para el desarrollo del caso. Introducción a uso de Grafos. Asistencia para el desarrollo del caso. Asistencia para el desarrollo del caso.
Presentación de desarrollos y cierre del curso.

Gestión de datos para la toma de decisiones



Gustavo Poggi

Licenciado en Ciencias de la Computación por la UBA (FCEN), magister en Administración y Políticas Públicas . Especialización en Negocios y Tecnología por la Universidad de San Andrés.

Gustavo es actualmente asesor en la Dirección de Analítica de Datos de AFIP. Docente de posgrado y consultor internacional. orientado a la gestión de datos públicos. Aprendizaje Automático, Minería de Datos y gestión de TI Pública en general.



Coppolillo, Néstor Raúl

Licenciado en Ciencias de la Computación por la UBA (FCEN), Especialista en Explotación de Datos y Descubrimiento del Conocimiento. FCEyN UBA (2021)

Néstor es especialista en gestión de datos para la toma de decisiones en temáticas tributarias y aduaneras, con una amplia trayectoria docente en ciencia de datos, datamininig y Python.

Cursos complementarios

Big data & Social analytics. Geo referenciación. SPSS



Principales contenidos.

- **Big Data y ciencias sociales.** Concepto de Big Data. ¿Qué incluye? ¿Qué se dice y qué se hace? Aplicaciones reales: Análisis sociológicos. Marketing. Opinión pública. Ciencia Política.
- **Google trends.** Los motores de búsqueda como herramienta métrica y de análisis. Sistematización de datos. Análisis de picos de búsqueda y graficación Comparación entre distintos términos. Wordclouds.
- **Open data / datos abiertos.** ¿Qué son los datos abiertos? Fuentes oficiales. Datos en tablas (.csv; tablas en excel), datos geolocalizables (shapefiles y jsons), datos no tabulados (pdfs con tablas pegadas). APIs - datasets - paquetes en R.
- **Web scraping y análisis de sentimiento.** Las nuevas herramientas y fuentes de datos para análisis sociales. El procesamiento del lenguaje natural. Límites y oportunidades. Estudios de casos. Nociones básicas de HTML
- **Social media listening y análisis de redes.** Estudios de mercado, community manager e investigación de problemáticas sociales. API de Twitter y Facebook.

Big data & Social analytics - Docentes



Principales contenidos.

- **Conceptos y componentes básicos del SIG.** Datos geográficos y representación espacial. Vínculos entre datos y mapas. Proyecciones. Topología. Vista, Tablas y Mapa. Capas, datos y propiedades.
- **Capas vectoriales y raster.** Diferencia entre SIG y CAD. Propiedades de la capa. Rango de escalas. Hiperenlace. Clasificación de la información temática. Etiquetado. Búsqueda y consulta. Localizador por atributo. Capa vectorial. Georreferenciación y puntos de control.
- **Administración de tablas y datos.** Propiedades de las tablas. Herramientas asociadas a tablas. Estadísticas. Filtros. Unión y enlace. Calculadora de campos. Edición de tablas interna y externa.
- **Edición gráfica de los mapas.** El área de dibujo. Sesión de edición. Procedimientos para la entrada de órdenes. Digitalización de información. Exportar e importar un SHP – KML. Propiedades de las capas temáticas. Propiedades del mapa. Administración de elementos cartográficos en un mapa. Operaciones con gráficos. Propiedades de un elemento insertado.

GIS - Docentes



Principales contenidos. Nivel I

- **Preparación de los datos.** El proceso de investigación. Concepto de matriz de datos, unidad de análisis, variables, categorías y niveles de medición. Interfaces y vistas en SPSS. Editor de Sintaxis y visor de Resultados.
- **Creación de nuevas variables.** Creación de nuevas variables como parte del proceso de análisis. Comandos recodificar, contar y calcular.
- **Estadísticas descriptivas.** Análisis univariado. Frecuencias, Medidas de tendencia central, posición y variabilidad. Análisis bivariado. Hipótesis y relaciones entre variables. Tablas personalizadas.
- **Unidades de Análisis individuales y colectivas.** Fusión de BBDD (variables/casos). De unidades individuales a colectivas. Reestructuración y agregación de variables.



Marcela Grinszpun

- Lic. en Sociología (UBA). Maestría en Generación y Análisis de Información Estadística (UNTREF)
- Docente investigadora en Metodología de la Investigación y Estadística aplicada (UBA/UNQ)
- Marcela se desempeña desde hace más de 15 años como docente investigadora, evaluadora de proyectos y consultora especializada en análisis cuantitativo tanto en el ámbito privado como en proyectos de organismos oficiales nacionales e internacionales .
- Sus principales competencias se orientan al diseño de componentes en estrategias teórico metodológicas cualitativas y cuantitativas, en instrumentos de recolección de información, y es experta en procesamiento de datos y elaboración de informes/presentaciones.

- **Dirección ejecutiva.**

Raúl Di Tomaso. Secretario de Extensión UNQ. Director PITEI - Comunicación Digital y Big Data.

- **Coordinación programática.**

Emilio Cabello. Coordinador ESEP – La Bernalesa.

Germán Reynolds. Director PITEI - La Universidad en la Era Digital. Aportes para la transformación de procesos educativos, de investigación y de gestión universitaria.

- **Observatorio ESEP La Bernalesa**

Marcela Grinszpun. Coordinadora Observatorio

- **Gestión administrativa**

CCD

2022/2023