



"2024 - 40 años de la Reapertura de la Universidad Nacional de Luján y 30 años del Reconocimiento Constitucional de la Autonomía Universitaria"



Universidad Nacional de Luján  
Departamento de  
Ciencias Básicas

DISPOSICION CONSEJO DIRECTIVO DEPARTAMENTAL DE CIENCIAS BÁSICAS DISPCD-CB : 338 / 2024

LUJAN, 9 DE SEPTIEMBRE DE 2024

VISTO: El programa de la asignatura Introducción a las Ciencias de Datos (11281) para la carrera Analista Universitario en Ciencias de Datos presentado por la División Estadística; y

CONSIDERANDO:

Que la Comisión Plan de Estudio ha tomado intervención en el trámite.

Que ha sido tratado y aprobado por el Consejo Directivo Departamental de Ciencias Básicas en su Sesión Ordinaria del día 5 de septiembre de 2024.

Por ello,

EL CONSEJO DIRECTIVO DEPARTAMENTAL  
DE CIENCIAS BÁSICAS

D I S P O N E :

ARTÍCULO 1°.- Aprobar el programa de la asignatura Introducción a las Ciencias de Datos (11281) para la carrera Analista Universitario en Ciencias de Datos presentado por la División Estadística que como anexo forma parte de la presente Disposición.-

ARTICULO 2°.- Establecer que el mismo tendrá vigencia para el año 2024.-

ARTÍCULO 3°.- Regístrese, comuníquese, cumplido, archívese.-

Lic. Ariel H. REAL - Secretario Académico - Departamento de Ciencias Básicas

Lic. Emma L. FERRERO - Directora Decana - Departamento de Ciencias Básicas

---

**DENOMINACIÓN DE LA ACTIVIDAD:** 11281 – INTRODUCCION A LAS CIENCIAS DE DATOS

**TIPO DE ACTIVIDAD ACADÉMICA:** Asignatura.

**CARRERAS:** ANALISTA UNIVERSITARIO EN CIENCIAS DE DATOS (20)

**PLAN DE ESTUDIOS:** 20.01

---

**DOCENTE RESPONSABLE:**

REAL ARIEL HERNAN – PROFESOR ADJUNTO

**EQUIPO DOCENTE:**

ACUÑA, Pablo	Ing. Industrial	Jefe de Trabajos Prácticos
BAQUEL, Rubén	Lic. en Estadística	Jefe de Trabajos Prácticos
GENOVA, Cristian Ezequiel	Lic. En Sistemas de Información	Jefe de Trabajos Prácticos

---

**ACTIVIDADES CORRELATIVAS PRECEDENTES:**

Para el Plan de Estudio: 20.01

PARA CURSAR:

11271 – INTRODUCCIÓN A LA PROGRAMACIÓN

11280 – ORGANIZACIÓN DE LA INFORMACIÓN

PARA APROBAR:

11271 – INTRODUCCIÓN A LA PROGRAMACIÓN

11280 – ORGANIZACIÓN DE LA INFORMACIÓN

**CARGA HORARIA TOTAL:** HORAS SEMANALES: 6 - HORAS TOTALES 96

**DISTRIBUCIÓN INTERNA DE LA CARGA HORARIA:**

TEÓRICOS 40%, 38 HORAS, PRESENCIALES

PRÁCTICAS 60%, 58 HORAS, PRESENCIALES

PERÍODO DE VIGENCIA DEL PRESENTE PROGRAMA: 2024
---

---

**CONTENIDOS MÍNIMOS O DESCRIPTORES**

Ciclo de vida de un proyecto de ciencias de datos. Formulación de preguntas relevantes en ciencia de datos. Conceptos formales de Machine Learning, Big Data y Deep Learning. Exploración de datos: técnicas básicas y herramientas iniciales de exploración, procesamiento y visualización de los datos.

---

**FUNDAMENTACIÓN, OBJETIVOS, COMPETENCIAS**

Esta asignatura fue concebida con el propósito de introducir al futuro graduado en las distintas etapas del ciclo de vida de un proyecto de ciencias de datos orientado a la toma de decisiones en interacción con profesionales de otras áreas de conocimiento. Para ello será necesario conocer las herramientas estadísticas para la exploración, procesamiento y visualización de datos, introduciendo algunos modelos estadísticos simples y algoritmos de programación en R, que debidamente aplicados permitirán construir conocimiento a partir de un conjunto de datos. Por último, se brindarán a nivel introductorio conceptos claves en las ciencias de datos, como Machine Learning, Big Data y Deep Learning. El enfoque tomado en esta asignatura no es puramente matemático, es decir, sin dejar de lado la formación académica profesional y considerando la complejidad de los fenómenos actuales que se quieran abordar, se busca dar las bases conceptuales del pensamiento lógico-inductivo, que le permitan al futuro profesional resolver con juicio crítico, problemas relacionados a la variabilidad en los datos y sus posibles explicaciones.

**OBJETIVOS GENERALES:**

Al completar el curso el alumno podrá: Conocer las etapas involucradas en el ciclo de vida de un proyecto de ciencia de datos. Podrá identificar variables relevantes y sus escalas de medición; adquirirá habilidades en el uso del software R que le permitan ordenar, clasificar y presentar información univariada y bivariada; realizar análisis exploratorio de datos e interpretarlos. Se espera también que pueda identificar valores atípicos o outliers y decidir sobre la estrategia más apropiada para su tratamiento. También podrá estimar un modelo de regresión lineal simple como herramienta de estimación de una variable dependiente o respuesta, en función de otra independiente. Finalmente, se espera que pueda reconocer las distintas técnicas a utilizar en la profesión de Analista Universitario en Ciencia de Datos, comprendiendo sus principales ventajas y desventajas, sin entrar aún en profundidad en la teoría que las sustentan, ni en su implementación.

**CONTENIDOS**

**UNIDAD 1**

El concepto de Ciencia. Acercamiento a una definición de Ciencia de Datos. El perfil del científico de datos y su evolución hasta hoy. Fundamentos de la ciencia de datos. Perfiles en un equipo de ciencia de datos (data-driven) Características de una compañía basada en datos. Ciclo de vida de un proyecto de ciencia de datos.

**UNIDAD 2**

Modelado. Distintas propuestas de solución para los mismos problemas. Evaluación del modelo. Tipologías de tareas a abordar con los modelos. Entrenamiento y Test. Evaluación de resultados. Ejemplos de aplicación exitosos.

**UNIDAD 3**

Exploración de Datos. Visualización. Componentes que no deben faltar en un gráfico. Propuestas gráficas según tipo de variables. Gráficos para variables cualitativas: circulares, de barras adyacentes o superpuestas, otros. Gráficos para variables cuantitativas: histogramas de frecuencias, polígonos, histogramas de frecuencias acumuladas, ojivas.

#### UNIDAD 4

Exploración de Datos. Análisis exploratorio de datos. Tipos de datos, clasificación y escalas de medición. Estructura de datos, características. Conceptos de registro, variables, tuplas, etc. Organización de los datos en tablas de frecuencias (no agrupadas o agrupadas). Concepto de distribución. Cálculo de medidas de resumen: medidas de posición central tradicionales: media, mediana, moda. Medidas de posición complementarias: cuartiles, deciles, percentiles. Gráfico Boxplot.

#### UNIDAD 5

Exploración de Datos. Concepto e importancia de la variabilidad en los datos. Principales medidas para su estudio: rango, varianza, desvío estándar, rango intercuartil, desvío absoluto mediana (MAD). Estudio de la distribución de datos según su forma: asimetría y curtosis. Concepto de normalidad, análisis de normalidad en nuestros datos. Datos outliers, concepto e impacto en el análisis de datos. Medidas de posición central no afectadas por los datos outliers. Datos faltantes o inconsistentes: análisis y tratamientos recomendados.

#### UNIDAD 6

La búsqueda de explicación al comportamiento de nuestras variables. Relación entre dos variables cualitativas: tablas de doble entrada, cálculo de frecuencias relativas por filas o por columnas. Gráficos de mosaicos. Relación entre una variable cuantitativa y otra cualitativa: cálculo de medidas por cada nivel de la variable cualitativa, gráficos comparativos (boxplot comparativos, gráficos de perfiles, etc). Otros gráficos utilizados: curvas de nivel, gráficos de radar, caras de Chernoff. Relación entre dos variables cuantitativas: medidas de asociación - correlación, la matriz de correlaciones, gráfico de dispersión. Análisis de Regresión Lineal Simple.

#### UNIDAD 7

Introducción al lenguaje R y a R-Studio. Paneles, paquetes, instalación de un nuevo paquete en R-Studio. Conceptos de data frame, vectores, matrices, factores, listas. Tipos de datos en R. Funciones para ordenar. Operaciones con vectores y matrices. Importar datos, exportar datos. Guardado de Script. Creación de una función. Expresiones condicionales. Loops. Librería Tidyverse. El paquete Tidyverse y las herramientas disponibles para facilitar el análisis exploratorio de datos. Herramientas para la manipulación, importación, exploración y visualización de datos. Paquetes que lo componen: readr; dplyr; ggplot2; tibble; tidyr; purr; stringr; forcats. Casos de aplicación.

---

### METODOLOGÍA

La modalidad de trabajo adoptada son las clases teórico – prácticas. Para ello:

- a) La asignatura se organiza a través del Aula Virtual diseñada en la plataforma E-ducative.
  - b) El material de estudio estará a disposición del estudiante por Unidad temática.
  - c) En el aula se incluye un cronograma de clases, que le permitirá al estudiante tener el seguimiento de los temas a desarrollados y por desarrollar.
  - d) Las consultas de los estudiantes fuera de los horarios de clases serán atendidas a través de foros de discusión por unidad, disponibles en el Aula Virtual de la asignatura.
  - e) Para la aprobación de la asignatura se deberán aprobar un parcial y un trabajo final a presentar en clase pública (el cual se desarrollará con seguimiento semanal).
-

---

**REQUISITOS DE APROBACIÓN Y CRITERIOS DE CALIFICACIÓN:**

**CONDICIONES PARA PROMOVER (SIN EL REQUISITO DE EXAMEN FINAL)**

DE ACUERDO AL ART.23 DEL RÉGIMEN GENERAL DE ESTUDIOS RESHCS 261-21

Las condiciones para que el estudiante pueda acceder a la modalidad de promoción, es decir aprobar sin el

requisito de examen final, son:

- a) Tener aprobadas las actividades correlativas al finalizar el turno de examen extraordinario de ese cuatrimestre.
- b) Cumplir con el porcentaje mínimo de asistencia que para cada actividad se fije en el programa vigente para ese año académico.
- c) Aprobar todos los trabajos prácticos, monografías, trabajos de campo y/o actividades académicas especiales previstas en el programa, pudiendo recuperarse hasta un veinticinco por ciento (25%) del total por ausencias o aplazos.
- d) Aprobar el cien por ciento (100%) de las evaluaciones previstas con un promedio final no inferior a seis (6) puntos, sin haber recuperado ninguna.
- e) Aprobar una evaluación integradora de la asignatura con calificación no inferior a siete (7) puntos. Esta evaluación puede ser el último parcial en caso en que la asignatura tenga como modalidad acumular los contenidos del programa en los exámenes parciales y el último integre a todos.

**CONDICIONES PARA APROBAR COMO REGULAR (CON REQUISITO DE EXAMEN FINAL)**

DE ACUERDO AL ART.24 DEL RÉGIMEN GENERAL DE ESTUDIOS RESHCS 261-21

Las condiciones requeridas para que el estudiante pueda obtener la condición de regular, debiendo aprobar la asignatura con examen final son:

- a) Estar en condición de regular en las actividades correlativas al momento de su inscripción al cursado de la asignatura.
- b) Cumplir con el porcentaje mínimo de asistencia fijado para cada actividad en el programa de la asignatura, vigente para ese año académico.
- c) Aprobar todos los trabajos prácticos, monografías, trabajos de campo, y/o actividades académicas especiales previstas en el programa vigente, pudiendo recuperar hasta un cuarenta por ciento (40%) por ausencia o aplazos.
- d) Aprobar todas las evaluaciones con una calificación no inferior a cuatro (4) puntos, pudiendo recuperar hasta el cincuenta por ciento (50%) de las mismas. Cada evaluación solo podrá recuperarse en una oportunidad.

Cumplidos los requisitos de correlatividades y los establecidos en los incisos anteriores, para aprobar la asignatura, el estudiante deberá presentarse a rendir examen final en condición de regular, con el programa vigente al momento de cursado.

**EXÁMENES PARA ESTUDIANTES EN CONDICIÓN DE LIBRES**

- a) Aquellos estudiantes que habiéndose inscripto oportunamente en la presente asignatura hayan quedado en condición de libres por aplicación de los artículos 22, 25, 27, 29 o 32 del Régimen General de Estudios, podrán rendirla en tal condición.
- b) Aquellos estudiantes que no cursaron la asignatura y se presenten en condición de libre, también podrán rendirla en esta condición.
- c) Las características del examen libre son las siguientes: se deberán rendir dos instancias, una teórica la que deberá aprobarse con el 60% de sus respuestas correctas, y otra práctica, la que también deberá ser aprobada con el 60% de sus respuestas correctas. Para acceder a la segunda instancia debe aprobarse la primera. La nota final del examen corresponderá a la nota de la última instancia rendida.

## **BIBLIOGRAFÍA**

### **BIBLIOGRAFÍA OBLIGATORIA:**

AGUIRRE, C. N., MIÑO, M. Fernanda., & SIMONETTI, E. F. (2005). Estadística aplicada a las Ciencias Sociales y Humanas (Editorial Universitaria de Misiones, Ed.).

ARANEDA, Patricio. (2021, April 27). Tidyverse para Data Análisis. <https://rpubs.com/paraneda/tidyverse>

DEL CAMPO Neira, P. César. (n.d.). Introducción a lenguaje de programación R y R-Studio. Retrieved August 3, 2024, from Introducción a lenguaje de programación R y Rstudio

DIEGO, I. M. D. E., & FERNÁNDEZ Isabel, A. (2020). Ciencia de datos para la ciberseguridad. RA-MA Editorial. <https://elibro.net/es/lc/unlu/titulos/222714>

LIND Douglas, MARCHAL William, & WATHEN Samuel. (2012). Estadística aplicada a los negocios y la economía.

MALDONADO, S. (2022). Analytics y Big Data: ciencia de los Datos aplicada al mundo de los negocios. RIL editores. <https://elibro.net/es/lc/unlu/titulos/225562>

MEDINA LA PLATA, E. H. (2023). Big data: los datos como generadores de valor. Universidad Peruana de Ciencias Aplicadas (UPC). <https://elibro.net/es/lc/unlu/titulos/231338>

MENOYO ROS D., GARCIA LOPEZ E., GARCIA CABOT A. Fundamentos de la ciencia de datos. Editorial Universidad de Alcalá. Año 2021.

MINGUILLÓN, J., Casas, J., & MINGUILLÓN, J. (2017). Minería de datos: modelos y algoritmos. Editorial UOC. <https://elibro.net/es/lc/unlu/titulos/58656>

TIDYVERSE. (n.d.). R packages for data science. Retrieved August 3, 2024, from <https://www.tidyverse.org/>

### **BIBLIOGRAFÍA COMPLEMENTARIA:**

ANDERSON, David R – SWEENEY Denis J. – WILLIAMS Thomas A. -. Estadística para Administración y Economía. Cengage Learning Editors. I.S.B.N. 13-978-970-686-825-1 y I.S.B.N. 10-970-686-825-9 10ª edición. Año 2011.

JOHNSON Robert. – KUBY Patricia. Estadística Elemental. Lo Esencial. Cengage Learning Editors. I.S.B.N. 13-978-607-481-855-0 11ª edición. Año 2012. Capítulo 3.

---

### **NÓMINA COMPLEMENTARIA DEL EQUIPO DOCENTE:**

No corresponde.

---

DISPOSICIÓN DE APROBACIÓN: CD[A COMPLETAR POR EL DEPARTAMENTO]

## Hoja de firmas