

**Facultad de Economía y Negocios / Ingeniería en Control de Gestión,
Mención Ciencia de Datos**

Nombre de la actividad curricular:	Proyecto Análisis de Datos Estadístico II – P-R
Código:	
Créditos:	10
Carácter:	Obligatorio
Prerrequisitos:	Estadística Modelos Regresión Múltiple - R y Métodos de Programación II (PI)
Tipo:	Laboratorio
Horas cronológicas de dedicación:	Docencia directa: 3 Trabajo autónomo: 7

I. DESCRIPCIÓN

El objetivo del presente curso de análisis de datos estadístico avanzado, es fortalecer las capacidades de los estudiantes en la programación con el lenguaje Python empleando las diversas bibliotecas de funciones estadísticas que este posee.

El curso permitirá al estudiante aplicar los diversos conceptos de estadísticas descriptiva, inferencial y de regresión múltiple en Python en torno a un proyecto de aplicación de características reales.

El curso posibilitará a los estudiantes desarrollar sus habilidades en el uso de bibliotecas estadísticas que facilitaran el desarrollo de las operaciones estadísticas, así como de diversas componentes que facilitaran su procesamiento, análisis y puesta a disposición empleando las funcionalidades gráficas.

El curso se desarrollará bajo una metodología del aprender haciendo, en cual todas las experiencias de aprendizaje serán prácticas y de aplicación en torno a un proyecto que será desarrollado a lo largo del curso.

Esta actividad curricular, ubicada en el octavo trimestre del plan de estudios, contribuye al logro de las siguientes competencias del Perfil de Egreso: “Evalúa la eficiencia, rendimiento y niveles de desempeño de procesos y gestión de recursos estratégicos, tácticos y operativos a nivel general y específico, para apoyar el proceso de toma de decisiones de la organización” y “Desarrolla procedimientos computacionales basados en ciencia de datos, tecnologías de la información y modelos de compliance para la generación y uso ético de la información y conocimiento de la organización”.

II. PROPÓSITOS DE APRENDIZAJE

Al término de esta actividad curricular los estudiantes serán capaces de:

1. Emplear funciones básicas de Python para la caracterización, limpieza y procesamiento de datos de una población.
2. Emplear los conceptos de correlación y regresión lineal para la caracterización de un conjunto de datos de una población usando Python.
3. Presentar información de un set de datos utilizando librería Matplotlib para graficar una población de datos.
4. Presentar información de un set de datos utilizando librería Seaborn para graficar una población de datos.
5. Presentar información agregada utilizando funciones de agrupación, agregación y pivoteo para el análisis de una población de datos.

III. CONTENIDOS

Unidad Temática N°1: Proyecto del Curso

El curso se desarrollará en torno a un proyecto de alto nivel de aplicación y contingencia que los alumnos deberán diseñar y resolver usando su conocimiento previo en materias de estadísticas y deberán emplear Python para su resolución de manera integral.

El propósito de la actividad es la integración del conocimiento estadístico adquirido y que esto puedan aplicarlo en Python.

El proyecto constará de 4 avances y los cuales deberán ser presentados por el grupo de alumnos que lo desarrolla.

Unidad Temática N° 2: Estadística Descriptiva en Python

Tema	Aprendizajes esperados	Contenidos		
		Saber conceptuales	Saber hacer procedimientos	Saber ser Actitudes

<p>Estadística Descriptiva en Python</p>	<p>Explica los principales elementos relacionados con la estadística descriptiva para la caracterización de un conjunto de datos de una población en Python.</p> <p>Utiliza medidas de tendencia central para caracterizar un conjunto de datos de una población en Python.</p> <p>Utiliza medidas de dispersión para caracterizar un conjunto de datos de una población en Python.</p> <p>Utiliza gráficos adecuados para caracterizar un conjunto de datos de una población en Python.</p>	<ul style="list-style-type: none"> ▪ Estadística descriptiva ▪ Casos y Variables ▪ Tipos de variable ▪ Matrices de datos y tablas de frecuencia <p>Gráficos y formas de distribución</p> <ul style="list-style-type: none"> ▪ Gráfico de torta y barras ▪ Histogramas <p>Medidas de tendencia central</p> <ul style="list-style-type: none"> ▪ Moda, Mediana y Media. <p>Medidas de dispersión</p> <ul style="list-style-type: none"> ▪ Rangos de dispersión (rango, rango intercuartil) y su representación en boxplot ▪ Varianza y desviación estándar ▪ Z-score 	<ul style="list-style-type: none"> ▪ Usar funcionalidades de Python para la generación de medidas de tendencia central. ▪ Usar funcionalidades de Python para la generación de medidas de dispersión. ▪ Usar funcionalidades de Python para la realización de análisis gráficos. 	<p>Compromiso con el aprendizaje de las herramientas que servirán para el procesamiento de información.</p> <p>Cuidado profesional en el tratamiento de datos, en su integridad, confiabilidad y privacidad</p>
--	--	--	---	---

Unidad Temática Nº 3: Regresión Lineal

Tema	Aprendizajes esperados	Contenidos		
		Saber conceptuales	Saber hacer procedimientos	Saber ser Actitudes

	<p>Realiza la correlación de variables para la caracterización de un conjunto de datos de una población</p> <p>Emplea la regresión lineal para la caracterización de un conjunto de datos de una población</p>	<p>Correlación</p> <ul style="list-style-type: none"> ▪ El concepto de correlación ▪ Graficando la correlación de variables ▪ Midiendo la correlación de variables con el indicador R-Pearson ▪ Causalidad v/s Correlación <p>Regresión Lineal</p> <ul style="list-style-type: none"> ▪ El concepto de regresión ▪ Determinación de los coeficientes de regresión ▪ Indicador de ajuste R2 	<p>Emplear la correlación de variables para la caracterización de un conjunto de datos de una población usando Python.</p> <p>Emplear la regresión lineal para la caracterización de un conjunto de datos de una población usando Python.</p>	<p>Compromiso con el aprendizaje de las herramientas que servirán para el procesamiento de información.</p> <p>Cuidado profesional en el tratamiento de datos, en su integridad, confiabilidad y privacidad</p>
--	--	---	---	---

Unidad Temática Nº 4: Introducción al Análisis Gráfico

Tema	Aprendizajes esperados	Contenidos		
		Saber conceptuales	Saber hacer procedimientos	Saber ser Actitudes
Introducción al Análisis Gráfico	<p>Identificar las principales características de un gráfico utilizando la librería Matplotlib</p> <p>Seleccionar el tipo de gráfico utilizando la librería Matplotlib que mejor representa las necesidades de información de acuerdo a la naturaleza del set de datos</p> <p>Construir un gráfico personalizando sus características para representar un set de datos utilizando la librería Matplotlib</p>	<p>Librería Matplotlib</p> <ul style="list-style-type: none"> ▪ Importación de la librería ▪ Componentes principales de un gráfico ▪ Tipos de gráficos <ul style="list-style-type: none"> ▪ Diagrama de dispersión ▪ Diagrama de barras ▪ Histograma ▪ Diagrama de caja ▪ Figuras y Subgráficos ▪ Posición y tamaño de gráficos 	<p>Utilizar el análisis gráfico empleando la librería Matplotlib que mejor representa las necesidades de información de acuerdo con la naturaleza del set de datos</p> <p>Generar gráficos personalizando sus características para representar un set de datos utilizando la librería Matplotlib</p>	<p>Compromiso con el aprendizaje de las herramientas que servirán para el procesamiento de información.</p> <p>Cuidado profesional en el tratamiento de datos, en su integridad, confiabilidad y privacidad</p>

		<ul style="list-style-type: none"> ▪ colores, marcadores y estilos ▪ Ticks, etiquetas y leyendas ▪ Fijando los límites de un gráfico ▪ Anotaciones y dibujos en un gráfico ▪ Utilizando orientación a objetos para graficar ▪ Guardando los gráficos en un archivo 		
--	--	--	--	--

Unidad Temática Nº 5: Análisis Gráfico Avanzado

Tema	Aprendizajes esperados	Contenidos		
		Saber conceptuales	Saber hacer procedimientos	Saber ser Actitudes
Análisis Gráfico Avanzado	<p>Identificar las principales características de un gráfico utilizando la librería Seaborn</p> <p>Seleccionar el tipo de gráfico utilizando la librería Seaborn que mejor representa las necesidades de información de acuerdo con la naturaleza del set de datos</p> <p>Construir un gráfico personalizando sus características para representar un set de datos utilizando la librería Seaborn</p>	<p>Librería Seaborn</p> <ul style="list-style-type: none"> ▪ Importación de la librería ▪ Tipos de gráfico ▪ Gráficos de distribución de observaciones <ul style="list-style-type: none"> ▪ Distplot ▪ Gráficos de dispersión y correlación de variables <ul style="list-style-type: none"> ▪ Joinplot ▪ Pairplot ▪ Gráficos de regresiones <ul style="list-style-type: none"> ▪ Regplot ▪ Gráficos de variables categóricas <ul style="list-style-type: none"> ▪ Barplot ▪ Countplot ▪ Boxplot ▪ Violinplot 	<p>Utiliza los diversos tipos de gráfico utilizando la librería Seaborn que mejor representa las necesidades de información de acuerdo con la naturaleza del set de datos</p> <p>Construir un gráfico personalizando sus características para representar un set de datos utilizando la librería Seaborn</p>	<p>Compromiso con el aprendizaje de las herramientas que servirán para el procesamiento de información.</p> <p>Cuidado profesional en el tratamiento de datos, en su integridad, confiabilidad y privacidad</p>

		<ul style="list-style-type: none"> ▪ Gráficos de matrices ▪ Heatmap ▪ Grillas de gráficos ▪ Pairgrid ▪ Facetgrid 		
--	--	---	--	--

Unidad Temática Nº 7: Introducción al Análisis Avanzado

Tema	Aprendizajes esperados	Contenidos		
		Saber conceptuales	Saber hacer procedimientos	Saber ser Actitudes
Presentación de información agrupada, agregada y con pivotes	<p>Caracterizar las principales herramientas para el análisis de la información agregada de un set de datos.</p> <p>Generar análisis estadísticos o funciones personalizadas a grupos de datos.</p>	<p>Agrupación de datos</p> <ul style="list-style-type: none"> ▪ División de un set de datos en grupos ▪ Iterando sobre los grupos ▪ Agrupación con funciones ▪ Agrupación con niveles de índices <p>Agregación de datos</p> <ul style="list-style-type: none"> ▪ Funciones estadísticas de agregación ▪ Funciones de agregación personalizadas <ul style="list-style-type: none"> ▪ Estrategia Split-apply-combine ▪ En qué consiste esta estrategia ▪ Tipos de problema que se pueden resolver ▪ Aplicando esta estrategia en Quantiles y Buckets ▪ Llenando datos perdidos con valores específicos de grupo 	<p>Emplear las diversas herramientas y componentes que posibiliten la generación de análisis de la información agregada para un set de datos.</p> <p>Diseñar e implementar funcionalidades que posibiliten un análisis estadístico empleando información agregadas.</p> <p>Desarrollar funciones personalizadas a grupos de datos.</p>	<p>Compromiso con el aprendizaje de las herramientas que servirán para el procesamiento de información.</p> <p>Cuidado profesional en el tratamiento de datos, en su integridad, confiabilidad y privacidad</p>

		<ul style="list-style-type: none"> ▪ Tablas pivote y tabulación cruzada ▪ Construcción de tablas pivote para presentar información ▪ Tabulación cruzada (crosstab) 		
--	--	---	--	--

IV. METODOLOGÍA

El curso se desarrollará en alto porcentaje con una metodología de aprendizaje basado en problemas o bien una metodología de proyectos, propiciando de esta forma, el “aprender haciendo”. Asimismo, realizarán actividades en donde se propicien prácticas como la “programación en grupos”, donde se produzca interacción y discusión entre los estudiantes, así como el desarrollo de trabajos en equipos para fortalecer el aprendizaje colaborativo.

El docente generará simulaciones basadas en casos reales para desafiar a los estudiantes a encontrar una solución a la problemática planteada, en dicho contexto tanto talleres como actividades grupales tienen como objeto poner en práctica el aprendizaje del estudiante en base sus conocimientos adquiridos empleando las herramientas y técnicas aprendidas a lo largo de su proceso de aprendizaje.

De igual forma se empleará el enfoque de clases expositivas, las cuales se reforzarán con prácticas de aprendizaje inverso, las cuales esperan generar un efecto de aprendizaje deductivo, lógico y construido de forma interactiva y participativa.

V. EVALUACIÓN DE APRENDIZAJES

El curso tendrá las siguientes evaluaciones:

- Pruebas parciales
- Controles, talleres y tareas cuyo fin es que los estudiantes realicen trabajos prácticos orientados al desarrollo de las competencias.
- Examen

VI. RECURSOS PEDAGÓGICOS

Bibliografía Básica:

- Python for Data Analysis. 2011. Wes McKinney.
ISBN: 978-1491957660. O'Reilly Media
- Python para todos. 2017. Elliott Hauser, Sue Blumenber
Licencia Creative Commons.

- Python Crash Course: A Hands-On, Project-Based Introduction to Programming. 2019. Matthes, Eric. ISBN: 978-1593279288. Editorial: Random House LCC US.

Bibliografía Complementaria:

- Python Cookbook, 2013. Brian K. Jones, David Beazley
ISBN: 978-1449340377. Editorial: O'Reilly Media.
- Dive into Python 3, 2009. Pilgrim, Mark.
ISBN: 978-1430224150. Editorial: Apress.