



**SOCIEDAD  
ECUATORIANA  
ESTADÍSTICA**

**PROGRAMA  
CIENCIA DE DATOS**

**ENERO 2024**

# SOBRE LA SEE

La Sociedad Ecuatoriana de Estadística (SEE), es una organización civil sin fines de lucro creada hace más de 24 años con el objetivo de promover la toma de decisiones basadas en datos, incentivar la investigación científica y la actualización de conocimientos en el ámbito de la estadística a nivel nacional.

La SEE organiza eventos académicos, profesionales y de difusión como las 11 ediciones del Seminario Internacional de Estadística Aplicada, 9 ediciones del Programa de capacitación en Ciencia de datos, R Weekend, Python Weekend, etc. Además realizamos actividades gratuitas con niños y adolescentes para promover la matemática y la estadística, por ejemplo desarrollamos la App Android Cuentos que Cuentan y Cuentan así como el encuentro de futuros científicos de datos.



**CAPACITACIÓN Y  
CONSULTORÍA**



**CULTURA  
ESTADÍSTICA**



**PROMOCIÓN E  
INVESTIGACIÓN**



**APOYO A LA  
EDUCACIÓN**

# ACERCA DEL PROGRAMA

Actualmente es común que las instituciones, desde PyMEs hasta grandes centros de investigación, acumulen grandes cantidades de datos correspondientes a las actividades que realizan, como ejemplo podemos mencionar a registros de ventas, compras, costos, mediciones de sensores, monitoreo de redes sociales. Todos estos datos son constantemente generados y almacenados en los sistemas transaccionales siendo necesaria la aplicación de técnicas estadísticas para describir, relacionar y analizar dichos datos.

La toma de decisiones basadas en datos es una tendencia mundial, al punto que en todo ámbito se resalta la importancia del Data Scientist, conocido como la persona con habilidades estadísticas, computacionales (puede programar) y de visualización de datos que lo llevan a encontrar los patrones que le servirán a la empresa para "capitalizar" la información recogida.



MSC, Gabriela Castro  
Presidenta

**Sociedad Ecuatoriana de Estadística**



Mgtr. Nestor Montaña  
Coordinador Programa

**Sociedad Ecuatoriana de Estadística**

## A QUIÉN VA DIRIGIDO

El programa internacional de ciencia de datos va dirigido a todo el que maneje información dentro de la empresa u organización, como:

a.- jefes o analistas de departamentos de inteligencia de negocios, Ciencia de datos, Analítica, Estadística, Big Data, Planificación de Demanda, Información

b.- analistas comerciales, analistas de mercado, analistas de marketing, analistas de abastecimiento, etc.

# OBJETIVO

El objetivo del presente programa es formar especialistas en ciencia de datos con capacidad para formular nuevas preguntas e implementar soluciones de alta tecnología que permitan tomar mejores decisiones. Al finalizar el programa el graduado tendrá los elementos necesarios para realizar Data Science: desde desarrollar visualizaciones o aplicaciones exploratorias utilizando herramientas como las shiny apps; conocerá cómo obtener información desde bases de datos relacionales (como Microsoft SQL o MySQL) de forma efectiva y tendrá las destrezas fundamentales para desarrollar scripts en R o Python con el objetivo de realizar análisis exploratorios, descriptivos, modelos lineales o aplicar Statistical, Machine o Deep learning a los problemas comunes que afrontamos en nuestra labor diaria.



# REQUISITOS



El programa es casi autocontenido puesto que empezamos desde lo más básico que es instalar el programa e importar los datos.

**Requisitos mínimos:** Conocimiento básico de matemáticas, habilidad para manejar herramientas como excel y experiencia previa con programación (macros de excel o C++, etc)

# COMPETENCIAS A ADQUIRIR

Manejo de conceptos para análisis de información

Conocimiento de cuándo y cómo usar cada una de las técnicas más importantes en Ciencia de Datos

Manejo fluido de R y Python para data science

Habilidad para desarrollar aplicaciones analíticas en Shiny o Dash



PRECIO NORMAL:

~~1500 USD~~

DESCUENTO  
ESPECIAL

990 USD

## Descuentos

### Publico General:

**5%** pronto pago (hasta 15 septiembre)

**5%** pago en efectivo

**10%** grupos de 3 o más

### Cientes y miembros SEE:

**5%** haber tomado un curso en SEE en este año

**10%** haber tomado 2 o más curso en SEE en este año

**20%** por haber tomado el programa de Ciencia de Datos

## MÉTODO DE PAGO

**Tarjeta de crédito: Online**

**Efectivo**

**Deposito/Transferencia**

Banco Pacífico

Número de cuenta: 7964730

Tipo de cuenta: Corriente

RUC: 1791709381001

Nombre de la organización:

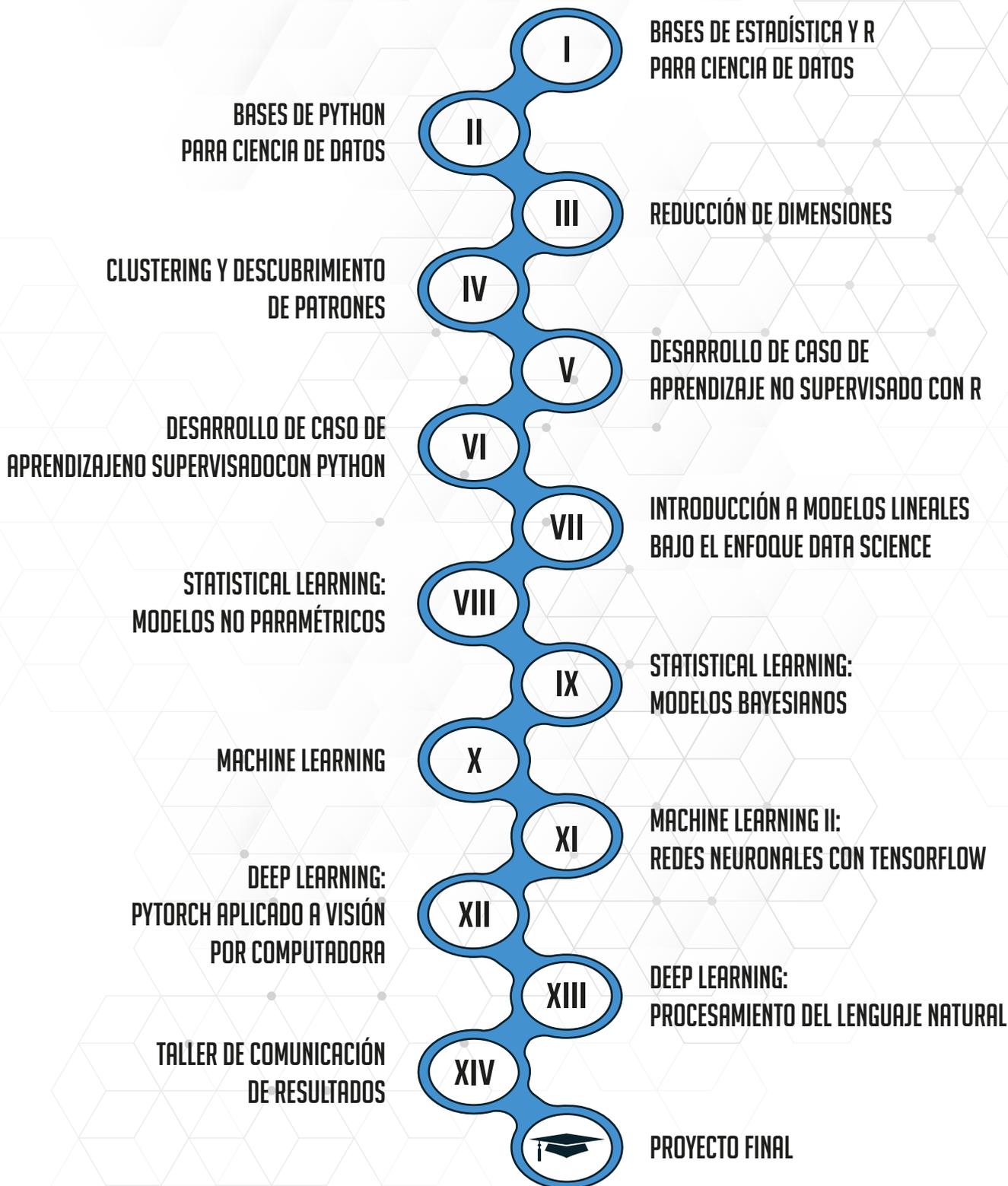
Sociedad Ecuatoriana de Estadística

# MALLA CURRICULAR

**ONLINE SINCRÓNICO**

**210 HORAS SINCRONICAS  
340 HORAS TOTAL**

**CLASES: ENERO - AGOSTO  
PROYECTO FINAL: SEPTIEMBRE**



## HORARIO

LUNES - JUEVES -- 19:00 - 21:30

SÁBADO - DOMINGO -- 08:30 - 13:30

I  
25 H

## BASES DE ESTADÍSTICA Y R PARA CIENCIA DE DATOS

Proporcionar los elementos básicos para empezar a resolver problemas aplicando Ciencia de Datos; esto incluye los conocimientos estadística necesarios para describir e inferir a partir de los datos mientras que se desarrollan las destrezas para utilizar R y Python desde RStudio y Anaconda (Spyder, Jupyter) respectivamente.

El curso incluye topicos como importación y exportación de información, manipulación de estructuras de datos utilizando la conocida colección de paquetes tidyverse en R o pandas en Python; se explica también la generación de gráficos y se desarrollan los conceptos básicos de estadística descriptiva e inferencial, tales como estimación puntual, estimación por intervalos y algunas pruebas de hipótesis tanto paramétricas como no paramétricas

## BASES DE PYTHON PARA CIENCIA DE DATOS

II  
15 H

Con el dominio de las bases para el análisis de series temporales, este curso es tu siguiente paso hacia la obtención pronósticos precisos. Descubrirás el proceso de modelamiento por completo, incluyendo la forma de validar tu modelo, las métricas para evaluar su calidad y cómo entrenarlo y afinar sus hiperparámetros todo esto mientras te familiarizas con librerías en R y Python que potenciarán tus análisis.

Las técnicas no paramétricas como la Descomposición STL y la Suavización Exponencial ETS se convertirán en herramientas esenciales en tu caja de herramientas. Explorarás modelos más avanzados que abordan múltiples fuentes de estacionalidad como la Descomposición MSTL y conocerás técnicas recientes como Facebook Prophet

III  
15 H

## REDUCCIÓN DE DIMENSIONES

En ocasiones el problema a analizar viene acompañado de una gran cantidad de observaciones y variables, esto si bien es beneficioso en cuanto a la disponibilidad de datos representa también un desafío puesto que puede ser computacionalmente intensivo. El presente curso se aprenden técnicas cuyo objetivo es reducir la dimensionalidad de los datos manteniendo la mayor cantidad de información posible.

**IV**

15 H

## CLUSTERING Y DESCUBRIMIENTO DE PATRONES

Desde las moléculas hasta los cuerpos celestes, existen fuerzas que hacen objetos cercanos más similares, esta cercanía o similaridad nos permite clasificar los objetos en grupos o clusters; un clúster es, por lo tanto, una colección de objetos que son "similares" entre ellos y son "diferentes" a los objetos que pertenecen a otros grupos.

Clustering se puede pensar definir como el proceso de organizar objetos en grupos cuyos miembros son similares de alguna manera, este es uno de los problemas de aprendizaje no supervisado más usados puesto que puede aplicarse para encontrar grupos de clientes de comportamiento similar, clasificar plantas o animales dadas sus características, identificación de grupos de titulares de pólizas de seguros con alto riesgo, etc.

## DESARROLLO DE CASO

### DE APRENDIZAJE NO SUPERVISADO CON R

**V**

10 H

En la SEE creemos en el Aprendizaje Práctico, por ello desarrollarás un proyecto real y prácticos que te preparará para enfrentar los desafíos del mundo real. En este módulo van a realizar todos los pasos de un proyecto de ciencia de datos donde se aplique aprendizaje no supervisado, esto es: entender los datos, explorarlos, ajustar el problema a resolver a una actividad de ciencia de datos y construir clusters o aplicar técnicas de reducción de dimensiones, todo esto usando R

**VI**

10 H

## DESARROLLO DE CASODE APRENDIZAJE NO SUPERVISADO CON PYTHON

Seguimos con el Aprendizaje Práctico, pero esta ocasión se lo realizará usando Python. De esta manera se llenará cualquier vacío que haya quedado de los módulos iniciales y se aumentará experiencia práctica a cada estudiante.

## INTRODUCCIÓN A MODELOS LINEALES BAJO EL ENFOQUE DATA SCIENCE

**VII**

30 H

Este curso se imparten las bases teóricas y prácticas para la resolución de problemas de predicción o clasificación (es decir: aprendizaje supervisado) en ciencia de datos; se desarrollan las estrategias train-test, remuestreo o croosvalidación usadas para la estimación de parámetros e hiperparámetros en Ciencia de Datos, todo esto empezando con modelos simples como regresión lineal múltiple, regresión logística para luego entrar en conceptos y modelos más complejos como regularización, balanceo d emuestras, modelos lineales generalizados o Regrsión Ridge, Lasso y elasticNet

**VIII**

15 H

## STATISTICAL LEARNING: MODELOS NO PARAMÉTRICOS

La estadística es la ciencia que facilita la toma de decisiones basadas en datos desde hace siglos, es por ello que mucho de los modelos de aprendizaje tienen fundamentos estadísticos muy fuertes.

Bases para la estadística no paramétrica  
Regresión cuantílica,  
Splines y Smoothing splines,  
Generalized Additive Model (GAM),  
Multivariate Adaptive Regression Splines (MARS)

## STATISTICAL LEARNING: MODELOS BAYESIANOS

**IX**

15 H

La estadística es la ciencia que facilita la toma de decisiones basadas en datos desde hace siglos, es por ello que mucho de los modelos de aprendizaje tienen fundamentos estadísticos muy fuertes.

Entendiendo diferencias entre Máxima verosimilitud, Bootstrap, Monte Carlo  
Introducción a métodos bayesianos,  
Cadenas de Markov de Monte Carlo  
Regresión bayesiana,  
Clasificación bayesiana

**X**

20 H

## MACHINE LEARNING

El boom de las computadoras y la consecuente facilidad para recolectar y almacenar de datos así como para procesarlos ha sido un factor trascendental para el auge de la Ciencia de Datos y con ello un sinnúmero de técnicas enfocadas a "obtener el jugo" de los datos. En este curso se sigue aumentando la caja de herramientas del científico de datos, esta vez con un conjunto de técnicas que se podría considerar vienen de la vena computacional de la ciencia de datos.

Las archi conocidas support vector machines, random forest y boosting son parte de los métodos impartidos en el presente curso que además termina con una introducción al Deep Learning y en particular a la visión por computadoras.

## MACHINE LEARNING II: REDES NEURONALES CON TENSORFLOW

**XI**

20 H

Las redes neuronales y en particular las redes aplicadas a la visión por computadoras (inicio del Deep Learning) han sido las causantes de gran parte del hype que tenemos alrededor de la Ciencia de datos y la inteligencia artificial hoy en día. En este curso se entiende la parte básica de las redes neuronales tanto para problemas con datos tabulares como con problemas de visión por computadora, todo esto usando Tensorflow desde R y Python

**XII**

10 H

## DEEP LEARNING: PYTORCH APLICADO A VISIÓN POR COMPUTADORA

En el curso anterior se sientan las bases para aplicar modelos de Deep Learning e incluso se resuelven los primeros problemas de clasificación de imágenes usando redes neuronales convolucionales; en este curso se continúa en el área de visión por computadora pero esta vez entrando en modelos que permitan generar imágenes sintéticamente. Además se adentra en el problema del procesamiento natural del lenguaje para lo cual se aprende sobre redes recurrentes, útiles también en problemas de series temporales.

**DEEP LEARNING:**

## PROCESAMIENTO DEL LENGUAJE NATURAL

**XIII**

15 H

Este curso permite comprender las bases de la computación en paralelo y los conceptos y destrezas claves para manejar Big Data; el objetivo es que el estudiante tenga interactúe con las herramientas más conocidas con son Hadoop con MapReduce, Spark y Cloudera

**XIV**

5 H

## TALLER DE COMUNICACIÓN DE RESULTADOS

Dicen que los profesionales técnicos tienen el problema de no saber comunicar sus desarrollos, por ello para finalizar el curso se realiza un pequeño taller de comunicación de resultados; enfocados en cómo vender un proyecto de Data Science y en cómo comunicar sus resultados

## PROYECTO FINAL

No hay mejor manera de aprender que mediante la práctica, es por ello que para aprobar el programa de Ciencia de Datos de la SEE se requiere un Proyecto de graduación. El mismo se empieza a realizar cuando ha transcurrido alrededor de la mitad de las materias, los estudiantes arman grupos y escogen un tema que van desarrollando mientras aprenden las últimas materias del programa y con la ayuda de un tutor asignado para que la culminación del proyecto sea exitosa.

# Capacitadores

## ZULEMMA BAZURTO

**Zulemma Bazurto Blacio es Magíster en Estadística por la Pontificia Universidad Católica de Chile e Ingeniera Comercial de la ESPOL. Fundadora de R-Ladies Guayaquil, Co-founder de R-Ladies Galápagos, Docente Titular de Estadística en la Facultad de Ciencias Administrativas de la Universidad de Guayaquil y miembro de la Directiva del Núcleo Guayas de la Sociedad Ecuatoriana de Estadística.**

Su mayor pasión es organizar e impartir conferencias sobre R para análisis de datos. Participa activamente en proyectos colaborativos en la Comunidad Internacional de R, ha formado parte del equipo traductor del famoso libro online "R for Data Science" de Hadley Wickham y Garret Golemund y actualmente forma parte de la Traducción colaborativa al castellano del libro "Teaching Tech Together. How to create and deliver lessons that work and build a teaching community around them" de Greg Wilson (2019, Taylor & Francis, ISBN 978-0-367-35328-5, <https://teachtogether.tech/>). Colabora con posts sobre temas actuales basados en análisis de datos para el grupo de investigación social cuantitativa "Los Cuantificadores" y recientemente fue una de las Keynote Speaker del 1er Women in Data Science (WiDS) Regional Guayaquil. Las líneas de investigación de su interés se centran en Modelos Lineales Generalizados, Métodos Robustos de Estimación, Teaching Data Science, Text Mining, Investigación Reproducible y Visualización de Datos con R.

## KATHERINE MORALES

**Ingeniera Matemática con mención en Estadística e Investigación Operativa por la Escuela Politécnica Nacional (EPN), realizó sus estudios de maestría en Ciencia de Datos en École Polytechnique - Francia.**

Se ha desempeñado como analista de datos en diferentes empresas del país, consultora independiente y capacitadora de R (avalada por la EPN). Adicionalmente, ha participado en varios retos empresariales en el área de ciencia de datos, en Francia y Ecuador. Cuenta con experiencia en aprendizaje bayesiano aplicado a la medicina, aprendizaje automático, procesamiento de imágenes y procesamiento de lenguaje natural.

Actualmente trabaja en el Departamento de Comunicaciones, Imágenes y Procesamiento de la Información de Télécom SudParis - Institute Polytechnique, y su área de estudio son las redes neuronales recurrentes y sus aplicaciones. Además, desarrollará su doctorado conjuntamente con el Hospital de Estrasburgo-Francia, en el tema de 'Inteligencia artificial para ayudar a la toma de decisiones en cirugías cardiovasculares'.

## KAREN CALVA

**Ingeniera Matemática especializada en estadística y ciencia de datos, graduada en la Escuela Politécnica Nacional. Actualmente, especialista de analítica avanzada en una de las instituciones financieras más grandes de Ecuador.**

Consultora independiente en temas de geo-estadística, gestión de procesos y logística para el levantamiento de información, desarrollo de aplicativos webs con motores analíticos y aprendizaje automático.

Docente para estudiantes de pregrado o profesionales, con más de seis años de experiencia dictando cursos y conferencias relacionadas a matemáticas, técnicas estadísticas y manejo de software libre como R y Python, en instituciones como la Sociedad Ecuatoriana de Estadística. Candidata al Máster en Dirección Logística de la Universidad Internacional de La Rioja.

## CHRISTIAN GALARZA

**Docente investigador en la facultad de Matemáticas y Ciencias Naturales de su alma mater, Escuela Superior Politécnica del Litoral - ESPOL, PhD en Estadística de la Universidad Estatal de Campinas, UNICAMP - Brasil, donde también obtuvo su Maestría en Estadística.**

Investigador entusiasta cuyos trabajos iniciales le han permitido ganar varios premios nacionales e internacionales, además de la oportunidad de difundir su investigación en congresos de todo el mundo. En la actualidad, Christian es árbitro de varias revistas científicas de alto impacto, miembro de varias sociedades estadísticas de todo el mundo, joven embajador 2020 del Instituto Internacional de Estadística (ISI) ante el IBC y autor de más de 10 paquetes R. Además, ha visitado 28 países, asistiendo a más de 40 congresos, con más de 30 contribuciones orales, 15 conferencias invitadas, 1 moderador de sesiones, 2 carteles y 3 cursos cortos impartidos. Sus áreas de interés estadístico son regresión de cuantiles, modelos de efectos mixtos lineales / no lineales, algoritmos EM y SAEM, distribuciones de cuantiles cero, mezcla de escalas de distribuciones normales sesgadas y modelos censurados e inflados con cero

## ROBERTO ESTEVES

**Ing. Sistemas por vocación y marketero por decisión, tiene 18 años de experiencia en Marketing Digital en banca, junto a manejo de innovación y tecnología en áreas como medios de comunicación y empresas de telecomunicaciones Subdirector del núcleo Guayas de la Sociedad Ecuatoriana de Estadística, cuenta con un Nanodegree de Udacity en Data Science y Procesamiento de Lenguaje Natural, certificado por la SEE, IBM, Platzi en Ciencia de Datos.**

Consultor de marketing y comunicación, se ha dedicado al estudio del comportamiento de los usuarios en redes sociales para encontrar insights que permitan elaborar mejores campañas. Estudiante de Postgrado en Data Science y Machine Learning en IEBS, la Escuela de Negocios de la Innovación y los Emprendedores, fellow AI Saturday, Udacity Scholarship en programas de Intel y Microsoft.

## NESTOR MONTAÑO

**Consultor privado con más de 10 años de experiencia en el sector privado desempeñando cargos como Jefe de Planificación de Demanda e Inteligencia de negocios y Planificación de demanda, Data Scientist o especialista en el área matemático-estadística o realizando consultorías en reconocidas empresas importadoras, mayoristas, de retail y consultoras. Es Ingeniero y Máster en estadística (ambos realizados en ESPOL), miembro electo del directorio nacional de la Sociedad Ecuatoriana de Estadística SEE y del R User Group Ecuador RUGE**

Además es capacitador de la Sociedad Ecuatoriana de Estadística SEE y diversas instituciones en temas de estadística, modelamiento estadístico o mediante aprendizaje automático, ciencia de datos e inteligencia artificial, esto mediante los lenguajes R o Python. Colabora constantemente con la academia a través de participaciones en eventos o dictado de diversos cursos y conferencias para estudiantes de pregrado o profesionales.

## CHRISTIAN TUTIVEN

**PhD in Automatic Control, Robotics and Control por la Universidad Politécnica de Catalunya, Máster en Administración de empresas por la UESS.**

Máster e Ingeniero en telecomunicaciones por la UCSG; docente e investigador de la Escuela Superior Politécnica del Litoral en áreas de machine learning aplicado a sistemas de control; líder de la comunidad Saturdays AI Guayaquil, cofundador de Steam Center S.A.; cuenta con más de 15 años de experiencia profesional tanto en la empresa privada, investigación y docencia a nivel superior.

## HUGO PORRAS

**Ingeniero en ciencias económicas y financieras de la Escuela Politécnica Nacional y máster en inteligencia artificial de la Universidad Internacional de la Rioja. Tiene 4 años de experiencia como científico de datos y 6 en el área de investigación.**

Actualmente se desempeña como Científico de Datos Senior en Banco Pichincha. Trabaja además como consultor e investigador independiente en temas relacionados a finanzas de bienes raíces, economía del bienestar, economía geográfica e IA aplicada a las ciencias sociales.



# SOCIEDAD ECUATORIANA ESTADÍSTICA

## PIDE MÁS INFORMACIÓN

[see.ecuador1@gmail.com](mailto:see.ecuador1@gmail.com)  
[capacitacion@see-ec.org](mailto:capacitacion@see-ec.org)

+593 98 334 9634



<https://see-ec.org/wordpress/programas/>