

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

COORDINACIÓN GENERAL DE FORMACIÓN PROFESIONAL

PROGRAMA DE UNIDAD DE APRENDIZAJE

I. DATOS DE IDENTIFICACIÓN

- 1. Unidad Académica:** Facultad de Ciencias Administrativas y Sociales, Ensenada; Facultad de Ciencias Administrativas, Mexicali; Facultad de Ciencias de la Ingeniería, Administrativas y Sociales, Tecate; y Facultad de Contaduría y Administración, Tijuana.
- 2. Programa Educativo:** Licenciado en Inteligencia de Negocios
- 3. Plan de Estudios:** 2021-2
- 4. Nombre de la Unidad de Aprendizaje:** Programación para la Extracción de Datos
- 5. Clave:** 39049
- 6. HC: 01 HT: 00 HL: 05 HPC: 00 HCL: 00 HE: 01 CR: 07**
- 7. Etapa de Formación a la que Pertenece:** Disciplinaria
- 8. Carácter de la Unidad de Aprendizaje:** Obligatoria
- 9. Requisitos para Cursar la Unidad de Aprendizaje:** Ninguno



Equipo de diseño de PUA

José Carlos Gallegos Mariscal
Francisco Reynan Beltrán
Esperanza Manrique Rojas
José de Jesús Parra Galaviz
Verónica Quizán García

Fecha: 08 de marzo de 2021

Vo.Bo. de subdirectores de las Unidades Académicas

Adelaida Figueroa Villanueva
Angélica Reyes Mendoza
Esperanza Manrique Rojas
Jesús Antonio Padilla Sánchez

II. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

La unidad de aprendizaje tiene como finalidad que el alumno aprenda las metodologías y algoritmos para la extracción de datos, utilizando herramientas para su análisis y la identificación de las diferentes fuentes disponibles. Todo esto con el objetivo de utilizar la información extraída para ayudar a la toma de decisiones orientadas a la inteligencia de negocios.
Esta unidad de aprendizaje pertenece a la etapa disciplinaria, es de carácter obligatorio, pertenece al área de conocimiento de Ciencia de Datos.

III. COMPETENCIA GENERAL DE LA UNIDAD DE APRENDIZAJE

Desarrollar programas de extracción, manipulación y visualización de datos aplicando los fundamentos y técnicas de lenguaje de programación para generar información que, de soporte en la toma de decisiones, de manera analítica, con confidencialidad y responsabilidad.

IV. EVIDENCIA(S) DE APRENDIZAJE

Aplicación de software que permita visualizar la información obtenida de diferentes fuentes de datos mediante el uso de técnicas para su extracción y manipulación.

V. DESARROLLO POR UNIDADES
UNIDAD I. Recolección de datos

Competencia:

Utilizar algoritmos y técnicas de extracción de datos, a través de herramientas, para su recolección en diferentes fuentes, mostrando actitud analítica, objetiva y honesta.

Contenido:

Duración: 3 horas

- 1.1 Estructuras para la extracción de datos.
 - 1.1.1 Arreglos.
 - 1.1.2 Listas.
 - 1.1.3 Tuplas.
 - 1.1.4 Diccionarios.
 - 1.1.5 Otras estructuras.
- 1.2 Introducción a las librerías en Python para la extracción.
- 1.3 Recolección de datos de archivos.
- 1.4 Web scraping and web crawling.

UNIDAD II. Transformación

Competencia:

Adaptar datos recolectados, por medio de su organización, normalización, exclusión y formato, para su almacenamiento en base de datos, con actitud creativa, organizada y profesional.

Contenido:

Duración: 3 horas

2.1 Almacenamiento en base de datos.

2.2 Pre-procesamiento de los datos.

2.2.1 Normalización.

2.2.2 Exclusión.

2.2.3 Formato.

2.2.3.1 Numérico.

2.2.3.2 Nominales.

2.2.3.3 Valores nulos.

2.2.3.4 Otros formatos.

UNIDAD III. Procesamiento y manipulación

Competencia:

Manipular datos, utilizando técnicas de comparación, filtrado y agrupación, con la finalidad de categorizar y dar formato a los datos para su posterior procesamiento, mostrando actitud analítica y de confidencialidad.

Contenido:

- 3.1 Agrupamiento y segmentación de datos.
- 3.2 Tablas de resumen.
- 3.3 Transformación de Dataframe.
- 3.4 Técnicas de búsqueda.
- 3.5 Indexación.

Duración: 5 horas

UNIDAD IV. Visualización

Competencia:

Presentar información generada a través del procesamiento de los datos aplicando herramientas para el análisis, con el propósito de apoyo a la toma de decisiones, mostrando actitud analítica y de honestidad.

Contenido:

Duración: 5 horas

- 4.1 Estadística básica.
- 4.2 Regresión.
- 4.3 Herramientas para análisis de datos.
 - 4.3.1 Matemáticas (numpy, scipy).
 - 4.3.2 Minería de datos (beautifulsoup, scrapy).
 - 4.3.3 Exploración y visualización de datos (pandas, matplotlib, plotly, seaborn)
- 4.4 Tipos de gráficas.

VI. ESTRUCTURA DE LAS PRÁCTICAS DE LABORATORIO

No.	Nombre de la Práctica	Procedimiento	Recursos de Apoyo	Duración
UNIDAD I				
1	Estructuras para la extracción de datos.	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de un ejercicio relacionado con la temática: Estructuras para la extracción de datos (Arreglos, listas, tuplas, diccionarios y otras) 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	3 horas
2	Recolección de datos de archivo	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Recolección de datos. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	5 horas
3	Web Scraping and web crawling	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Web scraping and web crawling 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	7 horas

UNIDAD II			•	
4	Almacenamiento en base de datos	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Almacenamiento de datos 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	3 horas
5	Procesamiento de datos Normalización	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de un ejercicio relacionado con la temática: Normalización, para preprocesamiento de datos. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	3 horas
6	Exclusión	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: de exclusión de datos y manejo de valores atípicos. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	5 horas
7	Formato	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: de formato (numérico, 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y 	4 horas

		<p>nominales, valores nulos y otros formatos)para preprocesamiento de datos.</p> <ol style="list-style-type: none"> 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	electrónicos	
UNIDAD III			•	
8	Agrupamiento y segmentación de datos	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: de agrupación y segmentación de datos. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	5 horas
9	Tablas de resumen	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Tablas de resumen. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados.67 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	5 horas
10	Transformación de Data Frame	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de tres ejercicios relacionados con la temática: de Transformación de Dataframe. 2. Analiza los problemas presentados. 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	7 horas

		<ol style="list-style-type: none"> 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 		
11	Técnicas de búsqueda	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de cinco ejercicios relacionados con la temática: de Técnicas de búsqueda. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	5 horas
12	Indexación	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: indexación. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Recursos bibliográficos y electrónicos 	3 horas
UNIDAD IV			•	
13	Estadística básica	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de cinco ejercicios relacionados con la temática: Estadística básica 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Python y librería statsmodels • Recursos bibliográficos y electrónicos 	4 horas
14	Regresión	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de 	<ul style="list-style-type: none"> • Computadora • Internet 	2 horas

		<p>dos ejercicios relacionados con la temática: Regresión.</p> <ol style="list-style-type: none"> Analiza los problemas presentados. Codifica los problemas expuestos. Ejecuta los códigos. Depura los errores presentados. Entrega la actividad al docente 	<ul style="list-style-type: none"> Plataforma IDE Recursos bibliográficos y electrónicos 	
15	Matemáticas (numpy, scipy).	<ol style="list-style-type: none"> Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Matemáticas (numpy, scipy). Analiza los problemas presentados. Codifica los problemas expuestos. Ejecuta los códigos. Depura los errores presentados. Entrega la actividad al docente 	<ul style="list-style-type: none"> Computadora Internet Plataforma IDE Python, librerías numpy o scipy Recursos bibliográficos y electrónicos 	5 horas
16	Minería de datos (beautifulsoup, scrapy)	<ol style="list-style-type: none"> Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Minería de datos (beautifulsoup, scrapy). Analiza los problemas presentados. Codifica los problemas expuestos. Ejecuta los códigos. Depura los errores presentados. Entrega la actividad al docente 	<ul style="list-style-type: none"> Computadora Internet Plataforma IDE Python, librerías beautifulsoup o scrapy Recursos bibliográficos y electrónicos 	6 horas
17	Exploración y visualización de datos (pandas, matplotlib, plotly, seaborn)	<ol style="list-style-type: none"> Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Exploración y visualización de datos (pandas, matplotlib, plotly, seaborn) Analiza los problemas presentados. Codifica los problemas expuestos. 	<ul style="list-style-type: none"> Computadora Internet Plataforma IDE Python, librerías pandas, matplotlib, plotly, o seaborn Recursos bibliográficos y electrónicos 	6 horas

		<ol style="list-style-type: none"> 4. Ejecuta los códigos. 5. Depura los errores presentados. 6. Entrega la actividad al docente 		
18	Tipos de graficas	<ol style="list-style-type: none"> 1. Atiende las instrucciones del docente y resuelve un mínimo de dos ejercicios relacionados con la temática: Tipos de gráfica. 2. Analiza los problemas presentados. 3. Codifica los problemas expuestos. 4. Ejecuta los códigos. 5. Depura los errores presentados. 	<ul style="list-style-type: none"> • Computadora • Internet • Plataforma IDE • Python, librerías pandas, matplotlib, plotly, o seaborn • Recursos bibliográficos y electrónicos 	2 horas

VII. MÉTODO DE TRABAJO

Encuadre: El primer día de clase el docente debe establecer la forma de trabajo, criterios de evaluación, calidad de los trabajos académicos, derechos y obligaciones docente-alumno.

Estrategia de enseñanza (docente):

- Técnica expositiva.
- Presenta información sobre los conceptos básicos.
- Presenta y resuelve ejercicios prácticos relacionados con las temáticas.
- Retroalimentación individual y grupal.
- Revisa y evalúa prácticas y actividades.
- Elabora y aplica evaluaciones.
- Proyección de códigos.

Estrategia de aprendizaje (alumno):

- Trabajo colaborativo.
- Exposiciones.
- Desarrollo de códigos.
- Participa activamente en clase.
- Desarrollo de la solución tecnológica en una organización a través del paradigma de la programación funcional que optimice los procesos de una organización.
- Elabora y entrega actividades y prácticas en tiempo y forma.

VIII. CRITERIOS DE EVALUACIÓN

La evaluación será llevada a cabo de forma permanente durante el desarrollo de la unidad de aprendizaje de la siguiente manera:

Criterios de acreditación

- Para tener derecho a examen ordinario y extraordinario, el estudiante debe cumplir con los porcentajes de asistencia que establece el Estatuto Escolar vigente.
- Calificación en escala del 0 al 100, con un mínimo aprobatorio de 60.

Criterios de evaluación

- Exámenes.....	30%
- Prácticas/tareas.....	50%
- Proyecto final.....	20%
Total.....	100%

IX. REFERENCIAS

Básicas

Arizmendi, L., Condor, I. (2019). *Fundamentos de estadística y probabilidades con aplicaciones: (en R, Python y otros softwares de tipo GNU/GPL)*. Edición Kindle.

Barry, P., & Suma, L. (2018). *Python*. Gliwice: Helion.

BRUCE, P. (2020). *PRACTICAL STATISTICS FOR DATA SCIENTISTS*. O'REILLY MEDIA, INC, USA.

Bruce, P., & Bruce, A. (2018). *Practical statistics for data scientists*. Beijing [i pozostate]: O'Reilly Media.

Deitel, P. (2020). *INTRO TO PYTHON FOR COMPUTER SCIENCE AND DATA SCIENCE: LEARNING TO PROGRAM WITH AI, BIG DATA AND THE*. UPPER SADDLE RIVER: PEARSON

Grus, J. (2019). *Data Science from Scratch*, 2nd Edition. O'Reilly Media, Inc.

Hilpisch, Y. (2018). *Python for Finance*, 2nd Edition (2nd ed.).

Johnson, M. (2018). *Concise Introduction to Programming in Python*, Second Edition. CRC Press LLC.

Krogh, J., Krogh., & Gennick. (2018). *MySQL Connector/Python Revealed*. Apress.

Thompson, E. (2020). *Programming for Data Science: 4 Books in 1. The Complete Beginners Guide you Can't Miss to Master the Era of the Data Economy, using Python, Java, SQL Coding*.

Complementarias

Goodrich, M., Tamassia, R., & Goldwasser, M. (2013). *Data structures and algorithms in Python*. Hoboken, N.J.: Wiley.

Hinojosa, G. (2016). *Python paso a paso*. Ra Ma.

Lutz, M. (2009). *Learning Python*. Beijing: O'Reilly.

Python. (2021). <https://www.python.org/>

Severance, Charles. (2016). *Python for Everybody: Exploring Data in Python 3*.

Vanderplas, J. (2016). *Python Data Science Handbook*. Sebastopol, CA: O'Reilly & Associates Inc.

X. PERFIL DEL DOCENTE

El docente que imparta la unidad de aprendizaje de Programación para la Extracción de Datos deberá ser Licenciado en Informática, Licenciado en Sistemas Computacionales, Licenciado en Ciencias Computacionales o carrera afín, Licenciado en Inteligencia de Negocios, o maestría en área afín. Con experiencia mínima de tres años en la docencia y profesional. Debe ser proactivo, analítico, que fomente el trabajo en equipo y la investigación.