

--Anexo--

Instrucciones

Esta tarea se trabaja en equipo de dos personas que previamente seleccionaron.

- Considere los siguientes conjuntos de datos, para responder las preguntas propuestas ó usted se puede formular las preguntas que pueda responder a partir de los datos.
- Recuerden que las preguntas se pueden responder o no depende de la calidad de los datos proporcionados.
- Pueden complementar con otros conjuntos de datos.
- Debe realizar un proceso CRISP-DM, por lo tanto, los pasos deben estar explícitos en la tarea.

Código

- **Proceso ETL:**
<https://colab.research.google.com/drive/1QFwUxX2Ddotynwjzva0cgDGln0RXO0go?usp=sharing>
- **K-means:** https://colab.research.google.com/drive/1vRkiwCATjNXMF1fKgm3oAlNwBIN-MqSL?usp=sharing#scrollTo=Rik5NfJU_Wd

Glosario

CRISP-DM: El proceso estándar transversal para la minería de datos (Cross Industry Standard Process for Data Mining, CRISP) es un marco metodológico ampliamente utilizado en el campo de la minería de datos para guiar y estructurar proyectos de análisis de datos. Proceso estándar transversal para la minería de datos (Cross Industry Standard Process for Data Mining)¹

LOCATEL: Locatel es un servicio de asistencia e información prioritariamente en la Ciudad de México y la Zona metropolitana del Valle de México.²

Dataset: Colección de información organizada en filas y columnas que se puede almacenar en un sistema de datos estructurados³

Campo: Columna. Eje Ya en un dataset

K-means: Método de agrupamiento utilizado en aprendizaje no supervisado para agrupar datos en k grupos basados en características similares.⁴

¹ Evangelista, M. (n.d.). Cross Industry Standard Process for Data Mining (CRISP-DM) en Minería de datos. Recuperado 2 de noviembre, 2024, de URL: <https://miguelevangelista.gitbook.io/mineria-de-datos/proceso-de-mineria-de-datos/cross-industry-standard-process-for-data-mining-crisp-dm>

² Gobierno de la Ciudad de México. (n.d.). Locatel: Proyectos. Sistema de Información y Atención a la Ciudadanía. <https://adip.cdmx.gob.mx/proyectos/locatel>

³ Ironhack. (2023, March 22). Una guía sobre datasets: qué son, cómo se utilizan y dónde encontrarlos. Ironhack. <https://www.ironhack.com/es/blog/una-guia-sobre-datasets-que-son-como-se-utilizan-y-donde-encontrarlos>

⁴ Evangelista, M. (n.d.). K-means. En Minería de datos. Recuperado 2 de noviembre, 2024, de URL: <https://miguelevangelista.gitbook.io/mineria-de-datos/tecnicas-de-mineria-de-datos/agrupamiento/k-means>

Aprendizaje Automatico: Rama de la inteligencia artificial (IA) que se encarga de crear algoritmos y modelos estadísticos para que las computadoras aprendan de los datos y mejoren con la experiencia⁵

Machine Learning: Idem

Cluster: Agrupamiento de datos que revela la diversidad y la coherencia inherentes en los datos, permitiendo identificar grupos de datos que comparten atributos comunes y se diferencian de los demás.⁶

Clustering: Proceso de agrupamiento de datos⁷

Modulo: Colección de funciones para tareas específicas reutilizables asociadas a un lenguaje de programación⁸

Algoritmo: Conjunto de instrucciones ordenadas y finitas que se utilizan para resolver un problema o realizar una tarea cuya aplicación permiten el desarrollo de *software* y sistemas⁹

Diccionarios

Diccionario 'fecha_alta'

- '2022-01-01' se sustituye por 1
- [...]
- '28/05/2023' se sustituye por 173932

Diccionario 'estado_usuario'

- | | | |
|------------------------|----------------------|---------------------|
| 1. AGUASCALIENTES | 19. DAKOTA DEL NORTE | 37. NAYARIT |
| 2. ALABAMA | 20. DELAWARE | 38. NEBRASKA |
| 3. ALASKA | 21. DURANGO | 39. NEVADA |
| 4. ARIZONA | 22. ESTADO DE MEXICO | 40. NUEVA JERSEY |
| 5. ARKANSAS | 23. FLORIDA | 41. NUEVA YORK |
| 6. BAJA CALIFORNIA | 24. GEORGIA | 42. NUEVO LEON |
| 7. BAJA CALIFORNIA SUR | 25. GUANAJUATO | 43. NUEVO MEXICO |
| 8. CALIFORNIA | 26. GUERRERO | 44. OAXACA |
| 9. CAMPECHE | 27. HIDALGO | 45. OHIO |
| 10. CAROLINA DEL NORTE | 28. ILLINOIS | 46. OREGON |
| 11. CAROLINA DEL SUR | 29. INDIANA | 47. PENNSILVANIA |
| 12. CHIAPAS | 30. IOWA | 48. PUEBLA |
| 13. CHIHUAHUA | 31. JALISCO | 49. QUERETARO |
| 14. CIUDAD DE MEXICO | 32. LOUISIANA | 50. QUINTANA ROO |
| 15. COAHUILA | 33. MICHIGAN | 51. SAN LUIS POTOSI |
| 16. COLIMA | 34. MICHOACAN | 52. SINALOA |
| 17. COLORADO | 35. MISURI | 53. SONORA |
| 18. CONNECTICUT | 36. MORELOS | 54. TABASCO |

⁵ IBM. (n.d.). Machine learning. IBM. <https://www.ibm.com/mx-es/topics/machine-learning/>

⁶ Immune Institute. (n.d.). ¿Qué es el clustering? Immune Institute. <https://immune.institute/blog/que-es-el-clustering/>

⁷ Immune Institute. (n.d.). ¿Qué es el clustering? Immune Institute. <https://immune.institute/blog/que-es-el-clustering/>

⁸ TechTarget. (n.d.). Module. TechTarget. <https://www.techtarget.com/whatis/definition/module>

⁹ Universidad de la República. (n.d.). ¿Qué son algoritmos? Universidad de la República. <https://ude.edu.uy/que-son-algoritmos/>

55. TAMAULIPAS
56. TENNESSEE
57. TEXAS
58. TLAXCALA

59. UTAH
60. VERACRUZ
61. VIRGINIA
62. WASHINGTON

63. WISCONSIN
64. WYOMING
65. YUCATAN
66. ZACATECAS

Diccionario 'servicio'

- **'JURIDICO'** se sustituye por 1
- **'PSICOLOGICO'** se sustituye por 2
- **'MEDICO'** se sustituye por 3

Diccionario 'sexo'

- **'FEMENINO'** se sustituye por 0
- **'MASCULINO'** se sustituye por 1