



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

FACULTAD DE INGENIERÍA ASIGNATURA: FUNDAMENTOS PARA EL ANÁLISIS DE DATOS

UNIDAD DE APRENDIZAJE 4: INTRODUCCIÓN AL ANÁLISIS DESCRIPTIVO EXPLORATORIO EDA

Introducción

La exploración y análisis descriptivo exploratorio (EDA, por sus siglas en inglés) es una fase fundamental en el proceso de análisis de datos y estadísticas. Su objetivo principal es obtener una comprensión inicial y profunda de un conjunto de datos, revelando patrones, tendencias, relaciones y características clave. Es el punto de partida para cualquier proyecto de análisis de datos y desempeña un papel crucial en la formulación de preguntas de investigación, la identificación de problemas de calidad de datos y la toma de decisiones informadas.

El EDA involucra una serie de técnicas y herramientas estadísticas y gráficas para examinar y resumir los datos de manera sistemática. Algunas de las actividades comunes que se realizan en el EDA incluyen:

Resumen de Datos: Se calculan estadísticas descriptivas básicas, como la media, mediana, moda, desviación estándar y cuartiles, para comprender las propiedades centrales y la dispersión de los datos.

Visualización de Datos: Se crean gráficos y visualizaciones, como histogramas, gráficos de barras, gráficos de dispersión y box plots, para representar los datos de manera efectiva. Estas visualizaciones ayudan a identificar patrones visuales y valores atípicos.

Análisis de Distribuciones: Se investiga la distribución de los datos para determinar si siguen una distribución normal u otra distribución específica. Esto es importante para seleccionar las pruebas estadísticas adecuadas.

Exploración de Relaciones: Se analizan las relaciones entre variables mediante correlaciones, matrices de dispersión y análisis de regresión. Esto puede ayudar a identificar variables predictoras y dependientes.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"

Detección de Datos Atípicos: Se buscan y manejan valores atípicos o errores en los datos que podrían sesgar los resultados del análisis.

Segmentación de Datos: Se agrupan los datos en segmentos o categorías relevantes para identificar patrones dentro de subconjuntos específicos.

Exploración Geoespacial: En el caso de datos geoespaciales, se utilizan mapas y visualizaciones espaciales para comprender la distribución geográfica de los datos.

Análisis Temporal: Si los datos tienen un componente temporal, se realizan análisis de series temporales para identificar tendencias y patrones a lo largo del tiempo.

El EDA no es un proceso lineal y puede implicar iteraciones a medida que se descubren nuevos conocimientos y se generan nuevas preguntas. Esta fase de análisis es esencial para guiar la selección de técnicas estadísticas posteriores y para asegurar que se obtenga una comprensión sólida de los datos antes de emprender análisis más avanzados o la construcción de modelos.

4.1. INTRODUCCIÓN AL ANÁLISIS DESCRIPTIVO EXPLORATORIO EN PYTHON

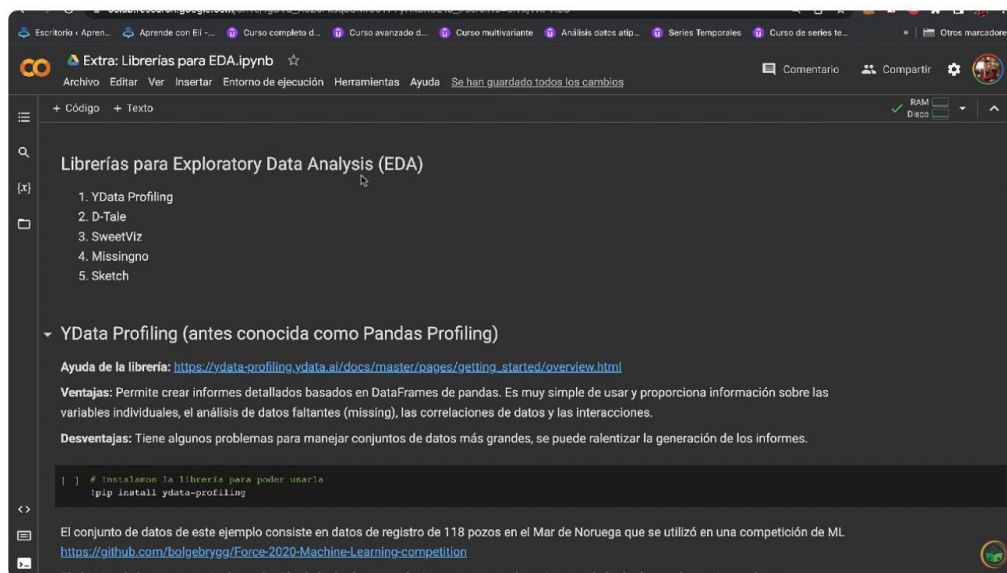
Figura 1. Fases del análisis exploratorio de datos



Nota: Imagen tomada de: <https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos>



Para el desarrollo de esta unidad de aprendizaje, que realmente se compone de creación de código, vamos a examinar el video:



Fuente: <https://www.youtube.com/watch?v=shcDnxhP12c>

4.2. Análisis descriptivo exploratorio - Caso Práctico 1

A través de un caso práctico (archivo en formato jupyter notebook) abarcaremos el análisis descriptivo exploratorio. Veamos:

<https://github.com/jaquimbayoc8/DiplomadoMLA2023/blob/main/2.%20AnalisisEDA/analisis.ipynb>

4.3. Caso Práctico 2

Se comparte el archivo en formato jupyter notebook con el análisis descriptivo exploratorio de un caso práctico. Enlace:

<https://github.com/jaquimbayoc7/DocumentacionDiplomadoML/tree/master/1.%20LimpiezaDatos>

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Bibliografía

- Agarwal, R., & Dhar, V. (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25(3), 443–448. <https://doi.org/10.1287/isre.2014.0546>
- Cao, L. (2017). Data science. *Communications of the ACM*, 60(8), 59–68. <https://doi.org/10.1145/3015456>
- Cleveland, W. (2014). Data science: An action plan for expanding the technical areas of the field of statistics. Undefined; <https://www.semanticscholar.org/paper/Data-science%3A-An-action-plan-for-expanding-the-of-Cleveland/915cd8e2b39eb02723553913d592b2237d4d9960>
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73. <https://doi.org/10.1145/2500499>
- Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766. <https://doi.org/10.1080/10618600.2017.1384734>
- Hayashi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. *Studies in Classification, Data Analysis, and Knowledge Organization*, 40–51. https://doi.org/10.1007/978-4-431-65950-1_3
- Loukides, M. (2011). The future belongs to the companies and people that turn data into products. https://www.gmsl.it/wp-content/uploads/2014/09/What_Is_Data_Science_.pdf
- Pereira, T., Hernández-Arteaga, S., & Alvarado, A. (2016). <https://doi.org/10.16925/9789587600490>
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>
- MIDAS. (2022). Michigan Institute For Data Science. MIDAS. <https://midas.umich.edu/about/>
- Shum, S. B., Hawksey, M., Baker, R., Jeffery, N., Behrens, J., & Pea, R. (2013). Educational data scientists: a scarce breed. Undefined; <https://www.semanticscholar.org/paper/Educational-data-scientists%3A-a-scarce-breed-Shum-Hawksey/6dddce1601e81cfe486e2864f32f1cfa33181982>
- Zhu, Y., & Xiong, Y. (2015). Defining Data Science. <https://arxiv.org/ftp/arxiv/papers/1501/1501.05039.pdf>

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"