

# REVOLUCIÓN DE LA INVESTIGACIÓN DE MERCADOS IMPULSADA POR LA IA



**CAROLINA UZCÁTEGUI**  
**DAVID ZALDUMBIDE**  
**EMANUEL LEITE**  
(ORGANIZADORES)



**PUCE**

**UMET**  
UNIVERSIDAD  
METROPOLITANA



**EDITORA  
ARTEMIS**

2024

# REVOLUCIÓN DE LA INVESTIGACIÓN DE MERCADOS IMPULSADA POR LA IA



**CAROLINA UZCÁTEGUI**  
**DAVID ZALDUMBIDE**  
**EMANUEL LEITE**  
(ORGANIZADORES)



**PUCE**

**UMET**  
UNIVERSIDAD  
METROPOLITANA



**EDITORA  
ARTEMIS**

2024



O conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons Atribuição-Não-Comercial NãoDerivativos 4.0 Internacional (CC BY-NC-ND 4.0). Direitos para esta edição cedidos à Editora Artemis pelos autores. Permitido o download da obra e o compartilhamento, desde que sejam atribuídos créditos aos autores, e sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

A responsabilidade pelo conteúdo dos artigos e seus dados, em sua forma, correção e confiabilidade é exclusiva dos autores. A Editora Artemis, em seu compromisso de manter e aperfeiçoar a qualidade e confiabilidade dos trabalhos que publica, conduz a avaliação cega pelos pares de todos manuscritos publicados, com base em critérios de neutralidade e imparcialidade acadêmica.

<b>Editora Chefe</b>	Prof. <sup>a</sup> Dr. <sup>a</sup> Antonella Carvalho de Oliveira
<b>Editora Executiva</b>	M. <sup>a</sup> Viviane Carvalho Mocellin
<b>Direção de Arte</b>	M. <sup>a</sup> Bruna Bejarano
<b>Diagramação</b>	Elisangela Abreu
<b>Organizadores</b>	Carolina Uzcátegui-Sánchez David Zaldumbide-Peralvo Emanuel Leite
<b>Imagem da Capa</b>	Dall-E
<b>Bibliotecário</b>	Maurício Amormino Júnior – CRB6/2422

#### Conselho Editorial

Prof.<sup>a</sup> Dr.<sup>a</sup> Ada Esther Portero Ricol, *Universidad Tecnológica de La Habana “José Antonio Echeverría”, Cuba*  
Prof. Dr. Adalberto de Paula Paranhos, *Universidade Federal de Uberlândia, Brasil*  
Prof. Dr. Agustín Olmos Cruz, *Universidad Autónoma del Estado de México, México*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Amanda Ramalho de Freitas Brito, *Universidade Federal da Paraíba, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Ana Clara Monteverde, *Universidad de Buenos Aires, Argentina*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Ana Júlia Viamonte, *Instituto Superior de Engenharia do Porto (ISEP), Portugal*  
Prof. Dr. Ángel Mujica Sánchez, *Universidad Nacional del Altiplano, Peru*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Angela Ester Mallmann Centenaro, *Universidade do Estado de Mato Grosso, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Begoña Blandón González, *Universidad de Sevilla, Espanha*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Carmen Pimentel, *Universidade Federal Rural do Rio de Janeiro, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Catarina Castro, *Universidade Nova de Lisboa, Portugal*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Cirila Cervera Delgado, *Universidad de Guanajuato, México*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Cláudia Neves, *Universidade Aberta de Portugal*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Cláudia Padovesi Fonseca, *Universidade de Brasília-DF, Brasil*  
Prof. Dr. Cleberton Correia Santos, *Universidade Federal da Grande Dourados, Brasil*  
Prof. Dr. David García-Martul, *Universidad Rey Juan Carlos de Madrid, Espanha*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Deuzimar Costa Serra, *Universidade Estadual do Maranhão, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Dina Maria Martins Ferreira, *Universidade Estadual do Ceará, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Edith Luévano-Hipólito, *Universidad Autónoma de Nuevo León, México*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Eduarda Maria Rocha Teles de Castro Coelho, *Universidade de Trás-os-Montes e Alto Douro, Portugal*  
Prof. Dr. Eduardo Eugênio Spers, *Universidade de São Paulo (USP), Brasil*  
Prof. Dr. Eloi Martins Senhoras, *Universidade Federal de Roraima, Brasil*  
Prof.<sup>a</sup> Dr.<sup>a</sup> Elvira Laura Hernández Carballido, *Universidad Autónoma del Estado de Hidalgo, México*



Prof.<sup>ª</sup> Dr.<sup>ª</sup> Emilas Darlene Carmen Lebus, *Universidad Nacional del Nordeste/ Universidad Tecnológica Nacional, Argentina*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Erla Mariela Morales Morgado, *Universidad de Salamanca, Espanha*  
Prof. Dr. Ernesto Cristina, *Universidad de la República, Uruguay*  
Prof. Dr. Ernesto Ramírez-Briones, *Universidad de Guadalajara, México*  
Prof. Dr. Fernando Hitt, *Université du Québec à Montréal, Canadá*  
Prof. Dr. Gabriel Díaz Cobos, *Universitat de Barcelona, Espanha*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Gabriela Gonçalves, Instituto Superior de Engenharia do Porto (ISEP), Portugal  
Prof. Dr. Geoffroy Roger Pointer Malpass, Universidade Federal do Triângulo Mineiro, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Gladys Esther Leoz, *Universidad Nacional de San Luis, Argentina*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Glória Beatriz Álvarez, *Universidad de Buenos Aires, Argentina*  
Prof. Dr. Gonçalo Poeta Fernandes, Instituto Politécnico da Guarda, Portugal  
Prof. Dr. Gustavo Adolfo Juarez, *Universidad Nacional de Catamarca, Argentina*  
Prof. Dr. Guillermo Julián González-Pérez, *Universidad de Guadalajara, México*  
Prof. Dr. Håkan Karlsson, *University of Gothenburg, Suécia*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Iara Lúcia Tescarollo Dias, Universidade São Francisco, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Isabel del Rosario Chiyon Carrasco, *Universidad de Piura, Peru*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Isabel Yohena, *Universidad de Buenos Aires, Argentina*  
Prof. Dr. Ivan Amaro, Universidade do Estado do Rio de Janeiro, Brasil  
Prof. Dr. Iván Ramon Sánchez Soto, *Universidad del Bío-Bío, Chile*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Ivânia Maria Carneiro Vieira, Universidade Federal do Amazonas, Brasil  
Prof. Me. Javier Antonio Albornoz, *University of Miami and Miami Dade College, Estados Unidos*  
Prof. Dr. Jesús Montero Martínez, *Universidad de Castilla - La Mancha, Espanha*  
Prof. Dr. João Manuel Pereira Ramalho Serrano, Universidade de Évora, Portugal  
Prof. Dr. Joaquim Júlio Almeida Júnior, UniFIMES - Centro Universitário de Mineiros, Brasil  
Prof. Dr. Jorge Ernesto Bartolucci, *Universidad Nacional Autónoma de México, México*  
Prof. Dr. José Cortez Godínez, Universidad Autónoma de Baja California, México  
Prof. Dr. Juan Carlos Cancino Díaz, Instituto Politécnico Nacional, México  
Prof. Dr. Juan Carlos Mosquera Feijoo, *Universidad Politécnica de Madrid, Espanha*  
Prof. Dr. Juan Diego Parra Valencia, *Instituto Tecnológico Metropolitano de Medellín, Colômbia*  
Prof. Dr. Juan Manuel Sánchez-Yáñez, *Universidad Michoacana de San Nicolás de Hidalgo, México*  
Prof. Dr. Juan Porras Pulido, *Universidad Nacional Autónoma de México, México*  
Prof. Dr. Júlio César Ribeiro, Universidade Federal Rural do Rio de Janeiro, Brasil  
Prof. Dr. Leinig Antonio Perazolli, Universidade Estadual Paulista (UNESP), Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Livia do Carmo, Universidade Federal de Goiás, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Luciane Spanhol Bordignon, Universidade de Passo Fundo, Brasil  
Prof. Dr. Luis Fernando González Beltrán, *Universidad Nacional Autónoma de México, México*  
Prof. Dr. Luis Vicente Amador Muñoz, *Universidad Pablo de Olavide, Espanha*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Macarena Esteban Ibáñez, *Universidad Pablo de Olavide, Espanha*  
Prof. Dr. Manuel Ramiro Rodríguez, *Universidad Santiago de Compostela, Espanha*  
Prof. Dr. Manuel Simões, Faculdade de Engenharia da Universidade do Porto, Portugal  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Márcia de Souza Luz Freitas, Universidade Federal de Itajubá, Brasil  
Prof. Dr. Marcos Augusto de Lima Nobre, Universidade Estadual Paulista (UNESP), Brasil  
Prof. Dr. Marcos Vinícius Meiado, Universidade Federal de Sergipe, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Mar Garrido Román, *Universidad de Granada, Espanha*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Margarida Márcia Fernandes Lima, Universidade Federal de Ouro Preto, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> María Alejandra Arecco, *Universidad de Buenos Aires, Argentina*  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Maria Aparecida José de Oliveira, Universidade Federal da Bahia, Brasil  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Maria Carmen Pastor, *Universitat Jaume I, Espanha*



Prof.ª Dr.ª Maria da Luz Vale Dias – Universidade de Coimbra, Portugal  
Prof.ª Dr.ª Maria do Céu Caetano, Universidade Nova de Lisboa, Portugal  
Prof.ª Dr.ª Maria do Socorro Saraiva Pinheiro, Universidade Federal do Maranhão, Brasil  
Prof.ª Dr.ª MªGraça Pereira, Universidade do Minho, Portugal  
Prof.ª Dr.ª Maria Gracinda Carvalho Teixeira, Universidade Federal Rural do Rio de Janeiro, Brasil  
Prof.ª Dr.ª María Guadalupe Vega-López, *Universidad de Guadalajara, México*  
Prof.ª Dr.ª Maria Lúcia Pato, Instituto Politécnico de Viseu, Portugal  
Prof.ª Dr.ª Maritza González Moreno, *Universidad Tecnológica de La Habana, Cuba*  
Prof.ª Dr.ª Mauriceia Silva de Paula Vieira, Universidade Federal de Lavras, Brasil  
Prof. Dr. Melchor Gómez Pérez, Universidad del Pais Vasco, Espanha  
Prof.ª Dr.ª Ninfa María Rosas-García, Centro de Biotecnología Genómica-Instituto Politécnico Nacional, México  
Prof.ª Dr.ª Odara Horta Boscolo, Universidade Federal Fluminense, Brasil  
Prof. Dr. Osbaldo Turpo-Gebera, *Universidad Nacional de San Agustín de Arequipa, Peru*  
Prof.ª Dr.ª Patrícia Vasconcelos Almeida, Universidade Federal de Lavras, Brasil  
Prof.ª Dr.ª Paula Arcoverde Cavalcanti, Universidade do Estado da Bahia, Brasil  
Prof. Dr. Rodrigo Marques de Almeida Guerra, Universidade Federal do Pará, Brasil  
Prof. Dr. Saulo Cerqueira de Aguiar Soares, Universidade Federal do Piauí, Brasil  
Prof. Dr. Sérgio Bitencourt Araújo Barros, Universidade Federal do Piauí, Brasil  
Prof. Dr. Sérgio Luiz do Amaral Moretti, Universidade Federal de Uberlândia, Brasil  
Prof.ª Dr.ª Silvia Inés del Valle Navarro, *Universidad Nacional de Catamarca, Argentina*  
Prof.ª Dr.ª Solange Kazumi Sakata, Instituto de Pesquisas Energéticas e Nucleares (IPEN)- USP, Brasil  
Prof.ª Dr.ª Stanislava Kashtanova, *Saint Petersburg State University, Russia*  
Prof.ª Dr.ª Susana Álvarez Otero – Universidad de Oviedo, Espanha  
Prof.ª Dr.ª Teresa Cardoso, Universidade Aberta de Portugal  
Prof.ª Dr.ª Teresa Monteiro Seixas, Universidade do Porto, Portugal  
Prof. Dr. Valter Machado da Fonseca, Universidade Federal de Viçosa, Brasil  
Prof.ª Dr.ª Vanessa Bordin Viera, Universidade Federal de Campina Grande, Brasil  
Prof.ª Dr.ª Vera Lúcia Vasilévski dos Santos Araújo, Universidade Tecnológica Federal do Paraná, Brasil  
Prof. Dr. Wilson Noé Garcés Aguilar, *Corporación Universitaria Autónoma del Cauca, Colômbia*  
Prof. Dr. Xosé Somoza Medina, *Universidad de León, Espanha*

**Dados Internacionais de Catalogação na Publicação (CIP)  
(eDOC BRASIL, Belo Horizonte/MG)**

R454 *Revolución de la Investigación de Mercados Impulsada por la IA [livro eletrônico] / Organizadores Carolina Uzcátegui-Sánchez, David Zaldumbide-Peralvo, Emanuel Leite. – Curitiba, PR: Artemis, 2024.*

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

Edição bilíngue

ISBN 978-65-81701-35-2

DOI 10.37572/EdArt\_081124352

1. Inteligência artificial. 2. Transformação digital. 3. Pesquisa de mercado. I. Uzcátegui-Sánchez, Carolina. II. Zaldumbide-Peralvo, David. III. Leite, Emanuel. IV. Título.

CDD 658.83

**Elaborado por Maurício Amormino Júnior – CRB6/2422**



## PRESENTACIÓN

En los últimos años, la investigación de mercados ha experimentado una profunda transformación gracias al avance de la IA. Esta tecnología, que inicialmente se centraba en la automatización y el análisis de datos, ha evolucionado para incluir herramientas capaces de comprender y procesar el lenguaje natural, permitiendo una interacción más fluida entre las máquinas y los seres humanos. El impacto de la IA en la investigación de mercados ha sido revolucionario, permitiendo a las empresas tomar decisiones basadas en análisis más precisos, eficiente, y eficaz, lo que ha redefinido la relación entre los negocios y sus consumidores.

El presente libro, ofrece una visión propositiva sobre cómo la IA está transformando las metodologías y enfoques en la investigación de mercados. Este libro se estructura en seis capítulos, cada uno abordando aspectos clave de esta evolución tecnológica.

**Capítulo 1. Transformación de la Investigación de Mercados con IA:** Este capítulo, escrito por Karine Rosália Felix Praça Gomes, Carolina Uzcátegui-Sánchez y Adriana Santamaría-Mendoza, explora cómo la IA ha cambiado el panorama de la investigación de mercados. Desde el uso de algoritmos de deep learning hasta el procesamiento del lenguaje natural, las autoras destacan las oportunidades y desafíos que estas tecnologías representan para los investigadores. La automatización de procesos y la capacidad de analizar grandes volúmenes de datos con precisión marcan un antes y un después en la investigación empresarial.

**Capítulo 2. Ingeniería de Prompts e Investigación de Mercados:** David Zaldumbide-Peralvo, Francisco Zagari-Forte y Tulio Carrión-González profundizan en el concepto de la ingeniería de prompts, una técnica que ha ganado relevancia con el uso de modelos de lenguaje como GPT. Este capítulo examina cómo la formulación de prompts efectivos puede optimizar la calidad de los resultados obtenidos en las investigaciones de mercado, permitiendo a las empresas obtener insights más precisos y contextualizados.

**Capítulo 3. Fundamentos y Aplicaciones de Analítica de Dato:** Emanuel Ferreira Leite, Karen Serrano Orellana y Antonella Coronel Ojeda nos presentan un análisis de cómo la analítica de datos ha pasado de ser una herramienta complementaria a un eje central en la toma de decisiones empresariales. El capítulo aborda tanto los fundamentos de la analítica descriptiva, predictiva y prescriptiva, como sus aplicaciones prácticas en diversos sectores del mercado.

**Capítulo 4. Investigación de Mercados y GPTs Personalizados:** En este capítulo, Carolina Uzcátegui, Arturo Cabezas Aguilar y María del Carmen Franco Gómez examinan el uso de modelos de lenguaje pre-entrenados (GPTs) personalizados en la investigación de mercados. La personalización de estos modelos permite a las empresas ajustar las

respuestas generadas por la IA a las necesidades específicas de su sector o mercado, optimizando así la eficacia de sus campañas y estrategias de marketing.

**Capítulo 5. Estrategias de Marketing en el Sector Bancario:** Carolina Uzcátegui, Andreinga González Ordoñez y Armando Urdaneta Montiel se adentran en el análisis de las estrategias de marketing aplicadas en el sector bancario, un sector que ha adoptado de manera significativa la IA para mejorar la experiencia del cliente y optimizar sus servicios. Este capítulo muestra cómo las tecnologías de IA están siendo empleadas para segmentar el mercado, personalizar ofertas y mejorar la retención de clientes en un entorno altamente competitivo.

**Capítulo 6. Competitividad de las Estrategias de Marketing en Pymes de Portoviejo:** Este es el último capítulo, donde David Zaldumbide Peralvo, David Morales López y Adriana Quevedo Espinoza estudian el impacto de la IA en la competitividad de las pequeñas y medianas empresas (pymes). A través de un estudio de caso en Portoviejo, Ecuador, los autores exploran cómo la adopción de estrategias de marketing basadas en IA puede mejorar la posición competitiva de las pymes en mercados locales e internacionales.

Este libro, además de ofrecer un análisis sugerente, se convierte en una obra de referencia para aquellos que deseen comprender los aspectos básicos de como la IA está transformando la investigación de mercados y las estrategias empresariales en la era digital.

Adicionalmente, se agradece la colaboración desinteresada y enriquecedora de cada colega que participó como par de cada uno de los capítulos, los pares involucrados en este trabajo fueron:

- Dra. Odalys Burgo Bencomo, Universidad Metropolitana, Sede Machala (Ecuador)
- Dra. Delia Esperanza García Vences, Universidad Autónoma del Estado de México (México)
- Mgs. Roberto Jácome Galarza, Universidad Internacional del Ecuador, Ecuador
- Mgs. María Teresa Mite Albán, Universidad de Guayaquil (Ecuador)
- Mgs. Raquel Irene Murillo Villacís, Universidad Técnica Particular de Loja (Ecuador)
- Mgs. Jose Gonzalo Narvaez Cumbicos, Universidad de Guayaquil (Ecuador)
- Mgs. Pamela Elizabeth Pasato Gualpa, Universidad Internacional de la Rioja (España)
- Dra. María José Pérez Espinoza, Universidad Metropolitana, Sede Machala (Ecuador)
- Dra. Gloria Ramírez Elías, Universidad Autónoma de Tlaxcala (México)

- Dr. Bil Serrano Orellana, Universidad Técnica de Machala (Ecuador)
- Dra. Izabele Sousa-Barros, Universidade de Pernambuco (Brasil)
- Ana de Lourdes Torralbas Blazquez, Universidad de Holguín (Cuba)

A todos ellos, nuestro más profundo reconocimiento por su dedicación y compromiso, que han hecho posible la finalización de este proyecto colectivo.

Carolina Uzcátegui-Sánchez

## PRÓLOGO

La investigación de mercados ha sido un pilar fundamental para las empresas en su búsqueda de la comprensión del comportamiento del consumidor, la competencia y las tendencias emergentes. Sin embargo, los avances en inteligencia artificial (IA) han transformado radicalmente la forma en que se lleva a cabo esta investigación, planteando nuevos desafíos y ofreciendo oportunidades sin precedentes. La IA ha cambiado el enfoque tradicional de la recopilación y análisis de datos, proporcionando herramientas más precisas y eficientes para capturar insights que antes resultaban inalcanzables con los métodos convencionales.

Históricamente, la investigación de mercados ha dependido en gran medida de encuestas, entrevistas y grupos focales para obtener información sobre las percepciones y preferencias de los consumidores. Si bien estas metodologías continúan siendo valiosas, presentan limitaciones significativas, como el sesgo del encuestado, el alto costo de implementación y el tiempo requerido para analizar los resultados. La IA, por otro lado, ha permitido superar estas barreras mediante el uso de algoritmos avanzados de aprendizaje automático y procesamiento del lenguaje natural (NLP, por sus siglas en inglés). Estas tecnologías no solo permiten la recopilación de grandes volúmenes de datos, sino que también facilitan un análisis más rápido y detallado de estos datos, transformando la investigación de mercados en un proceso más dinámico y adaptativo.

Se puede afirmar que centrarse en el estudio de la IA en la investigación de mercados es ineludible, dada la creciente relevancia que estas tecnologías tienen para la toma de decisiones empresariales. El mercado actual se caracteriza por su alta volatilidad y la rápida evolución de las preferencias de los consumidores. En este contexto, las empresas que no logren adaptarse a estos cambios mediante el uso de herramientas avanzadas de análisis, como la IA, corren el riesgo de quedar rezagadas. La IA proporciona una ventaja competitiva significativa al permitir a las empresas predecir con mayor precisión las tendencias futuras y personalizar sus estrategias de marketing en función de los comportamientos individuales de los consumidores.

Uno de los aspectos más prometedores de la IA en la investigación de mercados es su capacidad para analizar datos no estructurados, como comentarios en redes sociales, reseñas de productos y conversaciones en línea. Tradicionalmente, este tipo de datos ha sido difícil de cuantificar y analizar de manera efectiva debido a su naturaleza compleja y variada. Sin embargo, con el desarrollo de tecnologías de procesamiento del lenguaje natural, ahora es posible extraer insights valiosos de estas fuentes, permitiendo a las empresas comprender mejor las emociones, actitudes y preferencias de los consumidores en tiempo real. Esto ha llevado a un cambio en la forma en que las empresas interactúan con sus clientes, permitiendo una personalización a gran escala que antes era inimaginable.

Además, la IA está redefiniendo la relación entre el tiempo y la toma de decisiones en la investigación de mercados. En el pasado, los estudios de mercado solían requerir semanas o incluso meses para completarse, lo que a menudo resultaba en decisiones basadas en información que ya estaba desactualizada. Hoy en día, la IA permite que las empresas accedan a datos en tiempo real y tomen decisiones casi instantáneamente. Esto es crucial en un mundo donde las preferencias de los consumidores pueden cambiar de un día para otro, y donde la velocidad de adaptación es clave para mantener una ventaja competitiva.

El uso de IA también ha generado nuevas oportunidades para la segmentación de mercados. En lugar de depender exclusivamente de datos demográficos básicos, como la edad o el género, las empresas ahora pueden segmentar a sus clientes en función de patrones de comportamiento más complejos, como sus interacciones en línea, historial de compras y respuestas emocionales a diferentes productos o campañas publicitarias. Esta capacidad de segmentar a los consumidores con un nivel de precisión sin precedentes ha permitido a las empresas desarrollar campañas de marketing altamente personalizadas y dirigidas, aumentando la eficacia de sus esfuerzos de marketing y mejorando la experiencia del cliente.

A pesar de los beneficios significativos que ofrece la IA, es importante reconocer que su adopción en la investigación de mercados no está exenta de desafíos. Uno de los principales obstáculos es la necesidad de garantizar la privacidad y seguridad de los datos. La recopilación masiva de datos personales plantea preocupaciones éticas y legales, especialmente en un contexto donde los consumidores son cada vez más conscientes de cómo se utiliza su información. Las empresas deben ser transparentes en sus prácticas de recopilación de datos y asegurarse de que cumplen con las normativas de protección de datos, como el Reglamento General de Protección de Datos (GDPR) en Europa. Además, deben implementar medidas de seguridad robustas para proteger los datos contra accesos no autorizados y posibles ciberataques.

Otro desafío es la posibilidad de que la IA perpetúe sesgos existentes en los datos. Dado que los algoritmos de IA aprenden de los datos con los que son entrenados, si estos datos contienen sesgos, es probable que los algoritmos reproduzcan estos sesgos en sus predicciones y recomendaciones. Esto puede tener implicaciones negativas, especialmente en áreas como la segmentación de mercados y la personalización de campañas publicitarias. Por lo tanto, es fundamental que los investigadores de mercados sean conscientes de estos riesgos y trabajen activamente para mitigar los sesgos en los datos y garantizar que los algoritmos de IA se utilicen de manera ética y responsable.

Por último, la adopción de IA en la investigación de mercados requiere una inversión significativa en tecnología y formación. Las empresas deben estar dispuestas a invertir en la infraestructura tecnológica necesaria para implementar soluciones de IA, así

como en la formación de su personal para garantizar que comprendan cómo utilizar estas herramientas de manera efectiva. Esto incluye no solo a los equipos de investigación de mercados, sino también a otros departamentos clave, como marketing y ventas, que se beneficiarán de los insights generados por la IA.

María José Pérez Espinoza

## SUMÁRIO

### **CAPÍTULO 1..... 1**

#### TRANSFORMACIÓN DEL MARKETING CON IA

Karine Rosália Felix Praça Gomes

Carolina Uzcátegui-Sánchez

Adriana Santamaría-Mendoza

 [https://doi.org/10.37572/EdArt\\_0811243521](https://doi.org/10.37572/EdArt_0811243521)

### **CAPÍTULO 2..... 28**

#### INGENIERÍA DE PROMPTS E INVESTIGACIÓN DE MERCADOS

David Zaldumbide-Peralvo

Francisco Zagari-Forte

Tulio Carrión-González

 [https://doi.org/10.37572/EdArt\\_0811243522](https://doi.org/10.37572/EdArt_0811243522)

### **CAPÍTULO 3..... 58**

#### FUNDAMENTOS Y APLICACIONES DE ANALÍTICA DE DATOS

Emanuel Leite

Karen Serrano-Orellana

Antonella Coronel-Ojeda

 [https://doi.org/10.37572/EdArt\\_0811243523](https://doi.org/10.37572/EdArt_0811243523)

### **CAPÍTULO 4..... 91**

#### INVESTIGACIÓN DE MERCADOS Y GPTS PERSONALIZADOS

Carolina Uzcátegui Sánchez

Arturo Cabezas-Aguilar

María del Carmen Franco-Gómez

 [https://doi.org/10.37572/EdArt\\_0811243524](https://doi.org/10.37572/EdArt_0811243524)

### **CAPÍTULO 5..... 114**

#### ESTRATEGIAS DE MARKETING EN EL SECTOR BANCARIO: UN ANÁLISIS EMPÍRICO EN LA PROVINCIA DE EL ORO, ECUADOR

Carolina Uzcátegui-Sánchez

Andreina González-Ordoñez

Armando Urdaneta-Montiel

 [https://doi.org/10.37572/EdArt\\_0811243525](https://doi.org/10.37572/EdArt_0811243525)

**CAPÍTULO 6.....145**

COMPETITIVIDAD DE LAS ESTRATEGIAS DE MARKETING EN PYMES DE PORTOVIEJO

David Zaldumbide-Peralvo

David Morales-López

Adriana Quevedo-Espinoza

 [https://doi.org/10.37572/EdArt\\_0811243526](https://doi.org/10.37572/EdArt_0811243526)

**SOBRE OS ORGANIZADORES ..... 169**

**ÍNDICE REMISSIVO .....171**

## GLOSARIO

**AI o IA (Artificial Intelligence / Inteligencia Artificial):** Rama de la informática que se enfoca en la creación de sistemas y tecnologías capaces de realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de voz, la toma de decisiones, el aprendizaje y la resolución de problemas.

**Algoritmo:** Conjunto de reglas o instrucciones definidas paso a paso para resolver un problema o realizar una tarea específica.

**Análisis Predictivo:** Técnica de análisis de datos que utiliza modelos estadísticos para predecir futuros eventos o comportamientos basados en datos históricos.

**API (Interfaz de Programación de Aplicaciones):** Conjunto de herramientas y definiciones para construir software y permitir la interacción con otras aplicaciones.

**Automatización:** Uso de tecnología para realizar tareas sin intervención humana directa.

**Automatización del Marketing:** Uso de software y tecnología para gestionar campañas de marketing y procesos de forma automatizada, optimizando la eficiencia.

**Base de Datos:** Conjunto organizado de datos almacenados electrónicamente que puede ser accesible y gestionado por sistemas informáticos.

**Big Data:** Grandes volúmenes de datos que requieren de tecnologías avanzadas para su análisis y procesamiento.

**Business Intelligence (BI):** Proceso de recopilación, análisis y presentación de datos relevantes para ayudar en la toma de decisiones empresariales.

**Chatbot:** Programa de IA diseñado para simular conversaciones con usuarios humanos.

**Ciberseguridad:** Conjunto de medidas y prácticas destinadas a proteger sistemas informáticos y datos de ataques, daños o accesos no autorizados.

**Cloud Computing (Computación en la Nube):** Tecnología que permite el acceso remoto a recursos informáticos como almacenamiento y procesamiento de datos a través de internet.

**CRM (Customer Relationship Management):** Estrategia y software utilizados para gestionar y analizar las interacciones de una empresa con sus clientes.

**Customer Journey (Viaje del Cliente):** Conjunto de interacciones y experiencias que un cliente tiene con una empresa desde el primer contacto hasta la postventa.

**Data Analytics (Analítica de Datos):** Proceso de analizar grandes conjuntos de datos para obtener insights que apoyen la toma de decisiones.

**Data Mining (Minería de Datos):** Proceso de extraer patrones o conocimientos útiles de grandes volúmenes de datos mediante herramientas y técnicas de análisis avanzado.

**Data Science (Ciencia de Datos):** Disciplina que combina métodos matemáticos, estadísticos y de programación para analizar y extraer información de los datos.

**Datos Estructurados:** Datos organizados en un formato definido, como una base de datos relacional, lo que facilita su almacenamiento y análisis.

**Datos No Estructurados:** Información que no sigue un formato predefinido, como texto libre, imágenes o videos, y que requiere técnicas avanzadas para su análisis.

**Deep Learning (Aprendizaje Profundo):** Rama del aprendizaje automático que utiliza redes neuronales para procesar grandes cantidades de datos y descubrir patrones complejos.

**E-commerce (Comercio Electrónico):** Venta de productos o servicios a través de medios digitales.

**Experiencia del Usuario (UX):** Percepción y respuesta de un usuario ante la interacción con un producto o servicio, especialmente en términos de usabilidad y satisfacción.

**Feedback Loop (Bucle de Retroalimentación):** Proceso mediante el cual los resultados de una acción o sistema son retroalimentados al sistema para mejorar su rendimiento o comportamiento.

**GPT (Generative Pre-trained Transformer):** Modelo de lenguaje que utiliza técnicas de deep learning para generar texto similar al producido por humanos.

**Inteligencia Artificial (IA):** Tecnología que permite a las máquinas simular procesos cognitivos humanos como el aprendizaje y la resolución de problemas.

**Interfaz de Usuario (UI):** Espacio donde se produce la interacción entre el usuario y un sistema o software, como botones, menús y gráficos.

**KPI (Indicador Clave de Desempeño):** Métrica utilizada para evaluar el éxito o rendimiento de una actividad o proceso empresarial.

**Lead:** Persona o empresa que muestra interés en los productos o servicios de una organización, y que tiene el potencial de convertirse en cliente.

**LLM (Large Language Model):** Modelos de lenguaje de gran escala entrenados con grandes volúmenes de datos para realizar tareas como la generación de texto, la traducción automática y el análisis de sentimientos. Ejemplos de LLM incluyen GPT-3 y GPT-4.

**Machine Learning (Aprendizaje Automático):** Subcampo de la IA que permite a las máquinas mejorar su rendimiento en tareas mediante la experiencia.

**Marketing Digital:** Conjunto de estrategias empleadas en medios digitales para promocionar productos o servicios.

**NLP (Natural Language Processing / Procesamiento del Lenguaje Natural):** Subcampo de la inteligencia artificial que se enfoca en la interacción entre las computadoras y el lenguaje humano, permitiendo que las máquinas comprendan, interpreten y respondan al lenguaje natural.

**Omnicanalidad:** Estrategia de marketing que integra múltiples canales de venta y comunicación para ofrecer una experiencia uniforme al cliente.

**Optimización de Conversiones:** Proceso de mejorar la tasa de conversión de visitantes a clientes en sitios web o campañas de marketing digital.

**Personalización:** Proceso de adaptar productos o servicios para satisfacer las necesidades individuales de los consumidores.

**Prompt:** Instrucción o conjunto de palabras que se proporciona a un modelo de lenguaje, como GPT, para generar una respuesta. En el contexto de la inteligencia artificial, los prompts son esenciales para guiar el comportamiento y las respuestas de los modelos, optimizando su output de acuerdo con las necesidades del usuario.

**Prompt Engineering (Ingeniería de Prompts):** Técnica utilizada para mejorar la interacción con modelos de lenguaje mediante la optimización de las entradas textuales (prompts).

**Pyme (Pequeña y Mediana Empresa):** Empresas con un número limitado de empleados y un volumen de facturación moderado, que juegan un papel fundamental en la economía, especialmente en sectores como el comercio y los servicios.

**Red Neuronal:** Modelo computacional inspirado en el cerebro humano utilizado en el deep learning.

**ROI (Return on Investment):** Indicador financiero que mide el retorno de una inversión en relación con su costo, utilizado para evaluar la eficiencia y rentabilidad de inversiones empresariales.

**Startup:** Empresa emergente o de nueva creación que se caracteriza por su capacidad de crecimiento rápido y su enfoque en la innovación, generalmente en sectores tecnológicos o digitales.

**Segmentación de Mercado:** Proceso de dividir un mercado en grupos más pequeños con características y necesidades similares.

**SEM (Search Engine Marketing / Marketing en Motores de Búsqueda):** Estrategia de marketing digital que se centra en aumentar la visibilidad de un sitio web en los motores de búsqueda mediante el uso de anuncios pagados, como Google Ads. A diferencia del SEO, el SEM se basa en publicidad de pago por clic (PPC) y otras formas de publicidad paga en los motores de búsqueda.

**SEO (Search Engine Optimization / Optimización para Motores de Búsqueda):** Conjunto de técnicas y estrategias que se utilizan para mejorar la visibilidad y el posicionamiento de un sitio web en los resultados orgánicos de los motores de búsqueda, como Google, mediante la optimización de contenido, estructura y enlaces.

**Transformación Digital:** Proceso mediante el cual las empresas adoptan tecnologías digitales para mejorar su eficiencia y competitividad.

**Visualización de Datos:** Técnica utilizada para representar datos complejos de manera gráfica, facilitando su interpretación.

# CAPÍTULO 3

## FUNDAMENTOS Y APLICACIONES DE ANALÍTICA DE DATOS



Data de submissão: 28/09/2024

Data de aceite: 18/10/2024

**Emanuel Leite**

Universidade do Pernambuco, Brasil

<https://orcid.org/0000-0002-3086-2002>

**Karen Serrano-Orellana**

Universidad Metropolitana

Sede Machala, Ecuador

<https://orcid.org/0000-0001-8477-5716>

**Antonella Coronel-Ojeda**

Universidad Metropolitana

Sede Machala, Ecuador

<https://orcid.org/0009-0005-4520-4829>

**RESUMEN:** El presente capítulo explora de manera ilustrativa y ordena los principios esenciales y las aplicaciones prácticas del análisis de datos. Se abordan las definiciones y categorías clave del análisis de datos, incluyendo el análisis descriptivo, predictivo

y prescriptivo. La evolución histórica del Data Analytics se examina desde sus inicios en la estadística moderna hasta las tecnologías actuales de inteligencia artificial y análisis en la nube. Además, se diferencian claramente los conceptos de Business Intelligence, Data Science y Analítica de Datos, destacando sus aplicaciones específicas en diversos sectores. También se revisan las herramientas y metodologías más utilizadas, así como las etapas críticas del procesamiento y preprocesamiento de datos. Finalmente, se enfatiza la importancia de la visualización de datos para comunicar eficazmente los resultados del análisis y fundamentar decisiones estratégicas.

**PALABRAS CLAVE:** Analítica de Datos. Business Intelligence. Data Science. Machine Learning.

### FUNDAMENTALS AND APPLICATIONS OF DATA ANALYTICS

**ABSTRACT:** The present chapter explores in an illustrative and organized manner the

essential principles and practical applications of data analysis. It covers key definitions and categories of data analysis, including descriptive, predictive, and prescriptive analysis. The historical evolution of Data Analytics is examined, from its beginnings in modern statistics to current technologies such as artificial intelligence and cloud analytics. Additionally, clear distinctions are made between the concepts of Business Intelligence, Data Science, and Data Analytics, highlighting their specific applications across various sectors. The chapter also reviews the most commonly used tools and methodologies, as well as the critical stages of data processing and preprocessing. Finally, the importance of data visualization is emphasized as a means to effectively communicate analysis results and support strategic decision-making.

**KEYWORDS:** Data analytics. Business Intelligence. Data Science. Machine Learning.

## FUNDAMENTOS E APLICAÇÕES DA ANÁLISE DE DADOS

**RESUMO:** O presente capítulo explora de forma ilustrativa e organizada os princípios essenciais e as aplicações práticas da análise de dados. São abordadas definições chave e categorias de análise de dados, incluindo análise descritiva, preditiva e prescritiva. A evolução histórica da Análise de Dados é examinada, desde seus primórdios na estatística moderna até as tecnologias atuais, como inteligência artificial e análise em nuvem. Além disso, são feitas distinções claras entre os conceitos de Business Intelligence, Ciência de Dados e Análise de Dados, destacando suas aplicações específicas em diversos setores. O capítulo também revisa as ferramentas e metodologias mais comumente utilizadas, assim como as etapas críticas de processamento e pré-processamento de dados. Por fim, é enfatizada a importância da visualização de dados como um meio de comunicar eficazmente os resultados da análise e apoiar a tomada de decisões estratégicas.

**PALAVRAS-CHAVE:** Análise de dados. Business Intelligence. Ciência de Dados. Aprendizado de Máquina.

## 1 INTRODUCCIÓN

En la era de la información, el análisis de datos se ha convertido en una herramienta fundamental para las organizaciones que buscan obtener una ventaja competitiva en un mercado cada vez más dinámico y complejo. La Analítica de Datos permite transformar grandes volúmenes de datos brutos en conocimientos accionables, facilitando la toma de decisiones estratégicas y operativas basadas en evidencia (Marr, 2020). Este capítulo, ofrece una visión comprensiva y estructurada sobre los principios esenciales y las aplicaciones prácticas del análisis de datos.

La Analítica de Datos abarca una amplia gama de técnicas y metodologías que van desde el análisis descriptivo, que proporciona una visión general de eventos pasados, hasta el análisis predictivo, que utiliza datos históricos para prever eventos futuros, y el análisis prescriptivo, que sugiere acciones específicas para optimizar resultados (Bose, 2021). Este enfoque multifacético permite a las organizaciones no solo comprender lo que ha sucedido, sino también anticipar lo que podría suceder y cómo pueden actuar para alcanzar sus objetivos estratégicos.

La evolución de la Analítica de Datos ha sido impulsada por avances tecnológicos significativos a lo largo del tiempo. Desde los primeros desarrollos en estadística moderna y teoría de la probabilidad en el siglo XIX hasta la actual era de la inteligencia artificial y el aprendizaje profundo, cada etapa ha contribuido a la sofisticación y capacidades actuales del análisis de datos. Herramientas y tecnologías como los sistemas de gestión de bases de datos, herramientas de Business Intelligence, big data, machine learning y plataformas de análisis en la nube han transformado radicalmente la forma en que las organizaciones recopilan, almacenan, analizan y visualizan datos (Provost Y Fawcett, 2020).

Este capítulo también aborda la diferenciación entre la Analítica de Datos, Business Intelligence y Data Science, subrayando sus enfoques y propósitos específicos. Mientras que BI se centra en la recopilación y visualización de datos históricos para la toma de decisiones empresariales, Data Science combina técnicas avanzadas de estadística y computación para explorar y predecir eventos futuros. La Analítica de Datos, por su parte, actúa como un puente entre ambos, proporcionando un análisis integral que combina enfoques descriptivos, predictivos y prescriptivos (Lederer y Schmid, 2021).

Además, se exploran las diversas aplicaciones de la Analítica de Datos en diferentes sectores, destacando su capacidad para mejorar la eficiencia operativa, optimizar la cadena de suministro, predecir tendencias del mercado y personalizar ofertas de productos. Herramientas populares como Python, R, SQL, Tableau y Excel son esenciales para los analistas de datos, facilitando desde la recolección y limpieza de datos hasta su análisis y visualización.

El procesamiento y preprocesamiento de datos son etapas críticas que garantizan la calidad y precisión del análisis posterior. Técnicas como la eliminación de duplicados, la imputación de datos faltantes, la normalización y el escalado de datos aseguran que los datos estén en la forma adecuada para ser analizados (Sharma et al, 2024). El análisis exploratorio de datos (EDA) y las metodologías de modelado y evaluación, incluyendo la regresión, clustering, árboles de decisión y Random Forest, son fundamentales para construir modelos predictivos robustos y precisos. Adicionalmente, la visualización de datos juega un papel destacado en la comunicación efectiva de los resultados del análisis. La correcta elección de la representación gráfica no solo facilita la interpretación de los datos, sino que también resalta los insights obtenidos, permitiendo una comprensión más profunda y fundamentando decisiones estratégicas y operativas en un negocio.

Este capítulo proporciona una guía básica y accesible para entender los fundamentos y aplicaciones de la Analítica de Datos, preparando a los lectores para enfrentar los desafíos y aprovechar las oportunidades en este campo disciplinar.

## 2 FUNDAMENTOS DE LA ANALÍTICA DE DATOS

### 2.1 DEFINICIÓN

El Análítica de Datos es la disciplina que se enfoca en el análisis de datos brutos para extraer información significativa y útil. Este proceso involucra varias etapas, desde la recolección y limpieza de datos hasta su análisis y visualización. El objetivo final es obtener insights que puedan informar la toma de decisiones y estrategias de negocio. Análítica de Datos permite transformar grandes volúmenes de datos en conocimiento accionable mediante técnicas y herramientas estadísticas y computacionales (Varghese et al., 2023).

El alcance del análisis de datos incluye varias categorías de análisis:

- **Análisis descriptivo:** Proporciona una visión general de lo que ha sucedido en el pasado mediante la recopilación, organización y presentación de datos históricos. Este tipo de análisis utiliza medidas estadísticas como la media, la mediana, junto con herramientas como los gráficos de barras y los histogramas para proporcionar un resumen detallado y comprensible de los datos históricos (Vetter, 2017).
- **Análisis predictivo:** Se enfoca en predecir futuros eventos al identificar y analizar patrones en los datos históricos. Utiliza técnicas como regresión, series temporales y algoritmos de machine learning para anticipar tendencias y comportamientos futuros (Kumari y Aggrawal, 2022).
- **Análisis prescriptivo:** Va más allá de la predicción al proporcionar recomendaciones sobre acciones específicas que podrían maximizar los resultados deseados o minimizar riesgos. Utiliza técnicas de optimización y simulación para ofrecer recomendaciones precisas (Affenseller et al., 2022).

### 2.2 EVOLUCIÓN Y CONTEXTO

La evolución del Data Analysis ha sido impulsada por el desarrollo de nuevas tecnologías y metodologías. Aquí se presenta una breve línea de tiempo que destaca los hitos clave en su desarrollo:

- **Siglo XIX:** El desarrollo de la estadística moderna y la teoría de la probabilidad sentó las bases para el análisis de datos. Pioneros como Carl Friedrich Gauss y Adolphe Quetelet contribuyeron significativamente a estos campos (Friendly, 2008).

- **1960-1970:** Surgieron los primeros sistemas de gestión de bases de datos (DBMS), permitiendo la organización y almacenamiento eficiente de grandes volúmenes de datos (Ramos, 2006).
- **Años 1980:** La introducción de herramientas de Business Intelligence (BI) permitió a las empresas realizar análisis descriptivos más sofisticados. Herramientas como OLAP (procesamiento analítico en línea) y los primeros sistemas de informes facilitaron el acceso y la visualización de datos (Conesa, 2012).
- **Años 2000:** La expansión de internet y el crecimiento exponencial de los datos digitales llevaron al inicio de la era de big data. Tecnologías como Hadoop y bases de datos NoSQL permitieron el procesamiento y almacenamiento de datos masivos (Bengfort y Kim, 2016).
- **Década de 2010:** La popularización de machine learning y avances en procesamiento de datos a gran escala transformaron la capacidad de análisis predictivo y prescriptivo. Herramientas como Python y R se convirtieron en estándares de la industria para el análisis de datos (Lehrer et al., 2018).
- **Desde 2020:** En los últimos años, la evolución del análisis de datos ha sido marcada por el surgimiento de la inteligencia artificial (IA) avanzada y el aprendizaje profundo (deep learning). Herramientas y frameworks como TensorFlow y PyTorch han revolucionado el desarrollo y la implementación de modelos de IA. Además, la adopción del análisis de datos en la nube con plataformas como AWS, Google Cloud y Microsoft Azure ha permitido un acceso más amplio y escalable a capacidades avanzadas de análisis de datos. La integración de la analítica avanzada en procesos empresariales se ha acelerado, impulsada por la necesidad de resiliencia y adaptación en tiempos de incertidumbre global (Xu et al., 2021).

### 2.3 DIFERENCIA ENTRE ANÁLISIS DE DATOS, BUSINESS INTELLIGENCE Y DATA SCIENCE

Es importante distinguir entre Analítica de Datos, Business Intelligence (BI) y Data Science, ya que cada uno tiene un enfoque y propósito específico:

- **Business Intelligence (BI):** Se enfoca principalmente en la recopilación, almacenamiento y visualización de datos históricos, proporcionando una visión más detallada y comprensible del negocio, para ayudar en la toma de decisiones empresariales. Emplea una amplia gama de herramientas diseñadas para la obtención, visualización, gestión, transformación y

almacenamiento de datos (Sharda et al., 2022). Además, incluye herramientas especializadas en generación de reportes y procesamiento analítico en línea (OLAP). Entre estas herramientas se destacan los dashboards interactivos, informes y consultas SQL, que permiten presentar la información de manera comprensible. BI responde a preguntas como ¿Qué sucedió? y ¿Por qué sucedió?

- **Data Science:** Es un campo interdisciplinario que combina técnicas de estadística, informática y conocimientos de dominio para extraer conocimientos de datos estructurados y no estructurados (Roman y Joanne, 2022). Incluye técnicas avanzadas de machine learning, modelado predictivo y procesamiento de lenguaje natural (NLP). Data Science responde a preguntas como ¿Qué sucederá? y ¿Cómo podemos hacer que suceda?
- **Analítica de Datos:** Actúa como un puente entre BI y Data Science, enfocándose en analizar datos para obtener insights prácticos. Mientras que BI es más retrospectivo y Data Science más exploratorio y predictivo, Analítica de Datos combina ambos enfoques para proporcionar un análisis integral (Barón et al., 2021). Analítica de Datos responde a preguntas como ¿Qué podemos aprender de los datos? y ¿Qué acciones debemos tomar en base a estos insights?

## 2.4 APLICACIONES DEL ANALÍTICA DE DATOS

El Analítica de Datos, puede ser utilizado para varios propósitos, se presentan a manera algunas situaciones y escenarios de uso comunes:

- **Análisis Descriptivo:** Una empresa de retail puede utilizar análisis descriptivo para entender las tendencias de ventas pasadas, identificar productos más vendidos y desarrollar perfiles de los clientes usando datos demográficos y comportamentales, ayudando a comprender mejor quiénes son los clientes, cómo interactúan con la marca y analizar el comportamiento de compra (Sidel et al., 2018). Herramientas como Excel y Tableau pueden ser utilizadas para crear dashboards y gráficos interactivos que resumen estos datos.
- **Análisis Predictivo:** Una compañía de seguros podría utilizar técnicas de machine learning para predecir la probabilidad de siniestros basándose en datos históricos de clientes (Martín y Giménez, 2020). Algoritmos como regresión logística y árboles de decisión pueden ser implementados en Python usando bibliotecas como scikit-learn.

- **Análisis Prescriptivo:** Una empresa de manufactura puede usar análisis prescriptivo para optimizar su cadena de suministro. Utilizando modelos de simulación y optimización, la empresa puede identificar las mejores estrategias para minimizar costos y tiempos de entrega (Bertsimas y Van Parys, 2017).

### 3 METODOLOGÍAS Y TÉCNICAS DE LA ANALÍTICA DE DATOS

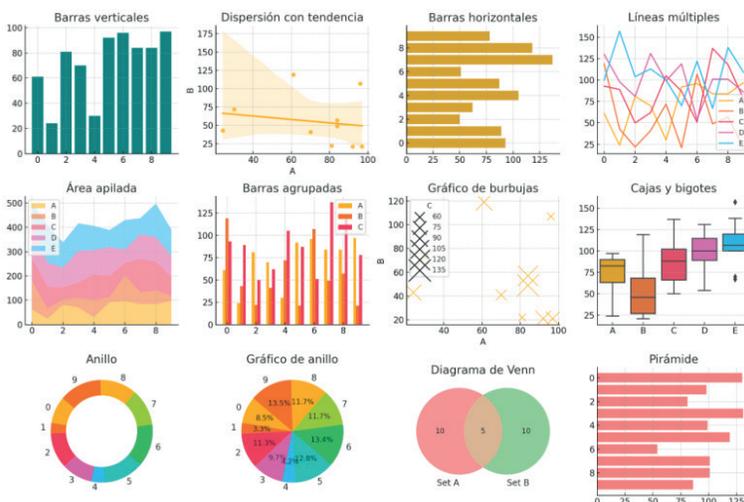
#### 3.1 DESCRIPCIÓN DE METODOLOGÍAS CLAVE

En Analítica de Datos, las metodologías se dividen principalmente en tres categorías: descriptiva, predictiva y prescriptiva. Cada una de estas metodologías tiene técnicas y herramientas específicas asociadas, y se utilizan para diferentes propósitos dentro del análisis de datos.

**1) Análisis Descriptivo:** El análisis descriptivo se centra en resumir y describir las características de un conjunto de datos. Utiliza estadísticas básicas y visualizaciones para proporcionar una visión clara de lo que ha sucedido en el pasado (Guevara et al., 2020).

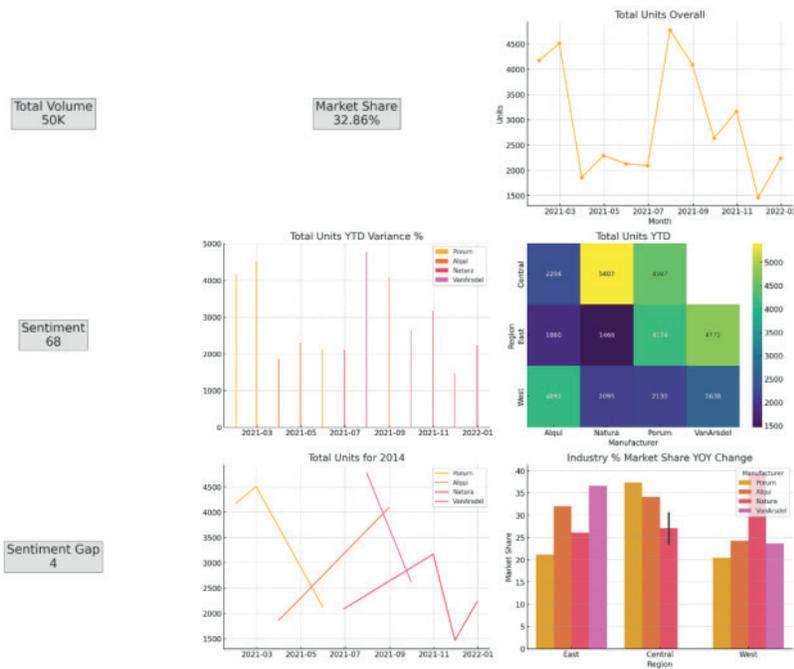
- **Estadísticas descriptivas:** Incluyen medidas fundamentales como la media, que proporciona el promedio de los datos; la mediana, que indica el valor central de un conjunto de datos ordenados; la moda, que representa el valor más frecuente; la desviación estándar, que mide la dispersión de los datos en relación con la media; y los percentiles, que dividen los datos en 100 partes iguales para mostrar la distribución.

Figura 1. Ejemplos gráficos de estadísticas descriptivas.



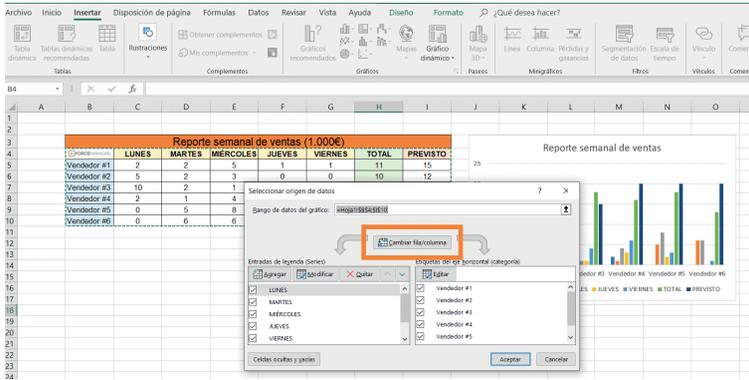
- **Visualizaciones de datos:** Utiliza diversas herramientas gráficas para representar datos de manera visualmente intuitiva. Los gráficos de barras comparan cantidades entre diferentes categorías, los histogramas muestran la distribución de una variable continua, los diagramas de caja y bigotes resaltan la mediana, los cuartiles y los valores atípicos de una variable, los gráficos de dispersión exploran la relación entre dos variables continuas, y los mapas de calor indican la intensidad de valores a través de una matriz de colores (Shu, 2020).

Figura 2. Dashboard resultado de análisis de datos.



- **Tablas dinámicas:** Son herramientas interactivas que permiten a los usuarios reorganizar y resumir grandes conjuntos de datos de forma dinámica. Facilitan el análisis de datos mediante la creación de resúmenes y la visualización de patrones y tendencias al agrupar, filtrar y ordenar la información de acuerdo a diferentes criterios. Su capacidad para convertir grandes volúmenes de datos en información clara y comprensible la convierte en una herramienta esencial para la toma de decisiones (Malek, 2023).

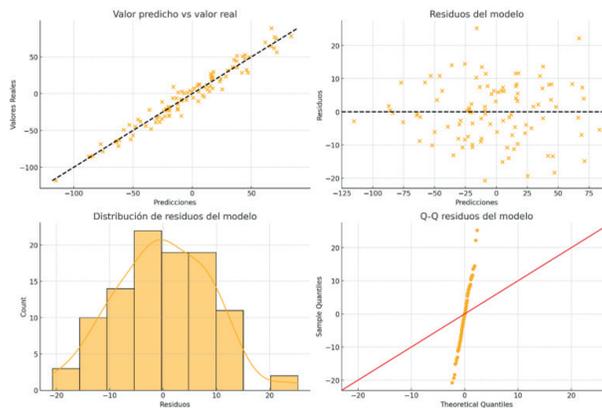
Figura 3. Tablas dinámicas utilizando MS EXCEL.



**2) Análisis Predictivo:** El análisis predictivo utiliza modelos estadísticos y algoritmos de machine learning para predecir futuros eventos basados en datos históricos.

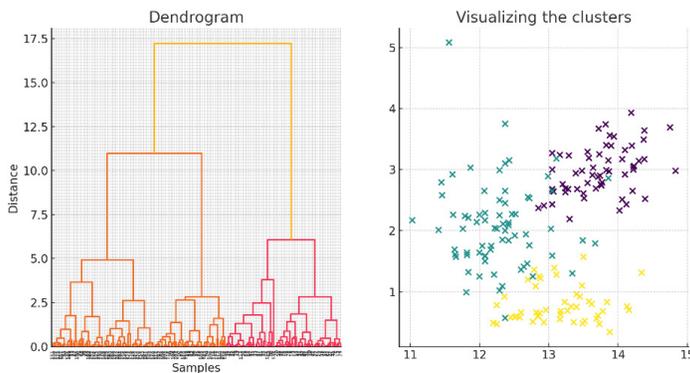
- Regresión:** La regresión es una técnica utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. Entre las técnicas de regresión, la regresión lineal es la más sencilla y común. Su objetivo es modelar la relación entre la variable dependiente y las variables independientes a través de una línea recta que se ajuste a los datos observados, siendo particularmente útil para predecir valores continuos (Zhang et al., 2020). La regresión logística, por otro lado, se utiliza para problemas de clasificación, permitiendo predecir clases binarias. Además, la regresión polinómica se aplica cuando la relación entre la variable dependiente y las variables independientes no es lineal, como una cuadrática o cúbica, proporcionando mayor flexibilidad en el modelado de datos complejos (Valadez y Ramírez, 2023).

Figura 4. Ejemplo gráfico del análisis de regresión.



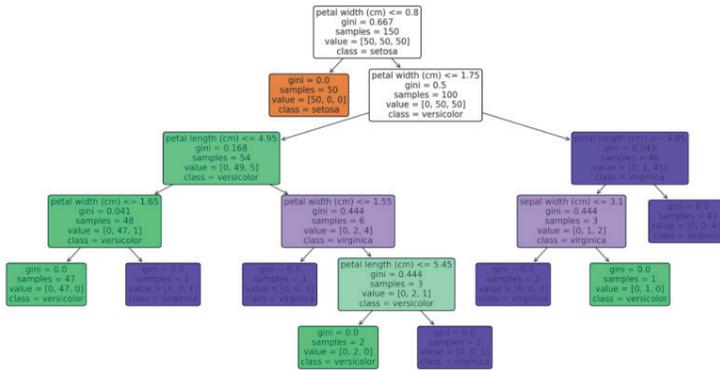
- Clustering:** El clustering es una técnica que organiza datos en grupos o clusters basados en la similitud de sus características. Entre las técnicas de clustering, K-means es especialmente popular debido a su capacidad para minimizar la variación dentro de cada cluster (Gao et al, 2023). Si bien es una de las formas más simples de agrupamiento, esta técnica también tiene sus limitaciones, debido a que se requiere especificar el número de clústeres previamente y además es sensible a los valores atípicos, lo que puede distorsionar los resultados (Raykov, 2016). Otras técnicas, como el clustering jerárquico, también agrupan datos similares, proporcionando diferentes enfoques para la identificación de patrones en conjuntos de datos. Esta técnica construye un conjunto de clústeres organizados en un árbol jerárquico llamado dendrograma, y a diferencia del K-means no requiere a priori una selección de número de clústeres (Shetty y Singh, 2021).

Figura 5. Ejemplos gráficos del análisis de clusters jerárquicos.



- Árboles de decisión:** Los árboles de decisión son modelos gráficos que dividen los datos en subconjuntos basados en valores de atributos, utilizados tanto para clasificación como para regresión (Matzavela y Alepis, 2021). Estos modelos emplean una estructura similar a un árbol para tomar decisiones basadas en reglas derivadas de los datos. Son fáciles de interpretar y pueden manejar de manera efectiva tanto problemas de clasificación como de regresión (Zeng, 2017). Los tipos más comunes de árboles de decisión son los árboles de clasificación y los árboles de regresión. El primero se utiliza cuando se desea predecir una clase o una categoría específica, mientras que el segundo se emplea cuando el objetivo es predecir un valor numérico continuo.

Figura 6. Ejemplo gráfico del análisis con árboles de decisión.



**3) Análisis Prescriptivo:** El análisis prescriptivo va un paso más allá al sugerir acciones específicas basadas en los resultados del análisis descriptivo y predictivo. Derivando en los siguientes análisis (Espino y Martínez, 2017):

- **Optimización:** Utiliza algoritmos de programación lineal y no lineal para encontrar la mejor solución a un problema, dadas ciertas restricciones.
- **Simulación:** Modela diferentes escenarios para evaluar posibles resultados y decisiones.
- **Algoritmos de recomendación:** Utilizan técnicas como la filtración colaborativa para sugerir productos o servicios a los usuarios.

### 3.1.2 Herramientas y software utilizados en Análisis de Datos

Varias herramientas y lenguajes de programación son esenciales para los analistas de datos. Algunas de las herramientas más populares incluyen:

- **Python:** Un lenguaje de programación versátil y poderoso, ampliamente utilizado en análisis de datos debido a su facilidad de uso y su extensa colección de bibliotecas especializadas. Con bibliotecas como pandas, numpy, scikit-learn y matplotlib que facilitan el análisis y visualización de datos (Chimarro et al., 2023).

Figura 7. Logo y página de Python.



<https://www.python.org/>

- **R:** Un lenguaje de programación y entorno de software especializado en estadística y análisis de datos. Es especialmente valorado en la comunidad académica y entre los estadísticos debido a sus capacidades avanzadas de análisis y visualización, con paquetes como ggplot2 para la visualización y caret para el machine learning (Fernández, 2020).

Figura 8. Logo y página de R.



<https://cran.rstudio.com/>

---

- **SQL:** Lenguaje estándar para la gestión y manipulación de bases de datos relacionales. Es esencial para la consulta, actualización e inserción de datos en bases de datos, permitiendo realizar operaciones complejas de manera eficiente y segura (Cruz et al., 2018). SQL es fundamental para extraer y preparar datos almacenados en sistemas de gestión de bases de datos relacionales (RDBMS) como MySQL, PostgreSQL y SQL Server.
- **Tableau:** Una potente herramienta de visualización de datos que permite a los usuarios crear dashboards interactivos. Tableau facilita la conexión a diversas fuentes de datos, la transformación y limpieza de los datos, y la creación de visualizaciones interactivas que pueden compartirse fácilmente (Milligan, 2022). Es muy utilizada en el entorno empresarial para el análisis visual y la toma de decisiones basadas en datos.
- **Power BI:** Es una herramienta de análisis y visualización de datos desarrollada por Microsoft, que permite a los usuarios convertir datos sin procesar en información significativa mediante la creación de informes y dashboards interactivos (Gonçalves et al., 2023). Esta herramienta es particularmente útil en el campo empresarial, cuando se desea analizar grandes volúmenes de datos y obtener información relevante en tiempo real.
- **Excel:** Herramienta de hoja de cálculo que sigue siendo popular para análisis de datos básicos y visualización. Además, Excel es accesible y conocido por una amplia base de usuarios, lo que facilita su integración en diversos procesos de negocio (Herruzo et al., 2019).

## 4 PROCESAMIENTO DE DATOS Y PREPROCESAMIENTO

El procesamiento y preprocesamiento de datos son etapas sensibles en el ciclo de vida de Analítica de Datos. Estas etapas aseguran que los datos estén en la forma adecuada para ser analizados y que cualquier problema de calidad de los datos se aborde de manera efectiva.

### 4.1 RECOLECCIÓN DE DATOS

La recolección de datos es la primera etapa del ciclo de vida de Analítica de Datos. Consiste en la recopilación de datos brutos que serán utilizados en el análisis (Stobierski, 02 de febrero de 2021).

- **Definición de objetivos:** Antes de comenzar la recolección de datos, es crítico definir claramente los objetivos del análisis. Esto ayudará a determinar qué datos son necesarios y de qué fuentes deben obtenerse.
- **Identificación de fuentes de datos:**
  - **Internas:** Sistemas de gestión empresarial, CRM, ERP, bases de datos transaccionales.
  - **Externas:** APIs públicas, datos de redes sociales, datos abiertos gubernamentales, datos de proveedores y clientes.
- **Metodología de recolección:**
  - **Automatización:** Utilizar scripts y herramientas automatizadas para la recolección de datos desde APIs y otras fuentes digitales.
  - **Manual:** Recolección manual de datos mediante encuestas, entrevistas y observaciones directas cuando la automatización no es posible.
- **Garantía de calidad de datos:**
  - **Verificación:** Asegurarse de que los datos recolectados son precisos y completos. Esto puede incluir la verificación cruzada con múltiples fuentes.
  - **Integridad:** Comprobar que los datos no están corruptos y que no faltan registros importantes.
- **Formato de datos:**
  - **Estandarización:** Convertir los datos recolectados en un formato estándar que sea compatible con las herramientas de análisis.

- **Documentación:** Mantener una documentación detallada de las fuentes de datos, los métodos de recolección y cualquier transformación inicial realizada.
- **Aspectos éticos y legales:**
  - **Privacidad:** Asegurarse de que la recolección de datos cumple con las leyes y regulaciones de privacidad, como el GDPR en Europa.
  - **Consentimiento:** Obtener el consentimiento informado de los individuos cuando se recolectan datos personales.

## 4.2 LIMPIEZA Y TRANSFORMACIÓN DE DATOS

La limpieza y transformación de datos es una etapa esencial en el ciclo de vida de Analítica de Datos, que implica la identificación y corrección de errores y anomalías en el conjunto de datos. Este proceso asegura que los datos estén en la forma adecuada para ser analizados, mejorando la calidad y la precisión del análisis posterior. Las principales actividades en esta etapa incluyen (Chu et al., 2016):

- **Eliminación de duplicados:** Identificar y eliminar filas duplicadas que se pueden haber introducido durante la recolección de datos. Los duplicados pueden sesgar los resultados del análisis y llevar a conclusiones incorrectas.
- **Corrección de valores atípicos:** Identificar y manejar outliers, que son valores que se desvían del resto de datos. Los outliers pueden ser corregidos mediante reemplazo con valores más representativos o la eliminación si se consideran errores. Se pueden emplear diferentes herramientas para la identificación de valores atípicos entre las cuales se incluyen la desviación estándar, el rango intercuartílico, así como los gráficos de caja y de dispersión. Los outliers pueden distorsionar el análisis estadístico y los modelos predictivos, afectando la fiabilidad de los resultados.
- **Manejo de datos faltantes:** Imputar o eliminar datos faltantes para asegurar un análisis robusto. Los datos faltantes pueden introducir sesgos y reducir la precisión del análisis si no se manejan adecuadamente.
- **Corrección de errores tipográficos y inconsistencias:** Estandarizar formatos y corregir errores en los datos, como inconsistencias en la denominación de categorías o errores de entrada. Las inconsistencias pueden dificultar el análisis y la integración de datos, afectando la calidad de los resultados.

**a) Técnicas de imputación de datos faltantes:** Existen varias técnicas para manejar datos faltantes, cada una adecuada dependiendo de la naturaleza de los datos y el contexto del análisis (Seu et al., 2022):

- **Eliminación de datos faltantes:** Eliminar filas o columnas con valores faltantes si la cantidad de datos faltantes es pequeña y no afecta significativamente el análisis. Esta técnica es simple y puede ser efectiva cuando los datos faltantes son mínimos, pero puede introducir sesgos si se eliminan demasiados datos.
- **Imputación media/mediana:** Rellenar valores faltantes con la media o mediana de la columna. Esta técnica es útil para mantener la integridad de los datos y es fácil de implementar, aunque puede suavizar las variaciones y no capturar patrones complejos en los datos, además no considera las relaciones entre diferentes variables, por lo que, si existen dependencias, este enfoque no podría detectarlas, lo que puede llevar a conclusiones erróneas en modelos predictivos.
- **Imputación mediante regresión:** Usar un modelo de regresión para predecir los valores faltantes basándose en otras variables del conjunto de datos. Esta técnica puede proporcionar estimaciones más precisas al considerar la relación entre variables, pero requiere un modelo bien ajustado y puede ser más compleja de implementar.
- **Métodos avanzados:** Como el uso de algoritmos de machine learning, árboles de decisión y técnicas de clústering, para predecir los valores faltantes. Estos métodos pueden capturar patrones más complejos y proporcionar estimaciones más precisas, aunque requieren conocimientos avanzados y más recursos computacionales.

**b) Normalización y escalado de datos:** La normalización y el escalado son técnicas utilizadas para ajustar los datos a un rango específico o a una distribución estándar. Esto es importante, especialmente cuando se utilizan algoritmos de machine learning que son sensibles a la escala de los datos (Peshawa et al., 2014).

- **Normalización (Min-Max Scaling):** Escala los datos a un rango específico, generalmente entre 0 y 1, transformando cada valor según la fórmula  $(X - X_{min}) / (X_{max} - X_{min})$ . Esta técnica es útil para algoritmos que dependen de distancias métricas, como el clustering y la regresión, asegurando que todas las variables contribuyan de manera equitativa. A pesar de su utilidad, presenta ciertas limitaciones dado que este método es sensible a los *outliers*.

- **Estandarización (Z-score Scaling):** Transforma los datos para que tengan una media de 0 y una desviación estándar de 1, utilizando la fórmula:  $(X-\mu) / \sigma$ , donde  $\mu$  es la media y  $\sigma$  es la desviación estándar. La estandarización es importante para algoritmos que asumen que los datos siguen una distribución normal, como la regresión lineal y las redes neuronales, permitiendo una mejor convergencia y rendimiento del modelo.

## 5 ANÁLISIS DE DATOS: ANÁLISIS EXPLORATORIO DE DATOS (EDA)

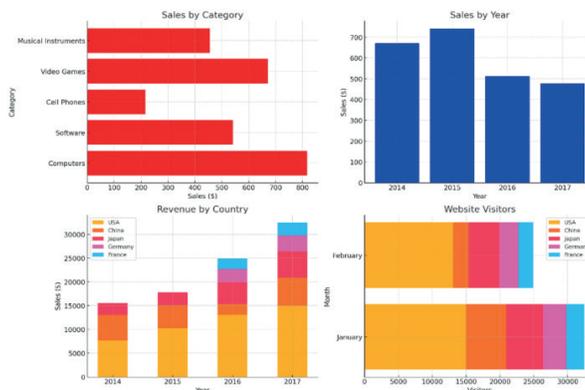
El Análisis Exploratorio de Datos (EDA) es una fase crítica en el proceso de Analítica de Datos. Su objetivo es resumir las principales características de un conjunto de datos y ganar una comprensión inicial de su estructura y relaciones subyacentes. EDA utiliza métodos visuales y estadísticos para descubrir patrones, detectar anomalías, probar hipótesis y comprobar supuestos (Rao et al., 2021). Este análisis preliminar es básico para identificar la naturaleza de los datos y determinar los pasos posteriores en el análisis más complejo.

### 5.1 VISUALIZACIÓN DE DATOS

Las visualizaciones de datos son herramientas poderosas para comprender datos complejos de manera intuitiva. A continuación, se describen algunas técnicas comunes de visualización y sus usos:

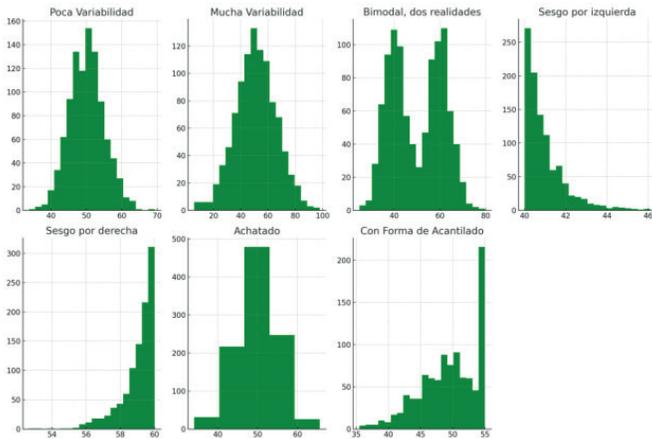
- **Gráficos de barras:** Utilizan barras rectangulares para representar datos categóricos. La longitud de cada barra es proporcional al valor que representa. Son útiles para comparar cantidades entre diferentes categorías, como las ventas por región, el rendimiento por departamento, o la frecuencia de eventos (Sheikh, 2023).

Figura 9. Gráfico de barras.



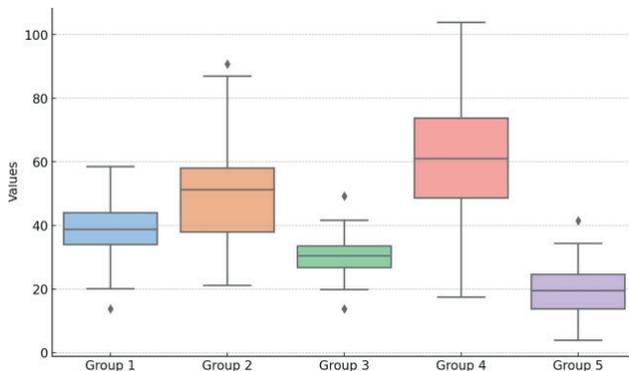
- **Histogramas:** Muestran la distribución de una variable continua dividiendo el rango de valores en intervalos (bins) y contando cuántos valores caen en cada intervalo, es adecuado para visualizar la distribución de datos numéricos continuos, identificar la presencia de sesgos, y detectar valores atípicos (Boels, L. et al., 2019).

Figura 10. Gráfico de barras.



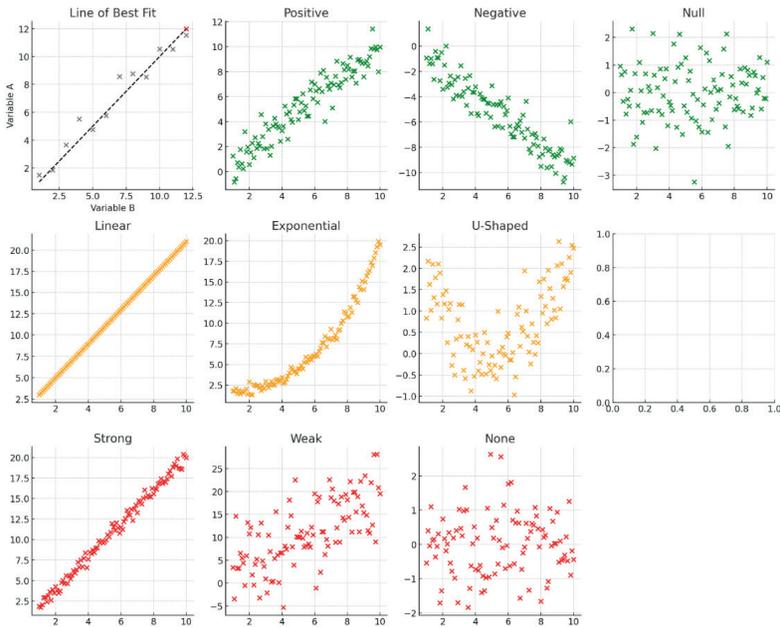
- **Diagramas de caja y bigotes (Boxplots):** Representan la distribución de datos en base a cinco números resumen: mínimo, primer cuartil (Q1), mediana (Q2), tercer cuartil (Q3), y máximo. Los bigotes indican el rango de los datos, y los puntos fuera de los bigotes son outliers. Resaltan la mediana, los cuartiles y los outliers de una variable, siendo útiles para comparar distribuciones entre diferentes grupos. Los boxplots facilitan la comparación rápida y precisa de las características estadísticas, así como la identificación de patrones o anomalías en los datos (Hu, k., 2020).

Figura 11. Gráficos de caja y bigotes.



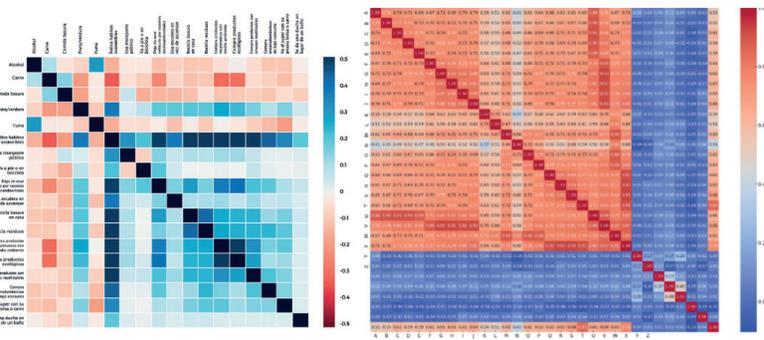
- **Gráficos de dispersión (Scatter plots):** Muestran la relación entre dos variables continuas mediante puntos individuales en un plano bidimensional. Utilizados para explorar la relación entre dos variables continuas, como por ejemplo la correlación entre la altura y el peso, o el análisis de regresión (Kiesel et al., 2023).

Figura 12. Gráfico de dispersión.



- **Mapas de calor (Heatmaps):** Utilizan colores para representar la intensidad de valores a través de una matriz de colores. Cada celda en el mapa de calor tiene un color que corresponde a un valor de la variable. Muestran la intensidad de valores en una matriz, siendo útiles para identificar patrones, correlaciones y áreas de alta concentración de datos, lo que resulta útil para detectar tendencias que podrían no ser evidentes en tablas o gráficos tradicionales (Peña y Trujillo, 2019).

Figura 13. Mapas de calor.



## 5.2 IDENTIFICACIÓN DE PATRONES Y TENDENCIAS

El análisis de tendencias y patrones ayuda a identificar relaciones importantes dentro de los datos. Algunas técnicas clave incluyen (Cueva et al., 2023):

- **Análisis de correlación:** Mide la fuerza y dirección de la relación lineal entre dos variables. La correlación se expresa mediante el coeficiente de correlación, que varía entre -1 y 1. Es útil para identificar relaciones lineales y puede orientar la selección de variables para modelos predictivos.
- **Detección de tendencias:** Usa gráficos de series temporales para identificar patrones a lo largo del tiempo, como tendencias ascendentes, descendentes o estacionales. Permite observar cómo cambian las variables a lo largo del tiempo, facilitando la previsión de tendencias futuras y la toma de decisiones estratégicas.
- **Análisis de segmentos:** Divide los datos en grupos o segmentos para encontrar patrones específicos dentro de subgrupos. Esto puede hacerse mediante técnicas como el clustering. Es útil para identificar segmentos de clientes con comportamientos similares, optimizar estrategias de marketing, y personalizar ofertas de productos.

## 5.3 TÉCNICAS ESTADÍSTICAS BÁSICAS PARA EDA

Además de las visualizaciones, el EDA emplea técnicas estadísticas para explorar los datos (Hussain y Athal, 2023):

- **Estadísticas resumen:** Entre las que se destaca:
  - **Media:** Calcula el promedio de los datos, proporcionando una medida central, aunque es sensible a valores atípicos.

- **Mediana:** El valor central que divide el conjunto de datos en dos mitades iguales. La mediana es particularmente útil para datos sesgados, ya que no se ve afectada por valores extremos.
- **Desviación estándar:** Mide la dispersión de los datos en torno a la media, indicando la variabilidad de los datos.
- **Percentiles:** Indican la posición relativa de un valor dentro de un conjunto de datos, útiles para entender la distribución y la dispersión.
- **Pruebas de normalidad:** Evalúan si los datos siguen una distribución normal. Las pruebas comunes incluyen la prueba de Shapiro-Wilk y la prueba de Kolmogorov-Smirnov. Es importante para determinar la aplicabilidad de ciertas técnicas estadísticas que asumen normalidad en los datos.
- **Análisis de varianza (ANOVA):** Compara las medias entre tres o más grupos para determinar si al menos una media es significativamente diferente de las demás. Es útil para evaluar diferencias significativas entre grupos en estudios experimentales y en la comparación de múltiples categorías de datos.

## 6 ANÁLISIS DE DATOS: MODELADO Y EVALUACIÓN

### 6.1 MODELOS MÁS COMUNES

El modelado y la evaluación son fases fundamentales en Analítica de Datos, donde se construyen modelos predictivos y se evalúa su rendimiento para asegurar su eficacia y fiabilidad. En esta sección, se explorarán los tipos de modelos más comunes y las técnicas utilizadas para evaluar su desempeño (Lamba y Madhusudhan, 2022).

**1) Regresión Lineal:** La regresión lineal asume una relación lineal entre la variable dependiente ( $y$ ) y una o más variables independientes ( $X$ ) (Shi, 2023).

- **Descripción:** El modelo se representa como  $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$ , donde  $\beta$  son los coeficientes que se ajustan durante el entrenamiento y  $\epsilon$  es el término de error que captura las variaciones no explicadas por el modelo.
- **Aplicaciones:** Predicción de valores continuos como precios de casas, ingresos, costos, etc.
- **Ventajas:** Simplicidad y facilidad de interpretación.
- **Limitaciones:** Supone una relación lineal y puede no funcionar bien con relaciones no lineales o con datos con muchos outliers.

**2) Árboles de Decisión:** Modelos que utilizan una estructura en forma de árbol para tomar decisiones basadas en reglas derivadas de los datos (Shu, 2020).

- **Descripción:** Utiliza una estructura en forma de árbol para dividir los datos en subconjuntos basados en valores de atributos, creando una serie de reglas de decisión.
- **Aplicaciones:** Clasificación de clientes, segmentación de mercado, predicción de resultados.
- **Ventajas:** Fácil de interpretar y visualizar, maneja relaciones no lineales.
- **Limitaciones:** Propenso al sobreajuste si no se poda correctamente, y es sensible a la variabilidad de los datos lo que puede comprometer la estabilidad del modelo.

**3) Random Forest:** Un conjunto de árboles de decisión que mejora la precisión mediante la agregación de múltiples modelos (Chen y Liu, 2024).

- **Descripción:** Un conjunto de múltiples árboles de decisión que mejora la precisión mediante el promedio de las predicciones de varios árboles (bagging).
- **Aplicaciones:** Clasificación de texto e imágenes, detección de fraude, predicción de enfermedades, análisis de crédito.
- **Ventajas:** Reduce el riesgo de sobreajuste, maneja relaciones no lineales y datos faltantes.
- **Limitaciones:** Más complejo y puede ser un poco más difícil de interpretar; y computacionalmente costoso.

**4) Regresión Logística:** Modelo utilizado para problemas de clasificación binaria, que estima la probabilidad de que una observación pertenezca a una clase particular (Campbell y Jacques, 2023).

- **Descripción:** Utilizado para problemas de clasificación binaria, estima la probabilidad de que una observación pertenezca a una clase particular utilizando una función logística.
- **Aplicaciones:** Predicción de eventos binarios como aprobaciones/rechazos, diagnósticos médicos.
- **Ventajas:** Simplicidad y facilidad de interpretación, útil para clasificación binaria.
- **Limitaciones:** No maneja bien problemas de clasificación con más de dos clases.

## 6.2 EVALUACIÓN Y VALIDACIÓN DE MODELOS

Una vez que se construye un modelo, es fundamental evaluar su rendimiento para asegurar que funcione bien no solo con los datos de entrenamiento sino también con

datos nuevos e inéditos (Tsiptsias et al., 2016). Se presenta a continuación, las técnicas de evaluación comunes:

**1) Matriz de Confusión:** Una matriz de confusión es una herramienta que se utiliza para evaluar el rendimiento de un modelo de clasificación. La matriz muestra las predicciones correctas e incorrectas desglosadas por cada clase. Contiene cuatro tipos de resultados: Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN) (Yeturu, 2020).

- **Aplicaciones:** Es especialmente útil para problemas de clasificación binaria, pero también se puede extender a problemas de clasificación multiclase.
- Interpretación:
  - TP (True Positive): Predicciones correctas de la clase positiva.
  - FP (False Positive): Predicciones incorrectas de la clase positiva.
  - TN (True Negative): Predicciones correctas de la clase negativa.
  - FN (False Negative): Predicciones incorrectas de la clase negativa.

## 2) Curva ROC y AUC:

- **ROC (Receiver Operating Characteristic):** La curva ROC es una gráfica que ilustra la capacidad de un modelo para distinguir entre clases. Representa la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) a varios umbrales de clasificación (Hajian, 2013).
- **AUC (Area Under the Curve):** El área bajo la curva ROC (AUC) proporciona una única medida de rendimiento para la evaluación de modelos. Un AUC de 1 indica un modelo perfecto, mientras que un AUC de 0.5 indica un modelo sin capacidad discriminativa (Çorbacioğlu, y Aksel, 2023).
- **Aplicaciones:** Evaluar modelos de clasificación binaria.
- **Interpretación:**
  - AUC cercano a 1: Excelente capacidad de discriminación.
  - AUC cercano a 0.5: No mejor que el azar.

## 3) Precisión, Recall y F1 Score:

- **Precisión (Precisión):** Es la proporción de verdaderos positivos entre todas las predicciones positivas. Indica la exactitud de las predicciones positivas del modelo (Li et al., 2021).

$$Precisión = \frac{TP}{TP + FP}$$

- **Recall (Sensibilidad):** Es la proporción de verdaderos positivos entre todos los casos positivos reales. Indica la capacidad del modelo para encontrar todos los casos positivos (Li et al., 2021).

$$Recall = \frac{TP}{TP + FN}$$

- **F1 Score:** Es la media armónica de la precisión y el recall. Es útil cuando se necesita un equilibrio entre precisión y recall (Li et al., 2021).

$$F1 = 2 \times \frac{Precisión \times Recall}{Precisión + Recall}$$

- **Aplicaciones:** Evaluar modelos de clasificación, especialmente cuando hay una clase desbalanceada (Li et al., 2021).
- **Interpretación:**
  - Alta precisión: Pocos falsos positivos.
  - Alto recall: Pocos falsos negativos.

#### 4) Error Cuadrático Medio (MSE) y R<sup>2</sup>:

- **MSE:** Es el promedio de los errores al cuadrado entre los valores predichos y los valores reales. Es una medida de la calidad de un modelo de regresión. Un MSE menor indica un modelo más preciso (Hodson et al., 2021).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- **R<sup>2</sup> (Coeficiente de Determinación):** Indica la proporción de la variabilidad de la variable dependiente que es explicada por las variables independientes en el modelo. Un valor de R<sup>2</sup> cercano a 1 indica un buen ajuste del modelo (Chicco et al., 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

### 6.3 TÉCNICAS DE EVITACIÓN DE SOBREAJUSTE

El sobreajuste ocurre cuando un modelo se ajusta demasiado bien a los datos de entrenamiento, capturando el ruido en lugar de los patrones reales. Esto resulta en

un rendimiento pobre en datos nuevos. Para evitar el sobreajuste, se pueden emplear las siguientes técnicas:

**1) Validación Cruzada:** La validación cruzada divide los datos en múltiples subconjuntos (folds) y entrena el modelo en diferentes combinaciones de estos subconjuntos. La técnica más común es la validación cruzada k-fold, donde los datos se dividen en k partes iguales (Wang et al., 2023).

- **Aplicaciones:** Evaluar el rendimiento de un modelo y asegurar que generaliza bien.
- **Ventajas:** Proporciona una estimación más fiable del rendimiento del modelo en datos no vistos.

**2) Regularización:** La regularización añade una penalización a la función de costo para evitar coeficientes excesivamente grandes. Las dos formas más comunes de regularización son la Lasso (L1) y la Ridge (L2) (Nagpal, 2017).

- **Aplicaciones:** Mejorar la generalización del modelo y evitar sobreajuste.
- **Ventajas:** Ayuda a simplificar el modelo y prevenir el sobreajuste.

**3) Poda en Árboles de Decisión:** La poda es el proceso de reducir el tamaño del árbol de decisión para eliminar ramas que tienen poca importancia y que pueden estar sobre-ajustando los datos de entrenamiento (Priyanka y Kumar, 2020).

- **Aplicaciones:** Evitar el sobreajuste en modelos de árboles de decisión.
- **Ventajas:** Mejora la capacidad del modelo para generalizar a datos nuevos.

## 7 VISUALIZACIÓN DE DATOS

### 7.1 CORRECTA ELECCIÓN DE REPRESENTACIÓN GRÁFICA

Un gráfico estadístico es una representación visual de datos que busca evidenciar una realidad concreta. Sin embargo, transmitir un conjunto de datos de manera precisa y captar la atención no es tarea sencilla. Es fundamental facilitar la comparación entre los datos, destacar tendencias, evitar errores de visualización e ilustrar el mensaje deseado (Divecha et al., 2023). Para ello, es necesario elegir el tipo de visualización adecuado según la relación entre los datos. A continuación, se desglosan los gráficos más recomendados según el tipo de asociación entre variables numéricas (Financial Times, 2016):

**a) Desviación:** Subraya las variaciones numéricas desde un punto de referencia fijo (habitualmente cero, pero también puede ser un objetivo o un promedio). Útil para mostrar sentimientos (positivo, neutral o negativo).

- **Barra Divergente:** Maneja valores de magnitud tanto negativos como positivos.
- **Tabla de Columna:** Divide un valor único en dos componentes contrastables (e.g., masculino/femenino).

**b) Correlación:** Muestra la relación entre dos o más variables. Es importante aclarar que las relaciones mostradas no siempre son causales.

- **Gráfico de Dispersión:** Muestra la relación entre dos variables continuas, cada una con su propio eje.
- **Línea del Tiempo:** Muestra la relación entre una cantidad (columnas) y un ratio (línea).

**c) Clasificación:** Clasifica variables numéricas cuando la posición de un elemento en una lista ordenada es más importante que su valor absoluto o relativo.

- **Gráfico de Barras:** Muestra los rangos de valores de forma sencilla cuando se ordenan.
- **Diagrama de Tira de Puntos:** Los puntos están ordenados en una tira, ahorrando espacio para diseñar rangos en múltiples categorías.

**d) Distribución:** Resalta una serie de valores dentro de un conjunto de datos y representa con qué frecuencia ocurren, ayudando a identificar valores atípicos y tendencias.

- **Histograma:** Muestra una distribución estadística, destacando la “forma” de los datos.
- **Gráfico de Cajas:** Visualiza distribuciones múltiples mostrando la mediana y el rango de los datos.
- **Pirámide Poblacional:** Muestra la distribución de la población por sexo, combinando dos gráficos de barras horizontales.

**e) Cambios en el Tiempo:** Enfatiza tendencias cambiantes a lo largo del tiempo.

- **Línea:** Muestra una serie temporal cambiante. Los datos irregulares pueden representarse con marcadores.
- **Mapa de Calor Calendario:** Muestra patrones temporales (diario, semanal, mensual) y requiere precisión en la cantidad de datos.

**f) Magnitud:** Visibiliza comparaciones de tamaño, ya sean relativas o absolutas.

- **Columnas:** Compara el tamaño de las cosas, con el eje siempre comenzando en 0.

- **Gráfico de Mekko:** Muestra tamaño y proporción de los datos simultáneamente, ideal para datos no muy complejos.
- g) Parte de un Todo:** Muestra cómo una entidad se divide en sus componentes, comúnmente utilizado para representar presupuestos o resultados electorales.
- **Gráfico de Pastel:** Muestra datos parciales o completos. No es fácil comparar el tamaño de los distintos segmentos.
- **Diagrama de Venn:** Muestra interrelaciones o coincidencias.
- h) Espacial:** Visualiza datos donde las ubicaciones precisas o los patrones geográficos son importantes.
- **Mapa Coroplético:** Coloca datos en un mapa.
- **Mapa de Flujo:** Muestra movimientos dentro de un mismo mapa, como movimientos migratorios.

Figura 14. Recomendación de utilización de gráficos.



Fuente: <https://datos.gob.es/es/blog/como-elegir-el-grafico-correcto-para-visualizar-datos-abiertos>, Basado de presentación más detallada del Financial Times: <https://github.com/Financial-Times/chart-doctor/blob/main/visual-vocabulary/Visual-vocabulary-es.pdf>

## 7.2 GENERACIÓN DE INSIGHTS Y TOMA DE DECISIONES

Conocer las diferentes opciones de representación estadística ayuda a crear visualizaciones más precisas, permitiendo una comprensión clara de la realidad evidenciada. En un contexto donde la información visual tiene cada vez más peso, es clave desarrollar herramientas que faciliten la transmisión de la información contenida en los datos a la ciudadanía, contribuyendo así a mejorar la sociedad.

La elección adecuada de la representación gráfica es fundamental para comunicar efectivamente los resultados de un análisis. Una correcta selección de gráficos no solo facilita la interpretación de los datos, sino que también resalta los insights obtenidos, permitiendo una comprensión más clara y profunda de la información. Esto, a su vez, es importante para informar de manera efectiva a los tomadores de decisiones. Con una visualización clara y precisa, se pueden identificar tendencias, patrones y anomalías que de otro modo podrían pasar desapercibidos. Este proceso es esencial para fundamentar decisiones estratégicas y operativas en un negocio, optimizando así los resultados y fomentando un enfoque basado en datos para la planificación y ejecución de estrategias empresariales.

## 8 IDEAS FINALES

Este trabajo hace un esfuerzo por proporcionar una visión comprensiva y estructurada sobre el Data Analysis, cubriendo los aspectos fundamentales hasta sus aplicaciones prácticas en diversos contextos empresariales. A continuación, se destacan las conclusiones más relevantes derivadas del contenido del capítulo:

- 1) Importancia del Analítica de Datos:** El análisis de datos se ha convertido en una herramienta esencial para las organizaciones, permitiendo transformar grandes volúmenes de datos en conocimientos accionables. Esto facilita la toma de decisiones informadas y estratégicas que pueden optimizar resultados y mejorar la eficiencia operativa.
- 2) Evolución Histórica:** La evolución del Analítica de Datos ha estado marcada por avances tecnológicos significativos desde el desarrollo de la estadística moderna en el siglo XIX hasta la reciente incorporación de inteligencia artificial avanzada y el análisis de datos en la nube. Cada etapa ha contribuido a la sofisticación y capacidades actuales de la analítica de datos.
- 3) Diferenciación Conceptual:** Es importante entender las diferencias entre Business Intelligence (BI), Analítica de Datos y Data Science, ya que cada uno tiene un enfoque y propósito específico. Mientras que BI se centra en la recopilación y visualización de datos históricos, Data Science explora y

predice futuros eventos mediante técnicas avanzadas, y Analítica de Datos actúa como un puente entre ambos, proporcionando un análisis integral.

- 4) **Metodologías Clave:** Las metodologías de análisis de datos se dividen principalmente en descriptiva, predictiva y prescriptiva. Cada una de estas metodologías utiliza técnicas y herramientas específicas para cumplir diferentes propósitos dentro del análisis de datos, desde describir lo que ha sucedido hasta predecir futuros eventos y recomendar acciones específicas.
- 5) **Herramientas y Software:** Herramientas como Python, R, SQL, Tableau y Excel son fundamentales para los analistas de datos. Estas herramientas facilitan desde la recolección y limpieza de datos hasta su análisis y visualización, permitiendo un manejo eficiente y eficaz de grandes volúmenes de información.
- 6) **Procesamiento y Preprocesamiento:** El procesamiento y preprocesamiento de datos son etapas críticas en el ciclo de vida del Analítica de Datos. Estas etapas aseguran que los datos estén en la forma adecuada para ser analizados y que cualquier problema de calidad de los datos se aborde de manera efectiva, mejorando la precisión y fiabilidad del análisis.
- 7) **Análisis Exploratorio de Datos (EDA):** El EDA es una fase esencial para entender la estructura y las relaciones subyacentes de un conjunto de datos. Utiliza métodos visuales y estadísticos para descubrir patrones, detectar anomalías y probar hipótesis, proporcionando una base sólida para análisis más complejos.
- 8) **Modelado y Evaluación:** El modelado y la evaluación de modelos predictivos son fundamentales para asegurar su eficacia y fiabilidad. Técnicas como la regresión, árboles de decisión, Random Forest y regresión logística, junto con métodos de evaluación como la matriz de confusión y la curva ROC, son esenciales para construir modelos robustos y precisos.
- 9) **Visualización de Datos:** La correcta elección de la representación gráfica es determinante para comunicar efectivamente los resultados de un análisis. Una visualización clara y precisa facilita la identificación de tendencias, patrones y anomalías, permitiendo una comprensión más profunda de la información y fundamentando decisiones estratégicas y operativas en un negocio.

En resumen, el análisis de datos es una disciplina multidimensional que abarca diversas técnicas y herramientas para extraer información significativa y útil de grandes volúmenes de datos. La evolución tecnológica continua, junto con la adopción creciente

de la analítica avanzada en la nube y la inteligencia artificial, seguirá transformando la manera en que las organizaciones utilizan los datos para tomar decisiones informadas y estratégicas. Este capítulo proporciona una guía sólida y accesible para entender los fundamentos y aplicaciones del Analítica de Datos, preparando a los lectores para enfrentar los desafíos y aprovechar las oportunidades en el campo de la analítica de datos.

## 9 AGRADECIMIENTOS

Este capítulo ha sido posible gracias al patrocinio del proyecto de investigación “Gestión de empresas y grupos de interés hacia la sostenibilidad desde la responsabilidad social empresarial”, proyecto que está adscrito a la carrera de Administración de Empresas de la UMET, Sede Machala; y se contó con la asistencia de la Red de Investigación Latinoamericana en Competitividad de Organizaciones (RILCO).

## REFERENCIAS

Affenzeller, M., Bögl, M., Fischer, L., Sobieczky, F., Yang, K., & Zenisek, J. (2022). *Prescriptive Analytics: When Data- and Simulation-based Models Interact in a Cooperative Way*. 2022 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 1-8. <https://doi.org/10.1109/SYNASC57785.2022.00009>

Barón R., García, E., y Sánchez, G. (2021). La inteligencia de negocios y la analítica de datos en los procesos empresariales. *Revista científica de sistemas e informática*, 1 (2) 37-53. <https://doi.org/10.51252/rcsi.v1i2.167>

Bengfort, B. & Kim, J. (2016). *Data analytics with hadoop*. O'Reilly Media, Inc.

Bertsimas, D., y Van Parys, B. (2017). Bootstrap robust prescriptive analytics. *Mathematical Programming*, 195(1), 39-78. <http://doi.org/10.1007/s10107-021-01679-2>

Boels, L., Bakker, A., Van Dooren, W., y Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, 28, 100291. <https://doi.org/10.1016/j.edurev.2019.100291>

Campbell, M., y Jacques, R. (2023). *Multiple Logistic Regression*. In *Statistics at Square Two* (Eds M. Campbell and R. Jacques), Chapter 3. <https://doi.org/10.1002/9781119401407.ch3>

Chen, M., & Liu, Z. (2024). Predicting performance of students by optimizing tree components of random forest using genetic algorithm. *Heliyon*. 10 (12). <https://doi.org/10.1016/j.heliyon.2024.e32570>

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>

Chimarro, J., Chuqui, F., Guamán, D., y Quishpe, C. (2023). El auge exponencial del lenguaje Python en el desarrollo tecnológico. *Revista Científica INGENIAR*, 6 (12) 240-256. <https://journalingeniar.org/index.php/ingeniar/article/view/152/221>

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016, June). *Data cleaning: Overview and emerging challenges*. In Proceedings of the 2016 international conference on management of data (pp. 2201-2206). <https://doi.org/10.1145/2882903.2912574>

Conesa Caralt, J. (2012). *Introducción al business intelligence. Introducción al business intelligence*. Ediciones UOC.

Çorbacioğlu, Ş. K., y Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23(4), 195-198. [https://doi.org/10.4103/tjem.tjem\\_182\\_23](https://doi.org/10.4103/tjem.tjem_182_23)

Cruz-Rojas, G. A., Molina-Blandón. M. A., y Valdiri-Vinasco. V. (2018). Vigilancia tecnológica para la innovación educativa en el uso de bases de datos y plataformas de gestión de aprendizaje en la universidad del Valle, Colombia. *Revista de investigación, desarrollo e innovación*, 9(2). <https://doi.org/10.19053/20278306.v9.n2.2019.9175>

Cueva Luza, T., Jara Córdova, O., Arias Gonzáles, J., Flores Limo, F., y Balmaceda Flores, C. (2023). Métodos mixtos de investigación para principiantes. *Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú S.A.C*. <https://doi.org/10.35622/inudi.b.106>

Divecha, C. A., Tullu, M. S., & Karande, S. (2023). Utilizing tables, figures, charts and graphs to enhance the readability of a research paper. *Journal of Postgraduate Medicine*, 69(3), 125-131. [https://doi.org/10.4103/jpgm.jpgm\\_387\\_23](https://doi.org/10.4103/jpgm.jpgm_387_23)

Espino, C., y Martínez, X. (2017). *Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso* [Tesis de grado, Universitat Oberta Catalunya] Repositorio UOC. <http://creativecommons.org/licenses/by-nc-nd/3.0/es/>

Fernández Lizana, M. I. (2020). Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales Rev. *Revista Científica de la UCSA*, 7(2), 97-111. <https://doi.org/10.18004/ucsa/2409-8752/2020.007.02.097>

Financial Times (2016). *Visual Vocabulary*. <https://github.com/Financial-Times/chart-doctor/blob/main/visual-vocabulary/Visual-vocabulary-es.pdf>

Friendly, M. (2008). *A Brief History of Data Visualization*. In: *Handbook of Data Visualization*. Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-33037-0\\_2](https://doi.org/10.1007/978-3-540-33037-0_2)

Gao, C. X., Dwyer, D., Zhu, Y., Smith, C. L., Du, L., Filia, K. M., ... y Cotton, S. M. (2023). An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research*, 327 (115265), 1-28. <https://doi.org/10.1016/j.psychres.2023.115265>

Gathani, S., Liu, Z., Haas, P. J., & Demiralp, Ç. (2022). *Understanding Business Users' Data-Driven Decision-Making: Practices, Challenges, and Opportunities*. arXiv preprint arXiv:2212.13643. <https://doi.org/10.48550/arXiv.2212.13643>

Gonçalves, C. T., Gonçalves, M. J. A., y Campante, M. I. (2023). Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform. *Information*, 14(11), 614. <https://doi.org/10.3390/info14110614>

Guevara, G., Verdesoto, A., y Castro, N. (2020). Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción). *Revista Científica mundo de la Investigación y el Conocimiento*, 4(3), 163-173. <https://recimundo.com/index.php/es/article/view/860>

- Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627–635. <https://doi.org/10.22088/cjim.4.2.627>
- Herruzo-Gómez, E., Hernández-Sánchez, B., Cadella, M., y Sánchez-García, J. (2019). *Emprendimiento e Innovación: Oportunidades para todos*. DYKINSON, S.L.
- Hodson, T. O., Over, T. M., & Foks, S. S. (2021). Mean squared error, deconstructed. *Journal of Advances in Modeling Earth Systems*, 13(12), e2021MS002681. <https://doi.org/10.1029/2021MS002681>
- Houtmeyers, K., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, 1-5. <https://doi.org/10.1123/ijsp.2020-0958>
- Hu, K. (2020). Become Competent within One Day in Generating Boxplots and Violin Plots for a Novice without Prior R Experience. *Methods and protocols*, 3(4), 64. <https://doi.org/10.3390/mps3040064>
- Hussain, T., y Athal, P. S. (2023). Visualization and Explorative Data Analysis. *International Journal of Enhanced Research in Science, Technology & Engineering*, 12(3). [https://www.researchgate.net/publication/369453742\\_Visualization\\_and\\_Explorative\\_Data\\_Analysis](https://www.researchgate.net/publication/369453742_Visualization_and_Explorative_Data_Analysis)
- Kiesel, D., Riehmman, P., y Froehlich, B. (2023). Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots. *In Computer Graphics Forum*, 42(6). <https://doi.org/10.1111/cgf.14923>
- Kumari, L., & Aggrawal, D. (2022). An Insight into Predictive Analytics Techniques. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2022.48071>.
- Lamba, M., y Madhusudhan, M. (2022). Predictive Modeling. *In Text Mining for Information Professionals: An Uncharted Territory* (pp. 213-242). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-85085-2\\_8](https://doi.org/10.1007/978-3-030-85085-2_8)
- Lederer, M., & Schmid, P. (2021). *Data Science for Business Analytics and Business Intelligence*, 495-508. <https://doi.org/10.4018/978-1-7998-3473-1.ch037>
- Lehrer, C., Wieneke, A., Vom Brocke, J. A. N., Jung, R., & Seidel, S. (2018). How big data analytics enables service innovation: materiality, affordance, and the individualization of service. *Journal of Management Information Systems*, 35(2), 424-460. <https://doi.org/10.1080/07421222.2018.1451953>
- Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *Int. J. Inf. Manag.*, 50, 57-70. <https://doi.org/10.1016/J.IJINFOMGT.2019.04.003>
- Li, Q., Zheng, S., Huang, Y., y Liu, D. (2021). Automatic Classification of NVST Short-exposure Data Based on Deep Learning. *Publications of the Astronomical Society of the Pacific*, 133(1020), 1–9. <https://www.jstor.org/stable/27123841>
- Malek, P. (2023, 3 de octubre). Your Guide to Using Looker Studio Pivot Tables. *COUPLER.IO by railsware*. <https://blog.coupler.io/looker-studio-pivot-table/>
- Marr, B. (2020). *Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things*. Kogan Page Publishers.

- Martín Romero, A., y Giménez Abad, M. (2020). *Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas* [Tesis de grado, Universidad Pontificia Comillas]. Repositorio Universidad Comillas. <http://hdl.handle.net/11531/45878>
- Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence*, 2, 100035. <https://doi.org/10.1016/j.caeai.2021.100035>
- Milligan, J. N., Hutchinson, B., Tossell, M., y Andreoli, R. (2022). *Learning Tableau 2022: Create effective data visualizations, build interactive visual analytics, and improve your data storytelling capabilities*. Packt Publishing Ltd.
- Nagpal, A. (2017, 13 octubre). *L1 and L2 Regularization Methods*. Towards Data Science. Towards Data Science. <https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c>
- Peña González, Z., y Trujillo Casañola, Y. (2019). Mapas de calor en sitios web. *Ediciones Futuro*, 12(9), 97-111. <https://publicaciones.uci.cu/index.php/serie/article/view/515>
- Peshawa, J., Muhammad A., Rezhna, F. (2014). Data normalization and standardization: a technical report. *Machines Learning Technical Reports*, 1(1), 1-6. [https://docs.google.com/document/d/1x0A1nUz1WWtMCZb5oVzF0SVMY7a\\_58KQulqQVT8LaVA/edit#heading=h.4ykexgmwxgcw](https://docs.google.com/document/d/1x0A1nUz1WWtMCZb5oVzF0SVMY7a_58KQulqQVT8LaVA/edit#heading=h.4ykexgmwxgcw)
- Priyanka, & Kumar, D. (2020). Decision tree classifier: a detailed survey. *International Journal of Information and Decision Sciences*, 12(3), 246-269. <https://doi.org/10.1504/IJIDS.2020.108141>
- Provost, F., & Fawcett, T. (2020). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Ramos, A. (2006). *Sistemas gestores de bases de datos. Grado Superior*. McGraw-Hill Interamericana de España. <https://www.mheducation.es/bcv/guide/capitulo/8448148797.pdf>
- Rao, S., Vardhan, V., y Shaik, H. (2021). Role of Exploratory Data Analysis in Data Science. *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1457-1461). IEEE. <https://doi.org/10.1109/ICCES51350.2021.9488986>
- Raykov, Y. P., Boukouvalas, A., Baig, F., y Little, M. A. (2016). What to do when K-means clustering fails: A simple yet principled alternative algorithm. *PLoS ONE*, 11(9), e0162259. <https://doi.org/10.1371/journal.pone.0162259>
- Roman, y Joanne. (2022). Data Science and Interdisciplinarity. *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, 35-49. [https://doi.org/10.1007/978-3-030-88389-8\\_3](https://doi.org/10.1007/978-3-030-88389-8_3)
- Seu, K., Kang, M. S., & Lee, H. (2022). An intelligent missing data imputation techniques: A review. *JOIV: International Journal on Informatics Visualization*, 6(1-2), 278-283. <http://dx.doi.org/10.30630/joiv.6.1-2.935>
- Sharda, R., Delen, D., y Turban, E. (2022). *Business Intelligence: A Managerial Perspective on Analytics* (4ª ed.). Pearson.
- Sharma, R., Sandhu, J., y Bharti, V. (2024). *Experimental Analysis of a Multimodal biometric System using Preprocessing and Feature Extraction Techniques and Their Impact on Analytical Results*. <https://doi.org/10.1109/ccict62777.2024.00043>
- Sheikh, M. (2023, 19 mayo). *Gráfico de barras: definición, ejemplos y cómo crear uno*. VISME. <https://visme.co/blog/es/grafico-de-barras/>

- Shetty, P., y Singh, S. (2021). Hierarchical Clustering: A Survey. *International Journal of Applied Research*, 7(4), 178-181. <https://doi.org/10.22271/allresearch.2021.v7.i4c.8484>
- Shi, Y. (2023). Application of Improved Linear Regression Algorithm in Business Behavior Analysis. *Procedia Computer Science*, 228, 1101-1109. <https://doi.org/10.1016/j.procs.2023.11.144>
- Shu, X. (2020). *Data Visualization. In Knowledge Discovery in the Social Sciences: A Data Mining Approach* (1st ed., pp. 70–90). University of California Press. <https://doi.org/10.2307/j.ctvw1d683.6>
- Shu, X. (2020). *Knowledge Discovery in the Social Sciences: A Data Mining Approach* (1st ed.). University of California Press. <https://doi.org/10.2307/j.ctvw1d683>
- Sidel, J. L., Bleibaum, R. N., y Tao, K. W. C. (2018). Quantitative descriptive analysis. In S. E. Kemp, J. Hort, & T. Hollowood (Eds.), *Descriptive analysis in sensory evaluation* (Cap. 8). <https://doi.org/10.1002/9781118991657.ch8>
- Stobierski, T. (2021, February, 02). *8 Steps in the data life cycle*. HBS Online. <https://online.hbs.edu/blog/post/data-life-cycle>
- Tsiptsias, N., Tako, A., & Robinson, S. (2016). Model validation and testing in simulation: a literature review. In *5th student conference on operational research (SCOR 2016)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik.
- Uskov, V., Bakken, J., Ganapathi, K., Gayke, K., Galloway, B., & Fatima, J. (2020). *Data Cleaning and Data Visualization Systems for Learning Analytics*, 183-197. [https://doi.org/10.1007/978-981-15-5584-8\\_16](https://doi.org/10.1007/978-981-15-5584-8_16)
- Valadez-García, A., y Ramírez Urquidy, M. A. (2023). *Modelos estadísticos para la investigación científica. Aplicaciones en las áreas económico-administrativas*. Editorial Comunicación Científica.
- Varghese, A., Joseph, C., Binoy, M., Jacob, R., & Joseph, A. (2023). Unleashing the Power of Data Analytics: A Pathway to Insightful Decision-Making. *International Journal of Engineering Technology and Management Sciences*. <https://doi.org/10.46647/ijetms.2023.v07i04.057>
- Vetter, T. (2017). Descriptive Statistics: Reporting the Answers to the 5 Basic Questions of Who, What, Why, When, Where, and a Sixth, So What?. *Anesthesia & Analgesia*, 125, 1797–1802. <https://doi.org/10.1213/ANE.0000000000002471>.
- Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2023). Spatial+: A new cross-validation method to evaluate geospatial machine learning models. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103364. <https://doi.org/10.1016/j.jag.2023.103364>
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, C., Wu, Y., Dong, F., Qiu, C., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., Roepman, R., & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4). <https://doi.org/10.1016/j.xinn.2021.100179>
- Yeturu, K. (2020). Machine learning algorithms, applications, and practices in data science. In *Handbook of Statistics*, 43, 81-206. <https://doi.org/10.1016/bs.host.2020.01.002>
- Zeng, J., Ustun, B., & Rudin, C. (2017). Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 180(3), 689–722. <http://www.jstor.org/stable/44682879>
- Zhang, X.-B., Xie, L., Zhang, X., y Hu, X. (2020). Regression Analysis. In *Energy Poverty and Household Heating Energy Transition in Rural China* (pp. 15–21). Environment for Development Initiative. <http://www.jstor.org/stable/resrep46934.7>

## SOBRE OS ORGANIZADORES



**Carolina Uzcátegui-Sánchez:** Ingeniera Empresarial por la Escuela Politécnica Nacional EPN, Magíster en Gestión Empresarial por la Universidad Técnica Particular de Loja UTPL, Doctorando en Ciencias Económicas, mención Administración en UNCUIYO (Argentina). Actualmente me desempeño como profesora titular agregado en Universidad Metropolitana, Sede Machala, y Coordinadora de la Maestría en Administración y Dirección de Empresas de la UMET Sede Machala. Profesora de pregrado y posgrado en Investigación de Mercados. Correo electrónico: cuzcategui@umet.edu.ec



**David Zaldumbide-Peralvo:** Candidato a Doctor en Ciencias Económicas por la Universidad Católica Andrés Bello y cuenta con una formación académica en investigación y economía. Tiene títulos en Metodologías de Investigación (Universidad de Córdoba, España), Economía Circular (Wageningen University, Países Bajos) y una Maestría en Economía (James Cook University, Australia). También es especialista en Gestión Ambiental y Marketing Digital (Universidad Blas Pascal, Argentina) e Ingeniero en Mercadotecnia (ESPE, Ecuador). Profesionalmente, ha sido Coordinador de Posgrado y Director Académico en la Asociación Ecuatoriana de Marketing y Comunicación. Su experiencia

abarca Administración, Estadísticas, Publicidad y Marketing. Como docente, ha impartido

cátedras en diversas áreas y es investigador registrado en SENESCYT. Además, ha publicado más de 60 artículos en SCOPUS, Web of Science, SciELO y Latindex, y ha participado en eventos científicos nacionales e internacionales.



**Emanuel Leite:** Posee una formación destacada con licenciatura, máster, doctorado y posdoctorado en instituciones de Brasil y Portugal. Ha servido en distintas capacidades académicas en la Universidad de Pernambuco, Universidad del Mar, Universidad Fernando Pessoa, entre otras. Ha escrito libros sobre emprendimiento y ha sido reconocido con varios premios, incluido el Premio Innovación de la Suerte y el Premio Emprendedor Profesor Emanuel Leite.

Tiene experiencia en administración, con especial interés en emprendimiento, innovación y empresas tecnológicas. Ha trabajado en la Oficina CEAG, fue miembro de consejos editoriales y núcleos educativos, y ha sido evaluador y revisor de diversas revistas académicas.

## ÍNDICE REMISSIVO

### A

Análisis competitivo 108, 114, 125, 129, 130, 134, 140

Análisis de datos 15, 16, 17, 21, 28, 33, 35, 36, 48, 50, 51, 58, 59, 60, 61, 62, 64, 65, 68, 69, 73, 77, 84, 85, 91, 93, 100, 101, 104, 110, 111, 116, 122, 129, 150, 165

Análítica de datos 58, 59, 60, 61, 62, 63, 64, 68, 70, 71, 73, 77, 84, 85, 86

Automatización en Marketing 1, 3

### B

Business Intelligence 58, 59, 60, 62, 84, 87, 88, 89

### C

ChatGPT 1, 2, 3, 8, 9, 10, 11, 12, 13, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 35, 36, 37, 38, 40, 41, 42, 43, 46, 48, 49, 50, 52, 53, 54, 55, 56, 94, 112, 113

Competitividad 23, 49, 52, 86, 102, 103, 111, 114, 120, 139, 140, 145, 146, 148, 150, 151, 153, 154, 156, 157, 165, 166, 168

### D

Data Science 58, 59, 60, 62, 63, 84, 88, 89, 90

Diferenciación 60, 84, 122, 123, 136, 138, 142, 145, 146, 148, 149, 158

### E

Estrategias 1, 2, 12, 13, 14, 15, 16, 17, 22, 33, 39, 40, 41, 42, 46, 48, 49, 50, 51, 61, 64, 76, 84, 99, 101, 102, 103, 105, 106, 107, 108, 110, 114, 115, 116, 117, 118, 119, 120, 121, 122, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 161, 162, 163, 164, 165, 166

### G

GPT personalizados 91, 92, 93, 94, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111

### I

IA 1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 25, 29, 30, 32, 33, 34, 37, 38, 39, 40, 43, 47, 48, 49, 50, 51, 52, 62, 91, 92, 93, 94, 97, 98, 99, 102, 103, 104, 105, 109, 111

Ingeniería de prompts 28, 38, 39, 40, 41, 42

Inteligencia artificial 2, 24, 28, 29, 48, 51, 53, 58, 59, 60, 62, 84, 86, 91, 92, 93, 94, 96, 116, 120

Investigación de mercados 1, 2, 18, 22, 28, 33, 34, 35, 36, 37, 38, 40, 48, 49, 50, 51, 91, 93, 95

## M

Machine Learning 5, 8, 22, 23, 26, 27, 30, 45, 53, 58, 59, 60, 61, 62, 63, 66, 69, 72, 89, 90, 112

Marketing digital 12, 14, 15, 17, 21, 22, 107, 140, 143, 145, 146, 148, 150, 151, 152, 153, 156, 157, 158, 159, 160, 162, 165, 166, 167, 168

Marketing estratégico 114, 115, 124, 127, 130, 134

Modelos de Lenguaje de gran tamaño (LLM) 91

## O

Optimización de prompts 28

## S

Sector bancario 114, 118, 124, 125, 127, 128, 129, 130, 132, 133, 138, 139, 140, 142, 143, 144