

Tabla de contenidos

Prólogo	xvii
Para quién es este libro	xvii
Cómo está organizado	xvii
Cómo usar este libro	xviii
Qué hace diferente este libro	xix
Un estilo deliberado: programación funcional aplicada	xix
Una nota sobre la terminología	xix
Alineación con el currículo oficial	xx
Las cajas de colores	xxi
El repositorio complementario	xxii
Una nota sobre el esfuerzo	xxii
I. Parte I: Fundamentos	1
1. El Detective de Datos	3
1.1. ¿Por qué la ciencia de datos cambia el mundo?	3
1.2. La Analogía del Detective	4
1.3. Roles Profesionales: Analista vs. Científico de Datos	5
1.4. Qué es la Ciencia de Datos	5
1.4.1. Una Fusión de Disciplinas	5
1.4.2. El Origen del Big Data	6
1.5. Las 8 Disciplinas de la Ciencia de Datos	7
1.5.1. La tabla maestra de disciplinas	8
1.5.2. Autoevaluación de las 8 Disciplinas	10
1.6. Casos de Éxito que Cambiaron Industrias	10
1.6.1. Caso 1: Moneyball - Datos vs. Tradición en el Béisbol	10
1.6.2. Caso 2: Netflix - El Motor Oculto del Entretenimiento	12
1.6.3. Caso 3: Detección de Fraude Bancario	13
1.6.4. Caso 4: Tinder - Algoritmos de Conexión Humana	13
1.6.5. Analiza un Caso de Éxito	14
1.6.6. Caso 5: Spotify Wrapped - Personalización a Gran Escala	15
1.7. El Flujo de Trabajo del Analista	16
1.7.1. El Ciclo de Vida de un Proyecto de Datos	16
1.7.2. Mapea un Problema al Flujo de Trabajo	17
1.8. Nuestro Laboratorio: Google Colab	18
1.8.1. ¿Qué es un Notebook?	18
1.8.2. Los Dos Idiomas del Notebook	19
1.8.3. Markdown: El Lenguaje de la Documentación	19
1.8.4. Tu primer notebook en Google Colab	20
1.8.5. El arte de la documentación (Markdown)	21

1.8.6. Gestión de ficheros: Google Drive	21
1.8.7. Configura tu entorno profesional	22
1.9. Roadmap del módulo	23
1.9.1. Tu viaje de aprendizaje	23
1.9.2. Competencias profesionales que desarrollarás	24
1.9.3. Perfiles profesionales de salida	25
1.10. Fuentes y lecturas recomendadas	25
1.11. Resumen de la unidad	26
2. NumPy: El motor invisible del Data Science	27
2.1. Introducción a NumPy: de listas de Python a computación científica	28
2.1.1. Instalación y Carga de la Librería	28
2.2. La necesidad de NumPy	28
2.2.1. El problema: ¿por qué no usar simplemente listas de Python?	28
2.2.2. El coste oculto de las listas	31
2.2.3. La analogía clave: cajón de sastre vs. pastillero	32
2.3. Anatomía de un ndarray	35
2.3.1. Conociendo nuestros datos: atributos fundamentales	35
2.4. Creación de Arrays: La Caja de Herramientas	37
2.4.1. Desde Estructuras Python Existentes	38
2.4.2. Arrays de Inicialización	38
2.4.3. Secuencias y Rangos	40
2.4.4. Datos aleatorios	41
2.5. Indexación y <i>slicing</i> : acceder a tus datos	44
2.5.1. Indexación básica (1D)	45
2.5.2. Indexación Multidimensional (2D)	45
2.5.3. Máscaras booleanas: el superpoder de la ciencia de datos	47
2.5.4. Fancy indexing: selección por índices	48
2.6. Operaciones matemáticas: vectorización en acción	49
2.6.1. Operaciones <i>element-wise</i> (elemento a elemento)	50
2.6.2. Funciones universales (ufuncs)	50
2.6.3. Agregaciones: del array a un escalar	51
2.7. Broadcasting: Magia de Shapes Compatibles	55
2.7.1. Reglas de Broadcasting	55
2.7.2. Aplicación: normalización de dataset por <i>z-score</i>	57
2.8. Manipulación de estructura: <code>reshape()</code> y <code>.T</code>	60
2.8.1. <code>reshape()</code> : cambiar la forma sin cambiar los datos	60
2.8.2. <code>.T</code> : transponer ejes	61
2.9. Errores comunes y <i>debugging</i>	62
2.10. Consolidación final: de NumPy a Pandas	68
2.10.1. Transición: ¿por qué necesitamos Pandas?	68
2.11. Fuentes y lecturas recomendadas	70
3. Pandas: El dialecto de los datos	71
3.1. Instalación y carga de librerías	71
3.2. ¿Por qué Pandas sobre Excel?	72
3.3. Las piezas fundamentales: Series y DataFrames	72
3.3.1. La Serie: la columna de datos con identidad	72

3.3.2. El DataFrame: nuestra tabla de datos en Python	75
3.4. Carga de datos desde archivos	79
3.4.1. Carga desde CSV	79
3.4.2. Otros Formatos Comunes	80
3.5. Selección de Datos	82
3.5.1. Selección de Columnas	83
3.5.2. Selección con <code>.loc</code> y <code>.iloc</code>	83
3.6. Filtrado booleano	86
3.6.1. Condiciones simples	86
3.6.2. Condiciones múltiples	86
3.6.3. Métodos de conveniencia para filtrado	88
3.6.4. Sintaxis moderna: filtrado con <code>.query()</code>	89
3.7. Ordenación de Datos	91
3.8. Agregaciones con <code>GroupBy</code>	92
3.8.1. Agregación Simple	93
3.8.2. Agregaciones Múltiples	94
3.9. Manejo de Valores Faltantes	96
3.9.1. Eliminación de nulos	97
3.9.2. Imputación de nulos	97
3.9.3. Detección y eliminación de duplicados	99
3.10. Transformación de datos	100
3.10.1. Creación de nuevas columnas: tres enfoques	100
3.10.2. Comparación: ¿cuál usar?	103
3.10.3. Aplicación de funciones personalizadas	104
3.10.4. Operaciones con <i>strings</i>	105
3.11. Combinación de <code>DataFrames</code>	108
3.11.1. Concatenación	108
3.11.2. Merge (Joins SQL)	109
3.12. Method Chaining: Flujo de Trabajo Profesional	112
3.12.1. Del código “paso a paso” al <code>method chaining</code>	112
3.12.2. El método <code>.pipe()</code> : transformaciones reutilizables	113
3.12.3. Guía de estilo: cuándo usar <i>method chaining</i>	114
3.13. Exportación de Datos	116
3.14. Fuentes y Lecturas Recomendadas	117

II. Parte II: Adquisición y Preparación 119

4. Adquisición de Datos 121	
4.1. El Paisaje de los Datos: Más Allá del CSV	121
4.1.1. El flujo profesional de datos	121
4.1.2. Dónde Viven los Datos	122
4.2. Ficheros Estructurados: CSV y Excel	123
4.2.1. Recordatorio: CSV que Ya Conoces	123
4.2.2. CSV: Problemas Reales	124
4.2.3. CSV con formato europeo	126
4.2.4. Excel: El Formato Empresarial	126
4.2.5. Guardar Resultados	127
4.2.6. Integrando CSV y Excel	128

4.2.7. Comparativa: CSV vs Excel	129
4.3. Ficheros Semi-estructurados: JSON	130
4.3.1. JSON Plano: El Caso Fácil	130
4.3.2. JSON Anidado: El Caso Real	132
4.3.3. La Solución: <code>pd.json_normalize()</code>	133
4.3.4. Aplanando una biblioteca JSON	135
4.3.5. Guardar JSON desde Pandas	136
4.4. Introducción a las APIs REST	137
4.4.1. La Metáfora del Restaurante	137
4.4.2. Antes de Empezar: ¿Qué es HTTP?	138
4.4.3. Anatomía de una Petición API	138
4.4.4. Métodos HTTP: El verbo de tu petición	139
4.4.5. Códigos de Estado HTTP (Status Codes)	139
4.4.6. Flujo de Trabajo Completo con APIs	140
4.4.7. Consumiendo tu primera API REST	142
4.4.8. APIs con Parámetros (Query Parameters)	142
4.5. Bases de Datos Relacionales: Conexión y SQL	143
4.5.1. El Modelo Relacional en un Vistazo	144
4.5.2. Conectar Python a una Base de Datos	144
4.5.3. Crear una BD de Ejemplo para Practicar	145
4.5.4. Creando tu primera base de datos SQLite	150
4.6. Consultas de Selección: El Puente SQL-Pandas	150
4.6.1. SELECT y WHERE	150
4.6.2. Consultas Básicas	151
4.7. Agregaciones: Sintetizando Información	152
4.7.1. Agregaciones	153
4.8. JOINS: Combinando Tablas	154
4.8.1. Practicando JOINS	155
4.9. Integrando fuentes: el enriquecimiento de datos	156
4.9.1. Caso práctico: enriqueciendo los barrios de Ames, Iowa	156
4.9.2. Paso 1: cargar CSV	156
4.9.3. Paso 2: enriquecer con API	157
4.9.4. Paso 3: consultar base de datos	158
4.9.5. Paso 4: combinar todo con <code>merge()</code>	158
4.9.6. Paso 5: validación y cierre	159
4.9.7. Enriquecimiento Completo	160
4.10. Errores Comunes y Buenas Prácticas	160
4.10.1. Buenas Prácticas	161
4.11. Resumen: Flujo de Adquisición de Datos	162
4.11.1. Checklist de Adquisición	163
4.12. SQL vs Pandas: ¿Cuándo Usar Cada Uno?	164
4.13. Fuentes y Lecturas Recomendadas	166
5. Limpieza y Calidad de Datos	167
5.1. Introducción: del <i>dataset</i> crudo al listo para modelar	167
5.1.1. Los datos del mundo real son desordenados	167
5.1.2. Diferencia: limpieza básica (UT3) vs. avanzada (UT5)	168
5.1.3. Configuración del entorno y carga del <i>dataset</i>	169
5.1.4. El <i>dataset</i> de referencia: Ames Housing	170

5.1.5. Diagnóstico de calidad en Ames Housing	171
5.2. Validación inicial: la primera línea de defensa	171
5.2.1. Paso 1: inspección de estructura	171
5.2.2. Paso 2: detección de valores faltantes	172
5.2.3. Paso 3: identificación de duplicados	174
5.2.4. Errores comunes en validación inicial	176
5.3. Manejo avanzado de valores faltantes	177
5.3.1. ¿Por qué faltan los datos?	177
5.3.2. Estrategias de imputación	178
5.3.3. Imputación por barrio (MAR)	181
5.4. Detección de <i>outliers</i> : el arte de distinguir errores de valores genuinos	184
5.4.1. ¿Qué es un <i>outlier</i> ?	184
5.4.2. Métodos para detectar <i>outliers</i>	184
5.4.3. Visualización de <i>outliers</i> con <i>boxplot</i> (diagrama de caja)	185
5.4.4. Detectando <i>outliers</i> en precios de vivienda	187
5.4.5. Estrategias para tratar <i>outliers</i>	187
5.5. Transformación de tipos de datos	191
5.5.1. Problema: fechas como <i>strings</i>	191
5.5.2. <i>Feature Engineering</i> temporal	191
5.5.3. Transformación de <i>strings</i> : normalización	193
5.5.4. Extracción con expresiones regulares (<i>regex</i>)	193
5.5.5. <i>Feature engineering</i> temporal con <code>YearBuilt</code>	194
5.5.6. Codificación de variables categóricas	195
5.5.7. <i>Target encoding</i> : codificar por el target	199
5.5.8. Codificación de variables categóricas: tres técnicas, una decisión	200
5.6. Normalización y Escalado	201
5.6.1. ¿Por qué escalar?	201
5.6.2. Comparando escaladores	205
5.6.3. ¿Qué método de escalado elegir?	206
5.6.4. Escalar después de <i>split</i> train/test	207
5.7. <i>Pipeline</i> completo: ensamblando todo	209
5.7.1. Usando el <i>pipeline</i>	210
5.7.2. <i>Pipeline</i> completo de Ames Housing	211
5.8. Resumen: mapa conceptual completo	212
5.8.1. Árbol de decisión: ¿qué técnica usar?	212
5.8.2. Comparativa: técnicas básicas vs. avanzadas (repasso)	213
5.9. Aplicación: preparar Ames Housing para modelado	214
5.10. Fuentes y lecturas recomendadas	215
5.11. Próximos pasos	215

III. Parte III: Análisis y Visualización

217

6. Análisis Exploratorio de Datos (EDA)

219

6.1. El rol del EDA en el ciclo de vida del proyecto	220
6.2. Las 4 preguntas clave del EDA	221
6.3. Caso de uso: predicción de precios de viviendas	221

6.4. Preparación del entorno	222
6.5. Carga del <i>dataset</i>	223
6.6. Análisis univariante: entendiendo variable por variable	223
6.6.1. Primera inspección del <i>dataset</i>	224
6.6.2. Variables numéricas: más allá de la media	224
6.6.3. Medidas de tendencia central revisitadas	224
6.6.4. Medidas de dispersión profundas	226
6.6.5. Visualizar la dispersión: boxplot y KDE	230
6.6.6. Forma de la distribución: asimetría y curtosis	232
6.6.7. Variables categóricas: frecuencias y proporciones	239
6.6.8. Tablas de frecuencia enriquecidas	240
6.6.9. Concentración de Categorías	241
6.6.10.Consolidación: análisis univariante	243
6.7. Análisis bivalente: relaciones entre variables	243
6.7.1. Numérica vs. numérica: correlaciones y patrones	244
6.7.2. Correlación de Pearson en Profundidad	244
6.7.3. Limitaciones de la correlación de Pearson	247
6.7.4. Correlación de Spearman: la alternativa robusta	248
6.7.5. Matrices de correlación	250
6.7.6. Categórica vs. numérica: análisis por grupos con <code>groupby</code>	254
6.7.7. Tablas de contingencia simples	256
6.7.8. Categórica vs. categórica: tablas de contingencia	259
6.7.9. Concepto intuitivo de independencia	261
6.7.10.Consolidación: análisis bivalente	264
6.8. Análisis multivariante: explorando interacciones	264
6.8.1. Estratificación: análisis condicional	264
6.8.2. Agregaciones múltiples y <i>pivoting</i>	266
6.8.3. <i>Pivot tables</i> avanzadas: múltiples agregaciones	268
6.8.4. Consolidación: análisis multivariante	271
6.9. Estadística inferencial: para ir más allá	271
6.10.Metodología de EDA: flujo de trabajo profesional	271
6.10.1.Registro de hallazgos clave	272
6.10.2.Preparación para modelado: <i>checklist</i> final	275
6.10.3.Errores comunes y cómo evitarlos	275
6.11.Del análisis a los datos: materializar los hallazgos del EDA	276
6.12.Fuentes y Lecturas Recomendadas	278
6.12.1.Reflexión final	278
7. Visualización de Datos	281
7.1. Los tres pilares de una buena visualización	281
7.2. Visualizaciones engañosas: ejemplos reales	282
7.3. El ecosistema de visualización en Python	283
7.4. Anatomía de un gráfico en Matplotlib	284
7.5. Preparación del entorno	287
7.6. <i>Dataset</i> principal: House Prices	288
7.7. Matplotlib: control total sobre gráficos	289
7.7.1. Fundamentos: la interfaz OOP	289
7.7.2. Personalización de ejes	290
7.7.3. Personalización de líneas y marcadores	294

7.7.4. Líneas de referencia y formato de ejes	295
7.7.5. Guardar gráficos	297
7.7.6. <i>Subplots</i> : múltiples gráficos en una figura	299
7.7.7. Tipos de gráficos esenciales en Matplotlib	301
7.7.8. Gráfico de barras: comparar categorías	303
7.7.9. <i>Boxplots</i> : visualización de cuartiles y <i>outliers</i>	305
7.8. Seaborn: gráficos estadísticos de alto nivel	308
7.8.1. Temas y paletas: estética profesional	309
7.8.2. Distribuciones: Visualizando Variables Univariantes	312
7.8.3. Relaciones: Visualizando Variables Bivariantes	318
7.8.4. Matrices y <i>heatmaps</i>	323
7.8.5. Gráficos categóricos	328
7.9. Personalización avanzada y mejores prácticas	334
7.10. Principios de <i>storytelling</i> visual	337
7.10.1. La estructura de una narrativa con datos	337
7.10.2. Checklist de un Buen Gráfico	338
7.11. Recapitulación: <i>toolbox</i> del analista visual	341
7.12. Fuentes y Lecturas Recomendadas	342
7.12.1. Reflexión final	342

IV. Parte IV: Machine Learning y Comunicación 345

8. Modelado y <i>Machine Learning</i>	347
8.1. Programación tradicional vs. <i>Machine Learning</i>	348
8.2. Tipos de aprendizaje automático	349
8.2.1. Aprendizaje supervisado	350
8.2.2. Aprendizaje no supervisado	351
8.2.3. Aprendizaje por refuerzo	351
8.3. Instalación y configuración	352
8.4. El flujo de trabajo de <i>Machine Learning</i>	354
8.4.1. ordenar el flujo ML	355
8.4.2. La importancia del <i>train/test split</i>	356
8.4.3. El concepto de «fit» y «predict»	359
8.5. Preparación de datos para <i>Machine Learning</i>	360
8.5.1. <i>Train/test split</i> avanzado	362
8.5.2. Preparación de datos para sklearn	364
8.5.3. Preparación de Datos Completa	368
8.6. Regresión Lineal: el modelo más interpretable	369
8.6.1. Intuición y Concepto	369
8.6.2. Implementación en Scikit-learn	371
8.6.3. Interpretación de coeficientes	371
8.6.4. Métricas de evaluación para regresión	373
8.6.5. Análisis de residuos	375
8.6.6. Regresión lineal aplicada	376
8.7. Clasificación: regresión logística	377
8.7.1. De regresión a clasificación	377
8.7.2. Implementación en Scikit-learn	378
8.7.3. Métricas de evaluación para clasificación	379

8.7.4. La curva ROC y AUC	382
8.7.5. ¿Qué variables importan más? Los coeficientes como medida de importancia	384
8.7.6. Clasificación y diagnóstico de errores	386
8.8. Árboles de Decisión: Modelos Interpretables No Lineales	387
8.8.1. Intuición y Concepto	387
8.8.2. Implementación en Scikit-learn	388
8.8.3. Visualización del árbol	389
8.8.4. <i>Feature importance</i>	391
8.8.5. Diagnóstico de <i>overfitting</i>	392
8.8.6. árboles de decisión con diagnóstico de <i>overfitting</i>	393
8.9. <i>Random Forest</i> : el poder de la sabiduría colectiva	394
8.9.1. Intuición: <i>ensemble learning</i>	394
8.9.2. Implementación en Scikit-learn	395
8.9.3. <i>Feature importance</i> robusto	396
8.9.4. Comparación de modelos	398
8.9.5. Comparación sistemática y selección de modelos	400
8.9.6. Síntesis: cuatro modelos para dos tareas	400
8.10.Caso práctico: predicción de precios en Ames, Iowa	402
8.11.Conceptos avanzados y próximos pasos	405
8.11.1. <i>Overfitting</i> (sobreajuste) vs. <i>underfitting</i> (subajuste): profundización	405
8.11.2.Regularización: Ridge y Lasso	407
8.11.3. <i>cross-validation</i> (validación cruzada)	411
8.11.4.Optimización de hiperparámetros	412
8.11.5. <i>Feature engineering</i> : conectar con el modelo	413
8.11.6.Interpretabilidad y explicabilidad	414
8.11.7.Ética y sesgos en <i>Machine Learning</i>	414
8.11.8. <i>Deployment</i> y monitoreo	415
8.11.9.Próximos pasos en tu aprendizaje	415
8.11.10Reflexión final	416
8.12.Fuentes y lecturas recomendadas	416
9. Dashboards y Comunicación de Resultados	417
9.1. Visión general: del <i>notebook</i> técnico al <i>dashboard</i> ejecutivo	417
9.2. El problema que resolvemos	418
9.3. Plotly Express – gráficos interactivos	418
9.3.1. Conceptos fundamentales	419
9.3.2. Primer gráfico interactivo	420
9.3.3. Gráficos esenciales	422
9.3.4. Personalización para comunicación profesional	424
9.3.5. Faceting – comparar múltiples categorías	425
9.3.6. Consolidación – Plotly Express	426
9.4. Streamlit – aplicaciones web en Python puro	426
9.4.1. Conceptos fundamentales de Streamlit	426
9.4.2. Elementos básicos	428
9.4.3. Componentes interactivos	429
9.4.4. Diseño profesional	431
9.4.5. Optimización – cacheo	432

9.4.6. Consolidación – Streamlit	434
9.5. Integración y práctica guiada	435
9.5.1. Cuadro de mando de Ames Housing – cierre del hilo narrativo	436
9.6. ANEXO: contenido avanzado (opcional)	440
9.7. Camino al proyecto final	442
V. Parte V: Proyecto Integrador	445
10. Proyecto Final Integrador: De Datos a Decisiones	447
10.1. El síndrome de la hoja en blanco	447
10.2. El caso de negocio: retención de clientes (<i>Churn</i>)	448
10.3. Estructura del proyecto (Hitos)	449
10.4. Checklist para Portfolio y Entrevista Técnica	454
11. Y ahora, ¿qué?	457
11.1. Lo que has construido	457
11.2. Tu perfil profesional	457
11.3. Qué aprender a continuación	458
11.4. Recursos para seguir	459
11.5. Una última nota	459
12. Sobre el Autor	461
Apéndices	463
A. Instalación y configuración del entorno	463
A.1. Instalación de Python vía Miniconda	463
A.1.1. Pasos para Windows:	463
A.1.2. Pasos para macOS:	464
A.1.3. Pasos para Linux:	464
A.2. Creación del entorno del curso	464
A.2.1. Paso a paso en la terminal:	464
A.3. Instalación de librerías (requirements.txt)	465
A.4. Configuración de Jupyter Notebook/Lab	465
A.4.1. Instalación y ejecución:	465
A.5. Verificación del entorno	466
A.6. Resolución de problemas comunes (<i>troubleshooting</i>)	466
A.7. Alternativa: Google Colab (Sin instalación)	467
B. Recursos y lecturas recomendadas	469
B.1. Documentación oficial (<i>core stack</i>)	469
B.2. Recursos por Unidad Temática	469
B.2.1. UT02: NumPy y Computación Vectorizada	469
B.2.2. UT03: Pandas y Manipulación de Datos	470
B.2.3. UT04: Adquisición de Datos	470
B.2.4. UT05: Limpieza y Calidad de Datos	470
B.2.5. UT06: Análisis Exploratorio de Datos (EDA)	471

B.2.6. UT07: Visualización de Datos	471
B.2.7. UT08: Machine Learning	472
B.2.8. UT09: Dashboards y Comunicación	472
B.3. Bibliografía general	473
B.3.1. Teoría y metodología	473
B.3.2. Técnica y código	473
B.4. Comunidades y aprendizaje continuo	473
B.5. Plataformas de práctica	473
C. Guía rápida de referencia (<i>cheat sheet</i>)	475
C.1. Computación numérica con NumPy (UT2)	475
C.1.1. Creación de <i>arrays</i>	475
C.1.2. Inspección y atributos	475
C.1.3. Manipulación de estructura	476
C.1.4. Agregaciones y estadísticas	476
C.1.5. Indexación y filtrado	476
C.2. Manipulación de datos con Pandas (UT3)	477
C.2.1. Carga y exploración	477
C.2.2. Selección y filtrado	477
C.2.3. Agregación y limpieza	478
C.3. Adquisición e integración de datos (UT4)	478
C.3.1. Carga de ficheros	478
C.3.2. APIs y servicios web	478
C.3.3. SQL básico (DML)	479
C.3.4. Integración (<i>merge/join</i>)	479
C.4. Limpieza y calidad de datos (UT5)	479
C.4.1. Valores faltantes (nulos)	479
C.4.2. Valores atípicos (<i>outliers</i>)	480
C.4.3. Transformación y codificación	480
C.5. Análisis exploratorio de datos (UT6)	480
C.5.1. Estadísticas de forma y dispersión	480
C.5.2. Relaciones entre variables	481
C.5.3. Flujo de trabajo EDA (5 fases)	481
C.5.4. Estructura de <i>notebook</i> de EDA	482
C.5.5. Función de análisis univariante – variables numéricas	483
C.5.6. Función de análisis univariante – variables categóricas	483
C.6. Visualización con Matplotlib y Seaborn (UT7)	484
C.6.1. Matplotlib: control de figuras y ejes	484
C.6.2. Seaborn: gráficos estadísticos rápidos	484
C.6.3. Gráficos multivariantes avanzados	485
C.7. <i>Machine Learning</i> con scikit-learn (UT8)	485
C.7.1. Flujo de trabajo y validación	485
C.7.2. Métricas de evaluación	486
C.7.3. Modelos y parámetros comunes	486
C.8. <i>Dashboards</i> y comunicación (UT9)	487
C.8.1. Plotly Express (interactividad)	487
C.8.2. Streamlit (app web)	487
D. Programación Funcional Aplicada al Análisis de Datos	489
D.1. Qué es el estilo funcional (aplicado)	489

D.2. Funciones puras e inmutabilidad	490
D.3. Map, Filter, Reduce en Python — y sus equivalentes en Pandas . . .	490
D.4. <i>Method Chaining</i> : el patrón central de Pandas	491
D.4.1. El método <code>.pipe()</code> : integrar funciones propias en la cadena	492
D.5. Funciones <i>lambda</i> y funciones anónimas	493
D.6. Cuándo NO usar el estilo funcional	494
D.7. Tabla resumen: imperativo vs. funcional	495
E. Glosario Bilingüe	497
E.1. Definiciones de términos	497
F. Git Esencial para Analistas de Datos	505
F.1. Comandos Básicos	505
F.2. Ramas (Branches)	505
F.3. Comandos Útiles	506
F.4. Flujo de Trabajo Típico	506
F.5. Resolución de Conflictos Básica	507
G. Bibliografía	509
Bibliografía	511
Índice Analítico	513