

Explotación de datos

Amazon Top 50 Bestselling Books

Visualización con Power BI: procesos

Extracción

El Dataset consiste en un solo archivo en formato CSV delimitado por coma y se compone de las columnas:

- Name (String)
- Author (String)
- User Rating (Int)
- Reviews (Int)
- Price (Int)
- Year (Int)
- Genre (String) - variable categórica

Name	Author	User Rating	Reviews	Price	Year	Genre
10-day Green Smoothie Cleanse	JJ Smith	47	17350	8	2016	Non Fiction
13/12/61: A Novel	Stephen King	46	2052	22	2012	Fiction
12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	47	18979	15	2018	Non Fiction
1984 (Signet Classics)	George Orwell	47	21424	6	2017	Fiction
5,000 Awesome Facts (About Everything) (National Geographic Kids)	National Geographic Kids	48	7665	12	2019	Non Fiction
A Dance with Dragons (A Song of Ice and Fire)	George R. R. Martin	44	12043	11	2011	Fiction
A Game of Thrones / A Clash of Kings / A Storm of Swords...	George R. R. Martin	47	29325	30	2014	Fiction
A Gentleman in Moscow: A Novel	Amor Towles	47	19699	15	2017	Fiction
A Higher Loyalty: Truth, Lies, and Leadership	James Comey	47	5983	3	2018	Non Fiction
A Man Called Ove: A Novel	Fredrik Backman	46	23848	8	2016	Fiction
A Man Called Ove: A Novel	Fredrik Backman	46	23848	8	2017	Fiction
A Patriot's History of the United States: From Columbus...	Larry Schweikart	46	490	2	2010	Non Fiction
A Station Life: A Memoir	Jayce Dugard	46	4349	12	2012	Non Fiction
A Wrinkle in Time (Time Quintet)	Madeline L'Engle	45	5153	5	2018	Fiction
Act Like a Lady, Think Like a Man: What Men Really Thi...	Steve Harvey	46	5013	17	2009	Non Fiction
Adult Coloring Book Designs: Stress Relief Coloring Boo...	Adult Coloring Book Designs	45	2313	4	2016	Non Fiction
Adult Coloring Book: Stress-Relieving Animal Designs	Blue Star Coloring	46	2925	6	2015	Non Fiction
Adult Coloring Book: Stress-Relieving Patterns	Blue Star Coloring	44	2951	6	2015	Non Fiction
Adult Coloring Books: A Coloring Book for Adults Featur...	Coloring Books for Adults	45	2426	8	2015	Non Fiction
Alexander Hamilton	Ron Chernow	48	9198	13	2016	Non Fiction

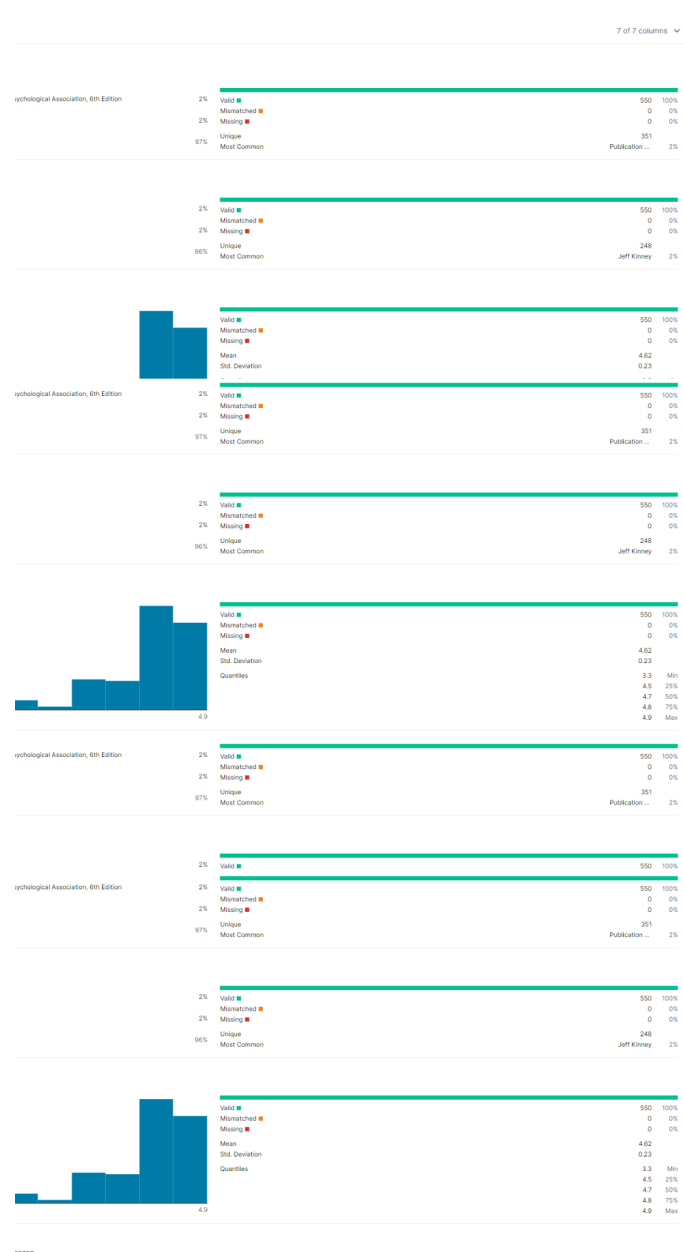
Esta tabla no dispone de otras relaciones y por tanto no es necesario establecer Primary Keys (PK):

Column	Icon
Author	
Genre	
Name	
Price	Σ
Profit	Σ
Reviews	Σ
User Rating	Σ
Year	Σ
Profit x Reviews	Σ
Sales	Σ

Transformación

Limpieza:

Hemos evaluado este dataset en la pasada lección con Python, verificando que existe integridad de datos y no necesitan de completado con medias o modificación de tipo por errores, ya que son válidos en un 100%.



Transformación básica:

Observamos que las columnas 'User Rating' (Int) y 'Price' (Int) necesitan de un casting adecuado para su futuro uso, por lo que es necesario pasar 'User Rating' a float dentro del rango del 0 al 5, 'Price' se cambia a float.

1.2 User Rating	1.3 Reviews	1.2 Price
4,7	17350	8

Además, aplicamos un cambio de formato a 'Price' por su tipo a moneda y representado en dólares US:

Herramientas de tablas		Herramientas de columnas	
Nombre	Price	Formato	Moneda
Tipo de datos	Número decimal		
Estructura		Formato	Propiedades
Name	Author	User Rating	Reviews
5,000 Awesome Facts (About Everything!) (National Geogr	National Geographic Kids	4,8	7665
Alexander Hamilton	Ron Chernow	4,8	9198

Transformación extra:

El departamento de Finanzas requiere de información de Ingresos y Beneficios en su dashboard, sin embargo, esta información no está incluida en el dataset.

Tras evaluar el proceso de opiniones de Amazon, podemos concluir que estas corresponden en un alto porcentaje a una venta real y verificada, por tanto, la simple multiplicación de 'Reviews' por 'Price' nos da como resultado los ingresos mínimos verificados por libro. Decidimos recoger este dato en un campo de medida nuevo denominado 'Sales':

Campos	
Buscar	
bestsellers with categories	
Author	
Genre	
Name	
Price	
Profit	
Profit x Reviews	
Reviews	
Sales	
User Rating	
Year	

Para el requerimiento de Beneficios, el dept. Financiero nos confirma que cada venta da como rendimiento un 0.35 por uno del PVP. Añadimos este cálculo en la columna 'Profit'. Pero el dato a

mostrar en el dashboard será el beneficio total por producto, por lo que guardamos en este dato en un nuevo campo de medida 'Profit x Reviews'

Carga

Finalmente, el resultado de carga es el siguiente:

	Name	Author	User Rating	Reviews	Price	Year	Genre	Profit
	5,000 Awesome Facts (About Everything!) (National Geographic Kids)	National Geographic Kids	4,8	7665	\$12,00	2019	Non Fiction	\$4,20
	Alexander Hamilton	Ron Chernow	4,8	9198	\$13,00	2016	Non Fiction	\$4,55
	Barefoot Contessa Foolproof: Recipes You Can Trust: A Cookbook	Ina Garten	4,8	1296	\$24,00	2012	Non Fiction	\$8,40
	Becoming	Michelle Obama	4,8	61133	\$11,00	2018	Non Fiction	\$3,85
	Becoming	Michelle Obama	4,8	61133	\$11,00	2019	Non Fiction	\$3,85
	Being Mortal: Medicine and What Matters in the End	Atul Gawande	4,8	11113	\$15,00	2015	Non Fiction	\$5,25
	Cabin Fever (Diary of a Wimpy Kid, Book 6)	Jeff Kinney	4,8	4505	\$0,00	2011	Fiction	\$0,00
	Can't Hurt Me: Master Your Mind and Defy the Odds	David Goggins	4,8	16244	\$18,00	2019	Non Fiction	\$6,30
	Creative Haven Creative Cats Coloring Book (Adult Coloring)	Marjorie Sarnat	4,8	4022	\$4,00	2015	Non Fiction	\$1,40
	Creative Haven Owls Coloring Book (Adult Coloring)	Marjorie Sarnat	4,8	3871	\$5,00	2015	Non Fiction	\$1,75

Visualización

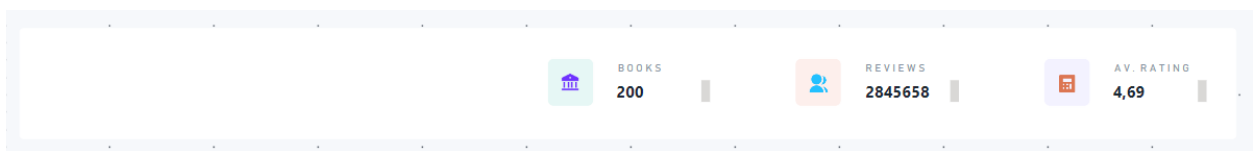
Se diseñan tres dashboards orientados a diferentes perfiles de usuario en base a sus KPIs y por tanto, priorizando la información que necesitan en sus informes. Se enfatiza en seguir los principios de Gestalt, para una representación sencilla, agradable y limpia. Estos dashboards se describen en detalle en el documento: Visualización con Power BI: audiencia y contexto.

Estos dashboards se componen de:

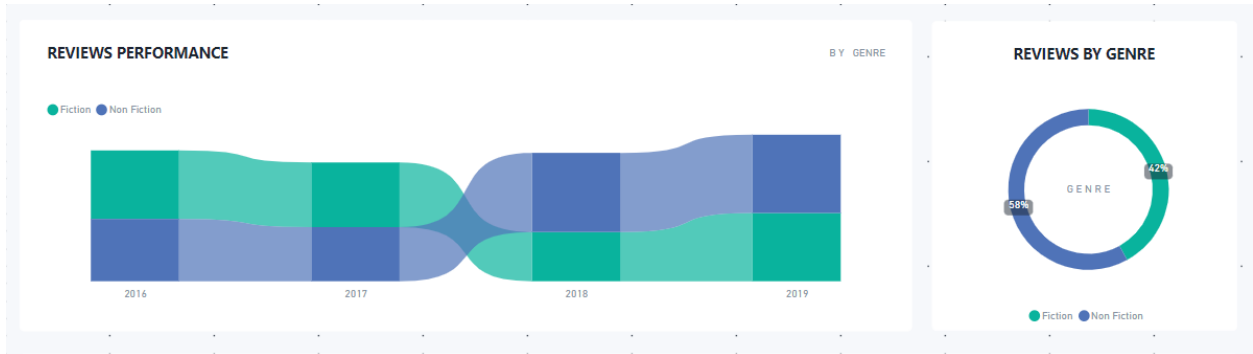
Barra lateral de navegación para un acceso rápido entre páginas:



Encabezado superior con los KPIs más importante del área:



Métricas evolutivas de performance:



Filtros en detalle: por año, por producto y ordenación en tabla:

The table displays a list of books with their respective review counts, genres, authors, prices, and user ratings. The table is filtered for the year 2016.

Book	Reviews	Genre	Author	Price	User Rating
Becoming	122266	Non Fiction	Michelle Obama	\$22.00	9.60
The 5 Love Languages: The Secret to Love That Lasts	102216	Non Fiction	Gary Chapman	\$32.00	19.20
The Four Agreements: A Practical Guide to Personal Freedom (A Toltec Wisdom Book)	93232	Non Fiction	Don Miguel Ruiz	\$24.00	18.80
Where the Crawdads Sing	87841	Fiction	Delia Owens	\$15.00	4.80