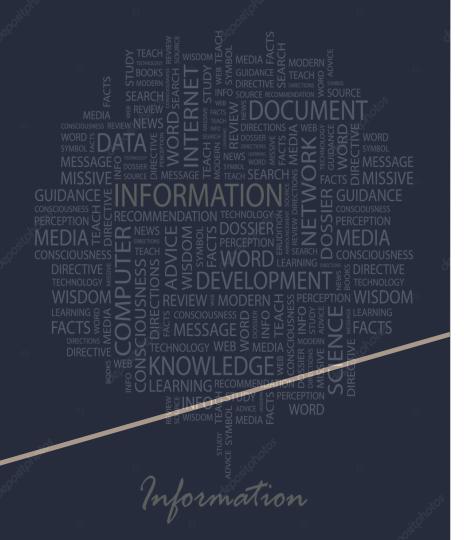
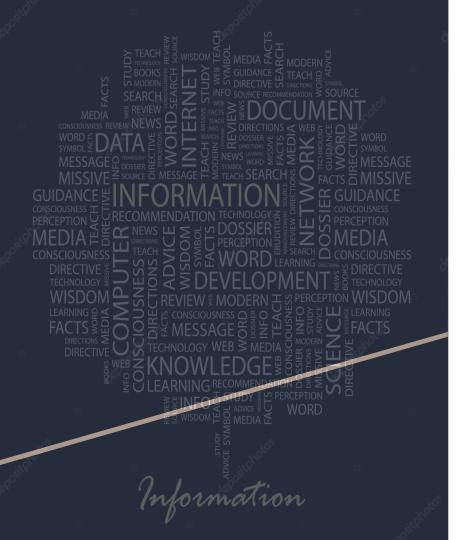
### Best Sellers in Books

Chen, Hung-Ting Hovhannisyan, Hayk Wang, Luxin Wang, Yudi

# TEAM 9



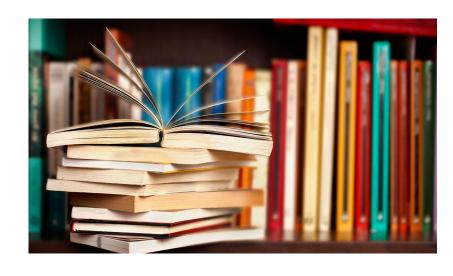
- 1. Motivation
- 2. Data Scraping
- 3. Analysis
- 4. Conclusion

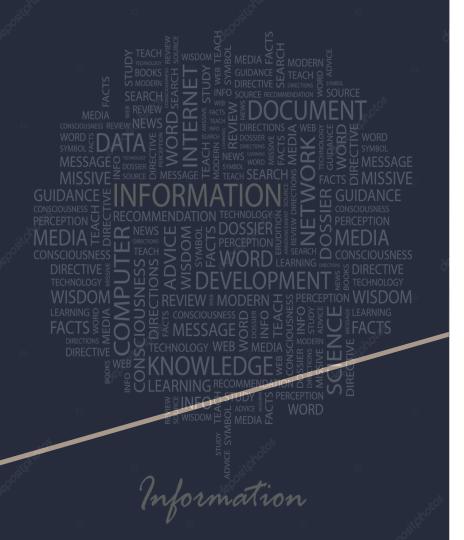


- 1. Motivation
- 2. Data Scraping
- 3. Analysis
- 4. Conclusion

### Motivation

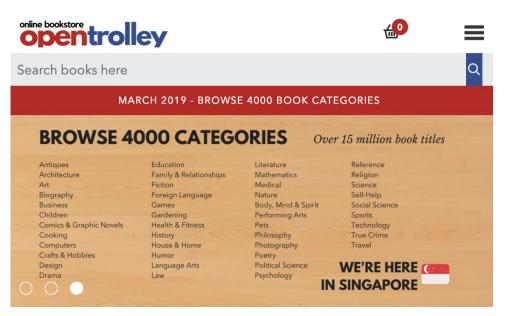
- What are the trends in best selling books.
- Get insight on what drive book price
- How are discounts given by the seller





- 1. Motivation
- 2. Data Scraping
- 3. Analysis
- 4. Conclusion

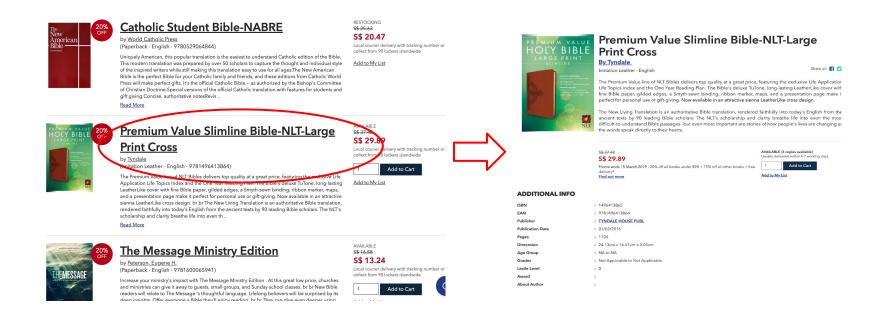
### Data Scraping



- Singapore Books Selling Website
- Each best seller category has been updated this February.
- Among all the categories we picked 14 which had the most books.

https://opentrolley.com.sg/Home.aspx

## Data Scraping



Method: Request package + XPath to scraping data from the HTML document

### Data Cleaning

#### Raw data:

name	author		price	original price	pages	publisher		publish type		publish date	available numbers	detail_url
['Catholic Student Bible-NABRE']	['\r\n	World Catholic Press']	['17.80']	['22.29']	['1408']	['WORLD BIBLE PUB CO	']	['(Paperback - English	- 9780529064844)']	['27/05/2011']	['0 copy available']	https://opentrolley
['Premium Value Slimline Bible-NLT-Large Print Cross']	['\r\n	Tyndale']	['29.89']	['37.42']	['1104']	['TYNDALE HOUSE PUBL	']	['(Imitation Leather - English	- 9781496413864)']	['01/03/2016']	['3 copies available']	https://opentrolley
['The Message Ministry Edition']	['\r\n	Peterson, Eugene H.']	['13.24']	['16.58']	['906']	['NAV PR ']		['(Paperback - English	- 9781600065941)']	['04/10/2016']	['910 copies available']	https://opentrolley
['Text Bible-NLT-Economy']	['\r\n	Tyndale']	['11.32']	['14.18']	['810']	['TYNDALE HOUSE PUBL	']	['(Paperback - English	- 9781414309477)']	['01/03/2006']	['202 copies available']	https://opentrolley
['KJV Mini Pocket Edition: Zippered Black']	['\r\n	']	['23.17']	['29.01']	['0']	['CHRISTIAN ART GIFTS	']	['(Imitation Leather - English	- 9781432102425)']	['01/10/2012']	0	https://opentrolley
['Premium Value Slimline Bible-NLT-Large Print Crown']	['\r\n	Tyndale']	['29.89']	['37.42']	['1104']	['TYNDALE HOUSE PUBL	']	['(Imitation Leather - English	- 9781496413871)']	['01/03/2016']	['227 copies available']	https://opentrolley
['Journaling Bible-ESV-Flowers']	['\r\n	Crossway Bibles']	['61.13']	['76.51']	['1066']	['CROSSWAY BOOKS	']	['(Hardcover - English	- 9781433548376)']	['30/09/2015']	0	https://opentrolley
['Haitian New Testament with Psalms-FL']	['\r\n	Haitian Bible Society']	['7.67']	['9.61']	['363']	['AMER BIBLE SOC ']	]	['(Paperback - Haida	- 9781585169443)']	['01/05/2010']	['103 copies available']	https://opentrolley

#### Adjustment:

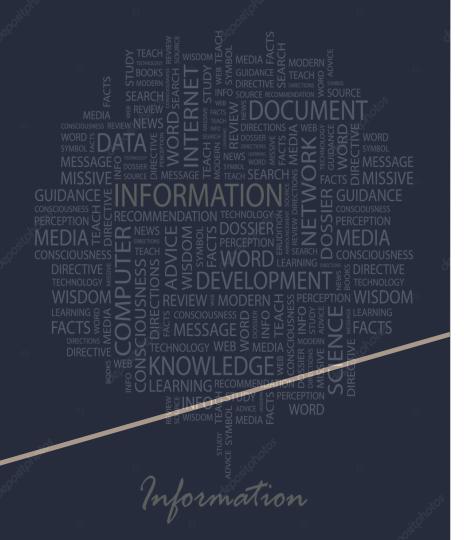
- Drop the unnecessary information
- Deal with the missing data
- Transform the price string into numbers
- Use publish year instead of the exact day
- Add the category tag

### Data Cleaning

#### Data after preprocessing:

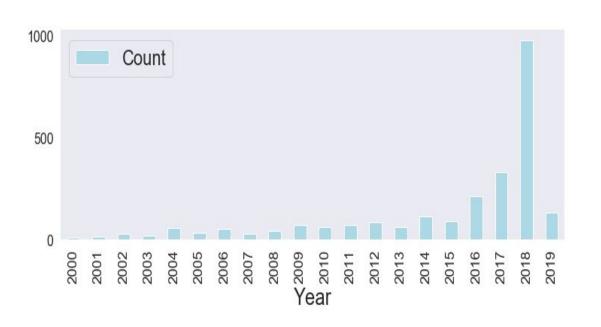
- 14 Categories
- 190 in each Category
- 2660 items in total
- Each item have 11 features

	name	author	price	original price	pages	publisher	publish type	available numbers	Category	publish year	publish month
0	['The Whole Smiths Good Food Cookbook: Whole30	\r\nSmith,Michelle	48.36	56.97	288	['HOUGHTON MIFFLIN ']	Hardcover	21	Cooking	2018	06
1	["Delish: Eat Like Every Day's the Weekend"]	\r\nEditorsofDelish	50.88	59.93	416	['HOUGHTON MIFFLIN ']	Hardcover	42	Cooking	2018	10
2	["Carla Hall's Soul Food: Everyday and Celebra	\r\nHall,Carla	48.35	56.95	336	['HARPER WAVE ']	Hardcover	90	Cooking	2018	10
3	['Great British Bake Off: Big Book of Baking']	\r\nCollister,Linda	70.85	83.46	320	['BBC BOOKS ']	Hardcover	61	Cooking	2014	09
4	['Tequila Mockingbird: Cocktails with a Litera	\r\nFederle,Tim	20.90	25.52	160	['RUNNING PR BOOK PUBL ']	Hardcover	116	Cooking	2013	04
5	['Ottolenghi Simple: A Cookbook']	\r\nOttolenghi,Yotam	59.63	70.24	320	['TEN SPEED PR ']	Hardcover	82	Cooking	2018	10
6	['The Kitchen Shortcut Bible: More Than 200 Re	\r\nWeinstein,Bruce	47.90	56.43	352	['LITTLE BROWN & CO INC ']	Hardcover	20	Cooking	2018	06
7	['Taste of Home Instant Pot Cookbook: Savor 17	\r\nTasteofHome	24.52	29.94	256	['READERS DIGEST CHILDRENS ']	Paperback	22	Cooking	2018	09
8	['The Pioneer Woman Cooks: Dinnertime: Comfort	\r\nDrummond,Ree	52.24	61.54	400	['WILLIAM MORROW & CO ']	Hardcover	56	Cooking	2015	10
9	['Cook Like a Pro: Recipes and Tips for Home C	\r\nGarten,Ina	54.59	64.30	272	['POTTER CLARKSON N ']	Hardcover	435	Cooking	2018	10
10	['The Instant Pot Electric Pressure Cooker Coo	\r\nRandolph,Laurel	22.43	27.39	176	['ROCKRIDGE PR ']	Paperback	100	Cooking	2016	04
11	['Cravings: Hungry for More']	\r\nTeigen,Chrissy	48.35	56.95	256	['POTTER CLARKSON N ']	Hardcover	281	Cooking	2018	09
12	['Cooking for Jeffrey: A Barefoot Contessa Coo	\r\nGarten,Ina	53.90	63.49	256	['POTTER CLARKSON N ']	Hardcover	30	Cooking	2016	10
13	['The Cookie Book: Decadent Bites for Every Oc	\r\nFirth,Rebecca	31.85	38.90	192	['PAGE STREET PUB ']	Paperback	24	Cooking	2018	09
14	['My Bread: The Revolutionary No-Work, No- Knea	\r\nLahey,Jim	54.02	63.63	222	['W W NORTON & CO INC ']	Hardcover	0	Cooking	2009	10



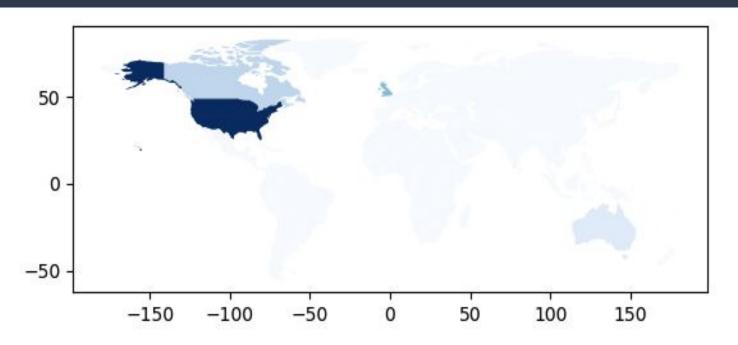
- 1. Motivation
- 2. Data Scraping
- 3. Analysis
- 4. Conclusion

### Analysis – Published Year vs Popularity



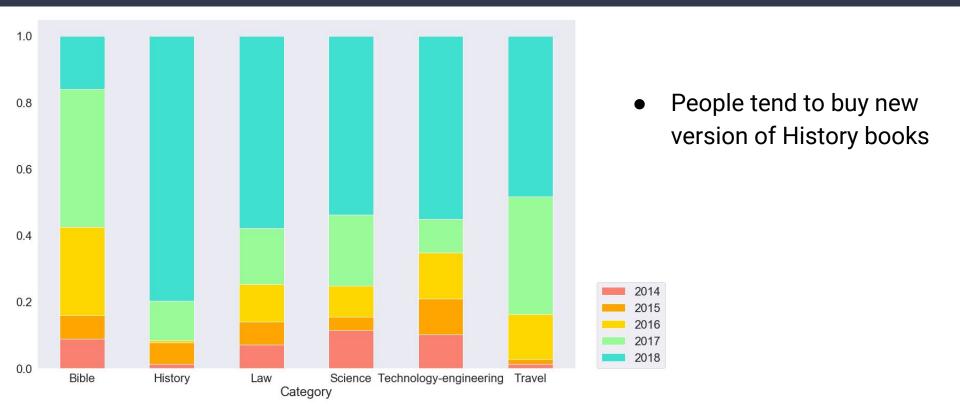
People like new releasing books!

### Analysis – Publishers vs Countries

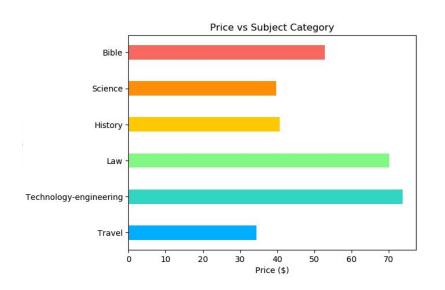


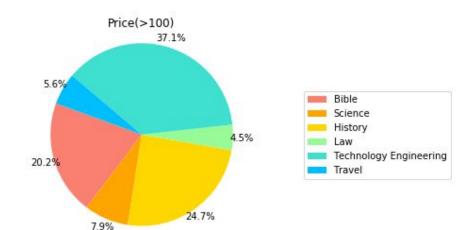
Dominated by English speaking Countries

### Analysis - Category vs Year

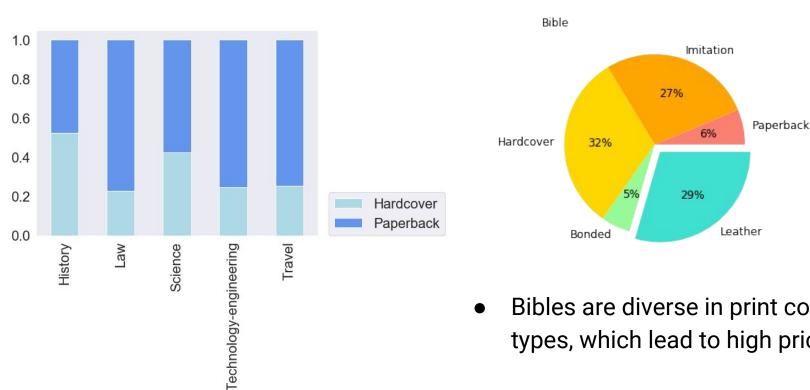


## Analysis - Price vs Category



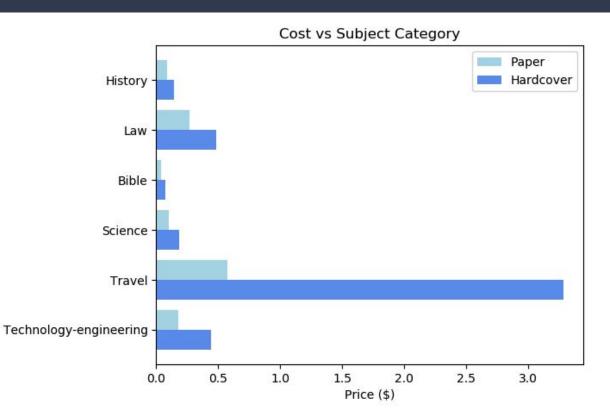


## Analysis - Category vs Print Cover Type



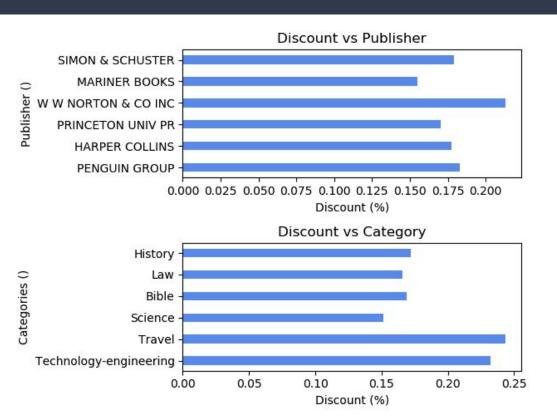
Bibles are diverse in print cover types, which lead to high prices.

## Analysis - Cost vs Subject Category



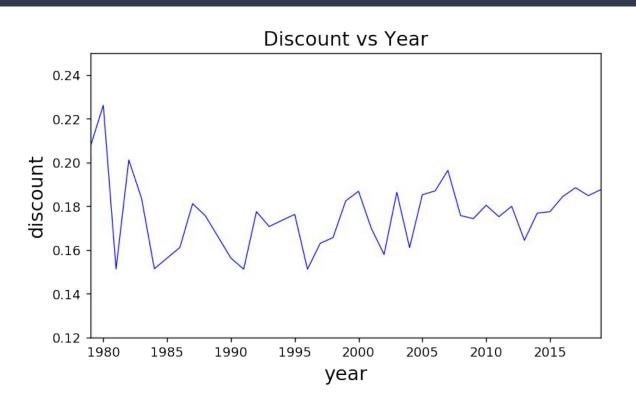
- Travel books cost more to make
- Selling Price is not associated with material cost

### Analysis – Discount vs Publisher



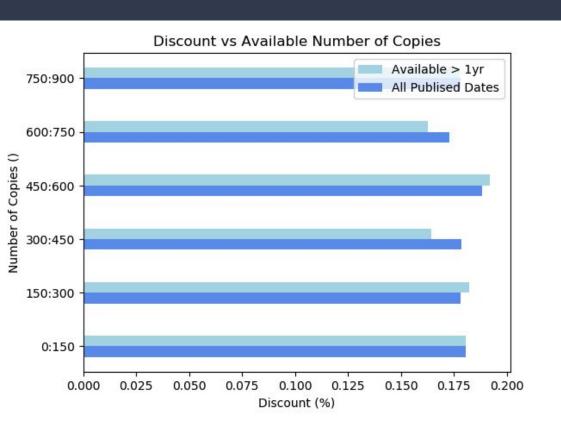
- Discount offer vs publisher seem to be uncorrelated
- Discount vs Categories also seems uncorrelated

### Analysis - Discount vs Published Year

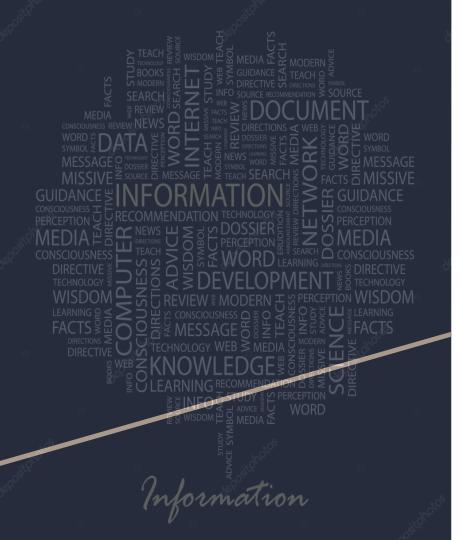


- Discount offer vs
  Published Year
- The discount seems uncorrelated with year before 2008 but becomes stable after 2008

### Analysis - Discount vs Available # of copies



- Discount vs Available number of copies also seems uncorrelated
- Correlation Coeff < 0.02</li>



- 1. Motivation
- 2. Data Scraping
- 3. Analysis
- 4. Conclusion

### Conclusion



- Cost of material is not the significant driver of price
- Discounts seem to be a marketing strategy