

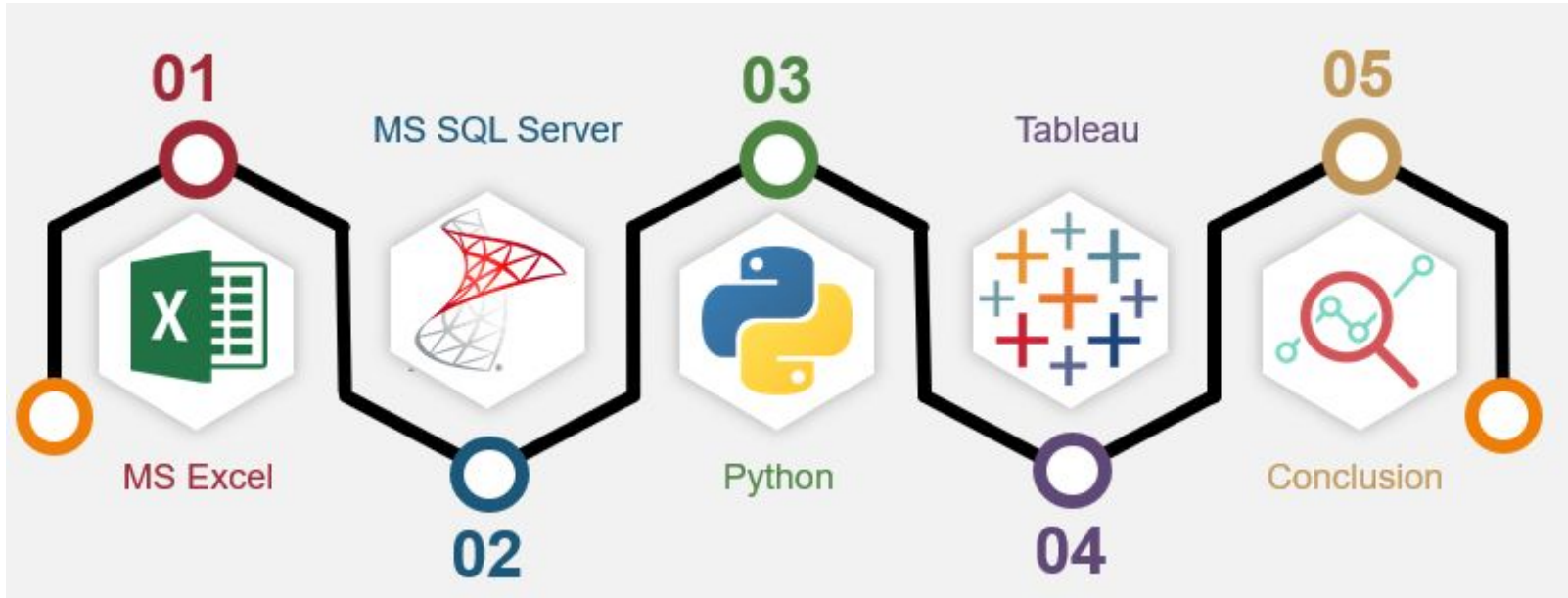


Scholastic Analytics Challenge

By: Kevin Gonzalez Garcia, Jerry Contreras, Tom Jaworski and Chudi Egbebike

[Full Tableau Story with All Visuals](#)

Technologies Utilized



What trends do you see in the data, among / demographics, genre/ theme, and price?



- There were originally 25 genres, that we grouped into 4 genres categories: Instructional, Fiction, Media & Miscellaneous, Non-Fiction
- It was clear that the genres in the instructional category, such as Reference & How-To had a much higher average unit price
- Additionally, there was not a lot of variation in price for the other genre categories

[Average Unit Price by Genre Category\(1\) Visual](#)

What trends do you see in the data, among / demographics, genre/ theme, and price?



- Originally, there were 326 themes that we grouped into 22 theme categories
- 2 of the top 4 theme categories by average unit price were Geography/History & Education/Industry Learning
- This continues to highlight that instructional and educational books have a higher average unit price

[Average Unit Price by Theme Category\(2\) Visual](#)

What trends do you see in the data, among / demographics, genre/ theme, and price?

- For Average Unit Price by State, we created three clusters
- The following trend was clear
 - Southern States have a higher Average Unit Price
 - Middle States fell in between Southern and Northern states for Average Unit Price
 - Northern States have the lowest Average Unit Price

	Centers
	Number of Items
	Sum of AverageUnitPrice
Cluster Low	13
Cluster High	23
Cluster Medium	16
	6.9799
	7.6289
	7.2848

[Average Unit Price by State Visual](#)


What does the data suggest about Scholastic's distribution channels, and how would you recommend structuring a distribution strategy?

The following are trends for the relationship between channels, the grade levels of the book, and revenue:

- Scholastic's main source of revenue comes from channel 2 which accounts for 64.9% of the revenue
- In regards to book grade level, Channel 2 is the main source of revenue for elementary grade level books (67.09%)
- Also, Channel 1 is the main source of revenue for middle/high school grade level books (60.59%)

[Revenue by Channel\(4\) & Revenue by Book Grade Level and Channel Grouped\(4.1\)](#)

What does the data suggest about Scholastic's distribution channels, and how would you recommend structuring a distribution strategy?



- The data suggests that channel 2 is a book fair, while channel 1 is an online platform
- “The undisputed leader in the field, Scholastic Book Fairs sells more than 100 million books to 35 million children each year. The division delivers more than 160,000 fairs in pre-K, elementary, and middle schools across the globe.”
- Book fairs are not as accessible for high school students, who are most likely to read high school grade level books

What does the data suggest about Scholastic's distribution channels, and how would you recommend structuring a distribution strategy?

- This reflects our data, which shows a decrease in revenue in channel 2 (book fair) around the middle school level and a further decrease in revenue for high school grade level books
- Channel 1 seems to be an online platform as it doesn't suffer the same constraints as channel 2 and is present through all book reading levels
- Following an selective distribution model, our channel specific distribution strategy involves channel 2 (book fairs) focusing on elementary to middle school book levels while channel 1(online) will be used as a general platform for all book levels

[Revenue by Individual Book Grade Level & Channel \(4.2\)](#)

Formulate a unified strategy for marketing between the two channels.



- Other than the Kindergarten book grade level, which is not as an effective book grade level to market to due to group reading within the classrooms in kindergarten, the grade level of books contributing to the lowest source of revenue are 4th-12th grade
- Our marketing strategy is to create a youtube campaign to promote and sell to the ages related to 4-12th grade level books: 9-17 as well as parents who have children within that age range

[Revenue by Individual Book Grade Level & Channel \(4.2\)](#)

Formulate a unified strategy for marketing between the two channels.



- Our unified marketing strategy is a youtube video advertisement, which promotes the scholastic brand across channel 1(internet platform) and channel 2(book fair)
- Our youtube video advertisement highlights the positivity of reading for book grade levels who appeal to an audience that may be distracted by technology

Formulate a unified strategy for marketing between the two channels.

US Youtube Scholastic Video Advertisement Campaign for One Month		
Total Market		50,000,000
Click Through Rate		3%
Viewers of Advertisement		1,500,000
Conversion Rate		4%
New Customers		60,000
Average Unit Price of Book	\$	7.42
Average Qty of Books Sold		2
Revenue	\$	890,400
Cost of Creating Ads	\$	75,000
Cost of Running Ads	\$	300,000
Profit	\$	515,400

- Total Market based on data of kids ages 9-17 that watch youtube and adults who have children ages 9-17 that watch youtube
- Click Through Rate & Conversion Rate based on Industry Averages
- Cost of running ads is a function of the viewers of the advertisement

Formulate a unified strategy for marketing between the two channels.



- In addition to youtube video advertisement, we will also be reaching out to influencers who have the highest overlap rate with the scholastic Youtube channel audience. We'll pay influencers on a per person basis dependent on the size of their audience
- We chose Youtube because of the percentage of young people who say they watch online videos "everyday" has more than doubled amongst both age groups, going from 24% to 56% among 8- to 12-year- olds, and from 34% to 69% among 13- to 18-year-olds

Formulate a unified strategy for marketing between the two channels.

Top 5 overlap audience channels



PopularMMOs

Area: United States Subscribers: 16.9M

10%

Overlap rate



FGTeeV

Area: United States Subscribers: 13M

9%

Overlap rate



MrBeast

Area: United States Subscribers: 29.6M

7.5%

Overlap rate



DanTDM

Area: United Kingdom Subscribers: 22.4M

6.5%

Overlap rate



Guava Juice

Area: United States Subscribers: 14.9M

6.5%

Overlap rate

Formulate a unified strategy for marketing between the two channels.

Martia



- Casual Youtube Video Watcher
- In her 40's
- Mother of 12 year-old Alex
- Fan of Scholastic
- Wants her son, Alex, to get off his iPhone and read a book
- Buys three books in the 6th grade reading level for Alex at the local book fair after watching the youtube video advertisement

Brian



- Watches Hours of Youtube A Day
- 17 years old
- Always on his phone & on social media
- Wants to get back into reading
- Fan of Scholastic
- Buys one book in the 12th grade reading level on the scholastic website after watching the youtube video advertisement

Where are there areas of significant overlap between the channels, and what strategy do you suggest to prevent unintentional competition between channels?



- There's significant overlap across the US through all channel sales
- A logistic regression model attempts to predict a binary response that would best suit Scholastic's needs by predicting which books could be better sold among the two masked distribution channels

[Channel Overlap by Zipcode](#)

Logit Regression Results

Dep. Variable:	CHANNEL	No. Observations:	200823			
Model:	Logit	Df Residuals:	200816			
Method:	MLE	Df Model:	6			
Date:	Sun, 09 Feb 2020	Pseudo R-squ.:	0.2064			
Time:	20:36:26	Log-Likelihood:	-84418.			
converged:	True	LL-Null:	-1.0637e+05			
Covariance Type:	nonrobust	LLR p-value:	0.000			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	1.8199	0.029	62.830	0.000	1.763	1.877
PROD_TYP	-0.5574	0.027	-20.791	0.000	-0.610	-0.505
SERIES	1.5813	0.015	103.472	0.000	1.551	1.611
NumberRange	-0.0003	2.95e-05	-8.599	0.000	-0.000	-0.000
UNIT_PRICE	-0.2601	0.003	-77.709	0.000	-0.267	-0.254
SCHOOL_TYPE	0.5496	0.017	32.879	0.000	0.517	0.582
StateID	0.0119	0.001	20.456	0.000	0.011	0.013
=====						

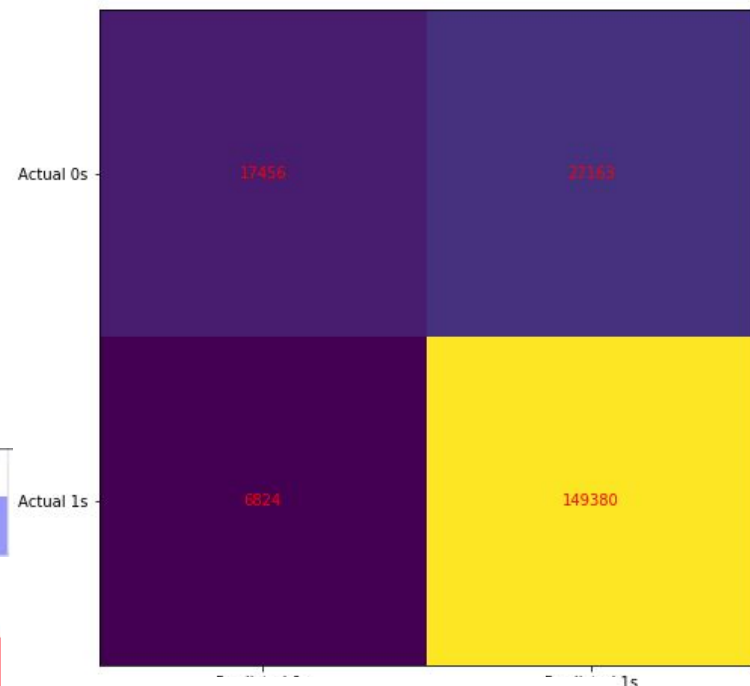
```
Z = pandas.DataFrame(numpy.array([[0, 1, 1000, 9.99, 0, 22],]),
                      columns=['a', 'b', 'c', 'd', 'e', 'f'])
```

```
Z = pandas.DataFrame(numpy.array([[1, 1, 1400, 14.50, 1, 1],]),
                      columns=['a', 'b', 'c', 'd', 'e', 'f'])
```

	0	1
0	0.324837	0.675163
1	0.553891	0.446109

Z represent the predictor variables, for the model, in the first example the regression recommend that you use channel 2 (1), the second example shows channel 1 (0) being recommend

Logistic Regression



$$R_{McFadden}^2 = 1 - \frac{\log(L_c)}{\log(L_{null})}$$

The Pseudo R-Squared != R-Squared
Pseudo R-Square = .2064, Good Model

Sources



<https://www.childstats.gov/americaschildren/tables/pop1.asp>

<https://www.statista.com/statistics/679812/number-of-households-with-children-by-age/>

<https://www.axios.com/80-percent-children-watch-youtube-dd3c76c6-7097-4f01-9ee2-d6b543888279.html>

<https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/>

<https://www.webfx.com/internet-marketing/how-much-does-youtube-advertising-cost.html>

<https://www.brandwatch.com/blog/youtube-stats/>

<https://support.google.com/youtube/answer/7628154?hl=en>

<https://www.wordstream.com/blog/ws/2014/03/17/what-is-a-good-conversion-rate>

<https://www.crunchsimplydigital.com/blog/influencer-marketing-conversion-rate/>

<http://investor.scholastic.com/static-files/10b155ba-8bfe-4a2a-aa04-fb7b4cf7b7c8>

<https://www.common sense media.org/sites/default/files/uploads/research/2019-census-8-to-18-full-report-updated.pdf>

<https://www.noxinfluencer.com/youtube/channel/UCR79JJpUPtZFwr7477TOMOg>

https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ft_19-04-10_socialmedia2019_facebookyoutube_2/

http://www.fldoe.org/core/fileparse.php/12087/urlt/GK_LanguageArts_Florida_Standards.pdf

Scholastic Analytics Challenge Description: (Full Story with all visuals)

Prior to conducting any analyses on the data, a precautionary measure was taken to ensure that no inaccuracies, biases, and misconceptions were present amongst the data. Since the dataset was large (two million records), an SQL database was created to access and query out our inquiries. Data exploration led us to clean the data using Python, and gain valuable insights through the utilization of Tableau.

Descriptive methods were applied to signal demographic trends among genres, themes, and prices. 25 genres were grouped into 4 unique genre categories: Instructional, Fiction, Media & Miscellaneous, and Non-Fiction. Comparing the average unit price to these genre categories shows that instructional books had the highest average unit price and that the remaining genre groups were relatively similar in terms of pricing.¹ From the 326 themes, 22 theme categories were created, these theme categories reflect the trend, shown by genres, that educational and instructional books were amongst the priciest on average.² A visualization with a cluster using the k-means algorithm was produced analyzing the average price of books per state. This visualization shows higher average prices in the southern US with the average price decreasing with northern mobility.³

A significant emphasis was placed on the identification of what the distribution channels were.⁴ So, an investigation examining the revenue by individual channels was conducted, and further drilled down into grade reading level book.^{4.1} The book reading levels (Lexile codes) were extracted from Scholastic's website to create grade reading level book classifiers. Starting from the 4th-grade reading level, there was an overall decline in revenue between both distribution channels.^{4.2} In support of this, Scholastic's website highlighted: "The division delivers more than 160,000 fairs in pre-K, elementary, and middle schools across the globe." These factors helped us conclude that channel 2 seems to exhibit the behaviors of a book fair. This assumption is backed by a near disappearance of revenue from channel 2 after the middle school reading level. The data would also suggest that channel 1 was an online platform, as revenue was spread throughout all reading levels. This insight leads to the recommendation of a selective distribution model involving channel 2 (book fairs) focusing on elementary to middle school book levels, while channel 1 (online) will be used as a general platform for all book levels.

We formulated a unified strategy for marketing between the two distribution channels. The overall view of channel revenue per reading level highlighted that the grade level of books contributing to the lowest source of revenue is 4th-12th grade.^{4.2} Our marketing strategy is to create a youtube campaign to promote and sell to the ages related to 4th-12th grade level books: 9-17 as well as parents who have children within that age range. Our unified marketing strategy is a youtube video advertisement, which promotes the scholastic brand across channel 1(internet platform) and channel 2(book fair). We created a cost structure of running youtube video advertisements for one month. The total market is based on data of children aged 9-17 that watch youtube and adults who have children ages 9-17 that watch youtube. The click-through rates & conversion rates are based on industry averages, and the cost of running ads is a function of the viewers of the advertisement. We ended up with 60,000 new customers and a total profit of \$515,400 in one month. We will also be using affiliate marketing. The purpose of a marketing strategy is to build brand reputation and awareness of the products and services available to the wider audience. A plan of action will involve generating traffic through partnered collaborations with media influencers that have a stronghold on our target demographic. More specifically, we will be reaching out to influencers who have the highest overlap rate with the scholastic youtube channel audience. We also included user stories to have a deeper understanding of our customer types for all aspects of the marketing campaign.

Going forward, Scholastic should attempt to minimize their overlap in their distribution channels by operating in distinct zip codes.⁵ Scholastic could use its historical data to create an algorithm based on a book's characteristics to make a recommendation as to which distribution channel a particular book is more likely to be sold in. A logistic regression model would be perfect because of attempts to predict a binary response, which would best suit Scholastic's needs by predicting which books could better be sold among the two masked distribution channels. In practice, a book would be scanned, and a distribution channel would be recommended, and unintentional competition would be prevented.

1 Average unit price by genre category

2 Average unit price by theme category

3 Average unit price by state

4 Revenue by channel, 4.1 Revenue by grouped book grade level and channel, 4.2 Revenue by individual book grade level and channel

5 Channel overlap by zip code