Big Data Final Project

Team Sixth Sense:
Jiayan Han, Ryan Chen, Sean Jung, Kasandra Woo

Agenda

- Background

- Data Description

- Insights and Analytics

- Q & A

Background

- Amazon users buy more books on average than other e-commerce sites

- Declining sales in the Books business segment on Amazon

- Learn more about this subset of customers

Data Description

Our dataset is from the Comscore data that represents a sample of US internet user's online purchases and demographics information

Column Name	Description
hoh_most_education	highest level of education completed by head of household
census_region	household's US geographic region
household_size	number of people in household
hoh_oldest_age	age of head of household
household_income	combined income of household
children	whether children present in home
racial_background	race of head of household
country_of_origin	Hispanic self-identification
zip_code	US postal code

Distribution by Categories

prod_category_id	total_purchase_amount
GENERAL BOOKS	4574.0
CHILDREN BOOKS	4423.0
TEXTBOOKS	2197.0
ART, MUSIC & PHOT...	717.0
LITERATURE & FICTION	531.0
CALENDARS	9.0
HEALTH, MIND & BODY	7.0
RELIGION	5.0
COOKING, FOOD & WINE	2.0
REFERENCE	1.0
BIOGRAPHY & MEMOIRS	1.0
HOME, HOBBIES & G...	1.0

| prod_category_id	total_purchase_price
TEXTBOOKS	64775.0
GENERAL BOOKS	64124.0
CHILDREN BOOKS	49468.0
ART, MUSIC & PHOT...	10235.0
LITERATURE & FICTION	6457.0
RELIGION	118.0
HEALTH, MIND & BODY	105.0
CALENDARS	84.0
COOKING, FOOD & WINE	21.0
BIOGRAPHY & MEMOIRS	5.0
REFERENCE	5.0
HOME, HOBBIES & G...	3.0

Zip code study

Textbooks

```
+--------+-----+
|zip_code|count|
+--------+-----+
|   91765| 62.0|
|   46590| 39.0|
|   75002| 33.0|
|   21229| 25.0|
|   77081| 20.0|
+--------+-----+
```
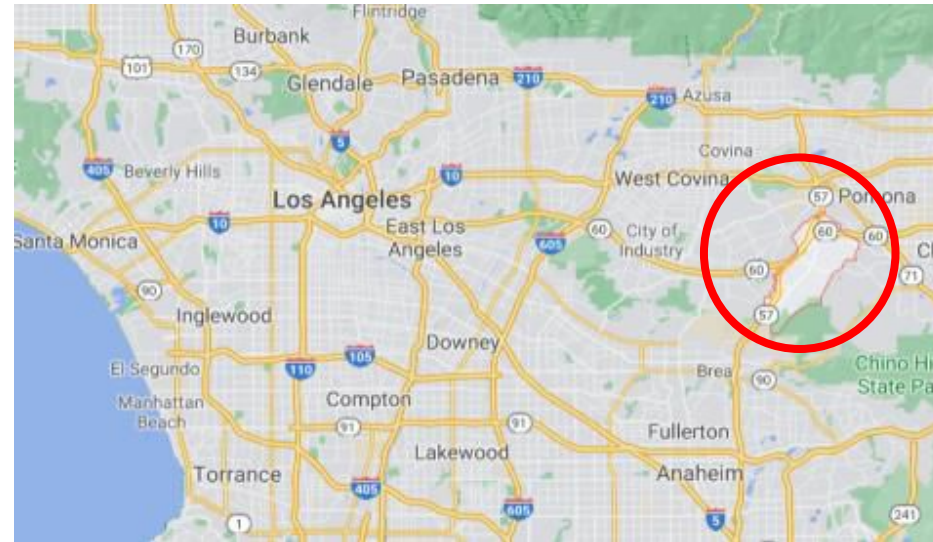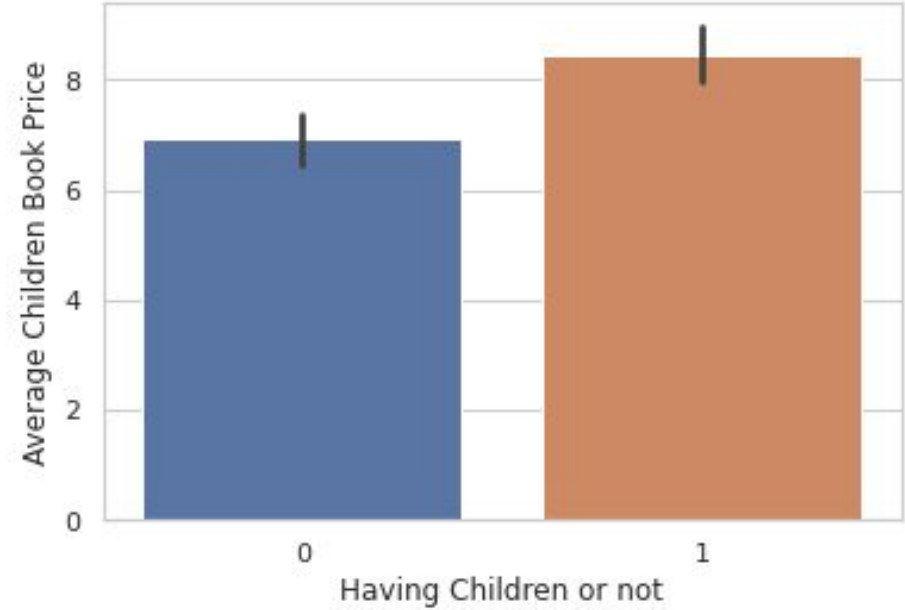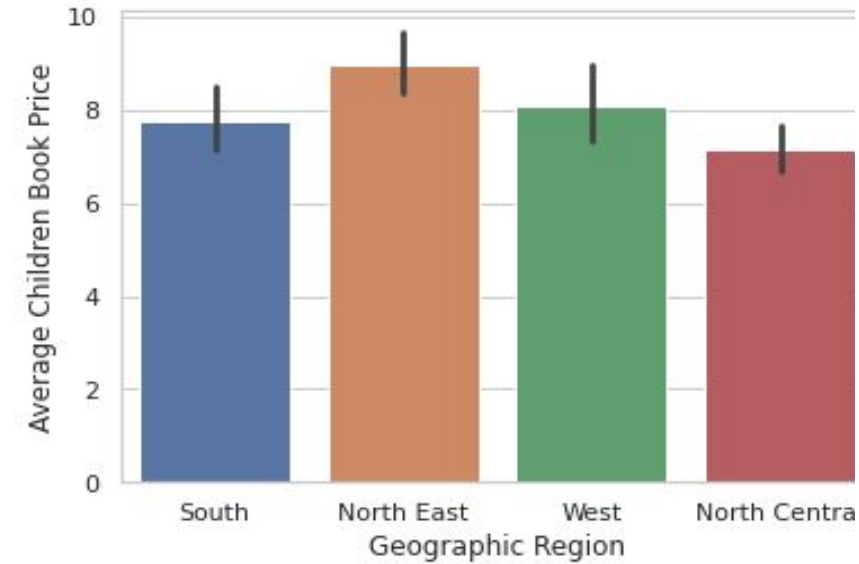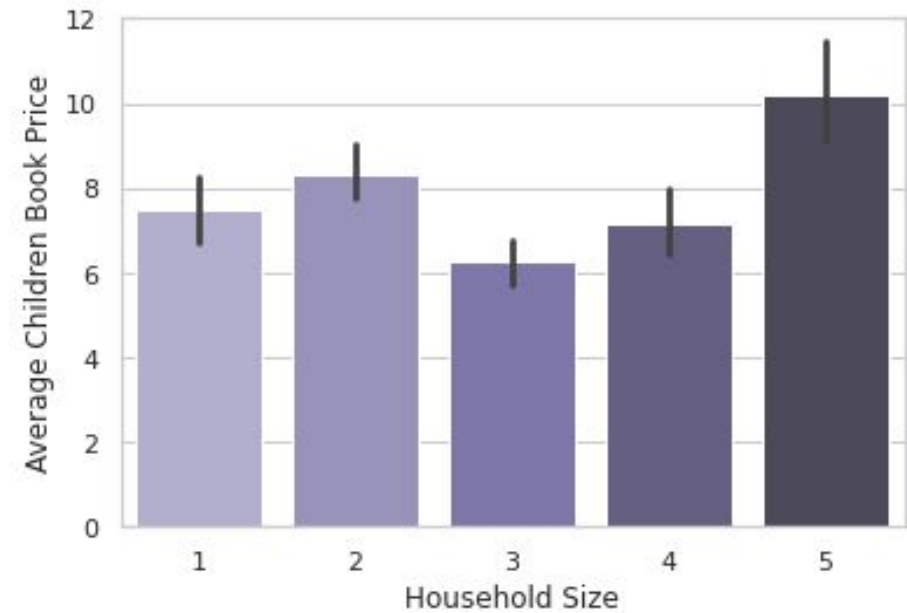

Race
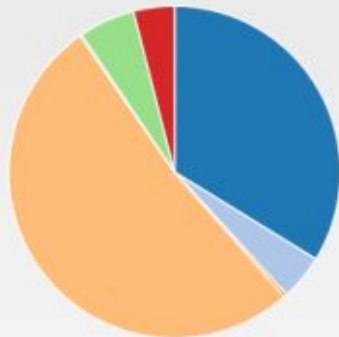
White		15,686	33.8%
Black Or African American		2,057	4.4%
American Indian Or Alaskan Native		154	0.3%
Asian		24,067	51.8%
Native Hawaiian & Other Pacific Islander		82	0.2%
Other Race		2,555	5.5%
Two Or More Races		1,856	4.0%

Children's Books

ID	Household Size
1	1 Person
2	2 People
3	3 People
4	4 People
5	5 / 5+ people

Children's Books vs Textbooks

Education Level
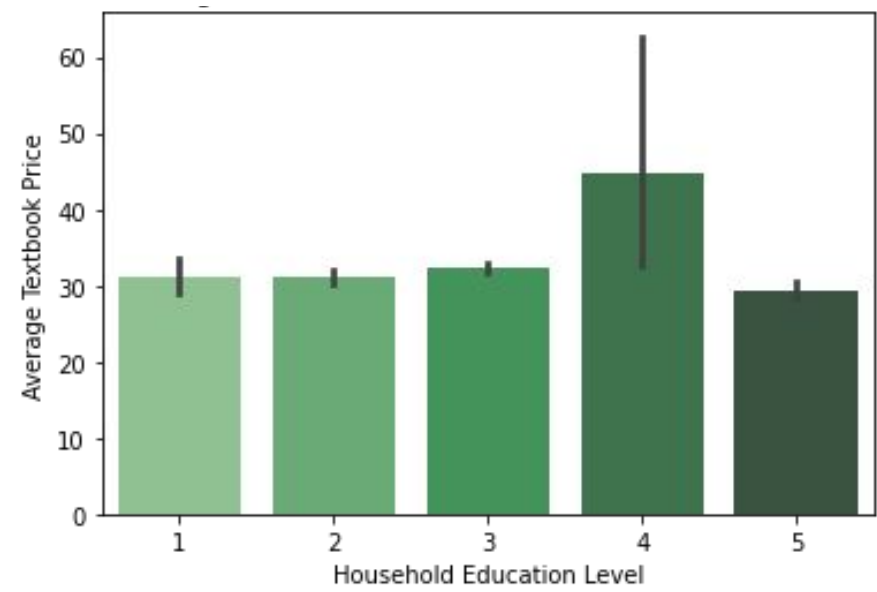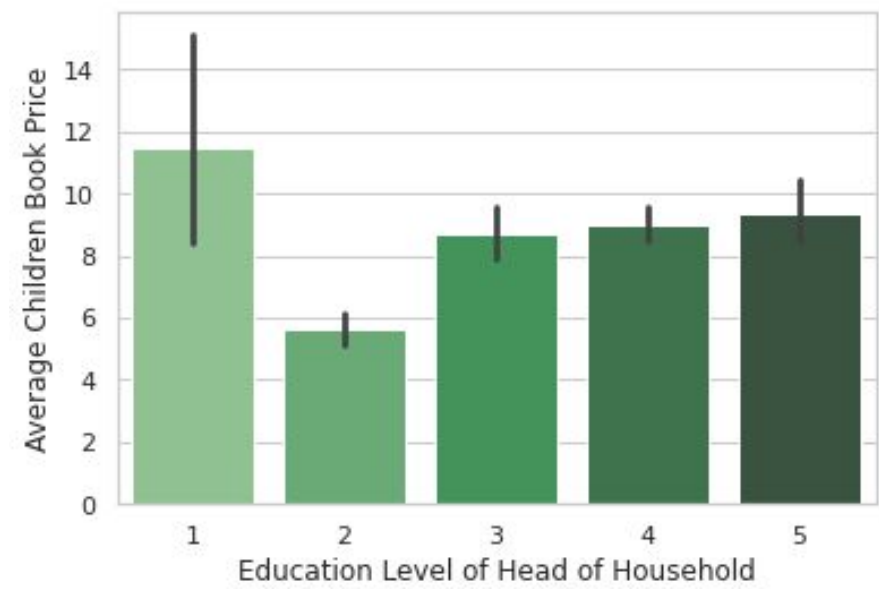
ID	Education Level
0	< High School
1	High School/GED
2	Some College, no degree
3	Associates Degree
4	Bachelor's Degree
5	Graduate Degree

Children's Books vs Textbooks

Household Income
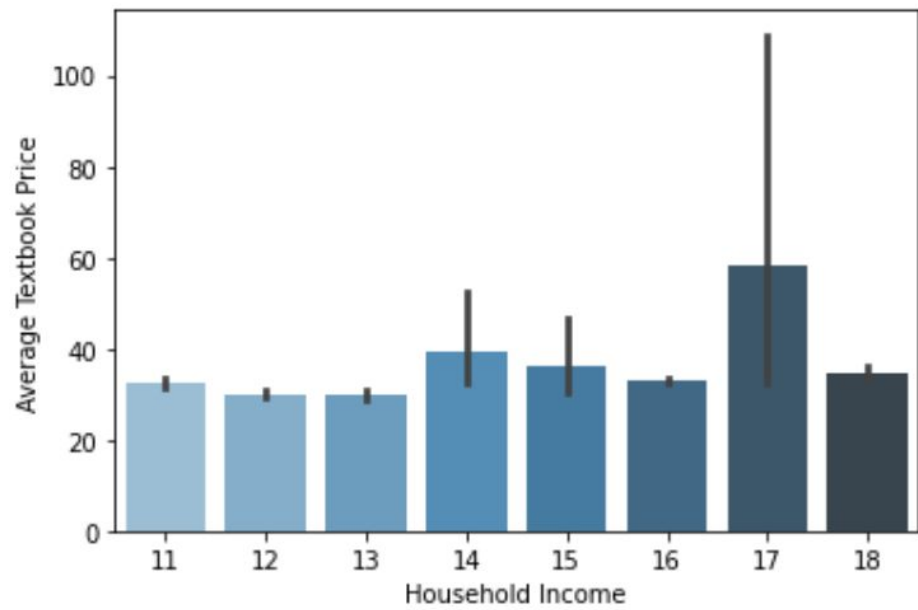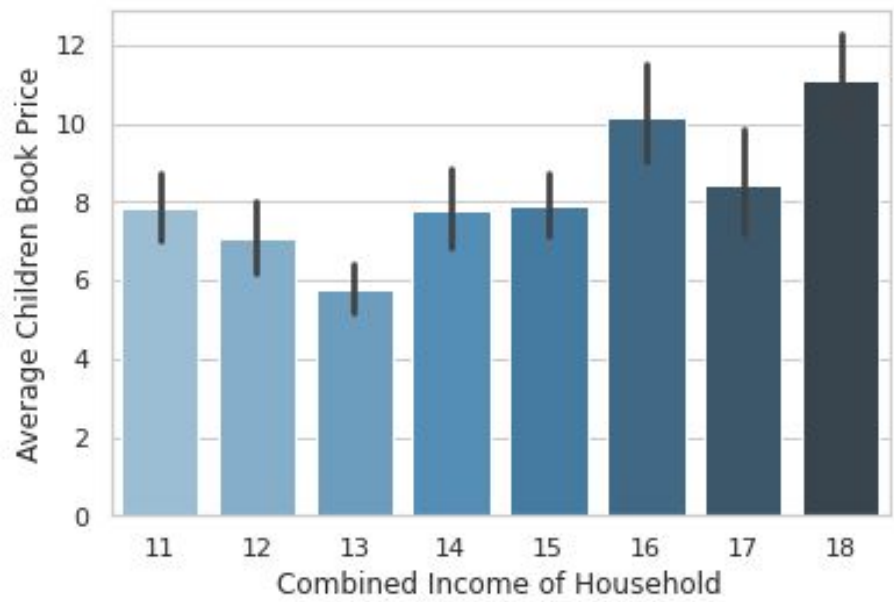
ID	Income
11	< $25,000
12	$25,000 – $39,999
13	$40,000 – $59,999
14	$60,000 – $74,999
15	$75,000 – $99,999
16	$100,000 – $149,999
17	$150,000 – $199,999
18	$200,000+

Children's Books vs Textbooks

Age of Head of Household
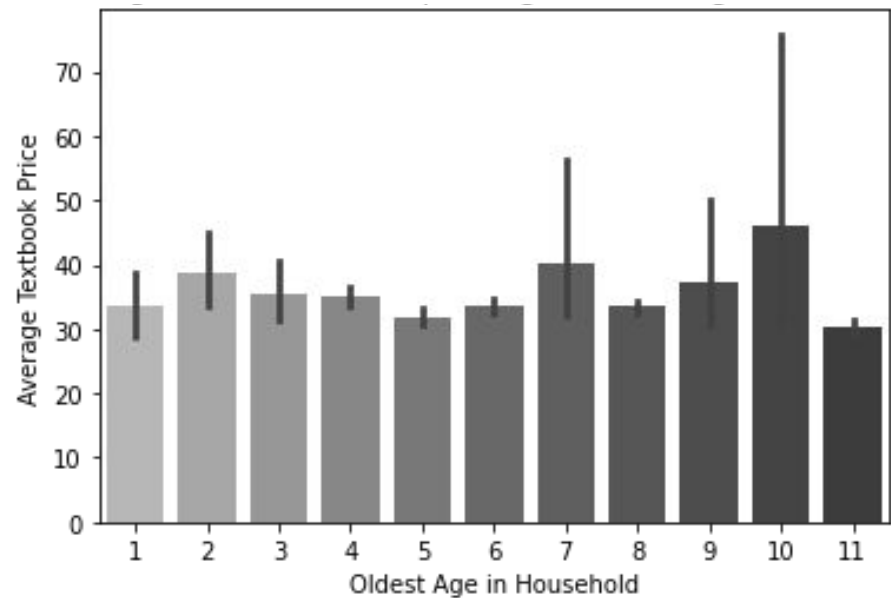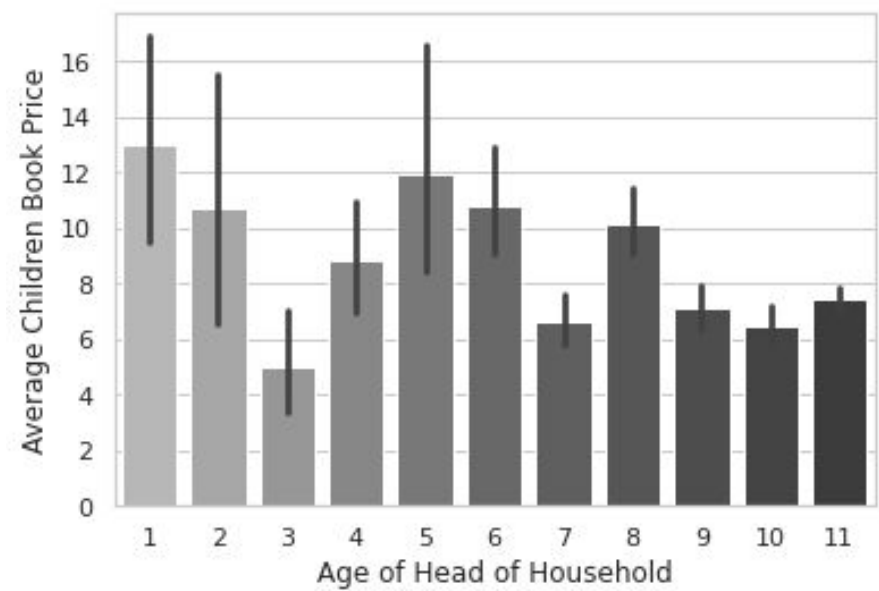
ID	Age Range
1	18-20
2	21-24
3	25-29
4	30-34
5	35-39
6	40-44
7	45-49
8	50-54
9	55-59
10	60-64
11	65 and Over

amazon

Challenges & Limitation

- Limited span of data

- Limited information on the individual consumers

- Limited categorical variables to describe customer's information

SWOT Analysis

- *Strengths (+)*
  - Correlation between each book category and different demographics
  - Correlation between each book category and geographic location
- *Opportunities (+)*
  - Allotting the budget effectively based off each identified demographic category and location
- *Weaknesses (-)*
  - Uncertainty about conversion rate
- *Threats (-)*
  - Uncertainty about the marketing investment and its ROI
  - Threats from competitors who can monetize such data first

amazon

Questions?