Olistの販売データから見る売上要因の推定

濱岡豊研究会 18期

窪田 禎之・前田 銀士

Outline

はじめに

先行研究　二次データ

付属資料

データ分析

回帰分析

考察

01

はじめに

- Kaggleが提供するデータをもとに、商品購入に関する分析を行う

- データ元: https://www.kaggle.com/olistbr/brazilian-ecommerce#olist_products_dataset.csv

02

はじめに – Olistとは

- ブラジルの電子商取引ブランドのひとつ
- 個人店主がインターネット上に出店することが可能なシステム　ECサイト

出所）https://angel.co/company/olist-com

03

はじめに

- 窪田の三田論で使用した画像枚数に関する先行研究が使える？

⇒画像枚数主体に研究

04

はじめに

- 注目したデータ
- product_photos_qty
  商品の画像枚数
- product_description_lenght
  商品説明の文字数
- product_category_name
  商品カテゴリー

05

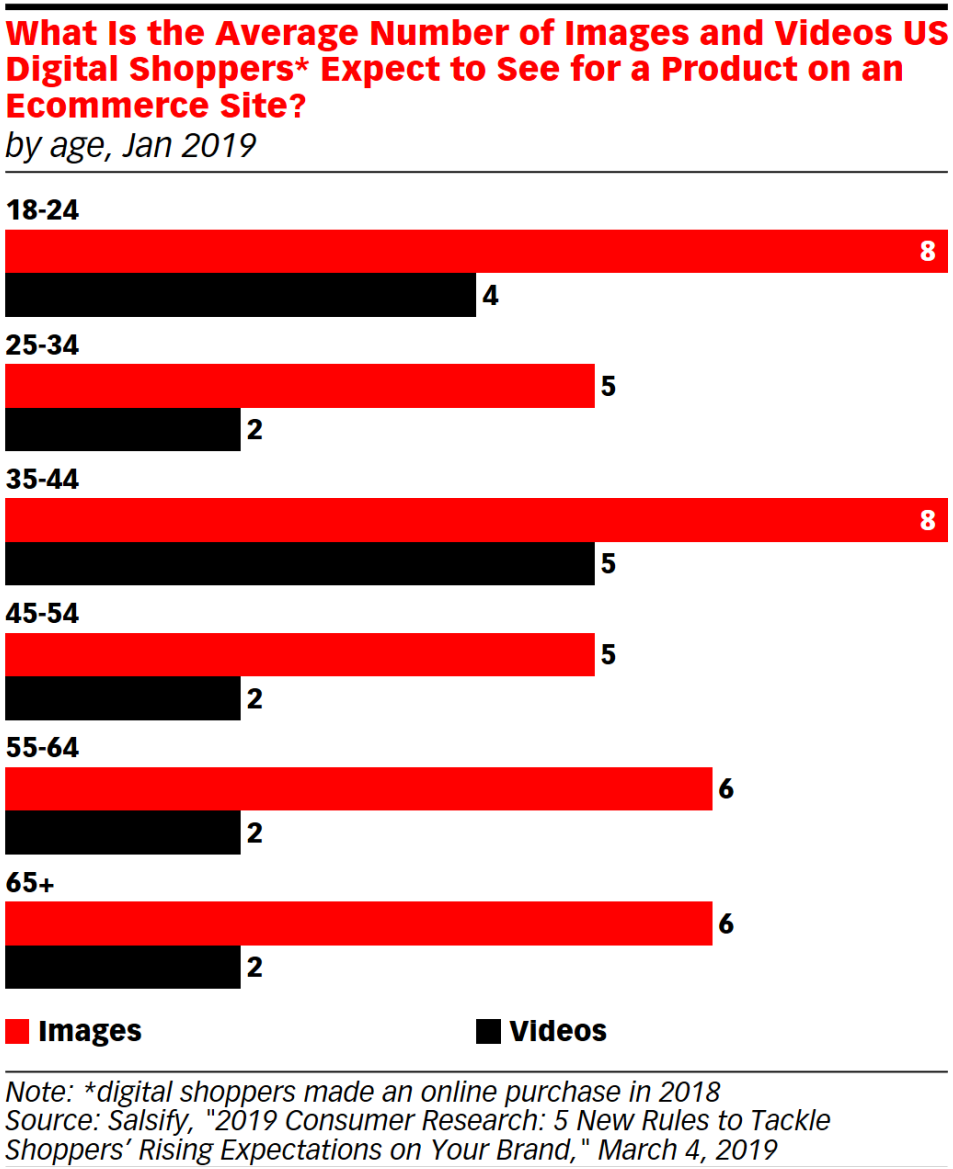
目的

- ECサイトolistにおいて、画像枚数と説明文量が販売個数にどう影響するのかを分析、考察する

先行研究＆2次データ

01 画像枚数

- アメリカのSalsifyが、ECサイトの画像と動画の閲覧数を調査した



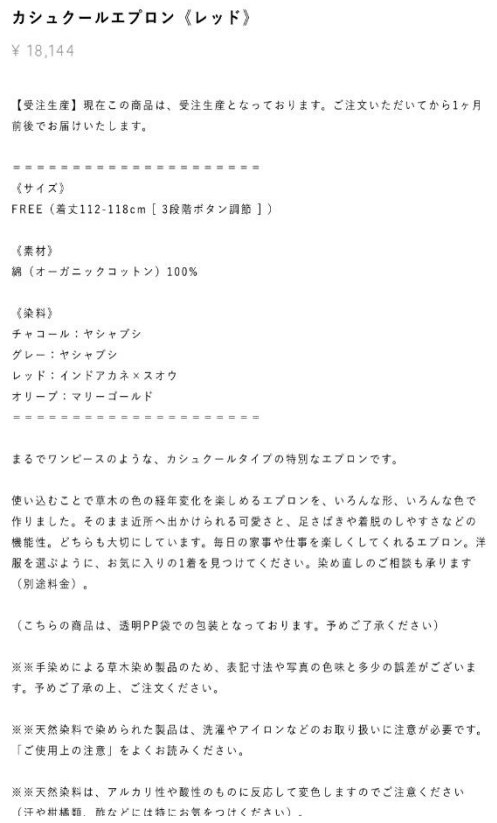
出所)　https://www.emarketer.com/chart/227373/async

02

画像枚数

- ネットショップを個人で作成可能なBASEでは1商品につき最大20枚まで商品画像を登録することができるが、直近12ヶ月間、毎月注文されているショップを当サイトが分析したところ、**売れているショップは1商品につき5枚以上の商品写真を登録している**ことがわかった。

出所）BASE https://baseu.jp/9184



カシュクールエプロン《レッド》

¥ 18,144

【受注生産】現在この商品は、受注生産となっております。ご注文いただいてから1ヶ月前後でお届けいたします。

＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝

《サイズ》
FREE（着丈112-118cm［3段階ボタン調節］）

《素材》
綿（オーガニックコットン）100%

《染料》
チャコール：ヤシャブシ
グレー：ヤシャブシ
レッド：インドアカネ×スオウ
オリーブ：マリーゴールド

＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝

まるでワンピースのような、カシュクールタイプの特別なエプロンです。

使い込むことで草木の色の経年変化を楽しめるエプロンを、いろんな形、いろんな色で作りました。そのまま近所へ出かけられる可愛さと、足さばきや着脱のしやすさなどの機能性。どちらも大切にしています。毎日の家事や仕事を楽しくしてくれるエプロン。洋服を選ぶように、お気に入りの1着を見つけてください。染め直しのご相談も承ります（別途料金）。

（こちらの商品は、透明PP袋での包装となっております。予めご了承ください）

※※手染めによる草木染め製品のため、表記寸法や写真の色味と多少の誤差がございます。予めご了承の上、ご注文ください。

※※天然染料で染められた製品は、洗濯やアイロンなどのお取り扱いに注意が必要です。「ご使用上の注意」をよくお読みください。

※※天然染料は、アルカリ性や酸性のものに反応して変色しますのでご注意ください（汗や柑橘類、酢などには特にお気をつけください）。

03 画像枚数

- Jifeng Maら(2019)

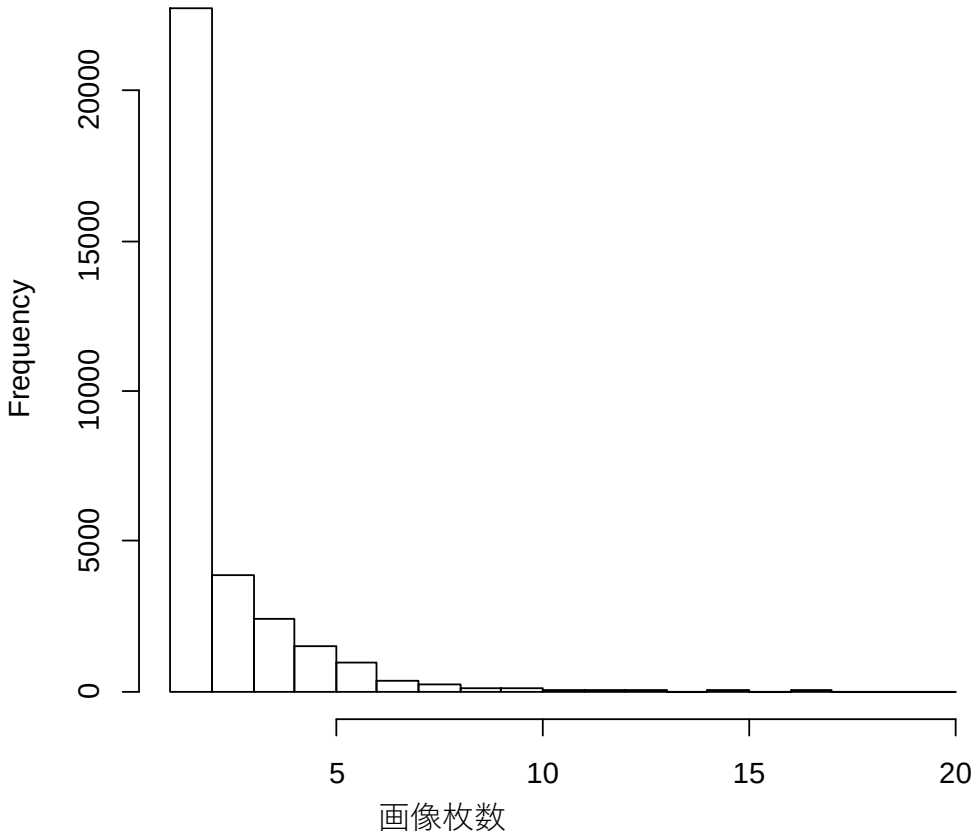
ユーザーイノベーションにおいて過度の記述的情報は情報過多につながるため評価プロセスに悪影響を与える。

プレゼンテーション資料の画像枚数が8枚になるまではイノベーションの採用可能性が上昇するが、それ以上だと減少していく。

出所)　Jifeng Ma, Yaobin Lu and Sumeet Gupta(2019) “User innovation evaluation: Empirical evidence from an online game community ”

データ分析

01 画像枚数



第1四分位:1.000

中央値　:1.000

第3四分位:3.000

平均:2.189

Max:20　Min:1

02

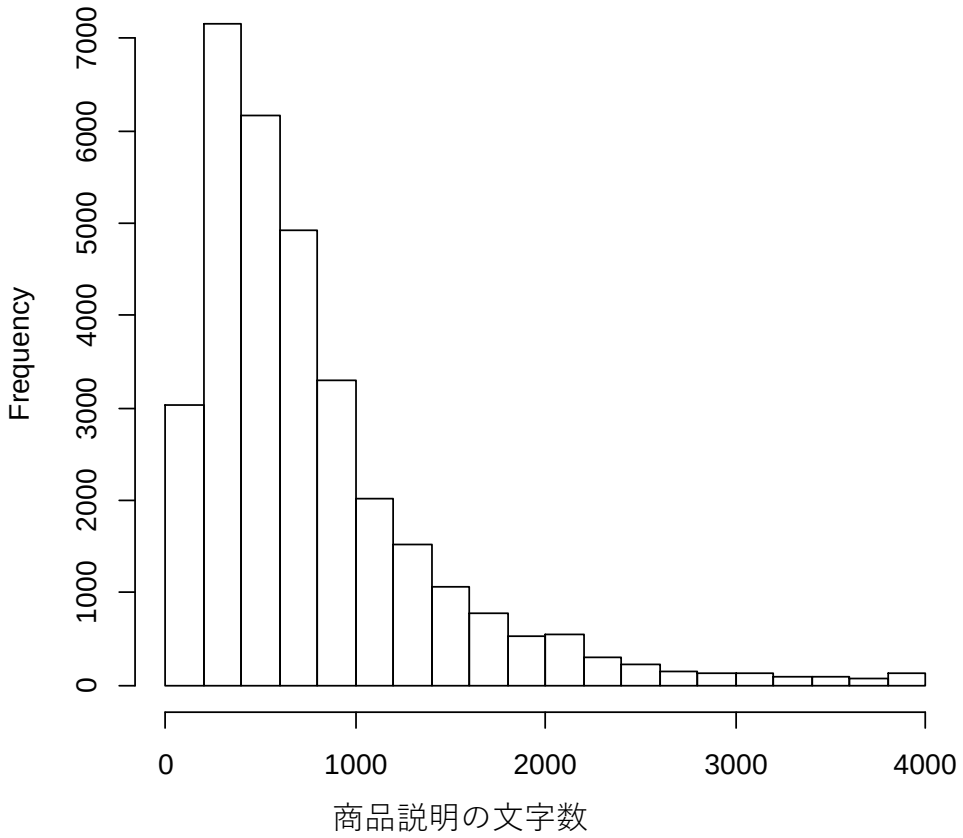
画像枚数

product_id	製品カテゴリー	画像枚数	seller_id	price	販売個数
f95d5d2156	おもちゃ	20	834f8533b2ecb6598dd004ff3de7203a	110.27	1
234495ab78	ベビー用品	19	834f8533b2ecb6598dd004ff3de7203a	138.94	2
e988004252	ペット用品	18	834f8533b2ecb6598dd004ff3de7203a	88.6	2
b659034bc6	ペット用品	18	834f8533b2ecb6598dd004ff3de7203a	88.6	2
f9aa001a85	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	89.15	2
26f4f1d683f	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	71.6	2
7f38cf4e517	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	77.7	2
28763a4fd1	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	71.6	1
801f0a5ea1	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	77.7	1
5948868c40	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	71.6	2
b085d8c884	ペット用品	17	834f8533b2ecb6598dd004ff3de7203a	71.6	1

- 特に画像枚数の多かった製品(17~20枚)を個別に調査

→すべて同じ販売元であったことが判明

03 商品説明の文字数



第1四分位:339.0

中央値　:595.0

第3四分位:972.0

平均:771.5

Max:3992.0　Min:4.0

04 商品説明の文字数

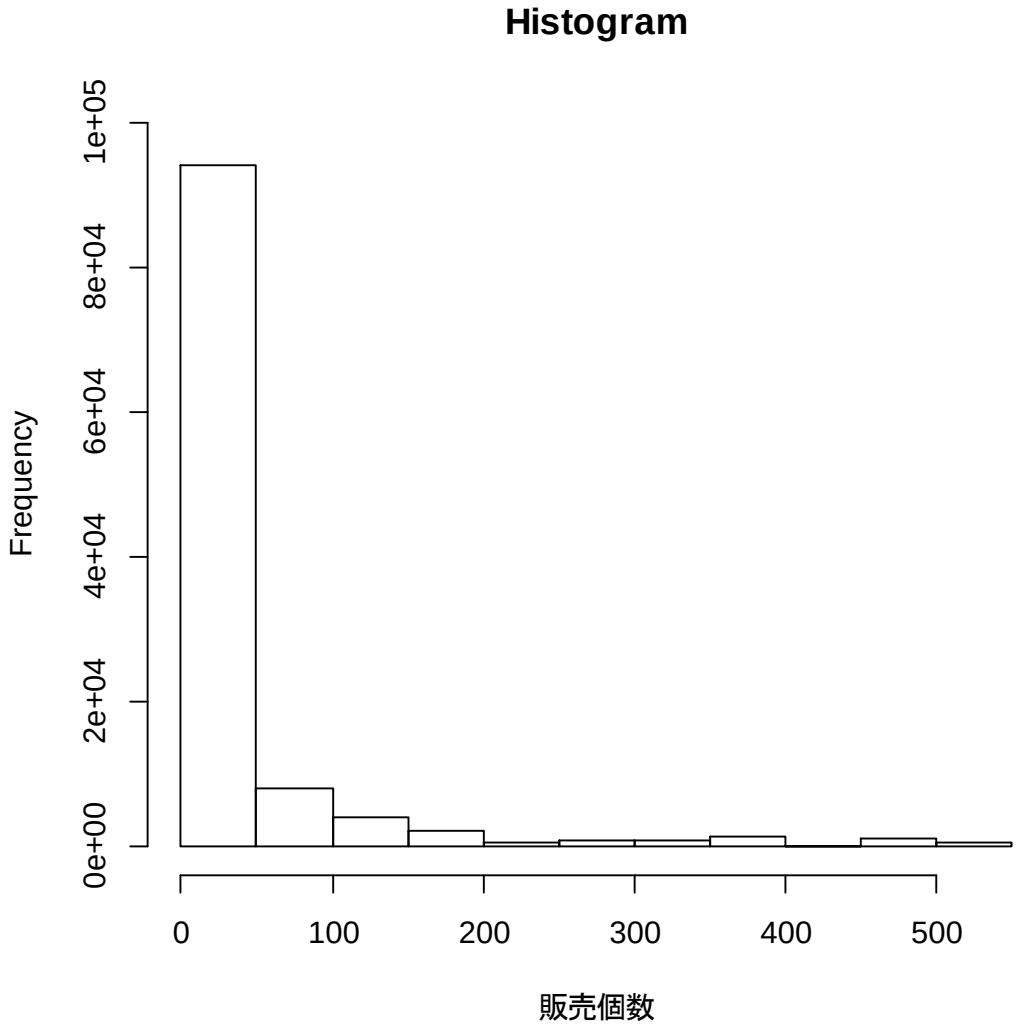
product_id	製品カテゴリ	説明文字数	画像枚数	重量(g)	縦(cm)	高さ(cm)	横(cm)	seller_id	price	販売個数
6d066313c	ベビー用品	4	2	855	36	23	28	9baf5cb77970f539089d09a38bcec5c3	148.99	1
67f95d5d2	家庭用品	4	2	681	39	7	28	933446e9a59dece7ae9175103820ca8f	64.99	1
fbf7215ba	家庭用品	4	2	5534	73	16	45	fe87f472055fbcf1d7e691c00b1560dc	299	1
dd5a5e7b2	家庭用品	4	1	1450	27	29	26	9baf5cb77970f539089d09a38bcec5c3	225.99	1
519f998a3	家庭用品	4	1	1045	39	16	38	9baf5cb77970f539089d09a38bcec5c3	107.99	1
0d17caa71	PCアクセサリ	8	1	825	30	16	16	0b90b6df587eb83608a64ea8b390cf07	163.73	2

- 説明文字数の少なかった製品(4~8文字)を個別に調査

⇒同じ販売元も含まれており、説明分量については販売元に影響を受けていると判明

05

各製品の販売個数

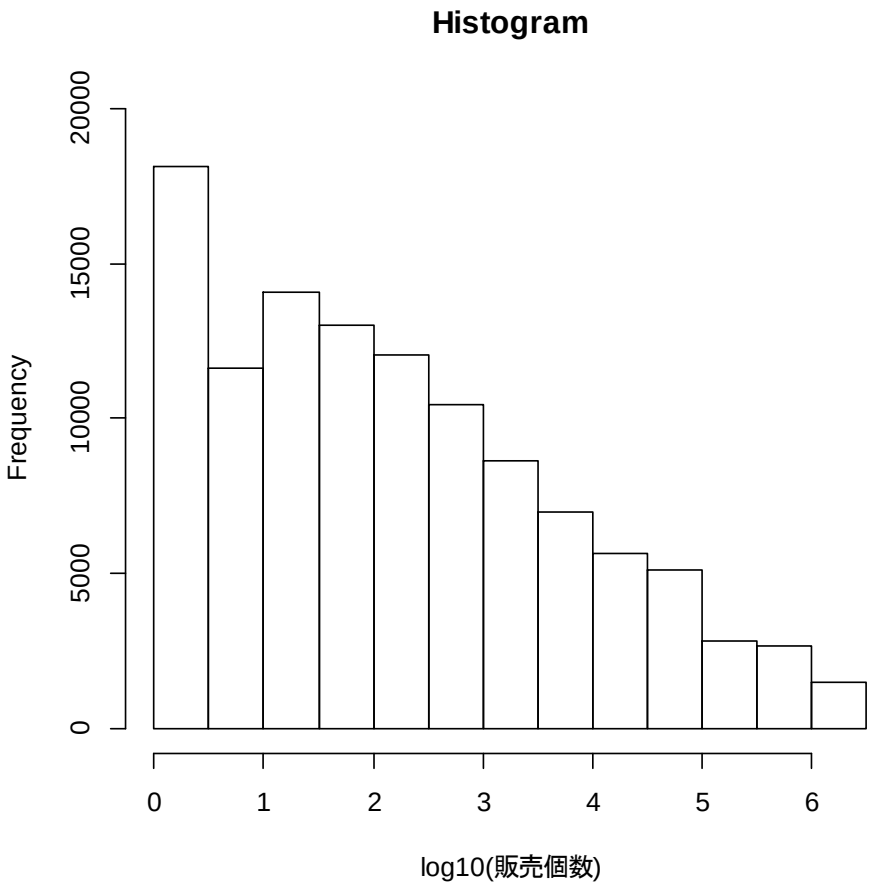


- 各製品の販売個数についてヒストグラムを作成した

⇒ほとんどの製品が1度しか購買されていないことが判明

06

各製品の販売個数



対数をとったヒストグラムも作成

⇒偏りはいくらか改善された

回帰分析

回帰分析

- 各製品の販売個数・画像枚数・説明文量　が一つのデータになるようマージ（Data3）
- 分析のため、各データを記号に置き換え

記号	データ名
a	販売個数
b	画像枚数
c	説明分量

01 重回帰分析①

res1<-glm(販売個数~画像枚数+説明分量,data=Data3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.834	0.476	79.476	2e-16***
画像枚数	0.956	0.143	6.671	2.55e-11***
説明分量	-0.004	0.000	-11.654	2e-16***

注）**有意水準** 0.1% : *** 1% : ** 10% : *
N=113425 , AIC=1292749 , R^2=0.001459

02 重回帰分析②

- 従属変数に対数をとる

res2<-glm(log10(販売個数)~画像枚数+説明分量,data=Data3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.946	0.004	228.606	2e-16***
画像枚数	-3.03e-04	0.001	-0.244	0.807
説明分量	6.40E-06	0.000	1.961	0.050*

注）**有意水準** 0.1% : *** 1% : ** 10% : *
N=113425 , AIC=238741 , R^2=1.664e-05

03 重回帰分析③

- 従属変数に対数,説明変数を2乗

res3<-glm(log10(販売個数)~(画像枚数)²+(説明分量)²,data=Data3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.950	0.003	361.962	2e-16***
(画像枚数)²	-4.25E-04	0.000	-2.841	0.005**
(説明分量)²	3.03E-09	0.000	2.853	0.004**

注）**有意水準** 0.1% : *** 1% : ** 10% : *
N=113425 , AIC=238730 , R^2= 0.0006061

04 ポアソン回帰分析①

res4 <- glm(販売個数 ~ 画像枚数+説明分量,data=Data3, family = poisson)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.639	0.001	3753.520	2e-16***
画像枚数	0.026	0.000	91.030	2e-16***
説明分量	-1.30E-04	0.000	-156.840	2e-16***

注）有意水準 0.1% : *** 1% : ** 10% : *
N=113425 , AIC =10330904 ,

05

ポアソン回帰分析②

- 説明変数を2乗

res5 <- glm(販売個数 ~ (画像枚数)² + (説明分量)² ,data=Data3, family = poisson)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.616	0.001	5900.440	2e-16***
(画像枚数)²	1.12E-03	0.000	33.630	2e-16***
(説明分量)²	-2.89E-08	0.000	-104.280	2e-16***

注）**有意水準** 0.1% : *** 1% : ** 10% : *
N=113425 , AIC =10349676 ,

06

回帰分析まとめ

	N	AIC	R^2
重回帰分析①	113425	1292749	0.001459
重回帰分析②	113425	238741	1.66E-05
重回帰分析③	113425	238730	0.0006061
ポアソン回帰分析①	113425	10330904	-
ポアソン回帰分析②	113425	10349676	-

⇒重回帰分析③の結果を採用

考察

01 考察

- 画像枚数は販売個数に対して逆U字型の影響を与える⇒採択(1%水準)
- 枚数が多いと売れなくなるという結果は想定外だが、視覚的な情報量が飽和する程、その商品の粗が目立ってしまうことがあることが要因なのではないかと推測した。

02 考察

- 商品説明の文字数は販売個数に対してU字型の影響を与える→採択(1%水準)
- 文字数が少ない商品は画像や認知度等により明らかに分かりやすい商品であり、多い商品は詳細に説明されているために分かりやすい、という解釈が可能だが、意外な結果となった。

03 今後の課題

- 提供されたデータが限られており、個々の商品について分析を行うことができなかった。
- 決定係数が非常に低く、採択された仮説についても、その影響が非常に小さくなってしまうことが懸念される。
- 商品カテゴリごとに分析を行うことにより、さらに興味深い結果が得られるのではないかとと考える。

謝辞

データセットを提供していただいたOlist社ならびに、分析の場を設けていただいたKaggle.com に感謝の意を表します。

Acknowledgements

We thanks to Olist for releasing this dataset. We also appreciate Kaggle.com for enabling analyze this data.

Reference

・Jifeng Ma, Yaobin Lu and Sumeet Gupta(2019) “User innovation evaluation: Empirical evidence from an online game community”（12月23日最終アクセス）

・Emarketer

(https://www.emarketer.com/chart/227373/async)（12月23日最終アクセス）

・Angellist

(https://angel.co/company/olist-com)（12月23日最終アクセス）

・BASE

(https://baseu.jp/9184)（12月23日最終アクセス）

付属資料

教授が出された課題

課題Ⅰ

- Q1　customer_idとcustomer_unique_idはそれぞれ何を意味するか考えなさい。

Customer_id

商品取引ごとに発行されるID

Customer_unique_id

顧客ごとに発行される識別子

- Q2　顧客が何回購買したかを算出するにはどうすればよいかを考えなさい。

それぞれのCustomer_unique_idがデータ内に存在する数

課題Ⅱ

- Q1　order_idとorder_item_idはそれぞれ何を意味するか考えなさい。

order_id

注文ごとに発行される識別子

order_items_id

一度の注文で商品をいくつ購入したか表す値

課題Ⅱ

- Q2　一回の注文で何アイテム購入されたかを算出するにはどうすればよいかを考えなさい。

order_items_id の数値の最大値を確認する

- Q3　一回の注文での支払い額は合計でいくらになるかを算出するにはどうすればよいかを考えなさい。

同一のorder_id 内のprice(商品価格)×order_items_id(個数)とfreight(配送料)の合計を算出

- Q4　格sellerの販売回数(注文回数)を算出するにはどうすればよいかを考えなさい。

同一seller_idのアイテム数を確認

- Q5　格productの販売回数(注文回数)を算出するにはどうすればよいかを考えなさい。

同一product_idのアイテム数を確認