7COM1079-0901-2019

Title:
The rating distribution across categories of sales. An analysis of the Brazillian E-commerce Dataset

Coursework - 53
18057684 Ruchi Pandey
18044900 Taimur Harris Ali Khan
18054909 Olorode Abiodun John
19000972 Felix Chukwueke
18057439 Timothy Arum

December, 2019

1. Introduction

The rapid growth of E-commerce in Brazil is led by Finance, manufacturing and retail sectors. Due to its extremely uneven income distribution to enable the citizens of Brazil to completely engage with B2C consumerism. As the industry develops the legislation and various government policies also have to be updated in order to accommodate it as a viable business. Incorporating its changing relationship to the world's economy. Other opportunities like selling imported goods and participation in value chains on a global scale (Tigre, 2003).

Olist is Brazil's largest department store for small businesses from every state to connect with one another. Merchants have the ability to sell products on Olist and ship them by partnering with various logistic partners. The dataset needed to be consolidated as it was split into files classified as per the information selected from the site. The categories of these sales are of wide ranges as they enable sellers from different states to make their products available all over the country. The ratings or review score should be more or less normally distributed but it isn't found to be so. Over a dataset of 100,000 values approximately, the data appears to fluctuate. Do the ratings of a particular category have an influence in the sales in any particular state in Brazil?

2. Background

Brazil's (Tigre, 2003) distribution of income is very unevenly distributed consisting of the one with the means to an income and infrastructure and the second being the poorer population which lacks the necessary tools to be a part of the "modern information economy". Liberating the economy helps other sectors to push through the global economy. A global supply chain comprising of a more inclusive environment and further involvement of e-commerce and other tech services. The idea of an Electronic Data Interchange (EDI) was first used in the finance sector and it took about another 10 years for other industrial sectors to catch up. Now, Brazil is most active internet user with a guesstimate of a third of the share of total users.

The effect (Cui, Lui and Guo, 2012) that a large number of online reviews have on the number of views of a product to showcase their experience with their product. There is a negative bias which has a sizable effect than the positive reviews do. The ratings of the products is usually prominently displayed on the site, thus the effect of the reviews greatly affects the customer's evaluation of the product. As consumers cannot get a feel of the product directly and the customer evaluations tend to be less symbolic of the product's quality. Thus the number of reviews as well as their volume is important for sales of a new product.

3. Research Question

How does the rating distribution across the categories affect their sales across the different states in Brazil?

4. Method

Data Consolidation

This dataset was found on kaggle titled "Brazilian E-Commerce Public Dataset by Olist"This particular dataset took some degrees of pre-processing. Gathering the data required from 9 sub data files and collect them, arrange them as well as translate them from Portuguese. Also, data about the population distribution in Brazil was collected open source site. MS Excel was mainly used to compile the data as per requirements. The order id uniquely identifying every case.
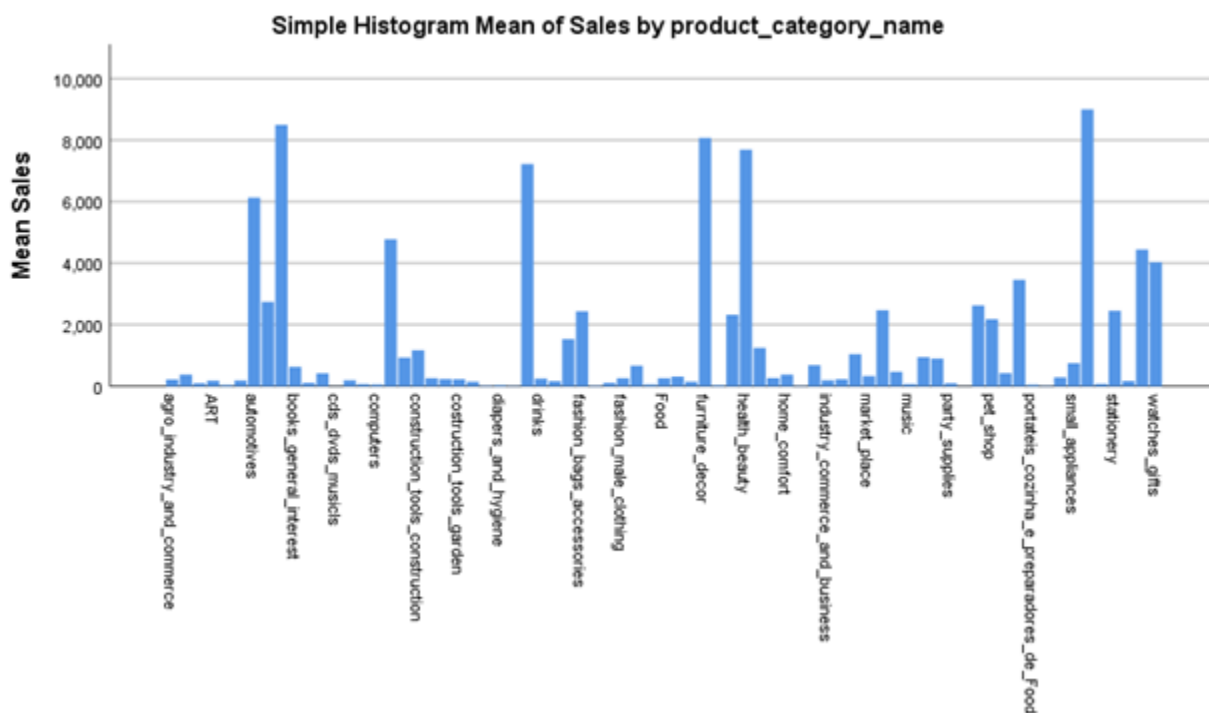
5. Results

Chart-01

The above histogram depicts the product categories most popular on the site as represented by the higher number of sales of which the highest was small appliances with a confidence interval between 8000 and 10000 sales with the least from agricultural and industrial, ecommerce, art, cds and dvds, fashion, food and music from 2016 to 2018.
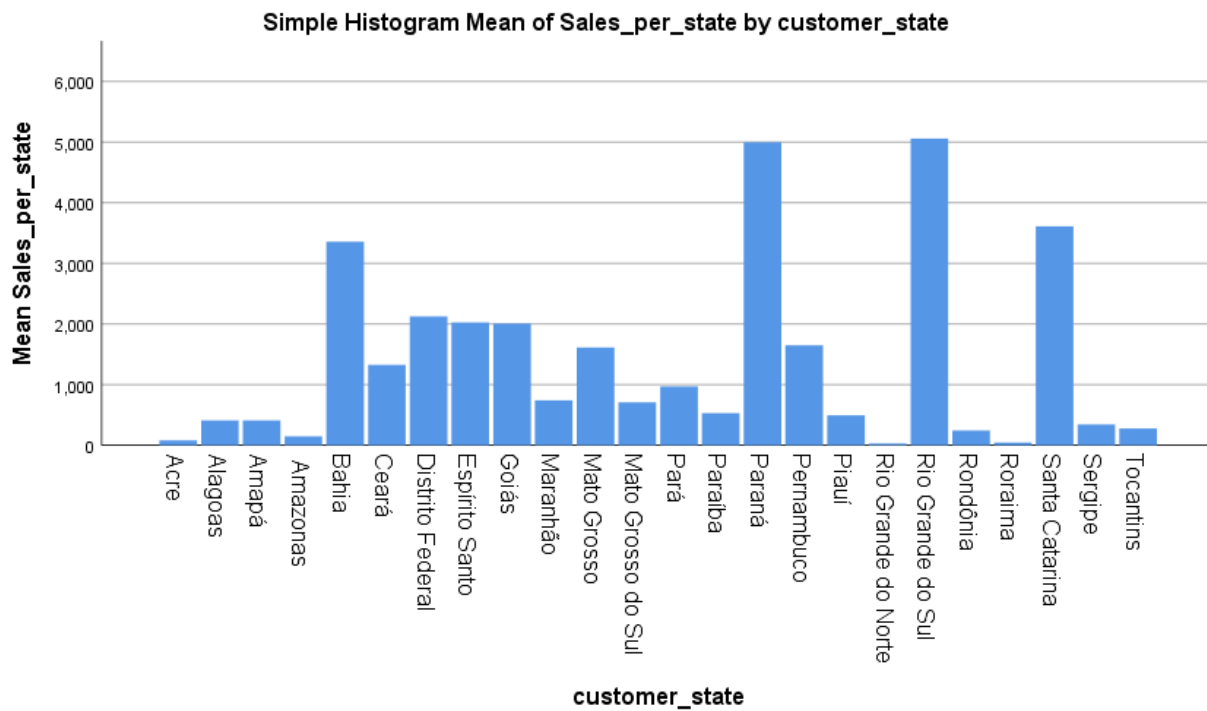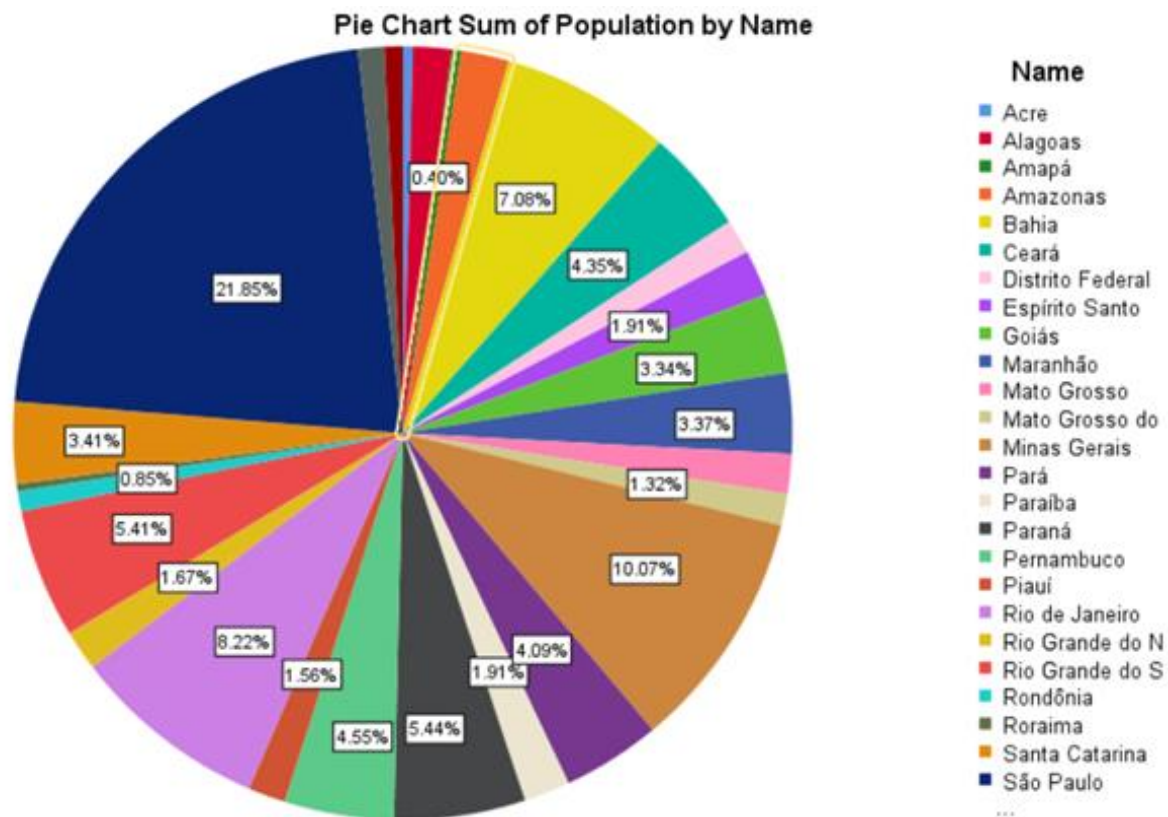

Chart-02

Pie Chart Sum of Population by Name

Name	
■	Acre
■	Alagoas
■	Amapá
■	Amazonas
■	Bahia
■	Ceará
■	Distrito Federal
■	Espírito Santo
■	Goiás
■	Maranhão
■	Mato Grosso
■	Mato Grosso do
■	Minas Gerais
■	Pará
■	Paraíba
■	Paraná
■	Pernambuco
■	Piauí
■	Rio de Janeiro
■	Rio Grande do N
■	Rio Grande do S
■	Rondônia
■	Roraima
■	Santa Catarina
■	São Paulo

Chart-03

From chart 2 and 3, excluding major states like Rio de Janeiro and Sao Paolo, we can see that the mean of the sales in Rio Grande Do Sul and Parana have higher sales even though they are less populated than other major states like Bahia.
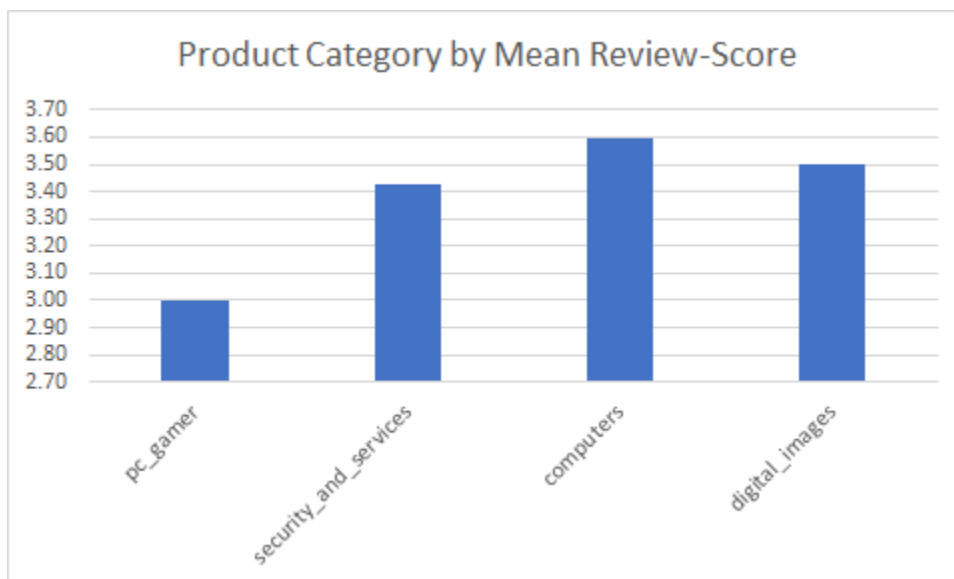
Product Category by Mean Review-Score

Chart-04

Chart 04 shows the product categories selected displaying the IT services like digital images, gaming with security services and computers are the lowest product categories while arts, music auxillary hardware and fashion for children are the highest reviewed products. There were only 4 sales of gaming products while CDs and DVDs have had only one sale with a 5 rating. Although, when we look at a distribution of sales over 100 the least rated.
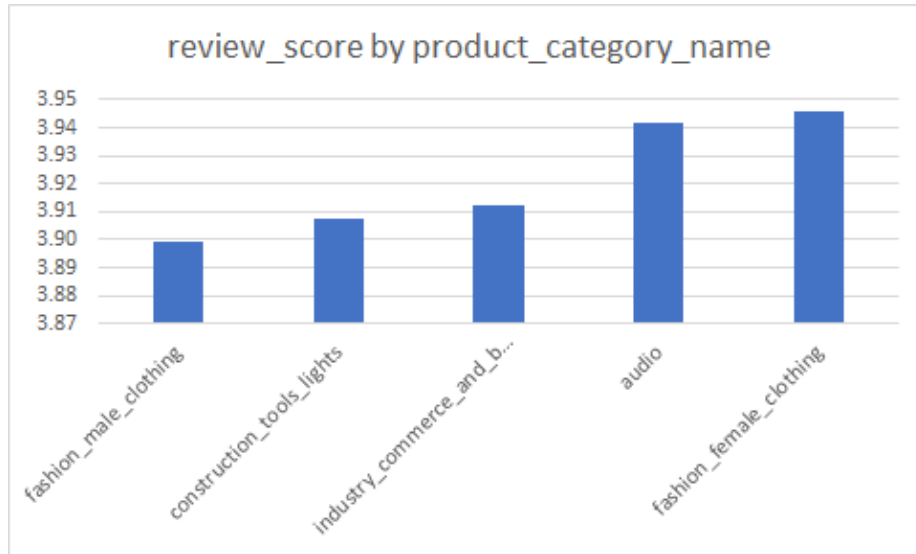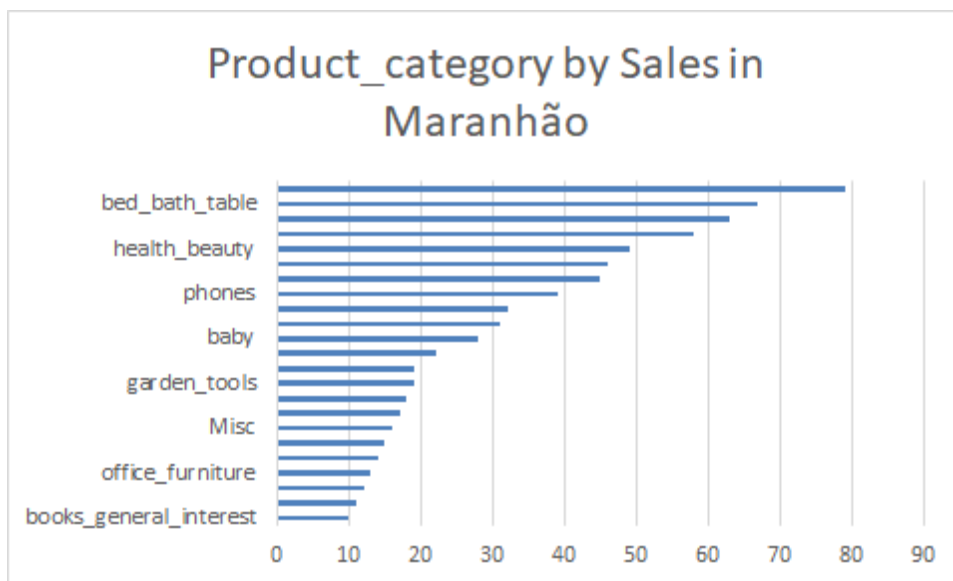

review_score by product_category_name

Chart-05

From chart 02 and 03 we observe that we know that Acre and Amapa have similar population distributions but Amapa has almost double the sales. This could be due to the fact that Olist services 21% more of their product categories Amapa.
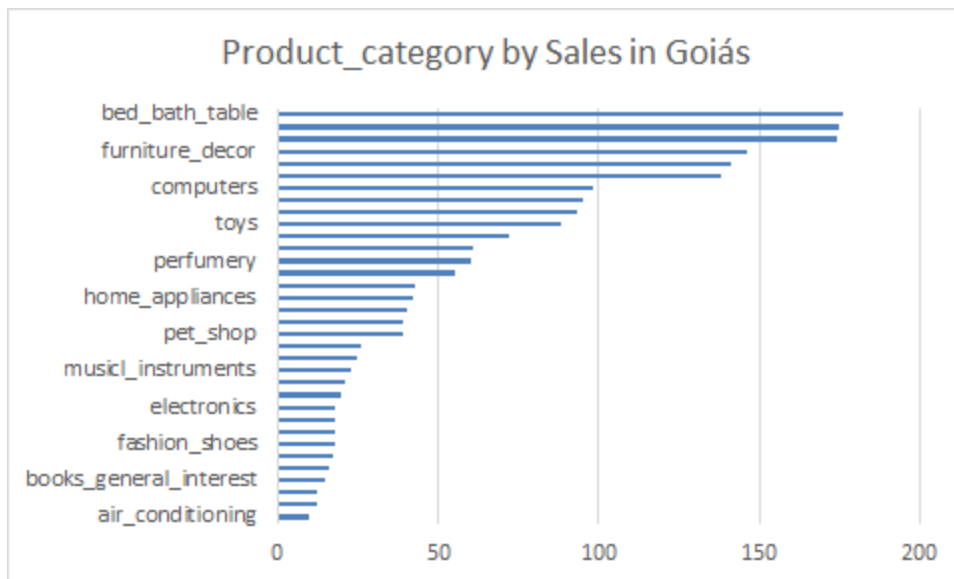

Product_category by Sales in Maranhão

Chart - 06

As seen from charts 05 and 06, even though the population is similar in Maranhao and Goias the sales are more than double of that of Maranhao. Even though the popularity of the product categories remain the same, the distribution of sales is drastically different. With only 740 sales in Maranhao and 2006 in Goias. This can be seen due to the higher popularity in usage of the site for a variety of products.

a)

review_score Mean by product category for Gaios
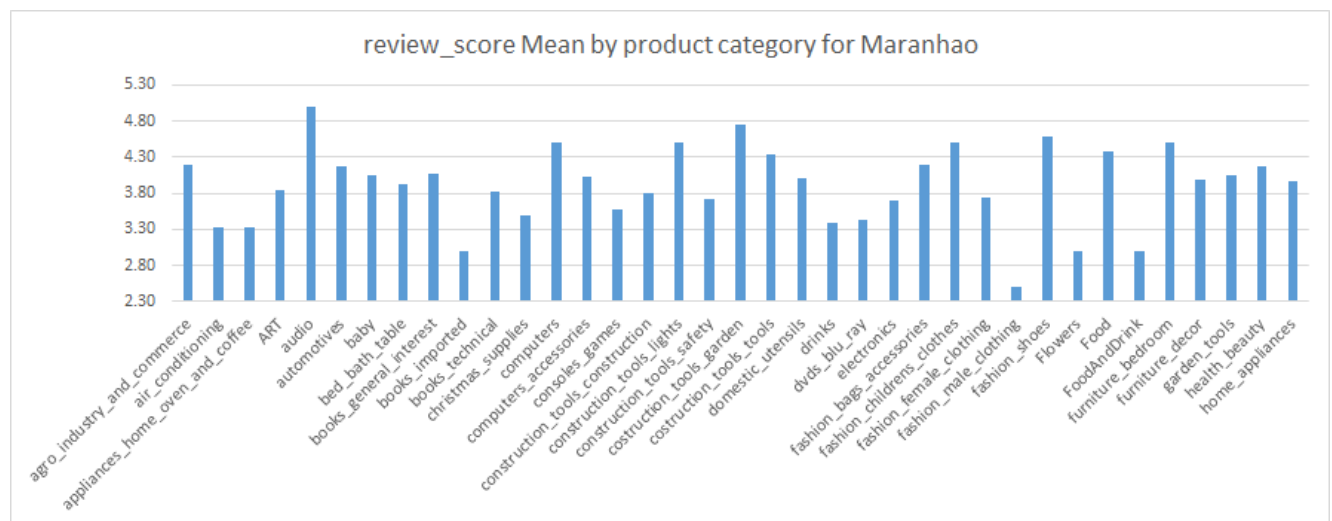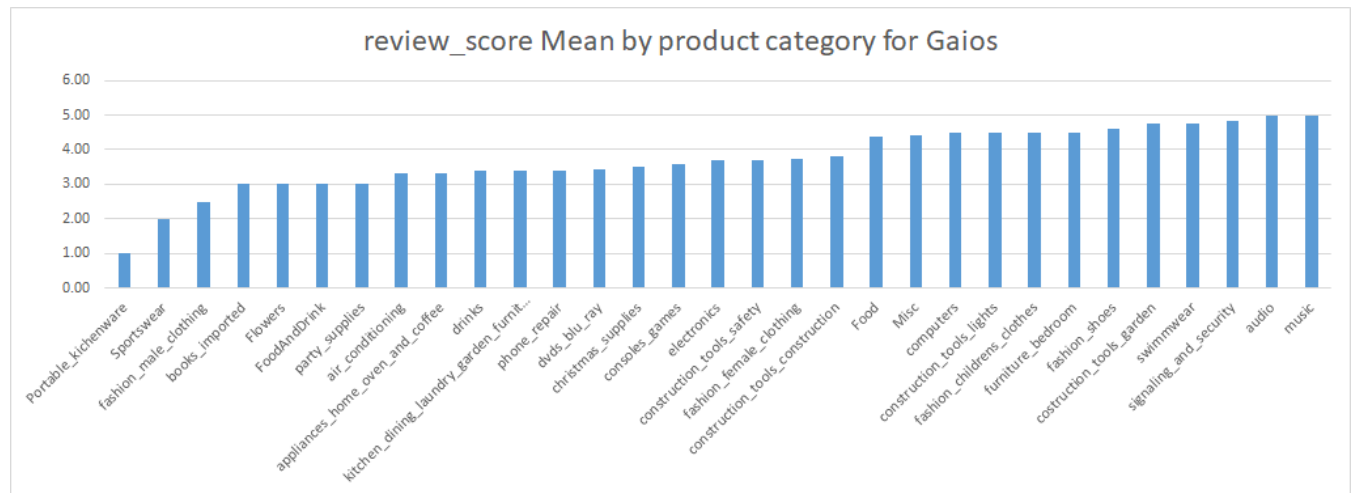
b)

Chart - 07

Chart - 07a shows the distribution of reviews of categories over Maranhao with the least number of sales for male fashion clothing while portable kitchenware has the least sales in Gaios seen chart 07b. The highest sales in audio for Maranhao and music for Gaios.

6. Discussions and Conclusions

The product categories with higher ratings have the highest sales across Brazil. In cases of Maranhao and Gaios, these states have equal population distribution yet the sales for Gaios were undoubtedly higher than the former.

Group Statistics

	customer_state binary	N	Mean	Std. Deviation	Std. Error Mean
sales_by_Cstate	4	2006	102.03	60.365	1.348
	5	740	102.82	59.432	2.185

		Levene's Test for Equality of Variances		t-test for Equality of Means
		F	Sig.	t
sales_by_Cstate	Equal variances assumed	.779	.377	-.305
	Equal variances not assumed			-.307

The difference in the means for the number of sales between the two customer states, Gaios (4) and Maranhao (5). There is a 0.76 % difference in means. Both states have similar mean sales

by category. The p-value is 0.761, our null hypothesis will be that there will not be a difference in means of sales. Since the p-value is greater than 0.05 we have failed to reject our null hypothesis thereby concluding that there is no significant difference.

Group Statistics

	customer_state binary	N	Mean	Std. Deviation	Std. Error Mean
review_score	4	2006	4.04	1.347	.030
	5	740	3.75	1.515	.056

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
review_score	Equal variances assumed	40.756	.000	4.798	2744
	Equal variances not assumed			4.546	1195.481

The difference of mean is 7.18% which is low indicating that both the states share similar reviews for different product categories, however, the p-value observed indicates that there is statistical significance i.e. a statistical difference does exist. Therefore we reject the null hypothesis that there is no difference between the means.

7. References

a. Tigre, P. (2003). Brazil in the Age of Electronic Commerce. *The Information Society*, 19(1), pp.33-43.

b. Cui, G., Lui, H. and Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), pp.39-58.

c. Ibge.gov.br. (2019). *IBGE | Portal do IBGE | IBGE*. [online] Available at: https://www.ibge.gov.br/en/home-eng.html [Accessed 20 Dec. 2019].

d. Kaggle.com. (2019). *Brazilian E-Commerce Public Dataset by Olist*. [online] Available at: https://www.kaggle.com/olistbr/brazilian-ecommerce/ [Accessed 20 Dec. 2019].

Appendix:

Bitbucket	https://bitbucket.org/cw53/cw53/src/master/
Trello	https://trello.com/invite/b/TK9D8PA9/720050511c916f336e0201e134f17ce5/coursework-53-team-research-and-development-project