



DATA ANALYSIS

This report is an investigative analysis of the fashion-product category sales data and industry performance of WISH. The data set was gotten from Kaggle and it consists of product listings, ratings, and sales performance of Wish Marketplace for the month of June, 2020.

It explores the correlations and patterns with regard to the product sales while incorporating features of descriptive, sales, statistical and predictive analysis.

2.1 Descriptive Analysis

This shows the distribution of the variables used for this analysis. It aims to investigate questions such as; what colours and sizes are the users buying the most on the platform? Is there a difference in the total in the sales per country on the platform?

In Table 2.1, the average units of product sold is 4339 with 1 and 100000 being the least and highest respectively. There is also a wide gap between the average retail price (23.29) and the price (8.33) of products and negative mean value of the gap-price. Wish can maximize AI in its price strategy in order to know when best to give a discount and to maintain prices as observed in other market place.

Table 2.1 Descriptive Statistics of Numeric Variables

N o.	Variabl e Name	Min	1 st Quartile	Media n	Mean	3 rd Quartile	Max	Standar d Deviatio n	Variance	Coeff. Of Variance
1.	Units Sold	1	100	1000	4339	5000	10000 0	9356.54	87544828	2.156379e+ 00
2.	Price	1.00	5.81	8.00	8.33	11.00	49.00	3.93	15.46	4.722948e- 01
3.	Retail Price	1.00	7.00	10.00	23.29	26.00	252.0 0	30.36	921.60	1.303549e+ 00
4.	Gap Price	-244	-18.00	-.20	-14.96	1.00	7.00	29.40	864.31	- 1.964753e+ 00
5.	Rating	1.00	3.56	3.85	3.82	4.11	5.00	0.52	0.27	1.348829e- 01
6.	Rating Count	0.00	24.00	150.0 0	889.00	855.00	20744	1983.93	3935974	2.229987e+ 00

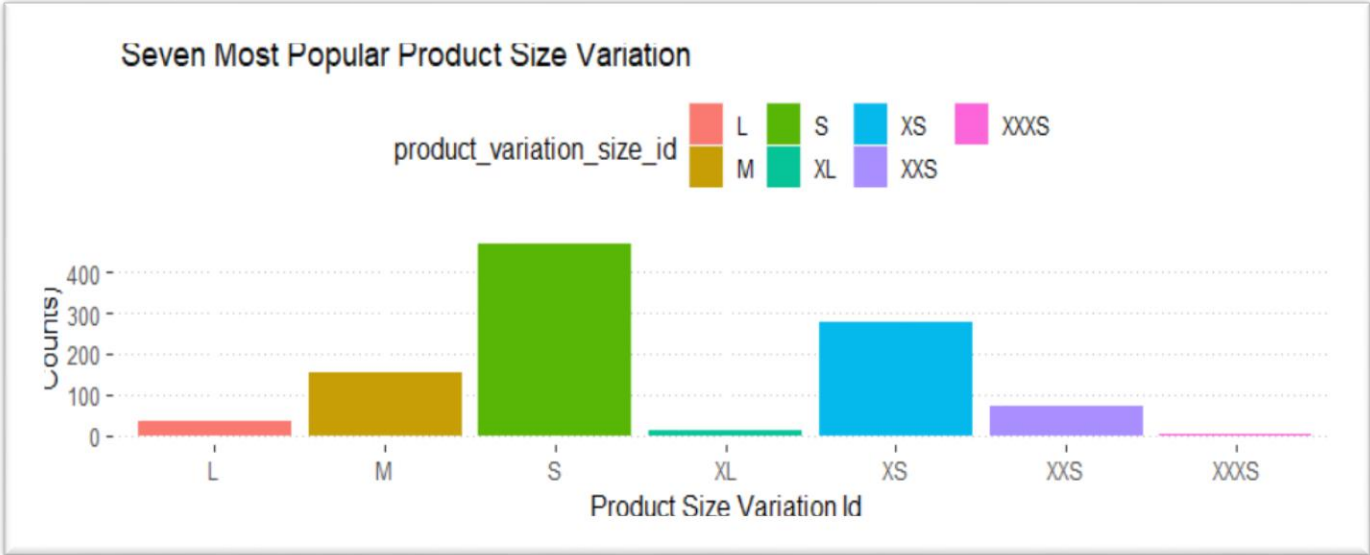


fig 2.1.1 Plot Of Seven Most Popular By Sizes

As shown in fig 2.1.1 and fig 2.1.2 size-variation S (small) and colour black are the most popular features buyers are purchasing on the platform. From fig 2.1.1, size 'S' has over 400 counts while XXXS, the least

bought size is barely above counts. Therefore, an automated analysis of sales data will improve trend and sales forecast.

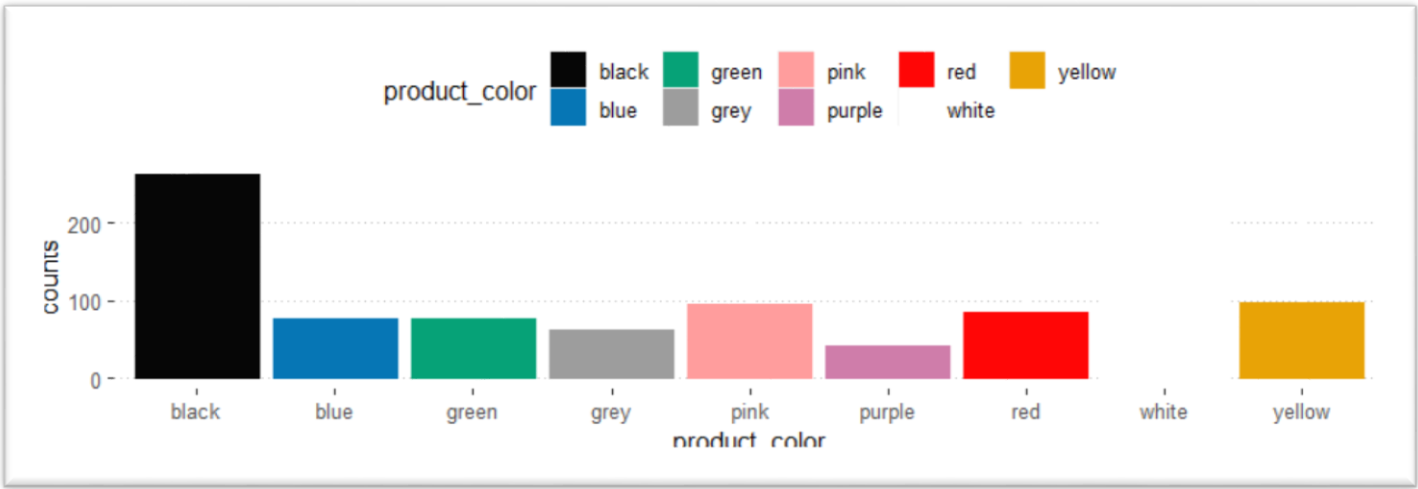


fig 2.1.2 Plot Of Top Products By Colour

In the same vein, product sales on Wish originate from different countries and fig 2.1.3 shows that over 90% of the products come from China. This is an essential feature to note because buyers on the platform come from different parts of the world. With AI, merchants on Wish can optimize supply chain and provide faster better shipping and delivery system for the users.

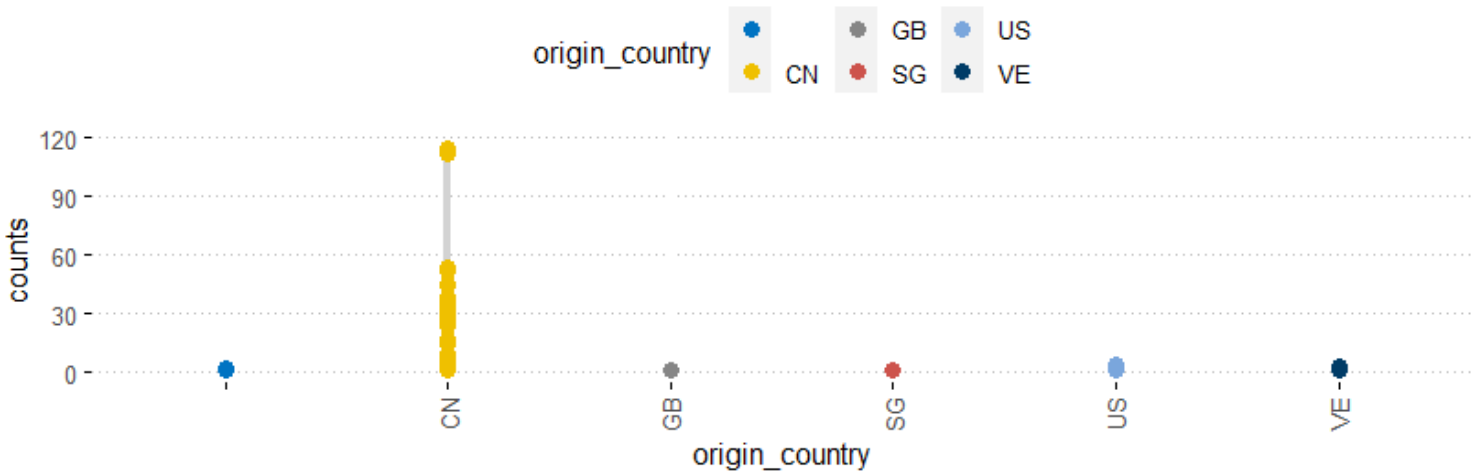


fig 2.1.2 Product Count By Origin Country

2.2. Sales Analysis

As stated in appendix, the gap price is the difference between the price on the platform and the retail price on other market places and fig 2.2.1 reveals that there is no stable trend on the platform. At a gap price of 0, about 50,000 products were sold and this was also observe at gap price of -100 and -50. However, the

highest number of goods were sold at a gap price between -50 to 0. With AI, the merchants can know the best time to drop prices so as to have the best sales.

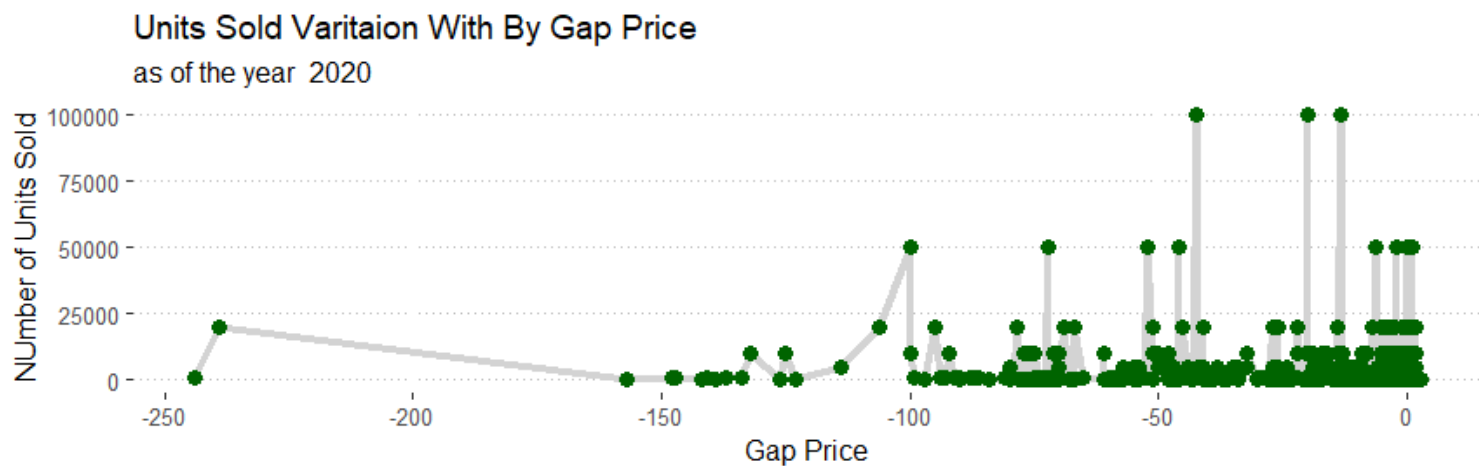


fig 2.2.1 Plot Of Variation in Units Sold By Gap Price

A look at the total and average sales on the platform reveals that does size ‘S’ have the highest count, it also has the highest average sales of 9625 units, followed by ‘M’(medium) and ‘L’(large) at 7036.4 and 5990.3 respectively. Fig fig 2.2.2 and Table 2.2.3 shows a breakdown of the units of products sold based on size variation. It can be inferred that most of the users that purchased clothing item on the platform in June 2020 wore small or medium clothing sizes. However, Wish can build a model to reveal how significant this relationship is in order to make better decisions.

	product_variation_size_id	total_sales
1	S	2260986
2	M	1069537
3	XS	712987
4	L	215650
5	XL	115500
6	XXS	108480
7	XXXS	1200

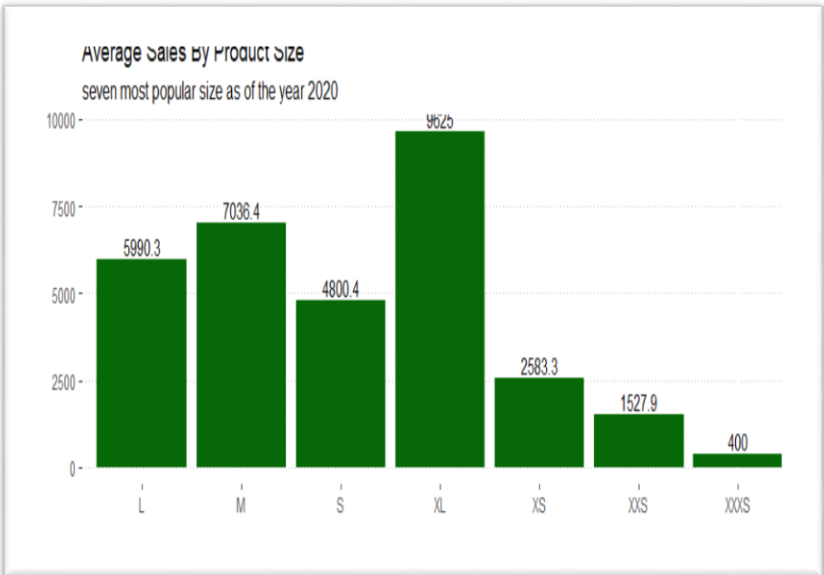


fig 2.2.2 Total Units Sold By Product Size

fig 2.2.3 Average Units Sold By Product Size

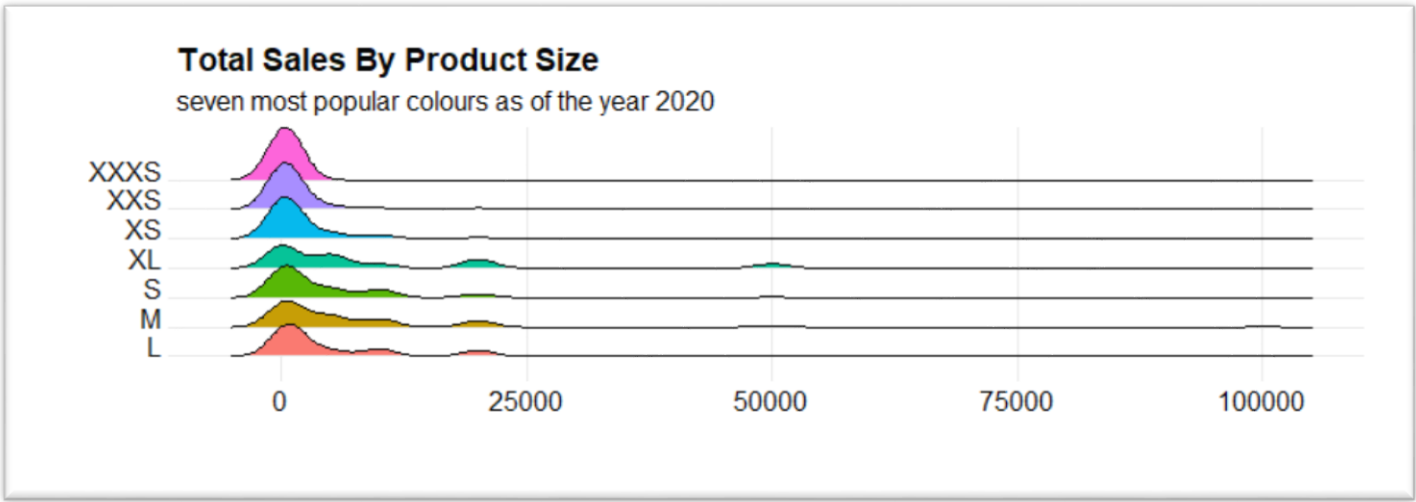


fig 2.2.4 Average Units Sold By Product Size

In terms of rating counts, fig 2.2.2 reveals an unstable trend in the rating count, consequently, the number of units sold is also not stable. Nevertheless, as the rating counts increased to 10000, there was a steady trend in units sold and then a sudden spike at around 17000 rating counts which led to an increase in sales to 100000. Hence, with AI, Wish can build a relationship with its users so as to prompt the users to give ratings as they make purchases on the platform.

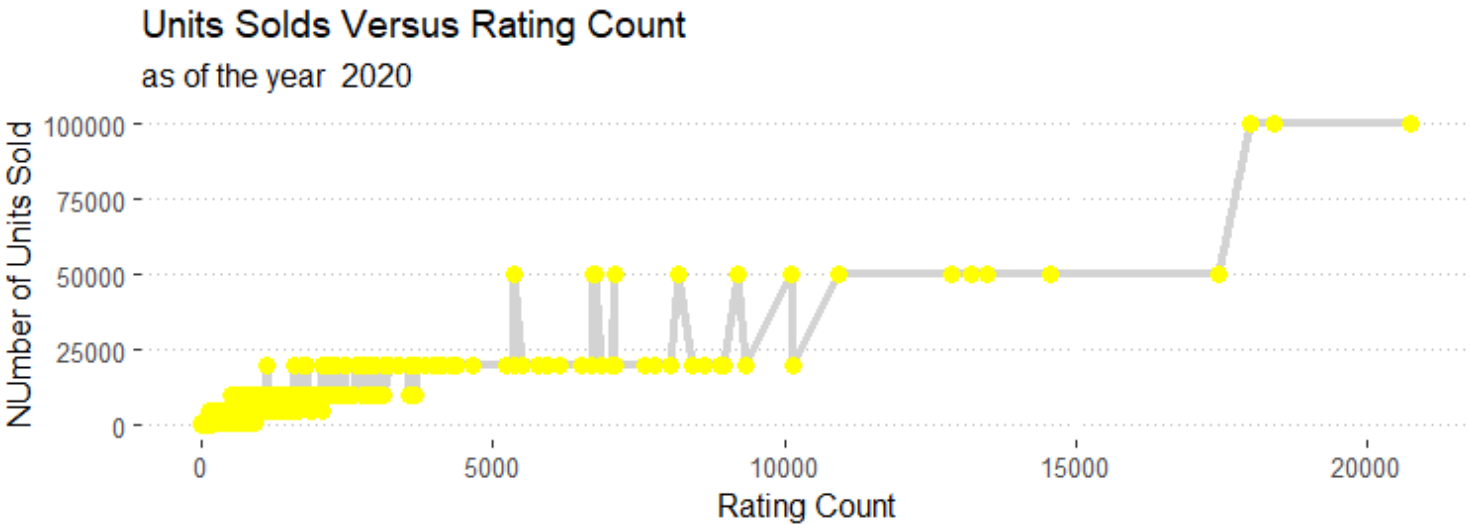


fig 2.2.5 Units Sold Versus Rating Count

fig 2.2.6, fig 2.2.7 and fig 2.2.8 gives a breakdown and sales pattern by product colour. Despite the fact that Black has the highest count (fig 2.1.2), it is however blue most sold item, with a total sales of 14553450 and an average sales of 5547.5. A predictive model can reveal whether or not product colour can predict units sold.

	product_color	total_sales
1	black	1453450
2	blue	258438
3	green	274713
4	grey	389970
5	pink	258190
6	purple	307100
7	red	290800
8	white	1037777
9	yellow	213902

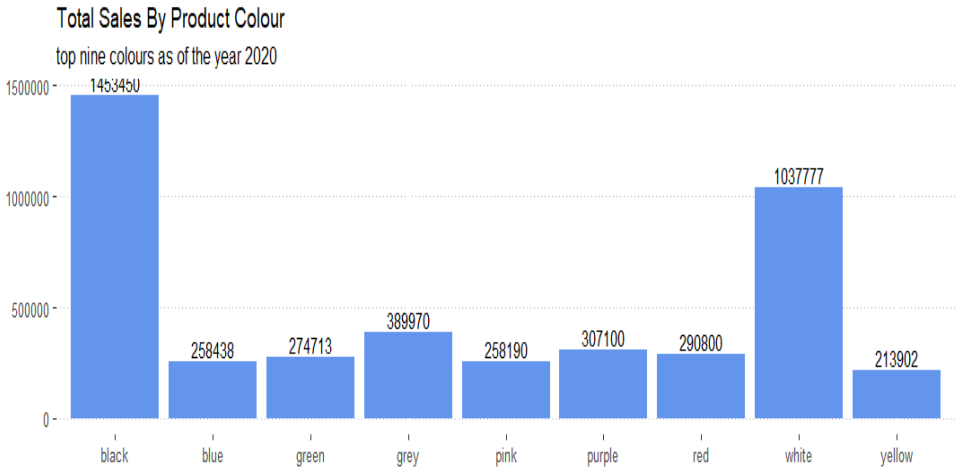


fig 2.2.6 Average Units Sold By Colour

fig 2.2.7 Total Units Sold By Colour



fig 2.2.8 Plot of Average Units Sold By Colour

The higher the rating count, the higher the rating of the product. fig 2.2.9 shows a there are more ratings between 3 and 5. It also reflects the influence of the rating counts. Hence investing in sentiment analysis and building customer relationship can help boost product sales on Wish marketplace.

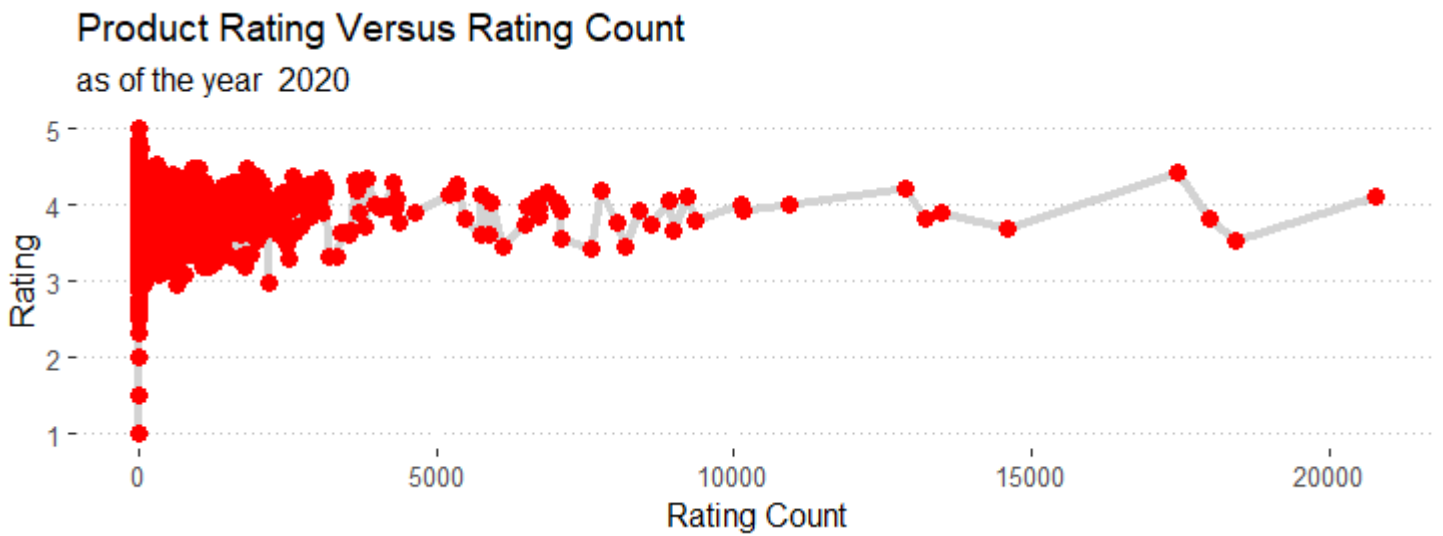


fig 2.2.9 Plot of Product Rating Versus Units Sold

Ideally, one would expect to see a higher sales when ads are used, however, that is not the case on Wish platform. fig 2.2.10 shows there is no substantial difference in the units of products sold with or without ads represented by 1 and 0 respectively. With AI, running the right ads has become easier and cheaper because business owners can direct ads to the right audience and leverage past data to make sales predictions.



fig 2.2.10 Plot of Average Units Sold By Ad Use

2.3. Statistical and Predictive Analysis

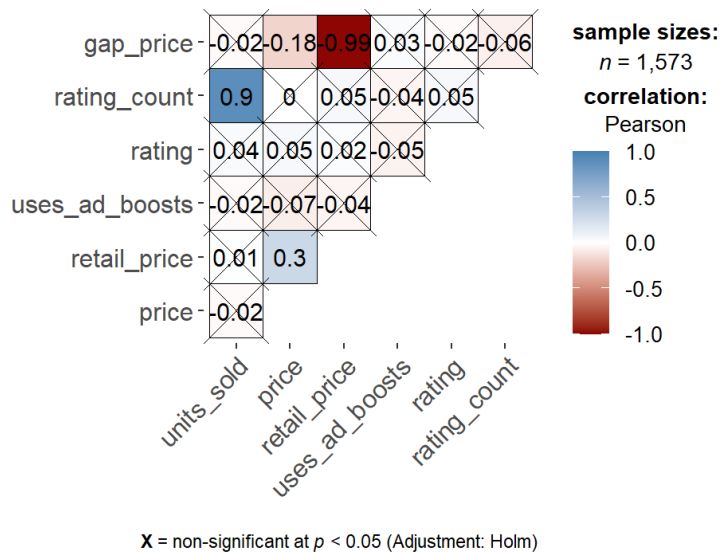


fig 2.3.1 Heatmap of Pearson Correlation Test

A Pearson Correlation test was performed to ascertain if the relationship between aforementioned variables and units sold are significantly significant. Table 2.3.1 reveals that rating count has the highest correlation of 0.90 and a $p_value < 2.2e-16$, hence it is statistically significant to the units sold.

Table 2.3.1 Table of Pearson Correlation Test Summary

Name	t-value	correlation	p_value
Price	-0.984	-0.025	0.325
Retail price	0.500	0.013	0.616
Rating	1.566	0.039	0.118
Uses ads boost	-0.636	-0.016	0.525
Rating count	81.582	0.90	2.2e-16
Gap price	-0.649	-0.016	0.517

Furthermore, 4 predictive models were built in order to examine which of the variables can predict units sold with a higher prediction correlation. Table 2.3.2 reveals the results of these models. Though all the

models have a low p_value, but Model 1 has a better accuracy and a further investigation can be performed on the variables used in building the model.

Table of Model Summary

Model	Adjusted R-Squared	Multiple R-Squared	p-value	Prediction Correlation
Model 1	0.362	0.364	< 2.2e-16	0.895
Model 2	0.374	0.416	< 2.2e-16	0.841
Model 3	0.416	0.493	< 2.2e-16	0.790

Recommendation

The advent of the COVID-19 pandemic has not only accelerated the consumers’ online purchase behaviour, it has also posed the problem of customer retention. Customers are willing to purchase goods online, however, they will rather do this on platforms with the most personalized experience, hence the need for Wish to harness sales data to take the right sales and marketing strategies, such as upsell, cross-sell, advertising campaigns, e.t.c (Simplify, 2021)

As shown in the analysis, rating count plays a significant role in sales prediction, therefore, Wish can invest in automated email marketing strategies that foster a relationship the users and also lead to more rating and rating counts. The relevance of Sentiment Analysis cannot be over-emphasized as doing this will enable Wish to know what its users are saying (Karthika et al., 2019)

Furthermore, Wish has not maximized the use of ads on the platform, hence no significant difference in sales when ads is used or not. AI techniques to boost sales through ads (Kaput, 2022)

Finally, different product features such as colour and size tend to have a trend with the units sold, however, a consistent analysis of the sales data will give better insights in to what is happening on the platform and an advanced predictive model and price optimization strategy can be built in order for the merchants to know when and what deals to offer on the platform.

CONCLUSION

Data is increasing tremendously and the traditional methods will not be sufficient to create insights or study the observed patterns. Hence, the need for the adoption of advanced methods to maximize the full potential of these. Reports has estimated the business investment into AI to reach \$191 billion by 2025 (MarketsandMarkets, 2019).Therefore, Wish can maximize AI techniques to improve sales and industry permanence.