**An Analysis of sellers of vintage musical instruments on Etsy**

The aim of this project is to provide buyers of vintage musical instruments, who are new to the Etsy website, a way to identify reliable sellers on the global marketplace. Connoisseurs of vintage items, especially those of musical instruments, are very particular about buying items. Purchasing something of such value from an internet marketplace can seem somewhat of a risk to the buyer, but given the right kind of inputs, they can be nudged towards the most appropriate sellers.

A seller’s popularity could be evaluated on the basis of the number of admirers (marked favorites) they have on Etsy. This can be analyzed on the basis of total sales made so far, total reviews, shop rating, number of years active with Etsy, favorable shop policies, shipping time and seller responsiveness.

My null hypothesis will be that count of customers who have marked the seller as a favorite depends on sales made so far, number of total reviews, shop rating, total years active on Etsy, shopping policies[[1]](#footnote-1) , shipping time , seller responsiveness and seller return score against my alternate hypothesis that it does not depends on these factors.

**Data Extraction**

Etsy does not provide seller data directly for analysis purposes. I had to visit the Etsy vintage musical instruments catalog and web scrape into each product and seller page to get required information.

**Data Cleansing**

Eliminated duplicate sellers (sellers with multiple products on display). Removed repetitive string values from count / total and year variables and extracted numerical values.

**Data Transformation**

Refer Data Appendix.

**Sample Characteristics**

The data consists of observations for 1,984 sellers. The average number of customers who have marked seller as a favorite is around 780. Average number of sales made by a seller is around 1115. In average, each seller has around 322 reviews. Average rating for a seller is around 4.49, with minimum rating being 0 and maximum being 5. On an average, sellers have been on Etsy for around 5 years, with some sellers having joined recently, and the oldest seller being 14 years on Etsy. Average shipping time for a seller is 6 days, with each seller taking a minimum of 1 day. Since some sellers did not provide exact shipping time, I assigned them a value of 99. This, assuming that these sellers have less clear shipping policies than other sellers who have given shipping estimates. A higher value will indicate higher shipping time. Average seller flexibility score is 2 points out of 7 with minimum being 0 and maximum being 7. I assigned a score of 0 for sellers who don’t offer returns, exchanges or cancellations, or have not provided any information. Average seller responsiveness is 1 point with 0 being the lowest and 2 being the highest. Average return score is 4 days, with minimum being 0 and maximum being 30 days. I assigned a score of 0 to sellers who have not provided return information. Graphical representations for some variables and predicted and observed values is provided in data appendix.

**Analysis**

I conducted an OLS regression at a significance level with alpha value 0.05 to find out the effect of total sales, total shop reviews, shop rating, years active with Etsy, seller shipping time, seller flexibility score, seller responsiveness and seller return score on the count of favorites for a seller. I took log of count of favorites, total sales and total shop reviews. All the coefficients of the independent variables are linear in parameters and the sample is random in nature as it is drawn from Etsy, there is no perfect collinearity among the explanatory variables. As can be seen from data appendix ,the expectation of residuals tends towards zero and the BP test which is check for heteroskedasticity is found to be 83 which is not a higher BP value but the lower p-value indicates that we can reject null hypothesis and accept alternate hypothesis that the model has heteroskedasticity. The plotting of residuals vs predicted values shows that the residuals are well within the standardized range of -3 to +3.

I re-specify my model into robust standard errors instead of OLS as there is evidence of heteroskedasticity. Weighted least squares cannot be conducted due to lack of ability to know the form of heteroskedasticity.

The coefficient of total sales, total shop reviews, years active with Etsy, seller shipping time and seller flexibility score came statistically significant at alpha value 0.05, that is these variables help to determine the popularity of a seller. The coefficient of total sales indicates that if there is 1 percent increase in total sales then the count of favorites will increase by 0.7 percent, holding everything else constant. The coefficient of total shop reviews indicates that if there is 1 percent increase in total reviews given to a seller then the count of favorites will increase by 0.09 percent, holding everything else constant. If a seller is in active with Etsy one more additional year then the count of favorites will increase by 15 units. If seller shipping time increased by 1 more additional day then count of favorites increased by 0.3 units, holding everything else constant. This is in contrast with the priori expectation which is that a seller will be more favorite if it takes less shipping time. This could be because handling and shipping musical instruments is more crucial and time consuming than other items. These could be stored in a different location than usual, and hence may take more shipping time than usual. Customers of these products would be aware of this fact and would expect shipping to take longer.

If the seller flexibility score is increased by 1 additional points then the count of favorites will increase by 5 units, holding everything else constant. This is in line with our expectation that as seller with favorable shopping policies will have more favorites compared to seller with less favorable shopping policies. Shop rating, seller responsiveness score and seller return score came as statistically insignificant. Though, the priori expectation was they will be positively related with count of favorites. Of these, seller responsiveness score is dependent on the seller response text on the website, which is easy to overlook and hence might not have been looked at by most users. Seller return score is also within the view shop policies navigation, and not directly visible. This could be the reason fewer customers might have explored it. The goodness of fit of the model suggest that the R-squared is 0.86 that means 86 percent of the variation in count of favorites is explained by total sales, total shop reviews, years active with Etsy, seller shipping time and seller flexibility score.

Hence the project is useful for those buyers of vintage musical instruments who would want to know the popularity of the seller based on the factors – total sales, total shop reviews , years active with Etsy, seller shipping time and seller flexibility score i.e purchase policies. This will also be useful to the Etsy admins to find out most popular sellers and also to provide feedback and incentives to seller and thus boost sales further.

**DATA APPENDIX**

**Data Sources**

**Starting URL:**

<﻿https://www.etsy.com/c/vintage/books-movies-and-music/music/musical-instruments?explicit=1>

**Datasets**

Dataset are created through web scraping through Etsy’s vintage musical instruments catalog and visiting each item and seller page.

**Variables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Transformed Variable** | **Description** | **Unit** |
| Count\_of\_favs | Log of Count\_of\_favs | **Dependent Variable.** Count of users who have marked seller as a favorite | Decimal |
| total\_sales | Log of total\_sales | **Independent Variable.** Total sales by seller | Decimal |
| total\_shop\_reviews | Log of total\_shop\_reviews | **Independent Variable.** Total shop reviews | Decimal |
| shop\_rating | NA | **Independent Variable.** Seller Rating | Decimal between 0 to 5 |
| years\_active\_with\_Etsy | Current Year –  Year of joining Etsy | **Independent Variable.** Number of years since joined Etsy | Integer |
| seller\_shipping\_time | Converting shipping time to days.  Weeks to be multiplied by 7  NA values converted to 99. | **Independent Variable.** Maximum days seller has committed to ship product in. 99 indicates seller did not provide any shipping details. | Integer between 0 to 99 |
| seller\_flexibility\_score | **Shop Policies Points (Out of 7) :-**  Accept Returns : 4  Accept Cancellations : 2  Accept Exchanges : 1  Don't Accept : 0  NA. : 0 | **Independent Variable.** Score created to gauge effectiveness of seller’s shipping policies based on offerings. | Integer between 0 to 7 |
| seller\_responsivness\_score | **Seller Responsiveness Score**  Seller Responds in few hours : 2  Seller Responds within 24 hours : 1  NA : 0 | **Independent Variable.** Score created to gauge effectiveness of seller’s response for a query. | Integer between 0 to 2 |
| seller\_return\_score | **Seller Return Score**  Days of Delivery  NAs are treated as 0 | **Independent Variable.** Number of days seller allows to return product. 0 indicates seller did not provide a return policy. | Integer |

**Descriptive Statistics**

**A screenshot of a cell phone

Description automatically generated**

**Regression Results**

**A close up of text on a white background

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

**Plots**

**Figure 1: Predicted vs residual and observed values**

**A close up of a map

Description automatically generated**

**Figure 2: Total sales vs count of favorites**

**A screenshot of a social media post

Description automatically generated**

**Figure 3: Total reviews vs count of favorites**

**A screenshot of a social media post

Description automatically generated**

**Figure 4: Years on Etsy vs count of favorites**

**A picture containing white

Description automatically generated**

**Figure 5 : Frequency Distribution of seller flexibility score**

**A picture containing screenshot

Description automatically generated**

1. I assigned a score out of 7 for each seller based on their shopping policy, i.e. whether they allow returns, cancellations or exchanges. Refer Data Appendix for Data Transformation details A score of 4 for returns, 2 for cancellations and 1 for exchanges will be assigned for each seller. Scores will be added for sellers providing at least two or more options, depending on options provided. [↑](#footnote-ref-1)