Wheaton College
Department of
Mathematics and
Computer Science

Investigating Top Merchants and Their Characteristics

Grace Flitsch

Background

The data for this project comes from AwanTunai, a company whose "vision is to bring prosperity to Indonesia's unbanked micro SMEs which create 60% of the workforce."[2] They work toward this goal by providing affordable loans to suppliers and merchants so they can invest in more merchandise and grow their business without being hindered by financial roadblocks.

Problem Statement

Since AwanTunai's mission is to help local businesses in Indonesia, the goal of this project is to provide to AwanTunai insights about their top merchants, or the store owners that are most valuable to invest in. We aim to derive insights that will enable AwanTunai to better predict which merchants to invest in in the future, and identify their top customers to encourage repetition of their behavior. In order to track behaviors and traits of top merchants, we first must define the term "top merchant." After that, we will note identifying characteristics that set top merchants apart from other merchants, so that we can incentivise those activities to current and future customers.

A typical Indonesian Corner Store[3]

A supplier warehouse[3]

Relevant data attributes

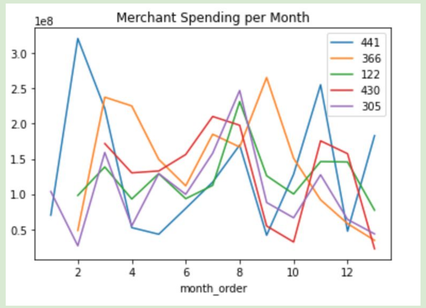
spent	A merchant's total spending from 8/1/2021- 8/15/2022
avgspent	A merchant's average spending per month since first purchase
order_id	The total number of unique order ids per merchant, per year
month_order	The number of months since 8/1/2021
merchant_id	A randomly assigned id given to each unique merchant
loyalty	The number of months since a merchant's first purchase

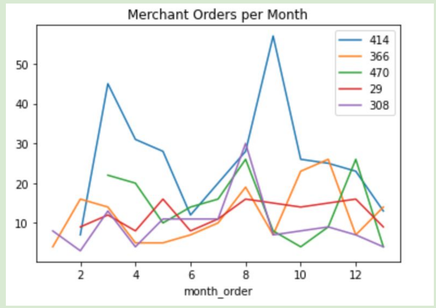
Methods

To answer our research question of how to define a "top merchant," we applied the industry standard RFM model to recommend the top five merchants.[1] This model can introduce bias since some intuitive decisions had to be made when it came to eliminating outliers. The process in inspecting outliers involved a general rule that required the merchant to have made at least one purchase per month since their first purchase. Next, we used a clustering model for a secondary test of top merchants. This model compared each column in a merchants dataframe that we created based on the original data to each other. Finally, we compared the two results and provided a definition and a few characteristics that we observed from top merchants.

RFM Model

Using the RFM (recency, frequency, monetary) model, we can rank the top five merchants over all time.[1] To provide the monetary ranking, we do an all-time and six month comparison to ensure fair treatment to newer customers. We disregard recency as a statistic given the removal of inconsistent customers as outliers. The top five according to this model are merchants 366, 414, 441, 470, 122.

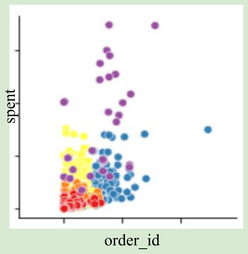


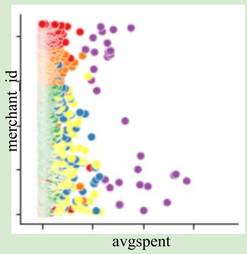


Using a model that removes the factor of "loyalty," or the number of months that a merchant has been making purchases, we re-rank the merchants. This is to make sure that inconsistencies in time since the first purchase don't affect the issue of whether a merchant is a high spender within their time as a customer. From this model, the top five merchants according to the "monetary" statistic are 366, 441, 430, 122, 305. Three of the five merchants in this group are also top five according to the original RFM model. We took these three as our overall "top merchants" and noted their tendency to be consistent spenders, not just one time high purchasers. *Note that data in month 13 only accounts for the first 15 days of the month, hence the drop off.

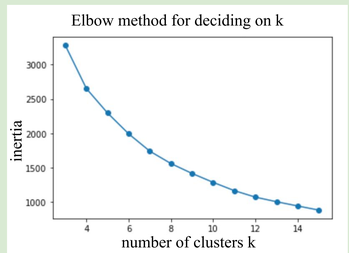
Clustering

The RFM model is an industry standard, but may include some bias because of the removal of outliers by intuition. By creating a cluster plot between each of the columns in the dataframe that was grouped by merchant_id it was found that the top 3 merchants identified by the RFM model were in the same cluster of every plot. This cluster (#3) clearly was the highest performing group, so data-driven model confirmed the RFM findings. This shows that RFM is a reliable model for AwanTunai to use to identify top merchants in the future. (Cluster 3 = purple)





Not only were the top 3 selected from the combination of the RFM and time-independent tests in cluster 3, but all of the merchants found in the top 5 from each model independently were also in the same cluster. The clustering relied on a find_elbow method that produced a plot that shows the efficiency of each option for total number of clusters. (Right) The choice of number of clusters is significant since each elbow option presents varied cluster sizes and members.



Conclusion

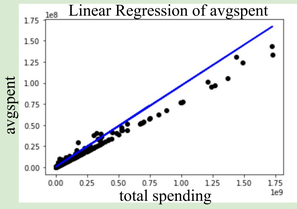
AwanTunai's goal as a company is to provide means for small merchants and suppliers to build value in their businesses. An important part of this goal is identifying which merchants are of high value and should be given larger lines of credit. To find these top value customers, or "top merchants," we used several models and discovered which merchants were significant customers regardless of the identifying statistic. From these models, we can conclude that merchants 336, 441, and 122 are consistently in the top 5 group of merchants. Looking at their characteristics, we gathered that merchants that will be valuable to this business tend to have high loyalty, spend consistently each month, and have made purchases recently. These top merchants also tend to be making larger orders rather than many small orders, since they did not end up in the top 5 when we investigated the number of total orders. In the future, AwanTunai can use both of these models to identify their current top merchants and incentivise them to continue being steady, high spending customers.

Further questions for research:

- Can these same methods be used to identify low-performing merchants in order to avoid high-risk loans?
- Can we predict future behavior of these top merchants to anticipate loan-repayment rate?
- Which factor in RFM has the most correlation with the outcome of the clustering model?

Caveat

As a small aside, we made a linear regression model that predicts the average monthly spending based on a merchant's total spending. This could be used by the business to predict spending habits over a long period of time for customers who have not been around as long as a typical top merchant. Below is the plot of the regression.



References

1.Serrano, S. (2022, July 11). *RFM analysis w/ segmentation examples*. Barilliance. Retrieved September 20, 2022.
2. Indonesia's Best Supply Chain Financing Services. AwanTunai. (n.d.). Retrieved September 20, 2022.
3. Inventory Purchase Recommendation for Merchants in Traditional FMCG Retail Business, Y. Li, D. M. Robani, V. Suciu, J. He-Yueya, The 9th IEEE International Conference on Data Science and Advanced Analytics
4. The pandas development team. (2020, Feb). *pandas-dev/pandas: Pandas.*

Acknowledgements

Thank you to Dr. Ying Li for facilitating and advising me in my research and presentation. I would also like to extend my gratitude to Wheaton College for enabling me to take this course with Dr. Li.