FINAL PROJECT REPORT: MIAMI DADE COUNTY PURCHASE ORDERS ASSESSMENT FOR SUPPLIER NEGOTIATION

Objective

Analyze the 2024 Purchase Order dataset from Miami Dade County to potentially negotiate with suppliers for better pricing opportunities. We will start with RFM methodologies in conjunction with K Means clustering to understand Miami Dade County's recency, frequency, and monetary value before moving onto assessing Monte Carlo simulations to further strengthen negotiation strategy.

Dataset

This dataset contains all purchases made by Miami Dade County in 2024. Included in this dataset is purchase order information such as purchase order unique identifier numbers, the amount of the purchase orders, the date of the purchase order, etc. This dataset also contains information such as item descriptions and supplier information which are vital for this analysis. There is a total of 43,058 rows and 31 columns. This dataset was taken from the Miami Dade County open-source data hub. Link: https://gis-mdc.opendata.arcgis.com/datasets/673f9be756ce4d2a95039bf16d3a3f7d_0/explore

Structural Investigation Summary

We started with 43,058 rows and 31 columns in the dataset. Seeing that we are only performing RFM, MBA analysis, and Monte Carlo simulations, we only need 5 columns: 'SUPPLIER_NAME', 'PO_DATE', 'PO_NUMBER', 'PO_AMOUNT', 'ITEM_DESCRIPTION', so the rest of the columns were dropped. 'SUPPLIER_NAME' was chosen as the unique supplier identifier because it was the only supplier identifier that didn't have any null values. There are 1,383 different suppliers and 31,069 different items.

Data Preprocessing

To preprocess the data, null values were dropped and converted the PO_DATE attribute into a datetime data type. Converted transaction dates in 'Recency', computed the number of purchases into 'Frequency', and computed the total spend per customer into 'Monetary.' To further maintain the integrity of the structure of the dataset, the 'ITEM_DESCRIPTION'

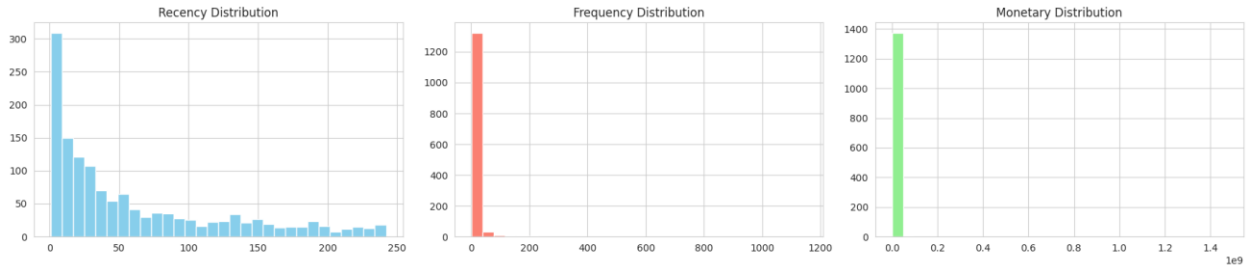
column was standardized and formatted to ensure accurate item results during further analysis.

RFM Analysis Summary

According to the decreasing trend in the Recency Distribution Chart, most purchases have a low recency value, meaning Miami Dade County has purchased a good number of items recently. The long tail on the chart suggests that there are still some instances where purchases were not made recently. The low recency value tells us that we might have some potential leverage in terms of recency when negotiating with suppliers.

According to the Frequency Distribution Chart, a huge number of purchase orders are going to the same suppliers. This means potential leverage in terms of frequency when negotiating with suppliers.

According to the Monetary Distribution Chart, it seems as if most suppliers are getting paid around the same amount of money. This means there might be some opportunity here to leverage the amount of money being spent when negotiating with suppliers.



Insights after reviewing the data from the top 5 suppliers for each respective metric category (recency, frequency, and monetary):

-There are a total of 51 companies with a recency value of 1, including the top 5 frequency suppliers. This means Miami Dade County has purchased from 51 companies within a days' time. There is not a lot to go on regarding recency alone.

-The top 5 monetary suppliers have very low frequency values, meaning Miami Dade does not purchase from them on a regular basis. There isn't a lot to leverage from these particular suppliers with the data we currently have.

-The top 5 frequency suppliers have a great amount of recency, frequency value, and monetary value. For the remainder of this project, we are going to analyze just the top frequency supplier, Genuine Parts Co, to see how we can make a compelling argument as to why Miami Dade County deserves better pricing.

K Means Clustering (Genuine Parts Co) Summary

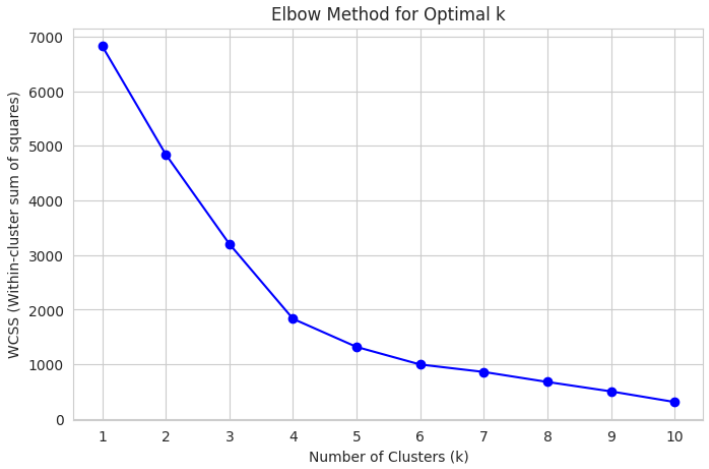
The dataset was filtered so that we can re-perform RFM solely on the data from Genuine Parts Co. The newly created data frame has 3,061 rows and 5 columns. We used the elbow method to find the optimum number of clusters (4 clusters) and proceeded to apply K Means clustering to properly assign items to groups. These are the results of the cluster:

-Cluster 0: 2,123 items, Recency of 32.2, Frequency of 1.3, and Monetary of 269.8. This tells us that most items in this cluster are fairly cheap and tend to be purchased once a month (more or less). This cluster also has the greatest number of items by far.

-Cluster 1: 4 items, Recency of 9.0, Frequency of 3.8, and Monetary of 195,317.3. This tells us that these 4 items are abnormally large purchases that aren't purchased too often.

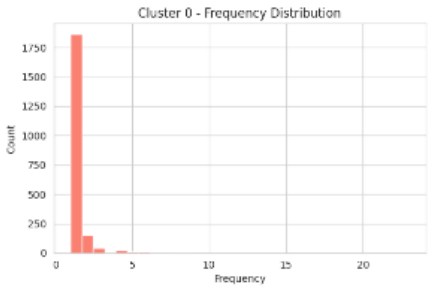
-Cluster 2: 148 items, Recency of 186.1, Frequency of 1.0, and Monetary of 3,374.3. This tells us that these items are items that aren't purchased often as the recency is the highest out of all clusters, the frequency is the lowest, and it has the 2nd least monetary value.

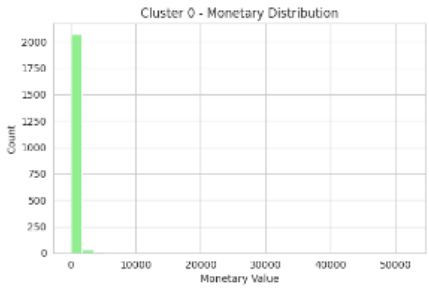
-Cluster 3: 1 item, Recency of 1.0, Frequency of 66.0, and Monetary of 6,911.7. This tells us that this 1 item gets purchased a lot as it has the highest frequency and the lowest recency.

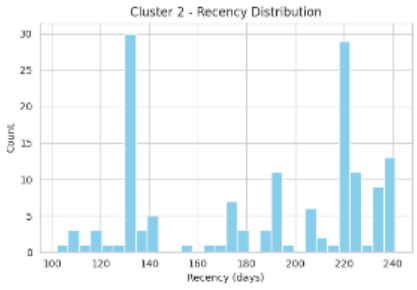


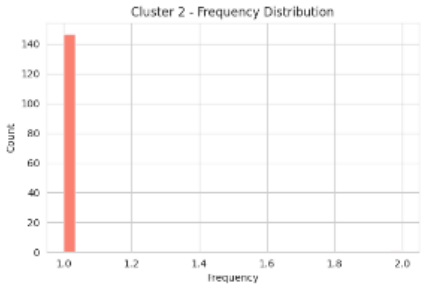

Customer Segments based on RFM
Monetary Value
Recency (days)
Cluster

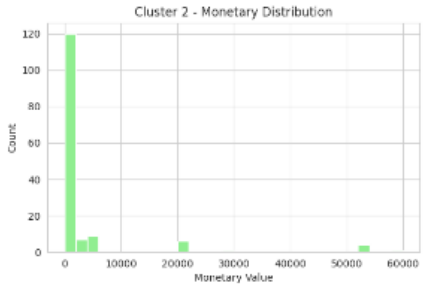

Cluster 0 - Recency Distribution
Count
Recency (days)


Cluster 0 - Frequency Distribution
Count
Frequency

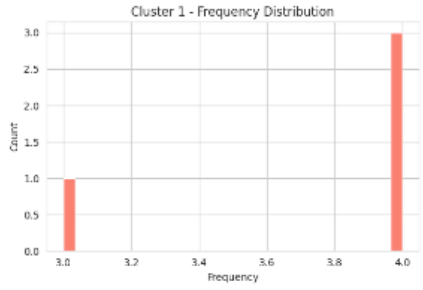

Cluster 0 - Monetary Distribution
Count
Monetary Value

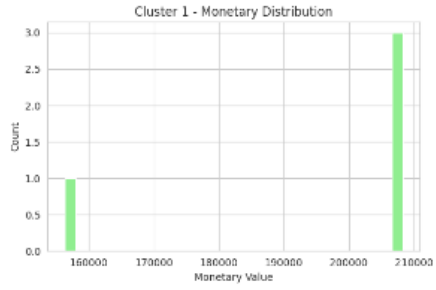

Cluster 2 - Recency Distribution
Count
Recency (days)

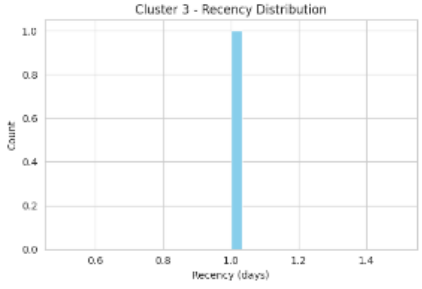

Cluster 2 - Frequency Distribution
Count
Frequency

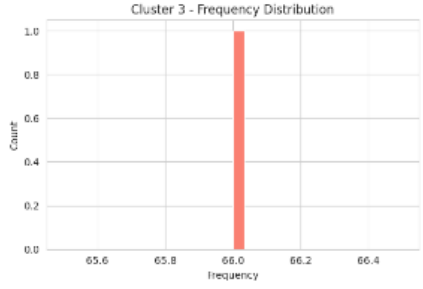

Cluster 2 - Monetary Distribution
Count
Monetary Value

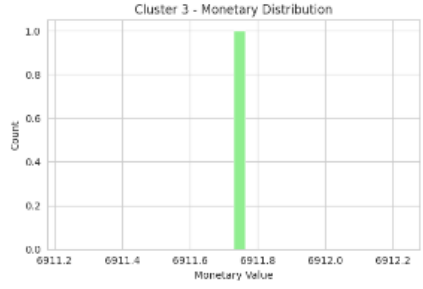

Cluster 1 - Recency Distribution
Count
Recency (days)


Cluster 1 - Frequency Distribution
Count
Frequency


Cluster 1 - Monetary Distribution
Count
Monetary Value


Cluster 3 - Recency Distribution
Count
Recency (days)


Cluster 3 - Frequency Distribution
Count
Frequency


Cluster 3 - Monetary Distribution
Count
Monetary Value

After reviewing all the data regarding clustering, it is safe to assume that Clusters 0 and 3 are clear candidates for bargaining due to their frequency and recency values. Whether or not the other clusters are also qualified for bargaining is at the discretion of an expert in the area. Cluster 3's sole item is a core that gets purchased 66 times a year with a high recency value and a monetary value of $6,911.70. Cluster 0 has 2,123 items with some of those items having low recency and high frequency values.

Monte Carlo Analysis Summary

This Monte Carlo analysis is meant to predict what spending would be next year to show suppliers the potential amount of money Miami Dade would be spending if Miami Dade were to continue business operations with them and what they would be losing if there is not a deal to be made.

Miami Dade County spent $876,747.18 at Genuine Parts Co in 2024. Using this amount as the base value, we ran 10,000 Monte Carlo simulations over the course of a business year (260 days). The expected value came out to $875,258.86 and a 95% confidence interval of $872,690.69 - $877,003.65. Yes, deals and discounts will lower this amount, but Genuine Parts Co still has a large amount of money to make, and a bigger amount to lose if Miami Dade were to go with another company.

Project Summary

According to the data, Miami Dade County spent $4,288,751,653.82 in 2024 on 43,058 transactions (31,069 distinctly different items) with 1,383 different suppliers. Using RFM we were able to gain insights as to where the money is going, how often, and how recent. Miami Dade County can use this data to possibly negotiate with suppliers to get better pricing options, given the vast spending capabilities of the county. We used the supplier Genuine Parts Co as an example of how you can use K Means Clustering to group items for strategical positioning when negotiating. Furthermore, we used a Monte Carlo analysis to show the potential spending capabilities of Miami Dade County for next year to emphasize the value Miami Dade County has to offer their business.

After assessing Genuine Parts Co's relationship with Miami Dade County, we were able to see that MDC (Miami Dade County) spent a total of $876,747.18 at GPC (Genuine Parts Co). Based on the data, we were able to estimate that MDC will spend $875,258.86 in total for 2025, with the least possible amount being $872,690.69 and the highest possible

amount being $877,003.65. Using RFM in conjunction with K Means Clustering we were able to see two out of four clusters with potential for negotiations based on hypothetical criteria (essentially, the higher the frequency and recency, the more leverage MDC has in purchasing power with frequently purchased items). There is a possibility that all four clusters have potential for negotiations, but that would have to be determined by an expert on the matter.

Cluster 0 has a monetary value of $572,734.08 and Cluster 3 has a monetary value of $6,911.73. Developing criteria for negotiations without being an expert in the field presents its challenges, but in this scenario, we are going to go with hypothetical thresholds and present recommendations to show an example of how this data can be used to possibly negotiate with GNC:

-MDC can ask for some type of overall discount for the expected $875,258.86 that MDC will be spending with GNC in 2025.

-MDC can request discounts be applied in brackets. For instance, when crossing the $100k threshold, there will be a 10% discount applied, when crossing the $300k threshold a 5% discount will be applied, etc.

- Cluster 0 has a monetary value of $572,734.08. Maybe an overall discount can be applied to all items in this cluster as it has a significant amount of monetary value?

-Cluster 3 has a monetary value of $6,911.73 for 1 item, a core, that was purchased 66 times in 2024. This one item could possibly be an item that receives a discount as it is purchased pretty frequently.

-Bargain with items within Cluster 0 that have a high frequency/recency combination. For instance, if the frequency of an item is 12 or more (indicating this item has been purchased 12 times within the year) and has a recency threshold of at least 30 or less (indicating it has been at least 30 days since the last purchase), we can assume that these items are being purchased monthly at the very least. With this information, MDC can request monthly discounts on these items. In this example, we were able to see 4 items that met that criterion: air filters ($9,552.18), oil filters ($3,770.41), fuel filters ($5,271.88), and 7565 ($7,231.12) for a total of $25,825.59.

In conclusion, this type of overall analysis can be run on any supplier in the dataset and can be used as a bargaining tool to negotiate better pricing. Getting better pricing can result in MDC saving potentially hundreds of thousands of dollars, maybe more depending on negotiations, that can be allocated for other county projects. This information can possibly also leverage negotiations with competing suppliers to attain better pricing if MDC's current suppliers do not wish to negotiate.

Further Investigation

More granularity can be attained with completed datasets from previous years so more accurate RFM and Monte Carlo predictions can be made. Price sheets from competitors would also be of value when negotiating. Categorical data to further strengthen negotiation strategy. Colaborating with experts in each supplier field or some type of comprehensive guide on supplier business relations for each sector would also be useful.