

# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024

## Project Objective and Motivation

For my final project, I aimed to analyze customer segmentation using a dataset containing simulated customer information. The objective was to identify distinct customer segments based on demographic and behavioral data to provide insights into targeted marketing strategies, improving customer satisfaction, and boosting sales. Initially, I planned to analyze Starbucks' customer sentiment from Twitter, but due to the limitations of the free version of the Twitter API and lack of recent data, I shifted my focus to a more feasible dataset.

## Data and Methodology

The dataset I used includes various customer attributes such as age, gender, income, spending score, membership years, purchase frequency, preferred category, and last purchase amount. Here's a breakdown of the steps and methodologies used:

### Data Collection and Preparation

I loaded the dataset into Google Colab, cleaned, and preprocessed it to ensure it was ready for analysis. The dataset was imported into the environment, and necessary preprocessing steps like handling missing values, encoding categorical variables, and scaling numerical features were performed.

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans, DBSCAN, AgglomerativeClustering
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

[23] # Load the dataset
file_path = '/content/customer_segmentation_data.csv'
df = pd.read_csv(file_path)

# Display the first few rows of the DataFrame
print(df.head())
```

	id	age	gender	income	spending_score	membership_years	
0	1	38	Female	99342	90	3	
1	2	21	Female	78852	60	2	
2	3	60	Female	126573	38	2	
3	4	40	Other	47099	74	9	
4	5	65	Female	140621	21	3	

	purchase_frequency	preferred_category	last_purchase_amount
0	24	Groceries	113.53
1	42	Sports	41.93
2	26	Clothing	424.36
3	5	Home & Garden	991.93
4	25	Electronics	347.08

```
[24] # Check for missing values
print(df.isnull().sum())

# Handle missing values if any
df = df.dropna() # Or use other methods to handle missing data

# Convert categorical variables to numeric if needed
df['gender'] = df['gender'].astype('category').cat.codes
df['preferred_category'] = df['preferred_category'].astype('category').cat.codes

print(df.head())
```

# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024

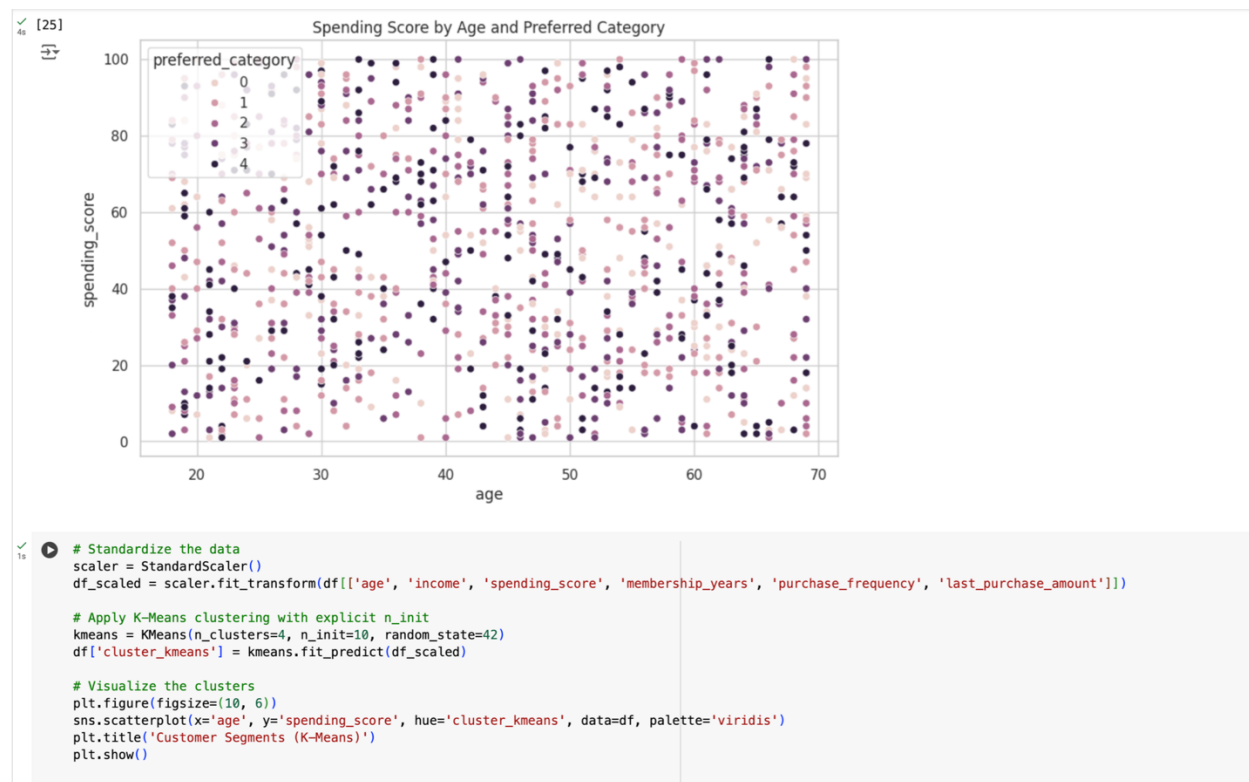
## Clustering Techniques

K-Means Clustering: Grouped customers into four clusters based on their features.

Used K-means clustering to segment customers into distinct groups based on their demographic and behavioral characteristics. This technique optimizes the within-cluster sum of squares.

Agglomerative Clustering and DBSCAN: Applied additional clustering techniques for comparison. Used agglomerative clustering (a hierarchical method) and DBSCAN (a density-based method) to validate the K-means results and ensure robust segmentation.

Principal Component Analysis (PCA): Reduced the number of dimensions in the data to make it easier to visualize. Performed PCA to reduce the dimensionality of the dataset, enabling visualization of the data in a 2D space while retaining most of the variance.



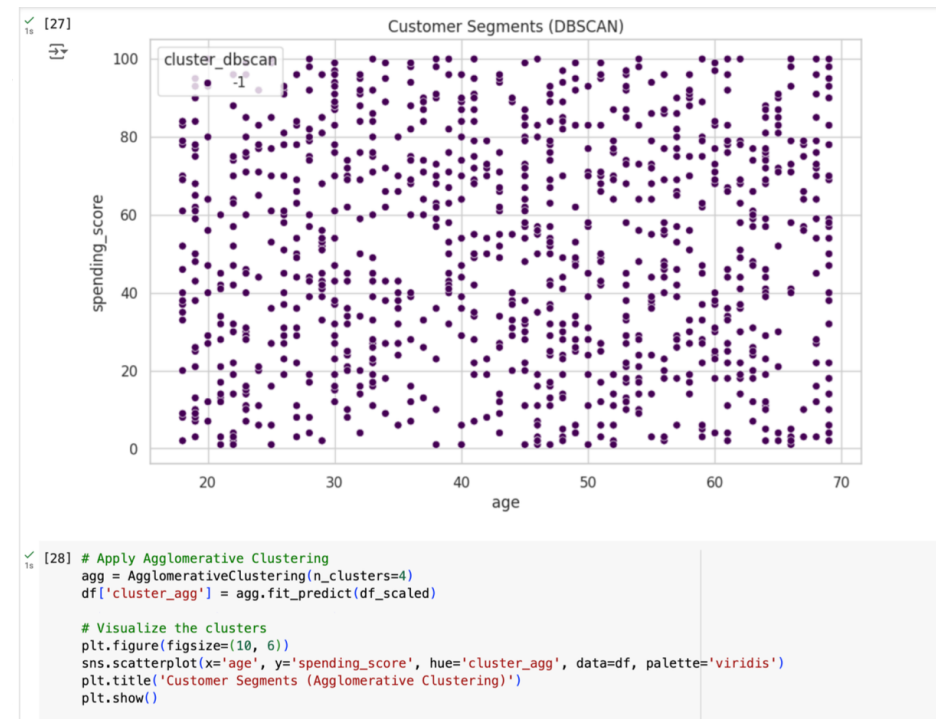
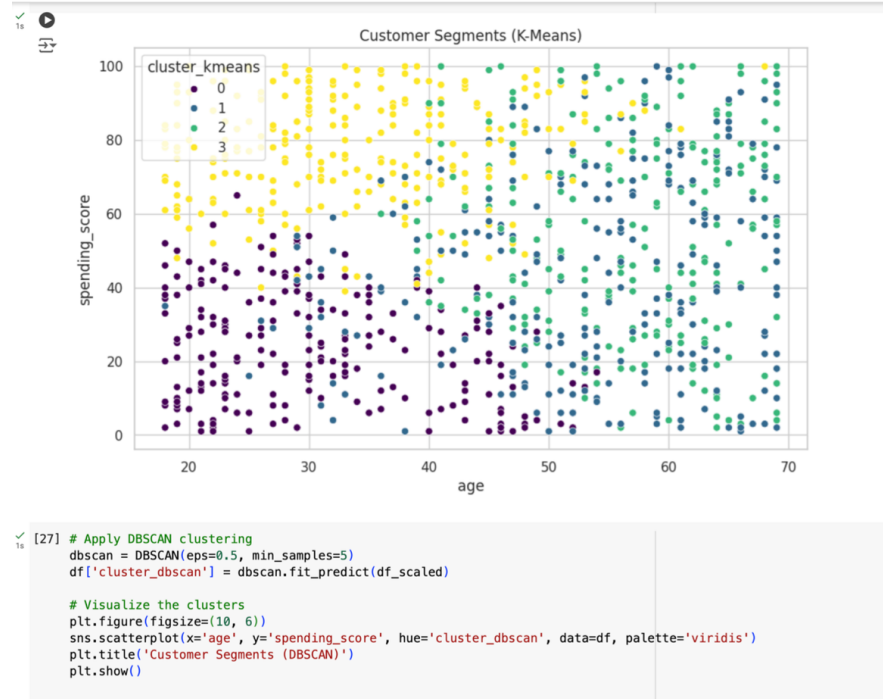
# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)

Professor Naina Chaturvedi

01:198:210 Data Management for Data Science

16 July 2024



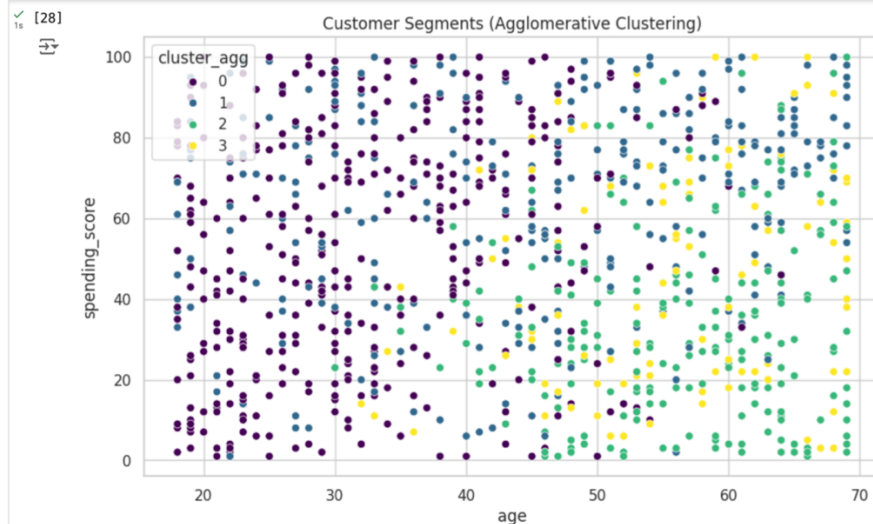
# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)

Professor Naina Chaturvedi

01:198:210 Data Management for Data Science

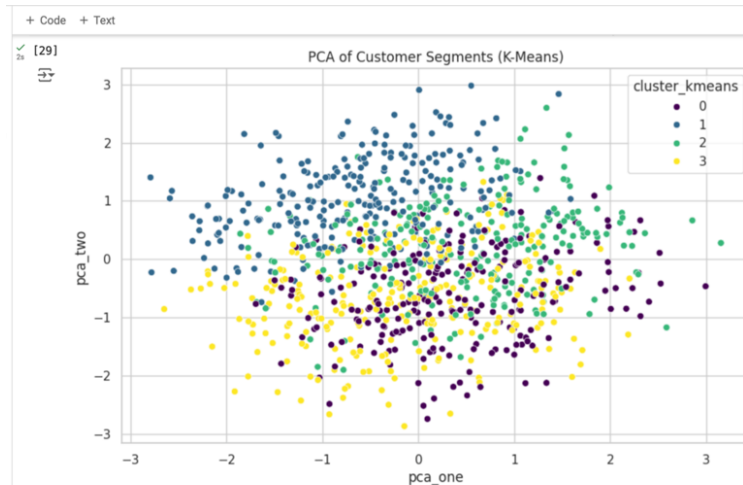
16 July 2024



✓ [29]

```
# Apply PCA
pca = PCA(n_components=2)
df_pca = pca.fit_transform(df_scaled)
df['pca_one'] = df_pca[:, 0]
df['pca_two'] = df_pca[:, 1]

# Visualize PCA components
plt.figure(figsize=(10, 6))
sns.scatterplot(x='pca_one', y='pca_two', hue='cluster_kmeans', data=df, palette='viridis')
plt.title('PCA of Customer Segments (K-Means)')
plt.show()
```



✓ [30]

```
# Analyze clusters
cluster_summary = df.groupby('cluster_kmeans').mean()
print(cluster_summary)

# Recommendations based on clusters
for cluster in cluster_summary.index:
    print(f"Cluster {cluster}:")
    print(f"Average Age: {cluster_summary.loc[cluster, 'age']}")
    print(f"Average Income: {cluster_summary.loc[cluster, 'income']}")
    print(f"Average Spending Score: {cluster_summary.loc[cluster, 'spending_score']}")
    print(f"Recommended Strategy: {'Provide discounts and offers' if cluster_summary.loc[cluster, 'spending_score'] < 50 else 'Offer premium products and services'}")
```

✓ RA Google Account  
Di KYLE MOLNAR  
kmm704@scarletmail.rutgers.edu

# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024

## Analysis and Insights

### Clustering Results:

Cluster 0: Profile: Middle-aged customers with moderate income and average spending scores.

Preferred Category: Groceries.

Recommendation: Offer premium products and services to cater to their spending potential.

Implement loyalty programs with cumulative benefits due to their high purchase frequency and longer membership years.

Cluster 1: Profile: Older customers with lower income and lower spending scores.

Preferred Category: Clothing.

Recommendation: Provide discounts and offers to encourage more spending. Tailor marketing campaigns to highlight new clothing lines and seasonal sales.

Cluster 2: Profile: Middle-aged customers with high income and high spending scores.

Preferred Category: Electronics.

Recommendation: Offer premium products and services to match their spending habits. Focus on new electronics product launches or exclusive deals to maintain their high spending and engagement.

Cluster 3: Profile: Younger customers with moderate income and average spending scores.

Preferred Category: Sports.

Recommendation: Provide discounts and offers to increase loyalty and spending. Highlight sports equipment and apparel in marketing campaigns.

# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024

+ Code + Text	
✓ [30]	
os	
cluster_kmeans	id age gender income \
0	501.321739 30.669565 1.013043 98862.330435
1	518.334630 53.264591 1.116732 55591.821012
2	504.741935 56.741935 0.979839 121045.322581
3	478.520755 33.841509 0.935849 80966.513208
cluster_kmeans	spending_score membership_years purchase_frequency \
0	23.839130 5.404348 28.386957
1	43.319066 5.661479 25.762646
2	54.358871 5.112903 25.455645
3	77.690566 5.671698 26.916981
cluster_kmeans	preferred_category last_purchase_amount cluster_dbscan \
0	2.130435 442.921174 -1.0
1	2.038911 665.218405 -1.0
2	2.028226 473.450685 -1.0
3	2.090566 385.282642 -1.0
cluster_kmeans	cluster_agg pca_one pca_two
0	0.556522 0.348996 -0.627431
1	1.762646 -0.613447 1.033744
2	1.483871 0.483641 0.231860
3	0.384906 -0.160590 -0.674960
Cluster 0:	
Average Age: 30.669565217391305	
Average Income: 98862.3304347826	
Average Spending Score: 23.839130434782607	
Recommended Strategy: Provide discounts and offers	
Cluster 1:	
Average Age: 53.264591439688715	
Average Income: 55591.82101167315	
Average Spending Score: 43.31906614785992	
Recommended Strategy: Provide discounts and offers	
Cluster 2:	
Average Age: 56.74193548387097	
Average Income: 121045.32258064517	
Average Spending Score: 54.358870967741936	
Recommended Strategy: Offer premium products and services	
Cluster 3:	
Average Age: 33.841509433962266	
Average Income: 80966.51320754716	
Average Spending Score: 77.69056603773585	
Recommended Strategy: Offer premium products and services	

## Additional Insights

Customer Age Distribution: Helps understand the age range of the customer base.

Age distribution analysis provides critical insights into age-targeted marketing strategies.

Income by Gender: Shows how income differs between genders.

Analysis of income by gender helps in designing personalized marketing efforts based on income levels.

Spending Score by Age and Preferred Category: Identifies spending patterns based on age and preferences.

Insights into how different age groups and their preferences affect spending behavior aid in more efficient marketing.

# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)

Professor Naina Chaturvedi

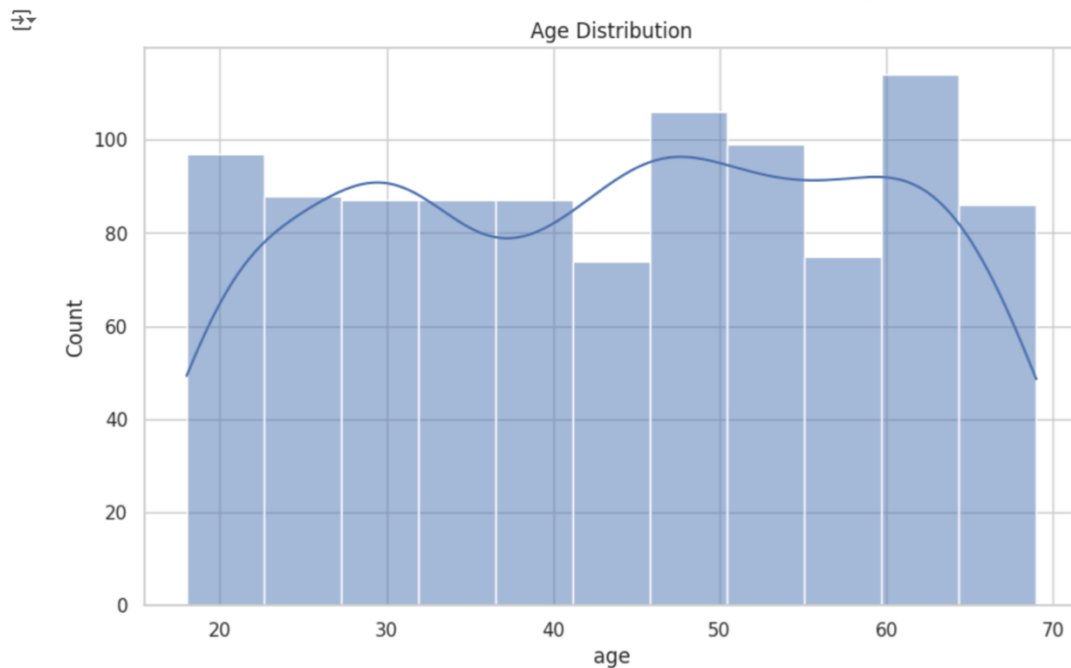
01:198:210 Data Management for Data Science

16 July 2024

```
[25] # Age distribution
plt.figure(figsize=(10, 6))
sns.histplot(df['age'], kde=True)
plt.title('Age Distribution')
plt.show()

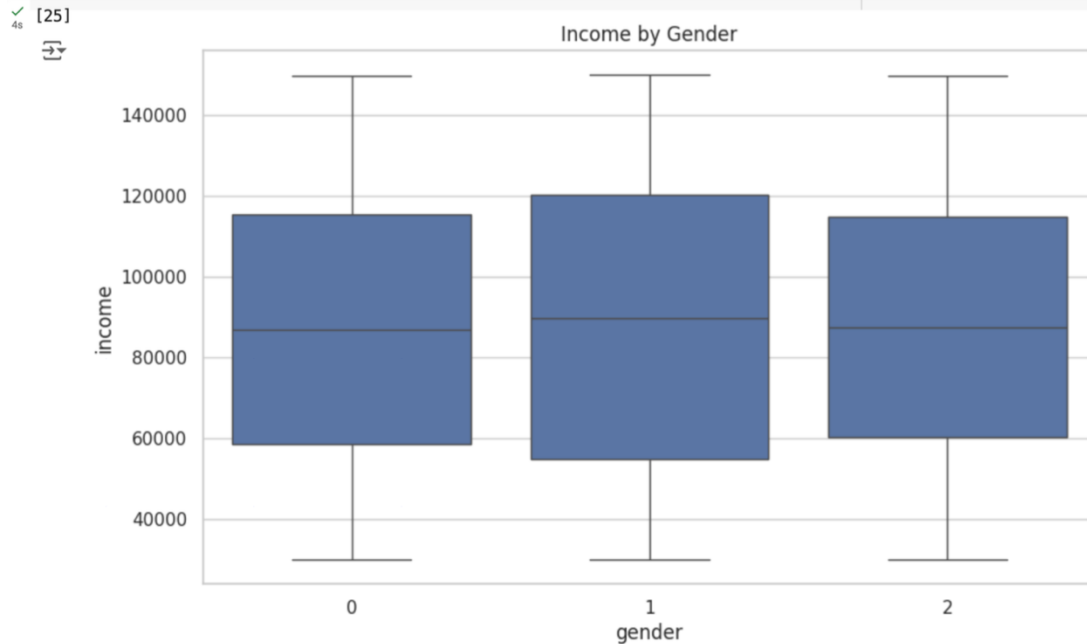
# Income by gender
plt.figure(figsize=(10, 6))
sns.boxplot(x='gender', y='income', data=df)
plt.title('Income by Gender')
plt.show()

# Spending score by age and preferred category
plt.figure(figsize=(10, 6))
sns.scatterplot(x='age', y='spending_score', hue='preferred_category', data=df)
plt.title('Spending Score by Age and Preferred Category')
plt.show()
```



# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024



## Importance and Problem Solving

**Targeted Marketing Strategies:** By identifying distinct customer segments, businesses can tailor their marketing campaigns to specific groups, leading to more effective promotions and increased sales.

**Improving Customer Satisfaction:** Understanding customer behavior and preferences allows businesses to provide personalized experiences, enhancing customer loyalty and satisfaction.

**Resource Allocation:** Businesses can allocate resources more efficiently by focusing on high-value customer segments and providing targeted offers.

## Conclusion

This project demonstrated how customer segmentation analysis could provide valuable insights into customer behavior, which is essential for developing targeted marketing strategies. By understanding different customer segments, businesses can tailor their products, services, and marketing efforts to meet specific needs, ultimately enhancing customer satisfaction and driving sales. The detailed methodologies, analyses, and insights derived from this project highlight the



# Customer Segmentation Analysis for Targeted Marketing Strategies: A Comprehensive Approach Using Clustering Techniques

Kyle Molnar / Net ID: kmm704 (solo project)  
Professor Naina Chaturvedi  
01:198:210 Data Management for Data Science  
16 July 2024

importance of data-driven decision-making in enhancing business strategies and achieving better outcomes.

## References

McKinney, W. (2022). Python for Data Analysis (3rd ed.)

Course materials from data management class.

Blog/resources from website: TechnologyAdvice on Customer Segmentation

Blog/resources from website: DiGGrowth on Customer Segmentation Analysis