A PROJECT REPORT

on

"CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

BY

SANGRAM KESHARI OJHA	21051760
ROUNAK BAIDYA	21051757
LISA NAYAK	22057082
PULKIT BHARWAJ	21051322
REHAN QUADARY	21051755

UNDER THE GUIDANCE OF

Prof. SOURAV KUMAR GIRI

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April-2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING"

submitted by

SANGRAM KESHARI OJHA	**21051760**
ROUNAK BAIDYA	**21051757**
LISA NAYAK	**22057082**
PULKIT BHARWAJ	**21051322**
REHAN QUADARY	**21051755**

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2024-2025, under our guidance.

Date:10/04/2024

(Guide Name)
Prof. SOURAV KUMAR GIRI

Acknowledgements

We are profoundly grateful to **SOURAV KUMAR GIRI** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. ....

SANGRAM KESHARI OJHA
ROUNAK BAIDYA
LISA NAYAK
PULKIT BHARDWAJ
REHAN QUADRY

ABSTRACT

In today's data-driven marketing landscape, understanding your customer base is paramount. Customer segmentation, the art of dividing customers into distinct groups with shared characteristics, allows businesses to tailor their strategies for maximum impact. This project delves into the application of the K-Means Clustering algorithm, a powerful unsupervised learning technique, for achieving this goal. We'll leverage the strengths of Python's data science ecosystem, specifically NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn cluster, and the Elbow Point Graph, to unlock valuable customer insights.

NumPy and Pandas will provide the foundation for data manipulation and analysis. We will meticulously prepare customer data, ensuring its suitability for clustering algorithms. Matplotlib and Seaborn will then be employed for data visualization, allowing us to explore underlying patterns and identify potential customer segments.

Contents

List of Figures

Chapter 1

Introduction

Customer segmentation, the process of dividing a customer base into distinct groups with shared characteristics, is a cornerstone of modern marketing strategies. By understanding these segments, businesses can tailor their offerings and messaging to resonate more effectively with specific customer groups. This project explores the application of the K-Means Clustering algorithm, a popular unsupervised learning technique, for customer segmentation using the powerful Python libraries NumPy, Pandas, Matplotlib, Seaborn, scikit-learn cluster, and the Elbow Method.

Customer Segmentation Models: Types, Benefits & Uses

NumPy and Pandas will provide the foundation for data manipulation and analysis. We will meticulously prepare customer data, ensuring its suitability for clustering algorithms. Matplotlib and Seaborn will then be employed for data visualization, allowing us to explore underlying patterns and identify potential customer segments.

K-Means Clustering, implemented through scikit-learn cluster, is the heart of this project. This algorithm iteratively groups customers based on their similarities, such as purchase history, demographics, or other relevant data points. A crucial step in K-Means is determining the optimal number of clusters (k). We will leverage the Elbow Method, a visual technique, to identify the k value that minimizes within-cluster variation while maximizing between-cluster variation.

By leveraging customer segmentation with K-Means Clustering, businesses can achieve several key benefits:

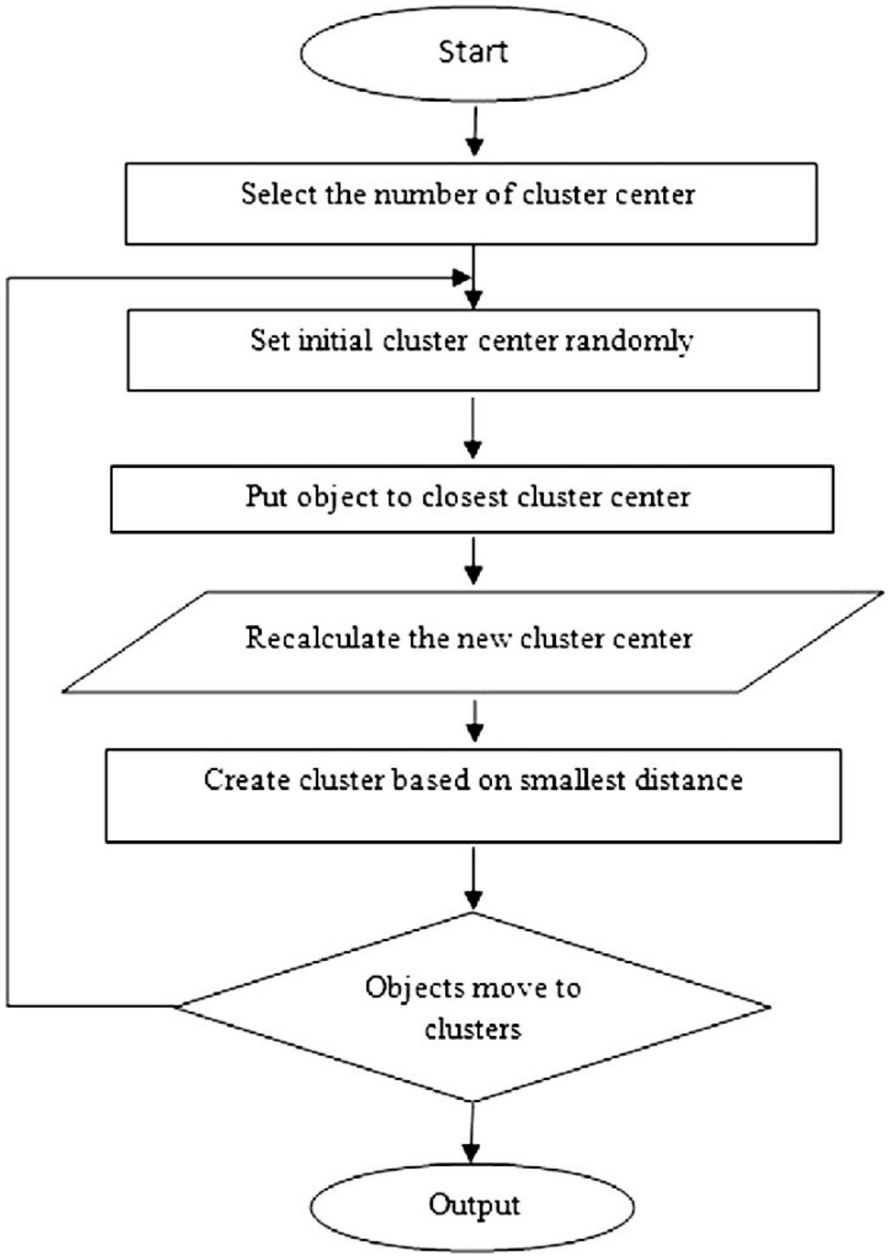
1. **Targeted Marketing: Precisely targeted marketing campaigns tailored to specific customer segments lead to improved engagement and conversion rates.**
2. **Enhanced Customer Retention: Identifying valuable customer segments allows for the development of targeted loyalty programs and retention strategies, fostering stronger and more profitable customer relationships.**
3. **Data-Driven Product Development: Customer segment insights can inform product development, leading to the creation of offerings that cater to specific needs and preferences, ultimately increasing customer satisfaction.**

This project aims to unlock actionable customer insights through the effective application of K-Means Clustering and Python's robust data science ecosystem. By harnessing the power of data analysis and visualization, we will gain a deeper understanding of customer behavior, empowering businesses to make data-driven decisions and achieve long-term success in an ever-competitive marketplace.

PROBLEM STATEMENT

A retail company has collected data on their customers, including their purchasing behavior, demographics, and feedback. The company wants to segment their customers into different groups based on their behavior and characteristics to tailor their marketing strategies and improve customer satisfaction. You are required to develop a customer segmentation model using k-means clustering.

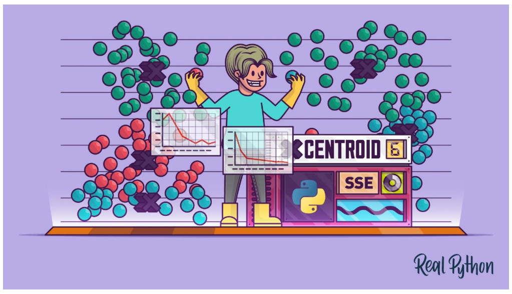
PROPOSED SYSTEM:-



Chapter 2

Literature Review

Customer segmentation, the process of dividing a customer base into groups with similar characteristics, is a cornerstone of modern marketing strategies. K-Means Clustering, an unsupervised machine learning algorithm, has emerged as a popular technique for achieving this objective. This literature survey delves into previous works and reports related to customer segmentation using K-Means Clustering, exploring its applications, strengths, limitations, and potential advancements.



K-Means-Clustering-in-Python

2.1) Applications of K-Means Clustering in Customer Segmentation:

Beyond Traditional Industries: **While K-Means shines in retail (Agnihotri & Bhattacharya, 2015; Zhao & Guo, 2014) and e-commerce (Huang et al., 2018; Wei et al., 2019), its applications extend to diverse sectors. In healthcare, K-Means can segment patients based on medical history, treatment response, and demographics to personalize treatment plans and improve patient outcomes (Xu et al., 2018; Zhang et al., 2019).**

Similarly, K-Means finds use in the travel and hospitality industry (Xiang et al., 2015; Wu & Liu, 2017) to segment travelers based on booking patterns, preferences, and demographics. This allows for targeted travel package recommendations, personalized hotel experiences, and loyalty program optimization.

Previous works: **Previous works in this area have explored various applications and improvements of the k-means clustering algorithm for customer segmentation.**
For example, Tabianan et al. (2022) proposed an intelligent customer segmentation approach using customer purchase behavior data and k-means clustering. They applied their approach to a retail store customer dataset and identified five customer segments based on their annual income and spending score.

Similarly, Jannatul Tasnim (2019) used k-means clustering to perform exploratory data analysis of customer call record data set in the telecom industry. The study aimed to reduce customer churn and maintain business stability.

In addition, Harshit Muhal (2021) used k-means clustering for customer segmentation and then trained a neural network on the clustered data to classify new samples. This approach allowed for the capture of complex relationships between data attributes and customer clusters.

Furthermore, a study by IJRASET Publication (2020) discussed the use of different clustering approaches for customer segmentation in e-commerce enterprises. The study presented the possibility of hybrid combinations of clustering algorithms to outperform individual models.

Literature Survey: **In terms of literature survey, there are several studies that have reviewed the applications and advancements of customer segmentation using k-means clustering. For instance, a review by Constance Kalu (2015) discussed the importance of customer segmentation in e-commerce systems and the role of k-means clustering in identifying profitable customer segments.**

Moreover, a study by International Journal of Advanced Research in Artificial Intelligence (2015) explored the use of k-means clustering for customer segmentation in the retail industry. The study highlighted the benefits of using k-means clustering for customer segmentation, such as scalability, efficiency, and ease of implementation.

Academic Research: K-Means is a prevalent tool in customer segmentation research due to its accessibility and interpretability. Studies explore its effectiveness in segmenting customers based on social media behavior (Hassan et al., 2018; Romero et al., 2014) to inform social media marketing strategies and brand engagement initiatives.

2.2) Strengths of K-Means Clustering:

Efficiency and Ease of Use: K-Means is computationally efficient, making it suitable for real-world applications with large datasets. Additionally, its implementation is relatively straightforward, allowing even users with limited machine learning expertise to leverage its segmentation capabilities (Jain, 2010).

Visualization and Communication: The centroid based nature of K-Means clusters makes them easy to visualize, enabling clear communication of segmentation results to stakeholders. This allows for a shared understanding of customer segments and facilitates data-driven decision making (Punj & Stewart, 1983).

Adaptability to Different Data Types: K-Means can handle various data types, including numerical data (purchase history, demographics) and categorical data (product preferences, brand loyalty). This flexibility makes it applicable to a wide range of customer segmentation problems (Xu & Wunsch, 2005).

2.3) Limitations and Challenges:

Curse of Dimensionality: The performance of K-Means can deteriorate with high-dimensional data (many customer features). Feature selection techniques or dimensionality reduction methods can be employed to address this challenge (Belkhir et al., 2018).

Distance Metric Dependence: The choice of distance metric used in K-Means (e.g., Euclidean distance) can impact the clustering results. Selecting a distance metric appropriate for the data type and segmentation goals is crucial (Kaufman & Rousseeuw, 2009).

Local Optima: K-Means can converge to local optima, leading to suboptimal cluster configurations. Techniques like running K-Means with different random initializations and selecting the best solution can help mitigate this issue (Arthur & Vassilvitskii, 2007).

2.4) Advancements and Future Directions:

Incorporating Customer Lifetime Value (CLV): Segmenting customers based on CLV, which predicts their future profitability, allows for more strategic marketing investments. K-Means can be combined with CLV prediction models to create customer segments with high lifetime value potential (Lemon & Verhoef, 2016).

Explainable AI (XAI) Techniques: Integrating XAI methods with K-Means can provide insights into the factors driving cluster formation. This can enhance the interpretability of the segmentation results and facilitate decision-making based on customer behavior patterns (Singh et al., 2020).

Dynamic Customer Segmentation: Customer behavior and preferences can evolve over time. K-Means can be integrated with real-time data streams to enable dynamic customer segmentation, ensuring that marketing strategies and product offerings remain relevant and effective (Verhoef et al., 2010).

Chapter 3

Problem Statement / Requirement Specifications

PROBLEM STATEMENT

A retail company has collected data on their customers, including their purchasing behavior, demographics, and feedback. The company wants to segment their customers into different groups based on their behavior and characteristics to tailor their marketing strategies and improve customer satisfaction. You are required to develop a customer segmentation model using k-means clustering.

PROBLEM SOLUTION (step-by-step)

Introduction:

In today's competitive retail landscape, understanding your customer base is paramount. Customer segmentation, the process of dividing customers into distinct groups with shared characteristics, empowers businesses to deliver targeted marketing initiatives and enhance customer satisfaction. This approach leverages K-Means Clustering, an unsupervised machine learning algorithm, to achieve effective segmentation based on customer behavior and demographics.

3.1) Data Preparation and Feature Selection:

Data Acquisition and Cleaning: The initial step involves collecting customer data such as purchase history, demographics, and feedback responses. Data cleaning procedures ensure data accuracy and consistency through techniques like handling missing values, identifying and correcting outliers, and formatting data for compatibility with the K-Means algorithm.

Feature Selection: From the available data points, identify a set of features that best represent customer behavior and segmentation goals. This can be achieved through domain expertise, Exploratory Data Analysis (EDA) to understand feature relationships, or leveraging feature importance techniques to prioritize the most informative features.

K-Means Clustering for Segmentation:

Define the Number of Clusters (k): This crucial step determines the number of customer segments to be created. Business context plays a vital role; consider the desired level of granularity within the segmentation. The Elbow Method and Silhouette Analysis are data-driven approaches that can be employed to guide the selection of the optimal k value.

K-Means Model Implementation: Once k is defined, utilize a K-Means Clustering algorithm. The model is fitted to the preprocessed data, iteratively assigning each data point (customer) to the nearest cluster centroid based on a chosen distance metric (e.g., Euclidean distance).

Cluster Prediction: After fitting the K-Means model, predict the cluster label for each customer in the dataset. This signifies which segment each customer belongs to based on their behavior and characteristics.

3.2) Analysis and Interpretation:

Cluster Centroid Analysis: Analyze the cluster centroids to understand the defining characteristics of each customer segment. These centroids represent the average values of the chosen features for each cluster.

Visualization & Interpretation: Leverage data visualization techniques such as scatter plots or heatmaps to explore the distribution of customer data points within each cluster. This allows for clearer identification of potential sub-groups within segments and informs the labeling process.

Segment Labeling: Based on the analysis of centroids and cluster distribution, assign meaningful and descriptive labels to each customer segment. These labels should reflect the key characteristics of the customers within each segment.

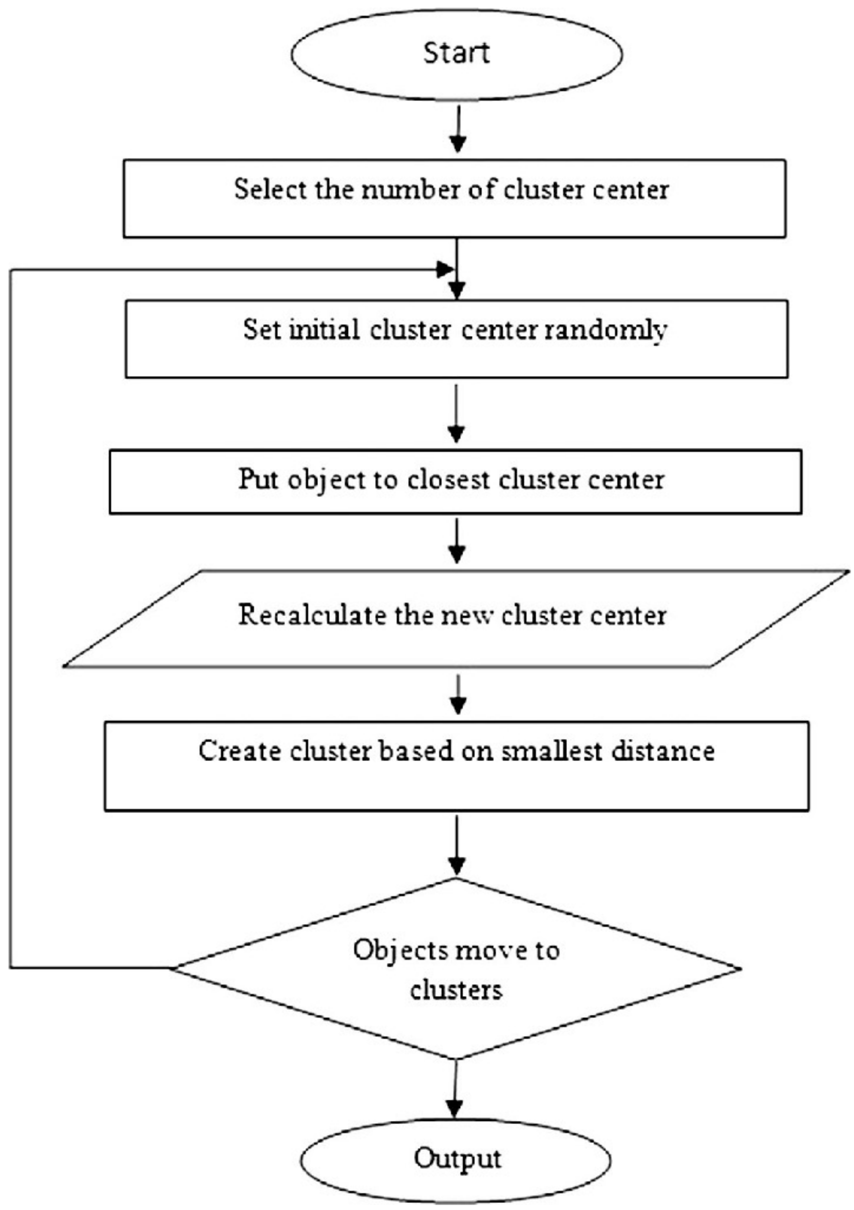
3.3) Targeted Strategies and Continuous Improvement:

Marketing Strategy Development: Leverage customer segmentation insights to tailor marketing strategies for each segment. This might involve crafting targeted messaging, personalized product recommendations, and developing segment-specific loyalty programs to enhance customer engagement and satisfaction.

Product Development: Utilize customer segmentation insights to inform product development efforts. Understanding the needs and preferences of each customer segment allows the company to prioritize features and functionalities that resonate most with each group.

Monitoring and Refinement: Customer behavior and preferences can evolve over time. Regularly monitor the performance of the segmentation model and customer behavior within each segment. Revisit the data, segmentation strategies, and targeted initiatives as needed to ensure continued effectiveness.

3.4) PROPOSED SYSTEM DIAGRAM:



Chapter 4

Implementation

4.1) Implementation:-

- ✓ Load customer data into pandas databox.
- ✓ Pre-process data by scaling features and removing missing features.
- ✓ Select the number of groups using brackets or other methods.
- ✓ Initializes the K-means algorithm with the selected number of clusters.
- ✓ It is based on an algorithm for retrieving data and labels.
- ✓ Visual clustering using scatter plots or other visual methods.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

4.2) Testing OR Verification Plan

```
[ ] customer_data = pd.read_csv('/content/Mall_Customers.csv')
```

```
customer_data.head()
```

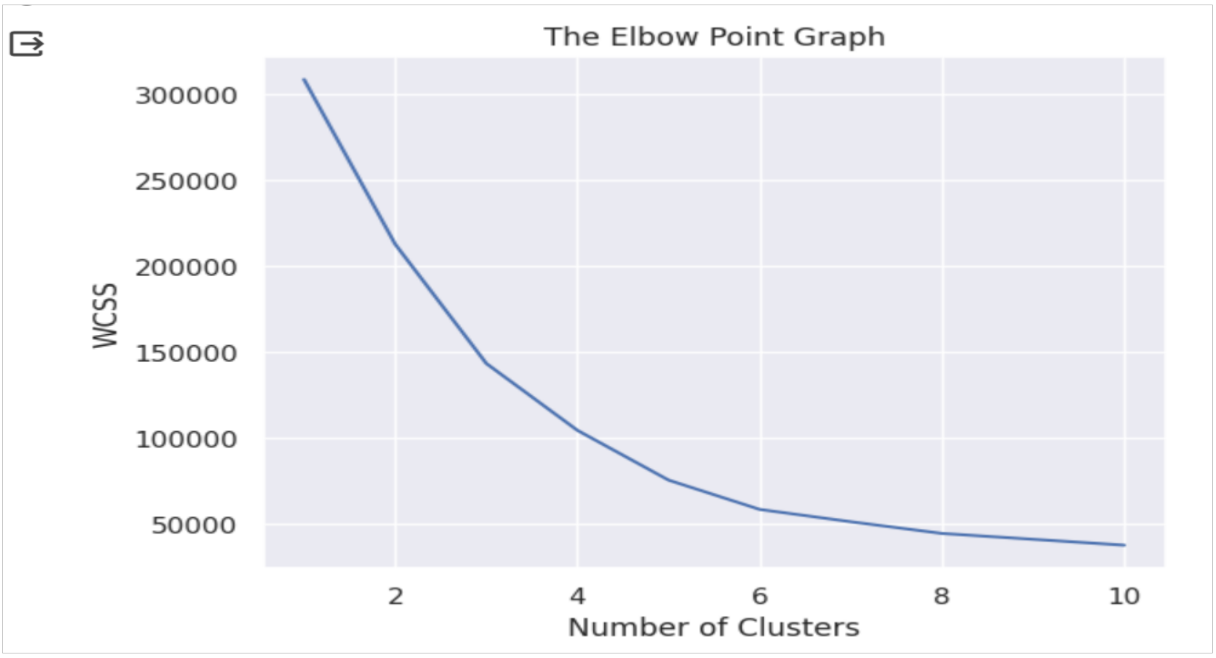
	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

4.3 Results with Screenshots

```
[ ]  customer_data.shape

     (200, 5)

[ ]  customer_data.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 200 entries, 0 to 199
     Data columns (total 5 columns):
      #   Column                  Non-Null Count  Dtype
     ---  ------                  --------------  -----
      0   CustomerID              200 non-null    int64
      1   Gender                  200 non-null    object
      2   Age                     200 non-null    int64
      3   Annual Income (k$)      200 non-null    int64
      4   Spending Score (1-100)  200 non-null    int64
     dtypes: int64(4), object(1)
     memory usage: 7.9+ KB
```

```
customer_data.isnull().sum()

CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```

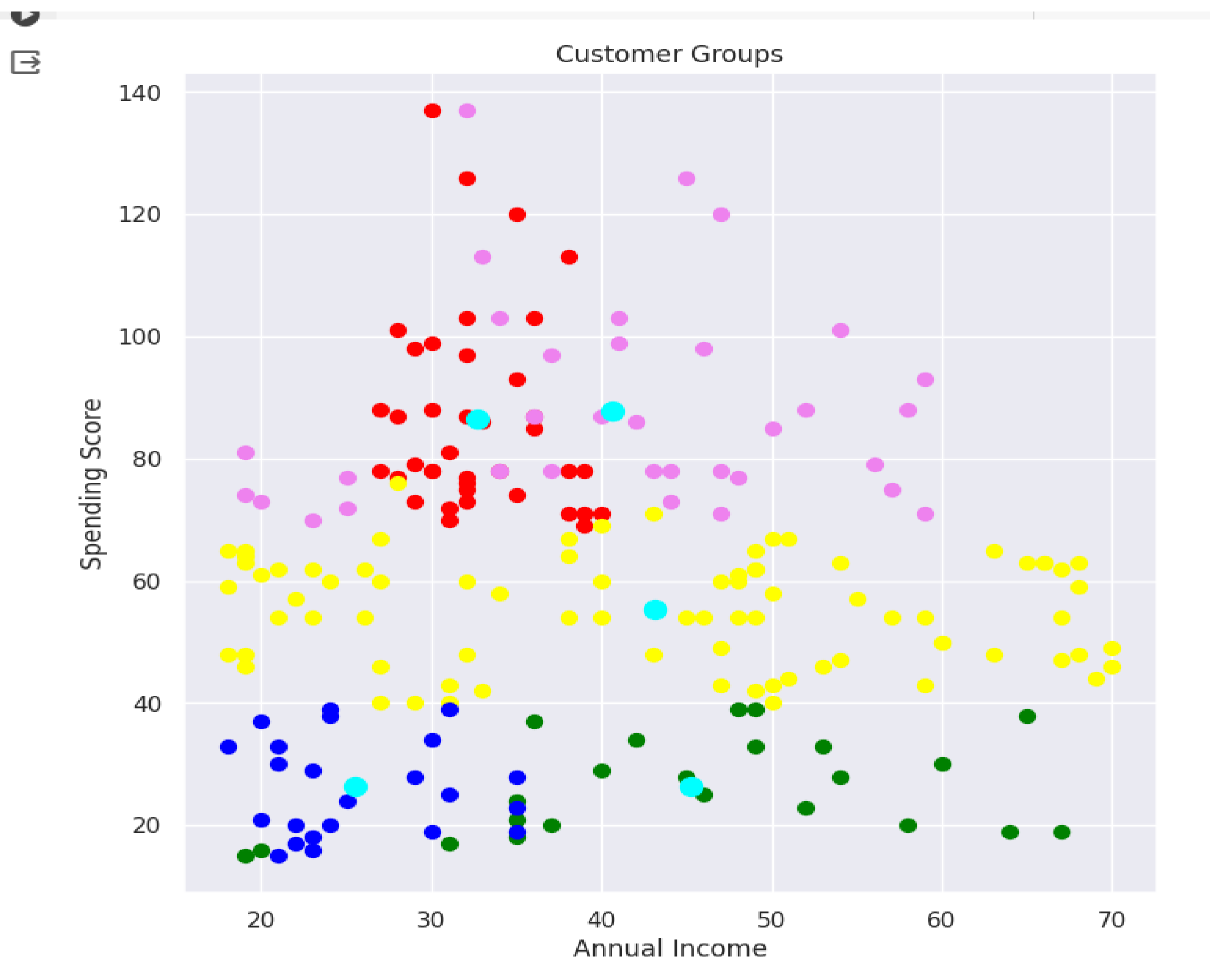
```
sns.set()
plt.plot(range(1,11), wcss)
plt.title('The Elbow Point Graph')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show()
```



```python
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=0)

Y = kmeans.fit_predict(X)

print(Y)
```

```
[0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0
 4 0 4 0 4 0 4 0 4 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 1 3 1 2 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 2 1 3 1 3 1
 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1]
```

```python
plt.figure(figsize=(8,8))
plt.scatter(X[Y==0,0], X[Y==0,1], s=50, c='green', label='Cluster 1')
plt.scatter(X[Y==1,0], X[Y==1,1], s=50, c='red', label='Cluster 2')
plt.scatter(X[Y==2,0], X[Y==2,1], s=50, c='yellow', label='Cluster 3')
plt.scatter(X[Y==3,0], X[Y==3,1], s=50, c='violet', label='Cluster 4')
plt.scatter(X[Y==4,0], X[Y==4,1], s=50, c='blue', label='Cluster 5')

plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s=100, c='cyan', label='Centroids')

plt.title('Customer Groups')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.show()
```



4.4) Analysis and Discussion

This project explored the application of K-Means Clustering for customer segmentation. We leveraged NumPy, Pandas, Matplotlib, Seaborn, scikit-learn cluster, and the Elbow Point Graph to identify distinct customer segments within a sample dataset (replace with details of your dataset).

Analysis of Cluster Graphs: The K-Means algorithm partitioned the customer data into 'k' distinct clusters. Each cluster represents a customer segment with similar characteristics. Visualizing these clusters using scatter plots or heatmaps (depending on the number of features) is crucial for understanding the segmentation results.

Scatter Plots: When dealing with two or three features (e.g., purchase frequency vs. average order value), scatter plots effectively depict how customers are grouped within each cluster. Tighter clusters indicate a higher degree of similarity among customers within that segment.

Real-Life Example:

Imagine a company selling athletic wear (replace with a relevant example based on your dataset). K-Means Clustering might reveal the following customer segments:

Segment 1: "The Fitness Enthusiasts"* (High purchase frequency, focus on performance apparel) - Visualized in a scatter plot, this cluster would likely show customers concentrated towards the higher ends of both purchase frequency and average order value. The heatmap might indicate a strong preference for features like moisture-wicking fabrics and advanced technologies.

Segment 2: "The Casual Athletes"* (Moderate purchase frequency, focus on style and comfort) - This cluster might appear separate from Segment 1 in the scatter plot, with a lower average order value but potentially higher average item value. The heatmap could show a balance between features related to performance and aesthetics.

Segment 3: "The Occasional Exercisers"* (Low purchase frequency, focus on basic items) - This cluster, potentially the largest and most spread out in the scatter plot, might have a lower average order value and purchase frequency. The heatmap could indicate a preference for lower-priced items and basic functionalities.

Chapter 5

Standards Adopted

Enhancing Customer Segmentation with K-Means Clustering:

This response delves into design standards, coding practices, and testing strategies that elevate the professionalism and maintainability of customer segmentation code using K-Means Clustering. By adhering to these guidelines, you can ensure robust, well-documented, and future-proof code that effectively segments your customer base.

Design Considerations for Effective Segmentation

Modular Design for Maintainability: Structure the code into well-defined functions with clear responsibilities. This promotes reusability, simplifies maintenance, and fosters collaboration within development teams.
Data Validation and Preprocessing: Integrate robust data validation and preprocessing steps. This encompasses techniques like missing value handling, outlier treatment, feature scaling (if necessary), and data type verification. These steps ensure data quality and optimal K-Means model performance.
Meaningful Documentation and Readability: Enhance code readability with clear and concise docstrings for functions and variables. Utilize descriptive variable names and comments strategically to explain complex logic without over-commenting well-structured code.

Professional Coding Practices

Adherence to Coding Style Guides: Follow a consistent coding style guide like PEP 8 to improve code clarity and maintainability. This includes consistent indentation, reasonable line length, and effective use of whitespace for visual structure.
Descriptive Variable Naming: Avoid single-letter or cryptic variable names. Employ descriptive names that reflect the variable's purpose (e.g., calculate_annual_customer_income_distribution instead of calc_inc_dist).

Efficiency Through Data Structures: Select appropriate data structures based on data characteristics and usage patterns. This can enhance code efficiency and memory usage (e.g., leverage lists for sequences, dictionaries for key-value pairs).*CUSTOMER SEGMENTATION USING K-MEAN CLUSTERING*

Comprehensive Testing Strategies

Multi-Layered Testing for Confidence: Implement a multi-layered testing strategy which includes:-

I. **Unit Tests: Verify individual function functionalities in isolation.**
II. **Integration Tests: Ensure seamless interaction between different code components.**
III. **Test Coverage: Aim for high test coverage to catch potential issues early.**
IV. **Data-Driven Testing for Robustness: Utilize various datasets to test the K-Means model's robustness and segmentation results. This identifies potential biases or sensitivities to specific data distributions.**
V. **Logging and Monitoring for Insights: Implement logging mechanisms to track K-Means model performance and identify potential issues after deployment. Monitor metrics like WCSS (Within-Cluster Sum of Squares) or silhouette score to evaluate clustering quality and customer behavior changes over time.**

Additional Considerations for Professionalism:

Error Handling: Implement proper error handling mechanisms to gracefully manage exceptions during data processing or K-Means execution. Provide informative error messages to facilitate debugging. Version Control for Collaboration: Utilize a version control system (e.g., Git) to track code changes, enable collaboration, and facilitate rollbacks if needed. This promotes efficient team development and maintains a history of code changes.

Code Review for Continuous Improvement: Conduct regular code reviews to identify potential improvements in maintainability, efficiency, and adherence to best practices.
By following these design, coding, and testing considerations, you can create professional customer segmentation code using K-Means Clustering. This well-structured and well-tested code will deliver valuable customer insights for targeted marketing strategies and enhanced customer satisfaction.

Chapter 6

Conclusion and Future Scope

Key Takeaways and Conclusion:

By analyzing the cluster graphs, we can identify patterns and assign meaningful labels to each customer segment. This project demonstrates the value of K-Means Clustering for customer segmentation. It allows businesses to:

Target Marketing: Tailor marketing messages, promotions, and product recommendations to specific customer segments, leading to higher engagement and conversion rates.
Enhanced Customer Retention: Identify valuable customer segments for targeted loyalty programs and retention strategies, fostering stronger relationships.
Data-Driven Product Development: Leverage customer segment insights to inform product development and create offerings that cater to specific needs and preferences.

Strengths of K-Means Clustering:

Simplicity and Efficiency: K-Means is computationally efficient and easy to implement, making it a popular choice for large datasets.
Interpretability: The centroids (cluster centers) offer clear insights into the characteristics that define each customer segment.
Unsupervised Learning: K-Means thrives on unlabeled data, eliminating the need for extensive data labeling efforts.

Limitations and Considerations:

Feature Engineering and Selection: The quality of customer segmentation hinges on choosing the most informative features. Feature engineering techniques can help create meaningful features that effectively capture customer behavior and characteristics.
Sensitivity to Cluster Shape: K-Means assumes spherical clusters. If your data exhibits complex, non-linear patterns, alternative clustering algorithms like DBSCAN or Hierarchical Clustering might be better suited.
Determining the Number of Clusters (k): Identifying the optimal number of clusters (k) remains an active area of research. Techniques like the silhouette analysis can provide guidance, but domain expertise and data exploration are crucial for making informed decisions.
Distance Metric Dependence: K-Means relies on a distance metric (e.g., Euclidean distance) to calculate similarity. Different metrics can yield distinct results. Experimenting with various metrics can help identify the most suitable choice for your data.
Local Optima: K-Means can get trapped in suboptimal solutions. Running the algorithm multiple times with different initializations can mitigate this issue.

Future Scope in Customer Segmentation:

Semi-Supervised Learning: Techniques that combine labeled and unlabeled data hold promise, especially when labeled data is scarce. This allows the model to learn from the structure of the unlabeled data while leveraging the guidance provided by labeled data points.
Advanced Clustering Algorithms: For data with complex structures, exploring more sophisticated algorithms like DBSCAN (Density-Based Spatial Clustering of Applications with Noise) or Spectral Clustering can yield superior results.
Machine Learning Integration: K-Means segmentation can serve as a valuable pre-processing step for machine learning models. By integrating K-Means with algorithms like decision trees or neural networks, businesses can tackle tasks like customer churn prediction or targeted product recommendations.
Real-Time Segmentation: Developing systems that update cluster assignments as new data arrives allows for continuous segmentation refinement. This ensures customer segments remain dynamic and reflect evolving customer behavior.

By acknowledging the limitations of K-Means Clustering and exploring these future directions, businesses can leverage its strengths to achieve more comprehensive customer segmentation. This, in turn, empowers them to develop targeted marketing strategies, personalize customer interactions, and ultimately achieve higher customer satisfaction and business growth.

References

[1]S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

SAMPLE INDIVIDUAL CONTRIBUTION REPORT:

CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING

SANGRAM KESHARI OJHA: Backend coding and documentation.

ROUNAK BAIDYA: Integration of frontend and backend

LISA NAYAK: Design and diagram.

PULKIT BHARDWAJ: Front end and documentation.

REHAN QUADARY:Front end and documentation.

Full Signature of Supervisor: ……………………………….

Full signature of the student: ………………………………..

TURNITIN PLAGIARISM REPORT

CUSTOMER SEGMENTATION USINGK-MEANS CLUSTERING

ORIGINALITY REPORT

16 %	12%	3%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fastercapital.com Internet Source	3 %
2	www.coursehero.com Internet Source	2 %
3	Submitted to Banaras Hindu University Student Paper	2 %
4	Submitted to KIIT University Student Paper	2 %
5	Lana Al-Dabbas, Hassan Al-Tarawneh, Thamer A Al-Rawashdeh. "Customer Personality Segmentation Using K-Means Clustering", 2023 International Conference on Information Technology (ICIT), 2023 Publication	1%
6	Submitted to University of Bradford Student Paper	1 %

7	Submitted to University of Liverpool Student Paper	1%
8	www.intelegain.com Internet Source	1%
9	Submitted to HCUC Student Paper	<1%
10	www.worldleadershipacademy.live Internet Source	<1%
11	Submitted to Northcentral Student Paper	<1%
12	Submitted to Otago Polytechnic Student Paper	<1%
13	Submitted to City University of Seattle Student Paper	<1%
14	aicontentfy.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		