

# Capstone Project Submission

## Instructions:

- i) Please fill all the required information.
- ii) Avoid grammatical errors.

**Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data companies can then out perform the competition by developing uniquely appealing products and services.

Here the main objective was to analyse and identify major customer segments using K Means algorithm and also using different verification method to confirm the result.

After loading the dataset I looked for duplicate values in the column. There were none. So I performed EDA. This process helped me figure out various aspects and relationships among the target and the independent variables. It gave me a better idea of which feature behaves in which manner compared to the target variable.

The dataset contains around 135080 null values in CustomerId which might tend to disturb the score so I dropped null values because of the high quantity of it.

After that I did Feature engineering so it is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved accuracy on unseen data.

And I converted the Invoice Date column into date time format and created new features like day, day\_num, month\_num, year, hours etc. From Invoice Date and again created Total Amount from Product of Quantity and Unit price columns. This can help improve machine learning accuracy since algorithms tend to have a hard time dealing with high cardinality columns.

For modelling I tried various algorithms like:

1. **K-means clustering**
2. **Hierarchical clustering**

**Contributor Roles:****Mohammed Javeed**

- 1. Data Summary**
- 2. Data Cleaning**
- 3. EDA on features**
- 4. Feature Engineering**
- 5. Feature Selection**
- 6. Modelling**
- 7. Evaluation Metrics of the model**
- 8. Final selection of the model**

Github Link:

<https://github.com/Javee01/Customer-Segmentation>

**Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Colab notebook, project presentation and video.**

Drive Link:

<https://drive.google.com/drive/folders/1HupY0HFBQiyFsKfr-uvND0qHYCYvpXaM?usp=sharing>