A dark blue vertical bar runs down the left side of the slide. A blue arrow points to the right from this bar, containing the date.

23-12-2022

Customer Segmentation

By- SHWETABH BISWAS

Several thin, curved lines in dark blue and light grey originate from the bottom left corner and sweep upwards and to the right.

biswas.6@iitj.ac.in
FEYNN LABS



CUSTOMER SEGMENTATION

A project report submitted by

SHWETABH BISWAS

As 1st project in a series of 3 projects for the position of
intern at

FEYNN LABS

CUSTOMER SEGMENTATION

AS A BUSINESS OPPORTUNITY

PROBLEM STATEMENT

Customer segmentation is the process of dividing / grouping customers based on their different attributes / characteristics. This can help the owner of any given business understand its customer better and making better offers to them and having a better shot at success. The modern era of Artificial Intelligence is capable of segmenting a given dataset based on its features using different algorithms like KNN, Support Vector Machines (SVM), etc. This category of problems is called classification.

Here in this report, we'll try to understand how to perform customer segmentation using a dataset from Kaggle. It contains multiple features to perform segmentation and we'll also interpret the inter-dependence of features on other features as well as segments.

BUSINESS NEED ASSESSMENT

Customer Segmentation can be used to develop various businesses. Some examples are –

- Marketing :
 - Improve Marketing focus – segments have different interests, values, tastes and reasons to purchase.
 - Identify most and least profitable customers.
 - Build loyal relationships.
 - Create representative customer.
- Product and brand :
 - Brands to appeal to particular segment.
 - Customize products and services.
 - Predict future purchasing patterns.
- Pricing :
 - Price products differently.
 - Willingness to pay for optimal value.
 -

TARGET SPECIFICATION AND CHARACTERIZATION

The dataset used here can be of customers from any domain like E-commerce, marketing services, IT sector etc. The features of the customers are – Gender, Marital Status, Age, Graduated, Profession, Work Experience, Spending tendency and Family Size. These are self explanatory.

EXTERNAL SEARCHES

The dataset used here is taken from Kaggle (1). It consists of 8068 rows. The analysis is independently done by me. The viewpoint of this as a business opportunity is applied for an E-commerce website.

BENCH MARKING

At present, only big companies like Amazon / Wallmart and social media giants like Meta / Instagram are using this technique of segmenting customers based on their features. This project will try implementing it on a smaller scale like a shopkeeper / fruit vendor.

APPLICABLE REGULATIONS

1. Must provide access to third – party websites for data collection and monitor the authenticity and behavior of the service.
2. Laws against data collection – some websites might have laws against customer data collection.
3. Privacy and delicacy of data collected must be ensured always.
4. Using the data for academic and research purposes by implementing different algorithms and techniques should be permitted.

APPLICABLE CONSTRAINTS

1. Continuous data collection and maintenance.
2. Lack of technical knowledge for the used.
3. Taking care of rarely brought products.

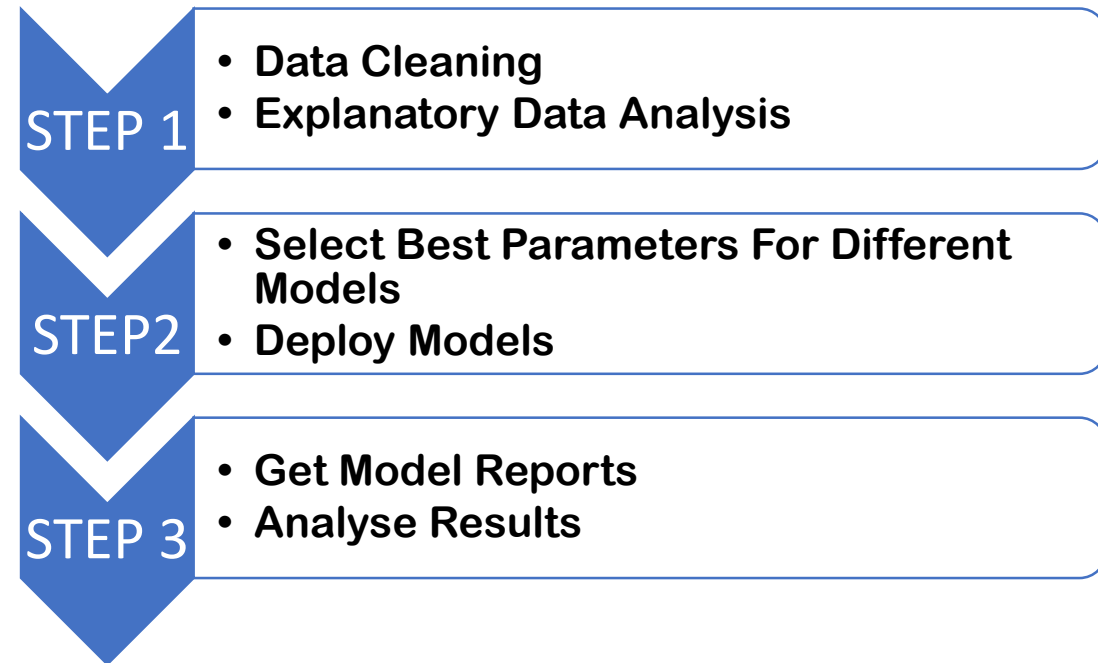
BUSINESS MODEL

The idea of segmenting customers may be helpful in generating extra revenue for the business in the following ways –

1. Recommending products based on the segment of the customer – If the user is recommended product based on the purchase of another user of the same group/segment, it's more likely to be appreciated.
2. Providing relevant discounts – If different products are offered as a collective discount, there is a much better chance it'll be accepted.
3. Social media campaigns based on segmented customers – if the user is shown ads based on preferences of others users of the same group, it'll definitely yield better results.

FINAL PRODUCT PROTOTYPE

The technique works by deploying various Machine Learning Models on training data to categorize data based on its features and select the model with the highest efficiency. The model categorizes data into segments having similar features. The model is then used to predict the groups of new users based on their features. The flow chart below demonstrates the workings –



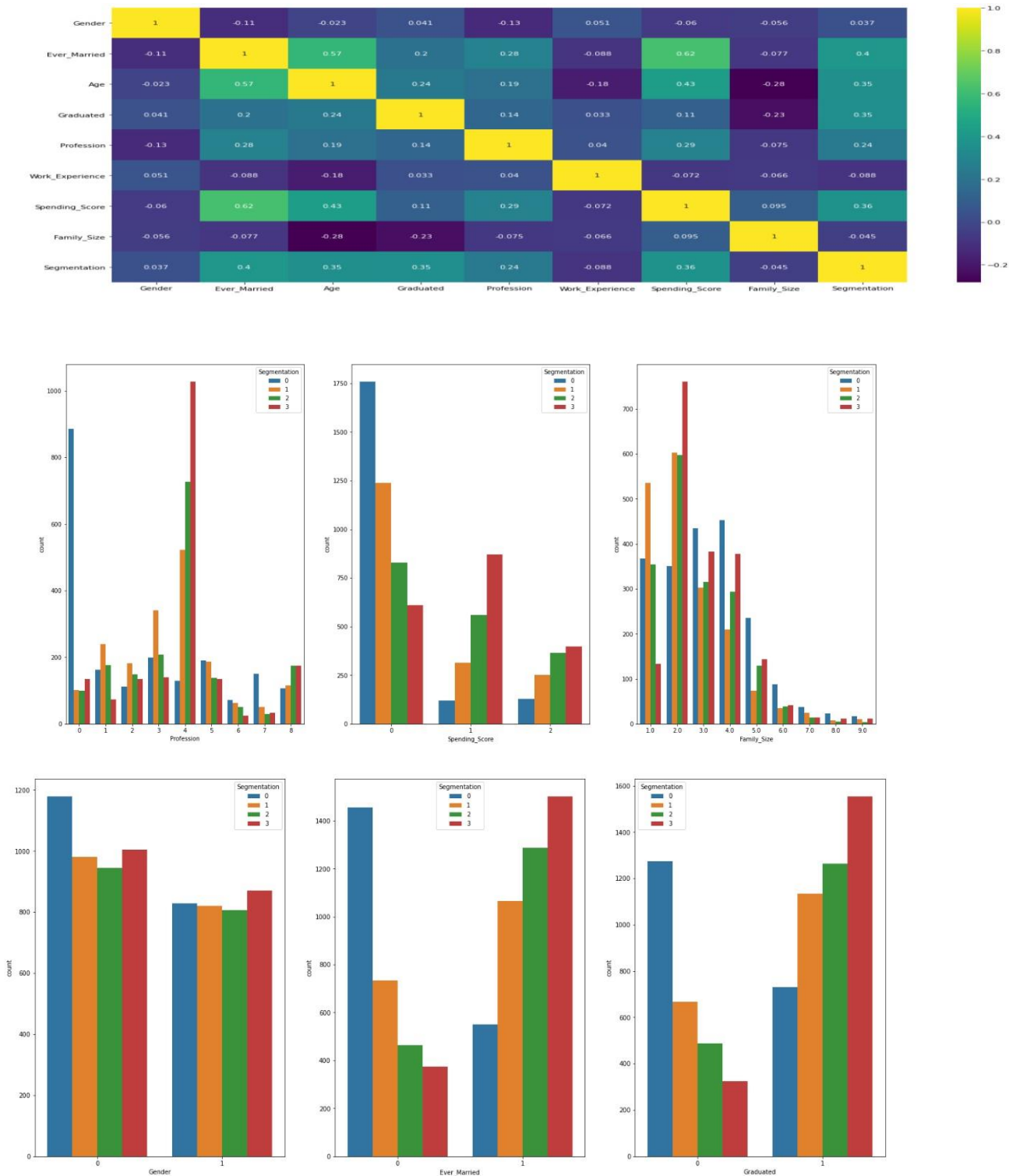
PRODUCT DETAILS AND CODE IMPLEMENTATION

A prototype of this idea was implemented by me and can be found in my GITHUB Repository (2). Let's now understand its working. First let's have a glimpse at the data after preprocessing –

	Gender	Ever_Married	Age	Graduated	Profession	Work_Experience	Spending_Score	Family_Size	Segmentation
0	Male	No	22	No	Healthcare	1.0	Low	4.0	D
1	Female	Yes	38	Yes	Engineer	NaN	Average	3.0	A
2	Female	Yes	67	Yes	Engineer	1.0	Low	1.0	B
3	Male	Yes	67	Yes	Lawyer	0.0	High	2.0	B
4	Female	Yes	40	Yes	Entertainment	NaN	High	6.0	A
...
8063	Male	No	22	No	NaN	0.0	Low	7.0	D
8064	Male	No	35	No	Executive	3.0	Low	4.0	D
8065	Female	No	33	Yes	Healthcare	1.0	Low	1.0	D
8066	Female	No	27	Yes	Healthcare	1.0	Low	4.0	B
8067	Male	Yes	37	Yes	Executive	0.0	Average	3.0	B

8068 rows × 9 columns

Next we'll try to find the correlation between various features and also between the target segments and the features.



Next, the hyper - parameters are needed to be tuned for the optimal performance of the specific model. The model score after fitting and predicting the segment for a given input is considered as optimal parameter for measuring the performance of the model. The different model scores for different models are –

MODEL	SCORE
GRADIENT BOOSTING CLASSIFIER	0.5172
RANDOM FOREST CLASSIFIER	0.449
LOGISTIC REGRESSION	0.516
K NEIGHBOURS CLASSIFIER	0.436
SUPPORT VECTOR CLASSIFIER	0.514

As per model scores, Gradient Boosting Method performs the best among all the used models and should be used for classifying users of unknown category.

CONCLUSION

Here in this project, I've tried to demonstrate the benefits of segmenting customers, the process of doing it and deploy a model for helping a small / medium scale business. At present, this technique of segmenting customers is used by big companies like E-commerce website Amazon/Flipkart, or educational websites like Byju's/PW etc. With proper implementation, this can also be used in small businesses like for shopping / grocery purposes. The opportunities are limitless and results are fruitful.

REFERENCES

1. <https://www.kaggle.com/code/georgevosorov/customer-segmetation-xgboost-0-5278/data?select=Train.csv>
2. <https://github.com/shwetabh-23/Customer-Segmentation.git>