Customer Segmentation for NPD Targeting

Abstract

In today's digital age, businesses are constantly interacting with customers across multiple touchpoints, collecting valuable customer data that can inform their overall strategy. This customer data is crucial for creating a successful business strategy, as it provides insight into customer demography and geography. Data-driven organizations understand the importance of gathering customer data and using it to improve the customer experience and optimize their business strategy. By collecting and analyzing customer data from various sources, businesses can gain a competitive advantage and stay ahead of the curve in today's dynamic market. The goal of the project is to highlight the significance of customer data in shaping a successful business strategy and emphasizes the importance of leveraging customer data to gain a deeper understanding of the customer base. The project's focus is selecting appropriate features for customer segmentation to help the business finalize a target demography and/or geography to increase their revenue.

Data source and Description

We have taken the data source from Kaggle (https://www.kaggle.com/datasets/iamsouravbanerjee/customer-segmentation-51k-records). It has 51,000 records with customer demography and geography information. The dataset consists of the following features:

No.	Attribute	Description
1	first_name	First Name of the customer
2	last_name	Last Name of the customer
3	title	Prefix of Customer Name
4	gender	Customer's gender
5	email	Email of the customer
6	city	City of residence
7	country	Country of residence
8	country_code	Country Code of Residence
9	latitude	Location parameter (from east to west)
10	longitude	Location Parameter (from north to south)
11	phone	Customer's contact number
12	street_address	Customer contact address
13	street_name	Street Name
14	street_number	Street Number
15	street_suffix	Street suffix
16	time_zone	Time Zone
17	company_name	Name of Customer's Employee
18	department	Department
19	job_title	Job Role of the customer
20	language	Customer's Language
21	university	Customer's University
22	linkedin_skill	Skills of the Customer
23	ip_address	Customer's Device IP Address

Goal of the Project

We will analyze the dataset and segment customers using appropriate features for business to target its new product development.

Data Exploration: We will use summary, statistics and visualization tools to examine the following:

- Check for duplicates, nulls and NaN

- Cleaning the data (remove duplicates and replace null/NaN)

- Segregating features into categorical and numerical

- Univariate Analysis (number and distribution of unique groups within each categorical variables)

- Demographic Feature Analysis (title, gender, company_name, department, job_title, language, university, linkedin_skill)

    o Distribution of number of customers based on title, gender and department

    o Identifying top universities, companies and profession

- Geographic Feature Analysis (city, country, latitude, longitude, street_name, street_number, street_suffix, time_zone)

    o Identifying the top countries from where the majority of the customers belong

    o Visualizing the distribution of customer count across the globe

    o Finding a geography which dominate the customer population

Clustering: Find out the target audience set to focus for the business to maximize sales based on existing customer base.