Team Member Details

Group Name: Elizabeth's Analytics

Name: Elizabeth Banning

Email: estall@hotmail.com

Country: USA

College: Western Governors University

Specialization: Data Science

Business Understanding

XYZ is a bank that offers various services to its customers. The bank has collected data on its customers. The dataset includes customer demographics (age, sex, resident status, and so on). The dataset also includes which types of accounts and services the customer uses, such as mortgage, securities, and credit card services. XYZ Bank would like to offer Christmas promotions to its customers. However, different promotions may apply to different types of customers. For example, a deal on mortgage rates is unlikely to be used by elderly customers who are not interested in moving. To that end, the bank would like to group its customers into categories and develop a different set of promotions for each category. For the sake of efficiency, XYZ Bank would like no more than five groups of customers; with too many groups, the support required to develop and market many promotional campaigns will not be cost-effective. Understanding the characteristics of each customer segment will allow XYZ Bank to personalize its offers for each group, ensuring that customers in the group will be more likely to take up the promotional offer.

Problem Description

In order to develop its promotional campaign, XYZ Bank needs to know the answers to the following questions:

- What is the best number of groups to divide customers into?
- What are the primary characteristics of each group?

To answer these questions, the k-means clustering algorithm will be used to segment the customers, and the inertia metric will be used to determine the optimal number of groups (k). Finally, the characteristics of each group will be summarized so that XYZ Bank can determine which offers to develop and target to each group.

Project Lifecycle

The project is anticipated to take approximately seven weeks according to the following timeline:

- Week 1 (due November 18, 2022): Business understanding and problem description

- Week 2 (due November 25, 2022): Problems and anomalies in the data
- Week 3 (due December 2, 2022): Data cleaning approaches
- Week 4 (due December 9, 2022): Exploratory data analysis and recommendation
- Week 5 (due December 16, 2022): Presentation and recommended model
- Week 6 (due December 23, 2022): Model and results
- Week 7 (due December 30, 2022): Final presentation of results

Data Intake Report

Name: Customer Segmentation

Report date: November 14, 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Elizabeth Banning

Data intake reviewer: N/A

Data storage location: N/A

Tabular data details:

cust_seg:

Total number of observations	1,000,000
Total number of files	1
Total number of features	48
Base format of the file	.csv
Size of the data	154 MB

Proposed Approach:

Duplicate values will be investigated using Python's .duplicated method.

Also note that upon reading the file into pandas, 16 columns were found to contain multiple datatypes. These columns will be investigated and the datatype will be specified for each of these columns.

Github Repository Link

ebanning/DataGlacierProject: This is a customer segmentation project for the DataGlacier Data Science virtual internship. (github.com)