Dataset Link:

The dataset used in the project can be found on Kaggle:

https://www.kaggle.com/datasets/sergeymedvedev/customer_segmentation

Description of Project:

The project aims to predict the customer lifetime value for an insurance company based on historical data. Customer lifetime value (CLTV) is a measure of the total worth of a customer to a business over the entire period of their relationship. It is an important metric for insurance companies as it helps them to optimize their marketing and customer acquisition strategies, as well as customer retention efforts.

A brief explanation of the outputs of the Predicting customer lifetime value Project

1. Data preprocessing: The code loads and preprocesses the data, which includes handling missing values, dropping unnecessary columns, and converting categorical variables to numerical variables.

2. Exploratory data analysis: The code performs exploratory data analysis to gain insights into the data. It generates various visualizations such as scatter plots, histograms, and box plots to understand the distribution, trends, and relationships between variables.

3. Feature engineering: The code performs feature engineering to create new features that can be used for modeling. It calculates features such as monetary value, recency, frequency, and tenure.

4. Model building and training: The code builds and trains a machine learning model to predict customer lifetime value using the xgboost library. It uses GridSearchCV to tune the hyperparameters of the model.

5. Model evaluation: The code evaluates the machine learning model using various performance metrics such as RMSE (Root Mean Squared Error), R2 score, and Mean Absolute Error. It also plots the actual and predicted values on a graph to compare the performance of the model.

6. Prediction: The code makes predictions on the testing data using the best-performing machine learning model.

7. Conclusion: The code interprets the results and discusses the strengths and weaknesses of the model. It provides suggestions for future improvements.

Description of Output:

The main output of the project is a machine learning model that can predict the customer lifetime value of new customers. The model takes in customer transaction history, demographic data, and purchase behavior as inputs and produces the predicted customer lifetime value as an output. The model can be used by e-commerce businesses to identify high-value customers and tailor their marketing efforts accordingly. The project also provides several evaluation metrics that measure the performance of the model, such as RMSE and R^2, which can be used to compare the performance of different models or to assess the model's accuracy. The project also includes several data visualizations that help to understand the data and the performance of the models.

Instructions on How to Run the Code/Project/File:

1. To run the code in the notebook, you will need to have Jupyter Notebook installed on your computer. Once you have installed the Jupyter Notebook, you can download the notebook from the Kaggle website and open it in the Jupyter Notebook.
2. Before running the code, you will need to make sure that you have downloaded the necessary data files and saved them in the correct directory. The notebook provides instructions on how to download the data files and where to save them.
3. Once you have downloaded the data files and opened the notebook in Jupyter Notebook, you can run each cell of the notebook by clicking on the cell and then clicking the "Run" button in the toolbar or by using the keyboard shortcut "Shift + Enter".
4. It is recommended that you run the code cells in order, as some cells depend on the output of earlier cells. The notebook also provides explanations and comments for each code cell, so beginners can follow along and understand the code.

Note: Make sure to update the file paths in the code cells to match the location of the downloaded dataset and kernel files on your local machine.