Jaime Aguilar

Rajib Biswas - Mentor

Springboard

Data Science Career Track

Capstone Final Report

KPI's of Consumer Electronics

Problem Statement

How may KPIs such as Lifetime Customer Value, and Cost per Customer Acquisition be determined for consumer electronics, so that
a Company may fully characterize and understand its customer base for driving greater growth in its marketing purposes?

Dataset 1

This dataset is synthetic and made for educational and hypothetical purposes to document consumer electronics sales alongside consumer characteristics, for an intended business intelligence analysis that aims to seek inferences from its consumers.

Data Wrangling Dataset 1

This dataset was released in 2024. In the product ID column, there are thousands of consumer electronics sales. There were no duplicate rows and there were no missing values so overall this dataset was fairly clean from the start.

Exploratory Data Analysis Dataset 1

The exploratory data analysis aims to draw inferences about consumer characteristics and their purchasing behavior of consumer electronics, to get a better understanding of the consumer themselves. Let’s start with seeing which product brands have the highest purchase frequency.

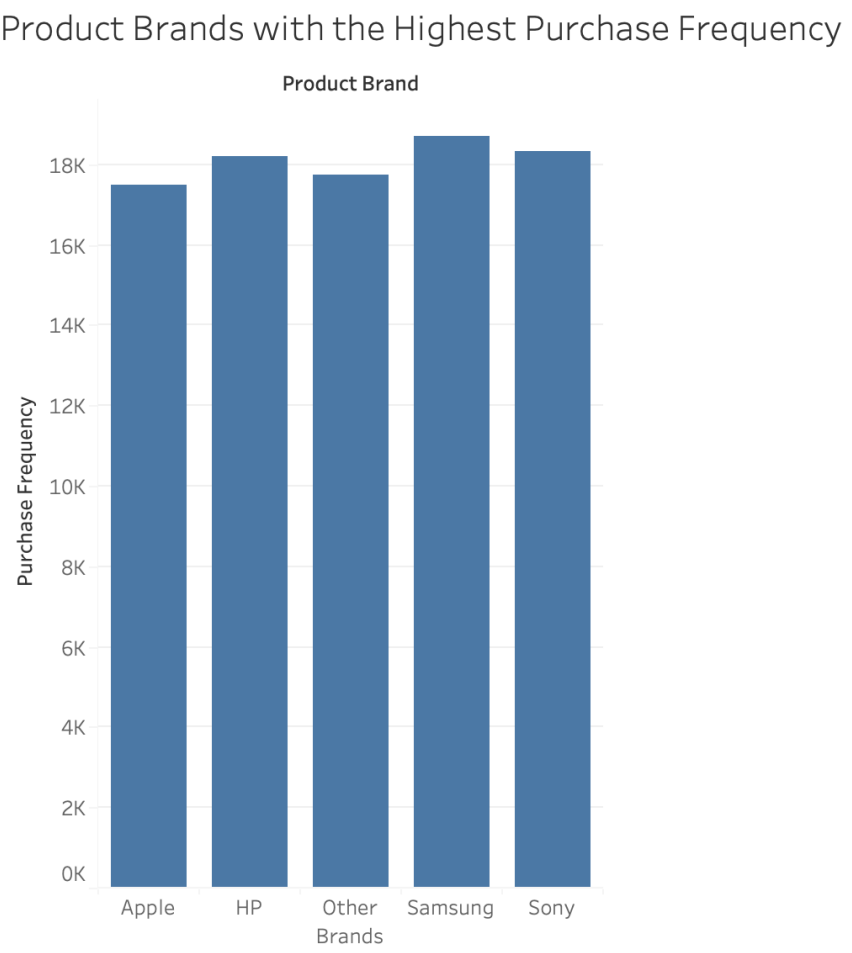

Figure 1. shows the consumer electronic brands with the highest purchase frequency.

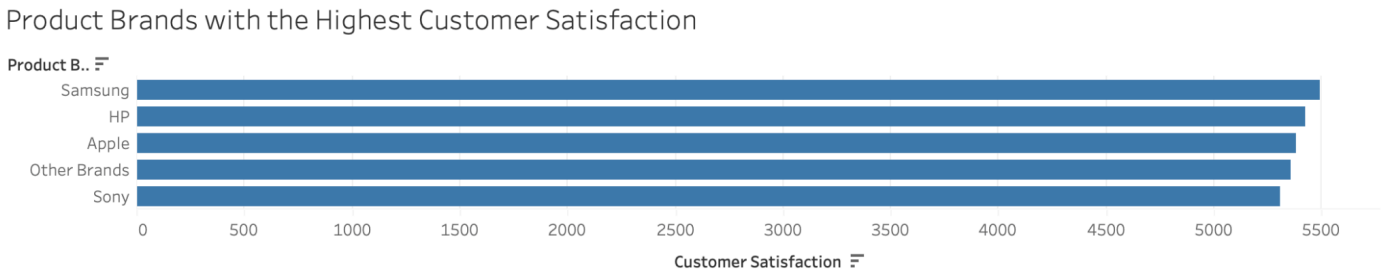

Figure 2. Shows which consumer electronic brands have the highest customer satisfaction.

It appears that not only is Samsung the most frequently purchased brand of electronics, but they also have the highest customer satisfaction ratings. Inferentially, consumers tend to purchase brands that they are highly satisfied with. Shockingly, brands like Apple and every other brand have higher customer satisfaction than Sony which is higher in purchase frequency than Apple. This points to the inference that customers also tend to purchase brands that they are loyal to, despite other brands (like Sony) grossing more sales than Apple or the other brands. Now let's look at the most sold product categories and the age groups that tend to buy certain products

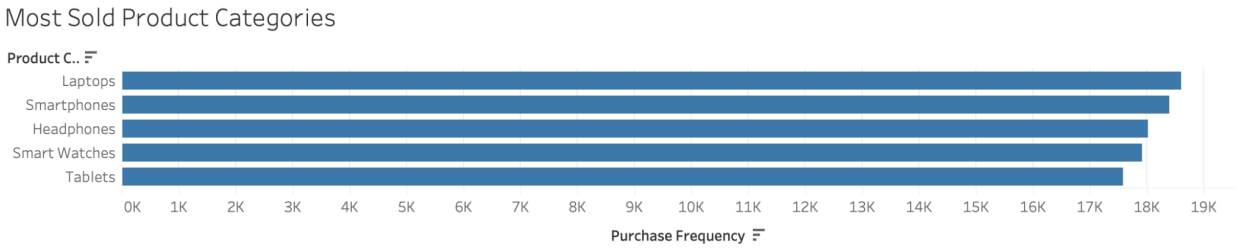

Figure. 3 shows the most sold consumer electronics categories

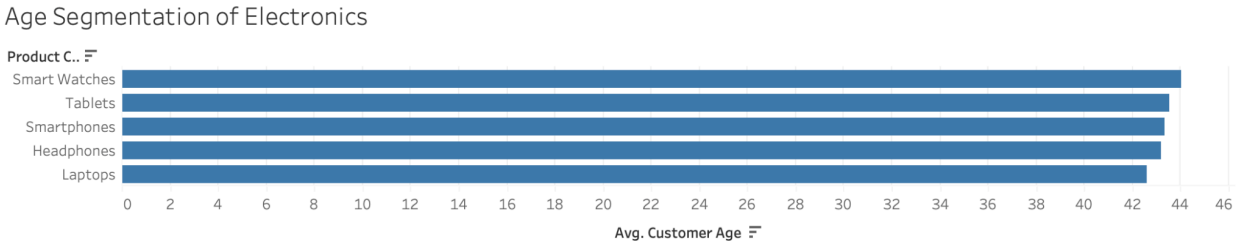

Figure 4. Shows the average age of consumers that tend to purchase a product category.

It appears that older consumers generally purchase smartwatches and tablets, while younger consumers purchase laptops and headphones. The most sold products are Laptops, smartphones, and headphones. Inferentially, younger people tend to make up the majority of consumer electronic purchases.

Preprocessing and Modeling Dataset 1

Certain features of dataset 1 where one hot encoded, such as product_category and product_brand. Several features were also standardized so the analysis may all be on the same scale. Two machine learning algorithms aimed to predict the KPI of Lifetime Customer value. The machine learning algorithms used were a Linear Regression model and a Random Forest. Let's look at the performance metrics of both models.

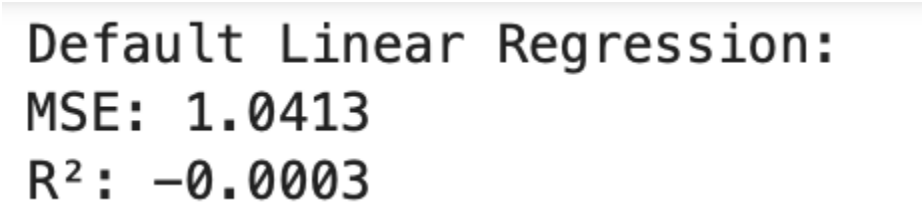

Figure. 5 shows the performance metrics of the Linear Regression of the consumer electronics dataset

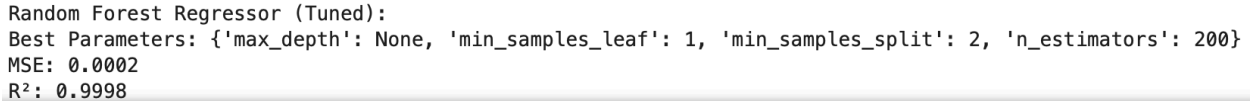

Figure 6. Shows a Random Forest regressor model with great metrics.

It appears that the random forest regressor, with hyperparameter tuning, performs far better than just a linear regression. The Mean squared error is almost perfect, and the R-squared shows high accuracy in being able to predict the variance within the dataset.

Now let's go ahead and make some numerical predictions of what the Lifetime Customer Value is for each category of electronics, based on how much our client should expect to make for the entire category of consumer electronics such as a smart watch or laptop.

AssignedCategory	Actual LCV (Unscaled)	Predicted LCV (Unscaled)
Laptops	2.028258e+08	2.027701e+08
Smart Watches	1.901177e+08	1.900759e+08
Smartphones	1.832840e+08	1.831955e+08
Tablets	2.021540e+08	2.021617e+08

Figure. 7 Shows Lifetime Customer Value KPI for each entire category in the millions.

Dataset 2

ABC Electronics is a hypothetical electronics e-commerce company. This dataset is synthetic and made for educational and hypothetical purposes to document electronics sales alongside with consumer characteristics, for an intended business intelligence analysis that aims to seek inferences from its consumers.

Data Wrangling Dataset 2

Dataset 2 was very clean from the start, it had no outliers, duplicates or null values. Data Wrangling dataset 2 was mainly held to fact-checking how "unclean" this dataset was, but this dataset was very clean from the start so there were little to no changes that needed to be made.

Exploratory Data Analysis Dataset 2

It was good to know a few basic things about this dataset, so let's start with finding out what are the lowest cost customer acquisition channels.

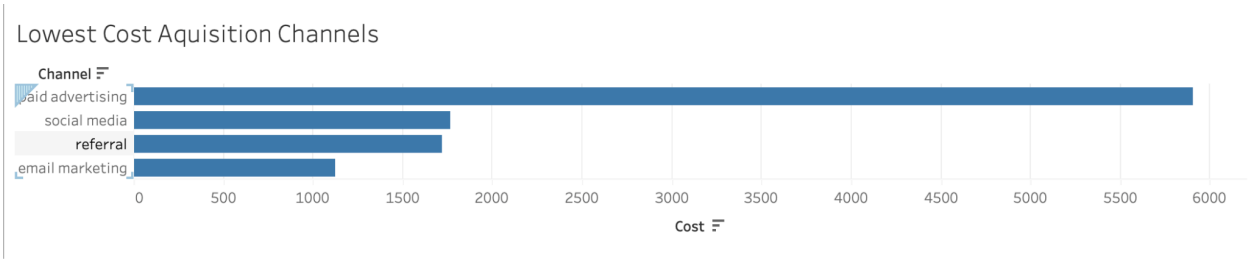

Figure. 8 shows the lowest cost acquisition channels being email marketing and referrals.

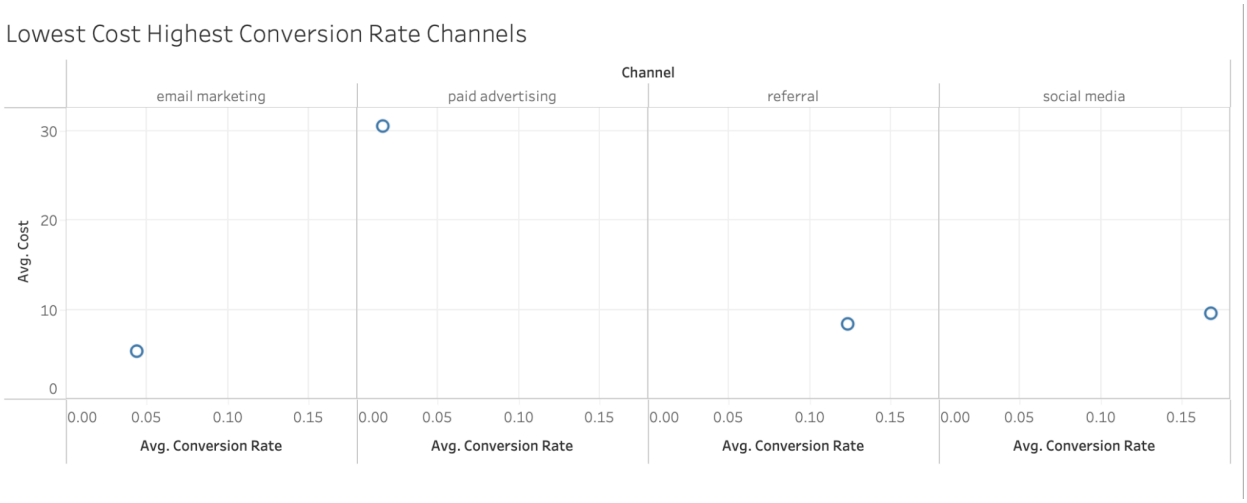

Figure. 9 shows The lowest cost and highest conversion rate channels in customer acquisition.

It appears that referrals have the highest conversion rate with the lowest cost and that social media is a close second in the cost-to-conversion rate ratio. Inferentially based on the two previous graphs, referrals and social media are a great way to acquire customers for a low cost. Thus, it is more beneficial for the company to invest more in these channels because of the high conversion rates and low costs that pertain to driving great business operations.

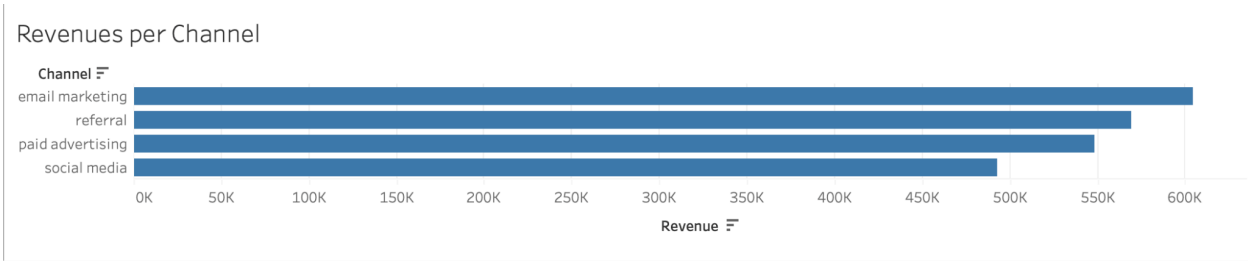

Figure. 10 shows revenue per channel being that email marketing has the highest revenue per channel.

Email marketing has the highest revenue per channel, but as you can see in Fig. 8 email marketing has a low cost and low conversion rate, which shows that it is truly not an effective strategy that leads to higher conversions for customer acquisition, despite the fact that email marketing is the highest revenue per channel. Referrals are the second highest revenue per channel, and they have a high conversion rate, so inferentially this company can invest greater resources to incentivize current to past clients for referrals to foster low-cost ways for a lot of customer acquisitions.

Preprocessing and Modeling Dataset 2

The Customer Acquisition Cost KPI (CAC) was calculated in this dataset, and several machine learning models were utilized to calculate this KPI. The dataset was made acceptable to the following machine learning models. A Random Forest Regressor, an XGBoost, and a gradient-boosting regression model. They were compared and contrasted with hyperparameter tuning to select the model with the best metrics. Figure. 10 below shows the results of running models on data set 2.

```
Model Performance (CAC Prediction) in Dollars:
MSE: 0.00 dollars²
R²: 1.0000
Cross-Validation MSE in Dollars: 0.00
Best Parameters from RandomizedSearchCV: {'model__xgb__n_estimators': 200, 'model__xgb__learning_rate': 0.01, 'model__rf__n_estimators': 100,
'model__rf__max_depth': 3, 'model__gb__n_estimators': 200, 'model__gb__max_depth': 5}
Future CAC Predictions (First 10 in Dollars): [  67.56529433   56.96161856   56.96161856   67.56529433 1863.37502211
   56.96161856  119.71694365  119.71694365   56.96161856   67.56529433]
```

Figure. 11 shows the results of a well-running Customer Acquisition Cost (CAC) model.

```
count     800.000000
mean      514.547619
std       764.026287
min        56.961619
25%        67.565294
50%       119.716944
75%       119.716944
max      1863.375022
Name: cac, dtype: float64
```

Figure. 12 shows summary statistics of the CAC for this dataset.

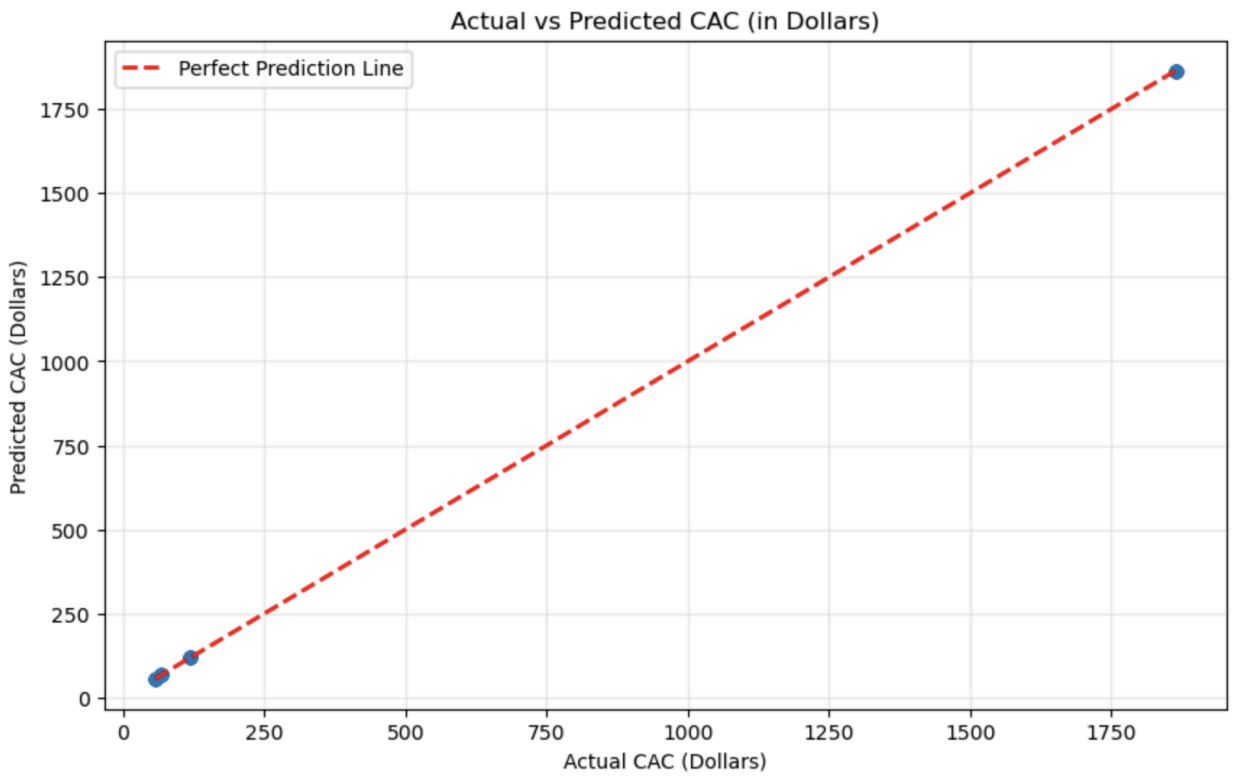

Figure. 13 Shows a perfect performance on a model that predicts the CAC KPI.

Perfect prediction certainly is a sign of concern, but is the model faulty or does it show data leakage? Technically, this is not data leakage. Data leakage occurs when the model has access to information during training that it would not have in real-world deployment or when predicting unseen data. We have a great model that perfectly predicts the CAC because CAC is a deterministic formula which is CAC = cost / conversion_rate. This model does overfit the data to a certain degree, but the KPI CAC is a very deterministic formula that explains why the model predicts so perfectly this KPI on this dataset.

Recommendations

Referrals and social media have the lowest cost and highest conversion rate of customer acquisition where the range of customer acquisition cost ranges from $56 to $1,800 which is several orders of magnitude higher than the lowest CAC. It is recommended that our company incentivizes their clients with rewards so that they may send over more referrals at a low cost to us. Social media is its own unique kind of channel which includes organic and non-organic means of outreach and customer acquisition so it is also recommended that our audience base is also incentivized to purchase our company's electronics with low-cost incentives that lead to more sales.

Younger people tend to buy more laptops and smartphones, and laptops have the highest lifetime customer value and are the most sold category of consumer electronics. Younger people tend to use social media more as well. Apple sells fewer products than Sony, but Apple has higher customer satisfaction than Sony. The highest customer satisfaction brands are Samsung and HP. It is recommended to target the younger generation demographic and to market more specifically laptops with high satisfaction such as Apple and HP to drive a fanatic loyalty that will lead to greater lifetime customer value and revenue generated by our company.

Future Work

In the future, a closer look at customer acquisition channels for the younger demographic can be expanded on. The younger generation is the driving force of the highest lifetime customer value categories of sold consumer electronics, so targeting channels such as social media (which has a very low Cost of Customer Acquisition) or perhaps holding seasonal sales in locations or areas where they tend to be around, such as college towns, will be an even greater way to reduce operational costs and increase revenue in a meaningful but mutually beneficial way.

MLA Works Cited

1. Biswas, Bhanupratap. *Customer Lifetime Value Analytics Case Study*. Kaggle, https://www.kaggle.com/datasets/bhanupratapbiswas/customer-lifetime-value-analytics-case-study. Accessed 3 Dec. 2024.

2. Elkharoua, Rabie. *Consumer Electronics Sales Dataset*. Kaggle, https://www.kaggle.com/datasets/rabieelkharoua/consumer-electronics-sales-dataset. Accessed 3 Dec. 2024.