**BUSINESS ANALYST REPORT**

**CONTEXT**

**PROJECT 1**

In our first project, the aim was to study the advertisement strategies and the cost of acquisition of customers. The idea was to optimize media strategy to minimise the acquisition cost. This included minimizing media campaign costs, targeting customer segments that were identified as having the lowest acquisition cost, and maximizing customer reach.

By making use of data analysis and machine learning solutions, we were able to identify the most important factors that influenced the cost of acquiring a customer. The main business insights gathered were:

-       The biggest factor that increased the cost of acquisition of customers was the **ratio of the space dedicated to meat in a convenience store**. The larger the meat area, the higher was the customer acquisition cost. This allowed us to conclude that meat can be categorised as an expensive product and a store selling a lot of meat is therefore more inclined to serving customers with more money.

-       Depending on the **media type used for marketing**, the type of store and even the gender the cost varied a lot. Having this knowledge was useful to allow us to better target our customers. Our study showed, for example, that daily paper and the radio were the advertisement media preferred by most customers, well ahead of the TV or the bulk mail, for example.

**PROJECT 2**

Based on the above, the store chain put in place some practices that successfully enabled them to:

-       **Optimize their market spending**: By predicting the cost of acquiring a customer, the store chain was able to better allocate its marketing spend and invest in channels and campaigns that are most likely to bring in new customers at the lowest cost. This helped maximize the return on investment for marketing activities and improve overall profitability.

-       **Increase customer lifetime value**: By predicting the cost of acquiring a customer and optimizing customer acquisition strategies, the store chain was able to attract new customers who are more likely to become loyal, high-value customers. This helping to increase the lifetime value of customers and improve overall customer retention and loyalty.

The business has hence realized the importance of gathering such data for analysis and modelling and is considering the expansion of their data base in the future by establishing an ongoing influx of data generated by new customers.

To do so, they realized that they may need to first **revisit their data management processes** to ensure that the data will be properly stored and cleaned. This would involve developing a proper data architecture, data storage systems, and data processing workflows.

Changing the data management processes can also lead to improved efficiency and cost savings, such as by reducing the time and resources required for data processing and analysis. This can involve implementing new data automation, as well as streamlining data workflows to ensure that data is processed and analysed more quickly and accurately.

In addition, since the former model developed led to significant positive impacts on the business, they are considering its **improvement** by employing methods such as hyperparameter tuning. The **use of automated tools and algorithms to automate the process of building and optimizing machine learning models** is also another point to consider. AutoML tools, for example, can automatically perform tasks such as data pre-processing, feature engineering, model selection, hyperparameter tuning, and ensembling.  By simplifying and accelerating the machine learning process, it becomes easier for businesses and individuals with limited technical expertise to build and deploy machine learning models.

**OBJECTIVES**

We can therefore list our objectives for this project as:

* **Automation of data processing tasks**, so that the store can reduce the manual effort required for data processing. By automating data preparation and preprocessing, the store can free up resources to focus on other high-value tasks, such as data analysis and business decision making.
* **Scalability, to handle large volumes of data**. This is important for stores that have many transactions, customers, and products. By building scalable data pipelines, the store can ensure that it can handle increasing data volumes without compromising performance.
* **Improved accuracy of predictions** from the machine learning model, so that the store can more reliably estimate the cost of acquiring new customers. This can help the store to better allocate its marketing budget and resources, and ultimately improve the return on investment (ROI) of its marketing efforts.
* Obtain an even **better understanding of the factors that influence customer acquisition cost**, such as customer demographics and behavior. By understanding these factors, the store can tailor its marketing strategies to better reach and engage its target audience, and ultimately acquire new customers more efficiently.

**KPIs**

To track the performance of the machine learning project, and ensure that it is generating value for the business, the following KPIs were determined:

1. **Cost of Customer Acquisition:** This will measure the total cost of acquiring a new customer.
2. **Accuracy of Predictions:** This will measure how accurate the machine learning model is at predicting the cost of acquiring a new customer.
3. **Cycle Time:** This will measure the time required to complete a process or task.

**OUTCOME**

**RESULTS**

**Modelling**

After having tried out a variety of methods to achieve the objectives listed above, the data team was able to yield the following results:

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Improve previous model** | **New model** | **Accuracy** |
| AutoML |  | x | RMSE = 1.407 |
| TPOT |  | x | **RMSE = 0.837** |
| HyperOpt | x |  | RMSE = 19.79 |
| Randomized Search | x |  | RMSE = 19.73 |
| Azure Automated ML |  | x | RMSE = 1.276 |
| Optuna | x |  | RMSE = 18.24 |

As seen above, building a new model gave better results than trying to improve the previous one.

The business decision of keeping the **TPOT** model was made, due to its highest accuracy of predictions, making in fact a very significant improvement on the model used in the previous project.

**Architecture**

The developed cloud architecture outlines a solution for a consistent, reliable machine learning framework. When the best model is ready for production, Azure Databricks deploys that model to the MLflow model repository. The storage layer Delta Lake and the machine learning platform MLflow also play significant roles. ​

Process Duration (from raw data to predictions): ~**52 minutes**

Estimated cost: **$0.30/hr** (when running)

Virtual machine sizeStandard\_DS12\_v2 (4 cores, 28 GB RAM, 56 GB disk, CPU)

**INSIGHTS & STRATEGIES**

**INSIGHTS**

**Feature Importances**

From the TPOT model, we found that **promotion name** is a very important predictor. This makes sense as the promotion names that convey a sense of value, for e.g., the words "free" or "super savers", must have been much more effective at attracting new customers and driving customer acquisition.

**Media type**, on the other hand, retains its importance (as obtained from our previous model). Through this project version, we found that cash register handouts tend to have a significant influence on the cost of acquiring new customers.

We now need to dive deeper into these insights to understand the way (whether positive or negative) these important features influence the acquisition cost.

**SHAP Values**

The following insights were obtained from the SHAP values graph generated for the TPOT model:

* Customers with a media type of "Cash Register Handout" or "Street Handout" have a higher acquisition cost in general. This makes sense, probably because of paper and printing costs incurred by producing these handouts. It is worth mentioning that high values of "Cash Register Handout" significantly increase customer acquisition cost while low values only slightly decrease it.​
* Customers acquired through "Price Slashers", "Two for One", or "Double Down Sale" have a lower acquisition cost.​

**STRATEGIES**

The business can leverage this newfound insight by employing these new strategies to optimize its marketing campaigns and drive customer acquisition more effectively:

* **Leveraging emotional appeal**: By creating promotion names that evoke positive emotions or a sense of excitement or anticipation, customers can feel motivated to engage with the store and hence drive customer acquisition.
* **Personalization**: By tailoring promotion names to specific customer segments or channels, promotions can be rendered more relevant and effective at driving customer acquisition. For example, the store could use different promotion names in its email campaigns compared to its social media campaigns to better target each audience.
* **Timing**: By carefully scheduling promotion names to coincide with specific seasons, holidays, or events that are relevant to the store's target audience, the store can capture customer attention and motivate them to act.

Regarding the media type, the business can still carry on with their previous successful practices (such as focusing on the daily paper and the radio) regarding this feature to reach new customers at a low cost. But in addition to these, they can also decrease handouts at the cash register and on the street since these methods are most likely incur extra costs when trying to acquire new customers.

**CONCLUSION**

The business can now expand their customer database without worrying about the efficiency of their data management processes or the scalability of their data architecture with regards to increasing volumes of data by leveraging the cloud solution.

Furthermore, the new model yielded promising results and allows the business to predict the cost of acquiring new customers more accurately. It also enables the business to come up with new strategies to further optimize their market spending and acquire new, loyal customers at low costs.