Weekly Churn Prediction Campaign Guide

Purpose: Help the marketing team understand, interpret, and act on churn prediction results.

Notebook: **20703562_ML_CW_MODEL _Implementation.html**

For Demonstration purpose,the latest churn prediction is already done for after 2022-03-22 ie between 2022-03-23 and 2022-04-21,is done in the notebook.

-Data Preprocessing:

> The dataset passed all structural quality checks and was ready for feature engineering and modeling without additional cleaning.
>
> > **Customer database:** 584/4312 'null' value of customer first name ('first') and date of birth ('dob'). Due to that, no direct analysis on customer profile was included in main report.
> >
> > **Instruction for Use** If any nulls or duplicates are found in future analysis, filter them out or investigate their cause before continuing with analysis.

1. What This Model Does

- Predicts which customers are most likely to churn (stop engaging or purchasing).
- Assigns each customer a churn probability (from 0% to 100%).
- Groups customers into rankings (1 = highest risk, 10 = lowest risk).
- Flags the top-risk groups (ranking 1 and 2) as targets for retention offers.

Automation Note

- The churn prediction process is fully automated.
- No manual input is required from the marketing team during weekly runs.
- **The only actions needed are**:
    - Load the input in the mentioned places(recent customer data)
    - Loading the target list (**`weekly_offer_target_list.csv`**) for campaign use.
    - **Optional tweaks to:**
        - Change the number of customer groups targeted (e.g., expand from ranking 1–2 to 1–3).
        - Align with new campaign goals or capacity.

The model runs weekly, updates all metrics and logs automatically, and flags customers based on the latest data.

2. Key Output: `weekly_offer_target_list.csv`

- Contains only customers flagged for offers based on high churn risk.
- Columns include:
    - `customer_id`: Unique customer identifier.
    - `churn_percent`: Likelihood of that customer churning.
    - `ranking`: Risk group (1 = highest risk).
- Sorted by highest churn risk first.

3. Summary Insights (Automatic)

- Shows how many customers are being targeted this week.
- Displays average churn risk for the targeted group.
- Provides segment-level summaries to understand group behavior.

4. Action for Marketing Team

1. Review the `weekly_offer_target_list.csv` file.
2. Launch retention campaigns (email/SMS/discount/etc.) for those customers.
3. Prioritize customers with the highest churn_percent.
4. Use `offer_sent_on` to ensure campaigns are linked to the correct week.

5. Data Access

To retrieve the current week's target list from the system:

```
SELECT customer_id, churn_percent, ranking, offer_sent_on

FROM weekly_churn_scoring_log

WHERE offer = 1

ORDER BY churn_percent DESC
```

Notes

- The "offer" flag is currently based on ranking 1 and 2 — this can be adjusted based on campaign size or strategy.
- Weekly predictions are updated automatically, with results logged in `weekly_churn_scoring_log`.

3. Figure and Table Mapping:

Report Element / Figure	Source
***Figure 3**. Target Class Proportions vs. Churn Definition Threshold*	ConsultingCorp_Report_ml76.pdf
***Figure 4.** Cumulative Distribution of Days Between Customer Visits*	ConsultingCorp_Report_ml76.pdf
***Figure 1**. Distribution of Active vs Inactive Customers*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
***Figure 2.**Churn vs Non-Churn Customer Counts Predicted*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
***Figure 5**. Windowing Strategy used for tempo feature construction*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
***Table 1.** Features generation approach*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
***Figure 6.**Correlation between input and output feature*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html

***Figure 7*.** *Top 10 Feature Importance Scores (Permutation Method)*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Table 2. *Final Feature sets*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
***Figure 8*.** *Model evaluation strategy*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Table 3. *Performance Comparison of Classifiers Across Key Metrics*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure *9.* *ROC Curve Comparisons*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 10. *Precision and Recall vs. Classification Threshold*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 11. *Confusion Matrix on Held-Out Test Set*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Table 4. *Summary Statistics of Churners vs non-Churners (30-Day Means)*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 12. *Customer Distribution by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 13. *SHAP Summary Plot*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 14: *Distribution of Frequency Variance by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 15: *Distribution of Spend Variance by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 16: *Customer Store Usage by Churn Label (Stacked)*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 17: *Average Basket Value (Capped) by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 18: *Average Number of Products by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html
Figure 19: *Average Number of Visits by Churn Risk Rank*	Databricks-notebook_20703562_ML_CW_MODEL_EVALUATION.html