MGMT 467 – Unit 2 Lab 2: Prompt Studio for AI-Assisted SQL + ML

This handout guides you step-by-step through constructing SQL queries via AI prompts to explore, engineer, and model churn prediction in BigQuery.

Task 0: Connect to BigQuery

Verify your BigQuery access and environment setup.

- Use %%bigquery in a Colab cell.
- Prompt example: 'Write a SQL query to return today's date and current user session.'
- Expected SQL: SELECT CURRENT_DATE() AS today, SESSION_USER() AS user;
- Checkpoint: Query returns one row with today's date and your user ID.

Task 1: Prepare ML Table

Create a modeling table with clean features and a churn label.

- Select columns like region, plan_tier, age_band, avg_rating, total_minutes, churn_label.
- Filter: churn_label IS NOT NULL.
- Prompt: 'Create a BigQuery table named churn_features from cleaned_features with selected features.'
- Checkpoint: Table exists and has non-null churn labels.

Task 2: Train Logistic Regression

Use BigQueryML to create a logistic_reg model.

- Use churn_features table.
- Prompt: 'Train a logistic regression model to predict churn_label from behavior features.'
- Checkpoint: Model appears in BigQuery; training completes.

Task 3: Evaluate Model

Use ML.EVALUATE to assess model performance.

- Metrics include accuracy, precision, recall, log_loss.
- Prompt: 'Evaluate my churn_model using ML.EVALUATE.'
- Checkpoint: Metrics appear in results table.

Task 4: Predict Churn

Generate predictions using ML.PREDICT.

Prompt: 'Predict churn using churn_model with user_id and probabilities.'

Checkpoint: Verify top users with highest churn risk.

■■ Feature Engineering Tasks

Task 5.0: Bucket a Continuous Feature

Bucket total_minutes into low, medium, high.

Prompt: 'Create a column watch_time_bucket based on thresholds.'

Explore: Churn rate by watch_time_bucket.

Task 5.1: Create Binary Flag

Flag users with total_minutes > 500.

Prompt: 'Create a binary column flag_binge based on total_minutes.'

Explore: Do binge-watchers churn more?

Task 5.2: Create Interaction Feature

Combine plan_tier and region.

Prompt: 'Create column plan_region_combo using CONCAT.'

Explore: Which plan-region combos are riskiest?

Task 5.3: Add Missing Value Flags

Flag missing age_band, avg_rating.

Prompt: 'Create column is_missing_age = 1 if age_band IS NULL.'

Explore: Are missing values predictive?

Task 5.4: Time-Based Feature

Add days_since_last_login.

Prompt: 'Create a column using DATE_DIFF with last_login_date.'

Explore: Recency of login vs. churn.

Task 5.5: Assemble Enhanced Feature Table

Combine all engineered features.

Prompt: 'Create churn_features_enhanced with all new and base columns.'

Checkpoint: No data loss; all features present.

Task 6: Retrain Model

Use engineered features in a new logistic_reg model.

Prompt: 'Train churn_model_enhanced using all enhanced features.'

Compare model metrics to base.

Task 7: Compare Performance

Compare churn_model vs. churn_model_enhanced.

Prompt: 'Evaluate both models and compare precision, recall, log_loss.'

Explore: Which features helped most?

■ Chain-of-Thought Exploration Prompts

Use these to guide deeper insights and reflection after each task:

- What assumptions does your model make? Are they valid for your data?
- Which features have the strongest correlation with churn?
- Did engineered features improve performance? Why or why not?
- Were there any surprises in your predictions?
- If you were Netflix’s retention team, what policy changes would you recommend based on your findings?