# Student Instructions: Lab 5 - Advanced Intelligence

Module: 5 | Time Estimate: 3 Hours

Role: Data Scientist

Tools: BigQuery ML, Google Cloud AI

### The Scenario

**RetroRetail Inc.** has a problem. Their supply chain team is flying blind.

* **The Quantitative Problem:** They buy inventory based on "gut feeling" rather than data. They need a forecast for next month's sales.
* **The Qualitative Problem:** Customers are leaving bad reviews, but the team can't read all 50,000 comments. They need to know *why* customers are unhappy.

**Your Mission:**

1. **Predict (Regression):** Use BigQuery ML to build a **Linear Regression** model that predicts next month's revenue based on historical trends.
2. **Analyze (Sentiment):** Use SQL string logic to parse unstructured customer reviews and extract "Sentiment" (Positive/Negative).
3. **Report:** Deliver a "Forecast vs. Reality" report to the Board.

### Part 1: The Prediction (Forecasting Revenue)

*We will use machine learning to predict the future.*

1. **Prepare Training Data:**
   * Navigate to BigQuery.
   * Create a View named view\_training\_data from your sales\_partitioned table (from Lab 4).
   * It must aggregate data by date: Select date, SUM(revenue) as total\_revenue.
2. **Train the Model (Create Model):**
   * Run the following SQL to "teach" BigQuery how to predict revenue based on the date.

SQL  
CREATE OR REPLACE MODEL `retro\_raw.model\_revenue\_forecast`  
OPTIONS(model\_type='LINEAR\_REG', input\_label\_cols=['total\_revenue']) AS  
SELECT  
 date,  
 total\_revenue  
FROM `retro\_raw.view\_training\_data`  
WHERE date < '2025-06-01'; -- Training on historical data only

* + *Note:* This takes ~1-2 minutes to run.

1. **Evaluate the Model:**
   * Run SELECT \* FROM ML.EVALUATE(MODEL retro\_raw.model\_revenue\_forecast, ...) to see how accurate your model is.
   * *Look for:* **Mean Absolute Error (MAE)**. This tells you, on average, how many dollars "off" your prediction is.
2. **Take Screenshot #1:** The evaluation results showing the MAE.

### Part 2: The Sentiment (Unstructured Data)

*The client sends you a new file: Customer\_Reviews.csv. It contains raw text comments.*

1. **Ingest:** Download Customer\_Reviews.csv (provided by the professor on Canvas), upload it to GCS, and load it into BigQuery as reviews\_raw.
2. **Analyze (The "Hack" Way):**
   * Real GenAI calls cost money. For this lab, we will simulate a "Rule-Based" AI model using SQL string logic.
   * Write a query that scans the review\_text column.
   * **Logic:**
     + If text contains 'bad', 'broken', 'slow' --> Sentiment = 'Negative'.
     + If text contains 'love', 'great', 'fast' --> Sentiment = 'Positive'.
     + Else --> 'Neutral'.
3. **Take Screenshot #2:** A query result showing the Review Text next to your calculated Sentiment column.

### Part 3: The Forecast (Predicting the Future)

*Now we use the model to generate new data.*

1. **Generate Future Dates:** Create a temporary table (or CTE) with dates for **July 2025** (dates the model has never seen).
2. **Predict:**
   * Use ML.PREDICT to generate revenue numbers for these future dates.

SQL  
SELECT \*  
FROM ML.PREDICT(MODEL `retro\_raw.model\_revenue\_forecast`,  
 (SELECT date FROM `retro\_raw.future\_dates`));

1. **Take Screenshot #3:** The result showing the predicted\_total\_revenue for July 2025.

### The Deliverable: Client Report #5

Submit a single PDF named LastName\_Lab5.pdf:

1. **The Forecast (Screenshots):**
   * Image of your Model Evaluation (MAE).
   * Image of your July 2025 Prediction results.
2. **The Sentiment (Screenshot):**
   * Image of your "Sentiment Analysis" SQL query results.
3. **Executive Summary:**
   * **Interpretation:** "Our model predicts revenue will be $X. Based on an MAE of $Y, we are confident in this number within a margin of error."
   * **Strategy:** "Sentiment analysis reveals that 'broken' is a top keyword. We recommend investigating shipping partners."