

BREAK  
THROUGH  
TECH

# Unwrapping Customer Delight

## Milestone #3 Meeting: The Analysis Phase

The Estée Lauder Companies

October 3, 2025



# Recap: Experimental Design

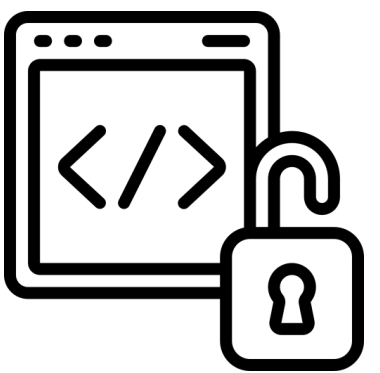
- **What is an RCT?**
  - Randomized Controlled Trial = split customers into two groups at random
  - **Control group (0)**: does *not* get the surprise gift
  - **Treatment group (1)**: *does* get the surprise gift
- **Why randomization matters:**
  - Makes groups *statistically comparable* on average (age, tenure, loyalty, spending history)
  - Prevents bias that could occur if we “cherry-picked” who gets gifts
- **Measuring the effect:**
  - Average Treatment Effect (ATE) = **difference in mean outcomes** between treatment and control
  - Formula:  $\Delta = \text{mean}(\text{revenue} \mid \text{treatment}=1) - \text{mean}(\text{revenue} \mid \text{treatment}=0)$
  - Example: If treated customers spend \$110 on average and control spend \$100, then ATE = +\$10
- **Big challenge:** Real data is noisy → harder to detect real differences without large enough samples





# Recap: Why Statistical Power Matters

- **Statistical power** = probability that our experiment correctly detects a real effect
- If power is too low → risk of *false negatives* (we conclude “no effect” when there actually is one)
- If sample size is too big → wasted resources, longer experiments, possible negative customer experiences
- **The 4 levers of power analysis:**
  - **Minimum Detectable Effect (MDE):** The *smallest* change we care about finding (e.g., +1% revenue lift)  
**Significance level ( $\alpha$ ):** Chance of false positive. Standard = 5%
  - **Power ( $1-\beta$ ):** Chance of catching a real effect. Standard = 80–90%
  - **Sample size (N):** How many customers we need in the experiment
- **How they interact:**
  - Smaller MDE → need larger N
  - Higher power (90% vs. 80%) → need larger N
  - Lower variance → smaller N is enough
- **Takeaway:** Designing experiments = balancing these levers to be efficient but still scientifically valid





# Recap: Variance Reduction w/ MLRATE

- **Why variance matters:**

- Outcome data (like revenue) has high natural variability
- This “noise” makes it harder to see the treatment effect

- **Solution: Regression adjustment (MLRATE)**

- Step 1: Use ML to predict expected revenue based on *pre-experiment covariates* (tenure, loyalty, prior spend, etc.)
- Step 2: Use those predictions ( **$G(x)$** ) as control variables in the treatment effect regression
- Step 3: Estimate treatment effect on the “leftover” variation (residuals) after accounting for predictable patterns

- **Key assumptions:**

- **$G(x)$**  is strongly correlated with revenue
- **$G(x)$**  is independent of treatment (since assignment is random)

- **Benefit:**

- Reduces unexplained variance → increases recall
- Means fewer customers are needed to detect the same effect

In practice: MLRATE is like turning down the “noise” so the gift signal comes through more clearly

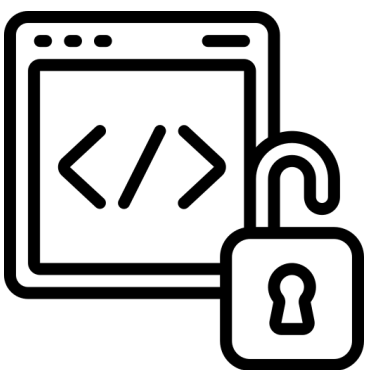






# What You Completed for Milestone #3

- You ran **two types of power analysis**:
  - 1. Standard t-test method**
    - Effect size =  $(\text{MDE}) \div (\text{std. dev. of revenue})$ .
    - Sample size based on *all variance* in the outcome
  - 2. MLRATE-adjusted method**
    - Effect size =  $(\text{MDE}) \div (\text{std. dev. of residuals})$ .
    - Sample size based on *reduced variance* (after regression adjustment)
- **Expected outcome:**
  - MLRATE gave you a *smaller required sample size*
  - Your residual distribution plot shows how much variance was explained by **G(x)**
- **Big takeaway:**
  - Power analysis tells us *how many customers* we need
  - MLRATE lets us run experiments more efficiently (same power, smaller N)





# Next Steps

## 1. Execution

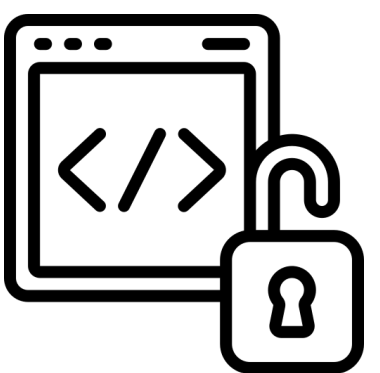
- We've determined the optimal sample size, given the data and business requirements
- Now, we apply our intervention. For example:
  1. For each user on the checkout page, randomly assign them into the treatment or control group.
  2. Offer a free gift to each customer in the treatment group.
  3. Repeat until we have reached the required sample size.
  4. Wait until  $t$  time has passed
  5. Collect the experiment data, wait until  $t$  time has passed:
    - a. Revenue: during the experiment period (i.e. time  $t$ ).
    - b. Covariates: The same ones we used in the power analysis, during the *pre-experiment period* (i.e. time  $t-1$ ).

**We will simulate the experiment execution phase and provide you with the above experiment data.**

## 2. Analysis

Now that we have the experiment data:

1. Conduct EDA to validate the data and get an initial, visual read of the treatment effect.
2. Estimate the treatment effect using MLRATE.





# Milestone #4: Moving to Experiment Data

- **What data you'll get** (via email):
  - File: `experiment_results.parquet`
  - Each row = a simulated customer in our experiment (aka gift giving strategy)
- **Columns included:**
  - `customer_id` → unique customer identifier
  - `aov` → average order value before the intervention
  - `days_since_last_purchase` → recency of last purchase (days)
  - `tenure_in_days` → how long the customer has been active
  - `loyalty_membership` → categorical/boolean flag for loyalty program
  - `assignment` → treatment assignment indicator (0 = control, 1 = treatment)
  - `revenue (t)` → post-intervention revenue (outcome variable of interest)





# EDA on Experiment Data

- **EDA goals:**
  - **Data quality checks:** row counts, datatypes, any missing values or duplicates?
  - **Univariate exploration:** histograms & boxplots for `revenue (t)`, `aov`, `tenure_in_days`, `days_since_last_purchase`.
  - **Treatment vs. control comparisons:**
    - Verify **covariate balance** across `assignment`
      - Expectation: Covariates should look *statistically similar*, since assignment was random
    - Compare distributions of `revenue (t)` between groups
      - This should give you an early visual sense of the treatment effect
  - **Relationships:** scatterplots (e.g., `aov` vs. `revenue (t)` colored by `assignment`), grouped box/violin plots (e.g., `revenue` by `loyalty_membership` × `assignment` to see heterogeneity of effects)
  - **Skew/outliers:** check for long-tail distributions
- **Meeting will also cover:**
  - A lecture on how to estimate the treatment effect (theory)





# Project milestones and timeline

These are the milestones for your Challenge Project. They include the [CRISP-DM](#) process steps you learned about in your ML Foundations course. In addition, there is an educational component in the front-end.

