A/B Testing Solution Design for Lead Utilisation Experiment

1. Preparatory Work

a. Define Hypothesis

- **Null Hypothesis (H0):** Mandating 4 leads per job does not lead to a significant improvement in lead utilisation.
- **Alternative Hypothesis (H1):** Mandating 4 leads per job significantly improves lead utilisation.

b. Define Key Metrics

- **Primary Metric:** Lead utilisation rate (number of accepted leads / number of available leads).

Potential Secondary Metrics	Rationale
Job Acceptance Rate (% of jobs with at least one accepted lead)	Indicates tradie engagement and platform efficiency. Higher rates suggest consumers are more likely to receive quotes, indirectly enhancing satisfaction.
Lead Saturation Index (Ratio of leads offered to leads claimed)	Identifies inefficiencies caused by oversupply of leads, helping evaluate the balance between supply and demand.
Revenue per Tradie Distribution (Distribution of revenue generated across tradies)	Ensures equitable revenue opportunities among tradies and identifies potential imbalances or disengagement risks.
Time to First Lead Claim (Average time from job posting to the first lead being claimed)	Indicates tradie responsiveness and engagement, which are crucial for consumer satisfaction and platform efficiency.

c. Sample Size Calculation

- Use a power analysis to determine the required sample size.
  - Inputs: Baseline lead utilisation rate, expected uplift (e.g., 5%), desired power (80%), and significance level ($\alpha = 0.05$).
- Ensure sufficient data collection period to achieve the required sample size.
- Use the sample size as a feasibility check for the experiment and to inform the statistical approach.

d. Data Preparation

- **Schema for Tracking Events:**
  - job_id: Unique identifier for each job.
  - group: A/B group identifier (Control or Test).
  - tradie_id: Unique identifier for each tradie.
  - lead_accepted: Boolean indicating whether the lead was accepted.
  - time_of_action: Timestamp of lead acceptance or job posting.

- **Event Taxonomy:**
    - **JobPosted**: Captures job details when a job is created.
    - **LeadOffered**: Captures details when a lead is sent to a tradie.
    - **LeadAccepted**: Captures details when a lead is claimed by a tradie.

e. Randomisation Plan

- Randomly assign jobs to two groups:
    - **Control Group:** Jobs with 3 leads (current setup).
    - **Test Group:** Jobs with 4 leads (new setup).
- Ensure randomisation is stratified by key factors like job category and location to account for variability.

f. Evaluate Statistical Approach: Frequentist vs. Bayesian

Frequentist Approach	**Bayesian Approach**
• Best suited for a clear, one-time decision (e.g., whether to roll out 4 leads permanently). • Provides p-values to determine statistical significance (e.g., whether the observed difference in lead utilisation rate is unlikely to occur by chance). • Requires a predefined sample size and does not incorporate prior knowledge or adapt during the experiment. • **Advantages:** Simple and widely understood. • **Limitations:** Does not quantify the probability of an uplift directly; relies on binary accept/reject outcomes.	• Best suited for ongoing decision-making and incorporating prior knowledge about lead utilisation rates. • Provides posterior probabilities for the likelihood of uplift, allowing for more intuitive interpretations (e.g., "There is a 95% probability that the new setup increases lead utilisation by at least 5%"). • Supports continuous monitoring and early stopping decisions. • **Advantages:** Flexible, provides richer insights, and adapts as data accumulates. • **Limitations:** More computationally intensive and may require additional expertise to implement.

- **Decision Criteria:**
    - Use **Frequentist** if: The goal is a straightforward, binary decision and the team is more familiar with p-values and traditional statistical tests.
    - Use **Bayesian** if: The goal is to gain probabilistic insights, adapt dynamically, or integrate prior knowledge into the analysis.
    - If unsure, consider piloting with both approaches on a smaller dataset to evaluate which provides more actionable insights for this specific experiment.

2. Technical Details of Execution

a. Experiment Implementation

- **Dynamic Lead Assignment:** The backend system dynamically assigns the number of leads as follows:
    - When a job is created, it is tagged with the randomly assigned group (Control or Test).
    - Based on the group:
        - If Control, the system retrieves and matches most relevant tradies and mandates a maximum of 3 leads.
        - If Test, the system retrieves and matches most relevant tradies and mandates a maximum of leads.
    - The relevance is determined using pre-existing job matching algorithms (e.g., based on location, category, and tradie activity).
- Ensure the assignment logic is logged for audit and troubleshooting purposes.

b. Data Tracking

- Implement event logging for JobPosted, LeadOffered, and LeadAccepted events.
- Store event data in a centralised analytics database.

c. Tooling and Infrastructure

- Use a cloud-based analytics platform (e.g., Snowflake or BigQuery) for data storage and analysis.
- Use Python or R for statistical analysis.
- Use a dashboarding tool (e.g., Tableau or Power BI) for real-time monitoring of metrics.

3. Measurement, Tracking, and Interpretation

a. Measurement and Analysis Plan

- Calculate incremental uplift across the previously defined primary and secondary metrics.
    - Identify and capture key data points:
        - Fields: job_id, group, tradie_id, lead_accepted, time_of_action.
        - Use timestamps to analyze time-sensitive metrics like time to first lead claim.
    - Ensure robust data validation:
        - Address missing values, outliers, and inconsistencies to maintain data quality.

Conduct statistical testing to gauge the confidence of the results based on the chosen approach.

Frequentist Approach	Bayesian Approach
• Conduct hypothesis testing (e.g., t-tests, chi-square tests) to evaluate statistical significance. • Use p-values to assess the likelihood of observed differences occurring by chance. • Final analysis performed after achieving the predefined sample size.	• Use Bayesian inference to estimate posterior probabilities and assess the likelihood of uplift. • Provide interpretable results (e.g., "95% probability that uplift exceeds 5%"). • Enable continuous monitoring and adaptive decision-making.

b. Interim Monitoring

- **Purpose:** Ensure the experiment progresses as expected, detect anomalies, and support early decision-making if justified.

Frequentist Monitoring	Bayesian Monitoring
• Use predefined checkpoints to avoid inflating false-positive rates. • Apply alpha-spending adjustments to maintain statistical rigor.	• Monitor metrics continuously without relying on fixed checkpoints. • Use updated posterior probabilities for dynamic progress assessment and early stopping decisions.

c. Interpretation of Results (post analysis)

- **Focus on Business Impact:**
    - Measure the incremental change in primary and secondary metrics (vs control group) and highlight the influence on revenue, consumer satisfaction, and platform efficiency.
    - Example: "A 10% increase in lead utilisation equates to an additional $X million in annual revenue."
- **Address Stakeholder Concerns:**
    - Tailor insights to different stakeholders:
        - **Product Teams:** Focus on tradie engagement and job acceptance rates.
        - **Finance Teams:** Emphasize revenue gains and cost efficiencies.
        - **Executives:** Highlight scalability and alignment with strategic goals.
- **Provide Context:**
    - Compare results against historical benchmarks or industry standards.
    - Example: "Our lead utilisation rate now surpasses the industry average by Y%."
- **Showcase Risks and Mitigations:**
    - Identify potential inefficiencies or risks:
        - **Low Lead Utilisation:** Oversaturation of leads or reduced tradie interest.
        - **Skewed Revenue Distribution:** Disengagement risks among low-performing tradies.
    - Propose actionable mitigations, such as revising lead allocation algorithms or offering incentives.
- **Recommend Clear Next Steps:**
    - Outline actionable and data-driven recommendations:
        - "Scale the 4-lead feature in high-performing regions first."
        - "Conduct tradie surveys to gauge satisfaction."

- **Support with Visual Storytelling:**
    - Use visualizations (e.g., bar charts, trend lines) to clearly illustrate key metrics and findings.
    - Example: A trend line showing the increase in lead utilisation over the experiment period.

d. Educating the Team

- **Tailored Communication:**
    - **Product Manager:** Use straightforward, goal-oriented language. Highlight how the experiment aligns with business goals like lead utilisation and revenue growth.
    - **Wider Team:** Simplify complex concepts (e.g., Bayesian inference) with visuals and relatable analogies to ensure understanding.
- **Visual Storytelling:**
    - Present key findings using bar charts, trend lines, and dashboards for metrics like lead utilisation and job acceptance rates.
    - Example: Show a bar chart comparing lead utilisation rates between control and test groups.
- **Experiment Integrity:**
    - Highlight the rigorous methodology, including randomisation, stratification, and data validation, to build trust in the results.
- **Stakeholder Alignment:**
    - Link findings to team-specific goals:
        - **Engineering Teams:** Emphasize backend achievements like dynamic lead assignment.
        - **Finance Teams:** Highlight revenue impacts and cost efficiencies.
- **Actionable Recommendations:**
    - Provide clear, next steps tailored to the audience:
        - "Scale the 4-lead feature in high-performing regions first."
        - "Refine lead allocation strategies based on identified inefficiencies."
- **Interactive Engagement:**
    - Conduct Q&A sessions to address team concerns and clarify results.
    - Use collaborative workshops to integrate team feedback into future iterations.

4. Implementation and Rollout

- If the results are favorable, roll out the 4-lead feature to all jobs.
- Continue monitoring metrics post-rollout to validate sustained improvement.
- Gather feedback from tradies and consumers to identify areas for further optimisation.

This solution ensures a robust and scientifically rigorous approach to validating the hypothesis, enabling data-driven decision-making for feature rollout.