**Marketing Audience Identification - Project Plan (Revised)**

## **1. ASK Phase Summary**

### **Business Task**

Identify specific professional occupations and age groups within those occupations that are likely to experience anxiety in the workplace and would value a discrete, instant-access anxiety management solution such as "Calm Button." This investigation will also explore current stress management strategies and unmet needs among these groups to inform targeted marketing approaches.

### **Key Questions**

1. **Primary Audience Segmentation:** Which occupations and age groups show the highest attack severity, lowest therapy attendance, highest stress levels, and indicate regular workplace occurrence of anxiety?
2. **Behavioral Pattern Analysis:** Among high-severity sufferers in target segments, what are common lifestyle patterns, therapy and medication usage, and stress management habits?
3. **Market Opportunity Analysis:** For each identified segment, what is the potential user base size, current management methods, attack frequency and severity, and professional context impact?

### **Metrics for Success**

* Actionable & Profession-Specific Segments
* Segment Sizing & Prioritization
* Differentiators & Workplace Context
* Data-Driven Channel Strategy Support
* Targeted Messaging & Value Proposition Alignment

## **2. PREPARE Phase**

### **Data Collection & Organization**

* Review anxiety\_attack\_dataset.csv (12,000 records)
* Create backup of raw data
* Document data dictionary and metadata (Appendix A & B of Data Analysis Request)
* Set up project directory structure (e.g., Raw\_Data, Cleaned\_Data, Analysis, Visualizations, Reports)
* Initialize version control repository (e.g., Git)

### **Initial Data Assessment**

* Verify data completeness for all 20 columns
* Check for missing values in each column
* Assess data types and formats for each variable
* Document potential biases in the dataset (acknowledging limitations of self-reported data, potential sampling bias)
* Identify any privacy concerns and ensure adherence to Privacy Requirements (Appendix C of Data Analysis Request)
* **Assess data against ROCCC criteria:**
  + **Reliability:** Assess data source credibility. (Self-reported survey data - acknowledge limitations)
  + **Originality:** Verify data source and collection method. (Assume original for case study purposes based on file description)
  + **Comprehensiveness:** Evaluate if data covers necessary variables for business questions. (Yes, for initial segmentation & behavior analysis)
  + **Currency:** Check data recency and relevance to current market conditions. (Assume recent enough for exploratory analysis, but note time sensitivity of anxiety trends)
  + **Citation:** Document data source and licensing information. (Note CCO license for publicly provided dataset)
* Verify and document adherence to Privacy Requirements (aggregate data only, no individual identification, minimum segment size 50, no location-specific analysis)
* Note data limitations (e.g., self-reported, potentially biased, no direct workplace context variables)

### **Tools Selection**

* Primary: R (tidyverse - dplyr, tidyr, ggplot2, lubridate, skimr, janitor, here) for statistical analysis, data manipulation, and visualization. Justification: R is powerful for statistical analysis, data wrangling, and creating high-quality visualizations, and is a standard tool in data analysis at Google.
* Secondary: Tableau Public for creating and sharing interactive dashboards for final presentation. Justification: Tableau is excellent for creating visually compelling and interactive dashboards, suitable for business stakeholders.
* Documentation: R Markdown for creating reproducible reports and technical documentation that integrates code, results, and narrative. Justification: R Markdown promotes reproducible research and clear communication of technical details.

## **2. PROCESS Phase**

### **Data Cleaning**

* Handle missing values: Document and justify chosen methods (e.g., imputation for some demographics, removal for records with excessive missing values if necessary - aiming for minimal data loss).
* Standardize occupation categories: Group similar occupations into broader, more actionable professional sectors (e.g., 'Doctor,' 'Engineer,' 'Teacher,' 'Student,' 'Unemployed,' 'Other').
* Clean text fields (if any text data is deemed relevant for analysis beyond the provided numerical/categorical data - though dataset may not have extensive text fields).
* Remove duplicates if any are identified based on 'ID' or other relevant identifiers.
* Format dates consistently (if date variables were present, though dataset appears to be cross-sectional without date variables).
* Document all cleaning steps in a data cleaning log (e.g., in R Markdown comments or a separate document).

### **Data Transformation**

* Create derived variables for analysis:
  + **Stress-Severity Index:** (Define as a combined metric using 'Stress Level (1-10)' and 'Severity of Anxiety Attack (1-10)' - e.g., average or weighted sum, to represent overall anxiety impact). Justification: To create a single, more holistic measure of anxiety impact.
  + **Treatment Engagement Score:** (Define based on 'Therapy Sessions (per month)' and 'Medication' variables - e.g., categorize into 'None,' 'Low,' 'Medium,' 'High' engagement based on therapy sessions and medication 'Yes/No'). Justification: To categorize and analyze different levels of treatment engagement within segments.
  + **Workplace Impact Indicator:** (Define as a proxy for workplace relevance, potentially using Occupation categories and Severity patterns, and perhaps inferring stressfulness of occupation from general knowledge or external data if feasible). Justification: To help prioritize occupations where anxiety is likely to be work-related and where "Calm Button" can offer workplace support.
* Bin age groups: Create age bands for segmentation (e.g., 18-24, 25-34, 35-44, 45-54, 55-64, 65+), ensuring minimum segment size > 50.
* Categorize occupations into professional sectors (e.g., Healthcare, Education, Tech, Finance, Service Industry, etc.) for higher-level analysis and actionable segmentation.
* Create flags for high-risk indicators: Create boolean flags for 'High Attack Severity' (e.g., Severity > 7), 'High Stress Level' (e.g., Stress Level > 7), 'Untreated' (Therapy Sessions = 0).

### **Quality Assurance**

* Verify cleaned dataset integrity by comparing summary statistics before and after cleaning.
* Cross-validate derived variables using logical checks and spot-checking against raw data.
* Document all transformations in data transformation log (in R Markdown or separate document).
* Create data quality report summarizing data cleaning steps, data quality metrics (missing values, duplicates addressed), and any remaining limitations.

## **3. ANALYZE Phase**

### **Exploratory Analysis**

* Basic statistical summaries (mean, median, standard deviation, percentiles) for key variables by occupation and age groups using R (dplyr, skimr).
* Distribution analysis of key variables (histograms, box plots) to understand data spread and identify outliers using R (ggplot2).
* Correlation analysis between Anxiety Attack Severity, Stress Level, Lifestyle Factors, Therapy Attendance, Medication Usage using R (corrr, psych).
* Initial segmentation exploration using cross-tabulation and grouped summaries in R (dplyr, tables).

### **Segment Analysis**

1. **Primary Audience Segmentation:**
   * Occupation & Severity analysis: Identify top occupations by average attack severity and stress level using R (dplyr, ggplot2).
   * Age & occupation cross-analysis: Analyze severity and stress within age groups *within* top occupations using R (dplyr, ggplot2).
   * Treatment engagement patterns: Analyze therapy attendance and medication usage across occupations and age groups using R (dplyr, ggplot2).
   * Stress level analysis by profession: Detailed breakdown of stress level distributions by occupation using R (dplyr, ggplot2).
2. **Behavioral Pattern Analysis:**
   * Lifestyle factors correlation: Examine correlations between lifestyle variables (sleep, activity, etc.) and anxiety severity within target segments using R (corrr, psych).
   * Treatment preferences: Analyze medication and therapy usage patterns within high-severity segments using R (dplyr).
   * Professional context impact: Interpret findings in the context of typical workplace stressors for identified occupations (qualitative analysis).
   * Demographic patterns: Describe the demographic composition (age, gender) of high-severity and underserved segments.
3. **Market Opportunity Analysis:**
   * Segment size calculations: Estimate the size of each identified target segment within the dataset (counts, percentages).
   * Severity-weighted prioritization: Prioritize segments based on a combination of size and anxiety severity metrics.
   * Treatment gap analysis: Quantify the "treatment gap" (low therapy attendance) within prioritized segments.
   * Professional context assessment: Qualitatively assess the workplace relevance and potential for "Calm Button" adoption within each prioritized segment.

### **Statistical Analysis**

* Cluster analysis (using R's kmeans or similar) of demographic and severity data to identify natural user segments.
* Chi-square tests (using R's chisq.test) to assess associations between categorical variables (e.g., Occupation and Therapy Usage).
* ANOVA (using R's aov) for group comparisons of means (e.g., comparing average Severity scores across occupations).
* Regression analysis (using R's lm or glm) to explore predictors of anxiety severity or therapy engagement (if appropriate and data allows).

## **4. SHARE Phase**

### **Data Visualization**

1. **Segment Profiles:**
   * Demographic composition charts (bar charts, pie charts) using Tableau and R (ggplot2).
   * Attack pattern visualizations (box plots, histograms) using Tableau and R (ggplot2).
   * Treatment preference graphs (bar charts, stacked bar charts) using Tableau and R (ggplot2).
   * Key pain point matrices (heatmaps, tables) summarizing segment characteristics using Tableau and R.
2. **Market Opportunity:**
   * Segment size treemap visualizing the relative size of target segments using Tableau.
   * Opportunity matrix (scatter plot or bubble chart) plotting segment size vs. anxiety severity/treatment gap using Tableau.
   * Digital Channel Preference Analysis: Visualizations (bar charts, tables) summarizing potential marketing channels based on inferred user behavior and professional context.
   * Customer Journey Point Mapping: Diagram/flowchart illustrating the anxiety attack experience and potential intervention points for "Calm Button" using presentation software (Google Slides/PowerPoint).

### **Deliverables Creation**

1. **Technical Documentation:**
   * Methodology Report (R Markdown document detailing the data analysis process, from Ask to Act phases).
   * Statistical Analysis Results (R Markdown document with code, outputs, interpretations of statistical tests).
   * Data Processing Documentation (R Markdown document or separate log file detailing data cleaning and transformation steps).
   * Code Repository (GitHub repository with all R scripts, R Markdown documents, data files (if permissible), and presentation materials).
2. **Business Deliverables:**
   * Executive Summary (concise 1-page summary of key findings and recommendations).
   * Segment Profiles (detailed descriptions of 3-5 target audience segments, including demographics, anxiety characteristics, behaviors, and pain points).
   * Marketing Recommendations (actionable recommendations for marketing channels, messaging, and positioning for each target segment).
   * Channel Strategy Guide (document outlining recommended marketing channels and rationale for each segment, considering professional context and digital behavior indicators).
3. **Presentation Materials:**
   * Executive Presentation (PowerPoint or Google Slides deck, max 30 minutes, focusing on key insights and business recommendations).
   * Detailed Findings Deck (supplementary slides with more in-depth data, visualizations, and technical details for deeper dives).
   * Interactive Dashboard (Tableau Public dashboard for stakeholders to explore segment data and visualizations dynamically).
   * Handover Documentation (brief guide for stakeholders on how to use the deliverables, access data/code, and implement recommendations).

**4. ACT Phase Planning**

### **Key Tasks**

* Develop action plan based on marketing recommendations (timeline, responsibilities, budget).
* Implement targeted marketing campaigns for prioritized segments.
* Establish a system for monitoring campaign performance and user acquisition within target segments (define key performance indicators - KPIs).
* Plan for A/B testing of different marketing messages and channels to optimize campaign effectiveness.
* Create feedback loop to gather user feedback from target segments and inform future app development and marketing iterations.