Analytics Startup Plan

Synopsis: ***This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

Project	*Understanding Key Drivers of Burnout in Remote Work Post-Pandemic*
Requestor	*Human Resources & Wellness Strategy Team*
Date of Request	*July 15, 2025*
Target Quarter for Delivery	*Q3 2025*
Epic Link(s)	*Employee Health, Remote Work Trends, HR Optimization*
Business Impact	*Help HR and corporate wellness teams mitigate remote work burnout by identifying key contributing factors to burnout, enabling proactive wellness planning.*

1.0 Business Opportunity Brief

Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve for. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment.

……

The specific ask:

Problem Statement
After the COVID-19 pandemic, many employees transitioned to permanent or hybrid remote work. Initial qualitative feedback suggests rising health concerns including screen fatigue, back pain, and work-life imbalance. However, we lack quantified insights to make evidence-based changes to remote work policies.

Business Ask
Develop a data-driven assessment that answers:

- How does remote work frequency relate to mental and physical health scores?
- What employee segments (e.g., age, job type, household size) are most at risk of adverse health impacts?
- Can we identify predictors of positive well-being among remote workers?

Opportunity Impact

Benefit	Impact
Healthier Workforce	Reduce reported burnout and fatigue by 30%+
Retention	Lower attrition rates among remote employees
Policy Improvement	Tailor WFH policies by segment (e.g., job role, age)
Productivity Uplift	Improve output by optimizing remote work practices

Stakeholder Alignment

- HR Team: Wants to reduce employee burnout and sick days
- IT & Facilities: Needs insight to optimize hybrid support tools
- Executives: Seeks evidence to guide long-term workplace planning

1.1 Supporting Insights

Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of market, nationally and regionally?

Trends & Research Findings

- **Remote Work Burnout**: Studies show over 60% of remote workers report higher stress levels (Harvard Business Review, 2024).
- **Musculoskeletal Issues**: Poor ergonomic setups are linked to back and neck pain in 40%+ of employees (NIOSH, 2023).
- **Mental Health**: Lack of boundaries and social interaction correlates with higher anxiety and sleep disruption (McKinsey, 2024).

Competitor Benchmarks

Company	Well-being Strategy
Google	Subsidized ergonomic setups & weekly mental check-ins
Microsoft	AI-driven workload balancing & wellness pulse surveys
Salesforce	Mandatory "Focus Fridays" and digital detox policies

1.2 Project Gains

Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? *This section is particularly key for prioritization against company goals and KPI's.*

Potential Outcomes:

- Evidence-backed wellness interventions.
- Reduced employee absenteeism and turnover.
- Improved targeting of corporate wellness programs.
- Use findings to support long-term hybrid work design

If we do nothing:

- Rising healthcare costs.
- Decreased productivity.
- Higher mental health claim rates in insurance plans.

Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority and data availability.

2.0 Analytics Objective

List the key questions, assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, however a recommended operating model or blueprint for a pilot etc.

Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed.

Primary Questions:

1. Is there a statistically significant relationship between remote work and self-reported stress or depression?
2. Are certain demographics (age, gender, income) more vulnerable to negative health impacts?
3. How do lifestyle habits (exercise, sleep, screen time) correlate with health outcomes among remote workers?
4. Which factors most influence health scores (e.g., hours of screen time, exercise frequency, household setup)?
5. Does remote work frequency negatively correlate with mental well-being?
6. Can we predict which segments are most at risk for health decline?

Hypotheses:

- H1: Remote workers have higher self-reported mental health issues post-pandemic.
- H2: Increased screen time and reduced physical activity are positively correlated with health issues.
- H3: Younger workers are more likely to report anxiety than older counterparts.
- H4: Higher remote work hours are associated with increased physical discomfort.
- H5: Employees without a dedicated home office experience worse mental health outcome.
- H6: Regular exercise or screen breaks mitigate health risks from remote work.

2.1 Other related questions and Assumptions:

List any assumptions that may affect the analysis

- Assume respondents are truthful in survey responses.
- Assume responses are representative of remote workers in North America.
- Some variables (e.g., income) may be self-reported and subject to bias.

2.2 Success measures/metrics

What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis.

- Correlation coefficients between remote work and health variables.
- P-values from statistical tests (e.g., Chi-square, ANOVA).
- Segmented insights by age, gender, work type.
- Clustering of health outcomes using unsupervised methods (keeping as an option)

2.3 Methodology and Approach

Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.

Type of Analysis: *logistic regression, linear regression, Chi-square test*

The initial approach will be to use a decision tree to determine which dealer level variables (size, region, segmentation...) are most significant related to a dealer's likelihood to churn. We will also use other techniques to verify our findings.

Methodology: *Key questions from 'Analytics objective' will be tackled in ascending order as outlined in '5.0 Timelines and deliverable section'.*

We will start by identifying all dealers that were active in the first quarter of 2018. We will then define the response variable to be a 1 if they are still active, and 0 otherwise. We will build a decision tree based on this sample, and observe which variables are the most important in determining whether these dealers are still active. We can then repeat this analysis using a sample based on the dealers that were active in the second quarter of 2018. The idea is to check if the same variables are being identified as the most important drivers of churn, or if the importance of variables change as we get closer to the present day.

Output: *The output will be a set of insights, rules and strategic recommendations that will help us to evaluate dealers based on likelihood to churn and positioning of sales-match.*

Type of Analysis

- Logistic Regression (Health risk classification)
- Decision Tree (to capture nonlinear relationships)
- Correlation Analysis (Pearson/Spearman)
- T-tests / ANOVA for group comparisons
- Clustering (for segment-based policy design)

Data Cleaning

- Encode categorical variables (e.g., "WFH Setup", "Job Role")
- Impute missing values (median for numerics, mode for categoricals)

- Address skew in continuous variables (e.g., log-transform screen time)

Data Preprocessing

- Normalize continuous variables (e.g., screen time, exercise hours)
- Feature engineering (e.g., combine sedentary hours + breaks)
- Handle multicollinearity

Modeling Plan

- Split data into training/test sets (80/20)
- Train binary and multiclass classification models
- Use SMOTE or class weighting for imbalance (if applicable)
- Evaluate using AUC-ROC, Precision, Recall, and F1-score

Expected Outcomes

- Identify top health risk indicators (e.g., lack of exercise, >6 hours screen time)
- Profiles of high-risk employee segments
- Actionable recommendations: better equipment subsidy, policy nudges

3.0 Population, Variable Selection, considerations

Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today.

Audience/population selection: Remote workers surveyed in 2024-2025.

Observation window: Post-pandemic period (2023–2025).

Inclusions: Remote/hybrid workers only.

Exclusions: Unemployed individuals or those not working remotely.

Data Sources: Kaggle: post_pandemic_remote_work_health_impact_2025

Audience Level: Individual-level survey responses.

Variable Selection:

Key Variables

Categorical

- Gender
- Age Group
- Job Role

- WFH Setup (Dedicated, Shared, None)
- Exercise Frequency Category

Numerical

- Weekly Remote Work Hours
- Screen Time (avg hours/day)
- Exercise Hours per Week
- Sleep Hours per Night
- Social Isolation Score (1–5)
- Salary Range

Target Variables

- **Burnout levels(low, medium or high)**

4.0 Dependencies and Risks

Identification of key factors that may influence the outcome of the project and likelihood of it happening:

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
Churn rate being inflated by counting multiple contracts from the same rooftop as individual observations.	*Low*		*Once analysis begins, we can quantify the inflation. However, this approach allows us to compare how the same dealer performed across different contracts and find useful patterns.*

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact

Self-reported mental health data inaccuracies	High	Medium	Model may misrepresent true health impacts
Missing data from hybrid workers	Medium	Medium	Reduced accuracy and generalizability
Misclassification of employment types	Medium	Low	Confounded conclusions in subgro analysis
Outdated data or pre-2021 behavior skew	Low	Low	Misleading trend projections
Limited SME availability for context validation	Medium	High	May delay interpretation and deployment

5.0 Deliverable Timelines

List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process.

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Kick-off / Formal Request	Align on project objectives, target (burnout prediction), stakeholders, and key outcomes.	Full	*1*	July 13
2.	Assessment / Triage	Evaluate data completeness and quality (burnout scores, screen time, WFH setup, etc.), confirm variable definitions, identify class imbalance or bias risks.	Full	2	July114
3.	Prioritization	Finalize modeling strategy (classification vs regression),	Full	2	July 15

		prioritize variables based on EDA, and confirm target as categorical or continuous.			
4.	Data Exploration & Analysis • Issues with duplicates • Issues with Spend data	Conduct EDA on burnout levels: identify correlations, outliers, and trends across demographics, lifestyle, and work-related variables.	Full	7	July 15-20
5.	Modeling Review	Select appropriate modeling frameworks (e.g., logistic regression, decision tree), define feature sets and validation methods.	Full	7	July 20
6.	Modeling	Train models to predict burnout levels, address imbalance if needed (e.g., SMOTE), validate using metrics like AUC, MAE, or F1-score depending on target type.	Full	7	July 21
7.	Modeling Peer Review	Conduct internal review to validate assumptions, feature relevance, and reproducibility. Compare alternative models.	Full	2	July 22
8.	Presentation Peer Review	Create stakeholder-ready narratives: visualize burnout drivers by segment and risk level; review for clarity and actionability.	Full	2	July 23
9.	Governance, Documentation	Document data transformations, modeling choices, assumptions, and limitations. Ensure reproducibility standards are met.	Full	7	July 28
10.	Portfolio	Prepare final presentation materials: dashboards, risk profiles, and strategic recommendations for HR and policy design.	Full	7	July 31
11.	Delivery & sign-off	Present final insights, validate with stakeholders, finalize burnout risk framework, and receive sign-off.	Full	3	August 11