# **Capstone Project Proposal (Data Analysis Pathway)[The-Productivity-of-Remote-vs-in-office-Workers]**

**Project Title:** *The-Productivity-of-Remote-vs-in-office-Workers*

**Proposed By:** *Tiffanie White*

**GitHub Repository: https://github.com/TiffanieWhite/The-Productivity-of-Remote-vs-in-office-Workers.git**

## **Project Overview**

Briefly describe the overall purpose of your project. What are you trying to analyze, explain, or uncover? Why does this matter? Write 2–3 sentences that summarize the big picture.

I changed my project direction after reading an article about how the current administration is reinforcing mandates to take away remote work. I will be trying to analyze how remote work has a bigger positive impact on workers and their production versus in office workers. It matters to me, personally, because I have two younger children (one is autistic and is completing his education at home and the other has a mild disability has gets out of school almost in the middle of the work day; I don’t have the income to send him to a daycare. I am a single parent and I have family support, but everyone works regular jobs.) I wish that those higher ups could step into the working class shoes and see how remote ties into people being able to maintain their homes, keep custody of their children, and how it’s linked to soo many other aspects of life to just be able to sustain a decent home life and happier, more productive, employees.

## **Data Sources**

List the datasets you will use. For each, include:

DATA SET# 1

* **Dataset name:** Employee Performance and Productivity Data
* **Source** Kaggle
* **Relevant fields:** all data
* **Link**: https://www.kaggle.com/datasets/mexwell/employee-performance-and-productivity-data

DATA SET #2

**Dataset name:** Remote Work Productivity

* **Source:** Kaggle
* **Relevant fields** : will try to include all fields to help support the argument to keep remote work
* **Link**: https://www.kaggle.com/datasets/mrsimple07/remote-work-productivity

*(You must include at least two datasets from different sources. Additional optional datasets may be listed here as well.)*

## **Research Objectives**

State your main questions and objectives. You should include:

* **Primary Question(s):** How much of a financial impact, for the company, is it to have a remote worker versus an in office worker?
* **Secondary/Exploratory Questions:** What is the productivity difference between the remote worker versus the in office worker? What is the turnover rate for the remote vs in office worker? Possibly delve into the over insurance/health difference between the two.

Explain briefly why these objectives are valuable or interesting.

These stats and the data is interesting so that companies can see and learn that not only is their a financial impact positively for the remote workers ,but it may actually cost the employer more in the long run to make everyone become in office workers.

## **Data Preparation Approach**

Describe how you will prepare the data for analysis. Be specific:

* How you will align or merge datasets (keys, years, geography).
* How you will address missing values (replacement, dropping rows, etc.).
* How you will address outliers or extreme values.
* What new variables or indices you may create.
* How you will structure your relational database (at minimum two tables from different sources, joined through a common key).

## **Current Status**

Summarize what you have already completed. Examples:

* Data acquired and inspected.
* Early cleaning or standardization steps.
* Prototype functions or exploratory plots.
* Evidence that the datasets can be successfully merged.

## **Deliverables (Remaining Work)**

List the specific things you still need to complete for your project. Break them into **required** tasks (must be done to meet project requirements) and **stretch** tasks (optional, if time allows).

* **Required Tasks (must be completed):***(Example: finish cleaning literacy dataset; implement 3 Python functions; build SQLite schema and load tables; produce 3 required visualizations; write README and data dictionary.)*
* **Stretch Goals (optional, if time allows):***(Example: add ACS demographics to analysis; build interactive dashboard; run regression model.)*

## **Project Timeline**

Outline the sequence of major steps or milestones leading up to the deadline. You may structure this as weeks, phases, or checkpoints. Examples:

* Phase 1: Acquire and clean data, set up database.
* Phase 2: Exploratory analysis, build functions, first draft visuals.
* Phase 3: Deeper analysis, refine visuals, draft report.
* Phase 4: Finalize deliverables, polish repo, record presentation.

Be as specific as possible about what you plan to complete at each step.

## **Additional Considerations**

Note any assumptions, limitations, or risks you anticipate. Examples:

* Geographic regions or variables you will exclude (and why).
* Additional datasets you may add if time allows.
* Possible challenges (e.g., very large files, inconsistent variable names, data sparsity).