KPMG 1E Productivity-Energy Tradeoff Optimizer

AI Studio Final Project

KPMG

BREAK THROUGH TECH

Meet Our Team

Kimberly Hernandez
University of
Maryland

Bettina George
University of North
Carolina at Chapel Hill

Ryan Vaseem
Stevens Institute of
Technology

Our AI Studio Coach & Challenge Advisors

Dr. Uohna June Thiessen
AI Studio Coach

Kathi Ray
Director

Sharad Nagariya
Director

Ryan Tuggle
Director

Nainika Narayanan
Associate

Brooke Bernstein
Associate

Presentation Agenda

Project Overview	4 minutes
Data Preparation	4 minutes
Data Understanding	4 minutes
Modeling and Evaluation	4 minutes
Conclusion	4 minutes

Core Values

1 **Integrity**
Do what is right.

Sustainability
AI solutions should be designed to be energy efficient, reduce carbon emissions, and support a cleaner environment.

2 **Excellence**
We never stop learning and improving.

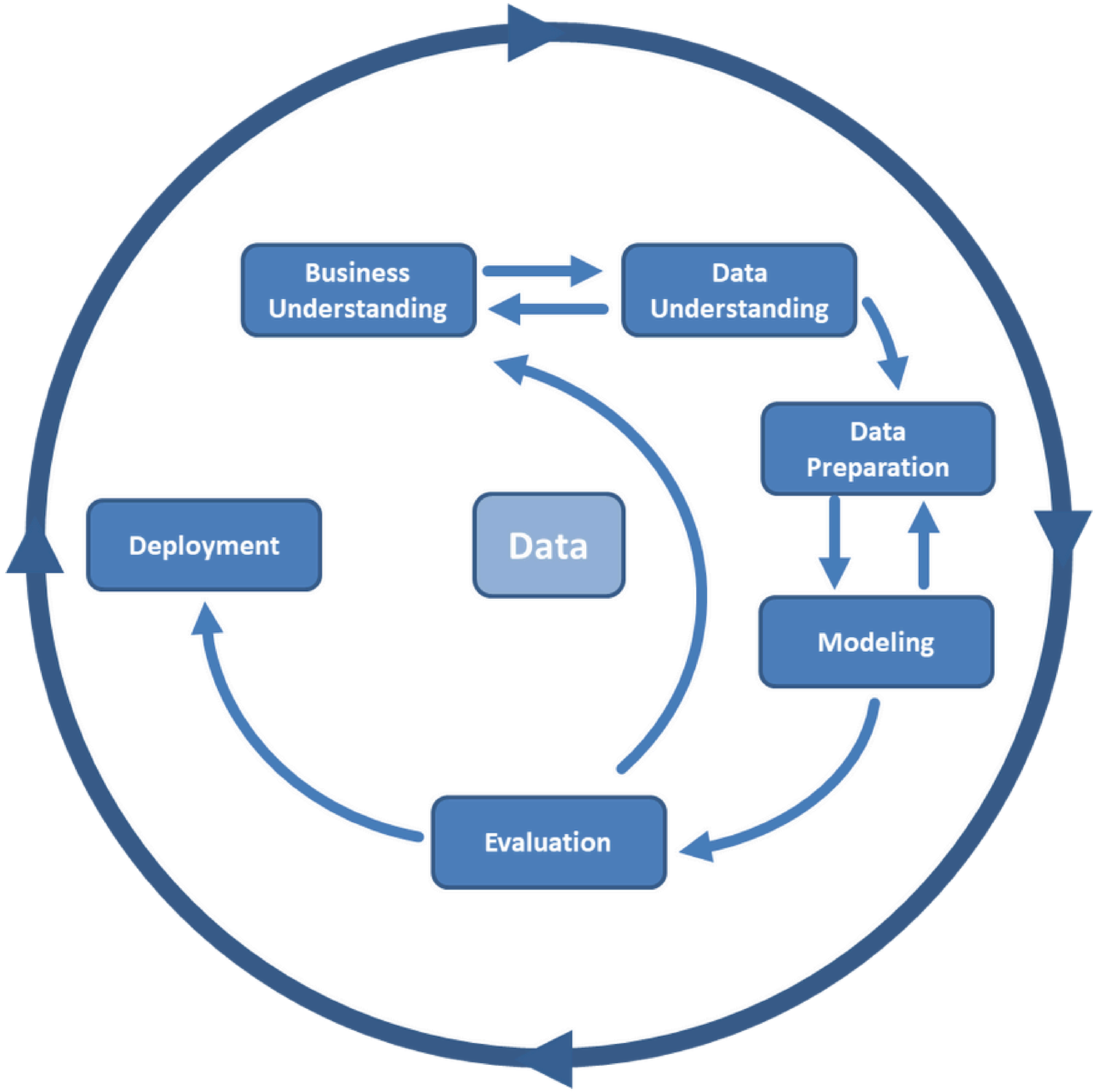
3 **Courage**
We think and act boldly.

4 **Together**
We respect each other and draw strength from our differences.

5 **For Better**
We do what matters.

KPMG US: Our Values

CRISP-DM Methodology: The Cross-Industry Standard Process for Data Mining



AI Studio Project Overview

Can we optimize the tradeoff between energy usage and employee productivity with AI?

Project Objective

Optimize Efficiency to Avoid Energy Waste

Help organizations to reduce unnecessary energy usage while maintaining current AI productivity levels.

Identify Workforce Usage of Gen AI Models

Assess which industries and job roles use AI in their everyday workflows.

Productivity Value

Assess the productivity value AI brings to the workforce based on the average time saved on work tasks.

Business Impact

Cut Costs & Boost Efficiency

Balance energy performance with sustainability to lower operational expenses.

Establish Smarter Workflows

Use AI to enhance employee productivity, collaboration, and task efficiency.

Scale Long-Term AI Growth

Drive long-term growth through AI-enabled systems that adapt and optimize enterprise performance.

Project Steps

Data Preparation

- Extracted data to analyze energy use by output tier.
- Cleaned and merged datasets to remove missing values.
- Selected Claude AI as an eco-efficient model for workforce impact analysis.

Exploratory Data Analysis

- Visualized patterns in energy usage and AI productivity across occupational groups.
- Identified key correlations between model efficiency and workforce impact.

Model Development and Evaluation Metrics

- Built a linear regression model to predict productivity and energy trade-offs.
- Applied the NSGA-II algorithm to optimize multi-objective outcomes (energy vs. productivity).

Final Results

- Obtained the optimal balance between energy efficiency and productivity using NSGA-II.
- Summarized findings into visual insights for sustainable and responsible AI adoption.

Project Strategy

August

Project Planning

Identify key goals of our AI Studio challenge, create a hypothesis, and outline project steps.

September

Data Collection & Exploratory Data Analysis

Extract and combine datasets from multiple sources to identify energy-use trends.

October

Model Development & Prediction Results

Train two linear regression models and apply the NSGA-II algorithm to balance energy efficiency and productivity, using a Pareto frontier.

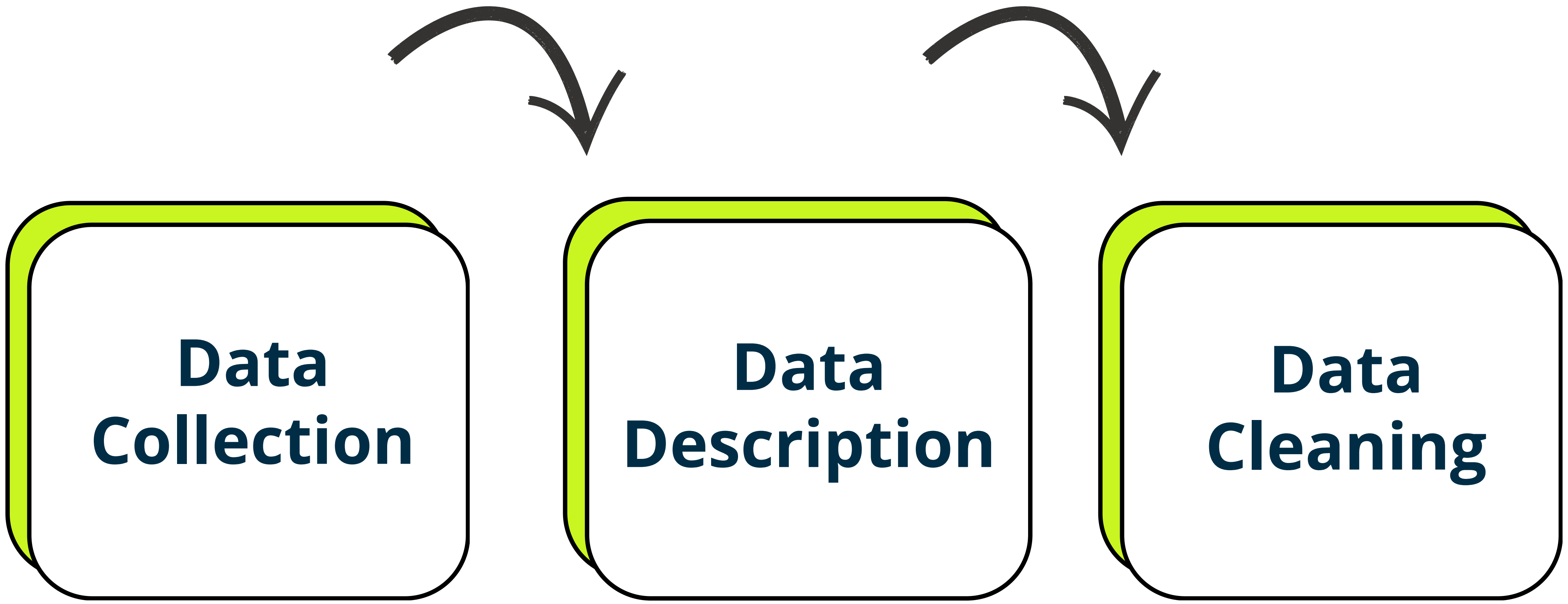
December

Final Presentation & GitHub Repository

Finalize insights, visualizations, and model results into a final presentation and GitHub repository.

Data Preparation

Overview



Data Collection

AI Energy Benchmark
Measures energy consumption across 28 AI models under low, medium, and high computational loads.

Claude AI Work Usage
Highlights U.S. worker adoption rates, occupations, industries, and AI-related task distributions.

Federal Reserve AI Productivity Report
Quantifies the average time saved per worker using AI tools based on usage intensity.

Data Description

Reviewed dataset structure using Pandas .info() and .describe() to summarize rows, columns, and data types.

Identified key variables including AI model, energy consumption (Wh), intensity level, and workforce productivity (time saved).

Conducted data profiling to check for completeness, consistency, and outliers, ensuring statistical accuracy before model training.

Visualized value distributions and correlation matrices to explore relationships between energy usage and AI model performance.

Data Cleaning

Remove duplicate and inconsistent entries in Claude_AI_Work_Usage and replace missing medium-tier energy values in AI_Energy_Benchmark with mean estimates.

Standardize all energy measurements to watt-hours (Wh) and normalize intensity levels (low, medium, high) for consistency across datasets.

Detect and clean extreme outliers, including AI models with energy usage above 300 Wh, and engineered new features such as energy_per_task for analysis.

Key Points

Data Collection

Focused on aligning common variables such as energy consumption, model type, and workforce productivity to prepare for regression and NSGA-II optimization.

Data Description

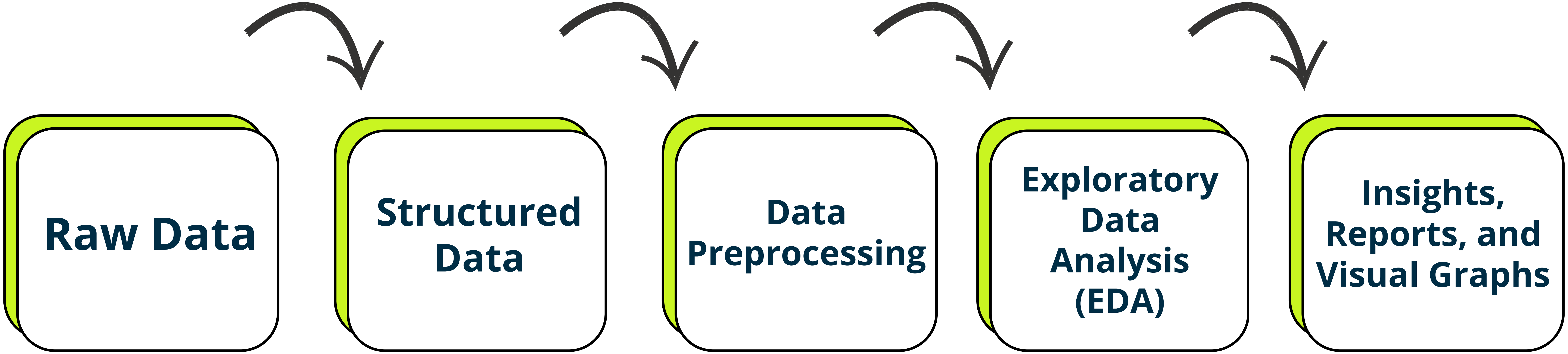
Reviewed dataset structures and statistical summaries to identify trends, key variables, and correlations between energy use, AI intensity, and workforce performance.

Data Cleaning

Removed missing and duplicate values, standardized energy units (Wh), and addressed outliers above 300 (Wh) to ensure consistent and reliable modeling inputs.

Data Understanding

Overview


Raw Data
Structured Data
Data Preprocessing
Exploratory Data Analysis (EDA)
Insights, Reports, and Visual Graphs

Raw Data

Datasets	Link to Study	Description	Total Rows	Total Columns
Dataset 1: Claude AI Workplace Usage	Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations (Academic Paper)	**Random Sampling of 2.8 million Claude AI conversations from November 2024 to December 2024.**	23 rows	8 columns
Dataset 2: AI Energy and Carbon Footprint	How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference (Academic Paper)	**Gen AI Energy consumption per query from models published from March 2023 to April 2025.**	29 rows	6 columns
Dataset 3: AI Productivity *(Time Saved Using AI)*	The Impact of Generative AI on Work Productivity (Published by the Federal Reserve)	**Employee Productivity Using AI, measured with time saved on work tasks. This source is being frequently updated.**	92,927 rows	2 columns

Structured Data

Standardization

Converted all three datasets into structured Data Frames, verified column consistency and data types, and standardized categorical variables for modeling readiness.

Merging Datasets

The datasets were then merged into a single structured table by aligning shared attributes, including AI model, energy consumption tier (low, medium, high), and workforce productivity levels.

Impact on the Model's Development

This merged dataset focused on our selected features and target label, removing any less important columns to promote efficiency for our linear regression model and NSGA-II algorithm.

Data Preprocessing

Coerced Energy Columns to Numeric

```python
# Coerce to numeric (handles 'object' High column)
for c in ['low_mean_wh','med_mean_wh','high_mean_wh']:
    energy_df[c] = pd.to_numeric(energy_df[c], errors='coerce')
```

Removed All Missing and Null Values

```python
fed_df = fed_df.dropna()
claude_df = claude_df.dropna()
energy_df = energy_df.dropna()
```

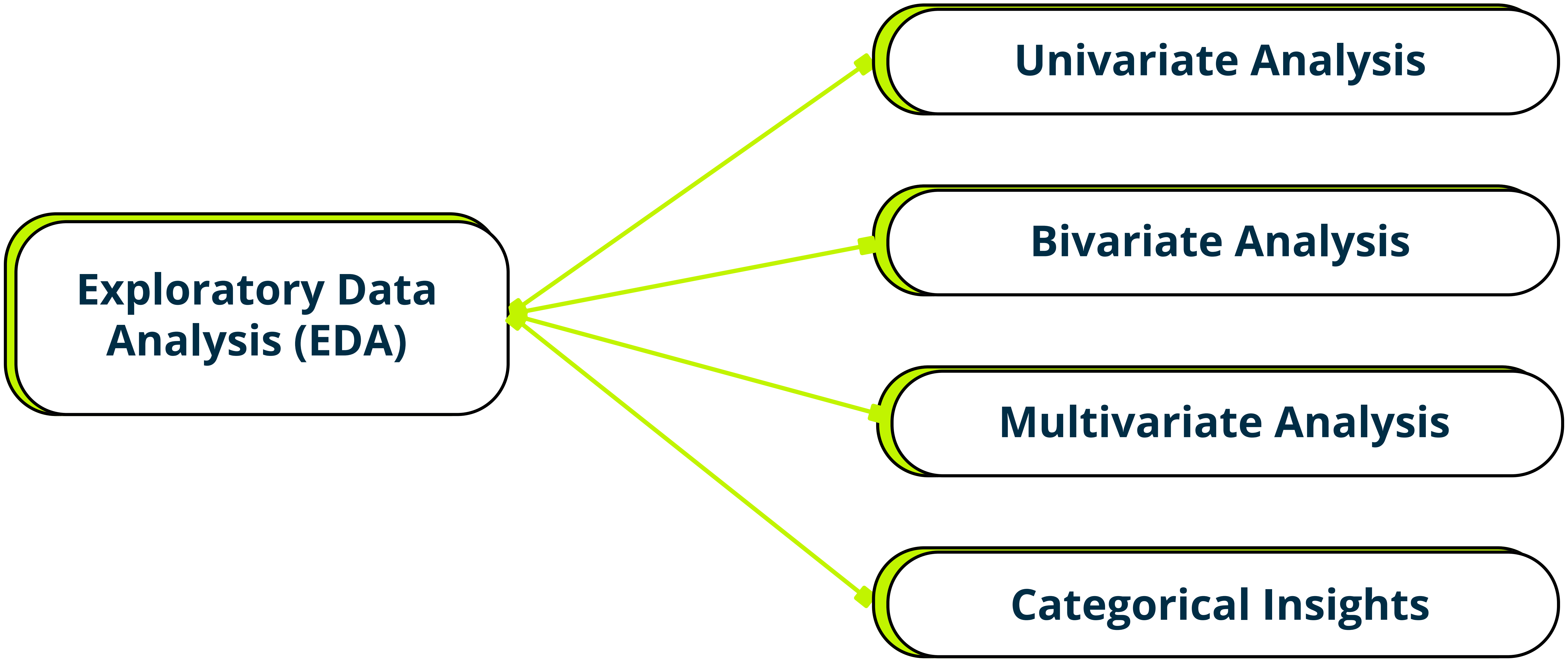
```
After Cleaning:
Fed Reserve rows: 92926, remaining nulls: 0
Claude AI rows: 18, remaining nulls: 0
Energy Benchmark rows: 25, remaining nulls: 0

All missing/null values removed successfully!
```

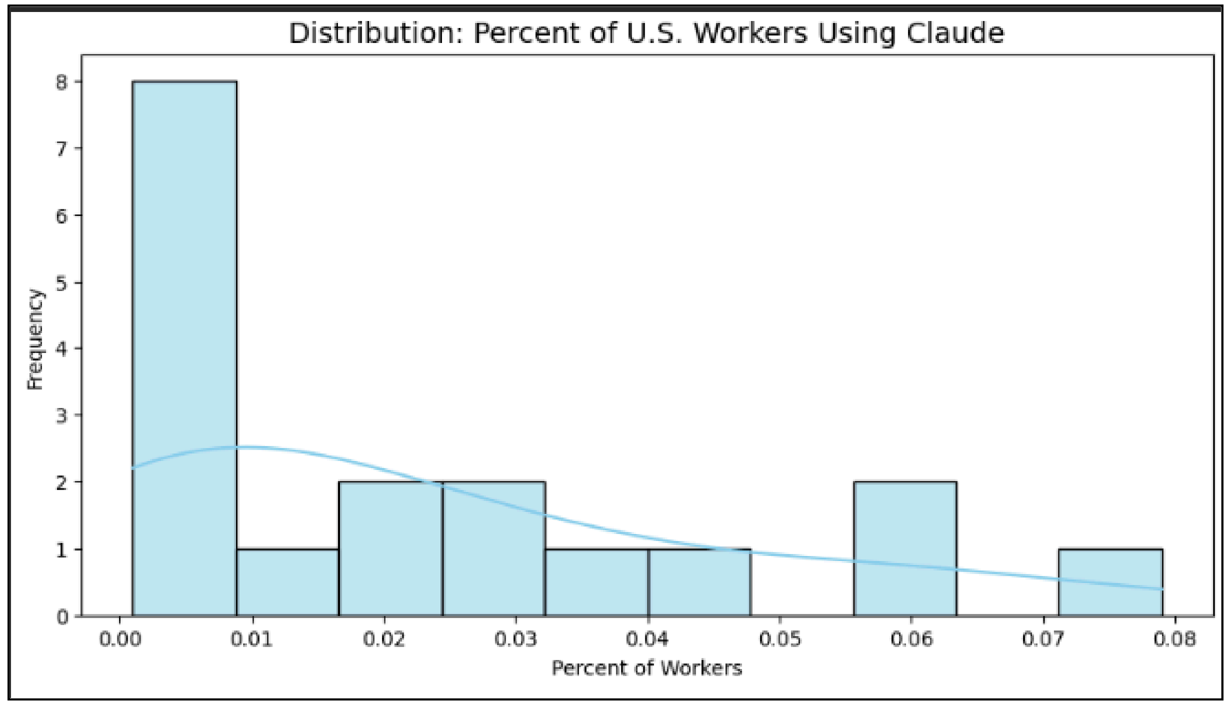
Rename Columns for Consistency

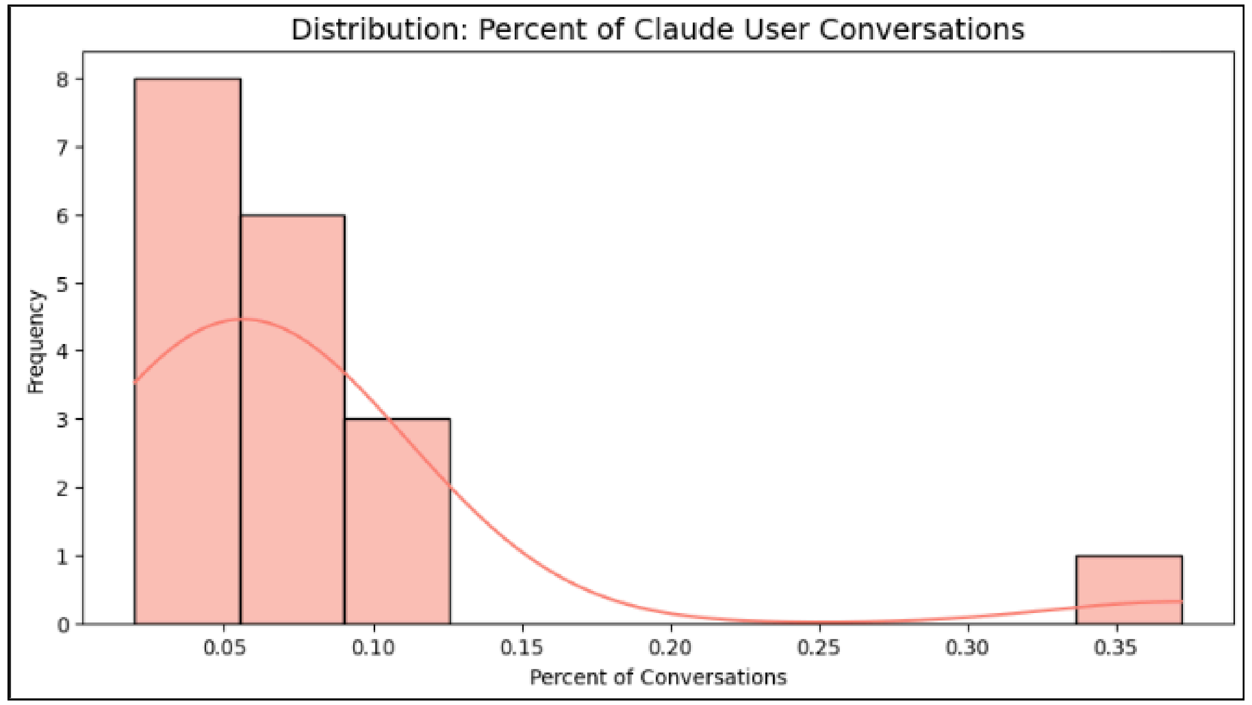
```python
fed_cols_map = {
    'Average time saved with AI (quantitative)': 'avg_minutes_saved',
    'frequency/intensity': 'intensity'
energy_cols_map = {
    'AI Model (28 different versions)': 'ai_model',
    'AI Model': 'ai_model',
    'Low Mean Energy Consumption (100 input - 300 output) (Wh)': 'low_mean_wh',
    'Low Std Dev Energy Consumption (100 input - 300 output) (Wh)': 'low_std_wh',
    'Medium Mean Energy Consumption (1k input - 1k output) (Wh)': 'med_mean_wh',
    'Medium Std Dev Energy Consumption (1k input - 1k output) (Wh)': 'med_std_wh',
    'High Mean Energy Consumption (10k input - 1.5k output) (Wh)': 'high_mean_wh',
    'High Std Dev Energy Consumption (10k input - 1.5k output) (Wh)': 'high_std_wh'
```

Exploratory Data Analysis

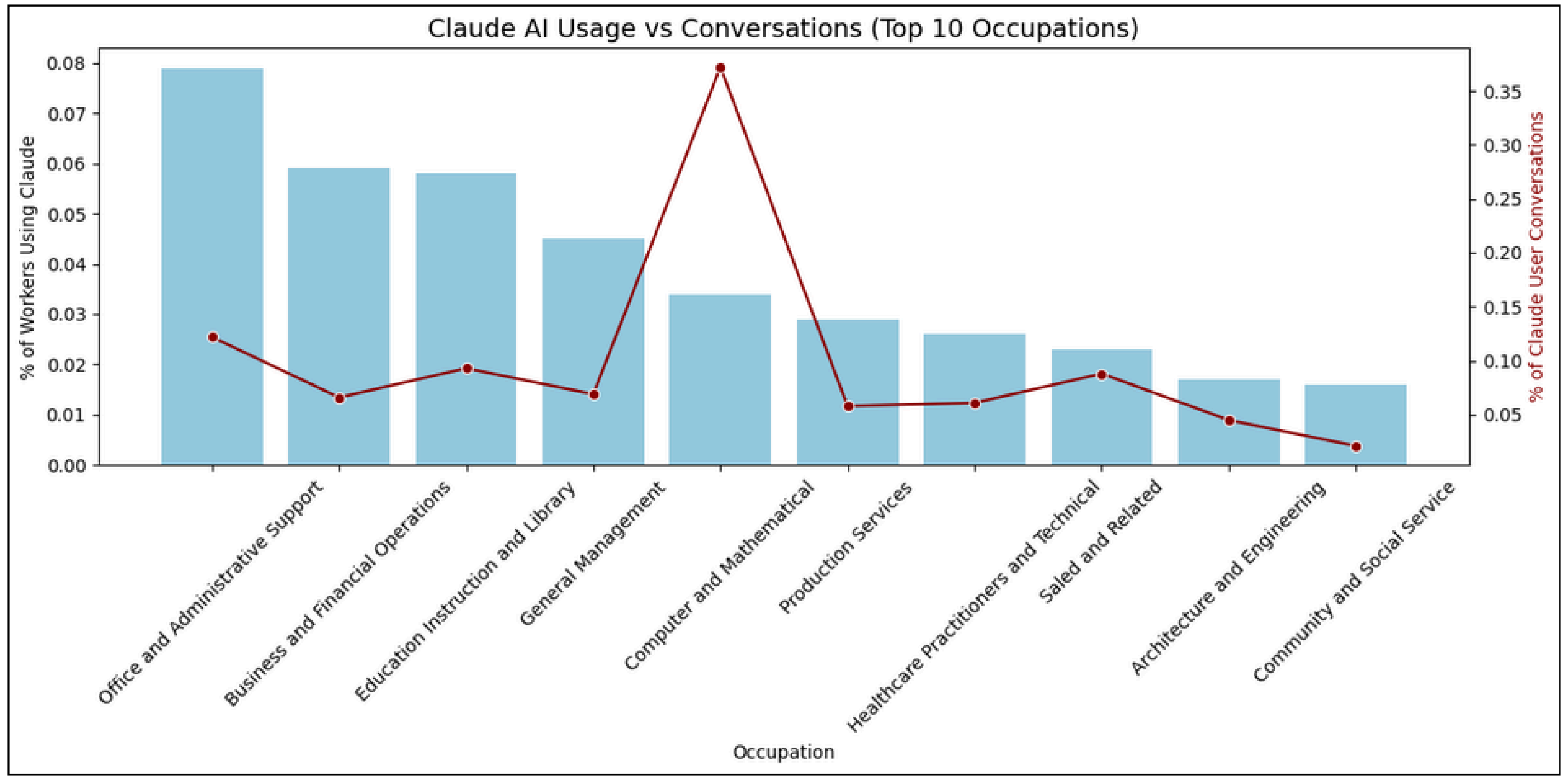


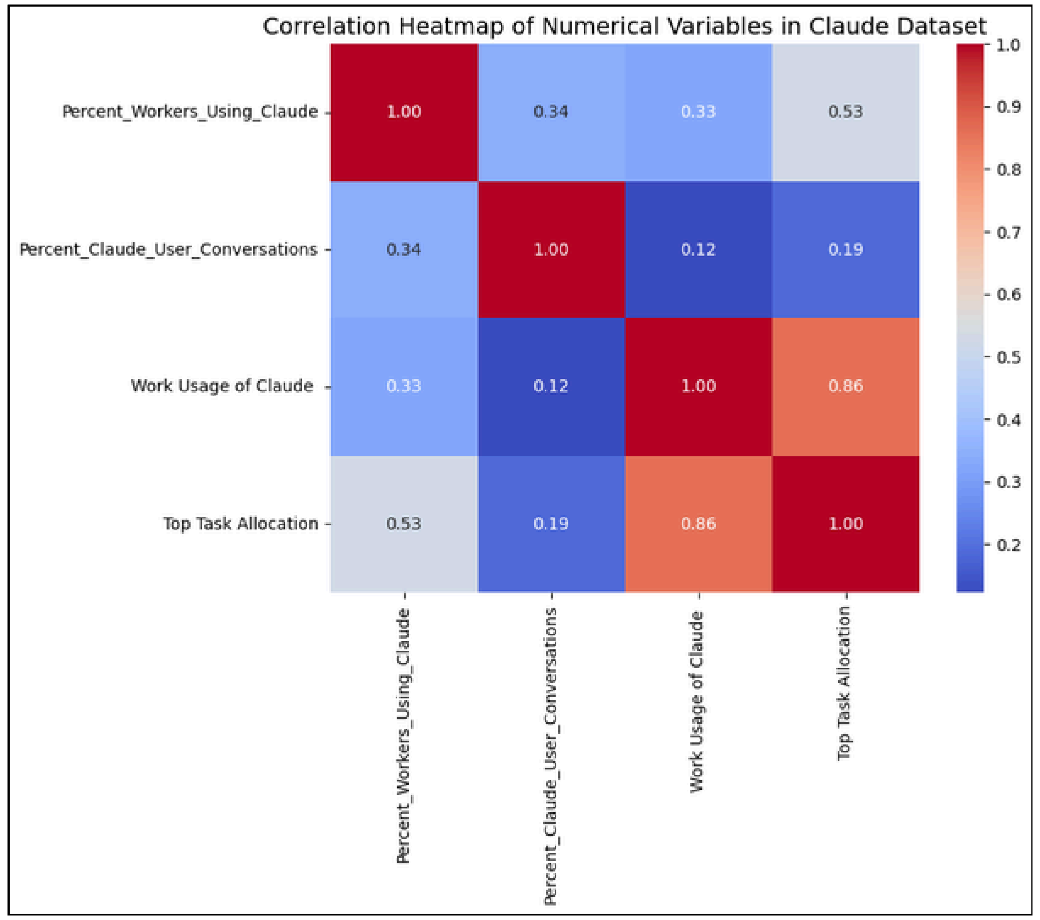
Univariate Analysis



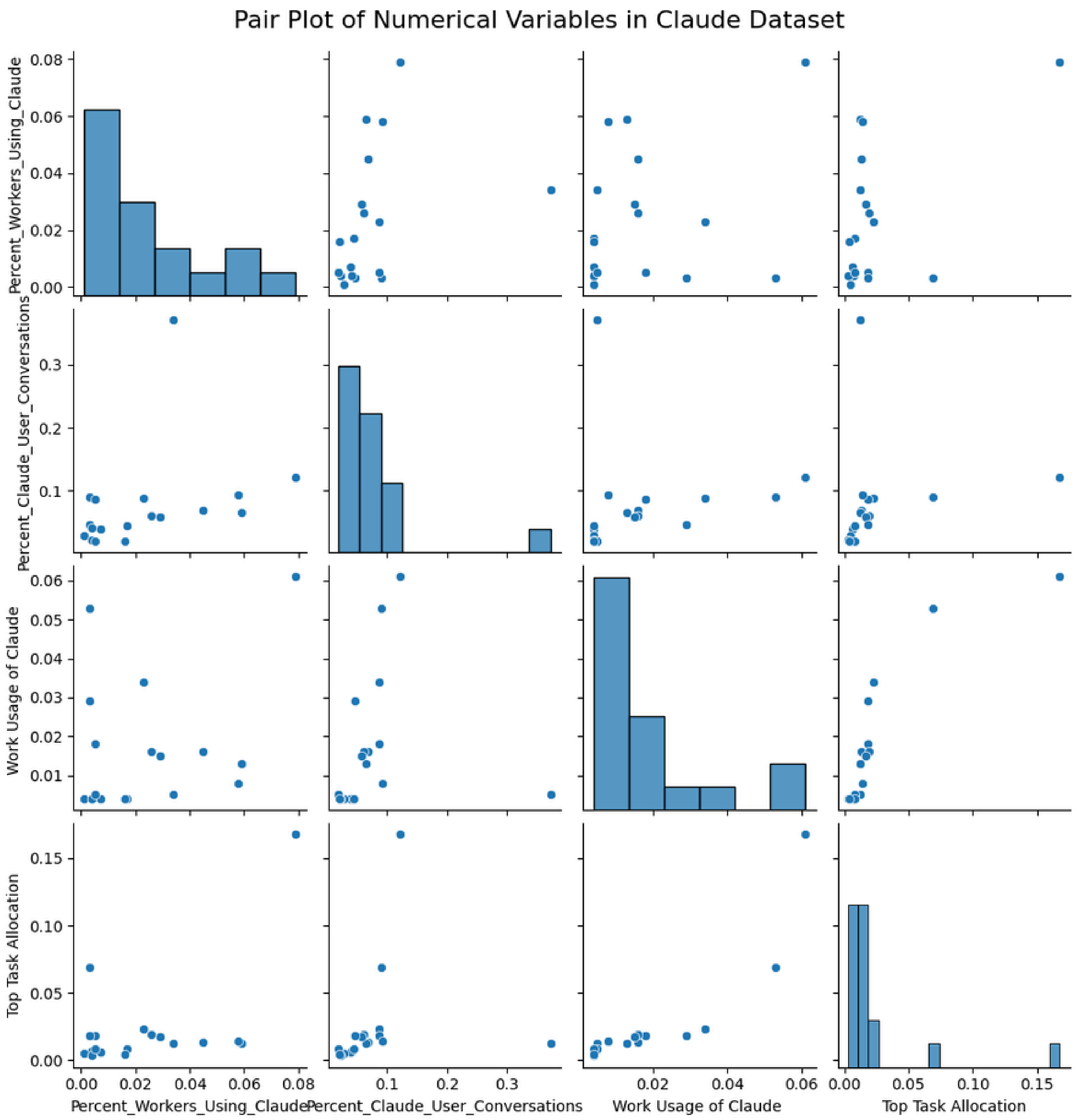


Bivariate Analysis

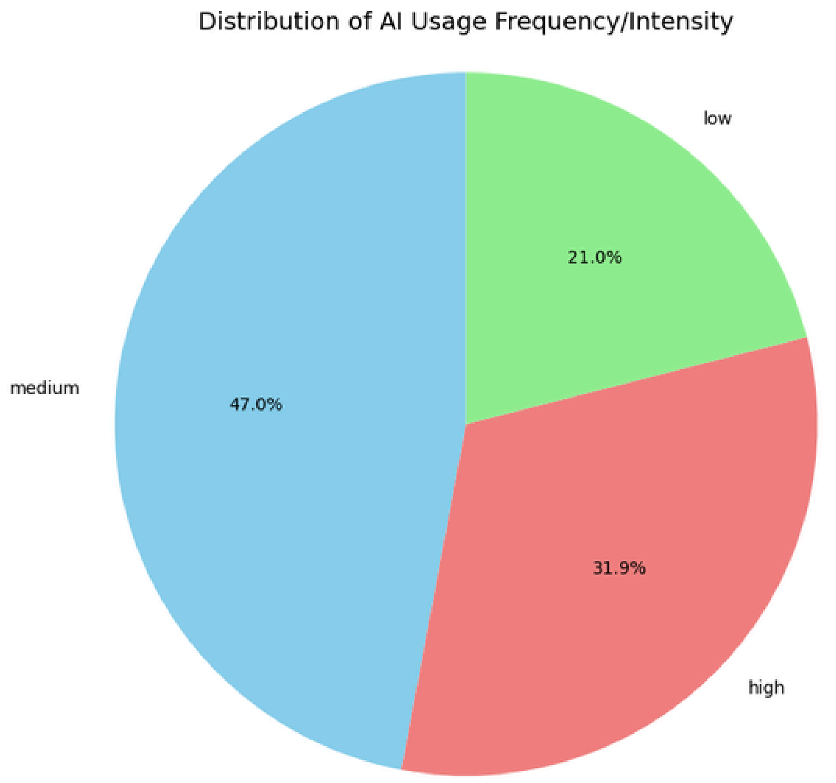




Multivariate Analysis



Categorical Insights



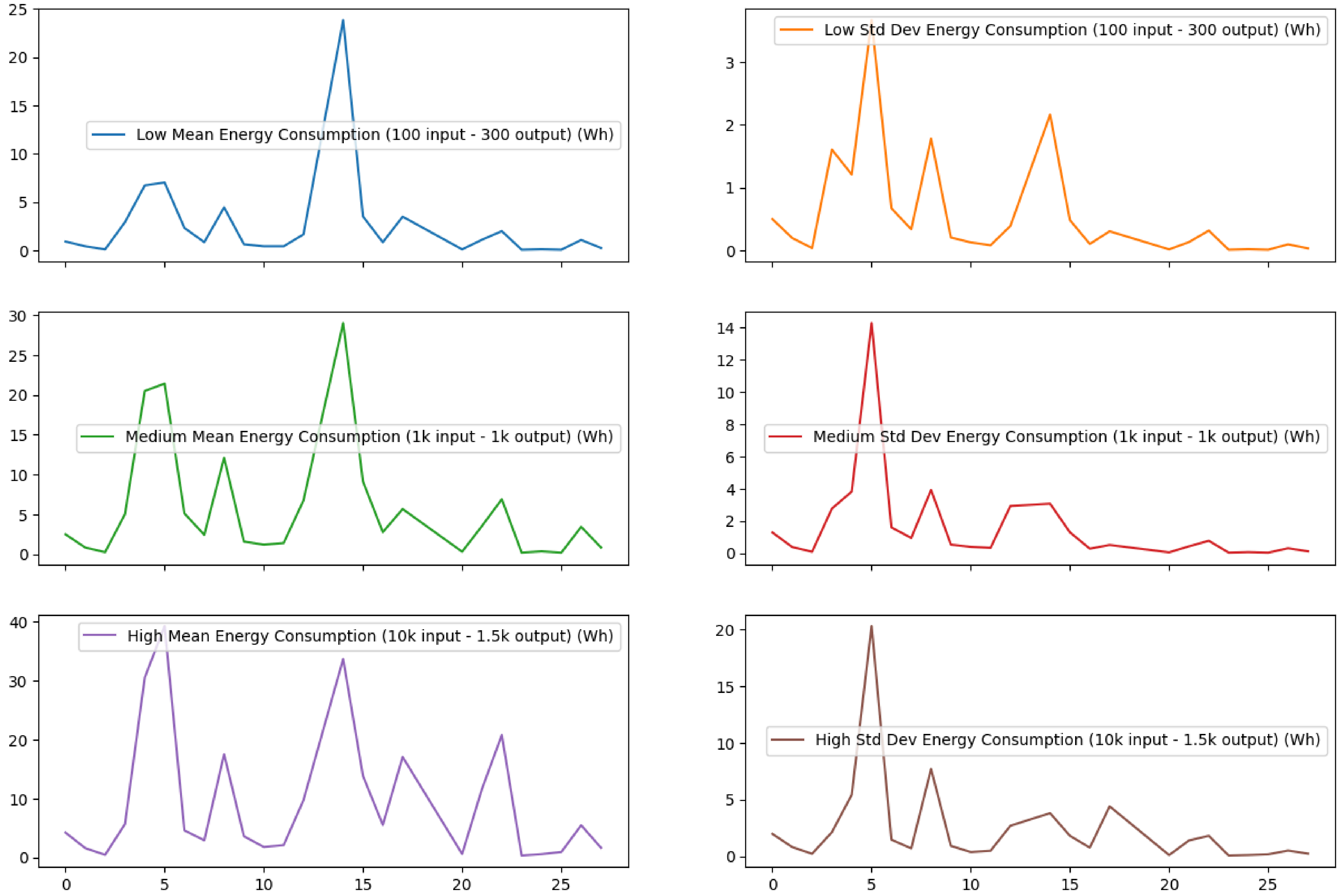
Low Use = 7 min avg.

Medium = 29.5 min avg.

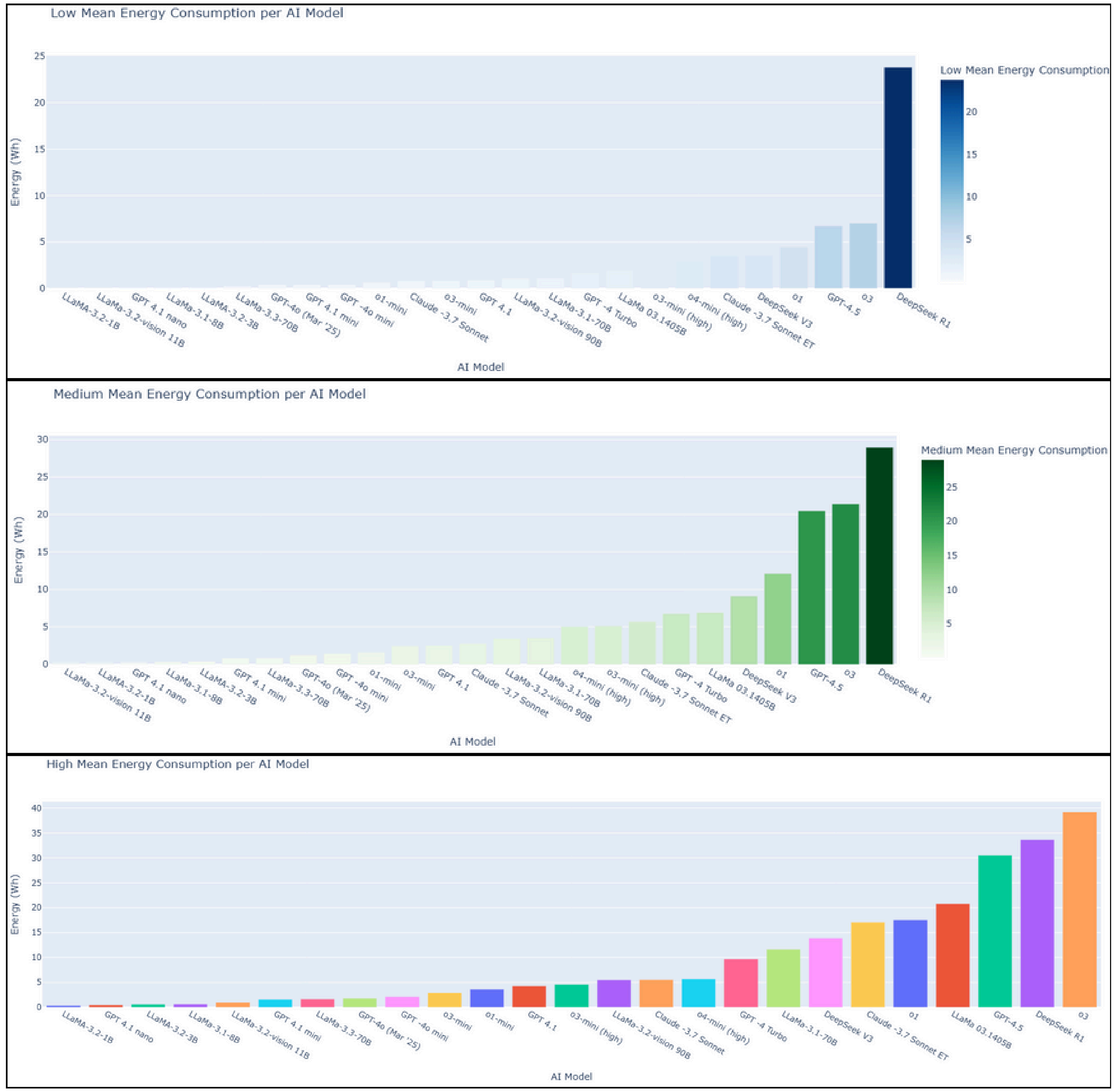
High = 120 min avg.

Visual Insights

Energy Consumption Trends



Mean Energy Per AI Model



Modeling & Evaluation

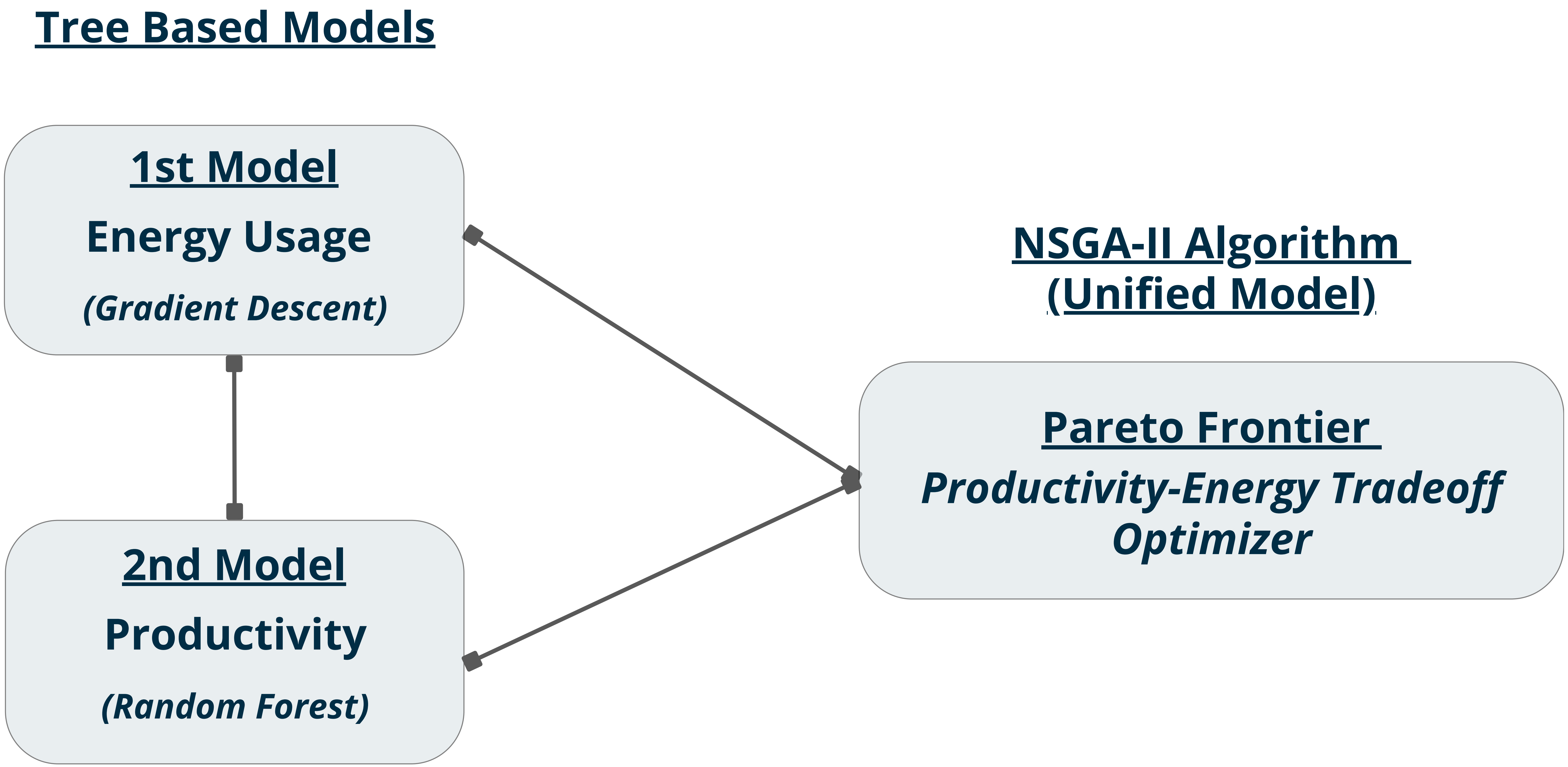
Overview

- Modeling
  - Data Transformation
  - Development Framework
  - Feature Selection
- Evaluation
  - Model Performance
  - Model Efficiency Insights

Data Transformation

I. Merged Dataset (Post-Cleaning)	• Energy Data (4 columns): Includes low, medium, and high median energy consumption (Wh), covering 28 AI model versions. • Productivity Data (2 columns): Intensity level and average minutes saved per task. • Claude Data (1 column): Workforce occupation type.
II. Weighting & Task Calculations	• Normalized worker weights based on conversation and usage percentages • Distributed 1,000 weekly tasks across occupation types • Computed weekly energy use (Wh) = Tasks × Mean energy per model × Occupation share
III. Added Calculated Features	• minutes_per_task: Mean time saved per task • intensity_share: Workload mix by AI usage level • occ_weight: Claude occupation weighting factor

Development Framework



Feature Selection

Model	Goal	Target Predictor (Y)	Selected Features (X)	Calculated Features (X)
Energy Consumption Linear Regression	Predict AI Energy Use, measured in Watts/hr	weekly_wh (weekly AI energy consumption in Watts/hr)	AI Model Type, Intensity Level, Occupation Type	intensity_share, occ_weight
Productivity Linear Regression	Predict AI Productivity in the workplace within various job roles.	estimated_time_saved (average minutes saved per task)	Intensity Level, AI Model Type, Occupation Type	intensity_share, minutes_per_task, occ_weight
Pareto Frontier	Find an optimized tradeoff balance between AI energy and productivity in the workplace.	Tradeoff between weekly_wh and minutes_per_task	Outputs from both regressions (weekly_wh, minutes_per_task)	Weighted combination of energy and productivity outputs to identify optimal efficiency points

Model Performance: Post-Outlier Filtering (Daily - Fix)

Energy Model

Data Groups	Results
Root Mean Square Error (RMSE)	13.73
Mean Absolute Error (MAE)	9.47
Weighted Avg. Percentage Error (WAPE)	7.95%
Mean Absolute Percentage Error (MAPE)	11.78%
R-Squared (R2)	0.99
Model Accuracy w/ 20% Error Margin	86.27%

Removed 83 Outliers (24.6% of Test Points).

Productivity Model

Data Groups	Results
Root Mean Square Error (RMSE)	1.7653
Mean Absolute Error (MAE)	1.10 minutes
R-Squared (R2)	0.9961

Model Performance: Post-Outlier Filtering

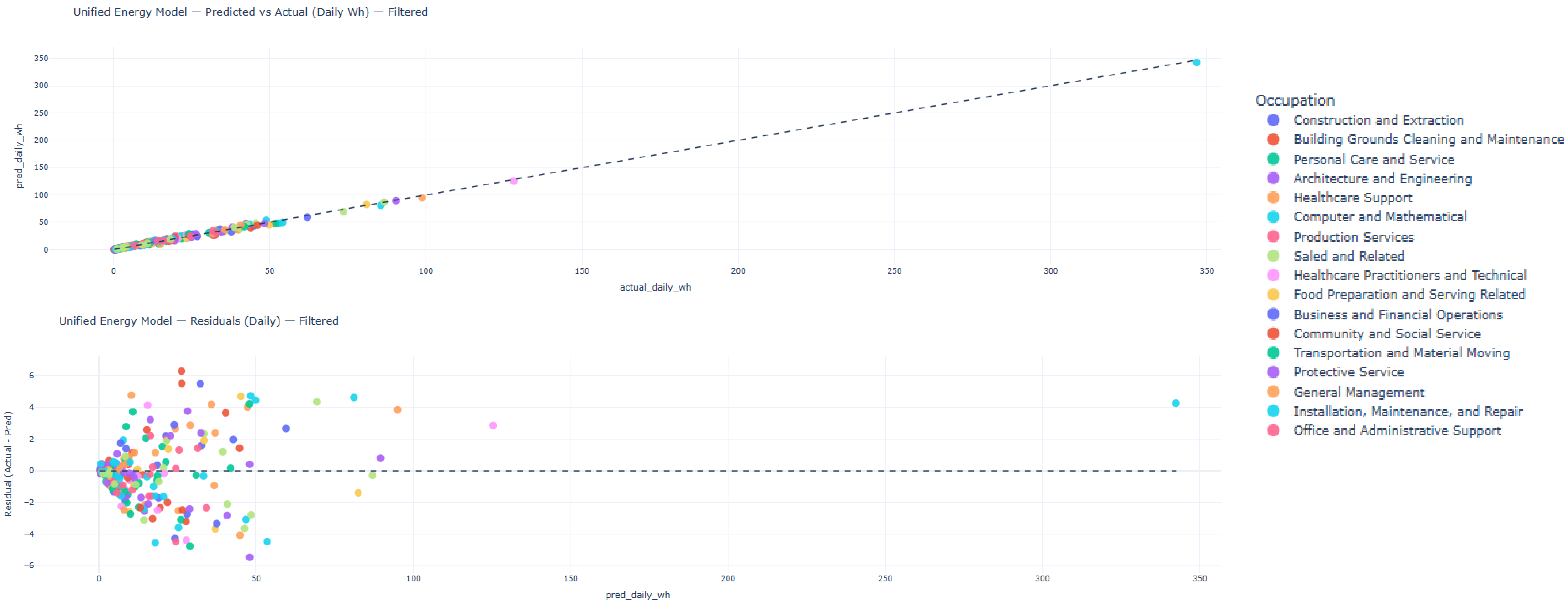
Occupation

- Construction and Extraction
- Building Grounds Cleaning and Maintenance
- Personal Care and Service
- Architecture and Engineering
- Healthcare Support
- Computer and Mathematical
- Production Services
- Saled and Related
- Healthcare Practitioners and Technical
- Food Preparation and Serving Related
- Business and Financial Operations
- Community and Social Service
- Transportation and Material Moving
- Protective Service
- General Management
- Installation, Maintenance, and Repair
- Office and Administrative Support

MODEL PERFORMANCE — ENERGY VS TIME SAVED

ai_model	weekly_energy_wh	weekly_minutes_saved
LLaMA-3.2-1B	22.648439	3856.760002
GPT 4.1 nano	29.411998	3833.156769
LLaMa-3.1-8B	36.898577	3846.468736
LLaMA-3.2-3B	38.451135	3856.085559
LLaMa-3.2-vision 11B	41.512660	3833.835505
LLaMa-3.3-70B	98.071411	3862.121343
GPT 4.1 mini	99.470414	3877.191159
GPT-4o (Mar '25)	123.059213	3876.326331
GPT -4o mini	142.811155	3876.898718
o1-mini	203.559909	3888.929860
o3-mini	226.233189	3924.065478
GPT 4.1	272.694260	3938.083989
Claude -3.7 Sonnet	324.611602	3963.307368
LLaMa-3.2-vision 90B	359.478317	4017.498944
o3-mini (high)	436.764184	4126.059174
o4-mini (high)	479.293635	4052.578008
LLaMa-3.1-70B	561.841444	3913.339848
GPT -4 Turbo	663.324478	4232.322450
Claude -3.7 Sonnet ET	885.010677	4100.226117
DeepSeek V3	945.234058	4418.282511
LLaMa 03.1405B	1029.805075	4265.301999
o1	1221.090052	4597.026479
GPT-4.5	2079.552243	5139.695515
o3	2407.623947	5216.469347
DeepSeek R1	2938.978211	5868.588331

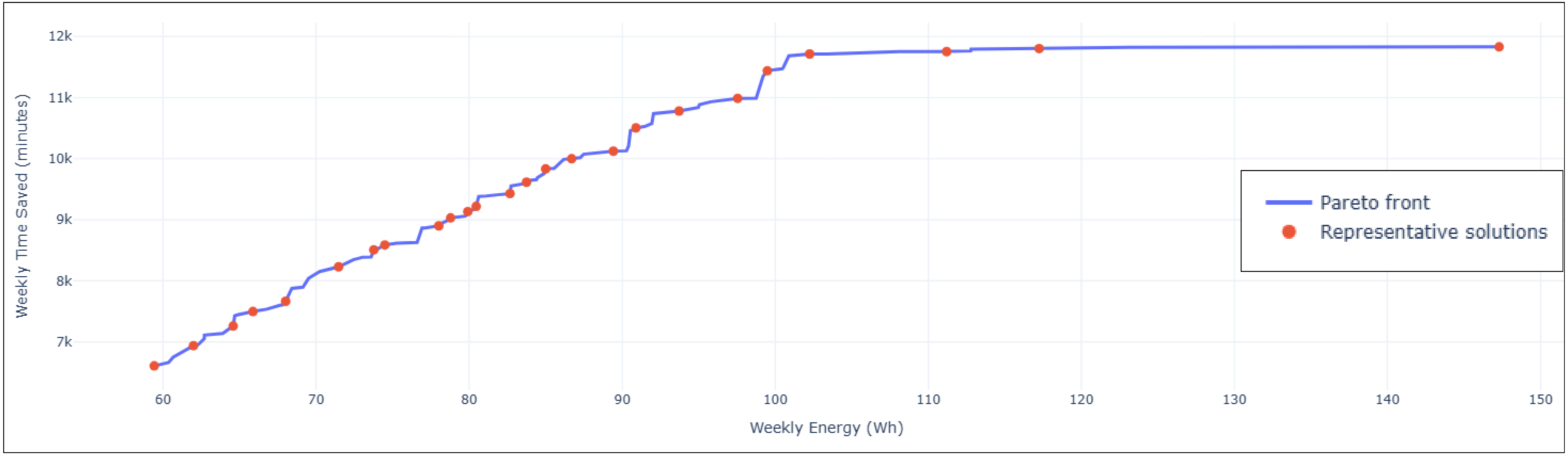
Model Performance: Post-Outlier Filtering



Pareto Frontier: Key Energy Models

Data Groups	Lowest Energy	Highest Productivity	Balanced Solution
Weekly Energy	111.1 Wh	714.9 Wh	193.9 Wh
Weekly Time Saved	5,267.6 minutes	9,232.4 minutes	8,584.1 minutes
Efficiency	46.41 minutes per Wh	12.914 minutes per Wh	44.26 minutes per Wh
Task Benchmark	100.00	100.00	100.00
Top 5 Results (model, intensity) Measured in tasks/week	<u>Medium</u> LLAMA-3.2-1B: 4.9 LLAMA-3.1-8B: 4.9 <u>High</u> LLAMA-3.2-3B: 4.9 LLAMA-3.2-vision 11B: 4.9 LLAMA-3.1-8B: 4.9	<u>High</u> LLAMA-3.2-vision 11B: 5.5 Deep Seek R1: 5.5 GPT 4.1 nano: 5.5 o3-mini: 5.5 o3-mini (high): 5.4	<u>High</u> LLAMA-3.2-1B: 8.5 GPT 4.1 nano: 8.5 LLAMA-3.2-vision 11B: 8.4 GPT 4.1 mini: 8.3 LLAMA-3.1-8B: 8.3

NSGA-II Pareto Frontier: Energy-Productivity Tradeoff (Weekly, 100 Tasks)

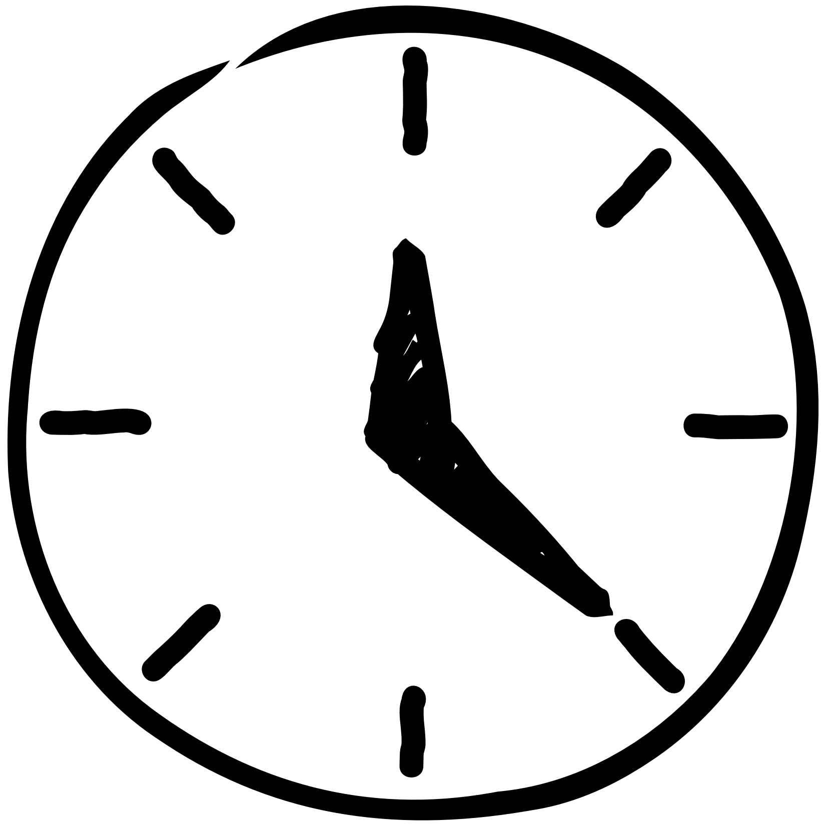


Model Efficiency Insights

Model Name	Strengths	Areas of Improvement
Pareto Frontier (Unified Model)	• Great accuracy Post-Outlier Filtering (R-squared = 0.995). • Low error metrics (MAPE/WAPE = 7 to 8%) across both daily and weekly predictions. • Removed 83 outlier points that were greatly skewed, leading to major improvements in model performance.	• Sensitivity to Outliers during the training phase of linear regression models. • Limited Generalization on Noisy Data

Conclusion

What We Learned



More Productivity Results in Higher Energy Use
More capable models can support complex tasks, but they often consume more energy to achieve higher performance.

Efficiency Varies Across Different AI Models
Context length, workload type, and efficiency strongly affect workplace productivity outcomes with AI.

Balancing Energy and Performance is Key
Choosing the right model depends on the task’s required quality and cost constraints.

Recommendations

Expand Model Coverage

Increase the number of AI models used in training to 50+ models to improve the linear regression model’s performance and Pareto Frontier results.

Enhance Feature Depth

Include additional features (i.e., compute efficiency, company size by total employees, and AI workplace dependency). This will provide better insights into the relationship between AI energy consumption and workplace productivity.

Potential Next Steps

Collect Additional Model Data

Gather energy and task performance metrics for newer AI models to broaden the dataset.

Utilize the Pareto Frontier to Cut Costs

Use the energy–productivity tradeoff optimizer to pinpoint where AI systems can deliver maximum productivity at minimal energy cost.

Identify Energy Limitation Use Thresholds to Cut Waste

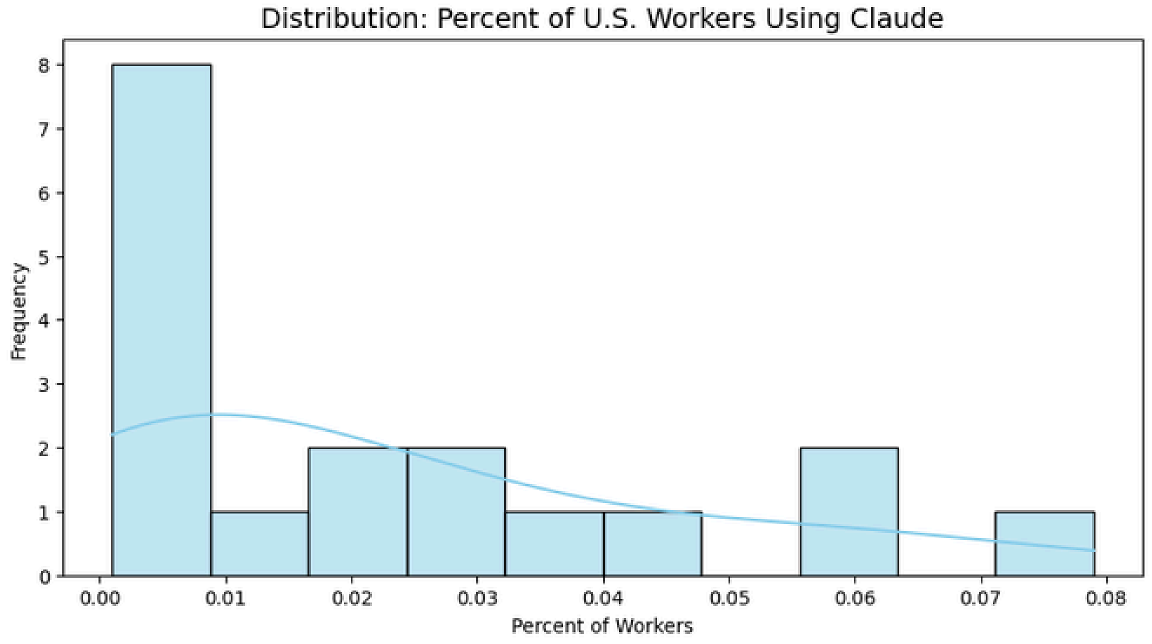
Leverage model predictions to detect when energy use becomes wasteful, reducing unnecessary consumption and promoting sustainability through responsible AI utilization.

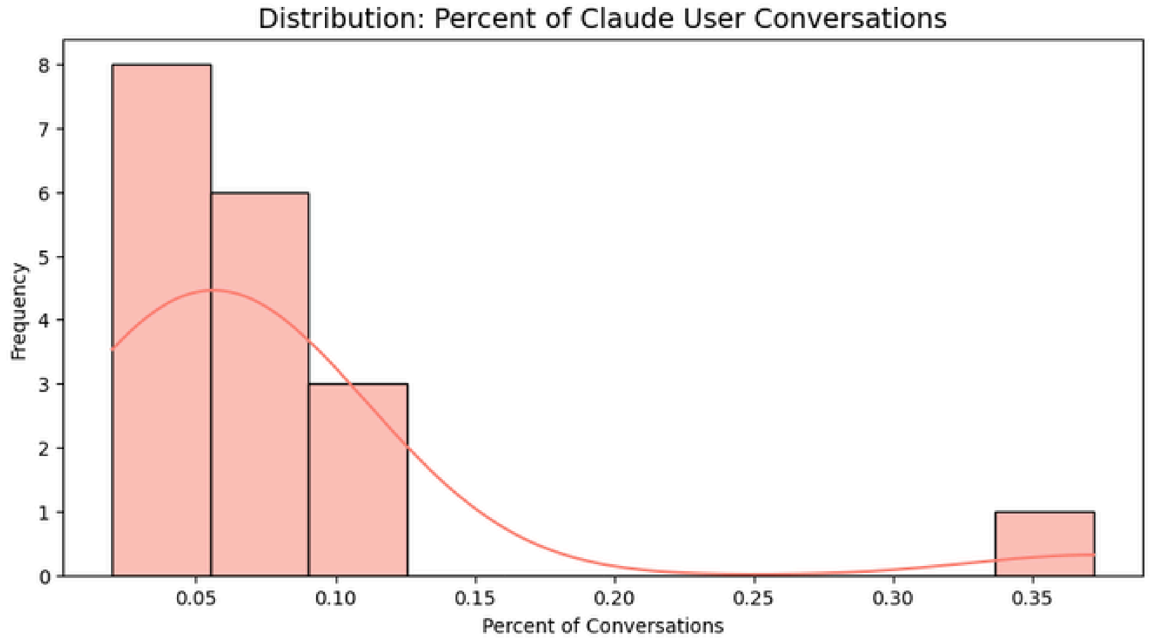
Thank you!

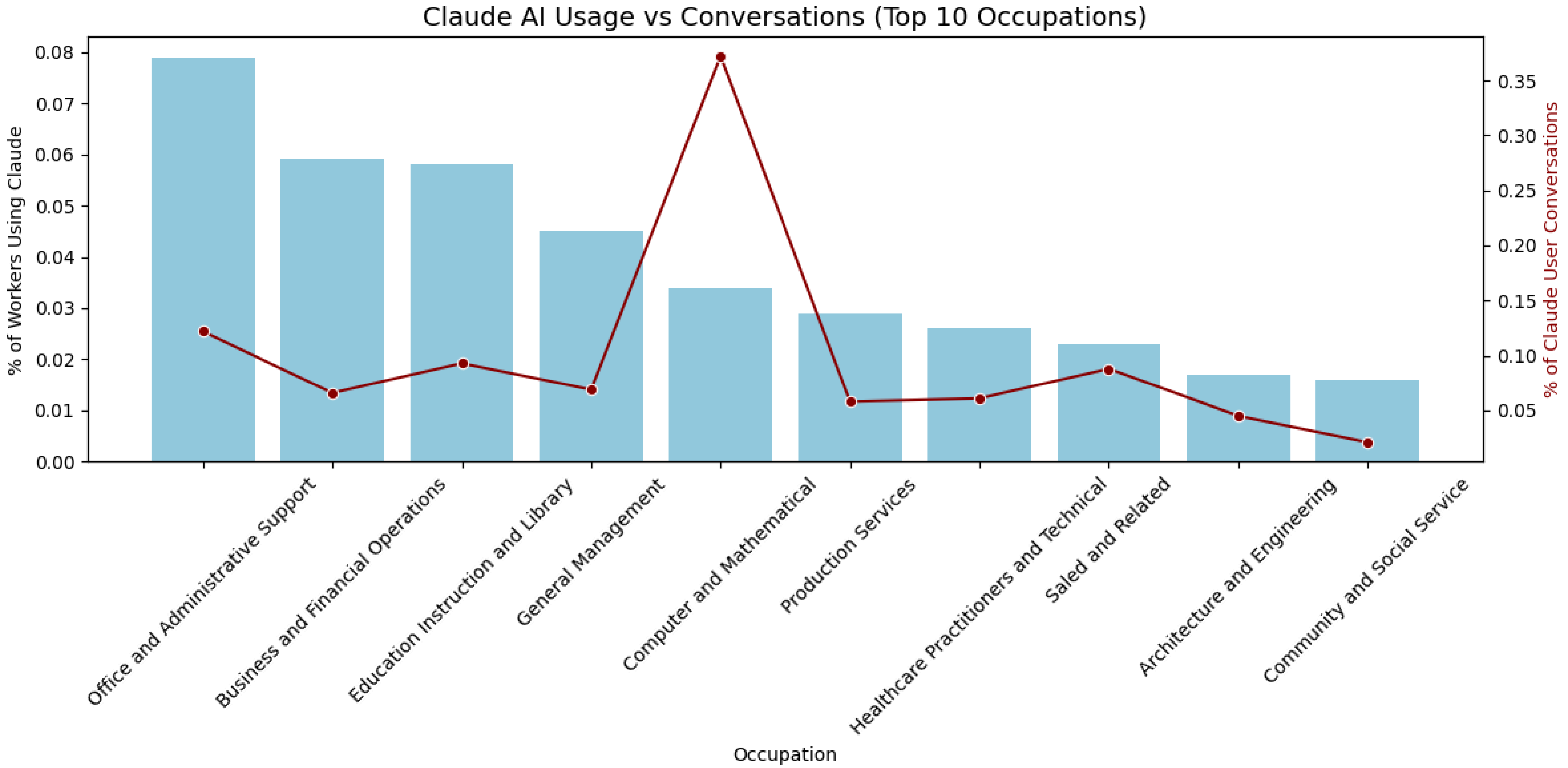
Questions?

APPENDIX

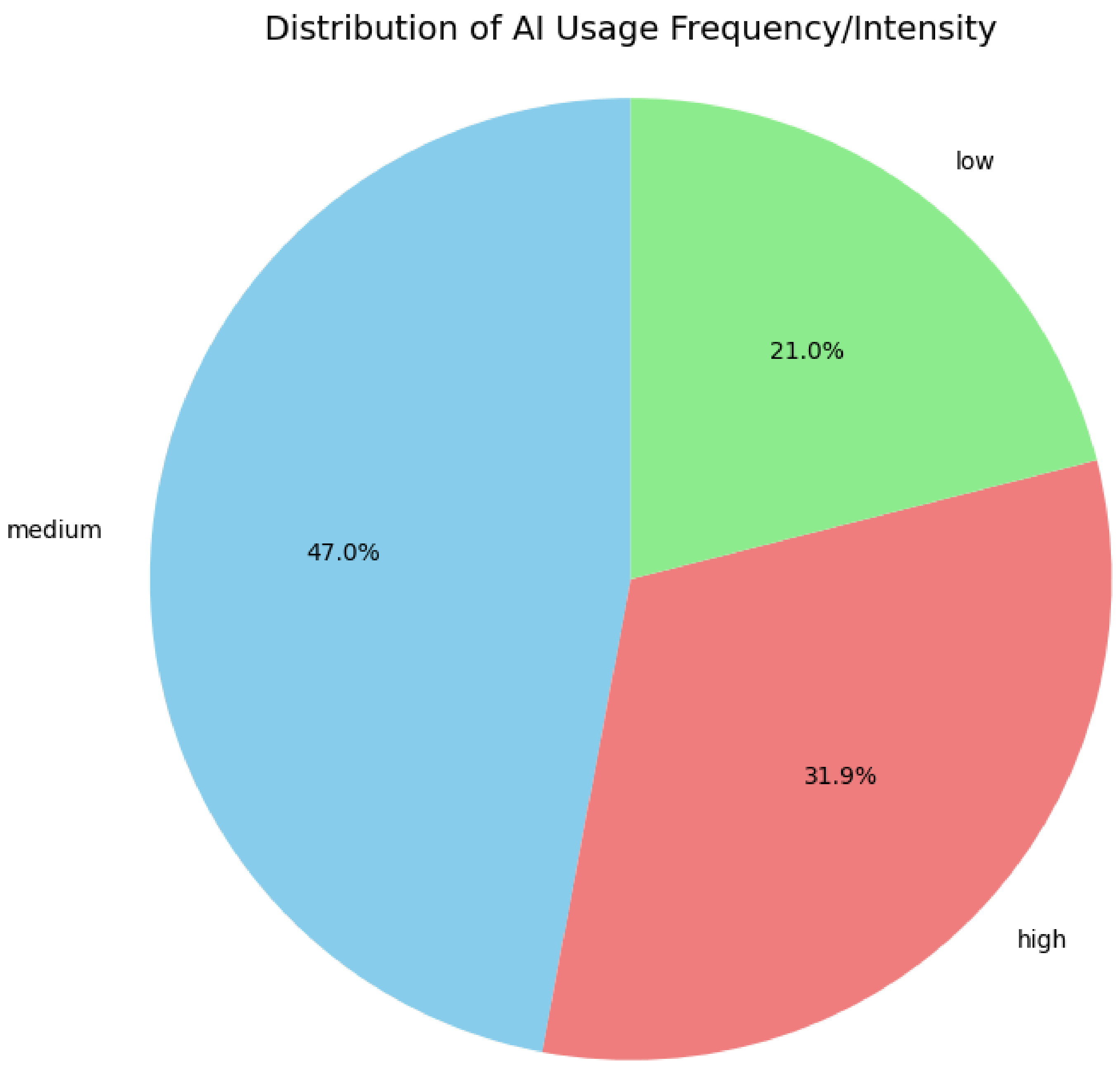
Claude AI Work Usage







Categorical Insights

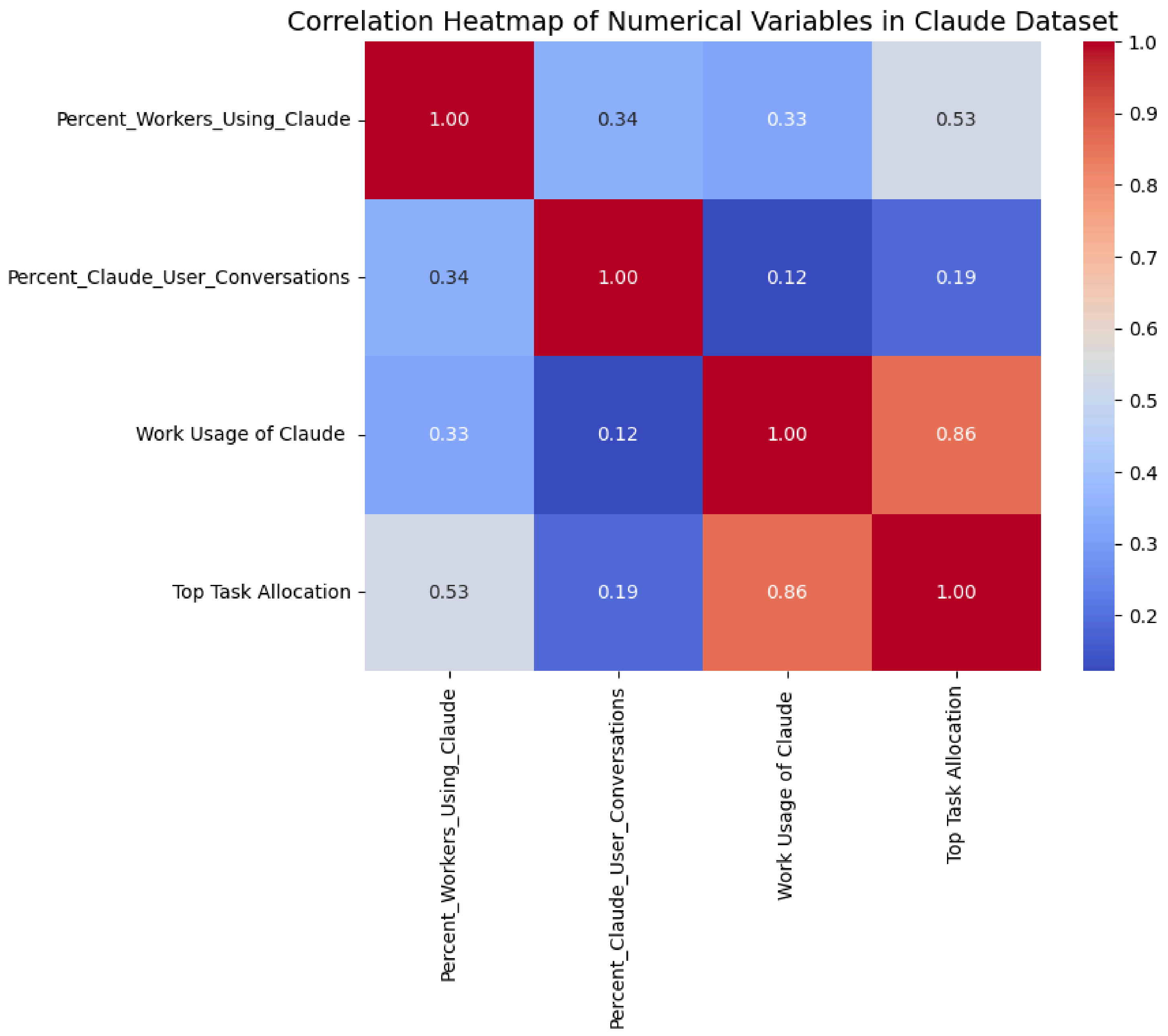


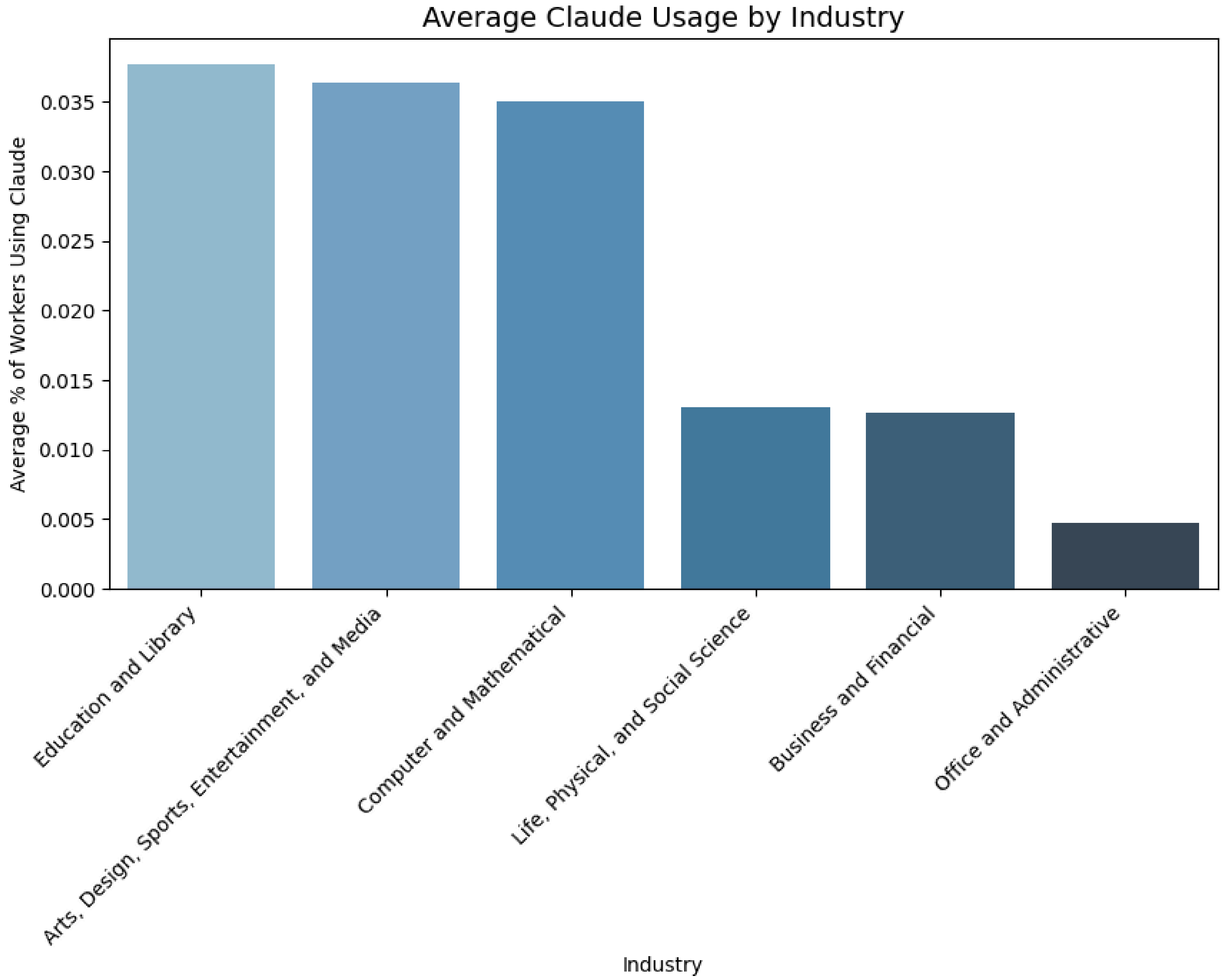
Low Use = 7 min avg.

Medium = 29.5 min avg.

High = 120 min avg.

Claude AI Work Results





Data Transformation (Detailed Ver.)

Merged Dataset (Post-Data Cleaning)

Energy Dataset (4 Columns)
- Low, Medium, and High Median Energy Consumption (Wh)
- AI Model (28 total versions).

Productivity Dataset (2 Columns)
- Intensity Level
- Average Minutes Saved

Claude Dataset (1 Column)
- Workforce Occupation Type

Data Transformation (Detailed Ver.)

Weighting & Task Calculations Based on Occupation

- Normalized worker weights from conversation and usage percentages.
- Calculated weekly task distributions across 1,000 tasks.
- Computed weekly_wh (watts/hr) = tasks × mean energy per model × occupation share.

Added Features (Calculated)

- minutes_per_task: mean time saved per task
- intensity_share: workload mix by AI usage level
- occ_weight: claude occupation weights

Data Sources (Detailed Version)

Dataset 1: Claude AI Workplace Usage

Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations

Academic Paper – Multiple Publishers

Random Sampling of 2.8 million Claude AI conversations from November 2024 to December 2024.

Dataset 2: AI Energy and Carbon Footprint

How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference

Academic Paper – Multiple Publishers

Gen AI Energy consumption per query from models published from March 2023 to April 2025.

Dataset 3: AI Productivity *(Time Saved Using AI)*

The Impact of Generative AI on Work Productivity

Published by the Federal Reserve

Employee Productivity Using AI, measured with time saved on work tasks. This source is being frequently updated.

Value Proposition & Potential Clients

Value Proposition

Strategic Focus Areas

Energy Monitoring & Data Centers

Work Productivity Using AI

Industry Value Reports

The Global Data Center Market was valued at $347.64 billion in 2024, with a projected annual growth rate of 11% by 2034.

(Yahoo Finance, 2025) [Link]

Digital workplace applications are expected to include embedded AI-driven algorithms by over 20% to enable greater workforce productivity

(Gartner, 2025) [Link]

Business Value for Our Clients

- Energy efficiency can be balanced with computational performance, supporting sustainable business operations while reducing carbon footprint.
- Save companies millions in costs as they continue to scale and use AI/LLMs, and Gen AI enterprise systems.

- AI-embedded work systems are transforming how we work by boosting the efficiency of time spent on tasks and promoting collaboration.
- Productivity is expected to continue to increase as we develop more intelligent systems.

Potential Clients

Cloud Services & Data Center Providers

aws

ORACLE

BROADCOM

EQUINIX

Financial Services & Consulting Firms

JPMorgan Chase & Co.

WELLS FARGO

CITADEL | Securities

TWO SIGMA

accenture

Defense, Energy & Y-Combinator Startups for AI Enterprise

DEPARTMENT OF DEFENSE · UNITED STATES OF AMERICA

U.S. AIR FORCE

Shell

conEdison

Julius

YOUR AI DATA ANALYST

Schneider Electric

cohere

Production & Development Overview

Timeline to Value	• 14 to 16 weeks to prototype. • 8 to 12 months to production. • Feasibility → Medium
Success Metrics	• Pareto frontier identification (3 to 5 solutions) • Productivity prediction accuracy above 75%. • Energy cost modeling within 25%
Strategic Priority	• Productivity-Energy Tradeoff Optimizer
Model Selection	• Markov Chain Monte Carlo (MCMC) method with a Bayesian Linear Regression • NSGA-II Algorithm applied for productivity & energy trade-off optimization

Problem Domain

Rising Energy Demand
As AI systems scale across industries, their energy consumption has grown exponentially.

High Operational Costs
Hosting and running AI models in data centers now drives significant energy expenses.

Need for Optimization
Organizations need a way to effectively measure and balance AI's energy use with workplace productivity gains to maintain cost-effective, eco-efficient, and responsible AI adoption.

Hypothesis Question

Can we optimize the tradeoff between energy usage and employee productivity with AI?

Beneficial Solutions

Optimize Energy Efficiency to Avoid Waste
Help organizations to reduce unnecessary energy usage while maintaining current AI productivity levels.

Improve Cost Management for Energy Spending
Predict operational expenses of AI usage in the workplace to budget corporate spending.

Promote Sustainable & Responsible Innovation
Supports the development of eco-efficient AI systems that align with global sustainability initiatives and corporate ESG strategies.

NSGA-II Recommended Solutions (Appendix)

```
  Weekly energy    : 111.1 Wh
  Weekly time save: 5,267.6 minutes
  Efficiency       : 47.411 minutes per Wh
  Sum of tasks     : 100.00 (should be ~100)
  Top 5 (model, intensity) allocations:
    - LLaMA-3.2-1B @ Medium: 4.9 tasks/week
    - LLaMA-3.2-3B @ High: 4.9 tasks/week
    - LLaMa-3.2-vision 11B @ High: 4.9 tasks/week
    - LLaMa-3.1-8B @ Medium: 4.9 tasks/week
    - LLaMa-3.1-8B @ High: 4.9 tasks/week
```

```
Highest productivity solution
  Weekly energy    : 714.9 Wh
  Weekly time save: 9,232.4 minutes
  Efficiency       : 12.914 minutes per Wh
  Sum of tasks     : 100.00 (should be ~100)
  Top 5 (model, intensity) allocations:
    - LLaMa-3.2-vision 11B @ High: 5.5 tasks/week
    - DeepSeek R1 @ High: 5.5 tasks/week
    - GPT 4.1 nano @ High: 5.5 tasks/week
    - o3-mini  @ High: 5.5 tasks/week
    - o3-mini (high) @ High: 5.4 tasks/week
```

```
Balanced solution (energy vs time)
  Weekly energy    : 193.9 Wh
  Weekly time save: 8,584.1 minutes
  Efficiency       : 44.265 minutes per Wh
  Sum of tasks     : 100.00 (should be ~100)
  Top 5 (model, intensity) allocations:
    - LLaMA-3.2-1B @ High: 8.5 tasks/week
    - GPT 4.1 nano @ High: 8.5 tasks/week
    - LLaMa-3.2-vision 11B @ High: 8.4 tasks/week
    - GPT 4.1 mini @ High: 8.3 tasks/week
    - LLaMa-3.1-8B @ High: 8.3 tasks/week
```

SLIDE DESIGN TEMPLATES