JOB SILL DEMAND
TREND ANALYSIS & SHORT -HORIZON FORECASTING

Name: Kirti Gupta
Course; Minor in AI
Date: August 2025

ABSTRACT

This project analyzes job skill demand trends using LinkedIn job postings. The pipeline includes data cleaning, salary harmonization, time-series construction, baseline forecasting (Naive, MA, SES, ARIMA), Prophet modeling, and a compact classification task predicting short-term skill demand shifts. Results are exported as reproducible artifacts, including metrics, forecasts, and polished plots.

DATA & PRE - PROCESSING

Source: postings.csv, jobs/job_skills.csv, mappings/skills.csv from **Kaggle**: linkedin-job-postings.
Cleaning included deduplication by job_id, filling missing values, clipping outliers, salary harmonization (annual USD conversion), and log-transform for salary EDA. Skills were joined in long format. Time series were constructed as counts of unique job IDs/day. Features engineered included rolling means, differences, z-scores, and MA ratios.

METHODS

Baselines: Naive, MA(k), SES, ARIMA(1,1,1) with MAE/RMSE/sMAPE/MASE metrics.
Prophet: daily models with weekly seasonality, additive mode, and changepoint tuning. Weekly Prophet attempted where possible.
Classification: Logistic Regression pipeline with accuracy-oriented threshold tuning and fallback walk-forward CV when hold-out was tiny.

RESULTS

Baselines vs Prophet: SES often outperforms Prophet on small, noisy series (e.g., Engineering: SES MAE≈1104 vs Prophet≈1145).
Adaptive CV: SES/MA(7) robust across multiple skills; ARIMA/Naive weaker.
Prophet daily: Writing/Editing MAE≈123, Quality Assurance≈150, Accounting≈471, Engineering≈1145, IT≈2298.
Classification: Engineering achieved ≈0.67 accuracy (walk-forward CV) with tuned thresholds.

DISCUSSION

SES and MA(7) provide reliable baselines. Prophet is competitive on steadier skills but less effective on volatile ones. Classification is limited by data sparsity but demonstrates threshold tuning.
Limitations: sparse series, weekly nonzero scarcity, noisy classification labels.
Next Steps: aggregate to weekly series, enrich features (holidays, regions), pooled/hierarchical models, and calibrated classification.

REPRODUCIBILITY & ASSETS

Artifacts: metrics_multi_skill_D.csv, advanced_adaptive_results.csv, prophet_daily/weekly metrics & forecasts, daily_skill_features.csv, engineering-focused comparison CSVs and plots, and PROJECT_CARD.json.
Notebook: end-to-end pipeline saving outputs to artifacts folder.

CONCLUSION

Short-horizon skill demand forecasting shows SES/MA(7) as robust benchmarks. Prophet is useful with richer signals.
Classification accuracy is constrained but improved with threshold tuning.
All results are reproducible with saved artifacts and documented pipeline.